DESIGN AND DEVELOPMENT OF IOT BASED ASSISTIVE DEVICE FOR NAVIGATIONAL REHABILITATION OF VISUALLY IMPAIRED PERSON

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Electronic and Communication Engineering

By

Nitin Kumar

Registration Number: 41800318

Supervised By

Dr. Anuj Jain (22631)

Electronic and Communication Engineering (Professor)

Lovely Professional University

LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2022

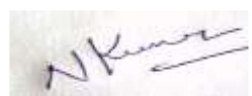
ACKNOWLEDGEMENT

I thank the Lord Almighty and my family members, whose showers of blessings have made this, project a reality.

I express my sincere thanks and deep sense of gratitude to my guide, **Dr. Anuj Jain,** Professor, Department of Robotics and Control and Technology, for his valuable support, suggestions, guidance and encouragement. And also, I thank him for providing me with the counsel and direction to help me complete this project.

My sincere thanks to **Dr. Manoj Kumar Shukla**, Assistant Professor and Head, Department of Robotics and Control for his kind support and for providing necessary facilities to carry out the project work and prepare the thesis.

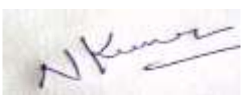
Last but not the least, I wish to record my sincere thanks to **Dr. Mahindra Swain** and other panel members for his valuable suggestions and critical reviews throughout the project.



Nitin Kumar
41800318

DECLARATION

I, hereby declared that the presented work in the thesis entitled "**Design and Development of IOT based Assistive device for Navigational Rehabilitation of Visually Impaired Person**" in fulfillment of degree of **Doctor of Philosophy (Ph. D.)** is outcome of research work carried out by me under the supervision of **Dr. Anuj Jain**, working as **Professor**, in the **School of Electronics and Electrical Engineering of Lovely Professional University, Punjab**, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.



Name of the scholar: Nitin Kumar
Registration No.: 41800318
Department/school: School of Electronics and Electrical Engineering
Lovely Professional University,
Punjab, India

CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled **Design and Development of IOT based Assistive device for Navigational Rehabilitation of Visually Impaired Person**" submitted in fulfillment of the requirement for the reward of degree of **Doctor of Philosophy (Ph.D.)** in the **School of Electronics and Electrical Engineering** , is a research work carried out by **Nitin Kumar , 41800318**, is bonafide record of his/her original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

(Signature of Supervisor)
Name of supervisor: Dr. Anuj Jain
Designation: Professor
Department/school: School of Electronics and Electrical Engineering
University: Lovely Professional University, Punjab, India

List of Figures

List of Tables

List of Abbreviations

SSD	Single Shot Detector
OCR	Optical Character Recognition
YOLO	You Only Look Once
TTS	Text to Speech
mAP	Mean Average Precision
VSLAM	Visual Simultaneous Localization and Mapping

Table of Contents

Abstract

This report, proposes a new approach in developing a deep learning-based prototyping model which can assist blind and visually disabled person to recognize their environments and navigate through them.

In the recent decades, the development of navigational devices has posed challenges for researchers to design smart guidance systems for visually impaired and blind individuals in navigating through known or unknown environments. Efforts need to be made to analyze the existing research from an historical perspective. Early studies of electronic travel aids should be integrated with the use of assistive technology based artificial vision models for visually impaired persons.

In this study, we, proposes a new approach in developing a deep learning-based prototyping model which analyzes model losses, recall value, Precision, and mean Average Precision (mAP) of the model.

With our system, a platform may be provided that may help those with visual impairments improve their quality of life and mobility. A higher level of participant confidence, accurate steering assistance with the environment may explain the increase their speed. Furthermore, the system can guide participants to key objects in an indoor and outdoor environment, and navigate a sequence of challenges both inside and outside.

CHAPTER 1

INTRODUCTION

Visual function is important for everyday life, which is not available to blind people. In the absence of vision, blind people are limited in their autonomy. A growing number of assistive systems are being developed using computer vision and machine learning. This work presents an assistive system based on visual recognition to compensate for visual impairment. To detect objects in the path of the visually impaired person on YOLOV5 networks, a pre-trained custom model was created with COCO and custom datasets. In comparison to existing techniques like SSD, and Faster RCNN, the model performs well, according to experimental results. Despite a dynamic change in the environment of the visually impaired person, the application maintains a reasonable performance level. A discussion of the system's limitations can be found in the conclusion section.

1. 1 Overview

In 2018, the World Health Organization estimated that there were 1.3 billion visually impaired people, 36 million of whom were blind. The number of people with vision impairments is expected to reach 2.2 billion by 2019 and 9.7 billion by 2030[109]. As per these statistics, there is an increasing number of blind people and visually impaired people every year. An individual who is visually impaired or blind typically finds it difficult to engage in day-to-day activities without hazel. Traveling through unfamiliar locations without close companionship becomes more difficult when it involves visiting unfamiliar locations. The high cost of guide dogs makes it difficult to obtain one to assist visually impaired individuals. Furthermore, traveling with no assistance in familiar surroundings could also pose a challenge because blind people cannot anticipate dynamic situations beforehand and cannot respond to them in real-time. As a result, it is not possible to take precautions to avoid such obstacles in the same way as someone with a good vision would. Consequently, such obstacles cannot be avoided similarly to those faced by a normal person with good vision.

As a result, it is not possible to take precautions to avoid such obstacles in the same way as someone with a good vision would.

Inn order visually impaired people to perform daily tasks like walking, they need assistance. To avoid obstacles, they need assistance to identify them and avoid them, especially when it comes to navigating securely. Currently, researchers are working on a support assistant for the visually impaired to address this problem.

The effectiveness of options such as white canes and seeing-eye dogs is limited, but they can be very helpful in some circumstances. Obstacles can be detected by a cane below the knee, but they can't be detected above the knee, and their range is limited. Visually impaired people can also benefit from seeing-eye dogs, but they need intensive training which can be difficult.

Sensors are embedded in walking assistants to improve the accuracy of vision devices in today's world. It is more accurate to detect obstacles with computer vision than with sensor-based traveling aids. With the assistance of computer vision, it is possible to detect and categorize obstacles that visually impaired people may encounter, and this can be extended to assist them in recognizing objects on their own.

The objects are categorized using computer vision-based object detection. An object recognition system based on deep learning, called YOLO - You Only Look Once, lets you spot objects in real-time with a camera. The YOLO algorithm uses a darknet, an adaptation of a convolutional neural network. By using YOLO, a regression-based model, the algorithm predicts classes in a single run for the whole image, thus speeding up computation.

Visually impaired people can also use voice guidance technologies to identify objects, which can be heard through the system. Visually impaired individuals can utilize this technology to learn what they cannot see. Through the proposed system the visually impaired can detect objects and be notified of their distance with a voice feedback system, which will act as a comfort for them and allow them to move independently.

1.2 MOTIVATION

Every aspect of our daily lives depends on vision. The human eye is used for many things, including navigating roads, finding objects and finding other people. However, blind people cannot see, which makes their lives quite challenging. In accordance with the latest WHO statistics, there are 2.2 billion people worldwide which suffer from vision impairment. It is predicted that by 2050 this number will increase threefold from its current level.

Approximately one thousand million people in the world experience vision impairment losses caused by eye diseases, such as cataracts, keratoconus (KC), or diabetic retinopathy. A resolution adopted in the 73rd assembly of the World Health Organization (WHO) in the year 2020, which aims to promote integrated, people-centered eye health care and prevention of blindness and vision impairment was passed in the year 2019. Until now, people with visual impairment have relied on low-cost traditional methods like white-cane, which is an effective touch-based technique but doesn't usually detect the shape of obstacles. It is also believed to be a low-cost navigational method if guided dogs are used, but dogs are less accurate due to their prospective mood while moving. It also uses sensor-based assistive devices like GPS a good alternative, but GPS is self-restrained in avoiding obstacles while on a mission. Electronic travel aids (ETA's) are also being used to guide people with visual impairments. One example of this is proposed in Electronic Mobility Cane (EMC) which constructs a logical map based on the surroundings. ETA’s have designed an IC tag system that is used for indoor walking which makes moving through a defined space easier. The floor color lanes were identified using an RFID tag and a one- chip processor. Deep convolutional neural networks (DCNN) for object detection reduce the amount of data and costs associated with hardware by using advanced technology.

Convolutional Neural Network (CNN) based algorithms are much more accurate and can be used both for previously known environments as well as unknown environments.

An assistive framework, such as the DEEP-SEE framework could be employed for distinguishing known faces from unknown ones. Scientists have developed systems with RFID tags, Bluetooth, ultrasonic, Wi-Fi, and a camera for assisting visually impaired people to navigate their environment. A powerful AI Navigation device for visually impaired people, such as NVIDIA Jetson TX2 embedded with DLSNF (Deep learning based Sensory Navigational Framework) provides an efficient device for navigation. In addition to detecting cracks in pavements, researchers have also discovered a way to detect currency. There are mobile apps that can help visually disabled people navigate a range of environments safely. With the assistance of acoustic waves, a person could also access information. YOLO algorithm can also be employed for checking the quality of objects, such as corn; the whole simulation was performed on NVIDIA TX2. By using tensor flow and the coco data set to train with 328K images, YOLO Object detection in the path of the user can be detected. Binaural sounds are generated via the HRTF (Head Related Transfer Function), which is used along with cameras to estimate and navigate efficiently in the environment.

Optical character recognition (OCR) and Text to speech (TTS) techniques were used to read text on the obstacles and smoothly navigate. To guide a user in map building, Simulation Localization and Mapping (SLAM) can be applied using sensors like RPLIDAR A2 and Kinect V1. The entire simulation can be run on Robot Operating System (ROS) with Turtlebot2. The validity of the YOLO algorithm has been verified with multiple data sets including COCO, VOC, and VisDrone.

Based on the above-mentioned research work, this research aims to

- Present a method to produce a module that feeds the user with vital data with greater accuracy, such as identifying individuals and detecting obstacles in the path.
- To facilitate practical implementation, a Deep Learning algorithm i.e. YOLO is proposed which is compatible with user-friendly hardware.

1.3 Evolution in Assistive Technology

There has been a significant increase in assistive technologies that promise to help blind people over the past few years. These technologies are based on Sensors, the Internet of Things, and Computer Vision. It is important to note that each of these systems has its advantages and limitations. Embedded systems like Raspberry Pi can be used to create portable assistive systems. In these systems, labels and text on packaged goods are recognized using computer vision. These systems can assist blind people not only in navigating from one place to another but also in shopping in supermarkets, reading textbooks, etc. These devices lacked to limited processing capabilities of these devices, there is also a concern about the power backup of these devices. In addition, these devices can read QR codes to identify objects, as they are compact and have built-in cameras. As part of this system, customers will also be able to shop in stores with QR codes on their products. This system would work if every product was labeled with a QR code, which is impossible to enforce. According to, blind persons can be assisted by identifying product labels using a mobile application. The OCR algorithm is used in this application, so it inherits both advantages and disadvantages. This application fails to detect product labels if they change in size, illumination, or scale.

The use of technological advancements in everyday products led to people taking advantage of those advantages in assistive tools as well. In order to help disabled people live a more fulfilling life, these tools are created as support for their daily activities. It was later termed Assistive Technologies to cover a wide range of such assistive tools. The term "assistance technology" refers to technologies, equipment, apparatus, services, processes, systems, processes and environmental modifications that assist people in living active, productive and independent lives as equal members of society through overcoming various physical, social, infrastructure, and accessibility barriers. People with disabilities are increasingly relying on assistive technologies, particularly in navigation, to assist with their daily lives. For example, Wayfindr 8 and Envision [15] are examples. Various possibilities for navigation systems are also being explored in light of the technological advances that have occurred in the mobile industry. As the de-facto standards for the implementation of assistive technologies, Csapóetal [16] reports that the largest and most common mobile platforms are rapidly evolving.

1.4 CURRENT DAY SCENARIO

It was around the years 2013-2014 that wearable technology began to become popular. Usually worn on the wrists and necks, these devices can be accessed remotely. Smart watches such as Fitbit, Pebble, and Google glasses are among the most popular wearable technology devices[113]. An example of such a method is Finger- Eye, which is proposed by [2]. In this system, the tip of each finger has an embedded camera that serves as an electronic wearable finger. Through the camera, the blind person is able to hear the word as it is scanned under their finger. However, the finger-eye test was only completed by using a table-mounted webcam and not a complete finger-eye test. There is a lot of interest in the concept of self-driving cars in the technology industry which has to be studied in order to assist blind people in their navigation.

A self-driving vehicle system based on event cameras is proposed in [10]. The output of an event camera is different from the output of a regular camera because it records the change in pixel intensity. The cameras act as motion detectors by detecting the movement of objects around them and removing redundant information.

1.5 PROBLEM STATEMENT

A primary objective of the model is to develop a real-time based system that would assist a visually impaired or blind person navigate through a known or unknown path and also allow VI or blind individuals to gain knowledge of their surroundings, avoid obstacles, recognize objects and navigate independently.

1.6 GAPS IDENTIFIED

1. Most of the research work does not training the data from the sensors, hence results in low accuracy in detection and recognition of objects [39].

2. To make navigation convenient for the visually impaired persons [12] by using lighter components throughout the model.

3. Design of the cost-effective and easy-handling device, will be in order to get easily access and affordable device for all [11].

4. To have low power consumption based device [9], the device can consume lowest power so that the device will be charged easily and VI person will be safe.

5. To design multi-environment way-finding device [8], such that the device could be used both for indoor as well as outdoor navigation.

1.7 OBJECTIVE

1. To identify the physical aspects of navigation to assist a visually impaired person.
2. Designing a Sensor-based model and analyzing the characteristics of the developed model.
3. Implementation and comparison with the existing models considering the future aspects of this technology.

1.8 ORGANIZATION OF THE REPORT

Chapter 1: The purpose of this chapter is to present the overall concept of the topic, the problem for which the study is conducted, and the research objectives.

Chapter 2: An overview of the literature is presented in this chapter, which is organized into four main sections: image segmentation, object detection, user interface, and user notification. A major focus of this thesis is on the user interface, so it is given priority over other topics.

Chapter 3: Research Methodology This study used a qualitative research methodology in its third chapter, which describes the research methods used in the study. A description of the different methods that are employed in prototype design will also be provided in this chapter, as will how the data will be collected for the data analysis, which is crucial to the conclusion of the thesis.

Chapter 4: The purpose of this chapter is to interpret and present the data gathered and the results of the study.

Chapter 5: A discussion and conclusion are drawn in the final chapter. There are recommendations and suggestions that are made for the future.

CHAPTER 2

LITERATURE SURVEY

Navigation plays a vital role in everyone's daily life. People use navigation for a variety of reasons, including work, education, and shopping. Most people recognize that vision is vital to navigation since it allows us to move from place to place. It is fairly easy to imagine how one would get around without vision in familiar environments, like our homes or offices. However, it is much more difficult to make your way through completely unfamiliar environments.

Being vision-impaired does not mean losing our freedom when it comes to getting to and fro wherever traversed. Even blind people or people with partial-sightedness can travel freely regularly which means that is most appropriate for them. According to Nicolas et al. [1], safely and efficiently navigating is one of the huge challenges to freedom for people with disabilities. For a safe and systematic navigation system, it is important to be aware that, it is better to provide travel skills to visually disabled persons to guide their vision [2]. However, there are still some challenges that visually impaired people must overcome during daily navigation [3]. There are several hazards to look out for, including pits in front of the lane, barriers on the path, steps, highway interchange, signposts, wet floors, oily or slippery walkways outdoors, etc. In the past, visually disabled people used sticks [2], guide dogs [3, 4], and volunteers [4, 5] as well as trained guides to navigate. People who become blind in their early lives frequently make use of their audial skills, such as echo-sounding, to navigate effectively [6-8]. It can significantly improve security for people on foot who require assistance in locating their exact location due to the presence of landmarks and clues close to public transportation stations. Orientation and mobility skills are important for individuals with vision impairments who want to navigate safely and efficiently [7]. An individual's mobility can be defined as their ability to move from one location to another efficiently and safely. A safe journey to a station includes navigating public transportation, crossing streets, and reaching a station without falling [10-12].

2.1 Various algorithms in navigation assistance for visually impaired persons

For object detection, the SIFT (Scale Invariant Feature Transform) algorithm is considered to be an image local feature description and one of the most popular algorithms that can extract key features from a frame. The solutions they developed for detecting and tracking objects in videos are presented. Using the improved k-means clustering algorithm, key features are grouped into groups that detect moving objects [23]. They used log-polar transform to stabilize the video [24]. They then compare the results of their proposed solution with the SIFT-ME algorithm to determine their affinity for transformation. Karami et al. [25] combined SURF and SIFT to detect objects and produce an object tracking algorithm that is robust. To speed up the identification of objects, SURF or Speeded up Robust Features was developed.A study of the SURF algorithm was performed by [25][26][27][28][29][30]. As part of this research, SURF performance enhancements are being investigated as well as comparative analyses with various algorithms such as SIFT, ORB, and BRIEF. By improving the SURF algorithm, illumination invariance and matching rates will be achieved.

One of the most widely used algorithms for identifying objects is optical character recognition (OCR). The main application of this technique has been to detect text in images. OCR algorithm can be used to identify objects according to the following works [31][32][33][34] discuss how further it could be used to accommodate more applications. Among these studies are experiments on food quality detection based on container label detection, manufacturing and expiry date detection, and retrieving individual information using ID card using the OCR-based model [35] provided a detailed description of their proposed technique for detecting elevator buttons in their paper. The purpose of this model is to assist assistive robots in navigating to their desired path As a result of combining OCR and Faster R-CNN, the authors came up with OCR-RCNN, a single neural network. In this case, compilation of custom dataset of elevator panels for applying to the OCR-RCNN algorithm is used several times.

Since its release, the algorithm such as YOLO (You Only Look Once) has become very popular. YOLO is a method of detecting objects that only need to be looked at once as its name suggests. The YOLO detector uses single shot detector, which means that instead of doing image classification and localization in two separate steps, it performs it in just one step. There is a minor degradation in accuracy resulting in speeding up the overall process of object recognition. In the past few years, the YOLO algorithm has undergone major upgrades: versions v2, v3, v4, and v5 while version 1 is still considered the original algorithm. Several studies are using YOLO to detect objects and compare them to FasterRCNN in [36][37] [38][40][41]. In these systems, a deep learning-based neural network is trained on a dataset to create a pre-trained model.

A vision-based system was proposed by [26] to assist visually impaired persons during navigation. In this system, there is a camera, a haptic feedback device, and an embedded computer to assist blind people in locating objects. As a wearable system, this system provides mobility to blind users. In this experiment, blind people wearing the system are made to walk through a maze while wearing it. Based on the results of the experiments, this system aids blind users in navigating a path without colliding with obstacles. However, navigation is slower using this system than when using a cane. In combination with a cane, this system can improve the speed of navigation for blind users. According to Liu et al. (2016) [34], an assistive system that uses the OCR algorithm can be used to help blind persons read. It consists of a glove that is equipped with an embedded camera index that is worn by the blind to assist in his navigation. The blind uses his index finger of the glove hand over the first sentence from left to right. As the picture of the text is processed by the camera beneath the finger, audio will be output. According to the experiment results, the finger can read the text in the image through the camera; however, the overall process is lagged as it creates a high-resolution image by combining several images of the same frame. Due to the experiments being conducted through a webcam mounted on a table for reading text, it is also necessary to evaluate the project's feasibility. Rajesh et al. (2017) [29] and Maolanon et al. (2018) [40] propose a model to aid blind users. An embedded computer is used by their system to process the text in images, which is handled by

Raspberry Pi. A comparison of Rajesh et al. (2017)[29] and Maolanon et al. (2018) uses the OCR algorithm, while a comparison of Maolanon et al. (2018) uses the YOLOv1 algorithm is shown [40]. It is possible to hear the text read out to blind people with the help of the Text to Speech (TTS) library. When it comes to reading text, the results of the experiment are impressive, but they do not perform well when it comes to recognizing faces. Its fast and accurate processing makes it the best in the industry, An accuracy of 83% can be achieved by YOLO's detection capabilities.

2.2 Assisting Blinds with Navigation Technologies

2.2.1 Vision-Based Imagery Systems

A variety of computer vision algorithms and optical sensors are used for visual navigation. To gather visual information about the environment, it uses different types of cameras. Through its use of visual features, this system detects obstacles and guides its users in navigating safely by providing directions. The integration of vision-based technologies into navigation systems has been attempted several times in the literature. These systems include stereo cameras, IP cameras with networks, and RGB-D cameras. Several noteworthy works in each category are also highlighted, as the works are organized according to the technology that captures visual images.

2.2.1.1 Stereo *Camera*

Using an intelligent assistant called Tyflos, a navigation method was presented in [61]. Visual camera systems are capable of capturing 3D data from an environment continuously or on demand, depending on the user's instructions. To establish connections with visually impaired users, the software converts these images into their verbal equivalents.

Using the navigation system proposed by [62], an inertial measurement unit (IMU) and earphones are integrated. Binaural rendering is an audio therapy technique that uses headphones to create sound from a detected object. When a detected object is detected at a specific position, it is converted into a sound source and transmitted to the user. sounds that can be localized both in distance and direction and can be used only in the outdoor surrounding.

2.2.1.2 Internet Protocol Camera

Computer vision algorithms may be used in remote processing systems and analyzed photos taken by IP cameras on the ceilings of each room in the Chaccour et al. system [63]. The user would be able to navigate safely to their destination by using a simple interactive application that could be downloaded on their smartphone. The figure depicts the overall flow of how the application works (see figure 2.1). It is primarily the expenses for installing IP cameras in the user's surroundings that cause issues with this system.

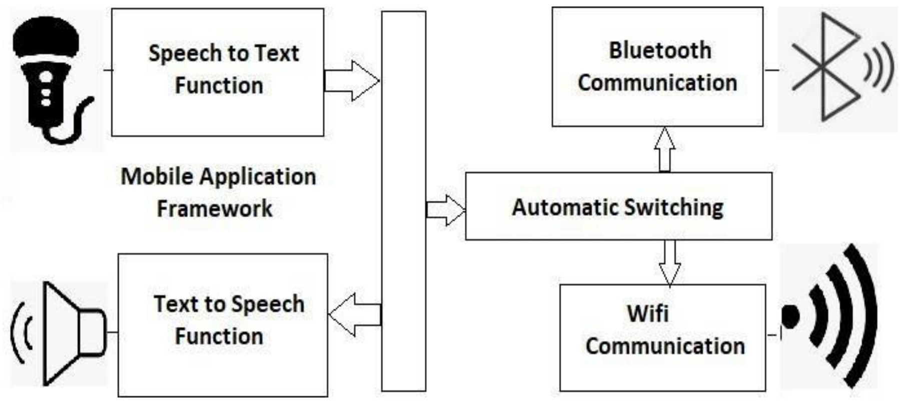


Figure 2.1 Block diagram of the VSLAM extended system [40]

2.2.1.3 Visual Simultaneous Localization and Mapping (VSLAM)

VSLAM (Visual Simultaneous Localization and Mapping) involves using visual inputs from a camera to detect and position objects simultaneously [64]. This technology is highly demanded in navigation design since it is easy to implement and requires only one camera sensor. Bai [65] reveals how to solve the problem of indoor localization and the construction of virtual blind roads using the VSLAM algorithm. Using a strategy of avoiding obstacles along the way, the system helps users navigate by selecting sub-goals dynamically. The same VSLAM approach was used in [66]. An android smartphone, a helmet equipped with stereo cameras, and a cloud server. Fig.

2.2 shows an overview of the system, including the cloud computing platform and the web-based application. The evaluations indicated the accuracy of object classification.

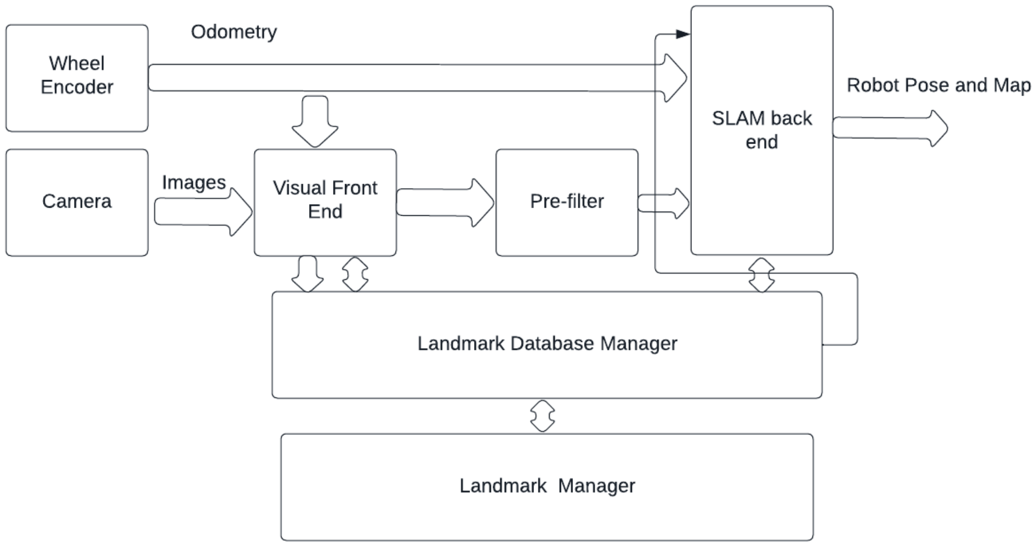


Figure 2.2 Overview of the system, including the cloud computing platform and the web-based application [66]

2.2.1.4 Depth (RGB-D) Cameras

This device uses a Google Tango tablet [67] as its mobile computing platform. When the electronic prototype of Smart Cane was developed, it used an intelligent situational awareness and navigation aid (ISANA) [68]. In addition to the RGB-D camera onboard, ISANA applied a Kalman filter approach for obstacle avoidance (TSM-KF). As part of the multimodal human-machine interface protocol, Xiao et al. [69] developed an electronic Smart Cane that integrates speech-to-audio interactions and robust haptic responses. The system is capable of providing both auditory and vibrotactile feedback. To do the computation, the system needs internet access; however, it can work indoors as well as outdoors.

An RGB-D camera was also used in the system proposed by [70] to facilitate visually impaired people in indoor navigation. Along with navigation software, the system includes a haptic feedback vest and a smartphone. A system may identify the location of both the start point and the destination point based on user voice commands. The system can not only store previously generated maps for navigation but also create new ones as the user is traveling. Figure 2.3 shows overall RGB based system.

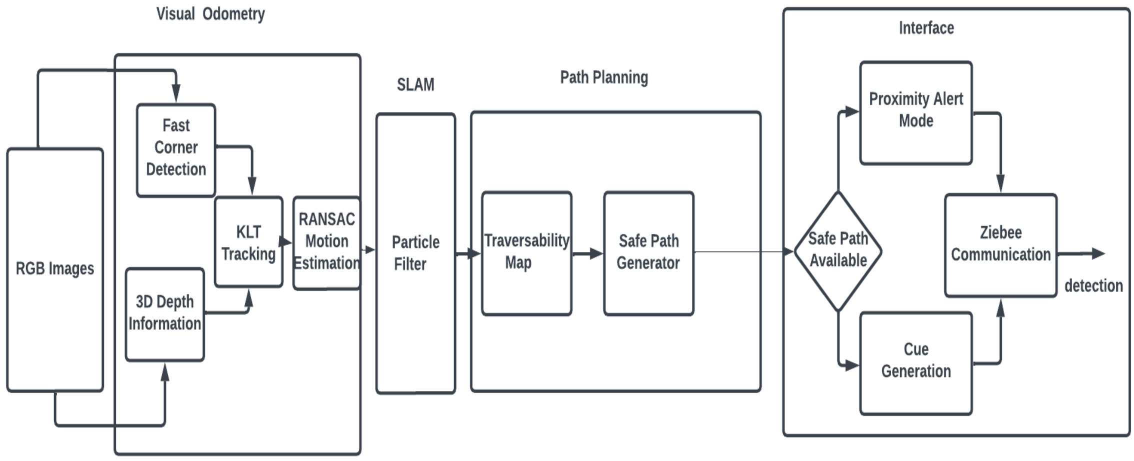


Figure 2.3 Overall RGB based system [70]

Based on wearable sensors, the model proposed in [71] provided navigation directions for blind users. Through audio and vibration feedback modes, the system detects line-of-sight objects using embedded sensors like depth sensors and an Inertial measurement unit in the device model. Social sensors (like Facebook posts, Twitter tweets, etc.) are used by users globally when making decisions. Using the social sensor, blind users can track the contents posted by others, and receive notifications about the occurrence of accidents at a certain spot. By utilizing this information, the navigation system can recommend a route so a blind person can follow it. Based on the changes in the environment, a route decision can be made.

2.2.1.5 Microsoft Kinect

In the design model used in navigation systems for the visually impaired, Microsoft Kinect has attracted great attention among researchers because it is designed to work with RGB-D cameras [72]. Consequently, it is worth mentioning the advancements that have occurred in this field separately. Motion sensing input devices such as Kinect is produced by Microsoft and can detect objects, so can be used for navigation.

A wide range of features is available in the devices, and they can work in low light.
Using Microsoft Kinect 360 for Xbox [74] data for input, [73] proposed a system that used an algorithm. By using this software, a 3D map of the indoor space can be

designed and distance can be measured between objects and people. A similar obstacle avoidance system was also developed for visually disabled people using a Kinect sensor [75]. Figure 2.4 shows the Kinect depth data processing of the system.

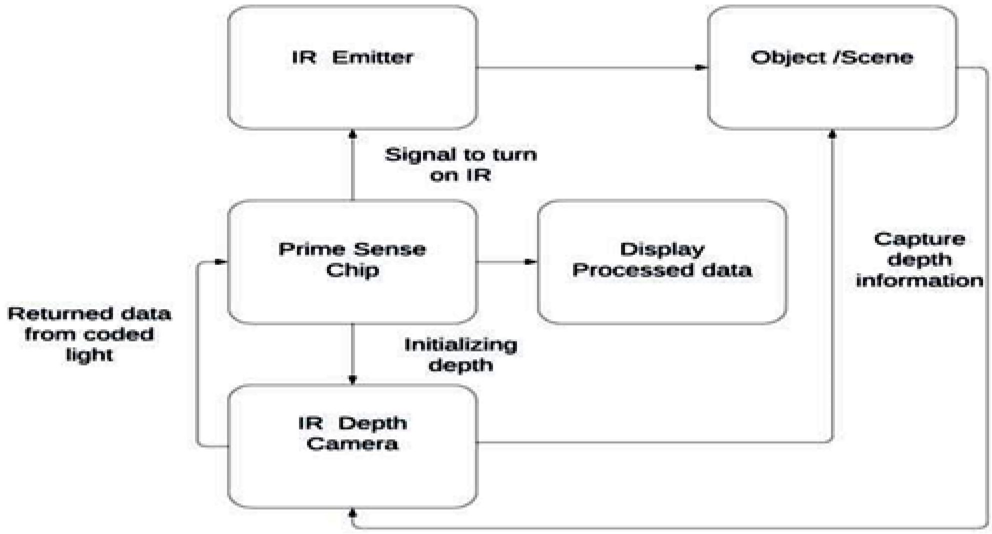


Figure 2.4 Kinect depth data processing [73]

A windowing-based mean method was applied to the Kinect depth images before they were used to detect obstacles. Upon recognizing an obstacle, the system transmits feedback via earphones.

2.2.1.6 LiDaR

Researchers explored LiDaR in navigating visually impaired individuals. With LiDaR Assist Spatial Sensing (LASS) [76], obstacles are detected and translated into stereo sounds of various pitches, which are then sent to a receiver using a LiDaR sensor. A sensor's spatial information is used to translate an obstacle's orientation and distance into relative pitch values. Additionally, it has been proposed in [77] to use the LiDaR sensor with a white cane. In addition to its size and weight, the system's scanning of the surroundings for obstacles is another disadvantage. LiDAR has been able to better exploit its advantages in this area with the release of smaller sensors.

2.3 Non-visual Data Systems

A discussion on varieties of nonvisual navigation systems that rely on information other than vision to guide them. There is a wide array of systems that rely on non-visual sensors like ultrasonic, beacons, and IR sensors to give navigational guidance to users. Although there are systems that rely mainly on features that will give navigational guidance to VI users.

2.3.1 BLE Beacons

There have been multiple reports of Bluetooth beacon-based systems in the literature [78-80]. Using Bluetooth Low Energy (BLE) beacons and Google Tango, Nair [81] proposes a hybrid positioning and navigation system to maximize its strengths and minimize its weaknesses. GuideBeacon uses Bluetooth beacons placed at different indoor locations to provide navigation instructions over smartphones via the GuideBeacon system [82]. The user interface and navigation modules of the proposed system could be improved. Figure 2.5 shows the building blocks of the GuideBeacon System.

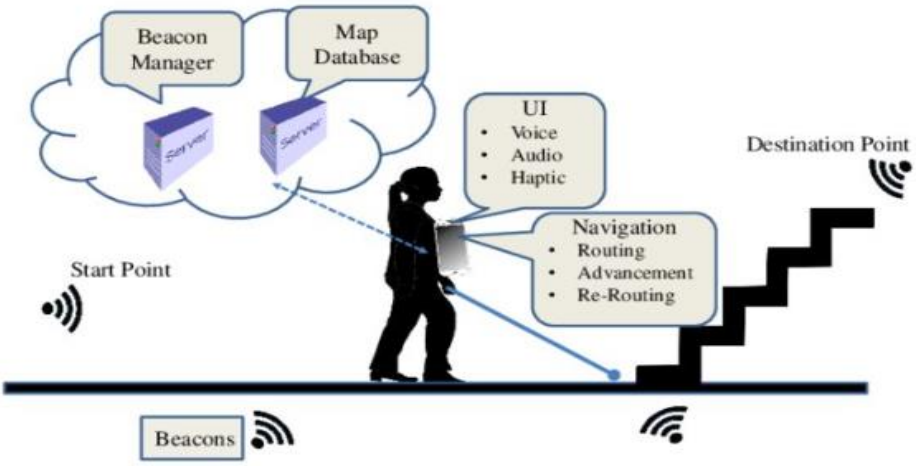


Figure 2.5 Building blocks and interaction of the GuideBeacon system [81]

2.3.2 IoT based

The Internet of Things (IoT) is the network of connected devices from which data can be transferred without any human or machine interaction. After the use of IoT in different applications, there are more and more reports in the literature describing navigation systems using the IoT concept [83, 84].

A handheld device called Indriya [85] is used along with a smart cane. Up to three meters away, the system can detect obstacles and humans can be distinguished with

80% accuracy. The system can also provide voice and vibration alerts before a possible collision. For its implementation of IoT, Indriya uses a minimal set of sensors based on its Android platform support. However, the system often fails to recognize slopes and steps. Figure 2.6 shows the prototype specific block diagram of Indriya.

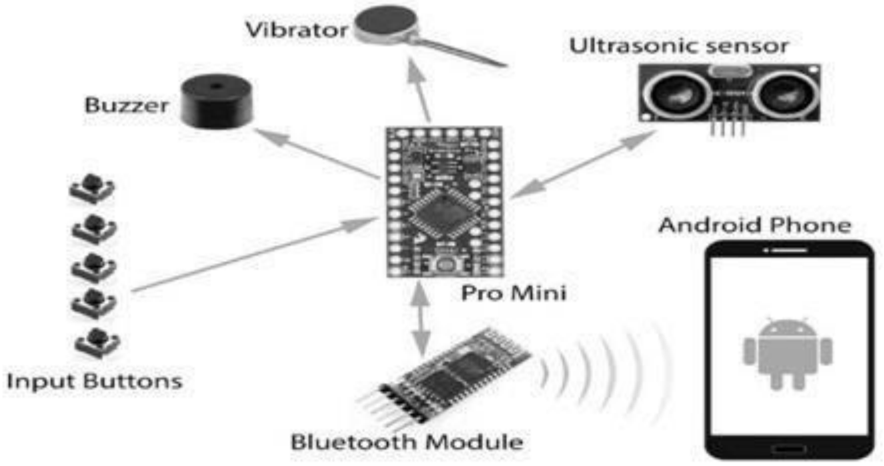


Figure 2.6 Prototype specific block diagram of Indriya[85]

In addition to taking advantage of the Internet of Things, The Blind Guide [86] utilizes other solutions and aids visually impaired individuals in navigating both indoors and outdoors. The Raspberry Pi board is notified when an obstacle is detected by the wireless sensor. The voice feedback system will inform the user of the distance and name of the object once it has been identified. As a prototype, the system depends on internet access for object recognition, making it dependent on Internet access in data networked locations.

2.3.3 Ultrasonic Sensors

Ultrasonic-based navigational systems are becoming one of the most popular choices to replace visual (camera) navigation systems in design. This technology is intended to work with Raspberry Pi or Arduino [87]. The ultrasonic blind stick described by [88] uses a proximity sensor (ultrasonic) and a Global Positioning System module. Utilizing ultrasonic sensors, As an obstacle avoidance system, NaviGuide[89] can recognize many different situations in the environment. Vibration and audio alerts may be used by the system to inform the user of priority information. As part of the design, the system can detect obstacles at knee level, floor level, and even wet floors. NavGuide cannot detect a pit or a downhill, which is a major limitation. Furthermore, NavGuide can detect wet floors only after a user step on them. In addition to ultrasonic sensors, the GuideCane [90] employed an embedded computer to detect obstacles during navigation and monitor the direction the system was moving as well as the user's motions. Overhanging obstacles such as tabletops and sidewalk borders, as well as important features such as sidewalk borders, are not detected by the Guide Cane.

2.3.4 Infrared Sensors

Infrared sensors are lower in cost and power consumption than ultrasonic sensors, so navigation experiments were conducted in this area. The development of a smart stick using infrared technology reported in [91] uses technologies such as Google Tango and Unity [92]. The solution presented in [93] uses infrared sensors to sense different objects such as walls, buildings, and animals. By placing the device on their arms, users can transmit navigation signals through vibrations. They are used to transmit navigation signals through movement and to notify the user of nearby threats.

2.3.5 Map-based Systems

When visually impaired individuals complete the O&M training, they use tactile materials, such as raised point maps, small-scale prototypes, or magnet boards. A variety of maps have been proposed to aid blind and visually impaired individuals in navigating. Using such tools allows blind and visually impaired people to develop spatial learning abilities. This includes the inability to update map content. To address this concern, accessible interactive maps have been designed [94]. An augmented reality map developed by the authors of [95] using participatory design can assist O&M classes. In this prototype, audio output is combined with projection and tactile tokens are used. Thus, people with visual impairments can explore and construct maps. Figure 2.7 shows system overview

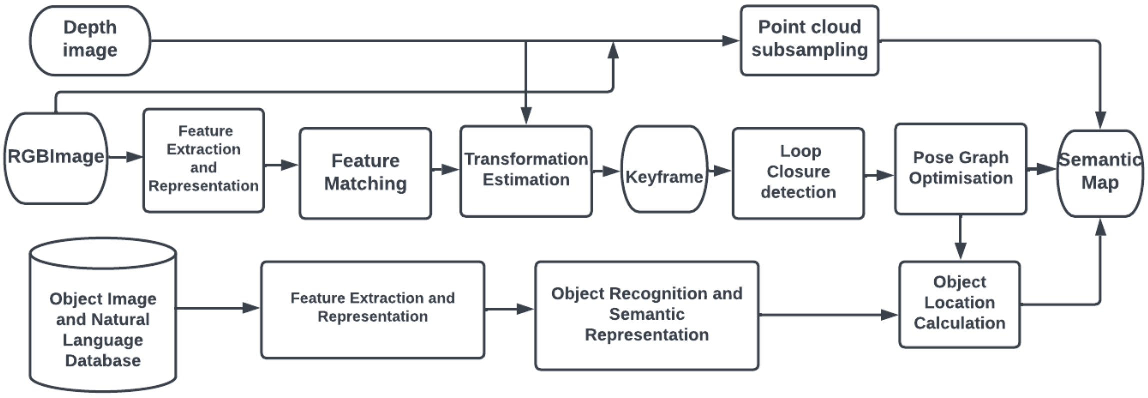


Figure 2.7 System overview [74]

Blind people could explore tactile maps using SmartTactMaps, a smartphone-based approach proposed by [96]. An RGB-D sensor was used to create a 3D environment map reported in [97]. Visually impaired people could navigate at home using the system's semantic information extracted from the RGB color images.

A tactile audiovisual map for eye-impaired people based on 3D printing called LucentMaps [98] has also been proposed, which will integrate mobile devices and physical maps more easily. With the VizMap system [99], various indoor information can be collected via computer vision and crowdsourcing. The model is created by volunteers to convert video into 3D spatial representations. A reconstructed 3D model that can query the environment is constructed from these video frames that have been semantically indexed.

2.3.6 Sound based systems

The model described in [100] functions by delivering a sensory representation of the environment to visually impaired individuals through a wearable sensory device. The user will be able to provide acoustic and haptic feedback. There are still several areas that need to be improved regarding usability and accuracy.

Electronic travel aid with stereo vision (SVETA) consists of stereo cameras molded into a headgear, which also has earphones. This report proposes a procedure for mapping stereo disparity imagery to musical sounds. The model performs well in indoor environments because each acoustic sound exhibits some information about the obstacle that the users. Users who are not familiar with the system may find it difficult to use it. Before using stereo musical sounds, the target users should be trained to understand their different meanings [101].

2.3.7 Smartphone based solutions

Navigation-based solutions use smartphones that offer users convenience and portability. There are a variety of solutions for visually impaired users on the smartphone platform described in this section. As shown in Figure 2.8, a speech operated system NavCog3 was a system for indoor navigation proposed by [102] which provides real-time feedback when incorrect orientation is detected and turn-by-turn instructions are provided. An application called PERCEPT-II proposed by Ganz et al. [103] allows mobile users to get GPS navigation instructions when they touch specified landmarks on the mobile device. Tags for Near Field Communication (NFC) were attached to the destination spots.

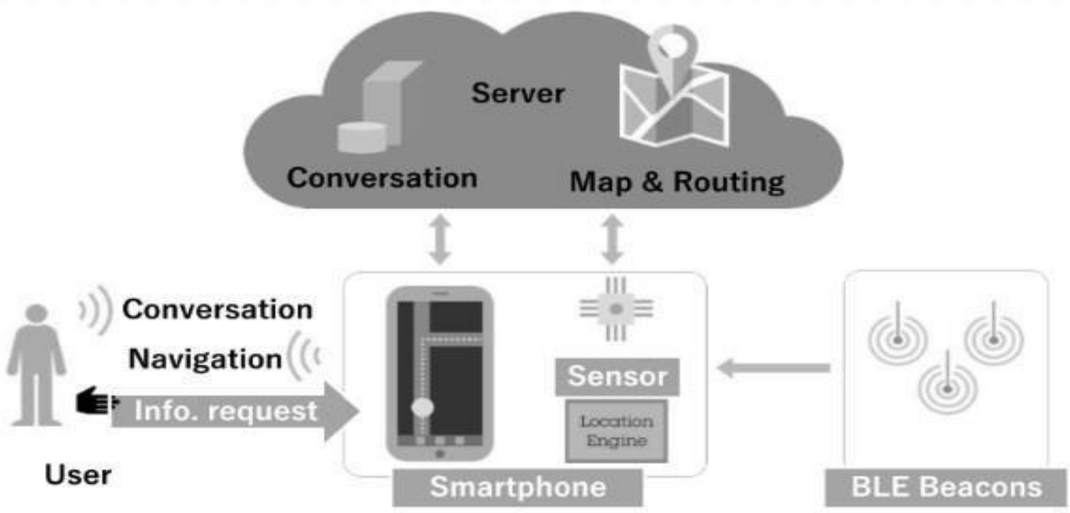


Figure 2.8 Overall system with speech-operated navigation [102]

According to Lin [4], an assisted navigation system can be developed by integrating a smartphone application with an image recognition system. A smartphone captures an image when the system is initiated and sent for processing. The overall system use deep learning algorithms [104, 105] but lacks many features, including high energy consumption, high network bandwidth requirements and many other problems.

A study by TARIUS et al. [106] aimed to improve the understanding of visual scenes in outdoor environments by using cameras. An application for mobile devices, a web server, and a remote assistance center are included in this system, as well as Bluetooth LE/iBeacon tags mounted along streets near points of interest. The system is plagued by issues such as the placement of Bluetooth beacons throughout the city.

The Envision [15] offers robust and accurate detection of static obstacles in live video streams captured by smartphones of average hardware capacities. Additional improvements could be achieved if the target users could understand the environment through obstacle recognition.

As part of the project called "Active Vision with Human-in-the-Loop for the Visually Impaired," [107] proposed a method for transmitting navigational information to the visually impaired. Vibration and audio cues are used to communicate with the user. ActiVis is currently implemented as an app using a Tango device [68] and a head-mounted display. The model will be more effective if the feedback parameters are adaptable. System performance can be improved by using Tactile Wayfinder [108]. This system includes a tactile belt in conjunction with a Personal Digital Assistant (PDA) that runs an application. This information is displayed once the path direction has been determined as a navigation aid, the belt vibrators provide navigational information to the user in the form of directions.

K. Sundar Srinivas et al. [13] wrote an article proposing a device for detecting obstacles and recognizing faces to assist visually impaired individuals socially. There was a device proposed that consists of smart glasses that had a raspberry pi, an ultrasonic sensor, and a Raspberry pi camera installed. The device which is connected with the raspberry pi, which can receive the ultrasonic transmissions from the sensors, process the data, and detect obstacles up to a distance of 1000 cm around the user. In order to identify visually impaired individuals in front of the camera, a database list is used to identify the person in front of the camera. It is easy to recognize people with this device. People can easily be identified with this device that is cost- effective and consumes little energy.

Pavan Hegde [14] developed, designed, and proposed a rugged, low-cost, low-powered, user-friendly, low-power smart eyeglass that would cater to people who are visually impaired. A simple eyeglass model has been developed that can be worn easily as an inexpensive solution. A sensor in the device emits ultrasonic waves that follow a person's movement and can detect them up to 6 meters away. It was improved to make it as safe as possible for users with basic image processing and computer vision approaches for each version. This device automatically generates an announcement through an earphone attached to the user once it determines what obstacles are present in the environment along with the number of counts of each obstacle.

Wu Tang et al., [15] presented a model for blind people to avoid those obstacles in an outdoor obstacle detection and recognition system. Their model creates a new obstacle-dataset (OD) with 15 common features for outdoor obstacle detection that has previously been considered as a benchmark. Their work evaluated three object detection algorithms including YOLO, SSD, and Faster RCNN to find the most effective one. The YOLO algorithm performed the best of the three algorithms.

Ahmad Sheikh Sadi et al. [16] were able to develop a prototype spectacle to assist people who are visually impaired. Using ultrasonic sensors, a walking guide was able to identify obstacles in various directions. In addition to the ultrasonic sensor, Convolutional Neural Network (CNN), which is based on Deep Learning algorithm, the algorithm can detect potholes on the road surface. The CNN is controlled by an embedded controller that can identify the surface of the road from obstacles. For real time classification, an embedded controller used a Convolutional Neural Network (CNN) that had been trained on the host computer system. Experimental analysis indicates that the system detects distances using ultrasonic sensors when 50 cm away from obstacles with 98.73% accuracy.

In a recent study, Annapoorani et al [17] demonstrated a Blind Sight-Object Detection system that utilizes state-of-the-art object detection methods and computer vision. For the system to function correctly, tasks that can be dealt with by the human visual system must be automated. The classification of images is achieved by identifying the features of the image and categorizing them into appropriate classes. Based on

COCO's 123,287 hand-labeled images, which are grouped into 80 categories, this project was largely based on the COCO dataset. A variety of spatial relationships are used to describe objects and their locations in space based on this data set. Object detection is used to detect objects in this system using a framework called You Only Look Once. In addition, the Indian currency denominations have been identified using a module developed to recognize the currency.

Dola Das et al [18] uses a Raspberry Pi, an SD card, vibrationators, headphones, and ultrasonic sonar sensors conducted an experiment. Through the PIR sensors, the system detects movable obstacles, such as people, vehicles, etc., in all directions around the user. Utilizing Ultrasonic Sonar sensors, the device measures the object's velocity as well as distance. Accordingly, the user receives vibrations and audio messages when approaching objects approach in the specified direction, velocity, and distance. It is also suggested that real-world applications can be made of the system by the author.

Saumya Yadav et al. [19] proposed a wearable assistive device that would assist visually impaired people in navigation automatically. Sensors, cameras, single board DSP chips, wet floor sensors, battery pack and a whole set of computing components are required for visually impaired people. In order to train the user to feel comfortable in the environment, they used a machine learning model for object recognition. For navigating staircases, potholes, speed bumps, wet floors, and narrow passages, this device provided navigation guidance. An alert was provided via vibration, to notify the user of impending obstacles.

The work of Mansi Mahendru and colleagues [20] proposed a system that detects a variety of commonly found objects and alerts the individual as to what is nearby as well as what is farther away. There are two different versions of YOLO algorithms being used in the development of the system, YOLOv1 and YOLOv3. In both versions, the algorithm was tested in a set of different scenarios to determine if it was accurate in every situation. YOLO and YOLOv3 are both tested for performance and accuracy using the same criteria. After the experiments were analyzed, it was discovered that YOLOv3 is significantly more powerful to detect small objects and distant objects than YOLO.

The deep learning technology proposed by V.N. Honmane et al. [21] enables blind individuals to detect objects In addition, a voice guidance technique is available to assist people with visual impairments in locating objects. In order to provide information about the objects via a deep learning model, a text-to-speech synthesizer is used to make the recognition easier for the person.

With the object- detection system, a person can locate objects within a specific space without assistance from others, and the system has been tested through several experiments.

Raghad Raied Mahmood et al., [22] proposed a method for assisting visually impaired people to locate objects in real time during their daily activities. It was possible to detect objects in an image whose location in its x-axis and y-axis was fixed by using an object detection framework. With the help of Google Text to Speech, you can convert the label of an object into audio. This results in the system becoming more reliable and accessible.

During this section, we discuss the research conducted on assistive technology for visually impaired individuals. The majority of the discussion in this thesis deals with these topics and the simulation. With the advent of convolutional neural networks-based deep learning algorithms, some of the best solutions are now available, and they are making a real impact, particularly when integrated with the Internet of Things. An overview of deep learning algorithms and their merits, drawbacks, and limitations is provided in this article.

2.4 RELATED WORK

In this section, we discuss, the relevant work related to assistive systems for helping blind persons and Object identification Algorithms. We discuss the offerings and drawbacks and/or limitations of the algorithms/systems.

2.4 Object Detection Algorithms

2.4.1 SIFT Algorithm

For object detection, the SIFT algorithm used to be one of the most popular algorithms before Deep Learning took over. In order to extract key features from a frame, the SIFT algorithm [23][24][25] is used. To detect moving objects, these key features are grouped by improved k-means clustering algorithm [37] they have stabilised the video frame with the log polar transform, which makes it resistant to scaling and rotation[41]. A comparison of their proposed solution with SIFT-ME is then conducted in order to evaluate their transformation accuracy. By combining SIFT and SURF algorithms [65] is able to detect objects reliably, resulting in a robust tracking algorithm. As a result of the above experiments, SIFT is a high-accuracy system, but it has a slow speed. Overall, their method performed well, but suffered from scaling issues, cluttering, and changes in illumination which make it difficult to accurately identify objects.

2.4.2 SURF

A high level of accuracy was achieved with SIFT, and it could handle scaling well. There was one drawback, which was the algorithm speed, which was very slow. An algorithm called SURF[30][27][26] or Speeded Up Robust Features[40] has been developed to increase the speed of object identification. These studies include enhancements to SURF performance as well as comparative studies with algorithms such as SIFT, ORB, and BRIEF, among others. To improve the matching rates and make the SURF algorithm illumination invariant, enhancements were made to the algorithm.

2.4.3 OCR

Object identification has also been widely performed using OCR algorithms. In most cases, it has been used to detect text in images. Researchers have used OCR to detect food labels, expiry dates, and ID card information. They have combined OCR [32][51][33]with Faster R-CNN into a single neural network called OCR-RCNN.

The results from all these experiments can be summarized as follows:

- OCR performs good in case of reading texts with accuracy as high as 70-90%

- OCR suffers with varying light and rotation
- reflective surface also pose a problem to the OCR algorithm
- OCR performs accurate retrieval of information

2.4.4 CNN

CNN, also known as convolutional neural networks, are subclasses of ANNs that are typically used in the analysis of visual images. In CNN models, there are a limited number of layers used to process input data, such as images at varying levels of abstraction. According to CNN, this network uses a mathematical operation called convolution. In this operation, the input signal f is convolved with the output signal g by means of a filter. By performing a scalar product operation, each position of the filter in the signal produces a scalar through the convolution. Each signal point is processed by the filter, and the scalar product produces a scalar as a result. In return, the convolution yields a complete signal (i.e. an image). CNN architecture is shown in the figure below.

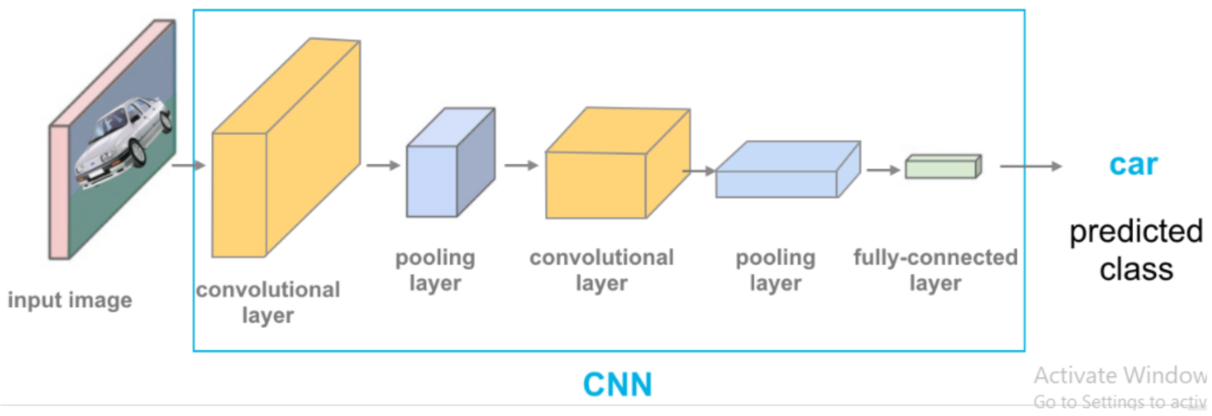


Figure2.9 Basic CNN architecture.

2.4.4.1 Convolution layers

Convolutional layers comprise the first layer of a CNN, which takes as input the image to be classified. Several convolutions of a single input are performed with different filters, and each convolution is completed with a non-linearity operation. An output of a non-linearity operation is also called a feature map. In order to quantify the resemblance between a local image and the filter, a scalar product is calculated between a filter and a set of local images defined by the filter size. Figure 2.10 shows an example of a kernel, which is a filter. Successive scalar products i.e. the convolution result of the filter with a 6x6 input matrix is illustrated in figure 2.11. In this case scalar products calculate resemblance degrees of sliding local matrices of size 3x3 (inside the 6x6 matrix) to the 3x3 filter, which is a prototype of vertical edges.

After every convolution, the pixel values obtained from the convolution are subjected to a non-linear operation. (Not illustrated) A convolution layer includes the non-linear operation. In the case of convolution output, the piecewise nonlinearity known as ReLu is an example. Images generate scalars when their filters are moved in the image, and scalars are images themselves.

1	0	-1
1	0	-1
1	0	-1

Figure 2.10: Example of 3x3 kernel

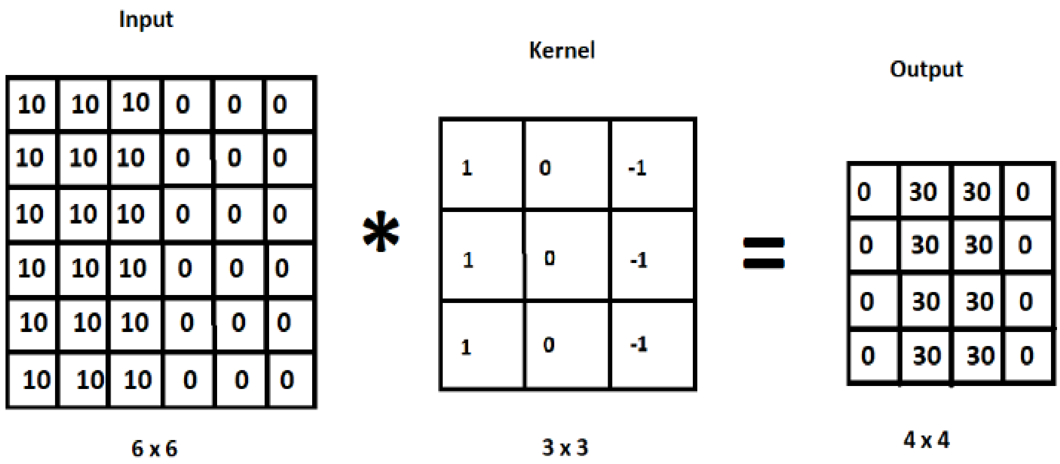


Figure 2.11: Example of convolutional operation with 6x6 image and 3x3 kernel.

2.4.4.2 Pooling layers

In CNNs, the pooling layer follows the convolution layer. A smaller version of the feature map is created by shrinking down the large map. Each step of the pooling retains the most dominant feature or data during the shrinkage. In the pooling operation, an operation similar to that performed in the convolution layer is performed. It is determined how much the map steps through the pooling and how much dominant data is extracted.

This pooling method is called max pooling as shown in figure 2.12.

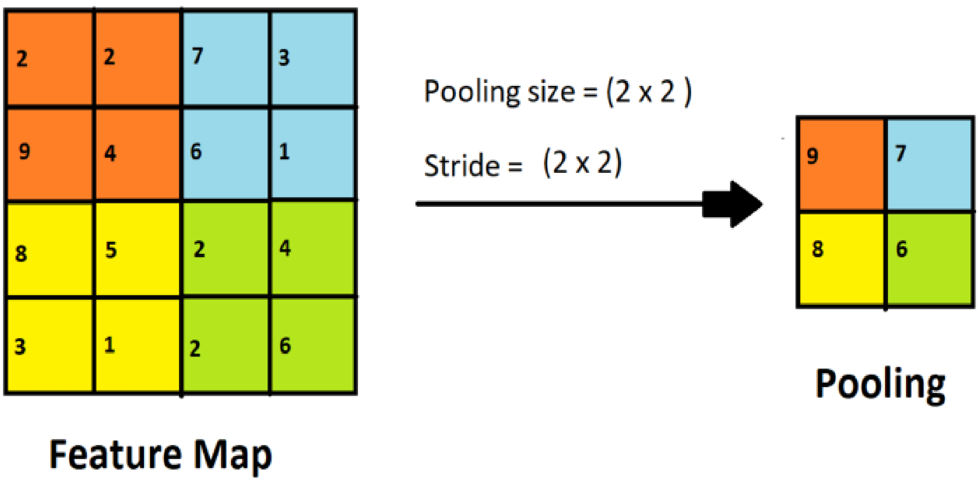


Figure 2.12: Examples of max pooling are 2x2, and it steps by 2x2.

2.4.4.3 Fully Connected Layer

Fully connected layers are the last layers in CNN architecture. Whenever a neuron in this layer communicates with another neuron in the previous layer, it will be called a layer of neurons. This layer flattens a matrix into a vector by scanning it in a predetermined way after receiving the last feature map. In order to generate the final output for CNN, this vector is fed into the fully connected layer.

2.4.5 YOLO

Since its release, the YOLO algorithm has gained popularity. YOLO as its name suggests only needs to look once for object detection. YOLO is a single shot detector, which means that it performs the localization and Image classification in only 1 step instead of doing it in two different steps. This speeds up the overall process of object recognition but introduces a minor degradation in accuracy . Over the years there has been 2 major up gradations to YOLO algorithm [43][36][50]: Yolo v2 and v3 and while the original algorithm is called as Yolo v1.

The results from their experiments are as follows:

- The results show that YOLO v3 is faster than its predecessors and is more accurate.
- YOLO and Faster-RCNN have comparative accuracy. However, Yolo v3 is better than Faster R-CNN in case of recall rate.
- YOLO3 inherits few problems from the old YOLO v1 such as the localization issues when multiple small objects are near to each other.

2.5 SUMMARY OF THE SURVEY

Our understanding and visualization of the entire thesis was greatly enhanced by the literature survey discussed above, which provided insight into the project's flow and architecture.

This research uses proximity sensors, such as ultrasonic ones, to detect obstacles by measuring distance between them and the user. This project also employs an object detection framework called Yolo for identifying objects. Several devices have been proposed to detect obstacles, determine their distance, and determine their category, but none combine these characteristics in combination using audio feedback. The techniques can also be used to develop a visual aid for people with visual impairments that is also useful for them as a supportive tool.

The comparative list between algorithm is shown in Table 2.1

Table 2.1 Comparative analysis of various algorithms

Algorithm	Author	Advantages	Limitations
SIFT (Scale Invariant Feature Transform)	Sharif et al. (2019) [23] Zheng et al. (2018) [24] Karami et al. (2015) [25] Jabnoun et al. (2015) [49] Liu et al. (2016) [34]	High Accuracy and invariance to rotation and scaling.	The algorithm is slow and performs poorly against blurring, occlusion and changes in illumination
SURF (Speeded Up Robust Features	Tareen et al. (2018) [30] Ding et al. (2018) [27] Wang et al. (2018) [26] Kim et al. (2017) [37] Geng et al. (2017) [29]	Fast, robust, requires little computation power and is robust to transformation	Inherently less accurate than SIFT, fewer key points are detected than other algorithms.
OCR (Optical Character Recognition)	Adriano et al. (2019) [32] Borisyuk et al. (2018) [51] Liem et al. (2018) [33] Kento et al. (2018) [26] Zhu et al. (2018) [35] Deshpande et al. (2016) [31] Liu et al.(2016)	Accurately and quickly recognize text in the images	Not light-resilient, low image detection accuracy.

Regional Proposed networks	Najibi et al. (2019) [47] Ammirato et al. (2019) [48] Chao et al. (2018) [46] Chen et al. (2018) [45] Hu et al. (2017) [42]	It is fast and capable of handling transformations as well as dealing with illumination changes well.	The training process is quite lengthy which includes multiple stages making it slower
YOLO (You Only Look Once)	Cao et al. (2019) [43] Benjdira et al. (2019),[36] Kim et al. (2017) [50] Maolanon et al. (2018) [40] Park et al. (2018) [43] Widhyastuti et al. (2018) [41] Redmon et al. (2018) [38]	The algorithm performs all of its predictions with the help of a single fully connected layer of neural network in one step, making it the fastest.	High Speed detection

CHAPTER 3
SYSTEM DESIGN

3.1 PROPOSED SYSTEM

Based on our review of various pieces of literature, an easy-to-use mobile application has to be designed which allow blind persons to use the designed system and navigate easily. Our system allows blind people to navigate the streets while wearing a face mask and holding a smart stick while using the application. In addition to identifying objects in the frame, the application will read out the name of each object through headphones worn by the blind user and announce it in audio. Both the name and the position of the object are announced in the audio announcement. In this way, a blind individual can navigate using this information.

The proposed system creates a visual aid as well as a voice feedback system that will help visually impaired individuals remain as mobile as possible. The user is provided with auditory feedback through headphones while the system captures the actual environment in real-time.

The Raspberry Pi 3 Model B+ was chosen for its affordability and portability as a functional device. In addition, this device is multiprocessing capable. It uses an ultrasonic sensor and a Raspberry Pi camera module to identify obstacles and classify objects. A improved YOLO algorithm detects the object and using eSpeak, a compact open-source speech synthesizer provides auditory feedback of what was detected.

The Stick acts a feedback to the proposed improved YOLO based model to be get improved with the output ranging in closure vicinity of the visually impaired person.

3.2 SYSTEM ARCHITECTURE

In real-time, data is collected using proximity sensors like the ultrasonic (HC-SR04) and the pi camera. When the obstacle is less than a threshold value, Raspberry Pi processes the data from the sensors for obstacle detection, and an audio feedback alert is then sent to the user informing of the obstacle's location and distance. If the distance between the obstacle and the user is greater than the threshold, the distance is continuously measured until it is below the threshold value. The proposed work features a software design as well as a hardware design, the software design being an Artificial Intelligence-based module that enables obstacle-aware navigation for people with disabilities.

As shown in Figure 3.1, the system architecture consists of a 1.2 GHz quad-core Cortex A53 processor, bone conduction headphones, and a 1.2 GHz 64-bit quad-core Cortex A53 processor. It is built to support the tracking and detection of moving objects. Datasets used to train the proposed model include real-time scenarios is associated with the trained CNN model [55] and feed with live camera. Once the model has been trained, the bounding box for the detected area is drawn if the object is detected. A path is chosen from Left, Right, or straight based on the live frame once the number of regions with the most likelihood of detection has been determined.

3.2.1 Hardware Design

The overall structure of a wearable mask and smart cane consists of components like a Raspberry Pi 3, an Ultrasonic Sensor, a GPS (Global Positioning System), a Camera, a headphone, and a power supply.

In comparison to previous Raspberry Pi versions, Raspberry Pi 3 is almost ten times faster due to its quad-core Cortex A53 processor running at 1.2 GHz, Wi-Fi, and Bluetooth 4.1 connectivity. It comes with 1 GB of RAM, an integrated camera interface, fitted with an Video Core IV 3-D of the graphics card chip attached to it..

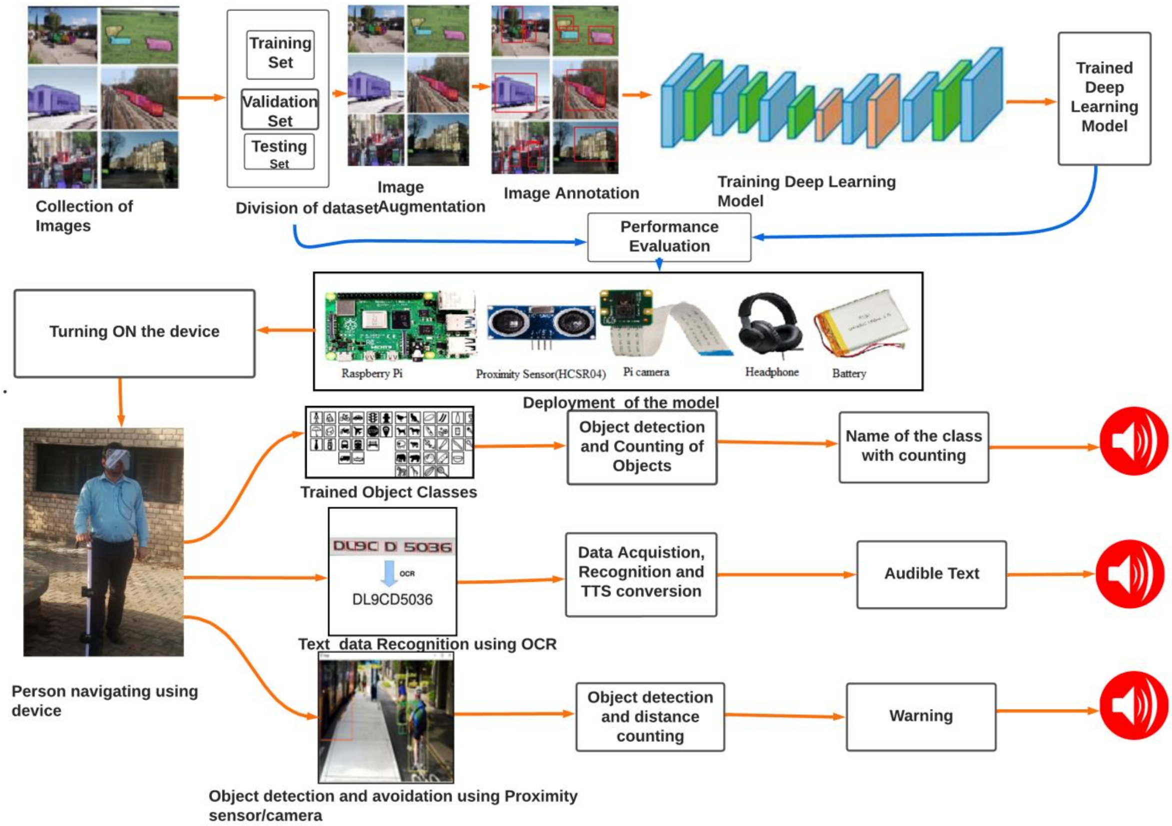


Figure 3.1 System Architecture

The YOLOv5 application is implemented through a Python script. This application reads frames from the video stream when it accesses our Open VINO environment. It performs near real-time object detection using a Raspberry Pi and YOLO.

The detection of objects will be done through an ultrasonic sensor (HC-SR04). It detects obstacles in the way, mainly by using ultrasound. A static current of 2 mA is produced by the HC-SR04 module when it is powered by 5 V DC. It is capable of measuring distances between 2 and 450 cm. The output ranges from Low to High when detecting an object. Our model uses a Pi Camera of 5-megapixel RGB (Red Green Blue) model B camera capable of taking static images of 2592 x 1944 pixels. Additionally, 1080p30 and 640x480p60 video formats are supported. It is supported with Audio Jack that supports 4 poles and 3.5 millimeters. A person who is visually disabled will be assisted by audio signals through headsets connected to this device.

The overall system uses *18650* lithium batteries of 3.7 V with a capacity of 2200 mAh. The overall system is supported by GPS module APM 2.5 NEO-M8N which is a low-cost, low- powered position sensor. It offers a 9600 baud rate and is useful in applications such as Data Logger for the position, velocity and time. The operating voltage of GPS is 3-5 Volts. The whole model was 3-D printed for encapsulating entire hardware components, Acrylonitrile butadiene styrene (ABS) is used in the model. The model is created in Fusion 360 and printed in ABS using 3-D printing.

3.2.2 3D Model designing of Wearable Mask and a Smart Cane

Hardware models are designed in Fusion 360 as shown in figure 3.2-3.14, a cloud-based software platform that supports 3-D modeling, CAD, CAM, CAE, and PCB for designing a variety of products. Our Wearable 3-D model has dimensions of 18 cm in length to 12 cm in breadth, the depth of the model is 4 cm to use with ease.

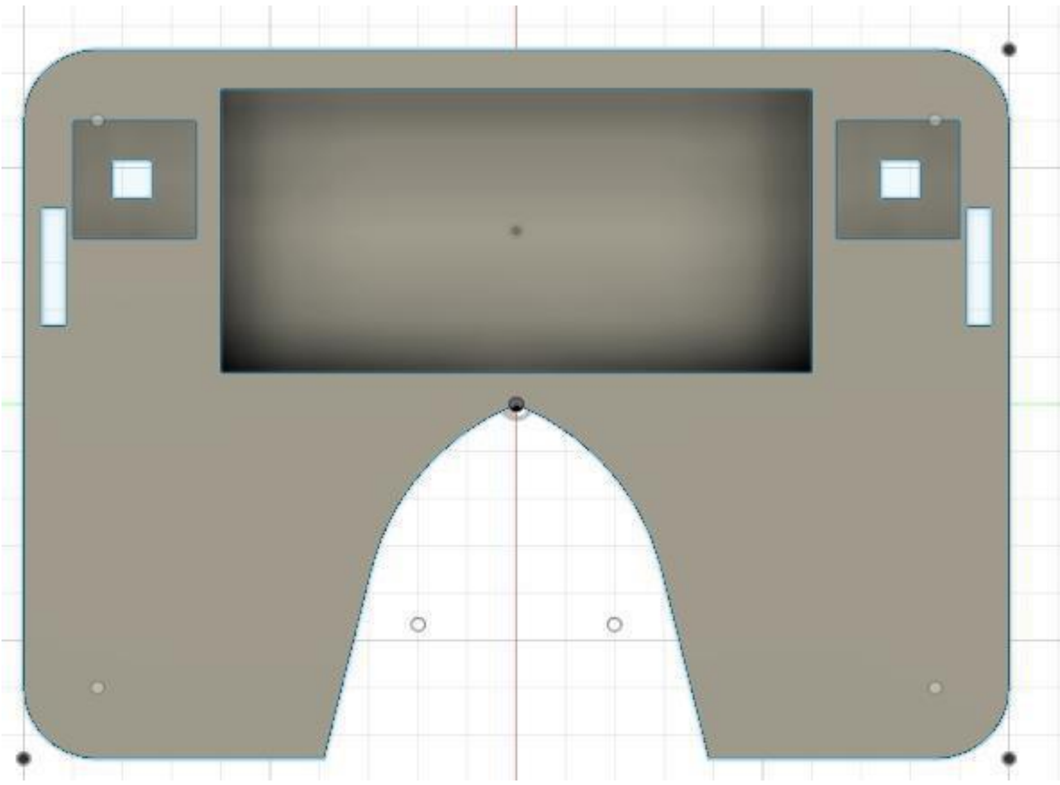


Figure 3.2 Rear part of the hardware model on Fusion 360 Software

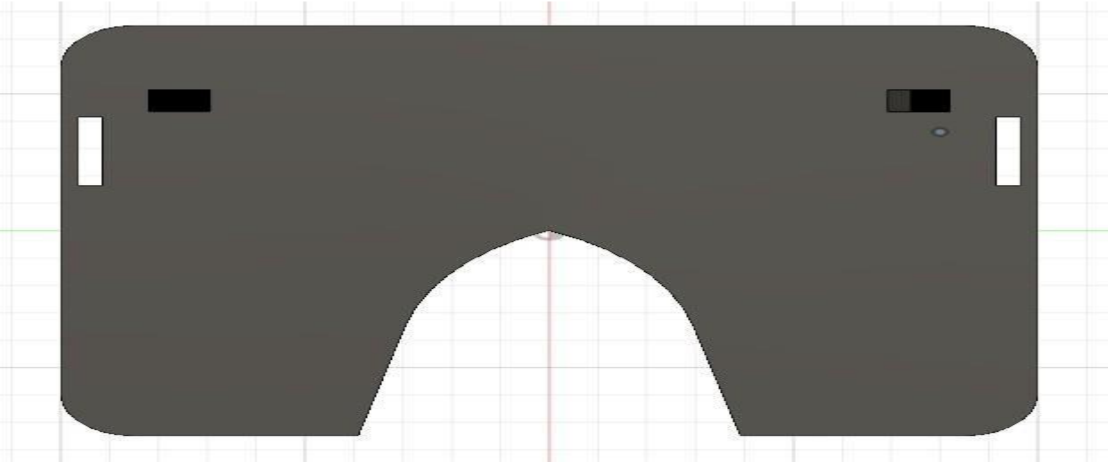


Figure 3.3 Front part of the hardware model on Fusion 360 Software

Figure 3.4 Actual 3D printed model as Virtual Eyes (LPU)

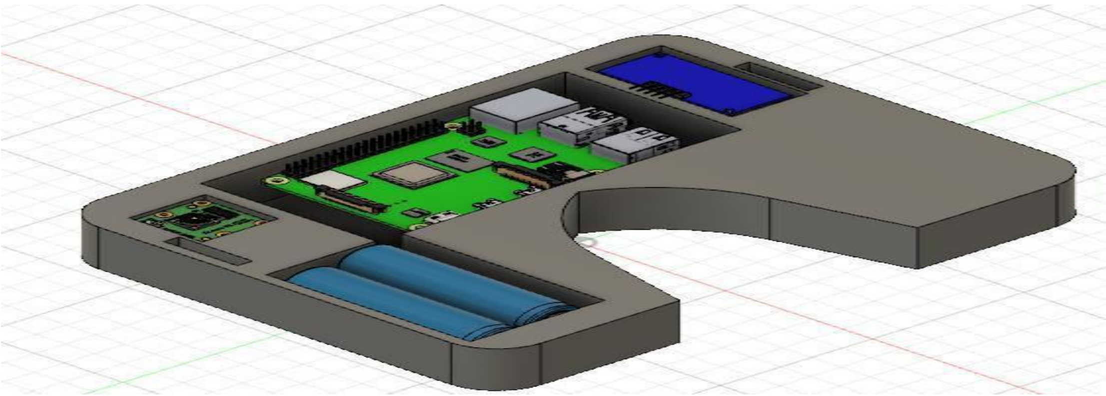


Figure 3.5 Virtually fitted hardware in 3D Printed face mask on Fusion 360

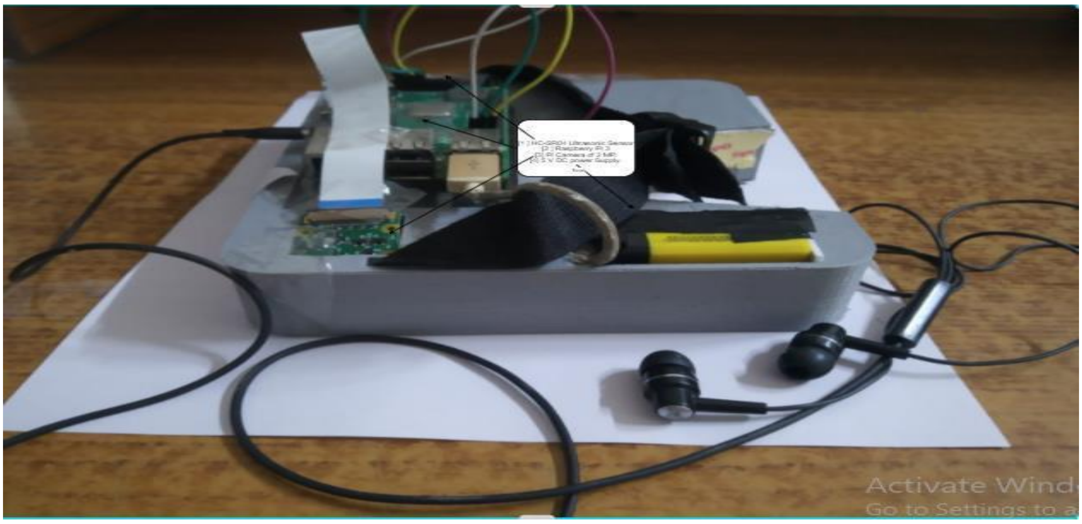


Figure 3.6 Assembled hardware in the 3D printed board

Figure 3.7 Face mask connection of the hardware model for programming in python

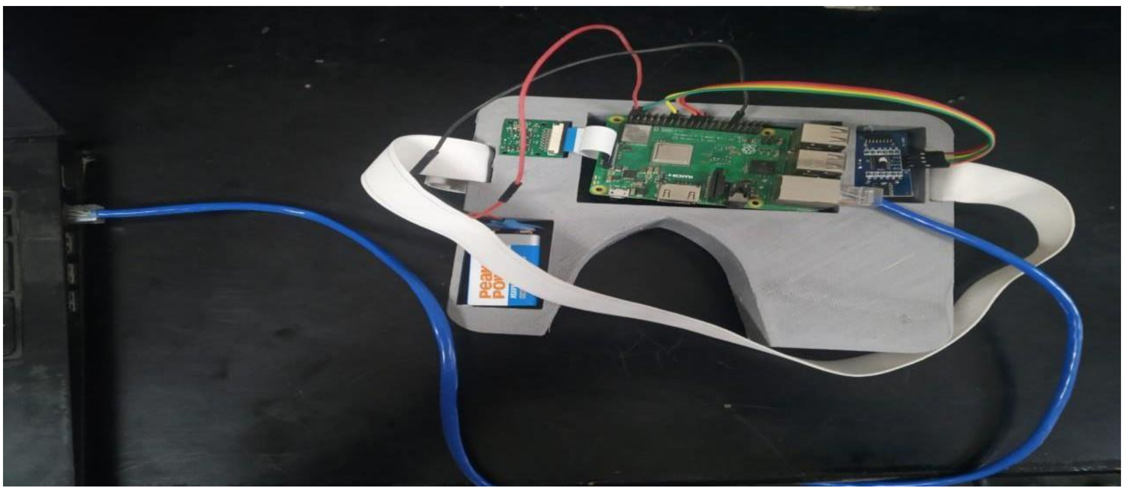


Figure 3.8 Top view of the actual hardware model

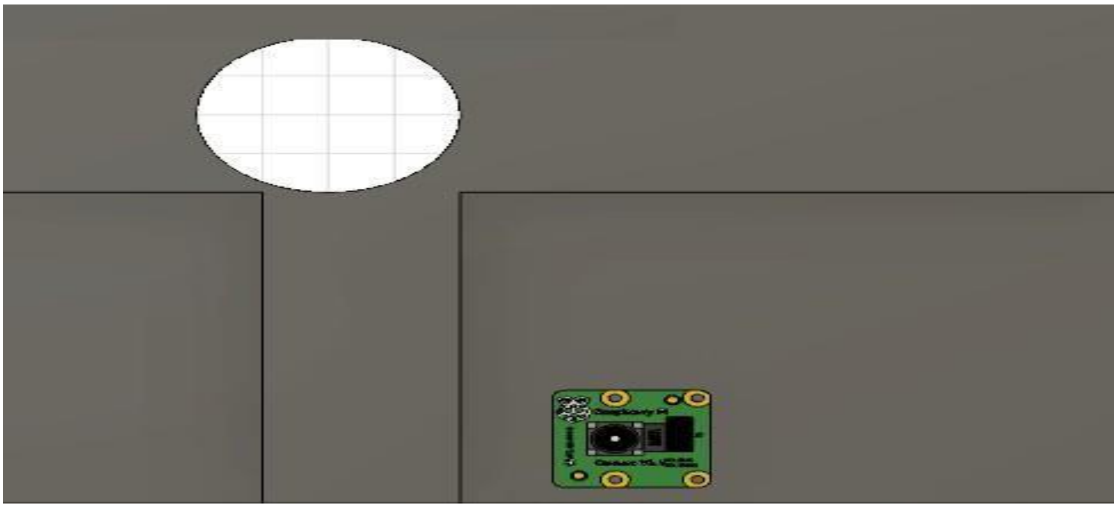


Figure 3.9 Rear view of the advanced stick hardware model on Fusion 360

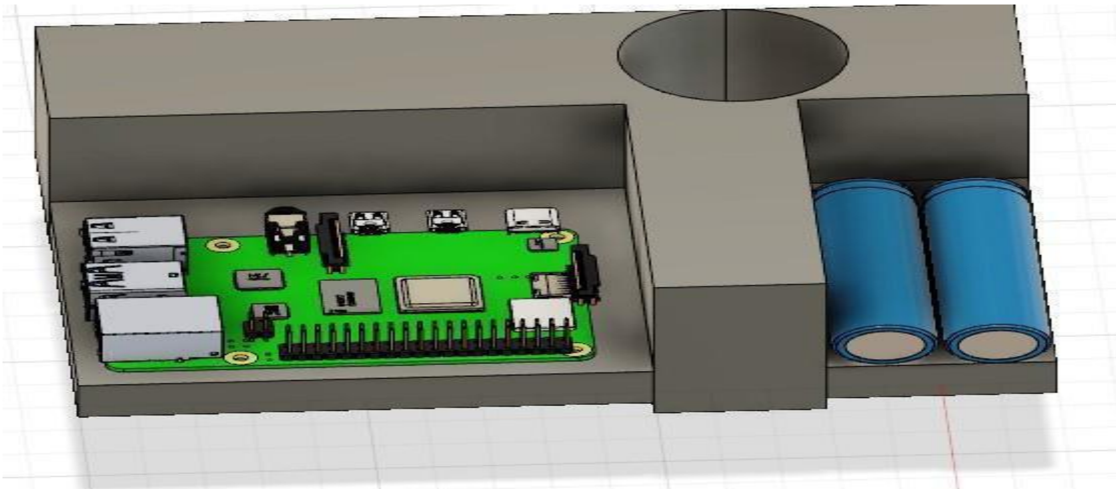


Figure 3.10 Hardware components in different using Fusion 360

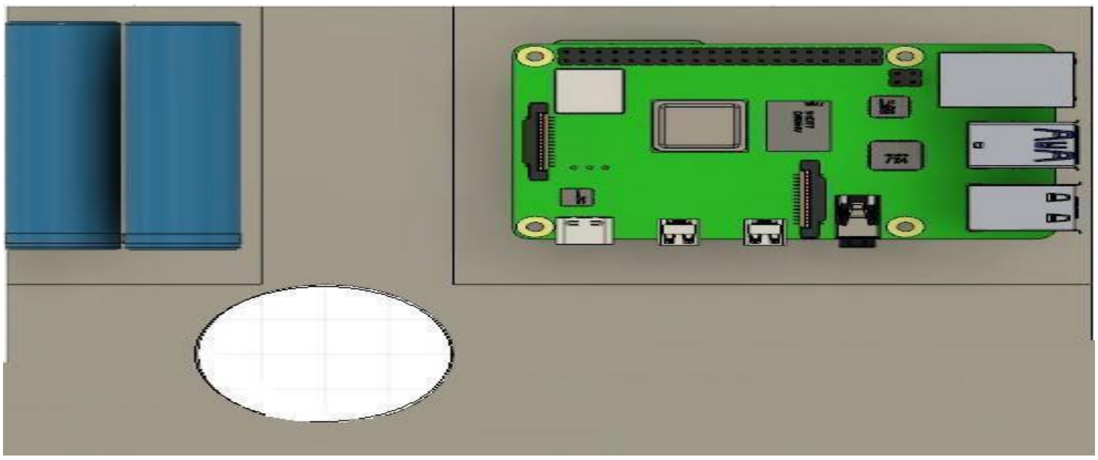


Figure 3.11 Top view of the hardware model designed on Fusion 360

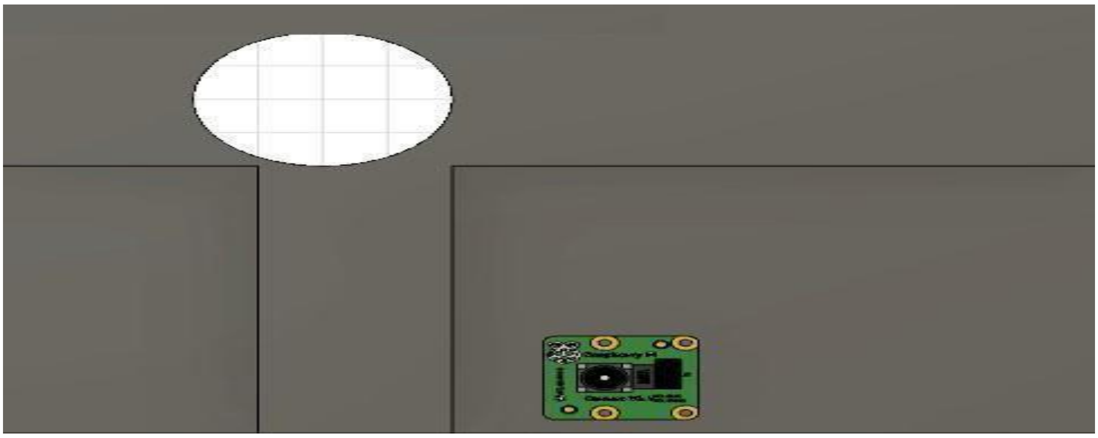


Figure 3.12 Back view of the hardware model designed on Fusion 360

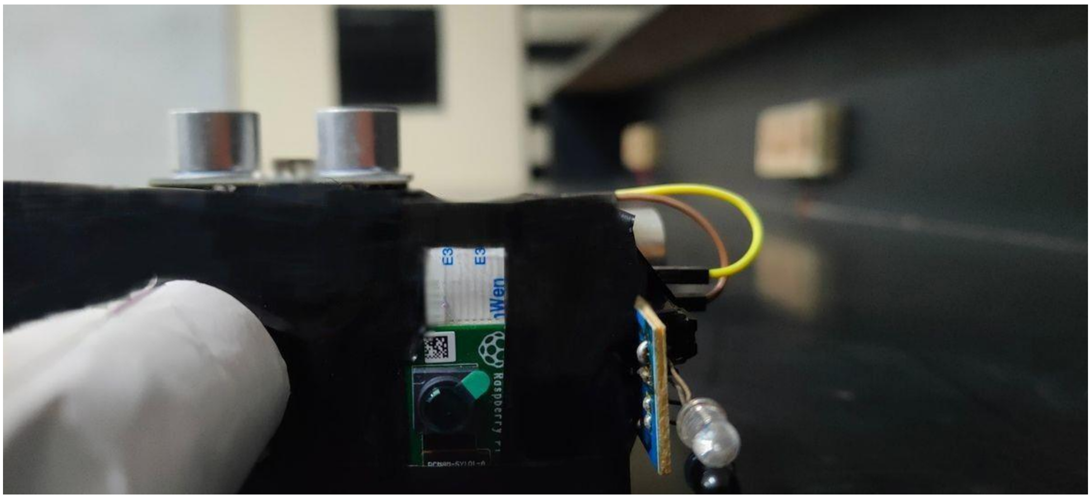


Figure 3.13 Actual 3D printed box with camera and proximity sensors

Figure 3.14 3D box connected with the conventional Stick

The front view consists of 2 openings in the design for camera input and the ultrasonic sensor. The opening for the camera is 1 cm and that for the ultrasonic sensor consists of 1.5 cm each. The rear view consists of spacing for capsulation of 4 objects namely ultrasonic sensor (45 mm X 20 mm X 1.5 mm), camera (25 mm X 23 mm X 9 mm), Raspberry pi (56.5 mm X 85.6 mm X 17 mm) and the batteries (50 mm X 29 mm). The model is worn on the head of the visually impaired person and rested on the nose of the user. The weight of the model is 240 grams which is easily worn all long day by the VI person.

Whereas for Smart Cane, a 3-D model is designed for our existing stick has dimensions of 65 mm X 120 mm X 50 mm. It will incorporate a Raspberry Pi (56.5 mm X 85.6 mm X 17 mm) and batteries (50 mm X 29 mm). The model is designed for a fixed diameter (25 mm) stick which is depicted in Figure 3.14. It will enhance more visibility of the scene.

3.3 Software Design

The system includes modules for acquiring images, pre-processing them, enhancing them, and annotating them.

A) Image Acquisition: The images are acquired through a camera model Raspberry Pi 3, which can record up to 60 frames per second (FPS) at 640x480p.

B) Image Augmentation: Acquired real time images were then enhanced using various methods such as flipping, brightness levels, noise levels, etc.

C) Image Annotation: A bounding box was placed around the detected objects after images were annotated with the LabelImg tool. During this process, images and bounding box positions were saved to a .xml file.

D) Data Sets: In terms of data sets, there are many existing datasets for path recognition such as PASCAL [252][253][254][255][256], but these are limited to a small number of classes.

E) Deep-learning model: There are many deep-learning models such as YOLO, SSD, FASTER-CNN, etc. and every technique has its advantages and disadvantages. With this model, YOLOv5-based model is utilized for detecting a road for people with disabilities.

3.3.1 Data Selection and Preprocessing

This matching was carried out using the COCO dataset, which is derived from [70].

The following features are included in the COCO dataset:

- Approximately 200K images are labelled of the total 330K
- It has almost eighty classes of object
- It contains 10 lakh class instances

To train computer vision models, the COCO dataset uses the MobileNet Convolutional Network. The model uses MobileNet v2 which has improved performance over its predecessor in every aspect. A significant improvement over MobileNet v1 is its speed, which is approximately 35% faster.

3.3.2 Design Specification

In the context of computer vision and machine learning, latest algorithms and developments and decided to implement the YOLO algorithm is discussed. Some algorithms have been evaluated, but YOLO is deemed to be the fastest and to offer the most robust capability for object detection [22][24][50][74][57]. As a result of successive improvements since YOLOv1, YoloV3 is now an improved algorithm. In comparison to YOLO V1 and V2, YOLO V3 overcomes the limitations of those versions. In addition, the algorithm provides better results regarding robustness and finding distinct objects within an image. To train the model, MobileNet Network was used. Pre-trained weights are stored in the COCO dataset for the pre-processing of images. Tensorflow is used to implement the YOLO algorithm [77].

3.3.3 Steps in YOLO

This application allows blind people to capture images with the rear camera in real-time. The YOLO algorithm is then applied to identify objects and announce them, shown are the steps required to recognize objects in real-time by the YOLO algorithm.

3.3.3.1 Image Classification

The process of classifying images consists of predicting what are in the image, for ex- Categorizing objects based on their appearance. Different objects such as human, car, dog, etc will give a value based on the likelihood of object appearing in any given image are classified.

3.3.3.2 Localization

Once the classification is performed on the different set of images, the next step is to locate these objects in the image. The process of localizing an object in an image is called as localization. YOLO can segment input images into the desired sizes by dividing them into NxN grids. There is a correlation between the grid size of the input image and the detection of an object. This leads to the term "regions" being used for these grids. By predicting the number of anchor boxes for a particular object, these regions can be used to determine how many anchor boxes will be drawn for it [68]. There are multiple actor boxes, each of which defines an object's boundaries. As long as there is a repeated localization and classification of the image.it would be possible to predict key objects present in the environment.

3.3.3.3 Non-Maximal Suppression

It is common for an object identification algorithm to identify the same object several times. To solve this problem, YOLO uses an algorithm called Non-Maximal Suppression. The bounding boxes are analyzed for confidence values and those with the greatest confidence values are kept, while those with low confidence values are removed. By repeating these steps, you can ensure that one object will only be identified once.

3.4 Class Diagram

Class diagram refers to the model used in the application as shown in figure 3.15. It includes classifiers, detectors, camera activity and algorithm model.

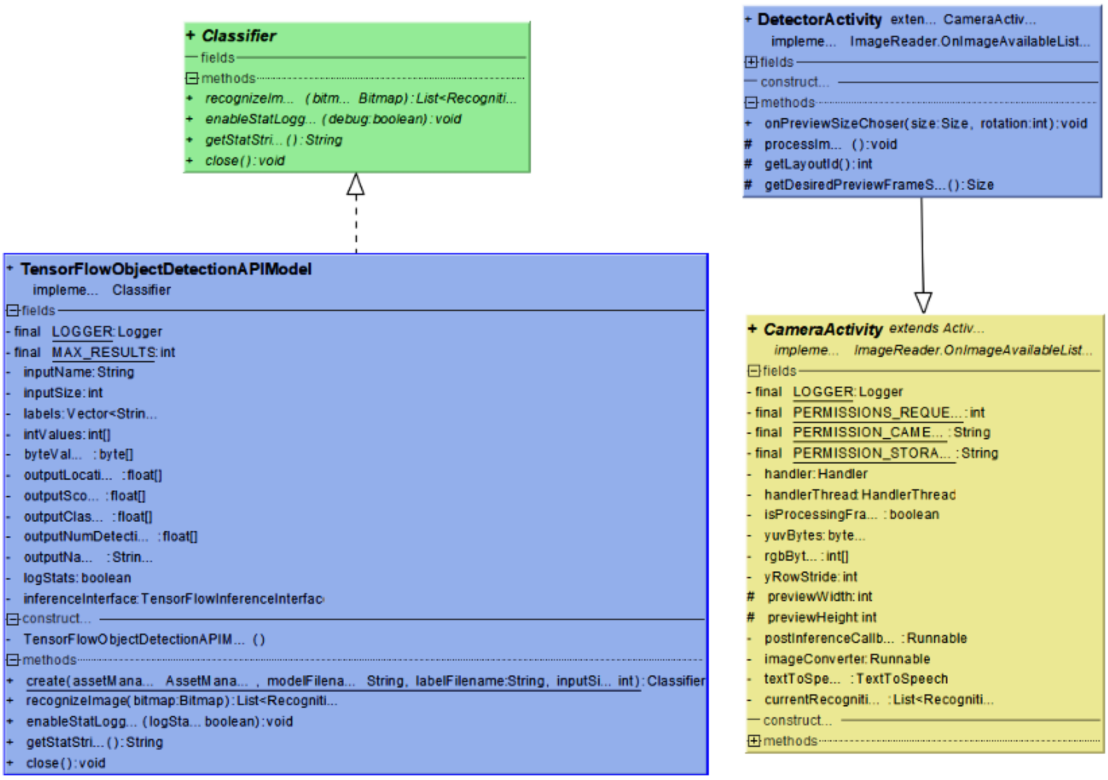


Figure 3.15: Class Diagram of the application

3.5 Data Acquisition

This project uses the Raspberry Pi 3 Model B+'s has general purpose input and output pins to detect an obstacle and the distance to it, as well as the ultrasonic sensor to use as a proximity sensor, it measures the distance between the user and the obstacle. GPIO pins on the Raspberry Pi are connected to four pins in the ultrasonic sensor. Raspberry Pi's VCC pin is connected to pin 2, its GND pin is connected to pin 6, its TRIG pin is connected to pin 12 which is GPIO18, and the ECHO pin is connected to pin 18 that is GPIO24. By analyzing the real-time data from the proximity sensor i.e., ultrasonic sensor, our program can detect an obstacle in real time and locate it automatically.

By using a video stream from the Pi camera, object detection data can be collected. Raspberry Pi camera modules are attached to the Raspberry Pi using the camera serial interface. The Raspberry Pi camera has a fixed focal length lens and has been designed to be onboard. The camera allows you to capture videos with 640x480 pixel resolution. Using RGB data retrieved from each video frame, our program recognizes objects from every frame that is already known by the system in real-time.

A video stream is collected by the Pi camera as part of the object detection process. With the dedicated camera serial interface, the Pi camera can be attached to the Raspberry Pi module. A fixed-focus lens Pi camera was designed for Raspberry Pi, which can be mounted onboard. This device is capable of capturing high-quality video of 640 pixels by 480 pixels. RGB data in real-time from every video frame, and our system recognizes objects based on that information.

3.6 MODULE DESCRIPTION

The system can be divided into three modules

➢ Module **1**: Detect Obstacles and their Distance

➢ Module **2**: Object Detection

➢ Module **3**: Generate Audio Feedback

3.6.1 Detect Obstacles and their Distance

To determine the distance between the user and an obstacle, ultrasonic sensor (HC-SR04) is used. In general, the ultrasonic sensor output (ECHO) will always give a LOW voltage, unless it is triggered, at which point it will output as HIGH voltage i.e. 5 Volt. Consequently, the sensor will be triggered by one GPIO pin and the ECHO voltage change will be detected by the other. As a result, the HC-SR04 must be triggered by a short pulse before the module will operate. In order to obtain an echo response, the sensor starts generating 10 kHz ultrasound bursts. Thus, the input pulse (trigger) is generated by setting the trigger pin from LOW to HIGH for time t (10) seconds and then setting it to LOW again. ECHO pin is set from LOW to HIGH the duration of the travel time for the pulse as soon as the sensor detects the traveling pulse and reflected signal. Once the echo pulse is received, the signal will change from 0 to 1 and remain at 1 throughout the pulse. Measurement of distance is done by calculating the difference between recorded time stamps between the ultrasound source and the reflecting object. Ultrasonic waves travel at a certain speed in different mediums and at different temperatures. The measured waves speed is 34300 cm/s, which is the speed of sound at sea level.

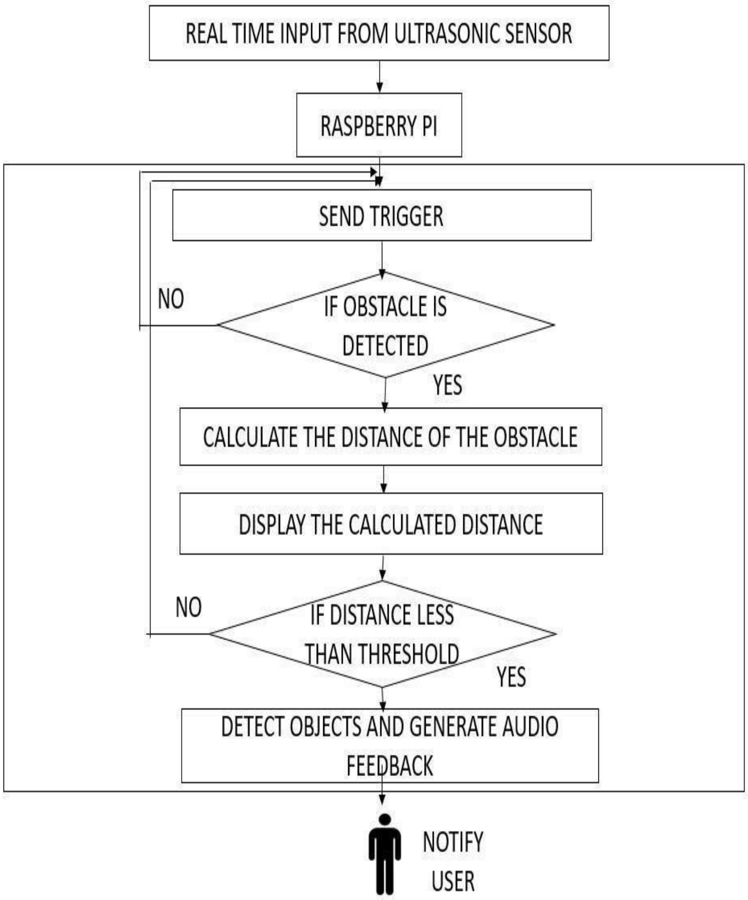


Figure 3.16 Flow diagram to detect obstacle and its distance

Therefore, the first step is to trigger the sensor, then calculate the distance of any obstacles detected before sending a trigger again. After calculating the distance, is displayed to the user. As shown in figure 3.16, the system detects the type of obstacle and elicits voice feedback when it comes across one within the given threshold value. This provides the blind person with information about their proximity to obstacles, alerting him or her to the danger and helping to avoid any potential accidents.

3.6.2 Object Detection

An object detection algorithm called You Only Look Once (YOLO) is used in this project. The algorithm You Only Look Once (YOLO) is the fastest object detection algorithms used in real-time detection. The YOLO algorithm of Darknet possesses 53 layers by default, but for detection purposes another 53 layers are added, making 106 layers in total. As a feature of YOLO, it samples the input image dimensions by 32,16 and 8 and detects objects at three different scales, sizes and locations within the network. This results in a detection kernel whose dimensions are 1 x 1 x 255.

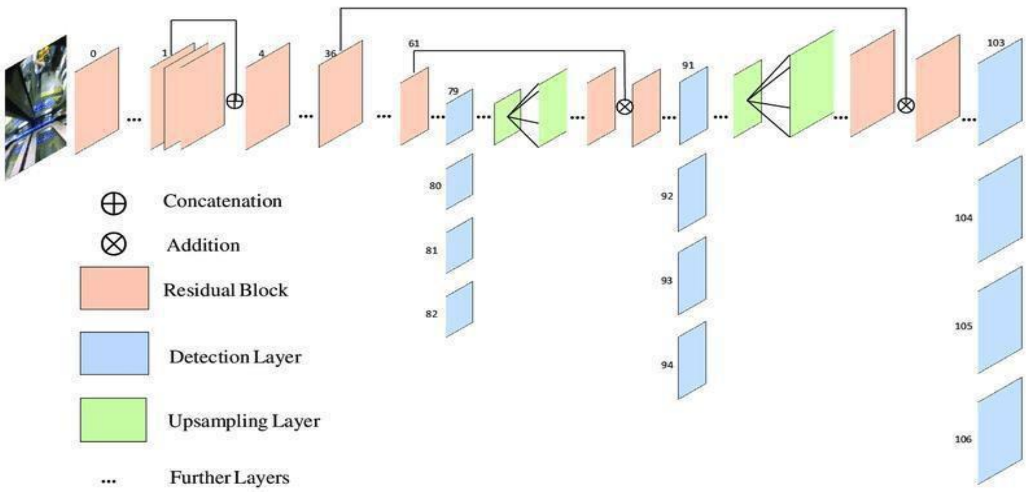


Figure 3.17 YOLO Algorithm Architecture

The first detection is made by the 82nd layer. As shown in figure 3.17, the first 81 layers of an image are down-sampled by the network, each layer having a stride of 32.

An image with a dimension of 416 x 416 will be transformed into a map of 13 x 13 based on input of the size of the image. With the 1 x 1 detection kernel, a single detection is made which giving us a 13 x 13 x 255 detection feature map.

This is followed by up sampling by two times to 26 x 26 dimensions of the feature map from the 79th layer before applying convolutions which then concatenate the feature map from the 61st layer to create a new feature map. The combined feature maps are then convolutional with 1x1 layers to combine the features from the earlier layers. A second detection is produced by the 94th layer, producing a feature map with 26 x 26 x 255 dimensions.

Additionally, the 36th layer feature map is concatenated with the 91st layer feature map after being subjected to convolutional layers. Following that, a 1 x 1 convolutional layer is applied to combine the previous layers information. Finally, the 106th layer is employed for final detection, resulting in a 52 x 52 x 255 feature map. Detection of large objects is performed by the 13 x 13-layer, detection of smaller objects is performed by the 52 x 52 layer, and detection of medium objects is performed by the 26 x 26 layer. Objects with multiple labels are predicted using a threshold value, which is calculated by logistic regression. Objects are assigned the classes that have the highest scores compared to the threshold.

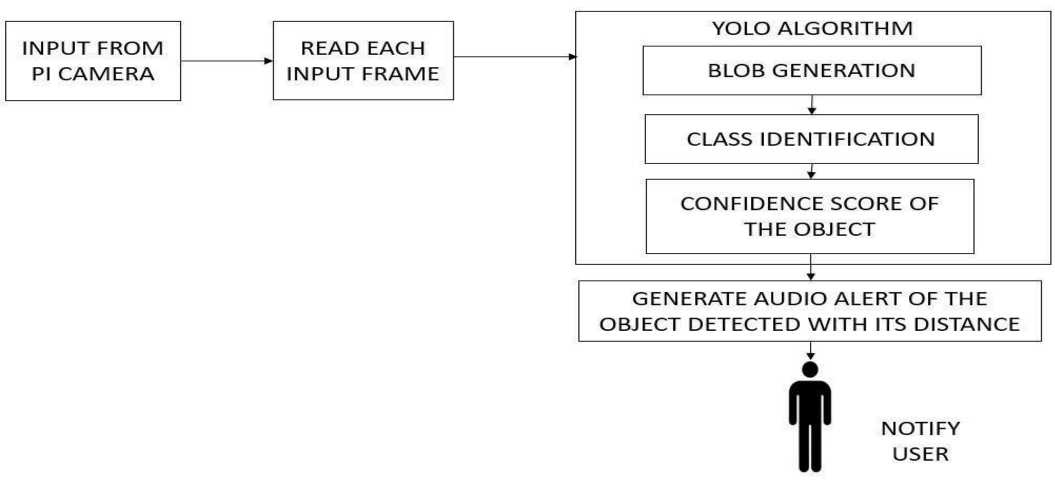


Figure 3.18 Object Detection in Raspberry Pi

The Pi camera is used in real time for object detection. In order to detect objects in video inputs, each object must be locally located in every frame. Object detection occurs on each video frame from the Pi camera after setting the paths to the model and label maps, loading the model into memory, and initializing the Pi camera. In order to detect common objects inside the user's view, the system detects the objects within the user's view as shown in figure 3.18.

3.6.3 Algorithm: Object Detection using YOLO

Input: Real time video stream from Raspberry Pi

Output: Label of the detected object

1. Load the existing YOLO model.
2. Process input stream frame by frame.
3. Initialize the variables class_id and confidence score.
4. Find the dimensions of each current frame.
5. Detect blobs in the frame.
6. Apply blob as input to the YOLO model.
7. Class_id is set to index of maximum value in the list of objects.
8. Confidence score is set to the highest value from the list.
9. If the confidence score is greater than threshold value, then the object 's class_id and their respective confidences of blobs are stored as a list.
10. Display the label of object detected with its confidence score.

3.6.4 Pseudo-code for Object Detection

The Pseudo-code for the model is wriiten in Python 3.6 version as shown in figure 3.19.

Running inference

```
# run inference through the network
# and gather predictions from output layers
outs = net.forward(get_output_layers(net))

# initialization
class_ids = []
confidences = []
boxes = []
conf_threshold = 0.5
nms_threshold = 0.4

# for each detetion from each output layer
# get the confidence, class id, bounding box params
# and ignore weak detections (confidence < 0.5)
for out in outs:
    for detection in out:
        scores = detection[5:]
        class_id = np.argmax(scores)
        confidence = scores[class_id]
        if confidence > 0.5:
            center_x = int(detection[0] * Width)
            center_y = int(detection[1] * Height)
            w = int(detection[2] * Width)
            h = int(detection[3] * Height)
            x = center_x - w / 2
            y = center_y - h / 2
            class_ids.append(class_id)
            confidences.append(float(confidence))
            boxes.append([x, y, w, h])
```

Figure 3.19 Pseudo-code for Object Detection

3.6.3 Generate Audio Feedback

In order to provide auditory feedback for the object type and distance between the object and the user, Raspberry Pi is used with the open-source speech synthesizer eSpeak for text-to-speech. A cross-platform text to speech library called pyttsx3 is also used in our system, along with eSpeak. Neither Python version 2 nor Python version 3 are compatible with it. This library works even in offline mode, making it an ideal solution for text-to-speech conversions.

Initializing the text-to-speech engine using the pyttsx3.init() function is the first step. You can then use the say () function to convert the detected object and its distance into an audio recording. In order to listen to the speech, the runAndWait() function makes it audible in the system. Without this function, the speech cannot be heard by the end user. In this way, the blind person is alerted to his or her distance from obstacles and is able to avoid any potential accidents by hearing the audio feedback generated.

3.7 Hardware and Software Requirements

3.7.1 Hardware

Raspberry Pi – Model 3 or above

Among the latest Raspberry Pi 3 models is the Raspberry Pi 3 Model B+ as shown in figure 3.20. The Broadcom BCM2837B0 is an ARMv8-based 64-bit Cortex-A53 SoC with a 1.4GHz clocked speed. It supports Bluetooth 4.2, BLE, 2.4GHz and 5GHz LPDDR2 SDRAM, IEEE 802.11b/g/n/ac wireless LAN, and 1GB of LPDDR2 SDRAM and communicate with a maximum throughput of 300 Mbps is achieved with Gigabit Ethernet over USB 2.0

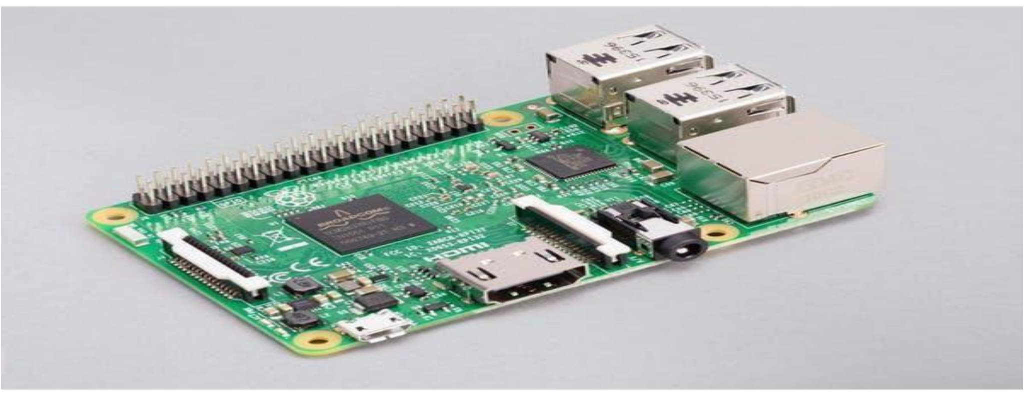


Figure 3.20 Raspberry pi 3 hardware

Ultrasonic sensor (HCSR04): An ultrasonic distance sensor, as shown in figure 3.21, such as the HCSR04, relies on sonar to detect distances to objects. It reads between 2 cm and 400 cm (0.8 inches and 157 inches) with an accuracy of 0.3 cm (0.1 inches).

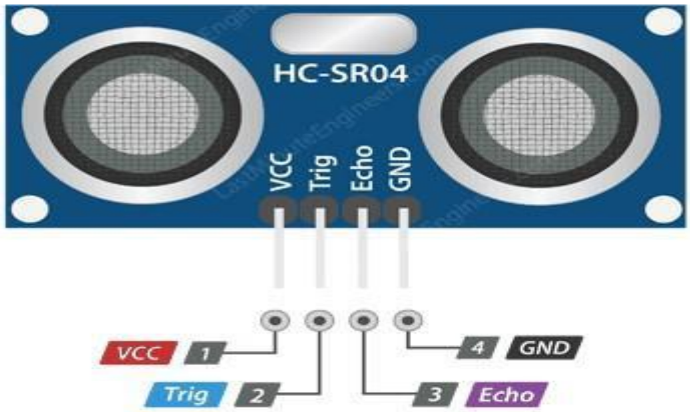


Figure 3.21 Ultrasonic Sensor

(HCSR04) Raspberry Pi camera:

The Raspberry Pi Camera module as shown in figure 3.22 can be used to take high-definition pictures and videos. It can be directly attached to Raspberry Pi Board via its CSI (Camera Serial Interface) interface. A 15-pin ribbon cable is used to connect the Pi Camera module to Raspberry Pi's CSI port.

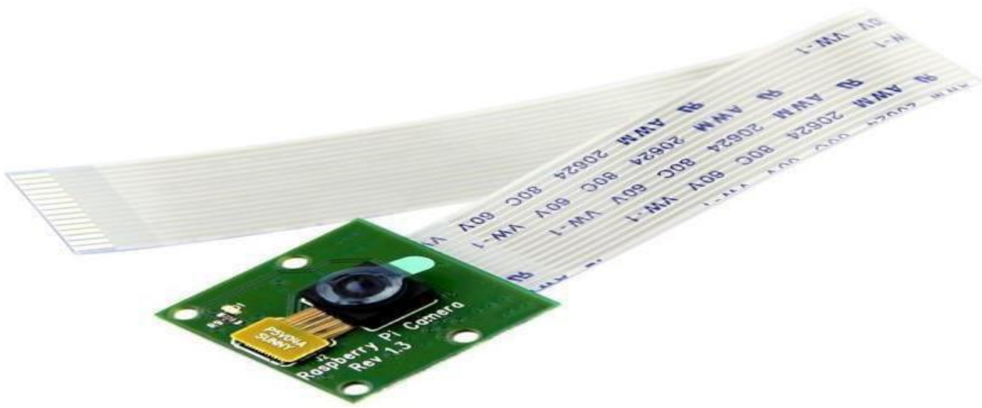


Figure 3.22 Raspberry pi camera

3.7.2 Software

Operating system - A Debian-based operating system for Raspberry Pi called Raspberry Pi OS (formerly Raspbian) is available for the device as shown in figure 3.23. As of 2013, it has been the official operating system for Raspberry Pi single-board computers provided by the Raspberry Pi Foundation.

Figure 3.23 Raspian OS in Raspberry pi 3

Programming language – Python

There are quite a few high-level and general-purpose programming languages available, but Python is one of the most popular. With the latest Python 3 version, Python programming language is being used for web development, machine learning applications, and for all cutting-edge techs used in the software industry.

Python IDE - Thonny IDE

This free Python-based Integrated Development Environment (IDE) was made specifically for beginners. Additionally, it includes a built-in debugger with which you can troubleshoot nasty bugs, as well as step through expression evaluation capabilities as shown in figure 3.24.

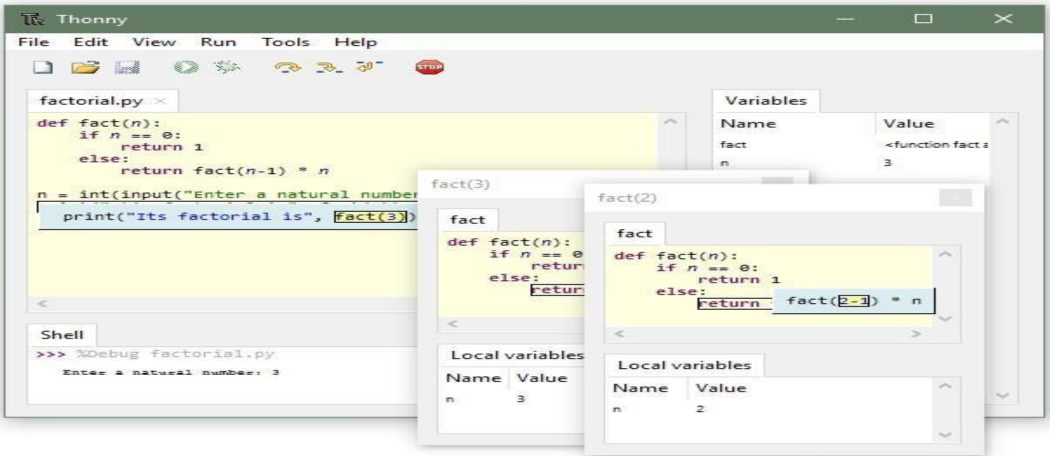


Figure 3.24 Thonny IDE in Raspberry pi 3

CHAPTER 4

IMPLEMENTATION AND RESULT

Our Proposed model is built through Raspberry Pi 3 Model B+ and is attached with an ultrasonic sensor interfaced with the Raspberry Pi 3 Model B+, other components such as camera attached to the Raspberry Pi 3 Model B+, headphones, and a power supply. In addition to evaluating each component of the system individually, the overall system is evaluated after all the components are assembled. Both obstacle detection and object detection were evaluated for the developed system.

In addition to an ultrasonic sensor and an RPi camera, headphones are included in this system and power supply for the Raspberry Pi 3. The individual components of the model were evaluated individually, and all the components had been assembled once the overall system is tested. Object detection and obstacle detection are evaluated with the developed system. A real-time test is conducted to ensure that the system is performing as intended.

4.1: Models Developed

As a result of the image taken by the camera, the model will be able to identify the objects. In addition to an audio notification of the names and locations of each object identified in the image, the output will include a label presenting the object's class and confidence level. Figure 4.1 shows the outputs of the application. The output shows various class model detection out of 80 pre-defined classes, such as remote, cup, Scissors, persons etc.

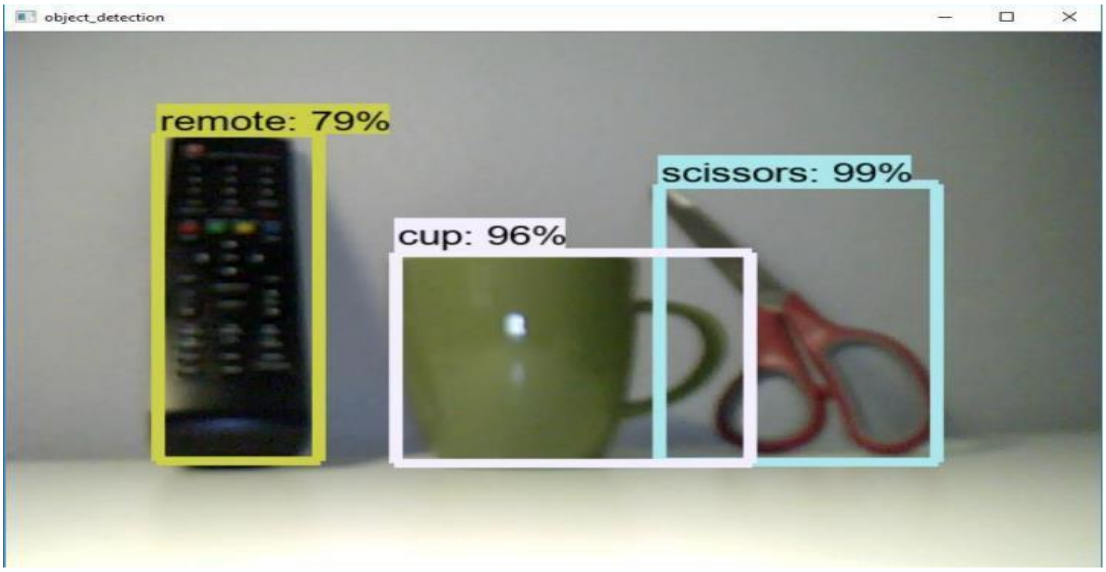


Figure 4.1 various identified objects

4.2 Code Developed

As the input image is processed, the algorithm uses a pre-trained dataset, called the COCO dataset, that has been trained using MobileNet-v1 Neural Network techniques to classify and locate all objects inside the image, iteratively.

After applying YOLO, a confidence score is calculated. Yolo uses 24 convolution layers that output to two fully connected layers. In a given image, this score indicates the probability of a certain object type being present. In this way, the class with the highest confidence is selected and the remaining classes are discarded. In this step, I convert the output class into speech by using a text-to-speech API. The implementation of text-to-speech for the detected objects can be seen in figure 5 below.

4.3 Finding relation between Object detection with the distance

In order to collect data, obstacles are placed in different orientations on an ultrasonic sensor. My calculation of the average value of data for each interval was based on three data points. Additionally, precision, error rate, standard deviation and variance are based on the observed data. This result represents the distortion in the observed distance compared to the actual distance resulting from the comparison between the actual distance and the observed distance.

The data is collected from single ultrasonic sensor by positioning obstacles in different orientation as shown in table 4.1, the model accuracy is calculated by averaging recorded data from each ultrasonic sensor 3 times with the same objects. The data is taken while the user is moving on a rounded bottleneck path which is guided by 2 side walls. From the data, it has been found that as the distance between the object and sensor increases the accuracy of the model decreases gradually. The average accuracy of the vision-based system is 98.6%.

In order to determine the distance that the sound has travelled, we use the formula: Distance = Time x Speed Of Sound / 2. Due to the fact that the sound must travel back and forth, the "2" is in the formula. In the first instance, the sound travels away from the sensor, then it bounces off a surface and returns.

Table 4.1 Data received from single ultrasonic sensor.

Actual Distance (cm)	Measured Distance (cms) Centimeters= Microseconds / 2*29.			Average	Accuracy (%)
	1	2	3		
25	24.64	24.3	24.8	24.58	98.32
50	49.25	49.75	48.9	49.30	98.6
100	98.70	98.81	98.5	98.67	98.67
150	148.6	148.8	147.26	148.22	98.81
200	198.8	198.6	197.14	198.18	99.09

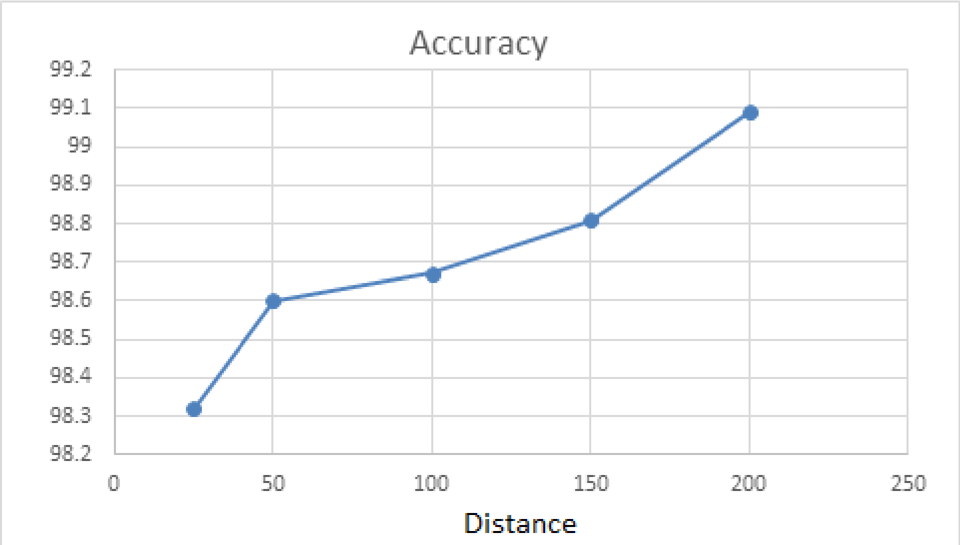


Figure 4.2 Graph between distance and accuracy

As depicted in figure 4.2, the highest accuracy turns out to be 99.09 when the user is 200 cm from the obstacle and the level of accuracy level falls when the user moves closure to the obstacle The accuracy is worst for the cases for very short distance less than 3 cm.

Table 4.2 shows variation in Standard Deviation and Variance with distance.

ACTUAL DISTANCE	AVERAGE OF OBSERVED DISTANCE	ERROR	STANDARD DEVIATION $\sigma = \sqrt{1/N \sum_{i=1}^{N} (x_i - \mu)^2}$	VARIANCE $1/N \sum_{i=1}^{N} (x_i - \mu)^2$
25	**24.58**	**1.68**	**0.27856**	**0.07**

50	**49.30**	**1.40**	**0.36**	**0.13**
100	**98.67**	**1.33**	**0.412**	**0.17**
150	**148.22**	**1.18**	**0.53**	**0.28**
200	**198.18**	**0.91**	**0.78**	**0.61**

Based on the table 4.2, it can be seen that the deformity is not severe, and the observed distance is acceptable. As the actual distance is increased, the distortion value rises in a positive direction. In general, accuracy increases as the distance increases.

Based on the table data, one can observe that the error rate decreases as distance increases. Among the different measures of deviation, standard deviation and variance are closely related. Variance is measured as the difference between each value and the mean. As a result of the variance value, the overall system has a greater range of data. As can be seen from the table data, standard deviations and variances are lowest when hindrances are close to users. A decrease in the distance between obstacles leads to an increase in these values. By using the sensor, the system is able to obtain the least standard deviation and variance of 0.21 and 0.07, respectively. A comparison between the actual distance and the observed distance can be found in figure 4.3.

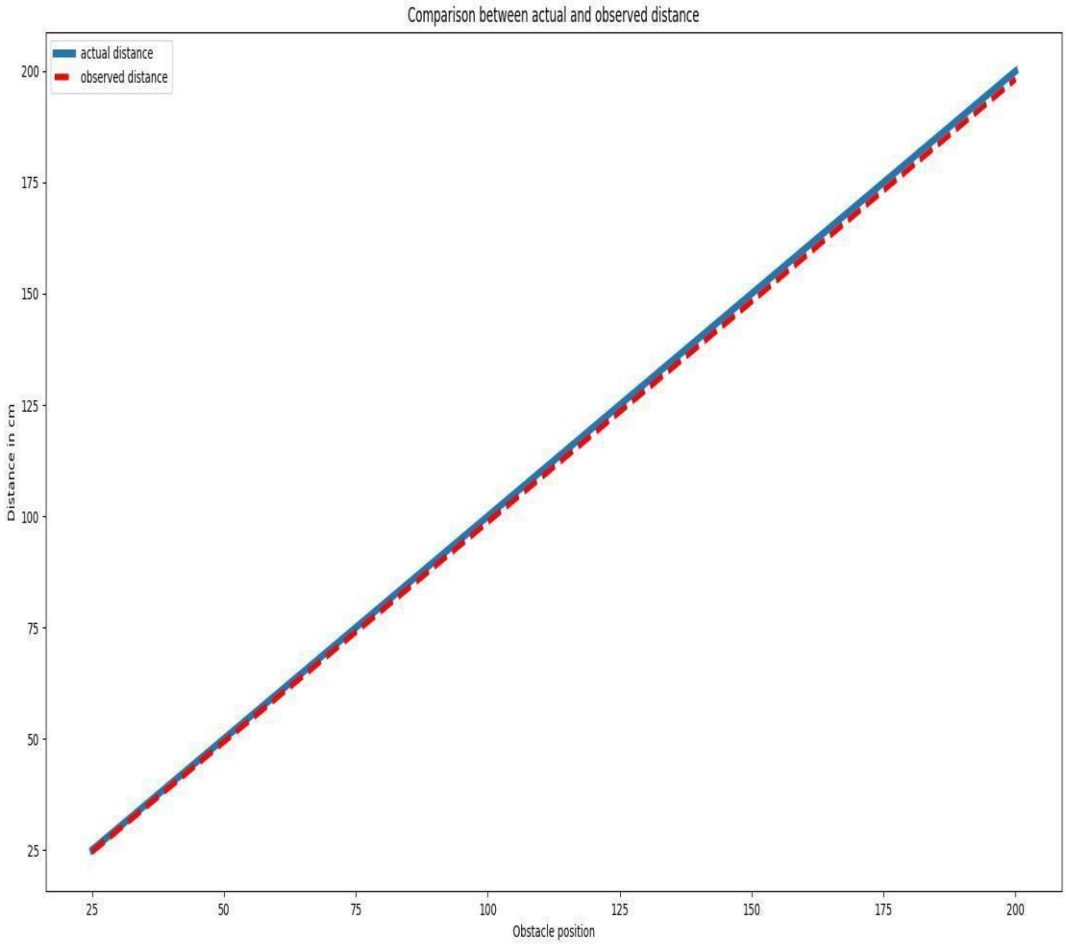


Figure 4.3 Comparisons between observed and actual distance

4.4 Detection of Object

Input from the Raspberry Pi camera captures a real-time video stream which is then split into various frames, and a width and height are determined for each input frame. As part of this system, objects are detected using the YOLOv5 algorithm. To obtain the correct prediction, a blob from the input frame is obtained using the OpenCV function blob From Image (), then it is sent to the YOLO pre-trained model Using each output layer's detection, calculation of the confidence score and class label are made. The final detection is determined by ignoring the objects with a confidence score of less than 0.6 and applying non-max suppression. In order to accurately identify the object, the class probability and confidence score are required.

4.5 Analysis of results with different datasets.

4.5.1 Analysis with COCO dataset:

The datasets used in this research work were collected from COCO (Common Objects in Context) data sets. It is one of the largest scale object detection datasets which comprises 330K images with 80 object categories some of them are listed as a person, bus, train, bicycle, car, etc. The other reliable datasets which can train our model are as follows COCO, COCO128, VOC, Argoverse, VisDrone, GlobalWheat, xView, Objects365, and SKU-110K. Figure 4.4 shows features of the COCO dataset and its collaborators.

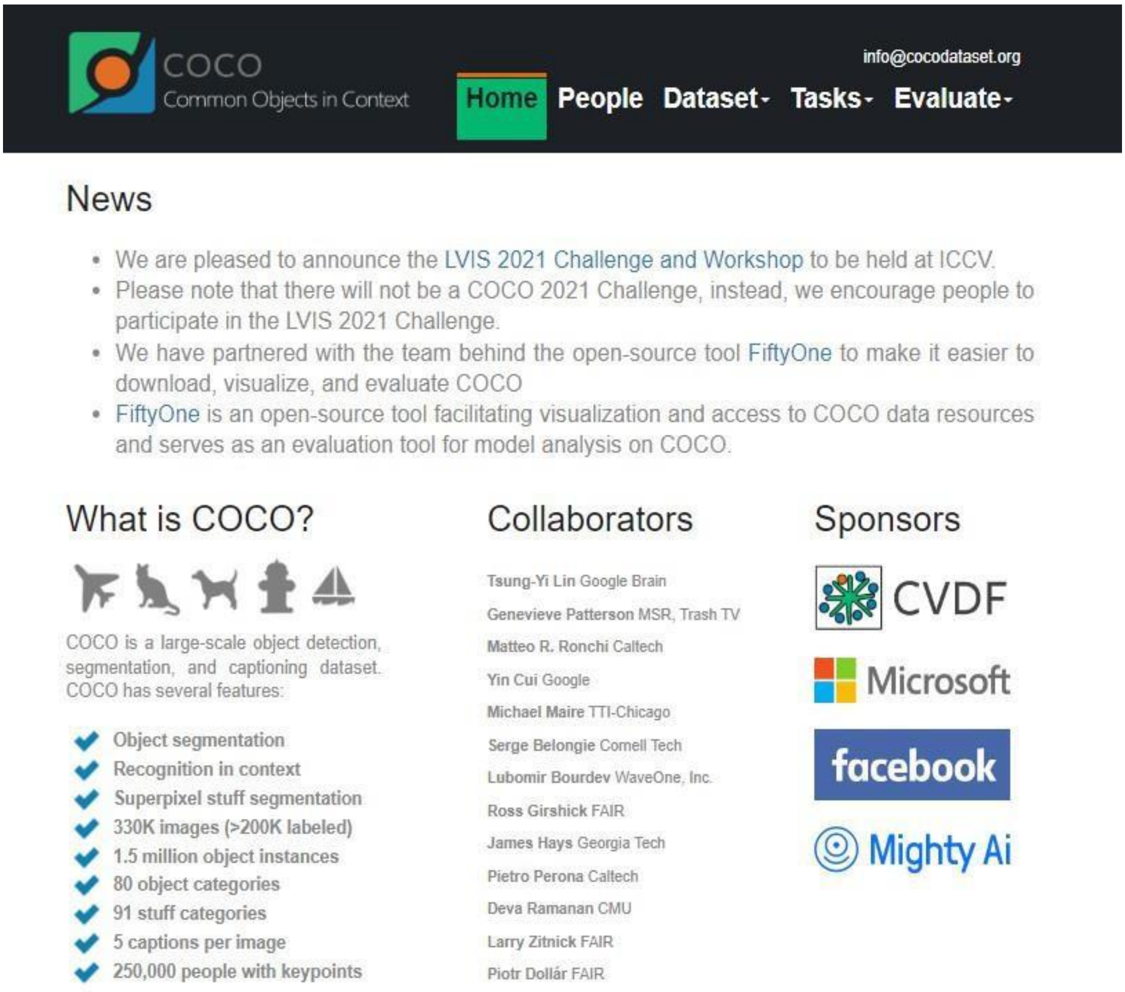


Figure 4.4 Features of the COCO dataset and its collaborators

person	fire hydrant	elephant	skis	wine glass	broccoli	dining table	toaster
bicycle	stop sign	bear	snowboard	cup	carrot	toilet	sink
car	parking meter	zebra	sports ball	fork	hot dog	tv	refrigerator
motorcycle	bench	giraffe	kite	knife	pizza	laptop	book
airplane	bird	backpack	baseball bat	spoon	donut	mouse	clock
bus	cat	umbrella	baseball glove	bowl	cake	remote	vase
train	dog	handbag	skateboard	banana	chair	keyboard	scissors
truck	horse	tie	surfboard	apple	couch	cell phone	teddy bear
boat	sheep	suitcase	tennis racket	sandwich	potted plant	microwave	hair drier
traffic light	cow	frisbee	bottle	orange	bed	oven	toothbrush

Figure 4.5 Classes used for classifications through the COCO dataset.

Various Classes of objects through the COCO dataset is shown in Figure 4.5 and Figure 4.6. The classes are default in general. Custom classes is preferred rather than default classes.

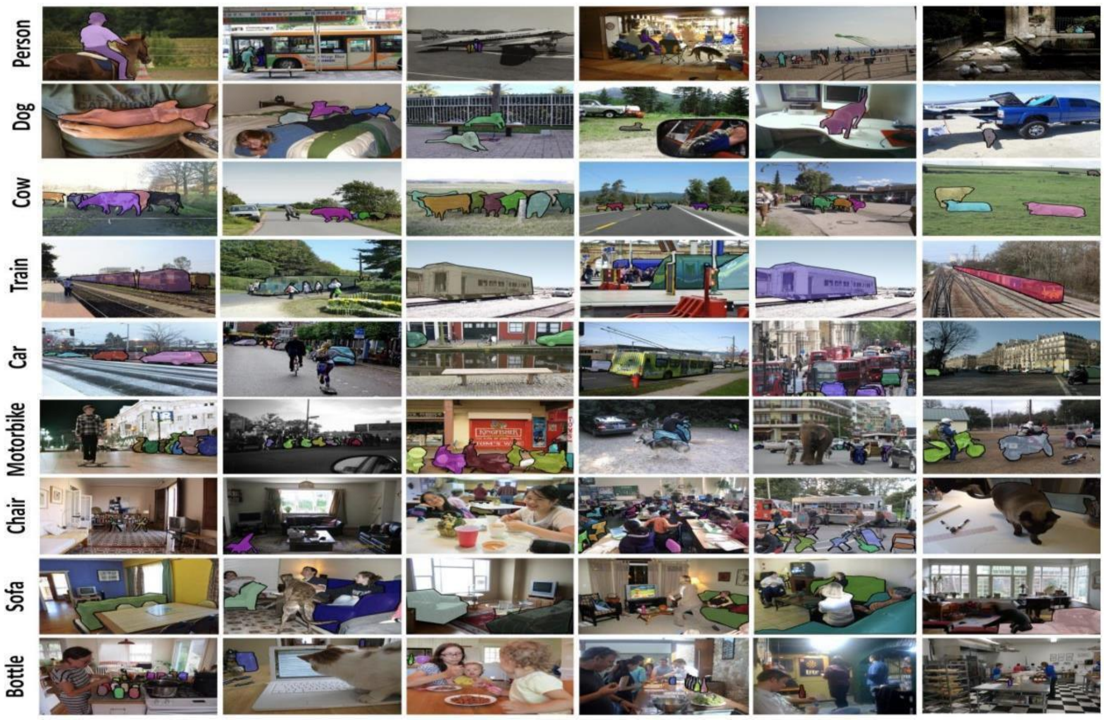


Figure 4.6 Object detection for various real-time frames.

As depicted in table 4.3, the model is tested for various known environments and the outcomes are among the 80 classes stated by the COCO model i.e. Traffic lights, cars, persons, trucks, cycles, etc. The amount of time taken to analyze this data varies from less than 1 sec to 3 sec, depending on the number of objects detected in the environment.

Table 4.3: Result analysis with known classes

Real-Time Environment data	Object detection in the environment	Time taken	Class Type	No of Classes
		2.874795 Seconds	Known	4 Classes namely: 'Car', 'Traffic light', 'Motorbike', 'Person'
		1.765487 Seconds	Known	4 Classes namely: 'Persons', 'Bus', 'Bicycle', 'Car'
		1.765487 Seconds	Known	3 Classes namely: 'Car', 'Traffic light', 'Truck'

4.6 Analysis with Manual Dataset:

The optimal distance between the object and the wearable device mounted on the visually impaired to around 1-2 meters. YOLOv5 model requires labeled data which comprises of class-label and position of all ground truth in images which could be automated using annotation which reduces further errors. With our data precision of 69.65, Recall 80.67, F1 as 74.77 and mAP as 80.69 % is determined as shown in figure 4.7.

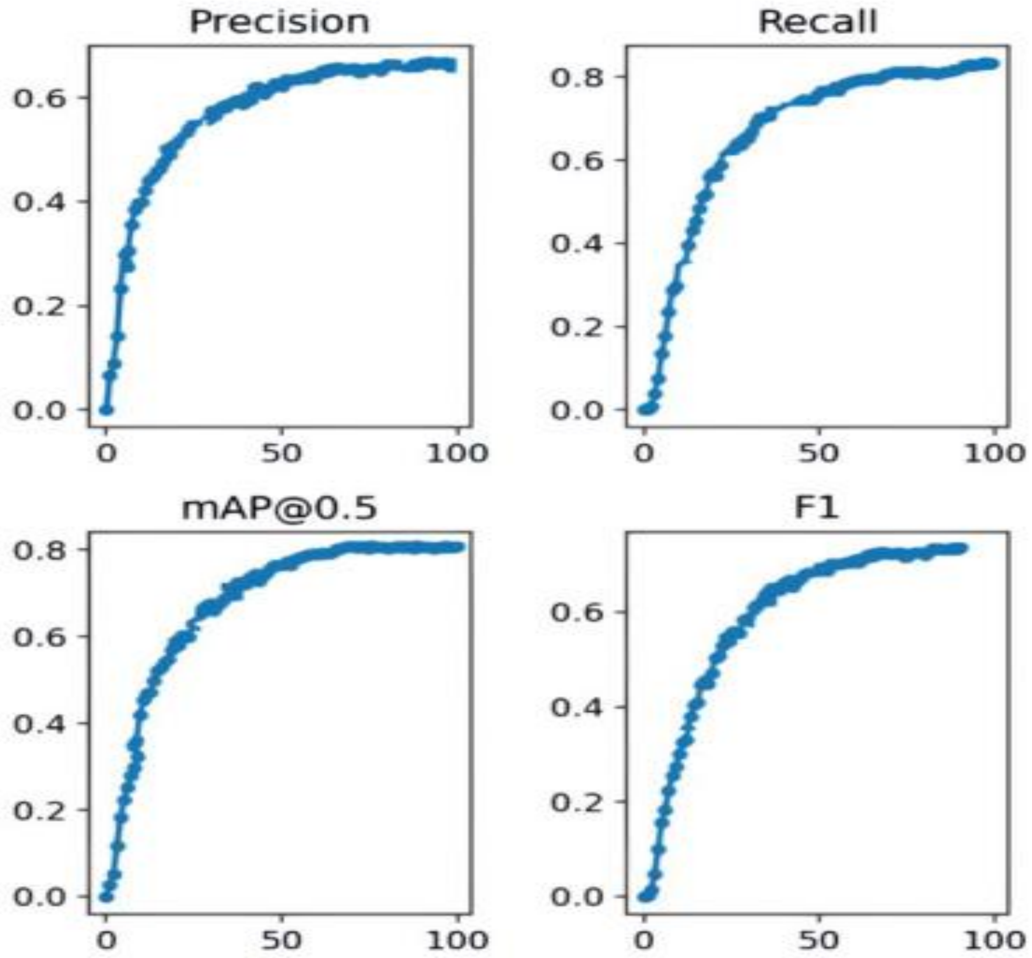


Figure 4.7: Graph of losses, Recall, Precision, and mAP with data training.

4.7 Training data:

For our model, 100 images each of the 3 classes namely Straight-Path, Left-Path, and Right-Path were manually captured. The images are captured in the daytime, through these 300 images dataset is divided into 90% for training the model and 10% is kept for testing the model. The images are then annotated our images in YOLO format (.yml) and trained across 60 epochs. In addition to enhancement of the images such as scaling, transformation flipping and data augmentation techniques were performed on the data as well as shown in figure 4.8.

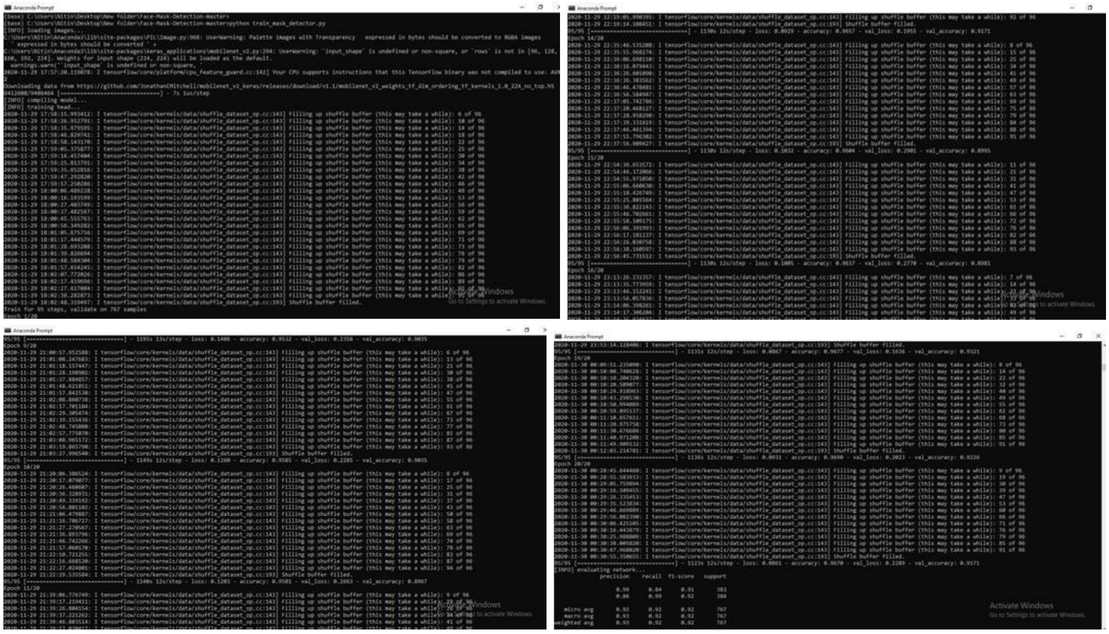


Figure 4.8: Training of dataset from known classes

Evaluation Parameter:

mAP (Mean Average Precision):

The Average Precision (AP) matrix is a popular measure of object detection accuracy. AP measures how different algorithms perform with the same dataset when it comes to accuracy. As a result, it is possible to compare the algorithm's performance with those of other algorithms. In order to calculate mAP, specific steps must be followed in order to arrive at the mean value of accuracy. An Intersection over Union calculation (IoU) determines the overlap between two bounding boxes, where one is the assumed box (Groundtruth) where the objects are located.

$$IoU = \frac{Prediction\ Result\ \cap Detection\ Result}{Prediction\ Result\ \cup Detection\ Result} = \frac{area\ 1}{area\ 1 + area\ 2 - area}$$

As part of Precision and Recall, True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) must be specified, and this is done by setting an IoU limit. Depending on the number of classes in the model, the limit may be reached.

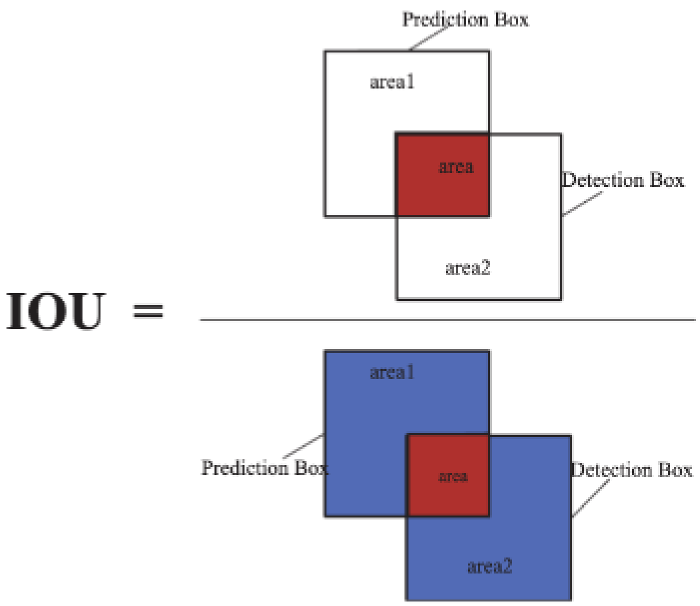


Figure 9: Intersection over Union (IoU)

For example, COCO with 80 classes should run number of times with AP @ [0.5: 0.05: 0.95], i.e. IoU = 0.5, IoU = 0.05 and IoU = 0.95 etc. Assuming that IoU = 0.95 means the following: IoU calculation aims to measure how equal two locations are, and when IoU approaches the value 1 they are identical. By giving a threshold IoU value, IoU decides whether two boxes are the same or not (when, for example, IoU is above 0.95). Thus, if one box is groundtruth, it is possible to determine the values of TP, FP, FN, and TN based on how often the algorithm's boxes have failed to predict the proper object.

- IoU $\geq$ 0.95 "True Positive"(TP)
- IoU $\leq$ 0.95 "False Positive"(FP)
- "False Negative"(FN) when the algorithm failed to detect objects
- "True Negative"(TN) when the algorithm detects the wrong object in the image.

The precision, recall, and accuracy are formulized as following:

$$Precision = \frac{\text{True Positive(TP)}}{\text{True Positive (TP)+ False Positive (FP)}}$$

$$Recall = \frac{\text{True Positive(TP)}}{\text{True Positive (TP)+ False Negative (FN)}}$$

$$\text{F1 (score)} = \frac{2\ (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

After calculating precision and recall, plotting shall occur, where Y-axis represents precision and recall for X-axis. By calculating the area under the graph, AP is calculated. This results in all data being available to calculate mAP.

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} APi$$

4.8 Analyzing data: The algorithm works based on the following four approaches:

- Residual blocks
- Bounding box regression
- Intersection Over Unions or IOU for short
- Non-Maximum Suppression.

Let's have a closer look at each one of them

1. Residual blocks: For analysing the image, the image has to be converted into a grid of cells. A grid of S X S cells is used to split each frame of the image into cells responsible for prediction. The cells in the grid determine the class of the object covered by their cells and the probability/confidence value associated with them, as shown in figure 4.9.

Figure 4.9: Residual Block over chosen path

2- Bounding box regression:

A bounding box outlines all the objects in an image and corresponds to rectangles defining the bounds. A bounding box can contain as many objects as there are in the image. The YOLO method determines the boundaries of bounding boxes using the following regression module, where Y represents the bounding box's final vector representation as shown in figure 4.10.

Y = [p_c, bx, by, bh, bw, c1, c2]

- p_c corresponds to the probability score of the grid containing an object. For instance, all the grids in red will have a probability score higher than zero. The image on the right is the simplified version since the probability of each yellow cell is zero (insignificant).

- bx, by are the x and y coordinates of the center of the bounding box with respect to the enveloping grid cell.

- bh, bw correspond to the height and the width of the bounding box with respect to the enveloping grid cell.

- c1 and c2 correspond to the two classes Player and Ball. As many classes can be added as your use case requires.

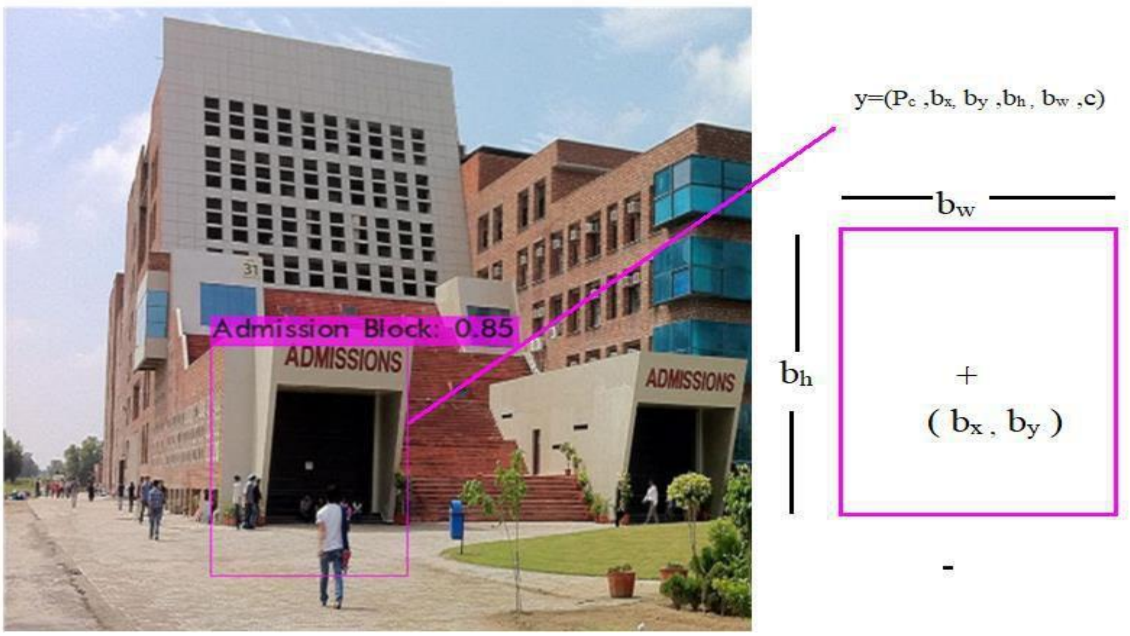


Figure 4.10: Bounding Box over chosen path

3- Intersection Over Unions or IOU:

In most cases, even though not all grid box candidates are relevant for prediction, one object in an image may have multiple grid box candidates. An IOU (a value between 0 and 1) discards those grid boxes that are irrelevant, and keeps only those that are relevant. This is how it works:

- The user defines its IOU selection threshold, which can be, for instance, 0.5.
- Then YOLO computes the IOU of each grid cell which is the Intersection area divided by the Union Area.
- Finally, it ignores the prediction of the grid cells having an IOU ≤ threshold and considers those with an IOU > threshold, as shown in figure 4.11.

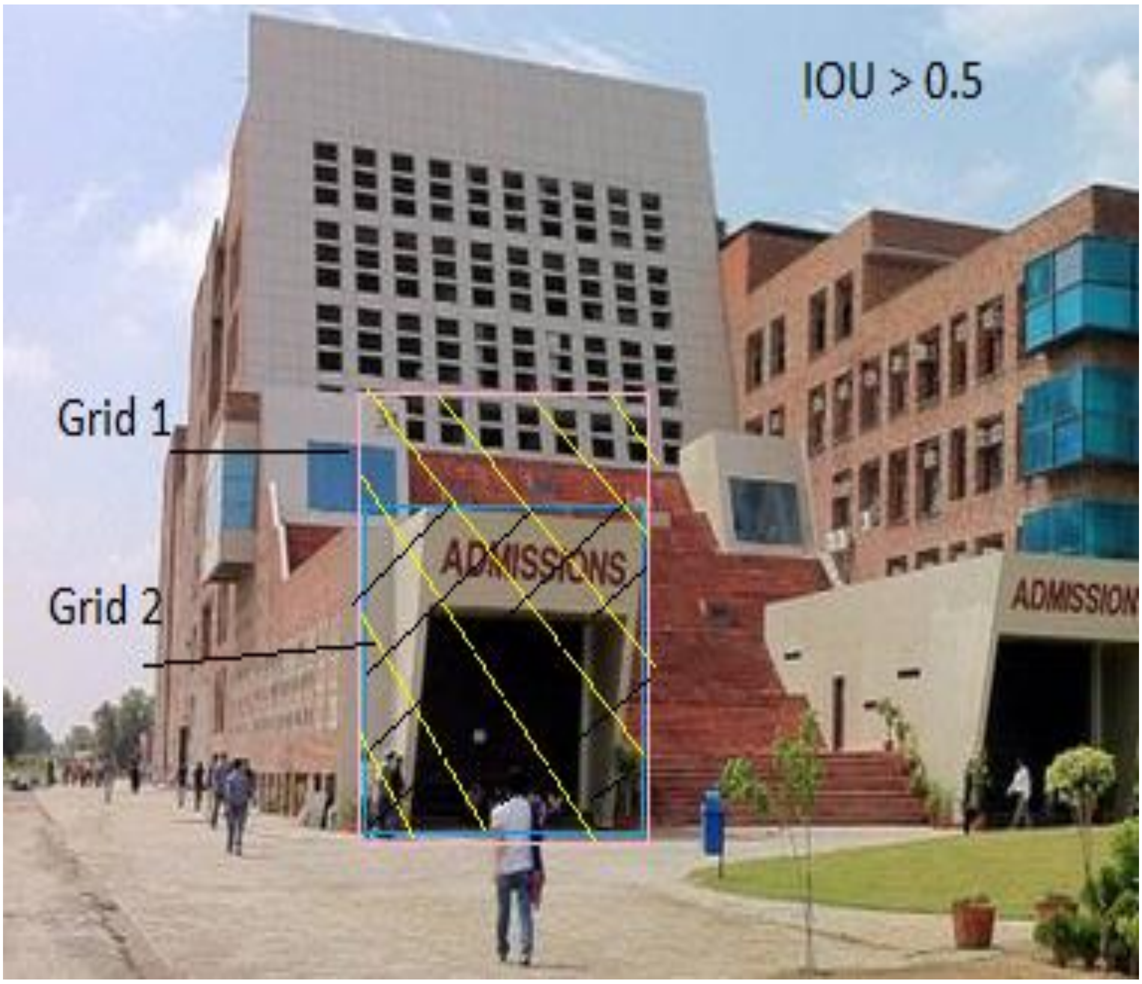


Figure 4.11: IOU value over chosen path

4- Non-Max Suppression or NMS

When an object has multiple boxes with IOUs above a threshold, setting a threshold is not always sufficient because leaving all of those boxes might generate noise.

Here NMS can be used to keep only the boxes with the highest probability score of detection.

The confidence score in the image can be given by equation (1)

$$\text{CS}= P_r(Obj) * IOU_{Groundtruth}^{Predicted} \quad \textbf{Eq. (1)}$$

Where CS = Confidence Score, $P_r(Obj)$ represents the probability of the object and the

IOU Predicted Ground truth represents the IOU of predicted and ground truth bounding boxes. A confidence score (CS) of Zero means there is no object in the cell. A Confidence score tending towards value 1 is considered to be the best.

The cost function or loss function of YOLOv5 can be given by equation (2) [52]

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[(x_{i-}\hat{x}_i)^2 + (y_{i-}\hat{y}_i)^2\right] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[((\sqrt{w}_i - \sqrt{\widehat{w_i}}^2 + (\sqrt{h}_{i-}\sqrt{\widehat{h_i}})^2\right] + \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[(C_{i-}\widehat{C_i})^2\right] + \lambda_{noobj} \quad \textbf{Eq. (2)}$$

In the above equation, 3 $\boldsymbol{\lambda}$ constants represent more than one aspect of the loss function, which represents the highest order.

YOLOv5 is a more advanced and upgraded version of YOLO and YOLOv5. A boundary box around an input frame is predicted by using logistic regression and Feature Pyramid Network (FPN) in YOLOv5. To detect objects, YOLOv5 uses 53 convolutional layers to extract features from Darknet-53.

4.9 Route Detection Module:

These days, special paths are specially designed for visually impaired people. These paths are tactile and support easy movements, in our project these tactile paths are converted into a smart path using augmented Quick Response (QR) code.

4.9.1 Tactile Surface Paving:

Tactile paving as shown in figure 4.12 is a walking surface indicator[44] [53] that can produce a warning when this is detected with long canes or by walking on it. Tile sizes are currently determined by ISO/FDIS 23599, which is designed as an assistive product for visually impaired and blind people. The tiles were different in their tactile characteristics, such as tiles with parallel blister lines or tiles with offset blister lines.

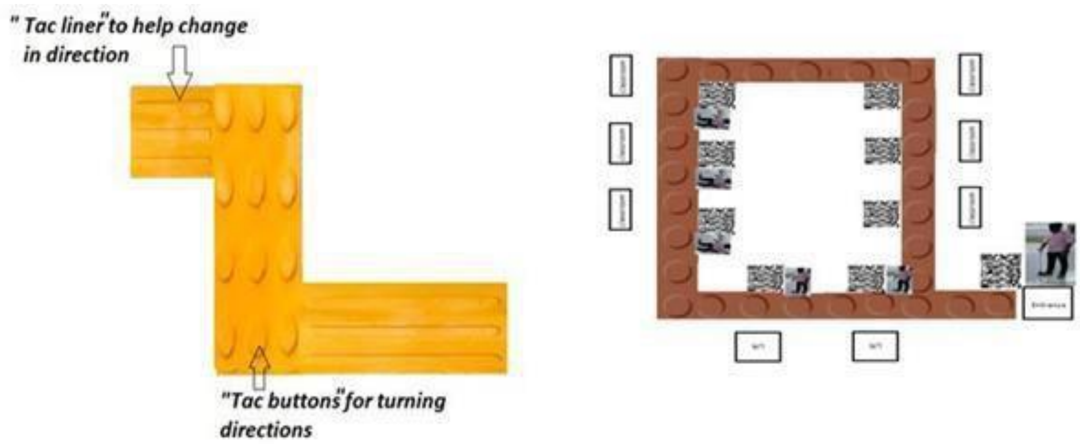


Figure 4.12. Corduroy Tactile used for tactile paving

In the below-mentioned table 4.3 majorly used tiles with application are discussed , although for navigation blister tiles are used.

Table 4.4. Various tiles used for tactile paving

Tile Name	Surface View	Features	Applications
Blister Tactile		Blisters	Railway Station
Corduroy Tactile		Parallel Bars	Pedestrian Path
Lozenge Tactile		Lozenge Tablet Shape	Edges of an street light rapid transit

4.10 Tracking Module:

The tracking module consists of two components namely Path Traversed Module and Mapper

4.10.1 Path Traversed Module: The VI user starts traversing the path from 36 blocks to 14 blocks in Lovely Professional University. This module records GPS data and stores it as a CSV file (Comma Separated Value). There are two columns in the file, namely Latitude and Longitude, which are values from the GPS module. The GPS data is appended every 30 seconds to the excel sheet as shown in Table 4.5.

Table 4.5. Test model Specifications and test conditions.

Time(min)	Place	Latitude	Longitude
0	**36 Block**	31.258189	75.707936
5	**Auditorium**	31.254047	75.70484
10	**13 Block**	31.254643	75.705323
15	**LPU Mall (14 blocks)**	31.255074	75.705666
20	**Baldevraj Hospital**	31.256661	75.70631

4.10.2 Mapper: It evaluates the current data and the previously stored data. In the event, if the user entered and moved through designated mapped steps then the output will be received by the user in the form of audio as "You are moving in the right direction" and if the user navigates off the path, then the message "It seems you have departed from the path" will be received through the microphone. As shown in figure 4.13, the path comprises movement through the tactile path, stairs, a round bottleneck path, and QR-enabled pavements in the path of the visually impaired user.

Figure 4.13 Movement of the visually impaired person through various environments.

Various sets of GPS data [54] from various test environments is included which varies from 36 blocks to 14 blocks in Lovely Professional University (LPU), Punjab(India). The set of data includes data from the 9th floor of 36 block building which has Blister-type tactile flooring and data from transient classrooms of that level to the movement in and around 14 blocks as shown in figure 4.14 and table 4.6.

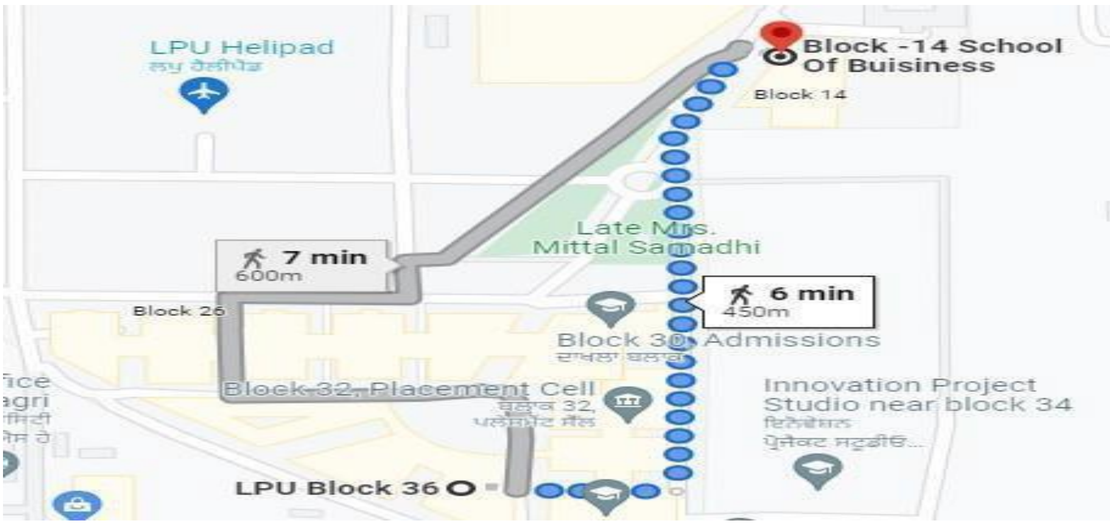
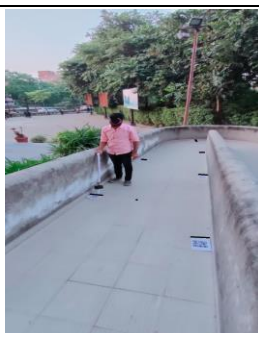


Figure 4.14 Mapping of a visually impaired person.

Table 4.6. Various paths used by the visually impaired person

	Testing Arena 1	Current GPS Value		Geo-location-based GPS Value		Accuracy	
Location		Latitude	Longitude	Latitude	Longitude	Latitude	Longitude
Entrance of Block 36		31.231	75.6065	31.196	75.515	85%	84.7%
Entering classroom		31.221	75.7047	31.1822	75.5977	82%	81%
Using Stairs to Block 14		31.247	75.7066	31.185	75.5753	75%	75.9%
Taking U shaped pavement to Block 14		31.252	75.7063	31.227	75.61	89.7%	87.5%

4.11 Testing and Evaluation:

4.11.1 Testing with Wearable mask and stick: To test the device, 3 places and 2 candidates has been chosen for testing environment, namely: 1) Garden Area 2) Classroom Area 3) Area enclosed from 36 blocks to 14 blocks. In the first environment hairs, trees, plants, polls, signboards, and more are detected. In the second environment i.e. Classroom environment, a tactile path is provided that contains various Quick Response codes tagged with black tape. The tapes provide useful information to VI users for easier movement. The last environment comprises the path from 36 blocks to 14 blocks in Lovely Professional University, Phagwara (India).

Table 4.7 Information detected with different devices.

Scene	User into scene	View from the device	Information detected	Device and Technique
Patent 1 in the garden area			3 Classes namely: 'Poll', 'Chair', 'Plant'	Wearable Device and YOLO
Patent 2 in the classroom area			Room no 102 'Seating right of the entrance'	Vision-Based Stick and OCR

4.11.2 Testing with Wearable mask for Known and Unknown Paths:

The Proposed model can be used for navigating any path i.e. Known or unknown path as shown in figure 4.15. The known path consists of 300 images of 3 classes namely straight, left and right which is supposed to get trained through this supervised model, whereas the unknown path consists of the real-time based unknown and untrained images.

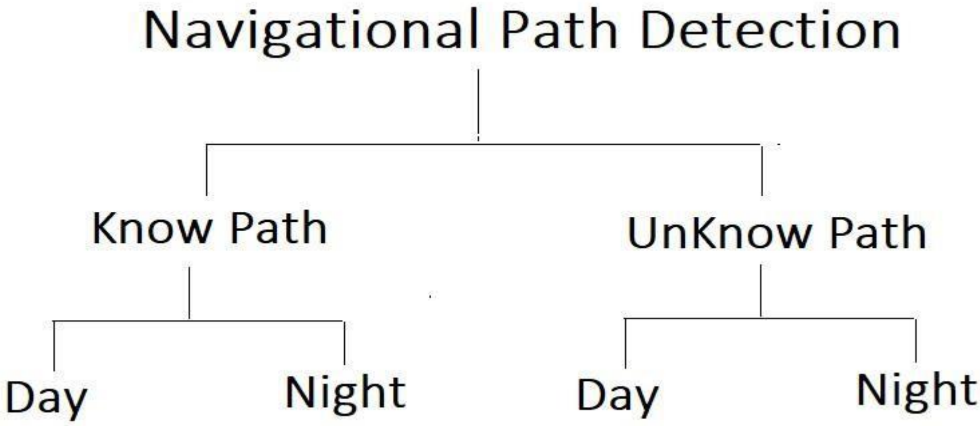


Figure 4.15 Feasibility of the proposed model

4.11.3 Testing in Known Path:

Testing in known path consists of known surroundings, so for that, 100 images of LPU Footpath behind 36 block and in front of 25 block are taken, the YOLO-based model has initially trained with the 100 each class images of Movable Path and Non-Movable Path. The annotation of a Non-Movable path consists of elements such as walls, pots, plants, barricades, boom barriers, parking post barriers, sign boards, park cement benches and park wooden benches The number of iterations used for training is 60 epochs. The model accuracy comes out to be 85.5%. In figure 4.16-4.18, detection for Non-Movable paths and Movable paths are shown with a certain probability depending on the IOU (Intersection over union).

Figure 4.16 Various IOU values for the Non-Movable and Movable path at the back of the 36 block.

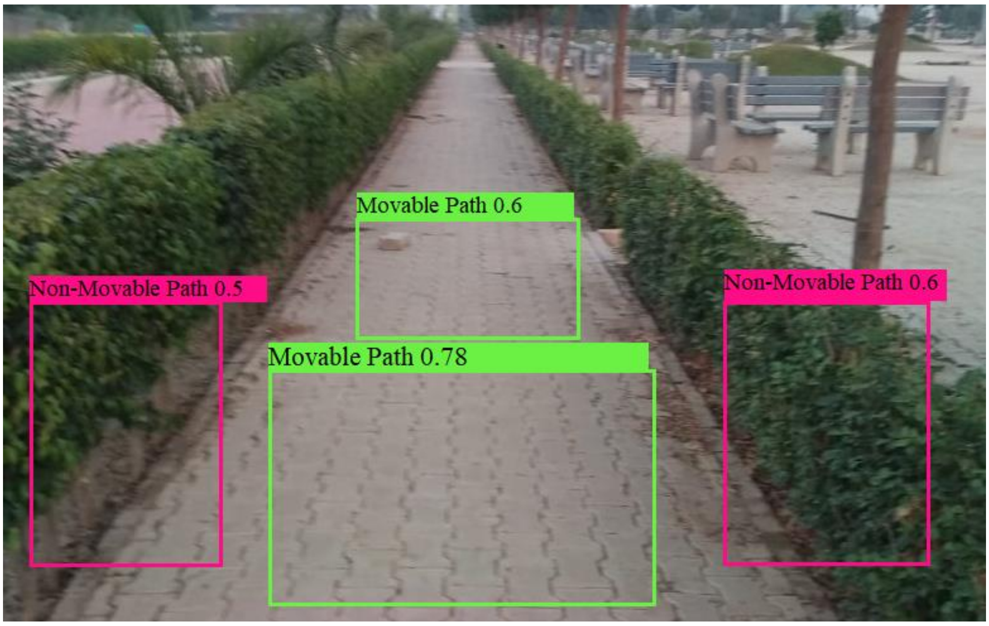


Figure 4.17 Various IOU values for the Non-Movable and Movable path in front of the 25 block.

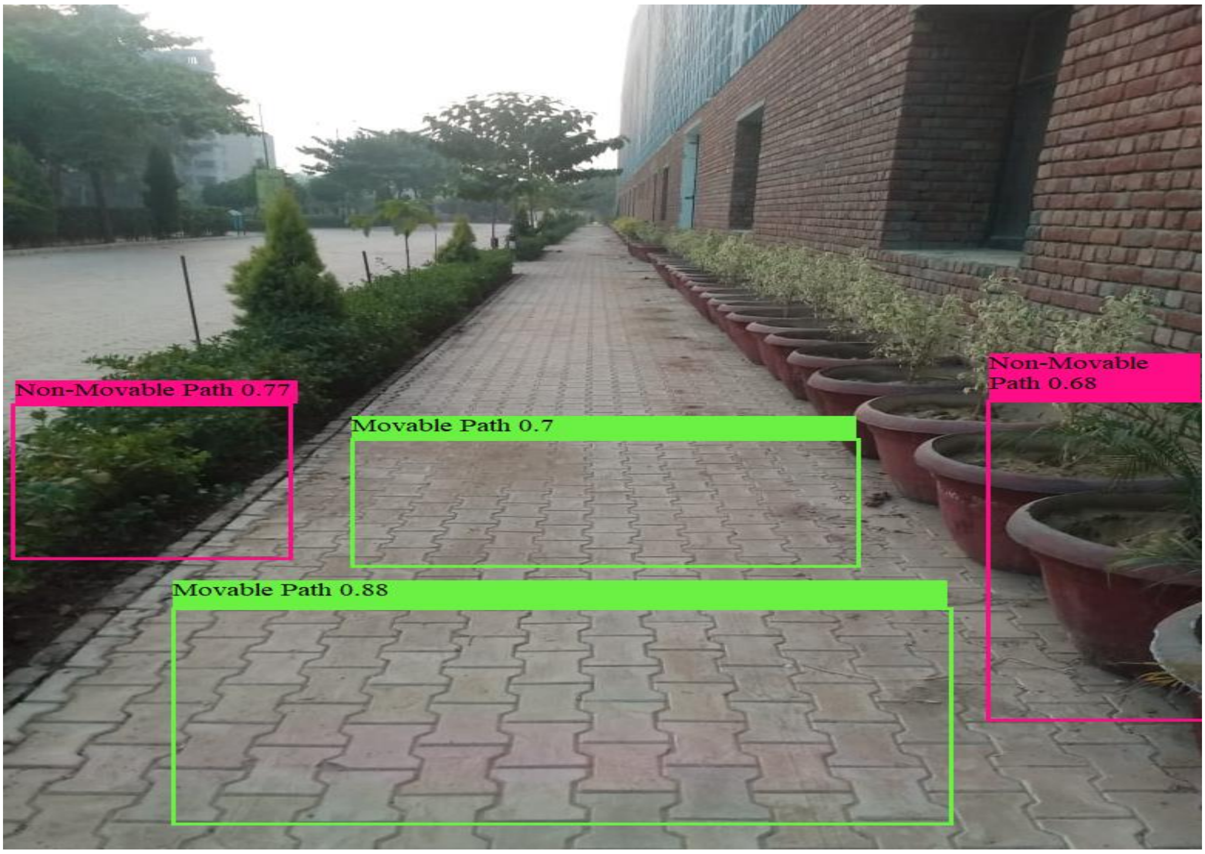


Figure 4.18 Various IOU values for the Non-Movable and Movable path in front of the 27 block.

Testing for Movable and Non-Movable classes consists of 100 images of Movable Path and Non-Movable Path, the mAP (minimum accuracy precision) for the system comes to be 81 %. Shown below in Table 4.8 are the 2 different classes with their parameters.

Table 4.8 Different parameters during the day

True/ False Predicted	→	
↓	Positive (MP)	Negative (NMP)
Positive (MP)	78 (TP)	6 (FP)
Negative (NMP)	2 (FN)	3 (TN)

Calculations for classes (Movable Class and Non-Movable Class) is as follows:

i) Recall = TP/ TP+FN

=78/78+2= 0.975= 97.5 %

ii) Precision = TP/ TP+FP

=78/78+6= 0.9285=92.85 %

iii) Accuracy = TP+TN/ TP+TN+FP+FN

=81/100= 0.81=81%

iv) F1 = 2 X Precision X Recall/ (Precision +Recall)

F1= 2 X 0.9285 X 0.975/(0. 9285 +0.975)

=0.9511

Furthermore, the Known model has been trained with daylight and artificial light (during the night). Shown below in Table 4.9 are the different classes and their accuracies during the night.

Table 4.9 Different parameters during the night

True/ False → **Predicted** ↓		
	Positive (MP)	Negative (NMP)
Positive (MP)	27 (TP)	32 (FP)
Negative (NMP)	25 (FN)	16 (TN)

Calculations for classes (Movable Class and Non-Movable Class) is as follows:

v) Recall = TP/ TP+FN

=27/27+25= 0.5191= 51.91 %

vi) Precision = TP/ TP+FP

=27/27+32= 0.457=45.7 %

vii) Accuracy = TP+TN/ TP+TN+FP+FN

=43/100= 0.43=43%

viii) F1 = 2 X Precision X Recall/ (Precision +Recall)

F1= 2 X 0.457 X 0.5191/(0.457+0.5191)

=0.4860

+

4.11.4 Testing in Unknown Path:

Testing in an unknown path consists of surroundings that are never seen or prior known to the person. This model is tested pre-trained on YOLO (You Only Look Once) with known images. The model accuracy comes out to be 96%. In figure 4.19-4.21, detection for Non-Movable paths and Movable paths are shown with a certain probability depending on the IOU (Intersection over union).

Figure 4.19 Various IOU values for Non-Movable and Movable path for unknown path1.

Figure 4.20 Various IOU values for Non-Movable and Movable path for unknown path2.

Figure 4.21: Various IOU values for Non-Movable and Movable path for unknown path3.

Table 4.10 Comparative result of different models in the day and artificial light

Algorithm	**Test Set Performance**		
	Number of the data set	**mAP**	**fps**
Yolo V3 (Artificial light)	**300**	**43.3**	**28**
Yolo V3 (Daylight)	**300**	**81**	**45**

As shown in Table 4.10, The model is trained with the YOLO algorithm with a set of 300 images, the Frames per second of the model is 45 FPS for daylight and 28 Frames per second for artificial light is giving better results for both categories. The results for daylight have 43.3 mAP and for daylight is 81 mAP. This shows that the mean average precision (mAP) for the path used by VI persons in the artificial light comes lesser than for the same path in daylight. As shown below in Figure 4.22-4.24

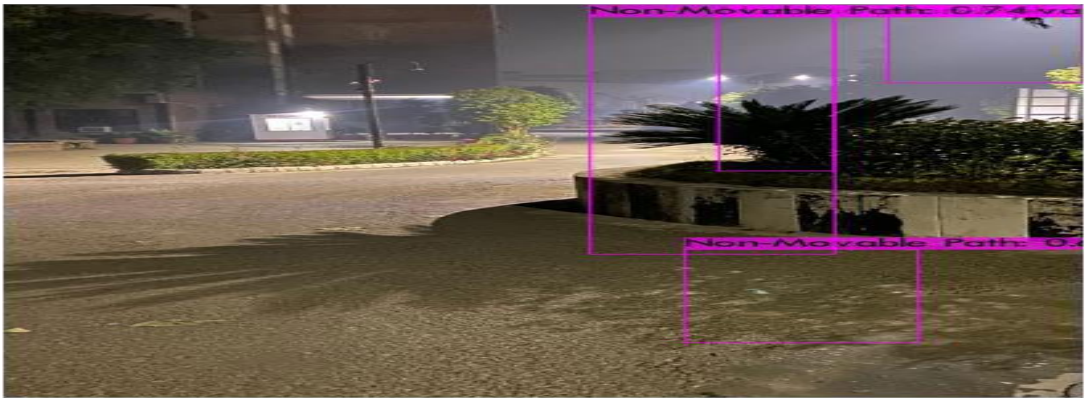


Figure 4.22 Various IOU values for Non-Movable and Movable path for path1 in artificial light.

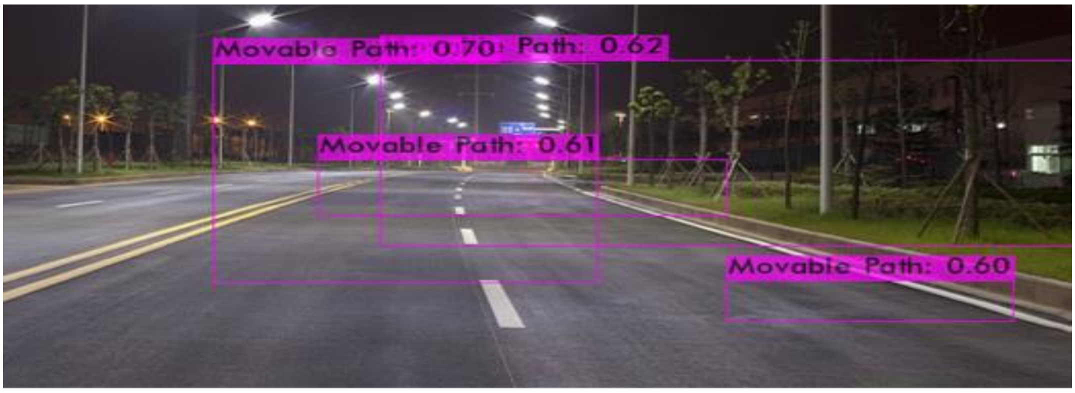


Figure 4.23 Various IOU values for Non-Movable and Movable path for path2 in artificial light.

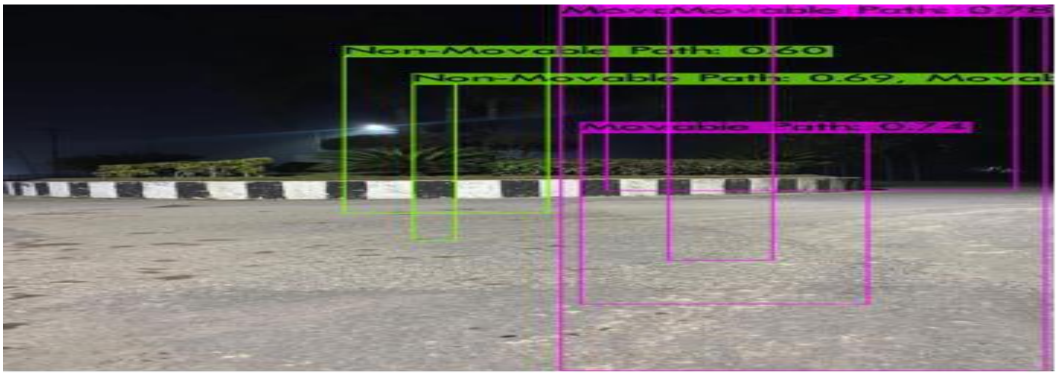


Figure 4.24 Various IOU values for Non-Movable and Movable path for path3 in artificial light.

For the model, 100 images for each class is collected, for the straight path mean average precision (mAP) comes to be 79%, for the right path 83% mean average precision (mAP), and for the left path 85% mean average precision (mAP). The model detected straight path takes about 42 milliseconds, detecting the right path takes 49 milliseconds while detecting the left path takes 47 milliseconds. The model also gives average accuracy of 82.33%. As our model also comprises smart cane the collective accuracy of the model increases to 89.24%.

As shown below in table 4.11, the comparison set of YOLOv3 and SSD (Single shot detector) is analyzed in the situation where a blind and visually impaired person is made to initially navigate through a known environment and later made to test through an unknown environment, for the environment, it is believed that the path can have 3 major movements as a straight path, left path and the right path.this environment is first with the YOLOv5 model [59] and then compare it with the SSD algorithm,As shown in in figure 4.25 the gradient of descent increases rapidly in the beginning, but then the Loss value gradually slows down, and eventually stabilizes. The results indicate better FPS (Frames per second) for the YOLOv3 as compared to SSD i.e., Approx. 51 FPS as compared to 32 FPS) as shown and figure 4.26

Table 4.11: Minimum Accuracy Precision (mAP) and time elapsed for different environments

Class	Number of images	YOLOv5						SSD(Single Shot Detector)					
		True	False	None	Accuracy	Time Elapsed (ms)	FPS	True	False	None	Accuracy	Time Elapsed (ms)	FPS
Straight Road	100	79	12	9	0.79	42	50	76	11	8	0.76	40	37
Left Turn	100	85	9	6	0.85	47	45	83	10	7	0.83	43	32
Right Turn	100	83	10	7	0.83	49	42	81	10	9	0.81	48	35

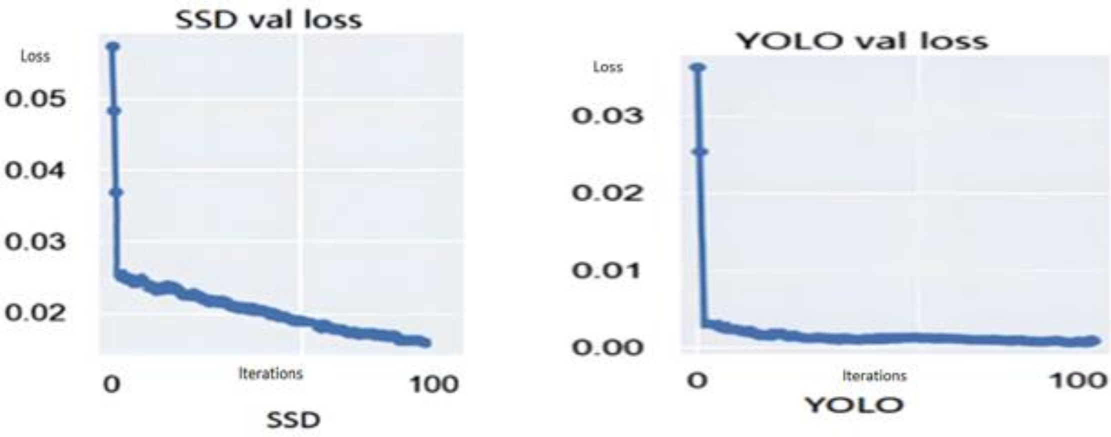


Figure 4.25 Comparison between SSD and YOLOv5 algorithm for losses

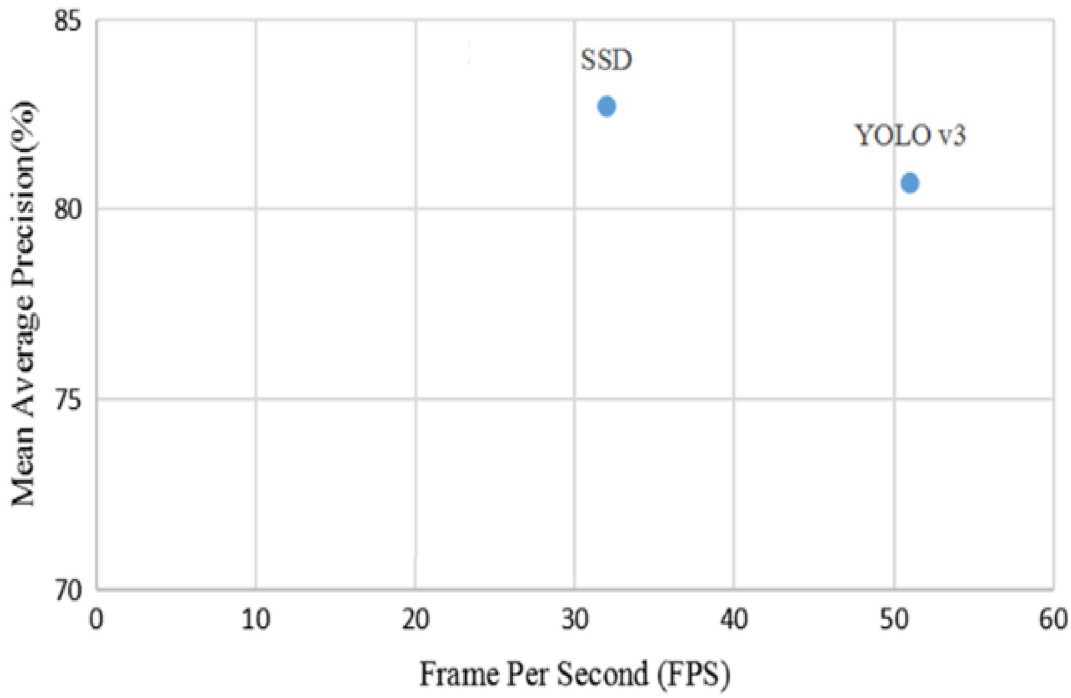


Figure 4.26 Comparison between SSD and YOLOv3 algorithm with mAP and FPS

4.12 Novelty:

This section shows the proposed improvement to the selected model. In this proposed method, the YOLO model was optimized using specific ROI. Fig 4.27 shows the classification through normal YOLO and Fig 4.28 shows classification through modified YOLO.

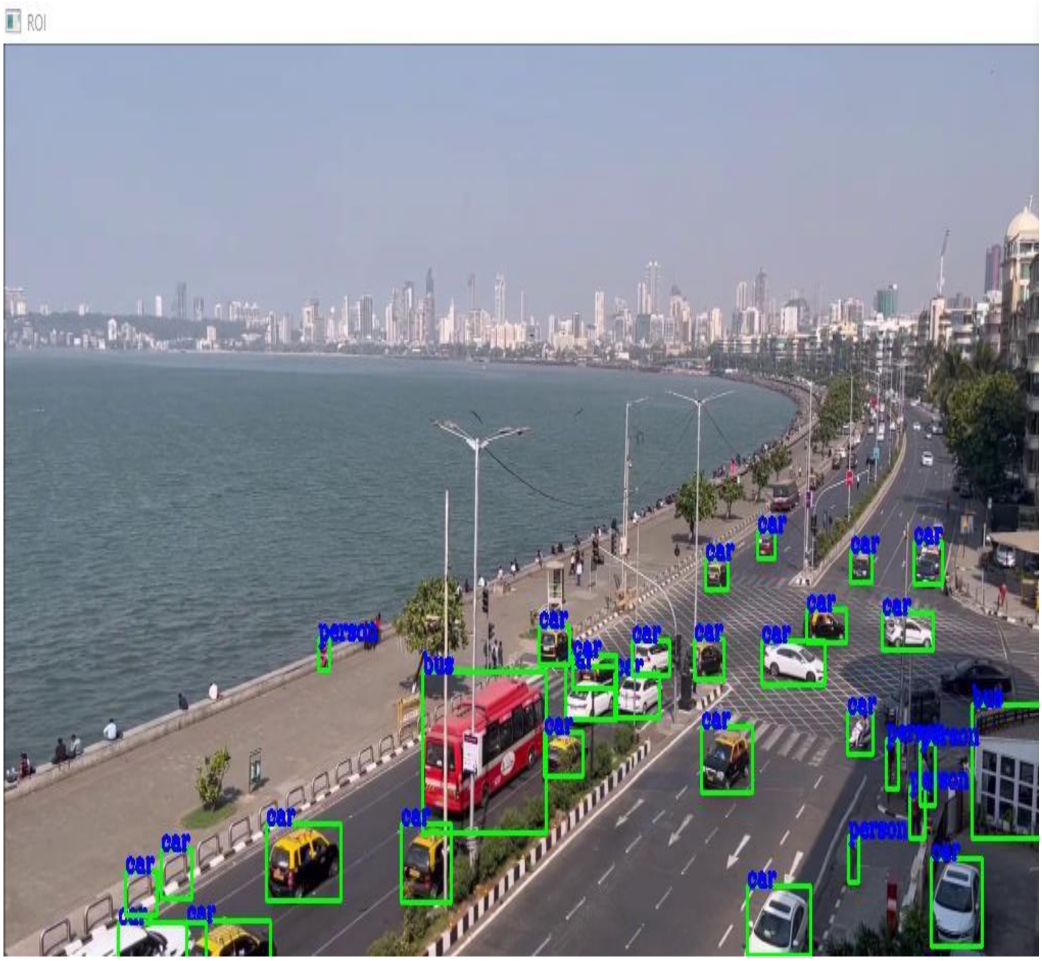


Figure 4.28: shows classified objects using YOLO.

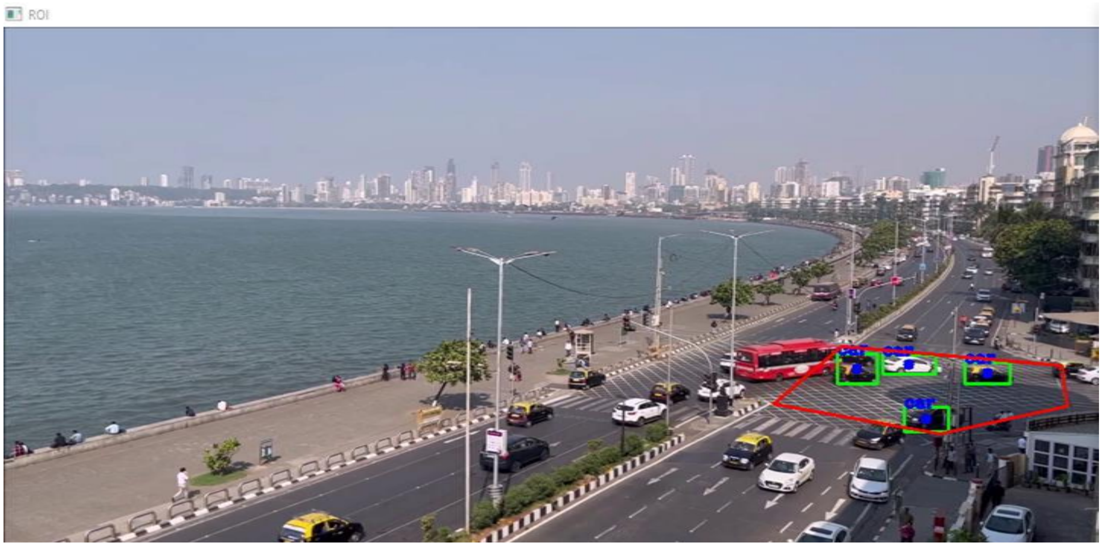


Figure 4.29: shows classified objects using modified YOLO.

The term region of interest (often shortened to ROI) is used to refer to a sample within a data set that is used in a particular way. It is commonly used in many areas of application to define a ROI. For Example as many as 30 objects are detected through YOLO, which are been reduced to 4 for a specific region of interest. ROIs can be viewed literally as polygonal selections on a map in geographic information systems (GIS). An object under consideration is defined by its ROI in computer vision and optical character recognition. There may be individual points of interest (POIs) within a ROI. As shown in figure 4.28 , a specific ROI is choose as pentagon with vertices as V1, V2, V3, V4 and V5 as (1250,400),(750,400),(700,800),(1200,800)(800,900)).

Different geometrical shapes for different ROI are as shown below

Table :4.12 shows different geometrical shapes for different ROI

Shapes	**ROI**
Polygon	Polygon Region of interest
Rectangular	Rectangular Region of interest
Circular	Circular Region of interest
Line	Line Region of interest
Polyline	Polyline Region of interest
Rectangular	Rectangular Region of interest
Ellipse	Ellipse Region of interest

4.13 Device Validation:

Validation is the scientific inquiry tradition of logical deduction and induction is typically rooted in the validation of engineering research. To confirm that the product operates as expected under the conditions where it will be used, validation testing would involve test cases, test suites, or even clinical trials. As shown in figure 4.28, the design validation tests are usually the last tests performed since they should be carried out on production or production equivalent models. The purpose of design validation is to ensure that the product meets the needs of the user.

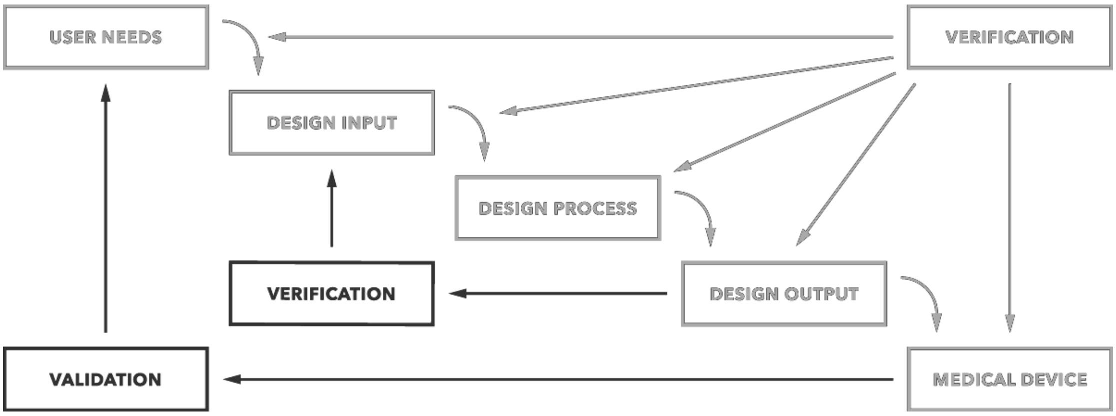


Figure 4.28 Block diagram for validation of the model

4.13.1 User Needs: To validate a specific device validation requires user needs, so a visit is planned to a government-approved disabled school in Agra. The school's name was –Agra Residential Blind School‖, school is a higher secondary and comprises 30 students and 5 teachers. At the time of the visit, 12 students (9 male and 3 female) and 2 teachers were present who were blind and partially blind. During interaction their lifestyle, reading and writing habits (braille) are revealed with various challenges are discussed which restrict them in moving out. Figure 4.29-4.30 shows the school and some school students.

Figure 4.29 Agra Residential Blind School.

Figure 4.30 Blind students during the interaction.

4.13.2: Design Input: For having the end user's design specification two students, Salim and Asif who were studying in 12^{th} class were agreed to assist in designing the product. In that process wearable mask specifications for Salim and Asif were taken as shown in Figure 4.31 Our Wearable 3-D model has dimensions of 18 cm in length to 12 cm in breadth, the depth of the model is 4 cm to use with ease and the model has an elastic wand for ease of use.

Figure 4.31 Blind students during design inputs.

4.13.3 Design Process Design Output and Verification: The Design process comprises designing the hardware and the software and the design output is the final model designed. This Section is already discussed in the section 3.2

4.13.4 Validation: Once the product with a wearable mask and smart stick is designed, some questionnaires are discussed with the students. Shown below in Figure 4.32-4.33 some school students going for validation of the device.

Figure 4.32 Blind students during testing.

Figure 4.33 Blind students during validation

Precision: Precision measures how accurate the model can predict. It is the ratio of True Positive data to the Total Positive data of the model. The Precision of the model comes to be 84%(as earlier calculated on page 71), as shown in figure 4.34.

Table 4.12 shows the precision of the device model

True/ False → Predicted ↓		
	Positive (MP)	Negative (NMP)
Positive (MP)	82 (TP)	6 (FP)
Negative (NMP)	2 (FN)	10 (TN)

.

Table 4.13 shows the Questionnaire for "Precision"

Participant	Class	Gender	Age	Questionnaire	Rating (0-10)
Salim	**12th**	**M**	**28**	**Precision**	**8**
Asif	**12th**	**M**	**29**	**Precision**	**9**
Vishal	**12th**	**M**	**27**	**Precision**	**8**
Rahul	**10th**	**M**	**25**	**Precision**	**9**
Santosh	**11th**	**M**	**24**	**Precision**	**8**
Aditi	**10th**	**F**	**22**	**Precision**	**8**
Riya	**9th**	**F**	**18**	**Precision**	**7**
Rihanna	**8th**	**F**	**16**	**Precision**	**9**

Overall Rating for device "Precision" > 8.25

Reason supporting Rating: Users were tested for different paths

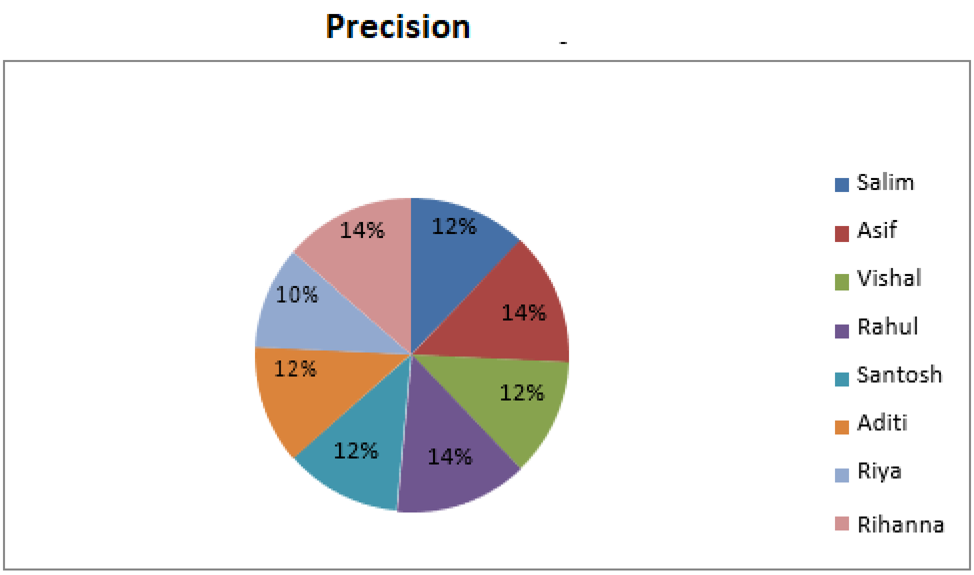


Figure 4.34 Pie chart for Precision Validation

Performance speed: Performance is the ability of a device to achieve its intended purpose as shown in figure 4.35**.** Here performance speed is in Frame per second, which varies from 15 FPS to 65 FPS as it depends on size of the image and computation power of the processor.

Table 4.14 shows the variation of image resolution with FPS

Image	Image Resolution	FPS
1	512 X 288	62
2	600 X400	58
3	640 X 480	56
4	1080 X 1920	35
5	1200 X 670	28

Table 4.15 shows the Questionnaire for "Performance Speed"

Questionnaire for "Performance Speed"

Participant	Class	Gender	Age	Questionnaire	Rating (0-10)
Salim	12th	M	28	Performance	9
Asif	12th	M	29	Performance	9
Vishal	12th	M	27	Performance	10
Rahul	10th	M	25	Performance	9
Santosh	11th	M	24	Performance	10
Aditi	10th	F	22	Performance	8
Riya	9th	F	18	Performance	9
Rihanna	8th	F	16	Performance	10

Overall Rating for device "Performance Speed" > 9.25

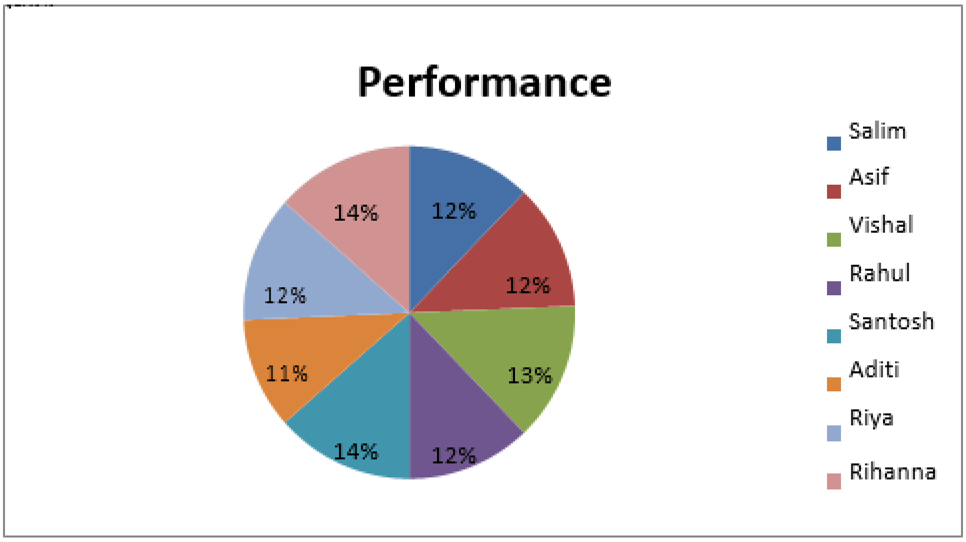


Figure 4.35 Pie chart for performance validation

User Friendliness of the device: User-friendly is a description of a device or software interface which can be used easily. The maximum rating for user friendliness was given by aditi owing to fact that she has keen for gadgets. Shown below in Table 4.16 is the table for questionnaire and pie chart shown in figure 4.36.

Table 4.16 shows the Questionnaire for "User Friendly"

Participant	Class	Gender	Age	Questionnaire	Rating (0-10)
Salim	12th	M	28	User Friendly	9
Asif	12th	M	29	User Friendly	8
Vishal	12th	M	27	User Friendly	8
Rahul	10th	M	25	User Friendly	7
Santosh	11th	M	24	User Friendly	8
Aditi	10th	F	22	User Friendly	10
Riya	9th	F	18	User Friendly	8
Rihanna	8th	F	16	User Friendly	9

Overall Rating for device User Friendliness > 8.375

Reason supporting Rating: Device helped VI person navigate easily.

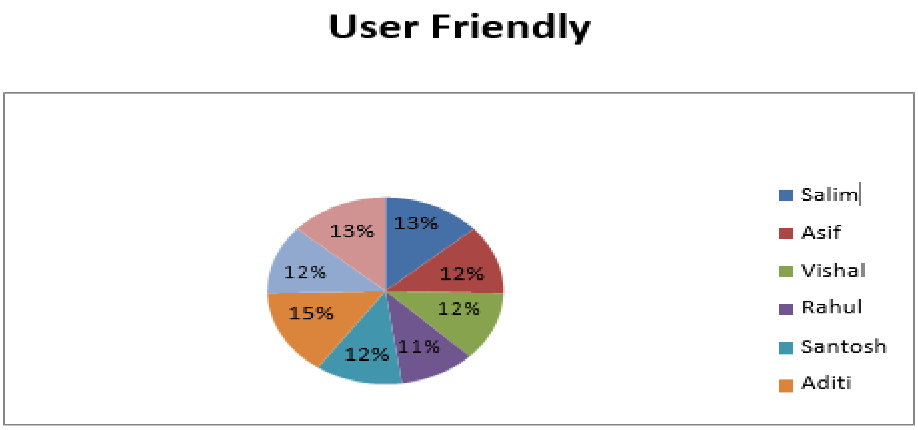


Figure 4.36 Pie chart for user friendly validation

Donning and Doffing time: Time taken to wear is called donning and time taken to wear out is called doffing time. The maximum rating for donning and doffing time was given by Salim, as the device was tested many a time on him.

Table 4.17 shows the Questionnaire for "Donning and Doffing time"

Participant	Class	Gender	Age	Questionnaire	Rating (0-10)
Salim	**12^{th}**	**M**	**28**	**Donning and Doffing time**	**9**
Asif	**12^{th}**	**M**	**29**	**Donning and Doffing time**	**7**
Vishal	**12^{th}**	**M**	**27**	**Donning and Doffing time**	**9**
Rahul	**10^{th}**	**M**	**25**	**Donning and Doffing time**	**7**
Santosh	**11^{th}**	**M**	**24**	**Donning and Doffing time**	**8**
Aditi	**10^{th}**	**F**	**22**	**Donning and Doffing time**	**8**
Riya	**9^{th}**	**F**	**18**	**Donning and Doffing time**	**7**
Rihanna	**8^{th}**	**F**	**16**	**Donning and Doffing time**	**9**

Overall Rating for device Donning and Doffing time > 8

Reason supporting Rating: Time taken to adjust the wand is different for different users.

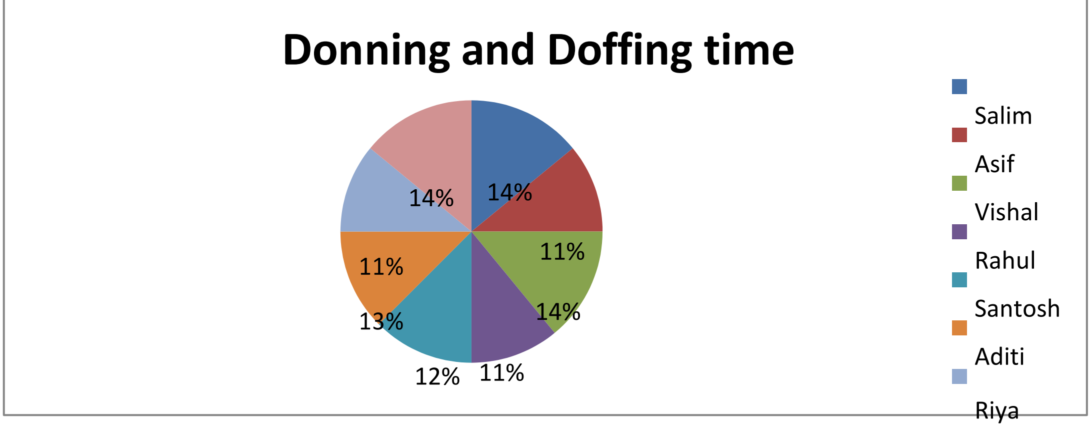


Figure 4.37 Pie chart for Donning and Doffing time validation

Product Size (Design): Product size is the specification of the product. It is made universal so that anyone can wear it as shown in figure 4.38

Table 4.18 shows the Questionnaire for "Product Size (Design)"

Participant	Class	Gender	Age	Questionnaire	Rating (0-10)
Salim	12th	M	28	Product Size (Design)	8
Asif	12th	M	29	Product Size (Design)	8
Vishal	12th	M	27	Product Size (Design)	9
Rahul	10th	M	25	Product Size (Design)	7
Santosh	11^{th}	M	24	Product Size (Design)	7
Aditi	10th	F	22	Product Size (Design)	7
Riya	9th	F	18	Product Size (Design)	7
Rihanna	8th	F	16	Product Size (Design)	8

Overall Rating for device "Product Size (Design)" > 7.625

Reason supporting Rating: Same Product is used by all the students

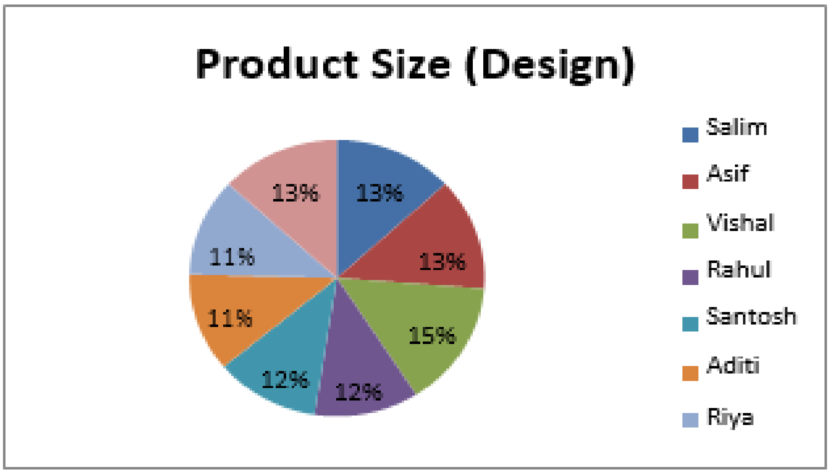


Figure 4.38 Pie chart for product size validation

Reason supporting Rating: Identified and classified many objects in the path of the user.

Table 4.19 shows Overall model Validation:

Parameters	Salim	Asif	Vishal	Rahul	Santosh	Aditi	Riya	Rihanna	Overall
Precision	8	9	8	9	8	8	7	9	8.25
Performance Speed	9	9	10	9	10	8	9	10	9.25
User Friendly	9	8	8	7	8	10	8	9	8.375
Donning and Doffing	9	7	9	7	8	8	7	9	8
Product Size (Design)	8	8	9	7	7	7	7	8	7.625

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

In this project, , a intend is to develop a computer vision-based system [111] that will assist visually impaired people in navigating their environments independently. Users can use the obstacle detection system to detect obstacles in their environment in front of them. A distance is determined between the user and the obstacle by using an ultrasonic sensor and the objects present in the path of the user could be detected, recognized and localized captured by Pi camera with the improved YOLO algorithm. In order to make it easier for users, the output is provided through audio feedback instead of vibration. With the proposed system, the visually challenged user is given accurate guidance about obstructions and objects in front of him or her, which is not possible with conventional guidance devices.

This study is aimed to develop an accurate and cost-effective solution that can be deployed to ease navigational accessibility for visually impaired people. Our model is unique due to its hybrid model that comprises a wearable device and a vision-based smart stick. Both the model is 3D designed in CAD model through Fusion 360 and thereafter printed. The wearable device is trained by a machine-learning algorithm to detect major objects that fall in the path of the user and a vision-based stick uses GPS, ultrasonic, and a Camera which adds accuracy to the existing model. The wearable model in a real-time navigational system was 3D modeled to achieve this objective and then a machine learning module was introduced to make the system more robust and adaptable to environmental changes. a low-cost wearable device is proposed that could assist visually impaired individuals in finding their way through an environment by identifying their surroundings such as cars, bicycles, persons, trucks, buses, etc. Our model has been tested with a known dataset provided by COCO[112] which has 80 classes and also with our custom data set which is a modification to the existing design.

The Custom design comprises three additional classes namely Straight-Path,Left-Path, and Right-Path which help a user in navigating a particular area. The proposed model combines a single neural network with a full image to use an entirely different approach. Using this Network, bounding boxes and probabilities of each region of the image are determined. As opposed to systems like R-CNN, which require thousands of bounding boxes for a single image, the bounding boxes for the predictions are weighted according to the probabilities. Improved YOLO-based models are very fast, over 1000x faster than R-CNN and 100x faster than Fast R-CNN [56][57[58][59]. These models are very different from other models in these categories. Moreover, YOLOv5 exhibits better accuracy and speed than other versions of YOLOv4 and YOLOv3[59][110].

The model is trained on over 300 images of various indoor and outdoor environments. The model by providing an anonymous path that contains segments class of left, right, and straight paths which supports the travel of people with visual impairments with accurate positional information and travel directions. Hence, this approach will be beneficial to visually impaired people. It is proposed that the model can solve the navigation problem of a visual person in indoor and outdoor environments and help them understand their environment with an overall accuracy of 89.24% as compared with previous models [60] This model improves precision as well as speed by eliminating past deficiencies of non-overlapping bounding boxes.

5.2 Future Scope

Future developments will incorporate under mentioned points

1. **Test new state-of-the-art lightweight object detectors:** The topic of object detection is hot right now, and new products are being launched every year that perform better than the previous ones. The model could be tested to Facebook detectors as well as other variants of new variants YOLO.

2. **Trying to improve the results of the OI (Open Image) Dataset** is definitely a possible future project, this could be achieved by possibly changing some configurations and letting the training run for more than 100 epochs.

3. **Prediction of distance by the object:** Alternatively, you can use more complex cameras with sensors that can measure the distance of an object, or modify your model so that it can predict also the distance of an object.

Since the pipeline developed in this thesis work is modular and robust for evaluating different algorithms, validated by the use of different algorithms and getting the complete performance analysis, it allows the user to evaluate an object detection algorithm for even more classes; also, the pipeline automatically considers the number of classes mentioned in the ground truth data, providing scalable architecture. Further, to improve detection algorithms ‘performance on the simulated dataset, the neural networks can be trained on the simulated dataset generated from the developed pipeline along with the real-world data. And the performance of such modules can then be easily evaluated on the developed pipeline. Apart from just evaluating the performance of detection modules on images, future work could be to use the same proposed architecture to evaluate the performance on video sequences. And, another metric can be added to quantify the performance of the pipeline such as versatility. Therefore, the scope of the proposed framework is not just limited to this thesis work and can be extended further to future course of study and development of the work.

REFERENCES

[1] A. G. Nicholas, and G. E. Legge, ―Blind navigation and the role of technology,‖ in The Engineering Handbook Of Smart Technology for Aging, Disability, and Independence,2008, pp. 479–500.

[2] Riazi A, Riazi F, Yoosfi R, Bahmeei F. Outdoor difficulties experienced by a group of visually impaired Iranian people. J Curr Ophthalmol. 2016 May 2; 28(2):85-90.

[3] R. Manduchi, S. Kurniawan, and H. Bagherinia, ―Blind guidance using mobile computer vision: a usability study,‖ in ASSETS, 2010, pp. 241–242.

[4] B.-S. Lin,C.-C.Lee, and P.-Y. Chiang, ―Simple smartphone-based guiding system for visually impaired people,‖ Sensors, Vol. 17, no. 6, pp. 1371, 2017.

[5] W. H. Hsieh, ―An Experimental Study on Cam-Controlled Planetary Gear Trains,Mechanism and Machine Theory, vol. 42, no. 5, pp. 513-525, May 2007.

[6] Y.Zhao, E. Kupferstein, D. Tal, andS. Azenkot,―It looks beautiful but scary: How low vision people navi gate stairs and other surface level changes,‖ in Proceedings of the 20th International ACM SIGAC-CESS Conference on Computers and Accessibility, ACM, 2018, pp.307–320.

[7] H. Devlin. Echolocation could help blind people learn to navigate like bats, Feb 2018. Acoustical Society of America. Exploring the potential of human echolocation, Jun 2017.

[8] M. I. Islam, M. M. H. Raj, S. Nath, M. F. Rahman, S. Hossen and M. H. Imam, "An Indoor Navigation System for Visually Impaired People Using a Path Finding Algorithm and a Wearable Cap," *2018 3rd International Conference for Convergence in Technology (I2CT)*, Pune, India, 2018.

[9] A. Ali and M. A. Ali, "Blind navigation system for visually impaired using windowing-based mean on Microsoft Kinect camera," *2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME)*, Beirut, Lebanon, 2017.

[10] M. C. Holbrook, and A. Koenig. History and theory of teaching children and youths with visual impairments. Foundations of Education. Volume I, 2000.

[11] J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu, "Smart guiding glasses for visually impaired people in indoor environment," in *IEEE Transactions on Consumer Electronics*, vol. 63, no. 3, pp. 258-266, August 2017, doi: 10.1109/TCE.2017.

[12] K. K. O. Sharma, N. Jain, A. K. Bhargav and S. B. Bhanu Prashanth, "Design & development of a bone conduction based navigation aid for the visually impaired," *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, India, 2017, pp. 1699-1702.

[13] Srinivas K.S., Sahithya K., Tejaswi G.L., Gopal K.H. and Karthik B.P., A new method for recognition and obstacle detection for visually challenged using smart glasses powered with Raspberry Pi., International Journal of Engineering Applied Sciences and Technology, Vol. 5, Issue 1, ISSN No.2455-2143, Pages 408-412, May 2020.

[14] Hegde P., Devathraj N., Sushma S.K. and Aishwarya P., Smart Glasses forVisually Disabled Person, International Journal of Research in Engineering and Science (IJRES), Volume 9 Issue 7, ISSN: 2320-9356, PP. 62-68, July 2021.

[15] Tang, W., Liu, D.E., Zhao, X., Chen, Z. and Zhao, C., A dataset for the recognition of obstacles on blind sidewalk. Universal Access in the Information Society, pp.1-14, August 2021.

[16] Islam, M.M., Sadi, M.S. and Bräunl, T., Automated walking guide to enhance the mobility of visually impaired people. IEEE Transactions on Medical Robotics And Bionics, Vol. 2, No. 3, pp.485-496, August 2020.

[17] Annapoorani, A., Kumar, N.S. and Vidhya, V., Blind-Sight: Object Detection with Voice Feedback, International Journal of Scientific Research & Engineering Trends Volume 7, Issue 2, ISSN: 2395-566X, pp.644-648, April-2021.

[18] Das, D., Saha, P. and Tabassum, N., Real Time Distance and Mobility Detection of Obstacles Using a Smart Multisensor Framework for Visually Impaired People, 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 590-593), June 2020.

[19] Yadav, S., Joshi, R.C., Dutta, M.K., Kiac, M. and Sikora, P., Fusion of Object Recognition and Obstacle Detection approach for Assisting Visually Challenged Person, 2020 43rd International Conference on Telecommunications and Signal Processing (TSP), (pp. 537-540), July 2020.

[20] Mahendru, M. and Dubey, S.K., Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3., 2021 11th International Conference on Cloud Computing, Data Science & Engineering (pp. 734-740), January 2021.

[21] Karmarkar, M.R.R. and Honmane, V.N., Object Detection System For The Blind With Voice Guidance, International Journal of Engineering Applied Sciences and Technology, Vol. 6, Issue 2, ISSN No. 2455-2143, Pages 67-70, June 2021.

[22] Mahmood, R.R., Younus, M.D. and Khalaf, E.A., Real Time Object Detection for Visually Impaired Person. Annals of the Romanian Society for Cell Biology, ISSN:1583-6258, Vol. 25, Issue 4, Pages. 14725 – 14732, April 2021.

[23] Sharif, M., Khan, S., Saba, T., Raza, M., & Rehman, A. (2019). Improved video stabilization using SIFT-log polar technique for unmanned aerial vehicles. *2019 International Conference on Computer and Information Sciences (ICCIS).*

[24] Zheng, L., Yang, Y., & Tian, Q. (2018). SIFT meets CNN: A decade survey of instance retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(5), 1224-1244.

[25] E. Karami, S. Prasad, and M. Shehata, –Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images," p. 6, 2015. [26] Wang, R., Shi, Y., & Cao, W. (2019). GA-SURF: A new speeded-up robust feature extraction algorithm for multispectral images based on geometric algebra.*Pattern Recognition Letters*, *127,* 11-17.

[27] Ding, P., Wang, F., Gu, D., Zhou, H., Gao, Q., & Xiang, X. (2018). Research on optimization of SURF algorithm based on embedded CUDA platform. *2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER).*

[28] Xingteng, J., Xuan, W., & Zhe, D. (2015). Image matching method based on improved SURF algorithm. *2015 IEEE International Conference on Computer and Communications (ICCC).*

[29] Geng, Z. X., & Qiao, Y. Q. (2017). An improved illumination invariant SURF image feature descriptor. *2017 International Conference on Virtual Reality and Visualization (ICVRV).*

[30] Tareen, S. A., & Saleem, Z. (2018). A comparative analysis of sift, surf, kaze, AKAZE, orb, and brisk. *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET).*

[31] Deshpande, S., & Shriram, R. (2016). Real time text detection and recognition on hand held objects to assist blind people. *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT).*

[32] Adriano, J. E., Calma, K. A., Lopez, N. T., Parado, J. A., Rabago, L. W., & Cabardo, J. M. (2019). Digital conversion model for hand-filled forms using optical character recognition (OCR). *IOP Conference Series: Materials Science and Engineering*, *482*, 012049.

[33] Liem, H. D., Minh, N. D., Trung, N. B., Duc, H. T., Hiep, P. H., Dung, D. V., & Vu, D. H. (2018). FVI: An end-to-end Vietnamese identification card detection and recognition in images. *2018 5th NAFOSTED Conference on Information and Computer Science (NICS).*

[34] Liu, Z., Luo, Y., Cordero, J., Zhao, N., & Shen, Y. (2016). Finger-eye: A wearable text reading assistive system for the blind and visually impaired. *2016 IEEE International Conference on Real-time Computing and Robotics (RCAR)*.

[35] Zhu, D., Li, T., Ho, D., Zhou, T., & Meng, M. Q. (2018). A novel OCR-RCNN for elevator Button recognition. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).*

[36] Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A., & Ouni, K. (2019). Car detection using unmanned aerial vehicles: Comparison between faster R-CNN and YOLOv3. *2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS).*

[37] Kim, K., Kim, P., Chung, Y., & Choi, D. (2017). Performance enhancement of YOLOv3 by adding prediction layers with spatial pyramid pooling for vehicle detection. *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS).*

[38] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517-6525.

[39] M. Maiti, P. Mallick, M. Bagchi, A. Nayek, T. K. Rana and S. Pramanik, "Intelligent electronic eye for visually impaired people," *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, Bangkok, Thailand, 2017, pp. 39-42.

[40] Maolanon, P., & Sukvichai, K. (2018). Development of a wearable household objects finder and localizer device using CNNs on raspberry Pi 3. *2018 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE).*

[41] Widyastuti, R., & Yang, C.-K. (2018). Cat‘s nose recognition using You Only Look Once (Yolo) and Scale-Invariant Feature Transform (SIFT). *Proceedings of the 2018 IEEE 7th Global Conference on Consumer Electronics, Nara, Japan, 55-56.*

[42] Hu, H., Gu, J., Zhang, Z., Dai, J., & Wei, Y. (2018, June). Relation networks for object detection. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA*, 3588-3597.

[43] Cao, C., Zheng, J., Huang, Y., Liu, J., & Yang, C. (2019). Investigation of a promoted you only look once algorithm and its application in traffic flow monitoring. *Applied Sciences*, *9*(17), 3619

[44] Park, J., Yun, Y., Eun, S., Cha, S., So, S., & Jung, J. (2018). Deep neural networks based user interface detection for mobile applications using symbol marker. *Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems*.

[45] Chen, L., Hermans, A., Papandreou, G., Schroff, F., Wang, P., & Adam, H. (2018, June). MaskLab: Instance segmentation by refining object detection with semantic and direction features. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Lake Tahoe, Nevada, USA*, 4013-4022.

[46] Chao, Y., Liu, Y., Liu, X., Zeng, H., & Deng, J. (2018, March). Learning to detect human-object interactions. *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision, Lake Tahoe, Nevada, USA*, 381-389.

[47] Najibi, M., Singh, B., & Davis, L. S. (2019, June). FA-RPN: Floating region proposals for face detection. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 7715-7724*.

[48] Ammirato, P., & Berg, A. C. (2019). A Mask-RCNN baseline for probabilistic object detection. *arXiv:1908.03621*

[49] Jabnoun, H., Benzarti, F., & Amiri, H. (2015, December). Object detection and identification for blind people in video scene. *Proceedings of the 2015 15th International Conference on Intelligent Systems Design and Applications, Marrakech, Morocco*, 363-367.

[50] Kim, Y., & Jung, H. (2018). Reconfigurable hardware architecture for faster descriptor extraction in SURF. *Electronics Letters*, *54*(4), 210-212.

[51] Borisyuk, F., Gordo, A., & Sivakumar, V. (2018). Rosetta. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 71-79.

[52] Song, Qisong, Shaobo Li, Qiang Bai, Jing Yang, Xingxing Zhang, Zhiang Li, and Zhongjing Duan. 2021. "Object Detection Method for Grasping Robot Based on Improved YOLOv5" *Micromachines* 12, no. 11: 1273.

[53] Xiao, J.; Hays, J.; Ehinger, K.A.; Oliva, A.; Torralba, A. SUN database: Large-scale scene recognition from abbey to zoo. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3485–3492.

[54] Rodríguez, Alberto, J. Javier Yebes, Pablo F. Alcantarilla, Luis M. Bergasa, Javier Almazán, and Andrés Cela. 2012. "Assisting the Visually Impaired: Obstacle Detection and Warning System by Acoustic Feedback" *Sensors* 12, no. 12: 17476-17496.

[55] Kumar, N., & Jain, A. (2021). Smart navigation detection using deep-learning for visually impaired person. *2021 IEEE 2nd International Conference On Electrical Power and Energy Systems (ICEPES).*252-256.

[56] K. Ding, X. Li, W. Guo and L. Wu, "Improved object detection algorithm for drone-captured dataset based on yolov5," 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE), 2022, pp. 895-899.

[57]M. Karthi, V. Muthulakshmi, *R. Priscilla, P. Praveen and K. Vanisri,* "Evolution of YOLO-V5 Algorithm for Object Detection*:* Automated Detection of Library Books and Performace validation of Dataset," *2021* International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems *(ICSES)*, 2021, pp. 1-6.

[58] *Z.* Wu et al.,*"*Using YOLOv5 for Garbage Classification,*"* 2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, 2021, pp. 35-38.*

[59] T. Zhang, B. Zhang, F. Zhao and S. Zhang, "COVID-19 localization and recognition on chest radiographs based on Yolov5 and EfficientNet," 2022 7th International Conference on Intelligent Computing and Signal Processing *(ICSP),* 2022, pp. 1827-1830.

[60] S. B. Kallara, M. Raj, R. Raju, N. J. Mathew, V. R. Padmaprabha and D. S. Divya, "Indriya — A smart guidance system for the visually impaired," 2017 International Conference on Inventive Computing and Informatics (ICICI), 2017, pp. 26-29.

[61] N. G. Bourbakis, and D. Kavraki, ―An intelligent assistant for navigation of visually impaired people,‖ in Proceedings- 2nd Annual IEEE International Symposium on Bioinformatics and Bioengineering (BIBE 2001), IEEE, 2001,pp. 230–235.

[62] T. Schwarze, M. Lauer, M. Schwaab, M. Romanovas, S.Böhm, and T. Jürgensohn, ―A camerabased mobility aid for visually impaired people,‖ KI-Künstliche Intelligenz,Vol. 30, no. 1, pp. 29–36, 2016.

[63] K. Chaccour, and G. Badr, ―Novel indoor navigation system for visually impaired and blind people,‖ in 2015 International Conference on Applied Research in Computer Science and Engineering (ICAR), IEEE, 2015, pp. 1–5.

[64] N. Karlsson, E. Di Bernardo, J. Ostrowski, L. Goncalves,P. Pirjanian, and M. E. Munich, ―The vslam algorithm for robust localization and mapping,‖ in Proceedings of the 2005 IEEE international conference on robotics and automation, IEEE, 2005, pp. 24–29.

[65] J.Bai,S.Lian,Z.Liu,K.Wang,andD.Liu,―Virtual-blind-road following-based wearable navigation device for blind people,‖ IEEE Trans. Consum. Electron.,Vol.64,no.1,pp.136–143, 2018.

[66] J.Bai,D.Liu,G.Su,andZ.Fu,―A cloud and vision-based navigation system used for blind people,‖ in Proceedings of the 2017 International Conference on Artificial Intelligence,Automation and Control Technologies, ACM,2017, pp. 22.

[67] B.Li,J.P.Munoz,X.Rong,Q.Chen,J.Xiao,Y.Tian,A.Arditi, and M. Yousuf, Vision-Based mobile indoor assistive navigation Aid for blind people,‖ IEEE Trans. Mob.Comput., Vol. 18, no. 3, pp. 702–714, 2019.

[68] E. Marder-Eppstein, ―Project tango,‖ in ACM SIGGRAPH 2016 Real-Time Live!, SIG-GRAPH ‗16, New York, NY,USA, Association for Computing Machinery, 2016, pp.25.

[69] J.Xiao,S.L. Joseph,X.Zhang,B.Li, X. Li, and J Zhang, ―An assistive navigation framework for the visually impaired,‖ IEEE Trans. Human-Mach. Syst., Vol. 45,no. 5, pp. 635–640, 2015.

[70] Y. H. Lee, and G. Medioni,Rgb-d camera based wearable navigation system for the visually impaired,‖ Comput. Vis. Image. Underst., Vol. 149, pp. 3–20, 2016.

[71] S.L.Joseph,J.Xiao,X.Zhang,B.Chawda,K.Narang,N.t. Rajput, S. Mehta, and L. Venkata Subramaniam, ―Being aware of the world:To-ward using social media to support the blind with navigation,‖ IEEE Trans. Human-Mach.Syst., Vol. 45, no. 3, pp. 399–405, 2015.

[72] A. Bhowmick, S. Prakash, R. Bhagat, V. Prasad, and S.M. Hazarika, ―Intellinavi: navigation for blind based on kinect and machine learning,‖ in International Workshop on Multi-disciplinary Trends in Artificial Intelligence,Springer, 2014, pp. 172–183.

[73] M. Sain, and D. Necsulescu, ―Portable Monitoring and navigation Control system for helping visually impaired people,‖ in Proceedings of the 4th International Conference of Control, Dynamic Systems, and Robotics (CDSR‗17), 2017, pp. 1–9.

[74] Z. Zhang, ―Microsoft kinect sensor and its effect,‖ IEEE Multimedia - IEEEMM, Vol. 19, pp. 4–10, 02 2012.

[75] A.Ali,and M. A. Ali, ―Blind navigation system for visually impaired using windowing-based mean on Microsoft Kinect camera,in International Conference on Advances in Biomedical Engineering, ICABME, 2017-October, 2017.

[76] C.Ton, A.Omar, V. Szedenko,V.H.Tran, A. Aftab, F.Perla, M. J. Bernstein, and Y. Yang, ―Lidar assist spatial sensing for the visually impaired and performance analysis,‖IEEE Trans. Neural Syst. Rehabil. Eng., Vol. 26, no. 9,pp. 1727–1734,2018.

[77] R. O‗Keeffe, S. Gnecchi, S. Buckley, C. O‗Murchu, A.Mathewson, S. Lesecq, and J. Foucault, ―Long range lidar characterisation for obstacle detection for use by the visually impaired and blind,‖ in 2018 IEEE 68th Electronic Components and Technology Conference (ECTC), IEEE,2018, pp. 533–538.

[78] M. Castillo-Cara, E. Huaranga-Junco, G. MondragónRuiz,A. Salazar, L. Orozco-Barbosa, and E. A. An-túnez,―Ray: smart indoor/outdoor routes for the blind using bluetooth 4.0 ble,‖in ANT/SEIT,2016,pp. 690–694.

[79] T.Ishihara, J. Vongkulbhisal, K. M. Kitani, and C.Asakawa, ―Beacon-guided structure from motion for smartphone-based navigation,‖ in 2017 IEEE Winter Conference on Applications of Computer Vision (WACV),IEEE, 2017, pp. 769–777.

[80] V. Nair, M. Budhai, G. Olmschenk, W. H. Seiple, and Z. Zhu, ―Assist: personalized indoor navigation via multimodal sensors and high-level semantic information,‖ in Proceedings of the European Conference on computer vision (ECCV), 2018.

[81] V.Nair,C.Tsangouri,B.Xiao,G.Olmschenk,Z.Zhu,and W. Seiple, ―A hybrid indoor positioning system for the blind and visually impaired using Blue-tooth and Google Tango,‖ J. Technol. Persons Disabil., Vol. 6, pp.62–82,2018.

[82]S.A. Cheraghi, V. Namboodiri, and L. Walker, ―Guidebeacon:beacon-based indoor wayfinding for the blind, visually impaired, and disoriented,‖ in 2017 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 2017, pp. 121–130.

[83] B. Vamsi Krishna, and K. Aparna, ―Iot-based in-door navigation wearable system for blind people,‖ in Artificial Intelligence and Evolutionary Computations in Engineering Systems, Springer, 2018, pp. 413–421.

[84] N. Sathya Mala, S. Sushmi Thushara, and S. Subbiah, ―Navigation gadget for visually impaired based on iot,‖ in 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), IEEE, 2017,pp.334–338.

[85] S. B. Kallara, M. Raj, R. Raju, N.-h. J. Mathew, V. R. Padmaprabha,and D. S. Divya, ―Indriya—a smart guidance system for the visually impaired,‖ in 2017 International Conference on Inventive computing and Informatics (ICICI), IEEE, 2017, pp. 26–29.

[86] D. Vera, D. Marcillo, and A. Pereira, ―Blind guide: Anytime,anywhere solution for guiding blind people,‖ in World Conference on Information Systems and Technologies,Springer, 2017, pp. 353–363.

[87] S. Gupta,I. Sharma, A. Ti-wari, and G.Chitranshi, ―Advanced guide cane for the visually impaired people,‖ in 2015 1st International Conference on Next Generation Computing Technologies (NGCT), IEEE, 2015, pp.452–455.

[88] A. Sen, K. Sen, and J. Das, ―Ultrasonic blind stick for completely blind people to avoid any kind of obstacles,‖ in 2018 IEEE SENSORS, IEEE, 2018, pp. 1–4.

[89] K. Patil, Q. Jawadwala, and F.-l. C. Shu, ―Design and construction of electronic Aid for visually impaired people,‖ IEEE Trans. Human-Mach. Syst, Vol. 48,no. 2, pp.172–182,2018.

[90] J. Borenstein, and I. Ulrich, –Applying mobile Robot technologies to assist the visual Impaired,‖ GuideCane, Vol.31, no. 2, pp. 131–136, 2001.

[91] A. A.Nada, M. A. Fakhr, and A. F.Seddik,―Assistive infrared sensor based smart stick for blind people,‖ in 2015 Science and Information Conference (SAI), IEEE, 2015,pp. 1149–1154.

[92] R.Jafri,R.L.Campos,S.A.Ali,andH.R.Arabnia,–Visual and infrared sensor data-based obstacle detection for the visually impaired using the google project tango tablet development kit and the unity engine,‖ IEEE. Access.,Vol. 6, pp. 443–454,2017.

[93] P. Marzec, and A. Kos, –Low energy precise navigation system for the blind with infrared sensors,‖ in 2019 MIXDES26th International Conference‖ Mixed Design of Integrated Circuits and Systems‖, IEEE, 2019, pp. 394–397.

[94] J. Ducasse, A. M. Brock, and C. Jouffrais, –Accessible interactive maps for visually impaired users,‖ in Mobility of Visually Impaired People, Springer, 2018, pp. 537–584.

[95] J. Albouys-Perrois, J. Laviole, C. Briant, and A. M. Brock,―Towards a multisensory augmented reality map for blind and low vision people: A participatory design approach,‖in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1–14.

[96] T. Götzelmann, and K. Winkler, –Smart-tactmaps: a smartphone-based approach to support blind persons in exploring tactile maps,‖ in Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, 2015,pp. 1–8.

[97] Q. Liu, R. Li, H. Hu, and D.-b. Gu, –Building semantic maps for blind people to navigate at home,‖ in 2016 8th Computer Science and Electronic Engineering (CEEC),IEEE, 2016, pp. 12–17.

[98] T. Götzelmann, –Lucentmaps: 3d printed audio visual tactile maps for blind and visually impaired people,‖ in Proceedings of the 18th International ACM Sigaccess Conference on Computers and ccessibility, 2016, pp. 81–90.

[99] C. Gleason, A. Guo, G. Laput, K. Kitani, and J. P. Bigham,–Vizmap: Accessible visual information through crowdsourced map reconstruction,‖ in Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility,2016, pp. 273–274.

[100] S.Caraiman,A.Morar,M.Owczarek,A.Burlacu,D.Rzeszotarski,N. Botezatu, P. Herghelegiu, F. Moldoveanu, P. Strumillo, and A. Moldoveanu, ―Computer vision for the visually impaired: The sound of vision system,‖ in Proceedings 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017, 2018- Janua, 2018,pp. 1480–1489.

[101]G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S.Yaacob, ―Wearable real-time stereo vision for the visually impaired,‖ Eng. Lett.,Vol.14,no.2,2007, pp. 6–14.

[102 D. Sato, U. Oh, K. Naito, H.-r. Takagi, K. Kitani, and C.Asakawa, ―Navcog3,‖ in Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS ‗17, 2017, pp. 270–279.

[103]A. Ganz, J. M. Schafer, Y. Tao, C. Wilson, and M. Robertson,―Perceptii: smartphone based indoor navigation system for the blind,‖ in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug 2014, pp. 3662–3665.

[104] S. Ren, K. He, R. Girshick, and J. Sun, ―Faster r-cnn:Towards real-time object detection with region proposal networks,‖ in Advances in Neural Information Processing Systems, 2015, pp. 91–99.

[105] J.Redmon, S.Divvala, R.Girshick, andA.Farhadi,―You only look once: Unified, real-time object detection,‖ in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.

[106] T.Mataŕo, F. Masulli, S. Rovetta, A. Cabri, C. Traverso,E. Capris, and S. Torretta, ―An assistive mobile systemsupporting blind and visual impaired people when are outdoor,‖in RTSI 2017 - IEEE 3rd International Forum on Research and Technologies for Society and Industry,Conference Proceedings, 2017, pp. 1–6.

[107] J. Lock, G. Cielniak,and N. Bellotto.A portable navigation system with an adaptive multimodal interface for the blind. AAAI Spring Symposium - Technical Report,SS-17-01 -:395–400, 2017.

[108] W. Heuten, N. Henze, S. Boll, and M. Pielot, ―Tactile wayfinder: a non-visual support system for wayfinding,‖ in Proceedings of the 5th Nordic Conference on Human computer Interaction: Building Bridges, ACM, 2008,pp.172–181.

[109] Ackland P, Resnikoff S, Bourne R. World blindness and visual impairment: despite many successes, the problem is growing. Community Eye Health. 2017;30(100):71-73. PMID: 29483748; PMCID: PMC5820628.

[110]. Adf M. Z. Hasan, S. Sikder and M. A. Rahaman, "Real-Time Computer Vision Based Autonomous Navigation System for Assisting Visually Impaired People using Machine Learning," *2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, Bangladesh, 2022, pp. 1-6.

[111]. U. R, H. P. S., R. N. G. and N. A, "Third Eye – A Smart Wearable Glass With Deep Learning Technology Powered With Artificial Intelligence for Visually Impaired People," *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, Bhopal, India, 2023, pp. 1-4

[112] D. Chaudhary, A. Mathur, A. Chauhan and A. Gupta, "Assistive Object Recognition and Obstacle Detection System for the Visually Impaired Using YOLO," *2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2023, pp. 353-358.

[113] J. Jayachitra, S. Muthulakshmy, M. A. Kumar and K. Elavarasi, "Design and Implementation of Smart Glove for Visually Impaired People," *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, 2023, pp. 1719-172.