Smart Visual Assistance for Visually Impaired

Neetha Das[1], Prajwal V[2], Pranav K Bharadwaj[3], Punith Kumar J[4] & Rahul V[5]

[1]Assistant Professor, Department of Computer science and Engineering
ACS College of Engineering, Bangalore, India
[2, 3,4 & 5]Student, Computer science and Engineering, ACS College of Engineering, India
***DOI**: https://doi.org/10.34293/acsjse.v2i1.27*

Abstract - *The visually impaired personnel face a variety problem in their day to day life. They come under very challenging situation on a daily basis. Walking with a stick and confusion in mind due to the trial and error method making it very hard to navigate in unfamiliar locations. With the advent of technology in this data driven world, the barrier for blind society can be cut down with innovative adaptations in technologies such as machine learning, deep learning and computer vision.*

QR codes can be used to inform some familiarized locations giving quick response in a closed environment for the blind person. This makes it is possible to help navigate the blind person in a closed environment.

The project attempts to help the blind people to get comfortable and be confident in unfamiliar locations through a speech assistive system. The projects consist of 3 modules namely:

(i) Object Detection and positioning.

(ii) Text to Speech Conversion.

The project helps the user to find obstacles in his path and avoid them, it also makes it possible to locates some of the locations in a closed environment, this is done with the help of CNNs and some APIs like Opencv, Yolo and Tensorflow, QRCode, pyzbar (used for generation of qrcode). The detected items are reported to the user in the form of speech with the help of some python libraries like pyttsx3.

***Keywords:** Deep Learning, Neural Network, Object Detection*

I INTRODUCTION

A. Machine Learning

There are more than quintillion bytes of data that are being generated on the daily basis, in this data driven world these data can be used to train machines so that they can learn themselves. This process of making machines/computers to learn themselves is stated as Machine Learning.

Machine Learning is a area of computer science and a sub - group of artificial intelligence (AI) which enables the computer the ability to learn from the statistical data. Machine Learning has played prominent roles in shaping today's data driven world, it provides us with the necessary analyzing tools to study the data.

Incorporating machine learning techniques to analyze large amounts of data that can help find out if hidden patterns were missing from raw data. This process of data analysis is known as data mining.

Machine learning in general can be summarized as:

- It is necessary to rearrange the list and the required list rules for problems that are already solved.
- Machine learning algorithm can make code easier and more efficient with code.
- Machine learning methods can find solutions to complex problems if there is no good solution.
- Machine learning can be flexible and can be used under Fluctuating environments.
- Machine Learning can get sit reps for complex problem with large amount of data.

Many sub- branches of Computer Science like Machine Leaning, Deep Learning, Optical Character Recognition (OCR), and Natural Language Processing, Computer Vision can be incorporated together to make human life simpler.

Initially, Machine Learning was generally used for playing games, simple learning problems and many problems of that kind. This might be because of lack of computing infrastructure and limited source of data. As days progress machine learning is much more than just gaming and simple learning program.

Natural language processing is used for designing or developing medical and psychological assistive systems. Computer Vision were used for digital processing of images, Computer Vision is used in the medical field for recognition of cancer cells, cell disorders, identification of fractures, blood related diseases with the advent of digitalisation of medical reports. Many government Agencies are using facial recognition technology in order to recognise and distinguish the criminal faces and locating missing persons.

Today's world machine learning is being a part of human life without which human life is unimaginable, this might be because of the production of absolutely huge amount of data and increasing computational infrastructure which are capable of processing those huge amount of data being produced.

In recent times, many companies are making use of the machine learning techniques for research purpose and development of smart solutions for complex problems. Online shopping portals are using customer information in order to provide smarter product recommendations, improving both customer experience and satisfaction with fulfilment of business requirement. Social networking sites use deep learning concepts for segregation of type of posts, users and personalising ads. These personalising ads are the prominent source of revenue of these social network sites. Spam messages are classified based on these similar kind of techniques. Facial Recognition technology, video recognition technology which are being predominantly used in mobiles and computer devices use deep learning concepts for their processes.

Every software application, sometimes operating systems are being automated for making user experience flawless and assistive. Recent development of detecting edge technology, companies are keen on developing automated vehicles and have achieved a certain amount of success in developing/Designing them. Since there is a lot of scope for development in this field companies are working on integrated system which can minimise the errors in the sense of security.

Machine Learning can also be used in researching medicines and their structures to get a clear understanding of the atomic structure of the molecules, these learnings are helpful in creating the cure for a disease. Smart assistive systems for people with hearing aid.

B. Deep Learning

Deep learning is a type or section of machine learning. Learning that employs artificial intelligence functions capable of simulating the inner workings of the human brain. This enables us to develop patterns and processes for making decisions. This is also possible with unsupervised data, and the data in general is unstructured.

How does Deep Learning Works?

In this digital age, in-depth learning has grown exponentially as the result of the large amount of data generated from many sources such as social media platforms, integrated search engines, e-commerce platforms, and many others are examples. The data is officially known as big data.

This data is also readily accessible by the applications that is cloud computing, by the way this is unstructured data that is so large that might take several years to process through. Companies have already started adapting artificial intelligence systems for automated processing of data which has huge potential in generating business.

Deep Learning can be used to find hidden patterns and possible vulnerabilities in the data to detect fraud or anomaly from happening.

C. Object Detection

At first, we tried to implement hardware or IOT based systems but due to ongoing lockdown we had to go implement the project using Third Party Applications.

1. The project is based on a system which is designed in such a way that an android application (assuming that it is to be implemented on an android device). The android device will capture real-time image frames and will send it to a Server integrated on a networked server based on a laptop where all necessary computations will take place
2. A Networked Server on a laptop will make use of an SSD (Single Shot Detector) detection model which is pre-trained which is indeed trained on COCO DATASETS. Then the testing is done to detect the output class with a precise accuracy metrics.
3. After the process of testing is done with the help of voice modules the class of the object will be converted into a default voice notes. The voice notes will then be sent to the blind victims which will assist them.
4. Along with the process of Object detection, we have made use of an alert system where approximation will can be calculated. If the visually impaired person is very close to the object in the frame or is in a safer distance, the system will generate a voice-based outputs along with the distance units from the frame to the object.
5. After the item is found, it's also critical to notify the blind person of the presence of something in his path.

D. PYTTS

In the voice production phase PYTTSX3 plays a crucial role. Pyttsx3 may be a Python translation library that Text is translated into speech. The whole library is consistent for both Python 2 and Python 3. Getting regard to pyttsx3. for instance an engine, a mill-hand called pyttsx3.init () requested by the appliance . Pyttsx3 may be a tool that converts text into easy speech. Pytts may be a text-to speech library available in python, which is employed to convert text to speech. This library works offline and it's compatible with python two and python three. Using pyttsx3, it is easy to convert the entered text which is generated into speech by this module. This module has high compatibility with systems. This module can assist with two voices first of which is a female voice and second being male voice which is providedbysapi5. The pyttsx3 module supports 3 text- to –speech engines which are:

Sapi5 (Speech Application Programming Interface): It's an API developed by Microsoft and which supports text to speech on windows systems.

Nsss (NSSpeech Synthesizer): This supports text to speech on mac os x developed by Apple.

Espeak: It's an open source software which supports text to speech on platforms like Linux and Windows.

E. Scope

As discussed before the population of blind is a few millions and many people have unaccountable problems of their own.

The smart vision assistance for visually impaired mainly aims to replace or beside the guiding sticks with and application that makes the blind personal life independent to some extent by detecting the objects and indoor navigation.

The project primarily focuses on developing user friendly to the end user with the minimal cost and high practicality when compared with all other existing systems and heavy, bulky systems.

This way the blind person across the globe could be independent on many inevitable situations on timely updates the smart visual assistance would become a part of the visually impaired person as it makes his life easy that ever before.

F. QR Code

The basic idea is to use a QR code to make a picture which will be scanned with any modern Smartphone (which will have a QR code reader embedded in it) which will be converted to something useful. QR codes are commonly wont to contain web address information, links, location tracking. But they will even be wont to direct Smartphone users to several other media (examples are videos, photos, product details etc). QR code is often used by advertisers to make the process easier, and therefore, attracts more people to visit a web page or content clip. This is why advertisers should always have a QR code on banners, business flash cards, posters, advertising magazines, and more.

In fact, the QR code applies to the same bar code used in the file, supermarkets. QR codes can be defined as a machine that can be machine-tested or a content clip which can be

quickly scanned and read using a Smartphone camera. Each one the QR code contains the dots and the number of black squares that represent certain information. As soon as the Smartphone scans the available QR code using the camera, translates the image into information that can be understood people.

Officially, a QR code can be called a piece of data or data entered. The a piece of information or data contained in a QR code can be a combination of alphanumerical, numerical, binary or Kanji (Kanji is actually the Chinese method characters used in modern Japanese writing system).

The usage of modern day QR code mainly consists of seven parts. Each one of these parts creates a sort of patterns of pixels that resemble the structure of a cross word puzzle. Each of the element of the pattern has a specific purpose that may convey certain information or data through the Code such as the direction of the print, timing, tolerance of error, and empty spaces to differentiate the Code from that surrounds it.

II LITERATURE REVIEW

Object Detection with Deep Learning: A Review

The paper mainly focuses on providing an insight of comparison on various object detection methods using deep learning. The authors have reviewed various existing approaches for object detection frameworks starting from the Generic object detection framework and some key modifications to improve the recognition further and Regression/Classification based framework. Experimental analysis and meaningful conclusions have been provided to draw useful information for an object recognition system. The experimental analysis, include results obtained from various convolution neural network architectures like R-CNN, Fast R-CNN, Faster R-CNN, SSP Net in the generic object detection framework. The Regression/Classification based methods included YOLO and SSD and their experimental analysis of the same have been provided. The paper gives an insight of the existing methods for object detection and their pipelines for better performance. All the experimental analysis was performed considering standard metrics. To compare object detection of the above methods they have used 3 bench mark datasets. From the experimental analysis of the generic and Regression framework the authors could identify that the convolutional neural network pipelines were able to perform better than the Regression Framework. Furthermore, the faster R-CNNs are considered to be the fastest among the faster R-CNNs while using the Microsoft Coco Dataset which composed of 300000 images. The authors have also extended their analysis on these methods on face detection and pedestrian detection. They have given a number of promising modifications that can be updated to the methods in order to optimize and improve cascading the model in the future.

Developing Walking Assistants for Visually Impaired People: A Review

The paper provides an insight for developing assistive system for the person with visual impairment. The authors have tried to review all the existing technologies in developing walking assistants in recent times for the blind. The authors discuss the innovative technologies for developing the walking assistant listing out the merits and demerits of each

of the framework. The update draws a schema of mobile assistants using computer vision, sensors and a smartphone-based program. The review also includes challenges for each of the frameworks listed above. The sensory assistants such as walk-in radar assistants, a wearable navigator, a smart stick, an electronic walking cane and an intelligent visual eye are reviewed experimental analysis in each program has been done and planned. The summaries are extremely easy to distinguish the performance factors of the frameworks that are used. This summary includes the salient features such as vital hardware components, detection range, cost , area coverage and weight of the device. The computer vision based walking assistants which included devices like wearable navigation system, smart guiding glasses and prototypes like eye sight have been reviewed and tabulations are provided. The review provides extreme prescience for the future scope of development for assistants for the visually impaired that can make them to easily navigate from and to locations. Smartphone based methods which used devices like Guide beacon, head-mounted device and Eyemate are also reviewed. From the tabulations that are provided by the authors the smartphone based methods and computer vision based frameworks are found to be cost efficient and precise. The authors have also tried to throw light on some modifications for the enhancement of these assistance systems and make them accessible.

Real-time RGB-based 3D Object Pose Detection using Convolutional Neural Networks

The authors of this paper have discussed the 3D pose detection of an object using innovative methods such as TQ-Net. The technique uses a single neural network to make the prediction about the quaternion and the translation vector is further converted to rotation matrix. Addition of normalization layer is done over the quaternion to get more precise results.The acquisition pipeline in this method is designed in such a way that initially the standard acquisition algorithm such as MTCNN,YOLO, SSD to consider the object and procure a boundary The bounding box is then resized to fit a particular proposal is inserted as input and location information into TQ-Net to further process the acquisition of the entry. The CNN architecture description and pictorial representations are very helpful in providing an insight of the approaches that are used. The dataset is divided 70% for training and the rest for testing. Then the model is optimized and the results are obtained. The positioning of the image is made using a parameter called projection error. Experiments on Line-Mod are made with ACVV database which consisted of over 18000 images with more than 15 different objects and translation vector, ground truth pose and rotation matrix. Furthermore, the computational efficiency is calculated considering parameters such as the effectiveness of q normalization, runtime efficiency and accuracy assessment. From the experimental analysis the method used is efficient. Various tabulations and graphical representations have been made to demonstrate the same.

Convolutional Neural Network-Based Real-Time Object Detection and Tracking for Parrot AR Drone

The authors used Convolutional Neural Networks (CNN) to track and detect the target object, moving or motionless, for a drone. Convolutional Neural Network(CNN) can be used

for object detection and object tracking and an AR Parrot Drone used arsliastitinto it. Basically, the convolutional network is trained on a single class of object. The architecture of the convolutional neural network used a single shot detection architectureas it combines the computational effect of the YOLO with the accuracy of regionselectors. The CNNs are trained for a single class of object (human in this case). The basic specification of the hardware is provided by the authors, the drone size It employs roughly 50cm x 50cm and weighs approximately 500g. The drone's design allows it to be used in both indoor and outdoor environments. The AR Drone 2 has two cameras, an Inertial Measurement Unit (IMU) with a 3-axis gyroscope, a 3-axis accelerometer, a 3-axis magnetometer, a pressure sensor, and an ultrasonic altitude sensor, as well as a pressure sensor and an ultrasonic altitude sensor. The drone would be linked to a Computer via Wi-Fi, and also the Robot OS (ROS) framework is used to execute object detection but also algorithm tracking. ROS is composed of two nodes: Node 1 is a tracking as well as tracking area, and Node 2 is an AR Drone driver package. The an object found in a picture; CNN returns the picture and therefore the border box to the thing . This boundary box ensures details on the form of the thing . Additionally, a mistake is calculated in between image centre but also, consequently, the object centre AR Drone will use a Wi-Fi linkage to send a photograph taken by the front camera to a PC. CNN, which has been trained to seek out images, transmits the received image to out what you would like (person). The CNN utilized in this project is predicated on the event of one shot detector (SSD).The results have been tabulated in the for a variety of distances and positions of the human. The experimental analysis and calculations will show that the single shot detection architecture is much more accurate to detect the object with 98% accuracy in comparison with the YOLO algorithm which delivered a accuracy of 96%. Future scope for improvement in terms of processing speed and many other parameters of the processing unit are identified. These system scan predominantly be implemented onanunnamedaerial vehicle.

Electronic Travel Aid System for Visually Impaired People

The paper discusses about the electronic travelling aid that act as navigation system for the blind community. The paper provides an insight about electronic travelling aid that are developed for navigation system in anun familiar environment. They have designed alert systems foremergencies, image processing is used for obstacles and distancing sensing through IR sensors. The alert system is designed to help the blind via web access. They built an electronic navigation system with three parts of the camera module and a sensor network that works on the grid. The camera module has two raspberry pi rev 1.3 cameras that can take high resolution images. The sensor list had three infrared levels SHARP GP2Y0A710K0F sense of proportion. Sound recognition systems display audio signals about objects or obstacles. The GY-NEO6MV2 GPS Module is employed to seek out the present GPS coordinates of the user and therefore the refore the Arduino Pro Mini ATmega328 controller is employed to trace the raspberry pi module and the existing links are sent to the Google MAP API to seek out the present location. The google direction API is used for finding the direction of the selected place. The two cameras present in the camera module scaptures

images in a similar way as that of human eye and produce disparity of images. Calculations and pictorial representations about the locations are represented in the paper. The flow diagram of each of the parts of the travelling aid provides an insight of the system. The experimental analysis which are group on the parameter height of the object show that the error percentage of the calculated distance with the actual distance of the object is less than 5% for the object of 30cm, less than 3% for the object of height 90cm, less than 6% for the object of 150cms. The authors have also discussed a number future scope of improvement in the electronic navigation system.

Smart Eye for Visually Impaired-An Aid to Help the Blind People

This paper represents the the concept of developing a smart system which can assist people with visual impairments in their daily activities. They suggest a "smart eye system", a device with Smart Visual Assistance for a Visually Impaired voice-enabled program that will directly identify a person with a challenge in their daily operations tasks. These device incorporates a spread of obtainable technologies and integrates into one device used for multiple functions which will be employed by the blind person. The written paper provides insight into a planning of such a system as well as therefore the difficulties associated in fence construction. Their main goal is to create a coherent system altogether aspects and supply an area for future extensions. The varied technologies and methods wont to verify these things Image processing and embedded technology are examples of such technologies. This app includes a route navigation system that is GPS-based. Voice detection detection, image processing to identify people who do not require an internet connection are also not supported Android features. The smart eye is equipped with a Raspberry Pi-like microcontroller, as well as a camera, sensor headphones, and other accessories. The system planning represents a much better understanding. Several features, including the camera, GPS module, Infrared (IR) sensor, and lightweight Dependent Resistor (LDR), were disrupted with the Raspberry Pi's proper performance. All items and sub-items are linked together using General Purpose Input Output (GPIO) anchors. Also it incorporates a frame for a 16GB Secure Digital (SD) data storage card. Its as a whole system is commonly divided into two small elements, such as route navigation and one to urge face - to - face. A barrier (object) detection system is part of the delivery module. The device guides the user to a previously saved location, and their path is saved. With the help of buttons, the user can select a destination. The system has four buttons from which the user can select the desired location, and the navigation route for these areas is provided as a voice command via the sound unit. Face and sight module makes use of the camera to take pictures of the person in front of the user's face and save them to the SD card, the last card inside the microcontroller. When a person is observed.

With the help of the Open Cv software, the capture and notification processes are completed. The proposed system has been tested for features such as routing and active human identification. They do, however, indicate that there is an expansion rate in many areas and integrate it with Google Maps for better GPS functionality.

III PROBLEM IDENTIFICATION

There are upwards of 286 million people who are visually impaired all throughout globe, with 40 million people anticipated to be totally blind. We come across many such people in day to day life and never know their scuffle.

There is enormous financial load globally associated with visual impairment which is estimated to be US$ 244 billion.

Navigating around places is perhaps the most difficult task for a blind person, especially one with total loss of vision.

This project aims to make their lives much easier than before and can make them individually independent.

In this world for the sighted, being blind puts a pressure to be normal, means that there will be few mishaps like stumbling upon the chair, tables or glass, these can be viewed as the inability of the visually blind to perform certain tasks but the reality is that they stem from the unreachable of our world.

IV OBJECTIVES

1. Our main goal is to simplify the user system and make communication between the two organizations as easy as possible. The whole project depends on the Smartphone App and its reliability. Do all the calculations again statistics. A separate data is created, where the description of the objects is found. At system level we can say that youth resides in a real-time app running on Smartphone..The smart vision assistance for visually impaired mainly aims to replace the guiding sticks with and application that makes the blind personal life independent to some extent.
2. The project primarily focuses on developing user friendly to the end user with the minimal cost and high practicality when compared with all other existing systems and heavy, bulky systems.
3. This way the blind person across the globe could be independent on many inevitable situations on timely updates the smart visual assistance would become a part of the visually impaired person as it makes his life easy that ever before.

V PROPOSED METHODOLOGY

A. Object Detection Layout

At first, we tried to implement hardware or IOT based systems but due to ongoing lockdown we had to go implement the project using Third Party Applications.

1. The project is based on a system which is designed in such a way that an android application (assuming that it is to be implemented on an android device). The android device will capture real-time image frames and will send it to Server such as a networked server based on a laptop where all necessary computations will take place
2. A pre-trained SSD will be used in the Laptop-Based Networked Server. (Single Shot Detector) detection model which is trained on COCO DATASETS. Then the test will be conducted and detect the output class with a precise accuracy metrics.

3. After the process of testing is done with the help of voice modules the class of the object will be converted into a default voice notes. The voice notes will then be sent to the blind victims which will assist them.
4. Along with the process of Object detection, we have made use of an alert system where approximation will can be calculated. If the visually impaired person is very close to the object in the frame or is in a safer distance, the system will generate a voice-based outputs along with the distance units from the frame to the object.
5. After an object is found, It is critical to alert the visually impaired person to the presence of the object in his or her path.

Technical Setup

B. Module Split Up

The four sections are:

Section 1 –Capturing The Image Frames.

Section 2 – Preprocessing The Image.

Section 3 – Identifying The Object and Its Distance from The Captured Image.

Section 4 –Generate an Audio Signal for The Identified Object.

Section-1

Capturing the Image

1. This section is responsible for capturing image frames through camera.

2. The system is implemented within in android or a web application enabling the user to capture and send it to the network connected laptop.
3. The setup is made in a way that it is capable of capturing real time video frames so that we can store and process the frames.

Section-2

Preprocessing the Image

1. The laptop consists of image preprocessing techniques like contrast enhancement, remove the noise, segmentation and many other techniques with have to done to images so the feature extraction could be done.
2. The images have to be resized before sending it to the object detection model in order to detect the object.

Section-3

Object and Distance Identification

1. The object detection models comprising of pre-trained models for the coco dataset, which is capable of identifying 90+ varieties of objects.
2. These object detection models use algorithms like Yolo, fast r-cnn, r-cnn or ssd to detect and identify the object.

These models take input images of a particular size and preprocessed.

Section-4

Generating an Audio Signal

1. This section generates audio signal to inform the user about the object detected.
2. The system is responsible for alert messages for the user about the object if it is very close to the user or exceeding the certain distance of safety.

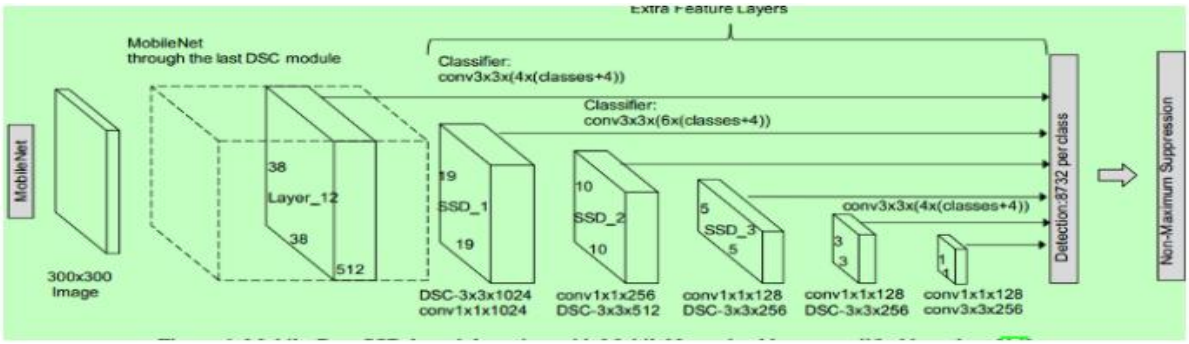
C. TENSORFLOW APIs

We have tried to implement it using the TensorFlow APIs. The advantages of using this one APIs is that it provides the set of common operations/options. By using this we don't have to start to program from scratch. This is the efficient way to implement or writing a code. They are a time saver because of the convenience. If we have to create a deep learning network, we have TensorFlow API that basically gives a structure to build the network hence providing the solution to the problem for object detection. We get the trained models in the framework it is also known as ModelZoo.The model results in a collection of COCO dataset, KITTI dataset, open Images Dataset. But for this project we make use of only the COCO dataset.

D. Single Shot Detection

We get a lot of pre-trained models with Tensorflow. We can use any of them for the requirement of the project. If we want the faster accuracy shall go with the SSD Detection

and say for the better accuracy use Mask RCNN but for the smooth performance, we go with SSD Mobile_Net DETECTION. Here we elaborate SSD Algorithm.



ASSD has 2 components: with SSD head but also a backbone model.

It's a feature extract or that is basically trained image classification network. With ResNet this is also a network trained on the Image Net from which we know that fully connected classification layer has been eliminated.

We have a single specification attached to the spine so the result is in the final stages of working with the results defined as binding boxes and categories of local objects. In this way we get a deep neural network that will give / remove semantic meaning or understanding with the input image provided but does not re-create the image location even though it has a low resolution. For example, with the input image that results in 2567*7 feature maps in the ResNet34. SSD uses the grid to divide the image and for object detection in the region the grid cell is responsible for that image. By Detecting object, we mean predict the location and class of the object that exists in that region.

E. Anchor box

For each cell grid on the SSD many anchors or previous boxes are provided. The configuration boxes provided are defined, because each shape and shape are responsible within the grid. While using the SSD simulation section, then there is a corresponding match from the anchor box to the binding boxes for each ground element of the object inside the given image. The highest level of distinction is responsible for the process of predicting an object's class and location Following network training, these assets are used for network training as well as predicting findings and their locations. The formalised paraphrase Each anchorbox has a different aspect ratio and zoom level. Different things have different forms; one is shorter, while the other is longer and can vary greatly in degree. SSD configuration, on the other hand, allows for some of the previously defined aspect ratios; these specified parameter ratios are associated with each grid cell zoom and / or scale scale.

F. Zoom Level

There is no strict rule for the anchor boxes to be having same size of the grid cell. The user is also interested in finding the varying sizes of object inside the grid cell. The zoom

parameter will be used to determine the how the size of anchor boxes is scaled upwards or downwards.

G. Mobilenet

This model has the concept of a Mobile_Net model but has been used to a cleverly divided depth and will create a featured Convolutions. It transforms a common basic variable into a deep intelligent combination. These 1 * 1 concepts are also known as incomprehensible convolutions.

If we want to make MobileNets work, each smart integration channel uses one universal filter. 1 * 1 convolutions are used in transparent convolutions which enables them to join deep intellectual compounds. With standard convolution both filters combine new input input and single-step effects. Depth of targeted identification depth is divided into 2 layers A layer designed to filter the purpose and a different layer of purpose integration. This factorization method will work to reduce computer computing by its model size.

H. Depth Estimation

Depth estimation or extraction features are the algorithms and techniques that aims to obtain the representation of the structure involving a spatial scene. Quite formally, depth estimation is used in calculation of the distance between two object. The prototype we have designed has been used for use by officials who appear to be prohibited from removing obstacles in order to deal with obstacles (objects) in their wniy. In finding this, we first need to find to find any offline object to the they are closely monitored in the immediate state of situation. After the hurdle (objects) is detected using real time frames A rectangular box-like formation is generated around the identified object..

Distance Approximations

If an acquired object occupies a large number of pixels on the frame, the system calculates the distance to the object from the individual based on certain constraints.. The following code usage is performed for detection items and return distance and location calculation.

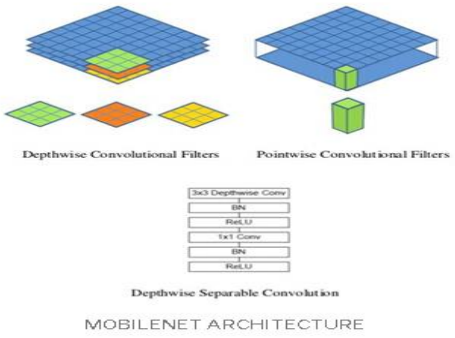
In this section, we have made an establishment of Tensorflow session which is comprised of the Crucial Features for the process of Object Detection. Therefore, for

upcoming analysis further iterations are carried out through the boxes. The Boxes can be defined as an array, inside of an array. Therefore, for the further iterations we need to define the following conditions.

The Index of the array boxes are represented by the term i. The analyzing of the score of the boxes is carried out by the index used. These indexes can be used to find different items. Using this meausurement of the width of detected object can be found. This is done by entering the maximum area in terms of pixel patterns.

Now that we have obtained the center of the two elements by dividing the elements by two and subtracting the same start axis coordinates. The center of our detected rectangle structure is calculatedungth method. dot drawn in the center of our rectangle at the last step. The default parameter for drawing boxes with 0.5 points, if scores [0] [i] is bigger than 0.5 (e.g. equal or >50 percent) and assume the thing has been found, if scores [0] [i] is bigger than 0.5.

Mid x is the centre of the X axis in the above formula, and mid y is the centre of the Y axis. If the apx distance range is set to is less than 0.5 and if mid_x is more than 0.3 and mid_y is less than 0.7 we can conclude that the find is very close to the visually impaired person. Using this code, the related distance to receipts from the visually impaired can be calculated. After the acquisition process is done the above code can be used to determine the corresponding distance of the acquisition from a visually impaired person. If the findings are too close, a warning or warning will be given to the visually impaired module.



MOBILENET ARCHITECTURE

I. Voice Generation Module

The Artificial production or generation of human speech can be termed as speech synthesis. This technique which is based on machine learning can be applicable to text-to-speech, generation of music, generation of speech, speech-enabled devices, systems used for navigation and can provide accessibility for assist visually impaired people.

Following the detection of an object, the most important thing is to inform the visually impaired person about the presence of the obstacle (object) in his/her path.. PYTTSX3 plays a vital and important role for purpose of voice generation module. PYTTSX3 may be a library utilized in Python which can are often used to converting text to speech. The PYTTSX3 library is compatible with both python 2 and python 3. A factory function is invoked by a

application to obtain the reference to a PYTTSX3 English instance. P YTTSX3 can be defined as a tool that is used to easily convert the given text into speech.

When the thing is detected, the algorithm we use works well, and the distance rating is calculated with the help of the cv2 library and cv2. The generated text is displayed on the screen using the putText () function. We will use Python-tesseract for character recognition to detect hidden text in image pixels. With a computer virus, the OCR (Selected Character Selection) can find the image in text and insert it in an easily recognisable form. Scanning and pixel analysis of the image completes the text acquisition process. Python-tesseract recognises and skims the embedded text within the images. Furthermore, these documents are displayed within the PYTTSX Engine, the factory function is requested by an application referred to as pyttsx3.init (). Over construction phase, pyttsx driver. Drive Representative object are often called a loading engine and talk driver from pyttsx. Drivers. After the development phase, the engine object is employed to register and register the return of events by system; create and stop speaking; find and found out speech engine properties; and starting and fixing event logs.

As a result of this process, audio commands can be performed as output. If the found item is too close, then the system creates a warning message that says “Warning: The item (item category) is too close to you. Stay alert! ”. When an acquired object is far safer than the system it produces the message "Object is far from safe". The production of voice messages can be achieved with the help of specific libraries such as pytorch, pyttsx3, pytesseract and engine.io.

The Pytorch is basically a library used in machine learning. The Pytorch library can be mainly applied to the audio domain. The Pytorch library helps in accessing the voice files in standard format such as the MP3 format. Therefore, the Pytorch can be used to manipulate the sound properties such as wave form, wavelength and frequency of the sound. The numerous options available for the synthesis of audio can be verified by taking a closer look at the Pytorch Functions.

J. Algorithm to be Implemented

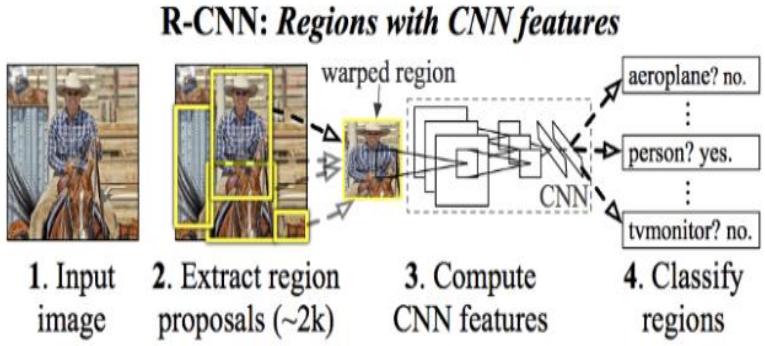
1. R-CNN (Region based convolutional neural network)

RCNN can be defines as approach which combines regions of rectangular proposals with the features of convolutional neural network. The RCNN algorithm proposes bunch boxes to an image pixel used to check if any of these boxes contain anything in the frame, instead of working in multiple regions. The RCNN method also uses selected searches that can be used to extract these boxes from the pixel image and these boxes known as regions.

The steps followed RCNN to detect objects can be summarised as:

1. Initially we look at a pre-trained convolutional neural network.
2. In the next step the model is saved. Final network layer according to the number of classes that need to be identified are trained.
3. The third phase involves the process of determining the region for each benefit picture. These regions are then redesigned to be able to match the input size of CNN (Flexible Neural Network).

4. After obtaining these regions, SVM (Support vector machine) is blocked which is not possible edit the findings and their origin. In each category, one banner SVM (support vector machine) needs training.
5. As a final step in the process, a straightforward reversal model is used produce strong bond boxes for each identified item found in picture frame.



RCNN Model Workflow

The working of RCNN can be summarised as follows:

1. The first phase CNN network is targeted to to perform tasks such as image classification.
   As an example, VGG or ResNet which can train on Image Netdataset. The process of task classifictitinn freely gives the number N of classes.
2. Selective search can have done by proposing category independent regions of interest (took a candidate per images). These regions may generally contain objects such as target objects and they can be of different sizes.
3. The CNN requires the region candidates to have a fixed size by wrapping them.
4. The best fix for CNN is the continue donwrapped proposal regions with K + 1 classes. Wnech of one class additional refers to background. During the fine adjustment of the process of, the learning rate should became smaller and took a mini-batch over samsmles the positive cases since most of the regions which are proposed are nothings just backgrounds.
5. Considering each and every imageregion, the propagation of forward press through the CNN results in the generation of feature vector. The binary SVM (Supportvector machine) consumes the generated feature vector for each class independently.
   The samples available are the IOU-proposed districts (union crossroads) that exceed the limit> = 0.3, and in the event of negative samples do not work.
6. In order to reduce and minimize local performance errors, a retrospective model that can be used to repair a window with predictive detection in offset box offset using CNN features can be trained.

Fast R-CNN (Region based Convolutional Neural Network)

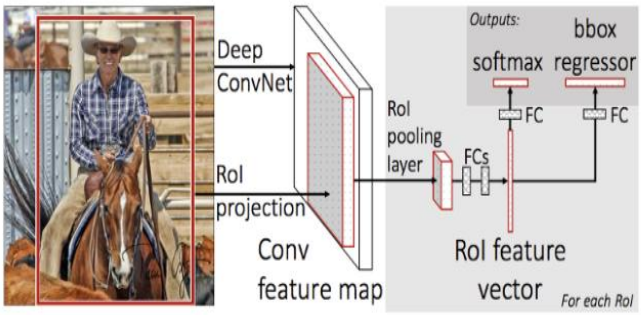
In order to deal with the issues related to RCNN, fast RCNN was proposed. The fast RCNN considers the whole image frames and region proposals which feeds as input in its CNN architecture in one forward propagation.

Fast RCNN contains CNN (usually pre-trained in the ImageNet segmentation process) where the final component can be replaced by the "ROI Pooling" layer and its final FC layer can be replaced by two-phase-specific binding box binding branch and layer (K + 1) section of softmax.

In each region suggestion we can extract the corresponding part from the feature map. This can be viewed as a map of the regional proposal feature. We are looking at a regional suggestion feature map from a given feature map and we can change the map size to the default size using the aggregate layer. This integration layer can be called the regional integration layer (ROI).

After this we soften the map of the feature of the regional proposal which is the size of the correction. This may produce an element vector, which can always be the same size.The output element vector can be used as an input in the last component. these feature carriers are fully connected layers that may be required to exit. The first result is a layer of softmax class, in which we can determine the category of the object we receive. the second output is the Bounding box Regressor, where we can remove the connecting box connecting to each class item.

The operations of convolutional features are performed only once in the fast RCNN because of which it need not extract all the 2000 features every time due to which it is more efficient when compared to RCNN.



Fast RCNN Model Workflow

The working off RCNN can be summarised as follows, many steps are same as that of RCNN:

1. Initially, CNN convolutional neural network is pre trained on image classification tasks.
2. Selective search is done on the proposed regions (2K candidates per images).
3. As the next phase, the CNN which is pre trained is altered.

Enter the final layer of CNN's Max integration are made to pre-train along the ROI layer. The ROI integration layer can generate regional suggestions for fixed lengths. Since many regional proposals for similar images are very scattered, CNN computer sharing makes perfect sense.

Instead of lasts of tmaxlayer (K classes) replaced the layer with a full softmax and received full layerover K + 1 classes.

As the last phase the model can be branched into two layers of outputs:

Softmax rating includes K + 1 classes (a feature similar to RCNN, +1 “background” category), in which different distribution opportunities are excluded per ROI.

A bounding-box regression model that can predict the relative off sets to the ROI which is original for each of the K classes.

A. QR code

A QR (Quick Response) code can be defined as a two dimensional image that is similar to a common barcode, that features storage capacity of large information and rapid readability. The QR code is typically are square shape image and it is comprised of two regions first being a functional region and the second being a coding region. The Functional region of the QR code is a combination of correcting, localising and seeking graphs. The coding region of the QR code is typically description of the format come version the characters of data. The QR code can be read and be decoded buy smart phones in a very manner which is straight forward.

In order to navigate the user from source to destination the QR codes can be embedded with grid locations. Costa-Montenegro has made a demonstration open navigation system working indoor by using the QR code. The requirements of this method are the deployment of locations and map servers which can increase the cost. Alghamdi has also proposed an indoor navigation method based on RFID and QR codes. Because of high cost the large scale deployment of this method has been hampered. Chiouhas also made similar kind of proposal using QR codes and PDR, but using this algorithm lower accuracy can be obtained when compared to other methods.

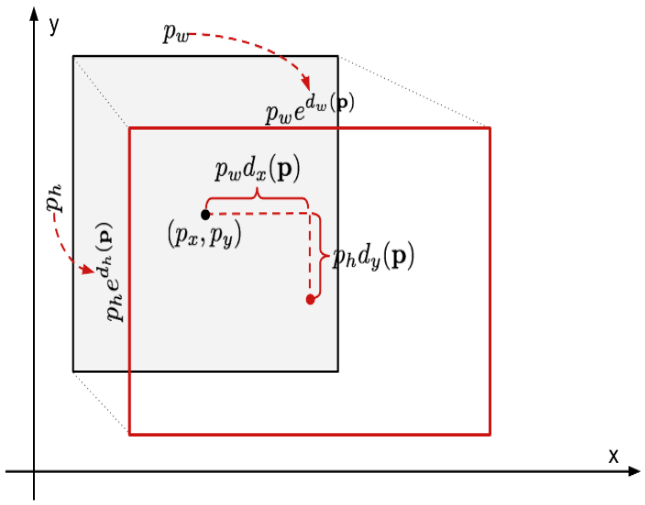
Among these methods indoor navigation by the QR code is the accessible one for most of the people. To read a QR code we need a scanner to capture the image and to process the image we can use a python library called zbar which supports both python 2 and python3.

B. Bounding Box Regression

By providing the link of the measured bond box p = (px, py, pw, ph) p = (px, py, pw, ph) (intermediate link, width, height) and the corresponding real-world link box g = (gx, gy,

gw, gh) g = (gx, gy, gw, gh), the regressor is set to read the continuous transition between the two centers and to adjust the penetration rate between width and height. All transformation functions take pp as input.

g^xg^yg^wg^h=pwdx(p)+px=phdy(p)+py=pwexp(dw(p))=phexp(dh(p))g^x=pwdx(p)+pxg^y=phdy(p)+pyg^w=pwexp(dw(p))g^h=phexp(dh(p))



A diagram of the transition between the predicted boxes and the reality below.

The obvious advantage of using such a change is that all binding box repair operations, di (p) di (p) where {x, y, w, h} {x, y, w, h}, which can take any possible values between the range - [-∞, + ∞]. The objectives they should study are:

txtytwth=(gx−px)/pw=(gy−py)/ph=log(gw/pw)=log(gh/ph)tx=(gx−px)/pwty=(gy−py)/phtw=log(gw/pw)th=log(gh/ph)

VI RESULTS

Figure 1: Screen Shows the Home Page of the Web Application

Figure 2: Screen Shows the Application Interface

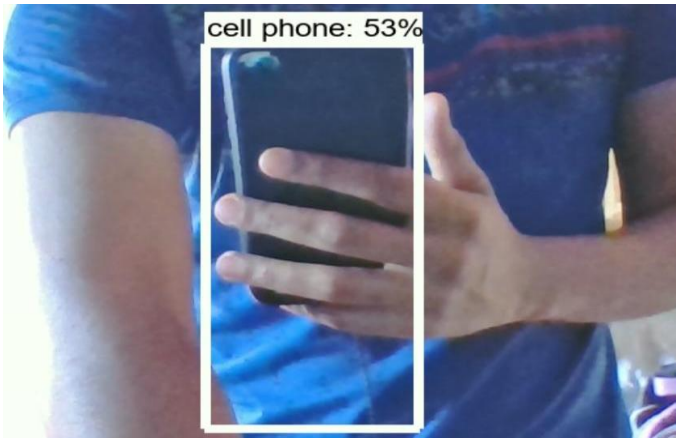


Figure 3: Object identified as cellphone giving Voice Feedback

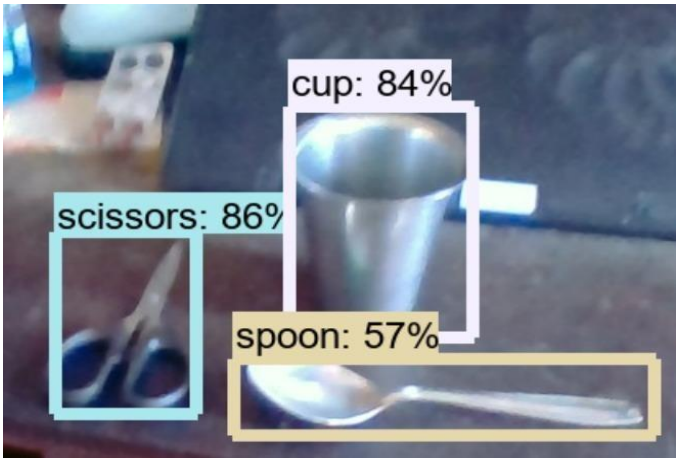


Figure 4: Object Identified and Respectively giving Voice feedback

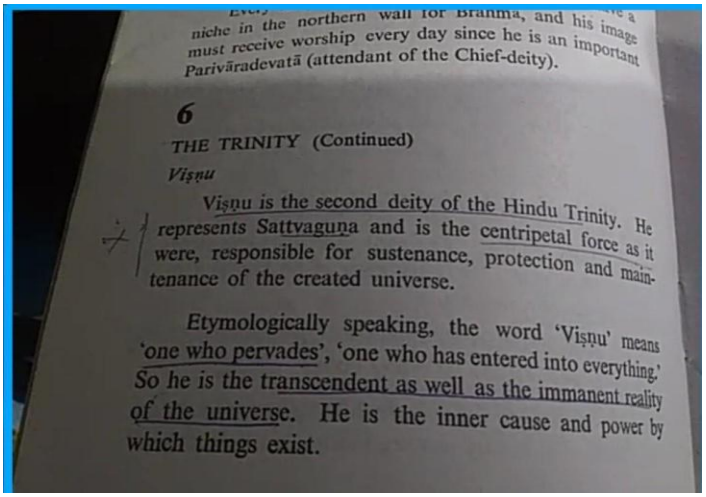


Figure 5: The Frame is Freezed and Continues to be Freezed Until all the Text in the Frame is Not Converted into text

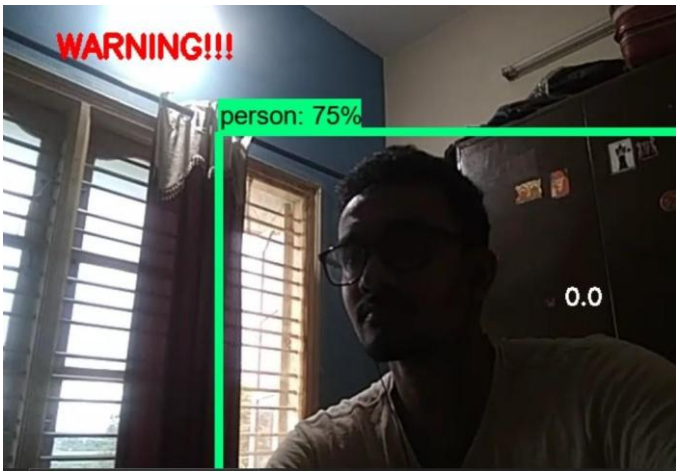


Figure 6: Person and Distance of the Person from the Camera is given as Feedback

VII CONCLUSION

This Project allows blind people or people who are having low vision to navigate around, without the help of others and avoid obstacles. This technology provides a good detection rates of identifying the objects that are in our surroundings. The limitation of this system is that it requires stability ,in order to improve the detection rate of the objects. The future Scope is to provide navigation support through the audio feedback system and natural language processing can be used to have a human like feeling for the blind This project provides experimental results on different methods for object detection and compares each method for their efficiencies.

VIII REFERENCES

[1] Object Detection with Deep Learning: A Review. Published on: 28 January 2019.By ad

[2] Zhong-Qiu Zhao; Peng Zheng; Shou-Tao Xu; Xindong Wu. Published in: IEEE Transactions on Neural Networks and Learning Systems (Volume: 30, Issue: 11, Nov. 2019)

[3] Real-time RGB-Based 3D Object Pose Detection using Convolutional Neural Networks. By Jin Liu; Sheng He; Yiting Tao; Daifei Liu. Published in: IEEE Sensors Journal (Volume: 20, Issue: 20, Oct.15, 15 2020)

[4] Developing Walking Assistants for Visually Impaired People: A Review.By Md. Milon Islam; Muhammad Sheikh Sadi; Kamal Z. Zamli; Md. Manjur Ahmed. Published in: IEEE Sensors Journal (Volume: 19, Issue: 8, April15, 15 2019)

[5] Convolutional neural network-based real-time object detection and tracking for a parrot AR-drone2. By Ali Rohan; Mohammed Rabah; Sung-Ho Kim. IEEE Access Date of Publication: 27 May 2019.

[6] Efficiently trainable text-to-speech system based on Deep convolutional networks with guided attention. By Hideyuki Tachibana; Katsuya Uenoyama; Shunsuke Aihara. Date Added to IEEE Xplore: 13 September 2018.

[7] Electronic Travel Aid System for Visually Impaired People. By P. S. Ranaweera; S. H. R. Madhuranga; H. F. A. S. Fonseka; D. M. L. D. Karunathilaka. Date Added to IEEE Xplore: 19 October 2017.

[8] Portable Camera-Based Assistive Text and Product Label Reading from Hand- Held Objects for Blind Persons. Published on: June 2014. [IEEE journal.]

[9] Textbooks: Hands on Machine Learning with Scikit Learn and Tensorflow. By Aurelien Geron.