Assistive Technology for Blind People for Object Detection and Description

Yasmin M A K
UG Student
Department of Electronics and Communication Engineering
Sri Sivasubramaniya Nadar College of Engineering
Chennai, India

Reshma G
UG Student
Department of Electronics and Communication Engineering
Sri Sivasubramaniya Nadar College of Engineering
Chennai, India

Saffar Subuhania T
UG Student
Department of Electronics and Communication Engineering
Sri Sivasubramaniya Nadar College of Engineering
Chennai, India

Edna Elizabeth N
Professor
Department of Electronics and Communication Engineering
Sri Sivasubramaniya Nadar College of Engineering
Chennai, India

***Abstract*—The development of deep learning has helped object detection to make rapid progress. In recent years, smart wearable technology has become the part of everyday life. Smart wearable technology includes watches, glasses and many other wearable items. There are plenty of smart applications built for us. There is a lack of technology to aid visually impaired. Visually impaired people have to rely largely on other senses such as hearing, touch and smell in order to understand their surroundings. It is really difficult for them to walk without knowing what lies ahead even with a stick. Giving blind people the great accessibility to their environment is the objective of the smart glass system. In this proposed project, a smart glass application system for visually impaired people based on deep learning was proposed. This system can provide voice automated output which describes the obstacles along with the distance of the object from the user using an ultrasonic sensor which will help them to navigate without hindrance. The system predicts the class of the object and text to speech conversion is performed by using python library. It gives them an opportunity to visualize things and also guides them to move freely without getting injured.**

***Keywords*: smart glass, machine learning, sensor, wearables**

I. Introduction

Machine Learning has gained attention since the introduction of high computing machines and the availability of huge amount of data also known as big data. Today, machine learning is used in many types of industries from medical image processing to autonomous car. Detecting objects in images has also become one of the important research areas and now computers are able to detect objects and also draw bounding boxes on it. This is also known as computer vision. Learning algorithms are implemented to detect objects and are used to aid visually impaired and blind persons. This paper explains how convolution neural network are trained on a dataset that can detect objects and narrate detected objects information to the visually impaired person via voice. In the voice guidance technology, the technique of synthesizing the name of the obstacle is done by using text-to-speech library. During the process of detection there are numerous obstacles that the blind person will encounter. The above described system will detect the object and provides a voice automated output. For instance, there is a chair in front of them. This obstacle will be detected and provided to the person to prevent them from injury.

II. RELATED WORKS

Taking different regions in an image and using CNN to classify the presence of an object has a problem because different objects have different spatial locations. Therefore algorithms like R-CNN, YOLO etc have been developed. YOLO out-performs other detection methods including R-CNN when generalizing natural images to other domains like artwork has been pointed out by Joseph Redmon et.al (2016). Pengchang Fang et.al (2018) proposed that Faster R-CNN is improved by adding more flexible context information fusion method to increase accuracy of small object detection but takes longer time for computation. It can be observed that Fuyan Lin et.al (2020) proposed that modifying YOLO network by adjusting the values of anchors will enhance the detection effect of small objects and YOLO v3 network has a high balance between detection speed and accuracy also more suitable for real time data. Rohit Agarwal et.al (2017) used SONAR sensor and Arduino for obstacle detection and alerting the user but there is no object recognition. Our proposed system detects objects and also recognizes the object. Detecting public signs using Intel Edison chip was proposed by Feng Lanet.al (2015). The detection is only restricted to public sign boards and not for general objects. Our system was specifically designed to help blind people and can detect the most common objects they come across everyday. Jyun- You L et.al (2019) have dealt at length about object detection using Epson BT-300 glasses, ultrasonic sensor and YOLO V3 based on tensorflow and keras where the average recognition rate is 96.3% and time for voice output is 3.83 seconds. Priyanka Malhotra et.al (2020) compared R-CNN, Fast R-CNN and YOLO and concluded that YOLO can be used for real-time object detection because of its speed but accuracy is compromised. Performance of Raspberry Pi Zero was evaluated by Diana Bezerra Correia Lima et.al (2019) which helped us to finalize Raspberry Pi 4B. Though Pi Zero is small and compact, it has only a single core processor whereas Pi 4B has four CPU cores. There are several components like Wifi, Ethernet capabilities have to be externally integrated for Pi Zero. Raspberry pi 4B is the best choice for our application. Usman masud et.al (2022) implemented object detection system to help the blind persons and mounted the components in the form of a stick which the blind person can take anywhere with them. They have used KNFB reader for text to speech conversion which is a third party app and it is not free. In our system we use python library to perform the conversion and the advantages are it is completely free and does not need an internet connection. Rather than a stick we are trying to implement this in glasses so that the angle of viewing the objects is optimum and is also very easy to wear.

III. PROPOSED METHODOLOGY

To achieve optimal performance, two different algorithms are used for training and the best one is finalized. MS COCO is large-scale object detection, segmentation dataset employed to provide the necessary information about the surrounding environment. A dataset of objects gathered from daily scenes is created for recognizing objects. Object detection is used to find objects in the real world from an image of the world such as bicycles, chairs, doors, or tables that are common in the scenes of the blind. The proposed method for the blind aims at expanding possibilities to people with vision loss to achieve their full potential. The main objective is to design and implement a real time object recognition system and mount them on spectacles. Our proposed project uses Raspberry Pi 4B, Camera, Ultrasonic sensor and headphones.

The proposed block diagram is shown in Fig 1. Our system measures the distance between the blind person and the object using an ultrasonic sensor. The obstacle that is in front of the blind person will be detected and the type of the object along with it's distance from the user will be converted to speech using a text-to-speech converter and the output is in audio format. When the user wants to know about the objects placed back, left or right to the user, then they have to turn in such a way that the camera points at the object. On doing so the object will be detected and the voice output will be generated. We have not used raspberry pi camera because of the cable length and it could be easily prone to damage so we went with USB camera which can be connected to the USB 2.0 or USB 3.0 ports of Raspberry pi. SSD Algorithm is employed here for object detection and after the video stream is processed and the bounding boxes are drawn, the output will then be converted to speech using text to speech converters. Python libraries can be used free of cost for this purpose. Pyttsx3 library proves useful. Raspberry pi has to be powered using an appropriate power source and immediately after supplying power to Raspberry pi these processes will initialize and the objects will be detected real-time. In case of continuous operation of pi it tends to heat up easily so we have mounted a fan in the Raspberry pi case to cool it down.

In Fig 2 the entire setup is given. The camera is mounted on the glasses with ultrasonic sensor and connected to the raspberry pi which is kept inside a pouch which can be placed on the blind person's shoulder or hip. The headphones are also connected to the pi. Raspberry pi is enclosed in a box with fan and placed along with the power bank inside the pouch as shown in Fig 3.

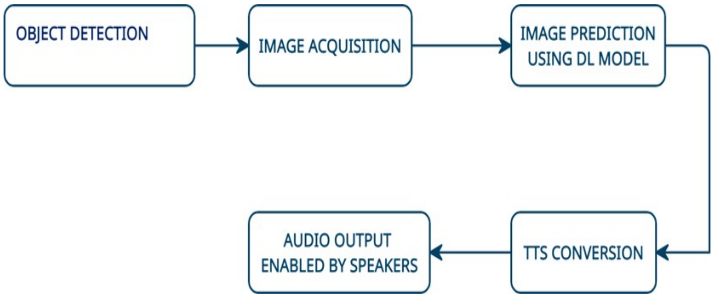

OBJECT DETECTION
IMAGE ACQUISITION
IMAGE PREDICTION USING DL MODEL
TTS CONVERSION
AUDIO OUTPUT ENABLED BY SPEAKERS

Fig. 1. Proposed block diagram of assistive navigation system

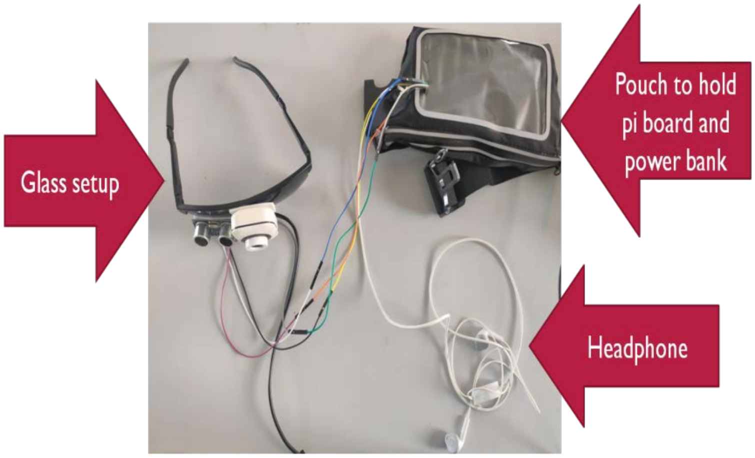


Fig. 2. The entire setup of our system

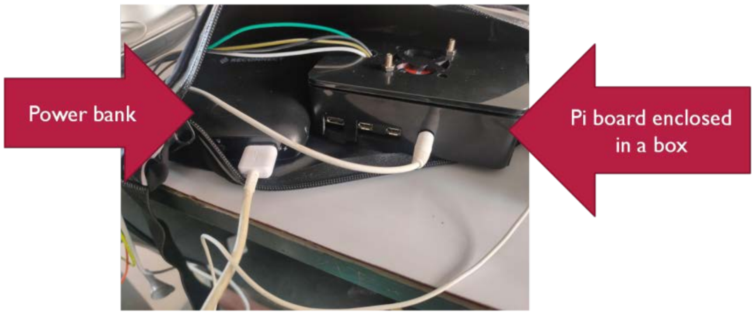


Fig. 3. Raspberry pi and power bank placed inside the pouch

IV. SOFTWARE IMPLEMENTATION

A. ALGORITHM

For detection based algorithms we need to draw a bounding box around the object to locate it. Also there might be many bounding boxes representing different objects of interest. We cannot proceed with this problem by building a standard CNN because the length of the fully connected layer is not constant. Taking different regions in an image and using a CNN to classify the presence of an object has a problem that different objects have different spatial locations. Therefore Algorithms like R-CNN, YOLO, SSD etc, have been developed. We have implemented YOLO and SSD and found that SSD was the better option.

1) YOLO: YOLO or You Only Look Once is an object detection algorithm much different from the region based algorithms. Training the model with YOLO has given good accuracy in correctly detecting the objects by drawing bounding boxes over it and assigning class probabilities for these boxes. YOLO is orders of magnitude faster (45 frames per second) than other object detection algorithms. YOLO algorithm is important because of the following reasons:

- Speed: This algorithm improves the speed of the detection because it can predict objects in real-time.
- High Accuracy: YOLO is a predictive technique that provides accurate results with minimal background errors.
- Learning capabilities: This algorithm has excellent learning capabilities that enable it to learn the representations of objects and apply them in object detection.

It works with the three techniques: Residual blocks, Bounding box regression, Intersection over union(IOU).

2) SSD: SSD stands for Single Shot multibox Detector. SSD takes only one shot to detect multiple objects present in an image using multibox. SSD is significantly faster in speed and also has high accuracy. SSD is a single stage object detection method that discretizes the output space of the bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. SSD runs a convolutional network on input image only one time and computes a feature map. So SSD is a better option as we can run it on a video. SSD's architecture builds on the venerable VGG-16 architecture, but discards the fully connected layers. Fig 4 shows the comparison between YOLO and SSD when implemented in both PC and Raspberry pi. When we tried YOLO in PC it was fast and accurate but when we implemented it in hardware YOLO took more time to process the output so we tried with SSD. SSD comparatively took less time. With Raspberry Pi, SSD gives higher FPS rate when compared to YOLO. YOLO in PC took approximately 0.2 seconds and SSD took 0.19 seconds. In PC they don't show much time difference but when implemented in pi YOLO took 3 seconds and SSD took 2.31 seconds which is less than YOLO so we implemented with SSD algorithm.

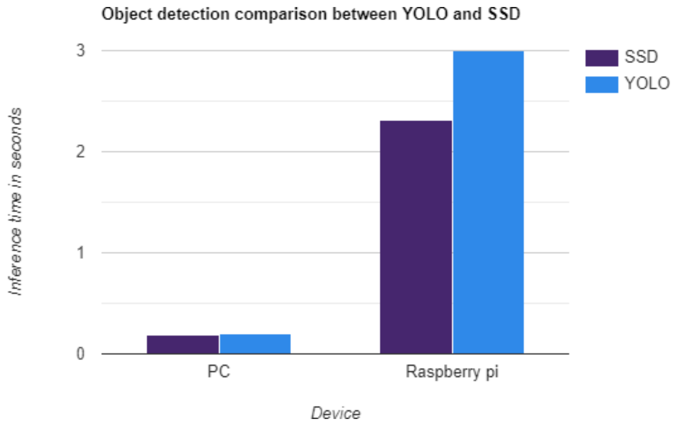


Fig. 4. Comparison of time taken for object detection for YOLO and SSD

B. DATASET

To train a deep learning model for object detection large number of photos are required. So, we made use of COCO dataset which is widely understood by state-of-the-art neural networks.

1) COCO DATASET: The COCO dataset is labeled, providing data to train supervised computer vision models that are able to identify the common objects in the dataset. Of course, these models are still far from perfect, so the COCO dataset provides a benchmark for evaluating the periodic improvement of these models through computer vision research. The COCO Dataset has 121,408 images. The COCO Dataset has 883,331 object annotations. The COCO Dataset has 80 classes. The COCO Dataset median image ratio is 640 x 480.

C. TEXT TO SPEECH CONVERTERS

IBM Watson API, Rev.ai API, Speechmatics API, Google Speech-to-text API, Robomatic.ai API, Amazon Polly API, Voicepods API, Microsoft Azure Cognitive Services API, Dialog Flow API, Ispeech API are the 10 API's we found to be the best and worth mentioning. All of these softwares are available online but with pricing. If free, only limited options are available. Either the time is restricted or words per day are fixed. This is definitely not going to help the blind people who will depend on this for a long period of time. We then decided to use python libraries for our project. Python provides many APIs to convert text to speech. One of such APIs is the Google Text-to-Speech API commonly known as the gTTS API. The other one is the pyttsx3.

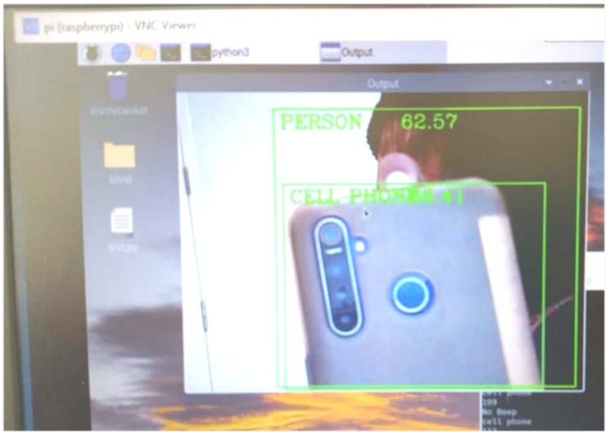
1) gTTS: gTTS is very easy to use. It is the tool which converts the text entered into audio which can be saved as a mp3 file. The gTTS API supports several languages including English, Hindi, Tamil, French, German and many more. The speech can be delivered in any one of the two available audio speeds, fast or slow. However, as of the latest update, it is not possible to change the voice of the generated audio. This works in any platform.

2) pyttsx3: It is a text to speech conversion library in Python, it looks for TTS engines pre-installed in your platform and uses them. These are the text-to-speech synthesizers that this library uses:

- SAPI5 on Windows XP, Windows Vista, 8, 8.1 and 10
- NSSpeechSynthesizer on Mac OS X 10.5 and 10.6
- espeak on Ubuntu Desktop Edition 8.10, 9.04 and 9.10

We implemented both the codes for all the objects listed in our dataset and both libraries worked fine. While the audio output from gTTS sounded more clear and humanlike and the one from pyttsx3 was somewhat robotic. Both were able to produce audio output for almost all the words and did well even for sentences. The major difference between them is that gTTS can only be used online and needs an internet connection to process the output which is not the case in pyttsx3 where the output is available even without internet connection. Internet connection cannot be relied on all times and it will affect the blind person so it is a better choice to go with pyttsx3.

V. RESULTS



```
109
No Beep
cell phone
```

Fig. 5. Cellphone and Person are detected with accuracies 66 percentage and 64 percentage

In Fig.5, the objects Cell phone and the Person holding it are identified. Though the person is not completely visible the system was able to detect them with an accuracy of 62 percentage. In Fig.6, the object Car was accurately detected with a bounding box around it and the detection accuracy was 65 percentage. The less accuracy may be due to the reason that the car was not fully shown and the wheels are not clearly visible. In Fig.7, the object shown is a Book and our system clearly identifies it as a Book. The cover of the book has many other elements like pen but our system is not misled and it detects the book with an accuracy 54 percentage. In Fig.8, Remote is detected by our system which is trained using SSD algorithm. Though Remote class is under represented in the COCO dataset our system detects it precisely.

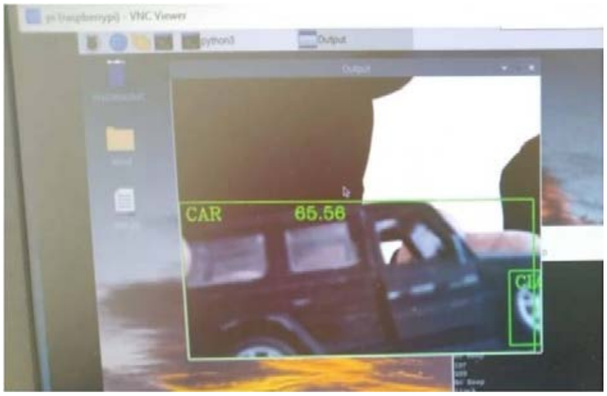
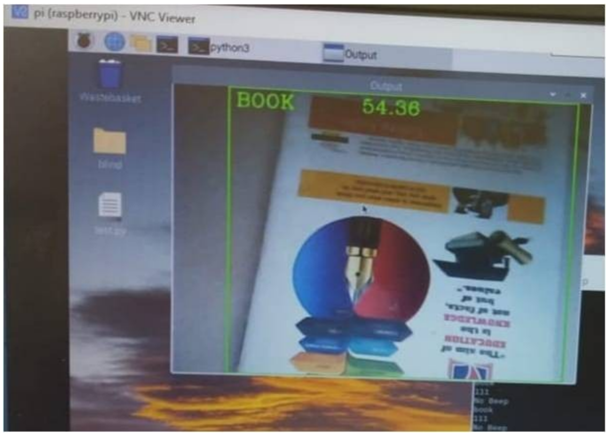


Fig. 6. Car is accurately detected with accuracy 66 percentage



```
111
No Beep
book
111
No Beep
```

Fig. 7. Book is detected with an accuracy of 54 percentage

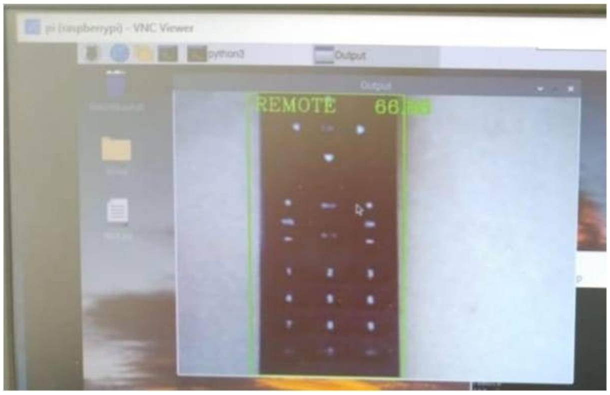


Fig. 8. Remote is detected using SSD algorithm with an accuracy of 67 percentage

In Fig. 9 both Laptop and Keyboard are detected. As we can see from the Fig.9, the keyboard is not fully covered in the frame but our system was able to identify it as keyboard and draws bounding box around it with an accuracy 55 percentage.

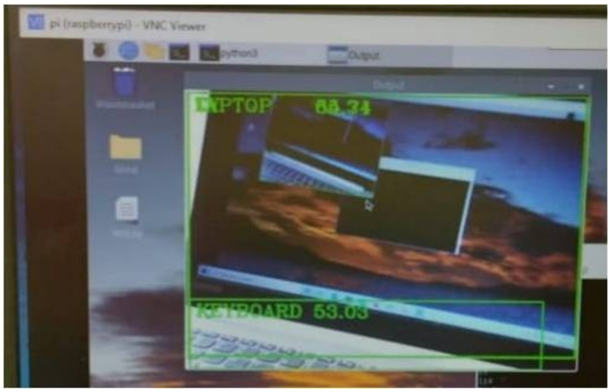


Fig. 9. Laptop and Keyboard are detected with accuracies 68 percentage and 55 percentage

VI. CONCLUSION

Our proposed system will be very beneficial for the blind people and people with vision impairment. They can live an independent life without having to depend on others to navigate. It gives them an opportunity to be able to detect obstacles and also guides them without getting injured. Because all the data is saved and processed on the Raspberry pi there is no need for an internet connection and the Python library used for text-to-speech conversion is also available offline. This is an additional benefit because internet connection cannot be relied on all times.

To extend this work, GSM module can be integrated to the system so that the location of the blind person can be continuously accessed by their relatives and close friends. In case of emergencies this feature can be very much useful for the blind person.

REFERENCES

[1] Diana Bezerra Correia Lima, Orlando Baiocchi, Cleonilson de Souza.”A Performance evaluation of Raspberry Pi Zero W Based Gateway running mqtt broker for IoT”.2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).

[2] Feng Lan, Guangtao Zhai , Wei Lin.” Lightweight Smart Glass System with Audio Aid for Visually Impaired People”.2015 TENCON 2015 - 2015 IEEE Region 10 Conference, 2015.

[3] Fuyan Lin,Xin Zheng,Qiang Wu.” Small object detection in aerial view based on improved YoloV3 neural network”. 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications(AEECA).

[4] Hasan U. Zaman, Saif Mahmood, Sadat Hossain, Iftekharul Islam Shovon.” Python Based Portable Virtual Text Reader”. 2018 Fourth International Conference on Advances in Computing, Communication Automation (ICACCA), 2018.

[5] Joseph Redmon,Santhosh Divvala, Ross Girshick, Ali farhadi.” You Only Look Once: Unified, Real-Time Object Detection”. In 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016.

[6] Jyun-You L,in ,Chi-Lin Chiang,Meng-Jin Wu,Chih-Chiung Yao, Ming-Chiao Chen.” Smart Glasses Application System for Visually Impaired People Based on Deep Learning”2020 Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN), 2020.

[7] Pengcheng Fang,Yijie Shi.” Small Object Detection Using Context Information Fusion in Faster R-CNN”. 2018 IEEE 4th International Conference on Computer and Communications (ICCC).

[8] Priyanka Malhotra, Ekansh Garg”Object Detection Techniques: A Comparison”. IEEE 7th International Conference on Smart Structures and Systems ICSSS 2020.

[9] Rohit Agarwal, Nikhil Ladha , Mohit Agarwal, Kuntal Kr. Majee, Abhijit Das,Subham Kumar, Subham Kr. Ra,,Anand Kr. Singh.” Low cost ultrasonic smart glasses for blind” 2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2017.

[10] Usman Masud, Tareq Saeed, Hunida M.Malaikah, Fezan Ul Islam and Ghulam Abbas. ”Smart Assistive System for Visually Impaired People Obstruction Avoidance Through Object Detection and Classification”.IEEE Access,vol. 10, pp. 13428 - 13441, 2022.

[11] Wei Wei” Small Object Detection Based on Deep Learning”.International Journal of Power, Intelligent Computing and Systems (ICPICS)19995791, 28-30 July 2020, pp.938-943.