Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

Assistive Object Recognition System for Visually Impaired

Shifa Shaikh
Electronics and Tele-communication
Vivekanand Education Society Institute of Technology
Mumbai, India

Vrushali Karale
Electronics and Tele-communication
Vivekanand Education Society Institute of Technology
Mumbai, India

Gaurav Tawde
Electronics and Tele-communication
Vivekanand Education Society Institute of Technology
Mumbai, India

***Abstract*— The issue of visual impairment or blindness is faced worldwide. According to statistics of the World Health Organization (WHO), globally, at least 2.2 billion people have a vision impairment or blindness, of whom at least 1 billion are blind. In terms of regional differences, the prevalence of vision impairment in low- and middle-income regions is four times higher than in high-income regions.[6] Blind people generally have to rely on white canes, guide dogs, screen-reading software, magnifiers, and glasses to assist them for mobility, however, To help the blind people the visual world has to be transformed into the audio world with the potential to inform them about objects as well as their spatial locations. Therefore, we propose to aid the visually impaired by introducing a system that is most feasible, compact, and cost-effective. So, we implied a system that makes use of Raspberry Pi in which you only look once (YOLO v3) machine learning algorithm trained on the coco database is applied. The experimental result shows YOLO v3 achieves state-of-the-art results of 85% to 95% on overall performance, 100% (person, chair, clock, and cell-phone) recognition accuracy. This system not only provides mobility to the visually impaired with that it provides the term that ahead is an XYZ object rather than a sense of obstacle.**

***Keywords*— Visual Impairment, Raspberry Pi, YOLO v3 Algorithm, Computer Vision, Object Recognition, voice output.**

I. INTRODUCTION

"ONLY BECAUSE ONE LACKS THE USE OF THEIR EYES DOES NOT MEAN THAT ONE LACKS VISION."

Eyesight is one of the essential human senses, and it plays a significant role in human perception about the surrounding environment. For visually impaired people to be able to provide, experience their vision, imagination mobility is necessary. The International Classification of Diseases 11 (2018) classifies vision impairment into two groups, distance and near presenting vision impairment.[6] Globally, the leading causes of vision impairment are uncorrected refractive errors, cataract, age-related macular degeneration, glaucoma, diabetic retinopathy, corneal opacity, trachoma, and eye injuries. It limits visually impaired ability to navigate, perform everyday tasks, and affect their quality of life and ability to interact with the surrounding world upon unaided. With the advancement in technologies, diverse solutions have been introduced such, as the Eye- ring project, the text recognition system, the hand gesture, and face recognition system, etc. However, these solutions have disadvantages such as heavyweight, expensive, less robustness, low acceptance, etc. [2] hence, advanced techniques must evolve to help them. So, we propose a system built on the breakthrough of image processing and machine learning.

The proposed system captures real-time images, then images are pre-processed, their background and foreground are separated and then the DNN module with a pre-trained YOLO model is applied resulting in feature extraction. The extracted features are matched with known object features to identify the objects. Once the object is successfully recognized, the object name is stated as voice output with the help of text-to-speech conversion.

The key contributions of the paper include:

- Robust and efficient object detection and recognition for visually impaired people to independently access familiar and unfamiliar environments and avoid dangers.
- Offline text-to-speech conversion and speech output.

II. RELATED WORK

1) Real-Time Objects Recognition Approach for Assisting Blind People:

In this paper, two cameras placed on blind person's glasses, GPS free service, and ultrasonic sensors are employed to provide information about the surrounding environment. Object detection is used to find objects in the real world such as faces, bicycles, chairs, doors, or tables that are common in the scenes of a blind. Here, GPS service is used to create groups of objects based on their locations, and the sensor detects an obstacle at a medium to long distance. The descriptor of the Speeded-Up Robust Features (SURF) method is optimized to perform the recognition. The use of two cameras on glasses can be sophisticated. [2]

2) Wearable Object Detection System for the Blind:

In this paper, the RFID device is designed as a support for the blind for the disclosure of objects; especially, it is developed for searching the medicines in a cabinet at home. This device can provide information about the distance of a defined object, how near or far it is and simplifies the search. For identifying the medicines, the device can provide the user with an acoustic signal to find the desired product as soon as possible. The

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

measure of the distance, in particular the movement of the antenna respect to tag, is made using the RSSI value. This application uses the RSSI (Received Strength Signal Indicator) value, measuring the power of the received signal of the tag. [3]

3) Smart Obstacle Detector for Blind Person:

Another system proposed in this paper focuses on giving information about what are the different types of obstacles in front of the user, their size, and their distance from the user. MATLAB Software is used for signal processing. The camcorder is used for recording videos. Video processing methods are used after that. The output of this system not only gives output in audio format but also vibration. A vibrating motor has been connected with an ultrasonic sensor. The ultrasonic sensor detects objects coming in its range and this makes the vibrating motor vibrate. Use of Camcorder, a stick with an Ultrasonic sensor makes this system bulky and dependent on the stick. [4]

III. BLOCK DIAGRAM

The figure given below is the block diagram of our system consisting of: Camera, Raspberry Pi, Speaker, and Power bank.

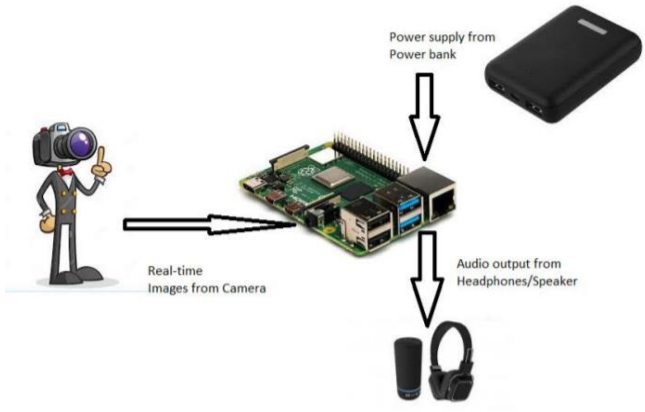

Fig 1: Block Diagram of User Side

Our system mechanism starts with an Image Acquisition need, which is done by the USB camera that is attached to the USB port of the Raspberry Pi (Rpi). Inside Rpi, we install a YOLO algorithm. A speaker is attached to one of the usb ports of Rpi as an output voice device. As we demand mobility, we are using the 5volts power bank as a power supply.

IV. ABOUT SYSTEM

Raspberry pi: The heart of our project is Raspberry pi, as we are going to have the result in the audio form we decided to use a speaker, also Raspberry pi supports high bass headphones. We are using the Raspberry pi (3 B+) design. To provide mobility to users, we decided to use a power bank as a power supply source to Raspberry pi. The reason for using Raspberry pi is one of the most popular single-board computers. All the major image processing algorithms and operations can be implemented easily with OpenCV on Raspberry Pi. We are using a 32 GB class 10 SD card for our Raspberry pi. Also, instead of using a Raspberry pi camera, we are using a USB camera as the cable of Raspberry pi camera is stiff and difficult to maintain.

Fig 2: Raspberry Pi model

YOLO: is an extremely fast, real-time, multi-object detection algorithm, and it satisfies the basic requirement of our system. YOLO applies a single convolutional neural network to an entire image and divides the image into an S x S grid and comes up with bounding boxes, which are drawn around images and predicts probabilities for each of these regions for object recognition, object localization, and object detection. [10]

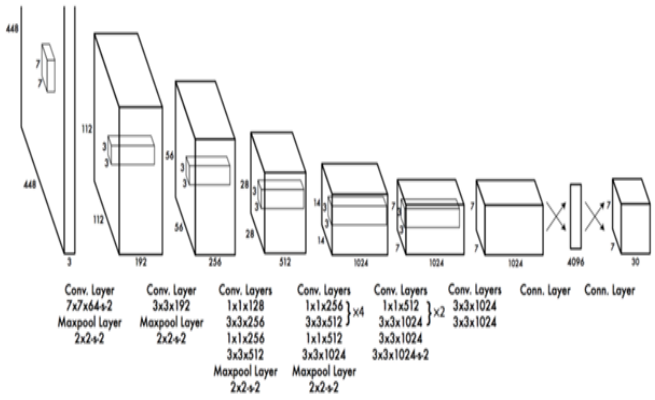

Fig 3: Network architecture of YOLO

YOLO predicts multiple bounding boxes per grid cell. For this, we select the highest IoU (intersection over union) with the ground truth. This strategy leads to specialization with the bounding box predictions. Each prediction gets better at predicting certain sizes and aspect ratios. YOLO uses a sum squared error between the predictions and the ground truth to calculate the loss. The loss function comprises:

A. Localization loss: measures errors between the predicted boundary box and the ground truth.

B. Classification loss: is the squared error of the class conditional probabilities for each class, when an object is detected.

C. Confidence loss: detects the objects present in the box.

The final loss adds localization, confidence, and classification losses. [11]

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

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

Fig 4: Loss function expression

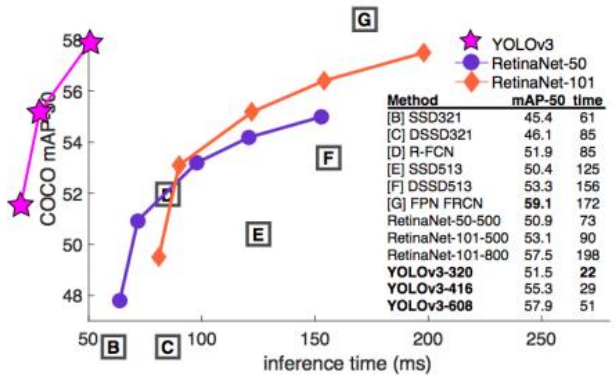

Fig 5: Comparison of YOLO with other algorithms.

Opencv (Open source computer vision): is a library of programming functions mainly aimed at real-time computer vision. The library has more than 2500 optimized algorithms. [12] These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, etc.

Dnn module (Deep Neural Network): dnn is the module in OpenCV, which is responsible for all deep learning related concepts.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig 6: DARKNET-53 Backbone Architecture

In our system, images are pre-processed, background and foreground are separated, and then the DNN module is applied, resulting in feature extraction. Darknet-53 is used as the feature extractor, as it achieves the classification accuracy 2x faster. And the extracted features are matched with known object features to recognize them. A DNN-based algorithm is more robust and accurate on a wide range of faces. DNN module only allows forward propagation on the pre-trained model.[13] In DNN module of OpenCV, it requires your input transform to a blob, or tensor in other neural network frameworks.

Pyttsx3 lib: For audio output, we use the "pyttsx3" library. The main advantage of using this library is that it does not require any internet, and it converts text-to-speech very fast and effectively.

V. WORKING & IMPLEMENTATION

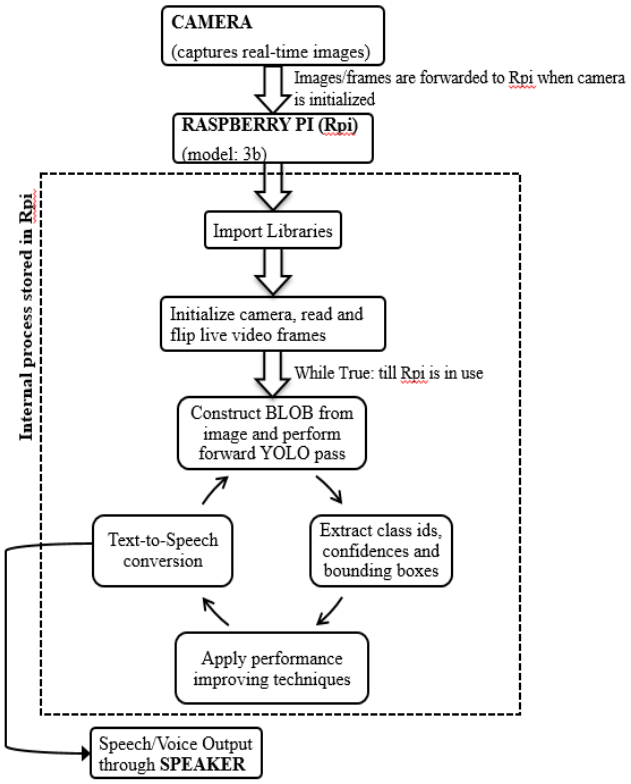

Fig 7: Flowchart of system

The Flowchart is a communication outline that shows how objects work with each other and in what order. The above Flowchart of our framework clarifies the stream that is first, the user starts and wears the system. Once the Raspberry Pi (Rpi) is on, it will implement its internal process/code. The code keeps on executing till the Rpi is on. Initially, Rpi will import all the libraries that are: OpenCV, Pyttsx3, Time, and NumPy and will read the text file containing class names, YOLO weights, and configuration files. After that, the code will initialize the camera connected to it. The camera will capture real-time frames at 1fps (frame per second), then the code will read the input image/frame and get its width and height to an adequate level. Then an object detection algorithm in our case YOLO is applied to this altered frame. Before forward passing this altered image to YOLO weights and YOLO configuration files, a '**BLOB from image**' is constructed. To obtain (correct) predictions from deep neural networks such as YOLO, you first need to pre-process your data. In the context of deep learning, feature extraction, and

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

image classification, we have used the OpenCV function blobFromImage. This function performs the following:
1. Mean subtraction - is used to help combat illumination changes in the input images in our dataset.
2. Scaling by some factor - is used to scale the input image space into a particular range
3. Optionally channel swapping. [8]

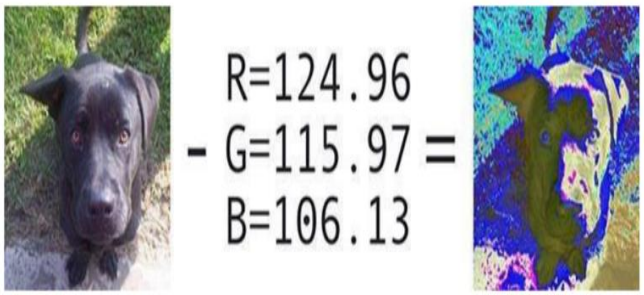

Fig 8: A visual representation of mean subtraction where the RGB mean (centre) has been calculated from a dataset of images and subtracted from the original image (left) resulting in the output image (right).

Then the code performs a forward pass of the YOLO object detector, giving us our bounding boxes, class ids, and associated class probabilities.
Another advantage of YOLO other than being fast is that it provides three methods to improve its performance:

- **Intersection over Union (IoU)** decides which predicted box is giving a good outcome. It calculates the IoU of the actual bounding box and the predicted bounding box.
- **Non-max suppression** suppresses weak, overlapping bounding boxes.
- **Anchor Boxes** detects multiple objects in a single grid.[7]

Further, the frames are divided into a 3x3 grid, which helps in finding the position of objects. Our system aims to produce an audio output for the visually impaired. The Detected object labels are converted into speech using the pyttsx3 library.
Lastly, Upon successful recognition of an object and as per grids, the system will provide speech output stating the name of the object along with its grid name, for e.x. ‘Mid left car’, ‘Mid right car’. Hence helping the visually impaired people in recognizing the objects in the field of view.

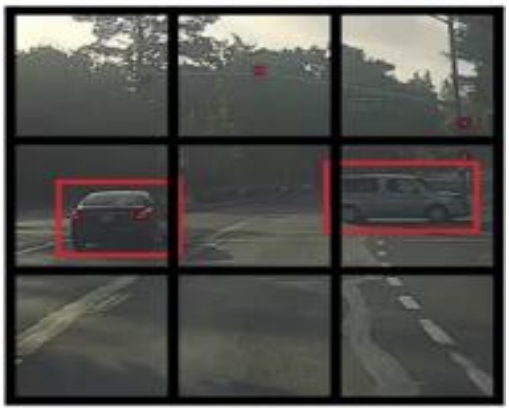

Fig 9: Division of Image into grids

VI. RESULTS AND DISCUSSION

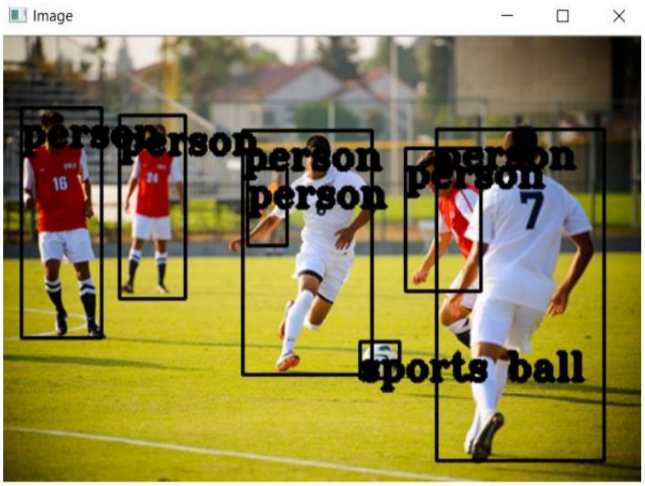

Fig 10: Image with bounding boxes and class label

Fig 11: Real-time Object Detection with Multiple Bounding Boxes 1

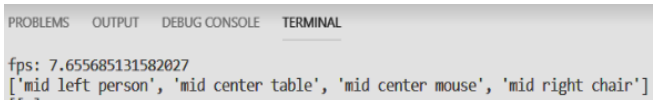

Fig 12: Real time YOLO Object detection with text output 1

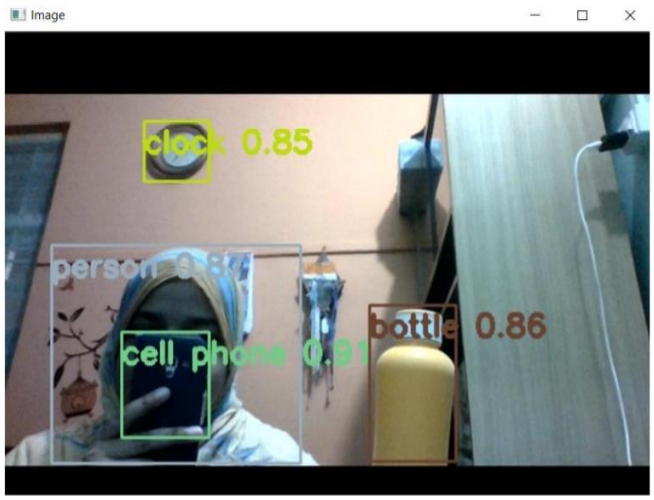

Fig 13: Real-time Object Detection with Multiple Bounding Boxes 2

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

```
fps: 7.733104841379943
['bottom left cell phone', 'mid left person', 'bottom center bottle', 'top left clock']
```

Fig 14: Real time YOLO Object detection with text output 2

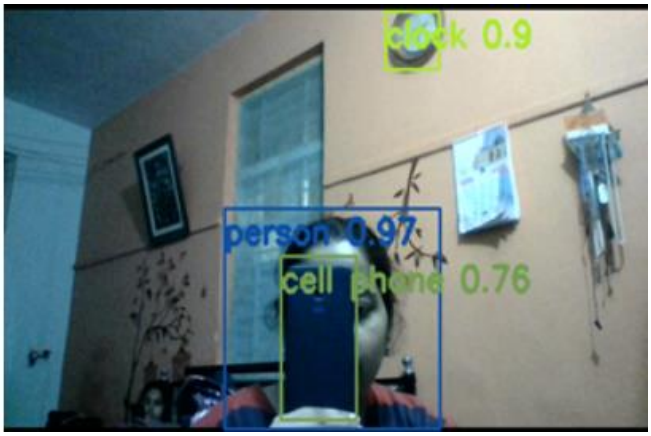

Fig 15: Real-time Object Detection with Multiple Bounding Boxes 3

```
fps: 8.380539030786958
['bottom center person', 'top center clock', 'bottom center cell phone']
```

Fig 16: Real time YOLO Object detection with text output 3

Fig 10 illustrates object detection and recognition of the already acquired image. The system has successfully recognized every object present in the image based on the trained coco dataset. Similarly, Fig 11, Fig 13, and Fig 15 illustrate real-time object detection and recognition, along with their confidences of class recognition. Also, Fig 12, Fig 14, and Fig 16 illustrate the text form, which is converted into speech. We have successfully achieved a speed of 7 fps to 9 fps in this CPU based system. The speed of detection and recognition can be increased with the use of a GPU based system.

VII. FUTURE SCOPE

The future perspective of this project is to increase the object recognition rate which can be achieved by using the TensorFlow library and to provide an exact distance measurement between the people and object. However, for developing an application that involves many objects that are fast-moving, you should instead consider faster hardware. Further, we can implement face recognition and text recognition in the same system. Thus, making the system compatible overall.

VIII. CONCLUSION

In recent years, some solutions have been devised to help blind or visually impaired in recognizing objects in their environment but they are not efficient. Our purpose is to provide a robust and comfortable system for the blind to recognize their surrounding objects. Our advanced system uses a USB camera to seize real-time images in front of the users. The machine learning and feature extraction technique used here is YOLO. The YOLO framework trades with object detection by choosing the entire image in a single instance, and splits the image into grids, then predicts the bounding box coordinates and class probabilities for these boxes. The biggest advantage of sing YOLO is its excellent speed – it's incredibly fast and YOLO also understands generalized object representation. This system will make visually impaired virtually visible also it innovatively uses the text-to-speech technology which provides audio descriptions of their surroundings and helps them to travel with self-confidence. The proposed system is mobile, robust, and efficient. Also, it creates a virtual environment and this system provides a sense of assurance as it voices the name of the object recognized.

REFERENCES

Book,

[1] Peter Harrington, "Machine Learning in Action, pp."by Manning Publications-1st edition.

Papers,

[2] Zraqou, Jamal & Alkhadour, Wissam & Siam, Mohammad. (2017). "Real-Time Objects Recognition Approach for Assisting Blind People." Multimedia Systems Department, Electrical Engineering Department, Isra University, Amman-Jordan Accepted 30 Jan 2017, Available online 31 Jan 2017, Vol.7, No.1

[3] A. Dionisi, E. Sardini and M. Serpelloni, "Wearable object detection system for the blind," 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings, Graz, 2012, pp. 1255-1258, doi: 10.1109/I2MTC.2012.6229180.

[4] Daniyal, Daniyal & Ahmed, Faheem & Ahmed, Habib & Shaikh, Engr & Shamshad, Aamir. (2014). "Smart Obstacle Detector for Blind Person." Journal of Biomedical Engineering and Medical Imaging. 1. 31-40. 10.14738/jbemi.13.245.

[5] Christian Szegedy Alexander Toshev Dumitru Erhan, "Deep Neural Networks for Object Detection."

[6] N.Saranya, M.Nandinipriya, U.Priya,"Real Time Object Detection for Blind People",Bannari Amman Institute of Technology, Sathyamangalam, Erode.(India).

[7] Rui (Forest) Jiang,Qian Lin,Shuhui Qu,"Let Blind People See: Real-Time Visual Recognition with Results Converted to 3D Audio",Stanford,2018.

Website,

[8] https://www.who.int/news-room/detail/08-10-2019-who-launches-first-world-report-on-vision

[9] https://www.analyticsvidhya.com/blog/2018/12/practical-guide-object-detection-yolo-framewor-python/

[10] http://datahacker.rs/face-detection-opencv-images/

[11] https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b

[12] https://medium.com/@jonathan_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088

[13] https://opencv.org/about/

[14] https://github.com/TheNsBhasin/DNN_Object_Detection