

# INDOOR AND OUTDOOR NAVIGATION ASSISTANCE SYSTEM FOR VISUALLY IMPAIRED PEOPLE USING YOLO TECHNOLOGY

AKILA S, MONAL S, PRIYADHARSHINI A, SHYAMA M, SARANYA M, ASSISTANT PROFESSOR,  
DEPARTMENT OF BIOMEDICAL ENGINEERING

MAHENDRA INSTITUTE OF TECHNOLOGY, NAMAKKAL

\*\*\*

**Abstract**— Good vision is a priceless gift, but vision loss is becoming more common these days. To assist blind individuals, the visual world must be turned into an aural world capable of informing them about objects and their spatial placements. Objects recognized in the scene are given names and then transformed to speech. Everyone deserves to live freely, even those who are impaired. In recent decades, technology has focused on empowering disabled people to have as much control over their lives as possible. In this paper, an assistive system for the blind is proposed, which uses YOLO for fast detecting objects within images based on deep neural networks for reliable detection, and Open CV under Python to let him know what is around him. The acquired results suggested that the proposed model was successful in allowing blind users to move around in an unknown indoor/outdoor environment using a user interface. Artificial intelligence was used in this project to recognize and evaluate items and convey information via speaker. In the suggested work, we employ darknet YOLO approaches to tackle the present system problem by reducing the recognize time of multiple objects in less time with the best time complexity.

**Keywords**— Assist Blind, Open CV, Python, Deep Neural Networks

## INTRODUCTION

Millions of people throughout the world struggle to comprehend their surroundings related to visual impairment. Despite their ability to discover alternative techniques to dealing with daily tasks, they have navigational challenges and also social difficulty. For example, finding a certain room in an unknown setting is quite difficult for them. Furthermore, it is difficult for blind and visually impaired people to tell if someone is speaking to them or to someone else during a conversation. The purpose of this research is to investigate the feasibility of employing the sense of hearing to comprehend visual objects. The visual and auditory senses are strikingly similar in that both visible objects and audio sounds could be spatially localized. Many individuals are unaware that we are capable of determining its spatial location upon a sound source simply through hearing with two ears. The goal of the project is to assist blind persons in navigating via output of a processor. Object Extraction, Feature Extraction, and Object Comparison are all part of the

approach used in this project. Artificial intelligence was used in this project to recognize and evaluate items and convey information via speaker. In the suggested work, we employ darknet YOLO approaches to tackle the present system problem by reducing the recognize time of multiple objects in less time with the best time complexity. We were inspired to create this project by the need for navigation assistance between blind people as well as a broader look in at advanced technology that is becoming available in today's environment. Technology is something that exists to make human tasks easier. As a result, we use solutions to address the difficulties of visually impaired persons in this initiative. The project's goal is to aid users in navigation through the use of technology, and our engineering profession encourages us to do so.

## RELATED WORKS

Many attempts at object identification and recognition have been performed with deep learning algorithms like CNN, RCNN, YOLO, and others. A literature review is undertaken in this study to better understand some of these algorithms. [1]

Using YOLOv3 and the Berkley Deep Drive dataset, Aleksa Corovi et al. (2018) developed a method for detecting traffic participants. This system can recognize five different object classes (truck, car, traffic signs, pedestrians, and lights) in various driving circumstances (snow, overcast and bright sky, fog, and night). The accuracy rate was 63%. [2]

Smart glasses were applied by Rohini Bharti, et al. (2019) to assist visually challenged persons. This system was built using CNN with Tensorflow, a custom Dataset, OpenCV, and a Raspberry Pi. This system can detect sixteen different classes. The technology has a 90 percent accuracy rate. [3]

Omkar Masurekar, et al. (2020) developed an object detection model to aid visually impaired individuals. We utilized YOLOv3 and a custom dataset with three classes (bottle, bus, and mobile). For sound generation, Google Text To Speech (gtts) is employed. The authors discovered that recognizing the objects in each frame took eight seconds and that the accuracy reached was 98 percent. [4]

Sunit Vaidya, et al. (2020) implemented a web application and an android application for object detection. YOLOv3 and coco dataset were used in these systems. The authors found that the maximum accuracy in web applications is 89 % and 85.5% in mobile phones. The required time for detecting the objects was two seconds and this time increased by increasing the number of objects. [5]

Many scholars have developed algorithms in recent years. In this application of computer vision, both machine learning and deep learning algorithms work. Since 2012, this section traces the evolution of the various methodologies utilized by the researchers in their research. The SVM technique is used by Histograms of Gradient Descent (HOG) to detect objects in real time.

You Just Look Once (YOLO) is an algorithm that predicts what objects are present and where they are by processing the image only once. A single convolutional network is used to predict multiple bounding boxes and class probabilities for those boxes [7]. The image is divided into grids to do this. The object's detection is handled by the grid cell at the object's center. When opposed to approaches that require object proposals, SSD is more easier. It gets rid of the proposal creation and following pixel or feature resampling phases. The output space of bounding boxes is discretized into a set of default boxes per feature map location using this method, which works with varied aspect ratios and sizes. All computing is encapsulated in a single network [8, 9].

Chen et al. [10] developed a system that includes glasses, the long stick cane, and a mobile application to aid in obstacle detection. Glasses and a long stick cane are used to detect the thing. The information is uploaded to the web platform and shown on the mobile device whenever the user falls.

**EXISTING METHOD**

Blind persons are currently employing an ultrasonic sensor in this system. If any object is detected by ultrasonic in front of blind persons, the buzzer will be activated. Object detection using OpenCV has been developed. However, it was unable to locate the object precisely. Mat lab software was used to recognize objects. But it's the result of a simulation.

**DISADVANTAGES**

- It's a simulation software. So we could not implement in real-time
- Output accuracy low

**PROPOSED METHOD**

We will use AI, OPENCV, and YOLO in this suggested system to detect objects in real time (You only look once). After the image is captured, it must be pre-processed as well as compressed. The model is trained using photos of many everyday things. It is learned by extracting the desired pattern from the image using feature extraction. The image is then compressed using feature fusion and dimension reduction for reliable and real-time performance. The classifier is then trained using the YOLO dataset. We select the best classifier by comparing the performance of several classifiers, and so the object recognition model is created. Now you can give this model any test image, and it will be categorized into one of the classes it has been trained in.

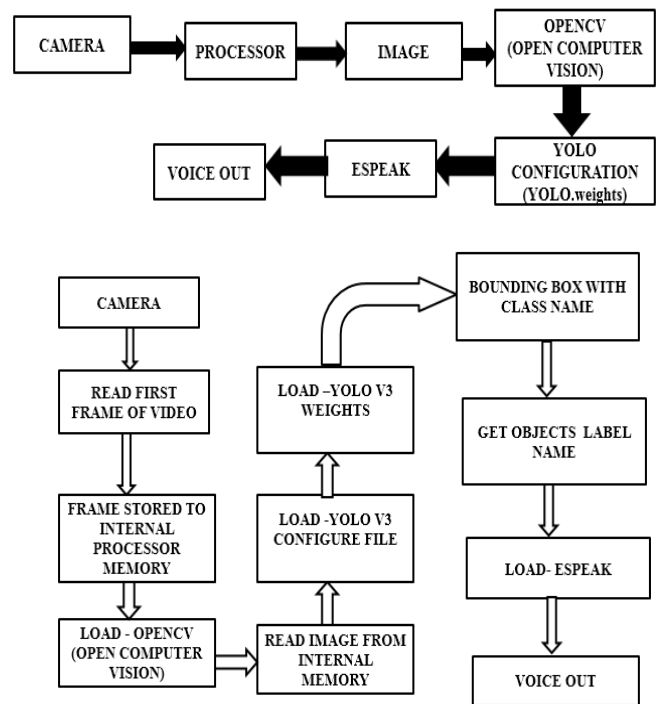


Fig.1. proposed block diagram

**MATERIAL AND METHODS**

**WEB CAM**

A webcam is a video camera that transmits or streams its image to or through a computer to computer network in real time. The video stream can be preserved, viewed, or forwarded to other networks via systems like the internet or email as an attachment once it has been "caught" by the computer. The video stream can be saved, viewed, or on sent when it is sent to a remote place. A webcam, unlike an IP camera (which connects over Ethernet or Wi-Fi), is typically attached via a USB connection or other similar

cable, or is incorporated into computer hardware, such as laptops. The phrase "webcam" (a clipped compound) can also refer to a video camera that is connected to the Internet indefinitely rather than for a specific session, and that provides a view for everyone who enters its web page over the Internet.

### PROCESSOR

A processor is a chip or logical circuit that responds to and interprets basic instructions in order to control a computer. The processor's primary functions include fetching, decoding, executing, and writing back instructions. The processor is also known as the brain of any system, including computers, laptops, smartphones, embedded systems, and so on. The two elements of the processors are the ALU (Arithmetic Logic Unit) and the CU (Control Unit). The Arithmetic Logic Unit executes all mathematical operations such as additions, multiplications, subtractions, divisions, and so on, while the control unit regulates the command or operation of the instructions, much like a traffic cop. Other components, such as input/output devices and memory/storage devices, communicate with the processor.

### OPENCV

OpenCV is indeed a cross-platform library that allows us to create real-time computer vision apps. It focuses primarily on image processing, video recording, and analysis, with capabilities such as face detection as well as object detection. We'll show you how to use OpenCV in your applications in this tutorial. One of the most widely used computer vision libraries is OpenCV. A good understanding of the fundamentals of OpenCV is essential to begin your path in the field of computer vision. In this essay, I'll try to explain the most fundamental and significant ideas of OpenCV in an easy-to-understand manner.

### YOLO CONFIGURATION

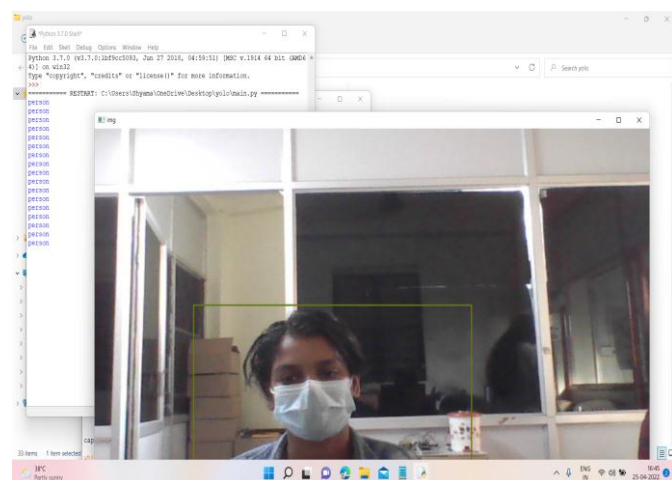
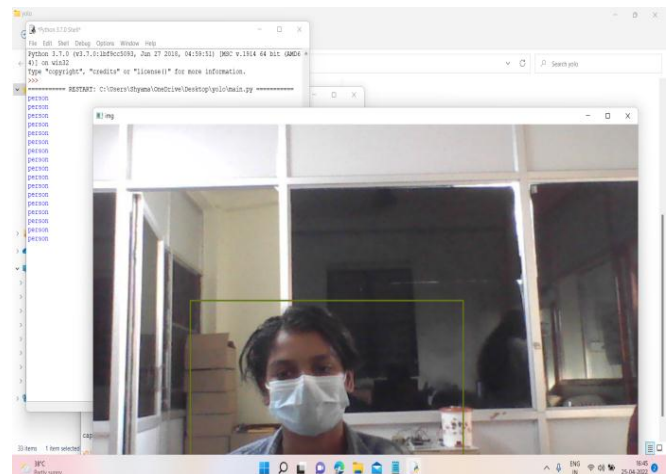
YOLO, a single CNN estimates multiple bounding boxes with class probabilities on these boxes all at the same time. YOLO improves detection performance by training on entire photos. Compared to other object detection approaches, the above system has a number of advantages: YOLO is an acronym for "you only look once." During training and testing, YOLO sees the complete image, thus it implicitly encodes contextual information about classes and also their appearance.

The YOLO (You Only Look Once) real-time object identification method is one of the most effective object recognition algorithms available, and it incorporates many of the most cutting-edge ideas from the computer vision research community. The capacity to detect objects is a crucial feature of autonomous vehicle technology. It's a field of computer vision that's blossoming and performing far better than it did only a few years ago.

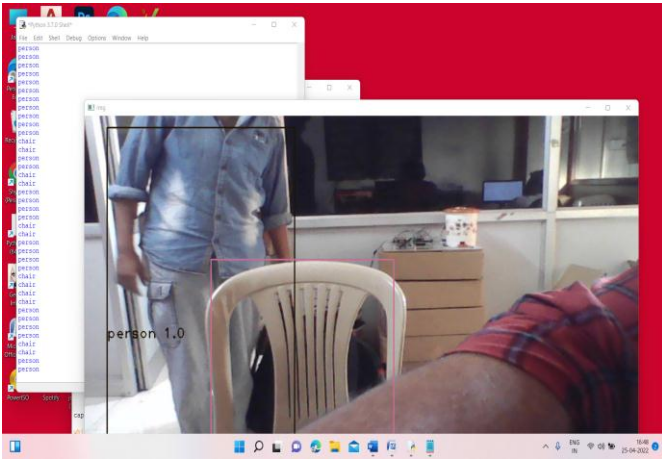
### ESPEAK

eSpeak is a small open source software voice synthesizer for Linux that supports English and other languages. Speak employs a technique known as "formant synthesis." This enables a large number of languages to be supplied in a little space. The voice is crisp and fast, but it lacks the naturalness and smoothness of larger synthesizers based on genuine speech recordings. eSpeak now supports over 50 languages, although it didn't support the Japanese language from the start of the project. eSpeak's speech is very flexible, but it lacks the naturalness of larger synthesizers are using unit-based synthesis and thus are based on human speech recordings. Voiced speech (e.g. vowels and sonorant consonants) is formed using formants in formant-based synthesis. Unvoiced consonants, on the other hand, are made with pre-recorded sounds. Additionally, voiced consonants are formed by combining formant-based voiced sounds with a pre-recorded unvoiced sound. The eSpeak system works using easy-to-understand text files called modular language data files.

### EXPERIMENTAL RESULTS







## CONCLUSION

The impetus and concept for the project came from a desire to help visually impaired persons overcome their challenges. Many techniques for implementing object detection were discovered, however the use of the Open CV Library and YOLO was indeed the best option. Based upon object recognition in frames, we describe a visual substitution method for blind persons. For object identification, this system uses Yolo configuration, weights, and feature matching. The experimental section is dedicated to putting the program to the test to detect some items in various frames under various settings. This research proposes a revolutionary navigation system for visually impaired individuals to assist them in securely and quickly reaching their destination. When it detects something, it sounds a bell to inform visually impaired persons.

## REFERENCES

- [1]A. Abdurasyid, I. Indrianto, and R. Arianto, "Detection of immovable objects on visually impaired people walking aids," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 17, no. 2, pp. 580–585, 2019, doi: 10.12928/TELKOMNIKA.V17I2.9933.
- [2]A. Corovic, V. Ilic, S. Duric, M. Marijan, and B. Pavkovic, "The Real-Time Detection of Traffic Participants Using YOLO Algorithm," *2018 26th Telecommun. Forum, TELFOR 2018 - Proc.*, 2018, doi: 10.1109/TELFOR.2018.8611986.
- [3]R. Bharti, K. Bhadane, P. Bhadane, and A. Gadhe, "Object Detection and Recognition for Blind Assistance," *International Research Journal of Engineering and Technology (IRJET)*, Volume: 06, Issue: 05, pp. 7085–7087, May 2019.
- [4]O. Masurekar, O. Jadhav, P. Kulkarni, and S. Patil, "Real Time Object Detection Using YOLOv3," *Int. Res. J. Eng. Technol.*, vol. 07, no. 03, pp. 3764–3768, 2020.
- [5]S. Vaidya, N. Shah, N. Shah, and R. Shankarmani, "Real-Time Object Detection for Visually Challenged People," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iciccs, pp. 311–316, 2020, doi: 10.1109/ICICCS48265.2020.9121085.
- [6] I. A. M. Zin, Z. Ibrahim, D. Isa, S. Aliman, N. Sabri, and N. N. A. Mangshor, "Herbal plant recognition using deep convolutional neural network," *Bull. Electr. Eng. Informatics*, vol. 9, no. 5, pp. 2198–2205, 2020, doi: 10.11591/eei.v9i5.2250.
- [7] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In 2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV (pp. 779–788).
- [8] Liu, W., et al. (2016). SSD: Single shot MultiBox detector. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer vision—ECCV 2016*. ECCV 2016.
- [9] Ning, C., Zhou, H., Song, Y., & Tang, J. (2017). Inception single shot MultiBox detector for object detection. In 2017 IEEE international conference on multimedia & expo workshops (ICMEW), Hong Kong (pp. 549–554).
- [10]L.-B. Chen, J.-P. Su, M.-C. Chen, W.-J. Chang, C.-H. Yang, and C.-Y. Sie, "An implementation of an intelligent assistance system for visually impaired/blind people," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2019, pp.