The Object Recognition Voice Assistant for Visually Impaired People

Mr. P. Ravinder Rao[1], G. Taruni[2], D. Meghanand Reddy[3], Md. Azeem Hussain[4]

[1]Assistant Professor, [2,3,4]Student, Dept. of Computer Science and Engineering, Anurag Group of Institutions, Hyderabad, India

Abstract - Image processing is the procedure of performing any task on a picture to make it suitable for retrieving relevant information. This relevant information consists mostly of properties or qualities of the picture sent into the camera module. The image processor's output is then used to detect different items in the image that used an object recognition computer vision approach. Object identification techniques use machine learning to find numerous patterns from picture feature extraction and discriminate between different items. As a result, we apply computer vision and advanced algorithms to give a model for recognizing an object for a blind person's environment and gives voice command for the concerned person This notion might benefit those struggling from vision, and other conditions. Sensing their surroundings and supporting them in knowing what is around them with an estimated object in the form of speech might aid them in choosing a path to move along. Although there are other applications that aid blind or visually impaired persons in navigating public spaces by employing additional personal assistance, Blind persons can use this application with voice commands to open camera and get output in the form of voice suggestions. A Interface will be developed between the two components. Using the live camera, the blind or visually impaired persons can use rear-facing camera, the sighted assistant may now aid the blind or visually impaired. Yet, in this project, we attempted to create a system that can function independently of any other individual.

Keywords – Recognition, Voice, visually, Impaired people.

I. INTRODUCTION

1.1 Introduction

According to Borne et al work.'s in Lancet Global Health [1], an estimated 217 million individuals have moderate to serious impaired vision, and 36 million are blind. Operational presbyopia impacts an estimated 1094.7 million people, with 666.7 million of them being over the age of 50. The growth in the number of elderly people will raise the amount of the population at risk of vision problems [1]. Furthermore, these visual impairments are not confined to the elderly; youngsters under the age of 15 who are during the prime of their life are also affected by these problems. Furthermore, the The World Health Organization (WHO) estimates that around 2.2 billion individuals worldwide suffer from vision impairment. Impairment. As a result, these persons are in desperate need of assistance to improve their eyesight and slow the advancement of their handicap in order to better feel the world [2]. In the middle of this, 15 million people are blind, makes The country the country with the world's largest blind community [3]. It is projected that 40 to 45 million people worldwide are blind and unable to walk alone help [4]. There are a few applications that aid blind or visually impaired persons in navigating public areas with the support of participants from these apps. Yet, these applications have the constraint of always relying on the support of volunteers. But, in this study, we provide a system, which also will exist freely of any volunteers, i.e., alone without support of anybody else.

Scope of the Project:

First and foremost, youngsters are the future of all nations. We can make the city more liveable. Traditional ways are being superseded by new methods as technology develops. But this progress is what the world desires. Several studies have been conducted to find solutions to everyday problems. Humans can distinguish items and their information in milliseconds. Finding information about items around them and mobility challenges are two of the most significant disadvantages that blind individuals face on a daily basis. Basic items are difficult for them to recognize, as is distinguishing objects of the same dimension and size.

1.2 Objective

Deep learning has been a more common approach in past few years for addressing these types of issues to recognize the entities [11, 12, 29, 27, 14]. Deep learning algorithms attain excellent average accuracy while being less expensive. To handle identification and recognition problems, many Convolutional Neural Network (CNN) approaches, such as Single Shot Detection (SSD) [13] and

 DOI : 10.35291/2454-9150.2023.0030

You Only Look Once (YOLO) [17], are applied. In this study, we will use a YOLO V5 Model to identify objects that used a camera. Moreover, a voice synthesis tool is required for converting identified visual speech into text. This is possible with gTTS and pyttsx3.

II. LITERATURE SURVEY

A Bootkoski et al [1] introduced the YOLOV4 design for remarkable object detection speed and precision. Next it provides many characteristics that are supposed to boost CNN efficiency before comparing and combining them to accomplish things on the Training data. Then it goes on to improve cutting-edge methods in order to make them more economical and fit for single GPU learning.

Daksha Janardhan et al [2] suggested a system for identifying objects with voice feedback based on the YOLOv4 model. It begins by stating the benefits of yolov4, and then provides a flowchart for its system design and implementation in the Android operating system. Its fundamental concept is to create a system that outputs speech in real-time using yolov4.

Wong et al. [33] suggested an actual CNN-based object recognition software for individuals with visual impairments in 2019. The objects group was captured in real time using a camera, but the photo function was disabled. A acoustic detector was then developed to identify the sight of people who are blind. Nasreen, Jawadi, and colleagues [18] described a method for helping visually challenged persons through the item's detection procedure. The created approach imports a rear camera image into a webpage and transmits it onto the website, where the YOLO model is used to identify the objects on the server side. Arjun et al. [20] introduced a wearable technology that includes wearable technology and footwear. Both smart shoes and smart glasses detect and avoid the barrier.an audible output to the user. Rahman, Ferdous, and colleagues [24] created a visually challenged object identification model using the YOLO algorithms and CNN.

The authors of proposed a process called Crop Selection Procedure in [4]. (CSM). The researchers describe a method for determining crop challenges, increasing yield potential net yield rates over weather conditions, and maximizing economic growth. The novelists looked into the different influential dimensions that can be used by various predictive model types for crops. The authors also describe algorithms and various ml methods. Crops were defined as seasonal crops, thought the entire crops, short-term horticultural products, and long-term plantation crops in the pro - posed Agriculture Selection.

Xinyi Zhou et al. [4] developed fully convolutional applications in different fields of object identification, and with the fast expansion of deep learning, its implementations have touched numerous study areas and obtained state-of-the-art achievements. And Android is said to be the most preferred choice among visually challenged users. Nonetheless, various specialist navigation and visual recognition gadgets are in use. Nevertheless, the biggest disadvantage is that they are relatively costly when compared to software. One of the key advantages of using deep convolutional neural network for object recognition is that it delivers excellent accuracy while trained on large datasets, however it is difficult to describe lengthy dependence using existing recurrent neural systems. [5]

Joseph Redone al. presented a "You Only Look Once: "Real-Time Object Recognition" detects things in real time. It was the only object tracking model presented at the time that employed analysis to dimensional characters segregate anchor boxes and their associated confidence score. It employed the Single Shot Detector (SSD) to estimate anchor boxes and conditional probabilities from pictures based on a single assessment. Because it is a network node model, its system is designed from start to finish. It made more localization errors than other detectors at the time, although it has a low risk of predicting negative detection when none occurs. Hrushevsky et al. [6] employed the Alex Net framework, which was the initial to win the ImageNet competition. It was a game-changing network that changed the emphasis from hand programming to DCNN. The GPU and The Rectified Linear Unit (RElu) was the device that prepared the way for the move from handcraft programming to features learning via the CNN model. After that, the researcher proceeded to investigate diverse uses of CNN and deep CNN in object tracking.

Geetha Priya. S et al [7] submitted a paper to properly appreciate YOLO. It compares many parameters such as speed, precision, complexity, and other attributes with various diversity estimate and object identification algorithms. Based on all of the criteria, the YOLO model promises to look at photographs differently from any other classifier. It forecasts anchor boxes and all class probability using a convolutional network, and it does so quicker than any other technique. It also provides information on many topics YOLO model might also be utilized in the following sectors.

III. OVERVIEW OF THE SYSTEM

3.1 Existing System

In existing work a system that uses an Android smartphone to assist a blind user with collision avoidance and routing. Smartphones are now available to anybody. In fact, smartphones have emerged as the most widely available instrument. As a result, this implementation uses a Smartphone with a camera that identifies items in its surrounding and produces audio output. The device's hearing capacity attempts to compensate for his lack of vision.

Sensor based fusion frame work is developed which users various sensors to help visual impaired.

 DOI : 10.35291/2454-9150.2023.0030

3.1.1 Disadvantages of Existing System

This application was developed on android operating system. It has limitations of battery and time of prediction. Users need to carry sensors and hardware kit which is not possible and only few objects can be detected in this method.

3.2 Proposed System

In proposed system we design a voice-based access system for visually impaired people where users can open camera with voice command and camera will detect live objects and give result in the form of voice. This system will help users

Advantages of Proposed System

- Using this application blind persons can know about objects with voice commands which help to understand in travel path.
- More than 100 objects can be detected from live video.

3.3 Proposed System Design

In this project work, I used five modules and each module has own functions, such as:

1. Data Collection
2. Yolo Object Detection
3. Object Capturing
4. Flask framework
5. Prediction

3.3.1 Data Collection

The proposed model is trained, validated, and tested using YOLOv5 on a custom produced dataset paired with the MS COCO 2017 Dataset [15]. MS COCO 2017 has 80 various object categories such as human, bike, table, car, truck, cycle, and so on. We have introduced 100 new class labels, including Which are included in the MS COCO 2017 Data source? (95 classes overall). These items are pertinent to the Indian environment. We contributed 30 - 50 photographs to the database for each item type, for a total of 500 photos. In all, 5000 photos are examined for feature extraction.

3.3.2 Yolo Object Detection

The "You Only Look Once" (YOLO) technique to object recognition focuses on finding objects in images and sorting those into a grid layout. Each square cell is responsible for locating things within its limits. YOLO v5 is now one of the finest models for detecting objects around. The wonderful thing about any of this Deep Neural Network is how simple it is to retraining it on our unique dataset [31]. The YOLO v5 variant takes up around 90% less capacity than the YOLO v4 variant. YOLO v5 is touted to be substantially quicker and lightweight than YOLO v4, with comparable to the YOLO v4 assessment. As a consequence, we went with YOLO v5.

3.3.3 Object Capturing

In this step opencv library is used with tk inter framework when user executes a python file which runs is listening mode. Application users' speech to text library which will convert user voice in to command and camera will open after detecting open camera command. Camera will capture live frames and preprocess using Opencv and process data to yolo model.

3.3.3 Predicting objects and voice Output

In this step trained yolo model is loaded in to application and each preprocessed frame is given as input to model which will detect multiple objects in the frame and draw bonding boxes for each object and processed to live streaming. Predicted objects data is in text format which is converted to voice output.

IV. ARCHITECTURE

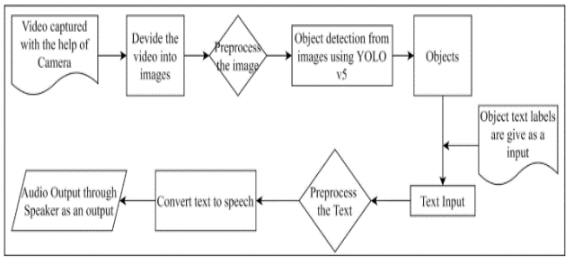


Fig 1: Frame work of object detection

Then, we must identify the things in the picture or video. Tensorflow API was utilised for object identification, which would be able to distinguish various items in the photos. "Human, bicycle, vehicle, motorbike, aircraft, and other more than 100 objects OpenCV was employed to input video sequence to the object tracking model in the case of videos. After identifying the item and classifying it, the parameters of the structuring element are assessed and used to estimate the object's position from the cameras. That is similar to the learning of neural network models, which we must explain or, to put it another way, train the algorithm using instances and estimating the voice commands are given a output of objects for other objects.

V. RESULTS SCREEN SHOTS

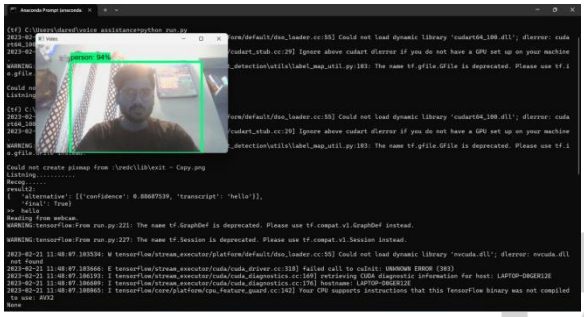
Main page

 DOI : 10.35291/2454-9150.2023.0030

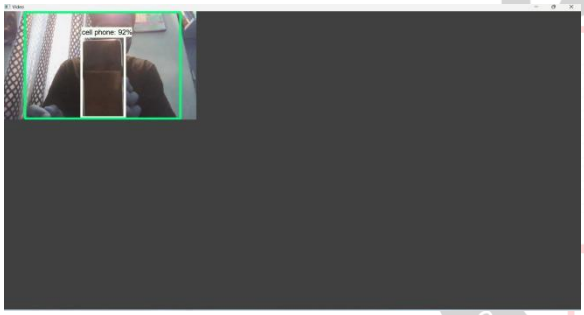
Voice command in Listening Mode:

Open camera speech to text:

Camera Open:



Detect object:



VI. CONCLUSION

For object detection, YOLOv5 is employed, while pyttsx3 and gTTS are used for speech production in the suggested. The suggested technology can reliably detect objects form live video stream and use speech to assist individuals with visual impairments who are disabled. With high degree of accuracy, the proposed model YOLOv5 can detect 100 distinct objects. Among the two Python libraries for voice creation, we discovered that pyttsx3 for voice to text conversion. Voice command is taken as input form user and camera will open. Live camera can track objects and send result in voice format which is easy for blind people to understand objects around them.

VII. FUTURE ENHANCEMENT

In future this application new objects can be trained and increase scope of object detection. This application can be integrated to spectacles for effective usage.

VIII. REFERENCES

[1] Bourne R, Flaxman S, Braithwaite T et al. (2017) Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. The Lancet Global Health 5:e888-e897. doi: 10.1016/s2214-109x(17)30293-0.

[2] World Health Organization (WHO). https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[3] India has the largest blind population (Kounteya Sinha, Oct 11, 2007). https://timesofindia.indiatimes.com/india/India-has-largest-blind-population/articleshow/2447603.cms

[4] World Health Organization (WHO). https://www.who.int/blindness/Vision2020%20-report.pdf.

[5] GOLLEDGE R, LOOMIS J, KLATZKY R et al. (1991) Designing a personal guidance system to aid navigation without sight: progress on the GIS component. International journal of geographical information systems 5:373-395. doi: 10.1080/02693799108927864

[6] Petrie H, Johnson V, Strothotte T et al. (1996) MOBIC: Designing a Travel Aid for Blind and Elderly People. Journal of Navigation 49:45-52. doi: 10.1017/s0373463300013084

[7] L. Ran, S. Helal and S. Moore, "Drishti: an integrated indoor/outdoor blind navigation system and service,"Second IEEE Annual Conference on Pervasive Computing and Communications, 2004. Proceedings of the, 2004, pp. 23-30, doi: 10.1109/PERCOM.2004.1276842.

[8] S. Chumkamon, P. Tuvaphanthaphiphat and P. Keeratiwintakorn, "A Blind Navigation System Using RFID forIndoor Environments," 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2008, pp. 765-768, doi: 10.1109/ECTICON.2008.4600543.

[9] D. Wahrmann, A. Hildebrandt, R. Wittmann, F. Sygulla, D. Rixen and T. Buschmann, "Fast object approximation for real-time 3D obstacle avoidance with biped robots," 2016 IEEE International Conference on Advanced Intelligent Mechatronics (AIM), 2016, pp. 38-45, doi: 10.1109/AIM.2016.7576740.

[10] Imran Ahmed, Misbah Ahmad, Joel J.P.C. Rodrigues, Gwanggil Jeon, Sadia Din, A deep learning-based social distance monitoring framework for COVID-19, Sustainable Cities and Society, Volume 65, 2021, 102571, ISSN2210-6707, https://doi.org/10.1016/j.scs.2020.102571.

[11 Howard, A.G., Zhu, M., Chen, B., Maleficence, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*

[12] Google's MobileNets on the iPhone (14 JUNE,2017). http://machinethink.net/blog/googles-mobile-net-architecture-on-iphone/

[13] MobileNet version 2 (22 APRIL 2018). http://machinethink.net/blog/mobilenet-v2/

 DOI : 10.35291/2454-9150.2023.0030