iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: II Month of publication: February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49107

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

Object Detection System with Voice Alert for Blind

Dr. M Y Babu[1], Akash Jatavath[2], G Yashwanth Kumar Reddy[3], Pittala Arun Kumar[4]
[1]*Associate Professor, Dept of Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad*
[2, 3, 4]*UG Student, Dept of Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad*

Abstract: ***As we can see, there are numerous blind persons nearby who encounter various challenges, such as difficulty in crossing roads and identifying objects in their environment. With the advancement in technology in several fields, human life is evolving to better standards. Unfortunately, those who are blind are unable to fully enjoy this kind of lifestyle. So, this project is one strategy for introducing blind individuals to a new way of living that makes them independent on others. The major goal of this project is to create a deep-learning algorithm that can be used to analyse the environment for people who are blind by using the rapidly evolving technology. We'll accomplish this using object detection and transform the data into speech alerts and warnings. Real-time object detection is one of the more challenging tasks since it requires continuous processing and takes a long time. The convolution neural network is the main backbone for any type of object detection (CNN). We can create algorithms based on photos and videos by employing a convolution neural network. We utilise the YOLO technique for object detection because it is simple and quick to process. In addition, for the voice warnings, we employed Text to Speech (TTS). The dataset used in this technique is the COCO dataset, which contains the names of things and objects in our daily lives. These algorithms have been thoroughly trained by the over 90 outdoor objects that we view every day in our daily lives.***
Keywords: Deep Learning; YOLO; TTS; CNN

I. INTRODUCTION

One of the most important organs in human body are eyes. We enjoy the beauty of nature, various types of books, and many other aspects of our lives. We can go anywhere independently and have fun with friends and family . What if we are blind? Forgot about enjoying, what if we don't even do our own work independently? What if we must depend on some others for regular daily works? It is difficult to think and imagine these kinds of situations. However, some of us in society are visually impaired. They must depend on others for their regular work. The ability to visualize the surroundings is a gift.

The visually impaired people face many difficulties in their day-to-day life. They face many difficulties in object detection and analysing of their surroundings. While Walking on the streets, they face many difficulties in identification and recognizing the objects .These cause them many injuries and accidents etc . So, to put an end to these difficulties we came with an idea of recognizing the objects around them by using object detection and converting them in to voice messages to identify and understand the situations around the person. The use of technology in finding and recognizing objects is huge because of these rapid increase in the technology like AI, ML and Deep Learning etc facilitate many tools and libraries for the development new ideas which are useful in contemporary society like smart sticks, navigators etc.

Huge number of research and developments are going in the domain of machine learning and object detection .Large number of new kind of tools are also came in to the existence .Few of those developments are similar to our idea .But all those projects implementation has distinctions and differences in object detection like using of different algorithms and different libraries for the processing .Our dataset contain nearly 90 object names which are useful and observed by a common man in our day to day life , Which is enough for the real time object detection. we use YOLO algorithm for object detection and Text to Speech conversion technique for voice alerts.

II. RELATED WORK

A. Literature Survey

1) Prof. Seema Udgirkar, Shivaji Sarokar, Sujit Gore, Dinesh Kakuste, Suraj Chaskar, "Object Detection System for Blind People"

The author of this work attempts to demonstrate how their smart vision, the goal of which is to move about the environment using a user-friendly interface system, was proposed. These writers created a system that can detect obstacles that are close to his head, particularly while entering through a door.

Simply put, it is designed to protect his head from harm. This item is made to help blind people traverse any area. It uses the user's buzzer and vibration as two output modes to direct the user toward an object and provide information about an impediment. There are two different operating modes: buzzer mode and vibration mode. These outputs are made available for blind people.

2) *Jigar Parmar, Vishal Pawar, Babul Rai, Prof. Siddhesh Khanvilkar," Voice Enable Blind Assistance System -Real time Object Detection".*

In this study, authors tried to identify a presented object in front of a webcam. They trained and tested the TensorFlow Object Detection API frameworks using made a model. To alleviate Input / Output concerns, a good frames per second solution is needed because reading a frame from a web camera causes a lot of issues. As a result, they concentrated on threading technology, which dramatically reduces processing time for each item while improving frames per second. The item detected box takes about 3-5 seconds to move over the next object in the video, even though the application correctly identifies everything in front of the webcam.

III. METHODOLOGY

A. *Object Recognition*

Though similar approaches to object identification, object discovery and object recognition operate differently. Even though both are widely used for images and video. Object detection is considered a subset of object recognition in processing. Object detection and recognition are often employed in a wide range of sectors, from personal security to workplace productivity. It is used for autonomous driving systems, machine inspection, surveillance, security, and image retrieval, among other computer vision applications. In general, the text to speech conversion capability cannot be recommended for non-operating system devices. Thus, Android- or iPhone-based smartphones are the most popular choice among smart phone users who are blind or visually impaired.
Object discovery is the phenomenon of finding instances of items in both still images and films that contain the objects in question. Bounding boxes and information about the identified items' locations inside the frame are highlighted. Technology related to both image processing and computer-aided vision is object detection. It classifies and identifies a wide range of things from movies and digital photos, including people, animals, and cars. Multiple things in a video or digital picture can be swiftly classified using object detection. Although object detection has been around for a while, it is currently more prevalent than ever across a variety of sectors. The object detecting system has been put into action using a variety of techniques.

1) *YOLO Algorithm*

The YOLO algorithm was initially proposed by Joseph Redmon and his colleagues. In 2015, he released a paper on YOLO under the heading "You Only Look Once" Real-Time recognition, and it became immensely successful right away. CNN is followed by YOLO. When making predictions, the algorithm only "looks once" at the image since there is only one propagation that occurs throughout the neural network. Compared to other methods of object identification, the YOLO model is the fastest and most effective. The main benefit of YOLO is its quickness. There are 45 frames per second in this. The model is constructed in a concise manner to acquaint its network with an abstract description of things. The primary goal of object detection is to identify one or more specific things in audio visual or digital pictures. Contrarily, object class recognition classifies items into a certain category or class. Every thing has unique qualities that make it easier to distinguish it from other objects in movies or pictures. Additionally, it sets them apart from other classes. Object detection is the process of finding and defining objects, such as people, animals, objects, cars, and so on. Combining the You Only Look Once (YOLO) architectural algorithm with the COCO dataset results in a quick and effective deep learning technique for object recognition.

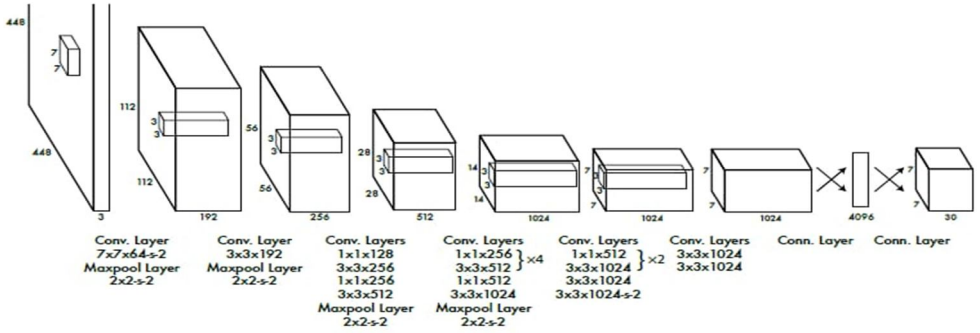


Figure 1: Yolo architecture

YOLO is made for comprehensive image processing and steadily raises the effectiveness of object detection. Frame identification is seen as a regression issue in this situation. In order to quietly store specialised information about groups and their looks, YOLO employs the entire background throughout training and testing periods while the networked is concentrated on recent images.
To continuously predict all bouncing boxes throughout all groups for a picture, it uses features from the full image. The method divides the informational image into a SxS example. The matrix cell can identify the point and choose the certainty scores for those containers when the focal point of an object falls within a network cell.

B. Estimation of Image Position

We must generate a bounding box for each identified object in order to approximate the position of the image. Using the specific bounding box's height and width in relation to the image frame. To estimate the position of an object within a bounding box, 5 values are used. The position of the item is shown by the first 4 values: bx, by, bw, bh, BC the fifth value, specifies how much of a box an object occupies.

$$BC=pr*IOU$$

Where IOU = Intersection over Union

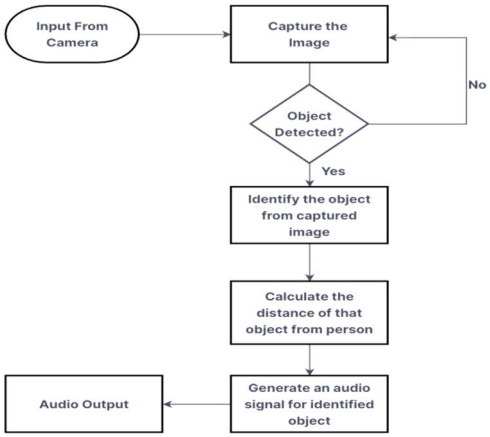
Pr = Object existing in box

This estimates the predictability and likelihood of the box containing an object of any class. If there is no object in the box, BC=0, otherwise BC=1.

C. Voice Generation

When the system locates the desired object, voice guiding is a feature that offers information to specific users, such as those who are blind, in a convenient manner. It is crucial to alert the blind person heading in the way of the presence of an object when it has been discovered.PYTTSX3 is an essential part of voice generation module. Text to speech conversion can be done using the Python library Pyttsx3. Python version 2 and 3 are both compatible with this package. A straightforward tool for text to speech conversion is Pyttsx3. We also used Google Text to Speech (GTTS) for voice alerts. Google Text to speech contain many inbuilt English accents for the users who are from different parts of the world. It is very easy to use, it converts the text into audio which can be saved as a mp3 file It also supports many regional languages which is also useful for those who do not able to understand English.

1) *Dataset:* Here we used COCO dataset. COCO stands for common object in context. The COCO dataset contains challenging, high-quality visual datasets for computer vision. The images in the dataset were gathered from commonplace locations that provided specific information. In real-world situations, multiple items or things may be contained within the same frame, and each one needs to be distinguished as a distinct object and properly segmented. The identification and segmentation of the objects visible in the photographs are contained in the COCO dataset. We used this information to develop our item recognition and detection technology for persons with disabilities. This dataset contains approximately 90 objects or items.
2) *Process Model:* The working of this system is represented in the below process model. Input is taken from the user's camera to capture the images. The system checks if any objects are detected in the image. If an object is detected, the system identifies the object. Then the system calculates the distance of the object from a person. Based on the calculated distance, the system generates an audio output.



IV. RESULTS

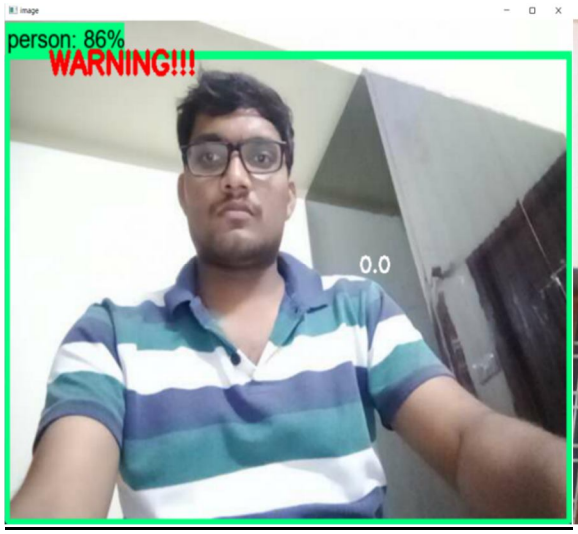


Figure 3: Detection of a person with 86% Accuracy

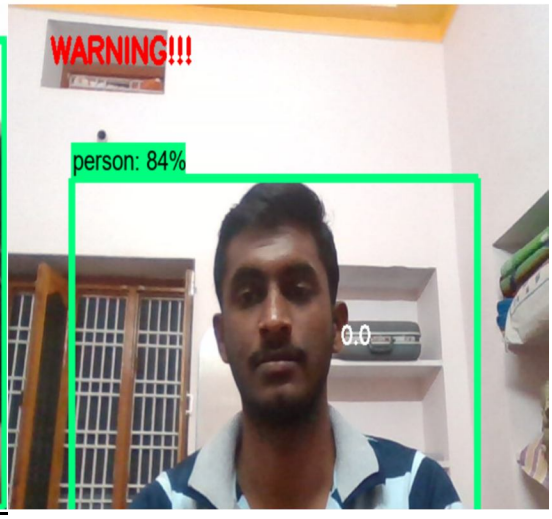


Figure 4: Detection of a person with 84% Accuracy

Figure 5: Detection of a Bottle with 63% Accuracy

Figure 6: Detection of a Bottle with 56% Accuracy

V. CONCLUSION AND FUTURE SCOPE

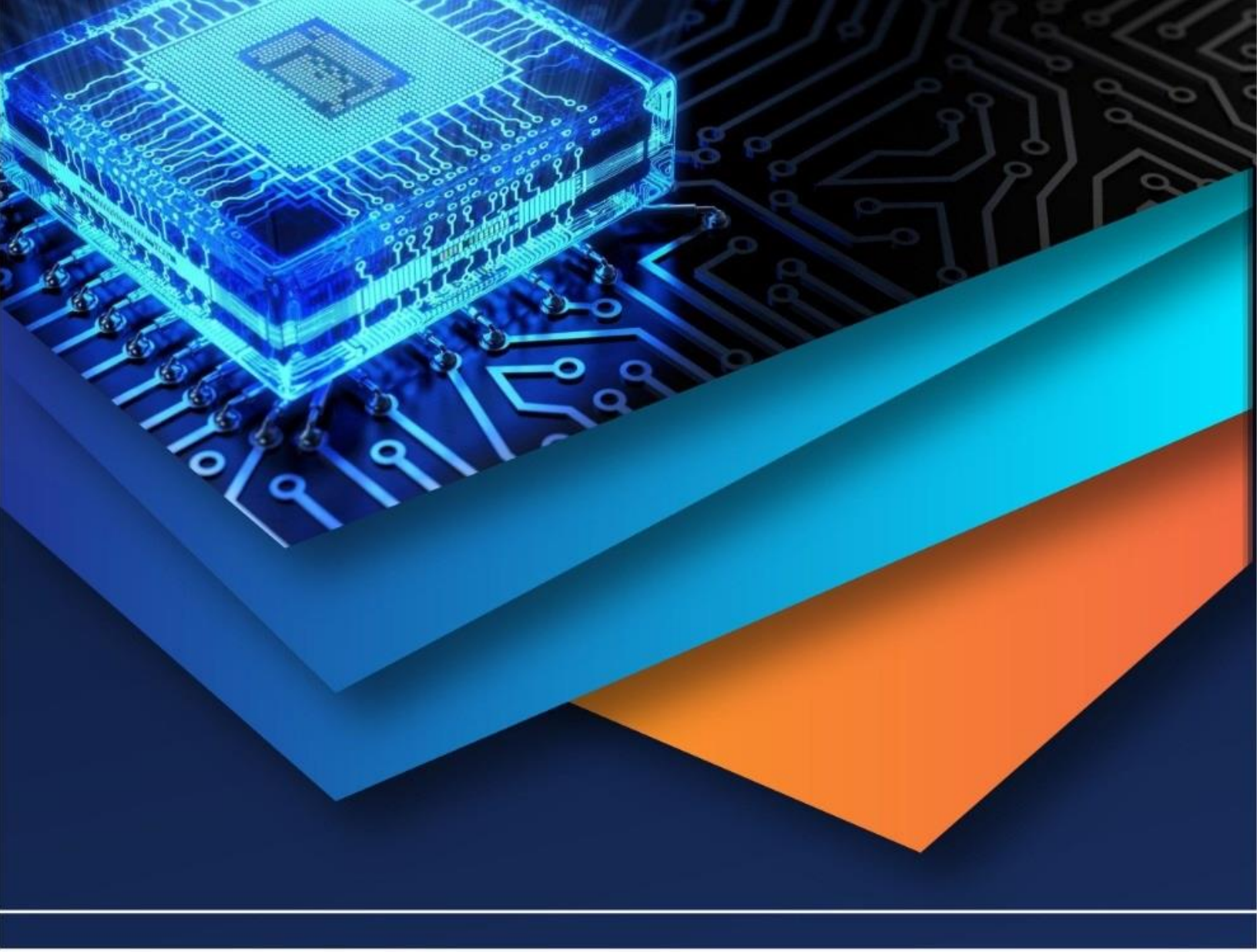
In this project we used image recognition, voice generation modules for the development of the project. As of now accuracy is good but in case if we want to increase the accuracy we have to train the model with more object/images in the dataset. This project is a small experiment which is useful for blind persons, can be able to find the objects which are surrounded by them, and they are in a position of taking care of themselves when they are outside.

The ability of the blind person to stand alone and carry out tasks independently makes this blind assistance device useful for object detection by voice warnings. The device's camera serves as the blind person's virtual eye, capturing every detail of their environment. The voice alerts keep the person informed about the surroundings so that accidents are decreased. Reduced rely on other parties.

There are so many people present in the world who are visually impaired and illiterate from different parts of the world .Some of them do not understand other languages except their local language in their local accent . so, one of the future scope for this project is to develop the idea in which voice alerts in such a way that they can use their own local language .

REFERENCES

[1] Miss Rajeshvaree Ravindra Karmarkar, Prof.V.N. Honmane "Object Detection System For The Blind With Voiceguidance"- Published Online June 2021 in Ijeast
[2] Jigar Parmar, Vishal Pawar, Babul Rai, Prof. Siddhesh Khanvilkar "Voice Enable Blind Assistance System -Real time Object Detection"- IRJET, Apr 2022
[3] Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam " Real-Time Object Detection with Yolo"- IJEAT,Feb 2019
[4] Mayuresh Banne, Rahul Vhatkar, Ruchita Tatkare "Object Detection and Translation for Blind People Using Deep Learning"- IRJET ,Mar 2020
[5] M. I. Thariq Hussan, D. Saidulu, P. T. Anitha, A. Manikandan and P. Naresh "Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People"- IJEER,June 2022
[6] Priya Kumari, Sonali Mitra, Suparna Biswas , Sunipa Roy, Sayan Roy Chaudhuri , Antara Ghosal , Palasri Dhar, Anurima Majumder "YOLO Algorithm Based Real-Time Object Detection"- IJIRT, June 2021
[7] N. V. N. Vaishnavi , Tummala Navya , Velagapudi Srilekha , Vinnakota Karthik , D. Leela Dharani "Blind Assistance In Object Detection And Generating Voice Alerts"- DRSR,Feb 2021
[8] Tanvir Ahmad , Yinglong Ma , Muhammad Yahya,Belal Ahmad,Shah Nazir , and Amin ul Haq "Object Detection through Modified YOLO Neural Network"- Published 6 June 2020 in Hindawi Scientific Programming
[9] Joseph Redmon, Santosh Divvala, Ross Girshick, "You Only Look Once: Unified, Real-Time Object Detection",The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
[10] Dumitru Erhan, Christian Szegedy, Alexander Toshev, "Scalable Object Detection using Deep Neural Networks", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2147-2154
[11] N. V. N. VAISHNAVI , TUMMALA NAVYA , VELAGAPUDI SRILEKHA , VINNAKOTA KARTHIK , D. LEELA DHARANI ," BLIND ASSISTANCE IN OBJECT DETECTION AND GENERATING VOICE ALERTS" ,in 2021



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY