Object Recognition using TensorFlow and Voice Assistant

Diya Baldota[1], Sunny Advani[1], Sagar Jaidhara[1], Prof.Amit Hatekar[1]
[1]Department of Electronics and Telecommunications Engineering
Thadomal Shahani Engineering College, Mumbai, India

Abstract—The world of innovation has prospered at a fast rate over the last decade with the rise of more up-to-date and more current developments. With the utilization of those more current advances, our lives got to be swifter. The fast advance of information and arrange innovation has advanced from the internet and mechanization frameworks that were initially utilized for regulatory workplaces and mechanical and commercial applications to the apparatus of those advances all over in life. The Internet has also become progressively well known. Each family has arranged a scope. Individuals started to seek a more helpful and superior living environment and started to ponder the application of portable gadgets, apps, and versatile systems in natural checking, machine automation, smart home, etc.

Proficient and exact object acknowledgment is a pivotal point inside the headway of computer vision frameworks. With the entry of machine learning and deep learning methods, the precision for object location has expanded radically. The project aims to incorporate an android application for object recognition with its localization to achieve high accuracy with real-time performance.

Keywords—Apparatus; versatile; automation; proficient; machine learning; deep learning.

I. INTRODUCTION

The fast progress of data and organize technology has advanced from the Internet and computerization systems that were initially utilized for authoritative offices and mechanical and commercial applications to the application of these innovations everywhere in life. Once you think of augmented reality, one of the key components to consider is object acknowledgment innovation, moreover known as object detection. This term alludes to a capacity to distinguish the frame and shape of diverse objects and their position in space caught by the device's camera.

It's a known reality that the evaluated number of visually disabled individuals within the world is almost 285 million, roughly break-even with the 20% of the Indian population.

They suffer-normal and consistent challenges in Navigation particularly when they are on their own. They are generally dependent on somebody for indeed getting to their fundamental day-to-day needs.

So, it's a very challenging errand and the mechanical arrangement for them is of most extreme significance and much required.

One such attempt from our side is that we came up with an Integrated Machine Learning Framework permits the Blind Casualties to distinguish and classify Genuine Time-Based Common Day-to-day Objects and produce voice feedbacks and calculates distance which produces warnings whether he/she is exceptionally near or distant absent from the object. The same framework can be utilized for Obstacle Detection Instrument.

The primary reason for object detection is to find vital things, draw rectangular bounding boxes around them, and decide the course of each thing found. Applications of object discovery emerge in numerous diverse areas counting recognizing people on foot for self-driving cars, checking rural crops, and indeed real-time ball following in sports

II. OBJECTIVE

The project aims to incorporate state-of-the-art techniques for object detection to achieve high accuracy with real-time performance. In this project, we use Python with an TensorFlow-based approach to solving the problem of object detection in an end-to-end fashion. The the resulting system is fast and accurate. A TensorFlow based application for an Android mobile the device, using its built-in camera is built for detecting objects, more specifically:
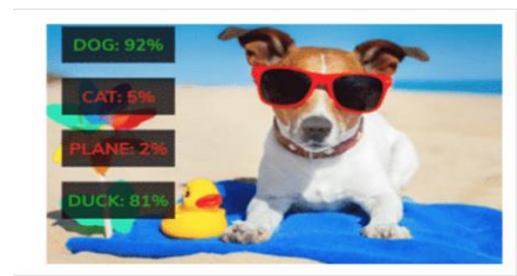

Fig. 1. These objects were chosen for object recognition and represent a dog and a duck on the beach

The framework is set up in such a way where an android application (assuming you're executing it on an Android gadget) will capture real-time outlines and will send them too the background of the application where all the computations take place.
• The background of the application is where the video stream is sent and is taken as an input,
which goes through the COCO DATASETS object detection model which tests and detects with accurate metrics.
• After testing with the assistance of voice modules the course of the object will be changed over into default voice notes which can at that point be sent to the blind victims for their help.

• Along with the object discovery, we have utilized an alarm framework where approximate will get calculated. In case the Blind Person is especially close to the diagram or is distant away at a more secure put, it'll produce voice-based yields alongside distance units.

III. LITERATURE SURVEY

To solve the visually impaired peoples' problem, number of methods and techniques are introduced.

In paper [1], it gives an effective demonstration of the method of detecting the object and analyzing the gesture of an object using machine learning and computer vision for choice making after surveying different researches within the field of pattern recognition.

In paper [2], the author proposed a well-known computer technology associated with computer vision and image processing that focuses on detecting objects or their instances of a certain course (such as humans, flowers, creatures) in computerized pictures and videos. There are different applications of object detection that have been well inquired about counting face detection, character recognition, and vehicle calculator. Object detection can be used for different purposes including recovery and surveillance. In this ponder, different essential concepts utilized in object detection whereas making use of the OpenCV library of python 2.7, progressing in the effectiveness and exactness of object detection are displayed.

In paper [3], the author described that everyone has the right to live independently, particularly the ones disabled, for the past decades, technology has helped the disabled to have control over their lives. In this research, an assistive system for the blind is proposed using YOLO for the detection of objects within images and video stream based on deep neural networks to make precise detection, and OpenCV under Python using Raspberry Pi3.The output obtained indicates the success of the proposed approach in providing blind users the capability to move around in unfamiliar indoor-outdoor environment, through a user-friendly device by person and object identification model.

IV. PROBLEM STATEMENT

The evaluated number of people visually impaired inside the world is 285 million, 39 million blind, and 246 million have moo vision. They are an important portion of our society. It's very difficult for them to every time perceive the outside world. In this present-day society, visually impaired people require supportive instruments in their day-to-day life. Existing devices cannot be utilized to prepare genuine-time data in the world. Our thought primarily centered on designing and actualizing an assistive framework for visually impaired people to detect objects effectively.

This study is aimed to create an application for visually impeded individuals so that they can utilize this application in real-time. All the plan and format are kept by considering the visually disabled individuals within the intellect. The point is to plan a low-cost and high-performance assistive application for the everyday activities of visually impaired people.

Fig. 2. Process of Image Detection

V. PROPOSED MODEL

The system is implemented in an android application that detects various objects in real-time.

A. System Overview

The system uses a smartphone to capture real-time input data. The app's camera is automatically accessed and it begins capturing the encompassing objects in case there's any present. Information is sent to the background where it is handled utilizing machine learning.

Fig. 3. System Overview

B. Implementation and Development Module

The framework is implemented by combing different technological stacks which are talked about below. Android studio is utilized for the advancement of the application because it is the official integrated development environment (IDE) planned particularly for android application improvement.

OpenCV library is utilized for picture capturing since it provides support to real-time applications. [5]

Python programming language is used for building the machine learning model. [6]

TensorFlow library is utilized for composing the machine learning application process. It gives high-performance numerical computation. It includes an adaptable design that makes simple deployment of computation over a variety of platforms possible.[7]

Android operating system, built on JetBrains' IntelliJ IDEA software and designed specifically for Android development.[8]

Java programming is utilized for building the back-end of the android application.[9]

1. Object Detection: The application makes use of the TensorFlow model for object detection. It uses a single neural network for the entire input picture. The network at that point separates the input image into a few diverse districts and predicting the bounding regions in form of boxes with their percentage score.
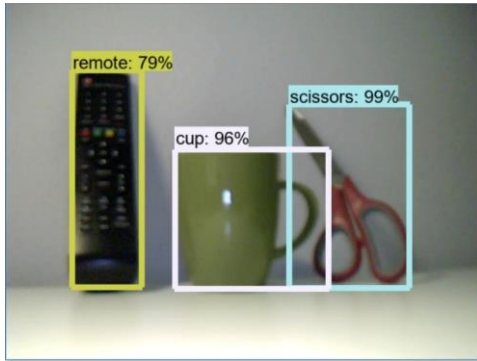

Fig. 4. Object Detection

2. Depth Estimation: Depth estimation or extraction include is nothing but the strategies and calculations which aim to get a representation of the spatial structure of a scene. In easier words, it is utilized to calculate the distance between two objects. Our model is utilized to help the blind individuals which aim to issue warning to the blind individuals around the obstacles coming on their way. To do this, we ought to discover that at how much distance the obstacle and individual are located in any real-time circumstance. After the object is detected, a rectangular box is produced around that object.

Fig. 5. Depth Estimation

3. Voice Assistance: Voice commands are generated as per the output and the application warns the visually impaired person if he/she is too close to the object or gives a generalized statement if they are at a safer distance from the object.

4. Dataset: In this project, Common Object in Context (COCO) dataset is used for training the TensorFlow model which recognizes 80 different categories.

VI. RESULTS AND ANALYSIS

The application home screen will directly open the camera of the device which will give the results according to the object/person/color that is present in front of the camera. It will also work when multiple objects are present in the frame and will give the preferred output of all objects in the frame.
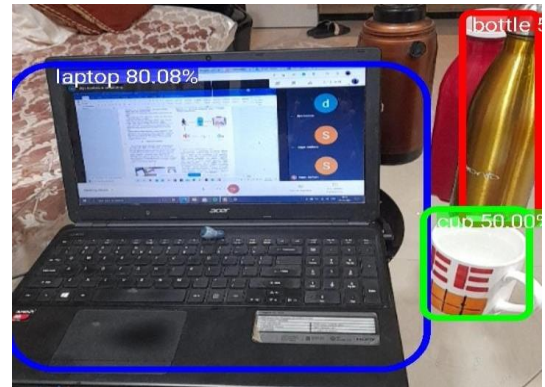

Fig. 6. Output of Multiple Objects

For example, in *Fig. 6.* we can see when multiple objects are captured by the camera, all the objects are detected by the application and then the output is given.

The application gives ease and flexibility within the field of app development. It brings efficiency and also makes a difference in the quick conveyance of the output. The application permits you to partition your work into parts and makes a difference for you to center on the center portion of the app or any framework. This methodology makes a difference in the improvement of the great and quality computer program. We can pass on the Features of the application to the system.

1) To begin with, we are capturing real-time pictures from the rear camera of the mobile the handset of blind individuals and sent to the system for further conclusions.

2) The system will test it using its COCO DATASETS and it detects the confidence accuracy of the image which it is testing. We reached 98% accuracy for certain classes like books, cups, remote, etc.

3)After testing the pictures, we are creating a yield on the framework and its expectation is being interpreted into voice with voice modules and sent to the blind individual with the help of wireless sound bolsters apparatuses.

Fig. 7. Output

VII.   CONCLUSION

An accurate and efficient object recognition system integrated in an Android Application has been developed which achieves comparable metrics with the existing state-of-the-art system. This project uses recent techniques in the field of computer vision. The prototype of the object recognition was successfully implemented using python programming OpenCV library and integrated onto the android application using java programming. However, during the nighttime or when there is no sufficient light, the accuracy of the application would lower accordingly. As per future work we will try to make an application for the IOS platform.

VIII.   ACKNOWLEDGMENT

IX.   REFERENCES

[1] Aditya Raj, "Model for Object Detection using Computer Vision and Machine Learning for Decision Making," *International Journal of Computer Applications,* 2019.

[2] Bhumika Gupta, "Study on Object Detection using Open CV Python," *International Journal of Computer Applications Foundation of Computer Science,* vol. 162, 2017.

[3] Abdul Muhsin M, "Online Blind Assistive System using Object Recognition," *International Research Journal of Innovations in Engineering and Technology,* vol. 3, pp. 47-51, 2019.

[4] A. Gupta, "BLIND ASSISTANCE SYSTEM: REAL TIME OBJECT DETECTION WITH DISTANCE AND VOICE ALERTS," *Medium,* 30 May 2020.

[5] "OpenCV," [Online]. Available: https://opencv.org/ .

[6] "Python programming language," [Online]. Available: https://www.python.org/.

[7] "TensorFlow," [Online]. Available: https://www.tensorflow.org/ .

[8] "Android Studio," [Online]. Available: https://en.wikipedia.org/wiki/Android_Studio..

[9] "JAVA," [Online]. Available: https://www.java.com/en/.