

Real-Time Object Detection System with Voice Feedback for the Blind People



Harshal Shah, Meet Amin, Krish Dadwani, Nishant Desai,
and Aliasgar Chatiwala

Abstract Lots of people suffer from temporary or permanent disabilities, and one of them is blindness or the visually impaired. According to the World Health Organization (WHO), more than 1 billion people suffer from blindness. Technology may assist blind and visually impaired persons in a number of ways; objects detecting is still a difficult process. There are so many techniques available for object detection. But, the accuracy and efficiency of detection aren't good. Therefore, in this paper, authenticate object detection in real time using the YOLO-v3 and deep learning techniques. The main focus of this paper is to create a smartphone application that is cost-efficient with high accuracy.

Keywords YOLO v3 · Object detection · Machine learning · Text to speech · Deep learning · Image processing · Video processing · Blind people · Computer vision · COCO dataset

1 Introduction

According to the WHO in 2021, around, 1 billion people will suffer from high levels of visual loss. Out of those 1 billion people, 640 million had very low vision, and 39 million was totally blind. Visually impaired persons have complicated challenges while executing things that normal people accept as normal, such as reaching for something, recognizing, and identifying items in their environment. Avoiding impediments and recognizing the things in their environment are a problem for them. The goal of this project is to develop APK deep learning and computer vision technologies that are being used to assist blind persons in overcoming obstacles. The gadget should be able to categorize and identify various things in real time, as well as alert the user of the object class via audio feedback. All of the design and layout considerations are made with visually impaired persons in mind. People's comments and thoughts are considered because they are the ones who will utilize this APK.

H. Shah (✉) · M. Amin · K. Dadwani · N. Desai · A. Chatiwala
Parul University, Vadodara, Gujarat, India
e-mail: harshal.shah@paruluniversity.ac.in

1.1 Deep Learning

Deep learning is a subset of machine learning in artificial intelligence (AI) that uses several layers of neural networks to describe human brain capabilities in data processing and pattern formation for important decision-making. Deep learning is an artificial intelligence (AI) function that comprises networks capable of unlabeled data from an unstructured dataset. Deep learning has grown significantly in the digital world, resulting in an avalanche of data in all kinds of formats and from every corner of the world. Big data can be gathered from a variety of online sources including e-commerce platforms, search engines, social media, and online streaming platforms and others. This huge volume of information is easily accessible and may be shared via cloud computing tools. However, the dataset utilized may well be unstructured and of such a large scale that understanding it, and extracting useful information and patterns might take decades for a human. Deep learning is an unsupervised learning technique that does not require data with specific variables and characteristics. Instead, it uses an iterative approach to get the desired outcome. When dealing with large amounts of data, deep learning typically performs admirably. Deep learning uses neural networks to automatically analyze a large dataset for patterns and tiny correlations. The learnt associations are utilized to interpret new knowledge once the model has been trained using the dataset.

1.2 Computer Vision

Computer vision is an associative subject that analyzes how computers can recognize digital pictures or videos (imagery) at a high level. Although computer vision techniques have been available since the 1960s, there has been a significant improvement in how well the software can explore this type of data due to recent advances in machine learning, as well as leaps forward in data storage technologies, much-improved computing capabilities, and cheap high-quality input devices. Any processing using visual material, such as photos, movies, icons, or pixels in general, is referred to as computer vision. However, there are a few important jobs in computer vision that serve as a framework. A model is trained on a dataset of distinct items in object categorization. The item is identified by the trained model as belonging to one or more of the trained classes. Because the model was trained on a dataset of an object, it should be possible to detect a specific instance of the object in object identification.

1.3 Object Detection

Object detection is a computer vision approach for detecting and detecting things in pictures and videos. Object detection may be used to numerical calculations in a scene, determine and monitor their specific location, and effectively label them using this type of identification and localization. Consider the following images: two books, one laptop, and one mobile phone. Object detection helps us to classify the sorts of objects we find while also locating them inside the picture. Bounding boxes are drawn around identified items using object detection. Object detection is strongly connected to other computer vision methods such as image recognition and image segmentation in that it assists in the comprehension and analysis of situations in pictures and video (Fig. 1).

2 Previous Work

There is some previous work already done in the past for helping blind people to help to visualize the surrounding objects with help of object detection techniques. There are some ideas shown below.

- Blind stick navigator
- Smart gloves for the blind
- Talking smart glass blind stick navigator.

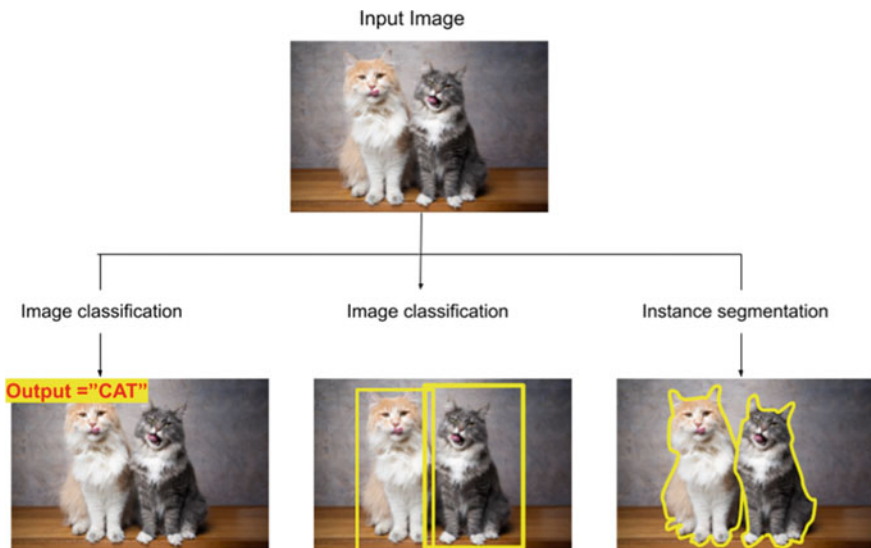


Fig. 1 Object detection

The objective of the present invention is to propose a blind-stick for guiding the blind comprising a location sensor, a micro-controller unit, a temperature sensor, water-level sensor, a heart pulse sensor had a voice module; the present invention relates to the blind stick which includes an electronic sensing mechanism for helping visually impaired people while walking on the road [1]. They have made it smart by having an RF remote which helps to find a stick whenever it is lost.

2.1 Problem and Weaknesses of Current System

- (1) Cost plays an effective role in everything; manufacturing devices include sensors, speakers, and many other tools which increase the cost of the system. Later on, such devices are not affordable for everyone.
- (2) Lifetime support: None of these products gives a lifetime guarantee as some or other day the product needs service. Device can be lost or any issue can come anytime.
- (3) Need to carry the stick or smart-box every time everywhere.
- (4) Low battery can also create an issue at some point.

2.2 Requirements of New System

Blind people or visually impaired people face issues whenever they need to find what they want, or later they need someone's guidance to avoid obstacles or barriers; sometimes, it may happen that guidance is not available; in that case, it becomes difficult for the blind person or their care taker because all of the items and devices mentioned above are expensive and cannot be affordable. To overcome this problem, we are designing an android application that will be available for free for any user. A user should have a smart android phone, and this application will be able to help them in each and every way covering corner things also. This application will work as a guide for blind people, which includes real-time object detection nearby them and informing the user through an assistant and will help them to know the object and the distance between user and object.

3 Objective

The technology uses object detection, and the software will send a message to the user if it finds any possible impediments in the path.

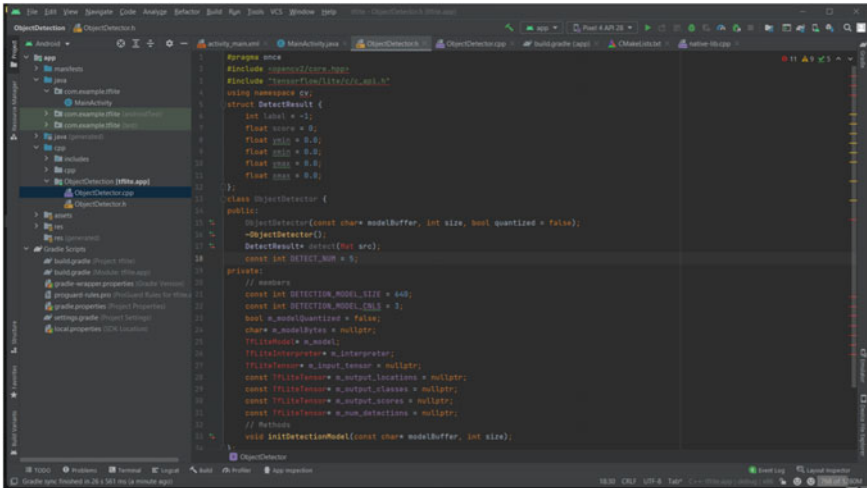


Fig. 2 Object detection code

3.1 Object Detection

For object detection, the tool uses the YOLO-v3 algorithm. For the whole input images, it utilizes a unique neural network. After that the network splits the input images into multiple regions, predicting the bounding regions in the form of boxes with their probability scores (Fig. 2).

3.2 Dataset

Microsoft’s Common Object in Context (COCO) dataset [2] is large-scale object detection, segmentation, and captioning dataset. COCO dataset is implemented in this research to train the YOLO model and text reader, which can detect 91 distinct classes. Features of COCO dataset are shown in below.

- 330 K images (>200 K labeled)
- 1.5 million object instances
- 80 object categories
- 91 stuff categories
- 5 captions per image.

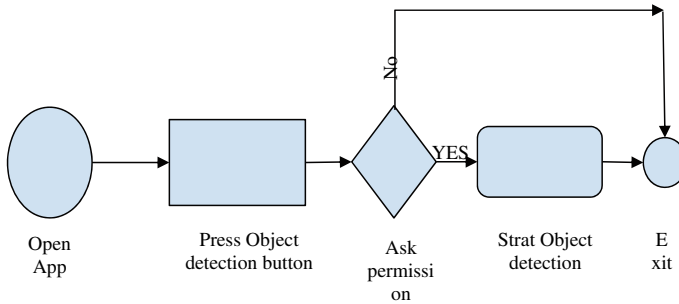


Fig. 3 System work flow

4 Proposed System

The system is built in an android APK that recognizes a variety of objects in real time.

4.1 System Overview

A smartphone is used to capture real-time input data. The APK [3] camera is instantly accessible, and it begins recording the surrounding items, which the application informed them of. To help them learn and navigate their system, the system speaks out about every activity (Fig. 3).

4.2 Implementation

The system is designed to use a combination of technology stacks, which are explained further below. Android Studio is the primary integrated development environment (IDE) built particularly for android application development; it is utilized to develop the app. TensorFlow lite is a lightweight version of TensorFlow that was developed with the mobile operating system and embedded devices in the brain. It gives mobile users a machine learning solution with minimal latency and tiny binary size. TensorFlow includes a collection of fundamental operators that have been optimized for use on mobile devices. Custom operations in models are also supported. The video or image is processed in real time using You Only Look Once (YOLO) [4]. It predicts the observed item in the form of bounding boxes using a single convolutional network that is quicker than the regional convolutional neural network (R-CNN).

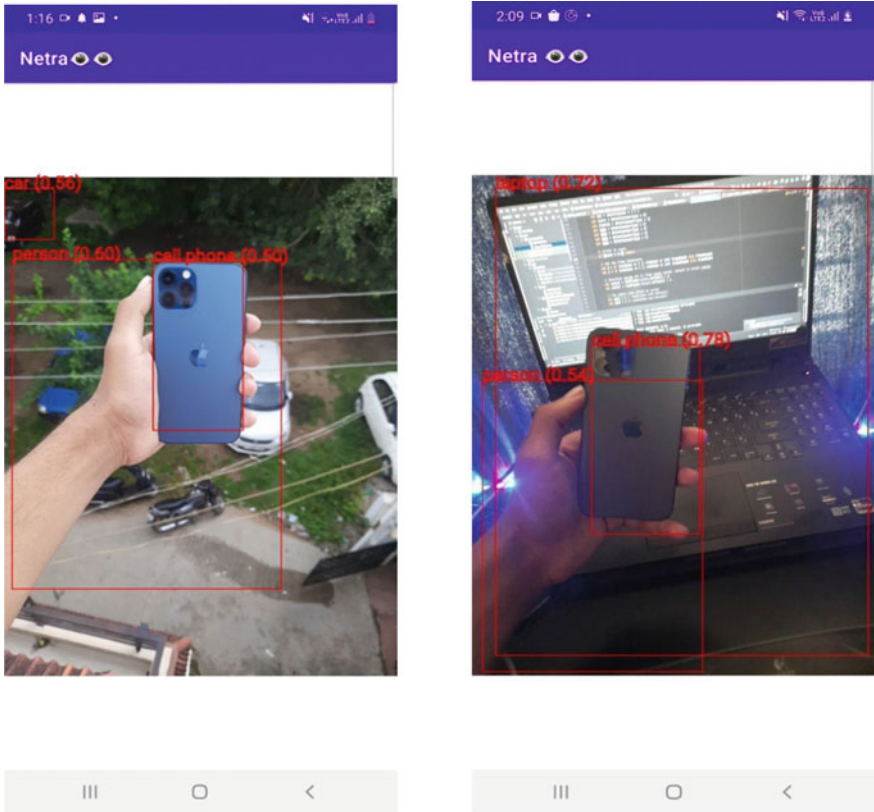


Fig. 4 Results

5 Experimental Results and Analysis

The APK [3] can identify and recognize a variety of obstacles that may be encountered when walking, as well as everyday items, and notify the user through assistance. A user must have a smart android phone, and this APK will be able to assist them in any manner possible, even inside and in non-crowded areas (Fig. 4).

6 Conclusion

The suggested solution would be extremely valuable and would improve the user's experience with their android smartphone by offering them assistance. With the user, auto-assistance will play a critical and significant role. Assistance can also be thought of as the user's third eye. Our major goal is to develop an APK [3] that allows

blind people or visually impaired people to experience their environment simply by listening to the assistance. As a result, it will aid in the prevention of such mishaps. The mobile devices may be conveniently carried, and the device's camera can be utilized to identify an object in the environment and provide an audio output.

References

1. R.R. Mahmood, M.D. Younus, E.A. Khalaf, N. Thirupathi Rao, Real Time Object Detection for Visually Impaired Person. *Annals of R.S.C.B* (2021)
2. Microsoft "COCO Dataset". <http://cocodataset.org/>
3. Android Developer. <https://developer.android.com/>
4. YOLO weights. <https://pjreddie.com/darknet/yolo/>