ABSTRACT

Vision loss is a serious factor that can affect one's course of life and quality of life, making them barred from independence, easy mobility, unnecessary tripping over objects, etc. As per World Health Organization's most recent studies, there are around 2.2 billion people on the planet who are visually impaired, and a major proportion of the population is over the age of 50 years. Population growth and aging are expected to increase the risk that more people acquire vision impairment. However, visual impairment in children can happen due to various factors.

With the advent of technology, a huge part of their day-to-day obstacles can be resolved if not eradicated. A handful amount of devices and facilities have been invented and designed for aiding the visually impaired. The main catch of which is to address the cost-effectiveness and accuracy ratio of the same.

The primary cornerstone of this project is about how a simple Object Detection model, done with the help of some python libraries and APIs, shall prove to be helpful for visually impaired people, and help with those subtle challenges. In our proposed methodology, to overcome the aforementioned drawbacks, we shall implement a Machine Learning Model whose key purpose is to detect the location of the object surrounding the person and provide voice feedback using Google Text-to-Speech API. To train the model around various aspects of the object, the MS-COCO dataset is being used.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

SERIAL NUMBER	ABBREVIATION	EXPLICATION
1	**YOLO**	You Only Look Once
2	**MS COCO**	Microsoft Common Objects in Context
3	**API**	Application Programming Interface
4	**OPENCV**	Open Computer Vision
5	**GTTS**	Google text to speech
6	**NMS**	Non Max Suppression
7	**IEEE**	Institute of Electrical and Electronics Engineering
8	**OCR**	Optical Character Recognition
9	**IC-APR 9600**	Integrated Circuit APR 9600
10	**R-CNN**	Region based Convulated Neural Network
11	**NUMPY**	Numerical Python
12	**GPU**	Graphics Processing Unit
13	**SSD**	Single Stage Detector

CHAPTER 1
INTRODUCTION

1.1 BRIEF

The fourth Industrial revolution that is also known as the technological revolution has almost gasped our industry. The foremost technology that is turning this revolution into reality is Artificial Intelligence, either it is in the form of Machine Learning or Deep Learning. There are many areas in the Machine Learning domain such as Speech Recognition, computer vision, and natural language processing. With the humongous availability of data around the world, it has become comparatively easier to train the algorithm, manipulate it and use it accordingly.

In this project, our prime concept is to simplify the lives of people affected visually by translating the perceptible surroundings around them into audio comments such that they can understand what is going on around them. Visual impairment is a serious cause that segregates people from the basic way of survival in times like these, making them go devoid of various opportunities and general chores that people don't even pay proper heed to. However, with the advent of technology and the technological revolution, the life of visually impaired people can be made comparatively smoother, helping them experience their surroundings in the most optimal process. Thus, for our proposed concept and to reach our required objective, we shall be using the aforementioned domain in ways that shall be explained hereunder.

For the same purpose, we use various Machine Learning libraries, of which OpenCV is a prime one. Open Source Computer Vision Library, generally termed as OpenCV is an open-source library that deals with image manipulation and various instantaneous operations. In this paper, we will be seeing the use of OpenCV for detecting objects. Other than that Google Speech API and a suitable training algorithm are used to get the proper outcome from the model.

1.2 ABOUT THE PROJECT

In this project, our goal is to build a prototype of a model that easily sees the objects around us and translates them in a way that visually impaired people can perceive it with zero to no difficulty. We use areas of machine learning such as speech recognition and computer vision to conduct this project. The dataset used to administer this project is the MS COCO dataset.

The MS COCO dataset is comprehensive object detection, segmentation, and labeling dataset published by Microsoft. COCO stands for Common Objects in Context, as the image data set was created with the aim of advancing image recognition, visual datasets for computer vision, mostly state-of-the-art neural networks. The COCO dataset offers a variety of features including object segmentation with detailed instance annotation, in-context detection, superpixel segmentation to name a few. Out of a total of 300,000 images, 200,000 are tagged. In addition, COCO also provides 80+ object categories called COCO Classes, 91 object categories called COCO Things, and provides 5 captions per image. Also, for the purpose of pose estimation, 250,000 people with 17 different pre-trained key points are generally used.

The project can be segmented into various steps for a better understanding of the work. The partial objective of the project is to detect objects around us and to identify them using the YOLOv3 algorithm that supports Darknet architecture, thus giving a wide spherical view of the object detection and recognition process as a whole. Furthermore, the latter half of the project aims on translating the recognized objects into speech through the Google-Text-to-Speech API of the python library. Combining both of these gives us the primary objective of the prototype.

Thereby, the initial designs of the project prototype are made, thus addressing the required and mentioned objectives.

1.3 PROJECT OUTLINE

This project is developed and structured using various python libraries for working with various domains of Machine Learning, such as speech recognition, computer vision, natural laguage processing. The whole project is made in Jupyter Notebook, which is a very usable platform for projects like such.

The project can be implemented after the successful implementation of OpenCV 3.4+. Since OpenCV 4 is still in beta right now and the official release has not been initiated yet, it is safer to use the version 3 of the OpenCV library. Apart from that, we need to install the YOLOv3 training dataset and MS COCO dataset to start our project implementation.

The whole project can be visualised in tree command of the terminal as 4 directories and 19 files initially. The 4 directories can be further enhanced as

–image : The path to the input image. We shall be detecting objects in this image using YOLO.

–yolo : This is the base path to the YOLO directory. The scripts will load the required YOLO files in order to perform object detection on the image.

–confidence : Minimum probability to filter weak detections is tracked in this directory. The default value is given as 0.5 and the value is open to experimentation.

–threshold : This is our non-maxima supppression (NMS) threshold with a default value of 0.3.

A variety of object detection techniques is used using the YOLOv3 algorithm which shall be explained further and the detected output is translated to audio for audio feedback. Implementing the same on a video based input meets the objective output that we have set to plan out.

CHAPTER 2
LITERATURE SURVEY

2.1 LITERATURE REVIEW

As a part of research for building our model, a literature review of a handful of IEEE published papers were made. We shall briefly be discussing the review hereunder with appropriate required details.

2.1.1 Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3:

The first paper is titled “Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3” and was published in the year 2021. This paper uses algorithms and techniques like the OpenCV library, Yolo, Yolo v3. The performance recorded in this paper indicates that it works better for smaller objects with future works mentioned as the expansion of the research on self explored dataset [1]

2.1.2 Reader and Object Detector for Blind:

The next paper is titled “Reader and Object Detector for Blind” which was published in the year 2020. This paper uses algorithms and techniques such as Raspberry pi, OCR, tesseract, and tensorflow for carrying out the project. Text reading and object detection was successful but not for smaller than 16 font size is what was recorded in the performance of the paper. As the objective for future works, making it available for multi languages is recorded as of now [2].

2.1.3 Obstacle Detection for Visually Impaired Patients:

The paper that was studied next is titled “Obstacle Detection for Visually Impaired Patients” which was published in the year 2014. The techniques and algorithms used in this paper are stereoscopic sonar system, sound buzzers, voice IC-APR 9600. Wearable optical detection system is provided that provides full body vibration effect on obstacle detection. However, the device has a very limited range when compared to its own size and is also found difficult for users to comprehend the guidance signals in time [3].

2.1.4 Voice Based Smart Assistive Device for Visually Challenged:

The paper that was studied next is titled “Voice Based Smart Assistive Device for Visually Challenged” and was published in the year 2020. The Raspberry Pi, Deep Learning, conversational AI, speech recognition, Assistive Technology, and algorithms and methodologies were described in the article. After being trained on only 50 photos of each object, the model has an accuracy of 83 percent and detects campus objects that are commonly available. However, because it was trained on 8000 photos from the Flickr 8k dataset, the accuracy drops as the image complexity grows.[4].

2.1.5 A Wearable Assistive Technology for the Visually Impaired with Door Knob Detection and Real-Time Feedback for Hand-to-Handle Manipulation:

The next paper is titled “A Wearable Assistive Technology for the Visually Impaired with Door Knob Detection and Real-Time Feedback for Hand-to-Handle Manipulation” and was published in the year 2017. Algorithms and techniques such as YOLOv2, Deep Learning, Neural Network were used. The performance of the device is increased to folds if the hand detection is stable. The biggest difficulty, however, is the consistency of the hand detection performance. More images will be added to the database in the future, and the door knob identification feature will be extended to more general door handles. [5].

2.1.6 VISION- Wearable Speech Based Feedback System for the Visually Impaired using Computer Vision:

The paper is titled as “VISION- Wearable Speech Based Feedback System for the Visually Impaired using Computer Vision”, published in 2020. It is a wearable device based on Raspberry pi, gTTs and YOLO. The text will be read out in English and at a slow speed that is saved as an mp3 file and future work is mentioned as location navigation that works in low-light conditions while remaining cost-effective.[6]

2.1.7 YOLO-compact:An Efficient YOLO Network for Single Category Real-time Object Detection:

This IEEE paper named, “YOLO-compact:An Efficient YOLO Network for Single Category Real-time Object Detection”, published in 2020, is an efficient way for a location navigation that works in low-light conditions while remaining cost-effective.. The model would be more precise if the depth, width and precision is improved. [7].

2.1.8 CPU based YOLO: A Real Time Object Detection Algorithm:

The next 2020 published paper having title “CPU based YOLO: A Real Time Object Detection Algorithm” is based on Faster R-CNN, YOLO, R-CNN, Fast R-CNN, SSD, Mask R-CNN, R-FCN, OpenCV and RetinaNet. The Model discovers objects from video at a pace of “10.12–16.29 frames per second” on many non-GPU platforms, with an accuracy of 80–99 percent. mAP of 31.05 percent is achieved by CPU Based YOLO with aforementioned future work as increment of FPS and mAP by optimizing the model.[8].