CHAPTER 1

INTRODUCTION

1.1 PROJECT OVERVIEW

Visually impaired people or in other words especially differently abled people are the ones who face a lot of difficulties even to accomplish their daily routine chores. Most of them even though they don't want, have to rely on other people for some kind of help. There are thousands of technologies being developed or have been developed for the assistance of these people. However, computer vision being one of these technologies is providing the most promising solute. Further, blind people find hard time to navigating around the street. Due to their inability to see world, they are often in danger of getting hit by obstacle and moving vehicles.

Consequently, eyesight is one of the essential human senses, and it plays a significant role in human perception about the surrounding environment. Thus, visually impaired people to be able to provide, experience their vision, imagination mobility is necessary. The International Classification of Diseases (2018) classifies vision impairment into two groups, distance and near presenting vision impairment. Globally, the leading causes of vision impairment are uncorrected refractive errors, cataract, age-related macular degeneration, glaucoma, diabetic retinopathy, corneal opacity, trachoma, and eye injuries. It limits visually impaired ability to navigate, perform everyday tasks, and affect their quality of life and ability to interact with the surrounding world upon unaided. Additionally, with the advancement in technologies, diverse solutions have been introduced such, as the Eye- ring project, the text recognition system, the hand gesture, and face recognition system, etc. However, these solutions have disadvantages such as heavyweight, expensive, less robustness, low acceptance, etc. hence, advanced techniques must evolve to help them. So, we propose a system built on the breakthrough of image processing and machine learning. The proposed system captures real-time images, then images are pre-processed, their background and foreground are separated and then the Deep Neural Network module with a pre-trained model is applied resulting in feature extraction. The extracted features are matched with known object features to identify the objects. Once the object is successfully recognized, the object name is stated as voice output with the help of text-to-speech conversion. The key contributions of the report include:

- Robust and efficient object detection and recognition for visually impaired people to independently access familiar and unfamiliar environments and avoid dangers.
- Offline text-to-speech conversion and speech output.

1.2 CHALLENGES AND MOTIVATION

This innovative blind assist system aims to empower individuals with visual impairments by seamlessly integrating cutting-edge computer vision and image processing technologies. By harnessing the capabilities of a pre-trained object detection model on the COCO dataset, the system not only identifies and narrates surrounding objects but also offers a unique book reading feature. This feature, supported by optical character recognition and text-to-speech technologies, allows users to explore and engage with written content independently. The project's motivation lies in fostering inclusivity, independence, and accessibility, providing a valuable tool for individuals with visual challenges to navigate and access information in their daily lives.

1.3 OBJECTIVES

- To build device which will be able to capture image of a book and identify the text, convert and play in audio format.
- The device should be able to detect the object using camera and play the object name in speaker
- Providing real-time object recognition and navigation assistance.
- Delivering auditory and tactile feedback for enhanced interaction.
- Offering accessible interfaces and customizable features.
- Improving independence, safety and quality of life.
- Develop a comprehensive understanding of the needs, preferences, and challenges of visually impaired and blind individuals through user-centered research.
- Design and implement hardware, software components that enable real-time assistance, object recognition, and navigation for visually impaired users.
- Integrate cutting-edge technologies, including computer vision and natural language processing, to enhance the functionality and accessibility of the developed solutions.

- Conduct rigorous testing and validation to evaluate the performance, reliability, and usability of the developed techniques in real-world scenarios.

- Address ethical and legal considerations related to privacy, data security, and accessibility rights in the development and deployment of the enhanced techniques.

1.4 SCOPE OF THE PROJECT

Here, an attempt to build device which will be able to capture image of a book and identify the text and convert and play in audio format. However, the device should be able to detect the object using camera and play the object name in speaker.

Thus, effort to evolve solution to the problem is to create a device which can recognize obstacle using camera and also say whether the traffic light is red or green, using voice alert. The system employees small compact arm computer raspberry pi, and the system is battery powered. The design of module is small in terms of implementation, also easy to carry.

This system will continuously record video of the surrounding and will convert it into frames. After analysing these frames, the system will alert the person about some obstacle or the surrounding. The main advantages are the portable, affordable and accessible system using image processing technologies is able to help visually impaired people. This system will help the visually impaired people to navigate their way through any obstacle and will give them a sense of visualization of world around them.

Technology Integration: Integrating computer vision, machine learning, and natural language processing technologies to develop innovative solutions tailored to the needs of visually impaired individuals.

- **Hardware and Software Development:** Designing and implementing both hardware components (e.g., cameras, sensors, wearable devices) and software applications (e.g., mobile apps, desktop software) that facilitate real-time assistance and navigation for visually impaired users.
- **Object Recognition and Navigation:** Developing capabilities for object recognition and navigation, enabling users to identify and interact with objects in their surroundings, navigate indoor and outdoor environments safely, and access relevant information about their surroundings.

- **Accessibility Features:** Incorporating a range of accessibility features, such as voice commands, auditory feedback, haptic feedback, and customizable user interfaces, to ensure effective interaction with the technology by visually impaired users.
- **User-Centered Design:** Employing a user-centered design approach, involving visually impaired individuals in the development process to understand their needs, preferences, and challenges better and to create solutions that are intuitive, inclusive, and empowering.
- **Testing and Validation:** Conducting rigorous testing and validation processes to evaluate the performance, reliability, and usability of the developed solutions in real-world scenarios, ensuring that they meet the needs and expectations of visually impaired users.

CHAPTER 2

LITERATURE SURVEY

1. Assisting Blind People Using Object Detection with Vocal Feedback

Publisher: IEEE

Heba Najm; Khirallah Elferjani; Alhaam Alariyibi

For visually impaired people, it is highly difficult to make independent movement and safely move in both indoors and outdoors environment. Furthermore, these physically and visually challenges prevent them from in day-to- day live activities. Similarly, they have problem perceiving objects of surrounding environment that may pose a risk to them. The proposed approach suggests detection of objects in real-time video by using a web camera, for the object identification, process. You Look Only Once (YOLO) model is utilized which is CNN-based real-time object detection technique. Additionally, The OpenCV libraries of Python is used to implement the software program as well as deep learning process is performed. Image recognition results are transferred to the visually impaired users in audible form by means of Google text-to-speech library and determine object location relative to its position in the screen. The obtaining result was evaluated by using the mean Average Precision (mAP), and it was found that the devised approach achieves excellent results when it compared to previous approaches.

Published in: 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)

https://ieeexplore.ieee.org/document/9837737

2. Robot Eye: Automatic Object Detection and Recognition Using Deep Attention Network to Assist Blind People

Publisher: IEEE

Ervin Yohannes; Paul Lin; Chih-Yang Lin; Timothy K. Shih

For visually impaired people, detection and recognition is a well-known topic in computer vision that still faces many unresolved issues. One of the main contributions of this research is a method to guide blind people around an outdoor environment with the assistance of a ZED stereo camera, a camera that can calculate depth information. In this paper, Author erected a deep attention network to automatically detect and recognize objects. The objects are not only limited to general people or cars, but include convenience stores and traffic lights as well, in order to help blind people cross a road and make purchases in a store. Since public datasets are limited, author also create a novel dataset with images captured by the ZED stereo camera and collected from Google Street View. When testing with images of different resolutions, author's method achieves an accuracy rate of about 81%, which is better than naive YOLO v3.

Published in: 2020 International Conference on Pervasive Artificial Intelligence (ICPAI)

https://ieeexplore.ieee.org/document/9302726

3. Visual Assistance for Blind Using Image Processing

Publisher: IEEE

B Deepthi Jain; Shwetha M Thakur; K V Suresh

An author has considered visually impaired people face lot of difficulties in their daily life. Many a times they rely on others for help. Several technologies for assistance of visually impaired people have been developed. Among the various technologies being utilized to assist the blind, Computer Vision based solutions are emerging as one of the most promising options due to their affordability and accessibility. This paper evolved a system for visually impaired people. The erected system aims to create a wearable visual aid for visually impaired people in which speech commands are accepted from the user. Its functionality addresses identification of objects and sign boards. This will help the visually impaired person to manage day-to-day activities and to navigate through people surroundings. Raspberry Pi enables the implementation of artificial vision with Python on the Open CV platform.

Published in: 2018 International Conference on Communication and Signal Processing (ICCSP)

https://ieeexplore.ieee.org/document/8524251

4. Design and Implementation of Obstacle Detection and Warning System for Visually Impaired People

Publisher: IEEE

Yusuf Sahabi Lolo; Kelechi Lawrence Ohammah; Amina Nna Alfa; Sadiq Abubakar Mohammed;

More recently, environmental information assists human beings to learn about the source that surround them, most visually impaired people make extensive use of the auditory environment not just to determine the presence of an obstacle, but also to successfully makeover around it. This paper discusses various methods for improving blind people's navigation by utilizing readily available technologies. The system includes the ability to detect obstacles for collision avoidance, as well as the ability to detect objects in up, down, and front directions using ultrasonic sensor. The other sensor detects water on the ground and is located near the bottom tip of the walking cane. The system's whole operation is controlled by a microcontroller-based circuit. In the case of a crisis or loss, the technology also allows the blind person to send an SMS message with his or her GPS position to the caretaker or family. These sensors are critical in detecting objects in all directions, allowing blind persons to be self- sufficient.

Published in: 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)

https://ieeexplore.ieee.org/document/9803138

5. Visual Assistance for Blind Using Image Processing

Publisher: IEEE

B Deepthi Jain; Shwetha M Thakur; K V Suresh

The recent observation of researchers, on visually impaired people face lot of difficulties in their daily life. Many a times they rely on others for help. Several technologies for assistance of visually impaired people have been developed. Among the various technologies being utilized to assist the blind, Computer Vision based solutions are emerging as one of the most promising options due to their affordability and accessibility. This paper emerges a system for visually impaired people. The system aims to create a wearable visual aid for visually impaired people in which speech commands are accepted from the user. Its functionality addresses identification of objects and sign boards. This will help the visually impaired person to manage day-to-day activities and to navigate through surroundings. Raspberry Pi is employed to implement artificial vision using python language on the Open CV platform.

Published in: 2018 International Conference on Communication and Signal Processing (ICCSP)

https://ieeexplore.ieee.org/document/8524251

6. Smart Assistive System for Visually Impaired People Obstruction Avoidance Through Object Detection and Classification

Publisher: IEEE

Usman Masud; Tareq Saeed; Hunida M. Malaikah; Fezan Ul Islam; Ghulam Abbas

Recent progress in innovation is making the life prosper, simpler and easier for common individual. The World Health Organization (WHO) statistics indicate that a large amount of people experience visual losses, because of which they encounter many difficulties in everyday jobs. Hence, author goal is to structure a modest, secure, wearable, and versatile framework for visually impaired to help them in their daily routines. Further, the plan is to make an effective system which will assist visually impaired people through obstacle detection and scenes classification. The methodology utilizes Raspberry-Pi 4B, Camera, Ultrasonic Sensor and Arduino, mounted on the stick of the individual. We take pictures of the scene and afterwards pre-process these pictures with the help of Viola Jones and TensorFlow Object Detection algorithm. The said techniques are used to detect objects. We also used an ultrasonic sensor mounted on a servomotor to measure the distance between the blind person and obstacles. The presented research utilizes simple calculations for its execution, and detects the obstructions with a notably high efficiency. When contrasted with different frameworks, this framework is a minimal effort, convenient, and simple to wear.

Published in: IEEE Access (Volume: 10)

https://ieeexplore.ieee.org/document/9691323

7. Design and Implementation of Obstacle Detection and Warning System for Visually Impaired People

Publisher: IEEE

Yusuf Sahabi Lolo; Kelechi Lawrence Ohammah; Amina Nna Alfa; Sadiq Abubakar Mohammed;

More recently, environmental information assists human beings to learn about the source that surround them, most visually impaired people make extensive use of the auditory environment not just to determine the presence of an obstacle, but also to successfully makeover around it. This paper discusses various methods for improving blind people’s navigation by utilizing readily available technologies. The system includes the ability to detect obstacles for collision avoidance, as well as the ability to detect objects in up, down, and front directions using ultrasonic sensor. The other sensor detects water on the ground and is located near the bottom tip of the walking cane. The system's whole operation is controlled by a microcontroller-based circuit. In the case of a crisis or loss, the technology also allows the blind person to send an SMS message with GPS position to the caretaker or family. These sensors are critical in detecting objects in all directions, allowing blind persons to be self- sufficient.

Published in: 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)

https://ieeexplore.ieee.org/document/9803138

CHAPTER 3

SYSTEM REQUIREMENT SPECIFICATION

3.1 FUNCTIONAL REQUIREMENTS

Functional requirements for a system that utilizes machine learning to assist visually impaired and blind individuals:

- Image Recognition: The system should be able to recognize objects, text, and scenes from images captured by a camera or provided as input.
- Object Detection: It should identify and locate common objects in the environment, such as chairs, doors, and stairs.
- Text-to-Speech Conversion: Convert text detected in images or from input text into spoken language for auditory feedback to the user.
- Voice Command Recognition: Enable users to interact with the system through voice commands for tasks such as navigation, information retrieval, and system control.
- Navigation Assistance: Provide turn-by-turn navigation instructions to help users navigate indoor and outdoor environments safely.
- Obstacle Detection and Warning: Alert users about obstacles in their path through auditory or haptic feedback to prevent collisions.
- Gesture Recognition: Recognize hand gestures or other physical movements to trigger specific actions or commands.
- User Personalization: Allow users to customize preferences such as speech rate, volume, and navigation settings to suit individual needs.

3.2 NON-FUNCTIONAL REQUIREMENTS

Non-Functional requirements for a system that utilizes machine learning to assist visually impaired and blind individuals:

- Accuracy: The system should have high accuracy in recognizing objects, text, and commands to ensure reliable assistance for visually impaired users.
- Real-Time Processing: The system should process inputs and provide feedback with minimal delay to support real-time interaction and navigation.

- Accessibility: Ensure the user interface is accessible and intuitive for individuals with visual impairments, including support for screen readers and alternative input methods.
- Robustness: The system should be resilient to variations in lighting conditions, environmental noise, and object occlusions to maintain performance in diverse settings.
- Privacy and Security: Safeguard user privacy by securely handling personal data and ensure that the system does not capture or transmit sensitive information without user consent.
- Scalability: Design the system to accommodate a large number of users and potential future expansions in functionality or user base.
- Compatibility: Ensure compatibility with a variety of devices and platforms commonly used by visually impaired individuals, such as smartphones, tablets, and wearable devices.
- Reliability: The system should operate reliably under different conditions and handle errors gracefully to minimize disruptions for users relying on its assistance.

3.3 HARDWARE REQUIREMENTS

System : Intel i5

RAM : 8 GB

Input devices : Mouse , Keyboard and Laptop camera

Hard disk : 1TB

3.4 SOFTWARE REQUIREMENTS

Operating system : 32 bit or 64 bit Microsoft Windows

Programming language : Python

Tool : Open CV

CHAPTER 4

SOFTWARE DESCRIPTIONS

4.1 Machine Learning

Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can effectively generalize and thus perform tasks without explicit instructions. Recently, generative artificial neural networks have been able to surpass many previous approaches in performance. Machine learning approaches have been applied to large language models, computer vision, speech recognition, email filtering, agriculture and medicine, where it is too costly to develop algorithms to perform the needed tasks.

The mathematical foundations of ML are provided by mathematical optimization (mathematical programming) methods. Data mining is a related (parallel) field of study, focusing on exploratory data analysis through unsupervised learning.

ML is known in its application across business problems under the name predictive analytics. Although not all machine learning is statistically based, computational statistics is an important source of the field's methods.

4.2 OPEN CV

OpenCV (Open-Source Computer Vision Library) is an open-source computer vision and machine learning software library. OpenCV was developed to provide a common infrastructure for computer vision applications and to accelerate the adoption of machine perception in commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 18 million. The library is used extensively in companies, research groups and by governmental bodies.

Along with well-established companies like Google, Yahoo, Microsoft, Intel, IBM, Sony, Honda, Toyota that employ the library, there are many startups such as Applied Minds, VideoSurf, and Zeitera, that make extensive use of OpenCV. OpenCV's deployed uses span the range from stitching streetview images together, detecting intrusions in surveillance video in Israel, monitoring mine equipment in China, helping robots navigate and pick up objects at Willow Garage, detection of swimming pool drowning accidents in Europe, running interactive art in Spain and New York, checking runways for debris in Turkey, inspecting labels on products in factories around the world on to rapid face detection in Japan.

It has C++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV leans mostly towards real-time vision applications and takes advantage of MMX and SSE instructions when available. A full-featured CUDA and OpenCL interfaces are being actively developed right now. There are over 500 algorithms and about 10 times as many functions that compose or support those algorithms. OpenCV is written natively in C++ and has a templated interface that works seamlessly with STL containers.

4.3 NUMPY

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object.
- Sophisticated (broadcasting) functions.
- Tools for integrating C/C++ and Fortran code.
- Useful linear algebra, Fourier transform, and random number capabilities.

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined using Numpy which allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

NumPy, which stands for Numerical Python, is a powerful Python library used for numerical computing. It provides support for arrays, matrices, and a wide range of mathematical functions to operate on these arrays efficiently. NumPy is a fundamental package for scientific computing in Python and is extensively used in various fields such as machine learning, data analysis, engineering, and scientific research.

NumPy targets the CPython reference implementation of Python, which is a non-optimizing bytecode interpreter. Mathematical algorithms written for this version of Python often run much slower than compiled equivalents due to the absence of compiler optimization. NumPy addresses the slowness problem partly by providing multidimensional arrays and functions and operators that operate efficiently on arrays; using these requires rewriting some code, mostly inner loops, using NumPy.

Using NumPy in Python gives functionality comparable to MATLAB since they are both interpreted, and they both allow the user to write fast programs as long as most operations work on arrays or matrices instead of scalars. In comparison, MATLAB boasts a large number of additional toolboxes, notably Simulink, whereas NumPy is intrinsically integrated with Python, a more modern and complete programming language. Moreover, complementary Python packages are available; SciPy is a library that adds more MATLAB-like functionality and Matplotlib is a plotting package that provides MATLAB-like plotting functionality. Although matlab can perform sparse matrix operations, numpy alone cannot perform such operations and requires the use of the scipy.sparse library. Internally, both MATLAB and NumPy rely on BLAS and LAPACK for efficient linear algebra computations.

Python bindings of the widely used computer vision library OpenCV utilize NumPy arrays to store and operate on data. Since images with multiple channels are simply represented as three-dimensional arrays, indexing, slicing or masking with other arrays are very efficient ways to access specific pixels of an image. The NumPy array as universal data structure in OpenCV for images, extracted feature points, filter kernels and many more vastly simplifies the programming workflow and debugging. Importantly, many NumPy operations release the global interpreter lock, which allows for multithreaded processing. NumPy also provides a C API, which allows Python code to interoperate with external libraries written in low-level languages.

4.4 PYCHARM

PyCharm is an integrated development environment (IDE) used for programming in Python. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems, and supports web development with Django. PyCharm is developed by the Czech company JetBrains.

It is cross-platform, working on Microsoft Windows, macOS, and Linux. PyCharm has a Professional Edition, released under a proprietary license and a Community Edition released under the Apache License. PyCharm Community Edition is less extensive than the Professional Edition.

- Coding assistance and analysis, with code completion, syntax and error highlighting, linter integration, and quick fixes
- Project and code navigation: specialized project views, file structure views and quick jumping between files, classes, methods and usages
- Python code refactoring: including rename, extract method, introduce variable, introduce constant, pull up, push down and others
- Support for web frameworks: Django, web2py and Flask
- Integrated Python debugger
- Integrated unit testing, with line-by-line coverage
- Google App Engine Python development
- Version control integration: Unified user interface for Mercurial, Git, Subversion, Perforce and CVS with change lists and merge.
- Scientific tools integration: integrates with IPython Notebook, has an interactive Python console, and supports Anaconda as well as multiple scientific packages including Matplotlib and NumPy.

4.5 SOLID STATE DRIVES(SSD)

The architecture of SSDs (Solid State Drives) is intricate and crucial for the functionality of a blind assistant system, particularly considering the specific needs of users with visual impairments. At the heart of SSD architecture lies flash memory, which serves as the primary storage medium. Unlike traditional hard disk drives (HDDs) that rely on spinning magnetic disks, SSDs store data in interconnected memory chips. This flash memory technology ensures that data is retained even when the power is turned off, making it ideal for storing critical system components such as the operating system, applications, and user data in a blind assistant system. The absence of moving parts in SSDs not only enhances reliability but also contributes to faster data access and transfer rates, essential for real-time processing and response, which are crucial for assisting individuals with visual impairments in navigating their surroundings and accessing information efficiently.

A fundamental component of SSD architecture is the controller, which acts as the brain of the SSD. The controller manages the flow of data between the host system and the flash memory chips, optimizing performance and ensuring efficient operation. In the context of a blind assistant system, an efficient controller is essential for minimizing latency and maximizing throughput, enabling seamless interaction with the user and facilitating quick responses to user inputs. Advanced controllers may incorporate features such as wear-leveling algorithms to evenly distribute data writes across the flash memory cells, prolonging the lifespan of the SSD and maintaining consistent performance over time.

NAND flash memory, the type of flash memory used in SSDs, is organized into a grid of cells, with each cell capable of storing multiple bits of data. This architecture enables SSDs to achieve high storage densities while maintaining fast access times. In a blind assistant system, NAND flash memory provides the storage capacity needed to store large datasets, including maps, navigation instructions, and user preferences. Additionally, the parallelism inherent in NAND flash memory allows SSDs to perform multiple data operations simultaneously, supporting multitasking capabilities essential for assisting individuals with visual impairments in performing various tasks concurrently, such as speech recognition, route planning, and accessing contextual information.

SSD architecture also incorporates features designed to enhance reliability and endurance, crucial considerations for blind assistant systems where system downtime or data loss can have significant consequences. Trim support, for example, is a feature that helps maintain SSD performance by marking data blocks that are no longer in use, allowing the SSD to reclaim storage space and optimize performance. Furthermore, SSDs typically have higher endurance ratings compared to HDDs, meaning they can withstand a greater number of read and write cycles before experiencing failures. This increased durability ensures that blind assistant systems can reliably serve users over an extended period without degradation in performance or reliability.

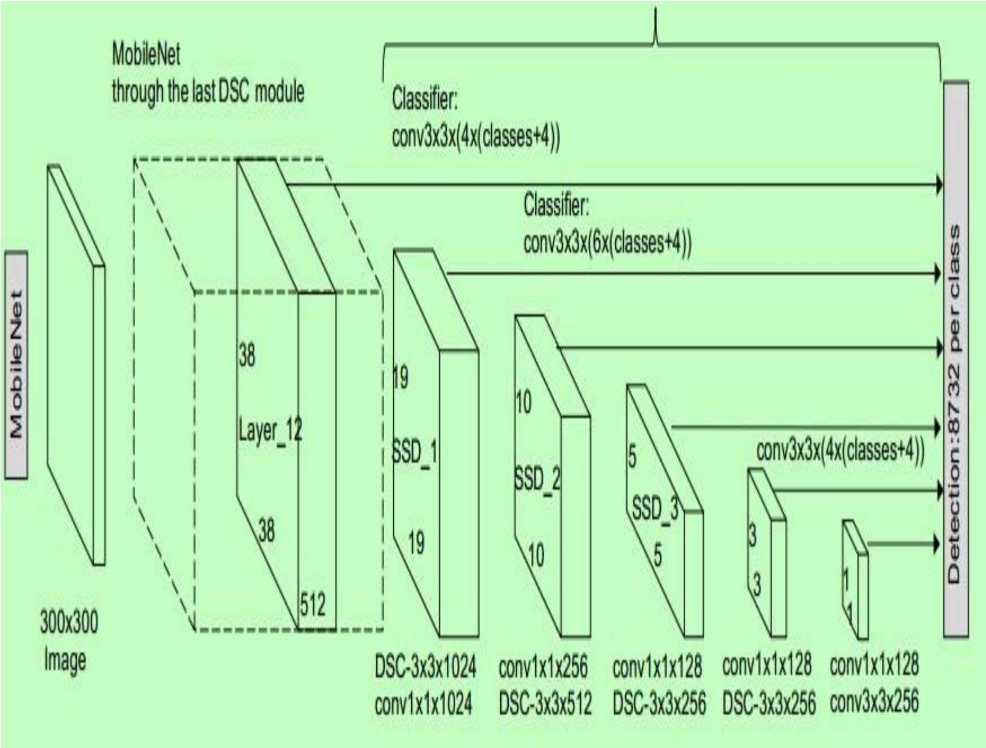

Fig. 4.5: Architecture of SSD

The figure 4.5 shows architecture od SSD. The SSD consists of two parts: an SSD head and a backbone model. As a feature extractor, the backbone model is essentially a trained image classification network. This is often a network trained on ImageNet that has had the final fully linked classification layer removed, similar to ResNet.

The SSD head is just one or more convolutional layers added to the backbone, with the outputs read as bounding boxes and classifications of objects in the spatial position of the final layer activations.

As a result, we have a deep neural network that can extract semantic meaning from an input image while keeping its spatial structure, although at a lesser resolution.

In ResNet34, the backbone produces 256 7x7 feature maps for an input picture. SSD divides the image into grid cells, with each grid cell being in charge of detecting things in that region. Detecting objects entails anticipating an object's class and placement inside a given region.

4.6 MOBILENET

This model is based on the MobileNet model's idea of depth wise separable convolutions and generates a factorised Convolutions. The depth wise convolutions are created by converting a basic conventional convolution into a depth wise convolution. Pointwise convolutions are another name for these 1 * 1 convolutions. These depth wise convolutions apply a general single filter-based notion to each of the input channels for MobileNets to work. These pointwise convolutions use 1 * 1 convolution to combine the depth wise convolutions' outputs. Both filters, like a typical convolution, combine the inputs into a new set of outputs in a single step. The depth wise identifiable convolutions partition this into two layers. One for filtering and the other for mixing. This approach of factorization has the effect of dramatically lowering computing time and model size.

4.7 VOICE GENERATION MODULE

Following the detection of an object, it is critical to inform the person on his or her way of the presence of that object. PYTTSX3 is a crucial component of the voice generation module. Pyttsx3 is a Python conversion module for converting text to speech. This library is compatible with Python 2 and 3. Pyttsx3 is a simple tool for converting text to speech.

Further, the technique works in the following way: every time an item is identified, an approximate distance is calculated, and the texts are displayed on the screen using the cv2 library and the cv2.putText() function. We utilise Python-tesseract for character recognition to find buried text in an image. OCR recognises text content on images and encodes it in a format that a computer can understand. The text is detected by scanning and analysing the image. As a result, Python tesseract recognises and "reads" text encoded in images. These texts are also linked to a pyttsx.

As an output, audio commands are generated. "Warning: The object (class of object) is too close to you," it says if the thing is too close. And if the object is at a safe distance, a voice is generated that states, "The object is at a safe distance." This is accomplished using libraries like as pytorch, pyttsx3, pytesseract, and engine.io.

Pytorch is a machine learning library first and foremost. Pytorch is primarily used in the audio field. Pytorch aids with the loading of the voice file in mp3 format.

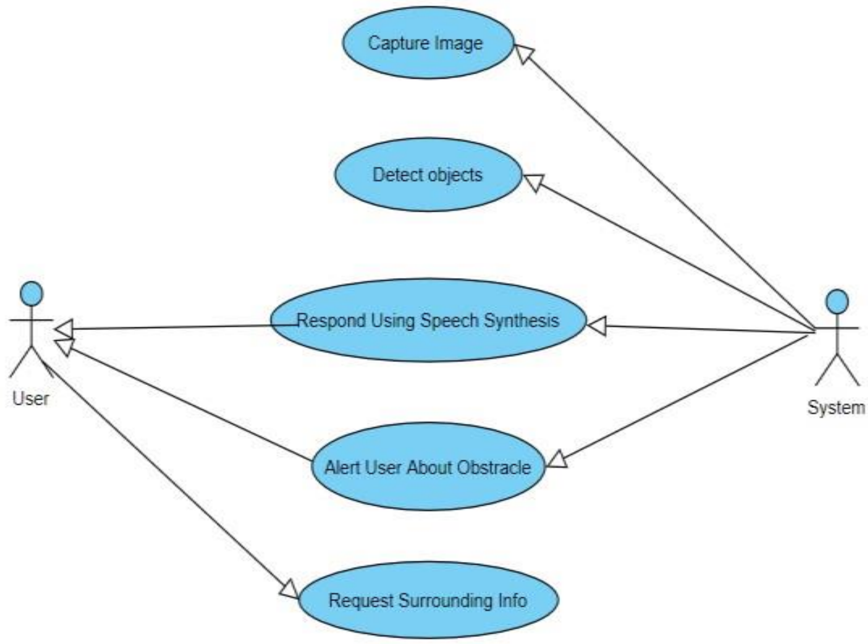

Fig. 4.7: Use Case Diagram

In Fig.4.7 there are two-character, user and system. User will send surrounding information system. And then system will capture the image, detect the object & respond using a speech synthesis. Moreover, in last system will give an alert to user about the particular obstacles.

4.8 DEEP NEURAL NETWORKS(DNN)

Deep Neural Networks (DNNs) have emerged as a transformative technology in assisting visually impaired and blind individuals, providing enhanced capabilities for navigating and understanding their surroundings. DNNs, a subset of machine learning, are composed of layers of interconnected neurons that process and learn from vast amounts of data. These networks can recognize patterns, make predictions, and perform complex tasks with high accuracy, making them ideal for assistive technologies.

- **Object Detection and Recognition:**

One of the primary applications of DNNs in this field is object detection and recognition. Advanced DNN models, such as Convolutional Neural Networks (CNNs), can analyze real-time video feeds from wearable cameras or smartphones to identify objects, people, and obstacles. This information can be conveyed to the user through auditory or haptic feedback, enabling them to navigate spaces safely and independently.

- **Text and Scene Understanding:**

DNNs excel at text recognition and scene understanding. Optical Character Recognition (OCR) technology, powered by DNNs, can convert written text into spoken words, aiding in reading printed materials or signs. Furthermore, Scene Text Recognition (STR) can read and interpret text in complex environments, such as street signs or product labels, providing contextual information to the user.

- **Facial Recognition and Social Interaction:**

Facial recognition capabilities enabled by DNNs can identify individuals and their emotions, assisting blind users in social interactions. This technology can describe the facial expressions and identities of people in the user's vicinity, enhancing social engagement and awareness.

- **Integration with Mobile Devices:**

Mobile applications leveraging DNNs offer portable and convenient solutions for the visually impaired. These apps can integrate various functionalities such as object detection, text reading, and facial recognition, providing comprehensive assistance through a single device.

- **Future Prospects:**

The future of DNNs in assisting the visually impaired looks promising with ongoing research and development. However, advances in computational power, data availability, and model optimization will further enhance the capabilities and accessibility of these technologies, making them an indispensable tool for improving the quality of life for visually impaired and blind individuals.

CHAPTER 5

SYSTEM ARCHITECTURE

5.1 SYSTEM ARCHITECTURE

1. User Interface (UI):

a) Input Interface:

Camera input for object detection and book reading, along with user controls for initiating these features, are included.

b) Output Interface:

Text-to-speech output provides information to the user, while audio feedback conveys details about detected objects and read text.

2. Object Detection Module:

a) Pre-trained Model:

Faster R-CNN, YOLO, or SSD trained on the COCO dataset, along with custom training for specific object classes such as books, are utilized.

b) Real-time Object Detection:

Integration with the camera input enables the identification and classification of objects in the user's environment.

3. Book Reading Module:

a) Text Extraction:

The OCR module, such as Tesseract, is employed for extracting text from images, with image preprocessing applied to enhance text recognition.

b) Text-to-Speech (TTS):

The TTS engine, like Google Text-to-Speech, is utilized to convert extracted text into speech, allowing adjustments to speech rate and pitch for an improved user experience.

4. Image Processing Module:

- Enhance image quality for better object detection and text extraction.
- Techniques include contrast adjustment, noise reduction, and sharpening.

5. Book Detection Module:

- Utilizes the object detection module to specifically identify books.
- Triggered when the user wants to initiate the book reading feature.

6. Control and Integration Layer:

- Orchestrates communication between different modules.

- Determines when to activate object detection or book reading based on user input.

7. **Feedback and Notification:**

- Audio feedback through TTS for detected objects and read text.
- Notifications or alerts for important information.

8. **Accessibility Features:**

- Voice commands for hands-free control.
- Consideration for alternative input methods (e.g., gestures, buttons).

9. **Security and Privacy:**

- Ensure secure handling of user data.
- Implement mechanisms to protect user privacy during image capture and processing.

10. **Testing and Debugging Tools:**

- Tools for testing the system's functionality, accuracy, and performance.
- Debugging tools for diagnosing issues in real-time processing.

5.2 EXISTING SYSTEM

There are lots of strategies or ways that have been adopted by visually impaired people to address this hassle of theirs. A traditional approach that has been used for years by the visually impaired humans is using dogs that could help so as to navigate through their paths or using walking canes to keep themselves away from any obstacles. Both of them are inexpensive or reachable but aren't error prone. Being error prone is what is wanted for blind people as even the slightest of blunders can motive a large damage. Another manner to cope with this problem is to provide blind people with clever rehabilitative shoes alongside the spectacles. Each such shoe is surmounted with ultrasonic transducers to detect objects at unique level of heights and spectacles have a pair of ultrasonic transducers mounted centrally stored above the bridge and with a buzzer at one of the ends. A major drawback of this sensor primarily based approach is that it is just beneficial to detect items in place of recognizing them and hence image processing gives a promising answer to address such situations. One more traditional method use stick to find obstacle, these sticks are exercised to find the optical Infront. The individual becomes aware of the obstacle in front of them when the stick makes contact with it.

5.3 DRAWBACK OF THE EXISTING SYSTEM

These blind sticks are useless when it comes to situation like crossing traffic lights as sticks can’t detect weather the light is red or green. Furthermore, with the stick, obstacle detection occurs only upon contact with the object, and the blind stick lacks the capability to differentiate between different types of obstacles.

5.4 PROBLEM STATEMENT

There are lots of strategies or ways that have been adopted by visually impaired people to address this hassle of theirs. A traditional approach that has been relied upon for years by visually impaired humans is employing dogs that could assist in navigating through their paths. One more traditional method employs stick to find obstacle, these sticks are used to find the obstacle in front. The individual becomes aware of the obstacle in front of them when the stick makes contact with it. But it cannot detect without object contact with the obstacle which sometimes is dangerous since we have to detect object bit farther away. Also, blind people cannot read normal books since they cannot see, they can only understand special books which are designed to blind but the problem is not every book is designed for blind people.

5.5 PROPOSED SOLUTION

Our solution to the problem is to create a device which can recognise obstacle using web camera and gives output using voice alert. The system uses web camera and the system speaker.

This system will continuously record video of the surrounding and will convert it into frames. After analysing these frames, the system will alert the person about some obstacle or the surrounding. The main advantages are the portable, affordable and accessible system using image processing technologies is able to help visually impaired and blind people.

This system will help the visually impaired and blind people to navigate their way through any obstacle and will give them a sense of visualisation of world around them.

CHAPTER 6

METHODOLOGY

6.1 PROPOSED METHODOLOGY

Project consists of two functionalities:

1. **Object detection and audio output :** for detecting the object we have camera attached from which live video acquisition will be taken then that video will be converted into frames of images which is then used to detect the object, for which we have pre-trained dataset called Company Owned Company Operated (COCO) dataset in which 91 object have been trained and by using that dataset we will compare the object obtained from the camera and then make the detection of the object, after detection the label of the object will be obtained and then we have ESPEAK library to make the audio conversion of the labeled object.
2. **Text recognition and audio conversion:** Live video acquisition via camera initiates the first step. The video will be converted into frames of images, which will then be utilized to detect text within the image. The image is then pre-processed to achieve required resolution using open cv library. Preprocesses image will be sent to Convolution Neural Network based text detector tesseract library which detects text and then we have ESPEAK library to make the audio conversion and announces using speaker.

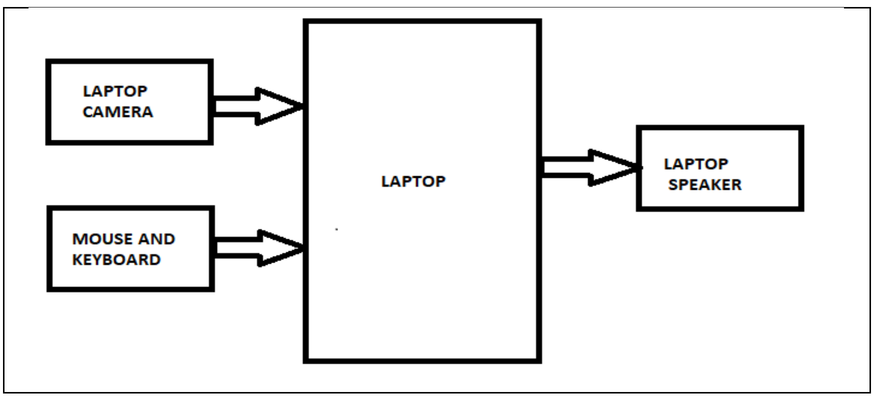

Fig. 6.1: Block diagram of Blind Assist System

The brain of system is laptop, all the processes are executed by laptop using ML. In the figure 6.1 the system consists of laptop camera which is used to take live video capture. The video is converted into frames, which are subsequently employed to detect the object. In this system we used a pretrained dataset called COCO dataset. In this dataset we already pretrained required number of objects. Utilizing that dataset, the object obtained from the camera is compared to facilitate object detection. After detection, the ESPEAK library is employed for audio conversion to announce the label of the object using the speaker.

6.2 METHODOLOGY

1. Objective:

The objective of this project is to implement real-time object detection using a webcam feed and provide audio feedback for detected objects.

2. Environment Setup:

- Python environment with required libraries: OpenCV, pyttsx3, matplotlib.
- Access to a webcam for live video feed.

3. Pre-trained Model Selection:

- Selected the SSD MobileNet V3 model for its balance between accuracy and speed in real-time object detection tasks.
- Chose the model trained on the COCO dataset due to its wide range of object classes.

4. Configuration and Initialization:

- Set the threshold for object detection confidence.
- Initialized the webcam feed and configured its properties (resolution, frame rate).
- Loaded the class names from the COCO dataset and the paths to the pre-trained model files.

5. Object Detection:

- Utilized OpenCV's `dnn_DetectionModel` class to load the pre-trained model.
- Configured input size, scale, mean, and color swapping to preprocess frames for inference.
- Implemented a helper function to calculate the distance to detected objects based on their bounding box width.

6. Real-time Processing Loop:

- Iterated over frames captured from the webcam in a continuous loop.
- Utilized the pre-trained model to detect objects in each frame.

- Annotated detected objects with bounding boxes, object names, confidence levels, and distances.
- Utilized Matplotlib to display annotated frames for real-time visualization.
- Provided audio feedback for detected objects using the `pyttsx3` library.

7. User Interaction:

- Enabled the user to exit the program by pressing the 'q' key.
- Provided real-time feedback through visual annotations and audio announcements.

8. Performance Optimization:

- Adjusted parameters such as input size, scale, and mean to optimize inference speed without compromising accuracy.
- Implemented multithreading for improved performance, separating the video capture and processing tasks.

9. Testing and Validation:

- Conducted extensive testing to ensure the accuracy and reliability of object detection.
- Validated the system's performance under various lighting conditions and object distances.
- Obtained feedback from users to identify areas for improvement and refinement.

10. Documentation and Reporting:

- Documented the entire process, including setup instructions, code explanation, and methodology.
- Prepared a detailed report outlining the project objectives, methodology, implementation details, results, and conclusions.

11. Future Enhancements:

- Investigate the integration of more advanced object detection models for improved accuracy.
- Explore the possibility of incorporating depth estimation techniques for more accurate distance calculation.
- Enhance the user interface for better user interaction and accessibility.
- Optimize the code for deployment on resource-constrained devices such as embedded systems or mobile platforms.

6.3 FLOWCHART FOR OBJECT DETECTION MODULE

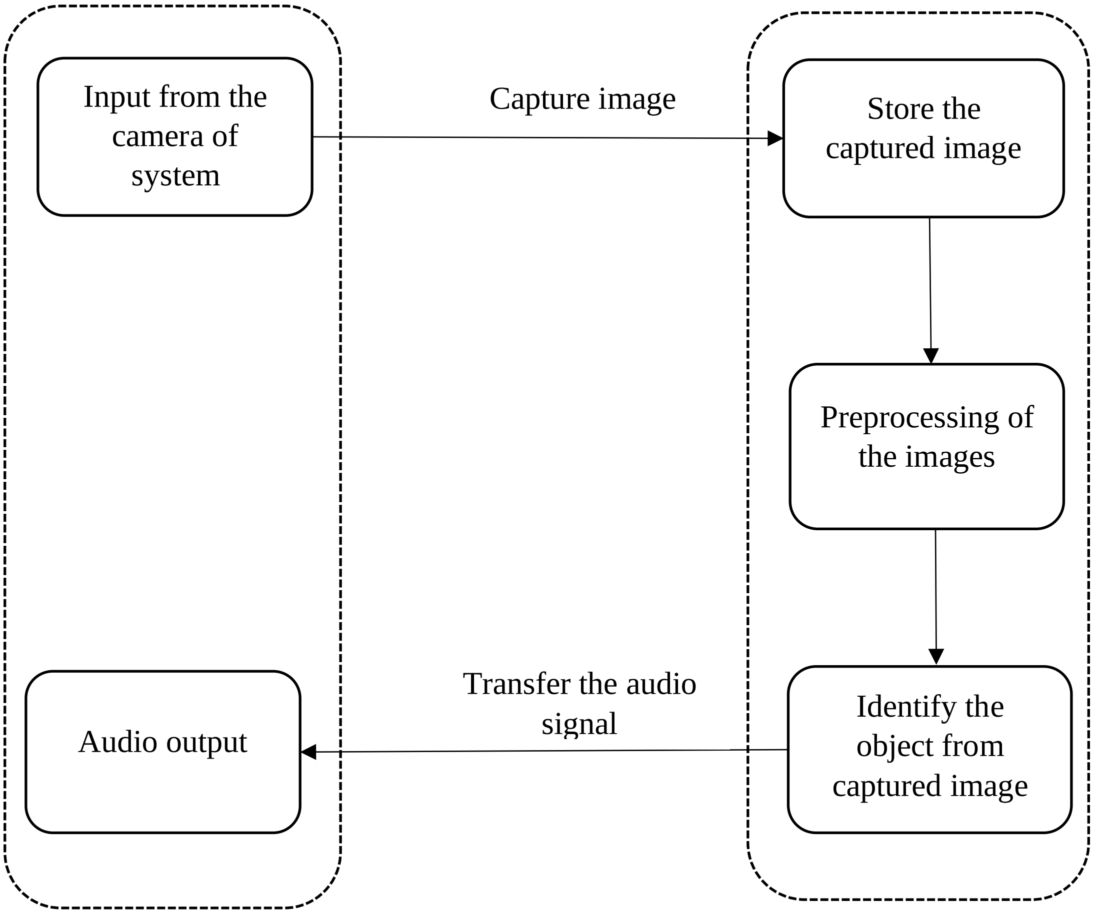

Fig. 6.3: Flow Chart for Object Detection Module

Visually challenged people, on the other hand, cannot readily go outside for work. They are completely reliant on others. As a result, when they wish to walk outside, they will want assistance. In figure 6.3 proposed design is based on the recognition of objects in the environment of a blind person. The obstacle detection technology functions in such a way that it requires various steps from frame extraction to output recognition. To detect items in each frame, a comparison between query frame and database objects is performed. A system is presented for recognizing and locating objects in photographs and videos. An audio file carrying information about each object detected is activated. As a result, both object detection and identification are addressed at the same time.

6.4 FLOWCHART FOR READING MODULE

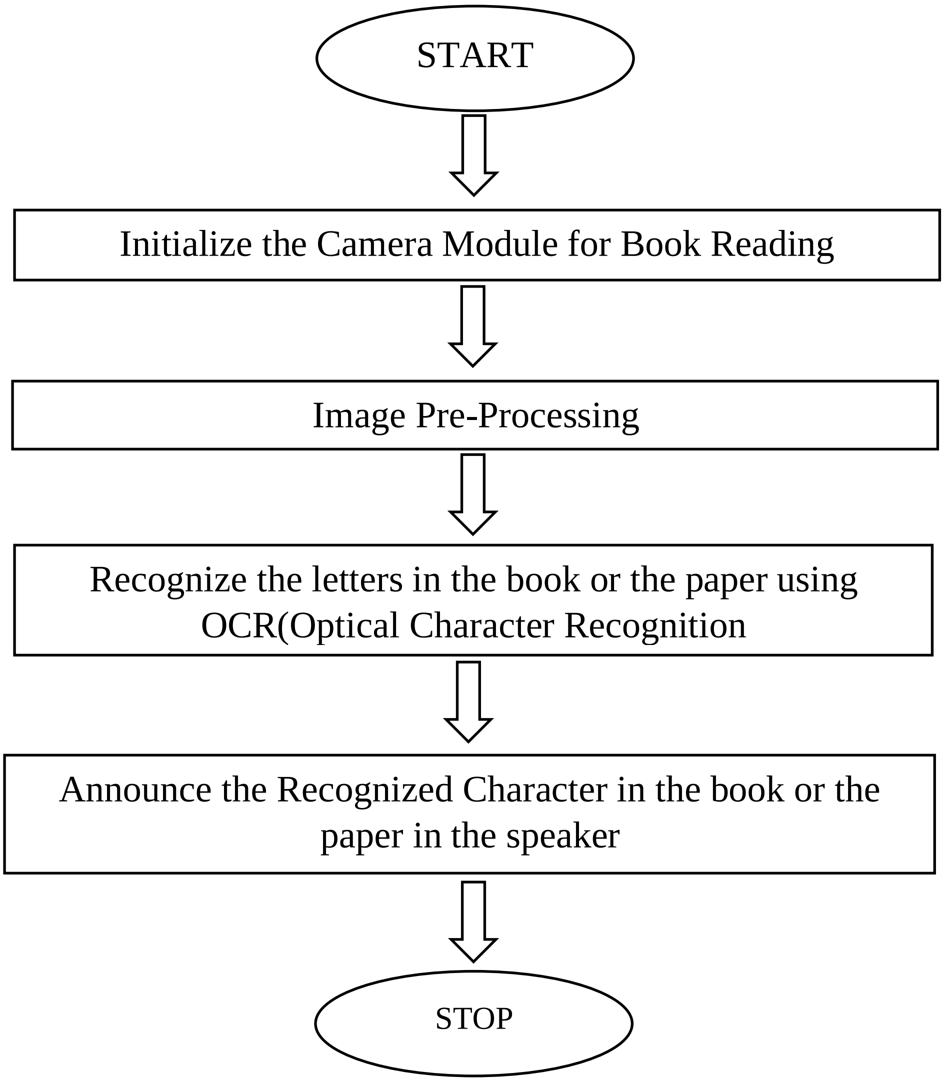

Fig. 6.4: Flowchart for Reading Module

1. **START:** In figure 6.4 the first step initiates the process of assisting visually impaired individuals in reading printed text from books or papers. It marks the beginning of the workflow.
2. **Initialize the Camera Module for Book Reading:** In figure 6.4 the second step involves setting up a camera or a specialized device equipped with a camera to capture images of the text from books or papers. The camera is positioned in a way that allows it to capture clear and focused images of the text.
3. **Image Pre-Processing:** Before performing optical character recognition (OCR) on the captured images, various pre-processing techniques are applied to enhance the

quality of the images. This may include steps such as resizing, noise reduction, contrast enhancement, and binarization to improve the readability of the text.

4. **Recognize the letters in the book or the paper using OCR (Optical Character Recognition):** Once the images are pre-processed, OCR software is used to analyze the images and extract text from them. OCR algorithms identify and interpret the characters present in the images, converting them into machine-readable text data.
5. **Announce the Recognized Character in the book or the paper in the speaker:** The recognized text is then converted into speech using text-to-speech (TTS) technology. The synthesized speech is then played through a speaker or a set of headphones, allowing visually impaired individuals to listen to the contents of the book or paper in real-time.
6. **STOP:** In figure 6.4 the last step marks the end of the process. Once the text has been successfully read aloud, the system may wait for further input or perform any necessary cleanup before shutting down or returning to standby mode.

CHAPTER 7

PROJECT DESCRIPTIONS

7.1 MODULES

- Object Detection module: Pre-trained COCO dataset.
- Book reading module: OCR module.
- Image processing module: To enhance image quality.

7.2 MODULES DESCRIPTIONS

1. Object Detection Module:

The image dataset was built with the objective of improving image recognition, therefore COCO stands for Common Objects in Context. The COCO dataset offers demanding, high-quality visual datasets for computer vision, with the majority of the datasets containing state-of-the-art neural networks.

COCO dataset was used that is available for objects detection, objects segmentation, etc. In this project, convolutional neuronal networks were trained to perform the process of discovering organisms. CNN has been trained in object detection using 90 objects such as car, book, apple, bear, mobile, bicycle, car etc.

There are a total of 90 predefined objects in this data set.

2. Book Reading Module:

OCR technology is used for detecting text and TensorFlow is used for detecting objects and recognizing people. Detected objects and text are converted to speech and is provided as an audio output to blind via a Bluetooth headphone.

Optical character recognition (OCR) systems provide persons who are blind or visually impaired with the capacity to scan printed text and then have it spoken in synthetic speech or saved to a computer file.

3. Image Processing Module:

For enhancing image quality in techniques aimed at assisting visually impaired and blind people, various image processing methods can be employed. While NumPy itself doesn't offer specific image processing functions, it serves as a foundational library for such tasks when combined with other libraries like OpenCV or scikit-image.

Here's a general outline of steps you might take using NumPy along with other libraries for image enhancement:

1. Loading the image: Use libraries like OpenCV or Pillow to load the image into a NumPy array.
2. Preprocessing: Preprocess the image, if necessary, which might involve steps like resizing, denoising, or normalization to make subsequent processing more effective.
3. Enhancement techniques: Apply enhancement techniques tailored to assist visually impaired individuals. Some common techniques include:

- Contrast enhancement: Adjust the contrast to improve visibility of details.
- Brightness adjustment: Modify the brightness of the image to make it more suitable for viewing.
- Color adjustments: Alter color channels to make specific elements more distinguishable.
- Edge enhancement: Enhance edges to make object boundaries more prominent.
- Noise reduction: Reduce noise in the image to improve clarity.
- Sharpness enhancement: Enhance image sharpness to improve visibility of edges and details.

4. Postprocessing: Perform any necessary postprocessing steps, such as resizing the image back to its original dimensions or converting it to a different format for display or storage.
5. Display or output: Display the enhanced image using appropriate tools or save it to a file for further use.

CHAPTER 8

RESULT

8.1 RESULTS OF OBJECT DETECTION

The following data represents the outcomes of object detection, including the identified objects and their corresponding accuracy levels. Audio output also will generate.

In fig. 8.1.1, shows the object detection of scissors and toothbrush with its corresponding accuracy and also the audio output of the same.

Fig. 8.1.1: Output 1

In fig. 8.1.2, shows the object detection of persons with their corresponding accuracy and also the audio output of the same.

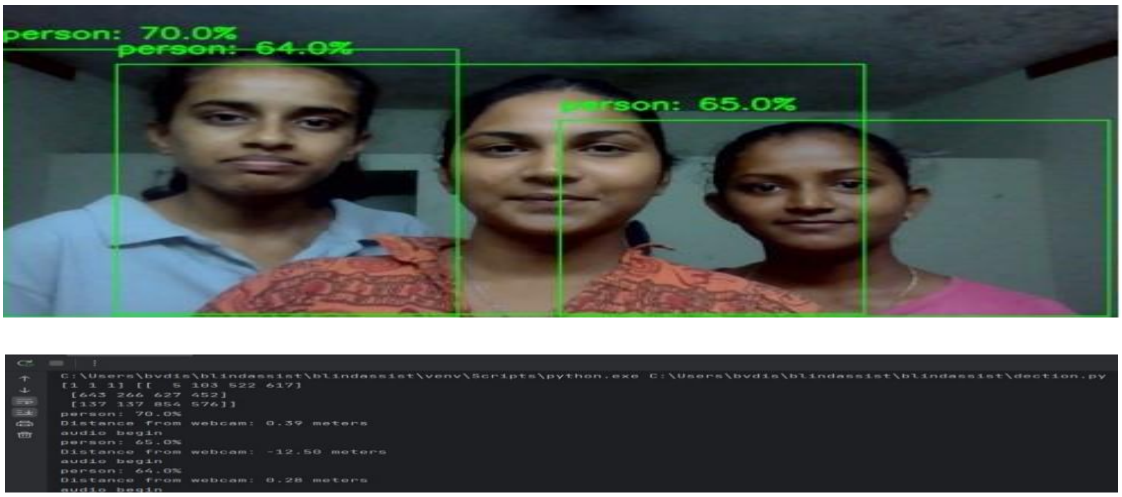

Fig. 8.1.2: Output 2

In fig. 8.1.3, shows the object detection of person and laptop with its corresponding accuracy and also the audio output of the same.

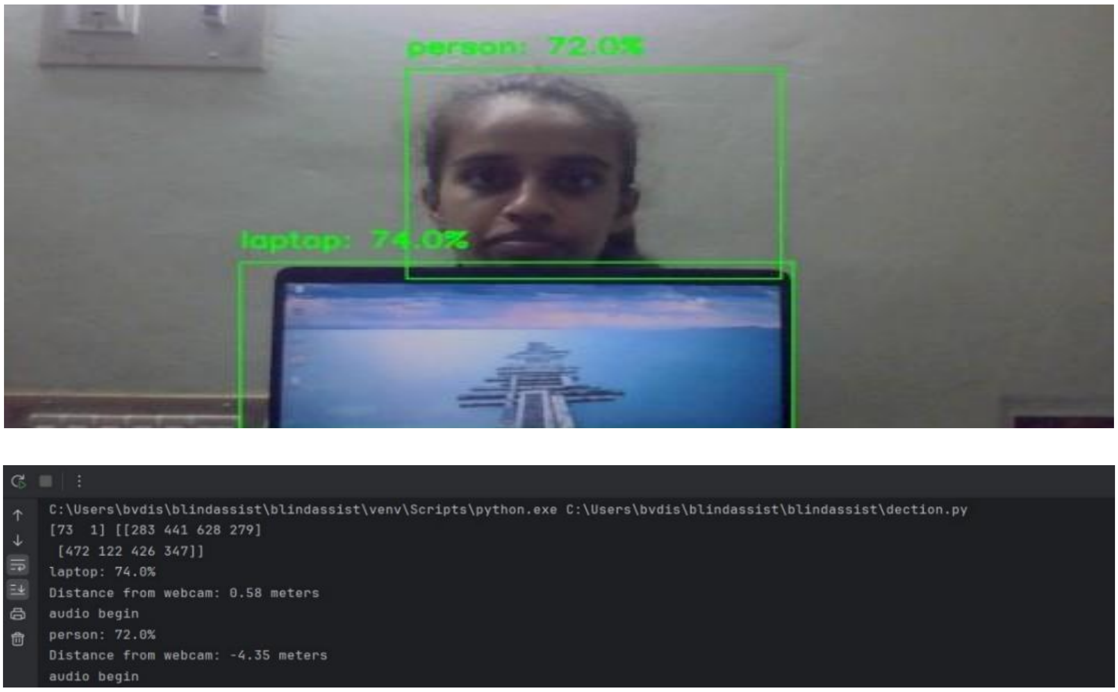

Fig. 8.2.3: Output 3

In fig. 8.1.4, shows the object detection of bottle and teddy bear with its corresponding accuracy and also the audio output of the same.

Fig. 8.1.4: Output 4

In fig. 8.1.5, shows the object detection of umbrella and person with its corresponding accuracy and also the audio output of the same.

Fig. 8.1.5: Output 5

In fig. 8.1.6, shows the object detection of cell phone and remote with its corresponding accuracy and also the audio output of the same.

Fig. 8.1.6: Output 6

In fig. 8.1.7, shows the object detection of person and cup with its corresponding accuracy and also the audio output of the same.

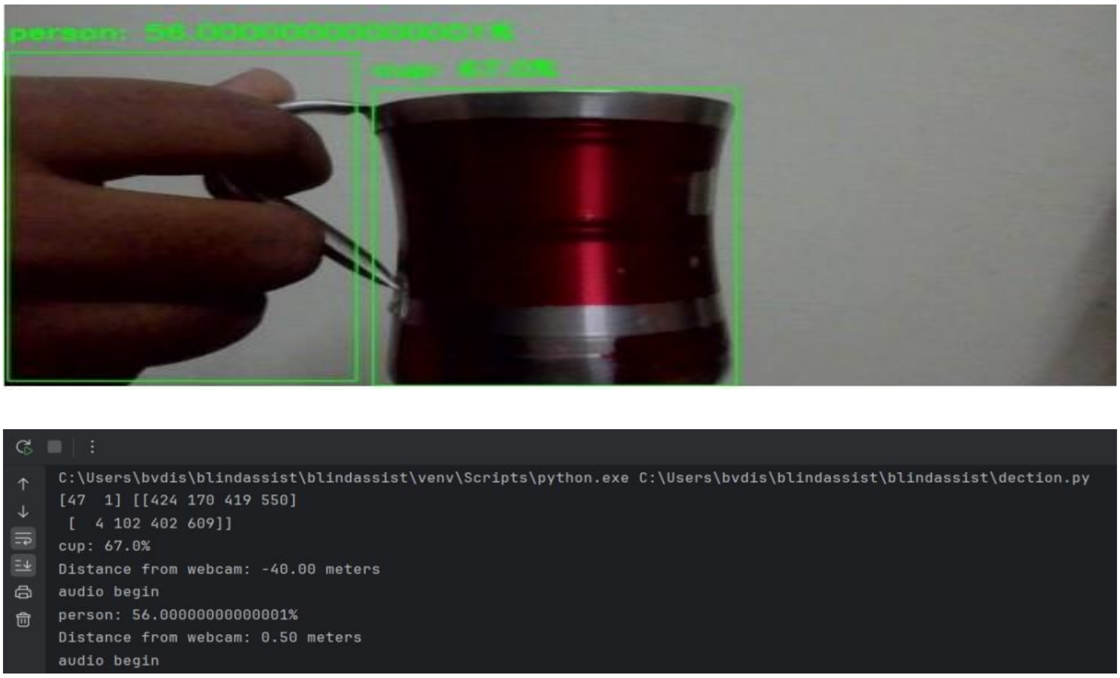

Fig. 8.1.7: Output 7

In fig. 8.1.8, shows the object detection of cell phone and cake with its corresponding accuracy and also the audio output of the same.

Fig. 8.1.8: Output 8

8.2 RESULTS OF READING MODULE

The following figures represents the outcomes of reading module, shows the text output. Audio output also will generate.

In fig.8.2.1, shows the reading of the given image in both text and audio format.

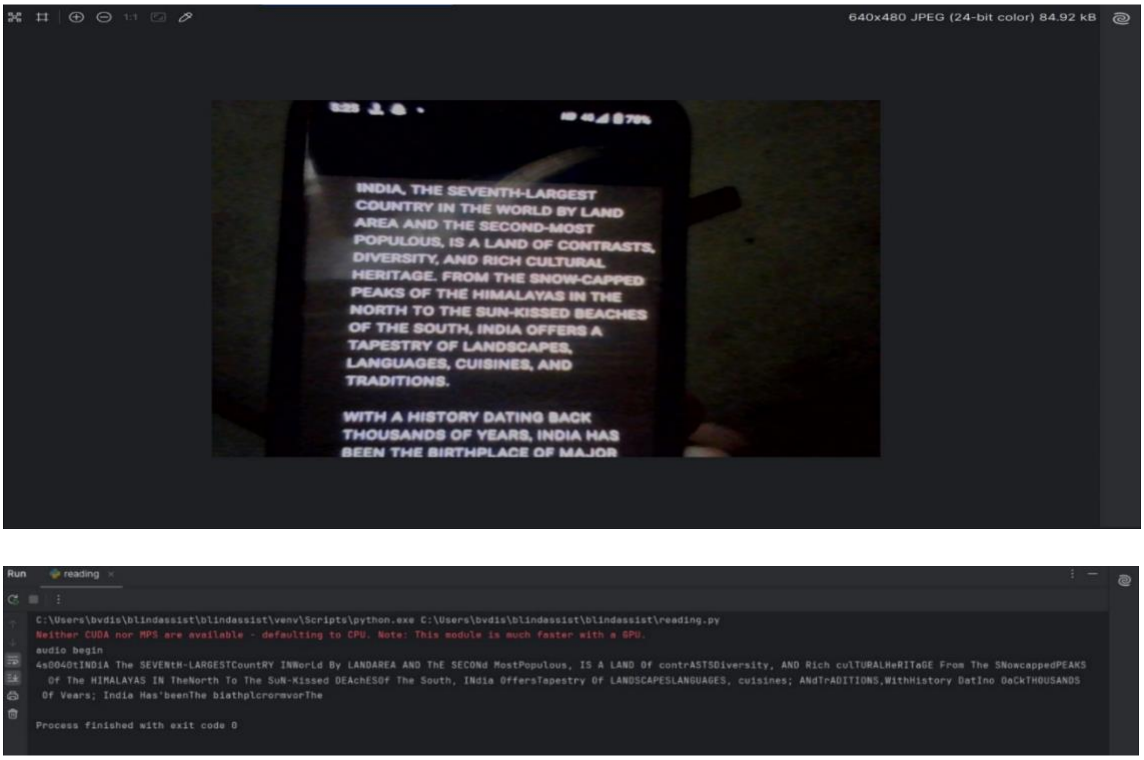

Fig. 8.2.1: Output 9

In fig. 8.2.2, shows the reading of the given image in both text and audio format.

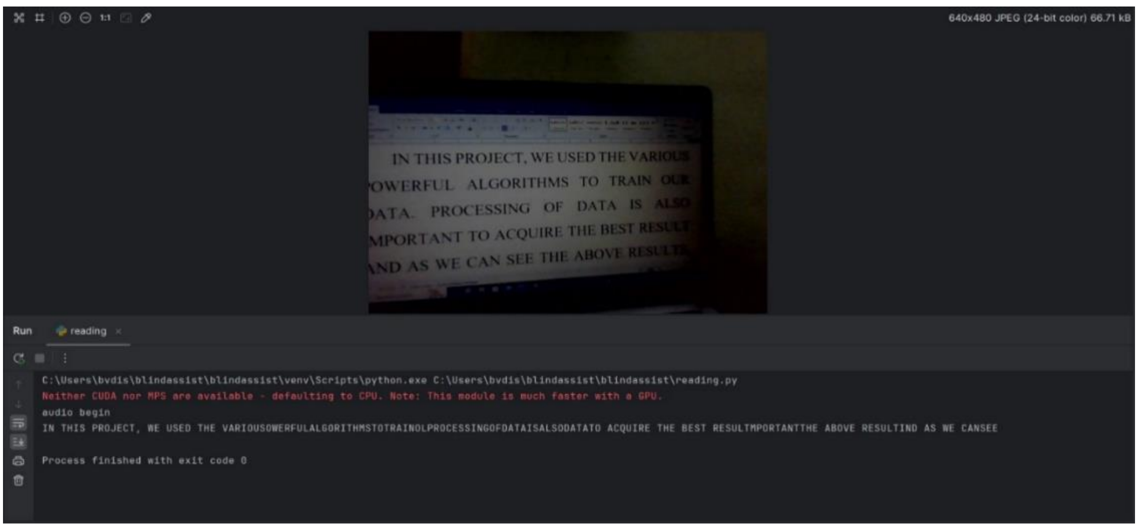

Fig. 8.2.2: Output 10

In fig. 8.2.3, shows the reading of the given image in both text and audio format.

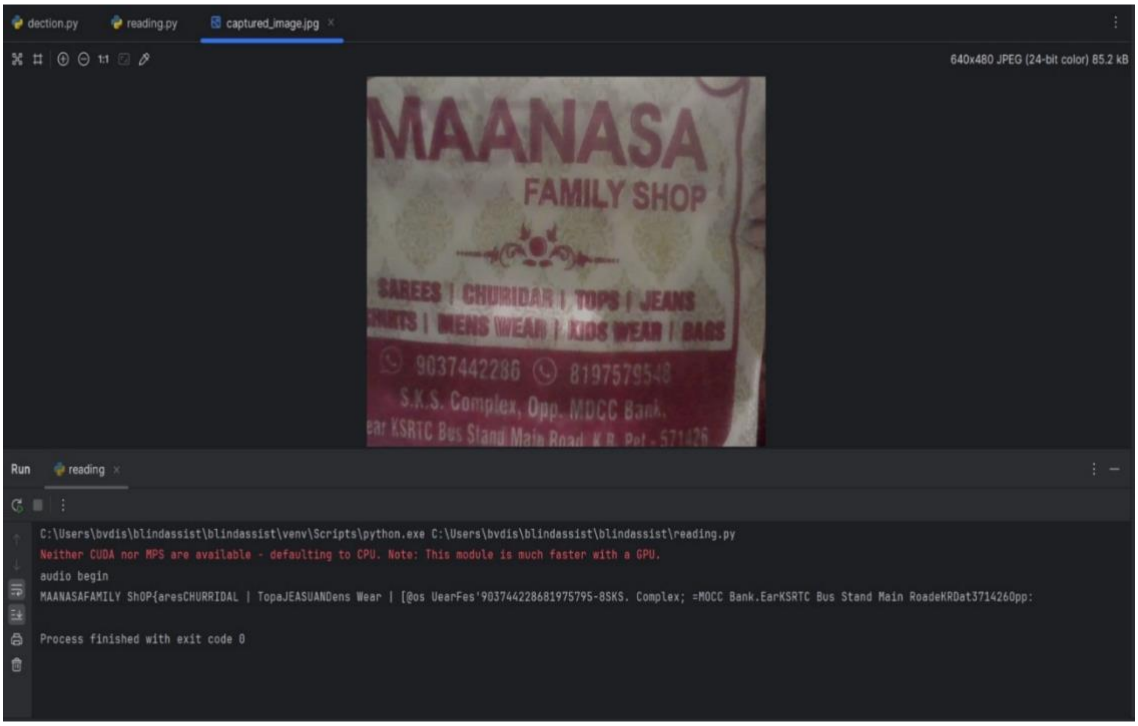

Fig. 8.2.3: Output 11

In fig. 8.2.4, shows the reading of the given image in both text and audio format.

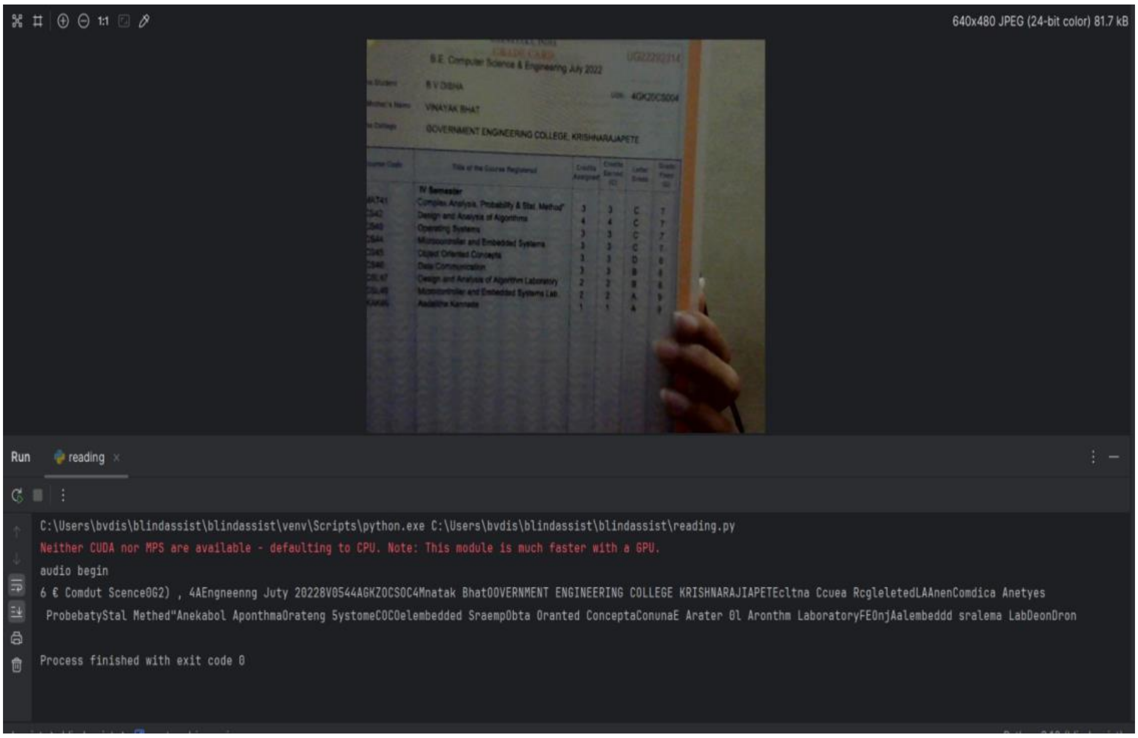

Fig. 8.4.4: Output 12

In fig. 8.2.5, shows the reading of the given image in both text and audio format.

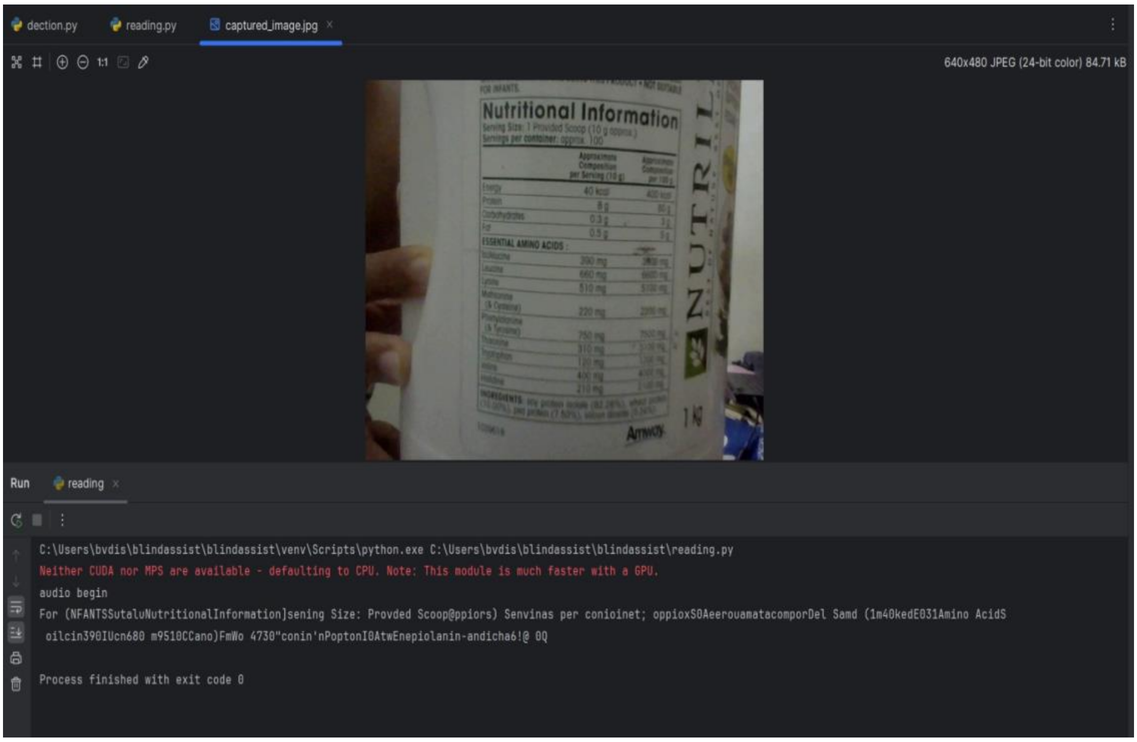

Fig. 8.2.5: Output 13

CHAPTER 9

CONCLUSION AND FUTURE SCOPE

In view of experimental emphasize, the blind assist system, incorporating object detection with the COCO dataset and a novel book reading feature, represents a significant stride towards fostering inclusivity and independence for individuals with visual impairments. However, empirical evidences encompass seamlessly integrating advanced computer vision and image processing technologies, the project empowers users to navigate their surroundings confidently through real-time object identification. Moreover, the innovative inclusion of a book or print media material reading capability, which supported by optical character recognition and text-to-speech functionalities, consequently opens new avenues for autonomous access to written content. This project not only addresses practical challenges faced by the visually impaired but also underscores the transformative potential of technology in enhancing accessibility, enriching educational experiences, and ultimately improving the overall quality of life for the community.

FUTURE SCOPE

Looking ahead, the future scope of this project involves several key areas of improvement and expansion to enhance its impact and usability for individuals with visual impairments:

1. **Enhanced Object Detection Accuracy:** Continuously refining and optimizing the object detection algorithms to improve accuracy and reliability in identifying various objects in different environments and lighting conditions.
2. **Expanded Object Categories:** Expanding the object categories beyond the COCO dataset to include a wider range of everyday objects, obstacles, and environmental elements that are relevant to the daily lives of visually impaired individuals.
3. **Advanced Navigation Features:** Integrating advanced navigation features, such as route planning, landmark recognition, and obstacle avoidance, to provide users with more comprehensive assistance in navigating complex environments both indoors and outdoors.
4. **Improved Book Reading Experience:** Further refining the book reading feature by enhancing the accuracy of optical character recognition (OCR) and text-to-speech

(TTS) functionalities, as well as adding support for additional languages and document formats.

5. **Integration with Wearable Devices:** Exploring integration with wearable devices, such as smart glasses or wearable cameras, to provide a more seamless and hands-free user experience for accessing real-time object detection and navigation information.

Eventually, by addressing these future directions and continuously iterating on the existing capabilities of the blind assist system, we aim to further advance its effectiveness, usability, and overall impact in fostering inclusivity, independence, and improved quality of life for individuals with visual impairments.

REFERENCES

[1] Rajani Suryakant Kolhe, Kajal Gajanan Dhole, Pratibha Sampat Thakre, Priyanka Sohan Prasad," Smart Stick for The Blind and Visually Impaired People", International Journal of Scientific Research in Science, Engineering and Technology, pp.207-210,2021.

[2] "Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People", June 2022, International Journal of Electrical and Electronics Research 10(2):80-86, June 202210(2):80-86, DOI:10.37391/ijeer.100205

[3] "Real time object detection for visually challenged people using machine learning "Prof. Sneha A. Khaire1, Dhanraj Nikam2, Ajinkya Pawar3, Chaitanya Devkar4, Shubham Pagar5, international journal of progressive research in engineering management and science (IJPREMS)Vol.03, Issue 05, May 2023, DOI: https://www.doi.org/10.58257/IJPREMS31126

[4] "Real time object detection system for blind people" Prof. V.N. Nirgude*1, Shwetark Kadam*2, Arihant Lodha*3, Rushikesh Kadam*4, Anirudha Kurhe*5 International Research Journal of Modernization in Engineering Technology and Science (Peer-Reviewed, Open Access, Fully Refereed International Journal) Volume:05/Issue:05/May-2023.

[5] Real time object detection and recognition system to assist the visually impaired 1dsouza elston ronald, 2dhanya bhat 3 john alsten tauro, 4s harshith, 5merlyn melita mathias, www.ijcrt.org © 2021 IJCRT | Volume 9, Issue 7 July 2021 | ISSN: 2320-2882

[6] Object Detection and Localization for Visually Impaired Dileep Reddy Bolla 1 Hrishita Rauniyar 2, Sowmya DN 3, Rakshitha.T.P 4 , Nabi Rasool 5, ITM Web of Conferences 57, 01014 (2023), https://doi.org/10.1051/itmconf/20235701014

[7] Object Detection using Artificial Intelligence: A Review 1 Shibani Singh, 2Mimi Ton Devi, 3Kandarpa Kalita 1 Student, 2 Student, 3Assistant Professor, © 2022 IJRTI | Volume 7, Issue 8 | ISSN: 2456-3315

[8] S. Alagarsamy, T. D. Rajkumar, K. P. L. Syamala, C. S. Niharika, D. U. Rani and K. Balaji, "An Real Time Object Detection Method for Visually Impaired Using Machine Learning," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-6, doi:10.1109/ICCCI56745.2023.10128388.

[9] K. G, R. P, N. N N, A. Beulah and R. Priyadharshini, "Detection of Electronic Devices in real images using Deep Learning Techniques", 2021 5th International Conference on Computer Communication and Signal Processing (ICCCSP), pp. 295-300, 2021.

[10] S. Alagarsamy et al., "Smart System for Reading the Bar Code using Bayesian Deformable Algorithm for Blind People", 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 424-429, 2022.Deep Learning Reader for Visually Impaired.by Jothi Ganesan 1ORCID, Ahmad Taher Azar 2,3, *ORCID,Shrooq Alsenan 2,Nashwa Ahmad Kamal 4,Basit Qureshi 2ORCID andAboul Ella Hassanien 5

[11] Research Paper on Blind Assist System Using Ml and Image Processing, Prof. Deepu G1, Adarsh N M 2, Bhuvan R3, Darshan A 4, Darshan R5, International Journal of Research Publication and Reviews, Vol 5, no 5, pp 9253-9256 May 2024

[12] ML Based Blind Assistance System Gudla Vaishnavi1, Varkala Prakash2, Sreebhashyam Pranatha3, Thimmanagari Sampurna4, B Rekha5, © April 2023| IJIRT | Volume 9 Issue 11 | ISSN: 2349-6002

APPENDIX-A

DATASET

COCO-Dataset Files:

1. person
2. bicycle
3. car
4. motorcycle
5. airplane
6. bus
7. train
8. truck
9. boat
10. traffic light
11. fire hydrant
12. street sign
13. stop sign
14. parking meter
15. bench
16. bird
17. cat
18. dog
19. horse
20. sheep
21. cow
22. elephant
23. bear
24. zebra
25. giraffe
26. hat
27. backpack
28. umbrella
29. shoe
30. eye glasses

31. handbag
32. tie
33. suitcase
34. frisbee
35. skis
36. snowboard
37. sports ball
38. kite
39. baseball bat
40. baseball glove
41. skateboard
42. surfboard
43. tennis racket
44. bottle
45. plate
46. wine glass
47. cup
48. fork
49. knife
50. spoon
51. bowl
52. banana
53. apple
54. sandwich
55. orange
56. broccoli
57. carrot
58. hot dog
59. pizza
60. donut
61. cake
62. chair
63. couch
64. potted plant

65. bed
66. mirror
67. dining table
68. window
69. desk
70. toilet
71. door
72. tv
73. laptop
74. mouse
75. remote
76. keyboard
77. cell phone
78. microwave
79. oven
80. toaster
81. sink
82. refrigerator
83. blender
84. book
85. clock
86. vase
87. scissors
88. teddy bear
89. hair drier
90. toothbrush
91. hair brush