Smart Navigation for Visually Impaired people using Artificial Intelligence

Rajvardhan Shendge[1*], *Aditya* Patil[2], and *Siddhi* Kadu[3]
[1]Department of Computer Engineering, Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Navi Mumbai
[2]Department of Computer Engineering, Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Navi Mumbai
[3]Department of Computer Engineering, Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Navi Mumbai

Abstract. This research introduces a blind aid that uses a live object recognition system. People who are blind or partially sighted rely significantly on their other senses, such as touch and auditory cues, to comprehend their surroundings. There is a need to deploy a technology that assists visually impaired persons in their daily routines, as there is now very little aid. Existing solutions, such as Screen Reading software and Braille devices, assist visually impaired individuals in reading and gaining access to numerous gadgets. However, these technologies are rendered worthless When the blind need to perform basic activities like recognizing the situation before them, such as recognizing people or objects, technologies become ineffective. This method will benefit blind or visually challenged all across the world. The goal is to help a person with total or partial blindness obtain a second set of eyesight without the assistance of a guardian, allowing them to live a better and more independent life. This project outlines working to create a more welcoming and inclusive environment, focusing on assistive technology that provides services, resources, and information to those with visual disabilities.

1 Introduction

One of the essential human senses is eyesight, and it plays a massive role in human life and helps in the perception of the surrounding environment. India is believed to include around 60 percent of the world's visually handicapped population. As our society progresses technologically, many new tools for people with disabilities have developed. The assurance of mobility for visually impaired persons is one of the many urgent supports. This innovative program will aid blind and visually impaired people all around the globe. The main goal is to let someone who is partially blind gain a second pair of vision without the assistance of a guardian, allowing them to live a worry-free life. This project expands on how it will aid in the creation of a more welcoming and inclusive environment, focusing on assistive technology that delivers services, resources, and data to those with visual difficulties. Independent living is a need in today's environment, yet visually impaired people confront significant limitations. The visually impaired are at a disadvantage since they need physical assistance to get information about their surroundings. As a result, there is a need to deploy technology that aids visually impaired people in their daily routines, as there is now very little help accessible. Screen Reading software and Braille devices are examples of existing solutions with different scopes. These technologies, on the other hand, are made useless. When visually impaired persons need to do essential fundamental functions like identifying the scene before recognizing people or things, such technologies are rendered ineffective. As a result, current technologies are more or less inaccurate, and if they fail, they may injure the user.Consequently, the user would be guided by this sophisticated program without the need for a personal human guide. This application would be self-contained, autonomous, and stand-alone since employing a fulltime guide may be costly and jeopardize one's privacy. It is straightforward for a vision-impaired person to carry out fundamental tasks since they can see items in their surroundings and can readily overcome any obstacles in their path. To avoid a mistake, visually impaired persons must be aware of their environment and their interactions with items. Finally, as an output of the object recognised within the frame with the highest confidence score among all other things present, the system will emit audio. The frame is chosen at a specific time interval to ensure that the audio output is not hampered.

In this paper, an ensemble approach is proposed to detect an object detected in a live camera frame and the audio output of the detected object's name and its actual position.The major contribution of the paper are providing smart android app for blind people which includes object detection with audio as output of

* Corresponding author: author@email.org

detected object in live camera frame as well as position of detected object.

The remaining of the paper is organized as follows: Section 2 is the survey of existing systems for detecting objects for a blind person—section 3 includes the analysis gap, followed by section 4 of system design. Section 5 demonstrates the proposed system along with the ensemble approach. The implementation details in 6 and the proposed system results are presented in section 7. Finally, the conclusion and future work is expounded in Section 8.

2 Literature Survey

There are various scholarly publications such as journal articles, research work etc. that are done in the field related to mobile navigating app for blind people. Many such articles and research reports were referred before making the project report. The reports/articles that have influenced the project, the most are :

WafaM.Elmannai, et al. [1] proposes a method intended to assist the visually impaired. The system combines sensor-based techniques with computer vision concepts to achieve an economically viable solution. Based on fuzzy logic and deep image information, the system uses an algorithm called New Obstacle. Using these techniques, visually impaired people can avoid obstacles by detecting objects in front of them. This system helps to support six blind people indoors and outdoors. System hardware requirements include a camera, GPS, Wi-Fi, microphone, compass, gyroscope, and microcontroller. This requires multisensory data and computer vision approaches at the software level. 96% detection accuracy and 100% the system offers obstacle avoidance. It helps to pass safely and gives high performance. The system is considered more dependable, straightforward, transportable, inexpensive, and accessible. This system also has some limitations when it comes to working. The system is compatible while detecting large objects because their size plays an essential role as they may not be detected in the frame. So it can be not easy to detect by finding the difference between background and background.

Qi-Chao Mao, et al. [2] proposes a method focuses on the autonomous mobility of the visually impaired. This is done by designing a wearable assistive device limited to blind people detecting traffic lights. The system is designed based on the AdaBoost algorithm. This approach is faster and more powerful in detecting objects. This system is enhanced with a flexible parallel architecture on the FPGA (Field Programmable Gate Array) platform. The main image and the confidence of the weak classifier are calculated in parallel. The parameters of the weak classifier are trained by the AdaBoost algorithm with MATLAB software and then configured on the FGPA platform. FPGA is said to be more flexible and consume less power. Testing shows that the system will detect traffic lights in videos at a frame rate of around 30 fps.

ZhenchaoOuyang, et al.[3] proposes a method for providing orientation and navigation capabilities, this paper proposes an electronic device called NavCane. This device helps visually impaired people move indoors and outdoors in a barrierfree journey. This device sends advance information about obstacles to its user without information overload, and the information is sent to that person through operational and auditory methods. It has many components such as an ultrasonic sensor and wet floor, GSM and GPS module, gyroscope, radiofrequency recognition vibrator motor, global module positioning system, and battery. The system has been evaluated by 80 visually impaired people and successfully in various situations. This NavCane device detects obstacles in known indoor environments, unlike other electronics. It is considered a low power consumption device in the vehicle and a low power consumption system. Analysis indicates that this NavCane improves barrier-free performance much more than white cane.

Wei Fang, et al.[4] This research focuses on knowledge-based data training and projected regions of interest to present a technique for developing a high-speed deep neural network for real-time video object recognition. The method creates a framework for training datasets across deep neural networks using limited sampling and cross-network knowledge prediction, which increases performance while reducing processing complexity. The training process is regulated by learn projection matrices by projecting knowledge and images representation of the teacher-level network from its middle layer to the middle layer sub-network layer. Experiments are being carried out to show that this system decreases the network's computational complexity by a factor of 16 and increases network performance significantly.

VidulaV.Meshram, et al.[5] proposes a method for embedded devices which have limited memory and processing power, making real-time object detection extremely difficult. Therefore, this paper proposes a design in which detection accuracy is not reduced. This task uses a lightweight object detection method called MiniYOLO v3. For darknet-based backbones, there are depthseparable folds and point-by-point group folds to reduce the size of parameters in the network. The process reducing the dimensions and then increasing the dimensions is adopted with the structure. The boundary effect of pointwise group convolution is suppressed by channel shuffle.

WenmingCao, et al.[6] proposes a method You Only Look Once, i.e., YOLO is one of the most well-known and widely used high-speed real-time object detection methods. To get better object detection results in a particular environment, this document proposes using the Tinier YOLO, which focuses improving better detection of results, performance and accuracy by reducing the size of the system . Focuses on network performance, detection speed, and accuracy to overcome all TinyYOLO issues. One of the biggest challenges in deploying fire modules in TinyYOLO V3

is to see the total number of these modules and their location in the system. Second, there is a problem with the style of connectivity between the modules to achieve better results in detection, accuracy, and performance. The system employs tight connections between launch modules to enhance mobility and maximize the flow of information between networks. However, beyond that, the model size cannot be reduced, as recognition accuracy can be significantly reduced. Therefore, this system uses a continuity component to address the issue of reducing model size. These path layers help you combine features extracted from the front layer to get the right features. Finally, the system suggests removing batch normalization from the module and overall performance to reduce computational cost factors.

3 Gap Analysis

Above mentioned papers propose a method of detecting the object in a live frame or detecting an object on the captured image by providing an audio output of the detected object in the captured image. At the same time, our system proposes a system in which an object is detected in the live frame, and the audio output of that detected object in the live frame and the position of the detected object in provided as output to the user.

4 System Design

In figure 1 the overall design is described as follow which is an MVC design Pattern and some people also termed it as architecture as well.

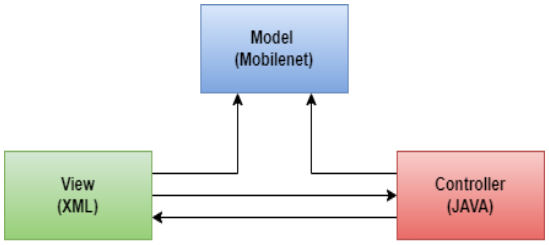

Fig. 1. System Design

4.1 Architecture Diagram

In figure 2 the system detects various objects in realtime producing real-time voice output. This system is divided into two parts: the user and the android application layers. Android Application receives input from the camera as input then the captured image is stored in the android application as a captured image then the pre-processing of captured image takes place , this process is called as pre-processing of the image the algorithm which identify the object from the image from the captured image, after detection of an object from the captured image it labels the object according to maximum accuracy and the application generates an audio signal for the identified object these audio signals are transferred to user layer as an audio output.

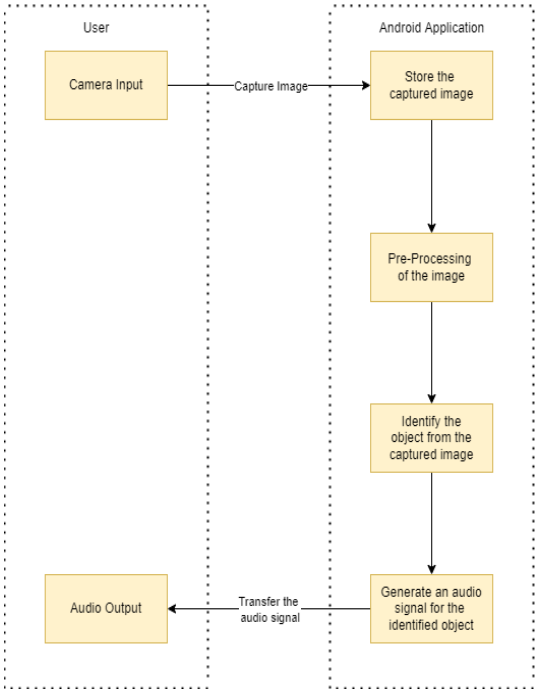

Fig. 2. Architecture Diagram

4.2 Object Detection

The particular thing detection plus recognition have in order to be accurate, because well as the particular detecting speed through the object ought to be high to ensure that the navigation regarding blind folks might be more simple. A fast R-CNN algorithm has already been implemented[10]. The particular system does not work on several regions; the R-CNN algorithm processes the particular image into the bundle of containers and checks when some of these types of boxes contain any object. The quick R-CNN algorithm offers to use the picky search methodology in order to extract these containers from a picture. The varying weighing scales, colors, textures, and enclosure would become the four areas that form a suitable object. The picky search identifies the particular patterns within the particular image, and based on those designs, they split the particular image into numerous regions. Object recognition is a software program related to pc vision and picture processing that detects the particular existence of items in digital pictures and videos along with a limitation package and types or even classes of items [2].

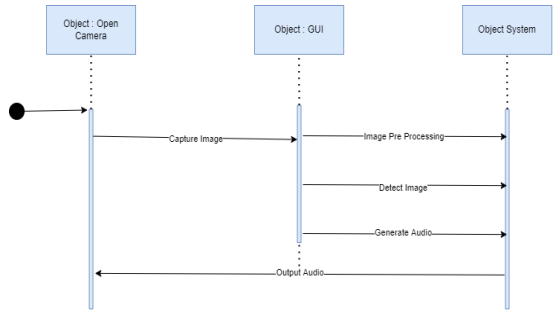

Fig. 3. Sequence Diagram for Object Detection

figure 3 shows the obejct is captured live through mobile camera and this captured image gets passed through object gui where image preprocessing process takes place, where object is detected and audio is generated for that particular detected object along with posiiton of that detected object in object system and this generated audio is given as ouput to user.

Aesthetically impaired users may understand their environment without difficulty plus remain independent simply by utilizing object recognition. Input: a picture or an image with one or even more objects. A single or more restricting boxes (e., Gary the gadget guy, defined by the dot, width, plus height) and a course label for every restricting box are created as output. Bounding boxes are produced, which predict the confidence score's particular certainty. This particular score lets us know that the bounding package will contain some items. For every bounding box, the cellular predicts a course of that item which provides the probability distribution amongst all the accessible classes in a given model. The particular confidence score plus the probability calculated gives all of us the final rating, which lets the particular consumer know exactly how likely it is that the particular bounding box consists of some specific item.

4.3 Tensorflow

In Figure 4, the TensorFlow Lite design is represented in a specific efficient transportable format known as FlatBuffers (identified by the. flite document extension), which offers several advantages over TensorFlow's protocol buffer design format, including reduced size (small programme code footprint) and faster inference (data is accessed directly without an additional parsing/unpacking step), allowing TensorFlow Lite to execute efficiently. For an automated era of pre- and postprocessing pipelines during on-device inference, a TensorFlow Lite model may optionally incorporate metadata with the human-readable model description and machine-readable information. This item detection approach makes use of the COCO mobile net SOLID STATE DRIVE v1 model, which has datasets for 80 different object categories that are regularly encountered in our world.

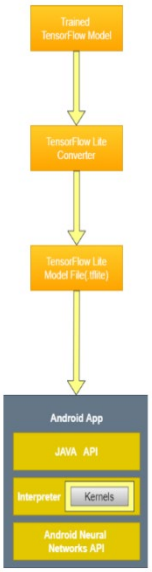

Fig. 4. Tensorflow Architecture

4.4 Image Pre-Processing

The first goal of Image and Video pre-processing is to drop frames from the video stream. A contemporary camera can record videos that contain at least 20 - 30 frames per second. Since the individual is not moving quickly, we may drop frames from the video to simplify the processing. We will transform the images/frames into a grayscale format to further reduce redundant processing. A picture comprises pixels, and every pixel stores its RGB color values. Reducing the number of values held by our pixels may minimize our processing by a substantial margin. YOLO (You Only Look Once) realtime object detection algorithm is one of the most successful object detection algorithms that comprises many of numerous concepts coming out of the computer vision research field. Object detection is a fundamental component of autonomous vehicle technology. It is an area of computer vision that's incredibly popular and functioning so much better than only a few years ago [8].

4.5 Android Studio

The particular Android SDK is used to create the application , in which aesthetically impaired users certainly discover objects and understand their environment. This platform is used for the front-side conclusion and backend of the program form. This system includes all the libraries and deals required to put this system together.

4.6 Dataset

The SSD MobileNet model was trained using the Common Object in Context (COCO) dataset, which recognises 80 different categories in this study.

5 Proposed System

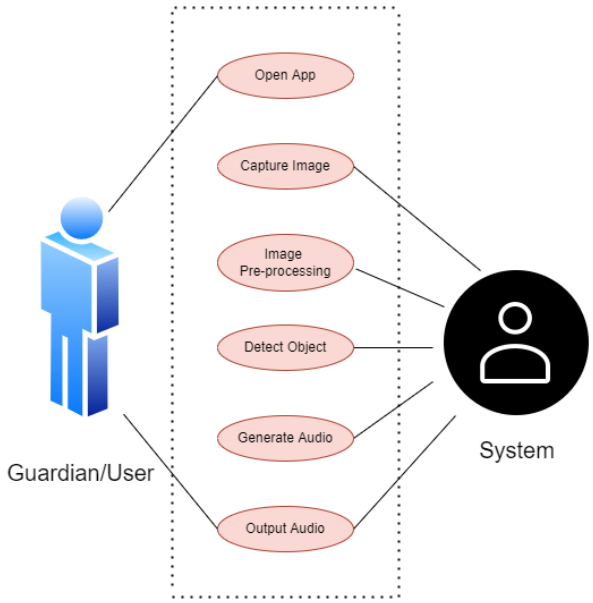

Fig. 5. Proposed System

In figure 5 Many difficulties are faced by the partially visually impaired person while performing basic tasks and most commonly during self-navigation in an unknown environment. They cannot detect the object around them, and in most cases, they cannot identify their objects accurately. So self-navigation difficulties can be overcome by this application, which will help them live everyday life like an average person. In an ordinary case, object detection cannot achieve high accuracy as other applications only detect objects in the captured image rather than object detection in live streaming video. So, the proposed system detects objects continuously in the live stream and not by taking photos every time. For detecting and recognizing objects in the live stream, the processing speed should be very high to detect all objects within the frame, but object detection and recognition are more accessible in the case of captured images. In this system, the android smartphone captures an image from the camera, and it is processed through an algorithm, and the object detection and voice output module produces the required output for the system. In object detection, the detected or recognized objects are first segmented, and where multiple segmented regions are formed, then the image is classified and identified using a fast RCNN algorithm. The voice module converts the detected object constraint area into text and produces output using speech processing. For conversion of text to speech, the system uses Google API. system proposes a system in which an object is detected in the live frame, and the audio output of that detected object in the live frame and the position of the detected object in provided as output to the user.

6 Implementation Details

In this system, the TensorFlow model is used, which in the backend uses the SSD algorithm and can work by speed, balance, and accuracy. This system accurately detects 81 objects approximately. The Mean Average Precision of this system is 74.3 mAP and is the highest among real-time processing models. This system produces speech feedback for the object detected in the frame. The same thing, on the other hand, may be called out multiple times when it is discovered. Even if the detection result is the same, repeating the same object's name is inconvenient. It's also improper if the names of two items are said at the same time, overlapping, or the user is unable to distinguish between them. To overcome this problem, even if one object is recognised in the first frame and is talking, the software will not mention that class for the next two seconds. At the same time, the problem of multiple detection of the same object has been fixed. In this way you get an idea of the exact performance of the model when the objects are correctly detected. Viable objects can be detected immediately, but only objects with an accuracy value more significant than the fixed threshold value will be reported to visually impaired users using output feedback by equals output. Voice. Multiple objects can even be detected precisely at the same time. The system is built by combining some of the technologies described below. Since Android Studio is a full-fledged integrated development environment (IDE) specially designed for Android applications, it is used to create apps. The Android framework allows taking photos and videos with the phone. Camera intent or Hardware.camera2 API. It's a video recording application that recognises objects and interprets text in real time. To create object identification patterns, the Tensorflow library is employed, which allows for rapid numerical computations. It offers a general-purpose design that allows calculations to be used across several platforms with ease. The SSDMobileNetCOCO paradigm allows for real-time video processing. The SSD is made up of a single composite network that learns to predict bounding box positions before the object that was recognised as bounding boxes [7].

The application detects objects using the MobileNetCOCO SSD model. The entire input image uses only one neural network. The network then divides the input image into multiple regions and predicts the boundary regions within the squares based on their probability scores [7]. This object parses images and extracts text from them. It can detect text in all types of images after it has been initialized. Google TexttoSpeech is used to implement the TTS feature, which talks items that have been discovered and identified, as well as providing guidance for those objects.

6.1 Object Detection Pipeline

In Figure 6 the camera will operate in a loop and take an image at a rate of about 2Hz. The picture will run through face detection, and any face boundbe drawn on

the screen as a teaser. The system then selects if it merely detects the largest face in the frame, and the complete image will be given to CNN. If the user opts to detect the largest face, then the head and shoulder region of the biggest face bound will be cropped out, and the new photo will be submitted to CNN for classification. CCN will produce an N dimension vector of confidences of detected features. Feature vector will be input into the logistic regression prediction function, then deliver probability/score for each of the classes.

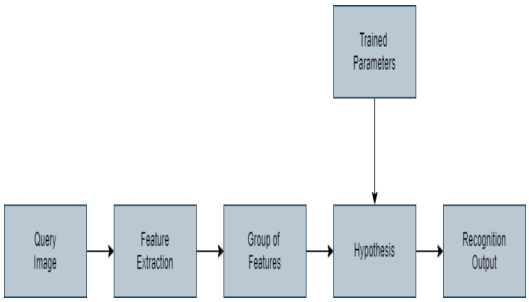

Fig. 6. Object Detection Pipeline

6.2 Convolutional Neural Network

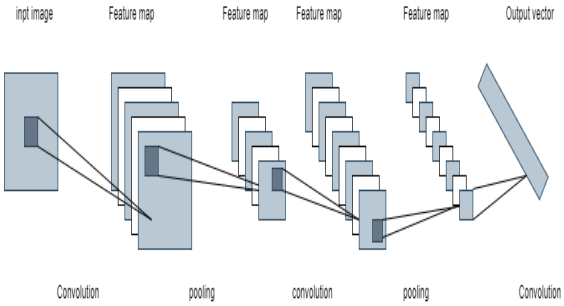

Fig. 7. Convolutional Neural Network [13]

Figure 7 displays a CNN that can recognise 2D shapes regardless of object rotation, scale, or other characteristics. The basic form of convolutional neural network, according to LeCun, is neurobiologically driven and comprises of three key activities:

- Extracting Feature.
- Mapping Feature.
- Subsampling.

7 Results and Analysis

In figure 8 and figure 9 the Tensorflow's object detection model was utilized in this application with the SSD method in the backend, and it works by balancing accuracy and speed. This sample successfully identified about 81 items. This model has the highest mAP value of 74.3 (mean accuracy) among models designed for real-time processing. After completing this endeavor, he was meant to respond to the identified item with a voice. However, the same item is recognized several times throughout time. Even though the detection findings are the same, it is not preferable to mention the same object name. Furthermore, it was unfavorable if two-item names were reported as identical or so near that the user could not tell the difference. If an item is spotted in the first picture and spoken about, this difficulty may be solved. The application will not speak about its class for the following five seconds, even if discovered. As a result, the difficulty of detecting an item many times has been addressed. Here are some examples of forecasts about an object's precise observed value. Please give us an idea of the model's accurate performance when detecting things. Multiple items can be recognised at once, but only those with a precision value better than a threshold value will be speech output feedback to visually impaired users. It is even possible to accurately identify many things at the same time.

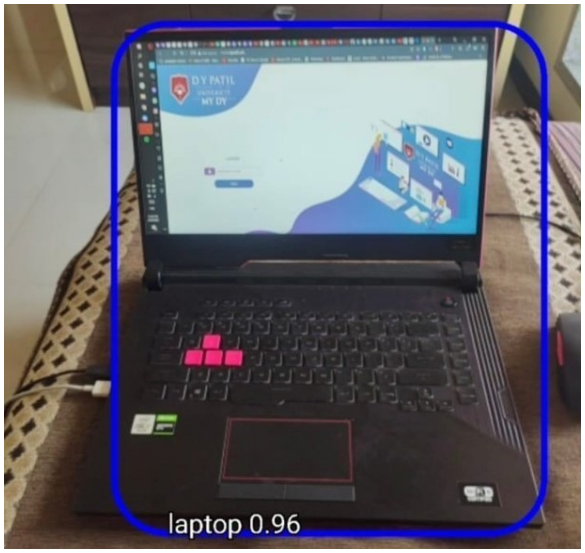

Fig. 8. Detection Of Objects

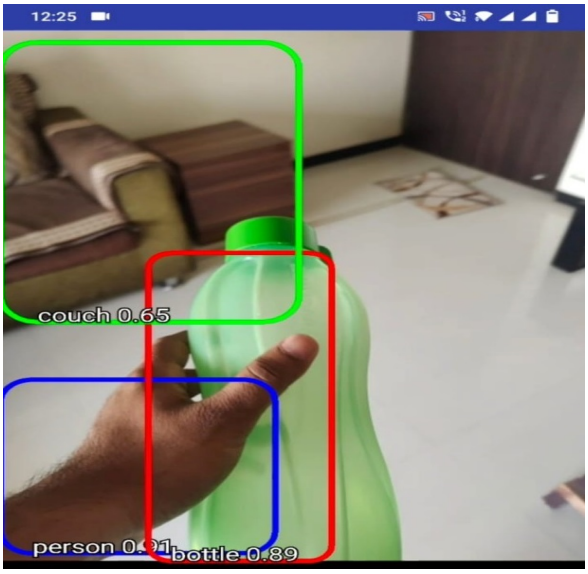

Fig. 9. Detection Of Objects

8 Conclusion

This study presents a computer vision-based navigation system for blind people and attempts to solve the problems presented in the introduction section. The suggested method is put to the test in various scenarios and against comparable apps on the App Store. The findings show that the application runs well under various conditions and is quicker and more efficient than its competitors. The program is also somewhat accessible to blind users since it enables them to launch it simply by putting an earphone jack into their mobile phones, which is not available in many other apps. This program can be used for navigation in its present version, but it has a few restrictions. This paper proposes a system that assists partially blind persons in navigation and detects the object surrounding of user. The proposed system is tested with many conditions and against applications of similar interest. It is noticeable from the observations that the applications perform pretty under the circumstances and is efficient, balanced, and fast. The application is easy to access for a partially blind person, and as the application provides output in the form of audio, it makes it more efficient for a partially blind person who is missing in other applications. In the current state of the application, this application can be used for navigation but has a few limitations. Being optimistic that it would cause people to be sympathetic with visually impaired people and they understand that even they should be able to live life like a normal person.

In the future, GPS can be added so that the device can guide the user where they want to travel and the device would direct them there. Use of the proximity sensor that will inform the visually impaired person of the object's or person's distance. An integrated reading mechanism, in which we can upload a book to the gadget and have it read by the device. Aside from that, device's battery consumption optimization can be imporved.

References

1. WafaM.Elmannai,KhaledM.Elleithy. “A Highly Accurate and Reliable Data Fusion Framework for Guiding the Visually Impaired”. IEEE Access 6 (2018) :33029-33054.
2. Qi-Chao Mao,Hong-Mei Sun,Yan-Bo Liu,RuiSheng Jia. “MiniYOLOv3:Real-Time Object Detector for Embedded Applications”.IEEE Access 7 (2019) :133529-133538.
3. ZhenchaoOuyang,JianweiNiu,YuLiu,MohsenGuizan i. “Deep CNN-Based Real-Time Traffic Light Detector for Self-Driving Vehicles”.IEEE Access 19 (2019):300-313.
4. Wei Fang,LinWang,Peiming Ren. “Tinier-YOLO:A Real-Time Object Detection Method for Constrained Environments”.IEEE Access 8 (2019) :1935-1944.
5. VidulaV.Meshram,KailasPatil,VishalA.Meshram, Fe lix Che Shu. “An Astute Assisstive Device for Mobility and object Recognition for Visually Impaired People”.IEEE Access 49 (2019) :449-460.
6. WenmingCao,JianheYuan,ZhihaiHe,ZhiZhang,Zhi q uan He. “Fast Deep Neural Networks With Knowledge Guided Training and Predicted Regions of Interests for Real-Time Video Object Detection”.IEEE Access 6 (2018): 8990-8999.
7. QiankunLiu,BinLiu,YueWu,WeihaiLi,Nenghai Yu. “Real-Time Online Multi-Object Tracking in Compressed Domain”.IEEE Access 7 (2019): 76489-76499.
8. Meimei Gong,Yiming Shu. “Real-Time Detection and Motion Recognition of Human MovingObjects Based on Deep Learning and Multi-Scale Feature Fusion in Video”.IEEE Access 8 (2020):25811- 25822.
9. Kiruthika, U., Somasundaram, T. S. (2018). Efficient agent-based negotiation by predicting opponent preferences using AHP. Journal of applied research and technology, 16(1), 22-34.
10. N. Senthil kumar, A. Abinaya, E. Arthi, M. Atchaya, M. Elakkiya, “SMART EYE FOR VISUALLY IMPAIRED PEOPLE”, International Research Journal of Engineering and Technology, Volume: 07 Issue: 06, June 2020.
11. Liang – Bi Chen, Ming-Che Chen, “An implementation of an intelligent assistace system for visually impaired/blind people, ”IEEE, 2018.
12. Shahed Anzarus Sabab, Md. Hamjajul Ashmafee, “Blind Reader: An IntelligentAssistant for Blind”, 19th International Conference on Computer and Information Technology, December 18-20, 2016, North South University, Dhaka, Bangladesh.
13. M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. Awwal, and V. K. Asari, “A state-of-the-art survey on deep learning theory and architectures,” Electronics, vol. 8, no. 3, p. 292, 2019.
14. Shreyash Patil, Oshin Gawande, Shivam Kumar, Pradip Shewale,“Assistant Systems for the Visually Impaired”, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 01 — Jan 2020.
15. Gagandeep Singh, Omkar Kandale, Kevin Takhtani, Nandini Dadhwal, “A Smart Personal AI Assistant for Visually Impaired People”, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 06 — June 2020.
16. Mingmin Zhao, FadelAdib, Dina Katabi Emotion Recognition using wireless signals.