Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

Developing Methodologies for Navigation using Multimedia to help the Visually Impared

[1]Vinod Biradar, [2]Karuna C. Gull

[1]Research Scholar, Department of Computer Science and Engineering, S. G. Balekundri Institute of Technology, Belagavi, Visvesvaraya Technological University, Belagavi-590018, Karnataka, India
[2]Professor, Department of Computer Science and Engineering, S. G. Balekundri Institute of Technology, Belagavi, Visvesvaraya Technological University, Belagavi-590018, Karnataka, India

Abstract: The objective of research into navigation assistance is to facilitate autonomous living for individuals who are visually impaired. Although numerous navigation aids employ contemporary technologies and methodologies, they do have certain drawbacks, including portability, object detection, convenience, and the need for extensive training. When developing a navigation aide for individuals with visual impairments, portability and user-friendliness without the need for extensive training are additional crucial factors. Certain navigation systems might fail to furnish precise details regarding the nature of hurdles that can be identified, a critical component for making well-informed decisions while in real-time travel. This research proposes the EfficientNet with entirely contorted Convolutional Neural Networks, for the development of a navigation assistance system in order to circumvent all of these challenges. The system's EfficientNet is a portable object detection model that permits the execution of multiple operations, including uniform expansion, resolution, depth, and breadth. The completely convolutional CNN, a pre-trained model, finds extensive application in object detection and computer vision tasks. The evaluation of the proposed system is conducted using the metrics of Distance, Mean Average Precision (mAP), and Accuracy.
Keywords: EfficientNet, Fully-Convoluted Convolutional Neural Network, Navigation Assistance, Object detection, Visually Impaired Persons.

1. Introduction

According to the World Health Organisation (WHO, 2021), the global population of individuals with vision impairments or blindness is estimated to be over 2.2 billion. Numerous research works have shown the challenges encountered by those with vision impairments while navigating. One aspect of a dignified existence is the ability to move about freely, and the main driving force behind this activity is the provision of instruments to facilitate that mobility. Individuals who are visually impaired usually rely on assistance from sighted individuals, guiding dogs, and white canes to help them navigate. When someone with sight assists someone with a visual impairment, the person feels secure and at ease. However, depending on someone else to guide you might prevent you from being independent.

Numerous navigation assistance systems have been suggested in the literature to solve difficulties like time-consuming training, portability issues, costly and continual dependency on a second person, etc. Certain products were made to be used indoors, while others were made to be used outside. But a lot of these gadgets are unpleasant, and adopting these solutions makes people feel like they're being socially shamed[1].

Researchers have looked at a numerous technologies, like artificial intelligence, to identify navigational aids that work for those who are visually impaired. Recently, there has been a growing exploration of deep learning models for navigation aid systems' obstacle detection[2-5]. Even if there are a lot of object detection models available, choosing one that is appropriate for a real-time navigational environment that requires minimal memory footprint and short inference time requires rigorous research and analysis. Many of these technologies are uncomfortable, and using them makes individuals feel embarrassed[6].

Researchers have examined machine learning alongside artificial intelligence to find visually impaired navigating assistance. Recent research has focused on deep learning frameworks for navigation assistance system obstacle detection[7]. Even though there are many object identification models, selecting one for real-time navigation with a small memory footprint and fast inference time takes much study and analysis.

A webcam and computational algorithms are used in the second approach. Even without a signal, the object sends data. After taking images, data is deleted and reviewed. Images may reveal pixel count, colours, and more. AI, deep learning, machine learning, and various additional technologies enable this. Computer vision is essential[8].

The proposed system should compute and display the user-camera gap when the user approaches a camera. This can be done using computer vision. OpenCV captures video, recognises faces, and converts coloured pictures to grayscale. Computer vision simplifies many computer science tasks[9].

The proposed device can recognise a moving human face, which has economic potential and is necessary for security surveillance. Distances are easier to calculate using the camera's focal length. Calculating research distance between fixed camera and detected face.

The screen displays item distances. Use of machine learning-based integrated OpenCV facial recognition to let users monitor their distance from the display. Few people utilise hardware tools.Lasers and radar aren't used in the system. You just need a camera to capture video. Users may talk to the system. This article proposes a novel mobile application solution that can be loaded on an Android laptop or mobile device. via the use of mobile phone cameras, the programme will allow users who have a direct visual impairment to explore the area via spoken messages. It also has the ability to recognise items by name. The system's architecture primarily makes use of a number of deep learning algorithms to identify objects and determine their proximity to the camera[10-12]. He or she must not only identify items but also see the surroundings and determine the direction and distance of objects. As a result, he or she will give orders for individuals to relocate. Lastly, the technology will provide voice control, enabling visually handicapped individuals to learn how to navigate their surroundings. Furthermore, it labels the items in front of individuals so that the partly sighted person may comprehend what they are looking at. On a worldwide scale, it is a new work. It is economical and accessible to all societal groups. With the current approach, a visually impaired individual would get audio instructions for the required job on a typical Android smartphone that has been loaded with the software and is fastened to their chest strap.

The following explanation will walk you through how the succeeding parts of the paper are organised. In the second section, a general summary is presented of the ongoing difficulties that people who have visual impairments have to deal with, as well as the potential remedies that have been suggested. Furthermore, this study encompasses an examination of scholarly literature pertaining to assistive technology, as well as an exploration of the seminal works that have significantly contributed to the advancement of research in this domain. Section 3 presents a comprehensive design for a mobile application, including a detailed examination of its numerous components and functionalities. The text provides a comprehensive explanation of the many algorithms that are available to do the necessary tasks. Section 4 presents the analysis of the experimental results. In conclusion, Section 5 serves to conclude the article and provide a summary of potential avenues for further research.

2. Literature Survey

There have been significant efforts made, with proper attention given to the relevance of making use of modern technological breakthroughs, to fulfil the requirements of the visually impaired people around the world. Researchers working in the area of computer vision have had access to a fresh viewpoint as a result of the execution of the Yolo project, which has resulted in the creation of a number of different applications.

This section reviews the work and solutions established for visually impaired people. In work [13] conducted a tactile memory set experiment using multimodal navigation assistance, including visual feedback as well as high contrast auditory signals. The gamepad is designed for use with the game. Blind children were taught memory games using vibrations instead of sounds or embossed graphicsThree sorts of auditory cues have been created for fully blind individuals to guide them. Visual enhancement is achieved via Augment Reality (AR) and traverse steering integration for visually impaired individuals. Smart guiding glasses enable vision challenged individuals in mobility[14]. This multi-sensor system employs ultrasonic sensors to identify impediments. Additionally, augment reality enhances visualisation. It will let visually challenged people move freely without problems.

Using 3D things in robot movements in the surroundings may help visually impaired individuals. The method divides points into many flat patches and removes inter-planar interactions . The procedure creates several High-Level Features for each patch according to object template IPRs. Each patch on the plane is assigned to a particular object pattern after being run through a plane classifier that is based on the Gaussian mixing pattern. The last step is recursively combining organised plans into model objects[15]. The recommended technique recognises objects based on their geometrical surroundings, making it resistant to visual alterations. Unfortunately, this technology can only identify structural items and cannot comprehend non-structural ones.

The RGB and depth images were processed by segmentation networks in order to extract semantic information from the data. Individuals who are visually impaired will be able to avoid obstacles with the help of the gadget since it offers accurate input [16].A disadvantage of the system might be its lack of mobility.

A Kinect-based navigation assistance using infrared depth measurements. The corner identification algorithm detects barriers in photos, and the depth detector measures their distance[17]. Users would benefit if the system provided obstacle information.

An assisted navigation system using a Microsoft Kinect level sensor. Speed-Up Robust Characteristics (SURF) model extract features for obstacle identification. Users would hear sounds using headsets[18]. The technology might give barrier information, however portability may be a problem.

Real-time video was captured using a USB camera. The device recognised impediments, measured distances employing a laser, and gave users audible feedback. Multimodal fusion-based quicker RCNNs detected and classified obstacles[19].

The device provides audible feedback and shares the location with family/friends employ Raspberry Pi-based YOLOv3 for obstacle detection. The technology recognises items and provides real-time audio feedback[20]. These systems do not include support for navigation-related features such as estimating distance, recognising scenes, or detecting moving obstacles, all of which might potentially enhance navigation. Despite the fact that the systems are brand new, the obstacle detection architectures that they use are not the most effective deep learning models in terms of accuracy, deduction length of time, and deployment in small portable devices such as smartphones.[21-25]The user interface of the system was created in such a way that it enables users to input their navigation purpose, and the system then provides intelligent directing via a succession of destination targets that are dependent on environmental processes.

RFID/BLE installation is expensive and complicated. In outdoor navigation, such a method is impractical[26]. The system's mobility and indoor operation are its key drawbacks. Hardware-based navigation aides are cumbersome and immobile. Power supplies and wiring make hardware computing boards as well as external cameras cumbersome to carry while travelling[27-29].

A transformer-based identification and recognition model for smart glass using computer vision as well as deep learning[30]. An remote server linked to a smartphone executes these models in the proposed system. The scientists say the algorithm can identify impediments in low-light and dark-scene photos to help users at night.

Smartphones were explored more in navigation aid research due to their rapid expansion. The navigation aid system relied on smartphone voice input. Stereo cameras sent environmental footage to a distributed computing infrastructure[31]. To operate, the system needed continual data network connection.

A smartphone app-integrated picture recognition system [32] offers online and offline operation depending on network availability. The information was delivered to a system on the back end after being acquired by a smartphone while it was in immediate reach to the user. The offline system uses the quicker R-CNN algorithm for improved accuracy. Instead, the algorithm known as YOLO was used online to speed up processing. User would get audio results after recognising barriers and their distance. The smartphone had a lightweight CNN that was able to determine the position and orientation of the obstacles [33]. The experts said the arrangement was tested indoors and outside, yet the connections interconnecting system components may impair navigation.

A real-time mobile application that uses machine vision, a 2D map, as well as the smartphone's IMU to estimate and monitor the user's interior position[34]. The app also requires an annoyance — walking with the smartphone held or worn so that the camera faces front.

Many portable solutions need network access and server processing. Navigators who use the internet may worry about security and confidentiality. Some have developed more accurate general-purpose object identification models, but they have not proven how to combine them into an adaptable framework with limited processing resources[35]. Object and obstacle detection techniques designed for real-time use need to have a fast inference time without sacrificing precision. Models of Objections, in General may be accurate yet slow to produce results, which might cause navigation mishaps owing to reaction time delays.

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

An autonomous walk guidance system in [36], which employs ultrasonic sensors to detect potholes and obstacles. The approach uses CNN. The front detector has 98.73 percent accuracy and 1.26 percent inaccuracy when obstacles are 50 cm away, according to experiments. This inquiry was limited by using an ultrasonic instrument and an above camera to photograph the road or items[37]. This work has drawbacks including ultrasonic equipment and an above camera to capture road or item images. The step also vibrates when an object is spotted.

3. Proposed Architecture

We examined our literature assessment and suggestions to determine navigation system limitations. We also evaluated our research on current cellphones' navigation aide capabilities. These investigations led us to numerous design elements that might enhance visual impairment users' navigation.. The user was an integral element of the evaluation of needs and subsequent testing. This section details our proposed system's architecture and implementation. Due to its excellent classification performance on a wide variety of data types, deep learning has become a prominent subject of machine learning. Convolutional neural networks (CNNs) are an effective deep learning method for labelling images. CNN, an image processing deep learning system, utilises pictures as input. The algorithm, which gathers and classifies visual characteristics using several operations, has become popular. These models from this work are detailed below.

3.1.R-CNN

The R-CNN architecture identifies image object types and bounding boxes. The R-CNN model selects visual object candidates via selective search. Selective search prioritises small regions from small to large. Equal regions get larger when merged. Recursion continues. Iterations create meaningful regions and group image items. 2,000 regions are separately entered into a CNN model to predict classes and bounding boxes. Selective search identifies R-CNN region potential. These district candidates feed CNN. After region nomination, CNN networks and about 2,000 regions are formed. Absence of a feature map lowers disc space. The softmax layer replaces SVM in area recommendation classification, improving speed and accuracy. The selective search technique for the R-CNN framework slows the Fast R-CNN model.Region proposal network (RPN) replaces selective search in Faster R-CNN.

3.2.EfficientNet

EfficientNet employs compound scaling to improve neural network performance. EfficientNet reduces parameters and FLOPs to boost performance and computational efficiency.Since there are many methods to scale up in CNN designs, raising layers is difficult. Manually selecting the ideal mixture is time-consuming. Compound Scaling and NAS (Neural Architecture Search) help EfficientNet scale up, as detailed in figure 1.

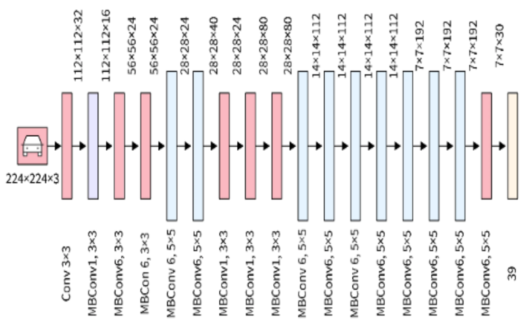

Fig 1: EfficientNet Working

3.3.YOLO

YOLO technique is named from the phrase "You Only Look Once," meaning to only glance once. YOLO can quickly identify and locate things in a picture. The YOLO approach is popular because to its excellent accuracy and real-time . The approach "looks only once" at the picture, requiring only one neural network forward propagation pass for prediction. Using non-maximum suppression, the object detection technique detects each item just once and outputs the recognised objects and bounding boxes. One CNN predicts several bounding boxes and class probabilities using YOLO. Full photos may be optimised using YOLO for detection.

YOLO uses CNN to execute these processes. YOLO model architecture includes 24 convolutional layers and two fully linked levels. A 7x7 (SxS) grid arrangement is used in the design.It requires 448x448x3 photos as input. Architecture generates 7x7x30 output.

The YOLO method has evolved. The initial version of YOLO V1 architecture, only supports the identical input quality during testing as the training picture due to the completely connected output layer. To improve upon YOLO V1 and maintain its popularity, the v2 architecture is more precise, quicker, more powerful.

The following are some design aspects that we discovered to be important in a navigation assistance for the visually impaired:

- Low-latency systems are thought to provide the greatest user experience.
- Complete data privacy is assured for both the user as well as the environment entities.
- It helps keep implementation costs down.
- They may help lessen the need for electrical energy.
- Many navigational aids omitted the important design consideration of portability.

Five components of this system are in coordination with one another. The components consist of the following: (1) Object detector; (2) Component for distance computation; (3) Component for direction computation; (4) Component for decision making; and (5) Component for voice commands. The sections that follow describe the roles of these components as illustrated in figure 2.

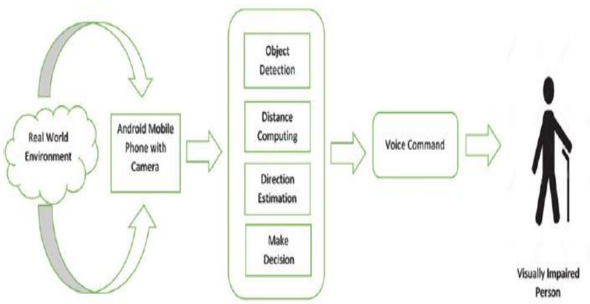

Fig 2: Proposed visually impaired mobile app architecture

The person in charge faces the camera first.Increases eyesight, focus length, and system input. Face is identified in every photo when the user advances in front of the camera.User-camera separation predicted. Face is rectangle-surrounded. The rectangle calculates focal length and distance from perspective width and height.The camera length of focus is calculated by admin to measure distance. Go before the camera. Cameras detect faces in every frame. Perspective width, height, as well as face detection determine the face's bounding box. Later, the length of the focal point is found out.To see how the distance varies as the user moves both forward and reverse on the screen, measure the distance in inches.

3.4.Detection of Objects

Figure 3 depicts the Object Detection system's capabilities. It depicts three vertical phases in object training and prediction. The first set of stages is for Data Construction, which is for creating a new collection of items that will be educated by the system. The second set of steps is for model training, while the third set is for object prediction.

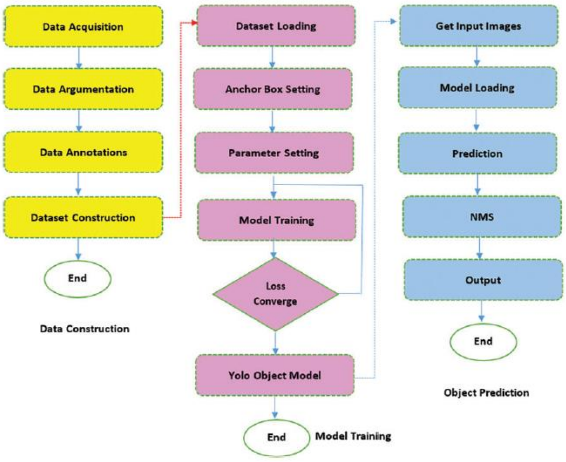

Fig 3: Standard Yolo 3 object prediction stages

The first two stages are unnecessary in this situation since we are using the MS-COCO (Microsoft Common Objects in Context) training dataset. The third stage was completed on its own.It uses photos from the surroundings to detect items. The EfficientDet model is tested on the COCO dataset, a general-purpose object/obstacle detection task. EfficientDet model outperformed similar-sized models in benchmark data sets with improved mean average precision (mAP) utilising fewer parameters and processing. Therefore, the model is quicker on GPU as well as CPU than other motion detectors.

The object detection technique in Yolo systems is done out employing the following procedures.

1. Recognising and categorising the things in the image.
2. Identifying the things in the photograph.
3. Detection of objects.

Non-max suppression selects the optimal limit box for an item while rejecting or "suppressing" any others. The NMS (Non-Maximal Suppressing) takes two elements into account.

1. The objectiveness score is determined by the model.
2. The IOU (bounding box intersection)

The goal is to identify items in every SS grid cell using the provided picture.

Each cell predicts B bounding boxes and confidence. Whether an item is in the grid cell along with the IoU of GT and projections may be determined by confidence. See Fig. 4 for the Yolo3 prediction model. Equation (1) expresses confidence:

Confidence = Prob *(*Object*)* × IoU *(*Ground Truth, Predict*)* (1)

where Prob(Object) ranges between 0 and 1

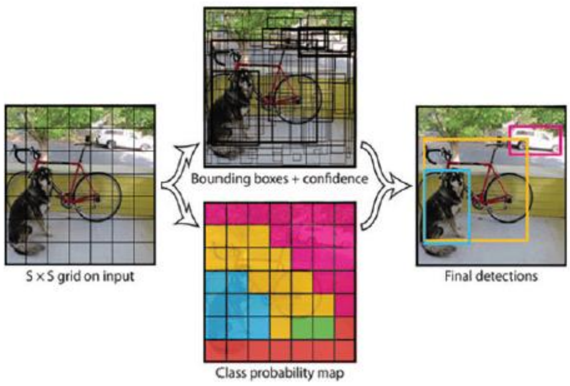

Fig 4: Yolo3 prediction model

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

3.4.1.Data Flow diagram

As depicted in the utilisation case diagram, the visually impaired user engages with the system. Five use cases exist, each of which functions within the system. As illustrated in Figure 3, it provides valuable insights into the correlation between spoken word use cases and object A predetermined distance D is maintained by the administrator standing in front of the surveillance camera. Video in real time is recorded by the camera.

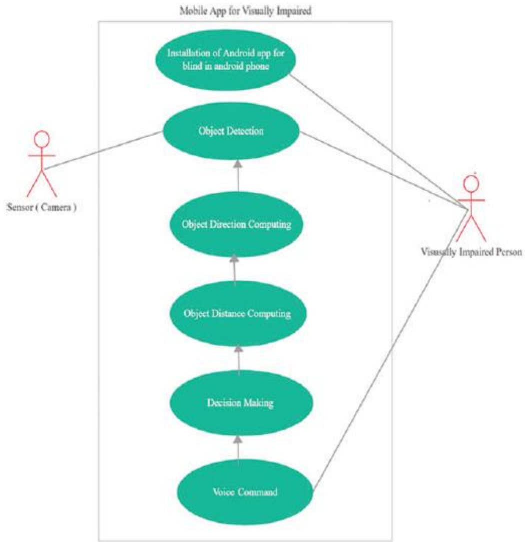

Fig 5: Use case diagram for visually impaired person mobile app

The programme receives the frame-by-frame output from the camera. determines the height and width of the perspective while detecting the visage. A perimeter is delineated around the identified visage. This perspective width is utilised by the administrator in order to compute focal length. Admin inputs this focal length into the system. Currently, the system is evaluating its focal length. As the user approaches the camera, they initiate motion. The acknowledged breadth, denoted as "standard width of the face" (w). The programme calculates the viewpoint width and height after identifying the user's visage and assigning it the value P, which represents the perspective width.

3.5.Estimate of distance

In video, distance estimation is the process of calculating the separation between a subject and the camera. Live video is captured by the camera as an individual walks in proximity to it. Footage is captured when a subject appears in front of the filming device and initiates motion in its direction. The live footage that was recorded is incorporated into a video frame pattern. These structures are addressed in a separate manner. For each frame, a face recognition technique is implemented. The detected visage is surrounded by a rectangle. The rectangle encircling the detected visage is utilised to compute its height and breadth. The term for this is the "perspective width." Focal length is determined by utilising the perspective breadth. The proposed system is determined by the focal length.

Researchers have tested distance estimate systems using external sensors and cameras. We studied smartphone-only distance calculation techniques for our design. We examined Law of 57 for our system's distance estimate module. Since the navigation aid may use the camera on the smartphone alone, the approach is portable and convenient.

The Rule of 57 states that an impediment with an angular dimension of 1◦ is 57 times further away than relevant (see Fig. 3). Thus, an obstacle's angular size (in degree) should match its real size to the 360-degree circle's diameter at that distance away from the viewer. Astronomical telescope picture distance and angles were used to develop this approach (Harvard, 2012)[39]. To calculate distances using telescopic pictures, you must understand that an obstacle's perceived angular size relates directly to its dimensions and its distance from the viewer. As the spectator moves away, the impediment seems smaller. Experiments indicate that it can measure obstacle distance even if the impediment is larger than 1◦ from the smartphone camera sensor's screen.

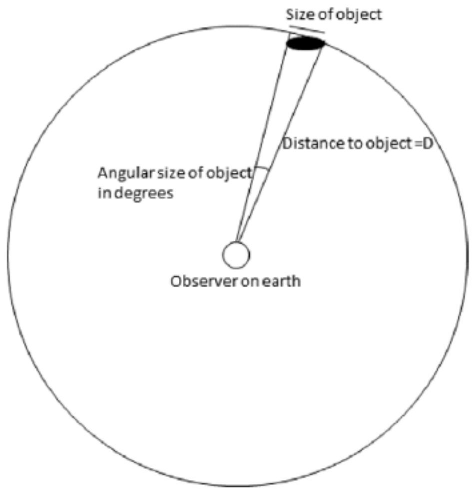

Fig 6: The Rule of 57.

Source: **Adapted from Harvard (2012).**

3.6.Estimating the position

For each impediment detected, the obstacle detecting model generates an array of four integers [top, left, bottom, right] that delineate a bounding rectangle encompassing the corresponding position. After dividing the image space into three different areas (left, centre, and right), the region that the detected obstacle occupies to the greatest extent is determined. As its position, the position calculation module outputs the area encompassing the obstacle's centre.

3.7.Motion monitoring

It consists of three primary stages. In the beginning, the image undergoes pre-processing through the implementation of simple spatial filtering using Gaussian and median filtering. For background modelling, a dual-mode SGM (single Gaussian model) is then implemented. Ultimately, the background movements are compensated for with modified motion utilising Kanade–Lucas–Tomasi (KLT) through the combination of models. It was demonstrated through testing on a smartphone that this method computes results faster than comparable methods.

3.8.Scene identification

The EfficientNet-Lite44 model was implemented using the method of transfer learning on a set of 20 custom scene classes that are representative of typical navigation environments both indoors and outdoors. A comprehensive description of the scene classes is provided in Section 3.3.7. In addition, to reduce the occurrence of false positives associated with unfamiliar scenes, we integrated a threshold parameter into our implementation. A scene is classified as unknown by the module when the probability value falls below a specified threshold. 0.7 was the finalised threshold following trial and error.

3.9.Output

The textual information obtained from multiple modules regarding the obstructions and the scene is converted to audio format by the output module. This audio is then transmitted to the user via a bone conduction earpiece. The output module implemented Pyttsx5, which is a Text-to-Speech (TTS) library for Python, to convert textual data into audio output. Offline, Pyttsx operates flawlessly across multiple platforms.

3.10.Customised datasets

Twenty distinct categories of impediments pertinent to indoor as well as outdoor navigation environments were compiled into a bespoke dataset for our obstacle detection module. Bench, bike, billboard, bookcase, cabinetry, automobile, chair, dog, door, fire hydrant, kitchen appliance, furniture, person, plant, stairs, traffic light, table, tree, and waste container are among the objects in question. Images from four distinct sources were utilised to compile the dataset: Google Open Pictures V6, ImageNet, the LISA Traffic Sign Dataset, and our own collection.

Google Open Pictures V6 is mostly used for object identification and segmentation research. ImageNet is crucial to computer vision as well as deep learning research. The LISA Traffic Indicator dataset contains annotated US traffic sign frames in pictures and movies. The 20 obstacle class images were derived from various sources. Many extracted pictures need the preprocessing phase such as relabeling. We also manually labelled locality photos using other techniques.6 We selected several common indoor and outdoor navigation scene categories for the scene recognition module. Images from MIT's Places365, Google Open Images V6, and Flicker make up the collection.7 Adding actual location photos to the dataset increases the quantity of images and improves model performance.

3.11.Android application

The app has a voice assistant. The voice assistant activates the app based on two voice commands. Activate Navigation activates the app in navigation mode as well as scans the surroundings. Real-time audio alerts from the app describe the hurdles ahead. The smartphone screen shows recognised impediments and relevant text from the app. Use the Identify scenario voice command to get the navigation environment scene from the app. Then the scene recognition module will recognise the scene and produce sounds again. Fig. 5 shows our android app in action. Fig. 6 shows app results from obstacle detection as well as scene recognition modules.

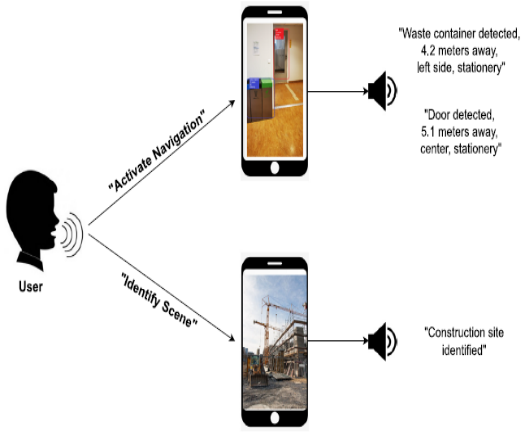

Fig 7: Working Process of the Application

4. Experimental Result Analysis

This part talks about the tests that were done to see how well the five main modules (obstruction identification, scene acknowledgment, distance estimating, detection of motion, and position estimation) worked. The next part talks about how the user testing was done and what the feedback was based on the pilot test using a user. Then, our guidance aid is put up against other systems that do similar things in terms of features and functions.

We trained obstacle detection and scene identification deep learning models using an Intel Xeon CPU with 64 GB RAM with an NVIDIA GeForce GTX 1080 Ti GPU. Test platform parameters include TensorFlow-GPU 2.4, NVIDIA CUDA toolkit 11.0, along with CUDNN 8.1. The data set is randomly shuffled and divided 80:10:10 to train, verify, and test the models. The accuracy metrics for the obstacle detection and scene identification models are shown in Tables 1 and 2. Few obstacle detection module tests are shown in Fig. 7. The findings show how the model finds impediments. The obstacle identification model is 87.8% accurate. Most learned obstacle dataset photos are accurate (above 80%), however cabinets and stairs require better (see Table 1). Fig. 7 shows the model's performance in several tests.

Table 1: Distance Estimation and Accuracy Percentage

Obstacle	Distance estimated when the actual distance					
	5m	Acc %	Err %	10 m	Acc %	Err %
Bench	2.7	99	1	4.9	99.3	1
Chair	2.5	99.3	1	3.9	97.3	2.66
Person	2.7	97.3	2.66	3.9	99.1	1
Billboard	1.9	99.1	1	3.9	99.3	1
Fire hydrant	2.2	99.3	1	3.6	99	1
Bookcase	2.5	96.3	3.9	4.2	99.3	1
Table	2.2	99.1	1	4.3	99	1
Waste container	1.9	97	3.1	3.8	99.3	1
Door	2.3	98	2.1	4.5	97.3	2.66
Tree	2.8	99.1	1	4.6	99.1	1

Table 1 shows the low mistake rate and good precision. This indicates that the distance calculation function is effective.

The model can recognise most obstacles. It may have been missed by the model due to the image's closeness to stairs. Both barriers are similar in colour and other aspects, which may have led the model to provide an inaccurate conclusion. Through studies, obstacles of various sizes were set at four distances (1, 3, 5, and 10 m) to assess the distance estimate module. Distance estimates from the distance calculation module are compared against group truth. Motion detection is tested using moving objects. Users will be notified of obstacle movement status, kind, distance, etc.System aspects are studied. 1. Object identification mode identifies various items. 2. Subject Distance Mode provides camera-object distance information. 3. Navigation Mode lets users navigate in specific environments.Showing the thump-up in the distance when the camera switches models. Therefore, assessing system operation in all three modes is crucial.

The navigation mode allows the system to provide commands while navigating around the surroundings. Second, the system can recognise items learned in the dataset created by COCO in object identification mode. Therefore, Precision, Recall, F1, and Accuracy are tested to accurately identify objects in the environment. Third, identifying the distance between the object and camera is a challenge in distance finding mode. The system measured distance and real distance can be contrasted for hundreds of items. In addition to the three analyses, two more are conducted to measure extra performance. Fourthly, usability aspects such as user pleasure, comfort, auditability, learnability, and usefulness are assessed via input from visually impaired individuals. The product is assessed for speed, weight, cost, and comfort to guarantee practical usage. These assessments demonstrate that the system is extremely satisfying and user-friendly.

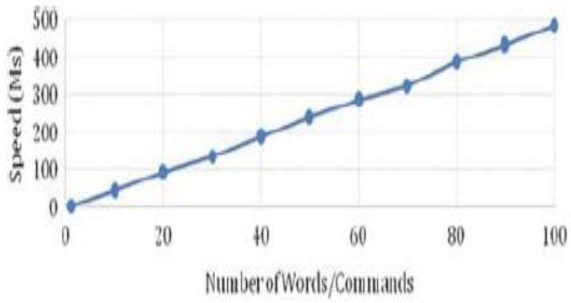

Fig 8: Number of words/commands *vs.* speed

5. Conclusion

The main objective of this project is to develop a smartphone app that helps vision-impaired individuals navigate their surroundings. Distance computing and object recognition are the main components of the created system. The item identification system will provide the names of objects. The distance sensing system guides users to move in the following directions: Proceed Front, Proceed Left, Advance Right, switch Left or Right, Proceed Front, Proceed Left, Move Right, or Do Not Move. Ultimately, the goal is to aid visually impaired and elderly individuals in identifying things and navigating their surroundings by using speech commands to determine distance. The project uses Yolo3 methods for object detection and distance computation. The device can only detect objects in front of the visually impaired individual and determine their distance from them. The mobile app is suitable and easy to use.Although the technology has limitations such as comprehensive route recognition and virtual mapping, obtaining the intent of the disabled person remains a major difficulty for the smartphone application system. The future plan may adapt to these developments.

References

[1] R. K. Katzschmann, A. Brandon and R. Daniela, “Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 3, pp. 583–593, 2018.

[2] Y. Idesawa, M. Takahiro, S. Masatsugu and O. Junji, “Effective scheme to control multiple application windows for screen reader users with blindness,” in *2020 IEEE Int. Conf. on Systems,Man, and Cybernetics (SMC)*, Toronto, Canada, pp. 2674–2679, 2020.

[3] K. Bhowmik, A. F. David, A. Phillip, B. Hal, N. C. Robert *et al.*, “Hear, now, and in the future: Transforming hearing aids into multipurpose devices,” *Computer*, vol. 54, no. 11, pp. 108–120, 2021.

[4] M. Butler, M. Keri and E. R. Susan, “Lighting prescriptions for low vision,” *Journal of Housing For the Elderly*, vol. 33, no. 2, pp. 189–203, 2019.

[5] K. Papadopoulos, C. Konstantinos, K. Eleni, K. Georgios, S. Rainer *et al.*, “Environmental information required by individuals with visual impairments who use orientation and mobility aids to navigate campuses,” *Journal of Visual Impairment & Blindness*, vol. 114, no. 4, pp. 263–276, 2020.

[6] F. Faisal, H. Mehedi, S. Samia, M. Z. Hasan and A. H. Siddique, “Voice activated portable braille with audio feedback,” in *2nd Int. Conf. on Robotics, Electrical and Signal Processing Techniques (ICREST)*,Dhaka, Bangladesh, pp. 418–423, 2021.

[7] S. K. Jarraya,W. S. Al-Shehri and S. A. Manar, “Deep multi-layer perceptron-based obstacle classification method from partial visual information: Application to the assistance of visually impaired people,” *IEEE Access*, vol. 8, pp. 26612–26622, 2020.

[8] M. M. Islam, S. S. Muhammad and B. Thomas, “Automated walking guide to enhance the mobility of visually impaired people,” *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, pp. 485–496, 2020.

[9] W. J. Chang, B. C. Liang, C. C. Ming, P. S. Jian, Y. S. Cheng *et al.*, “Design and implementation of an intelligent assistive system for visually impaired people for aerial obstacle avoidance and fall detection,” *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10199–10210, 2020.

[10] P. J. Duh, C. S.Yu, F. C. Liang-Yu, C.Yung-Ju and C.Kuan-Wen, "V-Eye:Avision-based navigation system for the visually impaired," *IEEE Transactions on Multimedia*, vol. 23, pp. 1567–1580, 2021.

[11] M. A. Khan, P. Pias, R. Mahmudur, H.Mainul and M. A. Ahad, "An AI-based visual aid with integrated reading assistant for the completely blind," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 6, pp. 507–517, 2020.

[12] S. Mahmud, H. S. Redwanul, I. Milon, L. Xiangxu and K. Jong-Hoon, "A vision based voice controlled indoor assistant robot for visually impaired people," in *2020 IEEE Int. IOT, Electronics and Mechatronics Conf. (IEMTRONICS)*, Toronto, Canada, pp. 1–6, 2020.

[13] C.Wan-Jung,C. Liang-Bi, H. Chia-Hao,C. Jheng-Hao,Y. Tzu-Chin *et al.*, MedGlasses:Awearable smartglasses- based drug pill recognition system using deep learning for visually impaired chronic patients,"*IEEE Access*, vol. 8, pp. 17013–17024, 2020.

[14] J. Connier, Z. Haiying, D. V. Christophe, L. Jian-Jin, S. Hongling *et al.*, "Perception assistance for the visually impaired through smart objects: Concept, implementation, and experiment scenario," *IEEE Access*, vol. 8, pp. 46931–46945, 2020.

[15] N. Rachbureeand and P. Wattana, "An assistive model of obstacle detection based on deep learning: YOLOv3 for visually impaired people," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3434, 2021.

[16] Khan,N. Shah and H.U. Khan, "Analysis of navigation assistants for blind and visually impaired people: A systematic review," *IEEE Access*, vol. 9, no. 9, pp. 26712–26734, 2021.

[17] A. Sundas, S. Badotra, Y. Alotaibi, S. Alghamdi and O. I. Khalaf, "Modified bat algorithm for optimal VM's in cloud computing," *Computers, terials & Continua*, vol. 72, no. 2, pp. 2877–2894, 2022.

[18] S. Sennan, G. Kirubasri, Y. Alotaibi, D. Pandey and S. Alghamdi, "EACR-LEACH: Energy-awareclusterbased routing protocol for WSN based IoT," *Computers, Materials & Continua*, vol. 72, no. 2, pp. 2159– 2174, 2022.

[19] Anuradha, N. Subramani, O. I. Khalaf, Y. Alotaibi, S. Alghamdi *et al.*, "Chaotic search and rescueoptimization based multi-hop data transmission protocol for underwater wireless sensor networks," *Sensors*, vol. 22, no. 9, pp. 2–20, 2022.

[20] X. R. Zhang, W. F. Zhang, W. Sun, X. M. Sun and S. K. Jha, "A robust 3-D medical watermarking based on wavelet transform for data protection," *Computer Systems Science & Engineering*, vol. 41, no. 3, pp. 1043–1056, 2022.

[21] X. R. Zhang, X. Sun, X. M. Sun, W. Sun and S. K. Jha, "Robust reversible audio watermarking scheme for telemedicine and privacy protection," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3035–3050, 2022.

[22] Y. Alotaibi, "A new meta-heuristics data clustering algorithm based on tabu search," *Symmetry*, vol. 14, no. 3, pp. 623, 2022.

[23] N. Subramani, P. Mohan, Y. Alotaibi, S. Alghamdi and O. I. Khalaf, "An efficient metaheuristic-based clustering with routing protocol for underwater wireless sensor networks," *Sensors*, vol. 22, no. 2, pp. 1–16, 2022.

[24] S. Rajendran, O. I. Khalaf, Y. Alotaibi and S. Alghamdi, "Mapreduce-based big data classification model using feature subset selection and hyperparameter tuned deep belief network," *Scientific Reports*, vol. 11, no. 1, pp. 1–10, 2021.

[25] R. Rout, P. Parida, Y. Alotaibi, S. Alghamdi and O. I. Khalaf, "Skin lesion extraction using multiscale morphological local variance reconstruction based watershed transform and fast fuzzy C-means clustering," *Symmetry*, vol. 13, no. 11, pp. 1–19, 2021.

[26] S. Bharany, S. Sharma, S. Badotra, O. I. Khalaf, Y. Alotaibi *et al.*, "Energy-efficient clustering scheme for flying ad-hoc networks using an optimized leach protocol," *Energies*, vol. 14, no. 19, pp. 1–20, 2021.

[27] Y. Alotaibi, "A new database intrusion detection approach based on hybrid meta-heuristics," *Computers, Materials & Continua*, vol. 66, no. 2, pp. 1879–1895, 2021.

[28] S. S. Rawat, S. Alghamdi, G. Kumar, Y. Alotaibi, O. I. Khalaf *et al.*, "Infrared small target detection based on partial sum minimization and total variation," *Mathematics*, vol. 10, no. 4, pp. 671, 2022.

[29] Mohan, N. Subramani, Y. Alotaibi, S. Alghamdi, O. I. Khalaf *et al.*, "Improved metaheuristics-based clustering with multihop routing protocol for underwater wireless sensor networks," *Sensors*, vol. 22, no.4, pp. 1618, 2022.

[30] Alsufyani, Y. Alotaibi, A. O. Almagrabi, S. A. Alghamdi and N. Alsufyani, "Optimized intelligent data management framework for a cyber-physical system for computational applications," *Complex&Intelligent Systems*, vol. 17, no. 4, pp. 1–13, 2021.

[31] Y. Alotaibi, "Automated business process modelling for analyzing sustainable system requirements engineering," in *2020 6th Int. Conf. on Information Management (ICIM), IEEE*, London, United Kingdom, IEEE Xplore, pp. 157–161, 2020.

[32] J. Jayapradha, M. Prakash, Y. Alotaibi, O. I. Khalaf and S. Alghamdi, "Heap bucketization anonymity-An efficient privacy-preserving data publishing model for multiple sensitive attributes," *IEEE Access*, vol. 10, pp. 28773–28791, 2022.

[33] Y. Alotaibi, "A new Secured e-government efficiency model for sustainable services provision," *Journal of Information Security and Cybercrimes Research*, vol. 3, no. 1, pp. 75–96, 2020.

[34] H. H. Khan, M. N. Malik, R. Zafar, F. A. Goni, A. G. Chofreh *et al.*, "Challenges for sustainable smart city development: A conceptual framework," *Sustainable Development*, vol. 28, no. 5, pp. 1507–1518, 2020.

[35] Y.Alotaibi, M. N. Malik, H. H. Khan,A. Batool, S. U. Islam*et al.*, "Suggestion mining from opinionated text of big social media data," *Computers, Materials & Continua*, vol. 68, no. 3, pp. 3323–3338, 2021.

[36] G. Suryanarayana, K. Chandran, O. I. Khalaf, Y. Alotaibi, A. Alsufyani *et al.*, "Accurate magnetic resonance image super-resolution using deep networks and gaussian filtering in the stationary wavelet domain," *IEEE Access*, vol. 9, no. 1, pp. 71406–71417, 2021.

[37] G. Li, F. Liu, A. Sharma, O. I. Khalaf, Y. Alotaibi *et al.*, "Research on the natural language recognition method based on cluster analysis using neural network," *Mathematical Problems in Engineering*, vol. 2021, no. 1, pp. 1–13, 2021.

[38] WHO (2021). Vision impairment and blindness. In *Blindness and vision impairment*. World Health Organization, URL https://www.who.int/news- oom/fact-sheets/ detail/blindness-and-visual-impairment. (Accessed 20 April 2022).

[39] Harvard, U. (2012). Measuring size from images: A wrangle with angles and image scale. In *Smithsonian astrophysical observatory*. Harvard University, URL https://lweb.cfa.harvard.edu/webscope/activities/pdfs/measureSize.pdf. (Accessed 01 October 2020).