iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VI Month of publication: June 2025

DOI: https://doi.org/10.22214/ijraset.2025.72792

# Object Detection and Distance Calculation for Visually Impaired Persons

Paruchuri Venkata Sudheer, Sirigiri Rohit Reddy, Karumajji Kyathisree

***Abstract: The study introduces a advanced assistive technology that enables precise distance calculation and real-time object identification for individuals with vision impairments. With the help of the MobileNetV3 trained model, the system can recognize up to 91 distinct items at once, providing a thorough awareness of the user's surroundings. In addition to identifying particular items, the system evaluates the distances between various things in the field of vision and between the user and those objects. It utilizes sophisticated computer vision algorithms to provide reliable object detection and depth estimates, therefore mitigating the difficulties encountered by those with visual impairments. By using crafted aural interface to provide distance information, the audio feedback function improves the user experience. Through the provision of accurate and timely spatial layout information, our technology seeks to enhance the ability of visually impaired people to navigate and to become independently mobile. In order to guarantee successful integration into the everyday life of the visually handicapped, the research also addresses the difficulties in designing such systems, such as the requirement for real-time processing, environmental adaptation, and user-friendly aural feedback.***
***Keywords: Visual impairment, Distance estimation, MobileNetV3, Spatial awareness, Multi Object detection and Human-computer interaction.***

## I. INTRODUCTION

One of the main tasks in computer vision is detection of objects. Using object detectors that are trained on a sub-image and applied exhaustively at all sizes and locations is a frequent approach to solve this issue. Using this paradigm, a discriminatively trained deformable part model (DPM) was able to attain state of the art performance on detection tasks. Even in this day of rapid technological progress, meeting the needs of those who are visually impaired whose disabilities make it difficult for them to navigate and understand their environment on their own remains a formidable task [1]. By developing a cutting-edge assistive device that makes use of deep learning and computer vision, our research aims to alleviate this problem. The main goal is to provide visually impaired people with enhanced spatial awareness so they can identify, comprehend, and notice items in their surroundings instantly [2]. The main goal of this project is to create a complete system that uses the trained MobileNetV3 model for precise distance estimate and multi-object identification. The system attempts to identify 91 different items at once using advanced computer vision techniques, giving users a comprehensive and rich view of their environment [3]. Our study focuses a distinct emphasis on comprehending the spatial interactions among detected items and between the user, in addition to individual object recognition. This all-encompassing method aims to improve the visually impaired navigating experience while encouraging their independence and self-sufficiency.

This research is important because it goes beyond technology and explores the societal impact that it aims to have. People with visual impairments frequently encounter difficulties in obtaining information that is essential to their daily existence, which can lead to a feeling of dependence. In order to address these issues, this research proposes a method that uses an understandable auditory feedback system to identify objects and provide important distance information [4]. Our technology aims to improve the quality of life for visually impaired people by encouraging independent movement, which will promote equality and inclusion in their contacts with the outside world [5]. The study introduces a cutting-edge assistive technology that enables precise distance calculation and real-time object identification for individuals with vision impairments. This study adds a number of significant insights to the field of assistive technology for people with vision impairments. Above all, our study presents a stable and effective system based on the MobileNetV3 trained model, which enables the simultaneous identification of a large number of objects (91 total).

For visually challenged users, this feature greatly broadens their awareness of their world and gives them a more thorough comprehension of their surroundings [6]. The system's capacity to precisely estimate distances in addition to identifying specific objects makes a significant contribution. Our work assures that visually impaired people may determine the spatial connections not just between themselves and the items they perceive, but also between several objects inside their range of view, by utilizing sophisticated computer vision algorithms.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

This dual feature provides a comprehensive navigation and environmental comprehension solution, significantly improving the system's usability and usefulness. Additionally, our study highlights the necessity of integrating an audible feedback mechanism within the system. This input is essential to converting the visual data that the system collects into a format that users with visual impairments can access and use. A more natural and engaging user experience is made possible by the auditory feedback, which not only offers information about objects that have been spotted but also gives real-time distance. Additionally, our study highlights the necessity of integrating an audible feedback mechanism within the system. This input is essential to converting the visual data that the system collects into a format that users with visual impairments can access and use. A more natural and engaging user experience is made possible by the auditory feedback, which not only offers information about objects that have been spotted but also gives real-time distance cues.

The goal of this project is to develop a cutting-edge assistive technology system specifically tailored to meet the needs of people who are visually impaired. This system is primarily concerned with implementing the trained MobileNetV3 model for reliable multi-object identification [7]. The system can recognize up to 91 different items at once, giving the user a thorough awareness of their surroundings. Moreover, the content includes complex distance estimation algorithms that allow the system to identify items individually as well as evaluate the spatial connections between objects and the user, as well as between several objects in the frame [8]. By converting visual information into easily understood cues, the user-friendly aural feedback system adds even more value to the material and gives visually impaired people improved navigation and spatial awareness.

## II. LITERATURE SURVEY

Considerable progress has been achieved in the area of assistive technology for people with visual impairments in the past, with particular attention on several facets of object detection, distance estimates, and auditory feedback systems. Cazzato [9] proposed a survey of computer vision methods for dimensional object detection from unmanned aerial vehicles for the visually aided people. Diwan [10] and team built an object detection technique using the YOLO model. The majority of early work focused on basic object detection using traditional computer vision methods. For example, several systems relied on color and form cues to identify objects at a basic level; nevertheless, the accuracy and scalability of these techniques were frequently compromised. Zou [11] proposed object detection using computer vision techniques that detect the limit number of objects with low accuracy of the model. Gehrig [12] introduced recurrent vision transformers for object detection with event cameras. With the development of deep learning, the area saw a sea change as researchers began using convolutional neural networks (CNNs) for more advanced object recognition. deep learning models, such as YOLO (You Only Look Once) and faster R-CNN, were first used in groundbreaking efforts for real-time object recognition. These advancements greatly increased precision, making it possible to recognize objects in a variety of situations with more dependability.

Yang [13] and team proposed a real time object detector based MobileNetV3 for UAV applications. Zuhair [14] proposed a comparison of tensor flow and tensor flow lite for object detection on RaspberryPi. Brownlee [15] and team used deep learning for computer vision for image classification, object detection, and face recognition in python. Wang [16] built object detection by combining both computer vision as well as deep learning techniques. Zhang [17] and team build object detection with efficient integral aggregation. Using advancements in object identification as a foundation, scientists worked to improve distance estimate techniques. Some of the first methods used stereo vision systems to measure distances by using depth information. Nevertheless, the expense, complexity, and real-time processing of these systems were frequently problematic. More precise distance measurements have been made possible by recent developments, such as the incorporation of depth-sensing technologies like LiDAR and structured light. In parallel, attempts were made to improve aural feedback devices that provide information to those who are blind or visually impaired.

Wang [18] built an omni-supervised object detection system with transformers. Su [19] and team proposed a unified transformer framework for group segmentation as well as co-segmentation, co-saliency detection and video salient object detection. Nagarajan [20] proposed a hybrid optimization enabling deep learning for indoor object detection and distance estimation to assist visually impaired persons. Kim [21] built a camera radar 3d object detection with spatial contextual fusion transformer. Mao [22] and team introduced 3D object detection for autonomous driving vehicles using computer vision and convolutional networks. Tasyurek [23] proposed a new approach for spatial street sign detection using deep learning object detection, distance estimation, rotation and projection system. Yang [24] built a symmetry driven unsupervised abnormal object detection for railway inspection. Prior research investigated the application of voice synthesis and spatialized sound to convey distance cues and object information. Despite their innovation, these systems frequently had trouble integrating exact spatial information with various object cues.

In addition, scholars realized how important it is for design to be inclusive, which prompted them to look at user interfaces that are especially made to accommodate those who are visually impaired. Research has investigated the use of wearable technology, haptic feedback, and customized interfaces to enhance user engagement and usability in general. Even with these developments, there are still difficulties in arriving at a thorough and timely answer. Current methods are not always able to address numerous objects in a frame at the same time or provide reliable distance information in a variety of settings. Beyond that, user-centric assessments and feedback systems are crucial elements that need more research to guarantee the effective incorporation of assistive technology into the everyday lives of people who are visually impaired. In order to improve accessibility and usability, the current study includes a system that not only combines powerful object identification and distance estimates, but also an aural feedback interface that is easy to use.

## III. DATA COLLECTION & PREPROCESSING

The procedure of gathering data for the purpose of training the MobileNetV3 model entails compiling a large and varied dataset [25]. This collection includes pictures of a variety of outdoor and interior environments, guaranteeing a broad range of things that people with vision impairments can come across in their daily lives. All of the images are painstakingly labeled with instances of the designated items and the bounding boxes that correspond to them. Common components including people, animals, cars, homes, and electronics are examples of objects of interest. The items chosen for the dataset are essential to the operation of assistive technology and encompass a wide variety of objects found in the surroundings of a visually impaired person. Based on its significance and possible influence on the user's safety and navigation, each object is given a particular weight. For example, because they are so important in metropolitan settings, things like "person," "car," and "traffic light" could have greater weights. This weighting scheme is crucial for maximizing the learning process of the model by giving more weight to things that are more significant in the context of visual navigation.

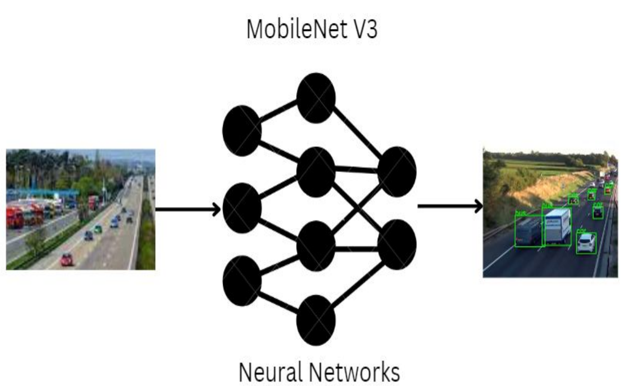



Fig.1 Data preprocessing for detection of objects

Given that people are of utmost significance in the environment of a visually impaired person, person-centric things receive extra consideration in this dataset (as shown in Fig.1). This category contains items like "hats," "backpacks," and "umbrellas" that are inextricably linked to human presence and interaction. By adding these items, the model is able to offer more detailed information about the surrounding area, which improves the user's safety and spatial awareness. The selection of objects for the collection is based on both their individual value and their contextual relevance across a range of scenarios. To improve the system's flexibility to various situations, items such as "bed," "mirror," and "dining table" are incorporated to solve scenarios that arise within interior spaces. This variety of contextual factors guarantees that the model can identify and comprehend items in a wide range of real-world situations. Iterative data gathering involves ongoing refinement depending on user input and model performance. The dataset is modified while the model is tested and trained in order to increase the accuracy and dependability of the system. The iterative process of data collection guarantees that the MobileNetV3 model is refined to identify and comprehend the designated objects with exceptional accuracy, which is consistent with the main objective of developing an assistive technology system that blends in seamlessly with the daily lives of people who are visually impaired.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

## IV. WORKING METHODOLOGY

The suggested assistive technology system's operation is based on an intricate fusion of deep learning and computer vision methods (as shown in Fig.2). The MobileNetV3 model is fundamentally the basis for both distance estimates and object detection. Using an input device, like a camera, the system first takes pictures of the user's surroundings in real time. The pre-trained MobileNetV3 model is then applied to these photos, allowing the system to concurrently recognize and classify several items from a preset list. Each detected object's bounding box coordinates are included in the model's output, which enables the system to determine an object's presence as well as its precise placement inside the frame. With the use of sophisticated computer vision algorithms, the distance between the user and the identified items is calculated, making it easier to accurately estimate the physical space. In order to properly communicate this information to those who are visually impaired, the system includes an aural feedback mechanism that translates visual information into audible cues.

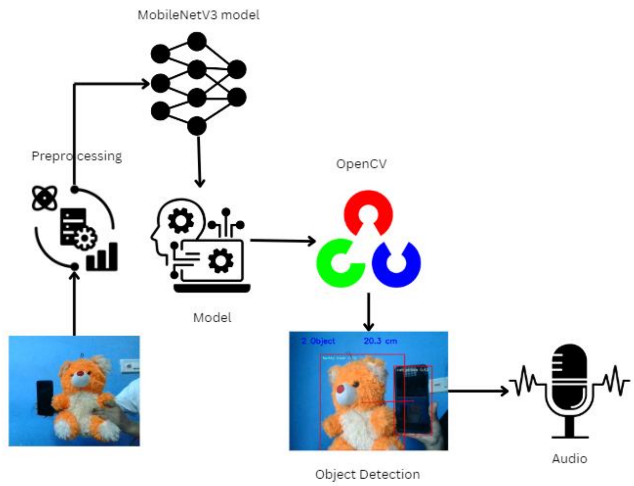



Fig.2 Working Architecture for object detection

Users may get real-time information on the kinds of items in their surroundings and the distances between them and themselves through the auditory feedback. Because of the system's easy design, users may easily understand and act upon the information provided by the audio feedback. The overall objective of improving spatial awareness and navigational abilities for those with visual impairments is in line with this user-centric approach. The whole working technique is based on a continuous loop in which real-time input is captured, processed, and provided by the system. The system's dependability and usefulness in a range of everyday settings are ensured by this iterative process, which enables adaptation to changing surroundings and different scenarios. All things considered, the assistive technology system operates on a well-coordinated framework that uses state-of-the-art technologies to enable people who are visually impaired to comprehend and navigate their environment.

### *A. MobileNet V3 pretrained model*

The convolutional neural network architecture known as MobileNetV3 model is the best option for the suggested assistive technology system as it has become well-known for being an effective tool for real-time image processing jobs. The MobileNetV3 pretrained model was chosen based on its unique design features, effectiveness, and performance attributes that meet the application's needs. The third version of the MobileNet series, known as MobileNetV3, was released to address the difficulties associated with implementing deep neural networks on mobile and edge devices that have constrained processing power. To balance model size, speed, and accuracy, Google developed MobileNetV3, which makes use of architectural design techniques including inverted residuals and linear bottlenecks (as shown in Fig.3). Because of this, it's especially appropriate for real-time applications and situations where computing performance is crucial.

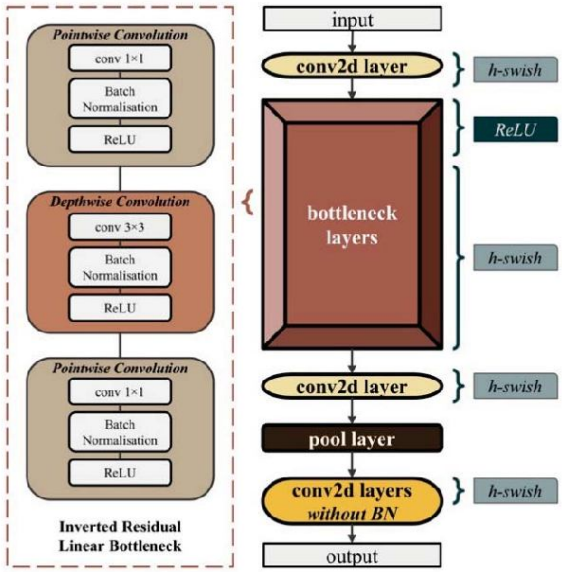



Fig.3 Mobile-Net V3 Architecture

The focus MobileNetV3 places on model efficiency without sacrificing accuracy is one of its main differentiators. The architecture is more computationally efficient since it uses a lightweight design that drastically lowers the number of parameters. For real-time applications, this efficiency is key because it guarantees that the model can process pictures quickly. This is a critical need for the proposed assistive technology since it provides users with visual impairments with timely information. Specifically designed to be deployed on mobile and edge devices, MobileNetV3 complies with the useful requirements of the suggested assistive technology. Because of its optimized design, which works well on systems with limited resources, it is the best option for applications that need to strike a compromise between computational efficiency and model complexity. Because of MobileNetV3's demonstrated effectiveness and accuracy in real-time image recognition tasks, it was chosen above alternative models. While other models, such Faster R-CNN or YOLO (You Only Look Once), perform well in object identification, MobileNetV3 is designed expressly for situations when speed and resource efficiency are critical. This is in line with the goals of the assistive technology system, which has to analyze visual input quickly and accurately in order to give people with visual impairments current information.

### *B. Distance Measurement*

One essential element that enables the system to determine the distance between the camera and identified objects is the distance measurement in the code that has been given. The focal length finder function and a subsequent distance estimate function are employed in this procedure. The purpose of the focal length finder function is to determine the focal length, which is the separation between the CMOS sensor and the camera lens. Three arguments are required for this function. The focal length is calculated using the formula,

$$Focal\ Length = \frac{Real\ Width}{Width\ in\ Reference\ Image\ X\ MeasuredDistance}$$

where it uses a reference item to determine the optical characteristics of the camera. The distance estimation formula is,

$$Distance\ = \frac{Face\ Width\ in\ Frame}{Real\ Face\ Width\ X\ Focal\ Length}$$

is used to calculate an object's distance from the camera based on how big it appears in the picture. The device then shows the determined distance on the screen, showing the distance in inches. The code also has the ability to generate lines denoting distance levels and rectangles surrounding items that are recognized. The distance between the camera and objects it has identified is displayed in real time thanks to the visual representation of this data on the screen. The algorithm also includes object tracking capability, which computes the Euclidean distance between objects by tracking their centroids. The screen displays this dynamic distance level, which provides more details on the spatial relationships between various items in the frame. Object tracking is used by the code to track the motion of identified objects across a series of frames.

The system records these items' locations dynamically by computing their centroids. In order to display this dynamic tracking, circles are drawn at the centroids of the objects that have been recognized. Next, in order to provide real-time information on the relative distances and motions of the monitored objects, the Euclidean distance between their centroids is computed. The Euclidean distance between monitored objects is represented visually by the dynamic distance level.

## V. RESULTS

This research produces the following outcomes: distance calculations, object identification in real time, and text-to-speech conversion for an audio response. The algorithm uses the video feed to identify items in real-time using a pre-trained neural network. Bounding boxes, class labels, and confidence ratings are appended to detected items to provide a visual depiction of the objects in the frame. Using a focal length that has already been determined, the algorithm determines the distance between the camera and objects that have been spotted. For every item that is spotted, distance information is calculated to provide spatial awareness of the object's proximity to the camera. The pyttsx3 library is used in the code to convert text to speech. When a new item is detected, the system creates a text string dynamically that conveys the object's kind and estimated distance. After that, the text string is transformed into audible speech, giving rise to a prompt and distinct audio response. The purpose of the audio feedback is to improve user interaction especially for those who are visually impaired. In order to help users navigate and comprehend their surroundings, verbal information about the identified items and their distances is provided.

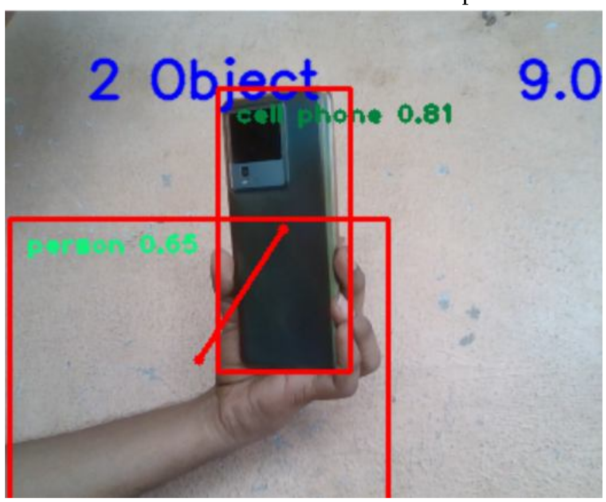



Fig.4 Object Detection and Distance Calculation

The code runs in an infinite loop, gathering frames, identifying items, and giving feedback in real time. As objects are identified or as their distances alter within the video stream, users are continuously provided with aural cues. Object tracking is a feature of the code that lets it track the motion of objects it detects across a series of frames. The display of dynamic distance levels on the screen gives users access to real-time data on the spatial relationships between the tracked objects (as shown in Fig.4). Upon detecting or tracking a new item, the system instantly produces and emits auditory replies. Based on the identified items and their distances, users may make well-informed judgments thanks to the spoken answers, which provide vital environmental information. The use of audio input guarantees successful interaction between the system and visually impaired persons. Real-time voice answers prioritize accessibility and inclusion, a design approach that is shown by the code.

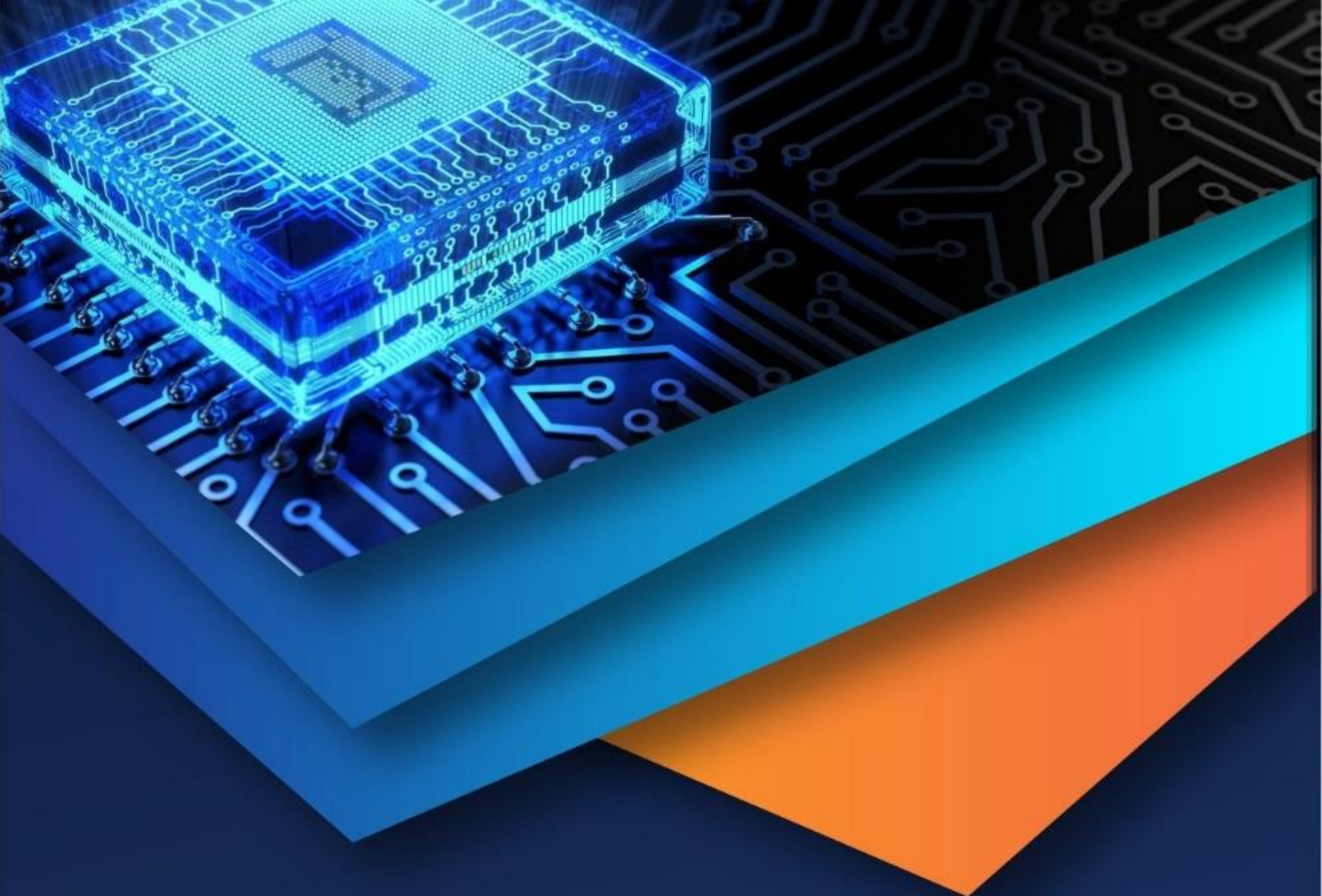
## VI. CONCLUSION

To sum up, this initiative is a noteworthy development in assistive technology that is designed to improve the navigational skills and spatial awareness of those who are visually impaired. The system gives users extensive information about their surroundings by integrating a dynamic distance level display, a real-time object identification system, and distance calculation using a computed focal length. An essential layer of accessibility is added by the text-to-speech technology, which allows the system to provide audio feedback along with dynamic communication of objects that are recognized and their distances. The real-time voice answers, object tracking, and constant monitoring all support a user-centric design that makes sure visually impaired people get timely and pertinent information. The project's flexibility, demonstrated by the utilization of the pyttsx3 library and the MobileNetV3 model, shows a dedication to tackling issues that the visually impaired confront in the actual world.

All things considered, this project not only demonstrates the technological capabilities of deep learning and computer vision, but also highlights how these fields have the potential to improve the lives of people who are visually impaired by giving them more inclusive and accessible ways to perceive and interact with their surroundings.

## REFERENCES

[1] Brady, Erin, et al. "Visual challenges in the everyday lives of blind people." Proceedings of the SIGCHI conference on human factors in computing systems. 2013.
[2] Bousbia-Salah, Mounir, Maamar Bettayeb, and Allal Larbi. "A navigation aid for blind people." Journal of Intelligent & Robotic Systems 64 (2011): 387-400.
[3] Zou, Zhengxia, et al. "Object detection in 20 years: A survey." Proceedings of the IEEE (2023).
[4] Wu, Jia-jie, et al. "Focal length measurement based on Hartmann–Shack principle." Optik 123.6 (2012): 485-488.
[5] Neal, Daniel R., et al. "Measurement of lens focal length using multicurvature analysis of Shack-Hartmann wavefront data." Current Developments in Lens Design and Optical Engineering V. Vol. 5523. SPIE, 2004.
[6] Koteswararao, M., and P. R. Karthikeyan. "Accurate and real-time object detection system using YOLO v3-320 in comparison with MobileNet SSD network." AIP Conference Proceedings. Vol. 2822. No. 1. AIP Publishing, 2023.
[7] Iyer, Rakkshab, et al. "Comparison of YOLOv3, YOLOv5s and MobileNet-SSD V2 for real-time mask detection." Artic. Int. J. Res. Eng. Technol 8 (2021): 1156-1160.
[8] Li, Wei, and Kai Liu. "Confidence-aware object detection based on MobileNetv2 for autonomous driving." Sensors 21.7 (2021): 2380.
[9] Cazzato, Dario, et al. "A survey of computer vision methods for 2d object detection from unmanned aerial vehicles." Journal of Imaging 6.8 (2020): 78.
[10] Diwan, Tausif, G. Anirudh, and Jitendra V. Tembhurne. "Object detection using YOLO: Challenges, architectural successors, datasets and applications." multimedia Tools and Applications 82.6 (2023): 9243-9275.
[11] Zou, Zhengxia, et al. "Object detection in 20 years: A survey." Proceedings of the IEEE (2023).
[12] Gehrig, Mathias, and Davide Scaramuzza. "Recurrent vision transformers for object detection with event cameras." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
[13] Yang, Yonghao, and Jin Han. "Real-Time object detector based MobileNetV3 for UAV applications." Multimedia Tools and Applications 82.12 (2023): 18709-18725.
[14] Zuhair, Muhammad Syihabuddin Az, Andi Widiyanto, and Setiya Nugroho. "Comparison of tensorflow and tensorflow lite for object detection on Raspberry Pi 4." AIP Conference Proceedings. Vol. 2706. No. 1. AIP Publishing, 2023.
[15] Brownlee, Jason. Deep learning for computer vision: image classification, object detection, and face recognition in python. Machine Learning Mastery, 2019.
[16] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
[17] Zhang, Shan, et al. "Kernelized few-shot object detection with efficient integral aggregation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[18] Wang, Pei, et al. "Omni-detr: Omni-supervised object detection with transformers." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
[19] Su, Yukun, et al. "A unified transformer framework for group-based segmentation: Co-segmentation, co-saliency detection and video salient object detection." IEEE Transactions on Multimedia (2023).
[20] Nagarajan, Anandh, and M. P. Gopinath. "Hybrid Optimization-Enabled Deep Learning for Indoor Object Detection and Distance Estimation to Assist Visually Impaired Persons." Advances in Engineering Software 176 (2023): 103362.
[21] Kim, Youngseok, et al. "Craft: Camera-radar 3d object detection with spatio-contextual fusion transformer." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37. No. 1. 2023.
[22] Mao, Jiageng, et al. "3D object detection for autonomous driving: A comprehensive survey." International Journal of Computer Vision (2023): 1-55.
[23] Taşyürek, Murat. "Odrp: a new approach for spatial street sign detection from exif using deep learning-based object detection, distance estimation, rotation and projection system." The Visual Computer (2023): 1-21.
[24] Yang, Taocun, et al. "Symmetry-driven unsupervised abnormal object detection for railway inspection." IEEE Transactions on Industrial Informatics (2023).
[25] Zhao, Liquan, and Leilei Wang. "A new lightweight network based on MobileNetV3." KSII Transactions on Internet & Information Systems 16.1 (2022)




10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

TOGETHER WE REACH THE GOAL


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY