

Advanced Object Detection System for Visually Impaired Individuals Using Deep Learning for Enhanced Mobility Assistance

¹Mr. K. Ramesh Babu,²Utti Vasanthi,³Salagala Anjula,⁴Thulluri Mohan Sandeep

¹Assistant Professor, Dept of Computer Science and Engineering, St. Ann's College of Engineering and Technology, Chirala-523187, India.

^{2,3,4}B. Tech Student, Dept of Computer Science and Engineering, St. Ann's College of Engineering and Technology, Chirala-523187, India.

ABSTRACT

Visually impaired individuals often face difficulties in identifying surrounding objects, which affects their safety and independent mobility. This project presents an advanced object detection system using deep learning to assist visually impaired users in real time. The system utilizes the YOLO algorithm to detect multiple objects accurately from live camera feeds or uploaded videos. A Flask-based web application is developed to handle video input and system control efficiently. Computer vision techniques are applied to process video frames and highlight detected objects. The names of detected objects are converted into speech using text-to-speech technology. This audio feedback helps users understand their surroundings without visual dependency. The system is designed to be user-friendly, cost-effective, and easy to operate. It enhances environmental awareness and reduces the risk of accidents. Overall, the project

demonstrates how artificial intelligence can improve assistive technologies for visually impaired individuals.

KEYWORDS: *Object Detection, YOLO, Deep Learning, Assistive Technology, Computer Vision*

INTRODUCTION

Visual impairment significantly affects an individual's ability to navigate and understand their surroundings safely. Traditional mobility aids such as white canes and guide dogs provide limited environmental awareness and cannot identify dynamic objects. Recent advancements in computer vision and deep learning have enabled machines to perceive and interpret visual information with high accuracy. Object detection algorithms, especially YOLO-based models, can identify multiple objects in real time from video streams. Integrating these models with assistive technologies offers new possibilities for enhancing independent mobility for visually impaired users. Audio feedback systems can convert visual information into meaningful voice alerts. Such systems help users recognize obstacles, objects, and surroundings

effectively. The proposed system leverages deep learning to perform real-time object detection through a camera input. Detected objects are processed and converted into spoken descriptions. This approach aims to improve safety, confidence, and independence for visually impaired individuals.

LITERATURE SURVEY

Object detection and assistive technologies for visually impaired individuals have gained significant attention due to advances in computer vision and deep learning. Early assistive systems relied on traditional image processing techniques, which were limited in accuracy and real-time performance. Redmon et al. (2016) introduced the YOLO (You Only Look Once) framework, enabling fast and real-time object detection. Bochkovskiy et al. (2020) further improved YOLO with YOLOv4, enhancing detection accuracy and speed. Several researchers have explored the use of deep learning-based object detection to assist visually impaired users in understanding their surroundings. Tapu et al. (2014) proposed a wearable vision system to detect obstacles and provide audio feedback. Yi et al. (2018) demonstrated that combining object detection with voice assistance improves user mobility and safety. Recent studies using YOLOv5 and YOLOv8 (2022–2023) show high accuracy in real-time environments. Integration of text-to-speech

systems has been proven effective for conveying object information audibly. Overall, literature indicates that deep learning-based object detection systems significantly enhance mobility assistance for visually impaired individuals.

RELATED WORK

Several researchers have explored assistive technologies to improve mobility and safety for visually impaired individuals using computer vision and deep learning techniques. Early systems relied on traditional image processing methods, which had limited accuracy in complex real-world environments. With the emergence of deep learning, Convolutional Neural Networks (CNNs) significantly improved object recognition performance. Redmon et al. introduced the YOLO (You Only Look Once) framework, enabling real-time object detection with high speed and accuracy. Later versions of YOLO further enhanced detection efficiency and robustness. Studies have integrated object detection with audio feedback systems to convey environmental information to visually impaired users. Researchers have also utilized OpenCV with deep learning models to process live video streams effectively. Some works focused on wearable camera-based systems to detect obstacles and objects in real time. Recent advancements emphasize lightweight

models for real-time deployment on edge devices. Overall, deep learning-based object detection systems have proven effective in providing reliable visual-to-audio assistance for enhanced mobility.

EXISTING METHODS

In the existing system, visually impaired individuals mainly rely on traditional mobility aids such as white canes and guide dogs to navigate their surroundings. These methods help in detecting nearby obstacles but provide very limited information about the type of object present in the environment. Some basic electronic aids and ultrasonic sensors are also used, which can only indicate the presence of an obstacle without identifying it. These systems lack intelligence, contextual awareness, and real-time object recognition capabilities. Moreover, they do not provide detailed audio descriptions, making it difficult for users to understand complex surroundings. As a result, existing solutions offer limited assistance in dynamic and crowded environments, highlighting the need for an advanced deep learning-based object detection system with audio feedback for improved mobility and safety.

PROPOSED METHOD

The proposed system is an advanced object detection and voice assistance solution designed to support visually impaired individuals in safe and independent

mobility. It uses a deep learning-based object detection model (YOLO) to identify objects from live camera input or uploaded videos in real time. The system processes video frames, detects objects accurately, and generates corresponding object labels, which are then converted into audio feedback using text-to-speech technology. This allows users to understand their surroundings through voice guidance without visual interaction. The system is easy to use, requires minimal user input, and operates efficiently in real-world environments. By providing real-time audio alerts, the proposed system enhances situational awareness and significantly improves mobility assistance for visually impaired users.

ARCHITECTURE

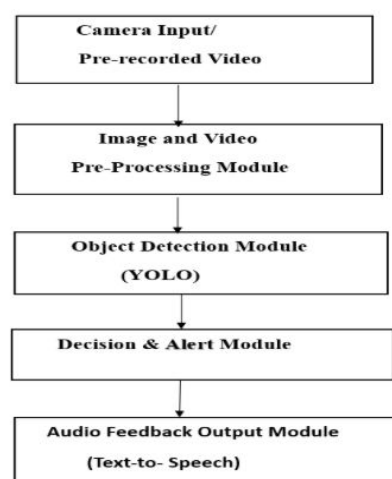


Fig 1: System Architecture

METHODOLOGY DESCRIPTION

Camera Input / Pre-recorded

Video:

This module captures real-time video through a camera or accepts pre-recorded video as input. It provides visual data of the surroundings for further processing and object detection.

Image and Video Pre-Processing Module:

This module enhances video frames by resizing, noise removal, and normalization. Pre-processing improves detection accuracy and ensures compatibility with the YOLO model.

Object Detection Module (YOLO):

This module uses the YOLO deep learning model to identify and classify objects in each frame. It outputs object names along with bounding boxes in real time.

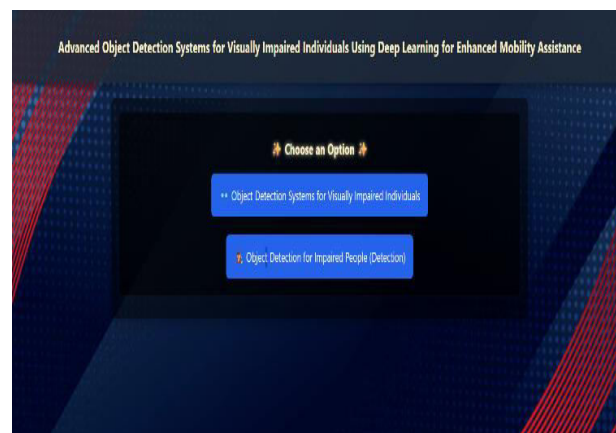
Decision & Alert Module:

This module analyzes detected objects and decides which information should be conveyed to the user. It filters relevant objects and prepares meaningful alert messages.

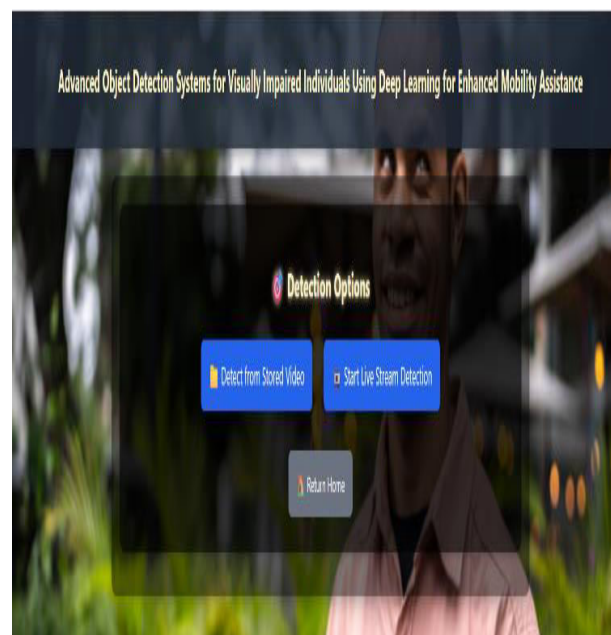
Audio Feedback Output Module (Text-to-Speech):

This module converts detected object names into spoken audio using a text-to-speech engine. It provides real-time voice feedback to

assist visually impaired users in understanding their surroundings.

RESULTS AND DISCUSSION**Fig2: Front page of the Output**

The image shows the home interface of the Advanced Object Detection System, providing users with two clear options to start object detection.

**Fig3: Interface after clicking object Detection for visually impaired people**

This page allows users to provide input either by uploading a stored video or by starting live camera streaming for object detection.

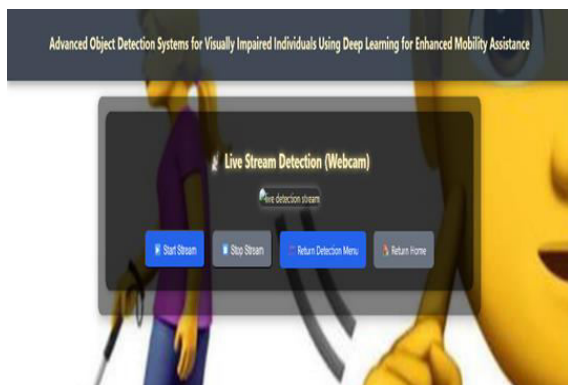


Fig 4: clicking on start live stream

This slide shows the live webcam-based object detection feature of the system. It captures real-time video and uses deep learning to detect objects and obstacles, helping visually impaired users move safely. Simple controls allow users to start or stop the stream and navigate easily, emphasizing real-time assistance and accessibility.

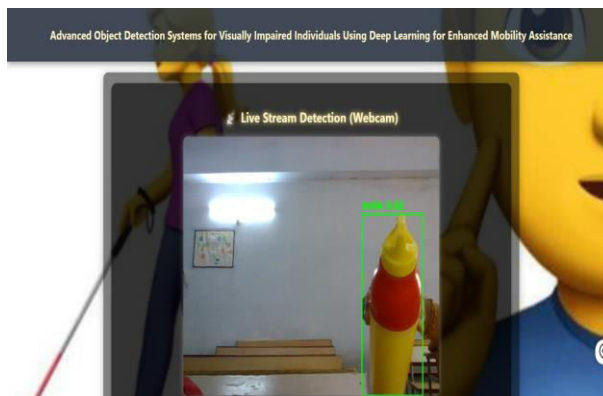


Fig 5: ongoing live stream

When the user clicks Start Detection, the system activates the webcam and begins real-time object detection. Detected objects are highlighted with bounding boxes and label names, and the system provides audio feedback announcing object name to assist visually impaired users.

CONCLUSION

This project successfully implements an advanced object detection system to assist visually impaired individuals using deep learning techniques. By integrating the YOLO object detection model with a Flask-based web application, the system accurately identifies objects from live camera input or uploaded videos. The detected objects are converted into real-time audio feedback using text-to-speech technology, enabling users to understand their surroundings without visual dependency. The results demonstrate that the proposed system improves environmental awareness, safety, and independence, proving the effectiveness of artificial intelligence in assistive applications.

FUTURE SCOPE

The proposed system can be further enhanced by integrating it into mobile applications and wearable devices such as smart glasses. GPS and navigation support can be incorporated to provide real-time route guidance and obstacle alerts. Multilingual voice output can be implemented to support users from different regions. Additionally, distance estimation and object prioritization can be added to deliver more accurate and meaningful audio feedback, thereby improving usability and real-world performance.

REFERENCES

- [1] Nandan, M. J., Sen, M. K., Harini, P., Sekhar, B. M., & Balaji, T. (2013, December). Impact of urban growth and urbanization on the environmental degradation of Lakes in Hyderabad City, India. In *AGU Fall Meeting Abstracts* (Vol. 2013, pp. B31E-0452).
- [2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149, June 2017.
- [3] H. Najm et al., "Assisting Blind People Using Object Detection with Vocal Feedback," *Int. J. Adv. Comput. Sci. Appl.*, 2021.
- [4] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable Vision-Based Assistance System for the Blind," *IEEE Access*, vol. 6, pp. 31102-31118, 2018.
- [5] S. Jain, A. Gupta, and P. Kumar, "Real-Time Object Detection for Visually Impaired People," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, 2020.
- [6] A. Kumar, R. Singh, and S. Kumar, "Vision-Based Navigation Assistance Using Deep Learning," *J. Ambient Intell. Hum. Comput.*, 2022.
- [7] M. Singh, A. Sharma, and V. Laxmi, "Intelligent Assistive System for the Visually Impaired," in *Proc. IEEE 5th Int. Conf. Comput. Commun. Autom. (ICCCA)*, 2020.
- [8] G. Jocher et al., "Ultralytics YOLOv8," 2023. [Online].
- [9] M. Grinberg, *Flask Web Development: Developing Web Applications with Python*. Sebastopol, CA: O'Reilly Media, 2018.
- [10] A. Rosebrock, *Deep Learning for Computer Vision with Python*. PyImageSearch, 2017.
- [11] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*. Sebastopol, CA: O'Reilly Media, 2008.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770-778.
- [13] T. Lin et al., "Microsoft COCO: Common Objects in Context," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 740-755.
- [14] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.
- [15] W. Liu et al., "SSD: Single Shot MultiBox Detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 21-37.
- [16] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257-276, March 2023.
- [17] P. J. Garrido-Cerezo et al., "Deep Learning for Visually Impaired People Assistance: A Survey," *Sensors*, vol. 22, no. 19, p. 7411, 2022.

- [18] S. Gupta, K. Agarwal, and S. Dev, "Audio-Based Navigation for Blind People Using Object Detection," *Wireless Pers. Commun.*, 2023.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Commun. ACM*, vol. 60, no. 6, pp. 84-90, June 2017.
- [20] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303-338, 2010.