

Object Detection for Blind People Using AI

P.Sreesudha, CV.Kiranmaiee, P.Santhoshini, Nadia Shareen, Megha Chandana, Harini Vadla

Electronics and Telematics Engineering G. Narayanamma Institute of Technology and Science Hyderabad, India

Abstract— Real-time object detection and environmental awareness are essential components in assistive technologies for visually impaired individuals. Traditional mobility aids provide limited information about surrounding objects and their proximity, making independent navigation difficult in complex environments. In this work, an AI-based assistive vision system is proposed that integrates the YOLOv8 deep learning model for real-time object detection, distance estimation techniques for proximity awareness, and text-to-speech output for auditory feedback. The system captures input from a camera, detects and classifies multiple objects in the environment, estimates their distance from the user, and converts the detected object labels along with distance information into speech output. This enables visually impaired users to understand nearby obstacles and objects more effectively while moving in indoor and outdoor environments. The proposed approach offers a practical, low-cost, and efficient assistive solution by combining computer vision and artificial intelligence to enhance user safety, independence, and confidence. Experimental observations indicate that the system performs effectively for common object categories and provides meaningful audio guidance in real time.

Keywords— Real-time object detection, assistive technology, visually impaired navigation, YOLOv8, computer vision, distance estimation, environmental awareness, text-to-speech (TTS), auditory feedback, obstacle detection, AI-based vision system, wearable assistive device, indoor and outdoor navigation, deep learning, user safety, independence enhancement, low-cost solution.

I. INTRODUCTION

Object detection is one of the most important tasks in computer vision and has widespread applications in surveillance, autonomous systems, healthcare, and assistive technologies. It involves identifying and localizing objects within an image or video stream, allowing machines to interpret their surroundings and respond intelligently. In recent years, deep learning-based object detection models such as R-CNN, SSD, and YOLO have significantly improved the speed and accuracy of detection systems, making real-time deployment feasible.

For visually impaired individuals, understanding the surrounding environment remains a major challenge. Traditional mobility aids such as white canes and guide dogs help in basic navigation, but they do not provide detailed information about nearby objects, their type, or how far they are from the user. As a result, visually impaired users may face difficulties in avoiding obstacles, recognizing important objects, and moving independently in unfamiliar or crowded environments.

To address this limitation, the proposed system combines object detection, distance estimation, and speech output into a single assistive framework. The system uses the YOLOv8 model to detect multiple objects from real-time camera input. In addition to identifying the object class, the system estimates the distance between the user and the detected object. The detected object names and corresponding distance information are then

converted into speech output, enabling the user to hear what objects are present and how close they are.

The primary contributions of this work are as follows:

1. Development of a real-time assistive vision system using YOLOv8 for multi-object detection.
2. Integration of distance estimation to provide spatial awareness of detected objects.
3. Implementation of speech-based feedback to improve accessibility for visually impaired users.
4. Design of a practical and low-cost framework suitable for future wearable and mobile assistive devices.

II. METHODOLOGY

This section presents the methodologies and concepts adopted from existing research works that contribute to the design of the proposed system, “AI-Based Object Detection and Distance Estimation for Visually Impaired Using YOLOv8.” The selected studies provide the foundation for real-time object detection, distance-aware perception, and speech-based assistive interaction.

A. Real-Time Object Detection Using YOLO-Based Models
YOLO (You Only Look Once) is one of the most widely used one-stage object detection frameworks for real-time applications. Unlike two-stage detectors such as Faster R-CNN, YOLO predicts object locations and class probabilities in a single forward pass through the network, resulting in

significantly faster inference speed. This makes it highly suitable for assistive systems where low latency is critical.

Among the YOLO family, YOLOv8 provides improved accuracy, better feature extraction, and enhanced performance in detecting objects of different scales. The model supports real-time inference while maintaining strong detection capability across a wide range of common object categories. Due to these advantages, YOLOv8 was selected as the core detection model for the proposed system

B. Distance Estimation for Spatial Awareness

Carion Object recognition alone is not sufficient for visually impaired users, as understanding the proximity of surrounding objects is equally important for safe movement. Therefore, distance estimation is incorporated into the proposed system to provide spatial awareness. The system estimates the distance of detected objects from the user based on visual cues derived from camera input, such as object size and relative positioning in the frame.

Although monocular distance estimation offers a lightweight and low-cost solution, its accuracy may vary depending on scene complexity and object scale. However, it remains practical for assistive applications where approximate distance awareness can still provide meaningful support. The methodology also allows future integration with advanced sensors such as stereo cameras, depth cameras, or LiDAR for more accurate depth perception.

C. Speech-Based Assistive Feedback

Speech output plays a critical role in assistive systems for visually impaired individuals because it transforms visual information into an accessible auditory format. After detecting objects and estimating their distance, the system generates descriptive text such as “Chair detected at 1.5 meters” or “Person detected at 3 meters.” This text is then converted into speech using a text-to-speech (TTS) module.

The use of speech feedback allows users to quickly understand their surroundings without relying on visual interfaces. This improves environmental awareness, obstacle recognition, and independent navigation. The integration of real-time object detection with audio feedback forms the core of the proposed assistive framework.

III. SYSTEM DESIGN

The proposed system is designed to detect multiple objects in real time, estimate their distance from the user, and provide corresponding speech output. The methodology is divided into

four major stages: camera input acquisition, object detection using YOLOv8, distance estimation, and audio feedback generation.

A. Overview of the System

The system continuously captures frames from a live camera. Each frame is processed by the YOLOv8 model to identify and classify the objects present in the environment. Once the objects are detected, the system estimates the distance of each object relative to the user. The detected object labels and their estimated distances are then converted into audio output so that visually impaired users can hear what objects are nearby and how far they are located.

The implementation is carried out in Python using computer vision and speech synthesis libraries. The overall framework is designed to be simple, efficient, and suitable for future deployment on portable assistive devices.

B. YOLOv8 for Object Detection

YOLOv8 is employed as the primary object detection model due to its improved speed, accuracy, and efficiency in real-time applications.

Model Loading: The YOLOv8 pre-trained model is loaded into the Python environment.

Frame Preprocessing: Each captured frame is resized and normalized according to model requirements.

Inference: The frame is passed through the YOLOv8 network to generate bounding boxes, confidence scores, and object class predictions.

Filtering: Low-confidence detections are removed using confidence threshold.

Object Labeling: The final detected objects are labeled with class names and bounding boxes for visualization.

The output of this stage is a list of detected objects along with their positions and confidence scores.

C. Audio Output Generation

To improve accessibility, the system converts the detected object names and their estimated distances into speech output. A descriptive sentence is generated for each detection or group of detections, such as:

“Person detected at 2 meters.”

“Chair detected at 1 meter.”

“Car detected at 5 meters.”

The generated text is passed to a text-to-speech module, which produces an audio output that can be played through speakers or earphones. This auditory feedback enables visually impaired users to interpret their surroundings effectively.

D. Implementation Details

The proposed system was implemented in Python using the following setup:

- Programming Language: Python
- Libraries: Open CV, NumPy, Ultralytics YOLOv8, and Text-to-Speech library
- Model: YOLOv8 pre-trained on the COCO dataset
- Input Source: Live camera feed or captured images
- Output: Annotated frame with detected objects and corresponding audio feedback

The system is capable of detecting multiple objects simultaneously and generating meaningful speech output in real time.

E. Work flow Summary

- Capture live frame from the camera.
- Preprocess the frame and feed it into the YOLOv8 model.
- Detect and classify objects present in the frame.
- Estimate the distance of each detected object.
- Convert object labels and distance information into speech.
- Output the annotated frame and corresponding audio feedback.

This workflow creates a practical and user-friendly assistive system for visually impaired individuals.

IV. RESULTS AND DISCUSSIONS

A. Experimental Setup

Object Detection For Blind People Using AI

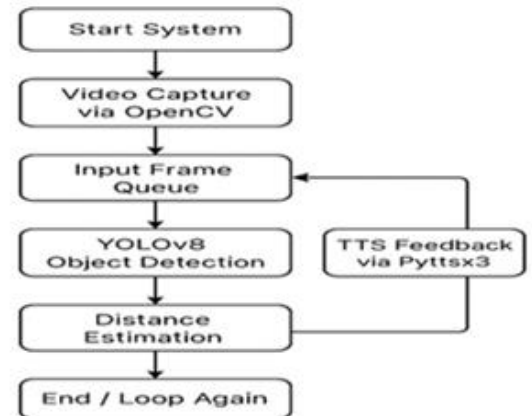


Figure. 1. Flowchart of the proposed YOLOv8-based object detection and audio output generation system.

The proposed system was implemented using Python, OpenCV, NumPy, and the YOLOv8 framework. The object detection model was based on pre-trained YOLOv8 weights trained on the COCO dataset. Experiments were conducted using live camera input in both indoor and outdoor environments containing common objects such as chairs, tables, people, vehicles, bottles, and electronic devices.

The evaluation focused on the following outputs:

- Object Detection Results — bounding boxes and labels for detected objects.
- Distance Estimation Results — approximate distance between the user and the detected object.
- Speech Output Results — audio feedback generated from detected object labels and distance values.

F. Detection Performance

The YOLOv8 model demonstrated strong performance in detecting multiple common objects in real time. The system successfully identified objects such as people, chairs, tables, bottles, and vehicles across different environmental conditions. The detection speed was suitable for assistive applications,

allowing users to receive immediate information about nearby objects.

Compared to earlier lightweight object detectors, YOLOv8 provided improved accuracy and more stable detection in scenes with multiple objects. The model was particularly effective in recognizing medium and large objects, while smaller or partially occluded objects showed slightly lower confidence in cluttered scenes.

G. Speech Output Evaluation

The text-to-speech module successfully converted the detected object labels and distance values into understandable speech output. The generated audio feedback was clear and informative, allowing users to receive real-time descriptions of nearby objects and their proximity.

For example, the system produced outputs such as:

- “Person detected at 2.3 meters.”
- “Chair detected at 1.1 meters.”
- “Bottle detected at 0.8 meters.”

This auditory feedback improved accessibility and made the system practical for visually impaired users in real-world situations.

QualitativeObservation



Figures2 and 3 depicts of equalitative detection results of the proposed YOLOv8 model in different real-world

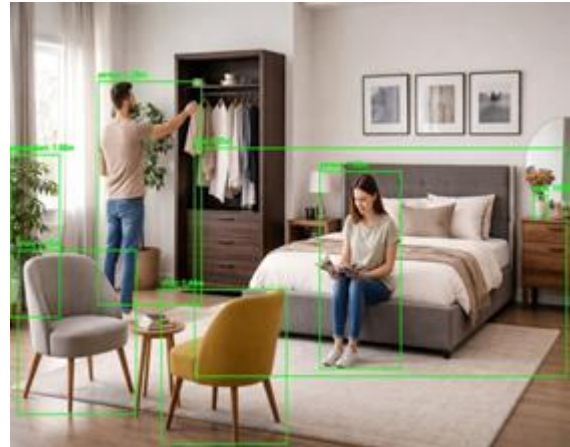


Figure.3.Detection in a crowded environment demonstrating robustness to background and lighting variations.

A. Conclusion

This project develops an AI-based assistive system for visually impaired people using computer vision and artificial intelligence. It uses YOLOv8 with the COCO dataset to detect common objects from camera input in real time. The system can identify multiple objects, estimate their distance from the user, and convert the object name and distance into audio output. This helps users understand nearby obstacles and move more safely. In the future, the system can be improved with better accuracy, wearable device integration, and faster processing for greater independence and safety.

B. Scope

The proposed system can be further improved by combining this system with wearable devices such as smart glasses or head-mounted cameras. This would make the system more portable and convenient for the visually impaired user, who can get real-time object and distance information while moving around. With small dimensions of the hardware and efficient processing, the system could become a useful assistance device for everyday use.

Another possible improvement can be to improve accuracy of object-detection and distance-estimation using advanced sensors like depth-camera or LiDAR to use with the YOLOv8. Training the system with larger and more diverse datasets outside of the COCO Dataset can also help the model to find more kinds of objects, and work in more complex environments, such as crowded streets or low-light conditions.

In the future, the system can also have voice command interaction and navigation assistance. Users could request the system to recognize certain objects or get instructions on

obstacles around and safe paths to take. Integrating navigation and mobile applications with GPS would make the system more intelligent and useful in the outdoor environment to help the visually impaired to move more independently and safely.

REFERENCES

1. Carion, N. et al. End-to-End Object Detection with Transformers (DETR), Facebook AI, 2020.
2. Zheng, Y. et al. LAM-YOLO: Drones-based Small Object Detection on Lighting-Occlusion Attention Mechanism YOLO, 2024.
3. Li, X. et al. A Novel Depth Map Restoration Strategy Based on Information Entropy-Guided, IEEE, 2023.
4. Wahyudi, D. et al. Toward Detection of Small Objects Using Deep Learning Methods: A Review, IEEE ICITEE, 2022
5. Deshpande, S. & Shriram, R. "Assistive Text on Hand Held Objects for Blind People." *Emerging Research in Computing, Information, Communication and Applications (ERCICA 2016)*, Springer, 2017, pp. 273–282.
6. Nagesh, A. N., Vijay, A. S., Salim, N. M., Vaisakh, M. C., John, J. "Assistive Technology for the Blind." *Advances in Communication Systems and Networks, Lecture Notes in Electrical Engineering*, vol. 656, Springer, 2020.
7. Nishajith, et al. "Deep learning based object detection and surrounding environment description for visually impaired people." PMID:PMC10360957, use of SSDLite MobileNetV2 + TTS on wearable glasses.
8. Adam, T. et al. "Enhanced Assistive Technology for Blind People Using Deep Transfer Learning."
9. *Scientific Reports, Nature*, 2025.
10. Islam, M. et al. "Low-Cost Deep Learning-Based Assistive System for Obstacle Detection for Visually Impaired People." *PubMed Central*, 2023.