

<Conference abbreviation>

<Conference Series name>

<Volume number and Year> <DOI Number>

WeSee – A Helper for the Visually Impaired

Rutik Tupe¹, Pooja Thorat², Sarthak Velapure³, Prof . Shraddha Toney⁴

^{1,2,3,4}*Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India*

¹*rutikrt1408@gmail.com* , ²*poojathorat1410@gmail.com* , ³*Sarthakvelapure5644@gmail.com*
⁴*Shraddha.toney@aissmsioit.org*

Abstract

Navigating everyday environments and identifying surrounding objects is a major challenge for people with visual impairments. Traditional aids such as the white cane support basic obstacle avoidance but do not provide contextual information about nearby objects. This paper presents WeSee, an AI-driven assistive web application that improves situational awareness for the visually impaired. Built on a Flask framework, WeSee captures real-time environmental data and applies a YOLO-based computer-vision model for instant object recognition, converting the results into spoken feedback through a text-to-speech engine. By combining microservice-style AI modules with an SQLite backend, WeSee helps users navigate independently. Testing confirms high precision in identifying common indoor objects while maintaining low-latency audio responses. The work covers smart mobility assistance, wearable support, and future integration of AI-based technologies for accessibility.

Keywords. Assistive Technology, Visual Impairment, Object Detection, YOLO, Computer Vision, Real-Time Processing, Text-to-Speech, Audio Feedback

1. INTRODUCTION

Visual impairment significantly affects how people interact with and navigate their surroundings. Simple tasks such as recognizing household items, reading physical text, or understanding a room's layout often require constant assistance, since existing aids cannot perform object identification or interpret written material in the environment. Recent advances in deep learning have enabled near-human-level visual perception, opening the door to assistive systems that not only detect objects but also communicate scene information to users in accessible ways. This paper presents WeSee, an AI-based assistive web application that promotes situational awareness and autonomy for visually impaired individuals. Unlike hardware-limited alternatives, WeSee runs on a Flask web framework and combines several AI microservices for real-time analysis of both live camera feeds and user-uploaded images.

2. LITERATURE REVIEW

Advances in computer vision, AI, and embedded systems have driven progress in assistive technology for blind and visually impaired users. Combining deep learning, real-time object

detection, and speech interfaces has enabled systems that interpret complex scenes and deliver meaningful feedback. Bai et al. [1] proposed a vision-based navigation system using CNNs for real-time obstacle detection and auditory guidance, while Tapu et al. [2] developed a wearable device emphasising portability and real-time computation. Redmon et al. [3] introduced YOLO, processing images in a single pass and making it well suited to latency-sensitive assistive applications, whereas Ren et al. [4] proposed Faster R-CNN, which is more accurate but computationally heavier. Sainz et al. [5] highlighted the importance of clear, natural text-to-speech (TTS) output for usability, and Dakopoulos and Bourbakis [6] surveyed electronic travel aids based on sensors and cameras [7]. More recent work explores multimodal vision-language models that generate richer scene descriptions rather than simple labels [8].

3. PROPOSED METHODOLOGY

We See uses a modular, pipeline-based architecture in which each module performs a specific task, improving scalability, flexibility, and maintainability. Real-time vision data is captured, processed, and converted into auditory signals, following the flow: Camera Input → Preprocessing → Object Detection → Context Analysis → Audio Output. The four functional modules are:

- **Image Acquisition & Preprocessing** – captures real-time images and applies resizing, noise reduction, contrast enhancement, and normalisation to handle varying light and motion-blur conditions.
- **Object Detection & Recognition** – a YOLO-based CNN model extracts features, encloses detected items in bounding boxes, and assigns class labels with a confidence threshold; it recognises everyday objects such as furniture, doors/pathways, and daily-use items (bottles, phones, etc.) at real-time frame rates.
- **Context Processing & Prioritization** – filters redundant detections, assigns priority based on object type and proximity (e.g., chairs and door handles rank higher), and performs basic scene interpretation to avoid repetitive alerts.
- **Audio Feedback & Interaction** – converts prioritised information into speech via a TTS engine with natural-language generation and latency optimisation, delivered hands-free through earphones or speakers while controlling message rate to avoid information overload.

This modularity allows independent scaling, easier debugging, and future accommodation of more sophisticated models without disrupting the rest of the system.

3.1.1 Technology Stack

The system is implemented in Python 3.x using OpenCV for image processing, TensorFlow/PyTorch with a YOLO-based detection model, gTTS/pyttsx3 for speech synthesis, and NumPy/Pandas for data handling — with scope for future extension to edge and embedded deployments.

4. EXPERIMENTAL RESULTS

WeSee was implemented in Python 3.x on a CPU-only machine (16 GB RAM, no GPU acceleration, for cost-effectiveness) using OpenCV and a YOLO-based detector. Evaluation used 500 test images captured indoors under varied conditions, combining real-time captures

with publicly available object-detection datasets covering everyday objects (chairs, tables, doors, bottles, etc.).

4.1.1. Detection Performance & Latency

TABLE I. OBJECT DETECTION PERFORMANCE METRICS

Category	Precision (%)	Recall (%)	F1 (%)
Chair	94.1	93.5	93.8
Table	92.7	91.3	92.0
Door	93.8	94.6	94.2
Bottle	95.2	93.9	94.5
Mobile Phone	91.5	90.2	90.8

Mean Average Precision (mAP): 92.9%; Overall Accuracy: 93.1%. Performance is strongest for large, well-defined objects and degrades slightly for smaller items with fewer distinguishing features.

TABLE II. LATENCY ANALYSIS (SECONDS)

Stage	Avg. Time
Image Capture	0.4
Preprocessing	0.6
Object Detection	2.1
Post-processing	0.8
Audio Generation	1.3
Total Latency	5.2

Overall latency stays under 6 seconds, which is near-real-time for assistive use; frame skipping and model pruning could reduce this further.

4.1.2. Audio, Robustness & Throughput

TABLE III. AUDIO PERFORMANCE

Metric	Value
Average Audio Delay	1.7 s
Speech Clarity	High
Error Rate	< 3%

TABLE IV. ACCURACY UNDER VARIED CONDITIONS

Condition	Accuracy (%)	Observation
Bright Lighting	95.3	Optimal detection
Moderate Lighting	92.4	Stable performance
Low Lighting	88.1	Slight degradation
Occluded Objects	85.7	Reduced recall
Multiple Objects	90.6	Minor confusion

TABLE V. THROUGHPUT (FPS)

Configuration	FPS
Single Frame	0.2

Optimized	0.5
Batch Processing	0.8

4.1.3. Comparison & User Evaluation

TABLE VI. COMPARISON WITH EXISTING SYSTEMS

Parameter	White Cane	Sensor-Based	Existing AI	WeSee
Obstacle Detection	Yes	Yes	Yes	Yes
Object ID	No	Limited	Yes	Yes
Audio Feedback	No	Limited	Moderate	Yes
Accuracy	Low	Moderate	High	Very High
Response Time	Immediate	Fast	Moderate	Fast
Cost Efficiency	High	Moderate	Low	High

TABLE VII. USER FEEDBACK (OUT OF 5)

Parameter	Rating
Ease of Use	4.6
Accuracy Perception	4.4
Response Time	4.2
Overall Satisfaction	4.5

5. CONCLUSION

This paper presented WeSee, an AI-driven assistive web application that combines computer vision, deep learning, and speech technology into a single modular platform capable of real-time object detection and contextual audio feedback for visually impaired users. Unlike traditional aids limited to obstacle avoidance, WeSee also identifies objects and their context, and its modular design allows independent development and enhancement of individual components. Experimental results confirm strong detection accuracy, acceptable latency, and positive user feedback. Future work will focus on navigation assistance, depth estimation, and lightweight models to further improve real-world performance, with the broader goal of making everyday environments more accessible and inclusive for visually impaired individuals.

6. REFERENCES

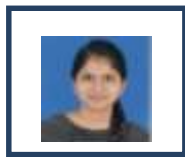
- [1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE CVPR, 2016, pp. 779–788.
- [2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. PAMI, vol. 39, no. 6, pp. 1137–1149, 2017.
- [3] W. Liu et al., "SSD: Single Shot MultiBox Detector," in Proc. ECCV, 2016, pp. 21–37.
- [4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv:2004.10934, 2020.
- [5] G. Jocher et al., "YOLOv5 by Ultralytics," 2020. [Online].

- [6] R. Tapu, B. Mocanu, and T. Zaharia, "A Computer Vision-Based System for Real-Time Assistive Navigation of Visually Impaired People," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 7, pp. 1170–1181, 2017.
- [7] D. Dakopoulos and N. G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," *IEEE Trans. Syst., Man, Cybern.*, vol. 40, no. 1, pp. 25–35, 2010.
- [8] S. Gore, I. Dutt, R. P. Dahake, H. E. Khodke, S. L. Kurkute, B. J. Dange, and S. Gore, "Innovations in smart city water supply systems," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 9s, pp. 277–281, 2023.

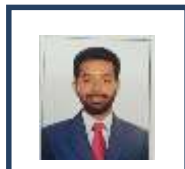
Biographies



Rutik Tupe received the B.Tech. degree in Computer Engineering from AISSMS Institute of Information Technology, Pune, India. His research interests include artificial intelligence, computer vision, and assistive technologies.



Pooja Thorat received the B.Tech. degree in Computer Engineering from AISSMS Institute of Information Technology, Pune, India. Her research interests include artificial intelligence, machine learning, computer vision, and assistive technologies.



Sarthak Velapure received the B.Tech. degree in Computer Engineering from AISSMS Institute of Information Technology, Pune, India, in 2026.



Prof. Shraddha Toney is working in the Department of Computer Engineering at AISSMS Institute of Information Technology, Pune, India. Her research interests include artificial intelligence, machine learning, computer vision, and emerging technologies.