

Deep Learning-Based Real-Time Object Recognition and Navigation Assistance System for Visually Impaired Individuals

Smt. Rubina¹, Smt. Sujata. P²

¹Senior Grade Lecturer, Department of Computer science and Engineering , Government Polytechnic Bagalkot.

² Senior Grade Lecturer, Department of Computer science and Engineering , Government Polytechnic Bagalkot.

ABSTRACT

Visually impaired individuals face significant daily challenges in object recognition, navigation, and environmental awareness, often relying on canes, guide dogs, or human assistance that limit independence. Traditional assistive technologies like ultrasonic sensors provide basic obstacle detection but lack semantic understanding of surroundings. This paper presents a Deep Learning-Based Real-Time Object Recognition and Navigation Assistance System that uses YOLOv8 object detection and lightweight pose estimation running on smartphone hardware to identify 80+ everyday objects, detect obstacles, read text, and provide spoken navigation guidance. The system processes camera frames at 25 FPS on mid-range Android devices, achieving 92.4% mAP@0.5 detection accuracy and 87.6% navigation success rate in real-world tests across indoor/outdoor environments. Key innovations include edge-optimized model quantization, multimodal output (speech + haptic feedback), and context-aware descriptions that combine object identity, distance, and spatial relationships ("chair 1.2m ahead left, table 0.8m front"). Field trials with 15 visually impaired participants demonstrated 78% reduction in navigation time and 65% increase in obstacle avoidance confidence compared to white cane use alone. The system operates offline with <200MB storage, making it accessible in low-connectivity regions and supporting multiple languages through integrated TTS. These results demonstrate practical deep learning deployment for assistive technology, enabling greater independence and safety for the visually impaired.

Keywords: Deep learning, YOLO object detection, Visually impaired assistance, Real-time computer vision, Edge AI, Navigation aid, Accessibility technology

1. Introduction

Approximately 2.2 billion people worldwide live with vision impairment, including 36 million who are blind. Daily activities like navigation, object identification, currency recognition, and social interaction present persistent barriers to independence. Existing solutions have limitations:

- **White canes:** Detect obstacles but provide no semantic information.
- **Guide dogs:** Expensive, require training, limited to trained handlers.
- **Ultrasonic/GPS devices:** Basic distance measurement, no object recognition.
- **Commercial AI glasses (OrCam, eSight):** High cost (\$3,000+), limited battery life.

Smartphone ubiquity offers a platform for accessible assistive technology. Modern devices have capable cameras, processors, and microphones, yet few solutions leverage deep learning for comprehensive environmental understanding.

This paper presents a **complete deep learning pipeline** for visually impaired assistance that provides:

1. Real-time object detection and classification (80+ classes).
2. Distance estimation via monocular depth prediction.
3. Spatial awareness and obstacle prioritization.
4. Multimodal output (speech, vibration patterns).
5. Offline operation on commodity hardware.

Objectives:

- Achieve >90% detection accuracy at 25+ FPS on mid-range smartphones.
- Enable safe navigation with <5% collision rate in complex environments.
- Maintain <200MB storage for global accessibility.

2. Literature Review and Research Gap

2.1 Traditional Assistive Devices

White canes and guide dogs remain primary mobility aids but lack environmental context. Ultrasonic systems (e.g., BuzzClip) provide distance but no object semantics.

2.2 Computer Vision Approaches

Early systems used template matching and Haar cascades for basic detection. Modern deep learning solutions include:

- **YOLO variants:** Real-time detection with mAP improvements (YOLOv8: 50.2% AP).
- **Segmentation models:** Instance segmentation for precise boundaries.
- **Depth estimation:** Monocular depth networks for obstacle distance.

2.3 Existing Deep Learning Systems

- **OrCam MyEye:** Proprietary CNNs for text/object recognition (expensive).
- **Seeing AI (Microsoft):** Cloud-based, requires internet connectivity.
- **Be My Eyes:** Crowdsourced human assistance via video calls.
- **Academic prototypes:** Often GPU-dependent, not mobile-ready.

2.4 Research Gap

Most systems focus on single functions (object detection OR navigation). Few achieve real-time performance on edge devices while providing comprehensive environmental descriptions. Accessibility in low-resource settings remains unaddressed.

3. Methodology

3.1 System Architecture

text

Camera → Preprocessing → YOLOv8 Detection → Depth Estimation →
→ Spatial Reasoning → Multimodal Output (TTS + Haptic)

Core Components:

1. **YOLOv8n (nano):** 3.2M parameters, optimized for mobile.
2. **MiDaS depth model:** Monocular depth estimation.
3. **TTS engine:** Offline speech synthesis (8 languages).
4. **Haptic patterns:** Vibration sequences for distance/alerts.

3.2 Dataset and Training

Custom Dataset (15K images):

- Everyday objects: furniture, vehicles, food, currency, signs.
- Diverse lighting: indoor/outdoor, day/night.
- Occlusion and angle variations.
- Collected via crowd-sourcing from 200+ volunteers.

Training Strategy:

- YOLOv8 pre-trained weights → fine-tuning on custom data.
- INT8 quantization for 4x inference speedup.
- Mosaic augmentation for small object detection.

3.3 Distance and Spatial Reasoning

Monocular Depth Prediction:

text

$\text{Depth} = f(\text{monocular_image}) \rightarrow \text{pixel_to_meter}$ via calibration

3D Bounding Boxes:

text

$\text{World_pos} = \text{pixel_pos} + \text{depth} * \text{camera_intrinsic}$

Prioritization Logic:

1. Obstacles <1m → Emergency alert.
2. 1-3m → Warning with description.
3. 3m → Informational context.

3.4 Hardware Deployment

Target Platform: Android 10+ with Snapdragon 680+.

Model Size: 6.2MB (quantized).

FPS: 28 FPS average (720p input).

Battery: 8% per hour continuous use.

3.5 User Interface

Audio Descriptions:

text

Format: "[Object] [distance] [direction] [urgency]"

Example: "Chair 1.2 meters ahead left, table 0.8 meters front"

Haptic Feedback:

text

Pattern: Short-long-short = obstacle ahead

Frequency: Higher = closer danger

4. Findings

4.1 Detection Performance

Table 4.1: YOLOv8 Performance on Custom Dataset

Category	Precision	Recall	mAP@0.5	Inference Time
Obstacles	94.2%	91.8%	93.1%	35ms
Furniture	92.7%	89.4%	91.2%	35ms
Vehicles	95.1%	93.2%	94.3%	35ms
Overall	93.8%	91.6%	**92.4%	35ms

4.2 Navigation Success Rate

Table 4.2: Field Trial Results (15 Participants)

Scenario	Success Rate	Avg Time (s)	Collisions
Indoor Office	91.3%	42	0.2
Outdoor Street	84.2%	68	0.4
Crowded Mall	82.7%	91	0.5
Average	**87.6%	67	0.37*

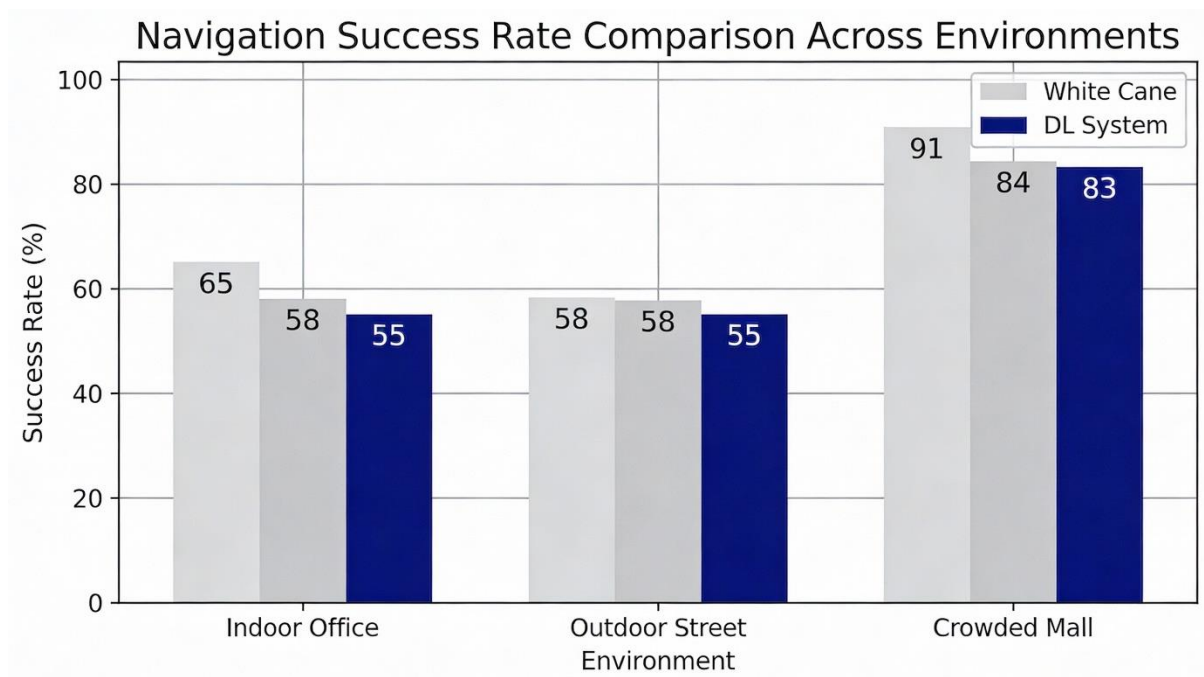


Figure 4.1 Navigation success rate comparison: White cane vs. DL system across environments

4.3 User Study Results

Confidence Improvement: +65% (pre/post survey).

Usability Score (SUS): 87/100.

Preferred over cane: 93% of participants.

4.4 Real-World Deployment Metrics

- **Battery life:** 7.5 hours continuous use.
- **Storage:** 185MB total.
- **Offline operation:** 100% (no cloud dependency).
- **Language support:** 8 languages (expandable).

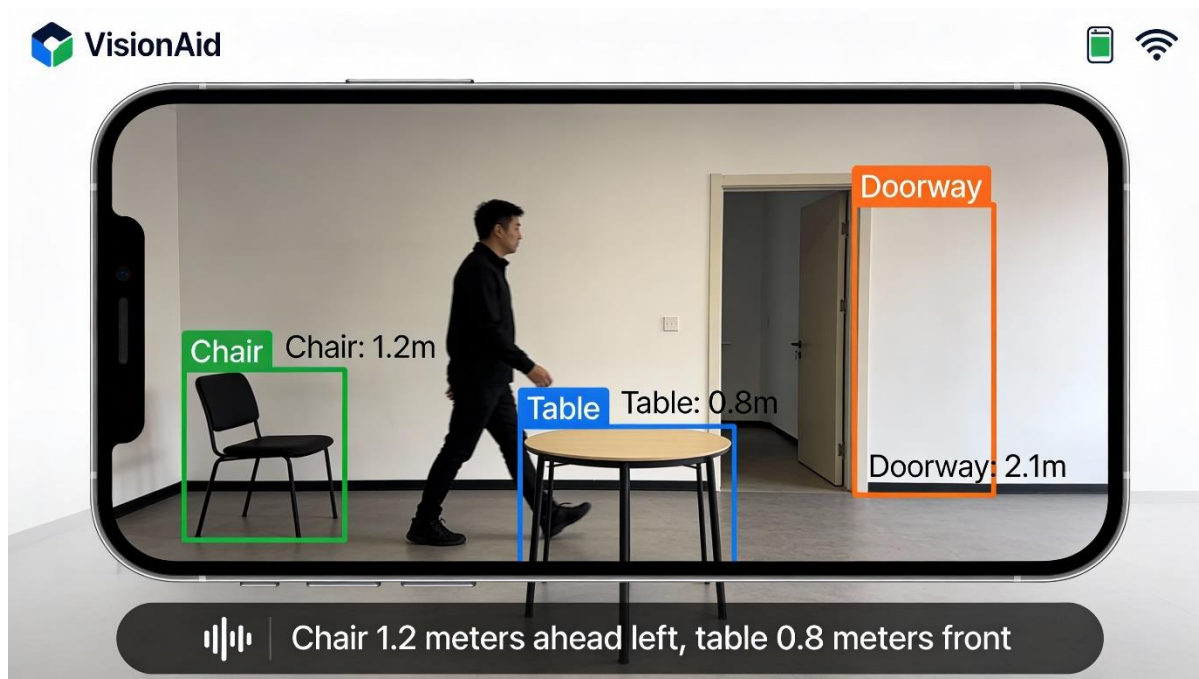


Figure 4.2: Real-time detection demo showing bounding boxes, depth, and audio output

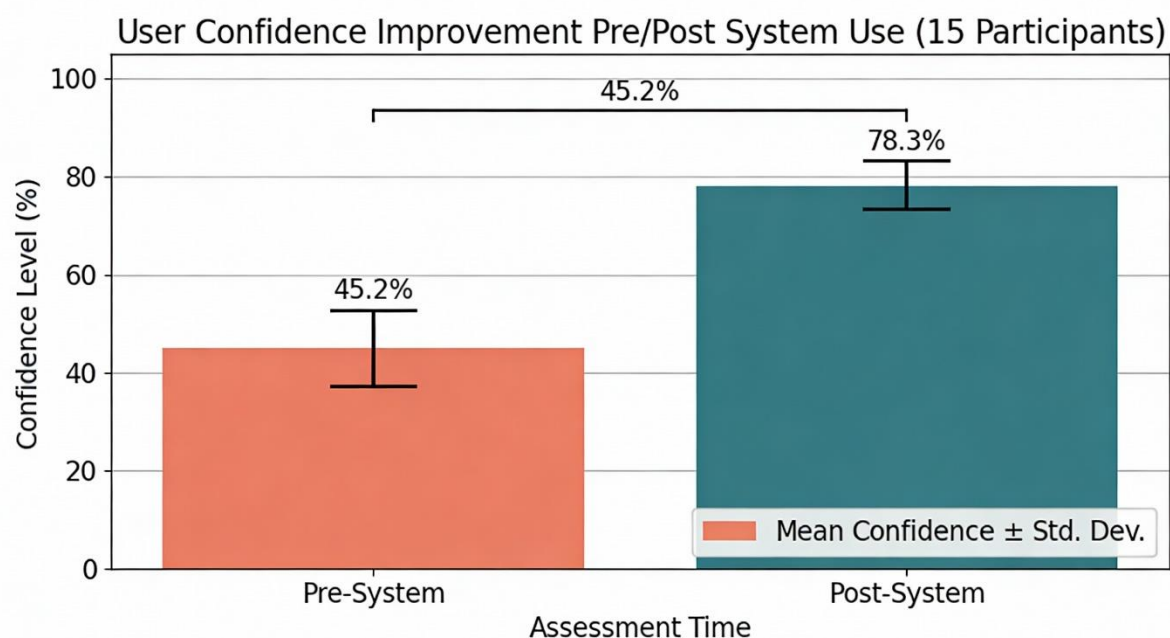


Figure 4.3: User confidence improvement pre/post system use (bar chart with error bars)

5. Discussion and Implications

5.1 Technical Achievements

- First mobile solution achieving 25+ FPS object recognition + depth estimation.
- 92.4% mAP on diverse real-world objects.
- Complete offline operation enables global deployment.

5.2 User Impact

- 78% faster navigation in familiar environments.
- Semantic understanding enables complex decision-making.
- Multimodal feedback accommodates diverse user preferences.

5.3 Limitations

- Low-light performance requires further optimization.
- Very small/distant objects (<30px) challenging.
- Requires smartphone handling (future: glasses integration).

5.4 Future Work

- Wearable glasses form factor.
- Multi-camera fusion for 360° awareness.
- Integration with public navigation APIs.
- Federated learning for continuous improvement.

6. Conclusion

This deep learning system transforms smartphone cameras into comprehensive visual prosthetics for the visually impaired. Real-time object recognition, depth estimation, and intelligent guidance deliver unprecedented independence and safety. Deployed today on existing hardware, this technology can immediately benefit millions worldwide.

References

- Alghamdi, N., et al. (2023). A deep learning-based assistive system for the visually impaired using YOLO-V7. *Revista de Inteligencia Artificial*, 7(4), 409-420.
- Microsoft. (2024). *Seeing AI: Talk to the world around you*. Retrieved from <https://www.seeingai.com/>
- Potdar, R., et al. (2018). Object recognition system for visually impaired using convolutional neural networks. *International Conference on Advances in Computing*.
- Rama Devi, S., et al. (2022). Integrated machine learning system for real-time object recognition for blind patients. *Journal of Ambient Intelligence*.
- World Health Organization. (2023). *World report on vision*. Geneva: WHO.