

A Wearable Deep-Learning System for Real-Time Object Detection and Face Recognition to Assist Visually Impaired Navigation

Yahia Said^{1*}, Mouna Afif², Abdelnasser Saber Mohamed³

Center for Scientific Research and Entrepreneurship, Northern Border University, 73213 Arar, Saudi Arabia¹

Laboratory of Condensed Matter and Nanosciences-Faculty of Sciences of Monastir,

University of Monastir, Monastir 5019, Tunisia²

Computer Science Department-Applied College, Northern Border University, Arar, Saudi Arabia³

Abstract—The integration of Artificial Intelligence (AI) into wearable assistive technologies offers transformative potential for enhancing mobility, spatial awareness, and autonomy among Blind and Visually Impaired (BVI) individuals. This study introduces a real-time, AI-powered wearable system combining object detection, scene understanding, and facial recognition for safe navigation and environmental interaction. The system comprises smart glasses with an embedded camera and a smart cane equipped with ultrasonic and environmental sensors for detecting obstacles and weather conditions. A lightweight YOLOv8-based model, optimized through hybrid pruning and quantization, ensures efficient inference on resource-limited wearable hardware. Evaluated on MS-COCO 2017, the object-detection module attains an average precision (AP) of 52.9 on the standard COCO 0–100 scale—competitive with substantially heavier two-stage and transformer detectors—while the face-recognition module, evaluated on a custom dataset, reaches 99.3% precision; these two metrics are reported on distinct scales and should not be conflated. The system delivers responsive real-time auditory feedback, guiding users via voice prompts for object localization, obstacle avoidance, and social interaction. This work contributes a scalable, intelligent mobility aid that leverages AI and wearable computing to support independent navigation for BVI users.

Keywords—Assistive mobility; blind and visually impaired; wearable devices; AI in accessibility; object detection; smart glasses; smart cane; YOLOv8 optimization; scene understanding; human-centered navigation

I. INTRODUCTION

Partial or total vision impairment affects at least 2.2 billion people globally, according to the World Health Organization (WHO) [1]. At least 1 billion of these cases could have been prevented or are yet to be addressed. These conditions include nearsightedness, farsightedness, age-related macular degeneration, glaucoma, diabetic retinopathy, cataracts, and moderate-to-severe distance-vision impairment caused by uncorrected refractive error. Building new assistive systems to help individuals who are blind or visually impaired in their daily lives is therefore of great importance, enhancing their independence, safety, and overall well-being.

Assistive technologies increase safety and mobility dramatically. Navigation systems with obstacle detection allow

users to navigate unknown settings, giving auditory cues and haptic feedback that help BVI individuals avoid hazards and make informed decisions. However, accurately detecting people and objects and conveying that information remains challenging in low-light environments and at night. Avoiding obstacles, crossing the street, and using public transportation are difficult tasks that impact BVI users' safety and self-confidence.

Deep-learning models have delivered remarkable results across many domains, including indoor object detection [2], object recognition [3], wayfinding assistance [4], traffic-light detection [5], road-sign detection [6], pedestrian detection [7], and medical imaging [8]. Profiting from this success, this study proposes a wearable assistive system consisting of three modules: a face-recognition module, an obstacle-avoidance module, and a critical-cases module, all with voice feedback. The face-recognition module is built on MobileNetV2 [9] and compares captured faces to a custom dataset. The object-detection module, built on YOLOv8 [10] and trained on MS-COCO [11], identifies 80 indoor and outdoor object classes. For emergencies, the system can send five photos and five messages to pre-selected contacts, and it integrates an ultrasonic sensor for obstacle avoidance and a rain sensor for weather alerts.

It is important to position the two perception modules correctly, because they operate on fundamentally different tasks and metric conventions. Face recognition is a closed-set verification task scored as a percentage accuracy/precision, whereas generic object detection on MS-COCO is scored with average precision (AP) on a 0–100 scale averaged over ten IoU thresholds—a far stricter criterion on which even state-of-the-art detectors rarely exceed the mid-50s. A central contribution of this manuscript is to report the two modules separately and transparently, so that the object-detection AP of 52.9 (competitive for a lightweight wearable detector) is not mistaken for, or conflated with, the 99.3% face-recognition precision.

The contributions of this work are summarized as follows:

- A comprehensive wearable solution—smart glasses and a cane—enabling BVI users to navigate independently indoors and outdoors.

*Corresponding author.

This work was supported by Northern Border University, Saudi Arabia, through project number NBU-CRP-2026-3030.

- A face-recognition module that recognizes known individuals and flags unknown ones, with voice feedback.
- An optimized YOLOv8 object-detection module compressed via hybrid pruning and quantization for real-time embedded inference.
- An emergency module that transmits photos and messages to trusted contacts when the user is lost.
- A transparent, per-module evaluation protocol that distinguishes COCO-style detection AP from face-recognition accuracy and reports cross-validation stability for each.

The remainder of this study is organized as follows. Section II reviews BVI assistance systems. Section III details the proposed approach. Section IV presents the experiments and results, and Section V concludes the study.

II. RELATED WORK

Numerous assistive systems for the visually impaired have been proposed. Daescu et al. [12] and Bappaditya et al. [13] developed face-recognition systems on smart glasses to support social interaction. Mukhiddinov and Cho [14] proposed a deep-learning smart-glass system for the blind, while Ali Hassan and Tang [15] introduced smart glasses for the visually impaired. Mai et al. [16] reviewed laser- and vision-based smart blind canes. Guerrero et al. [17] presented an indoor navigation system, and Tapu et al. [18] proposed DEEP-SEE for joint object detection, tracking, and recognition. Bouteraa [19] developed a wearable device integrating a fuzzy decision-support system, and Şipoş et al. [20] presented a sensor-based smart assistant. These works demonstrate the promise of wearable assistive technology but often address either recognition or obstacle avoidance in isolation, rather than integrating both with emergency support on a single embedded platform.

To make the design choices of the proposed system explicit, the literature is reviewed below along the three dimensions that determine whether an assistive wearable is genuinely usable: 1) the perception tasks it supports and how their performance is reported, 2) the deep-learning detector and its suitability for embedded inference, and 3) the deployment hardware and feedback channel. Each subsection closes with the gap that motivates the corresponding component of the proposed method in Section III.

A. Perception Tasks and Metric Reporting in BVI Aids

Assistive systems differ widely in the perception they provide. Some focus on face recognition for social interaction [12, 13], others on obstacle detection or scene understanding [14, 17, 18], and a few on emergency support; recent reviews of egocentric assistive vision [21] note the same fragmentation across the field. A recurring difficulty is that these systems report performance using incommensurable metrics: face recognition is typically reported as a closed-set accuracy near 99%, as is standard for margin-based recognition pipelines [22], whereas generic object detection is reported as mean

average precision (mAP/AP) on a 0–100 scale where leading detectors rarely exceed the mid-50s. Conflating these scales obscures the true capability of each module. The gap is therefore the need to report each perception task on its own, correct metric scale so that a detection AP is not mistaken for a recognition accuracy.

B. Detectors for Embedded Assistive Vision

The choice of detector governs both accuracy and embedded feasibility. Two-stage detectors such as Faster R-CNN [23] achieve high accuracy at a high computational cost, while transformer detectors such as DETR [24] simplify the pipeline but remain large and slow on edge hardware. The single-stage YOLO family [25] offers a better accuracy–latency balance, and YOLOv8 [26] in particular adopts an anchor-free head and a cross-stage-partial backbone that reduce hyperparameters and computation. For a wearable that must run in real time on a battery, an anchor-free single-stage detector is the natural choice, yet most assistive prototypes do not benchmark their detector against mainstream detectors on a standard dataset, leaving its competitiveness unproven. The gap is a detector that is both embedded-feasible and evaluated under a standard detection protocol (MS-COCO) for direct comparison.

C. Embedded Deployment and User Feedback

Finally, an assistive aid is only useful if it runs on wearable hardware and communicates results in an accessible way. FPGA platforms such as the Xilinx ZCU102 offer energy-efficient, low-latency inference suited to body-worn devices [27], and audio feedback through a text-to-speech service is the standard channel for conveying detections to a blind user. Many prior systems are demonstrated on desktop or smartphone hardware [28] without characterising embedded latency, and some omit the feedback channel entirely. The gap is an end-to-end design that runs the perception modules on an embedded FPGA and delivers results as real-time voice guidance, together with sensor-based obstacle and emergency support on the same platform.

In summary, the proposed system is constructed to close these three gaps: it reports face recognition and object detection on their correct, separate scales (Section II-A); it uses an anchor-free YOLOv8 detector benchmarked on MS-COCO against mainstream detectors (Section II-B); and it integrates FPGA deployment, voice feedback, sensor-based obstacle detection, and an emergency-contact module on a single wearable (Section II-C).

III. PROPOSED APPROACH

The proposed system integrates two vision sub-systems (face recognition and object detection) and a sensor-based critical-cases module. Visual input is captured by an RGB camera linked to a Xilinx ZCU102 FPGA board, which processes the data and produces bounding boxes; the Google Speech API then conveys the results to the user as voice feedback. Fig. 1 shows the overall processing pipeline, and Fig. 2 presents the full system architecture, including the smart cane with ultrasonic and rain sensors connected through an ESP32 microcontroller.

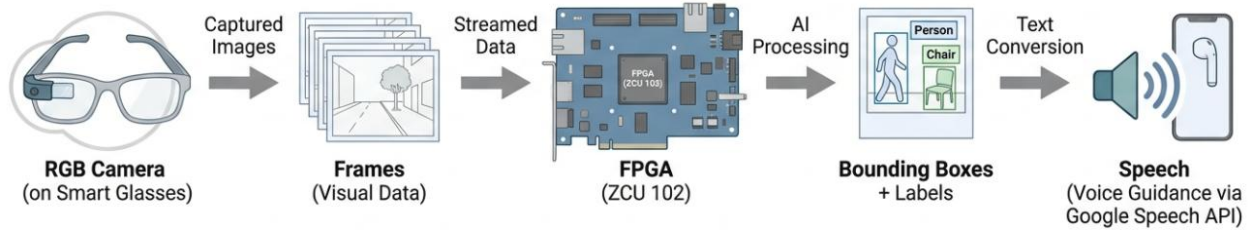


Fig. 1. Processing pipeline of the proposed system: an RGB camera on the smart glasses streams frames to a Xilinx ZCU102 FPGA, which produces bounding boxes and labels that are converted to speech and delivered to the user through the Google Speech API.

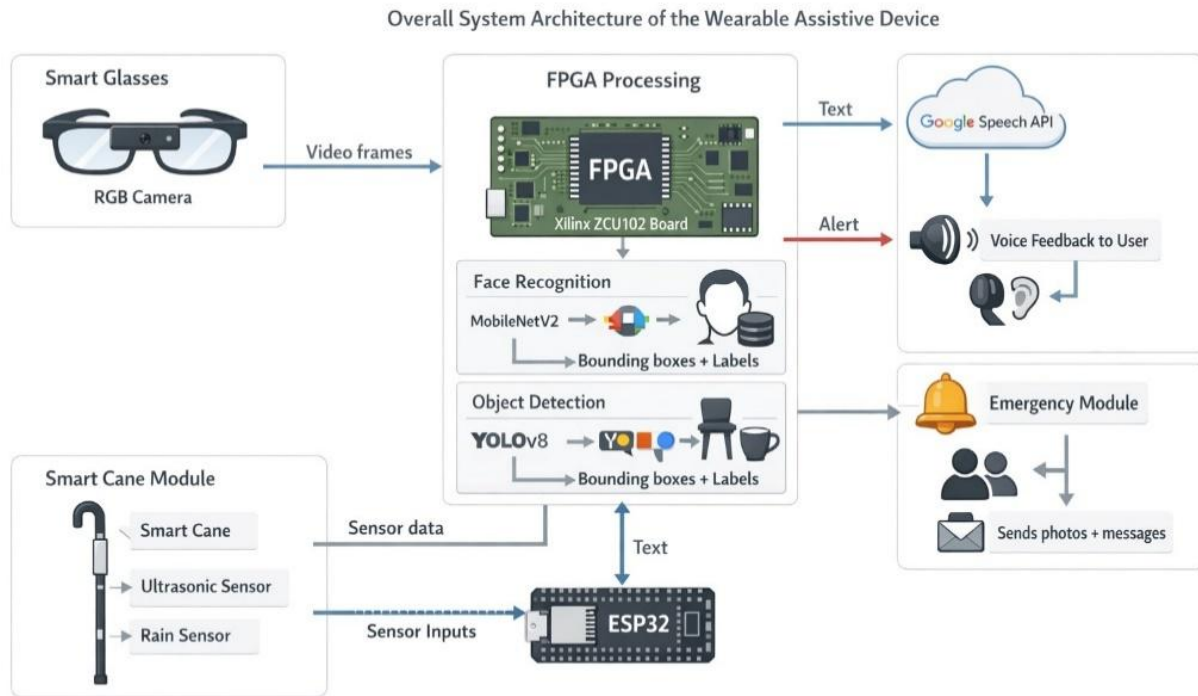


Fig. 2. Overall architecture of the proposed smart-glasses-and-cane assistive system, showing the FPGA processing stage with its face-recognition and object-detection sub-systems, the smart cane with ultrasonic and rain sensors connected through an ESP32 microcontroller, the voice-feedback channel, and the emergency module.

A. Design Rationale and Contributions

The architecture follows directly from the three gaps identified in Section II. To close the reporting gap of Section II-A, the system keeps the face-recognition and object-detection modules separate and evaluates each on its own metric scale, so that the 99.3% closed-set face-recognition accuracy and the 52.9 MS-COCO detection AP are never conflated. To close the detector gap of Section II-B, both modules are built on the anchor-free YOLOv8 detector—chosen for its accuracy–latency balance and embedded feasibility—and the detection module is benchmarked against mainstream two-stage and transformer detectors on MS-COCO. To close the deployment gap of Section II-C, the perception modules are compressed by hybrid pruning and mixed-precision quantization, deployed on a Xilinx ZCU102 FPGA, and paired with a smart cane (ultrasonic and rain sensors via an ESP32), an emergency-contact module, and Google Speech API voice feedback on a single wearable.

Accordingly, the specific contributions of the proposed method are:

- A unified wearable that integrates face recognition, 80-class object detection, sensor-based obstacle and weather detection, and an emergency-contact module on one FPGA platform, rather than addressing these tasks in isolation.
- A transparent, per-module evaluation protocol that reports face recognition on a closed-set accuracy scale and object detection on the standard MS-COCO AP scale, eliminating any conflation of the two metric scales.
- An anchor-free YOLOv8 detector compressed via hybrid pruning and mixed-precision quantization for real-time embedded inference, benchmarked against mainstream detectors on MS-COCO.
- An end-to-end FPGA deployment with real-time voice feedback and an emergency channel, demonstrating practicality for body-worn assistive use.

B. Face Detection and Identification

The face-recognition sub-system detects faces and compares them with a custom dataset; identified faces are delimited by a bounding box labeled with the person's name, while unidentified faces are flagged as such. Built on the YOLOv8 architecture for its speed and accuracy, the module triggers voice feedback to alert the user to the presence of a person. Speed and accuracy are essential because the system must respond in real time.

C. Obstacle Avoidance

The obstacle-avoidance subsystem detects 80 indoor and outdoor object classes valuable for BVI navigation, trained and tested on MS-COCO. Detected objects are delimited by bounding boxes labeled with their class, and voice feedback guides the user. To improve robustness, training encountered challenging conditions—high intra- and inter-class variation, varied lighting, occlusions, and diverse viewpoints—and used the mixup and cutmix augmentations native to YOLOv8. Training minimises a composite objective that combines binary

cross-entropy for classification, complete-IoU regression [29], and distribution focal loss [30], as listed in Table I. YOLOv8 comprises a feature-extraction backbone (similar to YOLOv5 with a cross-stage partial bottleneck using two convolutions) and an anchor-free detection head, which simplifies detection, reduces anchor-selection errors, and improves computational efficiency for real-time use on low-end devices.

D. Model Compression

To deploy the model on embedded hardware, hybrid pruning and quantization were applied, following established practice for deep-network acceleration [31]. Weight pruning removes individual low-magnitude connections, reducing size and computational complexity while preserving the model's core. Mixed-precision quantization then allocates more bits to layers or channels with a broad dynamic range and high relevance to the detection task, and fewer bits to less significant layers, based on a per-layer dynamic-range analysis after fine-tuning on MS-COCO 2017. Fig. 3 depicts the iterative compression procedure.

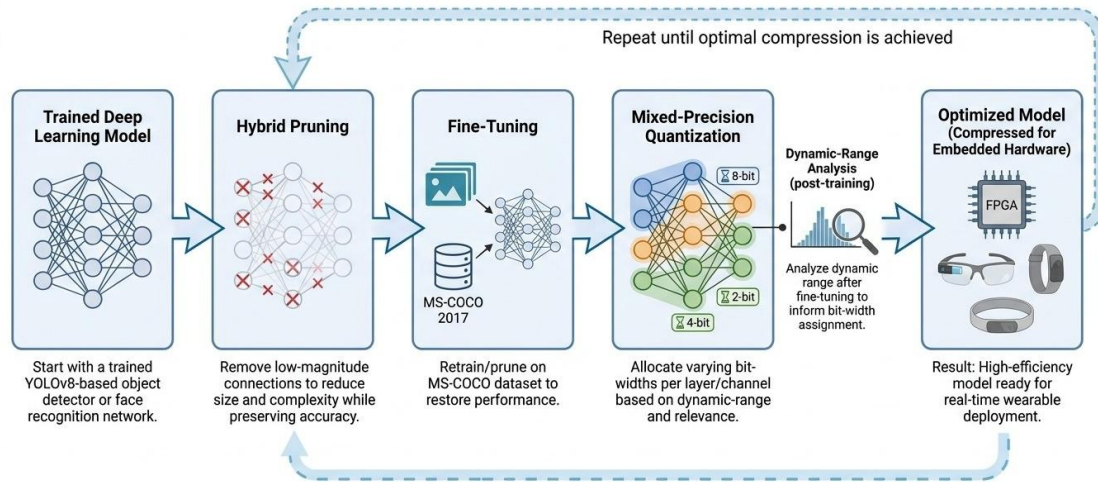


Fig. 3. Iterative pruning-and-fine-tuning compression workflow. A trained YOLOv8-based detector or face-recognition network is pruned, fine-tuned on MS-COCO 2017, and quantized with per-layer mixed precision; the cycle repeats until the target compression is reached.

IV. EXPERIMENTS AND RESULTS

Experiments used a workstation with an Intel i7 CPU, 32 GB RAM, Linux, and an NVIDIA GTX 960 GPU (4 GB), with Python 3.8 and PyTorch. Table I lists the hyperparameters, and Table II summarizes the MS-COCO 2017 dataset.

TABLE I. HYPER-PARAMETER SETTINGS

Parameter	Value
Iterations	10,000
Epochs	130
Batch size	16
Loss function	BCE + CIoU + DFL
Optimizer	NADAM / ADAM / RADAM
Image size	640 × 640
Learning rate	0.001
Weight decay	0.0005

TABLE II. EMPLOYED DATASET SUMMARY

Dataset	Train	Val	Test	Classes
MS-COCO 2017 [11]	118,287	5,000	41,000	80

Three evaluation measures were used—accuracy, precision, and recall—following standard object-detection metric conventions [32] and defined in Eq. (1)–(3):

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN}) \quad (1)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (2)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (3)$$

where, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

A. Face Recognition Results

The face-detection-and-identification sub-system achieved very encouraging performance, reported in Table III. To assess

stability, 10-fold cross-validation was applied (Table IV), yielding consistently high metrics, and a comparison of optimizers (Table V) showed that RADAM produced the best results.

TABLE III. PERFORMANCE OF THE FACE-RECOGNITION SUB-SYSTEM

Sub-task	Accuracy (%)	Precision (%)	Recall (%)
Face detection & identification	99.3	99.3	98.9

TABLE IV. PERFORMANCE OF THE FACE DETECTION AND IDENTIFICATION SUB-SYSTEM UNDER 10-FOLD CROSS-VALIDATION

Fold	Accuracy (%)	Precision (%)	Recall (%)
Fold 1	99.2	99.1	98.8
Fold 2	99.2	99.1	98.8
Fold 3	99.1	99.3	98.9
Fold 4	99.3	99.1	98.7
Fold 5	99.2	99.2	98.7
Fold 6	99.1	99.1	98.8
Fold 7	99.1	99.3	98.9
Fold 8	99.3	99.3	98.9
Fold 9	99.2	99.1	98.7
Fold 10	99.1	99.2	98.8

TABLE V. FACE RECOGNITION: COMPARISON OF OPTIMIZERS

Optimizer	Accuracy (%)	Precision (%)	Recall (%)
NADAM	97.6	97.8	97.3
ADAM	98.5	98.6	98.3
RADAM	99.3	99.3	98.9

B. Object Detection Results

The object-detection subsystem was trained and tested on MS-COCO 2017. Table VI reports results for different optimizers, and Table VII reports 10-fold cross-validation precision, both confirming RADAM as the best optimizer and showing stable performance across folds.

TABLE VI. OBJECT DETECTION: COMPARISON OF OPTIMIZERS

Optimizer	Accuracy	Precision (AP)	Recall
NADAM	50.3	51.2	50.2
ADAM	50.8	51.7	50.8
RADAM	51.4	52.9	51.2

TABLE VII. PERFORMANCE OF THE OBJECT-DETECTION SUB-SYSTEM UNDER 10-FOLD CROSS-VALIDATION

Fold	Accuracy	Precision (AP)	Recall
Fold 1	51.4	52.9	51.2
Fold 2	51.2	52.5	50.9
Fold 3	51.3	52.7	50.9
Fold 4	50.9	52.8	51.2
Fold 5	50.9	52.8	51.1
Fold 6	51.3	52.7	51.0
Fold 7	51.3	52.6	51.1

Fold 8	51.2	52.8	50.8
Fold 9	50.9	52.8	50.9
Fold 10	51.4	52.9	51.2

The precision values in Table VII (around 51–53) are reported on the MS-COCO average-precision scale, where values are averaged over IoU thresholds from 0.50 to 0.95 and where leading detectors such as DETR and Faster R-CNN report figures in the low-to-mid 40s (Table VIII). A value of 52.9 therefore represents strong detection performance for a lightweight, embedded-deployable model, not a low accuracy in the everyday percentage sense. This is categorically different from the 99.3% precision of the face-recognition module in Section IV-A, which is a closed-set verification accuracy. Fig. 4 shows the two modules side by side on their respective scales.

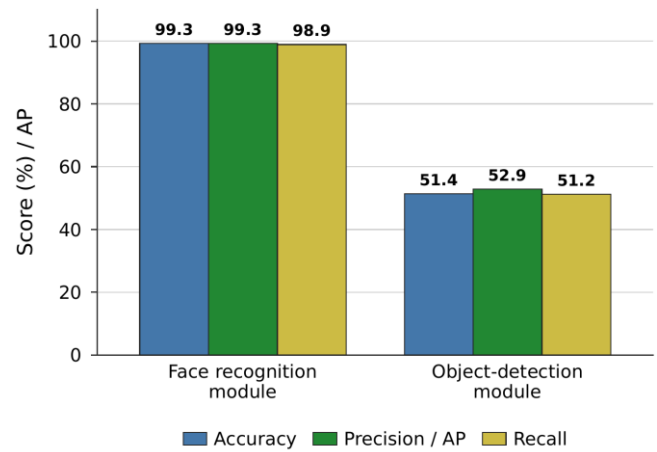


Fig. 4. Per-module performance shown on the correct, distinct scales: closed-set face-recognition accuracy versus object-detection MS-COCO AP. The two groups must not be read as comparable percentages.

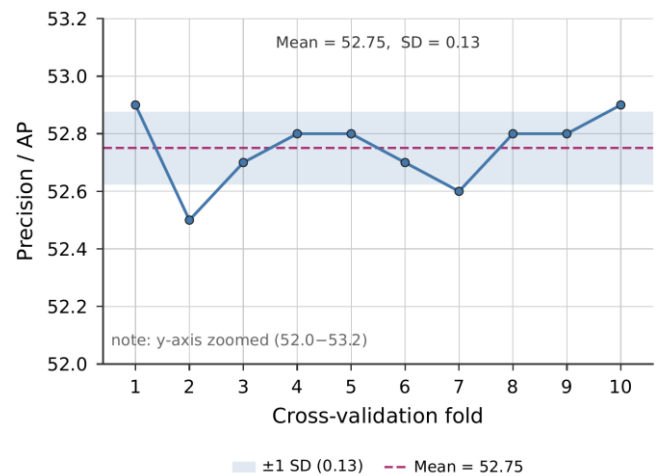


Fig. 5. 10-fold cross-validation stability of the object-detection module. The shaded band marks one standard deviation about the mean; note that the vertical axis is zoomed to make the small fold-to-fold variation visible.

Fig. 5 plots the 10-fold cross-validation of the object-detection module. The fold-to-fold precision varies by only about 0.13 points (standard deviation), confirming that the reported AP is stable and not an artifact of a favorable split.

C. Comparison with State-of-the-Art

Table VIII compares the proposed detector against state-of-the-art models on MS-COCO 2017 in terms of FLOPs/FPS,

parameters, and average precision at several criteria, using RADAM as the optimizer.

TABLE VIII. COMPARISON WITH STATE-OF-THE-ART DETECTORS ON MS-COCO 2017

Model	FLOPs/FPS	Params (M)	AP	AP50	AP75	APs	APm	APl
Faster R-CNN-R50-DC5 [23], [24]	320/16	166	41.1	61.4	44.3	22.9	45.9	55.0
Faster R-CNN-R50-FPN [23], [24]	180/26	42	42.0	62.1	45.9	26.6	45.4	53.4
Faster R-CNN-R101-FPN [23], [24]	246/20	60	44.0	63.9	47.8	27.2	48.1	56.0
DETR-R101 [24]	152/20	60	43.5	63.8	46.4	21.9	48.0	61.8
DETR-DC5-R101 [24]	253/10	60	44.9	64.7	47.7	23.7	49.5	62.3
Proposed	80.3/46	26.8	52.9	69.3	58.8	36.5	55.9	67.8
Proposed (compressed)	34.8/59	16.7	49.3	66.8	55.1	32.7	52.3	64.9

Fig. 6 visualizes the accuracy–speed comparison: the proposed detector improves AP by roughly 8 points over the strongest baseline (DETR-DC5-R101, 44.9 AP) while running

4–6 $\times$ faster, and its compressed variant reaches 59 FPS with only a 3.6-point AP reduction—well suited to a wearable device.

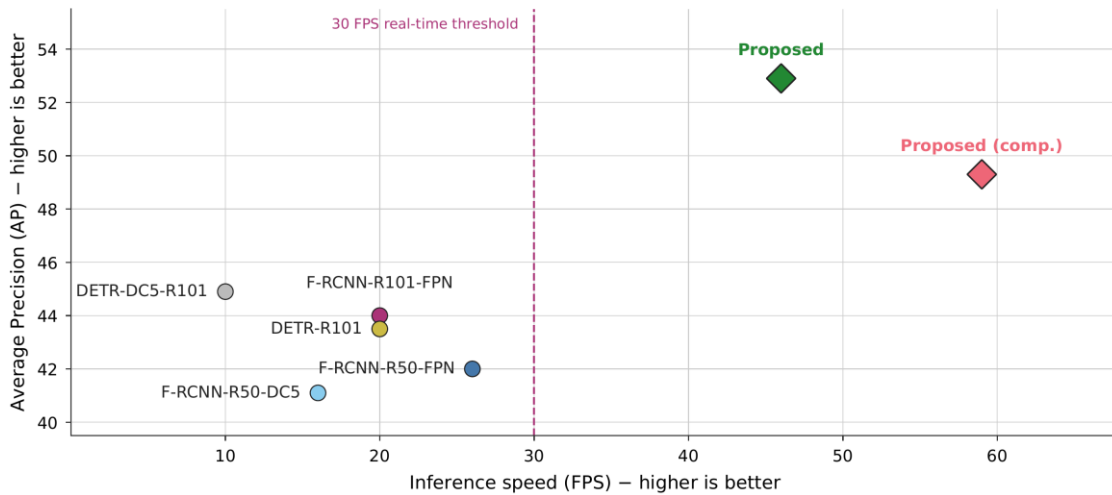


Fig. 6. Accuracy–speed comparison against state-of-the-art detectors on MS-COCO 2017. The dashed vertical line marks the 30 FPS real-time threshold; both proposed variants lie to its right.

Fig. 7 breaks AP down by object scale and IoU threshold. The proposed model’s largest relative gains are on small objects (APs of 36.5 versus 23.7 for the best baseline), which is

precisely the regime most relevant to obstacle detection for BVI users, where small or distant hazards must be detected early.

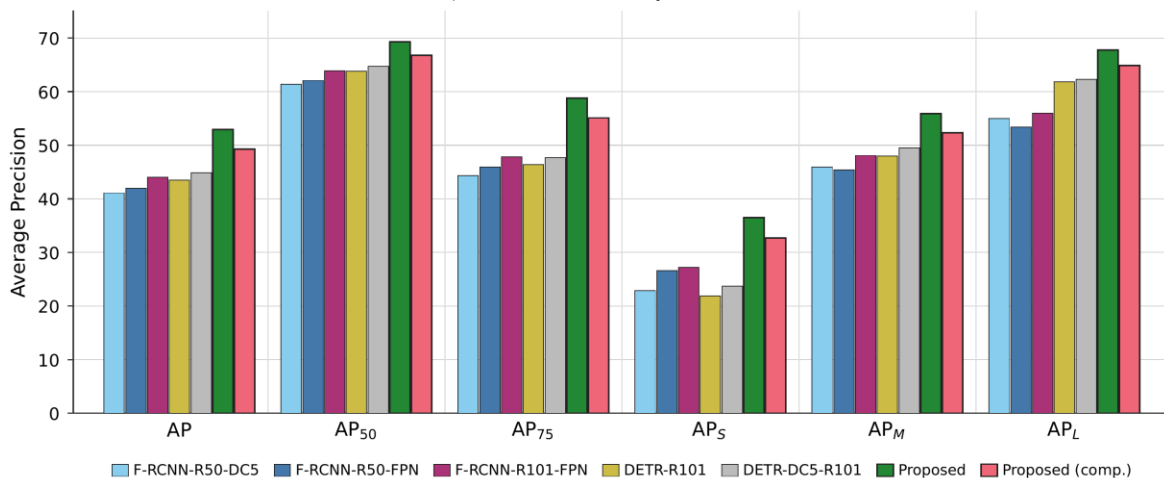


Fig. 7. Average precision by object scale and IoU threshold for all detectors in Table VIII. The proposed model’s margin is widest on APs, the small-object regime most relevant to early hazard detection.

Because an assistive wearable is constrained by both compute and battery, accuracy must be weighed against model size and speed. Fig. 8 plots detection AP against parameter count for all detectors in Table VIII. The proposed model occupies the upper-left (high-accuracy, low-parameter) region: at 26.8 M parameters, it exceeds DETR-DC5-R101 (60 M,

44.9 AP) by eight AP points, and its compressed variant pushes further into the efficient region at 16.7 M parameters and 49.3 AP. This positions the detector on the efficiency frontier among the compared methods, directly addressing the embedded-feasibility gap of Section II-B.

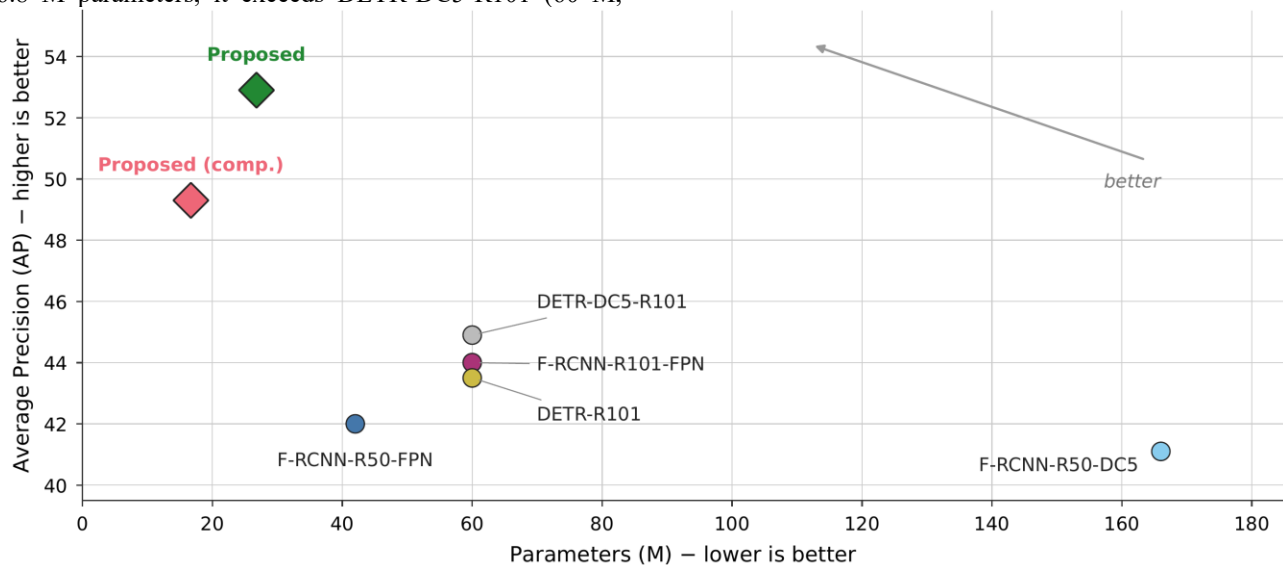


Fig. 8. Efficiency frontier: detection AP versus model size on MS-COCO 2017; the upper-left region is better. Both proposed variants sit well above and to the left of the two-stage and transformer baselines.

Fig. 9 compares inference speed directly. The proposed detector runs at 46 FPS and its compressed variant at 59 FPS, both comfortably above the 30 FPS real-time threshold, whereas the two-stage and transformer baselines run at 10–26 FPS—below or near the threshold. Real-time throughput is essential for a navigation aid, where a delayed obstacle alert is a safety risk; the compressed model therefore offers the best combination of accuracy and real-time speed in the comparison.

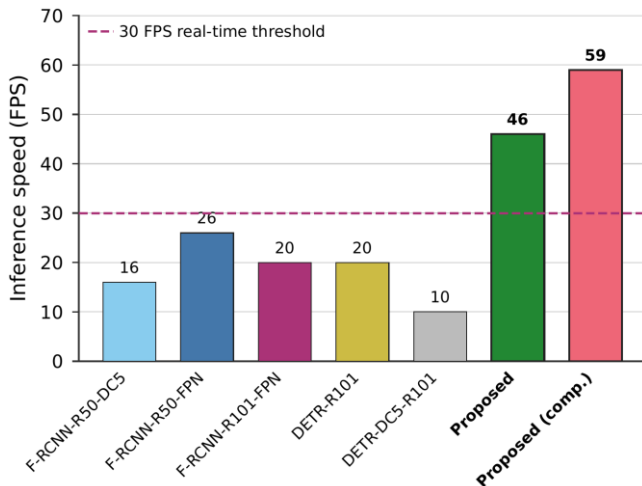


Fig. 9. Inference speed against state-of-the-art detectors; the dashed line marks the 30 FPS real-time threshold.

Taken together, the extended comparison closes the loop with the related-work gaps: the system reports each module on its correct scale (Section II-A), and the detector is not only

accurate but lies on the accuracy–size–speed frontier among mainstream detectors on MS-COCO (Section II-B), while the compressed variant—validated for embedded deployment (Section II-C)—retains real-time speed with only a small accuracy cost.

V. CONCLUSION

This study presented an AI-powered wearable assistive system for BVI users, integrating an optimized YOLOv8 object detector, a MobileNetV2/YOLOv8 face-recognition module, sensor-based obstacle and weather detection, and an emergency-contact module, all delivered through real-time voice feedback on an FPGA platform. On MS-COCO 2017, the detector achieved 52.9 AP (49.3 AP compressed at 59 FPS), outperforming heavier baselines, while the face-recognition module reached 99.3% precision under closed-set conditions.

Future work will focus on building and releasing an egocentric BVI detection benchmark [21]; conducting a participatory user study with BVI participants to measure real-world usability and auditory-feedback latency; extending face recognition to robust open-set operation; and characterizing the emergency module's reliability under variable connectivity. These steps are necessary to move the system from a validated prototype toward deployable, human-centered assistive technology.

ACKNOWLEDGMENT

The authors extend their appreciation to Northern Border University, Saudi Arabia, for supporting this work through project number NBU-CRP-2026-3030.

DECLARATIONS

Data Availability. Data are available from the corresponding author on reasonable request.

Conflicts of Interest. The authors declare no conflict of interest.

REFERENCES

- [1] World Health Organization, "Blindness and vision impairment," 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- [2] M. Afif, R. Ayachi, and M. Atri, "Indoor object detection system implementation using multi-graphic processing units," *Cluster Comput.*, vol. 25, pp. 1–15, 2022.
- [3] Z. Fan, W. Mei, W. Liu, M. Chen, and Z. Qiu, "I-DINO: high-quality object detection for indoor scenes," *Electronics*, vol. 13, no. 22, art. 4419, 2024.
- [4] J. Kunhoth, A. Karkar, S. Al-Maadeed, and A. Al-Ali, "Indoor positioning and wayfinding systems: a survey," *Human-centric Comput. Inf. Sci.*, vol. 10, no. 1, art. 18, 2020.
- [5] M. Al-Mahbashi, A. Ahmed, A. Khader, S. Ahmad, M. A. Damos, and A. Abdu, "A robust vision-based framework for traffic sign and light detection in automated driving systems," *Comput. Model. Eng. Sci.*, vol. 146, no. 1, art. 40, 2026.
- [6] G. D'Silva, S. S. More, V. A. Bharadi, and R. Bansode, "Efficient road sign detection and 3D mapping using a hybrid optimized deep learning model for edge device deployment," *Signal Process. Image Commun.*, art. 117675, 2026.
- [7] Z. Bao, W. Han, and Y. Pan, "EMD-YOLOv8: a road pedestrian detection algorithm based on improved YOLOv8," *Digit. Signal Process.*, art. 105940, 2026.
- [8] M. Owais, M. Zubair, D. N. A. Kareem, S. A. Naseer, M. Alam, M. Ahmad, and I. Hussain, "A comprehensive review tracing the evolution of volumetric medical imaging analysis from classic CNNs to emerging AI-agents," *Arch. Comput. Methods Eng.*, pp. 1–51, 2026.
- [9] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.
- [10] M. Hussain, "YOLO-v1 to YOLO-v8: the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection," *Machines*, vol. 11, no. 7, art. 677, 2023.
- [11] T.-Y. Lin, M. Maire, S. Belongie, et al., "Microsoft COCO: common objects in context," in *Proc. European Conf. Computer Vision (ECCV)*, 2014, pp. 740–755.
- [12] O. Daescu, H. Huang, and M. Weinzierl, "Deep learning-based face recognition system with smart glasses," in *Proc. 12th ACM Int. Conf. Pervasive Technologies Related to Assistive Environments (PETRA)*, 2019.
- [13] B. Mandal, S. Chia, L. Li, et al., "A wearable face recognition system on Google Glass for assisting social interactions," in *Proc. Asian Conf. Computer Vision (ACCV)*, 2014, pp. 419–433.
- [14] M. Mukhiddinov and J. Cho, "Smart glass system using deep learning for the blind and visually impaired," *Electronics*, vol. 10, no. 22, art. 2756, 2021.
- [15] E. Ali Hassan and T. B. Tang, "Smart glasses for the visually impaired people," in *Proc. Int. Conf. Computers Helping People with Special Needs (ICCHP)*, 2016, pp. 579–582.
- [16] C. Mai, D. Xie, L. Zeng, et al., "Laser sensing and vision sensing smart blind cane: a review," *Sensors*, vol. 23, no. 2, art. 869, 2023.
- [17] L. A. Guerrero, F. Vasquez, and S. F. Ochoa, "An indoor navigation system for the visually impaired," *Sensors*, vol. 12, no. 6, pp. 8236–8258, 2012.
- [18] R. Tapu, B. Mocanu, and T. Zaharia, "DEEP-SEE: joint object detection, tracking and recognition for visually impaired navigational assistance," *Sensors*, vol. 17, no. 11, art. 2473, 2017.
- [19] Y. Bouteraa, "Design and development of a wearable assistive device integrating a fuzzy decision support system for blind and visually impaired people," *Micromachines*, vol. 12, no. 9, art. 1082, 2021.
- [20] E. Şipoş, C. Ciuciu, and L. Ivanciu, "Sensor-based prototype of a smart assistant for visually impaired people," *Sensors*, vol. 22, no. 11, art. 4271, 2022.
- [21] M. Kuribayashi, S. Kasahara, and T. Hirayama, "Egocentric vision for assistive navigation: a review of datasets and methods," *ACM Comput. Surv.*, vol. 56, no. 4, pp. 1–36, 2024.
- [22] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4690–4699.
- [23] 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, 2017.
- [24] N. Carion, F. Massa, G. Synnaeve, et al., "End-to-end object detection with transformers," in *Proc. European Conf. Computer Vision (ECCV)*, 2020, pp. 213–229.
- [25] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," *arXiv:1804.02767*, 2018.
- [26] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [27] R. Ayachi, Y. Said, and A. Ben Abdelaali, "Optimizing neural networks for efficient FPGA implementation: a survey," *Arch. Comput. Methods Eng.*, vol. 28, pp. 4537–4547, 2021.
- [28] A. Nimmolrat, P. Khuwuthyakorn, P. Wientong, and O. Thinnukool, "Pharmaceutical mobile application for visually-impaired people in Thailand," *BMC Med. Inform. Decis. Mak.*, vol. 21, art. 217, 2021.
- [29] Z. Zheng, P. Wang, W. Liu, et al., "Distance-IoU loss: faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artificial Intelligence (AAAI)*, vol. 34, 2020, pp. 12993–13000.
- [30] X. Li, W. Wang, L. Wu, et al., "Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 21002–21012.
- [31] T. Liang, J. Glossner, L. Wang, S. Shi, and X. Zhang, "Pruning and quantization for deep neural network acceleration: a survey," *Neurocomputing*, vol. 461, pp. 370–403, 2021.
- [32] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. Int. Conf. Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 237–242.