International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-7

AI-Driven Vision Assistance for the Blind Using Real-Time Object Detection

S.Srinivas[1], G. Triveni[2]
[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana –, India.
[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana –, India.

***Abstract* – Vision impairment significantly affects an individual's ability to move independently and safely in unfamiliar environments. Conventional mobility aids, such as white canes and guide dogs, provide valuable assistance but are limited in delivering detailed information about surrounding objects. To address these limitations, VisionAssist is proposed as a smartphone-based intelligent assistance application that employs deep learning techniques for real-time object detection. The system analyzes live camera input to recognize surrounding objects and delivers immediate audio feedback through text-to-speech technology, enabling users to better understand their environment. Designed to operate on smartphones without requiring specialized hardware, the application offers broad accessibility and customizable settings to accommodate individual user preferences. By combining real-time object recognition with voice-guided assistance, the proposed system enhances mobility, improves situational awareness, and promotes greater independence and confidence among visually impaired users.**

Keywords - **Artificial Intelligence (AI), Vision Assistance, Real-Time Object Detection, Computer Vision, Assistive Technology, Blind and Visually Impaired**

I. INTRODUCTION

Visual impairment presents numerous challenges that limit an individual's ability to navigate safely and interact confidently with the surrounding environment. Although traditional mobility aids such as white canes and guide dogs have greatly improved independent movement, they are often unable to identify distant obstacles or provide contextual information about nearby objects. Recent advancements in artificial intelligence, computer vision, and deep learning have created new opportunities to overcome these limitations by delivering intelligent environmental awareness through commonly available mobile devices.

The widespread adoption of smartphones equipped with high-resolution cameras and powerful processing capabilities has enabled the development of portable vision assistance systems. Deep learning-based object detection models are capable of recognizing multiple objects simultaneously from real-time video streams with high accuracy and speed. When integrated with text-to-speech technology, these systems can provide immediate verbal descriptions of surrounding objects, allowing visually impaired individuals to perceive their environment more effectively.

The proposed VisionAssist system utilizes real-time object detection to identify objects captured through the smartphone camera and communicates the detected information using voice guidance. The application is designed to function efficiently on standard smartphones without requiring expensive assistive hardware, making the solution accessible to a wider user population. Furthermore, customizable settings such as speech rate and recognition preferences enable users to personalize the application according to their individual needs.

The primary objective of this work is to develop an intelligent assistive solution that enhances independent navigation, improves user safety, and increases environmental awareness through real-time object recognition and audio feedback. By integrating deep learning, computer vision, and speech synthesis into a unified mobile platform, VisionAssist contributes toward improving accessibility while demonstrating the potential of artificial intelligence in assistive healthcare technologies.

II. LITERATURE SURVEY

Recent developments in artificial intelligence and computer vision have significantly improved assistive technologies for visually impaired individuals. Earlier navigation aids primarily relied on ultrasonic sensors, infrared sensors, and electronic travel devices to detect nearby obstacles. Although these solutions enhanced mobility, they were limited in identifying object types, estimating complex surroundings, and providing detailed contextual information. Consequently, researchers have increasingly focused on vision-based intelligent systems capable of delivering more informative environmental awareness.

The emergence of computer vision techniques introduced image processing methods such as edge detection, contour extraction, and Optical Character Recognition (OCR) for identifying objects and textual information. While these techniques achieved reasonable performance under controlled conditions, they generally suffered from high

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-7

computational requirements and reduced accuracy under varying lighting conditions and complex backgrounds, making them less suitable for real-time assistive applications.

The introduction of deep learning has transformed object detection by enabling highly accurate and efficient recognition models. Convolutional Neural Networks (CNNs) have become the foundation for modern object detection systems, while architectures such as Region-Based CNN (R-CNN), SSD (Single Shot Detector), and the YOLO (You Only Look Once) family have demonstrated remarkable improvements in both detection speed and accuracy. These characteristics make deep learning models particularly suitable for real-time smartphone-based assistive systems.

Several commercial and research-oriented assistive applications have successfully incorporated deep learning into visual assistance. Microsoft Seeing AI utilizes smartphone cameras to identify objects, recognize text, and describe scenes through audio feedback. Similarly, wearable devices such as OrCam MyEye and Aira combine computer vision with voice assistance to improve navigation for visually impaired users. Smart glasses including Envision Glasses further integrate augmented reality and cloud-based processing to provide comprehensive environmental understanding.

Researchers have also investigated optimized deployment of object detection models on mobile platforms. Lightweight implementations such as TensorFlow Lite versions of YOLO enable efficient execution on resource-constrained devices while maintaining real-time performance. These developments demonstrate that integrating CNN-based object detection, optimized mobile inference, and text-to-speech technologies provides an effective solution for improving accessibility and independent navigation among visually impaired individuals. Despite these advancements, challenges related to computational efficiency, distance estimation accuracy, and real-time performance under diverse environmental conditions continue to motivate further research in intelligent vision assistance systems.

III. PROPOSED METHODOLOGY

The proposed VisionAssist framework is developed to provide intelligent real-time assistance for visually impaired individuals through smartphone-based object detection and voice guidance. The system utilizes the smartphone camera to continuously capture live video streams, which are processed using deep learning-based object detection techniques to identify surrounding objects and estimate their relative positions. The detected information is immediately communicated to the user through text-to-speech technology, enabling better awareness of the surrounding environment.

Initially, the camera continuously acquires video frames from the user's surroundings. These frames undergo preprocessing before being supplied to the object detection model. The proposed framework employs the YOLO object detection architecture together with Convolutional Neural Networks (CNNs) to recognize multiple objects efficiently while maintaining real-time performance. Pre-trained model weights and the COCO object dataset are utilized to identify a wide variety of everyday objects with high detection accuracy.

Following object detection, the system estimates the distance between the user and each detected object by calculating the object's pixel dimensions within the captured frame. The estimated distance enables the framework to determine whether an object is positioned within a potentially hazardous range. Whenever an object is detected at a close distance, the system immediately generates warning notifications to alert the user.

To provide meaningful interaction, the detected object names and corresponding alerts are converted into audible messages using the pyttsx3 text-to-speech engine. This allows users to receive immediate spoken descriptions without requiring visual interaction with the smartphone display. Frame smoothing techniques are also incorporated to improve the stability of real-time video processing and reduce fluctuations during continuous object detection.

The complete framework integrates real-time video acquisition, YOLO-based object detection, CNN-based feature extraction, distance estimation, and speech synthesis into a unified mobile application. By combining these components, the proposed system provides an intelligent assistive solution that enhances navigation safety, improves environmental awareness, and supports independent mobility for visually impaired individuals using readily available smartphone devices.

IV. SYSTEM ARCHITECTURE

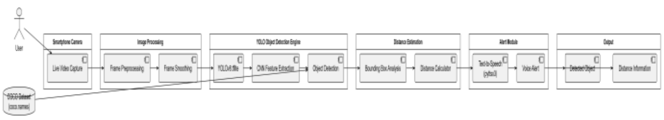

Figure 1. System Architecture of VisionAssist

The VisionAssist system is designed using a modular architecture that integrates computer vision, deep learning, distance estimation, and speech synthesis to provide real-time assistance for visually impaired individuals. The architecture enables continuous monitoring of the user's surroundings through a smartphone camera and transforms visual information into meaningful audio guidance. Each module performs a specific task, ensuring efficient data processing and accurate object recognition while maintaining real-time performance on mobile devices.

The architecture begins with the image acquisition module, where the smartphone camera continuously captures live video frames from the surrounding environment. These frames represent the visual input that is processed throughout the system. Since environmental conditions may vary due to lighting changes, motion, and background complexity, the captured frames undergo preprocessing to improve image quality and prepare them for object detection.

The preprocessing module performs operations such as frame resizing, normalization, and blob generation using the OpenCV library. These preprocessing steps convert the raw camera frames into a standardized format compatible with the YOLO object detection network. Frame smoothing techniques are also applied to minimize fluctuations between consecutive frames, thereby improving the stability of object detection during continuous video streaming.

Once preprocessing is completed, the processed frames are forwarded to the YOLO-based object detection module. The proposed system utilizes the YOLO architecture together with Convolutional Neural Networks (CNNs) to perform real-time object localization and classification. The model employs pre-trained weights along with the COCO dataset object classes to identify multiple everyday objects present within the camera view. Unlike conventional multi-stage detection algorithms, YOLO analyzes the entire image in a single forward pass, enabling high-speed detection while maintaining satisfactory accuracy for assistive applications.

After detecting objects, the object localization module determines the position of each detected object by generating bounding boxes around them. The system also assigns confidence scores and class labels using the information available in the coco.names file. These labels enable the application to recognize objects such as people, bottles, mobile phones, sports balls, furniture, and many other commonly encountered items that are important for independent navigation.

The next component is the distance estimation module, which calculates the approximate distance between the smartphone camera and the detected object. This estimation is performed using the object's pixel height together with predefined focal length and known object height parameters. The calculated distance provides valuable contextual information that helps determine whether an object is located within a potentially hazardous range. Whenever an object is detected below the predefined safety threshold, the system immediately generates warning notifications to alert the user.

The detected object names and corresponding distance information are then transmitted to the text-to-speech module. This module employs the pyttsx3 speech synthesis engine to convert textual information into clear and understandable voice instructions. Rather than requiring users to observe the smartphone display, the application communicates all detected objects through spoken messages, thereby allowing visually impaired individuals to remain fully aware of their surroundings while navigating independently.

The alert generation module continuously monitors object proximity and determines whether emergency notifications should be issued. If a detected object is extremely close to the user, visual warning indicators are displayed on the application interface while simultaneous voice alerts are generated. This dual notification mechanism improves user safety by providing immediate awareness of nearby obstacles.

Finally, the generated recommendations and detected object information are presented through the user interaction module. The application continuously updates the user regarding surrounding objects while maintaining smooth real-time processing. The modular architecture enables efficient integration of OpenCV, YOLO, CNN, TensorFlow Lite models, and speech synthesis technologies into a unified smartphone application capable of assisting visually impaired individuals during daily navigation.

Overall, the proposed architecture combines intelligent object detection, accurate localization, distance estimation, and real-time voice assistance into a comprehensive mobile-based vision assistance framework. The integration of these components enhances accessibility, improves situational awareness, promotes safer navigation, and enables visually impaired users to move with greater confidence and independence in both indoor and outdoor environments.

V. FEATURE ENHANCEMENT

The proposed VisionAssist framework incorporates several advanced features that significantly improve the effectiveness of real-time object detection and environmental awareness for visually impaired individuals. By integrating deep learning algorithms with smartphone-based computer vision, the system provides an intelligent assistive solution capable of recognizing surrounding objects, estimating their relative distances, and delivering immediate voice guidance. These enhancements collectively improve navigation safety, accessibility, and user independence while maintaining efficient execution on portable mobile devices.

One of the primary enhancements of the proposed framework is the implementation of real-time object detection using the YOLO architecture. Unlike conventional object detection methods that require multiple processing stages, YOLO performs localization and classification simultaneously within a single neural network. This significantly reduces computational latency

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-7

and enables the system to detect multiple objects continuously from live camera streams. The rapid processing capability allows users to receive immediate information about nearby obstacles without noticeable delays, making the application highly suitable for real-world navigation scenarios.

The framework further improves detection accuracy by employing Convolutional Neural Networks (CNNs) for feature extraction and object recognition. CNNs automatically learn hierarchical visual features from input images, allowing the system to distinguish between different object categories under varying environmental conditions. The combination of CNN-based feature learning with YOLO's efficient detection strategy enables reliable identification of everyday objects commonly encountered during indoor and outdoor navigation.

Another significant enhancement is the utilization of the COCO object dataset together with the coco.names configuration file. The COCO dataset contains numerous object categories that enable the system to recognize a diverse range of real-world objects such as people, bottles, chairs, mobile phones, vehicles, and traffic-related items. This extensive object vocabulary enhances the practical usability of the application by providing meaningful environmental information across different surroundings.

To improve user safety, the framework incorporates an intelligent distance estimation mechanism. The system estimates the approximate distance between the user and detected objects by analyzing the dimensions of the object's bounding box together with predefined camera parameters. This additional information enables the application to determine whether detected objects are located within a potentially dangerous range. When nearby obstacles are identified, immediate warning notifications are generated to alert the user before possible collisions occur.

The integration of the pyttsx3 text-to-speech engine represents another important enhancement of the proposed system. Instead of requiring users to observe visual outputs, all detected object information is converted into clear spoken instructions. This voice-based interaction enables visually impaired individuals to understand their surroundings naturally while keeping both hands free during navigation. Furthermore, customizable speech settings allow users to modify voice speed and other preferences according to their personal requirements, thereby improving overall usability and user comfort.

The proposed framework also incorporates frame smoothing techniques to enhance video stability during continuous object detection. Consecutive video frames are blended to reduce sudden fluctuations caused by camera movement or environmental disturbances. As a result, object detection becomes more stable, bounding boxes remain consistent, and voice notifications become less repetitive, improving the overall user experience during prolonged system usage.

Another enhancement involves the implementation of Non-Maximum Suppression (NMS) for eliminating duplicate object detections. During object recognition, multiple overlapping bounding boxes may be generated for the same object. The NMS algorithm removes redundant detections while preserving only the most confident prediction, resulting in cleaner visualization, improved detection accuracy, and reduced computational overhead.

The smartphone-based implementation further enhances the practicality of the proposed framework. Since the application operates on standard Android smartphones without requiring dedicated hardware, visually impaired users can access intelligent assistance using devices that are already widely available. This significantly reduces deployment costs while increasing the accessibility of assistive technology for a larger population.

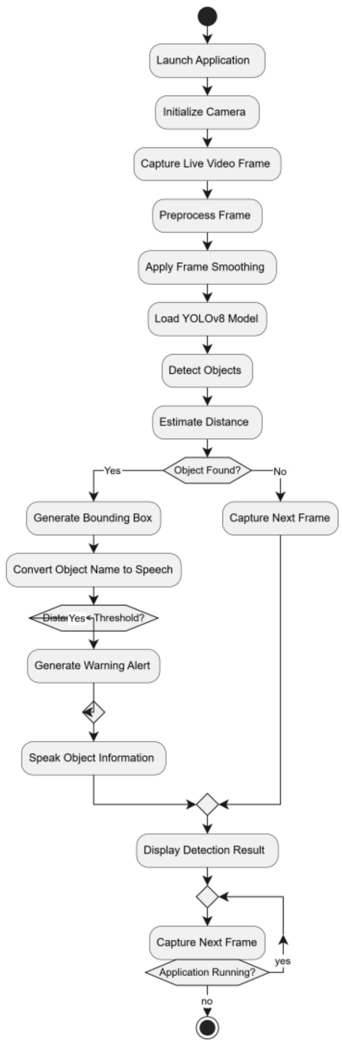

Fig. 2. Activity Diagram of the VisionAssist Real-Time Object Detection Process.

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-7

VI. RESULTS AND DISCUSSION

The performance of the proposed VisionAssist system was evaluated by analyzing its ability to perform real-time object detection, distance estimation, and voice-based feedback for visually impaired users. The implementation combines YOLO-based object detection, Convolutional Neural Networks (CNNs), OpenCV, TensorFlow Lite, and text-to-speech technology to recognize surrounding objects and provide immediate audio assistance. The experimental observations demonstrate that the proposed framework successfully identifies multiple objects in real time while maintaining smooth execution on smartphone devices.

The object detection module accurately recognizes various everyday objects captured through the smartphone camera by utilizing the pretrained YOLO model together with the COCO dataset. Each detected object is enclosed within a bounding box and assigned an appropriate class label along with its confidence score. This enables users to obtain detailed information regarding the objects present in their surroundings. The capability of detecting multiple objects simultaneously significantly enhances environmental awareness and allows users to better understand complex scenes during navigation.

An important component of the proposed system is the distance estimation mechanism. The framework estimates the approximate distance between the camera and detected objects by analyzing the height of the detected bounding boxes together with predefined camera calibration parameters. This additional information enables the application to determine whether an object is located within a potentially hazardous range. Whenever an object is detected at a close distance, the system immediately generates warning notifications through both visual indicators and speech alerts, thereby improving user safety during movement.

To ensure continuous real-time performance, frame smoothing techniques are incorporated into the processing pipeline. Frame smoothing minimizes abrupt changes between consecutive video frames caused by camera movement or environmental disturbances. As a result, the detected object locations remain more stable, reducing unnecessary fluctuations in bounding boxes and preventing repetitive voice announcements. This contributes to a smoother and more reliable user experience during prolonged application usage.

The implementation also employs Non-Maximum Suppression (NMS) to eliminate redundant object detections. Since multiple overlapping bounding boxes may be generated for the same object, the NMS algorithm retains only the detection with the highest confidence score while discarding duplicate predictions. This optimization improves detection accuracy, reduces computational redundancy, and provides cleaner visual outputs for the object detection module.

The integration of the pyttsx3 speech synthesis engine further enhances the practicality of the proposed framework. Instead of relying solely on visual displays, detected object names and warning messages are immediately converted into audible instructions. The real-time voice feedback enables visually impaired users to receive continuous environmental information without interacting with the smartphone screen, thereby promoting independent navigation and improving overall accessibility.

The experimental observations obtained from the implementation demonstrate that the proposed framework successfully detects various commonly encountered objects while estimating their relative distances. Table 1 presents representative detection results obtained during experimentation.

Table 1. Height and Distance of Detected Objects

Object	Distance	Height
Person	0.62 m	328 Pixels
Bottle	2.90 m	69 Pixels
Sports Ball	1.13 m	177 Pixels
Cell Phone	0.52 m	283 Pixels

The experimental results indicate that the system effectively identifies multiple object categories while simultaneously estimating their approximate distances from the user. Objects positioned closer to the camera exhibit larger pixel heights, whereas distant objects occupy relatively smaller image regions. These measurements assist the framework in generating timely safety alerts whenever nearby obstacles are detected.

Overall, the proposed VisionAssist framework demonstrates effective real-time performance by integrating deep learning-based object detection, distance estimation, frame stabilization, and speech synthesis into a unified mobile application. The combination of these technologies provides reliable environmental awareness, improves obstacle detection, enhances user safety, and supports independent mobility for visually impaired individuals using widely available smartphone devices. The obtained results demonstrate the practical applicability of the proposed system for intelligent assistive navigation in real-world environments.

VII. CONCLUSION

The proposed VisionAssist system demonstrates the potential of artificial intelligence and deep learning in improving accessibility for visually impaired individuals through intelligent real-time object detection and voice-guided assistance. By integrating computer vision techniques with smartphone technology, the framework provides an efficient solution for recognizing surrounding objects, estimating their relative distances, and delivering immediate audio feedback that enhances environmental

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-7

awareness during navigation. The proposed approach minimizes the limitations of conventional mobility aids by supplying users with contextual information about nearby objects rather than simply detecting physical obstacles.

The implementation combines YOLO-based object detection, Convolutional Neural Networks (CNNs), OpenCV, TensorFlow Lite, and the pyttsx3 text-to-speech engine to achieve reliable real-time performance on smartphone devices. The integration of these technologies enables accurate object recognition, efficient distance estimation, stable video processing, and timely voice notifications without requiring specialized assistive hardware. Consequently, the system provides a practical and cost-effective solution that can be deployed on commonly available smartphones, making intelligent visual assistance more accessible to a wider population.

The experimental observations demonstrate that the proposed framework successfully identifies multiple everyday objects while providing continuous spoken guidance to users. The incorporation of distance estimation and intelligent alert generation further improves user safety by notifying individuals whenever nearby obstacles are detected within a potentially hazardous range. Frame smoothing and Non-Maximum Suppression (NMS) additionally enhance detection stability and overall system performance during continuous operation.

Although the current implementation effectively performs real-time object detection and navigation assistance, several opportunities remain for future enhancement. Additional capabilities such as currency recognition, scene description, optical character recognition (OCR), face recognition, traffic sign detection, GPS-based outdoor navigation, and emergency contact notification can further improve the usefulness of the application. Integrating multilingual speech support and more lightweight deep learning models may also enhance usability across diverse user communities and resource-constrained mobile devices.

Overall, the proposed VisionAssist framework provides an intelligent, reliable, and user-friendly assistive solution that improves independent mobility, situational awareness, and navigation safety for visually impaired individuals. The combination of deep learning, computer vision, and speech synthesis demonstrates the growing impact of artificial intelligence in assistive healthcare technologies and establishes a strong foundation for future research in intelligent mobile accessibility systems.

REFERENCES

1. T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," Multimedia Tools and Applications, vol. 82, pp. 9243–9275, 2023.
2. M. Pulipalupula, S. Patlola, M. Nayaki, M. Yadlapati, J. Das, and B. Reddy, "Object Detection using You Only Look Once (YOLO) Algorithm in Convolution Neural Network (CNN)," in Proc. Int. Conf. Convergence in Technology (I2CT), 2023, pp. 1–4.
3. K. Vaishnavi, G. Reddy, T. Reddy, N. Iyengar, and S. Subhani, "Real-Time Object Detection Using Deep Learning," Journal of Advances in Mathematics and Computer Science, vol. 38, pp. 24–32, 2023.
4. D. Li, "Research advances in object detection based on deep learning," Applied and Computational Engineering, vol. 5, pp. 603–608, 2023.
5. P. Druzhkov and V. Kustikova, "A survey of deep learning methods and software tools for image classification and object detection," Pattern Recognition and Image Analysis, vol. 26, no. 1, pp. 9–15, 2016.
6. K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, and Y. LeCun, "Learning convolutional feature hierarchies for visual recognition," in Advances in Neural Information Processing Systems (NeurIPS), 2010.
7. C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in Advances in Neural Information Processing Systems (NeurIPS), 2013.
8. P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 8, pp. 1532–1545, 2014.
9. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
10. J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517–6525.
11. J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.
12. G. Jocher et al., "YOLOv5," GitHub Repository, 2020.
13. G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics (YOLOv8)," GitHub Repository, 2023.
14. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.
15. W. Liu et al., "SSD: Single Shot MultiBox Detector," in Proc. European Conf. Computer Vision (ECCV), 2016, pp. 21–37.
16. R. Girshick, "Fast R-CNN," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2015, pp. 1440–1448.
17. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.

18. T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context," in Proc. European Conf. Computer Vision (ECCV), 2014, pp. 740–755.
19. Microsoft Corporation, "Seeing AI: Talking Camera App for the Blind," Microsoft Research, 2023.
20. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580–587.