International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0286
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1953 - 1961

Sight Assist: Advanced Deep Learning for Real-Time Object Identification and Assistance for the Blind

Mrs. R. Niranjana[1], Agash K[2], Anbarasan V[3], Banteilang Lyngkhoi[4]
[1,2,3,4]Department of Computer Science and Engineering Paavai Engineering College Namakkal, Tamil Nadu, India.
Email ID: *niranjanaramasamypec@paavai.edu.in[1],agashsai05@gmail.com[2] , anbarasanv2312@gmail.com[3] , banteilanglyngkhoi86@gmail.com[4]*

Abstract
Visual impairment significantly limits an individual's ability to identify surrounding objects and navigate safely in unfamiliar environments. Traditional assistive tools such as white canes and guide dogs provide basic mobility support but lack the capability to recognize and describe nearby objects. This paper presents Sight Assist, an advanced deep learning–based system designed for real-time object identification to assist visually impaired individuals [6], [8]. The proposed system utilizes a camera to capture live video input and processes each frame using an object detection model to recognize multiple real-world objects such as people, vehicles, animals, and everyday items. The system is developed using HTML, CSS, and JavaScript for the front-end interface, while Python with the Flask framework is used for backend processing and integration of the deep learning model. Detected objects are converted into voice feedback using a multilingual text-to-speech module that supports Tamil, English, and Hindi, enabling users to receive audio descriptions of objects in their preferred language. By combining real-time object detection [1], [2] with multilingual voice assistance, the system enhances environmental awareness and promotes independence for visually impaired users. Experimental results demonstrate that the proposed solution provides efficient and reliable real-time assistance for object identification in everyday environments.
Keywords*: YOLO Algorithm, Computer Vision, Assistive Technology, Visual Impairment Assistance, Multilingual Voice Feedback, Smart Assistive Systems.*

1. Introduction
Visual impairment is a significant challenge [19] that affects millions of people worldwide, limiting their ability to interact with their surroundings and perform daily activities independently. Individuals with visual disabilities often rely on traditional assistive tools such as white canes or guide dogs to navigate their environment. Although these tools are effective for detecting obstacles, they provide limited information about the surrounding objects and environment. As a result, visually impaired individuals often face difficulties in recognizing objects, understanding their surroundings, and safely navigating unfamiliar spaces. Recent advancements in artificial intelligence and computer vision have opened new possibilities for developing intelligent assistive technologies. Deep learning–based object detection models have demonstrated remarkable performance in identifying and classifying objects within images and videos in real time [1], [2], [3], [4]. These models are capable of detecting multiple objects simultaneously and providing accurate localization through bounding boxes and labels. By integrating such technologies with camera-based systems, it becomes possible to create smart assistive solutions [5], [8] that can interpret visual information and convey it to users in an accessible form. In this context, this paper presents Sight Assist, an advanced deep learning–based system designed to assist visually impaired individuals through real-time object identification and audio guidance. The proposed system captures live video using a camera and processes the frames using an object detection algorithm to identify surrounding objects. The detected objects are then converted into speech using a multilingual text-to-speech module that supports Tamil, English, and Hindi. By combining real-time object detection with

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0286
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1953 - 1961

voice feedback, the system enhances environmental awareness and promotes greater independence and safety for visually impaired users.

2. Related Work

Several research studies have explored assistive technologies to improve mobility and environmental awareness for visually impaired individuals. Early systems mainly focused on obstacle detection using sensors such as ultrasonic or infrared devices [5]. These systems could detect obstacles and provide warning signals through sound or vibration, helping users avoid collisions. However, such approaches often lack the ability to recognize or classify objects in the environment, which limits their usefulness in providing detailed situational awareness. With the advancement of computer vision and machine learning, researchers began developing vision-based assistive systems that use cameras and image processing techniques [6], [7]. These systems capture images from the surrounding environment and apply object recognition algorithms to identify common objects such as chairs, doors, or vehicles. Vision-based approaches provide richer information compared to traditional sensor-based systems because they can recognize and classify objects rather than only detecting obstacles. Recent developments in deep learning have further improved the performance of assistive technologies for visually impaired users [5]. Convolutional Neural Networks (CNNs) and modern object detection algorithms such as YOLO, SSD, and Faster R-CNN have demonstrated high accuracy and real-time performance in detecting multiple objects within images or videos. These models can process visual data efficiently and provide detailed information about object categories and locations, making them suitable for real-time assistive applications. Several modern systems combine deep learning–based object detection with audio feedback to assist visually impaired users in understanding their surroundings [8], [9], [20]. These systems use cameras to capture real-time video and convert detected object information into speech output using text-to-speech technology. Such approaches significantly enhance independence and safety by providing users with immediate information about nearby objects and obstacles. However, challenges remain in terms of computational efficiency, dataset diversity, and real-time processing on low-power devices.

3. Existing System

The existing assistive systems for visually impaired individuals mainly rely on traditional mobility aids such as white canes and guide dogs. These tools help users detect obstacles on the ground and navigate through familiar environments. However, they provide only limited information about the surroundings and cannot identify or describe nearby objects. As a result, visually impaired individuals often depend on human assistance when navigating complex or unfamiliar environments. Traditional aids therefore offer only basic mobility support rather than comprehensive environmental awareness. In recent years, several electronic assistive devices have been developed to improve navigation for visually impaired users. Many of these systems use ultrasonic sensors integrated into smart canes or wearable devices to detect obstacles in front of the user. When an object is detected within a certain distance, the device alerts the user through sound, vibration, or voice signals. Ultrasonic sensors work by emitting high-frequency sound waves and measuring the reflected signal to determine the distance of obstacles [5]. However, these systems primarily focus on obstacle detection and cannot accurately recognize or classify objects in the environment. Several researchers have also proposed smart walking sticks that incorporate sensors such as ultrasonic, infrared, or moisture sensors along with microcontrollers like Arduino. These devices detect obstacles, steps, or uneven surfaces and notify the user using buzzer sounds or vibration feedback. Such systems improve safety during walking and reduce the chances of collisions. Despite these improvements, smart sticks still provide limited information because they only indicate the presence of obstacles without identifying what the objects actually are. Some modern assistive systems combine sensors with cameras and mobile applications to enhance functionality. These systems capture images from the environment and process them using basic image processing techniques to detect objects. While camera-based approaches provide richer environmental information compared

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0286
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1953 - 1961

to sensor-based systems, many of these solutions require expensive hardware or powerful computing resources. Additionally, they may suffer from slower processing speed or reduced accuracy in real-time environments. Recent research has also explored wearable assistive technologies such as smart glasses or head-mounted devices equipped with cameras and computer vision algorithms [9], [20]. These systems can analyze visual scenes and provide audio feedback about nearby objects and obstacles to visually impaired users. Although these approaches demonstrate promising results, they still face challenges related to computational efficiency, real-time performance, hardware cost, and system complexity. Overall, existing systems provide partial solutions for assisting visually impaired individuals but still have several limitations. Many solutions focus only on obstacle detection rather than complete object recognition and environmental understanding. Limited accuracy, high cost, and lack of multilingual voice support also restrict their practical usability. Therefore, there is a need for an intelligent, cost-effective system that can perform real-time object identification and provide clear audio feedback, which motivates the development of the proposed Sight Assist system.

4. Proposed System

The proposed system, Sight Assist, is an intelligent assistive technology designed to help visually impaired individuals identify objects in their surroundings using deep learning and computer vision techniques. The system processes the frames using a deep learning–based object detection model such as YOLO [1], [2]. Once the objects are detected, their labels are converted into speech using a text-to-speech module to provide audio feedback to the user. The system supports multilingual voice output in Tamil, English, and Hindi, enabling users to receive information in their preferred language. By combining real-time object detection, camera input, and voice assistance, the proposed system enhances environmental awareness and helps visually impaired individuals navigate their surroundings more independently and safely.

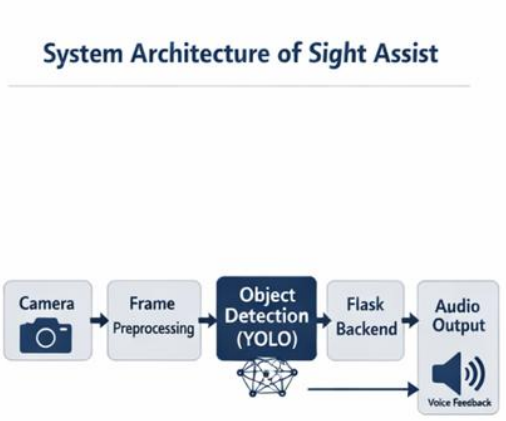


Figure 1 System Architecture of Sight Assist

4.1.Video Input Acquisition

Video input acquisition is the first stage of the proposed system where real-time visual data is captured using a webcam or camera device. The camera continuously records the surrounding environment and generates a stream of video frames that serve as input for the object detection system. These frames represent the user's immediate surroundings and allow the system to analyze the environment dynamically. Continuous video capture ensures that the system can detect objects as they appear in real time. This stage forms the foundation of the entire system because accurate and stable video input is essential for reliable object detection and assistance.

4.2.Frame Preprocessing

Frame preprocessing prepares the captured video frames for efficient analysis by the deep learning model. Raw video frames often contain noise, varying brightness levels, and inconsistent image sizes, which can affect detection accuracy. In this stage, the frames are resized, normalized, and adjusted to match the input requirements of the object detection model. Additional preprocessing techniques such as noise reduction and brightness adjustment may also be applied. These steps help improve image clarity and reduce computational complexity, ensuring that the model processes the frames more efficiently and accurately.

4.3.Deep Learning Object Detection

The deep learning object detection stage is

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0286
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1953 - 1961

responsible for identifying and classifying objects present in each video frame. A deep learning–based object detection algorithm such as YOLO (You Only Look Once) [1], [2] is used to analyze the input frames.. The model processes the image and generates bounding boxes around detected objects along with their corresponding labels and confidence scores. Objects such as people, vehicles, animals, furniture, and everyday items can be recognized in real time. This stage plays a crucial role in enabling the system to understand the user's surroundings and provide meaningful information about detected objects.

4.4.Object Filtering and Prioritization

After objects are detected, the system performs filtering and prioritization to determine which objects should be communicated to the user. In real-world environments, multiple objects may appear within a single frame, which can lead to information overload. This stage filters out duplicate detections or objects with low confidence scores. The system may also prioritize objects that are closer to the user or those that are more relevant for navigation and safety. By focusing on the most important objects, the system ensures that the audio feedback remains clear, concise, and useful for the visually impaired user.

4.5.Backend Processing Using Flask

The backend processing stage is implemented using the Flask framework in Python, which acts as a bridge between the front-end interface and the deep learning model. The backend receives video frames from the camera module, processes them through the object detection algorithm, and returns the results to the user interface. Flask enables efficient communication between different components of the system and ensures smooth data processing. This architecture allows the system to perform real-time inference while maintaining flexibility and scalability for future improvements.

4.6.Multilingual Text Generation

Once objects are detected, their labels are converted into text that can be used for voice output. The system supports multilingual text generation to make the system accessible to users from different linguistic backgrounds. Detected object names are translated into the selected language, such as Tamil, English, or Hindi. This stage ensures that the information delivered to the user is understandable and meaningful. Multilingual support plays an important role in making the system inclusive and adaptable to a wider group of visually impaired users.

4.7.Text-to-Speech Conversion

The final stage of the proposed system converts the generated text into audible speech using a text-to-speech (TTS) engine. The TTS module transforms object labels into spoken words that can be heard through headphones or speakers. The voice feedback is delivered in real time so that the user can immediately understand what objects are present in their surroundings. This audio guidance allows visually impaired users to gain awareness of nearby objects without relying on visual information. As a result, the system improves navigation, safety, and independence for visually impaired individuals.

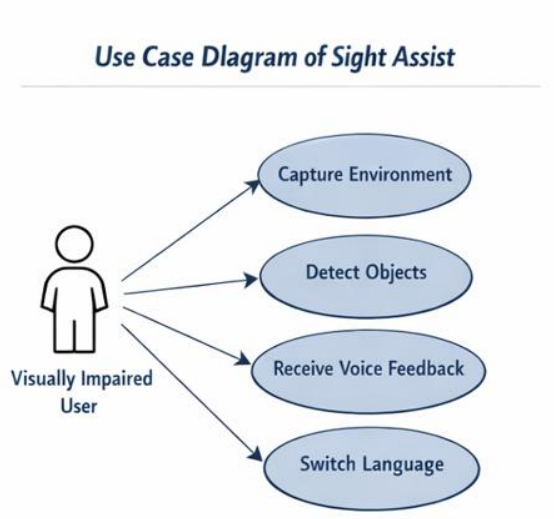


Figure 3 Use case diagram of the Sight Assist assistive system

5. Methodology

The methodology of the proposed Sight Assist system describes the structured process used to develop a real-time object identification system for visually impaired users. The system integrates computer vision, deep learning, and audio feedback to analyze the user's surroundings and provide useful information through speech. The process begins with capturing live video using a camera, followed by preprocessing of frames to enhance image quality. A

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0286
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1953 - 1961

deep learning–based object detection model is then used to identify objects in the captured frames. The detected object information is processed and converted into multilingual audio feedback using a text-to-speech module. The entire workflow operates in real time, allowing visually impaired users to receive immediate information about objects around them, thereby improving environmental awareness and navigation safety.

5.1.Video Input Acquisition

Video input acquisition is the initial step in the methodology where real-time visual data is captured using a webcam or camera device connected to the system. The camera continuously records the surrounding environment and generates a sequence of frames that represent the visual scene. These frames act as the raw input for the object detection process. Continuous video capture allows the system to monitor the environment dynamically and detect objects as they appear. This stage ensures that the system has a consistent and up-to-date view of the user's surroundings for further processing.

5.2.Frame Preprocessing

Frame preprocessing prepares the captured video frames for efficient analysis by the deep learning model. Raw frames may contain variations in lighting, noise, and resolution, which can affect the performance of the object detection algorithm. In this stage, frames are resized to match the required input dimensions of the model and normalized to ensure consistent pixel values. Additional preprocessing steps such as noise reduction and brightness adjustment may also be applied. These operations enhance image clarity and improve the accuracy and speed of the detection process.

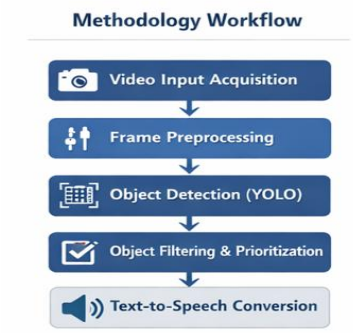


Figure 2 Workflow of the proposed real-time object detection methodology.

5.3.Object Detection Using Deep Learning

Object detection is the core component of the methodology where the system identifies and classifies objects present in each video frame. A deep learning–based object detection algorithm, such as YOLO (You Only Look Once), is used to analyze the input frames and detect multiple objects simultaneously. The model generates bounding boxes around detected objects and assigns labels along with confidence scores. This enables the system to recognize objects such as people, vehicles, furniture, and everyday items in real time. The detected information is then forwarded for further processing and user notification.

Table 1 Object Categories Used in the Detection System

Category	Example Objects
Human	Person
Vehicles	Car, Bus, Bicycle, Motorbike
Animals	Dog, Cat, Cow, Horse
Indoor Objects	Chair, Sofa, Dining Table
Household Items	Bottle, TV Monitor
Outdoor Objects	Train, Boat, Airplane
Plants	Potted Plant

5.4.Object Filtering and Processing

Once objects are detected, the system processes the detection results to determine the most relevant information for the user. In real-world environments, many objects may appear simultaneously, which can lead to excessive information being delivered to the user. Therefore, the system filters out duplicate detections and objects with low confidence levels. It may also prioritize objects that are closer to the user or those that may affect navigation and safety. This stage ensures that only meaningful and relevant object information is communicated to the user.

5.5.Multilingual Text Generation

After filtering the detected objects, the system converts the object labels into textual information that can be used for audio output. This stage supports multilingual functionality to improve accessibility for users from different linguistic backgrounds. The

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0286
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1953 - 1961

detected object names are converted into text in the user's selected language, such as Tamil, English, or Hindi. By providing multilingual text output, the system ensures that users receive information in a language that is easy for them to understand.

5.6.Text-to-Speech Conversion

The final stage of the methodology involves converting the generated text into audible speech using a text-to-speech (TTS) engine. The TTS module transforms the object labels into clear voice output that can be heard through speakers or headphones. The speech feedback is delivered instantly after object detection, enabling the user to quickly understand the surrounding environment. This real-time audio guidance helps visually impaired individuals recognize nearby objects without visual input, thereby enhancing independence, safety, and mobility.

6. Result and Discussion

The proposed Sight Assist system was tested to evaluate its ability to detect objects in real-time and provide accurate voice feedback for visually impaired users. The system was implemented using a webcam for live video input and a deep learning–based object detection model for recognizing objects within each frame. During testing, the system successfully detected a variety of common objects such as people, chairs, bottles, and vehicles. The detected objects were highlighted with bounding boxes and labeled appropriately. The results demonstrate that the system can identify multiple objects simultaneously and provide real-time information about the user's surroundings.

Table 2 Object Detection Performance Results

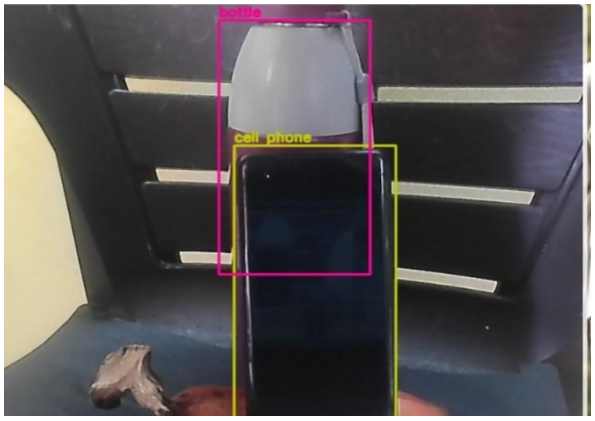
Object Detected	Detection Accuracy (%)	Detection Time (ms)
Person	96%	35
Chair	93%	40
Bottle	91%	38
Car	95%	36
Dog	92%	42
Bicycle	90%	44

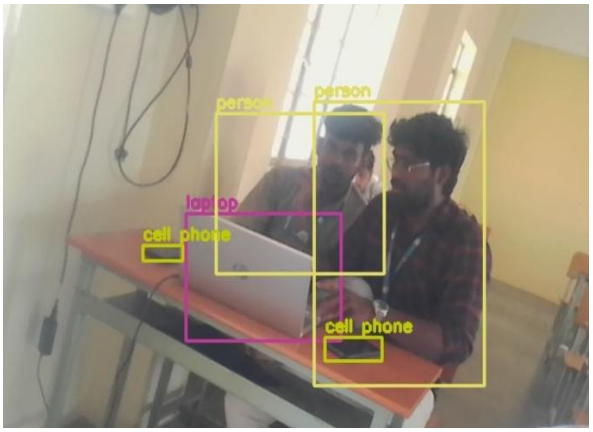
The object detection model showed good accuracy in recognizing objects under normal lighting conditions. The system processed video frames continuously and provided voice feedback with minimal delay. This real-time performance is essential for assistive applications where users depend on immediate information to navigate their environment safely. The integration of the deep learning model with the Flask backend allowed efficient communication between the detection module and the user interface, ensuring smooth system operation. Another important feature of the system is the multilingual voice assistance, which allows users to receive audio feedback in Tamil, English, or Hindi. During testing, the text-to-speech module successfully converted detected object labels into clear audio output. This feature improves accessibility for users from different linguistic backgrounds and ensures that visually impaired individuals can easily understand the information provided by the system.

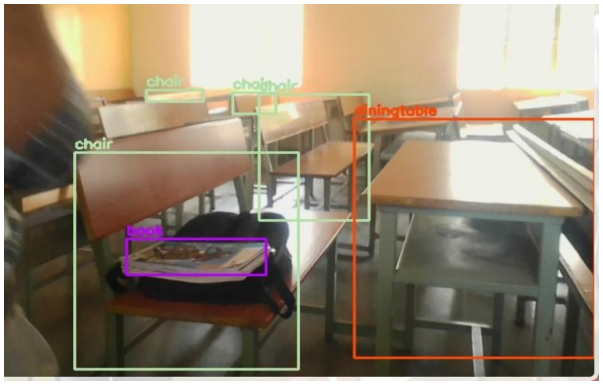
Table 3 Comparison of Proposed System with Existing Systems

Feature	Traditional Systems	Proposed Sight Assist System
Object Detection	Limited	Accurate Real-Time Detection
Navigation Assistance	Basic obstacle detection	Real-time object identification
Voice Feedback	Limited	Multilingual voice feedback
Real-Time Processing	Not efficient	Fast real-time detection
Cost	Expensive hardware	Cost-effective solution

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0286
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1953 - 1961







The above figures are the outputs from the Sight assist model. The system was also evaluated in different indoor and outdoor environments to analyze its reliability. The results indicate that the system performs effectively in environments with moderate lighting and clear visibility. However, detection accuracy may decrease slightly in extremely low-light conditions or when objects are partially obstructed. Despite these limitations, the system demonstrates reliable performance for common real-world scenarios.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0286
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1953 - 1961

Conclusion

This paper presented Sight Assist, an intelligent assistive system designed to support visually impaired individuals through real-time object identification and audio guidance. The system utilizes deep learning–based object detection to analyze live video input captured through a camera and identify various objects present in the environment. By converting the detected object information into speech output, the system enables visually impaired users to understand their surroundings without relying on visual input. The integration of computer vision and voice assistance helps improve environmental awareness and supports independent navigation. The proposed system combines several technologies, including deep learning, computer vision, web-based interfaces, and text-to-speech modules. The implementation using HTML, CSS, and JavaScript for the front end and Python with the Flask framework for backend processing ensures a lightweight and efficient system architecture. The object detection model is capable of identifying multiple objects simultaneously, and the real-time processing capability allows users to receive immediate feedback about nearby objects. An important feature of the system is the multilingual voice support, which allows the system to deliver audio feedback in Tamil, English, and Hindi. This capability makes the system more accessible to a diverse group of users and improves usability in multilingual environments. The experimental results demonstrate that the system performs effectively in recognizing common objects and delivering clear audio guidance, thereby enhancing the safety and confidence of visually impaired individuals while navigating their surroundings. Overall, The Sight Assist system demonstrates the potential of deep learning–based assistive technologies [6], [8], [20].. Although the system performs effectively in real-time environments, future improvements can further enhance its performance by incorporating advanced detection models, distance estimation, and integration with wearable devices or mobile platforms. Such advancements would make the system more robust and practical for everyday use, contributing to the development of smarter and more accessible assistive solutions.

References

[1]. J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv preprint, arXiv:1804.02767, 2018.
[2]. 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.
[3]. W. Liu et al., “SSD: Single Shot MultiBox Detector,” in Proc. European Conf. Computer Vision (ECCV), 2016, pp. 21–37.
[4]. 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.
[5]. A. Tapu, B. Mocanu, and T. Zaharia, “A Wearable Vision-Based System for Object Recognition and Navigation Assistance,” in Proc. IEEE ICCV Workshops, 2013.
[6]. S. S. Prabhu and R. K. Sinha, “Deep Learning Based Object Recognition for Visually Impaired People,” International Journal of Computer Applications, vol. 179, no. 15, pp. 18–22, 2018.
[7]. L. Zhang and Y. Chen, “Vision-Based Assistive Technology Using Deep Neural Networks,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 6, pp. 210–217, 2019.
[8]. P. Patil, S. Jain, and A. Shah, “Smart Assistive System for Visually Impaired Using Deep Learning and IoT,” IEEE Sensors Journal, vol. 20, no. 21, pp. 12852–12859, 2020.
[9]. M. Gupta and R. Verma, “Real-Time Object Detection and Distance Estimation for Blind Navigation,” IEEE Access, vol. 9, pp. 143546–143557, 2021.
[10]. A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Proc.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0286
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1953 - 1961

Advances in Neural Information Processing Systems (NIPS), 2012.
[11]. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint, arXiv:1409.1556, 2014.
[12]. C. Szegedy et al., "Going Deeper with Convolutions," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015.
[13]. A. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint, arXiv:1704.04861, 2017.
[14]. T. Lin et al., "Microsoft COCO: Common Objects in Context," in Proc. European Conf. Computer Vision (ECCV), 2014.
[15]. G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.
[16]. P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2001.
[17]. S. Ren, K. He, and J. Sun, "Object Detection Using Deep Learning Techniques: A Survey," IEEE Transactions on Neural Networks and Learning Systems, 2019.
[18]. S. Thrun et al., "Robotic Mapping: A Survey," Exploring Artificial Intelligence in the New Millennium, pp. 1–35, 2003.
[19]. World Health Organization, "Assistive Technology for Persons with Disabilities," WHO Publications, 2023.
[20]. [2S. Li, Z. Xu, and J. Zhang, "Deep Learning Based Real-Time Object Detection System for Assisting Visually Impaired People," Journal of Ambient Intelligence and Humanized Computing, vol. 14, pp. 7421–7433, 2023.