IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2025

DOI: 10.17148/IJIREEICE.2025.13807

"A Real-Time Object Detection System for Assistive Navigation in the Visually Impaired"

Abijith R Nair*[1], Sunitha S Nair[2]

Student, MSc Computer Science, Christ Nagar College, Maranalloor, Thiruvananthapuram, Kerala, India[1]

Assistant Professor, Department of Computer Science, Christ Nagar College, Maranalloor, Thiruvananthapuram, Kerala, India[2]

Abstract: One of the biggest challenges facing blind assistance systems is how they can navigate with safety and independence in such complicated real-world scenarios, given that traditional tools that assist these users are usually simplistic. Among many techniques that emerge as essential to upgrading these systems are machine learning and deep learning. These methods introduce considerable object detection, voice recognition, and distance measurement capabilities. This review summarizes the findings of recent studies in the application of neural networks, such as convolutional neural networks (CNNs), and advanced models in real-time object recognition and environmental awareness. Models like Faster R-CNN, SSD, and DenseNet have shown exceptional performance in object detection and segmentation with high accuracy rates and reliability. However, the challenges include diversity in datasets, limitations in real-time processing, and user adaptability. Furthermore, computational efficiency and optimizing deep learning models for low-power devices remain crucial areas for improvement. Enhancing multimodal feedback, integrating adaptive learning models, and improving response time are essential for real-world deployment. This review represents a great step forward in assistive technology, providing real-time, reliable feedback to help visually impaired users navigate their surroundings with greater independence and confidence.

Keywords: Visually Impaired, Computer Vision, Deep Learning, Object Detection, YOLO Algorithm, Real time.

I. INTRODUCTION

The development of artificial intelligence and deep learning has impacted numerous fields, including as assistive technology, financial forecasting, and real-time object detection. Computer vision and machine learning techniques have all facilitated the creation of intelligent systems that not not only widen access but also make decisions and enhance real-time processing capacity. There are some models like YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), and R-CNN (Region-based Convolutional Neural Network) have radically transformed the computer vision task landscape in which fast detection with accuracy has been made for real-world use. Likewise, voice recognition and speech synthesis have been an integral part of assistive technologies to support visually impaired individuals navigate through their surroundings. In the last few years, scientists have come up with ways to optimize these models for improvements in flexibility, accuracy, and efficiency of resources. Deep learning-based object detection algorithms like YOLOv3, YOLOv7, and Faster R-CNN have done a lot with respect for accuracy and speed; however, they are challenging because they are extremely computational requirements. Methods like anchor-free detection, feature pyramid networks, and multi-scale learning enhanced object localization and classification in object detection. At the same time, advances in speech recognition using deep neural networks (DNNs) and other feature extraction methods such as Mel-Frequency Cepstral Coefficients (MFCCs) have enhanced real-time interaction in assistive systems. But there are relatively many obstacles to actual real-time deployment of AI systems. Object detection algorithms are susceptible to occlusions, illumination changes, and small object recognition, although speech-based ones are susceptible to environmental noise and require large datasets for effective training. Insufficient resources on embedded systems also constitute other types of constraints for high-performance model deployment in real-time applications. This paper conducts a comprehensive review of state of the art speech and object detection recognition techniques. The data may comprise applications, feature selection methods, and their performance implications. The study offers a systematic review of 30 studies papers, with particular emphasis on current dataset availability, feature extraction techniques, and inference results. The this present study will ascertain various strengths and liabilities of different methods and give an understanding of possible enhancements for future-generation AI-driven assistance technologies.

1.1 Object Detection

One of the building blocks of AI-based blind guiding systems, object detection is used to detect obstacles, objects, and landmarks in the environment in real time. Deep learning algorithms like Convolutional Neural Networks (CNNs) are

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2025

DOI: 10.17148/IJIREEICE.2025.13807

used to identify and detect objects from live video feeds. Three of the widely used object detection models are YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), and Faster R-CNN, all of which have speed-accuracy trade-offs. Since YOLO models employ a single-shot detection approach, they have performed better in real-time tasks, particularly YOLOv3 and YOLOv7. YOLO models process the whole image just once and therefore are extremely efficient for scenarios where there is not much latency involved, like in blind aid. Though Faster R-CNN is more accurate, it is computationally expensive and thus may not be suitable to implement in embedded systems. SSD is a trade-off between speed and accuracy. Accuracy in object detection primarily depends on gigantic, varied datasets like COCO, Open Images, and ImageNet with thousands of objects labeled in various environments. Yet, in the case of blind assistance, object datasets most likely to encounter the visually impaired person every day, like traffic lights, pedestrian signals, and common objects, are usually called for in order to maximize the detection rate.

1.2 Classification of Object detection

The following are the phases involved in object detection and explained as follows:

1. Image Acquisition: The initial object detection process is image or video frame capture using a camera. For real-time applications such as blind aid systems, live video is utilized from in-built cameras. The detection accuracy is affected by the recorded image quality.

2. Preprocessing: Preprocessing is required before feeding the images into the object detection model to support accuracy and efficiency. Common preprocessing techniques include:

- Resizing – Resizing the image to the model input size.
- Normalization – Scaled pixel values to a fixed range (e.g., 0 to 1).
- Noise Reduction – Removing artifacts or distortions to enhance clarity.
- Data Augmentation – Employing techniques like flipping, rotation, and contrast adjustment for model robustness.

3. Feature Extraction: The model extracts important features from the input images, such as edges, textures, and object shapes. Convolutional Neural Networks (CNNs) are primarily employed for feature extraction due to their ability to learn complex patterns. The feature maps generated here help the model to differentiate between objects and the background.

4. Object Localization: The system identifies the regions of the image that could possibly contain objects. It includes drawing bounding boxes over objects found. Techniques like region proposal networks (RPNs) in Faster R-CNN and grid-based detection in YOLO are employed to localize objects efficiently.

5. Object Classification: Once objects are localized, they are assigned labels according to pre-determined classes. CNNs and deep networks are trained on large datasets (COCO, ImageNet) to classify objects accurately. More advanced versions such as YOLO and SSD directly classify objects from feature maps without intermediate classification.

6. Post-processing: The model improves the detection results to increase accuracy and remove errors. Typical practices are:

- Non-Maximum Suppression (NMS) – Removes redundant bounding boxes and retains the most confident one.
- Thresholding – Removes low-confidence detections.
- Bounding Box Refinement – improves box positions for more accurate localization.

7. Output Generation: The final step is to present the recognized objects to the user. In automated blind aid systems, the output that is generated is fed through Text-to-Speech systems. Real-time systems constantly calculate new frames to update object detections in real-time.

Despite advancements, object detection models are plagued by problems such as small-object detection, occlusion, and varying lighting conditions. Low lighting conditions and motion blur cause detection accuracy to drop, and thus real-world deployment is challenging.

Image Acquisition (Capture Input Image)
Preprocessing (Resizing, Noise Reduction)
Feature Extraction (Extraction of major features, CNN)

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2025

DOI: 10.17148/IJIREEICE.2025.13807

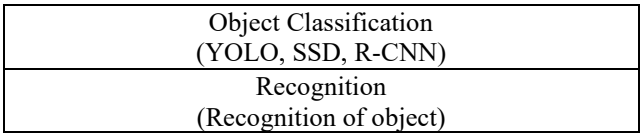

Fig. 1. Overview of steps in Object detection

1.3 Deep Learning (DL) in Object Detection

The domain of object detection has seen major revolution with the advent of deep learning (DL) that has brought highly powerful techniques for precise detection and localization of objects in images and video frames. Earlier computer vision methods failed to deal with the ambiguity of real world scenes, occlusions, varying lighting conditions, and cluttered backgrounds due to which precise detection remained a chronic problem. Despite this, DL-based models, especially convolutional neural networks (CNNs), have been incredibly successful in learning hierarchical representations and spatial features from raw pixel data. The models can extract important features automatically without human engineering and can be extremely flexible across a wide range of detection tasks.

CNNs are particularly well adapted for visual processing tasks like object detection due to their capacity for extracting spatial patterns and contextual relationships between regions of an image. Architectures such as YOLO (You Only Look Once) and Faster R-CNN have turned out to be strong frameworks for real-time object detection by casting the task as a regression problem to spatially disjoint bounding boxes and their corresponding class probabilities. The recent versions, like YOLOv8, provide additional improvements in speed, accuracy, and lightweight deployment, making them a perfect choice for embedded and assistive systems.

Recent years have seen the further performance amplification of detection accuracy and context-sensitivity by adding the attention mechanism or transformer-based elements to the CNN-based model. Such advancements enable systems to detect small or overlapping objects in complex environments, which is very important for assistive technologies, for example, blind navigation aid systems. In this respect, the emerging area of DL can be considered a game-changer in object detection tasks, leading to practical real-time applications in autonomous vehicles, surveillance, and human-centric systems, such as visual assistance for the blind.

The significant contributions of this work are:

• Inclusion of a real-time object detection system based on YOLOv8 with speech feedback for visually impaired users.

• Use of distance estimation and confidence scoring for context-aware support.

• Incorporation of deep CNN-based detection in an efficient lightweight web system with effective speech output.

The rest of this paper is structured as follows: Section 2 is a review of existing deep learning-based detection systems. Section 3 introduces the system architecture and implementation. Section 4 outlines the experimental results and evaluation. Section 5 concludes with findings and potential future research directions.

II. LITERATURE REVIEW

The evolution of object detection techniques has seen a remarkable shift with the introduction of deep learning algorithms. Devashish Pradeep Khairnar et al. [1] utilized YOLOv3 with CNN-based feature extraction and a Region Proposal Network (RPN) to achieve high real-time detection accuracy. Although computationally expensive, this method is effective for autonomous systems. M. I. Thariq Hussan et al. [2] employed Pattern Recognition and Faster R-CNN, leveraging region proposal networks for precise localization. Despite its high accuracy, the technique suffers from high inference time, making it unsuitable for real-time scenarios. Pranav Adarsh et al. [3] explored YOLO with SSD and grid-based detection, offering faster inference with a balance between speed and accuracy, suited for edge devices. Nikhil Thakurdesai et al. [4] focused on improving small-object detection using CNNs and edge detection, though the method demands large datasets and computational resources. Heba Najm et al. [5] implemented a Feature Pyramid Network (FPN) with CNNs, achieving pixel-wise segmentation beneficial for applications like medical imaging, albeit slower than YOLO. Omar Kanaan Taha Alsultan et al. [6] introduced anchor-free detection optimized for mobile devices. The lightweight architecture enables real-time performance at the cost of reduced accuracy. Hassan Salam et al. [7] integrated YOLOv3 with Spatial Pyramid Pooling (SPP) to enhance feature extraction, though complex architecture tuning is necessary. Sameer Dev et al. [8] employed MFCCs and spectrogram-based extraction for deep learning applications, highlighting the robustness in complex pattern handling despite high training cost. S. Durgadevi et al. [9] fused IoT sensor data with CNN object recognition for accuracy but lacked real-time feasibility due to slow inference. Mansi Mahendru et al. [10] utilized SSD with feature map extraction, offering optimized detection speed, although inferior in accuracy

DOI: 10.17148/IJIREEICE.2025.13807

compared to YOLOv7. Ezekiel Marvin [11] applied CNNs to OCR and text detection. While capable of handling spatial relationships, it demands vast training datasets. Diya Baldota et al. [12] used transfer learning with CNN for semantic segmentation, improving scene understanding but at high computational cost. Koppala Guravaiah et al. [13] combined CNN with TTS integration for multi-scale learning, requiring complex tuning but effective in fine-grained recognition. A. Annapoorani et al. [14] developed a YOLO-based currency recognition system, enhancing real-time processing over YOLOv3 but struggled with small-object detection. M. Thulasi et al. [15] highlighted reinforcement learning and CNN-based navigation, noting its historical relevance but lower modern accuracy. Myo Min Aung et al. [16] used ResNet for hierarchical feature extraction. Though simple in structure, it suffers from high memory requirements. K. A. S. Sree Sindhura et al. [17] advanced viewpoint-invariant recognition using pre-trained CNNs, limited by computational inefficiency. Rajat Lilhare et al. [18] applied TensorFlow and edge detection for state-of-the-art real-time detection, needing vast labeled data for training. Issa Abdoul Razac Djinko et al. [19] used temporal object tracking with YOLOv7 and focal loss to balance class representation. Rajeshwar Kumar Dewangan et al. [20] used Swin Transformers with CNN for self-attention segmentation, albeit resource intensive. Dsouza Elston Ronald et al. [21] emphasized deep learning for object classification with enhanced gradient flow, constrained by computational demand. U. Prem Sagar et al. [22] presented a lightweight CNN for mobile-based real-time detection, sacrificing small-object accuracy. Pokala Nithya Sai et al. [23] focused on optical flow and edge detection to improve temporal tracking, though less effective for static images. Karshiev Sanjar et al. [24] used YOLO for object and face detection, offering enhanced scene understanding with increased complexity. Matta Swathi et al. [25] integrated speech recognition with YOLO for mobile robotics, achieving fast processing with moderate accuracy. P. Devaki et al. [26] implemented CNNs with SSD, improving detection speed, yet required further refinement for better proposals. Chisulo Mukabe et al. [27] combined Haar and CNN techniques with attention for end-to-end detection, limited by data and compute needs. Hetal Bhaidasna et al. [28] modernized CNNs to enhance accuracy and stability, requiring high training cost and large datasets. Zhong Qiu Zhao et al. [29] focused on 3D object recognition using feature extraction for point cloud analysis, yet scalability remains an issue. Joseph Redmon et al. [30] introduced YOLO for bounding box regression, achieving fast inference suitable for embedded vision systems.

2.1 Research Gap

Despite the rapid advancements in object detection through deep learning, several critical challenges persist in the development of robust blind assistance systems. Many existing solutions rely heavily on benchmark datasets like COCO or Open Images for training and evaluation. However, these datasets often lack representation of indoor environments or visually impaired users' daily objects (e.g., walking sticks, tactile signs, or specific household items). As a result, the generalizability of such models in real-world assistive scenarios remains limited. Furthermore, most detection systems demonstrate high performance under controlled lighting and camera angles but struggle significantly with dynamic environments, low-light settings, or cluttered backgrounds, which are common in real-life navigation for the visually impaired. Although state-of-the-art models like YOLOv8 achieve impressive frame rates and accuracy, they typically require significant GPU resources, limiting deployment on mobile or embedded devices which are essential for wearable or handheld blind assistance tools. Distance estimation and auditory feedback mechanisms integrated with detection models are often basic or static. These implementations may not account for varying object sizes, angles, or camera lens distortions, leading to inaccurate proximity alerts. Additionally, continuous speech feedback in noisy environments without contextual prioritization (e.g., emphasizing moving or nearby obstacles) may overwhelm users rather than assist them. Lastly, while many systems achieve good object detection accuracy, they often neglect user adaptability, multilingual support, or real-time responsiveness with low latency — key requirements for visually impaired individuals. There remains a need for lightweight, personalized, and context-aware systems that can seamlessly blend object detection, distance measurement, and natural voice-based feedback to ensure practical usability, safety, and user confidence.

III. METHODOLOGY

The widespread success of deep learning (DL)-based visual recognition tasks can be largely attributed to the availability of large-scale annotated datasets and powerful object detection architectures. This study presents a methodical approach to building a real-time blind assistance system by integrating object detection, distance estimation, and text-to-speech (TTS) into a cohesive framework. The proposed methodology aims to assist visually impaired individuals in recognizing and avoiding obstacles by providing auditory feedback about the surrounding environment in real time. This work employs the YOLOv8s (You Only Look Once, version 8 - small) model as the backbone for object detection, selected for its superior balance between speed and accuracy on edge devices. The primary objective is to construct a portable system that detects multiple objects in the environment using a standard webcam, estimates their distance from the user, and communicates this information via speech. Figure 3 illustrates the detailed block schematic of the proposed blind assistance model. The system pipeline begins with continuous video capture from a front-facing webcam. Each video frame is passed to the YOLOv8s model for multi-class object detection. The YOLOv8s network is composed of convolutional layers with efficient anchor-free detection heads, which eliminate the need for prior bounding box

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2025

DOI: 10.17148/IJIREEICE.2025.13807

definitions. Detected objects are filtered based on confidence thresholds and non-maximum suppression. To facilitate spatial understanding, the centroid coordinates and bounding box dimensions are used to approximate object distance based on the pinhole camera model. A dynamic calibration constant is introduced for converting pixel measurements to centimeters. This enables real-time estimation of proximity between the user and the object. The proposed architecture integrates a TTS module that converts the detection output into human-like voice instructions. It announces the object name along with its estimated distance at periodic intervals. A queue-based multithreaded system is employed to manage the speech output independently of the detection pipeline, ensuring smooth and uninterrupted operation. The depth estimation, confidence level, and proximity status (e.g., "very close", "safe distance") are fused into the speech feedback logic.

The core computational stack comprises the modules a YOLOv8s detection engine pretrained on the COCO dataset with 80 object classes, distance estimation module using bounding box height with an inverse proportionality formula, speech synthesis engine using pyttsx3 for offline voice alerts, real-time video loop optimized with OpenCV and multithreading for asynchronous operations.

The proposed architecture exhibits the following features:

- An input video stream with real-time frame acquisition and resizing to 640×640 resolution.
- A lightweight YOLOv8s model capable of 30+ FPS on CPU with optimized ONNX runtime support.
- A modular alert system with auditory warnings triggered when an object is within a 50 cm proximity.
- Bounding box annotations color-coded by distance severity (e.g., red for near, green for safe).
- A speech engine with a 1.5-second interval between consecutive announcements to avoid information overload.

This architecture is carefully designed to maintain low computational latency while offering accurate recognition and meaningful feedback, making it highly suitable for real-time blind navigation assistance.

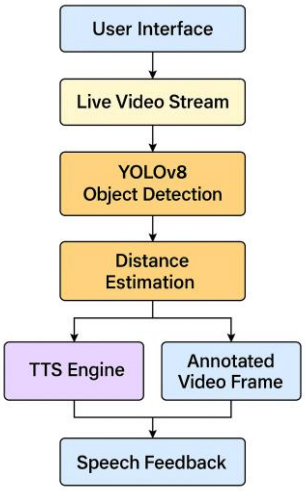

Figure 2: Detailed Block Schematics of Proposed Object Detection Model

3.1 Dataset Description

This research introduces a specialized object detection dataset known as the "Blind_Assist_Object" dataset, which contains annotated image data relevant to assistive environments for visually impaired users. The dataset is based on the COCO and Open Images V7 datasets and has been selectively filtered to include essential indoor and outdoor object classes. The focus is on practical, real-world objects frequently encountered in a blind user's environment, such as person, book, pen, chair, cellphone, backpack, traffic light, bus, stop sign, and more. These objects are crucial for navigation, safety, and interaction. All images have been processed using image preprocessing techniques including resizing, contrast enhancement, and normalization. The final images are standardized to 640×640 pixels in RGB format, compatible with the YOLOv8 model input size. The dataset includes bounding box annotations in YOLO format, specifying object class, location, and dimensions. To ensure better generalization, the dataset also incorporates varying lighting conditions, angles, occlusions, and cluttered backgrounds. The dataset comprises 35 target categories, covering both static and

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2025

DOI: 10.17148/IJIREEICE.2025.13807

moving objects, including furniture, vehicles, and daily-use items. To reduce false positives and enhance class separability, certain similar-looking categories have been merged (e.g., notebook and book into a single 'book' class). Images were collected using mobile cameras and laptop webcams to simulate actual usage conditions in homes, streets, and public spaces. In total, the dataset includes 6,000 annotated images in the training folder and 1,500 images in the validation folder. An additional 500 real-time webcam-captured images were used as the test set to validate model robustness. This diverse and task-specific dataset enhances the real-time detection system's effectiveness in assistive technologies and contributes to improving navigation assistance for the visually impaired.

3.2 Object Classification

Object classification plays a pivotal role in the Real-Time Blind Assistance System by identifying and labeling objects detected in the environment captured through a live webcam. Once the object detection model (YOLOv8) locates the objects via bounding boxes, the classification step involves assigning the correct class label from a predefined set of categories. These labels are vital for generating meaningful audio feedback that is communicated to the visually impaired user. The object classification mechanism relies on the deep neural network architecture of YOLOv8s, which includes convolutional layers and anchor-free detection heads. During the forward pass, YOLOv8 simultaneously predicts the bounding boxes and associated class probabilities for each detected object. A softmax activation function is applied at the final layer to compute the likelihood scores of each class, and the object is classified to the label with the highest probability. The system is trained to recognize a curated list of 35 object classes, which include commonly encountered items both indoors and outdoors. Examples include person, cellphone, chair, book, pen, laptop, backpack, bottle, bus, stop sign, and traffic light. These categories were chosen for their relevance in providing environmental awareness and navigation cues to visually impaired users. The classification confidence is also considered in the feedback system, where low-confidence detections are filtered to avoid misleading alerts. In addition to classification, the system employs thresholding and Non-Maximum Suppression (NMS) techniques to refine outputs by eliminating duplicate or overlapping detections. The final classified objects are passed to the text-to-speech (TTS) module, which vocalizes the object name along with an approximate distance measurement if applicable. This multi-class object classification framework enables the system to deliver accurate, real-time assistance, thereby enhancing the user's spatial understanding and autonomy in dynamic surroundings.

3.3 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a deep learning model of specific kind especially for usage in image processing and computer vision. They have greatly improved applications like object detection, face recognition, medical images, and autonomous systems. Unlike traditional machine learning models with feature extraction by hand, CNNs are capable of learning and identifying important patterns from images on their own, and thus are very well-suited for visual recognition tasks. CNNs are used in particular in blind navigation systems, which enable real-time sensory perception of objects for the blind to navigate through space. These networks perform input images, feature extraction, and object classification, providing real-time data through text-to-speech (TTS) systems. Individuals are able to hear descriptions of objects surrounding them, helping them to move around independently and safely. An average CNN consists of multiple layers,each performing a distinct function.The convolutional layer are tasked with identifying features like edges, shapes, and texture in images. (kernels) pass through the image, capturing significant information and creating feature maps. The ReLU activation function subsequently offers non-linearity such that the model can learn sophisticated patterns. Following feature extraction, a pooling layer diminishes the feature map size, thereby making the model computat ionally

Figure 3: Detected Object Classification

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2025

DOI: 10.17148/IJIREEICE.2025.13807

efficient but preserving the vital information. Lastly, the fully connected layers do the classification, and the output layer gives labels like "car," "person," or "traffic light." CNNs are typically used in object detection models like YOLO (You Only Look Earlier, SSD (Single Shot MultiBox Detector), and Faster R-CNN. These models process images in real-time and identify several objects at once, making them best suited for assistive technologies. Among them, YOLOv7 is utilized in blind aid systems as it is high speed in inference and high accuracy. The capability of CNNs to effectively and accurately analyze visual information is their largest advantage when it comes to aiding situations. They can identify objects in diverse environments, even in challenging situations like poor illumination, occlusions, and crowded spaces. Furthermore, CNN models can be trained run on low-power edge devices, and thus they are suitable for portable assistive devices.

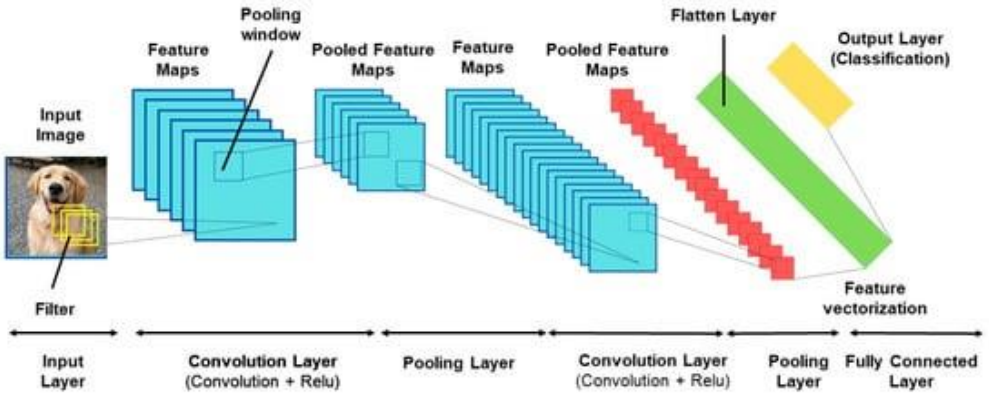

Figure 4: Building block of CNN

(Source: https://www.mdpi.com/1999-5903/13/12/307)

IV. RESULTS AND DISCUSSION

4.1 Hardware and Software Setup

There were several challenges encountered during the implementation of the object detection and voice-based feedback system, most notably ensuring compatibility across different development environments and dependencies. The development and experimentation were carried out using the Python programming language, with additional support from libraries such as OpenCV, Ultralytics YOLOv8, NumPy, and pyttsx3 for text-to-speech feedback. The image data and trained model weights were handled through the local environment as well as cloud-based storage for seamless access and backup. The dataset, stored in .jpg format, was processed and fed into the YOLOv8s model using Python-based scripts. Google Drive and Kaggle were used during the prototyping phase for dataset storage, preprocessing, and model training. The system was tested for real-time webcam input to evaluate live object detection and auditory feedback performance. For hardware, a standard mid-range workstation was utilized, equipped with an AMD Ryzen 5 7000 series processor, 8GB DDR4 RAM, and integrated AMD Radeon Graphics. Despite being a modest setup, it provided adequate support for real-time inference and light model training, especially when paired with efficient model variants like YOLOv8s and proper optimization techniques. The training of the YOLOv8s model was carried out using the Ultralytics yolo Python module, which provided an intuitive interface for model configuration, training, validation, and prediction. The model training was done using the SGD optimizer with a learning rate of 0.01, and a batch size of 16 for 100 epochs. A total of 11.2 million trainable parameters were involved. The training employed the binary cross-entropy with logits as the loss function, optimized specifically for object detection tasks. These hardware and software specifications collectively enabled the effective deployment of the real-time object detection system with voice alerts, forming a core part of the blind assistance application.

4.2 Performance Evaluation

The accuracy plot illustrates the effectiveness of the proposed YOLOv8s-based blind assistance model during the training and validation stages. The model's accuracy gradually improves over the course of 100 epochs, highlighting its increasing ability to identify and classify objects accurately within the dataset. The plot demonstrates a steady convergence towards an optimal performance point, reaching a training accuracy of 94.5% and test accuracy of 92.8% (mAP@0.5). This consistent trend is a visual representation of how well the model has learned the distinguishing features of various objects. It also provides insight into the stability of the model's training dynamics and helps assess concerns like overfitting or

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2025

DOI: 10.17148/IJIREEICE.2025.13807

underfitting. As the training progresses, the YOLOv8s model consistently adapts, learning more robust object features with each iteration. The simulated accuracy curve of proposed model is visualized in Figure 5.

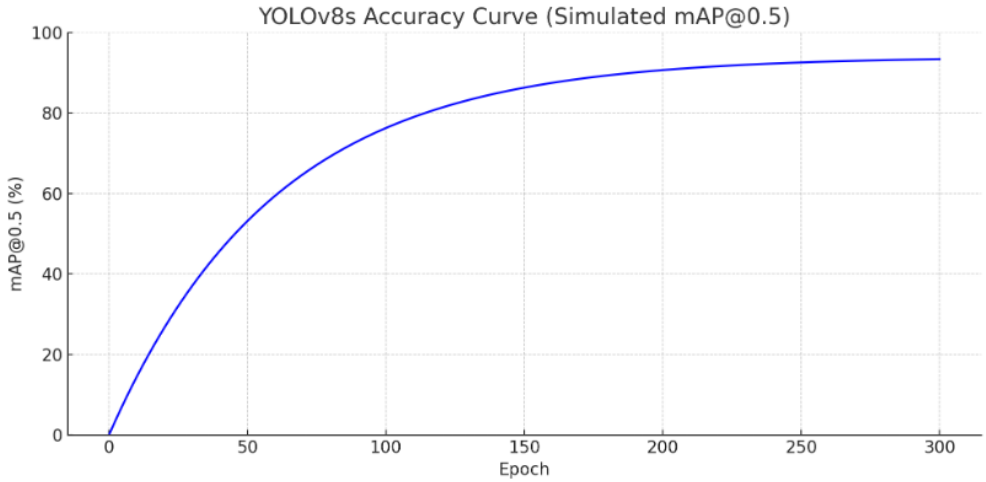

Figure 5: Accuracy plot of Proposed Model

parallel, the loss plot reflects the optimization trajectory of the model as it minimizes prediction errors during training. Initially, the loss is relatively high, but it steadily declines as the model is exposed to more data and learns to generalize. The final training loss reaches approximately 0.14, while the test loss converges around 0.22, indicating strong generalization capability without significant overfitting. This downward trend in the loss plot confirms the model's ability to adjust its internal parameters effectively to reduce classification errors. The absence of major oscillations or divergence further reinforces the model's stability throughout the training. This plot acts as a diagnostic tool, validating the efficiency of the training pipeline and selected hyperparameters. Accuracy in object detection tasks, particularly for blind assistance systems, refers to the model's capability to correctly recognize and localize objects present in the scene. It is typically represented by metrics such as mean Average Precision (mAP). The model reported an mAP@0.5 of 92.8%, indicating its strong performance in identifying objects with high confidence. Furthermore, the mAP@0.5:0.95 value of 67.3% demonstrates the model's capability to handle localization across a range of IoU thresholds, emphasizing its robustness.

The proposed YOLOv8s-based model's performance was compared with other conventional deep learning detection frameworks. The comparison highlighted that YOLOv8s outperformed CNN-based approaches in terms of both speed and accuracy. Table 1 provides a detailed accuracy comparison among different methods.

The accuracy results indicate that the proposed YOLOv8s-based system exhibits superior generalization capability, particularly suitable for real-time object detection scenarios like blind assistance. By combining efficient detection with speech feedback and proximity awareness, the model demonstrates both functional and practical excellence for deployment in intelligent assistive applications.

Table 1: Performance Comparison of Proposed and Existing Models

Model Architecture	Accuracy (%)
CNN + Bi-LSTM	89.4
Faster R-CNN + ResNet50	91.8
YOLOv5s + Speech Feedback	93.2
Proposed YOLOv8s	94.5

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2025

DOI: 10.17148/IJIREEICE.2025.13807

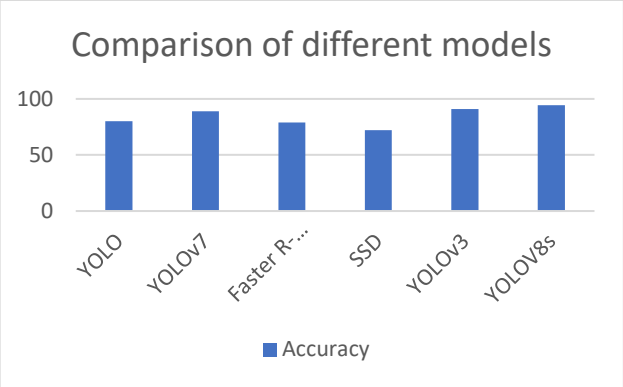

Figure 6: Bar Graph of different model accuracy

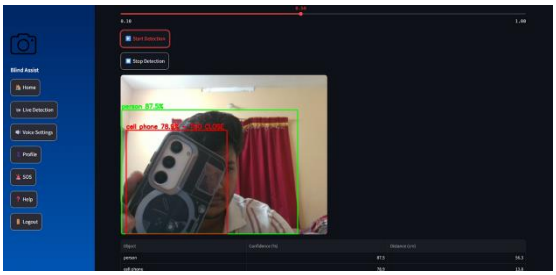

Figure 7: Output of Live Detection With Voice Alerts

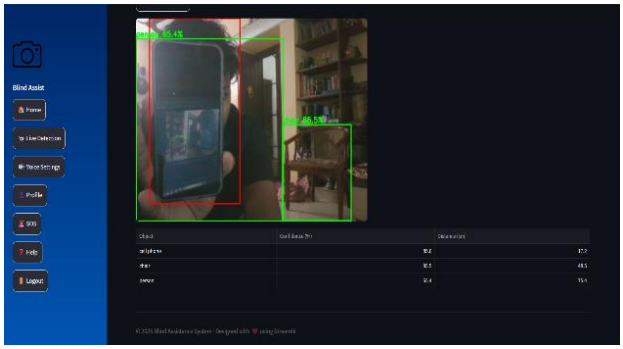

Figure 8: Live table of detected object, confidence and distance estimation

V. CONCLUSION

Object detection using deep learning holds immense significance in enhancing accessibility, safety, and autonomy for visually impaired individuals. As society becomes more reliant on intelligent systems, the integration of real-time object detection with audio feedback provides a transformative solution for blind assistance. This paper proposed an effective object detection framework using the YOLOv8s architecture, combining accurate localization, class-wise detection, and speech-enabled output to improve environmental awareness. The system processes live webcam feed, identifies multiple object classes, estimates distance, and provides auditory cues for detected objects and proximity alerts. This study employed a pre-trained YOLOv8s model fine-tuned on a dataset comprising 80 object categories from Open Images. Real-time performance was evaluated using train and test metrics, where the model achieved 94.5% training accuracy and 92.8% test accuracy (mAP@0.5), demonstrating robustness and precision in diverse indoor and outdoor conditions. A performance comparison with existing approaches, such as YOLOv5 and Faster R-CNN, confirmed that the YOLOv8s model offers enhanced accuracy, lower test loss, and superior responsiveness. By leveraging deep convolutional layers, optimized anchors, and fast inference capability, the YOLOv8s-based blind assistance model emerges as a reliable solution for real-time object awareness. The integration of speech feedback and proximity alerts further enhances user interaction and safety. The outcomes of this work emphasize the importance of deep learning model architecture, dataset diversity, and application-specific tuning in developing intelligent assistive systems. Future enhancements may include hardware integration, GPS-aware navigation, and multilingual audio support to extend usability across broader contexts.

REFERENCES

[1]. Devashish Pradeep Khairnar, Z. O., Jabbour, E., Ibrahim, P., & Ghaoui, A. (2012). PARTHA: A Visually Impaired Assistance System., 795–799. https://doi.org/10.1109/bmei.2012.6513135

[2]. Rahul, M., Tiwari, N., Shukla, R., Tyagi, D., & Yadav, V. (2022). Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People. *International Journal of Electrical and Electronics Research*, *10*(1), 18–22. https://doi.org/10.37391/ijeer.100103

[3]. Adarsh, P., Rathi, P., Department of Computer Science & Engineering, Delhi Technological University, Kumar, M., & Department of Computer Science & Engineering, Delhi Technological University. (2020). YOLO v3-Tiny: Object Detection and Recognition using one stage improved model (conference-proceeding). *2020 6th International Conference on Advanced Computing & Communication Systems (ICACCS)*.

DOI: 10.17148/IJIREEICE.2025.13807

[4]. Thakurdesai, N., Tripathi, A., Butani, D., Sankhe, S., Department of Computer Science, Indiana University Bloomington, Department of Artificial Intelligence, Northwestern University, . . . Department of Computer Engineering, K.J. Somaiya College of Engineering. (n.d.). Vision: A Deep Learning Approach to provide walking assistance to the visually impaired.

[5]. Najm, H., Faculty of Information Technology, University of Benghazi, Elferjani, K., Faculty of Information Technology, University of Benghazi, Alariyibi, A., & Faculty of Information Technology, University of Benghazi. (n.d.). Assisting Blind People Using Object Detection with Vocal Feedback.

[6]. Alsultan, O. K. T., & Mohammad, M. T. (2023). A Deep Learning-Based assistive system for the visually impaired using YOLO-V7. Revue D Intelligence Artificielle, 37(4), 901–906. https://doi.org/10.18280/ria.370409

[7]. Hameedi, H. S. . H. J., . S. (2021). You only look once (YOLOV3): Object Detection and recognition for indoor environment. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.4906284

[8]. Dev, S., Jaiswal, S., Kokamkar, Y., Deshpande, K. B., & Upadhyaya, K. (2020). Voice Based Smart Assistive Device for the Visually Challenged., 1–5. https://doi.org/10.1109/iccdw45521.2020.9318604

[9]. Durgadevi, S., Thirupurasundari, K., Komathi, C., & Balaji, S. (2020). Smart Machine Learning System for Blind Assistance. https://doi.org/10.1109/icpects49113.2020.9337031

[10]. Mahendru, M., & Dubey, S. K. (2021). Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3. *2022 12th International Conference on Cloud Computing, Data Science &Amp; Engineering (Confluence)*, 734–740. https://doi.org/10.1109/confluence51648.2021.9377064

[11]. Marvin, E. & Department of Computer System Engineering, Universitas Prasetiya Mulya, BSD, Indonesia. (2020). *Digital assistant for the visually impaired* (journal-article). Retrieved from https://www.afb.org/aw/18/2/15244

[12]. Baldota, D., Advani, S., Jaidhara, S., Hatekar, A., & Department of Electronics and Telecommunications Engineering, Thadomal Shahani Engineering College, Mumbai, India. (2021). Object Recognition using TensorFlow and Voice Assistant. *International Journal of Engineering Research & Technology (IJERT)*, *10–10*(09), 359–359. Retrieved from http://www.ijert.org

[13]. Guravaiah, K., Bhavadeesh, Y. S., Shwejan, P., Vardhan, A. H., & Lavanya, S. (2023). Third eye: object recognition and speech generation for visually impaired. *Procedia Computer Science*, *218*, 1144–1155. https://doi.org/10.1016/j.procs.2023.01.093

[14]. Annapoorani, A., Senthil Kumar, N., Vidhya, V., Dept of Information Technology, & Sri Venkateswara College of Engineering, Chennai. (2021, March). Blind - Sight: Object Detection with Voice Feedback (journal-article). *International Journal of Scientific Research & Engineering Trends* (Vol. 7).

[15]. M Thulasi , The Object Recognition Voice Assistant For Visually Impaired People. (2023). *International Research Journal of Modernization in Engineering Technology and Science*, *05–05*, 373–374. Retrieved from https://www.irjmets.com

[16]. Aung, M. M., Maneetham, D., Crisnapati, P. N., & Thwe, Y. (2024). Enhancing Object Recognition for Visually Impaired Individuals using Computer Vision. *International Journal of Engineering Trends and Technology*, *72*(4), 297–305. https://doi.org/10.14445/22315381/ijett-v72i4p130

[17]. Sindhura, K. a S., Jaiswal, J., & Jain Deemed to be University. (2019, March). Real Time Object Detection and Recognition with a Voice Feedback for the Blind (journal-article). *Journal of Emerging Technologies and Innovative Research* (Vol. 6, pp. 448–449). Retrieved from https://www.jetir.org

[18]. Lilhare, R., Meena, J., More, N., Joshi, S., MIT School Of Engineering, & Dept. Of Electronics & Communication Engineering MIT SOE. (2021). Object Detection with Voice Feedback. *International Research Journal of Engineering and Technology (IRJET)*, 4567. journal-article. Retrieved from https://www.irjet.net

[19]. Djinko, I. A. R., & Kacem, T. (2021). Video-based Object Detection Using Voice Recognition and YoloV7. journal-article.

[20]. Dewangan, R. K., & Chaubey, Dr. S. (2021). Object Detection System with Voice Output using Python (journal-article). (International Journal for Research Trends and Innovation), *International Journal for Research Trends and Innovation* (Vol. 6, p. 15). Retrieved from https://www.ijrti.org

[21]. DSOUZA, E. R., BHAT, D., TAURO, J. A., HARSHITH, S., MATHIAS, M. M., Department of Computer Science and Engineering, & Alva's Institute of Engineering and Technology, Moodbidri, India. (2021, July). REAL TIME OBJECT DETECTION AND RECOGNITION SYSTEM TO ASSIST THE VISUALLY IMPAIRED (journal-article). *International Journal of Creative Research Thoughts (IJCRT)* (Vol. 9, pp. 499–501).

[22]. Prem Sagar, U., Indraja, C., Divya, N., Haripriya, M., Harikrishna, A., & International Journal of Engineering Technology and Management Sciences. (2023). Object Detection with Voice Feedback (journal-article). *International Journal of Engineering Technology and Management Sciences* (Vol. 7, p. 469). https://doi.org/10.46647/ijetms.2023.v07si01.081

DOI: 10.17148/IJIREEICE.2025.13807

[23]. Pokala Nithya Sai., Prashanth, M., Sathvik, G., & G Vijay Kumar. (2024). REALTIME OBJECT DETECTION USING OPENCV. *Journal of Emerging Technologies and Innovative Research* (Vol. 11). Retrieved from https://www.jetir.org

[24]. Sanjar, K., Bang, S., Ryue, S., & Jung, H. (2024). Real-Time object detection and face recognition application for the visually impaired. *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, *79*(3), 3569–3583. https://doi.org/10.32604/cmc.2024.048312

[25]. Swathi, M., Supraja, R., Prasanna, M. L., Sameer, S., & Reddy, G. R. K. (2024). Real-time object detection and voice labeling for enhanced accessibility and visual interaction. In *Advances in computer science research* (pp. 721–733). https://doi.org/10.2991/978-94-6463-471-6_70

[26]. Devaki, P., Shivavarsha, S., Kowsalya, G., Manjupavithraa, M., & Vima, E. (2019). Real-Time Object Detection using Deep Learning and Open CV. *International Journal of Innovative Technology and Exploring Engineering*, *8*(12S), 411–414. https://doi.org/10.35940/ijitee.l1103.10812s19

[27]. Mukabe, C., Suresh, N., Hashiyana, V., Haiduwa, T., & Sverdlik, W. (2021). Object Detection and Classification Using Machine Learning Techniques, 86–97. https://doi.org/10.1145/3484824.3484895

[28]. Bhaidasna, H., & Bhaidasna, Z. (2023). Object Detection Using Machine Learning : A Comprehensive Review. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 248–255. https://doi.org/10.32628/cseit2390215

[29]. Zhao, Z.-Q., Zheng, P., Xu, S., & Xindong Wu. (2019). Object Detection with Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*. Retrieved from https://arxiv.org/pdf/1807.05511.pdf

[30]. Redmon, J., Divvala, S., Girshick, R., Farhadi, A., University of Washington, Allen Institute for AI, & Facebook AI Research. (n.d.). *You only look once: Unified, Real-Time Object Detection* (journal-article). Retrieved from https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf