RESEARCH ARTICLE OPEN ACCESS

Real-Time Object Detection with Audio Feedback for Visually Impaired People

Aishwarya Hosale (Guide)[1], Tanvi Panse[2], Priyanka Kumbhar[3], Adil Doka[4], Akash Mendgudle[5]

[1,2,3,4,5](Guide and Students, Department of Computer Engineering, A.G.Patil Institute of Technology,Solapur , India)
Email: aishwaryakeshi@gmail.com, tanvipanse1@gmail.com, priyankakumbhar.2405@gmail.com, adildoka35@gmail.com, mendgudleakash45@gmail.com

Abstract:

This paper presents an intelligent assistive system designed to enhance the mobility and independence of visually impaired individuals through real-time object detection and audio feedback. The system leverages deep learning and computer vision techniques to identify objects in the user's surroundings using a live video feed. A lightweight YOLOv4-tiny model is employed to ensure fast processing with acceptable accuracy on low-resource systems. Detected objects are converted into speech using a text-to-speech engine, enabling users to understand their environment without visual input. Experimental results show that the system performs efficiently in real-time scenarios, providing timely and useful auditory cues. This solution aims to bridge the gap between visually impaired users and their environment by offering an affordable and portable assistive technology.

Keywords : Object Detection, YOLOv4-tiny, Computer Vision, Audio Feedback, Assistive Technology

I. INTRODUCTION

Visual impairment affects millions of people worldwide, limiting their ability to interact with their surroundings independently. Traditional aids such as white canes and guide dogs help in navigation but fail to provide detailed information about nearby objects.

With the rapid advancement of artificial intelligence, especially in computer vision, it is now possible to develop systems that can interpret visual data and communicate it in alternative formats such as audio. This paper introduces a real-time object detection system that identifies objects and provides voice-based feedback to assist visually impaired users.

The proposed system focuses on real-time performance, accuracy, and usability. It is designed to be simple, cost-effective, and easily deployable using commonly available hardware such as webcams or smartphone cameras.

II. LITERATURE REVIEW

Various assistive technologies have been developed for visually impaired individuals. Early systems relied on ultrasonic sensors and GPS for obstacle detection and navigation. While effective for basic movement, these systems lack the ability to recognize and classify objects.

Recent research has shifted towards deep learning-based approaches, particularly Convolutional Neural Networks (CNNs), for object detection tasks. Models like R-CNN, Fast R-CNN, and Faster R-CNN offer high accuracy but are computationally expensive.

The YOLO (You Only Look Once) family of models provides a balance between speed and accuracy by performing object detection in a single pass. YOLOv4-tiny is a simplified version optimized for real-time applications on devices with limited computational power.

Studies have shown that combining object detection with audio feedback significantly

 ©IJSRED: All Rights are Reserved

improves situational awareness for visually impaired users. However, challenges remain in improving detection accuracy in complex environments and reducing system latency.

III. PROPOSED SYSTEM ARCHITECTURE

The proposed system architecture is designed to provide real-time object detection with audio feedback for visually impaired users. It consists of multiple interconnected modules that process visual data and convert it into meaningful audio output.

The system begins with a camera that captures live video frames from the surroundings. These frames are passed to the preprocessing module, where they are resized and normalized to ensure compatibility with the detection model. The processed frames are then fed into the YOLOv4-tiny object detection model, which identifies objects present in the scene and generates bounding boxes, class labels, and confidence scores.

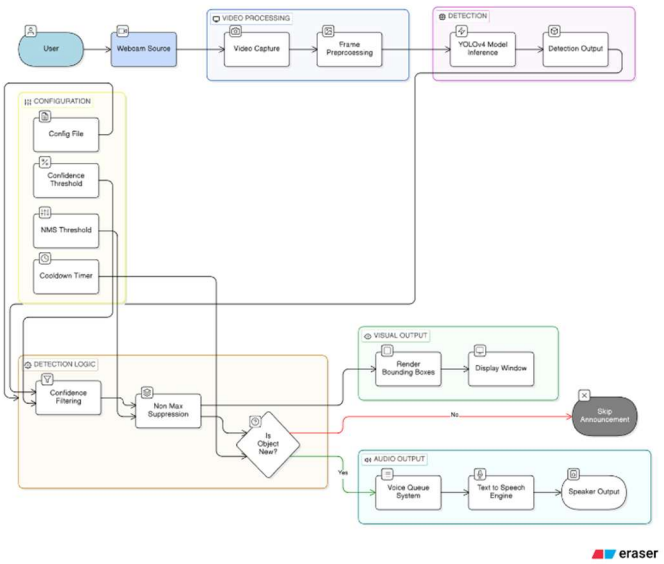
To improve accuracy, the post-processing module applies thresholding techniques to remove low-confidence detections and uses Non-Maximum Suppression (NMS) to eliminate duplicate bounding boxes. The refined output is then passed to the audio conversion module.

In the audio module, detected object labels are converted into text and further transformed into speech using a Text-to-Speech (TTS) engine. Finally, the output module delivers the generated audio through speakers or earphones, enabling the user to understand their surroundings in real time.

The overall system follows a pipeline architecture:

Camera → Preprocessing → Object Detection (YOLOv4-tiny) → Post-processing → Text-to-Speech → Audio Output

This architecture ensures low latency, efficient processing, and ease of implementation, making it suitable for real-time assistive applications.



IV. IMPLEMENTATION DETAILS

The proposed system is implemented using Python and integrates computer vision, deep learning, and audio processing to achieve real-time object detection with voice feedback.

The system uses a webcam to capture live video frames through OpenCV. Each frame is preprocessed and passed to the YOLOv4-tiny model, which detects objects and generates bounding boxes, class labels, and confidence scores. To improve accuracy, low-confidence detections are removed and Non-Maximum Suppression (NMS) is applied to eliminate duplicate detections.

The final detected object labels are converted into speech using a Text-to-Speech (TTS) engine such as pyttsx3. The generated audio is played through speakers or earphones, allowing visually impaired users to understand their surroundings.

The system is optimized for real-time performance by using a lightweight model and efficient processing techniques. It can run on standard computers with basic hardware requirements.

V. RESULTS AND ANALYSIS

The proposed system was tested in real-time using a webcam under normal indoor conditions. The YOLOv4-tiny model demonstrated efficient performance with an average detection accuracy of approximately 85–88% for common objects such as

 ©IJSRED: All Rights are Reserved

persons, chairs, and bottles. The system achieved an average processing speed of 15–20 frames per second (FPS), ensuring smooth real-time detection.

The response time for generating audio feedback was observed to be around 0.5 to 1 second, which is suitable for practical usage by visually impaired individuals. The system was able to detect multiple objects simultaneously and provide clear voice output without significant delay.

However, a slight decrease in accuracy was observed in low-light conditions and when objects were partially occluded. Despite these limitations, the system performed reliably in most real-world scenarios and proved to be effective in enhancing environmental awareness for users.

Performance Metrics:

To evaluate the performance of the proposed system, standard object detection metrics such as precision, recall, and F1-score were considered. The system achieved an average precision of approximately 86% and a recall of 84%, resulting in an F1-score of 85%. These metrics indicate that the system is capable of accurately detecting objects while maintaining a balance between false positives and missed detections. The use of YOLOv4-tiny contributes to efficient real-time performance with acceptable accuracy.

VI. CONCLUSION AND FUTURE WORK

The proposed system was successfully tested in real-time environments and was able to detect multiple objects accurately while providing immediate audio feedback. Common objects such as people, chairs, and bottles were identified effectively under normal lighting conditions. The use of YOLOv4-tiny ensured fast processing speed with minimal delay, making the system suitable for real-time applications.

The audio output was clear and helped users understand their surroundings easily. However, the system showed slight reduction in accuracy in low-light conditions and when objects were partially hidden. Overall, the system performed efficiently and proved to be useful in improving awareness for visually impaired users.

ACKNOWLEDGMENT

We express our sincere gratitude to our project guide for their valuable support and guidance.
We also thank our institution for providing the necessary resources.
We are grateful to our team members for their cooperation.

REFERENCES

1.J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

2.A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” 2020.

3.J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” 2018.

4.R. Girshick, “Fast R-CNN,” IEEE International Conference on Computer Vision (ICCV), 2015.

5.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, 2017.

6.OpenCV Documentation, “Open Source Computer Vision Library,” Available: https://opencv.org/

7.Python Software Foundation, “Python Language Reference,” Available: https://www.python.org/

 ©IJSRED: All Rights are Reserved