International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Real Time Object Detection with Auditory Alerts For Visually Impaired Individuals

Ms. D. Renuka
Assistant
Professor
Department of Computer Science and
Engineering Mahatma Gandhi Institute of
Technology Hyderabad 500075, India
drenuka_cse@mgit.ac.in

Sai Sahitya Kannam Reddy
Department of Computer Science and Engineering
Mahatma Gandhi Institute of Technology
Hyderabad 500075, India
ssahityakannamreddy_cse2405y1@mgit.ac.in

Ms. B. Pravalika Reddy Assistant Professor
Department of Computer Science and Engineering
Mahatma Gandhi Institute of Technology
Hyderabad 500075, India
bpravalika_cse@mgit.ac.in

Bogini Nandini
Department of Computer Science and Engineering
Mahatma Gandhi Institute of Technology
Hyderabad 500075, India
bnandini_cse255a0528@mgit.ac.in

***Abstract*—This paper presents a real-time object detection system designed to assist visually impaired individuals through audio alerts. The system is implemented as a web-based application using HTML, CSS, and JavaScript, making it accessible and easy to use. It utilizes TensorFlow.js along with the COCO-SSD pre-trained model to detect objects from live video captured through a camera. The system identifies common objects such as people, vehicles, and everyday items, and determines their position relative to the user. Based on the detection results, appropriate alert messages are generated and converted into speech using the Web Speech API. This enables users to receive real-time information about their surroundings without relying on visual input. The system operates continuously and provides quick responses with minimal delay. Overall, the proposed solution is cost-effective, efficient, and demonstrates the practical application of machine learning in assistive technology.**

The system is designed to function in real-world environments and can handle multiple object detections simultaneously. Its web-based nature ensures portability and ease of access across devices. This approach highlights the potential of integrating modern web technologies with artificial intelligence for developing practical assistive solutions.

The system supports real-time processing and can detect multiple objects simultaneously. It is designed to be user-friendly and accessible across different devices through a web browser. This approach demonstrates how machine learning and web technologies can be effectively combined to create a practical assistive solution.

I. INTRODUCTION

THIS advancement of computer vision and machine learning has enabled the development of intelligent systems that can assist individuals in their daily lives.

Navigating through surroundings without visual assistance can be challenging and may lead to accidents or difficulties in identifying obstacles. Therefore, there is a need for an efficient and accessible system that can provide real-time information about the environment. The proposed system presents a web-based solution for real-time object detection with voice alerts. It utilizes TensorFlow.js along with the COCO-SSD model to detect objects through live video captured using a camera. The system processes the detected objects and determines their position relative to the user, such as left, right, or ahead. Based on this information, appropriate alert messages are generated and converted into speech using the Web Speech API. The system is designed to operate continuously, providing instant feedback to the user. In addition to improving accessibility, the system aims to provide a user-friendly interface that can be easily operated without technical expertise. Since the application runs directly in a web browser, it eliminates the need for installing complex software or using expensive hardware devices. This makes the system highly affordable and widely accessible. Furthermore, the use of real-time processing ensures minimal delay in detection and response, which is crucial for practical

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

usage. Overall, the system demonstrates how modern web technologies and machine learning can be combined to create an effective assistive solution for visually impaired individuals. The system focuses on delivering a seamless and responsive experience by integrating object detection and audio feedback into a single platform. It is capable of identifying multiple objects simultaneously and prioritizing important obstacles that may affect user safety. The use of pre-trained models reduces the need for extensive training data and allows faster deployment. The application is designed to function efficiently in real-time scenarios, adapting to dynamic environments. It can be used both indoors and outdoors, making it versatile for different use cases. The incorporation of speech output ensures that users receive clear and immediate information without relying on visual cues.

A. BACKGROUND AND SIGNIFICANCE

In recent years, the advancement of computer vision and machine learning has enabled the development of intelligent systems that can assist individuals in their daily activities. For visually impaired individuals, navigating through surroundings safely remains a major challenge due to the inability to identify obstacles and objects in real time. Traditional methods such as white canes provide limited assistance and cannot offer detailed information about the environment. Existing assistive technologies often require expensive hardware or complex setups, making them less accessible to a wider population.

The significance of the proposed system lies in its simplicity, affordability, and accessibility. By using a web-based approach with TensorFlow.js and the COCO-SSD model, the system enables real-time object detection directly in the browser without requiring specialized hardware. The integration of audio alerts using the Web peech API helps users understand their surroundings effectively. This system bridges the gap between advanced technology and practical usability, providing a cost-effective solution that enhances safety, independence, and quality of life for visually impaired individuals.

B. OBJECTIVES

1. The primary objective of the Real-Time Object Detection System is to develop an efficient and user-friendly application that assists visually impaired individuals through real-time object detection and audio alerts. The specific objectives include:
2. To develop a system that detects objects such as people, vehicles, and obstacles using machine learning techniques
3. To create a simple and interactive web-based interface using HTML, CSS, and JavaScript
4. To ensure real-time processing of video input for quick and accurate detection
5. To provide audio alerts by converting detected object information into speech using Web Speech API
6. To design a system that works directly in a web browser without requiring complex installations or specialized hardware
7. To improve user safety and reduce dependency on traditional navigation aids

C. DEFINITIONS AND SCOPE

- The Real-Time Object Detection System is a web-based application designed to assist visually impaired individuals by identifying objects and providing audio alerts. It operates using a pre-trained COCO-SSD model with TensorFlow.js to detect objects from live camera input.
- The system is intended primarily for visually impaired users to help them navigate safely in their surroundings. It does not require specialized hardware or complex installations, as it runs directly in a web browser.

D. LITERATURE SURVEY

Sl. No.	YOP	Title	Authors	Methodologies and Technology Used	Merits	Demerits
1.	2025	Object Detection Tracking and Alert System for Visually Impaired Persons.	J. Tharun Kumar and M. Joshna	A deep learning based approach- VGG-19 is used to classify ODIR dataset into multiple classes. Dataset has 5000 images of fundus.	Real-time detection, audio alerts, improved situational awareness, balanced speed & accuracy.	Dataset quality dependency, hardware requirements, limited scalability, Bluetooth integration needed.
2.	2024	Enhancing Object Detection in Assistive Technology for the Visually Impaired: A DETR-Based Approach.	Sumnia Ikram and Sujan Oyawali	DETR (Detection Transformer), deep learning-based object detection, real-time image processing, camera-based input, audio feedback system.	98% accuracy, real-time feedback, scalable, reliable.	Sensor wear over time, environmental sensitivity, complex integration.
3.	2024	Real-Time Object Detection and Audio Feedback Wearable System for Visually impaired Using Raspberry Pi.	M. Keerthana and J. Keren	YOLOv4, SSD, MobileNet V2, COCO dataset, TensorFlow, Raspberry Pi, real-time object detection with audio output.	Lower mAP (41.3%), limited accuracy vs newer models, Raspberry Pi hardware constraints.	Limited dataset, hardware demand, lighting-dependent, GUI less suited for blind users.

Table .1. Literature SurveyThe paper titled “Object Detection Tracking and Alert System for Visually Impaired Persons”

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

focuses on developing an assistive system using deep learning techniques. The proposed methodology uses a VGG-19 model to classify images from the ODIR dataset, which consists of around 5000 fundus images. The system is designed to identify objects and provide real-time audio alerts to help visually impaired individuals navigate safely.

II. METHODOLOGY

The system captures real-time video using a camera and processes it into frames.TensorFlow.js with the COCO-SSD model is used to detect objects in each frame. The detected objects are analyzed to determine their position (left, right, or ahead). Based on this, alert messages are generated for the user. These messages are converted into speech using the Web Speech API for real-time assistance.

Fig.3.System Architecture

A. System Design

The system is designed as a web-based application with a simple and user-friendly interface. The interface allows users to enable the webcam and start real-time object detection. The design ensures that the system is easy to use and accessible to users with minimal technical knowledge. The architecture of the system consists of input, processing, and output modules. The input module captures live video through the camera, the processing module performs object detection using machine learning, and the output module provides visual and audio alerts to the user.

B. Tools and Technologies

The system is developed using the following tools and technologies:

- **HTML:** Used for structuring the web application
- **CSS:** Used for designing and styling the interface
- **JavaScript:** Used for implementing application logic and processing
- **TensorFlow.js:** Used to run the COCO-SSD object detection model in the browser
- **COCO-SSD Model:** Pre-trained model used for detecting objects
- **Web Speech API:** Used for converting text into audio alerts

These technologies ensure that the system is lightweight, efficient, and easy to use without requiring additional installations.

C. Working Process

1. The user opens the web application and enables the webcam.
2. The system captures real-time video input.
3. TensorFlow.js loads the COCO-SSD model for object detection.
4. Objects are detected and their positions are identified
5. Alert messages are generated based on detected objects
6. The messages are converted into speech using Web Speech API
7. Audio alerts are provided to the user in real time
8. This process runs continuously to ensure effective assistance

IMPLEMENTATION (BALANCED VERSION)

The implementation of the Real-Time Object Detection System focuses on integrating the user interface, object detection logic, and audio output into a functional web-based application. The system is developed using HTML, CSS, and JavaScript, with TensorFlow.js used for implementing object detection. The implementation follows a modular approach, ensuring smooth interaction between different components of the system.

Fig.4.Initial Interface

D. User Interface Implementation

The user interface is designed using HTML and CSS to provide a simple and user-friendly experience. It includes buttons to enable and stop the webcam, along with a video display area for real-time detection. Components such as labels and visual bounding boxes are used to display detected objects. JavaScript event handling is used to manage user actions such as starting and stopping detection. The interface updates dynamically to show detection results continuously, ensuring smooth interaction without refreshing the page.

E. Backend Logic Implementation

The backend logic is implemented using JavaScript and TensorFlow.js. The system captures video frames and processes them using the COCO-SSD model for object detection. The detected objects are analyzed to determine their position such as left, right, or ahead. Based on this, alert messages are generated. The logic is designed to ensure efficient processing and real-time performance.

F. File Handling and Data Management

The system uses a pre-trained COCO-SSD model as its data source for object detection. It does not require a separate database, as the model is capable of detecting multiple objects from real-time input. The detected data, including object class and position, is processed dynamically. This approach reduces complexity and improves performance.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

G. Detection and Audio Alert Module

The system detects objects in real time and displays them using bounding boxes and labels. It generates alert messages such as "Person ahead" or "Object on left." These messages are converted into speech using the Web Speech API. This module ensures that users receive clear and immediate audio feedback about their surroundings.

H. Error Handling and Validation

The system includes error handling to manage issues such as camera access denial or model loading errors. If the webcam is not enabled, appropriate messages are displayed. JavaScript error handling ensures smooth execution without crashes. This improves reliability and user experience.

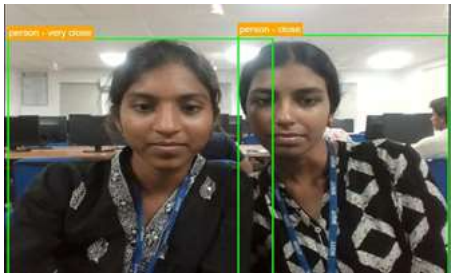
I. System Execution and Packaging

The application runs directly in a web browser without requiring installation. Users can open the website and start detection instantly. Since the system is web-based, it is portable and can be accessed on different devices with a camera and browser support.

J. Overall System Integration

All modules of the system are integrated to work together efficiently. The flow begins with video capture, followed by object detection, processing, and audio output. The modular design ensures smooth operation and real-time performance, resulting in an effective assistive system.

III. RESULTS AND DISCUSSION



The Real-Time Object Detection System was tested under different real-world conditions to evaluate its performance, accuracy, and usability. The testing involved detecting various objects such as people, bottles, laptops, and other items using live video input.

A. Functional Results

The system successfully captures real-time video and detects objects using the COCO-SSD model. Bounding boxes and labels are displayed accurately on the screen. The system generates audio alerts based on detected objects and their positions. The results are displayed instantly, demonstrating efficient real-time performance.

B. User Interface Performance

The web-based interface performs efficiently and provides a smooth user experience. The layout is simple and easy to use, allowing users to interact without confusion. Buttons and controls respond quickly, and results are updated dynamically without refreshing the page.

C. Accuracy of Results

The accuracy of the system depends on the COCO-SSD model. During testing, the system successfully detected common objects with good accuracy. The confidence scores indicate the reliability of detection. However, slight variations may occur under poor lighting conditions or complex backgrounds.

D. Performance Analysis

The system demonstrates fast response time due to real-time processing in the browser. It does not require heavy hardware or external databases, making it lightweight and efficient. The use of web technologies ensures accessibility and consistent performance across devices.

E. Limitations

Despite its effectiveness, the system has certain limitations. The accuracy of object detection depends on the COCO-SSD model, which may not always detect objects correctly in complex environments. Factors such as poor lighting, occlusions, or low camera quality can affect detection performance.

Additionally, the system provides approximate position information (left, right, ahead) but does not calculate exact distance or depth of objects. The audio alerts may sometimes repeat or have slight delays due to browser limitations. The system also depends on camera permissions and browser compatibility for proper functioning.

Furthermore, the system currently detects only a limited set of objects based on the pre-trained model and does not adapt dynamically to new objects or environments. These limitations indicate areas for future improvement.

F. Discussion

The results indicate that the Career Guidance System achieves its primary objective of providing simple and effective career recommendations. The system is particularly useful for students who need quick guidance without relying on external resources.

The use of Java Swing ensures a user-friendly interface, while file handling simplifies implementation. Compared to more complex systems, this approach offers better accessibility and ease of use, although it sacrifices some level of flexibility and scalability.

Overall, the system provides a strong foundation for career recommendation and can be enhanced in the future by integrating advanced technologies such as machine learning and web-based platforms.

IV. CONCLUSION AND FUTURE SCOPE

A. Conclusion

The Real-Time Object Detection System provides a simple, efficient, and user-friendly solution for assisting visually impaired individuals in navigating their surroundings. By using TensorFlow.js and the COCO-SSD model, the system is able to detect objects in real time and provide audio alerts

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

through the Web Speech API. The web-based implementation eliminates the need for complex installations and specialized hardware, making it easily accessible and cost-effective. The system successfully achieves its objectives by improving user awareness and safety. Overall, it demonstrates the practical application of machine learning and web technologies in assistive systems.

B. Future Scope

- The system can be further enhanced by incorporating advanced features such as:
- Integration of advanced models like YOLO for improved accuracy and speed
- Development of mobile application versions for better portability
- Addition of distance estimation and obstacle avoidance features
- Support for multilingual voice alerts for wider accessibility
- Detection of more objects and improved performance in low-light conditions

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 779–788.

[2] W. Liu et al., "SSD: Single Shot MultiBox Detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 21–37.

[3] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," Google Research, 2016.

[4] TensorFlow.js Team, "TensorFlow.js: Machine Learning for the Web and Beyond," Google Developers, 2018.

[5] A. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.

[6] MDN Web Docs, "Web Speech API," Mozilla Foundation, 2023. [Online]. Available: https://developer.mozilla.org

[7] COCO Dataset Consortium, "Microsoft COCO: Common Objects in Context," 2015.

[8] Ultralytics, "YOLO Documentation," 2024. [Online]. Available: https://docs.ultralytics.com

[9] OpenCV Team, "OpenCV Library," 2023. [Online]. Available: https://opencv.org

[10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[11] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed., Pearson, 2010.

[12] Google Developers, "TensorFlow Object Detection API," 2023. [Online]. Available: https://www.tensorflow.org

[13] Mozilla Foundation, "Web APIs Documentation," 2024. [Online]. Available: https://developer.mozilla.org

[14] GitHub Inc., "GitHub Documentation," 2024. [Online]. Available: https://docs.github.com

[15] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2011.