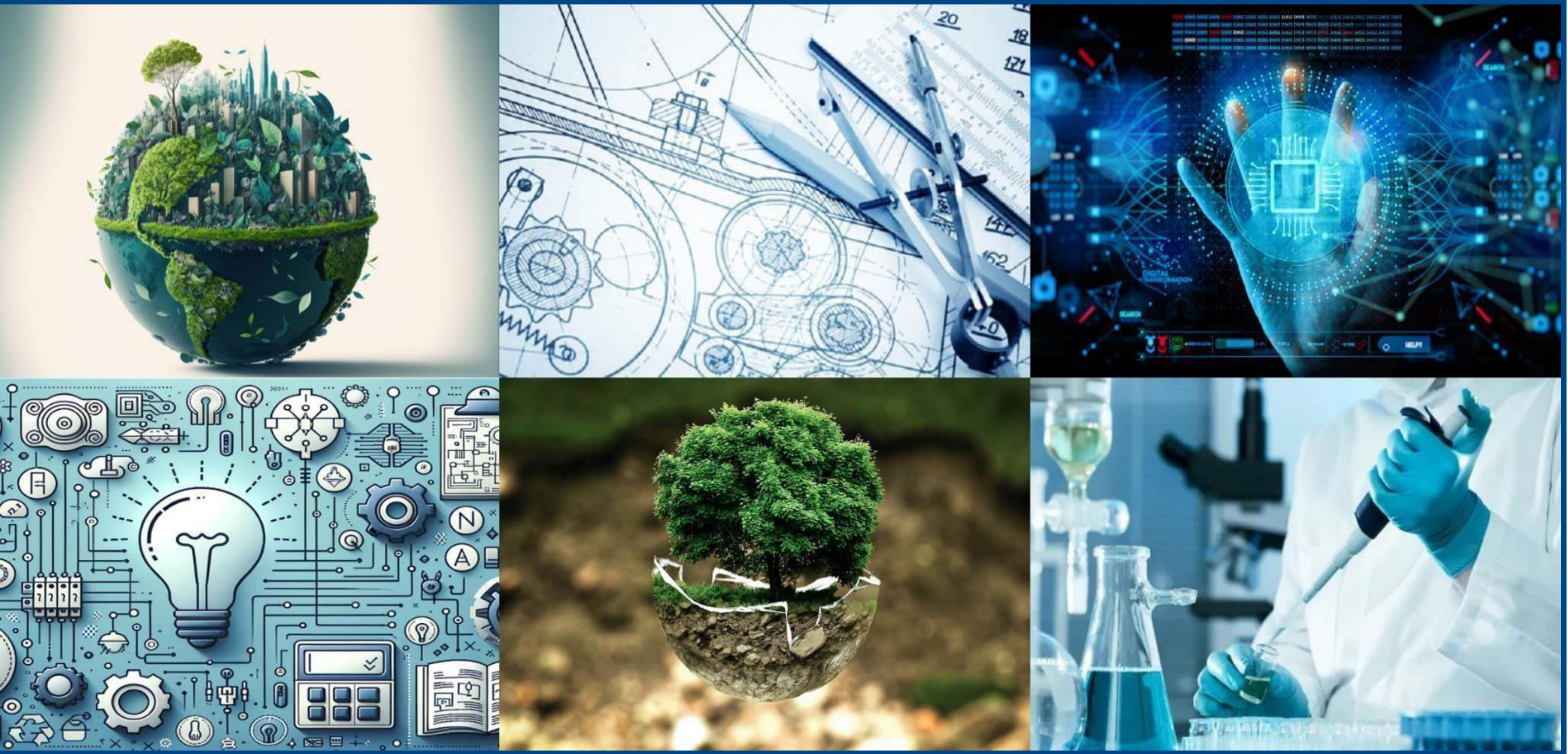
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808160

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VISIOTRACK: REAL-TIME OBJECT DETECTION AND DISTANCE ESTIMATION WITH VOICE GUIDANCE FOR THE VISUALLY IMPAIRED

Dr. Vidya Pol, Tejaswini G N

Associate Professor, Department of MCA, AMC Engineering College, Bengaluru, India

Student, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: The project “Visio track: real time object detection and distance estimation with voice guidance for the visually impaired” presents a real-time object detection system integrated with voice assistance using YOLOv5n. The YOLOv5n model accurately identifies and classifies objects in live video streams with high speed and efficiency. Detected objects are announced through a voice assistant, providing an interactive and accessible experience. The system is designed for applications in smart surveillance, assistive technology, and autonomous systems. It ensures low latency, high precision, and user-friendly operation.

KEYWORDS*:* real-time object detection, text-to-speech, voice assistant, visually impaired, deep learning, YOLOv5n

I. INTRODUCTION

This work aims to assist visually impaired persons who face a lot of hardship in recognizing objects. With advancements in technologies, object detection can be done using pre-trained models available. The work has utilized state of the art “You Only Look Once: Unified, Real-Time Object Detection” YOLOv5n algorithm to identify the object present before the person and COCO dataset is used to train this algorithm. The YOLO algorithm is a single-shot detector that analyses images utilizing a convolutional neural network (CNN). The objects detection has been done from real-time video taken from the webcam. The label of the object in the frame is recognized and then transformed into audio by using text to speech conversion which will be the anticipated output that can inform a blind person regarding the surrounds. This system provides users with real-time information about their surroundings in the form of voice to alert them and prevent any type of harm. Since object detection aids in the comprehension and analysis of scenes in pictures or videos, it is closely related to other related computer vision techniques like image recognition and image segmentation However, there are notable differences. While image segmentation provides a detailed pixel-by-pixel analysis of the components within a scene, image recognition merely assigns a class label to an identified object. Identifying object is distinct from these other jobs since it can locate items inside an image or video. We can then track such objects by counting them as a result.

II. LITERATURE SURVEY

Earlier also there have been attempts to identify and analyse the problem of object detection based on deep learning techniques like the Regions with CNN features model, developed by a team at Microsoft Research in the early 20th century which used a combination of region proposal algorithm and Convolutional neural networks (CNNs) to detect and localize objects. In recent years, numerous researchers have employed deep learning models, particularly convolutional neural networks (CNNs), to address object detection challenges, leading to the development of various state-of-the-art object detection models Two main categories can be used to classify the object detection models: two stage detectors such as R-CNN and gated R-NN Mask R-CNN and one-stage detectors such as YOLO SSD and YOLOR A two-stage detector operates in two distinct phases. Initially, it generates a set of candidate regions within the image. These regions are then processed in the second stage for object classification. The two-stage object detection methodology employs two rounds of analysis on the input image to precisely predict the presence and position of objects. In the first round, potential object locations are proposed, and in the second round, these proposals are refined to make final predictions. While the two-stage approach is generally regarded as more accurate than single-stage object detection, it is also more computationally intensive and time-consuming. The selection of the most suitable option hinges on some specific requirements and constraints of the particular use case. In general, real-world applications are better suited for single-shot object detection, while applications that value accuracy are better served by the two-shot approach. Additionally, object detection techniques can assist individuals with special needs, such as those who are visually impaired, in comprehending the content of images

EXISTING SYSTEM

There are already a number of assistive devices available to help people with visual impairments navigate their environment. To warn the user of items nearby, these systems frequently use ultrasonic sensors or simple obstacle detection. Without correctly determining the kind of item or its distance, the majority of conventional systems only offer a limited amount of information, such as the existence of an obstruction. Wearable cameras and computer vision are used by some sophisticated systems to identify objects, although many of them are pricy, large, or difficult to operate. They could also not be able to distinguish between different kinds of objects or estimate distance in real time. Even while these systems provide some help, more clever and affordable solutions are still required, as is mobility for those with vision impairments.

PROPOSED SYSTEM

A study of related work revealed that there are various existing models that do real time object detection with voice feedback but the main demerit of these models is that they use older algorithm for object detection like R-CNN, SSD, YOLOv3, YOLOv4, YOLOv5 or YOLOv6. The main drawback of these object detection algorithms is that the obtained accuracy and its real time speed does not go together, if the accuracy is good then the real time speed is slow and vice versa. The suggested approach has used YOLOv5n which is much faster and accurate because it uses a set of predefined boxes called anchor boxes and these boxes are with different aspect ratios, which are used to identify objects of different shapes, that helps to identify a broader range of objects compared to its previous versions, thus helping to reduce the number of false positive cases along with a better average precision. An outstanding feature of YOLOv5n is its exceptional computational efficiency, significantly surpassing other advanced object detection algorithms in image processing speed. Moreover, its capability to handle higher resolutions enables the detection of smaller objects with superior accuracy. Other than that, we have used Python3 for this work, the camera is initialized by using JavaScript and the camera starts capturing frames with the rate of 5 to 160 frames per second and feeds them to the algorithm. Then, the system employs YOLOv5n, trained on the COCO dataset, to recognize the object placed in front of the user. A Python library called gTTS is used to transform the recognized object into spoken words. gTTS acts as a bridge between your program and Google's text-to-speech service.

III. SYSTEM ARCHITECTURE

VisioTrack is an AI-powered assistive system developed to enhance mobility and environmental awareness for visually impaired individuals by providing real-time object detection and distance estimation. The system architecture integrates several functional modules to ensure accurate detection and effective user feedback.

The Input Module captures continuous video streams from a camera, which are then processed in the Pre-processing stage to enhance image quality and prepare data for detection. The processed frames are fed into the YOLOv10 deep learning model, which detects objects such as people, bottles, or mobile phones, and identifies their bounding boxes. The Post-processing module refines these detection results, calculates the approximate distance of each object from the user, and prioritizes relevant objects within a safe interaction range.

Detected object names are transmitted to the Voice Assistance Module, where a Text-to-Speech (TTS) engine converts the textual information into audible messages. These messages are then output through the Output Module, which consists of a speaker for real-time audio feedback and an optional display for showing bounding boxes. This dual feedback mechanism ensures that users receive timely and clear information about their surroundings, significantly improving navigation safety and independence.

The system’s modular design allows for scalability, enabling integration with advanced sensors or additional assistive features in the future.

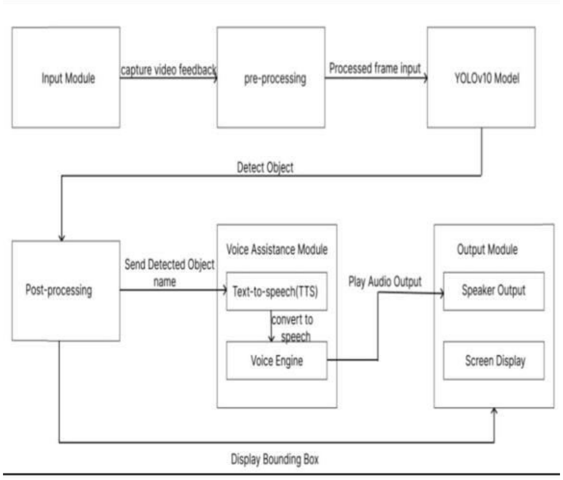


Fig 2.1 System Architecture

IV. METHODOLOGY

Modelling and analysis of object detection using YOLO (You Only Look Once) involve training a convolutional neural network to detect and classify multiple objects in an image in a single forward pass. YOLO divides the input image into a grid and predicts bounding boxes and class probabilities for each cell, enabling real-time performance. Its architecture balances speed and accuracy, making it ideal for applications like surveillance and autonomous driving. Training is done on annotated datasets using loss functions that consider objectless score, class probability, and bounding box regression. The Figure 2.1 architectural diagram represents a comprehensive system designed for real-time object detection with voice assistance and visual output, utilizing the YOLOv10 model. The process begins with the Input Module, which continuously captures video feed from a camera. These raw frames are then passed to the Preprocessing unit, where they are resized, normalized, and formatted to meet the input requirements of the YOLOv10 model. Once pre-processed, the frames are sent to the YOLOv10 Model, which performs object detection and returns data such as object class names, bounding box coordinates, and confidence scores. The output from YOLOv10 is then handled by the post-processing module, where redundant detections are as the filtered using techniques like non-maximum suppression, and the final detected object names and locations are extracted. The detected object names are simultaneously sent to the Voice Assistance Module, which converts them into spoken words using a Text-to-Speech (TTS) system and a Voice Engine. This audio output is routed to the Output Module, where it is played through a speaker, providing voice feedback to the user. Meanwhile, the bounding box and label information is also sent to the Output Module for screen display, allowing users to visually observe the detected objects on a display screen.

V. DESIGN AND IMPLEMENTATION

VisioTrack's design combines distance estimation and object detection into a small assistive device for people with vision impairments. A computing unit (such a Raspberry Pi or Jetson Nano), a stereo or monocular camera for visual input, and haptic actuators and audio output for user feedback are all part of the hardware configuration.

OUTCOME OF RESEARCH

The research led to the successful development of a system capable of detecting and identifying common objects such as people, bottles, and mobile phones using a camera and object detection algorithms. In addition to recognition, the system accurately estimates the distance between the user and detected objects in real time.

This outcome demonstrates the potential for computer vision to be effectively applied in assistive technologies for the visually impaired. By providing both object classification and distance measurement, the system offers improved spatial awareness and helps users safely navigate their environment. The solution is low-cost, efficient, and adaptable, making it a practical tool to enhance the independence and mobility of blind or visually impaired individuals.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. RESULT AND DISCUSSION

The implemented system was tested in various indoor and outdoor environments to evaluate its effectiveness in detecting objects and estimating their distance from the user

The distance estimation feature provided users with real-time feedback on how close or far the detected objects were. This proved to be especially useful for navigating obstacles and maintaining a safe distance. The system maintained consistent performance within a practical range (e.g., 0.5 to 3 meters), which is suitable for personal navigation.

the accuracy may vary depending on the angle of the object and the quality of the camera.

Overall, the results indicate that the system is effective for basic object recognition and distance estimation, showing strong potential as an assistive tool for visually impaired individuals. Future improvements could include better low-light performance, integration with audio feedback

Fig:(a) Actual Image

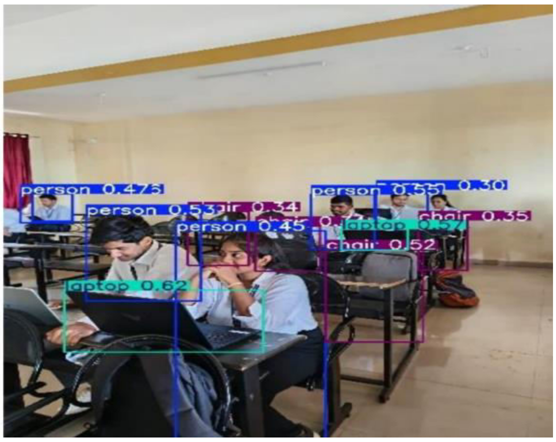


Fig:(b) Detected Image

VII. CONCLUSION

This work describes the development and implementation of an AI-based intelligent video surveillance system with the ability to detect and react to various real-time safety incidents. By interfacing a YOLO-based object detection platform with domain-specific modules for behavioral analysis, the system is able to identify important situations such as unusual movements, near misses, and unanticipated object or vehicle incidents without human involvement. Modular design makes every detection task run independently, enhancing flexibility, scalability, and maintainability in many different environments such as public institutions, health centres, and smart city infrastructures.

Performance tests show that the system attains high accuracy, low latency, and fast generation of alerts, confirming its real-time capability. The automated mail alert function also increases its value by facilitating fast reaction to perceived threats. In contrast to traditional systems that emphasize individual functions, this integrated approach facilitates concurrent multi-event monitoring using a single, streamlined pipeline.

Whilst performance under normal conditions is robust, there is still room for improvement under low-light and occluded scenes with potential to leverage in future using infrared or sensor fusion technology. The suggested framework as a whole makes an important contribution to intelligent surveillance by delivering a robust, deep learning-enabled system that further enables automated monitoring, extends situational awareness, and facilitates proactive safety management.

REFERENCES

1. S. Kumar, R. Patel, and A. Gupta. Real-time object detection and voice feedback for visually impaired using YOLOv5 and NLP. IEEE Transactions on Human-Machine Systems, 2023.
2. K. Tanaka, A. Smith, and R. Kumar. Adaptive YOLO-voice system for real-time marine debris detection. IEEE Journal of Oceanic Engineering, 2023.
3. Z. Liu, Q. Zhang, and Y. Chen. Privacy-preserving voice-activated object detection using YOLOv8
4. M. Leeuwen, E. P. Forking, W. Huizinga, J. Baan, and F. G. Heslinga. Toward versatile small object detection with temporal-YOLOv8.
5. Y. Qiu, Y. Lu, Y. Wang, and H. Jiang. IDOD-YOLOV Image-dehazing YOLOV5N for object detection in lowlight foggy traffic environments. Sensors, 2023
6. M. F. Musleh, M. A. Khan, A. Tariq, and M. Imran. Enhanced Real-Time Object Detection using YOLOv5n and MobileNetv3. Engineering, Technology & Applied Science Research, February 2025
7. M. Ahmed, T. Suzuki, and L. Zhou.
YOLO-Voice Net: Joint training of object detection and speech recognition for drone surveillance. IEEE Transactions on Multimedia, 2023.
8. A. Sharma, P. Nguyen, and D. Williams. YOLO-Voice: A low-cost assistive robot for elderly care using real-time detection and speech synthesis. IEEE Robotics and Automation Letters, 2024.
9. R. Singh, M. Fernández, and K. Yamamoto. Agri-Voice-YOLO: Real-time crop disease detection with voice feedback using YOLOv9. IEEE Transactions on Agri Food Electronics,

INNO SPACE
SJIF Scientific Journal Impact Factor

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com