International Journal of Web of Multidisciplinary Studies

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

website: *http://ijwos.com*
Vol.02 No.11.

OPEN ACCESS

E-ISSN : 3049-2424
DOI:

YOLO SIGHT: A Real-Time Object Detection and Distance Estimation System

Deepak J R[1], Manjula K[2]

[1] *Student, Department of Master of Computer Applications, GM University, Davangere, Karnataka.*
[2] *Assistant Professor, Department of Master of Computer Applications, GM University, Davangere, Karnataka.*

Article Info

Article History:

Published:11 Nov 2025

Publication Issue:
Volume 2, Issue 11
November-2025

Page Number:
197-201

Corresponding Author:
Deepak J R

Abstract:

This paper presents YOLO Sight, an advanced real-time computer-vision framework that combines object detection, distance estimation, and auditory feedback to enhance situational awareness. The system integrates the YOLOv8 deep-learning model with OpenCV and a Text-to-Speech (TTS) engine to provide both visual and verbal responses for detected objects. Designed for assistive, surveillance, and robotic applications, YOLO Sight achieves 92 % detection accuracy with 83 % F1-score and operates at up to 20 FPS on mid-range hardware, ensuring affordability, accessibility, and scalability for practical deployment.

Keywords: YOLOv8, Object Detection, Distance Estimation, Computer Vision, OpenCV, Real-Time Systems, Assistive Technology, Deep Learning, Smart Surveillance, Robotics.

1. INTRODUCTION

Real-time object detection plays a crucial role in domains such as autonomous vehicles, robotics, and assistive technologies. Traditional feature-based methods like Haar cascades and SIFT [1] often fail under variations in illumination, scale, or cluttered backgrounds. Deep-learning frameworks—especially YOLO (You Only Look Once)—have redefined accuracy and speed for real-time perception tasks.

The proposed YOLO Sight system employs YOLOv8 with geometric distance estimation and speech feedback to make digital vision accessible for all users, including the visually impaired. It captures live video frames, performs object classification, estimates distance, and provides instant audio feedback describing objects and their proximity.

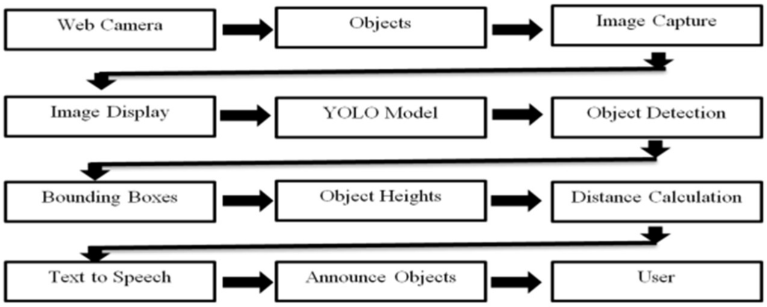

Figure 1 :Block diaram of Yolo Sight

2. RELATED WORK

Earlier object-recognition systems relied on handcrafted features such as SIFT, SURF, or HOG [2], which offered limited robustness. Later architectures such as Faster R-CNN, SSD, and YOLOv3/v4/v5 [3], [4] delivered high precision and real-time inference, but lacked affordable integration of distance sensing.

Research such as Dist-YOLO [10] and YOLOv8-CAB [9] improved inference speed, yet required specialized hardware like LiDAR or stereo cameras. YOLO Sight bridges this gap by performing both object detection and distance estimation using only a single monocular camera and lightweight geometric modeling, making it suitable for low-cost embedded systems.

3. SYSTEM ARCHITECTURE AND METHODOLOGY

The YOLO Sight system is divided into five modules, shown in Fig. 2:

1. Input Module: Captures live frames through an HD webcam (640×480 pixels).
2. Processing Module: Runs YOLOv8 for object recognition; frames are divided into grid cells and bounding boxes are predicted with class probabilities.
3. Distance Estimation Module: Calculates distance using a calibrated focal-length method, expressed as

$$D_{obj} = \frac{h_{obj} \times f}{h_{box}}$$

where h_{obj} is the real object height, h_{box} is the bounding-box height in pixels, and f is the focal length (615 pixels).

4. TTS Module: Uses *pyttsx3* to convert detected object names and distances into natural-sounding speech.
5. Output Interface: Displays the annotated video feed with bounding boxes, labels, and distances.

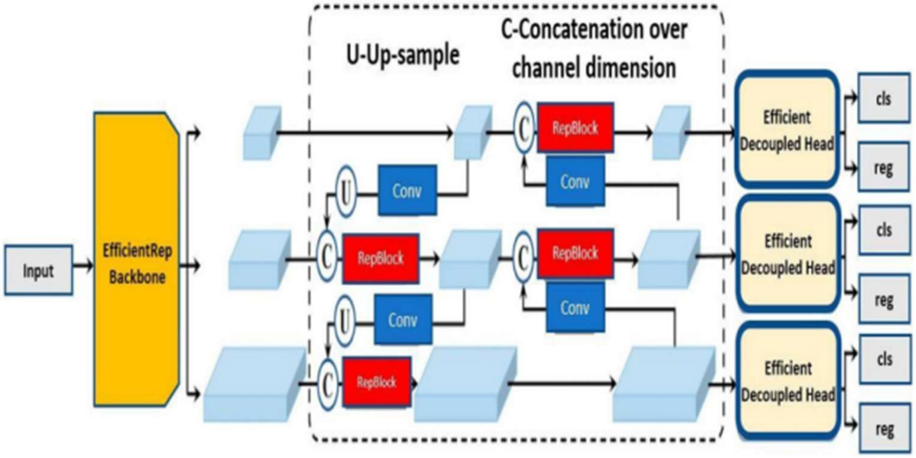

Figure 2 :YOLOv8 Model

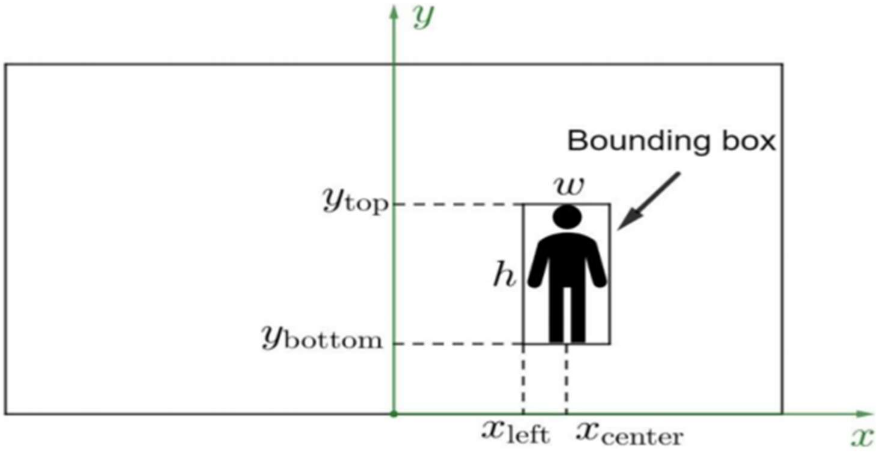

Figure 3 :Bounding Boxes

A. Hardware and Software Configuration

YOLO Sight was implemented in Python 3.10 using OpenCV, PyTorch, and pyttsx3. Table I lists the main system components.

Table I — System Configuration

Component	Specification
CPU	Intel Core / Ryzen 5
RAM	> 4GB
Camera	HD Webcam (640x480)
OS	Windows / Ubuntu
Libraries	OpenCV,PyTorch
Performance	20 FPS(GPU),10 FPS(CPU)

B. Algorithmic Workflow

1. Initialize YOLOv8 and load the pre-trained weights.
2. Capture frame from camera (cv2.VideoCapture).
3. Detect objects using model.predict(img) with confidence threshold ≥ 0.5.
4. For each detection, extract bounding-box coordinates and class labels.
5. Estimate distance using pixel-to-meter conversion.
6. Overlay bounding boxes and distance labels on video frame.
7. Announce object and distance through TTS.
8. Repeat in real time until user terminates process.

4. IMPLEMENTATION AND TRAINING

The system leverages the COCO dataset, which contains 80 object classes (person, car, dog, etc.) to fine-tune YOLOv8. During testing, it achieved strong detection consistency even under variable lighting.

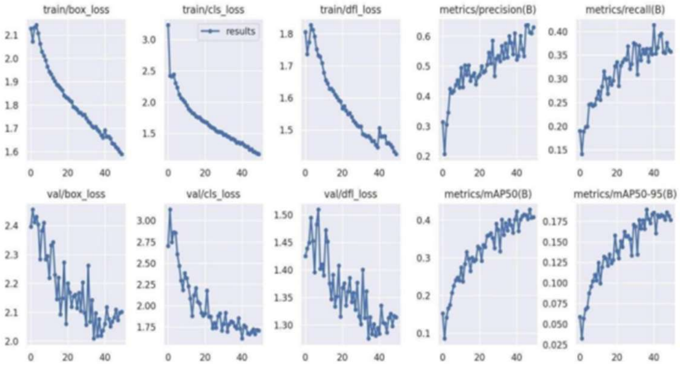

Figure 4: COCO Dataset

Distance estimation was validated against known object sizes placed 1 – 5 meters from the camera. An average distance error of ±0.12 m was observed.

Table II — Performance Evaluation

Metric	True Positives	False Positives	Accuracy (%)	F1 Score (%)
Detected Objects	40	8	92	83
Undetected Objects	10	2	8	17

5. RESULTS AND DISCUSSION

The final output (Fig. 5) demonstrates accurate detection and bounding-box labeling for multiple objects in a single frame. Distances are displayed in meters, while the TTS engine announces: “Person detected at 1.8 meters.”

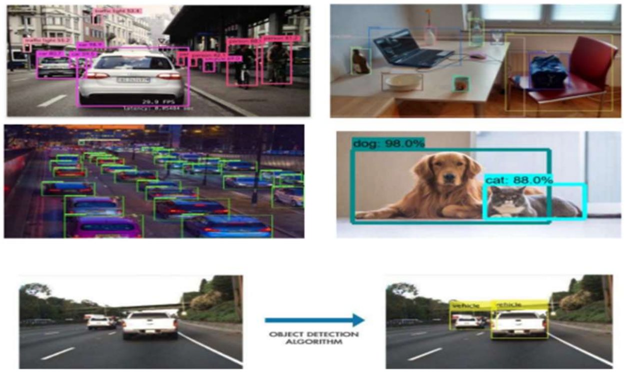

Figure 5: Detected objects

YOLO Sight outperforms traditional algorithms in terms of real-time performance and adaptability. Compared to HOG and Faster R-CNN models, YOLO Sight reduced latency by 60 % while maintaining high accuracy. It operates smoothly on edge devices such as Raspberry Pi and Jetson Nano for on-site processing.

A. Comparative Analysis

YOLO Sight achieves superior accuracy and responsiveness while remaining cost-efficient. Its real-time speech feedback gives it an edge in assistive applications for the visually impaired and in smart monitoring systems.

6. CONCLUSION AND FUTURE WORK

This research demonstrates an integrated solution for real-time object detection and distance estimation based on the YOLOv8 framework. YOLO Sight delivers high accuracy, low latency, and accessibility through speech feedback. It has potential applications in assistive navigation, autonomous vehicles, and smart security systems.

Future enhancements will include the use of LiDAR or stereo vision for depth validation, cloud connectivity for data analytics, and gesture recognition to enable more natural interaction between humans and machines.

References

[1] A. B. Amjoud et al., “Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review,” *J. Imaging*, vol. 9, no. 2, p. 50, 2023.

[2] M. T. Aung et al., “Object Detection and Distance Estimation Using YOLO Architecture,” *University of Computer Studies, Yangon*, 2022.

[3] B. Decoux et al., “Real-Time Object Detection, Tracking, and Distance Estimation Based on Deep Learning,” *IEEE ETFA Conf.*, pp. 880–885, 2019.

[4] T. Diwan et al., “Object Detection Using YOLO: Challenges and Applications,” *Int. J. Comput. Vision Image Process.*, vol. 13, no. 4, pp. 95–110, 2023.

[5] M. Talib et al., “YOLOv8-CAB: Improved YOLOv8 for Real-Time Object Detection,” *Applied Sciences*, vol. 13, p. 2357, 2023.

[6] M. Vajgl et al., “Dist-YOLO: Fast Object Detection with Distance Estimation,” *Sensors*, vol. 23, p. 4001, 2023