Available online @ https://jjem.jnnce.ac.in
https: www.doi.org/10.37314/JJEM.SP0275
Indexed in International Scientific Indiexing (ISI)
Impact factor: 1.395 for 2021-22
Published on: 08 December 2024

Smart Sight: Real-Time Object Detection and Distance Estimation

Jeevan U M[1*], Manjunatha H T[2]

[1*]Student, Department of MCA, Jawaharlal Nehru New College of Engineering, Shivamogga, India.

[2]Assistant Professor, Department of MCA, Jawaharlal Nehru New College of Engineering, Shivamogga, India

jeevanum23@gmail.com, manjudeepa@jnnce.ac.in

Abstract

Proposed methodology describes real time object detection and distance estimation system, designed to enhance user interaction and accessibility across various applications. YOLO which stands for you only look once is a well-known high-speed, highly accurate object detection model. The more recent model- YOLO v8 has gone a step further by improving the architecture and how it's trained thereby making it very effective especially when it comes to real time activities; on this part you can incorporate say an experiment related to recognition systems etc., another dataset so that people with some experience will be able to understand everything perfectly. Smart sight makes it possible for you to identify any object in front of the camera, similarly it relates to the distance using similar triangle formula. The system supports a variety of object categories such as people, animals; vehicle among others also provides real time spoken alerts with objects found and their distances. Due to this modular structure which operates on different platforms, it cannot only grow quickly in size, but can also be incorporated in many devices.

Keywords: *Similar Triangles Formula, Camera-based System, Surveillance Systems, Assistive Technology, Spatial Awareness, Custom Object Classes.*

1. Introduction

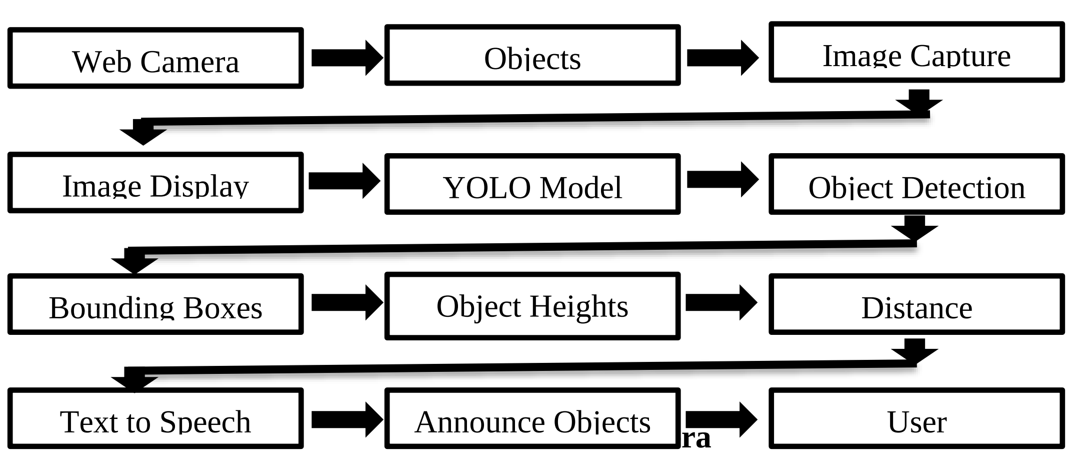
Computer vision needs object detection which supports various modes for recognizing and classifying objects in pictures or video streams. In order to improve interaction and functionality in multiple applications, systems are enabled to perceive and interpret the environment around them. In terms of speed and accuracy, the YOLO (You Only Look Once) family stands out among several object detection algorithms". Using a single pass, YOLO algorithms process images such that they achieve real-time performance, which is essential in applications that urgently require certain results This results in increased accuracy and efficiency of YOLO v8, which becomes more preferred for complicated real-time detection assignments. Therefore, we introduce Smart Sight, a complex object detection combined with distance measuring system aimed at improving user experience along with usability in different application areas. Combining the YOLOv8n model with Open CV or image processing with Pyttsx3, so that Smart Sight has an integrated system that recognizes objects and produces audible feedback. The system identifies and classifies objects within the camera's view and estimates their distances using the similar triangles formula. Smart Sight, which provides real-time voice announcements of detected objects and their distance, can identify individuals, vehicles, and animals and also everyday items among others.

2. Related Works

The related work has been reviewing various methodologies and techniques includes. Kharade et al [6] came up with the name "Virtual Voice Assistant Sara" in order to make it workable on computers when handling voice commands. This system though useful on a PC only does not do any sort of real time object detection unlike some others such as Phonesense developed by Khoury and Vorotnikovs, 2021 or distance assessments necessary for moving around physical spaces involving human beings. [5] Jones (1995) devised methods of computing the geometric distance between a point and a triangle in three-dimensional space the article "3D Distance from a point to a Triangle," The discussion focuses on theoretical geometric calculations and excludes practical issues such as object recognition and distance estimation that are important in improving mobility of the blind. [8] In his article, "Real-time Object Detection Using Deep Learning," Shaik (2023) has created a strategy that utilizes complex neural network ideas to enable one object to recognize others nearby. Decoux et al [3] in their work on" Real Time Object Detection, Tracking, Distance and Motion Estimation using Deep Learning" considered this particular research suitable for smart motion occluding for automobiles, and caution that it might require adaptations to serve as a daily navigation aid for the blind.". Rajesh et al [7] introduced a "Camcorder for Transportable Text to Speech Conversions". It assists in reading through converting camera captured text into voice. Vajgl et al [10] presented the "Dist-YOLO" model that improves YOLO by incorporating fast object detection and distance estimation. Talib et al [9] described, YOLOv8 was advanced by them into YOLOv8-CAB to make it work faster during the actual organization time for detecting objects. The betterment achieved through these changes does not deal with particular aspects of visually challenged people performing usual duties." [2] Aung (2022) integrates distance estimation into YOLO based detection in the article "Object Detection and Distance Estimation Using YOLO Architecture" But for real-time responsive systems, this project might not be completely optimized neither for particular applications like ones considered in Smart Sight. [1]," Amjoud and Amrouch (2023) looked at the progress made with deep learning towards enhancing object detection in their review article titled "Object Detection Using Deep Learning, CNNs, and Vision Transformers.

3. Methodology

The below block diagram describes the working process of Smart Sight. Since it is the real time project, it starts with webcam and displays the video output to the user.



Smart Sight Project mainly uses web cameras for photography since they are affordable and sold globally hence easy to access. Images can be sent in proper time when need arises using web cameras. Therefore, a web camera must be set to keep taking crucial photographs or frames for video at certain intervals before such content undergoes further processing in the specified order.

3.2. Objects

The Smart Sight system refers to any detectable items within the camera's field of view as objects. Objects present a major problem for object detection algorithms since they are so different from one another in terms of what they are; we therefore need strong flexible models that can detect and classify objects on them no matter how difficult the conditions become, particularly as they relate to illumination levels.

3.3. Image Capture

The web camera continually streams frames and for image capture the process entails extracting individual form that stream. The environment snapshot is taken at a specific point of time for every image captured. Object detection depends on this step because it serves as input data in its original form. Efficient systems for capturing images using devices should have been considered too though since they require both high-quality images and a bit rate to display them constantly in their real forms.

3.4. Image Display

The purpose of this step is providing immediate visual feedback confirming how well detection system functions object. It is also important that the display is clear and that it changes in time. Besides, it is necessary for this step to represent bounding boxes as well as labels around recognized objects so that it would be easy for users to get clear picture of what was found out by the system.

3.5. YOLOv8 Model

The Smart Sight project relies on the YOLO (You Only Look Once) model for object detection in images captured. The latest version of this model, YOLOv8, is famous because of its perfect balance of accuracy with high speed. YOLO divides an image into a grid and predicts bounding boxes and probabilities for each region

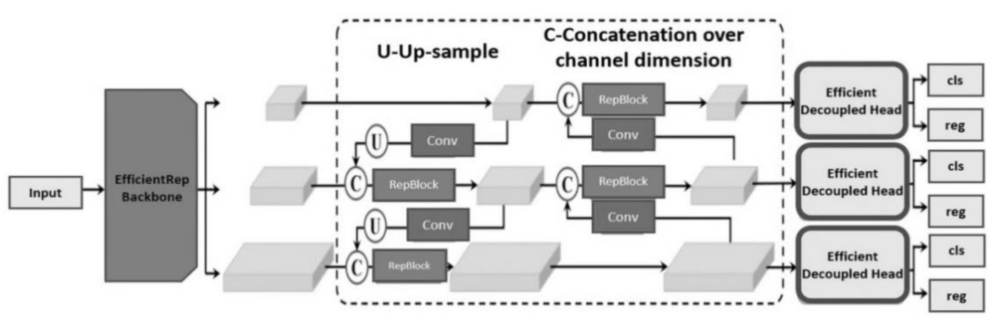


Figure 2: YOLOv8 Model

3.6. Object Detection

The core feature of the Smart Sight system is object detection where objects found in each frame are identified and located using YOLOv8 model. There are some aspects that are crucial during the detection phase such as pre-processing the image, using the YOLO model to make detections, and performing post-processing to remove detections with low confidence and rectify for regression and various occlusions such as translation or rotation in the object.

3.7. Bounding Boxes

Rectangular ovals called bounding boxes are useful for identifying items within pictures shown as graphs through shapes too as for showing how huge in size these detected items are situated. In the below figure 3.7 each one of these boxes consists out of two numbers representing position (x, y) and size (width, height) thereby representing whatever thing you detect on it so that we can see where it stands at first sight. Usually, they have category names alongside them plus a probability score that this is actually true.

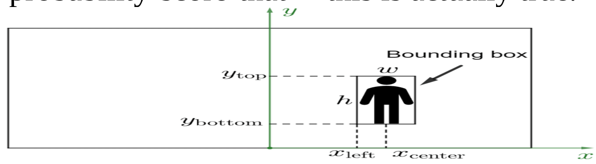


Figure 3 Bounding Boxes

3.8. Object Heights

Being able to figure out its distance from the enveloping (1,1) items however advanced this facility of Smart Sight system is in an effort towards understanding the environment. Techniques used to measure the height of objects include geometrical calculations that are based on the dimensionality of bounding boxes as well as parameters that are known about the camera setup. Accurate heights play a major role in the field of navigation where knowledge of one's terrain matter most.

3.9 Distance Calculation

One other important thing in Smart Sight project is calculation of the distance towards detected objects which allow users to determine their cameras' closeness to various items. As in its part accurate determination of distances has been claimed crucial for any usage that interacts with things or moves among them so that one has better positioned spatially.

$$D_{obj} = h_{obj} / h_{box} \times d_{box} \ \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where, D_{obj} is the normalized measurement of the object.

h_{obj} represents the height of the detected object.

h_{box} denotes the height of the bounding box.

d_{box} signifies a reference distance or dimension associated with the bounding box.

3.10. Text to Speech

In the Smart Sight system, the TTS technology converts found object textual information into audible speech, making it convenient and suitable for visually disabled individuals. The below figure 3.10 describes Pyttsx3 is the TTS library chosen for its simplicity, multilingual nature and independence from the internet because of its ability to work both online and off. It can be used for pronouncing sounds that are naturally sounding by being tuned in terms of speaking speed, loudness and pitch.

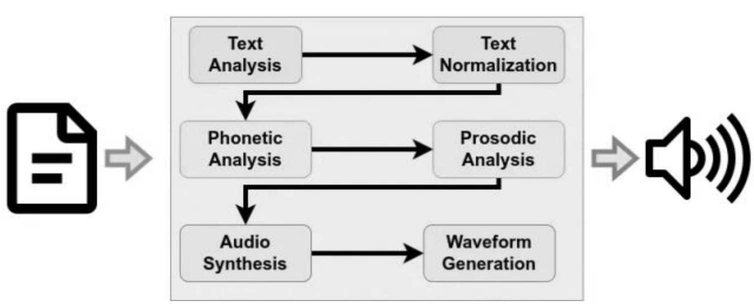


Figure 4: Text To Speech Using pyttsx3

3.11. Announce Objects

Integrating the object detection system with the text-to-speech functionality provided by pyttsx3 is how objects are announced in AI. When an object is detected and recognized, a descriptive message is created by the system and then sent to the user through TTS. Such an announcement usually contains the object's name, location besides may be the distance or even how big the object is.

3.12. User

The system aims at meeting their needs by provision of interactive and interpretable tools that allow them to effectively interact with, as well as understand their environment.

4. Experimental Results

The YOLOv8 model comes with pre-trained dataset such as COCO. This dataset provides a large base of object categories and enables to perform effective annotating huge range of object. The below table 4.1 describes these criteria indicate how well the model is actually able to find and place stuff on top of that about how much it can tell us which is true or not. As for future improvements reducing false positives and enhancing detection in more complicated situations may be targeted because they raise these metrics further.

Results (Objects)	True Positives (TP)	False Positives (FP)	True Negatives (TN)	False Negatives (FN)	Accuracy (%)	F1 Score (%)
Detected	40	8	N/A	9	92	83
Undetected	10	2	N/A	1	8	17

Table 1: Confusion Matrix of Performance Test

	Year	Algorithm Used	Limitations	Accuracy Percentage
[1]	2023	CNNs, Vision Transformers	Limited to image datasets, not real-time;	Not specified
[2]	2022	YOLO Architecture	Challenges in detecting small and occluded objects; less accurate in dynamic environments.	78% (for general scenarios)
[3]	2019	Deep Learning-based methods (YOLOv3)	Performance drops in low-light conditions; requires high computational power.	86%
[4]	2023	YOLO (various versions)	Difficulty in detecting small and fast-moving objects	81% (varies by YOLO)
[5]	1995	3D Geometric Algorithms	Focuses on 3D distance computation rather than real-time detection;	Not applicable
[6]	2023	Text-to-Speech Systems	Limited to desktop applications; requires predefined objects and environments.	Not specified
[7]	2023	Camera-based Text-to-Speech	Focuses on text detection rather than object detection;	Not specified
[8]	2023	Deep Learning (YOLOv5)	difficulty in detecting very small or overlapping objects.	89%
[9]	2023	YOLOv8-CAB	High sensitivity to changes in lighting and occlusions;	90%
[10]	2023	Dist-YOLO	Lower accuracy at larger distances;	88%

Smart Sight	2024	YOLOv8, OpenCV	Issues with detecting small, distant, and overlapping objects	92%

Table 2: Comparative Analysis

The Table 2 comparative analysis of the existing methodologies in which, the accuracy of Smart Sight is more than the below existing systems.The detected object is as shown in the figure 5

Figure 5: Detected Objects

together with recognition of their position in space is revealed. This technology could be further developed so that it becomes more precise while at once enhancing its performance and expanding functionality to meet different user requirements thereby making it practical and user friendly. Situational awareness needs, the system has also been coded using pyttsx3 so that they can listen rather than depend on visuals for information hence enhancing such awareness without any need for vision.

5. Conclusion

Through Smart Sight project uses the YOLOv8 model for real time object detects and achieves about 92% accuracy rate in COCO dataset. When it comes to visually impaired people’s objects so as to improve their perception and engagement with their surroundings. geometrical calculations used for estimating distance along with continuous webcam streams utilized in acquiring data continually from the environment around us; the practicality of Smart Sight in detecting of objects that are moving with high precision. The primary importance of this visual annotation is to help users to readily identify

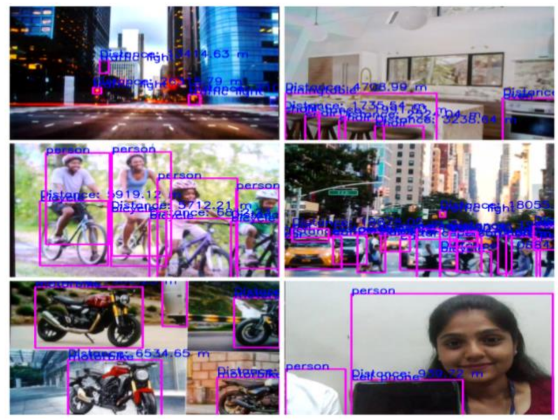


Figure 6: Detected Objects

Figure 6 shows the Smart Sight system working to highlight the objects within the view of the camera by encircling them with boxes.

References

1. Amjoud, A. B. Object Detection Using Deep Learning, CNNs and Vision

Transformers: A Review. Journal of Imaging, 9(2), 50. DOI: 10.3390/jimaging9020050.
2. Aung, M. T Object Detection and Distance Estimation Using YOLO Architecture. Dissertation, University of Computer Studies, Yangon. ResearchGate.
3. Decoux, B. Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility. 2019 IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 880-885. DOI: 10.1109/EST.2019.8806222.
4. Diwan, T., Anirudh, G. Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications. International Journal of Computer Vision and Image Processing, 13(4), 95-110. DOI: 10.4018/IJCVIP.20231001.
5. Jones, M. W ,3D Distance from a Point to a Triangle. Computer Graphics Forum, 14(3), 145-150. ResearchGate.
6. Kharade, S. and et al. (2023). Virtual Voice Assistant Sara for the Handicap and Blind People to Use Desktop. International Journal of Computer Applications, 175(5), 12-16. ResearchGate.
7. Rajesh, S, Transportable Camera Based Text to Speech Converter. International Journal of Engineering Research & Technology (IJERT), 12(6), 23-28. ResearchGate
8. Shaik, S, Real-time Object Detection Using Deep Learning.
Journal of Advances in Mathematics and Computer Science, 38(8), 45-56. DOI: 10.9734/jamcs/2023/v38i81787.
9. Talib, M, YOLOv8-CAB: Improved YOLOv8 for Real-time Object Detection. Applied Sciences, 13(5), 2357. ResearchGate.
10. Vajgl, M, Dist-YOLO: Fast Object Detection with Distance Estimation. Sensors, 23(8), 4001. DOI: 10.3390/s23084001