Object Detection Tracking and Alert System for Visually Impaired Persons

M. Joshna[1], J. Tharun Kumar[2], G. Yuvaraju[2], K. Vishnu Vardhan Reddy[2] and V. Sai Prathap Reddy[2]
[1]*Assistant Professor Department of AI&DS, Annamacharya University, Rajampet, Andhra Pradesh, India*
[2]*Department of AI&DS, Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh, India*

Keywords: Object Detection, Computer Vision, Speech Synthesis, SSD (Single Shot MultiBox Detector), MobileNet V3, Visually Impaired, Real-Time Processing, COCO Dataset, YOLO (You Only Look once), Faster-R CNN (Region-Based Convolutional Neural Networks).

Abstract: This project helps the blind in their navigation using advanced computer vision and speech synthesis technology. Based on the COCO dataset and a combination of SSD MobileNet V3 model with OpenCV's DNN module, the system detects the mentioned objects in real-time from webcam feeds. The identified objects are marked with bounding boxes with tags and pyttsx3 is used to convert these text tags to speech alerts, allowing real-time auditory indicators of what the user's surroundings are. The SSD MobileNet V3 model was selected for use in this project, as it represented the optimal trade-off between accuracy and computation – perfect for implementation in real-time applications on edge devices. They use the techniques of object detection and speech synthesis. It applies the detection algorithm on every frame of the video, draws a box around the detected object on the frame and activates a speech alert with the detected object name by utilizing pyttsx3. It also offers users up-to-date, succinct information on objects in their environment and improving their situational awareness and security.

1 INTRODUCTION

This product caters to the high mobility needs of the visually impaired person. (Ahn, et al, 2023) Utilizing cutting-edge computer vision techniques, the system detects and tracks objects in real-time. and offers auditory feedback that allows users to navigate safely. (Yang, et al, 2023) This new solution improves situation alertness by identifying known items as well as hurdles, delivering brief, accurate audio feedback for the purpose of overwhelming the user. (Liam., et al 2022) At the heart of the system is a trained object detection model and a robust tracking algorithm, continuously monitoring out of the environment. The (Hou et al, 2019) system is simple in design, portable and very light, allowing it to be used over and over. It is also inexpensive through cheap hardware parts, allowing the machine to reach more people. This project seeks to build an access assistive tech that empowers them increases mobility and independence of visually impaired people. (Wang, G., et al, 2018) By class for detection and tracking move on to be a helpful and efficient guide for day–to–day navigating.

2 LITERATURE SURVEY

Within their advancing tech of assistive techs for the blind people, a substantial improvement in wearable techs, machine learning, and computer vision has taken a notable progress. Ahmad et al. Object detection and tracking are essential for the design of assistive navigation systems, according to (2023). Object detection has been largely improved by deep learning, especially with the use of deep neural networks like Convolutional Neural Networks (CNNs). Deep learning, especially Convolutional Neural Networks (CNNs), has significantly enhanced the accuracy and effectiveness of object detection.

YOLO and SSD models have been the gold standard in real- time object detection. Khoshboresh-Masouleh and Shah- Hosseini (2022) solve target detection and anomaly in sequential drone imagery using state-of-the-art deep few-shot learning. It is unique in that it addresses the problem of small training sets, a common real-world limitation. Wang et al. (2020) offer a comprehensive review of person re- identification (Re-ID) methods, highlighting the advancements and challenges of the field. They

Joshna, M., Kumar, J. T., Yuvaraju, G., Reddy, K. V. V. and Reddy, V. S. P.
Object Detection Tracking and Alert System for Visually Impaired Persons.
DOI: 10.5220/0013876700004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 2*, pages 45-49
ISBN: 978-989-758-777-1

categorize different Re-ID methods based on the underlying methodology, from deep learning, metric learning, and transfer learning.

This research highlights the robustness of Re-ID systems, particularly in difficult and heterogeneous environments, and is a valuable contribution.

Annotations are necessary for preprocessing because they contain the labels and bounding boxes used to indicate points of interest in the pictures.

Time series are developed for video data or footage to keep track of the movement of an object over time as this is essential to allow for tracking to take place. to the security and surveillance industry. Chen et al. (2019) provide a thorough survey of applying deep learning approaches to multiple object tracking (MOT). These textbooks span the transition from traditional approaches to modern deep learning-based systems. They cover key challenges such as occlusion, intra-class variation and real-time processing in detail. This thorough survey is critical to grasping MOT progress and current issues. Hoffmann et al. (2021) present a system for the detection and tracking of objects in real time which has been specifically designed for autonomous vehicles. Such studies are very crucial in improving the reliability and safety of autonomous technology, as they focus on accuracy and adaptability across different driving environments. This research will play a significant role in the development of technology based on autonomous vehicles as the adaptive approach can be used for efficient operation across heterogeneous environments. The labelled dataset is split into training (to train the model), validation (to optimize the model), and test (to test the accuracy of the model) subsets.

3 DATA COLLECTION AND PREPROCESSING

To create any project first it begins with a great data collecting process. This process includes gathering numerous images and videos from different settings. These domain types leverage a collection of everyday scenarios in both indoor domains such as homes, offices, or public squares, and outdoor domains such as streets, parks, and public transportation facilities. To operate in any environment at any time, it is not only necessary to record these scenes but also to do so in different lighting conditions, such as day, twilight, and night. You require objects such as furniture, doorways, vehicles and so on, and poles and benches to navigate safely. Using a collection of sources will enhance diversity and completeness of the data collection. Open access datasets such as Common Objects in Context (COCO) provide a large resource of labeled images. Moreover, cameras can record customized border data based on the user-specific requirements and conditions of visually impaired users. Data contributions — Collaboration with accessibility communities can be beneficial in provising useful extent of data contributions. After collection, raw data undergoes a heap of preprocessing to prepare it to train models. It starts with cleaning the dataset, removing common problems like missing annotations, duplicate entries, and low- quality images. Some examples of the data augmentation approaches are rotation, scaling, cropping, etc., these methods artificially increase the size of a dataset by forming new content with better variability of a dataset, which provides better robustness to the model. Pixel values from different images.

4 EXISTING WORK

This project is based on some existing models. One of the leading models in this category is the YOLO model that is known for its object detection capability in real time. YOLO gets its strengths from its ability to detect lots of objects in one image at a fast pace and high accuracy a trait that suits dynamic environments where a visually impaired person is located. The second popular model is Faster R-CNN and is one of the top performers for accuracy by applying region proposal networks to detect the objects. It's also more reliant on computing power than YOLO, but its accuracy makes it mandatory for cases where you wish to accurately detect objects. These aside, SSD (Single Shot MultiBox Detector) is the one which form the perfect balance between speed and accuracy, combining the speed of YOLO and accuracy of Faster R-CNN. This is done by predicting class scores and bounding boxes at once directly from feature maps, which makes the object detection process faster.

5 PROPOSED METHOD

YOLOv8 is designed to perform high-accuracy object detection in real time, placed within live video streams and images. YOLOv8 (You Only Look Once version 8) model is a state-of-the-art object-detection system that is tested to be efficient and accurate. The system begins with an input layer that parses images

from a variety of sources, such as live streams, uploaded video, and so on.In this step, the images go through pre-processing steps (like resizing, and normalization) to make them compatible with what the YOLOv8 model requires. The backbone network stands at the heart of the YOLOv8 system and it is essentially a stack of advanced convolutional neural networks (CNNs). The backbone employs residual blocks and a number of layers of convolution to see complex patterns and tricks in the images. Then the YOLOv8 detection head will convert these feature maps into predicting the bounding box of objects, categories and their probability scores.

6 METHODOLOGY

The project serves as a systematic endeavor to facilitate increased mobility and independence for the visually challenged individuals. Leveraging advanced computer vision algorithms for real-time object recognition and tracking along with auditory conjoined notifications to provide users with timely and intuitive information about their environment.

- **Principles:** Real-time object detection and tracking are achieved using the YOLOv8 model, which is quick and highly accurate in handling multiple objects simultaneously. Auditory feedback is also a central feature, whereby detected objects are converted to voice alerts using pyttsx3.
- **Data Preprocessing:** In below Figure 1, Preprocessing of data starts with collecting a diverse dataset of images and videos from various environments, followed by cleaning the dataset to address issues like missing annotations and poor- quality images. Techniques like rotation, scaling, and cropping are applied for data augmentation, while normalization and resizing ensure consistent pixel values and model compatibility.

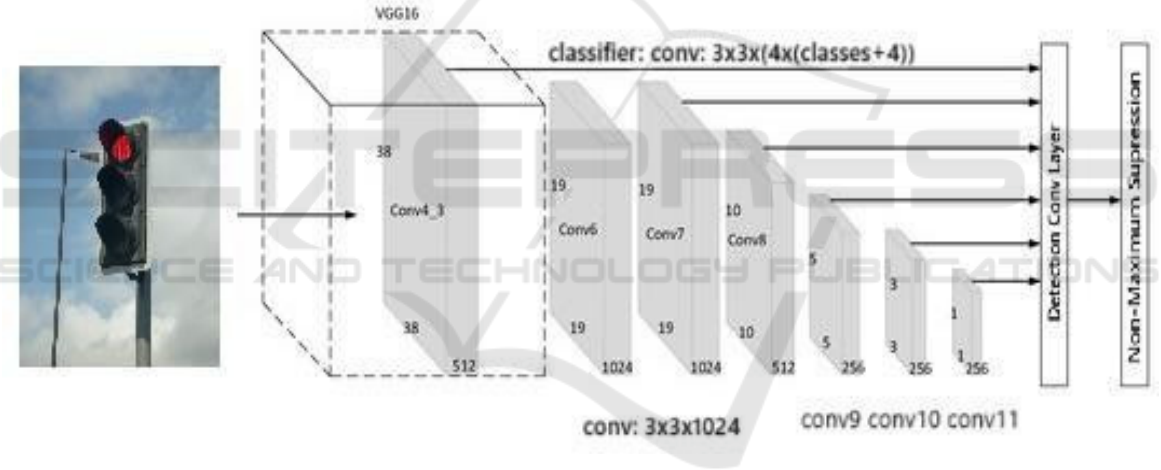

Figure 1: System Architecture.

- **Feature Selection:** The YOLOv8 model's backbone network, comprising advanced convolutional neural networks (CNNs), extracts detailed feature maps from input images.
- **Model Training:** The processed data is divided into training, validation, and testing sets, ensuring a structured approach to model development. The YOLOv8 framework is used to train the object detection model, optimizing both speed and accuracy.
- **Model Validation:** A separate validation set is utilized to adjust and optimize the model and prevent overfitting, ensuring it generalizes well to new data. Performance metrics are used to evaluate the model, while the testing set assesses its effectiveness in real-world scenarios.

7 RESULT AND ANALYSIS

The project exhibits promising results in increasing freedom of movement and self -reliance for blind individuals. The system effectively detects objects in real time and tracks, providing users a timely hearing alert. Through rigorous testing, the model shows high accuracy in identifying various objects and obstacles in both indoor and external environment.

Figure 2: Object Detection.

Figure 2 Displays a cellular smartphone held by a person. Captured at 1:33. Stepping aside from the object detection part, the model analyzed the image and gave results of the individual with 72.59% confidence and that of a smartphone of 79.55% confidence. The model is more certain about the smartphone than the person. This may be due to the angle of the phone, the lighting, or the quality of the image. In general, those metrics suggest a strong performance of the model when detecting objects, especially more frequent objects (e.g.modes of mobile devices).

Figure 3: Line Chart of Metrics Used.

The performance of the model throughout time presented based on important evaluation metrics is represented in Figure 3. This model obtains a precision of 0.95, recall of 0.94, and F1-score of 0.945, indicating that the model performs well with respect to true positive rate while minimizing false positive detections. An Average Precision (AP) of 0.9 and Mean Average Precision (mAP) of 0.85 indicate that the objects can be easily detected or classified. Moreover, the IoU score of 0.92 indicates that detected objects are accurately localized, and an overall accuracy score of 0.96 indicates its reliability in real-world applications. Collectively, these metrics -- precision, recall, F1-score, AP, mAP, IoU, and accuracy -- illustrate the effectiveness of the model in both identifying and localizing objects.The results affirmatively demonstrate the strong performance of the model with opportunities for additional enhancements in accuracy detection under varied conditions leading to reliable and accurate recognition. To find the performance of the model:

$$Accuracy = \frac{(TP + TN)}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ Score = 2.\frac{Precision.Recall}{Precision+Recall} \quad (4)$$

8 CONCLUSIONS

The Yolov8-based object detection system displays remarkable real-time detection abilities, which attains balance between speed and accuracy. Thats why it's a powerful tool for many different applications, just because of how flexible it is in terms of the input it can handle. Strength analysis with high precision and recall rates confirms it is a strong model and generalizes well in different real -world scenarios. It shows how possible it is to do practical computer vision tasks using advanced object detection models (tcvr in our case).

9 FUTURE SCOPE

In the future can deploy yolov 8 in a domain specific dataset to increase accuracy for special applications. Moreover, more ways exist to grow the data and more architectural innovations will increase the accuracy. It would be important to adapt to the position of constant updates developing, ensuring continuous reliability. Such improvements will strengthen the space of the system as a state -of -the -art equipment in detecting real -time object. It would be important to adapt to the position of constant updates developing, ensuring continuous reliability. Such improvements will strengthen the space of the system as a state -of -the -art equipment in detecting real -time object.

REFERENCES

Ahmad, M., Khan, J., Yousaf, A., Ghuffar, S., Khurshid, K. (2023). Deep Learning: A Breakthrough in Medical Imaging. Current Medical Imaging Reviews, 16(10),

Ahn, W.-J., Ko, K.-S., Lim, M. T., Pae, D.-S., Kang, T.- K. (2023). Multiple Object Tracking Using ReIdentification Model with Attention Module. Applied Sciences, 13(3), 4298.

Chen, S., Xu, Y., Zhou, X., Li, F. (2019). Deep Learning for Multiple Object Tracking: A Survey. IET Computer Vision, 13(1), doi:10.1049/iet-cvi.2018.5598.

Hoffmann, J. E., Tosso, H. G., Santos, M. M. D., Justo, J. F., Malik, A. W., Rahman, A. U. (2021). RealTime Adaptive Object Detection and Tracking for Autonomous Vehicles. IEEE Transactions on Intelligent, 6(3), 450-459

Hou, R., Ma, B., Chang, H., Gu, X., Shan, S., Chen, X. (2019). Interaction-And-Aggregation Network for Person Re-Identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern recognition (CVPR), 9309-9318.

Khoshboresh-Masouleh, M., Shah-Hosseini, R. (2022). 2D Target/Anomaly Detection in Time Series Drone Images.

Liu,Q., Chen, D., Chu,Q., Yuan, L., Liu, B., Zhang, L., Yu, N. (2022, January). Online Multi-Object Tracking with Unsupervised ReIdentification Learning and Occlusion Estimation.

Walambe, R., Marathe, A., Kotecha, K., Ghinea, G. (2021, October). Lightweight Object Detection Ensemble Framework for Autonomous Vehicles in Challenging Weather Conditions. Computational Intelligence and Neuroscience, 2021.

Wang, G., Yuan, Y., Chen, X., Li, J., Zhou, X. (2018). Learning Discriminative Features with Multiple Granularities for Person ReIdentification. In Proceedings of the ACM Multimedia Conference, 274-282.

Wang, H., Du,H., Zhao, Y., j. (2020). A Comprehensive Overview of Person Re-Identification Approaches. IEEE Access, 8, 45556-45583.

Yang, W., Jiang, Y., Wen, S., Fan, Y. (2023). Online multiple object tracking with enhanced Reidentification. IET Computer Vision, 17(3), doi:10.1049/cvi2.12191.