VisiCompanion: Comparitive study of Yolov8 and Yolov9 for Grocery detection

1st Payal Chhabra
Department of (CSE-AIML)
KIET group of institutions
Ghaziabad, India
payal49691@gmail.com

2nd Tanmay Arora
Department of (CSE-AIML)
KIET group of institutions
Ghaziabad, India
aroratanmay162@gmail.com

3rd Vibhore Jain
Department of (CSE-AIML)
KIET group of institutions
Ghaziabad, India
vjmj4005@gmail.com

4th Neeraj Gandhi
Department of (CSE-AIML)
KIET group of institutions
Ghaziabad, India
letsunitethisglobe@gmail.com

5th Swapnil Bhatnagar
Department of (CSE-AIML)
KIET group of institutions
Ghaziabad, India
swapnilbhatnagar2255@gmail.com

6thNitin Arora
Department of Computer Science and Engineering
Thapar Institute of Engineering and Technology
Patiala, India
nitin.arora@thapar.edu

***Abstract*—There are approximately 21 million visually impaired individuals who depend on other people for their basic work. This research adds to the current understanding of assistive technology by introducing a method for choosing and optimizing the best object detection models suited for accessibility environments. This research enhances the Visicompanion functions and assists visually impaired individuals in doing crucial daily tasks independently by selecting a suitable model. This research compares the analysis of two models i.e. YoLov8 [1] and YOLOV9 [2] and identifies which algorithm is best for the application. YOLOv8 is known for its rapid inference speed which is beneficial for real-time systems whereas YOLOv9 introduces advanced detection capability that improves accuracy but may increase problems for processing time and computational resource requirements. This research uses evaluation metrics like precision, recall, size of model, and resource utilization across different hardware configurations. The results showed thatYOLOv9 is slightly better than YOLOv8 with mAP reaching 95% and precision 96.8%.**

***Index Terms*—Yolov8, Yolov9, Visually impaired people [3], Computer vision [4]**

I. Introduction

The WHO estimates 70 million vision-impaired people, 0.24 million of them are children. They face difficulties in day-to-day life whether purchasing something. They always have to depend on others for their daily tasks. To resolve this dependency Visicompanion comes into play. This project aims to do a comparative analysis of the Yolov8 and Yolov9 models for object detection [5] and decide which model performs better and can be used for the final product.

You Only Look Once v8 [6], hereafter referred to as YOLOv8, is an improved version of the YOLO family that gained popularity over its speed and accuracy in the detection of objects. From its parents, YOLOv8 further develops enhanced architectural innovations regarding enhanced feature extraction mechanisms, effective model scaling, and optimized anchor-free detection for high precision with improved recall performance on various benchmark datasets while offering its hallmark features in real-time inference. Due to its lightweight nature, it can run on modern hardware and, thus, be the best choice for applications requiring efficiency and responsiveness: autonomous systems, surveillance, and assistive technologies. Yolov9 opens a new chapter for performance advances in the object detection field and beyond. Grounded in architecture improvement, the YOLOv9 encapsulates attention, dynamic computation, and neater backbone networks for accuracy and speed.

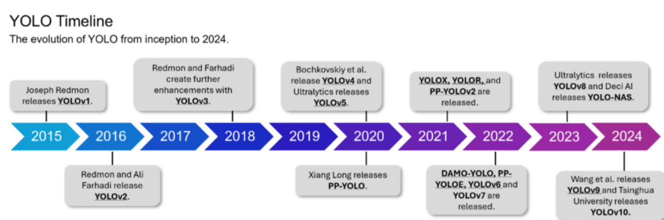

Fig. 1. Timeline of Yolo Models

A. *The problem statement*

This research addressed the challenges confronted by visually impaired people in supermarkets.

Most of the supermarkets in India do not have braille written anywhere, so visually impaired people have to depend on others for object identification.

B. *Primary objective*

The study accomplies the following goals:

1) Gathering different images to supermarket products in different angle to get every possible view and to make a dataset out of it.
2) Training Yolov8 and Yolov9 on that gathered image Dataset.
3) Determining which model performs better on supermarket dataset

II. LITRETURE SURVEY

Applications of deep learning models have transformed numerous sectors, including helping the blind and visually challenged. In particular, grocery detection [5] has drawn attention since it might greatly improve the shopping experience for those who are visually impaired by giving them aural feedback [7] about the objects in their environment. Detecting tiny things, which are frequent in shopping environments, is still difficult. YOLO-based models have been used in a number of publications to address the problem of small object recognition. By adding a Width Height Bounding Box Reconstruction (WHBBR) technique to YOLOv5, **Prabu Selvam et al. (2021)** suggested a modification that enhanced the model's performance in retail goods detection. On the **GroZi-120 dataset**, their method showed a notable improvement in both precision and recall, with precision reaching 86.3% and recall at 77.8%. For grocery detection applications, where identifying smaller products is essential to fostering an inclusive shopping environment for people with visual impairments, this enhancement is especially beneficial.

Since YOLOv8 and YOLOv9 are more sophisticated iterations of the YOLO model, their improved real-time tiny object identification skills make them ideal for supermarket detection jobs. These models offer better performance, speed, and accuracy, especially when handling little supermarket items that are sometimes overlooked in favor of larger goods on store shelves. Real-time feedback systems designed to help visually impaired people in grocery stores must be able to rapidly and properly identify these items.

In conclusion, there has been significant progress in improving the models' performance in real-world situations, and research on small object recognition utilizing YOLO-based models for supermarket detection is still ongoing [8]. The combination of YOLO's quick detection speed and sensory feedback tools, such as auditory signals is exceptional. Future research will probably concentrate on improving these models for even higher accuracy, cutting down on inference time, and strengthening their resilience to complex and crowded grocery store settings.

III. PERFORMANCE PARAMETER

For evaluating object detection models, a set of reliable performance metrics is needed to measure the accuracy of the model. Performance parameters like precision, recall, F1-score, and mAP will enable comparison between the YOLOv8 and YOLOv9 architectures under challenging conditions, including different occlusions, lighting, and object size variability.

A. *Mean average precision*

A high mAP says the model has learned to always detect objects precisely with few localization errors. mAP formula:

$$mAP = \sum_{n} (R_n - R_{n-1}) P_n$$

B. *Recall*

Recall measures how well a model can detect all possible instances of objects that might be present within an image and is calculated by:

$$Recall = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

C. *Precision*

Precision refers to the correctness of the model in terms of prediction, that is calculated by:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

IV. EXPERIMENTAL SETTING

This section discusses about the project environmental. First it provides details about the dataset, it provides details about the hyper parameter choosen for this experiment.

A. *Dataset*

This dataset includes 3586 pictures. 2972 pictures were trained, 415 validated, and 199 tested.

B. *Preprocessing*

Preprocessing step in this experiment include auto-orient and resize. The images of dataset has been resized to 640*640. A blur upto 2.5px is also applied.

C. *Hyper Parameter Tuning*

In this research below mentioned parameters are hypertuned to to achieve best performance from the version. Table 1. Shows the parameter with their values that are used during experiments for YOLOv8.

TABLE I
PARAMETERS FOR YOLOV8

Parameters	Values choosen
Learning rate	0.01
Batch size	32
Warm up epochs	3
No. of epochs	10
Input Size	640*640
patience	100
iou	0.7
momentum	0.937

whereas, Table 2. describes parameters choosen for Yolov9.

TABLE II
PARAMETERS FOR YOLOV9

Parameters	Values choosen
Learning rate	0.01
Batch size	16
Warm up epochs	3
No. of epochs	10
Input Size	640*640
patience	100
iou	0.7
momentum	0.937

V. RESULT

A. *Yolov8*

The performance and characteristics of the model have been comprehensively illustrated using various visualizations. Below we have provided the detailed analysis of each visualization and all the insights that can be extracted from it.

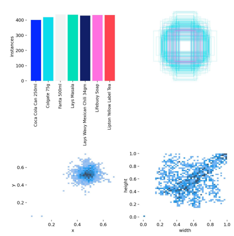

Fig. 2. Distribution Histogram (Top Left), Bounding Box Visualization (Top Right),Heat Map of Object Centers (Bottom Left),Heat Map of Bounding Box Dimensions (Bottom Right)

Figure 2 represents the distribution histogram, which shows the distribution at different instances across different classes. The top right image shows the detected bounding boxes as predicted. Each of the colored rectangles represents a detection. Bottom left: This is a heat map of object centers. It indicates the density of object centers throughout the image. The bottom right is the heat map for the bounding box dimensions: x-axis for the width, y-axis for the height normalized. These visualizations help diagnose model behavior, such as reporting a detection near the frame's center as small.

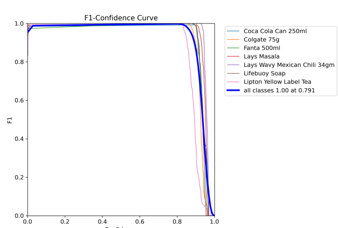

Fig. 3. Confidence Curve

Figure 3 for the F1 confidence curve. The confidence threshold and F1 score are on the x- and y-axes, respectively, from 0 to 1. The blue line shows the overall F1 score for all confidence threshold classes and also how the balance between precision and others with different confidence levels was achieved. The other grey lines represent the F1 score for the individual class as a function of the confidence threshold.

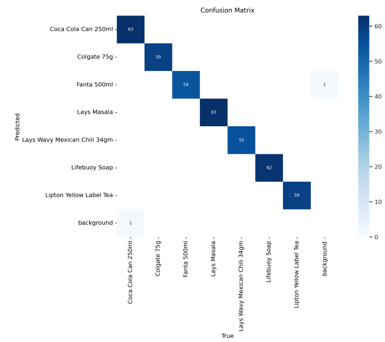

Fig. 4. Confusion Matrix

Figure 4. displays the confusion matrix for evaluating the classification performance of the YOLO v8 model. The diagonal line represents correct predictions where the true and predicted labels are the same. The off-diagonal elements correspond to misclassifications. The farther away an element is from the diagonal, the greater the error.

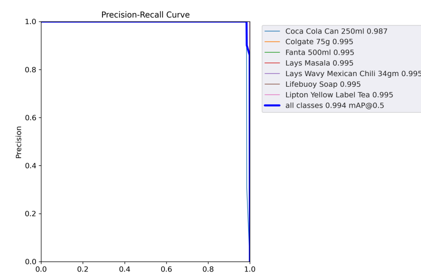

Fig. 5. Precision Recall Curve

Figure 5 evaluates the V8 model using the Precision-Recall Curve. It shows precision-recall tradeoffs at different thresholds. The mean average Precision (mAP) is 0.994, and the blue line shows overall training performance with an IoU threshold of 0.5. Individual class Precision-Recall curves are grey.The ideal curve is toward the top-right corner, indicating high precision and recall.

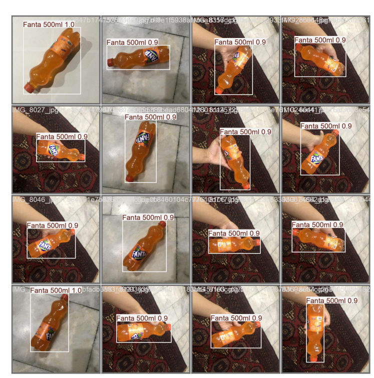

Fig. 6. Detected Imageset

The image provided above showcases the result of the object detection using the model YOLOv8. For example, for Lipton yellow label, fanta 500ml, etc. There is a diverse detection in this variety of objects. Moreover, there is a detection in different settings. The model has been tested on images which are taken in indoor and outdoor environments. Enhancement is made by applying non-maximum suppression which helps in improving the accuracy and reliability of the detection.

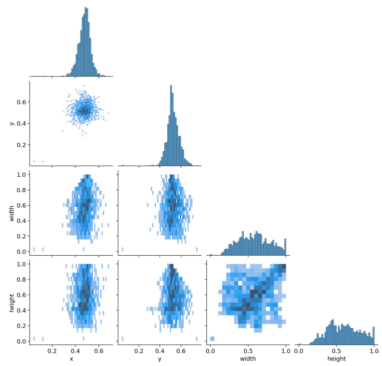

Fig. 7. Distribution of Individual Variables Curve

The provided pair plot image illustrates the various relationships between the variable's width and height. The summary of the results shown in the above image is the distribution of individual variables that is histogram on the diagonal, x, y, width, and height. Also, there can be a relationship between the pair of variables that is the scatter plot on the off-diagonal, x vs y, x vs width, y vs height, and width vs height. The key insights are object position, which is the uniform distribution of x and y values, the object size, and the inter-variable relationships. It demonstrates that the object detection model detects that objects are distributed evenly across the image frame.

B. *Yolov9*

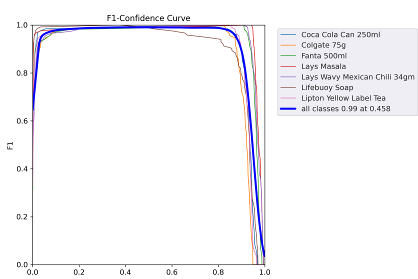

Fig. 8. Confidence Curve

The confidence Curve of F1-Score as shown in Figure 8 demonstrates the performance of an object detection model at different confidence thresholds. - X-Axis: Confidence threshold (0 to 1). - Y-Axis: F1 score (0 to 1), which represents the balance between precision and recall.

- The blue line indicates the all-class F1 score, peaking at 0.99 with 0.458 confidence. - Colored/gray curves indicate F1 scores for individual classes, such as "Coca Cola Can 250ml" and "Colgate 75g," showing the model's performance for each class. - Optimal F1 scores for most classes are achieved at thresholds between 0.8 and 1.0, where tuning these thresholds significantly improves prediction accuracy.

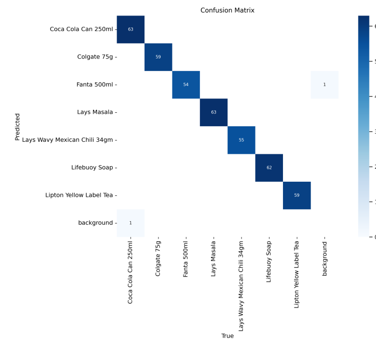

Fig. 9. Confusion Matrix

Figure 9 shows the confusion matrix of a classification model by comparing actual and predicted class labels. For example, for classes "Colgate 75g" and "Fanta 500ml," it scores 1.00 diagonal, meaning no misclassifications occurred for these classes. However, small mistakes are noted: "Coca-Cola Can 250ml" was misclassified occasionally as "background" with a 0.02 value.

There is a little confounding between the "background" class and product classes, suggesting places where the model has difficulties, such as overlapping values like 0.25. Overall, the matrix, therefore, highlights those classes where the model performs very well and even suggests improvement is needed in distinguishing between product classes and the background.

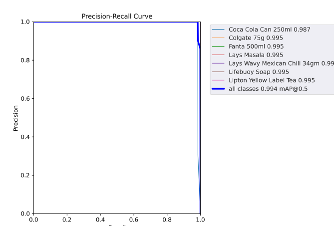

Fig. 10. Precision Recall Curve

Figure 10 demonstrates the precision and recall of a multi-class object detection model across different thresholds to assess its capability to predict the right positive instances accurately while capturing the actual positives well.

1. Class-wise Curves: Every curve is for one class like "Coca Cola Can 250ml" or "Colgate 75g" showing consistently high precision and recall.
2. Average Precision (AP): Most of the classes have an AP of 0.995, indicating perfect object classification.
3. Overall Performance: The blue curve shows strong detection reliability across all classes with a mean Average Precision (mAP) of 0.994.

The step and flat curves near the top show an excellent balance between precision and recall, and the uniformly high

AP values show that the model is robust and consistent across all classes.

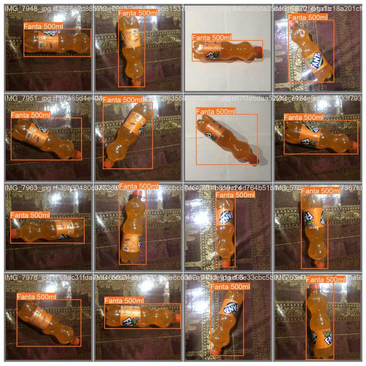

Fig. 11. Detected Imageset

In Figure 11, the image demonstrates the object detection capabilities of the YOLOv9 model, successfully identifying ”Fanta 500ml” bottles across various orientations, lighting conditions, and backgrounds. Each bottle is accurately labeled with bounding boxes, reflecting the model’s precision.

Overall, YOLOv9 excels in object detection, showing improved precision and robustness compared to its predecessors.

Detection Performance: The YOLOv9 model was able to detect all instances of ”Lifebuoy Soap” within the image; hence, robustness was confirmed in different angles, lighting, and partial occlusions. Bounding boxes appear accurate, as they tightly cover the objects without any overlap or misalignment; therefore, good localization is obtained. No obvious false positives and false negatives have been observed in this experiment.

Accuracy Metrics: The detections appear to have high confidence, and the model is able to correctly identify objects even with variations in object rotation and viewpoint. These results indicate that YOLOv9 performs exceptionally well in detecting objects with consistency and accuracy.

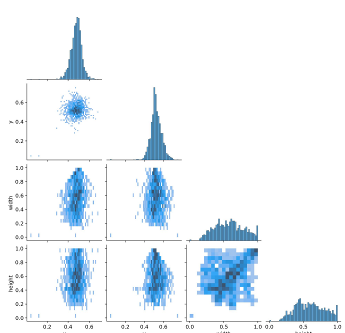

Fig. 12. Distribution of Individual Variables Curve

In Figure 12, visualization depicts the distribution and association of parameters of bounding boxes that are used within object detection tasks. Diagonal histograms give information on the distribution of parameters as separate features. The x and y coordinates remain centered on 0.5, thus suggesting that most objects are found centrally within images. However, the width and height parameters are more spread apart, indicating a drastic range of object sizes in the given data set.

In general, this visualization delivers insightful information about the spatial and dimensional characteristics of your dataset.

VI. Comparison Results

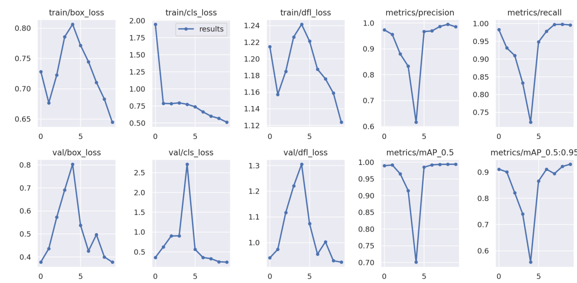

Fig. 13. Precision value curve

In Figure 13. The table shows the training and validation metrics over seven epochs. Key trends include:

Losses (box, classification, DFL): All losses decrease steadily, indicating the model’s improved performance in detecting and classifying objects both in training and validation.

Precision and Recall: Both metrics stabilize at 1.00 in the final epochs, showing the model effectively detects objects without false positives or missed detections.

mAP (Mean Average Precision): Performance metrics improve significantly, with mAP(0.5) and mAP(0.5):0.95 indicating high accuracy in object detection. The consistent improvement across metrics reflects a robust and well-generalized model.

Table III. Experimental Results to detect Grocery using different methods

Methods	Precision	Recall	mAP	F1 Score
YOLOv5	86.3%	77.8%	-	72.4%
HYOLONAS	82.3%	92.1%	96.8%	84.0%
Faster RCNN	0.47	0.51	90.46	-
Retina Net	0.51	0.65	25.54	-
YOLOV8(OUR PROPOSED)	87.7	84.8	91.4	
YOLOV9(OUR PROPOSED)	96.8	99.4	95	

References

[1] J. Qu, Q. Li, J. Pan, M. Sun, X. Lu, Y. Zhou, and H. Zhu, “Ss-yolov8: small-size object detection algorithm based on improved yolov8 for uav imagery,” *Multimedia Systems*, vol. 31, no. 1, pp. 1–17, 2025.

[2] M. M. Reda, D. M. El Sayad, N. Laban, and M. F. Tolba, “Performance evaluation of yolov8 and yolov9 for object detection in remote sensing images,” in *International Conference on Advanced Intelligent Systems and Informatics*, pp. 124–133, Springer, 2024.

[3] C.-E. Panazan and E.-H. Dulf, “Intelligent cane for assisting the visually impaired,” *Technologies*, vol. 12, no. 6, p. 75, 2024.

[4] P. Chhabra and S. Goyal, "A thorough review on deep learning neural network," in *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, pp. 220–226, IEEE, 2023.

[5] P. Chhabra and S. Goyal, "An efficient grocery detection system using hyolo-nas deep learning model for visually impaired people," 2024.

[6] A. Sharma, V. Kumar, and L. Longchamps, "Comparative performance of yolov8, yolov9, yolov10, yolov11 and faster r-cnn models for detection of multiple weed species," *Smart Agricultural Technology*, vol. 9, p. 100648, 2024.

[7] Y. Khandelwal, R. Goyal, and P. Negi, "User interface based text-to-speech synthesizer," in *2024 IEEE 1st Karachi Section Humanitarian Technology Conference (KHI-HTC)*, pp. 1–8, IEEE, 2024.

[8] J. I. A. Padilla, H. M. R. Pastor, A. C. T. Velasco, and J.-A. V. Magsumbol, "Performance evaluation of yolov8, yolov9 and yolov10 models in detecting fusarium wilt disease in banana leaf plants," in *2024 IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 1–7, IEEE, 2024.

[9] P. Chhabra and S. Goyal, "An examination of the feasibility of various deep learning object detecting techniques," in *2023 International Conference on Disruptive Technologies (ICDT)*, pp. 237–242, IEEE, 2023.

[10] A. Tasnim, B. Das, M. R. Islam, M. Amiruzzaman, M. R. Islam, and N. Ahmed, "Revolutionizing rose grading: Real-time detection and accurate assessment with yolov8 and deep learning models," *SN Computer Science*, vol. 6, no. 1, pp. 1–11, 2025.

[11] Y. Wu, Y. Yuan, *et al.*, "Research on oriented detection method for water gauge,"

[12] S. Alagarsamy, T. D. Rajkumar, K. Syamala, C. S. Niharika, D. U. Rani, and K. Balaji, "An real time object detection method for visually impaired using machine learning," in *2023 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–6, IEEE, 2023.

[13] M. Awais Ahmad, T. Ahmed, M. Aslam, A. Rehman, F. S. Alamri, S. A. Bahaj, and T. Saba, "Mathvision: An accessible intelligent agent for visually impaired people to understand mathematical equations," *IEEE Access*, vol. 13, pp. 6155–6165, 2025.

[14] J. Shajeena, R. Shiny, M. Vespa, and R. Kavitha, "Sign language translation and voice impairment support system using deep learning," in *2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, pp. 558–563, IEEE, 2024.

[15] M. S. H. Talukder and S. Akter, "An improved ensemble model of hyper parameter tuned ml algorithms for fetal health prediction," *International Journal of Information Technology*, vol. 16, no. 3, pp. 1831–1840, 2024.

[16] O. G. Ajayi, P. O. Ibrahim, and O. S. Adegboyega, "Effect of hyperparameter tuning on the performance of yolov8 for multi crop classification on uav images," *Applied Sciences*, vol. 14, no. 13, p. 5708, 2024.

[17] L. Ramos, E. Casas, E. Bendek, C. Romero, and F. Rivas-Echeverría, "Hyperparameter optimization of yolov8 for smoke and wildfire detection: Implications for agricultural and environmental safety," *Artificial Intelligence in Agriculture*, 2024.

[18] A. Maseleno, M. Huda, and C. A. Ratanamahatana, "Metaheuristic hyperparameter optimization and explainable deep learning approach for recognition of prohibited objects in baggage inspection," *Available at SSRN 4772555*.

[19] B. Karthika, M. Dharssinee, V. Reshma, R. Venkatesan, and R. Sujarani, "Object detection using yolo-v8," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–4, IEEE, 2024.