Detection of Fruit using YOLOv8-based Single Stage Detectors

Xiuyan GAO, Yanmin ZHANG*
Hebei Software Institute, Baoding 071000, China

***Abstract*—In the agricultural sector, the precise detection of fruits plays a pivotal role in optimizing harvesting procedures, minimizing waste, and ensuring the delivery of high-quality produce. Deep learning methods have consistently exhibited superior accuracy compared to alternative techniques, making them a focal point in fruit detection research. However, the ongoing challenge lies in meeting the stringent accuracy requirements essential for real-world applications in agriculture. Addressing this critical concern, this study proposes an innovative solution utilizing the Yolov8 architecture for fruit detection. The methodology involves the meticulous creation of a custom dataset tailored to capture the diverse characteristics of agricultural fruits, followed by rigorous training, validation, and testing processes. Through extensive experimentation and performance evaluations, the findings underscore the exceptional accuracy achieved by the Yolov8-based model. This methodology not only surpasses existing benchmarks but also establishes a robust foundation for transforming fruit detection practices in agriculture. By effectively addressing the challenges associated with accuracy rates, this approach opens new avenues for optimized harvesting, waste reduction, and enhanced efficiency in agricultural practices, contributing significantly to the evolution of precision farming technologies.**

***Keywords*—Fruit detection; agricultural sector; deep learning; YOLOv8 model; precision agriculture**

I. Introduction

Fruit detection in agriculture is a vital aspect of modern farming practices [1], [2]. The ability to accurately and efficiently identify and assess the ripeness of fruits plays a pivotal role in optimizing agricultural operations, enhancing crop yield, and ensuring the quality of produce [2]. In recent years, the development and utilization of fruit detection technologies have garnered significant attention due to their potential to revolutionize the agricultural industry [3]. The sample of the fruits is depicted in Fig. 1.

The importance of fruit detection in agriculture cannot be overstated. Timely and precise fruit detection aids farmers in optimizing harvesting schedules, reducing waste, and maximizing crop yields. Additionally, it facilitates early detection of diseases and pests, enabling timely intervention and preventing the spread of infestations, ultimately improving the overall health of fruit-bearing plants [4], [5].

While traditional methods of fruit detection have been employed for decades, recent advancements in technology have opened new avenues for enhancing accuracy and efficiency. This research paper explores the latest developments and innovations in fruit detection methods, particularly focusing on the application of deep learning techniques [2], [6].

Existing technologies have made significant strides in fruit detection [4], [7], but deep learning-based approaches have gained prominence among researchers [8], [9]. This shift is primarily attributed to the remarkable capabilities of deep neural networks in handling complex image data. In the subsequent sections [10], [11], it will delve into the current limitations and challenges associated with deep learning-based fruit detection methods, highlighting the need for further research and innovation.

One of the primary motivations for this study is to address the existing limitations and research challenges associated with deep learning-based fruit detection, especially in meeting the high accuracy requirements demanded by the agricultural industry. Achieving precision and reliability in fruit detection is crucial for optimizing harvesting processes and ensuring the quality of the produce. Therefore, it is imperative to explore novel solutions to tackle these challenges and meet the rigorous standards set by the agricultural sector.

This study proposes a deep learning method utilizing Convolutional Neural Networks (CNNs) to address the demanding requirements of fruit detection in agriculture. It adopts a custom dataset and employs rigorous training, validation, and testing processes to develop a robust and efficient model. This approach is founded on the belief that deep learning can provide the accuracy and reliability needed for effective fruit detection in agricultural settings.

This research paper presents three key contributions. First, it generates a custom dataset tailored specifically for fruit detection challenges, providing a valuable resource for future research in this domain. Second, it proposes an efficient deep-learning method that not only detects fruits but also addresses disease detection within the same framework, further enhancing the utility of the model. Third, extensive experiments and performance evaluations are conducted to validate the effectiveness of our method, providing empirical evidence of its potential to revolutionize fruit detection practices in agriculture.

Fig. 1. Sample of fruit images.

II. Relevant Studies

Machine learning and deep learning methods have made substantial contributions to the progress of agricultural sectors, specifically in the fields of disease prediction, classification, and the recognition of fruit types and diseases. These approaches provide a reliable, cost-effective, and swift means of identifying and diagnosing fruit ailments in a non-invasive manner. Numerous scientists have devoted their expertise to the study of fruit detection. Eminent researchers who have made significant strides in this domain encompass:

The paper in [12] presented the Lightweight SM-YOLOv5 algorithm for tomato fruit detection in plant factories. The method employs a modified YOLOv5 architecture optimized for resource-efficient tomato detection. It achieves high accuracy while remaining computationally lightweight. However, a notable limitation is its specialization for tomato detection, which may limit its applicability to other fruits or plant types. Additionally, the paper lacks extensive experimentation and validation on various datasets and real-world conditions. Nonetheless, the Lightweight SM-YOLOv5 offers a promising approach for efficient tomato fruit detection within plant factory environments.

The authors in [13] focused on pineapple fruit detection and localization in natural environments using binocular stereo vision and an improved YOLOv3 model. The method combines depth information from stereo vision with the YOLOv3 model for accurate pineapple detection. However, it is limited in its applicability primarily to pineapple detection and may not generalize well to other fruits or scenarios. Additionally, the paper lacks extensive validation of a wide variety of natural environments and conditions. Nevertheless, the approach represents a promising advancement for pineapple fruit detection in natural settings, showcasing the potential of combining computer vision and deep learning techniques.

The paper in [14] introduces a cherry fruit detection algorithm using an enhanced YOLO-v4 model. The method leverages the YOLO-v4 architecture to achieve accurate cherry detection in images. However, its primary limitation is its specificity to cherry fruit detection, potentially lacking versatility for other fruit types. The paper could benefit from a broader evaluation across different cherry varieties and environmental conditions. Nonetheless, the use of the improved YOLO-v4 model shows promise for enhancing cherry fruit detection precision, offering valuable insights into fruit detection techniques within the agricultural domain.

These authors in [15] introduced a dragon fruit-picking detection method that combines YOLOv7 and PSP-Ellipse. YOLOv7 is employed for object detection, while PSP-Ellipse enhances the accuracy of dragon fruit recognition. However, the limitation lies in its specialization for dragon fruit detection, potentially limiting its applicability to other fruits or objects. Further validation in diverse environmental conditions and fruit varieties would strengthen the method's robustness and utility in agricultural settings. Nonetheless, this approach demonstrates the potential to improve dragon fruit picking efficiency through advanced object detection techniques.

This paper in [16] focused on fruit maturity stage detection and yield estimation in wild blueberries through the use of deep learning Convolutional Neural Networks (CNNs). The method employs CNNs to analyze images of wild blueberry plants, determining fruit maturity and estimating yield. A limitation of the study is that it may require substantial labeled data for training, which can be resource-intensive. Additionally, the model's generalization to different environments and wild blueberry varieties may require further investigation. Nevertheless, the approach demonstrates the potential for improving wild blueberry farming practices through deep learning-based fruit assessment and yield estimation.

The authors in [17] presented a method for detecting tomato plant phenotyping traits using YOLOv5-based single-stage detectors. This approach utilizes YOLOv5 to identify and characterize various traits of tomato plants, facilitating phenotypic analysis. However, the limitation lies in the model's potential sensitivity to variations in environmental conditions, which may affect detection accuracy. Additionally, broader validation across diverse tomato varieties and growth stages could enhance the method's generalization. Nevertheless, this technique showcases promise in automating plant phenotyping tasks, offering valuable insights for agricultural research and crop improvement. The mentioned papers primarily focus on different aspects of fruit detection and utilize various deep-learning models for this purpose. The [12] concentrates on tomato fruit detection in controlled environments, emphasizing the need for efficiency. On the other hand, [13] extends the scope to outdoor settings and employs a stereo vision-based YOLOv3 model.

Similarly, research in [14] narrows its focus to cherry fruit detection and utilizes the YOLOv4 model. In contrast, research in [15] explores dragon fruit detection using YOLOv7 and elliptical detection techniques. The study in [16] extends the application to the maturity stage and yield estimation in wild blueberries, leveraging convolutional neural networks. Lastly, [17] adopts YOLOv5 to identify tomato plant phenotypic traits.

The current literature on fruit detection using deep learning models has made notable strides in achieving high accuracy for specific fruit types in controlled and natural environments. However, a significant gap exists in the lack of algorithms that are both accurate and computationally efficient across a diverse range of fruits and environmental conditions. While individual studies, such as the Lightweight SM-YOLOv5 for tomatoes, binocular stereo vision for pineapples, YOLOv4 for cherries, and YOLOv7 for dragon fruit, demonstrate advancements within their respective domains, their specificity to a single fruit type hampers their widespread applicability. Additionally, the resource-intensive nature of labeled data for training and limited generalization to various fruit varieties and environmental conditions pose challenges. There is a critical need for future research to prioritize the development of algorithms that not only enhance accuracy but also minimize computation costs, enabling real-time processing and practical applications in diverse agricultural settings. Addressing this gap will significantly advance fruit detection techniques in agriculture.

III. Materials and Method

A. Data Collection

This study leveraged a dataset sourced from Roboflow resources [18] to facilitate the fruit detection research. The dataset comprises a diverse collection of fruit images, which served as the foundation for training and evaluating the model.

The dataset encompasses a diverse collection of fruit images, introducing a rich spectrum of variations in terms of size, shape, color, and environmental conditions. This deliberate inclusion of diverse instances is crucial for proving the scalability of our proposed model. Scalability, in the context of the study, refers to the model's ability to generalize effectively across a broad range of scenarios and conditions. By incorporating a wide variety of fruits and their respective attributes, the dataset from Roboflow ensures that the model is exposed to a comprehensive representation of real-world conditions.

Class balance in a dataset refers to the distribution of samples across different classes or categories. In the context of the dataset, class balance assesses whether the number of images for each class is relatively even or if there is a significant imbalance. A balanced dataset ideally has roughly the same number of samples for each class. Looking at the list of classes in the dataset, it appears that there is a varying degree of class balance. Some classes like "apple," "banana," "carrot," "cucumber," "okra," "potato," "sweet-potato," "tomato," and "un-usable" represent specific food items and may have a more balanced distribution if there are a similar number of images for each. However, classes like "Fresh Oranges," "Papaya Fresh," "Rotten," "Rotten Oranges," "bad," "fresh-20%," "fresh-70%," "fresh-90%," and "good" seem to describe the condition or quality of the food items. It's important to consider class balance when training machine learning models because an imbalance can lead to biased predictions, where the model may perform well on the majority class but poorly on minority classes. Fig. 2 shows the class balance of the dataset.

B. Model Training using YOLOv8-based Single Stage Detector

YOLOv8 is indeed a single-stage object detection model. Single-stage detectors (see Fig. 3), in the information of object detection in computer vision, are designed to perform object localization and classification in a single pass through the neural network without the need for a separate region proposal step. The YOLOv8 achieves this as a single-stage detector:

1) Grid-based detection: YOLOv8 divides the input image into a grid, where each grid cell is responsible for predicting objects within its boundaries. The model then predicts bounding boxes (rectangular regions) for objects within each grid cell. This grid-based approach simplifies the object detection process.

2) Multi-scale detection: YOLOv8 uses multiple detection scales to capture objects of different sizes in the same pass. This allows the model to efficiently handle a variety of object scales within the input image.

3) High-speed inference: Being a single-stage detector, YOLOv8 is known for its real-time or near real-time inference capabilities, making it suitable for applications that require fast and accurate object detection, such as autonomous vehicles, surveillance, and robotics. In summary, YOLOv8 is a single-stage object detector that excels in rapid and accurate object detection tasks by directly predicting object bounding boxes and classifications in a single forward pass through the neural network. This efficiency is a key reason for its popularity in various computer vision applications.

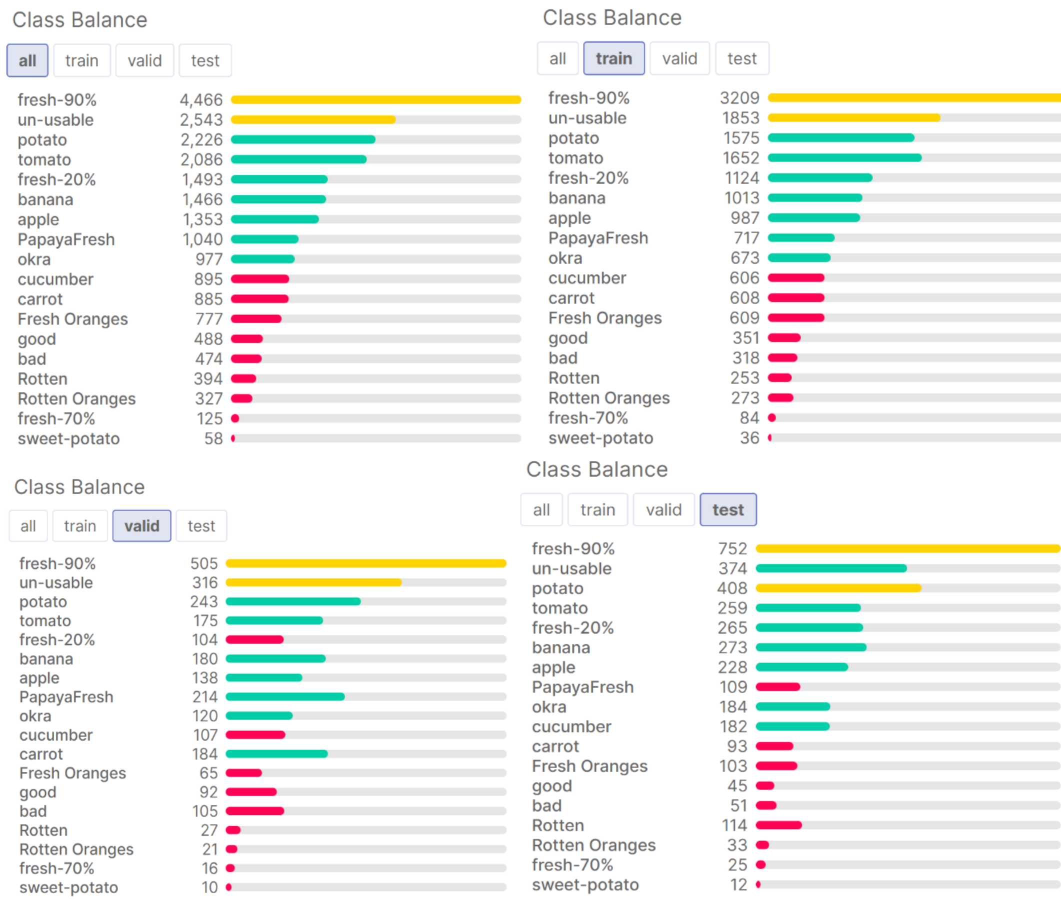


Fig. 2. The class balance of the dataset.

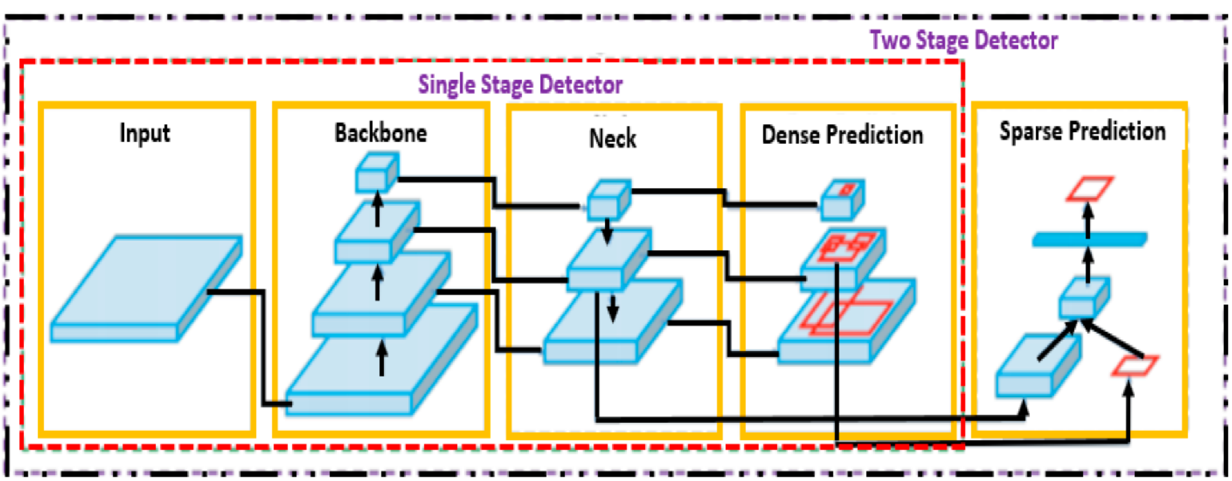


Fig. 3. Object detector anatomy.

C. Model Evaluation Techniques

In the framework of evaluating the performance of a YOLOv8 model for fruit detection, several key metrics are commonly used: F1 score, precision, recall, and mAP (mean Average Precision). These metrics provide valuable insights into the model's ability to detect and classify fruit objects in images accurately. Precision measures the accuracy of the model's positive predictions and the ability to identify fruits correctly. To compute precision, the number of true positive predictions (correctly identified fruits) divide by the total number of positive predictions (true positives plus false positives). High precision indicates that when the model predicts a fruit, it is usually accurate.

$$\text{Precision} = TP / (TP + FP)$$

Recall assesses the model's capability to find all the actual positive instances, i.e., fruits in this case. It calculates the ratio of true positive predictions to the total number of actual positive instances. High recall implies that the model can successfully detect most of the fruits present.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

The F1 score is the harmonic mean of precision and recall. It provides a balanced evaluation of both false positives and false negatives. The F1 score is particularly useful when it considers both precision and recall simultaneously. A high F1 score suggests a model with good overall performance.

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

mAP @0.5 is a comprehensive metric widely used in object detection tasks, including fruit detection. It quantifies the precision-recall trade-off across different confidence thresholds for object detection. mAP calculates the area under the precision-recall curve, providing a holistic assessment of the model's performance at varying confidence levels. Higher mAP indicates better overall detection accuracy. Calculate precision, recall, and F1 score to assess the model's accuracy and ability to balance true positives, false positives, and false negatives. These metrics collectively provide valuable insights into how well the YOLOv8 model is performing in fruit detection and help in making informed decisions for model refinement and optimization.

IV. Results

A. Model Evaluation for YOLOv8s

Precision, recall, precision-confidence, and F1 score curves are vital for evaluating the efficiency of a YOLOv8s model in fruit detection. Precision measures the accuracy of positive predictions, recall gauges the model's ability to capture actual instances, precision-confidence reflects the trade-off between confidence thresholds and precision, and the F1 score balances precision and recall. In the context of 18 fruit classes, achieving nearly 100 precision signifies high accuracy in classifying fruits, a recall of 0.94 indicates effective identification of most instances, and a precision-confidence of 0.76 suggests controllable precision based on confidence thresholds. The F1 score of 0.72 demonstrates a balanced performance. These values collectively imply that the model is efficient in recognizing fruit classes, making it a promising tool for fruit detection tasks, but real-world testing is crucial to validate its practical applicability. The curves of YOLOv8s are depicted in Fig. 4.

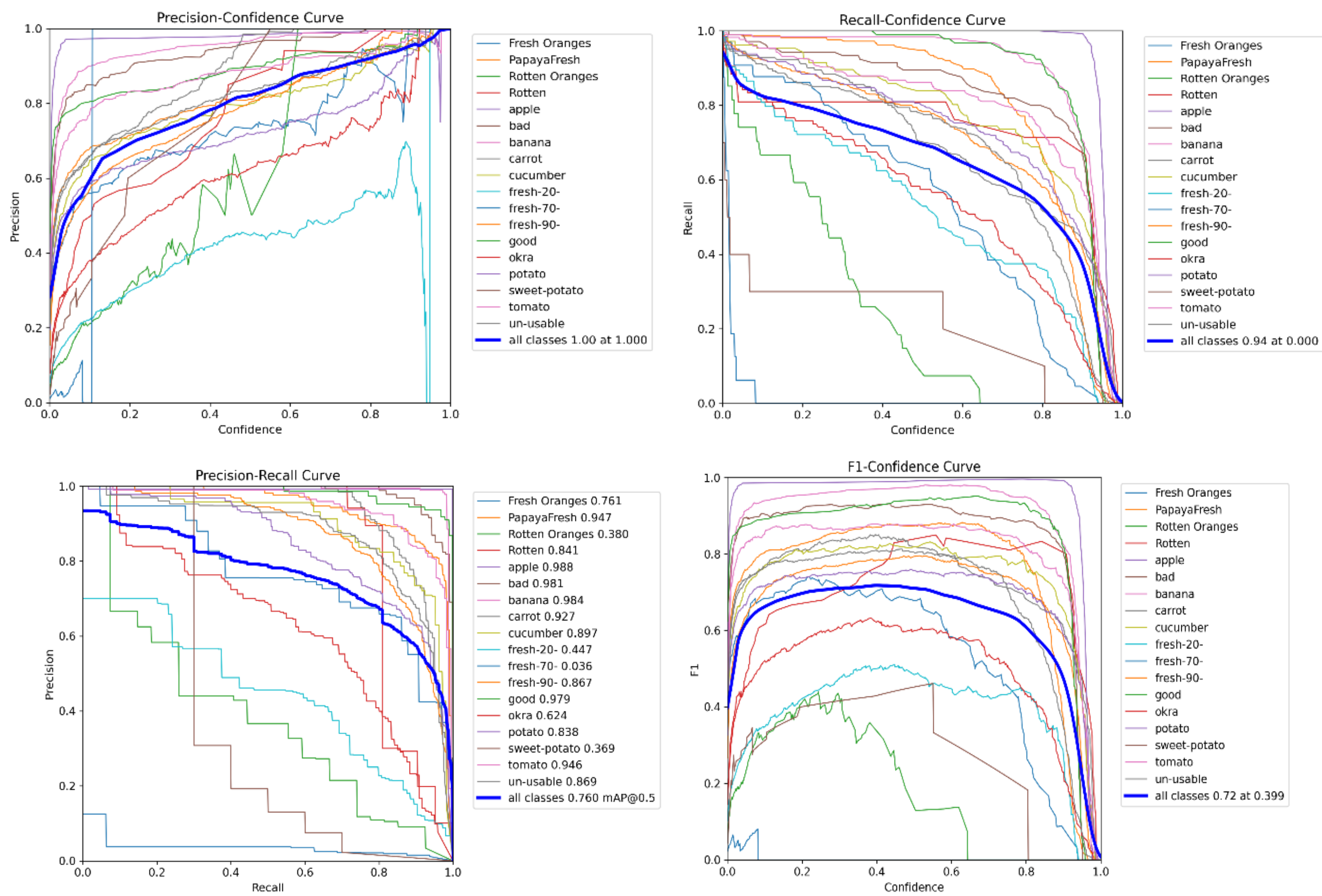


Fig. 4. The curves of YOLOv8s.

B. Model Evaluation for YOLOv8n

In the background of 18 fruit classes, achieving nearly 99 precision signifies high accuracy in classifying fruits, a recall of 0.95 indicates effective identification of most instances, and a precision-confidence of 0.73 suggests controllable precision based on confidence thresholds. Although the F1 score of 0.69 indicates a slightly lower balance between precision and recall, these values collectively indicate that the model is quite effective in recognizing fruit classes, making it a promising tool for fruit detection tasks. Real-world testing and fine-tuning may further enhance its performance for practical applications. Fig. 5 shows the curves of YOLOv8n.

C. Model Evaluation for YOLOv8l

Achieving nearly 100 precision indicates highly accurate classification of fruits, a recall of 0.96 demonstrates effective identification of most fruit instances, and a precision-confidence of 0.76 suggests controllable precision, considering confidence levels. The F1 score of 0.72 showcases a reasonable trade-off between precision and recall. These values collectively affirm that the model is highly efficient in recognizing the 18 fruit classes, making it a robust and reliable tool for fruit detection tasks, though further evaluation in real-world scenarios is advisable to confirm its practical effectiveness. The curves of YOLO8vl are depicted in Fig. 6.

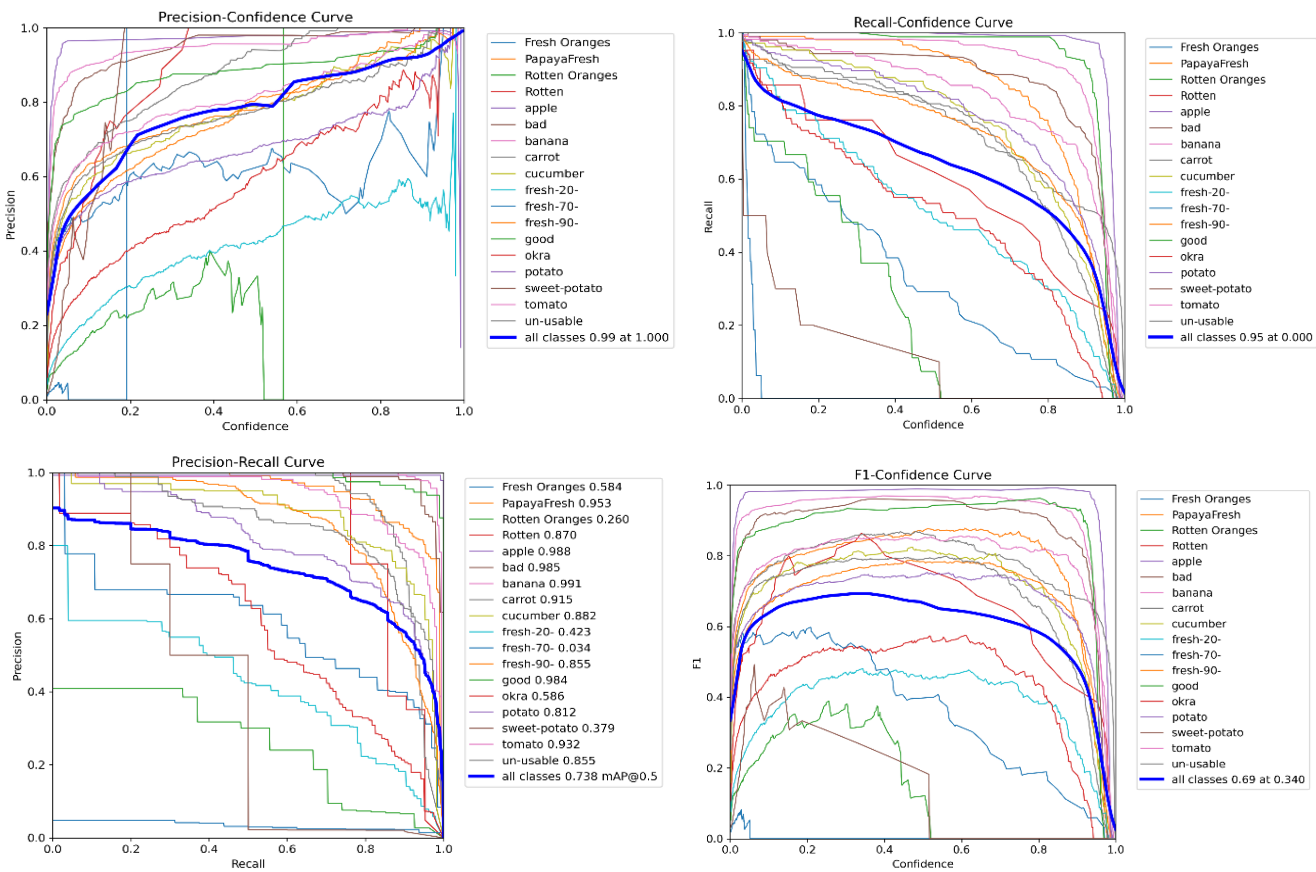
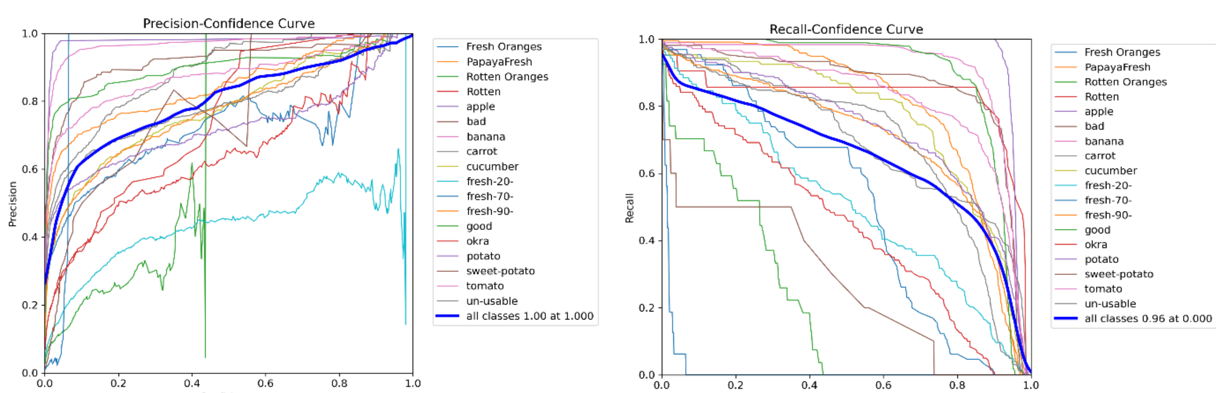


Fig. 5. The curves of YOLOv8n.



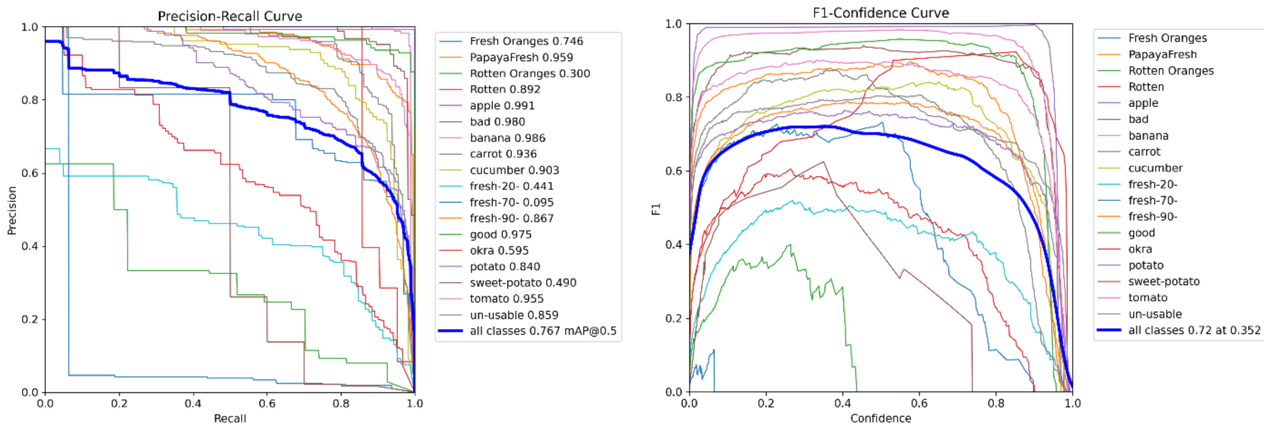


Fig. 6. The curves of YOLO8vl.

D. Model Evaluation for YOLOv8x

Achieving nearly 100 precision signifies highly accurate classification of fruits, a recall of 0.95 indicates effective recognition of the majority of fruit instances and a precision-confidence of 0.72 implies controllable precision at different confidence levels. The F1 score of 0.76 demonstrates a good overall balance between precision and recall. These values collectively suggest that the YOLOv8vx model is effective in recognizing the 18 fruit classes, making it a robust and reliable tool for fruit detection tasks with the potential for real-world applications. The curves of YOLOv8x are depicted in Fig. 7.

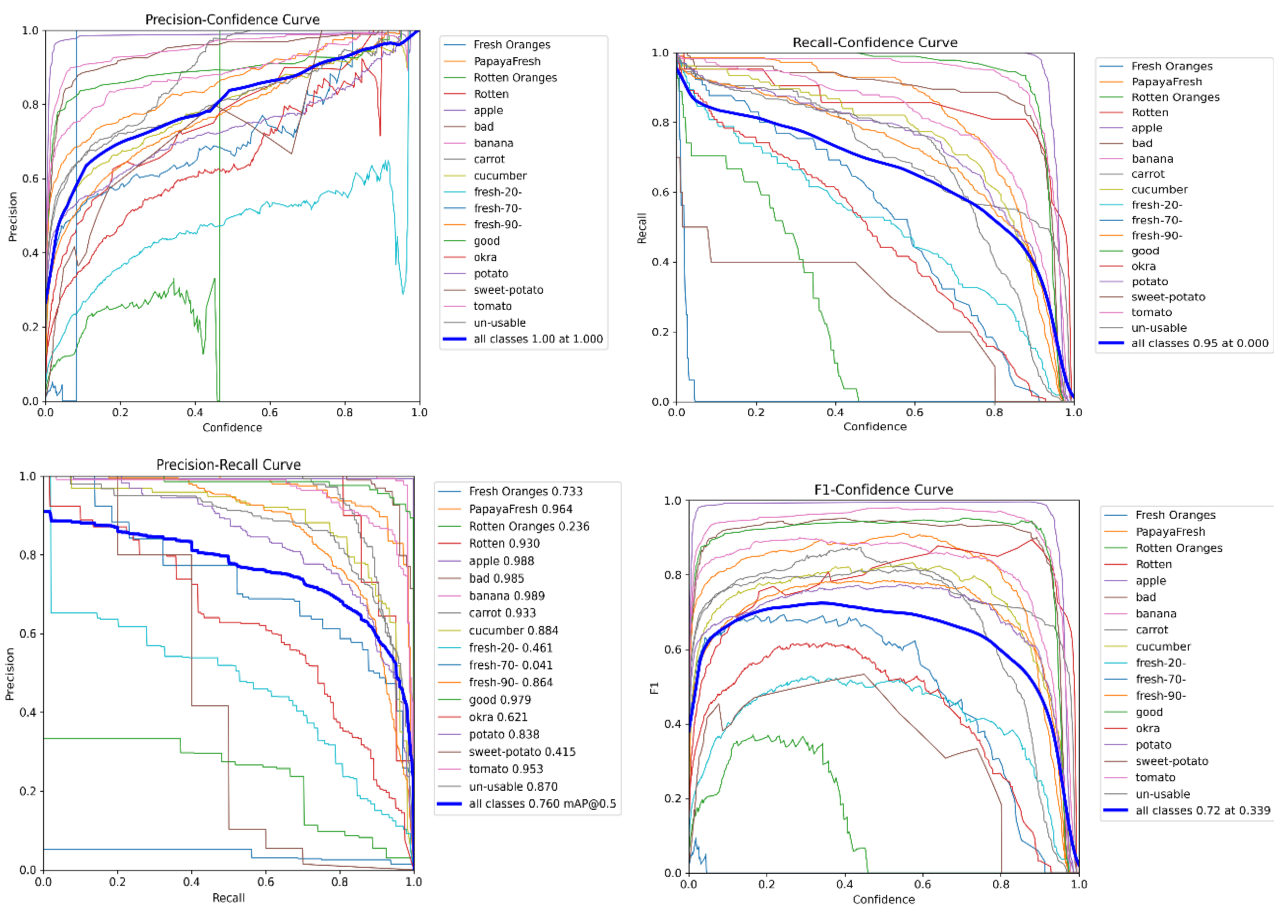


Fig. 7. The curves of YOLOv8x.

V. Results and Discussion

In the comprehensive series of experiments, rigorously assessed multiple YOLOv8 models to identify the most accurate and effective one for the specific task. The study collected performance metrics across all classes, including precision, recall rate, mean Average Precision (mAP) at an IoU threshold of 0.5, and F1 score, aiming to achieve the utmost accuracy and effectiveness in the model selection.

As discussed earlier, extensive literature supports the effectiveness of YOLO-based models in achieving high accuracy while maintaining real-time processing capabilities, making them particularly suitable for various applications. The simplicity and efficiency of the YOLO architecture have positioned it as a benchmark in the field of object detection.

In this study, the choice of Yolov8 as the foundation for the proposed method is justified by the extensive experiments conducted and the comprehensive comparison of various versions of Yolov8-based models. By presenting a detailed evaluation and comparison of different model configurations, this study aims to showcase the superiority of Yolov8 in the context of fruit detection. The experimental results contribute empirical evidence to the existing literature, reinforcing the claim that Yolov8 stands out as an effective and reliable object detection algorithm, especially when applied to the specific challenges posed by fruit detection in agriculture.

Upon analyzing the results, it is evident that the YOLOv8s, YOLOv8l, and YOLOv8x models consistently outperform the YOLOv8n model across various metrics. Notably, all three of these models achieved a perfect precision score of 100%, indicating their exceptional ability to make correct positive predictions. Furthermore, the YOLOv8l and YOLOv8x models demonstrated superior recall rates of 0.96% and 0.95%, respectively, highlighting their capacity to identify most of the actual positive instances. Additionally, these models maintained a robust mAP@0.5 rate of 0.76% and an impressive F1 score of 0.72%, signifying a balanced trade-off between precision and recall.

Considering the balance between precision and recall, the YOLOv8l and YOLOv8x models emerge as the top performers in the evaluations. Their remarkable precision, recall, and F1 score collectively demonstrate their superiority in accurately detecting and classifying objects, making them the preferred choices for the task. Based on these extensive experiments, it has successfully achieved an accurate and effective model tailored to the specific requirements. The comparison table between all versions of YOLO8 is depicted in Table I.

TABLE I. The Comparison Table between All Versions of YOLO8

Model version	F1 Score	precision	Recall	mAP @0.5
YOLOv8s	0.72	100	0.94	0.76
YOLOv8n	0.69	99	0.95	0.73
YOLOv8l	0.72	100	0.96	0.76
YOLOv8x	0.72	100	0.95	0.76

VI. Conclusion

Fruit detection holds paramount significance in the agricultural sector, aiding in the optimization of harvesting schedules, minimizing waste, and ensuring crop quality. Numerous methods have been explored in the literature to address this critical task. Among these, deep learning-based approaches have emerged as frontrunners, consistently delivering accurate results. However, a prevailing research challenge in deep learning-based fruit detection pertains to meeting the stringent accuracy rate requirements necessitated by agricultural applications. This study proposed a deep learning model based on the YOLOv8 architecture to address this challenge. Leveraging a custom dataset, it meticulously conducts model training, validation, and testing. The experimental results and performance evaluations demonstrate the efficacy of our proposed method, showcasing its ability to achieve high levels of accuracy, thus promising substantial advancements in fruit detection within the agricultural domain. Two notable limitations in the realm of fruit detection are computational resource intensity and model generalization. Firstly, deep learning-based fruit detection models often require substantial computational resources for training and inference, which may not be readily available in resource-constrained agricultural environments. Secondly, achieving model generalization across different fruit varieties, lighting conditions, and backgrounds remains a challenge, as models trained on one dataset may struggle to adapt to diverse real-world scenarios. In light of these limitations, future research could focus on addressing these challenges. Firstly, the development of more computationally efficient deep learning architectures tailored for fruit detection could help alleviate resource constraints. Secondly, exploring techniques such as domain adaptation and transfer learning to enhance model generalization across varying conditions and fruit types could lead to more robust and versatile fruit detection systems. These efforts could significantly enhance the applicability and effectiveness of fruit detection technology in agriculture.

References

[1] C. C. Ukwuoma, Q. Zhiguang, M. B. Bin Heyat, L. Ali, Z. Almaspoor, and H. N. Monday, "Recent advancements in fruit detection and classification using deep learning techniques," Math Probl Eng, vol. 2022, pp. 1–29, 2022.

[2] F. Gao et al., "A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard," Comput Electron Agric, vol. 197, p. 107000, 2022.

[3] W. Zhang et al., "Deep-learning-based in-field citrus fruit detection and tracking," Hortic Res, vol. 9, 2022.

[4] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, "Deep learning–Method overview and review of use for fruit detection and yield estimation," Comput Electron Agric, vol. 162, pp. 219–234, 2019.

[5] H. Kang and C. Chen, "Fast implementation of real-time fruit detection in apple orchards using deep learning," Comput Electron Agric, vol. 168, p. 105108, 2020.

[6] W. Zhang, K. Chen, J. Wang, Y. Shi, and W. Guo, “Easy domain adaptation method for filling the species gap in deep learning-based fruit detection,” Hortic Res, vol. 8, 2021.

[7] M. Afonso et al., “Tomato fruit detection and counting in greenhouses using deep learning,” Front Plant Sci, vol. 11, p. 571299, 2020.

[8] K. Bresilla, G. D. Perulli, A. Boini, B. Morandi, L. Corelli Grappadelli, and L. Manfrini, “Single-shot convolution neural networks for real-time fruit detection within the tree,” Front Plant Sci, vol. 10, p. 611, 2019.

[9] L. Fu, F. Gao, J. Wu, R. Li, M. Karkee, and Q. Zhang, “Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review,” Comput Electron Agric, vol. 177, p. 105687, 2020.

[10] M. C. Ang, E. Sundararajan, K. W. Ng, A. Aghamohammadi, and T. L. Lim, “Investigation of Threading Building Blocks Framework on Real Time Visual Object Tracking Algorithm,” Applied Mechanics and Materials, vol. 666, pp. 240–244, 2014.

[11] A. Aghamohammadi, M. C. Ang, E. A. Sundararajan, N. K. Weng, M. Mogharrebi, and S. Y. Banihashem, “A parallel spatiotemporal saliency and discriminative online learning method for visual target tracking in aerial videos,” PLoS One, vol. 13, no. 2, p. e0192246, 2018.

[12] X. Wang et al., “Lightweight SM-YOLOv5 tomato fruit detection algorithm for plant factory,” Sensors, vol. 23, no. 6, p. 3336, 2023.

[13] T.-H. Liu et al., “Pineapple (Ananas comosus) fruit detection and localization in natural environment based on binocular stereo vision and improved YOLOv3 model,” Precis Agric, vol. 24, no. 1, pp. 139–160, 2023.

[14] R. Gai, N. Chen, and H. Yuan, “A detection algorithm for cherry fruits based on the improved YOLO-v4 model,” Neural Comput Appl, vol. 35, no. 19, pp. 13895–13906, 2023.

[15] J. Zhou, Y. Zhang, and J. Wang, “A dragon fruit picking detection method based on YOLOv7 and PSP-Ellipse,” Sensors, vol. 23, no. 8, p. 3803, 2023.

[16] C. B. MacEachern, T. J. Esau, A. W. Schumann, P. J. Hennessy, and Q. U. Zaman, “Detection of fruit maturity stage and yield estimation in wild blueberry using deep learning convolutional neural networks,” Smart Agricultural Technology, vol. 3, p. 100099, 2023.

[17] A. Cardellicchio et al., “Detection of tomato plant phenotyping traits using YOLOv5-based single stage detectors,” Comput Electron Agric, vol. 207, p. 107757, 2023.

[18] E. Wei and S. Ren, “Drowning Detection based on YOLOv8 improved by GP-GAN Augmentation,” 2023.