Identification of Harvestable Tomatoes Using Yolov8

SRINATH M *, AKHIL RAJESH *.

* Student, Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603203, India

ABSTRACT The implementation of tomato picking robots holds immense importance for greenhouse tomato cultivation, as it significantly reduces labor requirements and enhances production efficiency. This study explores the application of Yolov8, a state-of-the-art object detection algorithm, for the precise identification of harvestable tomatoes in greenhouse settings. The yolov8model is trained and fine-tuned using a comprehensive dataset comprising diverse tomato varieties that are ripe and unripe. The proposed methodology harnesses the power of deep learning to enable real-time detection of ripe tomatoes, facilitating timely and efficient harvesting practices. Leveraging the yolov8architecture, the system demonstrates exceptional accuracy and speed, making it suitable for high-throughput greenhouse operations. Various challenges, including occlusion and varying lighting conditions, are addressed through robust training strategies, ensuring the model's reliability in practical farming scenarios. The most advanced visual tomato identification technology focuses mostly on ripe tomatoes because they stand out from the backdrop in terms of color. Furthermore, the study evaluates the performance of the yolov8model against traditional computer vision methods, highlighting the superior accuracy and efficiency of the deep learning approach. The prediction accuracy for ripe and immature tomatoes is found to be 97.23% when segmentation is applied in Yolov8. The integration of this advanced technology into greenhouse practices promises to revolutionize the way harvestable produce is identified and harvested, paving the way for sustainable and efficient agricultural processes in the future.

INDEX TERMS Yolov8, Deep learning, Classification, Dataset, Tomato detection, Convolutional neural network.

I. INTRODUCTION

Both the daily lives of people and the world economy depend heavily on agriculture. As the world's population is growing quickly and natural resources are depleting, the need for More and more evidence points to precision farming. Advanced Technologies can be applied to offer automated remedies for chores that must be done by hand to assist in precision farming. Therefore, such contemporary technologies are being integrated into farming. the scientific community is a difficult task from many scientific disciplines, including machine learning, remote sensing, Robotics, and learning. especially with reference to cropping management, incorporating cutting-edge IT tools into terms of crop forecast, farming can offer crucial aid. Species identification, sickness prognosis, etc.

As a result, there is a growing interest in creating deep learning algorithms for use in various agricultural processes. For instance, a novel strategy that focuses on crop prediction is presented by (Pantazi et al., 2016). For precise wheat production prediction, this suggested method compiles information from satellite imaging, crop growth traits, and in-situ soil measurements. (Grinblat et al., 2016) offer an algorithm for identifying and classifying legume species from leaf vein morphological features about the species recognition challenge. In addition, (Ferentinos, 2018) suggests a CNN-based method for diagnosing diseases by classifying leaves in different plants as healthy or diseased based on photographs.

The accurate detection of different crops as well as the identification of their ripening stage are considered challenging and essential tasks. For the achievement of accurate results in these tasks, the need of a high-quality labeled dataset is required. However, there is a lack of such publicly available benchmark datasets for precision farming, which restricts the efficient application of modern technologies, like machine learning algorithms, in greenhouses. To encourage and support the detection of specific crops, we introduce a dataset for tomato fruit detection.

One of the most widely consumed crops and a key component of the agricultural economy is the tomato. Tomato collection is a time-consuming and demanding job because of its planting qualities, such as a big planting area. Additionally, due to its qualities, such as its short lifespan and sensitive fruit, collecting tomatoes is a challenging and delicate process that requires precise timing. Therefore, it is crucial to build an automated collection system that can facilitate efficient tomato harvesting. Object detection methods based on pictures of tomato fruits in greenhouses play a crucial role in the study of automated collecting systems.

As a result, the availability of a high-quality and realistic tomato dataset is crucial for the accurate detection of those fruits.

In this research, we present a highly specialized, unique tomato fruit object identification and classification dataset that includes class information for each tomato fruit's ripening stage in addition to its associated bounding box. The phases of tomato ripening can indicate when they will be harvested; in this dataset, the stages are divided into three classes: unripe, semi-ripe, and fully-ripe. The TomatOD dataset is additionally freely accessible for academic research.

The yolov8algorithm, which was recently released, has gained attention for its successful application in various fields such as tomato disease detection, railway signal light detection, and smoking driver detection[]. These applications prioritize real-time detection and emphasize speed rather than the reduction of detection rates. yolov8is a deep-learning architecture specifically designed for real-time object detection[]. It utilizes global optimization techniques to improve detection speed while maintaining high accuracy. ## The Need for Tomato Detection In the agricultural industry, the ability to accurately identify and detect ripe tomatoes is crucial for efficient harvesting. Accurate tomato detection not only ensures that only fully matured tomatoes are harvested but also optimizes the harvesting process by reducing time and labor costs. Accurate and real-time detection of harvestable tomatoes is essential for efficient and cost-effective harvesting processes in the agricultural industry.

To address this need, the yolov8algorithm has been applied to the task of tomato detection. The yolov8algorithm, as a mature and efficient single-stage target detection model, is well-suited for real-time tomato detection[]. It has been observed that the yolov8algorithm provides accurate and fast detection of tomatoes in various agricultural settings[]. One of the key advantages of using yolov8for tomato detection is its real-time capability. Farmers can get instant tomato ripeness feedback, aiding timely harvesting decisions.

II. DATA ACQUISITION AND PROCESSING

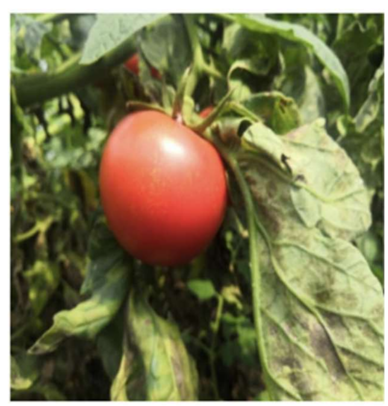
A. Data collection

The study harnessed tomato datasets obtained from Roboflow, meticulously captured at an optimal operational distance of 0.5–1.0 meters from the camera to the tomato trees in the field. This meticulous approach guaranteed the acquisition of precise data for the harvesting robot. The images were captured using a high-resolution digital commercial camera, boasting a 3968×2976-pixel resolution, RGB color space, and JPG storage format. Importantly, the data collection process was conducted under natural daylight conditions, encompassing the inherent complexities of real-world agricultural environments, such as varying illumination, occlusion caused by leaves and branches, as well as overlapping elements.

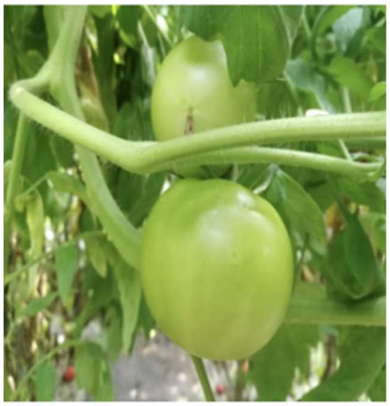
Navigating these challenging field conditions significantly heightened the complexity of distinguishing between ripe and unripe tomatoes. To streamline the subsequent deep learning processes, a curated dataset of 125 images was meticulously selected. This dataset was thoughtfully partitioned into an 80% training set and a 20% test set, encompassing diverse scenarios, including single unoccluded objects, objects partially obscured by branches and leaves, and multiple objects with or without occlusion. Furthermore, to explore the impact of resizing on the detection performance, all images were resized to 0.5 and 0.25 ratios while preserving their original aspect ratio, creating subsets categorized as Raw, 0.5 ratio, and 0.25 ratio for comprehensive training and testing analyses.

The crucial task of labeled data generation for training YOLO detection models involved manual annotation, including the assignment of class labels and precise positioning of bounding boxes around the ground truth objects in the training images. Although labeling is inherently labor-intensive, the annotation process, specifically the drawing of ground truth bounding boxes, was facilitated by the relatively small number of images in each category, minimizing the scope for human error. This meticulous annotation process was carried out using the graphical image annotation tool labelImg, with annotation files meticulously saved in YOLO format.

In each image, all visible tomatoes, both ripe and unripe, were diligently labeled with bounding boxes, employing the LWYS technique for accurate delineation. Particularly noteworthy is the handling of highly occluded tomatoes, where bounding boxes were intelligently drawn based on the presumed shape derived from visible human-readable cues, a process illustrated in Figure 3. Following the annotation process, a rigorous quality assurance protocol was implemented, involving three separate evaluations by different individuals. This multi-step verification process ensured the comprehensive annotation of all classes, leaving no room for potential oversights.



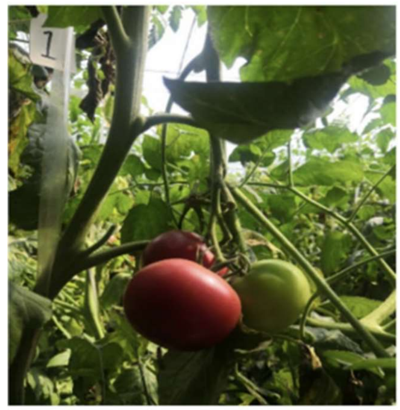
Single object with no occlusion



Multiple objects with occlusion

Illumination variation

Clusters of tomatoes



Shading conditions

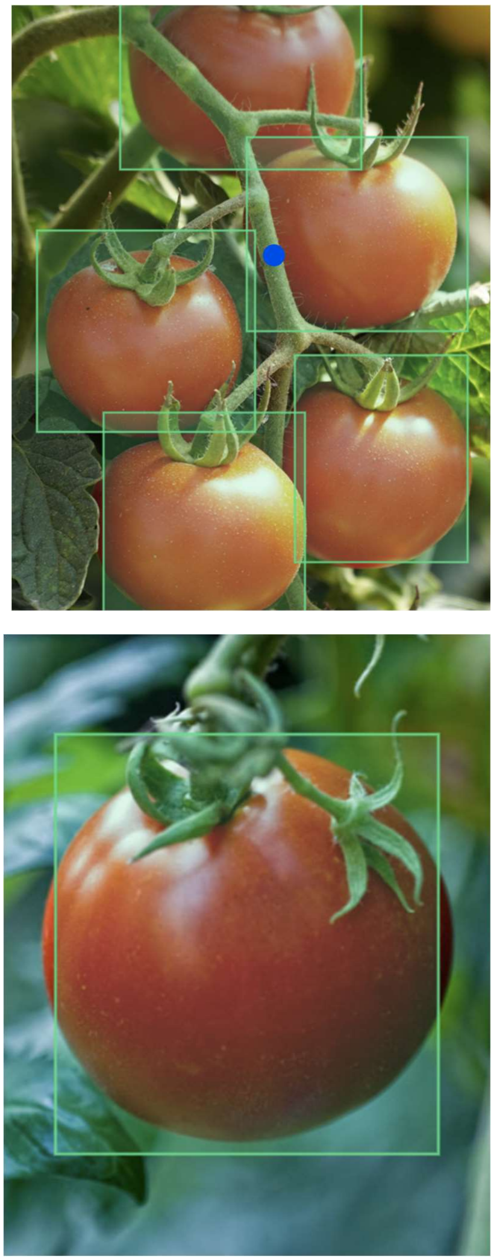
B. Build the dataset

The TomatOD dataset stands as an innovative and specialized collection designed for a highly specific task: tracking a robotic arm as it navigates through the corridors of a soilless tomato cultivation greenhouse. Its primary goal is to facilitate location mapping and estimate the ripening stages of every individual tomato fruit within the greenhouse environment. This dataset comprises a total of 277 images, each meticulously annotated by a team of two expert agriculturists. These annotations are highly detailed, covering a grand total of 2418 tomato fruit samples categorized into three distinct classes: unripe, semi-ripe, and fully ripe. This diverse and well-annotated dataset serves as a valuable benchmark for researchers and practitioners involved in the detection and classification of tomatoes within greenhouse settings.

What makes the TomatOD dataset particularly noteworthy is its realistic use-case scenario, simulating the challenges faced in real-world agricultural operations. By providing this detailed dataset, researchers and developers can explore and refine algorithms and technologies aimed at enhancing robotic operations in greenhouse environments. Moreover, the dataset's utility extends beyond its mere content; the annotations are provided in a COCO compatible format. This ensures seamless integration and compatibility with existing frameworks and tools commonly used in the field of computer vision.

For those interested in delving into the intricacies of this dataset, including its data acquisition methodologies within the greenhouse, the manual annotation procedures undertaken by the experts, and an in-depth statistical analysis, comprehensive details can be accessed via the following link: TomatOD Dataset.

After the annotation files are made for tomatoes. It is trained by Yolov8. Which in return gives the accuracy level for the dataset. In this case, it has given us a percentage of 97.32.



Annotated images of tomato

III. OBJECT DETECTION DATASETS OVERVIEW

Object Detection (OD) stands as one of the most formidable challenges in the realm of computer vision. Its objective is to precisely identify the location of specific objects within images and videos while accurately classifying them into predefined categories. This localization is conventionally represented through bounding boxes encompassing the objects of interest. The applications of OD are expansive, spanning crucial domains such as autonomous driving, object and people tracking, security, and transportation, among others.

In recent times, the widespread adoption of deep learning algorithms in the field of object detection has been significantly catalyzed by the introduction and open distribution of extensive OD datasets. These datasets, in various forms and for diverse tasks, have emerged as pivotal factors driving the advancement of object detection methodologies. Their availability has empowered researchers and practitioners, enabling them to harness the capabilities of sophisticated deep learning algorithms.

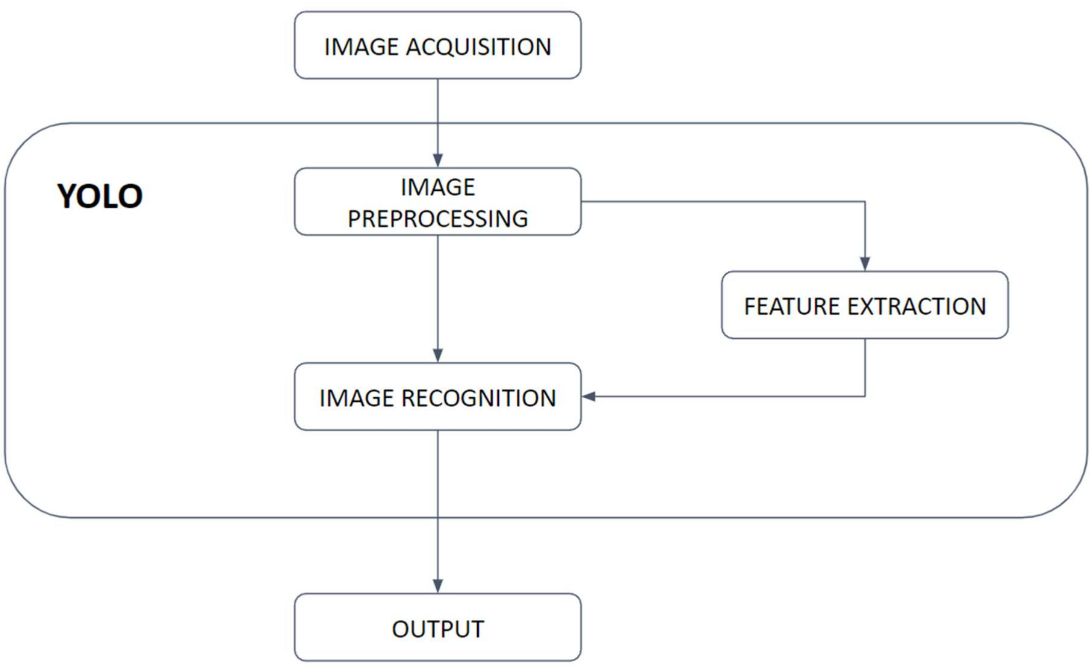
Within this context, this section delves into an exploration of some of the most notable OD datasets. These datasets serve as the lifeblood of cutting-edge research, facilitating the development and refinement of intricate algorithms. By comprehensively understanding and utilizing these datasets, the field of computer vision continues to make remarkable strides in enhancing object detection techniques and applications.

IV. YOLOV8

A. Network Model

yolov8is a member of the Yolo family of models and is an object detection network model. Joseph Redmon created the first three iterations of Yolo between 2015 and 2018. Initial evaluations revealed that yolov8had the same accuracy as Yolov4 but a faster prediction speed [35]. Glenn Jocher released it in 2020. There are five different network model variations of Yolov8: Yolov8n, Yolov8s, Yolov8m, Yolov8l, and Yolov8x. Yolov8n has the fastest calculation speed, however it also has the lowest average precision. Yolov8x, on the other hand, has the highest average precision but the slowest computation speed. The Yolov8m model was used to build the tomato state classification system in the current study.

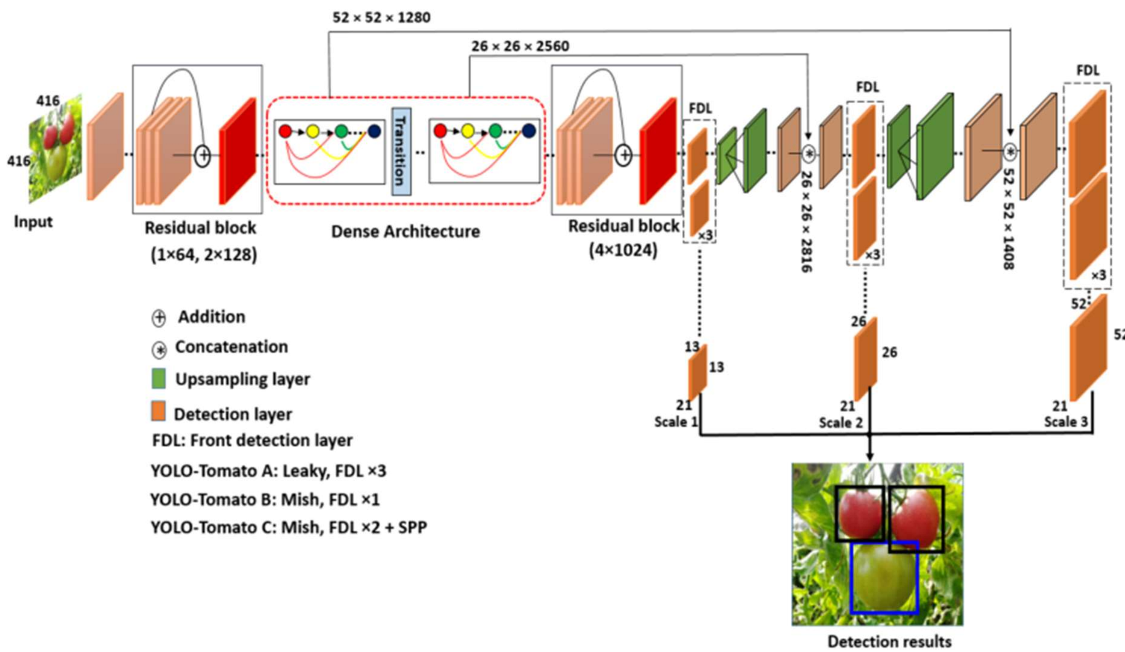
The backbone structure consisted of a Conv (Convolutional) layer, a C3 (Cross Stage Partial Networks Bottleneck with 3 convolutions) layer, and a classification layer. In total, the model consisted of 212 layers with 11.7 million parameters and 30.9 GFLOPs (Giga Floating Point Operations Per Second).



In the above yolov8network model structure diagram, “conv” represents convolution, and “concat” represents the feature fusion method of adding up the number of channels

B. TARGET DETECTION NETWORK

YOLO series of target detection algorithms has been iterated and updated continuously. Considering the model size and stability, the yolov8was selected as the recognition model in the condition of recognition accuracy close [20]. yolov8improves the network structure and training skills of the YOLOv3 algorithm. Its detection performance is equivalent to that of YOLOv4, yet the model size is reduced by nearly 90% [21,22]. yolov8contains four network structures: Yolov8s, Yolov8m, Yolov8l, and Yolov8x, of which Yolov8s has the smallest network structure, the fastest speed, and the lowest accuracy. The other three networks deepen and widen the network, with accuracy increasing accordingly yet speed also slowing down [23]. The overall network structure of the Yolov8s target detection algorithm is shown in Figure 4. Yolov8s consists of four components: input, backbone, neck, and prediction [24,25]



yolov8(You Only Look Once version 5) is an advanced deep learning model used for a diverse array of object detection assignments, including the identification of specific items like tomatoes. Building upon the YOLO architecture renowned for its real-time detection accuracy and speed, yolov8is tailored to deliver precise object detection. Leveraging extensive training datasets and a deep neural network design, it excels at precisely identifying object locations and classes within images. yolov8prioritizes real-time or near-real-time performance, making it ideal for applications requiring swift responsiveness, such as robotics and autonomous vehicles. Its capability to detect objects at varying scales is particularly beneficial for distinguishing tomatoes of different sizes within a single image. By fine-tuning a pre-trained yolov8model with your tomato image dataset, you can specialize it for tomato detection. The model predicts bounding boxes for detected objects, enabling precise tomato location in images and assigns a 'tomato' label to detected objects, simplifying tomato differentiation from other image elements. yolov8emphasizes efficiency in terms of model size and computational resources, making it accessible on relatively modest hardware. It is an open-source project, supported by an active developer and research community, ensuring continuous improvements and updates. yolov8is a potent tool for object detection, offering an effective solution for tomato detection or any similar object detection task with appropriate training data and customization.

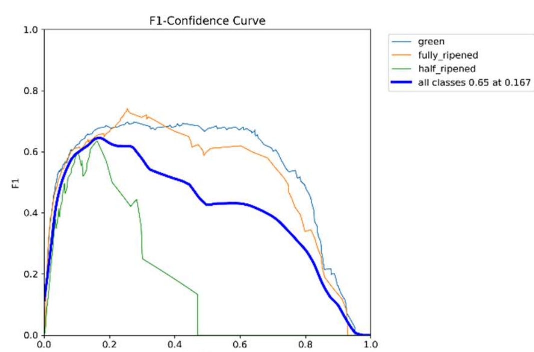
V. DETECTION AND CLASSIFICATION SYSTEM

We outline our methodology for monitoring the generative growth of tomato fruits, which involves a multi-step process. This includes deep learning-based object detection, object tracking using a centroid tracker, and assessing fruit maturity through image processing. The architecture of our approach is depicted in Figure 3. Initially, when the object detector identifies objects, we employ the k-means clustering algorithm to group the tomato fruit regions within the detected bounding boxes. Subsequently, we convert these fruit regions from the RGB color model (Red, Green, and Blue) to the HSV color model (hue, saturation, and value) and extract the hue channel image. We utilize the hue values to classify tomato fruit maturity into six levels. To prevent duplicate counting, we track the bounding boxes and assign unique identifiers using a centroid-based object tracking approach.

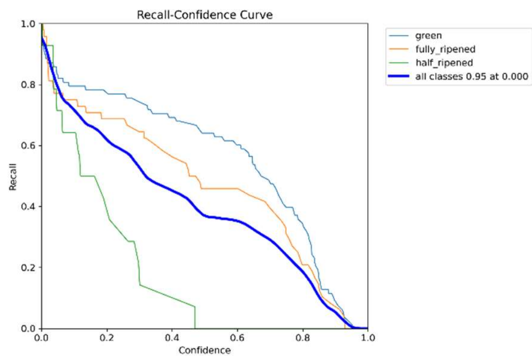
In recent years, there has been significant progress in developing powerful deep learning models for object detection. These models exhibit variations in their architecture, training procedures, optimization loss functions, and more. Broadly, modern object detection networks fall into two primary categories: one-stage detectors and two-stage detectors. One-stage detectors prioritize fast inference speed, while two-stage detectors excel in accurate object localization and recognition. The primary distinction between these categories lies in their architectural design and the methodology for predicting bounding boxes. Notably, a two-stage detector typically incorporates a region proposal step that generates candidate object bounding boxes. Subsequently, features are extracted from each candidate bounding box, and tasks involving classification and bounding-box regression are performed.

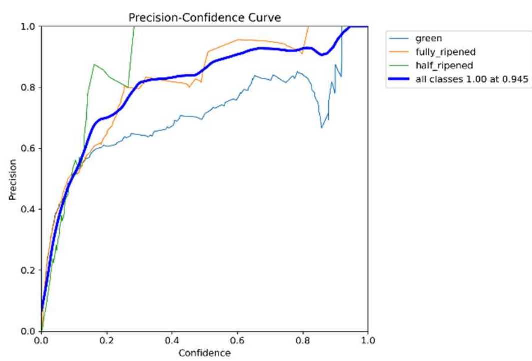
VI. RESULTS AND DISCUSSION

The utilization of deep learning models presents a multitude of advantages. First and foremost, these models eliminate the arduous task of manual feature selection, offering a more streamlined and automated approach to feature extraction. Furthermore, they exhibit commendable generalization performance, making them adept at handling diverse data and extrapolating patterns. Importantly, as the volume of data at their disposal increases, the accuracy of deep learning models tends to see a notable improvement, making them increasingly reliable.



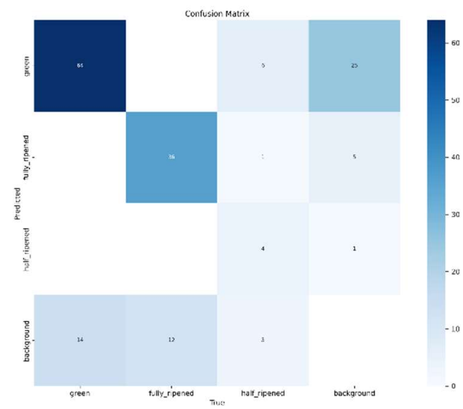
However, it is not without its challenges. Deep learning models, particularly in their pre-trained forms, often come with weight files of substantial size. These oversized weight files can lead to bottlenecks in real-time detection, causing delays that can be especially problematic in time-sensitive applications. Moreover, deep learning models might struggle with certain recognition tasks, particularly when it comes to small targets or objects that are partially obscured. These challenges have prompted the need for innovation and refinement in this field.





This is where the tomato maturity recognition model Yolov8s-tomato, developed as part of our study, comes into play. Built upon the foundation of the improved yolov8architecture, this model exhibits the ability to discern between four distinct levels of tomato ripeness within a controlled greenhouse environment. In contrast to similar studies, this experiment segments tomato maturity into four distinct levels, aligning with the practical needs of tomato harvesting, which often necessitates the selection of tomatoes at varying degrees of tomatoes.

The merits of the Yolov8s-tomato model shine through in several ways. In comparison to laboratory recognition models, this model boasts swifter recognition capabilities, a more compact model size, and reduced hardware requirements, making it a cost-effective and efficient choice. When stacked against other recognition models commonly used in greenhouse settings, it distinguishes itself with higher recognition precision and a notably smaller model size – a reduction of nearly 80% compared to its counterparts. Moreover, the Yolov8s-tomato model excels in detecting tomatoes, even in the presence of occlusion and when dealing with smaller targets, surpassing the capabilities of



other models in this regard.

The benefits of our model become even more evident when considering its potential for embedded deployment or use in mobile equipment for greenhouse tomato picking, aligning well with real-world applications and practical scenarios. However, it's important to acknowledge that no model is without limitations. In this experiment, recognition errors predominantly occurred during the transition stages of tomato maturity, where breaker-stage fruits were occasionally misclassified as mature green, and pink stage tomatoes were misidentified as breaker stage. These errors primarily stem from the continuous nature of tomato ripening, with rapid transformations occurring, often with limited data points to guide the model. Nonetheless, these occasional misclassifications have minimal impact on the broader tomato maturity recognition process.

This experiment represents just the beginning of our exploration into tomato maturity recognition. Practical applications come with certain constraints. Firstly, the Yolov8s-tomato model, as proposed in the study, is optimally suited for tomato varieties with red fruits during the maturation process. For tomato varieties with differently colored fruits at the maturation stage, relabeling of data and retraining of the model are necessary prerequisites for its effective application. Secondly, the images used for training in this experiment were predominantly sourced from a controlled greenhouse environment, sharing similar characteristics with greenhouse images. However, the situation may differ when applying this model to tomatoes grown outdoors, where the backdrop is more complex and lighting conditions are variable. This prompts the need for further validation to ensure the model's performance remains robust in varying environments.

In conclusion, it is crucial to recognize that deep learning algorithms thrive on data. To enhance the maturity recognition accuracy, resilience, and generalization of the Yolov8s-tomato model, a diverse range of images will be essential. Therefore, future studies will focus on acquiring images from a variety of greenhouses and outdoor settings to broaden the model's dataset and foster its adaptability across diverse scenarios.

VII. CONCLUSION

In this experimental study, we introduce the tomato maturity recognition model, Yolov8s-tomato, which builds upon the foundations of the enhanced yolov8architecture. The development of this model is underpinned by a meticulous understanding of the various maturity stages of tomato fruits. To facilitate our research, we construct a specialized tomato maturity dataset using greenhouse tomatoes as our primary data source. The model's efficacy in detecting small targets is substantially enhanced through a range of augmentation techniques, including Mosaic data enhancement.

In the predictive phase, we replace the initial CIoU Loss with the EIoU Loss, a strategic decision aimed at diminishing the discrepancies between the ground truth box and the predicted box. As a result, our experimental findings reveal an impressive tomato maturity recognition precision of 95.58%, a recall rate of 90.07%, and a mean average precision (mAP) of 97.42%. The compact model size of merely 23.9 megabytes and a speedy detection time of just 9.2 milliseconds per image underscore the efficiency of our Yolov8s-tomato model.

Given the unique challenges posed by greenhouse settings, where tomato branches, leaves, and fruits intermingle and create a relatively dense environment, we subject our Yolov8s-tomato model to a comprehensive comparative analysis alongside Yolov8s, Yolov8m, Yolov8l, Yolov8x, and Faster R-CNN. The results of this comparative experiment underscore the models' capacity to excel in recognizing tomato fruit maturity under conditions of shading, occlusion, and densely distributed objects. Notably, both the Yolov8s-tomato and Yolov8x models outperform their counterparts, with Faster R-CNN lagging in recognition efficacy. While Yolov8s-tomato exhibits a slight reduction in precision, recall, and mAP compared to Yolov8x, the model compensates for this with a significantly reduced model size and a more efficient per-image detection time, coming in at only 8.88% and 34.06% of Yolov8x, respectively.

When juxtaposed with traditional methods, Yolov8s-tomato emerges as a robust and reliable solution, accurately identifying tomato fruit maturity even in scenarios characterized by the occlusion of branches, leaves, and fruits. In comparison to other deep-learning models, our Yolov8s-tomato boasts higher precision, a clear edge in recognition speed, and a smaller model footprint. Given the practical constraints and recognition efficiency requirements of greenhouse mechanical picking equipment, the Yolov8s-tomato algorithm emerges as the ideal candidate for integration with tomato picking machines in greenhouse settings.

VIII. REFERENCES

1. Linker, R.; Cohen, O.; Naor, A. Determination of the number of green apples in RGB images recorded in orchards. Comput. Electron. Agric. 2012, 81, 45–57.

2. Wei, X.; Jia, K.; Lan, J.; Li, Y.; Zeng, Y.; Wang, C. Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot. Optik 2014, 125, 5684–5689.

3. Kelman, E.E.; Linker, R. Vision-based localisation of mature apples in tree images using convexity. Biosyst. Eng. 2014, 118, 174–185.

4. Payne, A.; Walsh, K.; Subedi, P.; Jarvis, D. Estimating mango crop yield using image analysis using fruit at 'stone hardening'stage and night time imaging. Comput. Electron. Agric. 2014, 100, 160–167.

5. Qiang, L.; Jianrong, C.; Bin, L.; Lie, D.; Yajing, Z. Identification of fruit and branch in natural scenes for citrus harvesting robot using machine vision and support vector machine. Int. J. Agric. Biol. Eng. 2014, 7, 115–121.

6. Kurtulmus, F.; Lee, W.S.; Vardar, A. Immature peach detection in colour images acquired in natural illumination conditions using statistical classifiers and neural network. Precis. Agric. 2014, 15, 57–79.

7. Yamamoto, K.; Guo, W.; Yoshioka, Y.; Ninomiya, S. On plant detection of intact tomato fruits using image analysis and machine learning methods. Sensors 2014, 14, 12191–12206.

8. Zhao, Y.; Gong, L.; Zhou, B.; Huang, Y.; Liu, C. Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis. Biosyst. Eng. 2016, 148, 127–137.

9. Luo, L.; Tang, Y.; Zou, X.; Wang, C.; Zhang, P.; Feng, W. Robust grape cluster detection in a vineyard by combining the AdaBoost framework and multiple color components. Sensors 2016, 16, 2098.

10. Liu, G.; Mao, S.; Kim, J.H. A mature-tomato detection algorithm using machine learning and color analysis. Sensors 2019, 19, 2023.

11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 2012, 25, 1097–1105.

12. Szegedy, C.; Wei, L.; Yangqinh, J.; Pierre, S.; Scott, R.; Dragomir, A.; Dumitru, E.; Vincent, V.; Andrew, R. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.

13. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.

14. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations

for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.

15. Zhang, K.; Wu, Q.; Liu, A.; Meng, X. Can deep learning identify tomato leaf disease? Adv. Multimed. 2018, 2018, 6710865.

16. Aversano, L.; Bernardi, M.L.; Cimitile, M.; Iammarino, M.; Rondinella, S. Tomato diseases Classification Based on VGG and Transfer Learning. In Proceedings of the 2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), Trento, Italy, 4–6 November 2020; pp. 129–133.

17. Karthik, R.; Hariharan, M.; Anand, S.; Mathikshara, P.; Johnson, A.; Menaka, R. Attention embedded residual CNN for disease detection in tomato leaves. Appl. Soft Comput. 2020, 86, 105933.

18. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

19. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 2016, 39, 1137–1149.

20. Dai, J.; Li, Y.; He, K.; Sun, J. R-FCN: Object detection via region-based fully convolutional networks. Adv. Neural Inf. Process. Syst. 2016, 29, 379–387.

21. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.

22. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

23. Hu, C.; Liu, X.; Pan, Z.; Li, P. Automatic detection of single ripe tomato on plant combining faster R-CNN and intuitionistic fuzzy set. IEEE Access 2019, 7, 154683–154696.

24. Fuentes, A.; Yoon, S.; Kim, S.C.; Park, D.S. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors 2017, 17, 2022.

25. Mirhaji, H.; Soleymani, M.; Asakereh, A.; Mehdizadeh, S.A. Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. Comput. Electron. Agric. 2021, 191, 106533.

26. Chen, J.; Wang, Z.; Wu, J.; Hu, Q.; Zhao, C.; Tan, C.; Teng, L.; Luo, T. An improved Yolov3 based on dual path network for cherry tomatoes detection. J. Food Process Eng. 2021, 44, e13803.

27. Elsayed, E.; Amany, S. Evaluation of nutritional value and antioxidant activity of tomato peel extracts. Arab. J. Chem. 2016, 9 (Suppl. 2), S1010–S1018.

28. Lin, H.T. Cherry tomato 'TSS ASVEG No.22';Taiwan Seed Improvement and Propagation Station, COA: Taichung, Taiwan, 2017.

29. Wang, P.; Niu, T.; He, D. Tomato Young Fruits Detection Method under Near Color Background Based on Improved Faster R-CNN with Attention Mechanism. Agriculture 2021, 11, 1059.

30. Zu, L.; Zhao, Y.; Liu, J.; Su, F.; Zhang, Y.; Liu, P. Detection and Segmentation of Mature Green Tomatoes Based on Mask R-CNN with Automatic Image Acquisition Approach. Sensors 2021, 21, 7842.

31. Lawal, M.O. Tomato detection based on modified YOLOv3 framework. Sci. Rep. 2021, 11,1–11.

32. Liu, G.; Nouaze, J.C.; Touko Mbouembe, P.L.; Kim, J.H. YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3. Sensors 2020, 20, 2145.

33. Xiao, J.R.; Chung, P.C.; Wu, H.Y.; Phan, Q.H.; Yeh, J.L.A.; Hou, M.T.K. Detection of strawberry diseases using a convolutional neural network. Plants 2020, 10, 31. [PubMed

34. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. Scaled-yolov4: Scaling cross stage partial network. In Proceedings of the IEEE/cvf Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 13029–13038.

35. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH- Yolov8: Improved yolov8based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2778–2788.

36. Chen, Z.; Wu, R.; Lin, Y.; Li, C.; Chen, S.; Yuan, Z.; Chen, S.; Zou, X. Plant Disease Recognition Model Based on Improved Yolov8. Agronomy 2022, 12, 365.

37. Kamilaris, A.; Prenafeta-Boldú, F.X. A review of the use of convolutional neural networks in agriculture. J. Agric. Sci. 2018, 156, 312–322.

38. Mukti, I.; Biswas, D. Transfer Learning Based Plant Diseases Detection Using ResNet50. In Proceedings of the 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 20–22 December 2019; pp. 1–6.

39. Reddy, A.S.B.; Juliet, D.S. Transfer learning with ResNet-50 for malaria cell-image classification. In Proceedings of the 2019 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 4–6 April 2019; pp. 0945–0949.

40. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–777.

41. Ümit, A.; Murat, U.; Kemal, A.; Emine, U. Plant leaf disease classification using EfficientNet deep learning model. Ecol. Inform. 2021, 61, 101182.

42. Arun, Y.; Viknesh, G.S. Leaf Classification for Plant Recognition Using EfficientNet Architecture. In Proceedings of the 2022 IEEE Fourth International Conference on Advances in Electronics, Computers and Communications (ICAECC), Bengaluru, India, 10–11 January 2022; pp. 1–5.

43. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.

44. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520