Detection of Pomelo in Overlapping Conditions Using Drones

Mahdaniar[a] and Indrabayu*[b]
Department of Informatics Hasanuddin University Makassar, Indonesia

Keywords: Detection, Pomelo, YOLOv11, CLAHE, Overlapping.

Abstract: Detection of Pomelo on trees in overlapping conditions and the similarity of colors between fruits and leaves are the main challenges in the implementation of smart farming systems. This study aims to develop an automatic detection system of Pomelo using a drone with a computer vision approach based on the YOLOv11 algorithm combined with the CLAHE (Contrast Limited Adaptive Histogram Equalization) image contrast enhancement technique. The research methodology includes image data collection, pre-processing, data labeling, model training, and evaluation using mAP, precision and recall. The initial results showed that YOLOv11 provide suboptimal performance in the detection process, resulting in the result, precision: 92%, recall: 81%, mAP50: 90%, mAP50-95: 72%. After YOLOv11 is integrated with CLAHE, the performance has been improved, achieving precision: 92%, recall: 84%, mAP50: 95%, mAP50-95: 67%.

1 INTRODUCTION

The Industrial Revolution 4.0 encourages the transformation of automation systems that are smarter, more efficient, and sustainable, driven by the development of digital technology, including in agriculture. One of the leading agricultural commodities is horticulture (Indrabayu et al., 2019a). Horticultural crops that have the potential to be developed in Indonesia are citrus commodities. Citrus (Citrus Sp) is one of the horticultural commodities that functions as a source of national income for the country, the fruit originating from Asia that can grow in tropical and subtropical areas (Addi et al., 2021). This is marked by the increasing consumption of pomelos in Indonesia from year to year. In general, pomelos are a o vitamin C source useful for human health. Pomelo juice contains 40-70 mg of vitamin C per 100 g (Khattak et al., 2021). In addition to pomelo flesh, the properties and benefits of pomelos are also contained in pomelo peels. The content of pomelo peel has benefits ranging from sedatives, skin smoothers, mosquito repellent (Adelina et al., 2020) (Sugadev et al., 2020).

The availability of computer vision technology, supported by the improvement of computer hardware capabilities, has become a major supporting factor in the development of automated farming systems. These systems are designed to solve various problems in the agricultural sector with high level of flexibility, effectiveness, and efficiency. This development has also led to an increase in the number of studies focusing on automatic fruit detection (Indrabayu et al., 2019b).

Previous research, a system for classifying ripe, immature and rotten pomelos quickly and efficiently by 335 Images taken using a 1024x768 pixel resolution digital camera are stored in JPG format using Naïve Bayes, Decision Tree and Neural Network methods (Wajid et al., 2018), automatic classification of pomelos at the fruit packaging factory with 120 images and three categories of pomelos, namely Bam, Payvandi, Thomson using the Gradient Descent, Stochastic Gradient Descent (SGD), RMS Prop and Adam methods, the results of the study obtained the best SGD method with an optimization of 0.95 (Pathak et al., 2020) detection lime fruit on a tree for estimation of yield using the YOLOv5-CS method using a dataset of 3000 images with a detection accuracy of 96.7% (Lyu et al., 2022a), pomelo fruit color detection based on frame selection detection on 811 pomelo fruit images using the YOLOv5 method obtained an accuracy rate of 99.4% [9], system for the classification of pomelo quality based on the characteristics specified on 953

[a] https://orcid.org/0009-0005-3135-5861
[b] https://orcid.org/0000-0003-2026-1809

Mahdaniar, and Indrabayu,
Detection of Pomelo in Overlapping Conditions Using Drones.
DOI: 10.5220/0014266900004928

In *Proceedings of the 1st International Conference on Research and Innovations in Information and Engineering Technology (RITECH 2025)*, pages 25-31
ISBN: 978-989-758-784-9

images of pomelo leaves using Convolutional Neural Network (CNN) with a 93.8% accuracy rate (Asriny et al., 2020) (Xu et al., 2023a), detection of leaf-covered raspberry fruits, which interfere with crop estimation and garden efficiency, with the limitations of traditional technology and challenges in the use of hyperspectral technology using Logistic Regression (LR), Random Forest (RF) methods (Chen et al., 2024), and detection of tomato fruit ripeness under conditions of changing lighting, obstruction, and fruit overlap, as well as models that are too large for limited devices using the GFS-YOLOv11 method obtained results of Precision (P): 92.0%, Recall (R): 86.8%, mAP50: 93.4%, mAP50-95: 83.6% (Wei et al., 2024).

Currently, most research focuses on improving detection by color and from different fruit targets. (Indrabayu et al., 2017) (Janowski et al., 2021), however, studies related to accurate detection of green pomelos in gardens have received less attention because they are complicated and not easy, especially the detection of green pomelos, which are very important for predicting garden yields (Lyu et al., 2022b).

Fruit detection with deep learning can be done using the computer vision method (Xu et al., 2023b), You Only Look Once (YOLO) is a single-stage object detector that has demonstrated excellent performance for detecting with high precision and accuracy (Liu et al., 2018), and with the most developed versions. YOLOv11, the latest model from Ultralytics, demonstrates superior performance on a wide range of computer vision tasks, including object detection, instance segmentation, feature extraction, pose estimation, object tracking, and classification, to real-time object detection (Hidayatullah et al., 2025). Studies to address the problem of low detection accuracy in complex plantation environments (such as varying lighting conditions, branch and leaf occlusion, fruit overlap, and small targets) used YOLOv11 in detecting occluded pears and the accuracy results showed precision, recall, mAP50, and mAP50–95 values of 96.3%, 84.2%, 92.1%, and 80.2%, respectively [Zhang et al., 2025).

Based on the observations in the pomelo field, the main obstacle is difficult to distinguish the color of pomelos and leaves, particularly in conditions of uneven illumination, overlapping fruits, or leaf occlusion. This makes manual detection difficult, time-consuming, and lowers the accuracy of fruit ripeness classification, which impacts harvest efficiency. To overcome this problem, a pomelo detection and classification system was developed using the YOLOv11 algorithm with the addition of CLAHE, to support smart farming practices and improve sustainability of agricultural production.

Beyond model development, the proposed system can be deployed on drones to enable real-time monitoring in orchards. Running inference on drones requires lightweight models that can operate under limited computational resources. Previous work has shown that compact models such as YOLOv4-tiny achieve significantly faster inference while maintaining satisfactory accuracy, making them suitable for onboard drone applications (Mpouziotas et al., 2023). This integration ensures practical field implementation while enhancing the scalability and efficiency of precision agriculture systems.

2 MATERIALS AND METHODS

The entire research process, from data preparation to fruit detection, is illustrated in the flowchart in Figure 1. This flowchart provides a comprehensive overview of the methodological workflow implemented in this study. Here is the workflow process from 1 to 8.

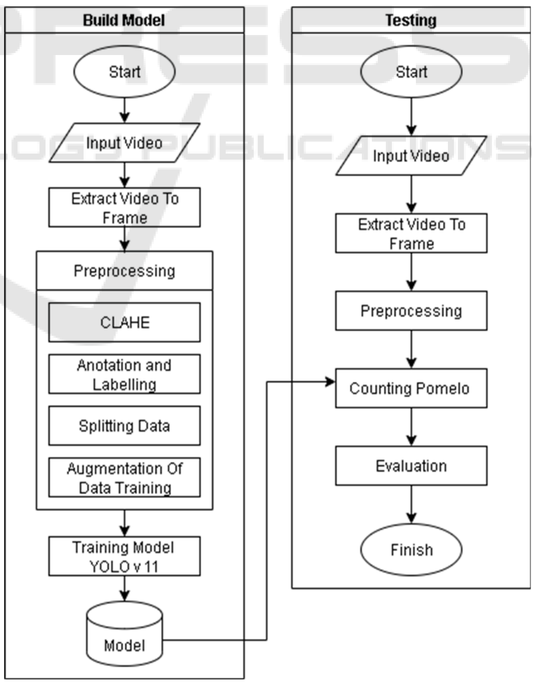

Figure 1: System Design Flow.

The dataset was partitioned into three subsets, consisting of 1,301 images for training, 124 images for validation, and 62 images for testing. However, the relatively small percentage of validation and testing data may introduce bias in the evaluation results, as limited samples can lead the model to appear to perform well while its performance may not generalize to more diverse real-world conditions (Brigato et al., 2020). The picture above illustrates the design of the proposed system for detecting pomelos on trees in challenging conditions where the color and edges of pomelos are similar to those of leaves, which can be divided into several main stages as follows:

2.1 Video Input and Extract Video Frames

The system starts by receiving video as input. This video likely contains footage that shows pomelos in various conditions, such as covered in leaves or stacked. From the given video, the system extracts the frame (frame-by-frame extraction). It means that the video is broken down into several static images so that it can be further processed in the next stage.

2.2 Preprocessing

After the frame is extracted, a preprocessing stage is carried out to improve the quality of the data before it is fed into the detection model. One of the methods mentioned in the diagram is CLAHE (Contrast Limited Adaptive Histogram Equalization), which is useful for improving image contrast, especially in suboptimal lighting conditions.

In this study, CLAHE was used with the parameters Clip Limit 2.0 and Tile Grid Size (32, 32). Clip Limit 2.0 controls the extent to which contrast is enhanced. This value was chosen because it provides a significant contrast enhancement without causing excessive effects such as the appearance of bright spots (noise amplification). Meanwhile, Tile Grid Size (32, 32) determines the size of the image division grid. With this size, the image is divided into small blocks with sufficient detail so that local contrast in certain areas can be enhanced. If the grid is too small, the result can increase noise, while a grid that is too large actually makes important details less visible. The combination of these two parameters allows CLAHE to produce images with clearer details, thus helping object detection models like YOLO to recognize oranges more accurately.

(a) (b)

Figure 2: Sample dataset (a) before applying CLAHE and (b) after applying CLAHE

2.3 Annotations

Data annotation refers to the process of labelling or adding information to different types of data, including text, images, and videos. At this stage, the processed frames are manually or semi-automatically annotated to mark the pomelo object. This annotation process is important to properly train the pomelo detection model.

2.4 Data Splitting

Data splitting is a method to divide a dataset into three parts: data training, validation, and testing. It is one of several factors that affects the performance of model. The proportions used were 70% for the training set, 20% for the test set and 10% for the validation set.

2.5 Training Model

The pomelo detection model was trained using YOLO (You Only Look Once), precisely the version of YOLOv11 mentioned in the diagram. This algorithm works by processing an image only once and immediately generating predictions in the form of bounding box locations, confidence scores, and object classes. YOLOv11 introduces significant improvements to its architecture, such as a more efficient backbone for feature extraction, a spatial attention module for capturing details of both small and large objects, and an anchor-free and decoupled detection head, resulting in more stable, faster, and more accurate predictions. These advantages make YOLOv11 highly effective for detecting oranges under various conditions, such as when the fruit is close together, covered by leaves, or in less than ideal lighting, making it highly suitable for automated fruit detection and counting tasks.

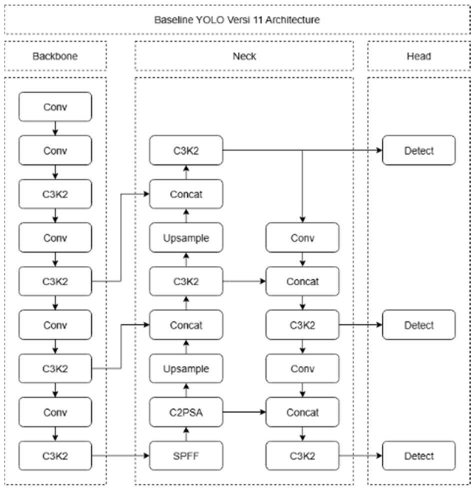

Figure 3: YOLOv11 detection architecture.

The YOLOv11 architecture consists of three main parts: the backbone, the neck, and the head. The backbone extracts feature from the input image through convolutional layers and C3K2 blocks, which efficiently capture visual patterns. Feature from the backbone are then processed in the neck, which combines information from multiple scales with unsampled and concat operations, as well as SPPF and C2PSA modules to expand the spatial context and emphasize important areas. The results of this processing are forwarded to the anchor-free, decoupled head, resulting in more stable class and bounding box location predictions. The head feature three detection branches at different scales, allowing YOLOv11 to detect small, medium, and large objects with high accuracy.

Table 1: Comparison of this study with previous studies.

Study / Method	Dataset / Target	Approach	Reported Performance
Anis Ilyana et al., 2025	Coffee	YOLO11	Ripe fruit: mAP50 = 77.4%, Precision = 64.5%, Recall = 81.2% Half-ripe fruit: mAP50 = 69.5%, Precision = 62.4%, Recall = 67.9%
Dhungana, P et al., 2025	subsea pipeline	YOLOv8n and YOLOv11n+CLAHE	YOLOv11n without enhancement: mIoU = 70.98%, Dice = 81.29% YOLOv11n with CLAHE: mIoU = 70.48%, Dice = 80.77%
Sapkota et al., 2025	Apple	YOLOv11 and CBAM	Result Trunk: 83% precision (with CBAM) vs 80% (without CBAM). Result Branch: 75% precision (with CBAM) vs 73% (tanpa CBAM)
Ours	Pomelo	YOLOv11 and CLAHE	Best performance YOLOv11 without CLAHE: Precision = 85%, Recall = 56%, mAP50 = 72%. Best performance YOLOv11 with CLAHE: Precision = 83,8%, Recall = 78,2%, mAP50 = 85,1%.

Table 2: Configuration of the training model.

Dataset	Epoch	Precision	Recall	Loss Train	Loss Validation	mAP 50	mAP 50-95
YOLOV8	100	0.7200	0.5400	0.8607	0.9366	0.6955	0.5014
	200	0.8300	0.5133	0.8755	0.9122	0.7144	0.5288
	300	08009	0.4703	0.8679	0.9324	0.6700	0.5000
Baseline (YOLOV11)	100	0.7353	0.5603	0.9435	0.9558	0.7019	0.5185
	200	0.8502	0.5377	0.9072	0.9393	0.7203	0.5387
	300	0.8124	0.4889	0.9021	0.9545	0.6740	0.5087
CLAHE + YOLOV11	100	0.7851	0.7795	1.0578	0.9916	0.8379	0.5626
	200	0.8233	0.7821	0.9950	0.9707	0.8471	0.5579
	300	0.8385	0.7665	0.9906	0.9677	0.8506	0.5741

2.6 Evaluation

After the model is trained, the evaluation stage aims to assess its performance in detecting pomelos. The assessment is conducted using metrics such as mAP (mean Average Precision) to measure how well the model can recognize objects in the image. In addition, a confusion matrix is employed to calculate precision and recall, providing a more detailed overview of the model's ability to distinguish between correct and incorrect detections.

2.7 Build Model Testing

The trained models were tested with new videos to see how well the models could detect pomelos in more real conditions. These trials aim to measure the reliability of the model before it is implemented in a real environment.

3 RESULTS AND DISCUSSION

3.1 Training Perfomance of Yolov11 and (Yolov11 + Clahe)

In this study, the YOLOv11 model was trained with parameter configurations designed to balance convergence speed and generalization capabilities. The initial learning rate was set at 0.01 with a gradual decrease scheme to maintain weight update stability during the training process. A momentum parameter of 0.937 was applied to accelerate convergence by utilizing the previous gradient direction, while a weight decay of 0.0005 was used as a regularization mechanism to reduce the risk of overfitting. Furthermore, a batch size of 16 was selected considering the limitations of CPU-based computing, while the training process was carried out for 300 epochs so that the model had the optimal opportunity to achieve the best performance. These parameters were used consistently in both the baseline YOLOv11 model training and the YOLOv11 model with the CLAHE method to ensure a fair comparison in performance evaluation.

Table 1 Based on the comparison table, it can be concluded that each study produced different performance outcomes due to variations in research objects, dataset conditions, and applied methods. Anis Ilyana et al. (2025) demonstrated that the ripeness level of coffee significantly affects detection accuracy, with ripe fruits yielding better results than half-ripe fruits. Dhungana et al. (2025) showed that the application of CLAHE in subsea pipeline detection does not always provide benefits, and in fact, slightly reduced performance compared to results without CLAHE. Meanwhile, Sapkota et al. (2025) proved that the addition of the CBAM module improved detection precision for apples, particularly in more complex areas such as the trunk and branches.

In this study, the use of CLAHE with YOLOv11 showed a significant improvement in recall and mAP50 for pomelo detection. This indicates that CLAHE helps the model identify more objects that might otherwise be difficult to detect, albeit with a slight compromise in prediction precision. Overall, this comparison highlights that the effectiveness of additional methods such as CLAHE or CBAM largely depends on the type of object, image quality, and dataset characteristics.

The table 2 shows the results of the evaluation of the YOLOv11 model on two types of datasets: original and those that have been processed using CLAHE. The model with the original dataset yielded the highest precision off 0.8502 and the 50th mAP of 0.72032 at the 200th epoch, but the recall was low (0.5377), indicating many undetected objects.

.In contrast, the use of CLAHE significantly improves performance. Precision and recall were more stable, reaching 0.8385 and 0.7665, respectively, on the 300th epoch. The values of mAP 50 and mAP 50-95 were also higher, at 0.8506 and 0.5741. These results prove that CLAHE preprocessing helps improve the detection of low- contrast objects, making the model more accurate and reliable

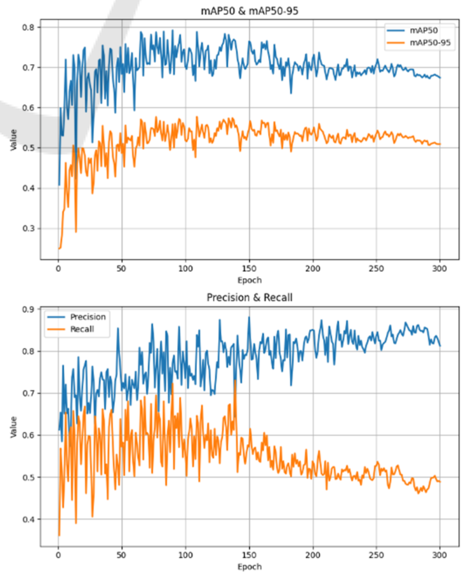

Figure 4: Yolov11 training results graph (mAP 50, mAP 50-95, precision, and recall) with 300 epochs.

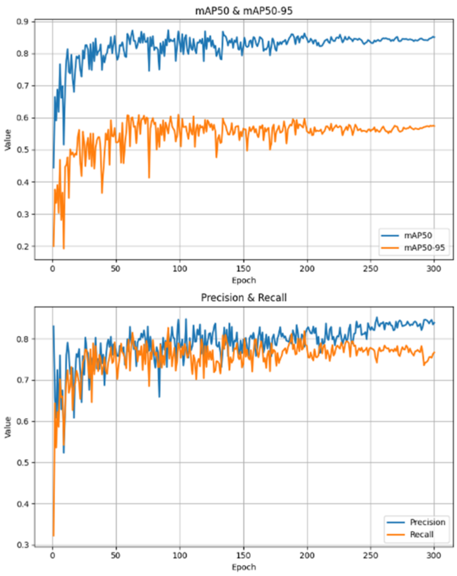

Figure 5: Graph of Yolov11+ Clahe training results (mAP 50, mAP 50-95, precision, and recall) with 300 epochs.

3.2 Testing Performance of Yolov11 dan (Yolov11 + Clahe)

The table 3 shows the results of the evaluation of the YOLOv11 model on two types of datasets: original and those that have been processed using CLAHE. The model with the original dataset yielded the highest precision of 0.8502 and the 50th mAP of 0.72032 at the 200th epoch, but the recall was low (0.5377), indicating many undetected objects.

In contrast, the use of CLAHE significantly improves performance. Precision and recall were more stable, reaching 0.8385 and 0.7665 respectively on the 300th epoch. The values of mAP 50 and mAP 50-95 were also higher, at 0.8506 and 0.5741. These results prove that CLAHE preprocessing helps improve the detection of low-contrast objects, making the model more accurate and reliable.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Explanation:
FN = The model fails to detect an object even though it exists.
TP = The model successfully
detects an object and the object does indeed exist.
TN = The model does not detect objects, and they do not exist.
FP = The model detects an object when it doesn't exist.

Table 3: Testing Configuration.

Dataset	Baseline (Yolov8)	Baseline (Yolov11)	Clahe + Yolov11
TP	315	328	412
FP	16	11	113
FN	145	118	34
Precision	0.921	0.929	0.921
Recall	0.803	0.815	0.841
mAP 50	0.894	0.907	0.936
mAP 50-95	0.717	0.721	0.67

4 CONCLUSIONS

This research successfully developed a pomelo detection system on trees using the YOLOv11 algorithm combined with the CLAHE contrast enhancement method. The results of the experiment showed that preprocessing using CLAHE was able to improve the accuracy of fruit detection, especially in conditions of overlap and color similarity between fruits and leaves. The increased mAP, precision, and recall values after the use of CLAHE prove the effectiveness of this approach in dealing with challenges in complex plantation environments. This system can be a supporting solution in the implementation of smart farming, especially to improve the efficiency and accuracy of the automatic pomelo harvesting process.

ACKNOWLEDGEMENTS

The authors would like to thank the Master's Program in Informatics Engineering, Hasanuddin University, and the Artificial Intelligent and Multimedia Processing (AIMP) Thematic Research Group for their support and facilities during this research. The authors also acknowledge colleagues who provided constructive feedback and assistance throughout the research process. In addition, the authors acknowledge the use of a generative AI tool in improving the clarity and grammar of this manuscript. All contents, analyses, and conclusions remain the sole responsibility of the authors.

REFERENCES

Addi, A. Elbouzidi, M. Abid, D. Tungmunnithum, A. Elamrani, and C. Hano, “An Overview of Bioactive Flavonoids from Pomelos,” *Appl. Sci.*, vol. 12, no. 1, p. 29, Dec. 2021, doi: 10.3390/app12010029.

Adelina and E. Adelina, “Identification of Morphology and Anatomy of Local Pomelos (Citrus sp) in Doda Village and Lempe Village, Central Lore District, Poso Regency.”

Asriny, S. Rani, and A. F. Hidayatullah, “Pomelo Fruit Images Classification using Convolutional Neural Networks,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 803, no. 1, p. 012020, Apr. 2020, doi: 10.1088/1757-899x/803/1/012020.

Brigato, L., & Iocchi, L. (2021, January). A close look at deep learning with small data. In *2020 25th international conference on pattern recognition (ICPR)* (pp. 2490-2497). IEEE.

Chen, J. Wang, R. Xi, and Z. Ren, “Analysis of Leaf cover on Raspberry Fruits Based on Hyperspectral Techniques Combined with Machine Learning Models,” Jul. 15, 2024, Springer Science and Business Media LLC. doi: 10.21203/rs.3.rs-4607290/v1.

Dhungana, P., Fresta, M., Tamrakar, N., & Dhungana, H. (2025, Juni 30). YOLO-Based Pipeline Monitoring in Challenging Visual Environments (arXiv:2507.02967v1). https://doi.org/10.48550/arXiv.2507.02967

Hidayatullah, N. Syakrani, M. R. Sholahuddin, T. Gelar, and R. Tubagus, “YOLOv8 to YOLO11: A Comprehensive Architecture In-depth Comparative Review.”

Ilyana, A., Nurdin, N., & Maryana, M. (2025). Real-Time Detection of Coffee Cherry Ripeness Using YOLOv11. Journal of Applied Informatics and Computing, 9(4). https://doi.org/10.30871/jaic.v9i4.9735.

Indrabayu, Mar’atuttahirah, and I. S. Areni, “Automatic Counting of Chili Ripeness on Computer Vision for Industri 4.0,” in *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, BALI, Indonesia: IEEE, Jul. 2019, pp. 14–18, doi: 10.1109/iciaict.2019.8784858.

Indrabayu, A. R. Fatmasari, and I. Nurtanio, “A Colour Space Based Detection for Cervical Cancer Using Fuzzy C-Means Clustering,” in *Proceedings of the 6th International Conference on Bioinformatics and Biomedical Science*, Singapore: ACM, Jun. 2017, pp. 137–141, doi: 10.1145/3121138.3121196.

Janowski, R. Kaźmierczak, C. Kowalczyk, and J. Szulwic, “Detecting Apples in the Wild: Potential for Harvest Quantity Estimation,” *Sustainability*, vol. 13, no. 14, p. 8054, Jul. 2021, doi: 10.3390/su13148054.

Khattak et al., “Automatic Detection of Pomelo and Leaves Diseases Using Deep Neural Network Model,” *IEEE Access*, vol. 9, pp. 112942–112954, 2021, doi: 10.1109/access.2021.3096895.

Liu, Y. Tao, J. Liang, K. Li, and Y. Chen, “Object Detection Based on YOLO Network.”

Lyu, R. Li, Y. Zhao, Z. Li, R. Fan, and S. Liu, “Green Citrus Detection and Counting in Orchards Based on YOLOv5-CS and AI Edge System,” *Sensors*, vol. 22, no. 2, p. 576, Jan. 2022, doi: 10.3390/s22020576.

Mpouziotas, D., Karvelis, P., Tsoulos, I., & Stylios, C. (2023). Automated wildlife bird detection from drone footage using computer vision techniques. *Applied Sciences*, *13*(13), 7787.

Pathak, H. Gangwar, and A. S. Jalal, “Performance Analysis of Gradient Descent Methods for Classification of Pomelos using Deep Neural Network,” in *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India: IEEE, Mar. 2020, pp. 68–72, doi: 10.23919/indiacom49435.2020.9083723.

Sapkota, R., & Karkee, M. (2025, Januari 26). Comparing YOLOv11 and YOLOv8 for instance segmentation of occluded and non-occluded immature green fruits in complex orchard environment (arXiv:2410.19869v3). https://doi.org/10.48550/arXiv.2410.19869

Sugadev, K. Sucharitha, I. R. Sheeba, and B. Velan, “Computer vision based automated billing system for fruit stores,” in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India: IEEE, Aug. 2020, pp. 1337–1342, doi: 10.1109/icssit48917.2020.9214101.

Wajid, N. K. Singh, P. Junjun, and M. A. Mughal, “Recognition of Ripe, Unripe and Scaled Condition of Pomelo Citrus Based on Decision Tree Classification.”

Wei et al., “GFS-YOLO11: A Maturity Detection Model for Multi-Variety Tomato,” *Agronomy*, vol. 14, no. 11, p. 2644, Nov. 2024, doi: 10.3390/agronomy14112644.

Xu, H. Zhao, O. M. Lawal, X. Lu, R. Ren, and S. Zhang, “An Automatic Jujube Fruit Detection and Ripeness Inspection Method in the Natural Environment,” *Agronomy*, vol. 13, no. 2, p. 451, Feb. 2023, doi: 10.3390/agronomy13020451.

Zhang, S. Ye, S. Zhao, W. Wang, and C. Xie, “Pear Object Detection in Complex sOrchard Environment Based on Improved YOLO11,” *Symmetry*, vol. 17, no. 2, p. 255, Feb. 2025, doi: 10.3390/sym17020255.