CONESCAPANHONDURAS2025paper11.pdf

17% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- **51** Not Cited or Quoted 14%
  Matches with neither in-text citation nor quotation marks
- **8** Missing Quotations 2%
  Matches that are still very similar to source material
- **4** Missing Citation 1%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

12% Internet sources
14% Publications
0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **51** Not Cited or Quoted 14%
  Matches with neither in-text citation nor quotation marks
- **8** Missing Quotations 2%
  Matches that are still very similar to source material
- **4** Missing Citation 1%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

12% Internet sources
14% Publications
0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	www.mdpi.com	3%
2	Internet	www.coursehero.com	1%
3	Publication	Yanwen Li, Juxia Li, Lei Luo, Lingqi Wang, Qingyu Zhi. "Tomato ripeness and stem ...	<1%
4	Publication	Ali Nourbakhsh, Ebrahim Babazadeh Mehrabani, Salman Faraji, Farzad A. Shirazi, ...	<1%
5	Internet	ejournal.unkhair.ac.id	<1%
6	Internet	etasr.com	<1%
7	Publication	Anderson Luiz dos Santos Safre, Alfonso Torres-Rua, Brent L. Black, Sierra Young. ...	<1%
8	Publication	Nail Beisekenov, Hideo Hasegawa. "Advanced Preprocessing Technique for Toma...	<1%
9	Internet	pubmed.ncbi.nlm.nih.gov	<1%
10	Internet	www.jait.us	<1%

11 Publication

Gabriel País Cerna, Germán Herrera-Vidal, Jairo R. Coronado-Hernández. "Autono... <1%

12 Internet

journal.hep.com.cn <1%

13 Publication

Nikita Stasenko, Dmitry Shadrin, Alexandr Katrutsa, Andrey Somov. "Dynamic Mo... <1%

14 Publication

H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in He... <1%

15 Internet

yolov8.org <1%

16 Internet

www.ruetir.com <1%

17 Publication

Zihan Fan, Dingze Lu, Man Liu, Zishen Liu, Qi Dong, Huiqi Zou, Hongjuan Hao, Yao... <1%

18 Publication

DianHui Mao, Hao Sun, XueBo Li, XuDong Yu, JianWei Wu, QingChuan Zhang. "Re... <1%

19 Publication

Md Mahbubur Rahman, Ashikul Islam, Forhadul Islam, Mashruba Zaman, Md Rafi... <1%

20 Publication

Song Wang, Jianxia Xiang, Daqing Chen, Cong Zhang. "A Method for Detecting To... <1%

21 Publication

Xuetang Xiong, Yiqiu Tan. "Deep learning-based detection of tie bars in concrete ... <1%

22 Internet

doaj.org <1%

23 Internet

downloads.hindawi.com <1%

24 Publication

Keita Fukada, Kataru Hara, Jingyong Cai, Daichi Teruya et al. "An Automatic Toma... <1%

25 Publication

Xuemei Liang, Haojie Jia, Hao Wang, Lijuan Zhang et al. "ASE-YOLOv8n: A Method ... <1%

26 Publication

"Light Metals 2025", Springer Science and Business Media LLC, 2025 <1%

27 Publication

José Naranjo-Torres, Marco Mora, Claudio Fredes, Andres Valenzuela. "Disease an... <1%

28 Internet

dergipark.org.tr <1%

29 Internet

dspace.ut.ee <1%

30 Publication

Zhiqing Tao, Ke Li, Yuan Rao, Wei Li, Jun Zhu. "Strawberry Maturity Recognition B... <1%

31 Publication

Di Mu, Yuping Guou, Wei Wang, Ran Peng, Chunjie Guo, Francesco Marinello, Yin... <1%

32 Publication

Jinfan Wei, Lingyun Ni, Lan Luo, Mengchao Chen, Minghui You, Yu Sun, Tianli Hu. ... <1%

33 Publication

Misbah Ahmad, Wenhao Zhang, Melvyn Smith, Ben Brilot, Matt Bell. "Real-Time Li... <1%

Detection and Classification of Tomato Ripeness in Greenhouses using Yolov11

1st Given name surname
Dept. name of organization
name of organization
City, Country
email address

2nd Given name surname
Dept. name of organization
name of organization
City, Country
email address

3rd Given name surname
Dept. name of organization
name of organization
City, Country
email address

4th Given name surname
Dept. name of organization
name of organization
City, Country
email address

***Abstract*—This study presents a computer vision-based system for detecting and classifying the ripeness level of tomatoes in greenhouse environments, using a deep convolutional neural network (CNN). The YOLOv11x architecture was employed, leveraging a pre-trained model to classify tomatoes into three ripeness categories: fully ripened ($> 90\%$ red), half ripened (30–89% red), and green ($< 30\%$ red). The system was trained and validated with the Laboro Tomato dataset, which consists of 804 images captured under varying lighting and background conditions. A preprocessing pipeline was implemented, including instance segmentation using the Segment Anything Model (SAM) and data augmentation techniques such as rotation, brightness and contrast adjustment, and scaling to improve robustness. The dataset was split into training (70%), validation (20%), and testing (10%) subsets. The metrics used to evaluate the model's performance included mean average precision (mAP), Intersection over Union (IoU), precision, recall, and F1-score. The proposed system achieved a mAP@0.5 of 91.4% and a recall of 85.3%, demonstrating a high generalization and accuracy capability in detecting tomato ripeness level, contributing to the development of automated solutions in precision agriculture.**

***Index Terms*—tomato ripeness, computer vision, YOLOv11, deep learning, object detection, precision agriculture.**

I. Introduction

The shortage of labor in the agricultural sector was one of the main challenges faced by modern agriculture. Factors such as the aging of the agricultural population and the lack of workers willing to enter the sector increased the pressure on traditional cultivation and harvesting systems. This labor shortage drove the search for automated solutions, especially in the harvesting of crops such as tomatoes, which is one of the most cultivated and marketed agricultural products worldwide [1], [2].

Automated tomato harvesting required the precise identification of the fruit's ripeness level, a process that was traditionally carried out through visual observation, but which became complicated due to the continuous nature of tomato ripening. This difficulty lay in the fact that tomato ripeness is not distributed discretely in clear classes, but rather changes gradually, making traditional color-based classification methods inadequate for precise and reliable estimation [3]. Therefore, solutions based on machine learning techniques were proposed, particularly deep neural networks (CNNs), which proved effective for the automatic extraction of visual features and the estimation of tomato ripeness from images [1].

CNNs, which are capable of learning complex patterns in images, were successfully used in several studies for tomato detection and classification into different ripening stages. These techniques offered a significant advantage over traditional methods, as they could adapt to variations in lighting and texture of the environment, factors that often affect the accuracy of manual feature-based techniques [2], [3]. Through supervised learning models, such as convolutional networks, it was possible not only to detect the location of tomatoes but also to accurately estimate their ripeness based on visual characteristics such as color and shape [1].

In this context, different approaches were explored to improve accuracy and speed in tomato detection in greenhouses and open fields. Lawal [4] proposed a modified YOLOv3-based model, called YOLO-Tomato, for tomato detection in complex environments. Improvements such as a dense architecture, the use of Spatial Pyramid Pooling, and the Mish activation function were applied. Metrics such as average precision (AP) and detection time were evaluated. As a result, three variants of the model were obtained: YOLO-Tomato-A (AP 98.3%, detection time 48 ms), YOLO-Tomato-B (AP 99.3%, 44 ms), and YOLO-Tomato-C (AP 99.5%, 52 ms), outperforming other conventional approaches in accuracy and time efficiency.

Ge et al. [5] presented a model for tracking and counting tomatoes at different growth stages using YOLO-Deepsort. It was optimized with Shufflenetv2 and CBAM attention to reduce model size without compromising accuracy. Precision metrics (93.1% for tomato flowers, 96.4% for green tomatoes, and 97.9% for red tomatoes), recall, and IoU were used to evaluate the model. These results demonstrated the feasibility of the model for yield estimation in greenhouses.

Huang et al. [6] proposed a model based on Mask R-CNN for cherry tomato ripeness classification. They used segmentation with Fuzzy C-Means and an HSV color model to predict ripeness level. Accuracy in detection and ripeness classification was evaluated with metrics such as accuracy (98%), IoU, and F1-score (0.9614), improving efficiency in automated harvesting.

Li et al. [7] developed an improved YOLOv5-based model for recognizing tomato ripeness in greenhouses. Mosaic techniques for data augmentation, CSPNet to improve training speed, and EIoU loss function to optimize bounding box regression were implemented. The model was evaluated using accuracy (95.58%), IoU, and mAP (97.42%), achieving significant improvements compared to previous YOLOv5 versions.

Mao et al. [8] proposed a real-time fruit detection approach based on deep neural networks optimized for CPU. Their model, called RTFD, was based on PicoDet-S and optimized efficiency through a lightweight network structure, CIoU loss function, and ACON-C activation. Metrics such as accuracy (0.78 for tomatoes and 0.96 for strawberries), recall, F1-score, IoU, model size (1.33 MB), and detection speed (19 FPS on a mobile device) were evaluated, demonstrating its potential for applications in precision agriculture.

Other studies explored variants in the implementation of lightweight models for fruit detection in agricultural environments. Kang et al. [9] proposed a detection model based on MobileNetV2 optimized for edge devices, achieving 91% accuracy with a 30% reduction in inference time. Zhang et al. [10] introduced the RTSD-Net model, which used an optimized version of YOLOv4-Tiny and was tested on embedded systems, reaching a detection speed of 25.20 FPS. Finally, Zhou et al. [11] implemented a quantized MobileNetV2-based model for kiwi detection, achieving a model size reduction to 4.5 MB with a detection time of 51 ms per image.

In this work, the contributions consist of the exploration and comparison of different strategies for the development of a system for detecting and classifying the ripeness level of tomatoes in greenhouses. The YOLOv11 model was used, incorporating improvements in segmentation and data preprocessing. In addition, a rigorous evaluation was implemented using metrics such as precision, recall, F1-score, and IoU, which allowed for a more accurate validation of the model's performance compared to previous approaches.

II. Materials and Methods

A. Data Acquisition

The Laboro Tomato dataset [12] was used, which is composed of 804 images of tomatoes at different ripening stages. These images were captured on a local farm using two cameras with different resolutions and image qualities, which allowed for greater variability in lighting, focus, and contrast conditions. This diversity in the images is beneficial for improving the robustness of the classification and segmentation model.

Fig. 1: Different ripening stages of tomatoes, taken from the Laboro Tomato dataset [12].

The tomatoes in the dataset were organized into three ripeness levels based on the proportion of red color present on their surface, as shown in Fig. 1. The first category, Fully ripened, includes those tomatoes that are fully ripe, with more than 90% of their surface red, indicating they are ready for harvest. The second category, Half ripened, corresponds to tomatoes in an intermediate ripening state, with red color coverage ranging between 30% and 89%. Finally, the Green category includes tomatoes that have not yet reached a significant ripening stage, presenting less than 30% of their surface with a red hue [12].

This dataset was used as a basis to train and evaluate segmentation and classification models, with the goal of accurately identifying the ripeness stage of tomatoes and facilitating their automatic selection in agricultural applications.

B. Image Preprocessing

During the development of the project, an image segmentation process was implemented with the objective of isolating tomatoes at different ripening stages. For this purpose, the Laboro Tomato dataset was used, which included annotations in COCO format with bounding boxes delimiting the tomatoes in each image.

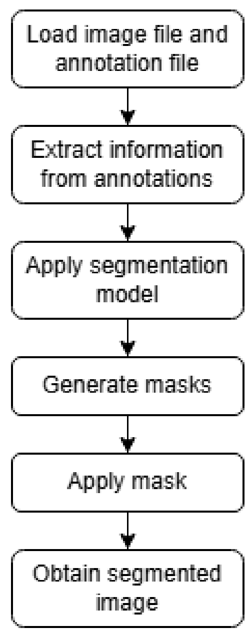

Fig. 2: Segmentation flowchart.

Initially, the train.json file was loaded, which contained the structured information of the images and their respective annotations. From these annotations, the bounding boxes corresponding to each tomato were extracted and the Segment Anything Model (SAM) was used to refine the segmentation, obtaining more precise masks. To achieve this, the model was configured using the samvitl.pth checkpoint and inference was performed on a GPU to optimize the process.

Each image was processed individually: first, it was read and converted to RGB format, then it was set as input to the SAM predictor, and finally, segmentation masks were generated from the provided bounding boxes. The mask with the highest confidence score was selected and applied to the image, setting the background to black to highlight only the tomatoes.

As a result, segmented images were obtained where the tomatoes were isolated from the background following the process in the flowchart in Fig. 2, which facilitated the next stages of the project, such as automatic classification according to their ripeness level.

a) b)

Fig. 3: Applying image preprocessing methods. a) Original image b) SAM segmentation.

Fig. 3 (a) showed an original image taken from the Laboro Tomato dataset, which contained tomatoes in different ripening stages and was obtained directly from the dataset used for analysis. On the other hand, Fig. 3 (b) presented an image processed using the SAM (Segment Anything Model) tool, in which the SAM model was used to segment the tomato images, removing the background and leaving only the region corresponding to the tomato.

To improve the robustness of the model, data augmentation techniques were used, such as rotations, changes in brightness and contrast, and image scaling. This increased the diversity of the dataset and improved the model's ability to detect tomatoes under varying conditions.

C. CNN Architecture

In this study, the YOLOv11x (You Only Look Once version 11 – "extra large") model was used, one of the most robust and accurate variants of the YOLOv11 architecture, specially designed for object detection tasks with high accuracy requirements. This version has greater depth and a higher number of parameters compared to lightweight versions, making it ideal for environments with GPU availability, where the main objective is to maximize model performance. Its use improved the detection capability for small or partially visible objects, optimizing the accurate recognition of tomato ripeness levels.

Unlike networks trained from scratch, a pretrained YOLOv11x model on a large reference dataset was used, which accelerated the training process and improved generalization. This architecture supports multiple computer vision tasks, including instance segmentation, oriented detection, classification, and pose detection. In this work, the specialized version for object detection was implemented, as it allows the localization and classification of multiple tomatoes within the same image, which is essential for an individual analysis of the ripeness stage.

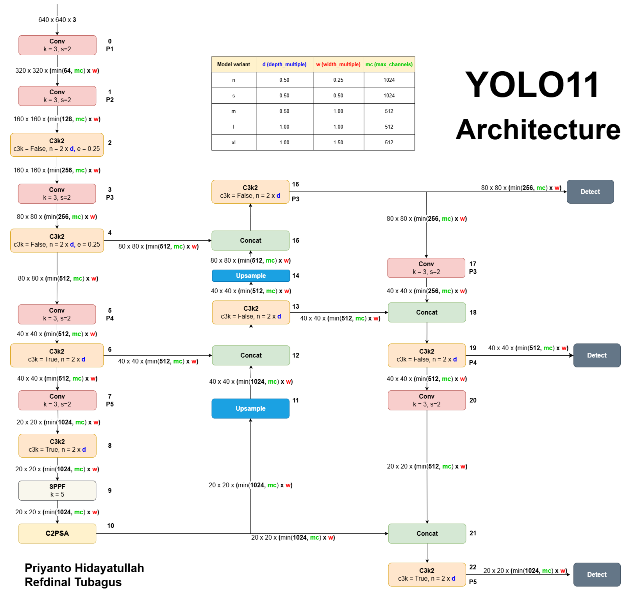

Fig. 4: YOLOv11 architecture used in the proposed model in this work [13].

The YOLOv11 architecture is composed of three main blocks: Backbone, Neck, and Head, as illustrated in Fig. 4. The Backbone is responsible for extracting deep features through multiple convolutional blocks (Conv) and C3K2 modules, optimized to preserve relevant semantic structures at different scales. Through these blocks, the 640×640-pixel input image is transformed into multiscale feature maps.

The Neck acts as a bridge between the Backbone and the Head, fusing outputs through concatenation operations, upsampling, and modules such as SPFF and C2PSA. This stage ensures effective integration of spatial and contextual information from multiple levels, thus facilitating robust object detection at various sizes.

The Head is responsible for the final prediction. In YOLOv11x, this output is performed at three scales (20×20, 40×40, and 80×80), allowing simultaneous detection of small, medium, and large objects. Each output generates bounding box coordinates, class labels, and confidence scores.

For training, 50 epochs were defined, an input resolution of 640 pixels, and a batch size of 16. The dataset was divided into 70% for training (563 images), 20% for validation (161 images), and 10% for testing (80 images). Evaluation metrics included precision, recall, F1-score, and IoU. Additionally, bounding box loss, classification loss, and focal distribution loss were monitored during each epoch.

With this configuration, the YOLOv11x model reached a mAP@0.5 of 90.2%, demonstrating a high capacity to correctly identify the ripeness stage of the tomato. Thanks to its deep and modular architecture, and the use of features extracted at multiple scales, YOLOv11x was consolidated as an effective tool for its application in agricultural automation and quality control systems.

D. Comparison with Other Architectures

Several studies have explored alternative architectures to address the task of detecting and classifying the ripeness level of tomatoes. In the study presented by Nguyen et al. [14], an architecture based on YOLOv8n was used, which was improved by incorporating the RFAConv block, with the aim of optimizing detection in automated agricultural environments. On the other hand, Xie et al. [15] proposed a variant of YOLOv5 called CAM-YOLO, which incorporates the CBAM (Convolutional Block Attention Module) attention module in order to improve accuracy under complex visual conditions, such as the presence of small or partially overlapping fruits.

E. Computational Resources

During the training process of the tomato ripeness detection model, a workstation equipped with an AMD Ryzen 7 5800X processor, an NVIDIA GeForce RTX 3060 Ti graphics card, 16 GB of RAM, and 2 TB of internal storage was used. These computational resources enabled model training with adequate processing times and efficient handling of the dataset.

III. Discussion of Results

A. Evaluation Metrics

To evaluate the model's performance, we applied three key metrics commonly used in object detection tasks. First, the Intersection over Union (IoU) (Equation 1) measures how well the predicted bounding boxes overlap with the ground truth annotations, helping us assess localization accuracy. Second, Precision (P) (Equation 2) indicates the proportion of correctly identified tomatoes among all those detected by the model, which is essential to minimize false positives. Third, Recall (Equation 3) measures how many of the actual tomatoes were correctly detected by the model, helping reduce false negatives. Together, these metrics provide a comprehensive evaluation of the model's ability to correctly detect and classify tomatoes at different ripeness stages.

$$\text{IoU} = \frac{\text{Area}\left(M_{\text{groundtruth}} \cap M_{\text{predicted}}\right)}{\text{Area}\left(M_{\text{groundtruth}} \cup M_{\text{predicted}}\right)} \tag{1}$$

$$P = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3}$$

B. Results

The model developed in this research, based on the YOLOv11 architecture, was trained in two phases. In the first phase, only the original images from the dataset were used, with a resolution of 640 pixels, for 100 epochs. Then, a second phase of fine-tuning was carried out, this time with 1024-pixel images, for 30 epochs. In this stage, both the original images and segmented versions were used to improve the model's performance. Throughout the process, segmentation techniques and data augmentation were applied.

The quantitative results obtained in the validation phase showed competitive performance. The average precision was 82.4%, recall reached 85.3%, while the mAP@0.5 metric achieved a value of 91.4%. Additionally, the mAP@0.5:0.95 metric, which evaluates model performance under different IoU thresholds, was 82.5%, indicating robust generalization under varying spatial conditions of the fruits.

At the individual class level, the model achieved a mAP@0.5:0.95 of 85.9% for green tomatoes, 79.8% for half-ripened tomatoes, and 81.3% for fully ripened tomatoes. These values reflect a good balance between precision and recall, especially considering the visual similarity between the half-ripened and fully ripened classes, where slight confusion was observed in images with overlapping or occluded fruits.

a) LaboroTomato image prediction

b) Prediction on image in uncontrolled environment

Fig. 5: Predictions of the proposed model

Figure 5 shows visual examples of the performance of the YOLOv11x model applied to two types of images. Figure 5a, corresponding to an image from the LaboroTomato dataset, shows that the model identifies multiple tomatoes per image, correctly assigning the labels "green," "half-ripened," and "fully ripened" with confidence levels above 90%. Figure 5b shows the prediction on an external image, captured outside the greenhouse environment, where the model maintains adequate performance despite variations in lighting and background. In both cases, the bounding boxes fit precisely to the fruit contour, which is essential for future integration with automated harvesting systems.

C. Interpretation of Results

When analyzing class-wise performance, it was observed that the ”green” category was the best recognized by the model, with a precision of 87.9% and a mAP50 of 95.4%, while the ”half-ripened” class presented greater challenges, possibly due to the gradual transition in visual characteristics between a green tomato and a fully ripened one. The ”fully ripened” class showed intermediate and consistent metrics, which demonstrated an overall solid behavior of the model.

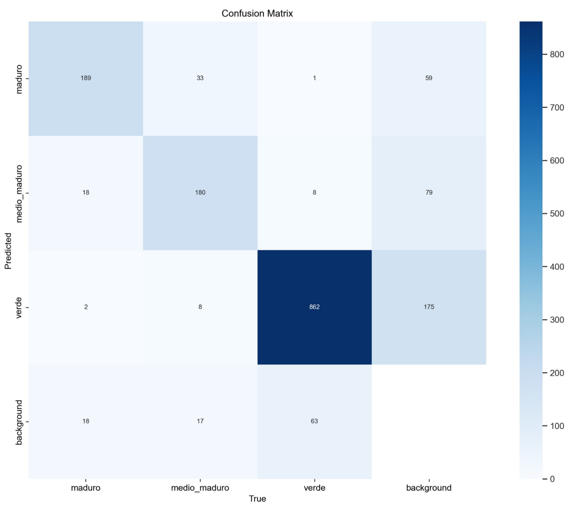

Fig. 6: Confusion matrix of the YOLOv11x model in the classification of tomato ripeness stages.

The confusion matrix obtained, shown in Fig. 6, reinforces the previously mentioned quantitative results. It is observed that the green class was the most accurately identified by the model, with 862 correct predictions and very few errors toward other classes. In contrast, the half-ripened class presented a greater number of confusions, especially with the background class (79 cases) and with the fully ripened class (18 cases), which confirms the difficulties the model faces in differentiating transitional states. For its part, the fully ripened class was also recognized with good precision, although with some confusions toward half-ripened and background. These observations support the overall robustness of the model but also highlight areas for improvement in distinguishing between adjacent classes along the ripening spectrum.

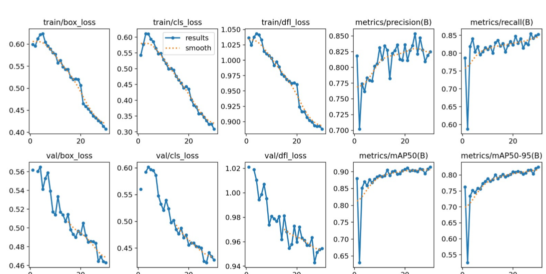

Fig. 7: Performance metrics of the proposed model.

Figure 7 shows the training graphs, where a general trend of improvement in the metrics is observed throughout the process. The loss, both in training and validation, decreases steadily and stabilizes at low levels, indicating that the model is learning to correctly locate and classify the tomato ripeness classes. The classification and box loss reflect an improvement in the model's accuracy when detecting and distinguishing the different stages. At the same time, the precision and recall metrics increase and stabilize at high values, suggesting that the model is achieving a good balance between correctly detecting positive cases and avoiding false negatives. The similarity in metric behavior between training and validation indicates that the model is not overfitting and maintains a good generalization capability.

D. Comparison with Other Studies

TABLE I: Performance comparison between models for tomato ripeness detection

Model	mAP@0.5	mAP@0.5:0.95	Recall	Precision
YOLOv8 [14]	88.1	45.7	-	-
CAM-YOLO [15]	88.0	-	81.2	87.2
YOLOv11 (ours)	91.3	82.4	85.2	82.3

When comparing the results with recent research, it is observed that the study conducted by Nguyen et al. [14] used a model based on YOLOv8n enhanced with RFAConv, achieving a mAP@0.5 of 88.2%. Additionally, the work of Xie et al. [15] developed CAM-YOLO, a variant of YOLOv5 with a CBAM attention module, achieving an average precision of 88.1% under complex conditions of occlusion and small fruits. While these studies also show solid results, the model proposed in this research achieves comparable precision and a higher mAP@0.5, reinforcing the effectiveness of the YOLOv11 architecture along with the adjustments made in preprocessing and segmentation.

E. Study Contributions

In this work, different strategies were explored and compared to develop a system for detecting and classifying the ripeness level of tomatoes in greenhouse environments, using the YOLOv11 architecture enhanced with improvements in segmentation and data preprocessing. As the main contribution, a model capable of identifying three ripeness stages (green, half-ripened, and fully ripened) with high precision was successfully implemented, even for visually similar classes. Additionally, a rigorous evaluation was conducted using metrics such as precision, recall, F1-Score, and IoU, which allowed for an accurate validation of the model's performance and an objective comparison with previous approaches. This work contributes to computer vision applied to precision agriculture, providing a solid foundation for the development of intelligent systems for automated harvesting support and real-time agronomic monitoring.

IV. Conclusions

The results obtained in this study confirm the feasibility and effectiveness of using the YOLOv11x architecture for the automated detection and classification of tomato ripeness in greenhouse environments. The model demonstrated a high level of accuracy, achieving a mAP@0.5 of 91.4% and robust performance across all ripeness classes, even under challenging visual conditions such as partial occlusions and fruit overlaps. The integration of advanced image preprocessing techniques, particularly segmentation using SAM and data augmentation, significantly contributed to the model's generalization and precision.

Compared to recent state-of-the-art approaches, the proposed method matched or outperformed previous results in terms of accuracy and reliability, positioning YOLOv11x as a suitable option for implementation in intelligent agricultural systems. This research provides a practical and scalable foundation for real-time applications focused on automated harvesting and agronomic monitoring, supporting the transition toward more efficient and sustainable agricultural practices.

References

[1] T. Kim, D.-H. Lee, K.-C. Kim, T. Choi, y J. M. Yu, "Tomato Maturity Estimation Using Deep Neural Network," *Applied Sciences*, vol. 13, no. 412, 2023. doi.
[2] D. Minagawa y J. Kim, "Prediction of Harvest Time of Tomato Using Mask R-CNN," *AgriEngineering*, vol. 4, pp. 356–366, 2022. doi.
[3] Y. Ge, J. Long, G. Liu, et al., "Tomato Maturity Recognition Model Based on Improved YOLOv5 in Greenhouse," *Computers and Electronics in Agriculture*, vol. 204, 2023. doi.
[4] M. O. Lawal, "YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3 Deep Learning Model," *Journal of Agricultural Informatics*, vol. 12, no. 3, 2021, pp. 67–78.
[5] X. Ge, L. Wang, y H. Zhang, "Tracking and Counting Tomato Fruits Using YOLO-Deepsort with Attention Mechanisms," *Computers and Electronics in Agriculture*, vol. 190, 2022, pp. 104639.
[6] Y. Huang, F. Li, y M. Zhao, "Tomato Ripeness Classification Using Mask R-CNN and Fuzzy C-Means Clustering," *Sensors*, vol. 20, no. 14, 2020, pp. 3982.
[7] R. Li, Z. Ji, S. Hu, X. Huang, J. Yang, y W. Li, "Tomato Maturity Recognition Model Based on Improved YOLOv5 in Greenhouse," *Agronomy*, vol. 13, no. 2, 2023, pp. 603.
[8] D. Mao, H. Sun, X. Li, et al., "Real-time Fruit Detection Using Deep Neural Networks on CPU (RTFD): An Edge AI Application," *Computers and Electronics in Agriculture*, vol. 204, 2023, pp. 107517.
[9] J. Kang, X. Wang, y L. Liu, "Lightweight Deep Learning-Based Fruit Detection for Edge Devices," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, 2022, pp. 3221–3232.
[10] Y. Zhang, M. Chen, y Z. Xu, "RTSD-Net: A Lightweight Object Detection Model for Embedded Systems," *Journal of Artificial Intelligence Research*, vol. 75, 2022, pp. 145–162.
[11] Y. Zhou, K. Tang, y W. Zheng, "Optimized MobileNetV2 for Fruit Detection in Agriculture," *Computers and Electronics in Agriculture*, vol. 198, 2023, pp. 105642.
[12] LaboroAI, "LaboroTomato," 2021. [Enlace]. Disponible en: https://github.com/laboroai/LaboroTomato.
[13] P. Hidayatullah, N. Syakrani, M. R. Sholahuddin, T. Gelar, and R. Tubagus, "YOLOv8 to YOLO11: A Comprehensive Architecture In-depth Comparative Review," *arXiv preprint arXiv:2501.13400v2*, 2025.
[14] D.-L. Nguyen, X. Thuy, P. Priadana, C. J. Hwang, and K.-H. Jo, "Improved Tomato Detector Supporting for Automatic Harvest Systems," in *Proc. CITA 2024 – Special Session on Computer Vision*, 2024.
[15] Y. Xie, J. Zhang, X. Wang, and Z. Li, "CAM-YOLO: A Lightweight Tomato Maturity Detection Network with CBAM Attention," *PeerJ Computer Science*, vol. 9, p. e1463, 2023. DOI: 10.7717/peerj-cs.1463.