Multi-Stage Tomato Flower Detection Using Deep Learning

Jennifer Dinh
Computer & Information Sciences Dept.
Webster University
jenniferdinh@webster.edu

Ayanna Miles
Natural Sciences & Mathematics Dept.
Webster University
ayannamiles@webster.edu

Abstract

Pollinator decline poses a significant challenge to greenhouse agriculture, particularly for crops such as tomatoes that rely on vibration-based pollination. Although robotic pollination systems have been explored, their effectiveness is highly dependent on accurate flower detection and stage identification. This work presents a multi-stage tomato flower detection framework using two state-of-the-art object detection models: YOLOv8m and RT-DETR-L. Unlike prior work that focuses on binary classification (e.g., flower vs. bud), we evaluate a three-stage scheme — Bud, Anthesis, and Post-Anthesis — provided by a publicly available dataset. Multi-stage classification has the potential to inform robotic decision-making, such as determining which flowers are ready for pollination, which should be revisited, and which are not yet suitable. The models are trained on this dataset and evaluated on both structured and real-world images. Results showed that YOLOv8m achieves higher detection performance with mAP@0.5 = 92.1%, compared to the 89.2% achieved by RT-DETR-L on the structured dataset, while performance drops to 71.1% and 64.2%, respectively, on real-world images. This performance gap highlights the complexity of flower stage classification and the limitations of current annotation practices. Inconsistent stage labeling and missing flower instances reduce the reliability of stage information; consequently, while this framework provides a reproducible pipeline for further evaluation, it is not yet ready for fully autonomous robotic pollination.

1 Introduction

Pollinator decline has been a long-standing issue with significant losses observed across wild bee colonies [1]. This issue is especially pronounced in greenhouse environments, where controlled conditions limit natural pollinator activity as shown in the Australian tomato greenhouse industry. Tomato plants (Solanum lycopersicum) present a unique challenge due to their poricidal anthers, which require vibration within a specific frequency to release pollen effectively. While tomatoes are capable of self-pollination, insufficient vibration can result in reduced fruit set, lower fruit weight, and diminished fruit quality. Traditional artificial pollination methods, such as tuning forks and electric toothbrushes, are labor-intensive and require skilled workers, making them difficult to scale [2, 3, 4, 5].

Recent advances in robotics and computer vision offer promising alternatives. The commercialization of Polly+, a buzz-pollinating robot, demonstrates the feasibility of robotic pollination [6], though its region-specific limitations highlight the need for more accessible and scalable solutions. Several systems remain under investigation [7, 8, 9, 10], underscoring that the field has yet to reach broad deployment. A potential bottleneck in advancing robotic pollination systems is reliably identifying flowers at the correct developmental stage. Variability in flower morphology and inconsistencies in labeled datasets can complicate accurate classification, which may affect the timing of pollination interventions.

Most existing approaches treat flower detection as a binary problem (flower vs. bud), neglecting the temporal dynamics of flower development. Flowering progresses through multiple stages, each with distinct biological characteristics. Pollination is most effective during anthesis, when pollen is available and the stigma is receptive [11]. Failure to distinguish these stages can lead to poorly timed interventions, reducing pollination success. Accurate flower stage recognition is critical for precision pollination systems, as it enables targeted and timely actuation while minimizing unnecessary interactions with non-viable flowers. Beyond operational efficiency, multi-stage classification provides meaningful biological context — informing

the system not only whether a flower is ready for pollination, but which stage of development it has reached. This distinction is essential for optimizing resource usage, improving fruit set, and advancing the development of autonomous agricultural systems.

To address this gap, this paper extends the binary detection framework presented in [12] by transitioning to a multi-stage classification scheme. While prior work successfully demonstrated the efficacy of YOLO-based architectures for general flower localization, a binary "flower vs. bud" approach lacks the temporal resolution required for autonomous pollination. We move beyond this limitation by categorizing tomato flowers into three biologically meaningful stages: Bud (pre-bloom), Anthesis (the optimal window for pollination), and Post-Anthesis (expired or pollinated).

We evaluate two modern detection architectures: YOLOv8m, a real-time convolutional detector and RT-DETR-L, a transformer-based real-time detection model. To the best of our knowledge, this is among the first works to apply RT-DETR to multi-stage tomato flower detection.

The main contributions of this work are:

- A three-stage tomato flower classification framework that extends prior binary detection approaches.
- A comparative evaluation of YOLOv8m and RT-DETR-L on multi-stage flower detection for robotic pollination.
- A real-world evaluation dataset collected from iNaturalist to assess model generalization beyond controlled conditions.

2 Related Work

To contextualize our approach, we survey prior work on tomato flower stage classification, YOLO-based detection, and the emerging use of transformer models in agricultural vision tasks.

2.1 Tomato Flower Development and Stage Classification

Understanding the biological stages of tomato flower development is critical for designing effective pollination systems. Hiraguri et al. [13] identify six distinct stages of tomato flower growth, with the fourth stage—anthesis—representing the optimal window for pollination due to pollen availability and stigma receptivity [11, 14]. Many computational approaches simplify this biological complexity into binary classification tasks. While this reduces annotation effort and model complexity, it fails to capture the temporal progression of flowering and limits the system's ability to make nuanced decisions. In practice, distinguishing between pre-anthesis (immature) and post-anthesis (expired) flowers is essential, as both are unsuitable for pollination but require different operational responses [14].

Recent work on pollination has begun to acknowledge the importance of stage-aware classification [15, 16]; however, fine-grained multi-stage labeling remains underexplored in the context of robotic pollination. This gap motivates the need for models capable of distinguishing multiple biologically meaningful stages.

2.2 Deep Learning for Flower Detection

Early flower detection methods relied on traditional computer vision techniques such as color thresholding, edge detection, and clustering. While computationally efficient, these approaches are highly sensitive to environmental variability, including lighting conditions, occlusion, and background complexity [17].

The introduction of deep learning has significantly improved robustness and accuracy in agricultural vision tasks. Convolutional Neural Network (CNN)-based object detectors such as Faster R-CNN, SSD, and the YOLO family have been widely adopted for plant phenotyping and flower detection [18]. Among these, YOLO-based models are particularly well-suited for real-time applications due to their single-stage architecture and high inference speed.

Singh et al. [12], provide a relevant benchmark for tomato flower detection by comparing YOLOv5s and YOLOv8s in a robotic pollination context. Their results show that YOLOv8s achieved superior performance, with a mean Average Precision (mAP) of 92.6% compared to 91.2% for YOLOv5s, while also offering faster

inference (0.7 ms per image). However, their approach is limited to binary classification, which restricts its applicability for precise pollination timing.

Similarly, other studies have applied YOLO-based architectures to detect flowers across different plant species [16, 19, 20], demonstrating strong performance but typically focusing on coarse-grained or binary stage detection tasks rather than fine-grained multi-stage labeling.

Prior work has also highlighted the limited availability of annotated agricultural flower datasets, proposing synthetic data generation techniques to improve detection performance. For example, synthetic flower images have been created by segmenting and compositing flower instances onto plant backgrounds, improving performance across models such as Faster R-CNN, YOLOv3, SSD, and CenterNet, with reported precision exceeding 91% [18]. However, these approaches primarily focus on improving detection accuracy rather than addressing stage-specific classification.

2.3 Transformer-Based Object Detection

Transformer-based models have recently emerged as a powerful alternative to CNN-based detectors by leveraging self-attention mechanisms to capture global context. DETR (DEtection TRansformer) and its variants eliminate the need for anchor boxes and non-maximum suppression, simplifying the detection pipeline [21, 22].

RT-DETR is a real-time adaptation of DETR that integrates efficient feature extraction with transformer-based attention, enabling competitive performance with reduced computational overhead [22]. Recent studies [20, 23, 24] have demonstrated the effectiveness of RT-DETR in general object detection tasks and some agricultural applications.

Despite these advances, the application of transformer-based detectors to flower stage classification, particularly in tomato pollination systems, remains limited.

2.4 Robotic Pollination and Vision Integration

For robotic pollinators to operate effectively, robust environmental mapping and autonomous navigation capabilities are essential. These enable robots to move efficiently within the greenhouse environment and precisely locate target flower clusters. Robotic pollination systems, including platforms such as BrambleBee and B-Droid, integrate perception, planning, and actuation to automate pollination tasks. These systems rely heavily on vision modules for detecting flower locations and determining pollination readiness.

A common limitation across these systems is the reliance on simplified detection outputs, often lacking detailed stage classification. This can lead to inefficient or mistimed pollination attempts, reducing overall system effectiveness. Furthermore, many systems are evaluated in controlled environments, limiting their generalizability to real-world greenhouse conditions [7, 8, 9, 10, 6].

3 Methodology

This section describes the experimental setup used to evaluate YOLOv8m and RT-DETR-L on multi-stage tomato flower detection. We outline the dataset, model configurations, training procedure, and evaluation protocol used across both test sets.

3.1 Dataset Preparation

This study utilized a publicly available tomato flower dataset from Kaggle [25], comprising 14,545 training images and 3,656 validation images totaling 18,201 images across three developmental stages based on flower morphology:

- **Bud:** The flower is closed; reproductive organs are protected and not yet ready for pollination.
- **Anthesis:** The petals are fully reflexed, exposing the anther cone. This is the biologically optimal window for pollination.
- **Post-Anthesis:** The petals have wilted or drooped forward, signaling the end of the pollination window.

Both training and validation splits were combined and redistributed using a stratified 80/15/5 train/validation/test split, where each image was assigned to its dominant class to ensure proportional class representation across all subsets, as shown in Table 1.

Table 1: Dataset Distribution Across Train, Validation, and Test Splits

Class	**Train**	**Validation**	**Test**	**Total**
Bud	7,849	1,472	491	9,812
Anthesis	5,550	1,041	347	6,938
Post-Anthesis	1,121	210	70	1,401
Total	**14,520**	**2,723**	**908**	**18,151**

Additionally, a supplementary real-world test set of 82 images were collected from iNaturalist using the search term *Solanum lycopersicum* and manually annotated using CVAT following the same labeling schema observed in the Kaggle dataset. Images with ambiguous stage classification were excluded from annotation to ensure label reliability. This test set captures diverse real-world conditions including varying lighting, outdoor wildlife settings, and indoor greenhouse environments

3.2 Model Selection and Training Configuration

Two distinct architectures were selected to evaluate the trade-off between localized feature extraction and global context modeling:

- **YOLOv8m (YOLO):** A CNN-based, anchor-free detector optimized for real-time inference and low-latency deployment.
- **RT-DETR-L (RT-DETR):** A Real-Time Detection Transformer leveraging a hybrid encoder to model global spatial relationships without requiring Non-Maximum Suppression (NMS).

Both models were implemented using the Ultralytics framework to ensure a consistent training pipeline and were initialized with COCO-pretrained weights via transfer learning. Training followed the hyperparameter configurations reported in the benchmark by Singh et al. [12], with 150 epochs at a resolution of 640×640, batch size 8, and weight decay of 0.0005. YOLOv8m was trained with SGD ($lr = 0.001$) following the benchmark, while RT-DETR-L used AdamW ($lr = 0.0001$) to account for the transformer's sensitivity to learning rate selection.

3.3 Evaluation

Models were evaluated on two test sets: the Kaggle test set as the primary in-distribution evaluation, and the iNaturalist test set — a real-world out-of-distribution test set compiled from publicly available citizen science imagery — to assess model generalization beyond controlled conditions. Performance was measured using Precision, Recall, Mean Average Precision at IoU threshold 0.5 (mAP@0.5), and Mean Average Precision averaged across IoU thresholds from 0.5 to 0.95 (mAP@0.5:0.95). Inference speed was also recorded to assess real-time deployment feasibility. Results are further analyzed through Precision-Recall curves and normalized confusion matrices.

4 Results

This section presents the quantitative and qualitative results for both models across the two test sets.

4.1 Kaggle Test Set

Both models performed well on the Kaggle test set Table 2. YOLOv8m achieved an overall mAP@0.5 of 92.1%, outperforming RT-DETR-L at 89.2%. Anthesis achieved the highest per-class mAP for both models,

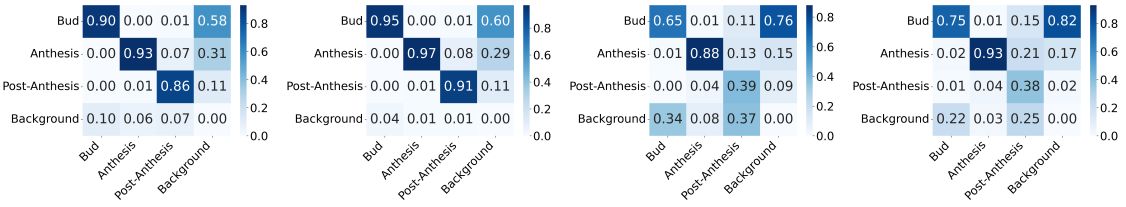
while Post-Anthesis was the weakest class, likely due to its underrepresentation in the training set relative to Bud and Anthesis instances. YOLOv8m also demonstrated a significantly faster inference time of 2.8ms compared to RT-DETR-L at 6.5ms, which is expected given its smaller model size, but is an important consideration for real-time robotic pollination deployment.

Table 2: Detection Performance of YOLOv8m and RT-DETR-L on Kaggle Test Set

Model	Class	P	R	mAP@0.5	mAP@0.5:0.95	Inf. (ms)
YOLOv8m	All	86.1%	86.4%	92.1%	84.6%	2.8
	Bud	87.5%	86.2%	93.0%	82.8%	–
	Anthesis	87.8%	90.2%	95.1%	89.1%	–
	Post-Anthesis	82.8%	82.7%	88.2%	82.0%	–
RT-DETR-L	All	83.6%	85.8%	89.2%	80.3%	6.5
	Bud	85.8%	85.6%	91.7%	80.1%	–
	Anthesis	86.0%	89.8%	93.5%	85.4%	–
	Post-Anthesis	79.1%	82.1%	82.4%	75.4%	–

The normalized confusion matrices (Figure 1a, 1b) show that RT-DETR-L achieves higher diagonal values across all three stages, with values of 0.95, 0.97, and 0.91 for Bud, Anthesis, and Post-Anthesis respectively, compared to YOLOv8m's 0.90, 0.93, and 0.86. However, this differs from the overall recall reported in Table 2, where YOLOv8m outperforms RT-DETR-L (86.4% vs 85.8%). This discrepancy likely arises because the confusion matrix reflects performance at a fixed operating point (i.e., specific confidence and matching criteria), whereas the overall recall summarizes overall detection performance across the dataset. Additionally, both models show a high false positive rate for the Bud class, with 58% and 60% of true background instances predicted as Bud for YOLOv8m and RT-DETR-L respectively, suggesting both models struggle to distinguish small bud-like structures from background.

Figure 2a and 2b show that YOLOv8m maintains higher precision across recall levels than RT-DETR-L, resulting in a larger area under the curve and a higher overall mAP@0.5.



(a) YOLOv8, KG (b) RT-DETR, KG (c) YOLOv8, iN (d) RT-DETR, iN

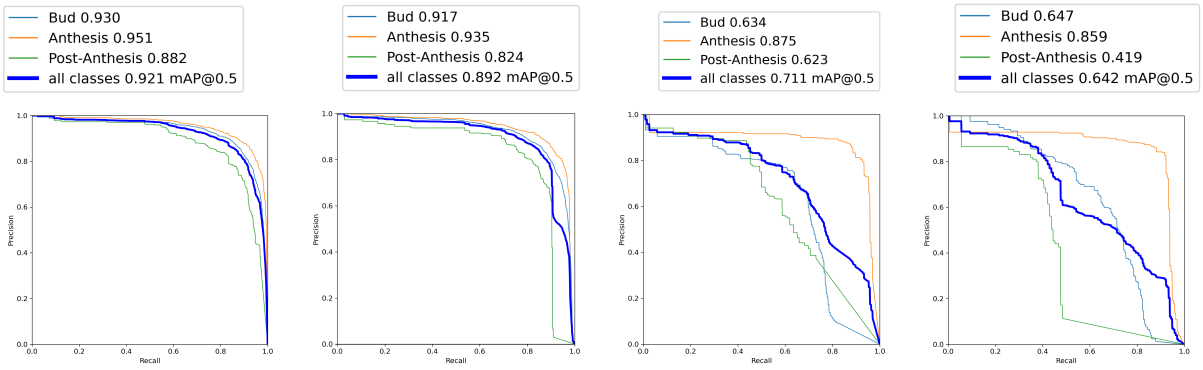
Figure 1: Normalized Confusion Matrices on Different Test Sets. KG: Kaggle, iN: iNaturalist.

4.2 iNaturalist Test Set

A notable performance drop on the iNaturalist test set reveals that model performance is highly sensitive to dataset variability and annotation quality.

The iNaturalist test set results follow a similar trend to the Kaggle evaluation, with both models struggling most with Post-Anthesis detection and YOLOv8m outperforming RT-DETR-L across all metrics. Inference times were notably higher than those reported in Table 1, as iNaturalist images were not preprocessed to a standard resolution prior to evaluation, resulting in increased computational overhead. Furthermore, the gap between mAP@0.5 and mAP@0.5:0.95 was approximately 20% on the iNaturalist set compared to approximately 8% on the Kaggle set, indicating that while both models successfully detect flowers, bounding box localization is considerably less precise under real-world conditions likely due to varied angles, object distances and occulsion present in the dataset.

The iNaturalist confusion matrices (Figure 1c, 1d) reflect similar trends to the Kaggle evaluation, with Anthesis achieving the highest per-class recall for both models. Notably, YOLOv8m achieved a higher



(a) YOLOv8, KG (b) RT-DETR, KG (c) YOLOv8, iN (d) RT-DETR, iN

Figure 2: Precision-Recall Curves on Different Test Sets. KG: Kaggle, iN: iNaturalist.

Post-Anthesis diagonal value (0.39), compared to RT-DETR-L (0.38), a reversal from the Kaggle results. Post-Anthesis instances were largely misclassified as Background or Anthesis, suggesting visual ambiguity between these stages under real-world conditions. The elevated background row values relative to the Kaggle matrices further indicate that both models produced more false positives on the iNaturalist set, which may be partially attributable to incomplete ground truth annotations, as some flower instances present in the images were not captured during manual labeling.

The PR curves in (Figure 2c, 2d) show a steeper precision drop for the Post-Anthesis class as recall increases, reflecting the visual ambiguity of this stage under real-world conditions. This is further compounded by inconsistent annotation criteria in the training set, where the absence of a Pre-Anthesis class led to transitional stage flowers being grouped under either anthesis or post-anthesis, making it difficult for the model to learn a precise decision boundary between the two classes.

Table 3: Detection Performance of YOLOv8m and RT-DETR-L on INaturalist Test Set

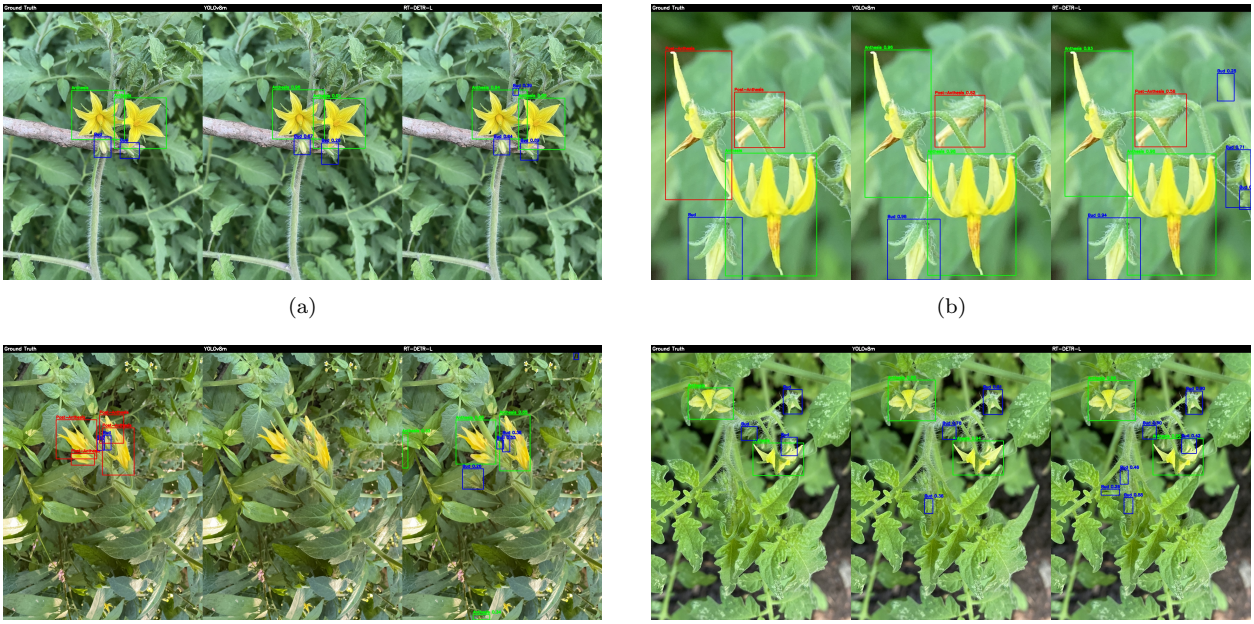
Model	**Class**	**P**	**R**	**mAP@0.5**	**mAP@0.5:0.95**	**Inf. (ms)**
YOLOv8m	All	80.7%	65.4%	71.1%	52.5%	24.2
	Bud	68.9%	64.3%	63.4%	42.2%	–
	Anthesis	85.8%	88.1%	87.5%	72.5%	–
	Post-Anthesis	87.5%	43.7%	62.3%	42.9%	–
RT-DETR-L	All	76.7%	63.7%	64.2%	46.7%	40.1
	Bud	64.6%	64.3%	64.7%	41.2%	–
	Anthesis	83.8%	91.8%	85.9%	70.1%	–
	Post-Anthesis	81.6%	35.2%	41.9%	29.0%	–

4.3 Qualitative Results

Figure 3a highlights the need for a Pre-Anthesis class. Although predictions align with the ground truth, the Anthesis label does not adequately represent this transitional flower stage. Both models demonstrate accurate localization, with predicted bounding boxes closely matching the ground truth.

In Figure 3b, an ambiguous flower on the far right was excluded from annotation due to uncertainty in stage classification. RT-DETR-L correctly predicted this instance as a bud, consistent with what manual inspection would suggest. Additionally, one flower in this image displays Anthesis morphology but with visibly shriveled anthers, suggesting reduced pollination viability. However, it was labeled as an Anthesis to maintain consistency with the Kaggle dataset's observed annotation criteria, further highlighting the biological validity limitations discussed in this study.

Figure 3c illustrates YOLOv8m's conservative detection behavior alongside RT-DETR-L's more permissive nature. The false positives produced by RT-DETR-L in this image are likely attributable to the absence



(a) (b)

(c) (d)

Figure 3: Qualitative detection results on the iNaturalist test set. Left: ground truth; middle: YOLOv8; right: RT-DETR-L. Green bounding boxes indicate Anthesis, red indicate Post-Anthesis, and blue indicate Bud. Subfigures adapted from: (a) [p4_fatima_zamarron], CC BY 4.0; (b) and (d) [suneholt], CC BY 4.0; (c) [jduns29], CC BY-NC 4.0. Modifications: added bounding boxes.

of clearly defined annotation guidelines for visually ambiguous classes such as Anthesis and Post-Anthesis in the training set.

Figure 3d demonstrates both models' ability to detect flower instances missed during manual annotation, with RT-DETR-L showing particular sensitivity in identifying unlabeled Buds. This suggests that the models may generalize beyond the constraints of the ground truth annotations.

5 Conclusion and Future Work

This study evaluated YOLOv8m and RT-DETR-L for multi-stage tomato flower classification on a controlled greenhouse dataset (Kaggle [25]) and a real-world iNaturalist dataset. While both models achieved high performance on the Kaggle dataset—YOLOv8m with an mAP@0.5 of 92.1% and RT-DETR-L with 89.2%—closer inspection revealed missing labels and inconsistent stage annotation, particularly between Anthesis and Post-Anthesis stages. Performance dropped substantially on the iNaturalist dataset (71.1% vs. 64.2%), highlighting the limitations of existing annotation practices and the challenges of generalizing to diverse, real-world conditions.

These results emphasize that the primary bottleneck for multi-stage flower classification is not model choice alone, but the lack of reliable, stage-specific annotation. Notably, both models detected flowers that were missed during manual labeling, with RT-DETR-L showing particular sensitivity, suggesting that model predictions can complement and even improve existing annotations.

Future work will focus on developing rigorous annotation criteria in collaboration with domain experts, and expanding datasets to capture environmental variability. Evaluating scaled variants of both YOLO and RT-DETR models will also facilitate deployment on resource-constrained robotic platforms while providing a more controlled comparison of CNN- and transformer-based approaches.

Acknowledgment

This work was partially funded by the Webster University Faculty Research Grant and the Webster University President's Student/Faculty Research Grant. We would like to thank our advisors, Dr. Gamage and Dr. Miller-Struttman, for their guidance throughout this project—Dr. Gamage for shaping the research direction and supporting the computer science aspects, and Dr. Struttman for providing valuable biological insights.

References

[1] C. J. Rhodes, "Pollinator decline – an ecological calamity in the making?" *Science Progress*, vol. 101, no. 2, pp. 121–160, 2018.

[2] A. Dingley, S. Anwar, P. Kristiansen, N. W. M. Warwick, C.-H. Wang, B. M. Sindel, and C. I. Cazzonelli, "Precision pollination strategies for advancing horticultural tomato crop production," *Agronomy*, vol. 12, no. 2, 2022. [Online]. Available: https://www.mdpi.com/2073-4395/12/2/518

[3] S. S. Greenleaf and C. Kremen, "Wild bees enhance honey bees' pollination of hybrid sunflower," *Proceedings of the National Academy of Sciences*, vol. 103, no. 37, pp. 13 890–13 895, 2006.

[4] P. A. De Luca and M. Vallejo-Marín, "What's the 'buzz' about? the ecology and evolutionary significance of buzz-pollination," *Current Opinion in Plant Biology*, vol. 16, no. 4, pp. 429–435, 2013.

[5] K. Hogendoorn, C. L. Gross, M. Sedgley, and M. A. Keller, "Increased tomato yield through pollination by native australian amegilla chlorocyanea (hymenoptera: Anthophoridae)," *Journal of Economic Entomology*, vol. 99, no. 3, pp. 828–833, 2006.

[6] Arugga, "Arugga technology," https://www.arugga.com/technology, 2024, accessed: 2026.

[7] J. Strader, J. Nguyen, C. Tatsch, Y. Du, K. Lassak, B. Buzzo, R. Watson, H. Cerbone, N. Ohi, C. Yang, and Y. Gu, "Flower interaction subsystem for a precision pollination robot," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 5534–5541.

[8] University of the West of England, "B-droid – a robot that's busy as a bee," 2013, accessed: 2026-03-25. [Online]. Available: https://info.uwe.ac.uk/news/uwenews/news.aspx?id=2676

[9] G. Ren, T. Wu, T. Lin, L. Yang, G. Chowdhary, K. C. Ting, and Y. Ying, "Mobile robotics platform for strawberry sensing and harvesting within precision indoor farming systems," *Journal of Field Robotics*, vol. 41, no. 7, pp. 2047–2065, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.22207

[10] R. T. Karunarathna, C. Wickramarathne, M. A. M. Akmal, C. S. W. Arachchi, K. Dissanayaka, and N. Silva, "Towards autonomous strawberry pollination via iot and robotics," in *2025 7th International Conference on Advancements in Computing (ICAC)*. IEEE, 2025, pp. 1–6.

[11] K. Esau, *Anatomy of Seed Plants*, 2nd ed. New York: John Wiley & Sons, 1977.

[12] R. Singh, A. Khan, L. D. Seneviratne, and I. Hussain, "Deep learning approach for detecting tomato flowers and buds in greenhouses on 3p2r gantry robot," *Scientific Reports*, vol. 14, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:272400830

[13] T. Hiraguri, H. Shimizu, T. Kimura, T. Matsuda, K. Maruta, Y. Takemura, T. Ohya, and T. Takanashi, "Autonomous drone-based pollination system using ai classifier to replace bee for greenhouse tomato cultivation," *IEEE Access*, vol. 11, pp. 99 352–99 364, 2023.

[14] H. Xiao, C. Radovich, N. Welty, J. Hsu, D. Li, T. Meulia, and E. van der Knaap, "Integration of tomato reproductive developmental landmarks and expression profiles, and the effect of SUN on fruit shape," *BMC Plant Biology*, vol. 9, p. 49, 2009.

[15] Q. Zhang, Z. Zhang, S. Manzoor, T. Li, C. Igathinathane, W. Li, M. Zhang, M. Mhamed, S. Javidan, and M. Abdelhamid, "A comprehensive review of autonomous flower pollination techniques: Progress, challenges, and future directions," *Computers and Electronics in Agriculture*, vol. 237, p. 110577, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169925006830

[16] R. Ren, H. Sun, S. Zhang, H. Zhao, L. Wang, M. Su, and T. Sun, "Fpg-yolo: A detection method for pollenable stamen in 'yuluxiang' pear under non-structural environments," *Scientia Horticulturae*, vol. 328, p. 112941, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304423824001018

[17] Y. Zhao, L. Gong, Y. Huang, and C. Liu, "A review of key techniques of vision-based control for harvesting robot," *Computers and Electronics in Agriculture*, vol. 127, pp. 311–323, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169916304227

[18] U. Rahim and H. Mineno, "Data augmentation method for strawberry flower detection in non-structured environment using convolutional object detection networks," *Journal of Agricultural and Crop Research*, vol. 8, pp. 260–271, 11 2020.

[19] P. A. Dias, A. Tabb, and H. Medeiros, "Apple flower detection using deep convolutional networks," *Computers in Industry*, vol. 99, pp. 17–28, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016636151730502X

[20] J. Lim, H. S. Ahn, M. Nejati, J. Bell, H. Williams, and B. A. MacDonald, "Deep neural network based real-time kiwi fruit flower detection in an orchard environment," 2020. [Online]. Available: https://arxiv.org/abs/2006.04343

[21] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 213–229.

[22] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs beat YOLOs on real-time object detection," *arXiv preprint arXiv:2304.08069*, 2023. [Online]. Available: https://arxiv.org/abs/2304.08069

[23] S. Bumbaca and E. Borgogno-Mondino, "On the minimum dataset requirements for fine-tuning an object detector for arable crop plant counting: A case study on maize seedlings," *Remote Sensing*, vol. 17, no. 13, 2025. [Online]. Available: https://www.mdpi.com/2072-4292/17/13/2190

[24] I. Pinheiro, G. Moreira, S. Magalhães *et al.*, "Deep learning based approach for actinidia flower detection and gender assessment," *Scientific Reports*, vol. 14, p. 24452, 2024.

[25] M. J. Karim, "Tomato-flower-3-class," 2025. [Online]. Available: https://www.kaggle.com/dsv/11224850