HOPSY: Harvesting Optimization for Production of StrawberrY using Real-time Detection with Modified YOLOv8-nano

Zijing Huang[1], Won Suk Lee[1], Niteesh Chowdary Takkellapati[2]

[1]Department of Agricultural and Biological Engineering, University of Florida, USA

[2]Department of Electrical and Computer Engineering, University of Florida, USA

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.
Enhancing the efficiency of strawberry harvesting is pivotal for boosting profitability and sustainability in agriculture. This study introduces the AgriBerry Vision (ABV) imaging module, an advanced portable real-time detection system designed to address significant challenges associated with uncertain yield fluctuations. The ABV system incorporates a modified YOLOv8-nano model, achieving impressive detection speeds with an average inference time of 961.4 ms on the NVIDIA Jetson Nano. It classifies strawberries into three stages of ripeness—flower, immature fruit, and mature fruit—and records these classifications along with precise GPS coordinates to create detailed yield maps. These maps are crucial for optimizing labor distribution, enhancing operational efficiency, and supporting timely, informed decision-making in strawberry field management. The user-friendly system requires minimal setup with two power and detection activation switches. It integrates seamlessly with agricultural machinery such as sprayers, making it a practical solution for widespread adoption in the field. From a technical standpoint, the ABV system's model exemplifies efficiency, with only 1,653,463 parameters and a modest computational demand of 4.9 GFLOPS. Despite its streamlined architecture, it maintains high detection accuracy with an overall mean Average Precision (mAP) of 0.912. The ABV system offers a solution for strawberry harvesting by integrating advanced machine learning algorithms with robust edge computing technology. Future enhancements will focus on refining the detection of flower stages and incorporating additional functionalities, such as canopy estimation, to further advance precision agriculture's capabilities, such as yield prediction.

Keywords.
Artificial Intelligence, Computer Vision, Edge Device, Yield Mapping, Yield Prediction

Introduction

The agricultural industry continuously seeks to improve efficiency and productivity, particularly in cultivating high-value crops such as strawberries. Strawberry harvesting presents a unique set of challenges that significantly impact the profitability and sustainability of its production. Among these challenges is the unpredictability of major fruit waves, which complicates planning and resource allocation (Xiong et al., 2020). Additionally, the variability in yield distribution and the difficulty in recognizing the correct ripeness stages of strawberries necessitate precise and timely decision-making. The high labor costs associated with manual harvesting compound these challenges, making pursuing cost-effective solutions imperative (Xiong et al., 2019). Although fully autonomous harvesting solutions have been explored, their adoption is hindered by high costs, complexity, limited user-friendliness, and maintenance requirements, underscoring a critical need for innovation in this domain (Rajendran et al., 2023).

In response to these challenges, the critical role of timely yield data becomes apparent. Such data can revolutionize field management by enabling more informed decision-making, optimizing labor scheduling and allocation, and facilitating the development of targeted harvesting strategies. In-time information on yield distribution and ripeness stages allows for adjusting field management plans and harvesting strategies in response to ever-changing field conditions, enhancing efficiency and reducing waste (Bonora et al., 2014; Kawamura et al., 2018; Yang et al., 2019). The importance of timely actions, underpinned by accurate and immediate data, cannot be overstated, as it directly influences the success of the harvesting operation and the overall yield quality (Peng et al., 2022).

The primary objective of this work was to develop a precise, lightweight, and user-friendly detection system capable of assisting in creating a strawberry fruit distribution map with GPS data. This system can leverage the proposed innovative AgriBerry Vision (ABV) imaging module equipped with a power switch and a system switch. It was designed to capture images, run a detection model, and record detection results along with the GPS coordinates of each image. The gathered information will then be processed to create a comprehensive yield map for the field, a vital tool for optimizing harvesting operations.

A novel detection model was proposed to meet the demand for a lightweight and efficient solution. This model is based on YOLOv8 (Jocher et al., 2023). It has been restructured with the inclusion of the coordinate attention (Hou, Zhou, and Feng, 2021) and the substitution of the C2f component with the FasterNet Block (Chen et al. 2022), which includes the integration of Efficient Multi-Scale Attention (EMA), leading to the creation of the C2f-Faster-EMA mechanism (Chen et al., 2024). The proposed model represents the lightest among the YOLOv8-nano-based models, offering significant processing speed and accuracy advantages. The limitations of current harvesting practices can be surmounted by integrating the proposed model with the imaging module, yielding a cost-effective and user-friendly solution. Such a system is designed to adapt to the dynamic nature of strawberry production.

The proposed system introduces a paradigm shift in strawberry harvesting, aiming to address critical challenges by applying advanced imaging and detection technologies. Integrating in-time yield data and GPS mapping facilitates a more strategic approach to harvesting, enabling significant improvements in efficiency, cost management, and yield quality. As moving forward, the potential of this technology to transform agricultural practices and contribute to the sustainability of high-value crop production is both promising and profound.

System Development and Methodology

Hardware and System Design

The AgriBerry Vision (ABV) imaging module was developed to provide a streamlined, effective, and user-friendly solution for strawberry production. Fig. 1 illustrates the hardware components of the system. The design is characterized by its simplicity and operational ease, comprising solely a power switch on the power bank and a toggle switch that governs the system's functionality.

The software architecture of ABV facilitates a high level of integration, thereby simplifying the operation of the hardware. The user activates the power to initiate the operation, and then the system switches. The ABV captures images every five seconds. Throughout the five-second interval, the pre-trained modified YOLOv8 model identifies the target and logs the detection and counting data for the number of strawberry flower and fruit, saved alongside the GPS coordinates in a CSV file on a flash drive. After completing the coverage of the designated area, the user deactivates the system power switches and retrieves the flash drive. The data extracted facilitate the mapping of yield distribution across the strawberry field and support subsequent yield prediction analyses.

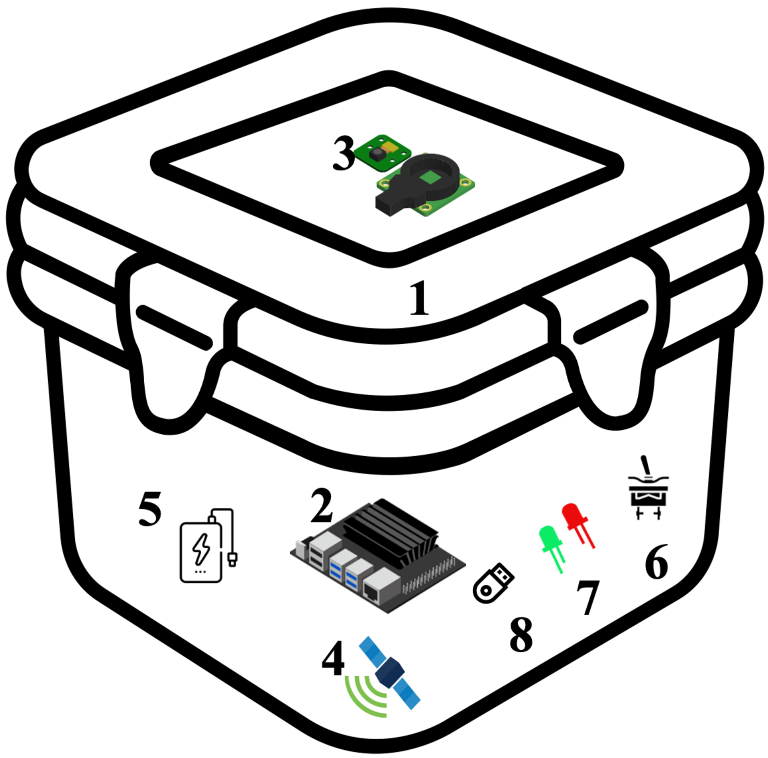


Figure 1. The Agriberry Vision Imaging Module (ABV) is equipped with various hardware components: (1) one ZH-080604-06 case (Polycase, Avon, OH, USA), (2) the processing unit is the JATSON NANO (NVIDIA, Santa Clara, CA, USA), (3) imaging is done using an IMX477 camera (ArduCam, Nanjing, China), (4) location tracking is enabled by a GPS Receiver BU-353N (GlobalSat, New Taipei City, Taiwan), (5) power is supplied by a power bank (Charmast, Shenzhen, China), (6) control is managed through an AST series Toggle Switch (CIT Relay % Switch, Roger, MN, USA), (7) and indication signal is provided by LED lights. (8) Processed data will be saved to a flash drive (note that this could be any flash drive the user prefers). Each piece of hardware enhances the functionality and efficiency of the module.

It is designed with the end-user in mind, specifically for farmers, and will be mounted on a sprayer for regular agricultural operation. During the testing phase, the ABV was mounted on the Autonomous Strawberry Imaging Robot (ASTIR) (Fig. 2) to facilitate initial trials.

Methodology

YOLOv8

A computer vision detection model is needed to monitor the growth of strawberries. In the rapidly advancing domain of object detection algorithms, YOLOv8 (Jocher, Chaurasia, and Qiu, 2023) emerges as a significant enhancement within the celebrated "You Only Look Once" (YOLO) series developed by Ultralytics. YOLOv8 presents improvements that elevate its performance, adaptability, and efficiency over its predecessors. YOLOv8-nano was chosen as the baseline model due to its real-time performance capabilities and efficacy on edge devices. It possesses the fewest parameters and demonstrates the fastest operational speed (minimal inference time) among other variants while maintaining detection accuracy within this research.

Figure 2. ABV (in the red circle) attached to ASTIR

This study considered three classes for simplicity and efficiency: flower, immature fruit, and mature fruit. It is widely recognized that flowers and fruits are small and often partially occluded. To tackle those specific challenges of this application, enhancements were concentrated on using two modules in the proposed model (Fig. 3): the Coordinate Attention (CA) (Hou, Zhou, and Feng, 2021) module and the C2f-Faster-EMA (CFE) module (Chen et al., 2024). The CA Module (Fig. 4f) bolsters the model's ability to discern critical spatial features while mitigating interference from irrelevant components. This functionality proves vital in settings where target objects are diminutive and may be camouflaged by foliage or other environmental elements. By incorporating location data into channel attention, the CA module creates an attention graph capable of capturing long-range dependencies and maintaining precise location information. This feature is particularly beneficial for accurately detecting small and occluded objects within the application. The module is designed to be straightforward and lightweight, allowing for easy integration into current mobile network frameworks such as MobileNetV2 (Sandler et al., 2019), MobileNeXt (Daquan et al., 2020), and EfficientNet (Tan and Le, 2020), with minimal additional computational load. Extensive experimental results reveal that the CA module excels in ImageNet classification tasks and significantly boosts performance in downstream applications like target detection and semantic segmentation. Modifications have been implemented in the CA module for cattle detection, which includes integrating deformable convolution and CA mechanisms within the YOLOv8 architecture to refine spatial information and highlight essential features in the detection process (Yang et al., 2023).

The C2f-Faster-EMA (CFE) Module (Fig. 4b) represents another critical component originally designed to enhance performance in grape detection tasks. This module amalgamates the FasterNet Block (Fig. 4c) and the Efficient Multi-Scale Attention (EMA) (Ouyang et al., 2023) mechanism (Fig. 4g), which effectively learns channel descriptions without reducing the channel dimension in convolutional operations. The EMA's capability to foster enhanced pixel-level attention augments feature representations in sophisticated feature maps, essential for identifying small, partially occluded objects like grapes amid foliage. By combining the FasterNet Block and EMA, the CFE module improves feature extraction efficiency and substantially elevates the model’s ability to discern critical features under complex conditions. This configuration enables the YOLOv8-GP model to execute precise, concurrent detection of grape bunches and picking points while maintaining a low computational demand (Chen et al., 2024).

These advancements facilitate the model's achievement of efficient target detection for agricultural automation scenarios, which require high levels of real-time performance and efficacy. This necessitated the initial selection of these specific enhancements. Specifically, all C2f modules in the YOLOv8 backbone were replaced with CA modules and those in the detection head. Furthermore, the feature dimensions in the backbone were reduced to ensure enhanced performance in detecting small objects.

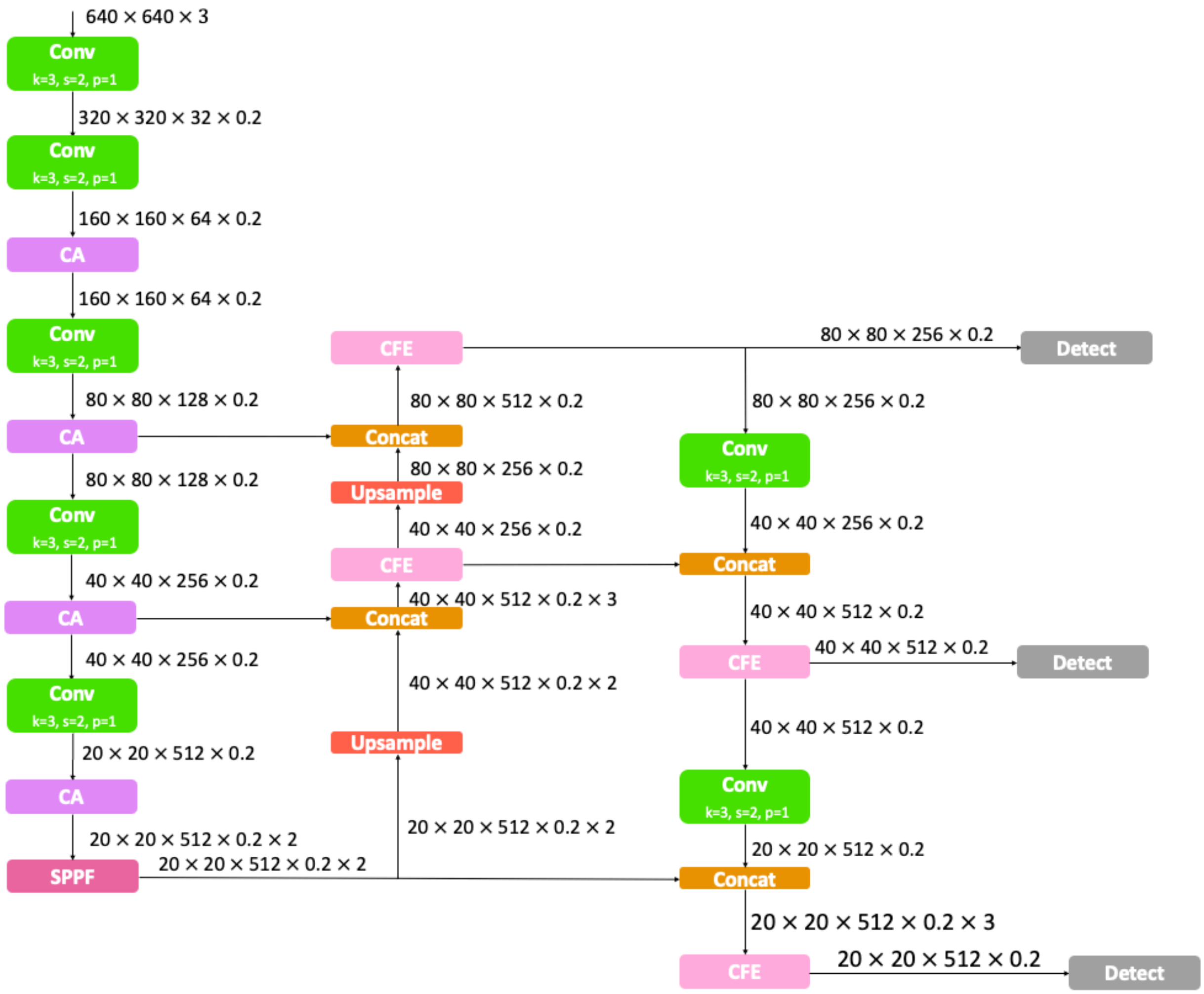
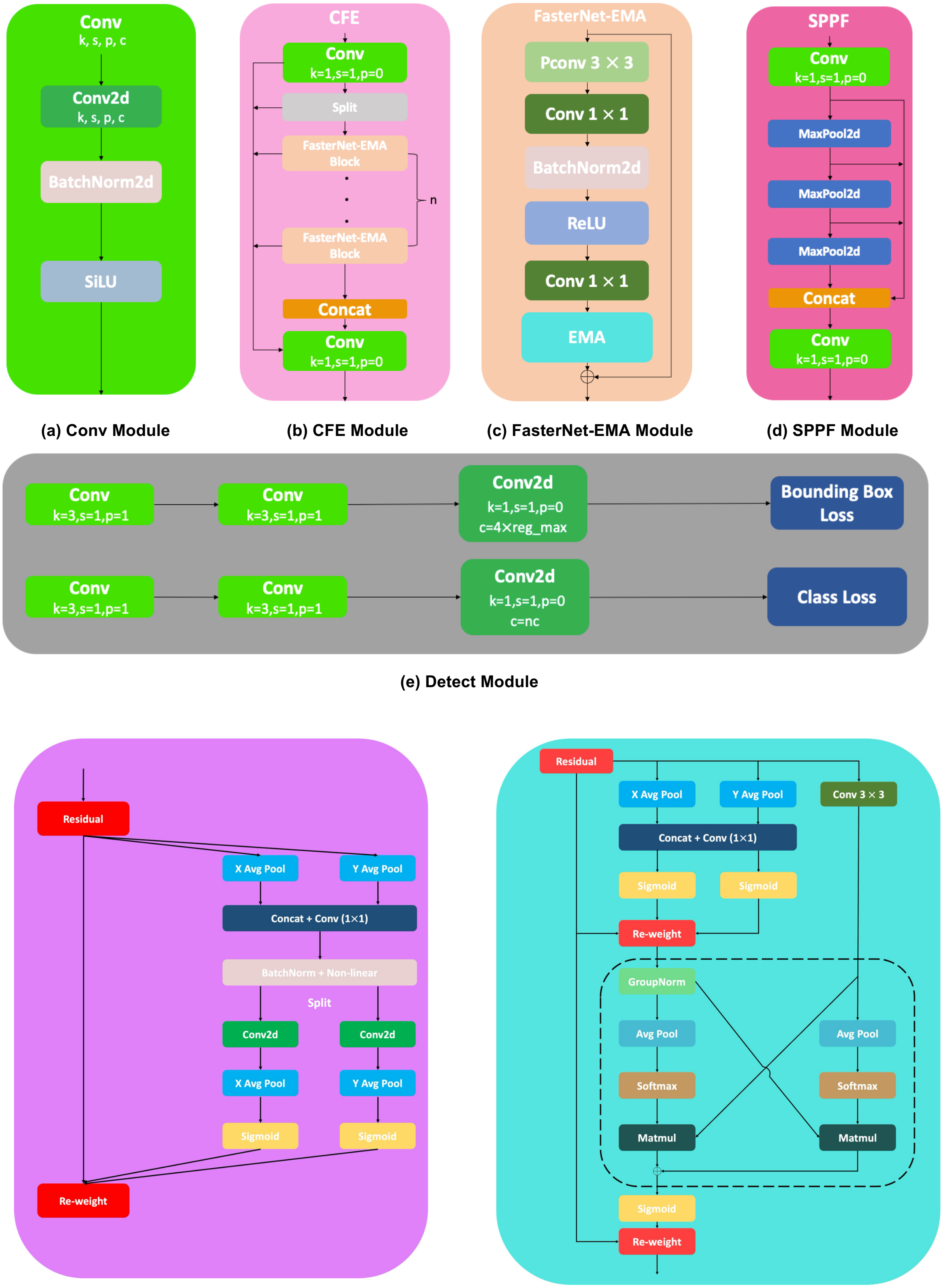


Figure 3. Proposed Modified YOLOv8 Architecture

Data Collection for Training

The data were collected at the University of Florida/IFAS Plant Science Research and Education Unit (PSREU) in Citra, FL, United States (29.40432° N, 82.14161° W). Collection occurred weekly from March to May 2023. An IMX477 camera (ArduCam, Nanjing, China) was used for the data collection. The dataset comprised 1,581 images, each with a resolution of 640 by 480 pixels. This dataset was divided into training, test, and validation subsets, respectively 80%, 10%, and 10%. The training subset was augmented into 3,333 images through vertical and horizontal flips. Within the study, three detection classes were established: flower, immature fruit, and mature fruit.



(a) Conv Module (b) CFE Module (c) FasterNet-EMA Module (d) SPPF Module

(e) Detect Module

(f) Coordinate Attention Module (g) EMA Module

Figure 4. Functional Modules of Proposed Modified YOLOv8 Architecture

Yield Mapping with ABV

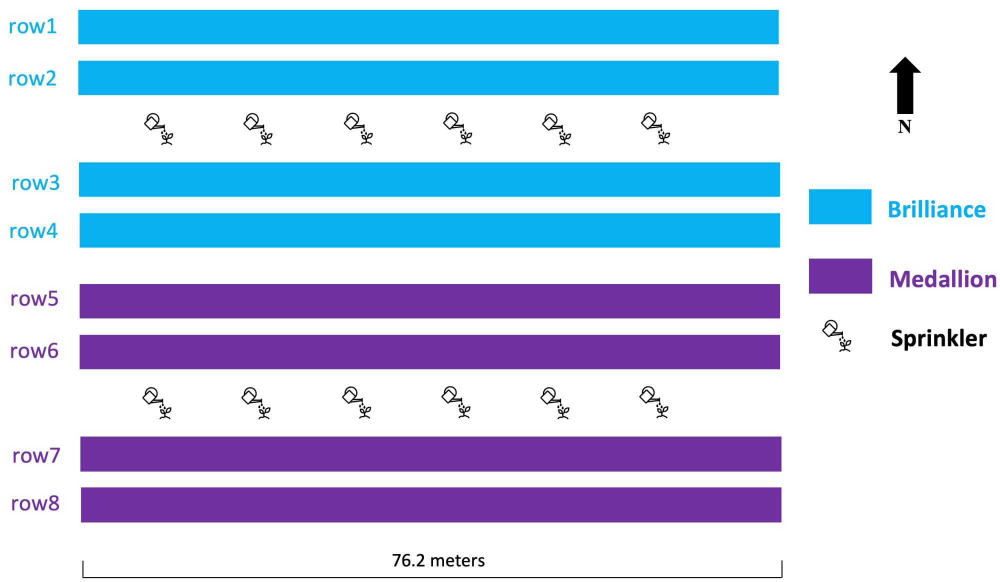


Figure 5. Layout of The Strawberry Field

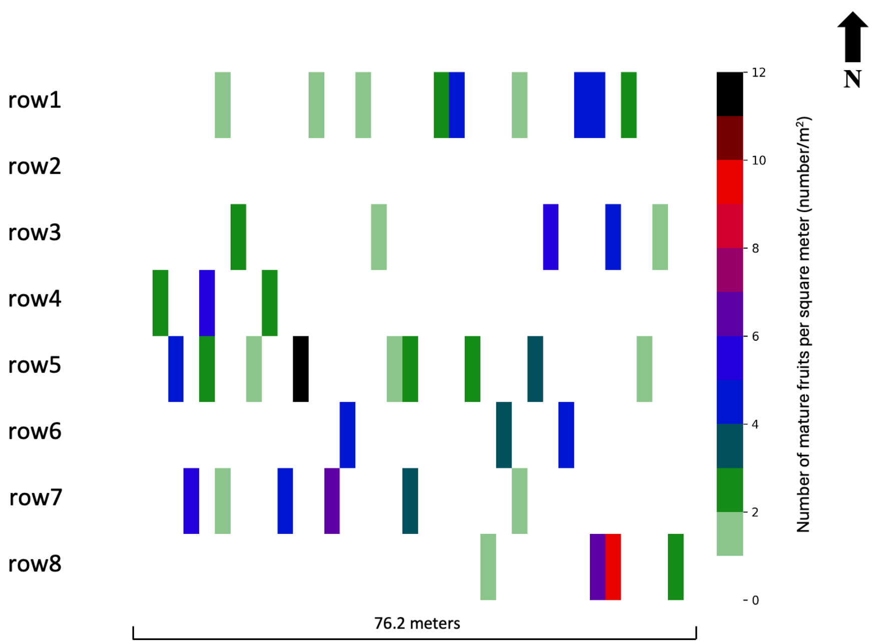


Figure 6. Yield Mapping Results from ABV

The field experiment was conducted at the University of Florida/IFAS Plant Science Research and Education Unit (PSREU) in Citra, Florida, United States (29.40432° N, 82.14161° W). Two strawberry cultivars, Brilliance and Medallion, both developed by the University of Florida, were included. The field was organized into ten rows, with the northern five rows containing the Brilliance cultivar and the southern five rows containing the Medallion cultivar. Rows with strawberry plants were indexed from one to eight, progressing north to south (Fig. 5).

Fig. 6 provides a yield map generated using ABV, offering valuable insights for farmers. Strawberry yield variability can be clearly observed, which is the main goal of yield mapping. However, the map includes an empty row (row 2). It exhibits uneven detection, even with a five-second interval set for data processing (collection and detection), making the detection not evenly distributed across the map. This issue primarily arises from the GPS system's limited accuracy, which has a positioning accuracy of one meter and impacts overall system performance. The limitation reflects a compromise between cost and positioning accuracy, given the project's aim to create an affordable and user-friendly system. Thus, affordability was prioritized over higher GPS accuracy.

Results

Detection Benchmark

Table 1. Comparisons between models with different architectures when taking YOLOv8-nano as the base model

Model	GFLOPS	Parameters	Jetson Nano Time (ms)	mAP	mAP50-95
YOLOv8-nano	8.1	3,236,505	1128.9	0.923	0.595
YOLOv8-nano (modified)	5.3	1,753,281	1000.4	0.911	0.562
Proposed	**4.9**	**1,653,463**	**963.0**	0.912	0.559

The proposed model balances performance and efficiency, incorporating CA and CFE modules. This model is evaluated on the NVIDIA Jetson Nano (NVIDIA, Santa Clara, CA, USA). It boasts the lowest computational demand at 4.9 GFLOPS and the fastest average inference time of 963 ms on the Jetson Nano while maintaining a high mean Average Precision (mAP) of 0.912 and a mAP50-95 of 0.559. Additionally, it has the smallest parameter count at 1,653,463, demonstrating a streamlined architecture. The modified version of YOLOv8-nano, with reduced feature dimension size to align with the proposed model, shows a slightly higher computational load of 5.3 GFLOPS and average inference times of 1000.4 ms on the Jetson Nano. It achieves mAP scores close to the proposed model at 0.911 and a slightly higher mAP50-95 at 0.562, with a parameter count of 1,753,281. The original YOLOv8-nano, requiring more computational resources at 8.1 GFLOPS and the largest parameter count at 3,236,505, delivers the highest mAP of 0.923 and mAP50-95 of 0.595, illustrating a trade-off between complexity and performance.

This comparison underscores the effectiveness of the CA and CFE modules in providing efficient performance metrics and balancing low computational requirements with high accuracy, which is crucial for applications that demand both speed and precision.

Conclusion

Developing the AgriBerry Vision Imaging Module (ABV), equipped with the modified YOLOv8-nano model, significantly advances strawberry yield mapping. This system provides a robust solution to the complexities of strawberry harvesting, marked by its user-friendly design and real-time data processing capabilities.

Incorporating the Coordinate Attention and C2f-Faster-EMA modules into the YOLOv8 framework has resulted in a model that excels in operational efficiency while maintaining accuracy, thus supporting its deployment in practical agricultural settings. The ability of the ABV system to classify strawberry ripeness along with their locations demonstrates the effective integration of machine learning and edge computing, optimizing farm management and labor allocation.

Future research will focus on refining the detection of more growth stages and incorporating additional functionalities to enhance the system's assessment of crop conditions. TensoRT could be included for model acceleration on the edge device. These improvements aim to broaden the applicability of the ABV system beyond strawberries, establishing a foundation for its adoption across various aspects of precision agriculture, such as enhanced yield prediction. Such progress is anticipated to catalyze a global shift toward more intelligent and environmentally conscious agricultural practices.

Acknowledgments

This study was supported by a USDA NIFA grant (Award number: 2023-67021-40618).

References

Bonora, Elisa, Massimo Noferini, Dario Stefanelli, and Guglielmo Costa. 2014. "A New Simple Modeling Approach for the Early Prediction of Harvest Date and Yield in Nectarines." *Scientia Horticulturae* 172:1–9. doi: 10.1016/j.scienta.2014.03.030.

Chen, Jiqing, Aoqiang Ma, Lixiang Huang, Hongwei Li, Huiyao Zhang, Yang Huang, and Tongtong Zhu. 2024. "Efficient and Lightweight Grape and Picking Point Synchronous Detection Model Based on Key Point Detection." *Computers and Electronics in Agriculture* 217:108612. doi: 10.1016/j.compag.2024.108612.

Daquan, Zhou, Qibin Hou, Yunpeng Chen, Jiashi Feng, and Shuicheng Yan. 2020. "Rethinking Bottleneck Structure for Efficient Mobile Network Design."

Hou, Qibin, Daquan Zhou, and Jiashi Feng. 2021. "Coordinate Attention for Efficient Mobile Network Design."

Kawamura, Kensuke, Hiroshi Ikeura, Sengthong Phongchanmaixay, and Phanthasin Khanthavong. 2018. "Canopy Hyperspectral Sensing of Paddy Fields at the Booting Stage and PLS Regression Can Assess Grain Yield." *Remote Sensing* 10(8):1249. doi: 10.3390/rs10081249.

Ouyang, Daliang, Su He, Guozhong Zhang, Mingzhu Luo, Huaiyong Guo, Jian Zhan, and Zhijie Huang. 2023. "Efficient Multi-Scale Attention Module with Cross-Spatial Learning." Pp. 1–5 in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Peng, Chen, Stavros Vougioukas, David Slaughter, Zhenghao Fei, and Rajkishan Arikapudi. 2022. "A Strawberry Harvest-Aiding System with Crop-Transport Co-Robots: Design, Development, and Field Evaluation." *Journal of Field Robotics* 39(8):1231–57. doi: 10.1002/rob.22106.

Vishnu Rajendran, Bappaditya Debnath, Bappaditya Debnath, Sariah Mghames, Willow Mandil, Soran Parsa, Simon Parsons, and Amir Ghalamzan-E. 2023. "Towards Autonomous Selective Harvesting: A Review of Robot Perception, Robot Design, Motion Planning and Control."

Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2019. "MobileNetV2: Inverted Residuals and Linear Bottlenecks."

Tan, Mingxing, and Quoc V. Le. 2020. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks."

Xiong, Ya, Yuanyue Ge, Lars Grimstad, and Pål J. From. 2020. "An Autonomous Strawberry-Harvesting Robot: Design, Development, Integration, and Field Evaluation." *Journal of Field Robotics* 37(2):202–24. doi: 10.1002/rob.21889.

Xiong, Ya, Cheng Peng, Lars Grimstad, Pål Johan From, and Volkan Isler. 2019. "Development and Field Evaluation of a Strawberry Harvesting Robot with a Cable-Driven Gripper." *Computers and Electronics in Agriculture* 157:392–402. doi: 10.1016/j.compag.2019.01.009.

Yang, Qi, Liangsheng Shi, Jinye Han, Yuanyuan Zha, and Penghui Zhu. 2019. "Deep Convolutional Neural Networks for Rice Grain Yield Estimation at the Ripening Stage Using UAV-Based Remotely Sensed Images." *Field Crops Research* 235:142–53. doi: 10.1016/j.fcr.2019.02.022.