

Pitfalls in Multispectral In-Field Strawberry Detection: Challenges and First Results for Ripeness Classification in Robotic Harvesting [★]

Timo Lange^{1,2} ✉, Tim Tiedemann¹, Kai von Luck¹, Peer Stelldinger¹, and Keshav Dahal²

- ¹ Faculty of Computer Science and Digital Society, Hamburg University of Applied Sciences, Hamburg, Germany
{timo.lange,Tim.Tiedemann,Kai.vonLuck,peer.stelldinger}@haw-hamburg.de
- ² School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley, Scotland
keshav.dahal@uws.ac.uk

Abstract. In agriculture, autonomous robotic harvesting of strawberries in unstructured outdoor fields faces significant challenges from variable lighting, occlusions, and fruit clustering. While RGB-based detection is common, multispectral imaging (MSI) promises enhanced ripeness classification, yet in-field applications remain underexplored due to data scarcity and technical pitfalls. This paper presents a novel MSI dataset from a commercial strawberry farm, comprising 188 recordings with 1690 strawberries, annotated as unripe, ripe, or overripe - extending beyond binary ripe/unripe schemes. We adapt YOLOv8 for MSI input (14 spectral channels from ultraviolet (UV) to short-wave infrared (SWIR)) and compare baselines against RGB counterparts, revealing key challenges like imperfect image registration and high dynamic range (HDR) artifacts, resulting in MSI underperformance (YOLOv8 mAP@50: 0.754 MSI vs. 0.783 RGB). Our findings highlight deployment pitfalls for robotic systems and provide a foundation for future in-field MSI advancements. To our knowledge, this is the first study analyzing high-resolution, aligned MSI data for multi-class strawberry ripeness in realistic field conditions.

Keywords: Deep Learning · Multispectral Imaging · Detection · Fruit Harvesting.

1 Introduction

Robotic systems offer significant potential to improve resource efficiency and food security, contributing to several Sustainable Development Goals (SDGs)³, including *zero hunger (2)*, *good health and well-being (3)*, *responsible consumption*

[★] This work was funded by the German Federal Ministry of Food and Agriculture (BMEL) within the RoLand project (grant no. 28DK103C20).

³ <https://sdgs.un.org/goals>

and production (12), and climate action (13). Accordingly, robotics is increasingly applied to precision agriculture (PA), particularly autonomous in-field fruit harvesting. Harvesting labour-intensive and delicate fruits such as strawberries remains challenging. Their sensitivity requires careful handling, precise localization, and accurate gripper guidance. In addition, fruit ripeness and quality must be assessed to distinguish saleable fruit from overripe, damaged, malformed, or diseased specimens. In unstructured outdoor fields, further challenges arise from occlusions by leaves and stems, overlapping fruit, and highly variable lighting conditions. Current research predominantly focuses on red, green, blue (RGB)-based detection for in-field or greenhouse environments, typically addressing fruit detection or segmentation and, less frequently, ripeness classification. In contrast, MSI and hyperspectral imaging (HSI) are mainly studied in laboratory settings, where they are used to assess ripeness, firmness, and disease under controlled lighting with bulky equipment and complex procedures unsuitable for field deployment [2]. Such studies often rely on small datasets and relatively simple pixel-based methods (e.g., support vector machine (SVM), multilayer perceptron (MLP), or random forests) and do not address the complexity of dynamic field conditions.

Recent work has proposed using MSI and machine learning (ML) for in-field strawberry harvesting [21], with first results on in-field ripeness classification reported in [16]. Building on these efforts, this study combines the advantages of MSI, promising improved ripeness discrimination, higher spatial resolution than typical HSI setups, and suitability for robotic deployment, with modern deep learning models.

We investigate the challenges and pitfalls of multi-sensor MSI data for multi-class strawberry detection under realistic field conditions. To this end, an in-field dataset was collected and used to compare RGB- and MSI-based object detection using YOLOv8, Real-Time DEtection TRansformer (RT-DETR), and a convolutional block attention module (CBAM)-enhanced YOLOv8 model.

2 Related Work

In recent years, numerous systems for fruit detection, classification, and harvesting have been proposed. Strawberry detection and segmentation predominantly rely on RGB imaging, while ripeness, pest, and quality assessment increasingly leverage additional spectral information such as MSI and HSI. Across these domains, deep learning (DL) methods dominate detection and segmentation, whereas laboratory-based classification tasks often also employ classical ML approaches, including SVM, clustering, and simple MLPs.

In-field, real-time capable systems using HSI or MSI data remain rare and are generally unsuitable for autonomous harvesting applications like [4]. Furthermore, publicly available strawberry datasets are scarce, typically limited to the visual (VIS) spectrum, small in size, or collected under laboratory conditions without challenging environment. A recent overview of trends in HSI- and DL-based fruit quality analysis is given by [12].

Several autonomous harvesting systems have been proposed [6,25,10]. For sensitive fruits such as strawberries, most work focuses on elevated-trough cultures, where structured environments simplify perception. In-field harvesting systems remain limited, as they are often bulky, costly, difficult to scale, or too slow for practical deployment. A survey on fruit ripeness classification is provided in [18]. Despite progress, the integration of robust perception with real-time, autonomous harvesting in unstructured environments remains challenging.

Most in-field strawberry detection systems rely on VIS-based sensing using RGB or red, green, blue, depth (RGB-D) sensors. Depth-enhanced systems have been explored in multiple harvesting platforms [6,5,25,10]. Ge et al. [6] use Mask R-CNN for instance segmentation in a table-top harvesting robot, while later work [5] investigates RGB-D and time of flight (ToF) sensors with YOLOv4 for ripeness detection and 3D localization.

YOLO-based architectures are widely adopted due to their real-time performance. Comparative and optimized YOLO variants have been proposed for ripe and unripe strawberry detection, peduncle and stalk estimation, and 3D localization [15,13,1,22]. These works explore architectural modifications such as attention mechanisms, deformable convolutions, feature pyramid enhancements, and pruning strategies to improve accuracy and inference speed. Mask R-CNN remains relevant for instance segmentation, including ridge-planting based systems with picking point estimation [26]. unmanned aerial vehicle (UAV)-based RGB imaging has also been applied for strawberry flower and maturity detection under varying acquisition conditions [29].

Several systems rely on classical image processing. These include color thresholding in RGB and Hue, Saturation, Value (HSV) spaces with adaptive illumination handling using infrared (IR)-sensors [25], R and G channel-based segmentation [10]. Feature-based classification using handcrafted descriptors combined with SVM has also been explored for damaged fruit detection [14].

MSI and HSI enable the analysis of internal fruit properties and subtle surface characteristics that are not accessible in the visible spectrum. Overviews of agricultural HSI and MSI methods, including ML, DL, and transfer learning approaches, are provided in [17,30]. Devassy et al. [2,3] demonstrate that strawberry firmness can be estimated accurately in laboratory conditions using 186-band HSI, with effective band reduction via SPA. Laboratory-based studies further investigate strawberry maturity, soluble solid content (SSC), quality parameters, defect detection, and disease identification using HSI and MSI combined with both DL and classical ML methods [20,23,28].

Despite promising laboratory results, in-field applications of HSI and MSI remain limited. First in-field results for strawberry ripeness classification using MSI are reported in [16]. Gao et al. [4] investigate early ripe and ripe classification using HSI in both laboratory and field settings with SVM and AlexNet, applying feature selection and principal component analysis (PCA). Jol et al. [7] evaluate in-field quality estimation using RGB, Orange-Cyan-Near-Infrared (OCN), and environmental data, finding that RGB performs best under uncontrolled lighting conditions. Kirk et al. [9] demonstrate improved detection performance by

fusing RGB and CIELab color spaces using a ResNet-18–RetinaNet architecture. These results highlight the practical challenges of deploying multi-spectral sensing outside controlled environments.

The reviewed literature reveals several persistent gaps. Overripe strawberries are rarely considered, despite their importance for quality control and disease prevention. MSI and HSI data are predominantly collected under laboratory conditions, while existing in-field systems are often unsuitable for robotic harvesting [4] or do not fully exploit spectral information for ripeness prediction [7]. Most multi-spectral studies rely on small datasets, whereas in-field approaches largely depend on RGB or RGB-D sensing. Additionally, the frequent use of low-resolution imagery limits the exploitation of fine-grained fruit characteristics such as surface texture.

3 Multispectral Strawberry Detection

This section outlines the methodology for investigating MSI-based strawberry detection under field conditions. The goal is to assess the applicability of MSI for improving strawberry detection and ripeness classification in uncontrolled outdoor environments, which remains largely underexplored.

Deploying MSI-based detection for robotic harvesting in standard outdoor fields requires defining the experimental setup, collecting data, and applying appropriate preprocessing for ML training. Based on these requirements, an MSI-enabled model is developed and compared against an RGB baseline. The following sections describe these steps in detail.

3.1 Experimental Setup

Strawberry detection targets standard outdoor fields with plants arranged in rows on ridges. The final sensor setup is intended for robotic deployment, similar to the platform proposed by Tiedemann et al. [21]. For initial experiments and data collection, a larger sensor setup with additional spectral bands is used to explore effective spectral combinations. The sensor configuration employed in this study is shown in Figure 1.



Fig. 1: MSI image acquisition setup. Filter wheels with bandpass filters are attached to the SWIR (left) and near infrared (NIR) (right) camera. The UV sensor is placed in the top center, the RGB sensor underneath.

Table 1: MSI strawberry dataset.

Data Split	MSI Images	Instances			
		All	Unripe	Ripe	Overripe
Training	148	1354	845	436	73
Validation	21	164	97	59	8
Test	19	172	104	59	9

Data collection uses four sensors covering the UV, VIS, NIR, and SWIR ranges. Two filter wheels with bandpass filters are mounted in front of the NIR and SWIR sensors, yielding a total of 14 spectral channels. The captured spectra are: **UV** (190 nm–1100 nm) **VIS** (400 nm–700 nm) **NIR** (400 nm–1000 nm, 730 nm–755 nm, 830 nm–865 nm, 928 nm–955 nm) **SWIR** (400 nm–1700 nm, 930 nm–1030 nm, 1290 nm–1310 nm, 1440 nm–1460 nm, 1485 nm–1645 nm).

3.2 Multispectral Strawberry Dataset

The initial dataset consists of 188 recordings with 148 training, 21 validation and 19 test samples. Detailed class distribution is provided in Table 1. For more robust evaluation, a 5-fold cross-validation dataset was compiled with 41, 36, 35, 42, and 34 samples per fold. Folds are stratified to preserve class distributions. Classes include unripe, ripe, and overripe strawberries, each annotated with bounding boxes.

Figure 2a shows a representative RGB image from the dataset. The corre-



(a) RGB image of one recording, showing some ripe and unripe strawberries. (b) False color image, showing the channel shift between the SWIR and RGB sensors.

Fig. 2: RGB and false color image of one recording.

sponding MSI channels are shown in Figure 3, where several dataset peculiarities

are visible. Subfigure 3a is blurred, likely due to a non-apochromatic, low-focal-length lens. In subfigures 3b and 3e, reflections of the lens in the bandpass filters appear, probably caused by short-shifting effects that allow light to pass through filters that should block it—a common issue with lenses shorter than 12 mm. Some filters employ *StableEDGE*⁴ technology to mitigate this, but the affected 940 nm and 1450 nm filters do not. Additionally, subfigures 3c and 3d show artifacts from HDR processing, which occur when exposure bracketing includes heavily overexposed images. Image registration is also not pixel-perfect,

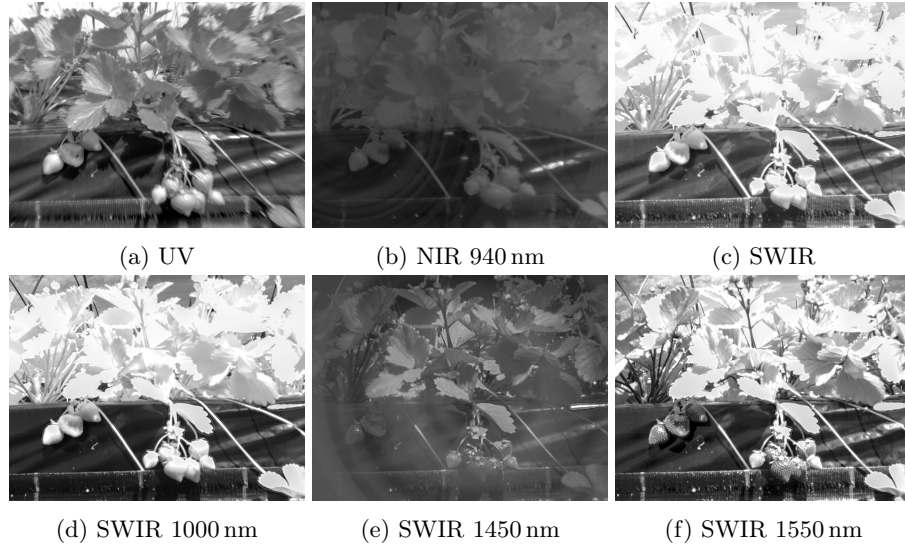


Fig. 3: Selected MSI channels of the recording shown in fig 2a.

as seen in the false-color image in Figure 2b, where the SWIR channel is mapped to B and the VIS channels to G and R. Misalignment is visible on stems and fruits.

Data Acquisition. The images are taken ~ 30 cm above the ground at a right angle and ~ 50 cm from the strawberries. Recordings are captured without artificial light on three different days with varying weather and lighting conditions. For each recording and spectrum, exposure bracketing is captured and processed into HDR images to account for lighting differences.

Image Registration. Images from all sensors and spectral bands are registered to the VIS camera at the center of the setup (section 3.1, figure 1). Registration is performed using a checkerboard calibration at the start of each strawberry row. Feature-based registration was explored, but finding distinctive features proved difficult, as stems and leaves appear very similar and non-linearities between spectra hinder reliable matching [19,11].

⁴ <https://midopt.com/stabledge/>

Preprocessing. To use the MSI images with ML models, further processing is required. First, the different exposure brackets of each spectral channel are combined into HDR images, ensuring no images are completely over- or underexposed. Second, the largest common overlap of all spectral images is determined, and the lower-resolution images are upsampled to match the resolution of the VIS sensor, which has the highest resolution. Finally, the aligned and cropped spectral images are fused into a single MSI cube.

Labelling. Segment Anything Model (SAM) [8] was used to assist, with manual corrections to improve mask precision. All segmentation masks are based on RGB images from the VIS sensor, to which all other MSI channels were registered (section 3.2). Individual masks per channel, e.g., for enhanced registration, were not considered. Annotations by one person were reviewed and corrected by another, with roles rotated to ensure quality.

3.3 YOLOv8 MSI Modifications

Related work shows that YOLO variants perform well in strawberry detection. As a baseline for RGB and MSI data, YOLOv8 was chosen. It already incorporates techniques effective for this domain, such as feature pyramids for detecting strawberries of varying sizes, focal loss for handling class imbalance, and extensive built-in data augmentation. The model is lightweight, supports multiple predefined sizes, and is suitable for deployment on robots with constrained resources. Its CNN-based architecture also reduces overfitting risk on small datasets. To use YOLOv8 with MSI data, the ultralytics implementation must be adapted, as it is designed for RGB input.

Data Loading. The data loading mechanism was modified to read MSI data from NumPy arrays and to allow configuration of the used channels, facilitating experiments with different spectral combinations.

Data Augmentation. Data augmentation is a key component of the ultralytics framework and was adapted to fully support MSI training. Built-in augmentations were modified for multispectral data, while techniques not applicable to MSI, such as hue adjustments in HSV space, were disabled.

Model Configuration. The YAML-based model configuration was adapted to support an arbitrary number of input channels. Additionally, configuration parsing was extended to initialize the CBAM layer.

CBAM. [24] was integrated into YOLOv8 to provide spatial and channel-wise attention, with best results achieved by inserting a single module after the first backbone *Conv* layer. The ultralytics implementation is used, replacing max-average pooling with PyTorch’s `adaptive_avg_pool2d` and the original MLP with a convolutional neural network (CNN).

RT-DETR. Additionally, an MSI-adapted version of RT-DETR [27] was compared with YOLOv8 to include a real-time optimized transformer model. The model combines CNNs and transformers, consisting of a ResNet50-like backbone, a hybrid encoder, and a transformer decoder with auxiliary prediction heads.

Table 2: MSI and RGB detection results.

Model	Precision		Recall		mAP50		mAP50-95	
	Val	Test	Val	Test	Val	Test	Val	Test
YOLOv8 RGB	0.876	0.699	0.688	0.763	0.809	0.762	0.597	0.472
YOLOv8 MSI	0.78	0.723	0.697	0.57	0.75	0.613	0.519	0.381
CBAM RGB	0.83	0.721	0.711	0.661	0.791	0.733	0.572	0.459
CBAM MSI	0.857	0.608	0.789	0.583	0.824	0.619	0.596	0.398

3.4 Training

For comparison, all models were trained from scratch, as only RGB-pretrained weights are available. Standard framework configurations were used with the *AdamW* optimizer and batch sizes of 32 for the n and s scales, 16 for m , and 8 for l and x . Batch sizes were fixed per scale and chosen to fully utilize available GPU *VRAM* at the highest resolution. Default data augmentations unsuitable for MSI, such as HSV color space modifications, were disabled. Early stopping was applied with a maximum of 1000 epochs and a patience of 100.

4 Results

Results on the initial dataset using the standard training, validation, and test split are shown in Table 2. Initial experiments were limited to YOLOv8, its MSI-adapted variant, and the CBAM-extended model, RT-DETR was evaluated only in the cross-validation setting. Because validation and test results differed substantially, additional experiments were conducted using 5-fold cross-validation. In total, seven models (YOLOv8 n , s , m , l , x and RT-DETR ResNet50, l) were trained and evaluated across three image resolutions (640, 1024, 2010) and two modalities (RGB, MSI), resulting in 210 trained models. As the full results comprise 42 rows (including standard deviations over five folds), only selected results are reported here. Table 3 presents the models with the best mAP50 per modality, along with their precision, recall, and mAP50–95. The best performance for each model and modality was consistently achieved at the highest resolution, with resolution having a greater impact than model size. The top-performing models were always trained at image resolutions 2010 or 1024 pixels.

5 Discussion

As shown in Section 4 (Table 2), the RGB YOLOv8 model outperforms the unmodified MSI version across all metrics. This is likely due to the partly heavy displacement between the different MSI channels (Section 3.2) and the ground

Table 3: Best detection results per model and modality for 5-fold-cross validation. The best results are always achieved with the highest image resolution.

Model	Precision		Recall		mAP50		mAP50-95	
	Val	Std	Val	Std	Val	Std	Val	Std
YOLOv8 RGB(m)	0.780	0.075	0.711	0.061	0.783	0.041	0.594	0.033
YOLOv8 MSI(s)	0.750	0.058	0.687	0.051	0.754	0.027	0.548	0.032
CBAM RGB(s)	0.820	0.036	0.688	0.063	0.786	0.049	0.593	0.042
CBAM MSI(s)	0.746	0.058	0.673	0.067	0.739	0.029	0.534	0.038
RT-DETR RGB(l)	0.796	0.047	0.651	0.056	0.723	0.040	0.525	0.045
RT-DETR MSI(l)	0.751	0.049	0.664	0.036	0.695	0.025	0.512	0.015

truth annotations on the VIS images. Both models share the same RGB information extracted from the full MSI cube, so the MSI model has the necessary information to match the RGB model. Therefore the MSI model must learn to use the RGB channels for the precise bounding box prediction while using additional channels to improve the classification, respectively take the displacement into account. This suggests that the MSI model is either not able to learn to primarily utilize the RGB channel for the bounding box and the full spectrum to classify the objects or because the larger parameter space makes learning more difficult, especially with the small dataset size.

Artifacts from HDR processing may further hinder the MSI model. Merging exposure brackets can distort the intensity semantics of each wavelength, whose responses are affected differently by varying weather conditions. For example, wavelengths strongly absorbed by water (e.g. 1450 nm) are more sensitive to illumination changes, whereas VIS wavelengths are more stable. This explains why RGB channels have an advantage under uncontrolled lighting and fewer HDR artifacts, especially compared to SWIR bands.

Certain SWIR bands highlight strawberry seeds (1550 nm, 1450 nm, and 1300 nm), whereby qualitative analyses showed a difference between overripe and ripe fruits, suggesting these wavelengths could improve class separation.

However, filter reflections (Section 3.2) obscure portions of the image and vary with lighting, making it harder for the model to learn features in these bands. Tailored preprocessing or additional data may be required to compensate.

Adding a CBAM module improves the RGB model in mAP50 (0.786 vs. 0.783) and precision (0.820 vs. 0.780) and substantially reduces precision variability (0.036 vs. 0.075). Unexpectedly, the same modification reduced performance for the MSI YOLOv8 across all metrics. The RT-DETR model performs worse overall. Its MSI variant also underperforms compared to its RGB counterpart, but interestingly, MSI data reduces variance in mAP50 (0.025 vs. 0.040), mAP50-95 (0.015 vs. 0.045), and recall (0.036 vs. 0.056). Training RT-DETR was more unstable, exhibiting erratic loss and validation behavior per epoch.

Despite the small dataset (188 samples), the models learn effectively. Low to medium cross-validation standard deviations indicate that the results are reliable and not due to a favorable data split. Nevertheless, the initial experiments using a standard training, validation, and test split demonstrate that data splitting can have a strong influence when datasets are small. This data efficiency likely stems from: a) high instance density per image, b) low background diversity, c) few classes. This suggests that adding a few hundred samples could substantially improve performance without requiring thousands of images. Training on MSI data, with its larger input space, may require more samples than RGB to fully leverage the additional spectral information.

The main limitation of this study is dataset size. While addressed via cross-validation, larger datasets are needed for more rigorous evaluation. Ablation studies are also necessary to investigate failure modes and their impact on model performance.

6 Conclusion & Outlook

This study investigated multispectral in-field strawberry detection using a modified YOLOv8, CBAM and RT-DETR. Initial experiments with a standard training, validation, and test split showed large performance variations, indicating strong sensitivity to data splitting, which was addressed using five-fold cross-validation. Unexpectedly, RGB-based models consistently outperformed their MSI counterparts. Further analysis identified several pitfalls likely limiting performance in MSI-based detection. a) insufficient MSI registration with notable channel displacement b) artifacts caused by HDR processing, possibly reducing intensity semantics c) spectral filter reflections obscuring image regions, d) the small dataset, which complicates learning in the higher-dimensional MSI input space. Despite these challenges, the models achieve solid performance using only 188 samples in the cross-validation dataset. The best RGB model, a YOLOv8 with an additional CBAM module, reached an mAP50 of 0.786 with precision 0.820 and recall 0.673 is a YOLOv8 with an additional CBAM module in RGB spectrum. The best MSI model, a YOLOv8, achieved an mAP50 of 0.754, precision 0.750, and recall 0.687. Future work should include ablation studies to quantify the impact of the identified error sources, along with improved image registration and HDR processing. Incorporating additional broadband illumination and adapted auto-exposure for in-field multi-sensor setups represents another promising direction. Given the increased acquisition effort of MSI data and high annotation cost, leveraging RGB-pretrained models for MSI and using generative models to augment datasets with synthetic MSI samples are promising avenues for future research. Finally, as this study focused solely on early fusion, further investigation into alternative data fusion strategies is warranted.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Dai, S., Bai, T., Zhao, Y.: Keypoint detection and 3d localization method for ridge-cultivated strawberry harvesting robots **15**(4), 372. <https://doi.org/10.3390/agriculture15040372>
2. Devassy, B., George, S.: Estimation of strawberry firmness using hyperspectral imaging: a comparison of regression models . <https://doi.org/10.1255/jsi.2021.a3>
3. Devassy, B.M., George, S.: Contactless classification of strawberry using hyperspectral imaging (2688)
4. Gao, Z., Shao, Y., Xuan, G., Wang, Y., Liu, Y., Han, X.: Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning **4**, 31–38. <https://doi.org/10.1016/j.aiia.2020.04.003>
5. Ge, Y., Xiong, Y., From, P.J.: Three-dimensional location methods for the vision system of strawberry-harvesting robots: development and comparison **24**(2), 764–782. <https://doi.org/10.1007/s11119-022-09974-4>
6. Ge, Y., Xiong, Y., Tenorio, G.L., From, P.J.: Fruit localization and environment perception for strawberry harvesting robots **7**, 147642–147652. <https://doi.org/10.1109/ACCESS.2019.2946369>
7. Jol, C., Wen, J., Van Gemert, J.: Non-destructive infield quality estimation of strawberries using deep architectures. In: 2023 IEEE/CVF ICCVW. pp. 515–524. <https://doi.org/10.1109/ICCVW60793.2023.00058>
8. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. In: 2023 IEEE/CVF ICCV. pp. 3992–4003. IEEE. <https://doi.org/10.1109/ICCV51070.2023.00371>
9. Kirk, R., Cielniak, G., Mangan, M.: L*a*b*fruits: A rapid and robust outdoor fruit detection system combining bio-inspired features with one-stage deep learning networks **20**(1), 275. <https://doi.org/10.3390/s20010275>
10. Klaoudatos, D., Moulianitis, V., Aspragathos, N.: Development of an Experimental Strawberry Harvesting Robotic System:. In: Proc. 16th Int. Conf. Informatics Control, Automation and Robotics. pp. 437–445. SCITEPRESS - Science and Technology Publications, Prague, Czech Republic (2019). <https://doi.org/10.5220/0007934004370445>
11. Lange, T., Babu, A., Meyer, P., Keppner, M., Tiedemann, T., Wolff, S., Vögele, T.: First lessons learned of an artificial intelligence robotic system for autonomous coarse waste recycling using multispectral imaging-based method. CISA Publisher
12. Lanke, N.P., Chandak, M.B.: Recent trends in deep learning and hyperspectral imaging for fruit quality analysis: an overview . <https://doi.org/10.1007/s12596-025-02881-7>
13. Lemsalu, M., Bloch, V., Backman, J., Pastell, M.: Real-time CNN-based computer vision system for open-field strawberry harvesting robot **55**(32), 24–29. <https://doi.org/10.1016/j.ifacol.2022.11.109>
14. Mahendra, O., Pardede, H.F., Sustika, R., Suryo Kusumo, R.B.: Comparison of features for strawberry grading classification with novel dataset. In: 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA). pp. 7–12. <https://doi.org/10.1109/IC3INA.2018.8629534>
15. Meng, Z., Du, X., Sapkota, R., Ma, Z., Cheng, H.: YOLOv10-pose and YOLOv9-pose: Real-time strawberry stalk pose detection models **165**, 104231. <https://doi.org/10.1016/j.compind.2024.104231>

16. Meyer, P., Lange, T., Trost, M., Tiedemann, T.: First results of combining RGB segments and multi- spectral pixel classes for strawberry ripeness detection. In: 45. GIL-Jahrestagung, Digitale Infrastrukturen für eine nachhaltige Land-, Forst- und Ernährungswirtschaft, pp. 333–338. Gesellschaft für Informatik e.V. https://doi.org/10.18420/giljt2025_38
17. Qin, J., Chao, K., Kim, M.S., Lu, R., Burks, T.F.: Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering* **118**(2), 157–171 (2013). <https://doi.org/10.1016/j.jfoodeng.2013.04.001>
18. Rizzo, M., Marcuzzo, M., Zangari, A., Gasparetto, A., Albarelli, A.: Fruit ripeness classification: A survey **7**, 44–57. <https://doi.org/10.1016/j.aiia.2023.02.004>
19. Saleem, S., Sablatnig, R.: A robust SIFT descriptor for multispectral images **21**(4), 400–403. <https://doi.org/10.1109/LSP.2014.2304073>
20. Su, Z., Zhang, C., Yan, T., Zhu, J., Zeng, Y., Lu, X., Gao, P., Feng, L., He, L., Fan, L.: Application of hyperspectral imaging for maturity and soluble solids content determination of strawberry with deep learning approaches **12**. <https://doi.org/10.3389/fpls.2021.736334>
21. Tiedemann, T., Cordes, F., Keppner, M., Peters, H.: Challenges of Autonomous In-field Fruit Harvesting and Concept of a Robotic Solution:. In: Proceedings of the 19th International Conference on Informatics in Control, Automation and Robotics. pp. 508–515. SCITEPRESS - Science and Technology Publications, Lisbon, Portugal (2022). <https://doi.org/10.5220/0011321300003271>
22. Wang, C., Wang, H., Han, Q., Zhang, Z., Kong, D., Zou, X.: Strawberry detection and ripeness classification using YOLOv8+ model and image processing method **14**(5), 751. <https://doi.org/10.3390/agriculture14050751>
23. Weng, S., Yu, S., Guo, B., Tang, P., Liang, D.: Non-destructive detection of strawberry quality using multi-features of hyperspectral imaging and multivariate methods **20**(11), 3074. <https://doi.org/10.3390/s20113074>
24. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: CBAM: Convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *Computer Vision – ECCV 2018*, vol. 11211, pp. 3–19. Springer International Publishing. https://doi.org/10.1007/978-3-030-01234-2_1
25. Xiong, Y., Ge, Y., Grimstad, L., From, P.J.: An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation **37**(2), 202–224. <https://doi.org/10.1002/rob.21889>
26. Yu, Y., Xie, H., Zhang, K., Wang, Y., Li, Y., Zhou, J., Xu, L.: Design, development, integration, and field evaluation of a ridge-planting strawberry harvesting robot **14**(12), 2126. <https://doi.org/10.3390/agriculture14122126>
27. Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J.: DETRs beat YOLOs on real-time object detection. In: 2024 IEEE/CVF CVPR. pp. 16965–16974. <https://doi.org/10.1109/CVPR52733.2024.01605>
28. Zhou, X., Ampatzidis, Y., Lee, W.S., Zhou, C., Agehara, S., Schueller, J.K.: Deep learning-based postharvest strawberry bruise detection under UV and incandescent light **202**, 107389. <https://doi.org/10.1016/j.compag.2022.107389>
29. Zhou, X., Lee, W.S., Ampatzidis, Y., Chen, Y., Peres, N., Fraisse, C.: Strawberry maturity classification from UAV and near-ground imaging using deep learning **1**, 100001. <https://doi.org/10.1016/j.atech.2021.100001>
30. Zhou, Z., Li, S., Shao, Y.: Crops Classification from Sentinel-2A Multi-spectral Remote Sensing Images Based on Convolutional Neural Networks. In: *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. pp. 5300–5303 (2018). <https://doi.org/10.1109/IGARSS.2018.8518860>