First results of combining RGB segments and multi-spectral pixel classes for strawberry ripeness detection

Philipp Meyer[1], Timo Lange[1], Matthis Trost[1], and Tim Tiedemann[1]

Abstract: Autonomous harvesting of strawberries is dependent on the detection of fruits and their ripeness. For harvesting in commercial fields, at least unripe, ripe, and overripe fruits must be distinguishable. In this paper, we employ visual image-wise segmentation with multi-spectral pixel-wise ripeness classification on real field recordings. Our results show the feasibility and potential of this combination. Furthermore, we identify problematic parts and necessary improvements.

Keywords: strawberry harvesting, autonomous robot, multi-spectral imaging, machine learning

1 Introduction

Autonomous harvesting of fruits with robotic systems needs reliable classification of ripeness in the field [Ri23]. For strawberry harvesting, at least the “ripe” fruit state must be distinguishable from all other states [Xi20]. Our strawberry harvesting robot [Ti22] is subject to stricter requirements for harvesting in the field. “unripe” strawberries must remain on the plant, “ripe” strawberries must be harvested for sale, and “overripe” strawberries must be harvested for disposal. Thus, at least “unripe”, “ripe”, and “overripe” states must be distinguished.

Detection of ripeness depends on sensors (e.g., hyper-spectral imaging [Ga20]) and algorithms (e.g., machine- and deep-learning [Ka24; Pe21]). Images in the visual spectrum (RGB) are common and easy to acquire, while data from further spectrums are rare and more complex to record. Therefore, we propose a combined classification based on RGB images with deep learning object segmentation and multi-spectral (MS) images with a multi-layer perceptron (MLP) neural network for pixel-wise ripeness classification. Offloading strawberry detection to an RGB model reduces the complexity and amount of necessary data for MS ripeness classification. Our recorded MS images include ultraviolet (UV), visual (VIS), near-infrared (NIR), and short-wave infrared (SWIR) spectra. We show the first results of the setup employed in commercial strawberry fields.

In Section 2, we describe the combined algorithm pipeline. In Section 3, we present the first results and challenges. We conclude in Section 4 with an outlook on future directions.

---

[1] HAW Hamburg, Department Informatik, Berliner Tor 7, 20099 Hamburg, philipp.meyer@haw-hamburg.de, timo.lange@haw-hamburg.de, matthis.trost@haw-hamburg.de, tim.tiedemann@haw-hamburg.de

2 Combining RGB segmentation and MS pixel classification

The first step is the recording of images. Our sensor setup consists of four cameras (UV, VIS, NIR, and SWIR). A respective filter wheel expands the NIR and SWIR cameras with additional channels. Therefore, the MS images contain 14 images (UV without filter 190-1100 nm, VIS blue channel, VIS green channel, VIS red channel, NIR without filter 400-1000 nm, NIR filter 740 nm, NIR filter 845 nm, NIR filter 880 nm, NIR filter 940 nm, SWIR without filter 400-1700 nm, SWIR filter 1000 nm, SWIR filter 1300 nm, SWIR filter 1450 nm, SWIR filter 1550 nm). The RGB image for the segmentation consists of the recorded visual blue, green, and red channels. For the field recordings, the sensor setup is placed next to the strawberry mounds and individual shots are taken at various positions at equal distances.

For the second step, segmentation and pixel-wise ripeness classification need their trained model. For the segmentation, we use the Segment Anything 2 model (SAM 2) [Ra24] with a public pre-trained checkpoint (sam2_hiera_large.pt). As pixel ripeness classifiers, we assessed, e.g., SVMs, clustering algorithms, and neural networks. In first tests, SVMs and clustering algorithms performed on a similar level as MLPs, but the training time became excessive with an increasing number of pixels. Therefore, in this work, we use an MLP with four hidden layers (see Tab. 1). Input is one 14-dimensional multi-spectral pixel vector, and the output is one of the four classes “no strawberry”, “unripe”, “ripe”, and “overripe”. We trained the classifier with 91 field recordings. The training data represents the distribution of ripeness classes on the field with additional random pixels of the environment (30% no strawberry, 28% unripe, 34% ripe, and 8% overripe). We sampled data multiple times to balance the classes to the same size before training.

Layer (Type)	**Output Shape**	**Activation**	**No. of Parameters**
Input (Dense)	512	ReLU	7,680
Hidden (Dense)	256	ReLU	131,328
Hidden (Dense)	128	ReLU	32,896
Hidden (Dense)	64	ReLU	8,256
Hidden (Dense)	32	ReLU	2,080
Output (Dense)	4	Softmax	132
	Total number of parameters		182,372

Tab. 1: Structure of the multilayer perceptron that forms the pixel-wise ripeness classifier

In the third step, both models combine their results with a merger module into predictions of segments containing strawberries and their ripeness. Figure 1 shows the implemented detection pipeline that combines segmentation and pixel-wise ripeness classification. (1) After recording, the RGB image is used as input for automatic segmentation with SAM 2 (32 points per side, 0.4 max bounding box overlap, 1,000 pixel minimum / 20,000 pixel maximum segment area). This results in multiple segments representing any shape in the image within a minimum and maximum size. Thus, strawberries are just a subset of those segments. The left example image in Figure 1 depicts all detected segments. (2) The 14 MS images are the input data for the pixel ripeness classifier that uses each 14-dimensional pixel vector as input and predicts the pixel ripeness. The classifier predicts all pixels to generate a new image containing ripeness per pixel. The right example image in Figure 1 shows an example of such a predicted image. (3) The segment ripeness merger combines both previous results and removes all segments where the resulting class is “no strawberry”. Only predicted strawberry segments remain. The bottom image in Figure 1 depicts the result after the merger.

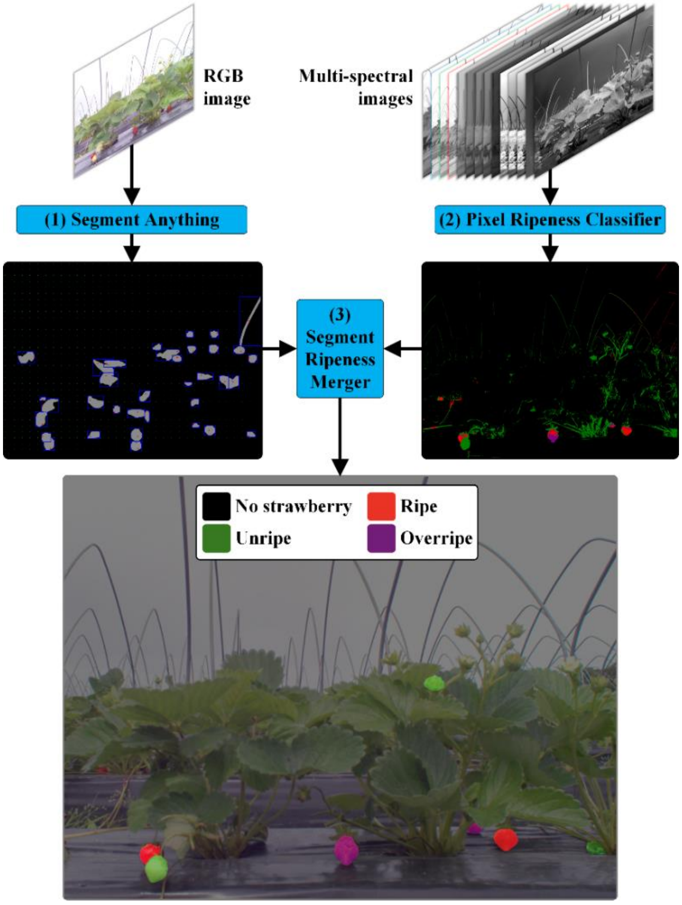

Fig. 1: The segmentation and pixel-wise ripeness classification combination with example images

The implemented merging strategy calculates the number of pixels per ripeness class and sets the class with the most pixels as the class for the whole segment. Refinements may be necessary to account for specific challenges: (A) Strawberries that are >50% ripe but also have a large unripe area. Those should be considered as “unripe”. (B) A small overripe area should directly lead to an “overripe” classification. (C) Individual pixel classifications can be wrong because of, e.g., water, dirt, and lighting. The following section discusses if and how merging strategy refinements could improve detection results.

3 First results

Our first results are based on two further datasets. One is a validation dataset for the pixel ripeness classifier (31 recordings with a total of 320,000 labeled pixels). The second dataset is a test dataset for the combined pipeline, which we recorded in a different field on another day (47 recordings with a total of 334 labeled strawberry segments).

Figure 2 shows the results with the two datasets as confusion matrices. The left matrix is the result of pixel ripeness classification on the validation data. The right matrix is the result of combined segment ripeness prediction of strawberries on the test data.

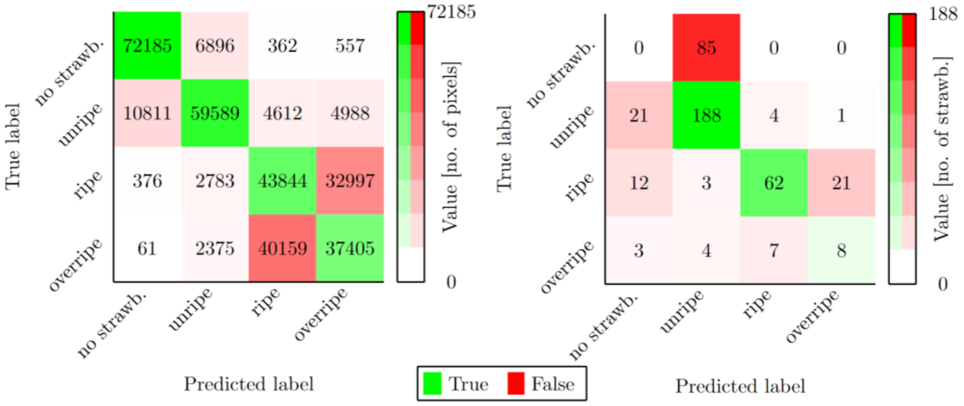

Fig. 2: Confusion matrices of pixel ripeness classification with pixel validation data (left) and combined segment ripeness prediction with strawberry test data (right)

The left confusion matrix (see Fig. 2) of pixel ripeness classification shows the classes “no strawberry” and “unripe” are highly distinguishable from each other and remaining classes. The classifier differentiated “ripe” and “overripe” clearly from “no strawberry” and “unripe” but highly confused them with each other. The main reason could be the low rate of the overripe class in the data (8%). Our data balancing just prevents models from betting on more common classes but is not able to compensate for missing variation of data. Experiments without balanced data have shown zero overripe predictions. Another challenge is that one strawberry in the training data consists of multiple pixels labeled with the same class. For precise learning of overripe areas on otherwise ripe strawberries, the training data labels must also include pixels with different classes on the same strawberry.

The combined segment ripeness results (see right confusion matrix Fig. 2) contain a peculiarity. In the test data, just strawberries are labeled so it cannot be measured how many “no strawberry” classifications are correct identifications as “no strawberry”. Despite that, it can still be measured how many predicted strawberries are “no strawberry” or how many strawberries are predicted as “no strawberries”.

The matrix shows that anytime a segment is falsely identified as a strawberry it is confused with the class “unripe”. This is due to the high similarity between e.g., stems and leaves with unripe strawberries. This could be improved by replacing SAM with a strawberry segmentation model. There is little confusion of “unripe” with the remaining classes.

Again, the number of overripe strawberries in the test data is low (6.6%). The portion of ripe strawberries classified as “overripe” (21.4%) and overripe strawberries classified as “ripe” (31.8%) is in the same order of magnitude and reflects the result of the pixel classifier. Confusion of “ripe” as “overripe” is lower than pixel-wise classification.

In total the pixel ripeness classification performs with an accuracy of 67%, while the combined segment ripeness accuracy is 61%. Another approach for strawberry ripeness classification reaches accuracies above 90% in controlled lab environments [Ka24]. Still, the f1-scores for “unripe” (0.76) and “ripe” (0.73) hint at potential for improvement.

At the current stage, merging refinements would bring hardly any improvement to the results: (A) Classifying strawberries with small unripe areas as unripe could just improve on four cases in the test data (1.9% of “unripe” predicted as “ripe”). (B) Classifying ripe strawberries with overripe areas would just move the error of falsely classified “overripe” as “ripe” to “ripe” as “overripe”. Still, an improved pixel ripeness classification model can benefit from tuned merging strategies. Furthermore, such strategies could be chosen to meet the demands of the respective market strategy or adapt to respective customers.

More training data, particularly with overripe strawberries and multiple classes labeled per strawberry are required to enable major improvements. More data would also enable the training of a segmentation model especially for strawberries.

4 Conclusion and outlook

Autonomous strawberry harvesting on commercial fields needs smart sensor technology that can detect strawberries and their ripeness to leave unripe strawberries on the plant and separate ripe and overripe strawberries.

We showed that the combination of visual segmentation and multispectral pixel-wise ripeness classification is a promising candidate with a high potential for strawberry detection in autonomous harvesting. Our results identify improvements that could lead to significant performance gains.

Future work is needed to increase performance to a level of practical usability: (1) Creation of reliable strawberry segmentation to reduce the number of objects that are falsely classified as strawberries. (2) Optimization of the training data labels to consider areas of different classes on the same strawberry. (3) Increase the size and balancing of the training data. (4) Optimize performance of the pixel ripeness classifier with advanced model architectures. (5) Furthermore, we plan a final demonstrator that is also able to identify malformed strawberries with low demand in commercial reselling.

Acknowledgments

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

Bibliography

[Ga20] Gao, Z. et al.: Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. AI in Agriculture vol. 4, pp. 31-38, 2020.

[Ka24] Karki, S. et al.: Classification of strawberry ripeness stages using machine learning algorithms and colour spaces. Horticulture, Environment, and Biotechnology vol. 65, Springer, Nr. 2, pp. 337-354, 2024.

[Pe21] Perez-Borrero, I. et al.: A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network. Neural computing and applications vol. 33, Springer, Nr. 22, pp. 15059-15071, 2021.

[Ra24] Ravi, N. et al.: SAM 2: Segment Anything in Images and Videos. In: arXiv preprint arXiv:2408.00714, 2024.

[Ri23] Rizzo, M. et al.: Fruit ripeness classification: A survey. Artificial Intelligence in Agriculture vol. 7, pp. 44-57, 2023.

[Ti22] Tiedemann., T. et al.: Challenges of Autonomous In-field Fruit Harvesting and Concept of a Robotic Solution. In: Proceedings of the 19th Int. Conference on Informatics in Control, Automation and Robotics – ICINCO, SciTePress, pp. 508-515, 2022.

[Xi20] Xiong, Y. et al.: An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation. Journal of Field Robotics vol. 37, Wiley Online Library, Nr. 2, pp. 202-224, 2020.