Weakly Supervised Panoptic Segmentation for Defect-Based Grading of Fresh Produce

Supplementary Materials

A. Dataset visualizations

Figure S.1. **Defect examples.** Example images of banana surface defects we aim to detect in this study. Bruises usually result from a dull impact, while scars are caused by a sharp impact. Old defects are usually darker than new defects as they oxidize over time.

B. Implementation details

All images were resized and padded to 1024^2 pixels resolution. Training images were augmented with a 50% Random Horizontal Flip and with additional random color-based augmentations (see Table S.1) to increase model robustness.

Table S.2 shows the hyperparameters used to train the Maskformer models. We evaluated the model every five epochs on the validation set and saved the model with the highest Panoptic Quality. All models were trained on a single NVIDIA RTX A5000 GPU with 24GB of memory.

Table S.1. **Random color augmentations applied to training images.** Values are uniformly sampled from the specified ranges.

Property	Sampling range
brightness	[0.9, 1.1]
contrast	[0.9, 1.1]
saturation	[0.9, 1.1]
hue	[-0.05, 0.05]

Table S.2. **Hyperparameters for Maskformer training.**

Batch size	2
Epochs	100
Evaluation frequency	every 5 epochs
Optimizer	Adam [12]
Learning rate	5×10^{-5}
Learning rate schedule	constant

C. Additional SAM evaluations

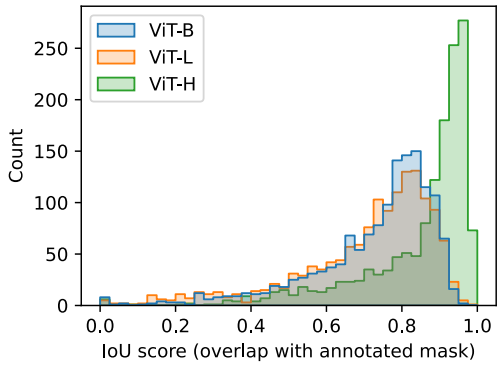

Figure S.2. **Distribution of IoU values when comparing annotated and SAM-generated defect masks.** While the ViT-B and ViT-L variants show a similar distribution, using the largest model size (ViT-H) leads to significantly higher overlap with hand-annotated masks.

D. Results using four defect categories

In this section, we describe our results when using four different defect classes instead of a single joint category. Our results clearly show that while the detection and segmentation of bruises and scars work well, a reliable categorization into one of the four predefined classes is not possible with the current approach and dataset. We hypothesize that this limitation stems from one or more of the following factors:

1. Ambiguous annotation: The classification of defects across the four predefined categories may be subjective to a large degree. The distinction between "old" and "new" defects is not always clear-cut, as the transition is gradual. Additionally, distinguishing bruises from scars can be ambiguous, especially for non-experts (see Figure S.1).
2. Image resolution: We use 1024^2 pixel resolution for images. While this is generally considered high for machine learning tasks, defects can be small and, thus, only be represented by a few pixels, making them harder to categorize.
3. The four defect types are unevenly represented in our dataset (37/182/387/834). A more balanced distribution of categories and a larger number of defect samples are likely to improve categorization accuracy. We recommend collecting more examples from the two underrepresented classes for future work.

Table S.3. **Results using multiple defect categories.** It is evident that our models are unable to categorize defect types. Most likely due to ambiguous annotations and/or limited training data. The configuration highlighted in blue is used for the visualizations in Figure S.3.

		Defects												
		Defect masks		Old Bruise		New Bruise		Old Scar		New Scar		Overall		
Model	PP	train	val	AP	IoU	AP	IoU	AP	IoU	AP	IoU	mAP	mIoU	PQ
Maskformer		Anno.	Anno.	.031 ± .059	.028 ± .043	.043 ± .024	.157 ± .037	.036 ± .018	.225 ± .097	.050 ± .018	.316 ± .054	.040 ± .019	.482 ± .024	.471 ± .023
Maskformer	✓	Anno.	Anno.	.030 ± .059	.013 ± .018	.088 ± .039	.114 ± .035	.034 ± .016	.223 ± .083	.017 ± .009	.066 ± .021	.042 ± .022	.436 ± .018	.431 ± .015
Maskformer		SAM-L	Anno.	.039 ± .055	.060 ± .080	.034 ± .016	.175 ± .074	.045 ± .034	.179 ± .022	.066 ± .018	.345 ± .046	.046 ± .012	.487 ± .018	.477 ± .013
Maskformer	✓	SAM-L	Anno.	.032 ± .049	.031 ± .038	.074 ± .017	.124 ± .051	.038 ± .021	.161 ± .009	.032 ± .014	.104 ± .039	.044 ± .015	.437 ± .018	.435 ± .019
Maskformer		SAM-L	SAM-L	.037 ± .053	.063 ± .081	.036 ± .016	.178 ± .075	.045 ± .033	.182 ± .023	.071 ± .019	.353 ± .049	.047 ± .012	.489 ± .018	.478 ± .013
Maskformer	✓	SAM-L	SAM-L	.032 ± .049	.027 ± .034	.080 ± .018	.127 ± .051	.036 ± .020	.163 ± .010	.034 ± .015	.108 ± .041	.046 ± .015	.437 ± .018	.436 ± .019
Maskformer		SAM-H	Anno.	.019 ± .026	.074 ± .078	.043 ± .020	.157 ± .092	.031 ± .021	.252 ± .063	.056 ± .016	.318 ± .047	.037 ± .014	.495 ± .010	.474 ± .010
Maskformer	✓	SAM-H	Anno.	.032 ± .056	.030 ± .029	.093 ± .022	.145 ± .074	.025 ± .029	.230 ± .041	.029 ± .011	.100 ± .050	.045 ± .016	.453 ± .013	.449 ± .009
Maskformer		SAM-H	SAM-H	.025 ± .035	.092 ± .105	.049 ± .022	.163 ± .091	.037 ± .024	.264 ± .064	.062 ± .018	.328 ± .047	.043 ± .015	.501 ± .010	.482 ± .011
Maskformer	✓	SAM-H	SAM-H	.045 ± .078	.029 ± .027	.106 ± .022	.149 ± .073	.027 ± .028	.239 ± .041	.034 ± .010	.104 ± .051	.053 ± .019	.455 ± .013	.456 ± .010

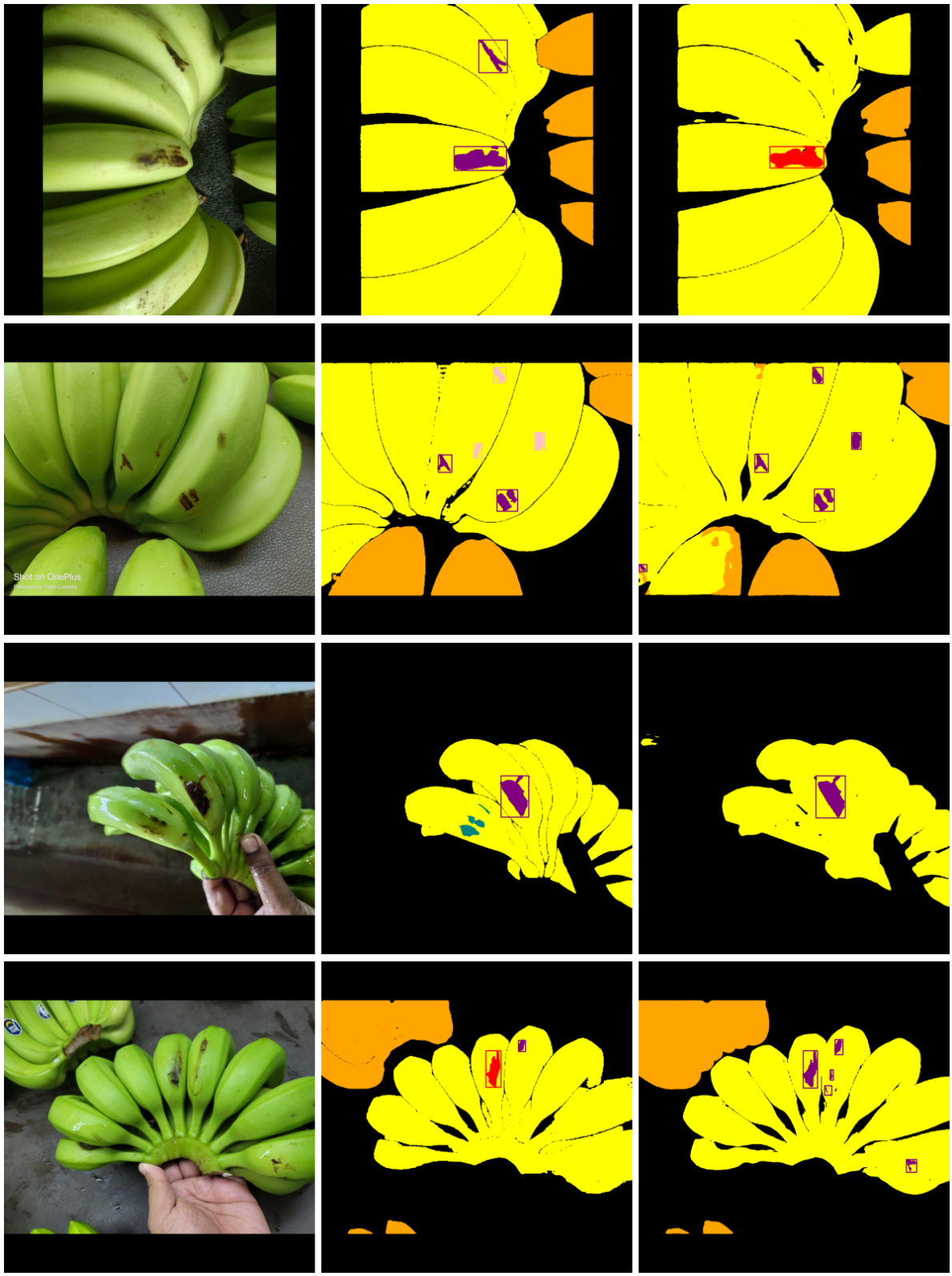

Figure S.3. **Example visualizations of annotated vs. predicted masks using four defect categories.** Left: Input Image, Mid: Annotation, Right: Maskformer Prediction. Segments are color-coded as follows: Foreground Banana, Background Banana, Old Bruise, Old Scar, New Bruise, New Scar.