SemanticSugarBeets: A Multi-Task Framework and Dataset for Inspecting Harvest and Storage Characteristics of Sugar Beets (Supplementary Material)

Gerardus Croonen Andreas Trondl Julia Simon Daniel Steininger
AIT Austrian Institute of Technology
Center for Vision, Automation & Control
{gerardus.croonen,andreas.trondl.fl,julia.simon,daniel.steininger}@ait.ac.at

This supplementary complements the main paper with a more detailed description of the image acquisition protocol, extended dataset statistics and additional discussion of results.

A. Image acquisition protocol

For the purpose of repeatability and consistency, we performed the following steps during image acquisition:

1. Compose a group of beets (3 to 5) to fit inside the camera frame, held in landscape mode
2. Put a folding ruler (or other object of known size) in the frame, ensuring its full visibility
3. From a standing position, take two (almost identical) photographs from a top-view perspective
4. Flip the beets and put them back in roughly the same position
5. Force a camera refocus by taking a photograph of a nearby object, such as your hand. This photo will also allow for the quick identification of separate beet groups and beet sides when viewing and meta-annotating the photos.
6. Repeat steps 2-3.

B. Extended dataset analysis

Tab. 1 provides a complete list of recording sessions and corresponding statistics and meta-parameters. The distribution of bounding box centers across all beet instances is depicted in Fig. 1. Representative examples for both classes of annotated reference markers are visualized in Fig. 2.

C. Extended methodology

As the original annotations usually contain multiple polygons of different fine-grained classes for a single sugar-beet, a method for automatic pre-processing is required to extract the final annotations compatible with instance segmentation. This process is visualized in Fig. 3. It consists of identifying the largest component of each beet and then extending it with all overlapping smaller components to derive annotations compatible with the coarse-grained instance-segmentation task described in Sec. 4.1 of the main paper.

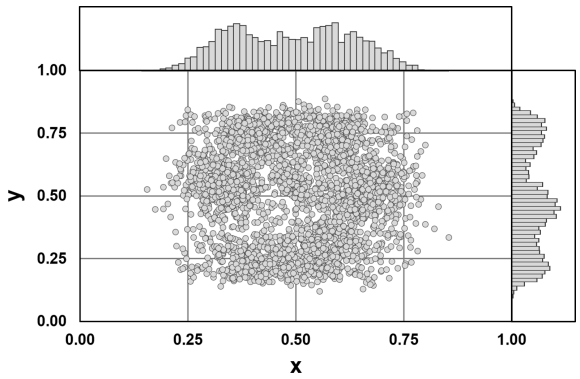

Figure 1. Distribution of normalized bounding-box centers of all annotated sugar-beet instances.

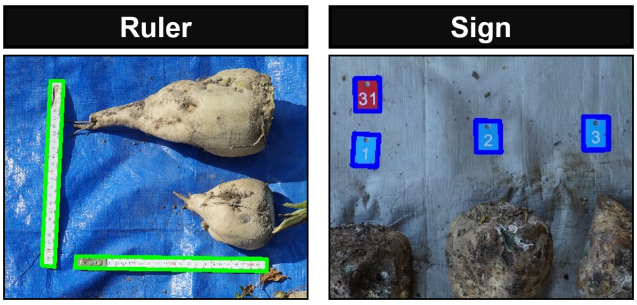

Figure 2. Representative examples of annotated reference objects.

D. Extended evaluation results

Tab. 2 provides the full evaluation details of our semantic-segmentation ablation study, a summary of which is provided in Fig. 5 of the main paper. Regarding the influence

SID	Images	Beets	B/I	Lighting	Moisture	Marker	Stage	Loc
0	33	165	5.0	Sunny	Dry	None	Sample	A
1	92	300	3.3	Sunny	Dry	Ruler	Sample	A
2	40	120	3.0	Diffuse	Dry	Ruler	Sample	A
3	40	120	3.0	Sunny	Dry	Ruler	Sample	A
4	4	12	3.0	Sunny	Dry	Ruler	Sample	A
5	31	93	3.0	Sunny	Wet	Ruler	Harvest	B/C
6	288	864	3.0	Diffuse	Wet	Ruler	Harvest	C
7	282	846	3.0	Sunny	Dry	Ruler	Harvest	D
8	116	319	2.8	Diffuse	Wet	Sign	Storage	E
9	28	83	3.0	Artificial	Wet	Sign	Storage	E
	954	**2922**	**3.1**					

Table 1. Parameters of recording sessions, including **S**ession **ID**, numbers of annotated **Images** and **Beets**, average ratios of **B**eets per **I**mage, **Lighting** conditions, beet **Moisture**, the presence of folding-ruler elements or plastic signs as **Marker** devices, processing **Stages** and cultivation **Loc**ations.

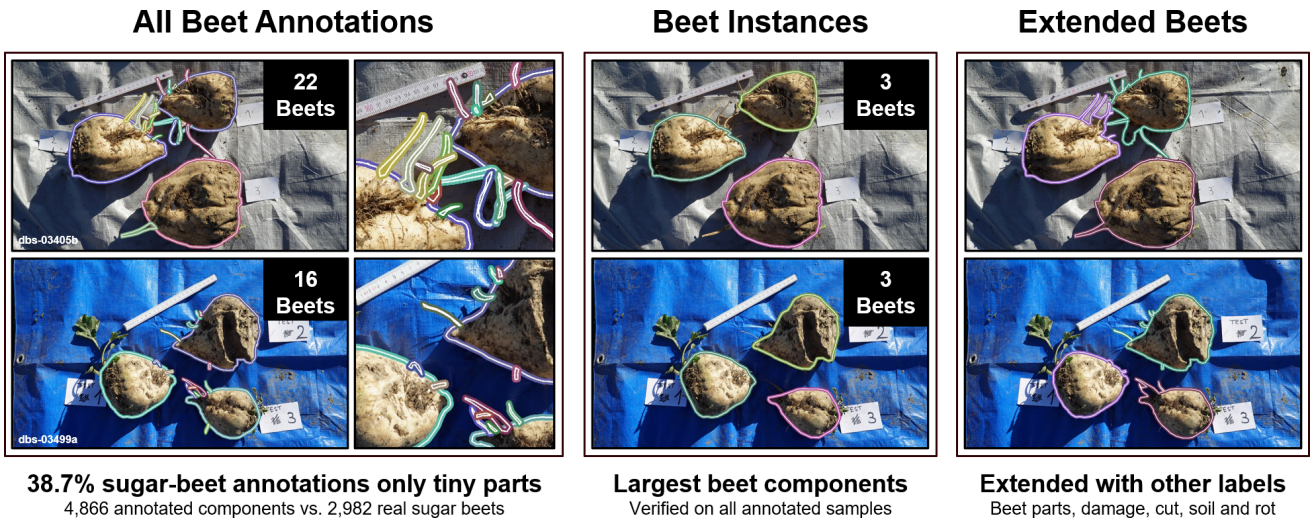

Figure 3. Annotation-synthesis pipeline to convert original annotations to instance-segmentation annotations of entire sugar beets.

of meta-parameters on performance, the exact values forming the basis for Fig. 7 of the paper is summarized in Tab. 3.

Architecture	Encoder	512	768	1024	Mean	t_{512}	t_{768}	t_{1024}
	EfficientNet	61.2	62.1	65.4		9.5	9.9	12.1
	MIT	64.0	64.5	65.6		**6.2**	**7.8**	13.9
MANet	MobileOne	59.5	61.2	61.3	62.8	14.4	16.6	19.6
	MobileNetV3l	60.6	64.4	60.9		6.5	8.3	**10.6**
	RegNetY	60.1	65.4	66.1		7.9	9.2	**10.6**
	EfficientNet	62.5	63.0	63.5		2.6	3.4	5.3
	MIT	63.1	63.9	63.9		4.5	5.3	9.6
PSPNet	MobileOne	63.4	64.7	64.6	62.2	6.1	6.5	8.2
	MobileNetV3l	61.6	62.5	62.5		2.3	2.6	3.3
	RegNetY	58.1	58.0	57.8		**1.7**	**1.9**	**2.3**
	EfficientNet	**66.7**	**68.3**	**68.6**		8.1	10.2	14.0
	MIT	65.6	66.7	66.3		**7.0**	8.7	15.5
U-Net	MobileOne	65.8	66.7	67.6	**67.0**	15.3	16.1	19.5
	MobileNetV3l	66.0	67.9	67.9		7.1	**8.3**	**11.4**
	RegNetY	65.9	66.9	67.5		**7.0**	8.9	12.6
	Mean	**62.9**	**64.4**	**64.6**				

Table 2. Results of semantic segmentation on the test set including all combinations of architectures, encoders and image sizes. Performance numbers are mIoUs, reported in percentages. The rightmost three columns show the inference time for each image size in ms.

		mIoU
Lighting	Sunny	**68.6**
	Diffuse	68.1
	Artificial	64.8
Moisture	Dry	**70.4**
	Wet	66.0
Stage	Sample	78.4
	Harvest	65.2
	Storage	**66.4**

Table 3. Test set performance of semantic segmentation, separated by meta-parameter.