Supplementary Material
GMT: Guided Mask Transformer for Leaf Instance Segmentation

Feng Chen[1]   Sotirios A. Tsaftaris[1]   Mario Valerio Giuffrida[2]

[1]IDCOM, School of Engineering, University of Edinburgh

[2]School of Computer Science, University of Nottingham

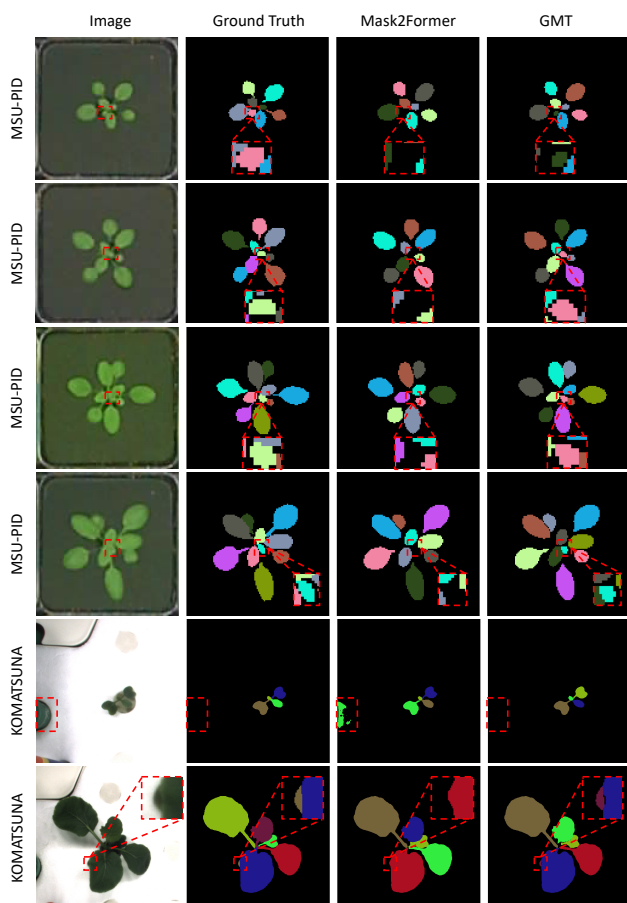feng.chen@ed.ac.uk, s.tsaftaris@ed.ac.uk, valerio.giuffrida@nottingham.ac.uk

Figure S1. Additional visual results on the MSU-PID [2] and KO-MATSUNA [3] test sets. Key differences between the predictions of Mask2Former [1] and our method (GMT) is highlighted in red boxes.

Figure S1 presents further qualitative comparisons on the MSU-PID [2] and KOMATSUNA [3] test sets, highlighting the differences between the predictions of Mask2Former [1] and our GMT model. In Figure S1, rows 1 to 4 show that Mask2Former misses some leaves, while GMT successfully captures and segments them. Row 5 presents that GMT correctly distinguishes leaves from other green objects. Row 6 highlights GMT's capability in segmenting overlapping leaves, even when they are small.

These findings are consistent as those in the main paper, demonstrating that GMT is able to handle challenging scenarios in plant image analysis, ultimately benefiting plant phenotyping and a variety of agricultural applications.

References

[1] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022. 1

[2] Jeffrey A Cruz, Xi Yin, Xiaoming Liu, Saif M Imran, Daniel D Morris, David M Kramer, and Jin Chen. Multi-modality imagery database for plant phenotyping. *Machine Vision and Applications*, 27:735–749, 2016. 1

[3] Hideaki Uchiyama, Shunsuke Sakurai, Masashi Mishima, Daisaku Arita, Takashi Okayasu, Atsushi Shimada, and Rin-ichiro Taniguchi. An easy-to-setup 3d phenotyping platform for komatsuna dataset. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 2038–2045, 2017. 1

1. Additional Visual Results

We provide additional qualitative results and discussion to complement Section 4.3 and Figure 5 in the main paper.