

Ground Truth

LeafOnlySAM

Ours

Ground Truth

LeafOnlySAM

Ours

Figure S1. **Comparison of ZeroPlantSeg and LeafOnlySAM.** Since LeafOnlySAM does not involve plant segmentation, we only show the leaf instance segmentation results on the PhenoBench and GrowliFlower datasets. The left three columns show the results on the GrowliFlower dataset, and the right three columns show the results on the PhenoBench dataset.

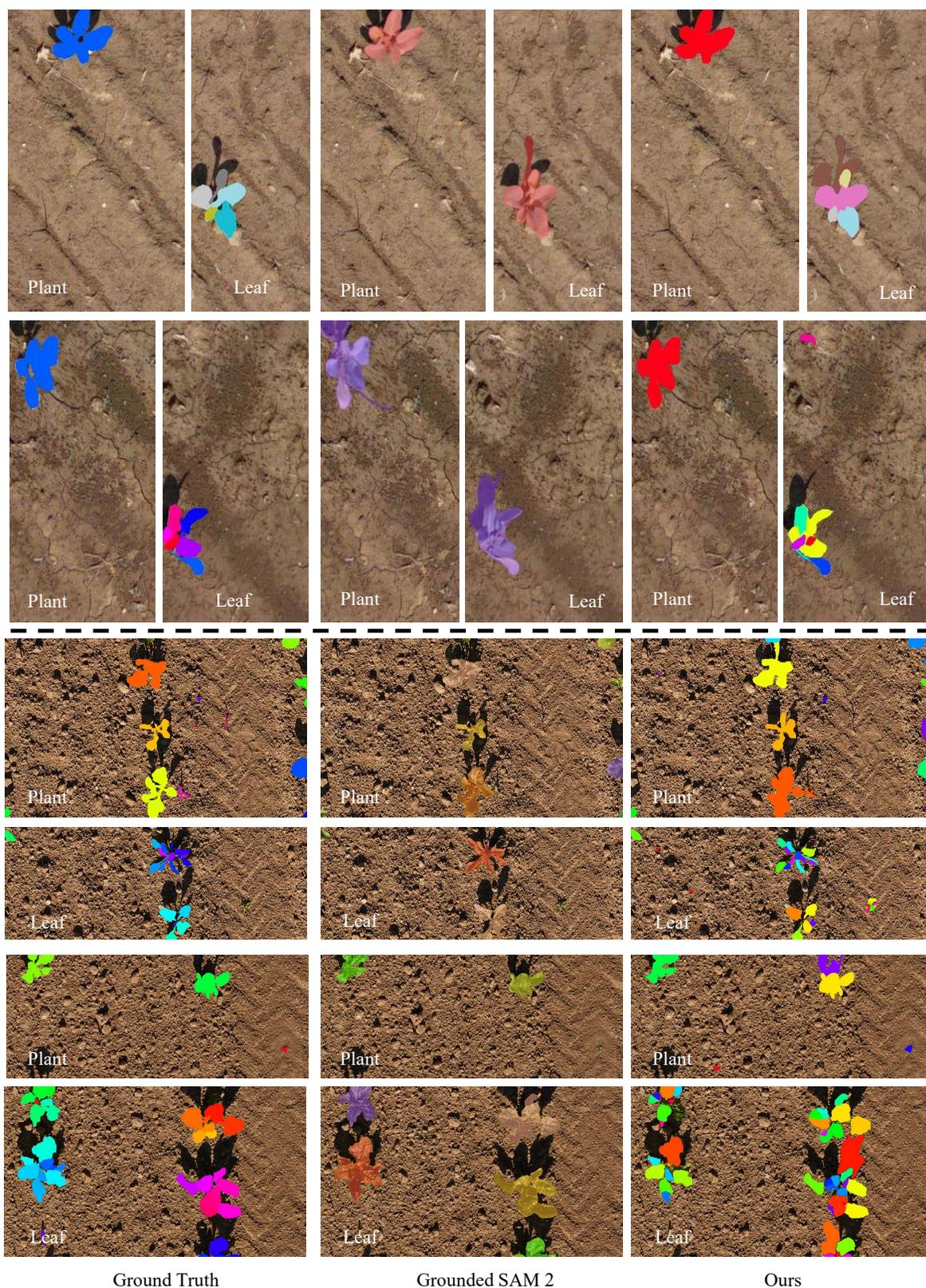
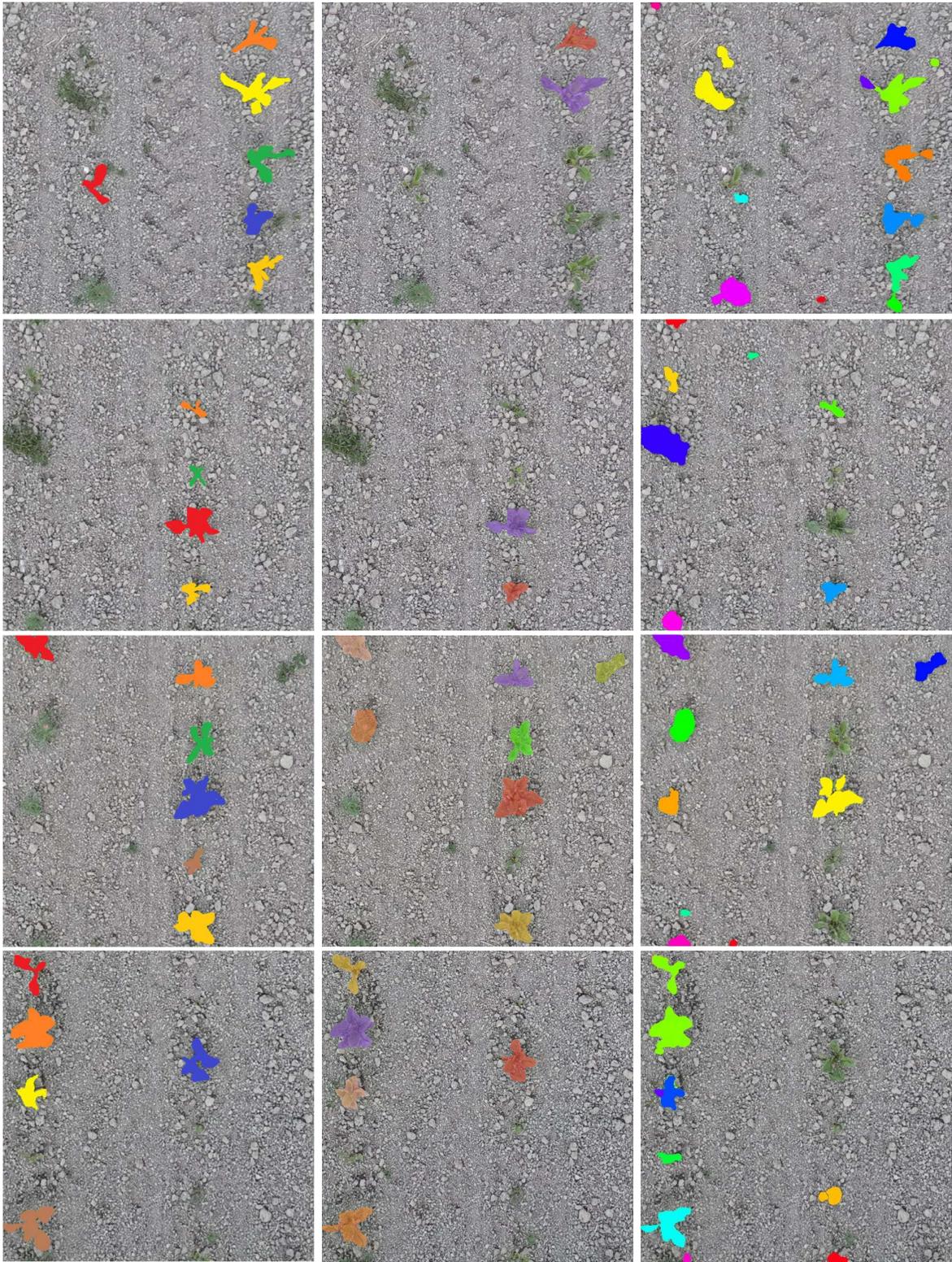


Figure S2. **Comparison of ZeroPlantSeg and Grounded SAM 2.** The results are divided into two parts. The top rows show the comparison on the GrowliFlower dataset, and the bottom rows show the comparison on the PhenoBench dataset. Each picture is denoted by “Plant” or “Leaf” to indicate whether it is a leaf instance segmentation or a plant individual segmentation result.



Ground Truth

Grounded SAM 2

Ours

Figure S3. **Comparison of ZeroPlantSeg and Grounded SAM 2 on the SB20 dataset.** Since SB20 only supplies annotations of individual plants, we only visualize the plant segmentation results.



Figure S4. **Failure cases of ZeroPlantSeg for leaf segmentation.** The top row displays the incorrect split. The bottom row shows an example of failure to recognize leaves when they are occluded.

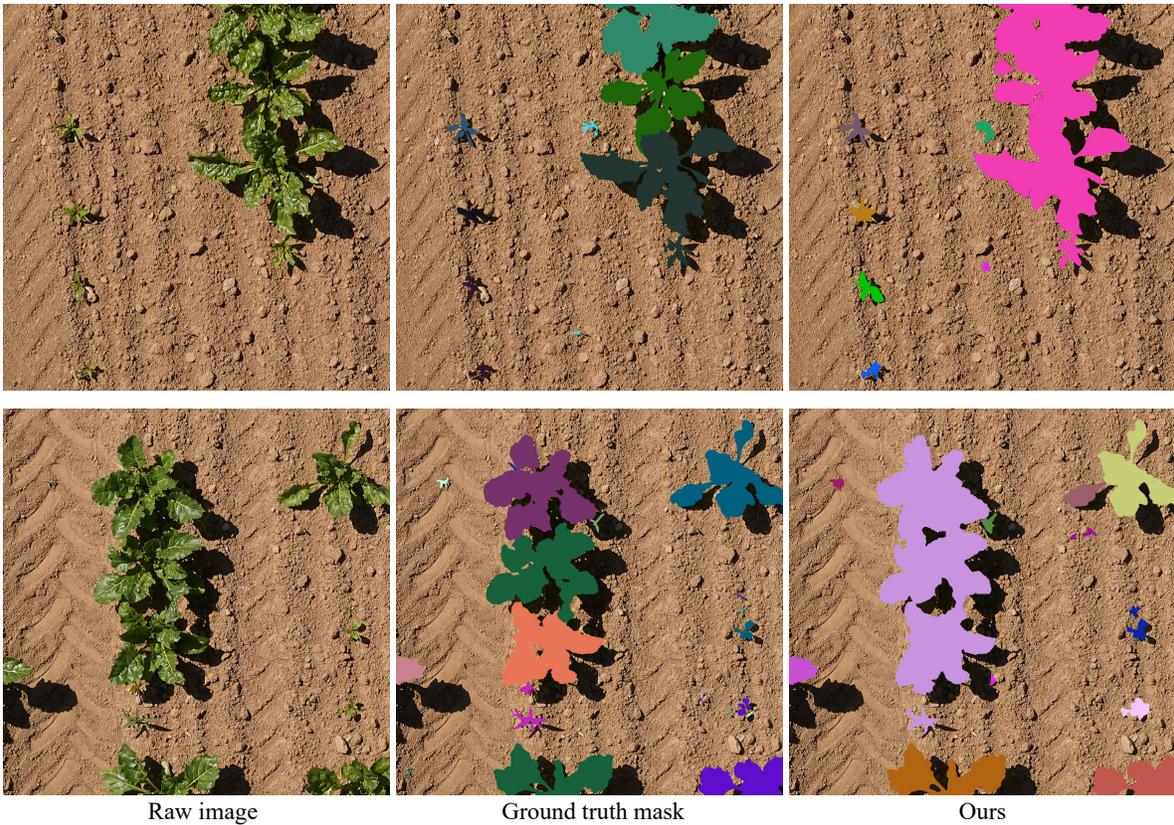


Figure S5. **Failure cases of ZeroPlantSeg for plant segmentation.** These show the situation where leaves from different plants are mistakenly seen as a whole.

B. Additional Results

While our method quantitatively surpasses these existing methods, Figs. S1, S2, and S3 show the visual results for the recent zero-shot models, namely, Grounded SAM2 [3] and LeafOnlySAM [4], compared to ours. Similar to the trend observed in the quantitative evaluation, our method more accurately estimates individual instances of leaves and plants than the state-of-the-art zero-shot segmentation models.

On the other hand, our method sometimes fails to recognize leaves where they are occluded by other leaves, or mistakenly splits one leaf into different parts (Fig. S4). This could be caused by the sliding windows strategy, as it would split one leaf into different windows and see them as individual leaves. In the meantime, it might also lead to incorrect plant segmentation by mixing leaves as if they were from the same plant. This could be caused by aggressive thresholds when executing Greedy DBSCAN (Fig. S5).

References

- [1] A Stewart Fotheringham, Chris Brunsdon, and ME Charlton. Geographically weighted regression. *The Sage Handbook of Spatial Analysis*, 1:243–254, 2009. 1
- [2] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, and Lei Zhang. Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 38–55, 2024. 1
- [3] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. SAM 2: Segment anything in images and videos. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2025. 6
- [4] Dominic Williams, Fraser Macfarlane, and Avril Britten. Leaf only SAM: A segment anything pipeline for zero-shot automated leaf segmentation. *Smart Agricultural Technology*, 8:100515:1–100515:6, 2024. 6