

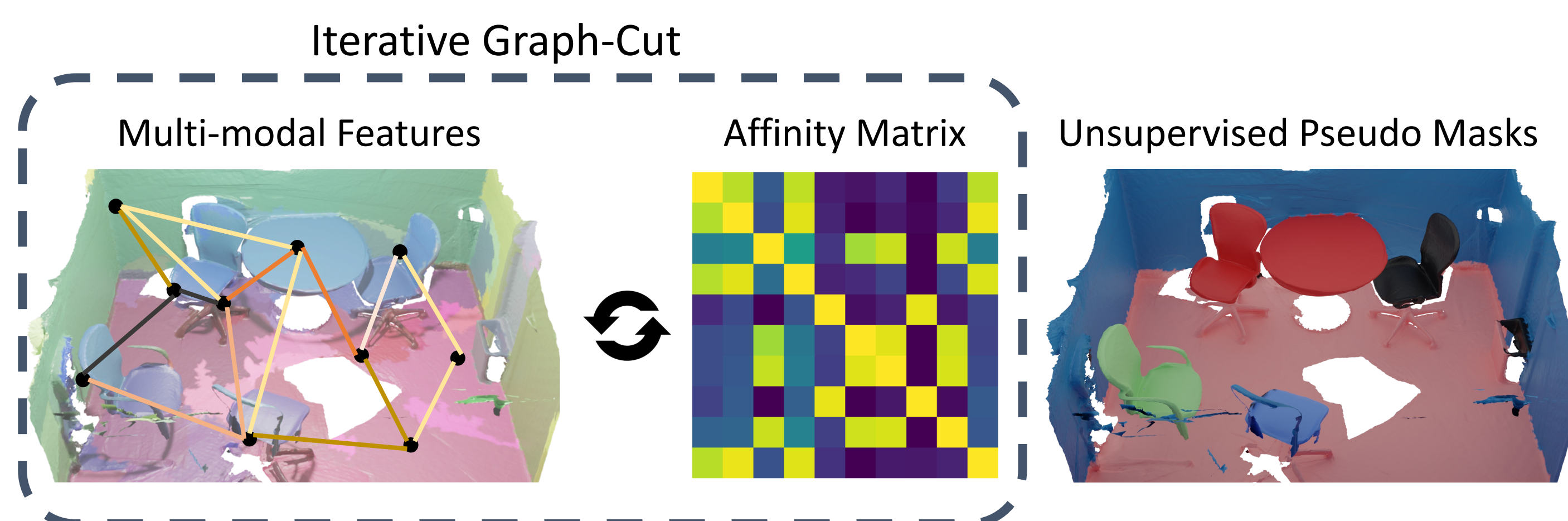
Unsupervised Instance Segmentation



With UnScene3D we propose:

- The first fully unsupervised 3D learning approach for class-agnostic 3D instance segmentation.
- For pseudo ground truth we leverage self-supervised color/geometry features aggregated on geometric segments, enabling efficient representation and learning on high-resolution 3D data.
- We refine the coarse proposals through a series of self-training iterations from the model's own predictions.

Pseudo GT Generation

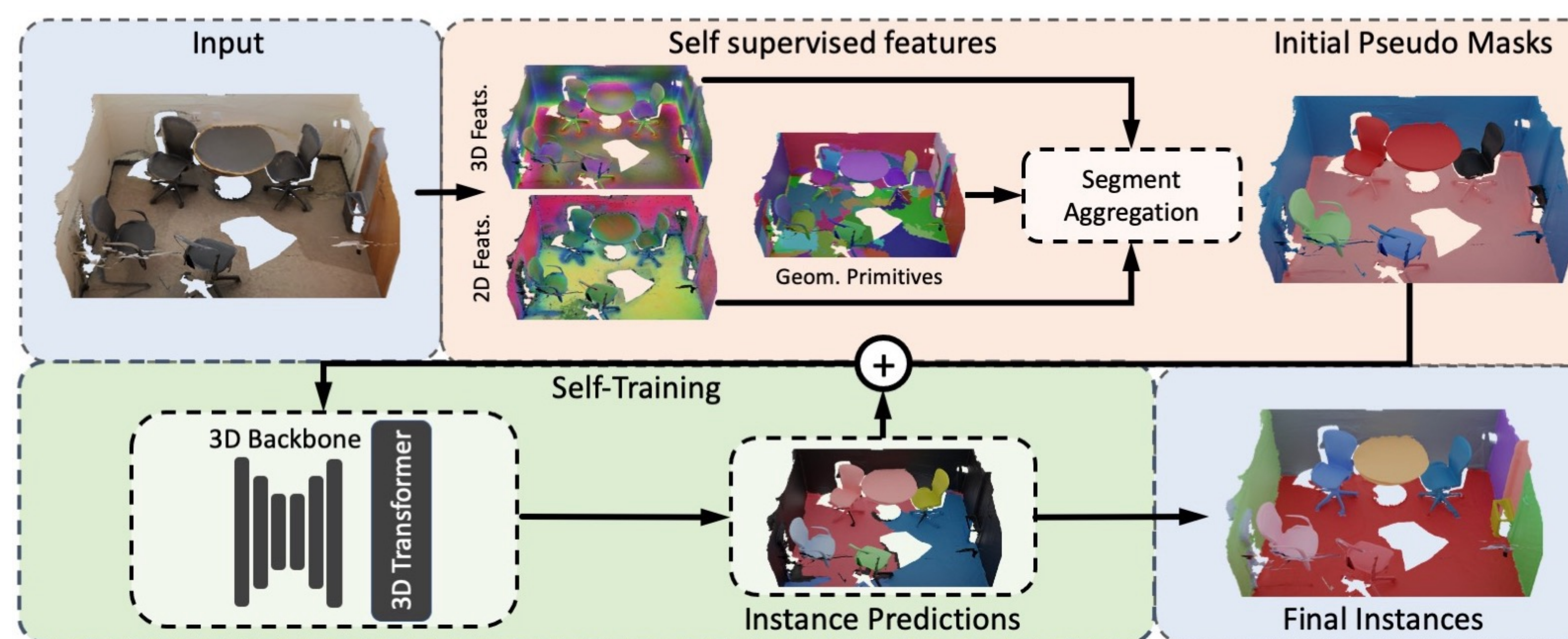


We utilize self-supervised features from the 2D/3D domains, and aggregate these on a geometric oversegmentation of the scenes. This enables us to efficiently represent both low- and high-level properties of the scene. From here, we iteratively separate foreground-background partitions using the Normalized Cut algorithm. For every iteration we mask out the already predicted nodes and continue the process until there are no more segments left in the scene.



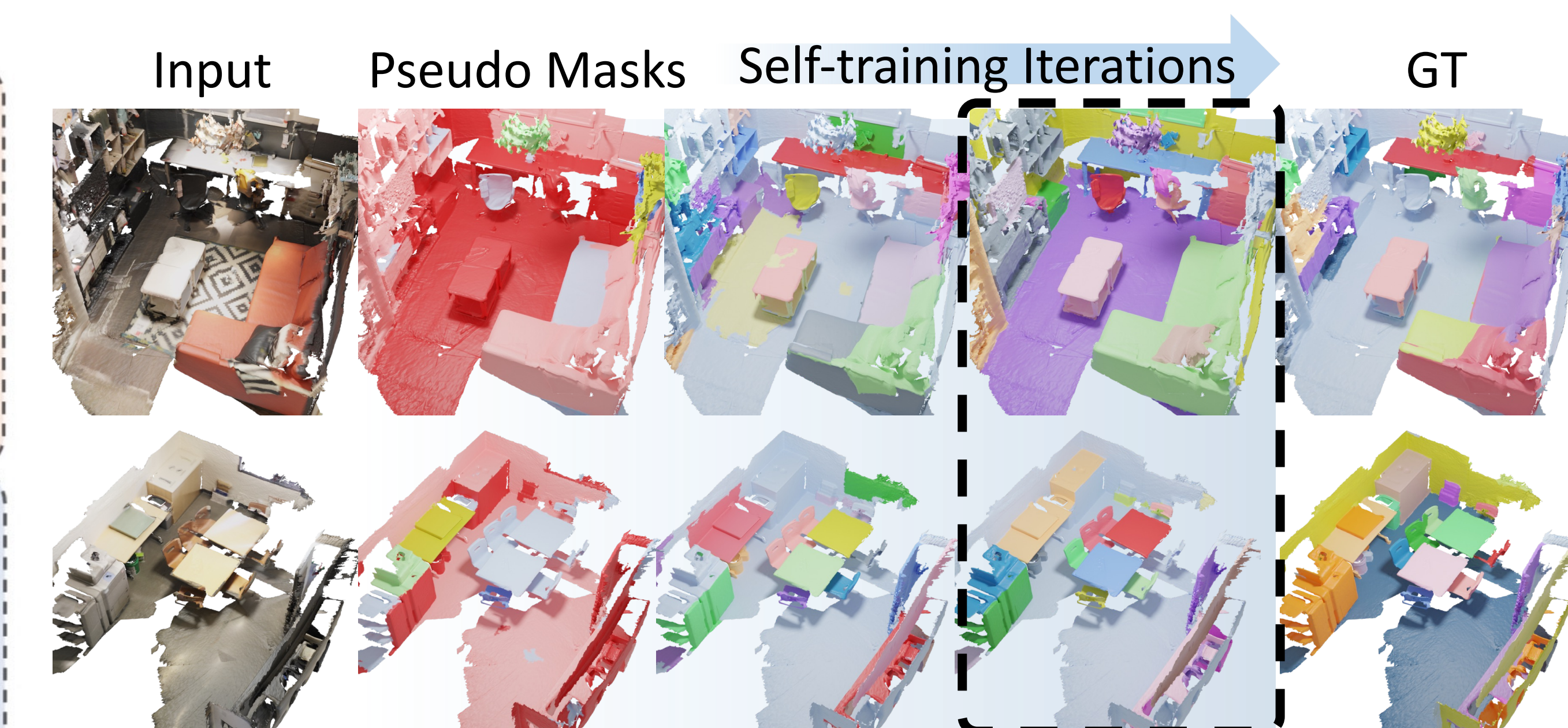
Check out our website for more results, full paper & code!

Self-training Approach



While pseudo masks provide a good set of initial proposals, we further refine them by training our model on its predictions. This self-training process can effectively densify and clean the originally sparse pseudo masks, resulting in a more accurate instance segmentation.

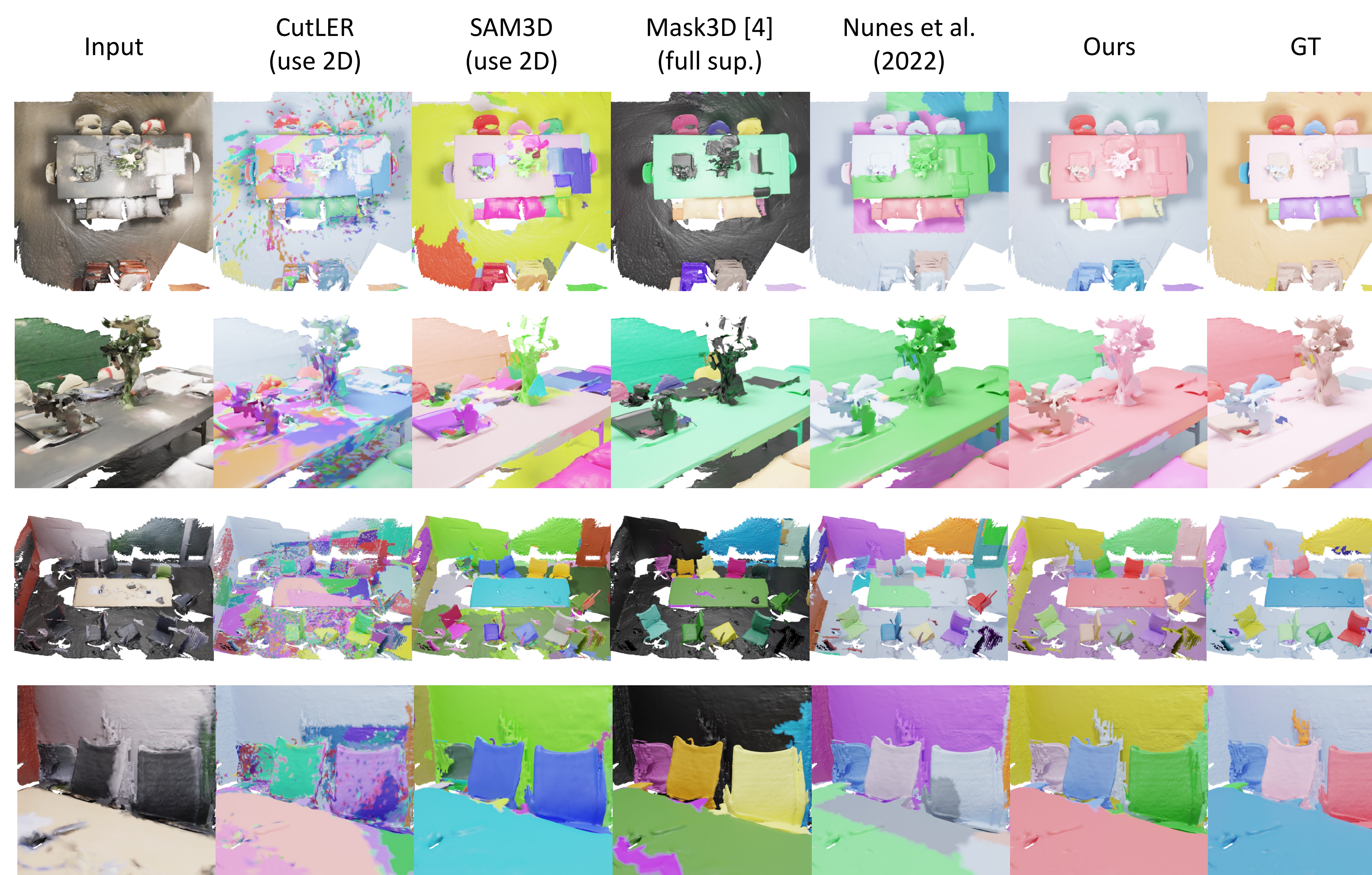
The Effect of Self-Training



After every training iteration we generate a new set of predictions and combine them with the initial mask to create a new, enriched version of pseudo training masks. We show that the initial sparse set of masks get better with every self-training iteration.

Unsupervised Instance Segmentation Results

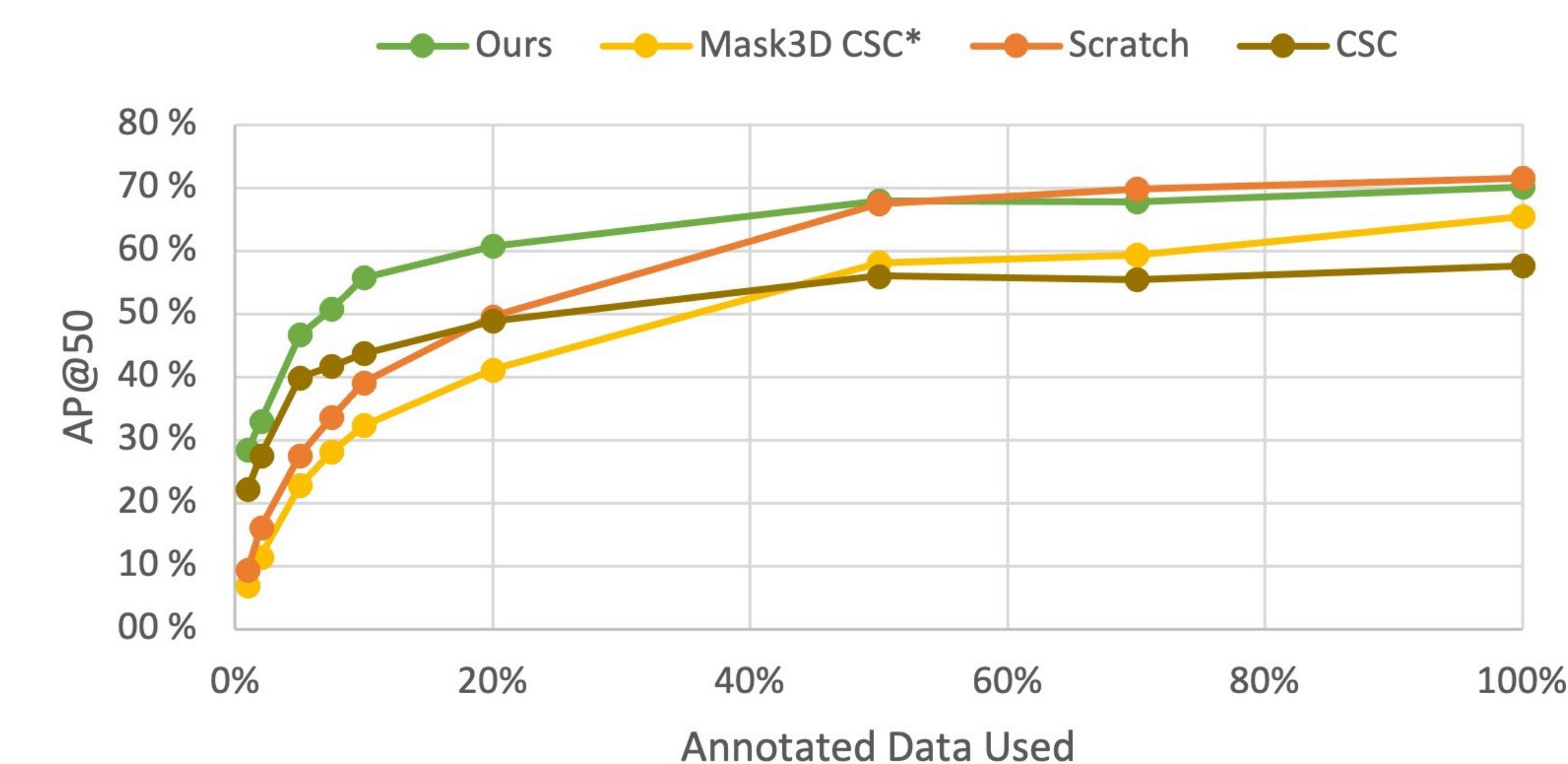
	Images	AP@25	AP@50	mAP
Nunes et al. [1]	X	30.5	7.3	2.3
CutLER (2D) [2]	✓	7.0	0.2	0.3
SAM3D [3]	✓	19.9	10	5.9
Ours	X	58.5	32.2	15.9



UnScene3D improves over unsupervised 3D instance segmentation methods by more than 300% Average Precision score, demonstrating effective instance segmentation in challenging, cluttered 3D scenes.

UnScene3D as Pretraining

ScanNet Limited Data



UnScene3D can learn powerful class-agnostic object properties, which can be used for downstream tasks such as data-efficient dense 3D instance segmentation. Our approach improves over the previous state-of-the-art self-supervised pretraining methods for limited data scenarios.

[1] Nunes, Lucas, et al. "Unsupervised class-agnostic instance segmentation of 3d lidar data for autonomous vehicles." *IEEE Robotics and Automation Letters* 7.4 (2022): 8713-8720.
 [2] Wang, Xudong, et al. "Cut and learn for unsupervised object detection and instance segmentation." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023.
 [3] Yang, Yunhan, et al. "Sam3d: Segment anything in 3d scenes." *arXiv preprint arXiv:2306.03908* (2023).
 [4] Schult, Jonas, et al. "Mask3d: Mask transformer for 3d semantic instance segmentation." *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023.