

Method Details

Geometric Oversegmentation

Graph clustering based on colors and normals as proposed by Felzenszwalb and Huttenlocher¹.

➔ Coarsen graph for more efficient processing.

[1] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004

Feature Extraction

2D Features:

Extracted using pixel-wise vision-language model¹ and projected to 3D.

3D Features:

Extracted using 3D contrastive learning model².

➔ Point-wise semantic and geometric features.

[1] B. Li, K. Q. Weinberger, S. Belongie, V. Koltun, and R. Ranftl, "Language-driven Semantic Segmentation," arXiv, Apr. 02, 2022

[2] J. Hou, B. Graham, M. Nießner, and S. Xie, "Exploring Data-Efficient 3D Scene Understanding with Contrastive Scene Contexts," arXiv, Jun. 25, 2021

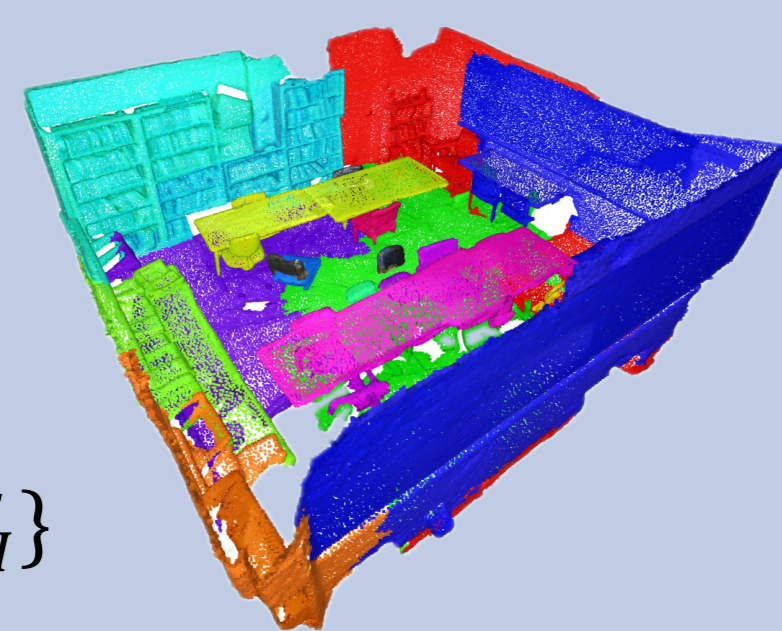
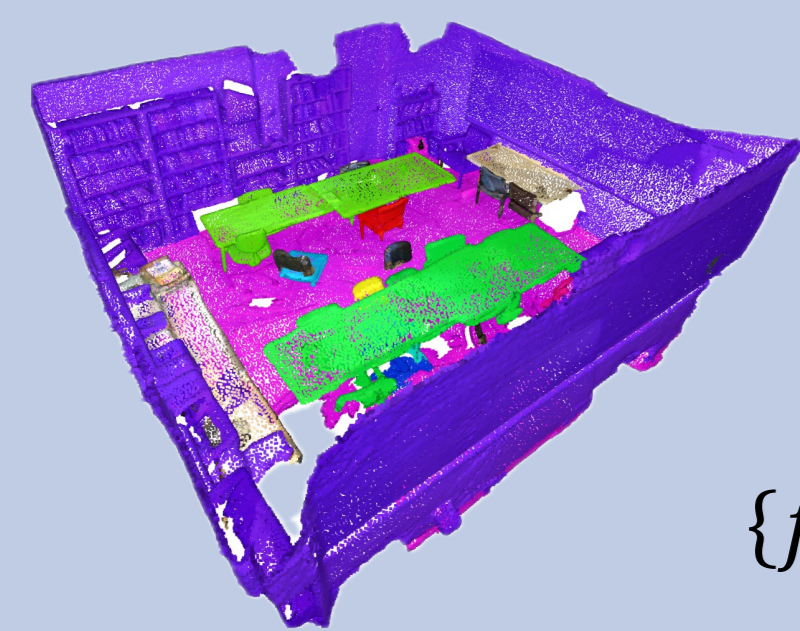
Graph Cut

Create initial pseudo segmentations using normalized cut and pseudo features by aggregating over instances.

➔ Initial supervision for self-training.

Iterative Normalized Cut¹

Spectral Clustering²



$\{f_1, \dots, f_I\}$

Method	mAP ²⁵	mAP ⁵⁰	mAP
Iterative*	16.3	9.4	5.7
Spectral Clustering**	17.0	6.2	3.3

[1] D. Rozenberszki, O. Litany, and A. Dai, "UnScene3D: Unsupervised 3D Instance Segmentation for Indoor Scenes," arXiv, Mar. 25, 2023

[2] U. Von Luxburg, "A tutorial on spectral clustering," *Stat Comput*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

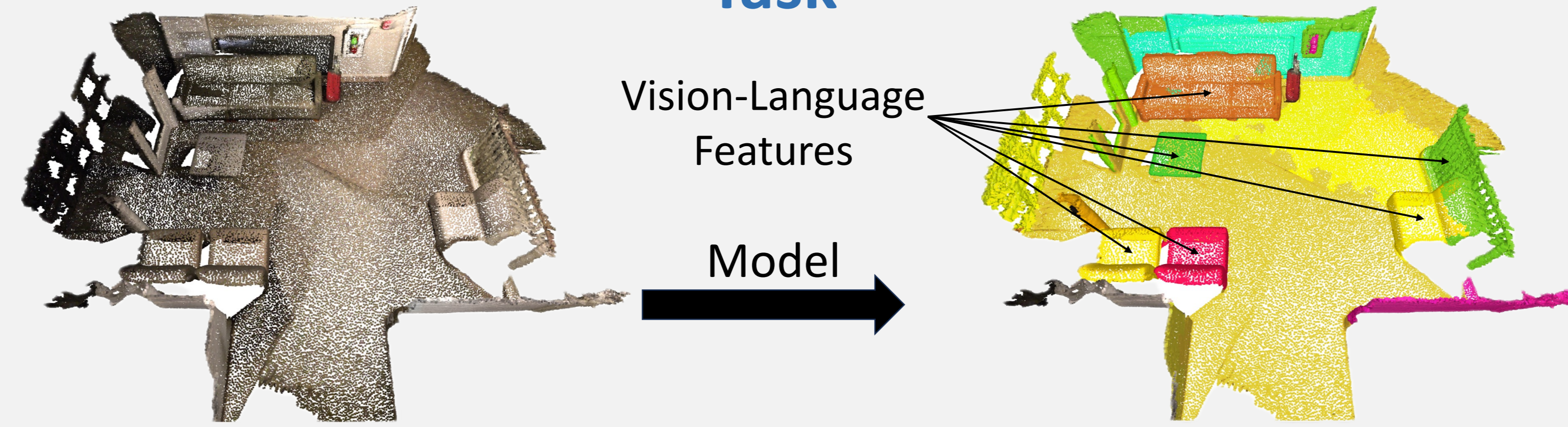
Self-Training

Train a model for open vocabulary instance segmentation using self-training bootstrapped by pseudo ground truth.

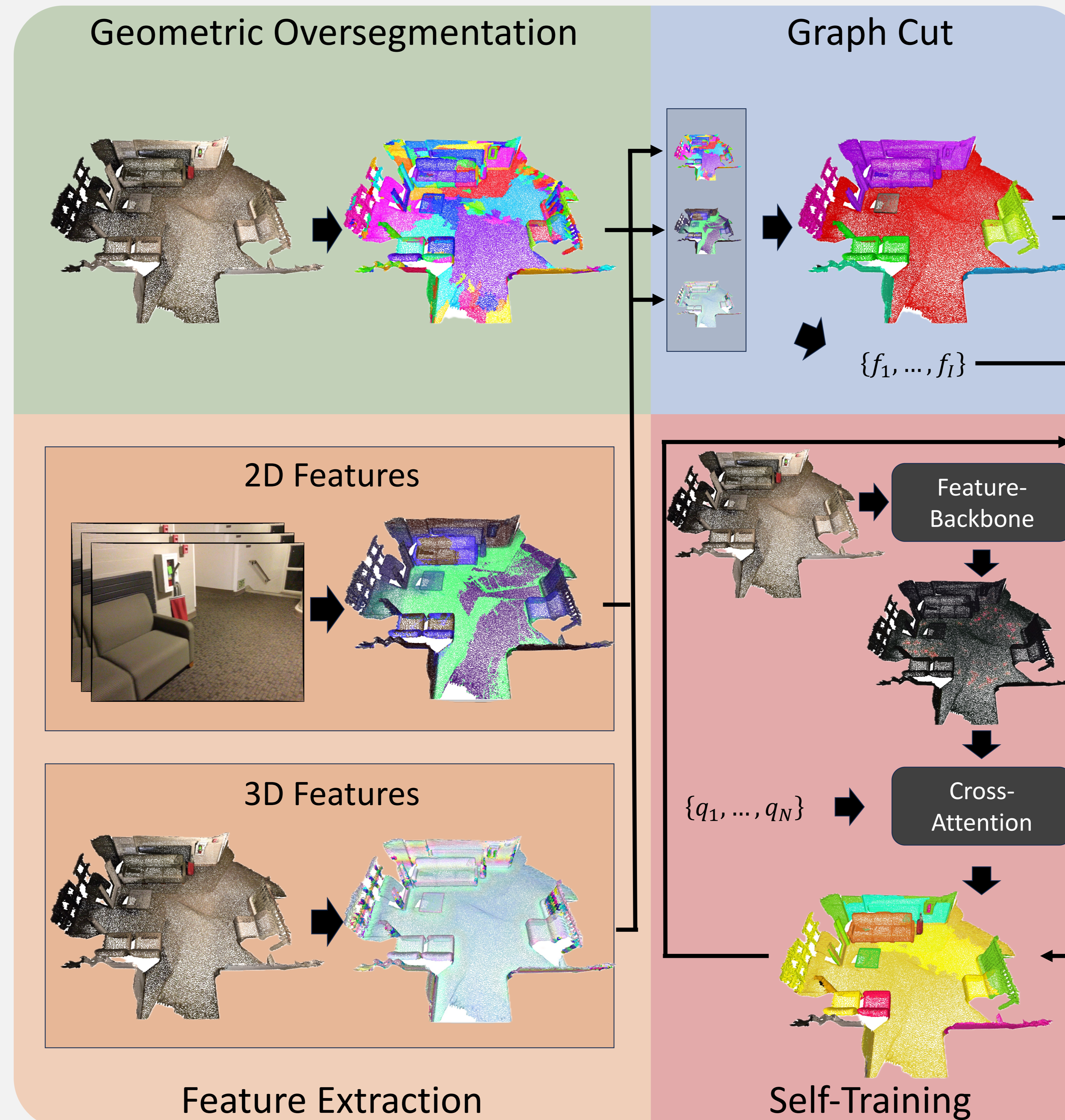
➔ Extract segmentations and features from the mesh.

➔ Fully Unsupervised Training

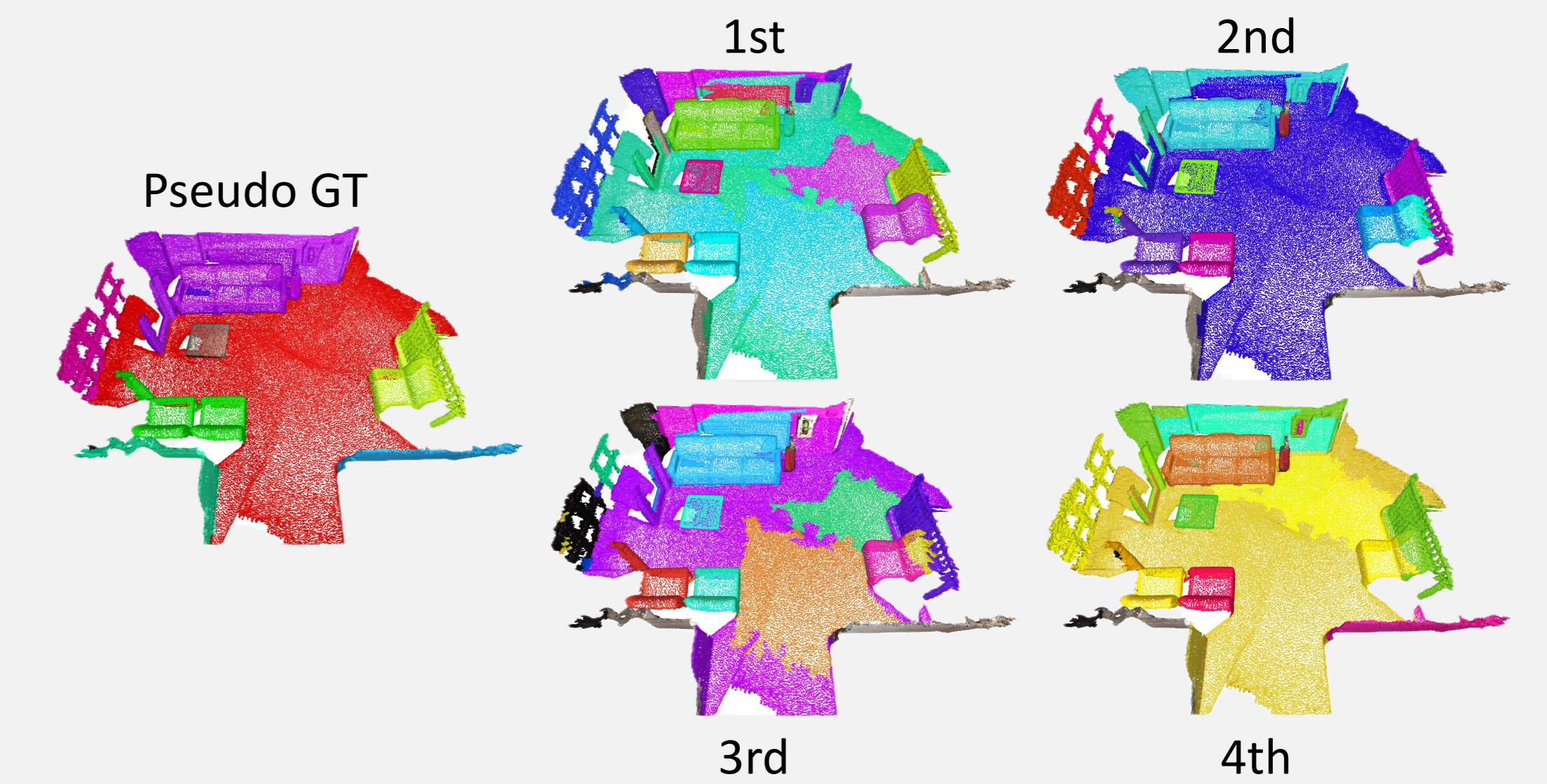
Task



Method Overview



Instance Segmentation Performance



Iteration	UnScene3D ¹			Ours		
	mAP ²⁵	mAP ⁵⁰	mAP	mAP ²⁵	mAP ⁵⁰	mAP
Pseudo GT	19.9	10.0	5.9	16.3	9.4	5.7
1st	52.9	23.2	10.4	19.8	13.1	8.2
2nd	56.9	29.8	15.0	18.2	12.0	7.0
3rd	58.8	31.9	15.9	19.1	12.1	6.9
4th	58.5	32.2	15.9	19.8	12.4	7.1

Other Baselines	mAP ²⁵	mAP ⁵⁰	mAP
HDBSCAN ²	32.1	5.5	1.6
Nunes et al. ³	30.5	7.3	2.3
Felzenszwalb ⁴	38.9	12.7	5.0
CutLer Projection ⁵	7.0	0.2	0.3

[1] D. Rozenberszki, O. Litany, and A. Dai, "UnScene3D: Unsupervised 3D Instance Segmentation for Indoor Scenes," arXiv, Mar. 25, 2023

[2] L. McInnes and J. Healy, "Accelerated Hierarchical Density Clustering," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 33–42

[3] L. Nunes et al., "Unsupervised Class-Agnostic Instance Segmentation of 3D LIDAR Data for Autonomous Vehicles," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 8713–8720, Oct. 2022

[4] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004

[5] X. Wang, R. Girshik, S. X. Yu, and I. Misra, "Cut and Learn for Unsupervised Object Detection and Instance Segmentation," arXiv, Jan. 26, 2023

Natural Language Queries

