OpenScene

3D Scene Understanding with Open Vocabularies

Songyou Peng* Kyle Genova Chiyu "Max" Jiang Andrea Tagliasacchi
Marc Pollefeys Thomas Funkhouser
* Work done during an internship at Google Research

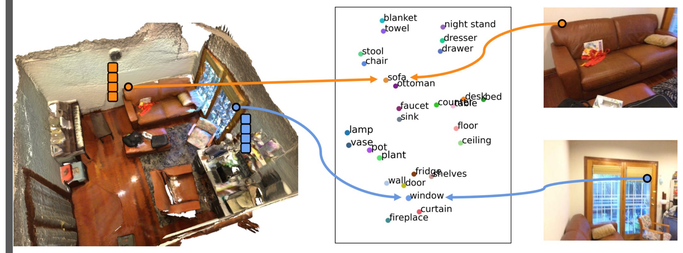
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

1. Introduction

Problem: Traditional 3D scene understanding only train and test on some **fixed common classes**

Goal: A **zero-shot** approach to perform novel 3D scene understanding tasks with **arbitrary queries**

Our Key Idea: Co-embed 3D features with CLIP text and image features

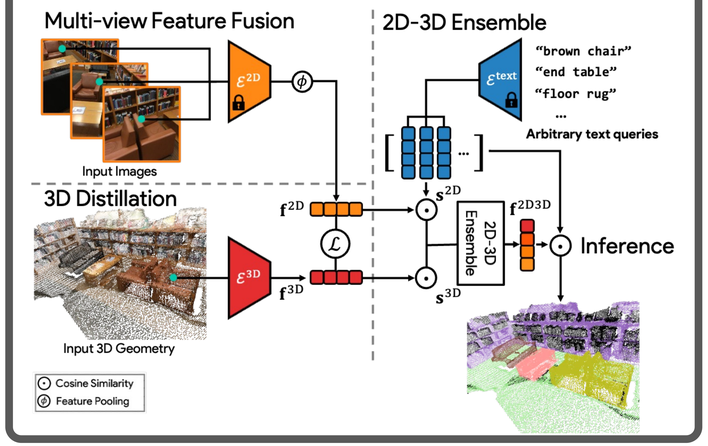


3D Geometry

CLIP Text Features
(visualize with T-SNE)

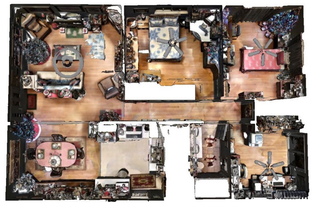
RGB Images

2. Method

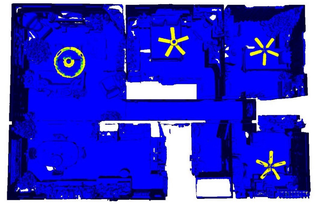


Paper, code, and real-time demo are available:
pengsongyou.github.io/openscene

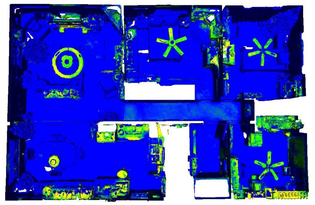
3. Zero-shot Open-vocabulary Scene Exploration



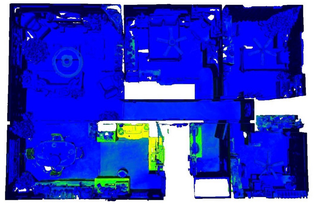
Input 3D Geometry



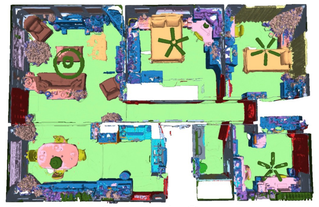
"fan" - Object



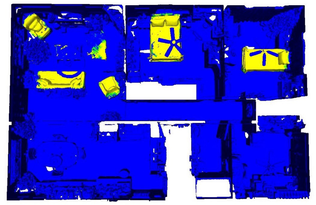
"metal" - Material



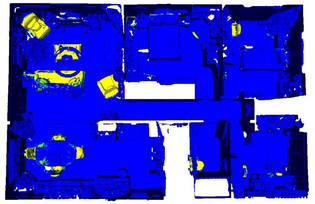
"kitchen" - Room Type



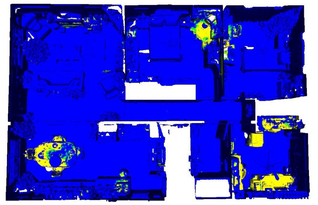
Zero-shot Semantic Segmentation



"soft" - Property



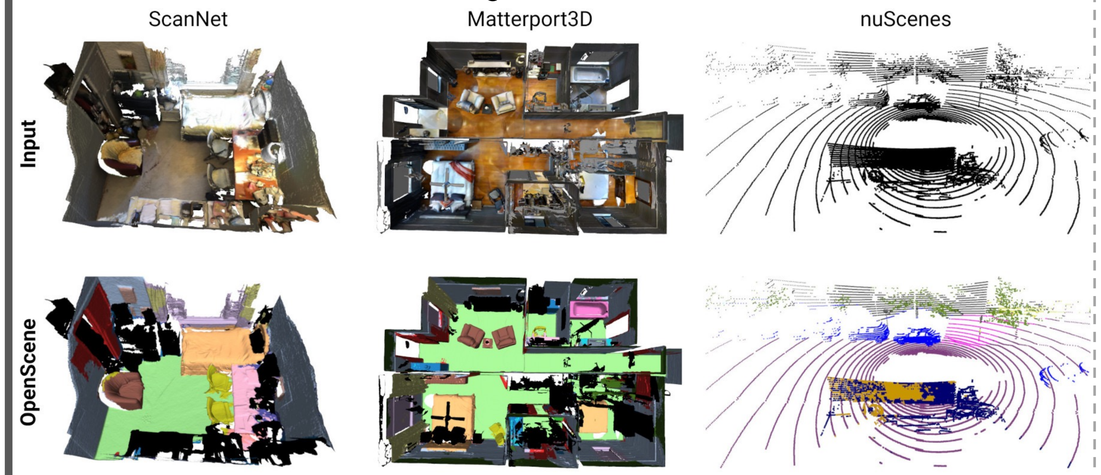
"sit" - Affordance



"work" - Activity

4. Additional Applications

3D Semantic Segmentation Benchmarks



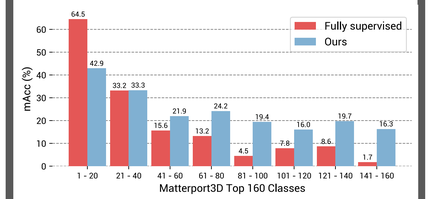
Object Retrieval

Image-based 3D Object Detection

Image Queries | Input 3D Geometry | Our Segmentation

5. More Studies

Robust to Finding Rare Object



Ablation Study

		ScanNet [11]		Matterport3D [4]	
		mIoU	mAcc	mIoU	mAcc
Ours LSeg	2D Fusion	50.0	62.7	32.3	40.0
	3D Distill	52.9	63.2	41.9	51.2
	2D-3D Ens.	**54.2**	66.6	**43.4**	53.5
Ours OpenSeg	2D Fusion	41.4	63.6	32.4	45.0
	3D Distill	46.0	66.3	41.3	55.1
	2D-3D Ens.	47.5	**70.7**	42.6	**59.2**