OpenScene

3D Scene Understanding with Open Vocabularies

Songyou Peng[1,3]* Kyle Genova[1] Chiyu "Max" Jiang[2] Andrea Tagliasacchi[1,5]
Marc Pollefeys[3] Thomas Funkhouser[1]

* Work done during an internship at Google Research

1 Google 2 Waymo 3 ETH zürich 4 5 SFU

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

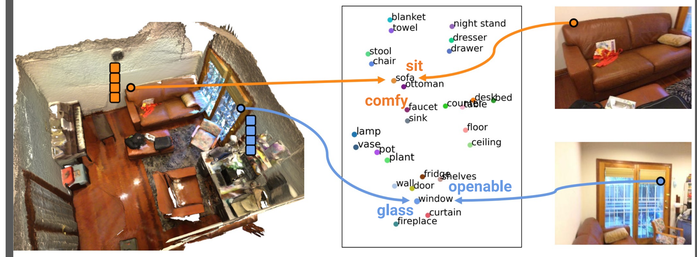
Paper, code, and real-time demo are available:
pengsongyou.github.io/openscene

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 **w/o annotation labels**

Key idea: Co-embed 3D features with CLIP image features → naturally also with CLIP text features



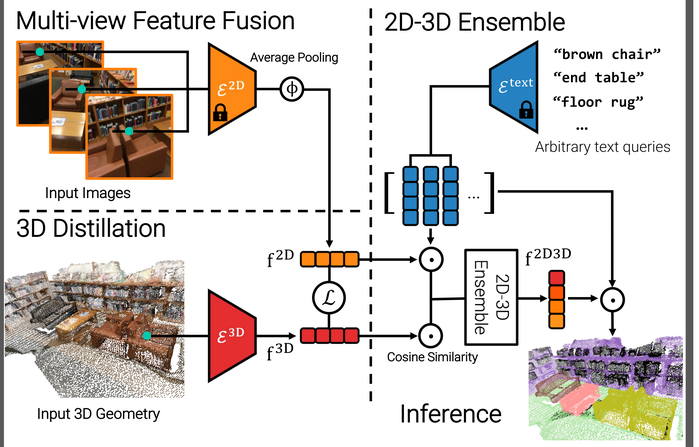
3D Geometry

CLIP Text Features
(visualize with T-SNE)

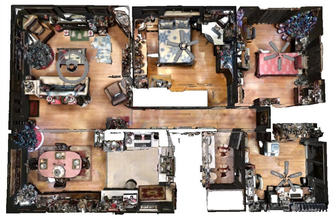
RGB Images

2. Method

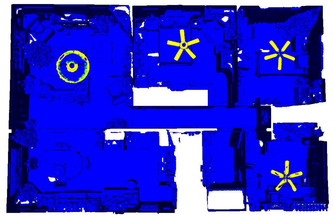
How to produce text-image-3D co-embedding?



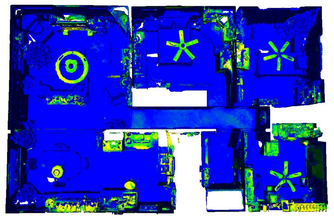
3. Zero-shot Open-vocabulary Scene Exploration



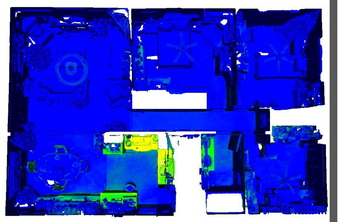
Input 3D Point Cloud



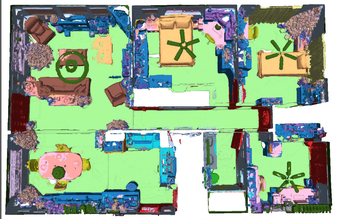
"fan" - Object



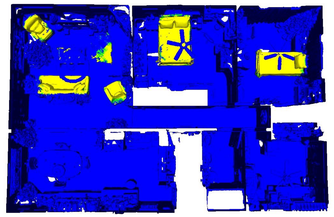
"made of metal" - Material



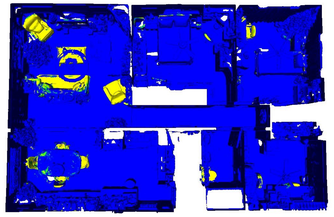
"kitchen" - Room Type



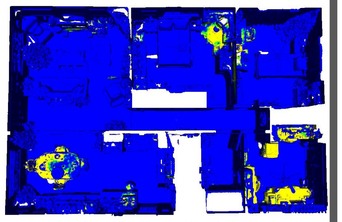
Zero-shot Semantic Segmentation



"anything soft" - Property



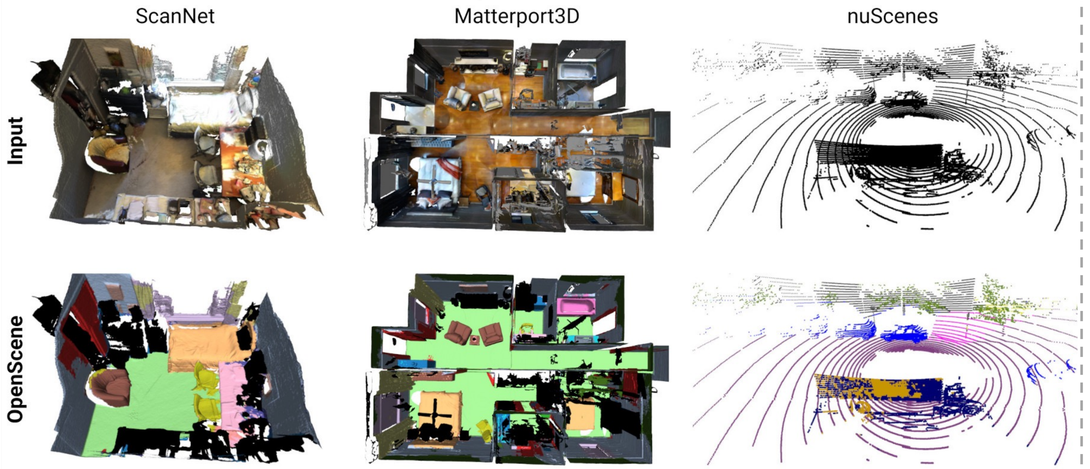
"where to sit" - Affordance



"work" - Activity

4. Additional Applications

3D Semantic Segmentation Benchmarks

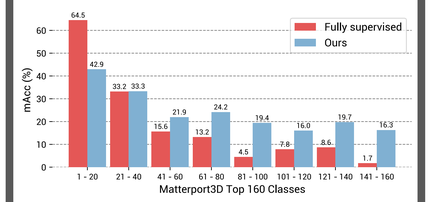


Rare Object Retrieval

Image-based 3D Object Detection

5. More Studies

Robust to Tailed Classes



Ablation Study