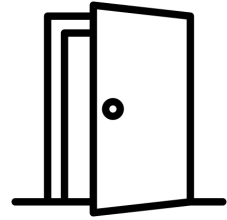




**ETH** zürich



# OpenScene

## 3D Scene Understanding with Open Vocabularies

CVPR 2023

Songyou Peng



Kyle Genova



Chiyu "Max" Jiang



Andrea Tagliasacchi



Marc Pollefeys



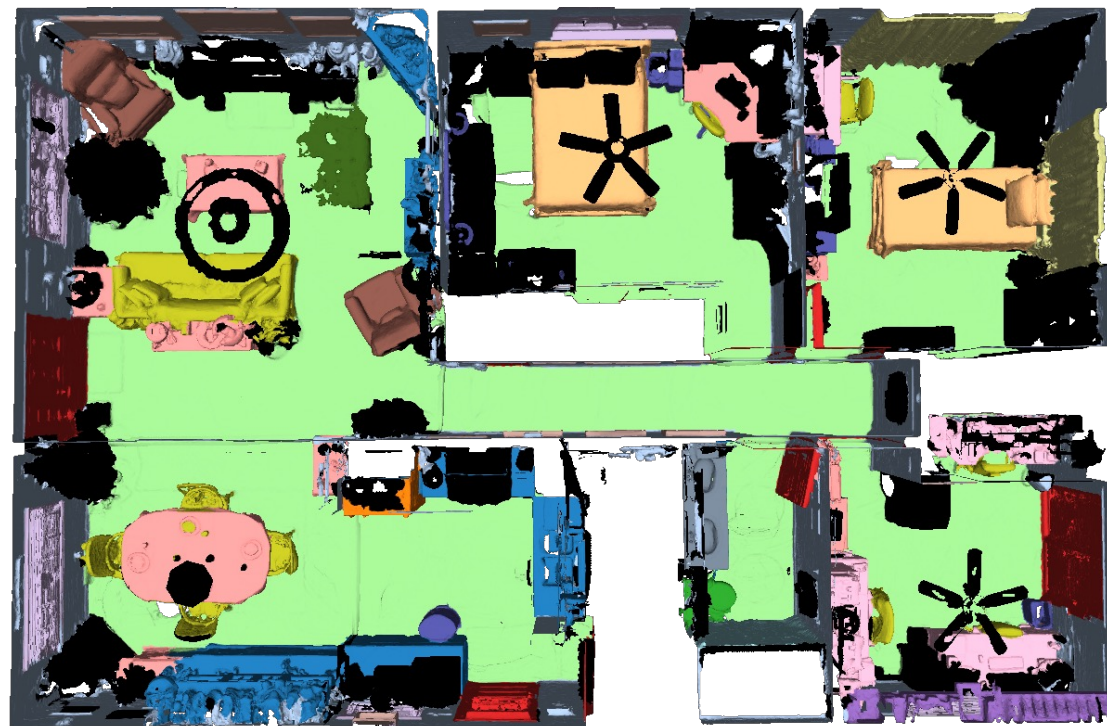
Tom Funkhouser





Input 3D Geometry

■ wall ■ floor ■ cabinet ■ bed ■ chair ■ sofa ■ table ■ door  
■ window ■ counter ■ curtain ■ toilet ■ sink ■ bathtub ■ other ■ unlabeled



Traditional Semantic Segmentation

Only train and test on a few common classes



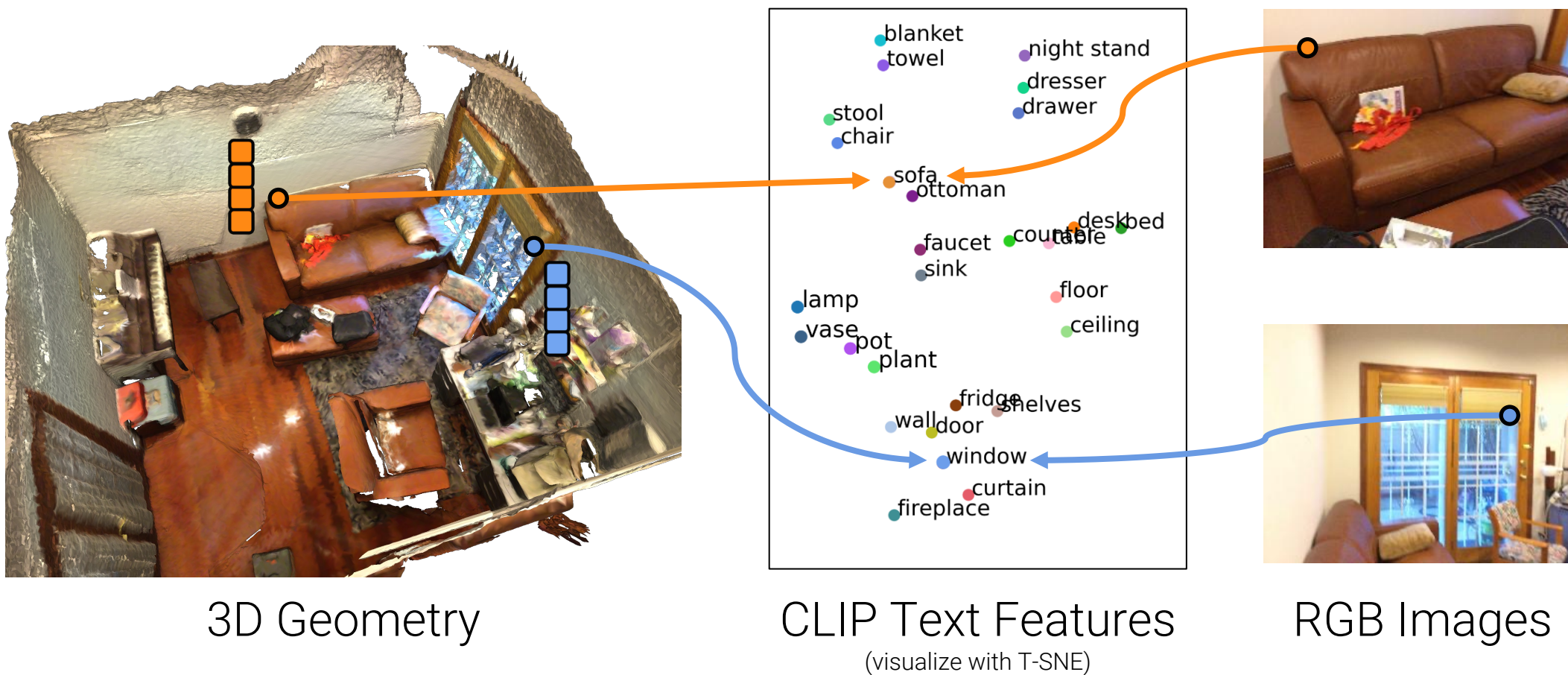


Input 3D Geometry

- Affordance prediction
- Material identification
- Physical property estimation
- Rare object retrieval
- Activity site prediction
- Fine-grained semantic segmentation
- Many more...

## 3D Scene Understanding Tasks w/o Labels

# Key Idea: Co-embed 3D features with CLIP features



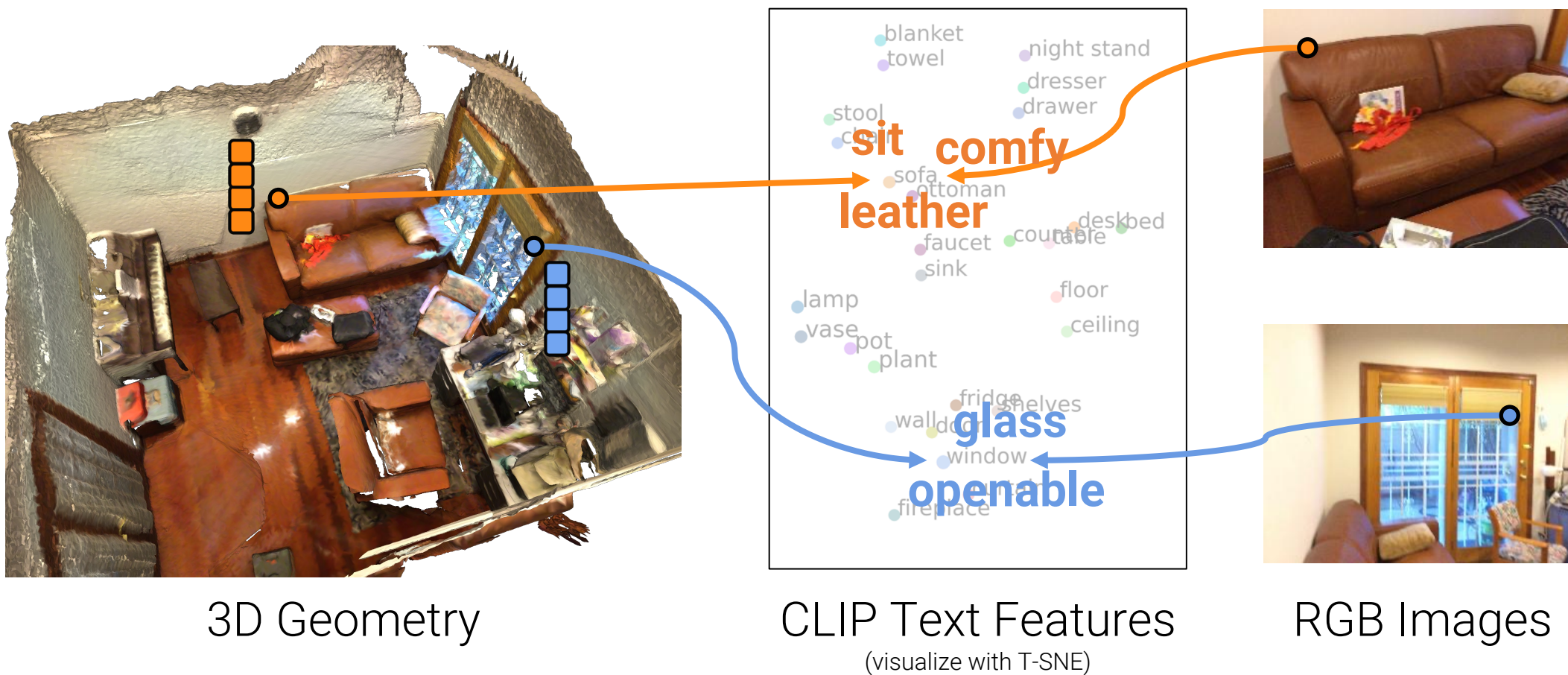
3D Geometry

CLIP Text Features  
(visualize with T-SNE)

RGB Images



# Key Idea: Co-embed 3D features with CLIP features



3D Geometry

CLIP Text Features  
(visualize with T-SNE)

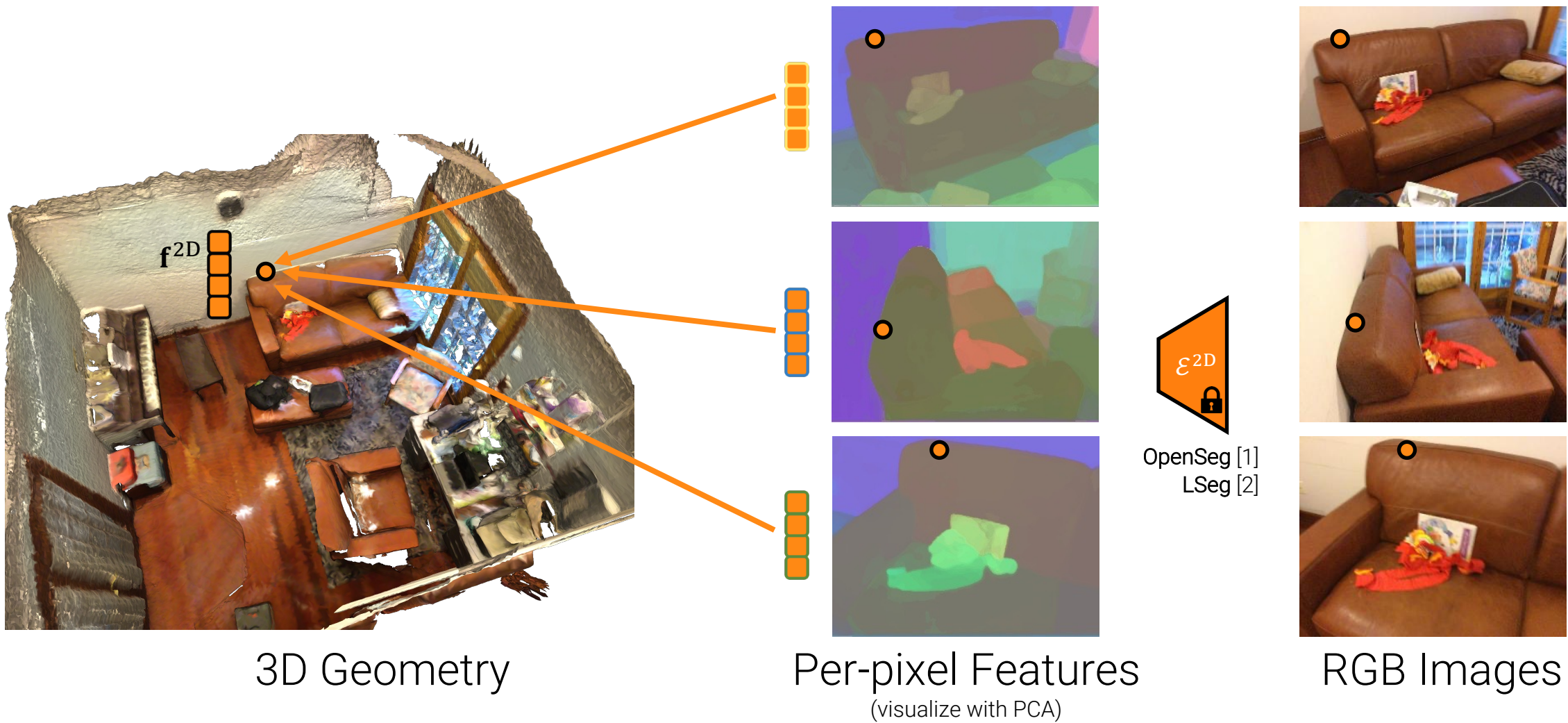
RGB Images

Note: bold word embeddings are approximate

How to Learn Such **Text-Image-3D Co-Embeddings**?



# Step 1: Multi-view Feature Fusion



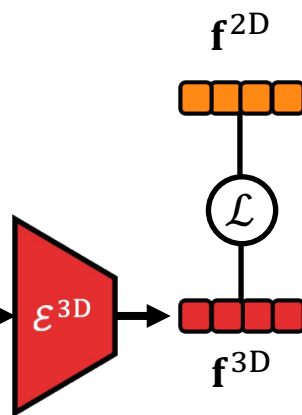
[1] Ghiasi, Gu, Cui, Lin: [Scaling Open-Vocabulary Image Segmentation with Image-Level Labels](#). ECCV 2022

[2] Li, Weinberger, Belongie, Koltun, Ranftl: [Language-driven Semantic Segmentation](#). ICLR 2022

# Step 2: 3D Distillation



3D Geometry



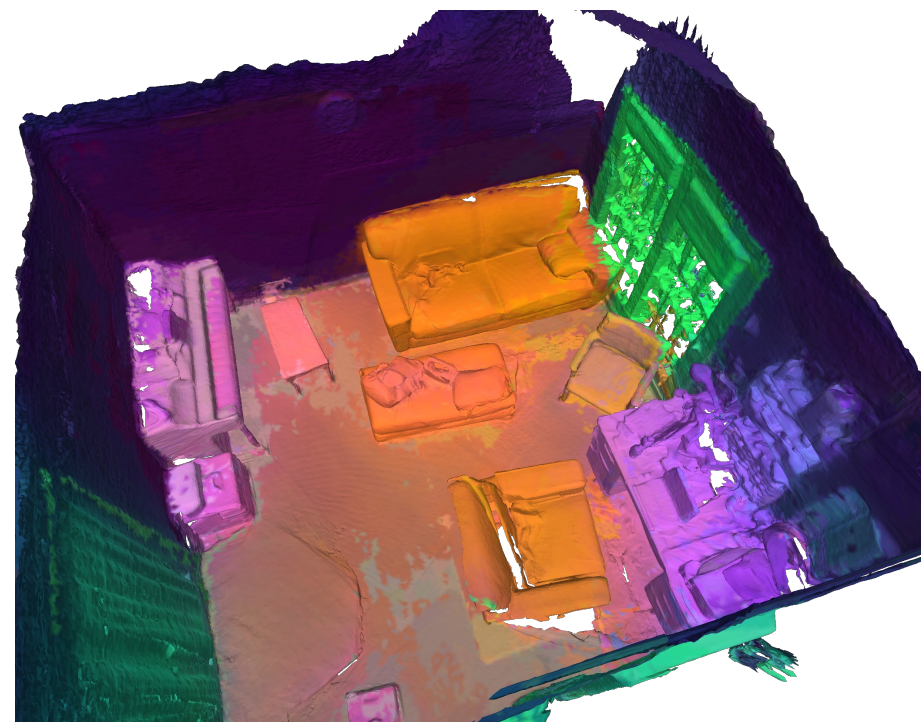
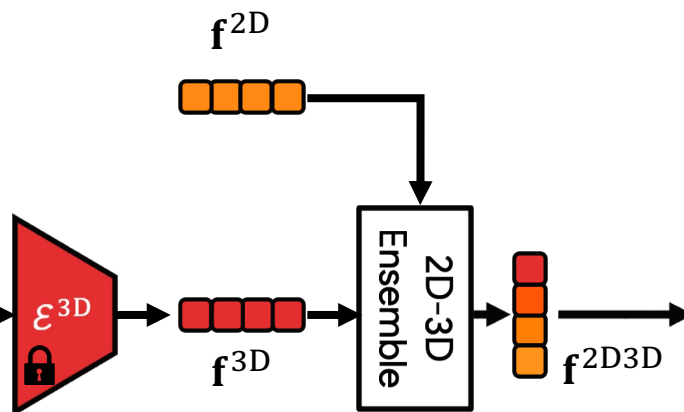
$$\mathcal{L} = 1 - \cos(\mathbf{f}^{2D} - \mathbf{f}^{3D})$$



# Step 3: 2D-3D Ensemble



3D Geometry



2D-3D Ensemble Features  
(visualize with PCA)

# Open-Vocabulary, Zero-shot 3D Semantic Segmentation





Input 3D Geometry



Our Zero-shot 3D Segmentation  
(20 classes)

■ wall ■ floor ■ cabinet ■ bed ■ chair ■ sofa ■ table ■ door ■ window ■ bookshelf ■ picture ■ counter ■ desk ■ curtain ■ refrigerator ■ shower curtain ■ toilet ■ sink ■ bathtub ■ other

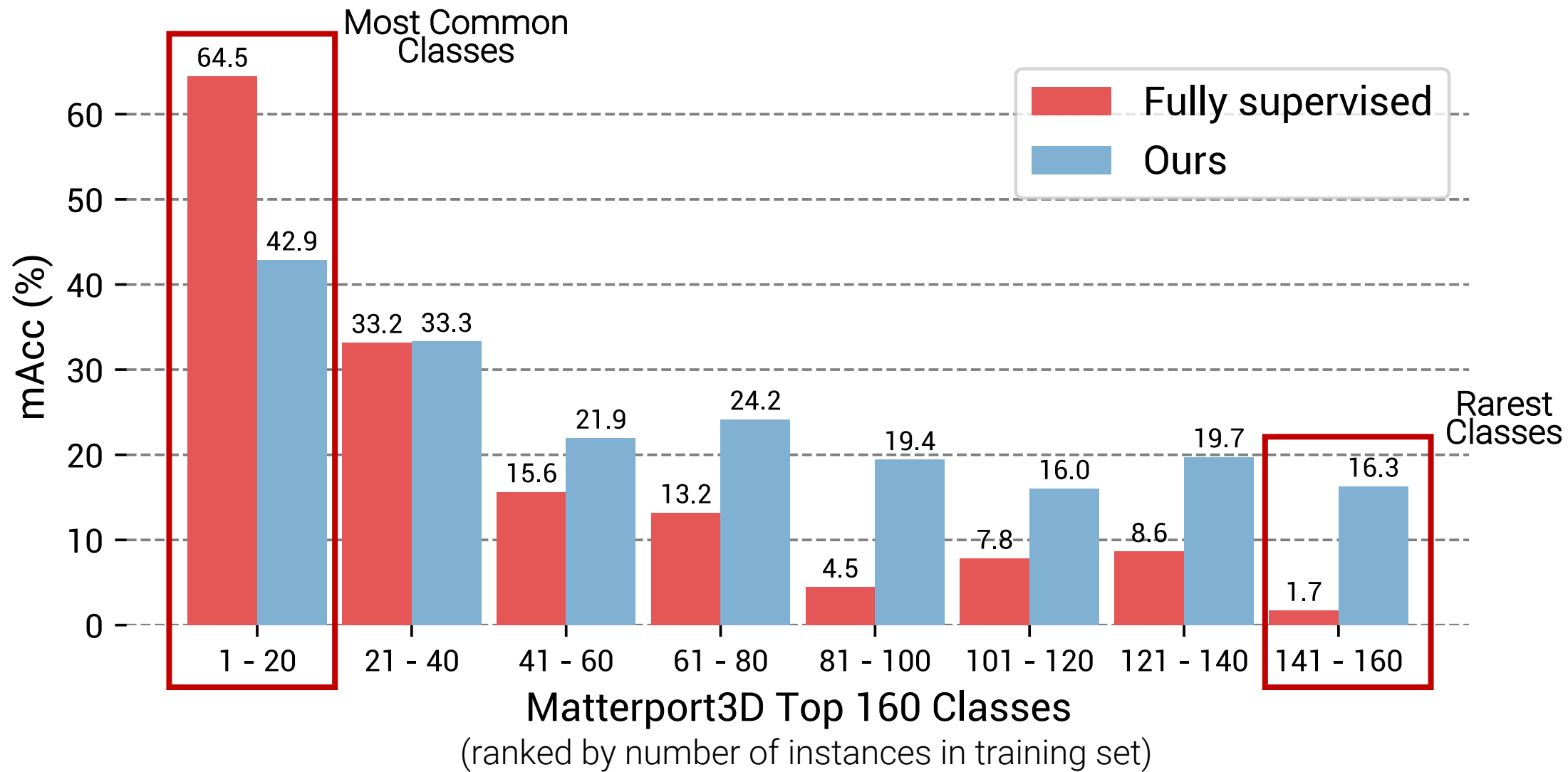




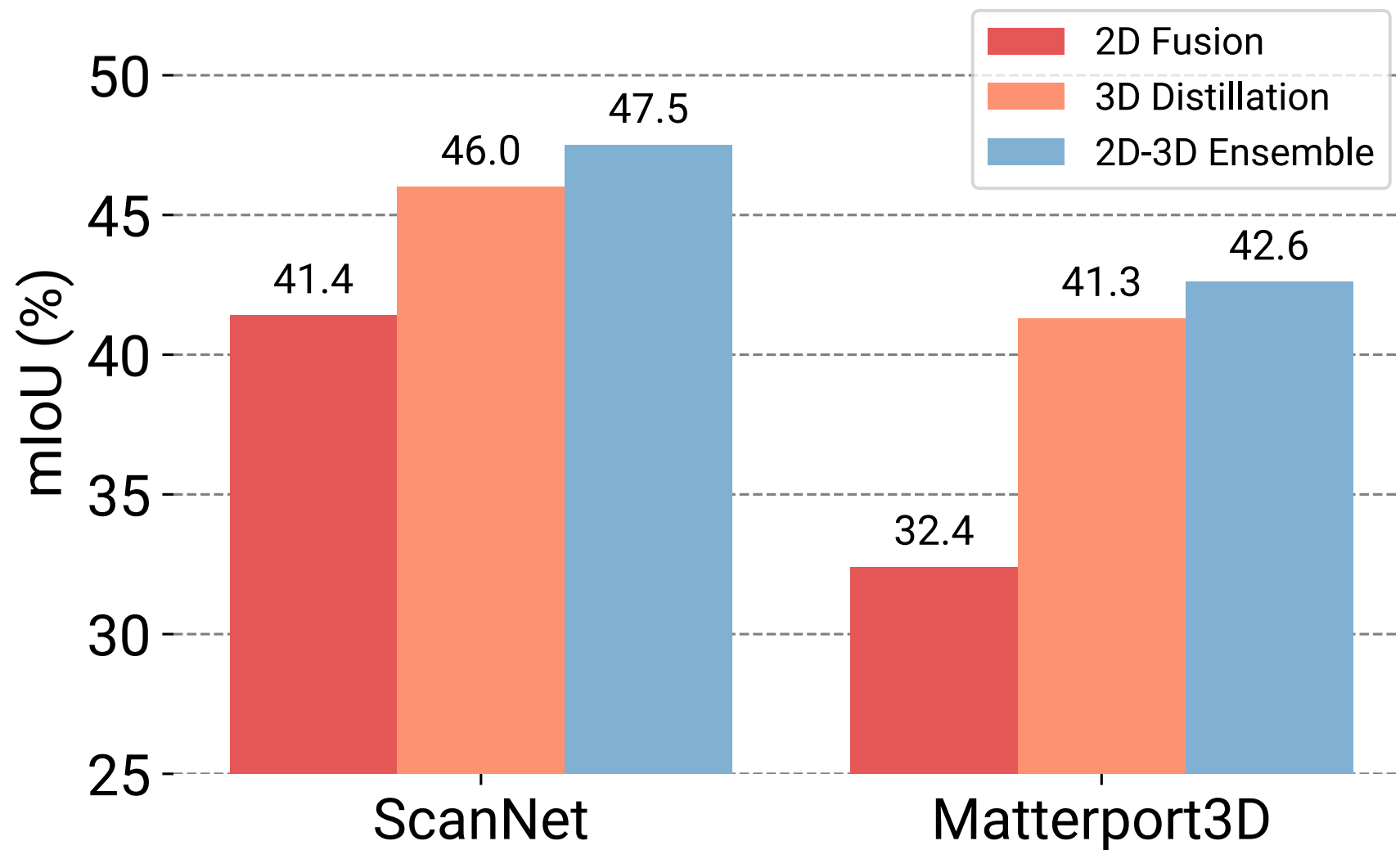
Our Zero-shot 3D Segmentation  
(160 classes)

- |           |           |               |                |                |                 |                |                   |              |                     |                      |                       |                  |                  |                         |             |
|-----------|-----------|---------------|----------------|----------------|-----------------|----------------|-------------------|--------------|---------------------|----------------------|-----------------------|------------------|------------------|-------------------------|-------------|
| ■ wall    | ■ cabinet | ■ bed         | ■ pot          | ■ bathtub      | ■ dresser       | ■ stand        | ■ clock           | ■ tissue box | ■ furniture         | ■ soap               | ■ cup                 | ■ hanger         | ■ urn            | ■ paper towel dispenser | ■ toy       |
| ■ door    | ■ curtain | ■ night stand | ■ desk         | ■ book         | ■ rug           | ■ drawer       | ■ stove           | ■ tv stand   | ■ air conditioner   | ■ thermostat         | ■ ladder              | ■ candlestick    | ■ plate          | ■ lamp shade            | ■ foot rest |
| ■ ceiling | ■ table   | ■ toilet      | ■ box          | ■ air vent     | ■ ottoman       | ■ container    | ■ washing machine | ■ shoe       | ■ fire extinguisher | ■ radiator           | ■ garage door         | ■ light          | ■ car            | ■ soap dish             |             |
| ■ floor   | ■ plant   | ■ column      | ■ coffee table | ■ faucet       | ■ bottle        | ■ light switch | ■ shower curtain  | ■ heater     | ■ kitchen island    | ■ paper towel        | ■ board               | ■ scale          | ■ jacket         | ■ toilet brush          | ■ cleaner   |
| ■ picture | ■ mirror  | ■ banister    | ■ counter      | ■ photo        | ■ refridgerator | ■ purse        | ■ bin             | ■ headboard  | ■ printer           | ■ sheet              | ■ piano               | ■ bag            | ■ bottle of soap | ■ drum                  | ■ computer  |
| ■ window  | ■ towel   | ■ stairs      | ■ bench        | ■ toilet paper | ■ bookshelf     | ■ door way     | ■ chest           | ■ telephone  | ■ bucket            | ■ rope               | ■ display case        | ■ display case   | ■ water cooler   | ■ whiteboard            | ■ knob      |
| ■ chair   | ■ sink    | ■ stool       | ■ garbage bin  | ■ fan          | ■ wardrobe      | ■ basket       | ■ microwave       | ■ blanket    | ■ glass             | ■ ball               | ■ toilet paper holder | ■ tea pot        | ■ range hood     | ■ paper                 | ■ projector |
| ■ pillow  | ■ shelves | ■ vase        | ■ fireplace    | ■ railing      | ■ pipe          | ■ chandelier   | ■ blinds          | ■ flower pot | ■ dishwasher        | ■ exercise equipment | ■ tray                | ■ stuffed animal | ■ candelabra     |                         |             |

# Comparison



# Ablation





# Image-based 3D Scene Query



Image Queries

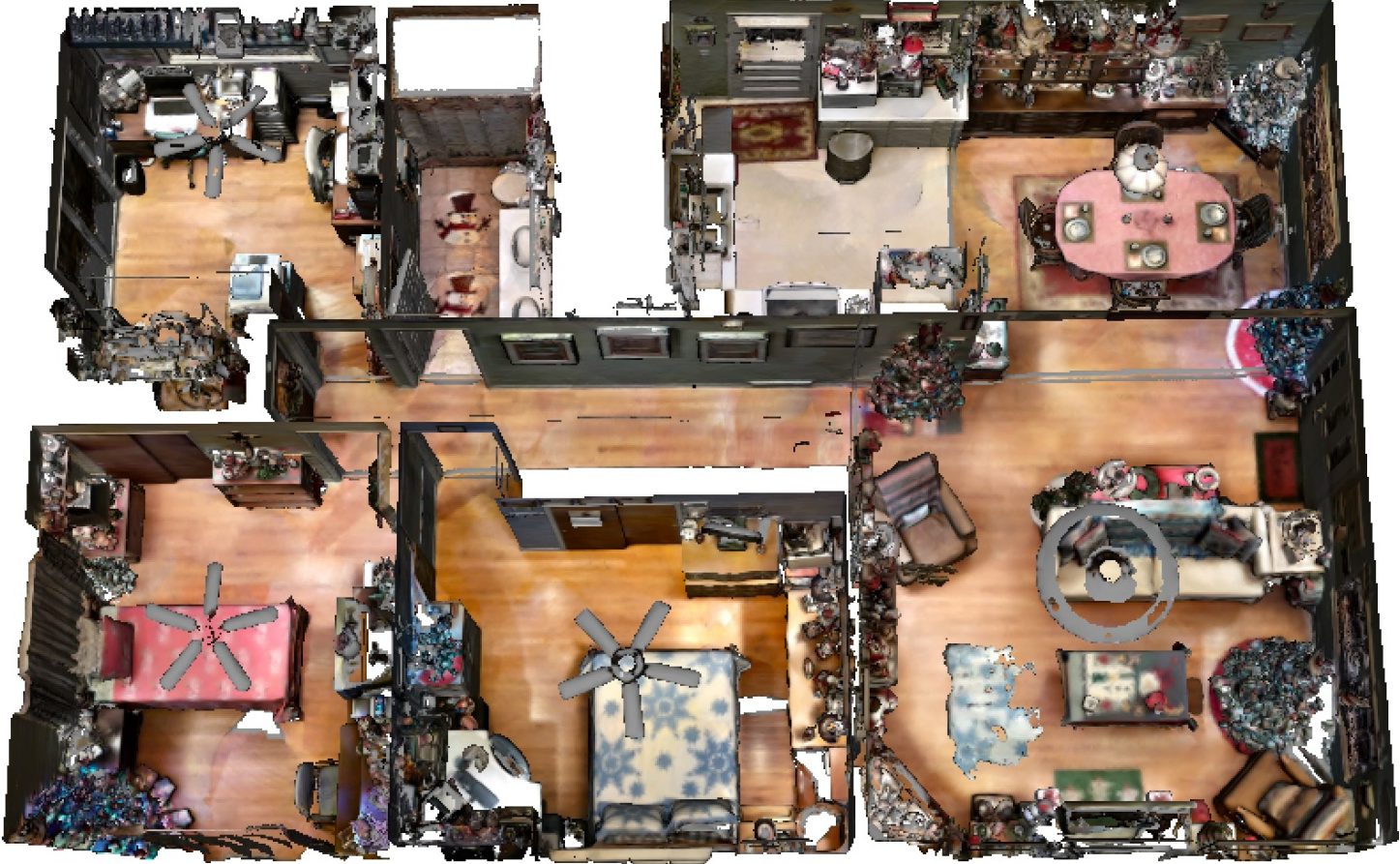
Given 3D Geometry

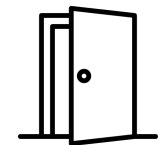
# **Interactive Demo**

Open-vocabulary 3D Scene Exploration



Text queries:

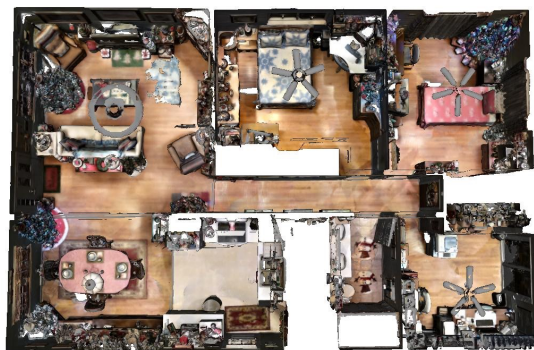




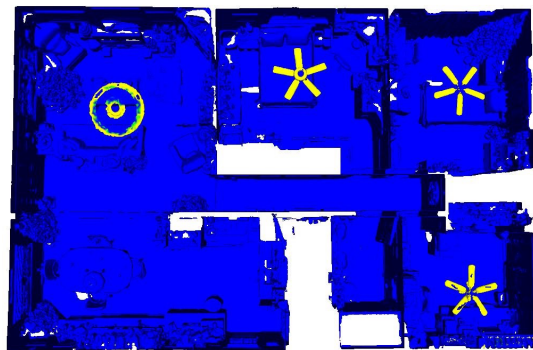
# OpenScene: 3D Scene Understanding with Open Vocabularies

Paper, code, and real-time demo at:

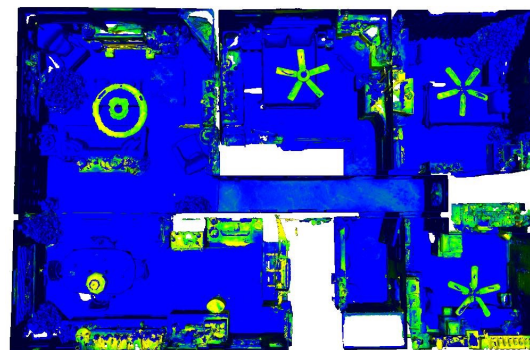
[pengsongyou.github.io/openscene](https://pengsongyou.github.io/openscene)



Input 3D Point Cloud



“fan” - Object



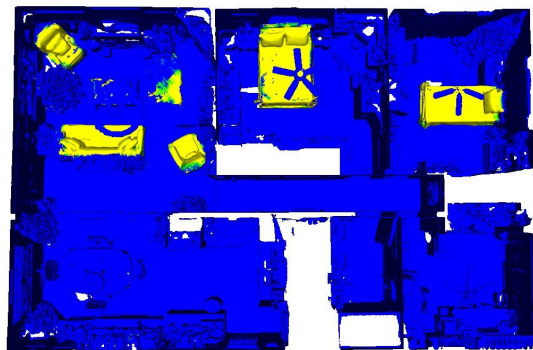
“made of metal” - Material



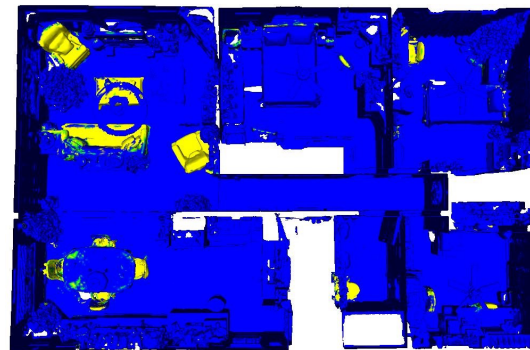
“kitchen” - Room Type



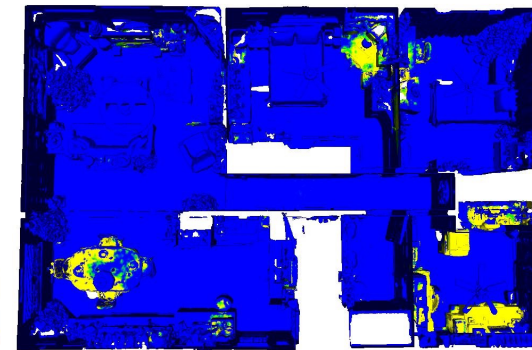
Zero-shot Semantic Segmentation



“anything soft” - Property



“where to sit” - Affordance



“work” - Activity