CVPR JUNE 17-21, 2024 SEATTLE, WA

https://quest-maps.github.io/

QueSTMaps

Queryable Semantic Topological Maps for 3D Scene Understanding

OpenSUN 3D

Yash Mehan*[1], Kumaraditya Gupta*[1], Rohit Jayanti*[1], Anirudh Govil[1], Sourav Garg[2], Madhava Krishna[1]

1 INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

IIIT Hyderabad

2 THE UNIVERSITY of ADELAIDE

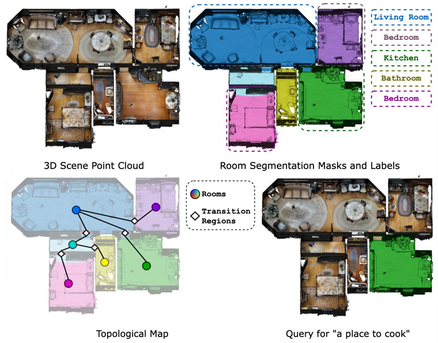
*Denotes authors with equal contribution

Overview: Building Room-Level Topology from Point Clouds

Problem: 3D scene understanding methods built solely upon object-level segmentation methods struggle to identify dense room-level regions.

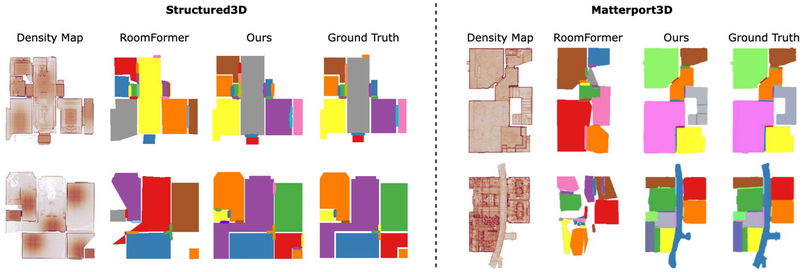
Our Solution: Assuming, only a posed RGB-D image sequence, we propose a two-step pipeline that comprises:

1. A room instance segmentation model based on a novel multi-channel representation comprising occupancy slices and a density map.
2. A room labeling transformer model which uses object features from ConceptGraphs to output a room label-aligned CLIP embedding.



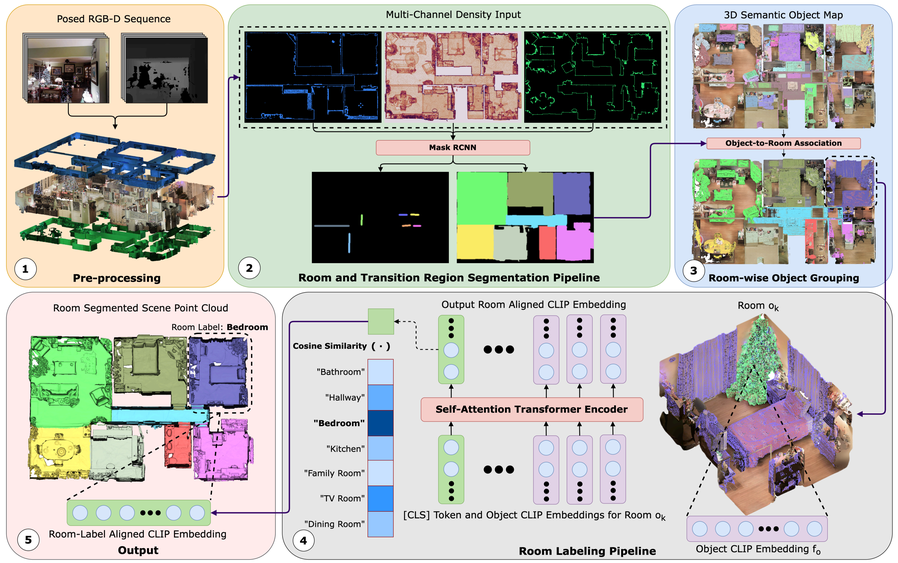
Our language-topology alignment supports natural language querying, e.g., a "`place to cook`" locates the "`kitchen`". We report an improvement of ~20% on room segmentation and ~12% on room classification over current state-of-the-art methods.

Room Segmentation Results



Our proposed method is able to handle adverse cases such as regions of low density (introduced by the scanning strategy) in the Stuctured3D dataset, as well as curved walls and clutter as seen in the Matterport3D dataset. **Note**: The colors only denote instance separation.

Model Pipeline: Room Segmentation and Labeling



Quantitative Results

Method	Struc3D		MP3D	
	TR (AP)	Rooms (AP)	TR (AP)	Rooms (AP)
HEAT [12]	-	94.24	-	30.78
RoomFormer [50]	76.94	96.93	24.85	64.42
Ours (Single Channel)	85.60	94.90	50.47	84.83
Ours (Multi Channel)	**87.99**	**98.86**	**60.92**	**88.47**

Table 1. Quantitative Results for Room Segmentation.

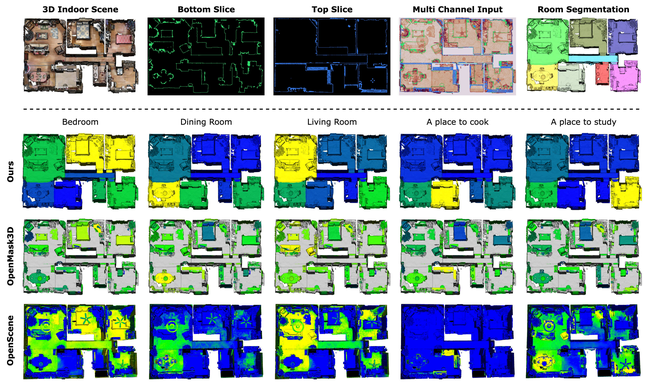
Method	Prec	Rec	F1	mAP
Statistical [13]	49.49	50.50	42.48	28.35
Zero Shot LLM [13]	46.69	45.34	43.00	48.07
Embedding-Based [13]	60.07	66.40	62.75	67.17
GNN Model [27]	58.79	65.59	61.38	67.45
Avg Object Embeddings	66.28	60.98	60.94	69.02
Ours (Logits)	75.31	75.61	74.63	79.04
Ours (Contrastive)	**77.97**	**76.83**	**75.43**	**79.12**

Table 2. Comparison of Room Labeling Models.

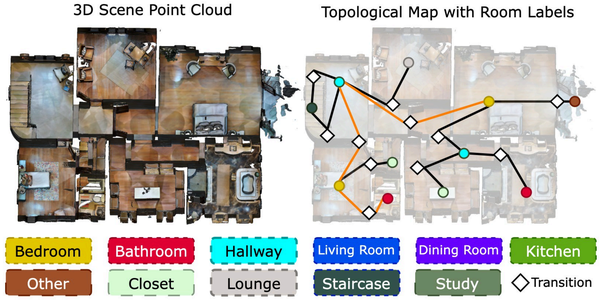
Method	mAP @ 0.5 IoU
RoomFormer [50]	23.52
GT Room Masks + Our Room labeling	79.12
Ours (Complete Pipeline)	74.02

Table 3. Complete Pipeline Evaluation.

Open-Set Natural Language Queries



Downstream Task: Navigation



Detected transition regions aid in building a room-level topological map which can be used for tasks such as room-to-room navigation using natural language commands.