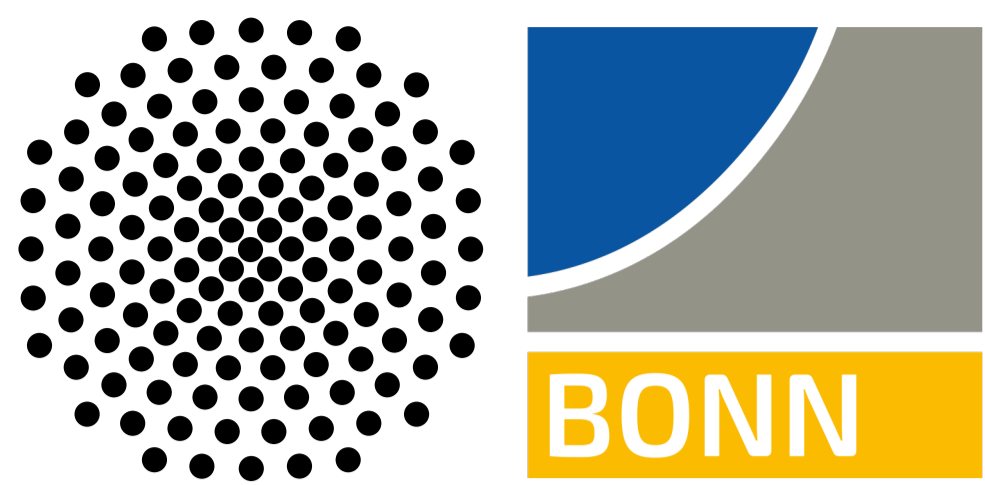


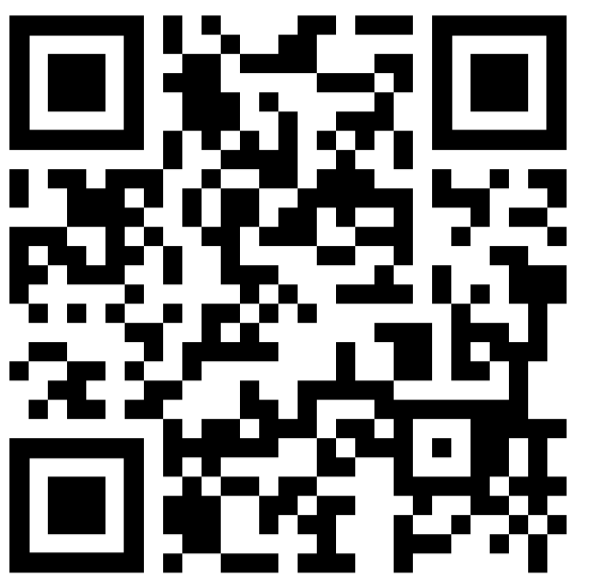
FunGraph: Functionality Aware 3D Scene Graphs for Language-Prompted Scene Interaction



Dennis Rotondi¹, Fabio Scaparro¹, Hermann Blum², Kai O. Arras¹

¹University of Stuttgart, Germany

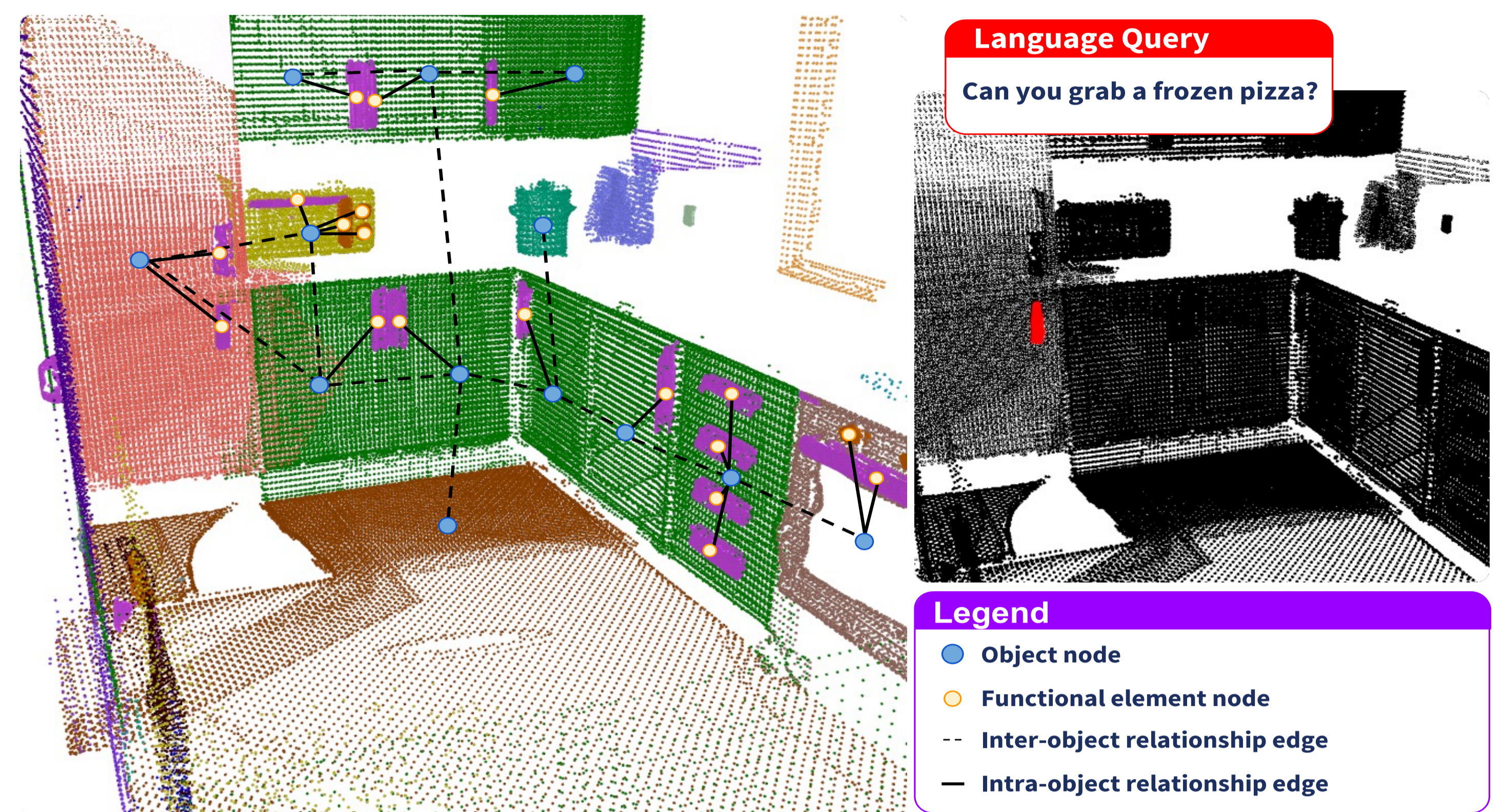
²University of Bonn, Germany



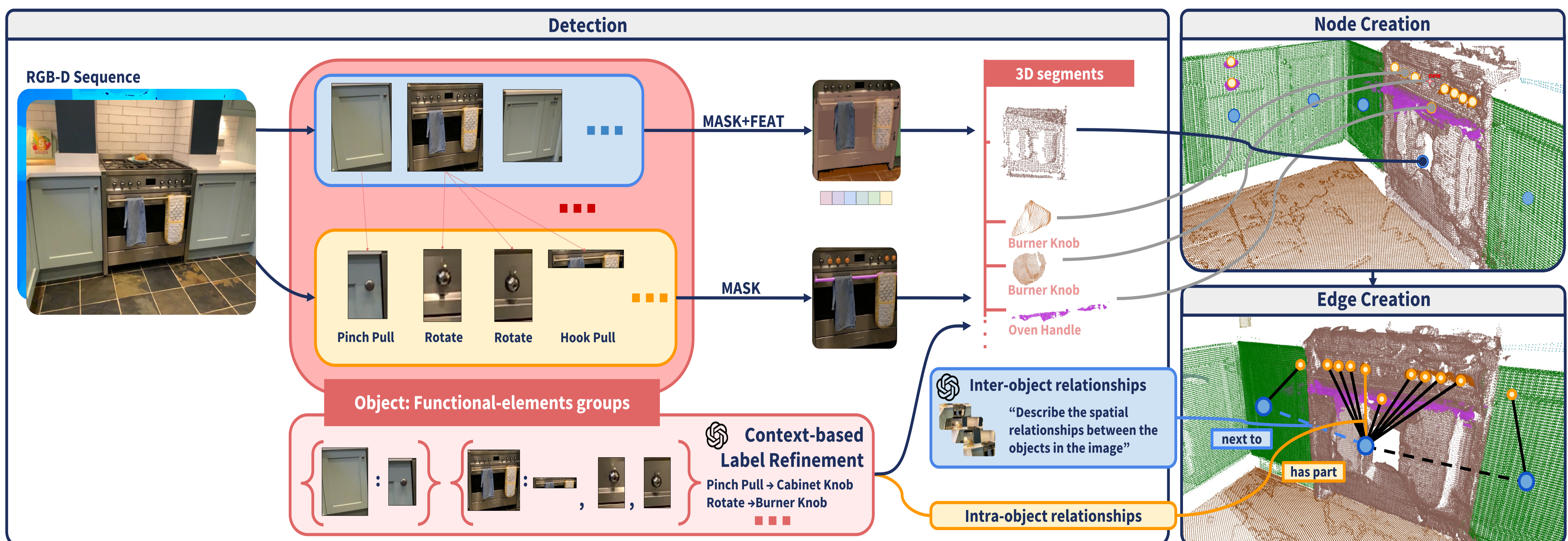
Contributions

- Introduces FunGraph, a novel framework for embedding functional interactive elements (FIEs) and their affordances into 3D scene graphs.
- Models not only inter-object relationships but also intra-object ones. For example, an oven is related to its control knobs, and a door to its handle.
- Addresses data scarcity and sensing challenges in fine-grained 3D object understanding by automatically generating 2D annotations from the SceneFun3D [1] dataset. This is particularly relevant for robotic incremental solutions that rely on RGB-D streams.
- Achieves state-of-the-art performance in FIE segmentation and improves accuracy in task-driven affordance grounding.

Applications

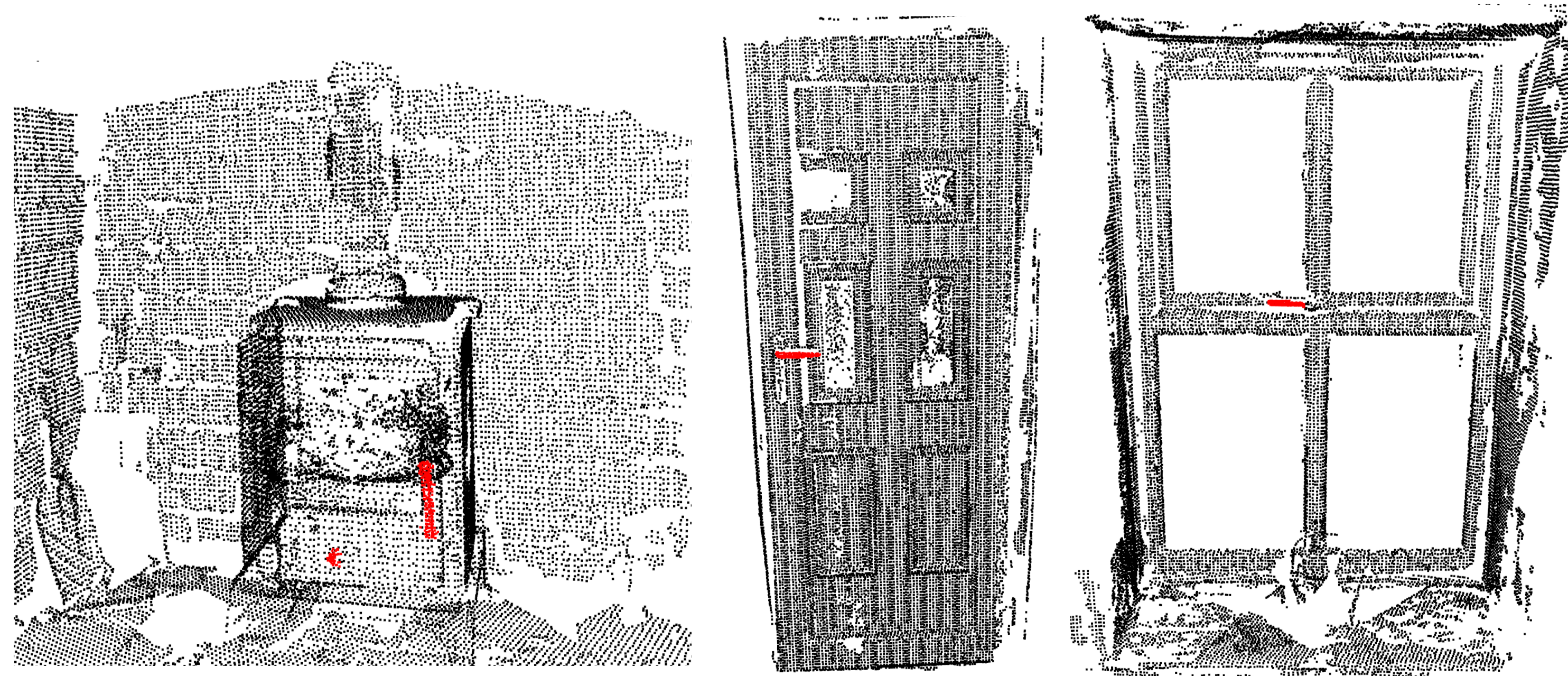


Method



Data Generation for Fine-tuning

- We produce data for 2D and 3D detection and segmentation.



Results

	CLIP			GPT-No-Context			GPT-Context		
	C	P	W	C	P	W	C	P	W
Handle	32.5	37.5	30.0	77.5	12.5	10.0	82.5	10.0	7.5
Knob	43.3	33.3	23.3	23.3	56.7	20.0	70.0	23.3	6.7
Button	10.0	15.0	75.0	55.0	20.0	25.0	85.0	10.0	5.0
Other	20.0	0.0	80.0	30.0	20.0	50.0	70.0	30.0	0.0
Total	30.0	28.0	42.0	52.0	28.0	20.0	78.0	16.0	6.0

We produce accurate OV labels for the 3D FIEs nodes.

Affordance	AP	AP ₅₀	AP ₂₅
FunGraph	5.9	16.0	30.3
FunGraph+PTv3	8.1	21.2	37.7
Mask3D-F [5, 26]	[7.9]	[18.3]	[26.6]

Our 3D FIEs segmentation is similar to SOTA 3D models.

Scene	#Queries	ConceptGraphs		FunGraph (ours)	
		AP ₂₅	AP _{>0}	AP ₂₅	AP _{>0}
423070	8	0.0	25.0	50.0	50.0
423306	3	0.0	0.0	33.3	66.7
423738	21	0.0	57.1	33.3	85.7
434892	5	0.0	40.0	40.0	40.0
435357	10	0.0	50.0	30.0	60.0
435715	12	0.0	8.3	33.3	75.0
435724	10	0.0	10.0	10.0	20.0
442392	8	0.0	25.0	37.5	37.5
464754	18	0.0	22.2	27.8	44.4
467330	4	0.0	50.0	100	100
Total	99	0.0	31.3	33.3	58.6

Compared to the SOTA, our method responds to task-driven affordance queries with higher accuracy.

