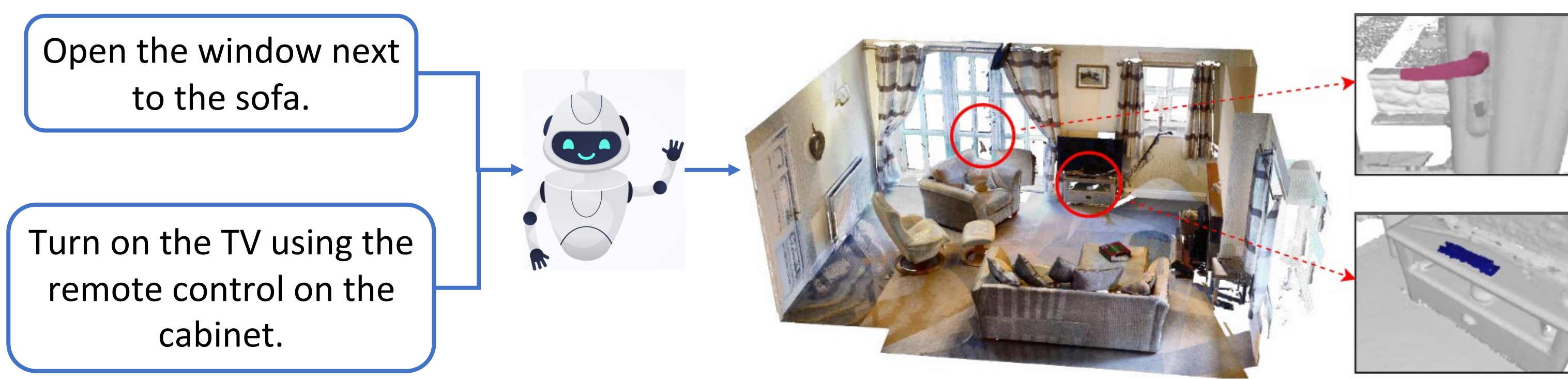


## Motivation

- Functionality segmentation provides actionable perception for manipulation.
- **Existing methods are inefficient** in terms of runtime and memory usage.



We introduce **T-FunS3D**, a **training-free open-vocabulary functional segmentation** method achieving SOTA performance on SceneFun3D dataset.

- Scene graphs are used to embed open-set instances and their spatial relations for efficient scene querying.
- Runtime and memory footprint are reduced by instance-to-part segmentation approach.
- Pre-trained large models are adopted to text understanding, semantic extraction, and functionality segmentation.

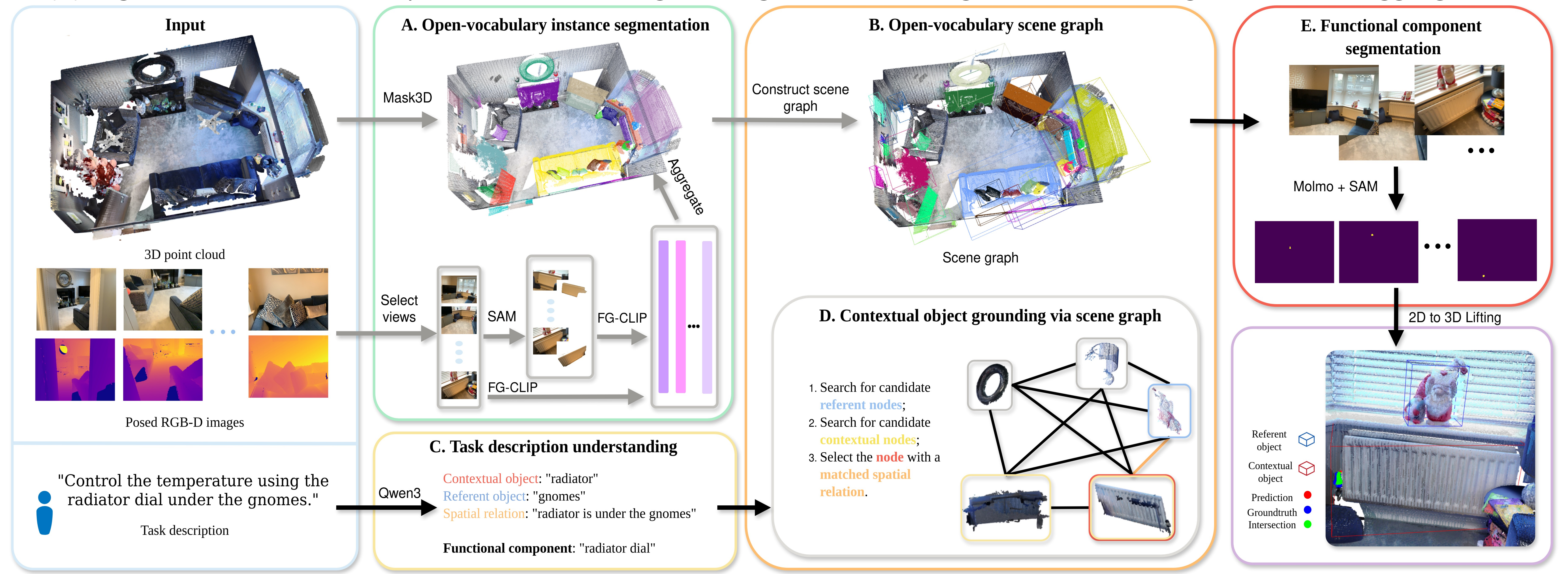
## Methodology

### Stage I: Open-Vocabulary Scene Graph:

- (A) associates open-vocabulary semantics with class-agnostic segmentation.
- (B) builds a scene graph with embeddings of instances and their spatial relations.

### Stage II: Task-Driven Functionality Segmentation:

- (C) analyzes task descriptions using LLM.
- (D) identifies contextual and referent objects based on visual-textual embedding similarities and retrieve images.
- (E) segments functional components on the images and generates 3D segmentation through multi-view aggregation.



## Experiment Results

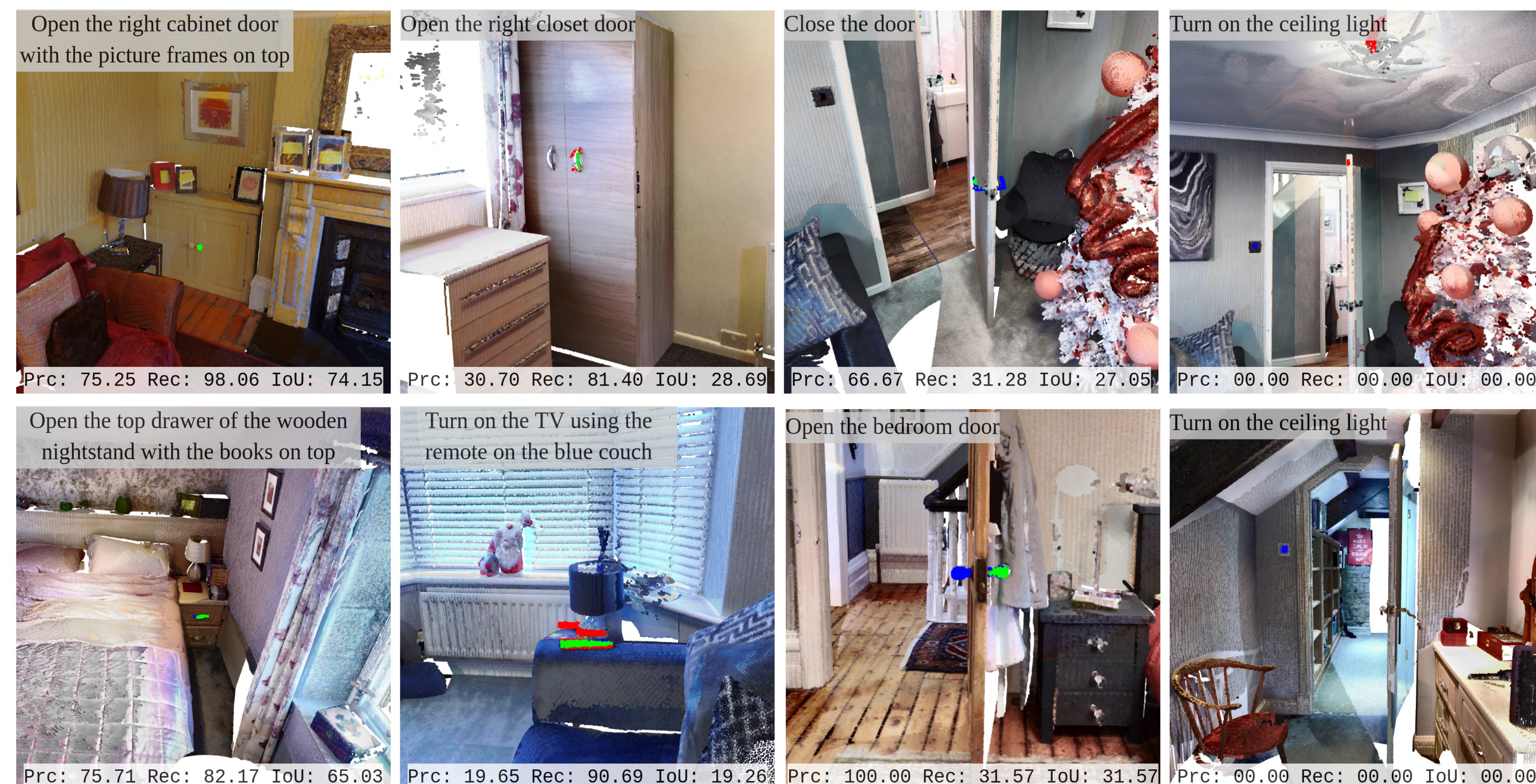


Table I. Performance comparison on SceneFun3D validation split

Methods	mAP	AP50	AP25	mAR	AR50	AR25	mIoU
OpenMask3D [11]	0.2	0.2	0.4	20.3	24.5	27	0.2
OpenIns3D [13]	0	0	0	<b>40.5</b>	<b>46.7</b>	<b>51.5</b>	0.1
LERF [16]	0	0	0	34.2	35.1	36	0
Fun3DU [5]	7.6	16.9	33.3	27.4	38.2	46.7	15.2
Fun3DU [5]†	4.4	10.3	23.4	30.9	42.3	49.7	12.0
<b>T-FunS3D (ours)</b>	<b>8.1</b>	<b>17.8</b>	<b>34.5</b>	23.8	35.8	46.9	<b>15.7</b>

Table II. Runtime comparison at instance and functionality segmentation

Methods	① OV Inst. Segm. (per-scene average)	② Func. Segm. (per-query average)
OpenMask3D [11]	Ⓐ 30s + Ⓑ 720s	-
Fun3DU [5]	1920s	167s
<b>T-FunS3D (ours)</b>	<b>Ⓐ 12s + Ⓑ 580s</b>	<b>78s</b>

- T-FunS3D achieves **30% higher mIoU** with **65% less runtime per task** than the previous SOTA on the SceneFun3D dataset.
- Unlike T-FunS3D, OpenIns3D [13] segments the entire objects rather than parts, leaving a larger AR margin.
- It remains challenging to segment physically unattached functional parts such as light switches.