SA3DIP: Segment Any 3D Instance with Potential 3D Priors

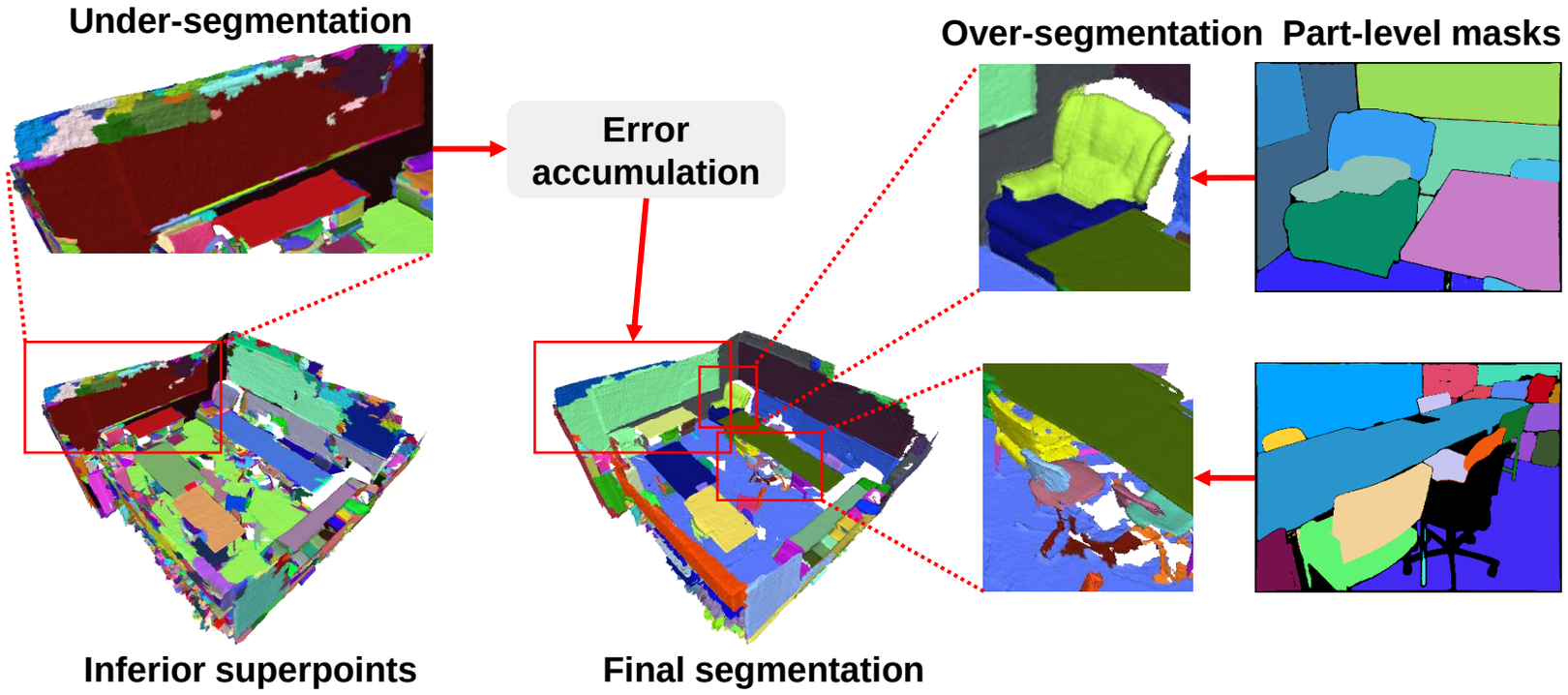
Xi Yang[1], Xu Gu[1], Xingyilang Yin[1*], Xinbo Gao[2]

[1] Xidian University, [2] Chongqing University of Posts and Telecommunications, [*] Corresponding author

Problem Background

Two major defects existing in the pipelines of the current works which utilizing 2D foundation model for **zero-shot 3D instance segmentation**:

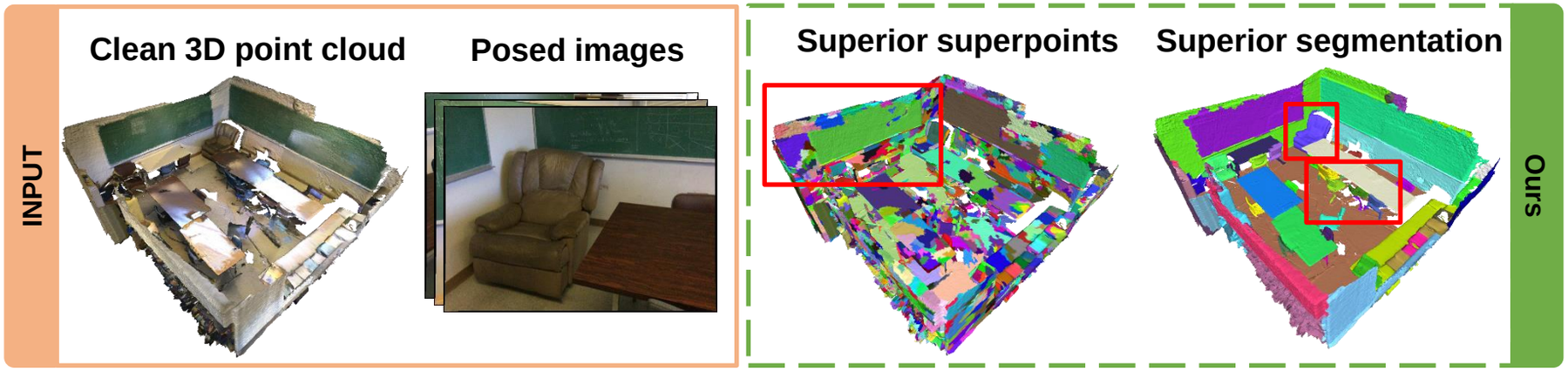
- **Under-segmented 3D primitives and subsequent error accumulation**
- **Part-level over-segment tendency of the 2D foundation segmentator**



Solution: SA3DIP

We therefore **designed two modules accordingly**, trying to alleviate these two major defects:

- **Complementary primitives generation module to generate more accurate and finer-grained primitives to avoid error accumulation**
- **Introducing the 3D space prior for providing instance-aware constraint, which was implemented by a 3D detector**

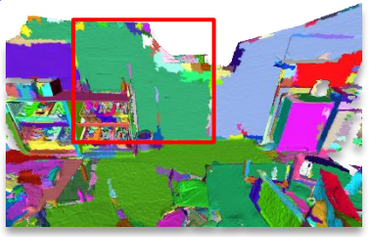


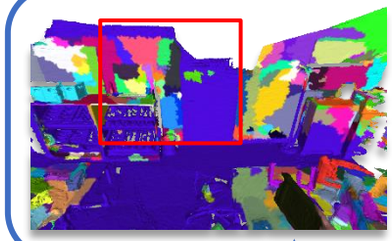
Method Overview

A Complementary Primitives Generation

- Finer-grained primitives $\{U_i\}_{i=1}^{N_U}$

$=$ $w_n \cdot$ Over-seg on geometry $\vec{n}_a$ $\oplus$ $w_c \cdot$ Over-seg on texture c_a

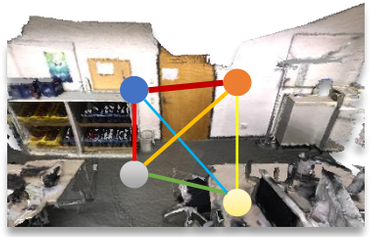




B. Scene Graph Construction

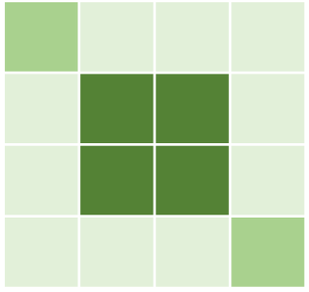
Graph nodes U_i

Edge weights A



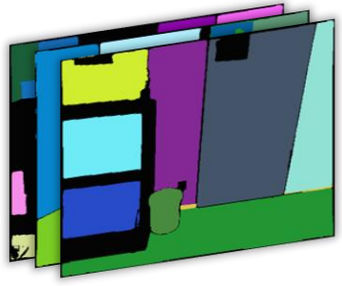
Superpoint graph

Affinity matrix



$A \in \mathbb{R}^{N_U \times N_U}$

Primitive-mask matching



Posed masks $\{S_m\}_{m=1}^{M}$

2D foundation segmentator

Input

Point cloud $P \in \mathbb{R}^{N \times 6}$

Posed RGB-D frames

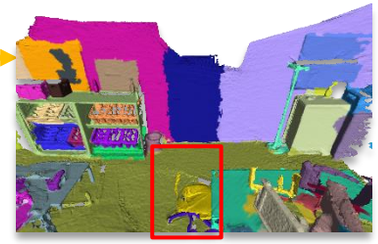
$\{I_m, D_m, K_m, E_m\}_{m=1}^{M}$

C. Region growing & Instance-aware refinement

Merging criteria

- Affinity score A
- Euclidean dis. $\mathbb{D}$
- Dynamic thre. δ_1

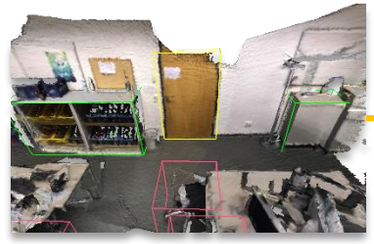
Region growing



Primary seg.

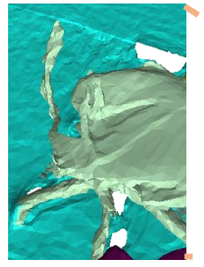
3D detector

3D boxes $\{bb_i\}_{i=1}^{N_{bb}}$



Instance-aware refinement

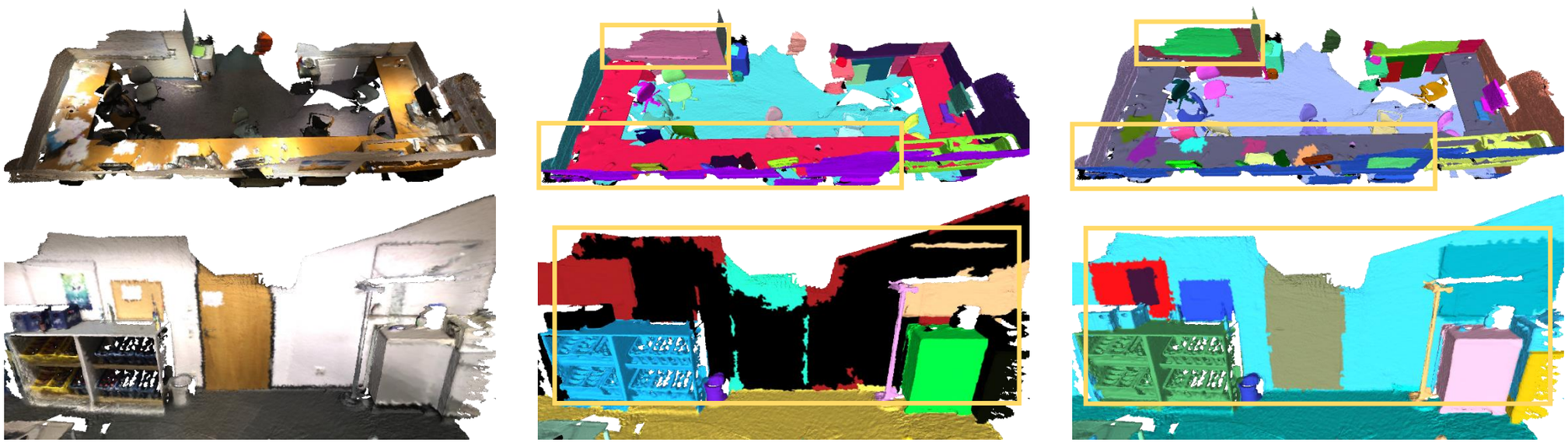
Output



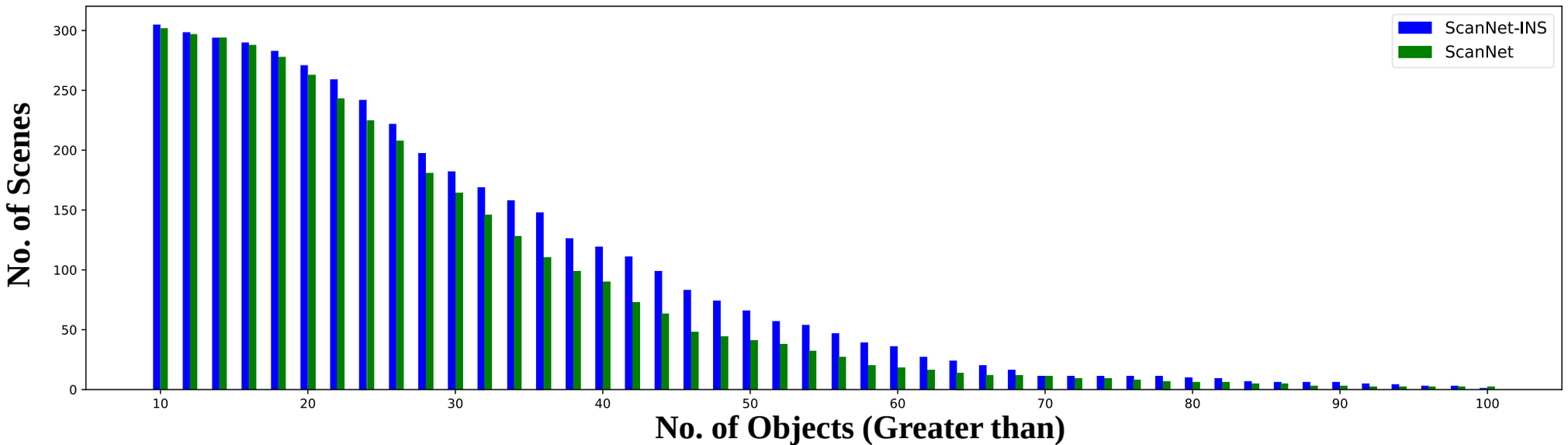
Refined seg.

Proposed Dataset

(a)



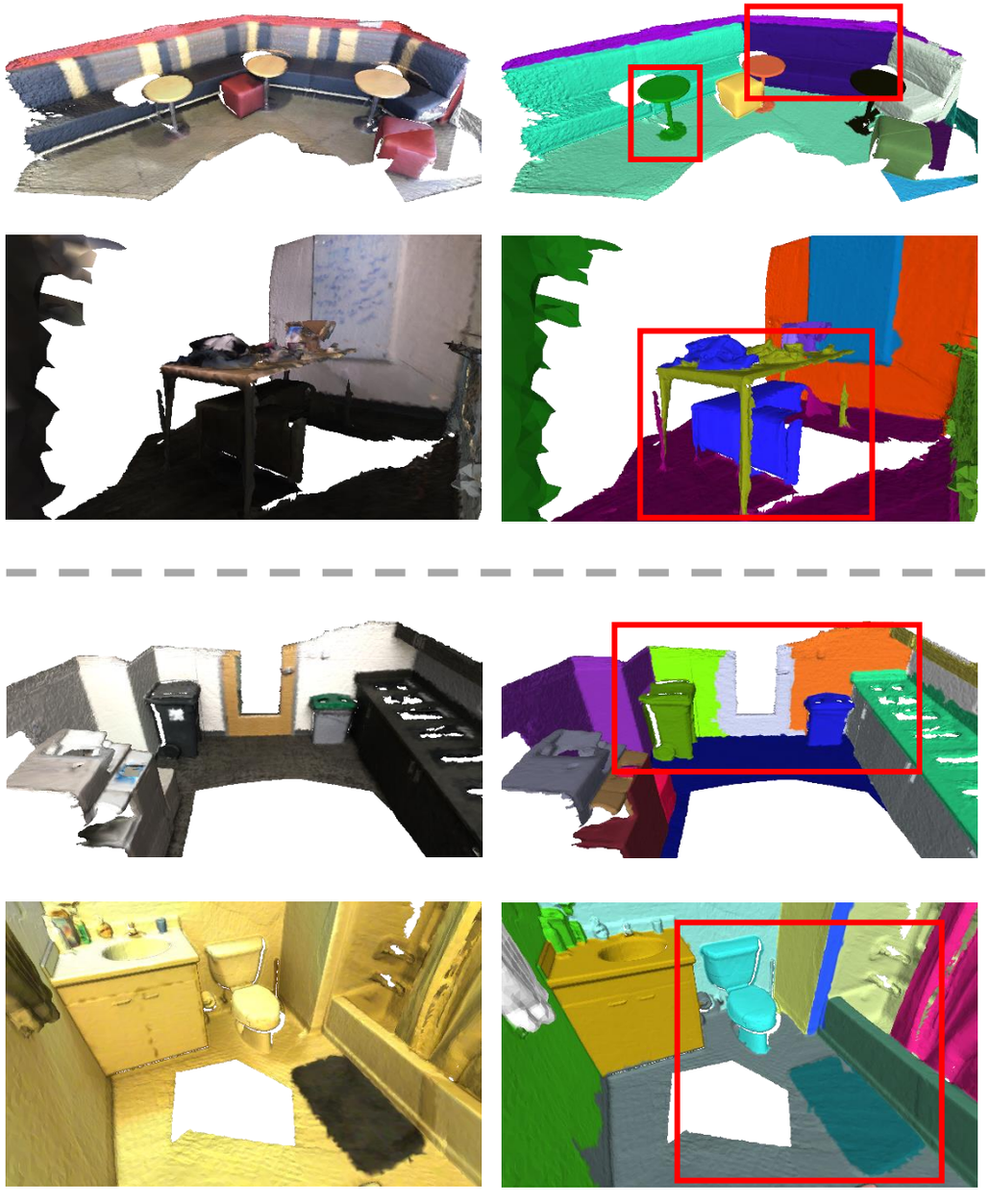
(b)



Results

Method	ScanNetV2			ScanNetV2-INS			ScanNet++		
	mAP	**AP$_{50}$**	**AP$_{25}$**	**mAP**	**AP$_{50}$**	**AP$_{25}$**	**mAP**	**AP$_{50}$**	**AP$_{25}$**
Closed-vocabulary									
Mask3D [2]	31.1	44.9	58.0	29.1	43.9	56.3	9.9	17.3	25.8
Open-vocabulary									
Felzenszwalb [38]	5.0	12.7	38.9	2.8	6.5	24.0	4.1	9.2	25.3
SAM3D [10] (w/o ensemble)	12.4	28.7	57.4	12.5	28.9	57.8	1.1	4.5	15.4
SAM3D [10] (w/ ensemble)	20.1	33.3	52.1	20.0	33.2	52.2	7.2	14.2	29.4
SAM-graph [7]	24.1	40.3	65.9	23.1	39.5	64.1	12.9	25.3	43.6
SAI3D [8]	30.8	50.5	70.6	28.9	49.2	69.7	17.1	31.1	49.5
SAMPro3D [9]	33.7	56.2	75.3	32.5	54.8	73.4	18.9	33.7	51.6
SA3DIP (ours)	**41.6**	**64.6**	**81.3**	**36.1**	**58.6**	**76.3**	**21.4**	**36.4**	**53.6**

w_n	w_c	3D Space Prior	ScanNetV2			ScanNetV2-INS		
			mAP	**AP$_{50}$**	**AP$_{25}$**	**mAP**	**AP$_{50}$**	**AP$_{25}$**
1	0	×	30.8	50.5	70.6	28.9	49.2	69.7
0	1	×	10.4	18.1	32.5	9.5	17.0	31.1
0.4	0.6	×	27.3	47.4	69.8	25.6	46.3	69.4
0.96	0.04	×	29.3	49.2	70.5	27.4	48.3	70.4
1	0	✓	40.8	63.6	80.7	35.9	57.8	75.4
0	1	✓	12.7	22.1	37.2	11.0	19.7	34.1
0.4	0.6	✓	39.1	62.7	80.2	33.5	56.3	75.0
0.96	0.04	✓	**41.6**	**64.6**	**81.3**	**36.1**	**58.6**	**76.3**



Input **Ours**