Qualcomm

CVPR Nashville JUNE 11-15, 2025

Any3DIS

Class-Agnostic 3D Instance Segmentation by 2D Mask Tracking

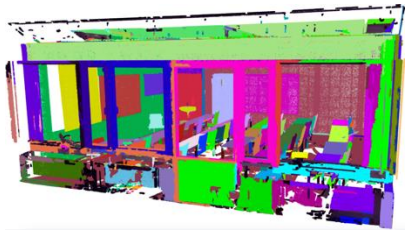
https://any3dis.github.io/

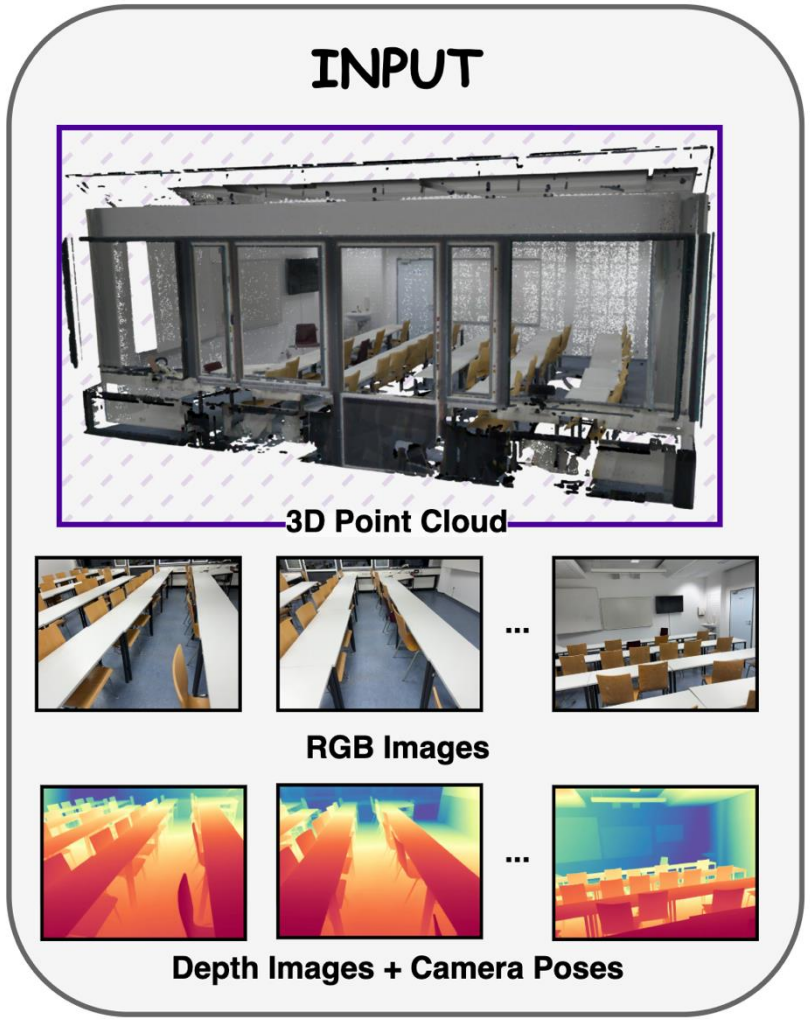
Phuc Nguyen **Minh Luu** **Anh Tran** **Cuong Pham** **Khoi Nguyen**

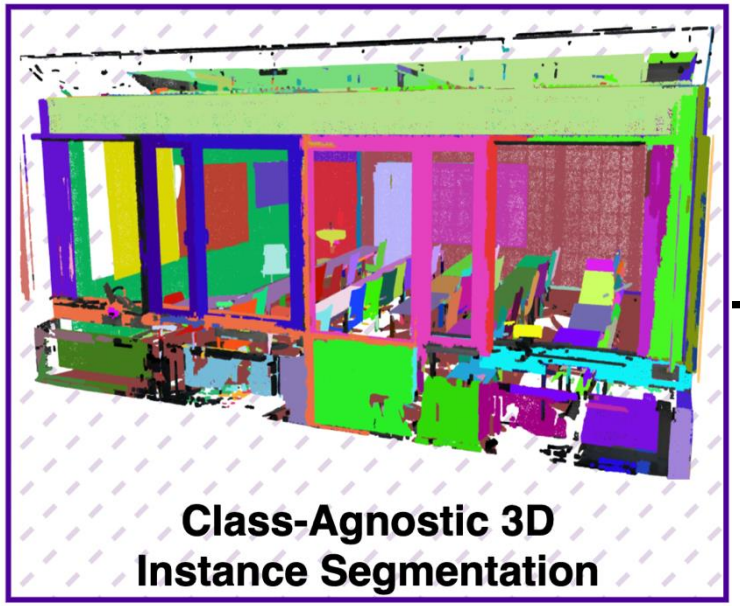


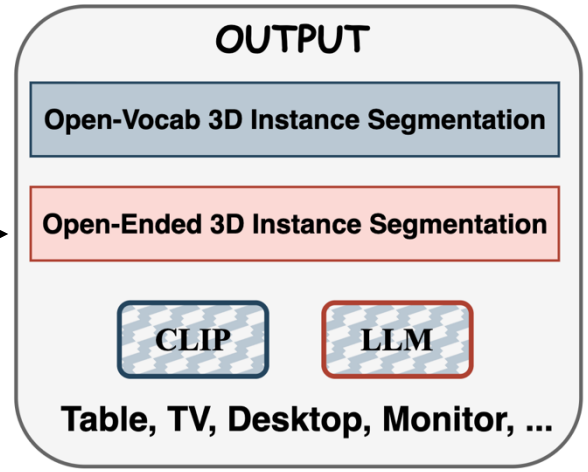


Our proposed Any3DIS



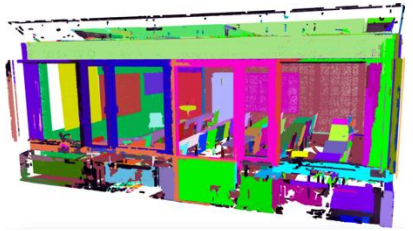






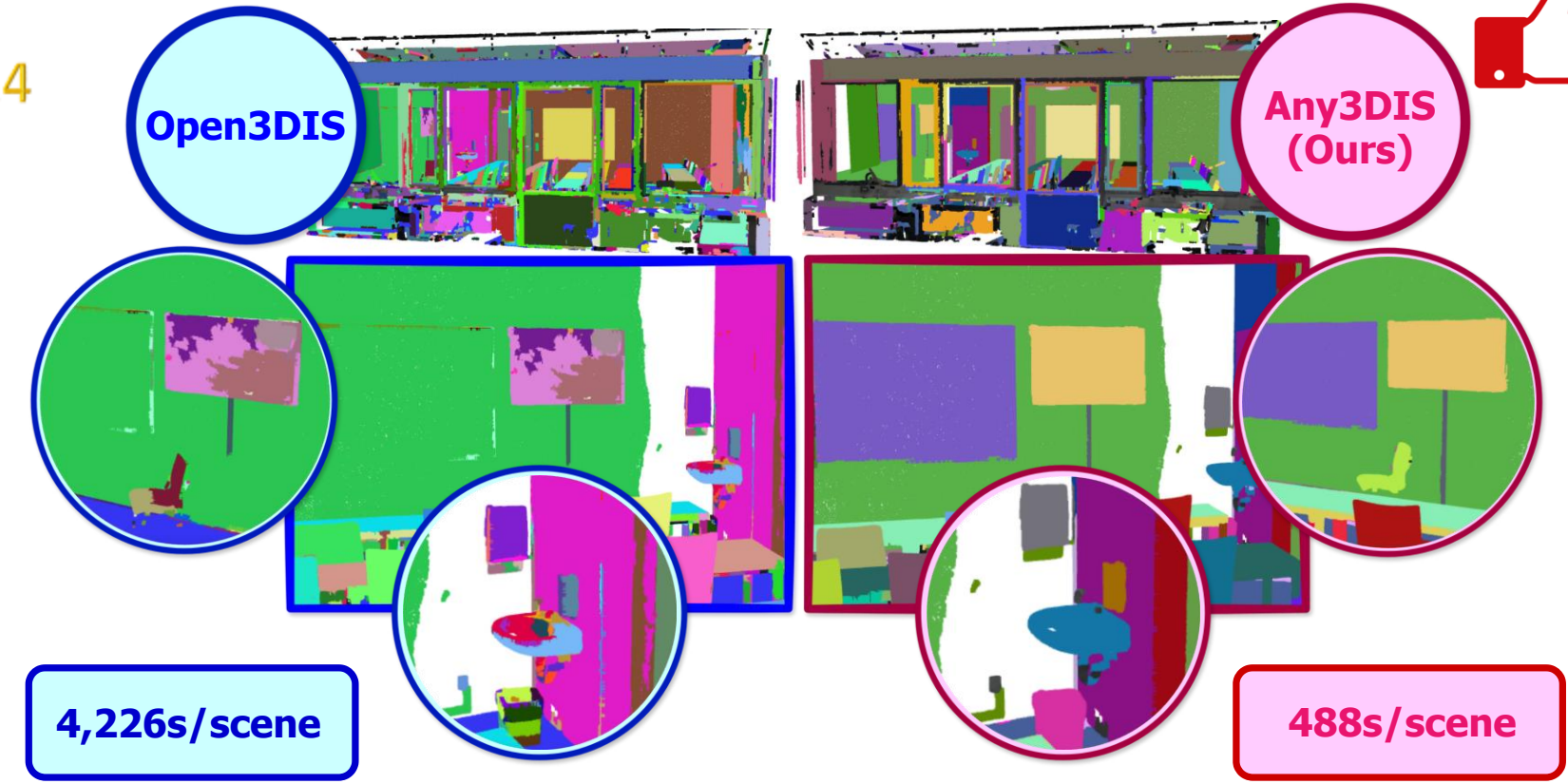
Given a **3D point cloud** with the corresponding **RGB-D sequences**, Any3DIS generates a set of **class-agnostic 3D proposals** and a **featurized point cloud** for open-vocabulary/open-ended queries

Motivations



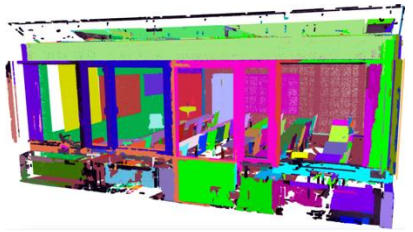
- **Any3DIS outperforms existing training-free SOTA methods on ScanNet200 and ScanNet++**
  - ✓ **about ~1.2x in average precision**
  - ✓ **about ~10x in speed**

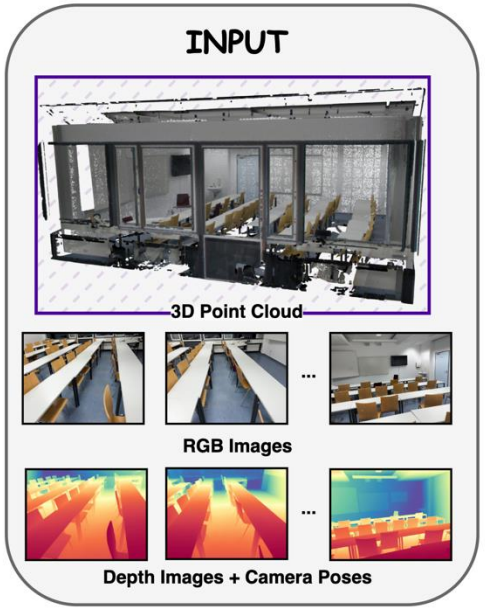
CVPR 2024



CVPR 2025

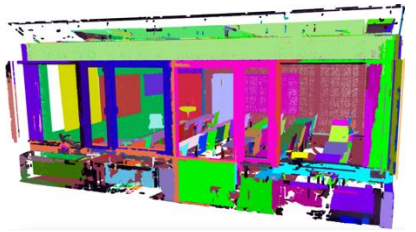
Our proposed Any3DIS: Step1

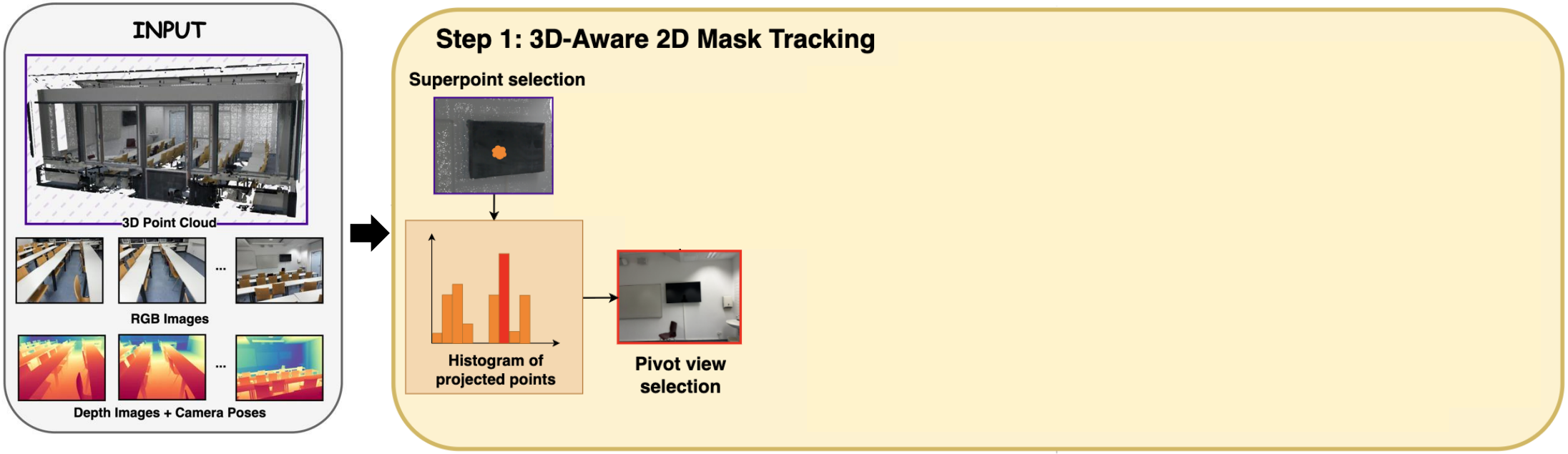




Step 1: 3D-Aware 2D Mask Tracking

Our proposed Any3DIS: Step1



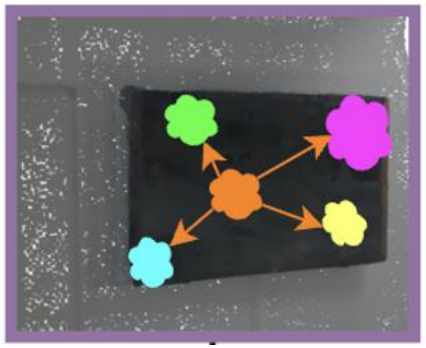


Camera Calibration

$$\rho_t^l = \Pi(\mathbf{S}_l, \mathbf{K}, \mathbf{E}_t, \mathbf{D}_t).$$

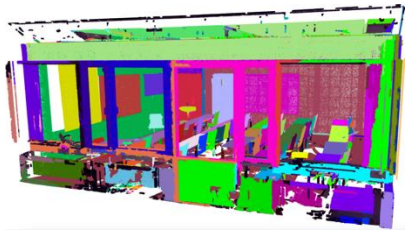
Weighted Distribution

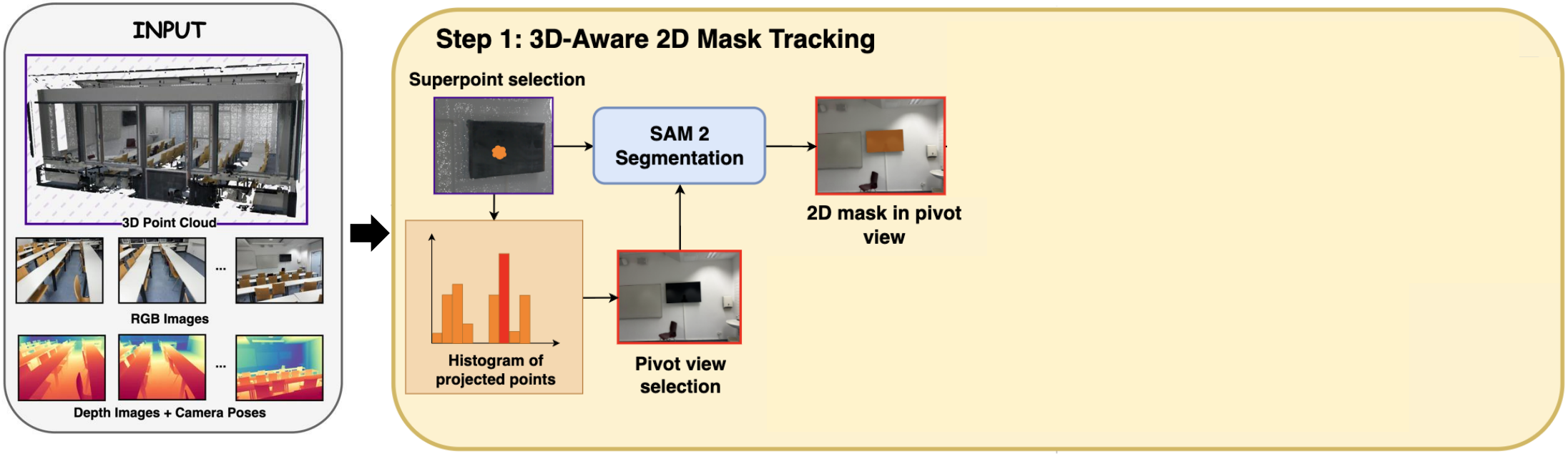
$$s_t^l = \frac{\sum_{k \in \mathrm{KNN}(\mathbf{S}_l)} \left|\rho_t^k\right| / N_{\mathbf{S}_k}}{\kappa}.$$

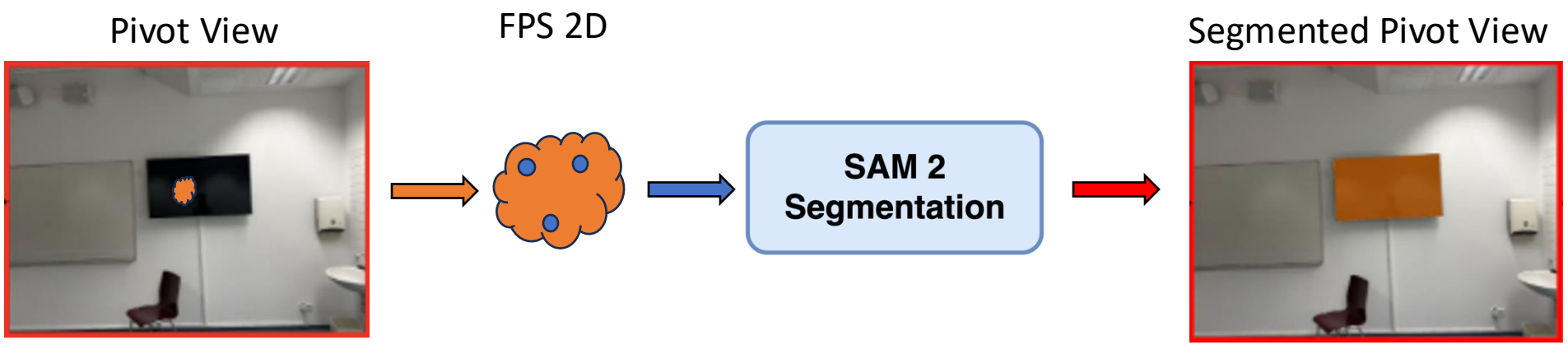


$$\mathrm{pivot}(l) = \mathrm{argmax}_t \, \psi_t^l.$$

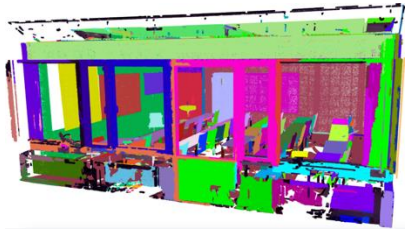
Our proposed Any3DIS: Step1

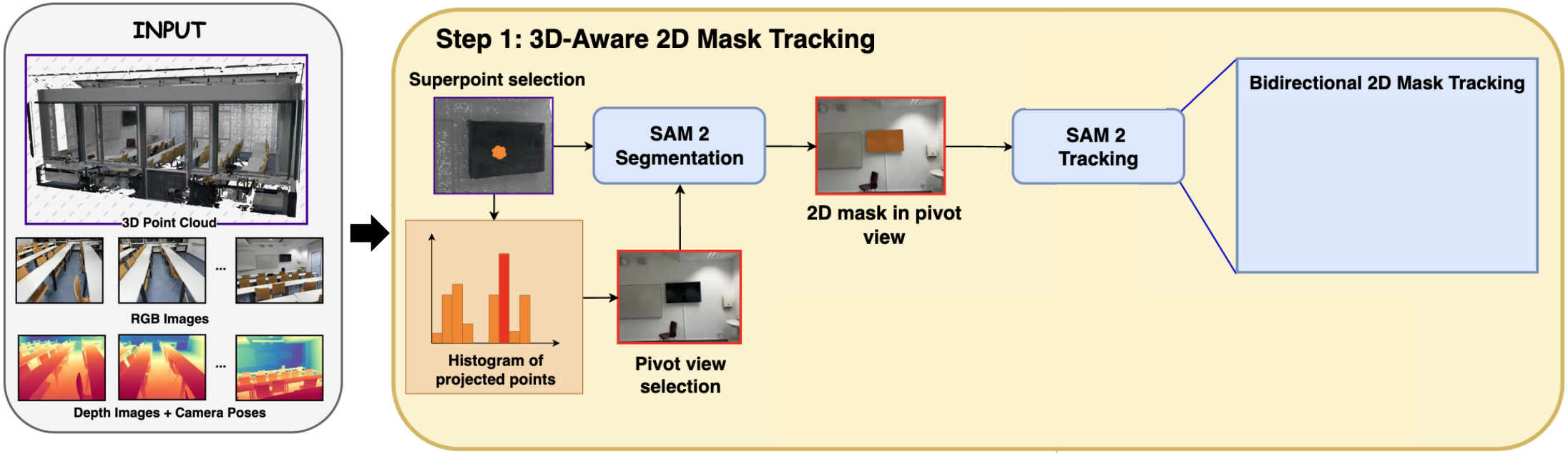




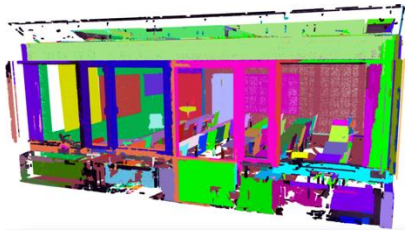


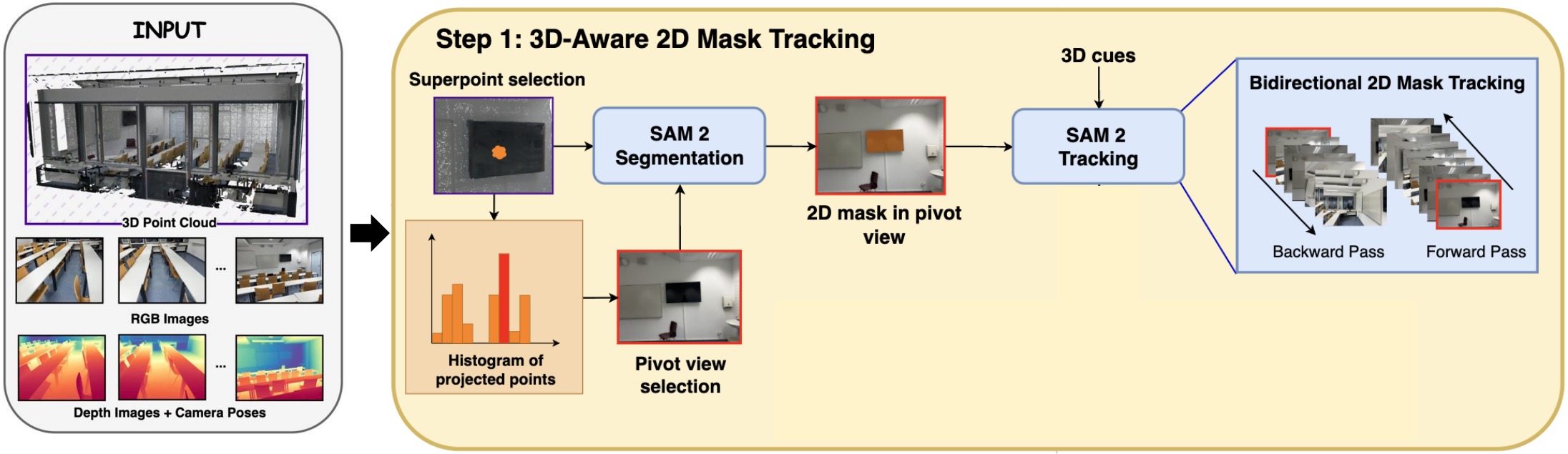
Our proposed Any3DIS: Step1

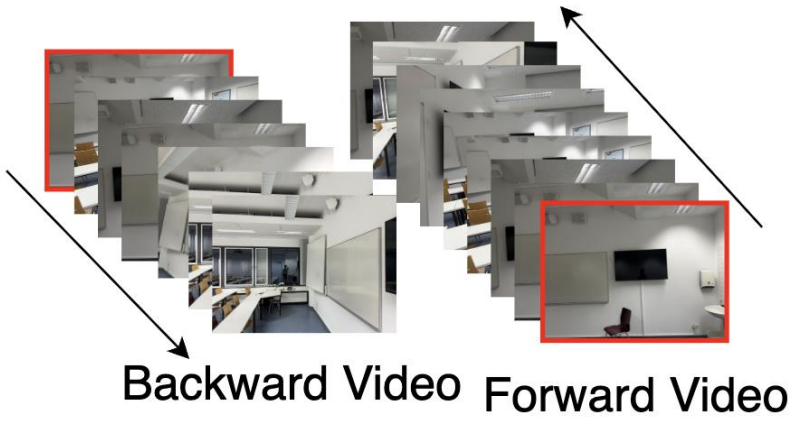


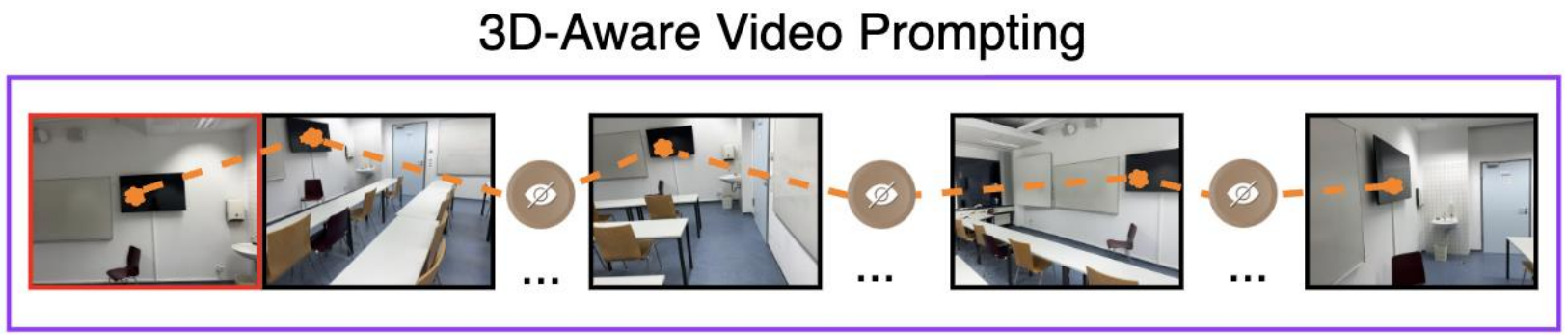


Our proposed Any3DIS: Step1

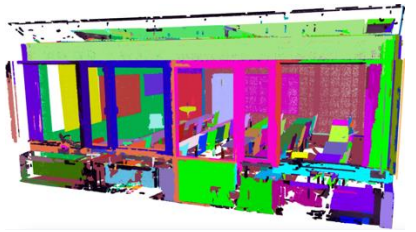


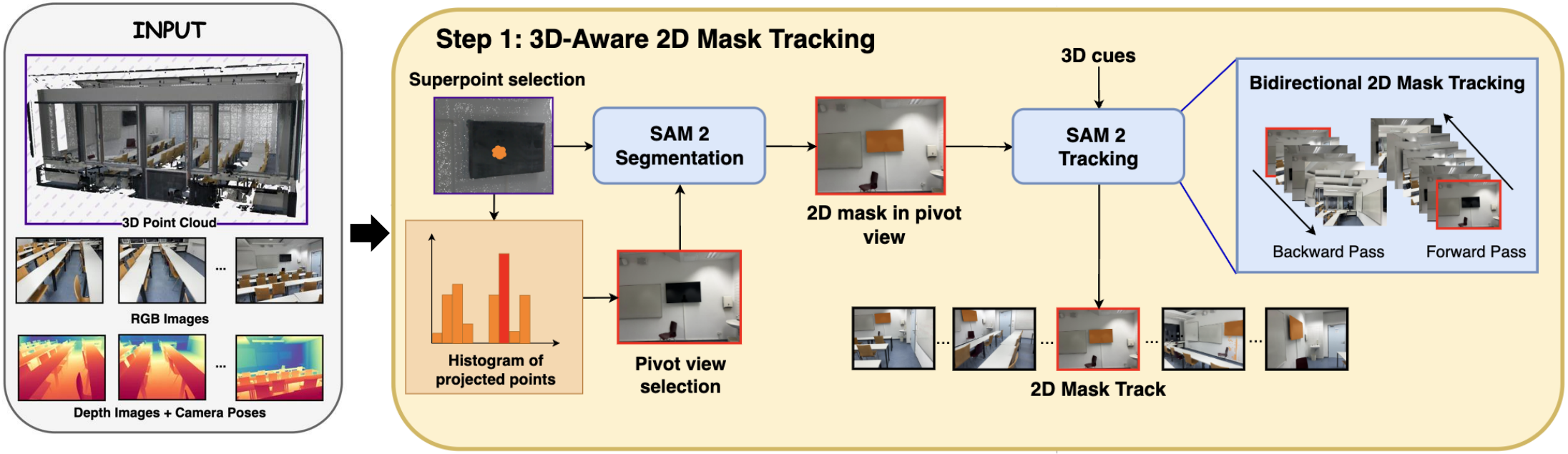




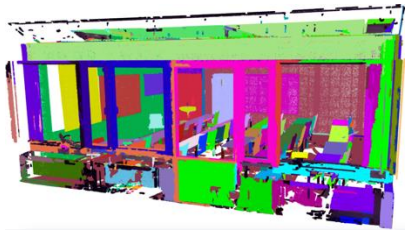


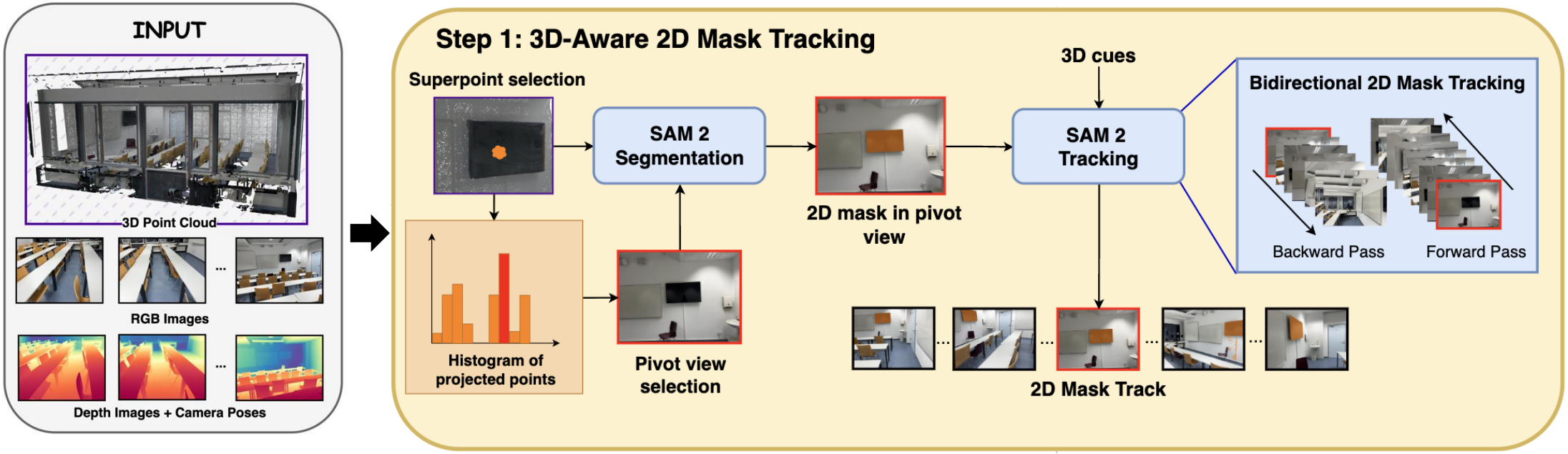
Our proposed Any3DIS: Step1



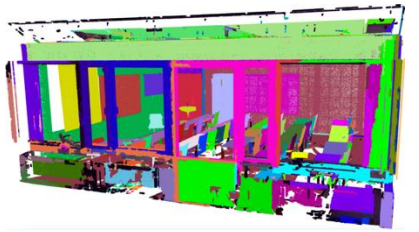


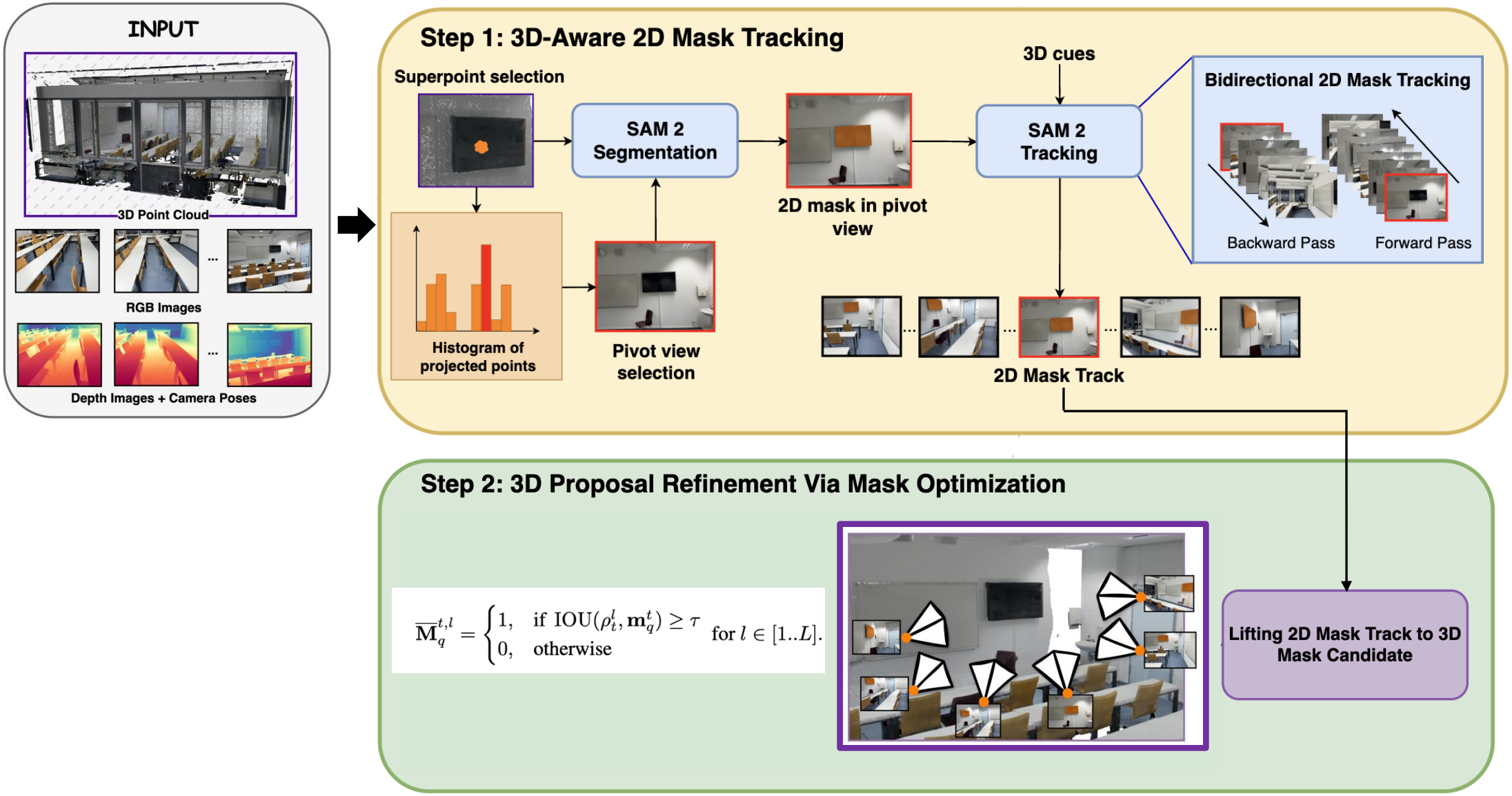
Our proposed Any3DIS: Step2



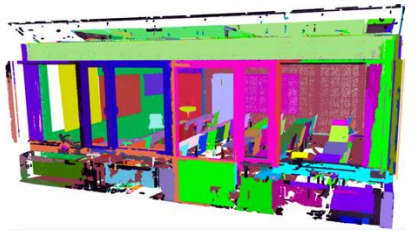


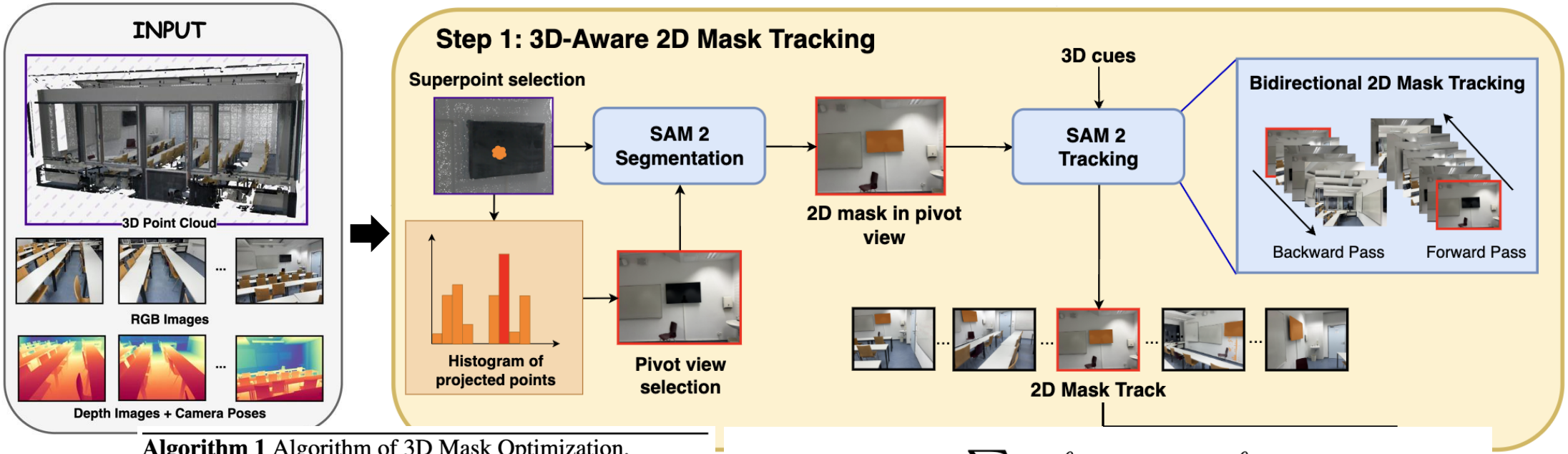
Our proposed Any3DIS: Step2





Our proposed Any3DIS: Step2



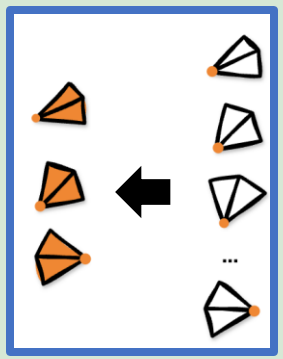


Algorithm 1 Algorithm of 3D Mask Optimization.

Input: Visibility vector $\overline{\mathbf{M}}_q^t$ in Eq. (3)

1: **initialize:** Objective value $\mathbf{C}_0 = \mathbf{0}$ and current solution $\theta_0 = \mathbf{0}$.
2: **for** $t = 1$ **to** T **do**
3: Option 1 - Retain current solution: $\theta_t^1 = \theta_{t-1}, \quad \mathbf{C}_t^1 = \mathbf{C}_{t-1}$
4: Option 2 - Add all superpoints in $\overline{\mathbf{M}}_q^t$ to current solution: $\theta_t^2 = \theta_{t-1} \vee \overline{\mathbf{M}}_q^t, \quad \mathbf{C}_t^2 = \mathcal{L}(\theta_t)$ in Eq. (4)
5: **If** $\mathbf{C}_t^1 > \mathbf{C}_t^2$ **then:** $\theta_t = \theta_t^1, \mathbf{C}_t = \mathbf{C}_t^1$ **else:** $\theta_t = \theta_t^2, \mathbf{C}_t = \mathbf{C}_t^2$
6: **end for**
7: **return** θ_T

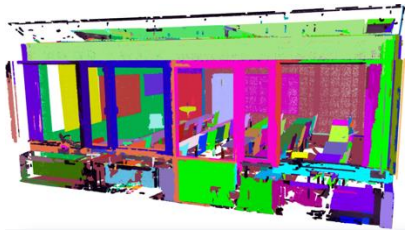
$$\max_{\theta} \mathcal{L}(\theta) = \max_{\theta} \sum_{t} \underbrace{\rho_t^{\theta} \otimes \mathbf{m}_t^q}_{\text{(A) Inside 2D Mask}} - \underbrace{\rho_t^{\theta} \otimes (\mathbf{1}_{H \times W} - \mathbf{m}_t^q)}_{\text{(B) Outside 2D Mask}},$$

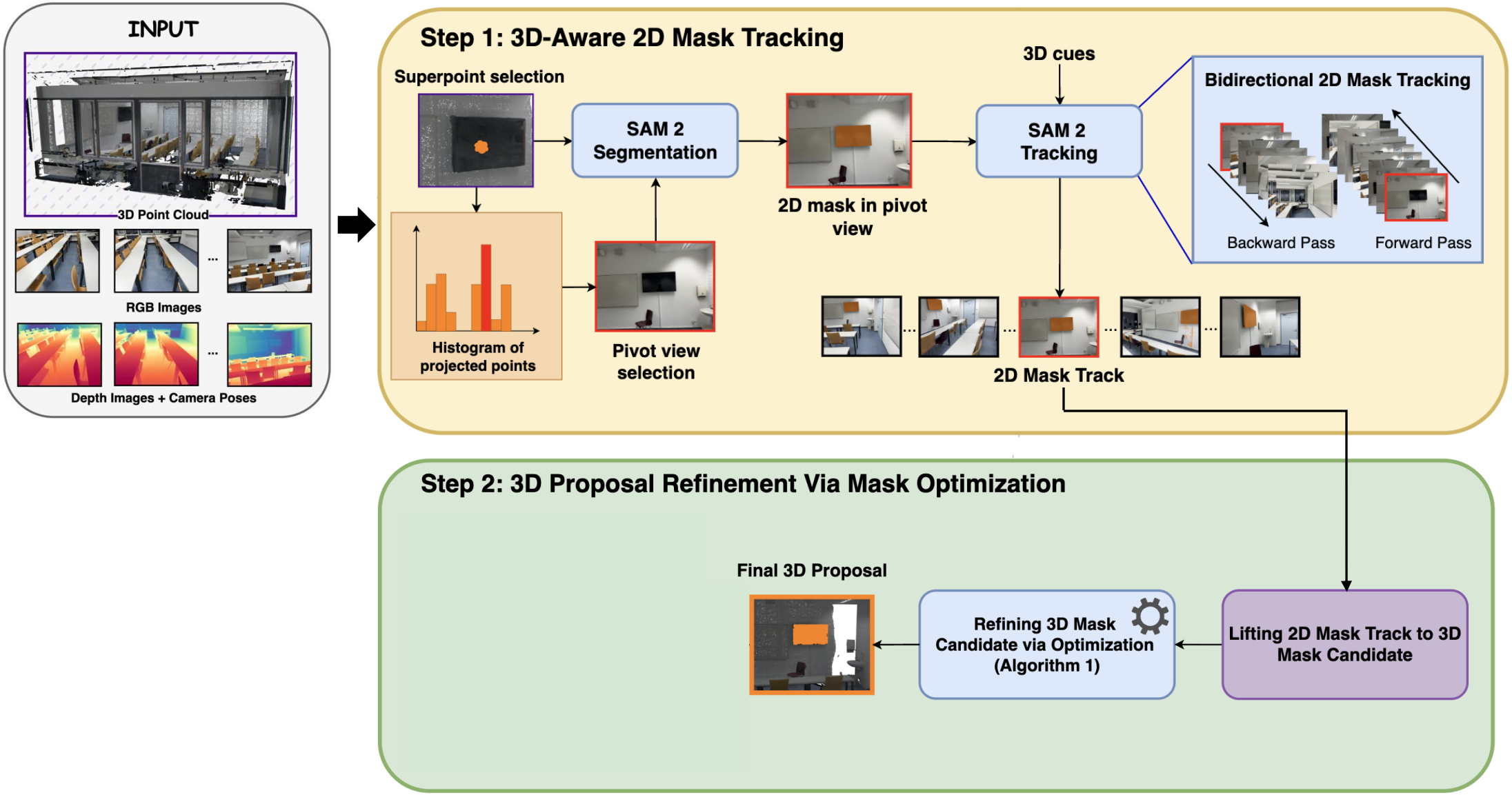


Refining 3D Mask Candidate via Optimization (Algorithm 1)

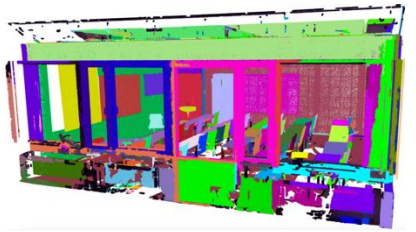
Lifting 2D Mask Track to 3D Mask Candidate

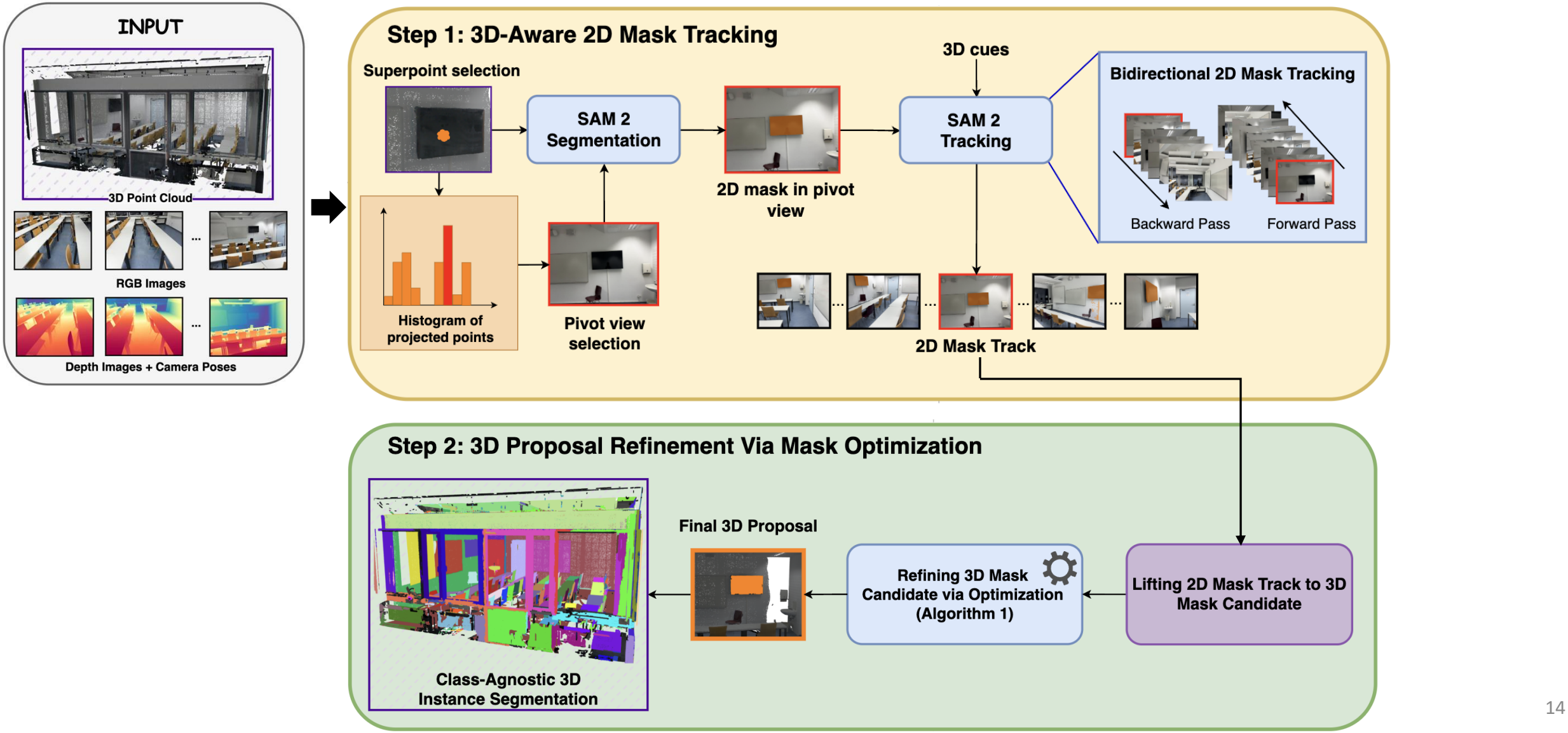
Our proposed Any3DIS: Step2



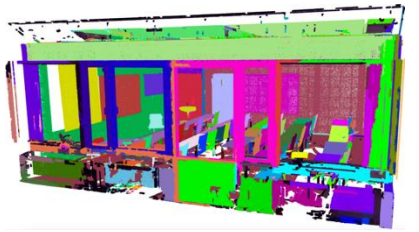


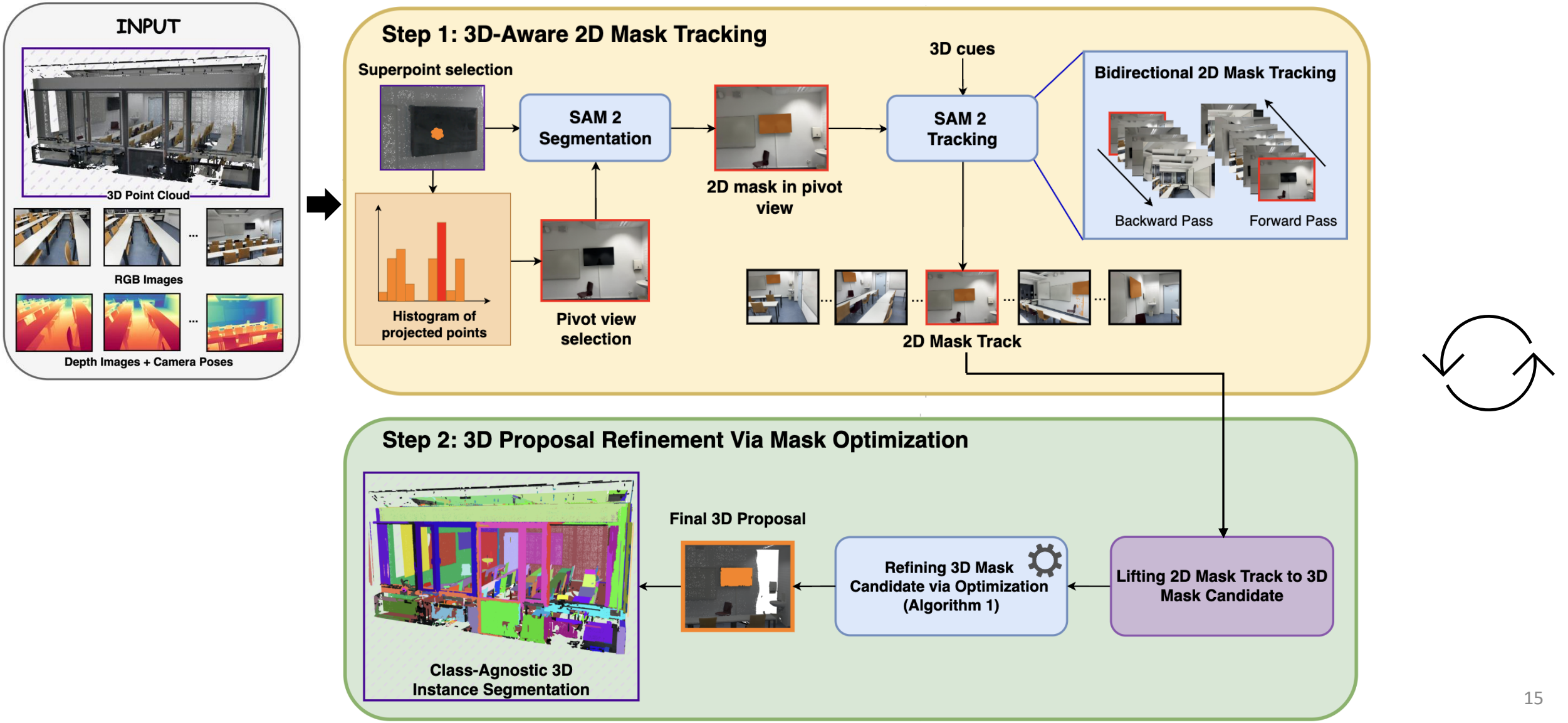
Our proposed Any3DIS: Step2



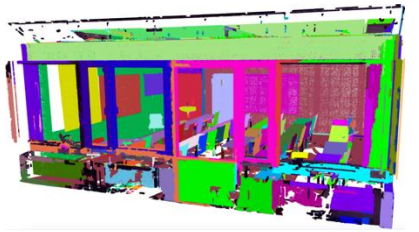


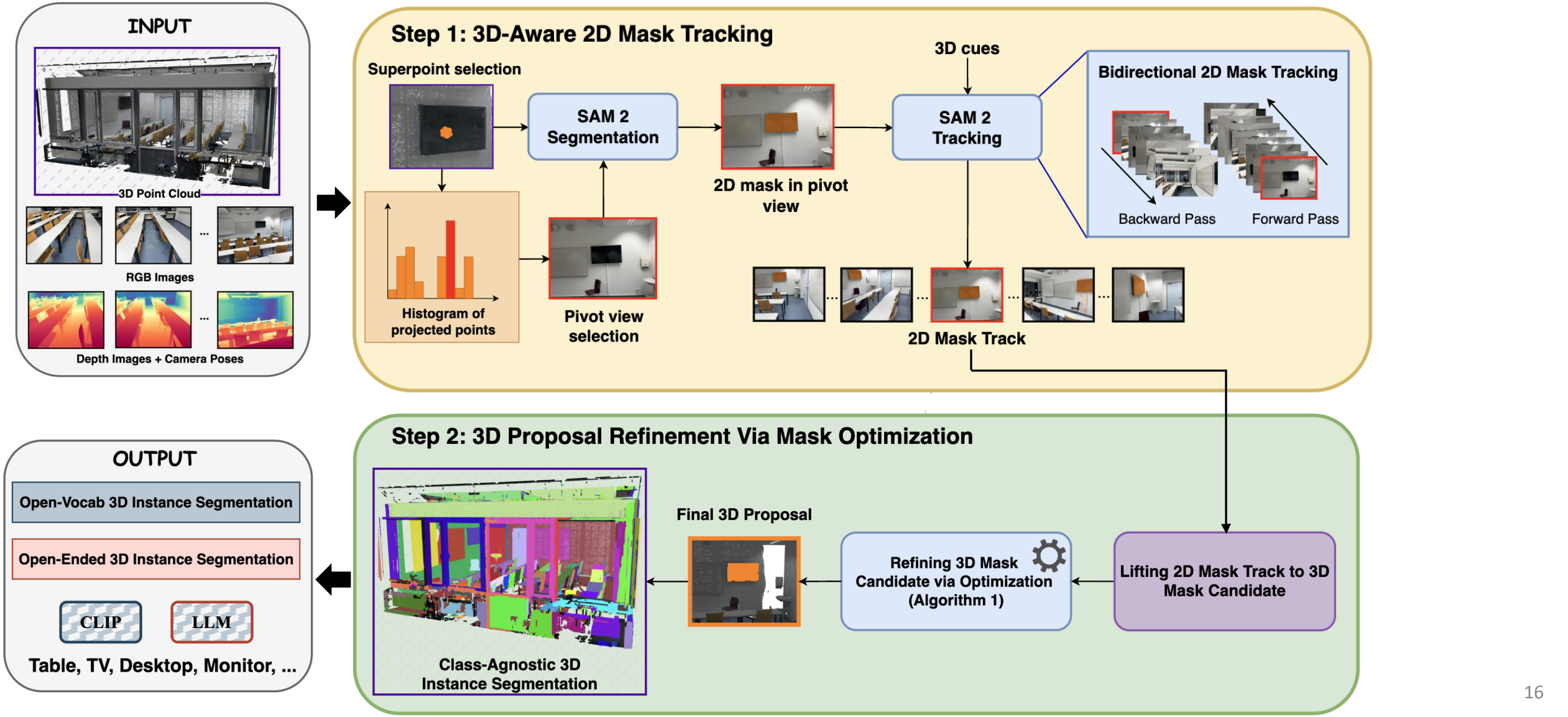
Our proposed Any3DIS: Step2



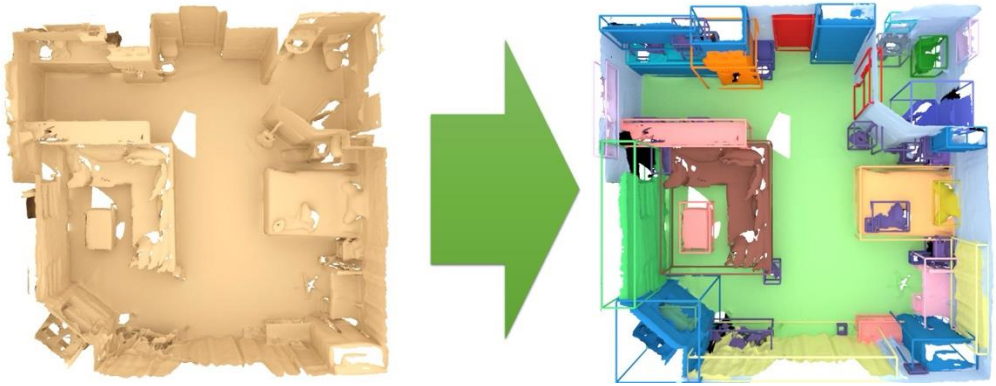


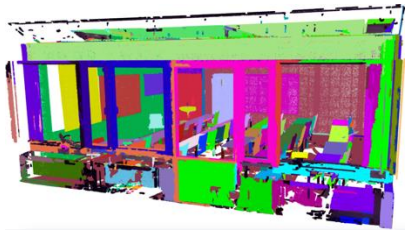
Our proposed Any3DIS





Result on ScanNet200





Method	2D Segmenter	AP	AP_{50}	AP_{25}
OVIR-3D [17]	SAM-HQ	14.4	27.5	38.8
Open3DIS [19]		31.5	45.3	51.1
Open3DIS [19]	SAM2-L	26.7	40.2	53.8
Any3DIS (ours)		**32.5**	**45.2**	**55.0**
ISBNet [29]	3D	40.2	50.0	54.6
Open3DIS [19]	3D + SAM-HQ	41.5	**51.6**	**56.3**
Any3DIS (ours)	3D + SAM2-L	**42.5**	51.2	54.5

Table 2. **Results of class-Agnostic 3DIS on ScanNet200,** benchmarking on 198 classes

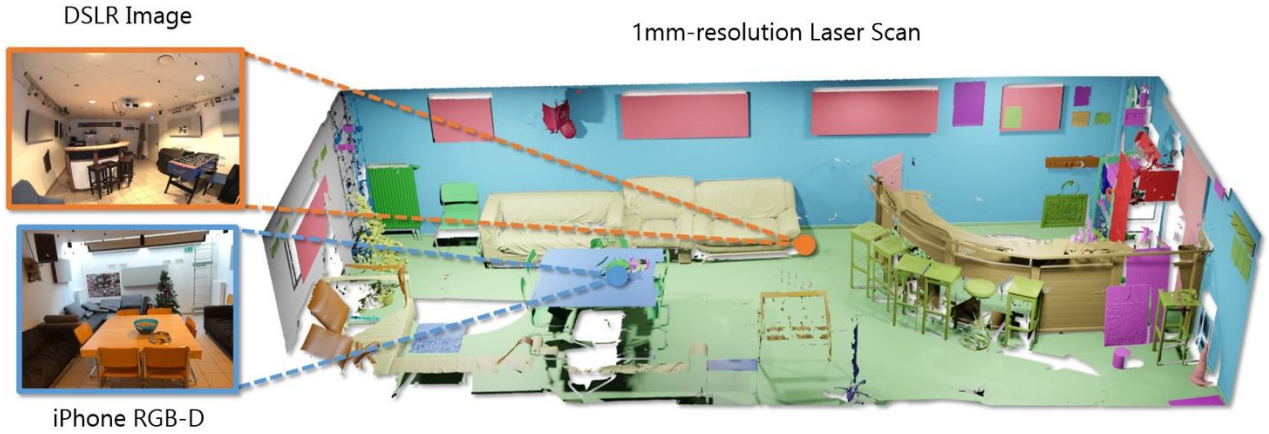
Method	Segmenter	AP	AP_h	AP_c	AP_t
Search3D [28]	Mask3D [25]	14.3	16.1	13.6	12.9
OpenMask3D [27]		15.4	17.1	14.1	14.9
SAI3D [35]	SAM-HQ	12.7	12.1	10.4	16.2
OVIR-3D [17]		11.2	12.6	11.5	10.3
Open3DIS [19]	SAM-HQ	18.2	**18.9**	16.5	19.2
Any3DIS (ours)	SAM2-L	**19.1**	17.5	**17.2**	**23.3**
Open3DIS [19]	3D + SAM-HQ	23.7	**27.8**	21.2	21.8
Any3DIS (ours)	3D + SAM2-L	**25.8**	27.4	**23.8**	**26.4**

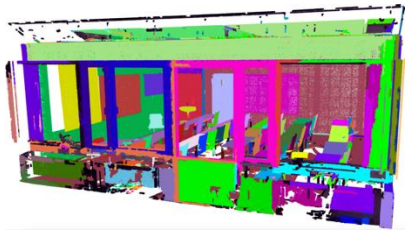
Table 4. **Results of Open-Vocabulary 3DIS on ScanNet200**, benchmarking on 198 classes.

Method	Setting	AP	AP_h	AP_c	AP_t
Large-Vocab [18]	21K+ classes	8.5	9.9	7.2	8.3
Image-Tagging [18]	RAM++	10.7	11.6	11.0	9.3
Mask-Wise [18]	OSM	14.4	18.9	13.5	10.2
Point-Wise [18]		16.0	**20.0**	14.3	13.2
Any3DIS (ours)	OSM	**19.1**	18.2	**16.9**	**22.8**

Table 6. **Results of Open-Ended 3DIS on ScanNet200**, benchmarking on 198 instance classes.

Result on ScanNet++





Method	2D Segmenter	AP	AP_{50}	AP_{25}
Segment3D [12]	SAM-HQ	19.0	29.7	41.6
SAM-Graph [10]		15.3	27.2	44.3
SAM3D [33]		8.3	17.5	33.7
SAMPro3D [31]		16.9	31.7	48.6
SAI3D [35]		17.1	31.1	**49.5**
Open3DIS [19]		20.7	**38.6**	47.1
Open3DIS [19]	SAM2-L	17.9	30.4	39.7
Any3DIS (ours)		**22.2**	35.8	47.0

Table 1. **Results of Class-Agnostic 3DIS on ScanNet++,** benchmarking on 1554 classes

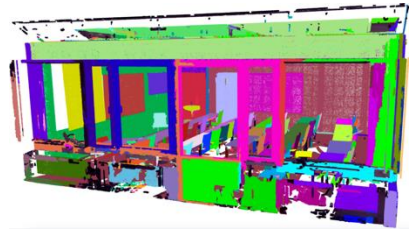
Method	2D Segmenter	AP	AP_{50}	AP_{25}
OVIR-3D [17]	SAM-HQ	3.6	5.7	7.3
Segment3D [12]		10.1	17.7	20.2
Open3DIS [19]		11.9	18.1	21.7
Any3DIS (ours)	SAM2-L	**12.9**	**19.0**	**21.9**

Table 3. **Results of Open-Vocabulary 3DIS on ScanNet++,** benchmarking a subset of 100 classes.

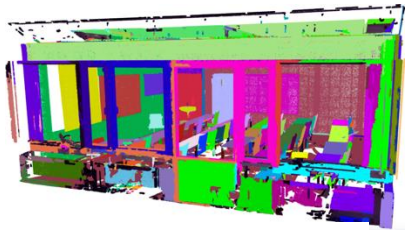
Method	Setting	AP	AP_{50}	AP_{25}
Large-Vocab [18]	21K+ classes	7.3	11.9	15.2
Image-Tagging [18]	RAM++	9.1	15.5	19.1
Mask-Wise [18]	OSM	16.3	24.8	29.0
Point-Wise [18]		18.4	29.4	33.6
Any3DIS (ours)	OSM	**20.1**	**30.4**	**35.5**

Table 5. **Results of Open-Ended 3DIS on ScanNet++,** benchmarking a subset of 100 classes.

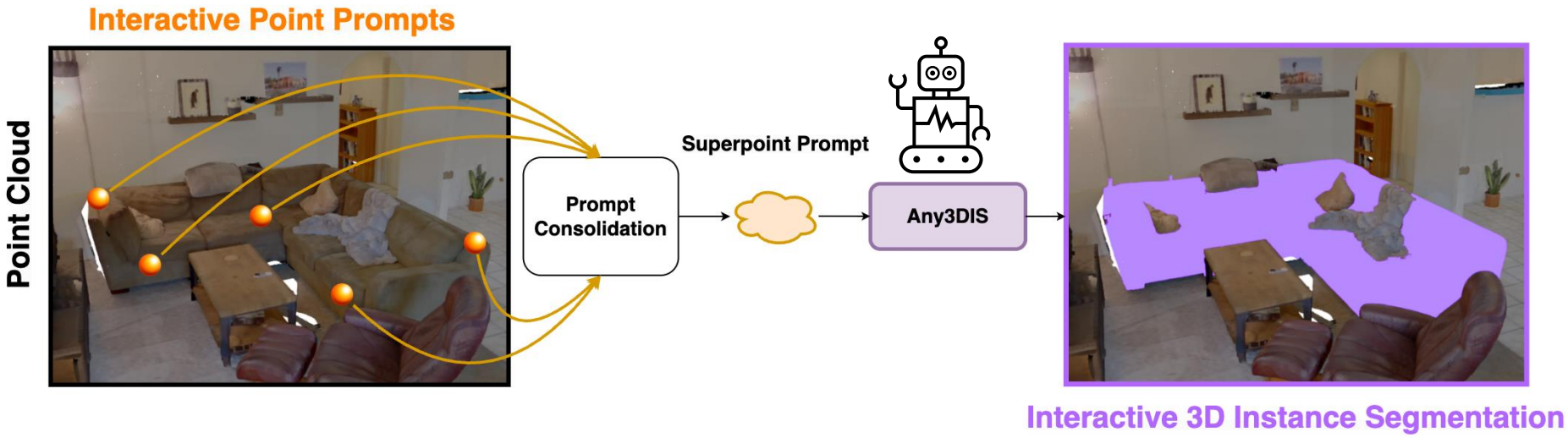
Qualitative Any3DIS

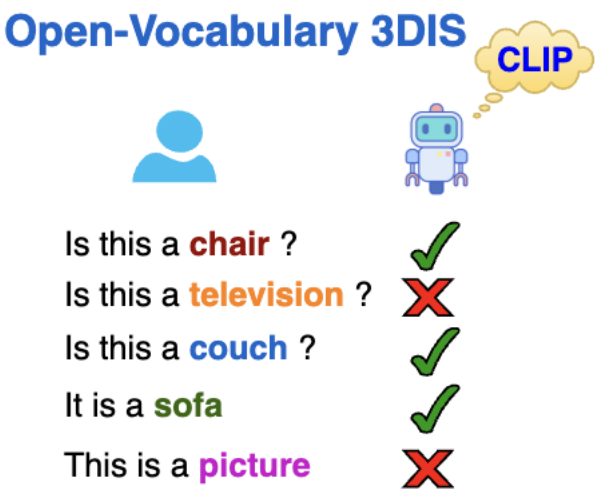


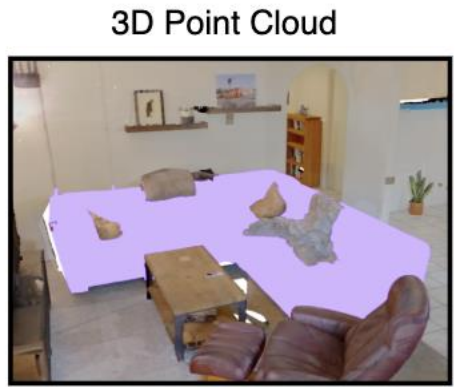
Application Any3DIS

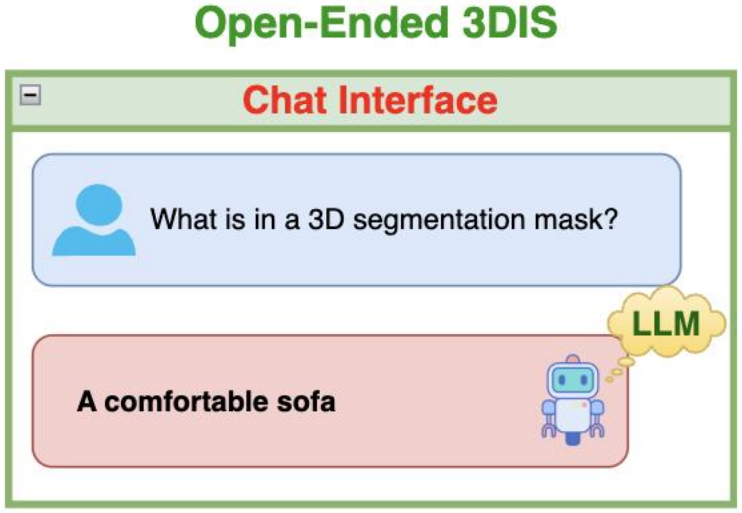


Embodied AI

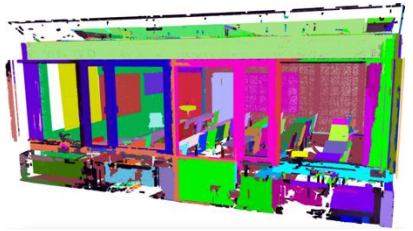






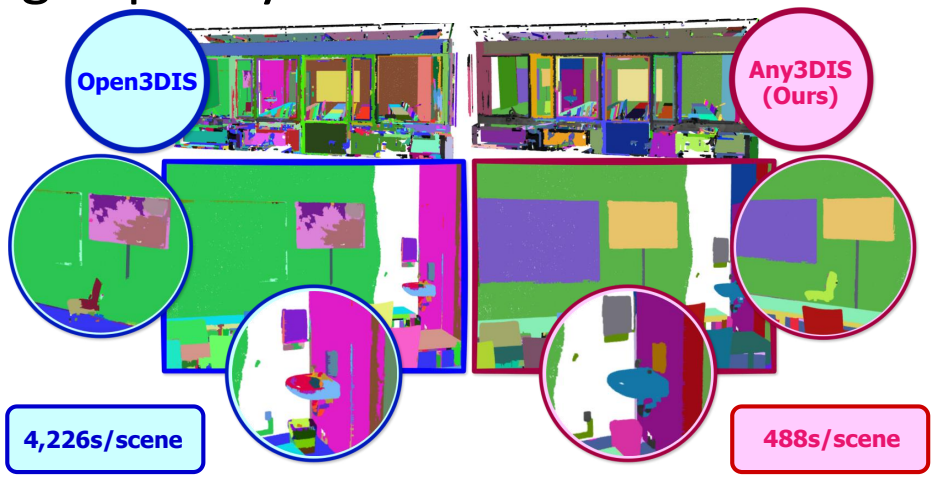


Summary



- We introduce **Any3DIS** to address the class-agnostic 3D instance segmentation task with a novel strategy to generate high-quality 3D proposals from pretrained 2D model.

- Our approach achieves **state-of-the-art performance** on 2 different indoor datasets.



- Future research on adapting **Any3DIS** for other 3D representation such as 3D Gaussian Splatting would be an interesting exploration.

https://any3dis.github.io/

Any3DIS

Thank you for your interest