Living Scenes

Creating and updating representations of evolving indoor scenes

Iro Armeni

Long-Term Perception for Autonomy in Dynamic Human-shared Environments: What Do Robots Need? – IROS 2024

Stanford | Gradient Spaces
Civil and Environmental Engineering

What are Living Scenes?

Living Scenes

Buildings are like living organisms, i.e., they evolve.

How can we realistically maintain an evolving representation throughout their lifespan?

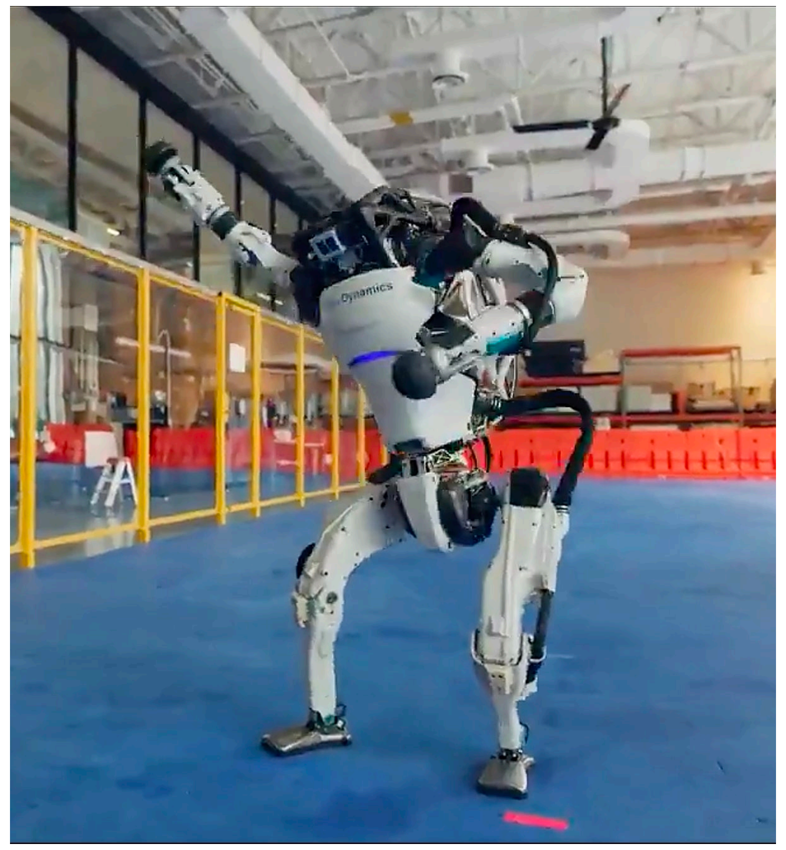
Image Credit: https://tii.imgix.net/production/articles/9001/60cb3783-e4da-468b-b75f-227dd859e0eb.gif?auto=compress&fit=crop&gif-q=50

[1] Fernanda Leite. Buildings as living things. https://cockrell.utexas.edu/news/archive/9042-buildings-as-living-things-6-questions-with-fernanda-leite, 2020

Agents in the Wild

Understanding of Surrounding Environment

Navigating



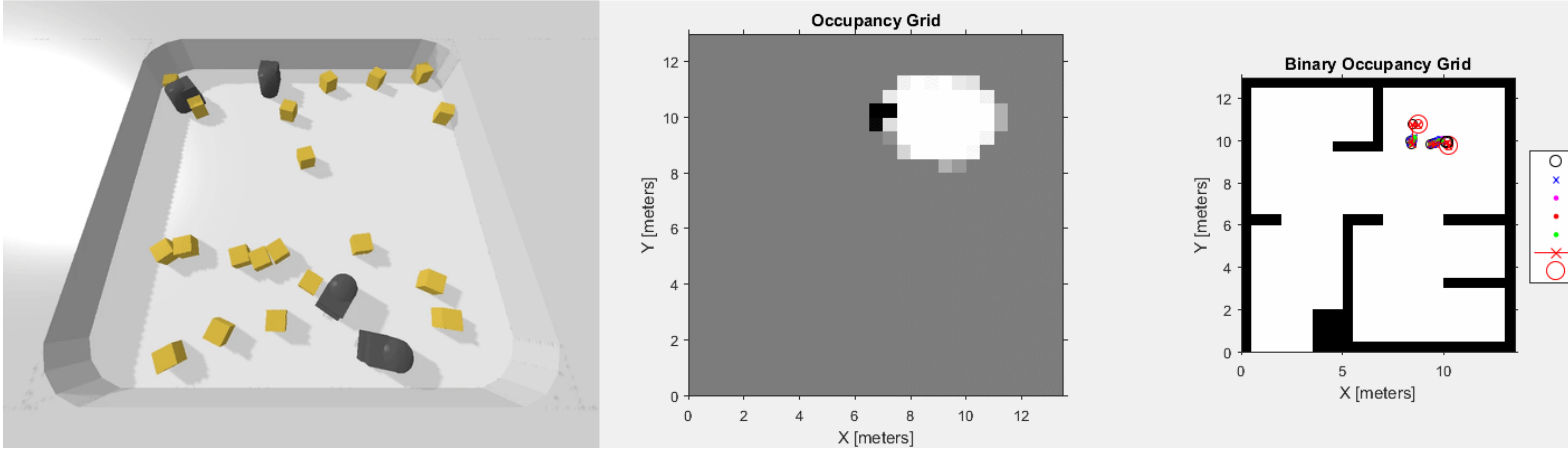
Acting

Interacting

Building a scene representation is the foundation for all tasks

https://storage.googleapis.com/pai-images/e99cfe0c3148408cbca91abd38350d19.jpeg
https://assets.entrepreneur.com/content/3x2/2000/1626727321-robot-dancing.jpg
https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcT95sHLHxLj4175-xrEiEMZVJC2hoplkTO6qg&usqp=CAU
https://www.theconstructsim.com/wp-content/uploads/2020/06/Robot-Navigation-TCS-2020.png

Agents in the Wild



We need to align and merge spatiotemporal data.

https://images.app.goo.gl/4zpVHxuQdb55CM937
https://images.app.goo.gl/8N7mFAcmjL5uymSv5

Living Scenes

Create & update replicas of geometry, semantics, & change using visual data*

Geometry-based	Scene Graph-based	Drastic Change
Living Scenes	**SGAligner**	**Nothing Stands Still**
Multi-object Relocalization and Reconstruction in Changing 3D Environments	3D Scene Alignment with Scene Graphs	A spatiotemporal benchmark on 3D point cloud registration

** while ensuring privacy and realistic implementations*

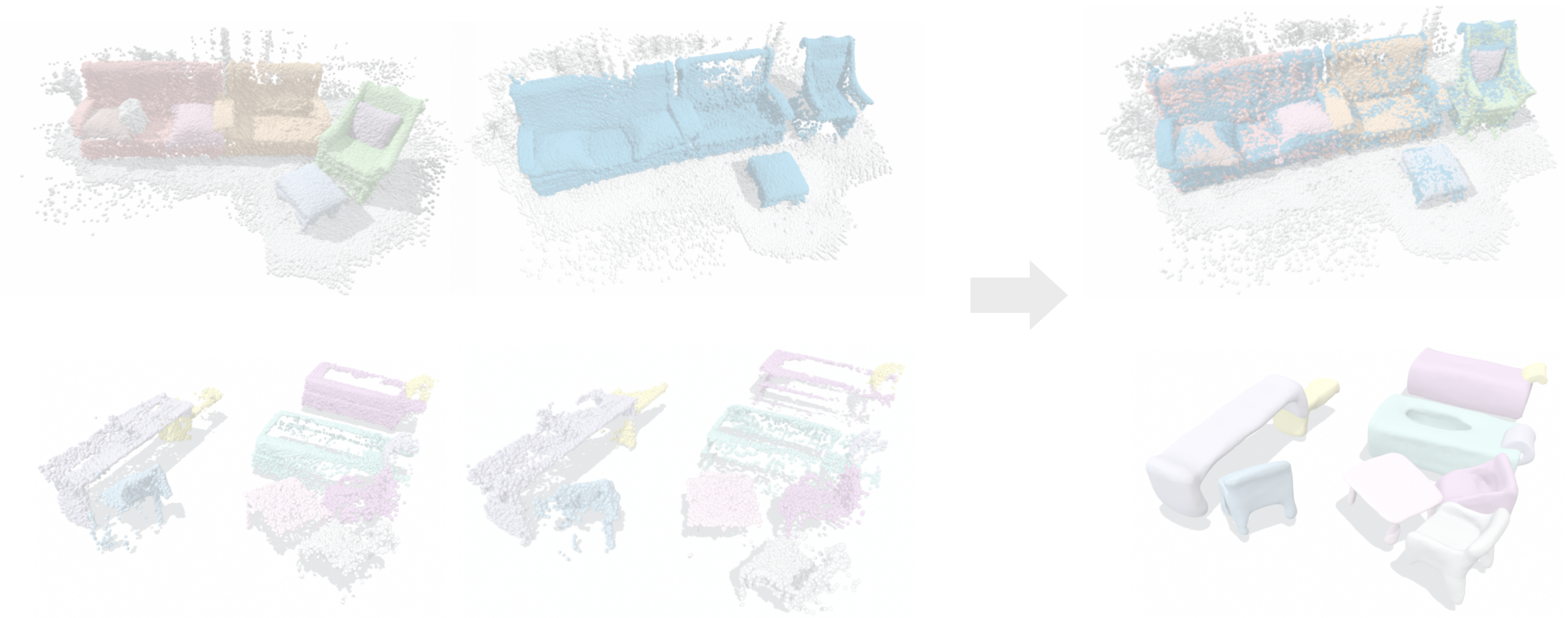
Living Scenes

Multi-object Relocalization and Reconstruction in Changing 3D Environments

Liyuan Zhu

Liyuan Zhu, Shengyu Huang, Konrad Schindler, Iro Armeni

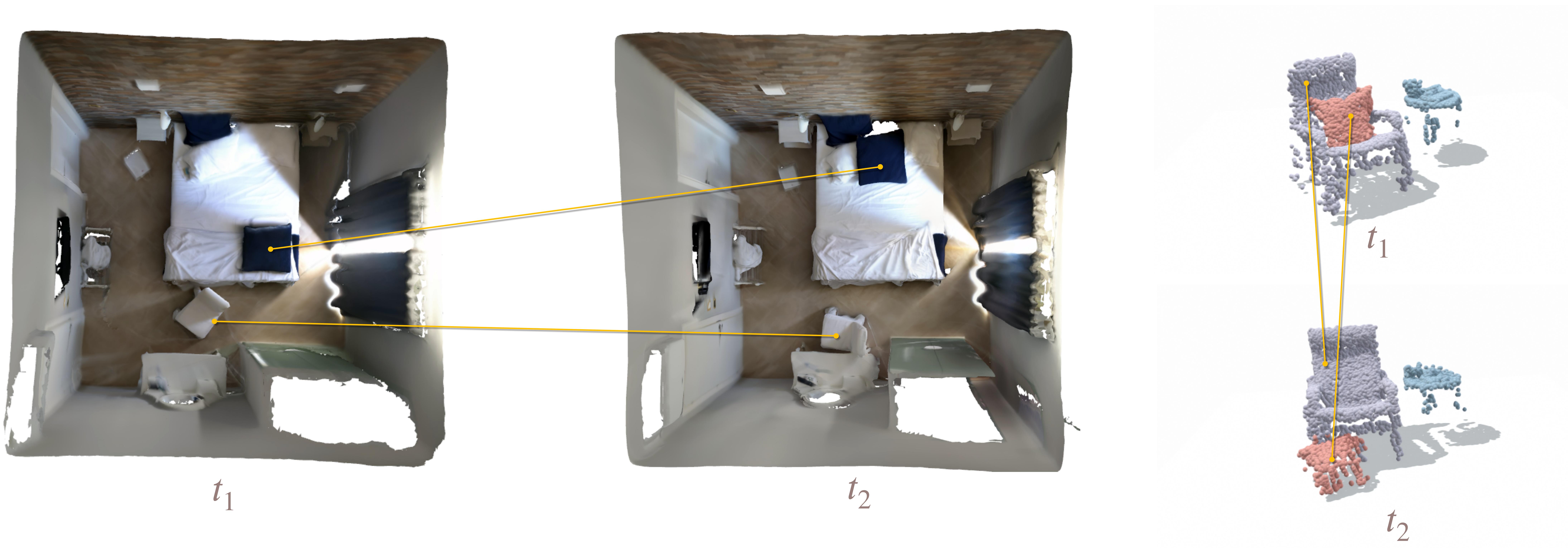
Spotlight



CVPR
SEATTLE, WA JUNE 17-21, 2024

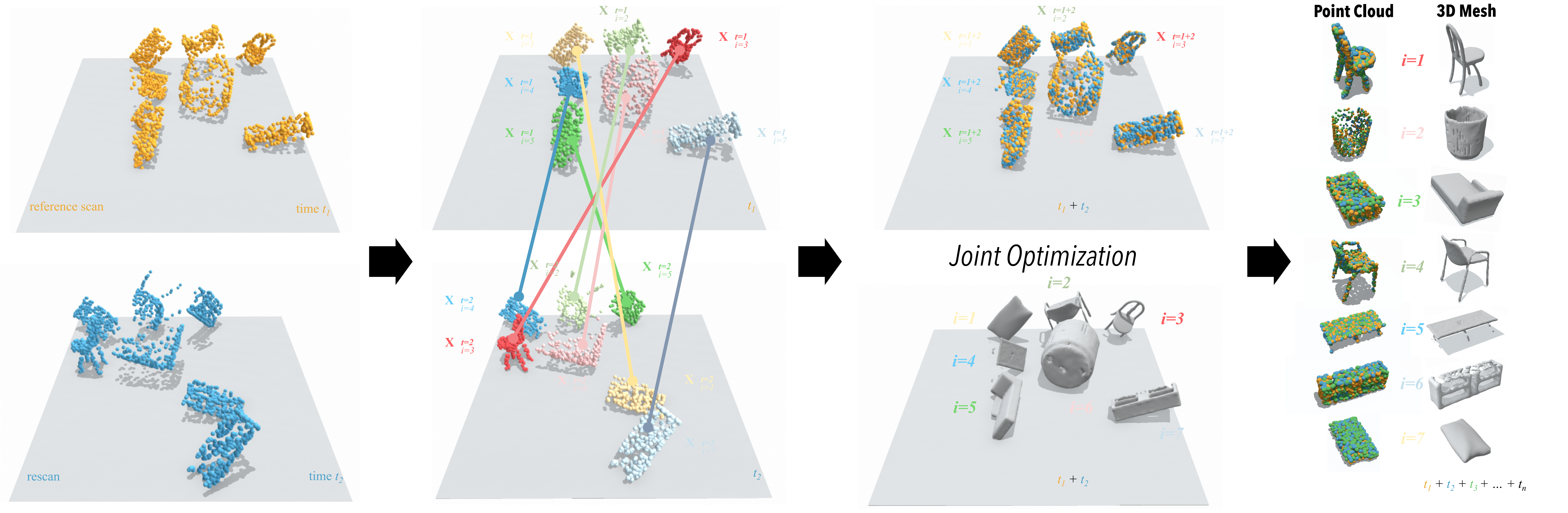
Living Scenes

Evolving environments with irregular change over long time spans and with sparse observations



Goal: Given sparse 3D observations, relocalize them over time and reconstruct them

Method Overview



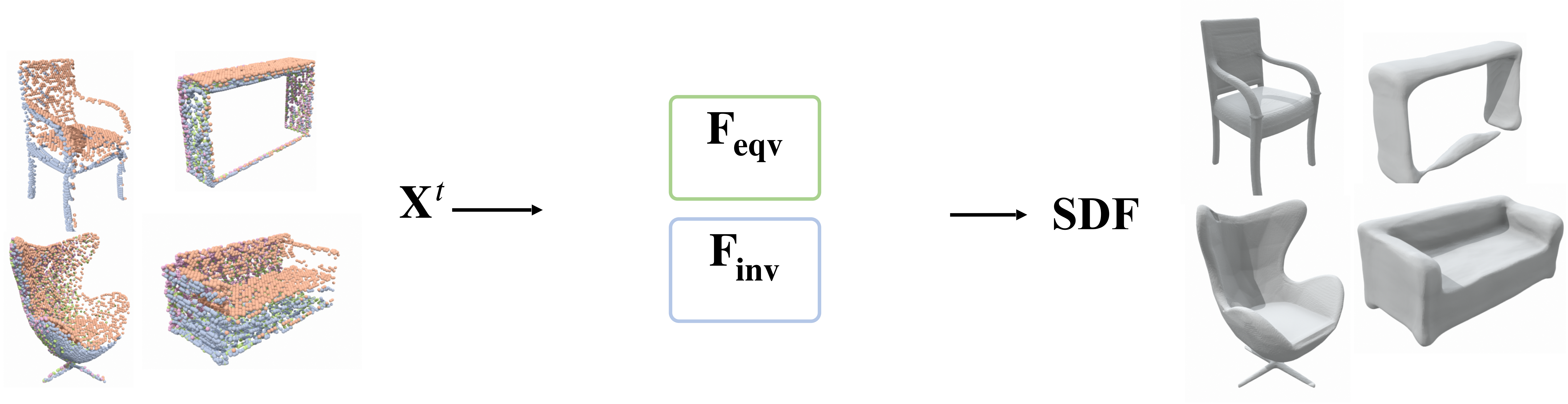
Instance Matching

Registration & Reconstruction

Instance-centric Accumulation

- 1 representation and embedding space for **all** tasks
- Trained only on synthetic data (object CAD models)
- Zero-shot evaluated on real-world noisy data
- Beyond what is seen: shape completion

Train on synthetic data

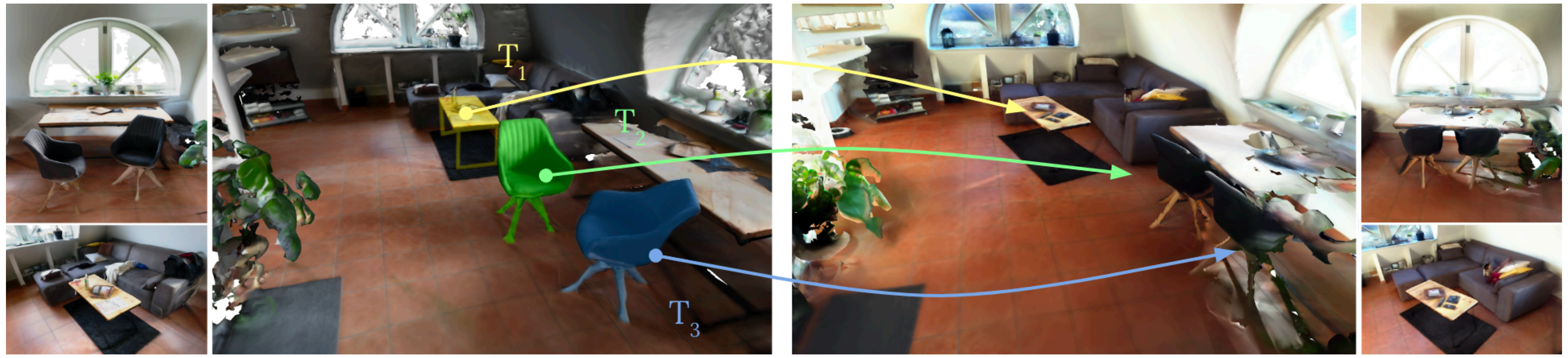


- VN-Encoder [2, 3] and DeepSDF decoder [1]
- Category-agnostic: trained on 7 classes on ShapeNet [4]
- SE(3)-equivariant and invariant embeddings

$$f(\mathbf{RX}) = \mathbf{R}f(\mathbf{X}), f(\mathbf{RX}) = f(\mathbf{X})$$

[1] Park et al., 2019. [2] Deng et al., 2021. [3] Assaad, et al., 2023. [4] Chang et al., 2015.

Evaluation on 3RScan Dataset



3RScan [1]

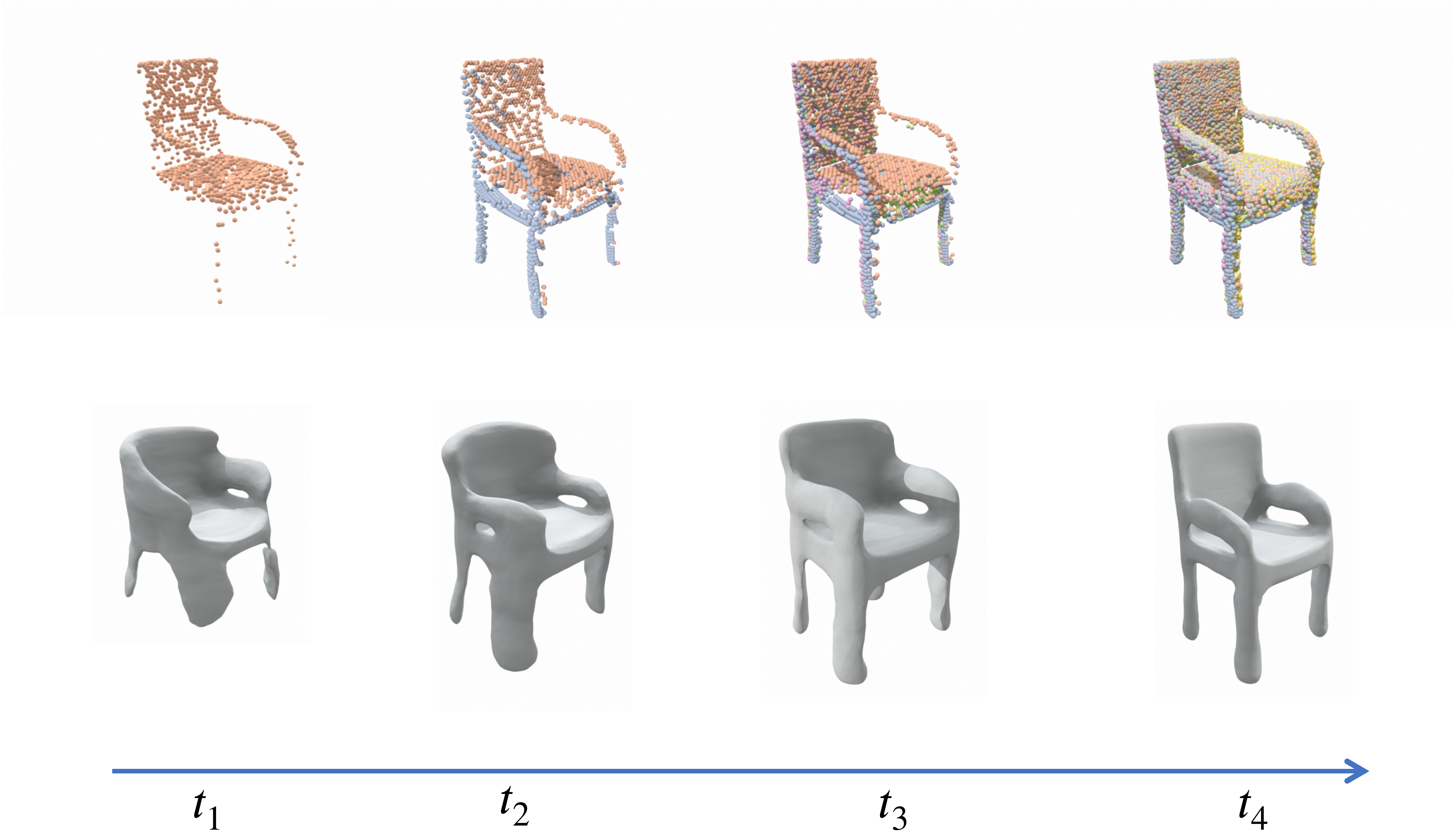
- 1482 3D reconstructed scenes
- 478 environments
- Temporal Change
- RGB-D sequences

Lived environments with sparse observation of object relocation, addition, or removal

[1] Wald et al, Rio: 3d object instance relocalization in changing indoor environments, ICCV 2019

Qualitative Results

Benefit of Accumulation



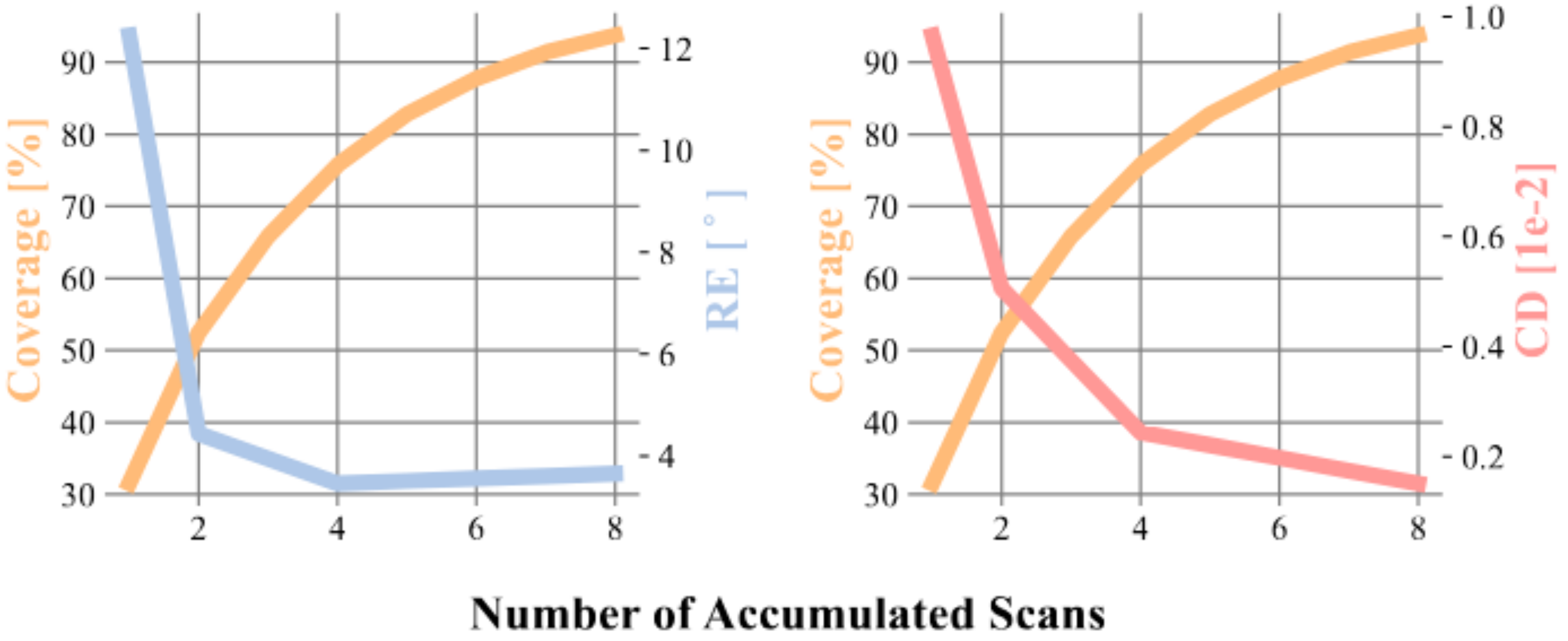


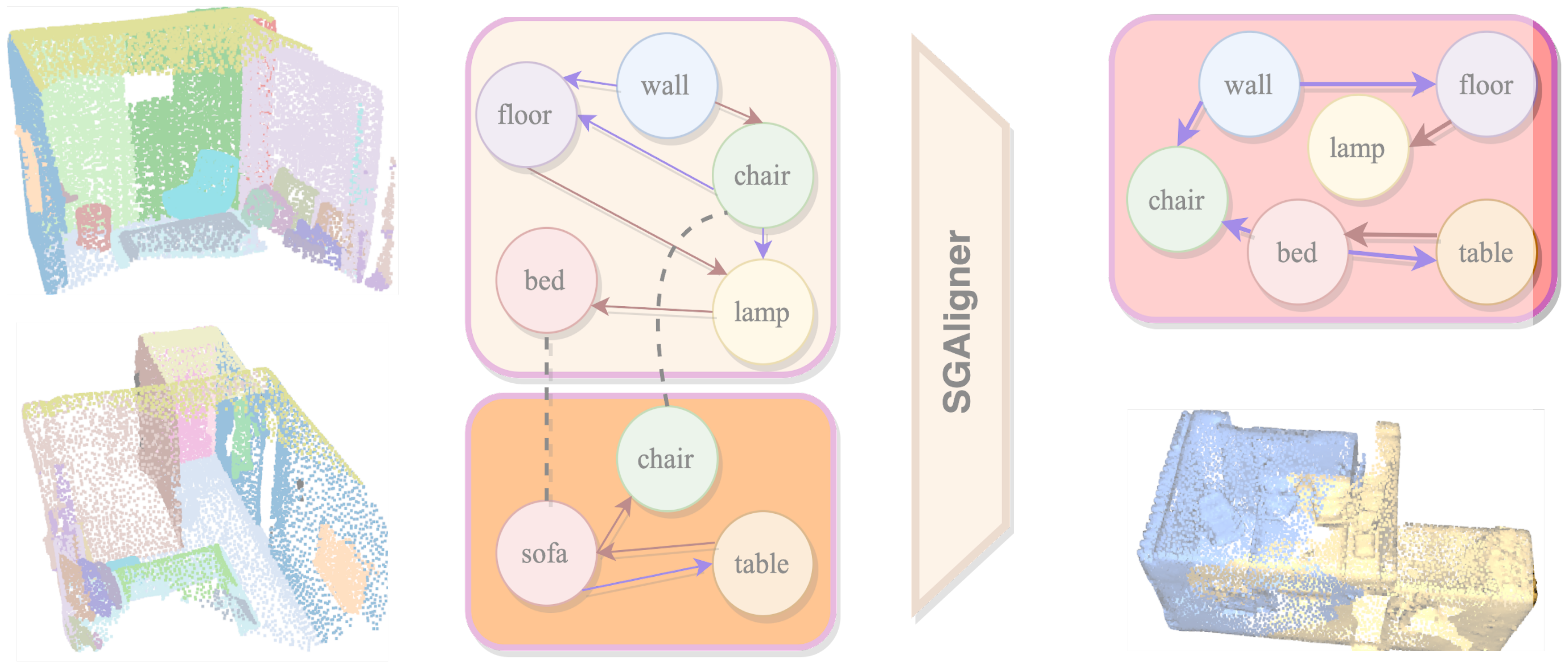
Figure 7. **Ablation study on point cloud accumulation.** The change of point cloud coverage, rotation error and chamfer distance w.r.t the number of accumulated scans.

SGAligner

3D Scene Alignment with Scene Graphs

Sayan Deb Sarkar, Ondrej Miksik, Marc Pollefeys, Daniel Barath, Iro Armeni

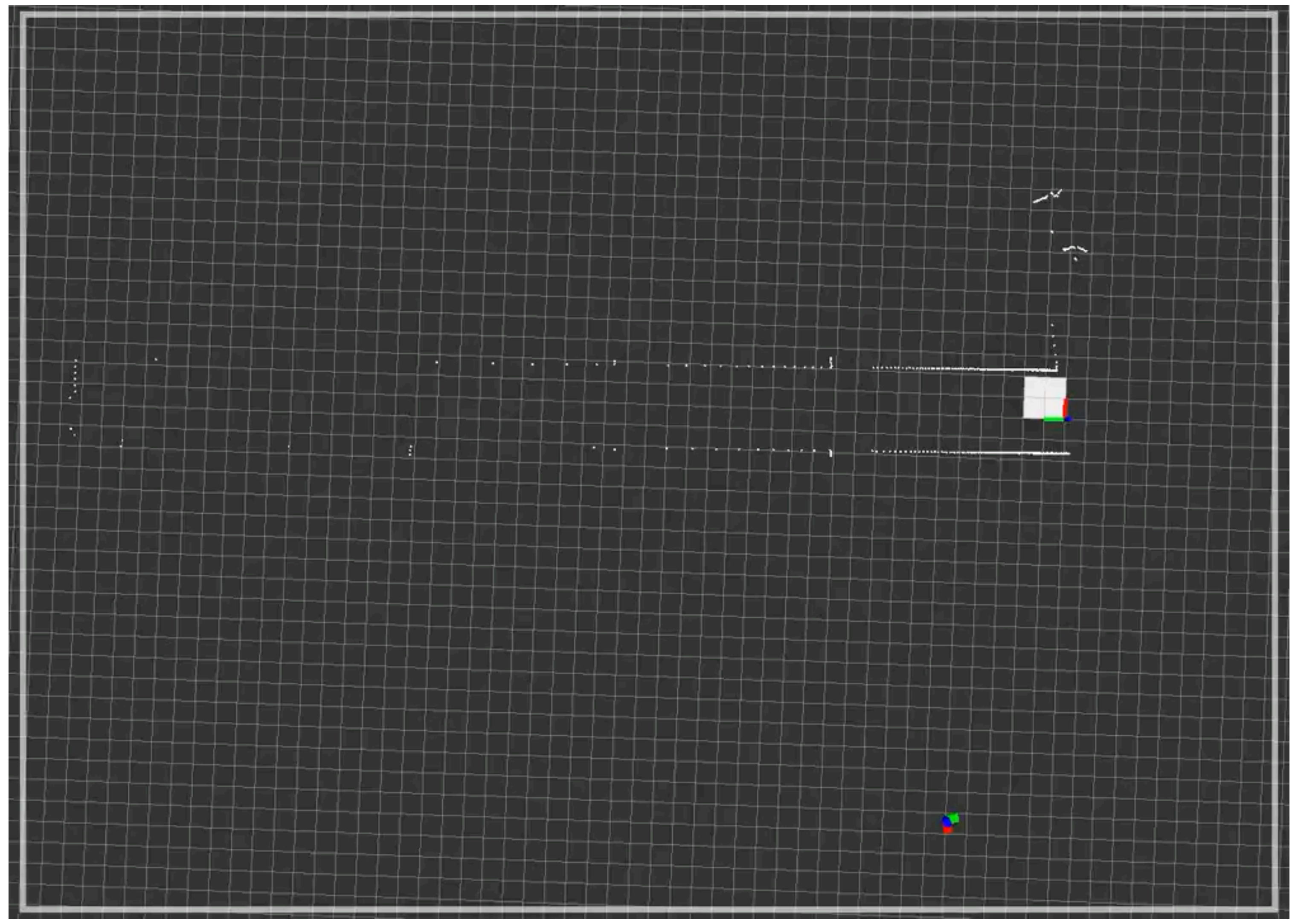
Sayan Deb Sarkar



ICCV23
PARIS

Low-Level

Building a ^ ___ Map



Representations

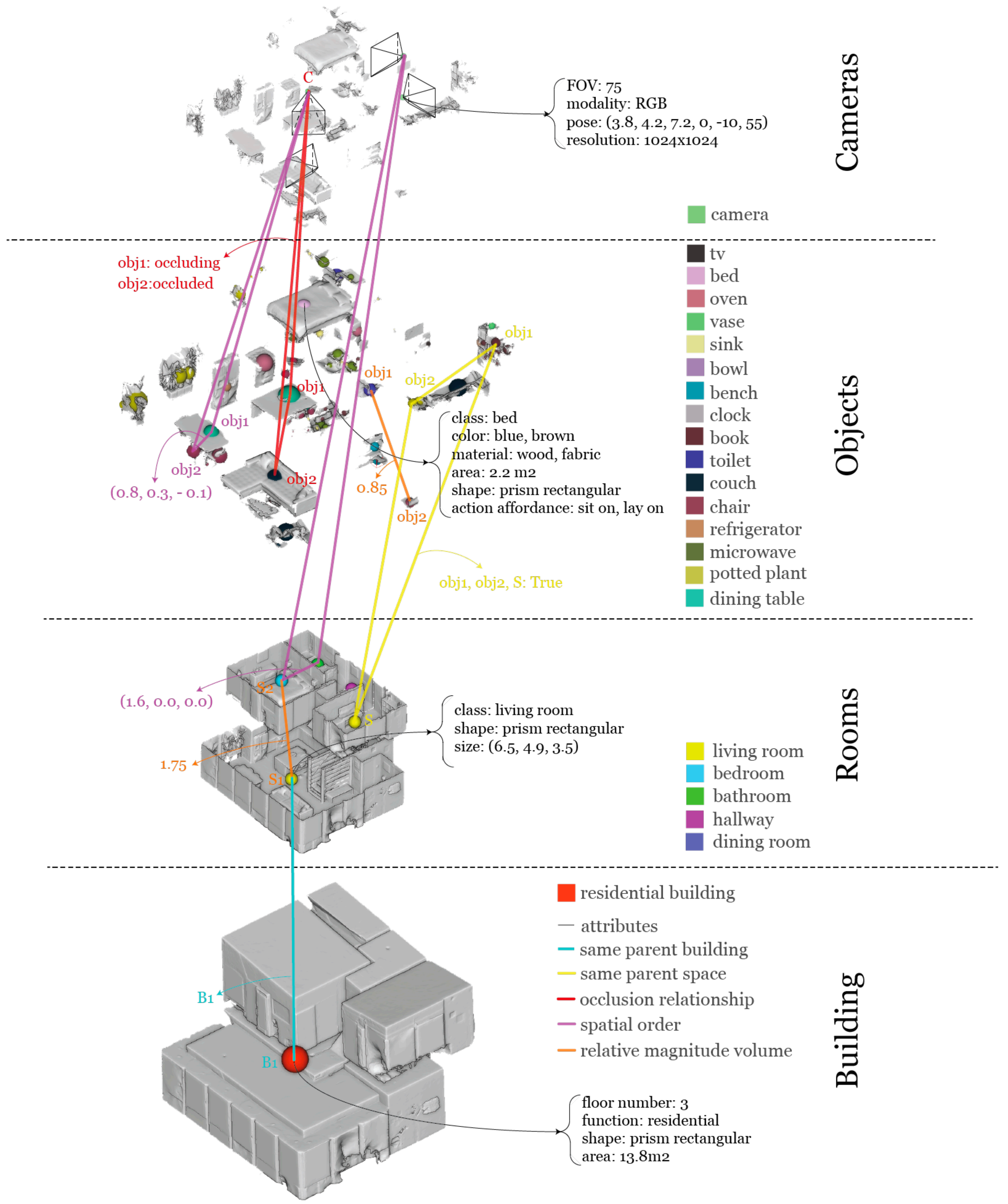
- Occupancy map (e.g., voxel grid, octomap, hash grid)
- Point cloud
- …

Caveats

- Decision making takes place on the metric space [Sepulveda et al., 2018]
- Semantic labels are attached directly to 3D geometry
- Limited higher-level understanding

https://www.youtube.com/watch?v=v4flz0AtENk&t=35s

Building a ^ High-Level Map



3D Scene Graphs

[Armeni et al., 2019; Kim et al., 2021; Rosinol et al., 2020; Wald et al., 2020]

- Includes both high- and low-level information
- Allows for decision making on the semantic space
- Light-weight [Chang et al, 2019]
- Privacy preserving [Li et al, 2022; Zhang et al., 2022]

Armeni et al., ICCV 2019

3D Scene Graphs in …

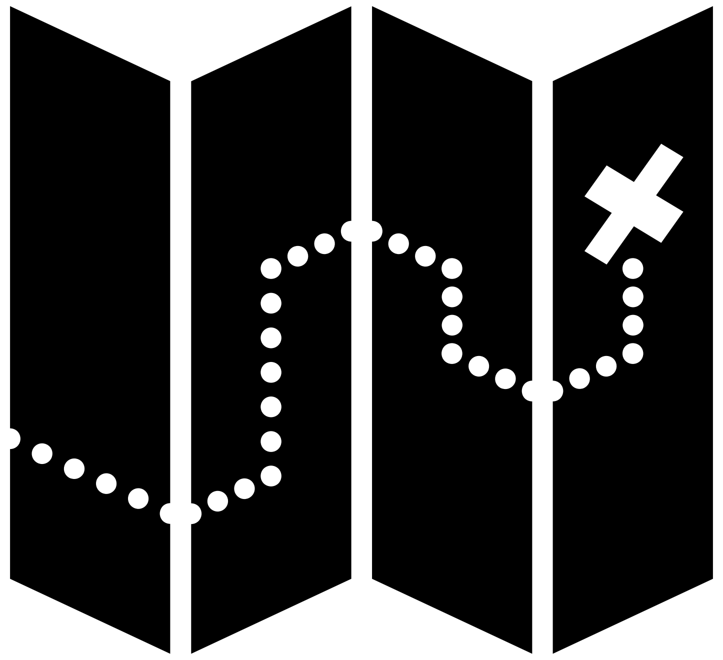
Use

- Increasingly used in agents, commonly built on the fly
- To perform:
  - *robotic navigation* [Sepulveda et al., 2019; Rosinol et al., 2021; Ravichandran et al., 2022; Li et al., 2022; Chang et al., 2022]
  - *task completion* [Gadre et al., 2022; Agia et al., 2022; Ravichandran et al., 2020; Jiao et al., 2022; Li et al., 2022]

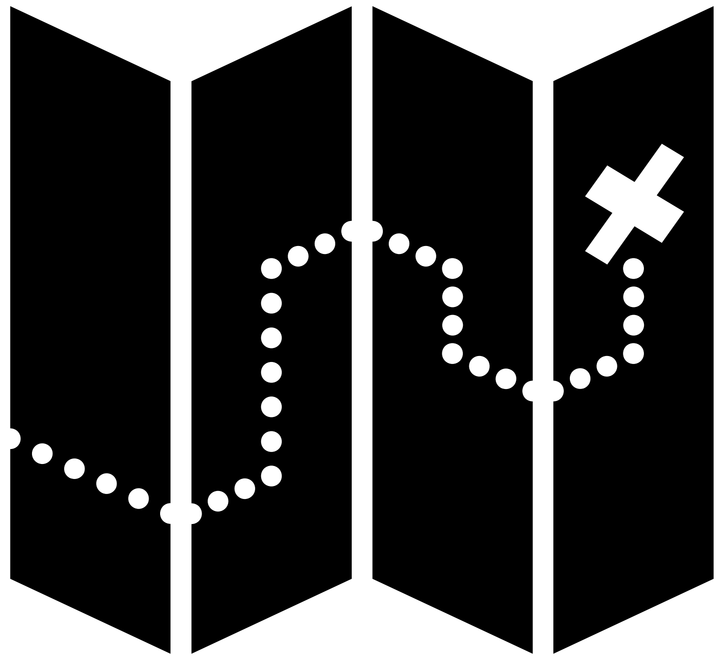
Prediction

- Online incremental building [Wu et al., 2021; Hughes et al., 2022]
- Offline based on RGBD images and/or 3D reconstructions [Armeni et al., 2019; Wald et al., 2020; Rosinol et al., 2021]

Can we leverage and recycle them
for creating 3D maps of environments?

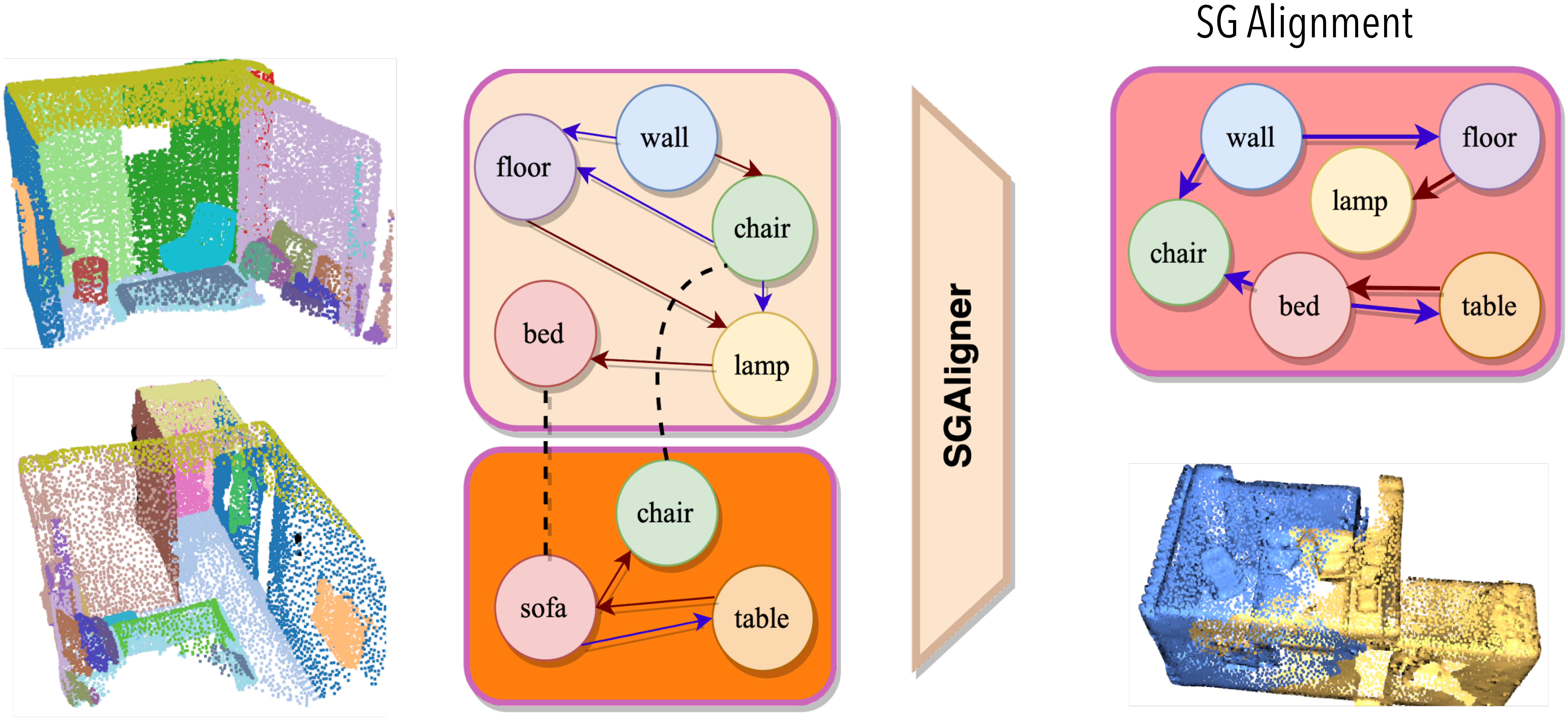


Can we leverage and recycle them for creating 3D maps of environments?



- Scene observations from one or multiple agents
- Static or changed scene
- Overlap from *zero* to partial or full

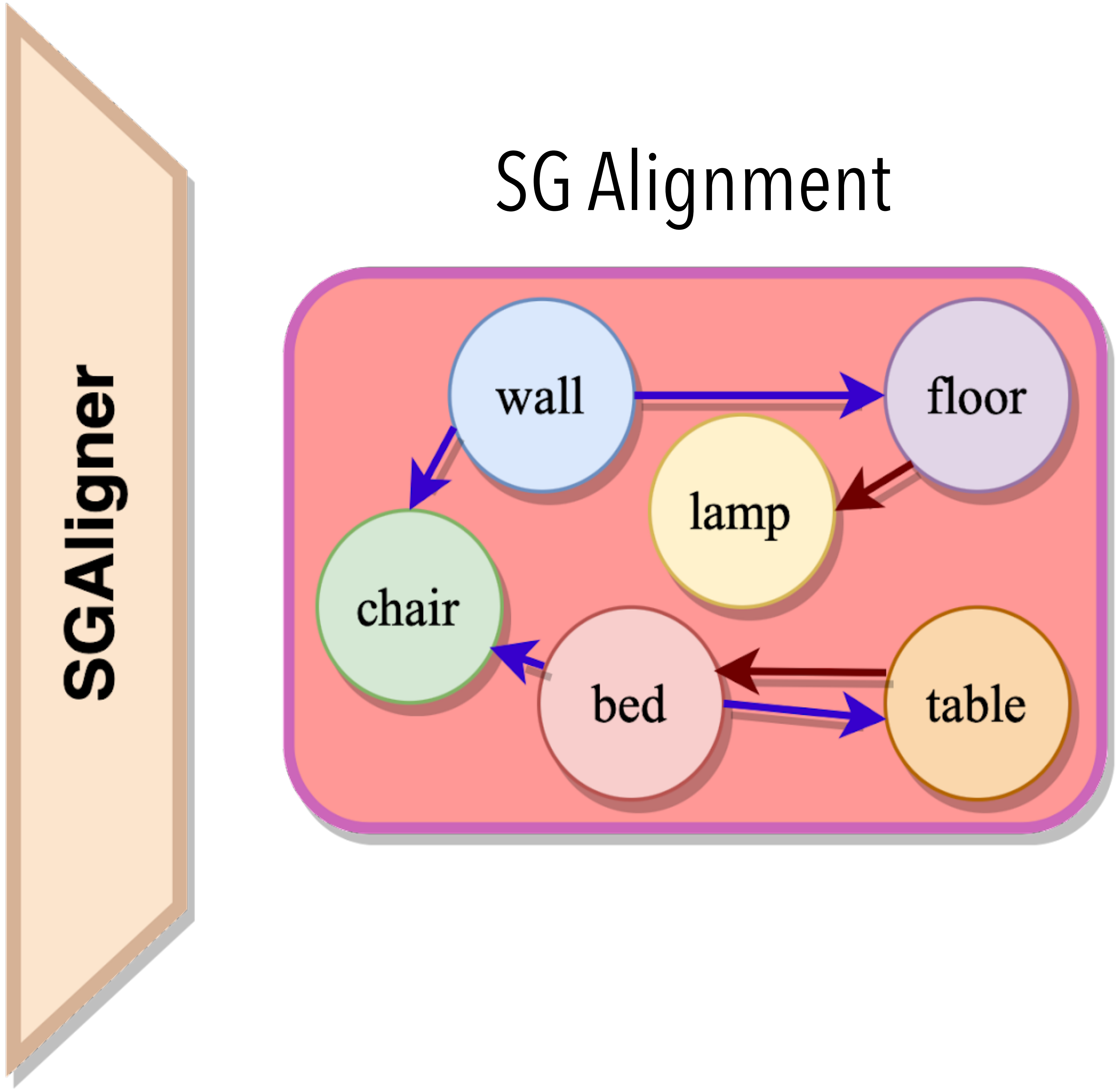
Aligning 3D Scene Graphs (SG) in the Wild



Pairs of PCs and their SGs

3D PC Registration

3D SG as Multi-modal Knowledge Graphs

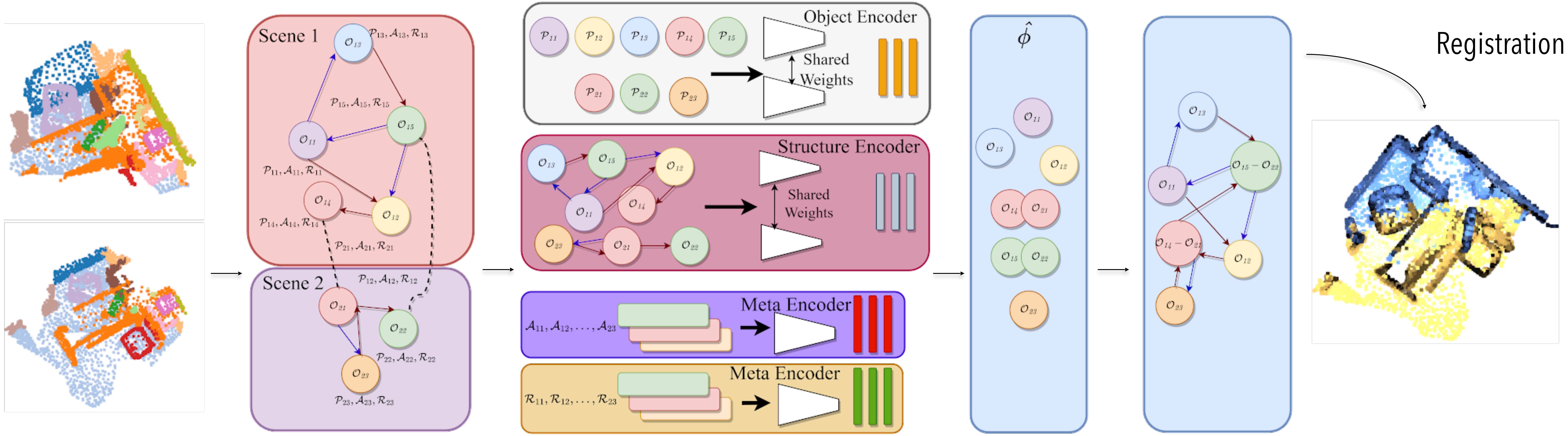


<u>3 Types of Information</u>

- semantic entities in the scene (e.g., object instances)
- their attributes (e.g., category, size, and material)
- relationships between the entities (e.g., relative position and attribute similarity)

Redesign entity alignment methods in multi-modality knowledge graphs for our setting

SGAligner Overview



Input 3D PCs

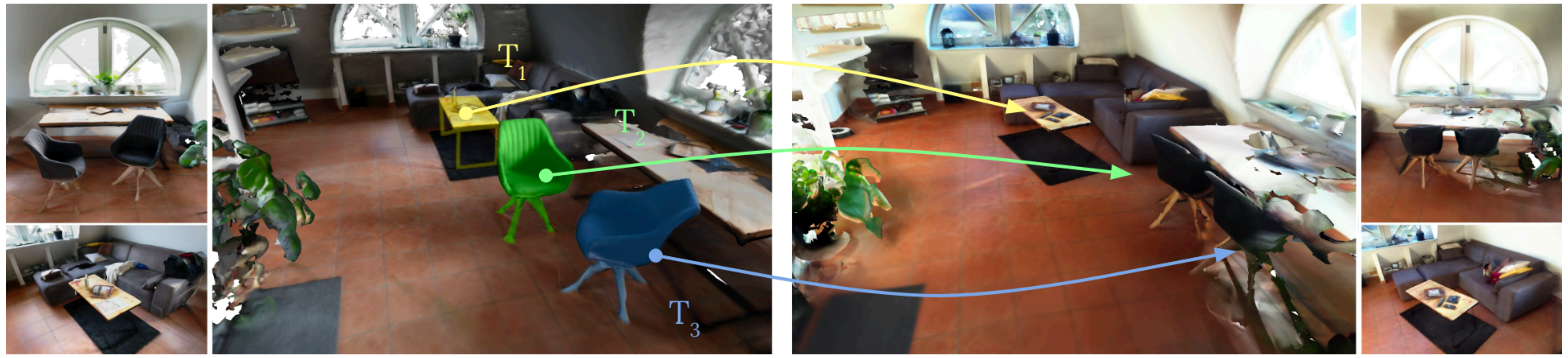
Input 3D SGs

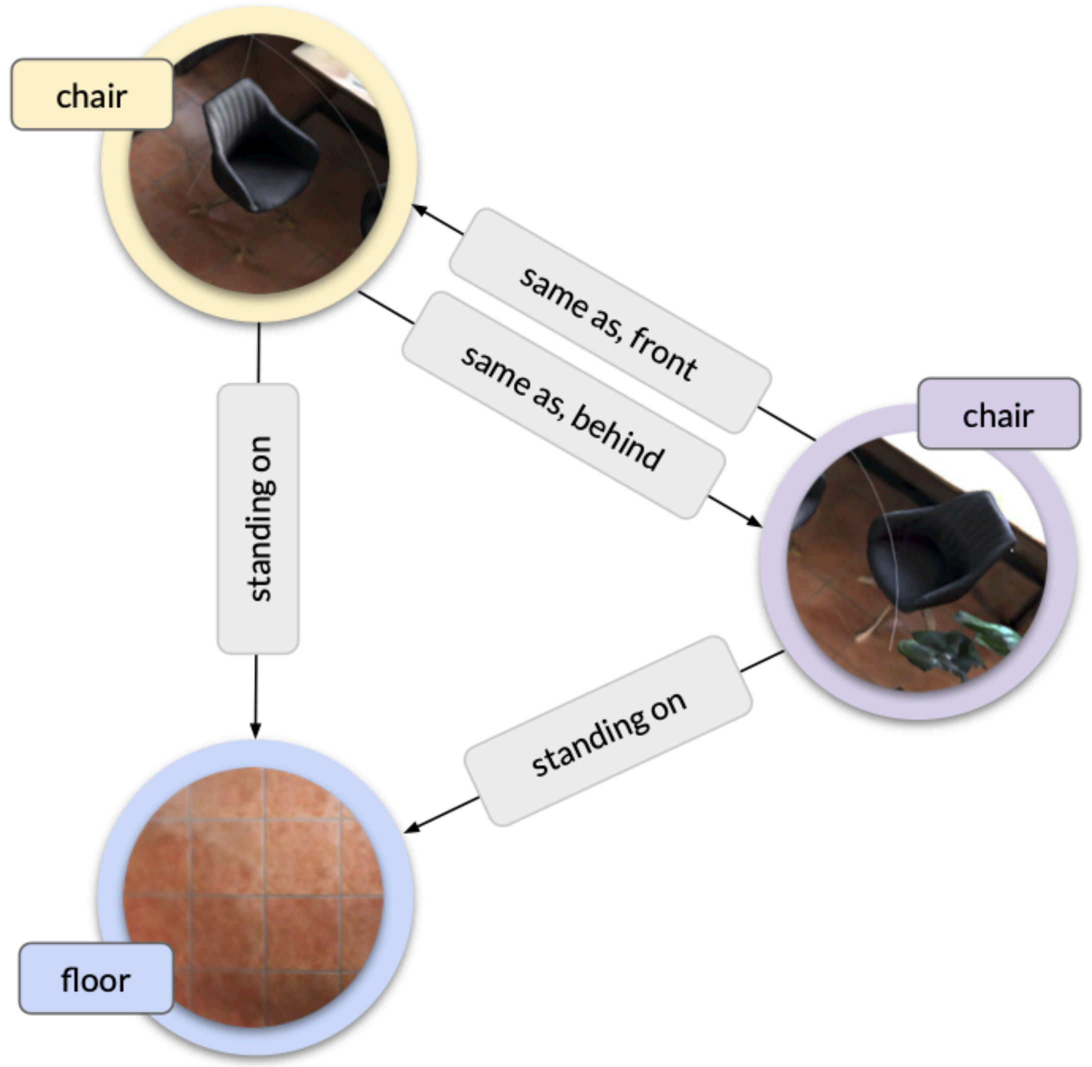
Uni-modal Embeddings

Joint Embedding Space

Aligned 3D SGs

3RScan + 3DSSG Datasets





3RScan [1]

- 1482 3D reconstructed scenes
- 478 environments
- Temporal Change
- RGB-D sequences

3DSSG [2]

- 3D Scene Graphs for 3RScan
- 48k nodes 544k edges
- Attributes: Static and Dynamic
- Relationships: Support, Proximity, Comparative

[1] Wald et al, Rio: 3d object instance relocalization in changing indoor environments, ICCV 2019
[2] Wald et al, Learning 3d semantic scene graphs from 3d indoor reconstructions, CVPR 2020

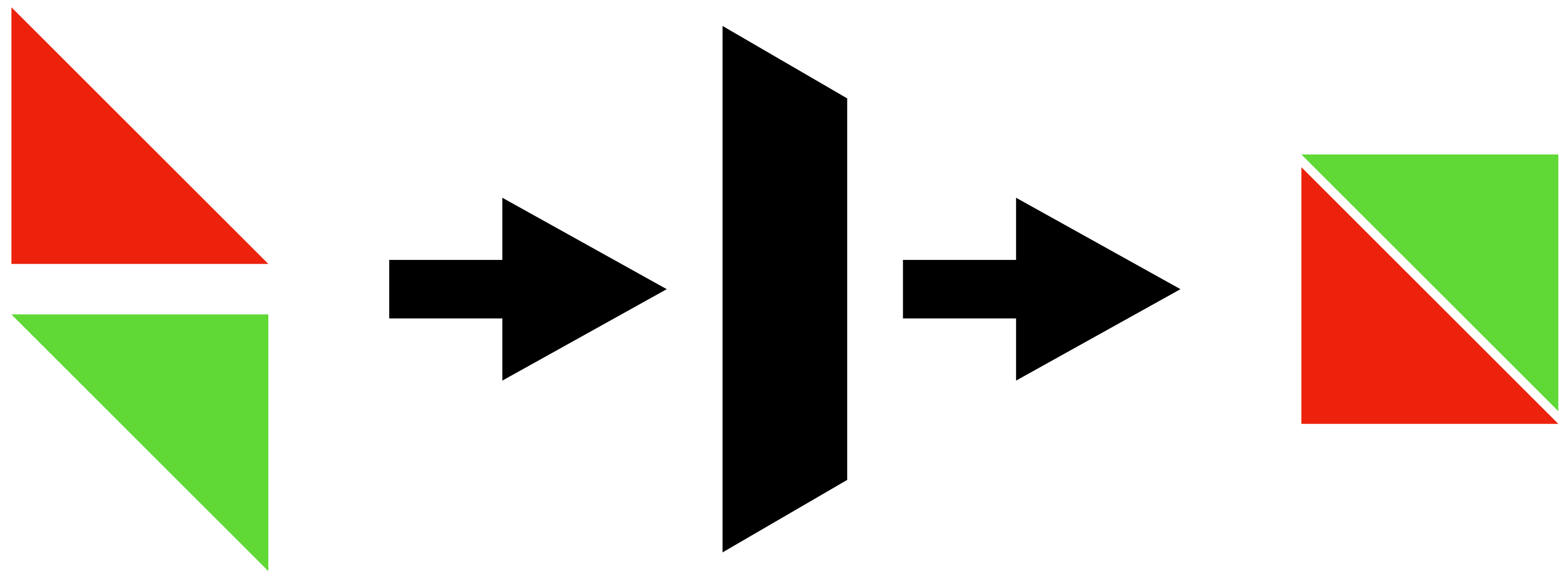
Evaluation on 3D Scene Graph Alignment

- ❖ Accurate matches regardless of input noise
- ❖ Meaningful results even in low spatial overlap
- ❖ Robust to temporal changes

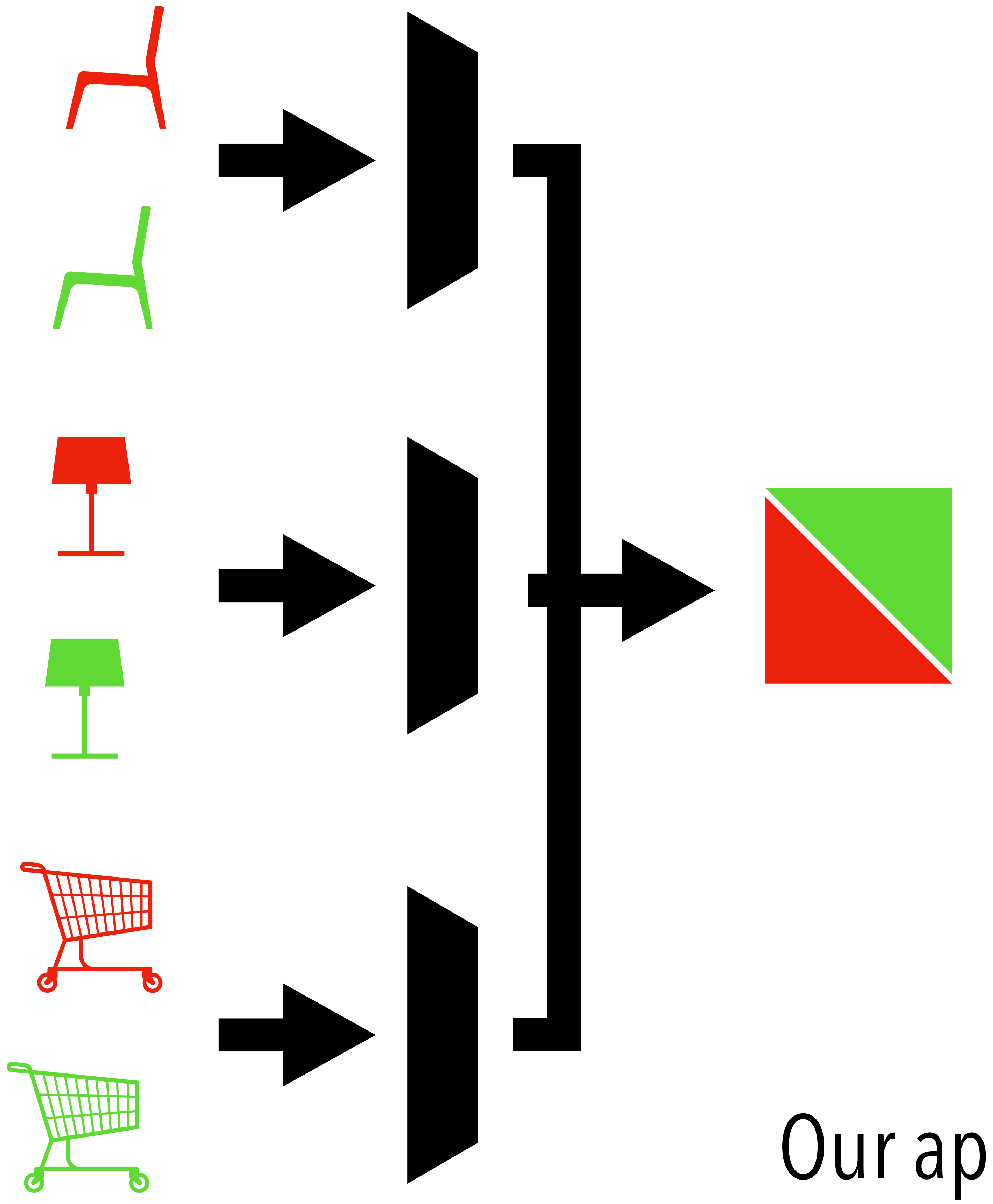
With temporal changes

Mean RR ↑	Hits @ ↑					No. of Pairs
	K = 1	K = 2	K = 3	K = 4	K = 5	
0.886	0.833	0.894	0.928	0.946	0.957	2262

Application : Point Cloud Registration



Common approaches



Our approach

[1] Qin et. al., Geometric Transformer for Fast and Robust Point Cloud Registration, CVPR 2022

Application : Point Cloud Registration

3D Point Cloud Registration

	Methods	CD ↓	RRE (°) ↓	RTE (cm) ↓	RR(%) ↑
	w/ Ground Truth 3D Scene Graphs				
	GeoTr	0.02247	1.813	2.79	98.49
Ours	K=1	0.01677	**1.425**	2.88	98.79
	K=2	**0.01111**	**1.012**	**1.67**	**99.40**
	K=3	**0.01525**	1.736	**2.55**	**98.81**
	w/ Predicted 3D Scene Graphs				
	GeoTr	0.06643	5.697	9.54	93.15
Ours	K=1	0.05041	2.49	3.86	94.95
	K=2	0.04251	1.725	3.36	98.33
	K=3	0.04863	2.194	2.55	97.96

Evaluation Per Overlap

	Overlap (%)	CD ↓	RRE ↓ (°)	RTE (cm) ↓	RR (%) ↑
GeoTr.	10-30	0.09788	8.830	13.56	92.25
	30-60	0.00584	0.156	0.24	97.36
	60-	0.00177	0.048	0.07	99.31
Ours	10-30	0.05160	5.660	8.48	95.35
	30-60	**0.00127**	**0.045**	**0.05**	**98.34**
	60-	**0.00046**	**0.018**	**0.02**	**99.93**

🔥 **49%** improvement in Chamfer Distance (CD)

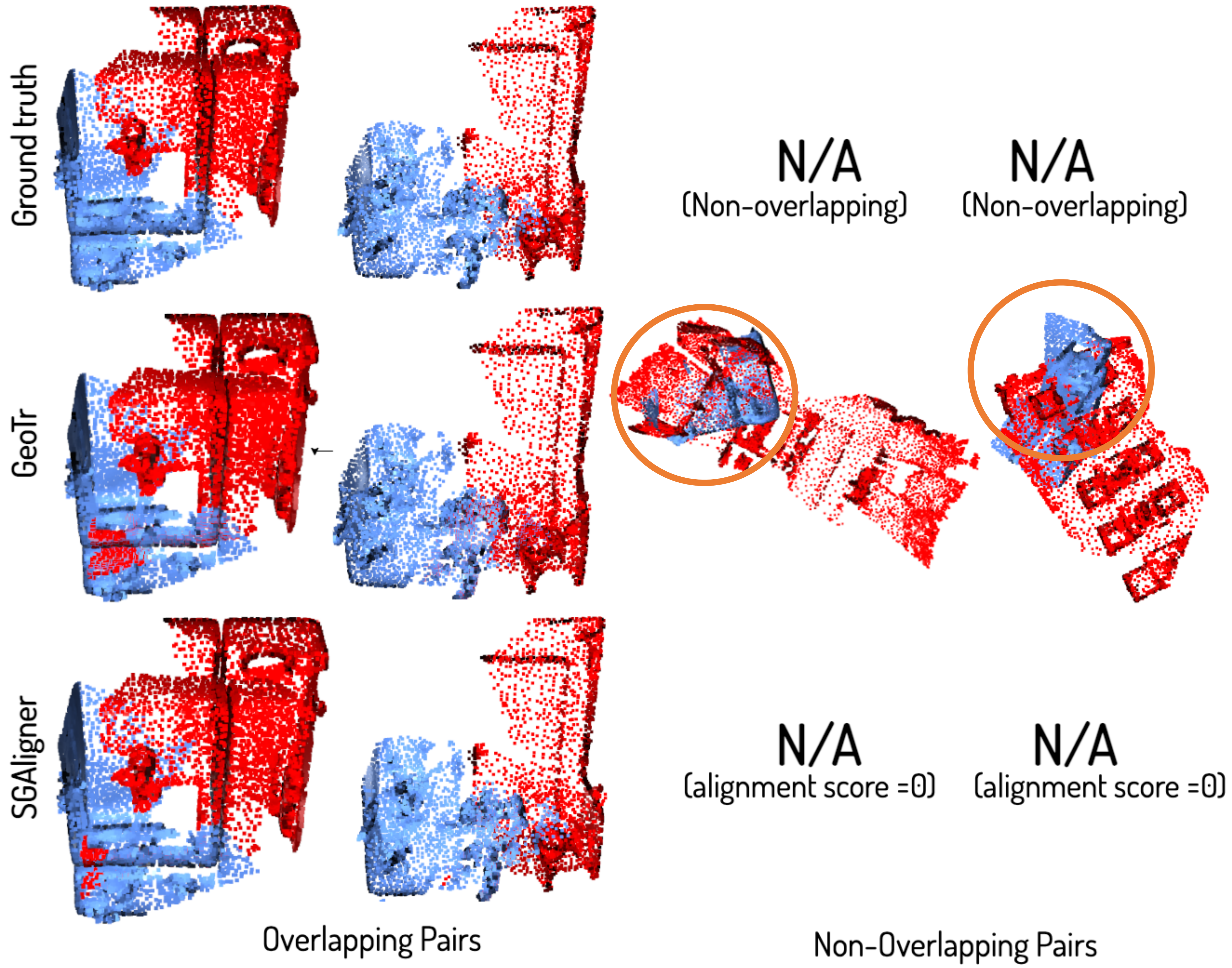
🔥 **40%** improvement in Relative Translation Error (RTE)

❖ Even higher gains, on noisy point cloud predictions

❖ Better than standard registration methods on low overlap scenarios

[1] Qin et. al., Geometric Transformer for Fast and Robust Point Cloud Registration, CVPR 2022

Registration : Finding Overlapping Scenes



Identifying (Non-)Overlapping Scenes

Method	Prec. (%) ↑	Recall (%) ↑	F1 (%) ↑	Average Time Per Scene (ms) ↓
Geotr	**99.63**	80.98	89.34	442.50
Ours	92.03	**90.94**	**91.48**	**139.64**

- 3 times faster than GeoTr
- Less computationally demanding
- Identify overlapping pairs more correctly

[1] Qin et. al., Geometric Transformer for Fast and Robust Point Cloud Registration, CVPR 2022

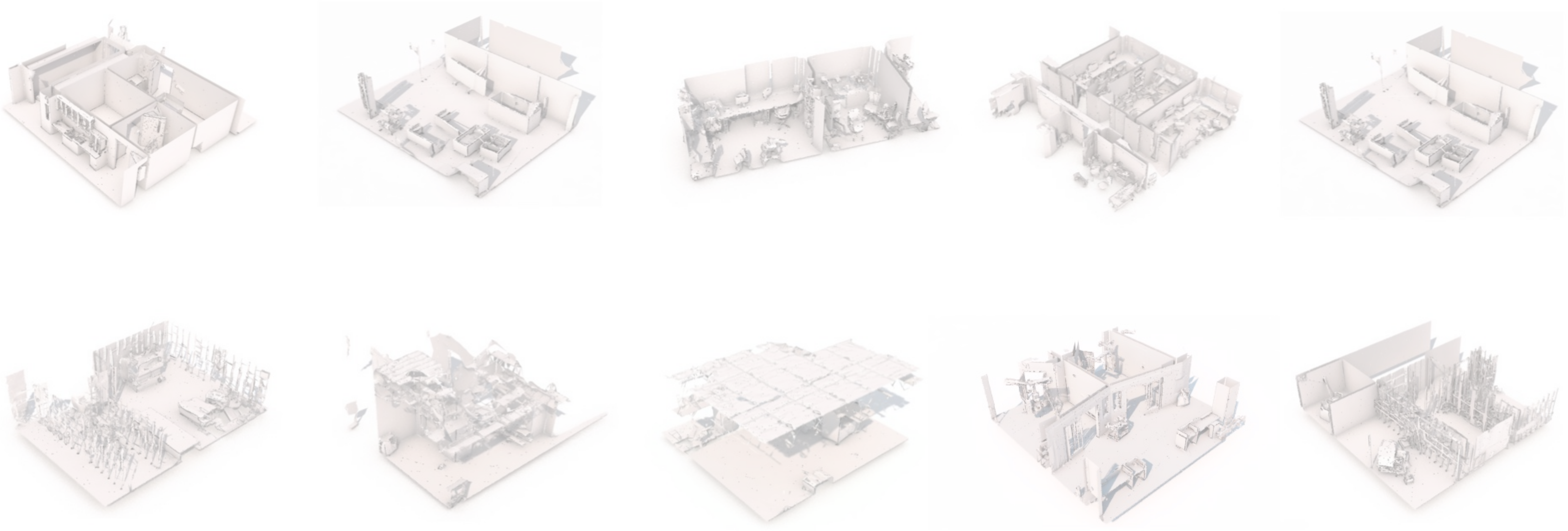
What happens when there are drastic changes?

Nothing Stands Still

Tao Sun

A Spatiotemporal Benchmark on 3D Point Cloud Registration Under Large Geometric and Temporal Change

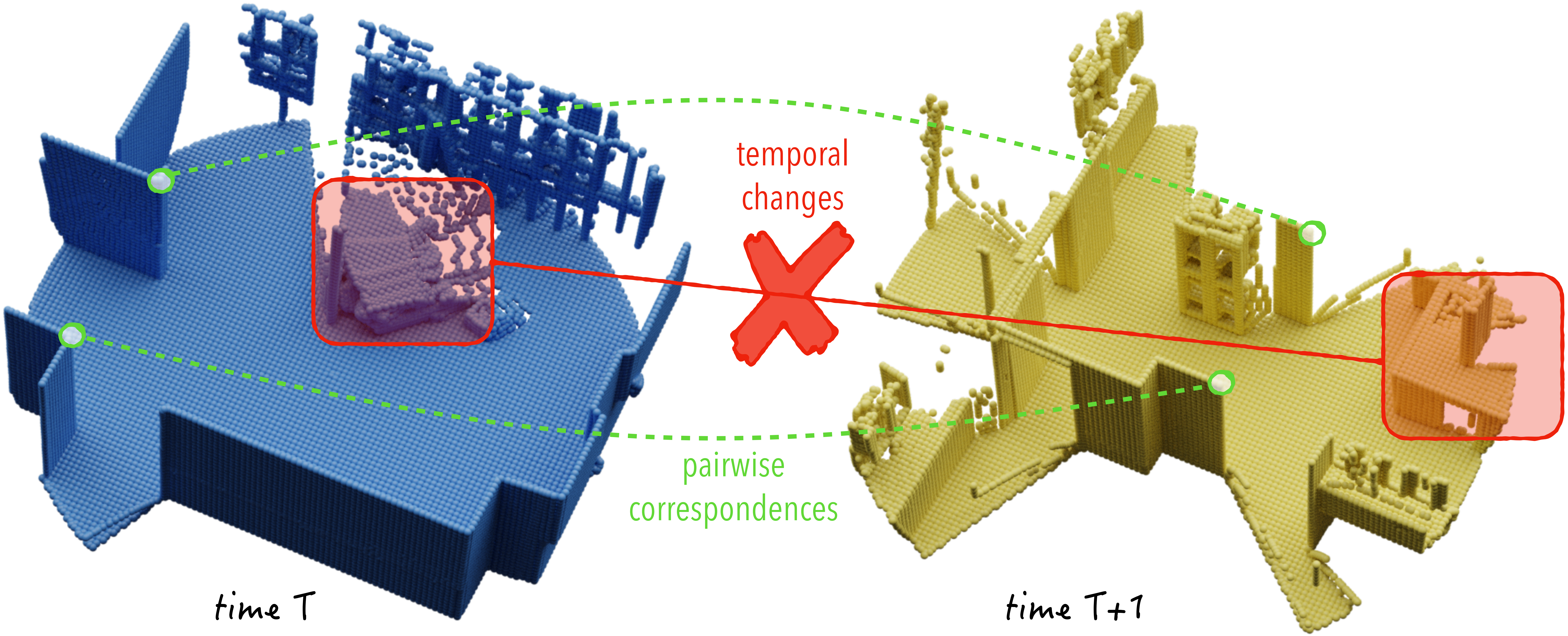
Tao Sun, Yan Hao, Shengyu Huang, Silvio Savarese, Konrad Schindler, Marc Pollefeys, Iro Armeni



Under Review

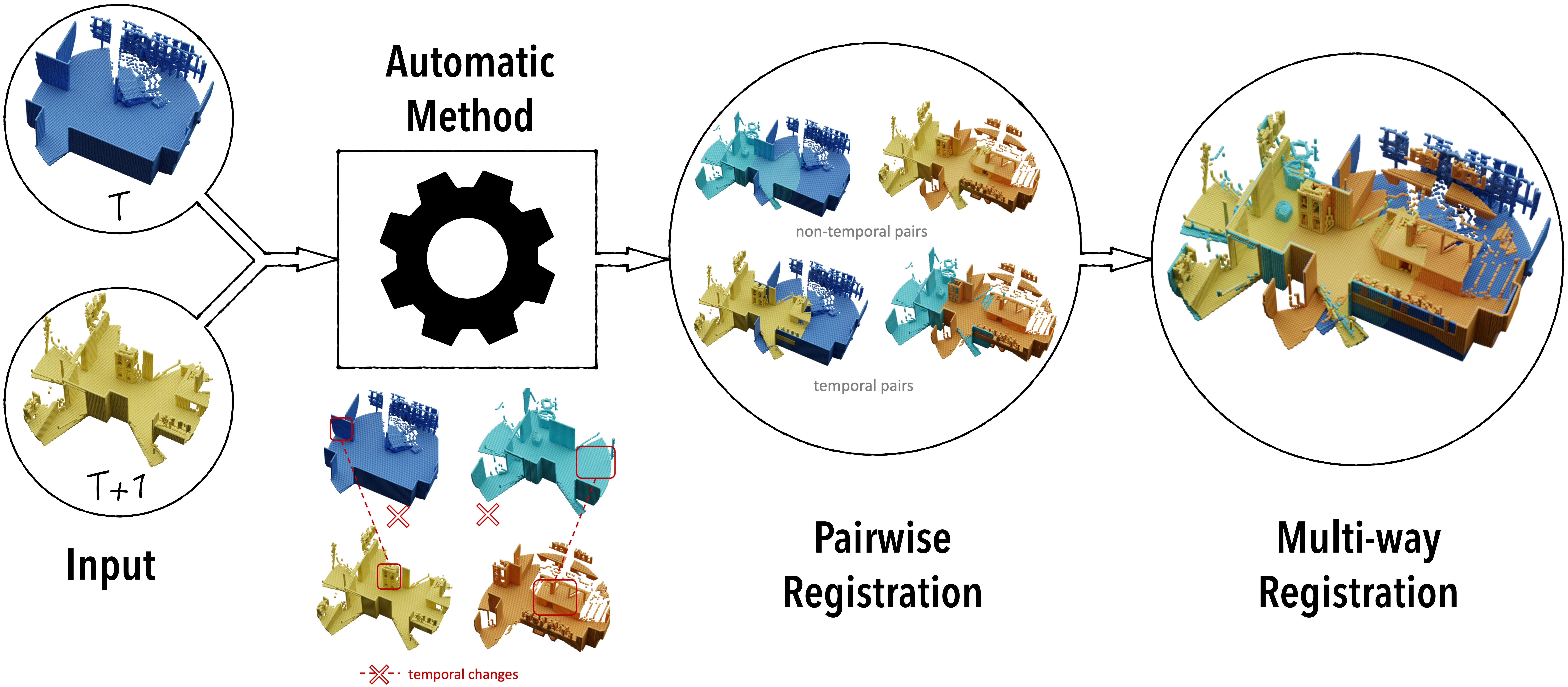
Spatiotemporal 3D Point Cloud Registration

Finding pairwise correspondences between changed scenes



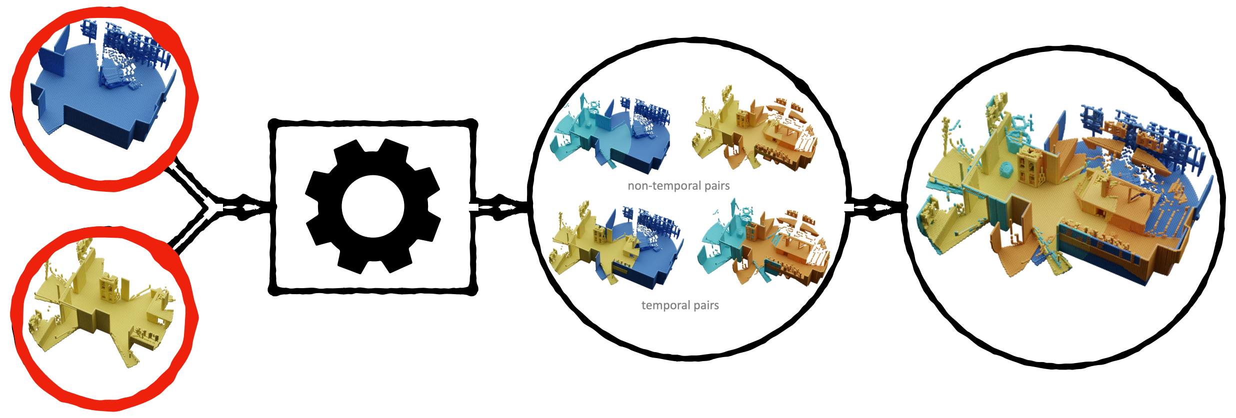
Nothing Stands Still

A Spatiotemporal Benchmark on 3D Point Cloud Registration Under Large Geometric and Temporal Change



Construction: Drastic Changes in Scenes

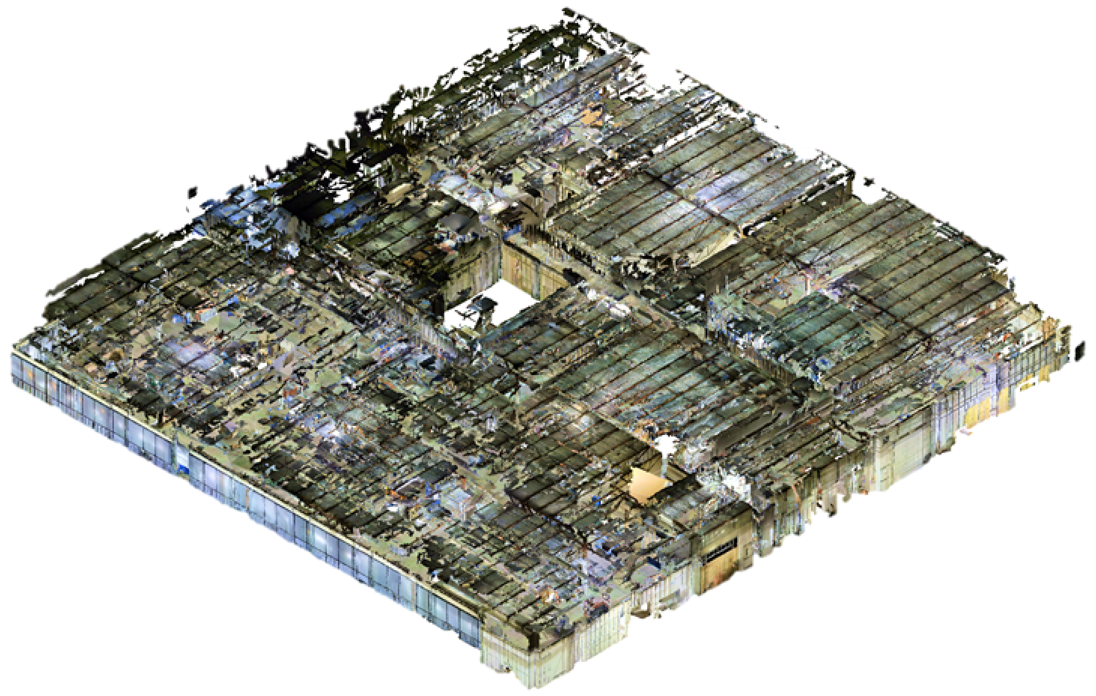
Video source: https://www.youtube.com/watch?v=O0CQspnjEoY



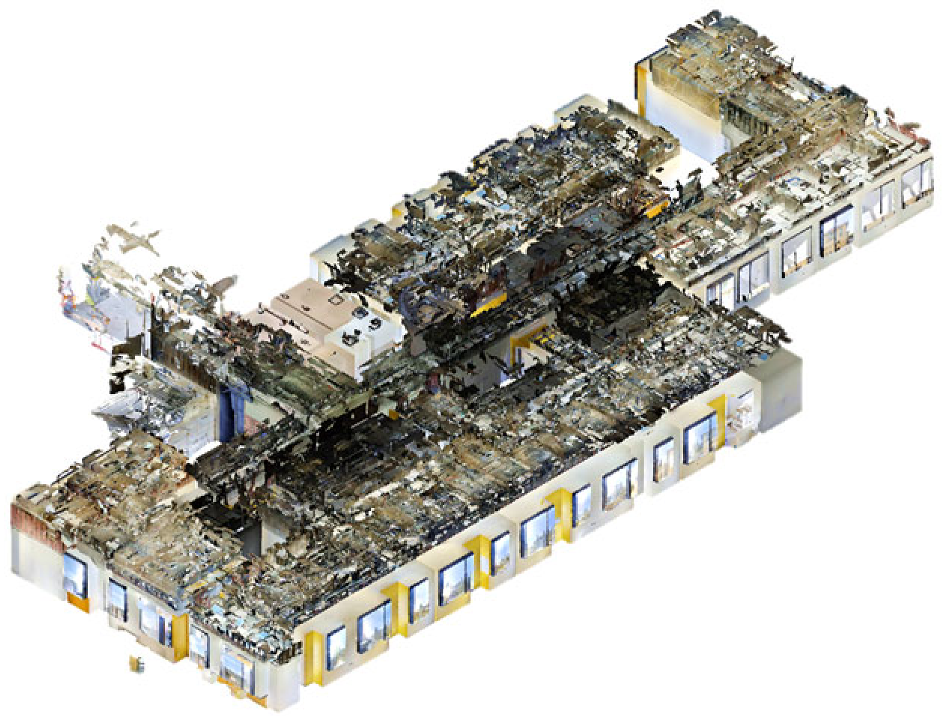
Nothing Stands Still

Data collected from 6 construction sites over months

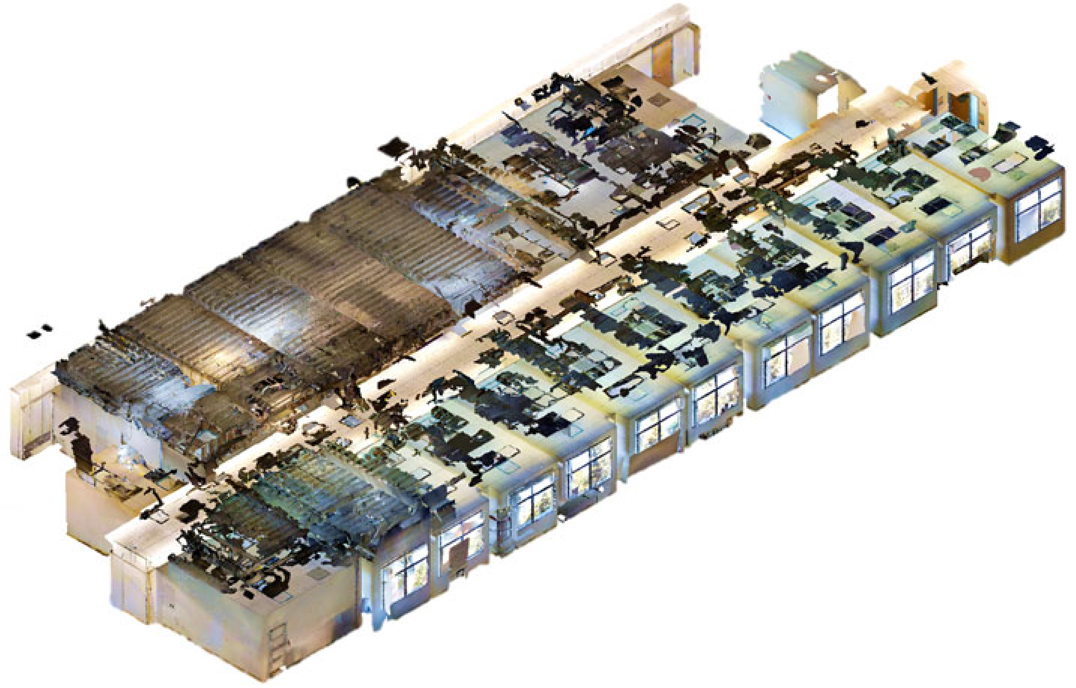
Area A



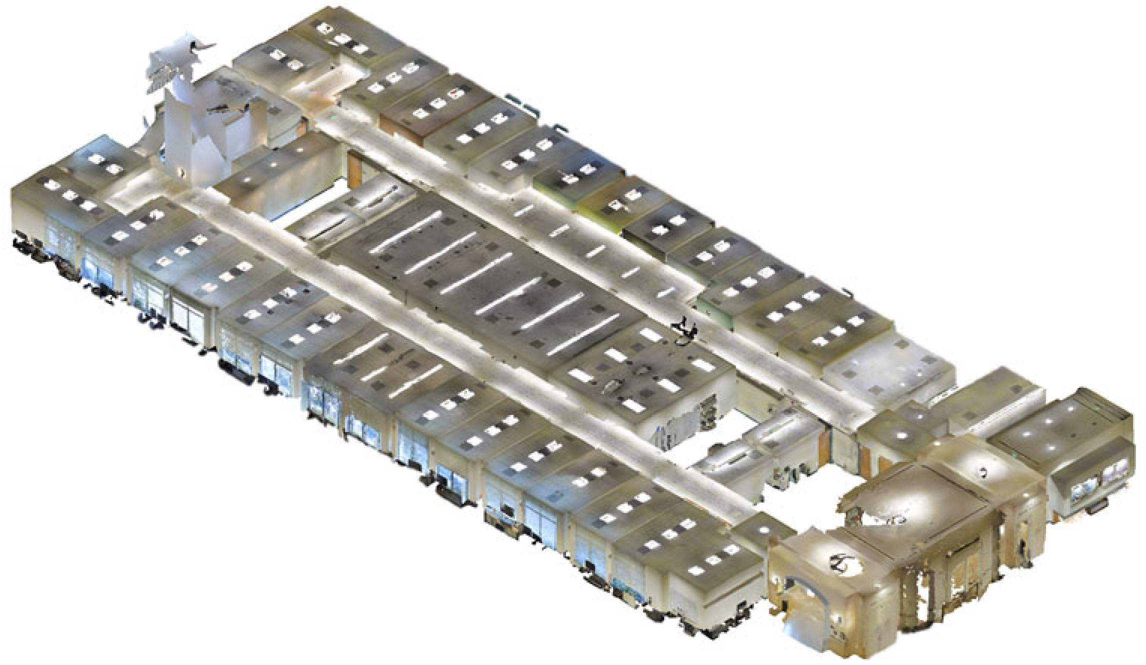
Area B



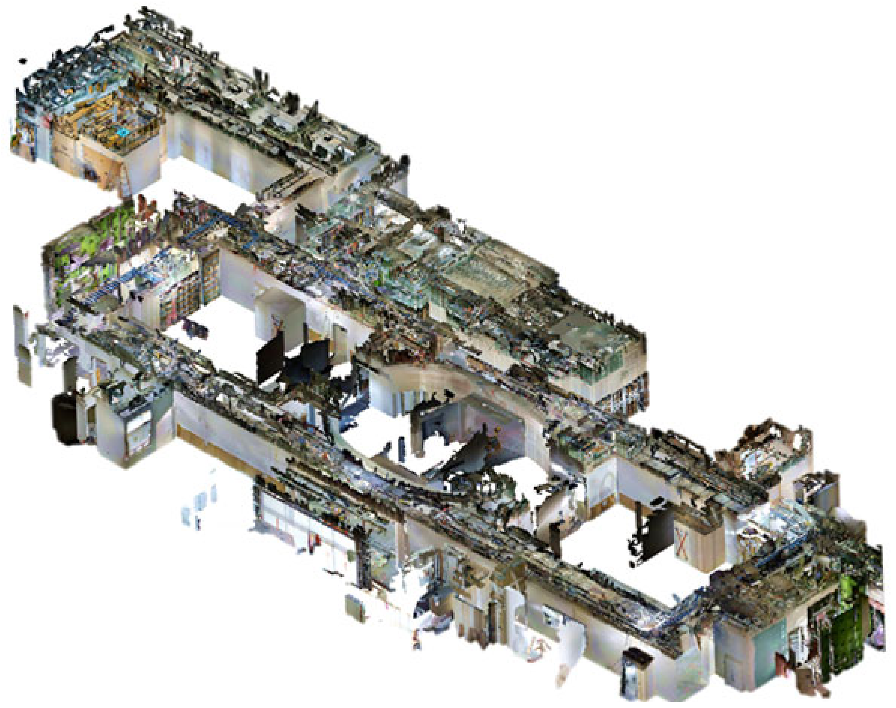
Area C



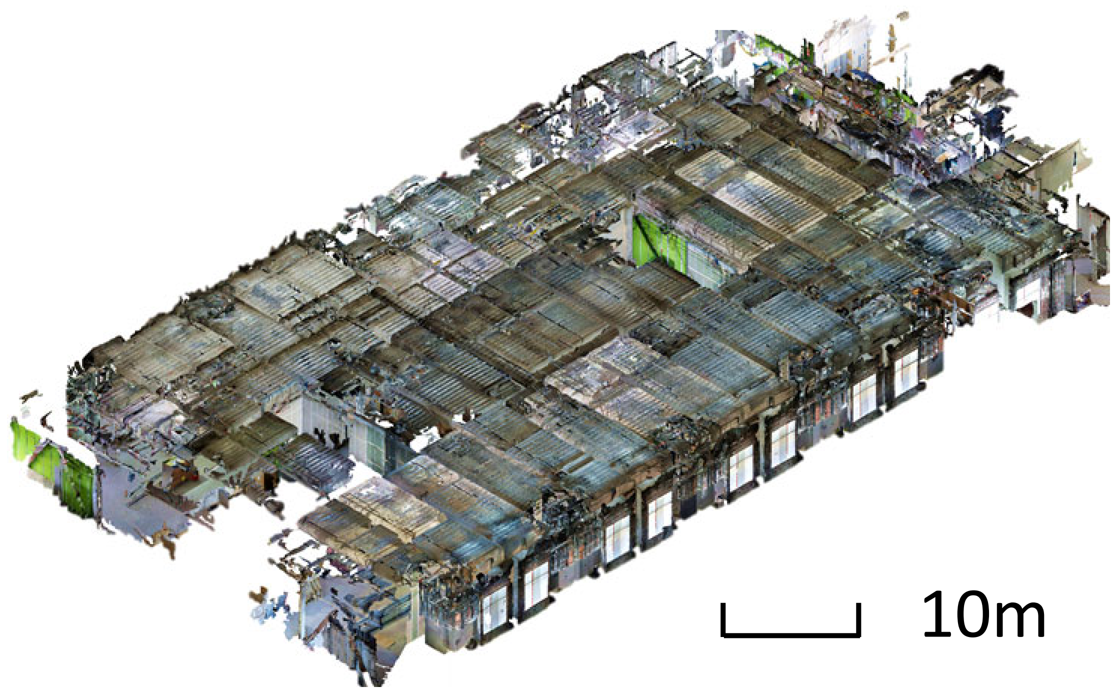
Area D



Area E

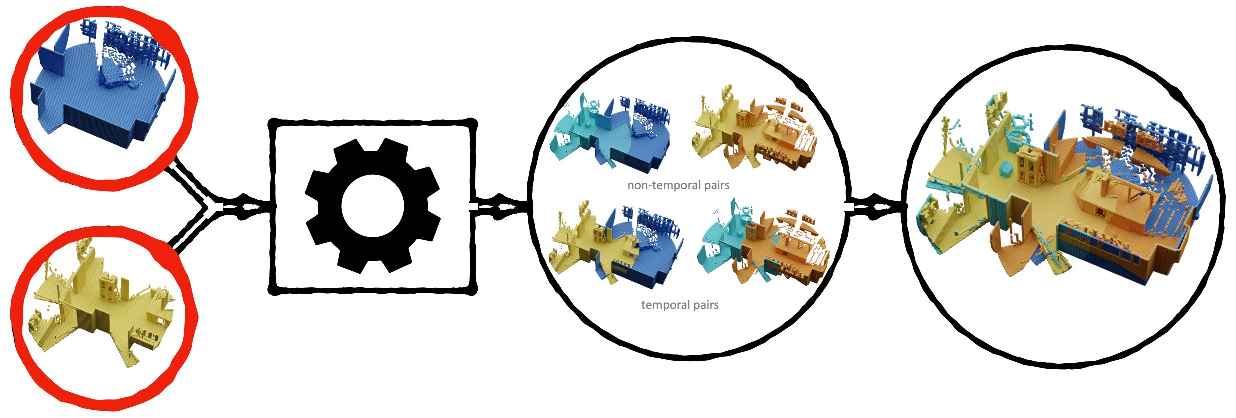


Area F

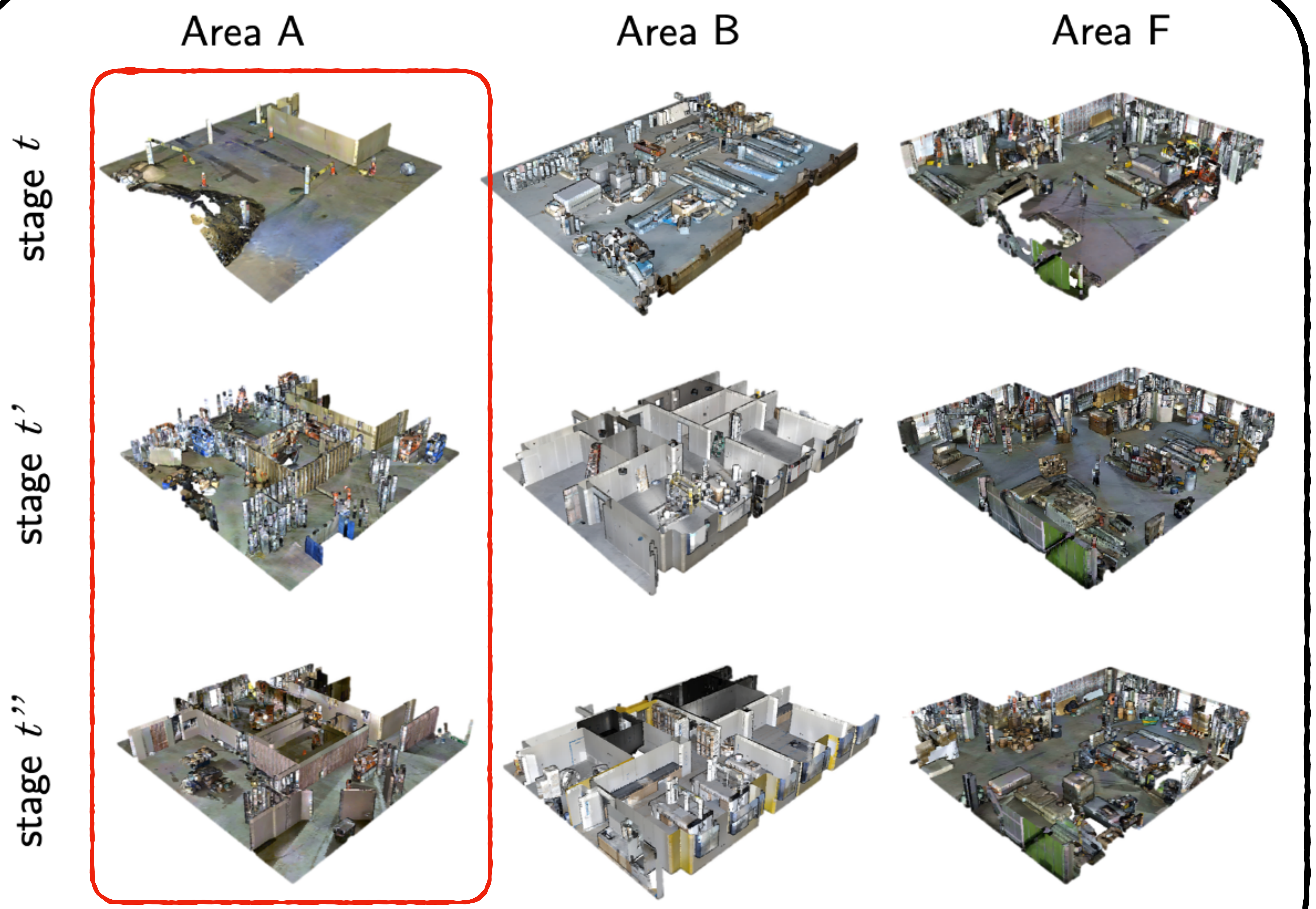


Total *spatiotemporal* area: 160,000 ft^2 (15,000+m^2)

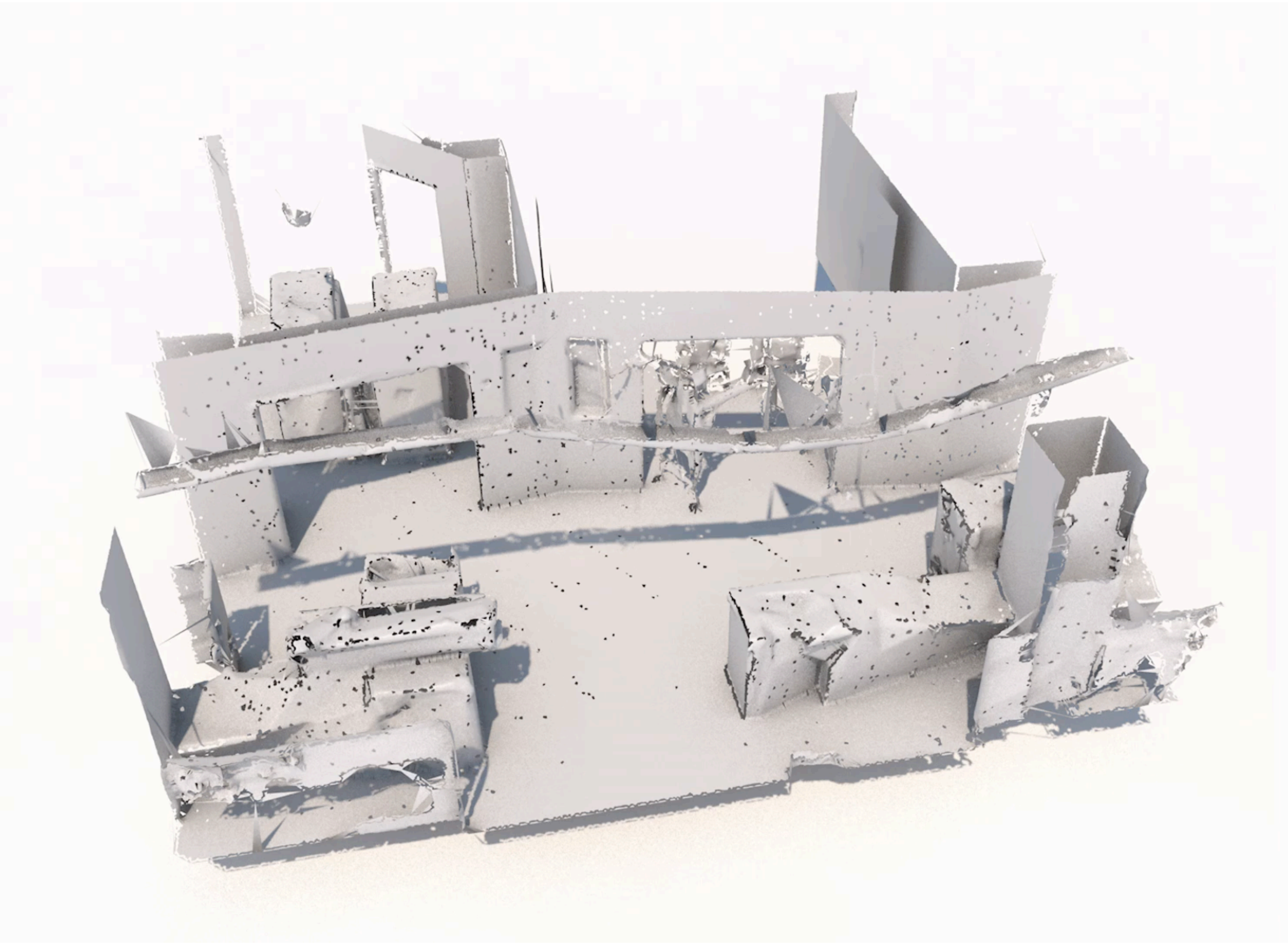
Nothing Stands Still



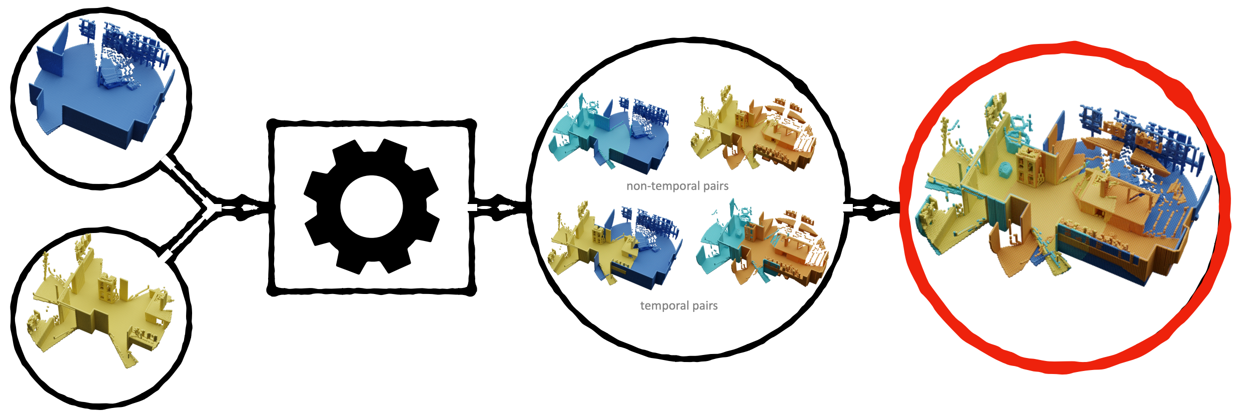
Construction sites demonstrate large changes in appearance, geometry, and topology



Sample areas of construction sites over different times



Sample area captured 4 times



Multi-way registration

Evaluation on Original Split

Ground Truth

Before Multiway

After Multiway

Living Scenes

Create & update replicas of geometry, semantics, & change using visual data*

Geometry-based	Scene Graph-based	Drastic Change
Living Scenes	**SGAligner**	**Nothing Stands Still**
Multi-object Relocalization and Reconstruction in Changing 3D Environments	3D Scene Alignment with Scene Graphs	A spatiotemporal benchmark on 3D point cloud registration

** while ensuring privacy and realistic implementations*

Living Scenes

Creating and updating representations of evolving indoor scenes

Iro Armeni

Long-Term Perception for Autonomy in Dynamic Human-shared Environments: What Do Robots Need? – IROS 2024

Stanford | Gradient Spaces
Civil and Environmental Engineering