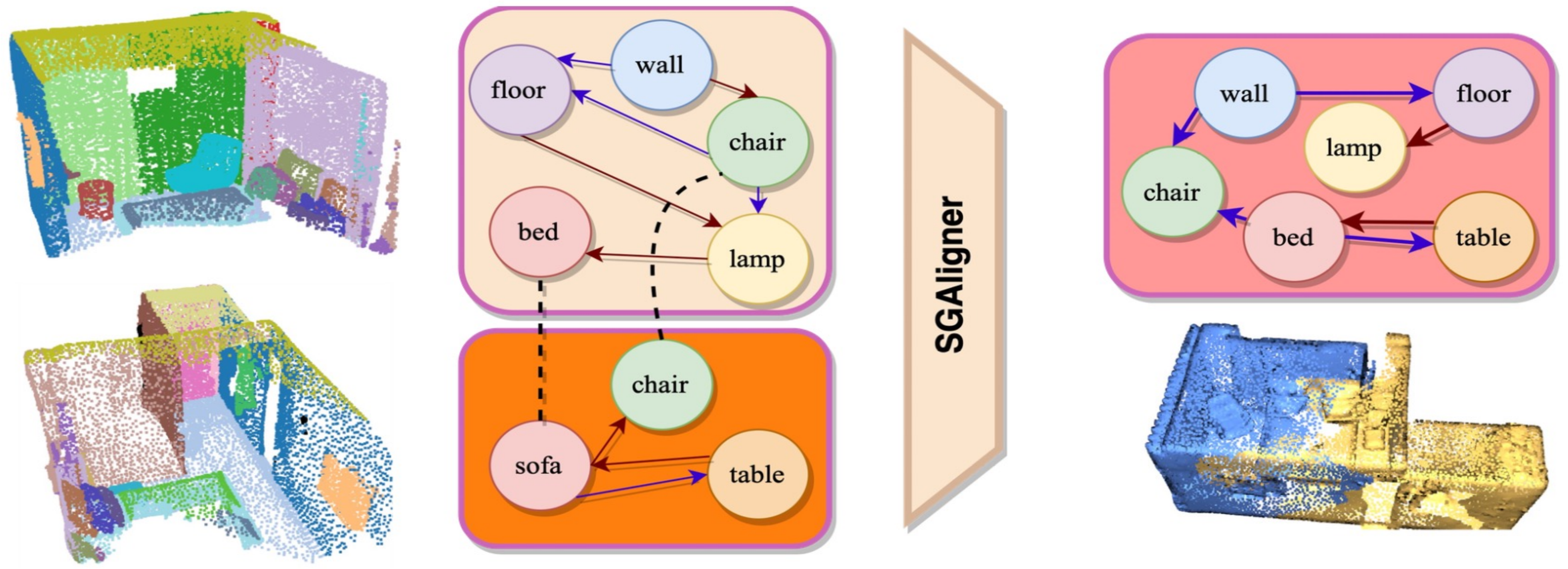
SGAligner : 3D Scene Alignment with Scene Graphs

Sayan Deb Sarkar[1], Ondrej Miksik[2], Marc Pollefeys[1,2], Daniel Barath[1], Iro Armeni[1]
[1]Deparment of Computer Science, ETH Zurich [2]Microsoft Mixed Reality & AI Lab, Zurich

Microsoft ETH zürich CVG Computer Vision and Geometry Lab
Project Webpage
ICCV23 PARIS

Problem Definition

Goal Scene alignment using multi-modal 3D scene graphs, for embodied agent operation



Standard Scene Alignment Challenges

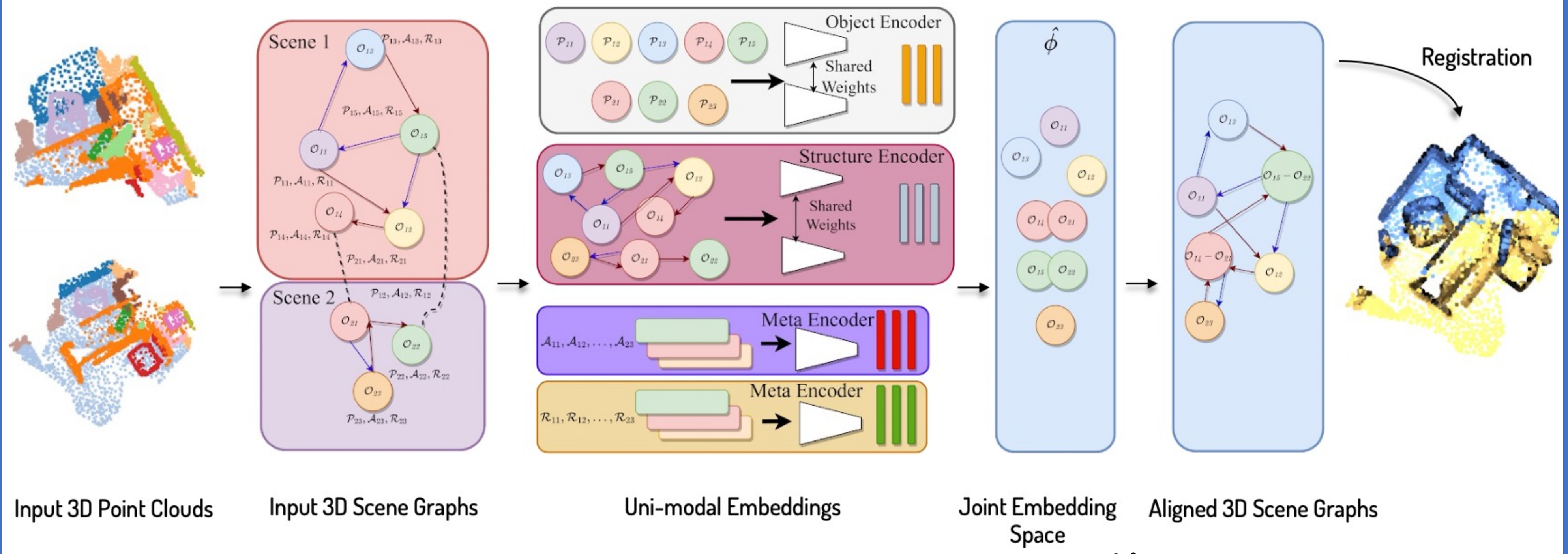
- Not robust to point cloud density/ extent
- Deciding if two scenes are non-overlapping
- Extra computation on raw metric space data
- $O(N^2)$ complexity for pairwise registration

Key Contributions

- Align 3D scene graphs in-the-wild
- Leverage results for 3D PC registration
- Robust to measurement noise
- Alignment of 3D scenes with changes

SGAligner Overview

- Inspired by works in entity alignment from language domain [1]
- Contrastive learning for interaction between unimodal and joint embedding spaces
- Operating on scene graphs is fast and robust in real-world scenarios
- Perform direct inference, without the need for any metric space data
- Map localization and registration in static + changed environments



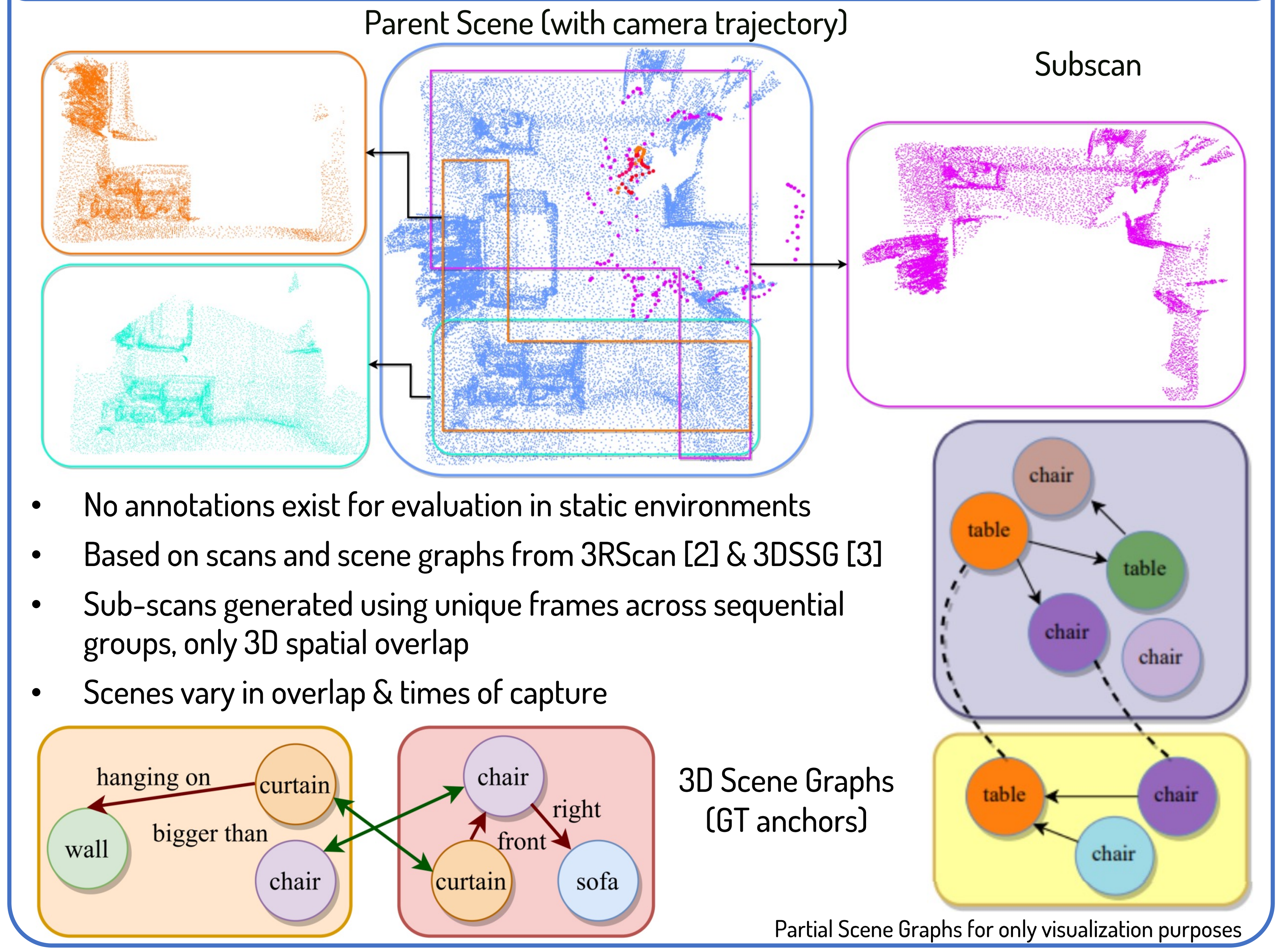
Embeddings

- Encode object, structure & relationship information into unimodal embeddings
- Joint embedding space to capture unimodal interactions

Alignment

- Match nodes to create aligned 3D scene graphs
- Use alignment for 3D point cloud registration

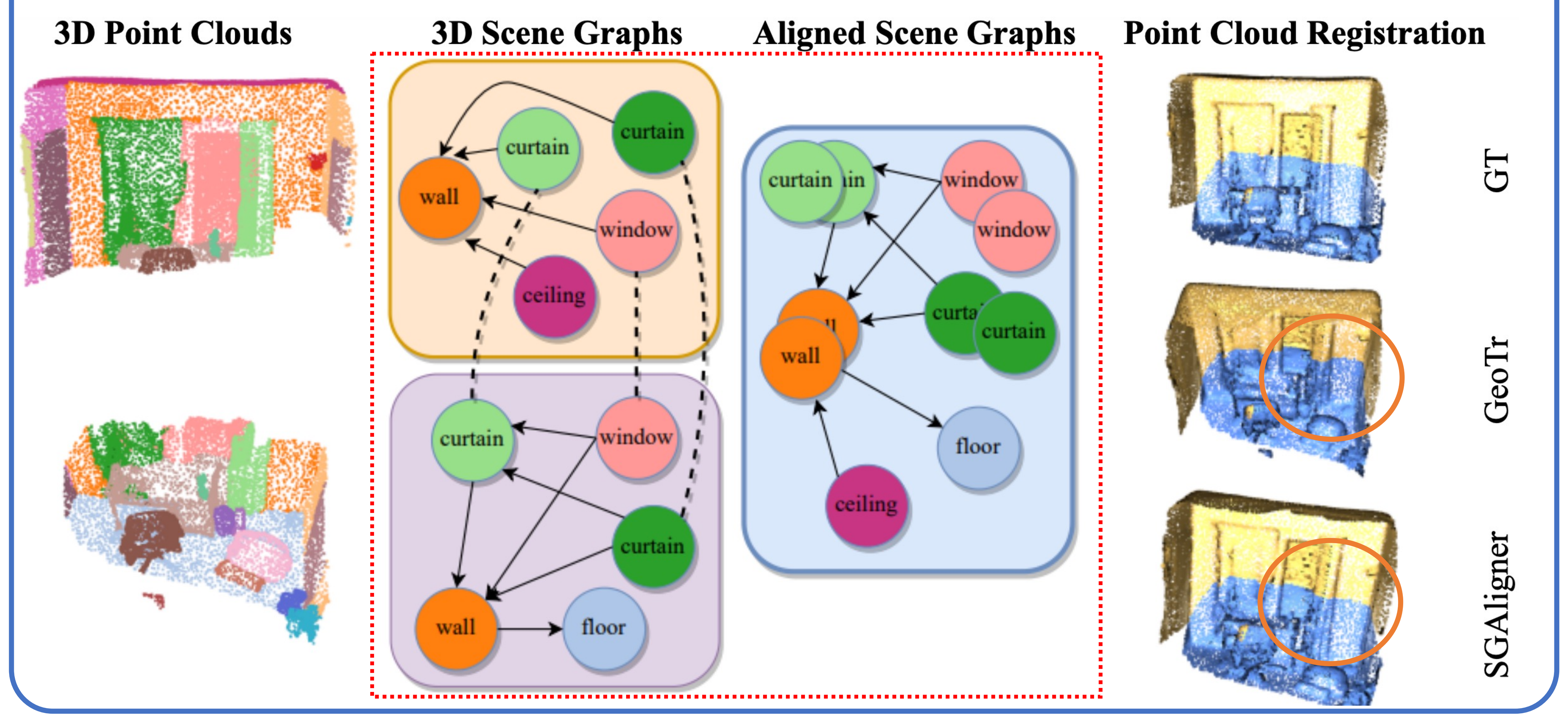
Dataset Generation



- No annotations exist for evaluation in static environments
- Based on scans and scene graphs from 3RScan [2] & 3DSSG [3]
- Sub-scans generated using unique frames across sequential groups, only 3D spatial overlap
- Scenes vary in overlap & times of capture

3D Scene Graph Alignment

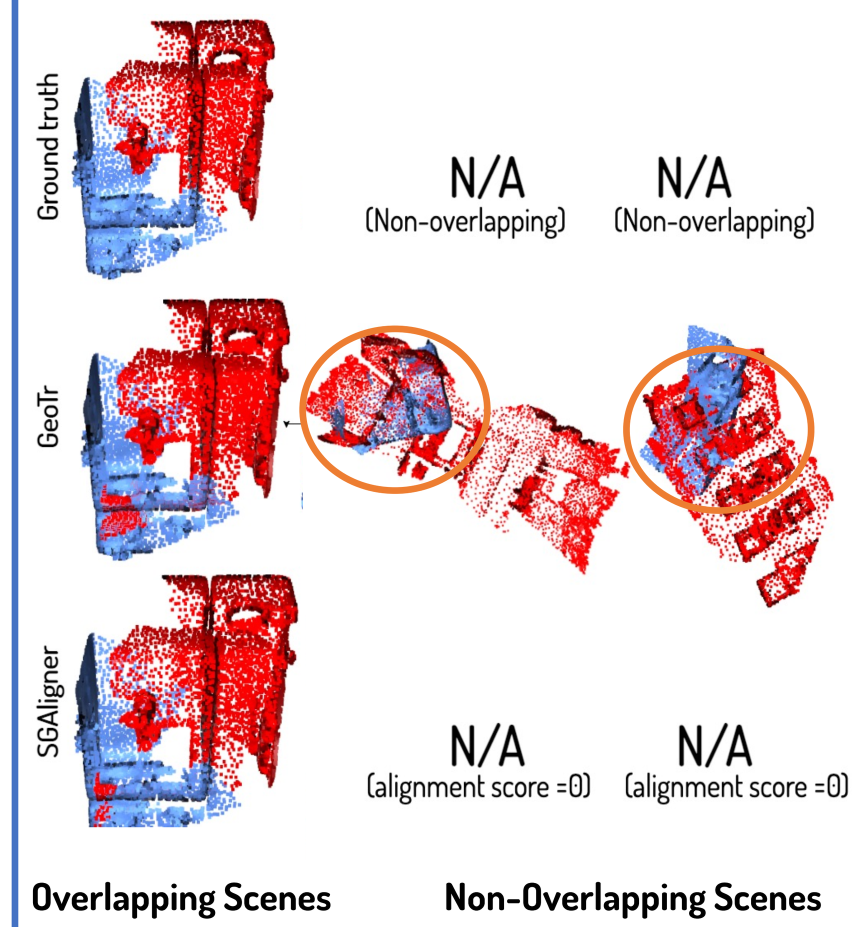
- Accurate matches despite input noise
- Meaningful results even in low spatial overlap
- Lightweight and privacy-aware, useful for autonomous agents building maps



Experimental Results

Key Results

- Performs registration on 3D correspondences from node-to-node matches
- Less sensitive to PC changes & incorrect matching techniques
- Better than standard registration methods in low overlap
- 3x faster to perform overlap check



Method	MRR ↑	Hits@1 ↑	Hits@3 ↑
w/ Ground Truth Scene Graphs			
EVA [4]	0.867	0.790	0.938
SGAligner	**0.950**	**0.923**	**0.974**
w/ Predicted Scene Graphs			
SGAligner	0.882	0.833	0.918

Table 1. **Evaluation on node matching**

Input Scene Graphs	Scene Graph Alignment Recall		
	R @ top-2	R @ top-50%	R @ All
Ground Truth	**0.964**	0.948	0.738
Predicted	**0.963**	0.856	0.450

Table 2. **Evaluation on 3D scene graph alignment**

Method	GeoTR[5]	K=1	K=2
CD ↓	0.02247	0.01677	**0.011111**
RRE ↓	1.813	1.425	**1.012**
RTE ↓	2.79	2.88	**1.67**
FMR ↑	98.94	99.85	**99.85**

Table 3. **3D Point Cloud Registration**

Method	GeoTR [5]	SGAligner
Precision	99.63	92.03
Recall	80.98	**90.94**
F1 Score	89.34	**91.48**
Time(ms)	442.50	**139.64**

Table 4. **Overlap Check for Point Cloud Registration**

Conclusion

- **First method** capable of aligning 3D scene graphs directly on the graph level
- Alignment directly helps improve **downstream tasks** (e.g., PC alignment) in terms of accuracy & speed
- Unlock agents to **leverage 3D scene graphs for creating maps** of the environments, further sharing and using it

[1] Lin. et. al, Multi-modal Contrastive Representation Learning for Entity Alignment, COLING 2022
[2] Wald. et. al., RIO: 3D Object Instance Re-Localization in Changing Indoor Environments, ICCV 2019
[3] Wald et. al, Learning 3D Semantic Scene Graphs from 3D Indoor Reconstructions, CVPR 2020
[4] Liu. et. al, Visual Pivoting For (Unsupervised) Entity Alignment, AAAI 2021
[5] Qin et. al, Geometric Transformer for Fast and Robust Point Cloud Registration, CVPR 2022