Hier-SLAM++: Neuro-Symbolic Semantic SLAM with a Hierarchically Categorical Gaussian Splatting

Authors: Boying Li, Vuong Chi Hao, Peter J. Stuckey, Ian Reid, and Hamid Rezatofighi

Workshop FMNS
ICRA ATLANTA 2025
IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION

Motivation

- **Semantic SLAM** jointly estimates ego-motion and global maps, enabling high-level robotic tasks in complex environments.
- **3DGS** is a novel representation beneficial for scene understanding.

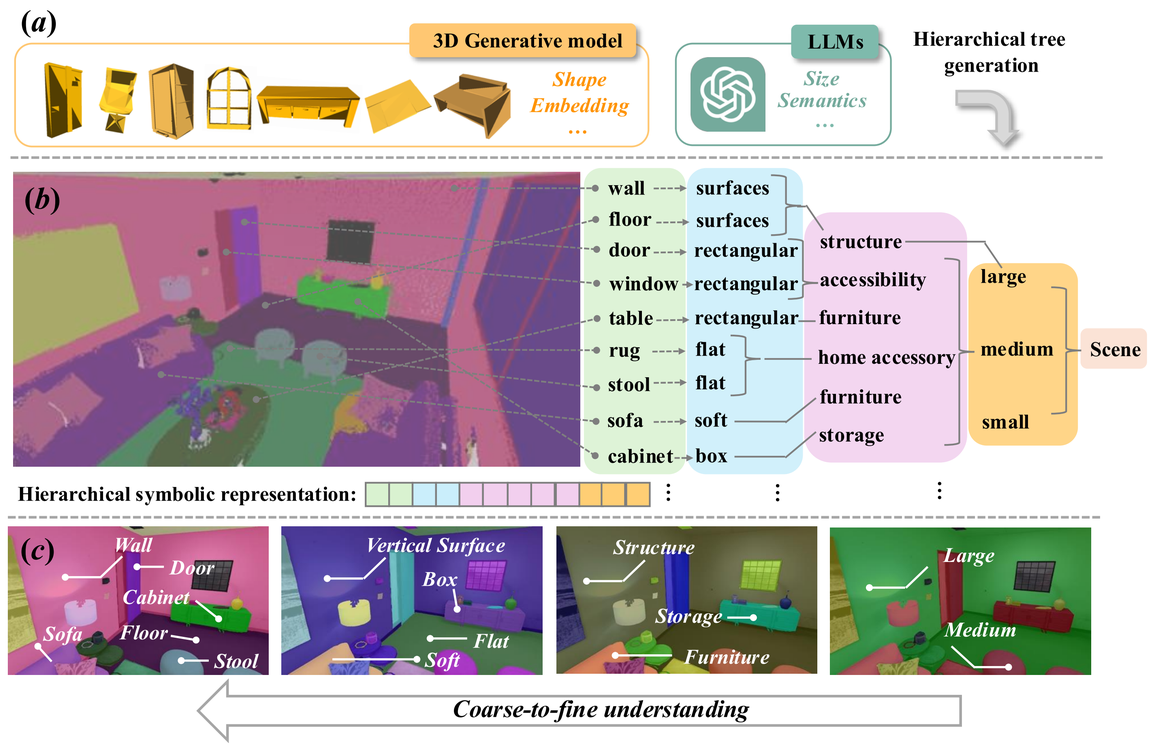
Main Challenges

- **Storage demands** and **processing time** increase significantly as the number of semantic classes grows.

Hier-SLAM framework

A) Semantic Representation:

- Semantics are inherently organized in a **hierarchical structure**. e.g., *wall* ← *plane* ← structure ← background ← scene *(see (b))*.
- We propose **a hierarchical representation** that combines embeddings from all levels:
  - **Embedding dimension** = max number of nodes per level.
  - **Significantly compacts** the original semantic space *(see (b), last row)*.
  - e.g., a 10-level binary tree can represent up to $2^{10} = 1024$ classes.
- Coarse-to-fine semantic understanding aligns with real-world observations — from distant to close-range views. *(see (c))*.
- Semantic information is **symbolically represented** and **learned in an end-to-end manner**.



B) Tree Generation with LLM and 3D Generative Models:

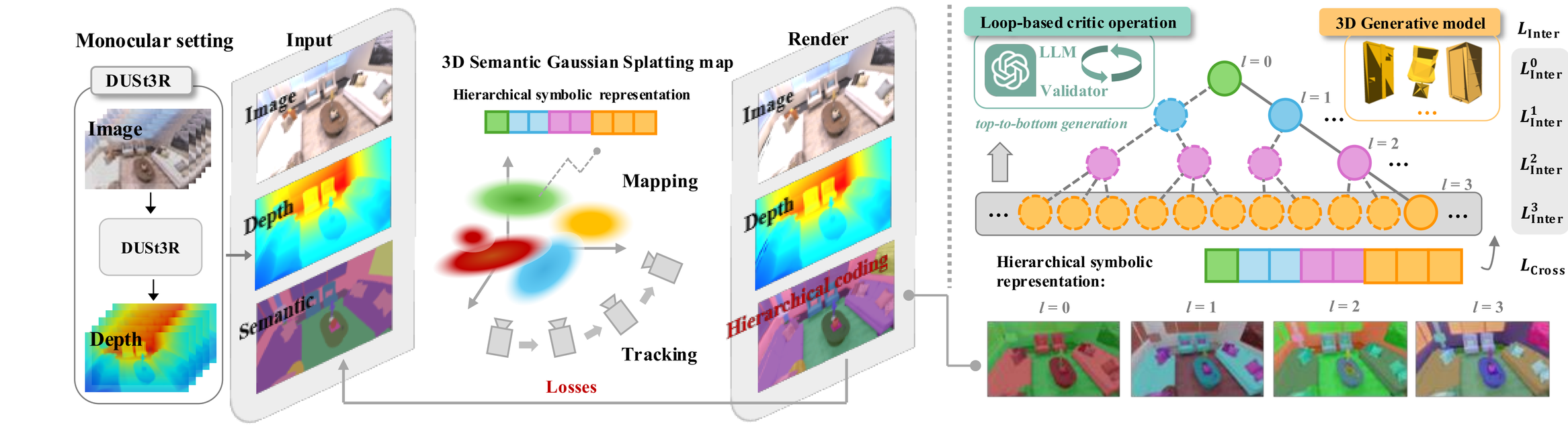
- We adopt **GPT-4 Turbo** for its strong reasoning and language capabilities.
  - The LLM **automatically** groups semantic classes **layer by layer** to form a hierarchical tree.
  - A **loop-based critic operation** validates and refines each clustering step.
- We also use **a text-to-3D Generative Model**
  - Semantic classes are clustered using **geometric embeddings** from the pre-trained model.
  - The LLM generates **language captions** for the clustering results.

C) Hierarchical Inter-level and Cross-level loss:

- **Inter-level**: CE loss at each hierarchy level.
- **Cross-level**: Refined MLP maps embeddings to original labels, followed by CE loss on original full classes.

D) Refine Semantic SLAM system:

- **3DGS**: Unifies forward and backward pipelines into a single module.
- An improved SLAM system supporting **both RGB-D and monocular** inputs using a **3D feed-forward model**.
- **Efficiency**: Achieves **2× faster** mapping and tracking speed.



Experiments

Results in Replica, ScanNet, TUM-RGBD datasets:

- Show **superior or on-par performance** in tracking, mapping, and semantic segmentation, with **2× operation speed-up**.
- **Scaling-up capability** to handle more than 500 semantic classes.

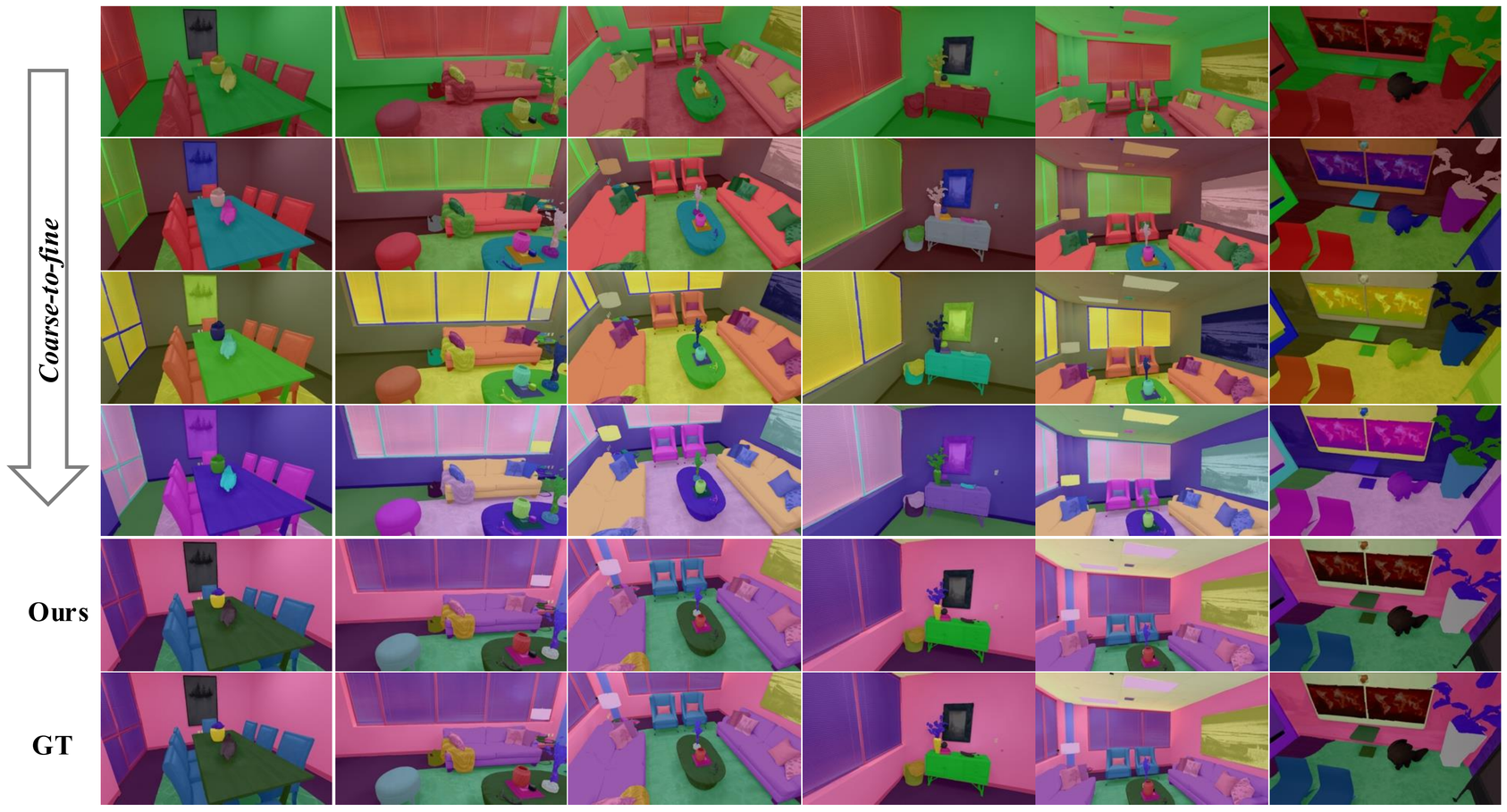


TABLE A.1
RGB-D TRACKING PERFORMANCE ATE RMSE (CM) ON THE REPLICA.
BEST RESULTS ARE HIGHLIGHTED AS **FIRST**, SECOND.

Methods	Avg.	R0	R1	R2	Of0	Of1	Of2	Of3	Of4
iMap [19]	4.15	6.33	3.46	2.65	3.31	1.42	7.17	6.32	2.55
NICE-SLAM [18]	1.07	0.97	1.31	1.07	0.88	1.00	1.06	1.10	1.13
Vox-Fusion [20]	3.09	1.37	4.70	1.47	8.48	2.04	2.58	1.11	2.94
co-SLAM [21]	1.06	0.72	0.85	1.02	0.69	0.56	2.12	1.62	0.87
ESLAM [22]	0.63	0.71	0.70	0.52	0.57	0.55	0.58	0.72	0.63
Point-SLAM [23]	0.52	0.61	0.41	0.37	0.38	0.48	0.54	0.69	0.72
MonoGS-RGBD [7]	0.79	0.47	0.43	0.31	0.70	0.57	0.31	0.31	3.2
SplaTAM [6]	0.36	0.31	0.40	0.29	0.47	0.27	0.29	0.32	0.55
SNI-SLAM [24]	0.46	0.50	0.55	0.45	0.35	0.41	0.33	0.62	0.50
DNS SLAM [2]	0.45	0.49	0.46	0.38	0.34	0.35	0.39	0.62	0.60
SemGauss-SLAM [25]	0.33	0.26	0.42	0.27	0.34	0.17	0.32	0.36	0.49
SGS-SLAM [26]	0.41	0.46	0.45	0.29	0.46	0.23	0.45	0.42	0.55
Hier-SLAM [8]	0.33	0.21	0.49	0.24	0.29	0.16	0.31	0.37	0.53
Hier-SLAM++ (one-hot)	**0.31**	**0.24**	**0.36**	**0.23**	**0.30**	**0.15**	**0.28**	0.39	0.51
Hier-SLAM++ (binary)	**0.31**	**0.23**	0.46	**0.23**	**0.29**	**0.15**	**0.27**	0.34	0.54

arXiv

Code Coming Soon.