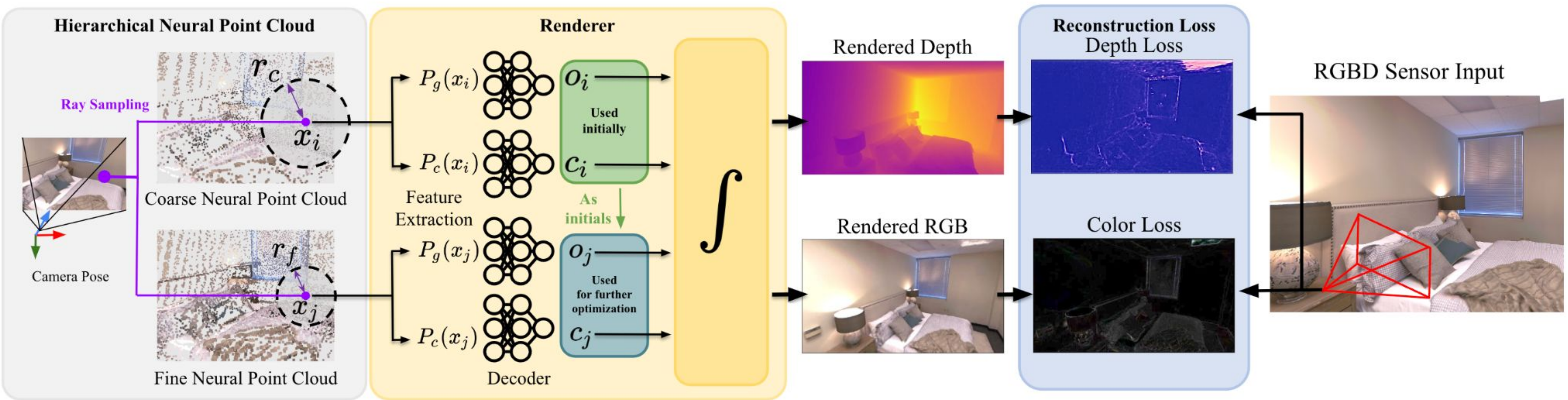
Hierarchical Dense Neural Point-based SLAM

Group members: Liuxin Qing[1], Longteng Duan[1], Guo Han[1], Shi Chen[1]
Supervisors: Erik Sandström[1], Martin R. Oswald[1, 2]
[1] ETH Zurich; [2] University of Amsterdam;



- ❏ A coarse-to-fine hierarchical neural point cloud with different dynamic sampling radius.
- ❏ In mapping: optimize the neural features in coarse and fine point clouds independently.
- ❏ In tracking: begin by optimizing camera pose using coarse-level features, and subsequently integrate fine-level features for more refined enhancements.

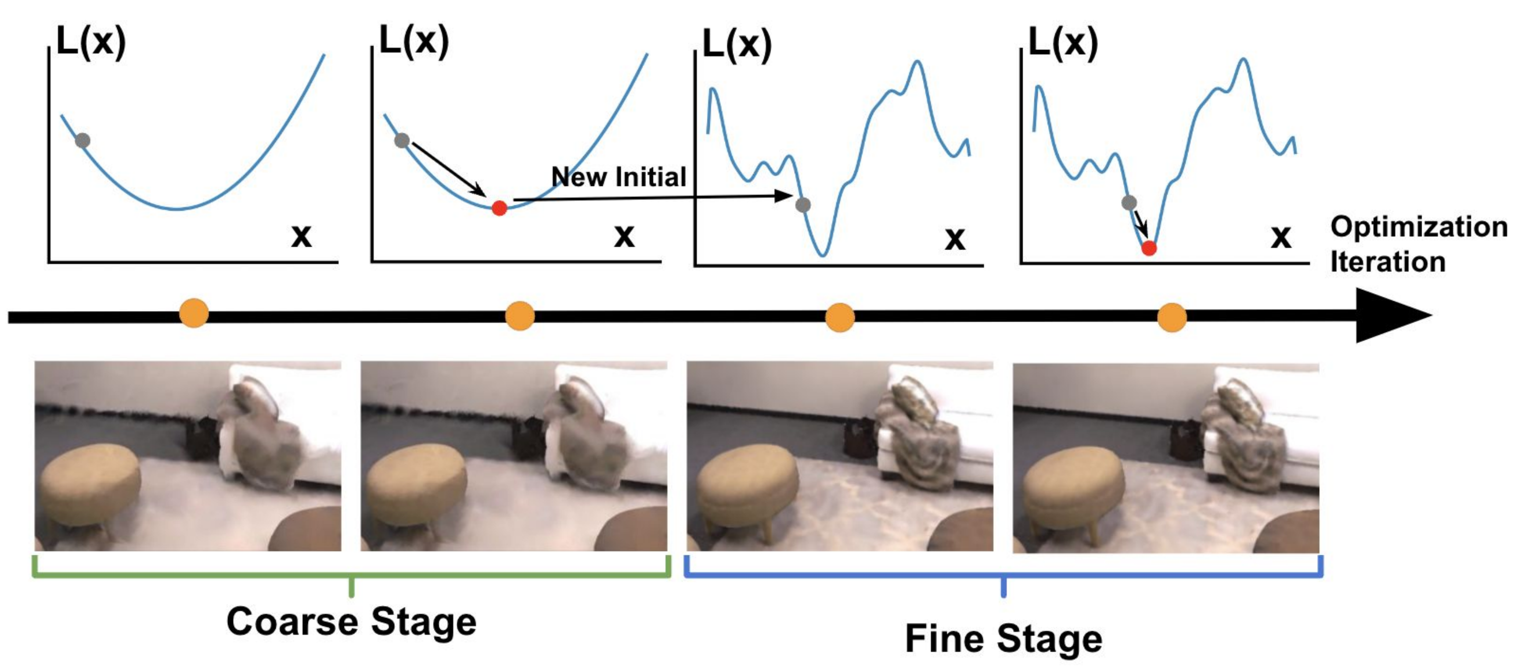
1 Introduction

• Background

Point-SLAM* is an approach developed for real-time capable SLAM using point-based neural scene representations. While Point-SLAM offers reliable performance in synthetic settings, it may suffer from unstable camera pose tracking in certain real-world conditions.

• Motivation

We aim to improve the robustness of camera pose tracker in real-world scenes using a coarse-to-fine method to prevent the tracker from being stuck at local minima. The figure below gives an intuition. The tracking process takes a multi-stage strategy, where the coarse stage gives a better initialization for the fine one.

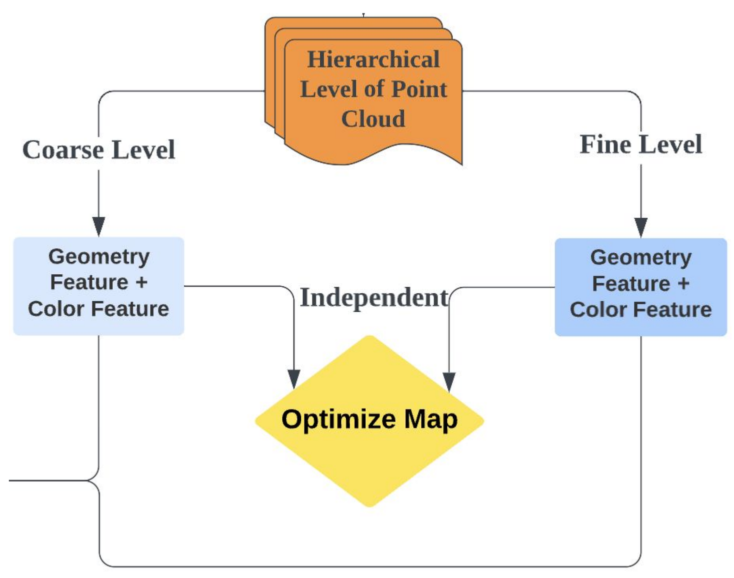


• Contributions

We have successfully integrated coarse neural features into the original Point-SLAM pipeline. It achieves comparable tracking performance on Replica dataset, as well as evident improvements on several ScanNet scenes.

• Independent (adopted)

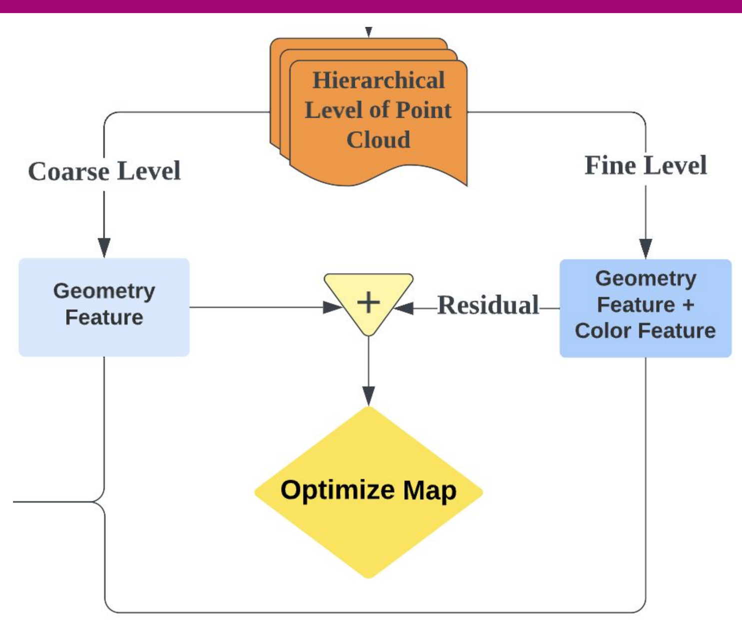
- ❏ Each point cloud as a complete scene representation, with features optimized independently.
- ❏ Proved to improve tracking stability.



2 Pipeline Design

• Residual

- ❏ Fine-level features as a complementary residual to coarse-level features.
- ❏ However, exacerbate tracking instability.



3 Results and Discussion

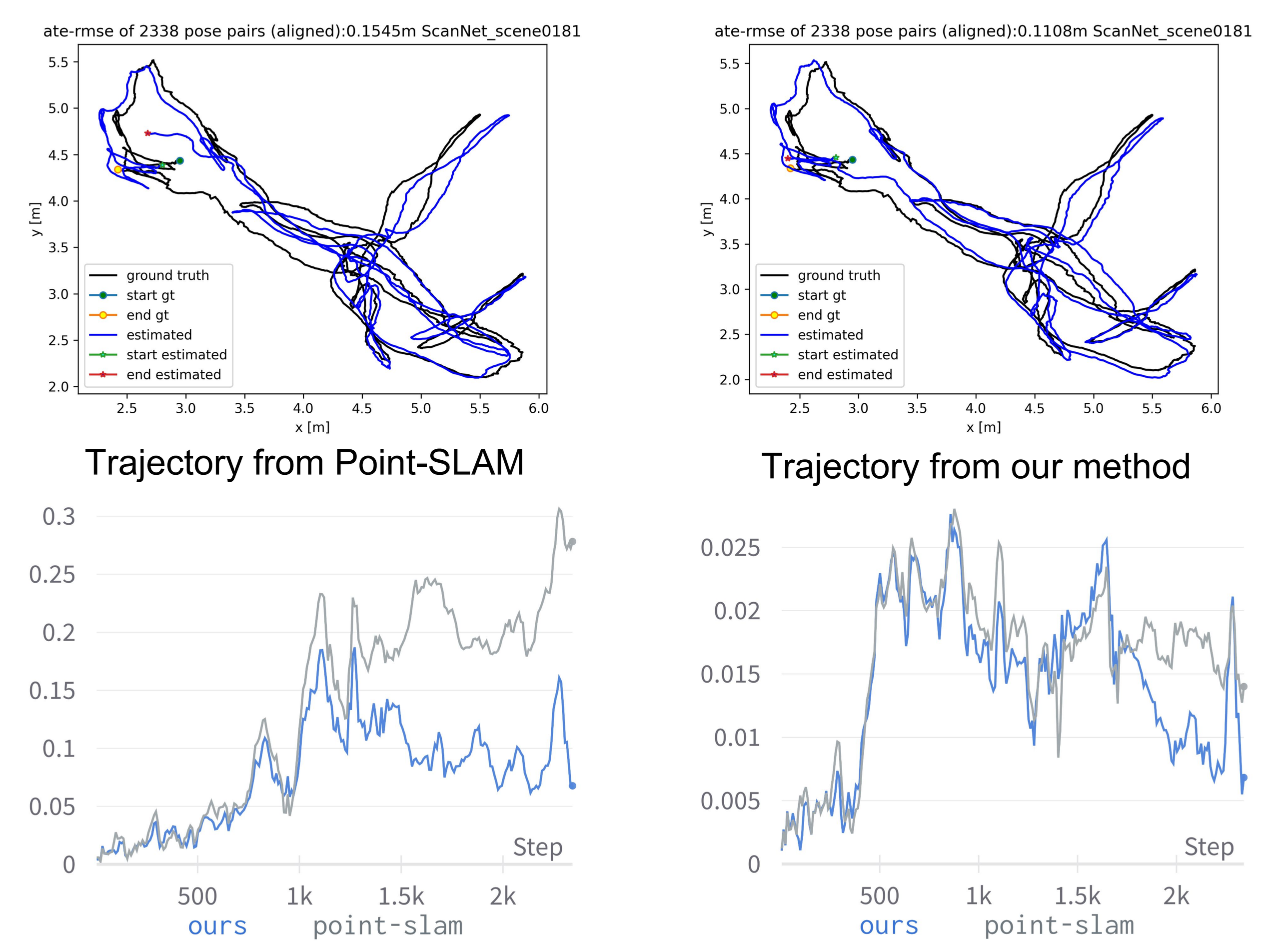


Figure: Camera tracking result for ScanNet scene 0181.

Scene	0025	0059	0062	0103	0106	0181	0207	Avg.
NICE-SLAM	10.11	14.00	4.59	4.94	7.90	13.40	6.20	8.73
Point-Slam	8.35	8.29	3.48	6.97	11.86	17.30	8.21	9.21
Ours	**6.13**	**7.06**	**3.54**	**6.73**	**12.99**	**11.08**	**4.53**	**7.44**

Table: Tracking performance (ATE-RMSE↓ [cm]) on ScanNet scenes. Best results are highlighted as **first**, **second**, and **third**.

- ❏ Average ATE-RMSE of our method is lower than baselines.
- ❏ Comparisons of tracking trajectories and camera positional/ rotational error illustrate the superiority of our method.

References

* Sandström, Erik, et al. "Point-SLAM: Dense Neural Point Cloud-based SLAM." *arXiv preprint arXiv:2304.04278* (2023).
* Zhu, Zihan, et al. "Nice-slam: Neural implicit scalable encoding for slam." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2022.