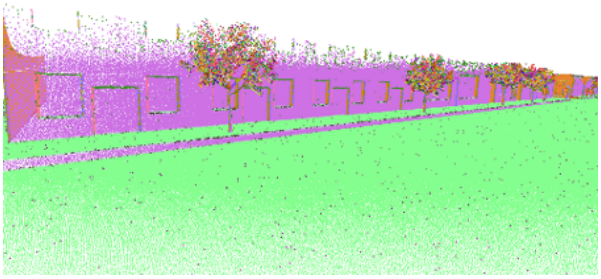
NeRF-LOAM: Neural Implicit Representation for Large-Scale Incremental LiDAR Odometry and Mapping Supplementary Material

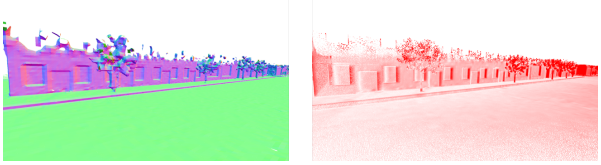
A. Simultaneously Odometry & Mapping Results

We present in Fig. 11 and Fig. 12 our odometry mapping results compared with Puma [36] herewith the corresponding ground truth map. To reconstruct more complete results, Puma uses polynomial function and normals to simulate the surrounding meshes, resulting in loss of detail (e.g., windows, curb) and unreal reconstruction (e.g., the connection of tree and wall), while our reconstruction can provide higher accuracy and neater results.

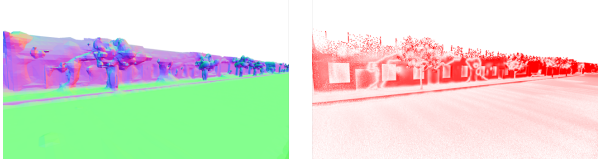
We also present in Fig. 13 and Fig. 14 the mapping results on MaiCity dataset of SHINE-Mapping [50] and Vdbfusion [37] combined with KissICP [38] odometry. These results also demonstrate that ours can provide a complete and smooth map.



(a) Ground Truth point cloud map



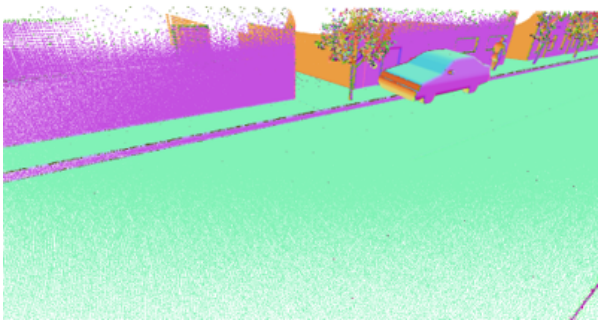
(b) Ours Odometry mapping result with error map



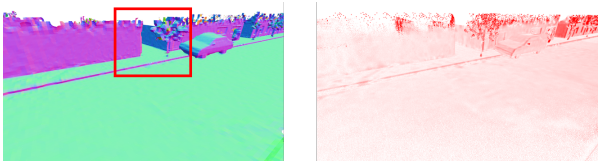
(c) Puma Odometry mapping result with error map

Figure 11. The simultaneously odometry & mapping results with error maps on the MaiCity [36] dataset. The error maps are with the ground truth map as a reference, where the redder points mean larger errors up to 20 cm.

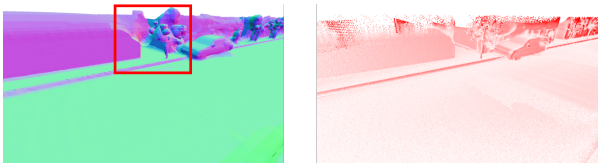
To demonstrate our odometry and mapping under large-scale environments. We provide in Fig. 15 the qualitative results of odometry mapping on the KITTI [11] dataset.



(a) Ground Truth point cloud map

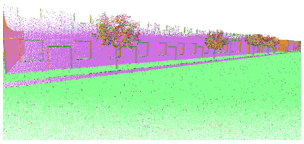


(b) Ours Odometry mapping result with error map



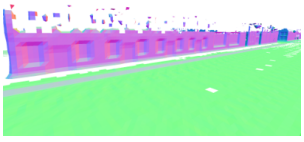
(c) Puma Odometry mapping result with error map

Figure 12. The simultaneously odometry & mapping results with error maps on the MaiCity [36] dataset. The error maps are with the ground truth map as a reference, where the redder points mean larger errors up to 20 cm. The red box illustrates the unreal reconstruction.

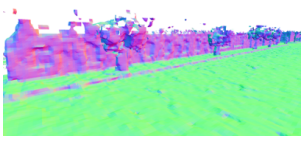


(a) Ground Truth point cloud map

(b) Ours with KissICP



(c) Vdbfusion with KissICP

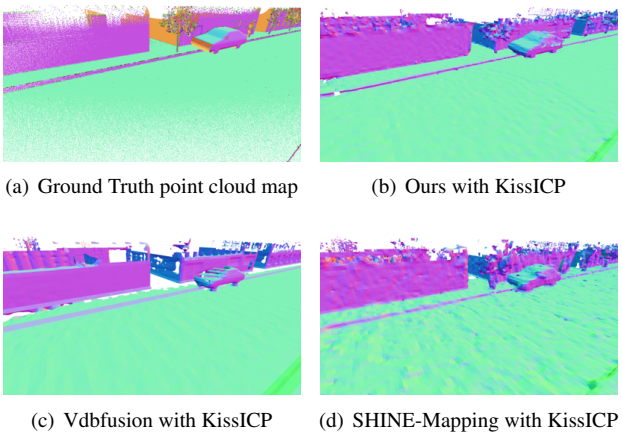


(d) SHINE-Mapping with KissICP

Figure 13. The mapping results combined with KissICP[38] odometry on the MaiCity [36] dataset.

Method	00		01		03		04		05		07		09		10	
	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}
ICP-po2po [3]	6.88	2.99	11.21	2.58	11.07	5.05	6.64	4.05	3.97	1.93	5.17	3.35	6.93	2.89	8.91	4.47
ICP-po2pl [30]	3.80	1.73	13.53	2.58	2.72	1.63	2.96	2.58	1.55	1.42	1.55	1.42	3.95	1.71	6.13	2.60
GICP [31]	**1.29**	0.64	4.39	2.58	1.68	1.08	3.76	1.93	1.02	0.54	0.64	0.46	1.97	0.77	**1.31**	**0.62**
SUMA [2]	2.93	0.92	4.05	1.22	1.43	**0.75**	11.90	1.06	1.46	0.79	1.75	1.17	1.92	0.78	1.81	0.97
PUMA(NN)[36]	2.15	1.14	4.32	1.04	**1.34**	1.07	2.09	1.46	1.56	1.07	1.88	1.36	1.80	0.82	2.24	1.67
PUMA(RC)[36]	1.55	0.74	3.38	1.00	1.60	1.10	**1.63**	**0.92**	**1.20**	**0.61**	0.72	0.55	1.51	0.66	1.38	0.84
DeLORA [26]	-	-	-	-	-	-	-	-	-	-	-	-	9.07	3.14	6.53	4.22
DeepPCO [41]	-	-	-	-	-	-	2.63	3.05	-	-	-	-	-	-	2.21	1.67
LONet [15]	1.47*	0.72*	1.36*	0.47*	1.03*	0.66*	0.51*	0.64*	1.04*	0.69*	1.70	0.89	1.37	0.58	1.80	0.93
PWCLONet [39]	0.78*	0.42*	0.67*	0.23*	0.76*	0.44*	0.37*	0.40*	0.45*	0.27*	**0.60**	**0.44**	**0.79**	**0.35**	1.69	**0.62**
Ours	1.34	**0.54**	**2.07**	**0.52**	2.22	1.57	1.74	1.00	1.40	0.65	1.00	0.65	1.63	0.57	2.08	0.69

Table 5. The odometry results on KITTI dataset [11]. t_{rel} and r_{rel} mean the average translational RMSE (%) and rotational RMSE (°/100 m) respectively on all possible subsequences in the length of 100, 200, ..., 800 m. "*" means the results on the training sequence, "-" means not provided, RC for ray casting, NN for nearest neighbor. The best results are bold whereas the results on training sequences are beyond our consideration.



(a) Ground Truth point cloud map (b) Ours with KissICP

(c) Vdbfusion with KissICP (d) SHINE-Mapping with KissICP

Figure 14. The mapping result combined with KissICP[38] odometry on the MaiCity [36] dataset.

Dataset	Grd	KS	GT	RMSE ↓	Acc. ↓	Comp. ↓	C-l1. ↓	F↑
MaiCity	✗	✗	✗	0.20	6.15	69.64	37.90	49.39
	✗	✓	✗	0.20	6.13	70.48	38.30	48.78
	✓	✗	✗	0.17	5.93	11.49	8.71	76.15
	✓	✓	✗	0.17	5.69	11.23	8.46	77.26
	✗	✗	✓	-	3.57	5.61	4.59	90.61
	✗	✓	✓	-	3.43	5.40	4.42	90.81
	✓	✗	✓	-	3.27	5.03	4.15	92.80
	✓	✓	✓	-	3.15	4.84	4.00	92.96
Newer College	✗	✗	✗	Failed	-	-	-	-
	✗	✓	✗		-	-	-	-
	✓	✗	✗	0.15	16.41	25.75	21.08	61.10
	✓	✓	✗	0.15	12.89	22.21	17.55	74.37
	✗	✗	✓	-	7.01	15.58	11.29	91.58
	✗	✓	✓	-	6.73	14.86	10.79	91.92
	✓	✗	✓	-	7.50	16.75	12.13	90.98
	✓	✓	✓	-	6.86	15.62	11.24	91.84

Table 6. Ablation study of our designs on Maicity [36], Newer College [29]. "-" stands for no meaning data. "Grd" means the ground separation, "KS" means the key-scan refine, and GT for application of ground truth pose

Our method can obtain a fine reconstruction of city environments without a loop. However, for trajectories with a loop, our method cannot maintain a consistent global map.

B. Mapping Quality

The ground truth pose is used in this section to compare our mapping ability with SHINE-Mapping [50] and Vdbfusion [37]. As the detailed reconstruction results are similar to the results of Sec. A, we provide in Fig. 16 the bird-eye view of reconstruction on Maicity [36] dataset. As we can see, SHINE-Mapping provides a relatively complete map but is not smooth enough. While Vdbfusion provides the smoothest map but the map is not complete. Our mapping process can provide the most complete and smooth result.

C. Odometry Evaluation on KITTI Dataset

In this section, we present the odometry evaluation on KITTI [11] dataset. As can be seen from Tab. 5, our odometry results show comparative performance compared to the non-learning-based method and outperform them on some sequences. Compared to the learning-based method, our method does not need to be pre-trained by numerous labeled data, and it can be directly employed in other environments, where some existing learning-based methods fail. This is important when we lack adequate data and ground truth labels or explore unknown environments. We also present our qualitative results on Fig. 17. Our odometry process shows the ability of generalization on different sequences and large-scale environments.

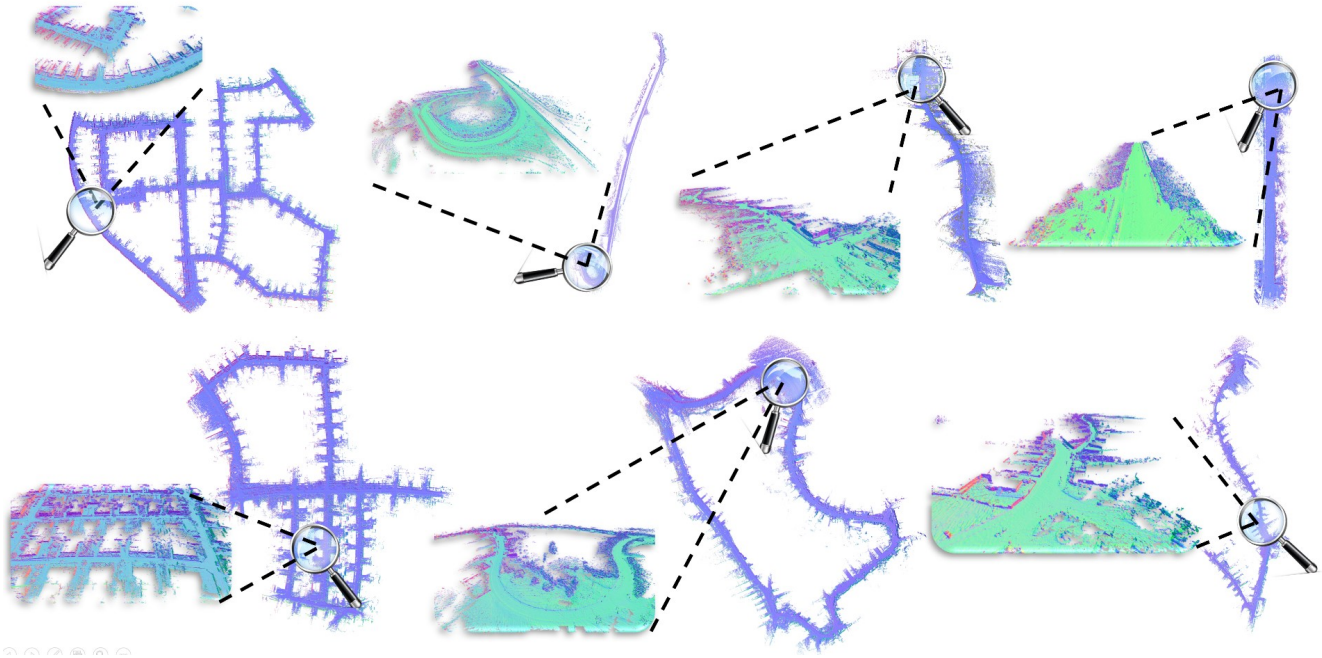
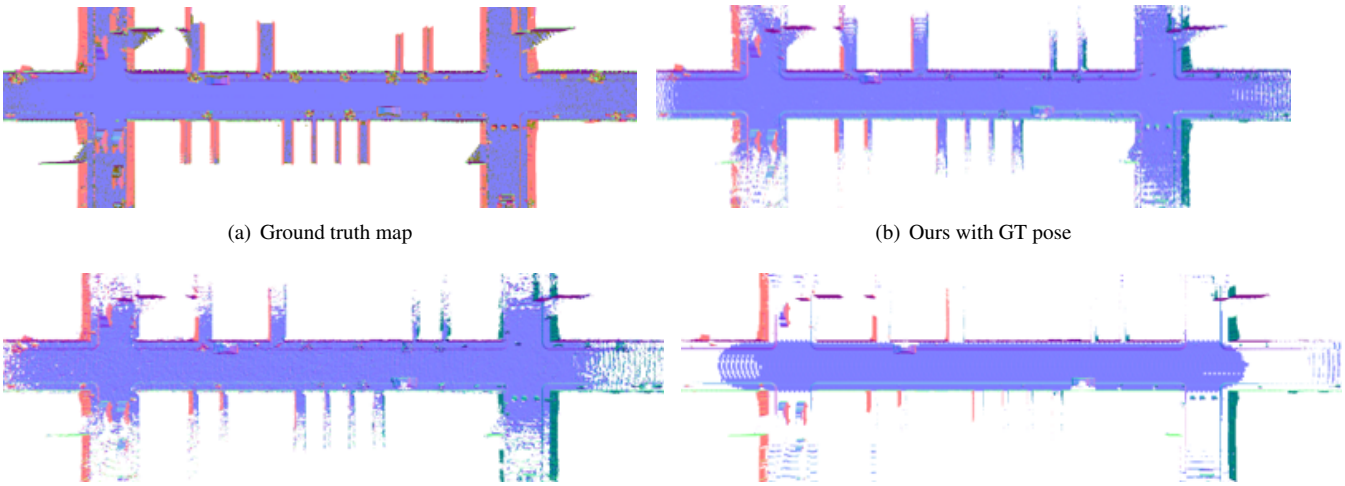


Figure 15. The qualitative result of our odometry mapping on KITTI [11] dataset. From left upper to right bottom, we list the results of sequences 00, 01, 03, 04, 05, 09, 10



(a) Ground truth map

(b) Ours with GT pose

(c) Shine With GT pose

(d) Vdbfusion with GT pose

Figure 16. The mapping result with ground truth pose on the MaiCity [36] dataset. We present the bird-eye view of the results, indicating that our method can reconstruct a complete and smooth map.

D. Additional Ablation Study

We show the full table of ablation study on Tab. 6 concerning the ground separation, key-scan refine, and application of ground truth pose. First, the ground separation can directly improve the odometry result, especially at the z-axis, where the ground separation takes effect. The qualitative results in Fig. 18 also prove its indispensable, and we can see that our method stays consistent at the z-axis on both datasets. Second, the key-scan refine can greatly improve the mapping quality when no ground truth pose is applied. Although this improvement becomes slight when the ground truth pose is applied, we still adopt this strategy as it can help us reconstruct a smooth and complete map. Third, the ground truth pose plays a significant role in mapping, especially for loops, which usually cause overlapping

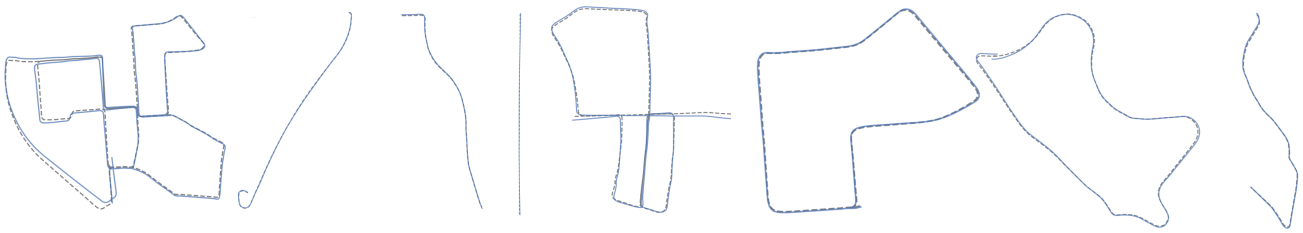
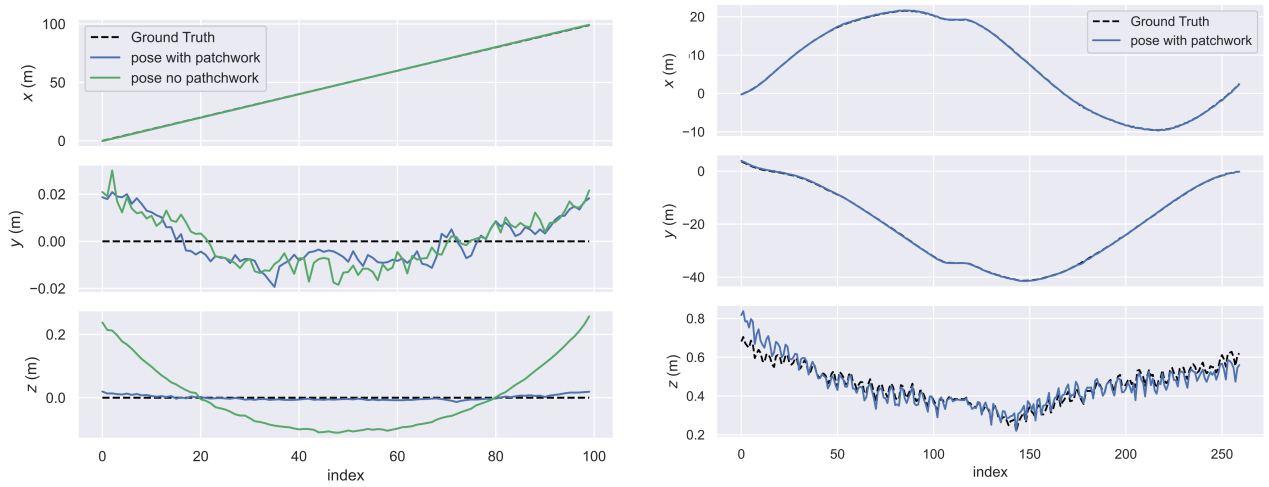


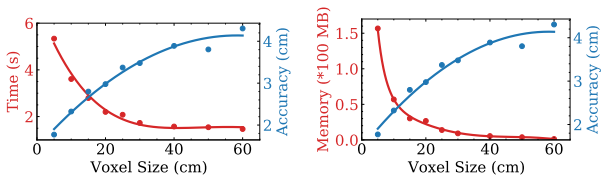
Figure 17. The qualitative results of our odometry on KITTI [11] dataset. From left to right, we list the results of sequences 00, 01, 03, 04, 05, 07, 09, 10. The dashed line corresponds to the ground truth and the blue line to our odometry method.



(a) Trajectory for MaiCity Dataset[36]

(b) Trajectory for Newer College Dataset[29]

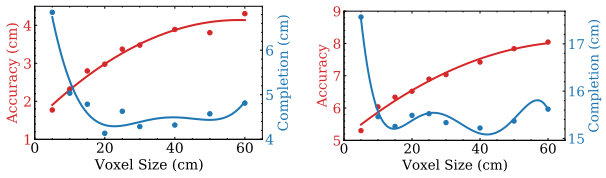
Figure 18. Ablation study for ground separation in terms of trajectory. The blue line is the trajectory with ground separation, and the green line is the one without ground separation. The dashed line represents the ground truth trajectory.



(a) Time vs Acc. on Maicity

(b) Mem. vs Acc. on Maicity

Figure 19. Study on voxel size v.s. processing time, memory consumption and accuracy distance on Maicity [36].



(a) Maicity

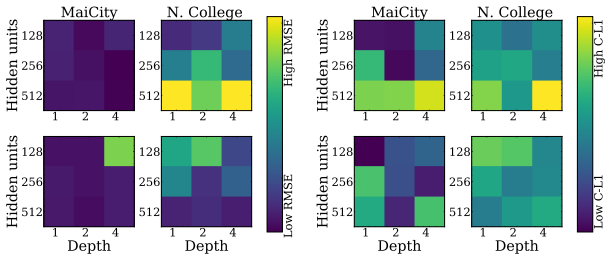
(b) Newer College

Figure 20. Study on voxel size v.s. accuracy distance and completion distance on Maicity [36] and Newer College [29]

meshes. Dealing with loop detection is an important task for our future work.

We then complement in Fig. 19 the effect of voxel size on Maicity[36] dataset on the processing time, accuracy distance, and memory consumption. The two lines cross at voxel size between 15 cm and 20 cm. We choose 20 cm as our choice for the reason that the processing time still decreases a lot while the accuracy remains. As indicated in Fig. 20, we find that the Chamfer-L1 distance stays almost invariant, as the completion distance decrease with a larger voxel size. A smaller voxel size brings finer reconstruction while a larger voxel size can make it more complete. Similarly, we choose 20 cm as the voxel size since the completion distance stays almost constant.

We explore here the influence of network architecture (i.e., hidden units and depth) and embedding length. Figure 21 show the normalized result of RMSE, Chamfer-L1 dis-



(a) Trajectory Accuracy (b) Mapping Accuracy

Figure 21. Ablation study for Network architecture and Embedding length. First low represent 8-embedding length, and second row is 16-length.

tance for various embedding length and network. During our study of the processing time, we found that a more profound and more hidden units network consumes more time while the embedding length affects little. We can also find from the figure that although a short embedding length can sometimes achieve exceptional results, it is unstable with the change of network. We choose 16 as our embedding length as it generalizes well and does not lower time efficiency. For the network, we use 2 layers deep and 256 hidden units architecture for it performs well in tracking as well as the mapping on two datasets while keeping time efficiency.

We use PCA to calculate the normals for points, where the eigenvector associated with the smallest eigenvalue is treated as normal. Considering efficiency and complex non-ground environments, we only apply normal rectification specifically for ground points (denoted as "ground") rather than for all points (denoted as "all"). We also use the cosine of the angle between normals and viewing angle as the weight to reflect the uncertainty (denoted as "weight"). As shown in Tab. 7, applying normal rectification to all points or using the weighted method, have no improvement in LOAM performance, while the cost time increases significantly.

	Time (ms)↓	RMSE↓	Acc.↓	Comp.↓	C-l1.↓	F↑
Ground	**127**	**0.10**	**3.04**	**5.21**	**4.12**	**92.9**
Weight-ground	**127**	**0.10**	4.00	8.40	6.20	88.6
All	689	0.15	3.25	5.52	4.38	92.5
Weight-all	689	**0.10**	4.17	8.40	6.29	89.0

Table 7. Ablation study of the ground separation on Maicity [36], Newer College [29].