3D LiDAR Mapping in Dynamic Environments Using a 4D Implicit Neural Representation

Supplementary Material

In this supplementary material, we present ablation studies that explore the impact of combining different types of losses and changing the number of basis functions. We also provide additional details on the implementation and experiments we conducted. Moreover, we showcase more qualitative results for dynamic object segmentation. We also open-source the scripts and code needed to reproduce all the experiments of our paper, which can be found at https://github.com/PRBonn/4dNDF

A. Ablation Study

The experiments on static mapping quality focus more on the result of the surface reconstruction. In order to study the reconstruction performance in the free space, we chose the KITTI seq. 05 from KTH-Dynamic-benchmark for the ablation study.

Influence of loss terms. Tab. 1 presents the impact of different combinations of losses influences the segmentation result. Note that we refer to the settings by the letters (A)–(F) in the first column of Tab. 1 in the following discussion. From the table, we can observe that only enabling L_{free}, *i.e.* case (A), alone results in a low score for DA, indicating that a large portion of dynamic objects remain unsegmented. However, adding of L_{eikonal} simultaneously, *i.e.*, case (B), can enhance the result. In case (C), the majority of dynamic objects can be eliminated by applying L_{certain}. However, since L_{certain} is only applied in the densely observed area, split by a hyperparameter r_{dense}, some distant dynamic points are preserved. By adding L_{free}, this problem can be solved, leading to further improvement in the result. In Fig. 1, we show qualitatively the influence of the different parts of the loss.

Influence of number of basis functions. Tab. 2 shows the impact of using different numbers of basis functions (K) on dynamic object segmentation result. $K = 1$ implies the map degenerates to 3D, which means the output of D_{mlp} is a single value representing time-independent signed distance. This leads to poor performance of both AA and DA. Increasing the value of K has a direct impact on the map's capacity, resulting in more accurate dynamic point segmentation. The optimal performance is achieved when K is set to 32. Further increasing the value of K does not lead to diminishing returns in terms of performance or even a degradation of the results. Therefore, we select $K = 32$ as the number of basis functions for all of our experiments in the main paper.

Table 1. Ablation study of losses combination on KITTI seq. 05 sequence from KTH-Dynamic-benchmark

	L_{free}	L_{certain}	L_{eikonal}	SA	DA	AA
A	✓			99.71	45.05	67.02
B	✓		✓	99.58	57.68	75.79
C		✓		99.96	89.11	95.38
D		✓	✓	99.11	92.42	95.71
E	✓	✓		99.44	99.50	97.45
F	✓	✓	✓	99.54	98.36	**98.95**

Table 2. Ablation study of the number of basis functions (K) on KITTI seq. 05 sequence from KTH-Dynamic-benchmark

K	SA	DA	AA
1	91.82	52.59	69.49
4	97.73	95.30	96.51
8	99.55	95.71	97.61
16	97.60	98.06	97.83
24	99.14	97.68	98.41
32	99.54	98.36	**98.95**
40	96.78	98.57	97.67
48	99.67	98.02	98.84

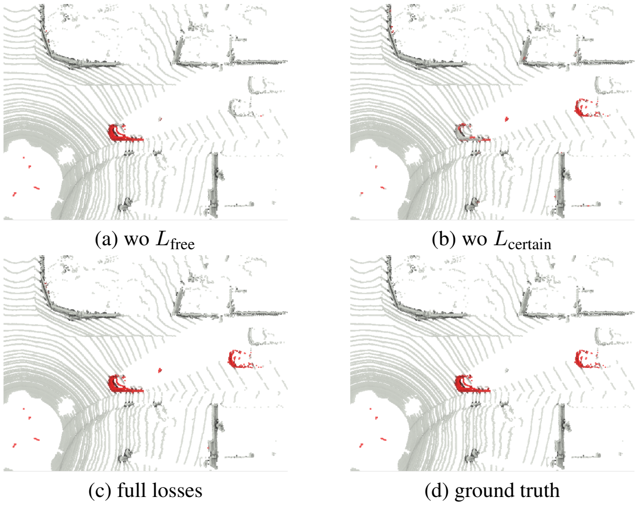


Figure 1. Segmentation results on one frame in the ablation study for losses. (a) shows the result when we turn off L_{free}, corresponding to case (D) in Tab. 1, we can see that the moving car far away from the sensor is not segmented. And turning off L_{certain} (case (B)) leads to poor dynamics segmentation in the dense observed area, which can be seen in (b).

B. Further Details on Experiments

B.1. Static Mapping Quality

Metrics. We represent the points sampled from ground-truth mesh or point clouds as P_{gt}, and represent the points sampled from estimated mesh as P_{es}, as detailed below for the different datasets. Then, we calculate the metrics as below:

$$\text{Comp} = \frac{1}{|P_{\text{gt}}|} \sum_{\mathbf{p}_{\text{gt}} \in P_{\text{gt}}} \min_{\mathbf{p}_{\text{es}} \in P_{\text{es}}} \left(\|\mathbf{p}_{\text{gt}} - \mathbf{p}_{\text{es}}\| \right), \tag{1}$$

$$\text{Acc.} = \frac{1}{|P_{\text{es}}|} \sum_{\mathbf{p}_{\text{es}} \in P_{\text{es}}} \min_{\mathbf{p}_{\text{gt}} \in P_{\text{gt}}} \left(\|\mathbf{p}_{\text{es}} - \mathbf{p}_{\text{gt}}\| \right), \tag{2}$$

$$\text{C-L1} = \frac{1}{2} \left(\text{Comp.} + \text{Acc.} \right), \tag{3}$$

$$\text{Precision} = \frac{\left|\left\{ \mathbf{p}_{\text{es}} \in P_{\text{es}} \mid \min_{\mathbf{p}_{\text{gt}} \in P_{\text{gt}}} \|\mathbf{p}_{\text{es}} - \mathbf{p}_{\text{gt}}\| < \xi \right\}\right|}{|P_{\text{es}}|}, \tag{4}$$

$$\text{Recall} = \frac{\left|\left\{ \mathbf{p}_{\text{gt}} \in P_{\text{gt}} \mid \min_{\mathbf{p}_{\text{es}} \in P_{\text{es}}} \|\mathbf{p}_{\text{gt}} - \mathbf{p}_{\text{es}}\| < \xi \right\}\right|}{|P_{\text{gt}}|}, \tag{5}$$

We report completeness (Comp.), accuracy (Acc.), Chamfer-Distance (C-L1) and F-score in the main text. For F-score, we use $\xi = 0.1\,\text{cm}$ in *ToyCar3* dataset and $\xi = 20\,\text{cm}$ in the real-world *Newer College* dataset.

Experiment settings. For *ToyCar3* dataset, we downsample the accumulated background point cloud with a resolution of $0.5\,\text{cm}$ and use the resulting point cloud as P_{gt}. We then uniformly sample the same number of points as P_{gt} on the mesh obtained by the methods and consider it as P_{es}. We used a resolution of $0.5\,\text{cm}$ for marching cubes to extract meshes in both our method and SHINE-mapping [4].

For *Newer College* dataset, we directly use the ground-truth point cloud collected by high-precision laser as P_{gt}. Because the coverage area of GT and the input data differ, we manually cropped the meshes to make their coverage regions as identical as possible. Then, similarly, we uniformly sample points with the same number of P_{gt} on the cropped mesh and use them as the P_{es} for evaluation. In this dataset, we use the resolution of $0.1\,\text{m}$ for marching cubes.

B.2. Dynamic Object Segmentation

Experiment settings. The KTH dataset contains four sequences in total, three of them (KITTI seq. 00, KITTI seq. 05 and Argoverse2) all use 64 beam LiDAR to collect data. We choose $r_{\text{dense}} = 15\,\text{m}$ for the experiments in these sequences. For the Semi-indoor sequence, the sensor is a 16-beam LiDAR, resulting in sparser scans. In this case, we set $r_{\text{dense}} = 8\,\text{m}$ to split the dense and sparse area. Fig. 2 and Fig. 3 demonstrate results of dynamic points removal in KITTI seq. 00 and KITTI seq. 05. Similar to the result depicted in the main text, Erasor [2] and Octomap [3] have a tendency to over-segment dynamic objects, which results in sparser static point clouds shown in zoomed view figures. Additionally, Removert [1] struggles with the complete removal of dynamic objects. Our approach achieves the best performance, with complete removal of dynamic objects while stably preserving static points. On the Semi-indoor sequence, there is an object that remains stationary for a long time, among the methods we tested, only Octomap successfully removed the object in the final map, thus achieving the highest score in this case.

C. Further Implementation Details

As mentioned in the main text, we encode geometric information using only two levels of feature hashing voxels, *i.e.*, $\mathcal{F}^l$, with different resolutions in all our experiments. The highest resolution of the voxel is determined by the scale of the scene. For the *ToyCar3* dataset, it is set to $2\,\text{cm}$, while for other LiDAR-based outdoor datasets, it is set to $30\,\text{cm}$. The resolution of the second voxel level is set to 1.5 times the highest resolution.

For the calculation of L_{eikonal}, we set the perturbation for the numerical gradient calculation ϵ as :

$$\epsilon = \frac{i(\epsilon_{\text{max}} - \epsilon_{\text{min}})}{I}, \tag{6}$$

where, i is the current epoch, I is the total number of epochs, which is 20. For outdoor LiDAR scenes, we set $\epsilon_{\text{max}} = 0.08\,\text{m}$, $\epsilon_{\text{min}} = 0.03\,\text{m}$. However, for *ToyCar3* dataset, as the scale is small, we set $\epsilon_{\text{max}} = 0.8\,\text{cm}$, $\epsilon_{\text{min}} = 0.3\,\text{cm}$.

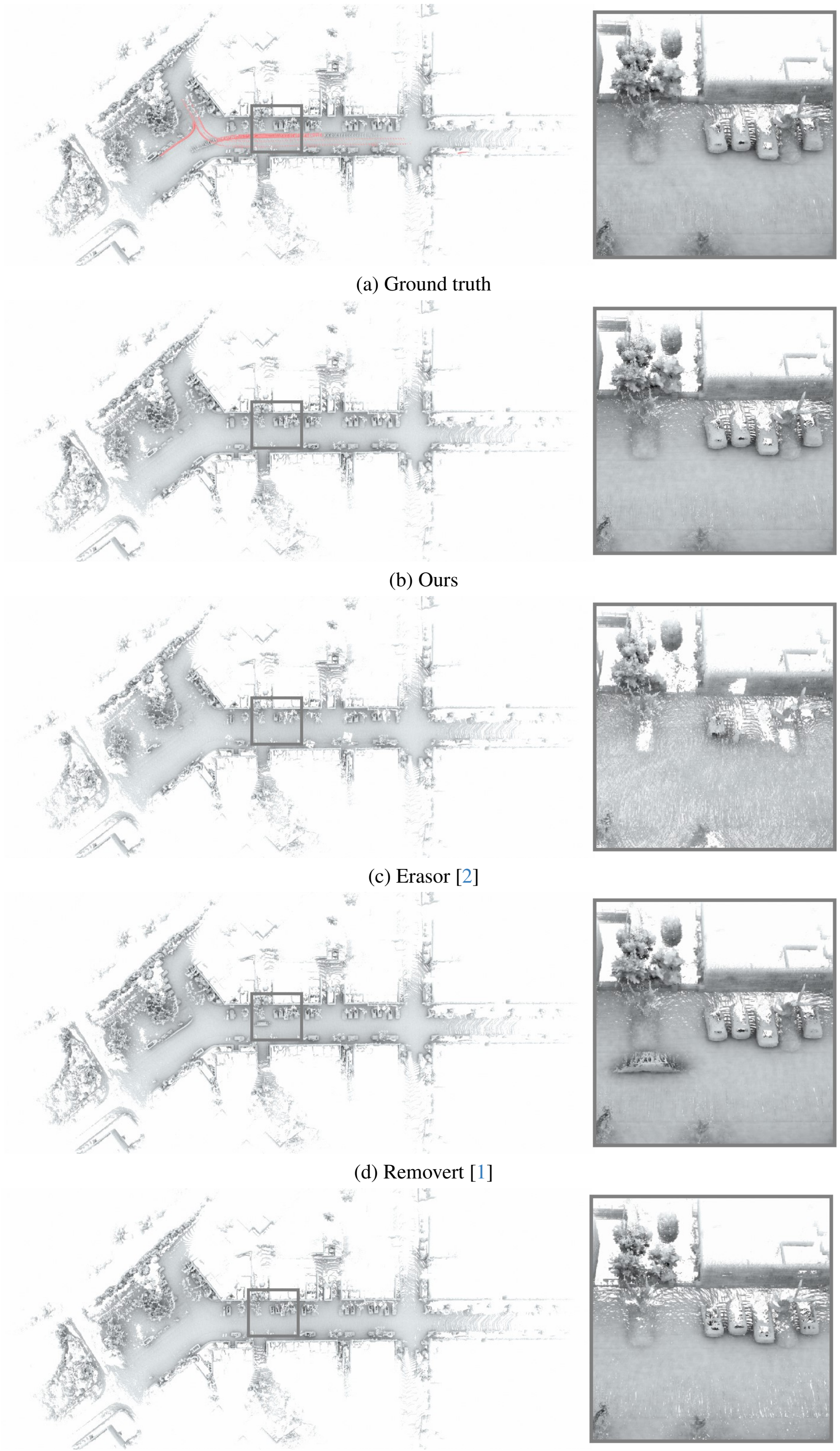
References

[1] Giseop Kim and Ayoung Kim. Remove, Then Revert: Static Point Cloud Map Construction Using Multiresolution Range Images. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2020. 2, 3, 4

[2] Hyungtae Lim, Sungwon Hwang, and Hyun Myung. ERASOR: Egocentric Ratio of Pseudo Occupancy-Based Dynamic Object Removal for Static 3D Point Cloud Map Building. *IEEE Robotics and Automation Letters (RA-L)*, 6(2):2272–2279, 2021. 2, 3, 4

[3] Qingwen Zhang, Daniel Duberg, Ruoyu Geng, Mingkai Jia, Lujia Wang, and Patric Jensfelt. A dynamic points removal benchmark in point cloud maps. In *IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, pages 608–614, 2023. 2, 3, 4

[4] Xingguang Zhong, Yue Pan, Jens Behley, and Cyrill Stachniss. SHINE-Mapping: Large-Scale 3D Mapping Using Sparse Hierarchical Implicit Neural Representations. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2023. 2



(a) Ground truth

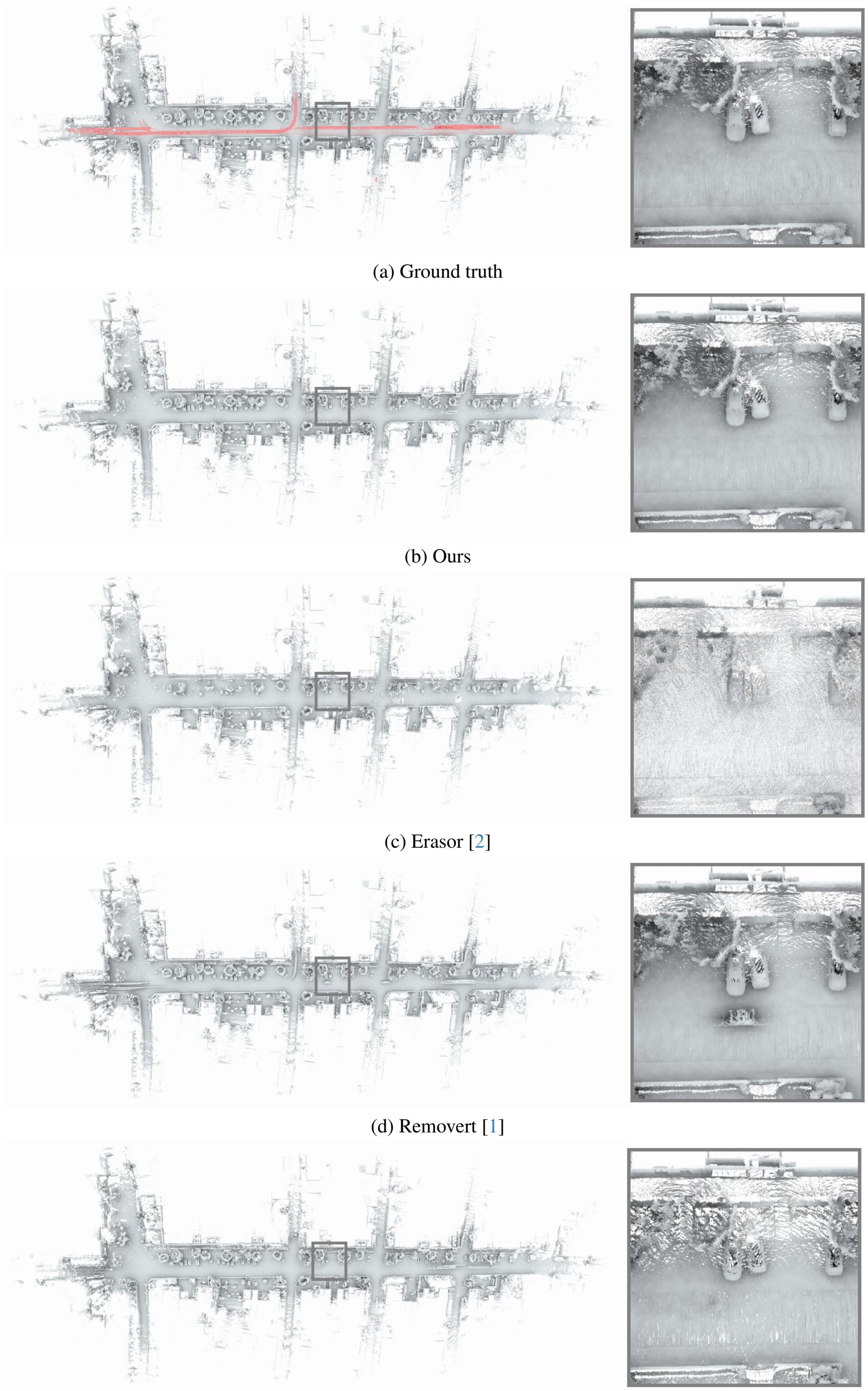
(b) Ours

(c) Erasor [2]

(d) Removert [1]

(e) Octomap* [3]

Figure 2. Comparison of dynamic object removal results produced by different methods on the KITTI seq. 00 of the KTH-benchmark. We display the bird's eye view of the complete point cloud with a zoomed view from the gray box. For the ground truth in (a), dynamic objects are marked in red in the bird's eye view, static points are depicted in the zoomed view for clearer comparison.



(a) Ground truth

(b) Ours

(c) Erasor [2]

(d) Removert [1]

(e) Octomap* [3]

Figure 3. Comparison of dynamic object removal results produced by different methods on the KITTI seq. 05 of the KTH-benchmark. We display the bird's eye view of the complete point cloud with a zoomed view from the gray box. For the ground truth in (a), dynamic objects are marked in red in the bird's eye view, static points are depicted in the zoomed view for clearer comparison.