LiSA: LiDAR Localization with Semantic Awareness

Supplementary Material

Sequence	Length	Tag	Training	Test
11-14-02-26	9.37km	sunny	✓	
14-12-05-52	9.22km	overcast	✓	
14-14-48-55	9.04km	overcast	✓	
18-15-20-12	9.04km	overcast	✓	
15-13-06-37	8.85km	overcast		✓
17-13-26-39	9.02km	sunny		✓
17-14-03-00	9.02km	sunny		✓
18-14-14-42	9.04km	overcast		✓

Table 1. **Dataset Descriptions on the QEOxford dataset.**

Sequence	Length	Tag	Training	Test
2012-01-22	6.1km	overcast	✓	
2012-02-02	6.2km	sunny	✓	
2012-02-18	6.2km	sunny	✓	
2012-05-11	6.0km	sunny	✓	
2012-02-12	5.8km	sunny		✓
2012-02-19	6.2km	overcast		✓
2012-03-31	6.0km	overcast		✓
2012-05-26	6.3km	sunny		✓

Table 2. **Dataset descriptions on the NCLT dataset.**

1. Dataset Details

As shown in Tab. 1 and Tab. 2, we report the training and testing trajectories on Oxford [2] (same as QEOxford [22]) and NCLT [6], along with their lengths and weather conditions.

2. Segmentation Models

As illustrated in Fig. 1, we showcase the segmentation results of the SPVNAS [39] and SphereFormer [20] models on the Oxford and NCLT datasets, respectively. It is evident that SphereFormer generally outperforms SPVNAS in terms of segmentation accuracy.

3. Additional Results

Runtime of LiSA and baselines. Inference time is a crucial metric in the localization task. Given the laser scanning rates of 20Hz and 10Hz for the Oxford and NCLT datasets respectively, a real-time algorithm needs to keep the inference time below 50ms and 100ms. LiSA does not incur additional time and computing after using additional semantic information, because of the superiority of its framework. Tab. 3 shows the inference times of LiSA and all baselines on QEOxford (equal to Oxford) and NCLT, respectively. The localization accuracy of LiSA is much better than other methods that satisfy real-time. Even compared to the time-consuming method SGLoc+PGO, LiSA can achieve similar or better accuracy.

Results on the scene with non-trivial angle change. In the main paper, only the orientation error of the state-of-the-art APR methods on the Oxford is slightly better than LiSA. We conjecture that the orientation change is small in this dataset, making the APR methods easy to remember. As shown in Fig.2, we selected an area in this dataset with a relatively complex orientation. In this case, the performance of the HypLiLoc [46] becomes worse than LiSA.

4. Network Architecture

The network architectures for Knowledge Distillation and Scene Coordinate Regression are illustrated in Fig. 3 and Fig. 4. Meanwhile, we provide changes in feature dimensions and our code and models will be made publicly available upon acceptance, ensuring the reproducibility of experimental results.

5. Benefit of diffusion-based distillation

Though any distillation method can be theoretically applied to LiSA, we apply diffusion-based distillation since it is state-of-the-art and empirically provides better performance than the regular L1 loss (Table 6 of the main paper). In Tab.4, we provide further comparisons to other distillation losses. Similar to the main paper, diffusion-based distillation performed better than other distillation losses, especially when the semantic segmentation model has limited accuracy (SPVNAS).

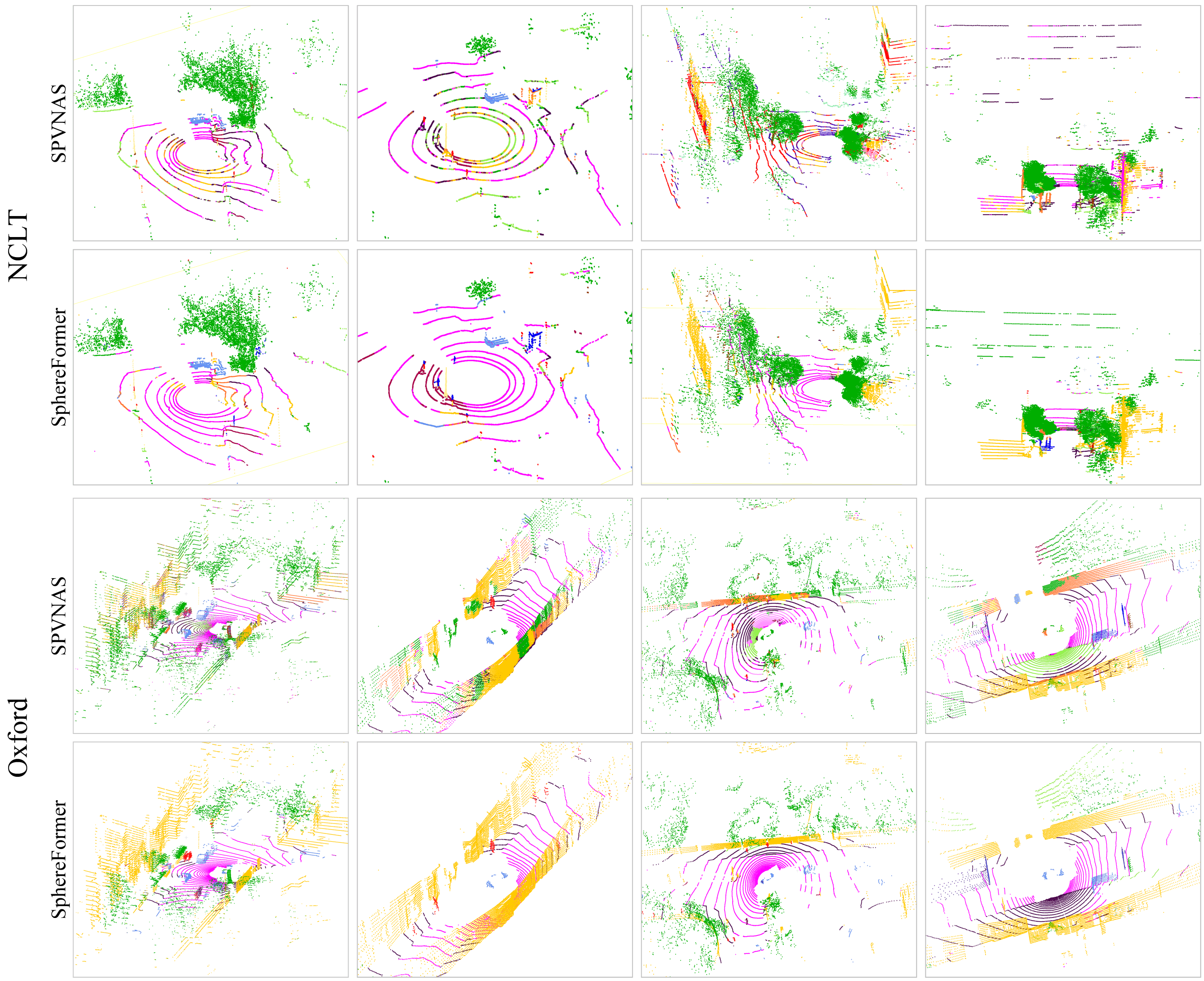

Figure 1. **Further qualitative results of segmentation on Oxford and NCLT.**

	QEOxford			NCLT		
Methods	localization results	running time	real-time	localization results	running time	real-time
PNVLAD	11.45m, 2.27°	6ms	✓	9.14m, 6.40°	6ms	✓
DCP	10.62m, 2.20°	3ms	✓	10.67m, 6.49°	3ms	✓
PointLoc	10.79m, 2.14°	625ms		7.57m, 4.60°	614ms	
PosePN++	5.13m, 1.69°	111ms		5.65m, 3.57°	108ms	
STCLoc	5.34m, 1.18°	97ms		5.15m, 4.18°	97ms	✓
HypLiLoc	3.89m, 1.27°	21ms	✓	1.95m, 3.16°	21ms	✓
SGLoc	1.53m, 1.60°	38ms	✓	1.83m, 3.54°	75ms	✓
SGLoc+PGO	1.31m, 1.12°	288ms		1.44m, 2.57°	325ms	
LiSA	0.95m, 1.14°	38ms	✓	1.51m, 2.34°	75ms	✓

Table 3. **Inference results on QEOxford and NCLT.** We report the localization accuracy and inference time for LiSA and all baselines on QEOxford and NCLT. LiSA far outperforms other methods to satisfy real-time performance and demonstrates sufficient competitiveness against SGLoc+PGO.

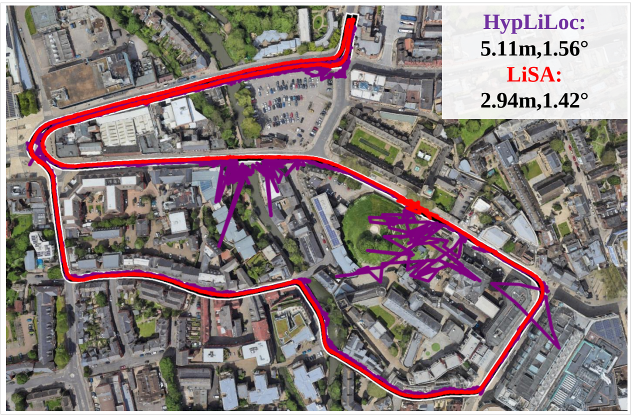

Figure 2. **Results on the scene with non-trivial angle change.**

quality of semantic features	loss function of KD	Mean Error (m/°)
low (SPVNAS)	L1	1.28m/1.53°
	L2	1.41m/1.76°
	Cos	1.27m/1.47°
	DDPM	**1.15m/1.40°**
high (SphereFormer)	L1	1.00m/1.15°
	L2	1.10m/1.26°
	Cos	0.98m/**1.13°**
	DDPM	**0.95m**/1.14°

Table 4. **Impact of the segmentation quality and the KD loss on QEOxford.**

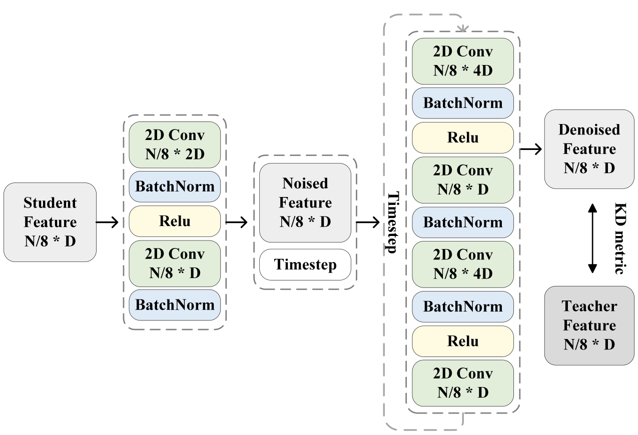

Figure 3. **The pipeline of knowledge distillation module.**

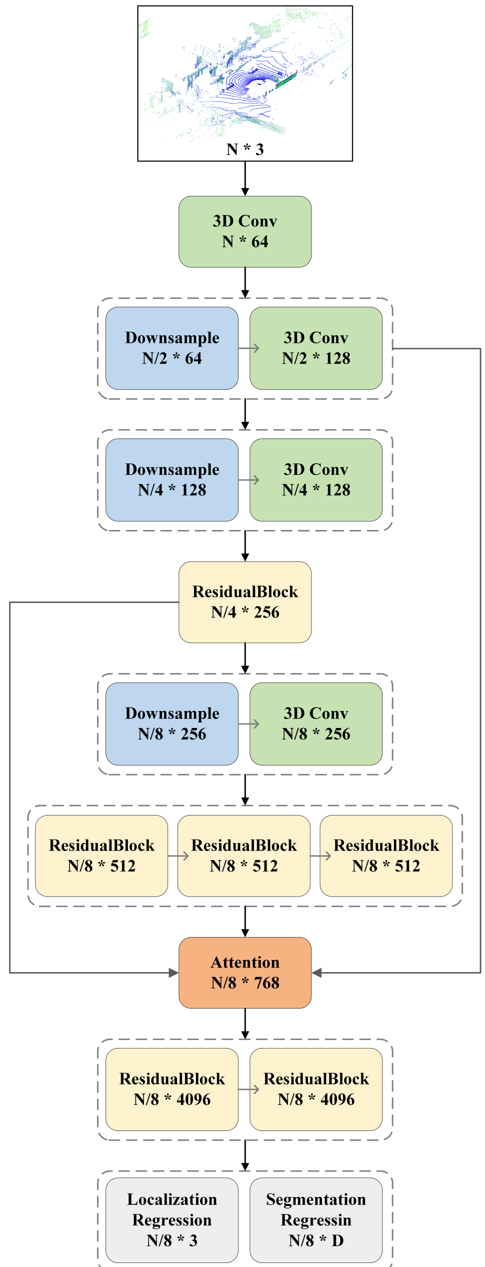

Figure 4. **The pipeline of scene coordinate regression module.**