Density-guided Translator Boosts Synthetic-to-Real Unsupervised Domain Adaptive Segmentation of 3D Point Clouds

Supplementary Material

In the following sections, we first compare two versions of the noise of DGT in Sec. 1. Then, we present the hyper-parameter sensitivity analysis in Sec. 2. After that, we present the t-SNE feature visualization of different methods in Sec. 3. Finally, we show more quantitative results for a better comparison against previous methods in Sec. 4.

1. Noise injection in DGT

As depicted in Fig. 1(e) and Fig. 1(f), points of the real-world scan show noticeable shifts in X and Y directions, and the points in the synthetic scan are integral and clean (Fig. 1(b) and Fig. 1(c)). However, the points in the red box show no significant shifts along the Z-axis in either synthetic or real-world scans. Thus, as stated in the main body of this paper, we do not inject noise on the Z-axis and only add random noise to the X and Y axes of the synthetic scan to enhance its realism. Moreover, we use PCAN and conduct experiments with two versions of DGT, *i.e.*, inject noise on the X and Y axes (XY-noise) and inject noise on the X, Y, and Z axes (XYZ-noise). As shown in Tab. 1, in comparison with XY-noise, XYZ-noise drops mIoU by 1.4% on SynLiDAR → SemanticKITTI (Syn → Sk).

2. Parameter sensitivity

***t*.** t is a hyper-parameter to control the update frequency of the teacher model. The larger t is, the more stable the teacher model is. In this study, we use LaserMix, fix α=0.99, and experiment with different t on Syn → Sk. As shown in Tab. 2, we get the best performance when t=100. A proper choice of t is between 100 and 200.

α. α is a hyper-parameter to control the update speed of the teacher model. A smaller α would render the training unstable, and a larger α would stabilize the model training but impede the student model from acquiring new target knowledge effectively. Here, we use LaserMix, fix t=100, and experiment with different α on Syn → Sk. As shown in Tab. 3, a proper choice of α is between 0.99 and 0.999.

Th_p. Th_p is the confidence threshold to select the pseudo labels. On the one hand, a smaller Th_p would yield many points with pseudo labels, but their accuracy cannot be guaranteed. On the other hand, a larger Th_p will filter out many incorrect pseudo-labeled points, but it is also possible to filter out correctly predicted points with smaller confidence. In this study, we experiment with different Th_p in our DGT-ST on Syn → Sk. We present the results in Tab. 4, among which we got the best performance when Th_p=0.4 and Th_p=0.5. However, for a fair comparison with CoSMix, we do not finetune this parameter and use Th_p=0.9 in the main body of this paper. Moreover, the final performance of DGT-ST is not sensitive to Th_p, and a proper choice of Th_p is between 0.4 and 0.7.

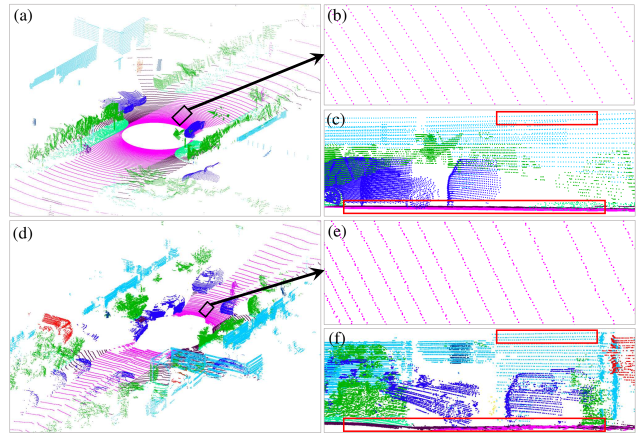

Figure 1. Comparison of synthetic and real-world scans. (a) and (d) show one scan of SynLiDAR and SemanticKITTI, respectively. (b) and (e) are zoomed-in visualizations of the road in the black box shown in (a) and (d). (c) and (f) are side-view visualizations of part of (a) and (d). The red boxes in (c) and (f) highlight that the points of synthetic and real-world scans do not exhibit significant shifts along the Z-axis.

PCAN	DGT with XY-noise	DGT with XYZ-noise
mIoU	**37.0**	35.6

Table 1. Comparison results of injecting noise on X and Y axes (XY-noise) and injecting noise on X, Y, and Z axes (XYZ-noise) in DGT on Syn → Sk.

t	1	100	200	300	400
mIoU	32.7	**36.0**	35.9	35.7	35.3

Table 2. Effect of t in the mean-teacher framework on Syn → Sk.

α	0.9	0.99	0.999	0.9999
mIoU	30.1	**36.0**	34.6	33.1

Table 3. Effect of α in the mean-teacher framework on Syn → Sk.

Th_p	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
mIoU	43.7	43.7	43.7	43.7	**43.8**	**43.8**	43.7	43.7	43.4	43.1	42.3

Table 4. Effect of confidence threshold Th_p for pseudo label selection in DGT-ST on Syn $\rightarrow$ Sk.

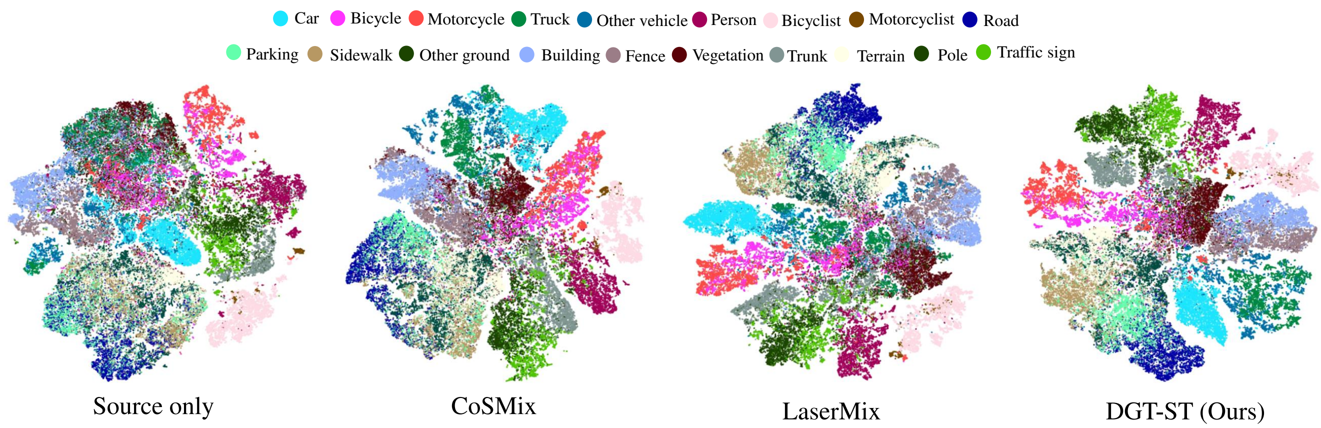

Figure 2. t-SNE visualization of the embedded features on Syn $\rightarrow$ Sk.

3. t-SNE visualization

In Fig. 2, we visualize the learned features of source only, CoSMix, LaserMix, and our DGT-ST by t-SNE [1]. We can observe that semantically similar categories are mixed together for all methods, *e.g.*, the features of road, sidewalk, and parking are mixed, and features of pole and traffic sign are mixed. In comparison, we can more easily separate different classes features of DGT-ST, *e.g.*, the trunk and other classes, the pole and traffic sign classes, and the building and fence classes. Therefore, we can conclude that DGT-ST extracts more discriminative features than the other works.

4. More qualitative results

In Fig. 3, we present more visualization results (error maps) on Syn $\rightarrow$ Sk, and compare our results with source only, PMAN, CoSMix, and the ground truth. Obviously, the incorrect predictions of DGT-ST are significantly fewer than other methods.

References

[1] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *J. Mach. Learn. Res.*, 9(11), 2008. 2

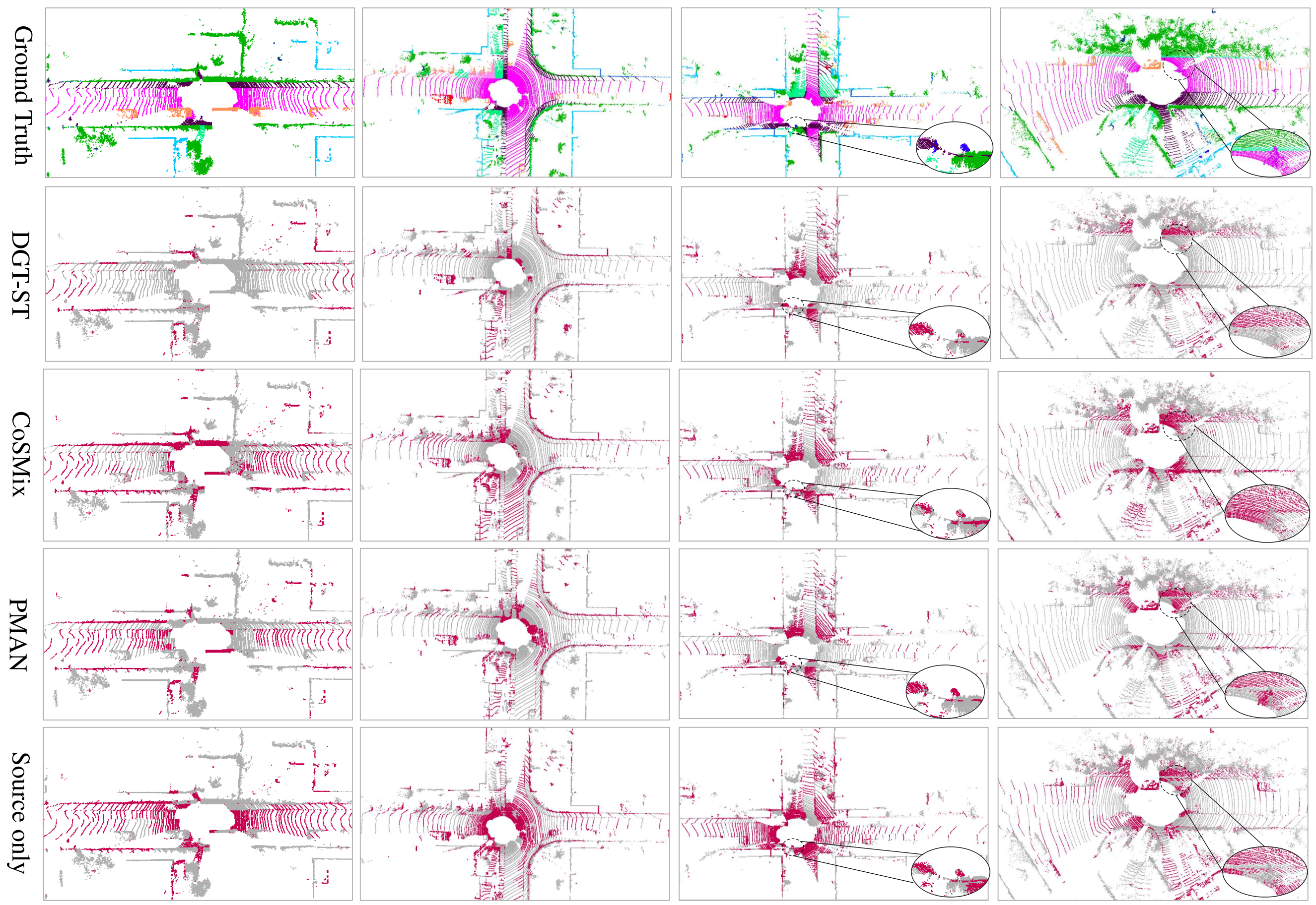

Figure 3. Additional qualitative results (error maps) on Syn $\rightarrow$ Sk. To highlight the differences, the correct and incorrect predictions are painted in gray and red, respectively.