Supplementary for Geometry-Aware Self-Training for Unsupervised Domain Adaptation on Object Point Clouds

Longkun Zou[1,2], Hui Tang[1], Ke Chen[1,3*], and Kui Jia[1,3,4*]

[1]South China University of Technology, [2]DexForce Technology Co., Ltd
[3]Peng Cheng Laboratory, [4]Pazhou Laboratory

{eelongkunzou, eehuitang}@mail.scut.edu.cn, {chenk, kuijia}@scut.edu.cn

A. Illustration of Our Self-Paced Learning

In Fig. A, we give an illustration for the proposed self-paced self-training (SPST) in an easy-to-hard learning manner, which selects target samples with pseudo labels from the most confident category predictions (cf. Sec. 3.1 in the main text).

B. More Experiment Results

B.1. More Results of Class-Wise Accuracy

In Table A, we report the class-wise classification accuracy on the task S→M, which shares the same settings as the synthetic-to-real task M→S* (see Table 2 in the main text). In general, we have similar observations: **(1)** our proposed GAST achieves a remarkable performance gain over all compared methods including the state-of-the-art DefRec + PCM, in terms of the average accuracy; **(2)** in almost all classes, our GAST performs better than or is comparable to DefRec + PCM; **(3)** all GAST variants except for the second one (*i.e*. RotCls only) outperform the very baseline w/o Adapt, but RotCls can bring benefits complementary to LocCls, *i.e*. LocCls+RotCls improves over LocCls; **(4)** combining all components (indicated by GAST) further enhances the performance. All these observations confirm the complementarity of the proposed three components and the superiority of our GAST.

B.2. Effect of Source Geometries

We examine the effect of geometric information from the source domain on classifying target instances. In Table B, we compare the results of GAST and GAST w/o Source Geometry Encoding (SGE) that removes both self-supervision of rotation and location on source data, on all adaptation tasks. It is observed that the performance on each task degrades significantly when GAST loses the awareness of source geometries during network training. This verifies the effectiveness of capturing domain-invariant geometric patterns.

B.3. Study of Rotation Invariance

We study an alternative strategy to capture domain-invariant geometric patterns, *i.e*. GAST w. Rotation Invariance Encoding (RIE) that categorizes an object point cloud with any rotation angle into one self-supervised class to achieve rotation invariance. Comparisons between GAST and GAST w. RIE are shown in Table A. We highlight the main observations below. **(1)** On average, GAST with rotation angle prediction outperforms GAST w. RIE with rotation angle confusion by a large margin. This verifies the significant effect and rationality of our used strategy. **(2)** On some classes like Bed and Bookshelf, GAST achieves a better performance than GAST w. RIE; on other classes like Cabinet, Lamp, and Monitor, GAST w. RIE surpasses GAST. This suggests that the two strategies can be complementary. Intuitively, GAST conducts a group-wise alignment and GAST w. RIE does a domain-wise one. In future, we will attempt to combine the two strategies.

B.4. Convergence and Generalization Analysis

We analyze convergence performance and generalization ability of our GAST by comparing it with the very baseline w/o Adapt in Fig. B, where we show the accuracy/loss curves over training epochs on held-out test sets of the target and source domains. Note that we empirically activate our proposed SPST component during the last 20 epochs, since SPST would introduce noises at the early stages of network training when the generated pseudo labels are extremely unreliable. We can observe that GAST after SPST works has a significant improvement in the test accuracy/loss on the target domain, converges more stably, and exceeds w/o Adapt.

*Corresponding authors

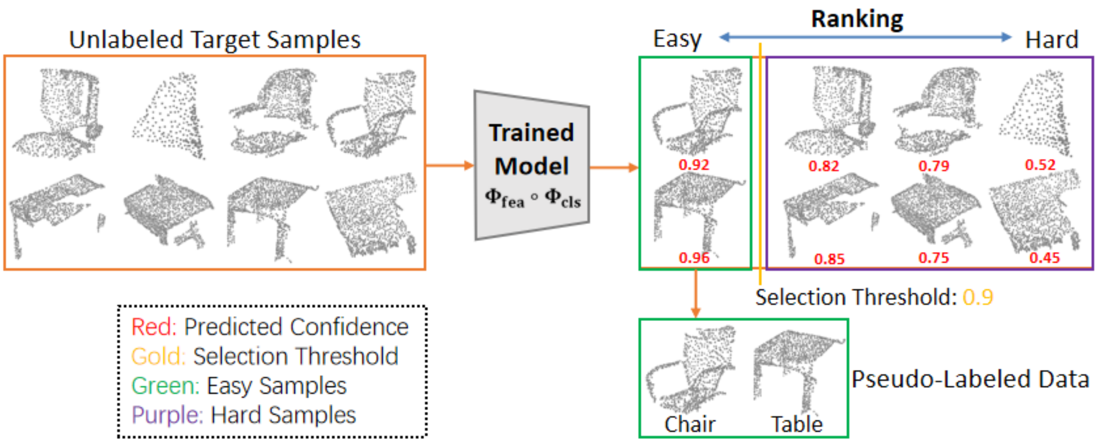


Fig. A. Illustration of our self-paced self-training (SPST), which selects target samples with pseudo labels at the higher levels of confidence.

	LocCls	RotCls	SPST	Bathtub	Bed	Bookshelf	Cabinet	Chair	Lamp	Monitor	Plant	Sofa	Table	Avg.
Supervised				96.0	95.0	97.0	91.9	100.0	100.0	100.0	93.0	98.0	100.0	97.1
w/o Adapt				96.0	34.0	49.0	22.1	99.0	**95.0**	**100.0**	67.0	**98.0**	**100.0**	76.0
DANN [2]				84.0	2.0	1.0	57.0	97.0	**95.0**	90.0	8.0	96.0	**100.0**	63.0
PointDAN [3]				92.0	33.0	5.0	22.1	96.0	85.0	88.0	42.0	96.0	96.0	65.5
DefRec + PCM [1]				**100.0**	**62.0**	41.0	36.0	**100.0**	90.0	98.0	71.0	**98.0**	**100.0**	79.6
	✓			**100.0**	25.0	41.0	51.2	**100.0**	**95.0**	96.0	65.0	97.0	**100.0**	77.0
		✓		94.0	22.0	33.0	53.5	**100.0**	85.0	95.0	44.0	**98.0**	**100.0**	72.4
GAST			✓	96.0	41.0	62.0	52.3	**100.0**	**95.0**	99.0	76.0	**98.0**	**100.0**	81.9
	✓	✓		98.0	25.0	66.0	40.7	**100.0**	85.0	96.0	65.0	**98.0**	**100.0**	77.4
	✓	✓	✓	98.0	**62.0**	67.0	44.2	**100.0**	90.0	92.0	**78.0**	**98.0**	**100.0**	**82.9**
GAST w. RIE	✓	✓	✓	**100.0**	2.0	48.0	**59.3**	**100.0**	**95.0**	98.0	**78.0**	**98.0**	**100.0**	77.8
GAST w. PCM	✓	✓	✓	98.0	41.0	**77.0**	46.5	**100.0**	**95.0**	94.0	77.0	**98.0**	**100.0**	82.7

Table A: Evaluation of class-wise classification accuracy (%) on the ShapeNet-10 to the ModelNet-10 (S→M). LocCls: distortion location prediction; RotCls: rotation angle prediction; SPST: self-paced self-training.

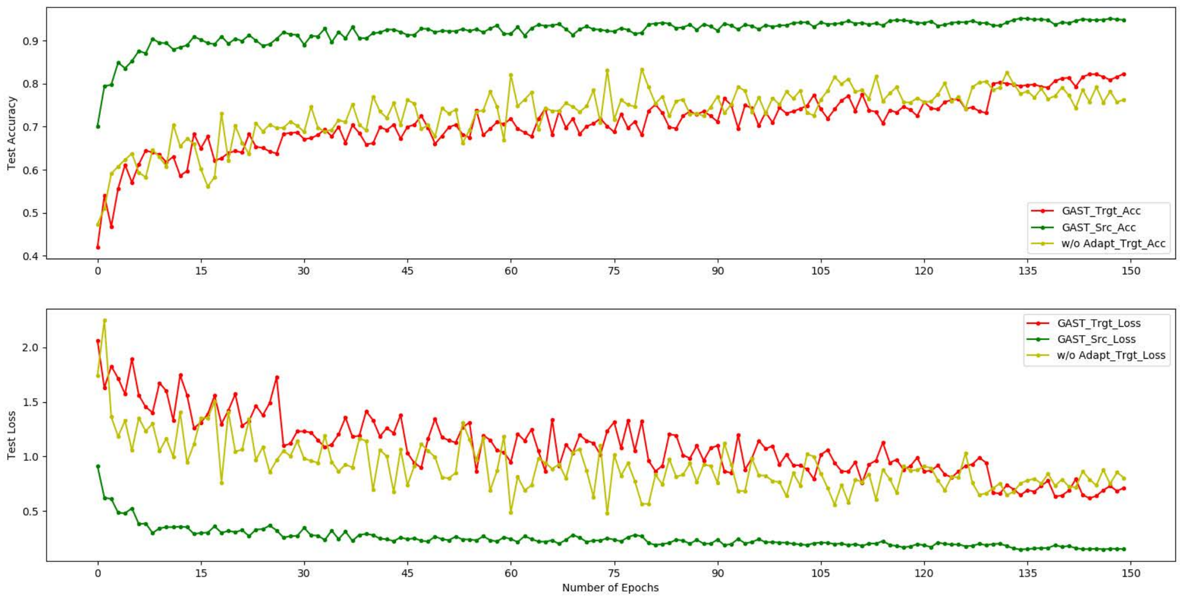
	M→S	M→S*	S→M	S→S*	S*→M	S*→S	Avg.
Supervised	93.9 ± 0.2	78.4 ± 0.6	96.2 ± 0.1	78.4 ± 0.6	96.2 ± 0.1	93.9 ± 0.2	89.5 ± 0.3
w/o Adapt	83.3 ± 0.7	43.8 ± 2.3	75.5 ± 1.8	42.5 ± 1.4	63.8 ± 3.9	64.2 ± 0.8	62.2 ± 1.8
DANN [2]	74.8 ± 2.8	42.1 ± 0.6	57.5 ± 0.4	50.9 ± 1.0	43.7 ± 2.9	71.6 ± 1.0	56.8 ± 1.5
PointDAN [3]	83.9 ± 0.3	44.8 ± 1.4	63.3 ± 1.1	45.7 ± 0.7	43.6 ± 2.0	56.4 ± 1.5	56.3 ± 1.2
RS [4]	79.9 ± 0.8	46.7 ± 4.8	75.2 ± 2.0	51.4 ± 3.9	71.8 ± 2.3	71.2 ± 2.8	66.0 ± 1.6
DefRec + PCM [1]	81.7 ± 0.6	51.8 ± 0.3	78.6 ± 0.7	54.5 ± 0.3	73.7 ± 1.6	71.1 ± 1.4	68.6 ± 0.8
GAST($R = 6$,$L = 8$)	85.0 ± 0.2	54.7 ± 0.3	69.2 ± 0.4	45.7 ± 0.2	69.7 ± 0.8	70.7 ± 0.1	65.8 ± 0.3
GAST w/o SGE	80.6 ± 0.2	53.9 ± 0.2	75.6 ± 0.4	47.6 ± 0.1	68.6 ± 0.3	71.0 ± 0.1	66.2 ± 0.2
GAST w. PCM	$\mathbf{85.1} \pm 0.2$	55.5 ± 0.3	79.6 ± 0.7	54.6 ± 0.2	78.0 ± 0.4	$\mathbf{75.4} \pm 0.4$	71.4 ± 0.4
GAST	84.8 ± 0.1	$\mathbf{59.8} \pm 0.2$	$\mathbf{80.8} \pm 0.6$	$\mathbf{56.7} \pm 0.2$	$\mathbf{81.1} \pm 0.8$	74.9 ± 0.5	$\mathbf{73.0} \pm 0.4$

Table B: Comparative evaluation in classification accuracy (%) averaged over 3 seeds ($\pm$ SEM) on the PointDA-10 dataset.

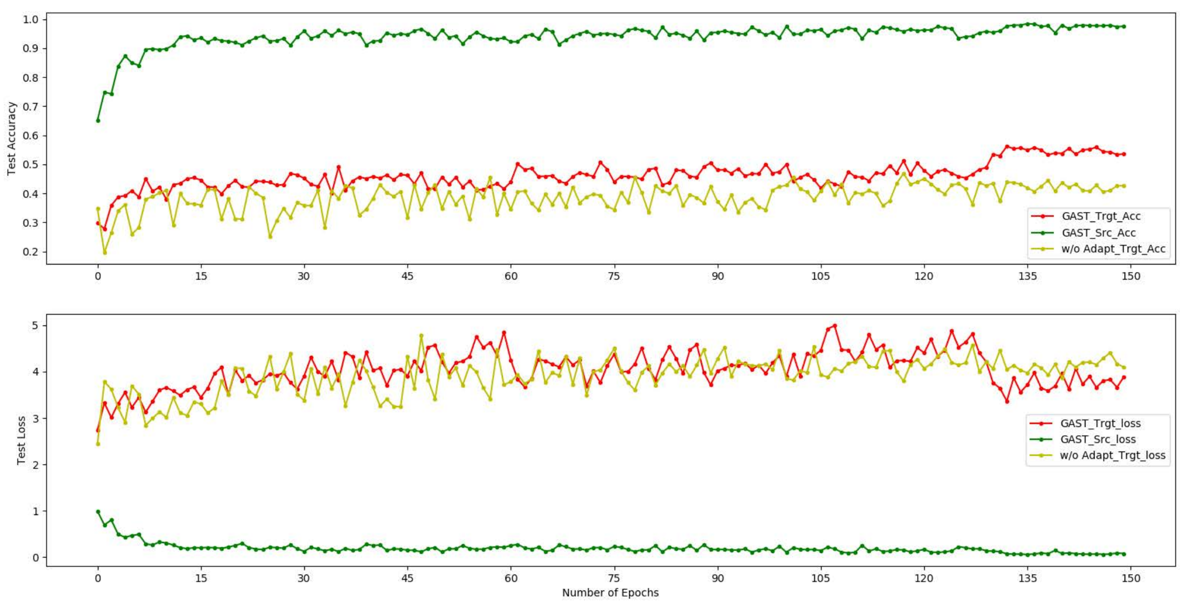
This verifies the effectiveness of our proposed SPST. It is worth noting that in the realistically significant synthetic-to-real task M→S*, GAST is always superior over w/o Adapt, demonstrating the usefulness and good generalization ability of our proposed self-supervised geometric information encoding.

B.5. Parameter Sensitivity

We examine the sensitivity of our proposed GAST to its hyper-parameters R (*i.e.* the number of rotation angles) and L (*i.e.* the number of split locations) by doing GAST experiments in another setting of $R = 6$ and $L = 2^3 = 8$ on all adaptation tasks. The results are reported in Table B.



(a) S→M



(b) M→S*

Fig. B. Convergence performance of GAST and w/o Adapt. Note that "Trgt_Acc"/"Trgt_loss" and "Src_Acc" / "Src_loss" indicate the accuracy/loss curves on held-out test sets of the target and source domains, respectively.

We can see that on almost all tasks, GAST ($R = 6$,$L = 8$) encounters a significant performance degradation, demonstrating the reasonability of the hyper-parameter setting (*i.e.* $R = 8$ and $L = 3^3 = 27$) adopted in the main text, although more various settings are certainly necessary to be examined.

B.6. Exploration of Point Cloud Mixup

We explore in our algorithmic framework the Point Cloud Mixup (PCM), a data augmentation technique adopted in [1], which may alleviate the long-tailed problem (*i.e.* class imbalance). In Table A, we also report the class-wise classification accuracy of GAST w. PCM that replaces our original loss with the mixup one on the labeled source data, on the task S→M. As we can see, PCM indeed improves the results on long-tailed classes, *e.g*. Cabinet and Lamp, but degrades those on major classes, *e.g*. Bed and Plant. This yields a slightly worse average result than our proposed GAST does. Such an observation is reflected in Table B as well, by comparing the results of GAST and GAST w. PCM on all adaptation tasks. Besides a slight improvement on M→S and S*→S, adding PCM to GAST results in degraded performance on other tasks. This encourages us to explore other techniques specific to the long-tailed problem, which can not only perform well on long-tailed classes but also remain superior on major ones.

References

[1] I. Achituve, H. Maron, and G. Chechik. Self-supervised learning for domain adaptation on point clouds. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 123–133, 2021.

[2] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.*, 17:2096–2030, 2016.

[3] C. Qin, H. You, L. Wang, C.-C. J. Kuo, and Y. Fu. Pointdan: A multi-scale 3d domain adaption network for point cloud representation. In *Adv. Neural Inform. Process. Syst.*, pages 7192–7203, 2019.

[4] Jonathan Sauder and Bjarne Sievers. Self-supervised deep learning on point clouds by reconstructing space. In *Adv. Neural Inform. Process. Syst.*, pages 12962–12972, 2019.