Self-Supervised Global-Local Structure Modeling for Point Cloud Domain Adaptation with Reliable Voted Pseudo Labels

SUPPLEMENTARY MATERIAL

1. Dataset Details

Dataset	Domain	Class
PointDA [27]	ModelNet-10 [37] ShapeNet-10 [2] ScanNet-10 [3]	bathtub, bed, bookshelf, cabinet, chair, lamp, monitor, plant, sofa, table
Sim-to-Real [16]	ModelNet-11 [37] → ScanObjectNN-11 [34]	bed, cabinet, chair, desk, display, door, shelf, sink, sofa, table, toilet
	ShapeNet-9 [2] → ScanObjectNN-9 [34]	bag, bed, cabinet, chair, display, pillow, shelf, sofa, table

Table 1. Details of the PointDA and Sim-to-Real datasets.

In this paper, we evaluate our method on two 3D domain adaptation benchmarks PointDA [27], which consists of 10 shared classes from ModelNet40 [37], ShapeNet [2] and ScanNet [3], and Sim-to-Real [16], which consists of 11 shared classes from ModelNet40 and ScanObjectNN [34], and 9 shared classes from ShapeNet and ScanObjectNN, respectively. The Sim-to-Real dataset is created for point-cloud-based meta-learning evaluation. In this paper, we also employ it to evaluate point cloud domain adaption methods. The details of the two datasets are listed in Table 1.

2. Self-Training

Method	M10 → S10	M10 → S*10	S10 → M10	S10 → S*10	S*10 → M10	S*10 → S10
w/o Adaptation	83.3	43.8	75.5	42.5	63.8	64.2
Self-Paced Self-Training [43]	84.4	45.9	**80.5**	**48.7**	64.8	70.4
Reliable Voting (ours)	**84.7**	**53.8**	77.7	48.3	**71.4**	**73.8**

Table 2. Self-training accuracy on the PointDA dataset.

In this section, we compare our reliable voting self-training method with Self-Paced Self-Training [43]. Experiments are conducted on the PointDA dataset. As shown in Table 2, our method achieves four better accuracies than Self-Paced Self-Training, showing its effectiveness on self-training.

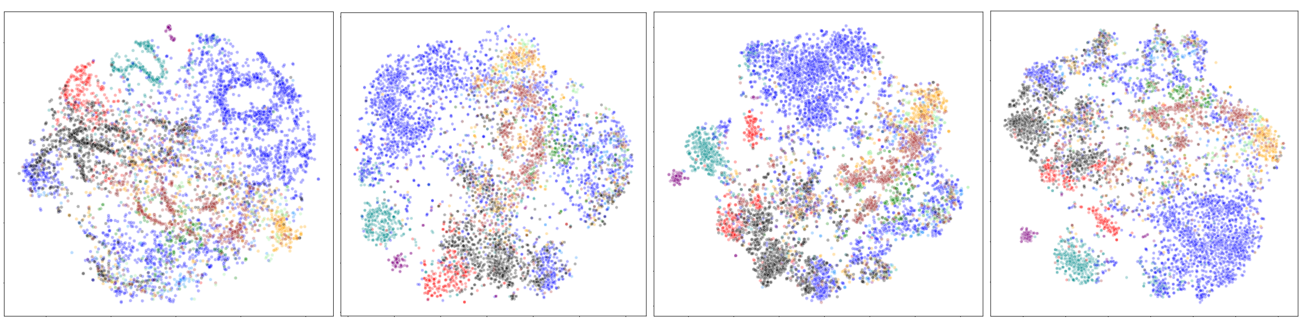
3. Visualization of the Target Training Data Distribution Change via t-SNE

After each training round, we perform reliable voting to assign accurate pseudo labels to the target point clouds. Because the amount of labeled target data increases, the adaptation is improved. We show the target training data distribution change using t-SNE in Fig. 1. As more and more reliable labeled target samples are voted and selected and used for supervised learning, the target features become more compact and discriminative.

In the figure, the blue color denotes the chair class, which is the largest class with 42.2% samples. Because the chair instances are diverse and some of them are similar to sofa and bed, it is challenging to accurately align all of them.

4. Examples of Reliable Voted Pseudo Label Generation

In this section, we visualize two reliable voted pseudo label assignment examples in Fig. 2. The reliable selection mechanism effectively avoids the noise of incorrect pseudo labels. Moreover, the voting method is able to adaptively add reliable target samples during training, which in return facilitates adaptation because the amount of labeled target data increases.



(a) round 1 (b) round 5 (c) round 10 (d) round 20

Figure 1. Visualization of the target training data distribution change during training via t-SNE. Each color denotes a class. Experiments are conducted on PointDA with the M10 $\rightarrow$ S*10 adaption scenario.

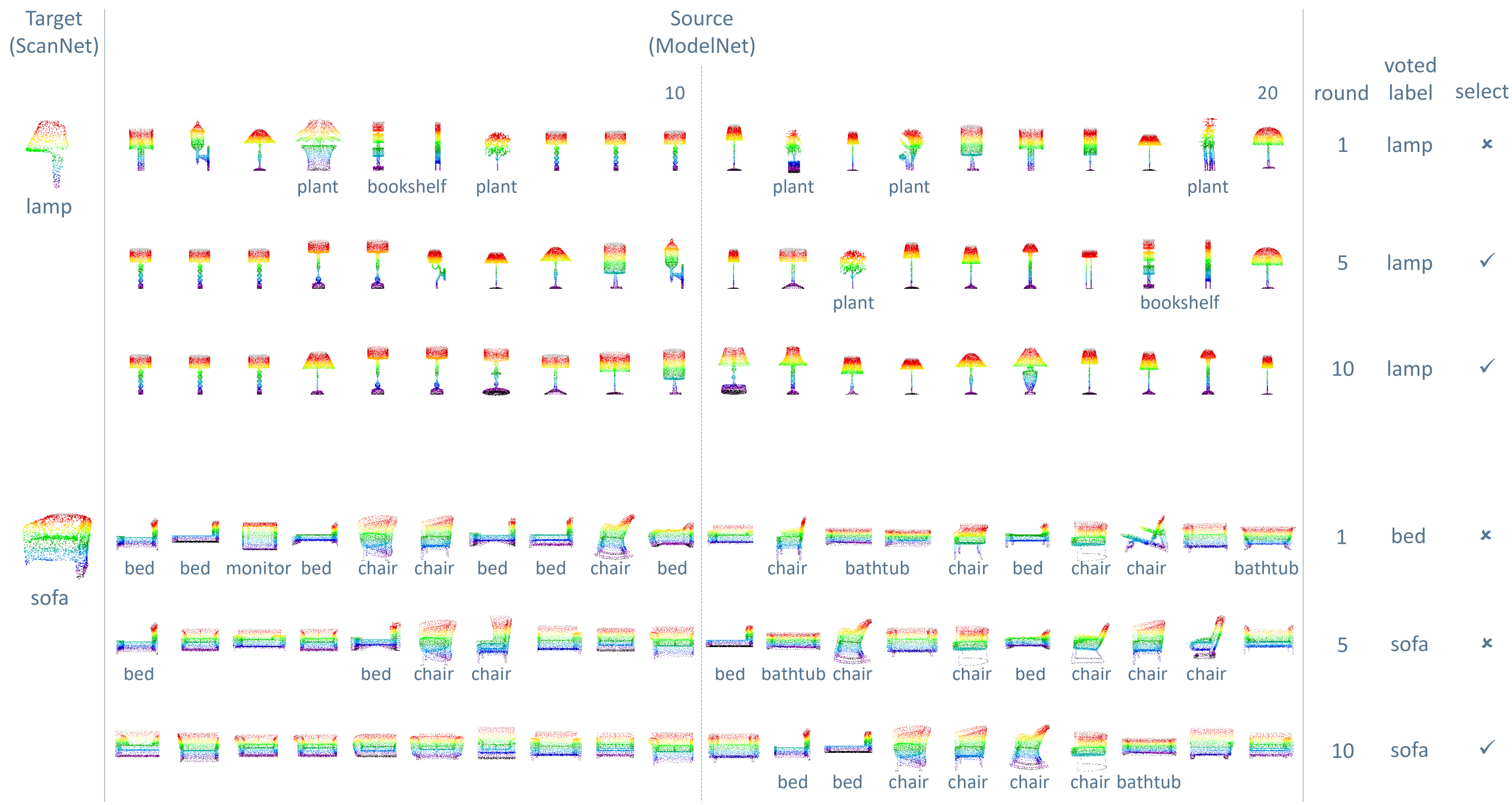


Figure 2. Visualization of two reliable voted pseudo label generation examples. Experiments are conducted on PointDA with the M10 $\rightarrow$ S*10 adaption scenario. The number (k) of source nearest neighbors for pseudo label voting is set to 10. The label consistency threshold (λ) is set to 1.0. The reliable selection mechanism effectively avoids the noise of incorrect pseudo labels. Moreover, as the network becomes stronger, the voting method is able to adaptively select an increasing number of target samples during training.