Supplementary Material for "LiDAR Distillation: Bridging the Beam-Induced Domain Gap for 3D Object Detection"

Yi Wei[1,2], Zibu Wei[1], Yongming Rao[1,2], Jiaxin Li[3], Jie Zhou[1,2], Jiwen Lu[1,2⋆]

[1]Department of Automation, Tsinghua University, China
[2]Beijing National Research Center for Information Science and Technology, China
[3]Gaussian Robotics, China
y-wei19@mails.tsinghua.edu.cn,
weizb18@mails.tsinghua.edu.cn,raoyongming95@gmail.com
lijx1992@gmail.com, jzhou@tsinghua.edu.cn,lujiwen@tsinghua.edu.cn

Target Domain Beams	density alignment	knowledge distillation	PointPillars					
			AP_{3D}			*Improvement*		
			Easy	Moderate	Hard	Easy	Moderate	Hard
32			80.79	65.91	61.09	-	-	-
	✓		82.80	67.01	63.82	+2.01	+1.10	+2.73
	✓	✓	86.00	70.15	66.86	+3.20	+3.14	+3.04
32*			74.59	57.77	51.45	-	-	-
	✓		78.74	63.02	58.94	+4.15	+5.25	+7.49
	✓	✓	82.83	66.96	62.51	+4.09	+3.94	+3.57
16			67.64	47.48	41.41	-	-	-
	✓		76.12	57.75	53.85	+8.48	+10.27	+12.44
	✓	✓	80.21	59.87	55.32	+4.09	+2.12	+1.47
16*			57.36	38.75	32.88	-	-	-
	✓		70.70	51.24	47.60	+13.34	+12.49	+14.72
	✓	✓	75.35	55.24	50.96	+4.65	+4.00	+3.36

Table 1. Component analysis on all target domains of KITTI dataset [2]. For **32*** and **16***, we not only reduce LiDAR beams but also subsample 1/2 points in each beam. The improvement is calculated in a progressive way.

Task	Method	PointPillars
		AP_{BEV} / AP_{3D}
KITTI → nuScenes	Direct Transfer	7.86 / 1.05
	SN[4]	14.96 / 5.28
	ST3D[6]	19.49 / 6.63
	Ours (naive downsample)	20.63 / 7.93
	Ours	**21.90 / 9.25**

Table 2. Results of KITTI → nuScenes adaptation.

⋆ Corresponding author

A More Dataset and Implementation Details

As a popular 3D object detection benchmark, KITTI [2] contains 3,712 training samples and 3,769 validation samples. Since KITTI dataset only provide the 3D bounding box labels for the objects within the field of view of the front RGB camera, we remove points outside of the front regions both for training and evaluation. According to the occlusion, truncation and 2D bounding box height, the objects are divided into three difficulty levels (Easy, Moderate and Hard).

The Waymo Open Dataset [3] is a large-scale dataset, which contains 1000 sequences in total, including 798 sequences (158,081 frames) in the training set and 202 sequences (39,987 frames) in the validation set. We used 1.0 version of Waymo Open Dataset. Same to ST3D [6], we also subsampled 1/2 training samples. Note that Waymo data is captured by a 64-beam LiDAR and 4 200-beam short-range LiDAR. The 200-beam LiDAR only captures data in a limited range and most of points come from 64-beam LiDAR. Thus, we only downsampled the points from 64-beam LiDAR.

The nuScenes dataset [1] consists of 28,130 training samples and 6,019 validation samples. The point clouds in nuScenes are 32-beam data while the equivalent beam to Waymo is 16^*. We only used nuScenes for evaluation.

The voxel size of SECOND and PV-RCNN are set to $(0.05m, 0.05m, 0.1m)$ on KITTI dataset and $(0.1m, 0.1m, 0.15m)$ on Waymo and nuScenes datasets. The models are trained on 8 RTX 2080 Tis.

B Component Analysis on Different Target Domains

To better understand the effects of point cloud density alignment and knowledge distillation, we conduct ablation studies on all synthetic target domains of KITTI [2] dataset. Table 1 shows the results. We observe that both point cloud density alignment and knowledge distillation contribute to the final results. On the one hand, when the domain gap is not large (*e.g.* $64 \rightarrow 32$), the knowledge distillation plays an more important role. On the other hand, when the domain gap is huge (*e.g.* $64 \rightarrow 16^*$), it is more crucial to align the point cloud density.

C Additional cross-dataset adaptation

As mentioned in ST3D [6], KITTI dataset lacks of ring view annotations and sufficient data. Due to these reasons, few methods select KITTI as source domain. To further demonstrate the effectiveness of our method, we add KITTI $\rightarrow$ nuScenes experiments with PointPillars backbone. During inference, we use the same field of view with that in KITTI dataset. However, we find that environmental domain gaps (such as object sizes) between these two datasets are huge and only using ST3D or our method cannot work well. Thus we combine ST3D and SN with our method and ST3D is also combined with SN. As shown in Table 2, our method can boost the performance of state-of-the-art methods with large margins. We also did ablation study on point cloud dowmsampling methods and

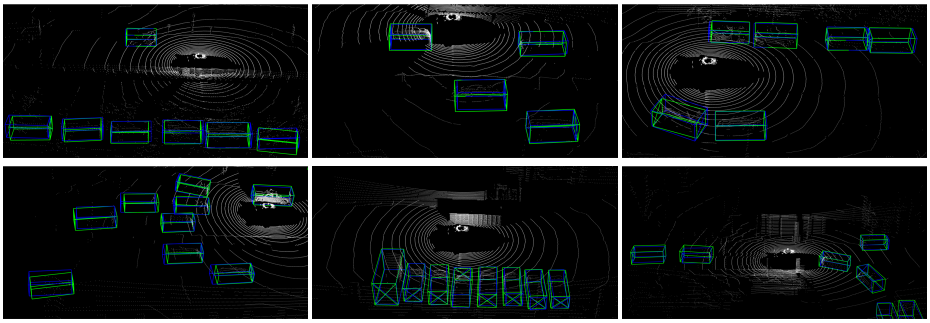

Fig. 1. Qualitative results of Waymo → nuScenes adaptation task. The green and blue bounding boxes represent detector predictions and groundtruths respectively.

we find that the proposed pseudo low-beam data generation method is better than naive downsampling method.

D The value to industry application

We finetune the PointPillars models on nuScenes datasets with different amount of groundtruth, which are pretrained with Waymo → nuScenes adaptation. In Table 3, the model pretrained with our method outperforms than other methods. With less groundtruth, the performance gains become larger. Surprisingly, with only 5% data, we can get higher performance than the model trained from scratch with 100% data. This experiment shows that we can use our method to reduce the need of expensive 3D labels, which is valuable to the industry.

Pretrained Method	AP_{BEV} / AP_{3D}		
	5%	10%	100%
Scratch	23.77 / 8.07	30.60 / 13.78	45.31 / 25.84
Direct Transfer	40.60 / 21.50	43.34 / 23.77	48.74 / 27.06
ST3D	43.66 / 24.03	45.98 / 25.72	49.70 / 29.34
Ours	**47.16 / 26.57**	**49.10 / 28.73**	**51.95/ 31.34**

Table 3. Results of finetuning experiments on nuScenes dataset.

E Qualitative Results

To better illustrate the superiority of our method, we finally provide some visualizations. Figure 1 shows qualitative results of cross-dataset adaptation equipped with SECOND-IoU [5]. We can see that our method can predict high-quality 3D bounding boxes.

References

1. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: CVPR. pp. 11621–11631 (2020)
2. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: CVPR (2012)
3. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., et al.: Scalability in perception for autonomous driving: Waymo open dataset. In: CVPR. pp. 2446–2454 (2020)
4. Wang, Y., Chen, X., You, Y., Li, L.E., Hariharan, B., Campbell, M., Weinberger, K.Q., Chao, W.L.: Train in germany, test in the usa: Making 3d object detectors generalize. In: CVPR. pp. 11713–11723 (2020)
5. Yan, Y., Mao, Y., Li, B.: Second: Sparsely embedded convolutional detection. Sensors **18**(10), 3337 (2018)
6. Yang, J., Shi, S., Wang, Z., Li, H., Qi, X.: ST3D: Self-training for Unsupervised Domain Adaptation on 3D Object Detection. In: CVPR. pp. 10368–10378 (2021)