Supplementary Material for "Density-Insensitive Unsupervised Domain Adaption on 3D Object Detection"

Qianjiang Hu Daizong Liu Wei Hu✉
Wangxuan Institute of Computer Technology, Peking University
No. 128, Zhongguancun North Street, Beijing, China
hqjpku@pku.edu.cn, dzliu@stu.pku.edu.cn, forhuwei@pku.edu.cn

In this supplementary material, we provide more ablation studies and visualizations omitted in our main paper due to the page limit, including

- Section S1: Additional ablation studies.
- Section S2: Qualitative results.
- Section S3: Performance under weakly supervised setting.

As in the main paper, all ablation studies and visualization results in this supplementary file are conducted on the domain adaption case of nuScenes → KITTI, using SECOND-IoU as the 3D detection backbone.

S1. Additional Ablation Studies

Sensitivity Analysis of pseudo labels' confidence threshold. As shown in Table 1, we investigate the effect of different confidence threshold c_{th} in Eq. (3) of our main paper for pseudo label generation. We can find that our method achieves the best performance when c_{th} is around 0.6. If c_{th} is even larger, the performance decreases significantly. This is because a larger c_{th} gives rise to a smaller number of positive examples that degenerate the self-training process.

c_{th}	AP_{BEV}	$\text{AP}_{\text{AP3}D}$
0.1	80.6	61.9
0.2	80.5	64.5
0.3	80.2	62.8
0.4	80.6	63.6
0.5	**81.4**	66.6
0.6	81.0	**67.2**
0.7	71.3	59.8
0.8	15.6	11.7

Table 1. Performance under different confidence thresholds c_{th}

✉ Corresponding author: W. Hu.

Sensitivity Analysis of the two terms in Edge-Level Consistency (ELC). Further, we investigate the importance of the two terms in ELC (Eq. (9) in the main body): the edge weight alignment and the GLR alignment. As shown in Table 2, the performance drops by 2.8 without the edge weight alignment (*i.e.*, $\gamma = 0.0$), and drops by 4.8 without the GLR alignment (*i.e.*, $\gamma = 1.0$). This indicates the importance of striking a good balance between the edge weight alignment and the GLR alignment.

γ	AP_{BEV}	$\text{AP}_{\text{AP3}D}$
0.0	81.2	64.9
0.1	81.7	65.1
0.2	81.6	63.7
0.3	81.6	64.6
0.4	81.7	65.4
0.5	**81.4**	66.6
0.6	**81.4**	**67.6**
0.7	81.3	65.7
0.8	81.2	63.9
0.9	81.0	64.1
1.0	80.4	62.9

Table 2. Performance under different γ

S2. Qualitative Results

Main results. As shown in Figure 1, we provide some qualitative results of our proposed DTS and competitive baselines (SN [1] and ST3D [2]) on the KITTI validation set. We observe that SN and ST3D produce a few negative predictions, while our predictions are clean and more accurate. This is because the teacher-student framework with both Node-Level Consistency (NLC) and ELC provides a stable and adaptive pseudo supervision to the detector.

Ablation results. As shown in Figure 2, we also provide some qualitative results of four ablation variants of the proposed DTS: Basic TS (basic teacher-student architecture, *i.e.*, DTS without NLC and ELC), DTS without NLC, DTS without ELC, and the complete DTS. We observe that with NLC and ELC introduced, our DTS reduces the number of negative predictions. Also, the complete DTS produces more precise predictions, as clearly demonstrated in regions marked with yellow circles in Figure 2(c).

S3. Performance under weakly supervised setting

Although our method is proposed for UDA, applying additional information (with SN or a few target-domain labels) can further improve the performance, as shown in Table 3. We observed one needs to provide around 50% labels to reach parity with the oracle detector, thus validating the potential applicability.

Method	AP_{BEV}	AP_{3D}	Method	AP_{BEV}	AP_{3D}
Ours	81.4	66.6	w/ 20% label	82.4	69.5
w/ SN	81.4	67.0	w/ 50% label	84.5	72.4
w/ 10% label	81.8	67.6	Oracle	83.3	73.5

Table 3. Adaptation performance comparison of unsupervised DA and semi-supervised DA, N $\rightarrow$ K.

References

[1] Yan Wang, Xiangyu Chen, Yurong You, Li Erran Li, Bharath Hariharan, Mark Campbell, Kilian Q Weinberger, and Wei-Lun Chao. Train in germany, test in the usa: Making 3d object detectors generalize. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11713–11723, 2020. 1

[2] Jihan Yang, Shaoshuai Shi, Zhe Wang, Hongsheng Li, and Xiaojuan Qi. St3d: Self-training for unsupervised domain adaptation on 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10368–10378, 2021. 1

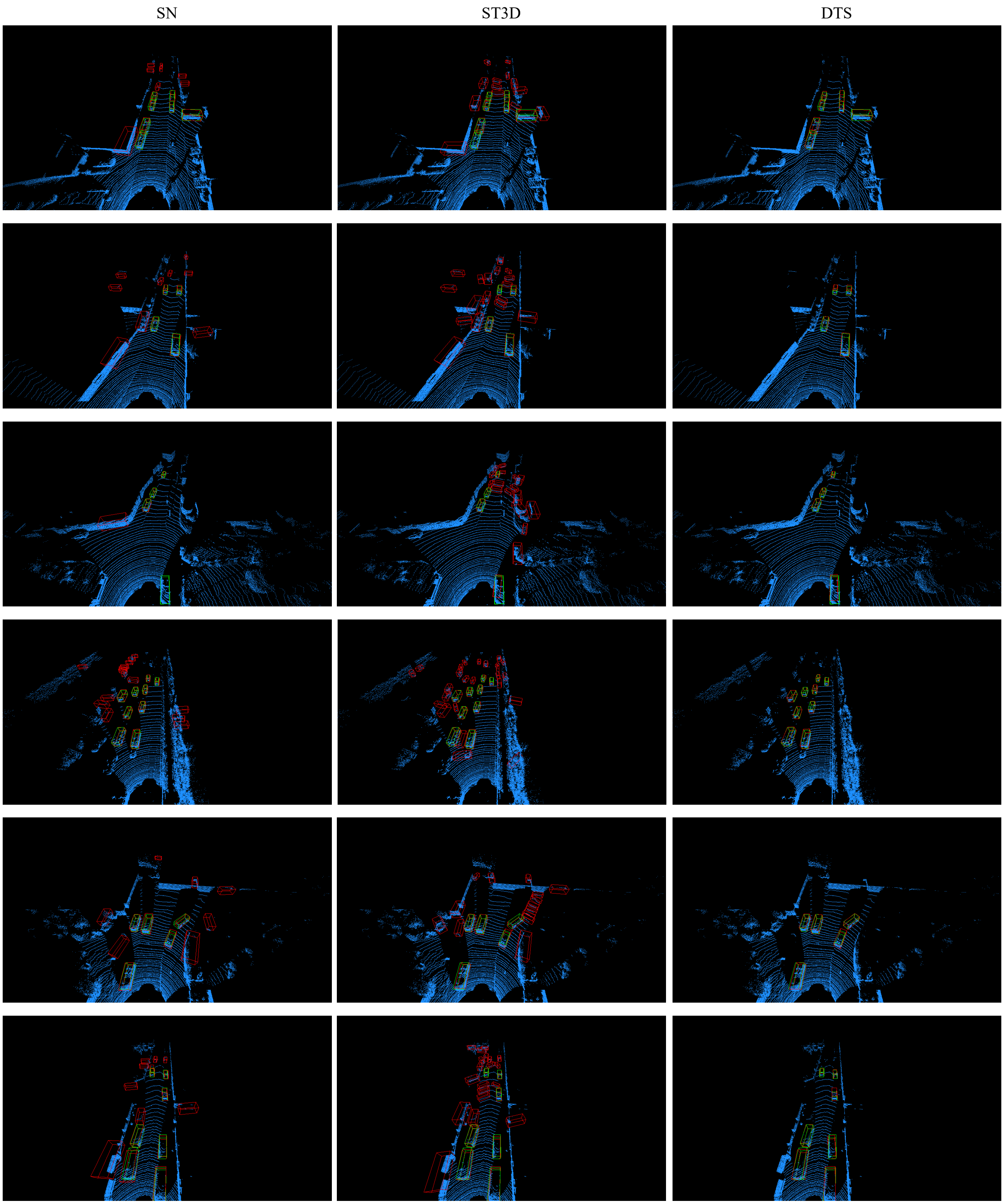

Figure 1. Qualitative results of our proposed DTS and competitive baselines on the KITTI validation set. The green boxes indicate the ground truth bounding boxes, while the red boxes indicate the predicted bounding boxes.

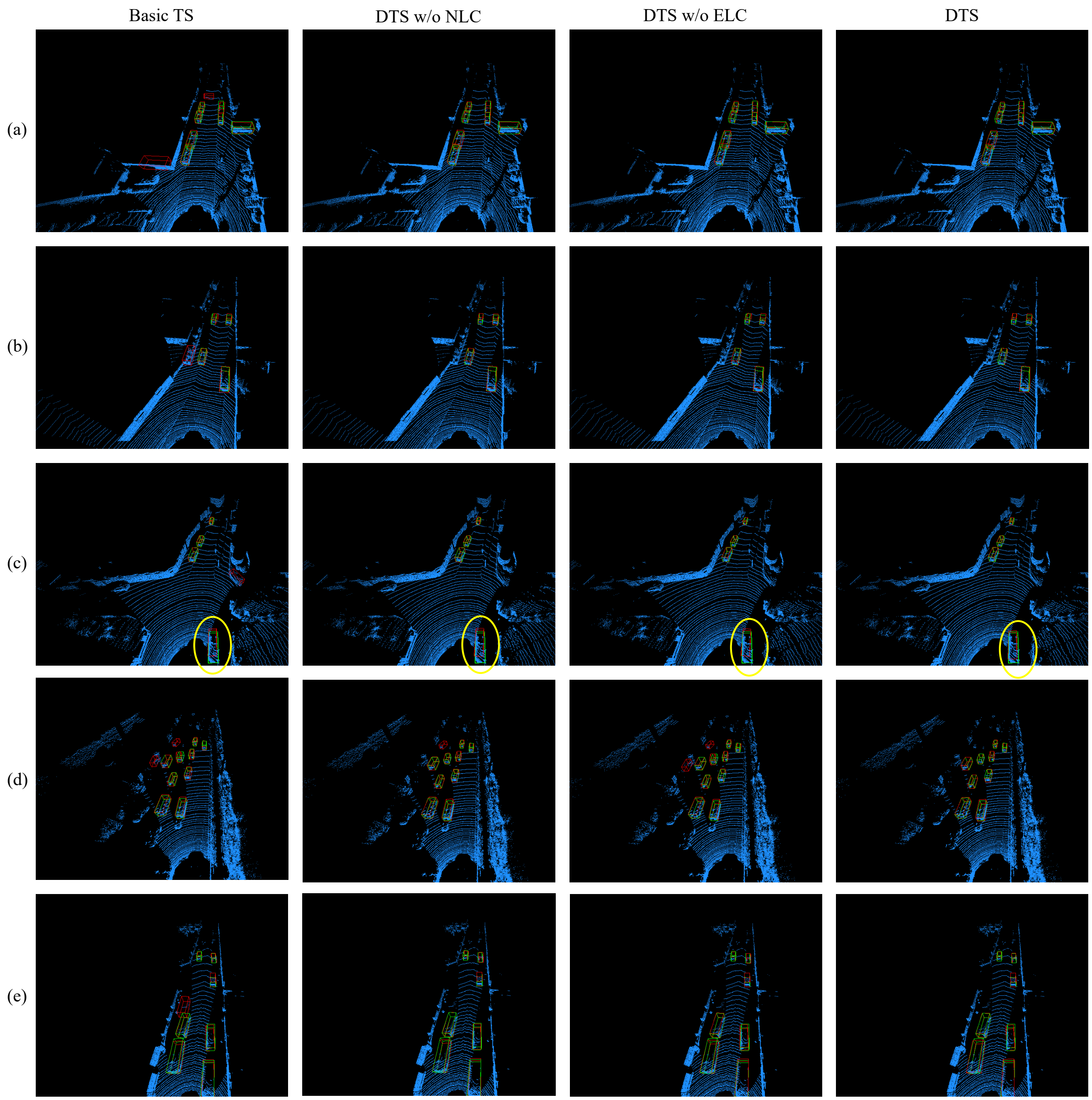

Figure 2. Qualitative results of our proposed DTS and ablation variants. The green boxes indicate the ground truth bounding boxes, while the red boxes indicate the predicted bounding boxes.