[Supplementary Document] Point-TTA: Test-Time Adaptation for Point Cloud Registration Using Multitask Meta-Auxiliary Learning

Ahmed Hatem
University of Manitoba
hatema@myumanitoba.ca

Yiming Qian
University of Manitoba
yiming.qian@umanitoba.ca

Yang Wang
Concordia University
yang.wang@concordia.ca

1. Learnable Balancing Weights

The final auxiliary loss is the weighted sum of our proposed three auxiliary losses. Instead of using the same fixed values of the auxiliary losses balancing weights, we propose to add the balancing weights to the learnable network parameters and learn their values during training. This enables the training algorithm to effectively balance the auxiliary tasks with optimal weights. At test-time, the learned weights are fixed.

We first initialize three trainable parameters $c_i (i = 1, 2, 3)$ with initial values of ones. These parameters are trained along with the auxiliary task parameters, i.e. $\{\theta^{aux}, c_i\}$, by optimizing the final auxiliary loss L_{aux}. To ensure that the learned balancing weights are in the appropriate range, we adapt the loss function in [5] to avoid large values of weights and biased losses. We further apply a softmax function to make balancing weights sum to 1. The learned parameters c_i are mapped to λ_i as following:

$$\lambda_i = Softmax(\frac{1}{2c_i^2}), \tag{1}$$

where i = (1,2,3). The final balanced auxiliary loss is defined as following:

$$L_{aux}(\{c_i\}_{i=1}^{3}, \theta_{aux}) = \lambda_1 \ell_{rec} + \lambda_2 \ell_{byol} + \lambda_3 \ell_{cc}. \tag{2}$$

2. Additional Results and Ablation Studies

2.1. Generalization to Unseen Datasets

We have conducted a cross-dataset generalization experiment on 3DMatch [7] and KITTI [3] datasets to evaluate the capability of our method in improving the generalization performance of point cloud registration networks. Due to the space limitation, we only report the registration results when training the model on 3DMatch [7] and evaluating on KITTI [3] in the main paper. Here we report the performance improvement of all the baselines when training on KITTI dataset [3] and testing on 3DMatch dataset [7]. As presented in Table 1, our method achieves the best performance across all the evaluation metrics with a good margin. This demonstrates the effectiveness of our method in boosting the generalization capability of the three backbones: DGR [2], DHVR [6], and PointDSC [1] to unseen datasets by enabling the networks to exploit the internal features of point clouds at test time.

Table 1: Results of cross-dataset generalization experiment. We train the models on KITTI dataset [3] and evaluate on 3DMatch dataset [7].

	Recall ↑	RE (deg) ↓	TE (cm) ↓
DGR [2]	87.39	2.71	7.58
Ours + DGR	**90.25**	**2.32**	**7.26**
DHVR [6]	87.16	2.74	7.43
Ours + DHVR	**90.48**	**2.25**	**7.04**
PointDSC [1]	89.42	2.15	6.89
Ours + PointDSC	**91.36**	**1.87**	**6.33**

Table 2: Ablation studies comparison between learnable and fixed balancing weights. Second row presents results with fixed balancing weights of ($\lambda_1 = 0.5$, $\lambda_2 = 0.3$, $\lambda_3 = 0.2$). Third row presents results with fixed balancing weights of ($\lambda_1 = 0.7$, $\lambda_2 = 0.1$, $\lambda_3 = 0.2$). Our framework with learnable balancing weights achieve better results in all evaluation metrics.

	Recall ↑	RE (deg) ↓	TE (cm) ↓
DGR [2]	91.31	2.43	7.34
DGR + fixed	92.03	1.76	6.48
DGR + fixed	92.26	1.73	6.41
DGR + learnable	**92.45**	**1.71**	**6.39**

2.2. Analysis of methodology components

We perform ablation study on each methodology component to further analyze our proposed method. We report the registration results when evaluating on 3DMatch dataset [7] in the main paper. Here we report results when testing on KITTI dataset [3] to investigate the contribution of each framework component. As shown in Table 3, the results of combining auxiliary learning with DGR [2]

Table 3: Ablation study on KITTI dataset. We report the contribution of each framework component.

	Recall ↑	RE (deg) ↓	TE (cm) ↓
DGR	95.24	0.44	23.25
DGR + Aux.	95.21	0.43	23.32
DGR + TTA (w/o meta)	96.63	0.39	22.30
DGR + Meta-Aux. (w/o TTA)	96.85	0.37	21.94
DGR + full framework	**97.36**	**0.34**	**21.16**

were slightly worse than baseline results. However, TTA and meta-learning significantly improve registration performance when combined with auxiliary learning. Finally, our final framework achieves the best registration results, which validate the contribution of each component.

2.3. Impact of Learnable Balancing Weights.

In this study, we report the impact of learning the balancing weights. We perform the experiments on the 3DMatch dataset and adapt DGR [2] as the baseline of the experiment. The results are shown in Table 2. In the second and third rows of Table 2, we report the results of using fixed balancing weights. Although the registration recall improved by 0.72% and 0.95%, respectively, it is hard to determine the optimal balancing weights without doing numerous experiments. Instead, training with learnable balancing weights effectively balances the auxiliary tasks and greatly improves all evaluation metrics.

3. More Qualitative Results

We present more qualitative results on 3DMatch (Figure 1 and Figure 2), 3DLoMatch (Figure 3), and KITTI (Figure 4 and Figure 5).

References

[1] Xuyang Bai, Zixin Luo, Lei Zhou, Hongkai Chen, Lei Li, Zeyu Hu, Hongbo Fu, and Chiew-Lan Tai. Pointdsc: Robust point cloud registration using deep spatial consistency. *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 1

[2] Christopher Bongsoo Choy, Wei Dong, and Vladlen Koltun. Deep global registration. *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 1, 2

[3] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research*, 2013. 1, 6, 7

[4] Shengyu Huang, Zan Gojcic, Mikhail (Misha) Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 5

[5] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[6] Junha Lee, Seungwook Kim, Minsu Cho, and Jaesik Park. Deep hough voting for robust global registration. *IEEE International Conference on Computer Vision*, 2021. 1

[7] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas A. Funkhouser. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 1, 3, 4

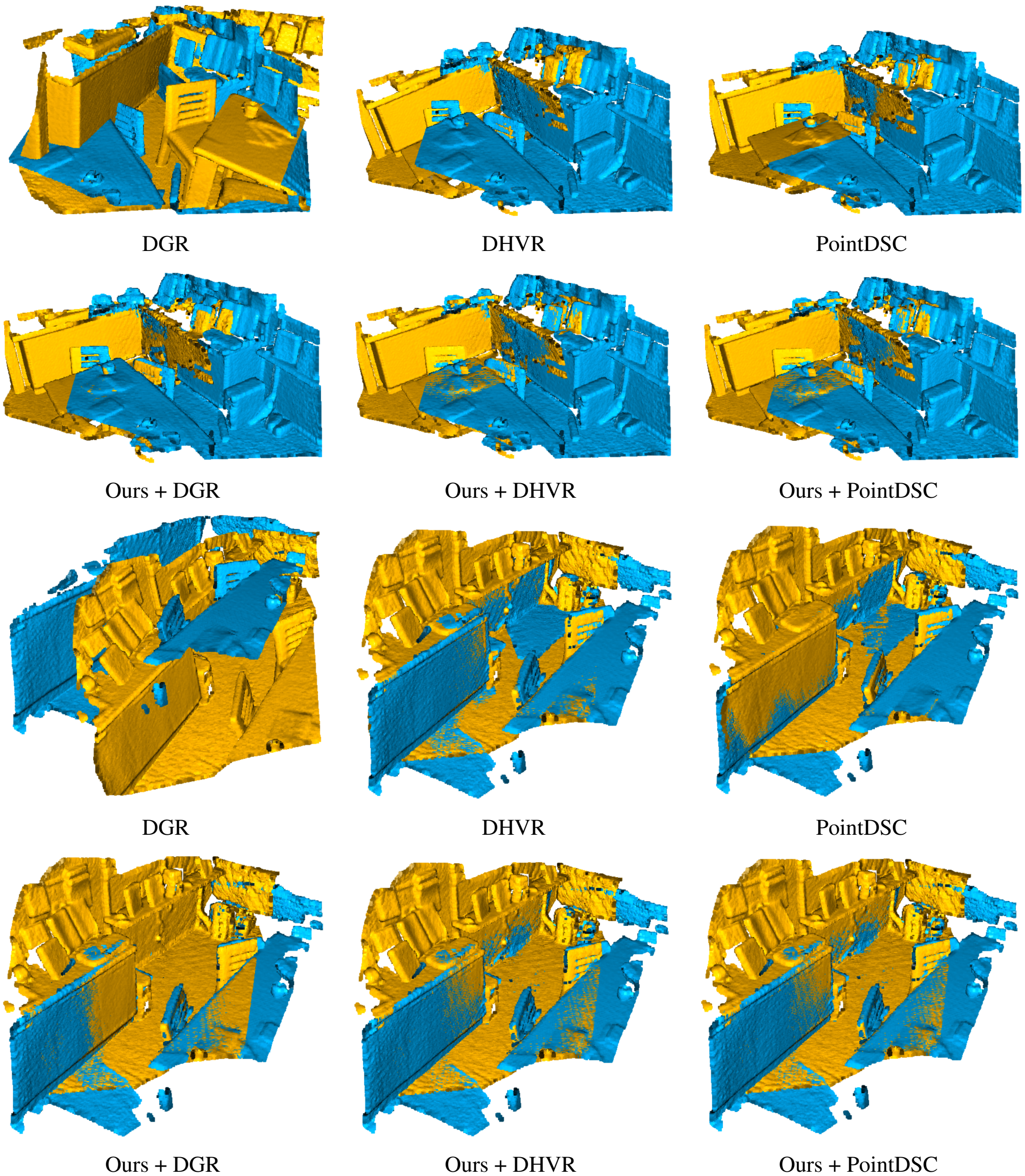

Figure 1: Qualitative results on 3DMatch dataset [7].

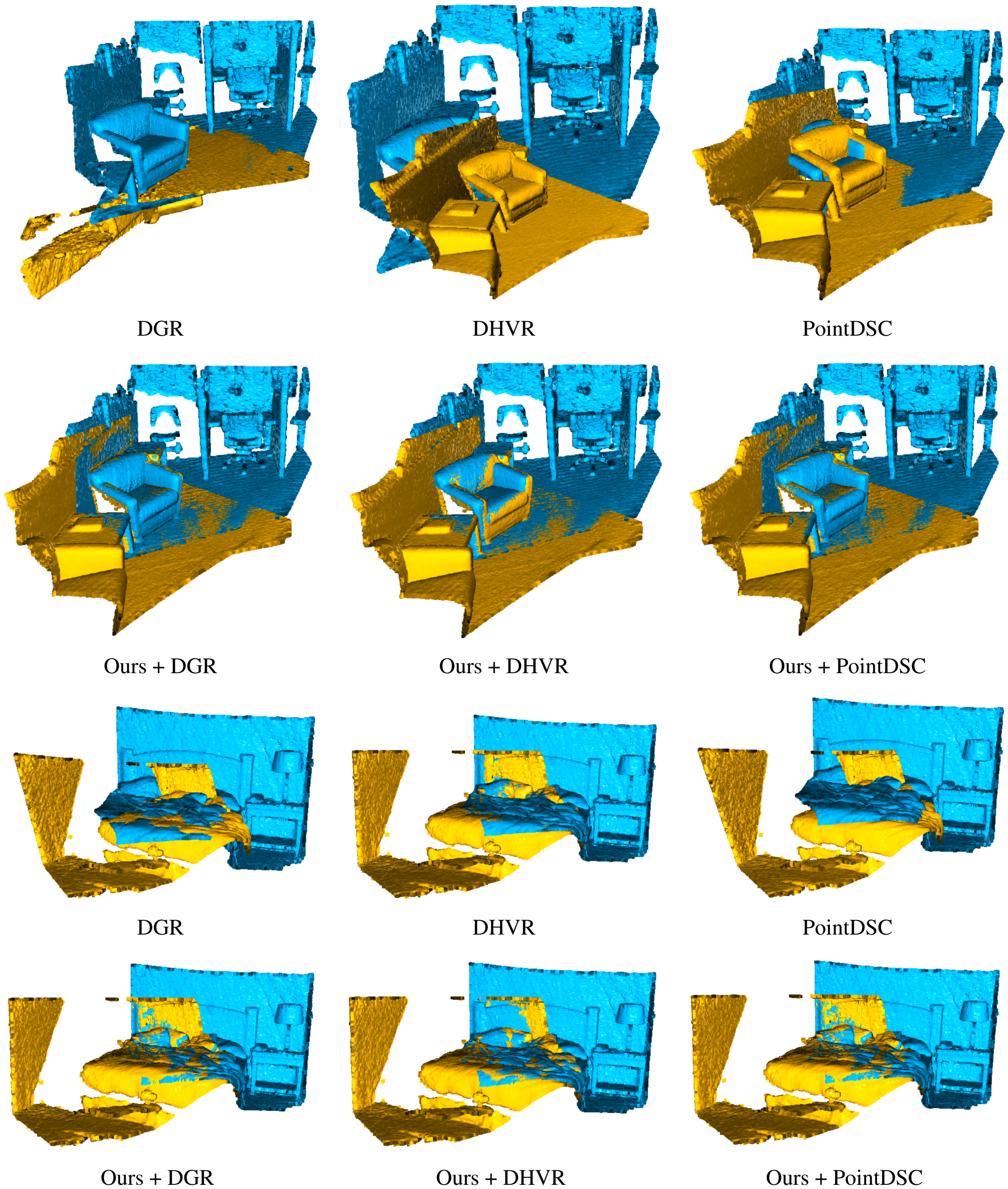

Figure 2: Qualitative results on 3DMatch dataset [7].

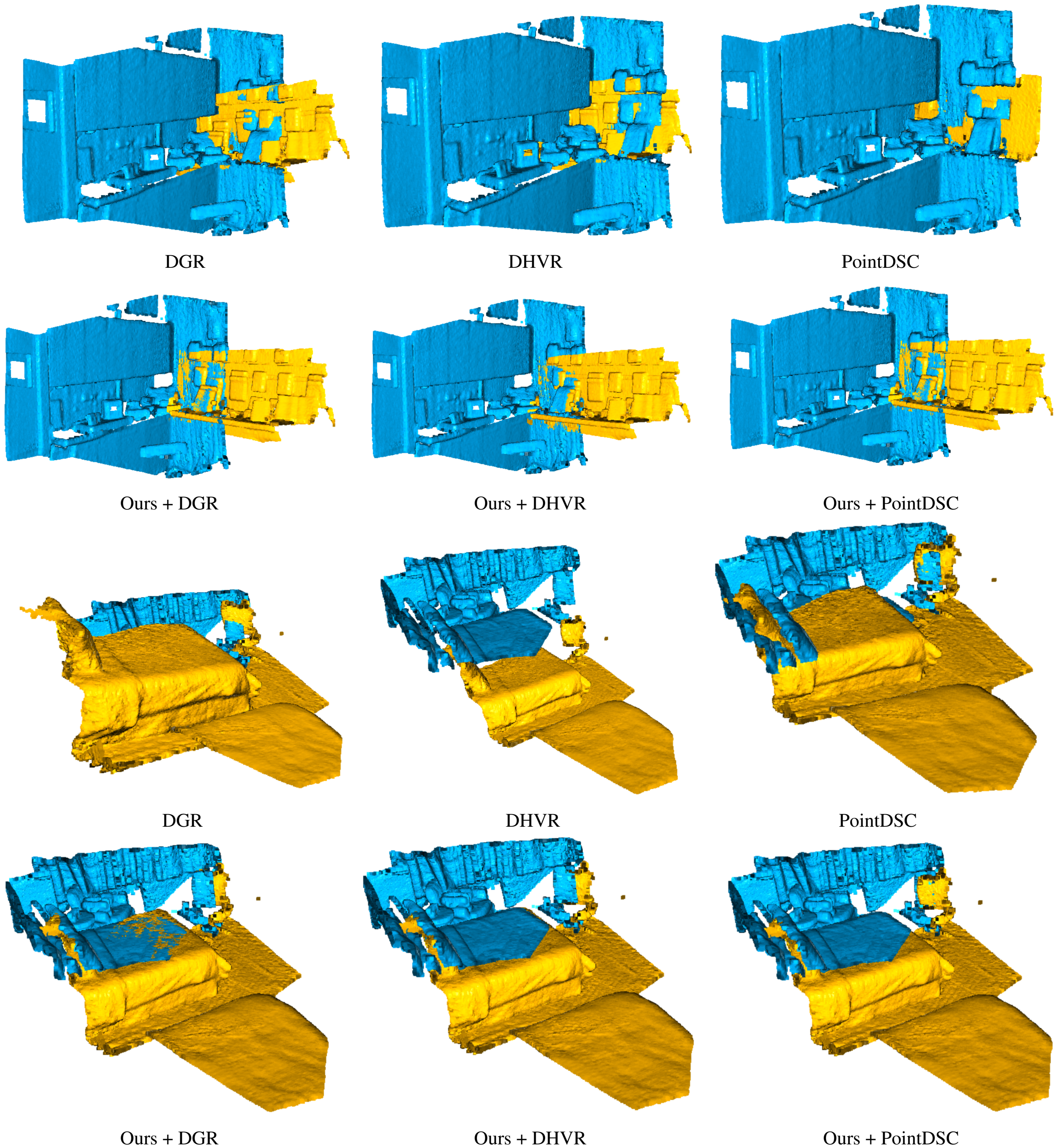

Figure 3: Qualitative results on low-overlapping 3DLoMatch dataset [4].

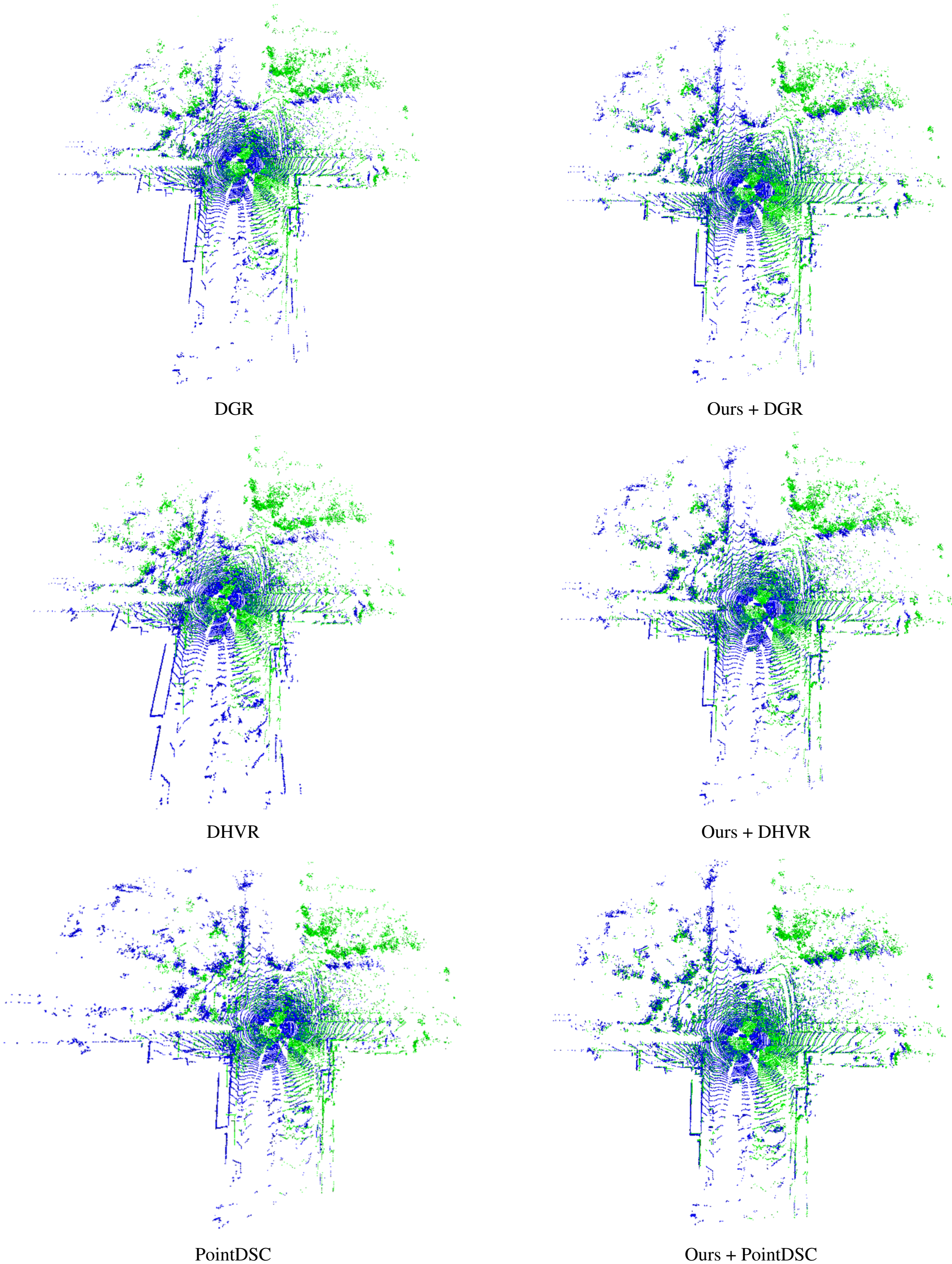

Figure 4: Qualitative results on KITTI dataset [3].

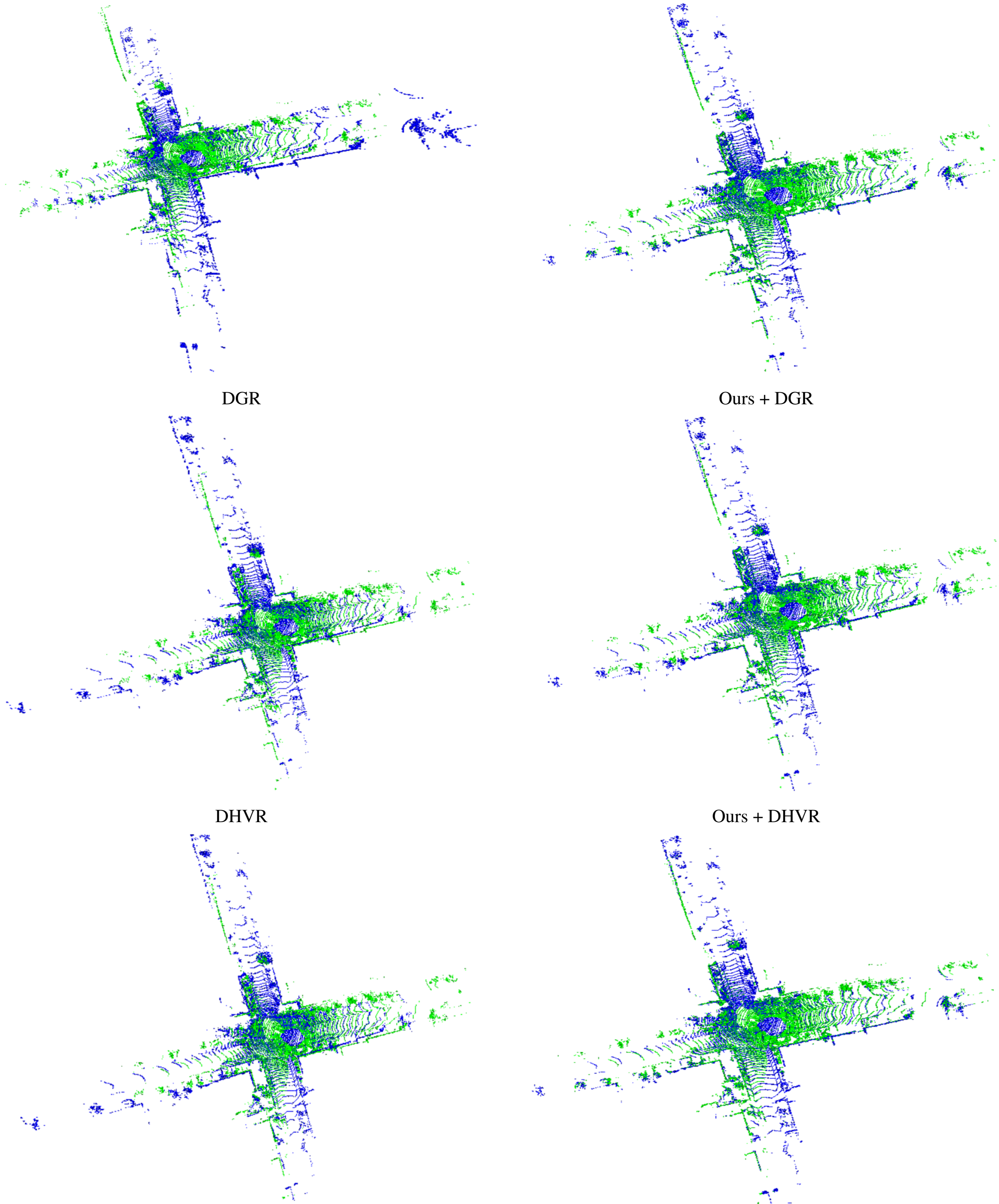

Figure 5: Qualitative results on KITTI dataset [3].