Supplementary Material for
LiDAR2Map: In Defense of LiDAR-Based Semantic Map Construction Using Online Camera Distillation

Song Wang, Wentong Li, Wenyu Liu, Xiaolu Liu, Jianke Zhu*
Zhejiang University
{songw, liwentong, liuwenyu.lwy, xiaoluliu, jkzhu}@zju.edu.cn

A. Training for Vehicle Segmentation

The training details for vehicle segmentation in Setting 1 and Setting 2 are slightly different from map segmentation. Also, we adopt Swin-Tiny [5] and PointPillars [3] as the feature extractors for image and LiDAR point cloud, respectively. The BEV feature pyramid decoder (BEV-FPD) uses a three-layer model with a trade-off between the accuracy and inference speed. We train the whole network for 15 epochs with 2 NVIDIA RTX 2080Ti GPUs. The learning rate is $1.5e^{-3}$, which decreases by a factor of 10 at the 10th epoch. The image size is set to 352×128 during training.

B. Additional Results

B.1. Map Segmentation

More Visual Results for BEV-FPD. We provide more visual results from the output of LiDAR2Map with different BEV-FPDs. In Fig. A1, the predicted semantic maps are gradually refined and become more accurate with the deepening of the number of layers, which further indicates the effectiveness of BEV-FPD on promoting our LiDAR2Map.

Comparison Under Different Weather and Light Conditions. As illustrated in Tab. A1, we compare LiDAR2Map with the state-of-the-art methods including HDMapNet-Fusion [4] and BEVerse [6] in different conditions. We employ PointPillars [3] as LiDAR backbone and 6-layer BEV-FPD for LiDAR2Map. Our method achieves the stable segmentation accuracy and outperforms other methods under different weather and light conditions. Fig. A2 provides the qualitative comparison in several typical scenarios. LiDAR2Map presents the superior capability in sunny, rainy and nighttime compared to HDMapNet-Fusion [4] and BEVerse [6]. Fig. A3 further reports more map predictions of our LiDAR2Map.

Method	Modality	Rainy	Night	All
HDMapNet-Fusion [4]	Camera & LiDAR	38.7	39.3	44.5
BEVerse* [6]	Camera	48.8	44.4	51.7
LiDAR2Map (Ours)	LiDAR	49.6	49.2	57.4

Table A1. Map segmentation results under different weather and light conditions on the nuScenes dataset. "$*$" means the results are obtained from its official public model.

B.2. Vehicle Segmentation

For vehicle segmentation, we provide the qualitative results on the nuScenes dataset with Setting 2 in Fig. A4. It obviously indicates that our method obtains the accurate vehicle predictions in different scenes.

C. Limitations and Future Work

The online Camera-to-LiDAR distillation scheme in our method incurs a certain amount of computation during the training, which increases the overall training time. Besides, the semantic map construction task relies on high-definition map annotations for the network training, which are only available in few datasets [1]. This limits the application of semantic map to more general autonomous driving scenarios. In the future, we will try to speed up the training process and explore the potential of LiDAR2Map with weakly-supervised forms, such as open street map [2].

References

[1] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, pages 11621–11631, 2020. 1

[2] Mordechai Haklay and Patrick Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, 7(4):12–18, 2008. 1

[3] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders

*Corresponding author is Jianke Zhu.

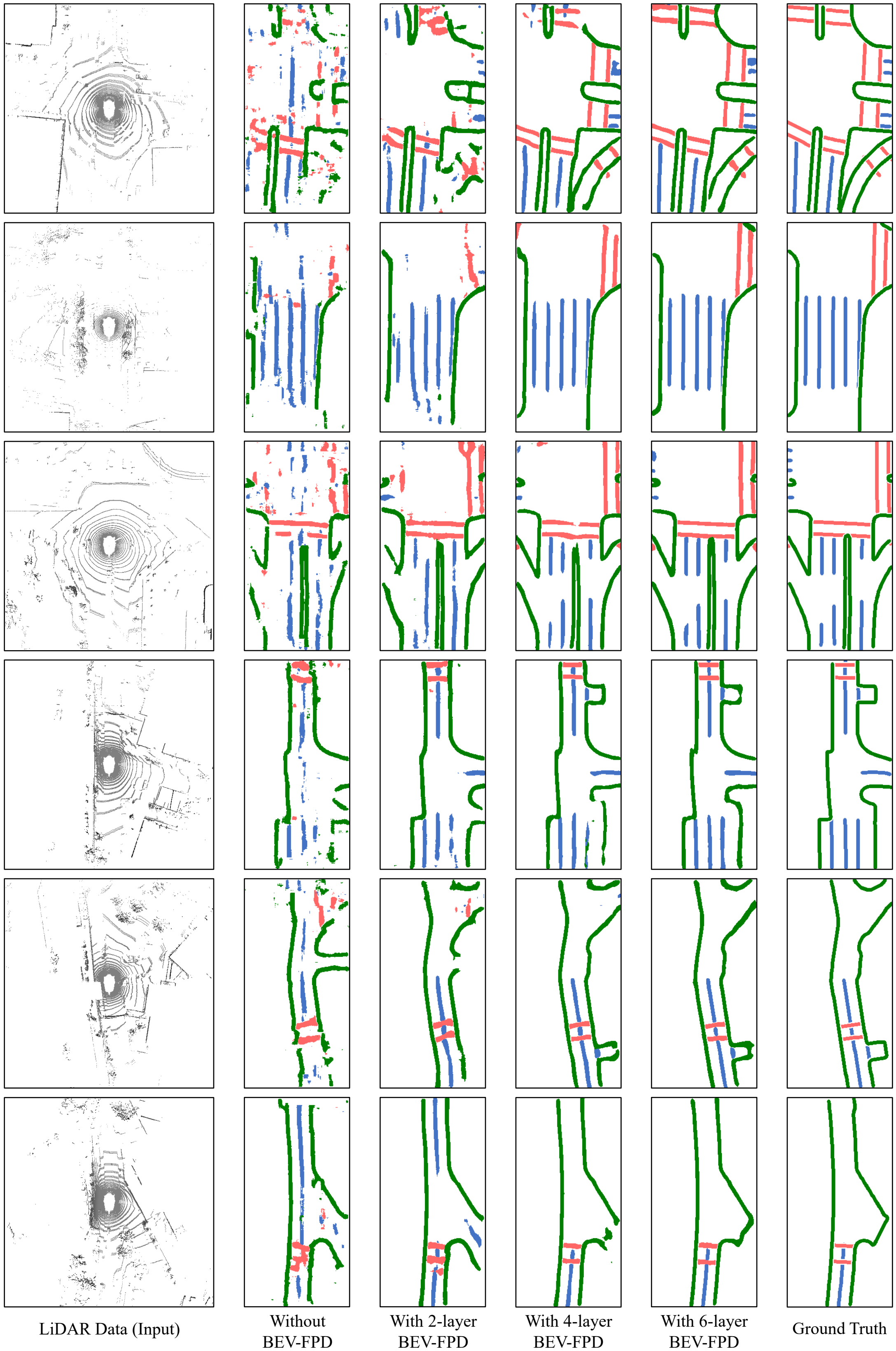


Figure A1. Additional visualization comparisons of LiDAR2Map with different BEV-FPDs on the nuScenes dataset.

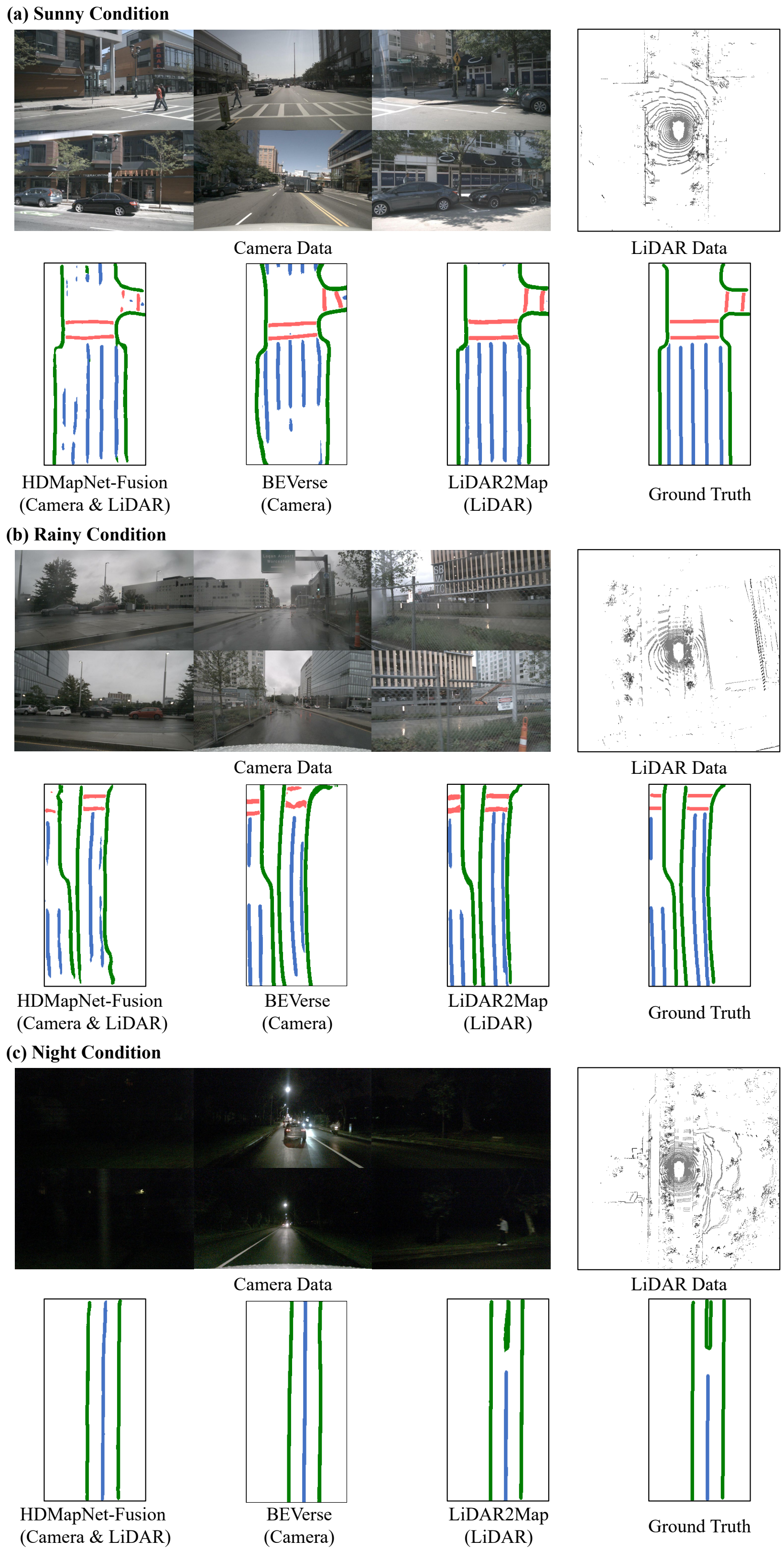


Figure A2. Qualitative results under various conditions. We compare our LiDAR2Map with other advanced approaches, including HDMapNet-Fusion [4] and BEVerse [6].

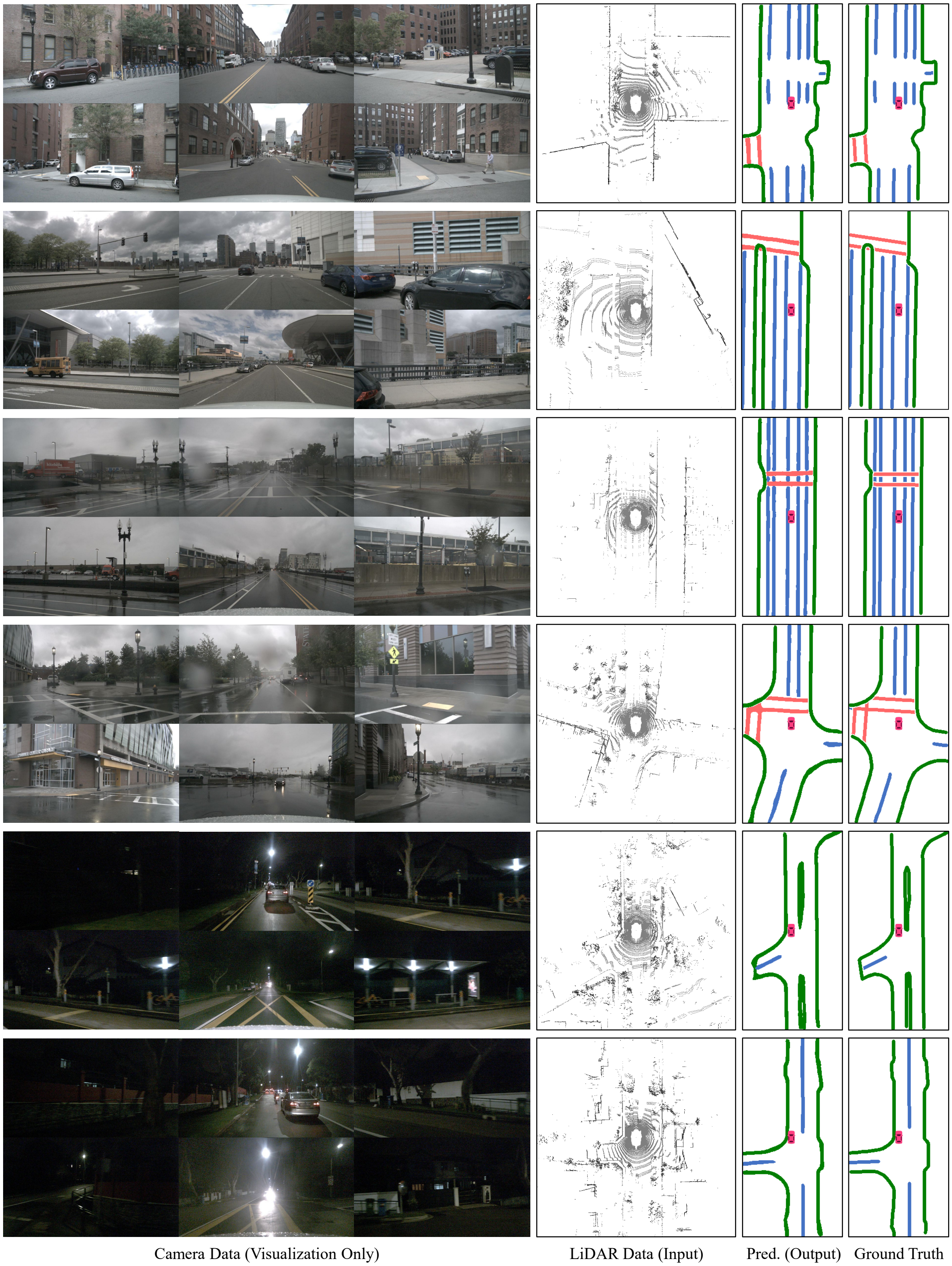
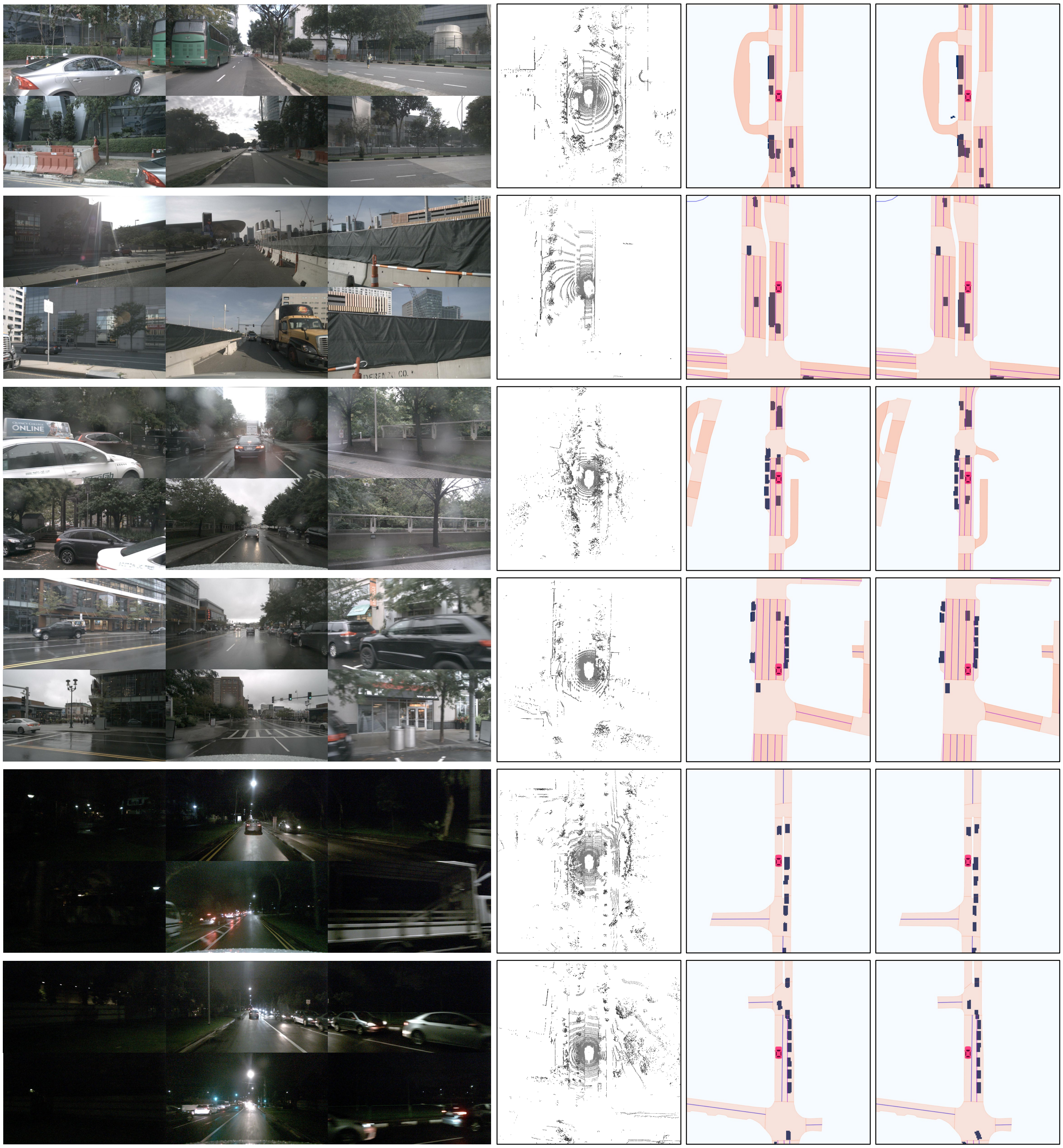


Figure A3. Additional visualization on map segmentation of LiDAR2Map with cloudy/rainy and day/night condition scenes.



Camera Data (Visualization Only) LiDAR Data (Input) Pred. (Output) Ground Truth

Figure A4. Visual vehicle segmentation of LiDAR2Map with cloudy/rainy and day/night condition scenes.

for object detection from point clouds. In *CVPR*, pages 12697–12705, 2019. 1

[4] Qi Li, Yue Wang, Yilun Wang, and Hang Zhao. Hdmapnet: An online hd map construction and evaluation framework. In *ICRA*, pages 4628–4634, 2022. 1, 3

[5] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*,

pages 10012–10022, 2021. 1

[6] Yunpeng Zhang, Zheng Zhu, Wenzhao Zheng, Junjie Huang, Guan Huang, Jie Zhou, and Jiwen Lu. Beverse: Unified perception and prediction in birds-eye-view for vision-centric autonomous driving. *arXiv preprint arXiv:2205.09743*, 2022. 1, 3