Supplementary Material: Vision-based Uneven BEV Representation Learning with Polar Rasterization and Surface Estimation

Zhi Liu [1,3 *] Shaoyu Chen [2,3 *] Xiaojie Guo [1†] Xinggang Wang [2]
Tianheng Cheng [2,3] Hongmei Zhu [3] Qian Zhang [3] Wenyu Liu [2] Yi Zhang [1]

[1] Tianjin University [2] Huazhong University of Science & Technology [3] Horizon Robotics

In this supplementary material, we provide more experiment details and visualizations about PolarBEV.

1 Experimental Details

Labels Generation The instance center labels are computed as a 2D Gaussian with standard deviation $\sigma_x = \sigma_y = 3$, centering at each instance center following previous method [1]. Each centerness label is a tensor of shape $1 \times H \times W$, indicating the likelihood of a pixel to be the center of an instance. Given an instance center, the offset labels are the vectors pointing from all other pixels of this instance to the center. We use the vehicle category to generate the segmentation label by projecting its 3D box annotations onto the ground plane.

Training Details We supervise the segmentation branch with a top-k cross entropy loss, while $L1$ loss for the centerness branch and offset branch. Instead of supervising the main output, we also calculate auxiliary loss for the output of each feature transformation with the same settings. After that, we add the auxiliary loss to the main loss with a coefficient α. Finally, we balance the three branches with a learnable weighting [2].

2 Visualizations

We provide more visualizations in Fig. 1. For each row, top shows the six camera views around the vehicle and the ground truth segmentation label, while the bottom shows the predicted centerness, offset, instance segmentation and segmentation results in turn.

References

[1] A. Hu, Z. Murez, N. Mohan, S. Dudas, J. Hawke, V. Badrinarayanan, R. Cipolla, and A. Kendall. Fiery: Future instance prediction in bird's-eye view from surround monocular cameras. In *ICCV*, 2021.

[2] A. Kendall, Y. Gal, and R. Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491.

6th Conference on Robot Learning (CoRL 2022), Auckland, New Zealand.

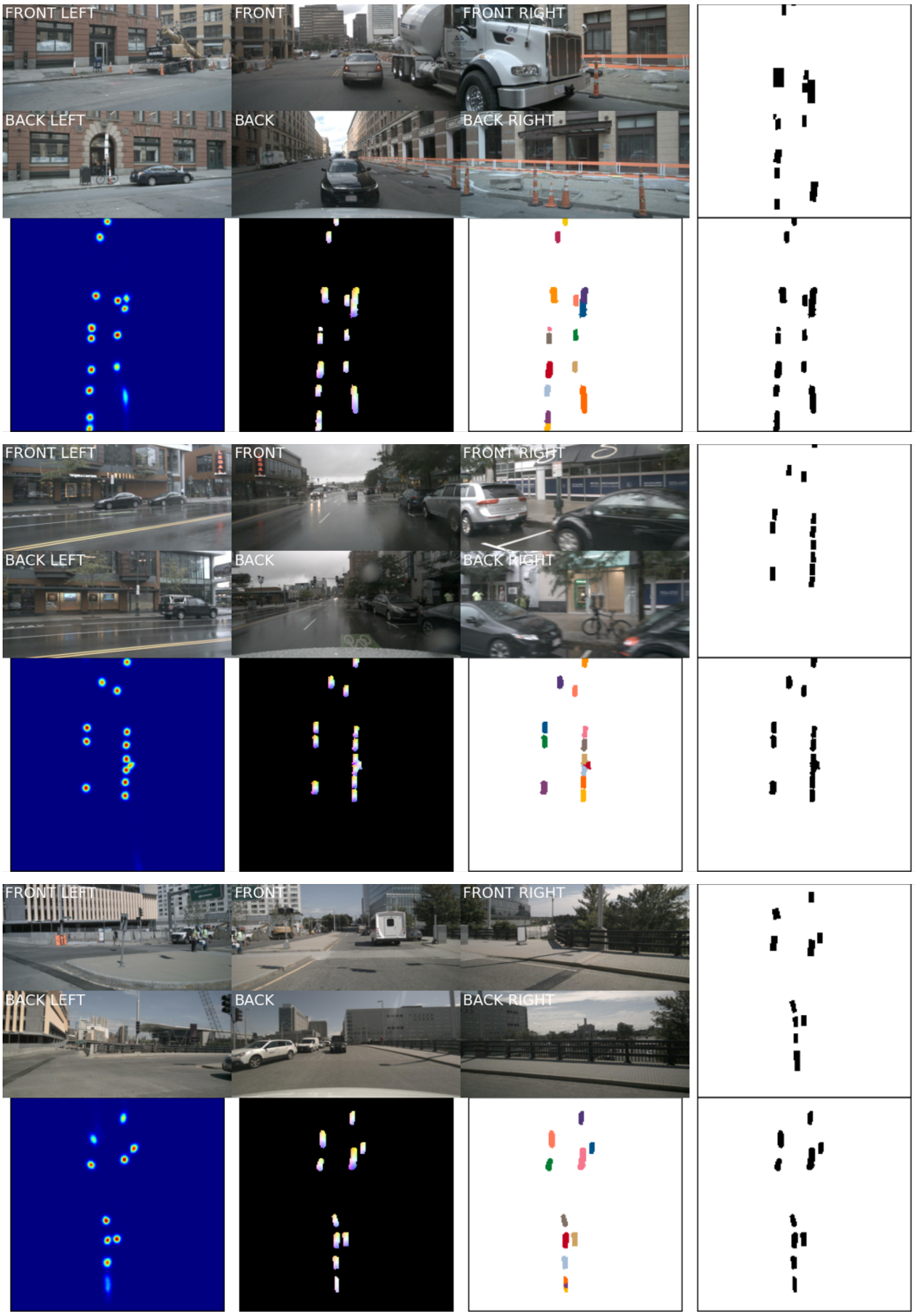

Figure 1: More visualizations on various scenes. Top shows the six camera views around the vehicle and the ground truth segmentation label. Bottom shows the predicted centerness, offset, instance segmentation and segmentation results in turn.