FB-BEV: BEV Representation from Forward-Backward View Transformations

Zhiqi Li[1,2], Zhiding Yu[2], Wenhai Wang[3], Anima Anandkumar[2,4], Tong Lu[1], Jose M. Alvarez[2]
[1]Nanjing University [2]NVIDIA [3]The Chinese University of Hong Kong [4]Caltech

1. Inference Speed

In Table 1, we show the inference speed of our proposed FB-BEV compared to BEVDet [1]/BEVDepth [2]. For the single-frame model, our FB-BEV is slightly slower than BEVDet. For models with multi-frame inputs, our model is faster than BEVDepth since our model owns smaller FLOPs.

2. Ablation of FRPN

We show that the IoU of the predicted foreground mask from FRPN in Table 2 and Table 3. From Table 2, we can observe that the temporal information can benefit the mask quality. In Table 3, we show the effect of foreground threshold t_f. Actually, the relationship between the IoU of the mask and final detection performance is is not positively correlated. Lowering the threshold will lead to lower IoU, but at the same time, more potential BEV grids will be input into the depth-aware backward projection module for optimization. Appropriately increasing the grid will improve performance, but increasing the transition will instead introduce interference features.

Table 1: Based on our experiments, we observe that FB-BEV yields slightly lower inference performance compared to BEVDet. However, we note that FB-BEV exhibits faster processing times than BEVDepth due to our model scale is much smaller than BEVDepth. The FPS is measured on RTX 3090 GPU.

Methods	**Temporal**	**Depth Sup.**	**mAP↑**	**NDS↑**	FPS
BEVDet	✗	✗	0.307	0.382	19.1
FB-BEV	✗	✗	0.312	0.406	16.1
BEVDepth	✓	✓	0.370	0.484	13.6
FB-BEV	✓	✓	0.378	0.498	15.3

Table 2: We show the IoU of the predicted foreground mask from FRPN.

Methods	**Temporal**	**Depth Sup.**	**mAP↑**	**NDS↑**
FB-BEV	✗	✗	0.312	0.406
FB-BEV	✓	✗	0.350	0.479
FB-BEV	✓	✓	0.378	0.498

Table 3: We show the effect of adopting different foreground threshold t_f.

Methods	**mAP↑**	**NDS↑**	t_f	IoU
FB-BEV	0.308	0.400	0.0	-
FB-BEV	0.312	0.406	0.3	31.2
FB-BEV	0.312	0.406	0.4	49.5
FB-BEV	0.312	0.406	0.5	49.3

References

[1] Junjie Huang, Guan Huang, Zheng Zhu, and Dalong Du. BEVDet: High-performance multi-camera 3d object detection in bird-eye-view. *arXiv preprint arXiv:2112.11790*, 2021. 1

[2] Yinhao Li, Zheng Ge, Guanyi Yu, Jinrong Yang, Zengran Wang, Yukang Shi, Jianjian Sun, and Zeming Li. BEVDepth: Acquisition of reliable depth for multi-view 3d object detection. *arXiv preprint arXiv:2206.10092*, 2022. 1