Enhancing 3D Object Detection with 2D Detection-Guided Query Anchors (Supplementary Material)

Haoxuanye Ji[2,*,♯] Pengpeng Liang[1,*,†] Erkang Cheng[2,‡]
[1]School of Computer and Artificial Intelligence, Zhengzhou University [2]Nullmax
jihaoxuanye@163.com, {liangpcs, twokang.cheng}@gmail.com

Appendix

A. More Implementation Details

We present the width range ($w_g^{\min}$, $w_g^{\max}$), height range ($h_g^{\min}$, $h_g^{\max}$), and length range ($l_g^{\min}$, $l_g^{\max}$) of each category g in Table 1. The interval used to sample width, height, and length candidates is 0.05m.

Category	$(w_g^{\min}, w_g^{\max})$	$(h_g^{\min}, h_g^{\max})$	$(l_g^{\min}, l_g^{\max})$
Car	(1.4, 2.8)	(1.2, 3.1)	(3.4, 6.6)
Pedestrian	(0.3, 1.0)	(1.0, 2.2)	(0.3, 1.3)
Bus	(2.6, 3.5)	(2.8, 4.6)	(6.9, 13.8)
Truck	(1.7, 3.5)	(1.7, 4.5)	(4.5, 14.0)
Trailer	(2.2, 2.3)	(3.3, 3.9)	(1.7, 14.0)
Construction vehicle	(2.1, 3.4)	(2.0, 3.0)	(3.7, 7.6)
Motorcycle	(0.4, 1.5)	(1.1, 2.0)	(1.2, 2.8)
Bicycle	(0.4, 0.9)	(0.9, 2.0)	(1.3, 2.0)
Traffic cone	(0.2, 1.2)	(0.5, 1.4)	(1.3, 2.0)
Barrier	(1.7, 3.6)	(0.8, 1.4)	(0.3, 0.8)

Table 1. The width, height, and length ranges of each class for anchor generation. The unit is meter.

B. More Visualization Results

Fig. 1 shows the visual comparison between BEVFormer-small-DAB3D [1] and its QAF2D-enhanced version. Case 1 and Case 3 show that QAF2D can make the detections more accurate. Case 2 demonstrates that QAF2D can help detect small objects that are missed by BEVFormer-small-DAB3D.

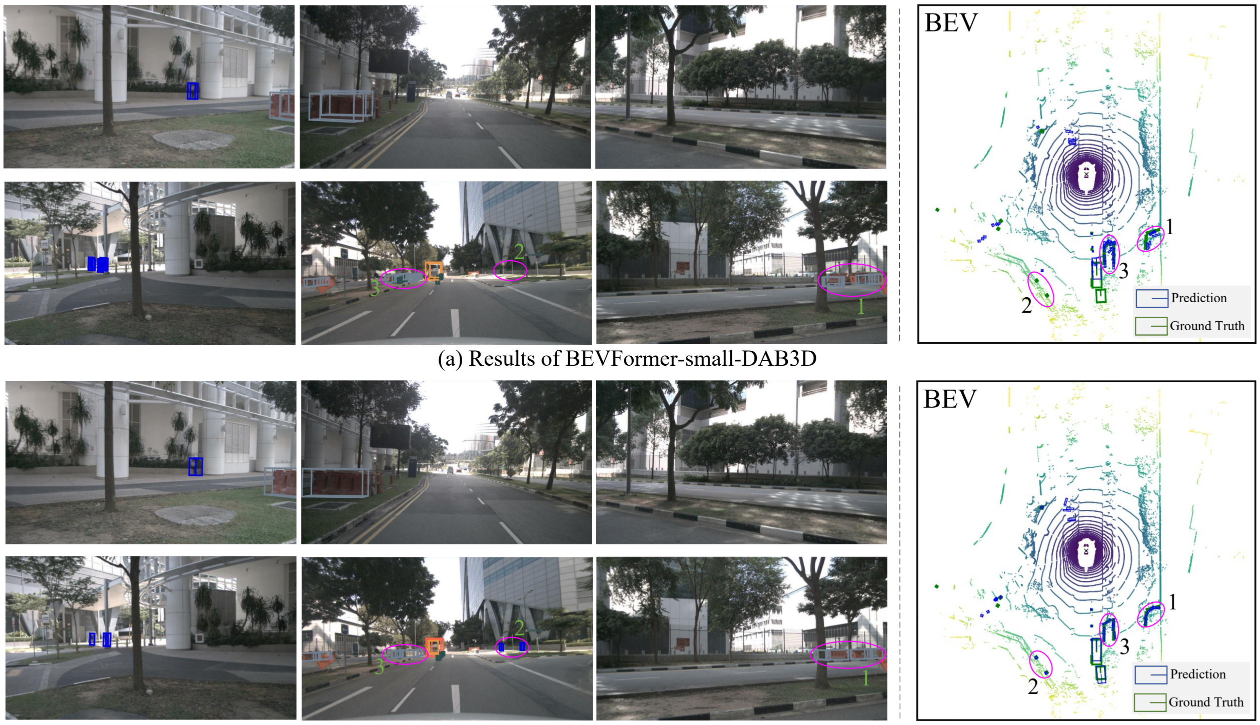
Fig. 2 shows the visual comparison between SparseBEV [2] and its QAF2D-enhanced version. Case 1 and Case 2 show that QAF2D is useful in improving the accuracy of detection results, and Case 3 demonstrates that QAF2D can alleviate the problem of missed detection of small objects.

[*] Equal contribution. [♯] Work done during an internship at Nullmax.
[†] Project lead. [‡] Corresponding author.

References

[1] Zhiqi Li, Wenhai Wang, Hongyang Li, Enze Xie, Chonghao Sima, Tong Lu, Yu Qiao, and Jifeng Dai. Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers. In *ECCV*, pages 1–18, 2022. 1

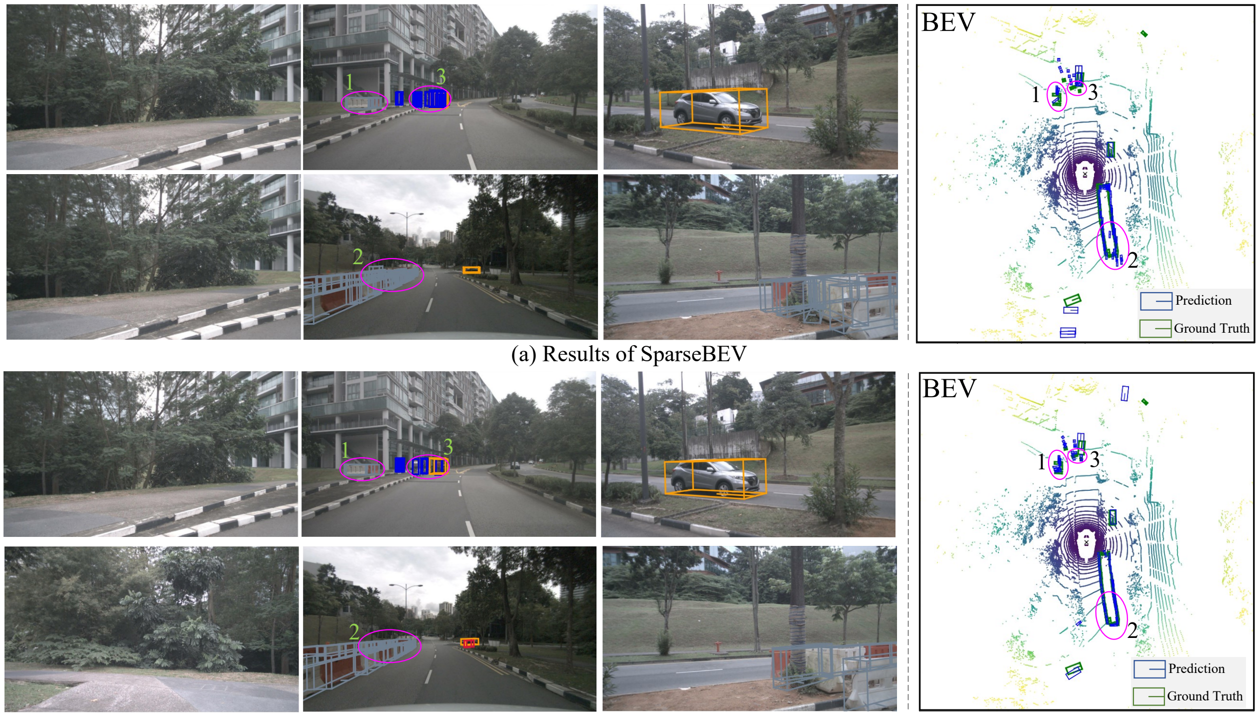
[2] Haisong Liu, Yao Teng, Tao Lu, Haiguang Wang, and Limin Wang. Sparsebev: High-performance sparse 3d object detection from multi-camera videos. In *ICCV*, pages 18580–18590, 2023. 1



(a) Results of BEVFormer-small-DAB3D

(b) Results of BEVFormer-small-QAF2D

Figure 1. Visualization results of BEVFormer-small-DAB3D and BEVFormer-small-QAF2D. The results in multi-camera images are shown on the left, and the corresponding results in bird's-eye-view are shown on the right.



(a) Results of SparseBEV

(b) Results of SparseBEV-QAF2D

Figure 2. Visualization results of SparseBEV and SparseBEV-QAF2D. The results in multi-camera images are shown on the left, and the corresponding results in bird's-eye-view are shown on the right.