Supplementary Materials for HINTED

Qiming Xia[1] Wei Ye[1] Hai Wu[1] Shijia Zhao[1] Leyuan Xing[1]
Xun Huang[1] Jinhao Deng[1] Xin Li[2] Chenglu Wen[1]* Cheng Wang[1]
[1]Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, Xiamen, China
[2]Section of Visual Computing and Interactive Media, Texas A&M University, Texas, USA

1. Evaluation on more datasets.

Besides the KITTI dataset evaluated in our paper, we've added experiments on the challenging nuScenes dataset. As shown in Table 1, our HINTED demonstrates significant advantages compared to the previous SOTA method CoIn.

Data	Method	mAP	NDS
nuScenes Sparsely-supervised	CenterPoint	8.09	25.77
	CoIn	12.47	33.79
	Ours	23.91	45.76

Table 1. The performance of HINTED on nuScenes dataset.

2. Discussing other density-related algorithms.

We have added a comparison with other density-related methods. As the results shown in Table 2, compared with previous methods, we took into account the relationship between the distribution of hard instances and distance in sparse settings. The designed MDS module is more suitable for detecting hard instances under sparse supervision. As a result, our approach achieves the best performance.

Cost	Method	Car-3D			Car-BEV		
		Easy	Mod.	Hard	Easy	Mod.	Hard
2%	Baseline	89.5	79.2	72.3	91.7	86.3	83.5
	DTS	85.6	76.7	72.4	90.0	85.2	82.3
	IASSD	89.7	80.1	76.9	94.6	88.7	85.8
	Ours	**94.3**	**82.5**	**78.7**	**95.7**	**90.1**	**87.1**

Table 2. Comparison with other density-related algorithms.

3. Comparison with PV-RCNN on Test Split of KITTI

Previous sparsely-supervised 3D object detection algorithms were only validated on the val split. In order to comprehensively assess the performance gap between our method and fully supervised algorithms, apart from validating on the val split, we also submitted the results obtained on the $test$ split to the KITTI official benchmark leaderboard. Tables 3 and 4 respectively present the comparison of our performance on the $test$ split with the fully supervised algorithm PVRCNN for 3D detection and BEV (Bird's Eye View) detection tasks. From the experimental results obtained on the $test$ split, our HINTED achieves over 90% performance compared to the fully supervised algorithms on both crucial detection benchmarks. This proves that our method's performance on the $test$ split aligns consistently with that on the val split.

*Corresponding author

4. The Detail of Fusion Module

Due to space constraints in the main text, a detailed explanation of the fusion process of mixed-density features is provided at this stage. As shown in Figure 1, inspired by SE block [1], we employ an attention mechanism to adaptively fuse features. For input feature map B_1, we initially down-sample its scale to match B_3 with average pooling. Subsequently, we further down-sample the feature map's scale to $1 \times 1 \times$ C with global average pooling. Finally, after passing through a fully connected layer and a sigmoid function, we obtain a weight λ_1. The calculation method for weights λ_2, λ_3, $\bar{\lambda}_1$, $\bar{\lambda}_2$ and $\bar{\lambda}_3$ follows a similar process as described above. The final mixed feature is obtained by combining these adaptive weights with the features.

5. More results

As shown in the Table 5, we present the results of the HINTED model for all evaluation metrics on the validation set in this section.

References

[1] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1

[2] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10526 – 10535, 2020. 2

Setting	Cost	Method	Car-3D			Ped.-3D			Cyc.-3D		
			Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard
Fully-supervised	100%	PV-RCNN[2]	90.25	81.43	76.82	52.17	43.29	40.29	78.60	63.71	57.65
Sparsely-supervised	2%	HINTED#	84.00	74.13	67.03	47.33	37.75	34.10	76.21	63.01	55.85
Percent(Avg=91.84%)			93.07%	91.03%	87.25%	90.72%	87.20%	84.63%	96.95	98.90%	96.87%

Table 3. Comparison with PV-RCNN on KITTI *test* split. The results are validated on 3D-detection benchmark. # indicates that TTA is not used.

Setting	Cost	Method	Car-BEV			Ped.-BEV			Cyc.-BEV		
			Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard
Fully-supervised	100%	PV-RCNN[2]	94.98	90.65 %	86.14 %	59.86 %	50.57 %	46.74 %	82.49 %	68.89 %	62.41 %
Sparsely-supervised	2%	HINTED	90.61 %	86.01 %	79.29 %	53.09 %	41.55 %	39.18 %	81.53 %	67.27 %	60.88 %
Percent(Avg=92.33%)			95.39%	94.88%	92.04%	88.69%	82.16%	83.82%	98.83%	97.64%	97.54%

Table 4. Comparison with PV-RCNN on KITTI *test* split. The results are validated on BEV-detection benchmark. # indicates that TTA is not used.

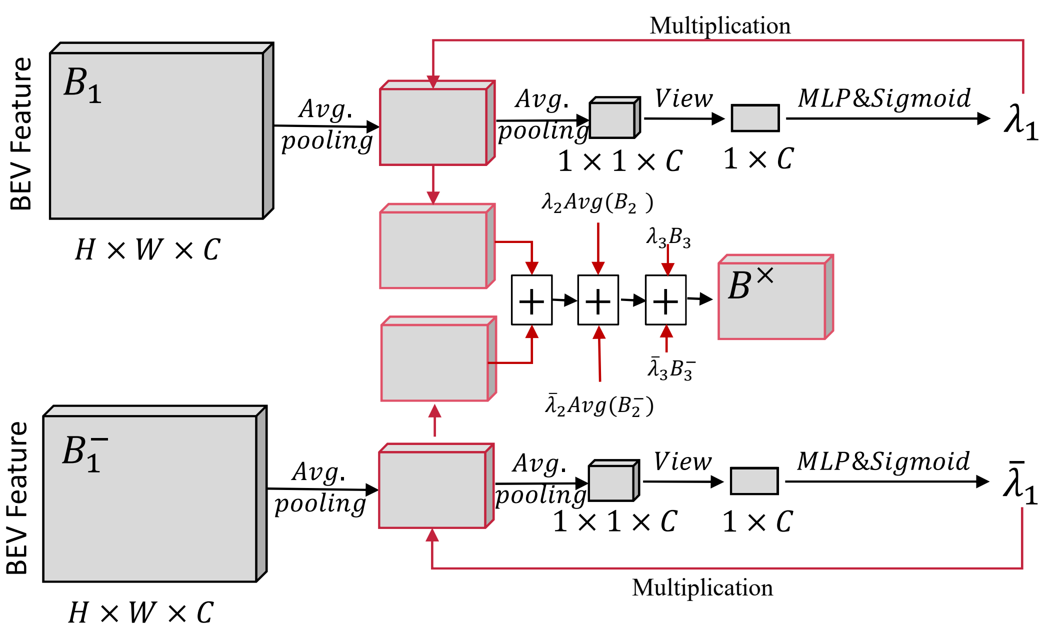

Figure 1. The illustration of fusion module.

IOU threshold	Metric	Car			Ped			Cyc		
		Easy	Mod	Hard	Easy	Mod	Hard	Easy	Mod	Hard
0.7, 0.5, 0.5	Bbox	98.74	91.76	88.78	74.50	68.43	62.24	96.35	82.22	79.31
	Bev	95.79	90.18	87.16	69.69	63.02	56.82	95.21	78.39	73.77
	3D	94.33	82.56	78.75	66.53	59.97	53.77	94.69	76.37	73.05
	Aos	98.20	90.99	87.85	70.35	63.80	57.91	96.17	81.44	78.51
0.5, 0.25, 0.25	Bbox	98.74	91.76	88.78	74.50	68.43	62.24	96.35	82.22	79.31
	Bev	98.70	93.41	90.86	79.68	72.95	66.63	95.23	78.51	75.25
	3D	98.59	91.91	90.50	79.49	72.78	66.50	95.23	78.51	75.25
	Aos	98.20	90.99	87.85	70.35	63.80	57.91	96.17	81.44	78.51

Table 5. Our method's results for all evaluation metrics on the *val* set.