This supplementary material contains the following three sections. Section A shows screenshots of the KITTI 3D and BEV leaderboard taken on the date of CVPR submission, showing the rank and time performance of our SE-SSD. Section B presents further ablation studies to analyze our ODIoU loss and shape-aware data augmentation on KITTI car dataset. Section C shows the 3D and BEV detection results of our baseline SSD and SE-SSD on the KITTI cyclist and pedestrian benchmarks. All our results on the KITTI *val* split are averaged from multiple runs and evaluated with the average precision of 40 sampling recall points.

A. KITTI Car Detection Leaderboards

As shown in Figures 1 and 2, our SE-SSD ranks $1st$ and $2nd$ on the KITTI BEV and 3D leaderboards of car detection [1], respectively, comparing with not only prior published works but also unpublished works submitted to the leaderboard. Also, our SE-SSD runs the fastest among the top submissions and achieved a balanced performance for the three difficulty levels, especially in 3D detection.

	Method	Setting	Code	Moderate	Easy	Hard	Runtime
1	SE-SSD			**91.84 %**	95.68 %	86.72 %	0.03 s
2	ADLAB			91.66 %	95.56 %	**86.92 %**	0.05 s
3	SPANet			91.59 %	95.59 %	86.53 %	0.06 s
4	PVGNet			91.26 %	94.36 %	86.63 %	0.05 s
5	SA-SSD		code	91.03 %	95.03 %	85.96 %	0.04 s
C. He, H. Zeng, J. Huang, X. Hua and L. Zhang: Structure Aware Single-stage 3D Object Detection fr							
6	MMLab PV-RCNN		code	90.65 %	94.98 %	86.14 %	0.08 s
S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang and H. Li: PV-RCNN: Point-Voxel Feature Set Abstr							
7	CN			90.50 %	94.51 %	85.86 %	0.04 s
8	Anonymous		code	90.46 %	92.83 %	85.94 %	0.05 s
9	DSA-PV-RCNN		code	90.13 %	92.42 %	85.93 %	0.08 s
10	Associate-3Ddet_v2			90.00 %	95.55 %	84.72 %	0.04 s

Figure 1. KITTI BEV (Bird's Eye View) car detection leaderboard, in which our SE-SSD ranks the $1st$ place.

	Method	Setting	Code	Moderate	Easy	Hard	Runtime
1	HRI-ADLab-HZ			**82.83 %**	89.00 %	76.00 %	0.1 s
2	SE-SSD			82.54 %	91.49 %	77.15 %	0.03 s
3	BorderAtt			82.33 %	87.77 %	77.37 %	0.08 s
4	HUAWEI Octopus			82.13 %	88.26 %	**77.41 %**	0.1 s
5	ADLAB			82.08 %	90.92 %	77.36 %	0.05 s
6	PV-RCNN-v2			81.88 %	90.14 %	77.15 %	0.06 s
7	RangeRCNN-LV			81.85 %	88.76 %	77.18 %	0.1 s
8	PVGNet			81.81 %	89.94 %	77.09 %	0.05 s
9	Anonymous		code	81.63 %	90.26 %	76.88 %	0.05 s
10	Voxel R-CNN			81.62 %	90.90 %	77.06 %	0.04 s

Figure 2. KITTI 3D car detection leaderboard, in which our SE-SSD ranks the $2nd$ place (HRI-ADLab-HZ is unpublished).

B. More Ablation Studies

Shape-aware data augmentation We analyze the effect of random dropout, swap, and sparsifying in our shape-aware data augmentation on KITTI *val* split for car detection, respectively. As Table 1 shows, all these operators (random dropout, swap, and sparsifying) boost the 3D moderate AP effectively, thus demonstrating the effectiveness of our proposed augmentation operators to enrich the object diversity.

Type	baseline	dropout	swap	sparsify	Full SA-DA
Moderate AP	83.22	83.46	83.48	83.43	**83.70**

Table 1. Ablation study on the operators (random dropout, swap, and sparsifying) in our shape-aware data augmentation (SA-DA).

ODIoU Loss Next, we try different values of γ in the ODIoU loss on KITTI *val* split for car detection. As Table 2 shows, the orientation constraint is an important factor to further boost the precision, so we finally set γ as 1.25.

γ	0.25	0.5	0.75	1.0	1.25	1.5	1.75
Moderate AP	83.47	83.65	83.73	83.78	**83.85**	83.58	83.52

Table 2. Ablation study on our ODIoU loss, in which we compare the 3D moderate AP of different settings of γ.

C. Experiments on KITTI Cyclist&Pedestrian

To validate the effectiveness of our SE-SSD framework, we further conduct experiments on the Cyclist and Pedestrian datasets in KITTI benchmark. In Tables 3 and 4, we compare the 3D and BEV average precisions between the baseline SSD and our SE-SSD on KITTI *val* split.

Cyclist & Pedestrian Results As Table 3 shows, our SE-SSD outperforms the baseline SSD by a large margin for both 3D and BEV cyclist detection, especially on the 3D moderate and hard subsets with an improvement of about 15 points. As Table 4 shows, our SE-SSD also outperforms the baseline SSD on both the 3D and BEV pedestrian detection by a large margin. These large improvements clearly show the effectiveness of our proposed SE-SSD framework.

Cyclist		Easy	Moderate	Hard
3D	SSD	75.73	55.86	51.97
	our SE-SSD	**80.07**	**70.43**	**66.45**
BEV	SSD	83.71	59.02	55.05
	our SE-SSD	**91.83**	**72.62**	**68.24**

Table 3. Comparison of 3D and BEV APs between our baseline SSD and SE-SSD on KITTI *val* split for "cyclist" detection.

Pedestrian		Easy	Moderate	Hard
3D	SSD	59.64	52.63	46.59
	our SE-SSD	**63.27**	**57.32**	**50.82**
BEV	SSD	63.53	57.29	51.36
	our SE-SSD	**67.47**	**61.88**	**55.94**

Table 4. Comparison of 3D and BEV APs between our baseline SSD and SE-SSD on KITTI *val* split for "pedestrian" detection.

[1] On the date of CVPR deadline, *i.e.*, Nov 16, 2020