SPG: Unsupervised Domain Adaptation for 3D Object Detection via Semantic Point Generation

Qiangeng Xu[1 †] Yin Zhou[2] Weiyue Wang[2] Charles R. Qi[2] Dragomir Anguelov[2]
[1]University of Southern California [2]Waymo, LLC
{qiangenx}@usc.edu {yinzhou,weiyuewang,rqi,dragomir}@waymo.com

In this supplementary material, we provide detailed analysis about the statistics of the Waymo Domain Adaptation Dataset in Section A; the robustness analysis of the foreground voxel classifier in Section B; the derivation of the dropout rate used in the RndDrop method in Section C; more results on the Waymo Domain Adaptation Dataset in Section D; more results on KITTI in Section E; and more visualization of the semantic point generation in Section F.

A. Statistics of the Waymo Domain Adaptation Dataset

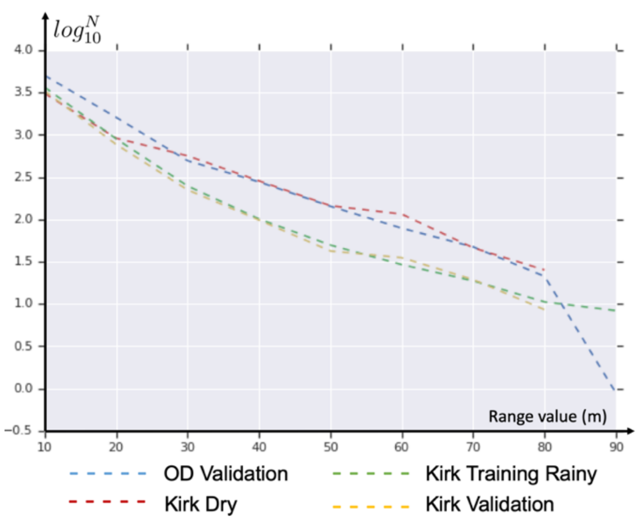


Figure 1: The average number of raw points per vehicle across different ranges. On the x axis, the range value stands for the distance between the center of a bounding box and the LiDAR sensor. The y axis shows the value after applying log_{10} on the number of points N. "Kirk Dry" is extracted from the Kirk Training set and contains frames captured in the dry weather.

We collect the statistics about the average number of points in a vehicle bounding box across different ranges. The range value is calculated as the euclidean distance between the LiDAR sensor and the center of a bounding box. We investigate four sets of point clouds:

- The OD Validation set, in which 99.5% of the frames are collected in the dry weather.
- The Kirk Dry set, which consists of all the frames with the dry weather condition from the Kirk training set.
- The Kirk Training Rainy set, which consists of all the frames with the rainy weather condition from the Kirk training set.
- The Kirk Validation set, in which all the frames are collected in the rainy weather.

As shown in Figure 1, the point clouds with similar weather conditions share similar numbers of points per object, even though they are collected at different locations. Specifically, the vehicle objects of the two "dry datasets", *i.e.,* the Kirk Dry set and the OD Validation set, have similar numbers of points across all ranges. The vehicle objects of the two "rainy datasets" *i.e.,* the Kirk Training Rainy set and the Kirk Validation set, share similar statistics.

In addition, the point clouds captured in **the dry weather** (the OD Validation set and the Kirk Dry set) have more points on each object than those collected in **the rainy weather** (the Kirk Training Rainy set and the Kirk Validation set). Please note that we have applied log_{10} to the number of points for better visualization. The difference in the number of points is substantial between two weather conditions across all ranges.

B. The Robustness of the Foreground Voxel Classifier

In order to generalize detectors to different domains, it is crucial to correctly classify foreground voxels so that semantic points can be reliably generated. Table 1 lists the evaluation results of the foreground voxel classifier. The results in Table 1 are averaged among all voxels in the foreground regions. Our SPG is trained on the OD training set. Then it is evaluated on the OD validation set and the Kirk validation set, respectively. The classification of a voxel is correct if its prediction score $\tilde{P}^f > 0.5$ when $y^f = 1.0$ or

Train	Eval	Accuracy	Precision	Recall	AP
OD Train	OD Val	99.3 %	90.9 %	92.9 %	86.7 %
OD Train	Kirk Val	98.9 %	88.4 %	88.2 %	78.3 %

Table 1: Foreground voxel classification results of our SPG. The model is trained on the OD training set and then it is evaluated on the OD validation set and Kirk validation set, respectively. The accuracy, precision and recall are evaluated by setting $\tilde{P}^f > 0.5$.

$\tilde{P}^f < 0.5$ when $y^f = 0.0$. The accuracy, precision and recall are all calculated under this setting. The AP is calculated using 40 recall thresholds. The results show that SPG achieves high performance in both domains.

C. Dropout Rate of the RndDrop Method

In the experiment section, we implement a baseline method RndDrop, where we randomly drop out 17% of points for point clouds from the source domain during training. This dropout ratio is chosen to match the ratio of missing points in the target domain. We calculate $(\overline{N}_{src} - \overline{N}_{tgt})/\overline{N}_{src} = 17\%$, where $\overline{N}_{src} = 121.2K$ is the average number of points per scene in the source domain and $\overline{N}_{tgt} = 100.4K$ is the average number of points per scene in the target domain.

D. More Results on the Waymo Domain Adaptation Dataset

The evaluation tool [16] provides the average precision for three distance-based breakdowns: 0 to 30 meters, 30 to 50 meters, and beyond 50 meters. The AP is calculated using 100 recall thresholds.

We perform two groups of model comparisons in the setting of UDA: Group 1. PointPillars vs. SPG + PointPillars; Group 2. PV-RCNN vs. SPG + PV-RCNN. We train all models on the OD training set and evaluate them on both the OD validation set and the Kirk validation set. Table 2 and 3 show the comparisons on vehicle 3D AP and vehicle BEV AP, respectively. Table 4 and Table 5 show the comparisons in pedestrian 3D AP and pedestrian BEV AP, respectively. In most cases, SPG improves the detection performance across all ranges for both vehicles and pedestrians.

E. More Results on KITTI

We provide more 3D object detection results on KITTI. There are two commonly used metric standards for evaluating the detection performance: 1) R11, where the AP is evaluated with 11 recall positions; 2) R40, where the AP is evaluated with 40 recall positions. In addition to the improvement on car and pedestrian detection, SPG also significantly boosts the performance in cyclist detection. Based on R11, Table 6 and Table 7 show the results in 3D AP and BEV AP for three object types, respectively. Based on R40, Table 8 and Table 9 show the results in 3D AP and BEV AP for three object types, respectively.

We show more comparisons on the KITTI test set in Table 10.

F. More Visualization of Semantic Point Generation

In Figure 3, we illustrate more augmented point clouds, where the raw points are rendered in the grey color and the generated semantic points are highlighted in red.

		Target Domain - Kirk				Source Domain - OD			
		Vehicle 3D AP (IoU = 0.7)				Vehicle 3D AP (IoU = 0.7)			
Difficulty	Method	Overall	0-30m	30-50m	50-Inf	Overall	0-30m	30-50m	50-Inf
	PointPillars	34.65	63.13	24.56	7.65	57.27	84.39	52.97	28.22
LEVEL_1	SPG + PointPillars	**41.56**	**68.26**	**31.91**	**13.08**	**62.44**	**86.18**	**58.13**	**35.40**
	Improvement	*+6.91*	*+5.13*	*+7.35*	*+5.43*	*+5.17*	*+1.79*	*+5.16*	*+7.18*
	PointPillars	31.67	59.26	22.09	7.08	52.96	82.30	50.74	24.6
LEVEL_2	SP + PointPillar	**38.15**	**64.57**	**28.66**	**11.96**	**58.54**	**85.75**	**56.02**	**31.02**
	Improvement	*+6.48*	*+5.31*	*+6.57*	*+4.88*	*+5.58*	*+3.45*	*+5.28*	*+6.42*
	PV-RCNN	55.16	76.68	47.96	27.59	74.01	91.39	70.94	49.51
LEVEL_1	SPG+PV-RCNN	**58.31**	**77.81**	**51.65**	**31.29**	**75.27**	**92.36**	**73.47**	**51.03**
	Improvement	*+3.15*	*+1.13*	*+3.69*	*+3.70*	*+1.26*	*+0.97*	*+2.53*	*+1.52*
	PV-RCNN	45.81	71.31	38.83	20.52	64.69	88.95	64.80	37.37
LEVEL_2	SPG + PV-RCNN	**48.70**	**72.41**	**42.16**	**23.52**	**65.98**	**91.62**	**65.61**	**39.87**
	Improvement	*+2.89*	*+1.10*	*+3.33*	*+3.00*	*+1.29*	*+2.67*	*+0.81*	*+2.50*

Table 2: The unsupervised domain adaptation vehicle detection results on both Waymo Open Dataset (OD) and Kirkland Dataset (Kirk). We show the vehicle 3D AP results in this table. The AP distance breakdowns are provided by the official evaluation tool.

		Target Domain - Kirk				Source Domain - OD			
		Vehicle BEV AP (IoU = 0.7)				Vehicle BEV AP (IoU = 0.7)			
Difficulty	Method	Overall	0-30m	30-50m	50-Inf	Overall	0-30m	30-50m	50-Inf
	PointPillars	51.88	75.56	46.04	25.55	72.26	92.23	71.35	51.11
LEVEL_1	SPG + PointPillars	**60.44**	**80.89**	**53.73**	**38.24**	**77.63**	**93.39**	**75.96**	**61.16**
	Improvement	*+8.56*	*+5.33*	*+7.69*	*+12.69*	*+5.37*	*+1.16*	*+4.61*	*+10.05*
	PointPillars	47.93	71.18	42.41	23.47	69.09	91.83	68.87	45.53
LEVEL_2	SPG + PointPillars	**56.94**	**77.13**	**49.99**	**35.04**	**74.90**	**93.06**	**73.96**	**54.51**
	Improvement	*+9.01*	*+5.95*	*+7.58*	*+11.57*	*+5.81*	*+1.23*	*+5.09*	*+8.98*
	PV-RCNN	70.38	84.27	65.31	52.98	85.13	95.99	84.02	72.19
LEVEL_1	SPG + PV-RCNN	**72.56**	**84.43**	**68.79**	**58.49**	**87.38**	**97.54**	**86.63**	**74.59**
	Improvement	*+2.18*	*+0.16*	*+3.48*	*+5.51*	*+2.25*	*+1.55*	*+2.61*	*+2.40*
	PV-RCNN	60.13	78.10	54.36	40.67	76.84	93.29	76.64	58.29
LEVEL_2	SPG + PV-RCNN	**62.03**	**78.86**	**56.47**	**44.94**	**78.05**	**94.45**	**80.25**	**59.56**
	Improvement	*+1.90*	*+0.76*	*+2.11*	*+4.27*	*+1.21*	*+1.16*	*+3.61*	*+1.27*

Table 3: The unsupervised domain adaptation vehicle detection results on both Waymo Open Dataset (OD) and Kirkland Dataset (Kirk). We show the vehicle BEV AP results in this table. The AP distance breakdowns are provided by the official evaluation tool.

		Target Domain - Kirk				Source Domain - OD			
		Pedestrian 3D AP (IoU = 0.5)				Pedestrian 3D AP (IoU = 0.5)			
Difficulty	Method	Overall	0-30m	30-50m	50-Inf	Overall	0-30m	30-50m	50-Inf
LEVEL_1	PointPillars	20.65	43.98	9.27	3.24	55.20	69.24	52.04	32.72
	SPG + PointPillars	**23.72**	**50.19**	**9.11**	**5.57**	**56.06**	**69.32**	**53.12**	**34.73**
	Improvement	*+3.07*	*+6.21*	*-0.16*	*+2.33*	*+0.86*	*+0.08*	*+1.08*	*+2.01*
LEVEL_2	PointPillars	17.66	40.67	7.40	2.32	51.33	65.85	49.32	29.29
	SPG + PointPillars	**19.57**	**46.42**	**7.44**	**3.99**	**52.33**	**65.63**	**50.10**	**31.25**
	Improvement	*+1.91*	*+5.75*	*+0.04*	*+1.67*	*+1.00*	*-0.22*	*+0.78*	*+1.96*
LEVEL_1	PV-RCNN	24.47	39.69	14.24	8.05	65.34	72.23	64.89	50.04
	SPG + PV-RCNN	**30.82**	**48.04**	**18.80**	**13.39**	**66.93**	**73.55**	**66.60**	**50.82**
	Improvement	*+6.35*	*+8.35*	*+4.56*	*+5.34*	*+1.59*	*+1.32*	*+1.71*	*+0.78*
LEVEL_2	PV-RCNN	17.16	36.39	9.64	3.51	56.03	66.88	56.58	35.76
	SPG + PV-RCNN	**22.05**	**44.07**	**12.91**	**5.77**	**57.68**	**68.28**	**58.29**	**37.64**
	Improvement	*+4.89*	*+7.68*	*+3.27*	*+2.26*	*+1.65*	*+1.40*	*+1.71*	*+1.88*

Table 4: The unsupervised domain adaptation pedestrian detection results on both Waymo Open Dataset (OD) and Kirkland Dataset (Kirk). We show the pedestrian 3D AP results in this table. The AP distance breakdowns are provided by the official evaluation tool.

		Target Domain - Kirk				Source Domain - OD			
		Pedestrian BEV AP (IoU = 0.5)				Pedestrian BEV AP (IoU = 0.5)			
Difficulty	Method	Overall	0-30m	30-50m	50-Inf	Overall	0-30m	30-50m	50-Inf
LEVEL_1	PointPillars	22.33	45.00	10.50	3.49	63.82	76.33	61.90	42.81
	SPG + PointPillars	**24.83**	**51.44**	**10.80**	**5.71**	**64.66**	**76.11**	**62.69**	**44.98**
	Improvement	*+2.50*	*+6.44*	*+0.30*	*+2.22*	*+0.84*	*-0.22*	*+0.79*	*+2.17*
LEVEL_2	PointPillars	18.40	41.63	8.58	2.49	60.13	73.34	58.77	38.83
	SPG + PointPillars	**20.67**	**47.56**	**8.98**	**4.11**	**60.93**	**72.94**	**59.54**	**41.11**
	Improvement	*+2.27*	*+5.93*	*+0.40*	*+1.62*	*+0.80*	*-0.40*	*+0.77*	*+2.28*
LEVEL_1	PV-RCNN	25.39	40.23	14.72	9.76	70.35	76.22	70.49	56.77
	SPG + PV-RCNN	**31.92**	**49.06**	**19.87**	**14.87**	**70.37**	**75.86**	**72.29**	**57.47**
	Improvement	*+6.53*	*+8.83*	*+5.15*	*+5.11*	*+0.02*	*-0.36*	*+1.80*	*+0.70*
LEVEL_2	PV-RCNN	17.88	36.89	9.97	4.23	60.81	69.22	61.86	41.32
	SPG + PV-RCNN	**22.65**	**44.57**	**13.48**	**6.38**	**60.88**	**70.62**	**63.65**	**43.27**
	Improvement	*+4.77*	*+7.68*	*+3.51*	*+2.15*	*+0.07*	*+1.40*	*+1.79*	*+1.95*

Table 5: The unsupervised domain adaptation pedestrian detection results on both Waymo Open Dataset (OD) and Kirkland Dataset (Kirk). We show the pedestrian BEV AP results in this table. The AP distance breakdowns are provided by the official evaluation tool.

	Car - 3D AP (R11)			Pedestrian - 3D AP (R11)			Cyclist - 3D AP (R11)		
Method	Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard
PointPillars[8]	86.46	77.28	74.65	57.75	52.29	47.90	80.05	62.68	59.70
SPG + PointPillars	**87.98**	**78.54**	**77.32**	**59.91**	**54.58**	**50.34**	**81.58**	**65.70**	**62.28**
Improvement	*+1.52*	*+1.26*	*+2.67*	*+2.16*	*+2.29*	*+2.44*	*+1.53*	*+3.02*	*+2.58*
PVRCNN[12]	89.35	83.69	78.70	64.60	57.90	53.23	85.22	70.47	65.75
SPG + PVRCNN	**89.81**	**84.45**	**79.14**	**69.04**	**62.18**	**56.77**	**86.82**	**73.35**	**69.30**
Improvement	*+0.46*	*+0.76*	*+0.44*	*+4.44*	*+4.28*	*+3.54*	*+1.60*	*+2.88*	*+3.55*

Table 6: Result comparisons on the KITTI validation set. The results are evaluated by the Average Precision with 11 recall positions. The baseline detectors, PointPillars and PV-RCNN, are directly evaluated by using the checkpoints released by [12, 17].

Method	Car - BEV AP (R11)			Pedestrian - BEV AP (R11)			Cyclist - BEV AP (R11)		
	Easy	Mod.	Hard	Easy	Mod.	hard	Easy	Mod.	Hard
PointPillars[8]	89.65	87.17	84.37	61.63	56.27	52.60	82.27	66.25	62.64
SPG + PointPillars	**90.07**	**88.00**	**86.63**	**65.16**	**59.86**	**56.07**	**86.02**	**71.93**	**65.69**
Improvement	*+0.42*	*+0.83*	*+2.26*	*+3.53*	*+3.59*	*+3.47*	*+3.75*	*+5.68*	*+3.05*
PVRCNN[12]	90.09	87.90	87.41	67.01	61.38	56.10	86.79	73.55	69.69
SPG + PVRCNN	**90.41**	**88.49**	**87.74**	**71.19**	**64.37**	**59.88**	**92.54**	**74.43**	**70.99**
Improvement	*+0.32*	*+0.59*	*+0.33*	*+4.18*	*+2.99*	*+3.78*	*+5.75*	*+0.88*	*+1.30*

Table 7: Result comparisons on the KITTI validation set. The results are evaluated by the Average Precision with 11 recall positions. The baseline detectors, PointPillars and PV-RCNN, are directly evaluated by using the checkpoints released by [12, 17].

Method	Car - 3D AP (R40)			Pedestrian - 3D AP (R40)			Cyclist - 3d AP (R40)		
	Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard
PointPillars[8]	87.75	78.39	75.18	57.30	51.41	46.87	81.57	62.94	58.98
SPG+PointPillars	**89.77**	**81.36**	**78.85**	**59.65**	**53.55**	**49.24**	**83.27**	**66.11**	**61.99**
Improvement	*+2.02*	*+2.97*	*+3.67*	*+2.35*	*+2.14*	*+2.37*	*+1.70*	*+3.17*	*+3.01*
PVRCNN[12]	92.10	84.36	82.48	64.26	56.67	51.91	88.88	71.95	66.78
SPG+PVRCNN	**92.53**	**85.31**	**82.82**	**69.66**	**61.80**	**56.39**	**91.75**	**74.35**	**69.49**
Improvement	*+0.43*	*+0.95*	*+0.34*	*+5.40*	*+5.13*	*+4.48*	*+2.87*	*+2.40*	*+2.71*

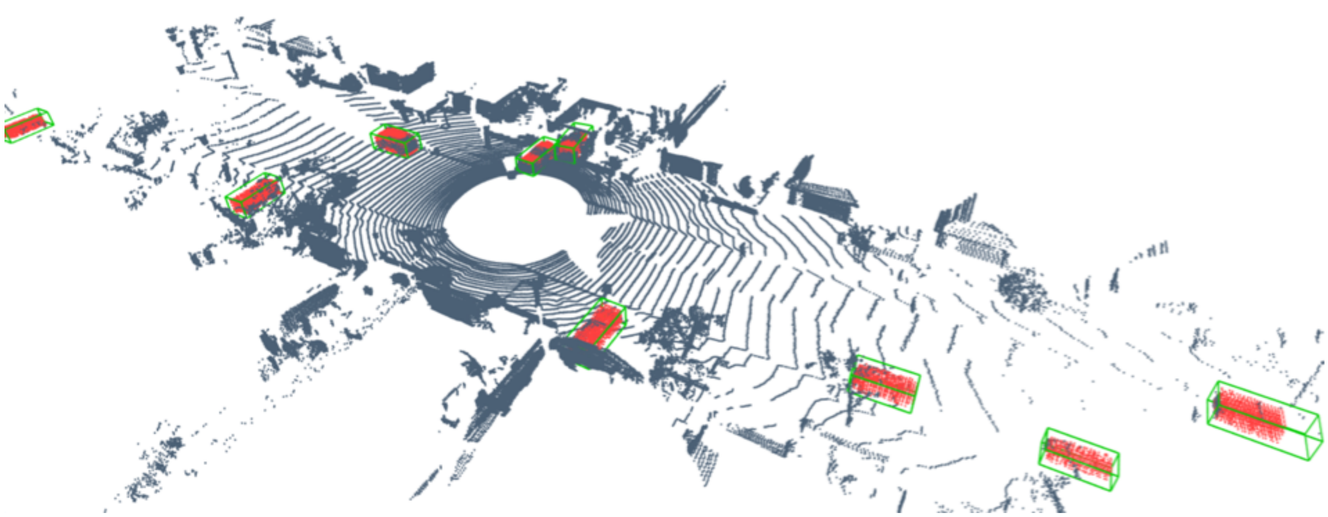
Table 8: Result comparisons on the KITTI validation set. The results are evaluated by the Average Precision with 40 recall positions. The baseline detectors, PointPillars and PV-RCNN, are directly evaluated by using the checkpoints released by [12, 17].

Method	Car - BEV AP (R40)			Pedestrian - BEV AP (R40)			Cyclist - BEV AP (R40)		
	Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard
PointPillars[8]	92.03	88.05	86.66	61.59	56.01	52.04	85.27	66.34	62.36
SPG + PointPillars	**94.38**	**89.92**	**87.97**	**65.38**	**59.48**	**55.32**	**90.29**	**71.43**	**66.96**
Improvement	*+2.35*	*+1.87*	*+1.31*	*+3.79*	*+3.47*	*+3.28*	*+5.02*	*+5.09*	*+4.60*
PVRCNN[12]	93.02	90.33	88.53	67.97	60.52	55.80	91.02	74.54	69.92
SPG + PVRCNN	**94.99**	**91.11**	**88.86**	**71.79**	**64.50**	**59.51**	**93.62**	**76.45**	**71.64**
Improvement	*+1.97*	*+0.78*	*+0.33*	*+3.82*	*+3.98*	*+3.71*	*+2.60*	*+1.91*	*+1.72*

Table 9: Result comparisons on the KITTI validation set. The results are evaluated by the Average Precision with 40 recall positions. The baseline detectors, PointPillars and PV-RCNN, are directly evaluated by using the checkpoints released by [12, 17].

Method	Reference	Modality	Car - 3D AP Easy	Mod.	Hard	Avg.
F-PointNet[11]	CVPR 2018	LIDAR & RGB	82.19	69.79	60.59	70.86
AVOD-FPN[7]	IROS 2018	LIDAR & RGB	83.07	71.76	65.73	73.52
F-ConvNet[18]	IROS 2019	LIDAR & RGB	87.36	76.39	66.69	76.81
UberATG-MMF[9]	CVPR 2019	LIDAR & RGB	88.40	77.43	70.22	78.68
EPNet[6]	ECCV 2020	LiDAR & RGB	89.81	79.28	74.59	81.23
CLOCs_PVCas[10]	IROS 2020	LiDAR & RGB	88.94	80.67	77.15	82.25
3D-CVF[25]	ECCV 2020	LiDAR & RGB	89.20	80.05	73.11	80.79
SECOND[20]	Sensors 2018	LiDAR	83.34	72.55	65.82	73.90
PointPillars[8]	CVPR 2019	LiDAR	82.58	74.31	68.99	75.30
PointRCNN[13]	CVPR 2019	LiDAR	86.96	76.50	71.39	77.77
3D IoU Loss[26]	3DV 2019	LiDAR	86.16	75.64	70.70	78.28
Fast Point R-CNNs[2]	ICCV 2019	LiDAR	85.29	77.40	70.24	77.64
STD[22]	ICCV 2019	LiDAR	87.95	79.71	75.09	80.91
SegVoxelNet[24]	ICRA 2020	LiDAR	86.04	76.13	70.76	77.64
SARPNET[23]	Neuro Computing 2019	LiDAR	85.63	76.64	71.31	77.86
HRI-VoxelFPN[24]	Sensor 2020	LiDAR	85.63	76.70	69.44	77.26
HotSpotNet[1]	ECCV 2020	LiDAR	87.60	78.31	73.34	79.75
PartA^2[14]	TPAMI 2020	LiDAR	87.81	78.49	73.51	79.94
SERCNN[27]	CVPR 2020	LiDAR	87,74	78.96	74.14	51.03
Point-GNN[15]	CVPR 2020	LiDAR	88.33	79.47	72.29	80.03
3DSSD[21]	CVPR 2020	LiDAR	88.36	79.57	74.55	80.83
SA-SSD[5]	CVPR 2020	LiDAR	88.75	79.79	74.16	80.90
CIA-SSD[19]	AAAI 2021	LiDAR	89.59	80.28	72.87	80.91
Asso-3Ddet[4]	CVPR 2020	LiDAR	85.99	77.40	70.53	77.97
Voxel R-CNN[3]	AAAI 2021	LiDAR	**90.90**	81.62	77.06	83.19
PV-RCNN[12]	CVPR 2020	LiDAR	90.25	81.43	76.82	82.83
SPG+PV-RCNN (Ours)	-	LiDAR	90.49	**82.13**	**78.88**	**83.83**

Table 10: Car detection result comparisons on the KITTI test set. The results are evaluated by the Average Precision with 40 recall positions on the KITTI benchmark website. We compare with the leader board front runner detectors that are associated with conferences or journals released before our submission. The Avg. AP is calculated by averaging over the APs of Easy, Mod. and Hard. difficulty levels.



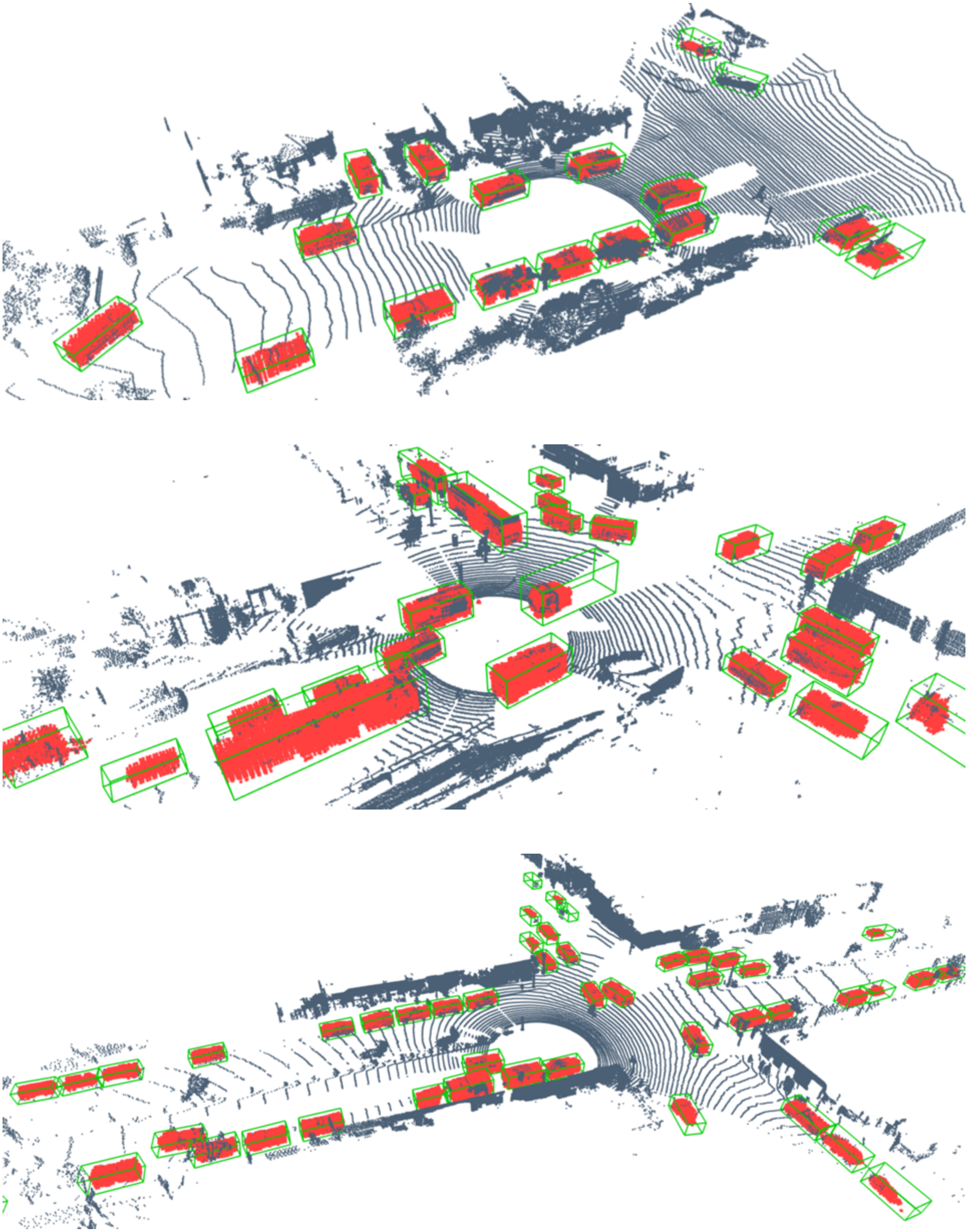


Figure 3: More visualization of generated semantic points. The grey points are original raw points. The red points are the generated semantic points. The green boxes are the predicted bounding boxes.

References

[1] Qi Chen, Lin Sun, Zhixin Wang, Kui Jia, and Alan Yuille. Object as hotspots: An anchor-free 3d object detection approach via firing of hotspots. In *European Conference on Computer Vision*, pages 68–84. Springer, 2020. 6

[2] Yilun Chen, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Fast point r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9775–9784, 2019. 6

[3] Jiajun Deng, Shaoshuai Shi, Peiwei Li, Wengang Zhou, Yanyong Zhang, and Houqiang Li. Voxel r-cnn: Towards high performance voxel-based 3d object detection. *arXiv:2012.15712*, 2020. 6

[4] Liang Du, Xiaoqing Ye, Xiao Tan, Jianfeng Feng, Zhenbo Xu, Errui Ding, and Shilei Wen. Associate-3ddet: Perceptual-to-conceptual association for 3d point cloud object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13329–13338, 2020. 6

[5] Chenhang He, Hui Zeng, Jianqiang Huang, Xian-Sheng Hua, and Lei Zhang. Structure aware single-stage 3d object detection from point cloud. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 6

[6] Tengteng Huang, Zhe Liu, Xiwu Chen, and Xiang Bai. Epnet: Enhancing point features with image semantics for 3d object detection. In *European Conference on Computer Vision*, pages 35–52. Springer, 2020. 6

[7] Jason Ku, Melissa Mozifian, Jungwook Lee, Ali Harakeh, and Steven L Waslander. Joint 3d proposal generation and object detection from view aggregation. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–8. IEEE, 2018. 6

[8] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12697–12705, 2019. 4, 5, 6

[9] Ming Liang, Bin Yang, Yun Chen, Rui Hu, and Raquel Urtasun. Multi-task multi-sensor fusion for 3d object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7345–7353, 2019. 6

[10] Su Pang, Daniel Morris, and Hayder Radha. Clocs: Camera-lidar object candidates fusion for 3d object detection. *arXiv preprint arXiv:2009.00784*, 2020. 6

[11] Charles R Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J Guibas. Frustum pointnets for 3d object detection from rgb-d data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 918–927, 2018. 6

[12] 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/CVF Conference on Computer Vision and Pattern Recognition*, pages 10529–10538, 2020. 4, 5, 6

[13] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointrcnn: 3d object proposal generation and detection from point cloud. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–779, 2019. 6

[14] S. Shi, Z. Wang, J. Shi, X. Wang, and H. Li. From points to parts: 3d object detection from point cloud with part-aware and part-aggregation network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020. 6

[15] Weijing Shi and Raj Rajkumar. Point-gnn: Graph neural network for 3d object detection in a point cloud. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1711–1719, 2020. 6

[16] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang,

Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset, 2019. 2

[17] OpenPCDet Development Team. Openpcdet: An open-source toolbox for 3d object detection from point clouds. https://github.com/open-mmlab/OpenPCDet, 2020. 4, 5

[18] Zhixin Wang and Kui Jia. Frustum convnet: Sliding frustums to aggregate local point-wise features for amodal 3d object detection. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1742–1749. IEEE, 2019. 6

[19] Sijin Chen Li Jiang Chi-Wing Fu Wu Zheng, Weiliang Tang. Cia-ssd: Confident iou-aware single-stage object detector from point cloud. In *AAAI*, 2021. 6

[20] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018. 6

[21] Zetong Yang, Yanan Sun, Shu Liu, and Jiaya Jia. 3dssd: Point-based 3d single stage object detector. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11040–11048, 2020. 6

[22] Zetong Yang, Yanan Sun, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Std: Sparse-to-dense 3d object detector for point cloud. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1951–1960, 2019. 6

[23] Yangyang Ye, Houjin Chen, Chi Zhang, Xiaoli Hao, and Zhaoxiang Zhang. Sarpnet: Shape attention regional proposal network for lidar-based 3d object detection. *Neurocomputing*, 379:53–63, 2020. 6

[24] Hongwei Yi, Shaoshuai Shi, Mingyu Ding, Jiankai Sun, Kui Xu, Hui Zhou, Zhe Wang, Sheng Li, and Guoping Wang. Segvoxelnet: Exploring semantic context and depth-aware features for 3d vehicle detection from point cloud. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2274–2280. IEEE, 2020. 6

[25] Jin Hyeok Yoo, Yecheol Kim, Ji Song Kim, and Jun Won Choi. 3d-cvf: Generating joint camera and lidar features using cross-view spatial feature fusion for 3d object detection. *arXiv preprint arXiv:2004.12636*, 3, 2020. 6

[26] Dingfu Zhou, Jin Fang, Xibin Song, Chenye Guan, Junbo Yin, Yuchao Dai, and Ruigang Yang. Iou loss for 2d/3d object detection. In *2019 International Conference on 3D Vision (3DV)*, pages 85–94. IEEE, 2019. 6

[27] Dingfu Zhou, Jin Fang, Xibin Song, Liu Liu, Junbo Yin, Yuchao Dai, Hongdong Li, and Ruigang Yang. Joint 3d instance segmentation and object detection for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 6