GACE: Geometry Aware Confidence Enhancement for Black-box 3D Object Detectors on LiDAR-Data

David Schinagl[1,2] Georg Krispel[1] Christian Fruhwirth-Reisinger[1,2] Horst Possegger[1] Horst Bischof[1,2]
{david.schinagl,georg.krispel,reisinger,possegger,bischof}@icg.tugraz.at
[1] Graz University of Technology [2] Christian Doppler Laboratory for Embedded Machine Learning

This supplemental material provides additional insights and quantitative results, as well as potential limitations of our confidence enhancement approach.

1. Geometric Dependencies

Figure 1 shows an additional example of how the detection performance depends on geometric properties. Here, we show how well a vanilla SECOND [6] model detects pedestrians on the Waymo dataset w.r.t. low-level statistics of the points within the detection bounding boxes. Besides the expected influence of the number of points (top), we found that the detection accuracy also strongly depends on the distribution of points along the z-axis of the bounding box, *i.e.* along the height of the suspected pedestrian. On the one hand, it can be seen that pedestrians are only detected reliably if points are present in the upper quarter of the bounding box, *i.e.* in the area of the head-shoulder silhouette (middle). On the other hand, the precision increases the better the points are distributed across the entire z-axis (bottom). The incorporation of these simple properties therefore enables an improved assessment of the confidence score.

2. Waymo Test Set Results

To demonstrate the reliability of GACE, we also evaluate our approach on the Waymo Open Dataset [5] test set using the official evaluation server. Note that for Waymo, detectors typically perform better on the test set compared to the larger validation set. Despite the already improved performance of the base detectors, GACE enables comparable and consistent improvements across the different detectors and classes, see Table 1.

3. Ablation Study Instance-Specific Properties

In Table 3, we show the complete list of all combinations of instance-specific properties and the impact on the overall performance. We analyze the contribution of each feature group within the instance-specific properties, namely *box properties* ($\mathbf{b}$, $\|\mathbf{c}\|$), *number of points* ($|\mathcal{X}_\mathbf{b}|$), *viewing angle* (α), and *point statistics* ($\mathcal{X}_\mathbf{b}^{\text{mean}}, \mathcal{X}_\mathbf{b}^{\text{std}}, \mathcal{X}_\mathbf{b}^{\text{min}}, \mathcal{X}_\mathbf{b}^{\text{max}}$).

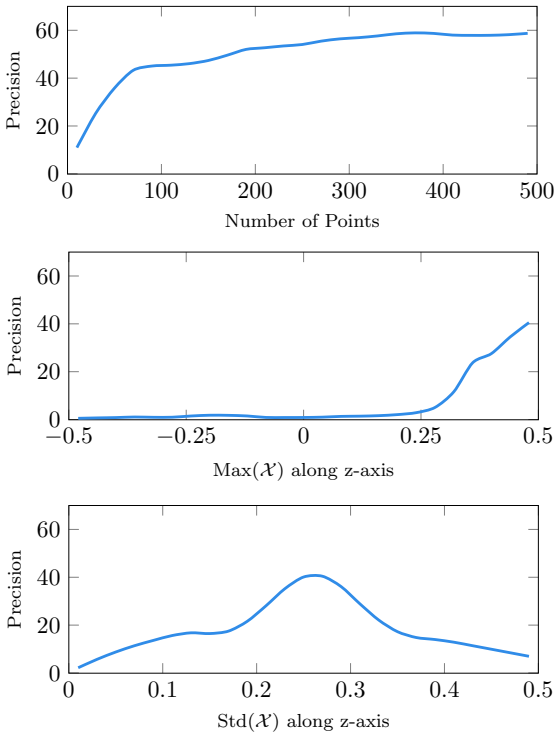

Figure 1. Precision of a SECOND [6] model on the Waymo Open Datase [5] for the pedestrian class on different low-level point statistics. Note that the statistics refer to zero-centered boxes on all axes, normalized to unit length.

Method	LEVEL_1		LEVEL_2	
	mAP	mAPH	mAP	mAPH
PART-A^2	71.66	67.99	65.84	62.47
+GACE	73.42$^{+1.76}$	69.65$^{+1.66}$	67.55$^{+1.71}$	64.08$^{+1.61}$
Centerpoint	74.39	71.67	68.92	66.38
+GACE	76.03$^{+1.64}$	73.39$^{+1.72}$	70.45$^{+1.53}$	67.97$^{+1.59}$
PV-RCNN	73.21	68.84	67.42	63.39
+GACE	74.61$^{+1.40}$	70.14$^{+1.30}$	68.79$^{+1.37}$	64.65$^{+1.26}$

Table 1. Performance gains using GACE on the Waymo Open Dataset [5] test set over all classes.

Method	Vehicle (LEVEL_2)			Pedestrian (LEVEL_2)			Cyclist (LEVEL_2)		
	0-30m	30-50m	50m-Inf	0-30m	30-50m	50m-Inf	0-30m	30-50m	50m-Inf
	mAP / mAPH	mAP / mAPH	mAP / mAPH	mAP / mAPH	mAP / mAPH	mAP / mAPH	mAP / mAPH	mAP / mAPH	mAP / mAPH
PointPillars [1]	88.10 / 87.58	61.88 / 61.18	34.78 / 34.01	67.80 / 49.03	58.32 / 40.98	41.77 / 27.60	69.40 / 66.17	45.71 / 43.43	35.00 / 31.39
+ GACE (Ours)	88.46 / 87.96	63.03 / 62.31	36.12 / 35.31	72.37 / 53.01	64.15 / 45.81	49.12 / 33.25	74.12 / 71.08	53.34 / 50.92	43.40 / 39.35
Improvement	**+0.36 / +0.38**	**+1.15 / +1.13**	**+1.34 / +1.30**	**+4.57 / +3.98**	**+5.83 / +4.83**	**+7.35 / +5.65**	**+4.72 / +4.91**	**+7.63 / +7.49**	**+8.40 / +7.96**
SECOND [6]	88.50 / 87.98	62.14 / 61.49	33.93 / 33.20	66.64 / 57.51	58.20 / 47.88	41.36 / 31.23	71.53 / 70.14	49.31 / 47.90	37.88 / 35.74
+ GACE (Ours)	88.81 / 88.28	62.98 / 62.32	35.01 / 34.23	71.23 / 63.27	62.65 / 52.81	45.64 / 35.58	75.90 / 74.61	54.65 / 53.26	44.41 / 42.35
Improvement	**+0.31 / +0.30**	**+0.84 / +0.83**	**+1.08 / +1.03**	**+4.59 / +5.76**	**+4.45 / +4.93**	**+4.28 / +4.35**	**+4.37 / +4.47**	**+5.34 / +5.36**	**+6.53 / +6.61**
Part-A^2 [4]	89.53 / 89.09	64.90 / 64.34	37.11 / 36.42	71.03 / 63.65	61.27 / 52.00	41.92 / 33.33	79.24 / 78.04	59.13 / 57.68	41.37 / 39.58
+ GACE (Ours)	89.87 / 89.43	65.72 / 65.12	38.10 / 37.30	72.50 / 64.91	63.65 / 53.90	46.25 / 36.71	82.49 / 81.22	64.52 / 62.75	50.82 / 48.80
Improvement	**+0.34 / +0.34**	**+0.82 / +0.78**	**+0.99 / +0.88**	**+1.47 / +1.26**	**+2.38 / +1.90**	**+4.33 / +3.38**	**+3.25 / +3.18**	**+5.39 / +5.07**	**+9.45 / +9.22**
Centerpoint [7]	88.88 / 88.39	65.26 / 64.72	37.05 / 36.50	74.44 / 69.48	66.95 / 60.42	51.89 / 44.21	80.56 / 79.50	64.15 / 62.85	50.40 / 49.09
+ GACE (Ours)	89.88 / 89.43	66.73 / 66.20	38.61 / 38.05	77.80 / 73.17	70.86 / 64.22	56.33 / 48.14	81.13 / 80.10	65.09 / 63.86	55.66 / 54.22
Improvement	**+1.00 / +1.04**	**+1.47 / +1.48**	**+1.56 / +1.55**	**+3.36 / +3.69**	**+3.91 / +3.80**	**+4.44 / +3.93**	**+0.57 / +0.60**	**+0.94 / +1.01**	**+5.26 / +5.13**
PV-RCNN [2]	89.87 / 89.40	66.39 / 65.74	39.73 / 38.88	71.84 / 63.21	63.47 / 52.72	46.60 / 35.87	77.58 / 76.12	57.86 / 55.85	41.20 / 38.56
+ GACE (Ours)	90.04 / 89.56	67.07 / 66.35	40.42 / 39.48	73.11 / 64.35	65.16 / 54.13	49.87 / 38.31	80.59 / 79.07	62.06 / 59.84	50.74 / 47.91
Improvement	**+0.17 / +0.16**	**+0.68 / +0.61**	**+0.69 / +0.60**	**+1.27 / +1.14**	**+1.69 / +1.41**	**+3.27 / +2.44**	**+3.01 / +2.95**	**+4.20 / +3.99**	**+9.54 / +9.35**
PV-RCNN++ [3]	90.99 / 90.58	69.74 / 69.24	43.18 / 42.56	78.71 / 73.77	70.48 / 63.60	54.87 / 46.64	79.54 / 78.58	63.35 / 62.28	48.02 / 46.57
+ GACE (Ours)	91.15 / 90.74	69.86 / 69.35	43.36 / 42.69	79.22 / 74.14	71.34 / 64.31	56.75 / 48.10	80.39 / 79.39	64.88 / 63.83	55.33 / 53.77
Improvement	**+0.16 / +0.16**	**+0.12 / +0.11**	**+0.18 / +0.13**	**+0.51 / +0.37**	**+0.86 / +0.71**	**+1.88 / +1.46**	**+0.85 / +0.81**	**+1.53 / +1.55**	**+7.31 / +7.20**

Table 2. Performance comparison on the Waymo Open Dataset [5] validation set across distance ranges for different baseline methods with and without our confidence enhancement module for difficulty LEVEL_2 and for all classes.

Box Properties	# Points	Viewing Angle	Point statistics	LEVEL_2 mAPH	Improvement
✓	✓	✓	✓	58.78	**+3.66**
✓	✓	✓		58.00	**+2.88**
✓	✓		✓	58.55	**+3.43**
✓	✓			57.33	**+2.21**
✓		✓	✓	58.46	**+3.34**
✓		✓		57.21	**+2.09**
✓			✓	58.49	**+3.37**
✓				56.98	**+1.86**
	✓	✓	✓	58.26	**+3.14**
	✓	✓		56.04	**+0.92**
	✓		✓	58.23	**+3.11**
	✓			55.96	**+0.84**
		✓	✓	58.30	**+3.18**
		✓		55.86	**+0.74**
			✓	58.15	**+3.03**

Table 3. Impact of instance-specific feature groups (SECOND base model / Waymo LEVEL_2 mAPH overall).

4. Range-based Evaluation

Table 2 presents the detailed results of the range-based evaluation on the Waymo Open Dataset [5] for all individual classes. Adjusting the confidence values using GACE leads to an improvement in performance for all classes and across all distance ranges. The detailed results for each class also show that our confidence enhancement approach achieves the highest performance gain at long ranges. These more stable detections are particularly important for safety, as they give autonomous vehicles more time to react, avoid collisions and plan a safe and efficient route. Nevertheless, as can be seen in the example of the pedestrian class for SECOND [6] and Centerpoint [7], performance can also be significantly improved for close-range detections.

5. Geometric Dependencies with GACE

In Figure 2 we show the impacts of GACE on the geometric dependencies of the precision of a SECOND model on the Waymo dataset using the vehicle length (left column) and the viewing angle (right column) as exmaples. Therefore, we thresholded the detections at different confidence thresholds of 0.3, 0.5, and 0.7 (rows) and show the precision before (blue) and after (orange) applying our confidence enhancement. GACE can especially improve the precision for challenging samples (such as trucks, *i.e.* long vehicles), which are typically underrepresented (and thus, suffer from low confidence scores). Thus, this is only partially reflected by the evaluation metrics (*e.g.* improvement for *vehicles* with SECOND is +0.56 LEVEL_2 APH).

6. Precision-Recall Plots

Figure 4 shows the remaining precision-recall plots of the evaluation on the Waymo validation set for each class at LEVEL_2 APH difficulty level. The plots confirm the performance improvements especially for the spatially smaller and therefore more difficult classes of pedestrians and cyclists. In addition, it is noticeable that the increase in precision is particularly high in the regions with a high recall, *i.e.* for detections with a low confidence score.

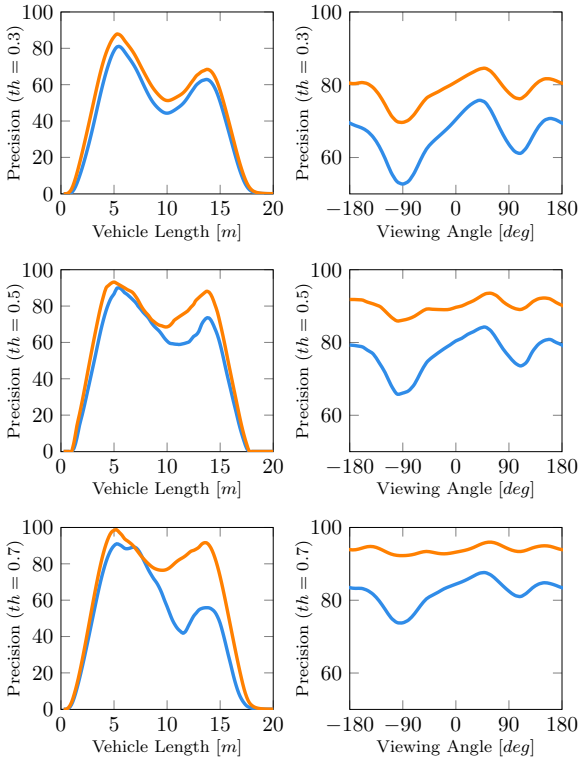

Figure 2. Precision of a SECOND model on the Waymo Open Dataset for the vehicle class as a function of the object length (left column) and of the viewing angle (right column), before (blue) and after (orange) applying GACE for different confidence score thresholds (rows).

7. Qualitative Results

To illustrate the behavior of GACE using qualitative examples, Figure 3 shows the two instances of each Waymo class in the validation set where GACE increased the confidence value the most compared to the baseline SECOND model. Conversely, Figure 4 shows the two instances per class where the confidence score decreased the most.

References

[1] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. PointPillars: Fast Encoders for Object Detection from Point Clouds. In *Proc. CVPR*, 2019. 2, 4

[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 *Proc. CVPR*, 2020. 2, 4

[3] Shaoshuai Shi, Li Jiang, Jiajun Deng, Zhe Wang, Chaoxu Guo, Jianping Shi, Xiaogang Wang, and Hongsheng Li. PV-RCNN++: Point-Voxel Feature Set Abstraction With Local Vector Representation for 3D Object Detection. *IJCV*, pages 1–21, 2022. 2, 4

[4] Shaoshuai Shi, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. From Points to Parts: 3D Object Detection from Point Cloud with Part-aware and Part-aggregation Network. *PAMI*, pages 2647–2664, 2021. 2, 4

[5] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In *Proc. CVPR*, 2020. 1, 2, 4

[6] Yan Yan, Yuxing Mao, and Bo Li. SECOND: Sparsely Embedded Convolutional Detection. *Sensors*, 18(10):3337, 2018. 1, 2, 4

[7] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Center-based 3D Object Detection and Tracking. In *Proc. CVPR*, 2021. 2, 4

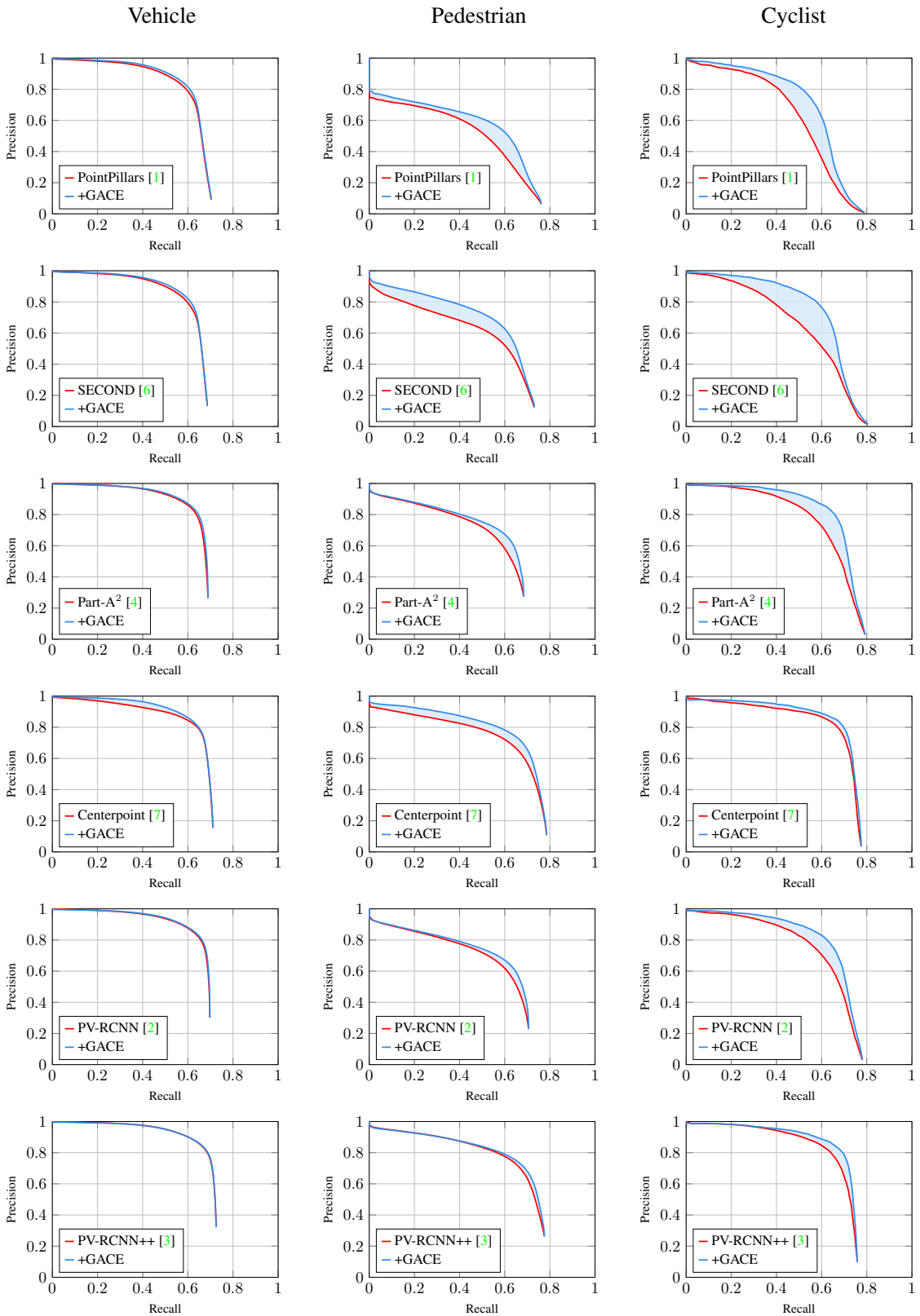

Figure 4. Precision-recall plots for all base models and all classes on the Waymo Open Dataset [5] validation set for LEVEL_2 APH. GACE allows us to better separate true and false positive hypotheses, which leads to more stable detection results, especially for the vulnerable road users.

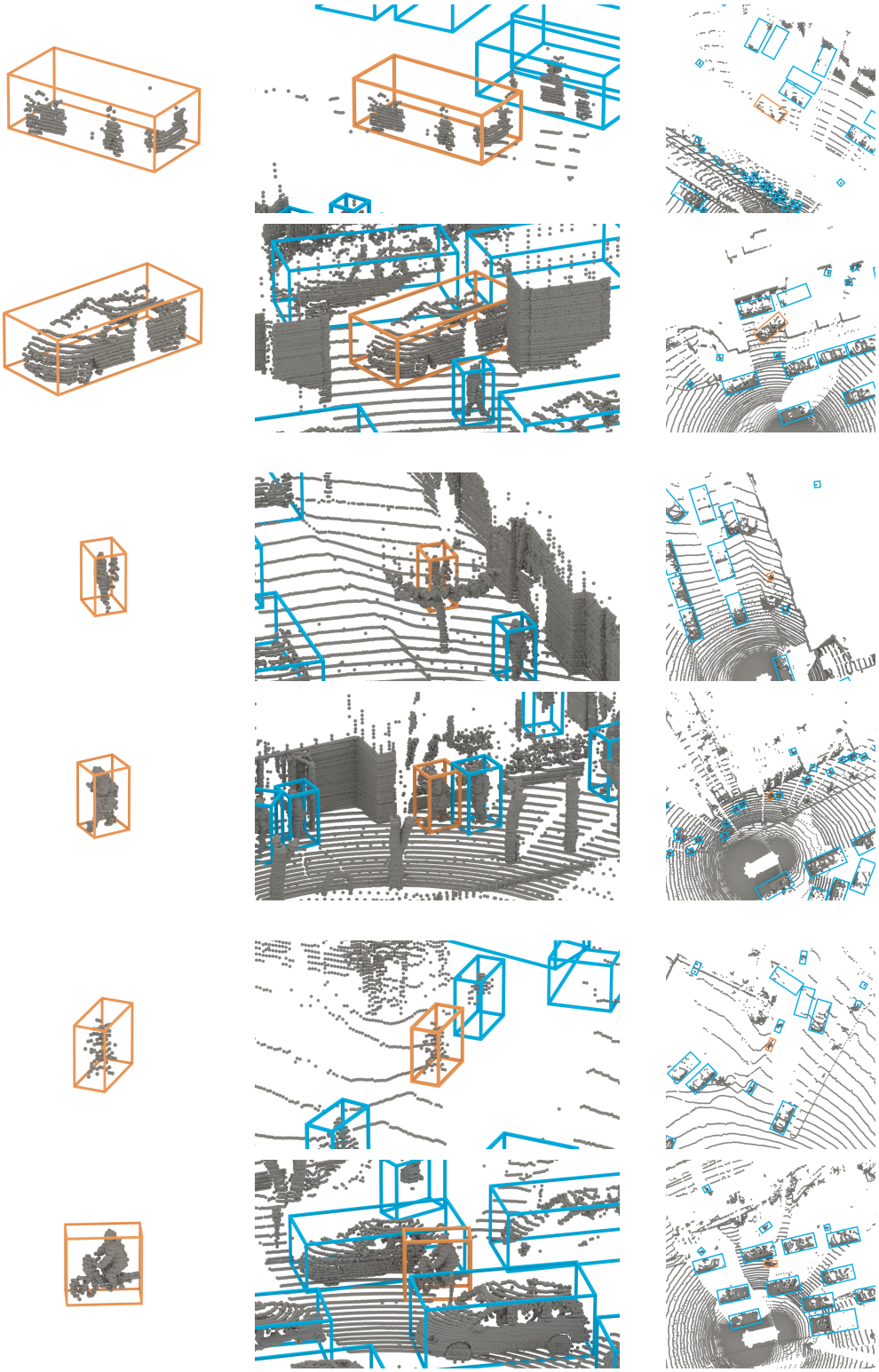

Figure 3. Qualitative Examples (Highest Score Increase): For each Waymo class, we show the two instances in the validation set where GACE increased the confidence value the most compared to the baseline SECOND model (vehicle at the top, pedestrian in the middle and cyclist at the bottom rows). Best viewed on screen.

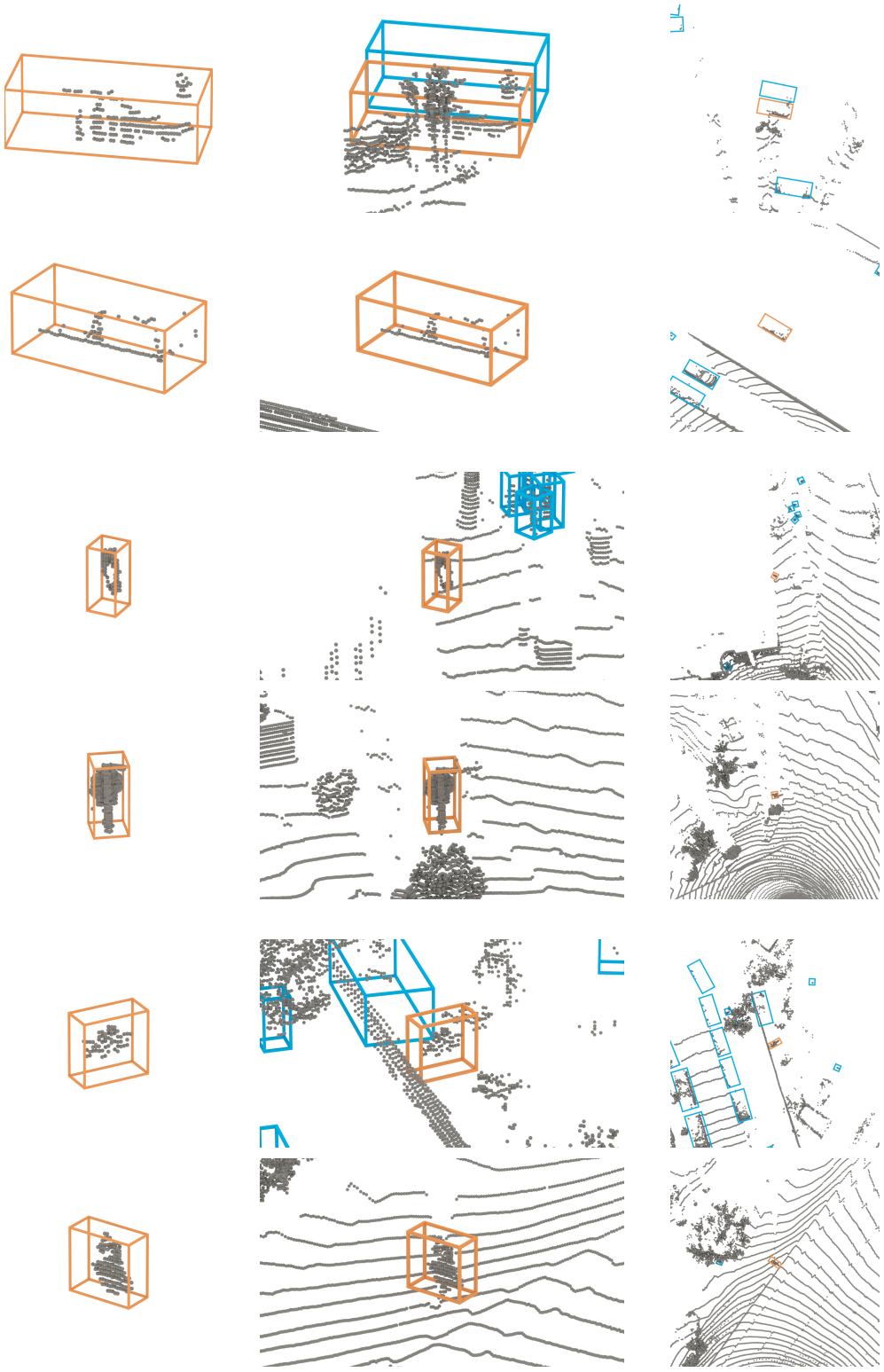

Figure 4. Qualitative Examples (Highest Score Decrease): For each Waymo class, we show the two instances in the validation set where GACE decreased the confidence value the most compared to the baseline SECOND model (vehicle at the top, pedestrian in the middle and cyclist at the bottom rows). Best viewed on screen.