Supplementary Material for SAILOR: Scaling Anchors via Insights into Latent Object Representation

Dušan Malić[1,2] Christian Fruhwirth-Reisinger[1,2] Horst Possegger[1] Horst Bischof[1,2]

[1]Institute of Computer Graphics and Vision, Graz University of Technology
[2]Christian Doppler Laboratory for Embedded Machine Learning

{dusan.malic, christian.reisinger, possegger, bischof}@icg.tugraz.at

In the following, we provide additional information for better clarification of our method as well as for easier reproducibility, which did not fit into the main manuscript due to the page limit. We first introduce our experimental setup (Section 1) followed by additional extensive experiments (Section 2). In Section 3, we present absolute anchor values obtained with SAILOR. To reproduce our weakly-supervised experiments, we present all required information in Section 4.

1. Implementation Details

We implement our approach as an OpenPCDet [2] extension. All datasets share the same configuration, where we limit lateral and longitudinal axes to $[-75.2, 75.2]$ and height to $[-2, 4]$ meters. We use voxel sizes of $[0.1, 0.1, 0.15]m$ for the three axis. Following [10], we shift KITTI, Lyft and nuScenes point clouds in the height axis by 1.6, 1.7 and 1.8 meters, respectively. We use four point features: the three-dimensional Euclidean position and the reflectivity. Since Lyft does not provide reflectivity values, we omit them in the respective experiments.

In cases when KITTI is the source model, we manually confine target models to the camera Field of View (FoV). Similarly, when either Waymo, nuScenes or Lyft is the source and KITTI is the target domain, we perform our calibration on 360°. The evaluation is still performed on the FoV data since this is the only region where KITTI is labeled. We do not accumulate LiDAR sweeps for the nuScenes and Lyft data, but instead use only a single sweep.

For the empirical evaluation, we opted for the Part-A^2 [5] model, even though our method is model-agnostic. The selected detector provides good semantics and differentiates the pooled features in the height axis due to the volumetric feature map. A detector that utilizes a Birds Eye View (BEV) feature map, *e.g.* SECOND [9], does not differentiate anchor height in feature space. Our method can not circumvent this lack of discrimination, thus would not be able to calibrate the anchor height. In this case, however, calibrating length and width is still possible. We train the model per source configuration and always evaluate the *last checkpoint* in a true unsupervised fashion.

During our anchor calibration, we use at most 2^{15} source features extracted using the predictions thresholded at confidence 0.5, 0.3, and 0.3 for the vehicle, pedestrian, and bicycle class, respectively. Moreover, the number of GMM parameters is set for each class separately, $K = [32, 8, 8]$. In order to balance speed and accuracy, we extract 1024 target features at each iteration. The individual linear search assesses the best overall fitness in a range of $\pm 30\%$ of the original anchor size with step size equal to the voxel size. Finally, DE jointly optimizes the output of the individual optimization. We define search bounds as $\pm$ voxel size and the stopping criterion is either number of iterations (1000) or fitness convergence (relative improvement ≤ 0.1).

2. Experiments

The experiments from the main manuscript are focused on different classes. However, different datasets offer additional evaluation protocols that could potentially provide additional insights. Therefore, in Section 2.1, we provide evaluation results for different difficulties (KITTI [3]) and different ranges (Waymo [6]) on the three main classes. Additionally, in Section 2.2, we provide further empirical reassurance for the correctness of our approach.

2.1. Extensive Evaluation

In Table 1, we present our evaluation when KITTI [3] is the target dataset. We report AP$_{BEV}$ / AP$_{3D}$ for the classes car, pedestrian and cyclist, and the three detection difficulties easy, moderate and hard. The source datasets are Waymo [6], nuScenes [1] and Lyft [4]. We compare our approach to Source-only Anchors (SA), Statistical Normalization (SN) [8], Output Transformation (OT) [8], Random Object Scaling (ROS) [10] and Target Anchors (TA).

The evaluation correlates with our findings in Table 1 of the main manuscript. In case where the average object size is different, *i.e*. for all three datasets the class car (Figure 1 of the main manuscript), we report substantial gains over all difficulties. We even surpass weakly-supervised approaches in the Waymo and Lyft case. For the classes pedestrian and cyclist, where the average object size does not vary much, the behavior depends on the source dataset. If a source model is powerful enough, *e.g*. Waymo $\rightarrow$ KITTI, we report slight improvements. In nuScenes $\rightarrow$ KITTI, we report slight precision decrease for the pedestrian class. This comes mainly from the sparsity of the source data, where a pedestrian object may contain just a few points which leads to poor latent semantics. Lyft dataset does not provide reflectivity values for the LiDAR points. We found that this influences latent semantics of smaller objects (pedestrians and cyclists) thus making them harder to optimize with our method. This is then further reflected in the final performance.

In Table 2 we show extensive evaluation on the Waymo dataset. We show $L1$_AP / $L2$_AP for different ranges and the three main classes. Our observations remain consistent with the findings in Table 1 of the main manuscript. We show slight improvement when the anchor sizes differ (KITTI $\rightarrow$ Waymo) and no improvement nor accuracy drop when the anchors are similar (nuScenes $\rightarrow$ Waymo). Since we modify only anchors, our method does not discover new objects. Instead our method allows more precise regression/classification and thus the most improvement comes from the objects which are in the close proximity from the sensor, *i.e*. in the $0m$ - $30m$ range.

2.2. Additional Verification

To further verify the correctness of our method, we opt to inspect the latent feature representation. With a model trained on Waymo, we extract latent features from Waymo and KITTI test datasets. Moreover, we extract features of the same KITTI objects using the anchors optimized with our method. With t-SNE [7], we reduce the high-dimensional representations to two dimensions, shown in Figure 1. We observe the affinity of source and target features obtained with the optimized anchors. The target features closer to the source features reduce the model's extrapolation requirements and thus can provide significantly better predictions.

High-quality predictions inevitably lead to precision improvement. To show that the improvement also correlates with our optimization objective, we conduct the following experiment: using a model trained on Waymo, we vary the anchor length and, at each step, evaluate the model on KITTI data and compute the fitness score. We depict the findings in Figure 3 of the main manuscript, where we report a clear correlation. Interestingly, we found that the optimal anchors are not the actual anchors from the target domain, which matches our findings in Table 1. We make the same observation in the point cloud space as depicted in Figure 2.

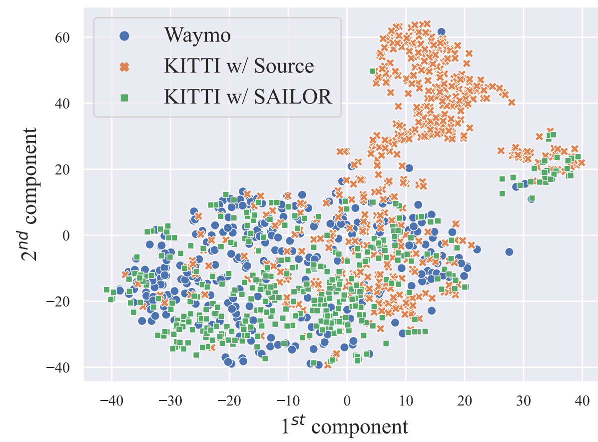


Figure 1: t-SNE [7] plot of latent features from Waymo and KITTI test dataset. We extract those features using a model trained on Waymo.

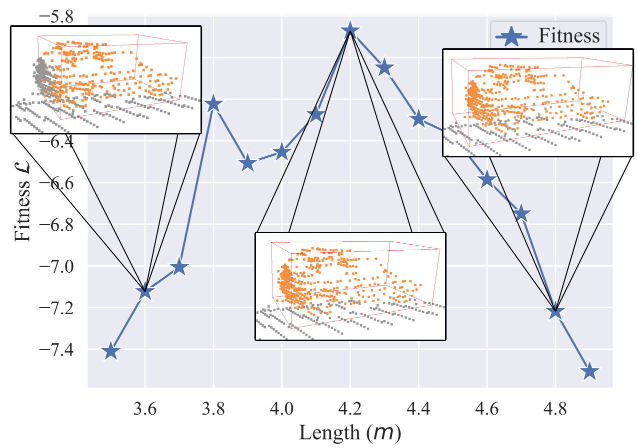


Figure 2: The influence of the length during our anchor calibration. The fitness is highest when the predicted bounding box captures the entirety of the object. Otherwise, either important cues are missing or we introduce background noise for the consecutive detection phase. Best viewed on screen.

3. Absolute Anchor Values

During our calibration, we keep all other anchor values fixed to the source configuration and vary only the length, width and height. In Table 3 we report our estimated anchor

Task	Method	Car			Pedestrian			Cyclist		
		Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
Waymo ↓ KITTI	SA	74.36 / 24.46	64.97 / 23.93	63.09 / 21.77	68.87 / 65.18	64.00 / 59.98	59.77 / 56.21	61.27 / 58.82	54.29 / 52.32	52.03 / 49.80
	SN†	80.91 / 27.03	68.77 / 24.11	66.06 / 22.44	**70.38 / 69.22**	**66.10 / 63.51**	**64.68 / 59.41**	64.06 / 60.54	54.63 / 52.83	52.65 / 50.97
	OT†	72.25 / 43.26	67.07 / 38.79	67.69 / 39.43	62.99 / 57.77	58.52 / 53.31	55.62 / 49.69	**73.38 / 72.08**	**63.14 / 61.82**	**60.65 / 57.27**
	ROS†	**83.41** / 45.94	74.57 / 43.00	74.71 / 41.76	68.88 / 66.52	63.01 / 62.39	59.59 / 57.84	59.70 / 57.17	51.85 / 51.09	49.22 / 48.49
	TA†	55.92 / 38.50	60.96 / 39.69	66.43 / 45.17	58.51 / 54.08	54.07 / 49.95	51.51 / 47.16	68.34 / 64.00	58.67 / 55.45	55.34 / 52.58
	Ours	82.53 / **61.10**	**79.90 / 58.02**	**80.13 / 58.88**	69.63 / 67.66	63.86 / 61.60	60.58 / 57.47	64.47 / 60.35	56.78 / 53.04	54.18 / 50.90
nuScenes ↓ KITTI	SA	83.00 / 34.21	68.48 / 26.37	67.30 / 25.38	16.62 / 11.95	14.90 / 10.37	12.60 / 9.43	37.80 / 31.68	27.97 / 22.90	26.80 / 22.00
	SN†	72.09 / 40.74	59.18 / 33.67	58.59 / 30.70	**37.79 / 34.12**	**34.92 / 31.67**	**32.53 / 29.80**	26.68 / 15.31	25.37 / 14.29	24.36 / 10.41
	OT†	83.54 / 48.02	72.29 / 36.42	69.32 / 35.00	22.51 / 19.34	20.71 / 18.15	19.90 / 17.40	28.58 / 12.43	24.04 / 8.84	23.54 / 8.74
	ROS†	**89.42** / 73.34	**75.97** / 56.28	**74.74** / 54.93	32.46 / 29.95	29.41 / 26.88	27.54 / 24.06	26.08 / 25.11	20.23 / 19.76	19.42 / 18.35
	TA†	82.98 / **68.29**	72.87 / **57.93**	72.11 / **56.48**	22.68 / 19.71	21.00 / 18.57	19.99 / 17.55	24.71 / 14.35	20.72 / 11.67	20.33 / 11.48
	Ours	83.71 / 67.72	72.76 / 55.10	71.31 / 53.52	5.56 / 4.64	5.53 / 4.64	5.40 / 4.65	**38.83 / 35.31**	**29.88 / 26.37**	**28.50 / 25.02**
Lyft ↓ KITTI	SA	**89.10** / 74.06	78.25 / 62.22	77.20 / 56.44	55.55 / 52.86	48.45 / 46.97	**46.30** / 42.74	60.74 / 59.20	53.74 / **51.54**	50.72 / **49.00**
	SN†	88.25 / 69.56	77.17 / 54.90	75.33 / 52.47	**56.21 / 53.59**	**48.61 / 47.32**	46.28 / **44.34**	61.41 / **59.22**	52.92 / 51.48	51.11 / **49.00**
	OT†	71.26 / 29.77	67.93 / 28.46	68.53 / 31.61	51.15 / 46.36	46.80 / 41.06	42.47 / 38.84	**62.36** / 49.50	**54.70** / 39.78	**51.89** / 38.09
	ROS†	86.02 / 62.77	72.47 / 48.95	67.11 / 43.79	54.92 / 52.50	48.38 / 46.44	45.27 / 42.15	62.13 / 58.74	53.12 / 49.98	49.64 / 46.48
	TA†	84.11 / 63.62	74.95 / 54.85	75.26 / 55.99	50.82 / 45.22	45.42 / 40.03	40.89 / 37.51	58.90 / 48.48	51.11 / 37.77	48.46 / 35.54
	Ours	88.55 / **83.87**	**84.00 / 73.85**	**82.05 / 71.06**	50.04 / 45.12	46.06 / 46.06	42.60 / 40.36	58.82 / 54.03	50.40 / 44.21	48.72 / 41.73

Table 1: Results of the adaptation from Waymo, nuScenes and Lyft to KITTI, for different classes and difficulties. We report AP_{BEV} / AP_{3D}. We compare our method to Source Anchors (SA), Statistical Normalization (SN), Output Transformation (OT), Random Object Scaling (ROS) and Target Anchors (TA). † denotes weakly-supervised methods.

Task	Method	Vehicle			Pedestrian			Cyclist		
		$0m$ - $30m$	$30m$ - $50m$	$50m - 75m$	$0m$ - $30m$	$30m$ - $50m$	$50m - 75m$	$0m$ - $30m$	$30m$ - $50m$	$50m - 75m$
KITTI ↓ Waymo	SA	5.41 / 5.38	4.05 / 3.54	0.59 / 0.42	11.23 / 10.10	9.83 / 8.39	1.06 / 0.74	16.82 / 16.70	6.65 / 6.27	0.35 / 0.33
	SN†	7.38 / 7.19	0.22 / 0.19	0.00 / 0.00	4.75 / 4.25	0.07 / 0.06	0.01 / 0.01	3.81 / 3.78	0.00 / 0.00	0.00 / 0.00
	OT†	16.48 / 16.10	8.00 / 7.00	1.21 / 0.87	**18.18 / 16.35**	**14.11 / 12.05**	**1.48 / 1.03**	**24.56 / 24.38**	**8.01 / 7.54**	0.43 / 0.40
	ROS†	7.31 / 7.14	1.65 / 1.44	0.08 / 0.06	7.31 / 7.14	1.65 / 1.44	0.08 / 0.06	5.29 / 5.25	0.06 / 0.05	0.00 / 0.00
	TA†	**19.73 / 19.28**	**10.71 / 9.37**	1.85 / 1.34	15.45 / 13.88	9.87 / 8.41	0.75 / 0.52	21.39 / 21.23	7.55 / 7.11	**0.46 / 0.43**
	Ours	12.12 / 11.83	9.18 / 8.04	**2.02 / 1.46**	11.36 / 10.22	10.19 / 8.69	1.13 / 0.79	16.20 / 16.08	6.37 / 6.00	0.40 / 0.38
nuScenes ↓ Waymo	SA	59.56 / 58.39	15.75 / 13.82	2.52 / 1.82	5.24 / 4.69	1.61 / 1.37	0.48 / 0.33	8.68 / 8.62	0.20 / 0.19	**0.04 / 0.04**
	SN	**66.82 / 65.52**	**22.93 / 20.15**	2.59 / 1.87	3.74 / 3.34	2.15 / 1.82	0.24 / 0.17	10.11 / 10.04	0.41 / 0.38	0.01 / 0.01
	OT†	63.51 / 62.23	16.80 / 14.75	2.53 / 1.83	13.63 / 12.22	**5.93 / 5.07**	**2.12 / 1.49**	4.30 / 4.27	0.12 / 0.12	0.03 / 0.02
	ROS†	58.56 / 57.41	22.05 / 19.40	**5.14 / 3.73**	**16.14 / 14.48**	1.30 / 1.11	0.19 / 0.13	**15.33 / 15.22**	**0.67 / 0.63**	0.03 / 0.03
	TA†	62.89 / 61.63	17.71 / 15.55	2.83 / 2.05	11.25 / 10.09	4.33 / 3.70	1.80 / 1.27	1.77 / 1.76	0.03 / 0.03	0.01 / 0.01
	Ours	60.55 / 59.32	14.02 / 12.30	1.87 / 1.35	3.83 / 3.43	1.02 / 0.87	0.27 / 0.19	8.01 / 7.95	0.19 / 0.18	0.02 / 0.02

Table 2: Results of the adaptation from KITTI and nuScenes to Waymo at different distances from the sensor. We show $L1$_AP / $L2$_AP for the three main classes. We compare our method to Source Anchors (SA), Statistical Normalization (SN), Output Transformation (OT), Random Object Scaling (ROS) and Target Anchors (TA).

sizes for the class Car / Vehicle on KITTI [3], Waymo [6] and nuScenes [1]. Additionally, we provide the source and target ground truth sizes for comparison. When our source and target data are from the same domain (the diagonal in the table), we manage to perfectly restore the sizes without any prior. We report the largest deviation in the case nuScenes ↔ nuScenes (again due to the LiDAR sparsity), however, this does not have a significant impact in the overall performance (see Table 1 in the main manuscript). In cases where source and target domain differ, we do not always match the target anchor sizes. However, as denoted in the main manuscript, this is not always desired as the actual target anchors usually degrade the precision of the detection model. Finally, we observe that our calibration method is not commutative, *e.g.* we obtain different anchor values for KITTI → Waymo and Waymo → KITTI. This is obvious, since SAILOR optimizes the anchors for the source-trained detector and thus, the sizes depend on the source data and the corresponding labeling policy. We observe similar findings for the classes Pedestrian and Cyclist in Table 4 and

Table 5, respectively.

4. Semi-supervised Parameters

We follow the protocol of Wang *et al*. [8] and ST3D [10], to reproduce the values of Statistical Normalization (SN), Output Transformation (OT) and Random Object Scaling (ROS). In our experiments, we always utilize the average anchor size between source and target domain, as in Table 7. SN requires retraining on the source data. Using the source label, we select the respective difference in the domain statistic from Table 8, and scale object bounding boxes and points inside accordingly. Since OT does not require retraining, during evaluation we add the respective difference to the predictions. We use the predicted class as the reference to select the appropriate residual. For ROS, we compute the size difference for all classes between the two domains and from this, derive the scale range employed during source retraining. Analogously to SN, we scale object bounding boxes and points inside using a random scale value. This value is drawn from a uniform distribution constrained to the range from Table 9.

References

[1] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A Multimodal Dataset for Autonomous Driving. In *Proc. CVPR*, 2020.

[2] OpenPCDet Development Team. OpenPCDet: An Open-source Toolbox for 3D Object Detection from Point Clouds. `https://github.com/open-mmlab/OpenPCDet`, 2020.

[3] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *Proc. CVPR*, 2012.

[4] R. Kesten, M. Usman, J. Houston, T. Pandya, K. Nadhamuni, A. Ferreira, M. Yuan, B. Low, A. Jain, P. Ondruska, S. Omari, S. Shah, A. Kulkarni, A. Kazakova, C. Tao, L. Platinsky, W. Jiang, and V. Shet. Level 5 Perception Dataset 2020. `https://level-5.global/level5/data/`, 2019.

[5] 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. *TPAMI*, 43(8), 2021.

[6] 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. In *Proc. CVPR*, 2020.

[7] Laurens Van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *JMLR*, 9(11), 2008.

[8] Yan Wang, Xiangyu Chen, Yurong You, Li Erran Li, Bharath Hariharan, Mark Campbell, Kilian Q Weinberger, and Wei-Lun Chao. Train in Germany, Test in The USA: Making 3D Object Detectors Generalize. In *Proc. CVPR*, 2020.

[9] Yan Yan, Yuxing Mao, and Bo Li. SECOND: Sparsely Embedded Convolutional Detection. *Sensors*, 18(10), 2018.

[10] Jihan Yang, Shaoshuai Shi, Zhe Wang, Hongsheng Li, and Xiaojuan Qi. ST3D: Self-training for Unsupervised Domain Adaptation on 3D Object Detection. In *Proc. CVPR*, 2021.

Source	Method	Target KITTI	Target Waymo	Target nuScenes
KITTI	Source Anchors	[3.90, 1.60, 1.56]	[3.90, 1.60, 1.56]	[3.90, 1.60, 1.56]
	Target Anchors	[3.90, 1.60, 1.56]	[4.70, 2.10, 1.70]	[4.63, 1.97, 1.74]
	Ours	[4.03, 1.58, 1.50]	[4.13, 1.80, 1.53]	[4.10, 1.60, 1.70]
Waymo	Source Anchors	[4.70, 2.10, 1.70]	[4.70, 2.10, 1.70]	[4.70, 2.10, 1.70]
	Target Anchors	[3.90, 1.60, 1.56]	[4.70, 2.10, 1.70]	[4.63, 1.97, 1.74]
	Ours	[4.29, 1.87, 1.62]	[4.77, 2.07, 1.78]	[4.90, 2.00, 1.60]
nuScenes	Source Anchors	[4.63, 1.97, 1.74]	[4.63, 1.97, 1.74]	[4.63, 1.97, 1.74]
	Target Anchors	[3.90, 1.60, 1.56]	[4.70, 2.10, 1.70]	[4.63, 1.97, 1.74]
	Ours	[3.90, 1.80, 1.60]	[4.60, 2.10, 1.80]	[4.20, 1.80, 1.50]

Table 3: Absolute anchor sizes (reported as length, width and height in meters) for class Car / Vehicle for KITTI [3], Waymo [6] and nuScenes [1]. We leave other anchor parameters to the default source value.

Source	Method	Target KITTI	Target Waymo	Target nuScenes
KITTI	Source Anchors	[0.80, 0.60, 1.73]	[0.80, 0.60, 1.73]	[0.80, 0.60, 1.73]
	Target Anchors	[0.80, 0.60, 1.73]	[0.91, 0.86, 1.73]	[0.73, 0.67, 1.77]
	Ours	[0.72, 0.66, 1.82]	[0.70, 0.70, 1.70]	[0.50, 0.70, 1.90]
Waymo	Source Anchors	[0.91, 0.86, 1.73]	[0.91, 0.86, 1.73]	[0.91, 0.86, 1.73]
	Target Anchors	[0.80, 0.60, 1.73]	[0.91, 0.86, 1.73]	[0.73, 0.67, 1.77]
	Ours	[0.87, 0.81, 1.70]	[0.82, 0.92, 1.69]	[0.80, 0.70, 1.60]
nuScenes	Source Anchors	[0.73, 0.67, 1.77]	[0.73, 0.67, 1.77]	[0.73, 0.67, 1.77]
	Target Anchors	[0.80, 0.60, 1.73]	[0.91, 0.86, 1.73]	[0.73, 0.67, 1.77]
	Ours	[0.50, 0.70, 1.90]	[0.80, 0.60, 1.90]	[0.70, 0.80, 1.70]

Table 4: Absolute anchor sizes (reported as length, width and height in meters) for class Pedestrian for KITTI [3], Waymo [6] and nuScenes [1]. We leave other anchor parameters to the default source value.

Source	Method	Target KITTI	Target Waymo	Target nuScenes
KITTI	Source Anchors	$[1.76, 0.60, 1.73]$	$[1.76, 0.60, 1.73]$	$[1.76, 0.60, 1.73]$
	Target Anchors	$[1.76, 0.60, 1.73]$	$[1.78, 0.84, 1.78]$	$[1.70, 0.60, 1.28]$
	Ours	$[1.66, 0.56, 1.72]$	$[1.69, 0.60, 1.74]$	$[1.40, 0.40, 1.40]$
Waymo	Source Anchors	$[1.78, 0.84, 1.78]$	$[1.78, 0.84, 1.78]$	$[1.78, 0.84, 1.78]$
	Target Anchors	$[1.76, 0.60, 1.73]$	$[1.78, 0.84, 1.78]$	$[1.70, 0.60, 1.28]$
	Ours	$[1.87, 0.78, 1.75]$	$[1.93, 0.82, 1.73]$	$[1.90, 0.70, 1.70]$
nuScenes	Source Anchors	$[1.70, 0.60, 1.28]$	$[1.70, 0.60, 1.28]$	$[1.70, 0.60, 1.28]$
	Target Anchors	$[1.76, 0.60, 1.73]$	$[1.78, 0.84, 1.78]$	$[1.70, 0.60, 1.28]$
	Ours	$[1.30, 0.40, 1.50]$	$[1.70, 0.50, 1.40]$	$[1.80, 0.40, 1.30]$

Table 5: Absolute anchor sizes (reported as length, width and height in meters) for class Cyclists / Bicycle for KITTI [3], Waymo [6] and nuScenes [1]. We leave other anchor parameters to the default source value.

Task	Method	**Car**	**Pedestrian**	**Bicycle / Cyclist**
KITTI $\rightarrow$ **Lyft**	Source Anchors	$[3.90, 1.60, 1.56]$	$[0.80, 0.60, 1.73]$	$[1.76, 0.60, 1.73]$
	Target Anchors	$[4.75, 1.92, 1.71]$	$[0.80, 0.76, 1.76]$	$[1.76, 0.63, 1.44]$
	Ours	$[4.10, 1.76, 1.60]$	$[0.80, 0.70, 1.70]$	$[1.70, 0.60, 1.50]$
Waymo $\rightarrow$ **Lyft**	Source Anchors	$[4.70, 2.10, 1.70]$	$[0.91, 0.86, 1.73]$	$[1.78, 0.84, 1.78]$
	Target Anchors	$[4.75, 1.92, 1.71]$	$[0.80, 0.76, 1.76]$	$[1.76, 0.63, 1.44]$
	Ours	$[4.60, 2.10, 1.70]$	$[0.90, 0.78, 1.76]$	$[1.90, 0.66, 1.77]$
Lyft $\rightarrow$ **KITTI**	Source Anchors	$[4.75, 1.92, 1.71]$	$[0.80, 0.76, 1.76]$	$[1.76, 0.63, 1.44]$
	Target Anchors	$[3.90, 1.60, 1.56]$	$[0.80, 0.60, 1.73]$	$[1.76, 0.60, 1.73]$
	Ours	$[4.10, 1.70, 1.70]$	$[0.70, 0.68, 1.70]$	$[1.70, 0.69, 1.61]$

Table 6: Absolute anchor sizes (reported as length, width and height in meters) for different classes. We leave other anchor parameters to the default source value.

Source	Class	Target KITTI	Waymo	nuScenes	Lyft
KITTI	Car	-	$[4.30, 1.85, 1.63]$	$[4.26, 1.78, 1.65]$	$[4.32, 1.60, 1.56]$
	Pedestrian	-	$[0.85, 0.73, 1.73]$	$[0.76, 0.63, 1.75]$	$[0.80, 0.68, 1.74]$
	Cyclist	-	$[1.77, 0.72, 1.75]$	$[1.73, 0.60, 1.50]$	$[1.76, 0.61, 1.58]$
Waymo	Vehicle	$[4.30, 1.85, 1.63]$	-	$[4.66, 2.03, 1.72]$	$[4.66, 2.00, 1.70]$
	Pedestrian	$[0.85, 0.73, 1.73]$	-	$[0.82, 0.76, 1.75]$	$[0.85, 0.81, 1.74]$
	Cyclist	$[1.77, 0.72, 1.75]$	-	$[1.74, 0.72, 1.53]$	$[1.77, 0.73, 1.61]$
nuScenes	Car	$[4.26, 1.78, 1.65]$	$[4.66, 2.03, 1.72]$	-	-
	Pedestrian	$[0.76, 0.63, 1.75]$	$[0.82, 0.76, 1.75]$	-	-
	Bicycle	$[1.73, 0.60, 1.50]$	$[1.74, 0.72, 1.53]$	-	-
Lyft	Car	$[4.32, 1.60, 1.56]$	-	-	-
	Pedestrian	$[0.80, 0.68, 1.74]$	-	-	-
	Bicycle	$[1.76, 0.61, 1.58]$	-	-	-

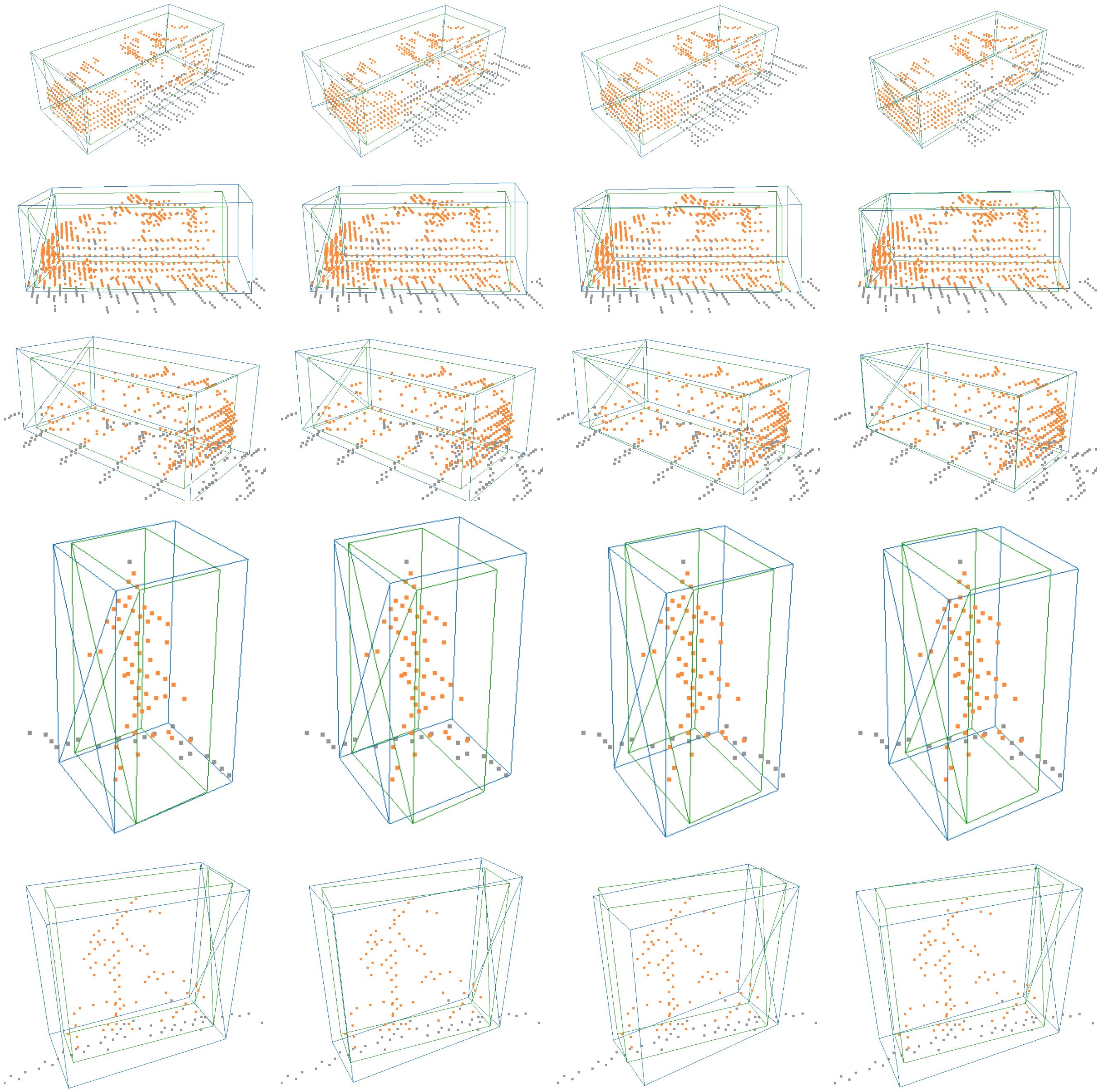
Table 7: The anchor sizes employed in the experiments with the weakly-supervised approaches SN, OT and ROS. They are the average anchor size between source and target domain. We leave other anchor parameters to the default source value.

Source	Class	Target KITTI	Waymo	nuScenes	Lyft
KITTI	Car	-	$[0.80, 0.50, 0.14]$	$[0.73, 0.37, 0.18]$	$[0.85, 0.32, 0.15]$
	Pedestrian	-	$[0.11, 0.26, 0.00]$	$[-0.07, 0.07, 0.04]$	$[0.00, 0.16, 0.03]$
	Cyclist	-	$[0.02, 0.24, 0.05]$	$[-0.06, 0.00, -0.45]$	$[0.00, 0.03, -0.29]$
Waymo	Vehicle	$[-0.80, -0.50, -0.14]$	-	$[-0.07, -0.13, 0.04]$	$[0.05, -0.18, 0.01]$
	Pedestrian	$[-0.11, -0.26, 0.00]$	-	$[-0.18, -0.19, 0.04]$	$[-0.11, -0.10, 0.03]$
	Cyclist	$[-0.02, -0.24, -0.05]$	-	$[-0.08, -0.24, -0.50]$	$[-0.02, -0.21, -0.34]$
nuScenes	Car	$[-0.73, -0.37, -0.18]$	$[0.07, 0.13, -0.04]$	-	-
	Pedestrian	$[0.07, -0.07, -0.04]$	$[0.18, 0.19, -0.04]$	-	-
	Bicycle	$[0.06, 0.00, 0.45]$	$[0.08, 0.24, 0.50]$	-	-
Lyft	Car	$[-0.85, -0.32, -0.15]$	$[-0.05, 0.18, -0.01]$	-	-
	Pedestrian	$[0.00, -0.16, -0.03]$	$[0.11, 0.10, -0.03]$	-	-
	Bicycle	$[0.00, -0.03, 0.29]$	$[0.02, 0.21, 0.34]$	-	-

Table 8: Difference between source and target dataset statistics, which we employ in our SN and OT experiments.

Source	Target KITTI	Waymo	nuScenes	Lyft
KITTI	-	$[1.05, 1.30]$	$[1.05, 1.30]$	$[1.05, 1.30]$
Waymo	$[0.70, 0.95]$	-	$[0.95, 1.00]$	$[0.95, 1.00]$
nuScenes	$[0.70, 0.95]$	$[1.00, 1.05]$	-	-
Lyft	$[0.70, 0.95]$	$[1.00, 1.05]$	-	-

Table 9: Scale ranges for different domains, which we employ in our ROS experiments.



(a) Source only (b) SN (c) ROS (d) Ours

Figure 3: Qualitative comparison of source only, Statistical Normalization (SN) [8], Random Object Scaling (ROS) [10] and our method on Waymo → KITTI case. We indicate the ground truth box in green and the predicted boxes in blue. The object points, according to the ground truth annotation, are shown in orange. Best viewed on screen.