Supplementary Material: Walking Your LiDOG: A Journey Through Multiple Domains for LiDAR Semantic Segmentation

Cristiano Saltori[1], Aljoša Ošep[2], Elisa Ricci[1,3], Laura Leal-Taixé[4]
[1] University of Trento, [2] TU Munich, [3] Federazione Bruno Kessler, [4] NVIDIA
cristiano.saltori@unitn.it

Abstract

We provide supplementary material in support of the main paper. We organize the content as follows:

- *In Sec. 1, we report the implementation details of our experiments, the additional experimental evaluations on Synth4D-nuScenes→Real, and an additional ablation of LiDOG.*
- *In Sec. 2, we report and discuss qualitative results comparing the performance of LiDOG with all the compared baselines and ground truth annotations.*

1. Experimental Evaluation

We provide the implementation details of our experiments in Sec. 1.1. In Sec. 1.2, we complete the set of experiments of the main paper and report the experimental results obtained on Synth4D-nuScenes→Real. In Sec. 1.3, we further ablate how the input resolution affects LiDOG performance.

1.1. Implementation details

We implement our method and all the baselines by using the PyTorch framework. We use MinkowskiNet [2] as a sparse convolutional backbone in all our experiments and train until convergence with voxel size 0.05m, total batch size of 16, learning rate 0.01 and ADAM optimizer [3]. In the experiments we use random rotation, scaling, and downsampling for better convergence of baselines and LiDOG. We set rotation bounds between $[-\pi/2, \pi/2]$, scaling between $[0.95, 1.05]$, and perform random downsampling for 80% of the patch points. We set projection bounds B^{3D} based on the input resolution. We set b_x and b_z to $50m$ in the denser source domains of Synth4D-KITTI and SemanticKITTI and b_x and b_z to $30m$ in the sparser domains of Synth4D-nuScenes and nuScenes. Quantization parameters x_q and z_q are computed based on the spatial bounds in order to obtain BEV labels of $168x168$ pixels. For downsampling dense features, we use a max pooling layer with window size 5, stride 3, and padding 1. The LiDOG 2D decoder is implemented with a series of three 2D convolutional layers interleaved by batch normalization layers. ReLU activation function is used on all the layers while softmax activation is applied on the last layer.

1.2. Synth4D-nuScenes→Real

In Tab. 1, we report the results for single-source *Synth4D-nuScenes→Real*. We observe a 41.82 mIoU gap (SemanticKITTI) between the *source* (19.71 mIoU) and *target* (61.53 mIoU) models on Synth4D-nuScenes→SemanticKITTI and a 23.92 mIoU gap (nuScenes), between *source* (24.57 mIoU) and *target* (48.49 mIoU) on Synth4D-nuScenes→nuScenes. Among the baselines, augmentation-based methods are the most effective with Mix3D achieving 31.64 mIoU (SemanticKITTI) and 31.23 mIoU (nuScenes). LiDOG reduces the domain gap between *source* and *target* models and outperforms all the compared baselines in all the scenarios. For example, on *Synth4D-nuScenes→Real* (Tab. 1), we obtain 34.79 mIoU, a +15.08 improvement over the *source* model.

1.3. BEV resolution

We study the impact of BEV image resolution on the LiDOG performance. We adapt the feature resolution by applying several pooling steps and the label resolution via the quantization step size. In Fig. 1, we report the results obtained on Synth4D-KITTI→Real, when using BEV features with 50% and 75% of the initial resolution (100%). Interestingly, on Synth4D-KITTI→nuScenes we observe a slight improvement with 75% resolution. However, we obtain consistently top performance on both domains with the full resolution (100%).

Table 1: **Synth4D-nuScenes→Real, single-source.** We train our model on Synth4D-nuScenes and test on SemanticKITTI and nuScenes. LiDOG improves over the *source* models by +15.08 mIoU on SemanticKITTI and by +9.21 mIoU on nuScenes. LiDOG outperforms all the compared baselines. Lower bound (*red*): a model trained n the source domain without the help of DG techniques. Upper bound (*blue*): a model directly trained on target data.

S	T	Info	Method	Vehicle	Person	Road	Sidewalk	Terrain	Manmade	Vegetation	mIoU
Synth4D-nuScenes	SemanticKITTI	Lower bound	Source	14.54	2.41	32.78	14.86	6.4	30.89	36.07	19.71
		Aug	Mix3D [5]	37.38	7.26	56.90	21.06	10.60	34.46	53.79	31.64
			PointCutMix [9]	27.51	4.32	56.04	21.37	7.02	24.38	45.35	26.57
			CoSMix [7]	16.13	6.42	39.71	14.63	13.04	23.54	30.57	20.58
		2D DG	IBN [6]	47.15	7.81	50.74	4.51	15.20	29.53	51.35	29.47
			RobustNet [1]	21.19	8.56	44.46	10.80	15.06	11.95	30.59	20.37
		3D UDA	SN [8]	11.56	2.38	37.50	10.86	5.19	20.70	39.23	18.20
			RayCast [4]	28,89	6.34	53.59	12.94	**15.86**	21.74	41.85	25.89
		3D DG	Ours	**55.08**	**11.42**	**59.48**	**26.10**	2.78	**34.83**	**53.82**	**34.79**
		Upper bound	Target	81.57	23.23	82.09	57.64	46.84	64.14	75.21	61.53
Synth4D-nuScenes	nuScenes	Lower bound	Source	17.73	8.94	36.53	7.28	5.78	52.13	43.62	24.57
		Aug	Mix3D [5]	**33.83**	17.85	47.74	8.93	9.71	56.35	44.18	31.23
			PointCutMix [9]	21.51	13.12	53.72	8.86	8.45	54.83	48.81	29.90
			CoSMix [7]	22.00	15.43	57.40	8.86	9.08	56.24	47.16	30.88
		2D DG	IBN [6]	23.07	14.13	44.96	7.10	9.83	53.70	**49.49**	28.90
			RobustNet [1]	21.59	12.29	48.52	8.14	6.22	51.33	47.68	27.97
		3D UDA	SN [8]	24.71	8.45	5.,08	5.66	**11.04**	47.00	39.05	26.57
			RayCast [4]	19.65	12.24	58.08	7.58	9.71	46.43	41.37	27.86
		3D DG	Ours (LiDOG)	26.79	**18.68**	**63.28**	**15.81**	6.57	**58.8**	46.5	**33.78**
		Upper bound	Target	47.42	20.18	80.89	37.02	34.99	64.27	54.67	48.49

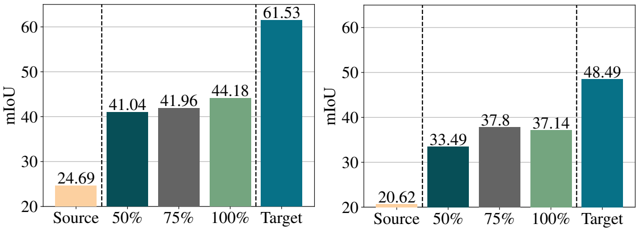

Figure 1: **BEV image resolution:** We compare the performance while changing the BEV image resolution on SemanticKITTI (*left*) and nuScenes (*right*), Source: *Synth4D-KITTI*

2. Qualitative evaluation

We report additional qualitative results for each baseline and in all the studied generalization directions. In Fig. 2-4 we show qualitative results in the Synth→Real setting: Synth4D-kitti→Real (Fig. 2), Synth4D-nuScenes→Real (Fig. 3) and Synth4D-kitti+Synth4D-nuScenes→Real (Fig. 4). In Fig. 5-6 we show qualitative results in the SemanticKITTI→nuScenes and nuScenes→SemanticKITTI directions, respectively. Source predictions are often incorrect and spatially inconsistent. Baselines consistently improve over the source model performance. LiDOG achieves the overall best performance with improved and more precise predictions. This can be seen in all the reported results, both *synth→real* and *real→real*.

References

[1] S. Choi, S. Jung, H. Yun, J.T. Kim, S. Kim, and J. Choo. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *CVPR*, 2021. 2

[2] C. Choy, J.Y. Gwak, and S. Savarese. 4D spatio-temporal convnets: Minkowski convolutional neural networks. In *CVPR*, 2019. 1

[3] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 1

[4] F. Langer, A. Milioto, A. Haag, J. Behley, and C. Stachniss. Domain transfer for semantic segmentation of LiDAR data using deep neural networks. In *IROS*, 2021. 2

[5] A. Nekrasov, J. Schult, O. Litany, B. Leibe, and F. Engelmann. Mix3D: Out-of-Context Data Augmentation for 3D Scenes. In *3DV*, 2021. 2

[6] X. Pan, P. Luo, J. Shi, and X. Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *ECCV*, 2018. 2

[7] C. Saltori, Fabio Galasso, G. Fiameni, N. Sebe, E. Ricci, and F. Poiesi. Cosmix: Compositional semantic mix for domain adaptation in 3d lidar segmentation. In *ECCV*, 2022. 2

[8] Y. Wang, X. Chen, Y. You, L.E. Li, B. Hariharan, M. Campbell, K.Q. Weinberger, and W.L. Chao. Train in germany, test in the usa: Making 3d object detectors generalize. In *CVPR*, 2020. 2

[9] J. Zhang, L. Chen, B. Ouyang, B. Liu, J. Zhu, Y. Chen, Y. Meng, and D. Wu. Pointcutmix: Regularization strategy for point cloud classification. *Neurocomputing*, 2022. 2

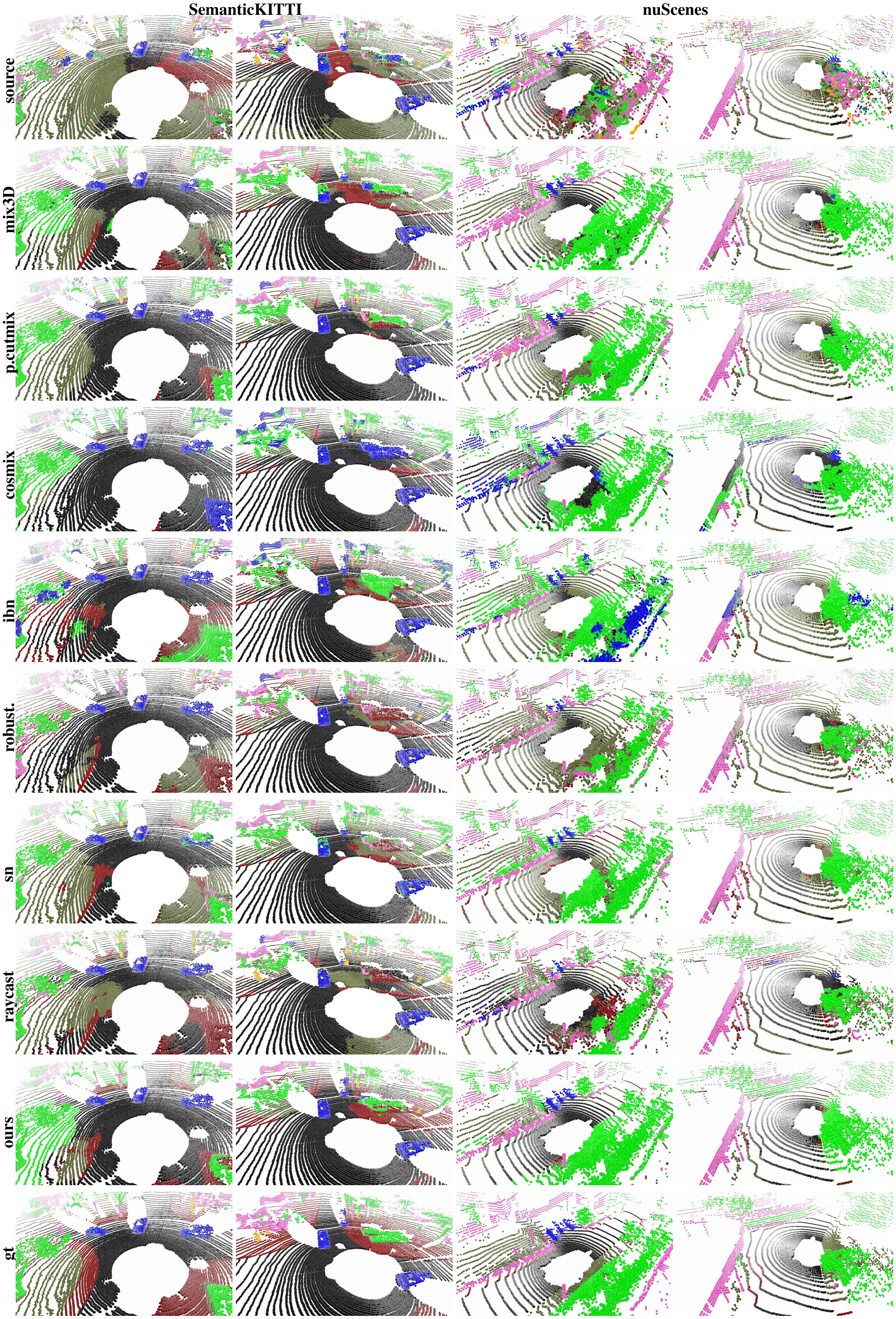

Figure 2: **Qualitative results.** *Top:* Synth4D-kitti→SemanticKITTI, *bottom:* Synth4D-kitti→nuScenes. LiDOG improves over source and baselines, *e.g.*, we observe the improvements of *road* in SemanticKITTI and *vehicle* in nuScenes.

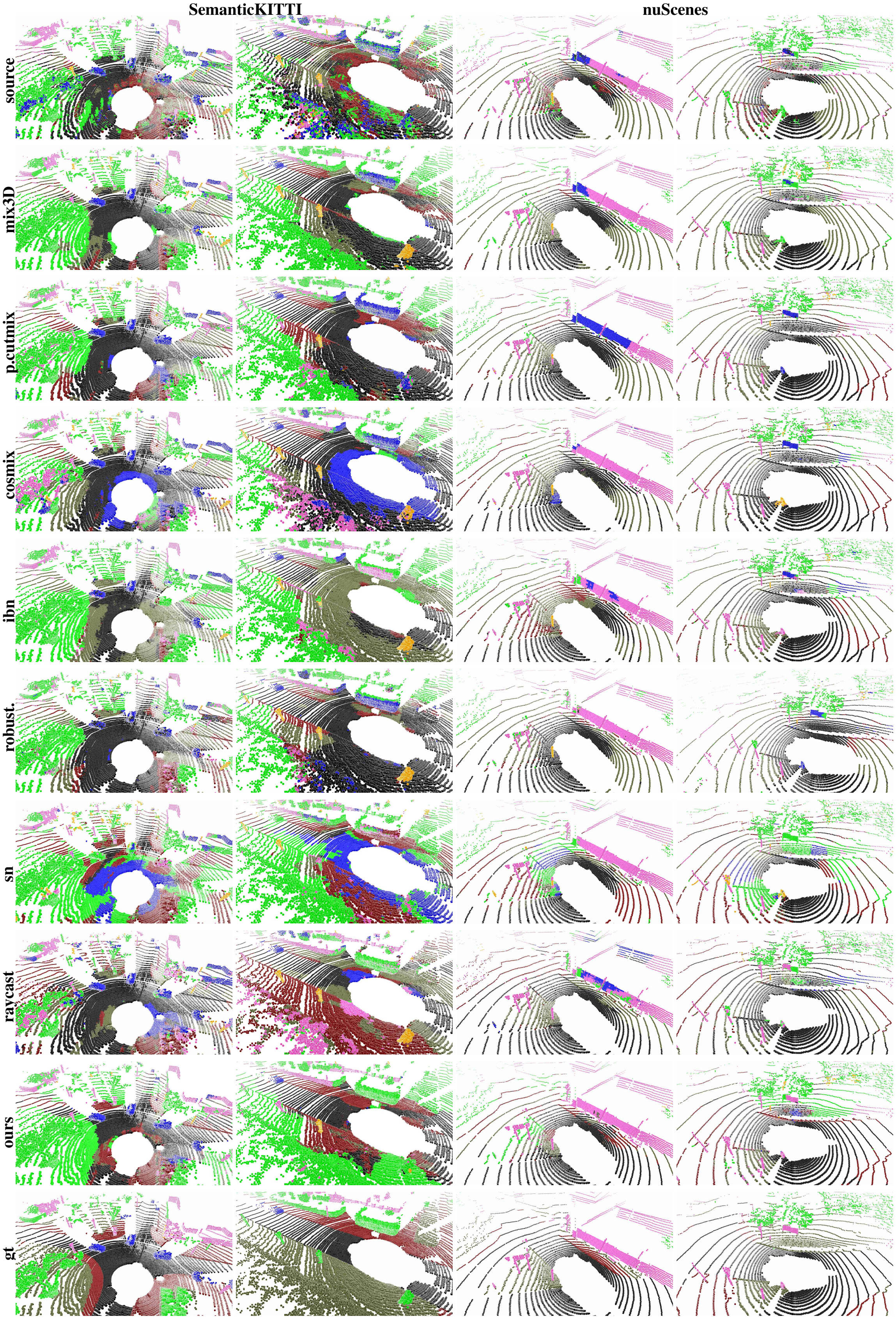

Figure 3: **Qualitative results.** *Left:* Synth4D-nuScenes→SemanticKITTI, *right:* Synth4D-nuScenes→nuScenes. LiDOG improves over source and baselines, *e.g.*, we observe the improvements of *road* in SemanticKITTI and *manmade* in nuScenes.

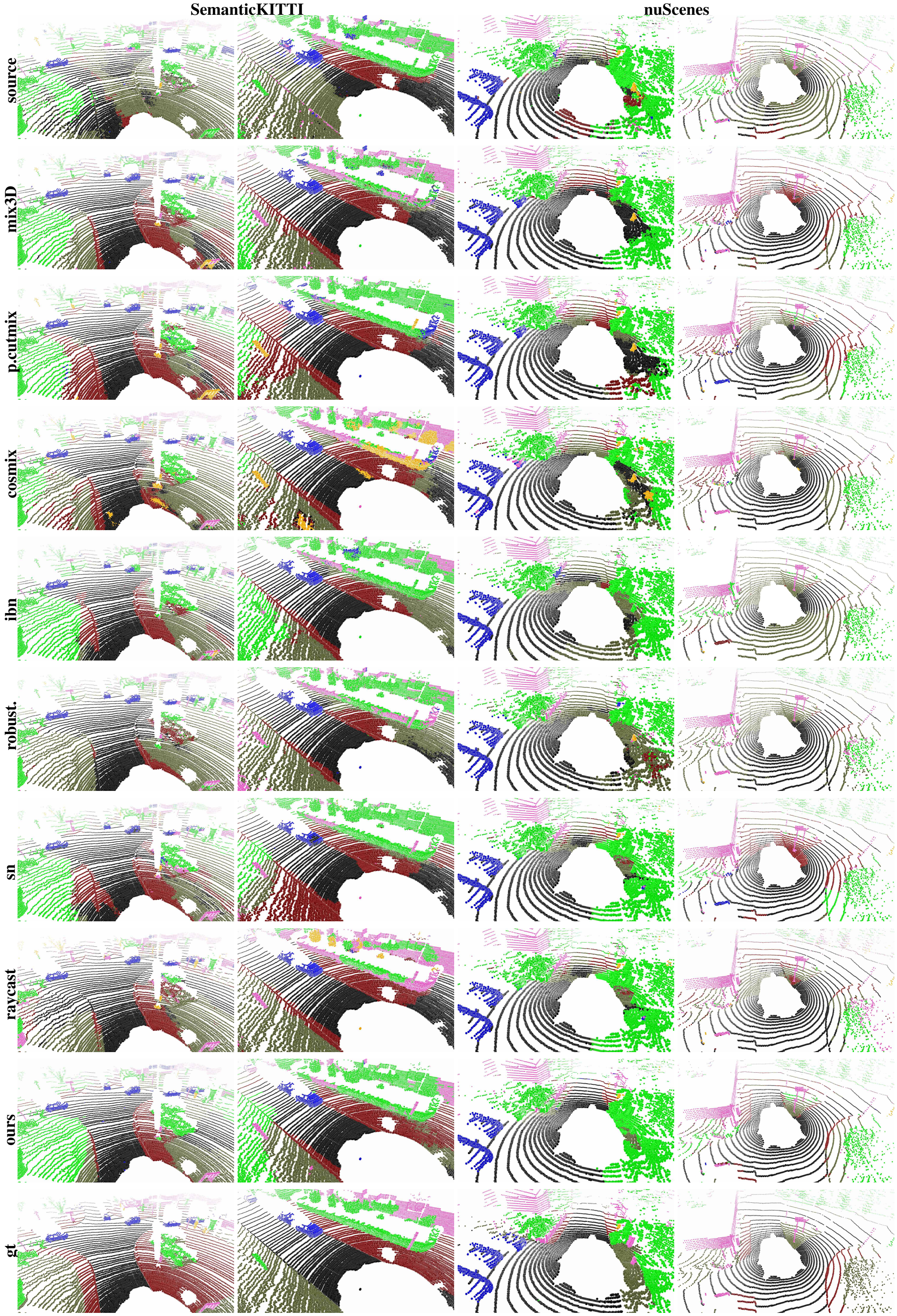

Figure 4: **Qualitative results.** *Top:* Synth4D-kitti+Synth4D-nuScenes→SemanticKITTI, *bottom:* Synth4D-kitti+Synth4D-nuScenes→nuScenes. LiDOG improves over source and baselines, *e.g.*, we observe the improvements of *vegetation* in SemanticKITTI and *road* in nuScenes.

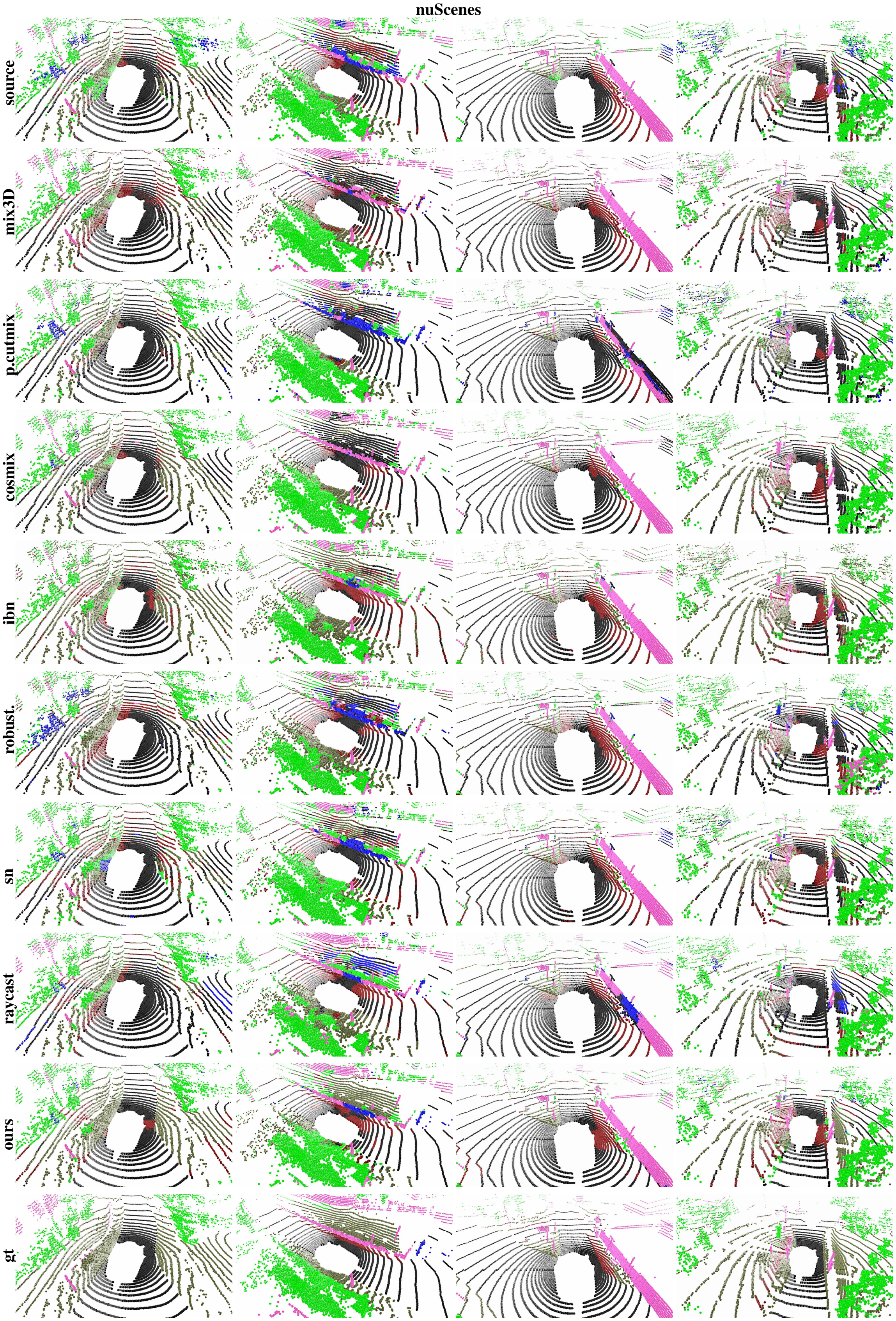

Figure 5: **Qualitative results.** *Top:* SemanticKITTI→nuScenes. LiDOG improves over source and baselines, *e.g.*, we observe the improvements in *terrain*, *road*, and *manmade*.

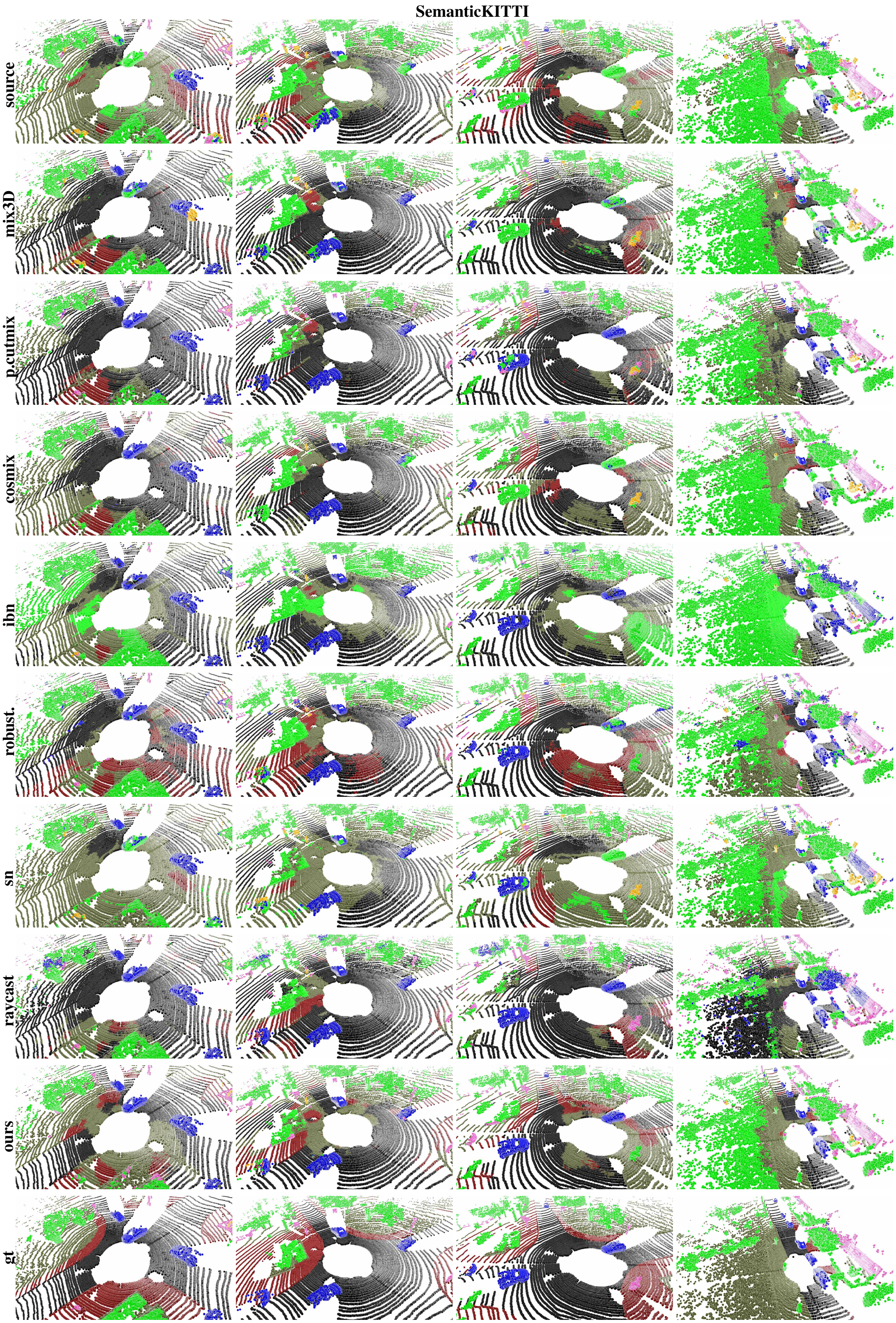

Figure 6: **Qualitative results.** *Top:* nuScenes→SemanticKITTI. LiDOG improves over source and baselines, *e.g.*, we observe the improvements in *sidewalk* and *road*.