LOSC: LiDAR Open-voc Segmentation Consolidator — Supplementary Material —

Nermin Samet[1], Gilles Puy[1], Renaud Marlet[1,2]

[1]Valeo.ai, Paris, France [2]LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France

We supplement the main paper with the performance on super-classes (Appendix A), classwise results (Appendix B), information about the quantity and quality of labeled points for training (Appendix C), details about prompts used in experiments (Appendix D), and a detailed study concerning the use of augmentations (Appendix E).

A. Results with super-classes

Some methods like SAL [5] and LeAP [4] evaluate on broader super-classes. For instance, SAL merges car, motorcycle, truck, etc., into a single 'vehicle' class. In Tab. 7, we recall results with the original fine-grained classes and also report our results adopting the same merging strategy as in SAL [5], applying class merging during inference to predict super-class labels. (As said in the experiment section of the paper, no comparison is possible with LeAP, which does not provides its mapping from 19 classes into 11.)

Besides, no code is available for SAL and the only reported metric is the mIoU on the panoptic segmentation benchmarks, which happens to slightly differ (< 1 pt in practice) from the mIoU on the semantic segmentation benchmark in the specific case of nuScenes, because of slightly different ground truths in both benchmarks. We thus report in Tab. 7 both mIoU_{sem} and mIoU_{pan}, for the nuScenes semantic and panoptic benchmarks, respectively. As can be seen in the table, LOSC clearly outperforms SAL on semantic segmentation, with both fine-grained classes and super-classes.

B. Classwise results

Tables 8 and 9 respectively present classwise semantic segmentation results on nuScenes [3] and SemanticKITTI [1] val sets, evaluated with networks initially finetuned on labelings L_{tim} or $L_{\text{aug}\otimes\text{tim}}$ (1st iter.), before iterated training (2nd and 3rd iter.). These tables detail the results reported in Table 3 of the main paper. They also provide classwise results for super-classes (cf. Appendix A).

For nuScenes, while the IoU% of most classes keeps on increasing as we iterate more, only a few classes see their performance stagnate or slightly drop. Therefore, the overall performance (mIoU%) is clearly increasing even after 3 iterations.

For SemanticKITTI, where the overall performance is lower and increasing at a lower rate as we iterate, compared to nuScenes, a higher number of classes have a best performance at the 1st or 2nd iteration. Still, the overall performance after the 3rd iteration remains the best one.

The results for the method we call LOSC (Table 5 of the main paper) correspond to three iterations.

C. Quantity and quality of labeled points for training

Table 10 below lists the number of labeled points per class in the nuScenes training set for the different labelings reported in Table 4 of the main paper. Colors in these tables illustrate the choices made in Eq. (5) to select labels that are more likely to be trustworthy, among labels from L_{tim} or L_{aug}.

Table 11 provides the classwise quality (IoU%) of labeled points in the nuScenes training set for the different labelings reported in Table 4 of the main paper. As in Table 10, color in this table illustrates the choices made in Eq. (5) to select labels from L_{tim} or L_{aug} that are more likely to be trustworthy. Note that points that receive no label with our methods are excluded when computing the metric. Hence, from one column to another (different labelings), the set of points on which the IoU is computed differs.

D. Text prompts

In this section, we provide detailed class-specific prompts for both our minimal and rich text prompt settings. Table 12 lists the *minimal* prompts employed in our final models for nuScenes and SemanticKITTI. Additionally, Table 13 presents the class-specific rich prompts used for nuScenes in the benchmarking of 2D open-vocabulary VLMs, as reported in Table 1 of the main paper.

Method	mIoU_{sem}	mIoU_{pan}	$mIoU^{(sc)}_{sem}$	$mIoU^{(sc)}_{pan}$	mIoU	$mIoU^{(sc)}$
SAL [5]	-	33.9	-	*52.6*	28.7	*52.8*
LOSC (ours)	49.3	**49.8**	*68.4*	-	**35.2**	***67.4***

Table 7. **Semantic segmentation results on nuScenes and SemanticKITTI validation sets for both original fine-grained classes and super-classes.** For nuScenes, mIoU_{sem} is computed with semantic segmentation ground truth while mIoU_{pan} is computed with the panoptic segmentation ground truth. $\text{mIoU}^{(\text{sc})}$, in blue, corresponds to a mapping of all classes into six "super-classes" (sc), as defined in SAL [5].

nuScenes	L_{tim}			$L_{\text{aug}\otimes\text{tim}}$			Full
Classes (default)	1st iter.	2nd iter.	3rd iter.	1st iter.	2nd iter.	3rd iter.	supp.
barrier	2.8	3.2	3.6	4.3	4.0	5.2	79.2
bicycle	28.2	28.0	27.7	21.2	20.4	21.3	53.2
bus	72.5	72.0	71.0	80.2	82.2	80.5	92.5
car	63.7	65.2	66.7	64.4	66.6	67.5	88.1
construction vehicle	2.4	4.6	5.9	9.3	9.8	10.6	50.4
motorcycle	57.0	57.6	58.7	62.6	63.5	64.1	87.8
pedestrian	67.1	67.9	68.2	70.8	72.1	72.7	83.4
traffic cone	39.7	44.9	46.5	42.3	45.5	46.3	70.3
trailer	0.0	0.1	0.1	0.1	0.0	0.0	73.5
truck	57.9	59.5	59.7	56.5	57.7	58.2	84.1
driveable surface	89.4	89.6	89.6	90.5	90.4	90.1	96.9
other flat	0.0	0.0	0.0	0.0	0.0	0.0	73.3
sidewalk	48.6	49.3	49.6	50.0	50.7	51.0	75.5
terrain	59.5	60.2	60.4	61.3	61.7	61.5	75.4
manmade	74.4	75.8	76.3	74.6	76.5	77.6	89.9
vegetation	78.4	79.4	79.7	79.5	81.0	81.7	86.5
mean	46.4	47.3	47.7	48.0	48.9	49.3	78.7
Superclasses							
object	5.3	6.0	6.2	6.6	6.4	7.6	
vehicle	79.9	81.8	83.3	80.5	82.4	83.0	
human	67.1	67.9	68.2	70.8	72.1	72.7	
ground	90.1	90.6	90.8	90.5	90.9	91.0	
nature	76.7	77.6	77.9	77.3	78.2	78.5	
structure	74.4	75.8	76.3	74.6	76.5	77.6	
mean	65.6	66.1	67.1	66.7	67.8	68.4	

Table 8. **Classwise semantic segmentation results (IoU%) of LOSC on nuScenes val set** when finetuning a 3D network [6, 7] on labelings L_{tim} or $L_{\text{aug}\otimes\text{tim}}$, and iterating training interleaved with time-based consolidation. "Full supp." is when training with ground-truth labels [6, 7]. Best results are in blue.

E. Augmentations

The following Table 14 presents detailed classwise results of zero-shot OpenSeeD [8] applied to the nuScenes dataset, before and after using our set of 10 distinct data augmentation methods from the Albumentations library [2]. We observe that the results remain largely consistent with the original, which helps us eliminate obviously incorrect labels. We also present the results of several other data augmentations in Table 15. These augmentations perform significantly worse than the no-augmentation baseline and result in a substantial loss of points.

SemanticKITTI	L_{tim}			$L_{\text{aug}\otimes\text{tim}}$			Full
Classes (default)	1st iter.	2nd iter.	3rd iter.	1st iter.	2nd iter.	3rd iter.	supp.
car	77.1	79.9	79.6	79.1	80.8	80.8	94.7
bicycle	35.6	37.4	37.5	33.3	35.5	34.7	44.7
motorcycle	58.4	61.3	63.4	58.0	56.0	56.7	64.8
truck	10.5	10.6	10.3	17.3	16.5	16.0	82.0
other-vehicle	4.8	2.7	2.3	30.8	32.3	34.1	37.4
person	52.8	54.5	55.0	55.0	56.7	56.5	71.4
bicyclist	0.0	0.0	0.0	0.0	0.0	0.0	85.9
motorcyclist	0.0	0.0	0.0	0.0	0.0	0.0	0.0
road	75.6	76.0	75.5	75.5	75.7	75.3	95.7
parking	0.0	0.0	0.0	0.0	0.0	0.0	49.2
sidewalk	55.2	55.9	55.3	55.2	55.3	54.9	83.6
other-ground	0.0	0.0	0.0	0.0	0.0	0.0	0.2
building	69.2	73.7	76.0	69.3	73.9	76.5	89.3
fence	8.0	8.9	9.3	7.3	9.2	9.0	56.0
vegetation	80.6	83.1	84.1	80.5	82.8	83.8	88.2
trunk	0.0	0.0	0.0	0.0	0.0	0.0	69.9
terrain	66.7	67.9	67.8	67.6	67.9	66.9	74.6
pole	0.0	0.0	0.0	0.0	0.0	0.0	66.0
traffic-sign	16.2	15.4	15.5	21.4	22.9	22.9	51.5
mean	32.1	33.0	33.2	34.2	35.0	35.2	63.4
Superclasses							
object	7.6	8.4	8.8	7.2	9.1	8.9	
vehicle	83.9	86.5	86.6	84.0	86.6	87.0	
human	65.6	66.0	65.7	63.8	63.9	63.1	
ground	84.0	84.3	84.0	84.3	84.1	83.6	
nature	84.4	85.9	86.3	84.3	85.6	85.7	
structure	69.2	73.7	76.0	69.3	73.9	76.5	
mean	65.8	67.5	67.9	65.6	67.2	67.4	

Table 9. **Classwise semantic segmentation results (IoU%) of LOSC on SemanticKITTI val set** when finetuning a 3D network [6, 7] on labelings L_{tim} or $L_{\text{aug}\otimes\text{tim}}$, and iterating training interleaved with time-based consolidation. "Full supp." is when training with ground-truth labels [6, 7]. Best results (besides full supervision) for each labeling are with a blue background.

Class	L_{vlm}	L_{tim}	L_{abc}	L_{aug}	$N_{\text{aug}}/N_{\text{tim}}$	$L_{\text{aug}\otimes\text{tim}}$	FoV_{gt}	1st iter.	2nd iter.	3rd iter.
barrier	2.0	2.1	0.5	0.4	*0.20*	2.1	7.7	4.8	4.8	4.7
bicycle	0.3	0.3	0.1	0.1	*0.46*	0.3	0.1	0.5	0.6	0.6
bus	3.7	3.7	2.5	2.4	*0.65*	2.4	3.8	8.1	7.7	7.9
car	27.4	27.5	24.4	24.0	*0.87*	24.0	32.0	93.8	102.2	103.7
construction vehicle	0.1	0.1	0.0	0.0	*0.10*	0.1	1.3	0.4	0.3	0.3
motorcycle	0.3	0.3	0.2	0.2	*0.70*	0.3	0.3	0.3	0.3	0.3
pedestrian	2.3	2.2	1.8	1.7	*0.75*	1.7	1.9	1.8	1.7	1.7
traffic cone	0.2	0.2	0.1	0.1	*0.45*	0.2	0.6	0.5	0.5	0.5
trailer	0.2	0.1	0.0	0.0	*0.06*	0.1	4.5	0.9	1.1	1.2
truck	20.4	20.4	15.7	15.5	*0.76*	15.5	13.3	58.7	54.8	49.6
driveable surface	156.3	149.4	142.1	134.3	*0.90*	134.3	260.9	298.0	299.6	298.9
other flat	0.1	0.1	0.1	0.1	*0.47*	0.1	6.6	0.7	0.3	0.3
sidewalk	51.7	50.0	38.4	36.5	*0.73*	36.5	57.6	69.5	68.2	68.1
terrain	38.6	37.7	28.1	27.0	*0.72*	27.0	56.7	54.1	50.1	47.5
manmade	128.6	128.5	109.9	108.7	*0.85*	108.7	146.5	168.9	152.2	147.0
vegetation	94.8	92.4	79.0	77.1	*0.83*	77.1	100.6	215.8	232.3	244.4

Table 10. **Quantity of labeled points used in nuScenes train set.** The number of points is given in millions. Values under 50k appear as "0.0". Values not meeting the conditions in Eq. (4) are indicated with a yellow background. Labels from L_{tim} or L_{aug} that are selected in $L_{\text{aug}\otimes\text{tim}}$ as specified in Eq. (5) are shown with a green background. Column "FoV_{gt}" gives the number of points in the field of view of the cameras for each class according to the ground truth, for comparison purposes. The following columns (1st to 3rd iter.) provide results of trained 3D networks and thus concern all scanned points, not just those visible from the cameras.

Class	L_{vlm}	L_{tim}	L_{abc}	L_{aug}	$L_{\text{aug}\otimes\text{tim}}$	1st iter.	2nd iter.	3rd iter.	F. supp.
barrier	4.2	4.6	2.2	2.1	5.4	5.0	4.8	4.8	79.2
bicycle	21.0	22.0	26.5	28.0	23.6	17.8	15.7	15.9	53.2
bus	57.5	57.5	74.4	74.9	74.9	66.4	67.2	67.7	92.5
car	63.4	63.5	69.8	70.5	70.4	73.3	76.1	77.3	88.1
construction vehicle	5.4	5.2	1.4	1.1	7.2	11.9	12.1	12.4	50.4
motorcycle	41.8	43.0	50.5	52.2	48.5	57.2	59.1	59.6	87.8
pedestrian	44.0	47.0	52.5	56.6	56.4	70.5	71.7	72.0	83.4
traffic cone	24.3	26.0	25.8	26.8	34.6	45.5	48.0	48.8	70.3
trailer	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	73.5
truck	41.3	41.5	48.4	49.2	49.1	47.7	48.4	48.7	84.1
driveable surface	84.4	84.2	90.7	91.0	91.0	89.8	89.8	89.7	96.9
other flat	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	73.3
sidewalk	48.4	48.6	57.9	59.0	59.0	53.5	54.3	54.8	75.5
terrain	52.3	52.6	65.2	66.6	66.4	57.0	57.5	57.5	75.4
manmade	69.5	70.3	77.8	79.1	78.3	74.2	75.7	76.5	89.9
vegetation	71.4	72.1	79.8	81.0	81.0	79.3	80.6	81.2	86.5
mean	39.3	39.9	45.1	46.1	46.6	46.8	47.6	47.9	78.7

Table 11. **Quality (IoU%) of labeled points used in nuScenes train set.** Labels from L_{tim} or L_{aug} selected in $L_{\text{aug}\otimes\text{tim}}$ are shown with a green background. Columns "1st iter." to "3rd iter." provide results of trained 3D networks and thus concern all scanned points, not just those visible from the cameras as in the previous columns. "F. supp." is when training with ground-truth labels [6, 7].

Table 12. Our classwise *minimal* text prompts used to obtain segmentation maps from OpenSeed [8].

Class	Text prompts
	nuScenes
pedestrian	person, pedestrian
bicycle	bicycle
bus	bus
car	car
construction vehicle	bulldozer, excavator, concrete mixer , crane, dump truck
motorcycle	motorcycle
trailer	trailer, sem trailer, cargo container, shipping container, freight container
truck	truck
barrier	barrier, barricade
traffic cone	traffic cone
driveable surface	road
other flat	curb, traffic island, traffic median
sidewalk	sidewalk
terrain	terrain, grass, grassland, lawn, meadow, turf, sod
manmade	building, wall, pole, awning
vegetation	tree, trunk, tree trunk, bush, shrub, plant, flower, woods
	SemanticKITTI
car	car
bicycle	bicycle
motorcycle	motorcycle
truck	truck
other vehicle	trailer, semi trailer, cargo container, shipping container, freight container, caravan, bus, bulldozer, excavator, concrete mixer, crane, dump truck, train, tram
person	person, pedestrian
bicyclist	bicyclist, cyclist
motorcyclist	motorcyclist
road	road
parking	parking, parking lot
sidewalk	sidewalk, curb, bike path, walkway, pavement,footpath,footway,boardwalk, driveway
other ground	water, river, lake, watercourse, waterway, canal, ditch, rail track, traffic island, traffic median, median strip, roadway median, central reservation
building	building, house, garage, wall, railing, stairs, awning, roof, bridge
fence	fence, barrier, barricade
vegetation	tree, bush, shrub, plant, flower
trunk	tree trunk, trunk, woods
terrain	terrain, grass, soil, grassland, hill, sand, gravel, lawn, meadow, garden, earth, peeble, rock
pole	pole
traffic sign	traffic-sign

Table 13. Our *rich* text prompts used to obtain segmentation maps from OpenSeed [8] on nuScenes.

Class	Text prompts
	nuScenes
pedestrian	pedestrian, person, human, man, woman, adult, child, stroller, wheelchair
bicycle	bicycle
bus	bus, autobus, motorbus, omnibus, double-decker, jitney, minibus, motor coach, school bus, tour bus, sightseeing bus, shuttle bus, shuttle, bendy bus
car	car, automobile, van, SUV, sedan, hatchback, wagon, minivan, convertible, jeep,ambulance
construction vehicle	bulldozer, excavator, concrete mixer, crane, dump truck, dozer, digger, dumper, tipper truck, tipper lorry, front loader, loader, backhoe, trencher, compactor, forklift
motorcycle	motorcycle, motorbike, choper, motor scooter, scooter, vespa
trailer	trailer, semi trailer, cargo container, shipping container, freight container
truck	truck, pickup, pickup truck, lorry, autotruck, semi, semi-tractor, semi-truck, motortruck
barrier	barrier, jersey barrier, barricade
traffic cone	traffic cone, safety pylon, road cone, highway cone, safety cone, caution cone, channelizing device, construction cone, cone, road delineator
driveable surface	road, parking, parking lot, highway, expressway, freeway, thoroughfare, route, thruway, turnpike, roadway, macadam, lane
other flat	water, river, lake, watercourse, waterway, canal, ditch, railway track, traffic island, traffic median, median strip, roadway median, central reservation
sidewalk	sidewalk, curb, bike path, walkway, pavement, footpath, footway, boardwalk
terrain	terrain, grass, grassland, hill, soil, sand, gravel, lawn, meadow, garden, earth, peeble, rock
manmade	building, skyscraper, house, wall, stairs, awning, roof, pole, streetlight, traffic light, traffic sign, bench, fire hydrant, hydrant, guard rail, guardrails, bollard, fence, railing, drainage, flag, banner, street sign, circuit box, traffic signal, stoplight, lamppost, bus shelter, bus stop, bin, mailbox, newspaper box, newsbox
vegetation	tree, bush, plant, shrub, potted plant, hedge, branch, tree trunk, flower, woods, forest

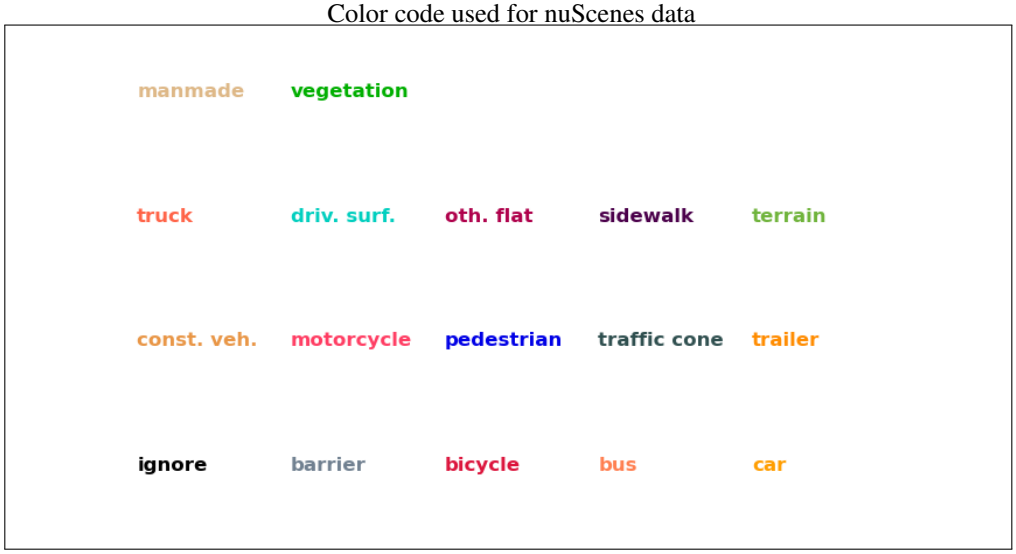
Class	no aug.	horizontal flip	hue saturation	blur	color jitter	auto-contrast	sharpen	chromatic aberration	emboss	fancy pca	clahe
barrier	1.6	1.9	2.1	1.7	1.7	2.6	1.7	1.6	1.7	1.6	1.4
bicycle	24.3	24.2	23.9	23.5	24.3	24.0	24.2	23.8	23.7	24.3	24.0
bus	56.7	57.2	54.6	57.4	56.2	49.7	53.0	56.0	56.6	56.8	55.6
car	47.9	48.0	48.0	48.4	48.1	46.3	47.8	47.7	47.2	47.9	48.3
construction vehicle	5.2	5.1	4.0	3.7	4.2	3.1	3.7	3.3	3.8	5.2	4.2
motorcycle	31.2	31.2	30.7	31.2	30.8	29.9	30.8	30.6	31.1	31.2	29.9
pedestrian	40.4	40.5	40.3	40.6	40.6	39.5	40.2	40.3	39.7	40.4	39.0
traffic cone	20.5	20.4	19.2	21.3	20.3	22.8	18.3	19.8	19.0	20.5	16.6
trailer	0.1	0.1	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.1
truck	45.7	45.5	45.0	46.1	45.8	44.3	45.2	45.1	45.1	45.7	46.0
driveable surface	50.6	50.6	50.3	50.2	50.4	50.2	50.2	50.1	50.3	50.6	50.4
other flat	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
sidewalk	34.4	34.4	33.9	33.6	34.1	33.8	34.2	33.0	33.8	34.4	33.9
terrain	43.9	43.8	42.7	43.4	43.7	44.0	43.6	43.3	43.2	43.8	43.6
manmade	66.7	66.7	65.9	66.9	66.4	65.2	65.7	66.4	65.7	66.7	66.6
vegetation	66.7	66.6	65.7	67.4	66.2	65.6	64.4	66.8	64.9	66.7	67.7
mean IoU	33.4	33.5	32.9	33.4	33.3	32.5	32.6	32.9	32.8	33.4	32.9
mean Acc	43.4	43.4	42.8	43.3	43.1	42.8	42.4	42.7	42.4	43.4	42.9

Table 14. **Classwise results on the nuScenes validation set, before and after applying our 10 selected data augmentations.** These augmentations yield results comparable to the no-augmentation baseline and help eliminate unreliable predictions.

Class	no aug.	salt-and pepper	gauss noise	random sun-flare	random snow	random rain	glass blur	shot noise	superpixels	vertical flip
barrier	1.6	1.6	1.0	1.3	1.3	1.1	0.5	0.9	0.9	0.0
bicycle	24.3	21.6	17.8	24.0	20.8	23.5	14.3	14.4	14.1	14.8
bus	56.7	46.9	50.0	45.5	50.5	42.9	35.2	37.3	36.8	22.8
car	47.9	48.2	48.2	45.3	48.0	46.2	47.5	45.5	39.6	36.5
construction vehicle	5.2	1.2	1.0	2.2	0.9	2.4	0.0	1.0	0.7	0.0
motorcycle	31.2	27.9	27.2	29.2	27.6	26.6	18.5	23.1	21.0	5.4
pedestrian	40.4	37.4	36.3	39.7	39.7	40.1	37.7	31.2	33.4	31.6
traffic cone	20.5	17.8	19.0	18.2	14.7	20.5	17.6	17.2	14.0	0.0
trailer	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.7
truck	45.7	43.0	45.5	42.5	44.0	39.9	37.8	35.4	35.4	22.3
driveable surface	50.6	48.4	48.7	48.9	48.7	48.2	45.1	46.6	46.6	17.1
other flat	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
sidewalk	34.4	29.8	28.9	31.0	28.8	27.4	16.5	25.2	26.4	7.5
terrain	43.9	41.0	41.6	38.9	37.6	42.6	40.0	37.3	40.5	36.6
manmade	66.7	63.2	61.3	62.7	65.3	60.1	59.9	56.3	61.4	40.3
vegetation	66.7	60.8	60.8	59.6	64.7	60.2	62.8	55.2	59.0	60.0
mean IoU	33.4	30.5	30.4	30.5	30.7	30.1	27.0	26.6	26.8	18.4
mean Acc	43.4	39.3	39.8	40.5	40.1	40.0	35.0	35.4	34.6	27.1

Table 15. **Classwise results on the nuScenes validation set, comparing performance before and after applying several additional data augmentations.** These augmentations significantly degraded performance compared to the baseline without any augmentation. Due to their negative impact and deviation from the original scores, we excluded them from our pipeline to ensure sufficient points in the scenes.

Color code used for nuScenes data



Color code used for SemanticKITTI data

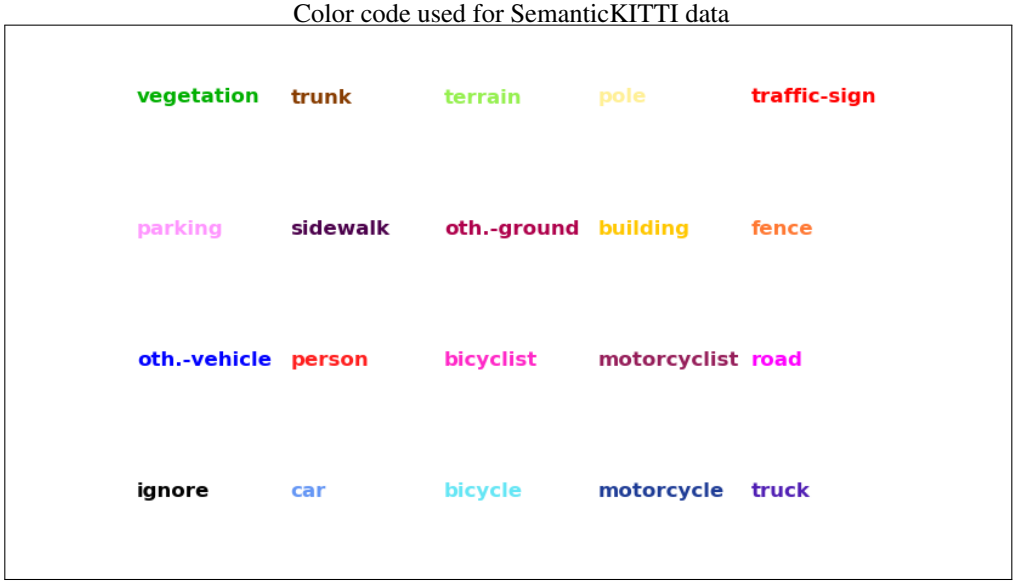


Figure 3. Color code used to represent each class on nuScenes (top) and SemanticKIITI (bottom) in Figure 3.

References

[1] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. In *ICCV*, pages 9297–9307, 2019.

[2] A. Buslaev, A. Parinov, E. Khvedchenya, V. I. Iglovikov, and A. A. Kalinin. Albumentations: fast and flexible image augmentations. *ArXiv e-prints*, 2018.

[3] 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 *CVPR*, 2020.

[4] Simon Gebraad, Andras Palffy, and Holger Caesar. Leap: Consistent multi-domain 3d labeling using foundation models. *arXiv preprint arXiv:2502.03901*, 2025.

[5] Aljoša Ošep, Tim Meinhardt, Francesco Ferroni, Neehar Peri, Deva Ramanan, and Laura Leal-Taixé. Better call sal: Towards learning to segment anything in lidar. In *European Conference on Computer Vision*, pages 71–90. Springer, 2024.

[6] Gilles Puy, Alexandre Boulch, and Renaud Marlet. Using a waffle iron for automotive point cloud semantic segmentation. In *ICCV*, 2023.

[7] Gilles Puy, Spyros Gidaris, Alexandre Boulch, Oriane Siméoni, Corentin Sautier, Patrick Pérez, Andrei Bursuc, and Renaud Marlet. Three pillars improving vision foundation model distillation for lidar. In *CVPR*, 2024.

[8] Hao Zhang, Feng Li, Xueyan Zou, Shilong Liu, Chunyuan Li, Jianwei Yang, and Lei Zhang. A simple framework for open-vocabulary segmentation and detection. In *ICCV*, 2023.