Supplementary Materials:
UniSeg: A Unified Multi-Modal LiDAR Segmentation Network and the OpenPCSeg Codebase

Youquan Liu[1,2,*]   Runnan Chen[1,3,]   Xin Li[1,4]   Lingdong Kong[1,5]   Yuchen Yang[1,6]
Zhaoyang Xia[1,6]   Yeqi Bai[1,†]   Xinge Zhu[7]   Yuexin Ma[8]   Yikang Li[1,†]   Yu Qiao[1]   Yuenan Hou[1,†]

[1]Shanghai AI Laboratory   [2]Hochschule Bremerhaven   [3]The University of Hong Kong   [4]East China Normal University
[5]National University of Singapore   [6]Fudan University   [7]The Chinese University of Hong Kong   [8]Shanghai Tech University

In this file, we supplement additional materials to support our findings, observations, and experimental results. Specifically, this file is organized as follows:

- Sec. A provides additional information on the OpenPCSeg codebase and summarizes the reproduced and reported performance.

- Sec. B elaborates on additional implementation details of the proposed methods and the experiments.

- Sec. C supplements additional quantitative results, including class-wise IoU scores and PQ scores for our comparative study and ablation study.

- Sec. D attaches additional qualitative results.

A. Additional Information of OpenPCSeg

The OpenPCSeg codebase supports tasks of LiDAR semantic segmentation and LiDAR panoptic segmentation. It includes range-image-based, voxel-based, fusion-based, point-based and BEV-based algorithms, as well as recent 3D data augmentation techniques. Range-image-based methods include SqueezeSeg [44], Squeeze-SegV2 [45], RangeNet++ [28], FIDNet [57], CENet [5] and SalsaNext [8]. Voxel-based algorithms have Minkowski-iNet [7], Cylinder3D [60], and DS-Net [15]. Fusion-based algorithms include RPVNet [48] and SPVCNN [39]. Point-based algorithms contain PointTransformer [56]. BEV-based algorithms including PolarNet [55], and Panoptic-PolarNet [58]. We also have three useful data augmentation algorithms, LaserMix [20], PolarMix [46], Mix3D [29]. A summary of supported features compared to the existing codebase is provided in Table A. OpenPCSeg supports more datasets and more features than other codebases. A detailed comparison between the reproduced and reported performance of different algorithms is summarized in Table B. Besides, we provide MinkowskiNet [7] and SPVCNN [38] variants are shown in Table C. More popular LiDAR segmentation algorithms, such as Panoptic-PHNet [23] and LidarMultiNet [52], will be added to this codebase in the future. We elaborate on more details of the benchmarked models, techniques, and datasets as follows.

A.1. Supported LiDAR Segmentation Model

A.1.1   Range View

- SqueezeSeg [44]: a classic 3D segmentor which can be trained end-to-end, proposed in 2017.

- SqueezeSegV2 [45]: an improvement over Squeeze-Seg by the Context Aggregation Module (CAM) to mitigate the impact of dropout noise, proposed in 2018.

- RangeNet++ [28]: a classic and widely used range view LiDAR semantic segmentation method which equips with GPU-enabled post-processing, proposed in 2019.

- SalsaNext [8]: a range-view solution for LiDAR semantic segmentation task which brings a Bayesian treatment to compute the *epistemic* and *aleatoric* uncertainties for each point, proposed in 2020.

- FIDNet [57]: a 3D segmentor with an improved post-processing method (NLA) over RangeNet++ and equips with an FID module for upsampling, proposed in 2021.

- CENet [5] a powerful range view method embedding multiple auxiliary segmentation heads for LiDAR segmentation task, proposed in 2022.

Table A: Supported features of existing LiDAR segmentation codebases. " ✓ " / " ✗ " denotes a supported / not supported feature. Symbol "△" denotes a feature that is to be supported in future updates.

Type	Feature	MMDetection3D[*]	3D-SemSeg[†]	lidarseg3d[‡]	Open3D-ML[§]	OpenPCSeg (Ours)
Task	Semantic Segmentation	✓	✓	✓	✓	✓
	Panoptic Segmentation	✗	✗	✗	✗	✓
	4D Panoptic Segmentation	✗	✗	✗	✗	✓
Dataset	SemanticKITTI	✓	✓	✓	✓	✓
	nuScenes	✗	✓	✓	✗	✓
	Waymo Open	✗	✗	✗	✗	✓
	ScribbleKITTI	✗	✗	✗	✗	✓
Model	SqueezeSeg	✗	✗	✗	✗	✓
	SqueezeSegV2	✗	✗	✗	✗	✓
	RangeNet++	✗	✗	✗	✗	✓
	SalsaNext	✗	✓	✗	✗	✓
	FIDNet	✗	✗	✗	✗	✓
	CENet	✗	✗	✗	✗	✓
	PolarNet	✗	✗	✗	✗	✓
	Panoptic-PolarNet	✗	✗	✗	✗	✓
	RandLA-Net	✗	✗	✗	✓	✗
	KPConv	✗	✗	✗	✓	✗
	SparseConvUnet	✗	✗	✗	✓	✗
	PointTransformer	✗	✗	✗	✓	△
	PointNet++	✓	✗	✗	✗	✗
	PAConv	✓	✗	✗	✗	✗
	DGCNN	✓	✗	✗	✗	△
	MinkowskiNet	✗	✗	✗	✗	✓
	Cylinder3D	✗	✓	✗	✗	✓
	DS-Net	✗	✗	✗	✗	✓
	4D-DS-Net	✗	✗	✗	✗	✓
	RPVNet	✗	✗	✗	✗	✓
	SPVCNN	✗	✗	✗	✗	✓
	2DPASS	✗	✓	✗	✗	△
	COARSE3D	✗	✓	✗	✗	△
	SDSeg3D	✗	✗	✓	✗	✗
	MSeg3D	✗	✗	△	✗	✗
Augmentation	Mix3D	✗	✗	✗	✗	✓
	LaserMix	✗	✗	✗	✗	✓
	PolarMix	✗	✗	✗	✗	✓
	# Supported Features	5	7	5	6	28

- COARSE3D [24]: a weakly supervised LiDAR semantic segmentation framework with a compact class-prototype contrastive learning scheme, proposed in 2022.

posed in 2020.

A.1.2 Bird's Eye View

- PolarNet [55]: a classic 3D segmentor which quantizing points into polar bird's-eye-view (BEV) grids, pro-

- Panoptic-PolarNet [58]: learn both semantic segmentation and class-agnostic instance clustering in a single network using a BEV representation to perform LiDAR panoptic segmentation task, proposed in 2021.

Table B: Comparisons between the reproduced performance in the **OpenPCSeg codebase** (mIoU-rep, PQ-rep) and reported performance from the original papers (mIoU-ori, PQ-ori). We benchmark various popular LiDAR semantic segmentation methods and LiDAR panoptic segmentation methods on the validation sets of SemanticKITTI [3] and nuScenes [4]. Note that we only report range-view methods with sizes 64×2048 and 32×1920 for SemanticKITTI and nuScenes, respectively.

Model	Type	SemanticKITTI				nuScenes			
		mIoU-ori	mIoU-rep	PQ-ori	PQ-rep	mIoU-ori	mIoU-rep	PQ-ori	PQ-rep
Mix3D [29]	Aug	–	–	–	–	–	–	–	–
LaserMix [20]		–	–	–	–	–	–	–	–
PolarMix [46]		–	–	–	–	–	–	–	–
SqueezeSeg [44]	Range	31.6	33.0(+1.4)	–	–	–	–	–	–
SqueezeSegV2 [45]		41.3	44.5(+3.2)	–	–	–	–	–	–
RangeNet$_{21}$ [28]		47.2	49.8(+2.6)	–	–	–	–	–	–
RangeNet$_{53}$ [28]		50.3	53.3(+3.0)	–	–	–	–	–	–
RangeNet$_{53}$++ [28]		52.2	54.0(+1.8)	–	–	–	65.8	–	–
SalsaNext [8]		55.8	58.2(+2.4)	–	–	–	68.1	–	–
FIDNet [57]		58.8	60.4(+2.6)	–	–	–	71.8	–	–
CENet [5]		62.6	63.7(+1.1)	–	–	–	73.4	–	–
PolarNet [55]	BEV	57.2	58.3(+1.1)	–	–	–	71.4	–	–
Panoptic-PolarNet [58]		–	–	59.1	59.5(+0.4)	–	–	67.7	67.8(+0.1)
MinkowskiNet [7]	Voxel	61.1	68.8(+7.7)	–	–	–	73.2	–	–
Cylinder3D [60]		65.9	66.9(+1.0)	–	–	76.1	76.2(+0.1)	–	–
DS-Net [15]		–	–	57.7	58.0(+0.3)	–	–	42.5	61.0(+18.5)
RPVNet [48]	Fusion	68.3	68.8(+0.5)	–	–	77.6	77.6(+0.0)	–	–
SPVCNN [39]		63.8	68.7(+3.9)	–	–	–	74.8	–	–

A.1.3 Point View

- PointTransformer [56]: a powerful 3D network that is constructed with the Transformer architecture [42], proposed in 2021.

- DGCNN [43]: a classic and widely used segmentation and classification method constructed by using EdgeConv, proposed in 2018.

A.1.4 Voxel & Cylinder

- MinkowskiNet [7]: a classic and widely used LiDAR segmentation and classification method, proposed in 2019.

- Cylinder3D [60]: a cylindrical and asymmetrical 3D convolution network for LiDAR semantic segmentation, proposed in 2021.

- DS-Net [15]: adopts consensus-driven fusion module and the dynamic shifting module for LiDAR panoptic segmentation, proposed in 2021.

- 4D-DS-Net [14]: an extensive network of DS-Net to perform 4D panoptic LiDAR segmentation via temporally unified instance clustering on the aligned adjacent LiDAR frames, proposed in 2022.

A.1.5 Fusion

- SPVCNN [39]: a powerful 3D segmentor adopt point-voxel fusion, proposed in 2020.

- RPVNet [48]: a multi-view LiDAR semantic segmentation method which includes range-point-voxel fusion, proposed in 2021.

- 2DPASS [49]: a new framework for LiDAR semantic segmentation via 2D prior-related knowledge distillation, proposed in 2022.

A.2. Supported Data Augmentation Technique

- Mix3D [29]: a data augmentation technique for segmenting large-scale 3D scenes which build new training samples by mixing two augmented scenes, proposed in 2021.

- PolarMix [46]: a data augmentation technique that cuts, edits, and mixes point clouds along the scanning direction from two scenes, proposed in 2022.

- LaserMix [20]: a powerful data augmentation technique that intertwines laser beams from different LiDAR scans, proposed in 2022.

A.3. Supported LiDAR Segmentation Dataset

- SemanticKITTI [3]: a large-scale outdoor dataset for semantic scene understanding of LiDAR sequences

Table C: Comparisons among the variants of MinkowskiNet[7] and SPVCNN[39] in the **OpenPCSeg codebase**. Results are on the validation sets of SemanticKITTI [3], nuScenes [4] and Waymo Open [37]. Symbol *mk* denotes the number of layers of the network; Symbol *cr* is the channel expansion rate. Note that the default setting of *mk* and *cr* are 18 and 1.0, respectively, for MinkowskiNet[7] and SPVCNN[39].

Model	Variant	Type	#Param	SemanticKITTI		nuScenes		Waymo Open	
				mIoU-ori	mIoU-rep	mIoU-ori	mIoU-rep	mIoU-ori	mIoU-rep
MinkowskiNet [7]	mk18cr0.5	Voxel	5.5 M	58.9	68.7$_{(+9.8)}$	–	–	–	–
MinkowskiNet [7]	mk18cr1.0		21.7 M	61.1	68.8$_{(+7.7)}$	–	73.2	–	66.7
MinkowskiNet [7]	mk34cr1.0		37.9 M	–	70.1	–	75.7	–	–
MinkowskiNet [7]	mk34cr1.6		96.5 M	–	70.1	–	76.2	–	68.2
SPVCNN [39]	mk18cr0.5	Fusion	5.5 M	60.7	68.7$_{(+8.0)}$	–	–	–	–
SPVCNN [39]	mk18cr1.0		21.8 M	63.8	67.6$_{(+3.8)}$	–	74.8	–	66.8
SPVCNN [39]	mk34cr1.0		37.9 M	–	69.0	–	76.1	–	–
SPVCNN [39]	mk34cr1.6		96.7 M	–	68.4	–	76.8	–	68.6

collected from the 64-beam scan sensor, proposed in 2019.

- nuScenes [9, 4]: a large-scale benchmark with support for various tasks, including camera images and LiDAR scans, and the point clouds are collected from the 32-beam scan sensor, proposed in 2020.

- Waymo Open [37]: A large-scale outdoor dataset consisting of well-synchronized and calibrated high-quality LiDAR and camera data, and the point clouds are collected from the 64-beam scan sensor, proposed in 2020.

- ScribbleKITTI [41]: is a recent variant of the SemanticKITTI dataset, which contains the same number of scans but is annotated with line scribbles (approximately 8.06% valid semantic labels) rather than dense annotation, proposed in 2022.

B. Additional Implementation Details

Network Structure. For the image branch, the input image size is 376×1241 on the SemanticKITTI [3] dataset. For the multi-camera images of nuScenes [4, 9] and Waymo Open [37] datasets, the image size is 900×1600 and 640×960, respectively. For the range branch, the input range-image size on the SemanticKITTI, nuScenes and Waymo Open datasets are 64×2048, 32×1920, and 64×2688, respectively. To construct a robust point-voxel-range fusion network for the point cloud branch, we first construct the point-voxel backbone based on the Minkowski-UNet34 [7]. Then, we add the range-image branch, i.e., SalsaNext [8], to the point-voxel network and perform point-voxel-range fusion by the **L**earnable cross-**V**iew **A**ssociation module (LVA). Range and voxel branches are UNet-like architectures with four down-sampling stages and four up-sampling stages. The dimensions of these nine

stages are 32, 32, 64, 128, 256, 256, 128, 96, and 96, respectively, and the point branch includes 4 MLPs with channel dimensions being 32, 256, 128, and 96, respectively. In addition, to increase model capacity, the channel expansion ratio is set as 1.75, 1.6, 1.6 for SemanticKITTI, nuScenes and Waymo Open datasets, respectively. We use ImageNet-pretrained ResNet-34 [13] as the feature extractor for the image backbone. The image backbone can be flexibly selected from off-the-shelf networks.

Data Augmentation and Test-Time Augmentation. We take different data augmentation strategies for the point cloud and image branches. For the image branch, we do not perform data augmentation. For the point cloud branch, we perform random flip (τ_{flip}) along with the X axis, Y axis and XY axis, and random translation (τ_{trans}) within the normal distribution of $[0, 0.1]$ as well as LaserMix [20] and PolarMix [46]. Global scaling (τ_{scal}) and global rotation (τ_{rot}) are also adopted. The scaling factor and rotation angle are randomly selected within $[0.9, 1.1]$ and $[0, 2\pi]$ for random scaling and random rotation. To further improve the performance of our model on the online leaderboard, we fine-tune our trained model on both train and validation set for 12 or 24 epochs with cosine annealing schedule [26] on the SemanticKITTI and nuScenes datasets, respectively, and adopt new Test-Time Augmentation (TTA) strategy as in [22]. Specifically, given an input LiDAR scan $\mathbf{p} \in \mathbb{R}^{N \times 3}$ in a LiDAR point cloud with coordinates (p^x, p^y, p^z). We apply the above four data augmentation transformations for $\mathbf{p}$ in a compound way $\tau_{comp}(\mathbf{p}) = \tau_{trans}(\tau_{flip}(\tau_{scal}(\tau_{rot}(\mathbf{p}))))$. The input scan is augmented into a set of $\{\mathbf{p}, \mathbf{p}_{comp,i}\}$, where i is the index of the augmented samples in the set. After that, the output of the prediction from multiple augmented of input LiDAR scan $\mathbf{p}$ are summed and performed the *argmax* to generate the final predictions at the inference stage. Note that the rotating angles are $\{0, \pm\frac{\pi}{8}, \pm\frac{\pi}{4}, \pm\frac{3\pi}{4}, \pm\frac{7\pi}{8}, \pi\}$ for yaw rotation in test-time.

Table D: Quantitative results of UniSeg and state-of-the-art **LiDAR semantic segmentation** methods on the *test* set of **SemanticKITTI** [3].

Model	mIoU	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic
PointNet [31]	14.6	46.3	1.3	0.3	0.1	0.8	0.2	0.2	0.0	61.6	15.8	35.7	1.4	41.4	12.9	31.0	4.6	17.6	2.4	3.7
PointNet++ [32]	20.1	53.7	1.9	0.2	0.9	0.2	0.9	1.0	0.0	72.0	18.7	41.8	5.6	62.3	16.9	46.5	13.8	30.0	6.0	8.9
Darknet53 [3]	49.9	86.4	24.5	32.7	25.5	22.6	36.2	33.6	4.7	91.8	64.8	74.6	27.9	84.1	55.0	78.3	50.1	64.0	38.9	52.2
RandLA-Net [17]	50.3	94.0	19.8	21.4	42.7	38.7	47.5	48.8	4.6	90.4	56.9	67.9	15.5	81.1	49.7	78.3	60.3	59.0	44.2	38.1
RangeNet++ [28]	52.2	91.4	25.7	34.4	25.7	23.0	38.3	38.8	4.8	91.8	65.0	75.2	27.8	87.4	58.6	80.5	55.1	64.6	47.9	55.9
PolarNet [55]	54.3	93.8	40.3	30.1	22.9	28.5	43.2	40.2	5.6	90.8	61.7	74.4	21.7	90.0	61.3	84.0	65.5	67.8	51.8	57.5
SqueezeSegv3 [47]	55.9	92.5	38.7	36.5	29.6	33.0	45.6	46.2	20.1	91.7	63.4	74.8	26.4	89.0	59.4	82.0	58.7	65.4	49.6	58.9
KPConv [40]	58.8	96.0	32.0	42.5	33.4	44.3	61.5	61.6	11.8	88.8	61.3	72.7	31.6	**95.0**	64.2	84.8	69.2	69.1	56.4	47.4
Salsanext [8]	59.5	91.9	48.3	38.6	38.9	31.9	60.2	59.0	19.4	91.7	63.7	75.8	29.1	90.2	64.2	81.8	63.6	66.5	54.3	62.1
FusionNet [54]	61.3	95.3	47.5	37.7	41.8	34.5	59.5	56.8	11.9	91.8	68.8	77.1	30.8	92.5	69.4	84.5	69.8	68.5	60.4	66.5
KPRNet [19]	63.1	95.5	54.1	47.9	23.6	42.6	65.9	65.0	16.5	93.2	73.9	80.6	30.2	91.7	68.4	85.7	69.8	71.2	58.7	64.1
TORNADONet [12]	63.1	94.2	55.7	48.1	40.0	38.2	63.6	60.1	34.9	89.7	66.3	74.5	28.7	91.3	65.6	85.6	67.0	71.5	58.0	65.9
RangeViT [1]	64.0	95.4	55.8	43.5	29.8	42.1	63.9	58.2	38.1	93.1	70.2	80.0	32.5	92.0	69.0	85.3	70.6	71.2	60.8	64.7
AMVNet [25]	65.3	96.2	59.9	54.2	48.8	45.7	71.0	65.7	11.0	90.1	71.0	75.8	32.4	92.4	69.1	85.6	71.7	69.6	62.7	67.2
GFNet [33]	65.4	96.0	53.2	48.3	31.7	47.3	62.8	57.3	44.7	**93.6**	72.5	**80.8**	31.2	94.0	**73.9**	85.2	71.1	69.3	61.8	68.0
JS3C-Net [50]	66.0	95.8	59.3	52.9	54.3	46.0	69.5	65.4	39.9	88.9	61.9	72.1	31.9	92.5	70.8	84.5	69.8	67.9	60.7	68.7
SPVNAS [39]	66.4	97.3	51.5	50.8	59.8	58.8	65.7	65.2	43.7	90.2	67.6	75.2	16.9	91.3	65.9	86.1	73.4	71.0	64.2	66.9
WaffleIron [30]	67.3	96.5	62.3	64.1	55.2	48.7	70.4	77.8	29.6	90.5	69.5	75.9	24.6	91.8	68.1	85.4	70.8	69.6	62.0	65.2
Cylinder3D [60]	68.9	97.1	67.6	63.8	50.8	58.5	73.7	69.2	48.0	92.2	65.0	77.0	32.3	90.7	66.5	85.6	72.5	69.8	62.4	66.2
AF2S3Net [6]	69.7	94.5	65.4	**86.8**	39.2	41.1	**80.7**	80.4	**74.3**	91.3	68.8	72.5	**53.5**	87.9	63.2	70.2	68.5	53.7	61.5	71.0
RPVNet [48]	70.3	97.6	68.4	68.7	44.2	61.1	75.9	74.4	73.4	93.4	70.3	80.7	33.3	93.5	72.1	86.5	75.1	71.7	64.8	61.4
SDSeg3D [22]	70.4	97.4	58.7	54.2	54.9	65.2	70.2	74.4	52.2	90.9	69.4	76.7	41.9	93.2	71.1	86.1	74.3	71.1	65.4	70.6
GASN [53]	70.7	96.9	65.8	58.0	59.3	61.0	80.4	**82.7**	46.3	89.8	66.2	74.6	30.1	92.3	69.6	87.3	73.0	72.5	66.1	**71.6**
PVKD [16]	71.2	97.0	67.9	69.3	53.5	60.2	75.1	73.5	50.5	91.8	70.9	77.5	41.0	92.4	69.4	86.5	73.8	71.9	64.9	65.8
2DPASS [51]	72.9	97.0	63.6	63.4	61.1	61.5	77.9	81.3	74.1	89.7	67.4	74.7	40.0	93.5	72.9	86.2	73.9	71.0	65.0	70.4
UniSeg (Ours)	**75.2**	**97.9**	**71.9**	75.2	**63.6**	**74.1**	78.9	74.8	60.6	92.6	**74.0**	79.5	46.1	93.4	72.7	**87.5**	**76.3**	**73.1**	**68.3**	68.5

Table E: Quantitative results of UniSeg and state-of-the-art **LiDAR panoptic segmentation** methods on the *test* set of **SemanticKITTI** [3].

Methods	PQ	PQ†	RQ	SQ	PQTh	RQTh	SQTh	PQSt	RQSt	SQSt	mIoU
RangeNet++ [28] + PointPillars [21]	37.1	45.9	47.0	75.9	20.2	25.2	75.2	49.3	62.8	76.5	52.4
LPASD [27]	38.0	47.0	48.2	76.5	25.6	31.8	76.8	47.1	60.1	76.2	50.9
KPConv [40] + PointPillars [21]	44.5	52.5	54.4	80.0	32.7	38.7	81.5	53.1	65.9	79.0	58.8
SalsaNext [8] + PV-RCNN [35]	47.6	55.3	58.6	79.5	39.1	45.9	82.3	53.7	67.9	77.5	58.9
KPConv [40] + PV-RCNN [35]	50.2	57.5	61.4	80.0	43.2	51.4	80.2	55.9	68.7	79.9	62.8
Panoster [10]	52.7	59.9	64.1	80.7	49.9	58.8	83.3	55.1	68.2	78.8	59.9
Panoptic-PolarNet [59]	54.1	60.7	65.0	81.4	53.3	60.6	87.2	54.8	68.1	77.2	59.5
DS-Net [15]	55.9	62.5	66.7	82.3	55.1	62.8	87.2	56.5	69.5	78.7	61.6
EfficientLPS [36]	57.4	63.2	68.7	83.0	53.1	60.5	87.8	60.5	74.6	79.5	61.4
GP-S3Net [34]	60.0	69.0	72.1	82.0	65.0	74.5	86.6	56.4	70.4	78.7	70.8
SCAN [49]	61.5	67.5	72.1	84.5	61.4	69.3	88.1	61.5	74.1	81.8	67.7
Panoptic-PHNet [23]	64.6	70.2	74.9	**85.7**	66.9	73.3	**91.5**	63.0	76.1	81.5	68.4
UniSeg (Ours)	**67.2**	**72.1**	**78.1**	85.5	**67.5**	**75.7**	89.0	**67.0**	**79.8**	**83.0**	**73.8**

Panoptic Head. We follow the instance head design in [58] to predict the instance centers and offsets for each BEV pixel. During the training phase, we encode the ground-truth center map by a 2D Gaussian distribution around each instance's mass center and create an offset map where the offset measures the distance to its corresponding instance's mass center. The size of the center map and the offset map is 480×360. The semantic segmentation predictions are utilized to create the foreground mask to form instance groups. Then, we conduct 2D class-agnostic instance grouping by predicting the center heatmap and offset for each point on the XY-plane. Finally, each instance group is assigned a unique label via majority voting to create the final panoptic segmentation. For the nuScenes panoptic segmentation, we follow [52] to refine the instance segmentation results via the predicted bounding boxes of the TransFusion detector [2]. For the panoptic segmentation evaluation, we evaluate the predicted instance with a minimal point of 30, and 50 as a valid instance on the nuScenes and SemanticKITTI datasets, respectively.

Evaluation Metrics. The definition of Panoptic Quality (PQ) [18], Segmentation Quality (SQ), and Recognition

Table F: Quantitative results of UniSeg and state-of-the-art **LiDAR semantic segmentation** methods on the *test* set of **nuScenes** [4].

Model	mIoU	barrier	bicycle	bus	car	construction	motorcycle	pedestrian	traffic-cone	trailer	truck	driveable	other	sidewalk	terrain	manmade	vegetation
PolarNet [55]	69.4	72.2	16.8	77.0	86.5	51.1	69.7	64.8	54.1	69.7	63.5	96.6	67.1	77.7	72.1	87.1	84.5
JS3C-Net [50]	73.6	80.1	26.2	87.8	84.5	55.2	72.6	71.3	66.3	76.8	71.2	96.8	64.5	76.9	74.1	87.5	86.1
PMF [61]	77.0	82.0	40.0	81.0	88.0	64.0	79.0	80.0	76.0	81.0	67.0	97.0	68.0	78.0	74.0	90.0	88.0
Cylinder3D [60]	77.2	82.8	29.8	84.3	89.4	63.0	79.3	77.2	73.4	84.6	69.1	97.7	70.2	80.3	75.5	90.4	87.6
AMVNet [25]	77.3	80.6	32.0	81.7	88.9	67.1	84.3	76.1	73.5	84.9	67.3	97.5	67.4	79.4	75.5	91.5	88.7
SPVCNN [39]	77.4	80.0	30.0	91.9	90.8	64.7	79.0	75.6	70.9	81.0	74.6	97.4	69.2	80.0	76.1	89.3	87.1
AF2S3Net [6]	78.3	78.9	52.2	89.9	84.2	77.4	74.3	77.3	72.0	83.9	73.8	97.1	66.5	77.5	74.0	87.7	86.8
2D3DNet [11]	80.0	83.0	59.4	88.0	85.1	63.7	84.4	82.0	76.0	84.8	71.9	96.9	67.4	79.8	76.0	**92.1**	89.2
GASN [53]	80.4	85.5	43.2	90.5	**92.1**	64.7	86.0	83.0	73.3	83.9	75.8	97.0	71.0	**81.0**	**77.7**	91.6	**90.2**
2DPASS [51]	80.8	81.7	55.3	92.0	91.8	73.3	86.5	78.5	72.5	84.7	75.5	97.6	69.1	79.9	75.5	90.2	88.0
LidarMultiNet [52]	81.4	80.4	48.4	**94.3**	90.0	71.5	87.2	**85.2**	80.4	**86.9**	74.8	**97.8**	67.3	80.7	76.5	**92.1**	89.6
UniSeg (Ours)	**83.5**	**85.9**	**71.2**	92.1	91.6	**80.5**	**88.0**	80.9	76.0	86.3	**76.7**	97.7	**71.8**	80.7	76.7	91.3	88.8

Table G: Quantitative results of UniSeg and state-of-the-art **LiDAR semantic segmentation** methods on the *val* set of **Waymo Open Dataset** [37]. Methods with * are our implementations.

Model	mIoU	car	truck	bus	other vehicle	motorcyclist	bicyclist	pedestrian	sign	traffic light	pole	construction	bicycle	motorcycle	building	vegetation	tree trunk	curb	road	lane marker	other ground	walkable	sidewalk
P-Transformer* [56]	63.3	93.1	58.8	61.4	25.4	0.0	67.9	85.5	72.3	36.2	71.4	66.4	58.7	54.3	93.7	90.0	64.7	65.2	90.4	48.2	42.8	74.5	71.7
Cylinder3D* [60]	66.0	**95.1**	59.6	74.1	28.7	**2.4**	62.3	86.8	71.5	33.6	73.4	65.2	62.0	76.5	95.1	**91.0**	66.6	65.5	92.3	49.9	47.1	**79.0**	75.1
SPVCNN* [39]	67.4	94.3	59.8	78.5	27.5	0.0	70.8	87.8	74.9	39.2	74.4	69.5	70.4	79.4	94.8	90.8	66.9	66.6	91.7	50.9	43.9	77.2	72.7
UniSeg (Ours)	**69.6**	94.4	**60.4**	79.6	**40.6**	0.0	**73.2**	**89.0**	**75.7**	**43.3**	76.1	**70.2**	**75.5**	**80.8**	**95.2**	91.0	**68.2**	**68.7**	**92.6**	**53.9**	**48.3**	78.8	**75.8**

Quality (RQ) is given as follows:

$$
\text{PQ} = \underbrace{\frac{\sum_{(i,j)\in TP} \text{IoU}(i,j)}{|TP|}}_{\text{SQ}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{RQ}}.
$$

(1)

The aforementioned three metrics are also calculated separately on *things* and *stuff* classes which produce PQ^{Th}, SQ^{Th}, RQ^{Th}, and PQ^{St}, SQ^{St}, RQ^{St}. In addition, we report $\text{PQ}^{\dagger}$ which is defined by swapping PQ of each *stuff* class to its IoU and then averaging over all classes.

C. Additional Quantitative Result

We provide a more comprehensive comparison between UniSeg and competitive LiDAR segmentation networks. Table D shows the class-wise IoU scores of different LiDAR semantic segmentation methods on the *test set* of SemanticKITTI [3]. Among all the LiDAR segmentation algorithms, UniSeg achieves compelling results. Table E shows the PQ, RQ, SQ, mIoU scores of different LiDAR panoptic segmentation methods on the *test set* of SemanticKITTI [3]. We can observe a clear advantage of UniSeg over other solutions. Table F shows the class-wise IoU scores of different LiDAR semantic segmentation methods on the *test set* of nuScenes [9, 4].

UniSeg yields high mIoU scores than the SoTA solution of LidarMultiNet [52], which demonstrates again the advantage of UniSeg. In addition, we provide detailed performance on the Waymo Open [37] *val set* in Table G. It shows UniSeg obtains higher efficacy.

D. Additional Qualitative Result

We provide more visual comparisons of UniSeg with baseline algorithm (single modal) in Fig. A, Fig. B, and Fig. C on the validation set of SemanticKITTI [3], nuScenes [9, 4] and Waymo Open [37], respectively. To highlight the differences in the error map, the correct/incorrect predictions are painted in gray/red, respectively. For the ground truth, different colors represent different classes. The single-modal baseline has higher prediction errors than our UniSeg, especially on small objects, *e.g.*, pedestrians. For example, in Fig. A, the baseline mistakenly predicts the person and fence and has higher prediction errors on the road boundaries. By contrast, UniSeg makes much better predictions on both person and fence, as well as the road boundaries, which is attributed to the comprehensive information provided by camera images and all views of the point cloud. In a nutshell, UniSeg can make more accurate point-wise predictions regardless of the distance and point density variation than the baseline.
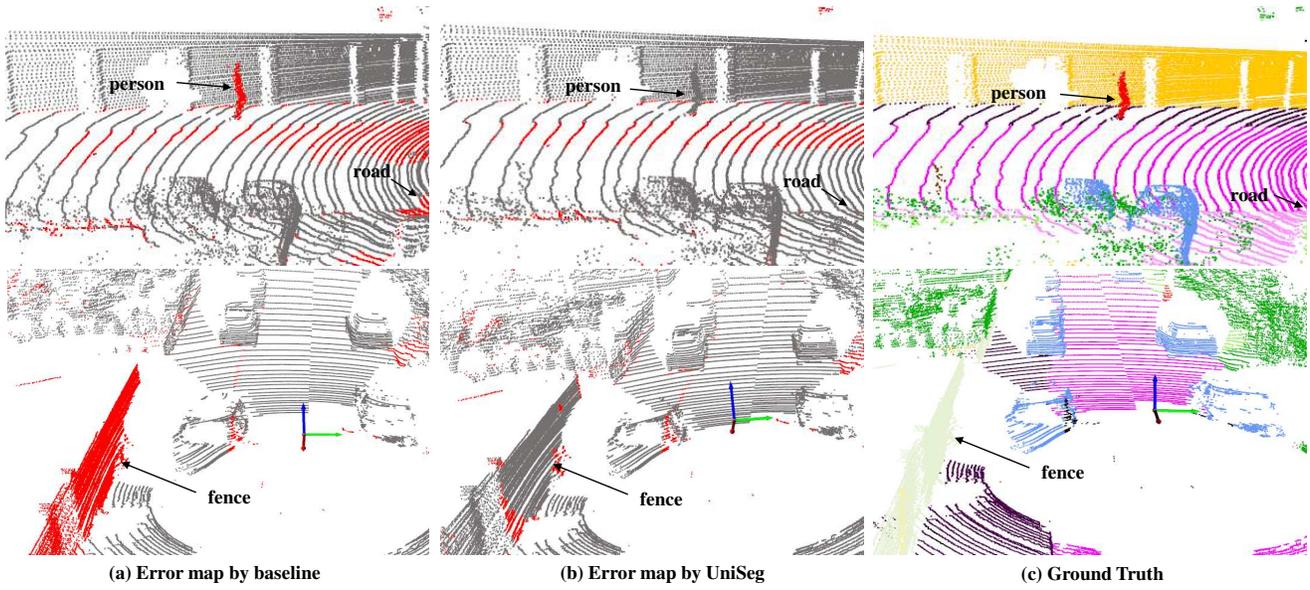
(a) Error map by baseline   (b) Error map by UniSeg   (c) Ground Truth

Figure A: Qualitative results of UniSeg on the SemanticKITTI validation set.

(a) Error map by baseline   (b) Error map by UniSeg   (c) Ground Truth
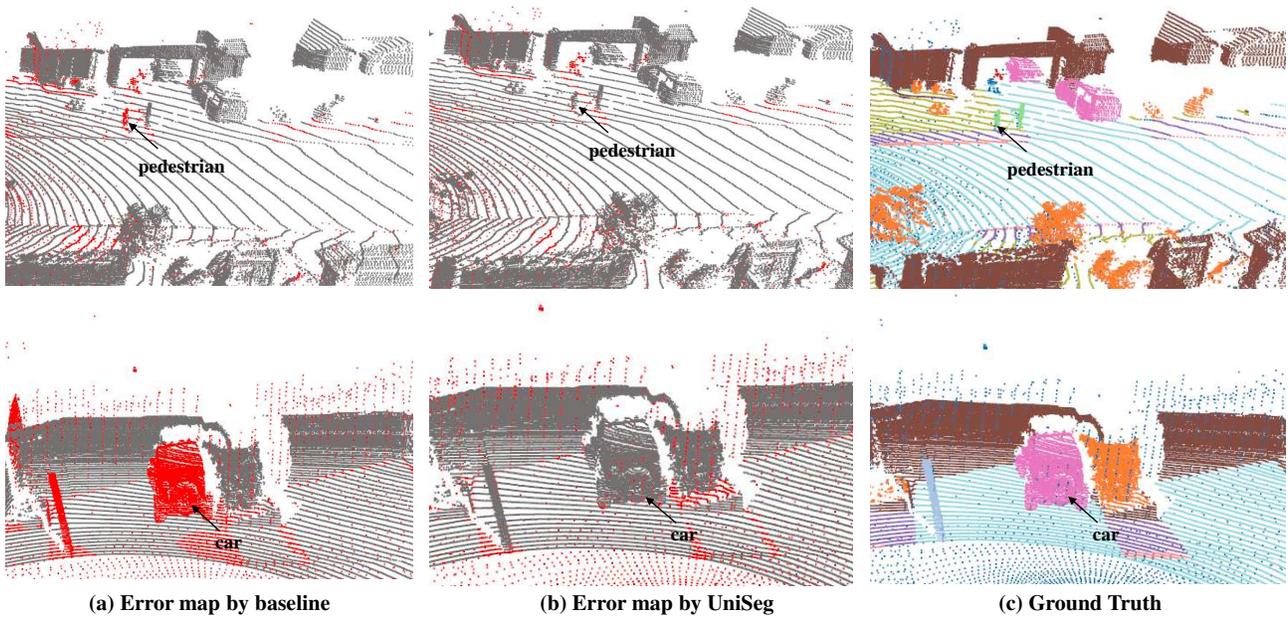
Figure B: Qualitative results of UniSeg on the Waymo Open validation set.

(a) Error map by baseline   (b) Error map by UniSeg   (c) Ground Truth
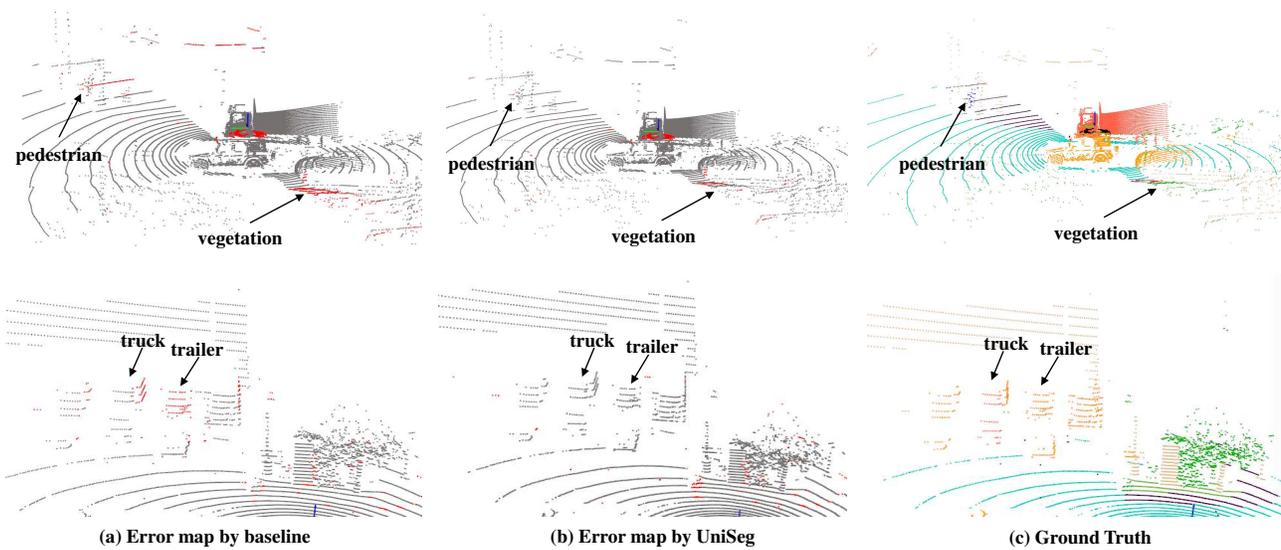
Figure C: Qualitative results of UniSeg on the nuScenes validation set.

References

[1] Angelika Ando, Spyros Gidaris, Andrei Bursuc, Gilles Puy, Alexandre Boulch, and Renaud Marlet. Rangevit: Towards vision transformers for 3d semantic segmentation in autonomous driving. *arXiv preprint arXiv:2301.10222*, 2023. 5

[2] Xuyang Bai, Zeyu Hu, Xinge Zhu, Qingqiu Huang, Yilun Chen, Hongbo Fu, and Chiew-Lan Tai. TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1090–1099, 2022. 5

[3] 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 *IEEE International Conference on Computer Vision*, pages 9297–9307, 2019. 3, 4, 5, 6

[4] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, pages 11621–11631, 2020. 3, 4, 6

[5] Hui-Xian Cheng, Xian-Feng Han, and Guo-Qiang Xiao. CENet: Toward Concise and Efficient LiDAR Semantic Segmentation for Autonomous Driving. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*, pages 01–06. IEEE, 2022. 1, 3

[6] Ran Cheng, Ryan Razani, Ehsan Taghavi, Enxu Li, and Bingbing Liu. (af)2-S3Net: Attentive Feature Fusion with Adaptive Feature Selection for Sparse Semantic Segmentation Network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 12547–12556, 2021. 5, 6

[7] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3075–3084, 2019. 1, 3, 4

[8] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. SalsaNext: Fast, Uncertainty-Aware Semantic Segmentation of lidar Point Clouds. In *International Symposium on Visual Computing*, pages 207–222. Springer, 2020. 1, 3, 4, 5

[9] Whye Kit Fong, Rohit Mohan, Juana Valeria Hurtado, Lubing Zhou, Holger Caesar, Oscar Beijbom, and Abhinav Valada. Panoptic nuScenes: A Large-Scale Benchmark for LiDAR Panoptic Segmentation and Tracking. *IEEE Robotics and Automation Letters*, 7:3795–3802, 2022. 4, 6

[10] Stefano Gasperini, Mohammad-Ali Nikouei Mahani, Alvaro Marcos-Ramiro, Nassir Navab, and Federico Tombari. Panoster: End-to-end panoptic segmentation of lidar point clouds. *IEEE Robotics and Automation Letters*, 6(2):3216–3223, 2021. 5

[11] Kyle Genova, Xiaoqi Yin, Abhijit Kundu, Caroline Pantofaru, Forrester Cole, Avneesh Sud, Brian Brewington, Brian Shucker, and Thomas Funkhouser. Learning 3d semantic segmentation with only 2d image supervision. In *2021 International Conference on 3D Vision (3DV)*, pages 361–372. IEEE, 2021. 6

[12] Martin Gerdzhev, Ryan Razani, Ehsan Taghavi, and Liu Bingbing. Tornado-net: Multiview Total Variation Semantic Segmentation with Diamond Inception Module. In *IEEE International Conference on Robotics and Automation*, pages 9543–9549. IEEE, 2021. 5

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 4

[14] Fangzhou Hong, Lingdong Kong, Hui Zhou, Xinge Zhu, Hongsheng Li, and Ziwei Liu. Unified 3d and 4d panoptic segmentation via dynamic shifting network. 3

[15] Fangzhou Hong, Hui Zhou, Xinge Zhu, Hongsheng Li, and Ziwei Liu. LiDAR-Based Panoptic Segmentation via Dynamic Shifting Network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13090–13099, June 2021. 1, 3, 5

[16] Yuenan Hou, Xinge Zhu, Yuexin Ma, Chen Change Loy, and Yikang Li. Point-to-Voxel Knowledge Distillation for LiDAR Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8479–8488, 2022. 5

[17] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient Semantic Segmentation of Large-Scale Point Clouds. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 11108–11117, 2020. 5

[18] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9404–9413, 2019. 5

[19] Deyvid Kochanov, Fatemeh Karimi Nejadasl, and Olaf Booij. KPRNet: Improving Projection-based Lidar Semantic Segmentation. *arXiv preprint arXiv:2007.12668*, 2020. 5

[20] Lingdong Kong, Jiawei Ren, Liang Pan, and Ziwei Liu. Lasermix for Semi-Supervised LiDAR Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2023. 1, 3, 4

[21] 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/CVF conference on computer vision and pattern recognition*, pages 12697–12705, 2019. 5

[22] Jiale Li, Hang Dai, and Yong Ding. Self-distillation for robust lidar semantic segmentation in autonomous driving. In *European Conference on Computer Vision*, pages 659–676. Springer, 2022. 4, 5

[23] Jinke Li, Xiao He, Yang Wen, Yuan Gao, Xiaoqiang Cheng, and Dan Zhang. Panoptic-PHNet: Towards Real-Time and High-Precision LiDAR Panoptic Segmentation via Clustering Pseudo Heatmap. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 11809–11818, 2022. 1, 5

[24] Rong Li, Anh-Quan Cao, and Raoul de Charette. Coarse3d: Class-prototypes for contrastive learning in weakly-supervised 3d point cloud segmentation–supplementary material. 2022. 2

[25] Venice Erin Liong, Thi Ngoc Tho Nguyen, Sergi Widjaja, Dhananjai Sharma, and Zhuang Jie Chong. Amvnet: Assertion-based multi-view fusion network for lidar semantic segmentation. *arXiv preprint arXiv:2012.04934*, 2020. 5, 6

[26] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 4

[27] Andres Milioto, Jens Behley, Chris McCool, and Cyrill Stachniss. Lidar panoptic segmentation for autonomous driving. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8505–8512, 2020. 5

[28] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. RangeNet++: Fast and Accurate Lidar Semantic Segmentation. In *IEEE International Conference on Intelligent Robots and Systems*, pages 4213–4220, 2019. 1, 3, 5

[29] Alexey Nekrasov, Jonas Schult, Or Litany, Bastian Leibe, and Francis Engelmann. Mix3d: Out-of-context data augmentation for 3d scenes. In *2021 International Conference on 3D Vision (3DV)*, pages 116–125. IEEE, 2021. 1, 3

[30] Gilles Puy, Alexandre Boulch, and Renaud Marlet. Using a waffle iron for automotive point cloud semantic segmentation. *arxiv:2301.10100*, 2023. 5

[31] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017. 5

[32] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *Advances in Neural Information Processing Systems*, 30, 2017. 5

[33] Haibo Qiu, Baosheng Yu, and Dacheng Tao. Gfnet: Geometric flow network for 3d point cloud semantic segmentation. *Transactions on Machine Learning Research (TMLR)*, 2022. 5

[34] Ryan Razani, Ran Cheng, Enxu Li, Ehsan Taghavi, Yuan Ren, and Liu Bingbing. Gp-s3net: Graph-based panoptic sparse semantic segmentation network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16076–16085, 2021. 5

[35] 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. 5

[36] Kshitij Sirohi, Rohit Mohan, Daniel Büscher, Wolfram Burgard, and Abhinav Valada. Efficientlps: Efficient lidar panoptic segmentation. *IEEE Transactions on Robotics*, 2021. 5

[37] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2446–2454, 2020. 4, 6

[38] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *International Conference on Machine Learning*, 2019. 1

[39] Haotian Tang, Zhijian Liu, Shengyu Zhao, Yujun Lin, Ji Lin, Hanrui Wang, and Song Han. Searching Efficient 3D Architectures with Sparse Point-Voxel Convolution. In *European Conference on Computer Vision*, pages 685–702. Springer, 2020. 1, 3, 4, 5, 6

[40] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and Deformable Convolution for Point Clouds. In *IEEE International Conference on Computer Vision*, pages 6411–6420, 2019. 5

[41] Ozan Unal, Dengxin Dai, and Luc Van Gool. Scribble-supervised lidar semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2697–2707, 2022. 4

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 3

[43] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019. 3

[44] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. Squeezeseg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D liDAR Point Cloud. In *IEEE International Conference on Robotics and Automation*, pages 1887–1893, 2018. 1, 3

[45] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *IEEE International Conference on Robotics and Automation*, pages 4376–4382, 2019. 1, 3

[46] Aoran Xiao, Jiaxing Huang, Dayan Guan, Kaiwen Cui, Shijian Lu, and Ling Shao. Polarmix: A general data augmentation technique for lidar point clouds. *arXiv preprint arXiv:2208.00223*, 2022. 1, 3, 4

[47] Chenfeng Xu, Bichen Wu, Zining Wang, Wei Zhan, Peter Vajda, Kurt Keutzer, and Masayoshi Tomizuka. Squeezesegv3: Spatially-Adaptive Convolution for Efficient Point-Cloud Segmentation. In *European Conference on Computer Vision*, pages 1–19. Springer, 2020. 5

[48] Jianyun Xu, Ruixiang Zhang, Jian Dou, Yushi Zhu, Jie Sun, and Shiliang Pu. RPVnet: A Deep and Efficient Range-Point-Voxel Fusion Network for Lidar Point Cloud Segmentation. In *IEEE International Conference on Computer Vision*, pages 16024–16033, October 2021. 1, 3, 5

[49] Shuangjie Xu, Rui Wan, Maosheng Ye, Xiaoyi Zou, and Tongyi Cao. Sparse cross-scale attention network for efficient lidar panoptic segmentation. *arXiv preprint arXiv:2201.05972*, 2022. 3, 5

[50] Xu Yan, Jiantao Gao, Jie Li, Ruimao Zhang, Zhen Li, Rui Huang, and Shuguang Cui. Sparse single sweep lidar point cloud segmentation via learning contextual shape priors from

scene completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 3101–3109, 2021. 5, 6

[51] Xu Yan, Jiantao Gao, Chaoda Zheng, Chao Zheng, Ruimao Zhang, Shuguang Cui, and Zhen Li. 2DPASS: 2D Priors Assisted Semantic Segmentation on LiDAR Point Clouds. In *European Conference on Computer Vision*, 2022. 5, 6

[52] Dongqiangzi Ye, Zixiang Zhou, Weijia Chen, Yufei Xie, Yu Wang, Panqu Wang, and Hassan Foroosh. LidarMultiNet: Towards a Unified Multi-task Network for LiDAR Perception. *arXiv preprint arXiv:2209.09385*, 2022. 1, 5, 6

[53] Maosheng Ye, Rui Wan, Shuangjie Xu, Tongyi Cao, and Qifeng Chen. Efficient point cloud segmentation with geometry-aware sparse networks. In *European Conference on Computer Vision*, pages 196–212. Springer, 2022. 5, 6

[54] Feihu Zhang, Jin Fang, Benjamin Wah, and Philip Torr. Deep FusionNet for Point Cloud Semantic Segmentation. In *European Conference on Computer Vision*, pages 644–663. Springer, 2020. 5

[55] Yang Zhang, Zixiang Zhou, Philip David, Xiangyu Yue, Zerong Xi, Boqing Gong, and Hassan Foroosh. Polarnet: An Improved Grid Representation for Online Lidar Point Clouds Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 9601–9610, 2020. 1, 2, 3, 5, 6

[56] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point Transformer. In *IEEE International Conference on Computer Vision*, pages 16259–16268, 2021. 1, 3, 6

[57] Yiming Zhao, Lin Bai, and Xinming Huang. FIDNet: LiDAR Point Cloud Semantic Segmentation with Fully Interpolation Decoding. In *2021 IEEE International Conference on Intelligent Robots and Systems (IROS)*, pages 4453–4458. IEEE, 2021. 1, 3

[58] Zixiang Zhou, Yang Zhang, and Hassan Foroosh. Panoptic-PolarNet: Proposal-free LiDAR Point Cloud Panoptic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 13194–13203, 2021. 1, 2, 3, 5

[59] Zixiang Zhou, Yang Zhang, and Hassan Foroosh. Panoptic-polarnet: Proposal-free lidar point cloud panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13194–13203, 2021. 5

[60] Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. Cylindrical and Asymmetrical 3D Convolution Networks for Lidar Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 9939–9948, 2021. 1, 3, 5, 6

[61] Zhuangwei Zhuang, Rong Li, Kui Jia, Qicheng Wang, Yuanqing Li, and Mingkui Tan. Perception-Aware Multi-Sensor Fusion for 3D LiDAR Semantic Segmentation. In *IEEE International Conference on Computer Vision*, pages 16280–16290, 2021. 6