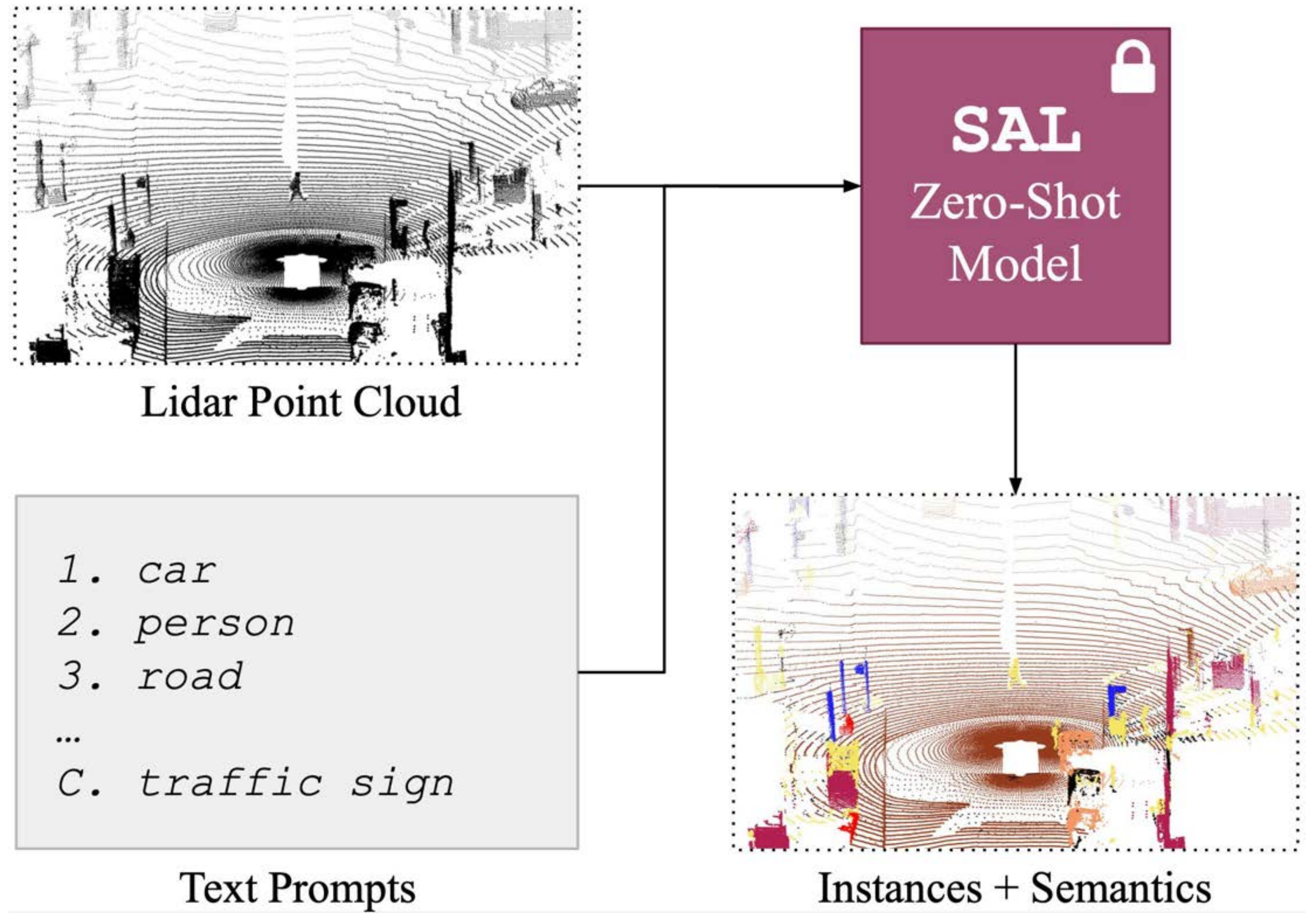
Better Call SAL: Towards Learning To Segment Anything in Lidar

Aljoša Ošep*, Tim Meinhardt*, Francesco Ferroni, Neehar Peri, Deva Ramanan, and Laura Leal-Taixé

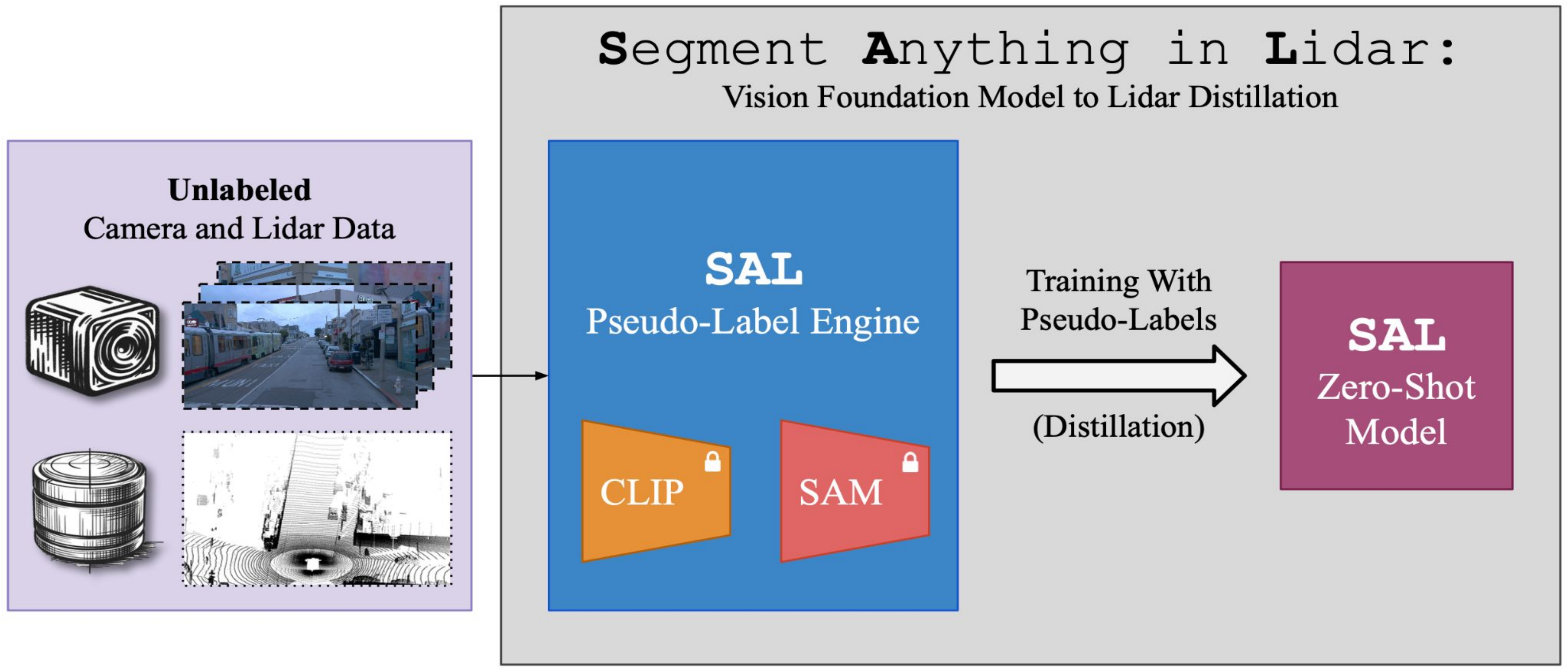
* Equal contribution

The Goal

Segment and classify **any** object



How To Build Zero-Shot Panoptic Segmentation Models For Lidar?

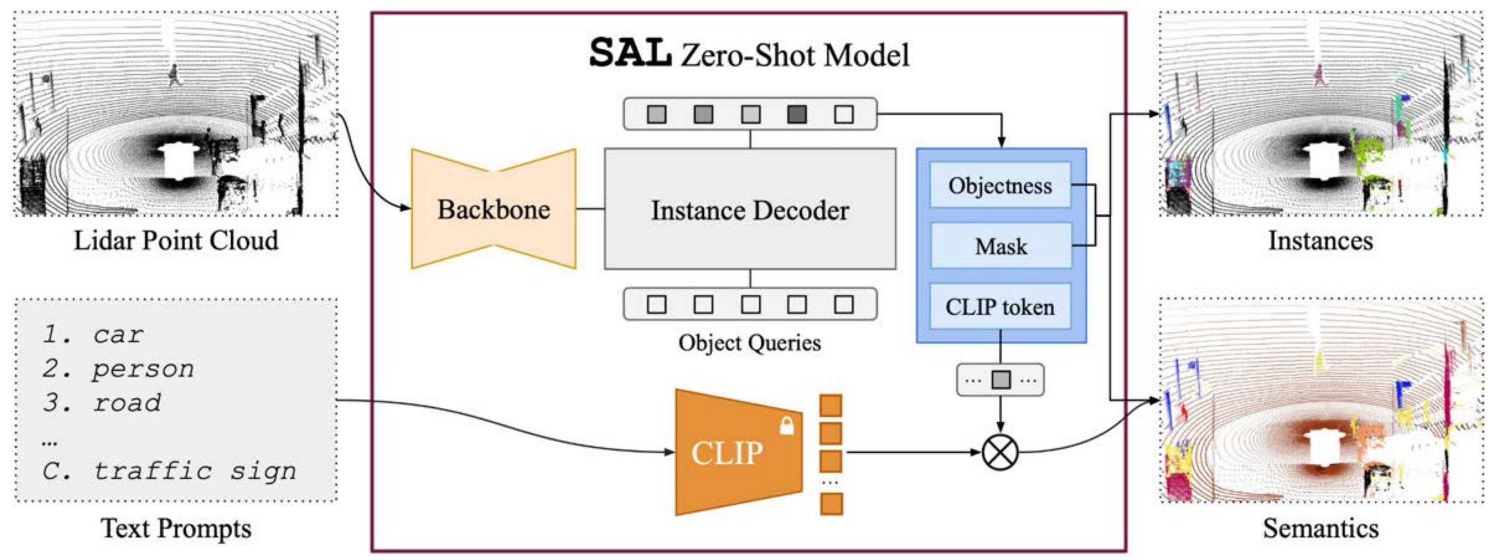


The Zero-Shot Model

Class-agnostic segmentation

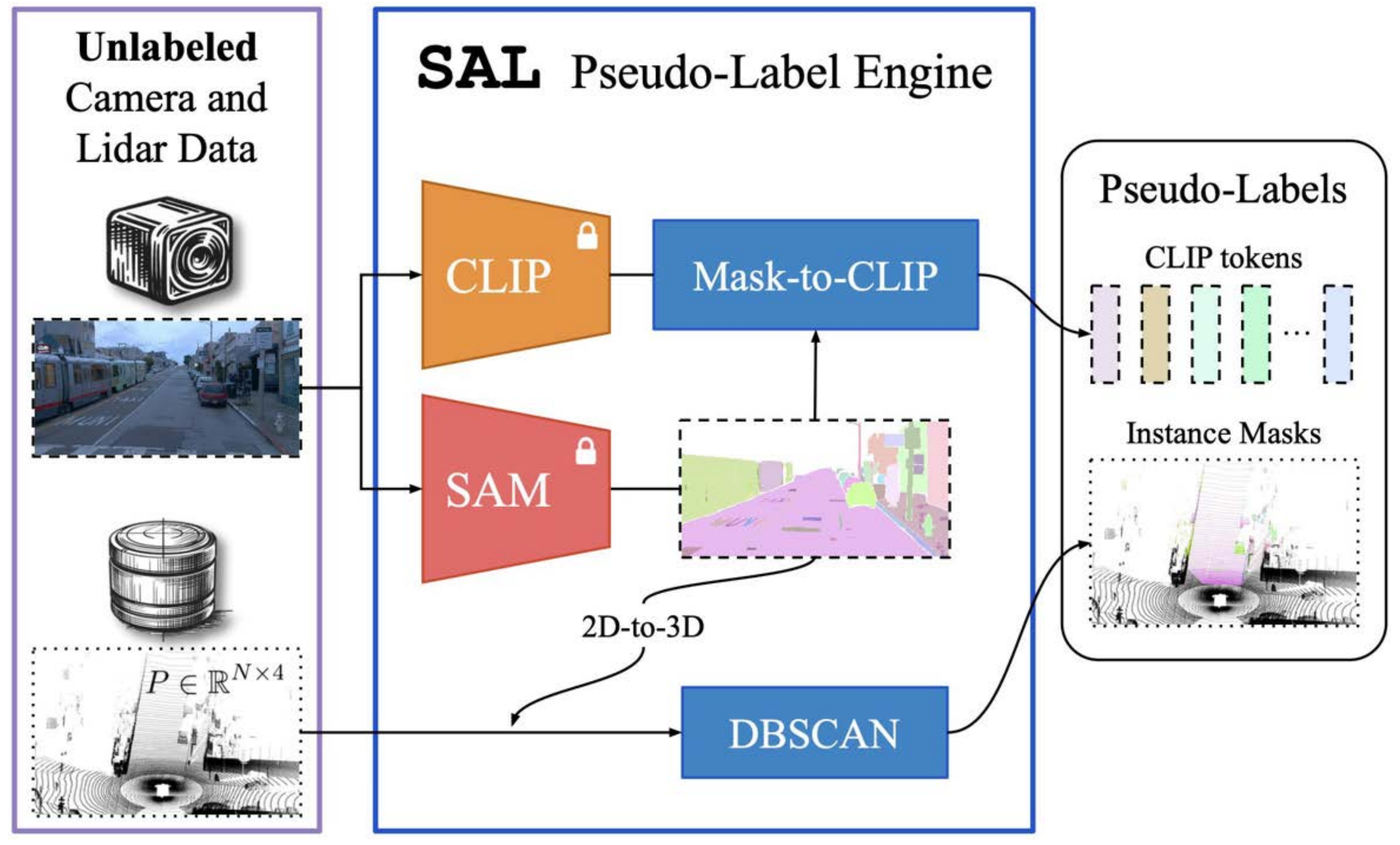
+

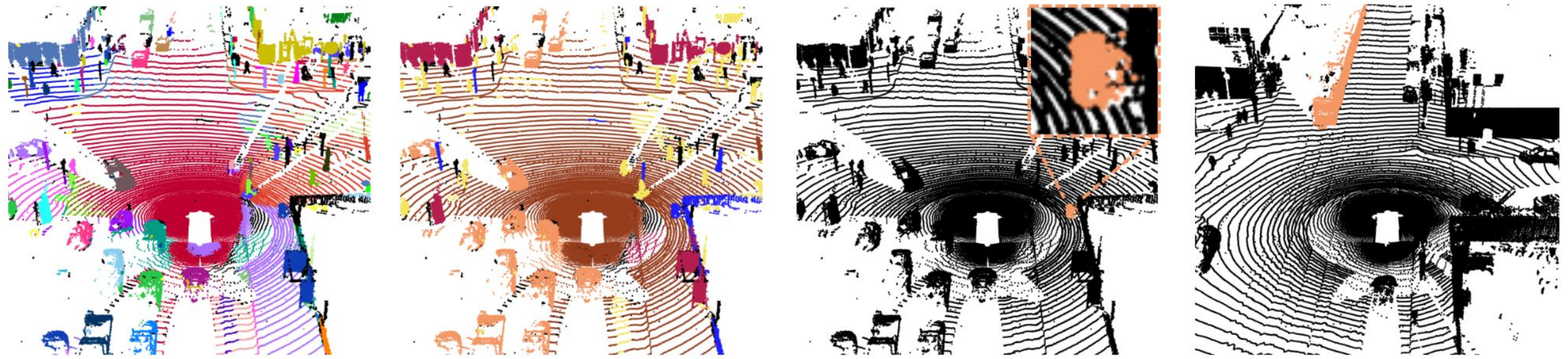
Zero-shot classification



The Pseudo-Label Engine

2D foundation models to **3D** labels





Class-agnostic instance segmentation | Text prompts: {car, building, …} | Text prompts: {trash bin} | Text prompts: {streetcar}

➔ The **SAL** model performs class-agnostic instance segmentation and zero-shot classification via text prompting.

➔ Our training pipeline does not utilize any Lidar or image ground truth labels.

➔ The pseudo-label engine generates 3D pseudo-labels via 2D foundation models.

Key Results

Pseudo-labels vs. zero-shot model

Method	Frust. Eval	Image Feat.	Default classes PQ	SQ	PQ^{Th}	PQ^{St}	mIoU	Super classes PQ	SQ	PQ^{Th}	PQ^{St}	mIoU
Class-agnostic Segmentation (Semantic Oracle)												
SAM	✓	✓	46.0	72.1	49.7	43.4	–	–	–	–	–	–
SAM+Erosion	✓	✓	42.2	69.4	45.6	39.6	–	–	–	–	–	–
SAM+DBS (filter)	✓	✓	46.7	70.3	76.8	24.8	–	–	–	–	–	–
SAM+DBS (replace)	✓	✓	48.7	73.7	53.1	45.4	–	–	–	–	–	–
SAL	✓	✗	70.7	81.9	75.4	67.3	–	–	–	–	–	–
Zero-Shot Lidar Panoptic Segmentation												
SAM+DBS+CLIP	✓	✓	27.5	71.5	31.7	24.5	30.6	51.1	77.5	71.2	41.0	54.3
SAL	✓	✗	33.1	71.4	22.8	40.5	33.5	63.9	84.2	88.3	51.7	66.4
SAM+DBS+CLIP	✗	✗	8.2	56.4	18.6	0.6	7.5	11.5	47.6	0.0	17.3	11.2
SAL	✗	✗	24.8	66.8	17.4	30.2	28.7	48.5	78.8	80.4	32.6	52.8

Lidar panoptic segmentation benchmarks

Dataset	Method	Supervision	Default classes PQ	RQ	SQ	PQ^{Th}	PQ^{St}	mIoU	Super classes PQ	RQ	SQ	mIoU
SemanticKITTI	DS-Net [26]	Full	57.7	68.0	77.6	61.8	54.8	63.5	–	–	–	–
SemanticKITTI	PolarSeg [98]	Full	59.1	70.2	78.3	65.7	54.3	64.5	–	–	–	–
SemanticKITTI	EfficientLPS [74]	Full	59.2	69.8	75.0	58.0	60.9	64.9	–	–	–	–
SemanticKITTI	GP-S3Net [70]	Full	63.3	75.9	81.4	70.2	58.3	73.0	–	–	–	–
SemanticKITTI	MaskPLS [45]	Full	59.8	69.0	76.3	–	–	–	78.4	87.1	88.2	84.5
SemanticKITTI	SAL	Full	59.5	69.2	75.7	62.3	57.4	63.8	81.7	90.0	89.2	85.9
SemanticKITTI	SAL	Zero-shot	24.8	32.3	66.8	17.4	30.2	28.7	48.5	59.4	78.8	52.8
nuScenes	PHNet [36]	Full	74.7	84.2	88.2	74.0	75.9	79.7	–	–	–	–
nuScenes	DS-Net [26]	Full	51.2	59.0	86.1	38.4	72.3	73.5	–	–	–	–
nuScenes	GP-S3Net [70]	Full	61.0	72.0	84.1	56.0	66.0	75.8	–	–	–	–
nuScenes	EfficientLPS [74]	Full	62.0	73.9	83.4	56.8	70.6	65.6	–	–	–	–
nuScenes	PolarSeg [98]	Full	63.4	75.3	83.9	59.2	70.4	66.9	–	–	–	–
nuScenes	MaskPLS [45]	Full	57.7	66.0	71.8	64.4	52.2	62.5	71.5	81.0	86.2	80.6
nuScenes	SAL	Full	70.5	80.8	85.9	79.4	61.7	72.8	74.2	82.7	87.1	84.0
nuScenes	SAL	Zero-shot	38.4	47.8	77.2	47.5	29.2	33.9	52.6	63.5	77.3	52.6