HYUNDAI NGV | Open R&D Seminar

Robust & Data-Efficient Learning for 3D Scene Understanding

Speaker: Lingdong Kong

Dec. 18th, 2023

Topic

1. Overview & Background
2. Data-Efficient 3D Perception
3. Robustness in 3D
4. Segment Any Point Cloud Sequences

Topic

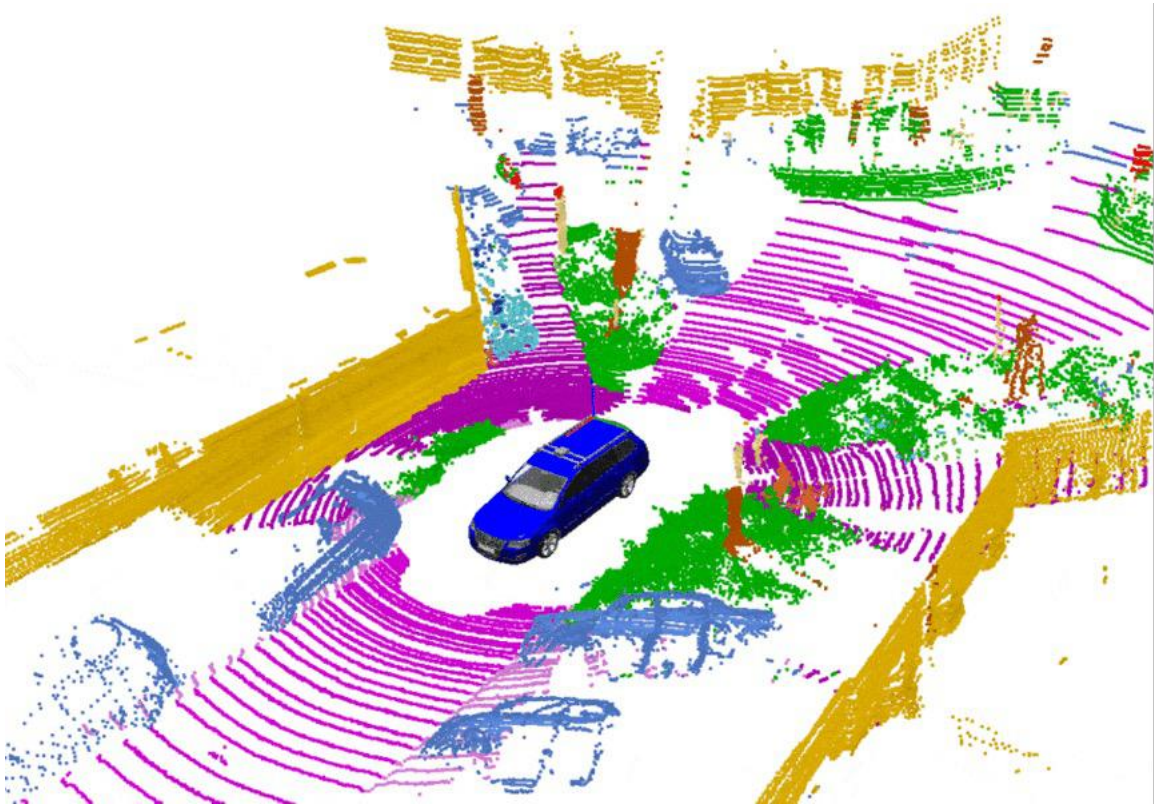
1. Overview & Background
2. Data-Efficient 3D Perception
3. Robustness in 3D
4. Segment Any Point Cloud Sequences

3D Scene Understanding

Indoor

Dai, et al. ScanNet V2. [CVPR 2017]

Outdoor



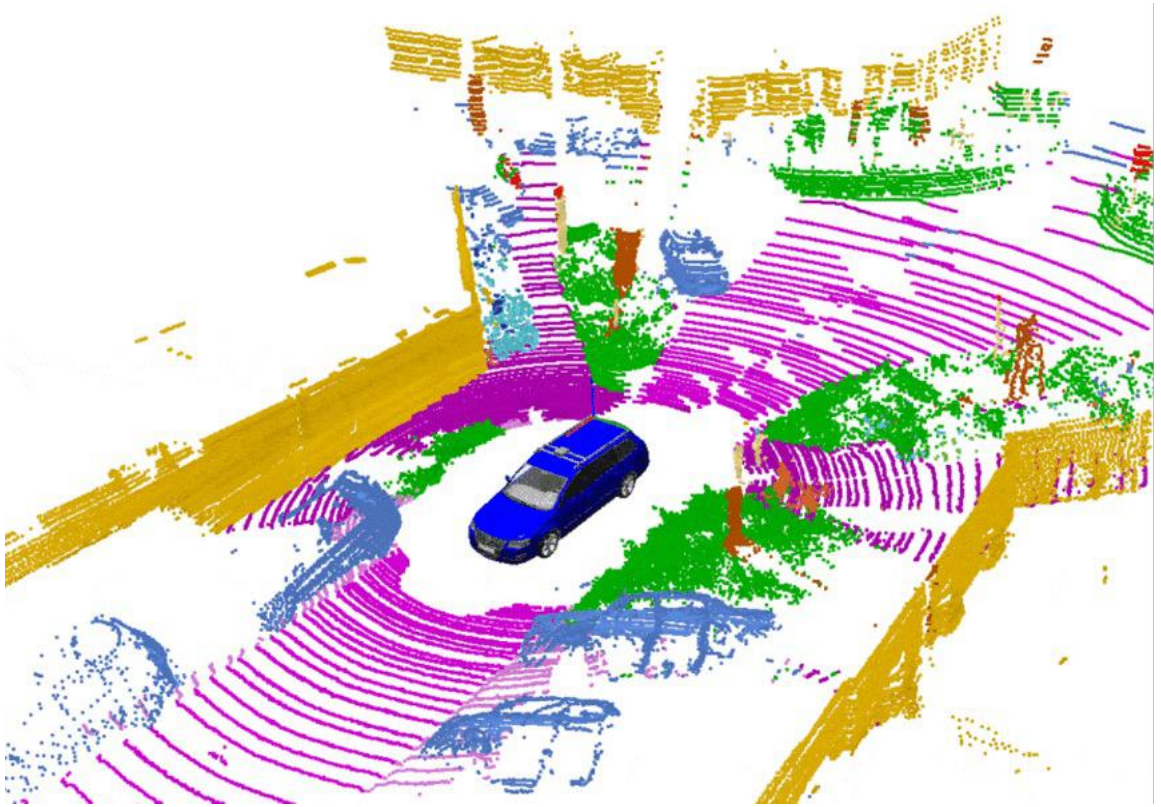
Behley, et al. SemanticKITTI. [ICCV 2019]

3D Scene Understanding

Main Features

- Autonomous driving perception
- Larger in scale (>120k points per scan)
- Sparser, more diverse
- Rich semantic categories
- Different sensor setups (32, 64, 128)
- Real-world distribution
- Standard benchmarks: nuScenes, SemanticKITTI, Waymo Open, etc.

Outdoor



Behley, et al. SemanticKITTI. [ICCV 2019]

MMDetection3D

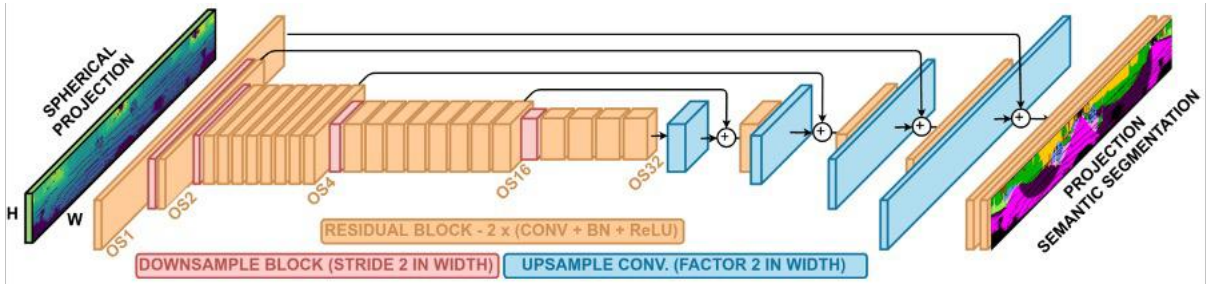
- MMDetection3D is an open-source toolbox based on PyTorch
- MMDetection3D is the next-generation platform for general 3D object detection
- As well as [NEW] 3D Semantic Segmentation

3D Object Detection	Monocular 3D Object Detection	Multi-modal 3D Object Detection	3D Semantic Segmentation

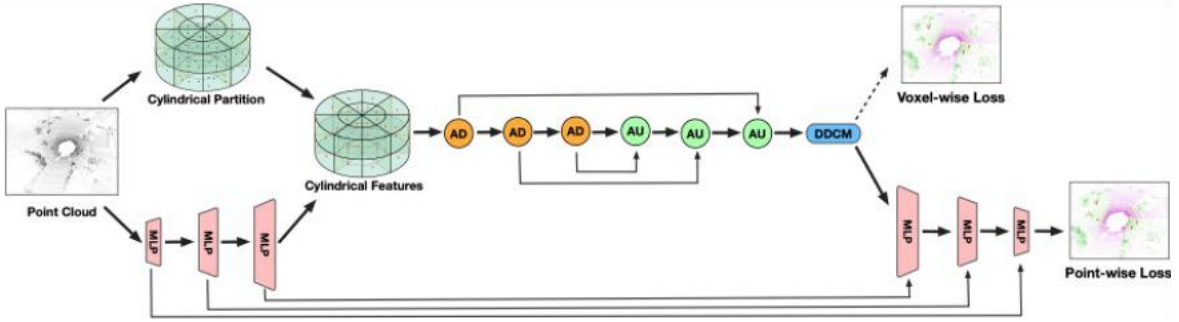
• **Outdoor**
  • SECOND (Sensor'2018)
  • PointPillars (CVPR'2019)
  • SSN (ECCV'2020)
  • 3DSSD (CVPR'2020)
  • SA-SSD (CVPR'2020)
  • PointRCNN (CVPR'2019)
  • Part-A2 (TPAMI'2020)
	• **Outdoor**
	  • ImVoxelNet (WACV'2022)
	  • SMOKE (CVPRW'2020)
	  • FCOS3D (ICCVW'2021)
	  • PGD (CoRL'2021)
	  • MonoFlex (CVPR'2021)
	• **Indoor**
	  • ImVoxelNet (WACV'2022)
		• **Outdoor**
		  • MVXNet (ICRA'2019)
		• **Indoor**
		  • ImVoteNet (CVPR'2020)
			• **Outdoor**
			  • MinkUNet (CVPR'2019)
			  • SPVCNN (ECCV'2020)
			  • Cylinder3D (CVPR'2021)
			• **Indoor**
			  • PointNet++ (NeurIPS'2017)
			  • PAConv (CVPR'2021)
			  • DGCNN (TOG'2019)

The MMDetection3D Codebase: https://github.com/open-mmlab/mmdetection3d

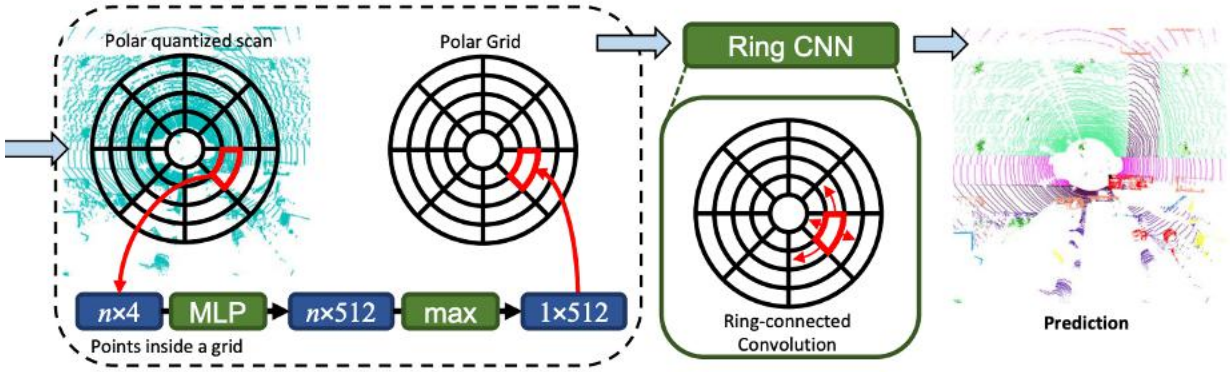
MMDetection3D v1.1.1



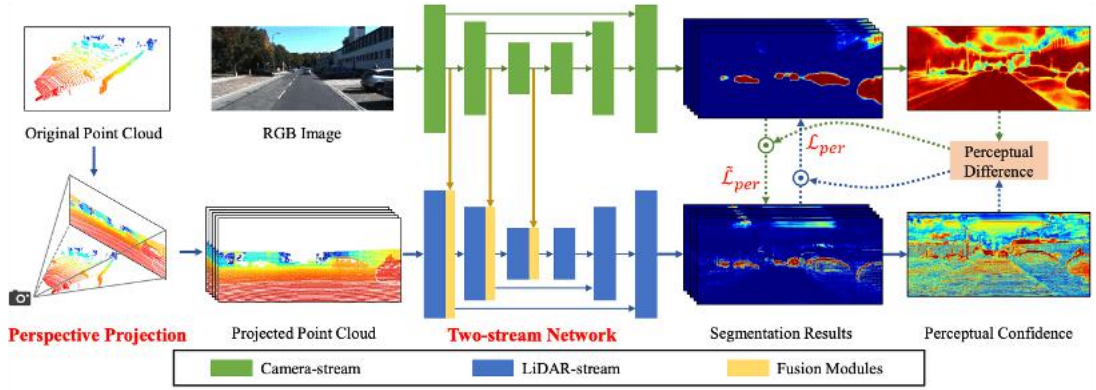
Milioto, et al. RangeNet++. [IROS 2019]



Zhu, et al. Cylinder3D. [CVPR 2021]

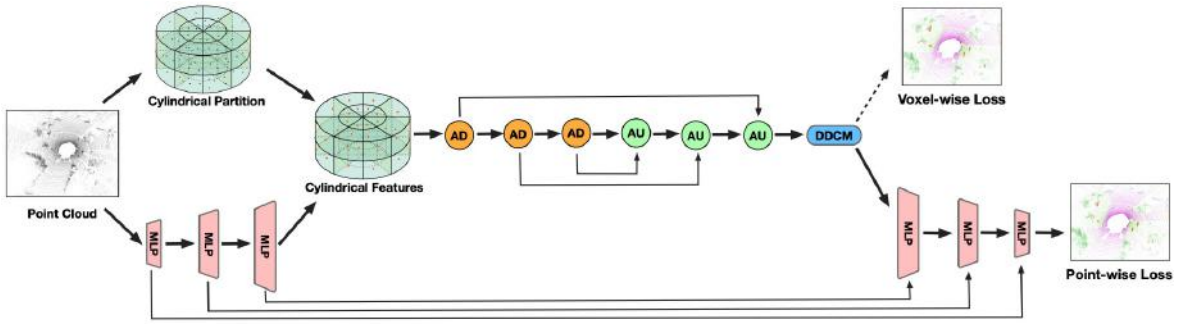


Zhang, et al. PolarNet. [CVPR 2020]

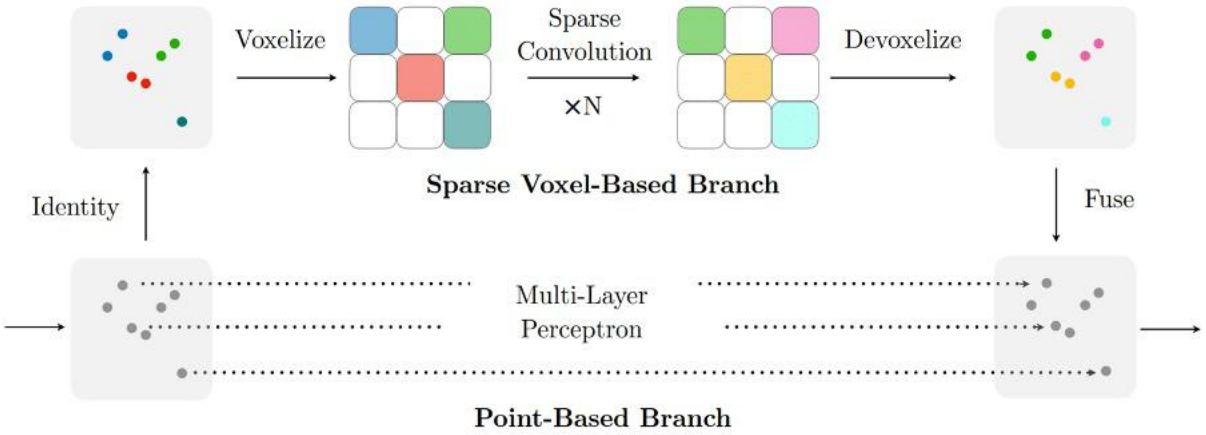


Zhang, et al. PMF. [ICCV 2021]

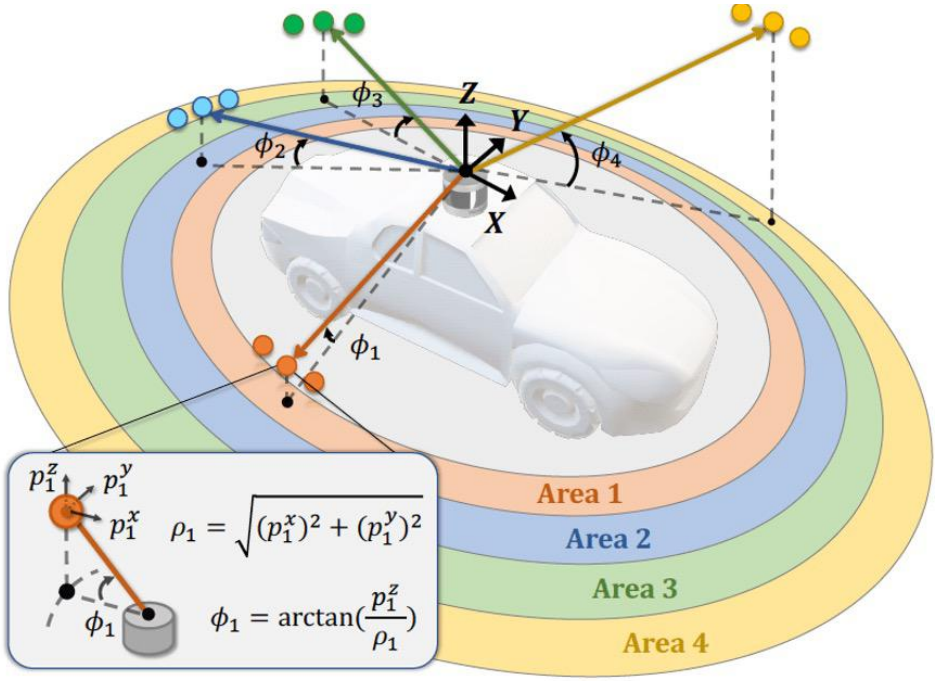
MMDetection3D v1.1.1



Zhu, et al. Cylinder3D++. [TPAMI 2022]



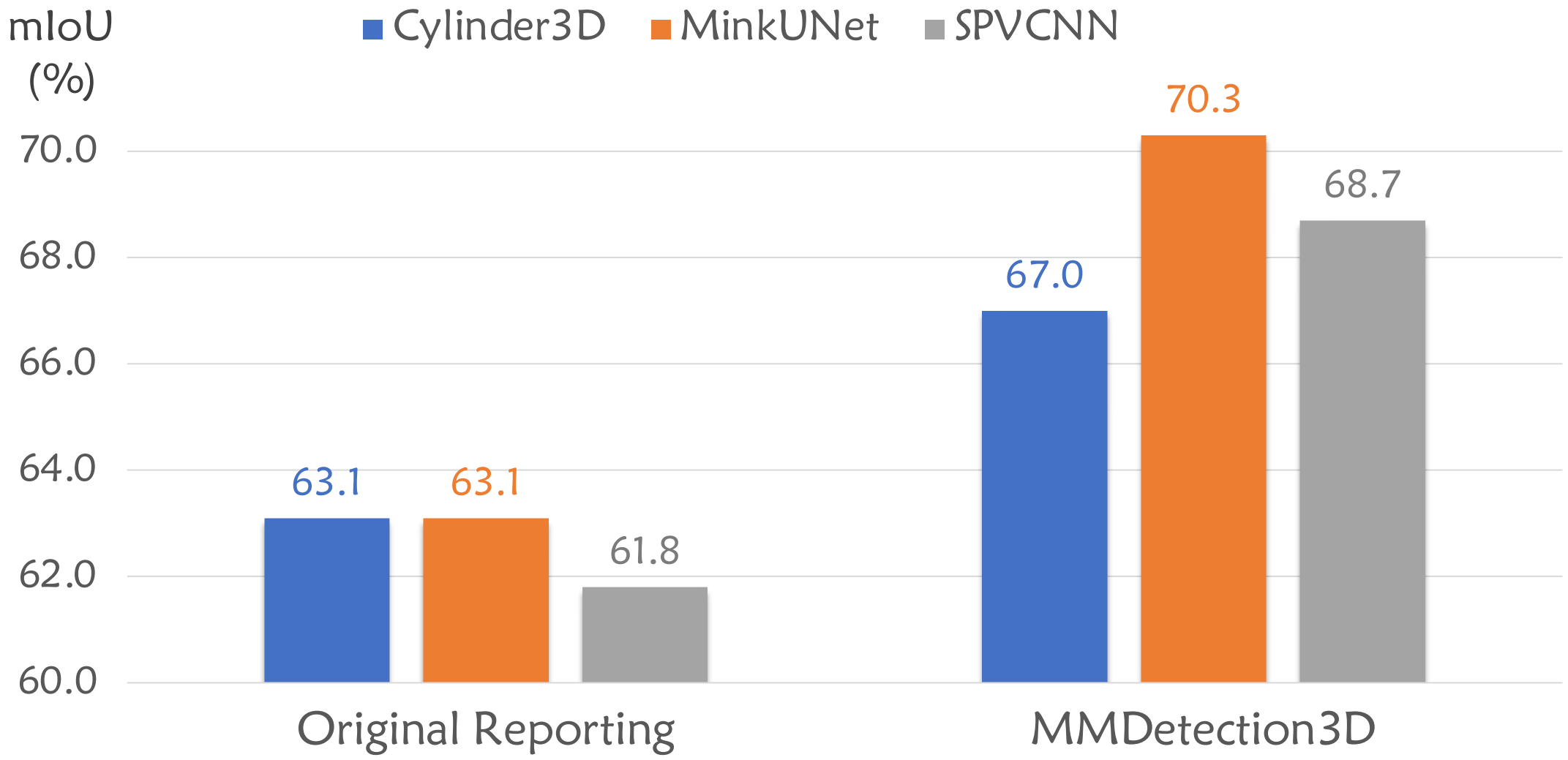
Tang, et al. SPVCNN. [ECCV 2020]



Kong, et al. LaserMix. [CVPR 2023]

The MMDetection3D Codebase: https://github.com/open-mmlab/mmdetection3d

MMDetection3D v1.1.1



The MMDetection3D Codebase: https://github.com/open-mmlab/mmdetection3d

MMDetection3D v1.1.1

Method	Backend	Lr schd	Amp	Laser-Polar Mix	Mem (GB)	Training Time (hours)	FPS	mIoU	Download
MinkUNet18-W16	torchsparse	15e	✓	×	3.4	-	-	60.3	model | log
MinkUNet18-W20	torchsparse	15e	✓	×	3.7	-	-	61.6	model | log
MinkUNet18-W32	torchsparse	15e	✓	×	4.9	-	-	63.1	model | log
MinkUNet34-W32	minkowski engine	3x	×	✓	11.5	6.5	12.2	69.2	model | log
MinkUNet34-W32	spconv	3x	✓	✓	6.7	2	14.6*	68.3	model | log
MinkUNet34-W32	spconv	3x	×	✓	10.5	6	14.5	69.3	model | log
MinkUNet34-W32	torchsparse	3x	✓	✓	6.6	3	12.8	69.3	model | log
MinkUNet34-W32	torchsparse	3x	×	✓	11.8	5.5	15.9	68.7	model | log
MinkUNet34v2-W32	torchsparse	3x	✓	✓	8.9	-	-	70.3	model | log

Major Features

- Supports different sparse convolution backends
- Supports most recent 3D augmentation techniques
- Achieves SoTA 3D semantic segmentation performance

The MMDetection3D Codebase: https://github.com/open-mmlab/mmdetection3d

Topic

1. Overview & Background

2. Data-Efficient 3D Perception

3. Robustness in 3D

4. Segment Any Point Cloud Sequences

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

LaserMix for Semi-Supervised LiDAR Semantic Segmentation

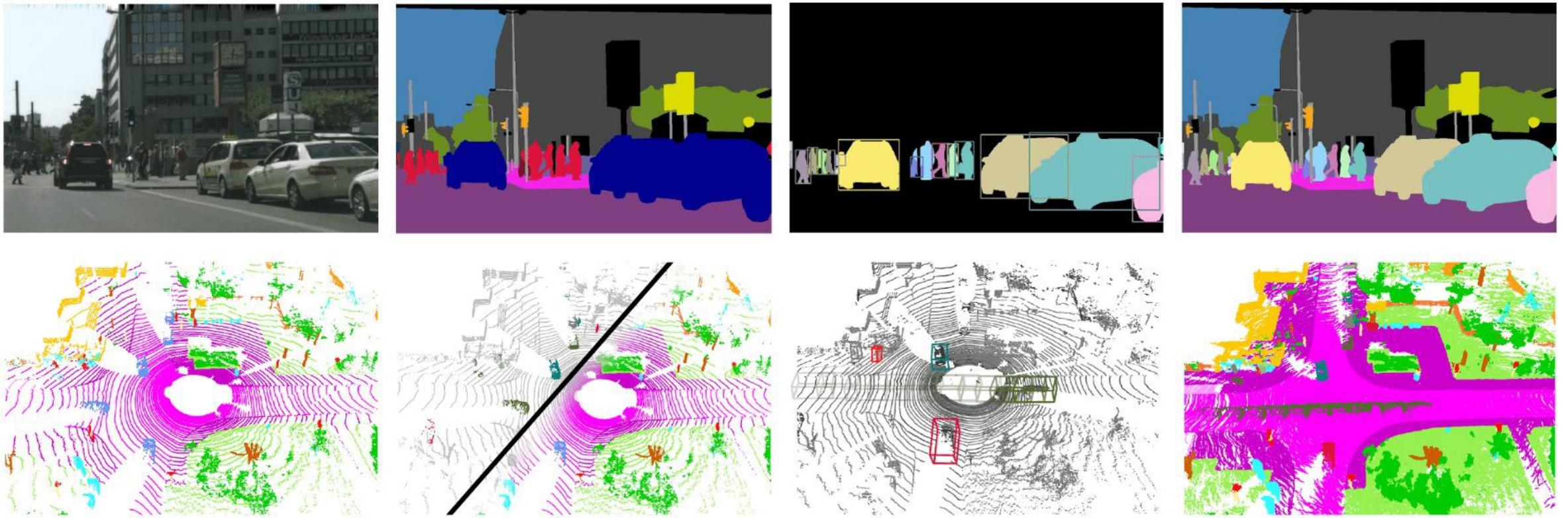
Lingdong Kong[1,2,3] Jiawei Ren[1] Liang Pan[1] Ziwei Liu[1]

[1] S-Lab, Nanyang Technological University

[2] National University of Singapore

[3] CNRS@CREATE

Autonomous Driving Perception



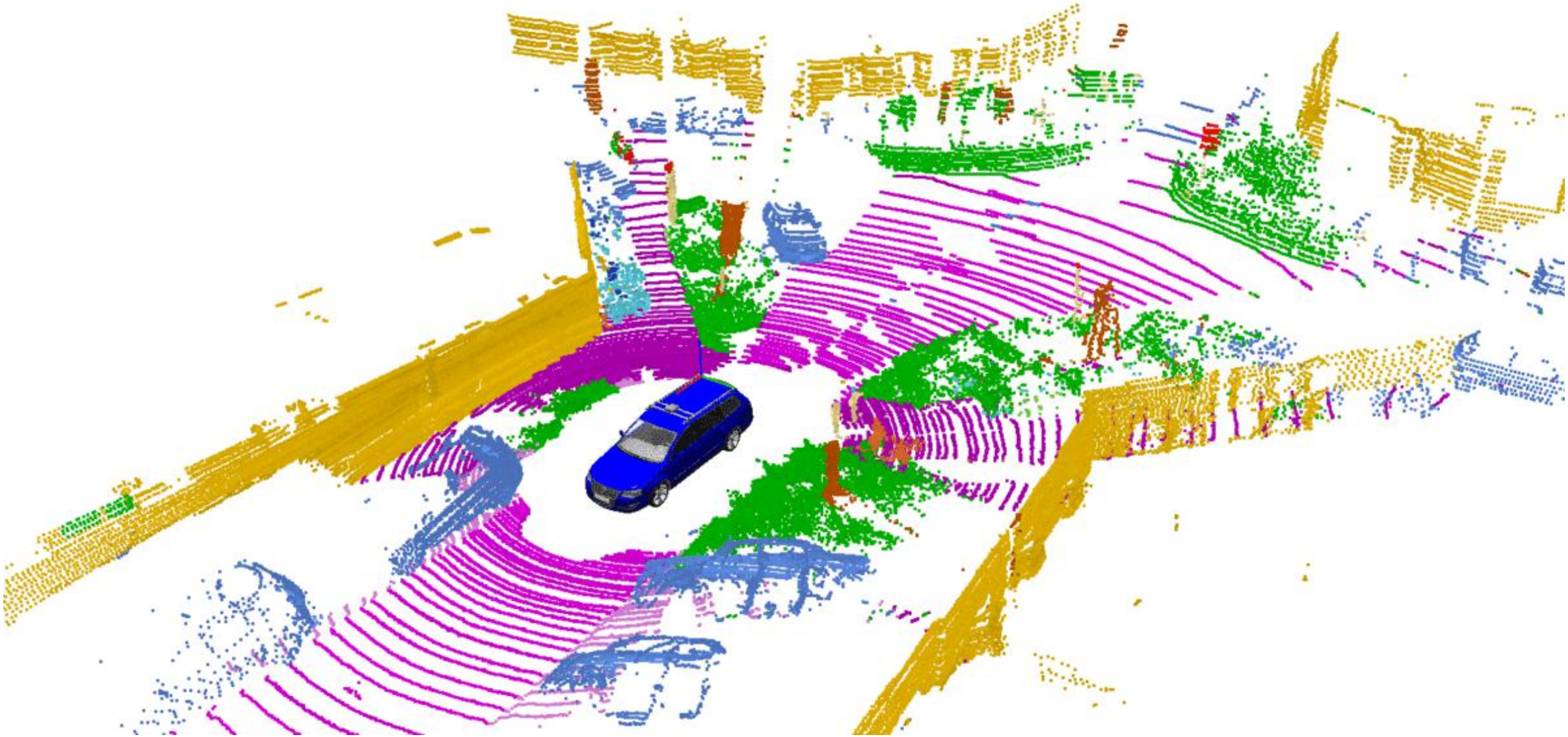
From left to right:

- LiDAR semantic segmentation
- LiDAR panoptic segmentation
- 3D object detection
- 4D LiDAR panoptic segmentation

Why LiDAR sensors?

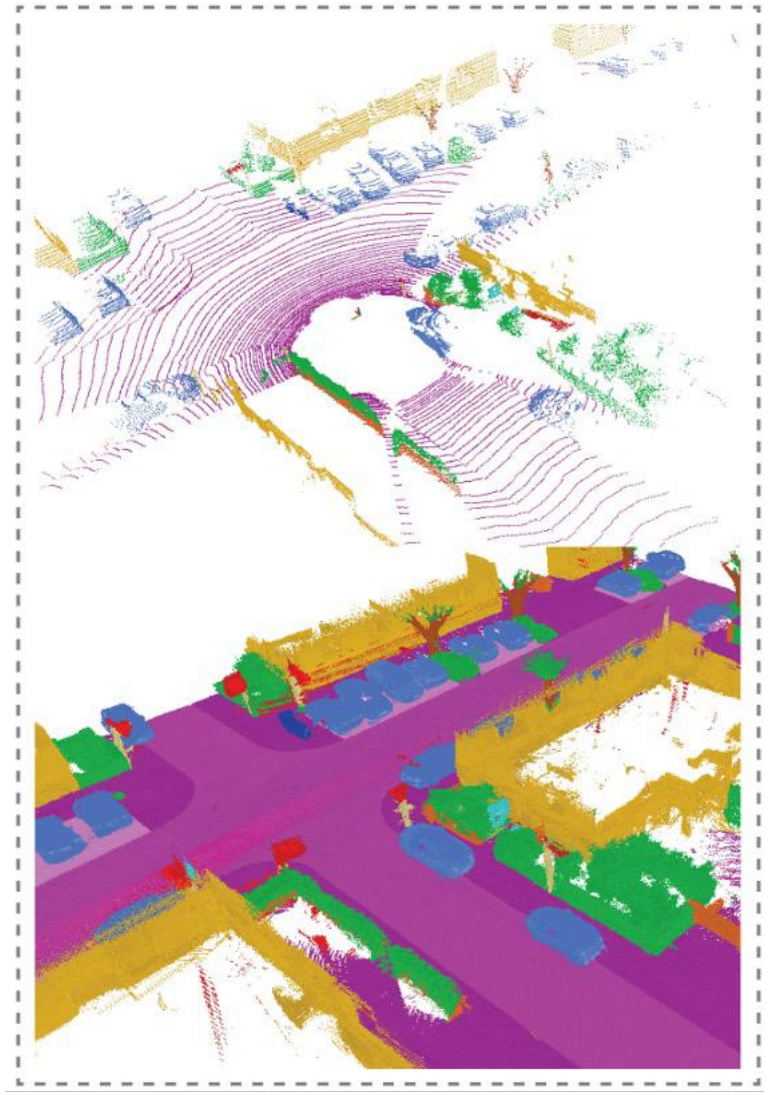
- Accurate depth sensing
- Robust at low-light conditions
- Dense perceptions
- …

LiDAR Semantic Segmentation



A. Milioto, et al. RangeNet++: Fast and accurate LiDAR semantic segmentation, IROS, 2019.

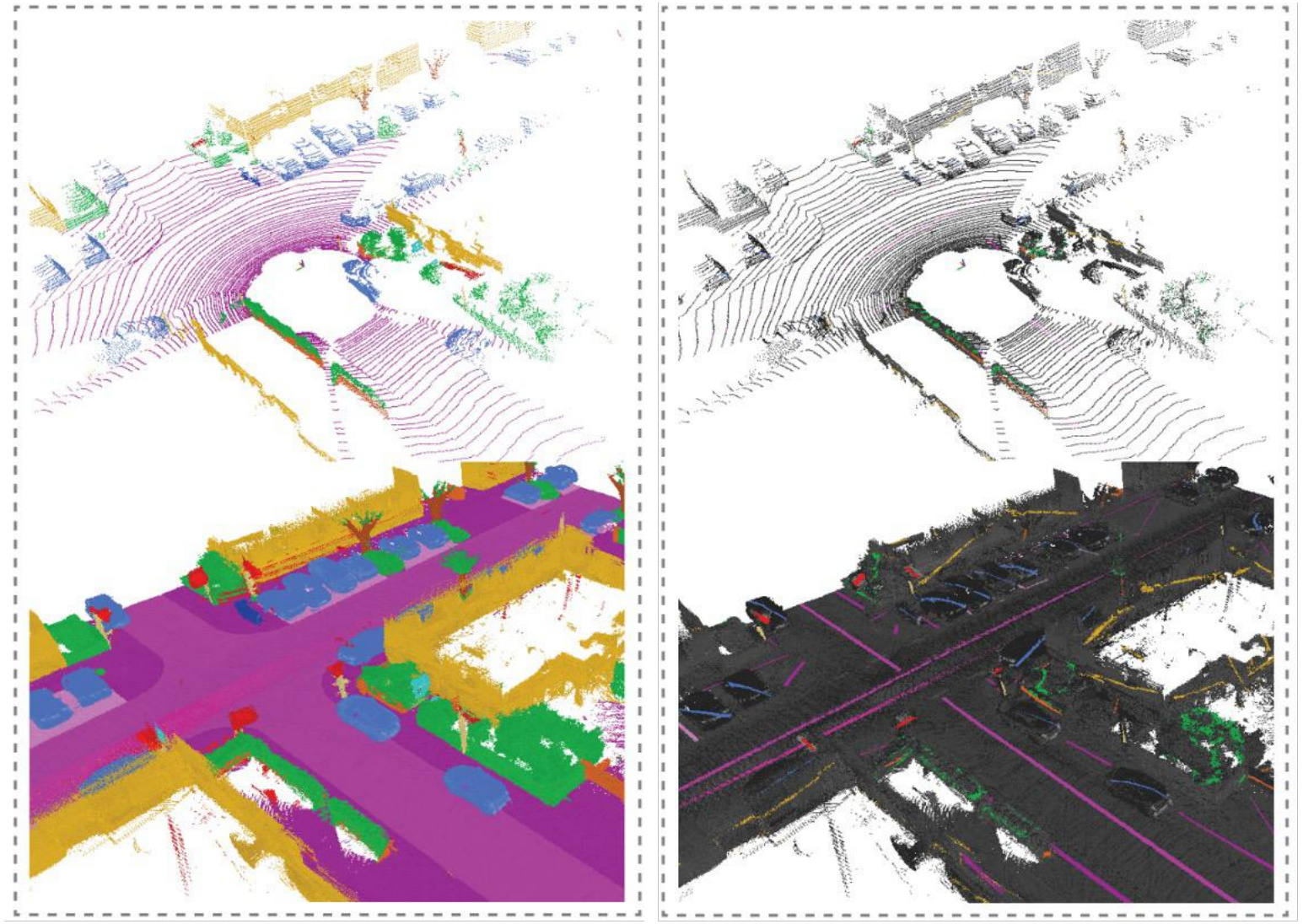
Fully-Supervised LiDAR Semantic Segmentation



- **SemanticKITTI**
  - Full labels (100%)
  - 19 semantic classes
  - 100 m x 100 m
  - Up to 4.5 hours

J. Behley, et al. SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences, ICCV, 2019.

Weakly-Supervised LiDAR Semantic Segmentation



- **SemanticKITTI**
  - Full labels (100%)
  - 19 semantic classes
  - 100 m x 100 m
  - Up to 4.5 hours
- **ScribbleKITTI**
  - Weak (scribble) labels (8.06%)
  - 19 semantic classes
  - 100 m x 100 m
  - 10 - 25 min per scan
  - 90% time saving

O. Unal, et al. Scribble-supervised LiDAR semantic segmentation, CVPR, 2022.

Semi-Supervised LiDAR Segmentation

Objective (This Work)

- We target on the less-explored semi-supervised LiDAR semantic segmentation

Semi-Supervised LiDAR Segmentation

Objective (This Work)

- We target on the less-explored semi-supervised LiDAR semantic segmentation
- Our goal is to leverage the abundant raw LiDAR scans for training accurate segmentation models

Semi-Supervised LiDAR Segmentation

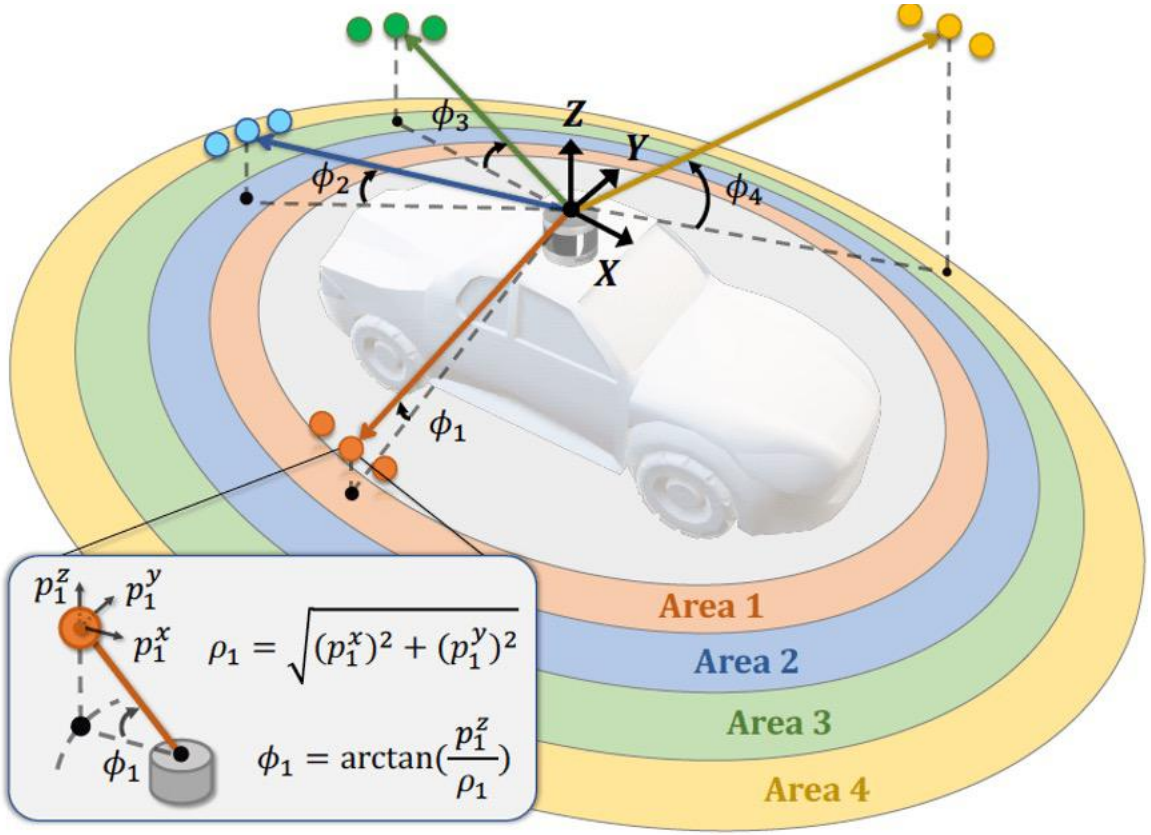
Objective (This Work)

- We target on the less-explored semi-supervised LiDAR semantic segmentation
- Our goal is to leverage the abundant raw LiDAR scans for training accurate segmentation models
- We propose **LaserMix** to make advantages of the spatial prior in LiDAR scenes for effective learning with semi supervisions

TL;DR

LaserMix is a data-efficient learning framework designed for LiDAR segmentation that:

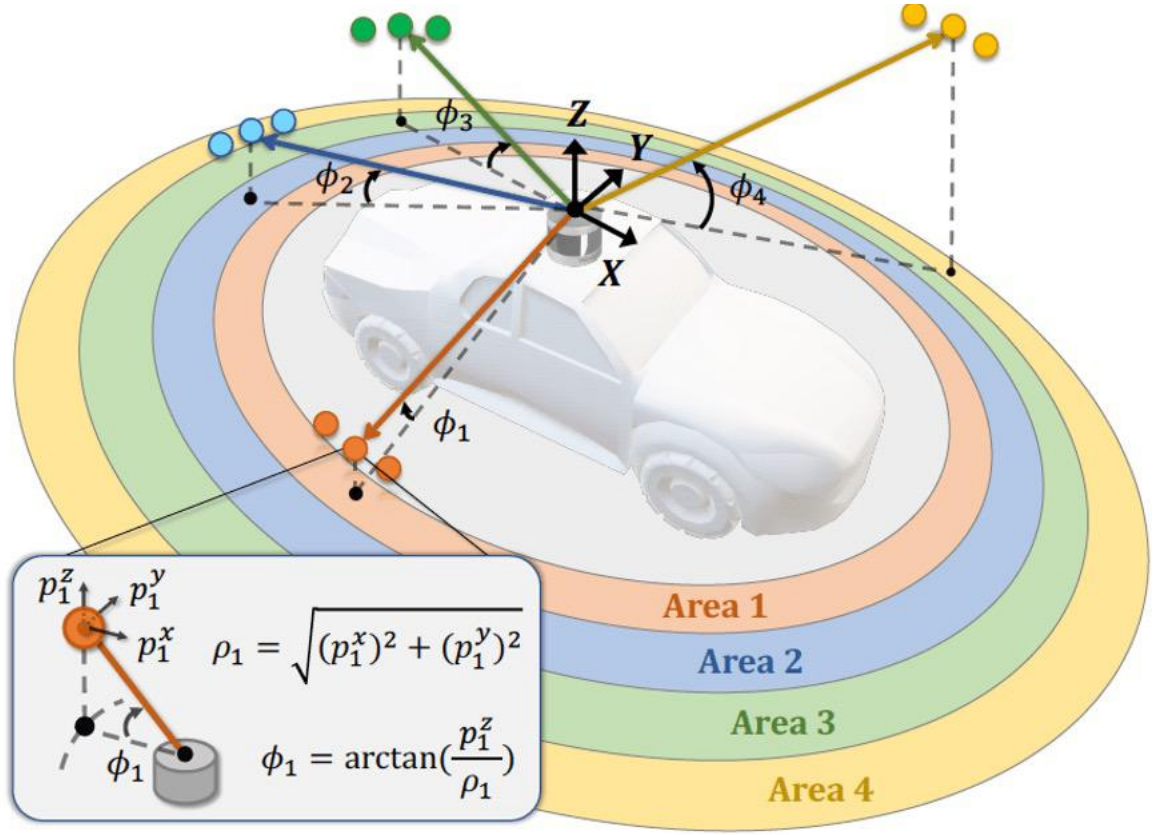
- Leverages the spatial prior in driving scenes for data-efficient learning



TL;DR

LaserMix is a data-efficient learning framework designed for LiDAR segmentation that:

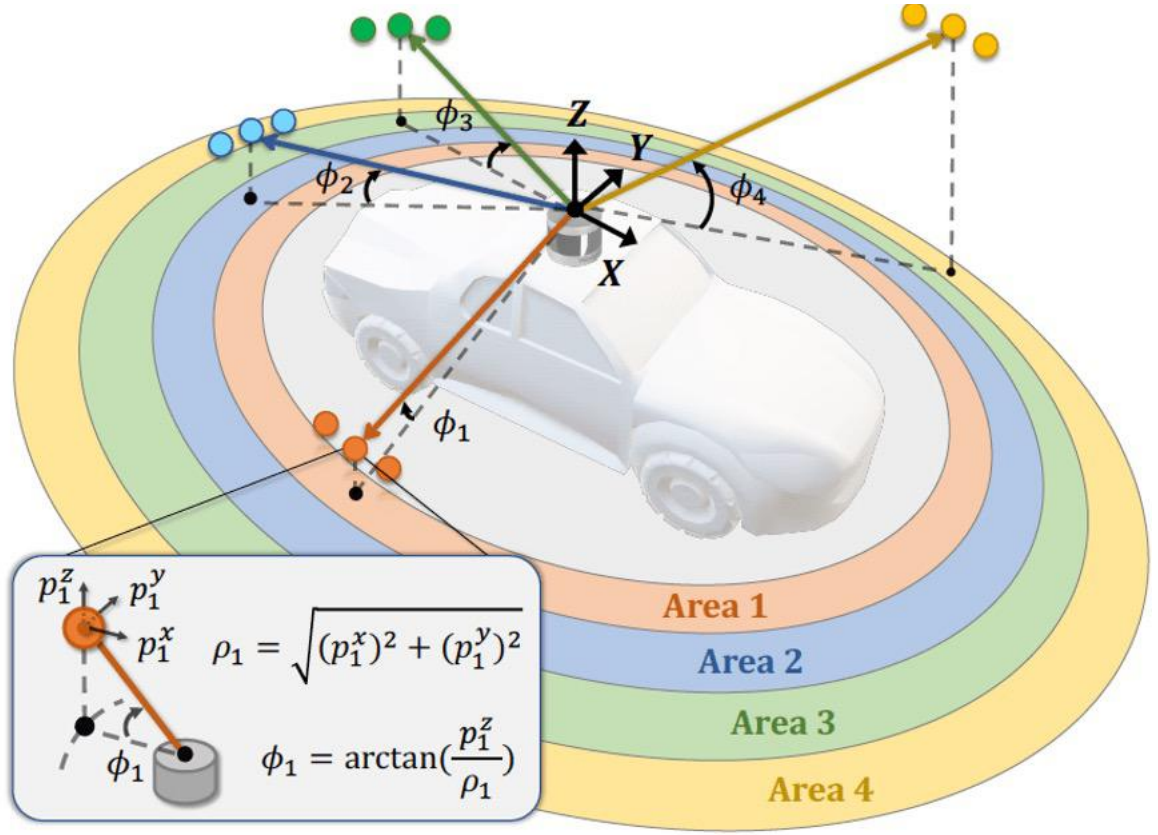
- Leverages the spatial prior in driving scenes for data-efficient learning
- Constructs low-variational areas via laser beam mixing



TL;DR

LaserMix is a data-efficient learning framework designed for LiDAR segmentation that:

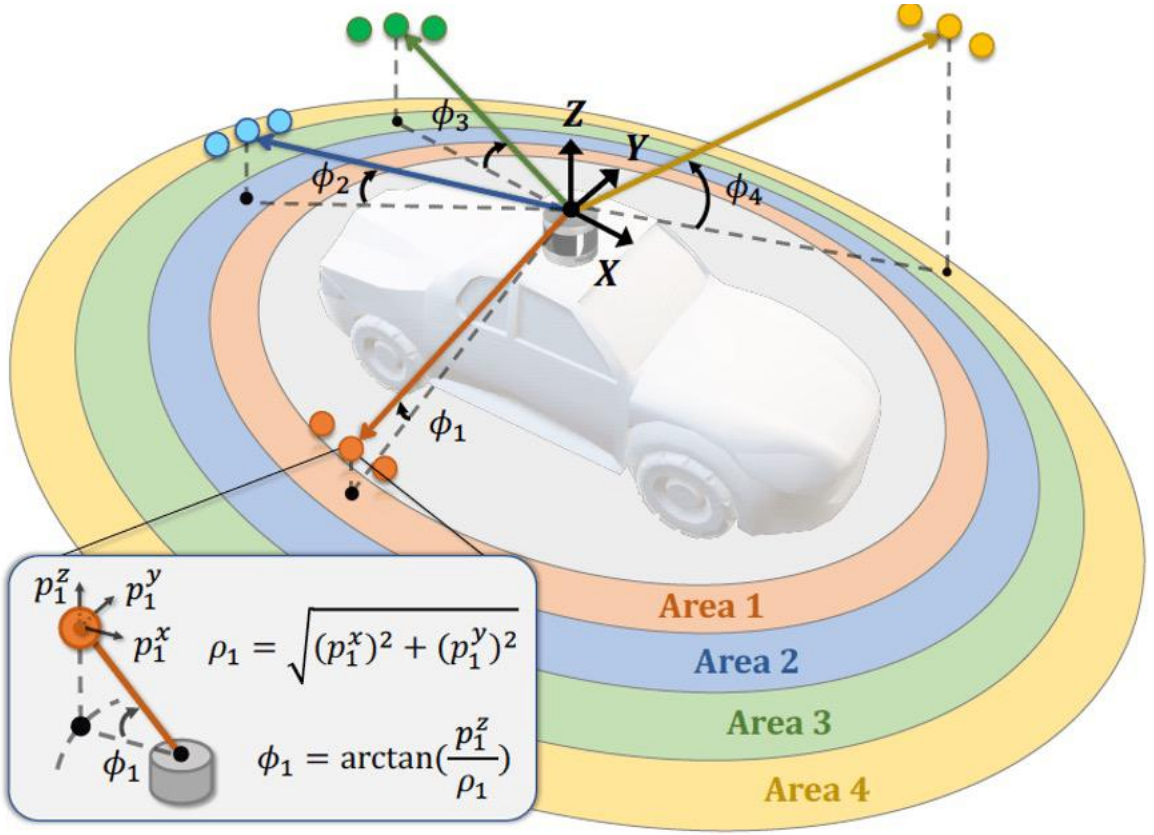
- Leverages the spatial prior in driving scenes for data-efficient learning
- Constructs low-variational areas via laser beam mixing
- Encourages the model to make confident and consistent predictions before and after mixing



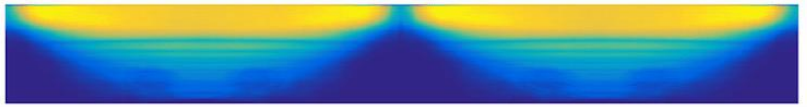
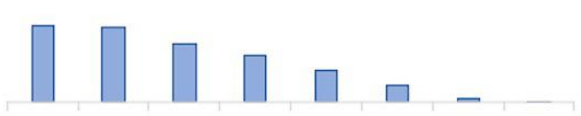
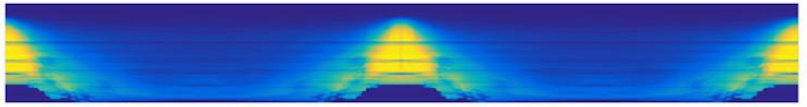
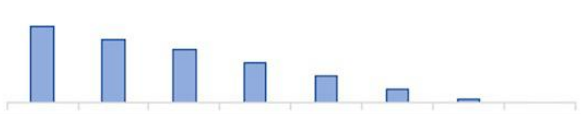
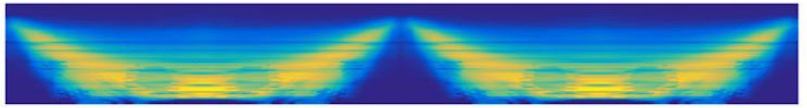
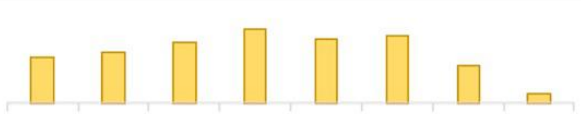
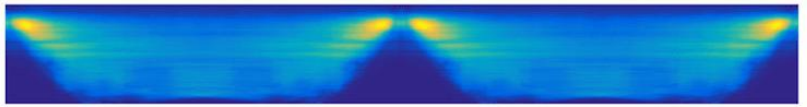
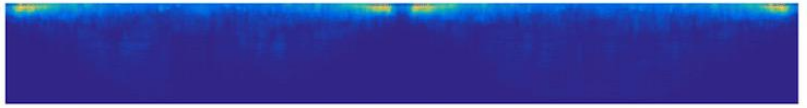
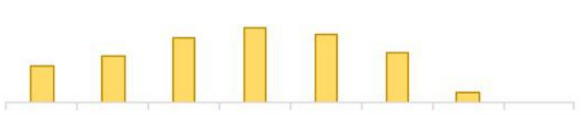
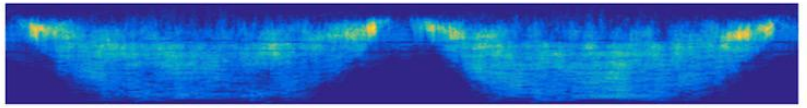
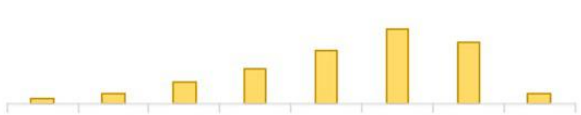
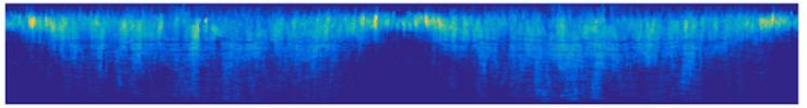
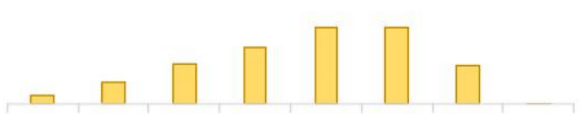
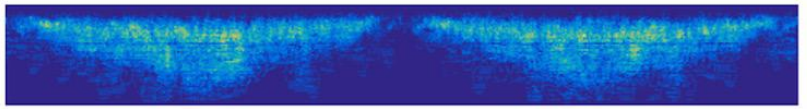
TL;DR

LaserMix is a data-efficient learning framework designed for LiDAR segmentation that:

- Leverages the spatial prior in driving scenes for data-efficient learning
- Constructs low-variational areas via laser beam mixing
- Encourages the model to make confident and consistent predictions before and after mixing
- Achieved competitive results over full supervision counterparts with 2x to 5x fewer annotations

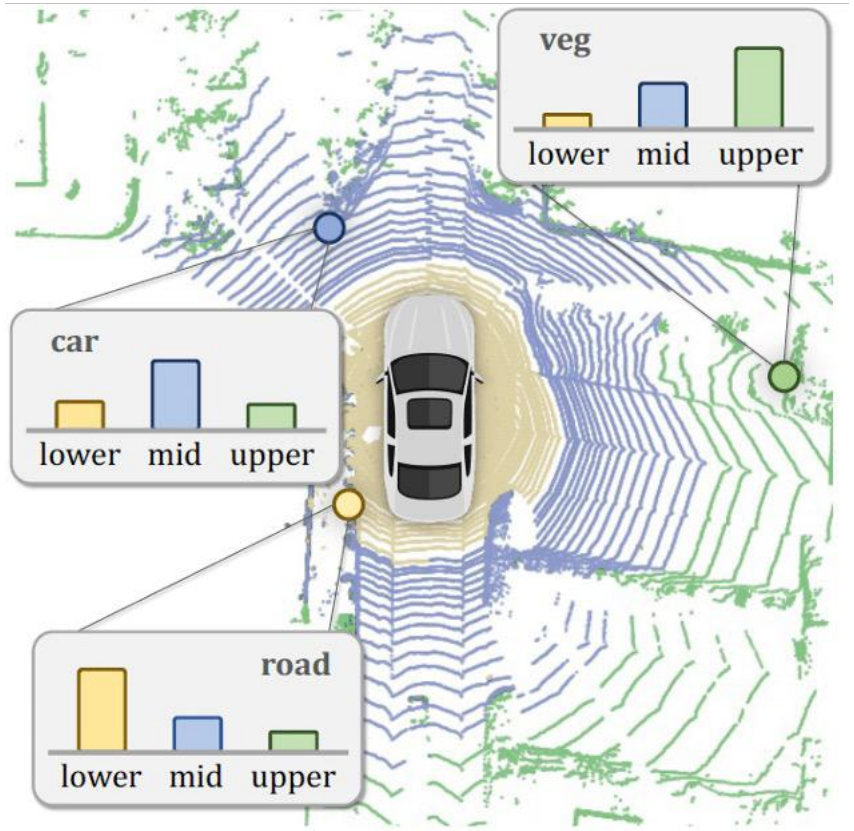


What is Spatial Prior ?

Class	Type	Proportion	Distribution	Heatmap
vegetation	static	24.825%		
road	static	22.545%		
sidewalk	static	16.353%		
car	dynamic	4.657%		
traffic-sign	static	0.061%		
motorcycle	dynamic	0.045%		
person	dynamic	0.036%		
bicycle	dynamic	0.018%		

Certain semantic class tends to appear at certain areas around the ego-vehicle!

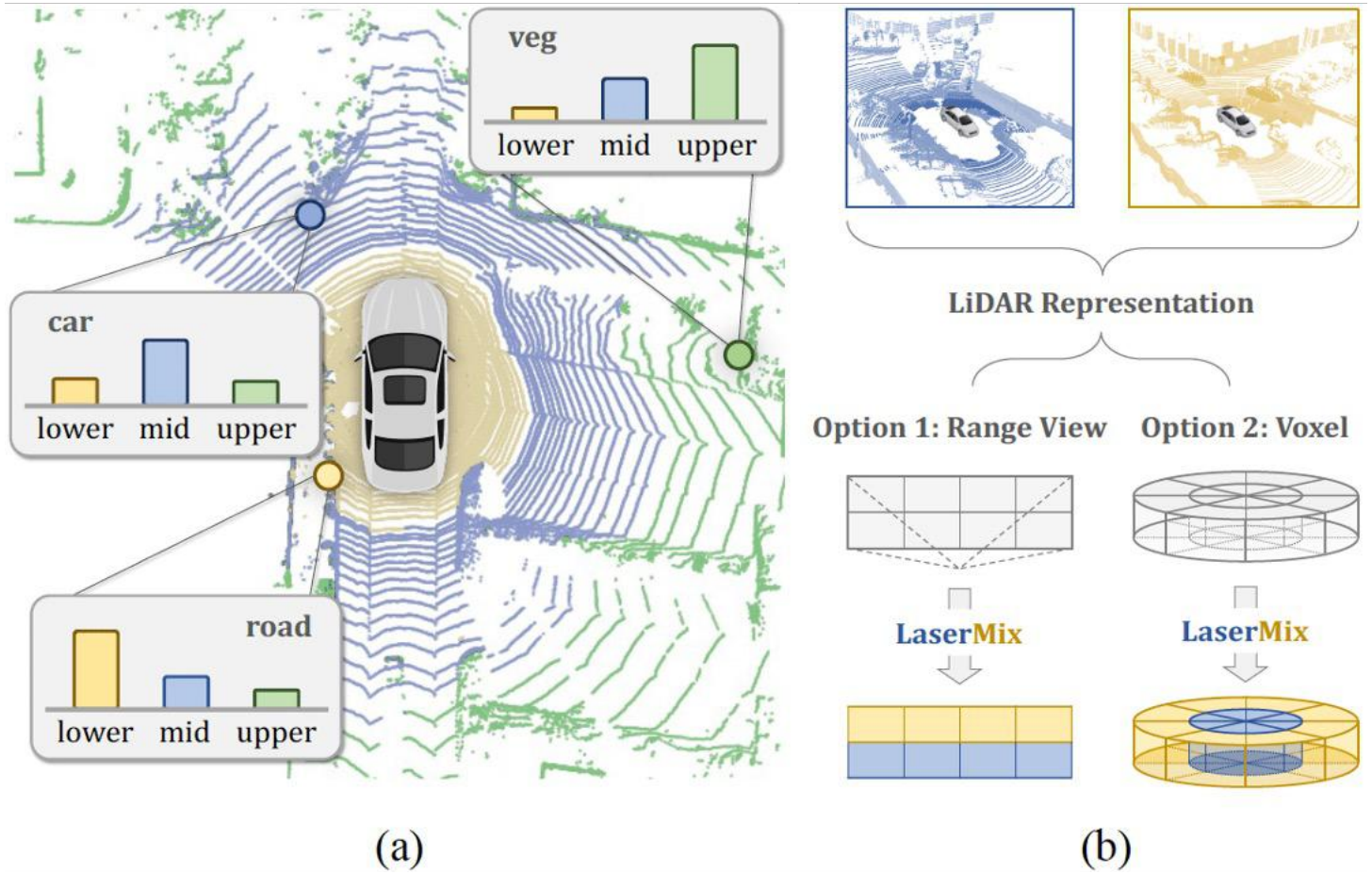
LaserMix: Overview



(a)

(a) Motivation. Semantic scene priors are overt for each category in LiDAR point clouds.

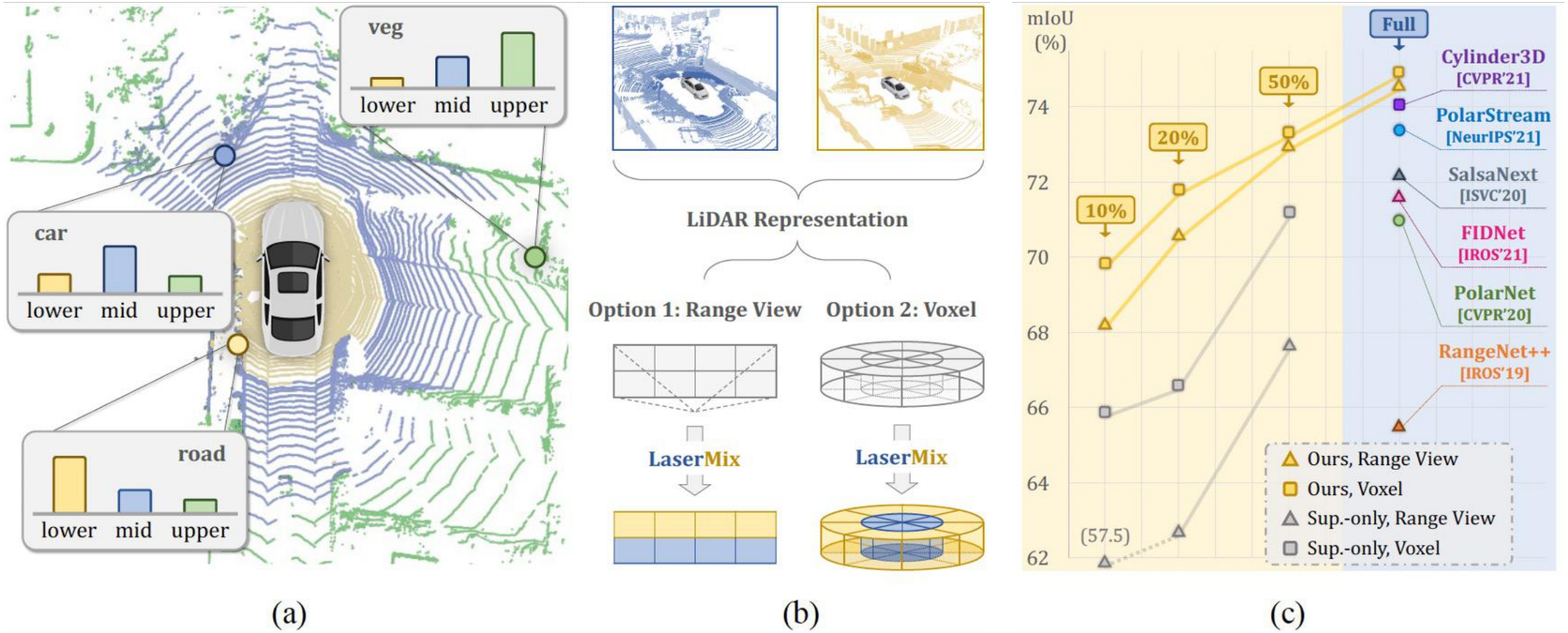
LaserMix: Overview



(a) **Motivation.** Semantic scene priors are overt for each category in LiDAR point clouds.

(b) **Generalizability.** LaserMix can be added into various popular LiDAR representations.

LaserMix: Overview

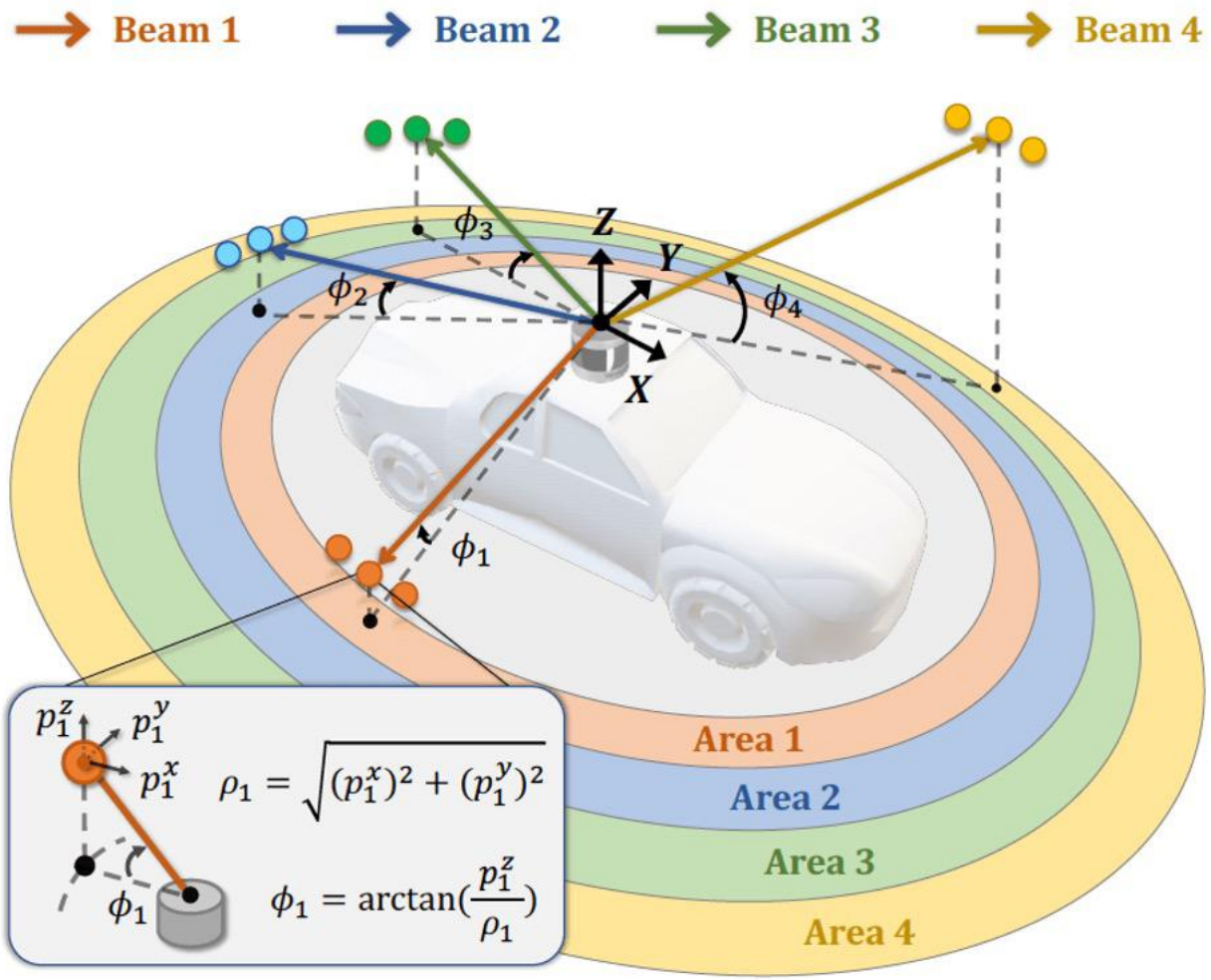


(a) **Motivation.** Semantic scene priors are overt for each category in LiDAR point clouds.

(b) **Generalizability.** LaserMix can be added into various popular LiDAR representations.

(c) **Effectiveness.** LaserMix helps to improve both semi- and fully-supervised settings.

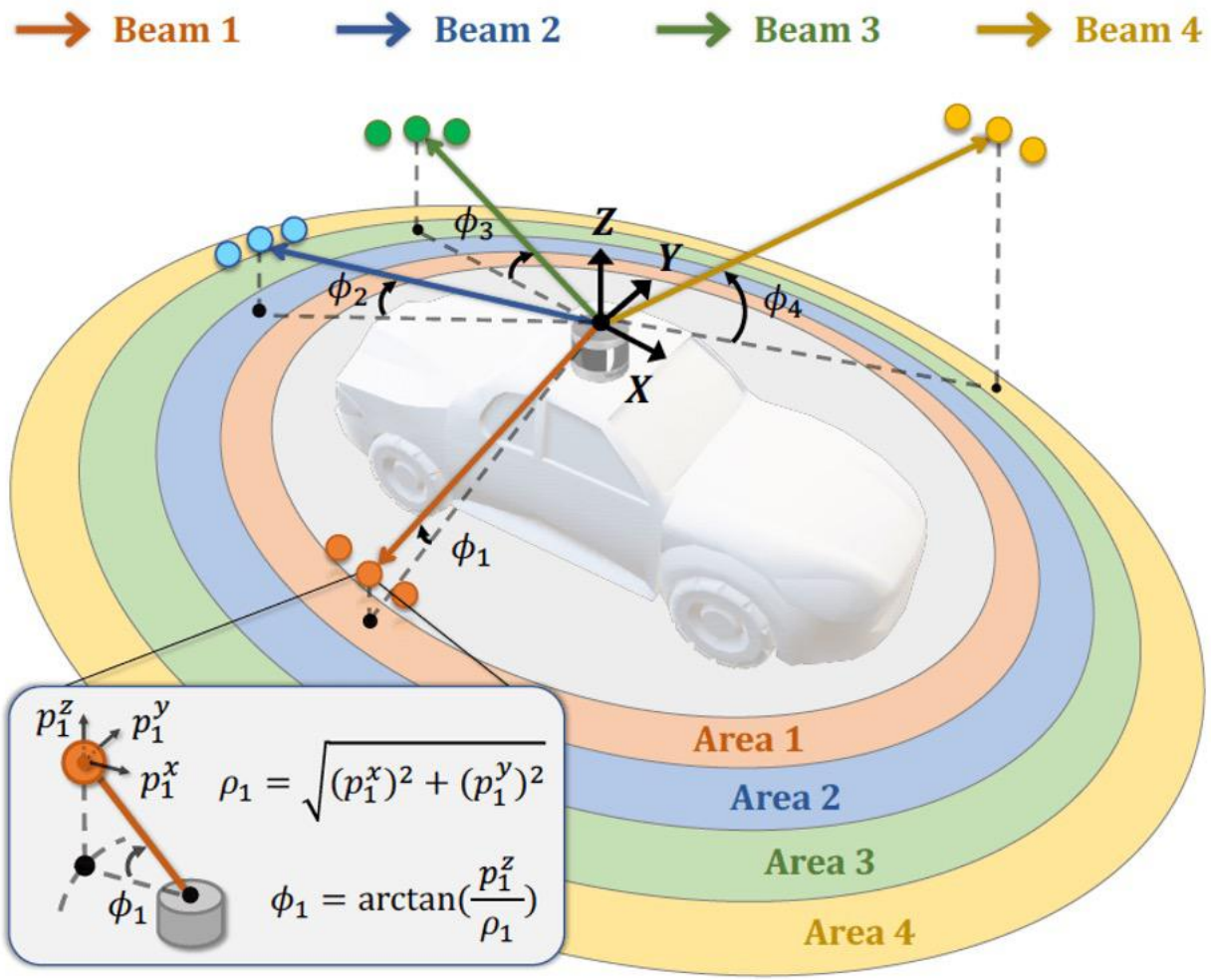
LaserMix: Laser Partition & Mixing



Three-Step Procedure

1. Partitioning the captured LiDAR scan into low-variational areas

LaserMix: Laser Partition & Mixing

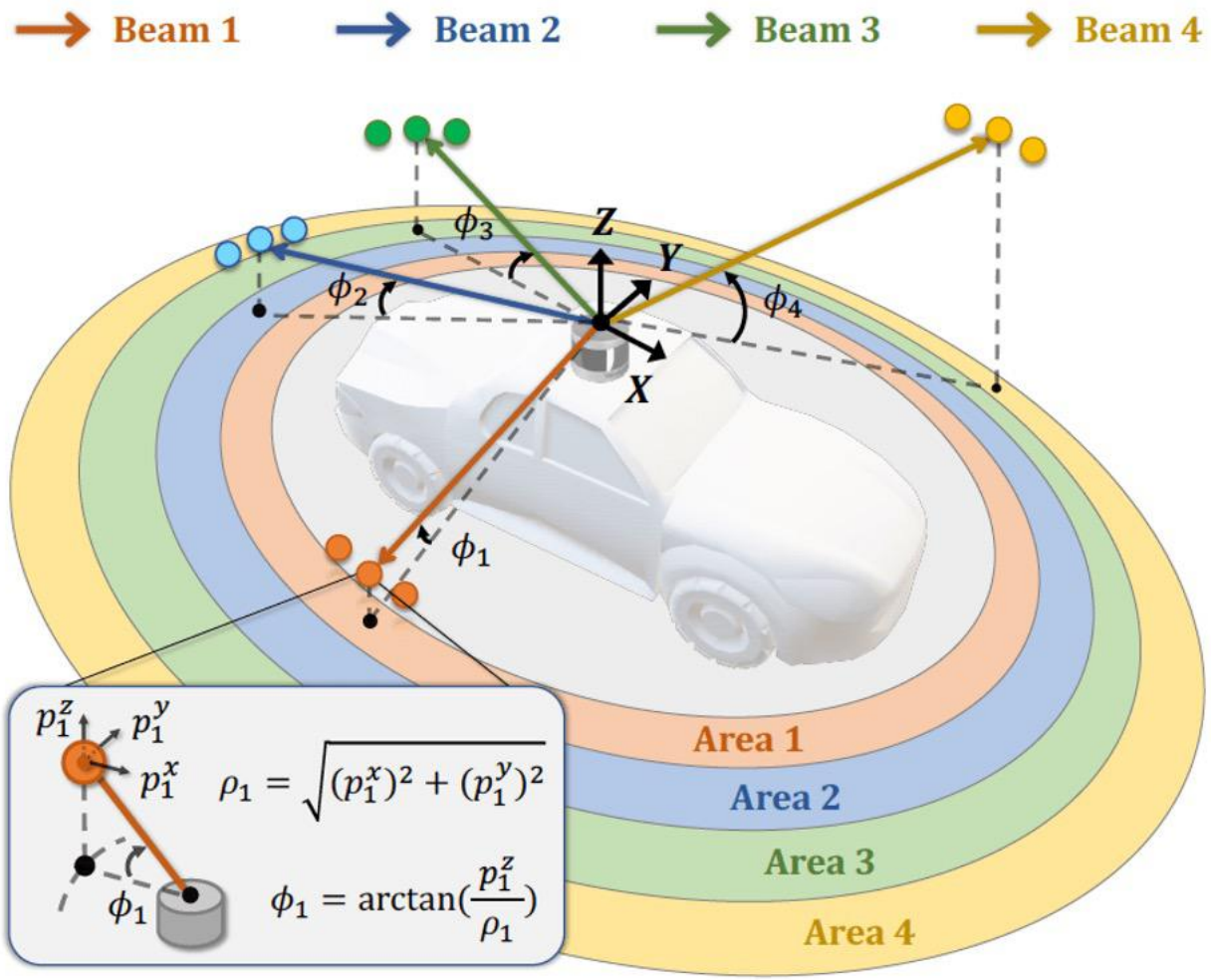


Three-Step Procedure

1. Partitioning the captured LiDAR scan into low-variational areas

2. Efficiently mixing every area in the LiDAR scan with foreign data

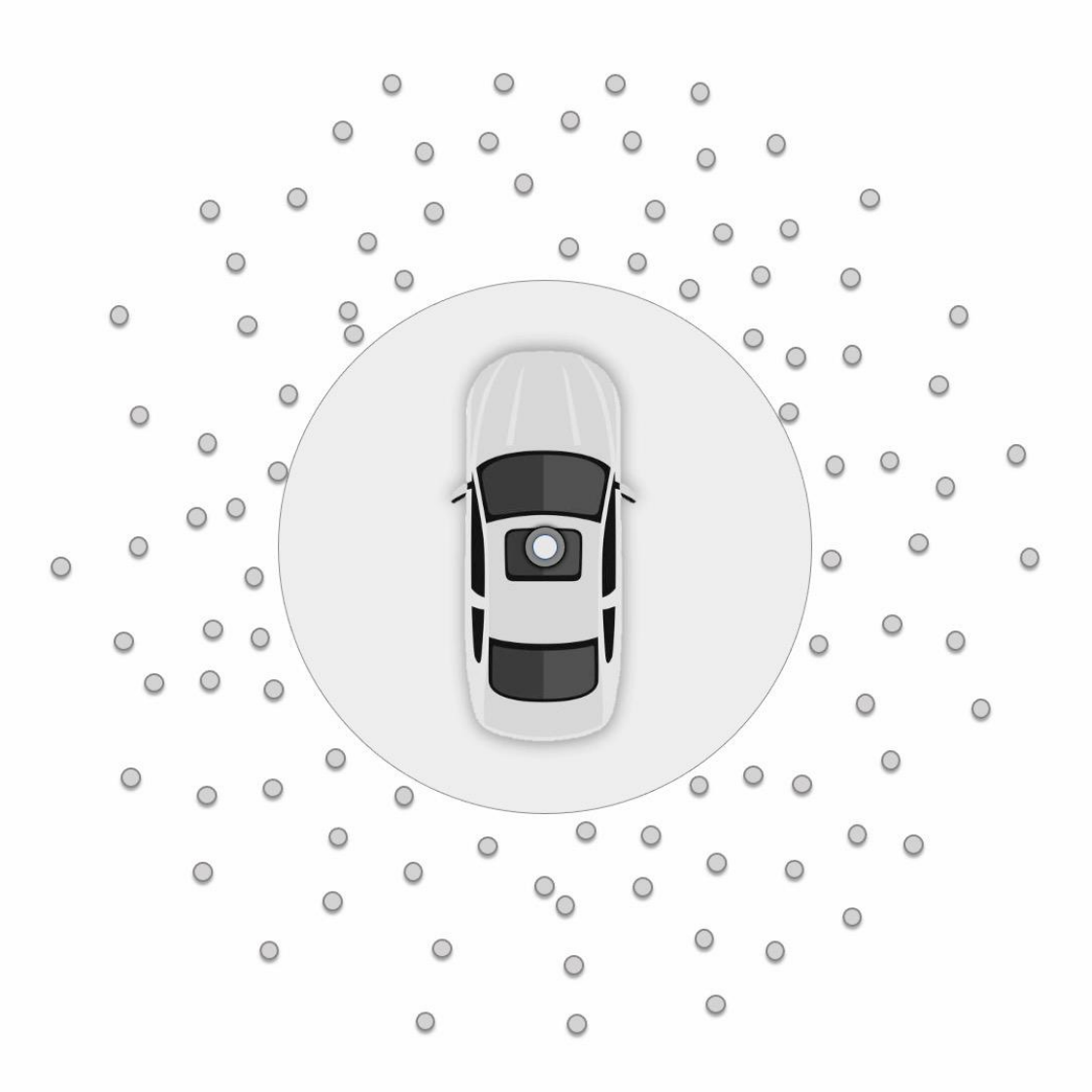
LaserMix: Laser Partition & Mixing



Three-Step Procedure

1. Partitioning the captured LiDAR scan into low-variational areas
2. Efficiently mixing every area in the LiDAR scan with foreign data
3. Encouraging the LiDAR segmentation models to make confident and consistent predictions on the same area in different mixing

LaserMix: Laser Partition & Mixing



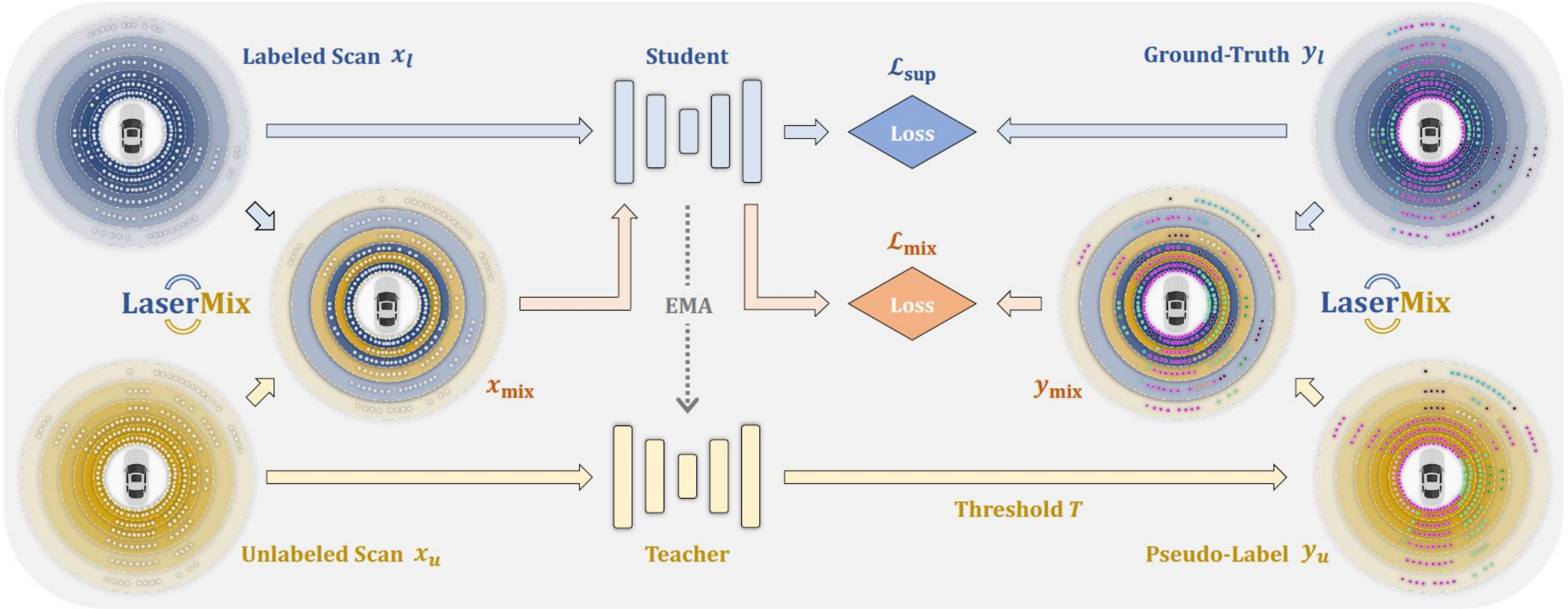
- **Inclination:**

$$\phi_i = \arctan\left(\frac{p_i^z}{\sqrt{(p_i^x)^2+(p_i^y)^2}}\right)$$

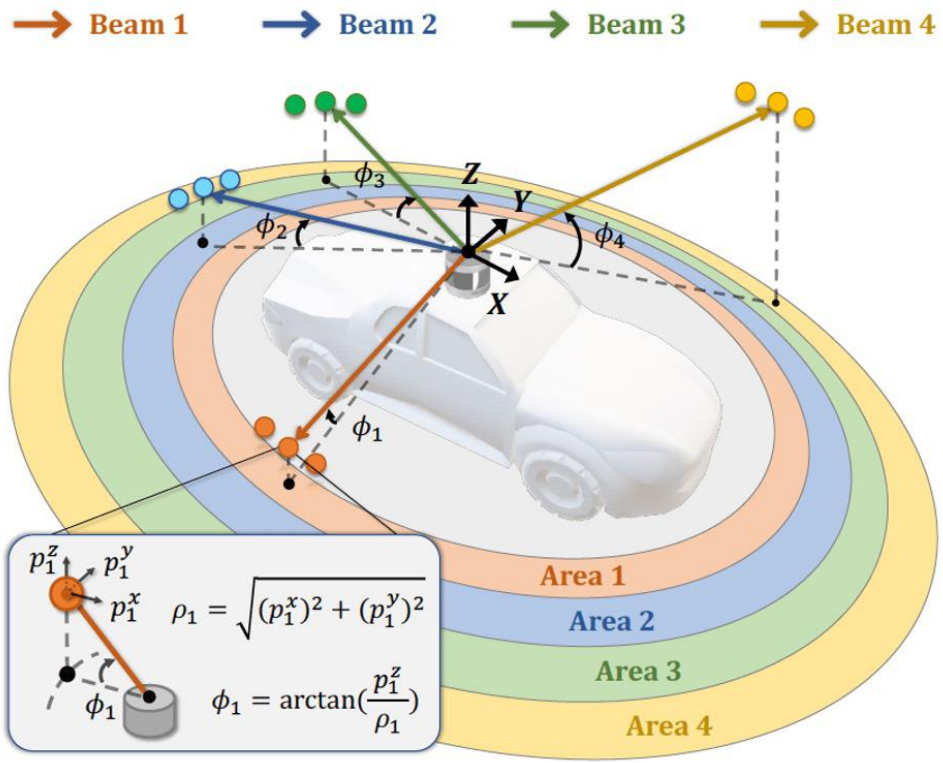
- **Depth:** $\rho_i = \sqrt{(p_i^x)^2+(p_i^y)^2}$

- **Azimuth:** $\alpha_i = \arctan\left(\frac{p_i^y}{p_i^x}\right)$

LaserMix: Consistency Regularization



Proof & Derivation (See Our Paper)



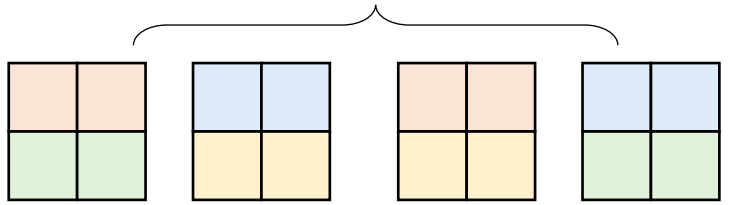
$H(X_{\text{in}}, Y_{\text{in}} \mid A = \text{Outer Rings})$

LiDAR data and labels strongly correlate with the area A

$H(X_{\text{in}}, Y_{\text{in}}|A)$ is low

A Simplified Case:

Color correlates with the row; each row has two colors

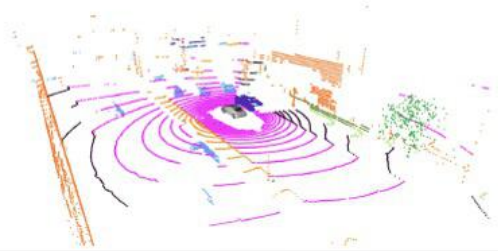
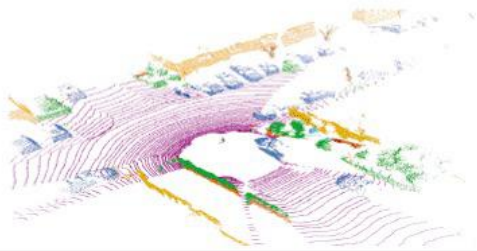
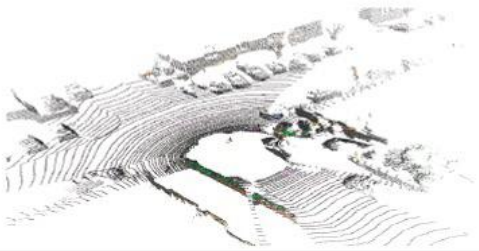
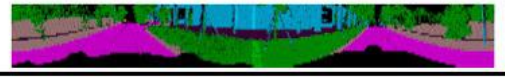
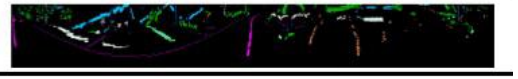


$$H(X, Y|A \in \{2\times2\}) = \log 4$$

$$H(X, Y|A \in \{1\times2, 1\times2\}) = \log 4$$

$$H(X, Y|A \in \{2\times1, 2\times1\}) = \log 2$$

Experimental Settings

	nuScenes [15]	SemanticKITTI [16]	ScribbleKITTI [4]
Vis.			
#Class	16	19	19
#Train	29130	19130	19130
#Val	6019	4071	4071
Res. (RV)	32×1920	64×2048	64×2048
Res. (voxel)	$[240, 180, 20]$	$[240, 180, 20]$	$[240, 180, 20]$
#Beam	32	64	64
$[\phi_{\text{up}}, \phi_{\text{low}}]$	$[10°, -30°]$	$[3°, -25°]$	$[3°, -25°]$
$[p^x_{\text{max}}, p^x_{\text{min}}]$	$[50m, -50m]$	$[50m, -50m]$	$[50m, -50m]$
$[p^y_{\text{max}}, p^y_{\text{min}}]$	$[50m, -50m]$	$[50m, -50m]$	$[50m, -50m]$
$[p^z_{\text{max}}, p^z_{\text{min}}]$	$[3m, -5m]$	$[2m, -4m]$	$[2m, -4m]$
#Label	100%	100%	8.06%
Intensity			
Range			
Semantics			

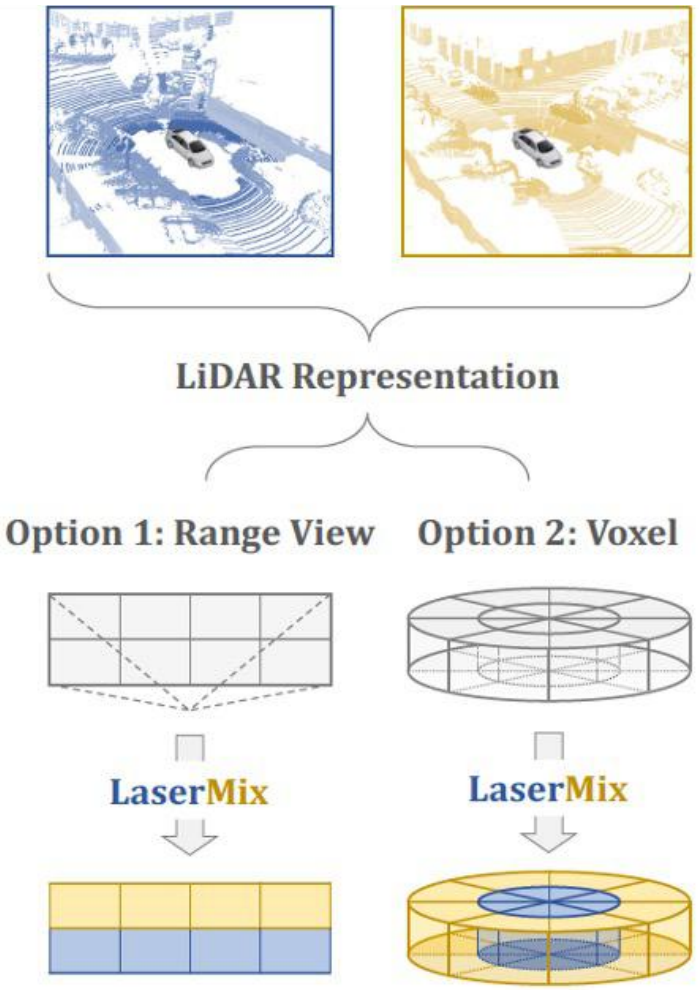
High-res LiDAR:
- SemanticKITTI
- Denser scenes

Low-res LiDAR:
- nuScenes
- Sparser scenes

Weak supervision:
- ScribbleKITTI
- Sparse labels

Experimental Settings

- **Range View**
  - Backbone: FIDNet [IROS' 21]
  - # Param: 6.05M
  - 6 x 32 x 1920 (nuScenes)
  - 6 x 64 x 2048 (SemanticKITTI/ScribbleKITTI)
- **Voxel**
  - Backbone: Cylinder3D [CVPR' 21]
  - # Param: 28.13M
  - [240, 180, 20]
- **Data Split**
  - 1%, 10%, 20%, 50% (labeled)
  - Random sampling
  - Assume the remaining ones are unlabeled



Y. Zhao, et al. FIDNet: LiDAR point cloud semantic segmentation with fully interpolation decoding, IROS, 2021.
X. Zhu, et al. Cylindrical and asymmetrical 3D convolution networks for LiDAR segmentation, CVPR, 2021.

Comparative Study

Repr.	Method	nuScenes [15] 1%	nuScenes [15] 10%	nuScenes [15] 20%	nuScenes [15] 50%	SemanticKITTI [16] 1%	SemanticKITTI [16] 10%	SemanticKITTI [16] 20%	SemanticKITTI [16] 50%	ScribbleKITTI [4] 1%	ScribbleKITTI [4] 10%	ScribbleKITTI [4] 20%	ScribbleKITTI [4] 50%
Range View	*Sup.-only*	38.3	57.5	62.7	67.6	36.2	52.2	55.9	57.2	33.1	47.7	49.9	52.5
	MeanTeacher [26]	42.1	60.4	65.4	69.4	37.5	53.1	56.1	57.4	34.2	49.8	51.6	53.3
	CBST [30]	40.9	60.5	64.3	69.3	39.9	53.4	56.1	56.9	35.7	50.7	52.7	54.6
	CutMix-Seg [29]	43.8	63.9	64.8	69.8	37.4	54.3	56.6	57.6	36.7	50.7	52.9	54.3
	CPS [13]	40.7	60.8	64.9	68.0	36.5	52.3	56.3	57.4	33.7	50.0	52.8	54.6
	LaserMix (Ours)	**49.5**	**68.2**	**70.6**	**73.0**	**43.4**	**58.8**	**59.4**	**61.4**	**38.3**	**54.4**	**55.6**	**58.7**
	$\Delta\uparrow$	+11.2	+10.7	+7.9	+5.4	+7.2	+6.6	+3.5	+4.2	+5.2	+6.7	+5.7	+6.2
Voxel	*Sup.-only*	50.9	65.9	66.6	71.2	45.4	56.1	57.8	58.7	39.2	48.0	52.1	53.8
	MeanTeacher [26]	51.6	66.0	67.1	71.7	45.4	57.1	59.2	60.0	41.0	50.1	52.8	53.9
	CBST [30]	53.0	66.5	69.6	71.6	48.8	58.3	59.4	59.7	41.5	50.6	53.3	54.5
	CPS [13]	52.9	66.3	70.0	72.5	46.7	58.7	59.6	60.5	41.4	51.8	53.9	54.8
	LaserMix (Ours)	**55.3**	**69.9**	**71.8**	**73.2**	**50.6**	**60.0**	**61.9**	**62.3**	**44.2**	**53.7**	**55.1**	**56.8**
	$\Delta\uparrow$	+4.4	+4.0	+5.2	+2.0	+5.2	+3.9	+4.1	+3.6	+5.0	+5.7	+3.0	+3.0

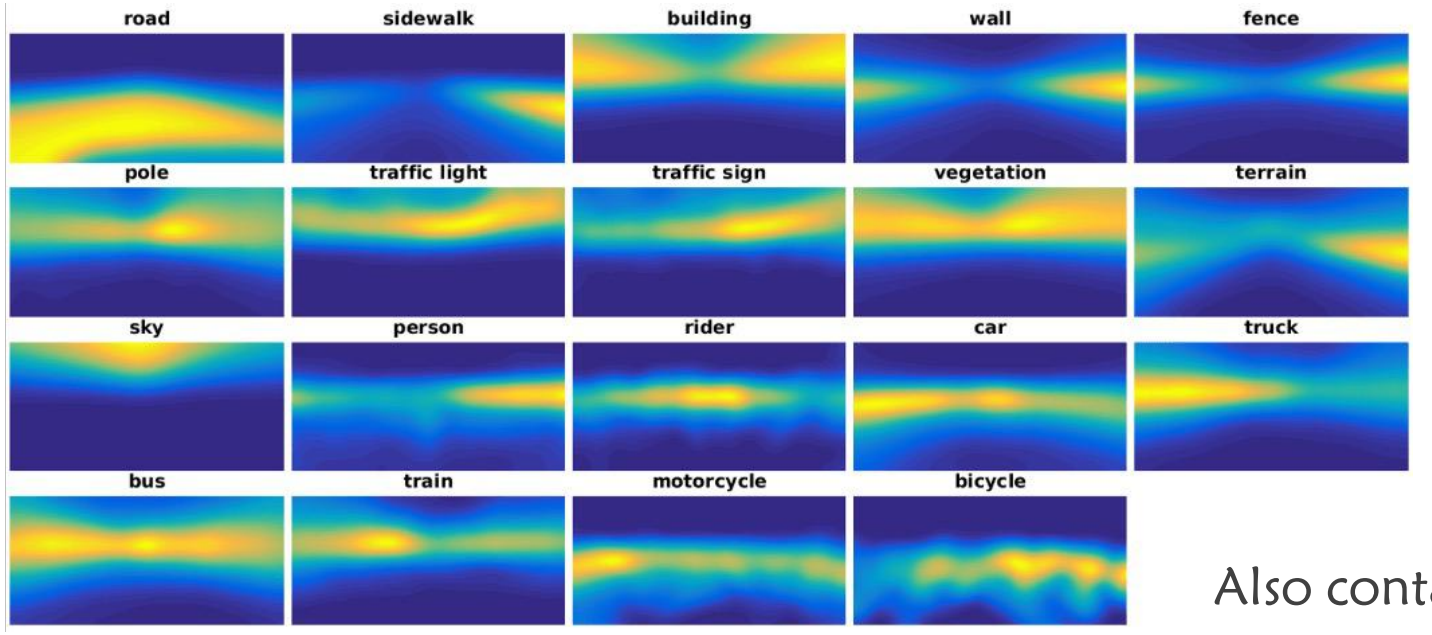
A. Tarvainen and H. Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, NeurIPS, 2017.

G. French, et al. Semi-supervised semantic segmentation needs strong, high-dimensional perturbations, BMVC, 2020.

Y. Zou, et al. Domain adaptation for semantic segmentation via class-balanced self-training, ECCV, 2018.

X. Chen, et al. Semi-supervised semantic segmentation with cross pseudo supervision, CVPR, 2021.

Comparative Study



Method	1/16	1/8	1/4	1/2
MeanTeacher [26]	66.1	71.2	74.4	76.3
***w/* Ours**	68.7	72.3	75.7	76.8
$\Delta\uparrow$	+2.6	+1.1	+1.3	+0.5
CCT [11]	66.4	72.5	75.7	76.8
GCT [12]	65.8	71.3	75.3	77.1
CPS [13]	69.8	74.4	76.9	78.6
CPS-CutMix [13]	74.5	76.6	77.8	78.8
***w/* Ours**	75.5	77.1	78.3	79.1
$\Delta\uparrow$	+1.0	+0.5	+0.5	+0.3

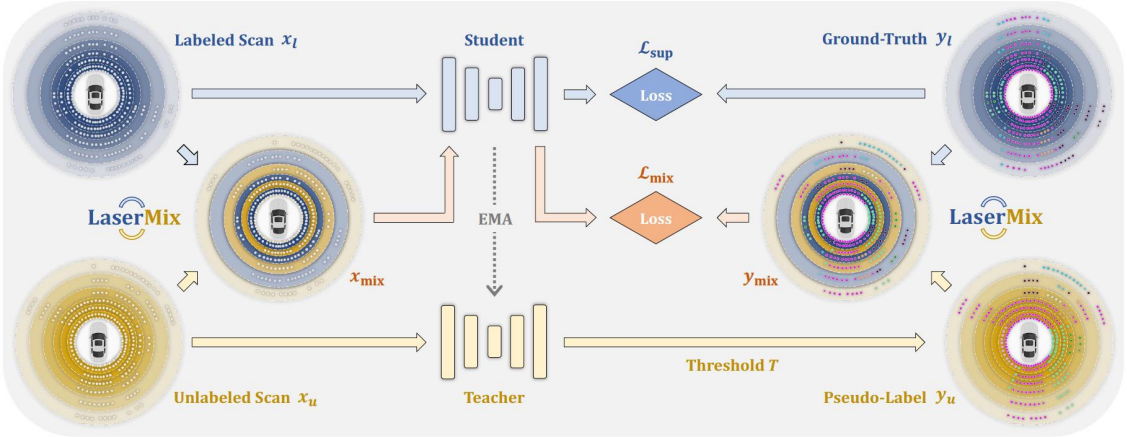
Also contains spatial priors in scenes!

Y. Ouali, et al. Semi-supervised semantic segmentation with cross-consistency training, CVPR, 2020.

Z. Ke, et al. Guided collaborative training for pixel-wise semi-supervised learning, ECCV, 2020.

Ablation Study

#	$\mathcal{L}_{\text{mt}}$	$\mathcal{L}_{\text{mix}}$	SS	TS	1%	10%	20%	50%
(1)	✓				42.1	60.4	65.4	69.4
(2)		✓	✓		45.6	64.3	67.8	71.6
	✓	✓	✓		47.0	65.5	69.5	72.0
(3)		✓		✓	46.0	64.1	69.5	72.3
	✓	✓		✓	49.5	68.2	70.6	73.0

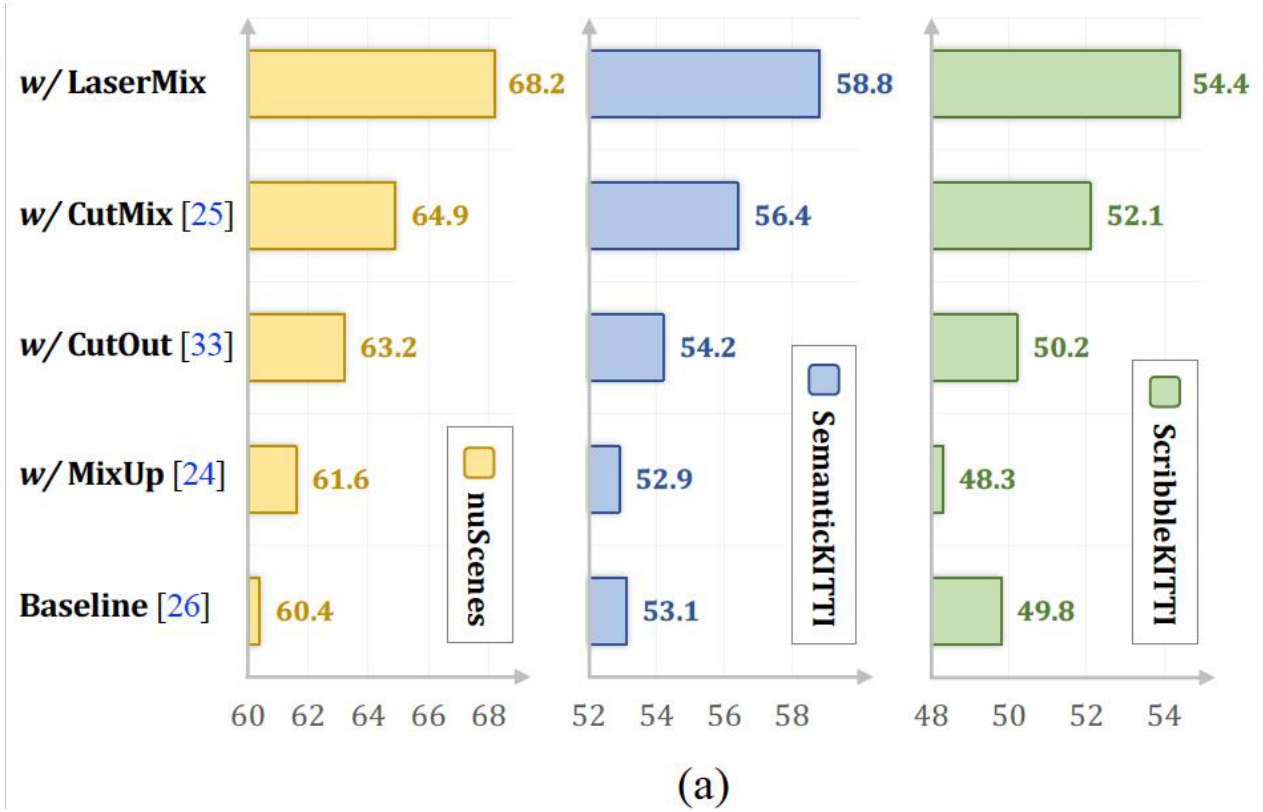


(1) Results of MeanTeacher.

(2) Results of LaserMix w/ student supervisions; much better than the counterpart.

(3) Results of LaserMix w/ teacher supervisions; much better than the counterpart.

Ablation Study



(a) Comparisons among different mixing techniques.

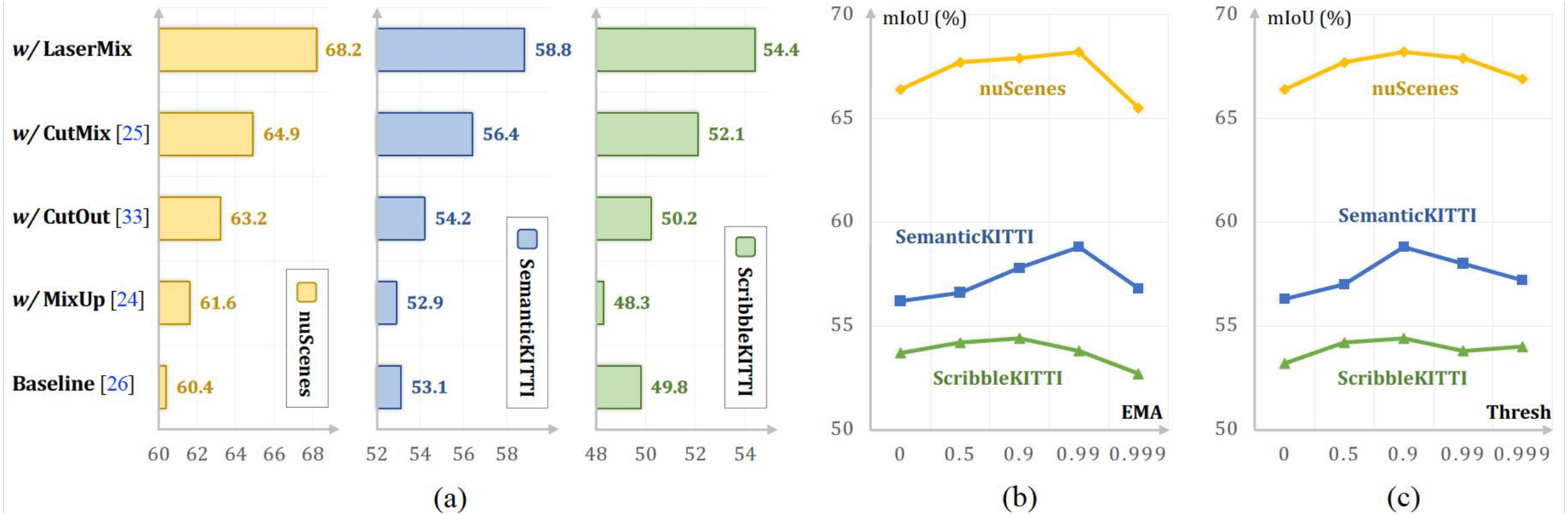
A. Nekrasov, et al. Mix3D: Out-of-context data augmentation for 3D scenes, 3DV, 2021.

S. Yun, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features, ICCV, 2019

T. DeVries and G. W. Taylor. Improved regularization of convolutional neural networks with cutout, arXiv, 2017

H. Zhang, et al. Mixup: Beyond empirical risk minimization, ICLR, 2018.

Ablation Study



(a) Comparisons among different mixing techniques. (b) EMA. (c) Confidence threshold.

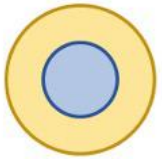
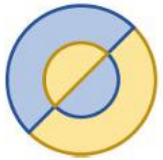
A. Nekrasov, et al. Mix3D: Out-of-context data augmentation for 3D scenes, 3DV, 2021.

S. Yun, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features, ICCV, 2019

T. DeVries and G. W. Taylor. Improved regularization of convolutional neural networks with cutout, arXiv, 2017

H. Zhang, et al. Mixup: Beyond empirical risk minimization, ICLR, 2018.

Ablation Study

Baseline	$(1\alpha, 2\phi)$	$(1\alpha, 3\phi)$	$(1\alpha, 4\phi)$	$(1\alpha, 5\phi)$	$(1\alpha, 6\phi)$
					
60.4	$63.5_{(+3.1)}$	$65.2_{(+4.8)}$	$66.5_{(+6.1)}$	$66.2_{(+5.8)}$	$65.4_{(+5.0)}$
$(2\alpha, 1\phi)$	$(2\alpha, 2\phi)$	$(2\alpha, 3\phi)$	$(2\alpha, 4\phi)$	$(2\alpha, 5\phi)$	$(2\alpha, 6\phi)$
					
$61.5_{(+1.1)}$	$63.3_{(+2.9)}$	$65.9_{(+5.5)}$	$66.1_{(+5.7)}$	$66.7_{(+6.3)}$	$65.3_{(+4.9)}$
$(3\alpha, 1\phi)$	$(3\alpha, 2\phi)$	$(3\alpha, 3\phi)$	$(3\alpha, 4\phi)$	$(3\alpha, 5\phi)$	$(3\alpha, 6\phi)$
					
$60.9_{(+0.6)}$	$64.2_{(+3.8)}$	$65.9_{(+5.5)}$	$66.3_{(+5.9)}$	$66.0_{(+5.6)}$	$65.2_{(+4.8)}$
$(4\alpha, 1\phi)$	$(4\alpha, 2\phi)$	$(4\alpha, 3\phi)$	$(4\alpha, 4\phi)$	$(4\alpha, 5\phi)$	$(4\alpha, 6\phi)$
					
$60.9_{(+0.6)}$	$64.7_{(+4.3)}$	$65.3_{(+4.9)}$	$65.6_{(+5.2)}$	$65.7_{(+5.3)}$	$65.2_{(+4.8)}$

- **Inclination:**

$$\phi_i = \arctan\left(\frac{p_i^z}{\sqrt{(p_i^x)^2+(p_i^y)^2}}\right)$$

- **Depth:** $\rho_i = \sqrt{(p_i^x)^2+(p_i^y)^2}$

- **Azimuth:** $\alpha_i = \arctan\left(\frac{p_i^y}{p_i^x}\right)$

Public Resources

- Paper: https://arxiv.org/abs/2207.00026
- Code: https://github.com/ldkong1205/LaserMix
- Project Page: https://ldkong.com/LaserMix

Topic

1. Overview & Background

2. Data-Efficient 3D Perception

3. Robustness in 3D

4. Segment Any Point Cloud Sequences

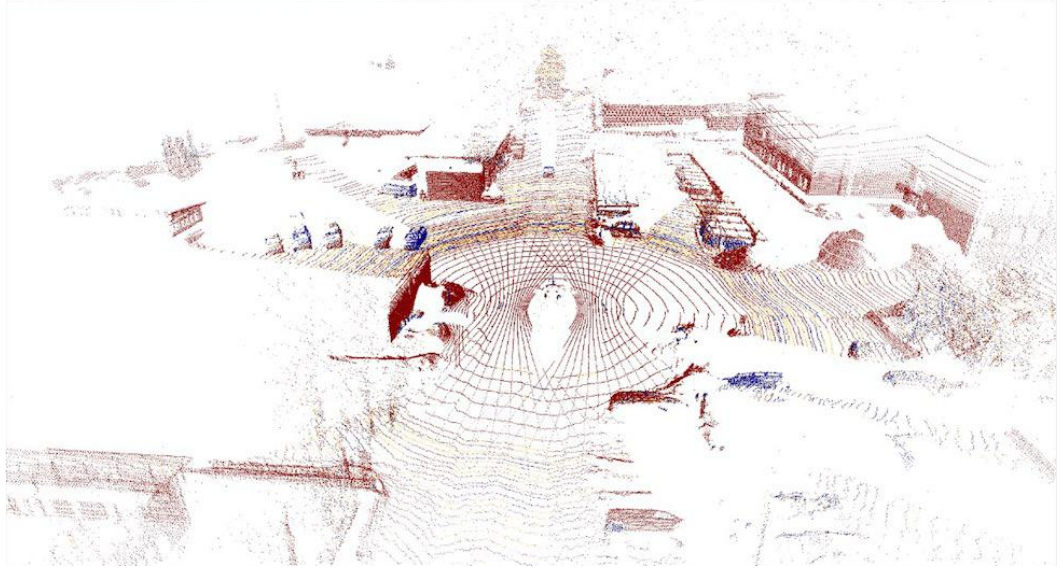
ICCV23
PARIS

Robo3D: Towards Robust and Reliable 3D Perception against Corruptions

Lingdong Kong[1,2,*], Youquan Liu[1,3,*], Xin Li[1,4,*], Runnan Chen[1,5], Wenwei Zhang[1,6]
Jiawei Ren[6], Liang Pan[6], Kai Chen[1], Ziwei Liu[6]

[1]Shanghai AI Lab [2]NUS [3]Hochschule Bremerhaven [4]ECNU [5]HKU [6]S-Lab, NTU

Complex Driving Environment



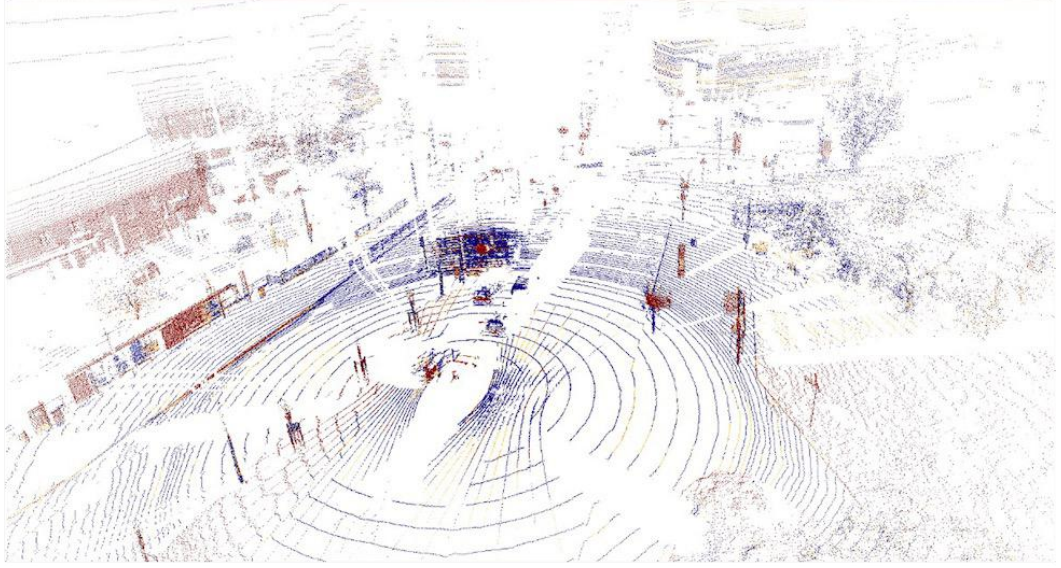


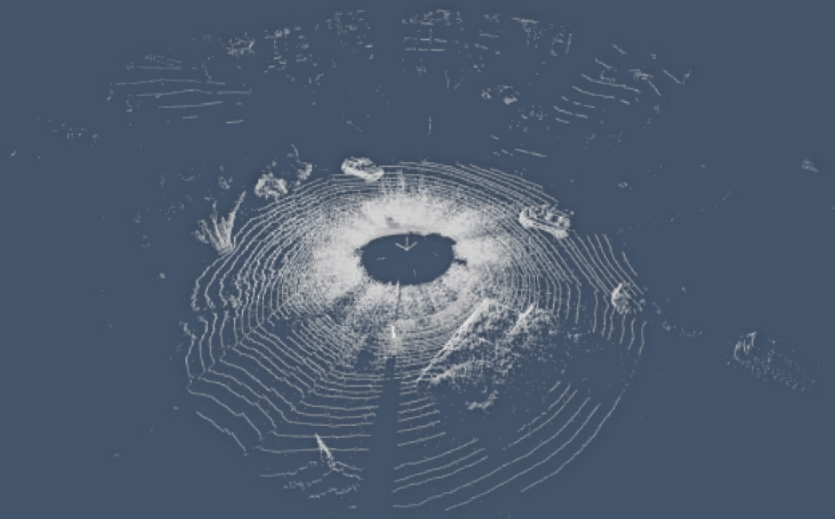
Image credit: https://zod.zenseact.com

Robo3D

ICCV23
PARIS

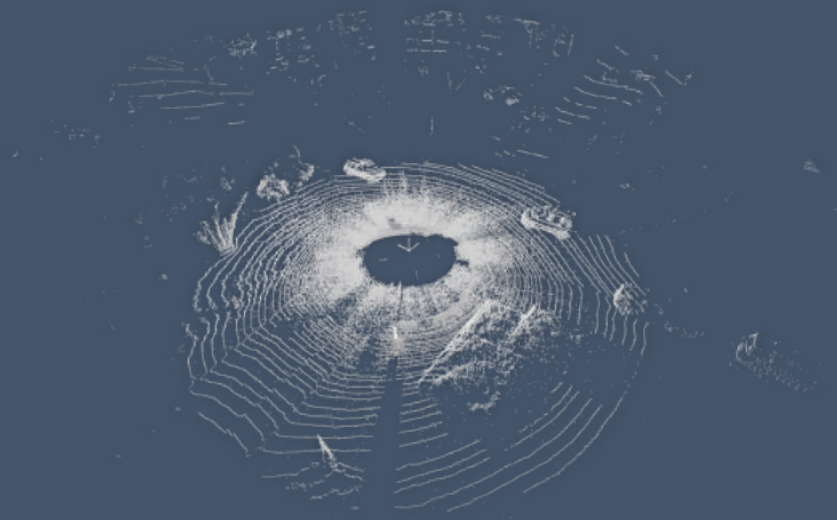
TL;DR

- We introduce Robo3D, the first systematically-designed robustness evaluation suite for LiDAR-based 3D perception under corruptions and sensor failures



Robo3D

ICCV23
PARIS

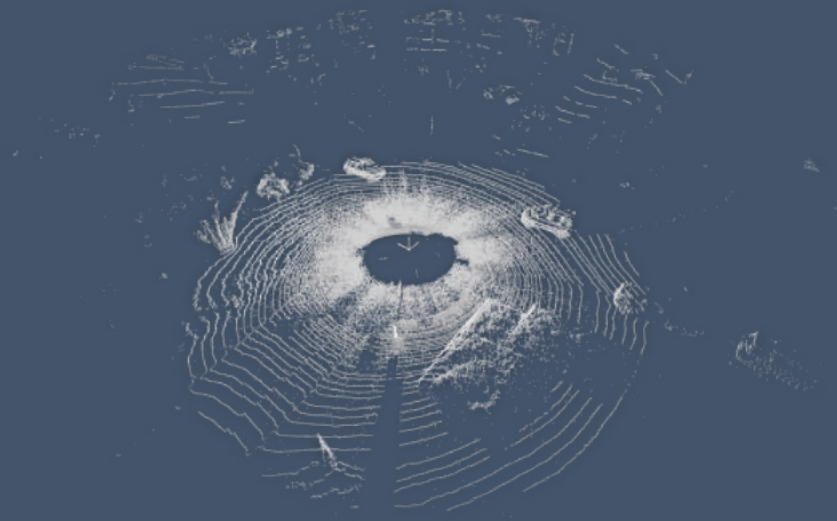


TL;DR

- We introduce Robo3D, the first systematically-designed robustness evaluation suite for LiDAR-based 3D perception under corruptions and sensor failures
- We benchmarked 34 perception models for LiDAR-based semantic segmentation and object detection tasks, on their robustness against corruptions

Robo3D

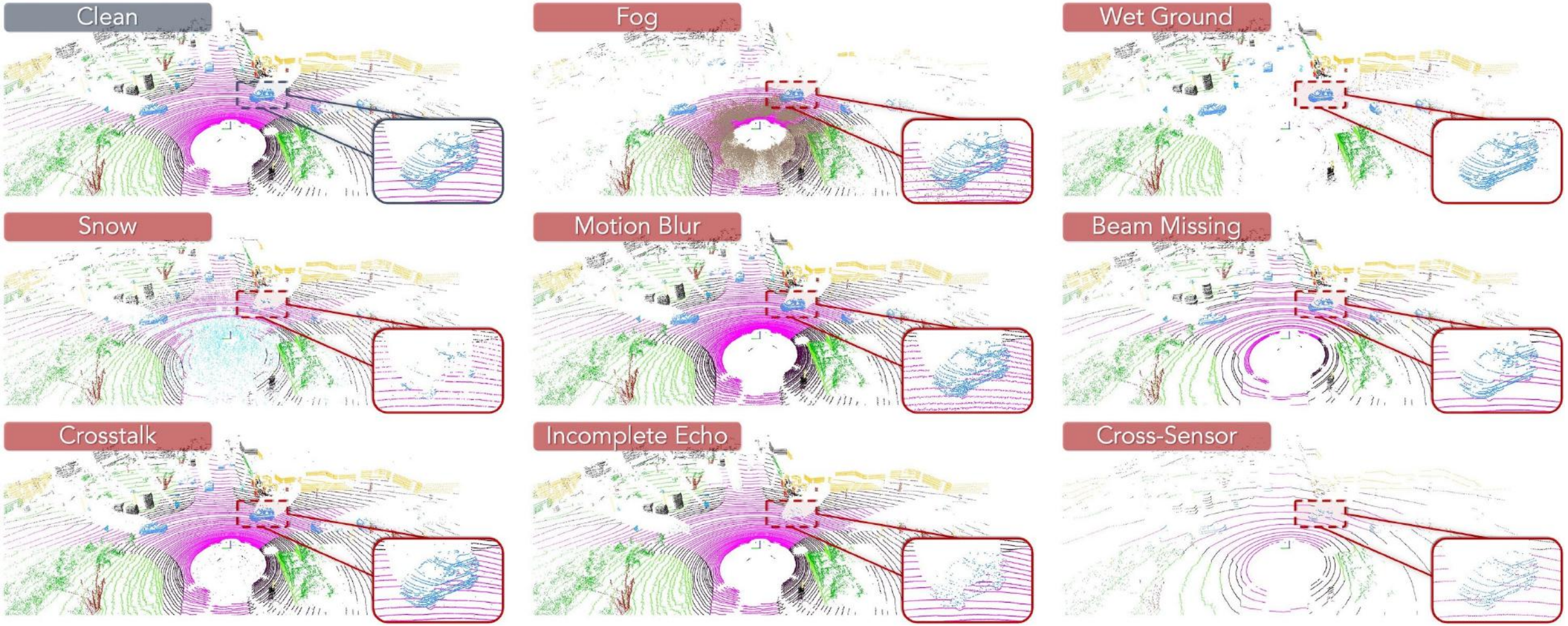
ICCV23
PARIS



TL;DR

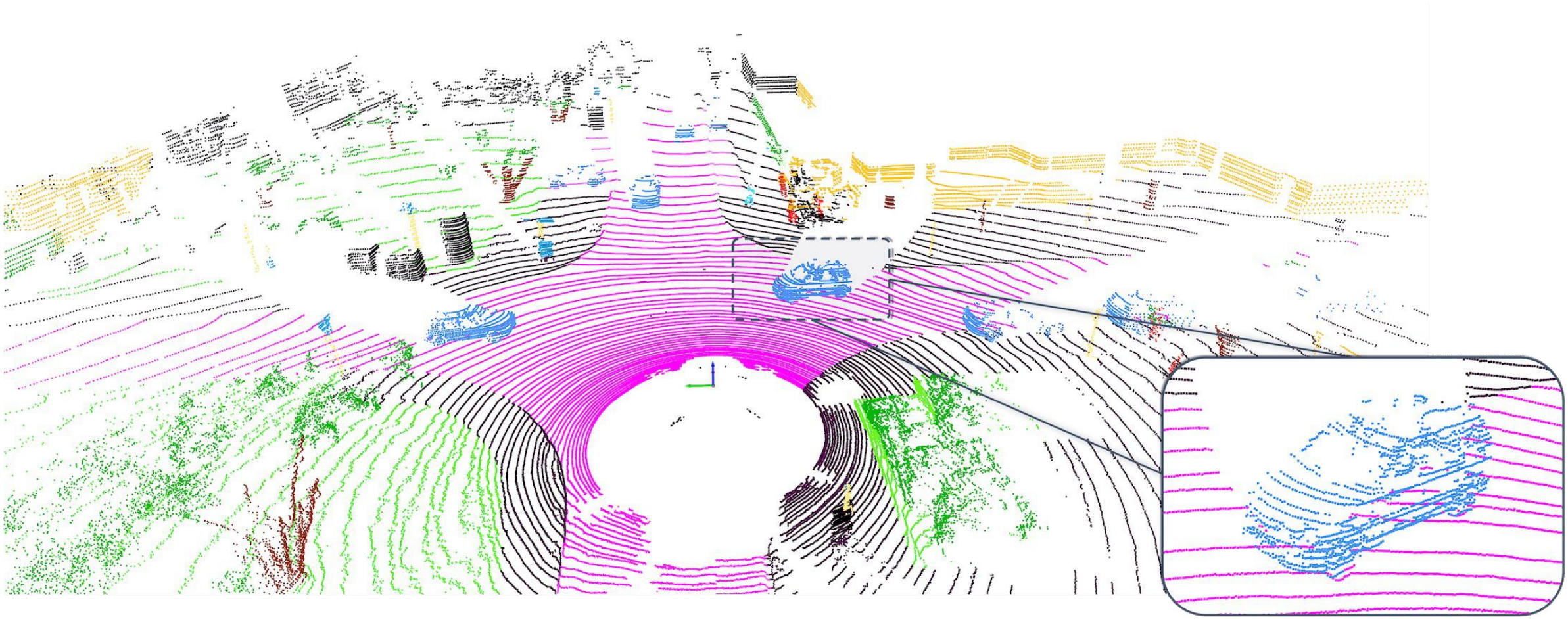
- We introduce Robo3D, the first systematically-designed robustness evaluation suite for LiDAR-based 3D perception under corruptions and sensor failures
- We benchmarked 34 perception models for LiDAR-based semantic segmentation and object detection tasks, on their robustness against corruptions
- Based on our observations, we draw in-depth discussions on the receipt of designing more robust and reliable 3D perception models

Robo3D: Taxonomy

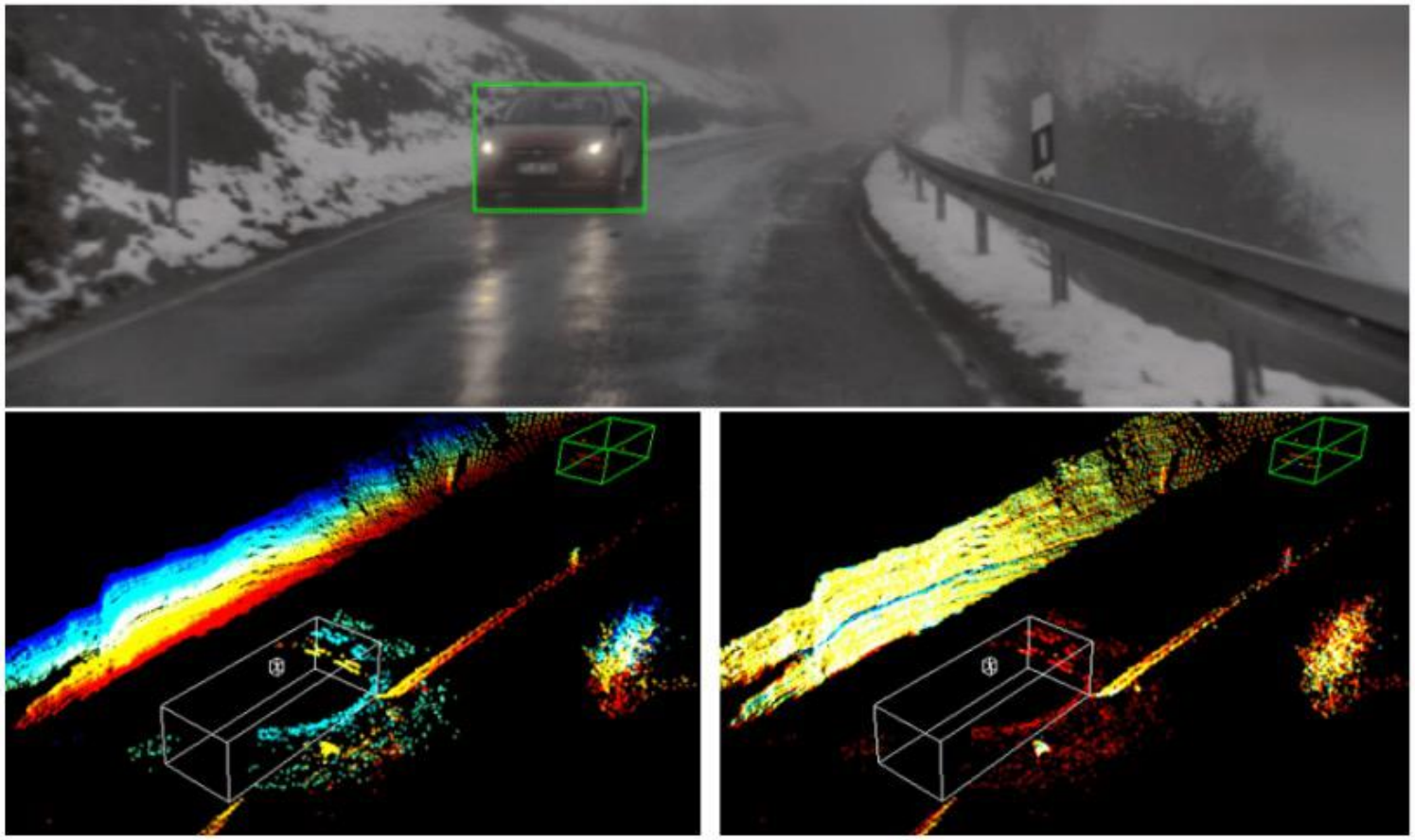


*More examples at: https://ldkong.com/Robo3D

3D Corruptions: Clean



3D Corruptions: Fog

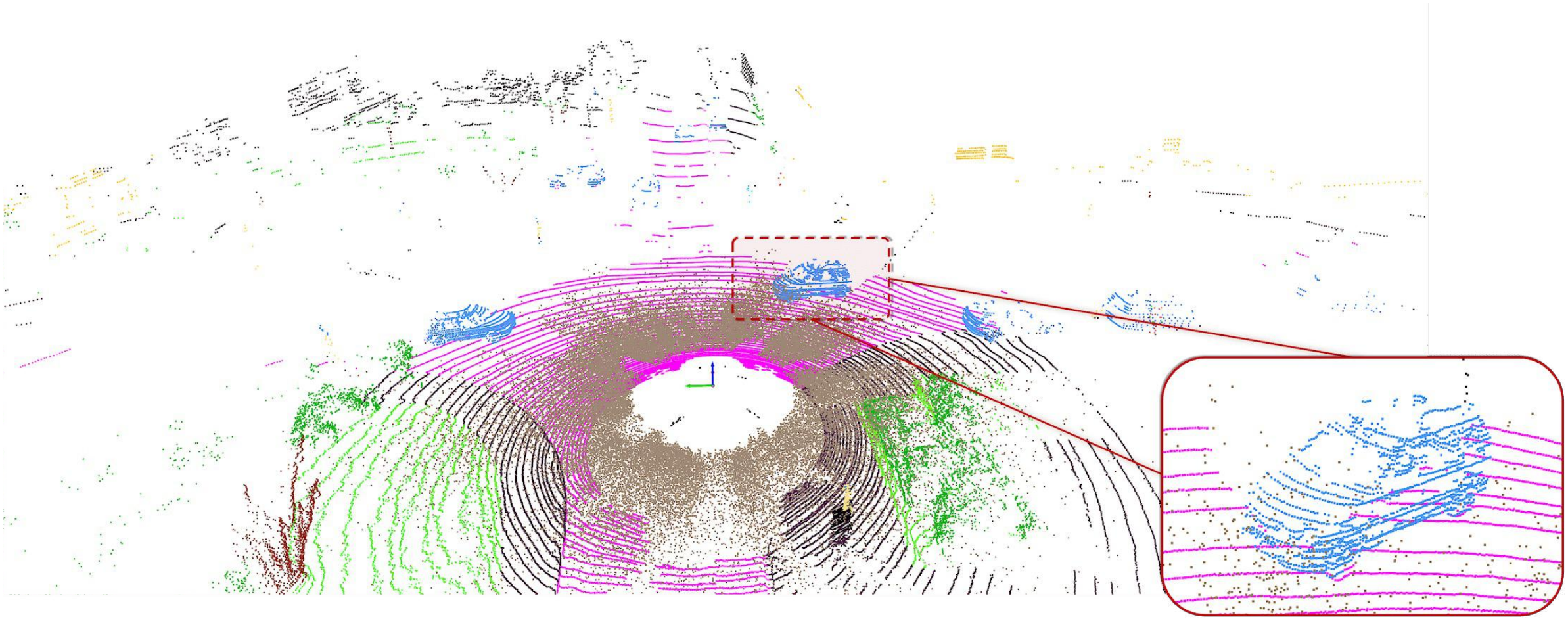


Fog Simulation

- The foggy weather mainly causes back-scattering and attenuation of LiDAR pulse transmissions
- This results in severe shifts of both range and intensity for the points in a LiDAR point cloud

M. Hahner, et al. Fog simulation on real LiDAR point clouds for 3D object detection in adverse weather, ICCV, 2021.

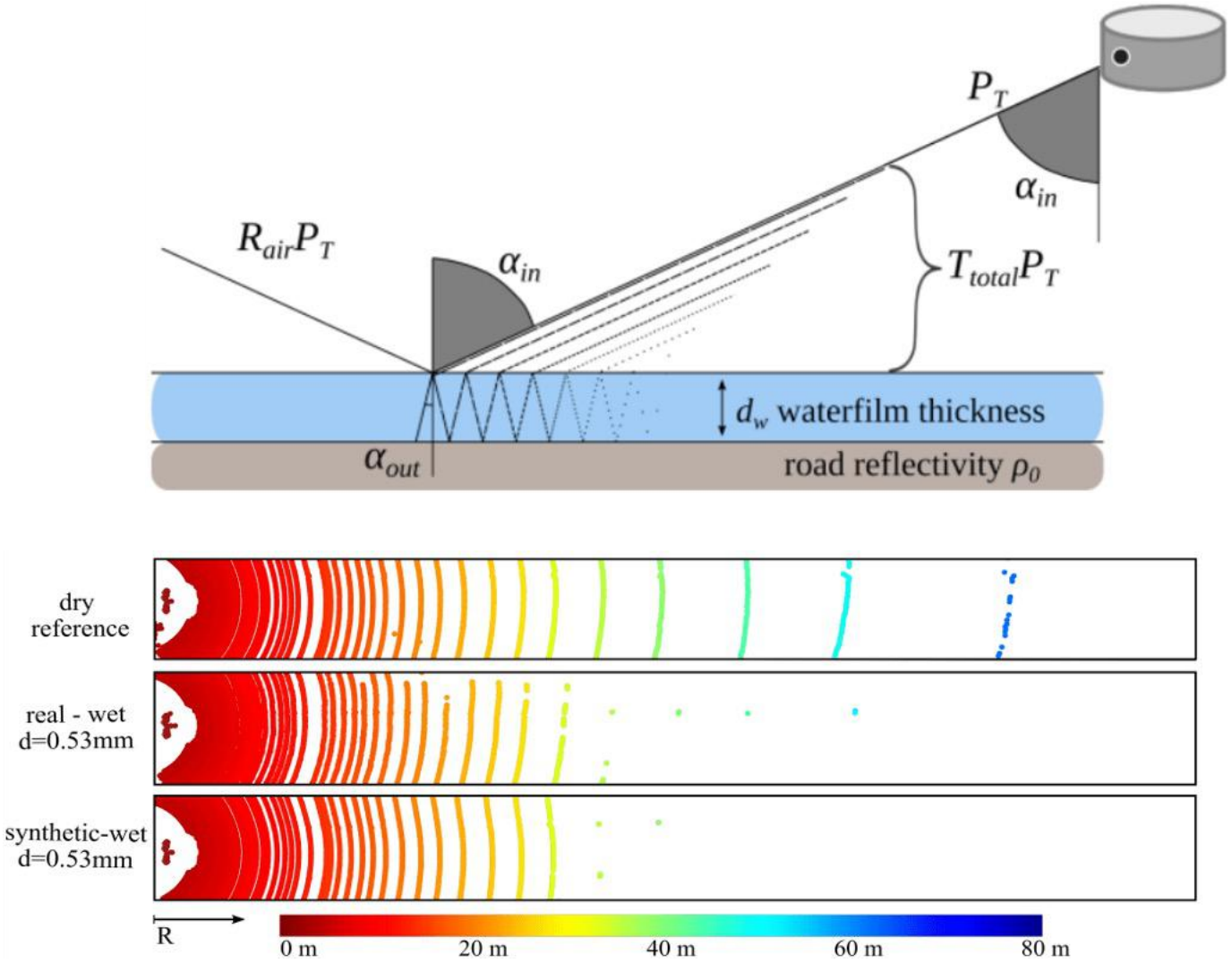
3D Corruptions: Fog



3D Corruptions: Wet Ground

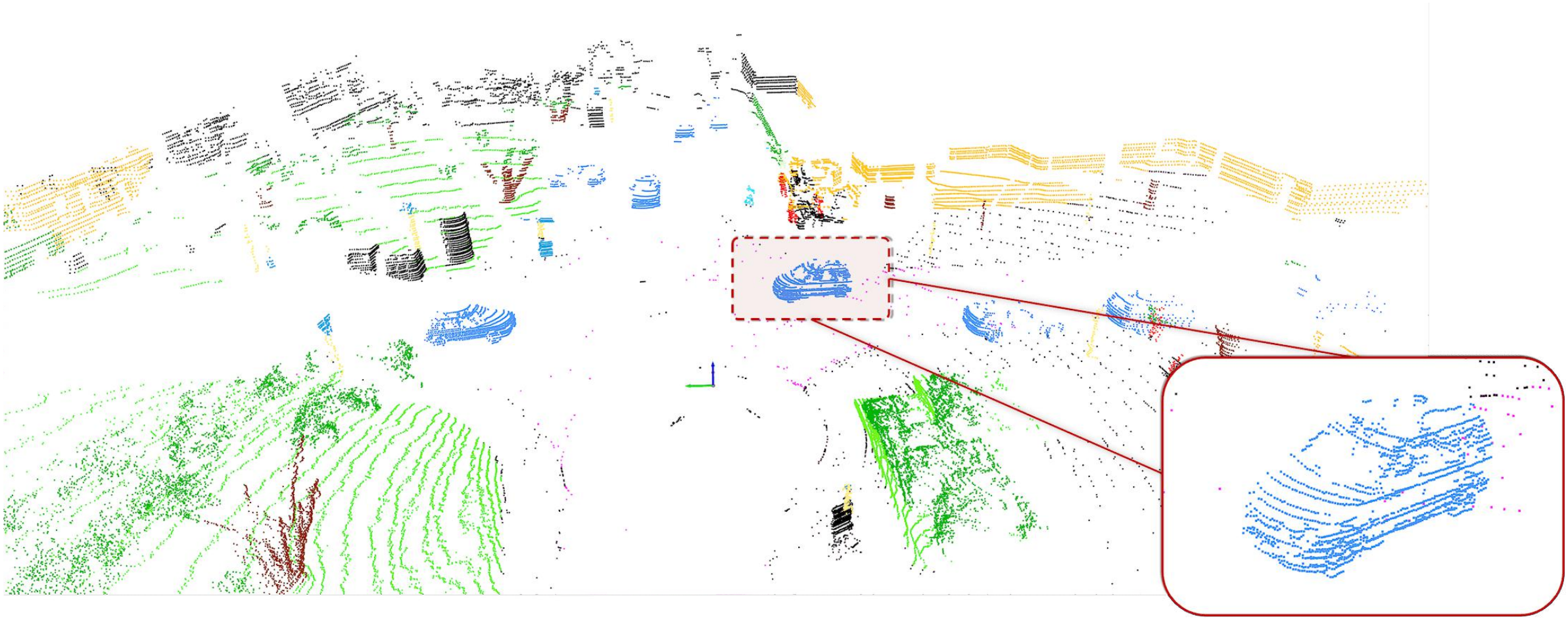
Wet Ground Simulation

- The emitted laser pulses from the LiDAR sensor tend to lose certain amounts of energy when hitting wet surface
- This cause significantly attenuated laser echoes depending on the water height and mirror refraction rate

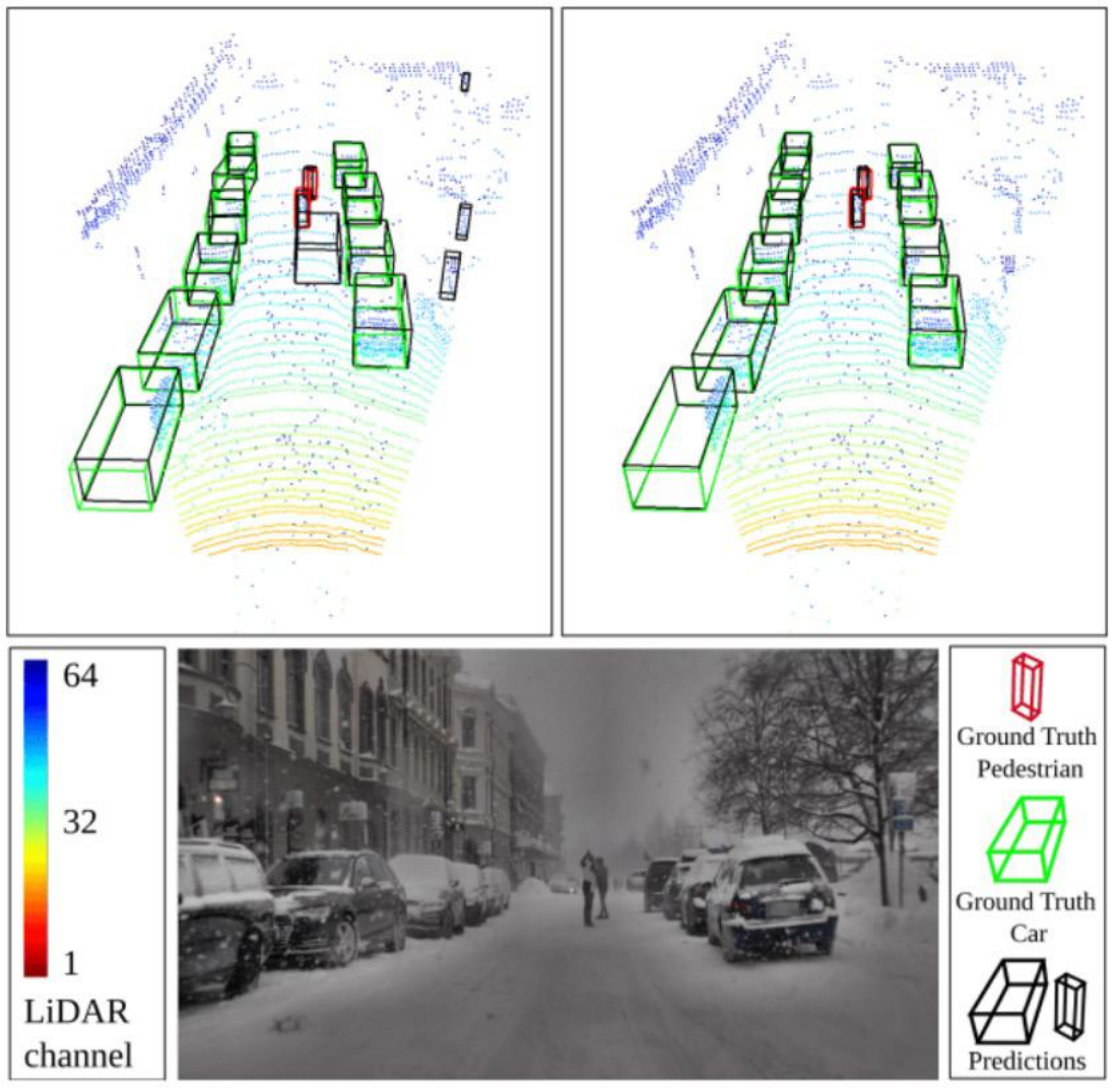


M. Hahner, et al. LiDAR snowfall simulation for robust 3D object detection, CVPR, 2022.

3D Corruptions: Wet Ground

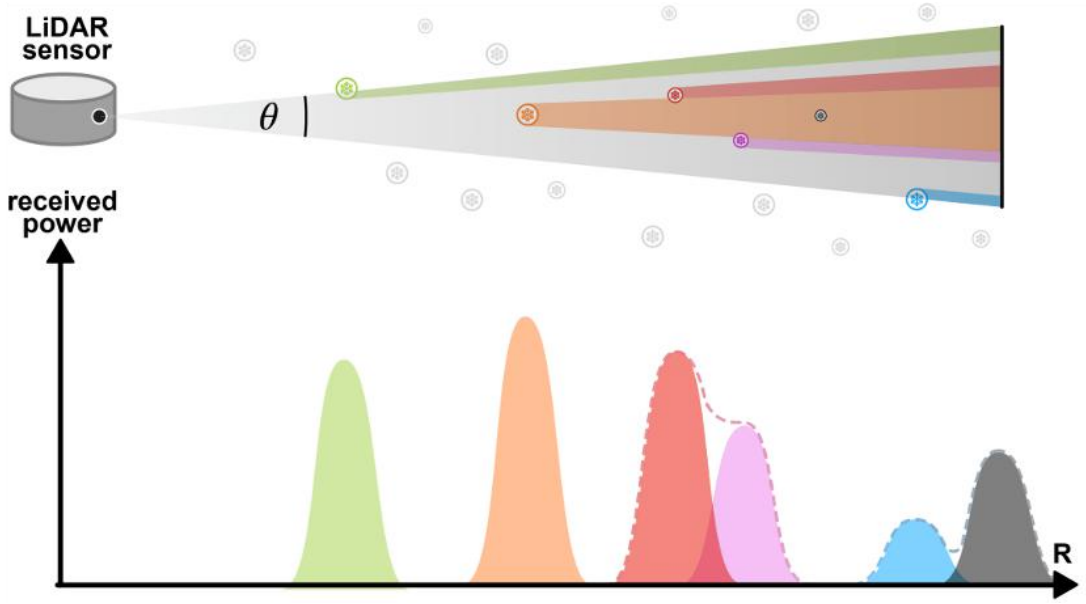


3D Corruptions: Snow



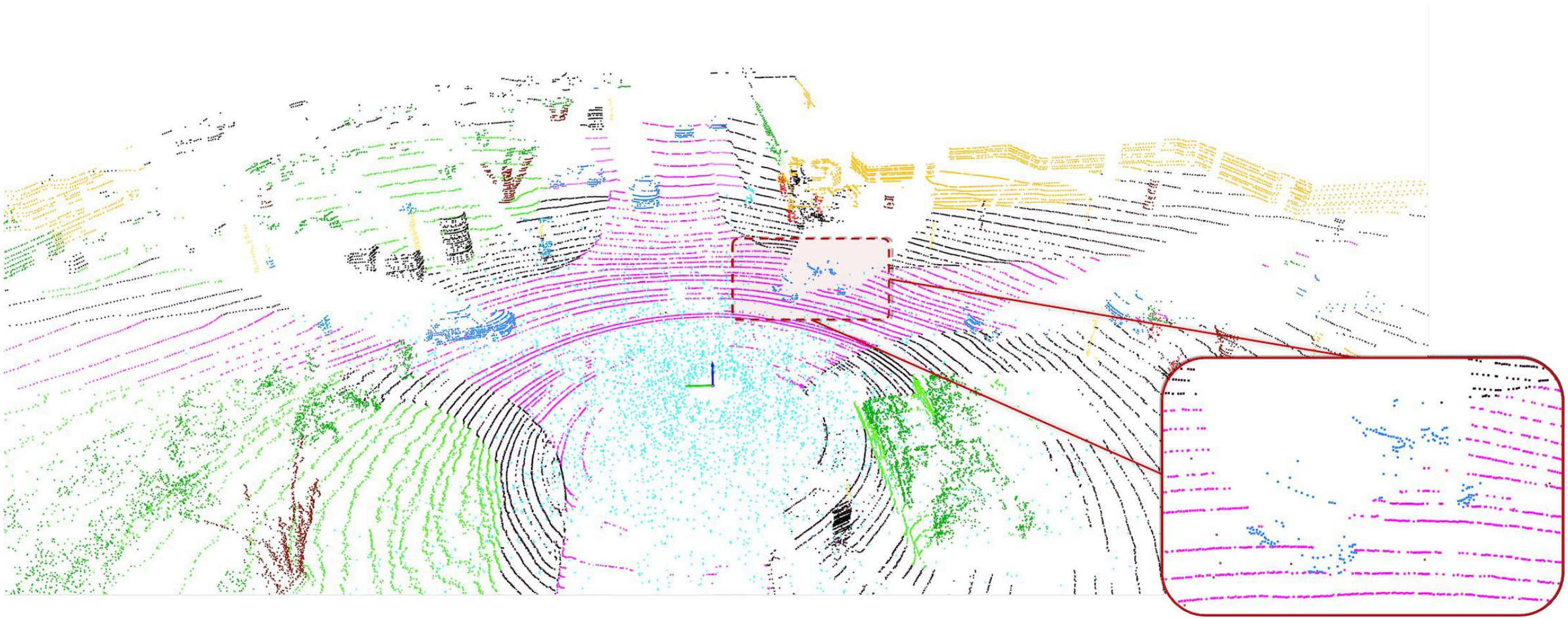
Snow Simulation

- For the laser beam in snowy weather, the set of particles in the air will intersect with it and derive the angle of the beam cross-section that is reflected by each particle, taking potential occlusions into account



M. Hahner, et al. LiDAR snowfall simulation for robust 3D object detection, CVPR, 2022.

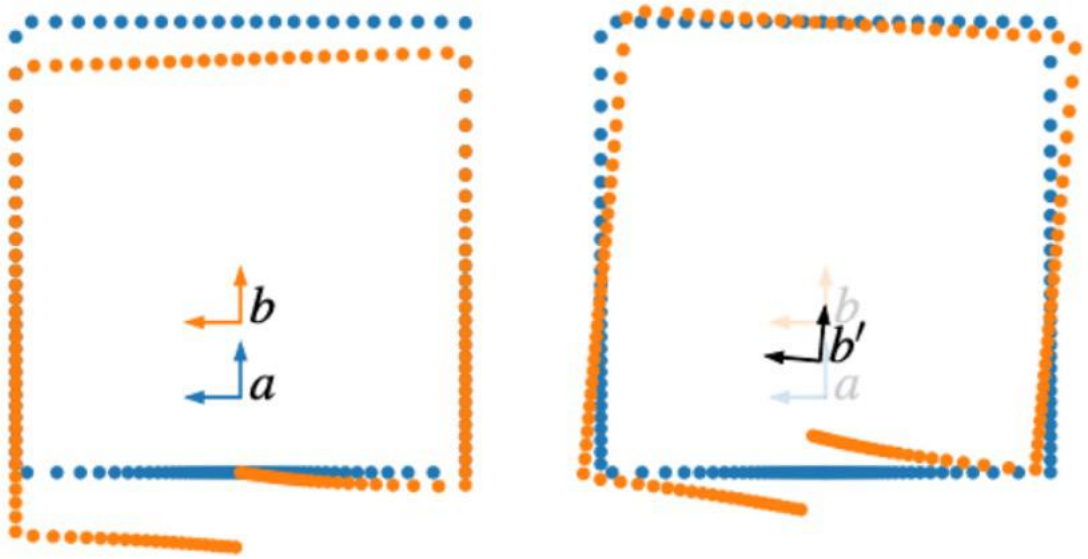
3D Corruptions: Snow

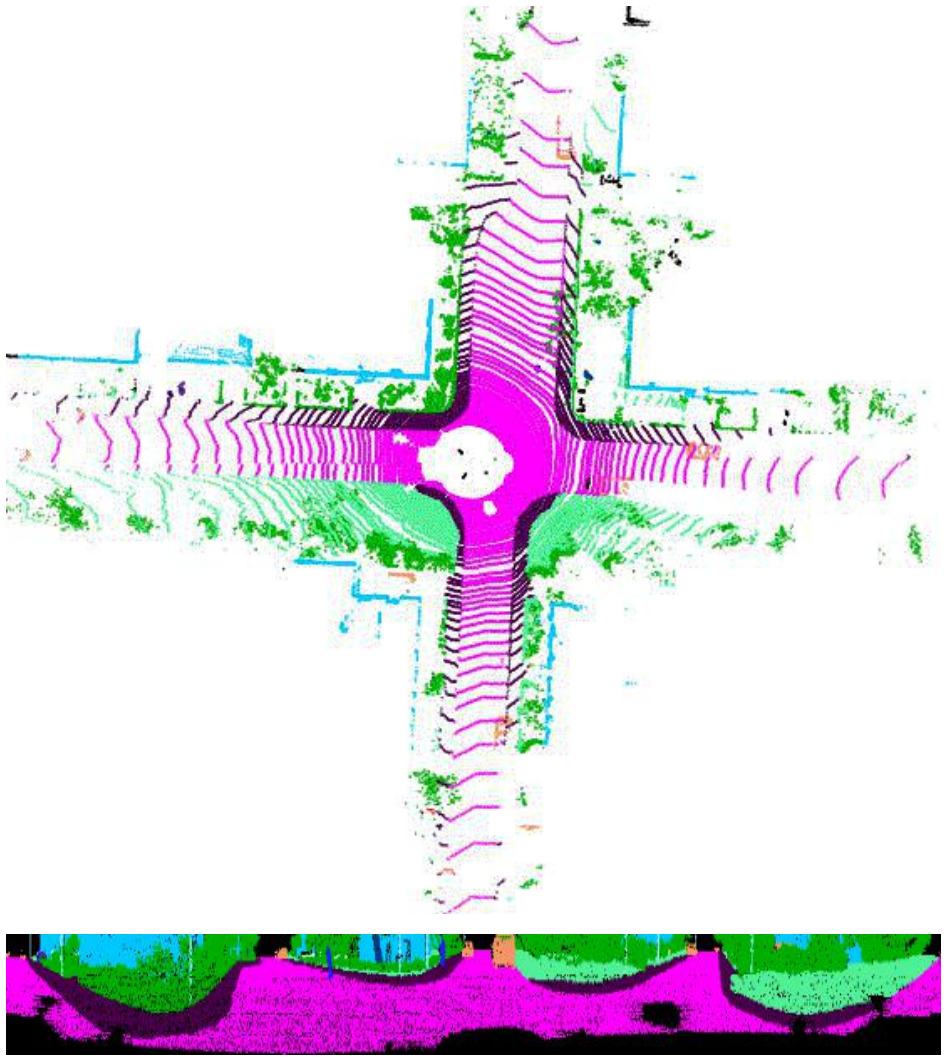


3D Corruptions: Motion Blur

Motion Blur Simulation

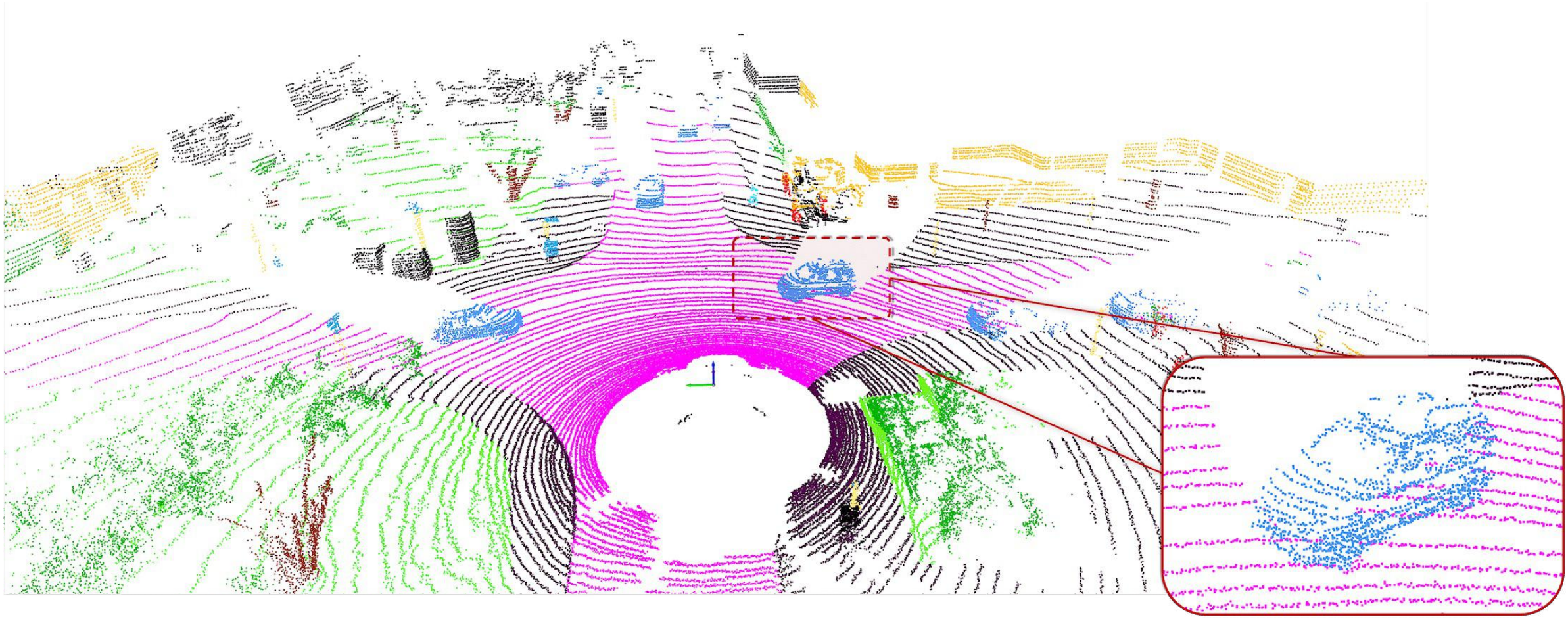
- LiDAR is often mounted on the rooftop or side of the vehicle and inevitably suffers from the blur caused by vehicle movement, especially on bumpy surfaces or during U-turning



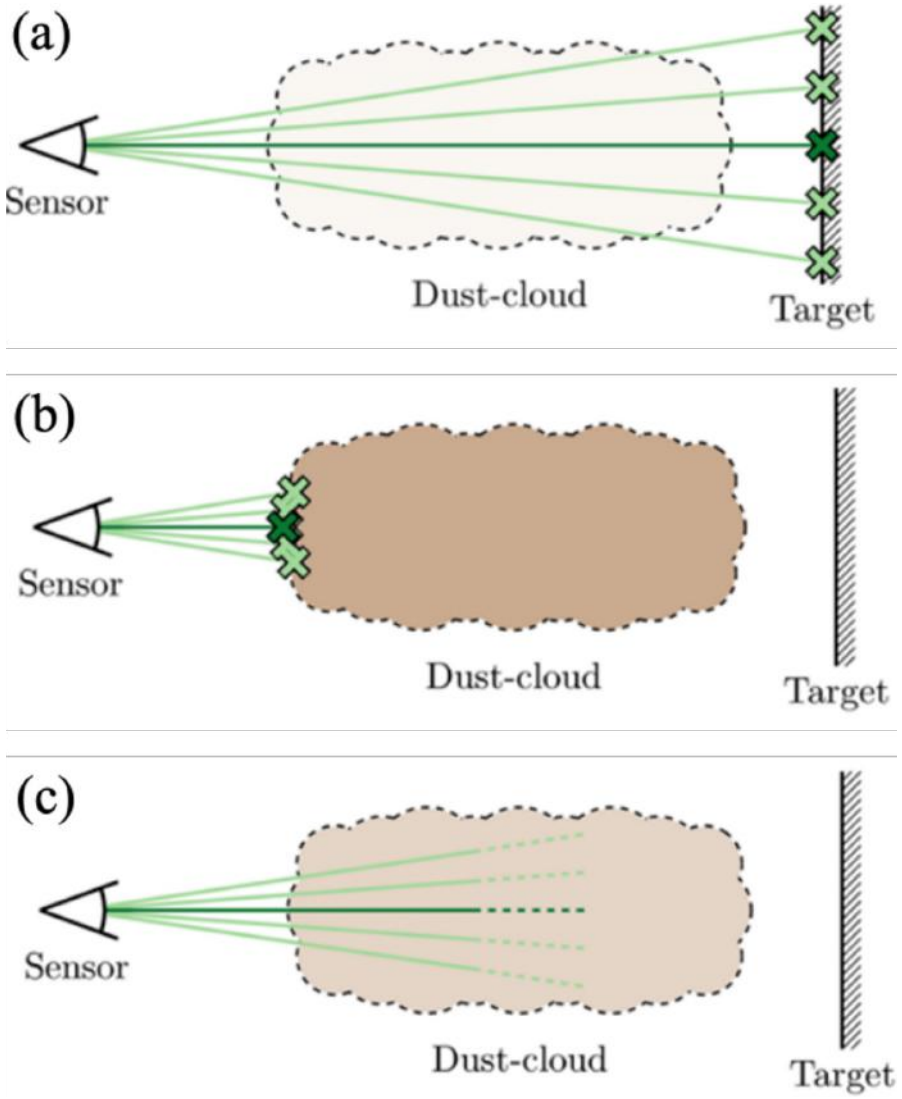


S. P. Deschenes, et al. LiDAR scan registration robust to extreme motions, ICRV, 2021.

3D Corruptions: Motion Blur



3D Corruptions: Beam Missing

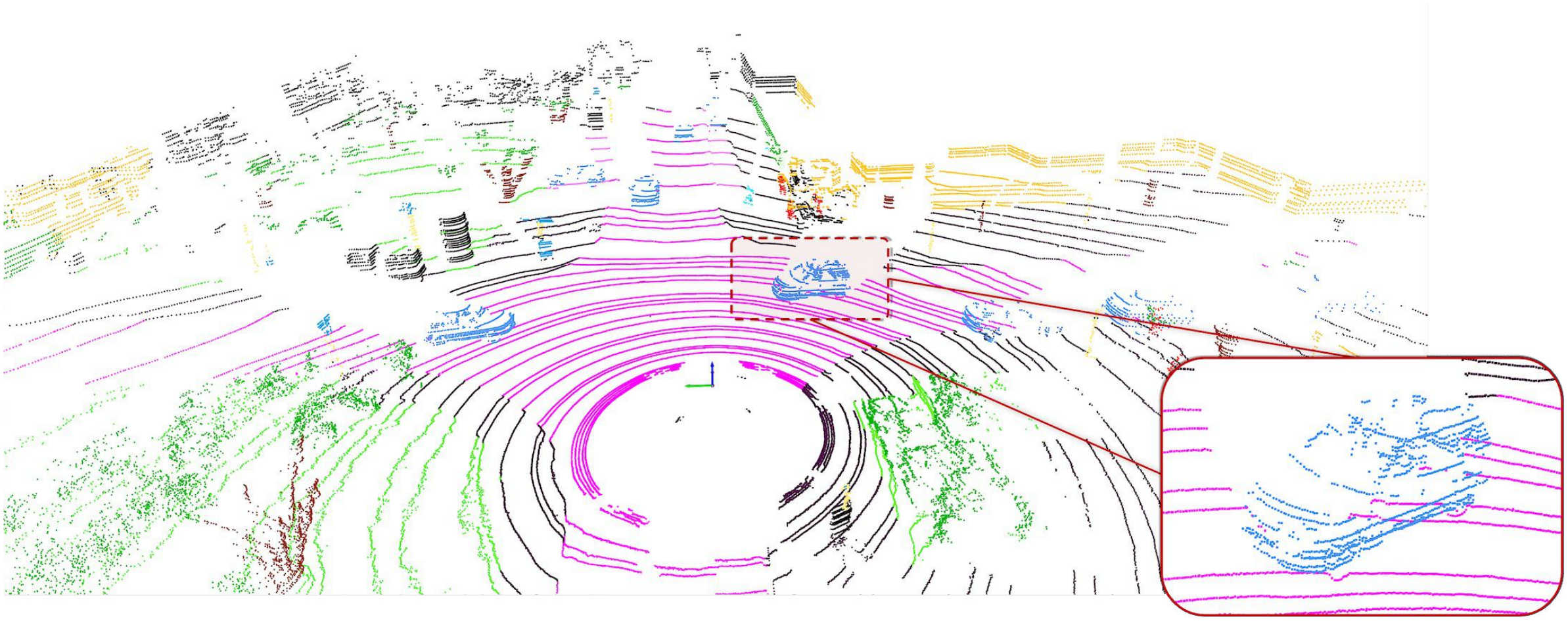


Beam Missing Simulation

- The dust and insect tend to form agglomerates in front of the LiDAR surface and will not likely disappear without human intervention, such as drying and cleaning
- This type of occlusion causes zero readings on masked areas and results in the loss of certain light impulses

T. G. Phillips, et al. When the dust settles: The four behaviors of LIDAR in the presence of fine airborne particulates, JFR, 2017.

3D Corruptions: Beam Missing

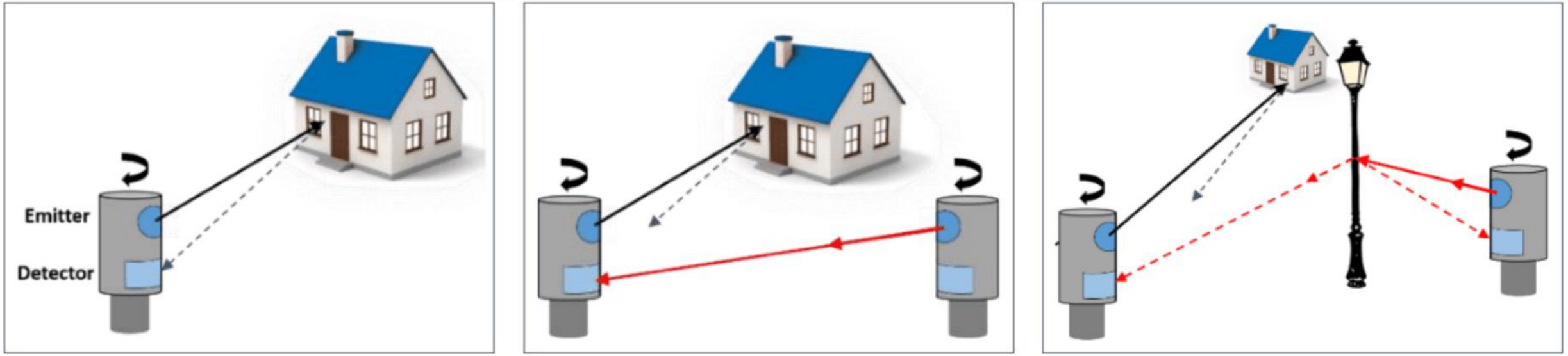


3D Corruptions: Crosstalk

Crosstalk Simulation

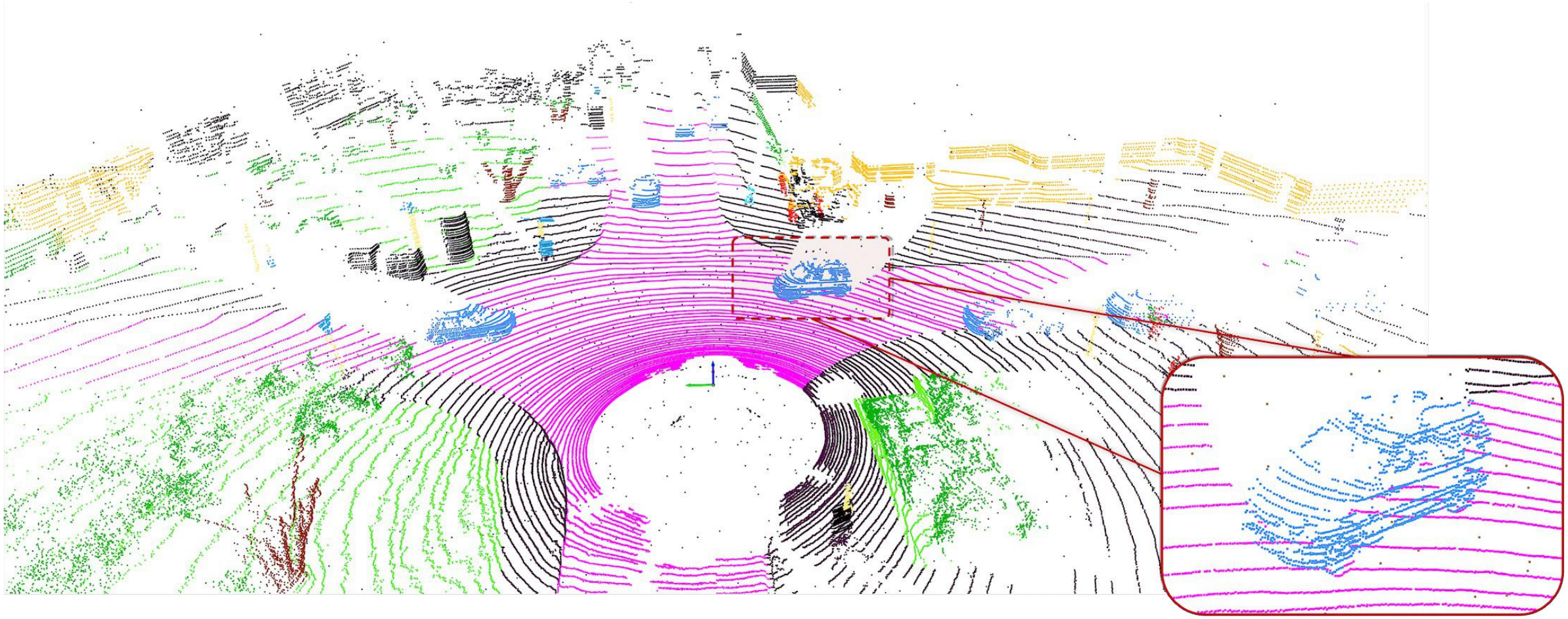
- The time-of-flight of light impulses from one sensor on a vehicle might interfere with impulses from other sensors from other vehicles within a similar frequency range

- Such a crosstalk phenomenon often creates noisy points within the mid-range areas in between two (or multiple) sensors



A. L. Diehm, et al. Mitigation of crosstalk effects in multi-LiDAR configurations, Electro-Optical Remote Sensing XII, 2018.

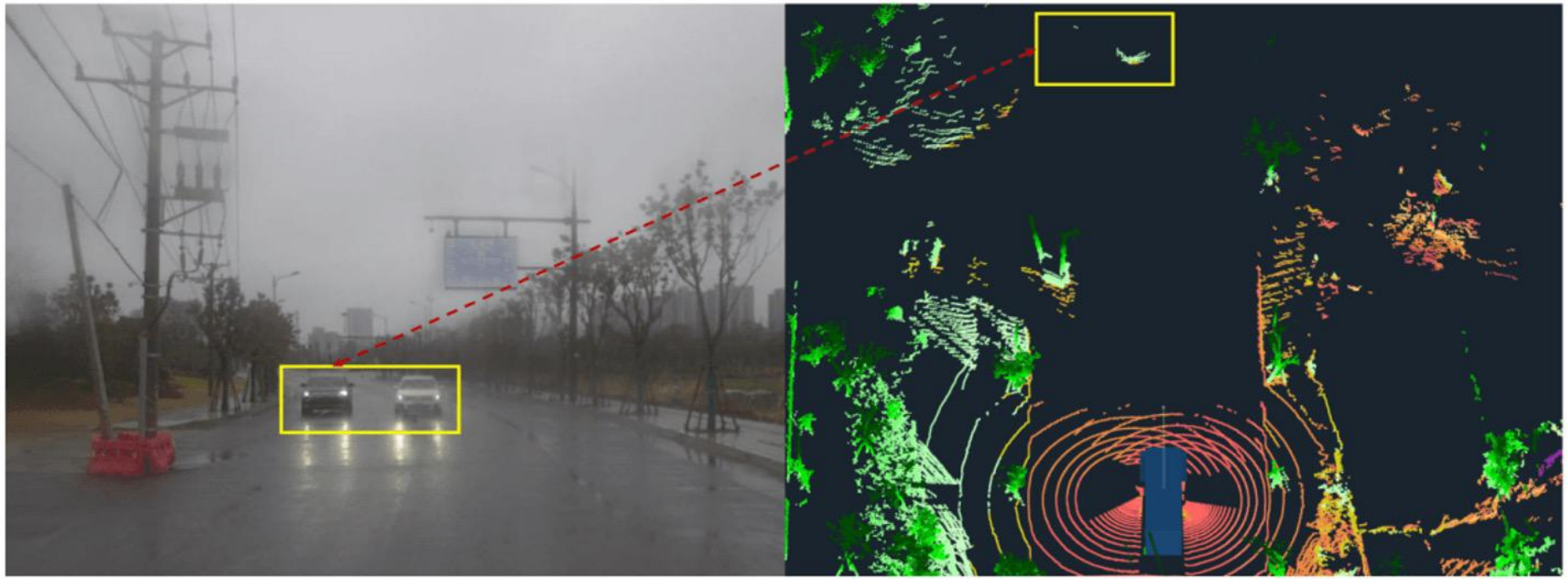
3D Corruptions: Crosstalk



3D Corruptions: Incomplete Echo

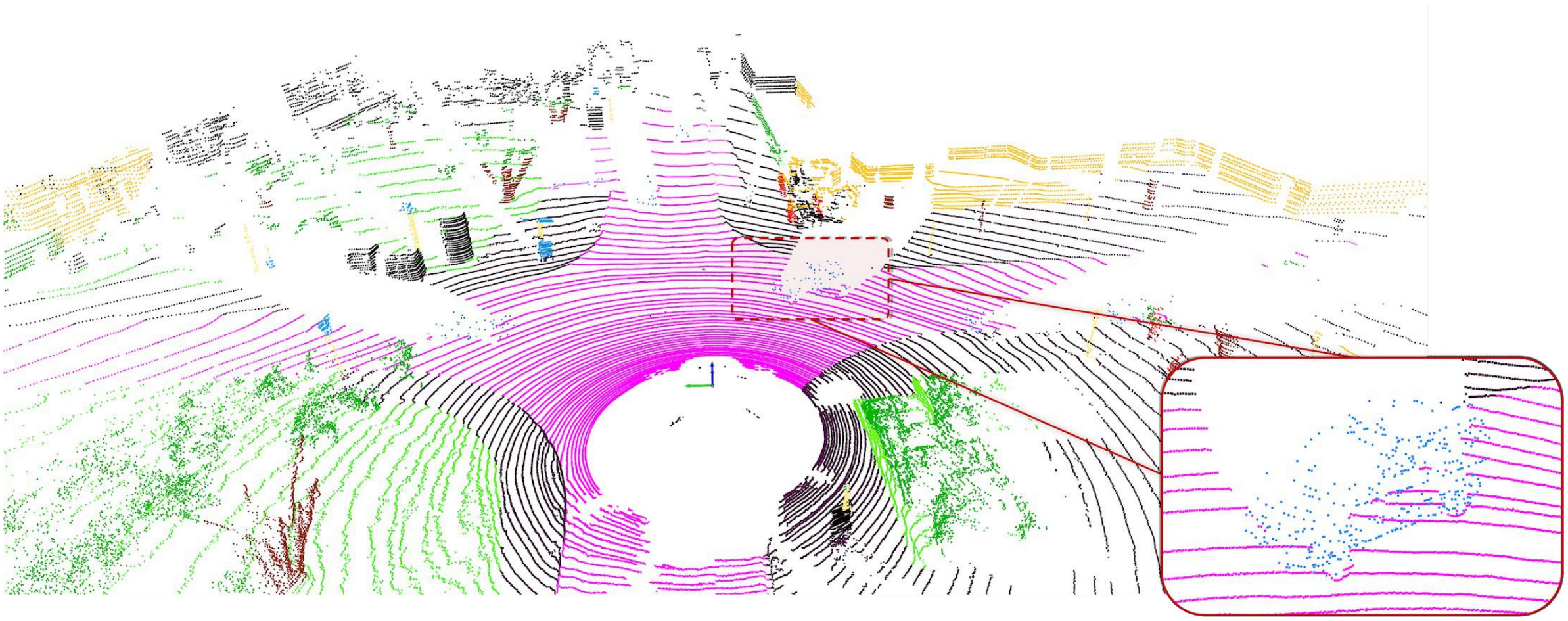
Incomplete Echo Simulation

- The near-infrared spectrum of the laser pulse emitted from the LiDAR sensor is vulnerable to vehicles or other instances with dark colors. The LiDAR readings are thus incomplete in such scan echoes, resulting in significant point miss detection

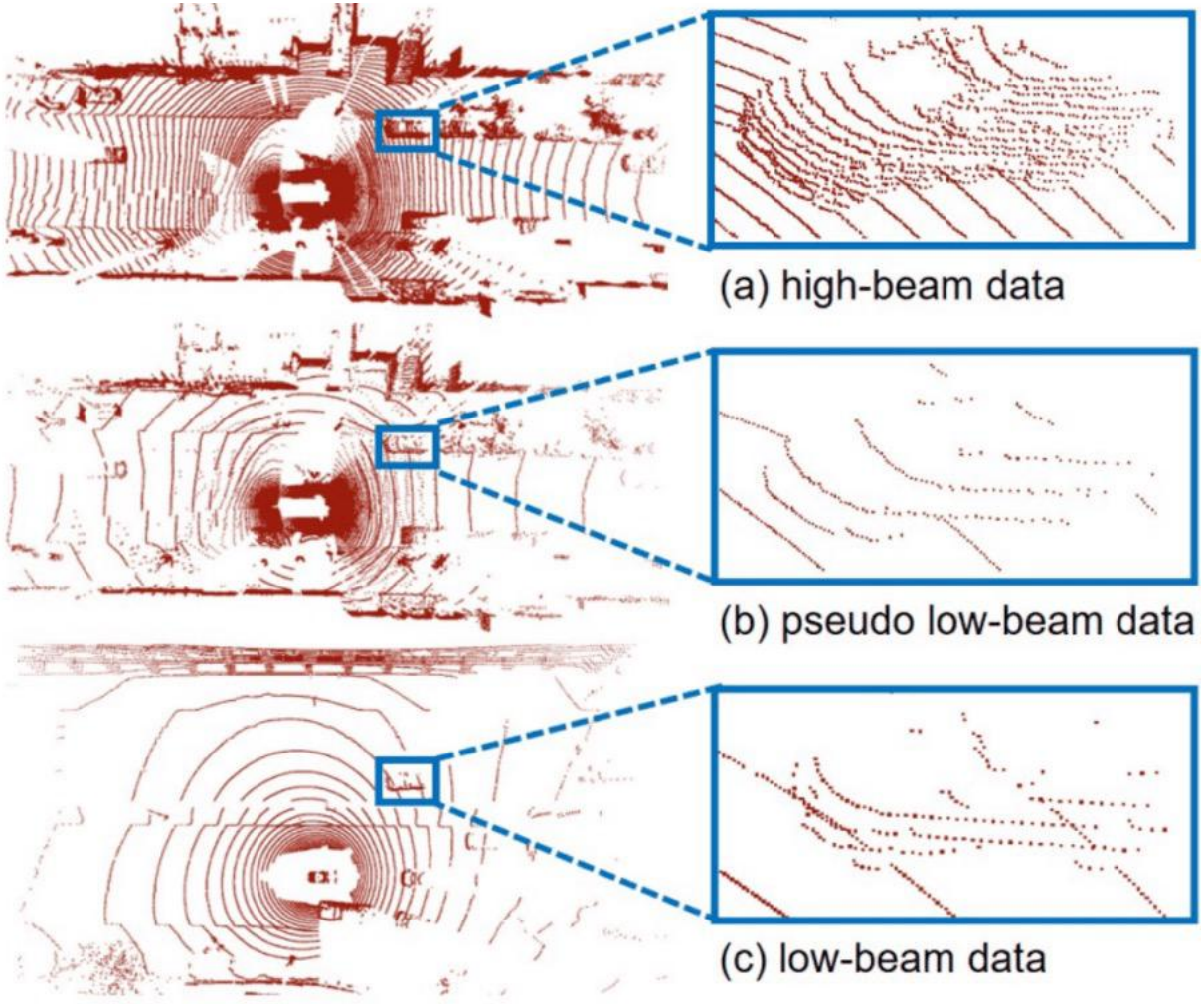


K. Yu, et al. Benchmarking the robustness of LiDAR-camera fusion for 3D object detection, arXiv, 2022.

3D Corruptions: Incomplete Echo



3D Corruptions: Cross-Sensor



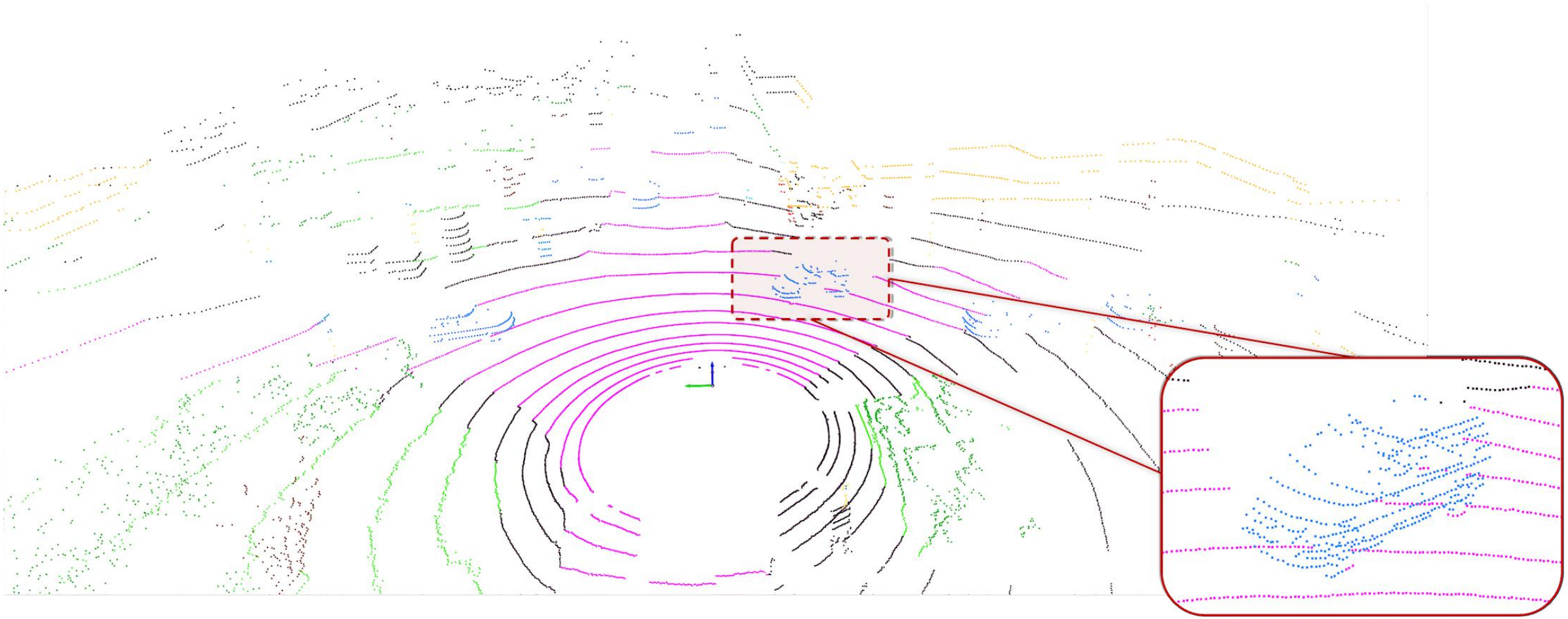
Cross-Sensor Simulation

- Due to the large variety of LiDAR sensor configurations (beam number, FOV, and sampling frequency), it is important to design robust 3D perception models that are capable of maintaining satisfactory performance under cross-device cases

- While previous works directly form such settings with two different datasets, the domain idiosyncrasy in between (e.g., different label mappings and data collection protocols) further hinders the direct robustness comparison

Y. Wei, et al. LiDAR distillation: Bridging the beam-induced domain gap for 3D object detection, ECCV, 2022.

3D Corruptions: Cross-Sensor



Robo3D: Statistics

Corruption Types

- Include 8 corruption types, each with 3 severity levels (Easy, Moderate, and Hard)

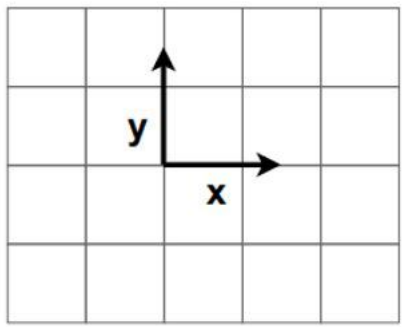
Datasets (6 different collections)

- LiDAR Semantic Segmentation: [1]SemanticKITTI-C, [2]nuScenes-C (Seg3D), [3]WOD-C (Seg3D)
- 3D Object Detection: [4]KITTI-C, [5]nuScenes-C (Det3D), [6]WOD-C (Det3D)

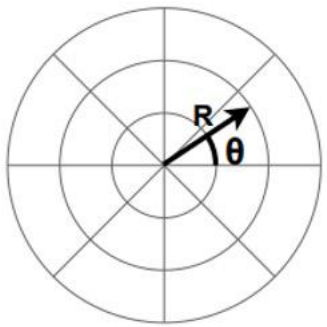
Model & Algorithm (34 perception models)

- LiDAR Semantic Segmentation: 22 segmentors
- 3D Object Detection: 12 detectors
- Data Augmentation: 3 augmentation techniques

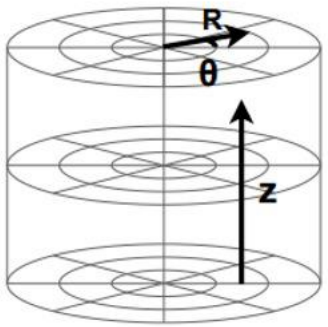
Robo3D: Representations



Cartesian Coordinates



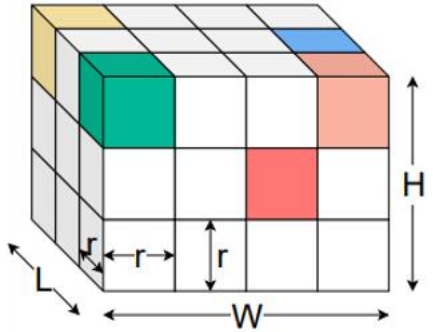
Polar Coordinates



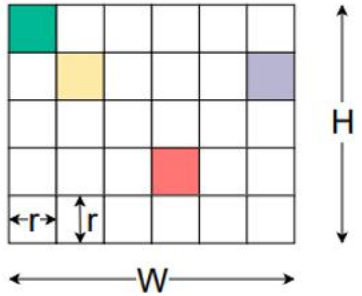
Cylinder Coordinates

Representation

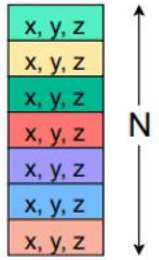
- **2D:** range view, birds eye view
- **3D:** cubic voxel, cylinder voxel



3D Voxels



2D Projection



1D Bag-of-Points

Operator

- **3D:** Conv3d, SparseConv, etc.
- **2D:** Conv2d, Linear, etc.
- **1D:** Conv1d, Linear, etc.

M. Uecker, et al. Analyzing deep learning representations of point clouds for real-time in-vehicle LiDAR perception, arXiv, 2022.

Robo3D: Metrics

Task-Specific Accuracy (Acc)

- LiDAR Semantic Segmentation: mean IoU (**mIoU**)
- 3D Object Detection: mean AP (**mAP**), nuScenes Detection Score (**NDS**)

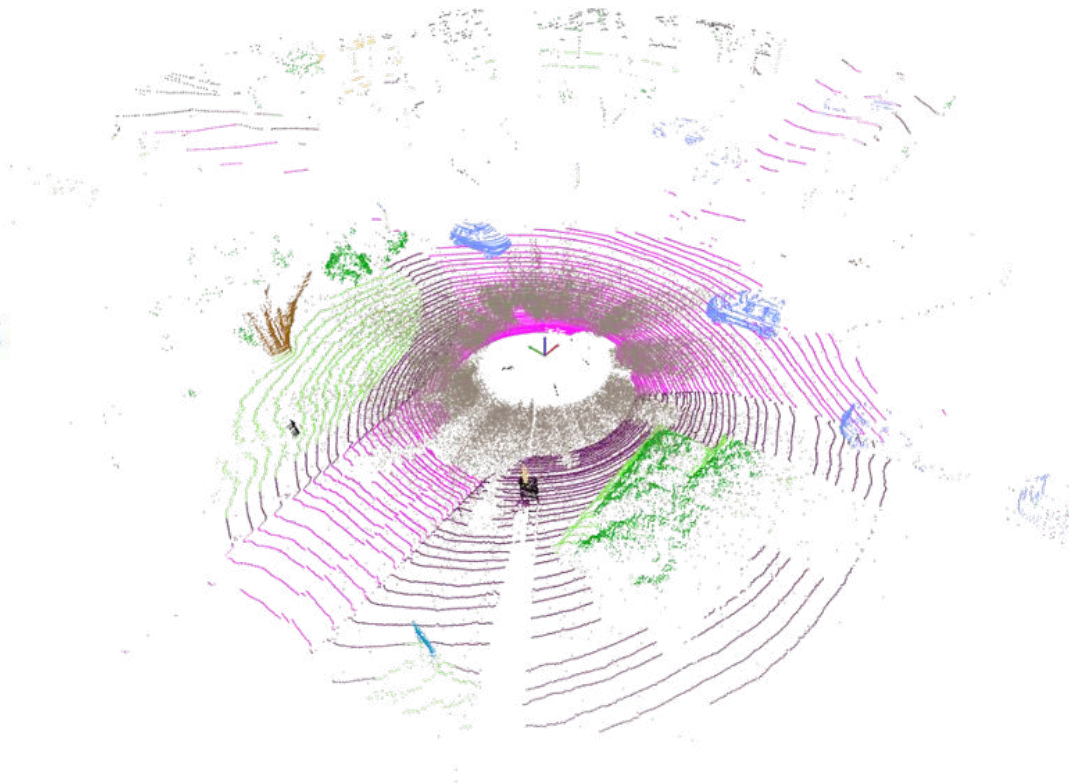
Robustness Metrics

- Mean Corruption Error (**mCE**)

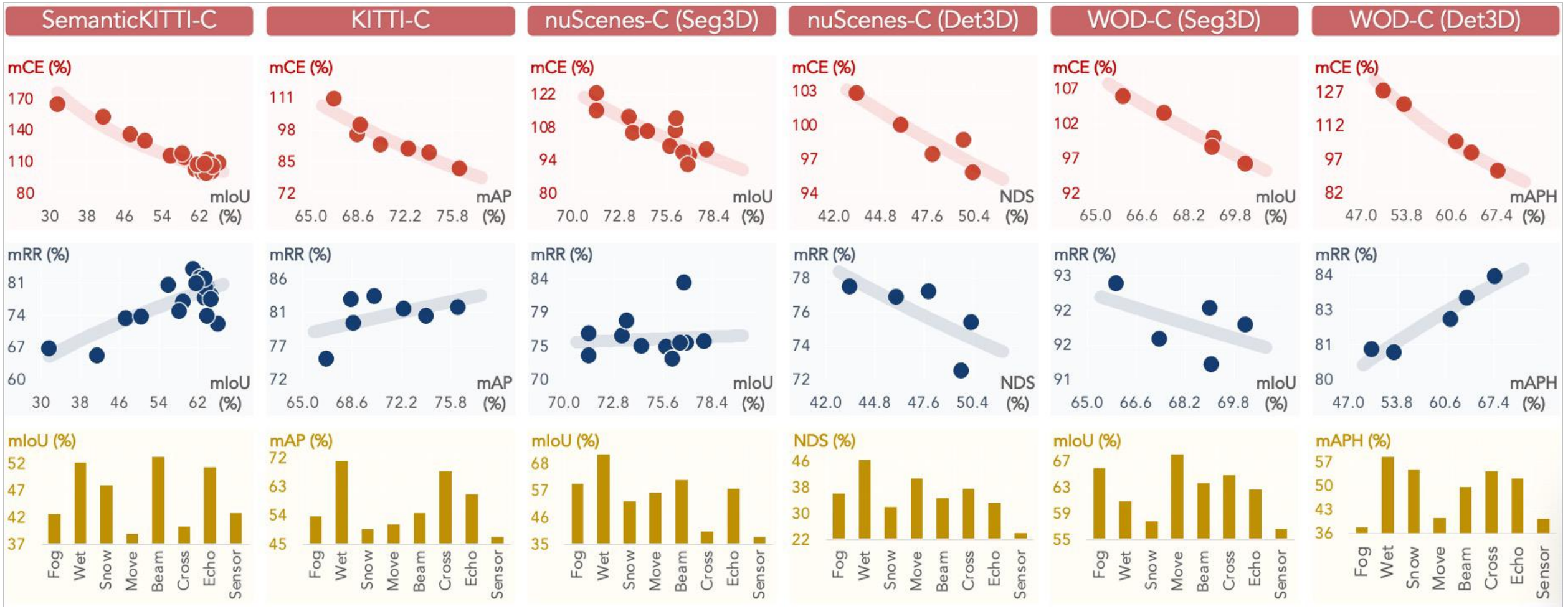
$$\mathrm{CE}_i = \frac{\sum_{l=1}^{3}(1 - \mathrm{Acc}_{i,l})}{\sum_{l=1}^{3}(1 - \mathrm{Acc}_{i,l}^{\mathrm{baseline}})}, \quad \mathrm{mCE} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{CE}_i$$

- Mean Resilience Rate (**mRR**)

$$\mathrm{RR}_i = \frac{\sum_{l=1}^{3}\mathrm{Acc}_{i,l}}{3 \times \mathrm{Acc}_{\mathrm{clean}}}, \quad \mathrm{mRR} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{RR}_i$$



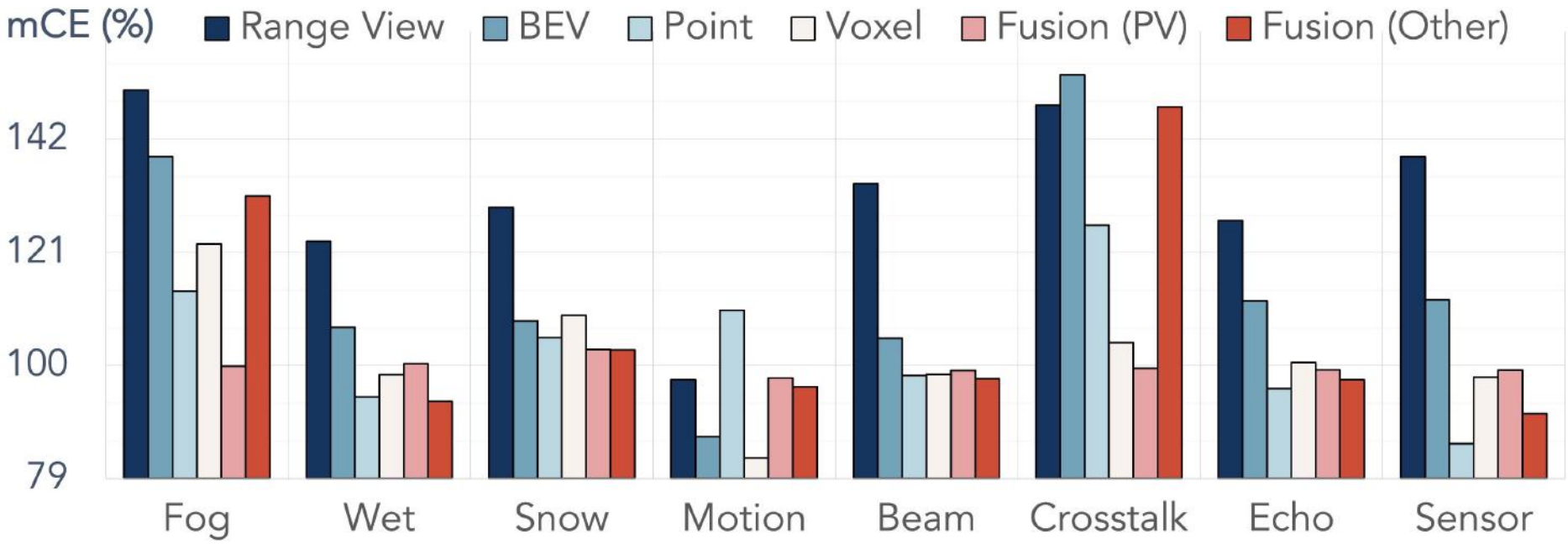
Robo3D: Benchmark Results



*More results and analysis at: https://github.com/ldkong1205/Robo3D

Robo3D: Key Observations

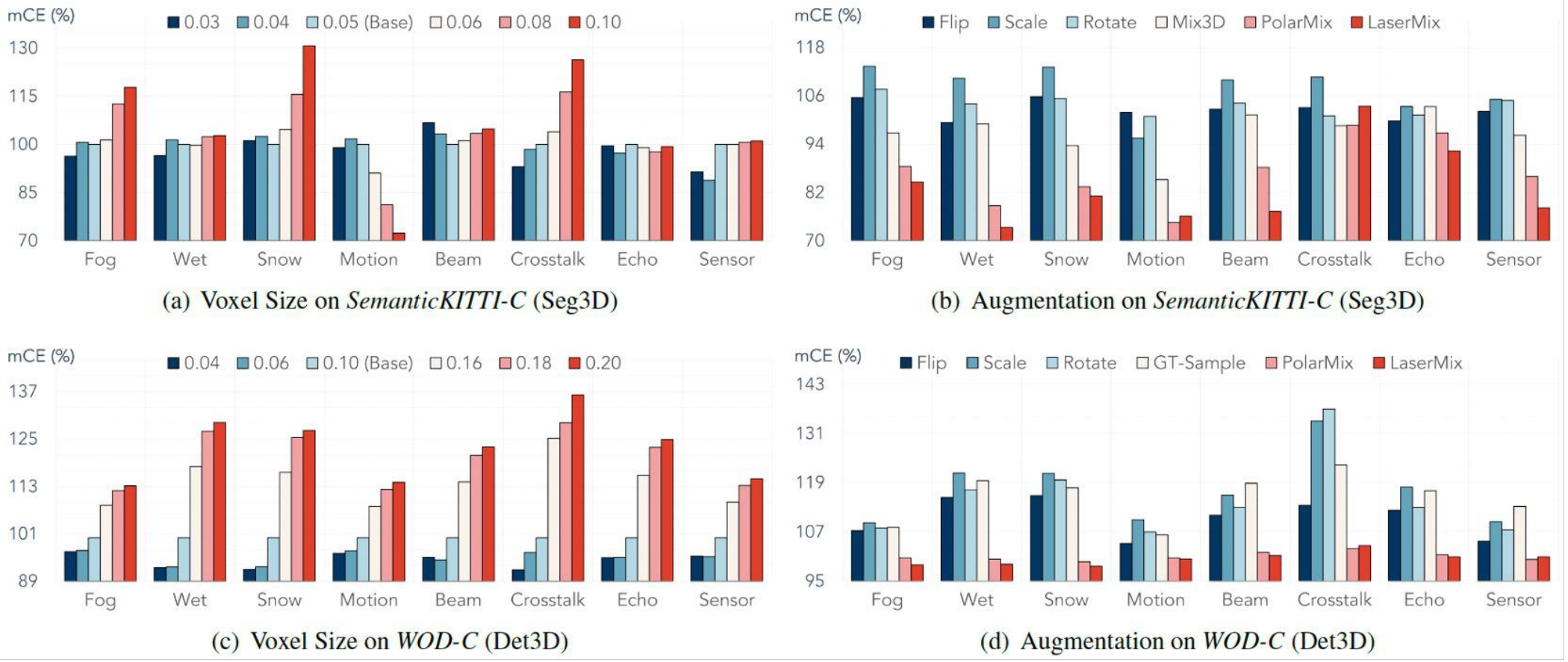
1. Existing 3D detectors and segmentors are **vulnerable** to real-world corruptions
2. Models trained with LiDAR data from different sources (sensor setups) exhibit **inconsistent sensitivities** to each corruption type
3. Representing the LiDAR data as **raw points**, **sparse voxel**, or the **fusion** of them tend to yield better robustness



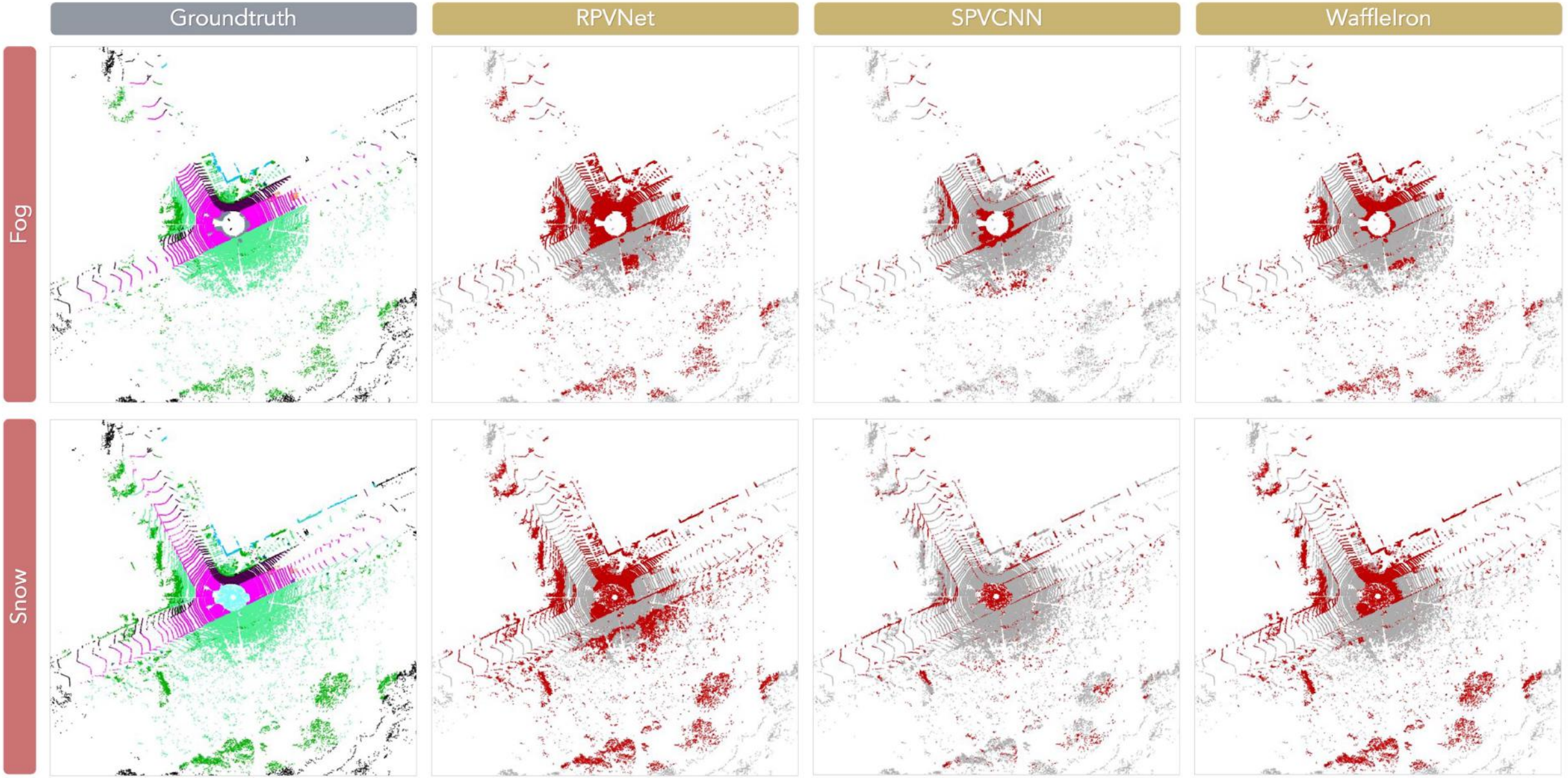
Robo3D: Key Observations

4. The 3D detectors and segmentors show different **sensitivities** to corruption scenarios

5. The recent **out-of-context augmentation techniques** improve 3D robustness by large margins; the flexible rasterization strategies help learn more robust features

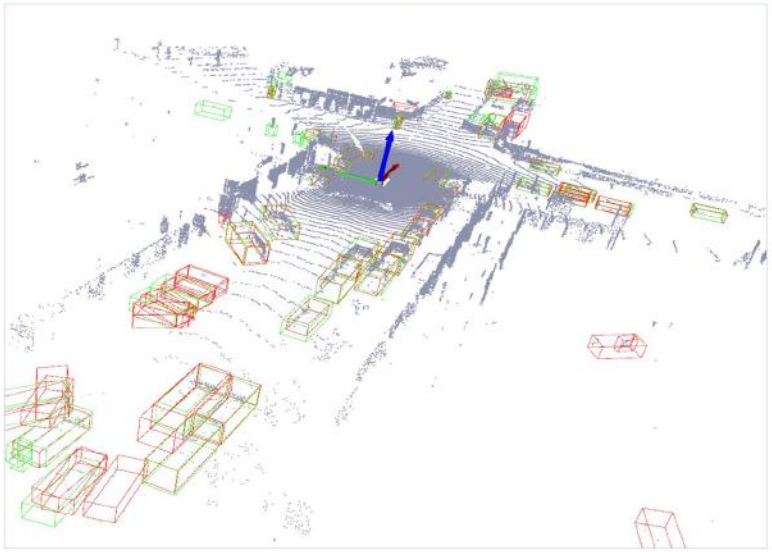


Robo3D: Qualitative Assessments

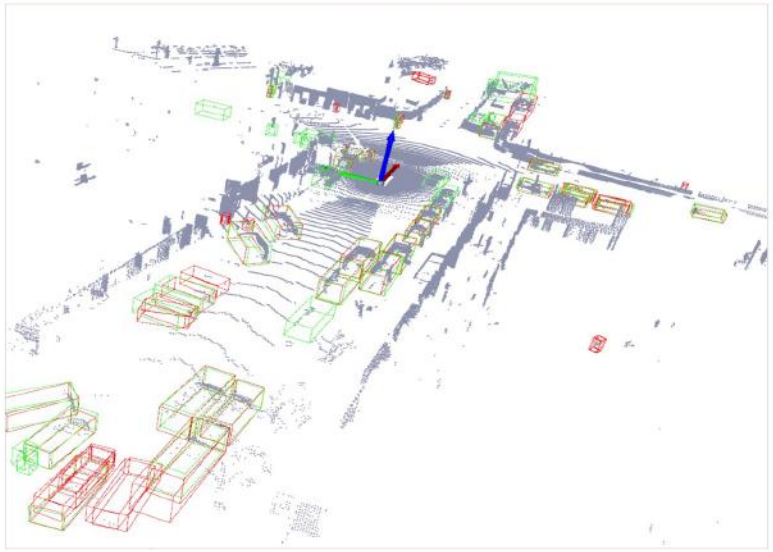


Robo3D: Qualitative Assessments

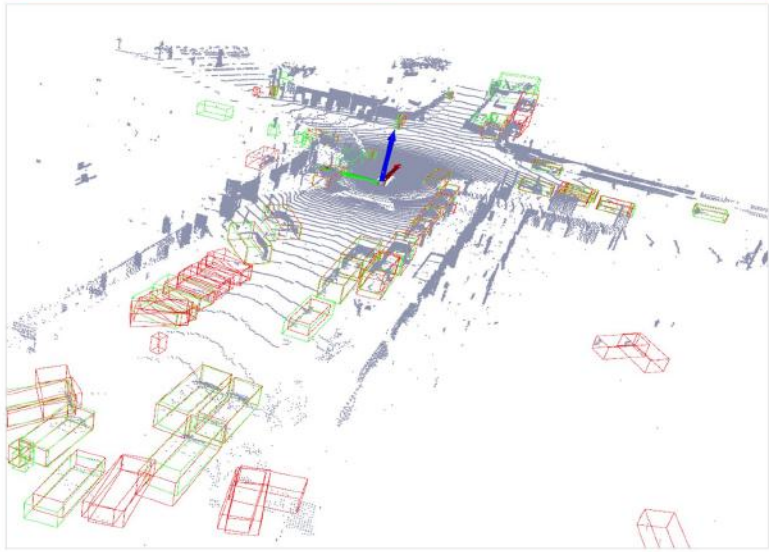
Fog



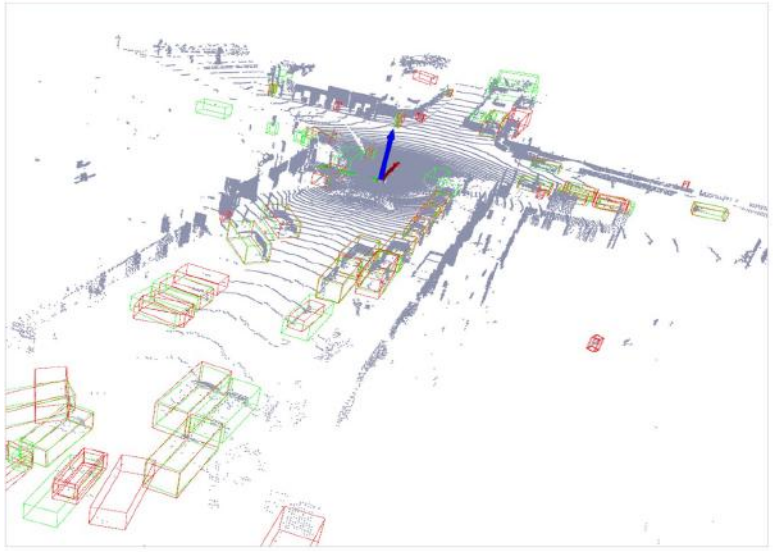
Wet Ground



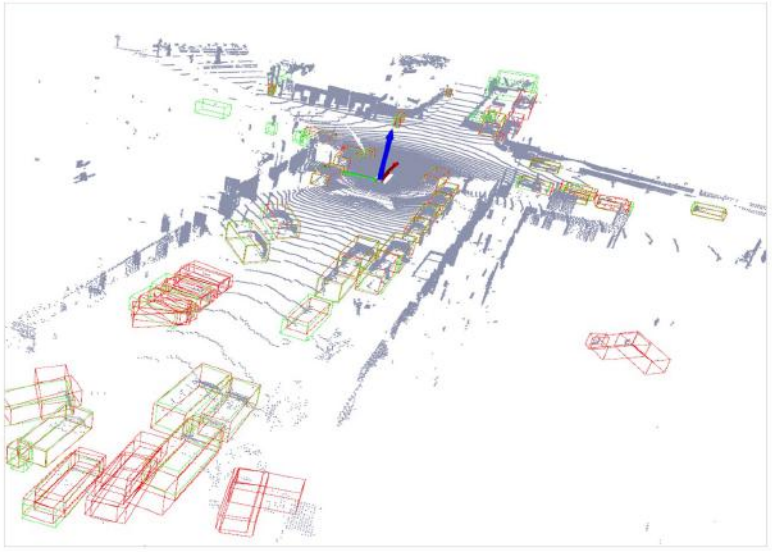
Crosstalk



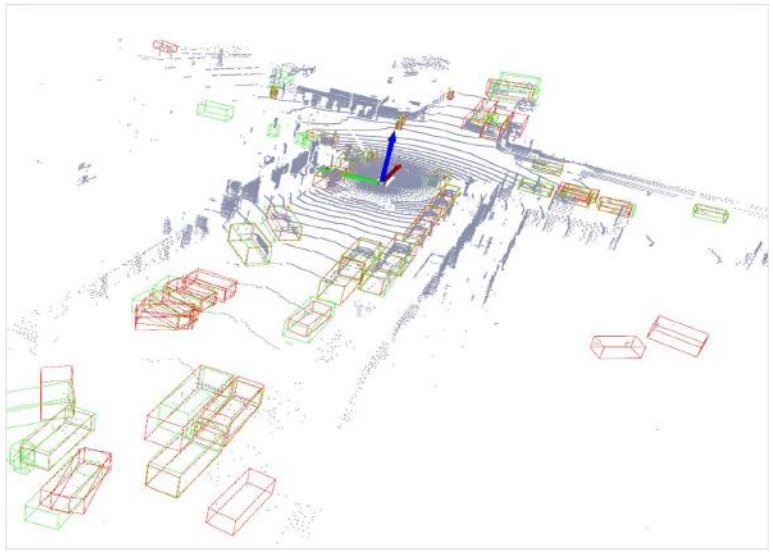
Snow



Motion Blur



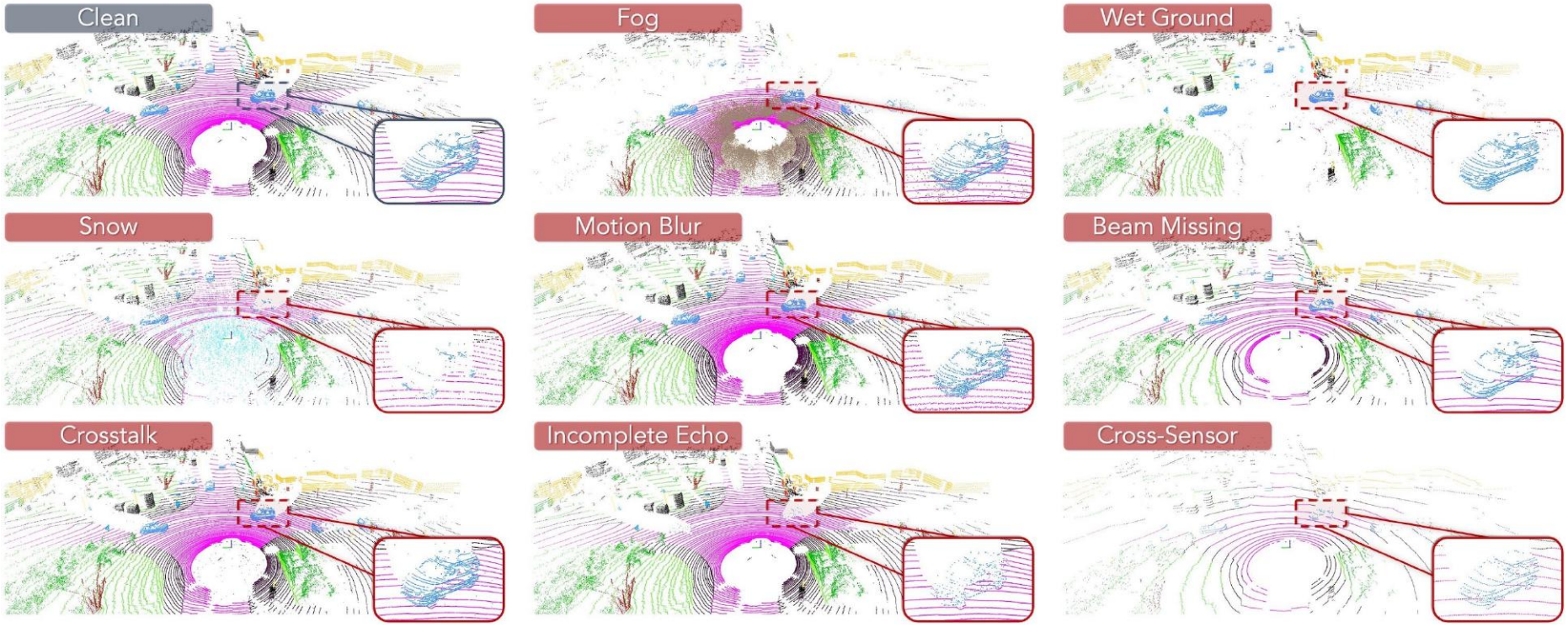
Cross-Sensor



Robo3D: Robustness Enhancement

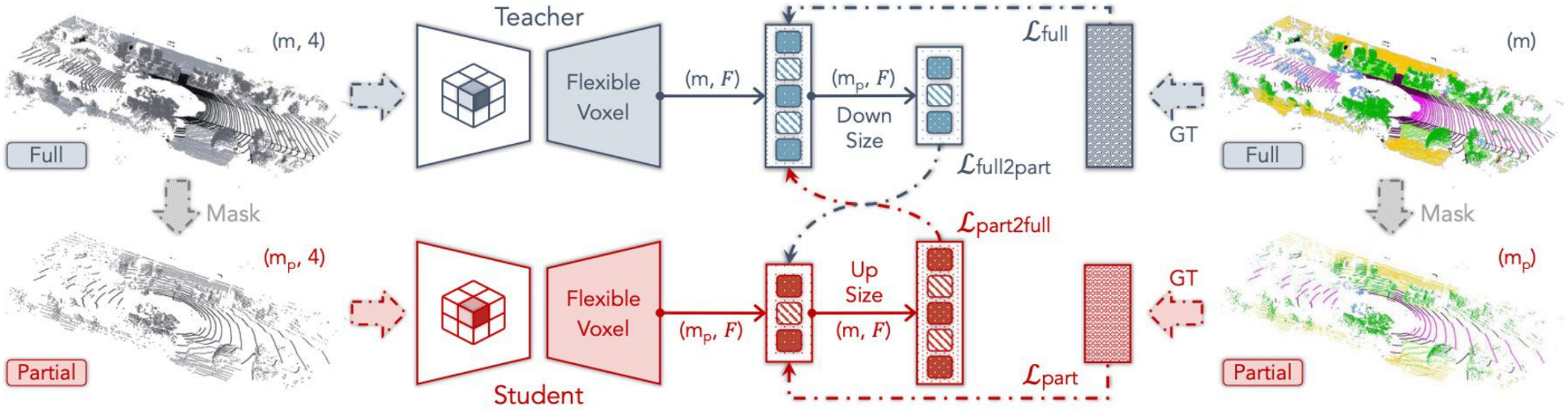
Motivation

- The natural corruptions often cause severe occlusion, attenuation, and reflection of light impulses, resulting in the unavoidable loss of LiDAR points in certain regions around the ego-vehicle
- For example, the wet ground absorbs energy and loses points on the surfaces; the potential incomplete echo and beam missing caused by reflection or dust/insect occlusion may lead to serious object failure



Robo3D: Robustness Enhancement

Density-Insensitive Training



- Completion loss: $\mathcal{L}_{\text{part2full}} = || \mathcal{G}_\theta^{\text{tea}}(x), \texttt{interp}(\mathcal{G}_\theta^{\text{stu}}(\tilde{x})) ||_2^2 .$

- Confirmation loss: $\mathcal{L}_{\text{full2part}} = || \texttt{subsample}(\mathcal{G}_\theta^{\text{tea}}(x)), \mathcal{G}_\theta^{\text{stu}}(\tilde{x}) ||_2^2 .$

ICCV23
PARIS

Open-Source Resources

Robo3D

Paper

Code

MMDet3D

https://github.com/ldkong1205/Robo3D

Topic

1. Overview & Background

2. Data-Efficient 3D Perception

3. Robustness in 3D

4. Segment Any Point Cloud Sequences

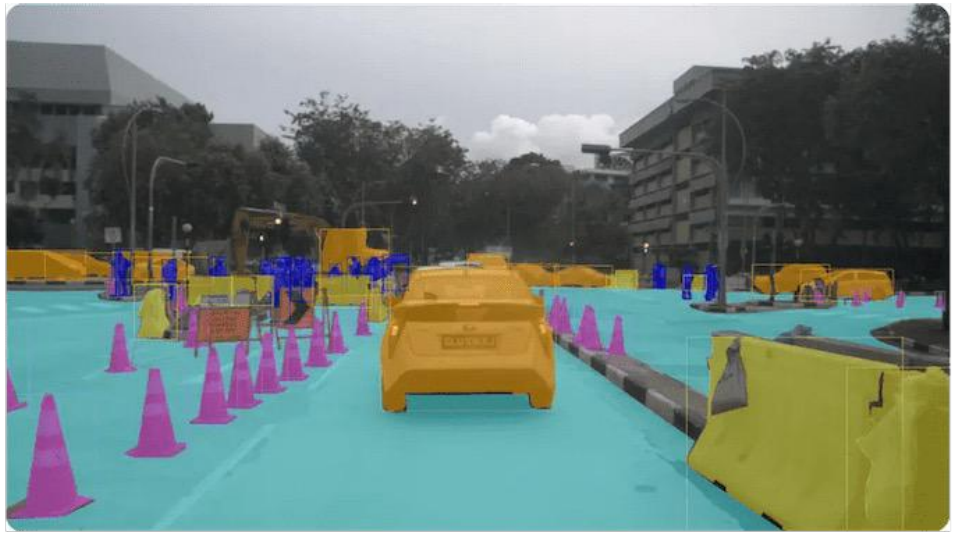
NEURAL INFORMATION PROCESSING SYSTEMS

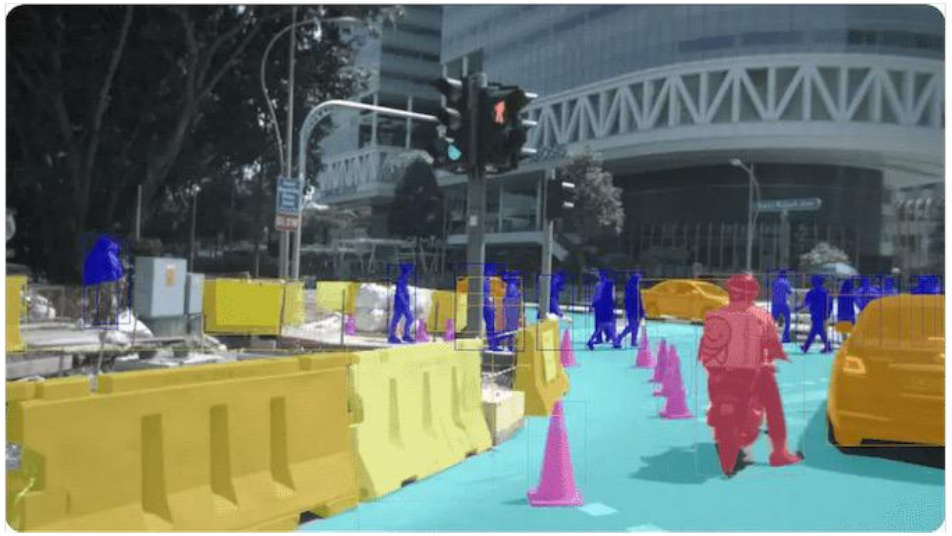
Segment Any Point Cloud Sequences by Distilling Vision Foundation Models

Youquan Liu[1,*], Lingdong Kong[1,2,*], Jun Cen[1,3], Runnan Chen[1,4]
Wenwei Zhang[1,5], Liang Pan[5], Kai Chen[1], Ziwei Liu[5]

[1]Shanghai AI Lab [2]NUS [3]HKUST [4]HKU [5]S-Lab, NTU

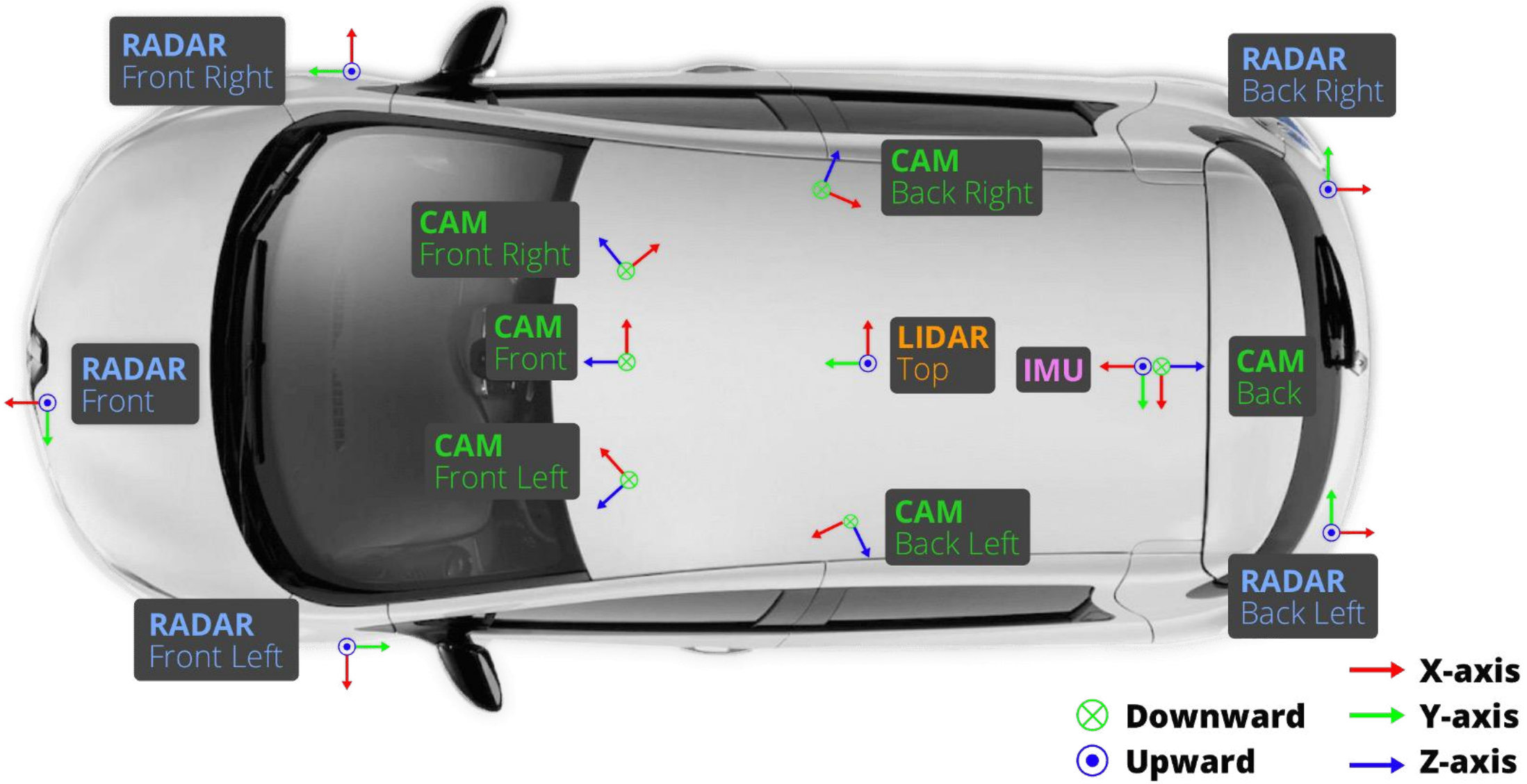
Visual Perception System





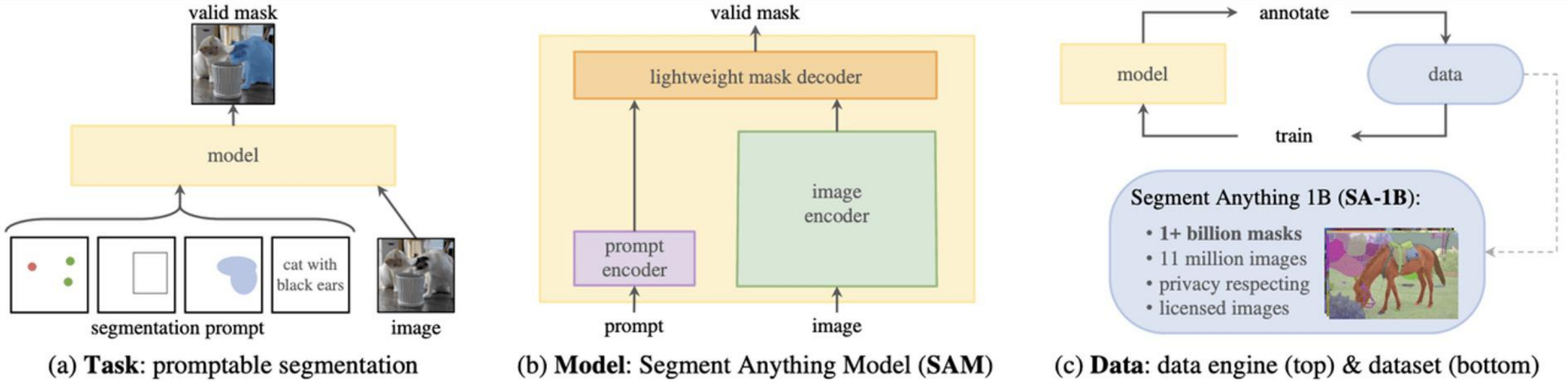
H. Caesar, et al. nuScenes: A multimodal dataset for autonomous driving, CVPR, 2020.

Autonomous Driving Perception System



H. Caesar, et al. nuScenes: A multimodal dataset for autonomous driving, CVPR, 2020.

Segment Anything

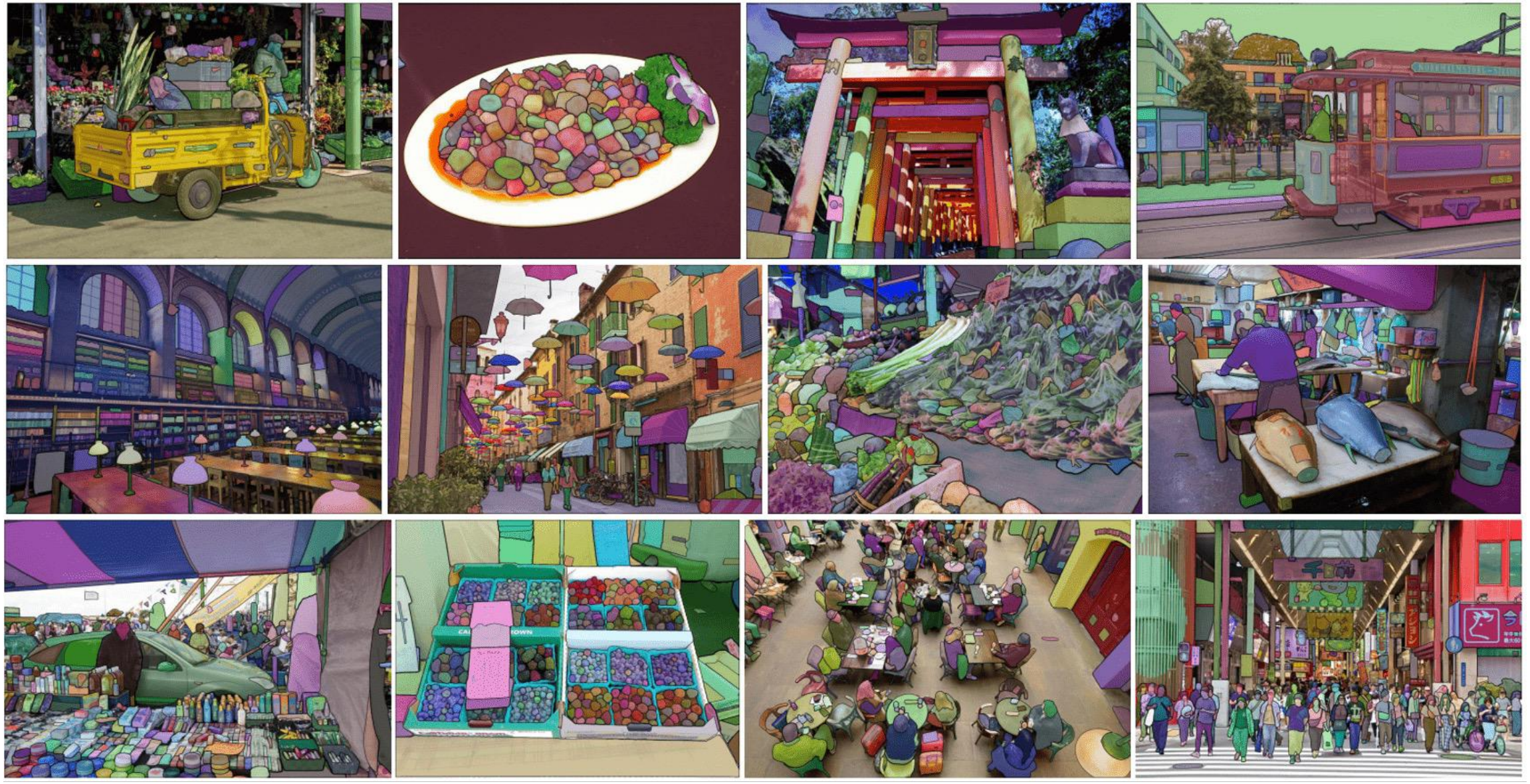


(a) **Task**: promptable segmentation (b) **Model**: Segment Anything Model (**SAM**) (c) **Data**: data engine (top) & dataset (bottom)

- **Training Set:** SA-1B, over 1B masks on 11M images
- **Model:** ViT-H SAM model
- **Demo:** https://segment-anything.com/demo

A. Kirillov, et al. Segment anything, ICCV, 2023.

Segment Anything

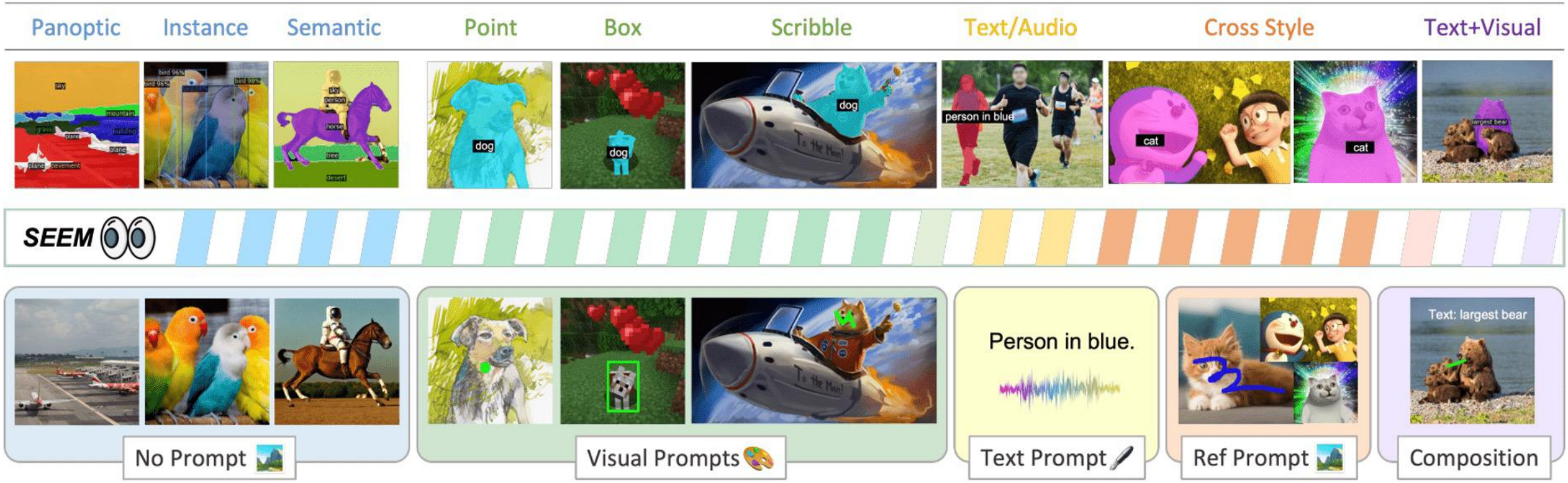


A. Kirillov, et al. Segment anything, ICCV, 2023.

X-Decoder

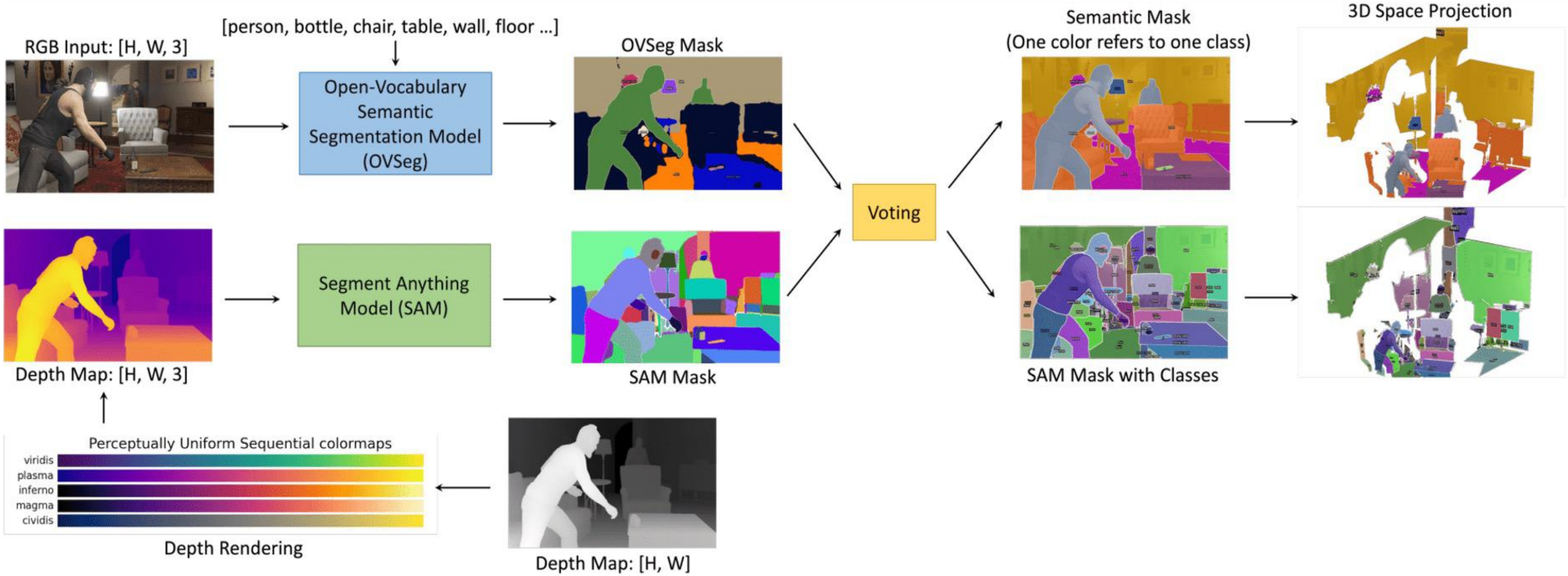
X. Zou, et al. Generalized decoding for pixel, image, and language, CVPR, 2023.

SEEM



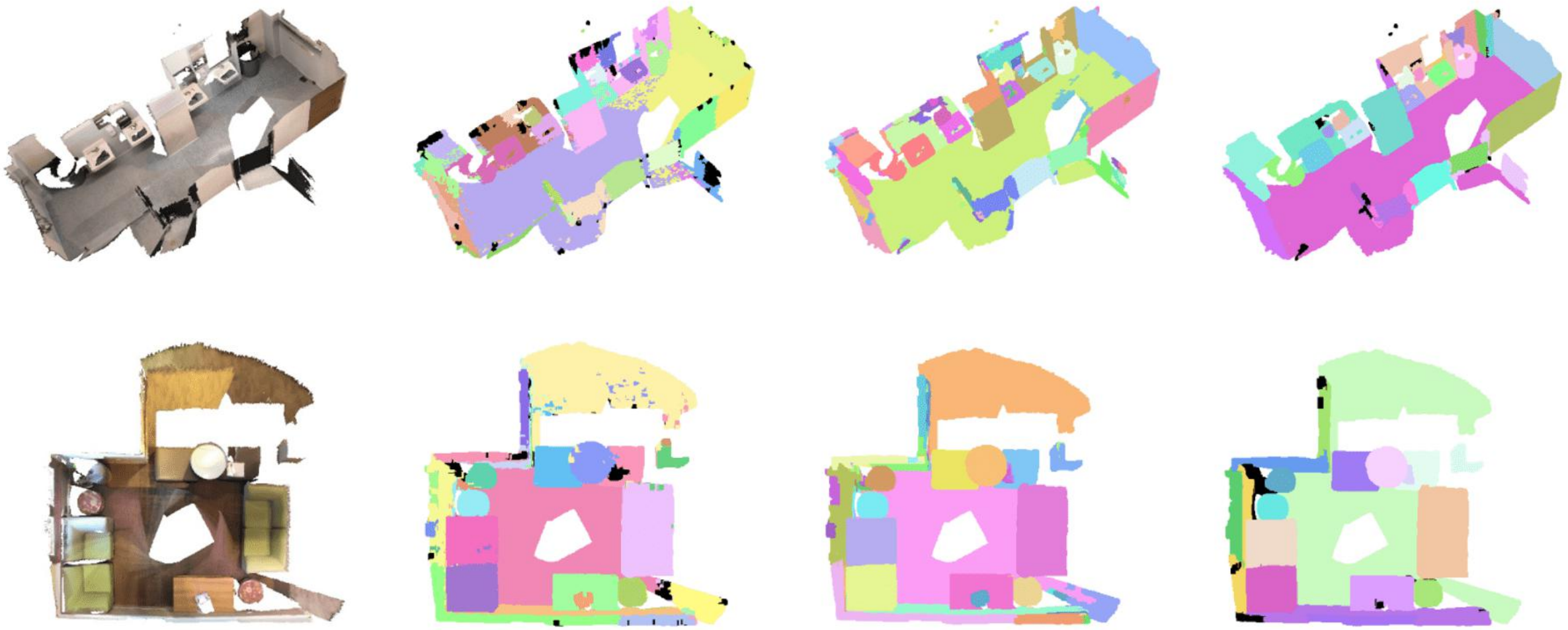
X. Zou, et al. Segment everything everywhere all at once, NeurIPS, 2023.

Segment Any RGB-D



J. Cen, et al. SAD: Segment any RGBD, NeurIPS Workshop, 2023.

SAM3D



Y. Yang, et al. SAM3D: Segment anything in 3D scenes, arXiv, 2023.

Seal

segment any point cloud

TL;DR

- Seal is a versatile self-supervised learning framework capable of segmenting any automotive point clouds

Seal

segment any point cloud

TL;DR

- Seal is a versatile self-supervised learning framework capable of segmenting any automotive point clouds
- Seal leverages off-the-shelf knowledge from vision foundation models (VFMs) and encouraging spatial and temporal consistency from such knowledge during the representation learning stage

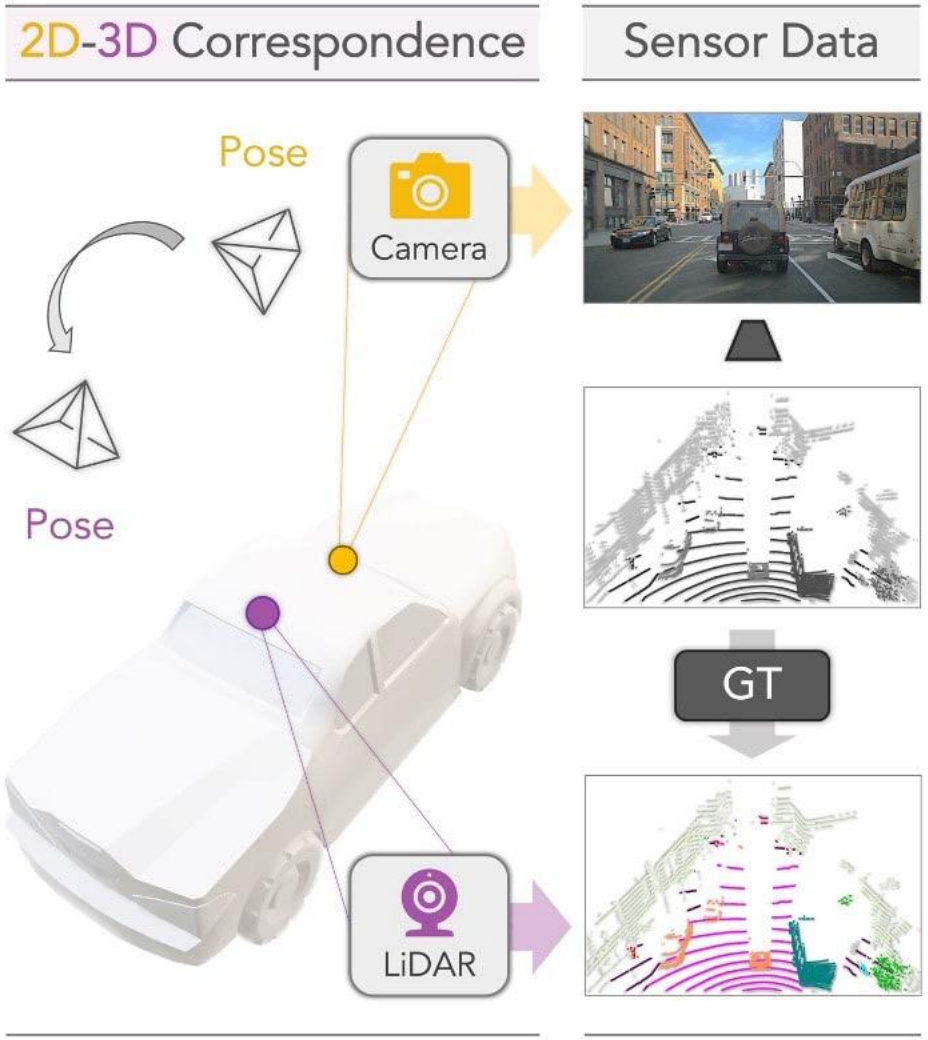
Seal

segment any point cloud

TL;DR

- **Seal** is a versatile self-supervised learning framework capable of segmenting any automotive point clouds
- **Seal** leverages off-the-shelf knowledge from vision foundation models (VFMs) and encouraging spatial and temporal consistency from such knowledge during the representation learning stage
- **Seal** enables knowledge transfer in an off-the-shelf manner to downstream tasks involving diverse point clouds, including those from real/synthetic, low/high-resolution, large/small-scale, and clean/corrupted datasets

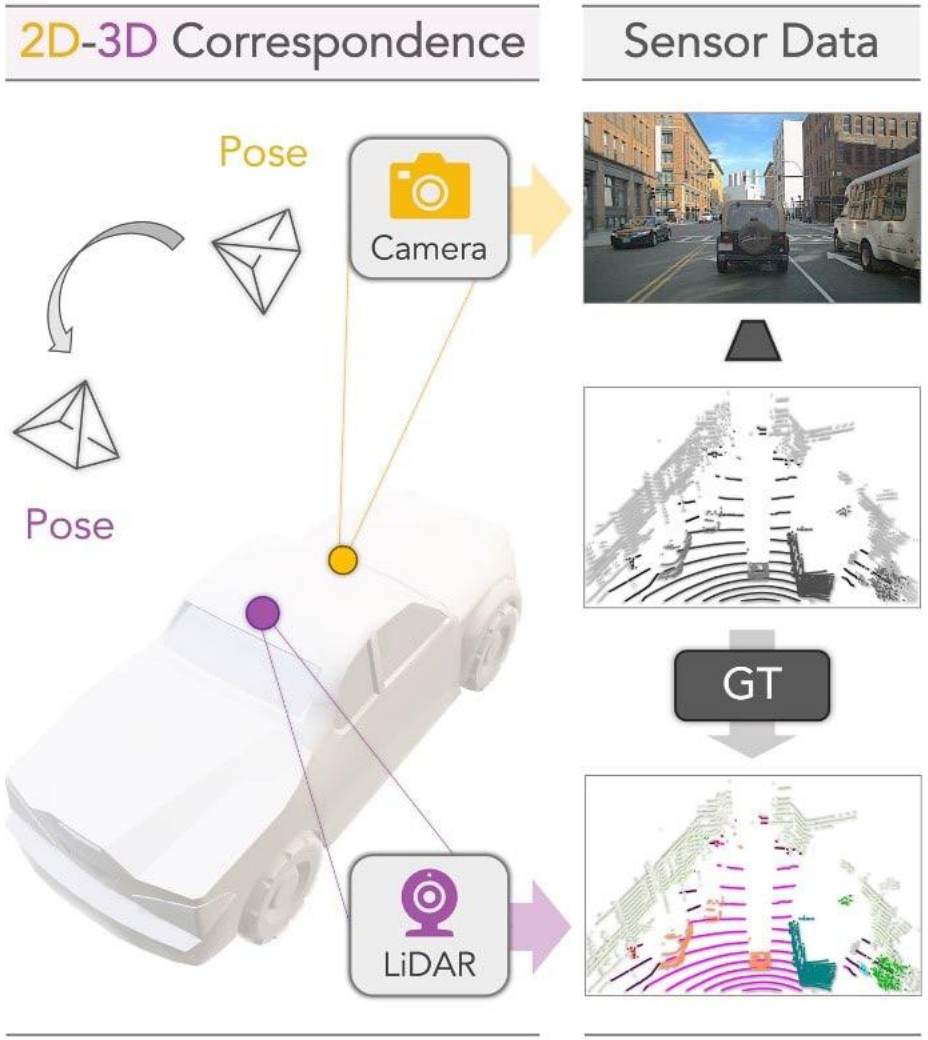
Seal: Challenges & Motivations



Challenges

- Camera views ≠ LiDAR views
- Automotive point clouds have unique label mappings
- LiDAR and cameras are not perfectly synchronized

Seal: Challenges & Motivations



Challenges

- Camera views ≠ LiDAR views
- Automotive point clouds have unique label mappings
- LiDAR and cameras are not perfectly synchronized

We hope to have a framework that …

- Conducts self-supervised learning on automotive point clouds
- Enforces spatial and temporal consistency during representation learning
- Can be generalizable to diverse downstream tasks

Superpixels

What is Superpixel

- Grouped pixels of perceptually meaningful atomic regions, which can be used to replace the rigid structure of the pixel grid

Fig. 1: Images segmented using SLIC into superpixels of size 64, 256, and 1024 pixels (approximately).

R. Achanta, et al. SLIC superpixels compared to state-of-the-art superpixel methods, TPAMI, 2012.

Superpixels

What is Superpixel

- Grouped pixels of perceptually meaningful atomic regions, which can be used to replace the rigid structure of the pixel grid

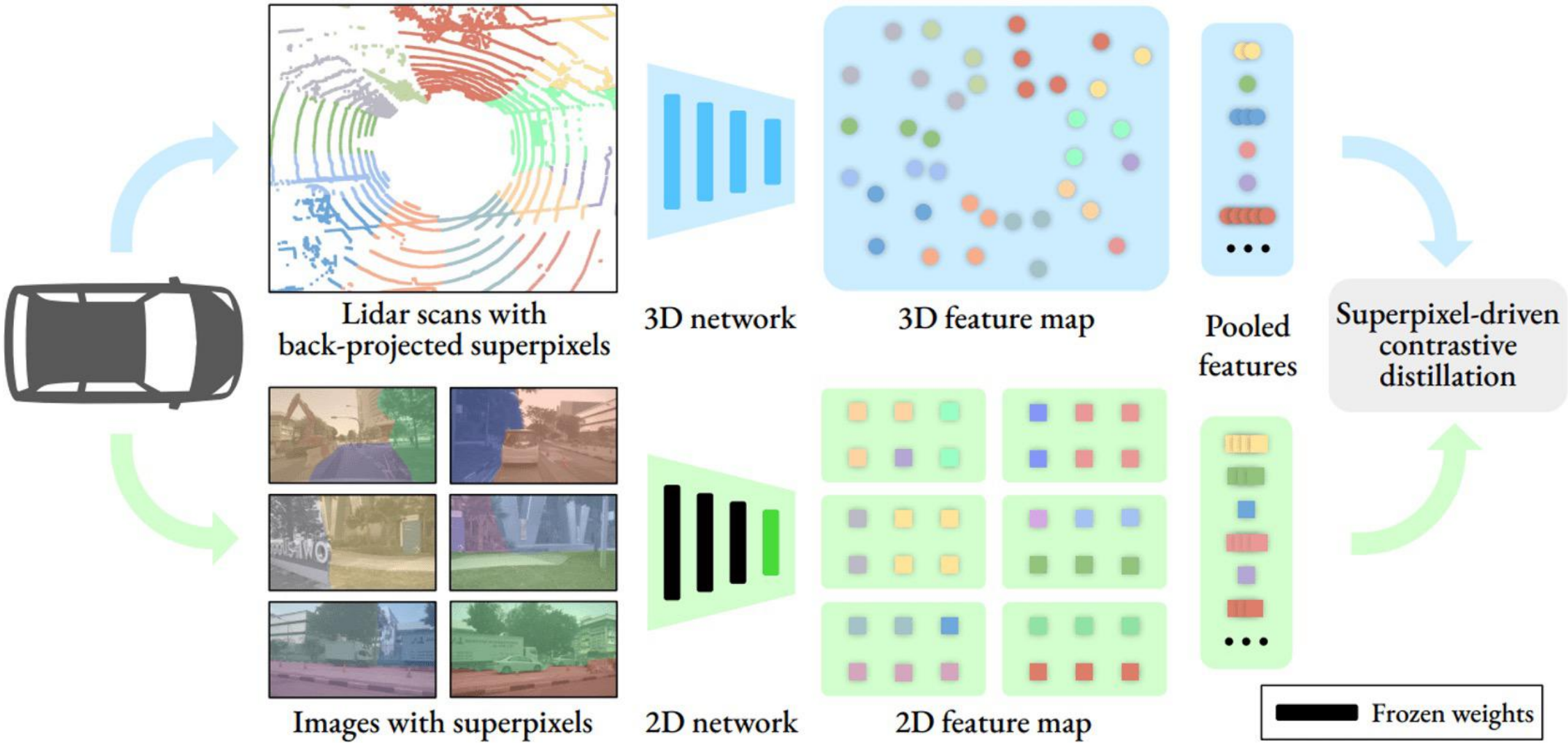
Key Features of Superpixel

- Captures image redundancy
- Provides a convenient primitive from which to compute image features
- Greatly reduces the complexity of subsequent image processing tasks

Fig. 1: Images segmented using SLIC into superpixels of size 64, 256, and 1024 pixels (approximately).

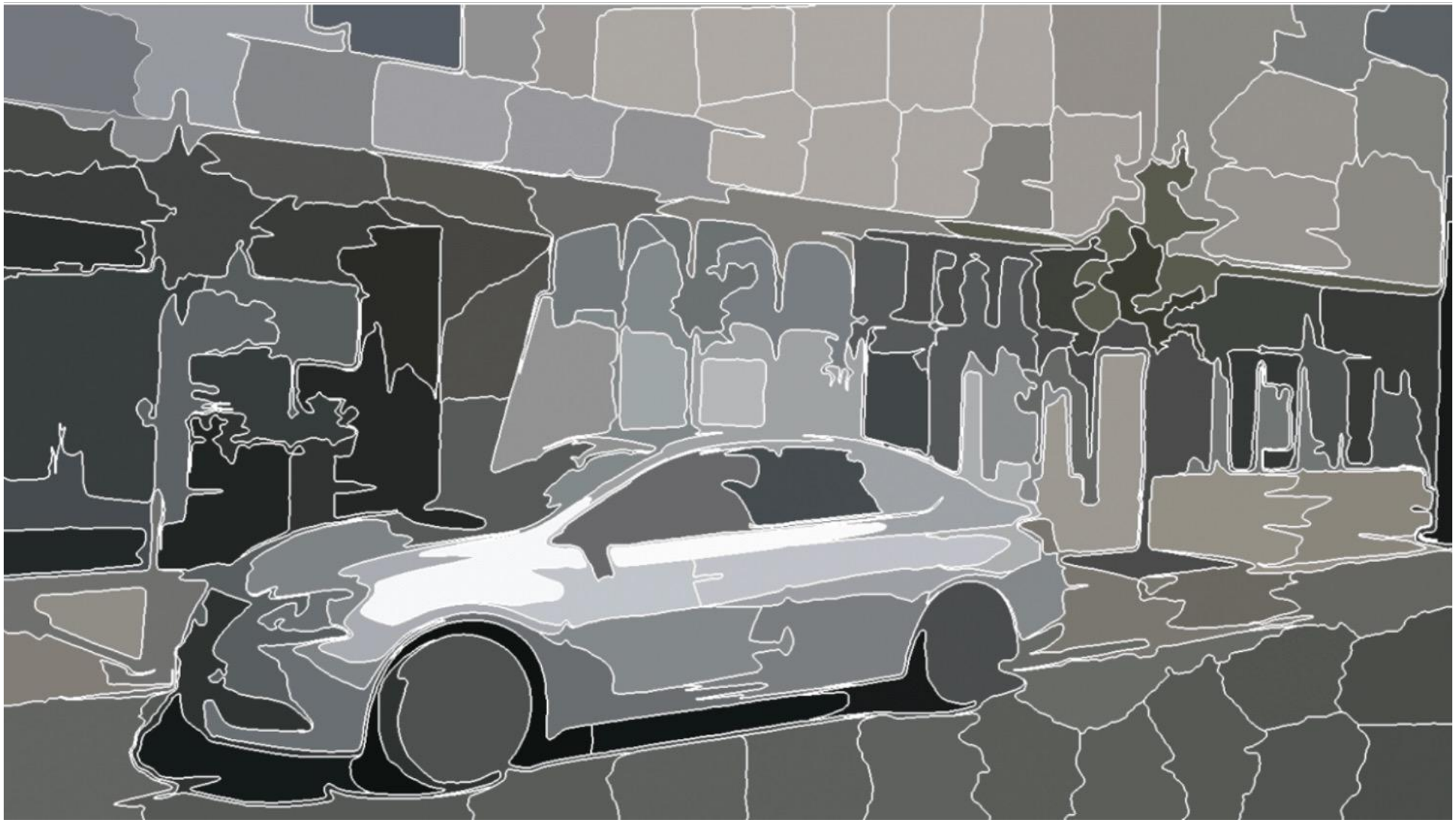
R. Achanta, et al. SLIC superpixels compared to state-of-the-art superpixel methods, TPAMI, 2012.

SLidR: Superpixel-Driven LiDAR Representation



C. Sautier, et al. Image-to-LiDAR self-supervised distillation for autonomous driving data, CVPR, 2022.

SLidR: Superpixel-Driven LiDAR Representation



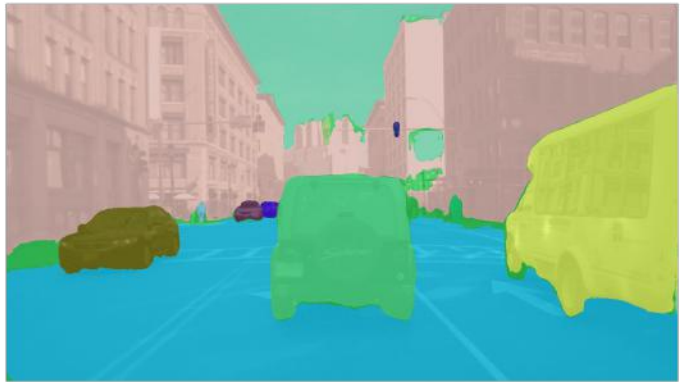
C. Sautier, et al. Image-to-LiDAR self-supervised distillation for autonomous driving data, CVPR, 2022.

Semantic Superpixels

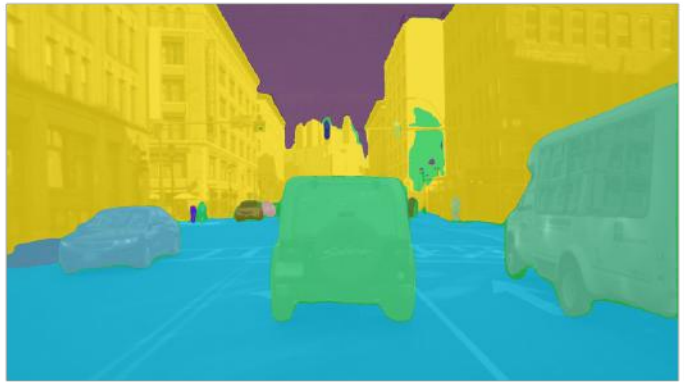
Camera View
(Front)

Superpixel
(SLIC)

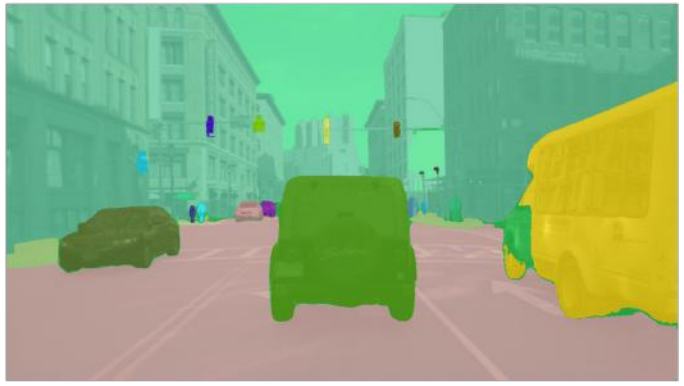
Semantic Superpixel
(SAM)



Semantic Superpixel
(X-Decoder)



Semantic Superpixel
(OpenSeeD)



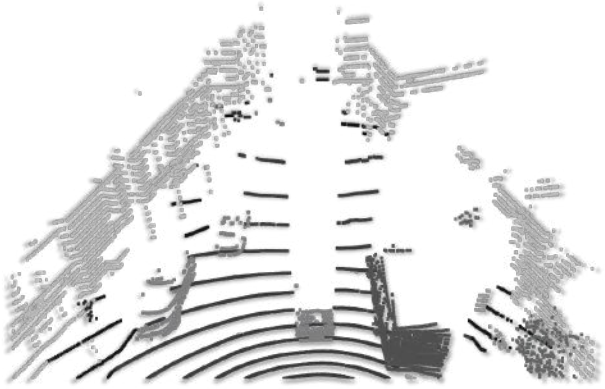
Semantic Superpixel
(SEEM)

Superpixel to Superpoint

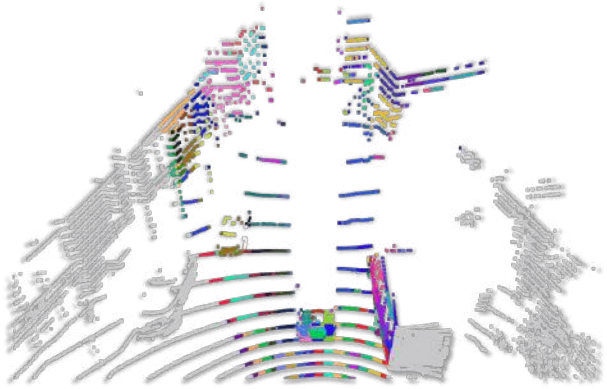
Camera View
(Front)

Superpixel
(SLIC)

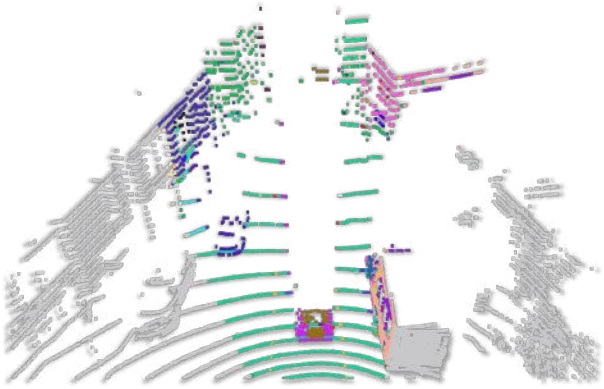
Semantic Superpixel
(SAM)



LiDAR View
(Front)

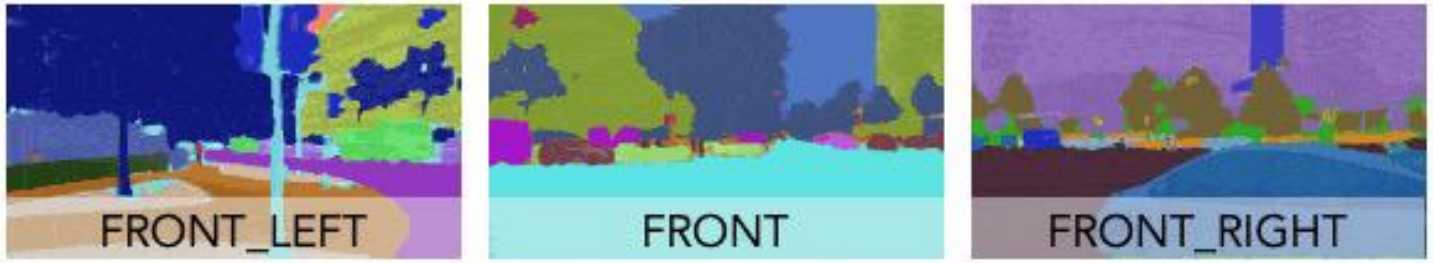


Superpoint
(SLIC)



Semantic Superpoint
(SAM)

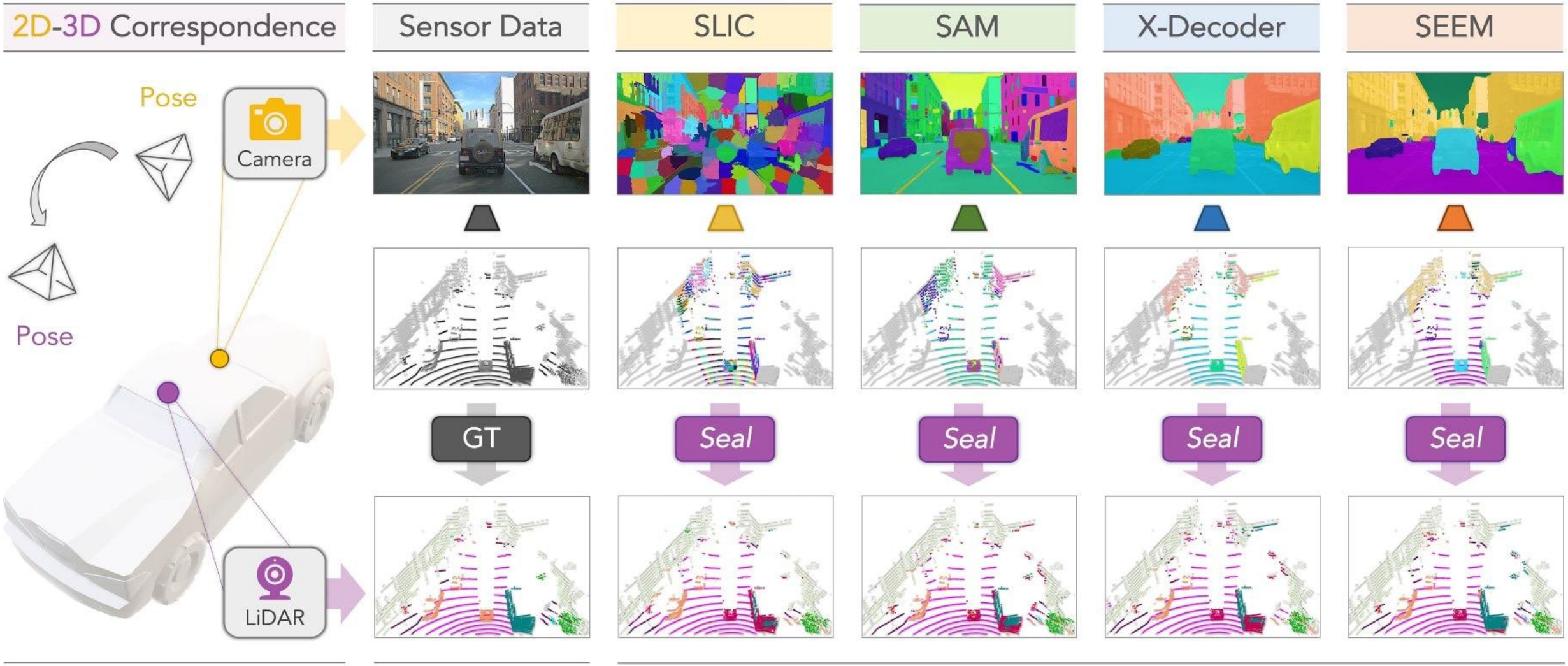
Superpixel to Superpoint



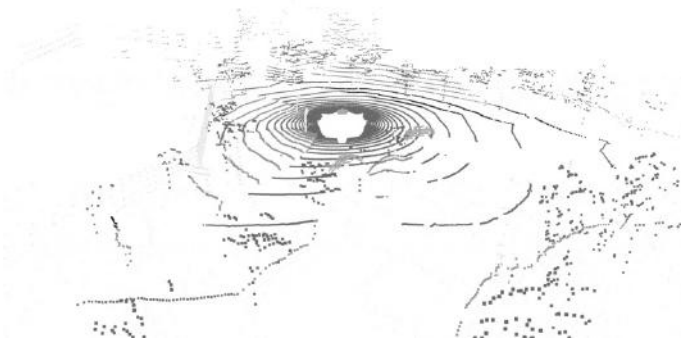
LIDAR (RAW UNLABELED POINT)

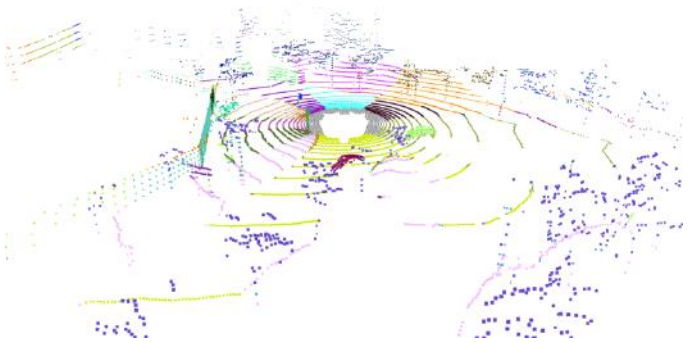
LIDAR (SEMANTIC SUPERPOINT)

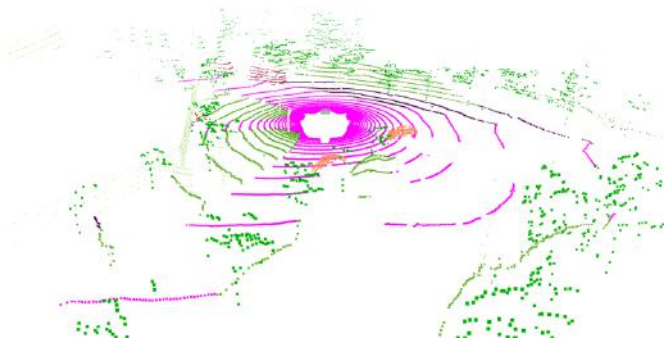
Seal: Vision Foundation Models

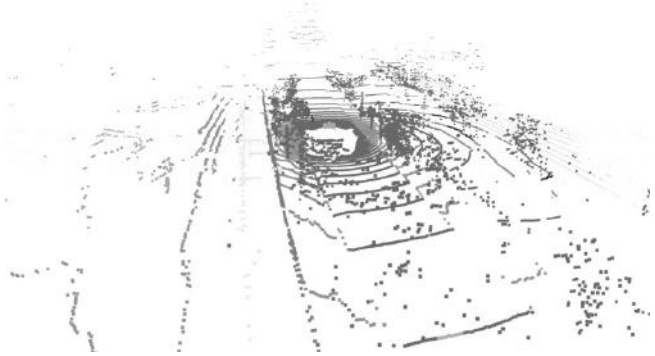


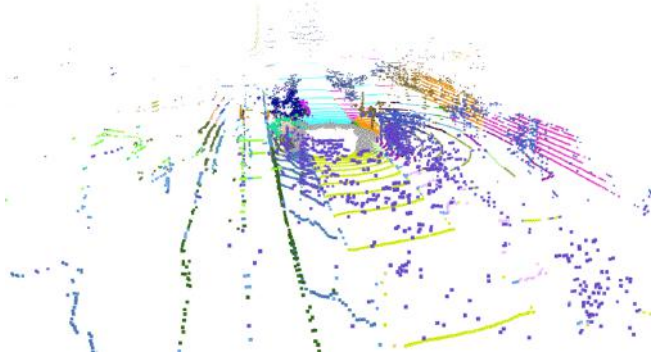
Seal: Semantic Superpoints

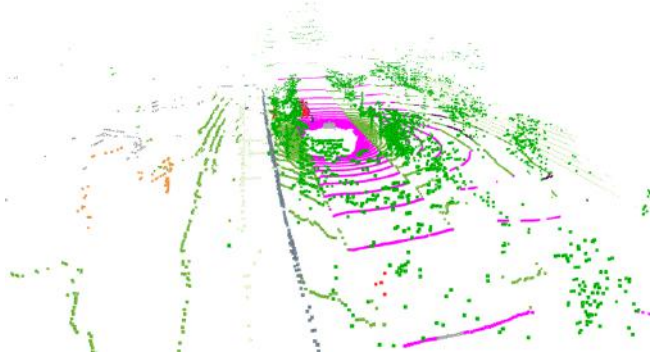










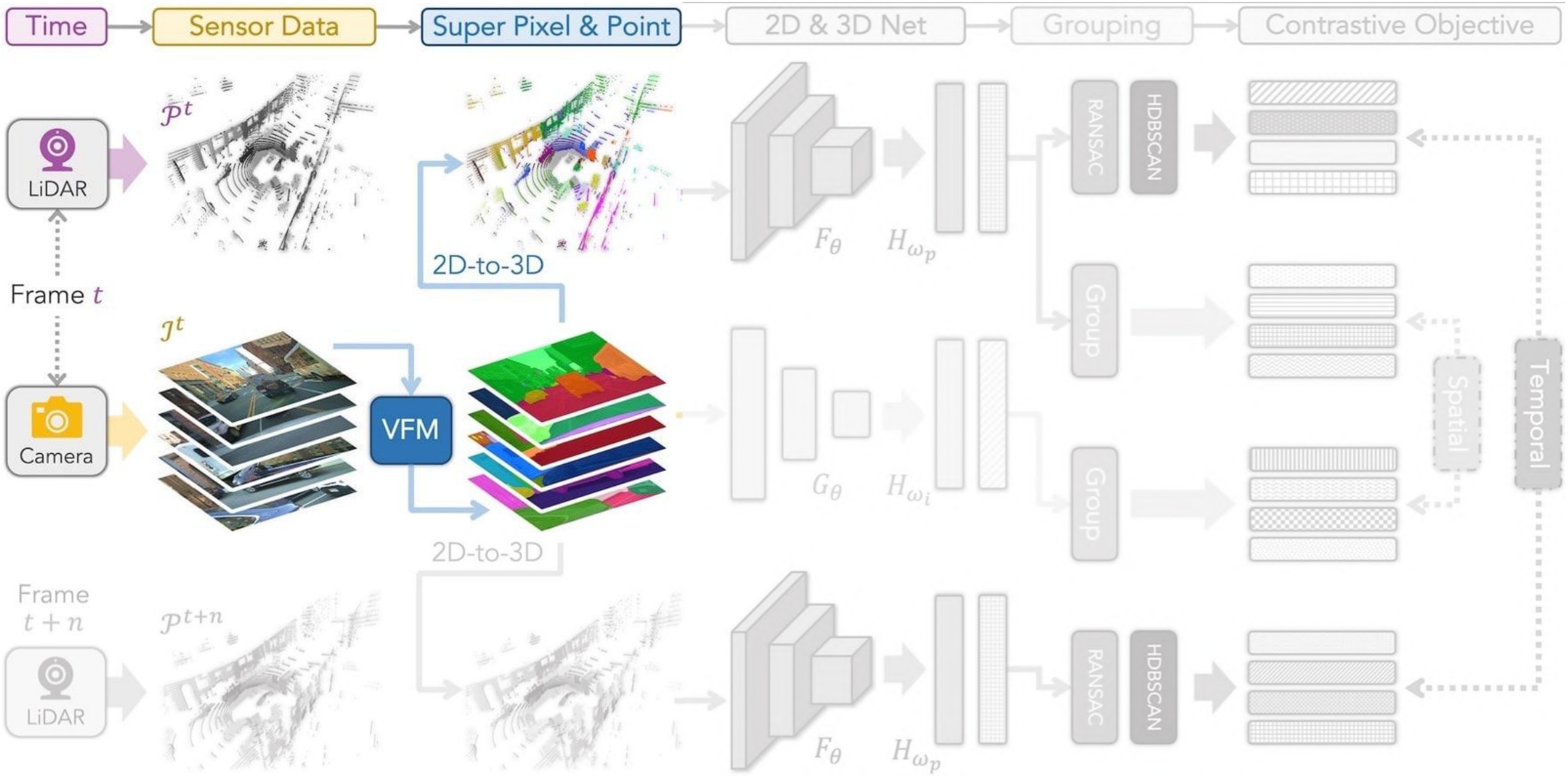


Raw Point Cloud

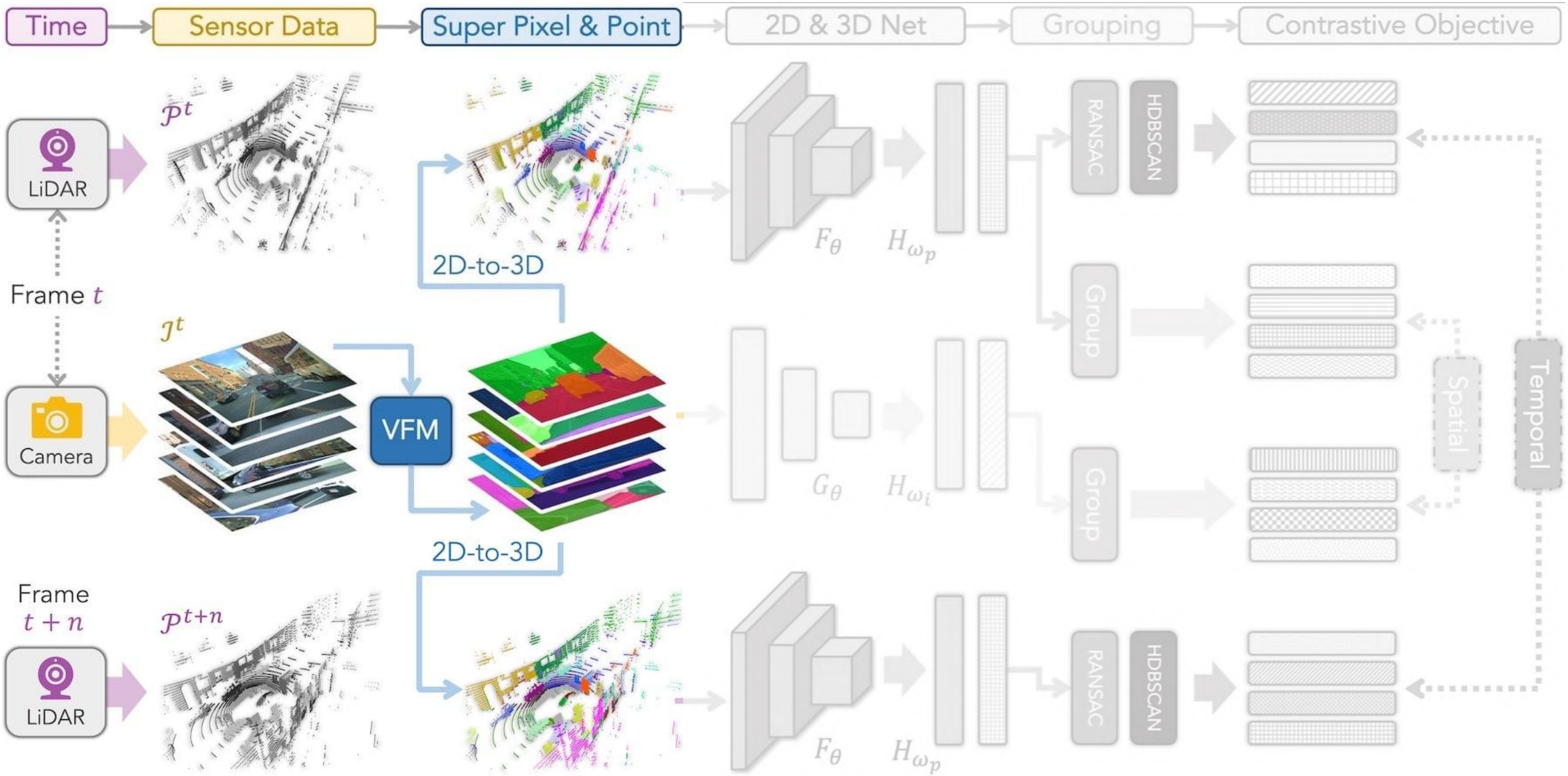
Semantic Superpoint

Ground-Truth

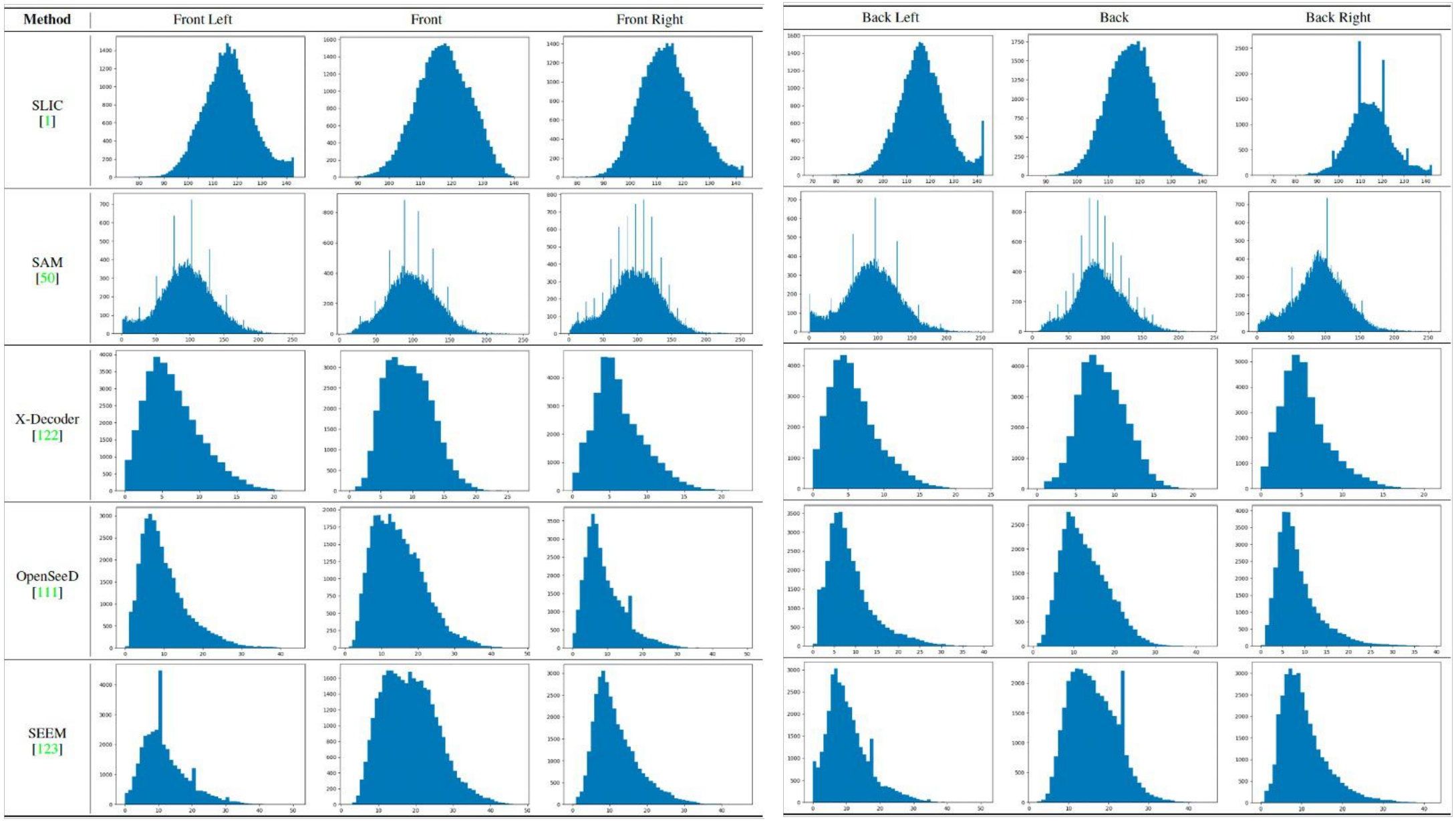
Seal: Framework



Seal: Framework



Seal: Superpixels Statistics



SLIC
(80~140)

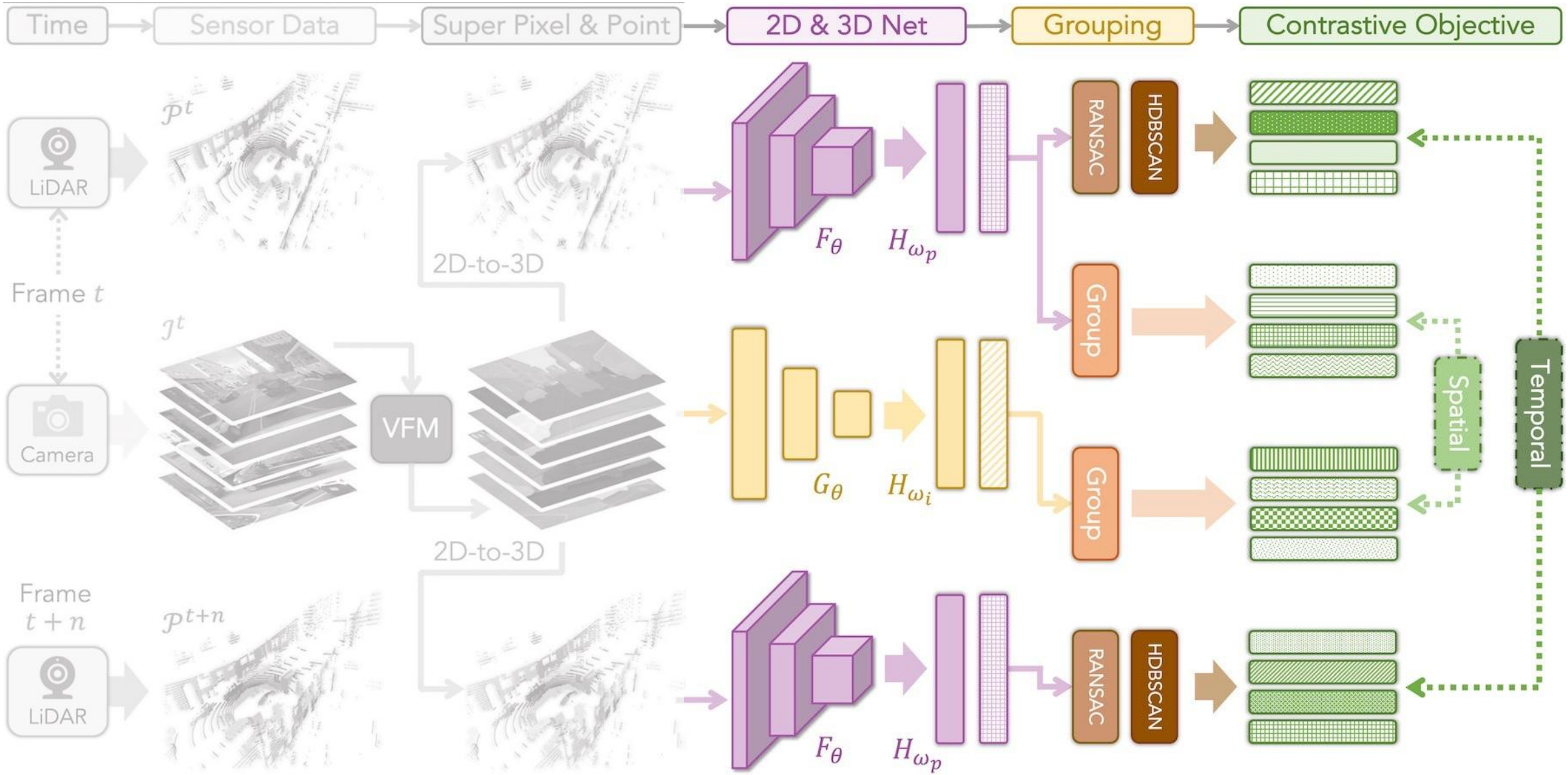
SAM
(5~250)

X-Decoder
(1~20)

OpenSeeD
(1~40)

SEEM
(1~50)

Seal: Framework



Seal: Spatial & Temporal Consistency

Spatial Consistency

- Cross-modality contrastive learning:

$$\mathcal{L}^{vfm} = \mathcal{L}(\mathbf{Q}, \mathbf{K}) = -\frac{1}{M}\sum_{i=0}^{M}\log\left[\frac{e^{(\langle \mathbf{q}_i, \mathbf{k}_i\rangle/\tau)}}{\sum_{j\neq i} e^{(\langle \mathbf{q}_i, \mathbf{k}_j\rangle/\tau)} + e^{(\langle \mathbf{q}_i, \mathbf{k}_i\rangle/\tau)}}\right]$$

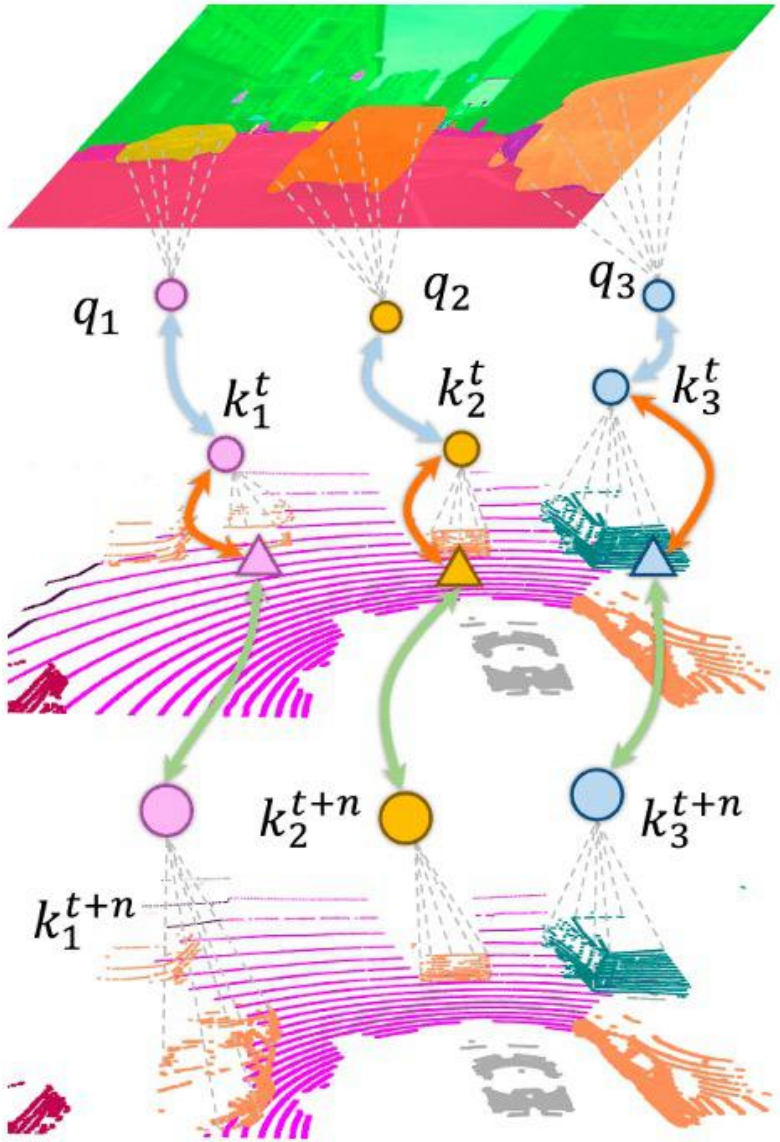
Temporal Consistency

- Superpoint temporal consistency:

$$\mathcal{L}^{t\to t+1} = -\frac{1}{M_k}\sum_{i=0}^{M_k}\log\left[\frac{e^{(\langle \mathbf{f}_i^t, \mathbf{f}_i^{t+1}\rangle/\tau)}}{\sum_{j\neq i} e^{(\langle \mathbf{f}_i^t, \mathbf{f}_j^{t+1}\rangle/\tau)} + e^{(\langle \mathbf{f}_i^t, \mathbf{f}_i^{t+1}\rangle/\tau)}}\right]$$

- Point-to-segment regularization:

$$\mathcal{L}^{p2s} = -\frac{1}{M_k}\sum_{i=0}^{M_k}\log\left[\frac{e^{(\langle \mathbf{f}_i^t, \mathbf{c}_i^t\rangle/\tau)}}{\sum_{j\neq i} e^{(\langle \mathbf{f}_i^t, \mathbf{c}_j^t\rangle/\tau)} + e^{(\langle \mathbf{f}_i^t, \mathbf{c}_i^t\rangle/\tau)}}\right]$$



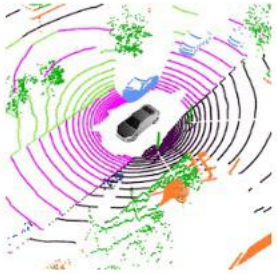
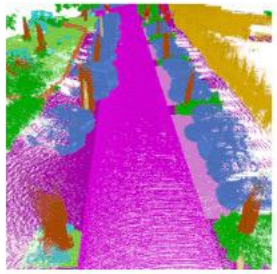
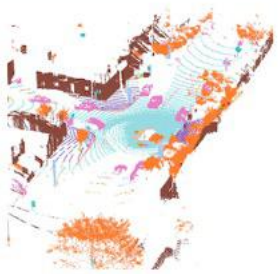
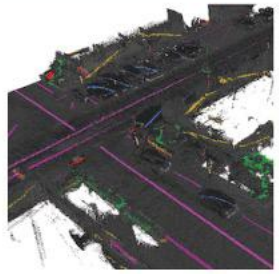
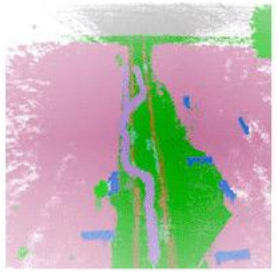
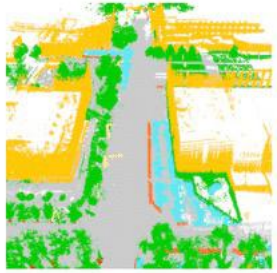
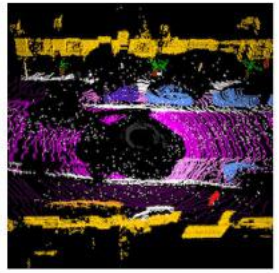
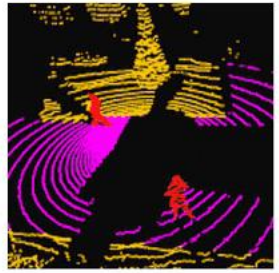
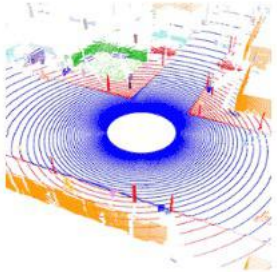
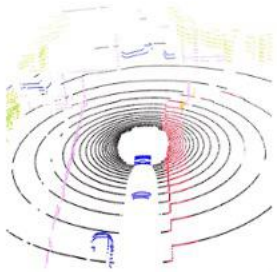
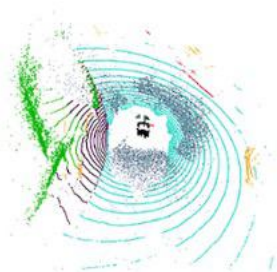
Seal: Experiments

Datasets

- Pretrained on nuScenes
- Linear probing with frozen backbone
- 20 downstream tasks on 11 point cloud datasets

Backbones

- 2D: ResNet-50, pretrained with MoCoV2
- 3D: MinkUNet with cylinder voxels of size 0.1m as the input
- 2x A100 GPUs

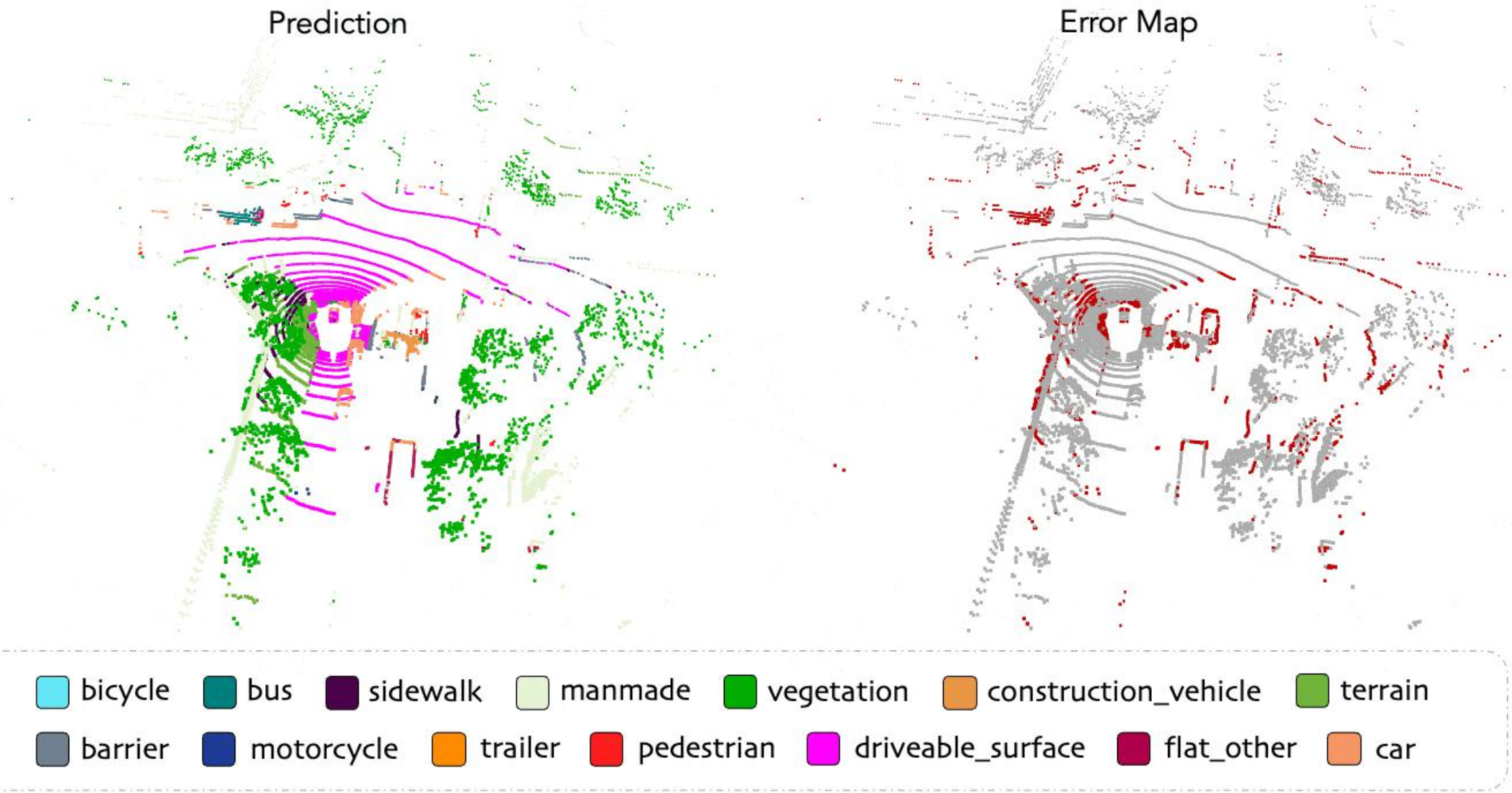
nuScenes	SemanticKITTI	Waymo Open	ScribbleKITTI
			
RELLIS-3D	**SemanticPOSS**	**SemanticSTF**	**DAPS-3D**
			
SynLiDAR	**Synth4D**	**nuScenes-C**	
			

Seal: Comparative Study

Table 1: Comparisons of different pretraining methods pretrained on *nuScenes* [26] and fine-tuned on *nuScenes* [7], *SemanticKITTI* [3], *Waymo Open* [88], and *Synth4D* [82]. **LP** denotes linear probing with frozen backbones. Symbol † denotes fine-tuning with the LaserMix augmentation [55]. Symbol ‡ denotes fine-tuning with semi-supervised learning. All mIoU scores are given in percentage (%).

Method & Year	nuScenes						KITTI	Waymo	Synth4D
	LP	1%	5%	10%	25%	Full	1%	1%	1%
Random	8.10	30.30	47.84	56.15	65.48	74.66	39.50	39.41	20.22
PointContrast [ECCV'20] [103]	21.90	32.50	-	-	-	-	41.10	-	-
DepthContrast [ICCV'21] [116]	22.10	31.70	-	-	-	-	41.50	-	-
PPKT [arXiv'21] [65]	35.90	37.80	53.74	60.25	67.14	74.52	44.00	47.60	61.10
SLidR [CVPR'22] [85]	38.80	38.30	52.49	59.84	66.91	74.79	44.60	47.12	63.10
ST-SLidR [CVPR'23] [66]	40.48	40.75	54.69	60.75	67.70	75.14	44.72	44.93	-
Seal (Ours)	**44.95**	**45.84**	**55.64**	**62.97**	**68.41**	**75.60**	**46.63**	**49.34**	**64.50**
Seal † (Ours)	-	48.41	57.84	65.52	70.80	77.13	-	-	-
Seal ‡ (Ours)	-	49.53	58.64	66.78	72.31	78.28	-	-	-

Seal: Linear Probing



Seal: Downstream Generalization

Table 2: Comparisons of different pretraining methods pretrained on *nuScenes* [26] and fine-tuned on different downstream point cloud datasets. All mIoU scores are given in percentage (%).

Method	ScribbleKITTI		RELLIS-3D		SemanticPOSS		SemanticSTF		SynLiDAR		DAPS-3D	
	1%	10%	1%	10%	Half	Full	Half	Full	1%	10%	Half	Full
Random	23.81	47.60	38.46	53.60	46.26	54.12	48.03	48.15	19.89	44.74	74.32	79.38
PPKT [65]	36.50	51.67	49.71	54.33	50.18	56.00	50.92	54.69	37.57	46.48	78.90	84.00
SLidR [85]	39.60	50.45	49.75	54.57	51.56	55.36	52.01	54.35	42.05	47.84	81.00	85.40
Seal (Ours)	**40.64**	**52.77**	**51.09**	**55.03**	**53.26**	**56.89**	**53.46**	**55.36**	**43.58**	**49.26**	**81.88**	**85.90**

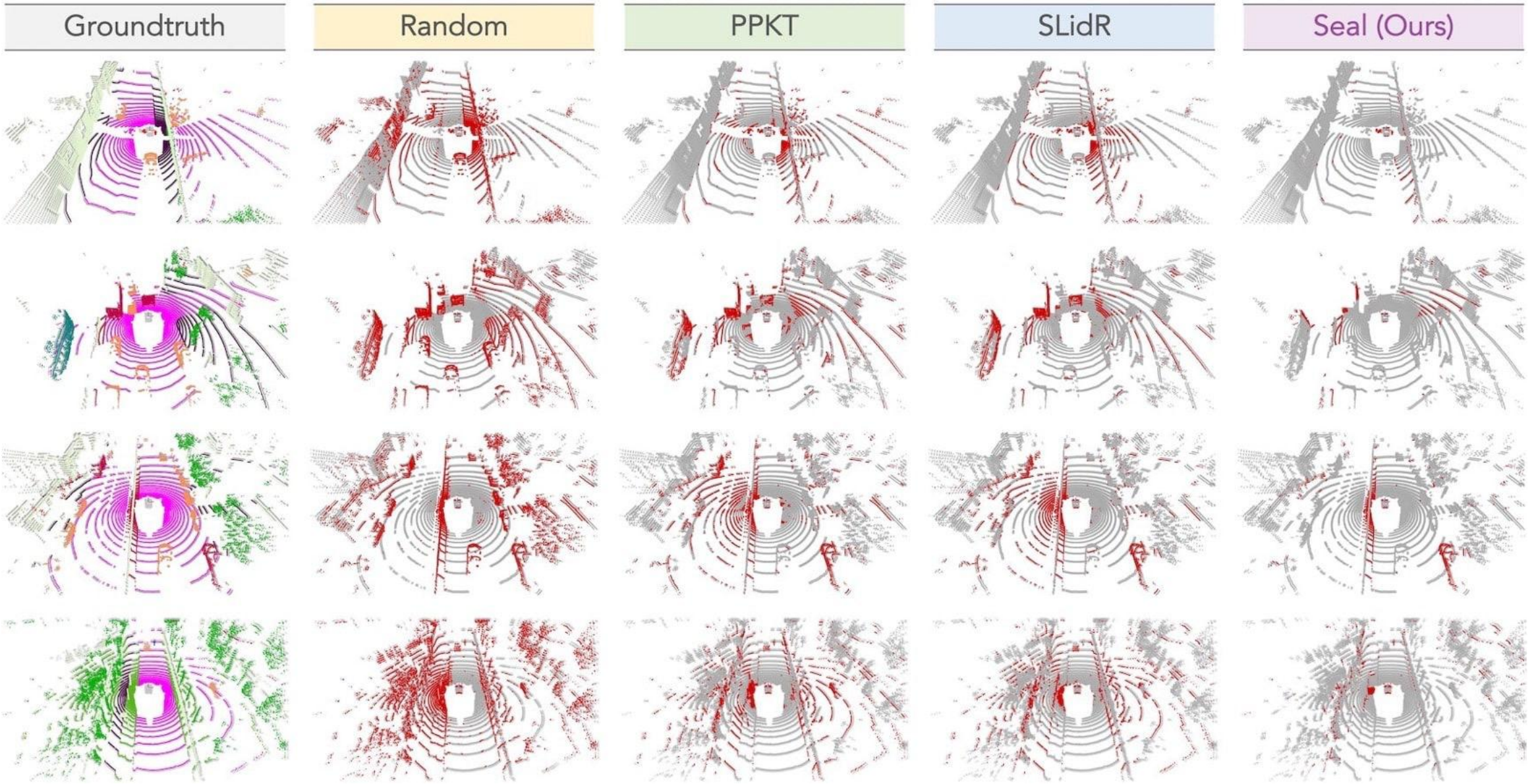
Comparative Study: Robustness Probing

Table 3: Robustness evaluations under eight out-of-distribution corruptions in the *nuScenes-C* dataset from the Robo3D benchmark [53]. All mCE, mRR, and mIoU scores are given in percentage (%).

	Initial	Backbone	mCE ↓	mRR ↑	Fog	Wet	Snow	Move	Beam	Cross	Echo	Sensor
LP	PPKT [65]	MinkUNet	183.44	**78.15**	30.65	35.42	28.12	29.21	32.82	19.52	28.01	20.71
	SLidR [85]	MinkUNet	179.38	77.18	34.88	38.09	**32.64**	26.44	33.73	**20.81**	31.54	21.44
	Seal (Ours)	MinkUNet	**166.18**	75.38	**37.33**	**42.77**	29.93	**37.73**	**40.32**	20.31	**37.73**	**24.94**
Full	Random	PolarNet	115.09	76.34	58.23	69.91	64.82	44.60	61.91	40.77	53.64	42.01
	Random	CENet	112.79	76.04	67.01	69.87	61.64	58.31	49.97	**60.89**	53.31	24.78
	Random	WaffleIron	106.73	72.78	56.07	73.93	49.59	59.46	65.19	33.12	**61.51**	44.01
	Random	Cylinder3D	105.56	78.08	61.42	71.02	58.40	56.02	64.15	45.36	59.97	43.03
	Random	SPVCNN	106.65	74.70	59.01	72.46	41.08	58.36	65.36	36.83	62.29	**49.21**
	Random	MinkUNet	112.20	72.57	62.96	70.65	55.48	51.71	62.01	31.56	59.64	39.41
	PPKT [65]	MinkUNet	105.64	76.06	64.01	72.18	59.08	57.17	63.88	36.34	60.59	39.57
	SLidR [85]	MinkUNet	106.08	75.99	65.41	72.31	56.01	56.07	62.87	41.94	61.16	38.90
	Seal (Ours)	MinkUNet	**92.63**	**83.08**	**72.66**	**74.31**	**66.22**	**66.14**	**65.96**	57.44	59.87	39.85

L. Kong, et al. Robo3D: Towards robust and reliable 3D perception under corruptions, ICCV, 2023.

Comparative Study: Qualitative Assessment

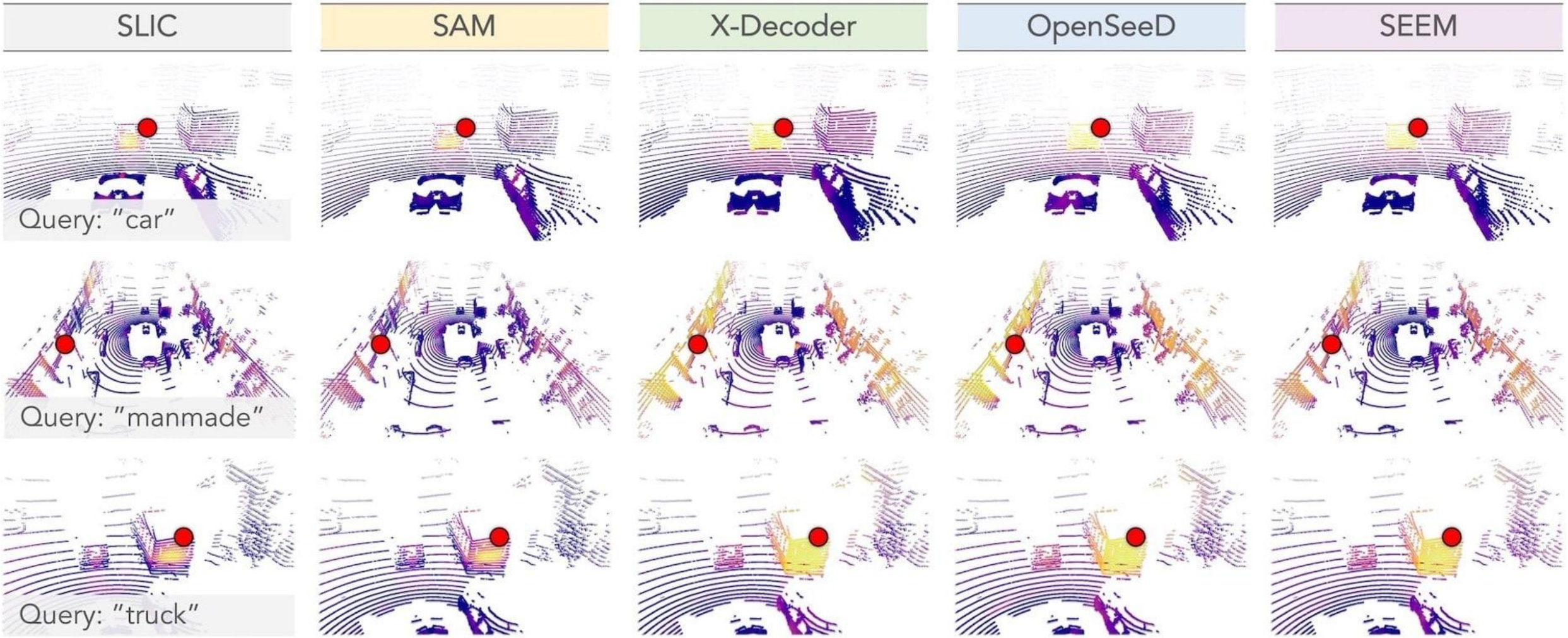


Ablation Study: Frameworks

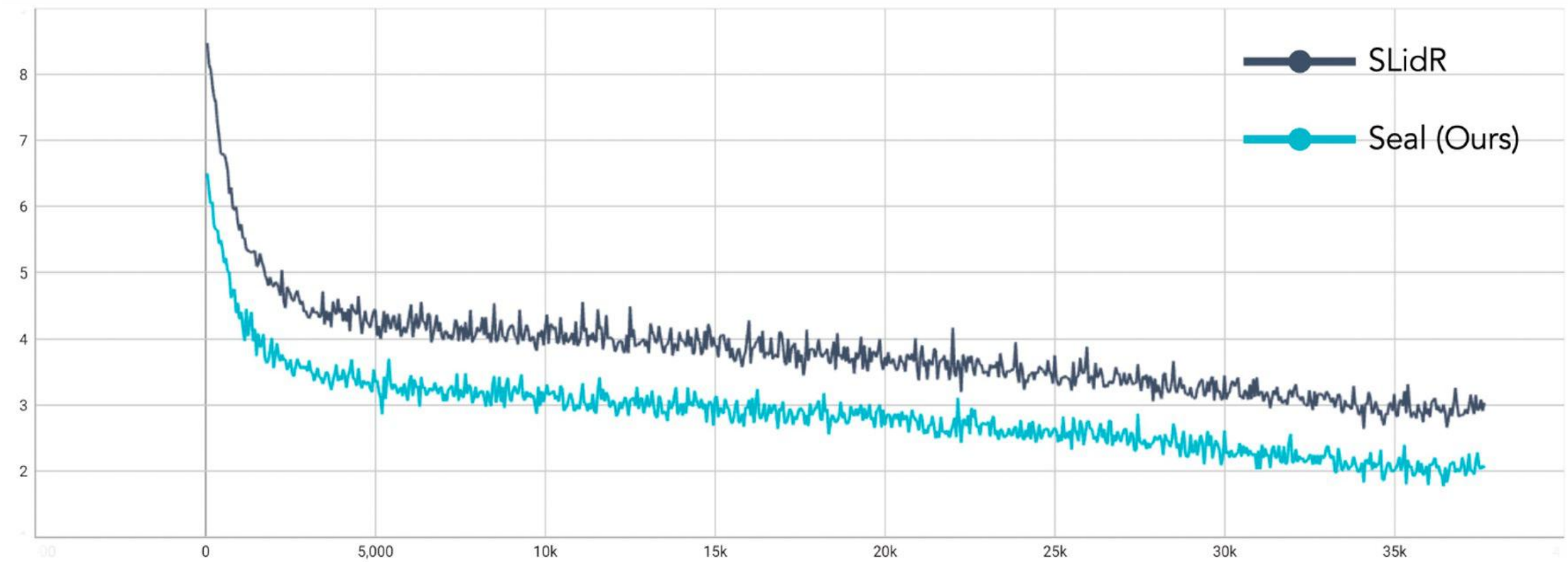
Table 4: Ablation study on pretraining frameworks (ours *vs.* SLidR [85]) and the knowledge transfer effects from different vision foundation models. All mIoU scores are given in percentage (%).

Method	Superpixel	nuScenes LP	nuScenes 1%	nuScenes 5%	nuScenes 10%	nuScenes 25%	nuScenes Full	KITTI 1%	Waymo 1%	Synth4D 1%
Random	-	8.10	30.30	47.84	56.15	65.48	74.66	39.50	39.41	20.22
SLidR	SLIC [1]	38.80	38.30	52.49	59.84	66.91	74.79	44.60	47.12	63.10
	SAM [50]	41.49	43.67	**55.97**	**61.74**	**68.85**	**75.40**	43.35	48.64	63.15
	X-Decoder [122]	41.71	43.02	54.24	61.32	67.35	75.11	45.70	48.73	63.21
	OpenSeeD [111]	42.61	43.82	54.17	61.03	67.30	74.85	**45.88**	48.64	**63.31**
	SEEM [123]	**43.00**	**44.02**	53.03	60.84	67.38	75.21	45.72	**48.75**	63.13
Seal	SLIC [1]	40.89	39.77	53.33	61.58	67.78	75.32	45.75	47.74	63.37
	SAM [50]	43.94	45.09	**56.95**	62.35	**69.08**	**75.92**	46.53	49.00	63.76
	X-Decoder [122]	42.64	44.31	55.18	62.03	68.24	75.56	46.02	49.11	64.21
	OpenSeeD [111]	44.67	44.74	55.13	62.36	69.00	75.64	46.13	48.98	64.29
	SEEM [123]	**44.95**	**45.84**	55.64	**62.97**	68.41	75.60	**46.63**	**49.34**	**64.50**

Ablation Study: Cosine Similarity



Ablation Study: Convergence Speed



The convergence rate comparison between SLidR and the proposed Seal framework

Ablation Study: Components

Table 5: Ablation study of each component pretrained on *nuScenes* [26] and fine-tuned on *nuScenes* [26], *SemanticKITTI* [3], and *Waymo Open* [88]. **C2L:** Camera-to-LiDAR distillation. **VFM:** Vision foundation models. **STC:** Superpoint temporal consistency. **P2S:** Point-to-segment regularization.

#	C2L	VFM	STC	P2S	nuScenes LP	nuScenes 1%	nuScenes 5%	nuScenes 10%	nuScenes 25%	nuScenes Full	KITTI 1%	Waymo 1%
(1)	✓				38.80	38.30	52.49	59.84	66.91	74.79	44.60	47.12
(2)	✓		✓		40.45	41.62	54.67	60.48	67.61	75.30	45.38	48.08
(3)	✓	✓			43.00	44.02	53.03	60.84	67.38	75.21	45.72	48.75
(4)	✓	✓	✓		44.01	44.78	55.36	61.99	67.70	75.00	46.49	49.15
(5)	✓	✓		✓	43.35	44.25	53.69	61.11	67.42	75.44	46.07	48.82
(6)	✓	✓	✓	✓	**44.95**	**45.84**	**55.64**	**62.97**	**68.41**	**75.60**	**46.63**	**49.34**

Open-Source Resources

Seal

Paper

Code

MMDet3D

https://github.com/youquanl/Segment-Any-Point-Cloud