Paper Tag: THU-PM-105

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

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

LaserMix for Semi-Supervised LiDAR Semantic Segmentation


Lingdong Kong


Jiawei Ren

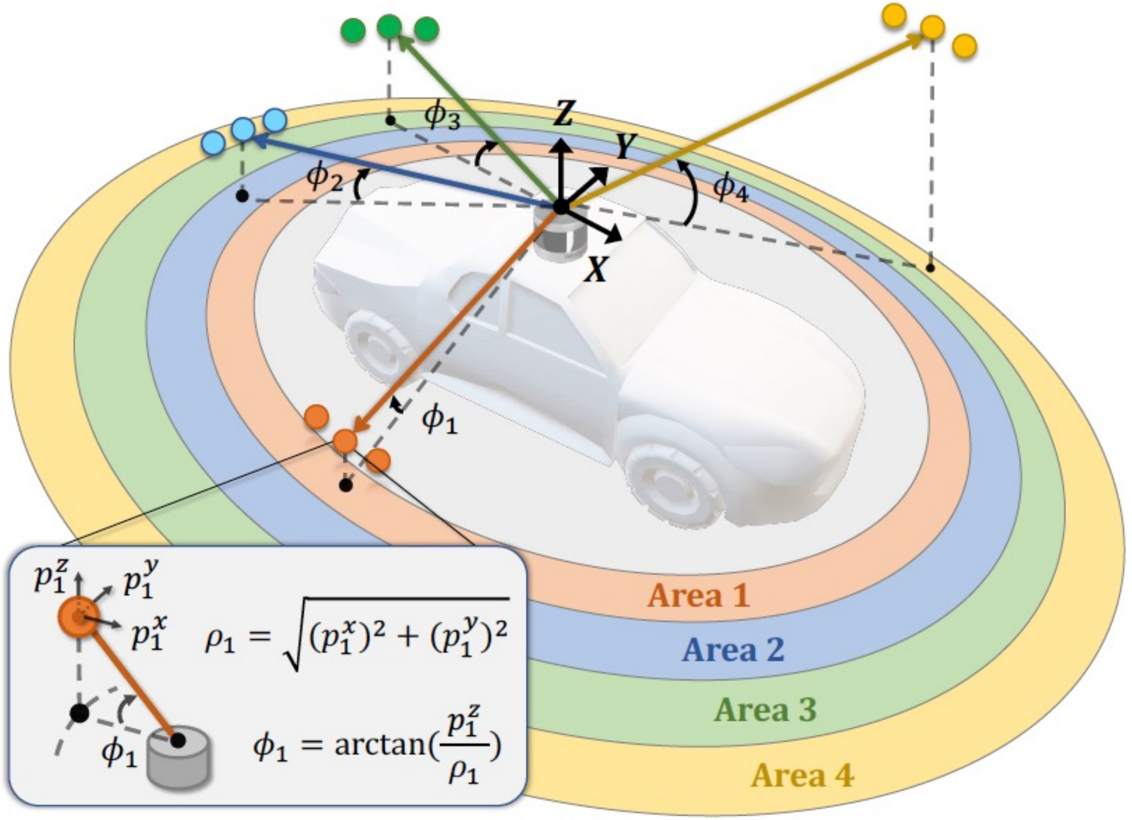

Liang Pan


Ziwei Liu

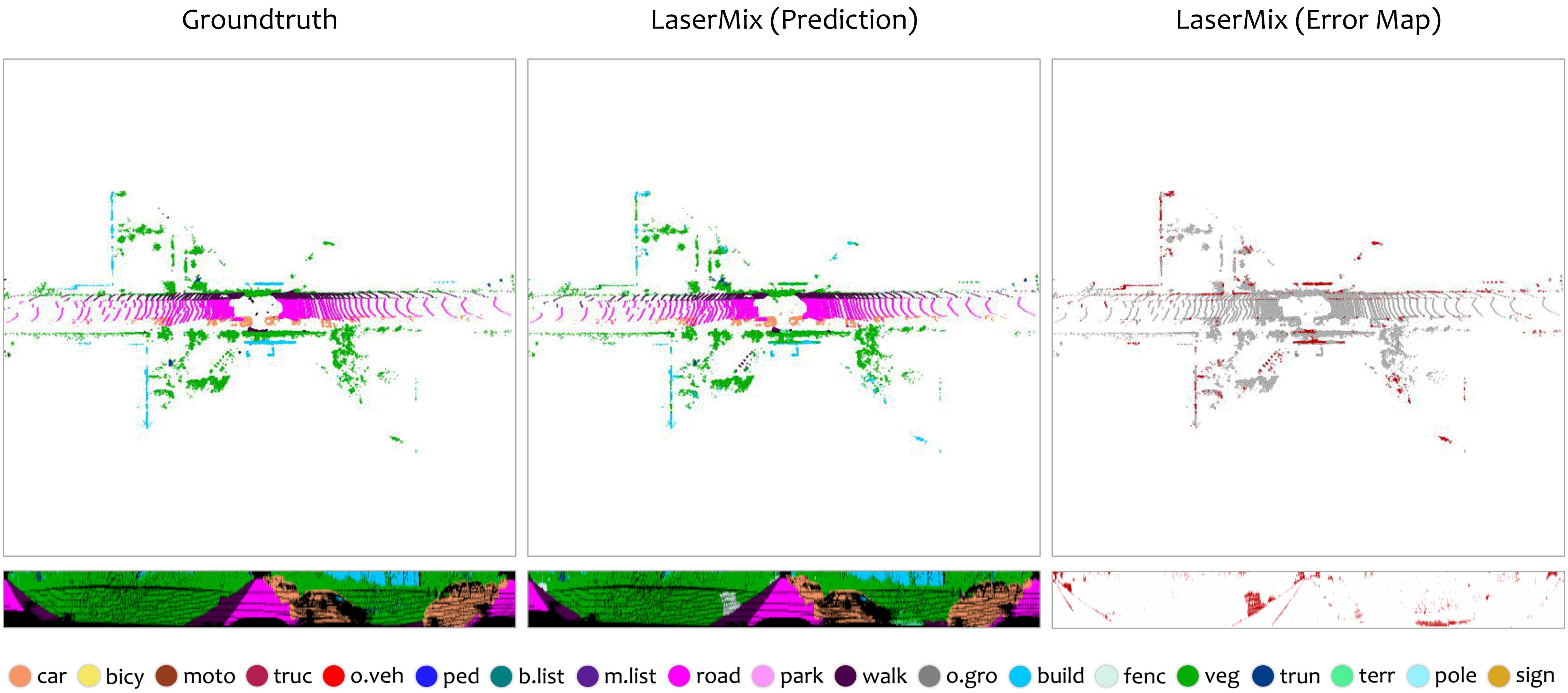
TL;DR

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

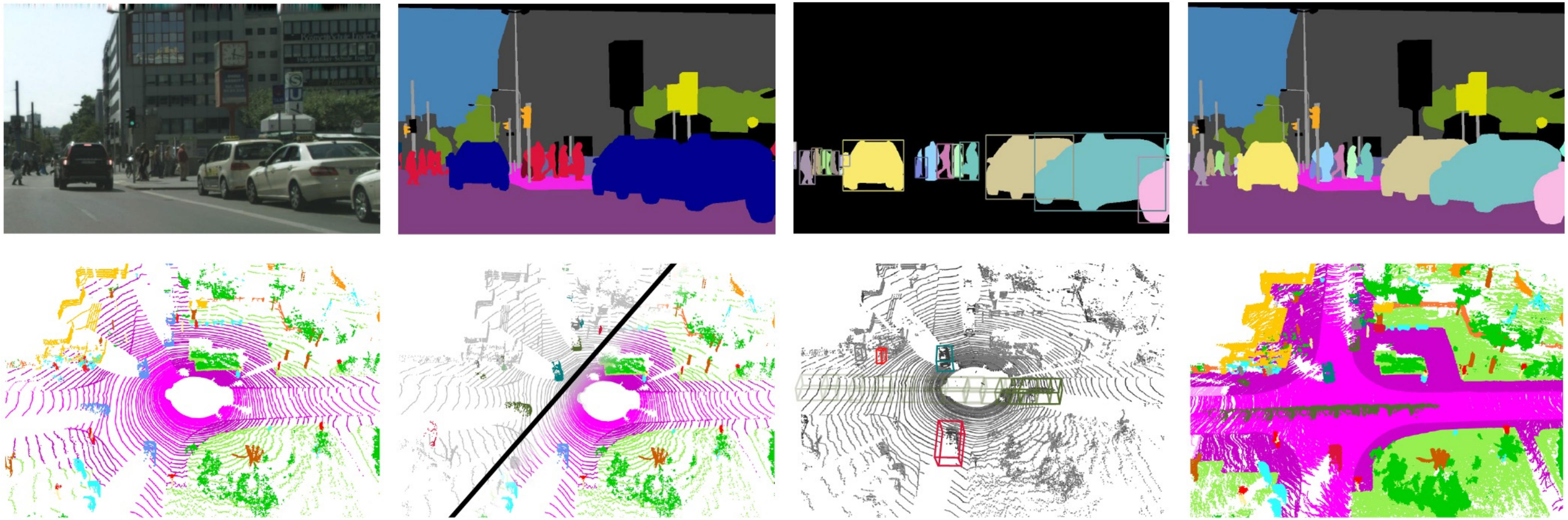
- Leverages the spatial prior in driving scenes for data-efficient learning;
- Constructs low-variational areas via laser beam mixing;
- Encourages the model to make confident and consistent predictions before and after mixing;
- Achieves competitive results over full supervision counterparts with 2x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



TL;DR


Groundtruth
LaserMix (Prediction)
LaserMix (Error Map)
car
bicy
moto
truc
o.veh
ped
b.list
m.list
road
park
walk
o.gro
build
fenc
veg
trun
terr
pole
sign

Autonomous Driving Perception



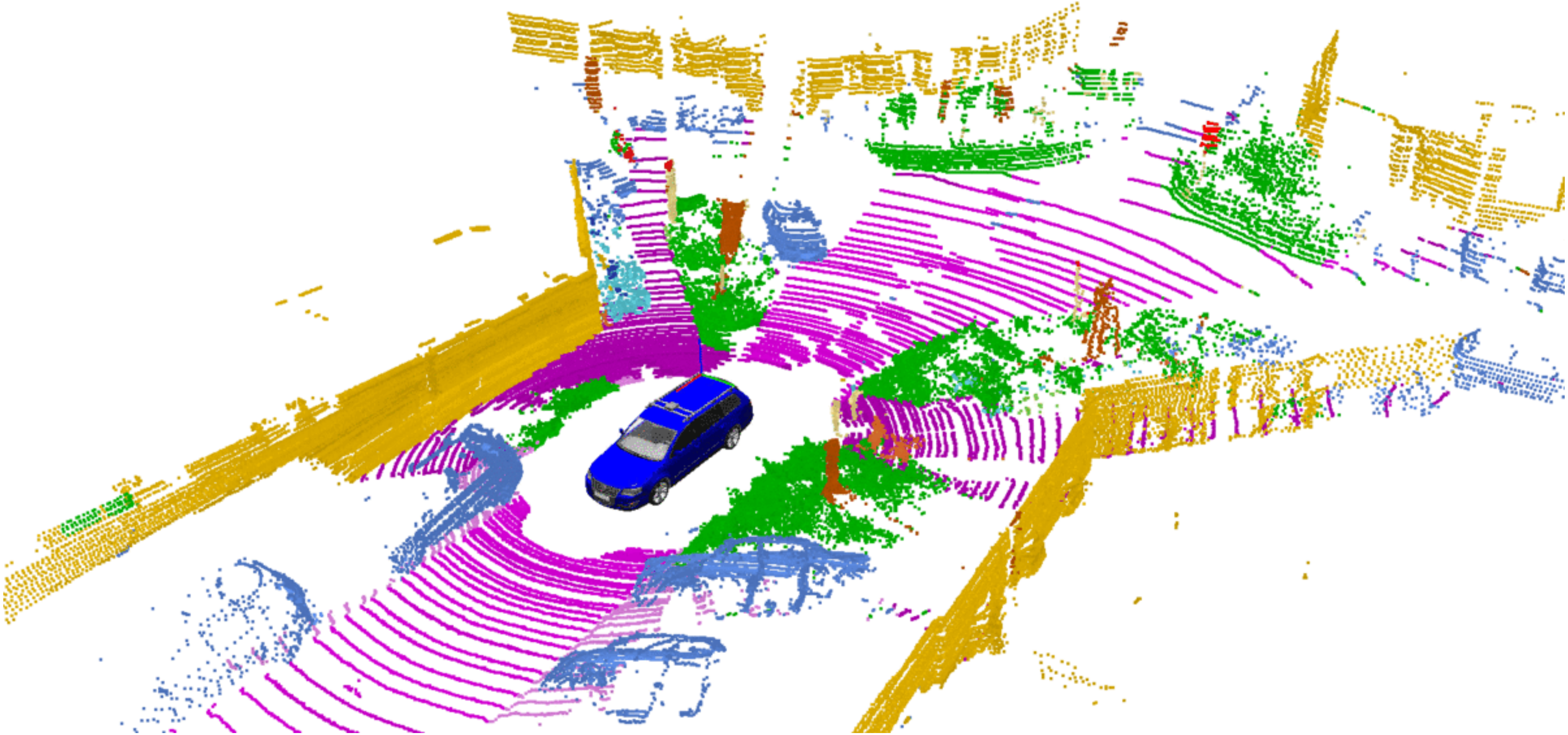
From left to right:

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

Why LiDAR sensors?

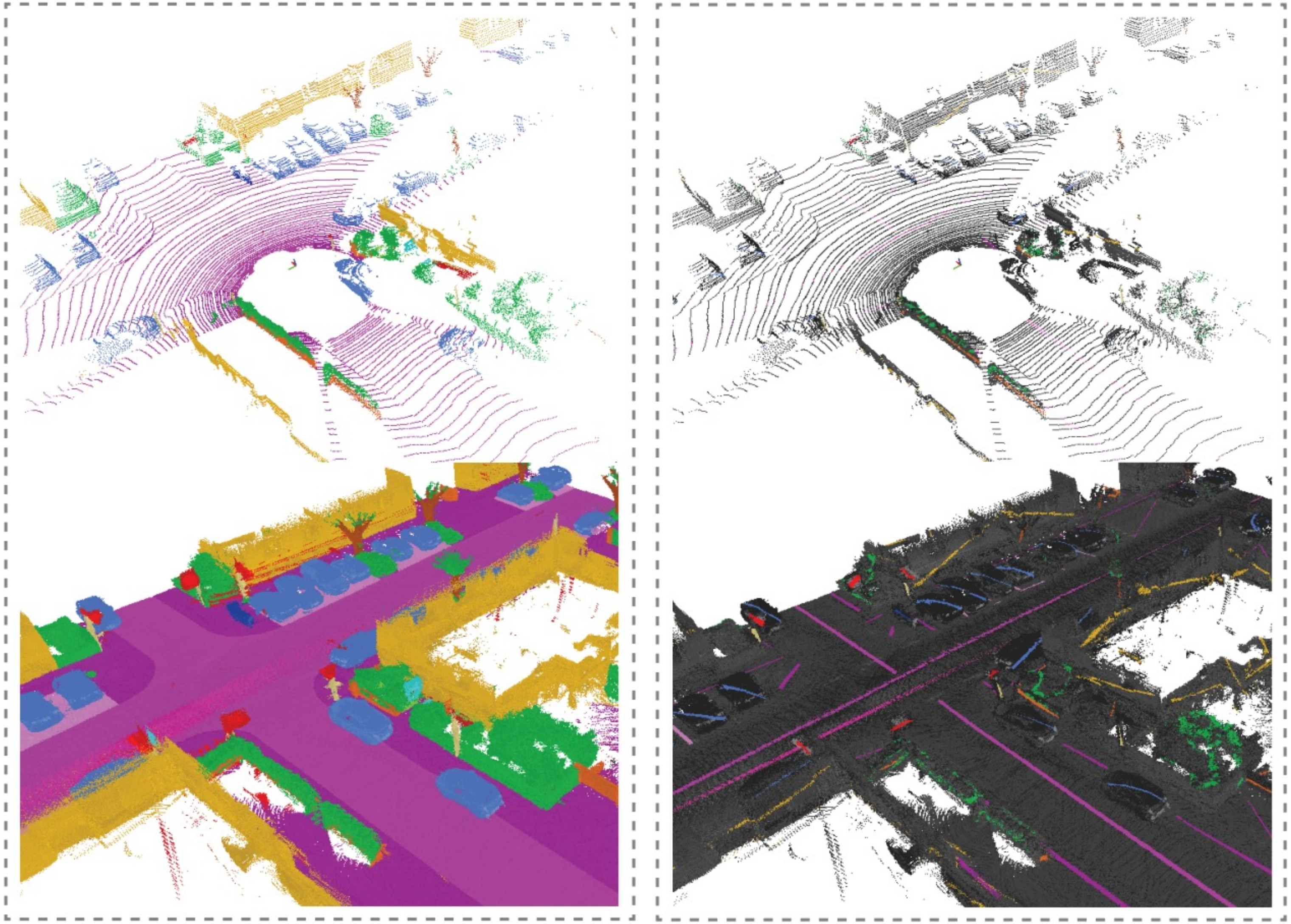
- Accurate depth sensing
- Robust at low-light conditions
- Dense perceptions
- …

LiDAR Semantic Segmentation



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

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

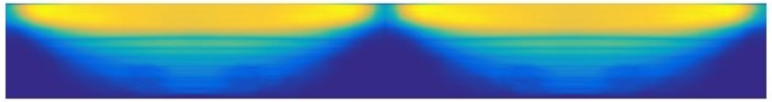
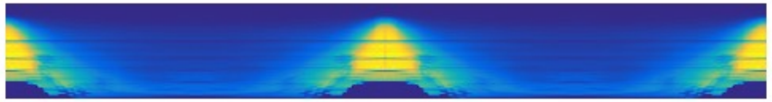
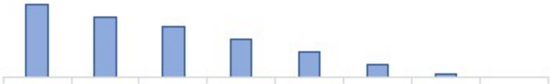
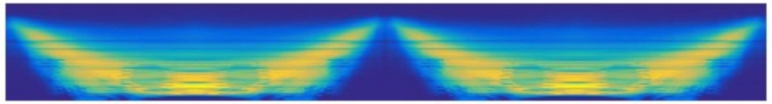
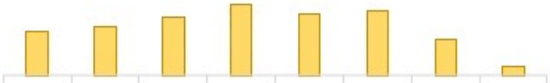
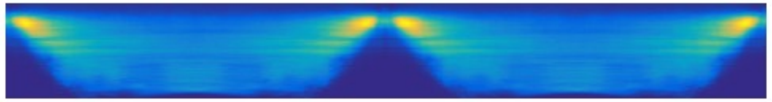
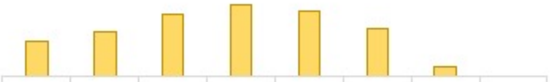
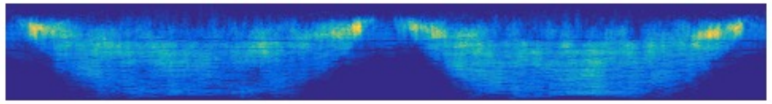
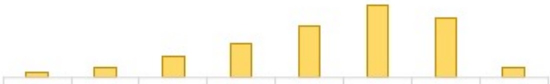
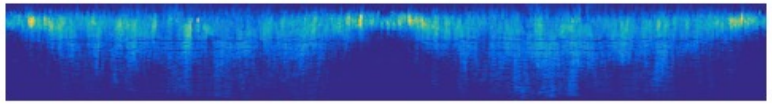
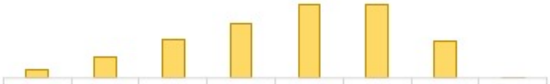
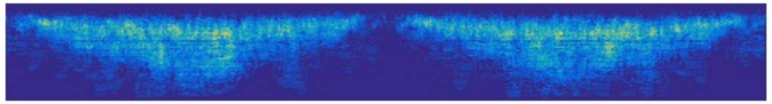
O. Unal, et al. "Scribble-supervised LiDAR semantic segmentation," CVPR, 2022.

Semi-Supervised LiDAR Segmentation

Objective

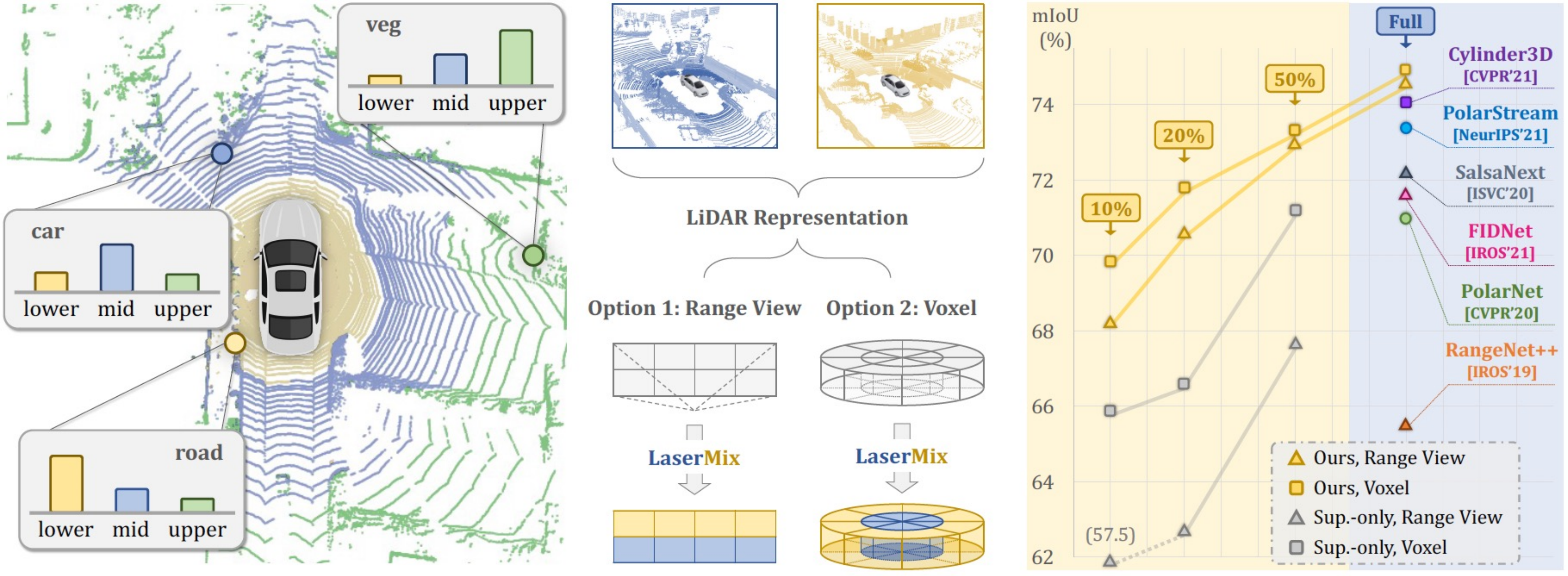
- We target on the less-explored semi-supervised LiDAR semantic segmentation.
- Our goal is to leverage the abundant raw LiDAR scans for training accurate segmentation models.
- We propose **LaserMix** to make advantages of the spatial prior in LiDAR scenes for effective learning with semi supervisions.

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

Overview



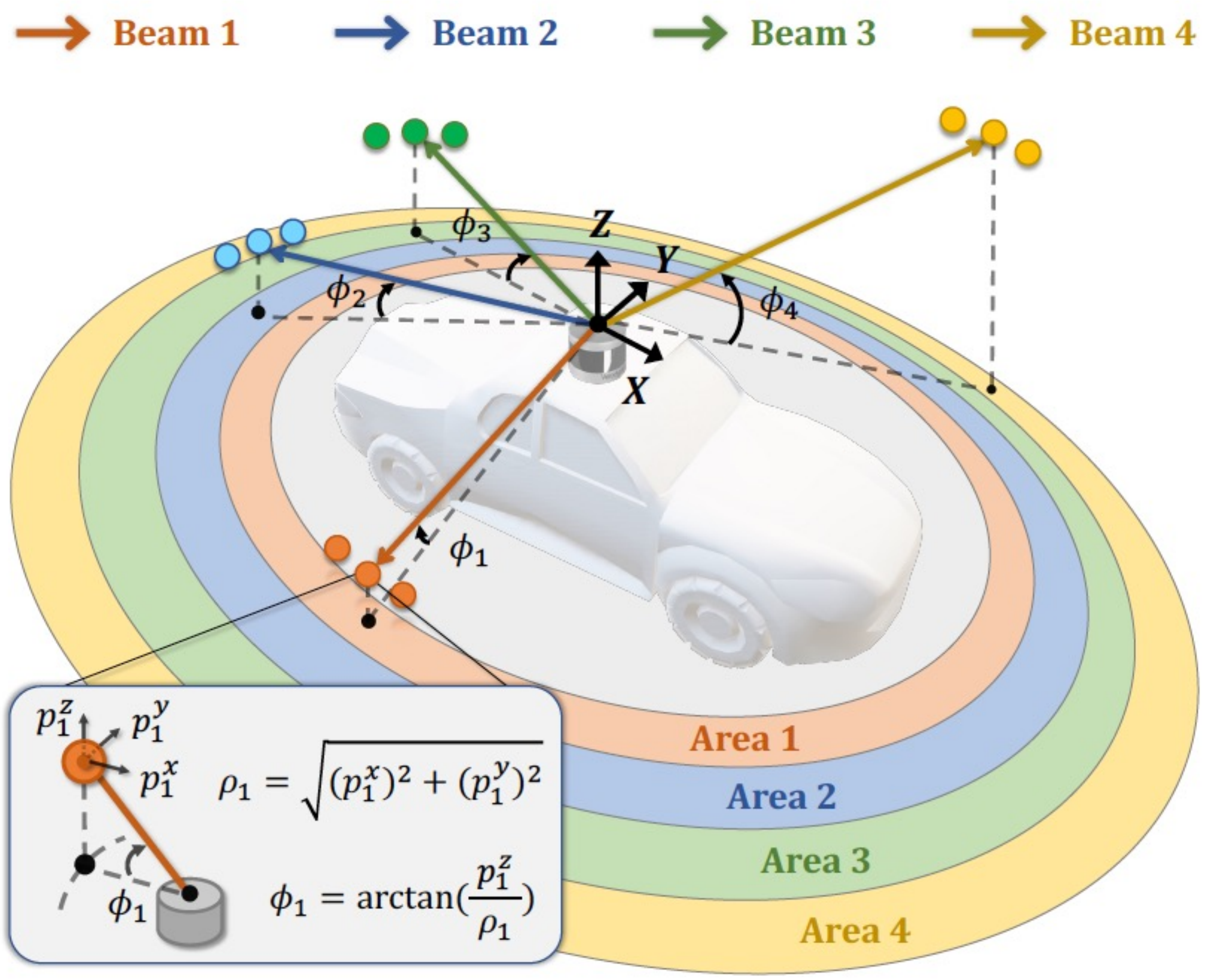
(a) (b) (c)

(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.

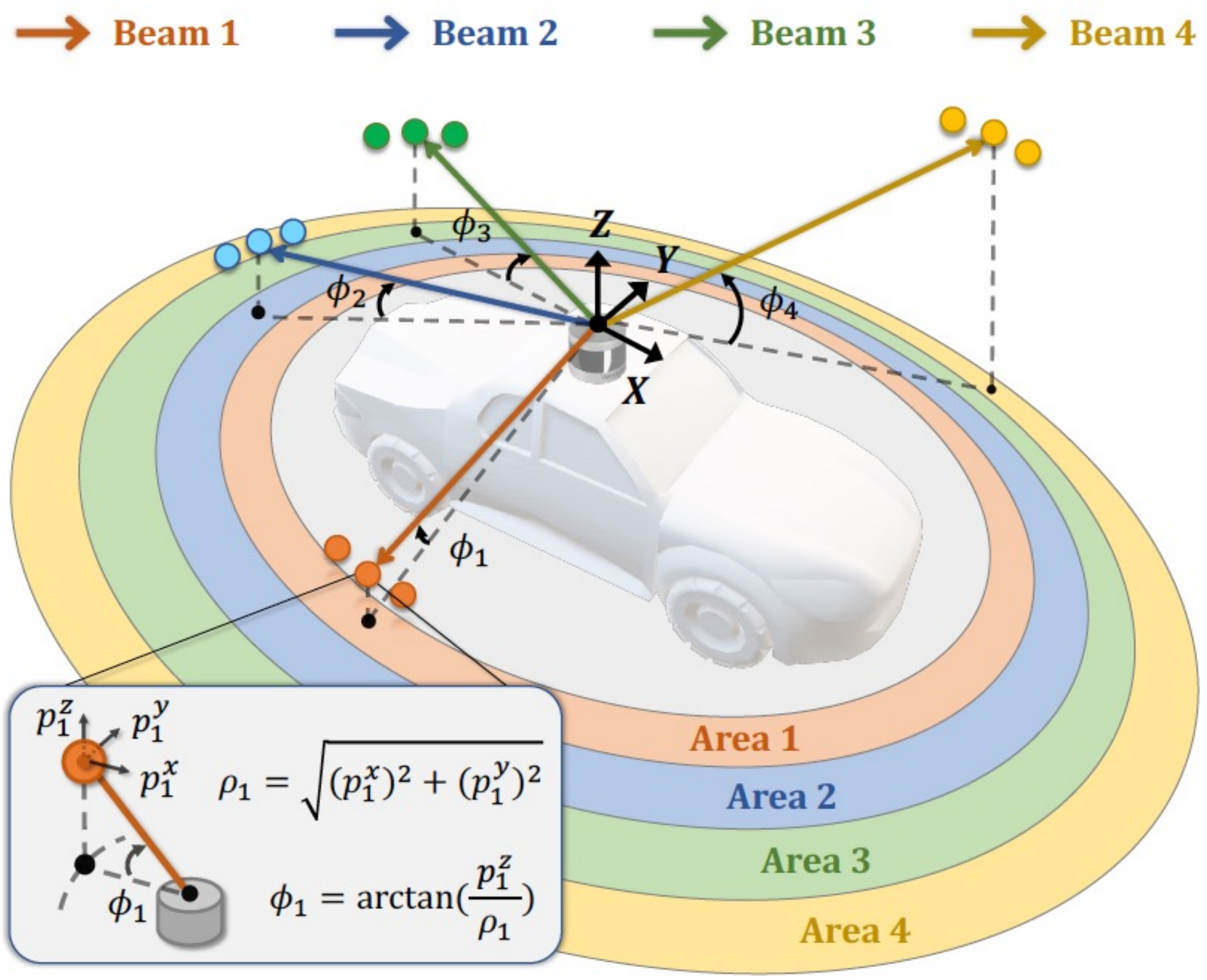
Laser Partition & Mixing



Three-Step Procedure

1. Partitioning the captured LiDAR scan into low-variation 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.

Laser Partition & Mixing



- Inclination:

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

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

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

Laser Partition & Mixing

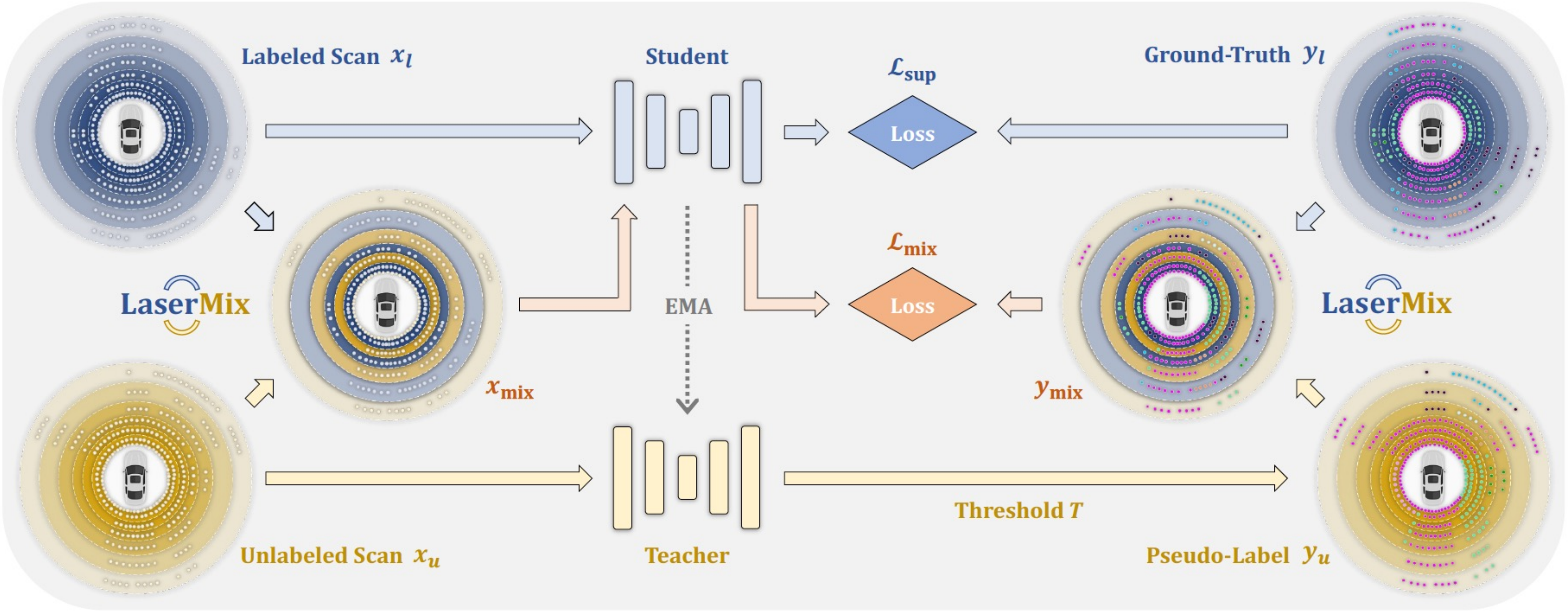
- Inclination:

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

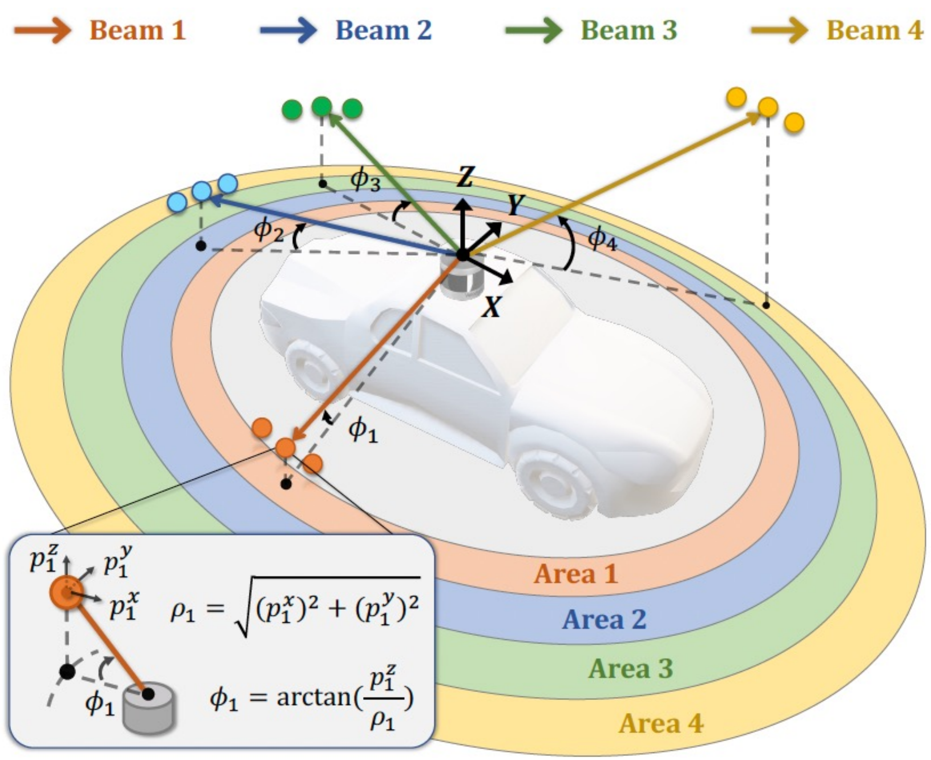
- Depth: $\rho_i = \sqrt{(p_i^x)^2+(p_i^y)^2}$

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

Consistency Regularization



Derivation (See Our Paper)



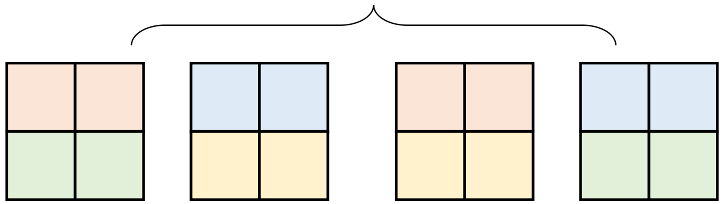
$H(X_{\text{in}}, Y_{\text{in}} \mid A = \text{Outer Rings})$

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

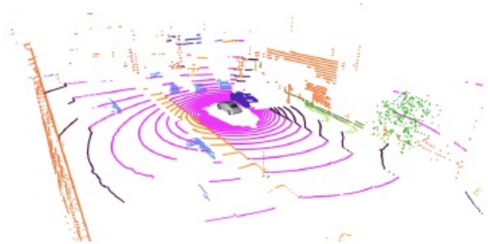
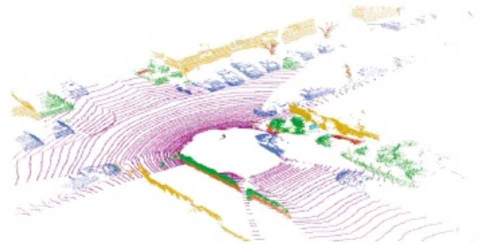
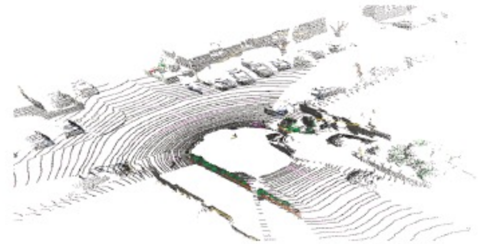
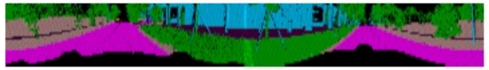
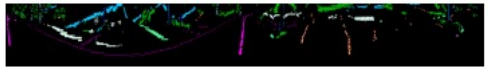


$$H(X, Y|A \in \{2\times2\}) = \log 4$$

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

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

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			

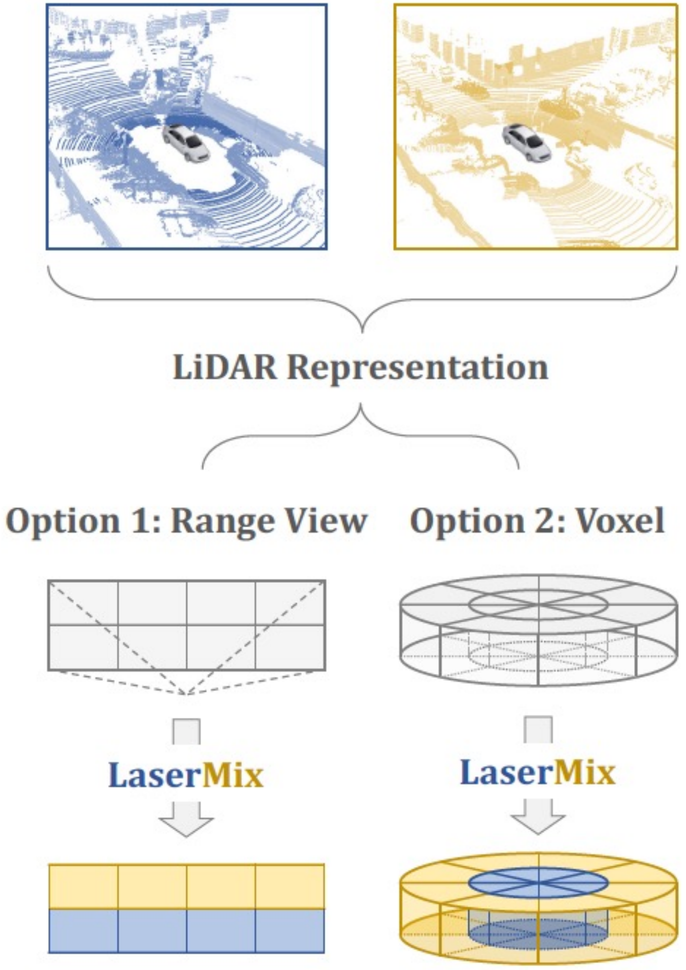
High-res LiDAR:
- SemanticKITTI
- Denser scenes

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.

Experimental Results

Repr.	Method	nuScenes [15] 1%	10%	20%	50%	SemanticKITTI [16] 1%	10%	20%	50%	ScribbleKITTI [4] 1%	10%	20%	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

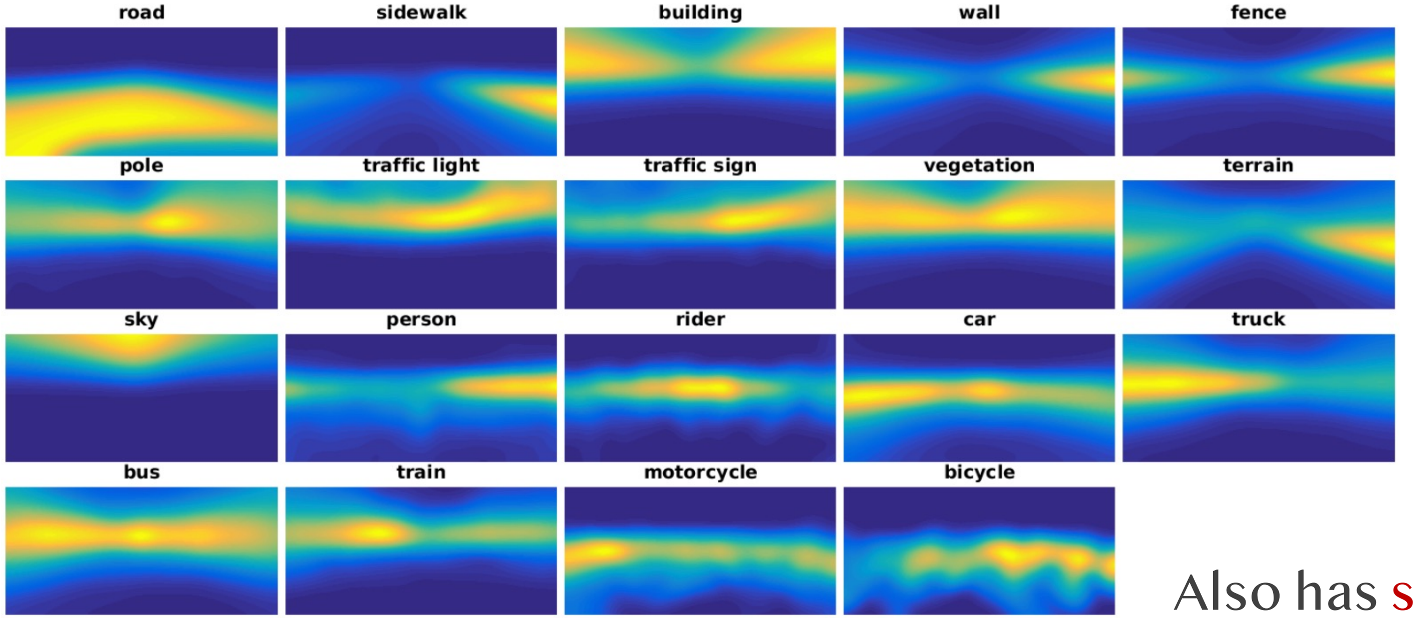
A. Tarvainen and H. Valpola. "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," NeurIPS, 2017.

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.

Experimental Results



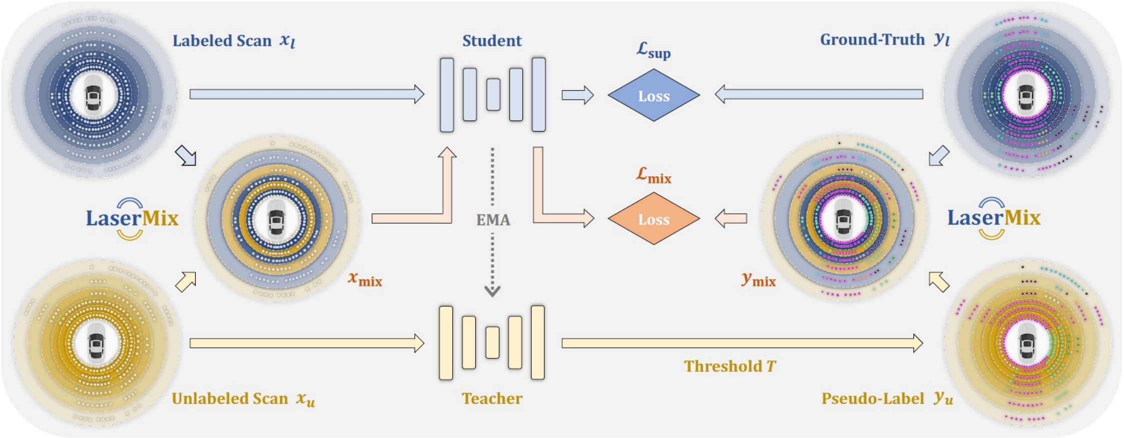
Method	1/16	1/8	1/4	1/2
MeanTeacher [26]	66.1	71.2	74.4	76.3
***w/* Ours**	68.7	72.3	75.7	76.8
$\Delta\uparrow$	+2.6	+1.1	+1.3	+0.5
CCT [11]	66.4	72.5	75.7	76.8
GCT [12]	65.8	71.3	75.3	77.1
CPS [13]	69.8	74.4	76.9	78.6
CPS-CutMix [13]	74.5	76.6	77.8	78.8
***w/* Ours**	75.5	77.1	78.3	79.1
$\Delta\uparrow$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

Also has 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

#	$\mathcal{L}_{mt}$	$\mathcal{L}_{mix}$	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

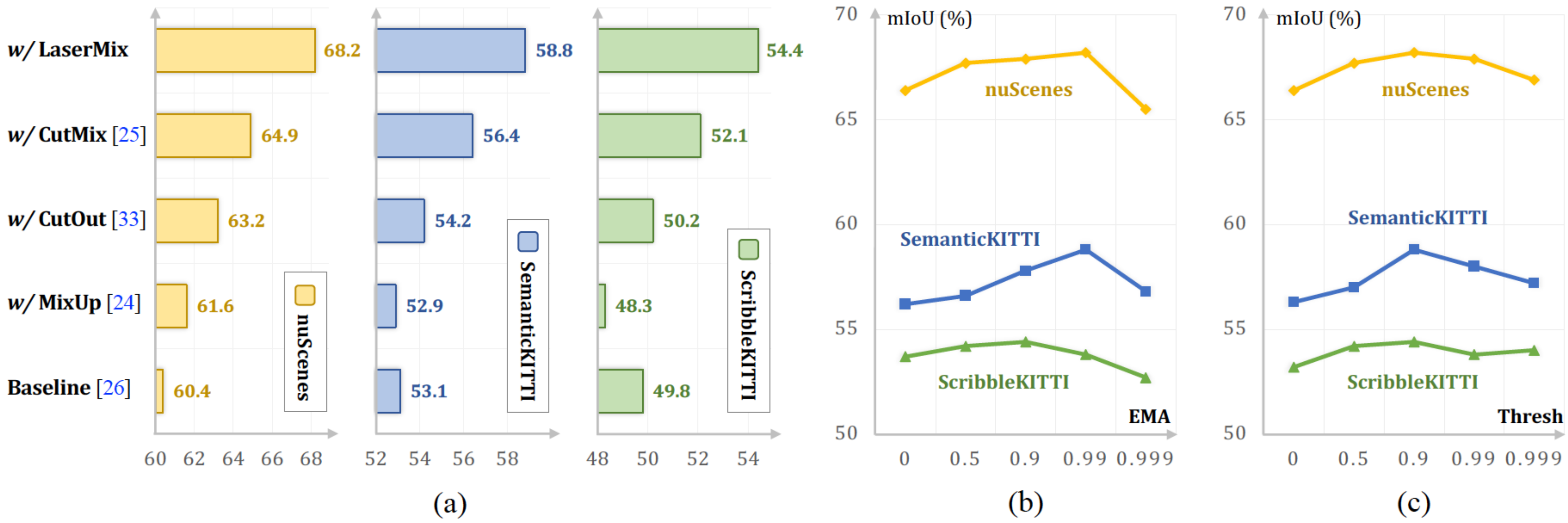


(1) Results of MeanTeacher.

(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. (b) EMA. (c) Confidence threshold.

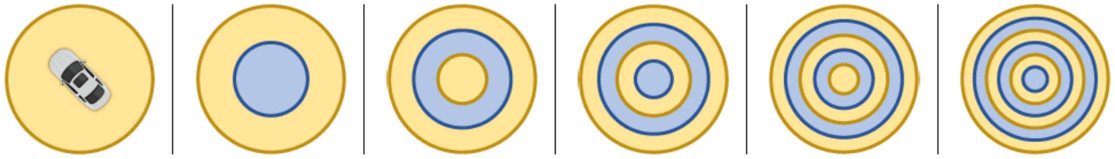
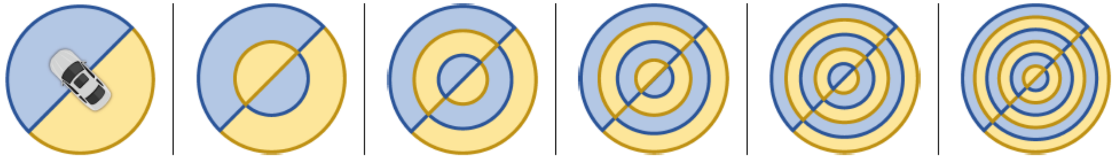
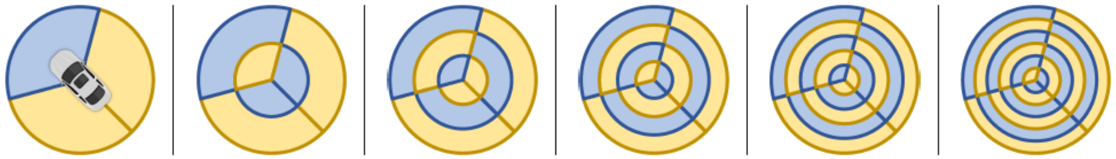
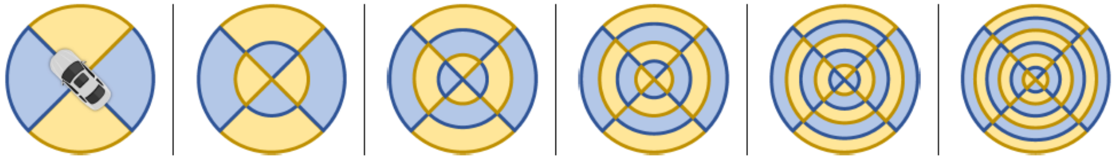
A. Nekrasov, et al. "Mix3D: Out-of-context data augmentation for 3D scenes," 3DV, 2021.

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
- **Tutorial:** https://zhuanlan.zhihu.com/p/528689803
- **Project Page:** https://ldkong.com/LaserMix