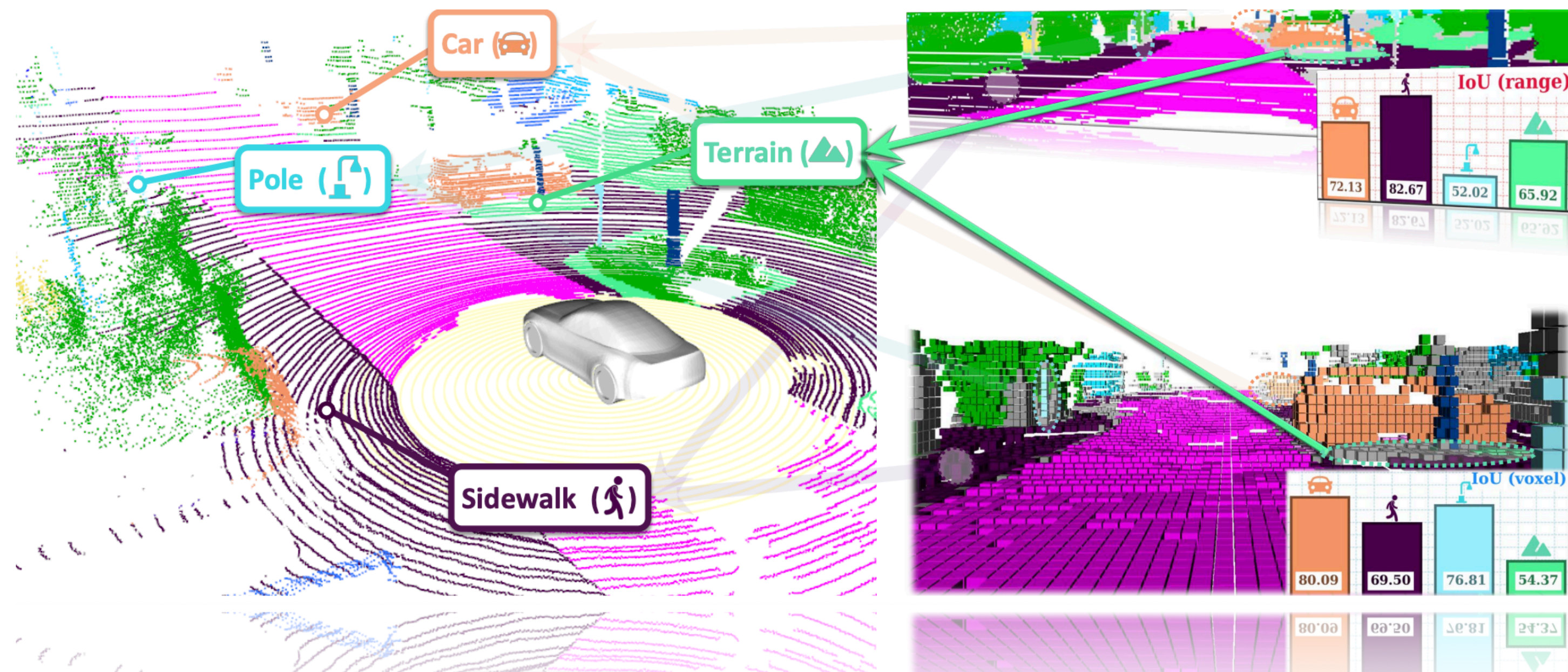


Motivation



◆ Clustering hypothesis: the data (▲) share same semantic meaning in various forms should be clustered together.

- I. Range view suffers from the information loss.
- II. Voxel grids exhibit reduced accuracy in distant dense areas.

◆ Dense point-wise Contrastive Learning.

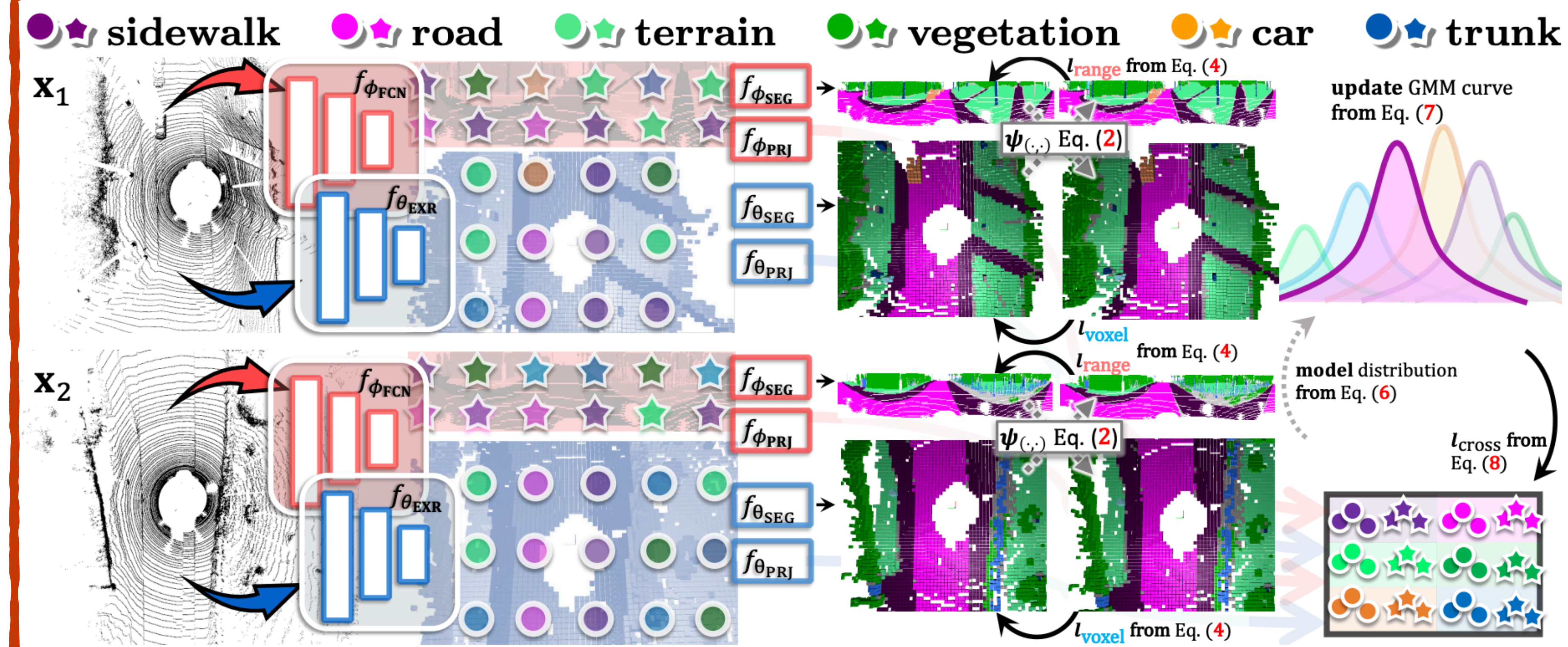
- I. A limited number of embeddings extracted from randomly selected samples.
- II. Potential noisy data or model bias from one network can adversely affect others

Contribution

◆ **ItTakesTwo** A semi-supervised LiDAR segmentation framework leveraging peer perturbation.

◆ A novel contrastive learning method samples positive/negative embedding pairs from Gaussian Mixture Model distributions based on two representations.

Methodology



Experiments (measured by mIoU)

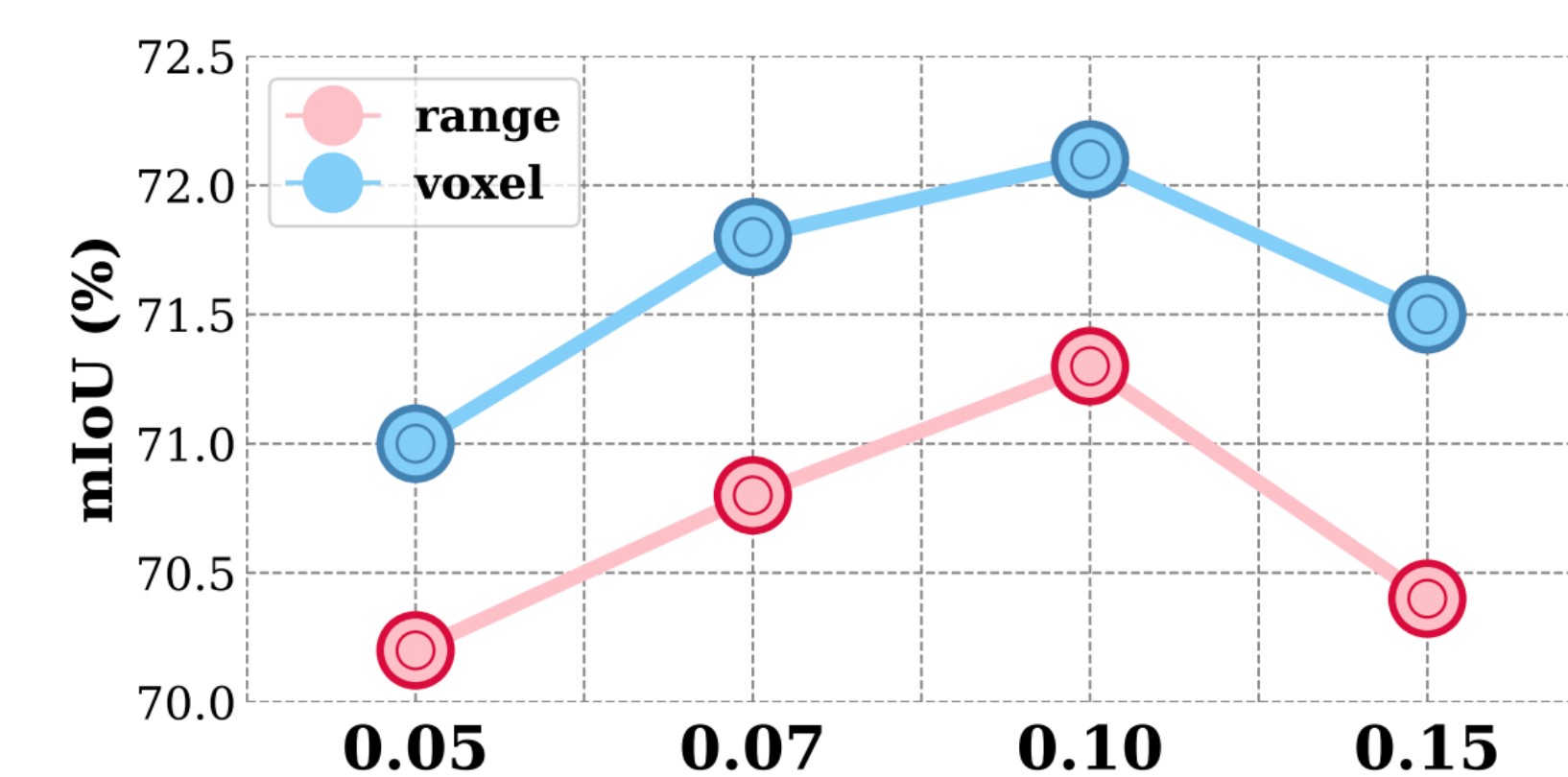
Tab). Uniform sampling results across all benchmarks, where the range view is based on FidNet and the voxel grid representation is based on Cylinder3D.

Repr.	Method	nuScenes [4]				SemanticKITTI [2]				ScribbleKITTI [55]			
		1%	10%	20%	50%	1%	10%	20%	50%	1%	10%	20%	50%
Range	sup.	38.3	57.5	62.7	67.6	36.2	52.2	55.9	57.2	33.1	47.7	49.9	52.5
	MT [53]	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 [71]	40.9	60.5	64.3	69.3	39.9	53.4	56.1	56.9	35.7	50.7	52.7	54.6
	CPS [6]	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 [27]	49.5	68.2	70.6	73.0	43.4	58.8	59.4	61.4	38.3	54.4	55.6	58.7
	IT2 (Ours)	56.5	71.3	73.4	74.0	51.9	60.3	61.7	62.1	46.6	57.1	57.3	58.6
Voxel	sup.	50.9	65.9	66.6	71.2	45.4	56.1	57.8	58.7	39.2	48.0	52.1	53.8
	MT [53]	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 [71]	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 [6]	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 [27]	55.3	69.9	71.8	73.2	50.6	60.0	61.9	62.3	44.2	53.7	55.1	56.8
	IT2 (Ours)	57.5	72.1	73.6	74.1	52.0	61.4	62.1	62.5	47.9	56.7	57.5	58.3

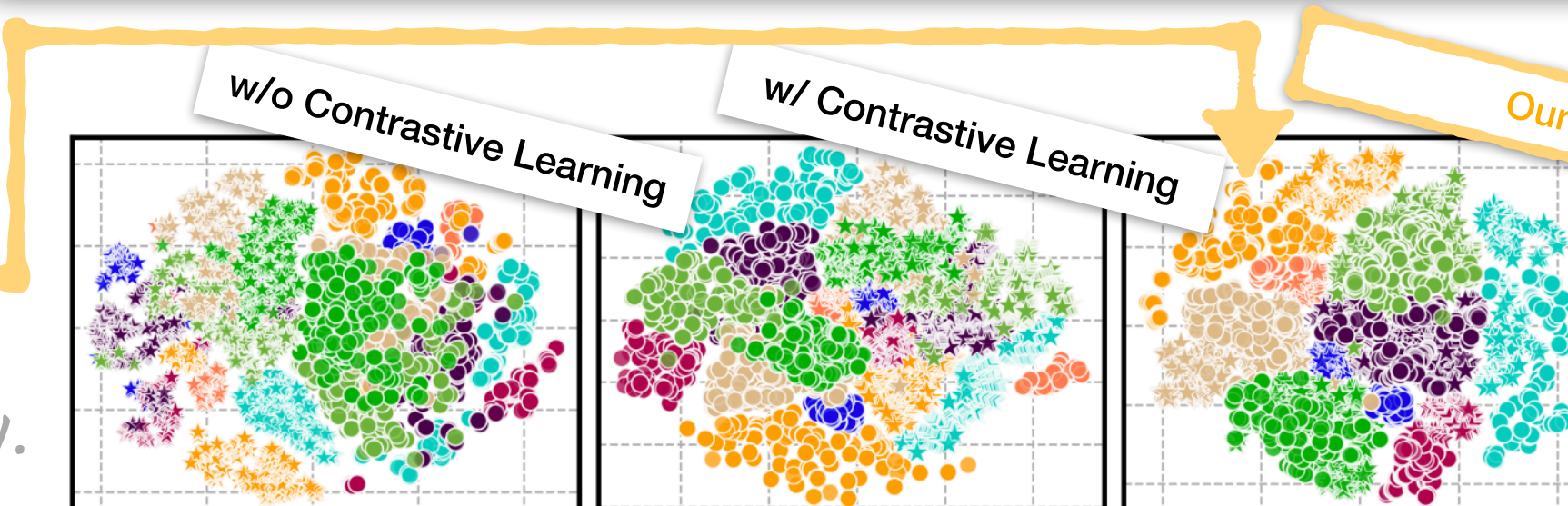
Ablation Studies

Tab). Ablation studies for components of our methods in the nuScenes dataset. Best results are marked in red.

Repr.	IT2	Ctrs.	Aug.	nuScenes [4]		
				10%	20%	50%
Range	✗	✗	✗	60.8	64.9	68.0
	✓	✗	✗	67.3 (6.5↑)	70.8 (5.9↑)	72.6 (4.6↑)
	✓	✓	✗	70.3 (9.5↑)	72.6 (7.7↑)	73.5 (5.5↑)
	✓	✓	✓	71.3 (10.5↑)	73.4 (8.5↑)	74.0 (6.0↑)
Voxel	✗	✗	✗	66.3	70.0	72.5
	✓	✗	✗	70.4 (4.1↑)	71.9 (1.9↑)	73.1 (0.6↑)
	✓	✓	✗	71.6 (5.3↑)	72.9 (2.9↑)	73.9 (1.4↑)
	✓	✓	✓	72.1 (5.8↑)	73.6 (3.6↑)	74.1 (1.6↑)



Curve). Ablation studies for different temperatures in our proposed cross-distribution contrastive learning.



- ★ IT2 significantly outperforms the single repr. baseline.
- ★ Cross-distribution Ctrs. enhances latent space layout and results (via GMM curve).
- ★ Representation specific Aug. further improves accuracy.

Visualisation

Fig). Error maps visualised from LiDAR points (top) and range view (bottom). The correct predictions are in green and the mistakes are highlighted in red.

