Less is More

Reducing Task and Model Complexity for 3D Point Cloud Semantic Segmentation

Li Li, Hubert P. H. Shum, Toby P. Breckon, Durham University, UK

Durham University
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Main Contributions

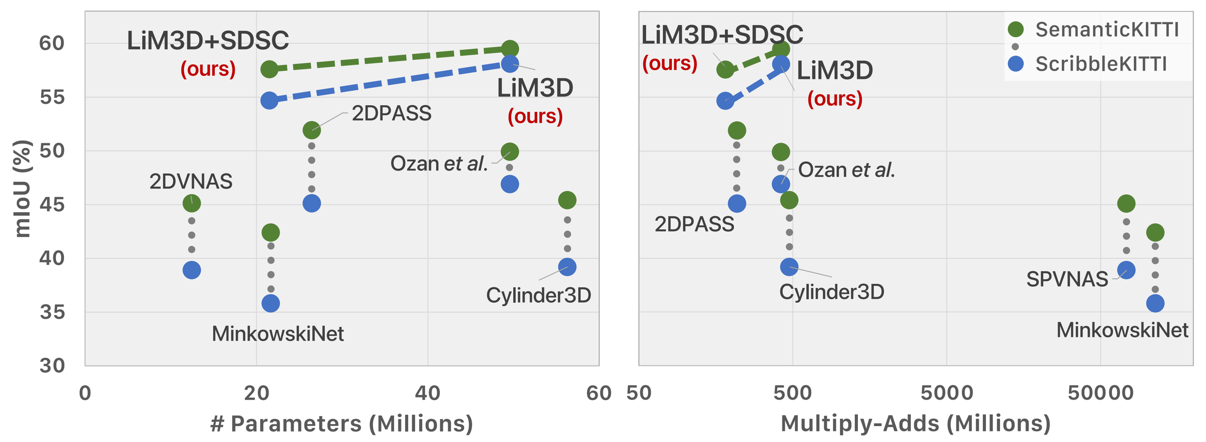


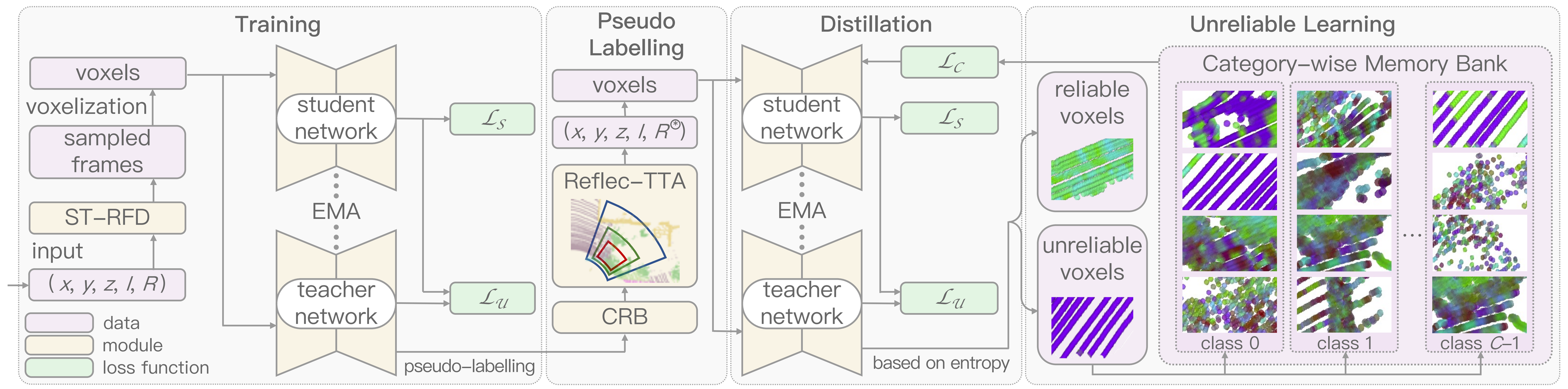
Figure 1. mIoU (%) against parameters and Mult-Adds @5% sampling protocol.

- **LiM** A novel methodology for semi-supervised 3D LiDAR semantic segmentation that uses significantly **Less parameters** and offers (**More**) **superior accuracy**.
- **SDSC** A novel Sparse Depthwise Separable Convolution (SDSC) module, to **reduce trainable network parameters**, and to both reduce the likelihood of over-fitting and facilitate a deeper network architecture.
- **ST-RFD** A novel Spatio-Temporal Redundant Frame Down-sampling (ST-RFD) strategy, to extract a maximally diverse data subset for training by **removing temporal redundancy** and hence future training requirements.
- **Reflec-TTA** **UPL** A novel soft pseudo-labeling method informed by **LiDAR reflectivity** as a proxy to in-scene object material properties, facilitating effective use of limited data annotation.

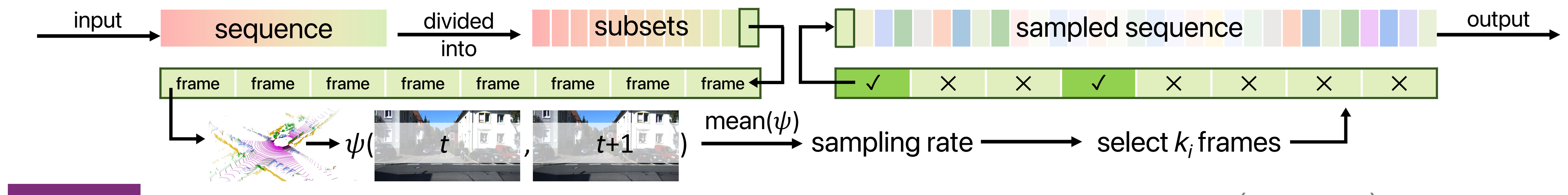
Method	# Parameters	# Mult-Adds	SeK [7]	ScK [46]
Cylider3D [63]	56.3	476.9M	45.4	39.2
Unal *et al.* [46]	49.6	420.2M	49.9	46.9
2DPASS [58]	26.5	<u>217.4M</u>	51.7	45.1
MinkowskiNet [13]	21.7	114.0G	42.4	35.8
SPVNAS [44]	**12.5**	73.8G	45.1	38.9
LiM3D+SDSC (ours)	<u>21.5</u>	**182.0M**	<u>57.6</u>	<u>54.7</u>
LiM3D (ours)	49.6	420.2M	**59.5**	**58.1**

Table 3. The computation cost and mIoU (%) under 5%-labeled training results on SemanticKITTI (SeK) and ScribbleKITTI (ScK).

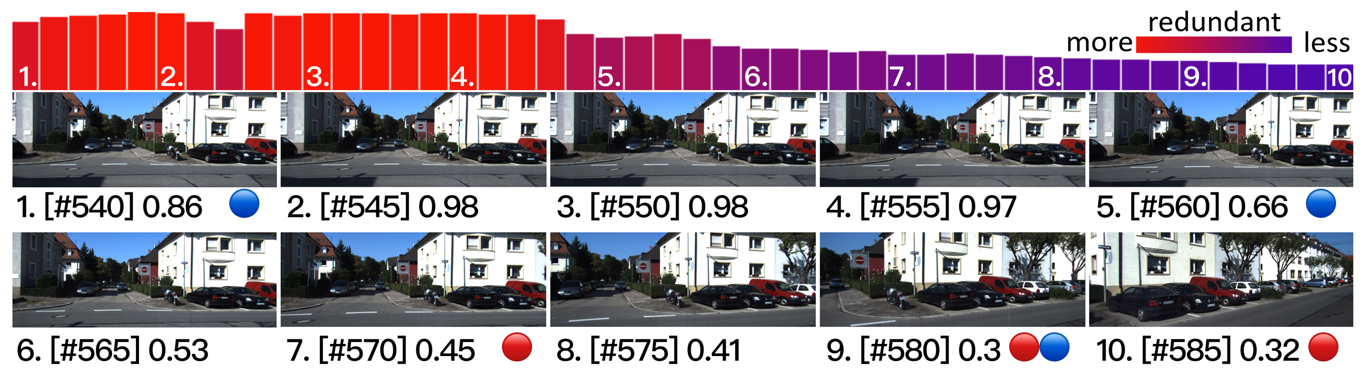
Methodology



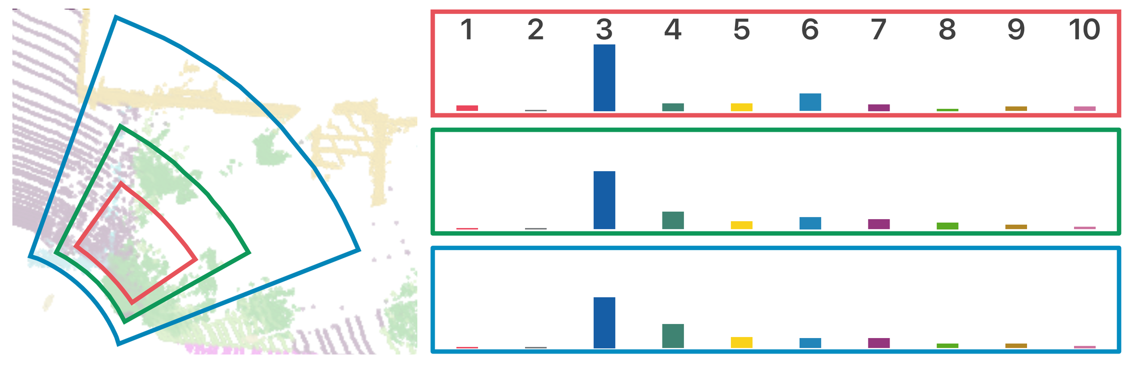
LiM *Figure 2.* Our proposed architecture for unreliable pseudo-labels LiDAR semantic segmentation involves three stages: training, pseudo-labeling, and distillation with unreliable learning.



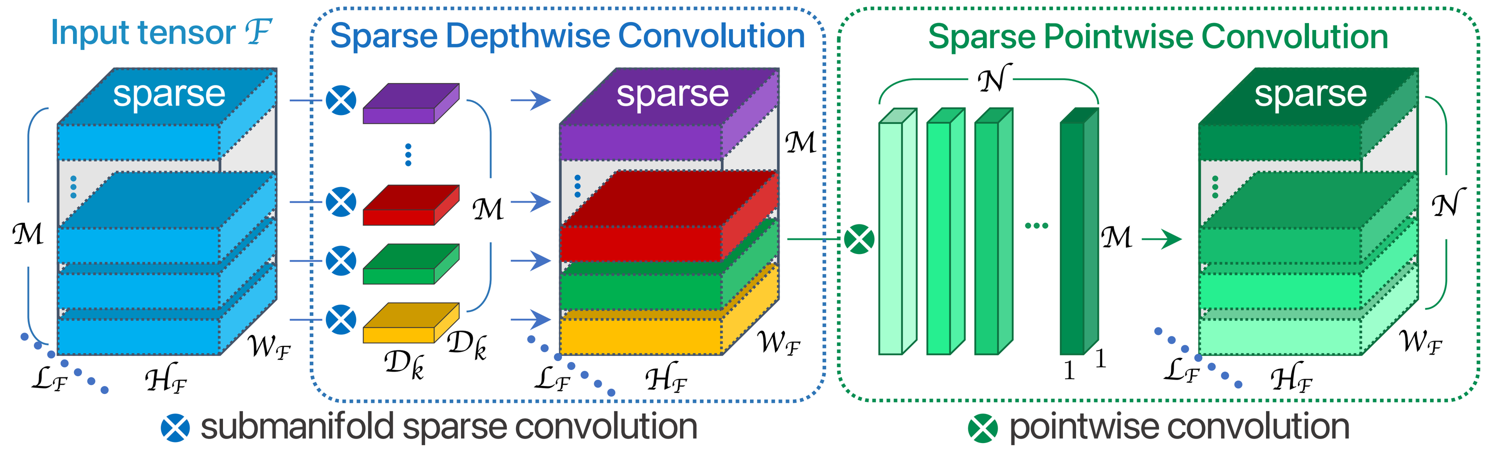
ST-RFD *Figure 6.* Overview of Spatio-Temporal Redundant Frame Downsampling (ST-RFD) approach.



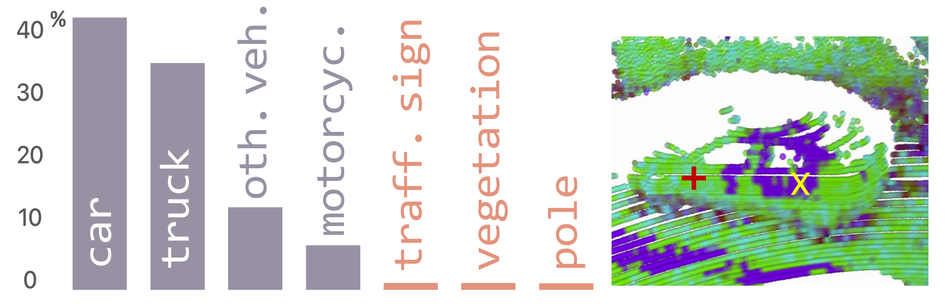
ST-RFD *Figure 5.* Illustration of LiDAR frame temporal correlation as [# frame ID] redundancy with 5% sampling on SemanticKITTI [7] using uniform sampling (selected frames in 🔵) and ST-RFD strategy (🔴).



Reflec-TTA *Figure 4.* Coarse histograms of Reflec-TTA bins.



SDSC *Figure 7.* Illustration of the Sparse Depthwise Separable Convolution (SDSC) module.



UPL *Figure 3.* Illustration on Unreliable Pseudo-Labels (UPL). Left: entropy predicted from an unlabeled point cloud — greener: lower entropy. Right: Category-wise probability of an unreliable prediction X.

Evaluation & Results

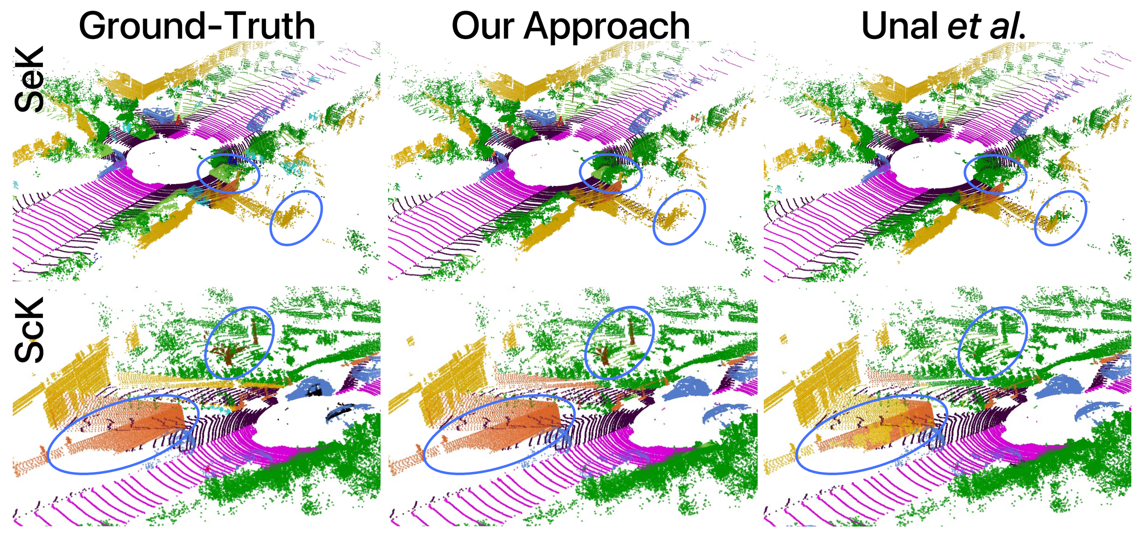


Figure 8. Comparing the 10% sampling split with ground-truth (left), our approach (middle) and Unal *et al.* [46] (right).

UP	RF	RT	ST	SD	Training mIoU (%) 5%	10%	20%	40%	Validation mIoU (%) 5%	10%	20%	40%	#Params (M)
					82.8	87.5	87.8	88.2	54.8	58.1	59.3	60.8	49.6
✓					–	–	–	–	55.9	58.8	59.9	61.2	49.6
✓	✓				83.6	88.3	88.7	89.1	56.8	59.6	60.5	61.4	49.6
✓		✓			–	–	–	–	57.5	59.8	61.2	62.6	49.6
✓	✓	✓			–	–	–	–	58.7	61.3	62.4	62.8	49.6
✓	✓	✓	✓		**85.2**	**89.1**	**89.5**	**89.7**	**59.5**	**62.2**	**63.1**	**63.3**	49.6
✓	✓	✓	✓	✓	83.8	88.6	89.0	89.2	57.6	61.0	61.7	62.1	**21.5**

Table 2. Component-wise ablation of LiM3D (mIoU as %, and #parameters in millions, M) where UP, RF, RT, ST, SD denote Unreliable Pseudo-labeling, Reflectivity, Reflec-TTA, ST-RFD, and SDSC.

Repr.	Samp.	Method		SemanticKITTI [7] 1%	5%	10%	20%	40%	50%	100%	ScribbleKITTI [46] 1%	5%	10%	20%	40%	50%	100%
Range	U	LaserMix [32]	(2022)	43.4	–	58.8	59.4	–	61.4	–	38.3	–	54.4	55.6	–	58.7	–
Voxel	U	Cylinder3D [63]	(CVPR'21)	–	45.4	56.1	57.8	58.7	–	67.8	–	39.2	48.0	52.1	53.8	–	56.3
	U	LaserMix [32]	(2022)	50.6	–	60.0	<u>61.9</u>	–	62.3	–	44.2	–	53.7	55.1	–	56.8	–
	P	Jiang *et al.* [29]	(ICCV'21)	–	41.8	49.9	58.8	59.9	–	65.8	–	–	–	–	–	–	–
	U	Unal *et al.* [46]	(CVPR'22)	–	49.9*	58.7*	59.1*	60.9	–	<u>68.2</u>*	–	46.9*	54.2*	56.5*	58.6*	–	<u>61.3</u>
	S	LiM3D+SDSC	(ours)	<u>57.2</u>	<u>57.6</u>	<u>61.0</u>	61.7	<u>62.1</u>	<u>62.7</u>	67.5	<u>55.8</u>	<u>56.1</u>	<u>56.9</u>	<u>57.2</u>	<u>58.9</u>	<u>59.3</u>	60.7
	S	LiM3D	(ours)	**58.4**	**59.5**	**62.2**	**63.1**	**63.3**	**63.6**	**69.5**	**57.0**	**58.1**	**61.0**	**61.2**	**62.0**	**62.1**	**62.4**

Table 1. Comparative mIoU for Range- and Voxel-based methods using Uniform sampling (U), sequential partition (P) and ST-RFD sampling (S): **bold**/<u>underlined</u> = **best**/<u>2nd</u> best.

Sampling	SemanticKITTI [7] 5%	10%	20%	40%	ScribbleKITTI [46] 5%	10%	20%	40%
Random	58.5	61.6	62.6	62.7	57.1	60.3	<u>60.5</u>	60.9
Uniform	58.7	61.3	62.4	62.8	56.9	60.6	60.3	61.0
ST-RFD-R	<u>59.1</u>	**62.4**	<u>62.9</u>	**63.4**	<u>58.0</u>	<u>60.7</u>	**61.2**	<u>61.8</u>
ST-RFD	**59.5**	<u>62.2</u>	**63.1**	<u>63.3</u>	**58.1**	**61.0**	**61.2**	**62.0**

Table 4. Effects of ST-RFD sampling (mIoU as %).

Ratio	Unreliable mIoU	Unreliable SS/FF	Reliable mIoU	Reliable SS/FF	Random mIoU	Random SS/FF
5%	**59.5**	**85.6**	57.2	82.3	56.4	81.2
10%	**62.2**	**89.5**	60.8	87.5	59.7	85.9
20%	**63.1**	**90.8**	61.4	88.3	60.5	87.1
40%	**63.3**	**91.1**	62.8	90.4	61.3	88.2

Table 5. Effects of differing reliability using pseudo voxels on SemanticKITTI validation set, measured by the entropy.

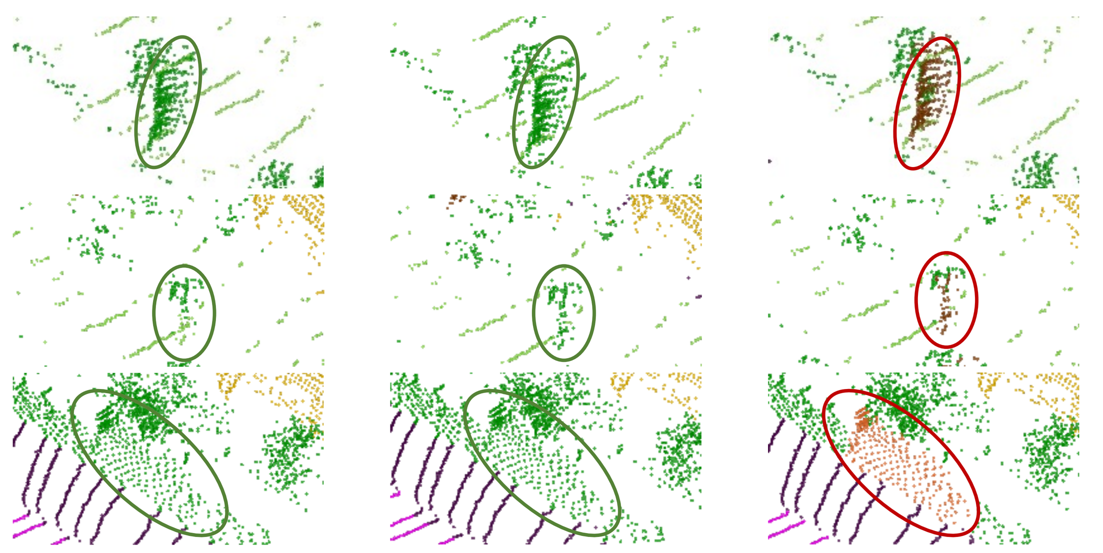


Figure A4. Magnification of regional details.

TTA	SemanticKITTI [7] 5%	10%	20%	40%	ScribbleKITTI [46] 5%	10%	20%	40%
Intensity	56.2	59.1	59.8	60.9	55.7	57.5	57.9	59.2
Reflectivity	**59.5**	**62.2**	**63.1**	**63.3**	**58.1**	**61.0**	**61.2**	**62.0**

Table 6. Reflectivity (Reflec-TTA) vs. Intensity (intensity-based TTA) on SemanticKITTI and ScribbleKITTI *validation* set (mIoU, %).