Hyundai-NGV Open R&D Seminar

Efficiencies in 3D Environment Understanding for Future Autonomous Driving

Li Li

Department of Computer Science
Durham University

li.li4@durham.ac.uk | https://www.luisli.org

Durham University

- **3rd oldest** university in England (1832)
- **World leading** university (top 100)
- **Top 20** in the world for sustainability
- **UK ranking: top 10**
- Computer Science
- **Computer Science**
    - NVIDIA CUDA Research Centre
    - Intel Parallel Computing Centre

NVIDIA CUDA RESEARCH CENTER

Intel Parallel Computing Centers

Our Research Team

Embracing **EDI**: United in **Diversity**, Committed to **Equity**, and Fostering **Inclusion** for All

Durham University

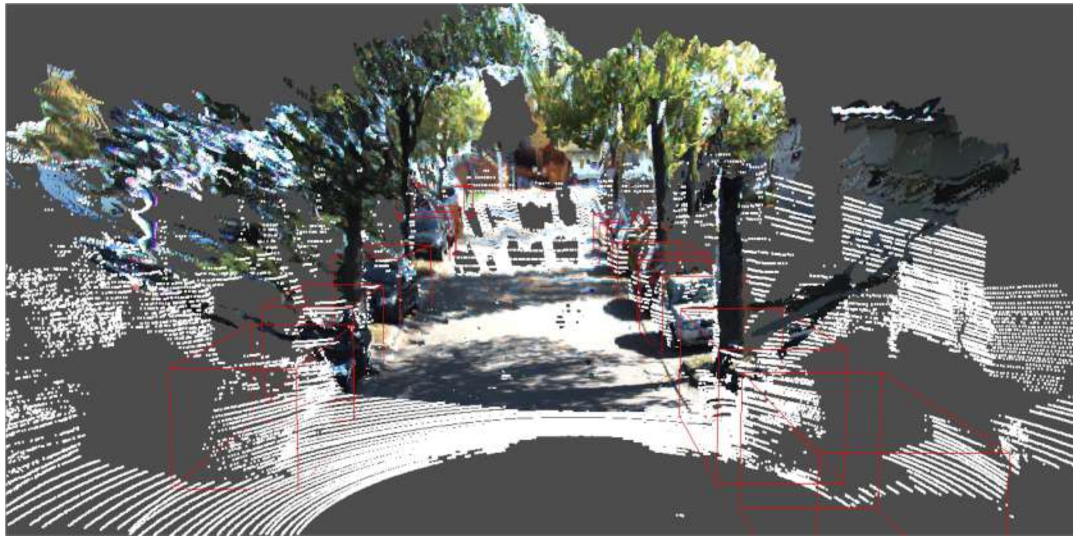
Current Team: Toby Breckon, Joshua Podmore, Jack Barker, Neelanjan Bhowmik, Yona Gaus, Brian Isaac-Medina, Seyma Yucer-tektas, Hiroshi Sasaki, Li (Luis) Li, Richard Boulderstone, Jiaxu (Judge) Liu, Wenke (Tom) E, Ghada Alosaimi, Yixin Sun, Xingyu Liu

Gone but not forgotten: Marcin Eichner, Stuart Barnes, Jiwan Han, Anna Gaszczak, Najla Megherbi**,** Ioannis Katramados, Greg Flitton, Andre Mouton, Marina Magnabosco, Olegs Mise, Alex Richardson, Oliver Hamilton, Dereck Webster, Chris Holder, Sheraz Shahid, Pedro Cavestany, Mikolaj Kundegorski, Micheal Devereux, Samet Akcay, Amir Atapour-Abarghouei, Khalid Ismail, Qian Wang, Grégoire Payen de La Garanderie, Bruna Maciel-Pearson, Nik Khadijah Nik Aznan, Philip Adey, Naif Alshammari, Will Prew, Hiroshi Sasaki, Aishah Alsehaim, Matt Poyser




…. Openday tours of key research topics

Virtual Lab Tour: *Applied Computer Vision*

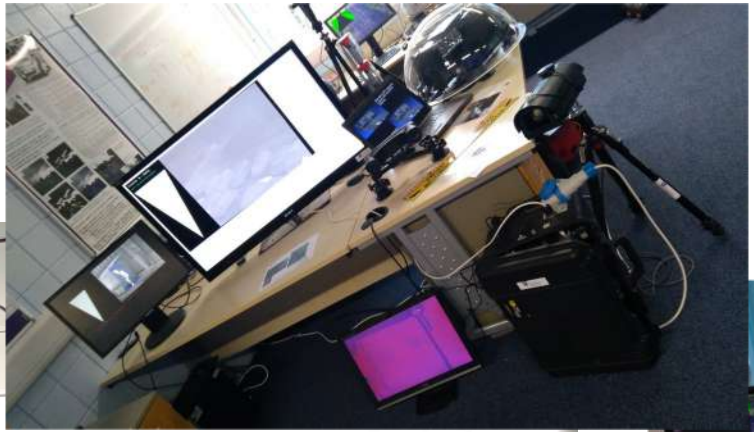


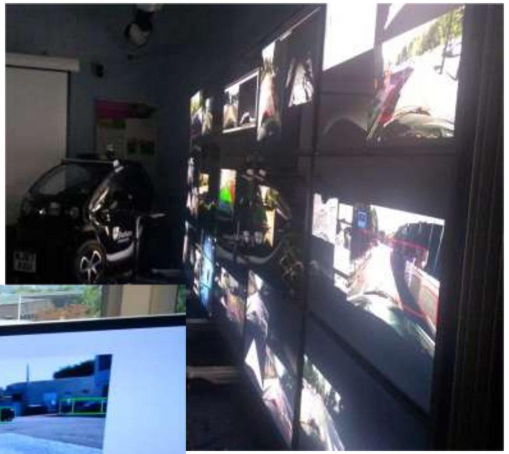
Department of

Computer Science

Prof. Toby Breckon

- Our research lab ...



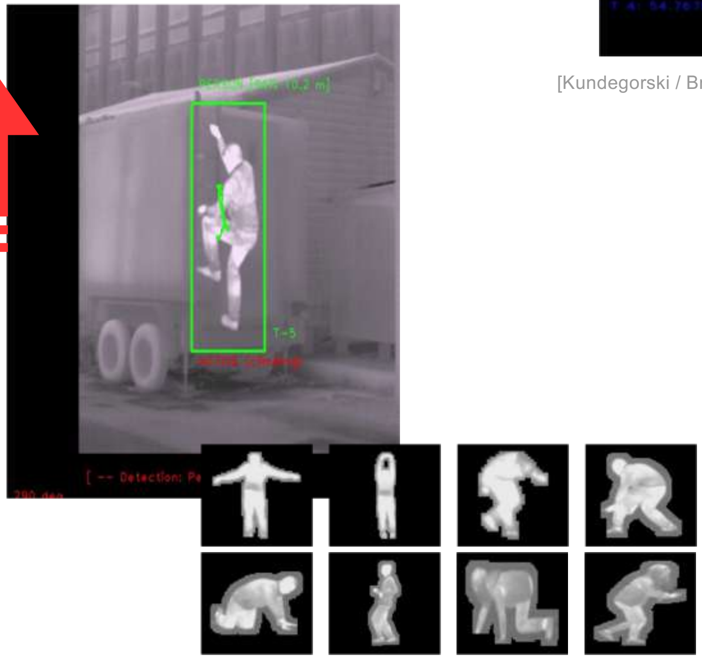


algorithms

for processing visual information

Automated Visual Surveillance

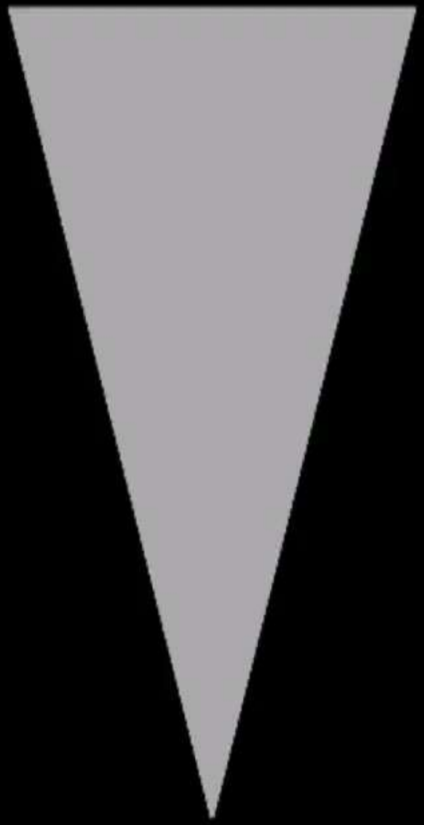
[Kundegorski / Breckon et al. '14]



[Kundegorski / Breckon et al. '15]

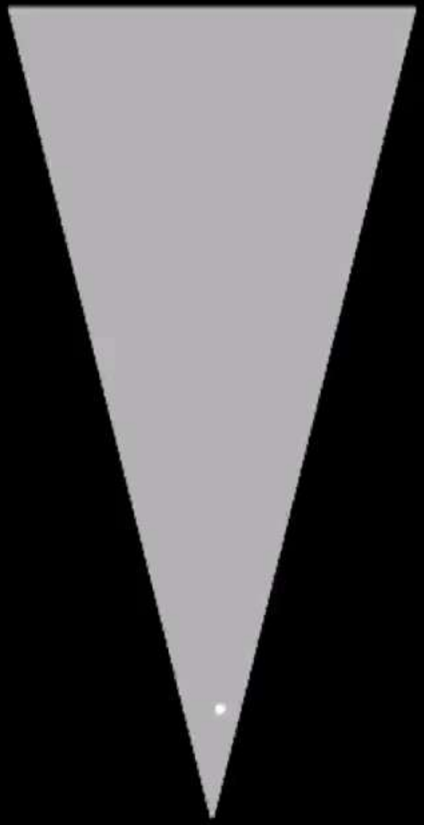
[Kundegorski / Breckon et al. '16]

Working with:



Cam: Lat:54.767N, Lon:-1.57006E, 290deg

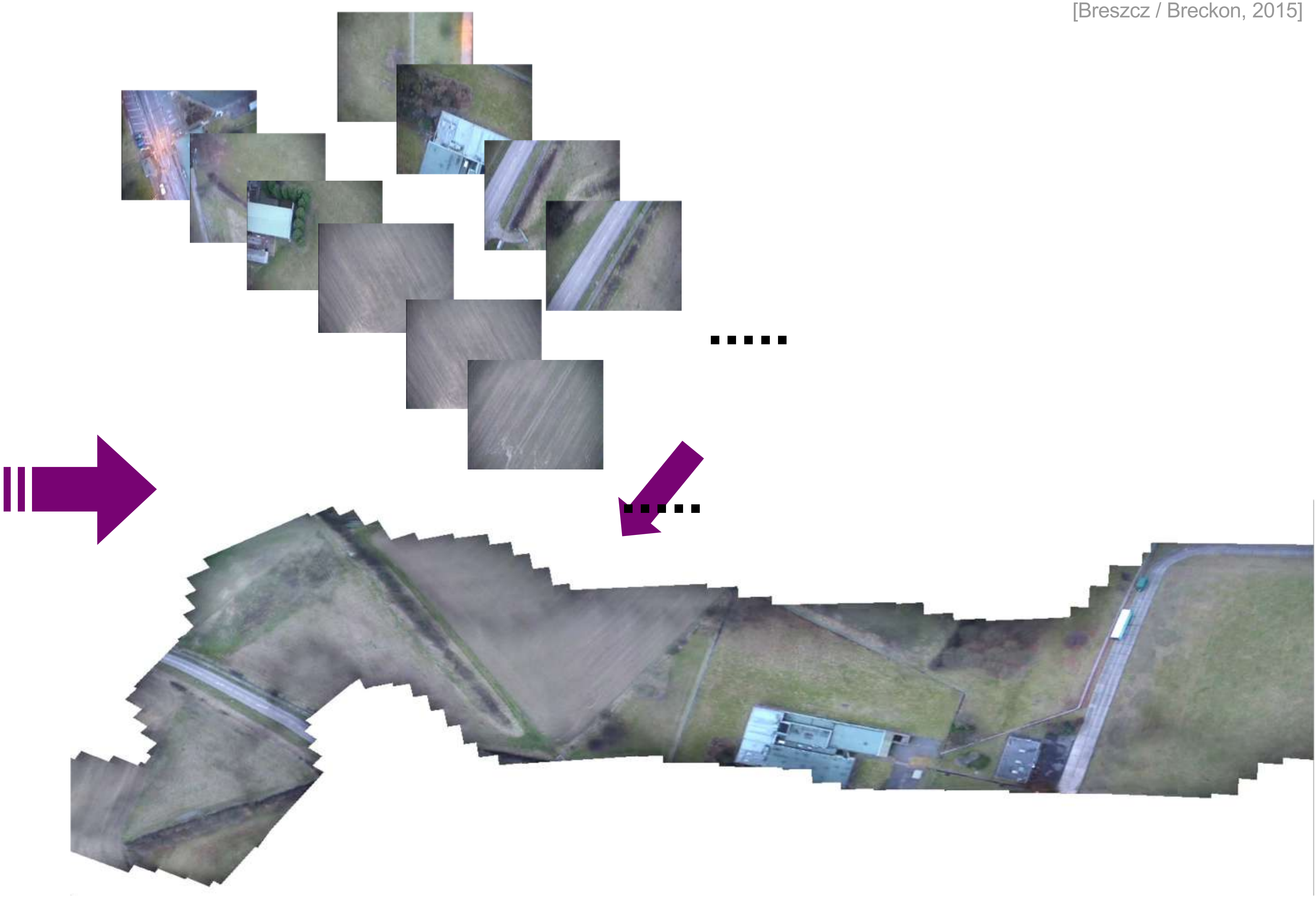
Working with: [dstl]



Cam: Lat:54.7668N, Lon:-1.57E, 100 deg

Automated Wide Area Search

[Breszcz / Breckon, 2015]



Working with:

BLUEBEAR

[Maciel-Pearson, Akcay, Atapour-Abarghouei,Holder, Breckon, 2019]

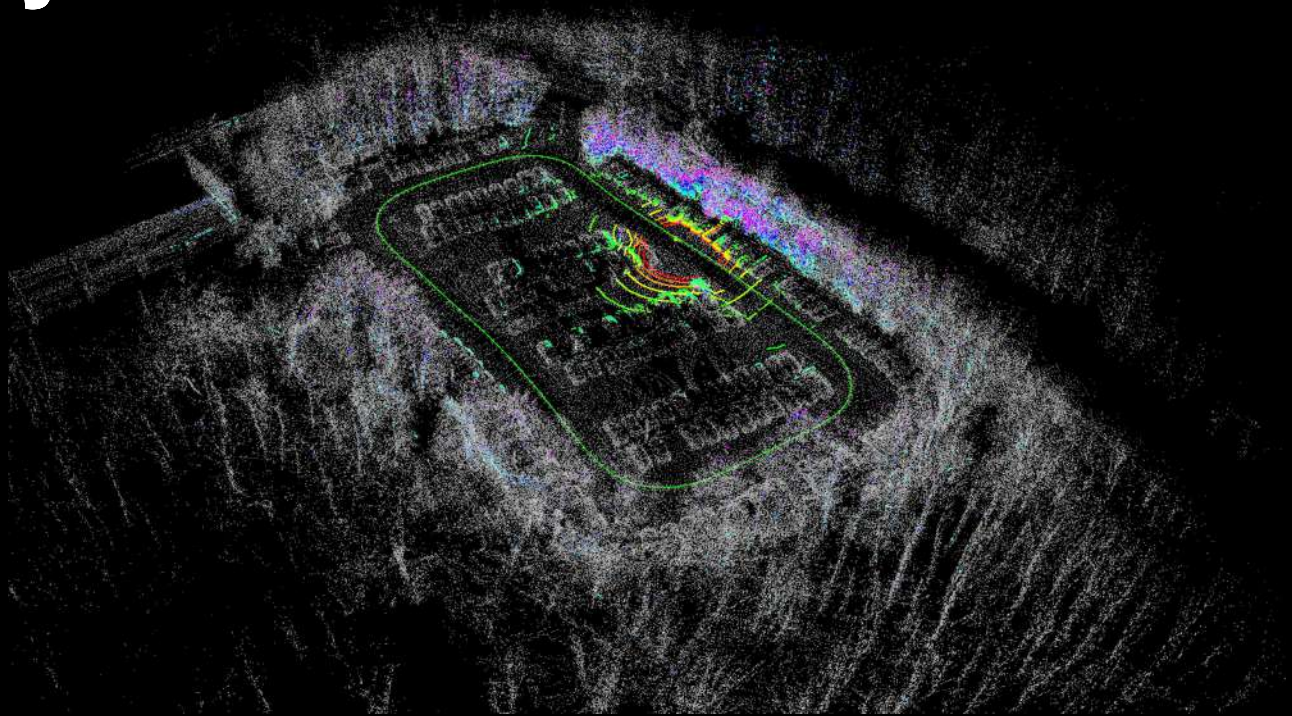
Working with:

Future Vehicle Autonomy

[Hamilton, Breckon, 2013]

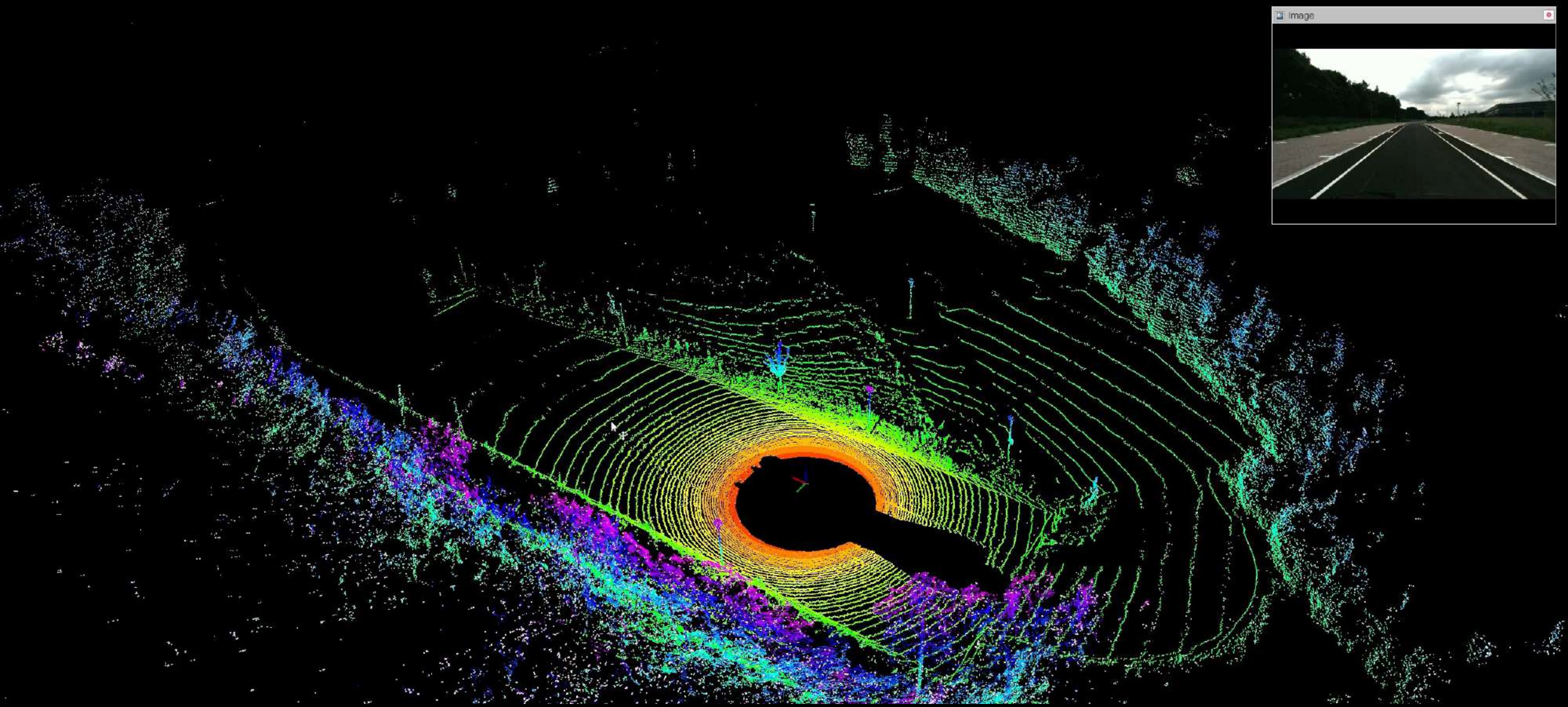
Working with: conekt powered by TRW

Future Vehicle Autonomy



Working with: Machines WithVision JAGUAR LAND ROVER

Future Vehicle Autonomy

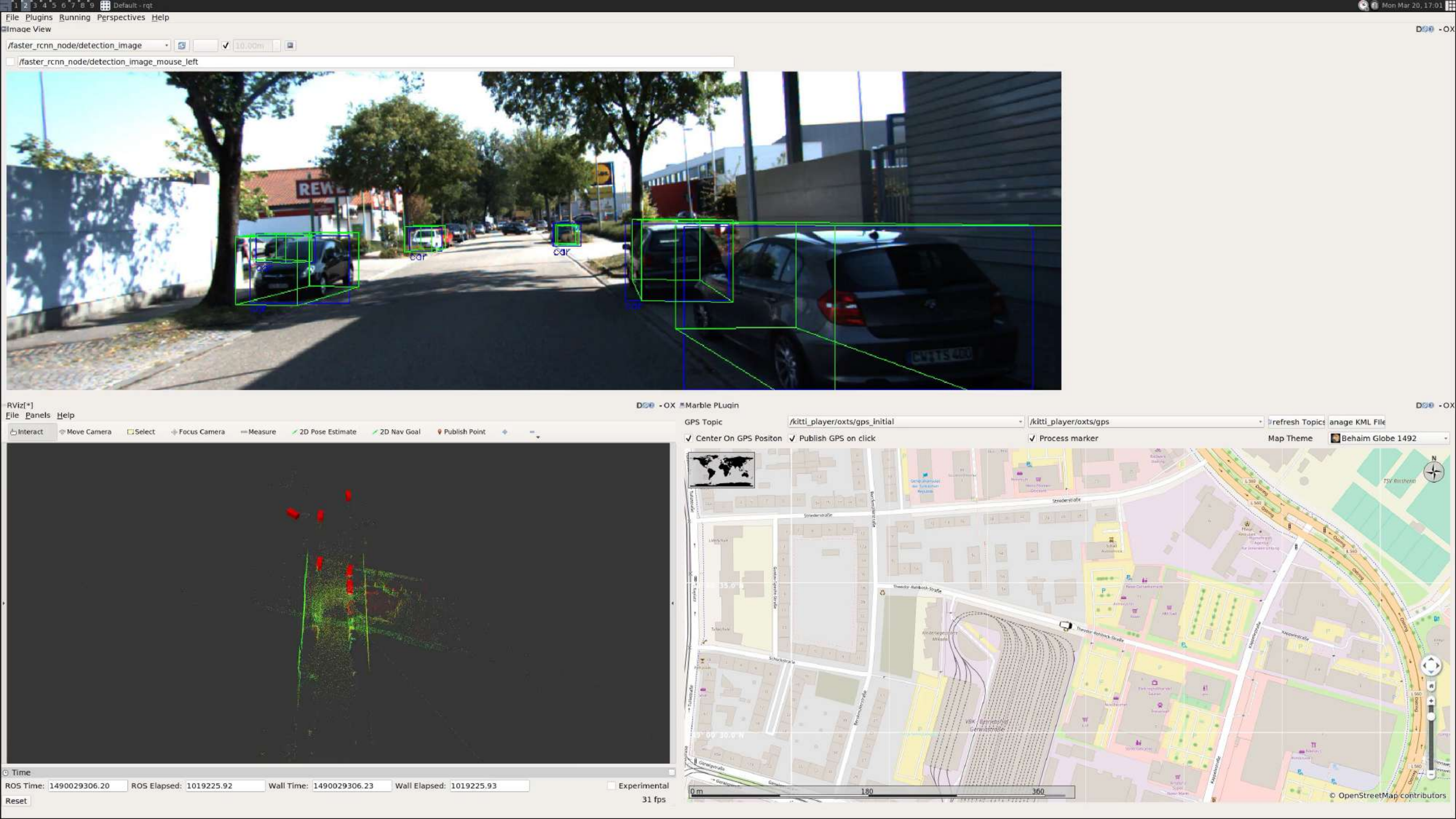


Efficiencies in 3D Environment Understanding for Future Autonomous Driving

[Hamilton, Breckon, 2016]

Working with: coneKt powered by TRW

Generalized Dynamic Object Removal for Dense Stereo Vision Based Scene Mapping using Synthesised Optical Flow


Default - rqt
Mon Mar 20, 17:01
File Plugins Running Perspectives Help
Image View
/faster_rcnn_node/detection_image
/faster_rcnn_node/detection_image_mouse_left
REWE
car
car
RViz[*]
File Panels Help
Interact
Move Camera
Select
Focus Camera
Measure
2D Pose Estimate
2D Nav Goal
Publish Point
Marble PLugin
GPS Topic
/kitti_player/oxts/gps_initial
/kitti_player/oxts/gps
refresh Topics
anage KML File
Center On GPS Positon
Publish GPS on click
Process marker
Map Theme
Behaim Globe 1492
Time
ROS Time: 1490029306.20
ROS Elapsed: 1019225.92
Wall Time: 1490029306.23
Wall Elapsed: 1019225.93
Experimental
Reset
31 fps
0 m
180
360
© OpenStreetMap contributors

Future Aviation Security

- via connected screening capability
  - Durham software intercepts image made available on network

Working with:

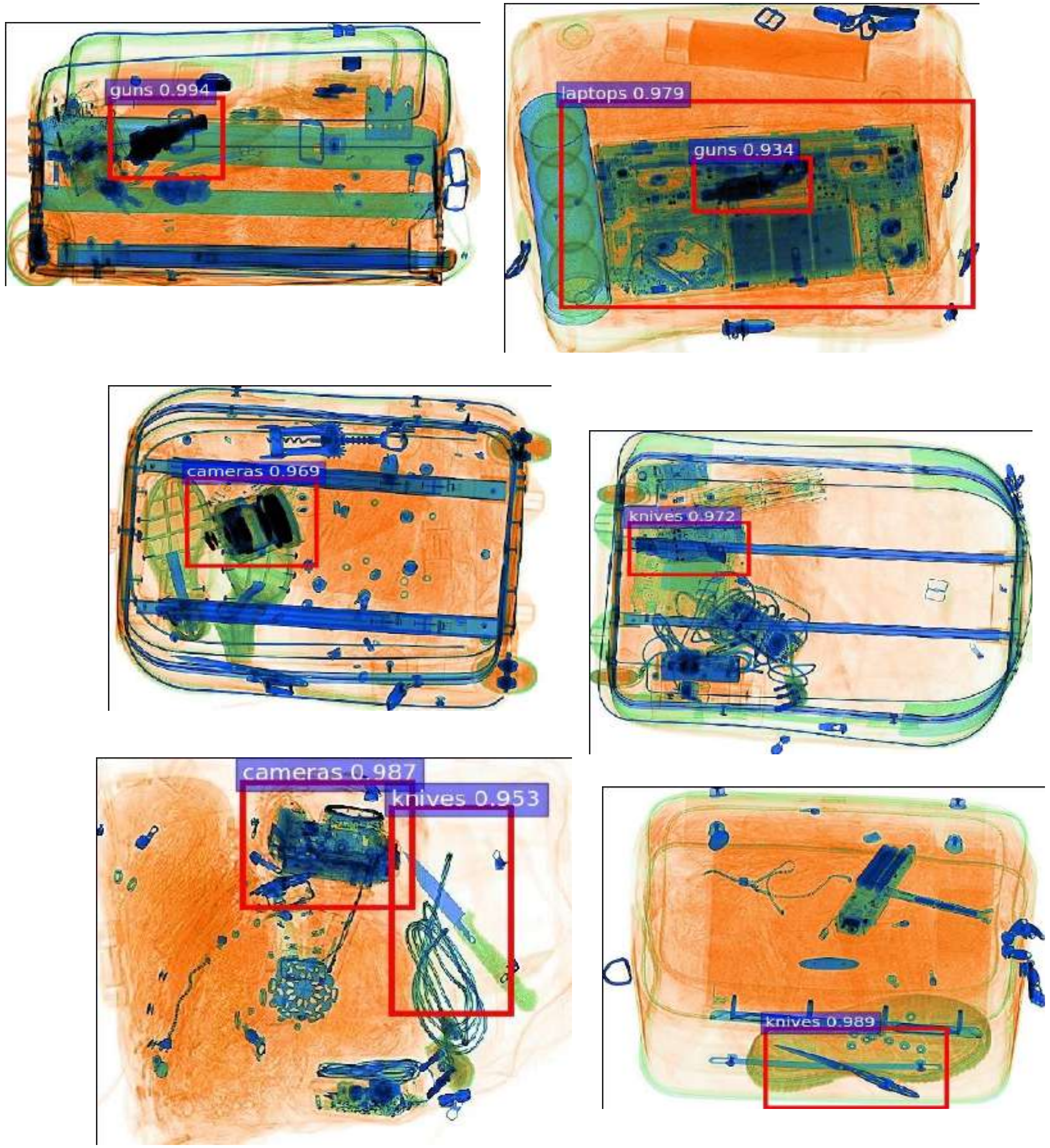
DETECTNOLOGY
SPECIALISED SECURITY SYSTEMS

GILARDONI

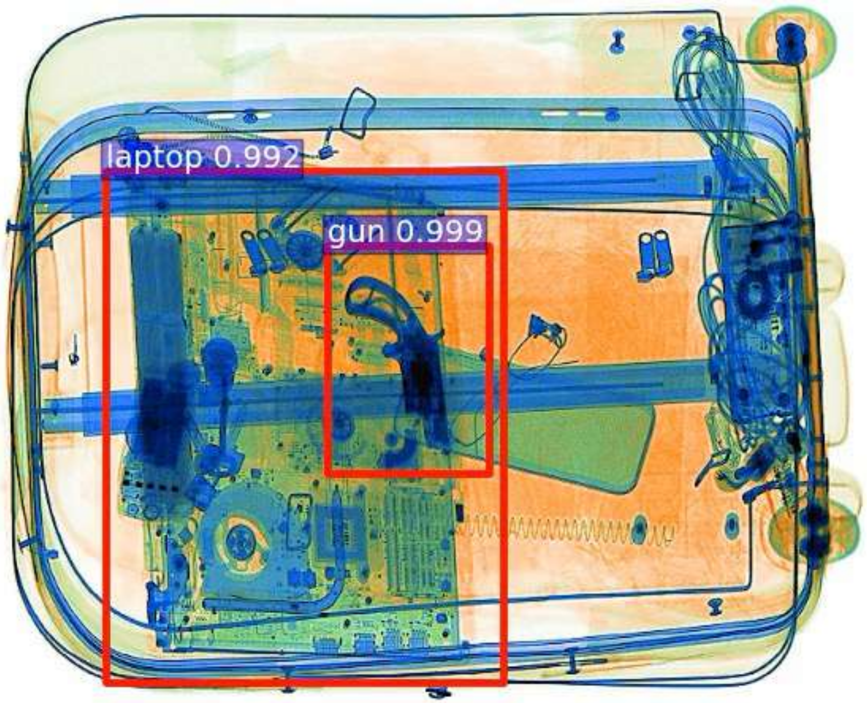
[video] - small

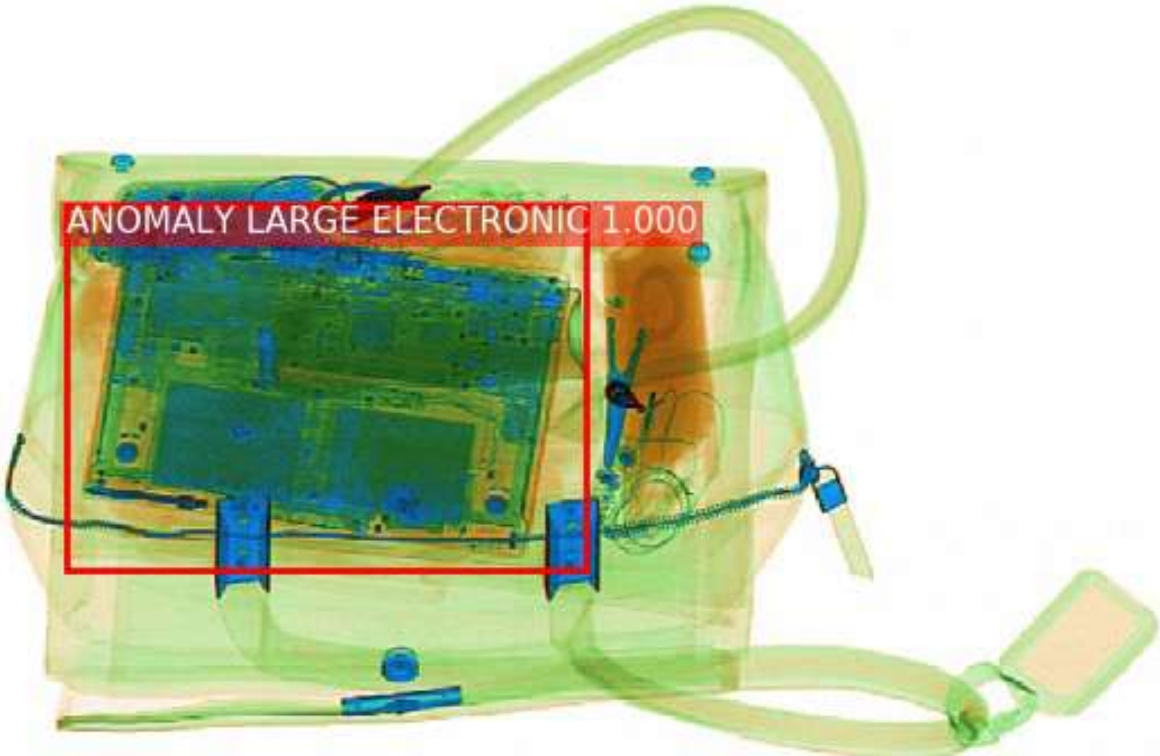
[video] - large

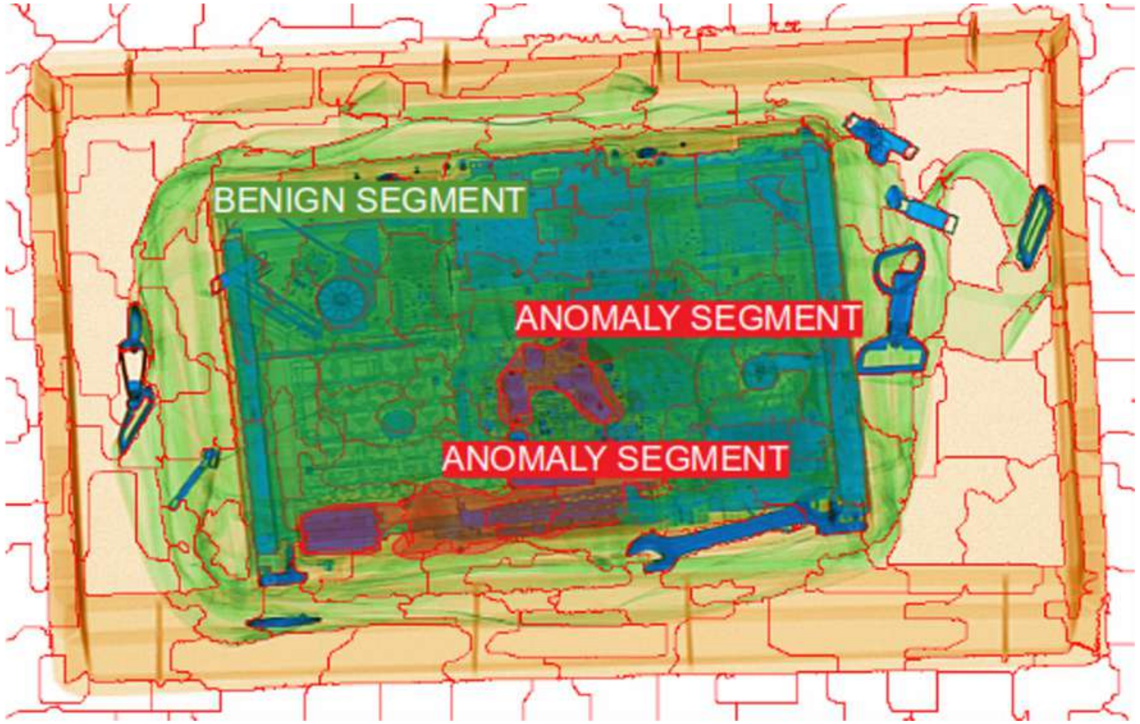
[Akcay, Kundegorski, Willcocks, Breckon – 2018]



Working with: Home Office | Department for Transport

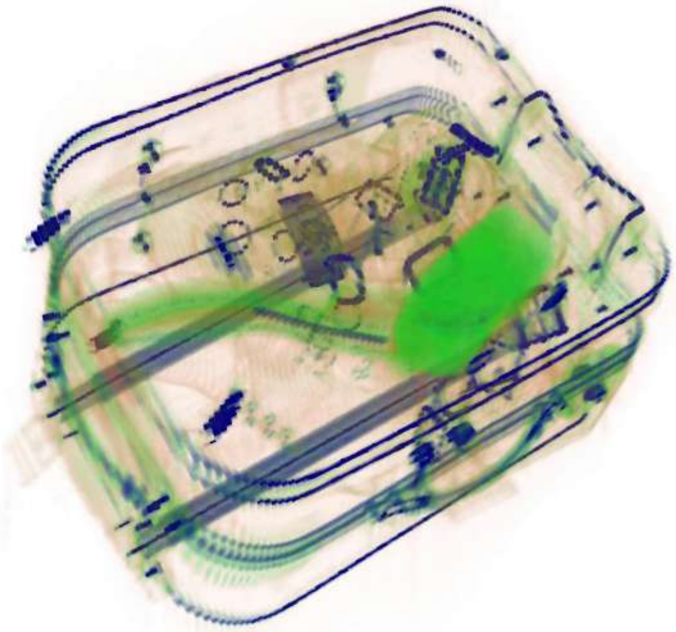

laptop 0.992
gun 0.999





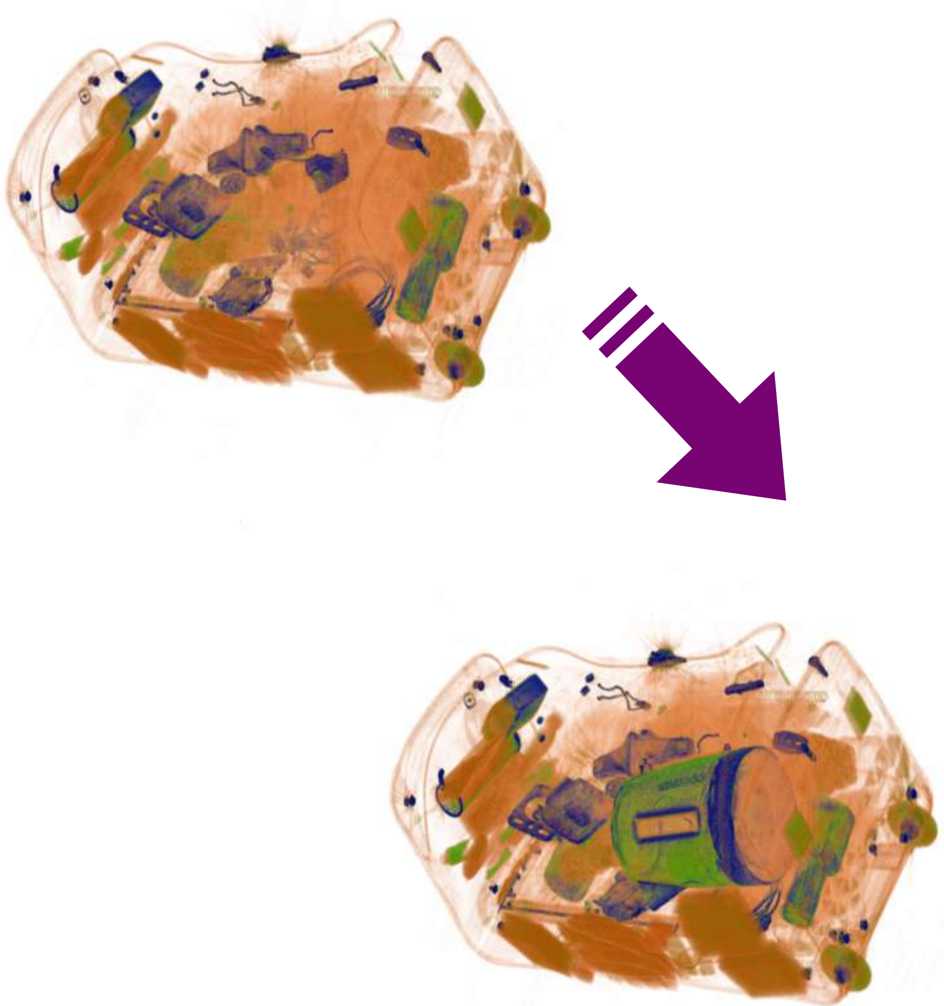
Working with: Home Office Department for Transport

[Wang, Bhowmik, Breckon – 2020]

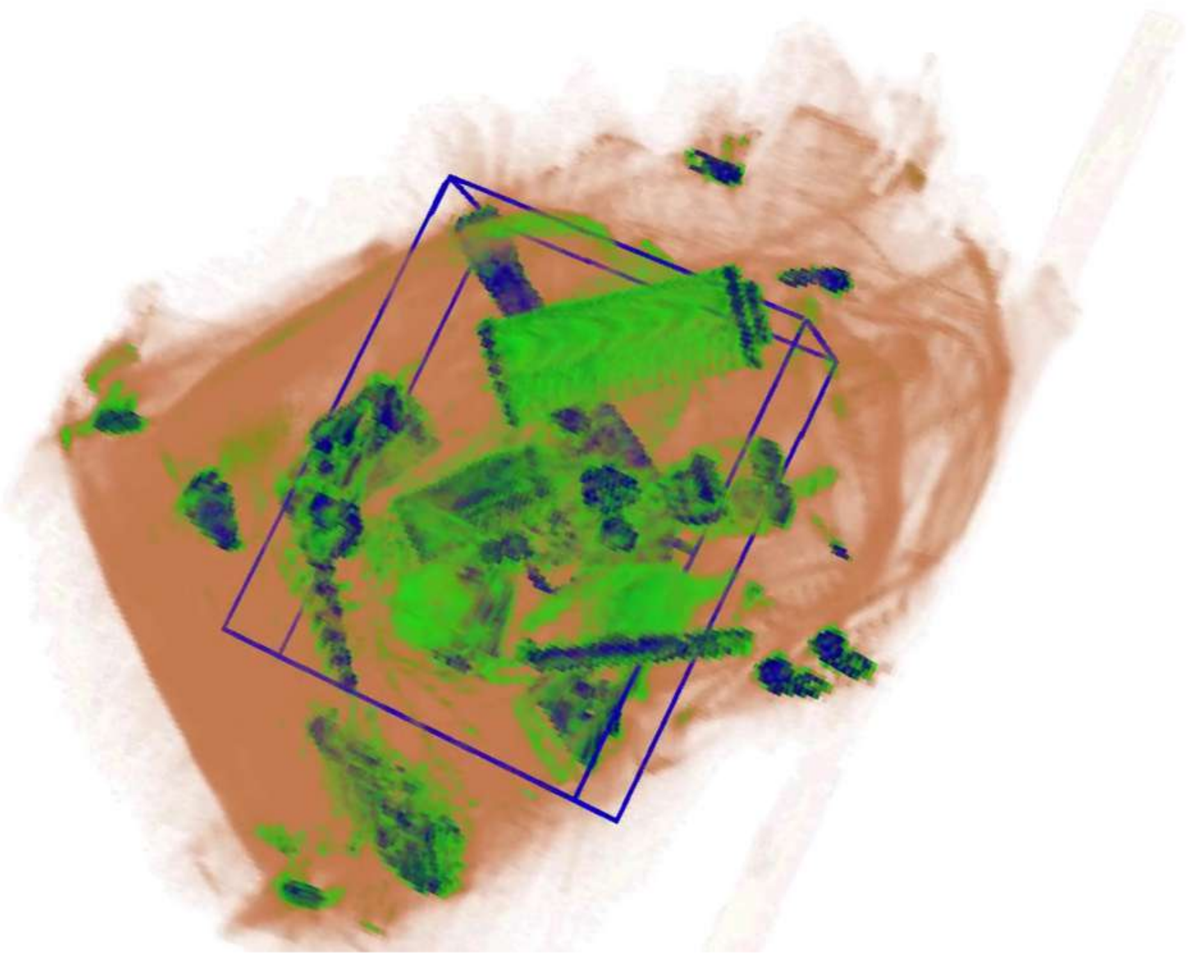


Working with: Home Office | Department for Transport

[Wang, Megherbi, Breckon – 2020]



[Wang, Bhowmik, Breckon – 2020]



Working with: Home Office | Department for Transport

algorithms

for processing visual information

Durham University

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

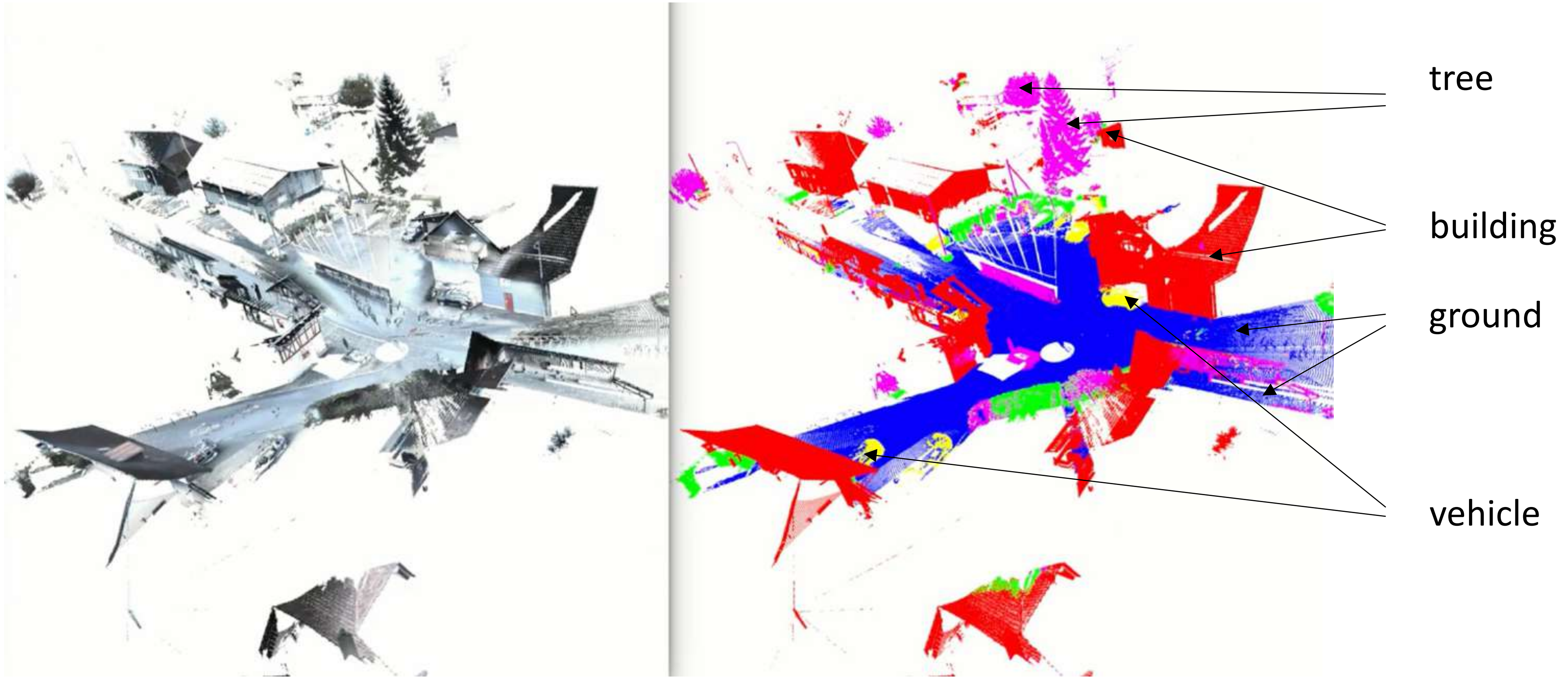
Less is More
Reducing Task and Model Complexity for 3D Point Cloud Semantic Segmentation

Li Li[1], Hubert P. H. Shum[1], Toby P. Breckon[1,2]

Department of {Computer Science[1] | Engineering[2]}
Durham University

li.li4@durham.ac.uk

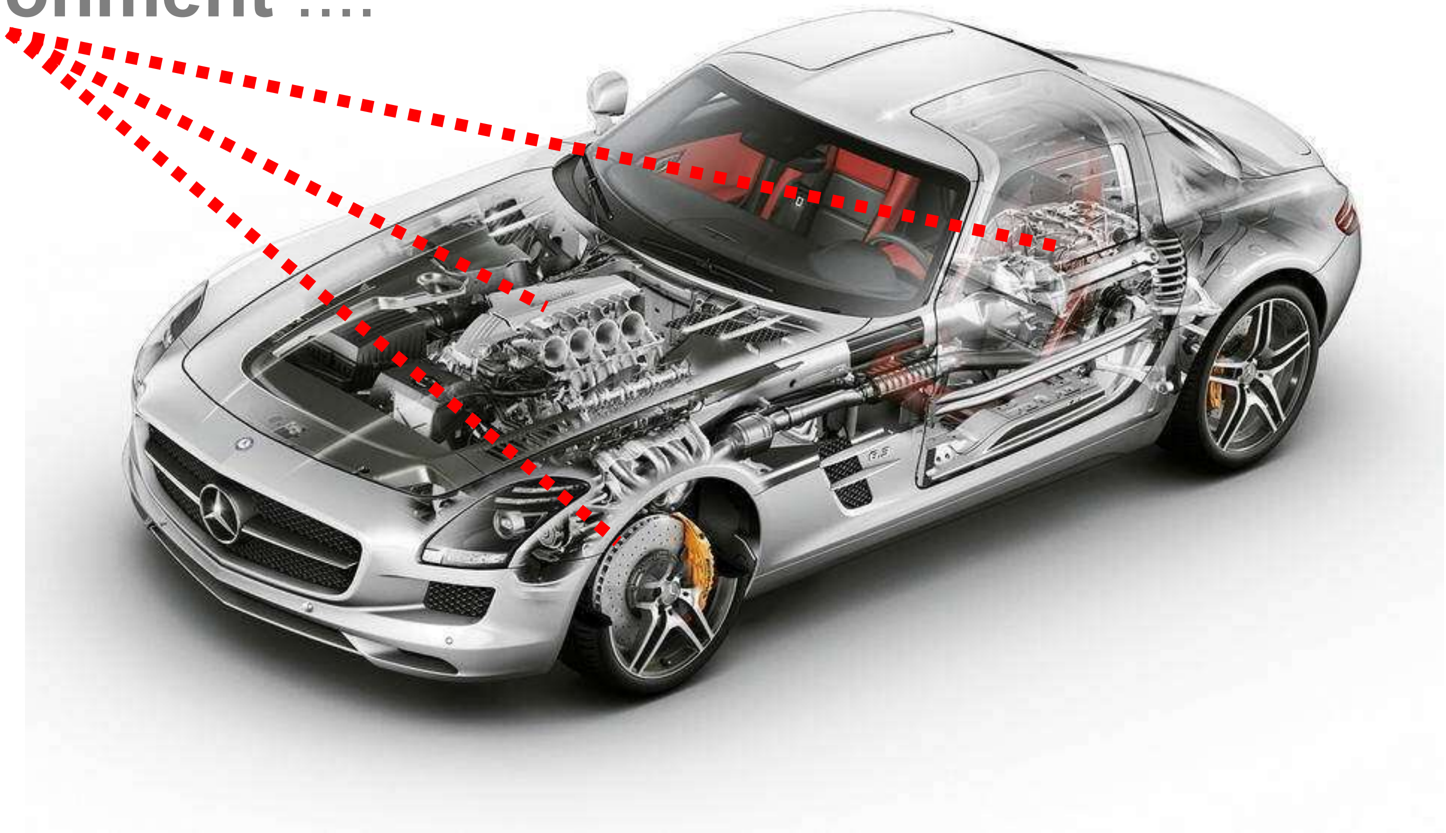
Point Cloud Semantic Segmentation



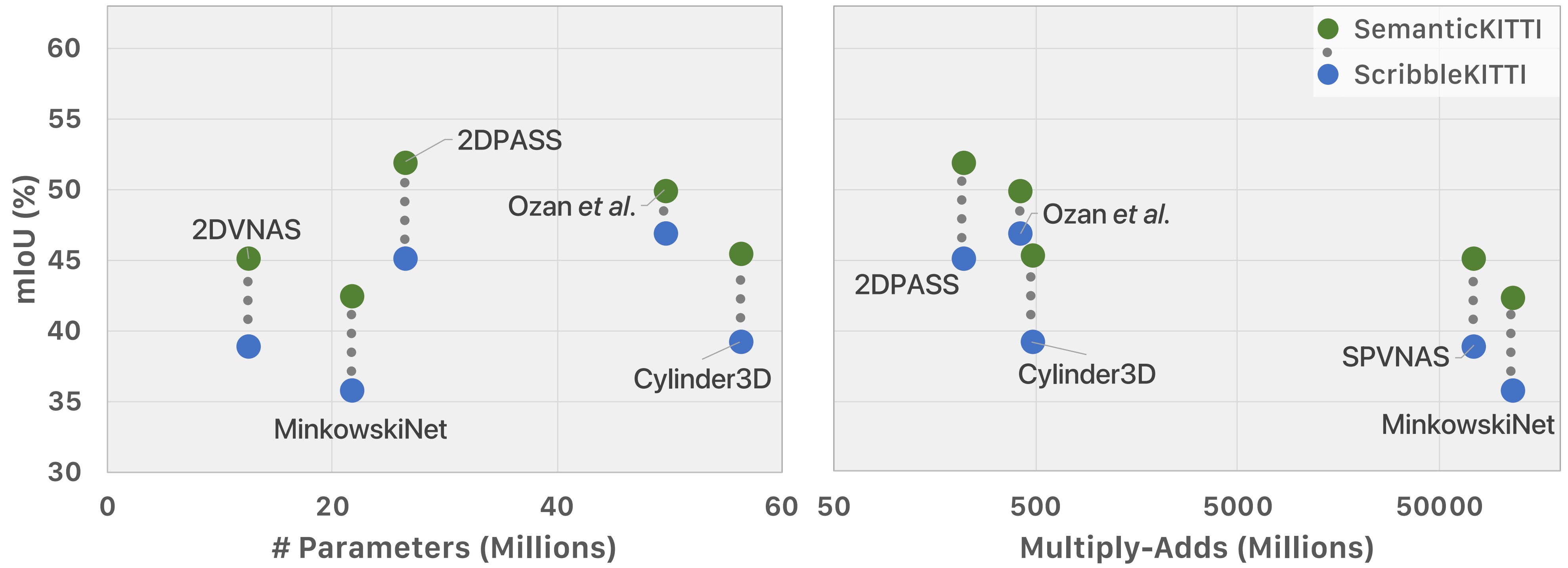
Input: point cloud data **Output**: semantic segmentation prediction

Future Vehicle Autonomy

Modern vehicles contain a **range of dynamics tunable to the road environment** ....

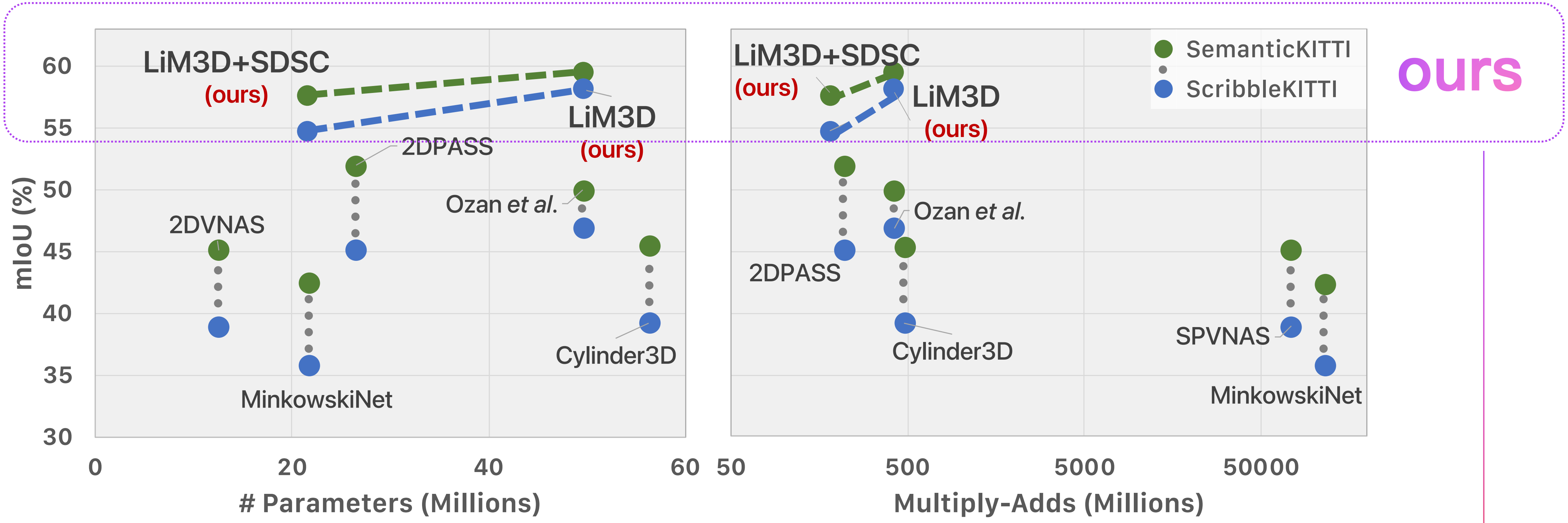


LESS IS MORE



Previous methods

LESS IS MORE

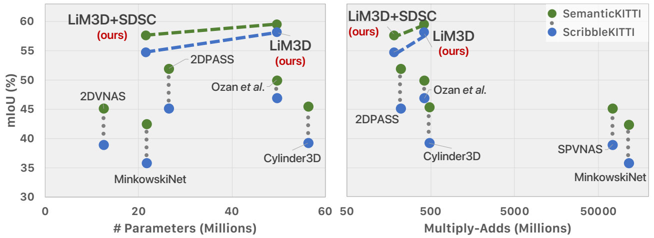


2.3x model size reduction

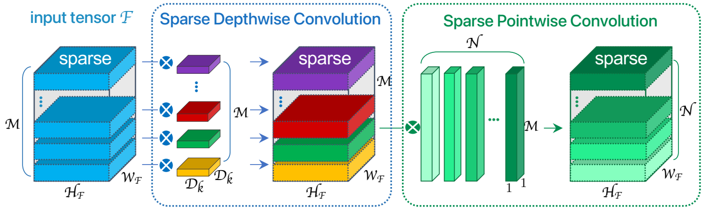
641x fewer multiply–adds

Contributions

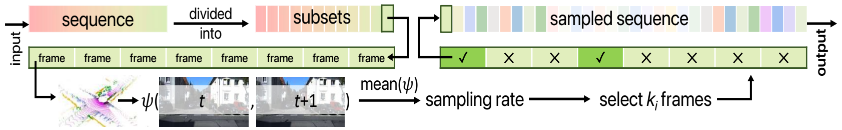
1 semantic segmentation: *less* **parameters** and **(*more*) superior accuracy**.



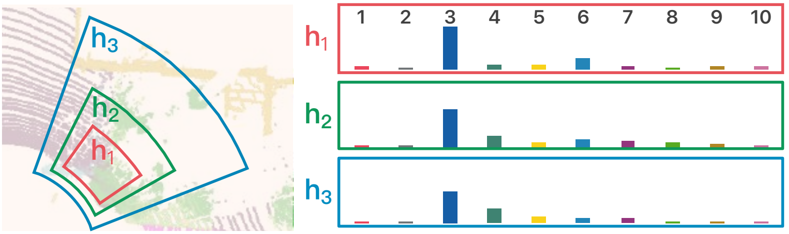
2 **Sparse Depthwise Separable Convolution (SDSC)**: to reduce trainable network without loss.



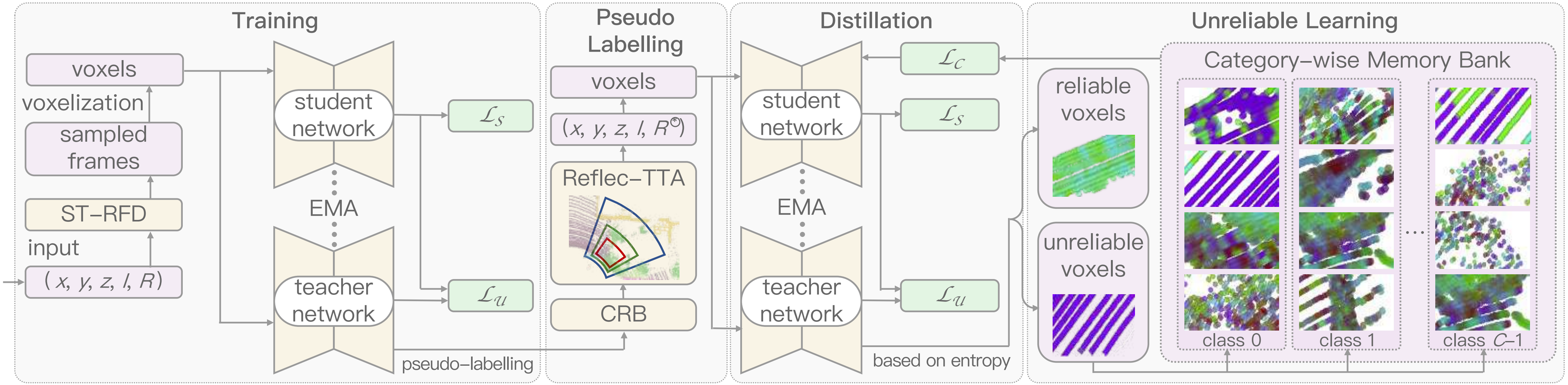
3 **Spatio-Temporal Redundant Frame Downsampling (ST-RFD)**: to remove temporal redundancy.

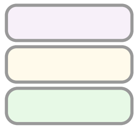


4 Soft pseudo-labeling method informed by **LiDAR reflectivity**: to use limited data annotation effectively.



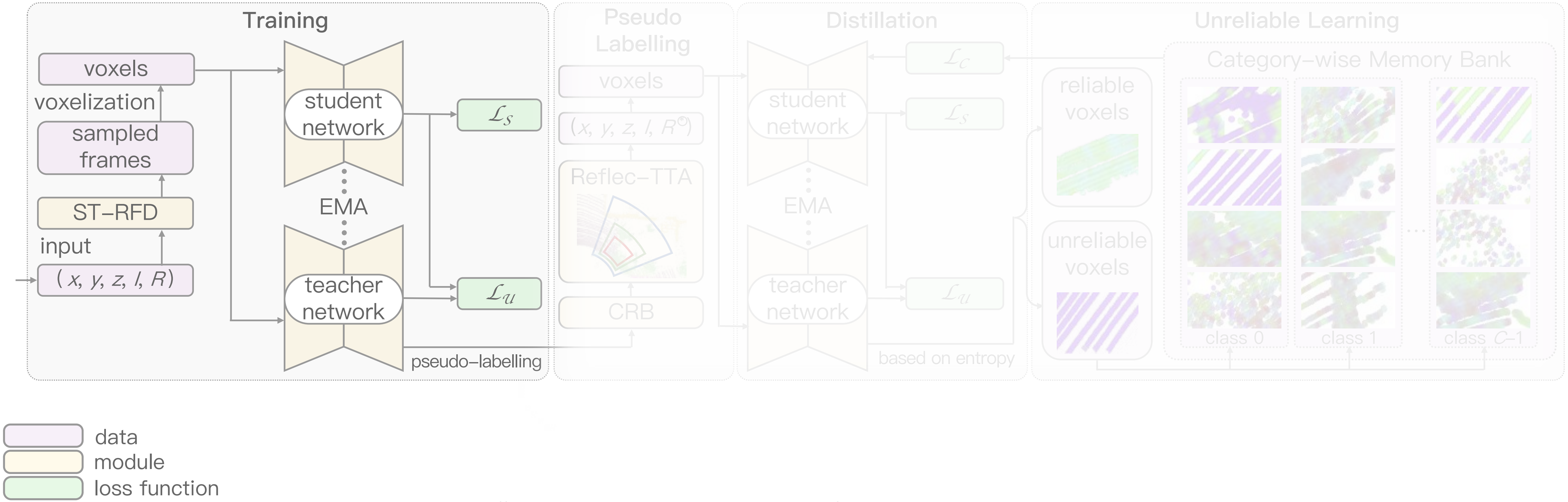
Our Proposed Architecture





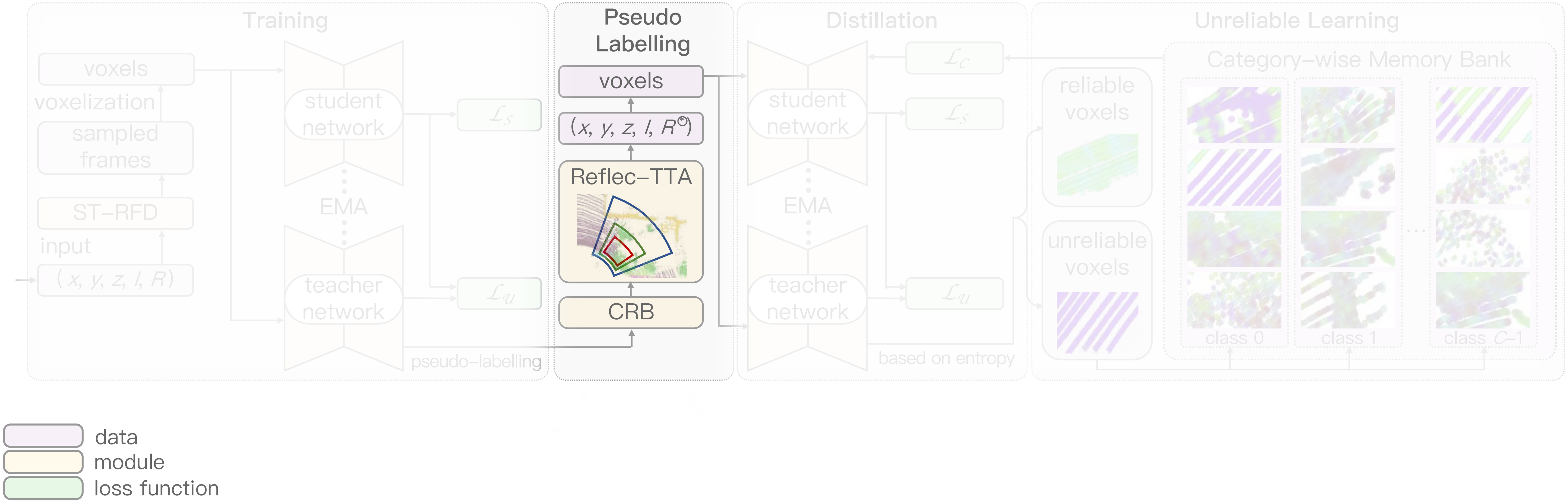
Our Proposed Architecture

training > to utilize **reflectivity–prior descriptors** and adapt the **Mean Teacher** framework to generate high–quality pseudo–labels



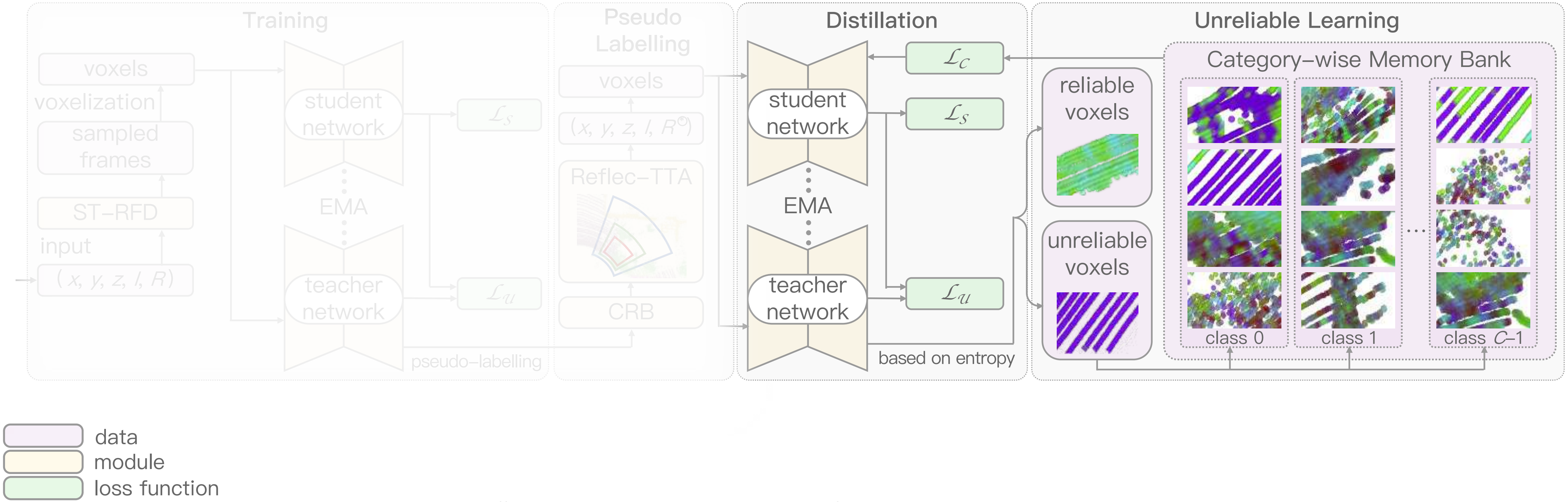
Our Proposed Architecture

pseudo labelling > to fix the trained teacher model prediction in a **CRB** manner, expanding dataset with **Reflec–TTA** during test time



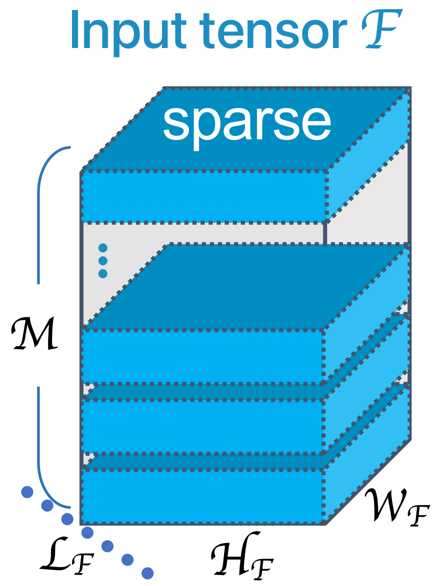
Our Proposed Architecture

distillation & unreliable learning > to train on the generated pseudo-labels, and **utilize unreliable pseudo-labels** in a memory bank for improved discrimination



Sparse Depthwise Separable Convolution

to reduce trainable network without loss



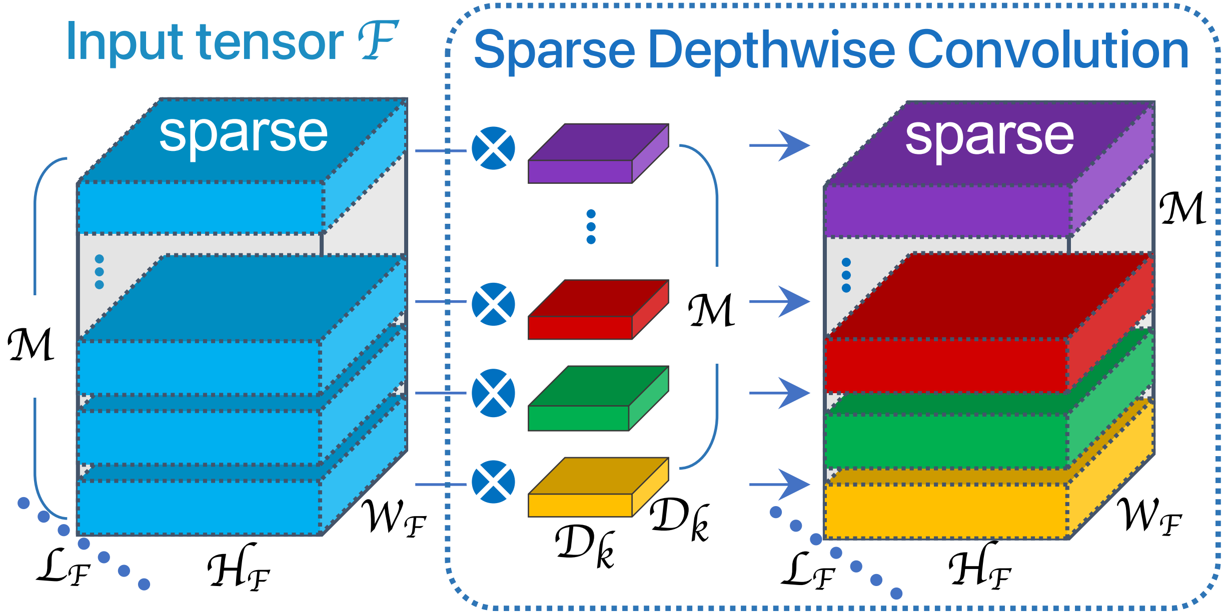
1 taking 3D voxels as input

submanifold sparse convolution

pointwise convolution

Sparse Depthwise Separable Convolution

to reduce trainable network without loss

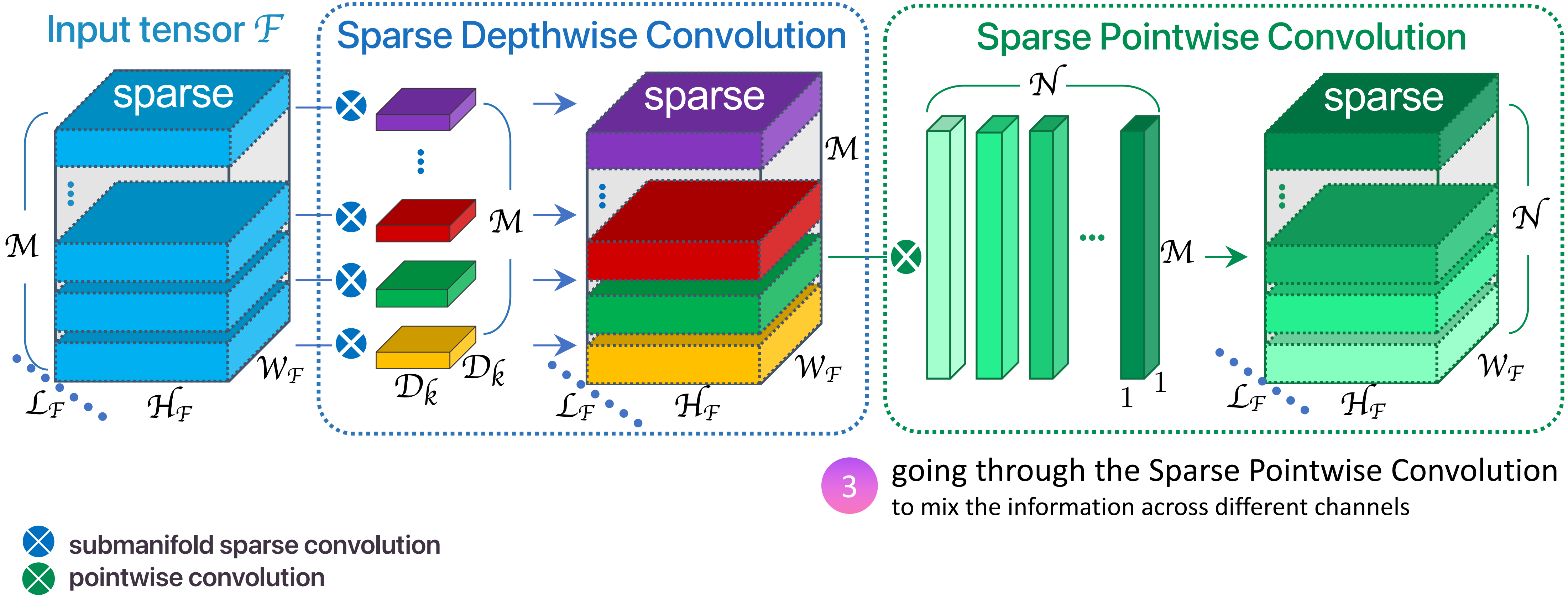


2 going through the Sparse Depthwise Convolution
to perform convolution with the trainable parameter reduction

submanifold sparse convolution
pointwise convolution

Sparse Depthwise Separable Convolution

to reduce trainable network without loss

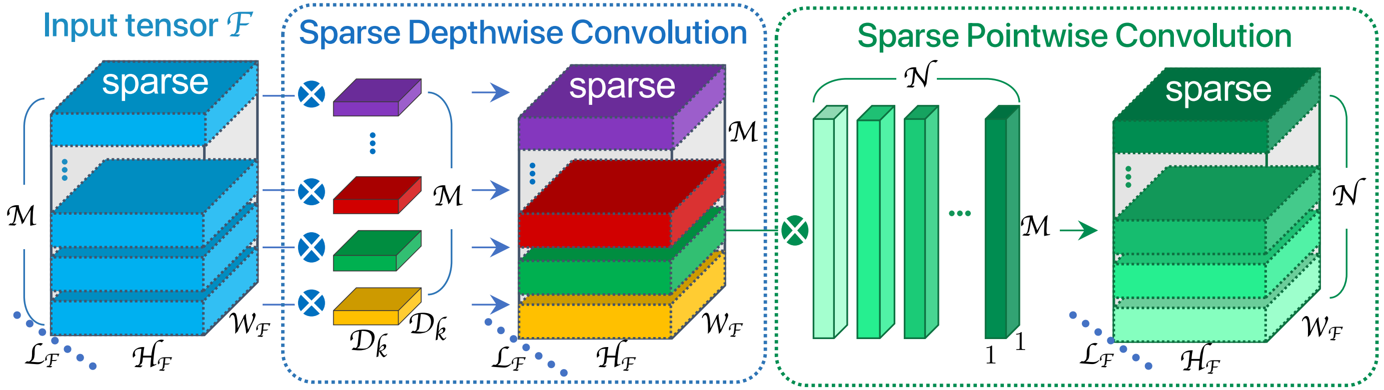


3 going through the Sparse Pointwise Convolution

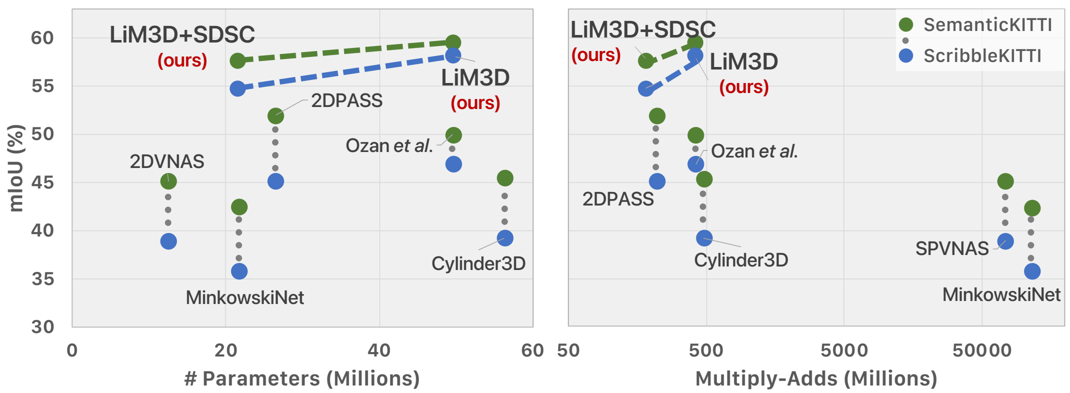
to mix the information across different channels

Sparse Depthwise Separable Convolution

to reduce trainable network without loss



submanifold sparse convolution
pointwise convolution



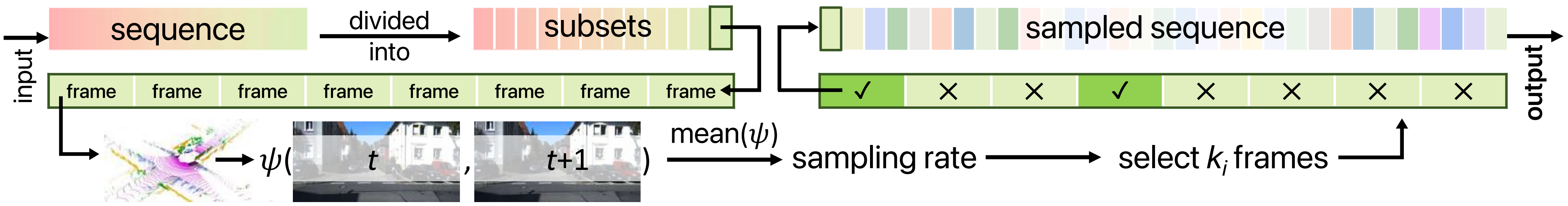
with our **Sparse Depthwise Separable Convolution** we can achieve:

2.3x model size reduction

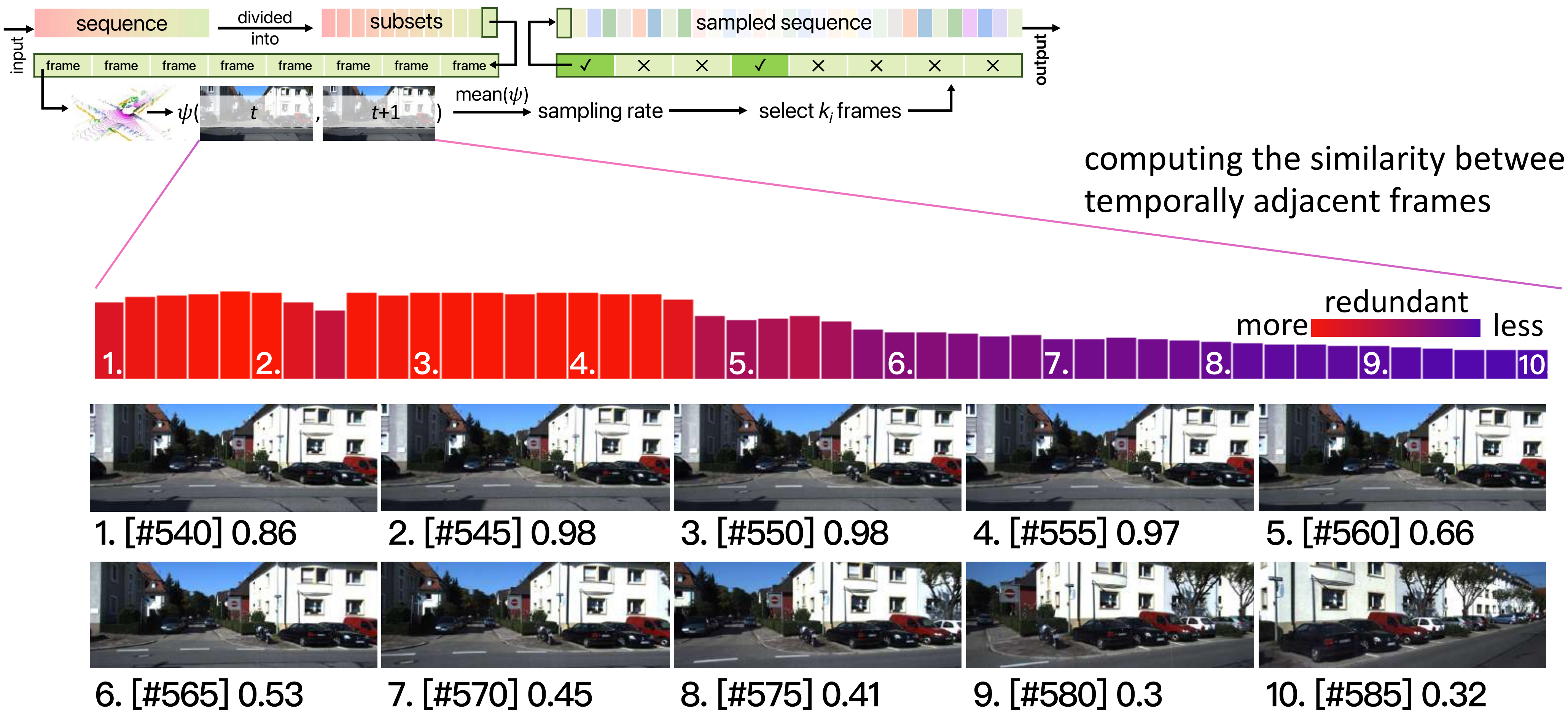
641x fewer multiply–adds

Spatio-Temporal Redundant Frame Downsampling (ST-RFD)

Using ST-RFD to extract a maximally diverse data subset for training by **removing temporal redundancy** and hence future **annotation requirements**

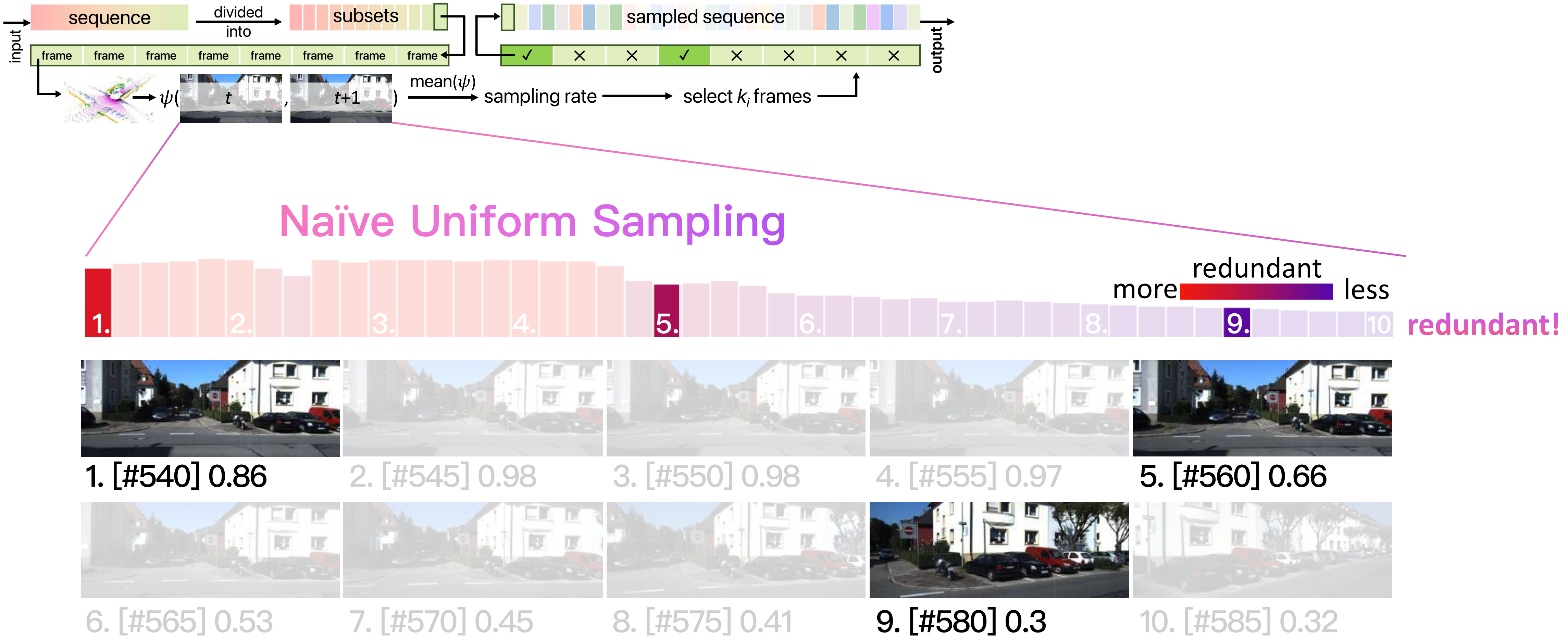


Spatio-Temporal Redundant Frame Downsampling (ST-RFD)



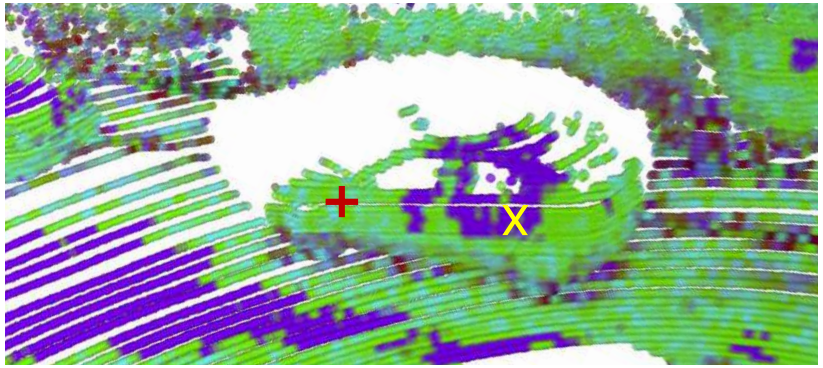
computing the similarity between temporally adjacent frames

Spatio-Temporal Redundant Frame Downsampling (ST-RFD)



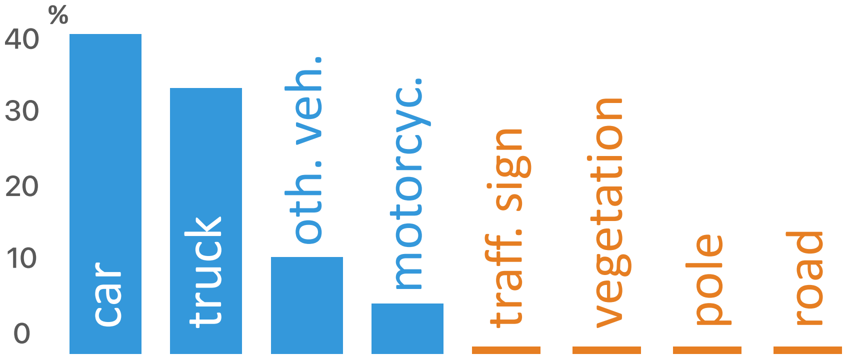
Spatio-Temporal Redundant Frame Downsampling (ST-RFD)

Using Unreliable Pseudo-labels
to Make Full Use of All Available Labels

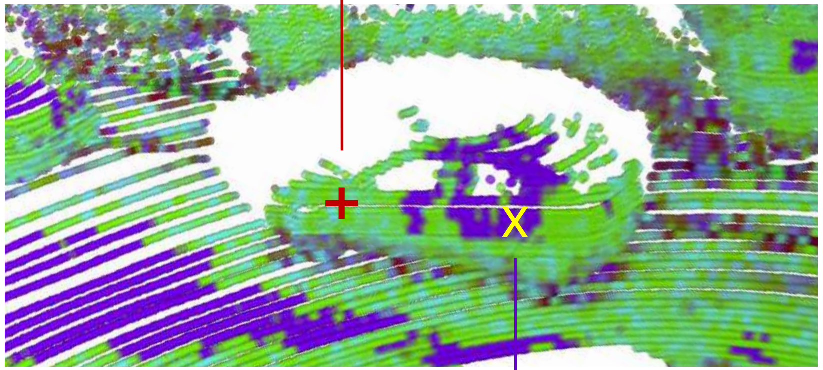


$$\mathcal{L}_C = -\frac{1}{C}\sum_{c=0}^{C-1} \mathop{\mathbb{E}}_{\mathbf{E}_c}\left[\log \frac{f(\mathbf{e}_c, \mathbf{e}_c^+, \tau)}{\sum_{\mathbf{e}_{c,j}^- \in \mathbf{E}_c^-} f(\mathbf{e}_c, \mathbf{e}_{c,j}^-, \tau)}\right]$$

$$= -\frac{1}{C}\sum_{c=0}^{C-1} \mathop{\mathbb{E}}_{\mathbf{E}_c}\left[\log \frac{\exp(\langle \mathbf{e}_c, \mathbf{e}_c^+\rangle/\tau)}{\exp\left(\langle \mathbf{e}_c, \mathbf{e}_c^+\rangle/\tau\right) + \sum_{j=1}^{N-1} \exp\left(\langle \mathbf{e}_c, \mathbf{e}_{c,j}^-\rangle/\tau\right)}\right]$$



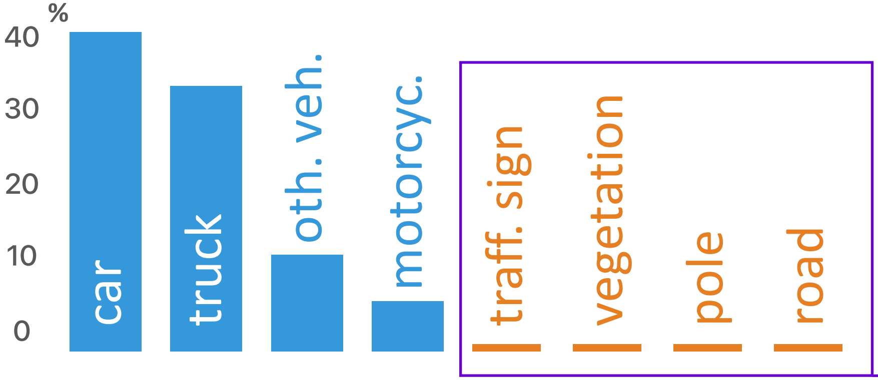
Using Unreliable Pseudo–labels
to Make Full Use of All Available Labels



positive sample

$$\mathcal{L}_C = -\frac{1}{C}\sum_{c=0}^{C-1} \mathbb{E}_{\mathbf{E}_c}\left[\log \frac{f(\mathbf{e}_c, \boxed{\mathbf{e}_c^+}, \tau)}{\sum_{\mathbf{e}_{c,j}^- \in \mathbf{E}_c^-} f(\mathbf{e}_c, \mathbf{e}_{c,j}^-, \tau)}\right]$$

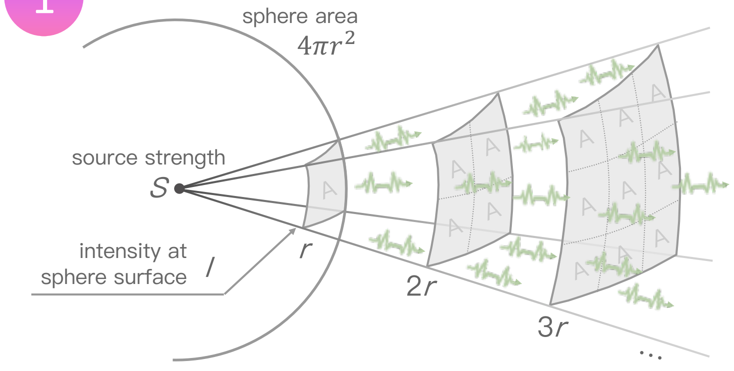
$$= -\frac{1}{C}\sum_{c=0}^{C-1} \mathbb{E}_{\mathbf{E}_c}\left[\log \frac{\exp(\langle \mathbf{e}_c, \mathbf{e}_c^+ \rangle / \tau)}{\exp\left(\langle \mathbf{e}_c, \mathbf{e}_c^+ \rangle / \tau\right) + \sum_{j=1}^{N-1} \exp\left(\langle \mathbf{e}_c, \boxed{\mathbf{e}_{c,j}^-} \rangle / \tau\right)}\right]$$



negatives sample

Using reflectivity–based Test Time Augmentation
to enhance performance of false or non–existent pseudo–labels

1



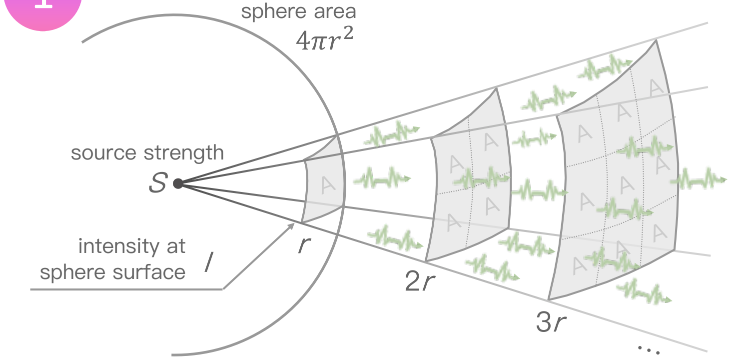
reflectivity

$$R = Ir^2 = \frac{S}{4\pi r^2} \cdot r^2 \propto S$$

Reflectivity is
a **distance–normalized intensity** feature

Using reflectivity-based Test Time Augmentation
to enhance performance of false or non-existent pseudo-labels

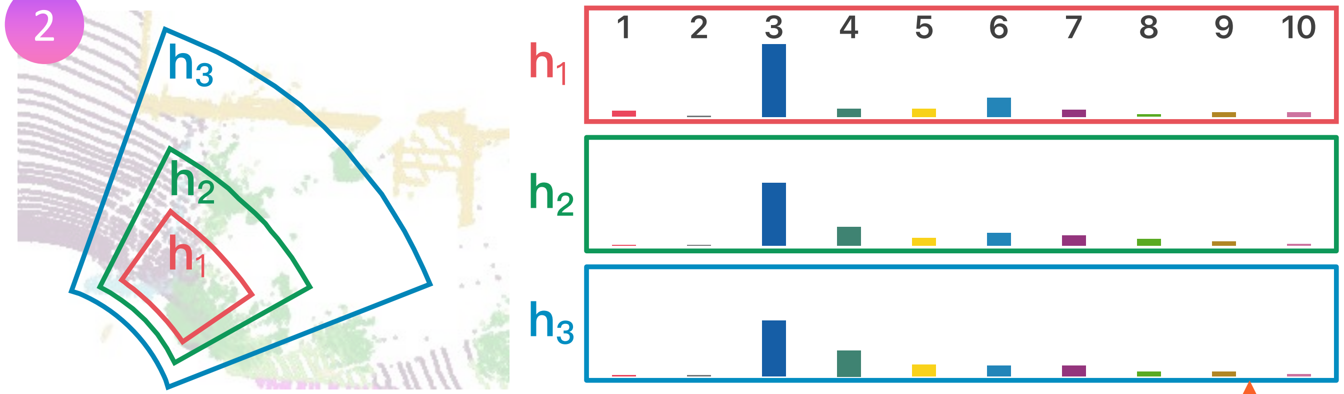
1



reflectivity

$$R = Ir^2 = \frac{S}{4\pi r^2} \cdot r^2 \propto S$$

2



$$\mathbf{h}_i = \left\{ h_i^{(k)} \mid k \in [1, N_b] \right\} \in \mathbb{R}^{N_b}, i \in [1, s],$$

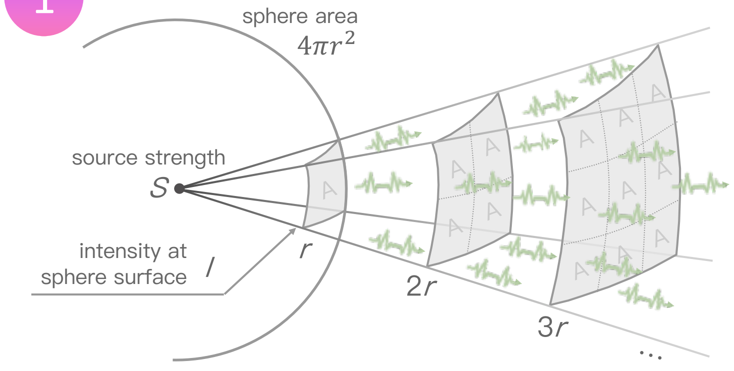
$$h_i^{(k)} = \# \left\{ \boxed{R_j} \in r_k, \forall j \mid p_j \in b_i \right\}$$

$$r_k = [(k-1)/N_b, k/N_b), k \in [1, N_b].$$

we apply various sizes of bins in cylindrical coordinates to analyze the intrinsic point distribution at varying resolutions (shown in h_1, h_2 and h_3).

Using reflectivity-based Test Time Augmentation
to enhance performance of false or non-existent pseudo-labels

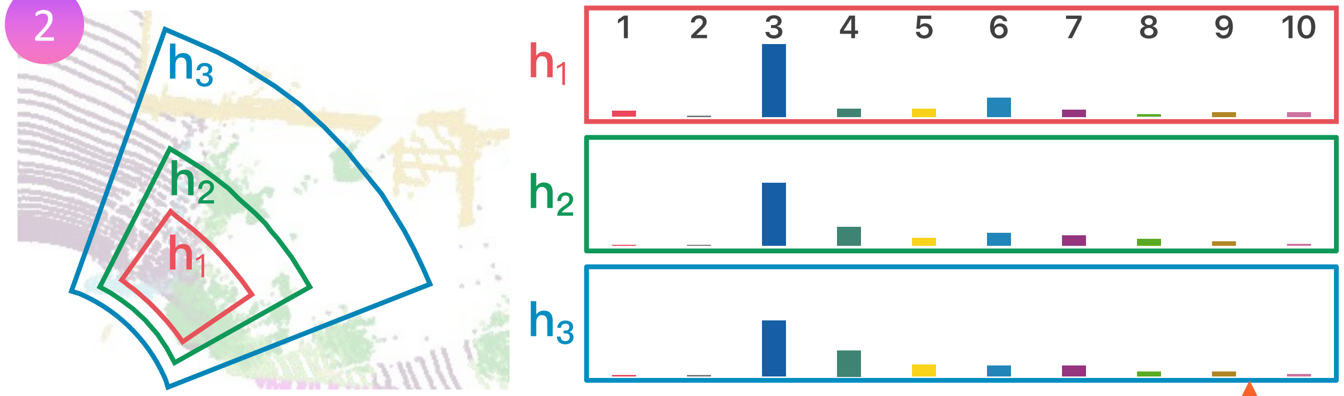
1



reflectivity

$$R = Ir^2 = \frac{S}{4\pi r^2} \cdot r^2 \propto S$$

2



$$\begin{aligned} \mathbf{h}_i &= \left\{ h_i^{(k)} \mid k \in [1, N_b] \right\} \in \mathbb{R}^{N_b}, i \in [1, s], \\ h_i^{(k)} &= \# \left\{ \boxed{R_j} \in r_k, \forall j \mid p_j \in b_i \right\} \\ r_k &= [(k-1)/N_b, k/N_b), k \in [1, N_b]. \end{aligned}$$

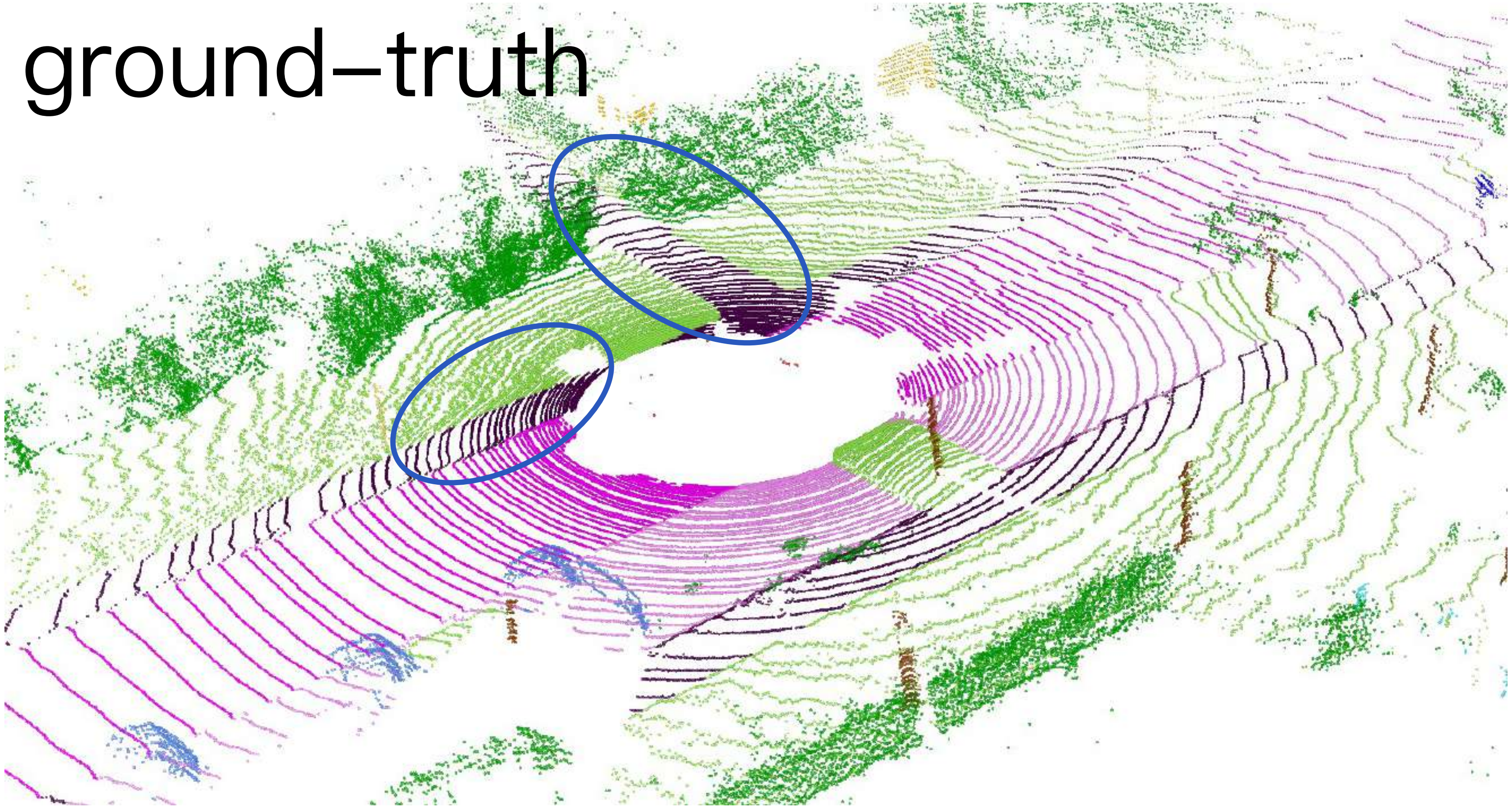
3

normalize
augmentation

$$\begin{cases} R^{\circledast} = \{\mathbf{h}_i / \max(\mathbf{h}_i) \mid i \in [1, s]\} \in \mathbb{R}^{sN_b} \\ P^{\circledast} = \{p \mid (x, y, z, I, R^{\circledast}) \in \mathbb{R}^{sN_b+4}\} \end{cases}$$

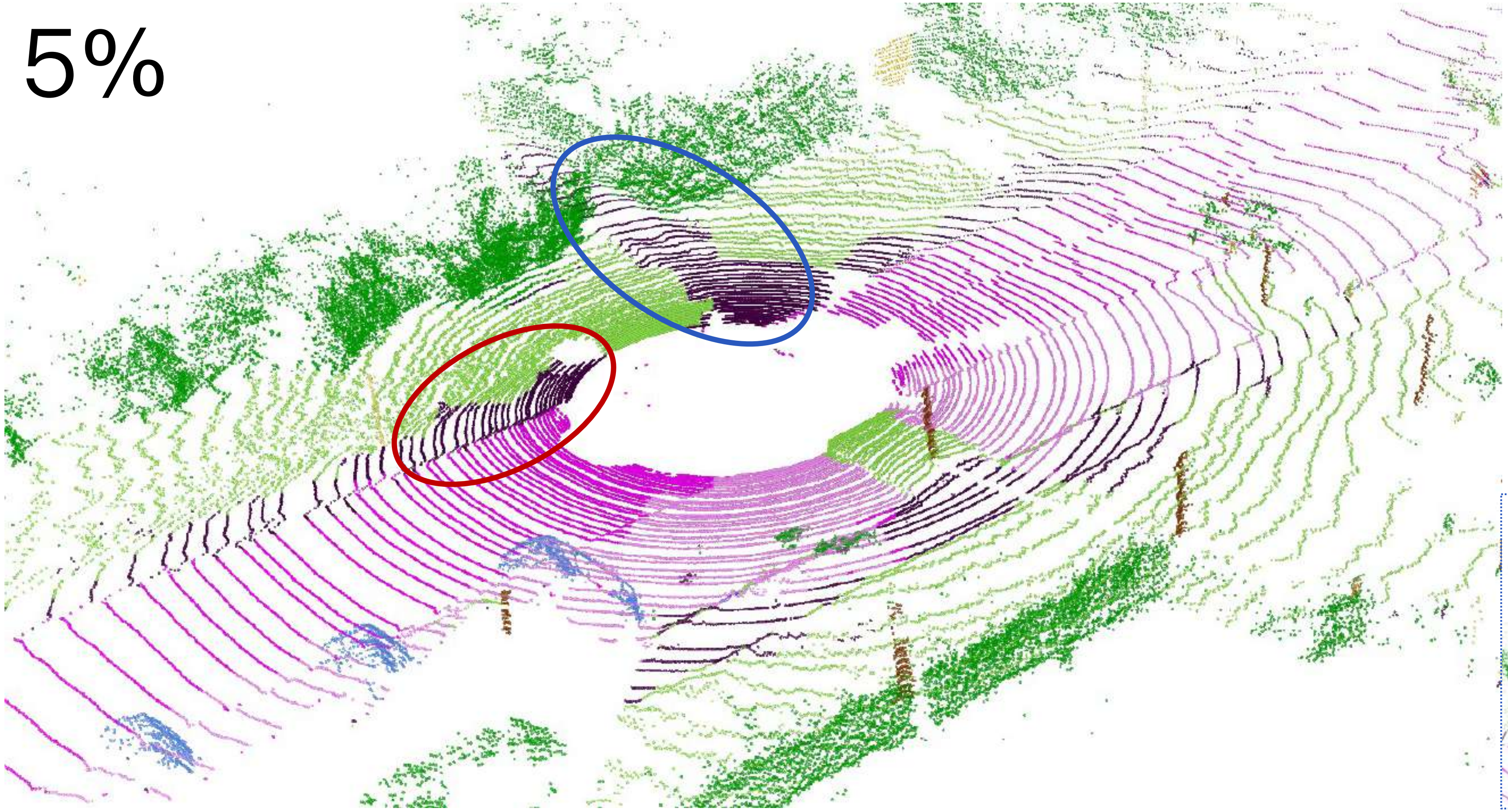
we then normalize it and append to the point set

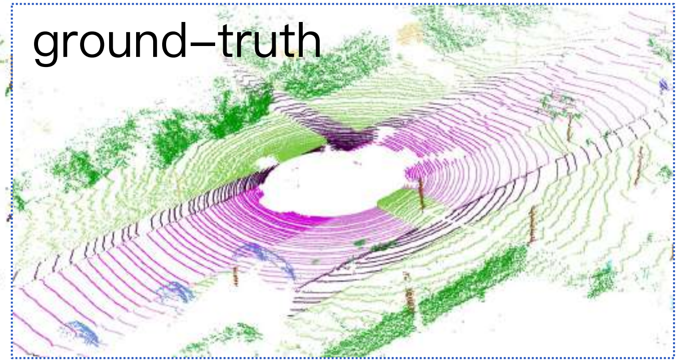
Qualitative results
Comparing {5%, 10%, 20%, 40%} labeled splits



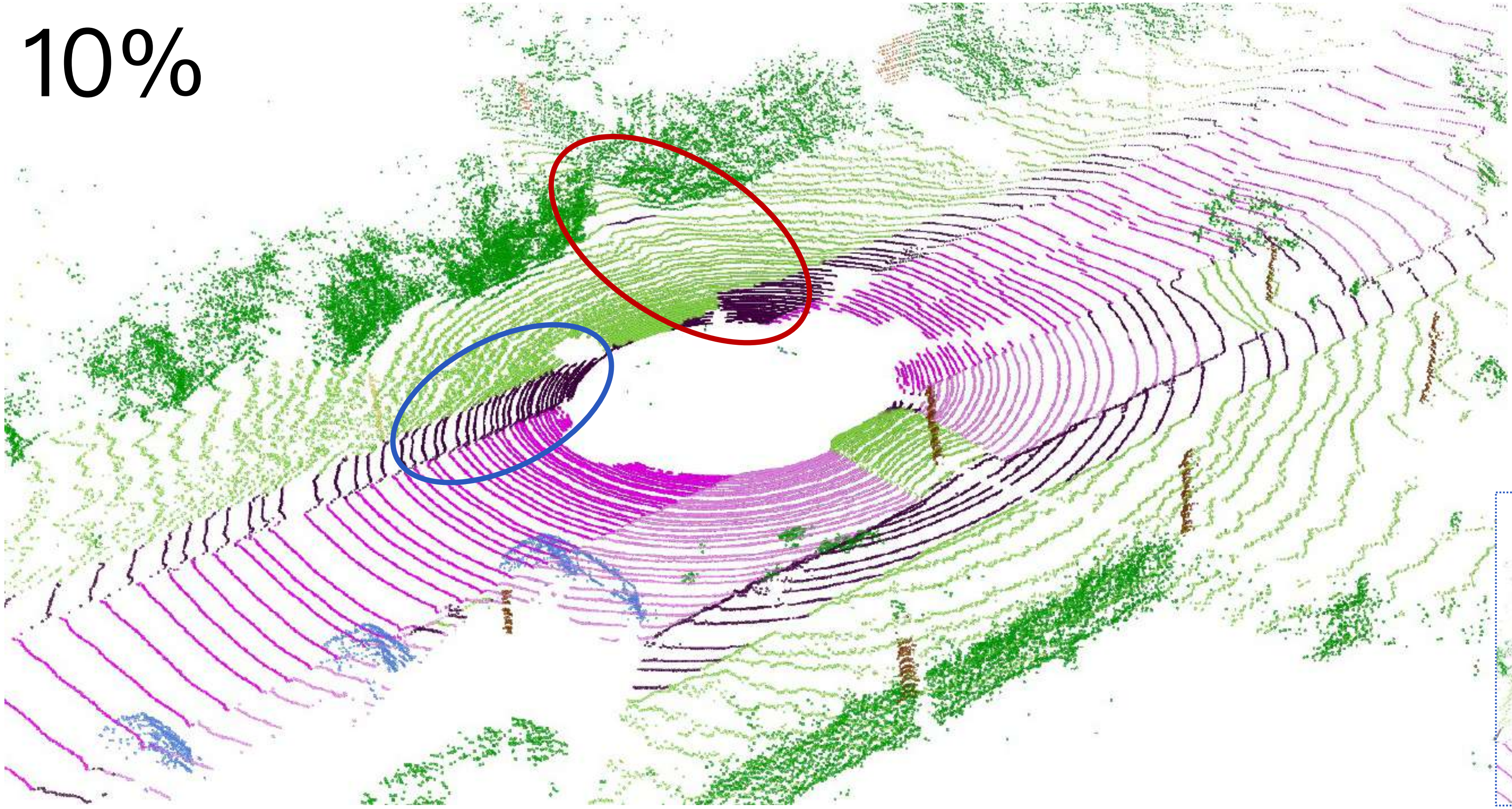
Qualitative results
Comparing {5%, 10%, 20%, 40%} labeled splits

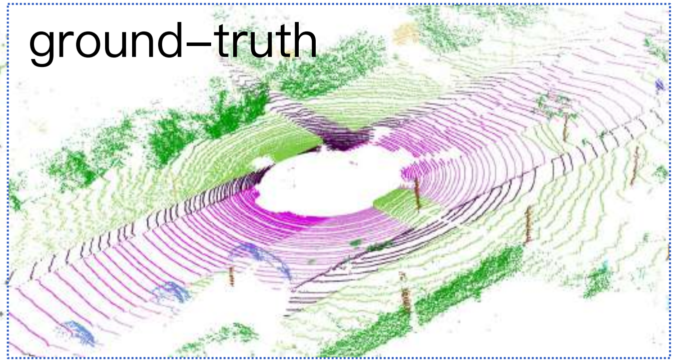
5%





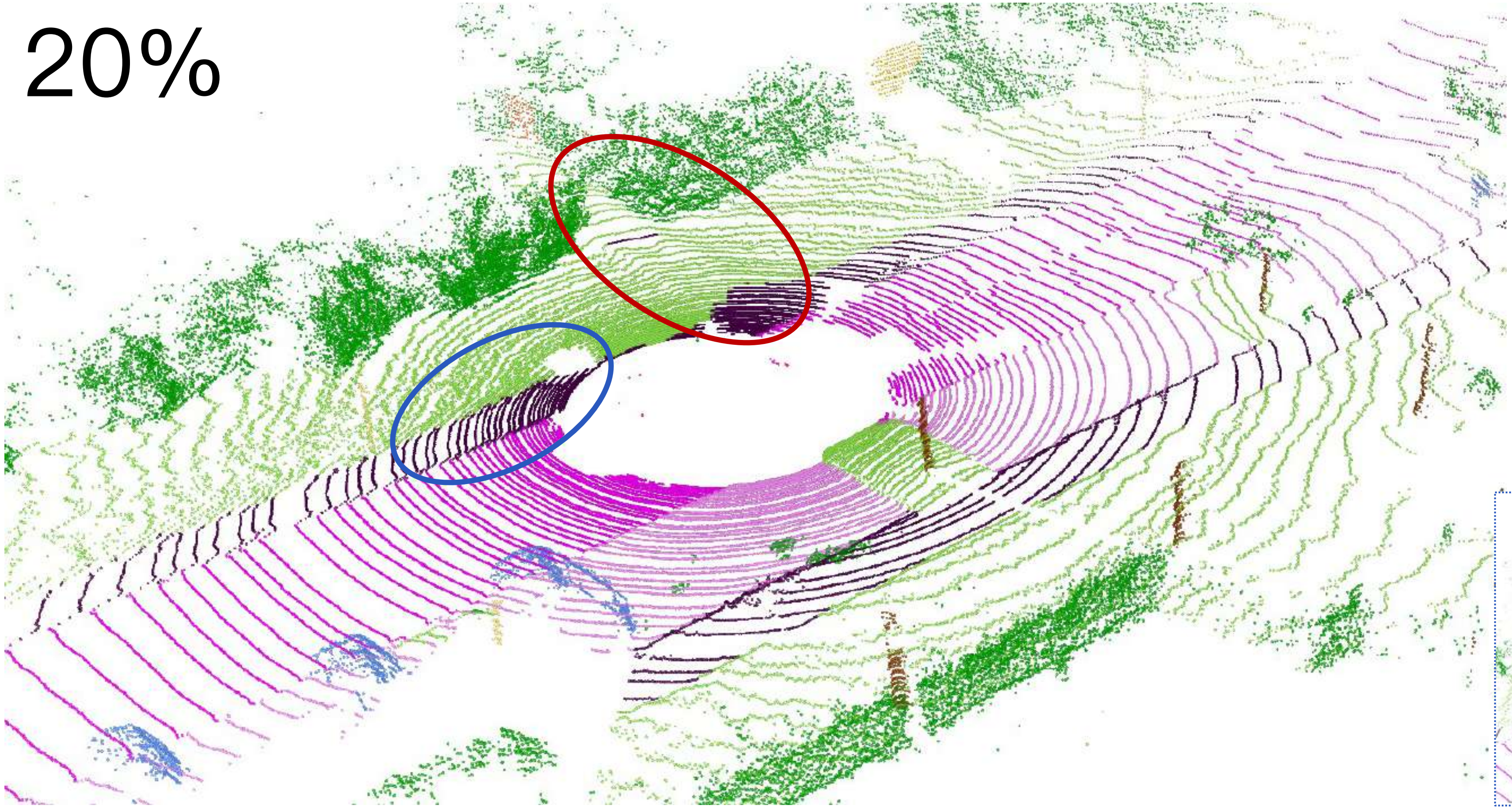
Qualitative results
Comparing {5%, 10%, 20%, 40%} labeled splits

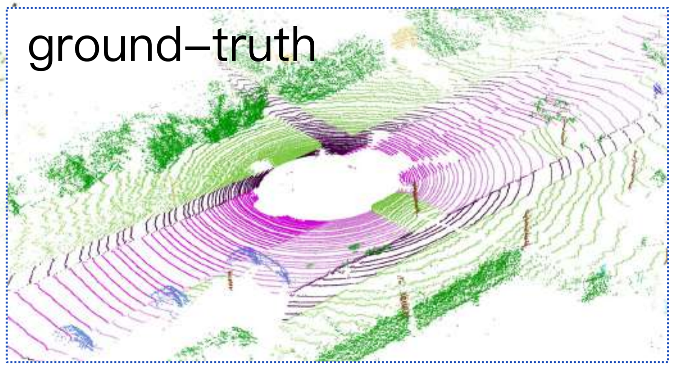




Qualitative results
Comparing {5%, 10%, 20%, 40%} labeled splits

20%

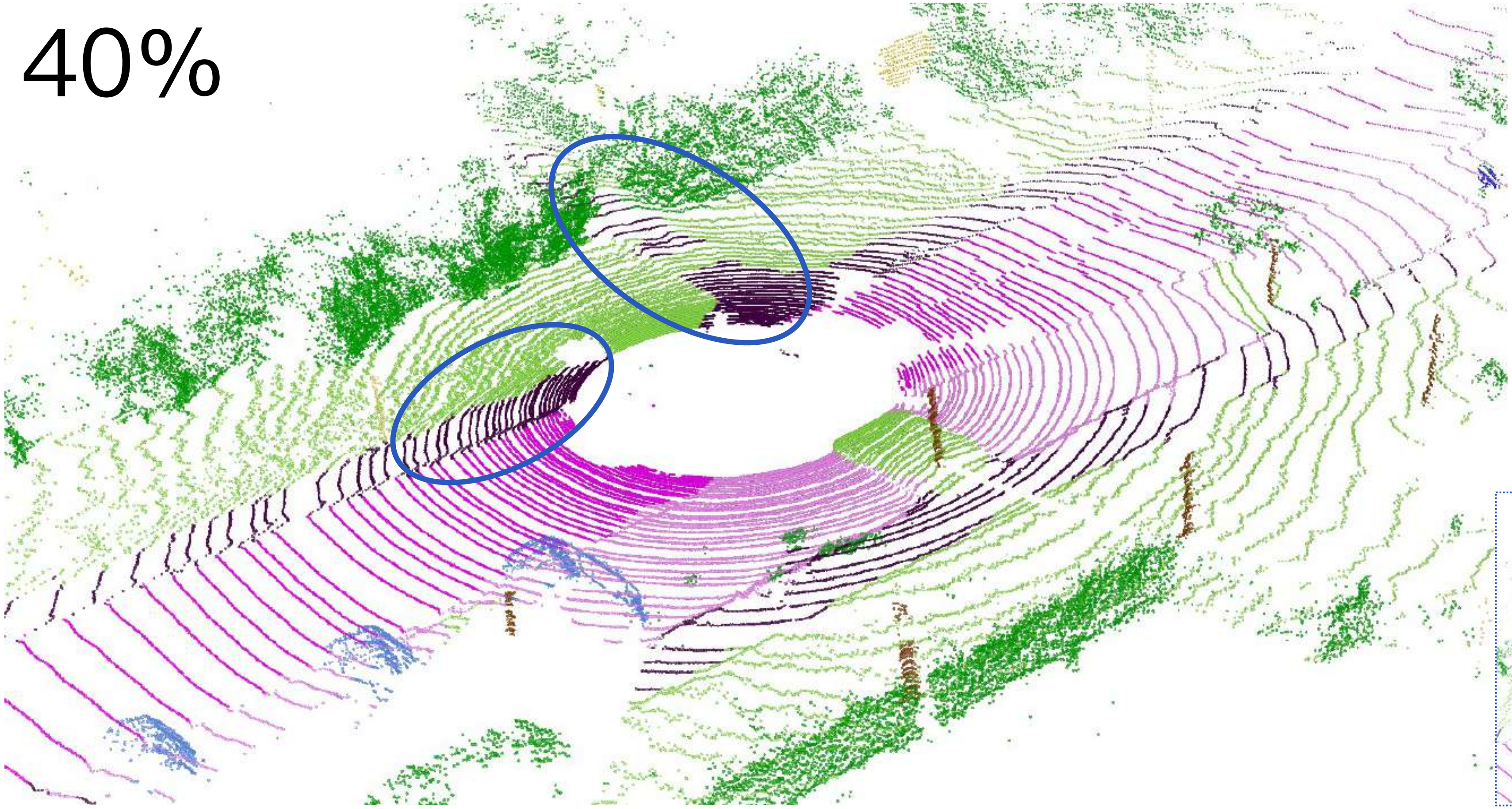


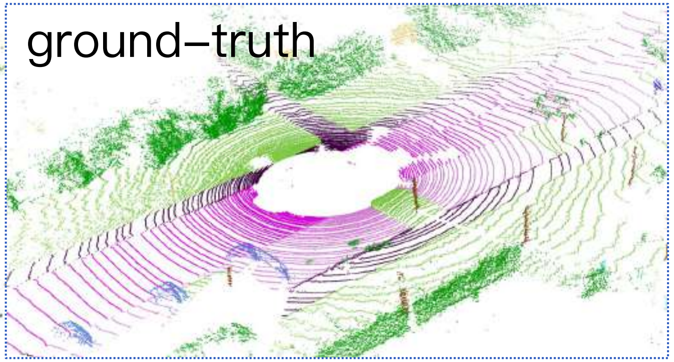


Qualitative results

Comparing {5%, 10%, 20%, 40%} labeled splits

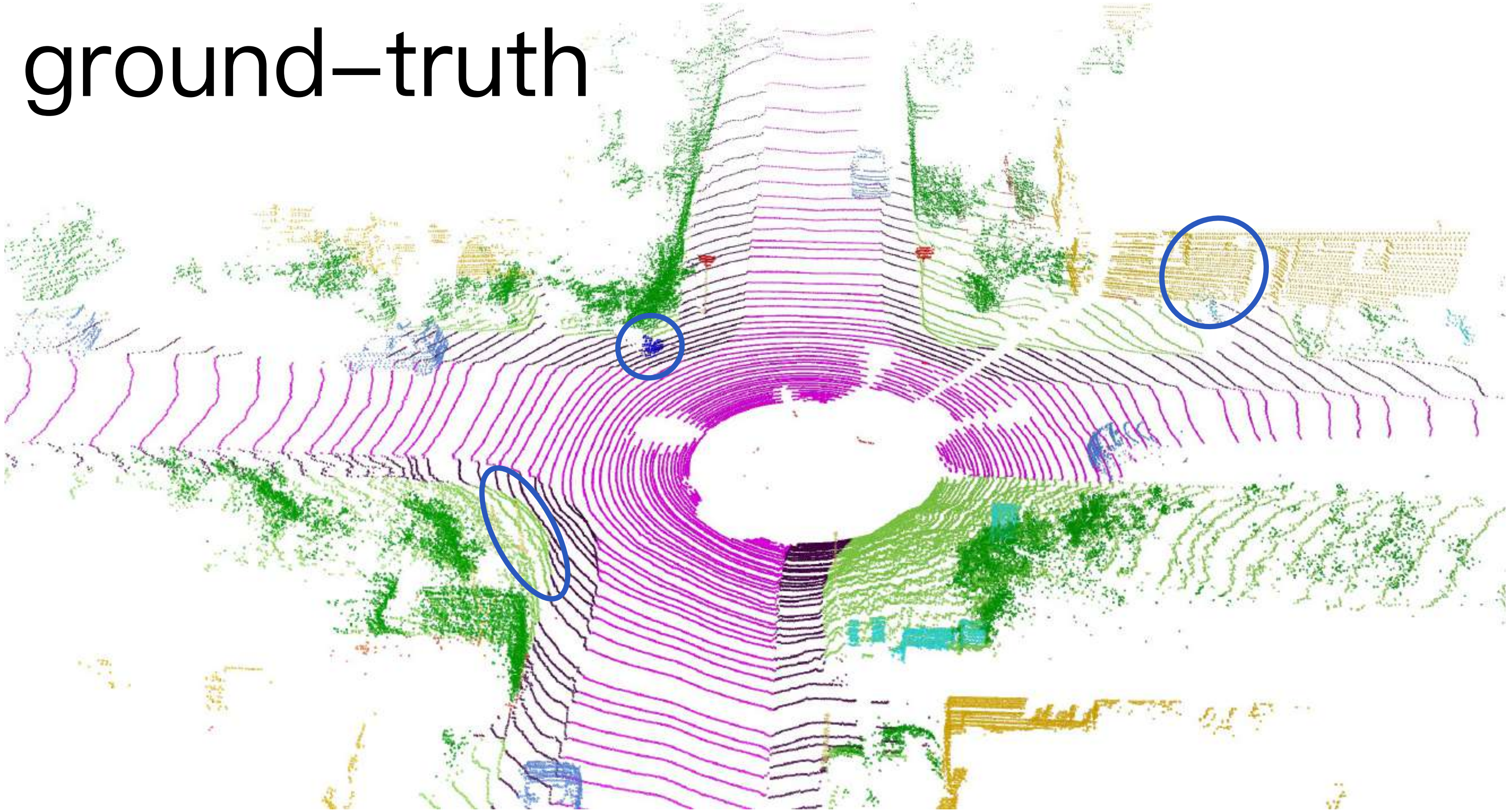
40%





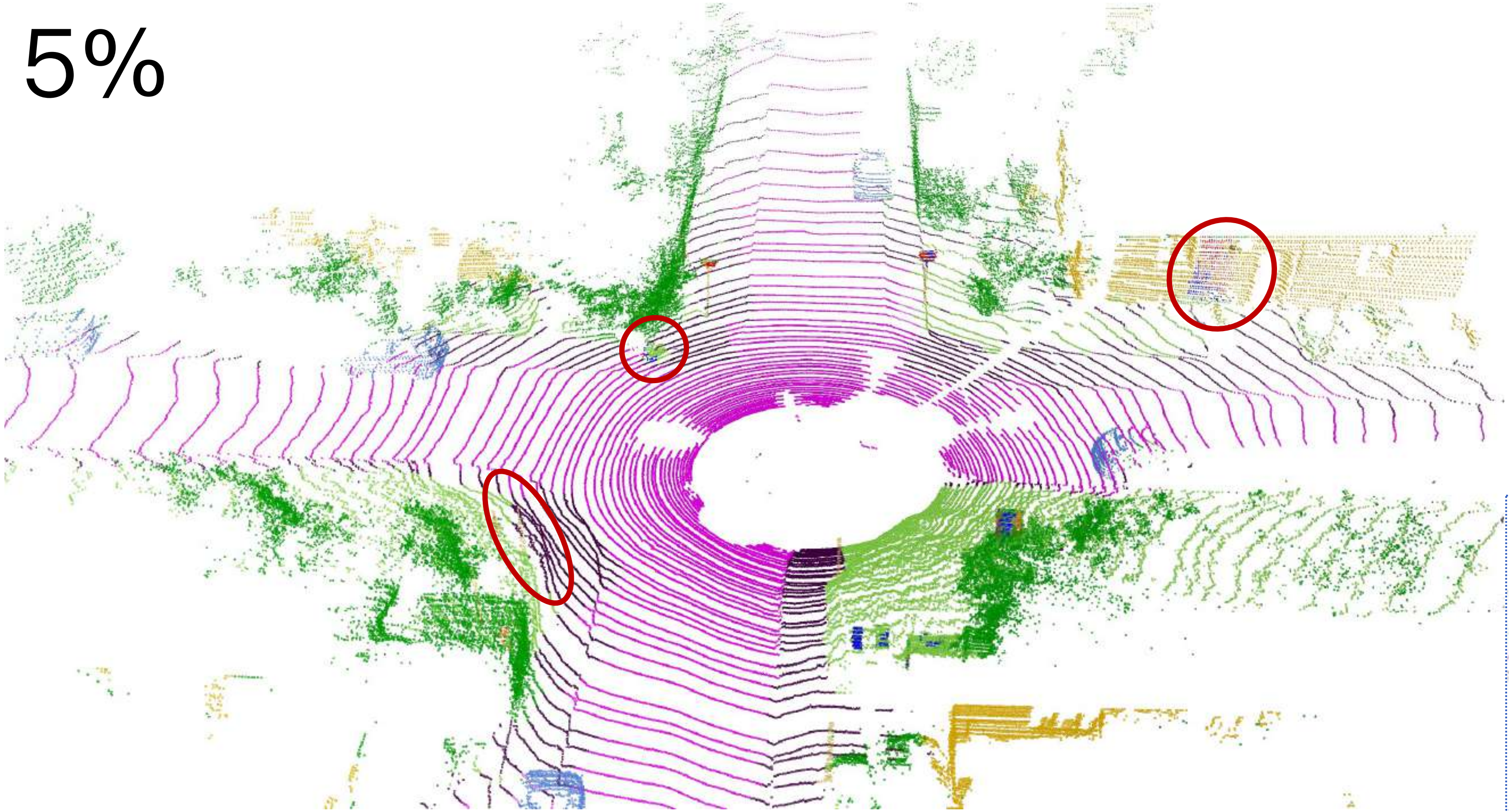
Qualitative results

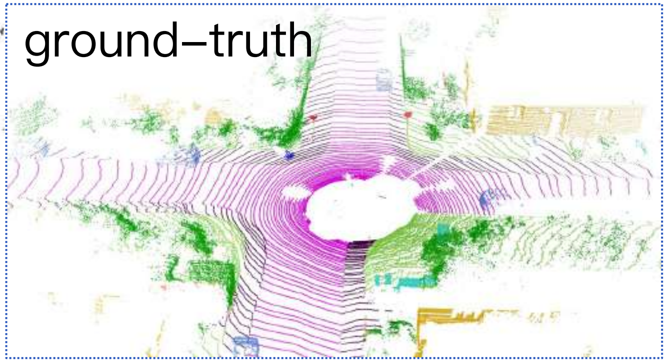
Comparing {5%, 10%, 20%, 40%} labeled splits



Qualitative results
Comparing {5%, 10%, 20%, 40%} labeled splits

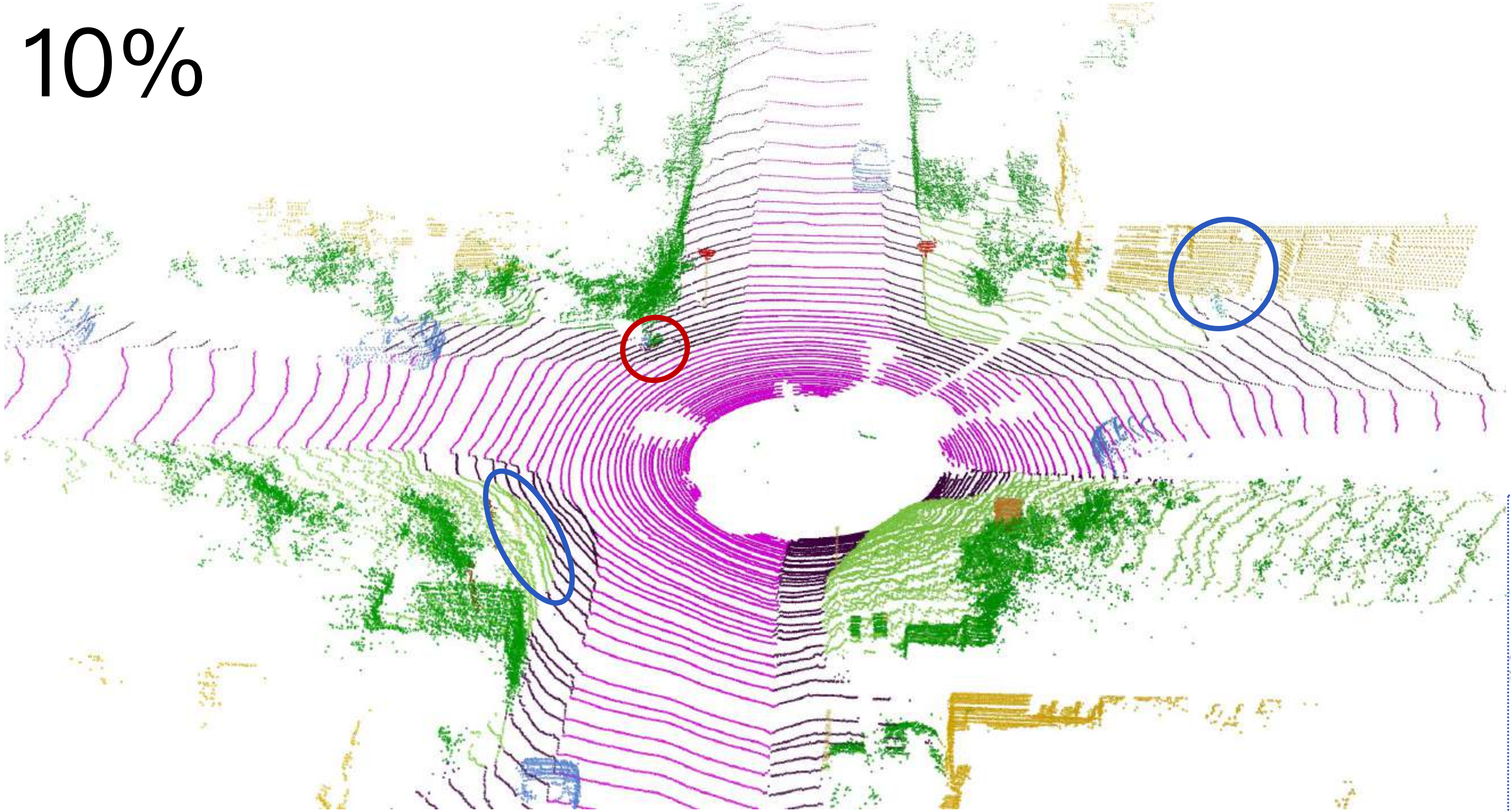
5%

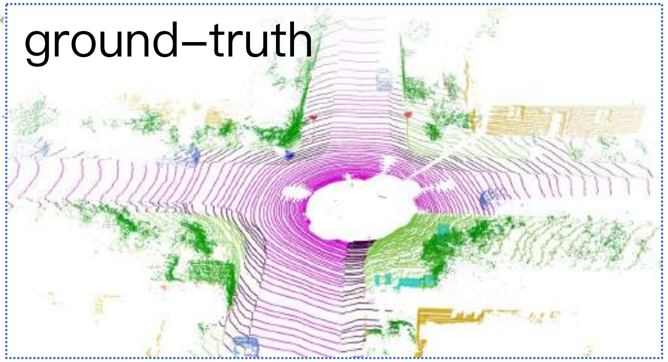




Qualitative results
Comparing {5%, 10%, 20%, 40%} labeled splits

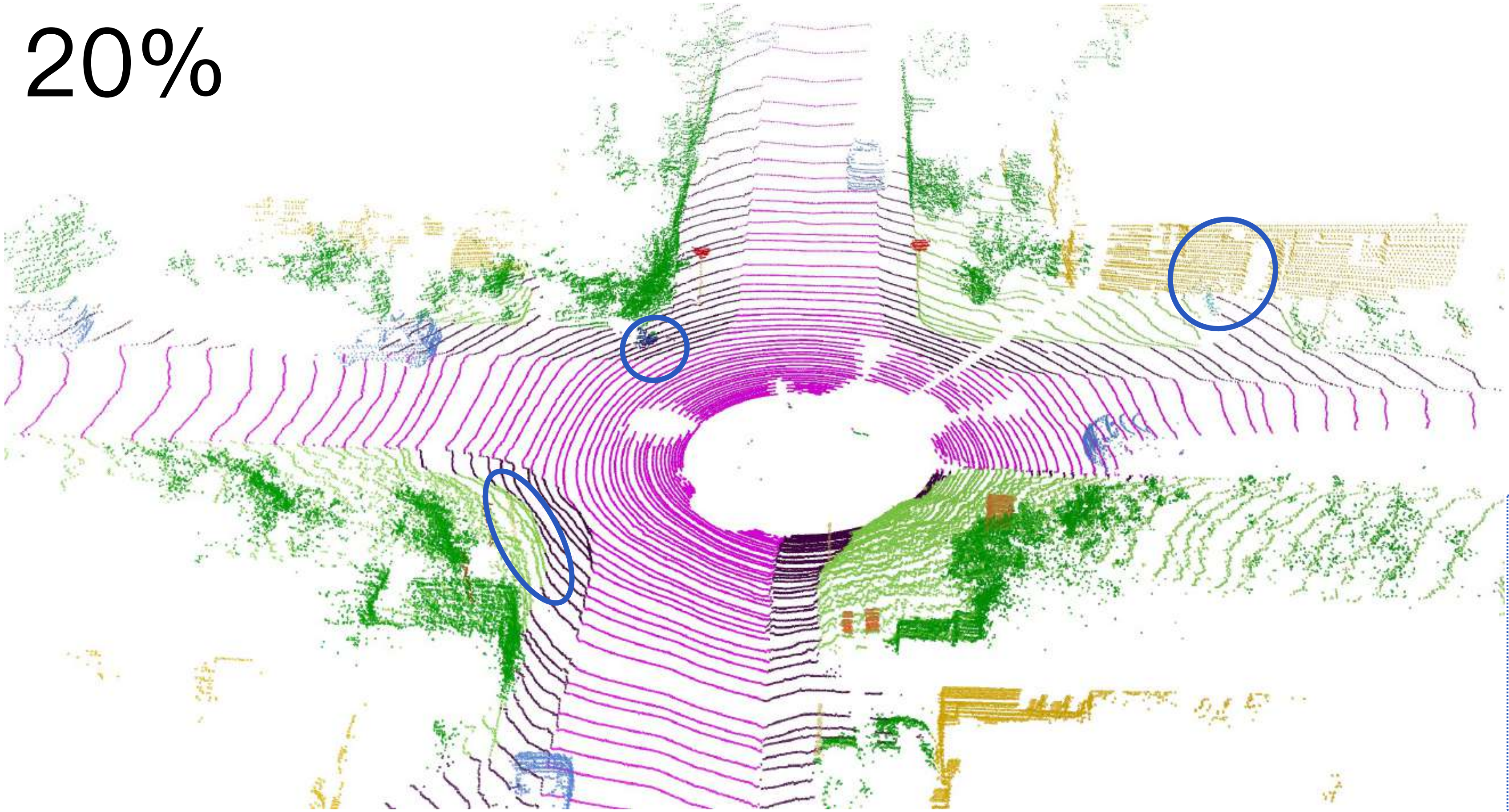
10%

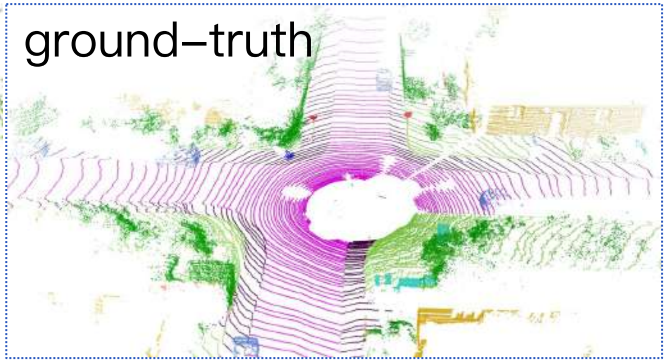




Qualitative results
Comparing {5%, 10%, 20%, 40%} labeled splits

20%

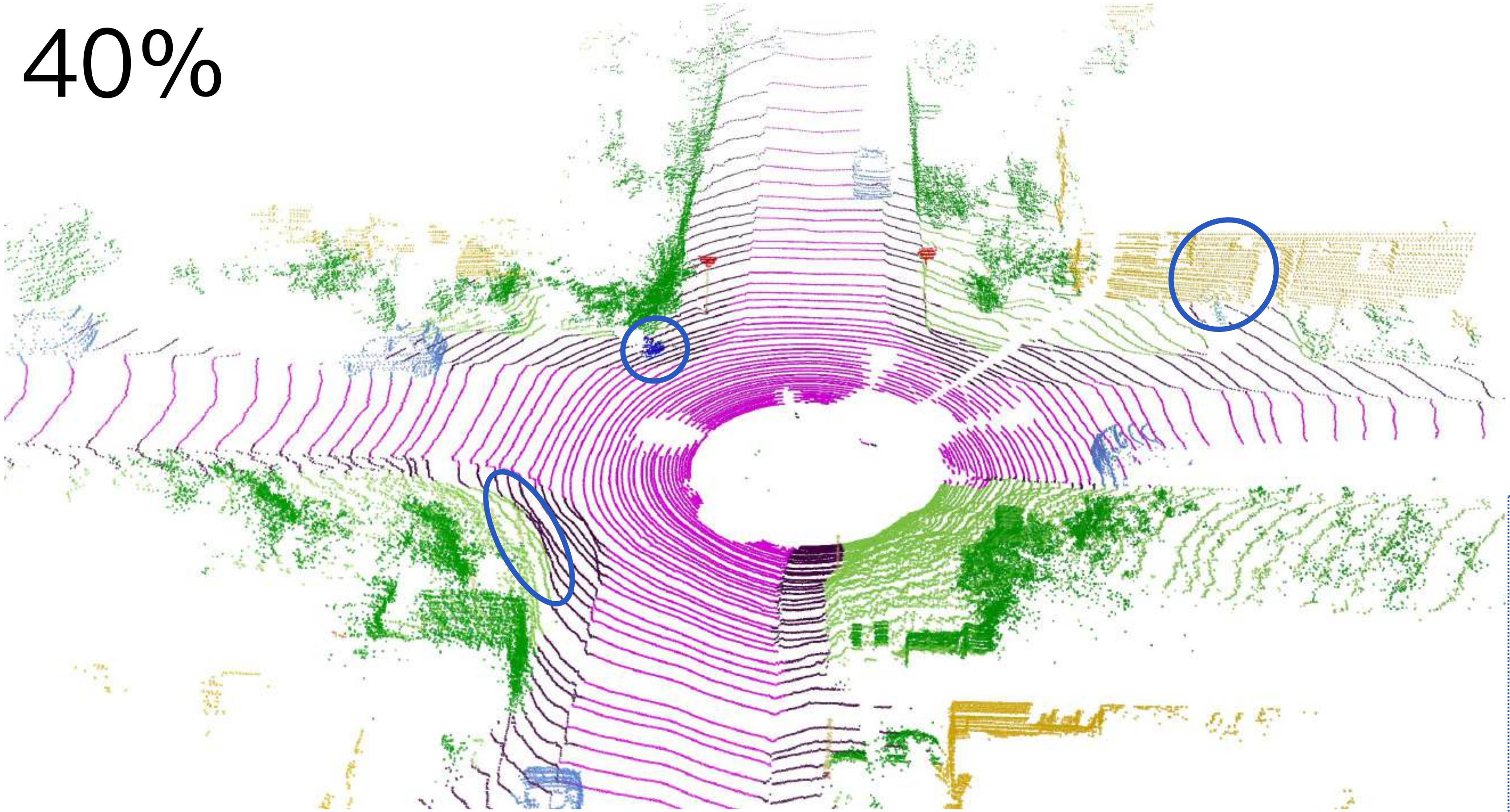


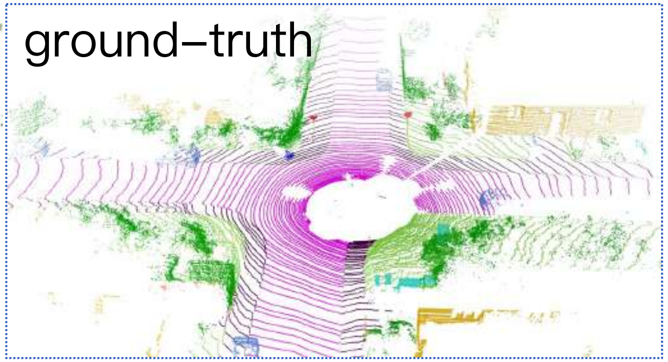


Qualitative results

Comparing {5%, 10%, 20%, 40%} labeled splits

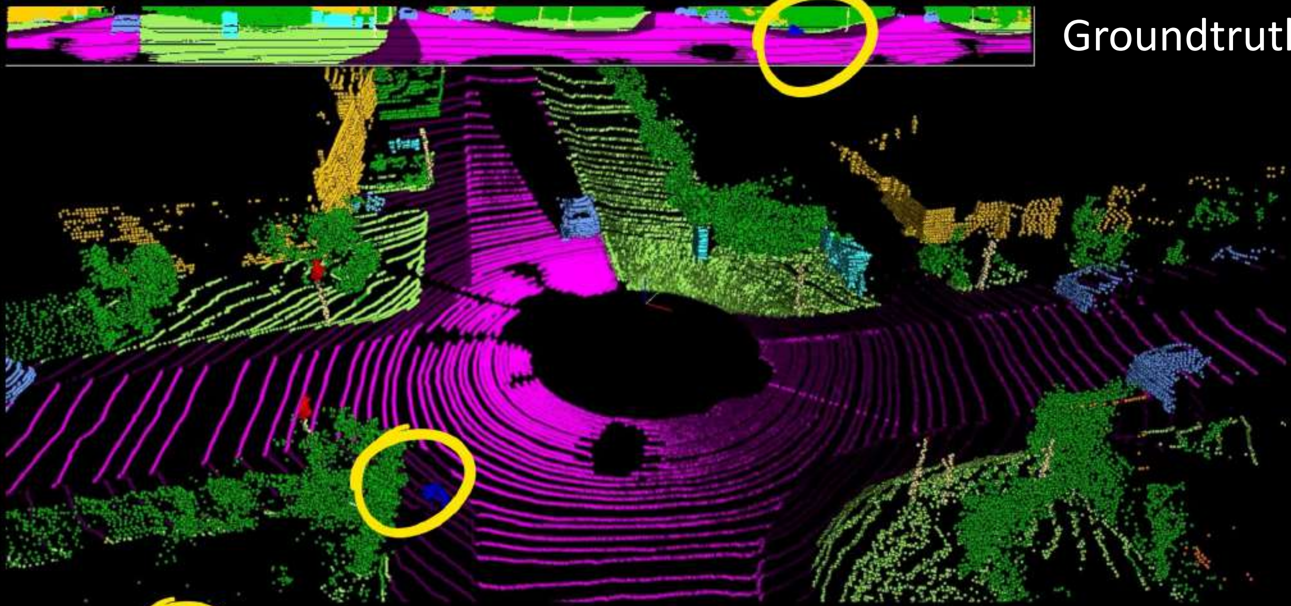
40%



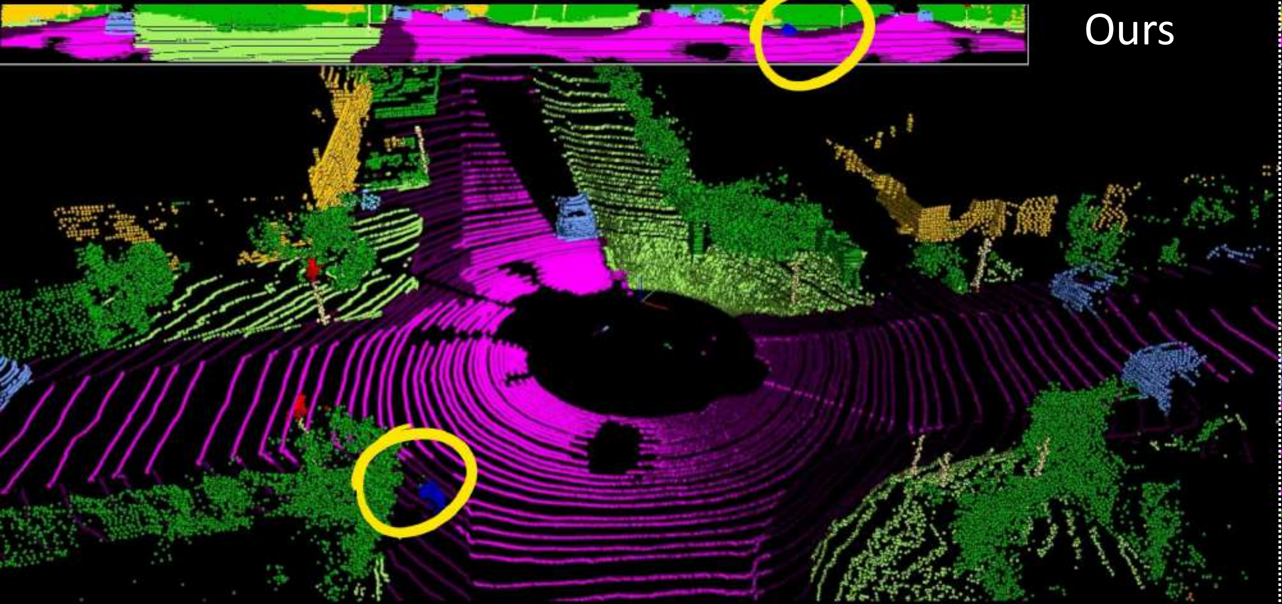


Qualitative results

5%-Labeled Frames



Groundtruth



Ours

(Ozan et al)

Comparative mIoU for Semi-supervised Methods

Repr.	Samp.	Method		SemanticKITTI [7]							ScribbleKITTI [46]						
				1%	5%	10%	20%	40%	50%	100%	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	61.9	–	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	–	68.2*	–	46.9*	54.2*	56.5*	58.6*	–	61.3
	S	LiM3D+SDSC	(ours)	57.2	57.6	61.0	61.7	62.1	62.7	67.5	55.8	56.1	56.9	57.2	58.9	59.3	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**

Comparative mIoU for Semi-supervised Methods

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	61.9	–	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	–	68.2*	–	46.9*	54.2*	56.5*	58.6*	–	61.3
	S	LiM3D+SDSC	(ours)	57.2	57.6	61.0	61.7	62.1	62.7	67.5	55.8	56.1	56.9	57.2	58.9	59.3	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**

Component-wise Ablation (Ours)

	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
LiM3D	✓	✓	✓	✓		**85.2**	**89.1**	**89.5**	**89.7**	**59.5**	**62.2**	**63.1**	**63.3**	49.6
LiM3D+SDSC	✓	✓	✓	✓	✓	83.8	88.6	89.0	89.2	57.6	61.0	61.7	62.1	**21.5**

UP Unreliable Pseudo labeling

RF Reflectivity Feature

RT Reflec–TTA

ST ST–RFD

SD SDSC module

Component-wise Ablation (Ours)

	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
LiM3D	✓	✓	✓	✓		**85.2**	**89.1**	**89.5**	**89.7**	**59.5**	**62.2**	**63.1**	**63.3**	49.6
LiM3D+SDSC	✓	✓	✓	✓	✓	83.8	88.6	89.0	89.2	57.6	61.0	61.7	62.1	**21.5**

UP Unreliable Pseudo labeling

RF Reflectivity Feature

RT Reflec-TTA

ST ST-RFD

SD SDSC module

The Computation Cost and mIoU

Under 5%–labeled Training Results

Method	# Parameters	# Mult-Adds	SeK [7]	ScK [45]
Cylider3D [61]	56.3	476.9M	45.4	39.2
Ozan *et al.* [45]	49.6	420.2M	49.9	46.9
2DPASS [56]	26.5	<u>217.4M</u>	51.7	45.1
MinkowskiNet [13]	21.7	114.0G	42.4	35.8
SPVNAS [43]	**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**

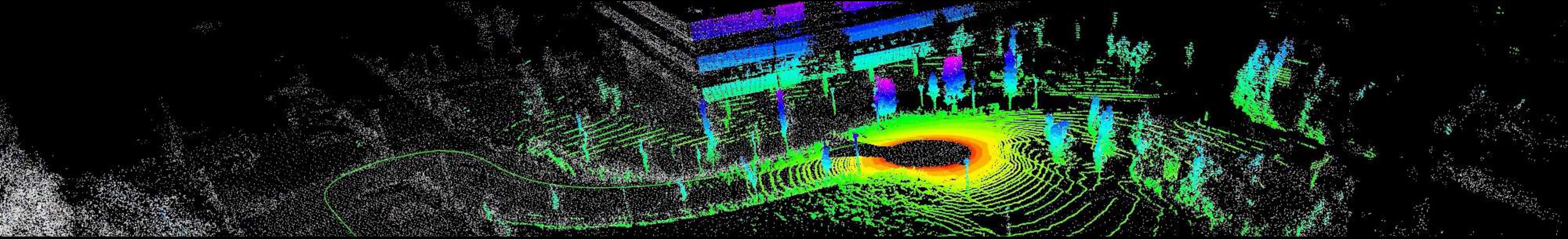
The Computation Cost and mIoU

Under 5%-labeled Training Results

Method	# Parameters	# Mult-Adds	SeK [7]	ScK [45]
Cylider3D [61]	56.3	476.9M	45.4	39.2
Ozan *et al.* [45]	49.6	420.2M	49.9	46.9
2DPASS [56]	26.5	217.4M	51.7	45.1
MinkowskiNet [13]	21.7	114.0G	42.4	35.8
SPVNAS [43]	**12.5**	73.8G	45.1	38.9
LiM3D+SDSC (ours)	21.5	**182.0M**	57.6	54.7
LiM3D (ours)	49.6	420.2M	**59.5**	**58.1**

2.3x model size reduction

641x fewer multiply-adds



DurLAR: A High-Fidelity 128-Channel LiDAR Dataset with Panoramic Ambient and Reflectivity Imagery for Multi-Modal Autonomous Driving Applications

<u>**Li Li**</u>

Khalid N. Ismail

Hubert P. H. Shum

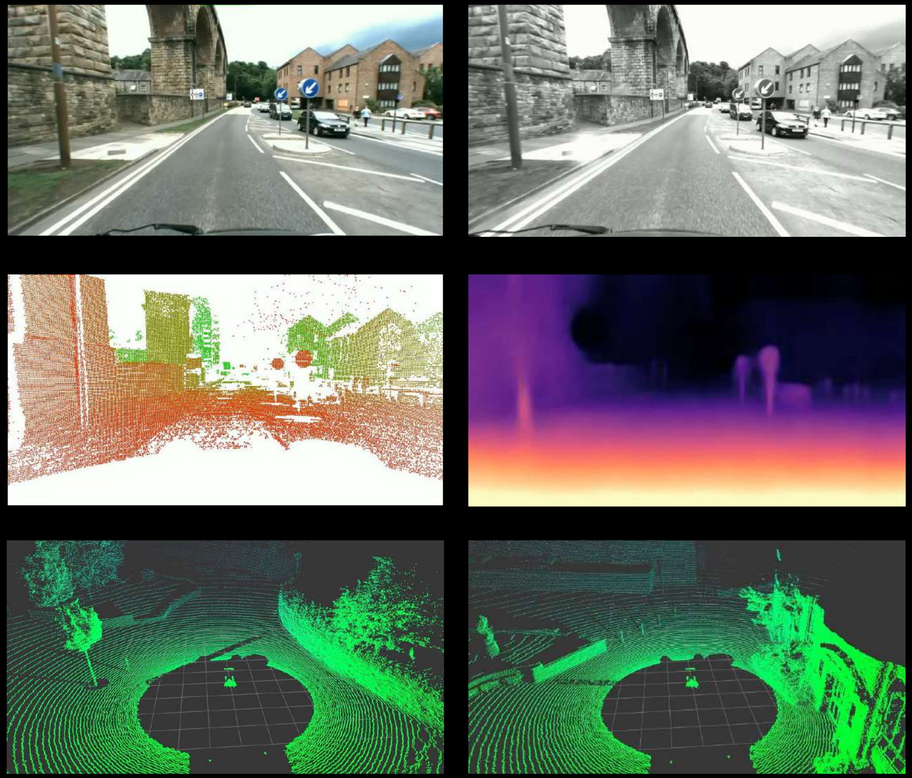
Toby P. Breckon

International Conference on 3D Vision, 2021

li.li4@durham.ac.uk

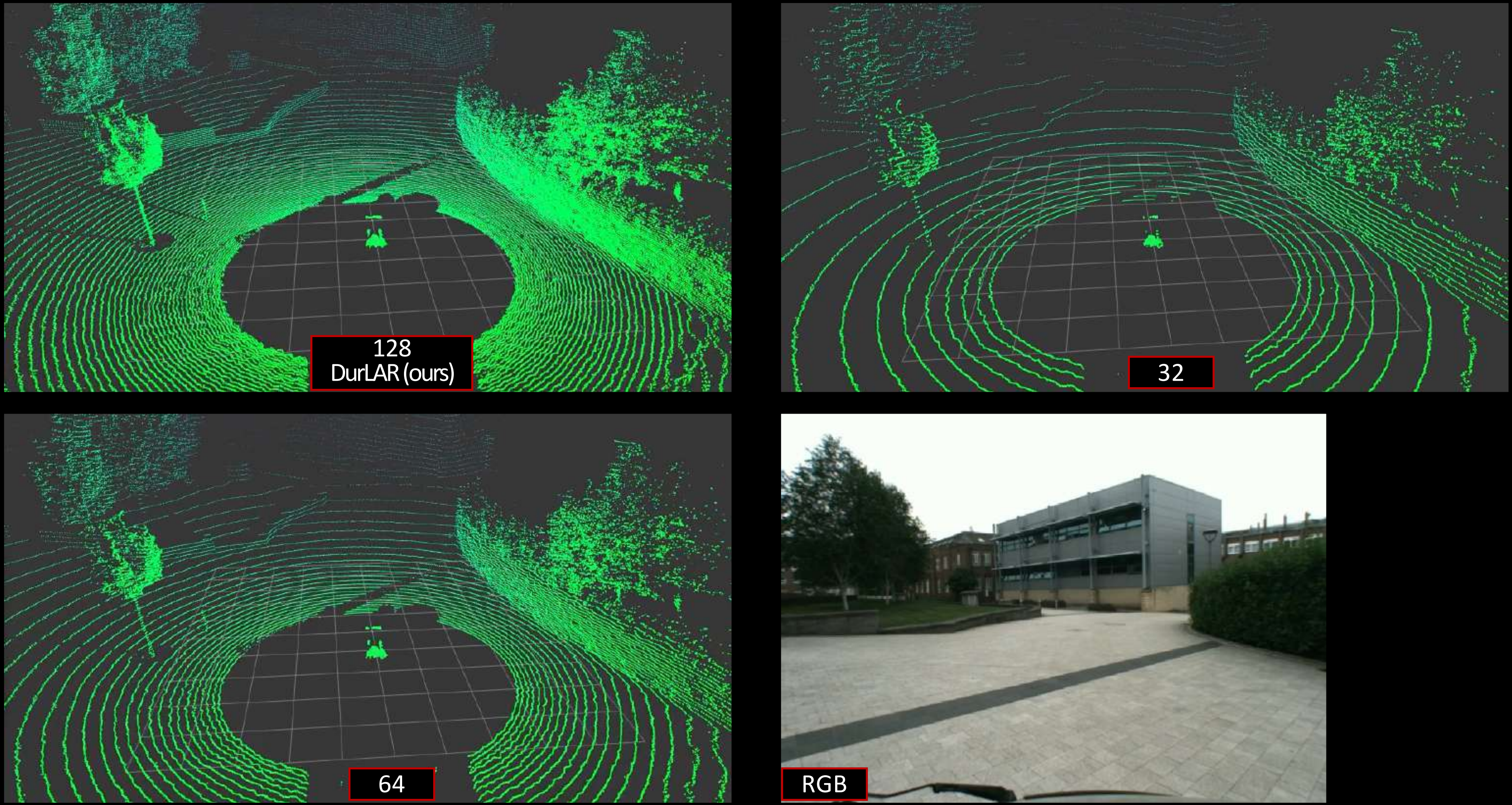
Durham University

DurLAR Dataset - Overview

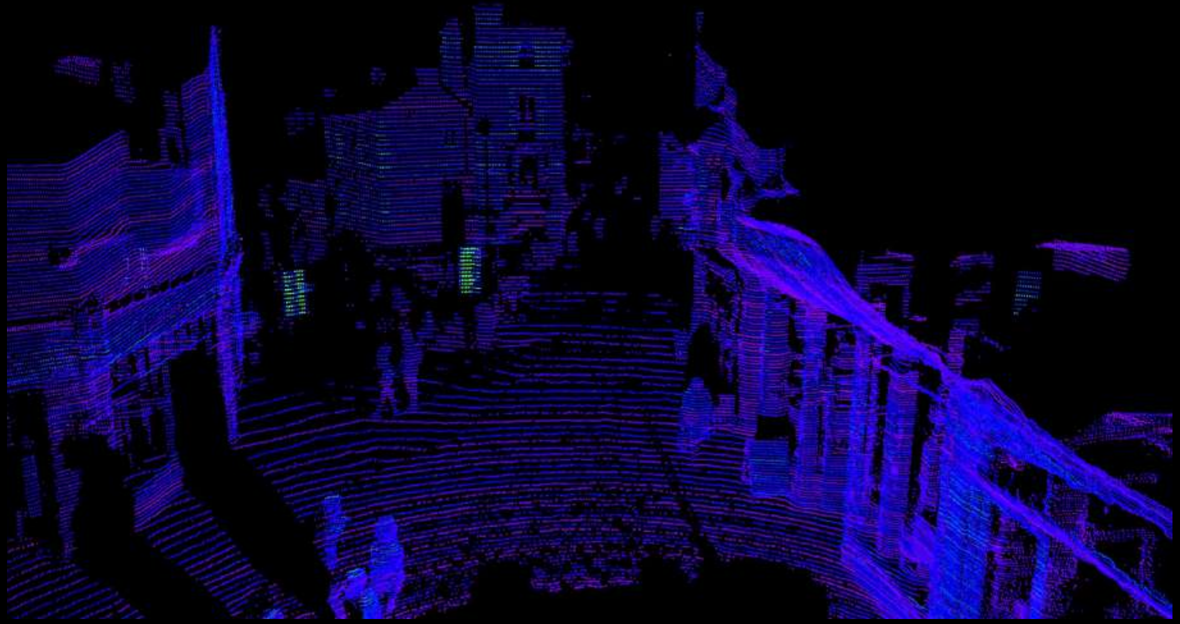


- A High-fidelity **128-channel LiDAR Dataset**
  - 100k+ frames
  - Synchronised at 10Hz
- First dataset with **LiDAR panoramic imagery**
  - Ambient imagery
  - Reflectivity imagery
- **Diversity**: time of day, repeated locations, weather
- **Monocular Depth Estimation - benchmark test**
  - Self-supervised ManyDepth
  - Supervised/Self-supervised ManyDepth

Higher Fidelity: 128 vs. 64/32 channel LiDAR

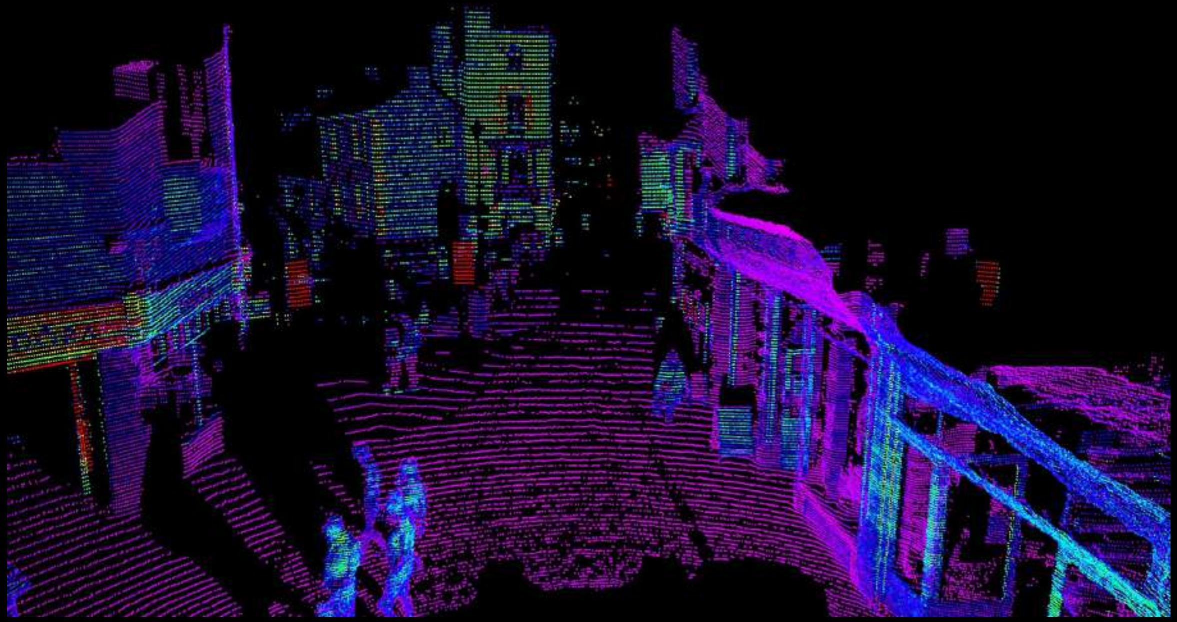


LiDAR Panoramic Imagery



Ambient

day/night scene visibility in the near-IR spectrum



Reflectivity

information indicative of the material properties of the object itself and offer good consistency across illumination conditions and range.

Diversity of Dataset Environments

Comparison with Existing Public LiDAR Datasets

Dataset	Resolution	Range/m	Diversity	Image	#Frames	Other sensors
DENSE	64	120	E/W/T	I	1M	D/M/F/T/B
H3D	64	120	E	I	28k	G/M
KITTI | SemanticKITTI	64	120	E	I	93k	N/S/G/M/B
KITTI-360	64	120	E	I	320k	N/S/G/M/B
LiVi-Set	32	100	E	I	10k	
Lyft Level 5	64	200	E/W/T	I	170k	D/B
nuScenes	32	100	E/W/T	I	1M	M/D/B
Oxford RobotCar	4	50	E/W/T	I	3M	N/S/G/M/B
Stanford Track	64	120	E	I	14k	M
Sydney Urban Objects	64	120	E	I	0.6k	
DurLAR (ours)	128	120	E/W/T/L	I/A/R	100k	U/N/S/G/M/B

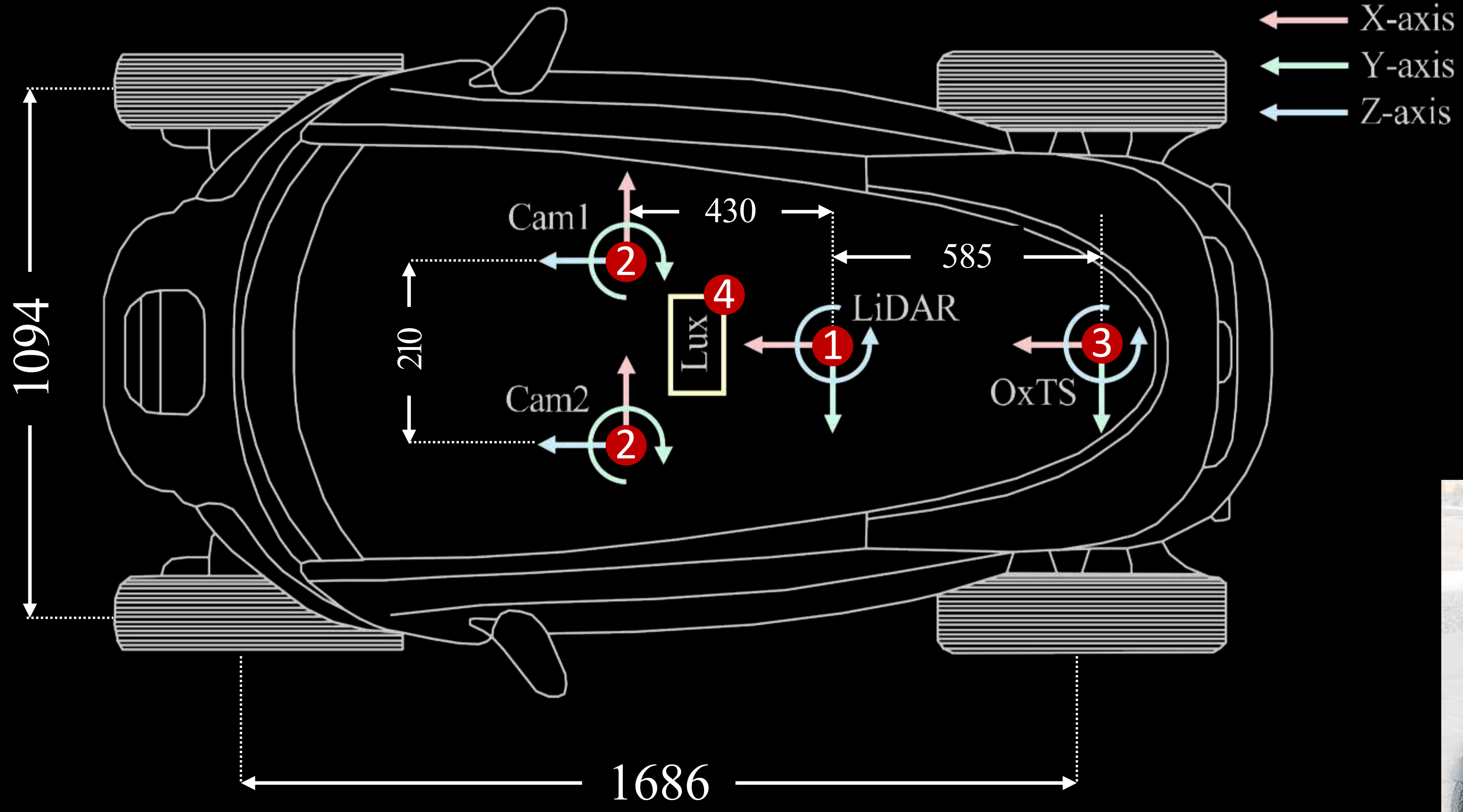
Image	Refer to	Diversity	Refer to	Sensors	Refer to	Sensors	Refer to
I	intensity	E	environments	D	radar	M	IMU
A	ambient	W	weather condition	U	lux meter	F	FIR camera
R	reflectivity	T	times of day	N	GNSS	T	Near IR camera
		L	repeated location	S	INS	B	stereo camera
				G	GPS		

Comparison with Existing Public LiDAR Datasets

Dataset	Resolution	Range/m	Diversity	Image	#Frames	Other sensors
DENSE	64	120	E/W/T	I	1M	D/M/F/T/B
H3D	64	120	E	I	28k	G/M
KITTI | SemanticKITTI	64	120	E	I	93k	N/S/G/M/B
KITTI-360	64	120	E	I	320k	N/S/G/M/B
LiVi-Set	32	100	E	I	10k	
Lyft Level 5	64	200	E/W/T	I	170k	D/B
nuScenes	32	100	E/W/T	I	1M	M/D/B
Oxford RobotCar	4	50	E/W/T	I	3M	N/S/G/M/B
Stanford Track	64	120	E	I	14k	M
Sydney Urban Objects	64	120	E	I	0.6k	
➡ DurLAR (ours)	128	120	E/W/T/L	I/A/R	100k	U/N/S/G/M/B

Image	Refer to	Diversity	Refer to	Sensors	Refer to	Sensors	Refer to
I	intensity	E	environments	D	radar	M	IMU
A	ambient	W	weather condition	U	lux meter	F	FIR camera
R	reflectivity	T	times of day	N	GNSS	T	Near IR camera
		L	repeated location	S	INS	B	stereo camera
				G	GPS		

On Vehicle Sensor Placement



1 Ouster LiDAR

2 Stereo camera

3 GNSS/INS antenna

3 GNSS/INS (inside)

4 Lux meter

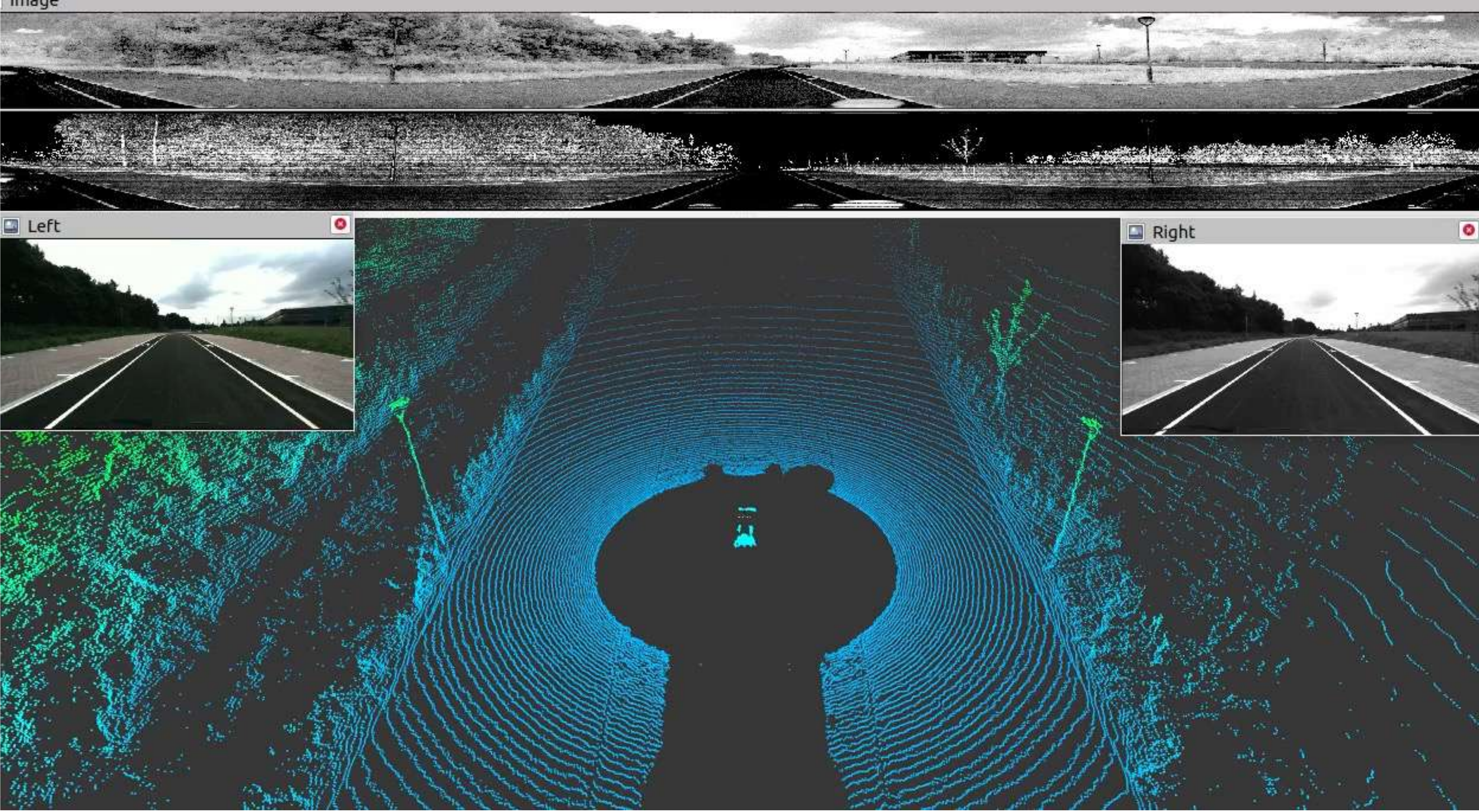
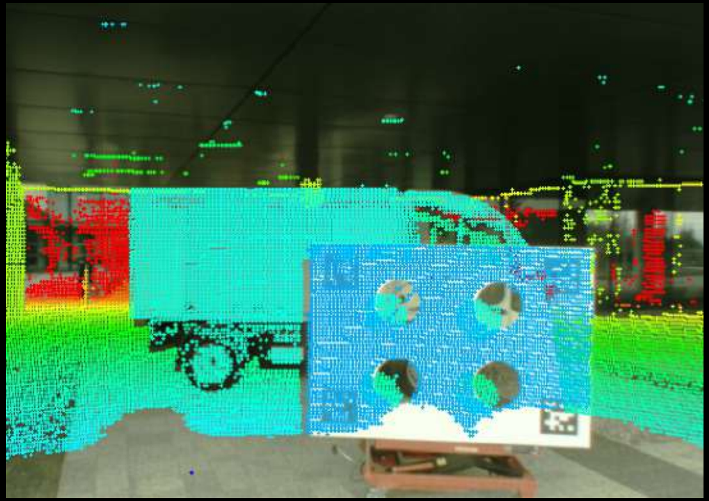
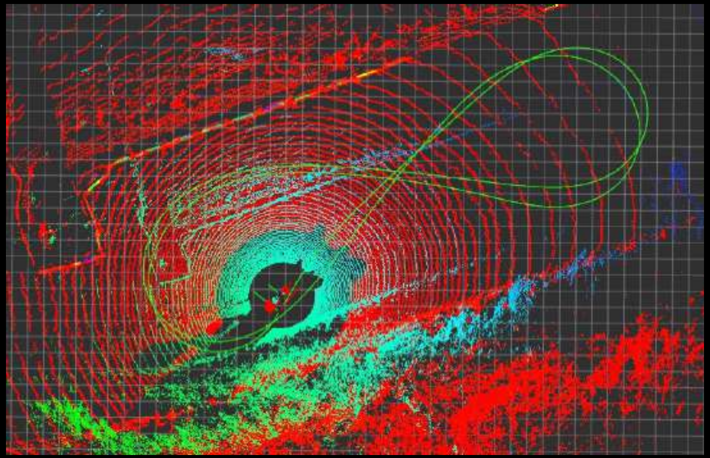


Image
Left
Right

Calibration and Synchronisation

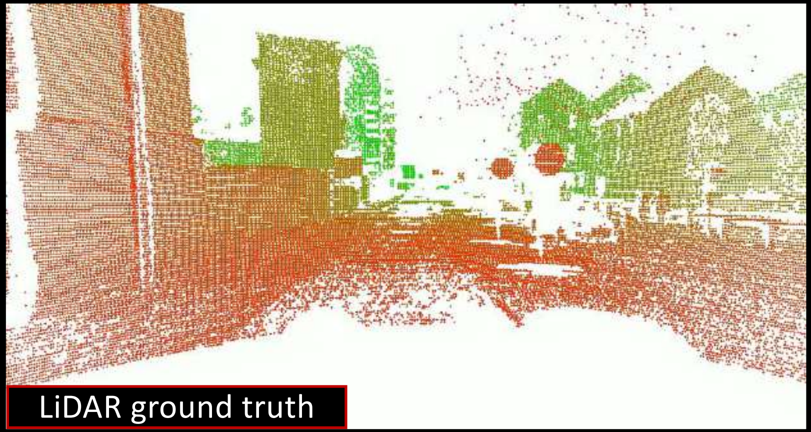




All sensor synchronisation is performed at a rate of **10 Hz**, using ROS (version Noetic) timestamps operating over a Gigabit Ethernet backbone to a common host (Intel Core i5, 16 GB RAM).

Sensor	Collecting rate	External calibration
LiDAR	10Hz	(a) Stereo; (b) GNSS/INS
GNSS/INS	100Hz	(b) LiDAR; stereo
Stereo	30Hz	(a) LiDAR; GNSS/INS
Lux meter	30Hz	

DurLAR Exemplar Environment
- City -

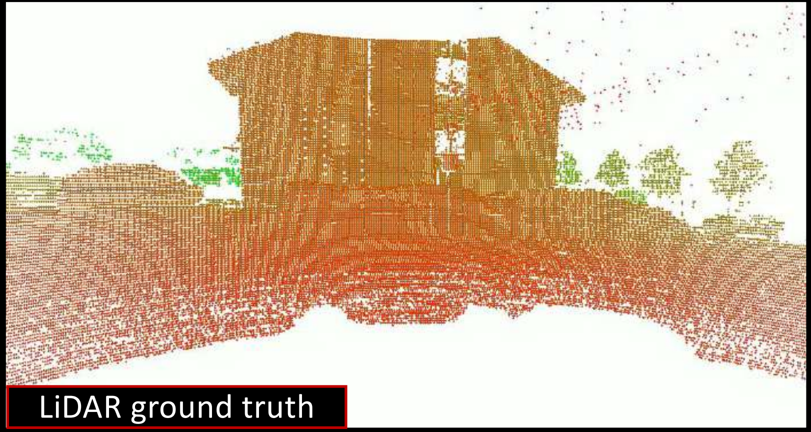


DurLAR Exemplar Environment

- Campus -

Left

Right

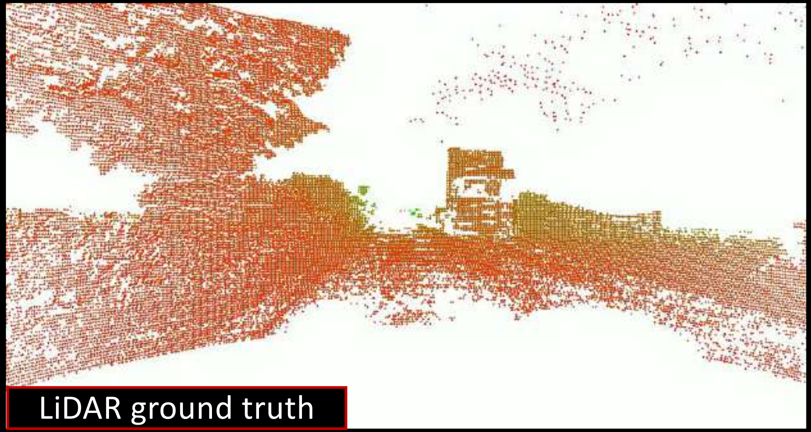


LiDAR ground truth

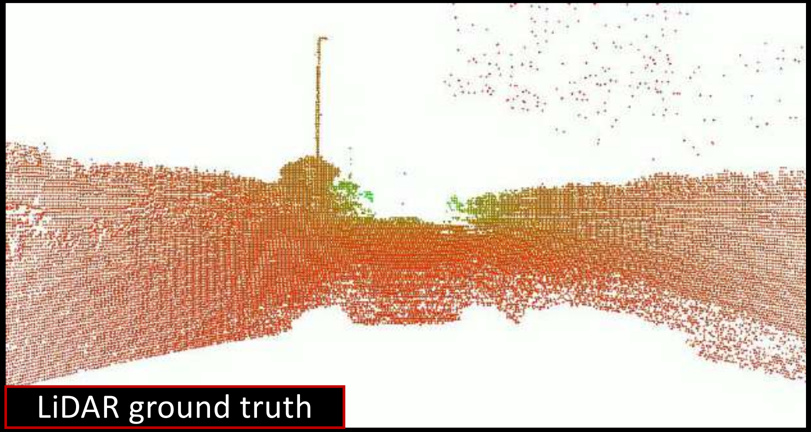
ambient

reflectivity

DurLAR Exemplar Environment
- Highway -

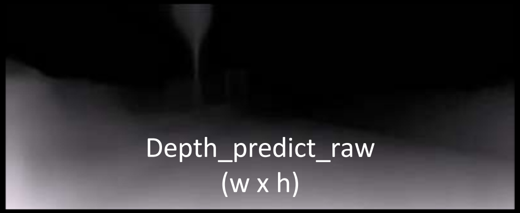


DurLAR Exemplar Environment
- Suburban -



Benchmark Task: Monocular Depth Estimation

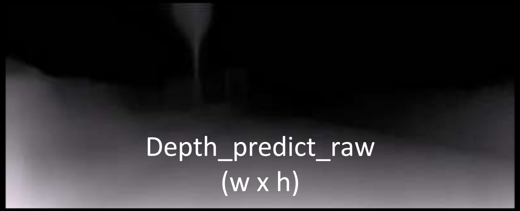
Supervised/self-supervised ManyDepth



Benchmark Task: Monocular Depth Estimation

Supervised/self-supervised ManyDepth

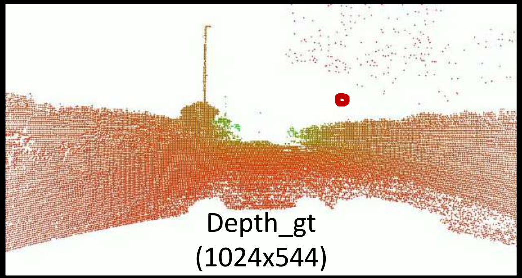
Update weights



Point cloud projection

$$\mathcal{L}_{\mathrm{Berhu}}(d, d^*) = \begin{cases} |d - d^*| & \text{if } |d - d^*| \le \delta \\ \dfrac{(d - d^*)^2 + \delta^2}{2\delta} & \text{if } |d - d^*| > \delta \end{cases}$$

Interpolate to match the size

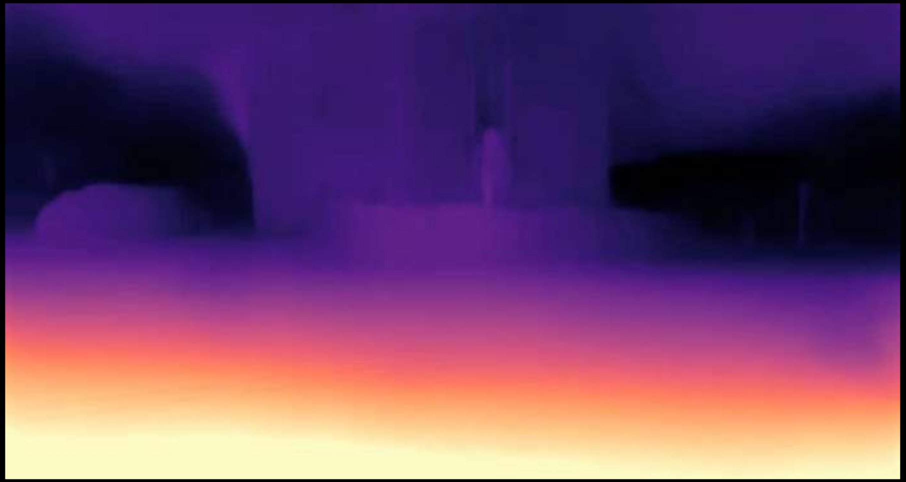


Compute per-pixel errors

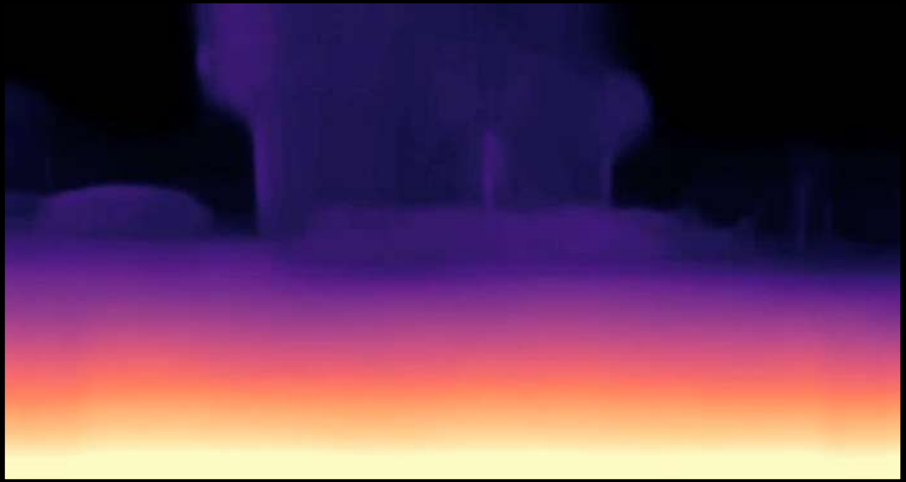
Size must be the same

Qualitative Evaluation

Monocular Depth Estimation

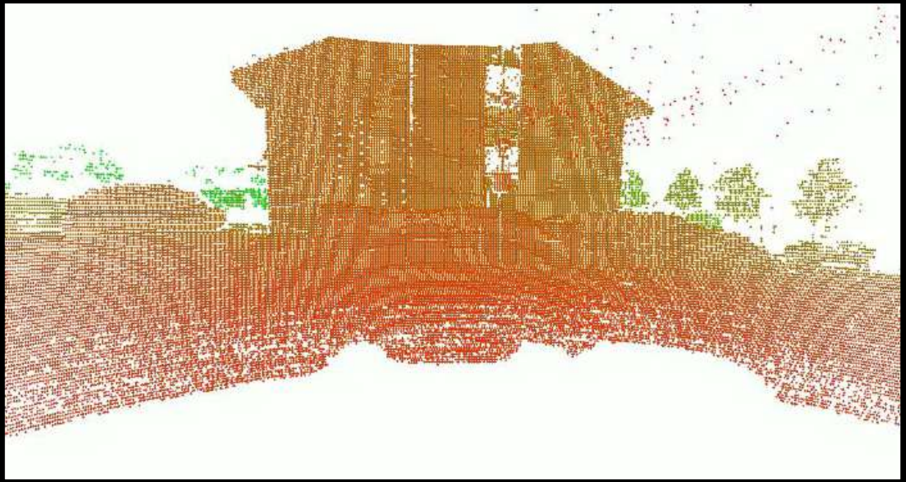


ManyDepth [Watson *et al.*, 2021]



ours

RGB camera frame



LiDAR depth ground truth

Qualitative Evaluation

Monocular Depth Estimation

	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
ManyDepth	0.109	1.111	3.875	0.177	0.901	0.966	0.984
Ours	**0.104**	**0.936**	**3.639**	**0.171**	**0.906**	**0.969**	**0.986**

ManyDepth
[Watson *et al.*, 2021]

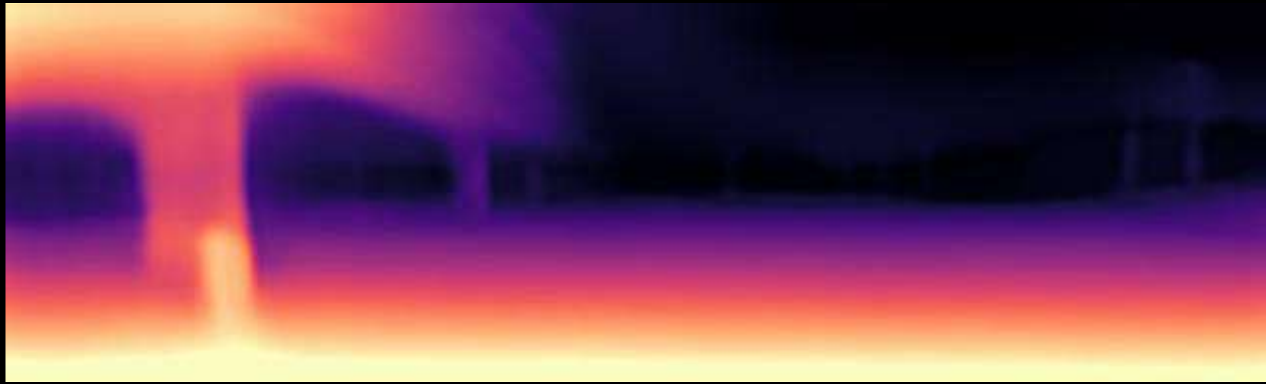
Ours

Qualitative Evaluation

Monocular Depth Estimation

	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
ManyDepth	0.109	1.111	3.875	0.177	0.901	0.966	0.984
Ours	**0.104**	**0.936**	**3.639**	**0.171**	**0.906**	**0.969**	**0.986**

ManyDepth
[Watson *et al.*, 2021]



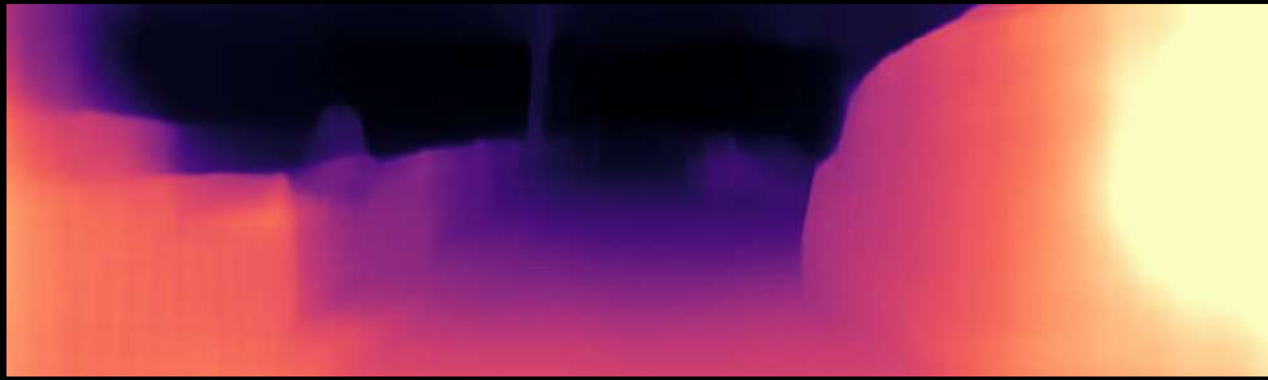
Ours

Qualitative Evaluation

Monocular Depth Estimation

	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
ManyDepth	0.109	1.111	3.875	0.177	0.901	0.966	0.984
Ours	**0.104**	**0.936**	**3.639**	**0.171**	**0.906**	**0.969**	**0.986**

ManyDepth
[Watson *et al.*, 2021]



Ours

Qualitative Evaluation

Monocular Depth Estimation

	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
ManyDepth	0.109	1.111	3.875	0.177	0.901	0.966	0.984
Ours	**0.104**	**0.936**	**3.639**	**0.171**	**0.906**	**0.969**	**0.986**

ManyDepth
[Watson *et al.*, 2021]

Ours

Quantitative Evaluation

Monocular Depth Estimation

Dataset	Method	+S	WxH	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
KITTI	ManyDepth (MR)	x	640x192	0.098	0.770	4.459	0.176	0.900	0.965	0.983
	ManyDepth (MR)	x	1024x320	0.093	0.715	4.245	0.172	0.909	0.966	0.983
Cityscapes	ManyDepth	x	416x128	0.114	1.193	6.223	0.170	0.875	0.967	0.989
DurLAR	Depth-hints	x	640x192	0.122	1.070	4.148	0.211	0.870	0.946	0.972
	Depth-hints	√	640x192	0.121	1.109	4.121	0.210	0.874	0.946	0.972
	MonoDepth2	x	640x192	0.111	1.114	4.002	0.187	0.895	0.960	0.981
	MonoDepth2	√	640x192	0.108	1.010	3.804	0.185	0.898	0.963	0.982
	ManyDepth (MR)	x	640x192	0.115	1.227	4.116	0.186	0.892	0.962	0.982
	ManyDepth (MR)	√	640x192	0.109	0.936	3.711	0.176	0.895	0.964	0.984
	ManyDepth (HR)	x	1024x320	0.109	1.111	3.875	0.177	0.901	0.966	0.984
	ManyDepth (HR)	√	1024x320	0.104	0.936	3.639	0.171	0.906	0.969	0.986

+S=√ Supervised/self-supervised ManyDepth

+S=x self-supervised ManyDepth

Quantitative Evaluation

Monocular Depth Estimation

Dataset	Method	+S	WxH	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
KITTI	ManyDepth (MR)	x	640x192	0.098	0.770	4.459	0.176	0.900	0.965	0.983
	ManyDepth (MR)	x	1024x320	0.093	0.715	4.245	0.172	0.909	0.966	0.983
Cityscapes	ManyDepth	x	416x128	0.114	1.193	6.223	0.170	0.875	0.967	0.989
DurLAR	Depth-hints	x	640x192	0.122	1.070	4.148	0.211	0.870	0.946	0.972
	Depth-hints	√	640x192	0.121	1.109	4.121	0.210	0.874	0.946	0.972
	MonoDepth2	x	640x192	0.111	1.114	4.002	0.187	0.895	0.960	0.981
	MonoDepth2	√	640x192	0.108	1.010	3.804	0.185	0.898	0.963	0.982
	ManyDepth (MR)	x	640x192	0.115	1.227	4.116	0.186	0.892	0.962	0.982
	ManyDepth (MR)	√	640x192	0.109	0.936	3.711	0.176	0.895	0.964	0.984
	ManyDepth (HR)	x	1024x320	0.109	1.111	3.875	0.177	0.901	0.966	0.984
	ManyDepth (HR)	√	1024x320	0.104	0.936	3.639	0.171	0.906	0.969	0.986

+S=√ Supervised/self-supervised ManyDepth

+S=x self-supervised ManyDepth

Cross-Dataset Tests

Monocular Depth Estimation - ManyDepth

Config.	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
K	0.159	1.536	5.101	0.244	0.798	0.923	0.963
D	0.189	1.764	5.580	0.264	0.758	0.908	0.959
K+D	0.188	1.941	5.182	0.262	0.769	0.912	0.958
D+K	0.151	1.123	4.744	0.233	0.805	0.927	0.967

Notation	The Training Configuration
K	KITTI only
D	DurLAR only
K+D	KITTI then fine-tuning with DurLAR
D+K	DurLAR then fine-tuning with KITTI

Cross-Dataset Tests

Monocular Depth Estimation - ManyDepth

Config.	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
K	0.159	1.536	5.101	0.244	0.798	0.923	0.963
D	0.189	1.764	5.580	0.264	0.758	0.908	0.959
K+D	0.188	1.941	5.182	0.262	0.769	0.912	0.958
D+K	0.151	1.123	4.744	0.233	0.805	0.927	0.967

Notation	The Training Configuration
K	KITTI only
D	DurLAR only
K+D	KITTI then fine-tuning with DurLAR
D+K	DurLAR then fine-tuning with KITTI

Ablation Results

Monocular Depth Estimation

vRes	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
32/+S	0.115	0.908	3.677	0.179	0.888	0.966	0.985
64/+S	0.107	0.918	3.735	0.175	0.895	0.967	0.986
128/-S	0.109	1.111	3.875	0.177	0.901	0.966	0.984
128/+S	0.104	0.936	3.639	0.171	0.906	0.969	0.986

+S **Supervised/self-supervised ManyDepth**

-S self-supervised ManyDepth

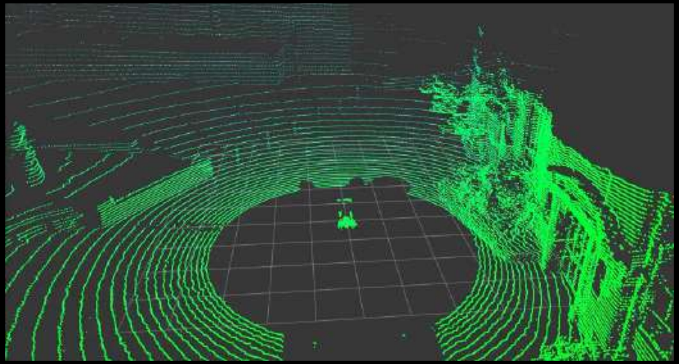
Ablation Results

Monocular Depth Estimation

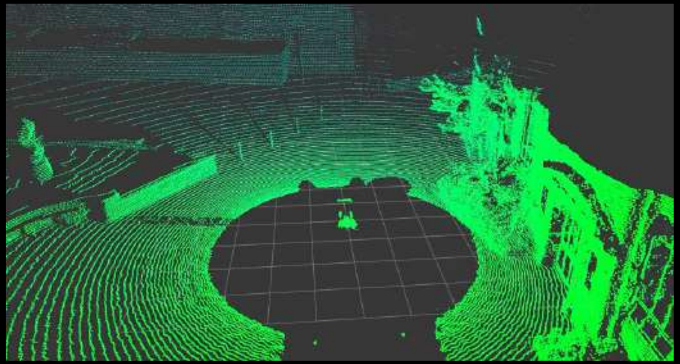
vRes	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
32/+S	0.115	**0.908**	<u>3.677</u>	0.179	0.888	0.966	<u>0.985</u>
64/+S	<u>0.107</u>	<u>0.918</u>	3.735	<u>0.175</u>	0.895	<u>0.967</u>	**0.986**
128/-S	0.109	1.111	3.875	0.177	<u>0.901</u>	0.966	0.984
128/+S	**0.104**	0.936	**3.639**	**0.171**	**0.906**	**0.969**	**0.986**

+S Supervised/self-supervised ManyDepth

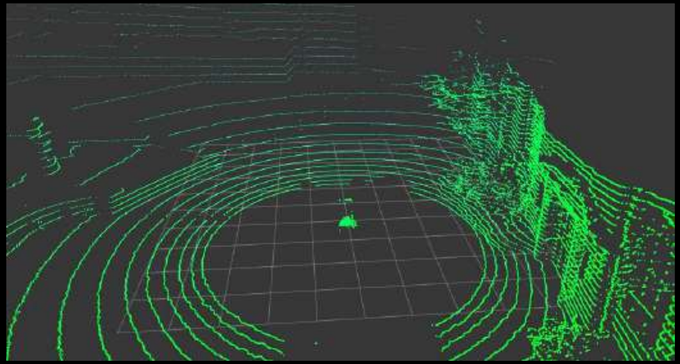
-S self-supervised ManyDepth



64 channels
50%

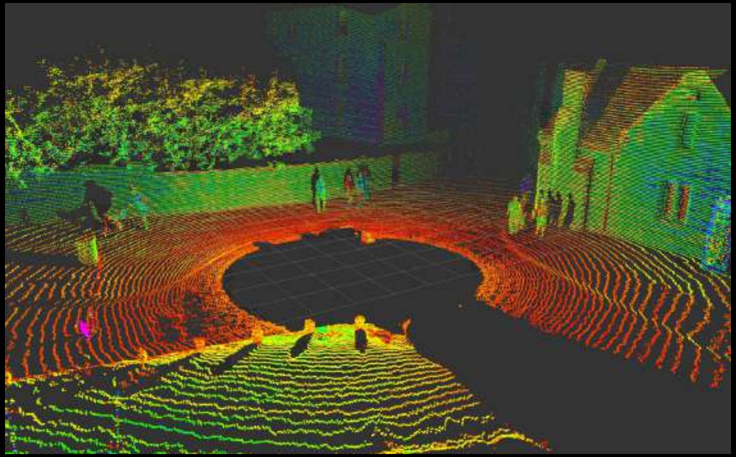


128 channels
DurLAR (ours)

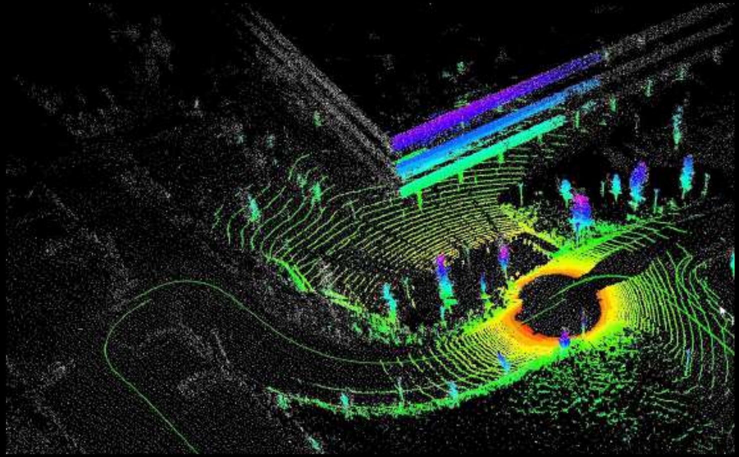


32 channels
75%

Future Applications



reflectivity/ambient



SLAM

Driver Attention Monitoring

Hyundai-NGV Open R&D Seminar

Efficiencies in 3D Environment Understanding for Future Autonomous Driving

Li Li

Department of Computer Science
Durham University

li.li4@durham.ac.uk | https://www.luisli.org