Spotting the Unexpected (STU): A 3D LiDAR Dataset for Anomaly Segmentation in Autonomous Driving

Alexey Nekrasov[1] Malcolm Burdorf[1] Stewart Worrall[2]
Bastian Leibe[1] Julie Stephany Berrio Perez[2]

[1]RWTH Aachen University, [2]The University of Sydney

THE UNIVERSITY OF SYDNEY

Visual Computing Institute
Computer Vision
Prof. Dr. Bastian Leibe

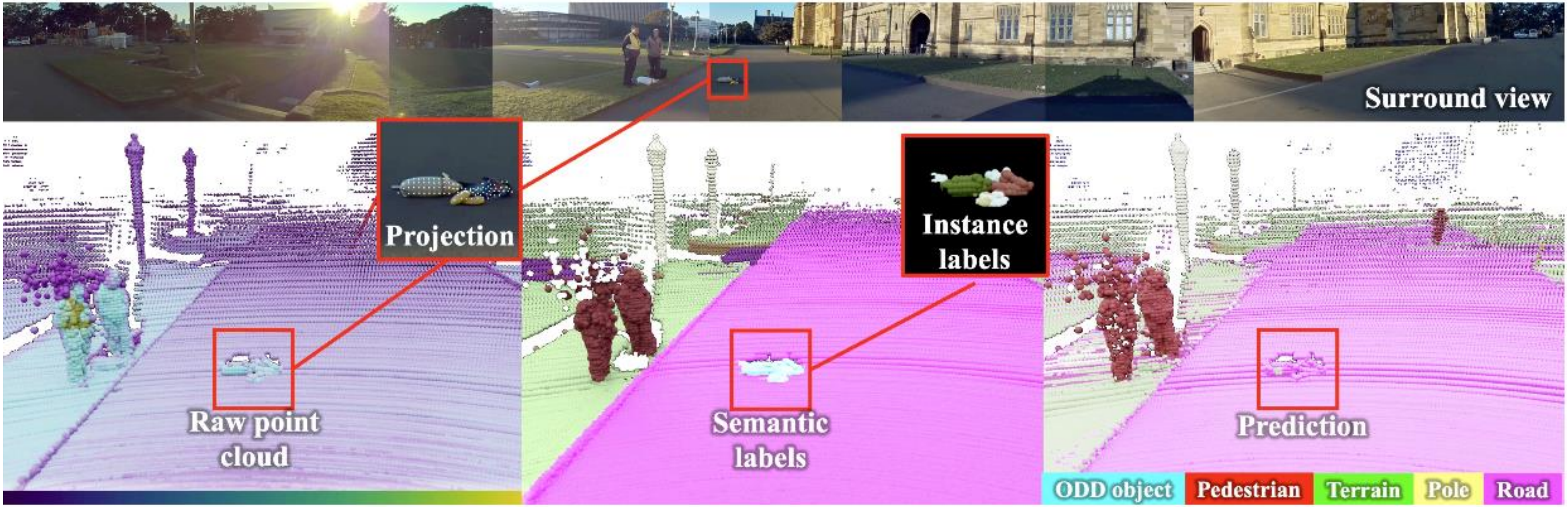
RWTH AACHEN UNIVERSITY

Challenges of Autonomous Driving

[1] YouTube: @AIAddict
[2] YouTube: @MarkRober
[3] YouTube: @followglobalnetwork
[4] YotTube: @newsflare

Related Works

Dataset	Camera	LiDAR Beams	Source	ODD Classes	Label	Size Test/Val	Temporal	Environments Test Sequences	Related Dataset
Fishyscapes									
FS Lost and Found	1 RBG	✗	Staged	1	SM	275/100	✗	n/a	Lost and Found
FS Static	1 RGB	✗	Augmentated	1	SM	1000/30	✗	n/a	CityScapes
CAOS									
StreetHazards	1 RGB	✗	Simulated	1	SM	1500	✓	2 sim towns - 150 seq	
BDD-Anomaly	1 RGB	✗	Validation Set	3	SM	810	✗	n/a	BDD100K
SegmentMeIfYouCan									
RoadAnomaly21	1 RGB	✗	Web Sourced	1	SM	100/10	✗	n/a	
RoadObstacle21	1 RGB	✗	Staged	1	SM	382/30	✓	9 streets - 90 seq	
CODA									
CODA-KITTI	1 RGB	✓- 64	Validation Set	6	BB	309	✗	n/a	KITTI
CODA-nuScenes	1 RGB	✓- 32	Validation Set	17	BB	134	✗	n/a	NuScenes
CODA-ONCE	1 RGB	✓- 40	Validation Set	32	BB	1057	✗	n/a	ONCE
CODA22-ONCE	1 RGB	✓- 40	Validation Set	29	BB	717	✗	n/a	ONCE
CODA22-SODA10M	1 RGB	✗	Validation Set	29	BB	4167	✗	n/a	SODA
Wuppertal ODD									
SOS	1 RGB	✗- dist	Staged	13	IM	1129	✓	9 streets - 20 seq	
CWL	1 RGB	✗- dist	Simulated	18	IM	1210	✓	26 streets - 26 seq	
Lost and Found	1 Stereo	✗	Staged	42	SM	1068	✓	5 streets - 5 seq	
SOD	1 RGB	✓- 16	Staged	1	SM	460/530	✗	2 streets - 5 seq	
WD-Pascal	1 RGB	✗	Simulated	1	SM	70	✗	n/a	
Vistas-NP	1 RGB	✗	Class exclusion	4	SM	11167	✗	n/a	Mapillary Vistas
STU (Ours)	**8 RGB**	**✓- 128**	**Staged**	**1**	**IM**	**8022/1960**	✓	**6 streets - 51 seq**	SemKITTI, P-CUDAL



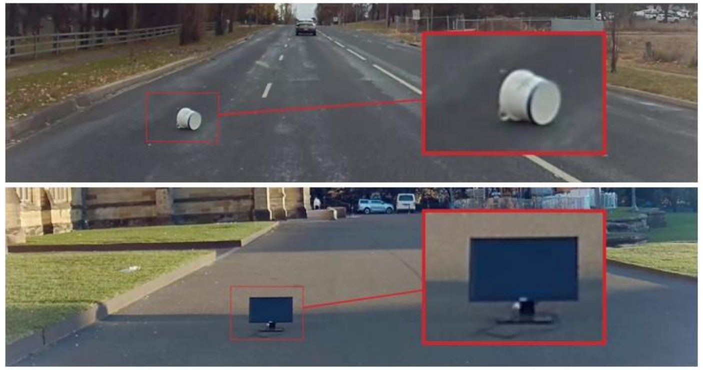
Spotting the Unexpected (STU): A 3D LiDAR Dataset for Anomaly Segmentation in Autonomous Driving

THE UNIVERSITY OF SYDNEY

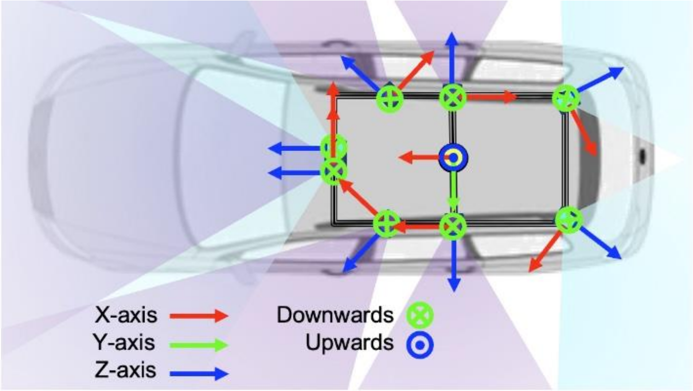
Visual Computing Institute
Computer Vision
Prof. Dr. Bastian Leibe

RWTH AACHEN UNIVERSITY

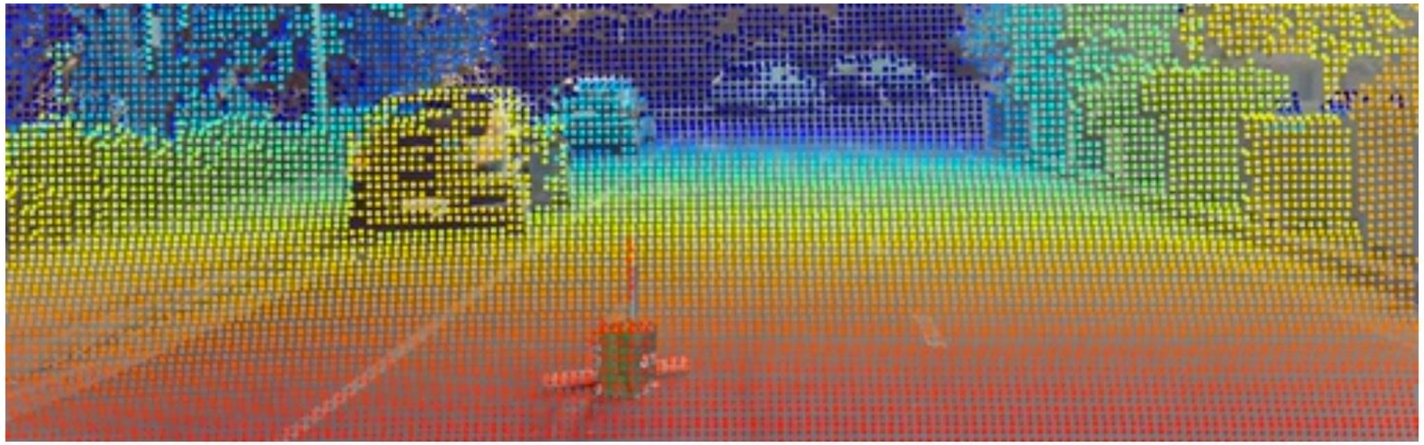
Data Capture



Data collected in both naturalistic (real-world) and controlled (staged) environments.

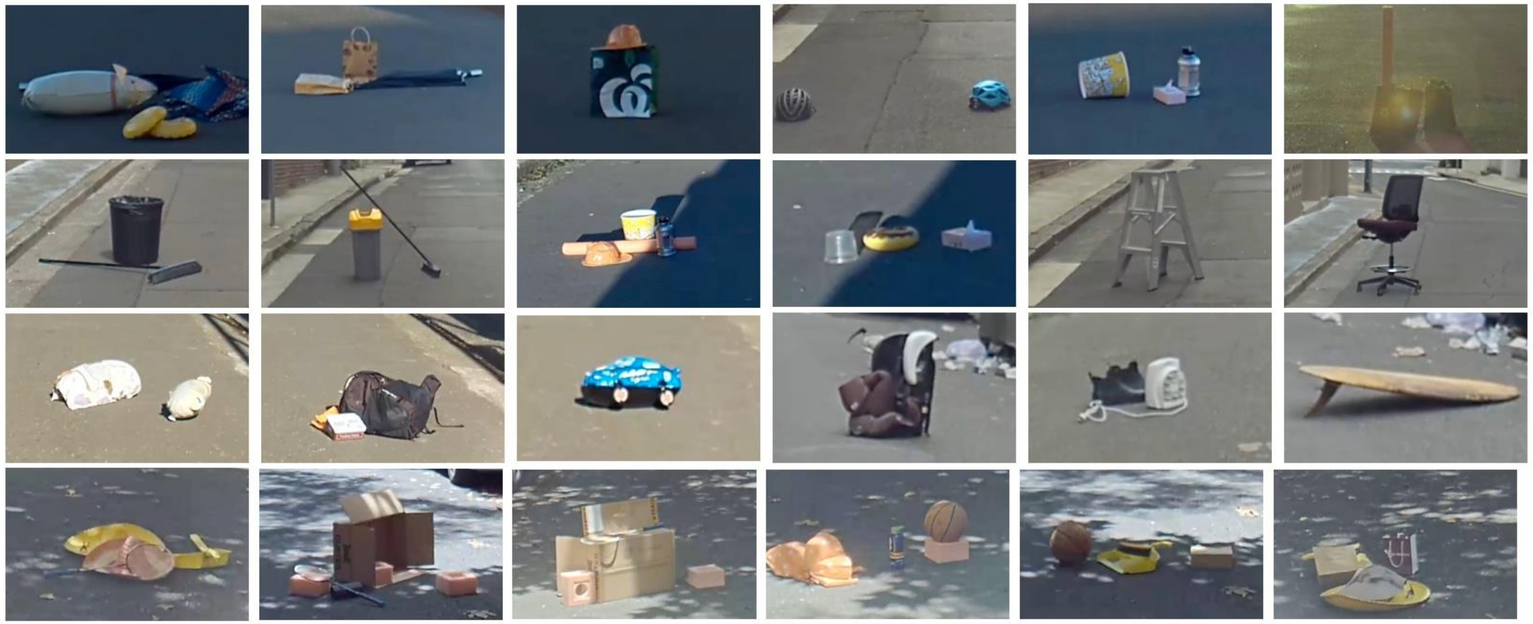


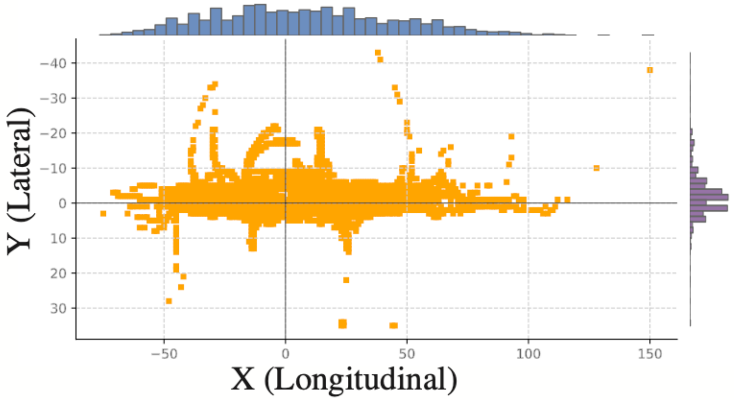
The field of view for the 60-degree and 120-degree cameras.



LiDAR-camera projection. Point clouds are colorized by distance to the camera.

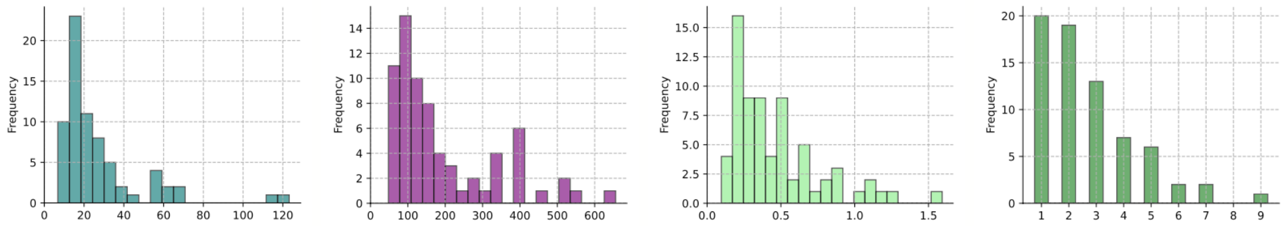
Data Collection



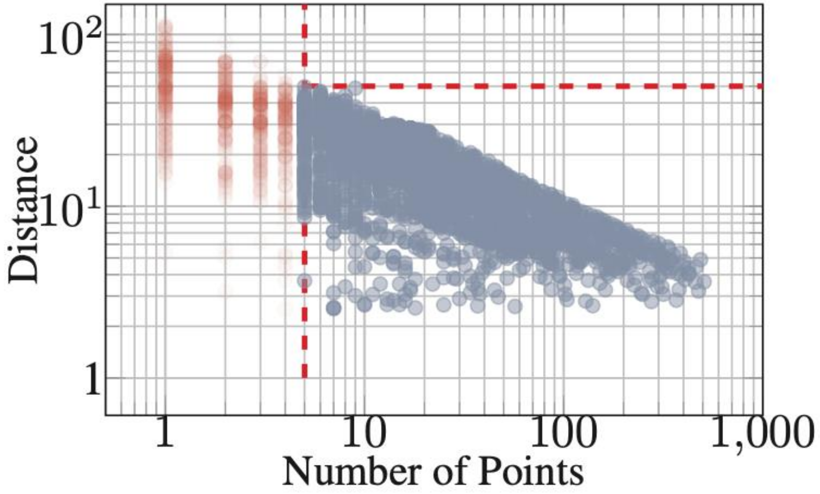


THE UNIVERSITY OF SYDNEY
Visual Computing Institute
Computer Vision
Prof. Dr. Bastian Leibe
RWTH AACHEN UNIVERSITY

Data Collection



A typical anomaly object has on average less then 50 points per sequence (a), with less then 300 (b) points at maximum, and a maximum height below one meter (c). We record up to nine individual anomaly instances in the same sequence (d).



Baselines and Results

Method	OoD Data	Point-Level OOD			Object-Level OOD				
		AUROC ↑	FPR@95 ↓	AP ↑	RecallQ	SQ	RQ	UQ	PQ
Deep Ensemble	✗	**86.74**	**58.05**	**5.17**	16.75	84.49	**10.43**	14.16	**8.81**
MC Dropout	✗	61.51	82.37	0.11	2.25	**86.72**	1.95	2.14	1.86
Max Logit	✗	84.53	81.49	0.95	**26.14**	83.06	2.13	**21.71**	1.77
Void Classifier	✓	85.99	78.60	3.92	17.64	84.40	8.19	14.89	6.91
RbA	✗	66.38	100.0	0.81	24.04	83.28	3.23	20.02	2.69

THE UNIVERSITY OF SYDNEY
Visual Computing Institute
Computer Vision
Prof. Dr. Bastian Leibe
RWTH AACHEN UNIVERSITY

Metrics

Object Metrics We adopt Panoptic Quality for a class c:

$$\mathrm{PQ}_c = \underbrace{\frac{\sum_{(p,g)\in TP_c} \mathrm{IoU}(p,g)}{|TP_c|}}_{\text{Segmentation Quality (SQ)}} \times \underbrace{\frac{|TP_c|}{|TP_c| + \frac{1}{2}|FP_c| + \frac{1}{2}|FN_c|}}_{\text{Recognition Quality (RQ)}}$$

To evaluate the recall of anomaly objects, we adopt the Unknown Quality (UQ) metric:

$$\mathrm{UQ} = \underbrace{\frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP|}}_{\text{Segmentation Quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + |FN|}}_{\text{Recall Quality (RecallQ)}} .$$

The UQ metric does not penalize FP and allows for the evaluation of object recall.

Point-Level Metrics. Similar to 2D anomaly segmentation, we evaluate the performance of the anomaly segmentation at the point level. We use the common metrics:

- Average Precision (**AP**);
- False-Positive Rate at 95% True-Positive Rate (**FPR@95**);
- Area Under Receiver Operating characteristic Curve (**AUROC**).

Points marked as "ignore" are excluded prior to evaluation. Erroneous predictions in ignored regions are not penalized.

Method	SemanticKITTI			STU-inlier Dataset		
	PQ	SQ	RQ	PQ	SQ	RQ
Mask-PLS*	59.9	76.4	69.1	39.6	67.0	50.5
Mask-PLS	55.6	76.0	64.5	50.9	69.4	62.2
Mask4Former-3D*	**61.9**	81.3	71.6	42.8	70.6	53.2
Mask4Former-3D	60.7	81.0	70.4	**52.7**	73.8	64.7

We introduce a few labeled scenes with no anomalies for joint training, and show small domain gap to SemanticKITTI.

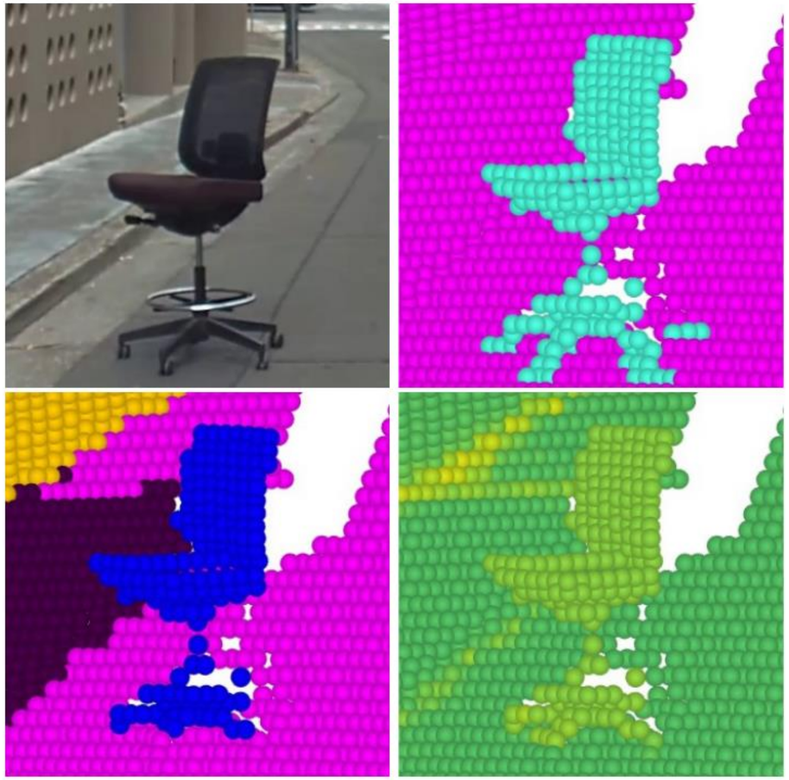


Figure 7. Example of a failure case anomaly segmentation. For the chair on the road, labeled as anomaly in upper left, model predicts "other-vehicle" ■ class, in lower left, with a high certainty, as indicated by MaxLogit scores on the bottom right.

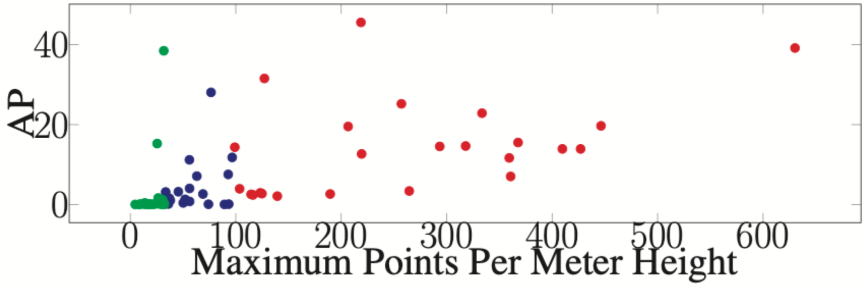
Small objects in the distance

Method	0–10m	10–20m	20–30m	30–40m	40–50m
Deep Ensemble	7.63	8.49	3.42	0.38	0.03
MC Dropout	0.16	0.53	0.06	0.04	0.01
Max Logit	2.25	1.53	1.20	0.27	0.01
Void Classifier	2.95	1.78	1.98	0.28	0.03
RbA	1.85	1.28	0.73	0.15	0.01

Performance of baseline methods at different distances. At distances above 30 meters, performance is poor.

Size	Deep Ensemble AP
● Large	14.35
● Medium	4.30
● Small	2.50

Mean AP for the Category.



Deep Ensemble performance degrades with increased distance and smaller anomaly size.

2D Evaluations on Validation Set

Method	Aux Data	AUROC↑	FPR@95↓	AP↑
DenseHybrid	✗	87.09	76.36	26.63
RbA	✗	89.58	75.21	37.12
UNO	✓	89.52	62.29	37.10
Mask2Anomaly	✓	90.54	78.09	36.38

We investigated 2D anomaly segmentation models in the front-view of the validation set. Domain shift increases false-positive rate.

Conclusion

Spotting the Unexpected (STU): A 3D LiDAR Dataset for Anomaly Segmentation in Autonomous Driving

Alexey Nekrasov[1] Malcolm Burdorf[1] Stewart Worrall[2] Bastian Leibe[1] Julie Stephany Berrio Perez[2]
[1] RWTH Aachen University, [2] The University of Sydney