Understanding the Challenges When 3D Semantic Segmentation Faces Class Imbalanced and OOD Data

Yancheng Pan <panyancheng@pku.edu.cn>, ***Huijing Zhao** <zhaohj@pku.edu.cn>

POSS-Lab, http://www.poss.pku.edu.cn/

School of Intelligence Science and Technology, Peking Universi

An Outline of the Semantic Segmentation Research at POSS Lab

- **New methods**
  - Semantic Segmentation of 3D LiDAR Data in Dynamic Scene Using Semi-Supervised Learning, T.ITS2020
  - Incorporating Human Domain Knowledge in 3D LiDAR-based Semantic Segmentation, T.IV2020
  - Scene-Adaptive Off-Road Detection Using a Monocular Camera, T.ITS 2018
  - Off-Road Drivable Area Extraction Using 3D LiDAR Data, IV2019
  - Fine-Grained Off-Road Semantic Segmentation and Mapping via Contrastive Learning, IROS2021
  - An Active and Contrastive Learning Framework for Fine-Grained Off-Road Semantic Segmentation, arXiv2022
- **Survey and analysis**
  - Are We Hungry for 3D LiDAR Data for Semantic Segmentation? A Survey of Datasets and Methods, T.ITS2021
  - **Understanding the Challenges When 3D Semantic Segmentation Faces Class Imbalanced and OOD Data, arXiv2022**
- **New dataset**
  - **SemanticPOSS** , IV2020

The leading students of these works!


Jilin Mei

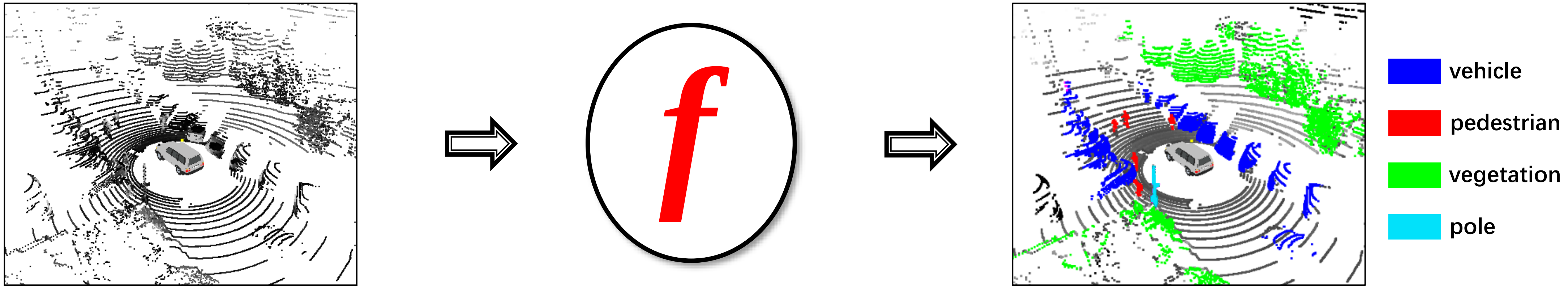

Biao Gao


Yancheng Pan

SemanticPOSS <http: //www.poss.pku.edu.cn/semanticposs.html>

3D Semantic Segmentation (3DSS)

Problem Formulation



3D Point Cloud

3DSS Model

Point-wise Semantic Labels

Applications: a key technique for a mobile agent to traverse at complex environments

Deep Learning methods have been the focus of the studies in solving the problem.

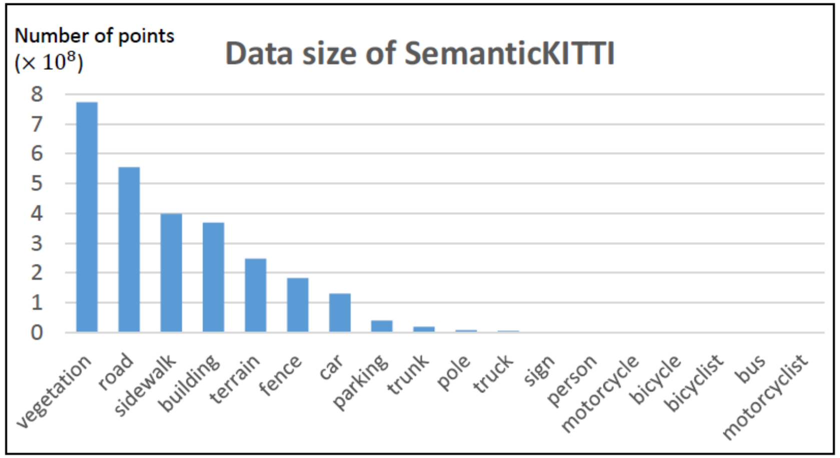
Challenges of Deep Learning-based 3DSS

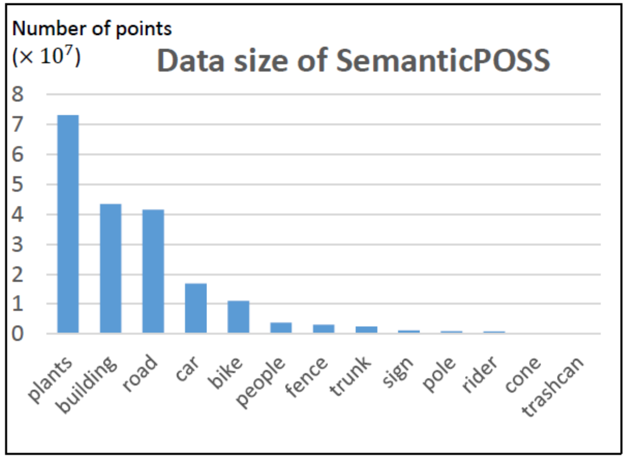
Deep learning methods are mostly data-driven

- **Data hunger problem**
  - Even severe for 3DSS task!

(Are We Hungry for 3D LiDAR Data for Semantic Segmentation? A Survey of Datasets and Methods? T.ITS2021)

- **Class-imbalanced (Long-tailed) data**
  - Real world is class-imbalanced!
- **Out-of-distribution (OOD) data**
  - The open world problem!
- **Aware its unsureness**
  - A key issue when deploying an AI system to safety-critical applications!
  - **Trust scoring** by thresholding on e.g. softmax confidence, ODIN etc..





When 3DSS Face Class-Imbalanced and OOD Data

Questions:

- How does the class imbalance problem affect 3DSS model performance?
- Can 3DSS model be aware of its unsureness?
- ➢ Can it detect whether the category prediction is correct or not, or whether the input sample is an ID or OOD?

A Logical Map of the Challenges and Questions

Datasets

Class Imbalance

- classes of **sufficient** training samples
- classes of **few** training samples

OOD Data

- classes of **no** training samples

3DSS Model

Trust Scoring

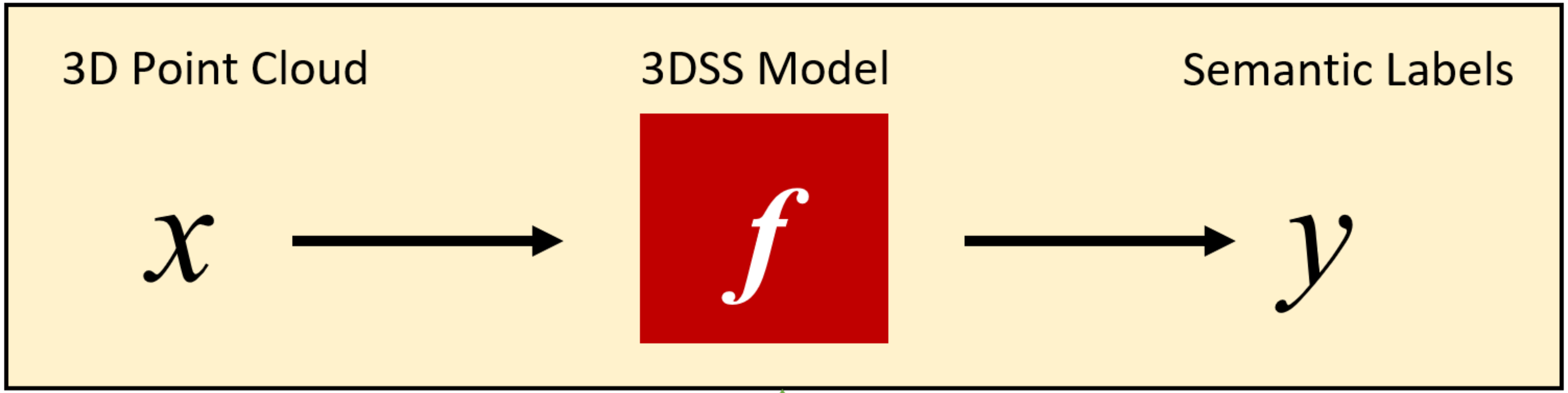
How does the class imbalance problem affect 3DSS models' performance?

Can the model be aware of its unsureness?

- whether the category prediction is correct or not?
- whether the input sample is an ID or OOD?

Experiment 1

Q: How class-imbalance problem affect model performance?



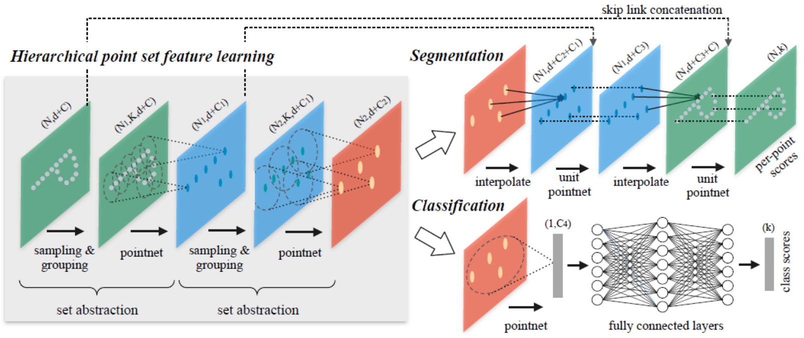
Train/Test: SemanticKITTI
3DSS Models: PointNet++, Cylinder3D, RandLA-Net

3DSS Models

Traditional 3DSS model

PointNet++

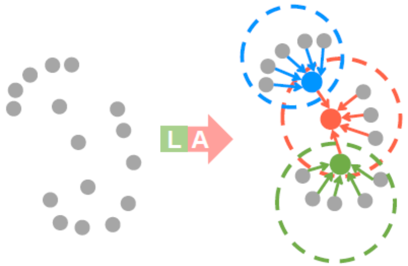
Point-based method



State-of-the-art 3DSS models

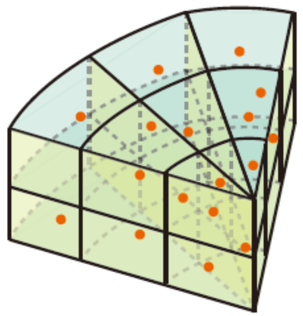
RandLA-Net

Point-based method

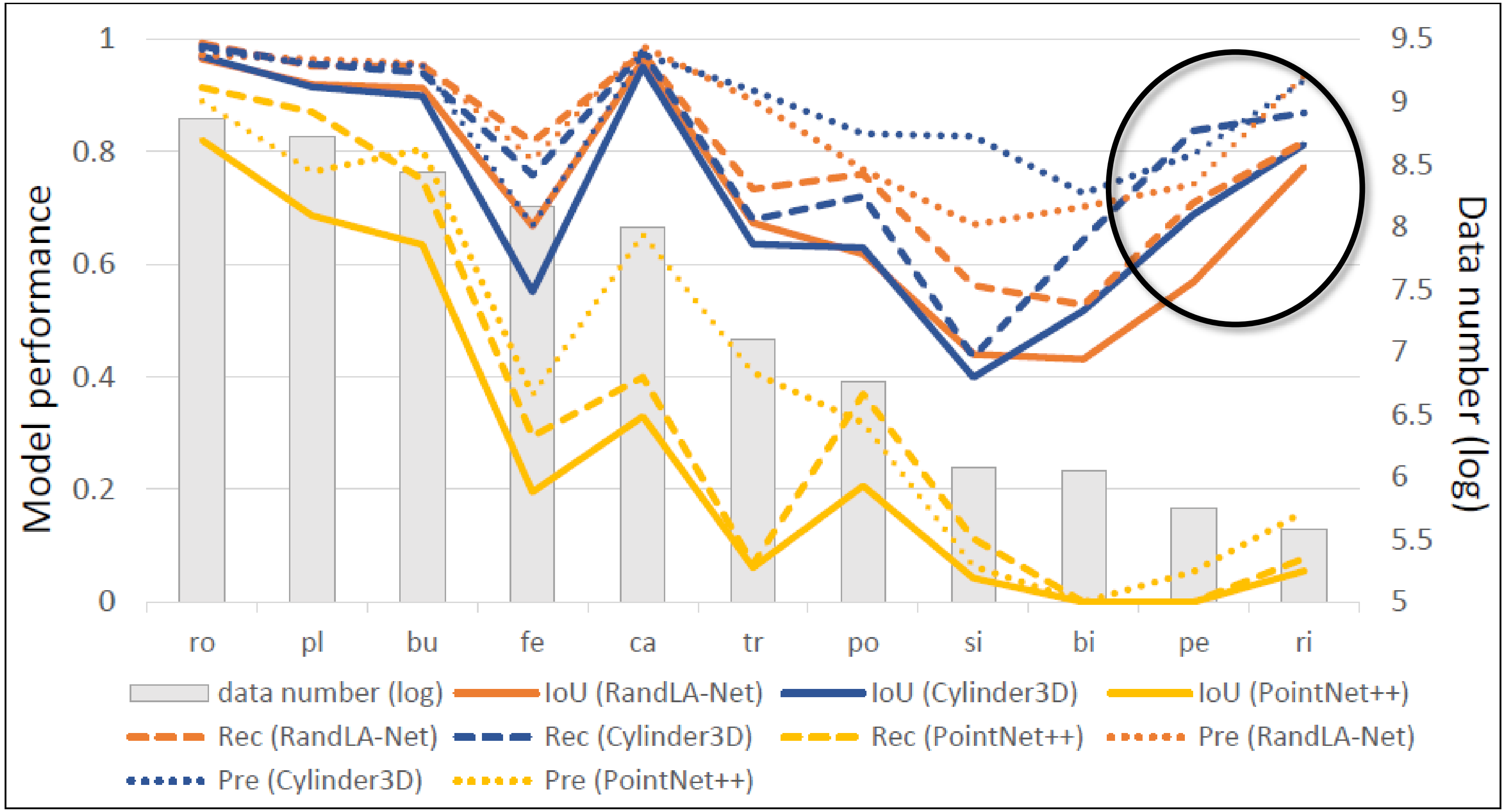


Cylinder3D

Voxel-based method

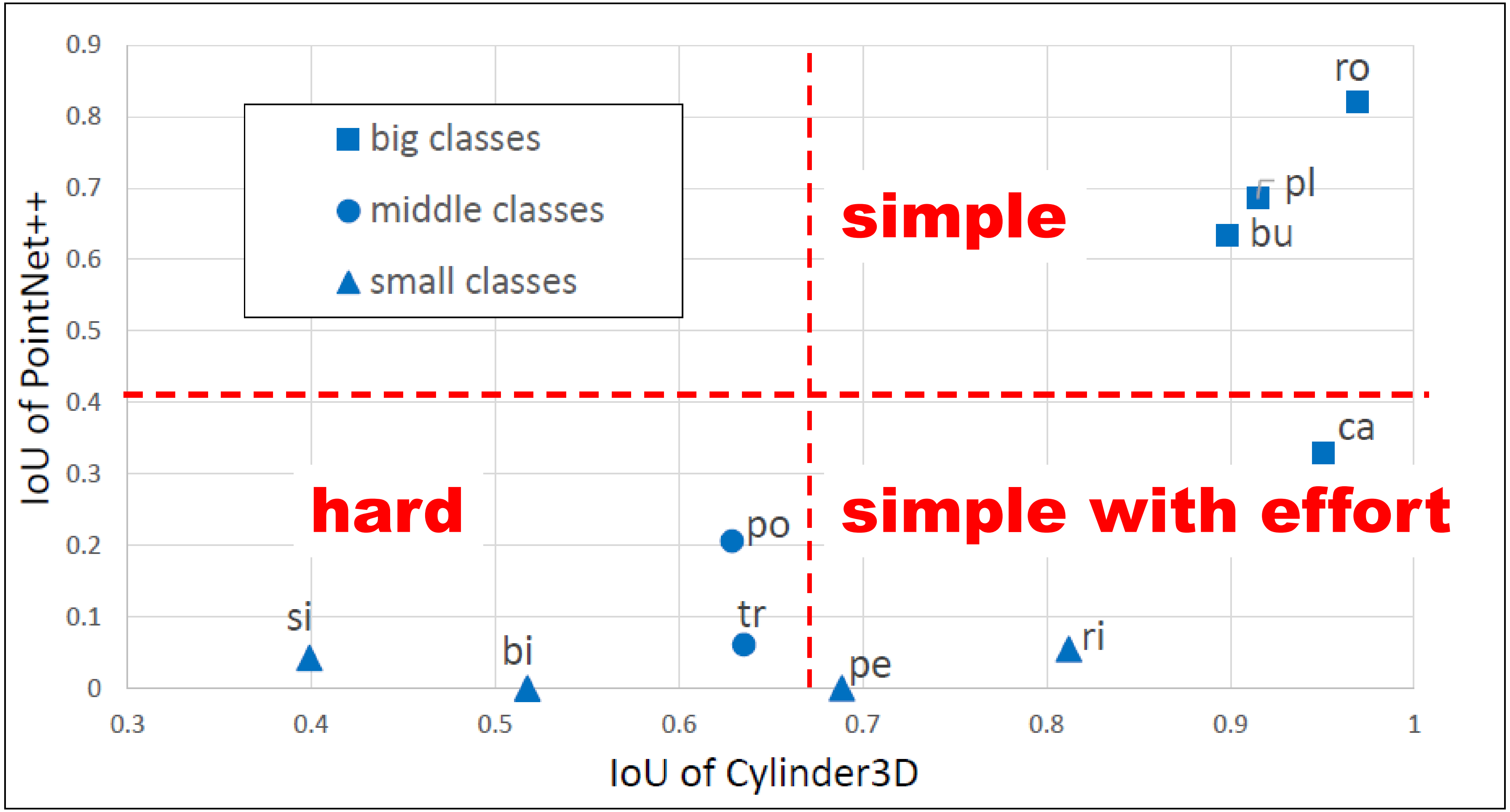


Results of Experiment 1



- The performance of PointNet++ has certain correlation with data size.
- The performance of some small classes has been greatly improved by RandLA-Net and Cylinder3D.

Accuracy Analysis

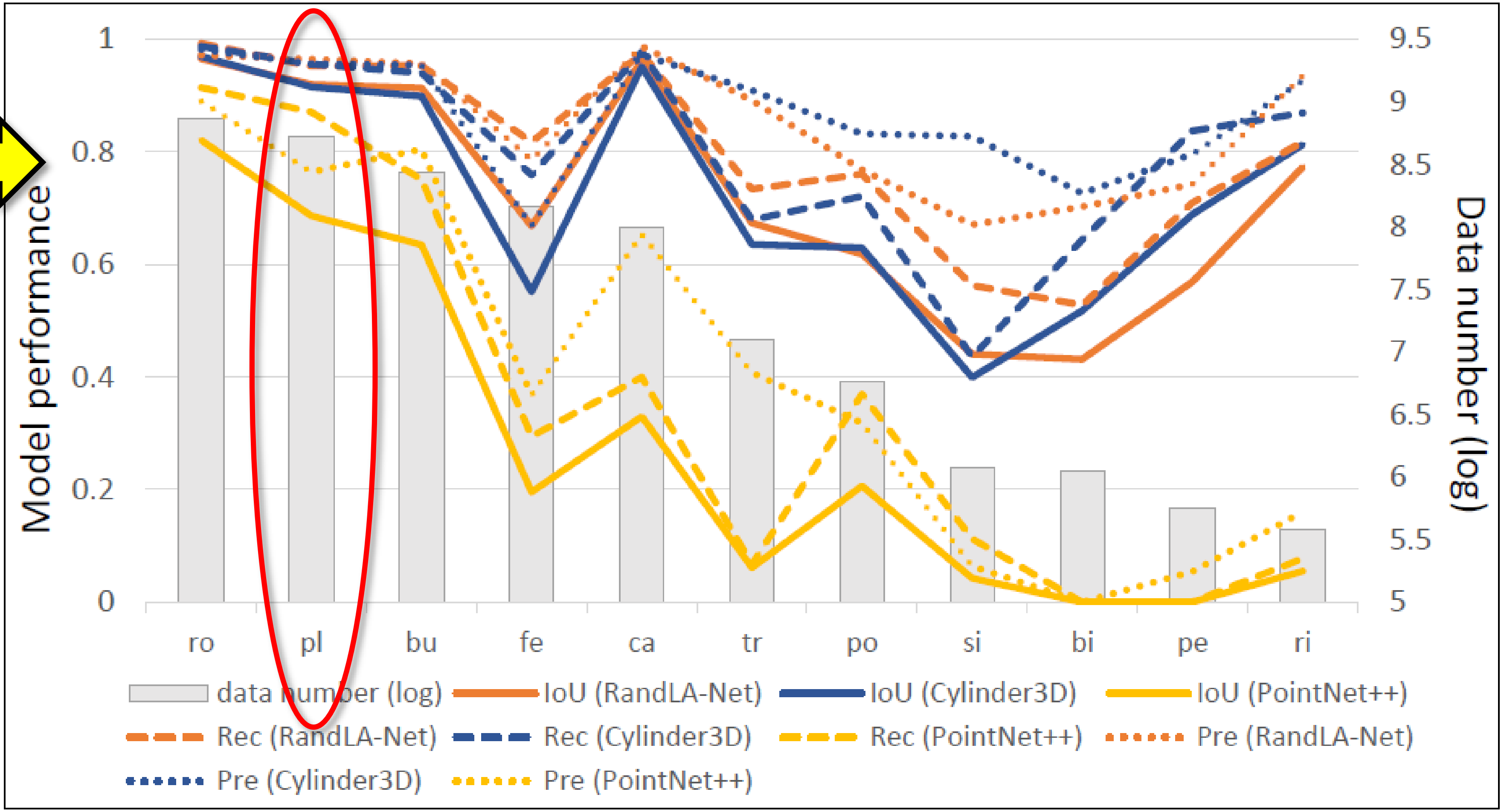


- The performance is not only related with data size.
- The performance of some small classes can be improved, but some are hard.

Confusion Analysis

PD = Plants

PD / GT	pe	ri	ca	si
pe	0.00	0.03	0.11	
ri	0.00	0.08	0.31	0.01
ca	0.00	0.00	0.40	0.00
si	0.00	0.01	0.03	0.11
tr	0.00	0.00	0.06	0.02
pl	0.00	0.00	0.02	0.00
po	0.00	0.01	0.03	0.01
fe	0.00	0.00	0.05	0.00
bu	0.00	0.00	0.00	0.00
bi	0.00	0.01	0.18	0.01
ro	0.00	0.00	0.02	0.00
wPre	0.14	0.57	0.33	0.64

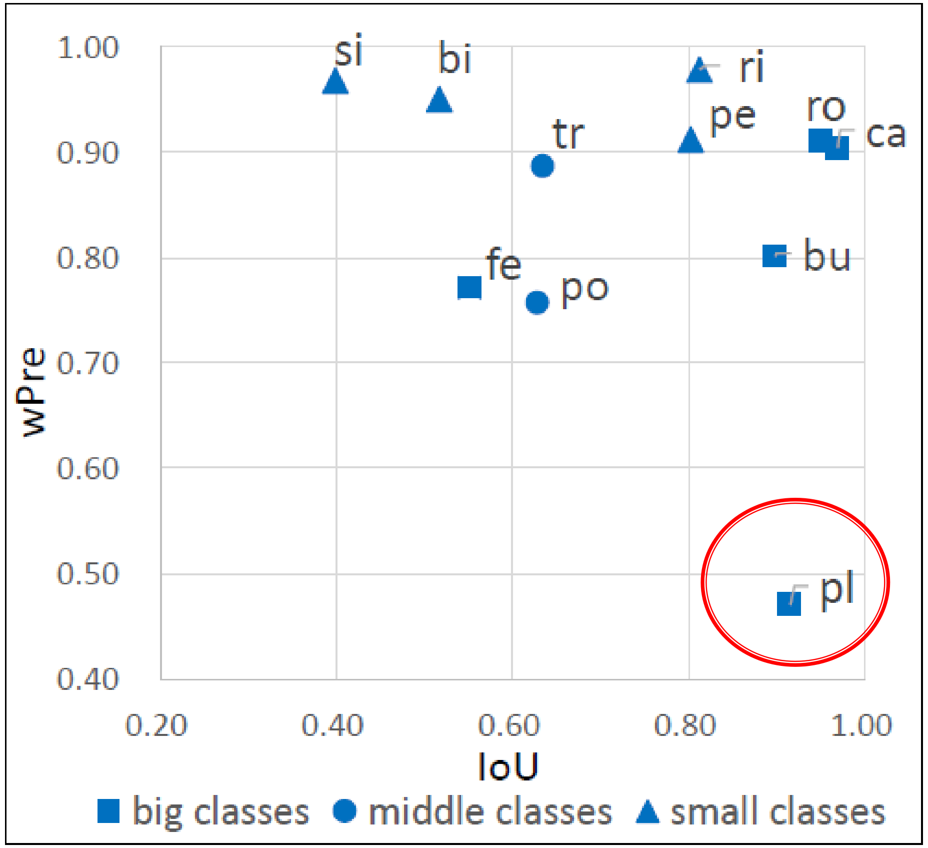


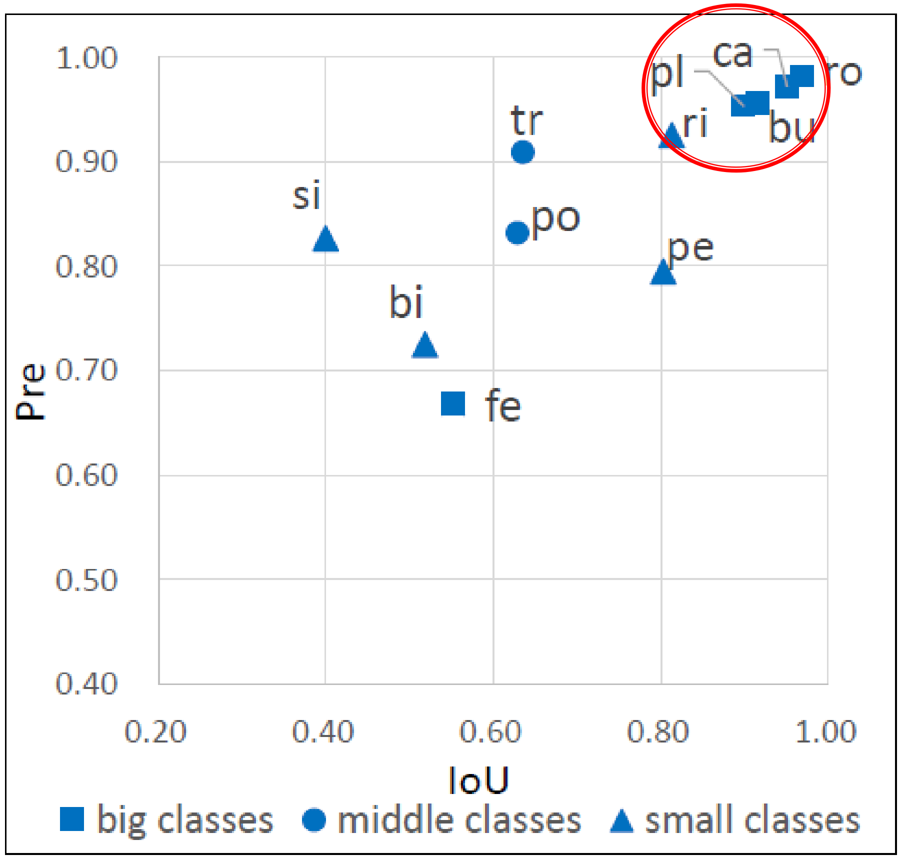
der3D

pl	po	fe	bu	bi	ro	Rec
.07	.00	0.01	0.03	0.01	0.01	0.84
.03	.00	0.00	0.00	0.02	0.01	0.87
.01	.00	0.00	0.00	0.00	0.01	0.91
.19	.22	0.11	0.02	0.00	0.00	0.44
.30	.00	0.00	0.00	0.00	0.01	0.68
.96	.00	0.02	0.00	0.00	0.02	0.96
.12	.72	0.02	0.03	0.00	0.02	0.72
.09	.00	0.76	0.14	0.00	0.01	0.76
.03	.00	0.02	0.94	0.00	0.00	0.94
.24	.00	0.03	0.01	0.64	0.02	0.64
.01	.00	0.00	0.00	0.00	0.99	0.99
.47	0.76	0.77	0.80	0.95	0.90	

- wPre (*weighted Precision*): A new metric to account for imbalanced class size.

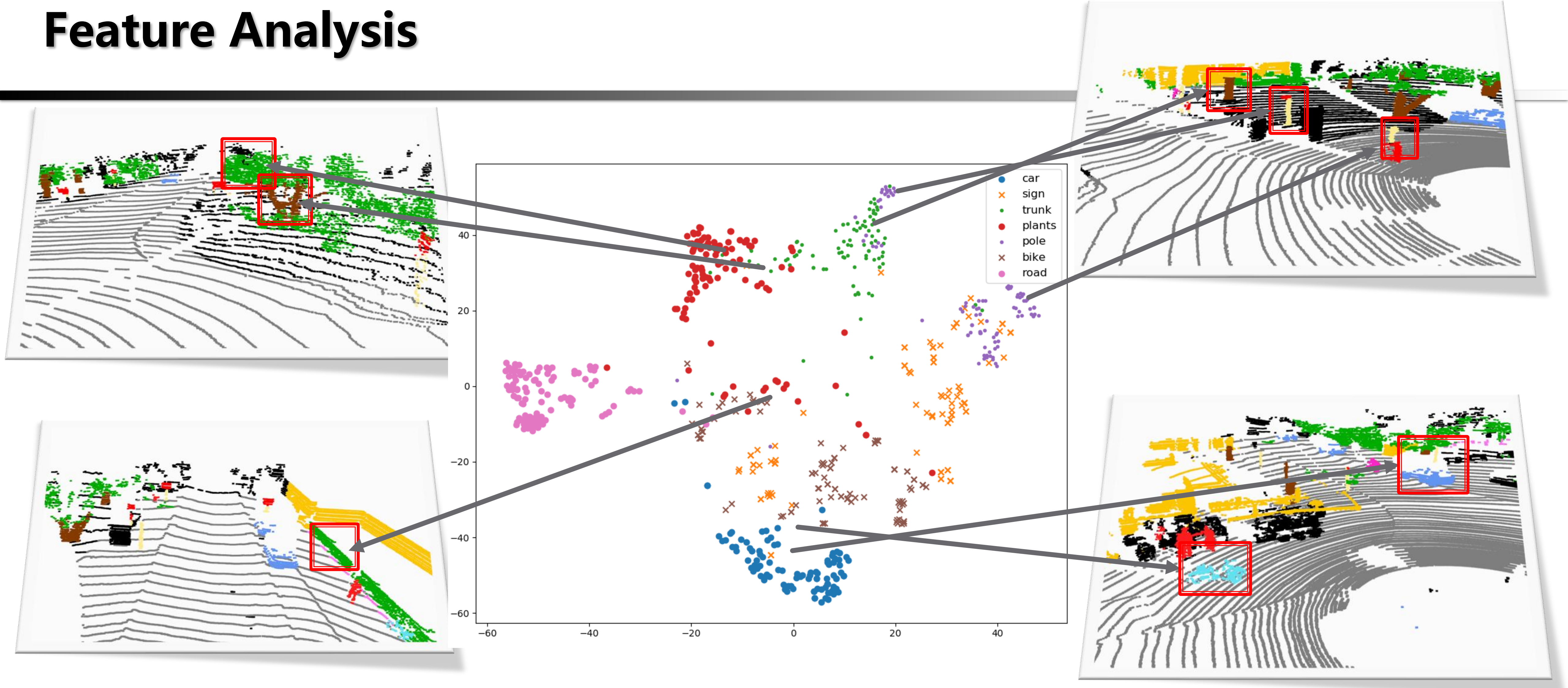
Confusion Analysis





- Plants has high-accuracy but easy to be confused.
- wPre (*weighted Precision*) can evaluate this property by accounting for imbalanced class size.

Feature Analysis



- There are **intra-class diversity** and **inter-class ambiguity**, who are the main reason of confusing.
- The classes are not only imbalanced on data size, but also their nature, who has been less studied in literature.

Experiment 2

Q: Can 3DSS model be aware of its unsurene
OOD data? → Can the model be aware its c

$$z = \begin{cases} 0, \text{if} & g(x) \leq \delta \\ 1, \text{if} & g(x) > \delta \end{cases}$$

3D Point Cloud	3DSS Model	Trust Scoring	I/O; C/W Prediction
X →	f →	g →	Z

Train: SubKITTI; **Test**: AugKITTI
3DSS Models: PointNet++, Cylinder3D, RandLA-Net
Trust Scores: Softmax confidence, Uncertainty, ODIN, MD

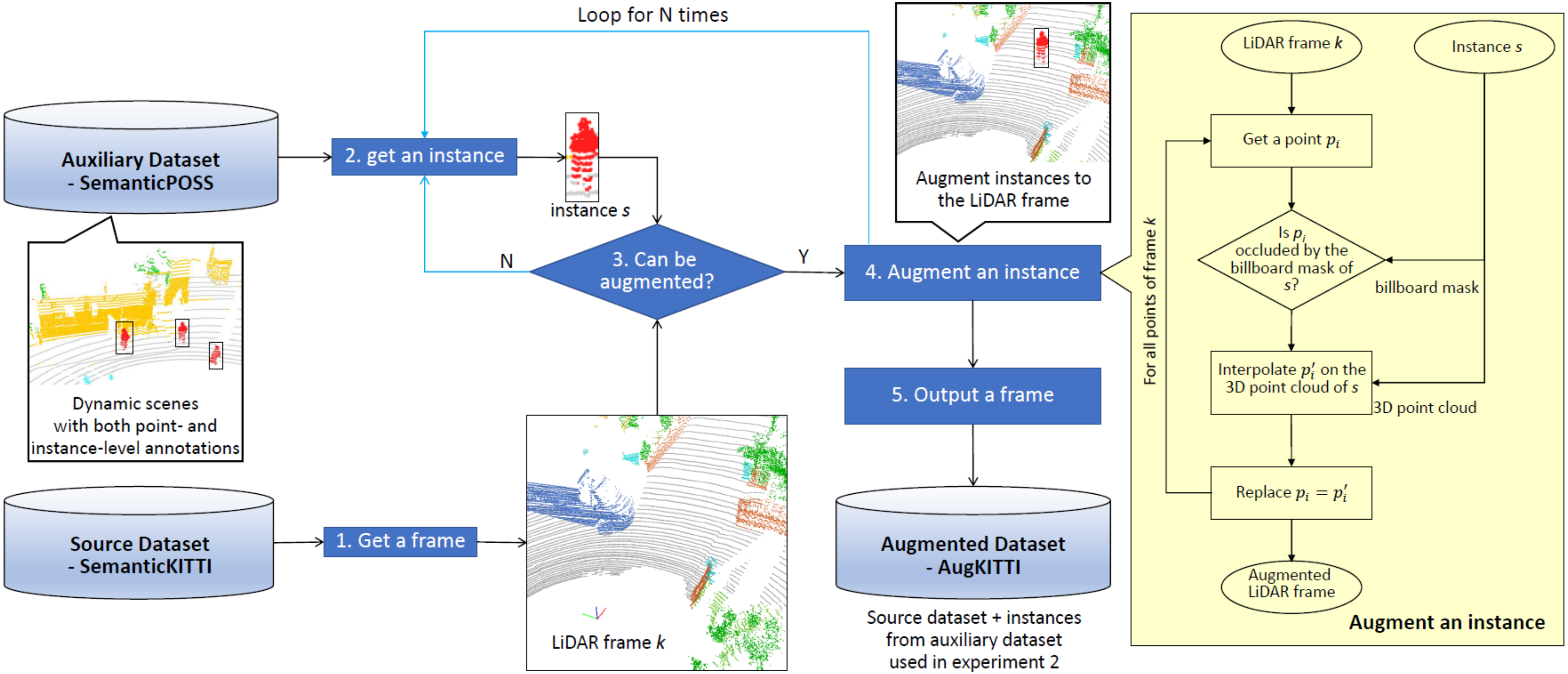
Dataset

- **OOD classes**: people, rider
- **ID classes**: others

- **Train dataset - SubKITTI**
  - SemanticKITTI frames that have no people and rider data.

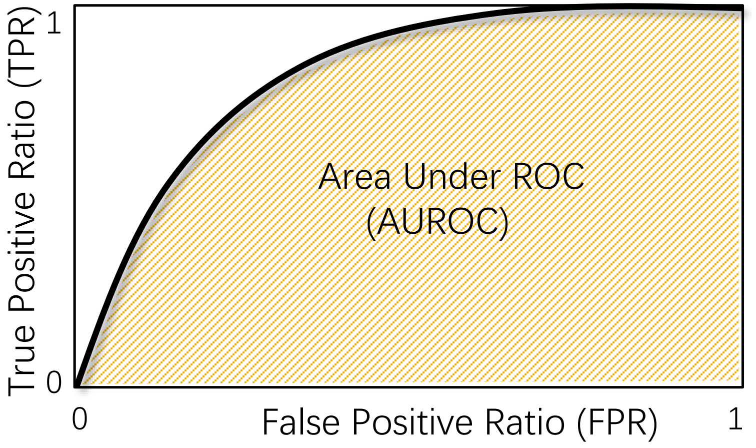
- **Test dataset - AugKITTI**
  - SemanticKITTI frames that are augmented with the people and rider data from SemanticPOSS.

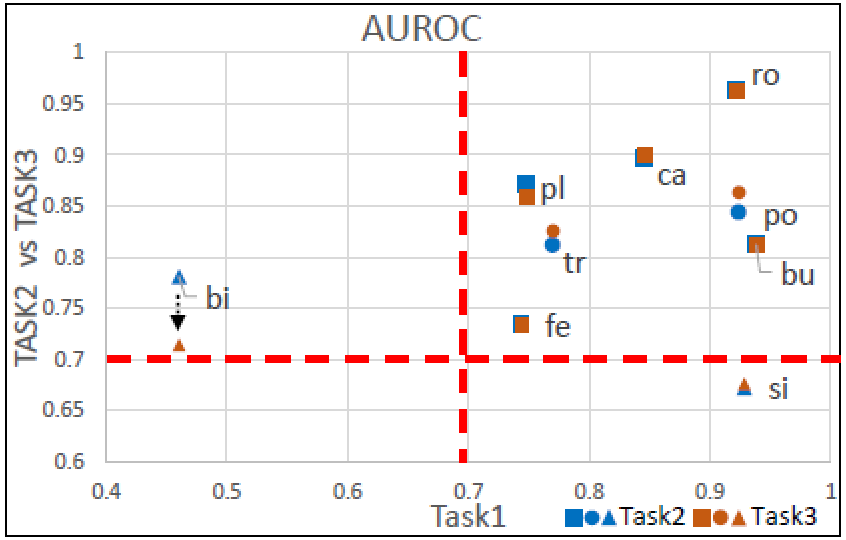
Dataset Augmentation

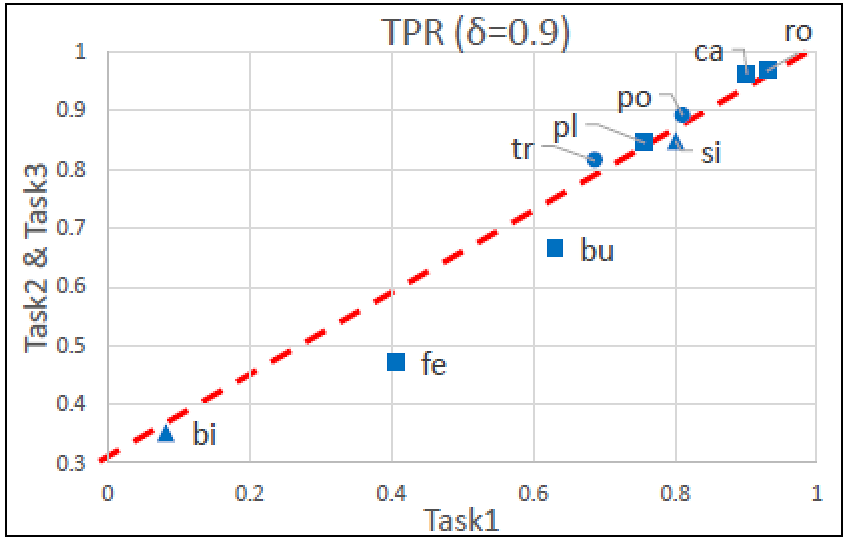


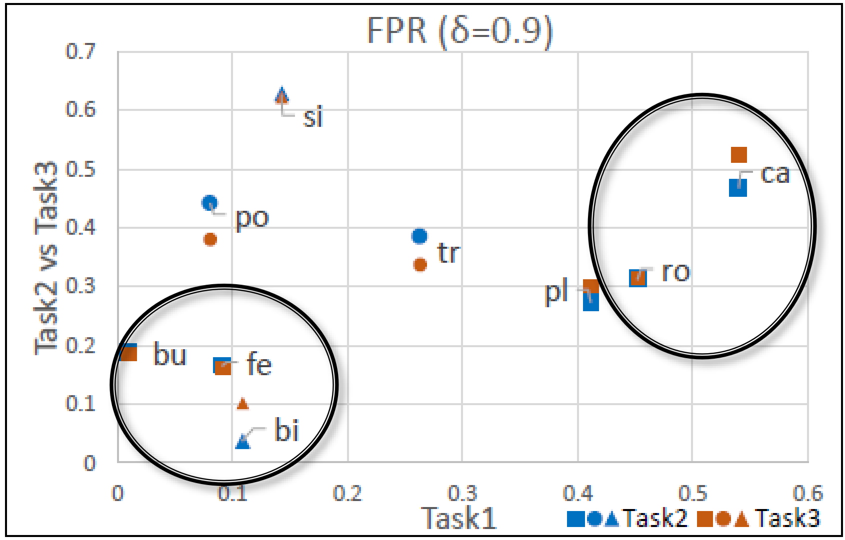
Results of Experiment 2

- **Task1 – I/O**: discriminate whether the data is ID or OOD
- **Task2 – C/W**: discriminate whether the predicted semantic class is Correct or Wrong **without** OOD
- **Task3 – C/W with OOD**: discriminate whether the predicted semantic class is C/W **with** OOD









Trust score: Softmax confidence; 3DSS model: Cylinder3D

Confusion Analysis

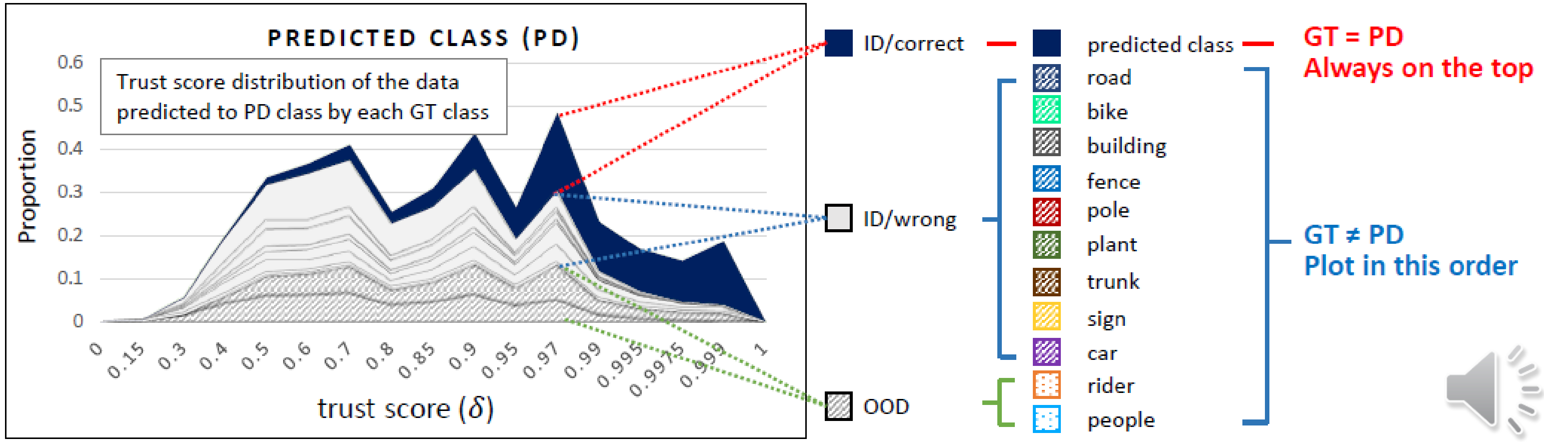
$$p(r,c) = \frac{|y_{GT} = r \wedge y_{PD} = c|}{|y_{GT} = r|}$$

$$q(r,c,\delta_i) = \frac{|y_{GT} = r \wedge y_{PD} = c \wedge \delta_i < g(x) \le \delta_{i+1}|}{|y_{GT} = r|}$$

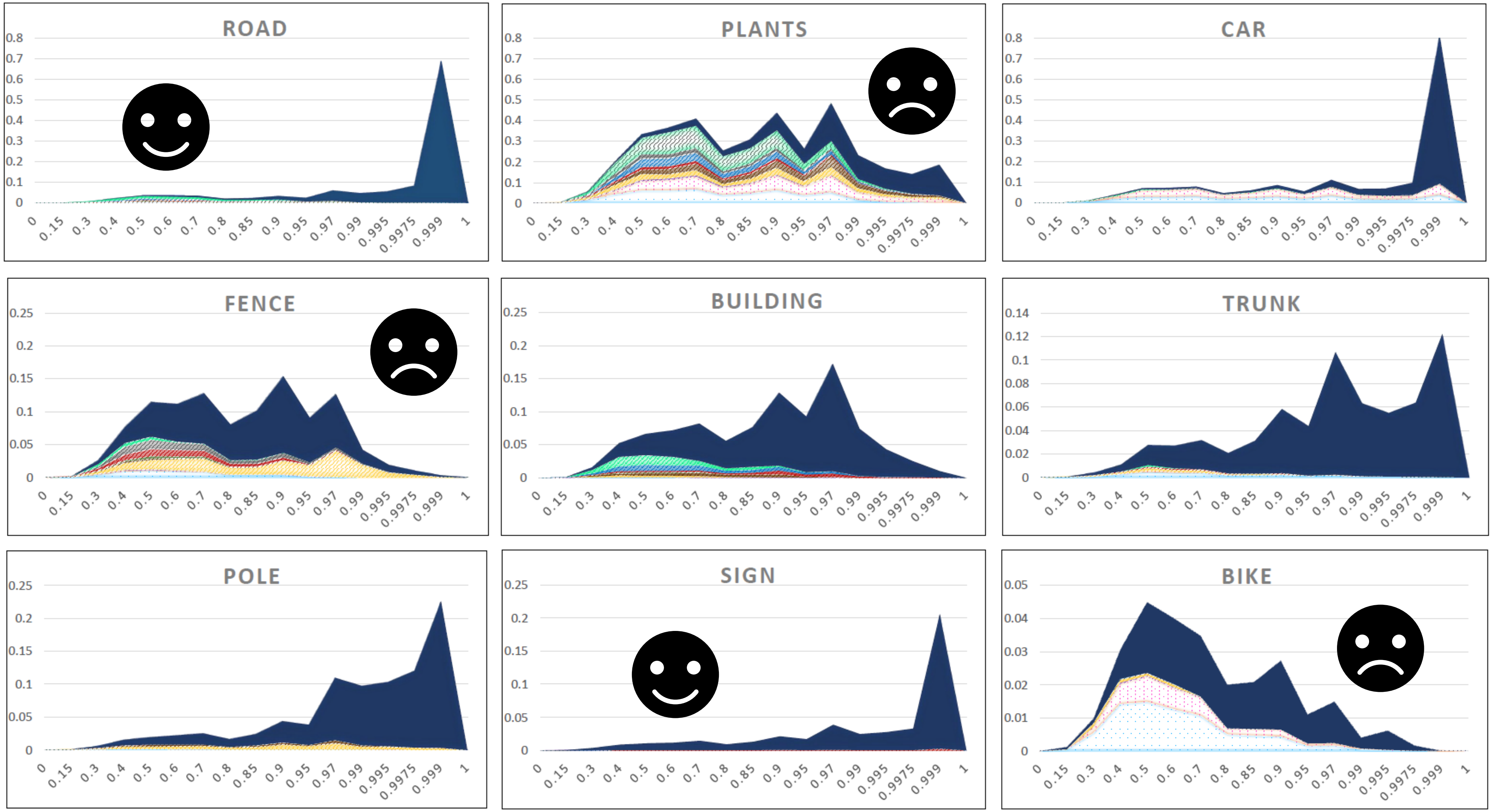
GT \ PD	pe	ri	ca	si	tr	pl	po	fe	bu	bi	ro	Rec
pe	0.84	0.02	0.01	0.00	0.00	0.07	0.00	0.01	0.03	0.01	0.01	0.84
ri	0.06	0.87	0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.02	0.01	0.87
ca	0.00	0.00	0.98	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.91
si	0.01	0.00	0.00	0.44	0.01	0.19	[illegible]	0.11	0.02	0.00	0.00	0.44
tr	0.00	0.00	0.00	0.00	0.68	0.30	0.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
pl	0.00	0.00	0.00	0.00	0.00	0.96	0.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
po	0.00	0.00	0.00	0.01	0.07	0.12	0.72	0.02	0.03	0.00	0.02	0.72
fe	0.00	0.00	0.01	0.00	0.00	0.09	0.00	0.76	0.14	0.00	0.01	0.76
bu	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.02	0.94	0.00	0.00	0.94
bi	0.00	0.00	0.06	0.00	0.00	0.24	0.00	0.03	0.01	0.64	0.02	0.64
ro	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.99	0.99
wPre	0.91	0.98	0.91	0.97	0.89	0.47	0.76	0.77	0.80	0.95	0.90	

PD=pl																	
GT \ δ	0	0.15	0.3	0.4	0.5	0.6	0.7	0.8	0.85	0.9	0.95	0.97	0.99	0.995	0.9975	0.999	p
pe	0.00	0.01	0.14	0.41	0.59	0.60	0.66	0.38	0.45	0.62	0.36	0.50	0.14	0.06	0.02	0.00	4.93
ri	0.00	0.00	0.04	0.19	0.47	0.52	0.63	0.38	0.47	0.71	0.45	0.83	0.38	0.26	0.21	0.21	5.77
ca	0.00	0.00	0.01	0.04	0.06	0.06	0.06	0.04	0.04	0.05	0.03	0.03	0.01	0.00	0.00	0.00	0.43
si	0.00	0.02	0.08	0.23	0.27	0.22	0.24	0.14	0.20	0.32	0.22	0.41	0.18	0.11	0.06	0.07	2.77
tr	0.00	0.00	0.03	0.10	0.22	0.25	0.31	0.19	0.24	0.37	0.26	0.52	0.25	0.17	0.12	0.07	3.10
pl	0.00	0.00	0.01	0.04	0.17	0.22	0.34	0.26	0.41	0.83	0.72	1.81	1.15	0.98	0.94	1.48	9.37
po	0.00	0.00	0.04	0.09	0.13	0.13	0.14	0.08	0.10	0.12	0.07	0.09	0.02	0.01	0.01	0.00	1.03
fe	0.00	0.00	0.04	0.17	0.40	0.38	0.40	0.23	0.26	0.31	0.15	0.18	0.05	0.03	0.01	0.01	2.63
bu	0.00	0.00	0.04	0.15	0.21	0.20	0.21	0.11	0.13	0.15	0.07	0.08	0.03	0.01	0.01	0.01	1.41
bi	0.00	0.00	0.14	0.56	0.81	1.07	1.10	0.74	0.77	0.89	0.32	0.39	0.11	0.05	0.01	0.00	6.96
ro	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04

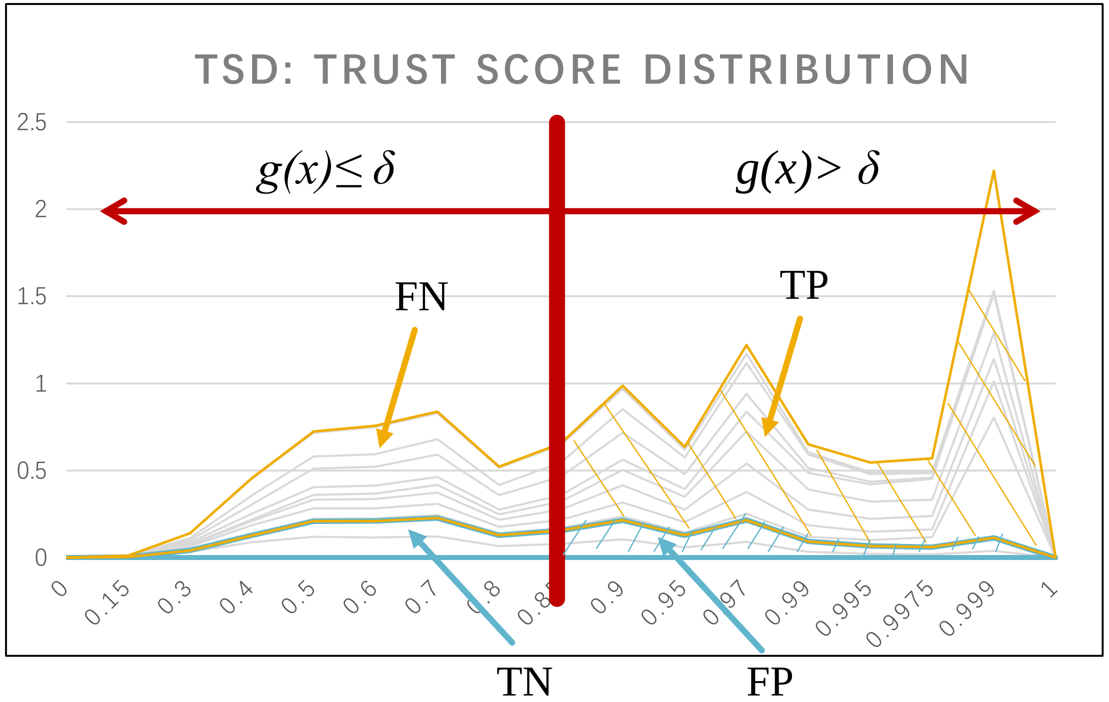
TSD: Trust score distribution

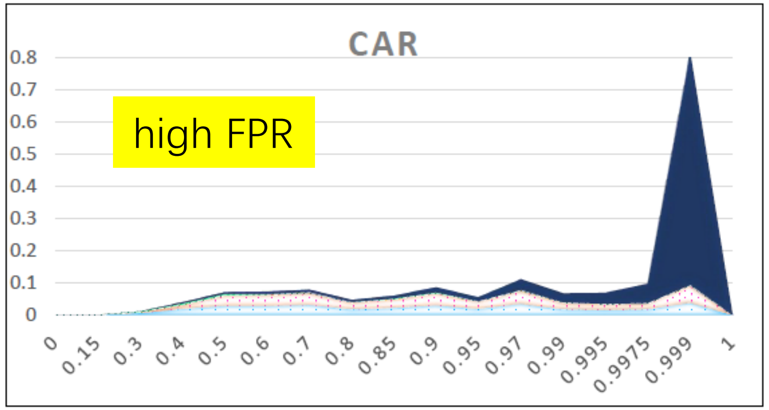


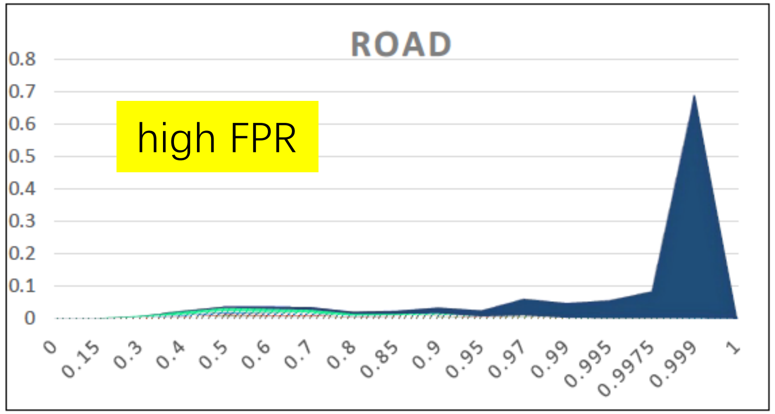
TSD: Trust score distribution



TSD: Trust score distribution





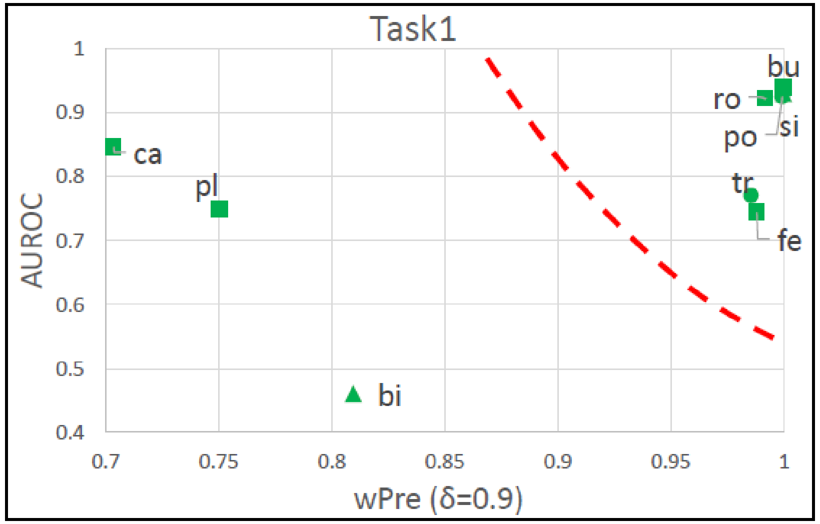


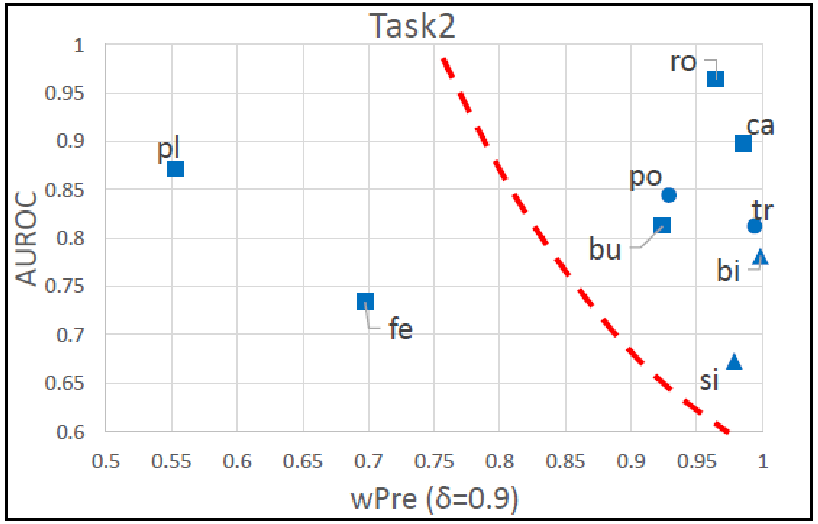
$$\mathrm{TPR}(c,\delta) = \frac{\mathrm{TP}(c,\delta)}{\mathrm{TP}(c,\delta)+\mathrm{FN}(c,\delta)}$$

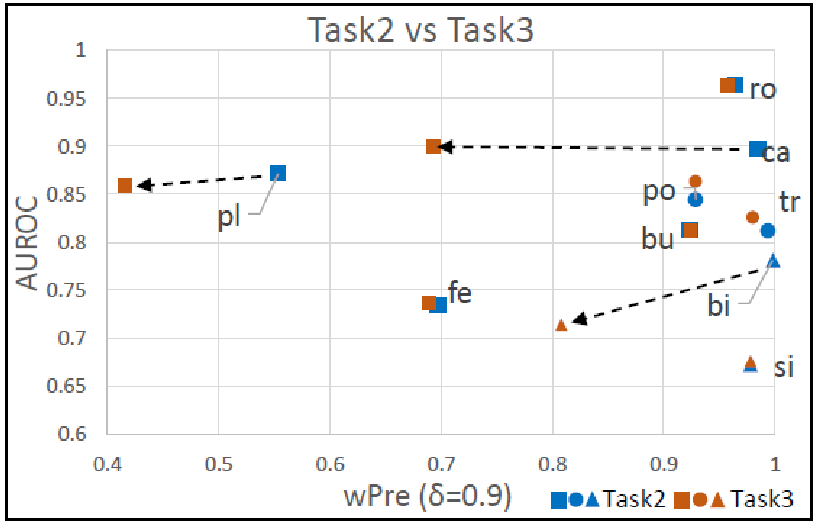
$$\mathrm{FPR}(c,\delta) = \frac{\mathrm{FP}(c,\delta)}{\mathrm{FP}(c,\delta)+\mathrm{TN}(c,\delta)}$$

- Some classes have very small FP and TN, and even a small FP could yield a high FPR
- If classes are highly imbalanced, TPR, FPR and AUROC may not sufficiently evaluate the performance.

AUROC with wPre







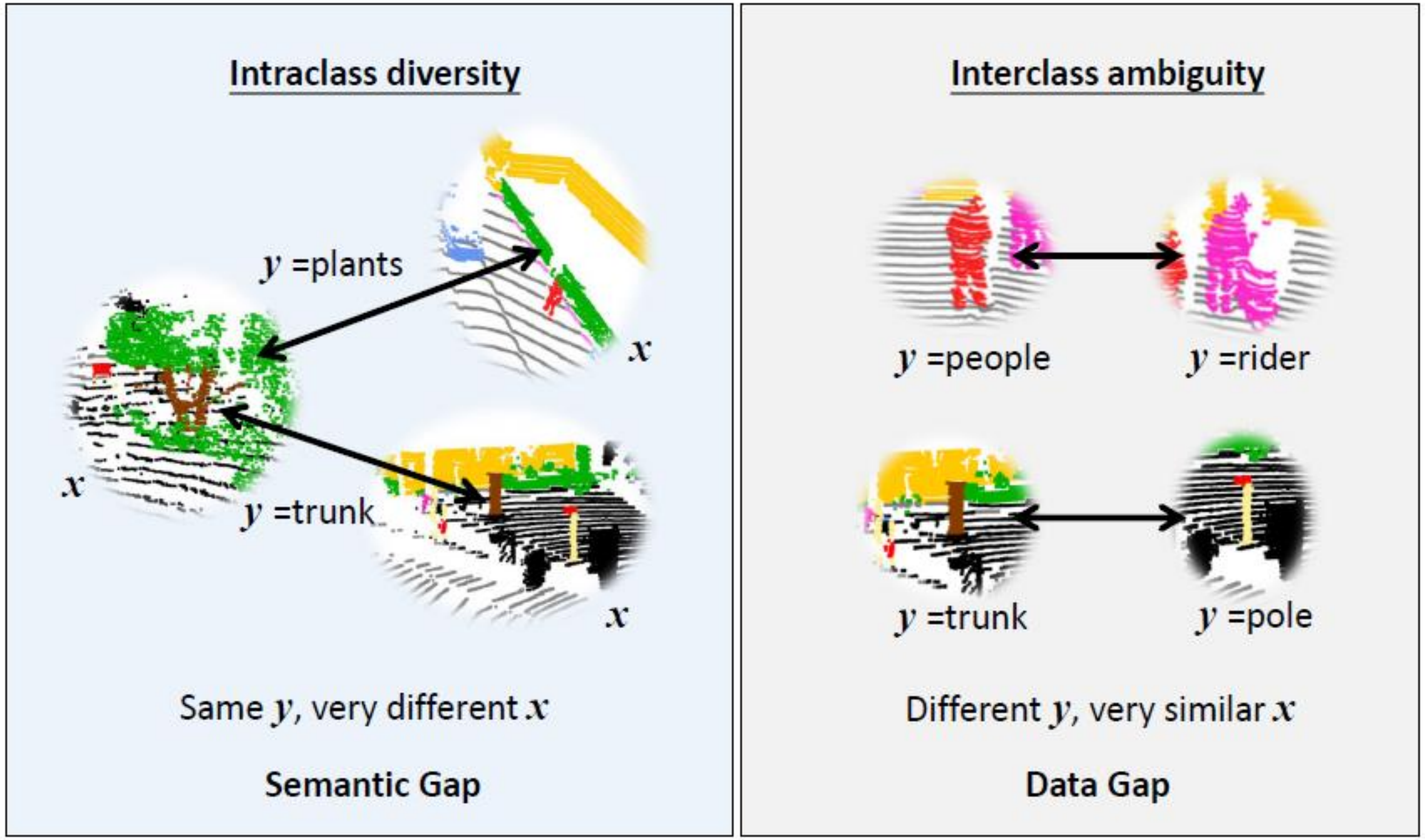
- Task1, car, plants and bike have poor precisions.
- Task2, plants and fence have poor precisions.
- Task3, car, plants and bike are the most affected by OOD.

$$\mathrm{wPre}(c,\delta) = \frac{\mathrm{wTP}(c,\delta)}{\mathrm{wTP}(c,\delta) + \mathrm{wFP}(c,\delta)}$$

Conclusion

- This work conducted experimental studies to understand the challenges of deep 3DSS models facing class imbalanced and OOD data.
- Two experiments are conducted with intensive analysis, and a 3D LiDAR dataset augmentation method, evaluation metrics that accounting for class-imbalance problem, a visual analysis method are developed.

Future Works



- Classes are not imbalanced only on data size.
- Intraclass diversity and interclass ambiguity need to be faced to improve the trustfulness of 3DSS, where semantic gap and data gap need to be studied at real-world scenes.

More results:

Understanding the Challenges When 3D Semantic Segmentation Faces Class Imbalanced and OOD Data, arXiv2022

POSS dataset:

http://www.poss.pku.edu.cn/download.html

More information of POSS-Lab:

http://www.poss.pku.edu.cn/

Contact:

Yancheng Pan, panyancheng@pku.edu.cn

Huijing Zhao , zhaohj@pku.edu.cn