UpCycling: Semi-supervised 3D Object Detection without Sharing Raw-level Unlabeled Scenes

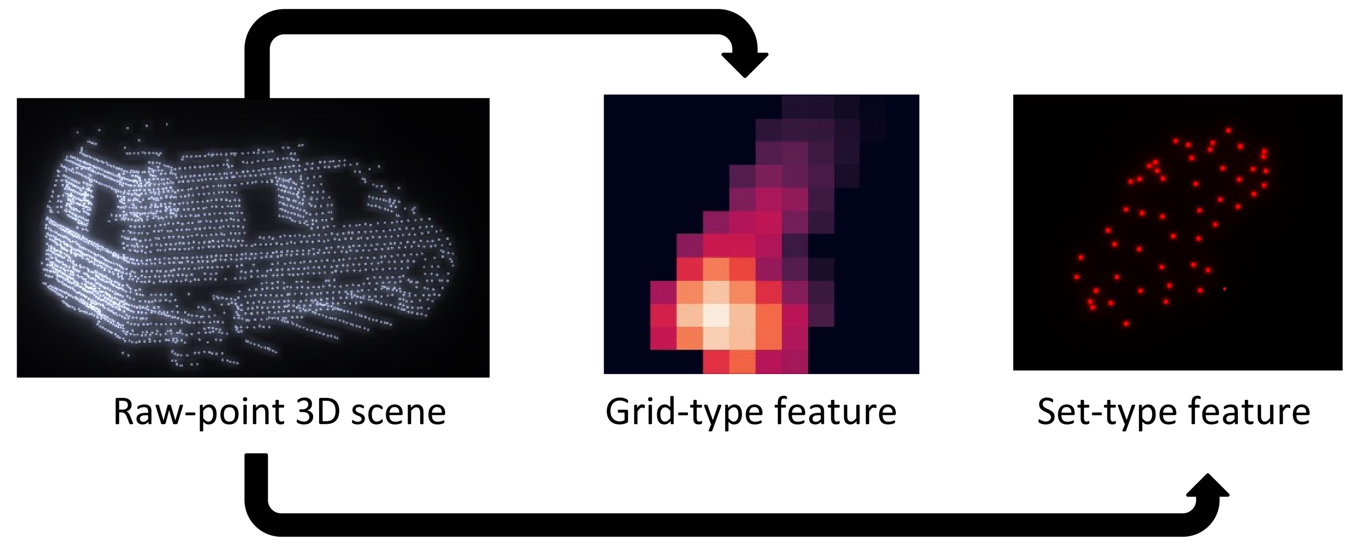
Sunwook Hwang[1], Youngseok Kim[1], Seongwon Kim[2], Saewoong Bahk[1], Hyung-Sin Kim[1]
[1]Seoul National University, [2]SK Telecom

ICCV23 PARIS

Introduction

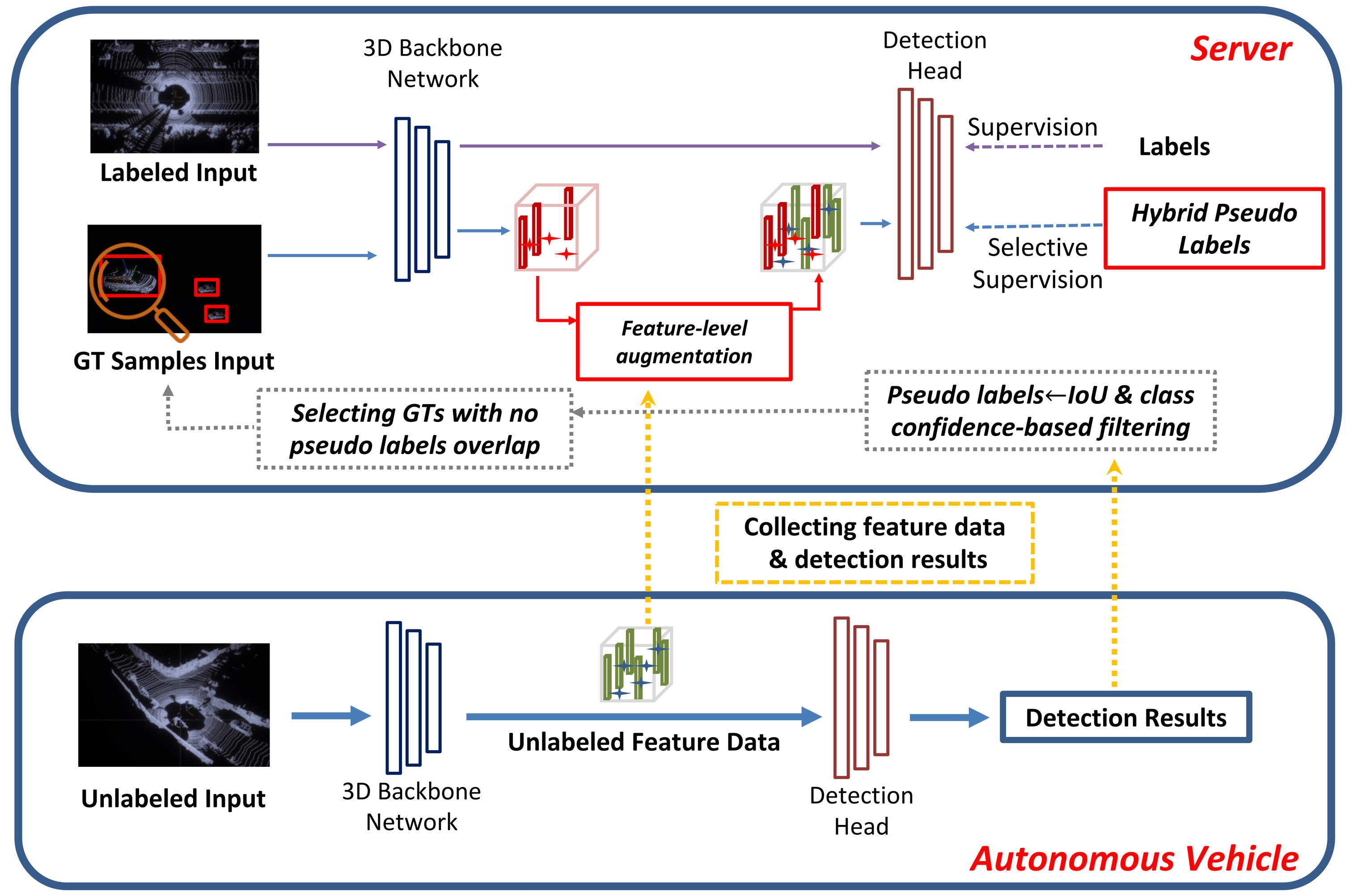
Motivation

- Problem: Collecting and annotating large datasets of 3D scenes
  - ➢ Exposing **private information** on the road
  - ➢ **Labeling costs** for vast amounts of data
- Semi-supervised learning: Using *de-identified*, *unlabeled* intermediate features



UpCycling System Overview

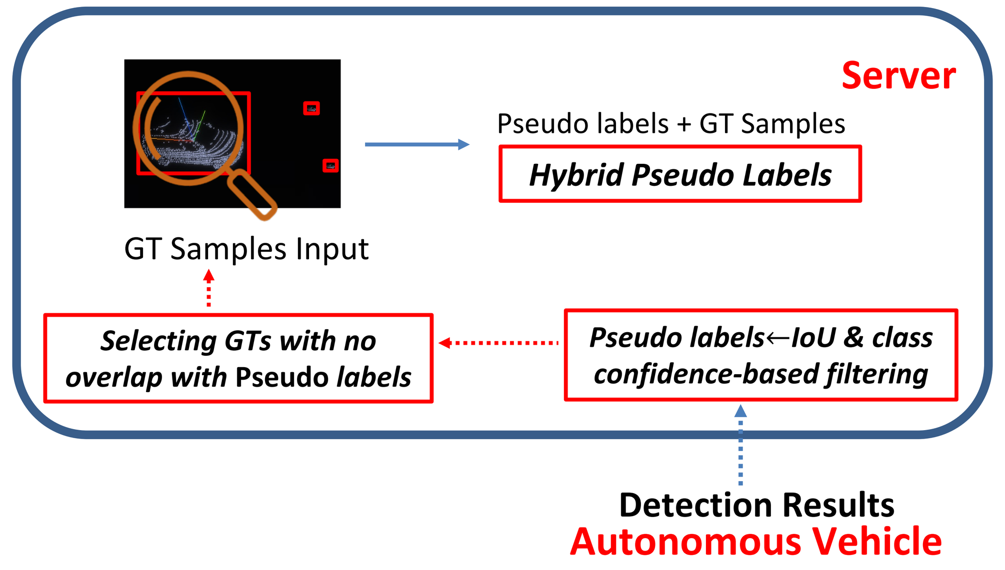
- ***AV-side***: zero additional computation
  - ➢ Sharing de-identified feature data and the detection results
- ***Server-side***: impossible to identify the original data
  - ➢ Semi-supervised learning with collected data and secured data



Essence of *UpCycling*

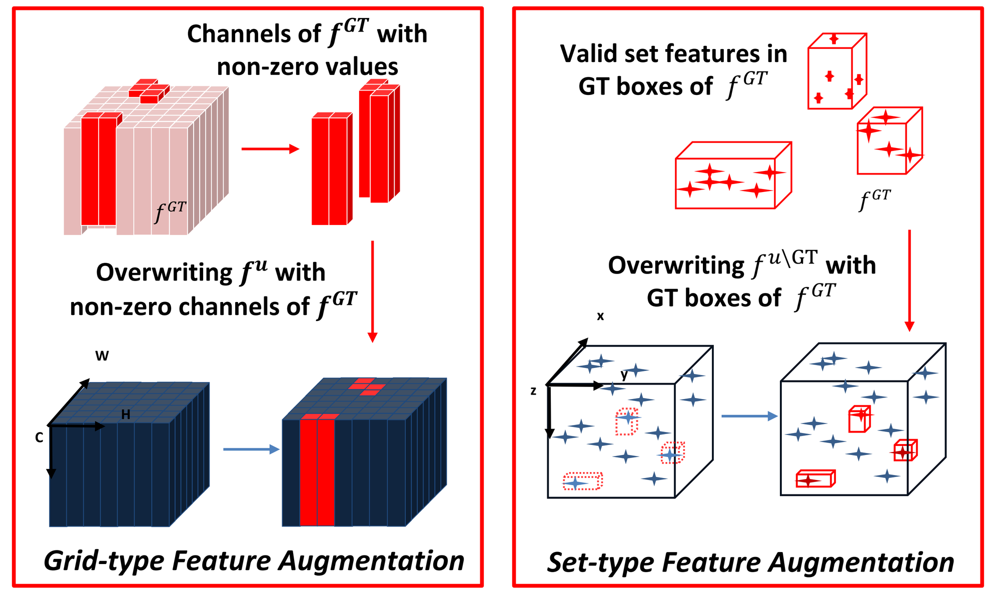
Hybrid Pseudo Label

- **Confidence-based pseudo label filtering**
- **Pseudo-label-aware GT sampling**
- **Hybrid pseudo labels**
  - ➢ **GT samples**: Enabling powerful supervision
  - ➢ **Pseudo labels**: Extending the training data



Feature Augmentation

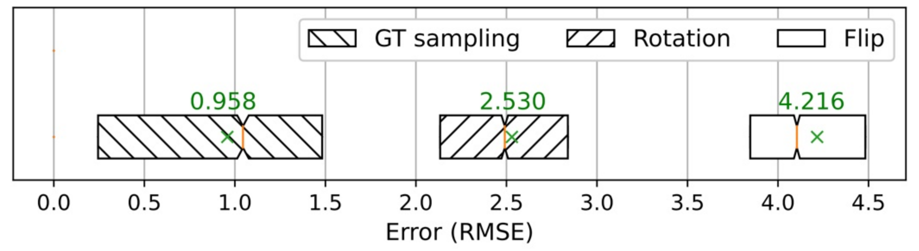
- **Versatile application**
  - ➢ Regardless of feature type
- **Augmentation methods**
  - ➢ **F-GT**: feature-level GT sampling
  - ➢ Overwriting GT features



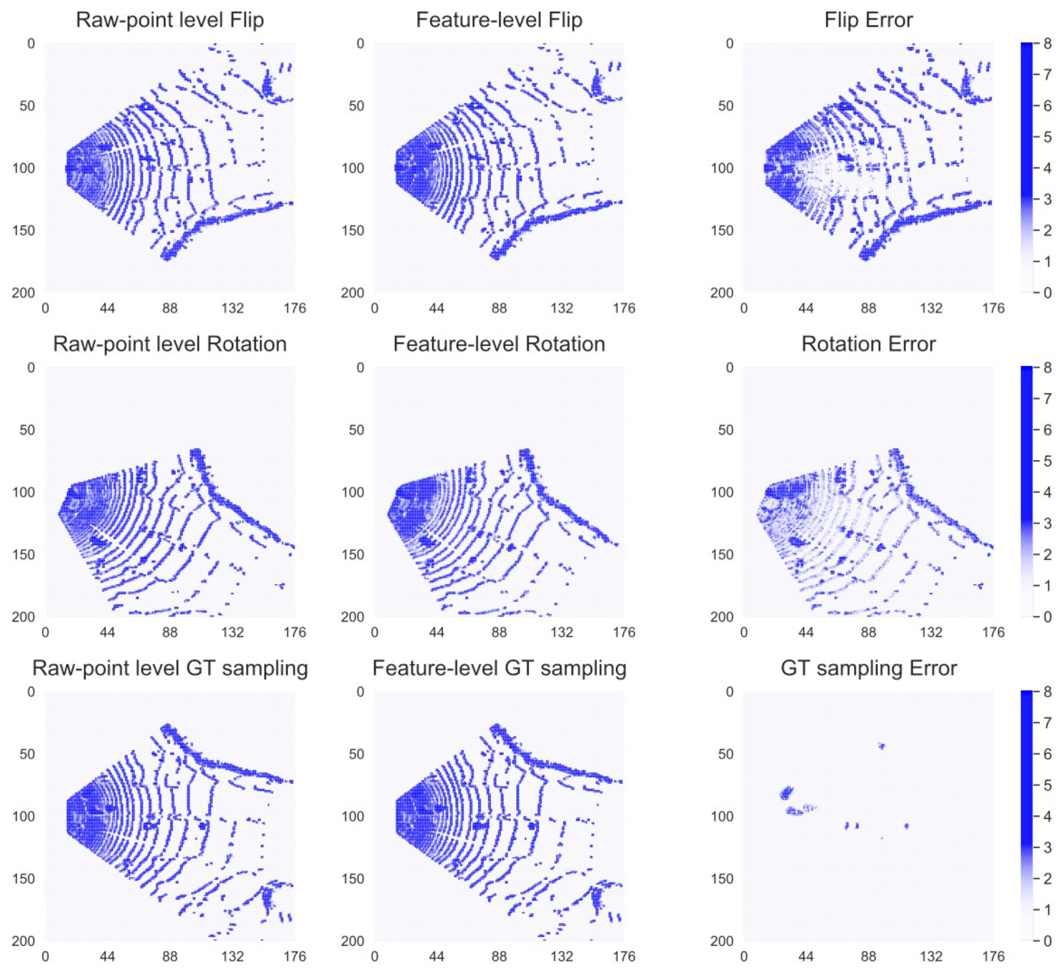
How does it Work?

- **GT sampling**
  - ➢ Error in restricted areas only
- **Flip, Rotation**
  - ➢ Widely spread error over the entire map
  - ➢ **Flip**: breaking apart geometric relationship severely
  - ➢ **Rotation**: Incurring perturbation

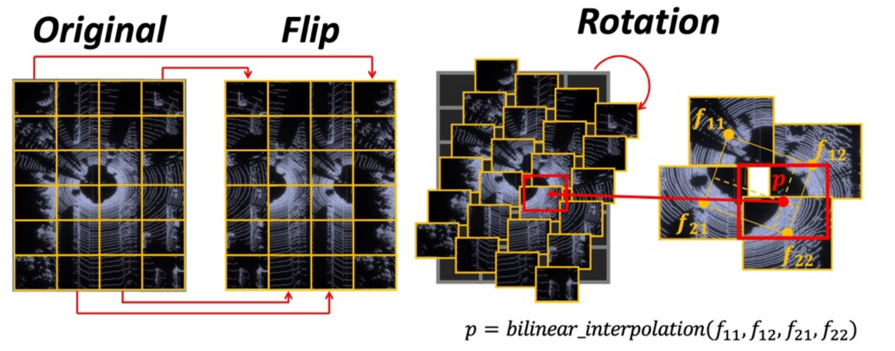
RMSE between raw- and feature-level augmentations



Heatmaps of Bird-eye View (BEV) features



Conceptual images of feature-level augmentation



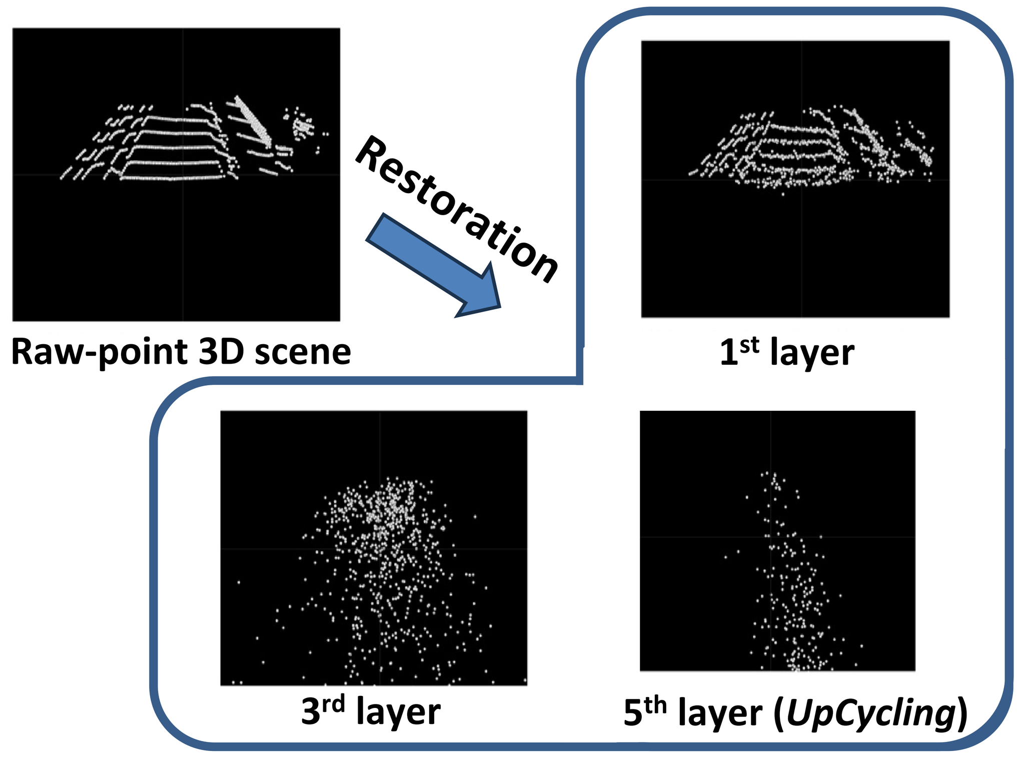
Effect of feature augmentation methods for SSL

Policy #	*Flip*	*Noise*	*RS*	*Rot.*	*F-GT*	AP_{3D} Easy	Mod	Hard
Baseline						70.58	56.00	47.94
1	✓					−16.31	−20.09	−19.79
2		✓				+0.03	+0.13	−1.23
3			✓			+2.47	−0.96	+0.63
4*	✓		✓			−11.69	−13.75	−13.32
5				✓		+4.80	+5.42	+7.96
UpCycling					✓	**+7.81**	**+7.87**	**+8.14**

Results

Privacy Protection of Feature Sharing

- Privacy leaks from inversion attack
- Restoration from backbone network
  - ➢ **1st, 3rd,** and **5th** convolution layers
- Privacy Protection via ***UpCycling***
  - ➢ Assured through deepest-layer features usage



Domain Adaptation

- **Achieving SOTA accuracy**
  - ➢ Regardless of the model, dataset, and detection task

Settings		
Source domain (SD)	Waymo	
Target domain (TD)	Lyft, KITTI	
Baseline	Train → SD	Test → TD
Oracle	Train → TD	Test → TD

Dataset	Method	SECOND-IoU AP_{BEV} / AP_{3D}	PV-RCNN AP_{BEV} / AP_{3D}
Lyft	Baseline	30.20 / 21.32	33.00 / 24.49
	SN	28.38 / 19.25	33.44 / 25.64
	ST3D	60.53 / 29.90	62.28 / 42.63
	UpCycling	**68.83 / 45.66**	**63.38 / 46.83**
	ST3D (w/ SN)	52.86 / 21.25	60.15 / 44.02
	UpCycling (w/ SN)	**65.10 / 49.24**	**63.58 / 49.35**
	Oracle	76.70 / 61.70	78.68 / 64.54
KITTI	Baseline	54.14 / 10.16	62.24 / 9.24
	SN	60.80 / 37.30	60.08 / **38.86**
	ST3D	**70.90 / 40.16**	**66.19** / 23.26
	UpCycling	58.26 / 11.71	62.09 / 11.35
	ST3D (w/ SN)	80.97 / 57.68	54.30 / 48.79
	UpCycling (w/ SN)	**84.12 / 67.65**	**85.90 / 61.12**
	Oracle	90.36 / 82.02	90.84 / 84.56

Partial-label Scenario

- Superiority to the SOTA method in most cases
- Less mature, large-backbone model underperform with scarce labels (≤ 10%)

AP_{3D}		2% Easy	2% Mod	2% Hard	10% Easy	10% Mod	10% Hard	25% Easy	25% Mod	25% Hard
SECOND-IoU	Baseline	56.69	44.11	37.19	70.58	56.00	47.94	84.47	71.06	62.87
	3DIoUMatch	63.57	49.58	43.00	71.76	57.01	50.08	81.71	68.51	60.92
	improved (%)	12.13	12.39	15.62	1.67	1.80	4.47	-3.26	-3.59	-3.11
	UpCycling	70.19	59.97	44.83	76.09	60.41	51.84	85.22	72.87	63.93
	improved (%)	**23.81**	**35.96**	**20.54**	**7.81**	**7.87**	**8.14**	**0.89**	**2.55**	**1.69**
PV-RCNN	Baseline	68.10	53.27	46.20	81.23	68.67	60.32	87.63	76.03	68.62
	3DIoUMatch	81.04	65.77	58.83	85.26	70.64	63.32	85.08	72.37	65.02
	improved (%)	**19.00**	**23.47**	**27.34**	**4.97**	**2.87**	4.98	-2.91	-4.81	-5.25
	UpCycling	76.46	61.44	52.94	83.64	69.60	63.53	88.05	76.61	70.80
	improved (%)	12.28	15.34	14.59	2.97	1.35	**5.32**	**0.48**	**0.76**	**3.18**

Conclusion

- A novel SSL framework by gathering de-identified & unlabeled data: Labeling cost, privacy leakage, and AV-side computation burden altogether
- Verified the effectiveness feature-based learning
- Achieving SOTA accuracy with large margins in various experiments

Network Laboratory in SNU NETLAB
AIoT Lab