MuDDoS:

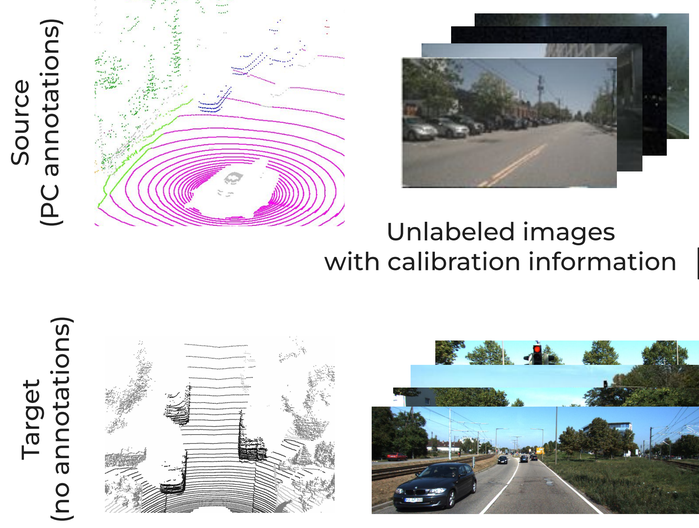
Improving Multimodal Distillation for 3D Semantic Segmentation under Domain Shift

Björn Michele[1,3], Alexandre Boulch[1], Tuan-Hung Vu[1], Gilles Puy[1], Renaud Marlet[1,2] and Nicolas Courty[3]

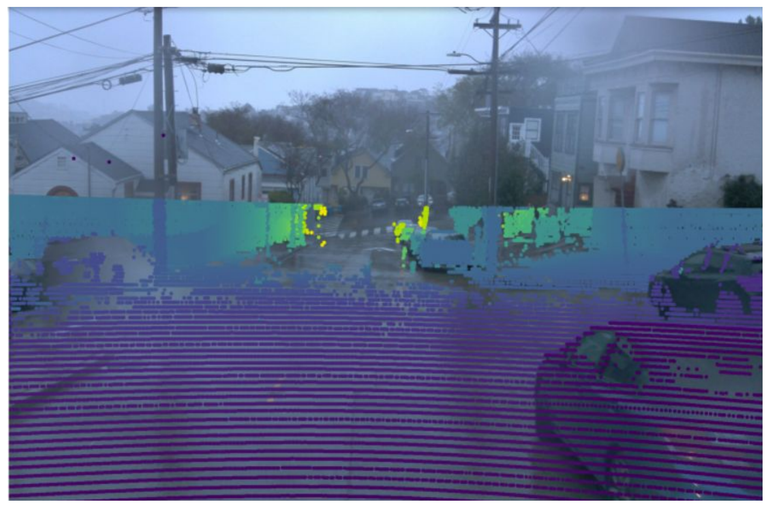
[1]valeo.ai, Paris, France. [2]LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, France.
[3]CNRS, IRISA, Univ. Bretagne Sud, Vannes, France.

BMVC 2025

The challenge: multimodal Unsupervised Domain Adaptation

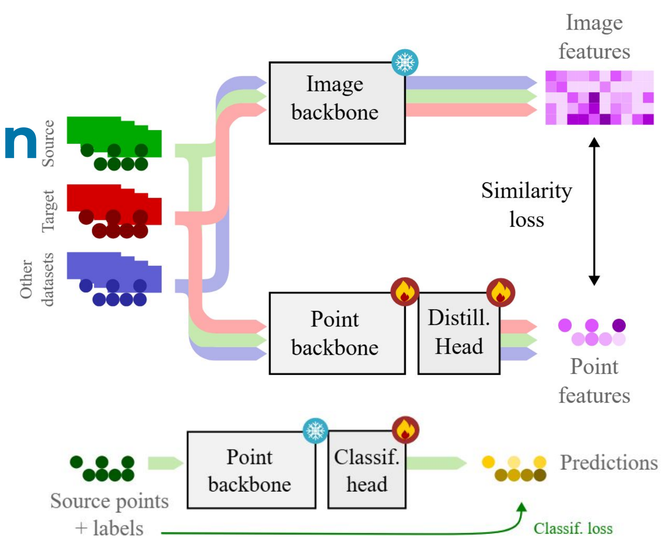


The data: Image + Point Cloud



The base: Joint VFM distillation

Step 1: Distilling a vision foundation model (VFM) into a point cloud backbone



Step 2: Training a classifier with source data

Our study: Get the best out of VFMs for 3D UDA

- Best practices of unimodal UDA
- Choice of a VFM for pretraining
- Choice of pretraining datasets
- Downstream training

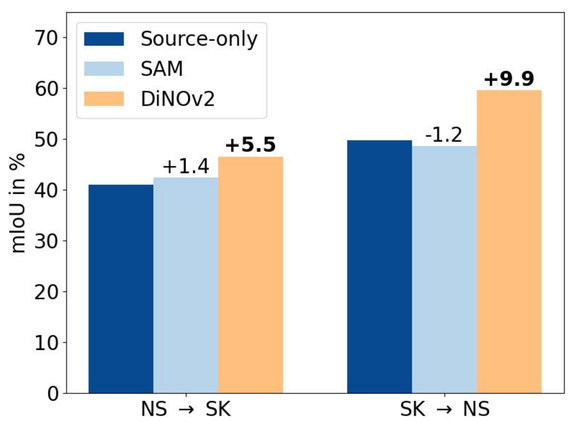
The experiments

Eval. in 4 settings, Real-to-Real, different sensor:
nuScenes to SemanticKITTI (32 to 64 Beams): NS→SK
nuScenes to Waymo (32 to 64 Beams): NS→WO
SemanticKITTI to nuScenes (64 to 32 Beams): SK→NS
Waymo to nuScenes (64 to 32 Beams): WO→NS

Best practices of unimodal UDA

✔️ No intensity
✔️ LayerNorm (LN) instead of BatchNorm (BN)
✔️ WaffleIron(WI) instead of Mink.-U-Net (MUNet)
✔️ Bigger than smaller

Choice of VFM



✔️ The used VFM matters (DiNOv2 >> SAM)

Choice of pretraining data

Back-bone	Pretrain S	Pretrain T	Pretrain E	Frozen back-bone	Avg. mIoU in %
WI-256	✗	✗	✗	✗	47.0
	✓	✗	✗	✓	33.3
	✓	✓	✗	✓	56.8
WI-768	✗	✗	✗	✗	50.4
	✓	✓	✗	✓	61.6
	✓	✓	✓	✓	**63.6**

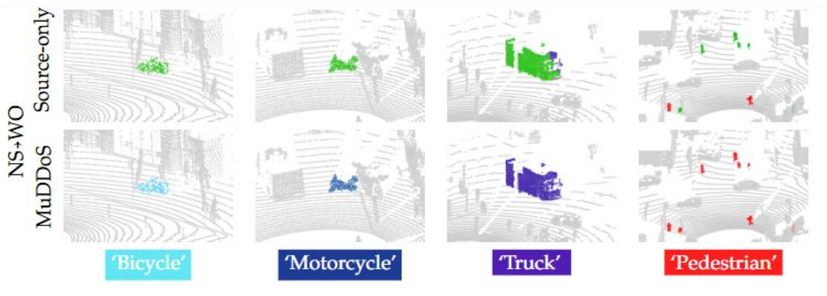
✔️ Joint distillation on source (S) and target (T) makes it work

Downstream training

Downstream training	NS→SK	SK→NS
Frozen Back. + Linear	43.4	<u>60.1</u>
Frozen Back. + MLP	<u>47.5</u>	**62.1**
Full finetuning	**50.5**	58.9
Joint distill. & classif.	44.9	56.0

✔️ Freeze the backbone to keep the distilled features

Qualitative results



Comparison to SOTA

Method	Back.	Mod.	Avg. mIoU in %
Source-only	MUNet	-	26.8
Source-only	WI-768	-	50.4
PL	MUNet	U	28.3
CosMix	MUNet	U	30.5
MM2D3D	MUNet	M	34.6
Adapt-SAM	MUNet	M	46.1
MuDDoS (ours)	MUNet	M	50.6
Self-Training	WI-768	U	44.2
MuDDoS (ours)	WI-768	M	**64.5**

MuDDoS Takeaways

☑ **Unimodal UDA best practices**
☑ **Careful VFM selection**
☑ **Joint distill. of source and target**
☑ **Careful downstream training**