Nuages de Points et Modélisation 3D

7 - Machine learning IV

Overview

Machine learning courses

- Surface reconstruction
- Descriptors and machine learning
- Image based processing
- Geometric deep learning
- Convolutional and Transformer based architectures
- Tasks and corresponding architectures

Today

Course overserview

- Tasks
  - Classification and segmentation
  - Detection from point cloud
  - Scene completion
  - Generation
  - Surface reconstruction
- Self-supervised learning
  - Point cloud only pretraining
  - Distillation
- Domain adaptation
- Open world

Course overserview

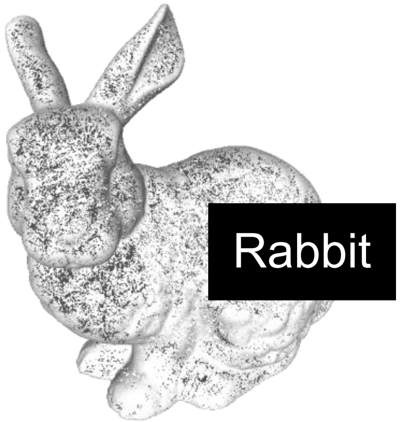
- Tasks
    - Classification and segmentation
    - Detection from point cloud
    - Scene completion
    - Generation
    - Surface reconstruction
- Self-supervised learning
    - Point cloud only pretraining
    - Distillation
- Domain adaptation
- Open world

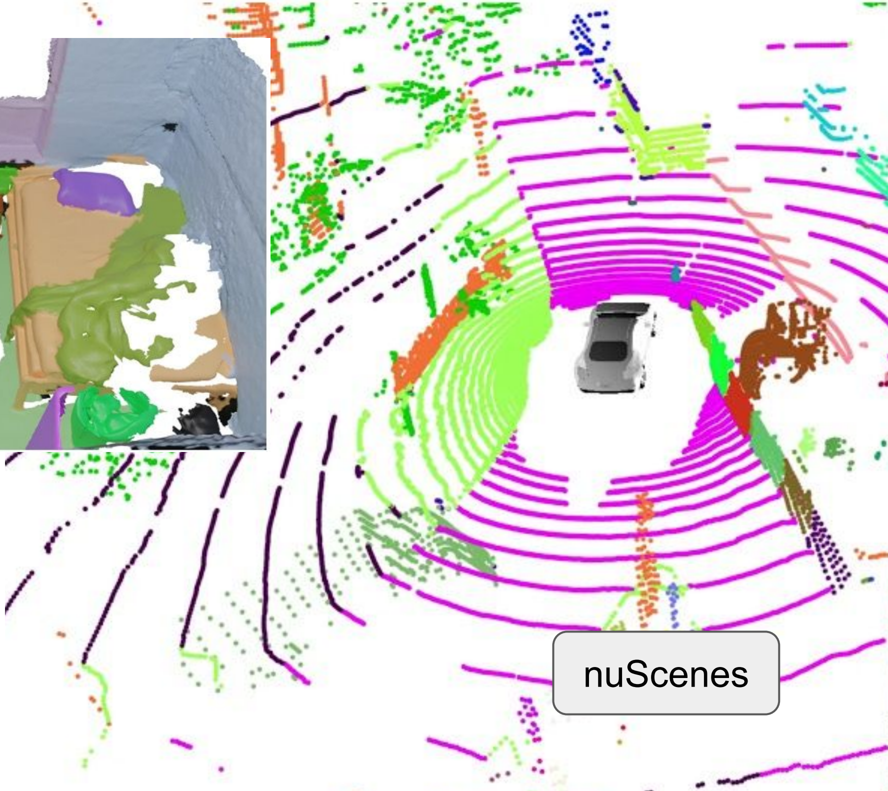
Next week: QCM + (opening on multiview and rendering)

I - Tasks

Classification and semantic segmentation

I - Tasks





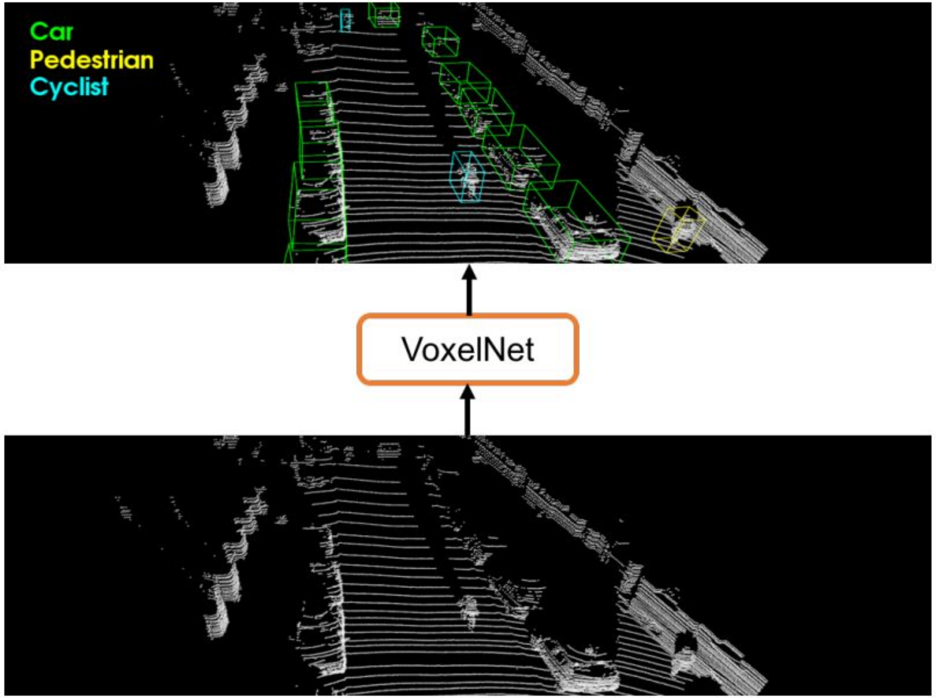
Classification → one label per point cloud

Segmentation → one label per point

Detection

I - Tasks

Put oriented boxes around objects of interest

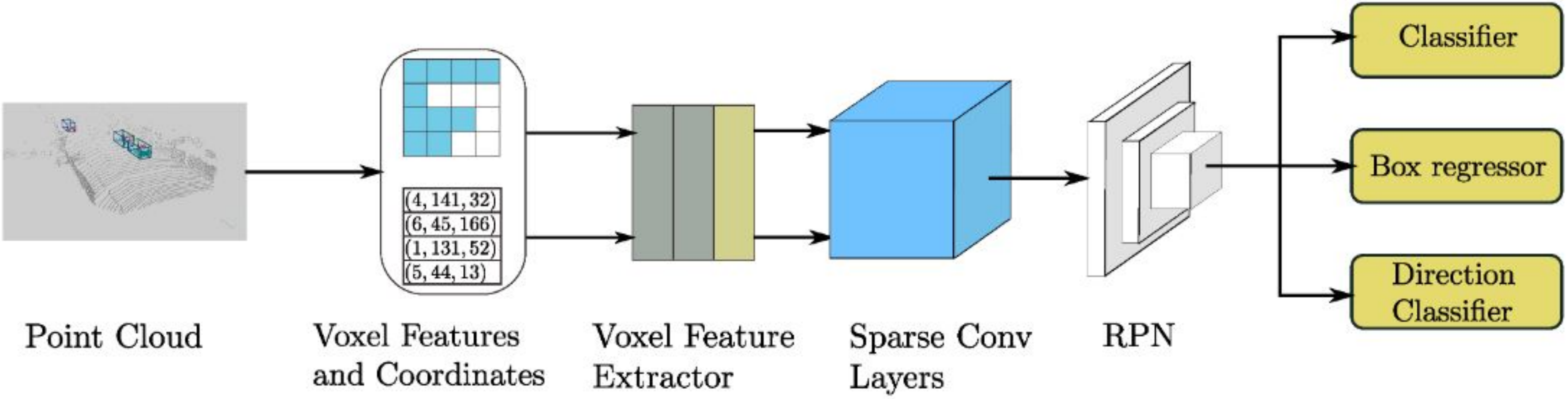


Yan, Yan, Yuxing Mao, and Bo Li. "Second: Sparsely embedded convolutional detection." Sensors 18.10 (2018): 3337.

Detection

I - Tasks

Second



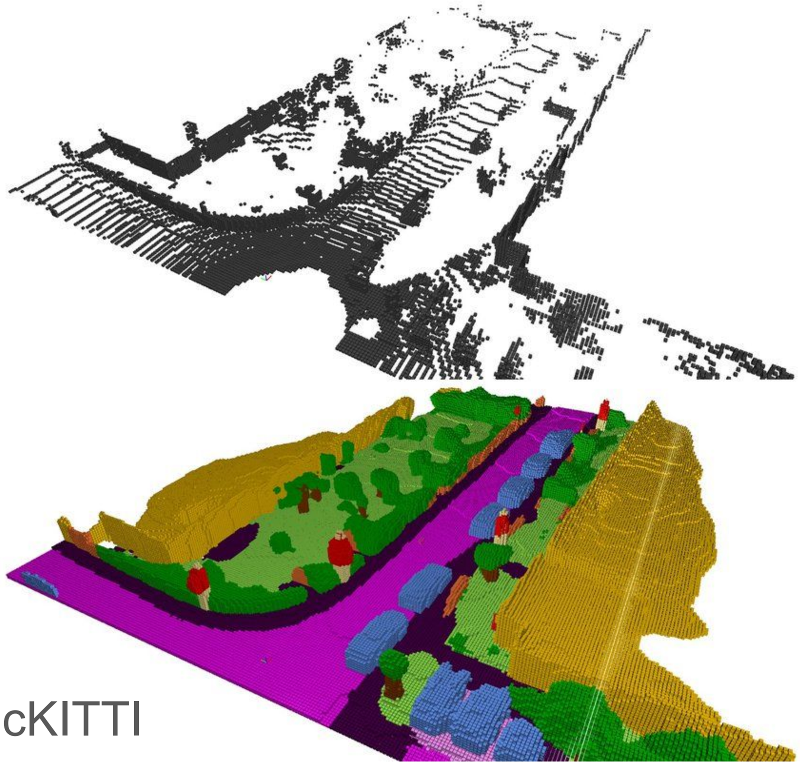
Yan, Yan, Yuxing Mao, and Bo Li. "Second: Sparsely embedded convolutional detection." Sensors 18.10 (2018): 3337.

Scene completion

I - Tasks

Input: lidar scan

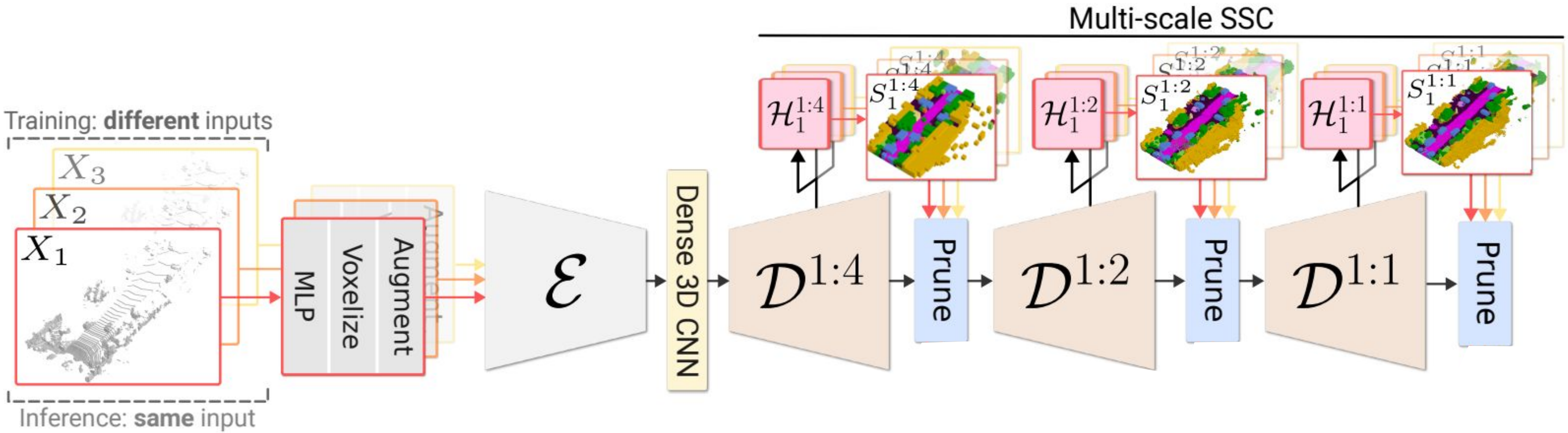
Output: completed voxel scene with semantics



SemanticKITTI

Scene completion

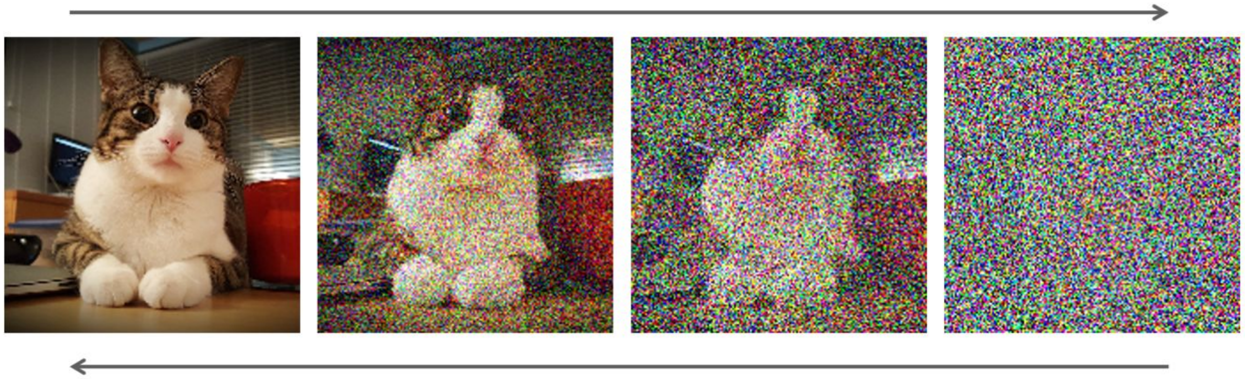
I - Tasks



Cao et al "Pasco: Urban 3d panoptic scene completion with uncertainty awareness." CVPR 2024.

Generation (e.g. diffusion models)

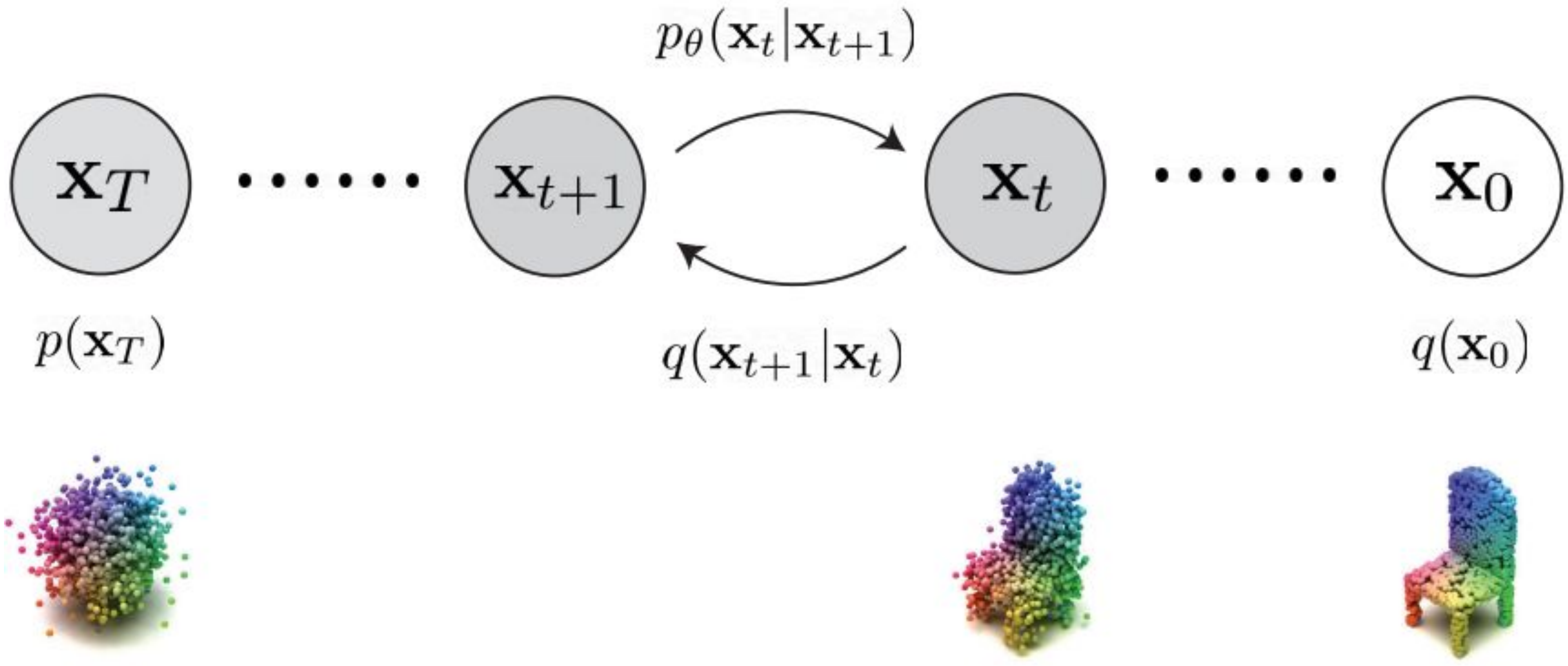
I - Tasks



https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-2/

Generation (e.g. diffusion models)

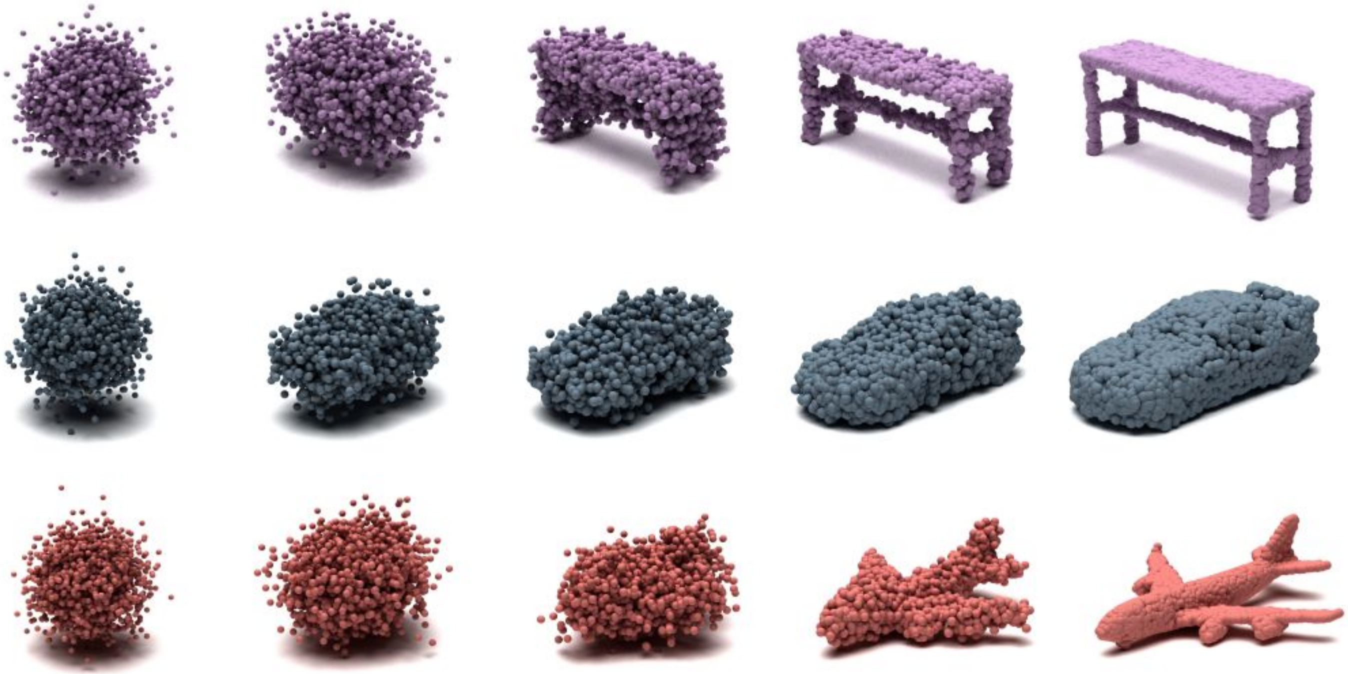
I - Tasks



Zhou et al. "3d shape generation and completion through point-voxel diffusion." ICCV. 2021.

Generation (e.g. diffusion models)

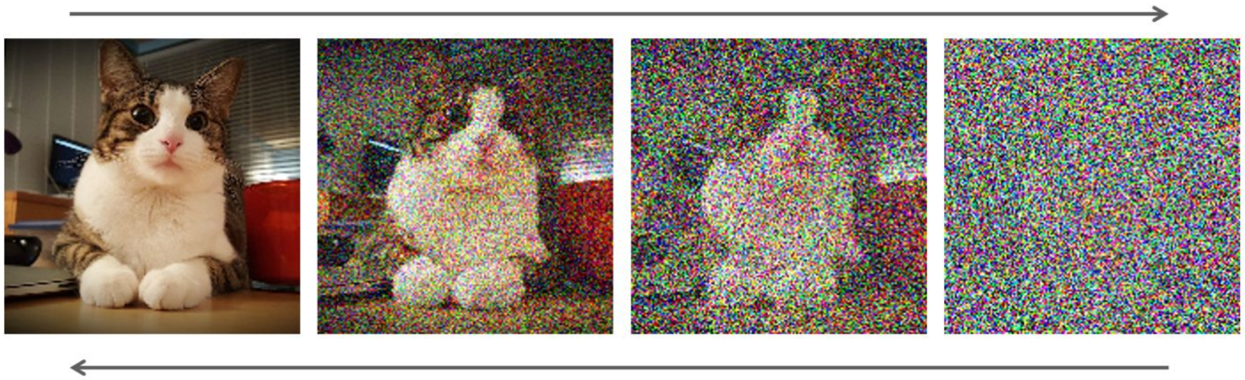
I - Tasks

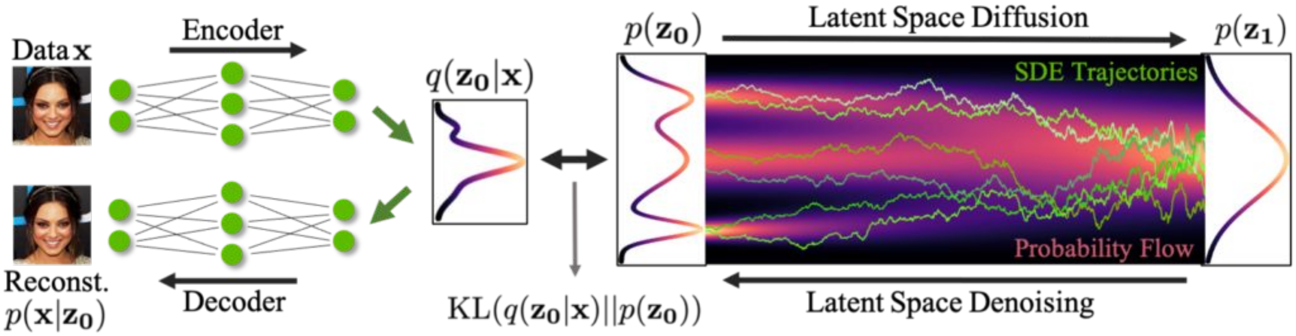


Zhou et al. "3d shape generation and completion through point-voxel diffusion." ICCV. 2021.

Generation (e.g. diffusion models)

I - Tasks

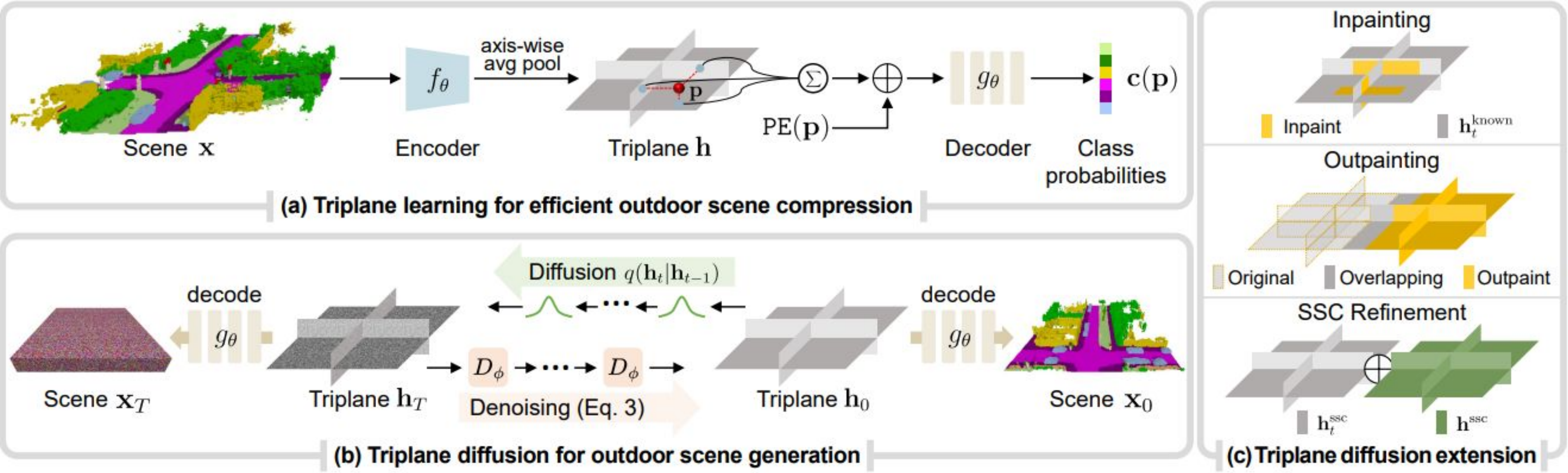




https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-2/

Generation (e.g. diffusion models)

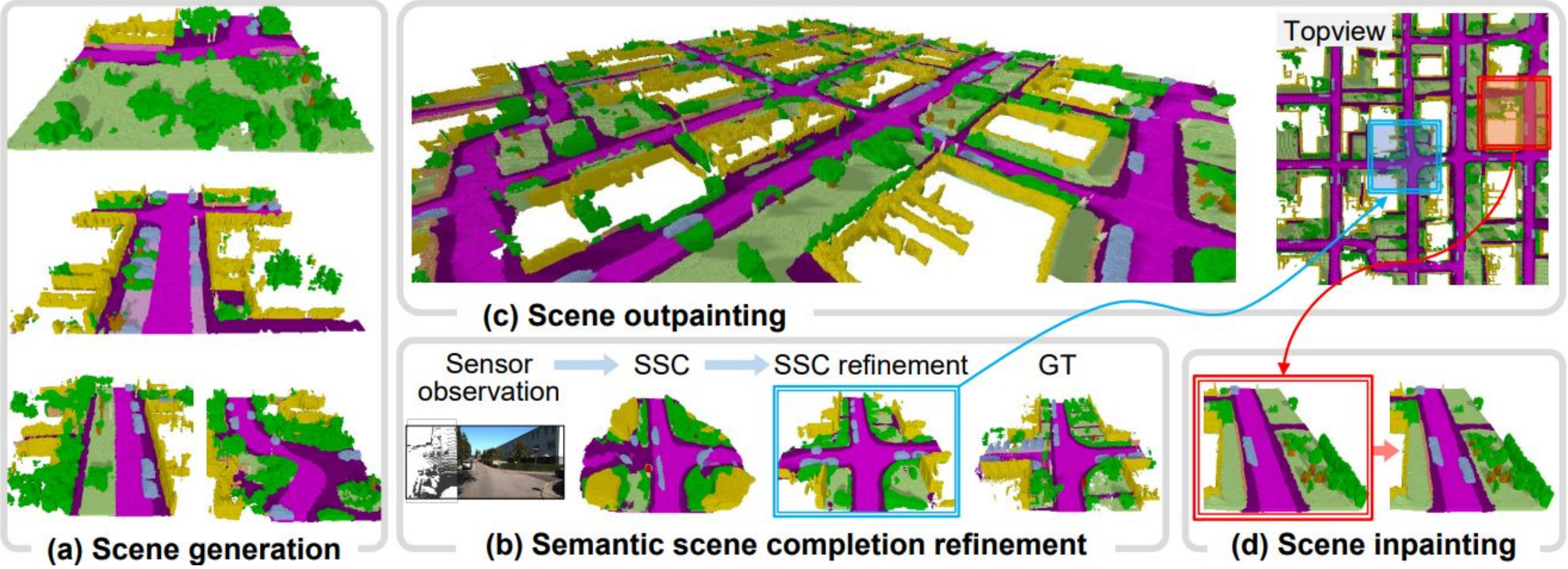
I - Tasks



Lee, Jumin, et al. "Semcity: Semantic scene generation with triplane diffusion." CVPR. 2024.

Generation (e.g. diffusion models)

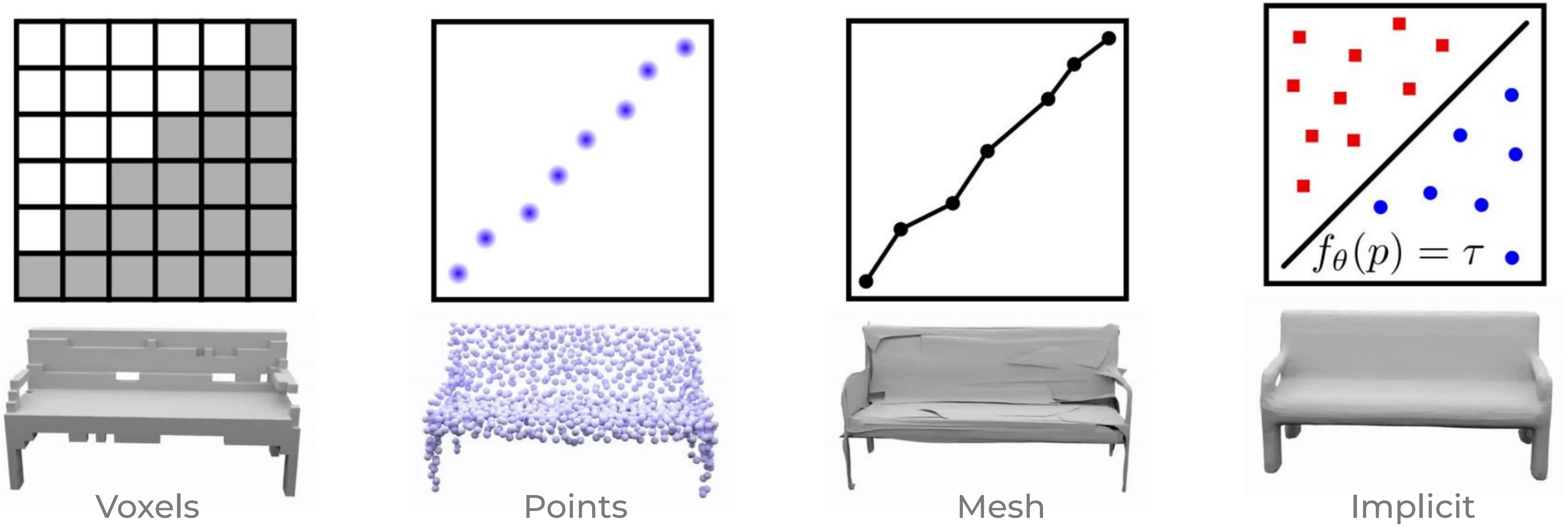
I - Tasks



Lee, Jumin, et al. "Semcity: Semantic scene generation with triplane diffusion." CVPR. 2024.

Surface reconstruction

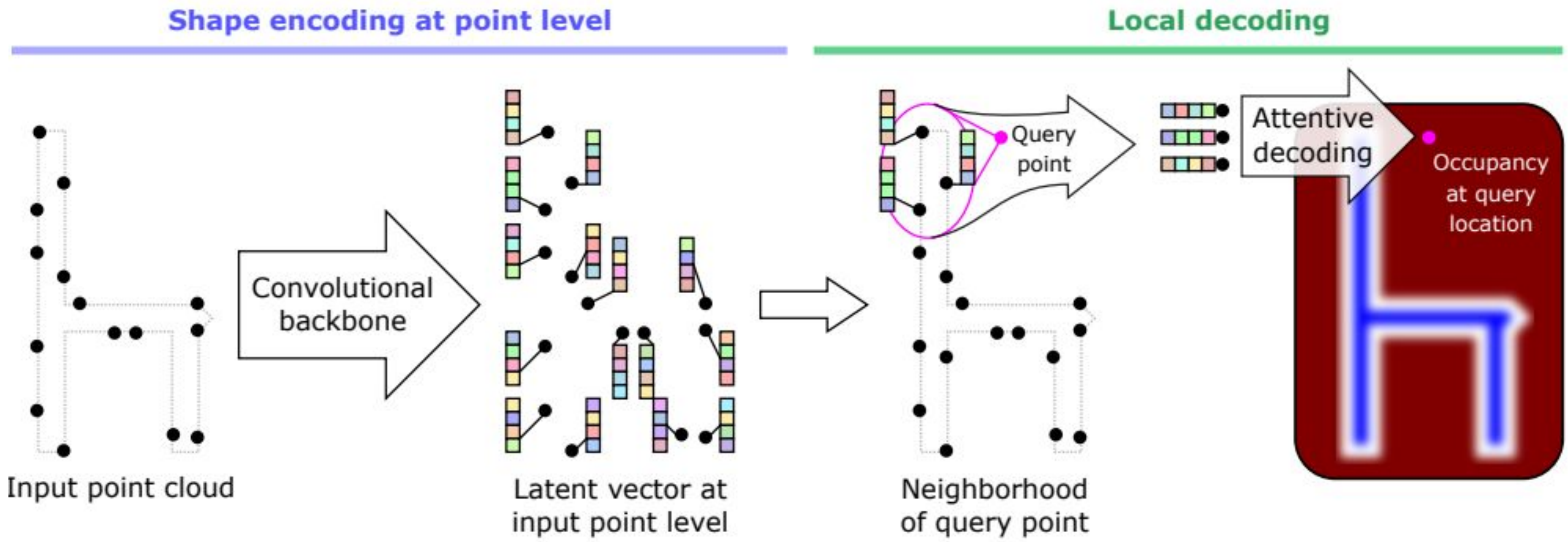
I - Tasks



Occupancy Networks: Learning 3D Reconstruction in Function Space
Mescheder, Lars and Oechsle, Michael and Niemeyer, Michael and Nowozin, Sebastian and Geiger, Andreas
Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2019

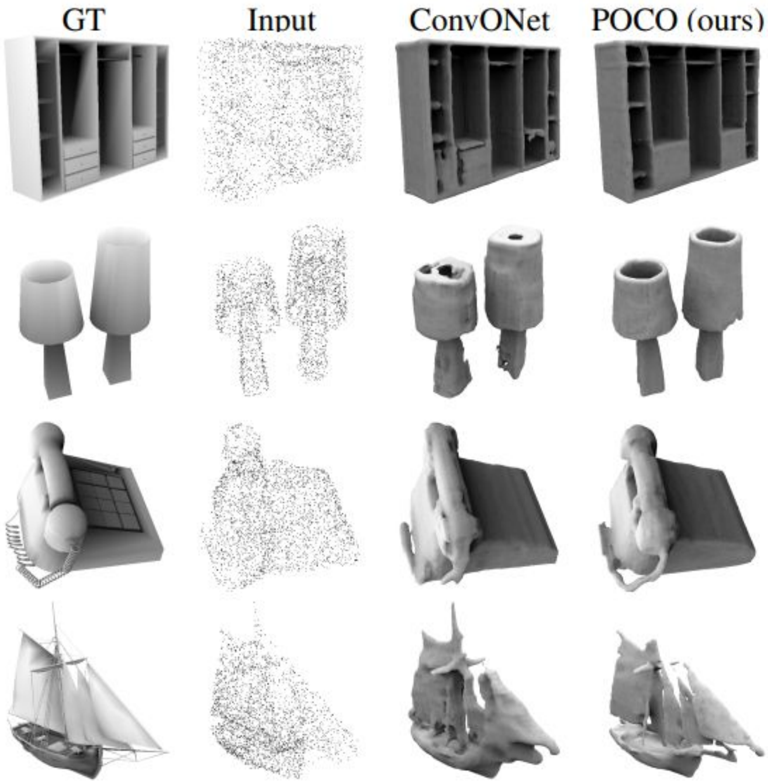
Surface reconstruction

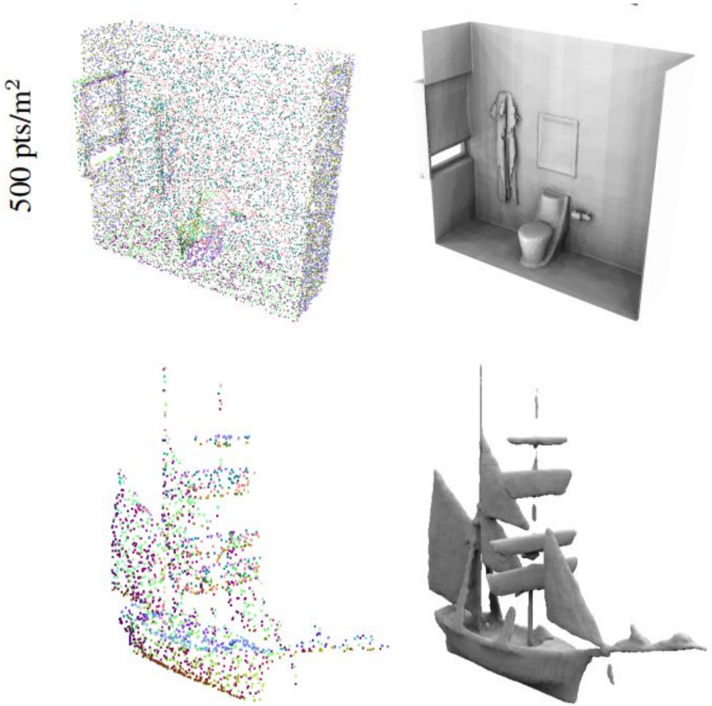
I - Tasks



Surface reconstruction

I - Tasks

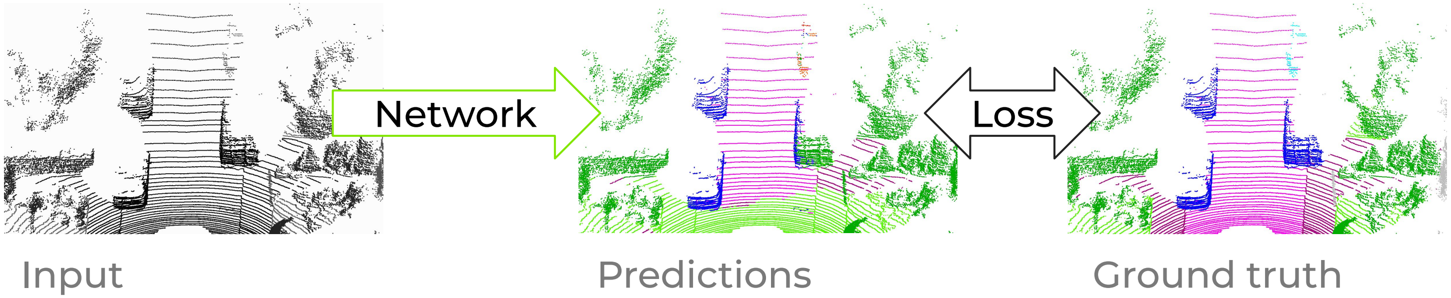




II - Self-supervised learning

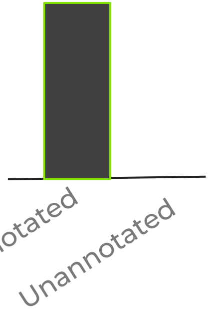
Supervised learning

Self-supervised learning



Train with annotated data

- Annotations are costly
- Limitation of the database size



Frugal learning

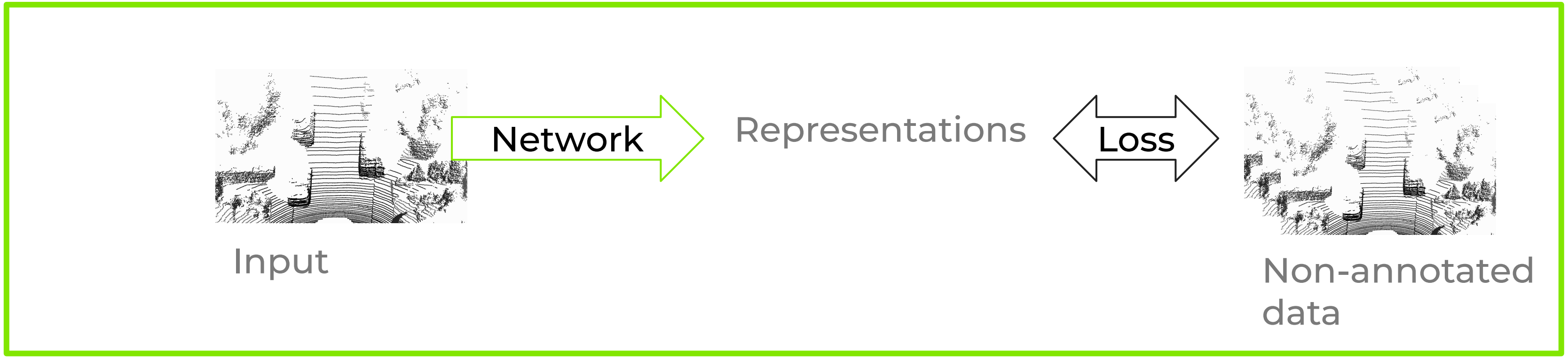
Self-supervised learning

Objective: learning with less annotations

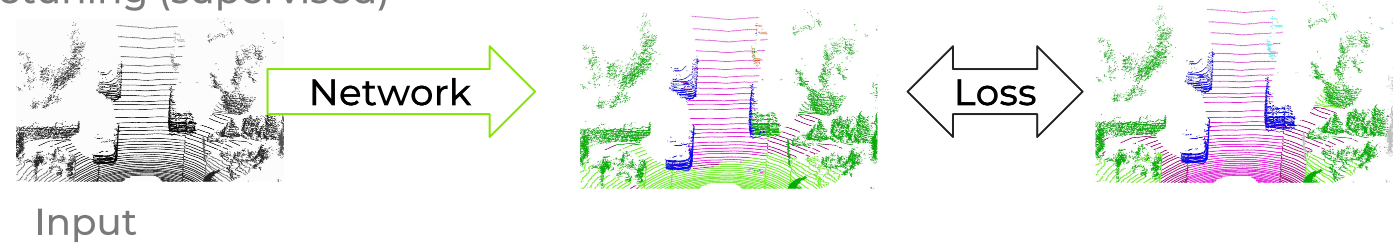
Self-supervision

What? learn useful representations without annotations

Why? better performance when finetuning / data-efficiency



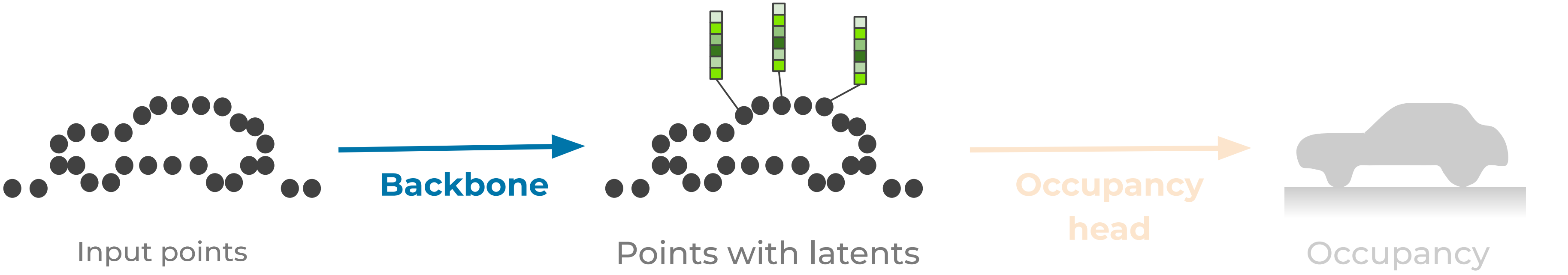
Step 2: finetuning (supervised)

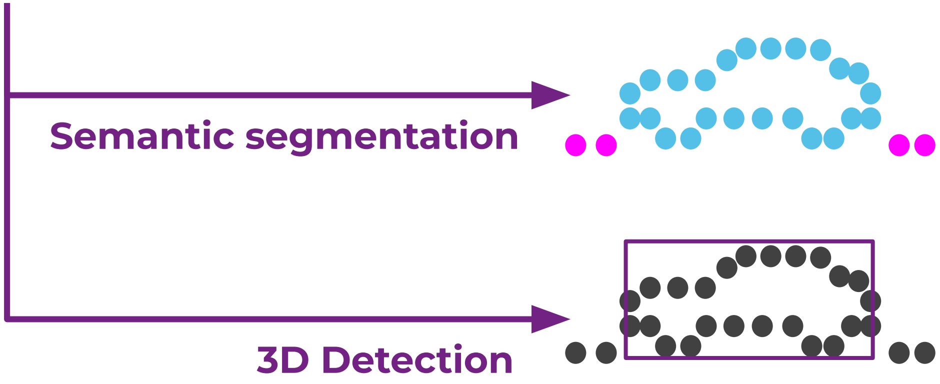


ALSO: Automotive Lidar Self-Supervision by Occupancy Estimation

valeo.ai

ALSO





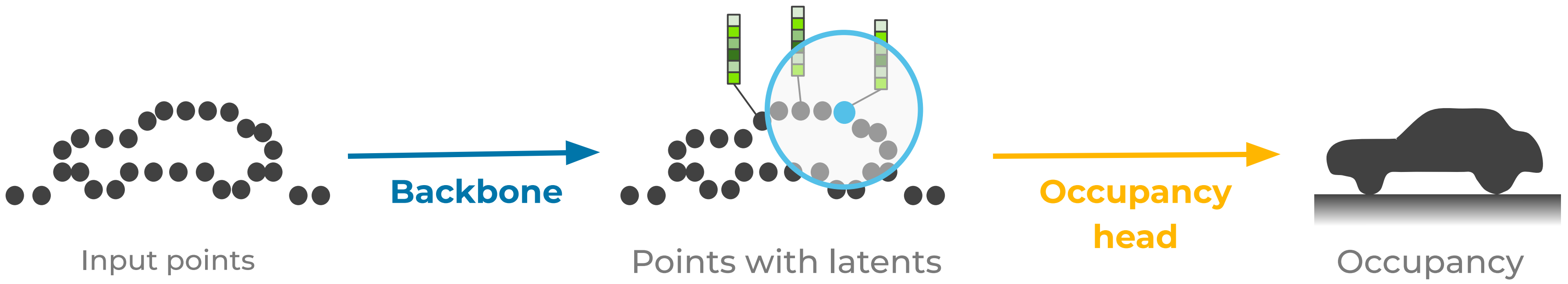
Pretext task: scene reconstruction

Intuition

- Reconstruction information (geometry) is useful for semantic tasks

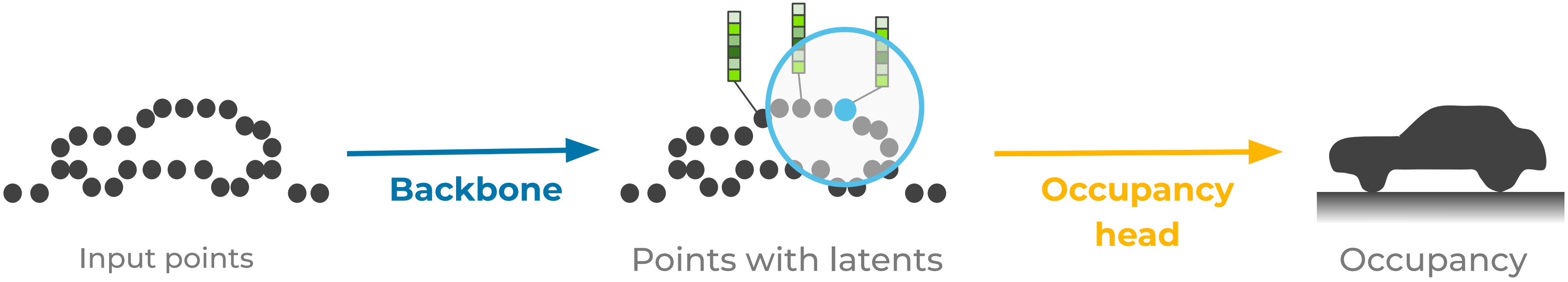
Context reconstruction vs local reconstruction

ALSO



Context reconstruction vs local reconstruction

ALSO



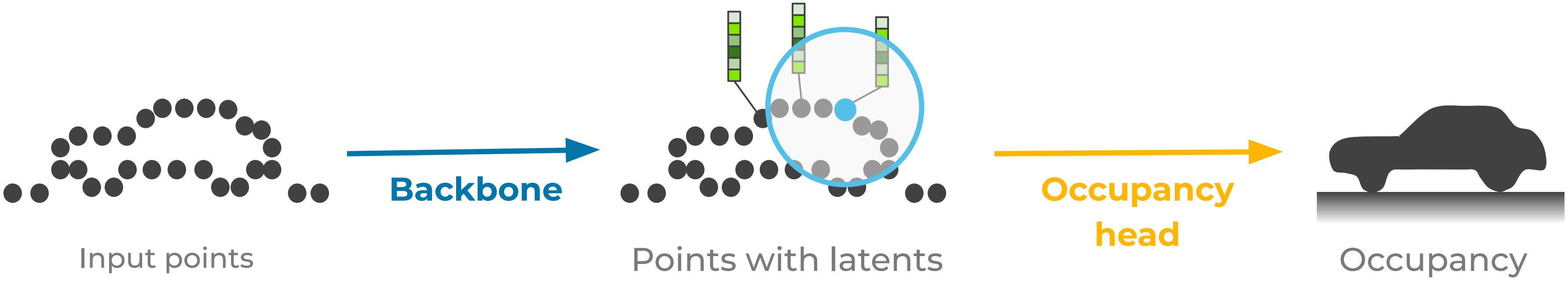
Local reconstruction [POCO head]
- everywhere, from features of neighboring points
⇒ (too) detailed geometry

Context reconstruction [ALSO head]
- of a 1 meter ball, from each single feature point
⇒ rough geometry, more suited for object recognition

POCO: Point Convolution for Surface Reconstruction, A. Boulch, R. Marlet, CVPR 2022

Context reconstruction vs local reconstruction

ALSO



Local reconstruction [POCO head]

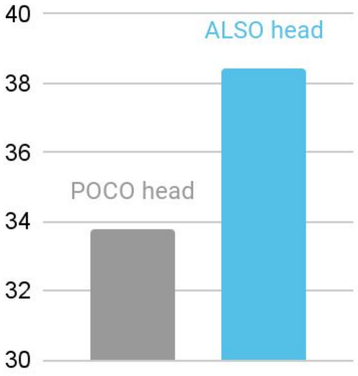
- everywhere, from features of neighboring points

⇒ (too) detailed geometry

Context reconstruction [ALSO head]

- of a 1 meter ball, from each single feature point

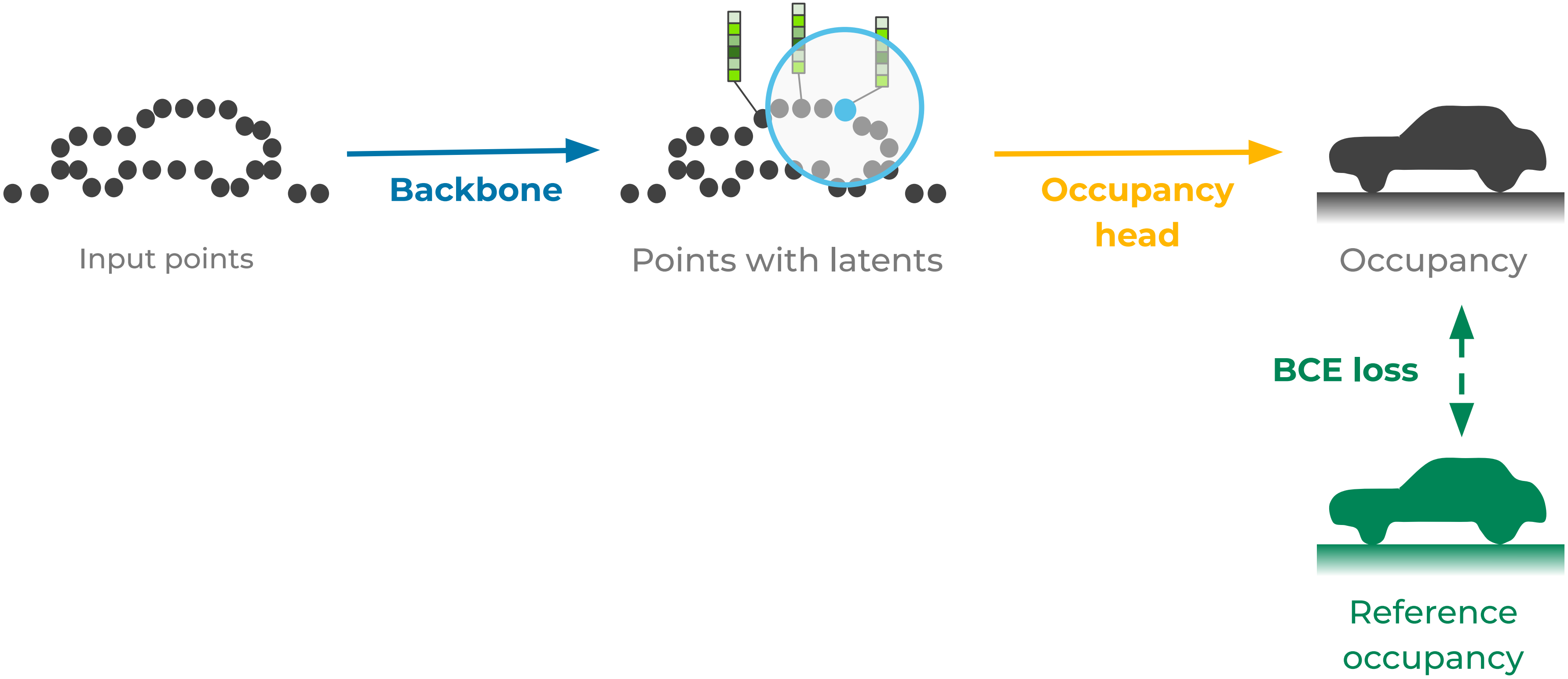
⇒ rough geometry, more suited for object recognition



POCO: Point Convolution for Surface Reconstruction, A. Boulch, R. Marlet, CVPR 2022

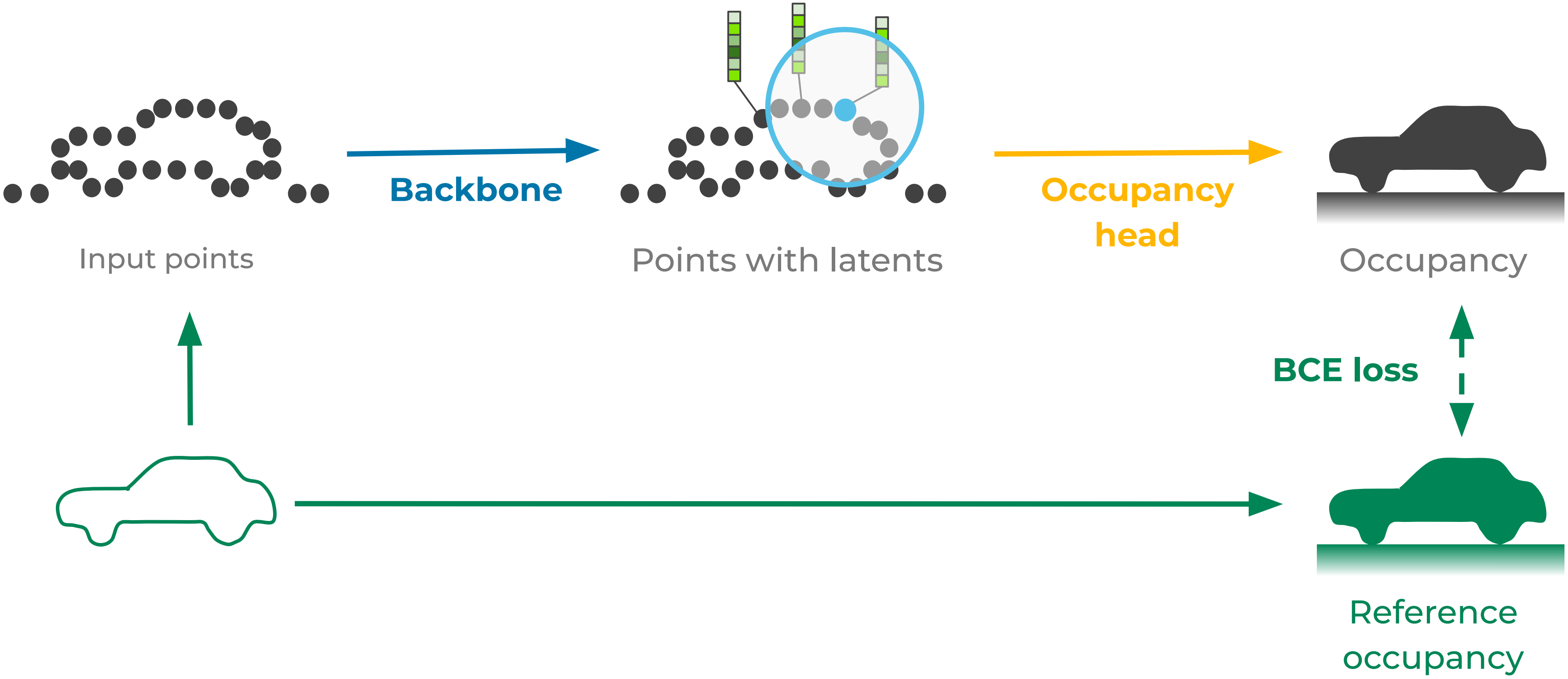
Supervision

ALSO



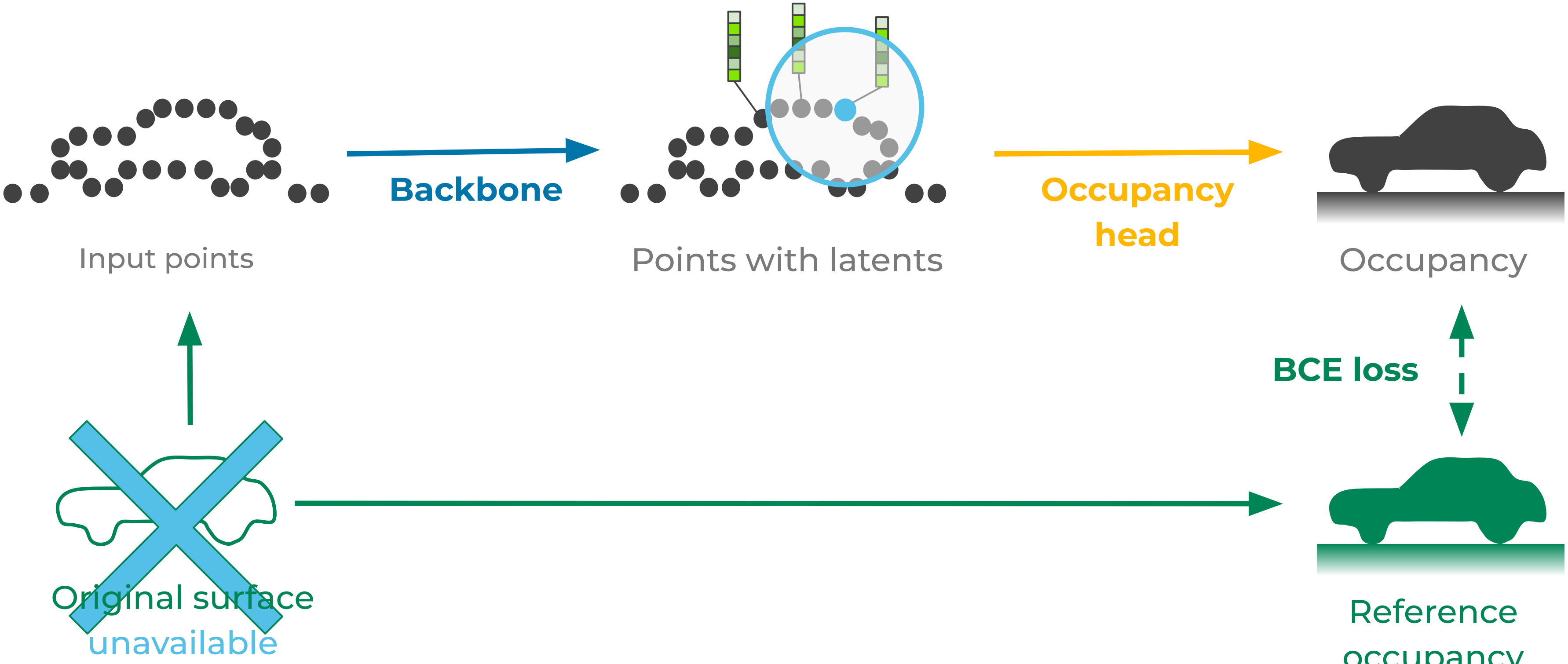
Supervision

ALSO



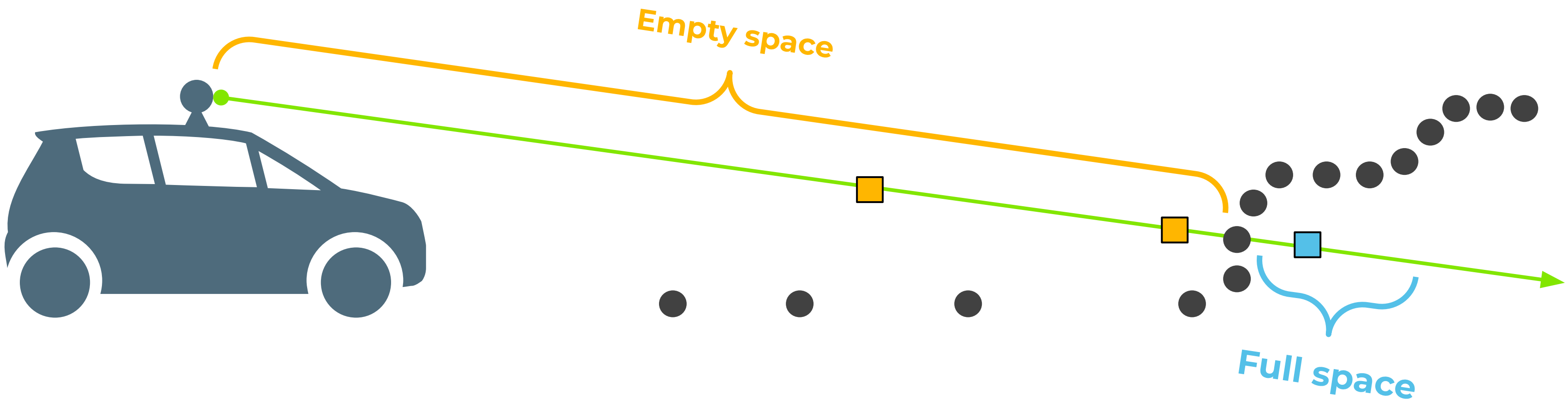
Supervision

ALSO



Self-supervised occupancy - Query point generation

ALSO



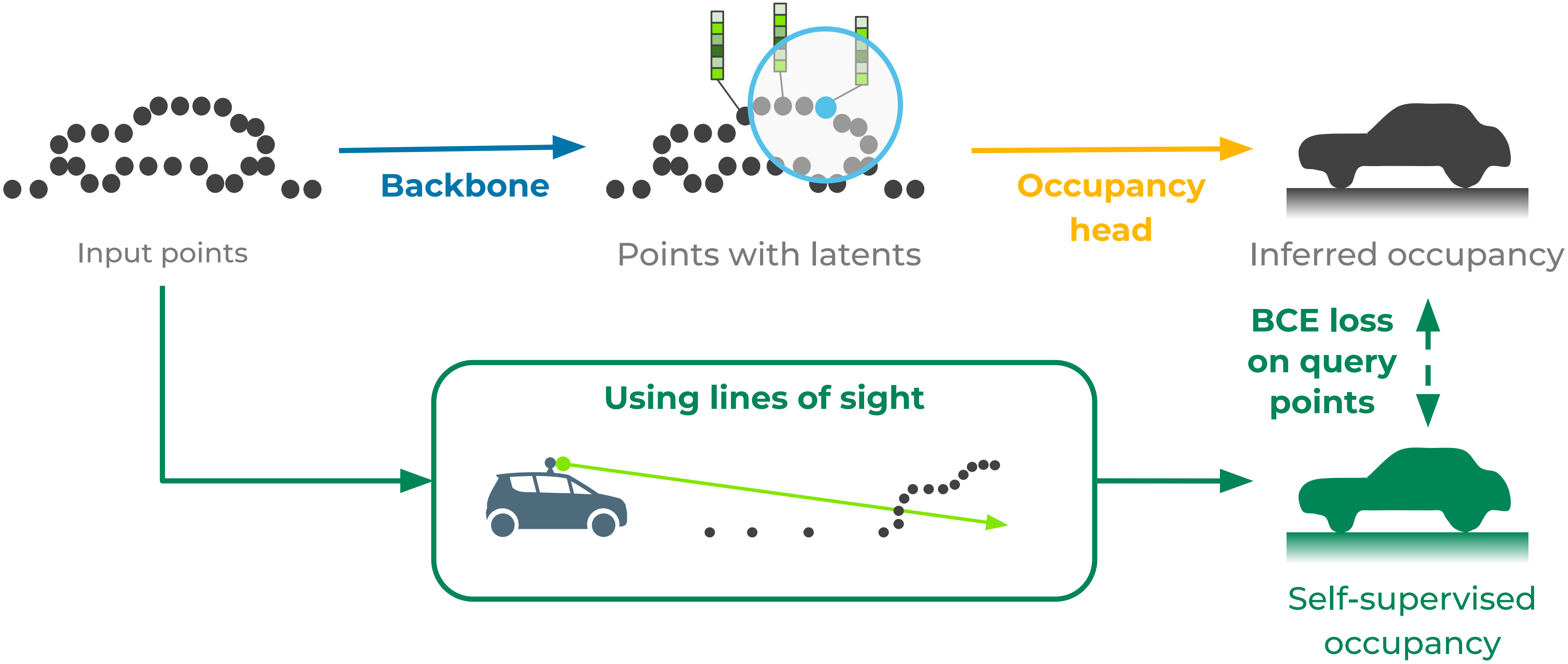
Along lidar lines of sight

- **Empty queries:** from sensor to observed point
- **Full queries:** just behind the point (max distance δ = 0.1 m)

Self-supervision

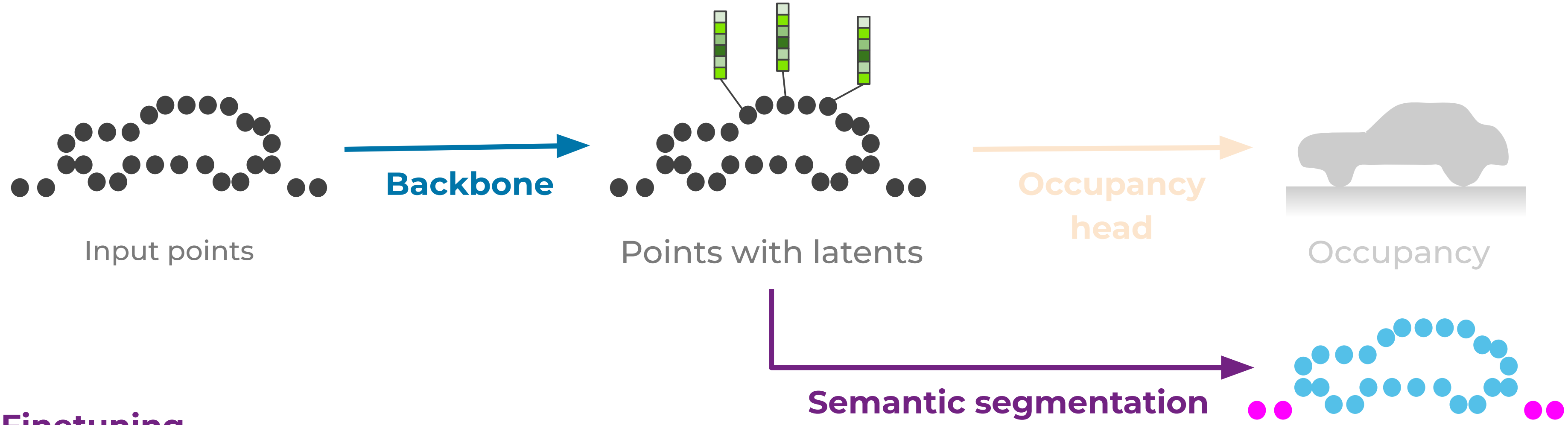
valeo.ai

ALSO



Downstream tasks

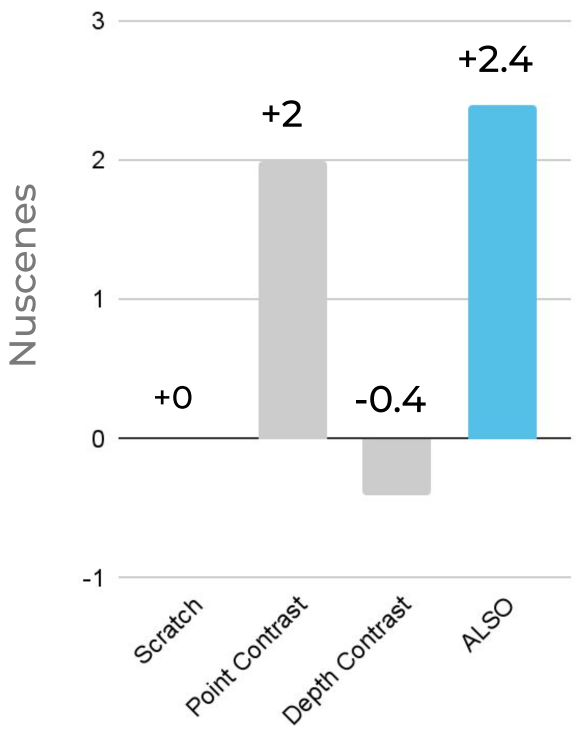
ALSO

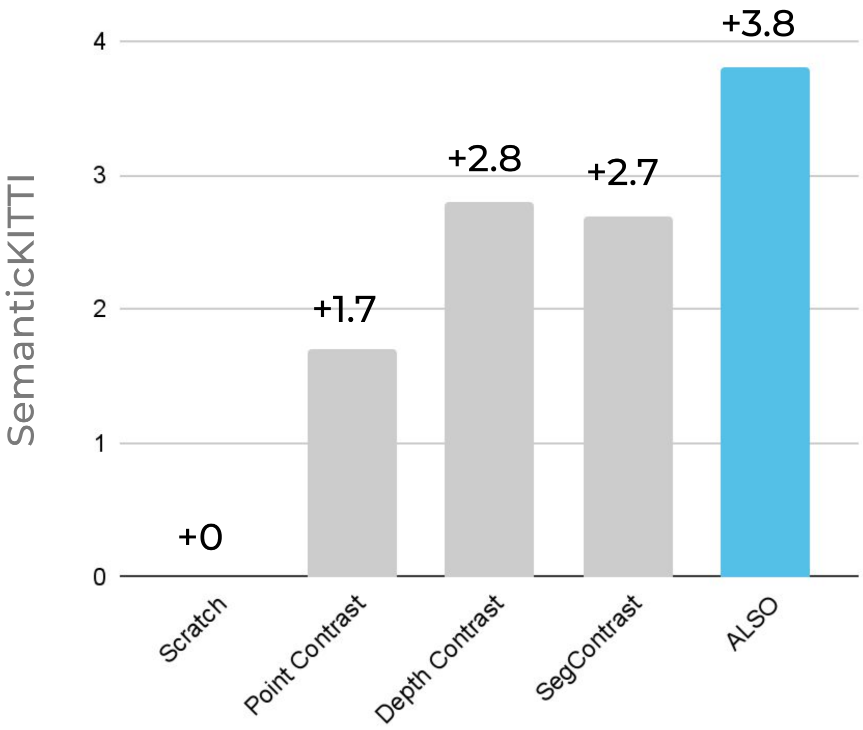


Finetuning
- remove occupancy head
- add a single linear layer
- finetune the whole network

Semantic segmentation - 1% annotated data

ALSO



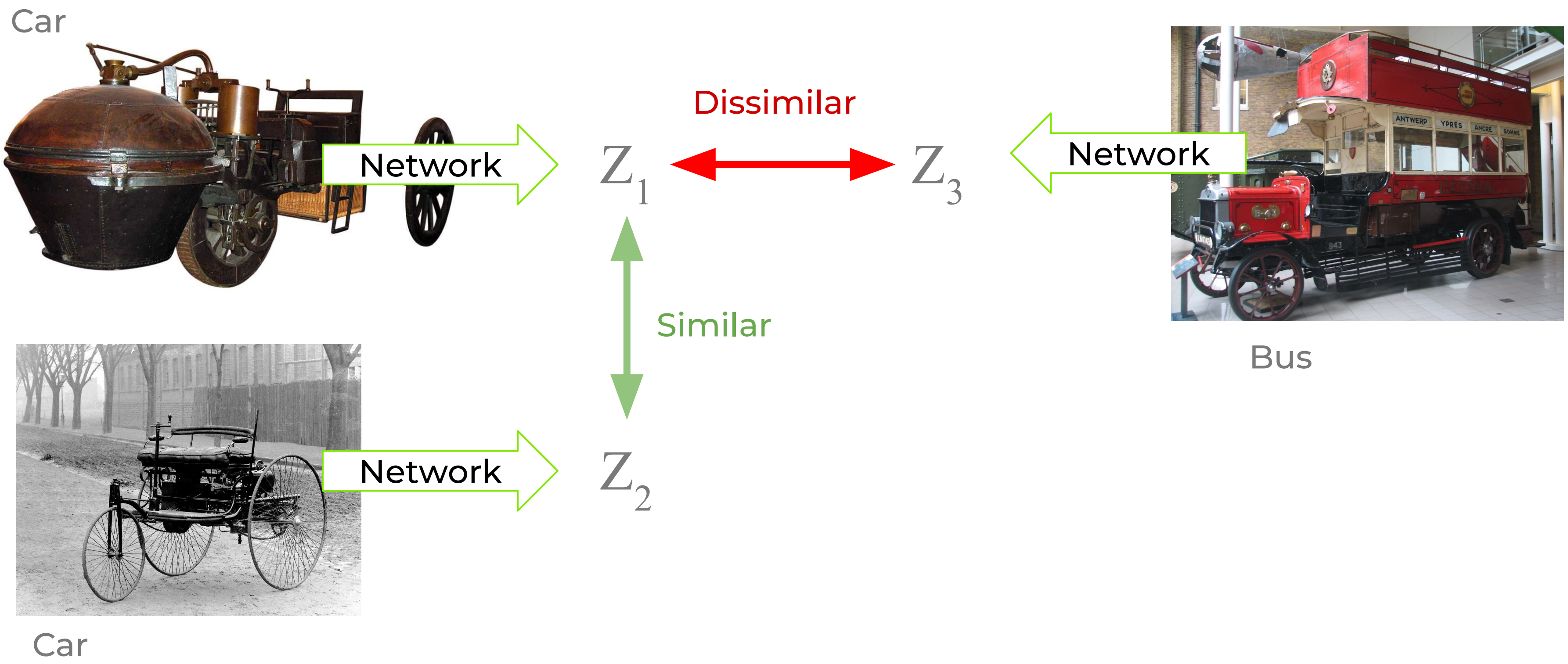


II - Self-supervision

Contrastive learning

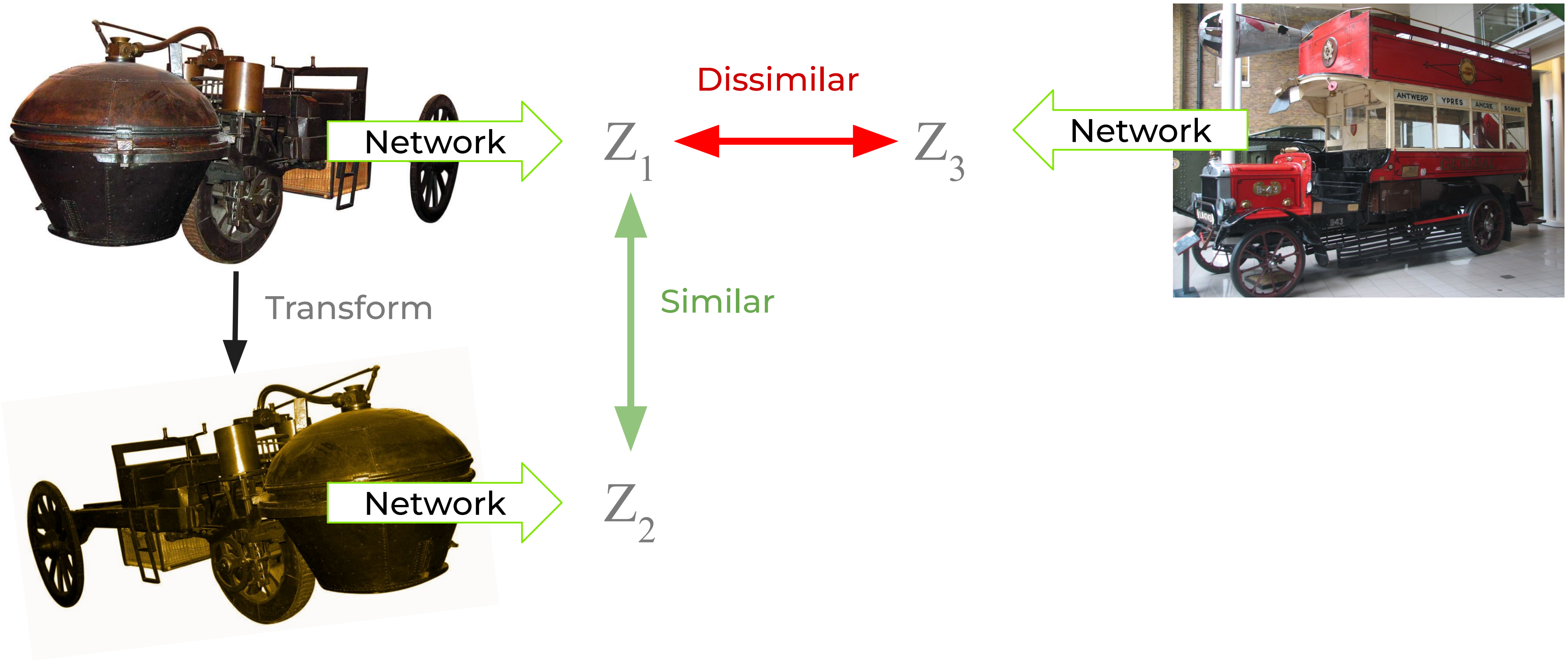
Contrastive learning

Contrastive methods



Contrastive learning

Contrastive methods



Contrastive learning

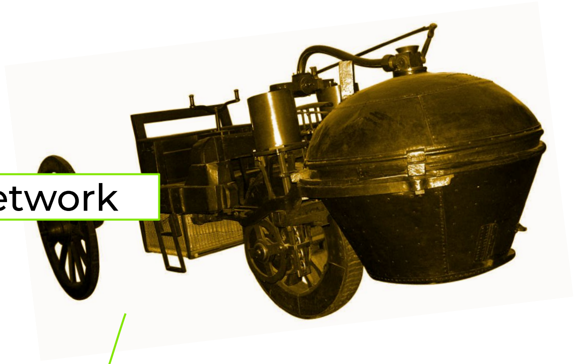
Contrastive methods

Network

q

k_+

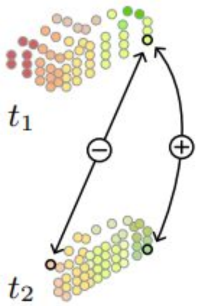
Network



$$\mathcal{L}_q = -\log \frac{\exp(q\cdot k_+/\tau)}{\sum_{i=0}^{K} \exp(q\cdot k_i/\tau)}$$

Contrastive self-supervised learning for point clouds

Contrastive methods



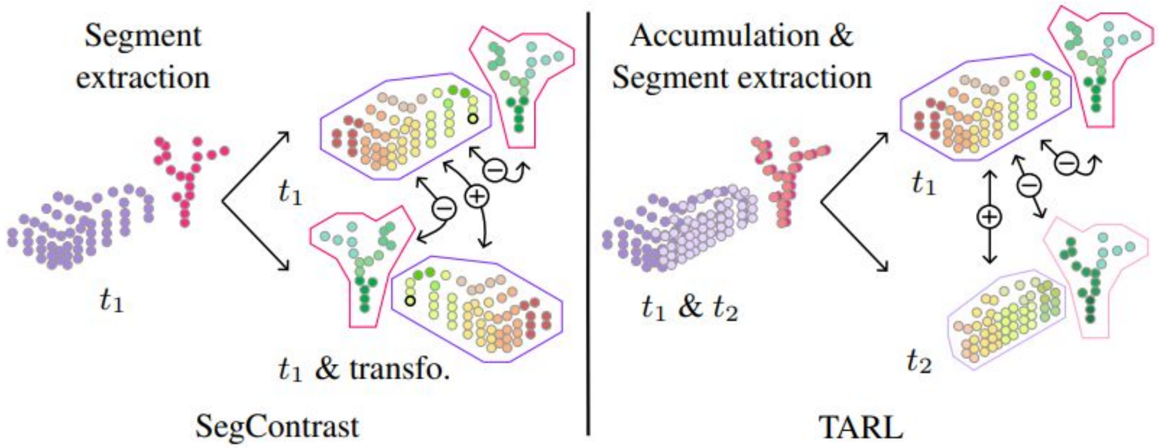
PointContrast

PointContrast

+++ Easy to implement
- - - Contrast in the same object

Contrastive self-supervised learning for point clouds

Contrastive methods

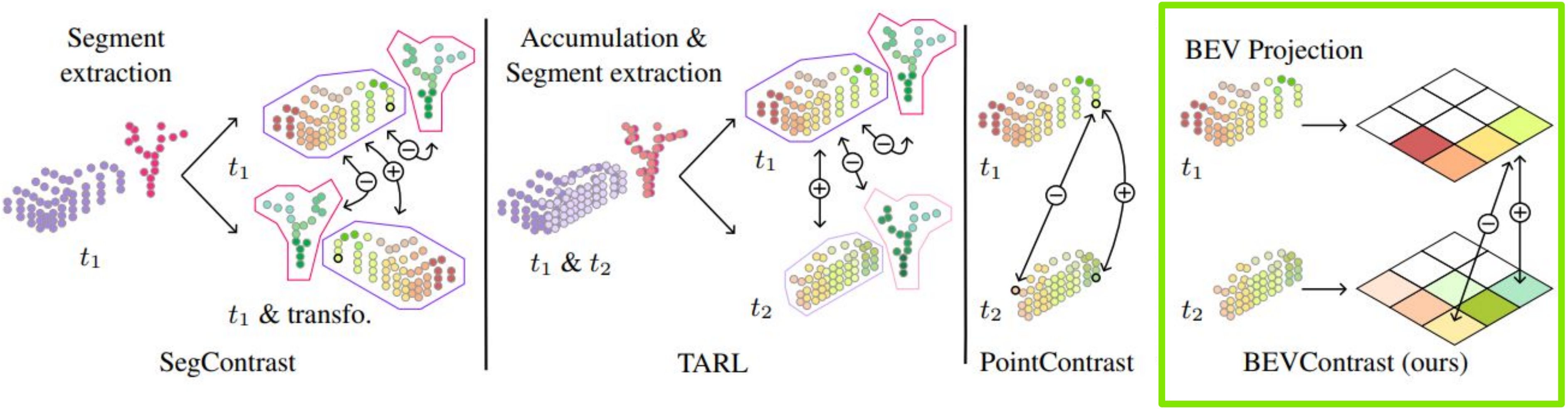


SegContrast / TARL

+++ Efficient thanks to object segmentation (temporal for TARL)
- - - Difficult to set up → rely on HDBScan (hyperparameters)

Contrastive self-supervised learning for point clouds

Contrastive methods

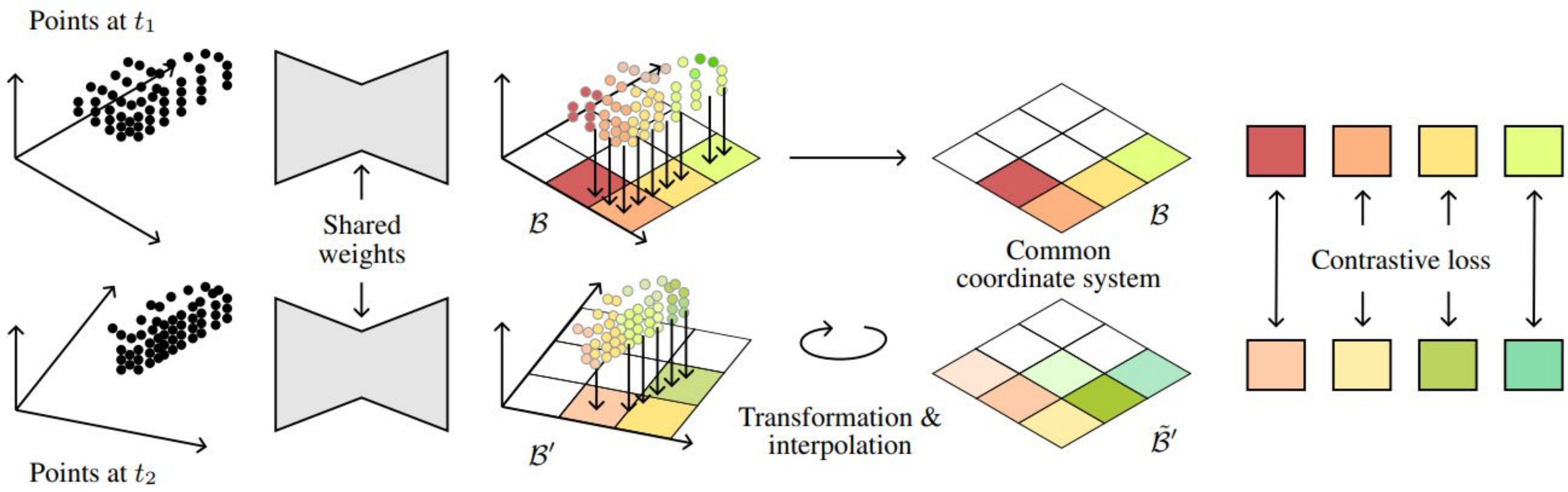


BEVContrast

+++ Simple → easy projection in BEV
+++ Object separation approximation with BEV cells

BEVContrast

Contrastive methods



BEVContrast

valeo.ai

Contrastive methods

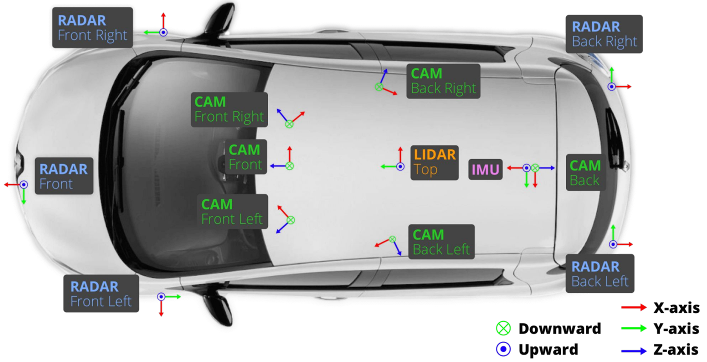
Dataset	Method	0.1%		1%		10%		50%		100%	
nuScenes	No pre-training	21.6	±0.5	35.0	±0.3	57.3	±0.4	69.0	±0.2	71.2	±0.2
	PointContrast† [40]	27.1	±0.5	37.0	±0.5	58.9	±0.2	69.4	±0.3	71.1	±0.2
	DepthContrast†[46]	21.7	±0.3	34.6	±0.5	57.4	±0.5	69.2	±0.3	71.2	±0.2
	ALSO [3]	26.2	±0.5	37.4	±0.3	59.0	±0.4	69.8	±0.2	71.8	±0.2
	BEVContrast (ours)	26.6	±0.5	37.9	±0.4	59.0	±0.6	70.5	±0.2	72.2	±0.1
SemanticKITTI	No pre-training	30.0	±0.2	46.2	±0.6	57.6	±0.9	61.8	±0.4	62.7	±0.3
	PointContrast‡ [40]	32.4	±0.5	47.9	±0.5	59.7	±0.5	62.7	±0.3	63.4	±0.4
	SegContrast [29]	32.3	±0.3	48.9	±0.3	58.7	±0.5	62.1	±0.4	62.3	±0.4
	DepthContrast† [46]	32.5	±0.4	49.0	±0.4	60.3	±0.5	62.9	±0.5	63.9	±0.4
	STSSL [39]	32.0	±0.4	49.4	±1.1	60.0	±0.6	62.9	±0.7	63.3	±0.3
	ALSO [3]	35.0	±0.1	50.0	±0.4	60.5	±0.1	63.4	±0.5	63.6	±0.5
	TARL [30]	37.9	±0.4	52.5	±0.5	61.2	±0.3	63.4	±0.2	63.7	±0.3
	BEVContrast (ours)	39.7	±0.9	53.8	±1.0	61.4	±0.4	63.4	±0.6	64.1	±0.4

II - Self-supervision

Distillation

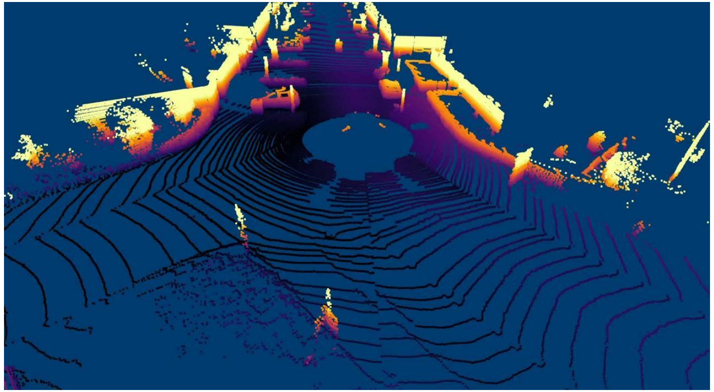
valeo.ai

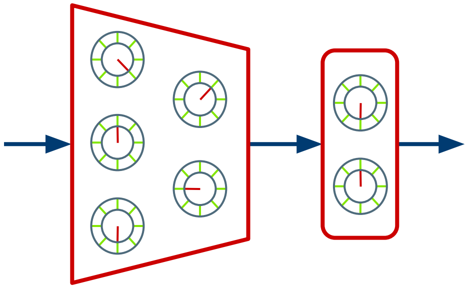
Sensor setup



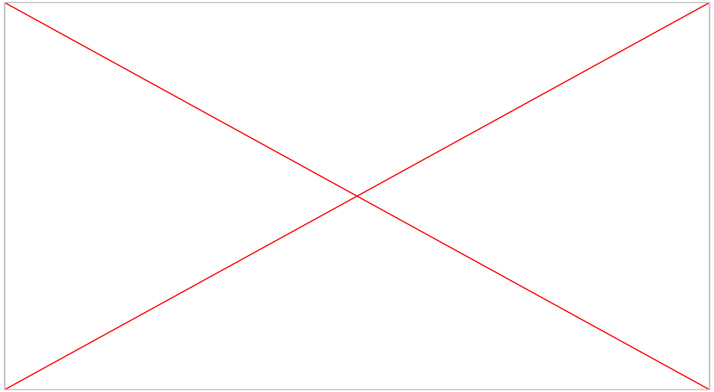
Example from nuScenes

Task of interest: point cloud semantic segmentation



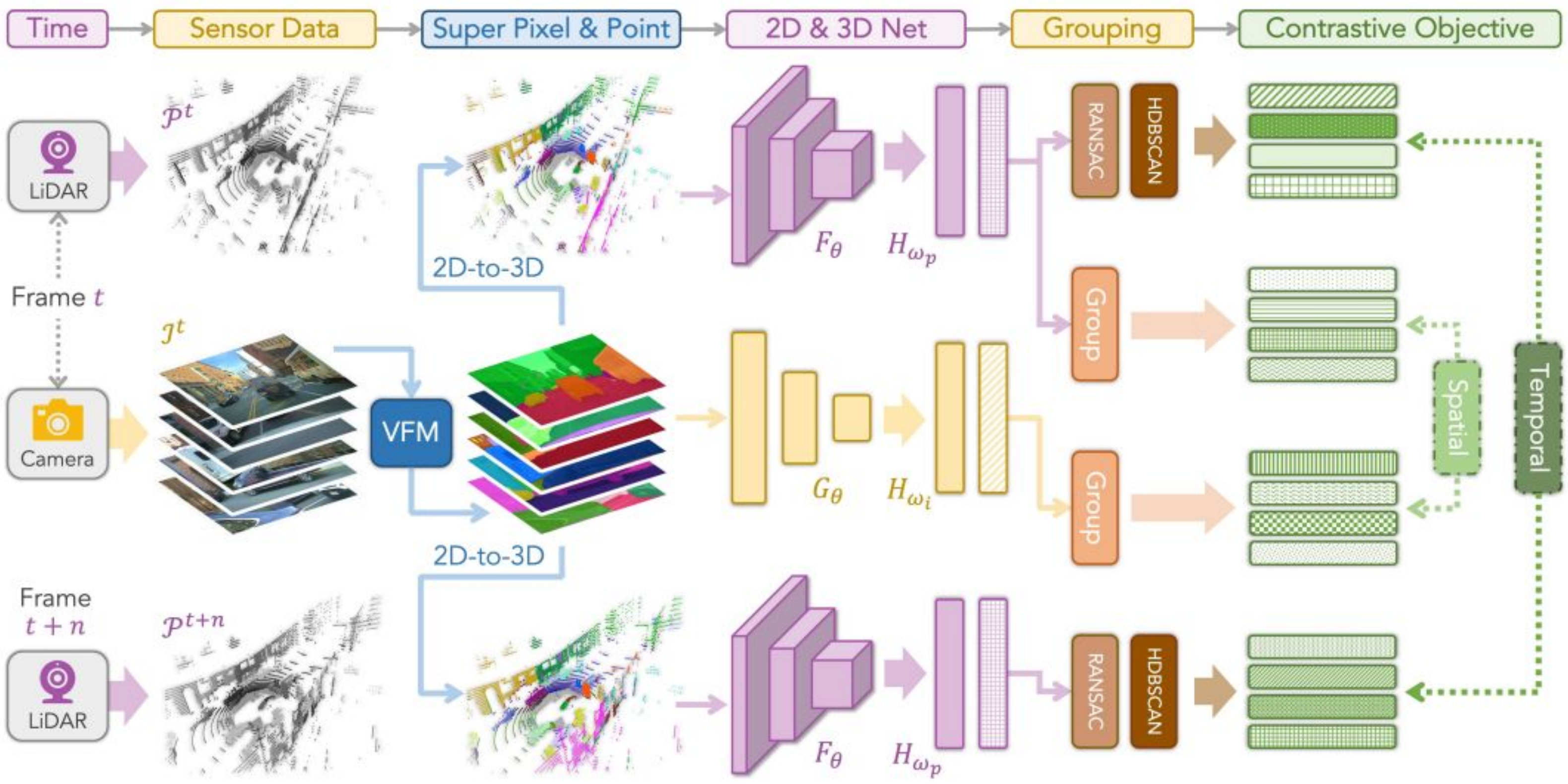


Example from SemanticKITTI



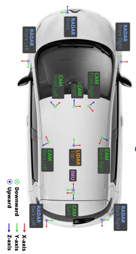
Credit Gilles Puy

SEAL - Segment any point cloud



Segment Any Point Cloud Sequences by Distilling Vision Foundation Models, Liu et al.

ScaLR



nuScenes

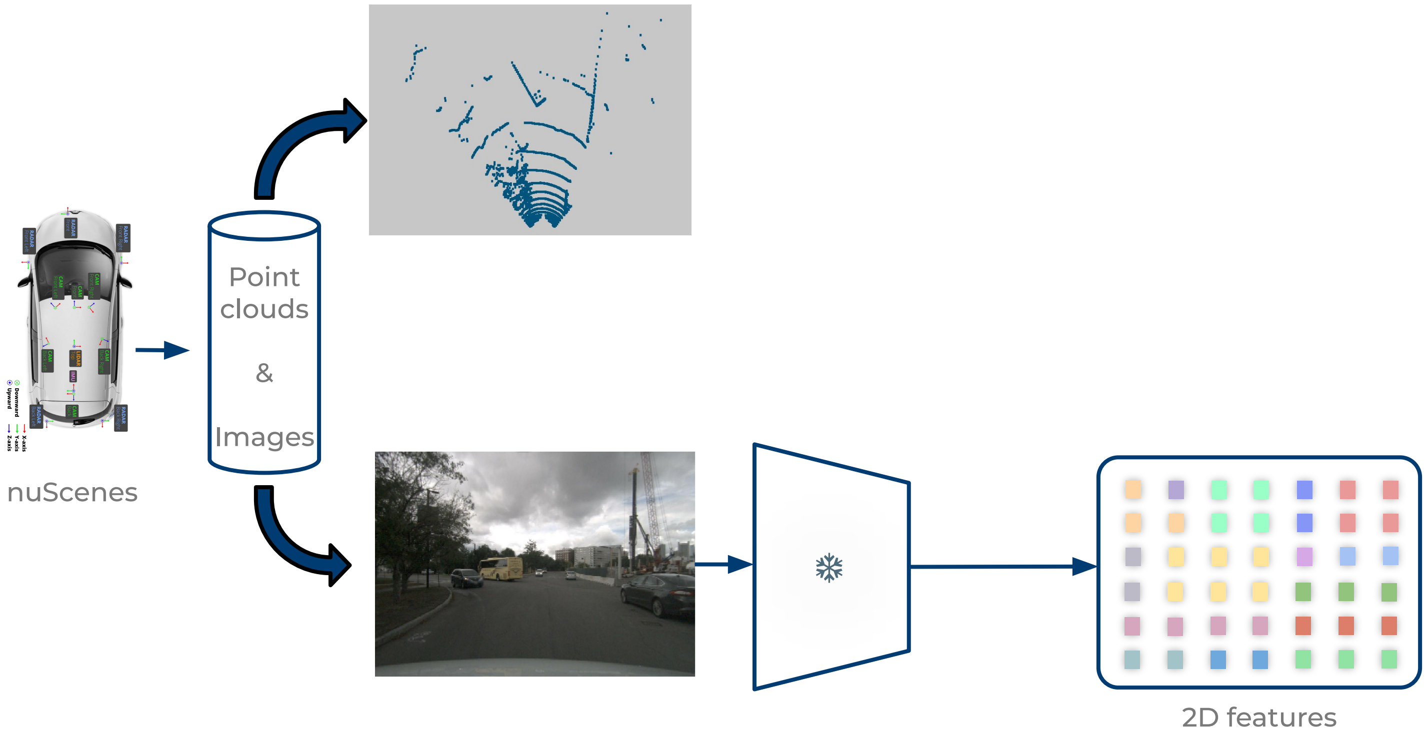
Point clouds

&

Images

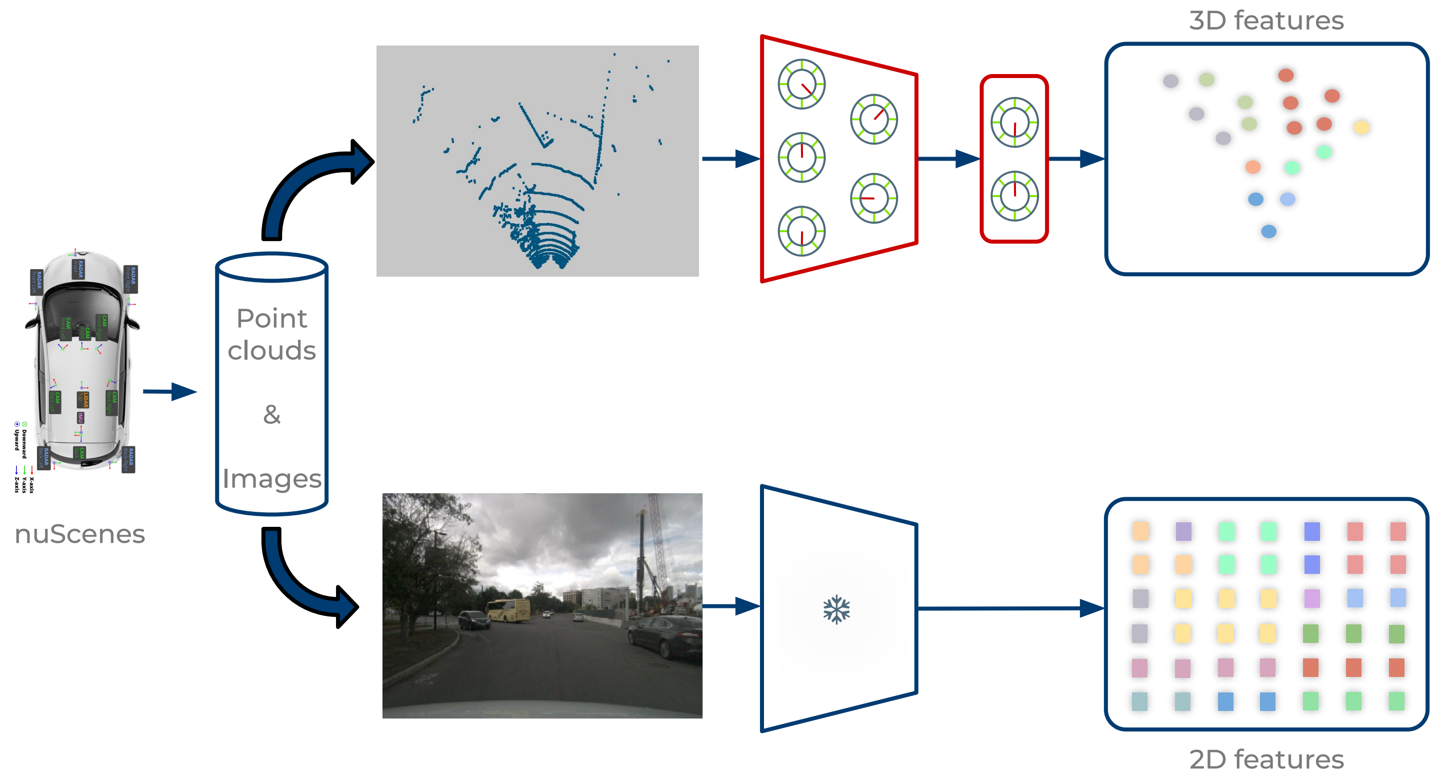
Credit Gilles Puy

ScaLR



Credit Gilles Puy

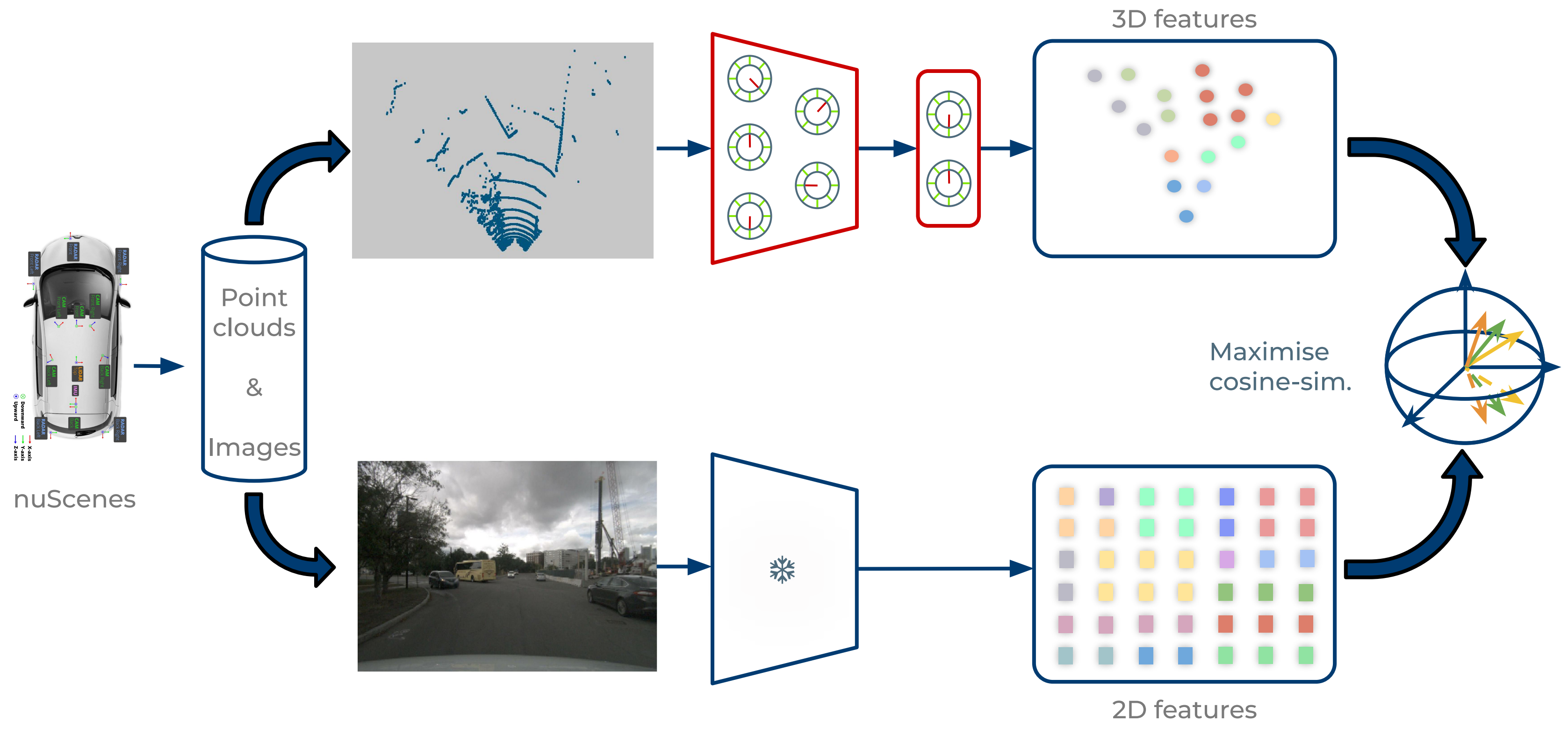
ScaLR



Credit Gilles Puy

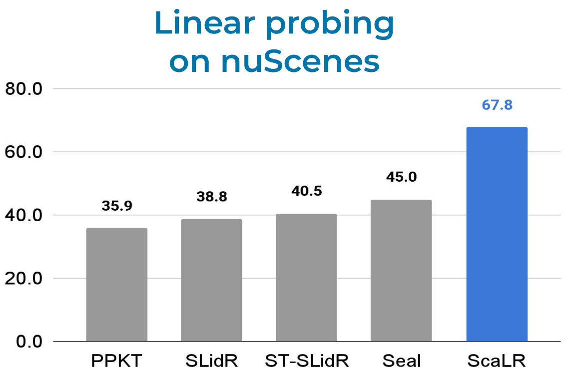
ScaLR

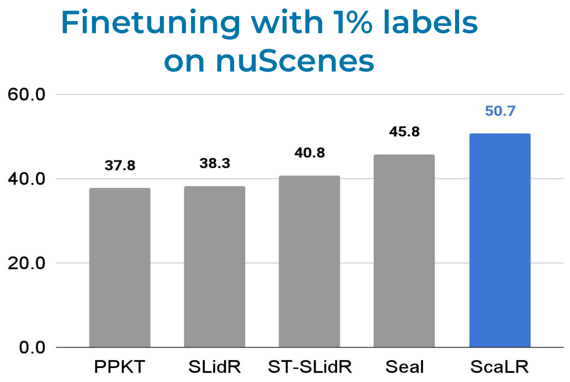
valeo.ai

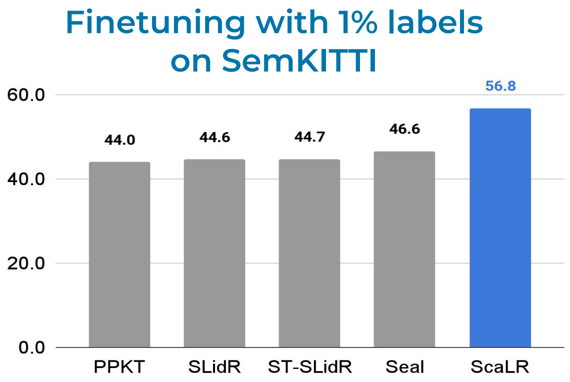


Credit Gilles Puy

ScaLR



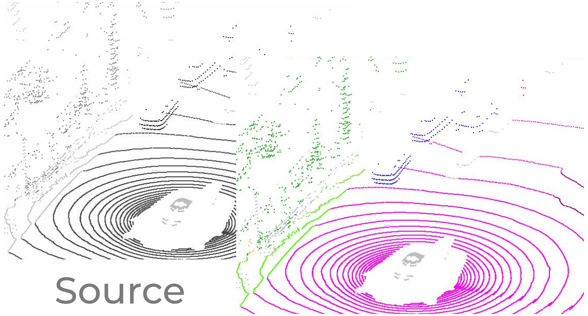




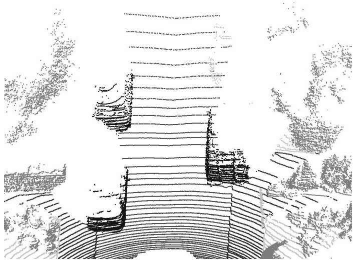
Credit Gilles Puy

III - Domain adaptation

Unsupervised domain adaptation



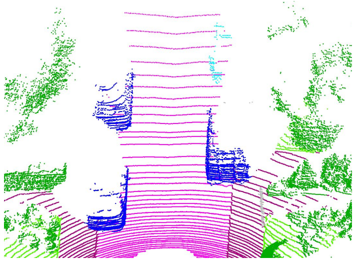
Source
(with annotations)



Target
(no annotations)

Network

Objective:
predict on target



Domain gap

Different sensors

Different locations

Different objects

⇒ source model performs poorly

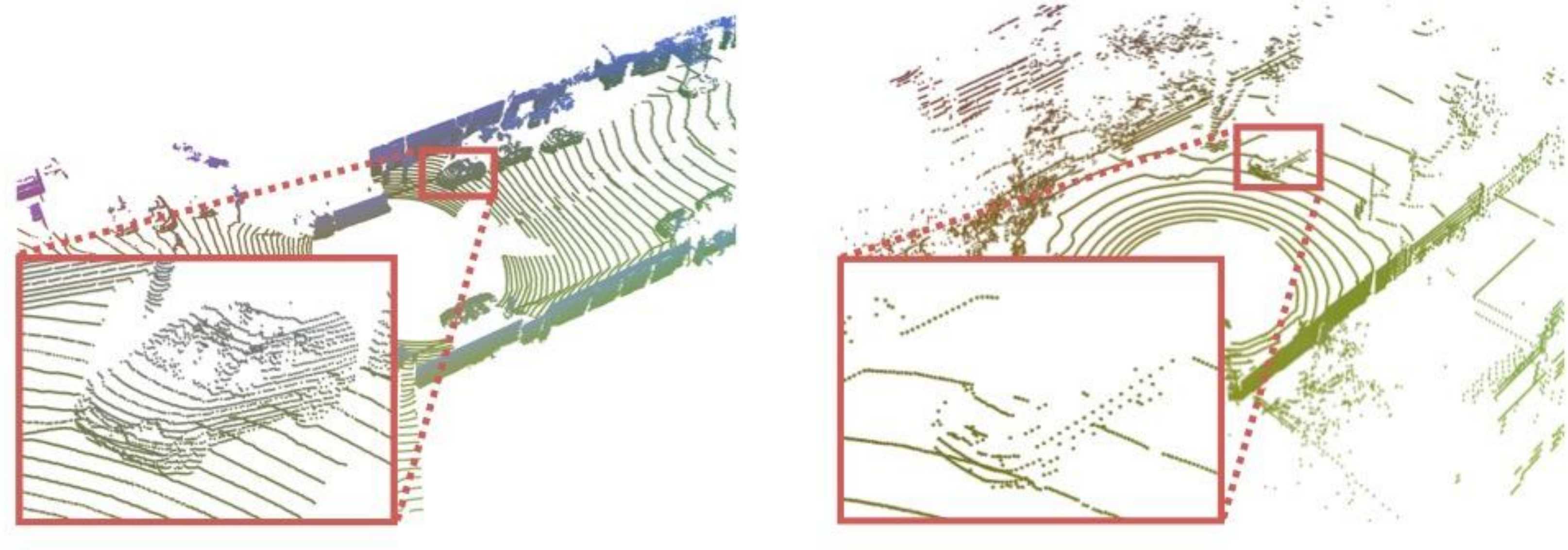
Adaptation

Use target data to enhance performances

⇒ prevent collapse

Example - sensor gap

Domain adaptation

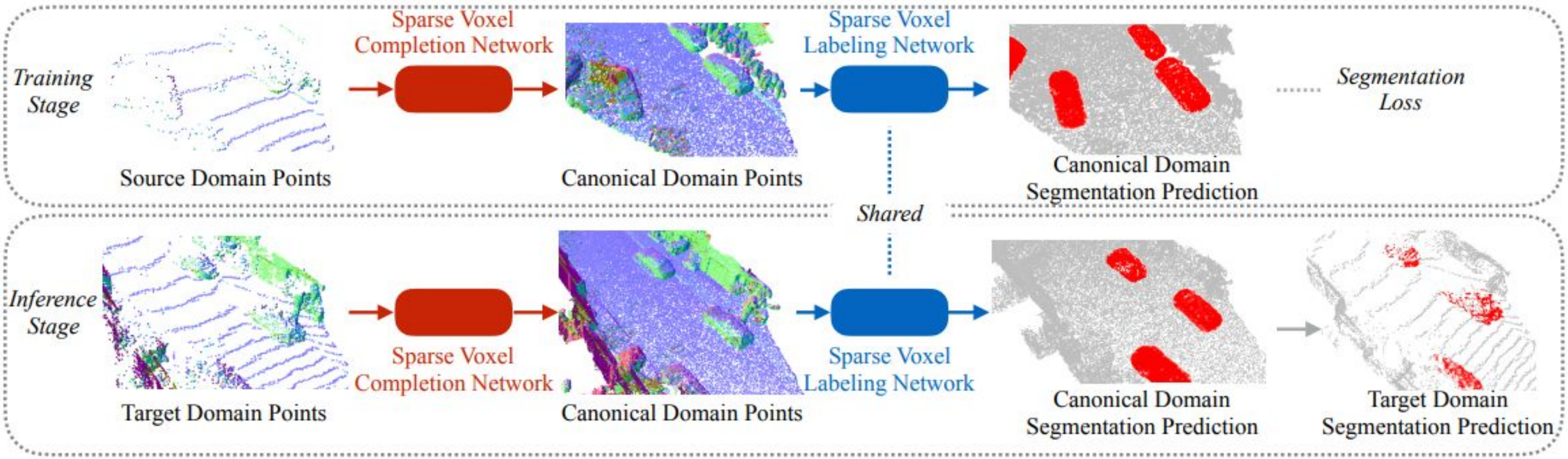


(a) captured by a 64-beam LiDAR (b) captured by a 32-beam LiDAR

(Yi et al. CVPR 2021)

Complete and label

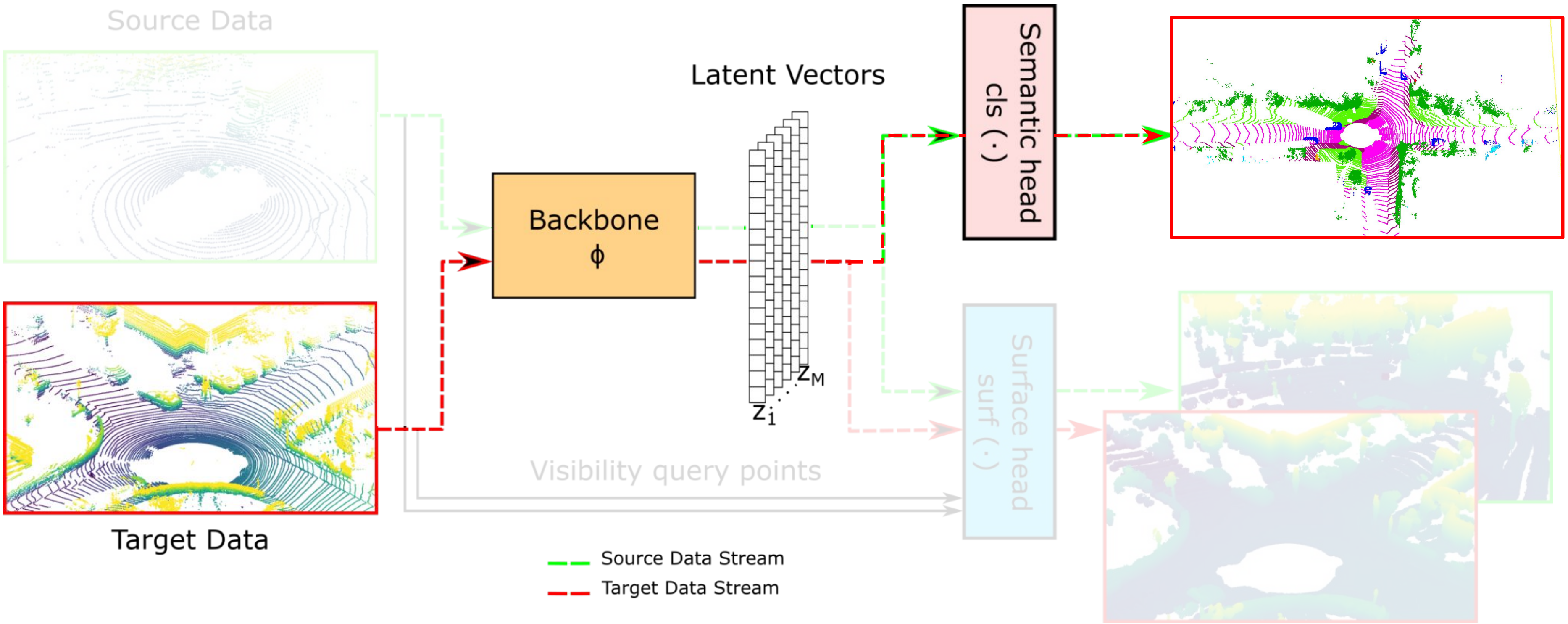
Domain adaptation



(Yi et al. CVPR 2021)

SALUDA

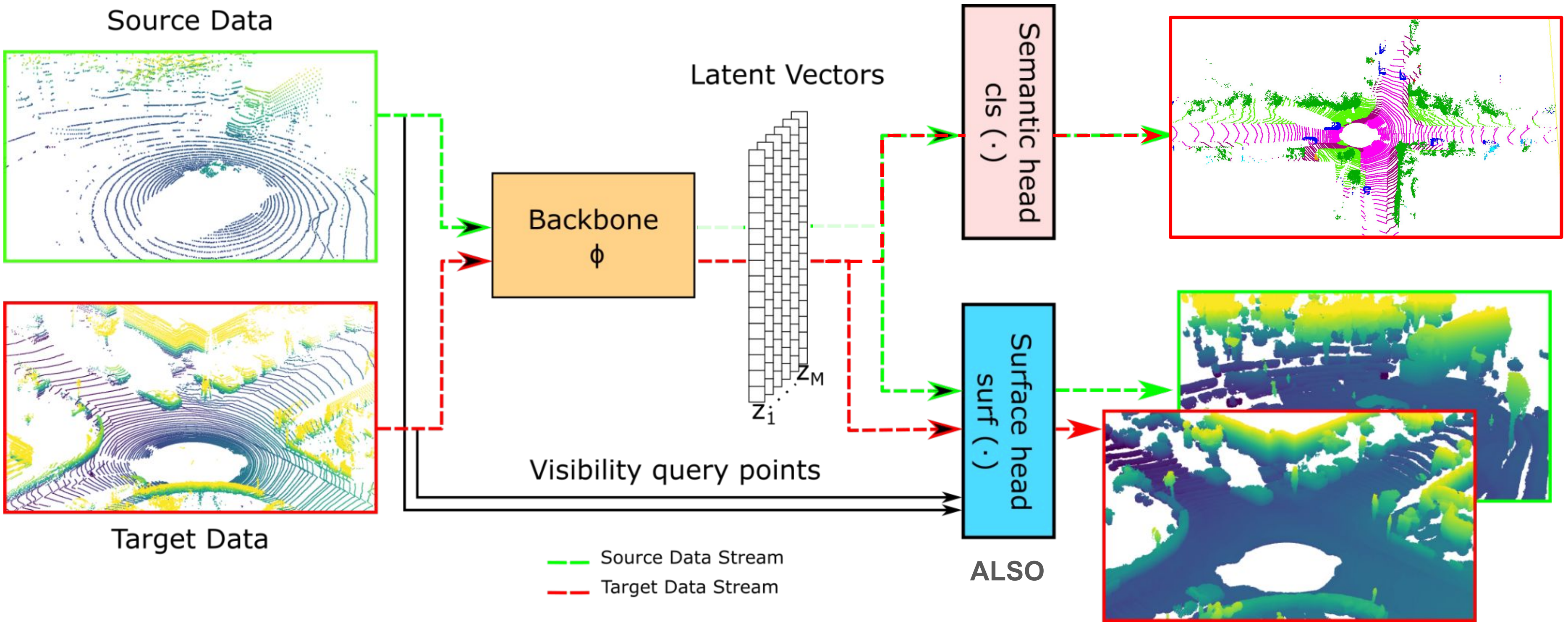
Domain adaptation



Michele et al.

SALUDA

Domain adaptation



Michele et al.

III - Open World

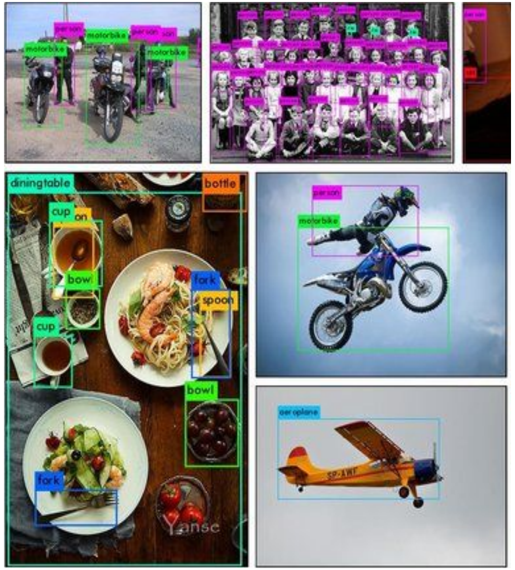
Open vocabulary

Classic benchmarks

- **Closed vocabulary** setup
- With definition of a **finite set** of classes
- Training of model in **fully-supervised** fashion
  - Requires a lot of annotation

Cityscapes dataset, **30 classes**

[road, sidewalk, parking, rail track, person, rider, car, truck, bus, on rails, motorcycle, bicycle, caravan, trailer, building, wall, fence, guard rail, bridge, tunnel, pole, traffic sign, traffic light, vegetation, terrain, sky]



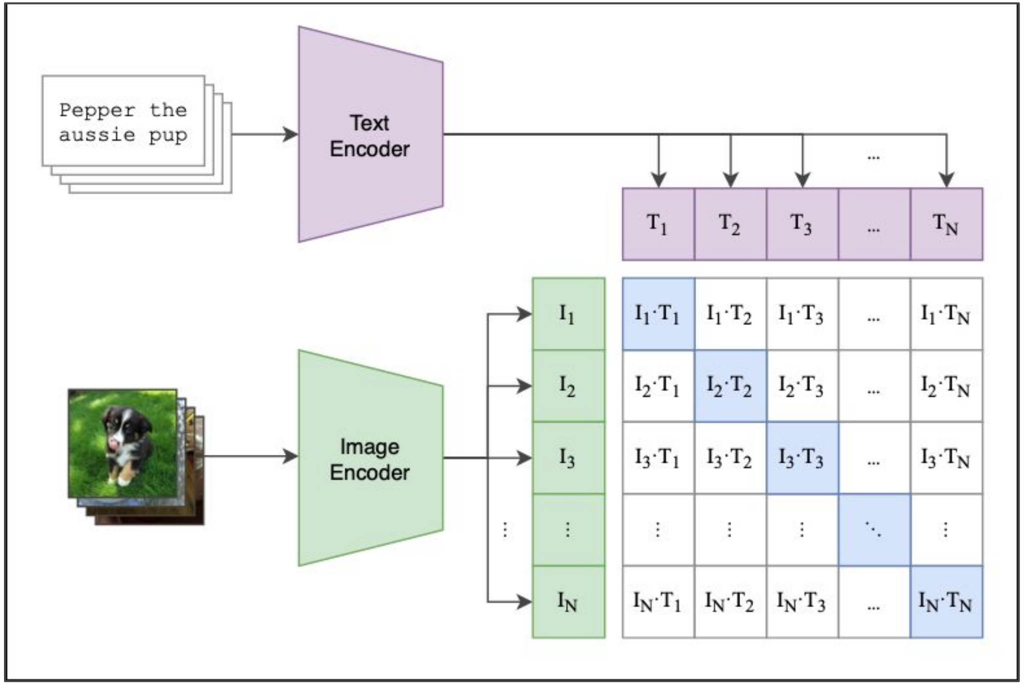
Pascal VOC dataset, **20 classes**

[person, bird, cat, cow, dog, horse, sheep, aeroplane, bicycle, boat, bus, car, motorbike, train, bottle, chair, dining table, potted plant, sofa, tv/monitor] valeo.ai

Open vocabulary

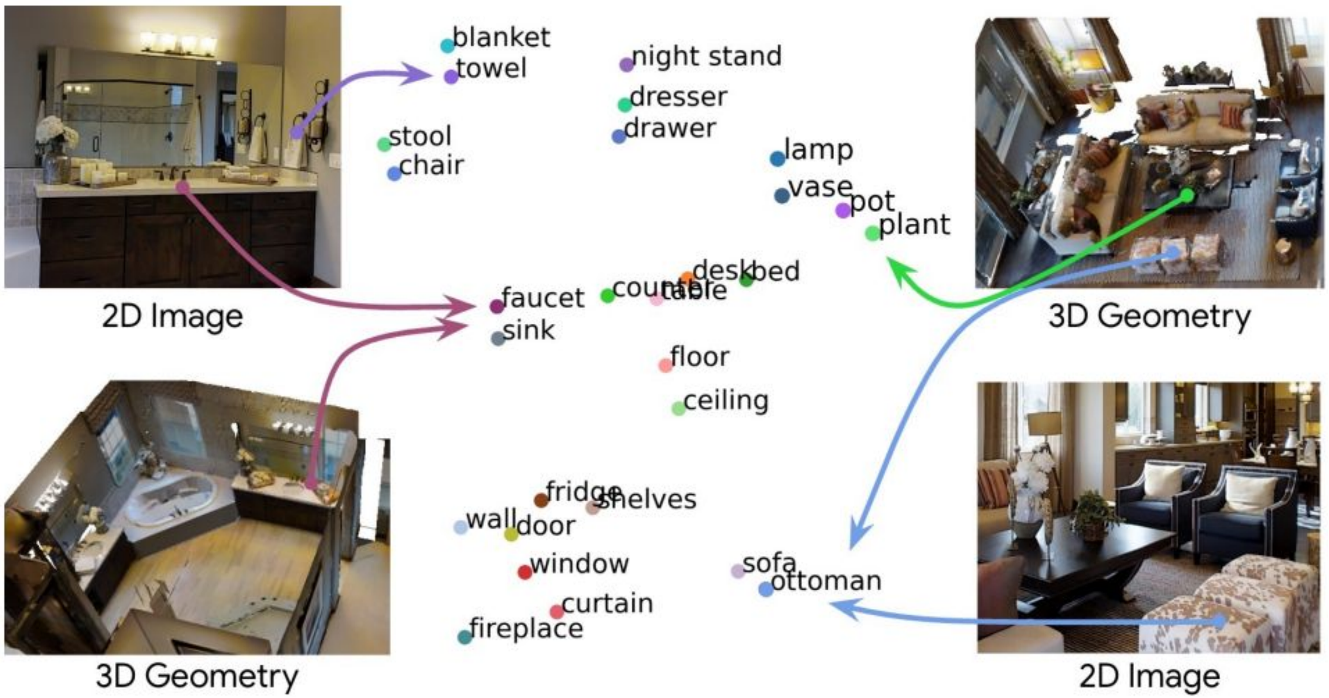
The world is complex

CLIP



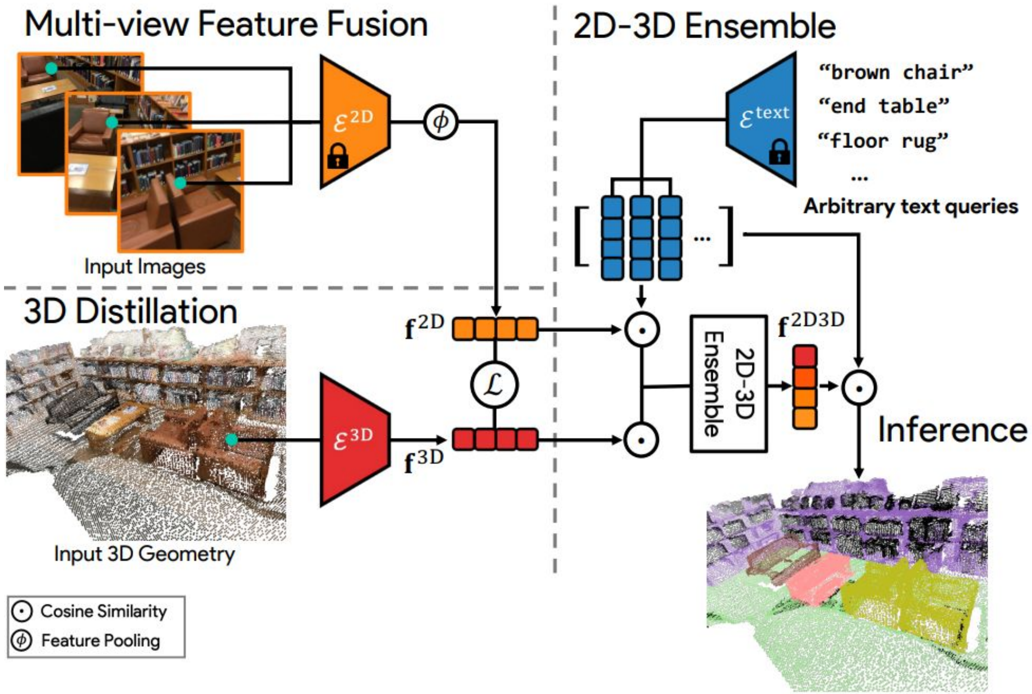
- Contrastive learning
  - Contrast positive/negative pairs
- Trained using 400 millions (**image / text**) pairs extracted from internet
  - Meta-data
  - Legends

OpenScene



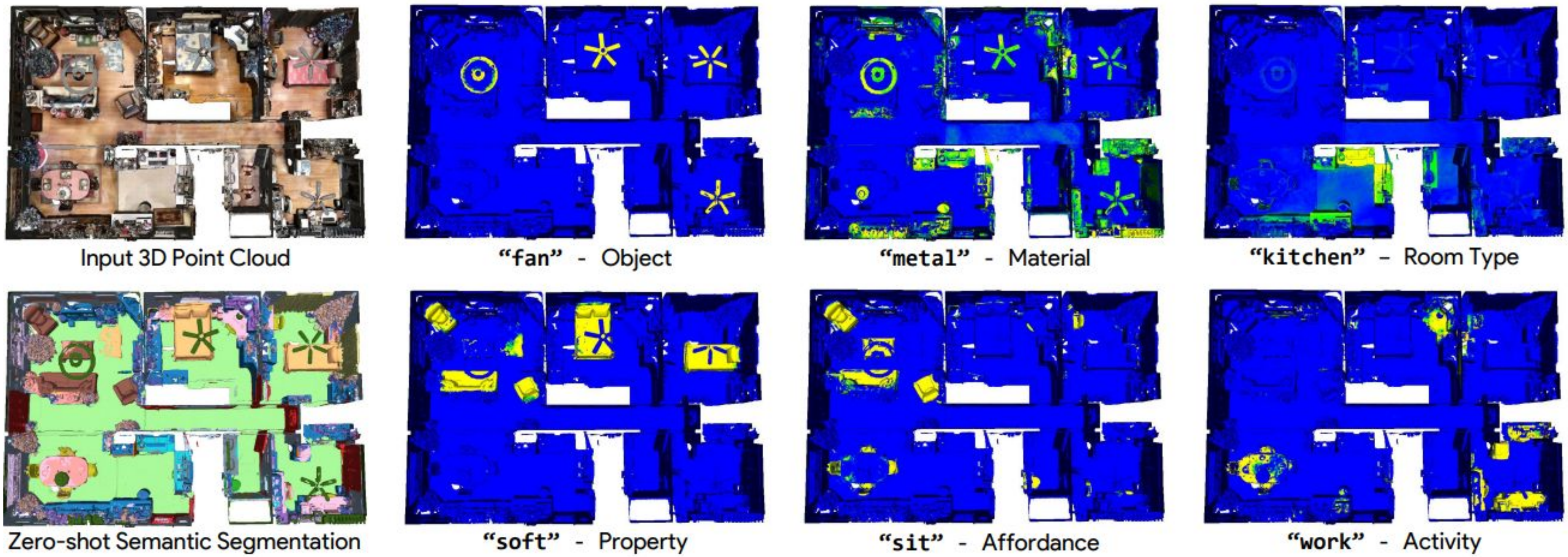
Peng, Songyou, et al. "Openscene: 3d scene understanding with open vocabularies." CVPR. 2023.

OpenScene



Peng, Songyou, et al. "Openscene: 3d scene understanding with open vocabularies." CVPR. 2023.

OpenScene



Peng, Songyou, et al. "Openscene: 3d scene understanding with open vocabularies." CVPR. 2023.

OpenScene

Link to video

https://pengsongyou.github.io/openscene

Peng, Songyou, et al. "Openscene: 3d scene understanding with open vocabularies." CVPR. 2023.

Conclusion and practical session

Conclusion

A very short overview of some tasks

- There are many other
- Only few methods were presented, not state-of-the-art anymore

Practical session

Open Vocabulary on point cloud with MaskCLIP

Peng, Songyou, et al. "Openscene: 3d scene understanding with open vocabularies." CVPR. 2023.