ALSO

Automotive Lidar Self-Supervision by Occupancy Estimation

Alexandre Boulch, Corentin Sautier, Björn Michele, Gilles Puy, Renaud Marlet

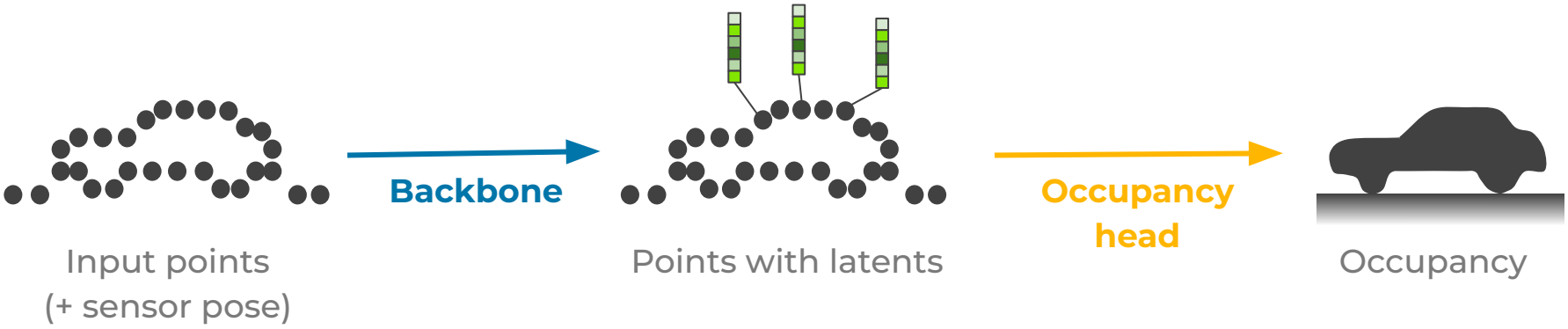
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Valeo

Gaspard · Monge
Laboratoire d'informatique

valeo.ai

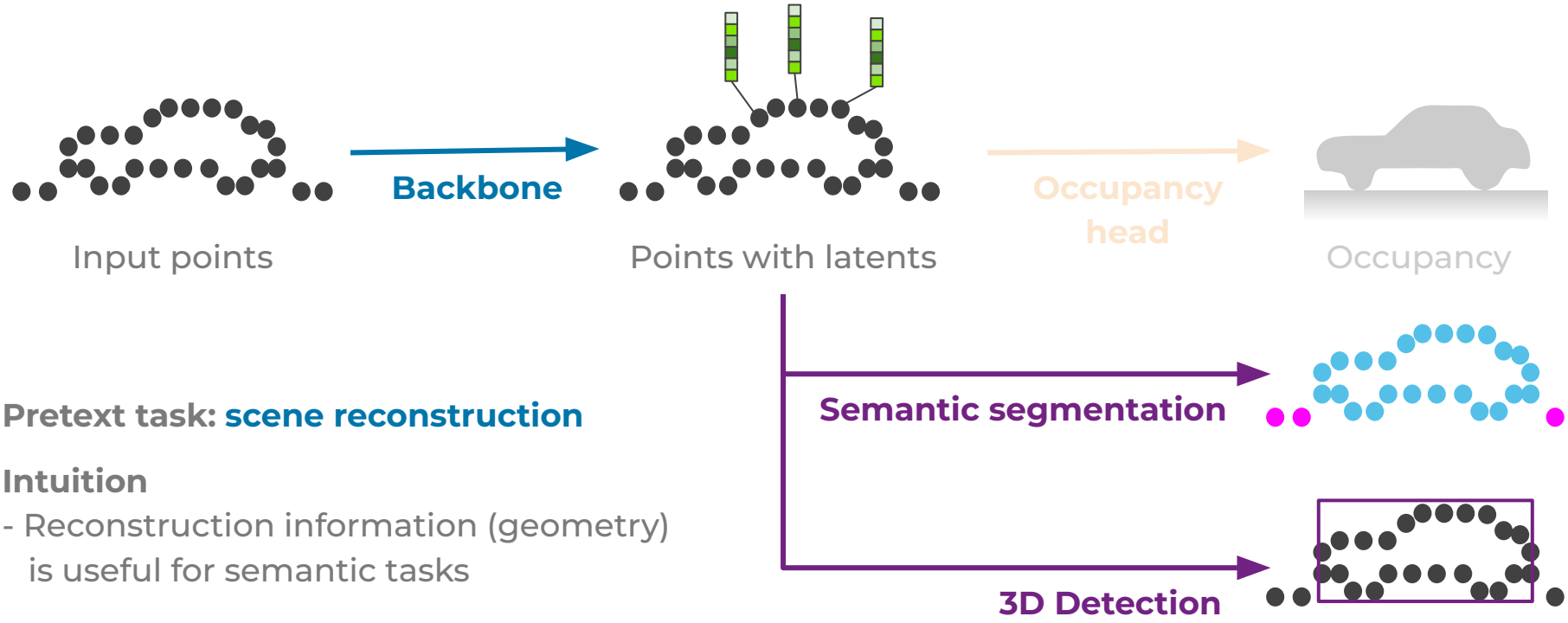
ALSO: Automotive Lidar Self-Supervision by Occupancy Estimation



Pretext task: scene reconstruction

- Implicit representation of surface (latent vectors)
- Occupancy decoding

ALSO: Automotive Lidar Self-Supervision by Occupancy Estimation

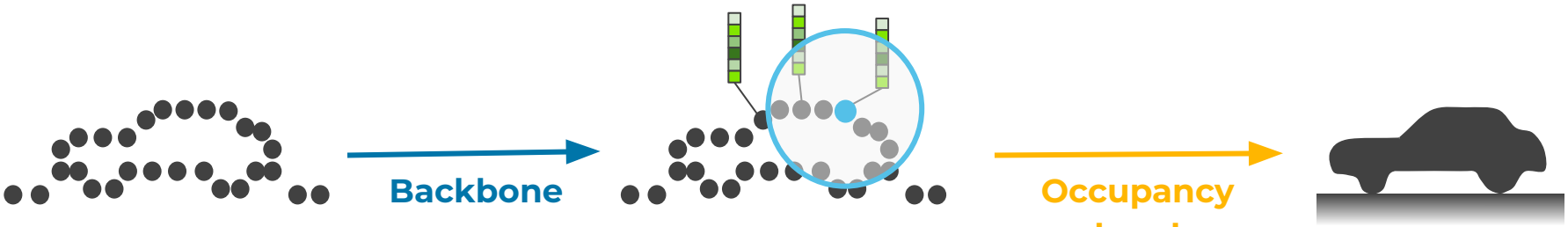


Pretext task: scene reconstruction

Intuition

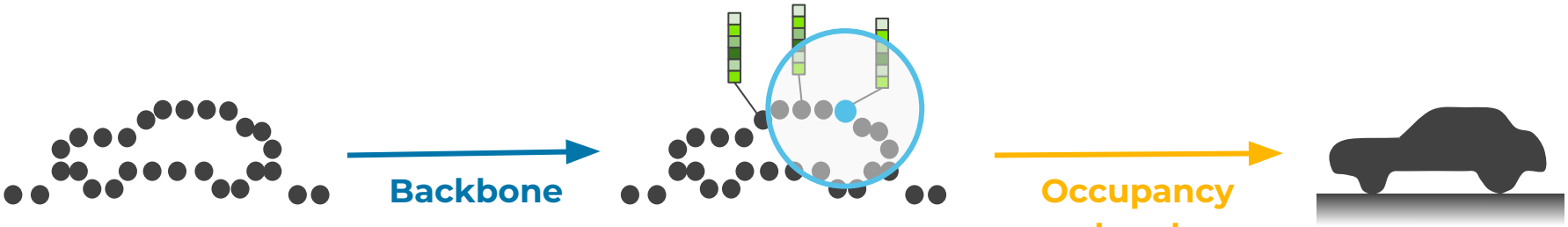
- Reconstruction information (geometry) is useful for semantic tasks

Context reconstruction vs local reconstruction



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

Context reconstruction vs local reconstruction



Local reconstruction [POCO head]

- everywhere, from features of neighboring points

⇒ (too) detailed geometry

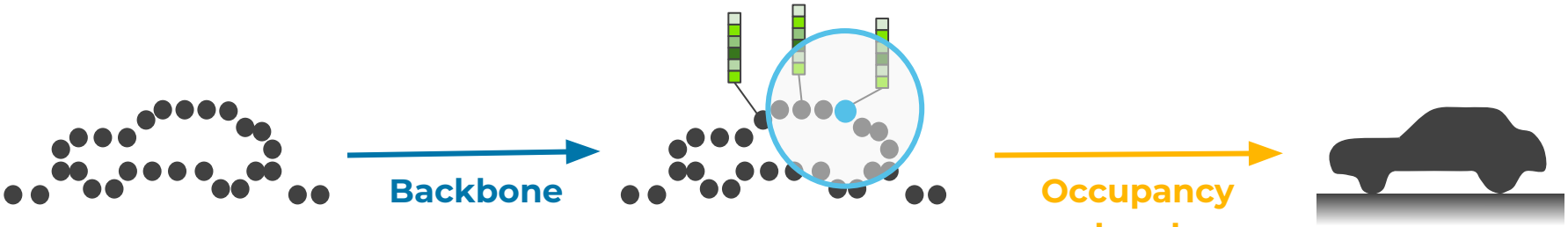
Context reconstruction [ALSO head]

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

⇒ rough geometry, more suited for object recognition

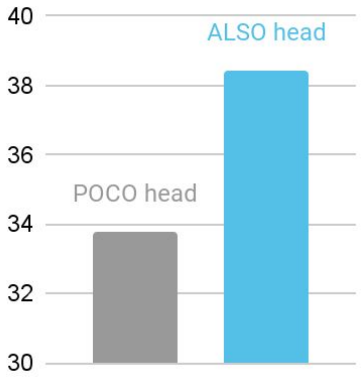
POCO: Point Convolution for Surface Reconstruction, A. Boulch, R. Marlet, CVPR 2022

Context reconstruction vs local reconstruction



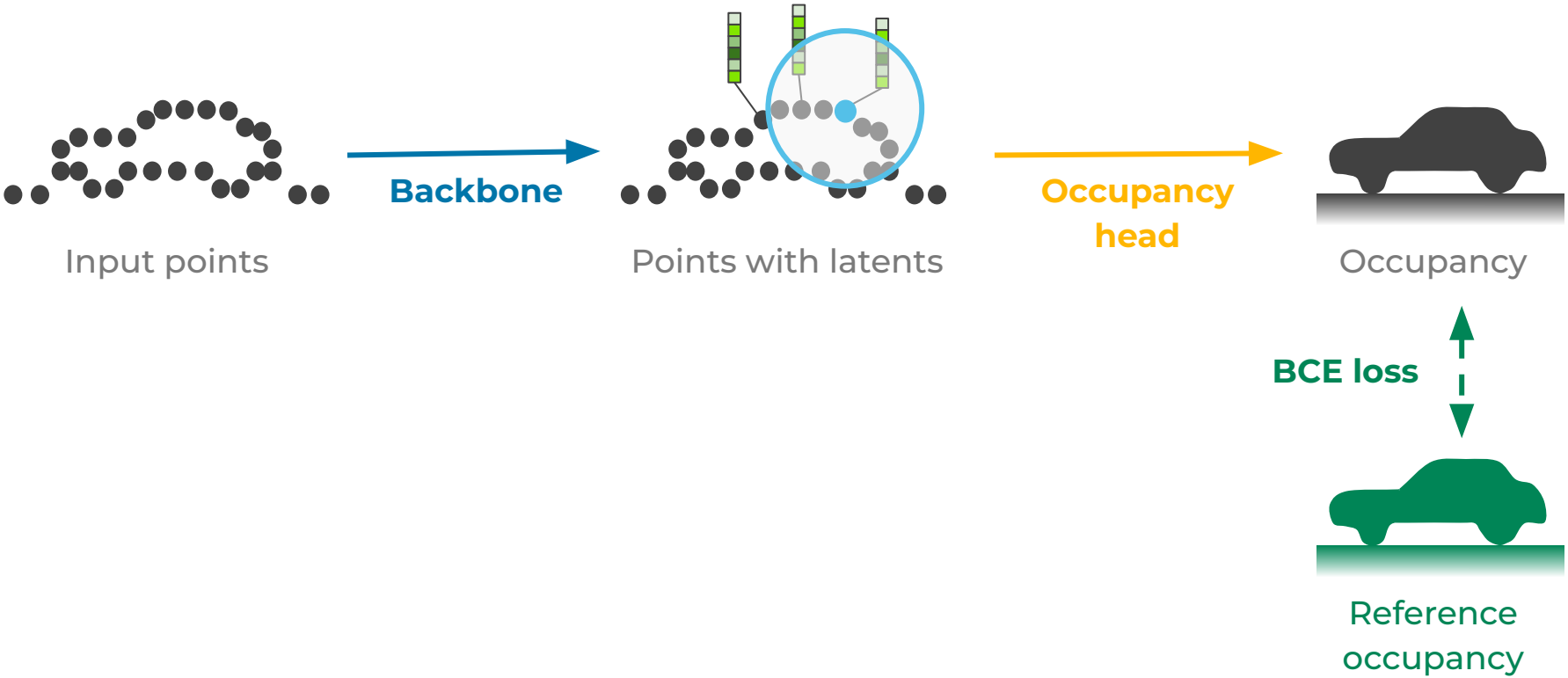
Local reconstruction [POCO head]
- everywhere, from features of neighboring points
⇒ (too) detailed geometry

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

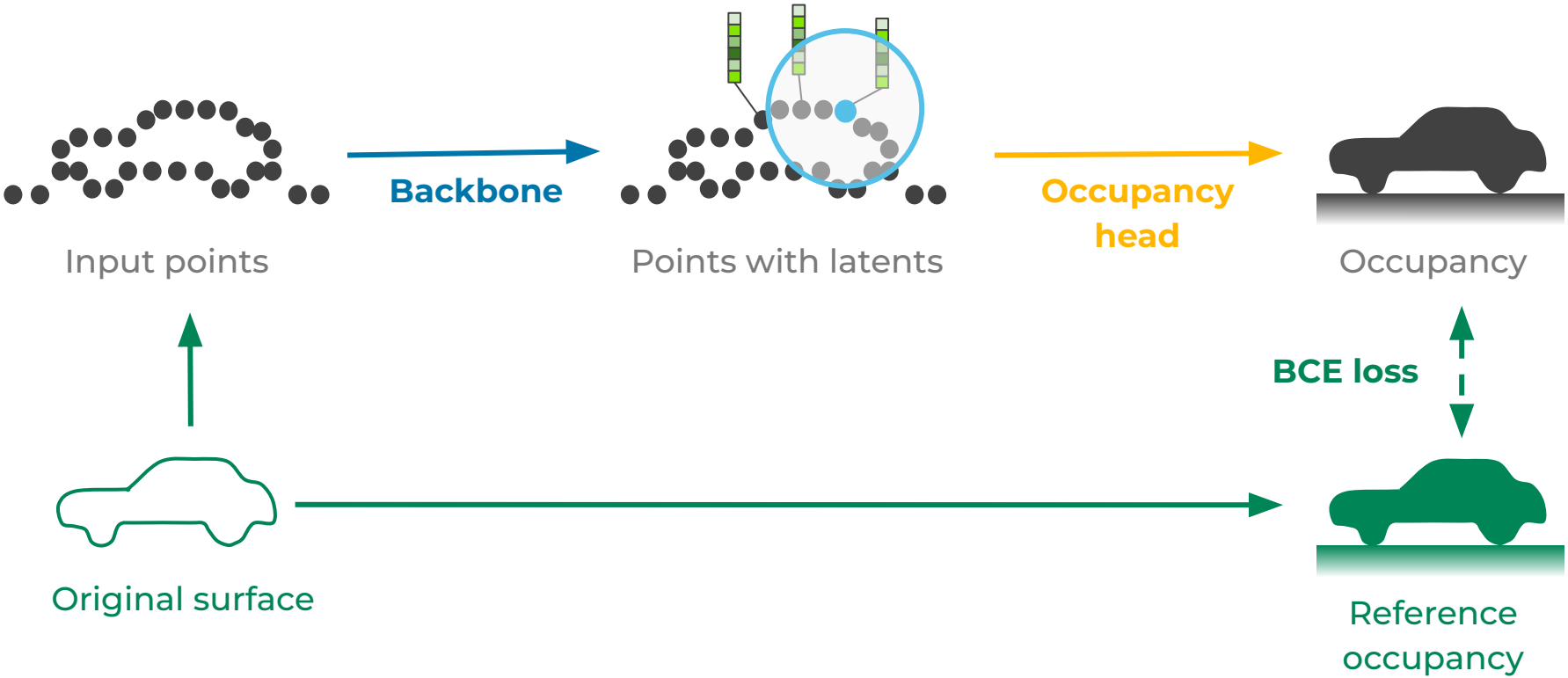


POCO: Point Convolution for Surface Reconstruction, A. Boulch, R. Marlet, CVPR 2022

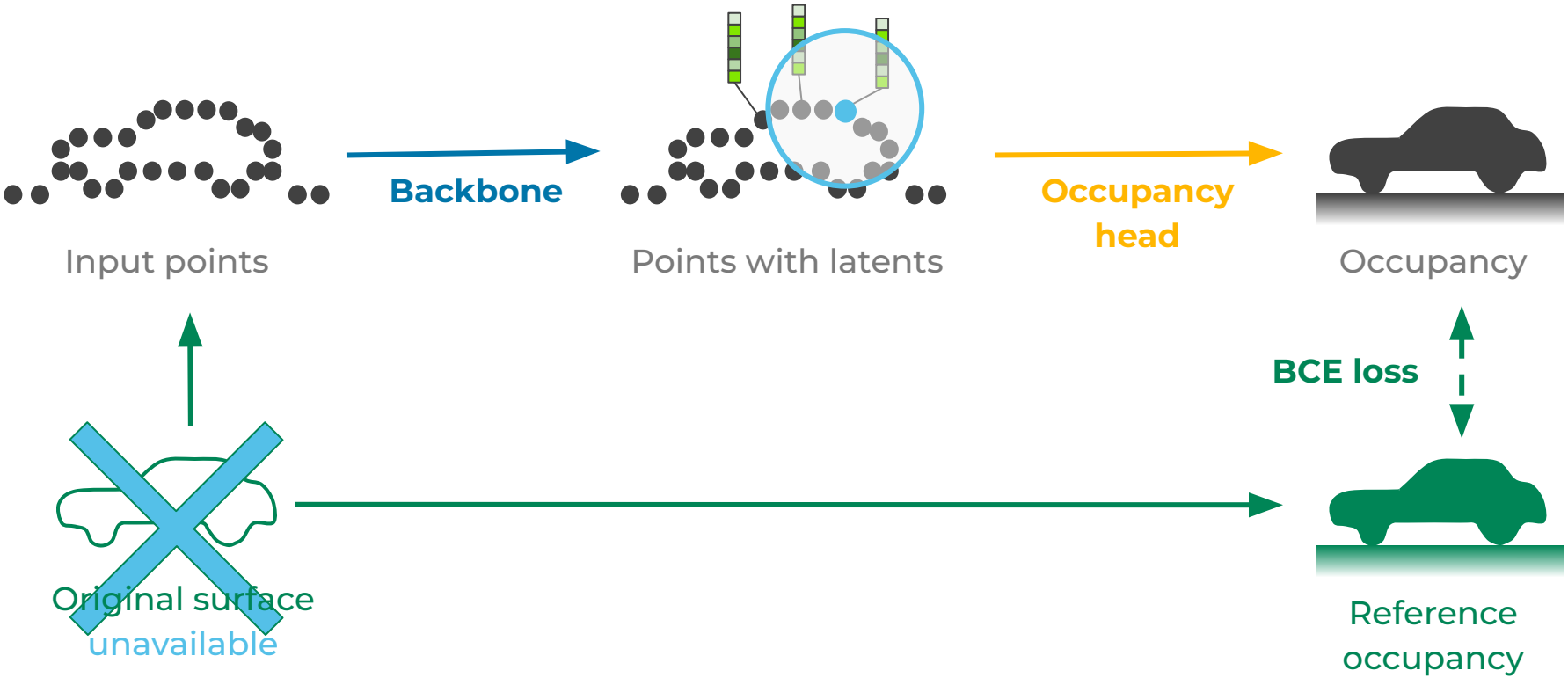
Supervision



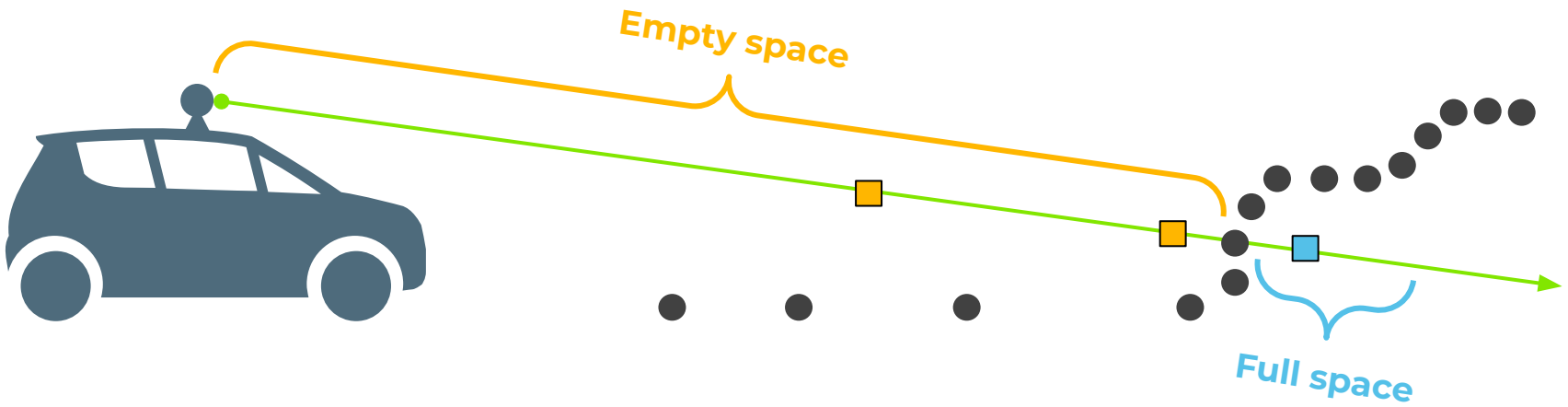
Supervision



Supervision



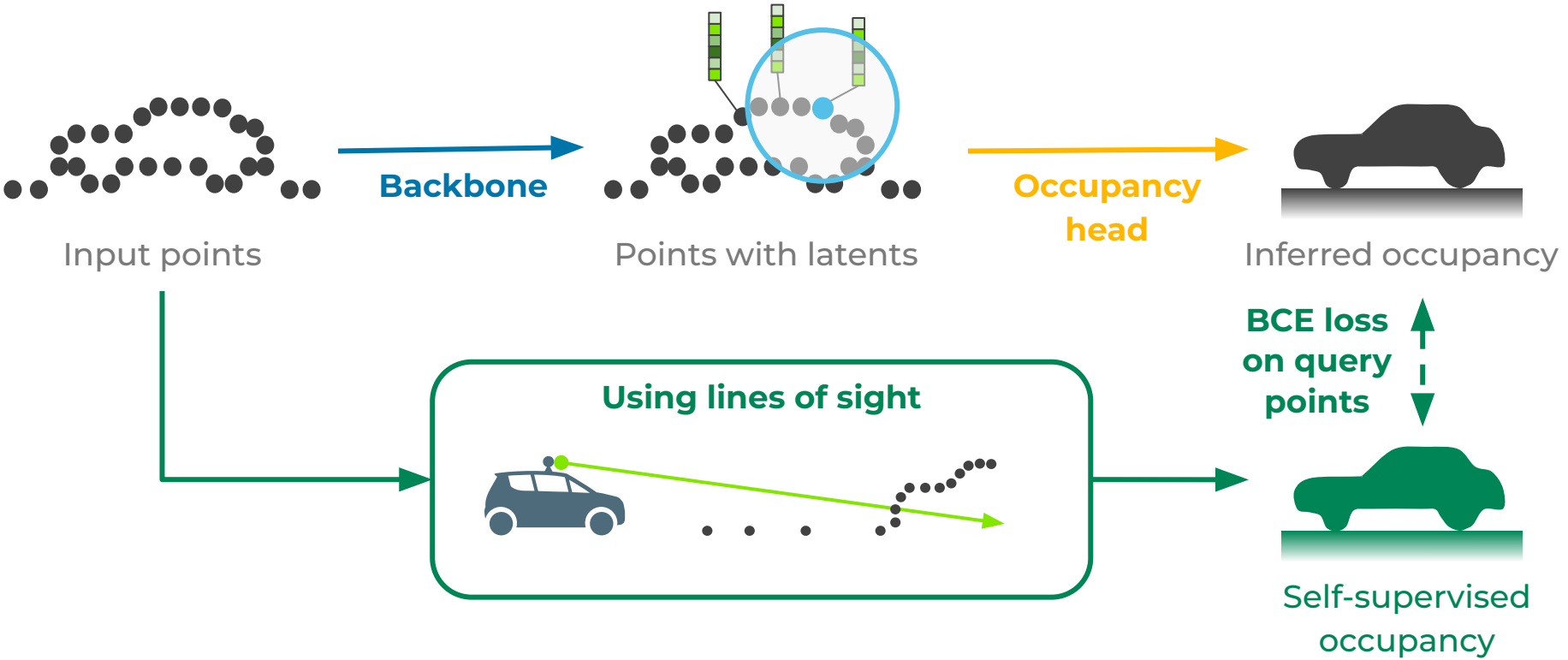
Self-supervised occupancy
Query point generation



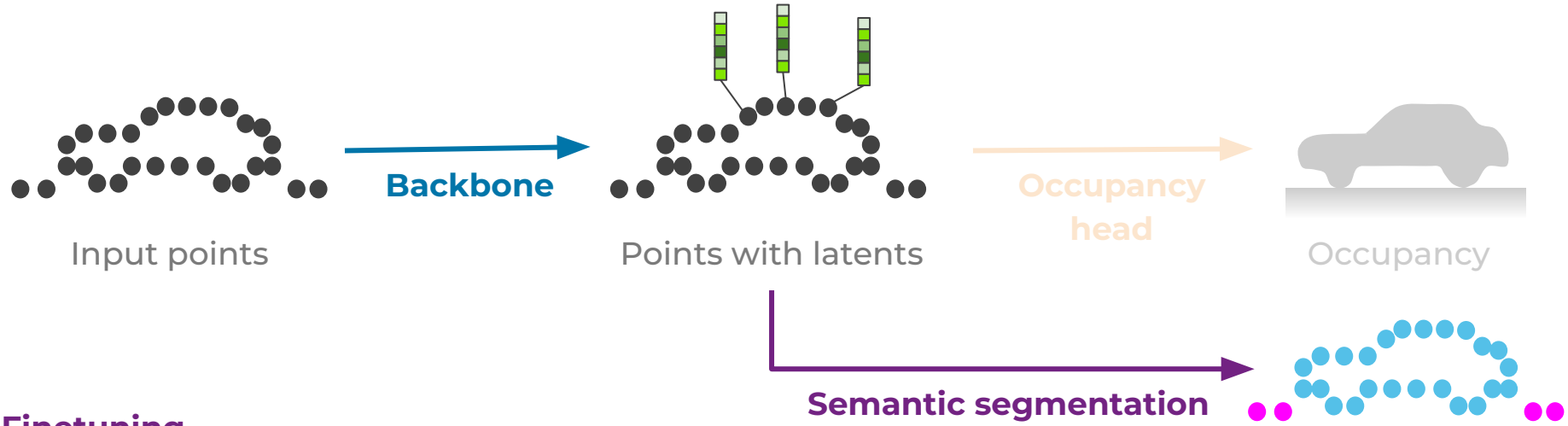
Along lidar lines of sight

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

Self-supervision

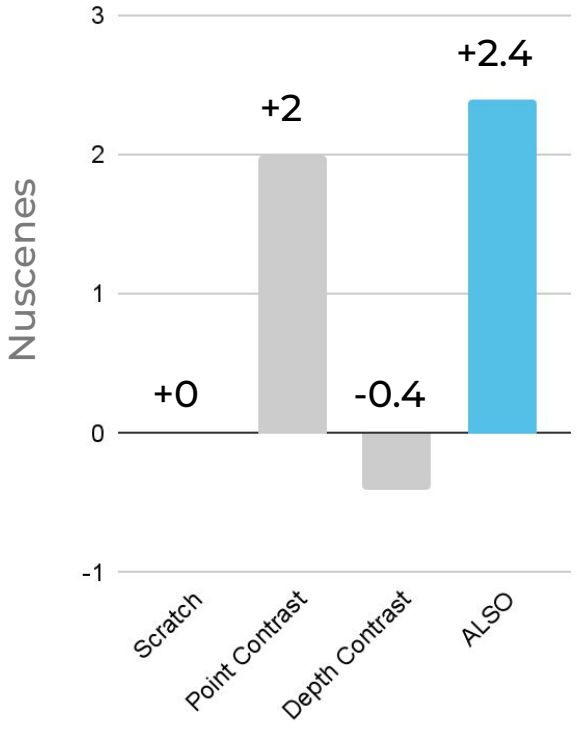


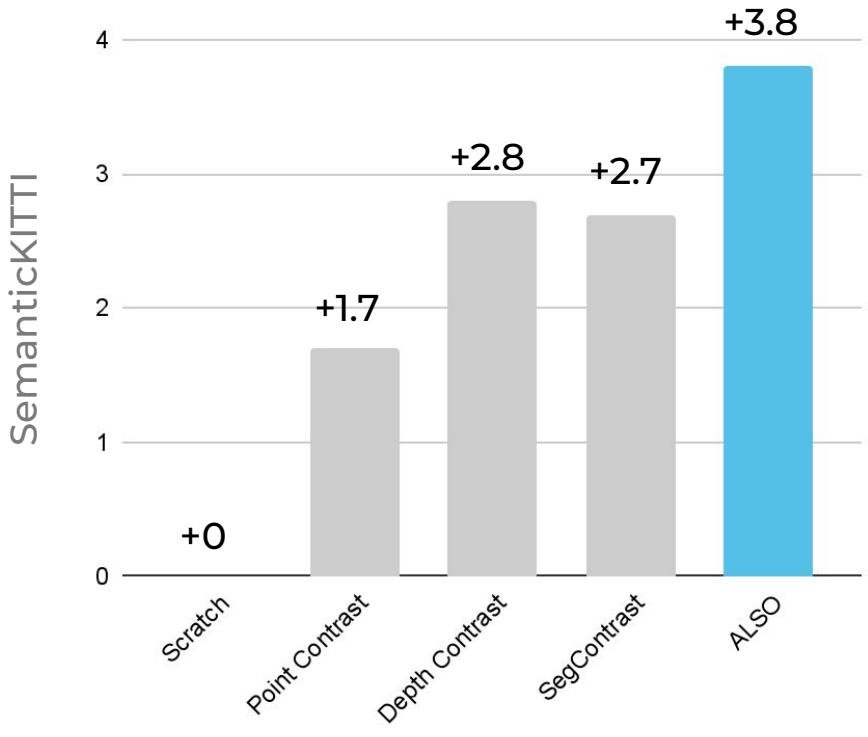
Downstream tasks
Semantic segmentation



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

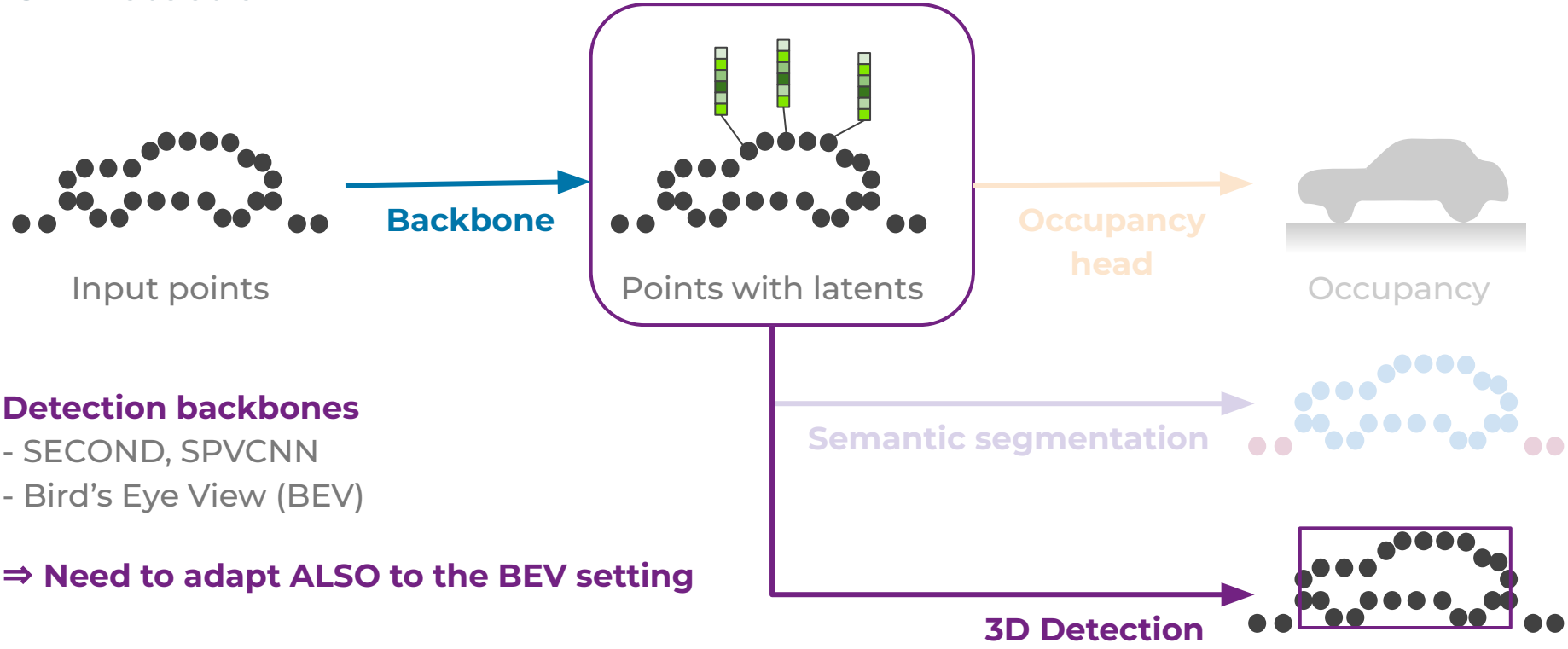
Semantic segmentation
1% annotated training data





Downstream tasks

3D Detection



Detection backbones

- SECOND, SPVCNN
- Bird's Eye View (BEV)

⇒ Need to adapt ALSO to the BEV setting

Pretraining for downstream tasks

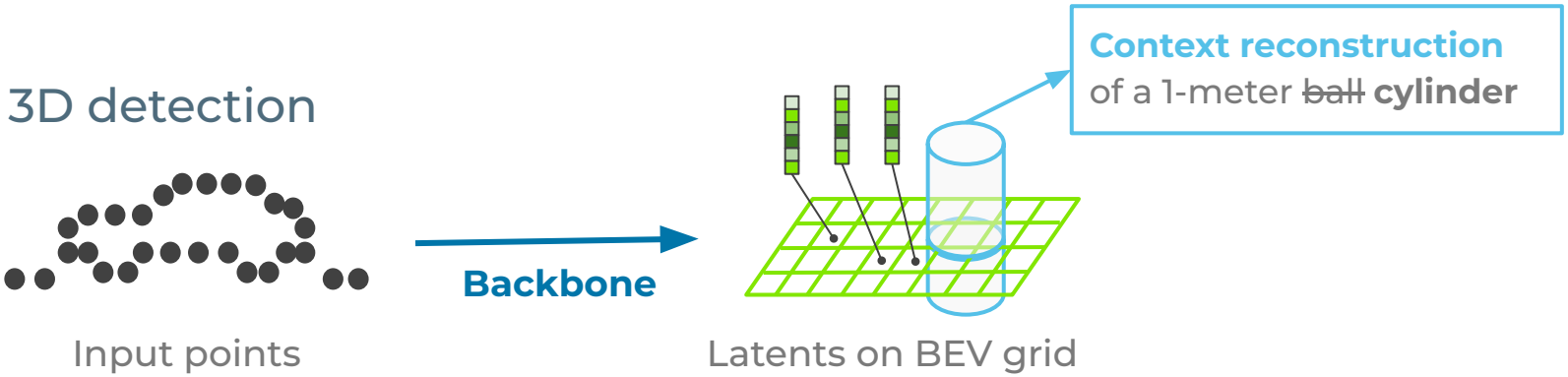
Semantic segmentation

3D detection

Pretraining for downstream tasks

Semantic segmentation

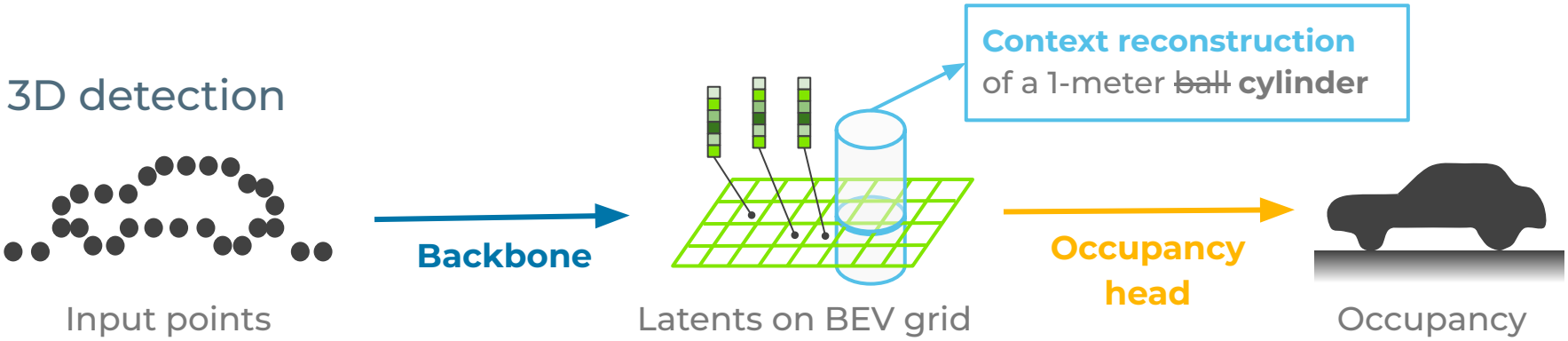
3D detection



Pretraining for downstream tasks

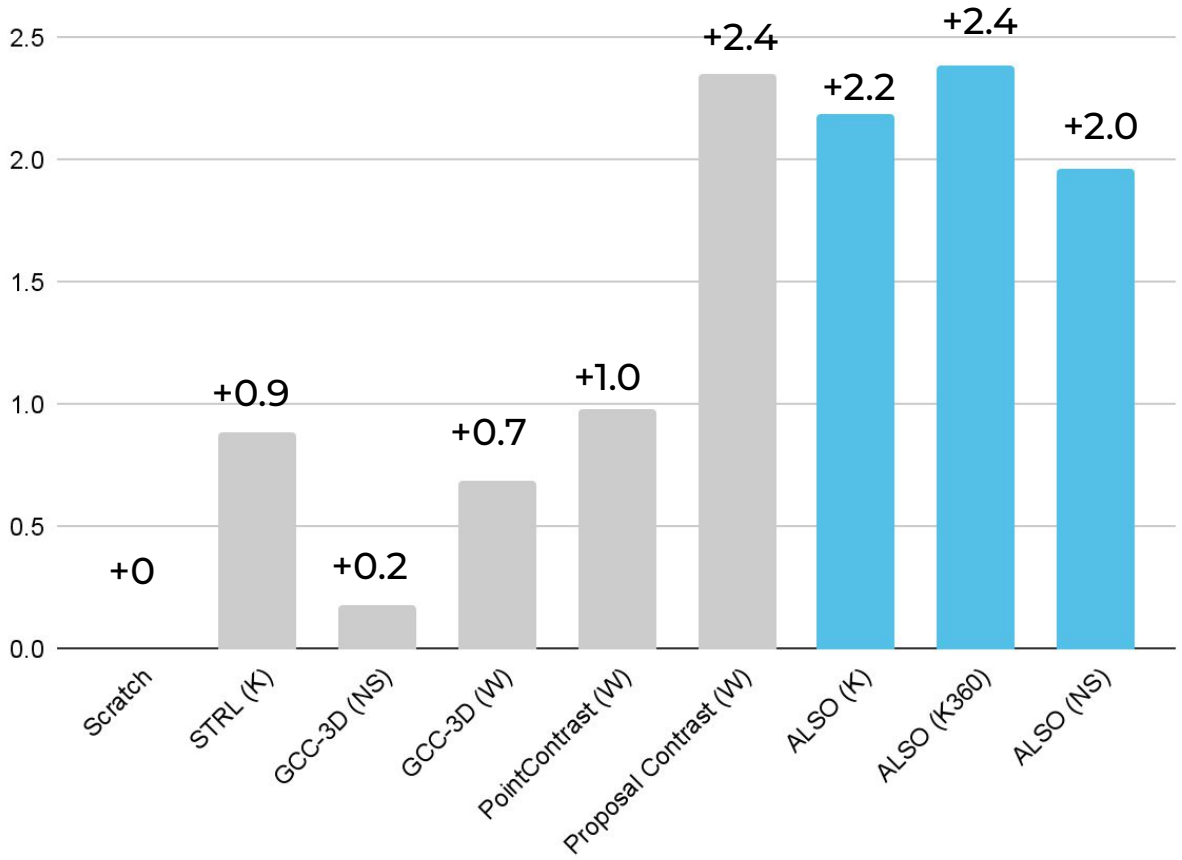
Semantic segmentation

3D detection

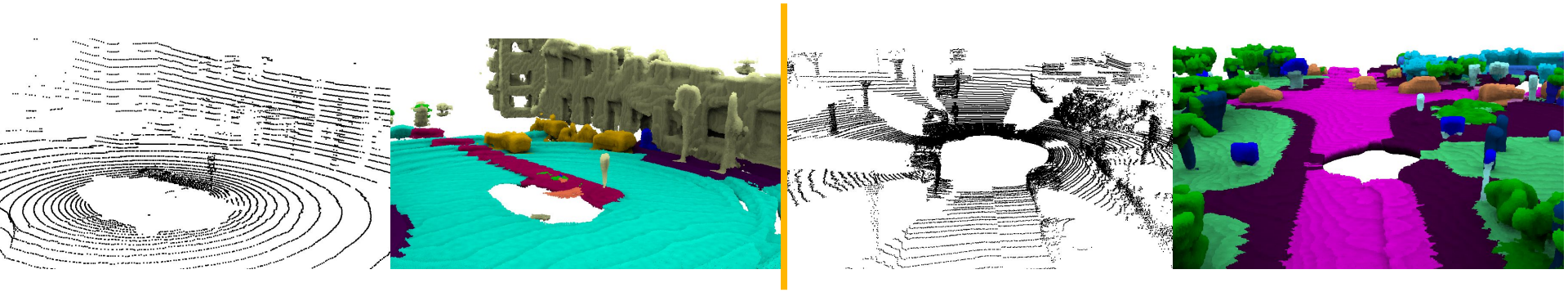


3D Detection

KITTI Benchmark



Conclusion



Semantic labels on top of estimated occupancy (nuScenes and SemanticKITTI)

Scene reconstruction as a self-supervised pretext task

✔ Competitive results

✔ Single stream
⇒ memory efficient

✘ Latent space structure not suited for direct linear probing

Personal page: www.boulch.eu

Github: https://github.com/valeoai/ALSO

Project page: https://boulch.eu/publications/2023_cvpr_also

Team publications: https://valeoai.github.io/blog/publications/