POP-3D: OPEN-VOCABULARY 3D OCCUPANCY PREDICTION FROM IMAGES

ICCV23 PARIS

Antonin Vobecky[1,2], O. Siméoni[2], D. Hurych[2], S. Gidaris[2], A. Bursuc[2], P. Pérez[2], J. Sivic[1]

[1] CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE

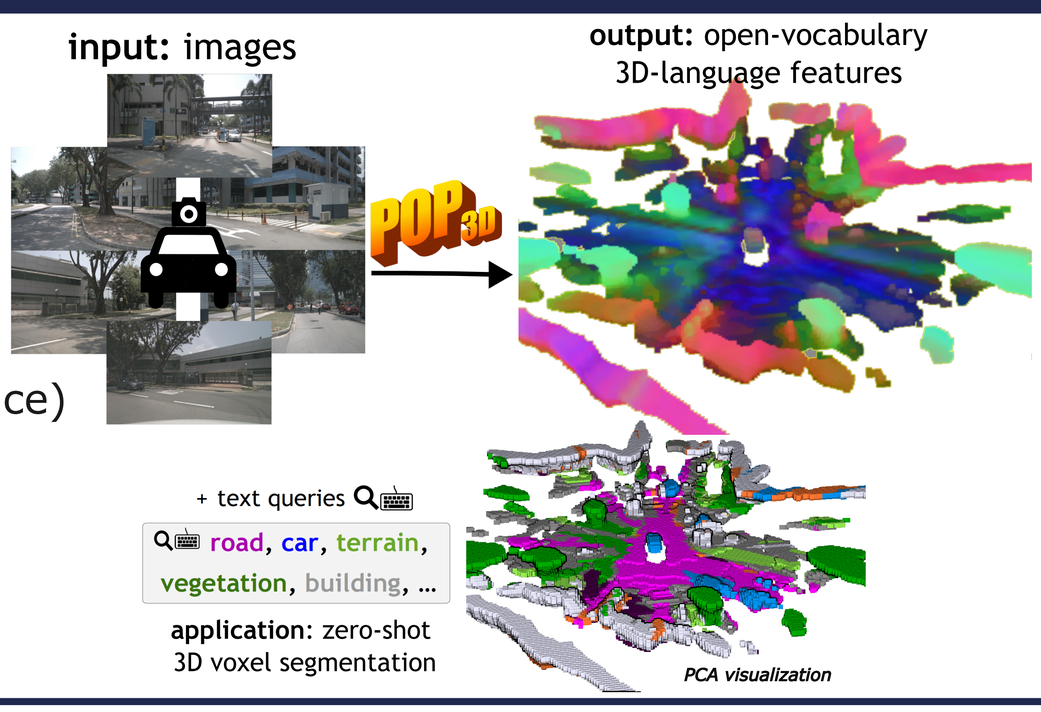
[2] valeo.ai

Motivation and Goal

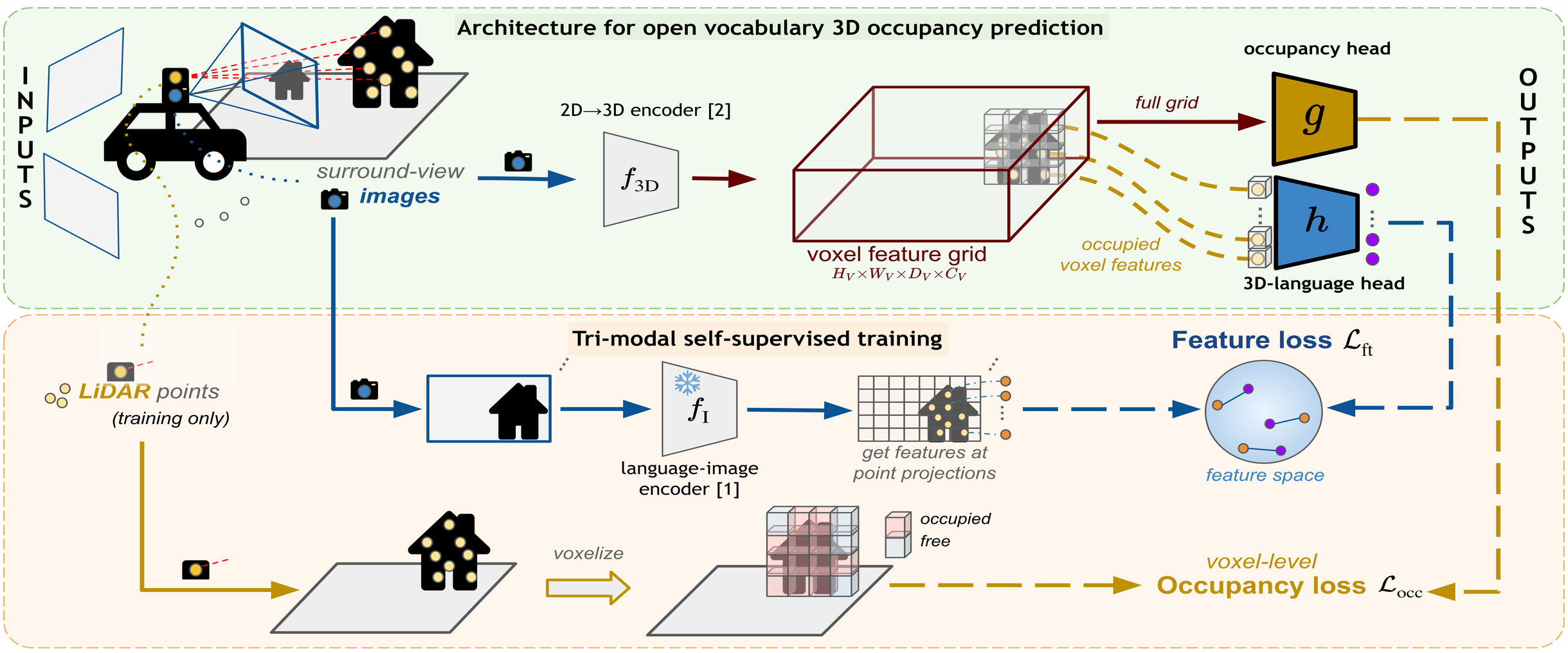
Task: **open-vocabulary 3D occupancy prediction**
- generate 3D occupancy maps from surrounding images, labeling occupied voxels with semantic classes defined using open-form text

Motivation: Tackle the following issues:
- annotations constrained by closed-vocabulary class definitions
- limited availability of sparse LiDAR annotations (expensive to produce)

Goal:
- open-vocabulary 3D semantic occupancy prediction
- *training*: unlabeled image-LiDAR data + pre-trained image-language model
- *inference*: images only



Training

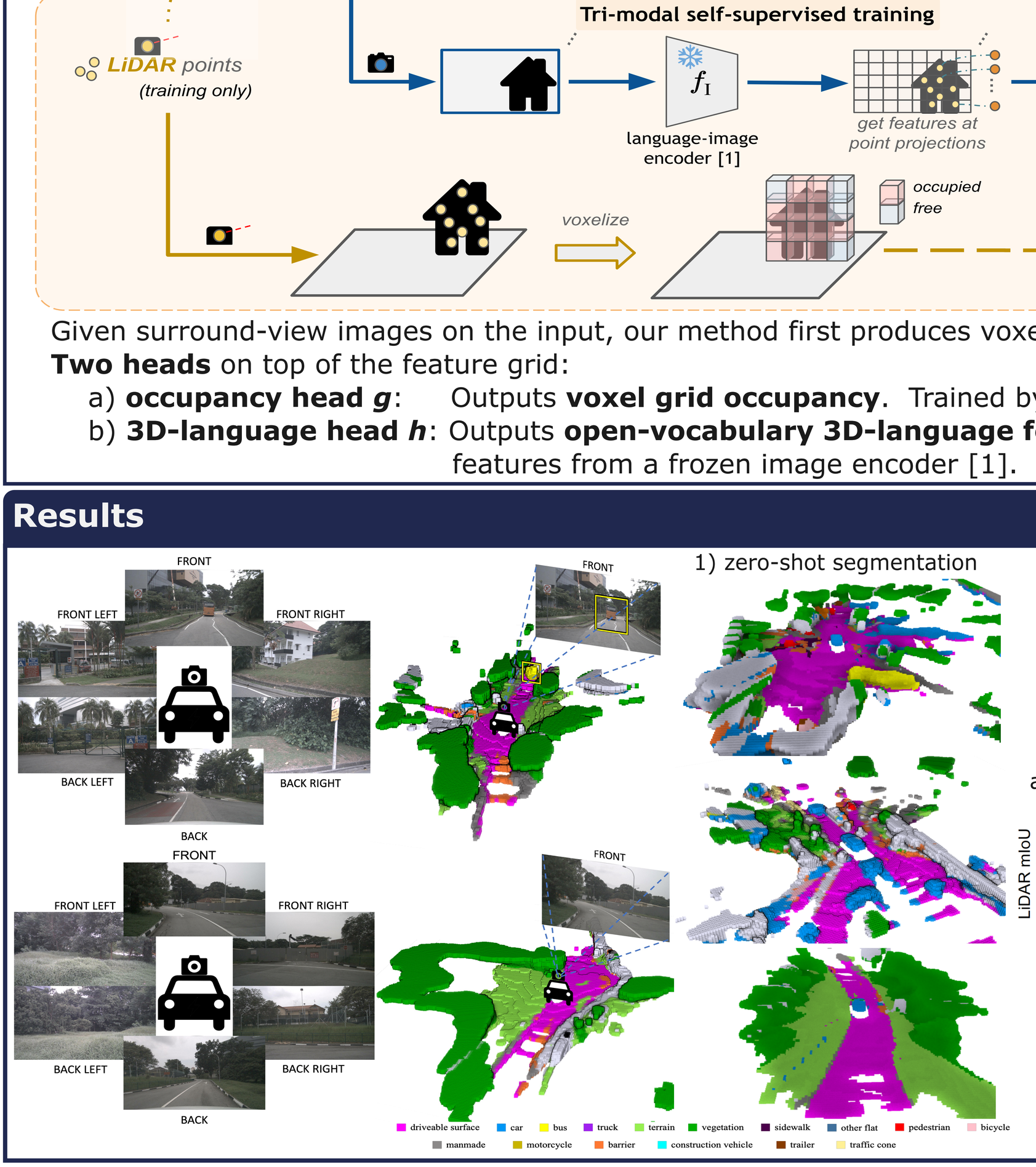


Given surround-view images on the input, our method first produces voxel feature grid using 2D-3D encoder.
Two heads on top of the feature grid:

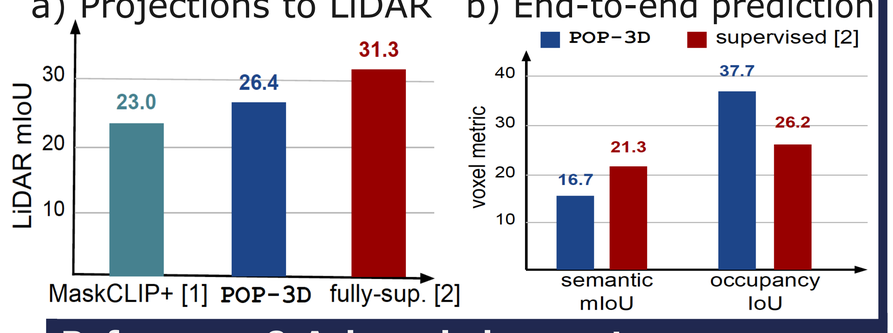
a) **occupancy head *g***: Outputs **voxel grid occupancy**. Trained by occupancy from LiDAR points.

b) **3D-language head *h***: Outputs **open-vocabulary 3D-language features**. Trained by language-image features from a frozen image encoder [1].

Results



Quantitative results



References & Acknowledgements

[1] Chong Zhou, Chen Change Loy, and Bo Dai. Extract free dense labels from clip. ECCV'22.
[2] Yuanhui Huang et al. Tri-perspective view for vision-based 3d semantic occupancy prediction. CVPR'23.

This work was partly supported by the European Regional Development Fund under the project IMPACT (reg. no. CZ.02.1.010.00.015 0030000468), and the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90140). Antonin Vobecky was partly supported by the CTU Student Grant Agency (reg. no. SGS21184OHK33T37). This research received the support of EXA4MIND, a European Union´s Horizon Europe Research and Innovation programme under grant agreement N° 101092944. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

[1] Czech Technical University in Prague, Czech Institute of Informatics, Robotics and Cybernetics