3D Semantic Occupancy Prediction from Monocular Camera - First report

Max Henking & Baptiste Poffet

6 April 2023

Deep Learning for Autonomous Vehicles, Spring Semester 2023

Definition

This project involves estimating the occupancy and semantic labels of the objects in a 3D scene using sensor data monocular camera images. The occupancy refers to whether a particular voxel or region in the scene is occupied by an object or not, while the semantic labels refer to the specific object categories, such as cars, pedestrians, buildings, and trees.

Inputs

- **Monocular camera images**

Figure 1: Example of a monocular camera image.

Outputs

3D semantic occupancy map of the scene. The 3D occupancy map will be represented as a voxel grid, where each voxel will be labelled with a semantic class. The model also predicts the occupancy probability of each voxel in the voxel grid.

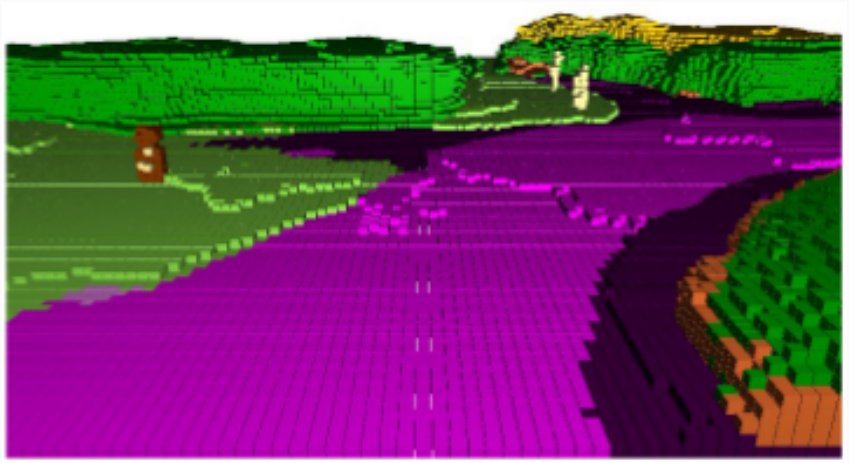

Figure 2: Example of a 3D semantic occupancy map.

Methodology

- Expliquer l'architecture du code/framework
- Ressources necessaires pour run le code.

Why this method

- Lire les articles.
- Expliquer pourquoi avoir choisi cet article/code/méthode, comparer avec d'autres méthodes trouvées dans la literature.

Literature

Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., and Gall, J. (2019).
SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences.
(arXiv:1904.01416).

Li, Y., Yu, Z., Choy, C., Xiao, C., Alvarez, J. M., Fidler, S., Feng, C., and Anandkumar, A. (2023).
VoxFormer: Sparse Voxel Transformer for Camera-based 3D Semantic Scene Completion.
(arXiv:2302.12251).

Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., and Terzopoulos, D. (2021).
Image Segmentation Using Deep Learning: A Survey.
IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 1–1.

To do