Explicit model pre-training for 3D occupancy prediction

Pontus Brandin,
Albert Henrik Mattsson

Master's thesis
2026:E32

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematics

CENTRUM SCIENTIARUM MATHEMATICARUM

Can Cars Learn to See in 3D on Their Own?

Self-driving cars need a three-dimensional understanding of their surroundings to navigate safely. Can they learn to interpret images correctly without relying on expensive annotated data or pre-existing AI knowledge?

Many modern AI systems learn from large collections of labeled data or from powerful pre-trained models that have already been trained on vast image datasets. While effective, these approaches create dependencies on external resources and may inherit limitations or biases from the data and models they rely on. Using labeled data also significantly reduces scalability, since annotations are expensive to produce. Reducing these dependencies could make it easier to develop and adapt autonomous driving systems to new environments.

In this thesis, we investigate whether a vehicle can learn a three-dimensional understanding of its surroundings directly from camera images. Rather than training a model for a single task, our goal is to pre-train a general 3D representation that can later be reused and fine-tuned for downstream tasks, such as object detection. While our model only uses camera input, we investigate the effect of using LiDAR data while training. Training the model with LiDAR makes use of the geometric information it provides, while avoiding the need for expensive LiDAR sensors in every car. This signal proved to help our model interpret the camera input correctly.

To evaluate the quality of the learned representation, we measured how well the model could predict the three-dimensional structure of a scene. Our results show that meaningful 3D representations can be learned without relying on pre-trained vision models during deployment, and that learning from input data alone remains possible even when external supervision is greatly reduced. Although these methods do not yet match the performance of approaches that make heavier use of external supervision, they still capture important information about the surrounding environment.

One of the most interesting findings was the strong connection between geometry and semantics. Better understanding of the spatial structure of a scene helped the model distinguish between different types of objects, while stronger semantic features improved its ability to reconstruct the three-dimensional world. This suggests that geometric and semantic understanding reinforce one another.

Our findings show that self-supervised learning is a promising path toward more independent and scalable 3D perception systems for autonomous driving.

Master's Theses in Mathematical Sciences 2026:E32
ISSN 1404-6342

LUTFMA-3618-2026

Mathematics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/