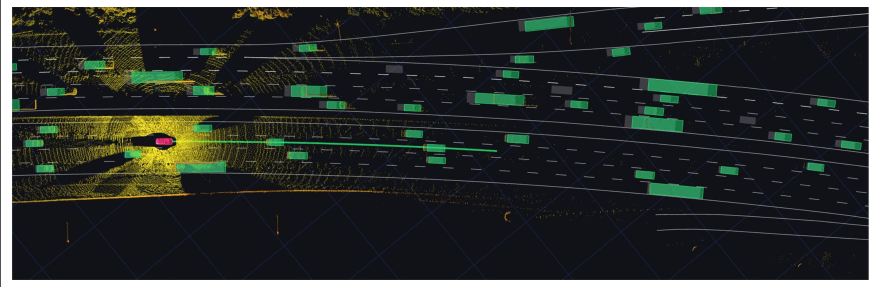
QuAD: Query-based Interpretable Neural Motion Planning for Autonomous Driving

Sourav Biswas*, Sergio Casas*, Quinlan Sykora, Ben Agro, Abbas Sadat, Raquel Urtasun

waabi · University of Toronto

Background

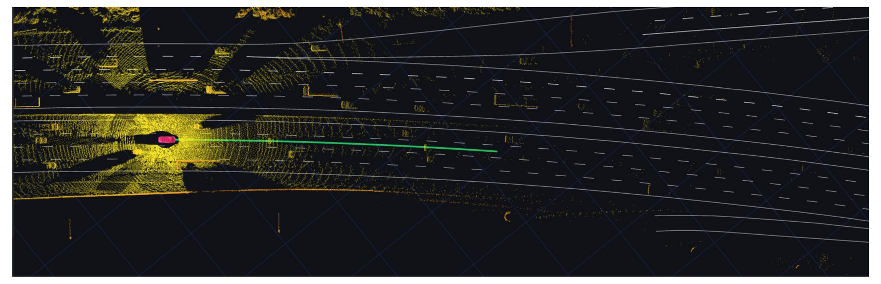
- **Object-based** perception and prediction detect objects in the scene & predict the trajectories



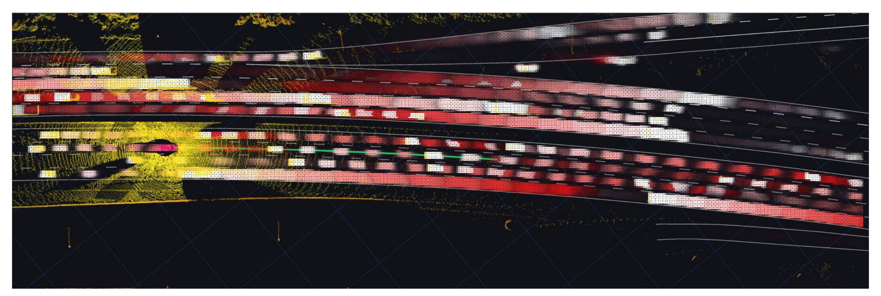
↑ Interpretable outputs
↓ Thresholding
↓ No uncertainty

- **Sensor-to-plan** methods map sensor data to plans

↑ No thresholding
↑ Uncertainty
↓ Less interpretable
↓ Distribution shift



- **Object-free** autonomy discretize space and future time into a spatio-temporal grid, and predict occupancy and flow at each cell

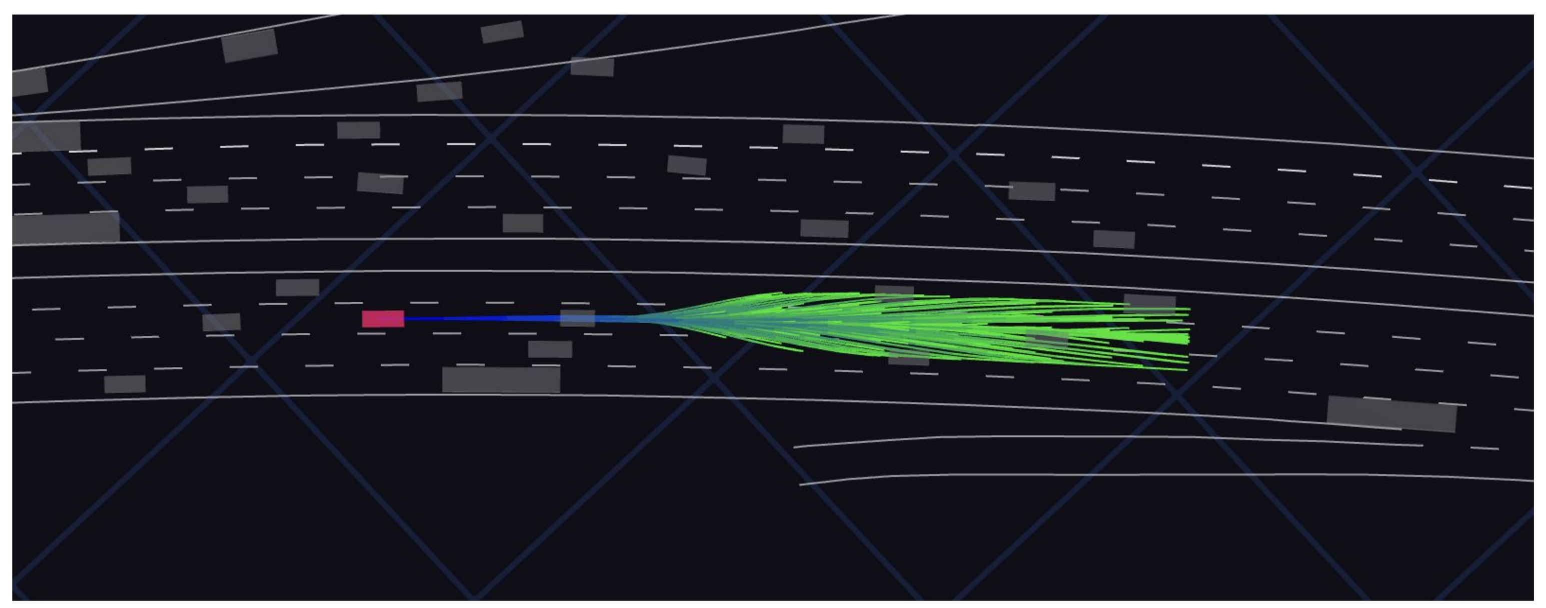


↑ No thresholding
↑ Uncertainty
↑ Interpretable
↓ Wasted computation

Our Approach

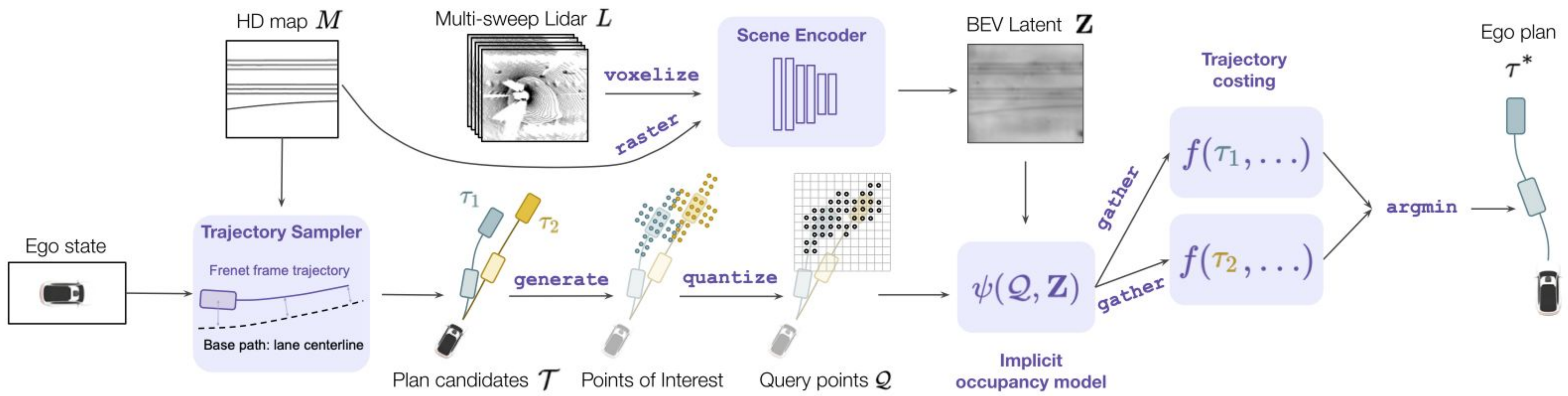
QuAD builds upon two observations:

1. the plans' reachable space is much smaller than the full spatio-temporal volume and
2. many ego states throughout the trajectories are in close proximity to each other.



Architecture

- **Intuition:** We can limit our queries for occupancy to areas relevant for motion planning which improves the efficiency of our system while maintaining high driving quality.



1. Given the ego state and the map, the trajectory sampler generates candidate plans.
2. Leveraging multi-sweep LiDAR and HD map, a scene encoder builds a BEV latent representation **Z**.
3. We generate points of interest for motion planning that cover the relevant areas around the ego vehicle future positions.
4. Since the points of interest are in close proximity to each other, we quantize them at a certain spatial resolution to create query points to send to our implicit occupancy model. *This ensures our approach runs within practical runtimes.*
5. Gather occupancy relevant to each trajectory, cost them, select the lowest cost trajectory.

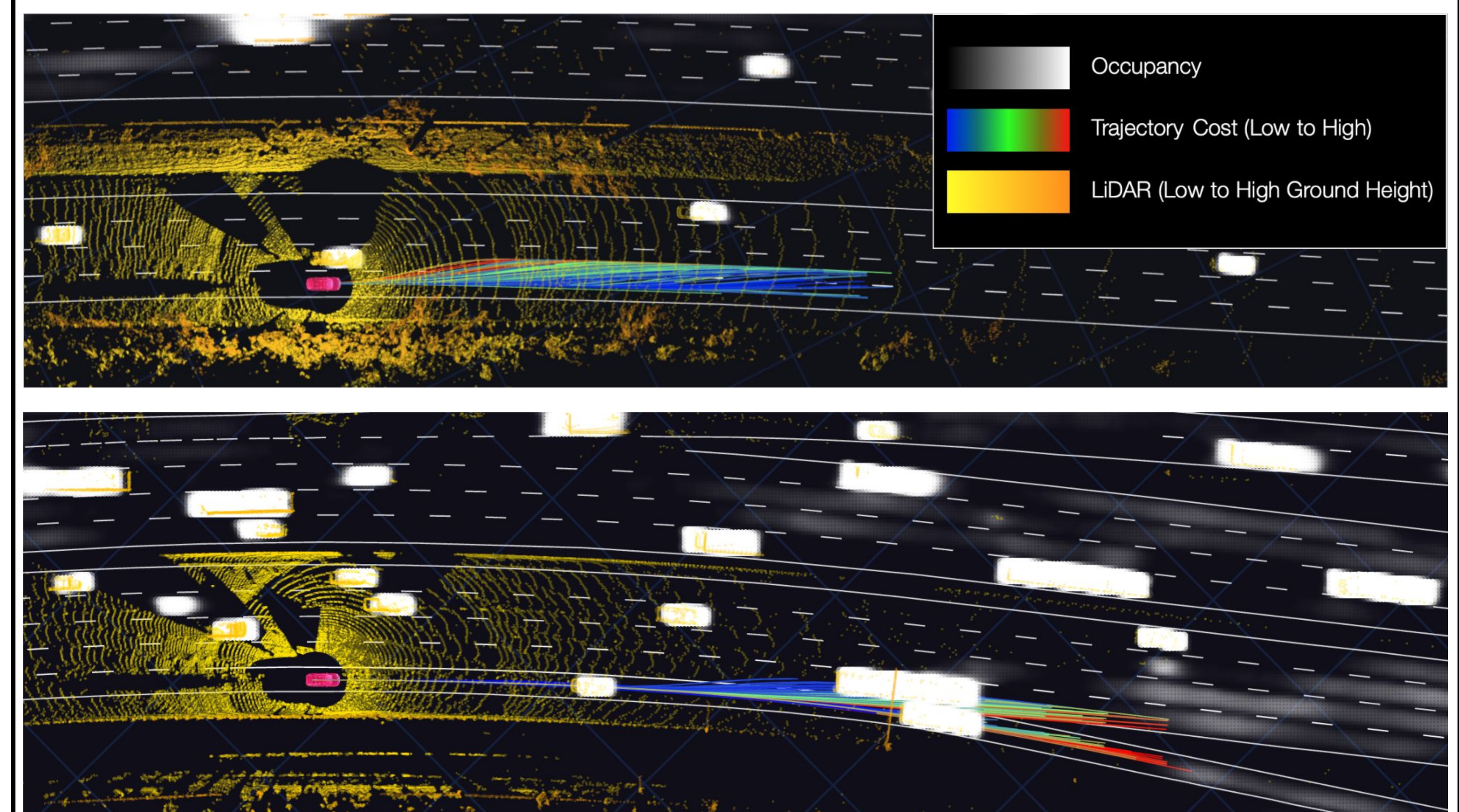
Training

We train our motion planner in two stages:

1. First we train the implicit occupancy model with binary cross entropy loss
2. Freeze the occupancy model & train cost aggregation weights with max margin to imitate an expert

$$\mathcal{L}_w = \max_{\tau} \left[\Delta J_r(\tau, \tau_e) + l_{\text{im}} + \sum_t \left[\Delta J_c^t(\tau, \tau_e) + l_c^t \right]_+ \right]_+$$

Qualitative Results



Experiments