

DeTra: A Unified Model for Object Detection and Trajectory Forecasting

Sergio Casas*, Ben Agro*, Jiageng Mao*[†], Thomas Gilles, Alexander Cui[†], Thomas Li, Raquel Urtasun

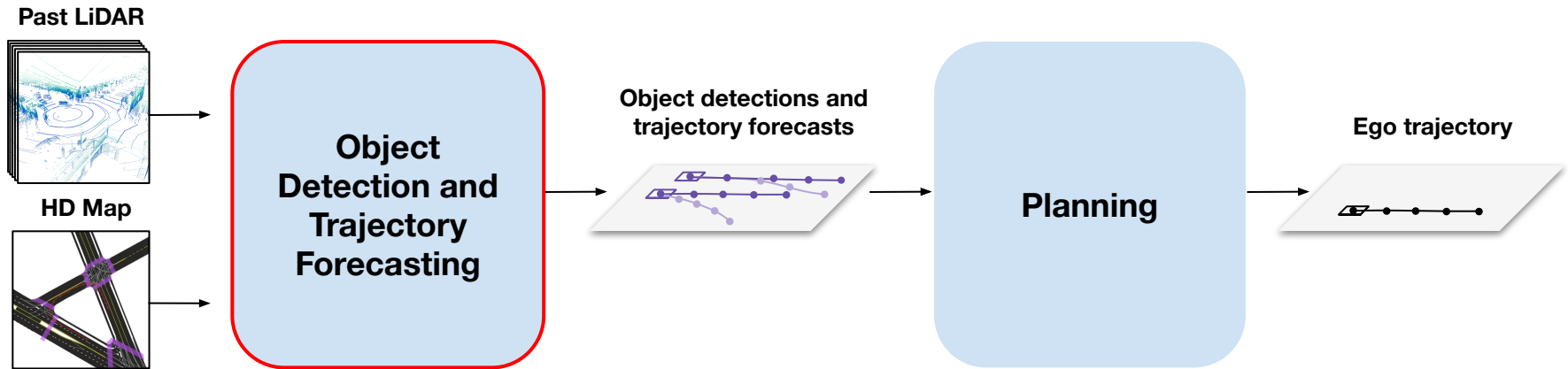


* denotes equal contribution

[†] work done while at waabi

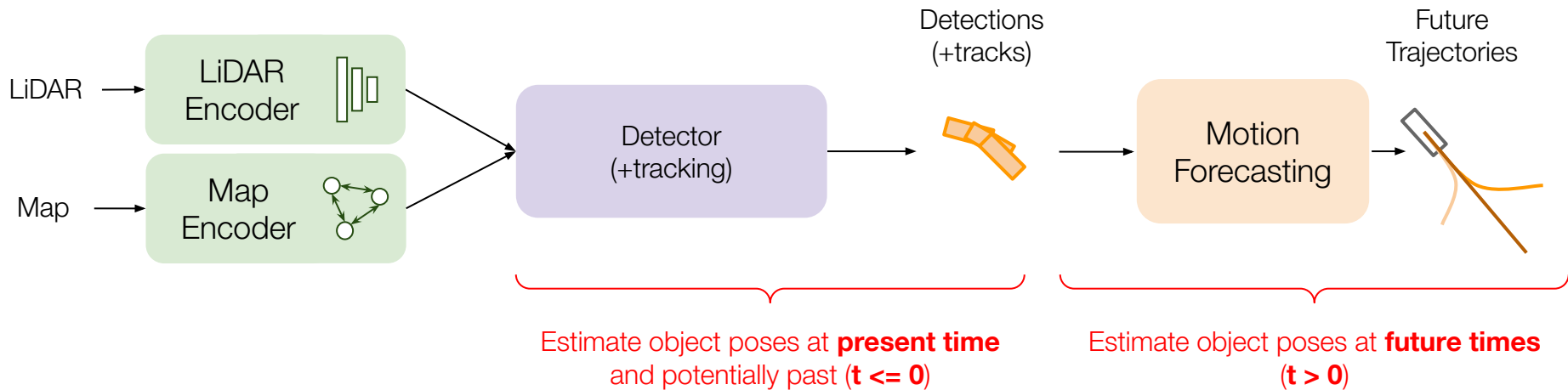
Task: Object Detection and Trajectory Forecasting

- Input: past LiDAR sweeps and high-definition maps
- Output: Object detections and trajectory forecasts (multiple hypothesis)



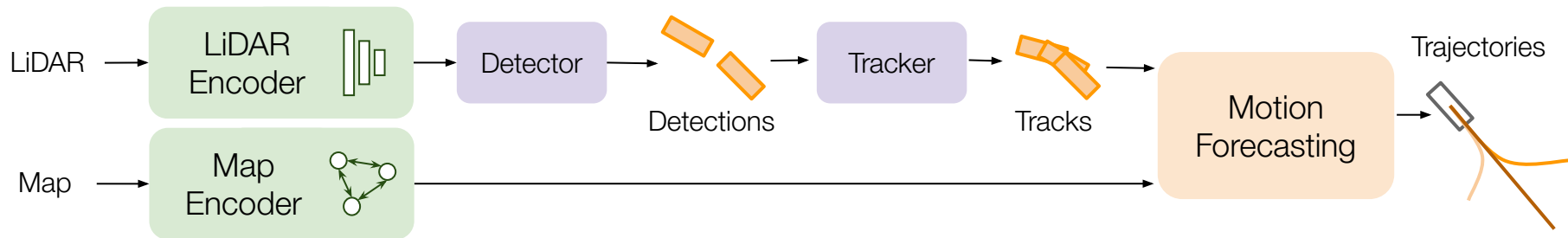
Background: Previous approaches

- Split reasoning about the present and future into separate cascading modules
- Suffer from narrow interfaces and compounding errors



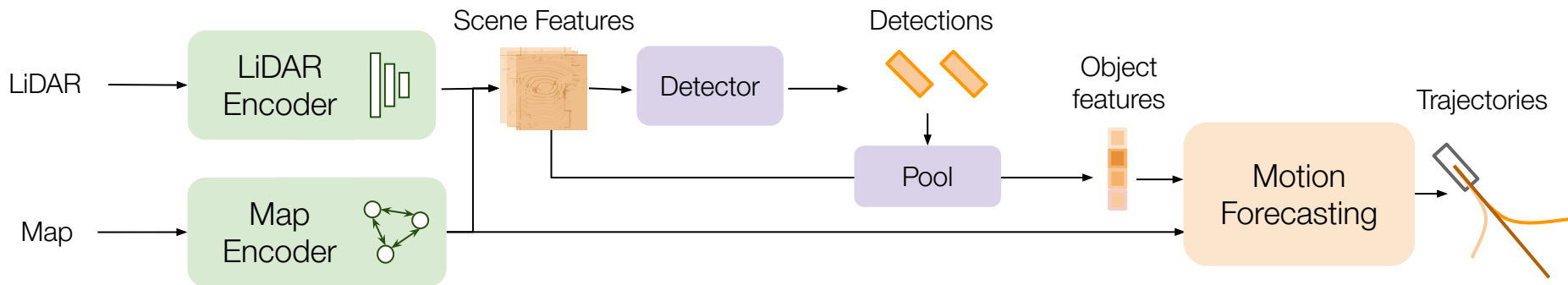
Background: Modular detection-tracking-forecasting

- A detection, tracker, and forecasting modules are cascaded
- These approaches suffer from narrow interfaces and compounding errors



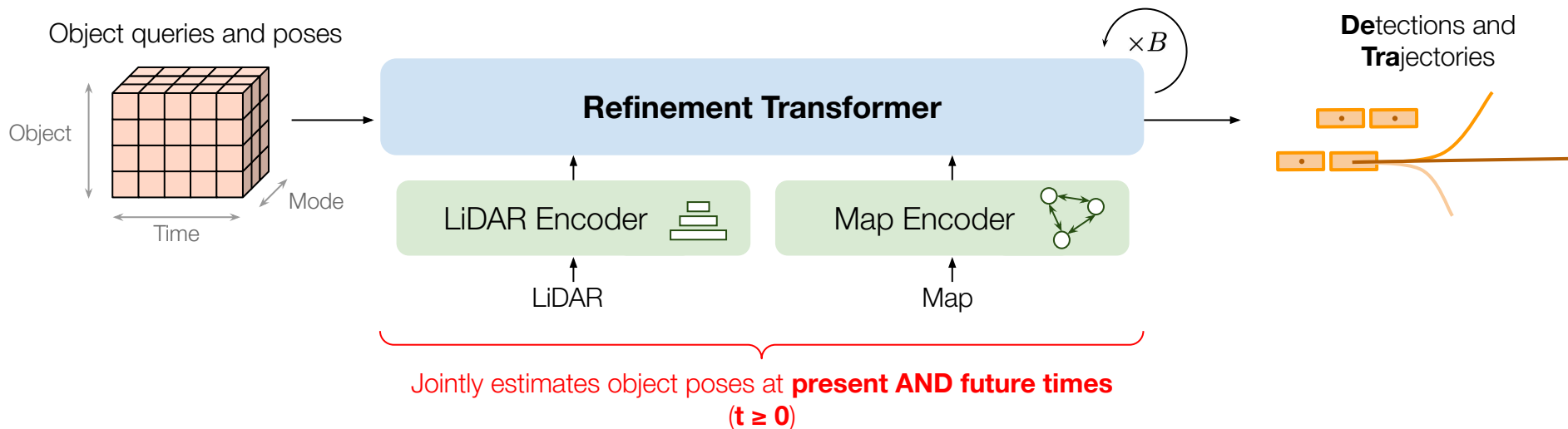
Background: Prior end-to-end detection and forecasting

- Tracker is replaced by object features from the LiDAR backbone
- These approaches have wider interfaces, but are still cascading

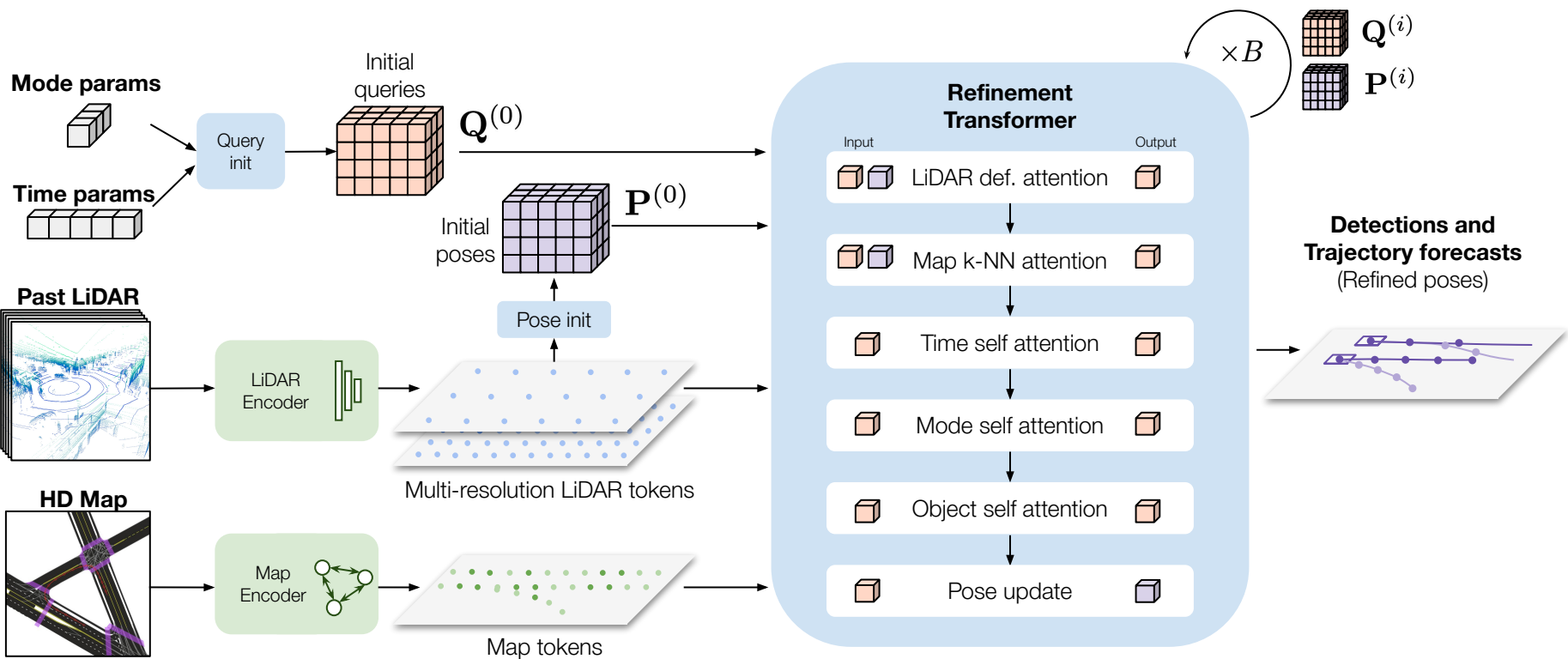


DeTra: Formulating the problem as trajectory refinement

- Object queries (features) and poses (coordinates) represent trajectories
 - $t = 0$ corresponds to the detection
 - $t > 0$ corresponds to the forecasts
- These are refined jointly over multiple refinement transformer blocks

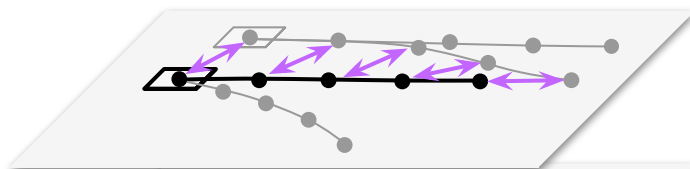


DeTra: Model architecture

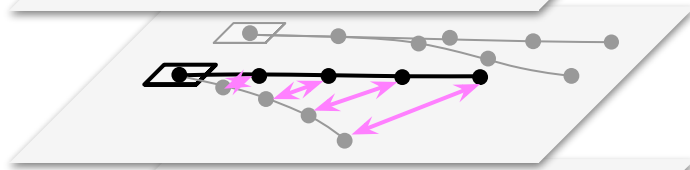


DeTra: Attention layers

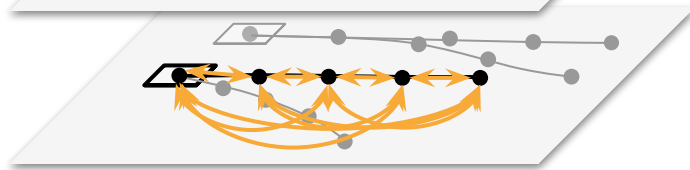
Object self
attention



Mode self
attention



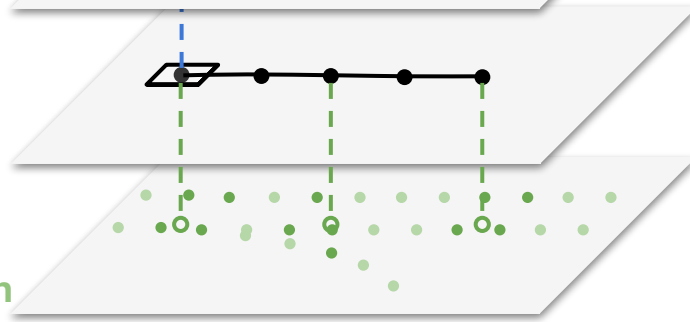
Time self
attention



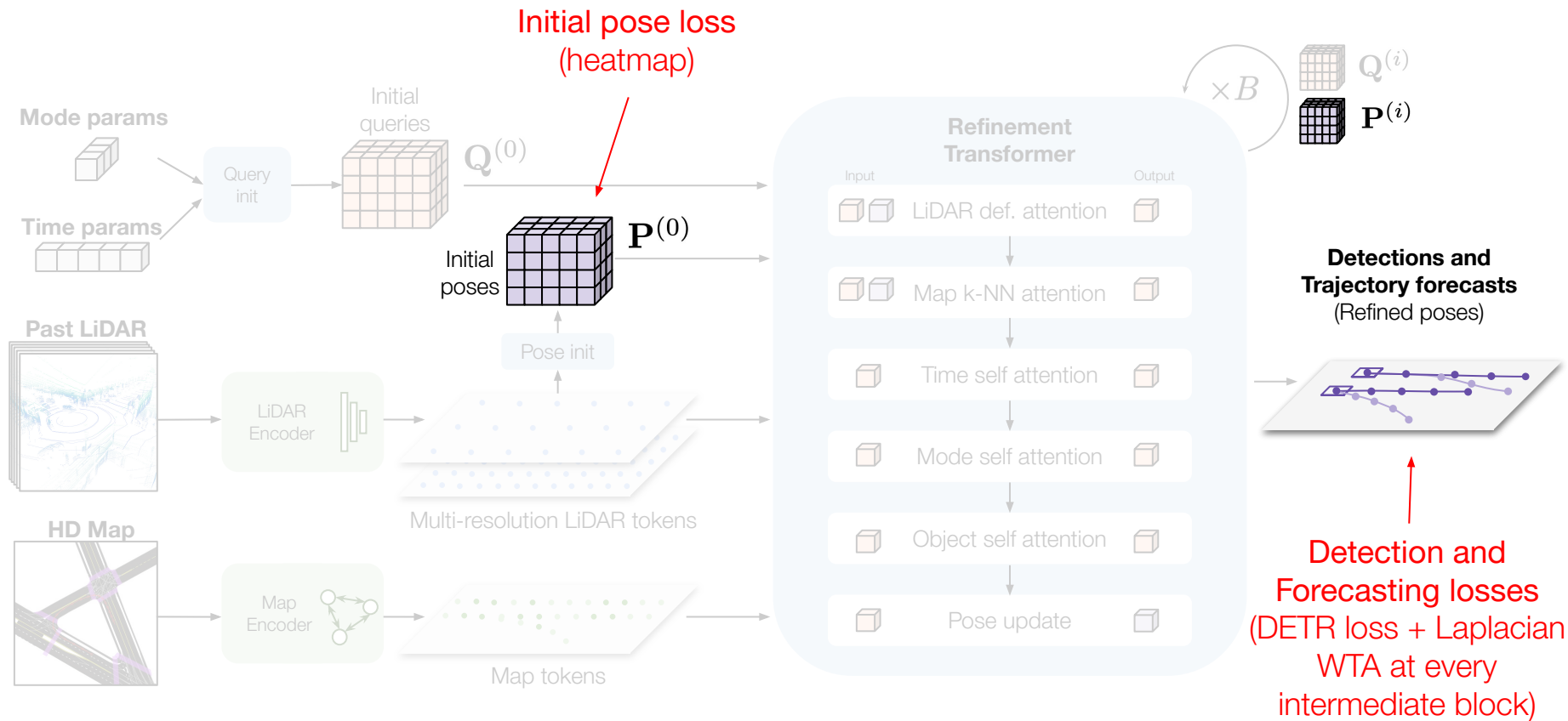
Deformable
LiDAR
attention



k-NN
Map
attention



DeTra: Training



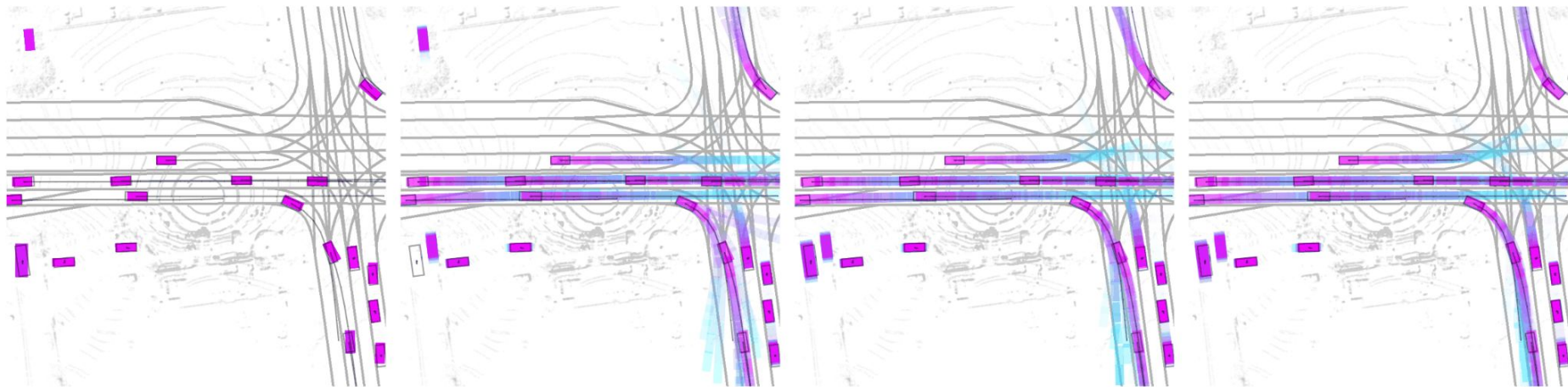
DeTra: Refinement Results

DeTra $i = 0$

DeTra $i = 1$

DeTra $i = 2$

DeTra $i = 3$



0s

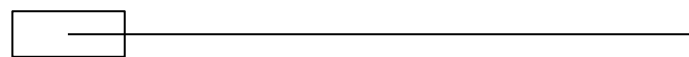
5s



Pred

0s

5s



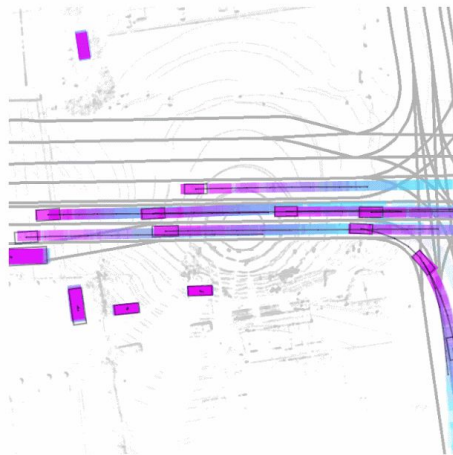
Ground Truth

DeTra: Refinement Results

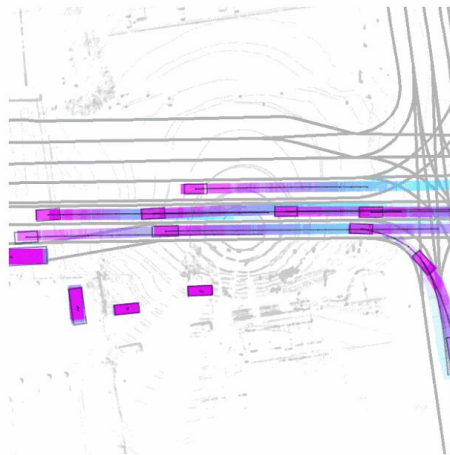
DeTra $i = 0$



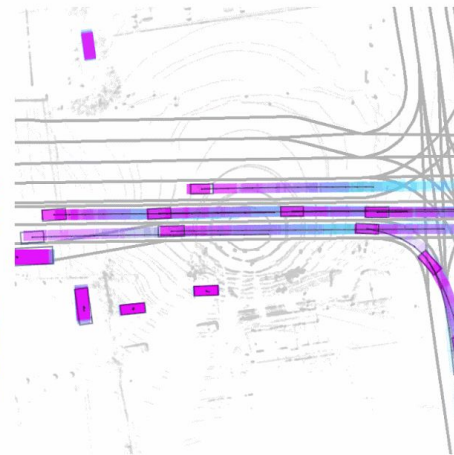
DeTra $i = 1$



DeTra $i = 2$



DeTra $i = 3$



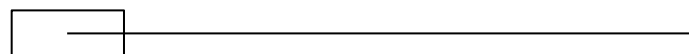
0s



5s

Pred

0s

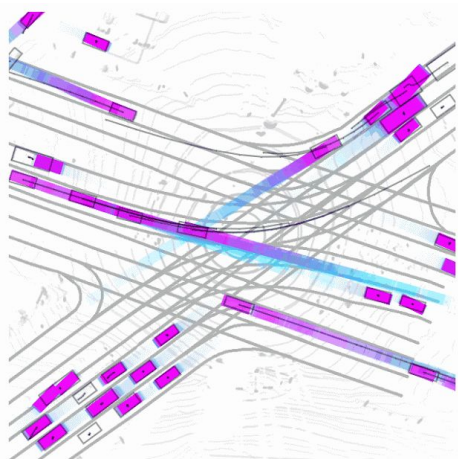


5s

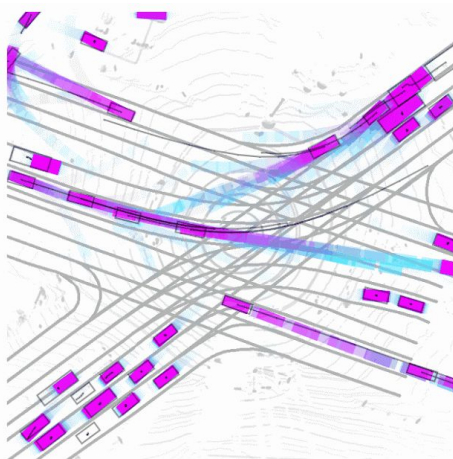
Ground Truth

DeTra: Comparison results

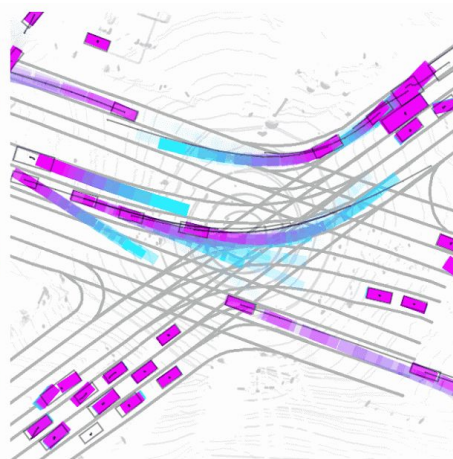
Scene Transformer



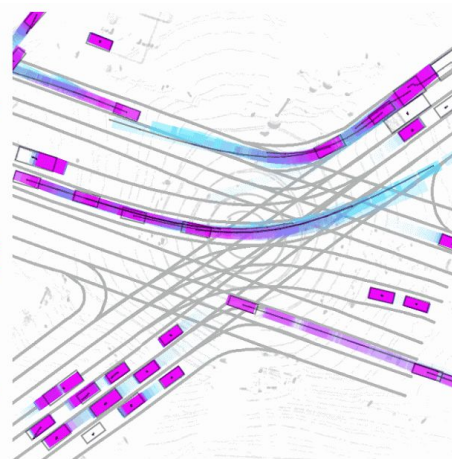
Multipath



GoRela



DeTra

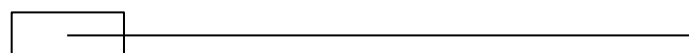


0s



Pred

0s



5s

Ground Truth

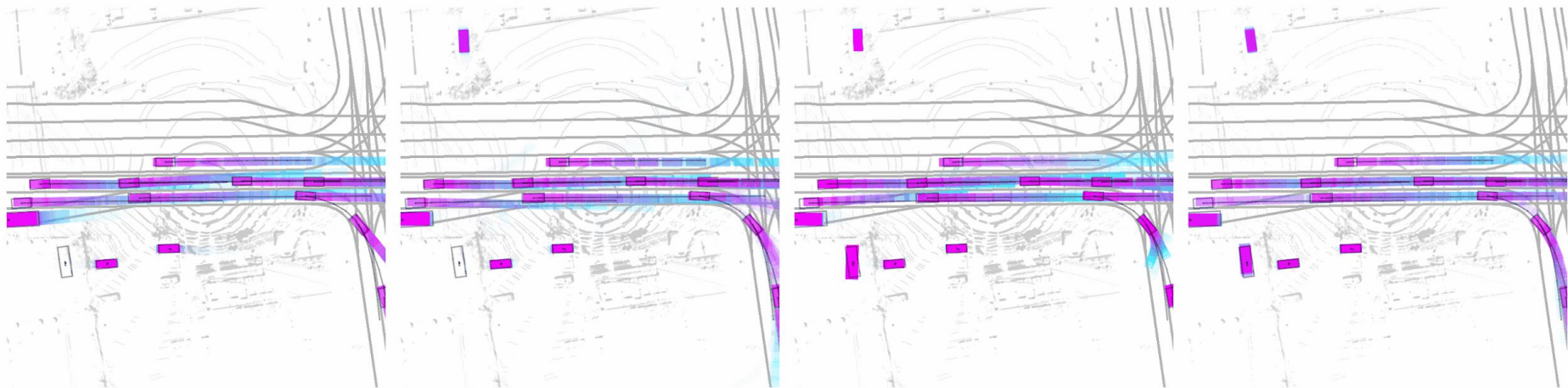
DeTra: Comparison results

Scene Transformer

Multipath

GoRela

DeTra



0s

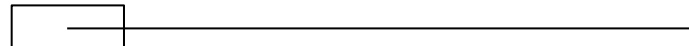
5s



Pred

0s

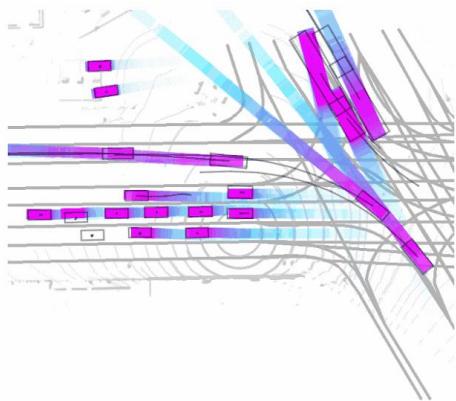
5s



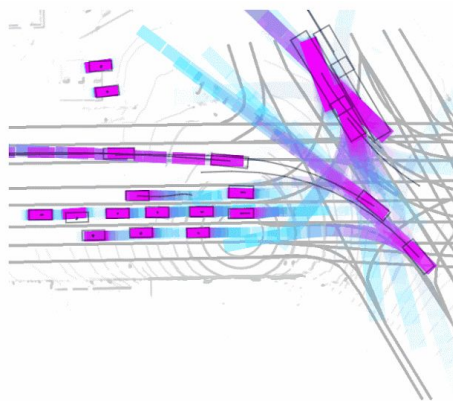
Ground Truth

DeTra: Comparison results

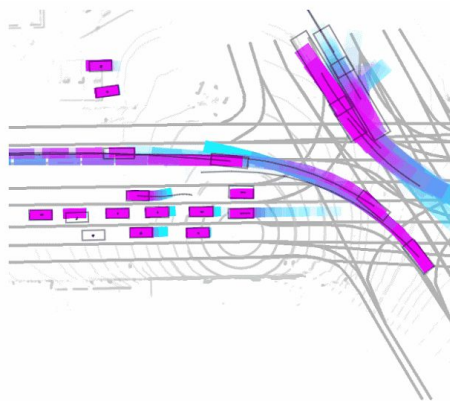
Scene Transformer



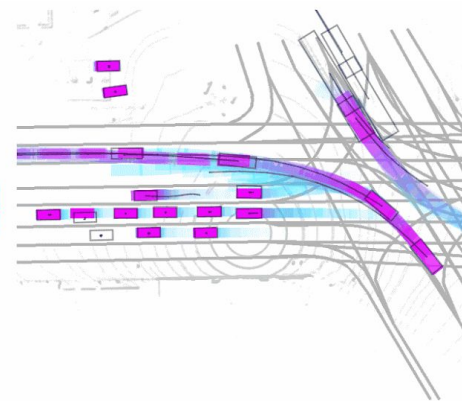
Multipath



GoRela



DeTra

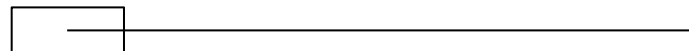


0s



Pred

0s



5s

Ground Truth

Conclusion

We introduce DeTra, a unified model for object detection and trajectory forecasting

- Formulates detection and forecasting as a single trajectory refinement problem
- Flexible architecture that can handle heterogeneous inputs
- Performs strongly in Argoverse 2 Sensor and Waymo Open Dataset
- Design choices are validated through extensive ablations
 - Refinement is key
 - Leveraging geometric priors in cross-attention is important
 - Every proposed component has a positive contribution