端到端自动驾驶

算法解读

关于自动驾驶内容

I. 端到端自动驾驶技术路线是什么方式?

- 从感知、规控的独立模型演进到端到端?
- 直接从零开始重新设计端到端?

- 采访多个路人甲，对自动驾驶的看法（一本正经胡说八道）

01. 传统级联与端到端算法

传统自动驾驶系统

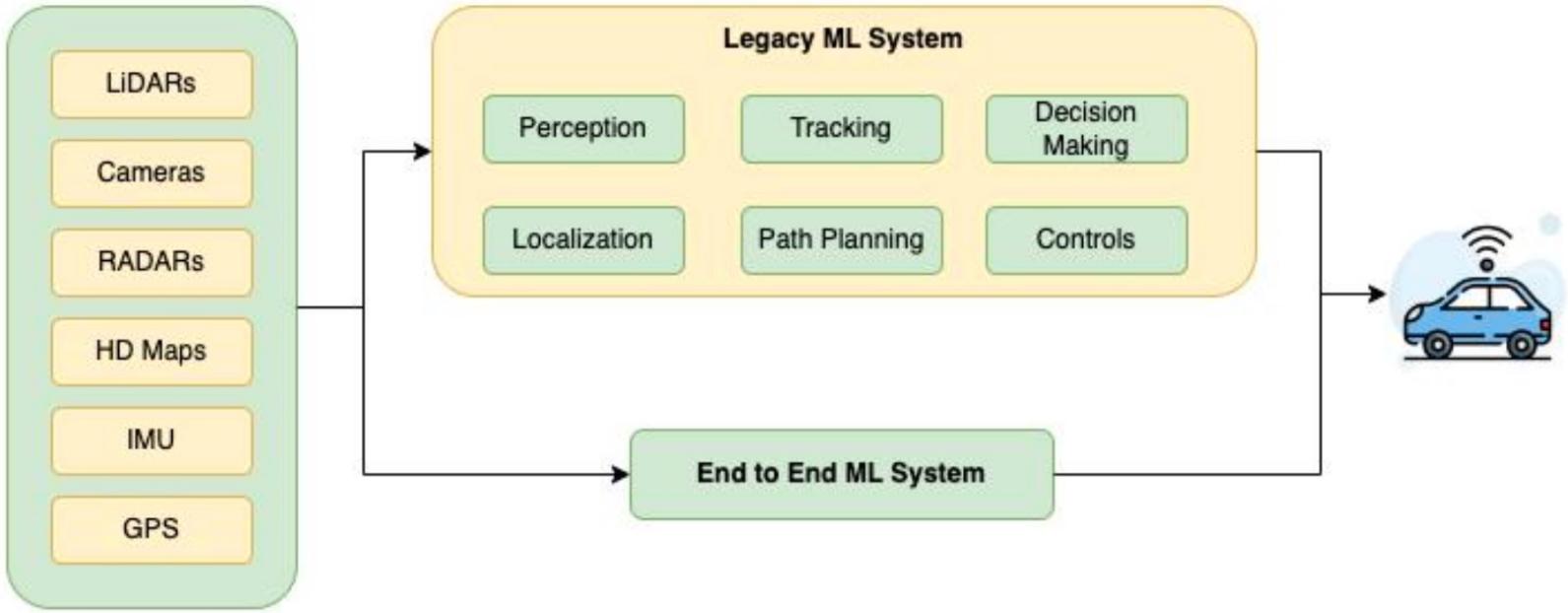
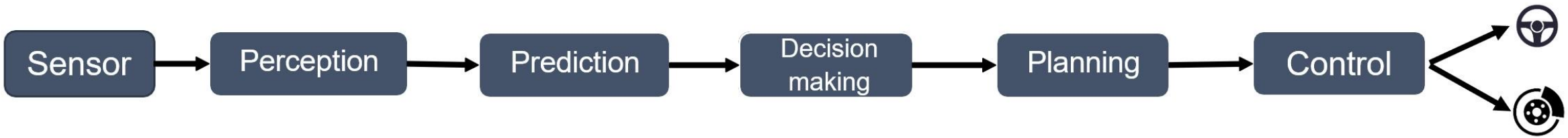


Fig. 2. Comparison of Modular Autonomous driving tasks vs single-model of end-to-end autonomous driving task.

1. 高精地图（Mapping）
2. 感知（Perception）
3. 定位（Localization）
4. 目标跟踪（Tracking）
5. 行为规划（Predicting）
6. 路径规划（Path Planning）
7. 决策（Decision Making）
8. 控制（Controls）
9. 仿真和测试（Simulation）

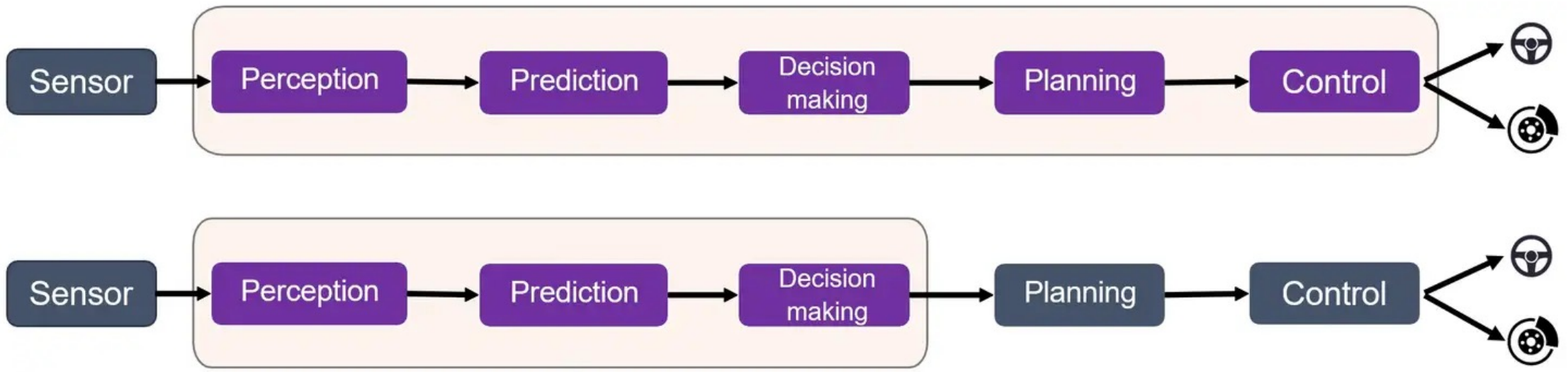
传统自动驾驶系统：级联式

- 每个模型都要专门进行训练、优化、迭代，随着模型不断进化，参数量提高和数据量增加，所需研发投入大，研发成本高。
- 模块化架构可以看做级联流水线，模型的输入参数，是前级模型的输出结果。如果前级模型输出结果有误差，会影响下一级模型输出，导致累计级联误差，最终影响系统性能。



端到端方式

- 通过一个模型实现流程中多个模型的功能。该模型可以接收传感器数据（图像、激光雷达等）作为输入，并输出车辆控制指令（如方向盘角度、刹车和加速等）。通过大规模数据集和训练算法，模型能够学习从感知到控制的完整驾驶策略。



级联方案与端到端方案对比

	级联	端到端
算法类型	模型算法 + 规则判断	模型算法 + 数据驱动
安全性	高	未知
可解释性	高	低
响应时延	中	高
算法难度	中	高
训练难度	中	高
评测手段	相同	相同
累计误差	高	无

02. E2E 算法解读

2.1 测评方法

研究方式

- 端到端自动驾驶学术研究主要分为两类：
  1. **闭环方式：**模拟器（如CARLA ）中进行验证，规划下一步指令可以被真实执行。
  2. **开环方式：**在已经采集现实数据上进行端到端研究（主要指模仿学习，IL）。

- **开环 vs 闭环：**闭环是算法控车，开环是算法不控车。用输出规划轨迹和人实际开出轨迹的差距来评分，数学本质是预测误差 Predict Error。

研究方式

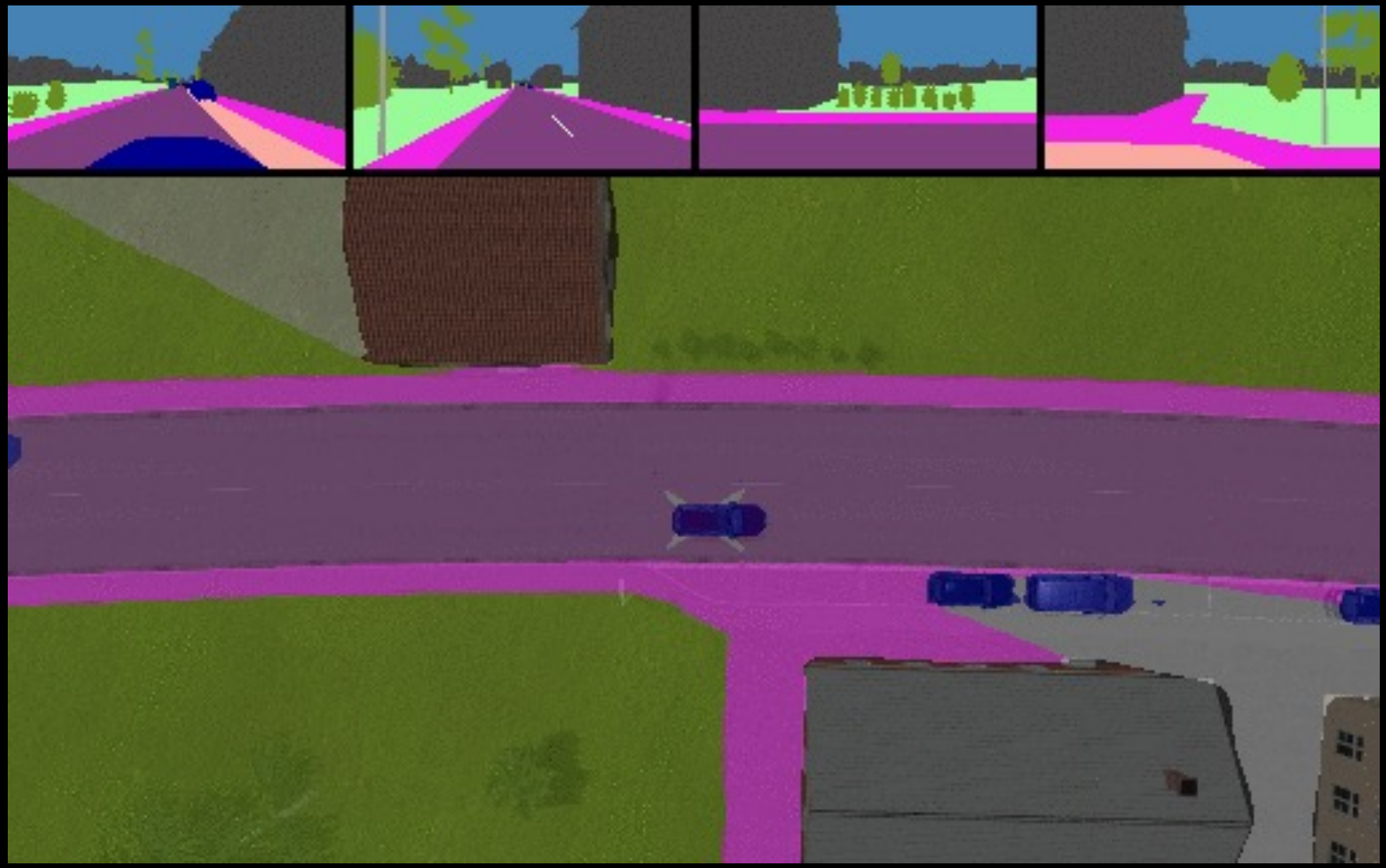
- 开环方式不能真正看到自动驾驶预测指令执行后效果。由于不能得到反馈，开环测评受限，文献中常用指标是：

1. **L2 距离**：计算预测轨迹 vs 真实轨迹之间 L2 距离，来判断预测轨迹的精度；
2. **Collision Rate**：通过计算预测轨迹和其他物体发生碰撞概率，来评价预测轨迹的安全性。

2.2 BEV Net

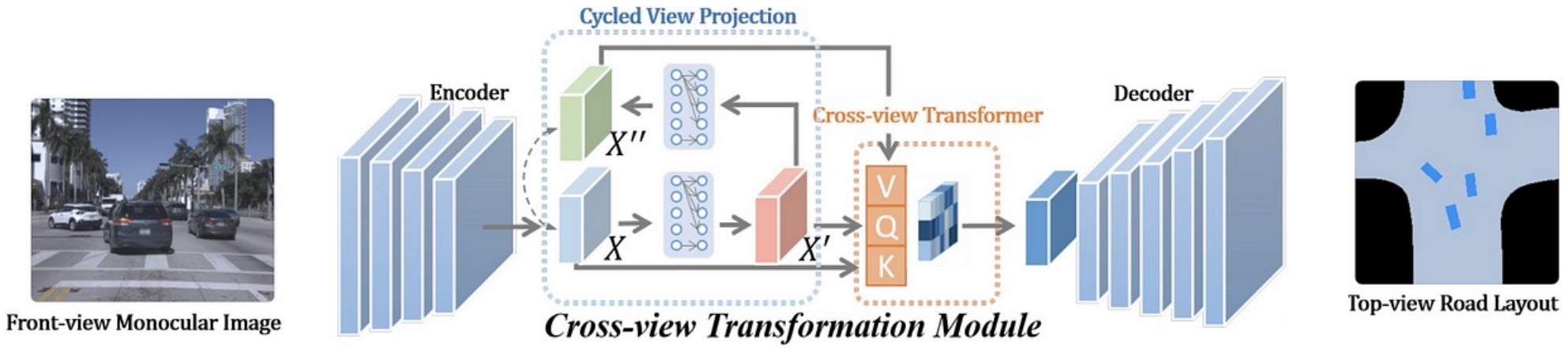
Birds-eye-view鸟瞰图

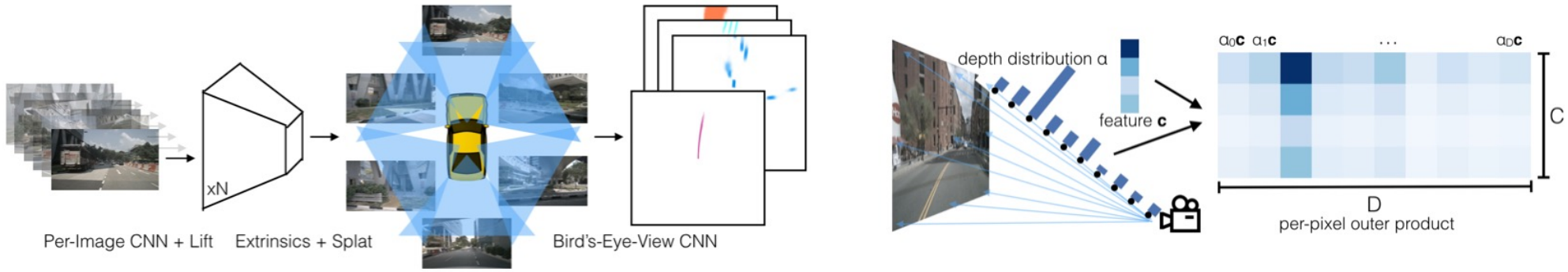
Birds-eye-view鸟瞰图



Birds-eye-view鸟瞰图

- Monocular BEV Perception with Transformers in Autonomous Driving



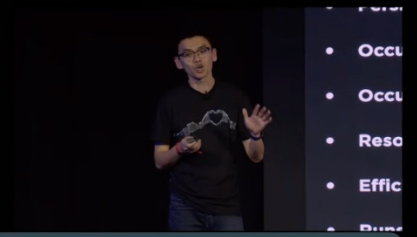


2.3 栅格网络

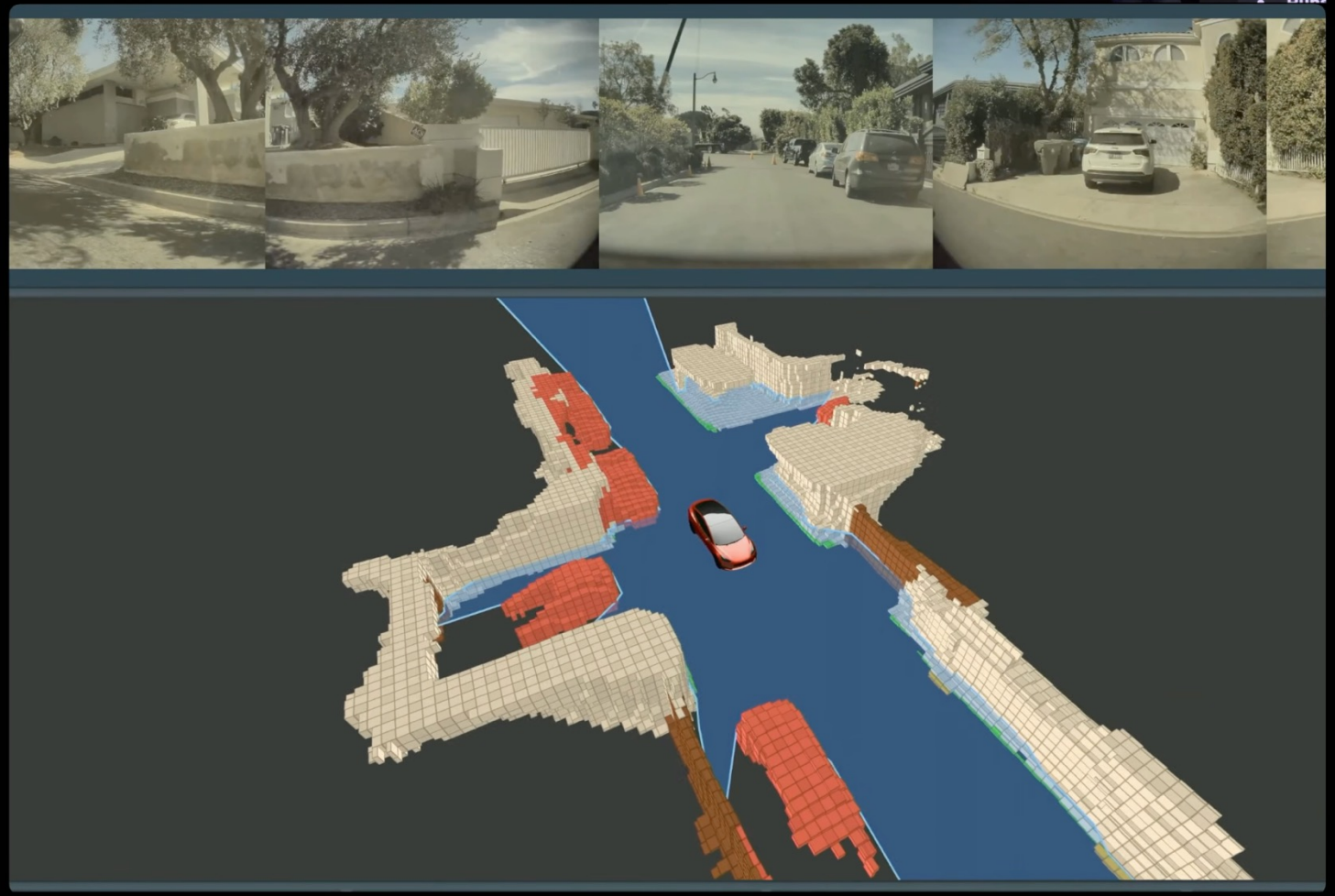
Occupancy Network

Occupancy Network

Properties



- Volumetric Occupancy
- Multi-Camera & Video Context
- Persistent Through Occlusions
- Occupancy Semantics
- Occupancy Flow
- Resolution Where It Matters
- Efficient Memory and Compute
- Runs in ~10 Milliseconds

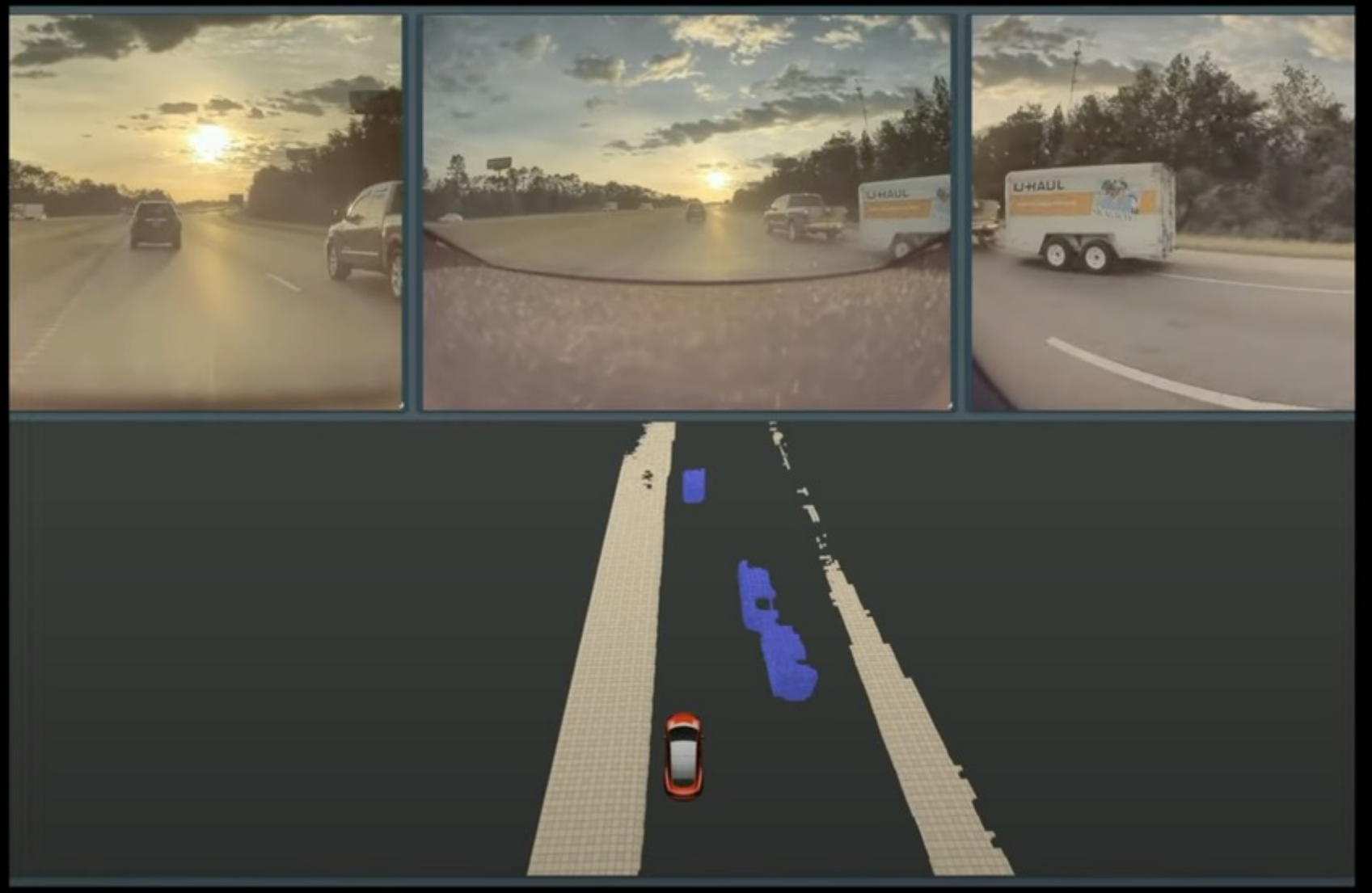


TESLA LIVE

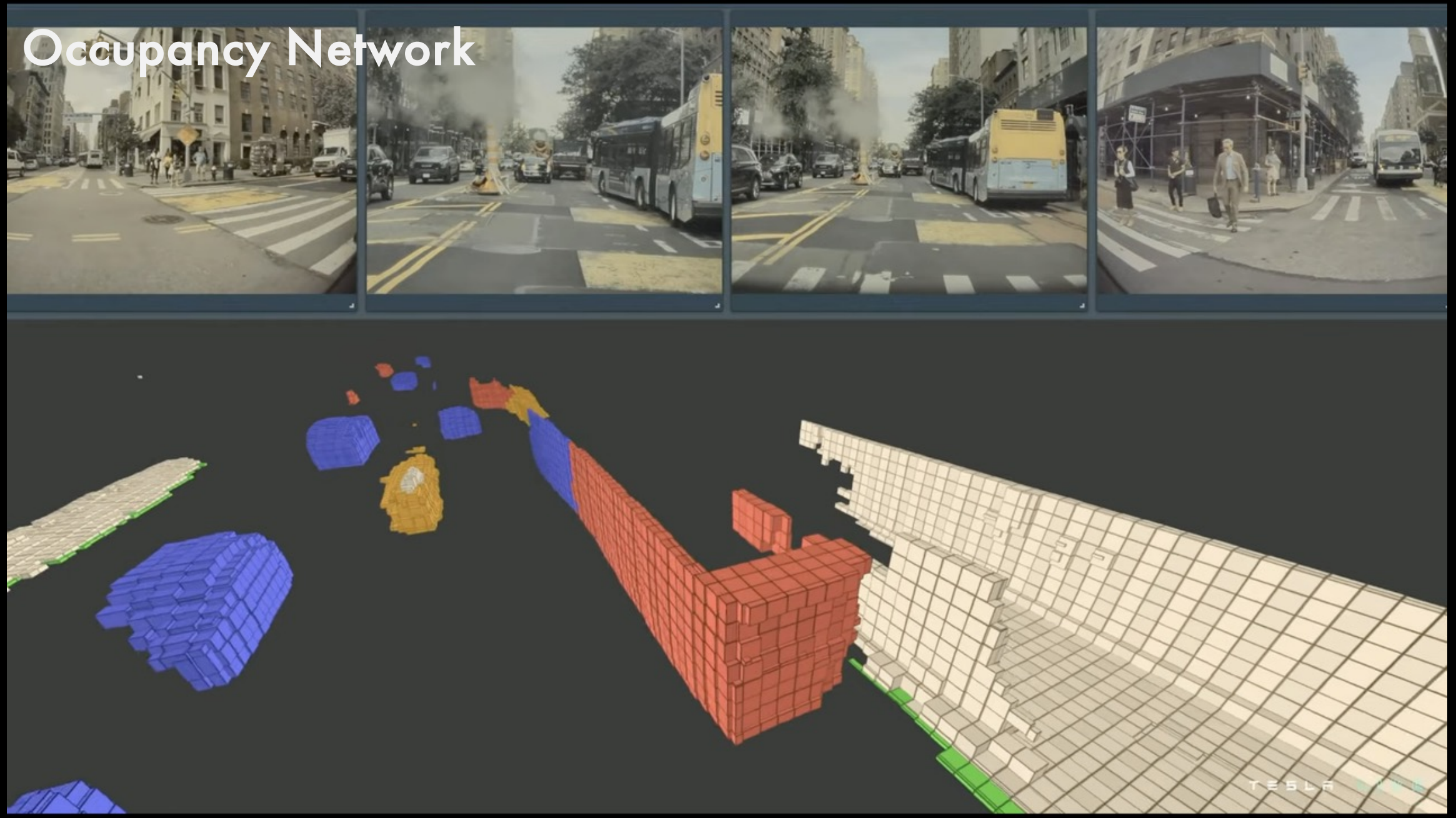
Occupancy Network

Properties

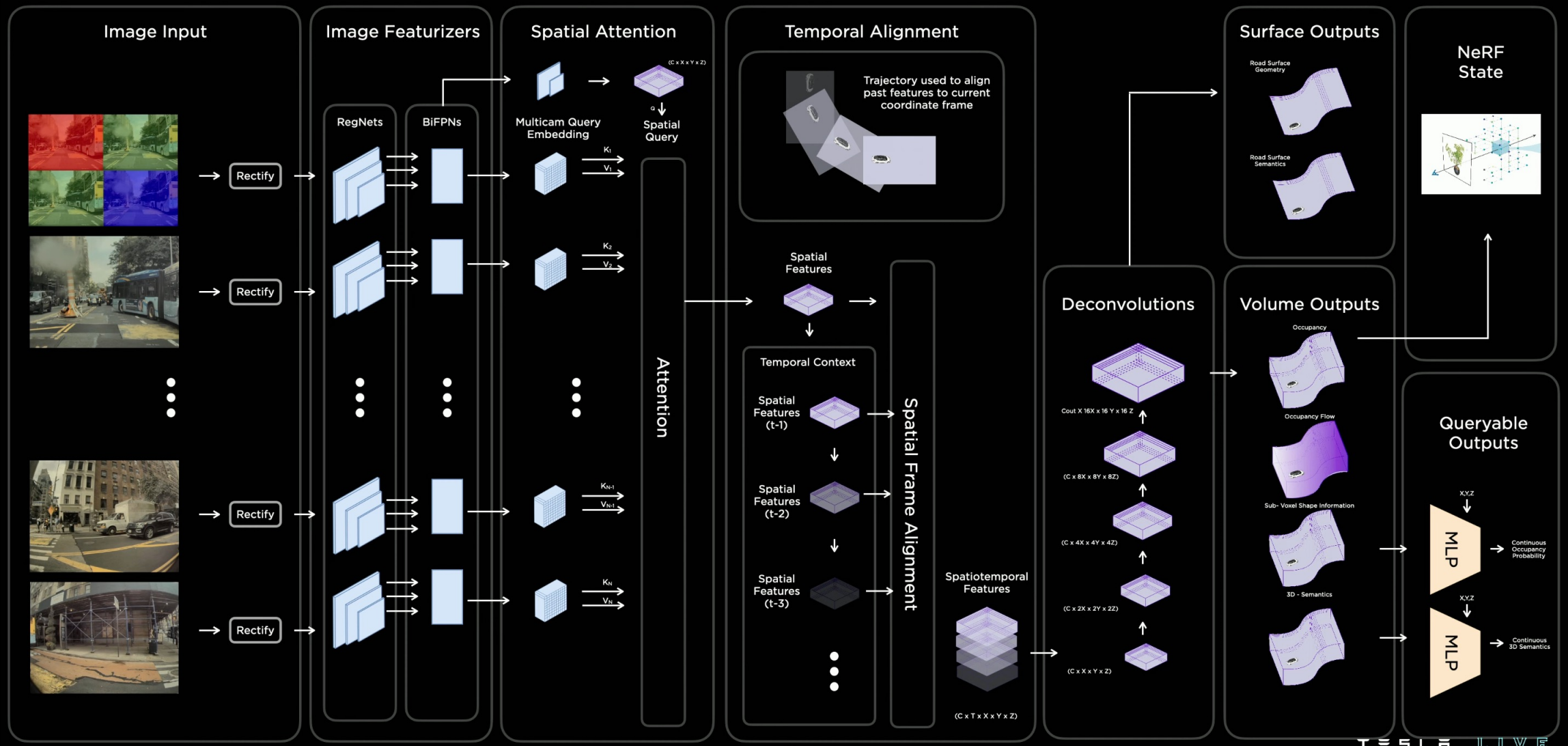
- Volumetric Occupancy
- Multi-Camera & Video Context
- Persistent Through Occlusions
- Occupancy Semantics
- Occupancy Flow
- Resolution Where It Matters
- Efficient Memory and Compute
- Runs in ~10 Milliseconds



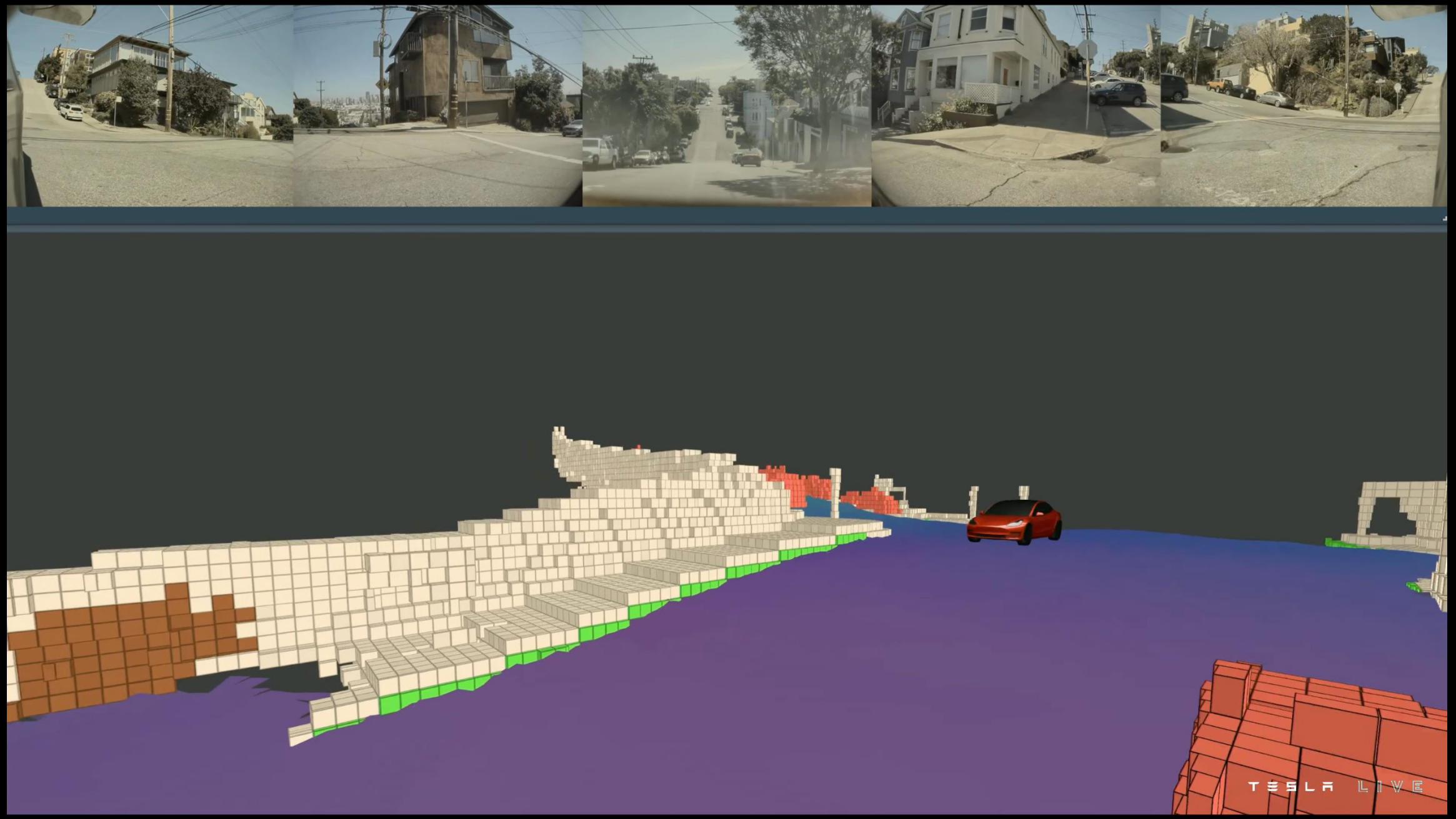
Occupancy Network



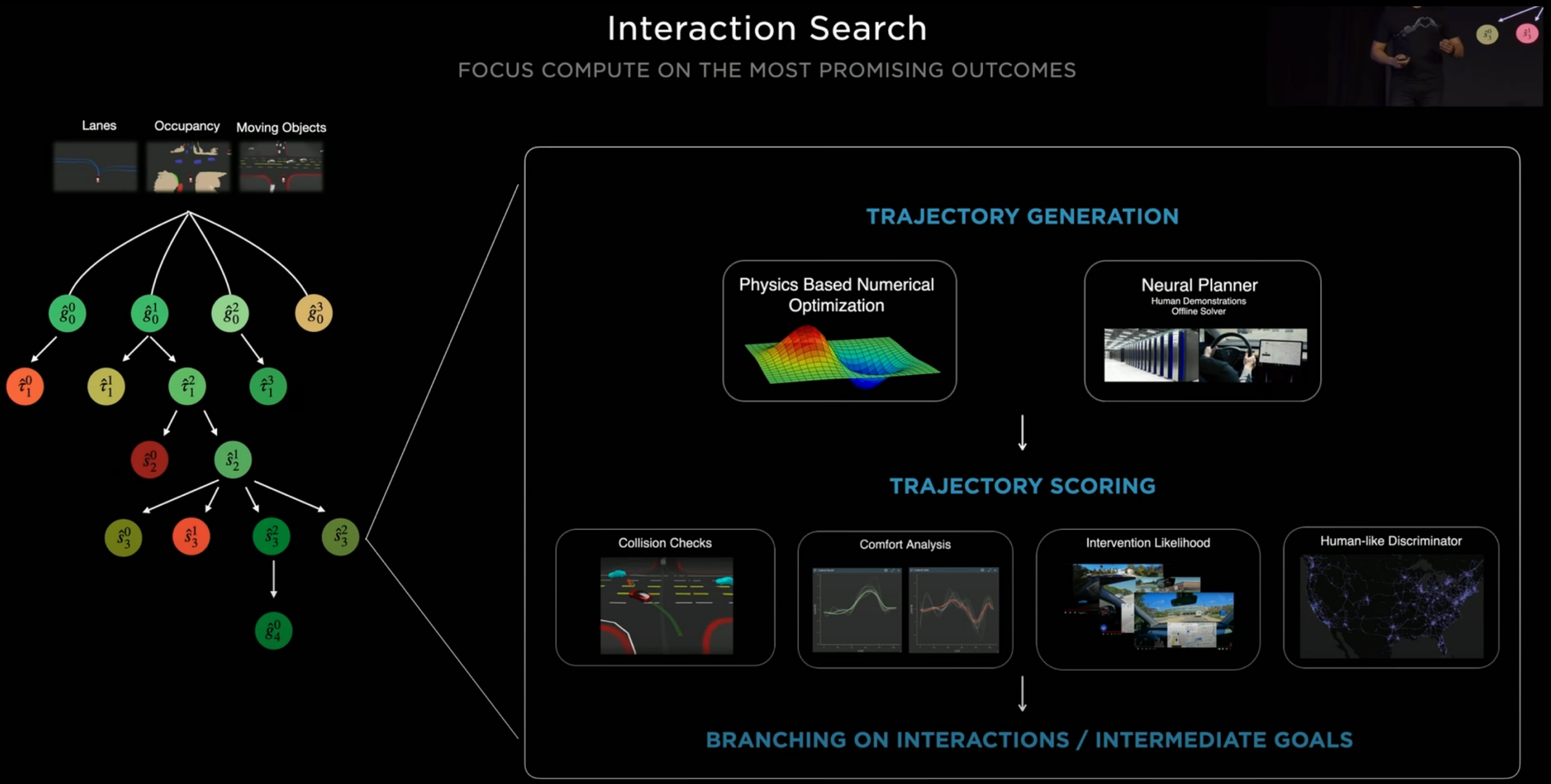
Occupancy Network



[1] Plenoxels: Radiance Fields without Neural Networks


TESLA LIVE

基于Vector Space的FSD路径规划



2.4 E2E 最新算法

理想汽车+清华大学 DriveVLM

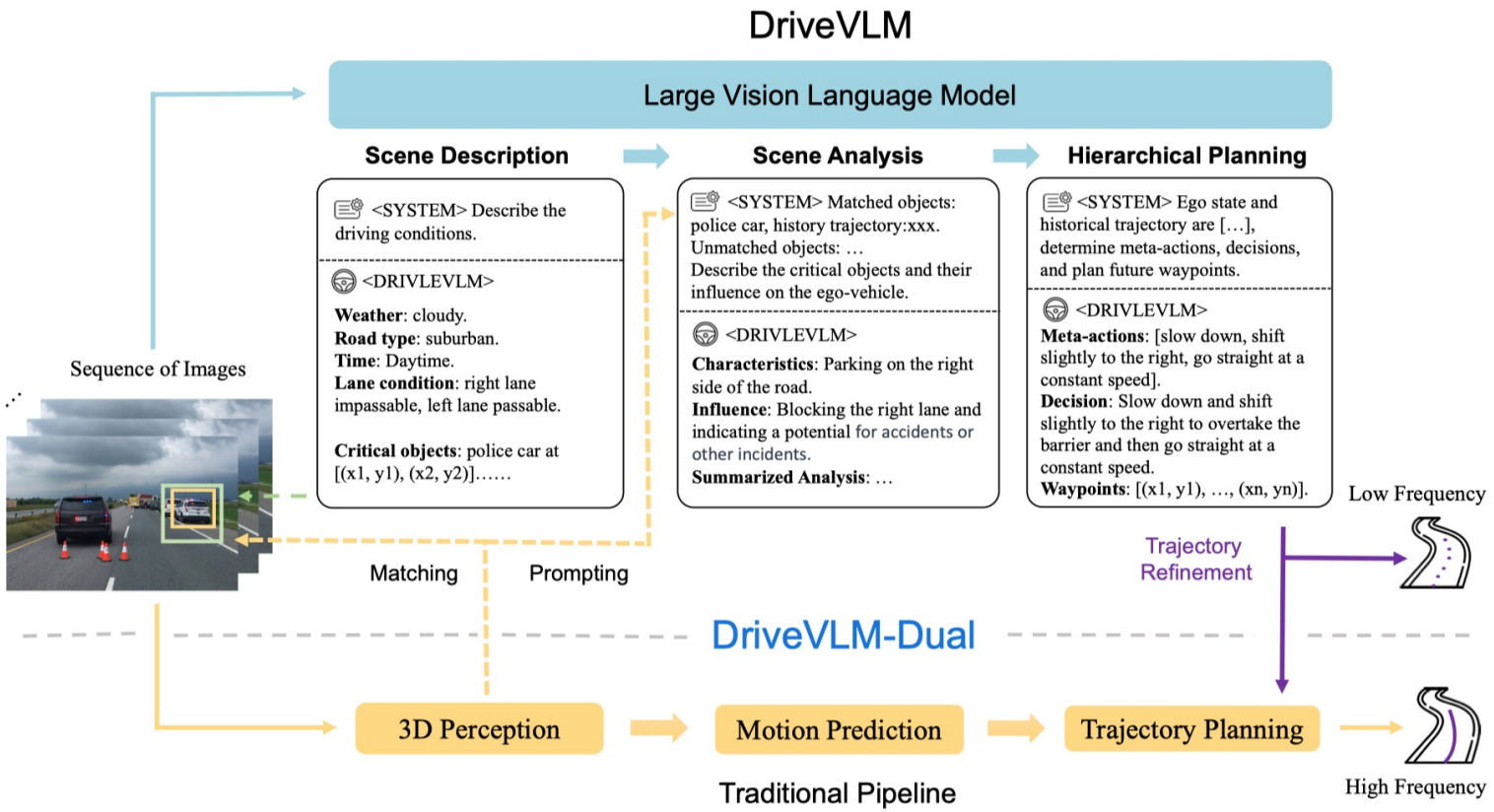


Figure 1: **DriveVLM and DriveVLM-Dual model pipelines.** DriveVLM takes images as input and, through a Chain-of-Thought (CoT) mechanism, outputs scene description, scene analysis, and hierarchical planning results. DriveVLM-Dual further incorporates traditional 3D perception and trajectory planning modules to achieve spatial reasoning capability and real-time trajectory planning.

理想汽车+清华大学 DriveVLM

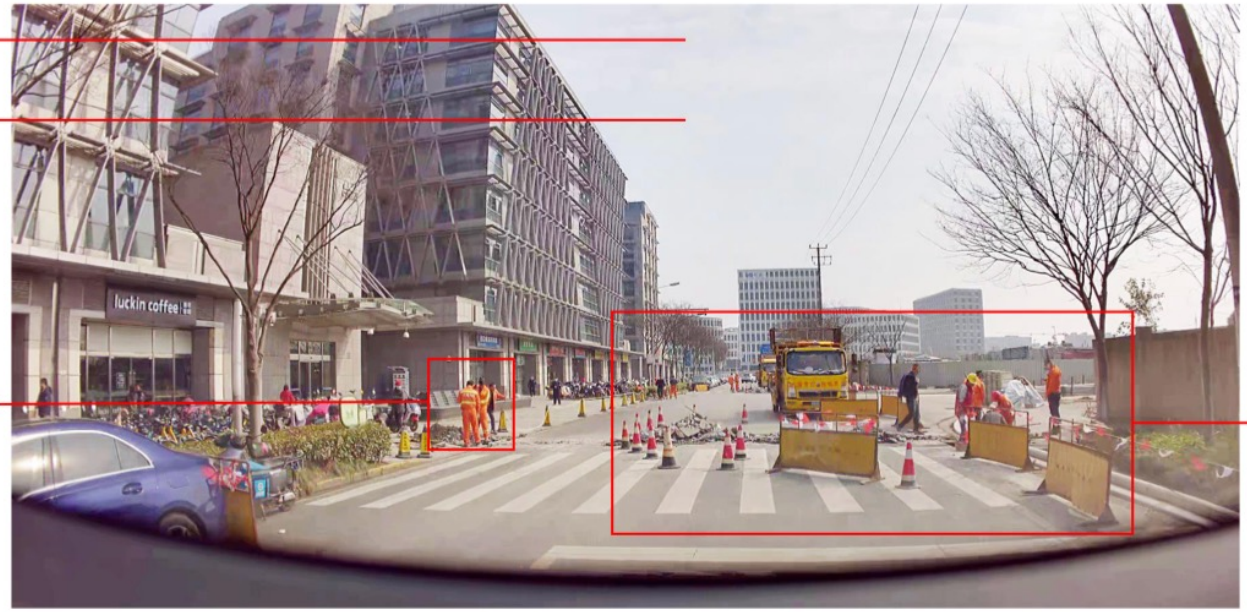
Scene Summary: The ego vehicle is moving at a constant speed along the current lane, with ongoing road construction work ahead; there are three construction workers working on the left side of the lane at the roadside.

Weather: Sunny

Time: Daytime

Critical Object 1:

Class:	Three Construction Workers
Characteristic:	Construction work on the side of the lane to the left of the host vehicle
Influence:	Affects the normal speed of the host vehicle



Road Condition: Construction

Lane Condition: Own Lane

Critical Object 2:

Class:	Construction Zone
Characteristic:	Road repair in front of the host vehicle lane
Influence:	Affects the host vehicle to drive straight normally

Meta Action: ["Slow down", "Change lane to the left", "Go straight slowly"]

Decision Description: Decelerate and change lanes to the left, keeping a safe distance from the construction workers on the left front side.

Figure 2: **An annotated sample of the SUP-AD dataset.**

地平线 VADv2

- VADv2: End-to-End Vectorized Autonomous Driving via Probabilistic Planning

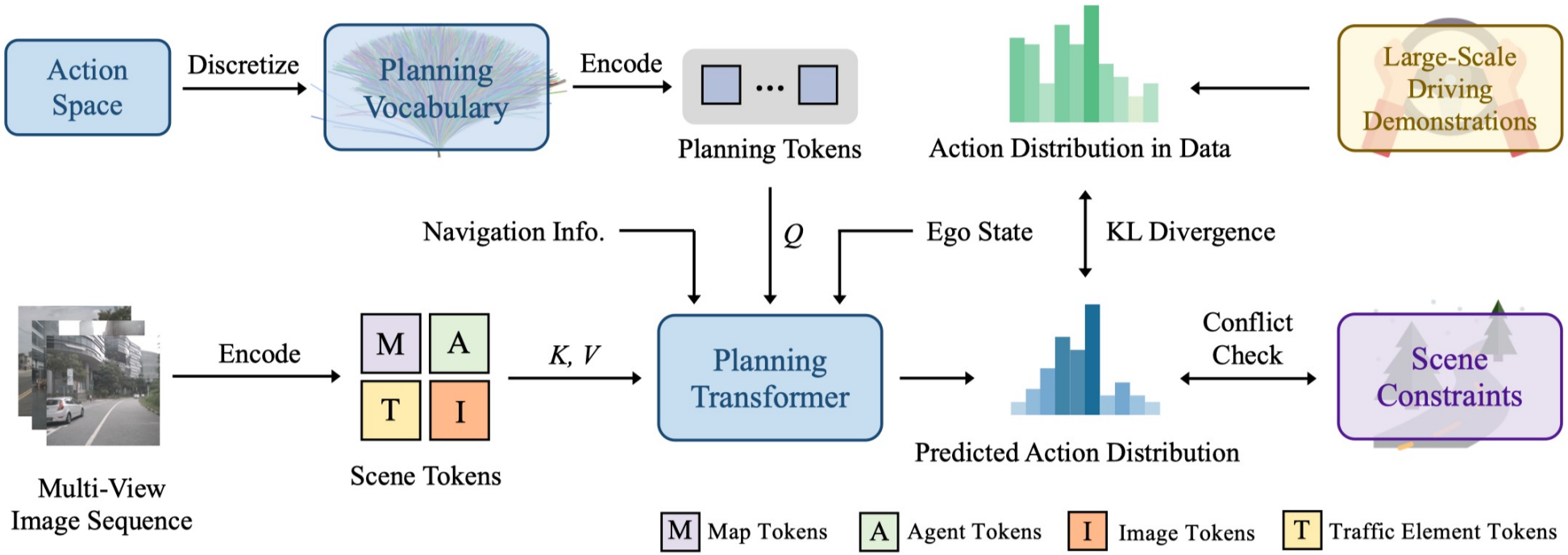
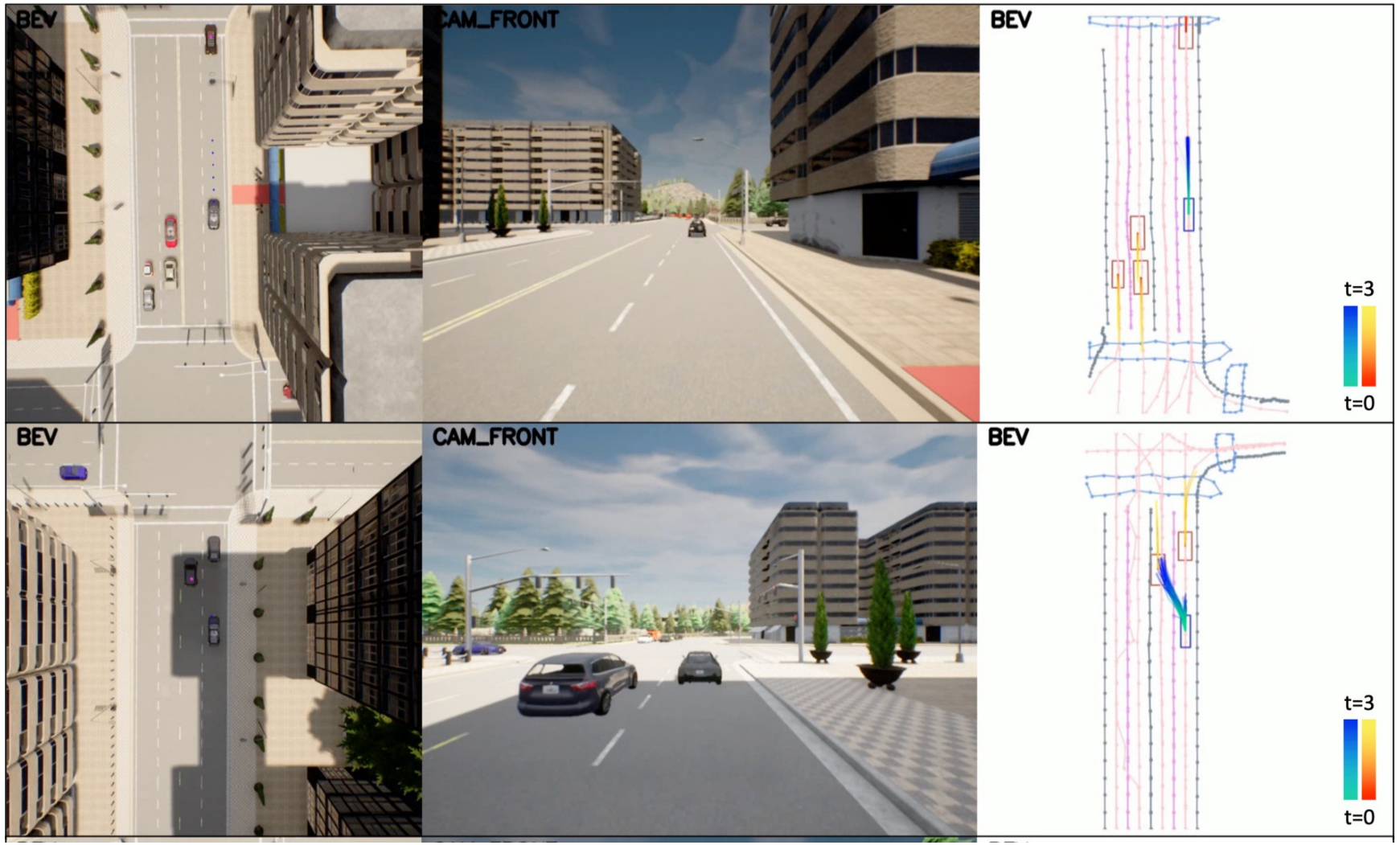
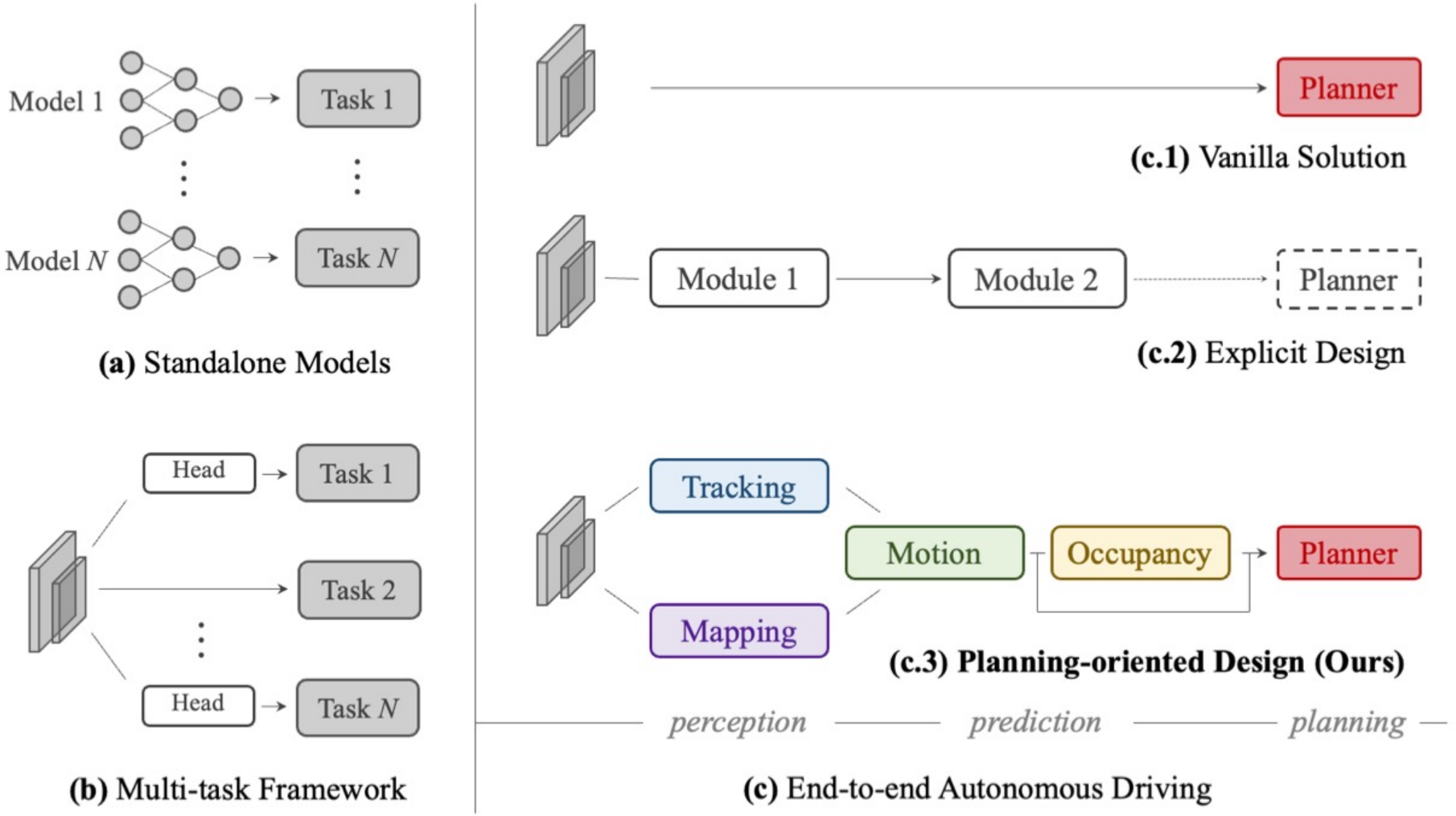


Figure 2. **Overall architecture of VADv2.** VADv2 takes multi-view image sequence as input in a streaming manner, transforms sensor data into environmental token embeddings, outputs the probabilistic distribution of action, and samples one action to control the vehicle. Large-scale driving demonstrations and scene constraints are used to supervise the predicted distribution.

地平线 VADv2



UniAD 2023 CPVR Best Paper



- UniAD 4个模块：
  - 特征提取
  - 感知模块
  - 预测模块
  - 规划模块
- Planning 为导性的网络模型。
- 单模型统一传统AD技术方案。

UniAD 2023 CPVR Best Paper

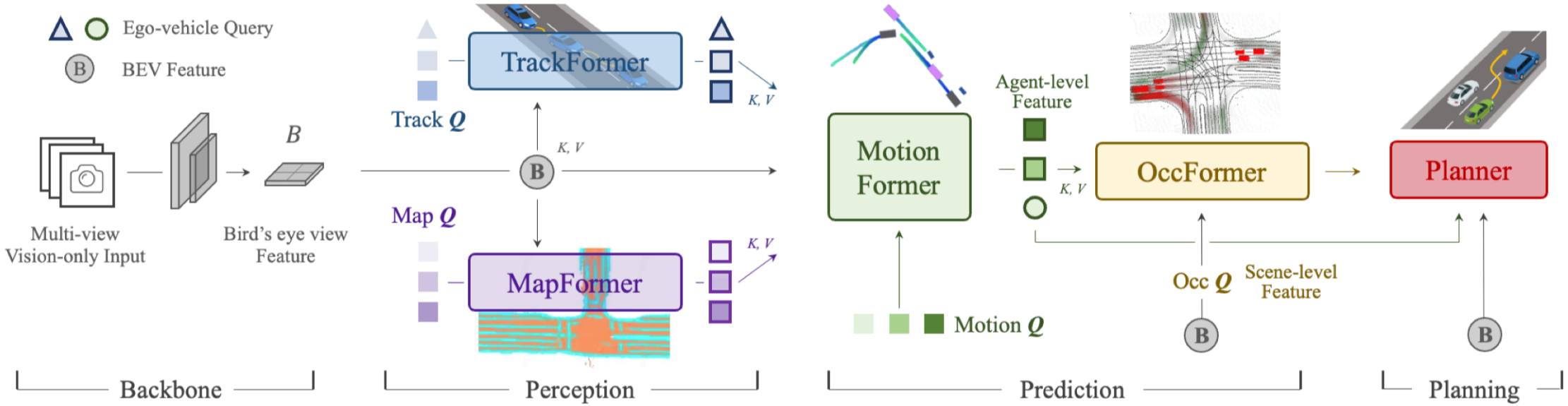


Figure 2. **Pipeline** of Unified Autonomous Driving (UniAD). It is exquisitely devised following planning-oriented philosophy. Instead of a simple stack of tasks, we investigate the effect of each module in perception and prediction, leveraging the benefits of joint optimization from preceding nodes to final planning in the driving scene. All perception and prediction modules are designed in a transformer decoder structure, with task queries as interfaces connecting each node. A simple attention-based planner is in the end to predict future waypoints of the ego-vehicle considering the knowledge extracted from preceding nodes. The map over occupancy is for visual purpose only.

UniAD 2023 CPVR Best Paper

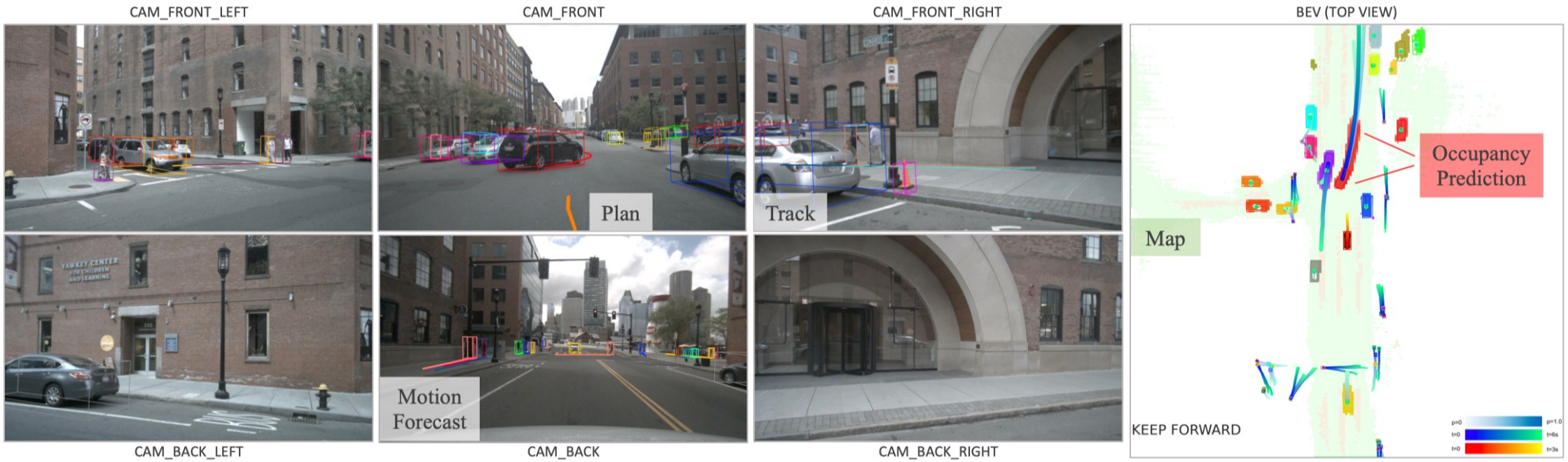
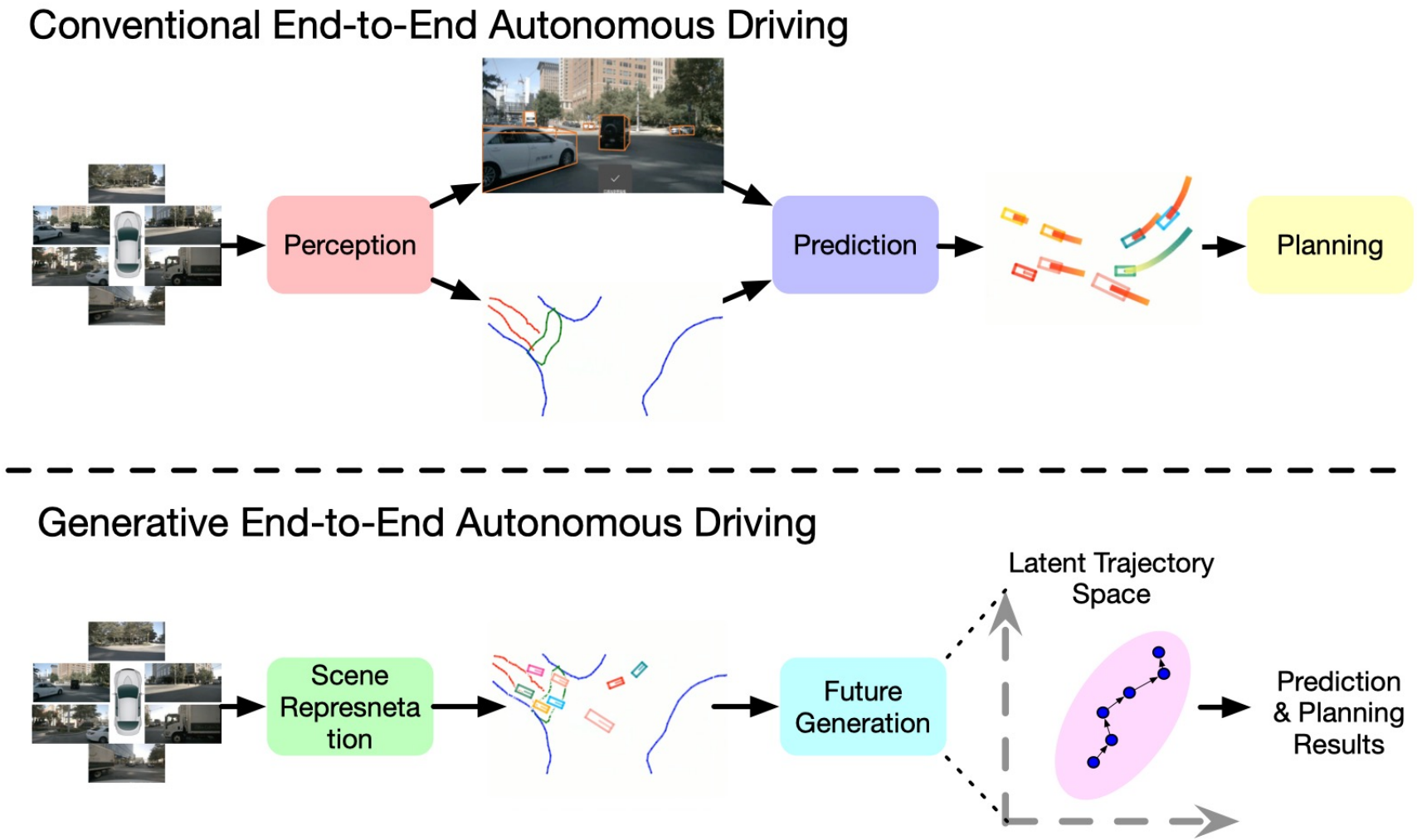


Figure 3. **Visualization results.** We show results for all tasks in surround-view images and BEV. Predictions from motion and occupancy modules are consistent, and the ego vehicle is yielding to the front black car in this case. Each agent is illustrated with a unique color. Only top-1 and top-3 trajectories from motion forecasting are selected for visualization on image-view and BEV respectively.

GenAD: Generative End-to-End Autonomous Driving



GenAD: Generative End-to-End Autonomous Driving

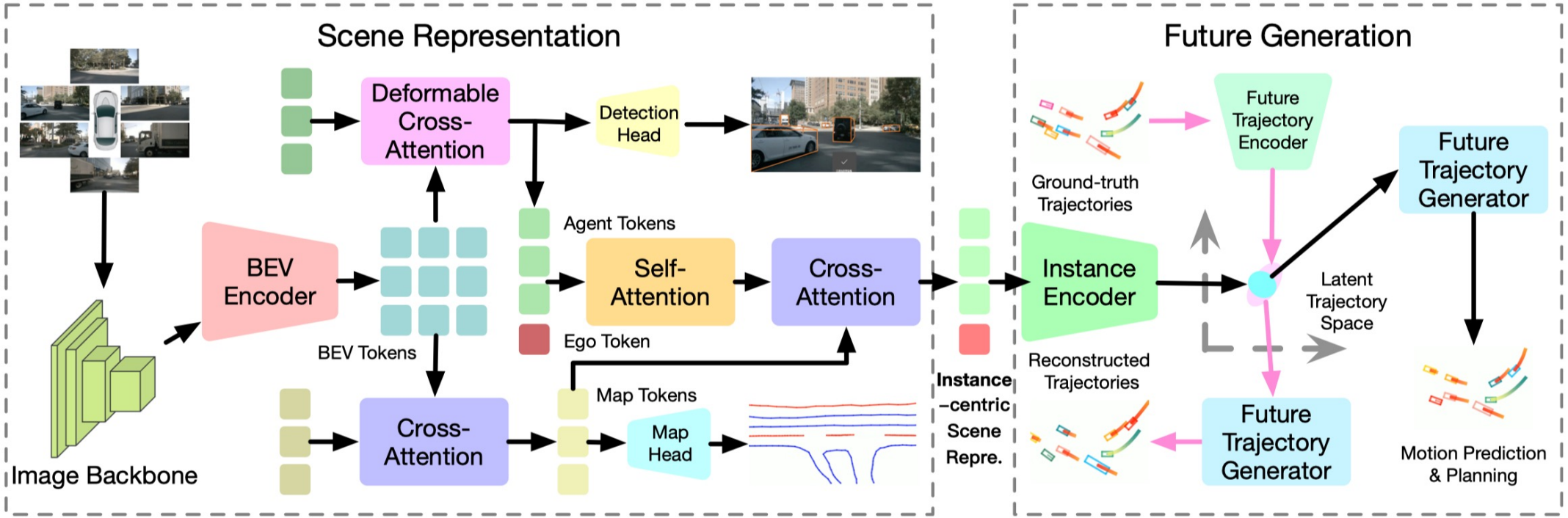
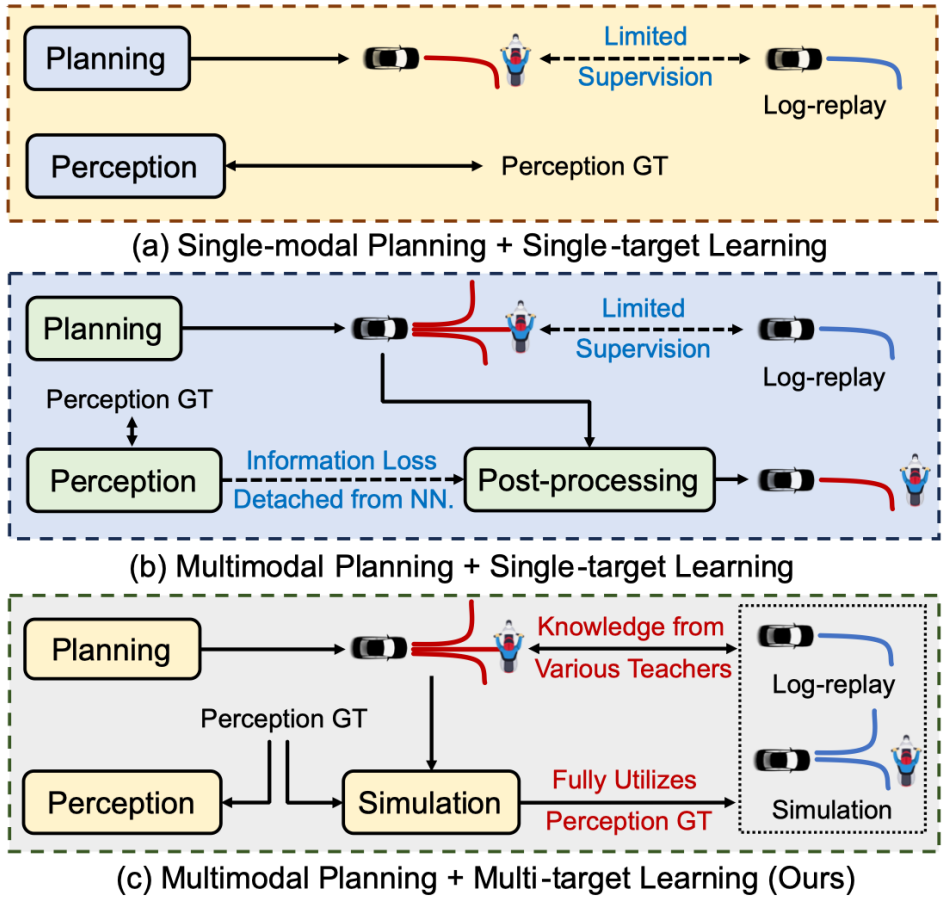


Figure 2. **Framework of our generative end-to-end autonomous driving.** Given surrounding images as inputs, we employ an image backbone to extract multi-scale features and then use a BEV encoder to obtain BEV tokens. We then use cross-attention and deformable cross-attention to transform BEV tokens into map and agent tokens, respectively. With an additional ego token, we use self-attention to enable ego-agent interactions and cross-attention to further incorporate map information to obtain the instance-centric scene representation. We map this representation to a structural latent trajectory space which is jointly learned using ground-truth future trajectories. Finally, we employ a future trajectory generator to produce future trajectories to simultaneously complete motion prediction and planning.

NVIDIA Hydra-MDP



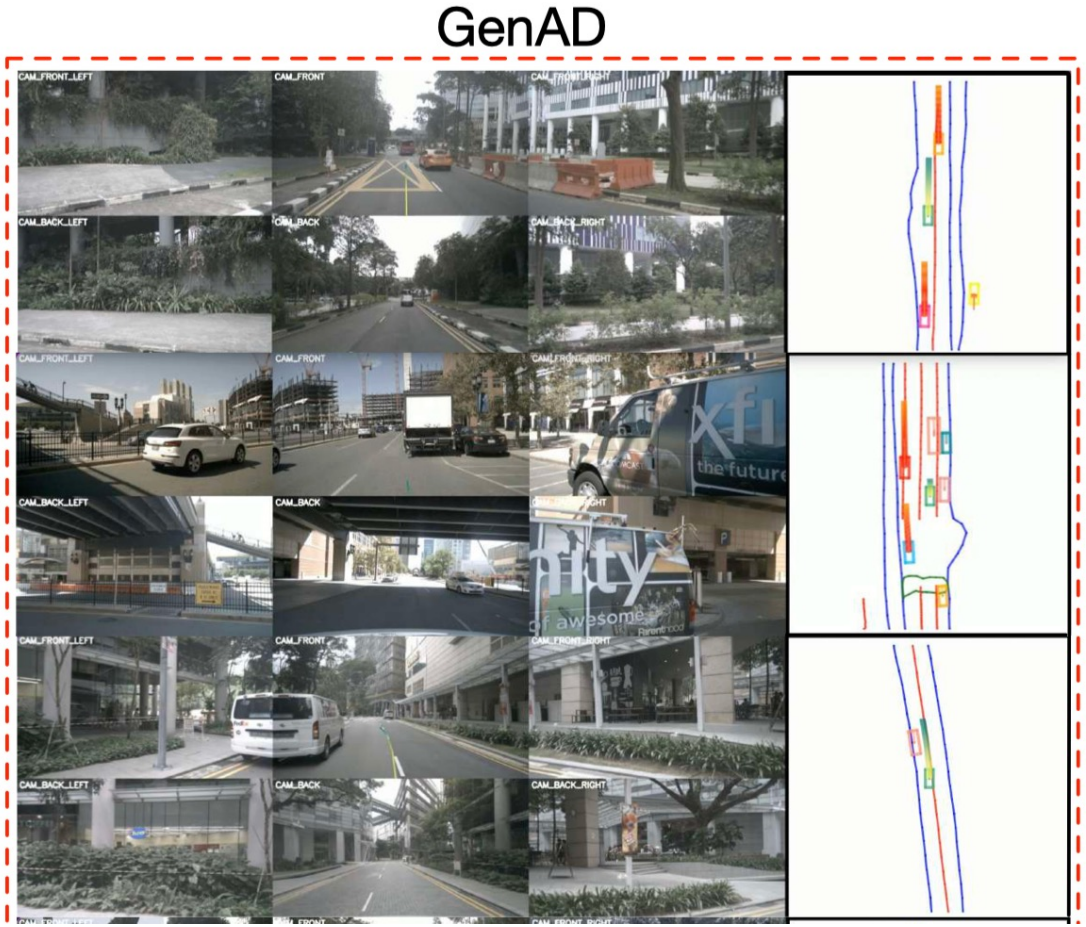


Figure 1. **Comparison between End-to-end Planning Paradigms.**

NVIDIA Hydra-MDP

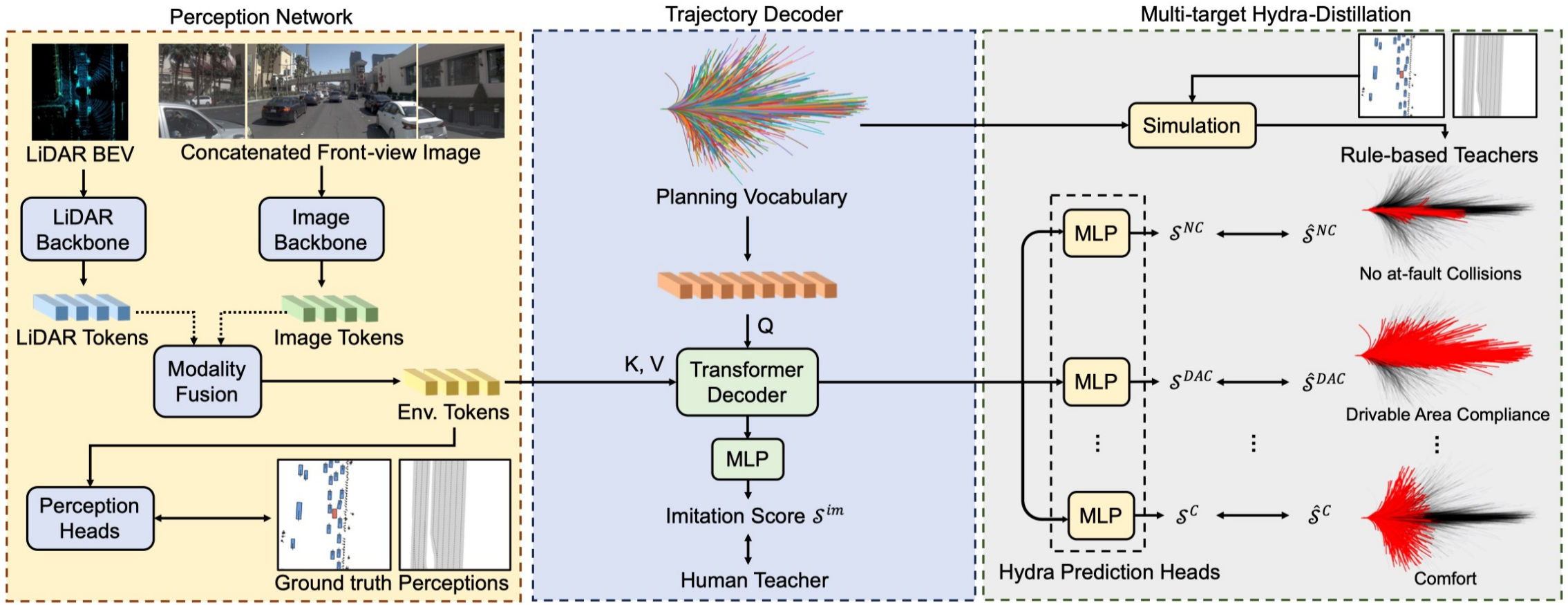
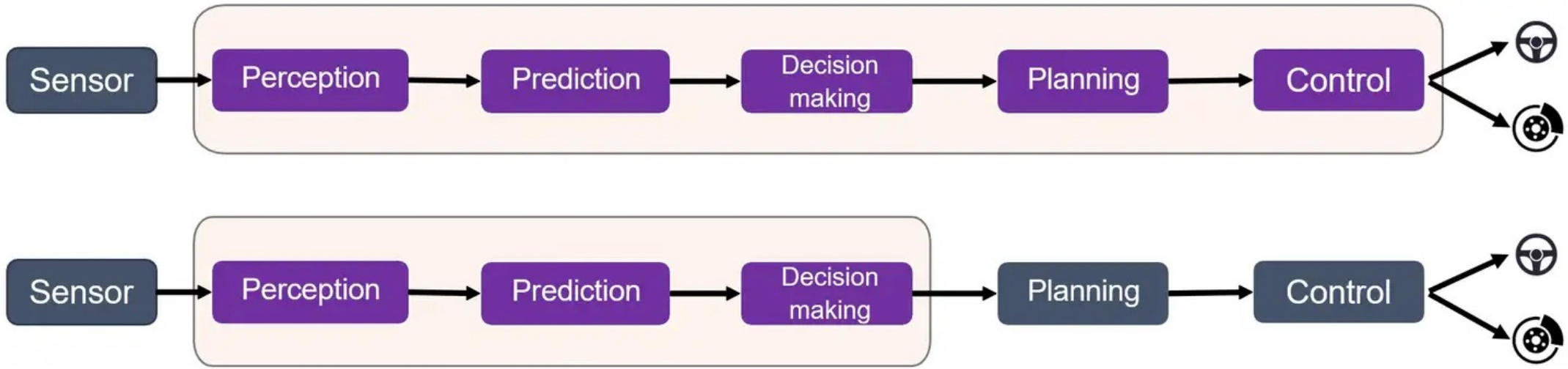


Figure 2. **The Overall Architecture of Hydra-MDP.**

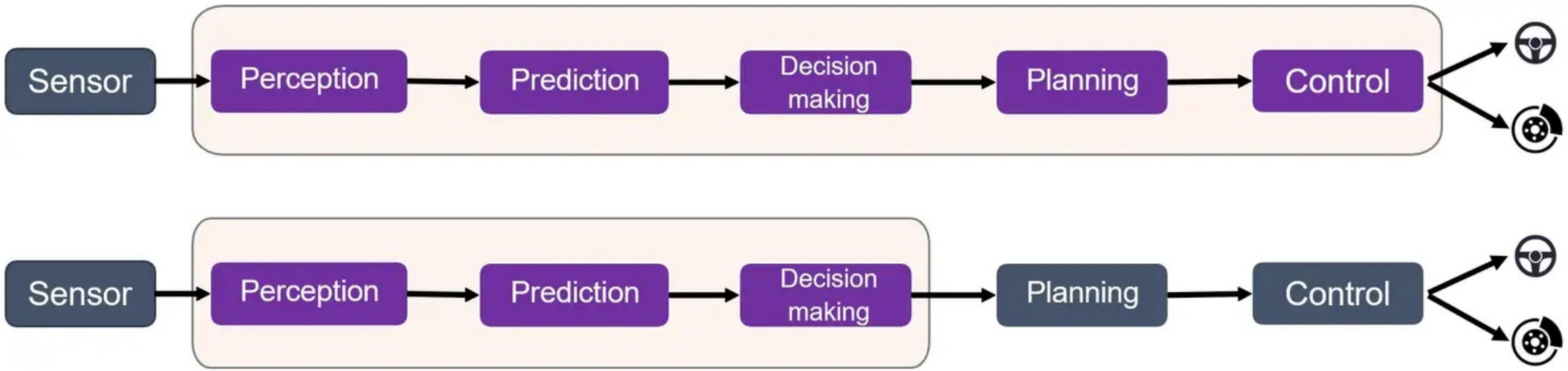
端到端式的技术路线：从感知出发

- **从感知出发，先解决感知与规划间潜层信息丢失的问题：**
  - 规划想要使用感知的特征信息必须要有有规划模型，但业界规划效果特别明显的模型少。为了让规划模型化，学术界适用模型来直接输出规划轨迹。



端到端式的技术路线：从规划出发

- **先解决规划问题，因其对驾驶效果影响最直接的模块：**
  - 方案 1：先用业界比较成熟的预测模型出个主车的粗规划结果，再用规则 rule-base 来修正；
  - 方案 2：使用多模型进行评价，输出大量待选轨迹，通过规则来实现轨迹初选。
  - 方案 3：上述方案结合，预测和多模型打分适用一个融合模型。



03. 技术思考

思考点

1. 端到端感知，端到端决策规划都可以算作是端到端自动驾驶?

- 学术没有给出明确结论，无论是感知后处理，还是规划候选轨迹评分，甚至是安全兜底策略，引入了规则代码（if else 等），系统便不是全可微，损失了端到端通过训练获得全局优化的优势。

2. 端到端是对传统 AD技术的推倒重来?

- 包括特斯拉所在，传统 AD 技术积累并没有被抛弃，从台前迁移到幕后。1）端到端可以从原有技术基础上，逐步减少规则代码实现端到端可导。2）遇到未知情况仍然由传统 AD 接管。

Thank you

把AI系统带入每个开发者、每个家庭、每个组织，构建万物互联的智能世界

Bring AI System to every person, home and organization for a fully connected, intelligent world.

ZOMI

Course chenzomi12.github.io

GitHub github.com/chenzomi12/DeepLearningSystem