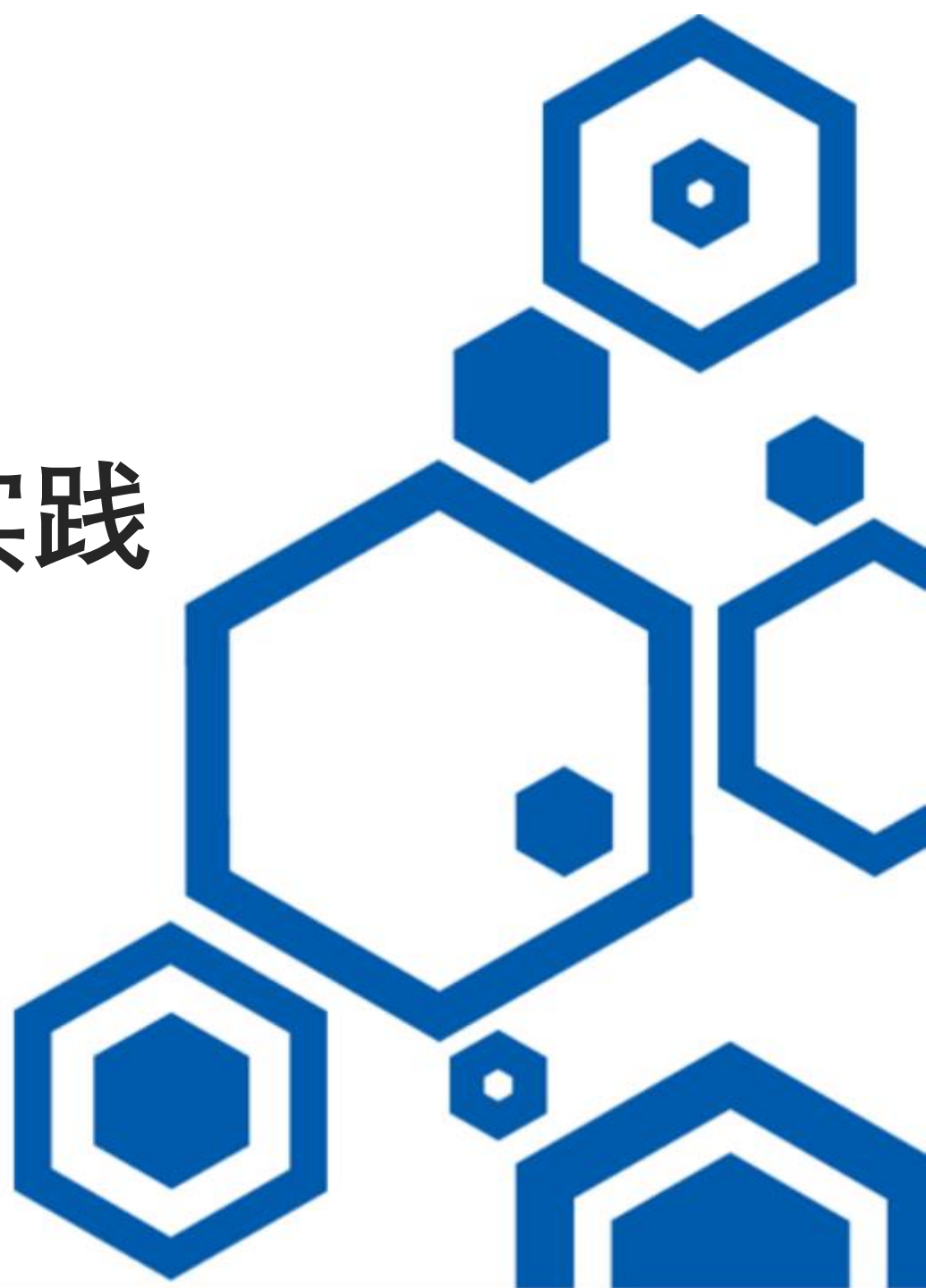


BEV感知系统理论与实践

Dongxu Fu





- 第一章. 从 L2 到 L2+ ，需求驱动技术发展
- 第二章. 从 Multi-View 到 BEV ，构建新的感知模型
- 第三章. 从 BEV 到 X ，让模型扩展到更多的任务
- 第四章. 从模型到量产：BEV模型设计及部署实践
- 第五章. 数据闭环，量产迭代引擎：如何更好的支持BEV？
- 第六章. 从 L4 到 L2 ，再到 L2+ ，探索未来技术趋势



第一章：从L2到L2+，需求驱动技术发展

- ◆ 当智驾需求从 L2 走到 L2+
- ◆ L2的感知系统及迭代方式，是否满足L2+？
- ◆ Tesla 是如何解决的？



一切源于智驾功能的新需求

L2 面向高速高架的辅助驾驶

L2 传感器主要关注前后向

L2

L2+

L2+ 进一步扩展高速高架场景功能，同时延伸至更加复杂的城区场景

L2+ 增加了上下匝道/自主超车（基于ALC的高级辅助驾驶功能），无保护左转功能

L2+ 传感器必须要关注周视



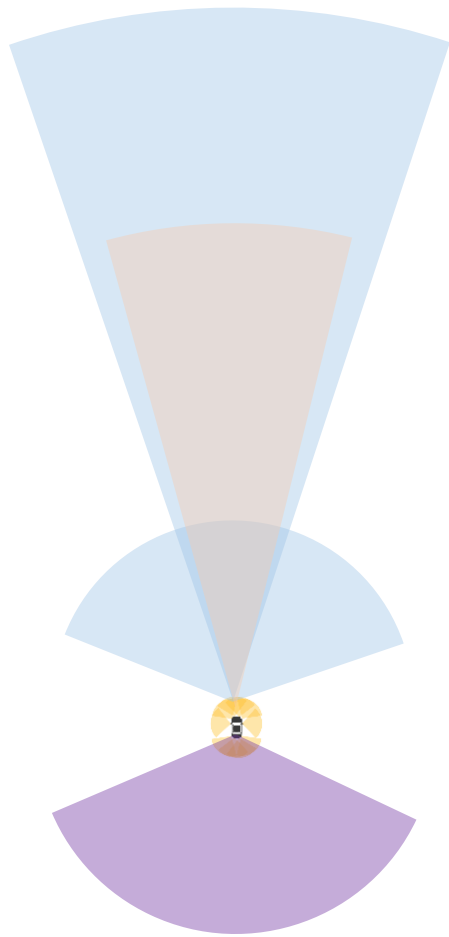
• 自适应巡航控制系统（ACC）



• 车道保持辅助系统（LKA）

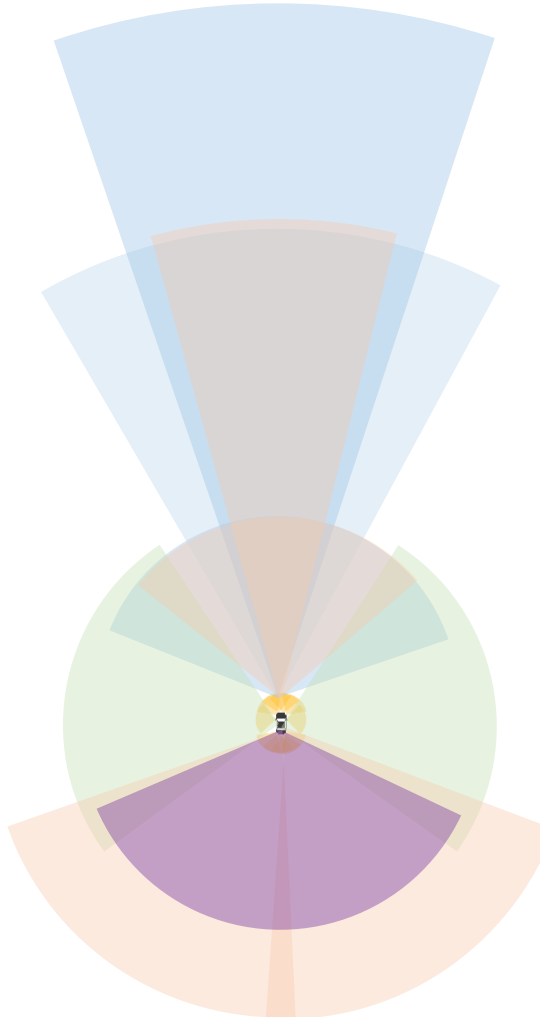


• 自动紧急刹车系统（AEB）



<20 TOPS

算力



200 -- 1000 TOPS



• 自动变道辅助系统（ALC）



• 自动汇入汇出

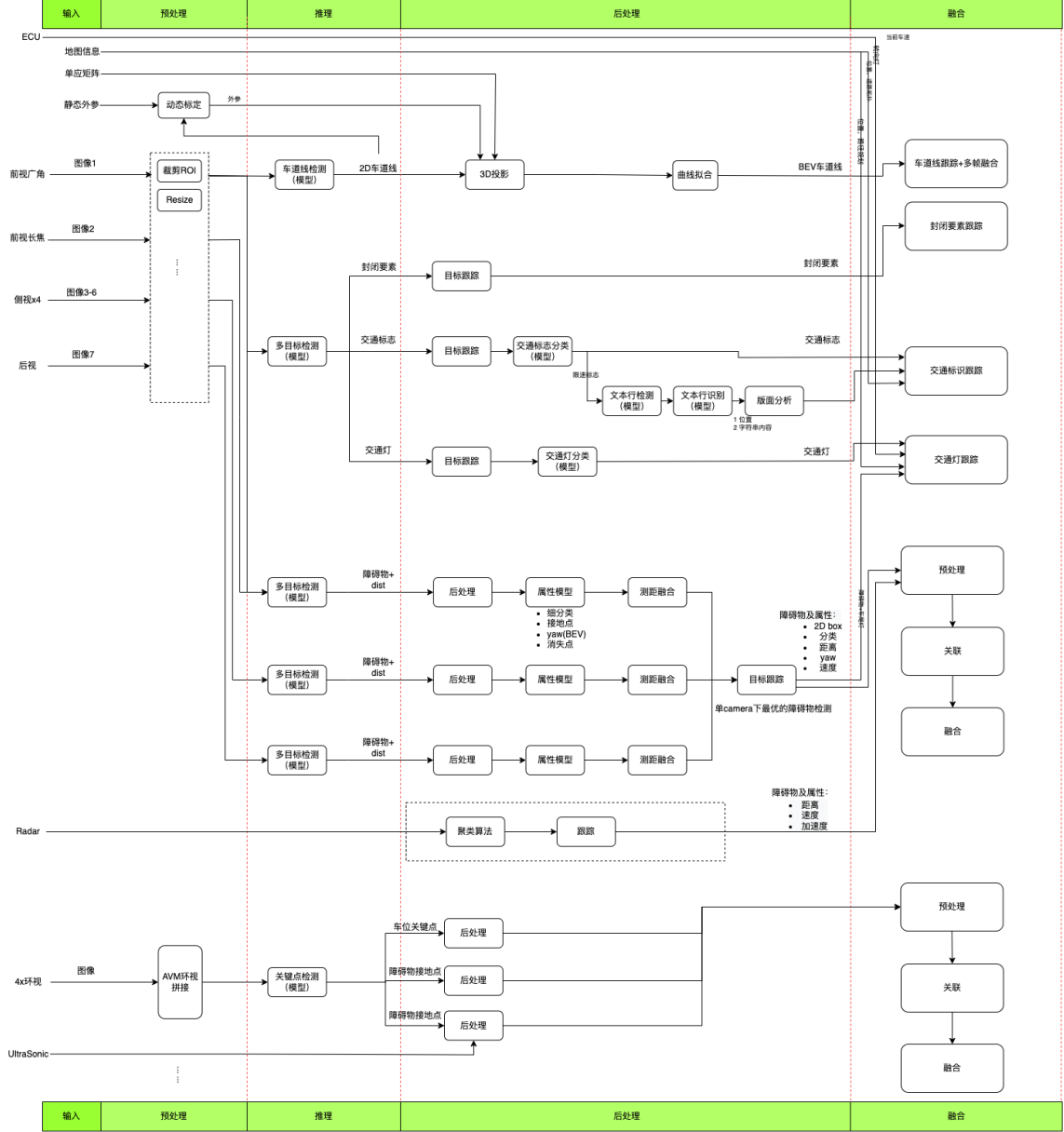


• 城区路口左转

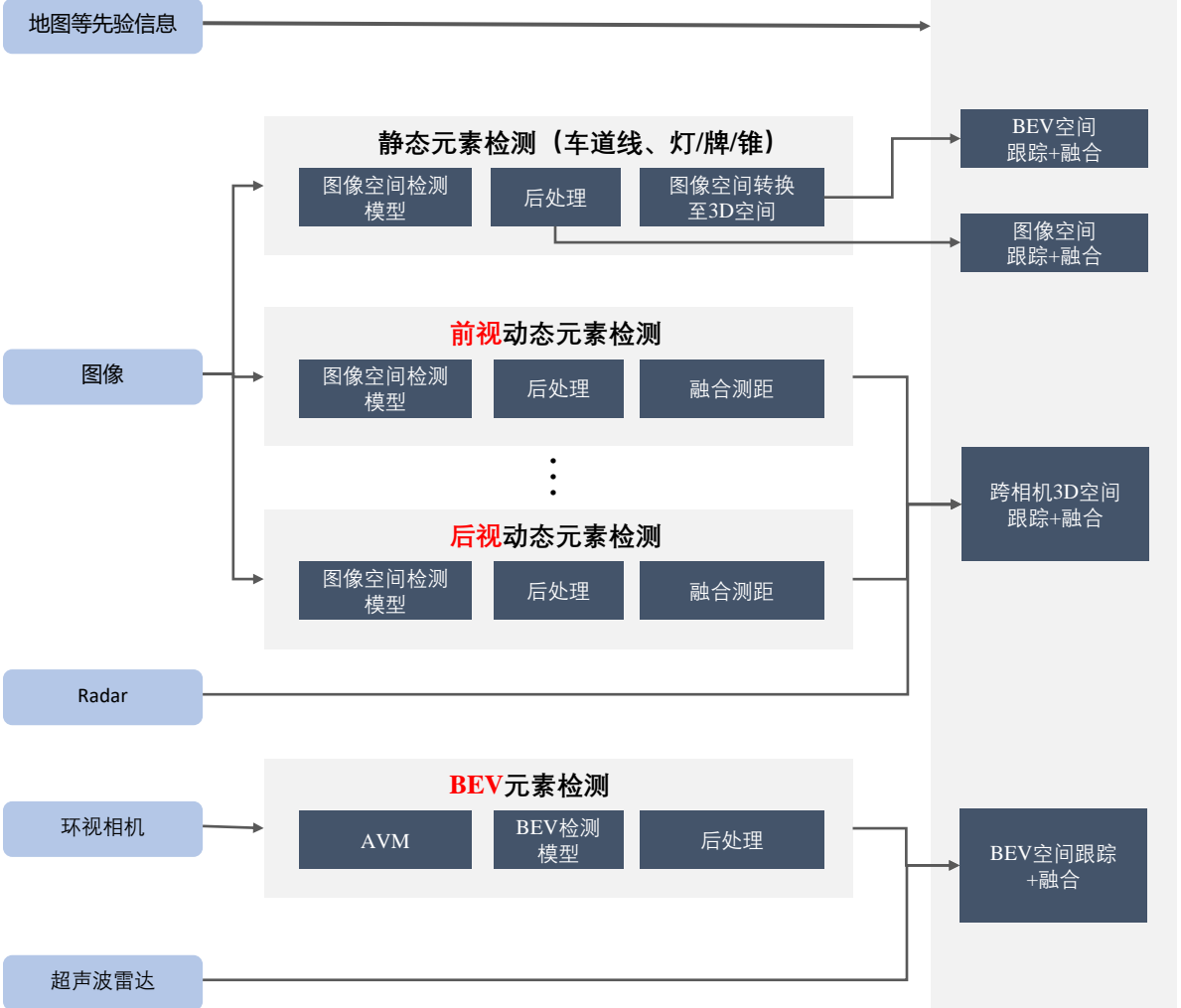


L2的感知系统能够支持吗？

真实的L2行泊感知系统

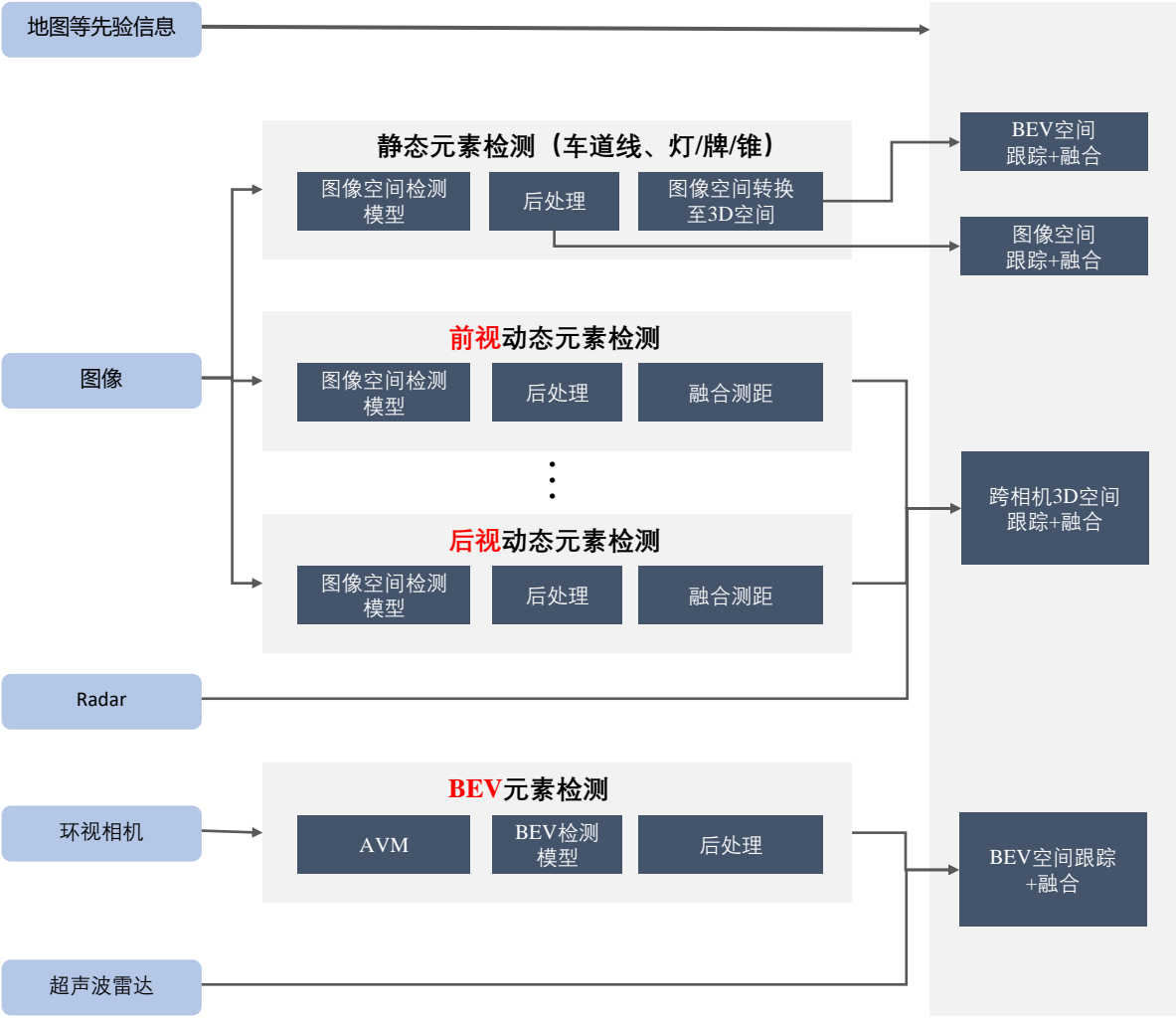


L2感知系统抽象



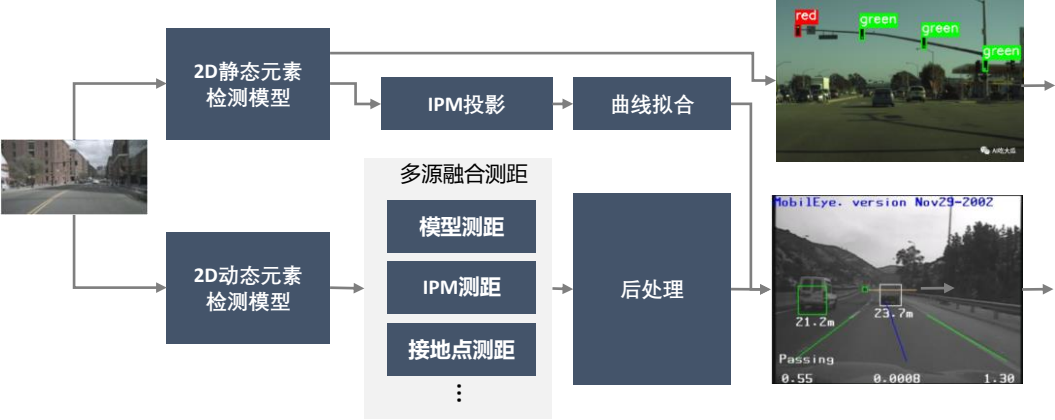


L2的感知系统能够支持吗？



算法实现细节1——以后融合为框架的感知，pipeline流程长，规则多

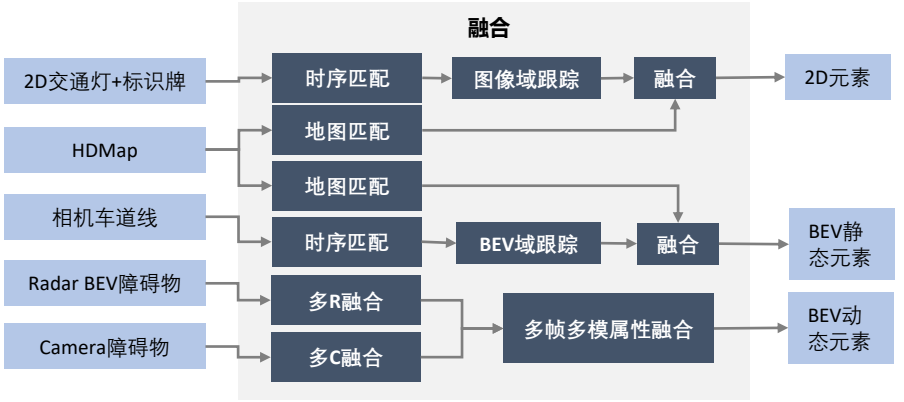
- 单目：障碍物、车道线检测 及其3D还原 + 交通灯 & 标识牌检测



- Radar检测



- 融合

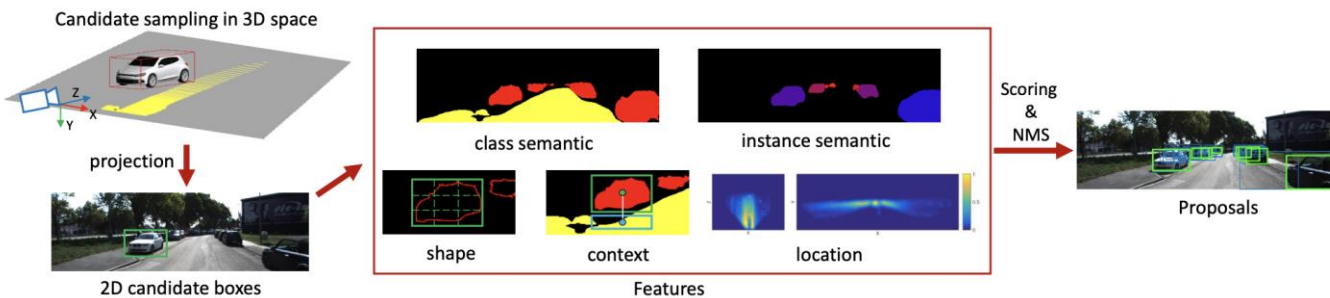




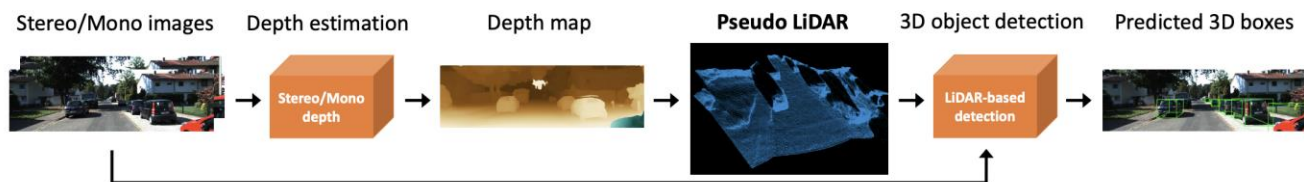
L2感知系统能够支持吗？

模型测距方法的性能不够

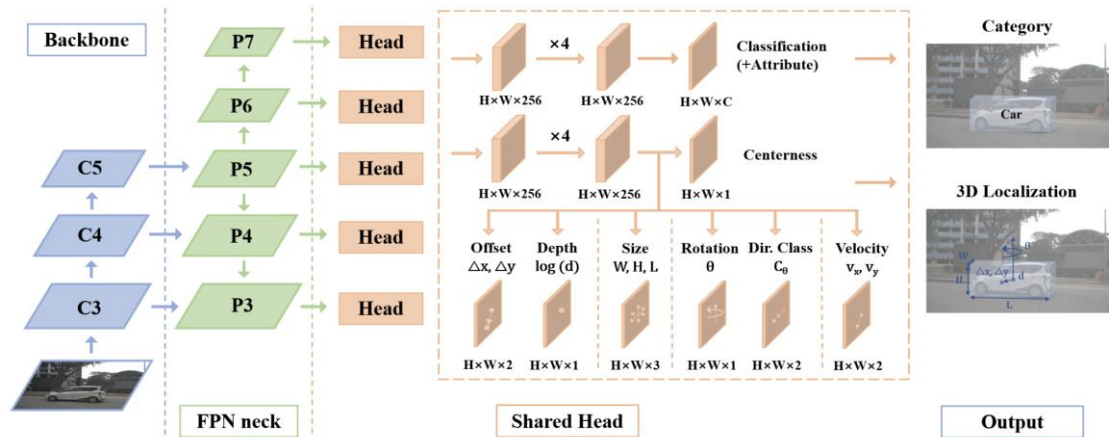
- Mono3D: 精度低, 鲁棒性差



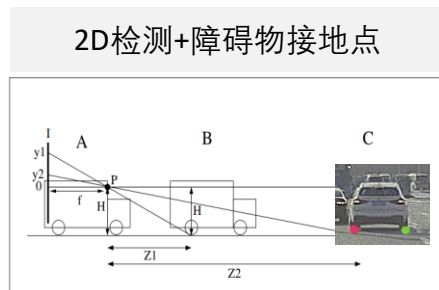
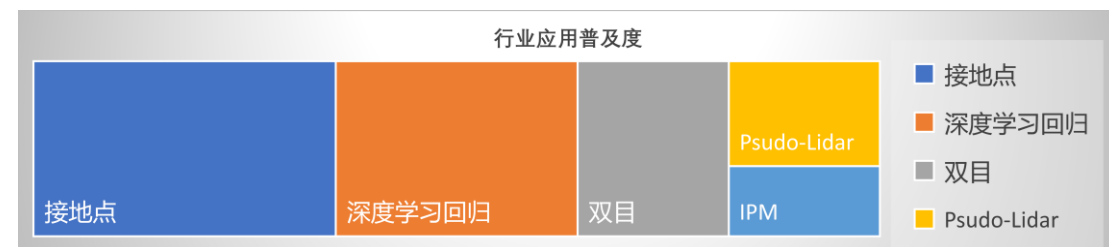
- Pseudo LiDAR: 算力开销大, 精度低



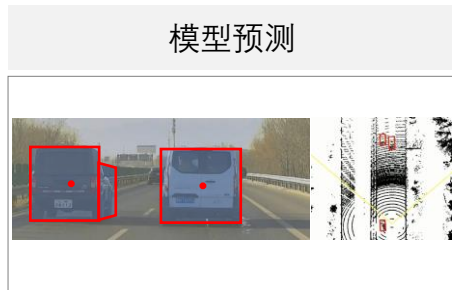
- FCOS3D: 数据量要求大



测距任务是重中之重, 但实现复杂, 规则多



- 地面平行假设
- 路面曲率, 车辆颠簸影响



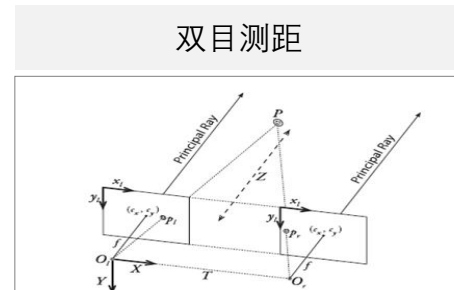
- 数据量要求高, 可解释性差
- 远距测距性能差 (GT不足)



- 地面平行假设
- 失真严重, 不适合障碍物



- 显存和算力消耗大
- 稠密点云真值难获取



- 视宽固定, 远近测距折衷
- 双目标定复杂, 成本高

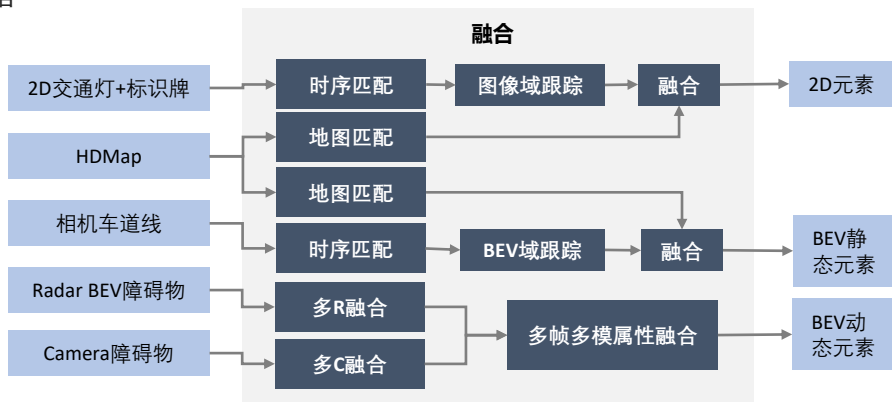
然而, 我们的目标, 四路环视如何加入?



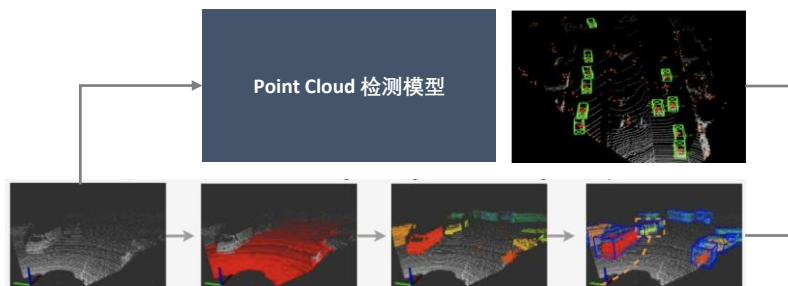
部分L2+系统加入LiDAR，怎么支持？

后融合？

- 融合



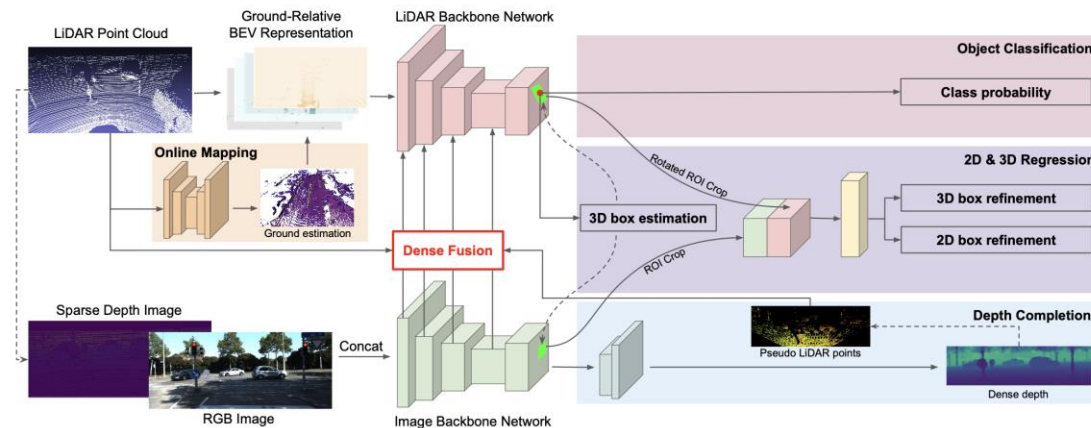
- LiDAR常用的算法：模型 + 通用聚类分割算法



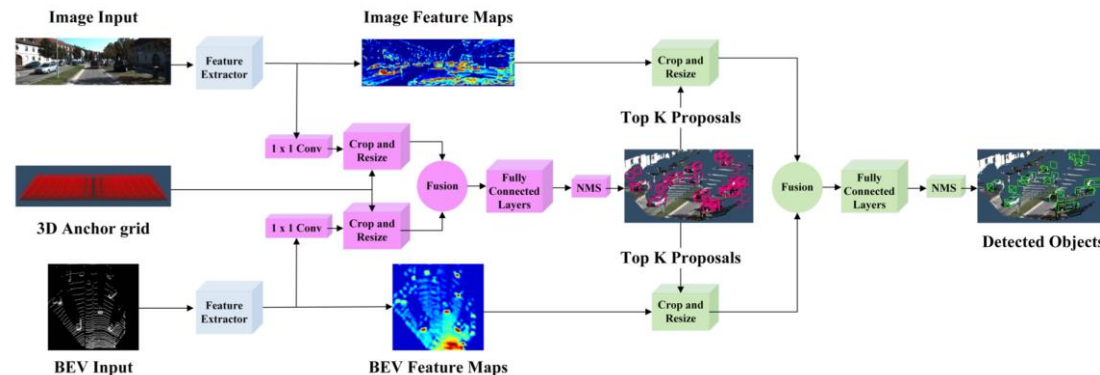
- 后融合需要一个更加复杂的融合模块，扩展性不好
- 前融合方法太重，且无论投影到2D平面或者3D空间，均不能最大化效果

可以支持LiDAR和Camera前融合的方法

- Feature projection-based方法：通过3D特征投影到2D（或者反向）实现特征融合，如 MMF

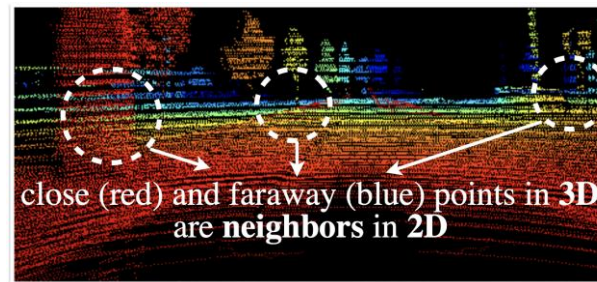


- Proposal-based方法：通过Proposal去截取2D和3D特征，并结合两者做检测，如AVOD

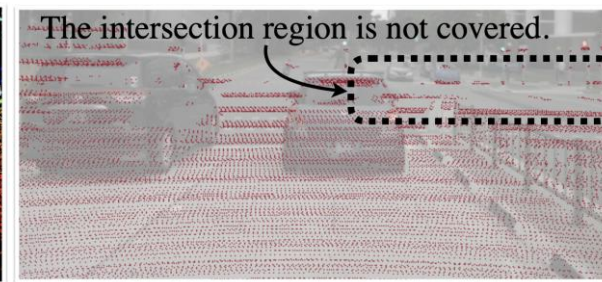


总结

- Geometric Lossy：点云投影到图像



- Semantic Lossy：像素集成到点云

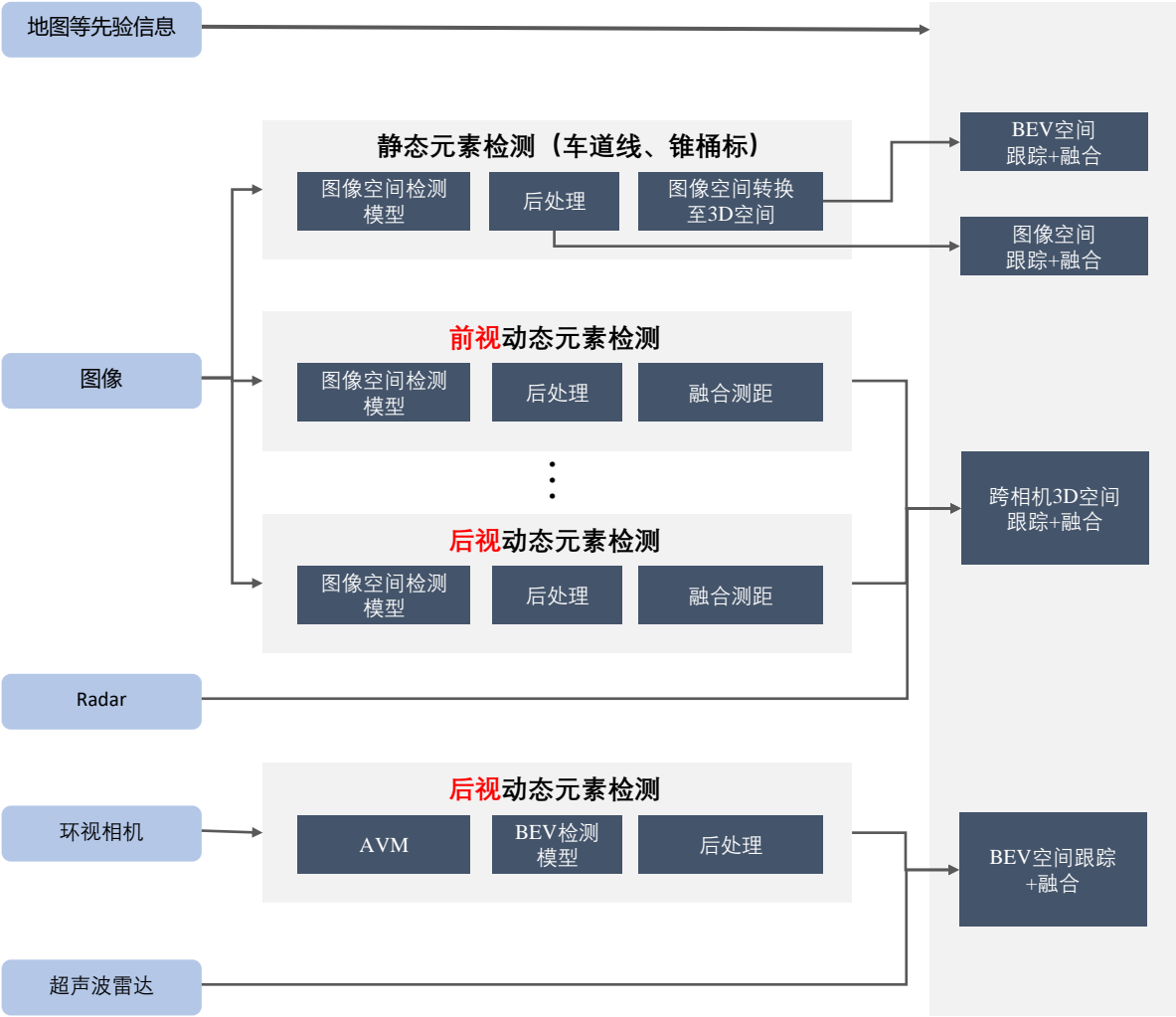


- 模型尺寸都很大，不便于部署

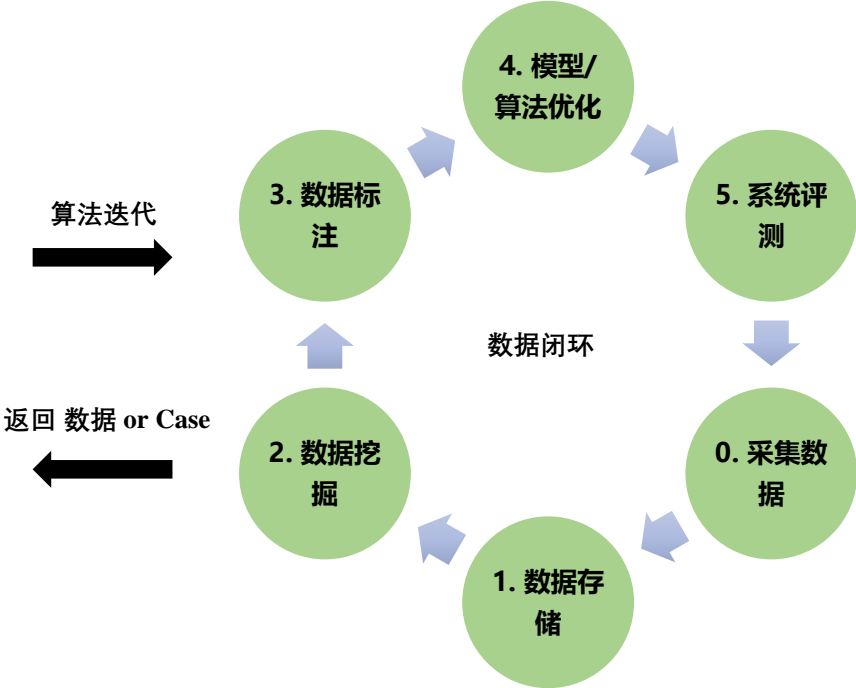


L2系统基于数据闭环来迭代的方式已成型，但其难以克服的问题

L2系统特点



迭代方式：数据闭环



问题：

- 系统迭代隔离：且融合是规则驱动，常见某种算法的提升，无法体现到感知输出层

结论1：L2感知算法的扩展性差，对环视感知的支持能力弱

- L2感知系统，pipeline长 + 规则 (rule-based)，支持四环视效果开发工作量大
- 跨视野的一致性会比较差，随着系统迭代，规则会更进一步增加，会让系统开发不可持续

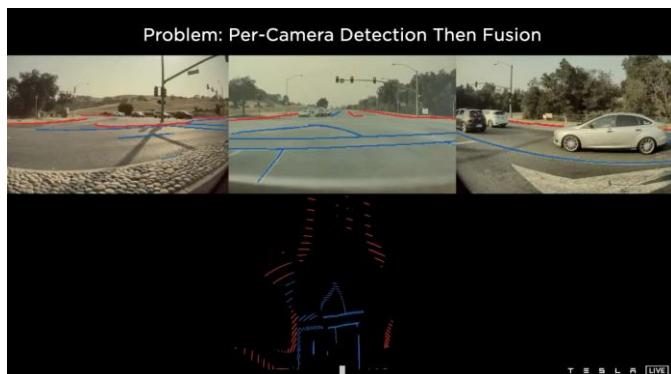
结论2：数据闭环可支撑思路环视，但由于算法设计问题，有放大系统泛化性问题的风险

- 新增四路环视，会让系统迭代更加隔离，融合迭代的难度增加

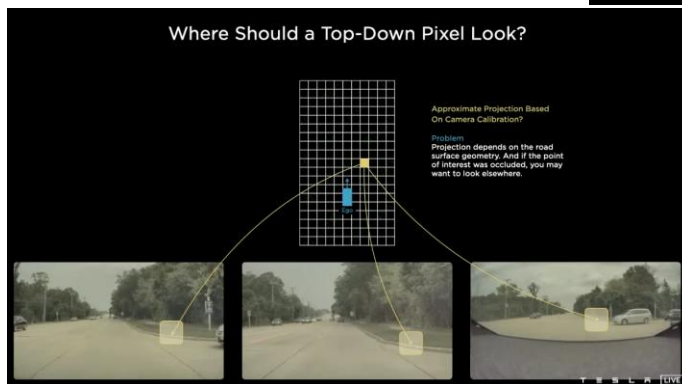
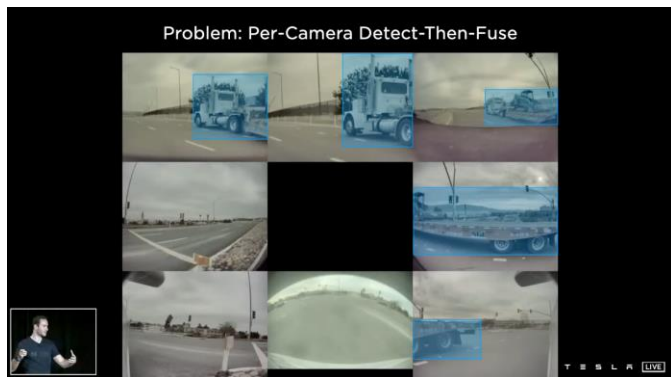


Tesla早于我们，遇到类似的问题@2021

- 跨相机道路结构检测



- 跨相机车辆检测



- 速度估计和遮挡检测问题



- 交通提示信息是提示未来的，如何在特征中保留



- 统一视角：似乎是更直观以及一致性更好的方式
- 时序叠加：在统一视角后下，进行时序叠加似乎可以保留完整时序信息

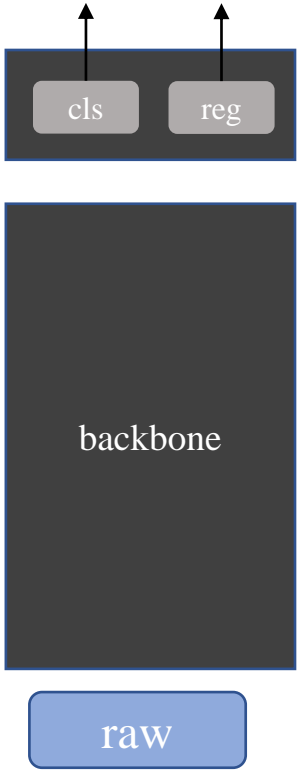


Tesla出现了，Tesla Vision的进阶之路

- Algorithm
- Product/Software/Version
- Labelling

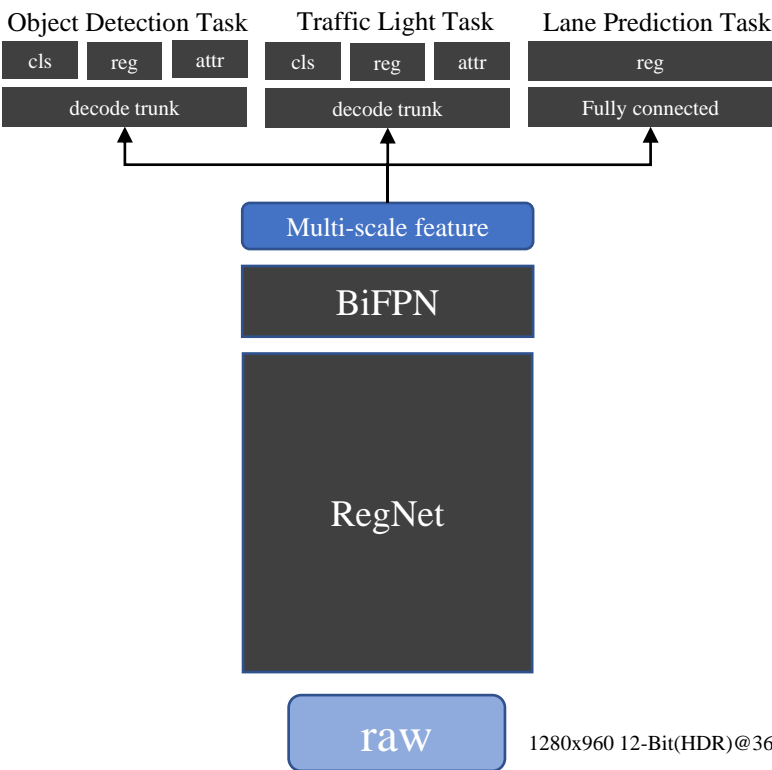
~2017

- General 2D detector
- Software 1.0
- Manual labelling
- Image space



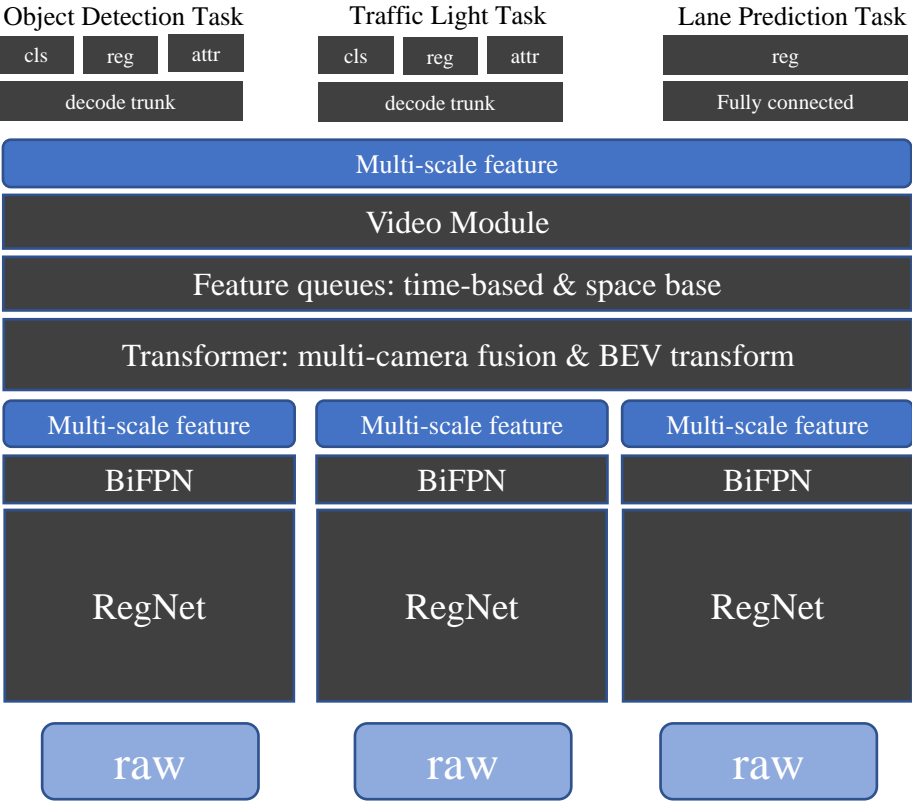
2017~2019

- HydraNet(multi-task)
- AutoPilot 4.0
- Manual labelling
- Vector space



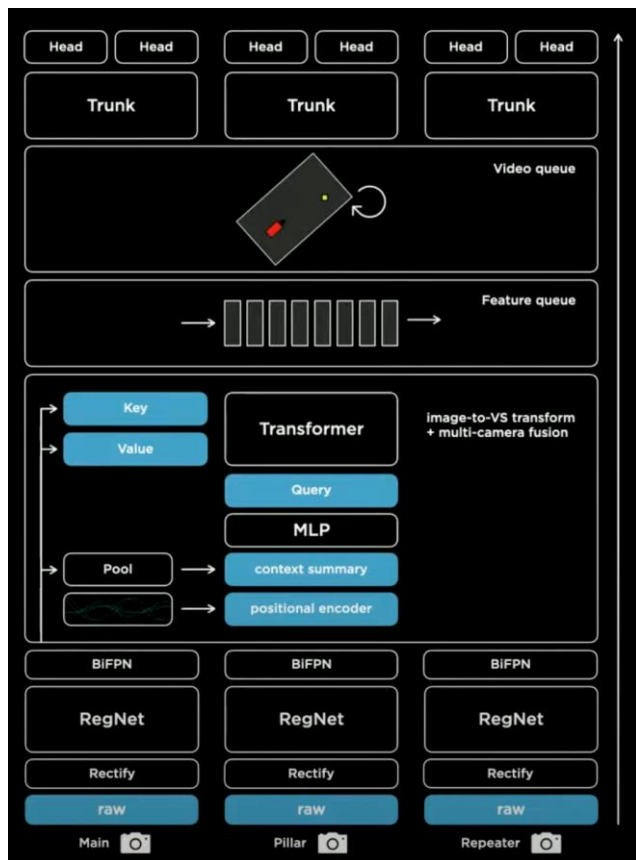
2019~2021

- HydraNet + spatial-temporal modules
- FSD beta 9.0/10.0
- Auto labelling
- Vector space



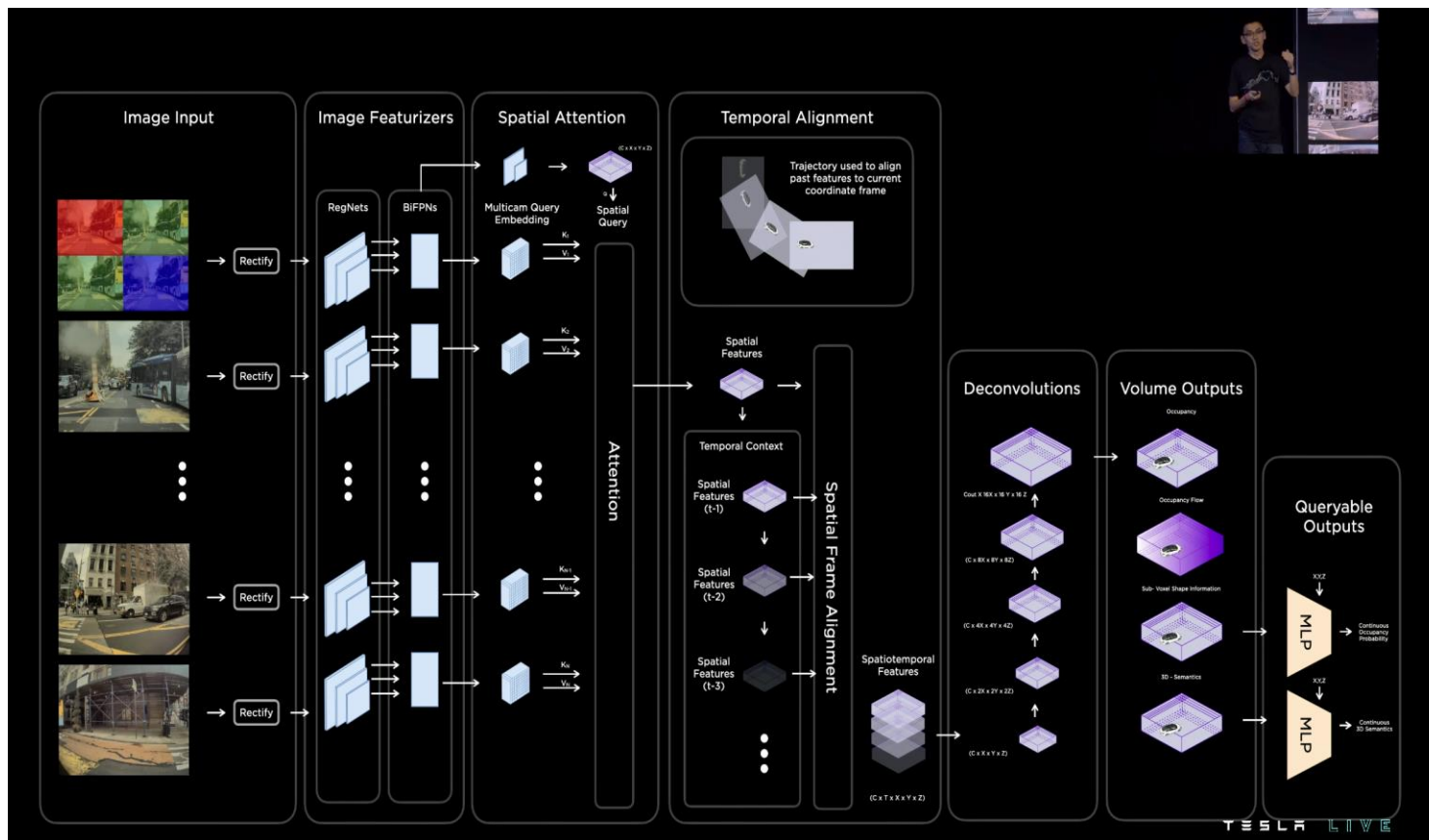
BEV概念开始出现，以及发展轨迹@Tesla AI Day 2021 & 2022

Tesla AI Day 2021



- **Transformer**: 将特征从图像域提升至3D Vector Space (类似于BEV)
- **Feature Queue**: 缓存空间和时序特征
- **Video Module**: 进行空间和时序特征融合

Tesla AI Day 2022



- **View Transformation**: 用Spatial Attention强化图像特征到Vector Space特征的转换
- **Temporal Fusion**: 直接融合连续帧的Vector Space特征
- **多任务扩展1**: 新增Occupancy 输出任务，以3D占据格形式（感知算力需求上升一个数量级）
- **多任务扩展2**: 进一步通过模型进行Planning

Tesla历经两年迭代的方案：View Transformation和Temporal Fusion，并在此基础上扩展新的任务

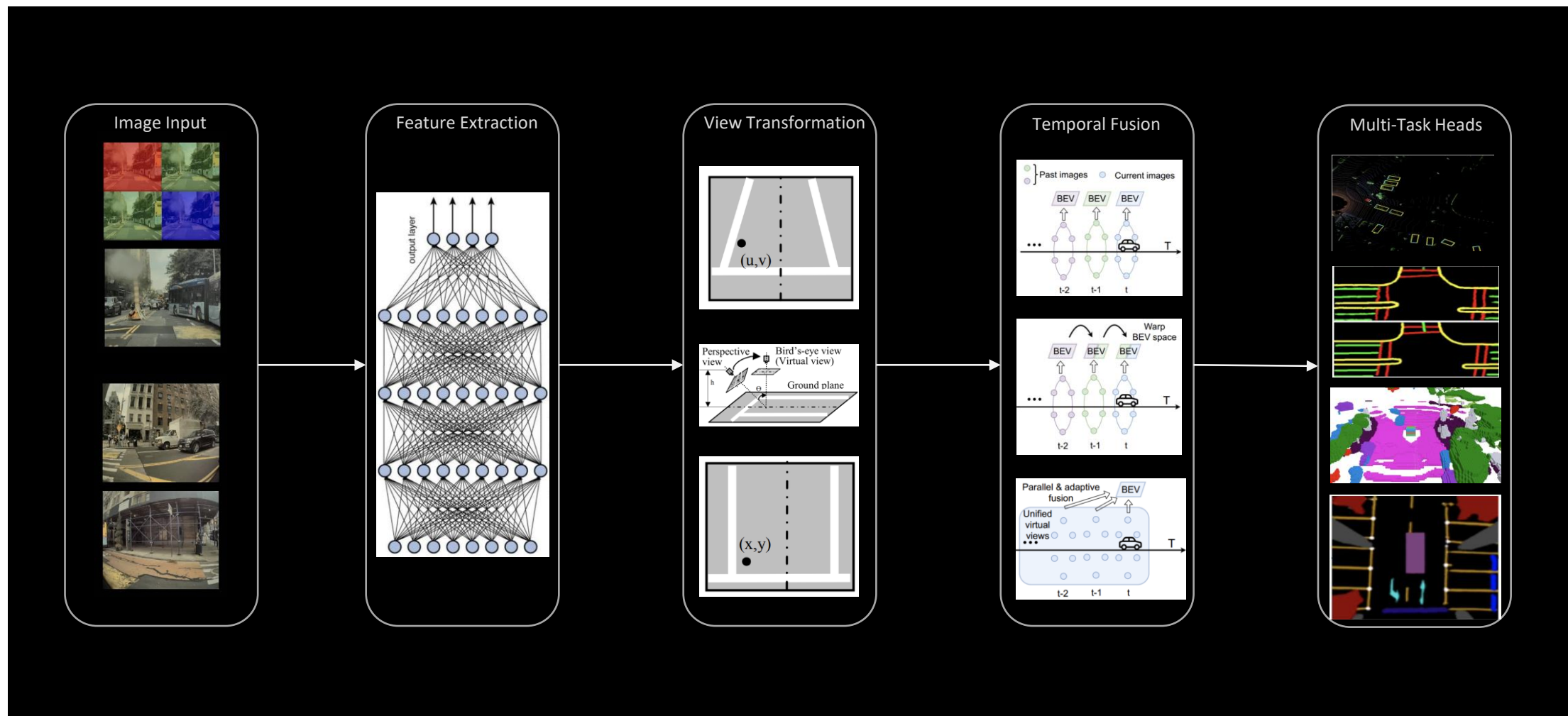
BEV概念开始出现，以及发展轨迹@Tesla AI Day 2021 & 2022

- ▣ 基于Vector Space方案，唯一一家不使用高精地图，仅依赖纯视觉，实现城区NOA功能的公司
- ▣ 强大的数据闭环能力，支撑其极快的迭代速度



FSD v10.12 @202212效果

Tesla眼中，一个好的L2+（甚至是FSD）感知系统，是什么样呢？



同时，系统依然可以通过支持 Vector Space的数据闭环驱动迭代



第二章：从 Multi-View 到 BEV，构建新的感知模型

为什么是BEV（Bird Eye View）？

View Transformation方法

- IPM：最传统的方法
- LSS（2D to 3D/从底向上）：基于深度分布估计的方法
- Transformer（3D to 2D/从顶向下）：基于注意力机制的方法

BEV动静态元素检测任务

- 基于LSS的动静态元素检测任务
- 基于Transformer的动静态元素检测任务