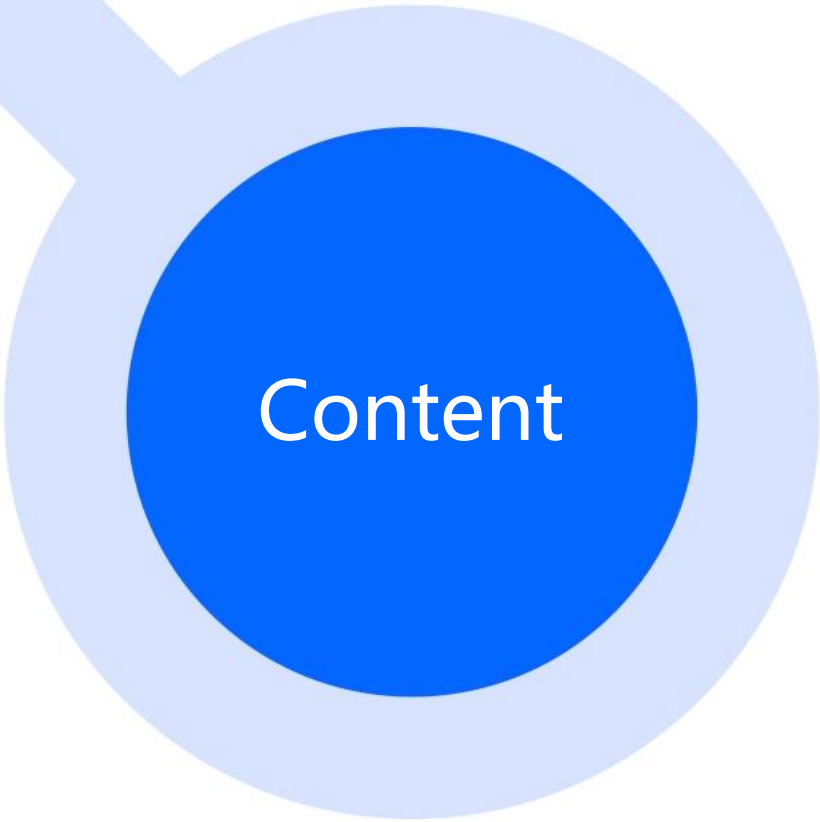


自动驾驶感知新范式

张家馨



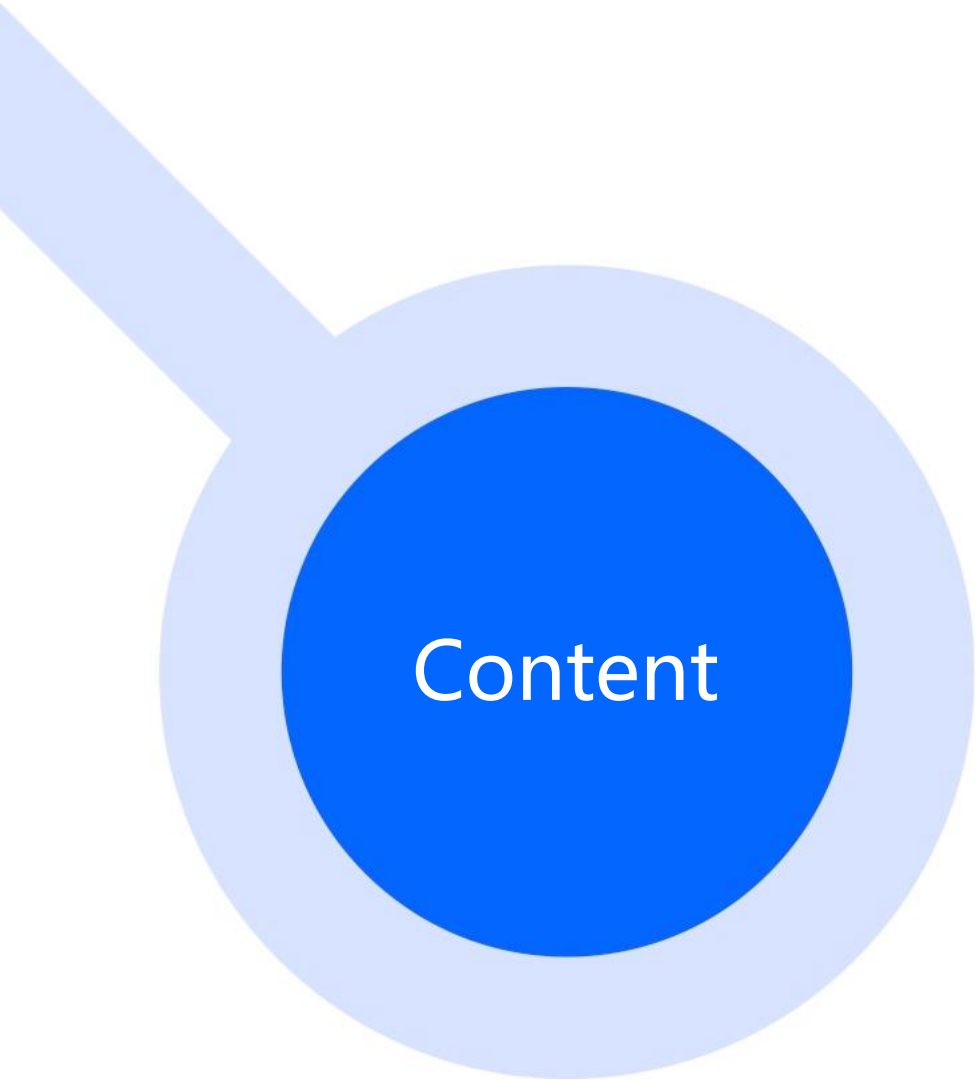
Content

上半节

- 01. 自动驾驶感知技术发展
- 02. BEV感知框架
- 03. BEV感知应用

下半节

- 04. 云端自动标注
- 05. 仿真系统
- 06. 端到端自动驾驶



Content

04. 云端自动标注

- (a) 激光雷达自动标注
- (b) 纯视觉自动标注
- (c) 系统工程

04. 云端自动标注

- 4D 主要包含3D 空间+ 时序信息
- 以BEV为代表的感知任务输出空间从2D 透视图像空间转换到3D空间，相应地标注从空间也转换到3D空间
- BEV感知输入时序video数据，输出时序信息，例如tracking、prediction、velocity、accelerated velocity等



图片来自于Tesla AI Day 2021

04. 云端自动标注

图商的HDMap

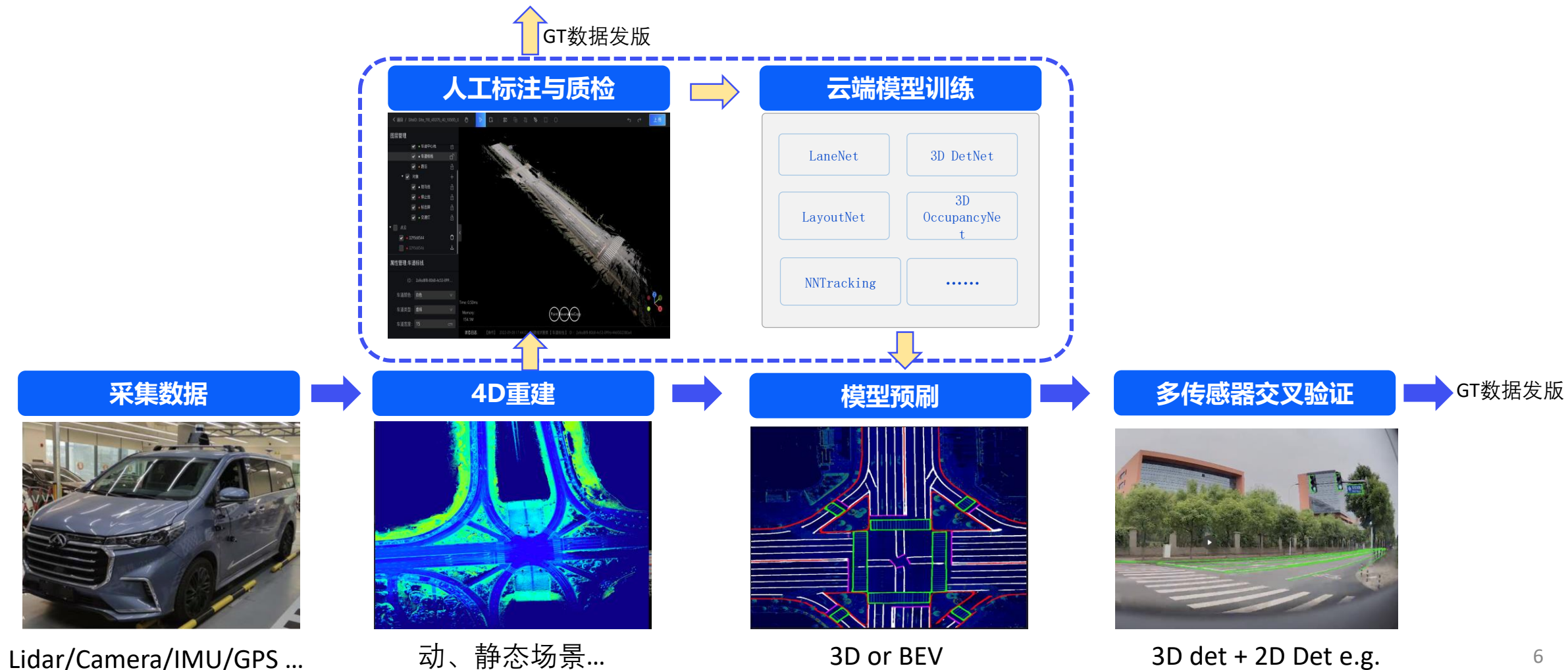
- 高精度传感器，成本高，可扩展性差
- 受地图鲜度影响，地图更新慢
- 定位精度差，横向误差10cm，纵向误差 50cm
- 完整度高

4D Label

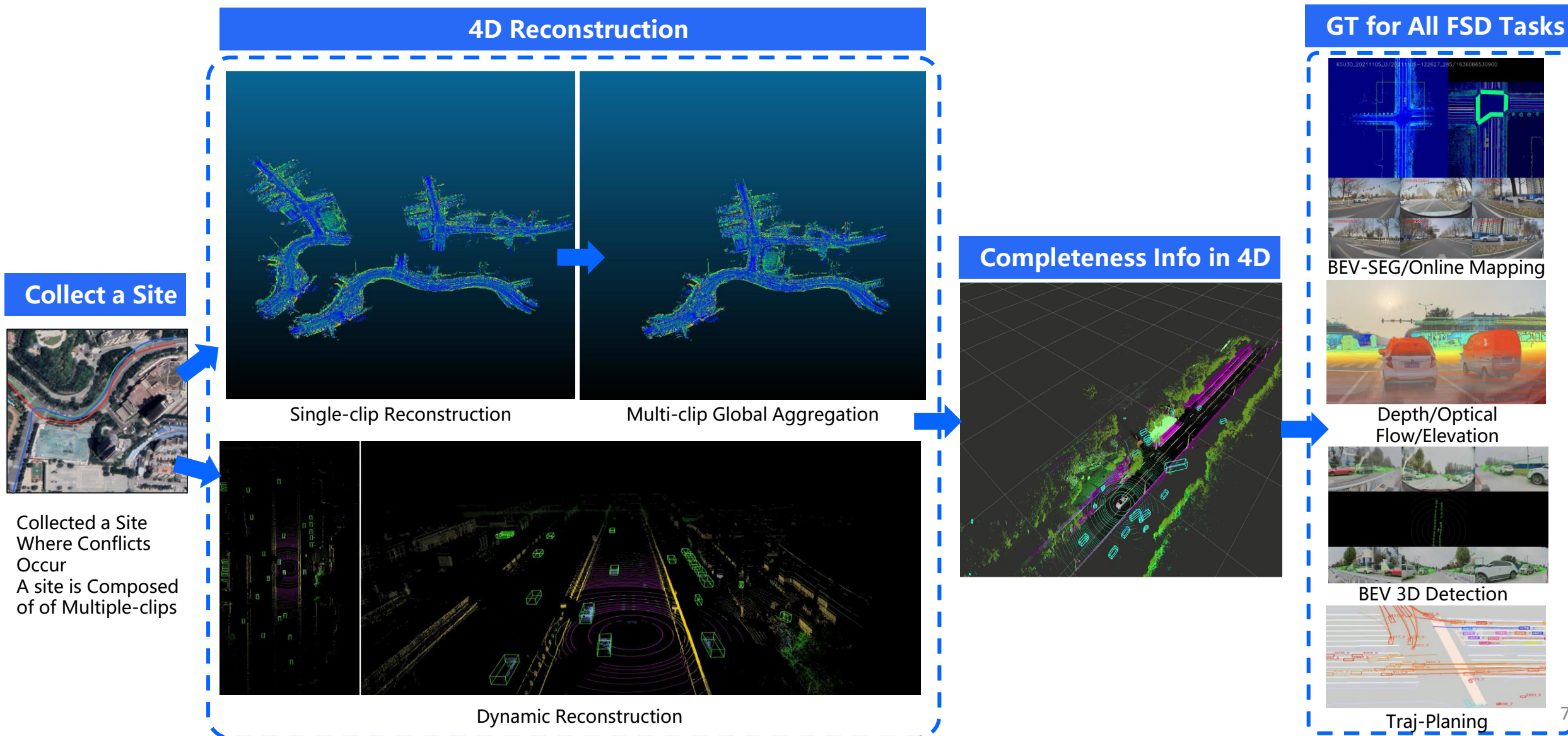
- 一般精度传感器，适用于采集车
- 即建即用，不存在地图鲜度的问题
- 定位精度高，横纵向<10cm
- 局部的建图，可快速覆盖

04. 云端自动标注-整体方案

- 通过4D重建实现点云级别或者object级别的重建，通过人工标注积累原始数据
- 随着数据积累到一定程度，可以训练云端大模型逐步替换人工标注，可提升80%+的标注效率



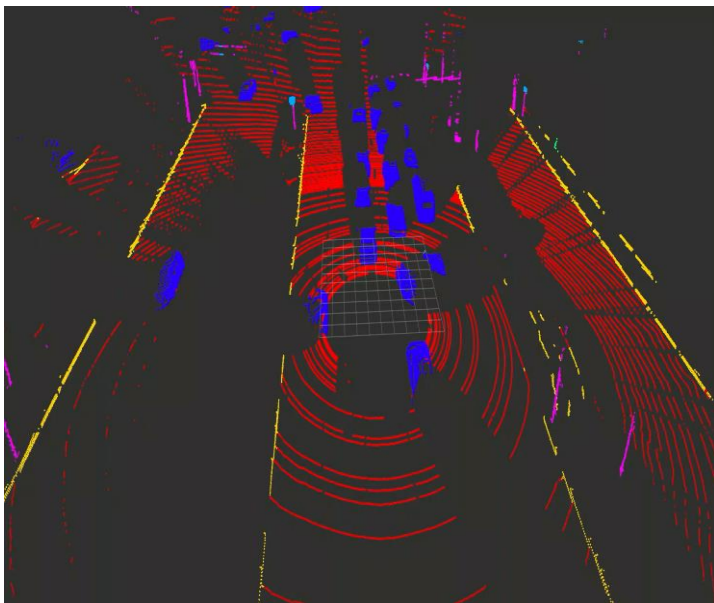
04. 云端自动标注-整体方案



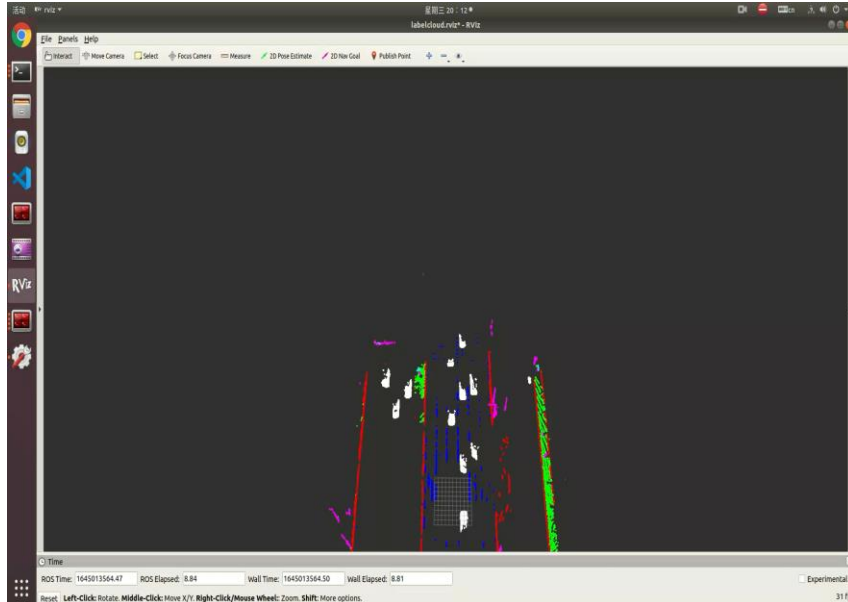
04. 云端自动标注-局部建图

- Lidar 定位建图依然面临着挑战，例如车辆高速行驶导致初始化困难、回环检测失败率高等
- 采用多传感器融合结合Lidar-seg语义信息提升单次建图的精度和鲁棒性、实现cm级别的建图精度
- 采用Learning based features用于提升全局聚合的成功率到90%以上

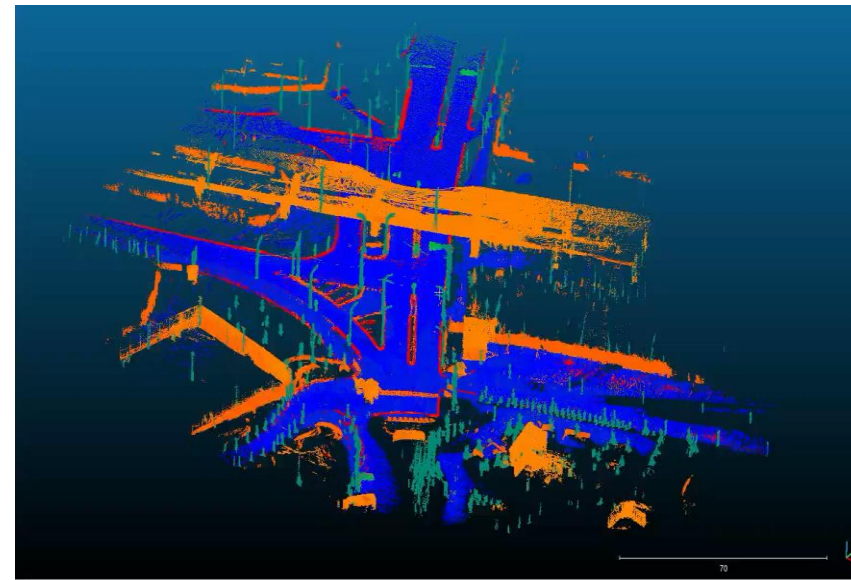
单帧Lidar-seg的效果



单次重建的效果

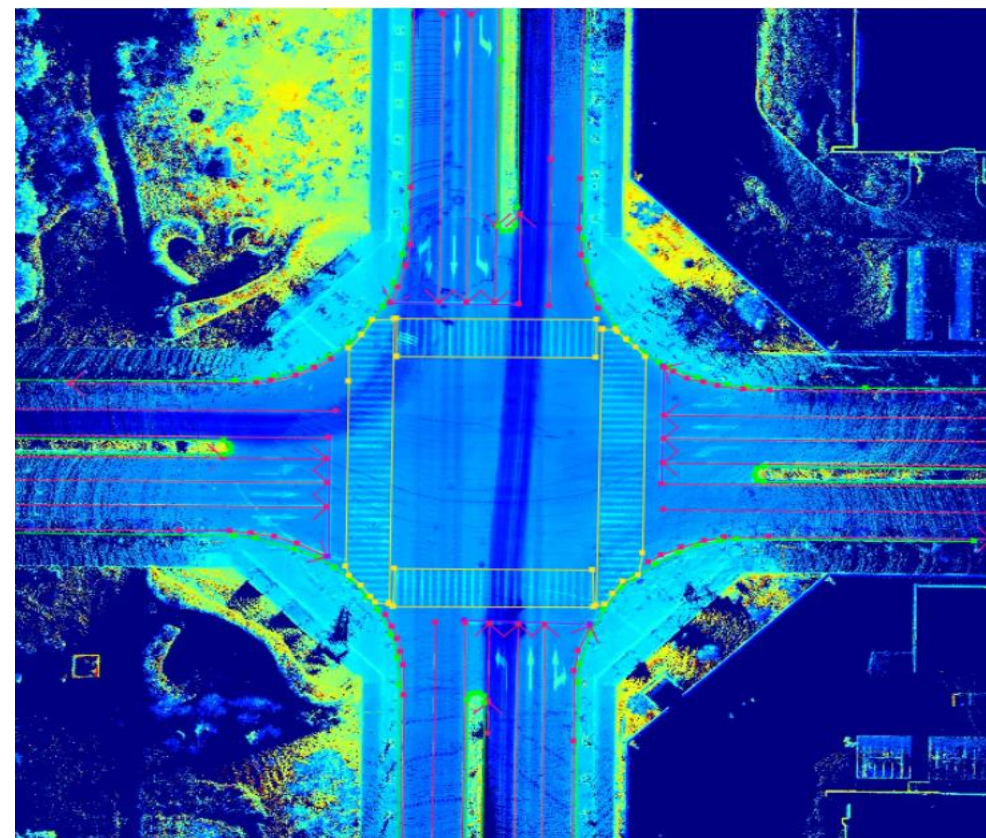
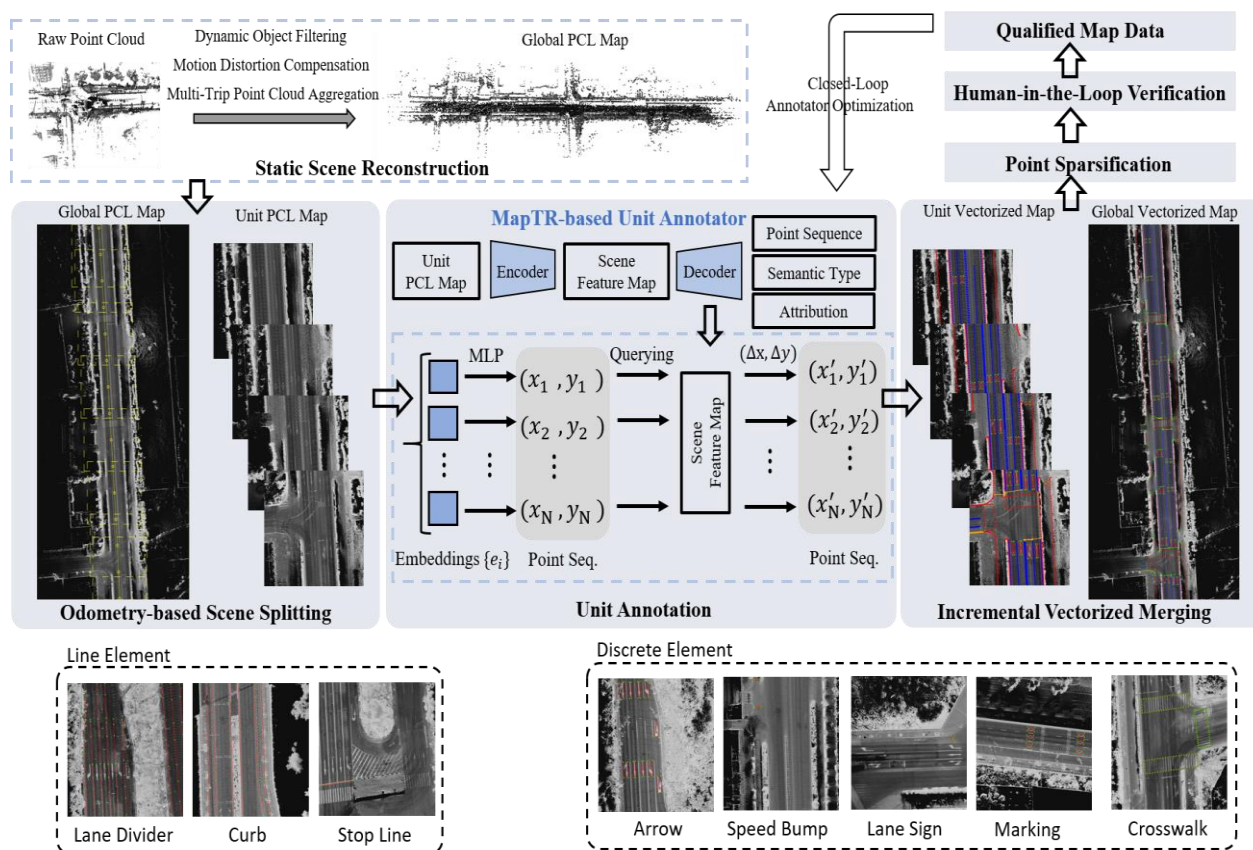


全局聚合的效果



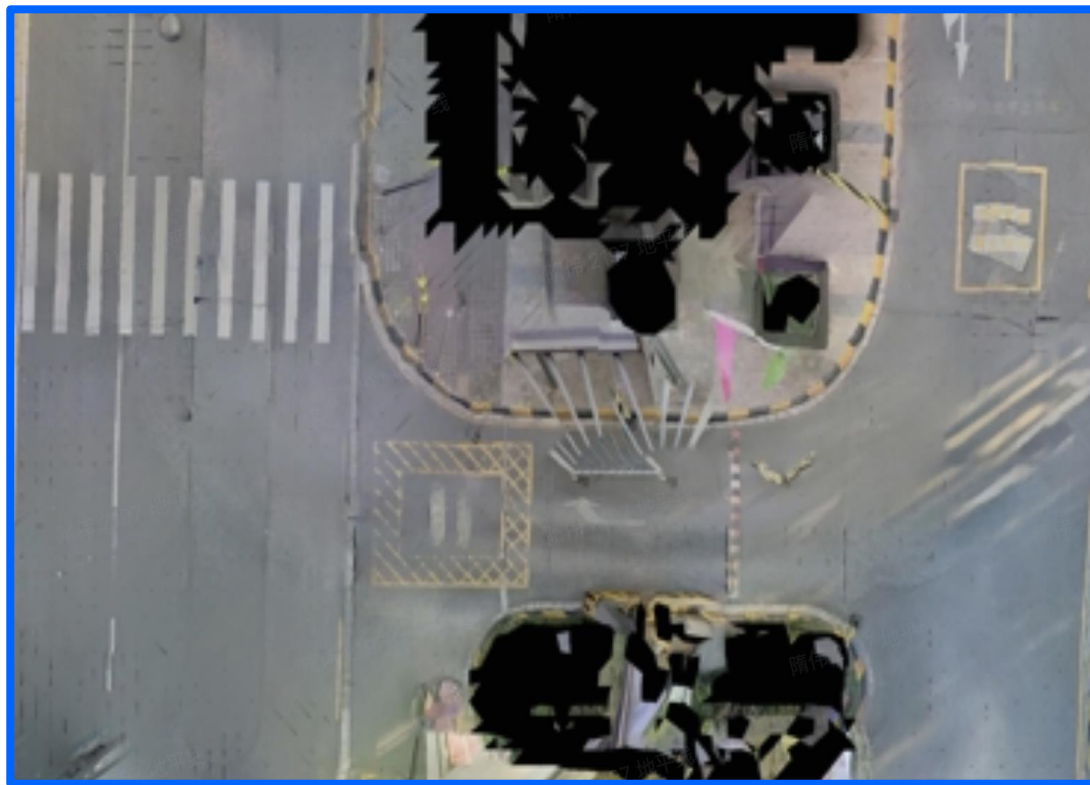
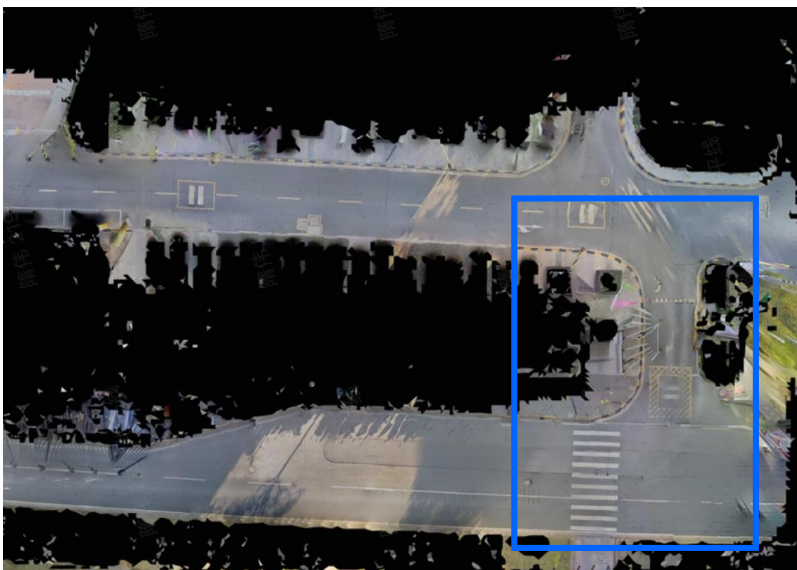
04. 云端自动标注-矢量元素提取

- 通过云端模型实现高度自动化标注，可节省**80%以上**的人工标注成本
- 模型支持离散要素、几何、属性、连接关系的标注



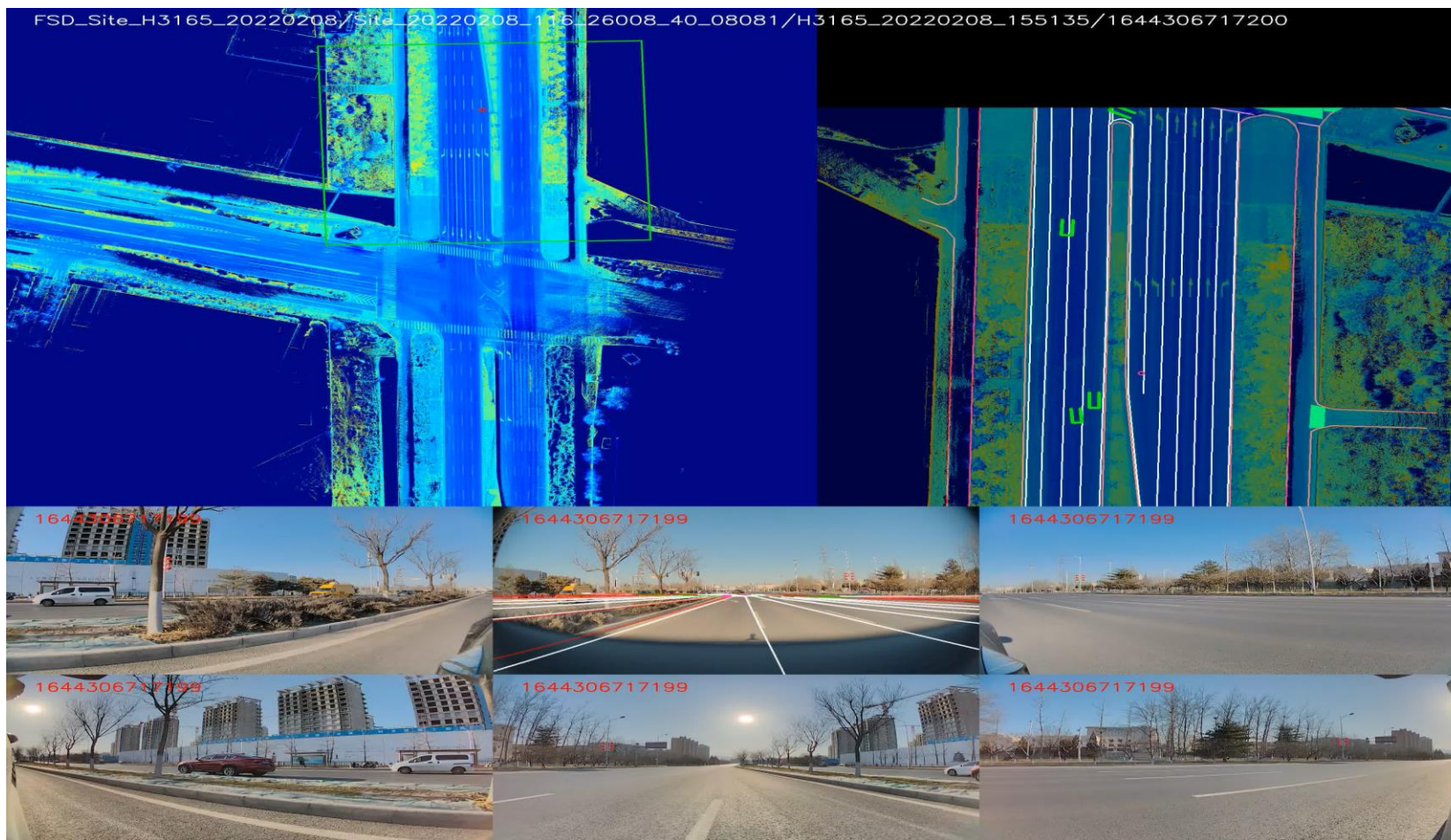
04. 云端自动标注-矢量元素提取

- 通过云端模型实现高度自动化标注，可节省**80%以上**的人工标注成本
- 模型支持离散要素、几何、属性、连接关系的标注



04. 云端自动标注-重投影一致性

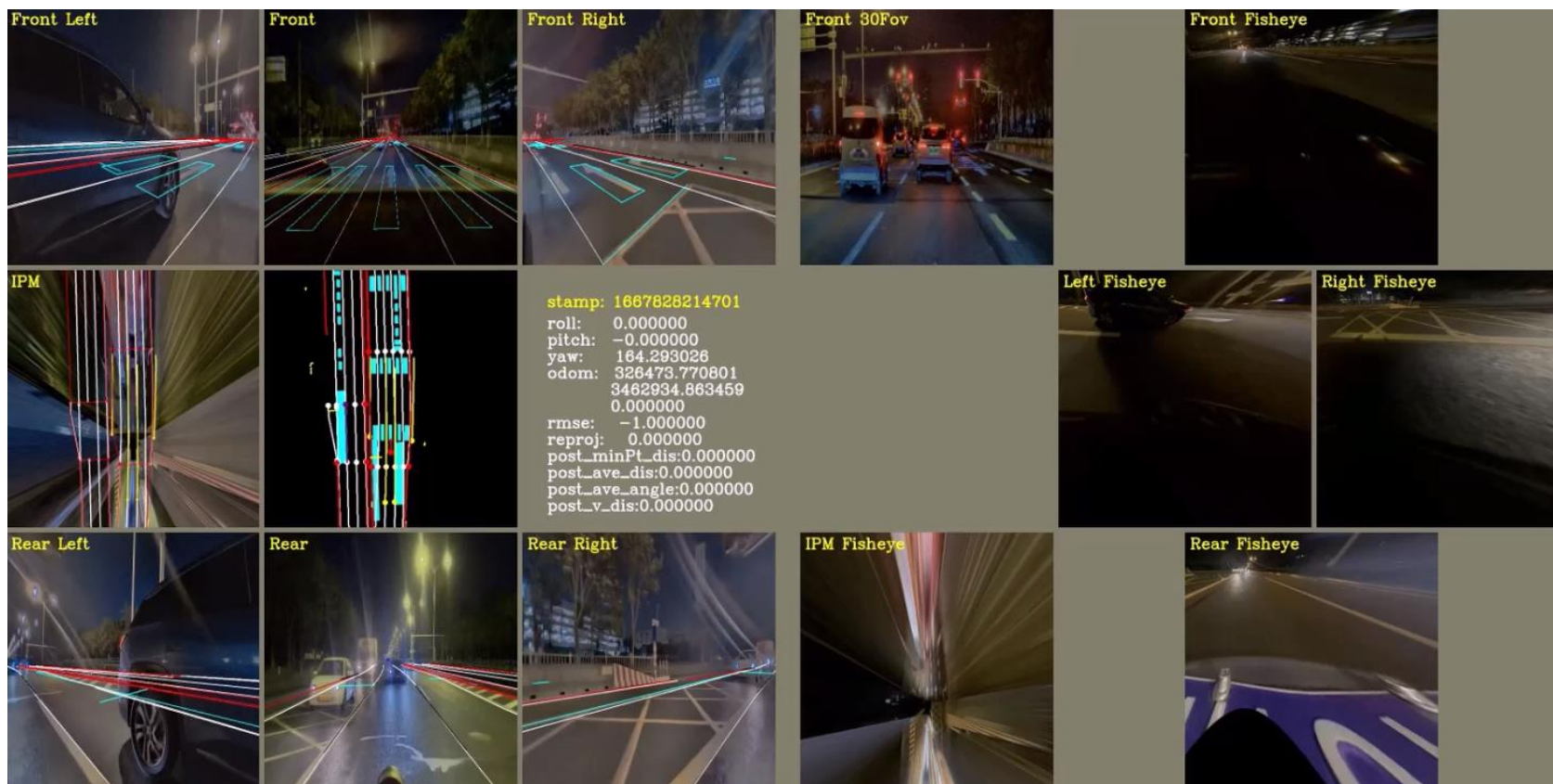
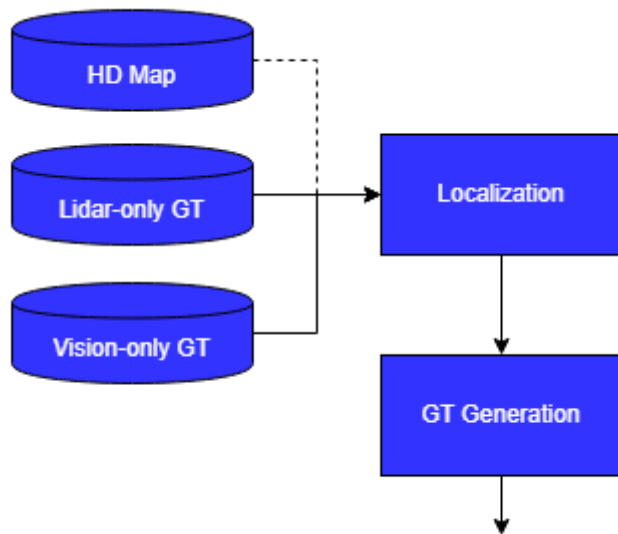
- 通过多个clip的聚合，不但可以保证局部建图的完整性，以满足远距离感知的需求，同时可以通过共享标注的方式减少标注成本，平均重投影误差 <4 px



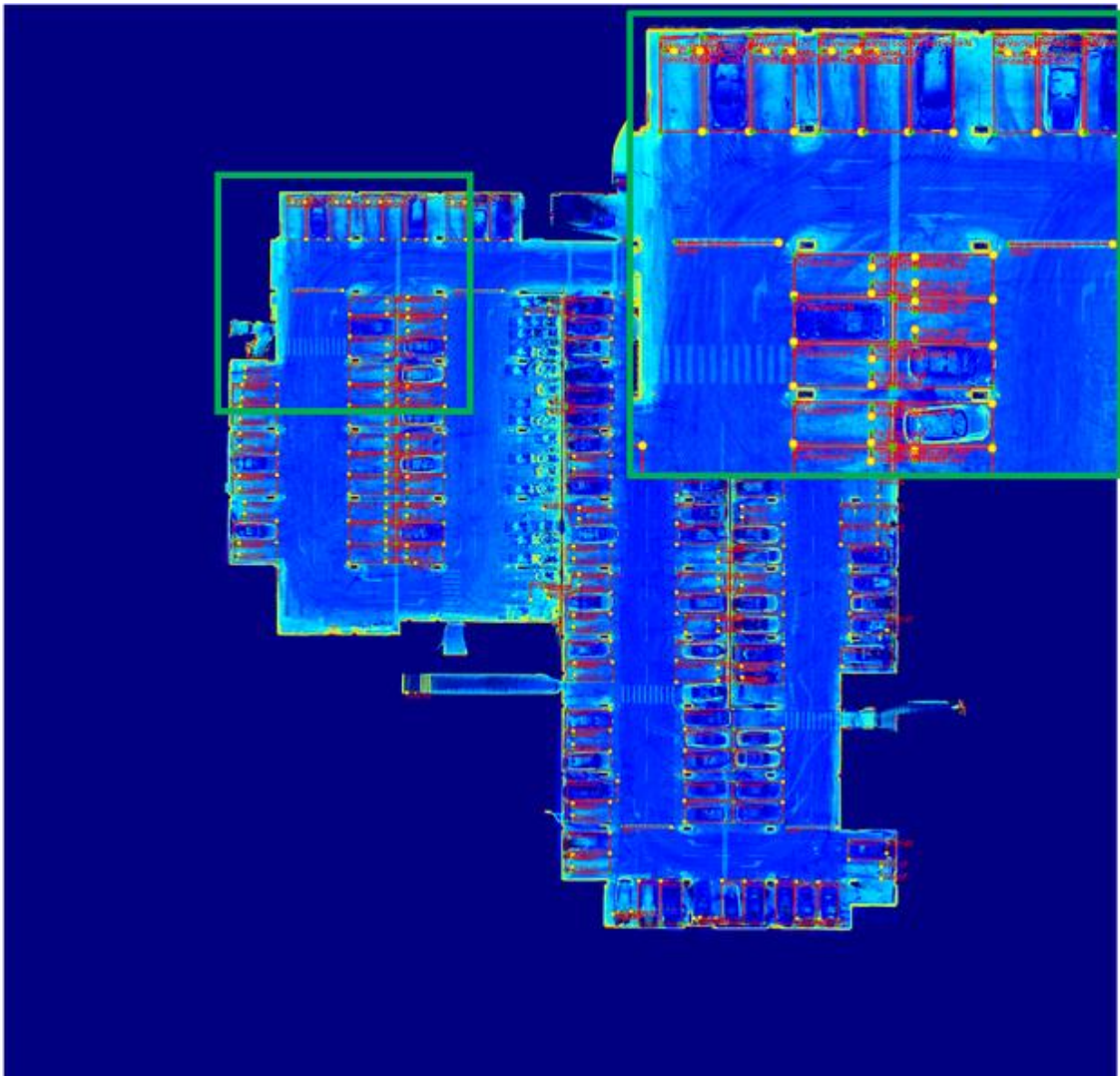
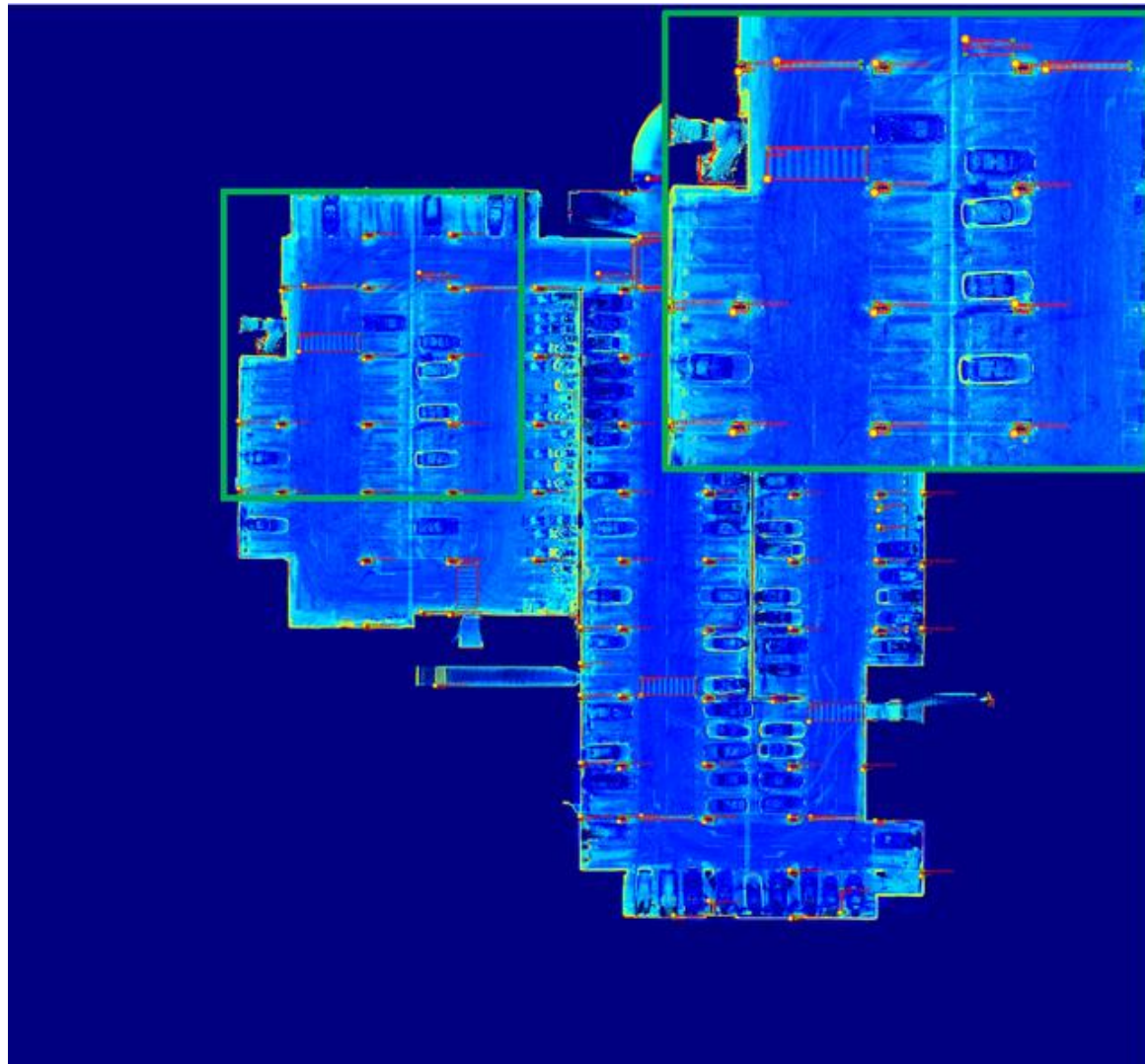
3段clip聚合的效果，包含静态路面要素和动态标注

04. 云端自动标注-特殊场景重投影

- 对于已经生成的真值，采用重定位技术低成本快速实现跨车、跨传感器之间的真值复用
- 能够有效地处理极端情况下的真值生成问题，平均重投影误差小于4 px

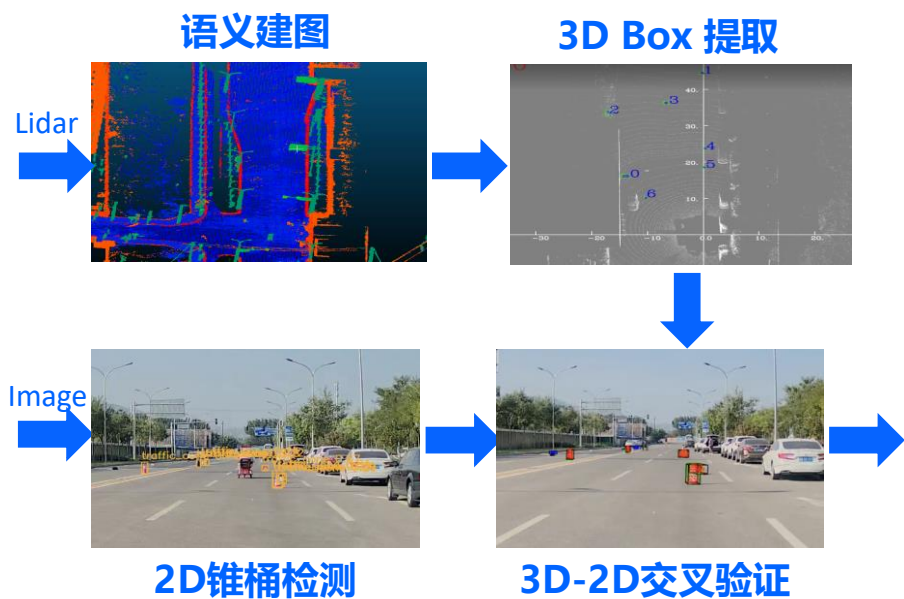


04. 云端自动标注-泊车场景



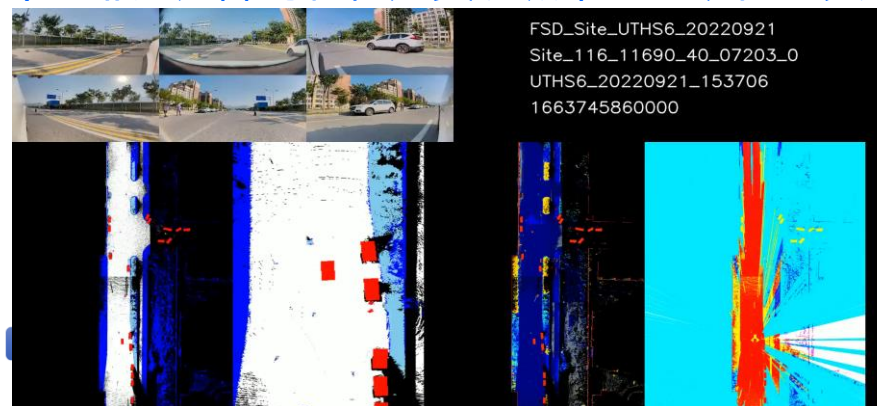
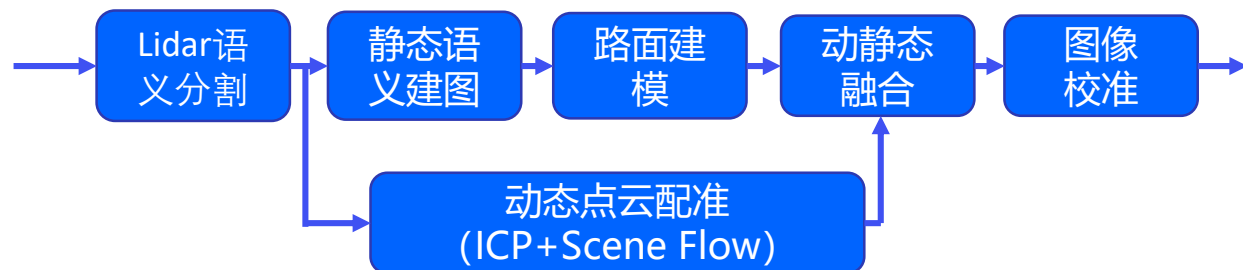
04. 云端自动标注-锥桶3D自动标注

- 采用Lidar+camera后融合的方式实现锥桶的3D真值全自动标注

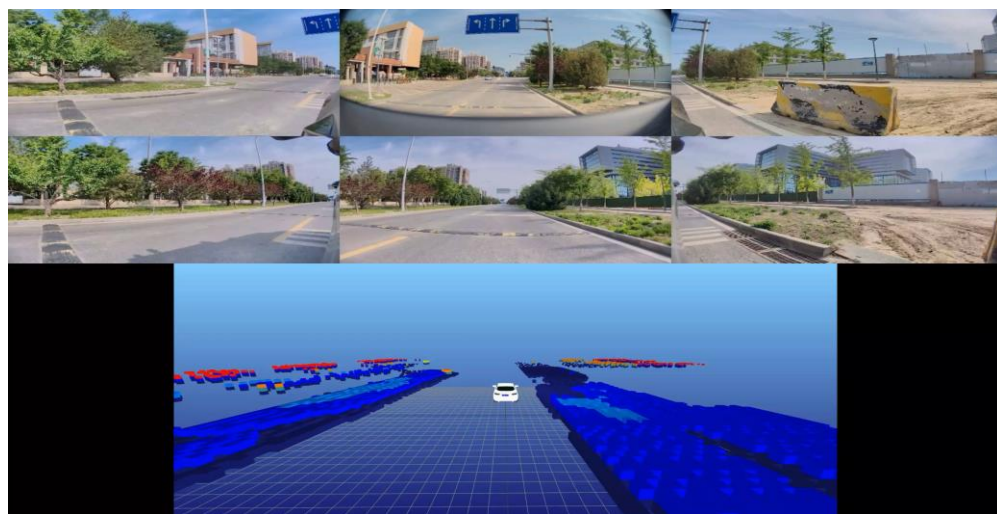


04. 云端自动标注-Occupancy

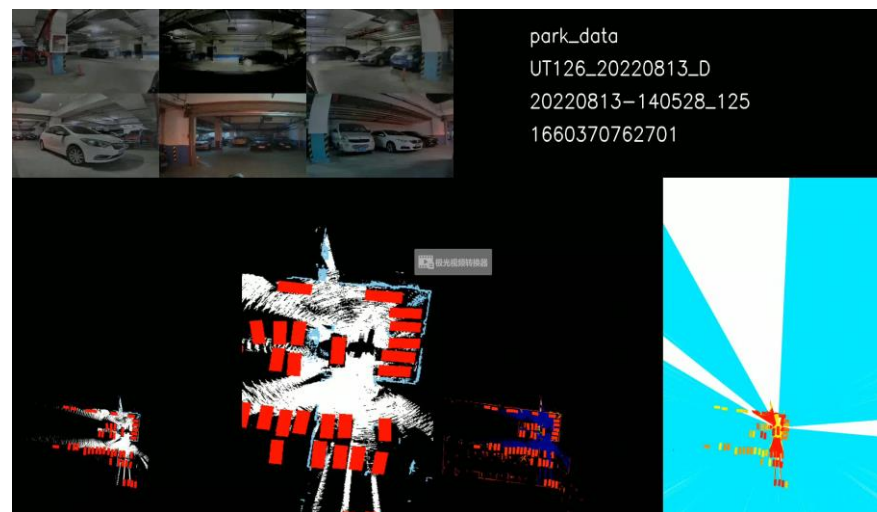
- Occupancy反映了场景中动静态物体的占据情况，用于识别场景中的非通用障碍物，是实现城市NoA功能的关键



行车场景 2D Occupancy 真值



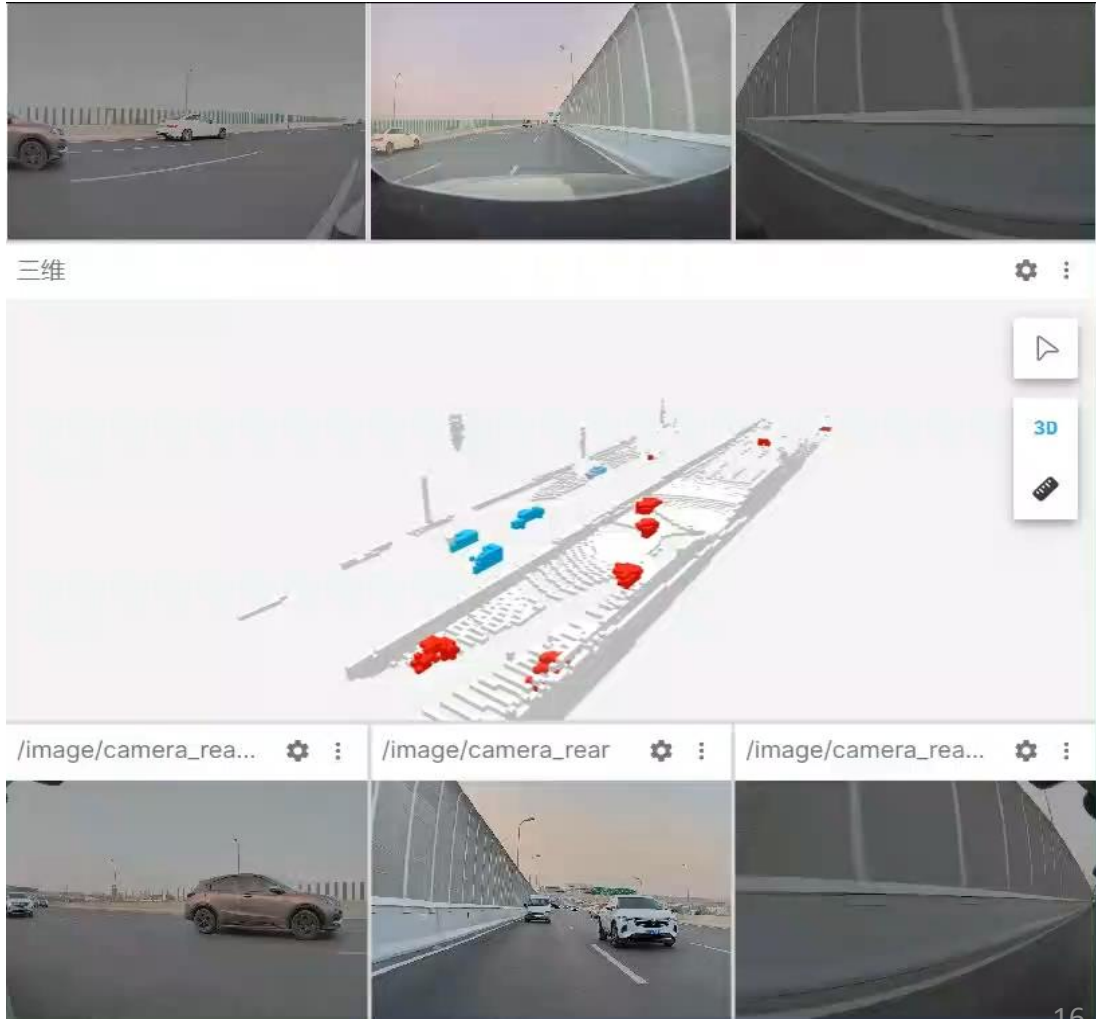
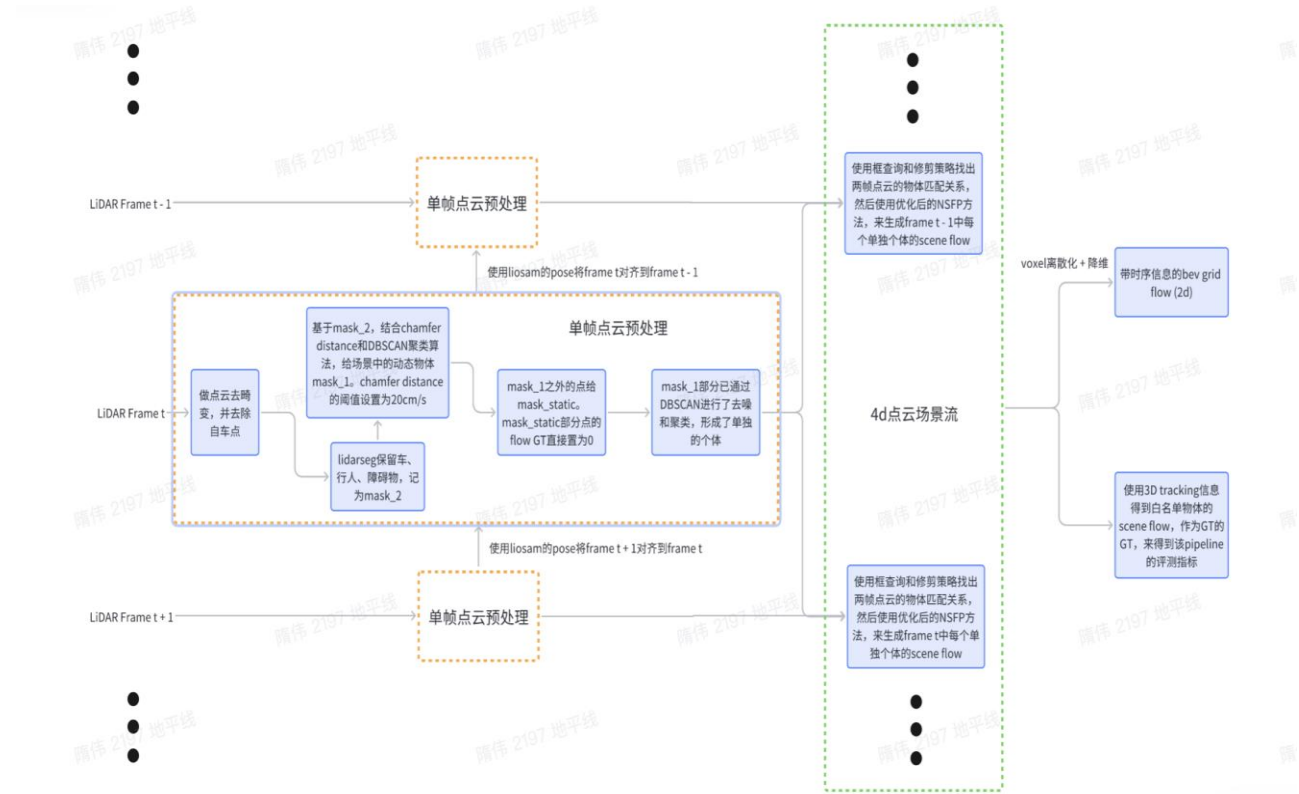
行车场景3D occupancy 真值



地库场景2D Occupancy 真值

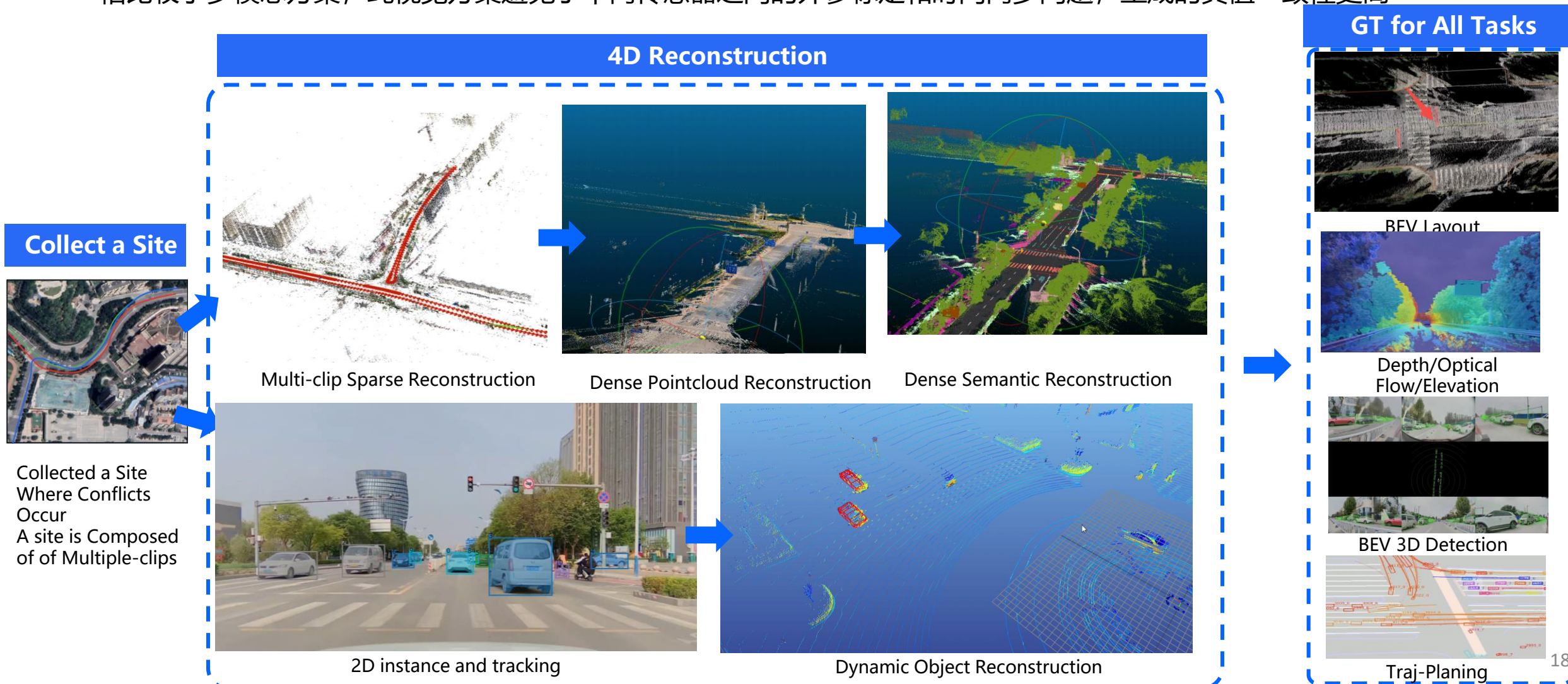
04. 云端自动标注-Occupancy Flow

■ Occupancy flow 反映了场景中动静态物体的运动情况，对通用障碍物的占据情况进行预测



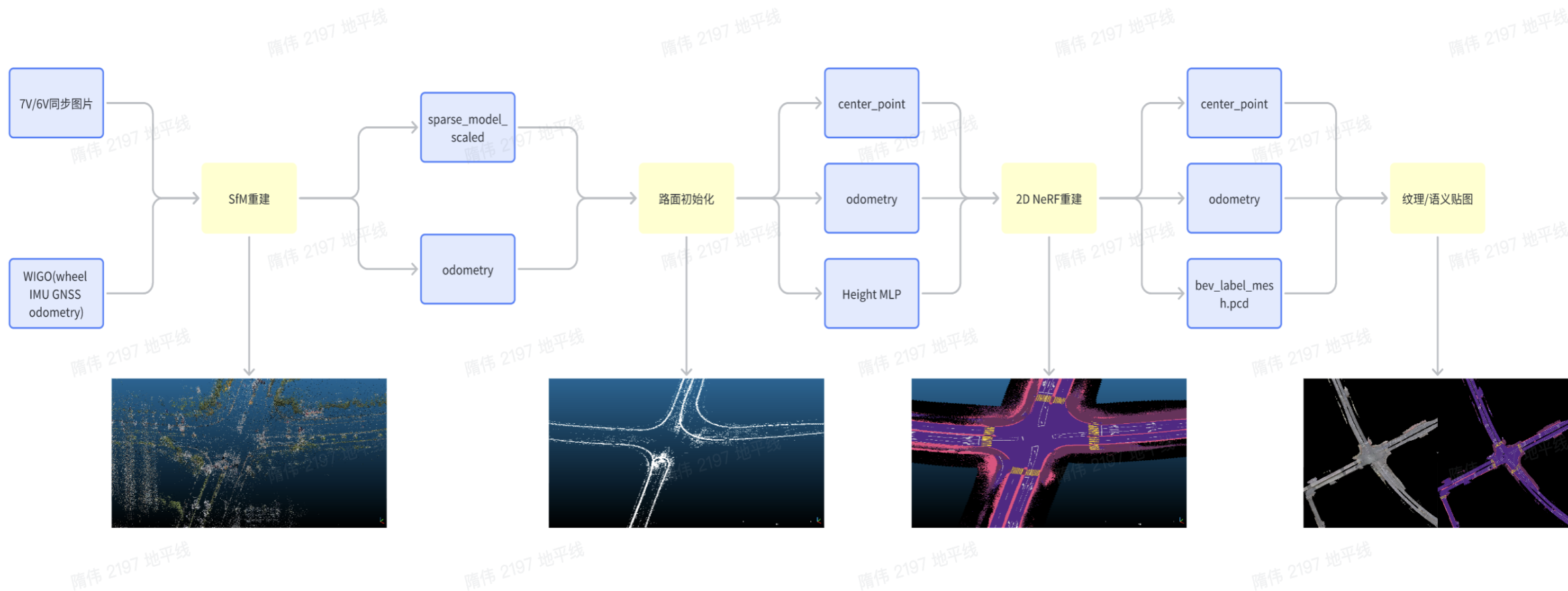
04. 云端自动标注-纯视觉方案

- 纯视觉方案在4d重建尤其是动态物体的重建非常具有挑战性
- 相比较于多模态方案，纯视觉方案避免了不同传感器之间的外参标定和时间同步问题，生成的真值一致性更高



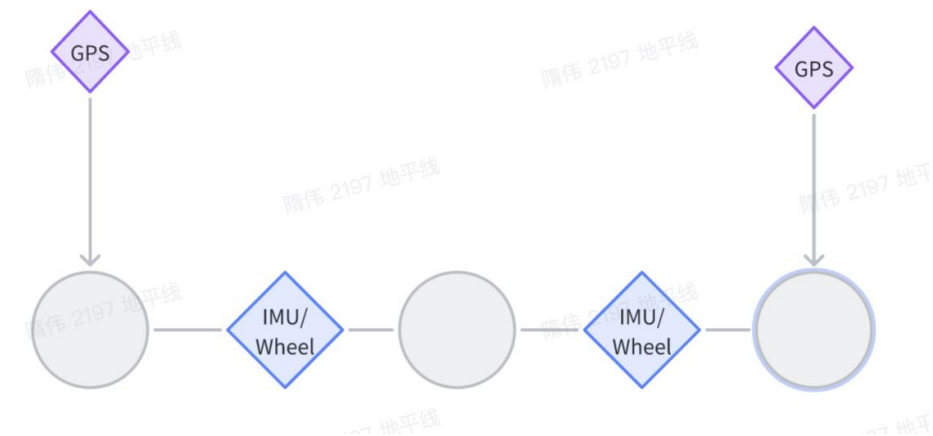
04. 云端自动标注-纯视觉路面重建

■ 路面重建的整体技术路线



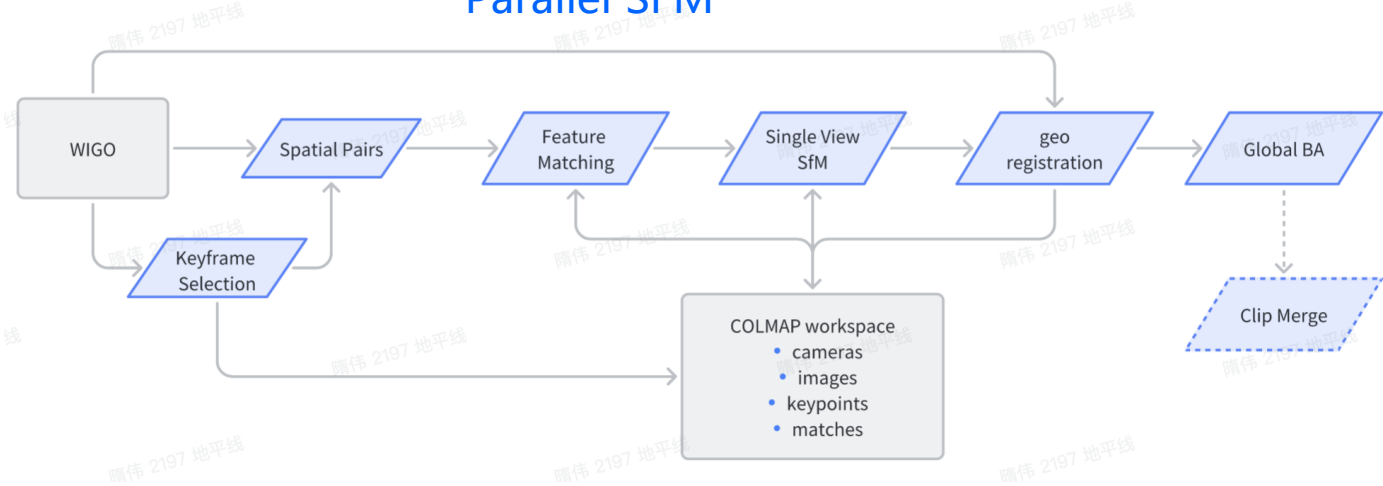
04. 云端自动标注-纯视觉路面重建

WIGO初始化



RTE	2.5 %
RRE	0.8 deg/100m

Parallel SfM



RTE	0.5 %
RRE	0.2deg/100m

04. 云端自动标注-纯视觉稠密重建

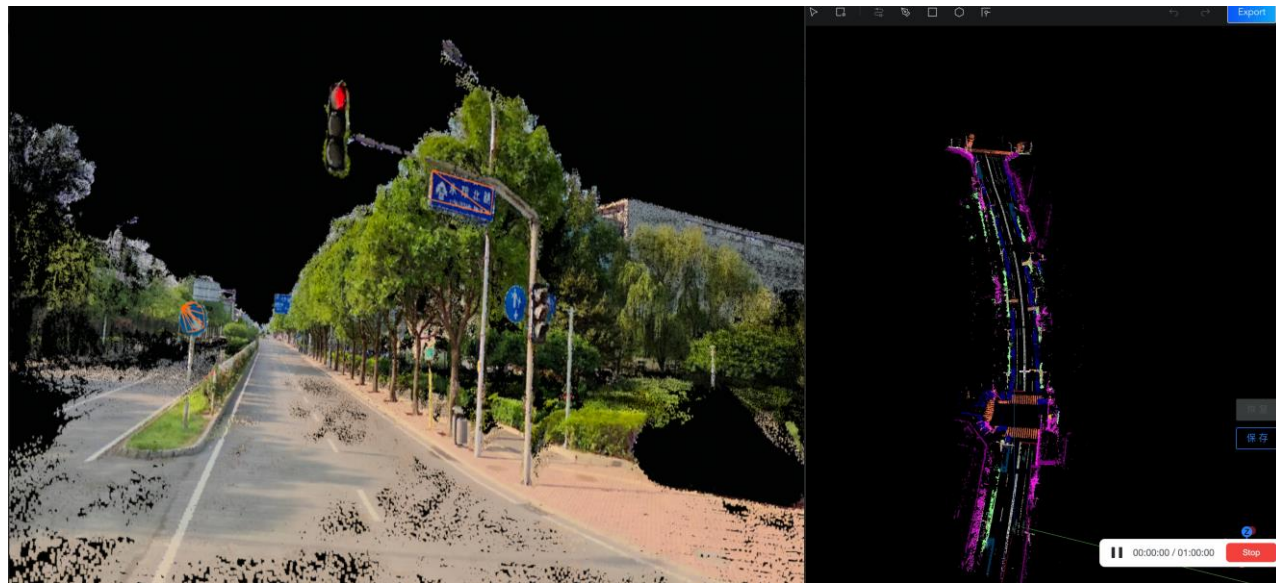
- 对传统的增量incremental SFM进行优化，采用GNN+Wheel+IMU优化后的姿态作为初始值，直接进行Multiple-camera及时序上的三角化，可以提升接近10倍的计算效率， $RTE < 0.5\%$
- 采用Learning based features (Superpoint/Superglue) 提升再弱纹理区域的重建鲁棒性
- 采用语义引导的MVS提升了点云的重建精度

Multi-Cam Structure From Motion



重建精度 $RTE < 0.5\%$ (对比Tesla 1.5%)

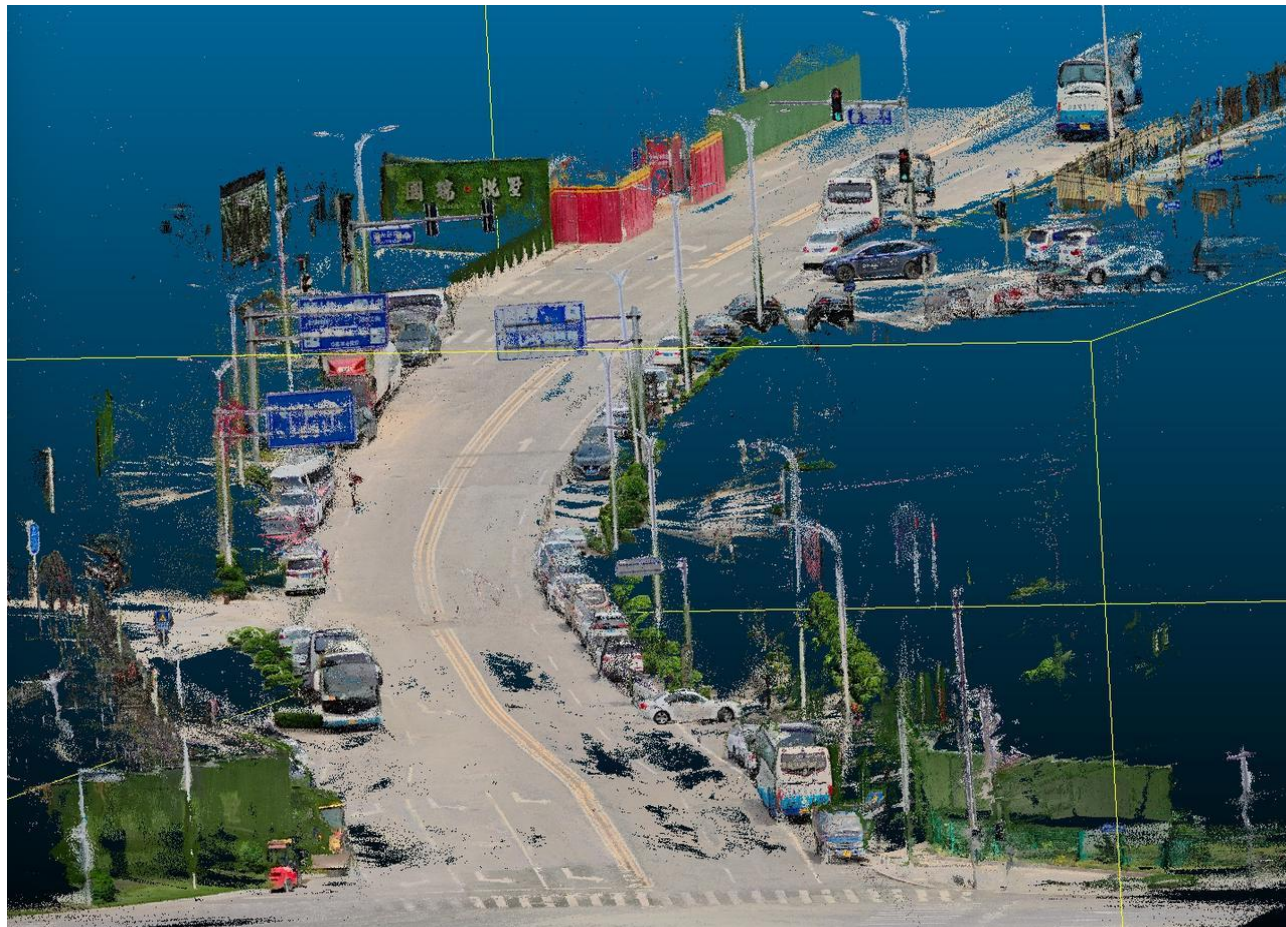
Semantic Info Guided Multi-view Stereo



纯视觉静态点云重建，精度 $abs.rel < 5\%$

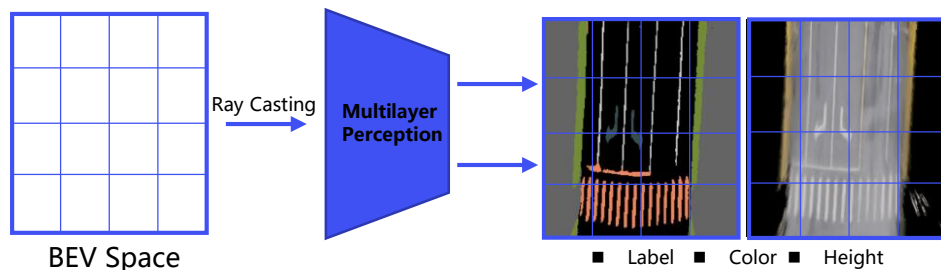
04. 云端自动标注-纯视觉路面重建

- 视觉稠密点云重建计算量太大，成本太高

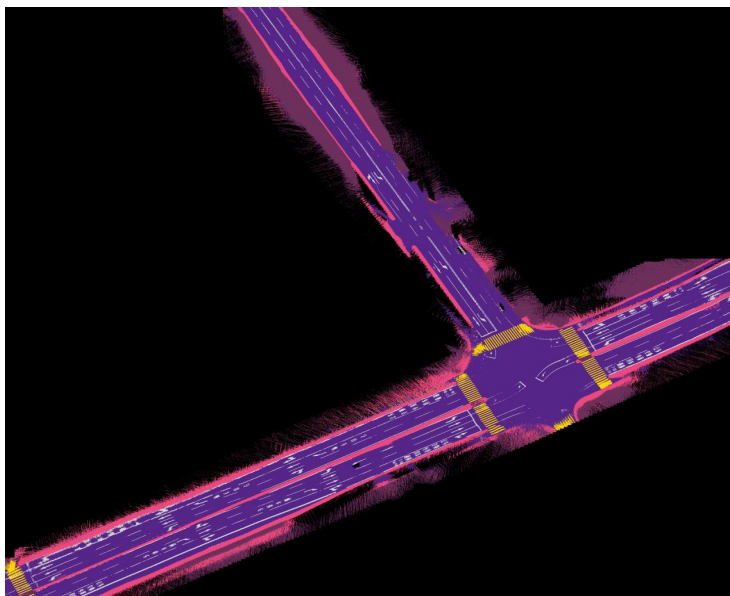


04. 云端自动标注-基于NeRF路面重建

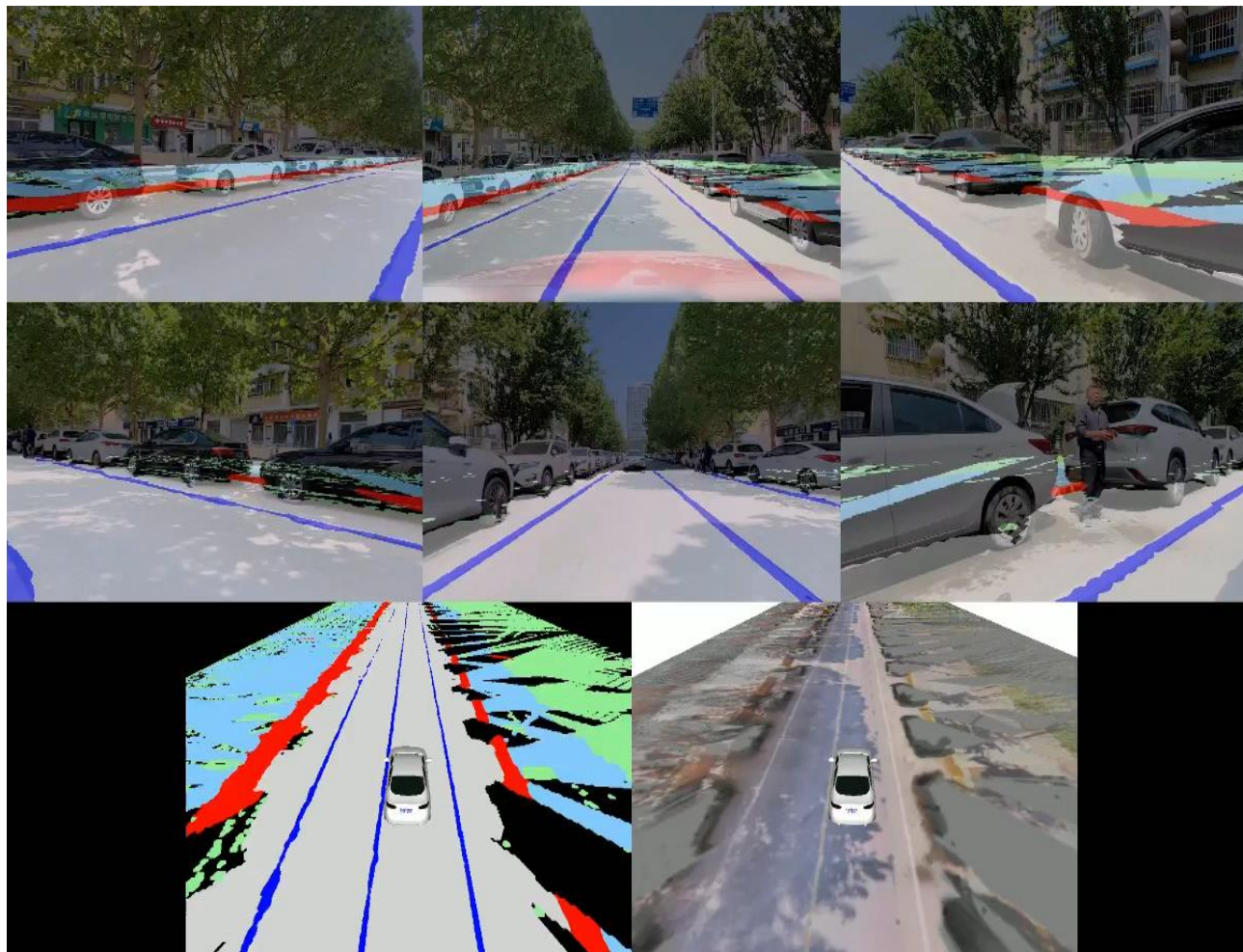
- 采用Nerf对路面进行重建，同时获取路面高度和纹理信息，利用语义信息进行监督，可以有效抵制弱纹理的影响



RoMe Framework

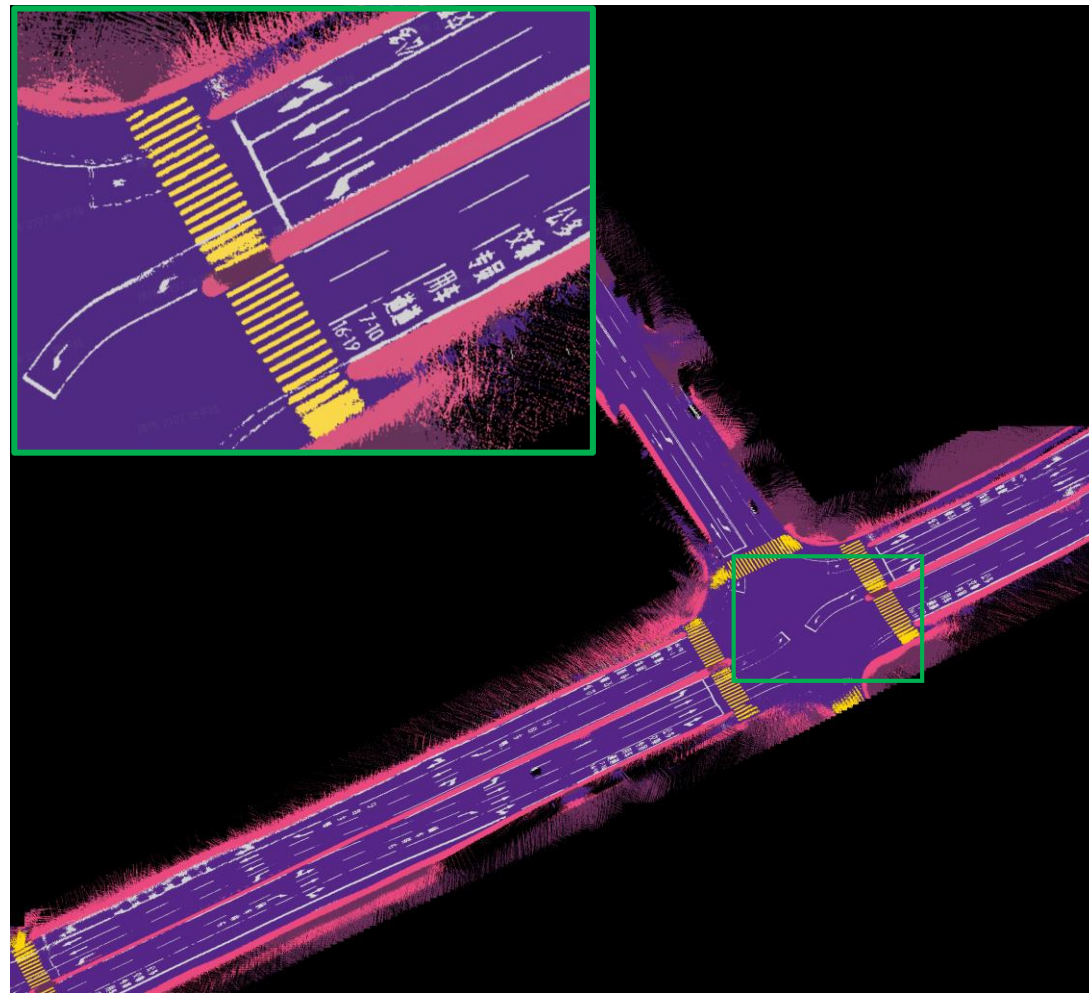
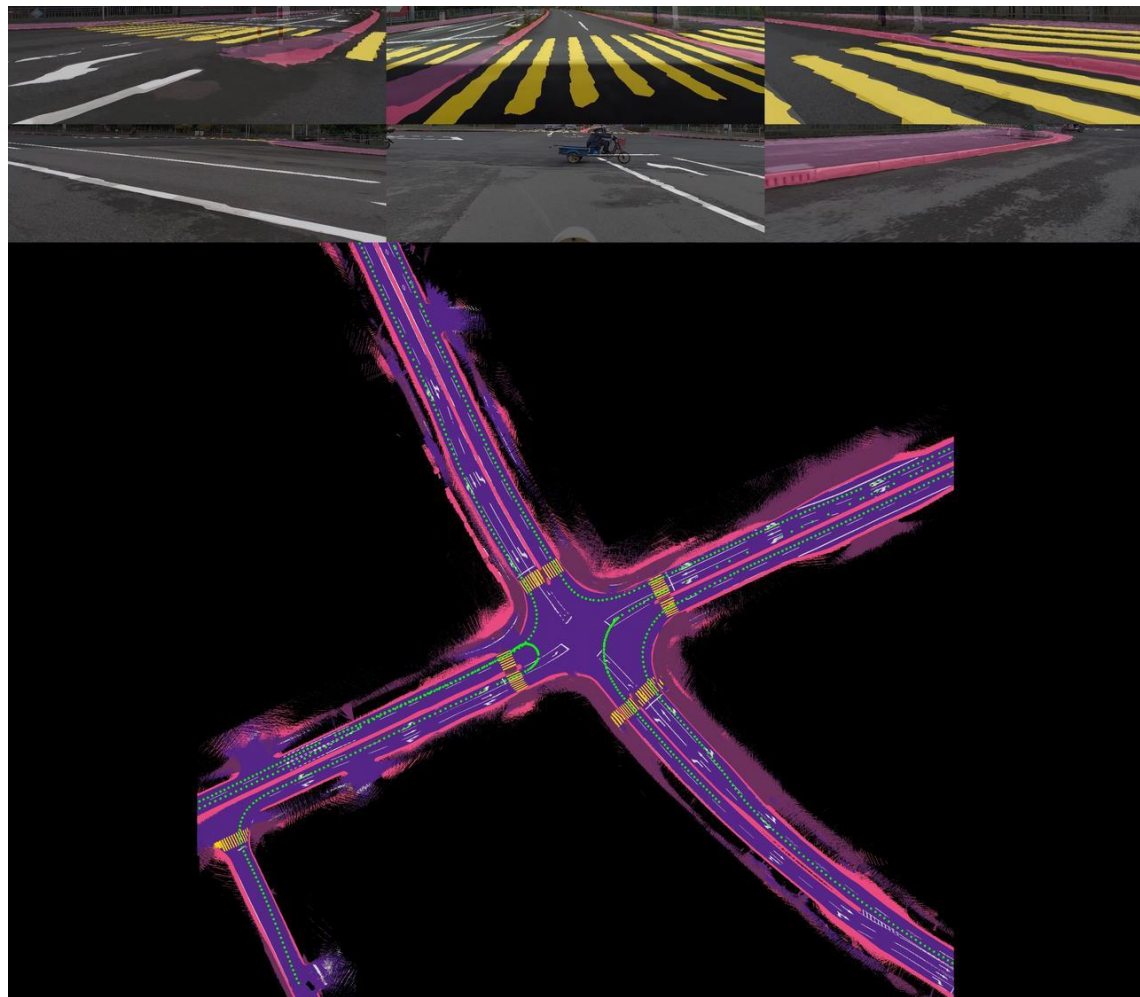


路面语义重建效果



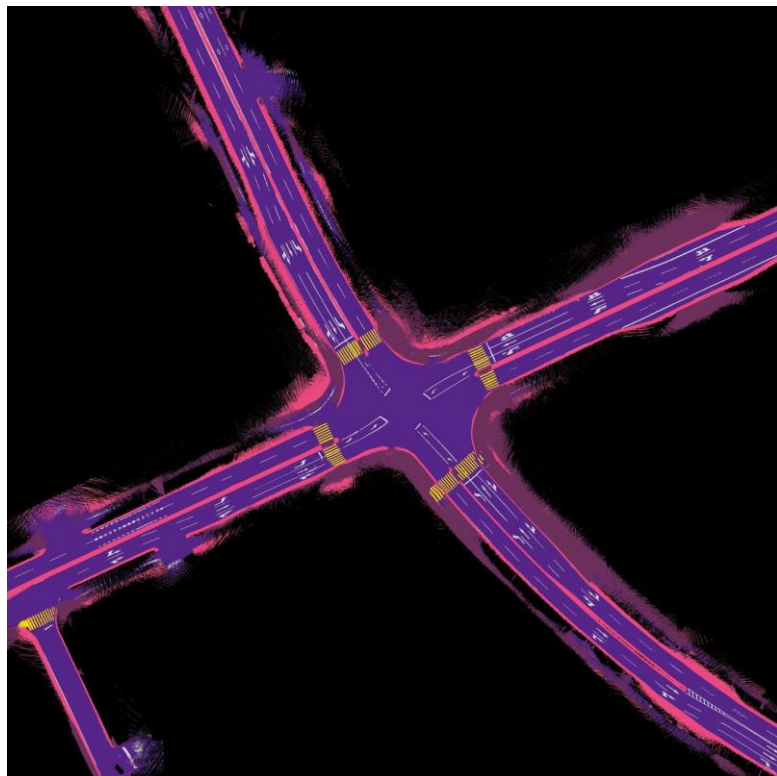
NeRF 用于路面元素自动标注

04. 云端自动标注-基于NeRF路面重建

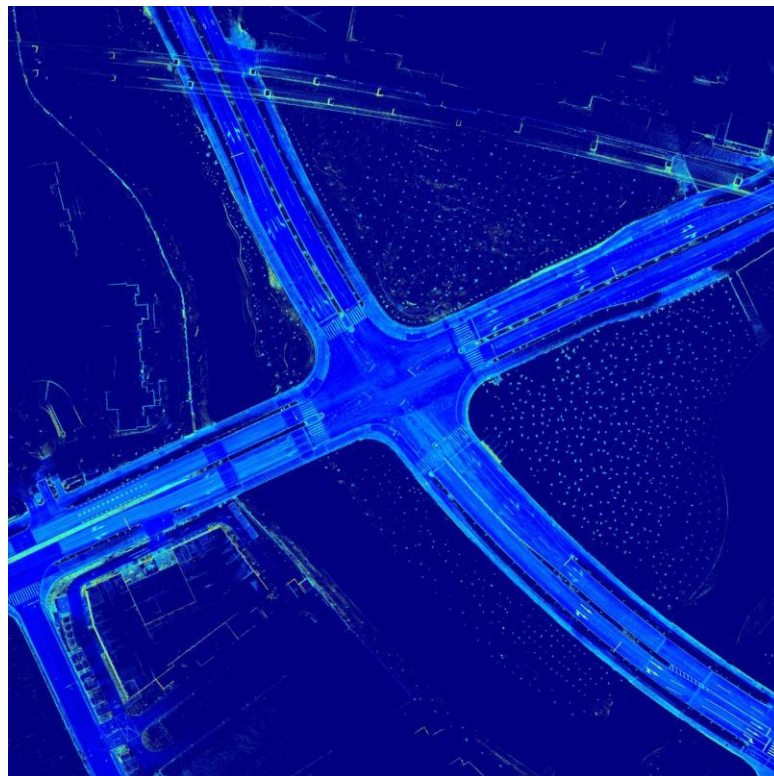


04. 云端自动标注-基于NeRF路面重建

高一致性多图层重建



语义图层



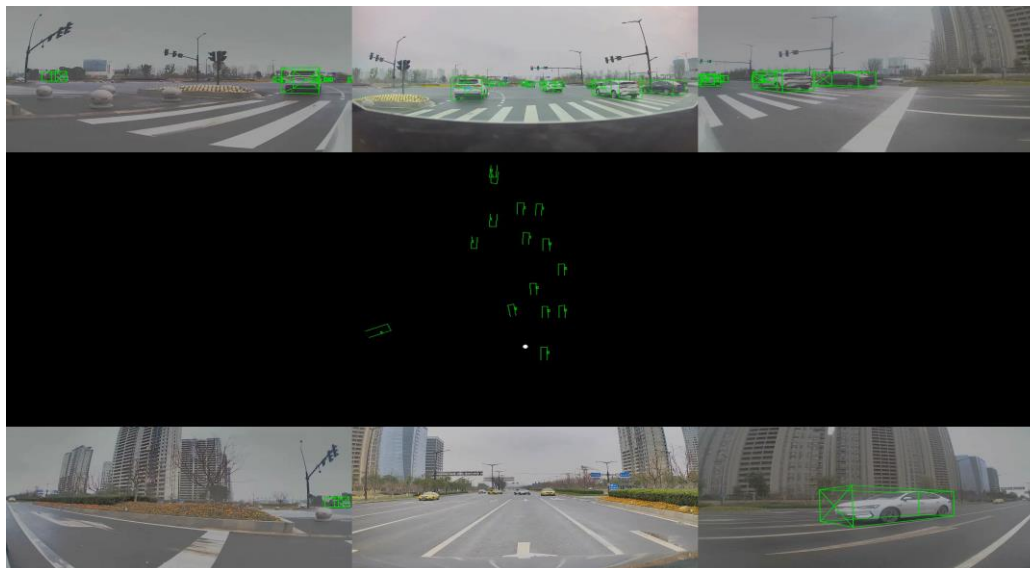
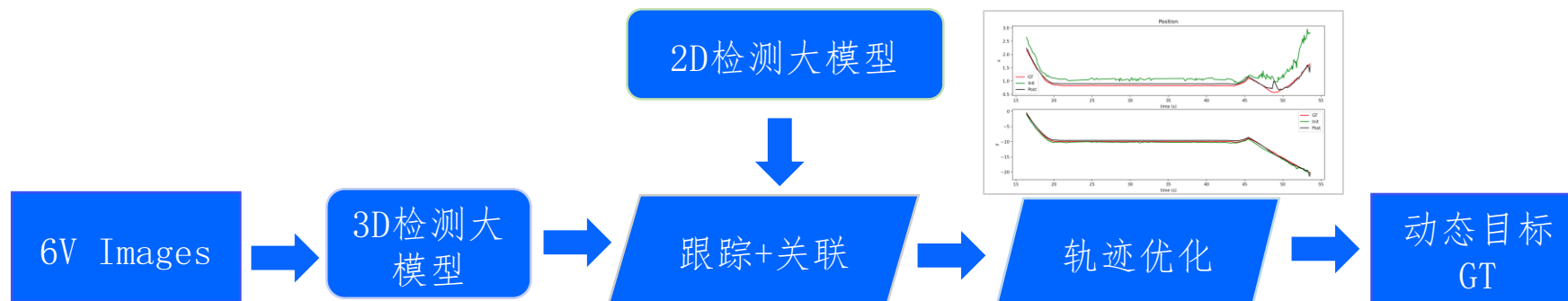
激光点云图层



视觉图层

04. 云端自动标注-纯视觉动态物体标注

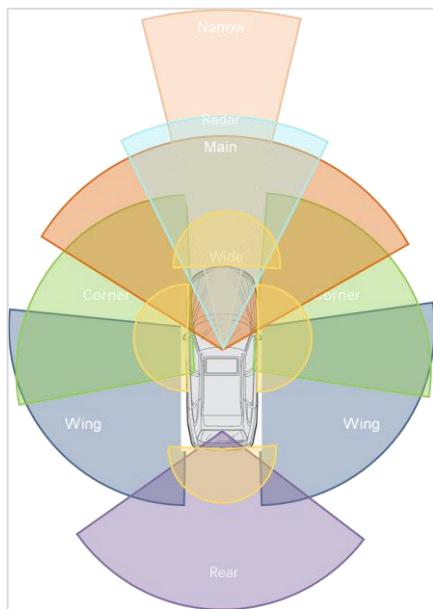
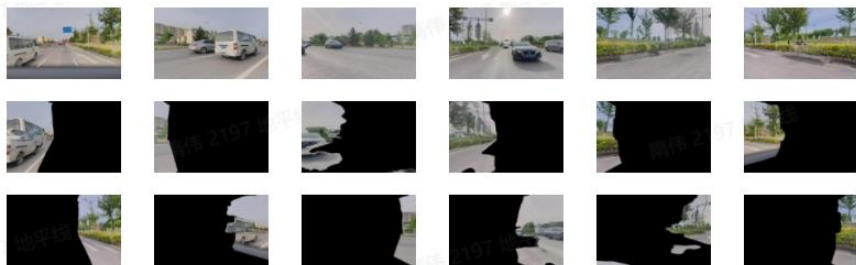
- 云上大模型可以采用更大容量的模型以及性能更高的时空融合策略
- 当云端大模型性能指标高于端上10个点以上，便具备作为真值预刷的可行性



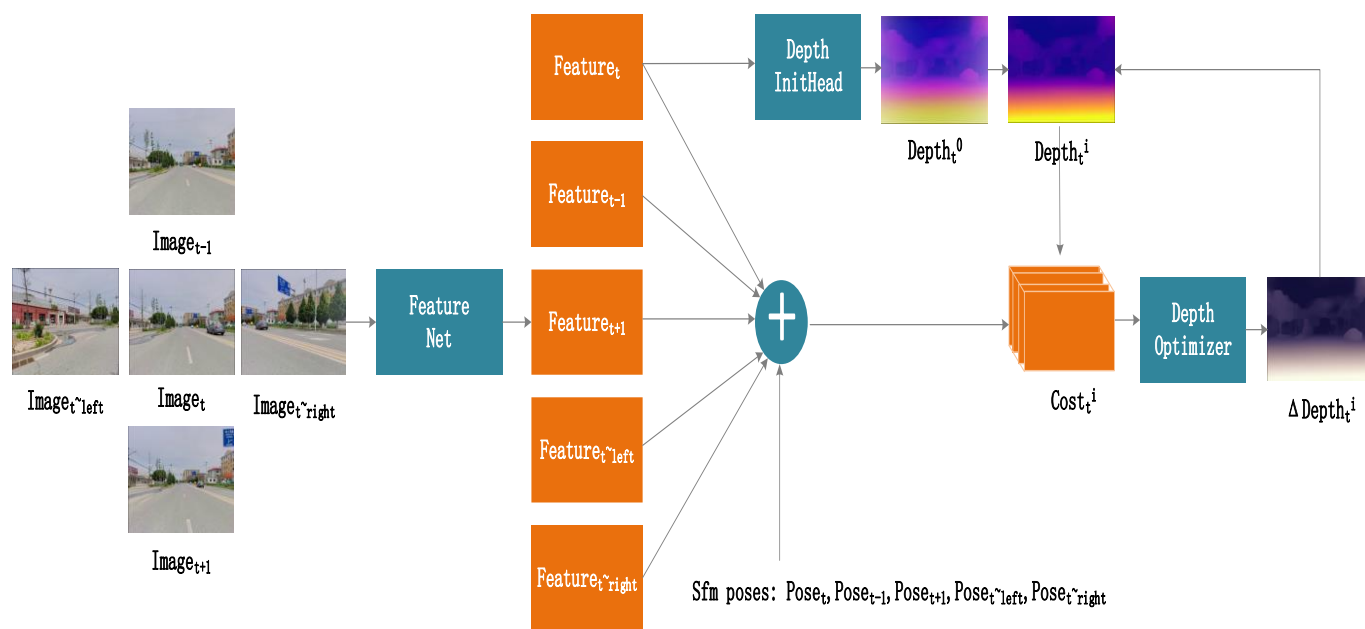
04. 云端自动标注-Visual LiDAR (VIDAR)

- 采用多传感器之间的空间约束和时序约束进行自监督学习
- 采用DRO 的优化模块提升网络的泛化性能

Multi-cam 时空约束

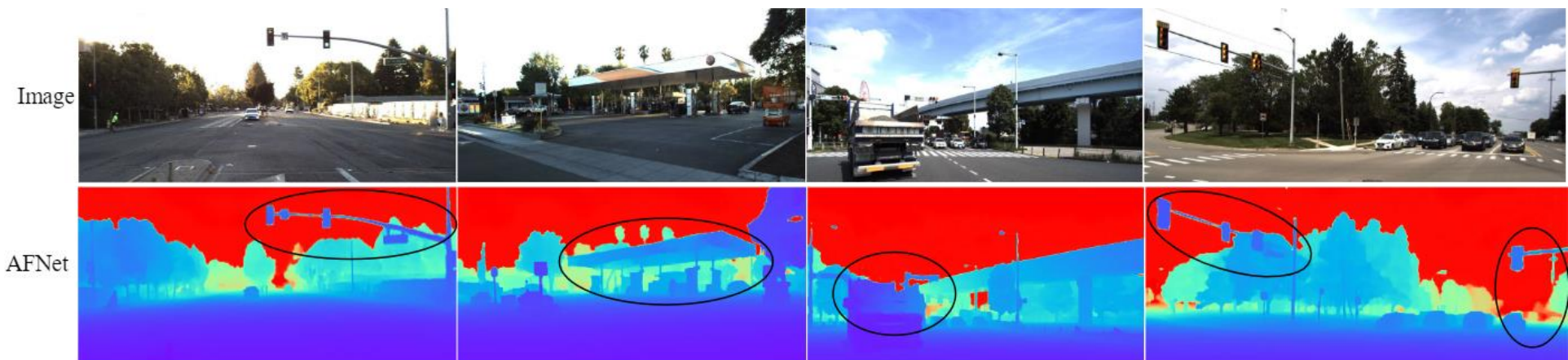
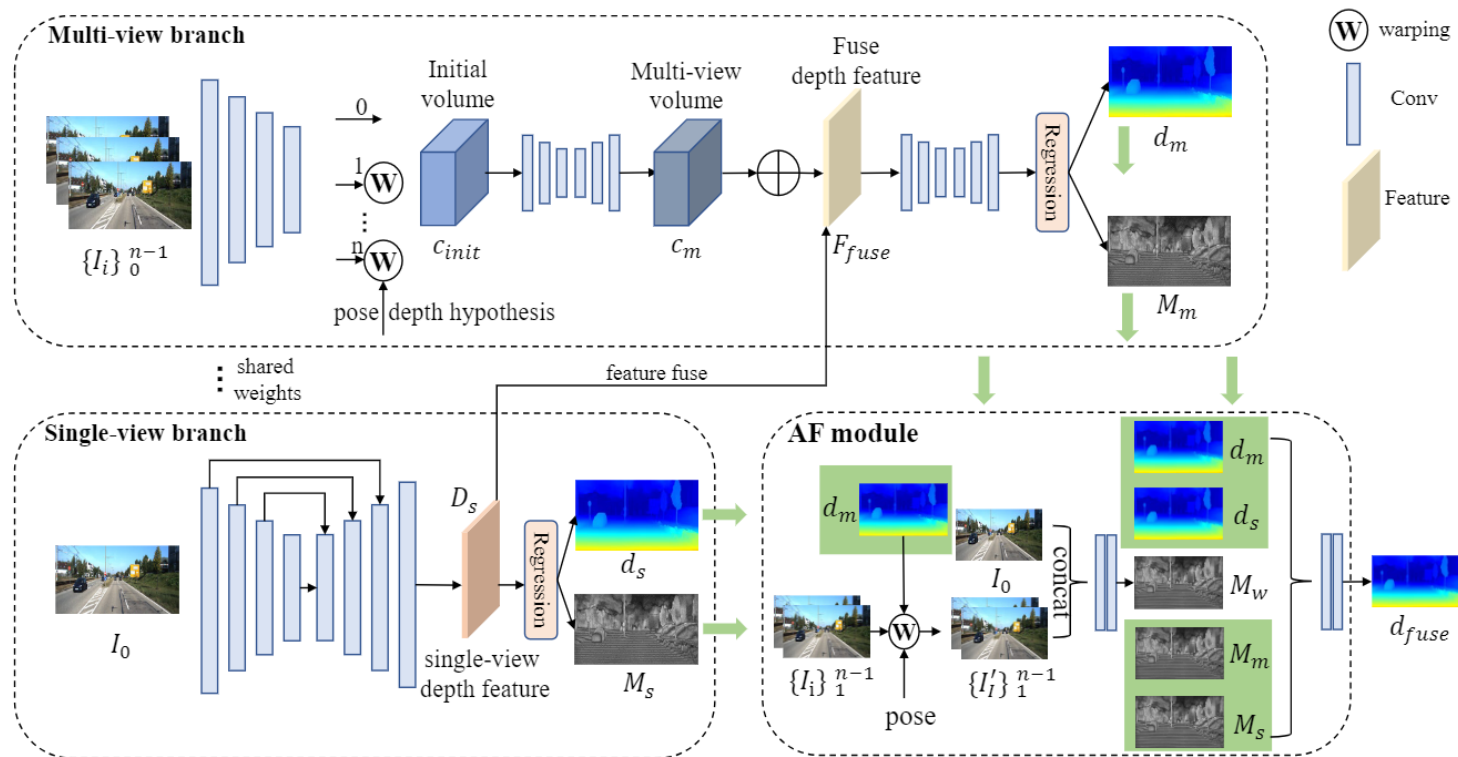


模型采用DRO



04. 云端自动标注-Visual LiDAR (VIDAR)

- AFNet (CVPR2024)
- 利用车辆周视相机形成时空一致性监督



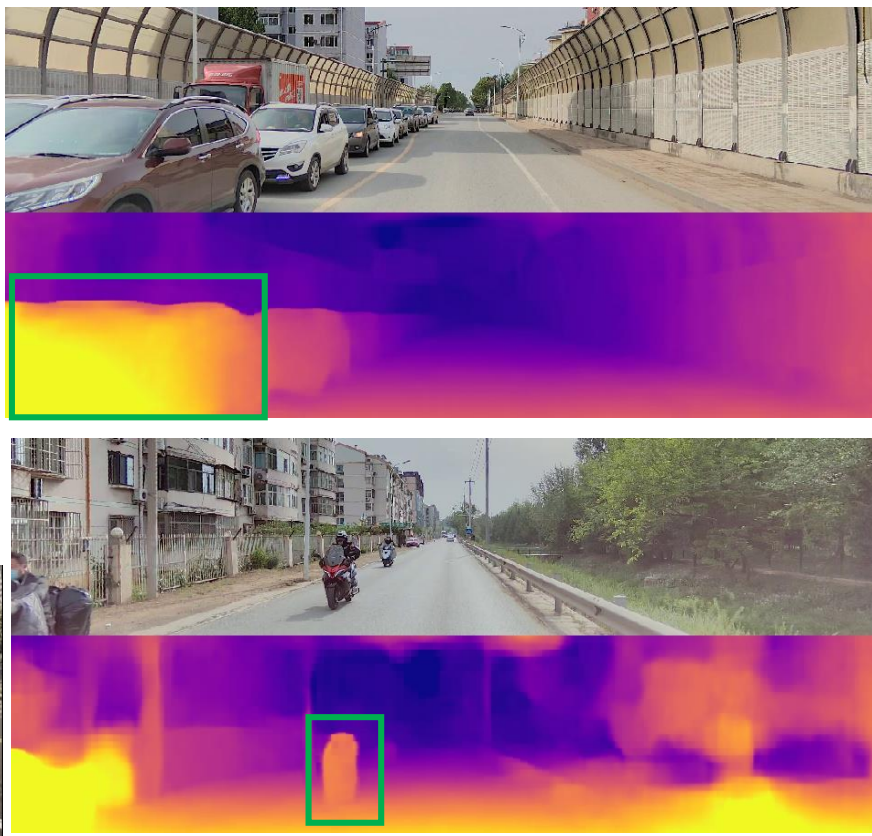
04. 云端自动标注-Visual LiDAR (VIDAR)

- 对于depth的估计，采用仿真数据，可以有效地提升动态物体边缘的感知感知效果

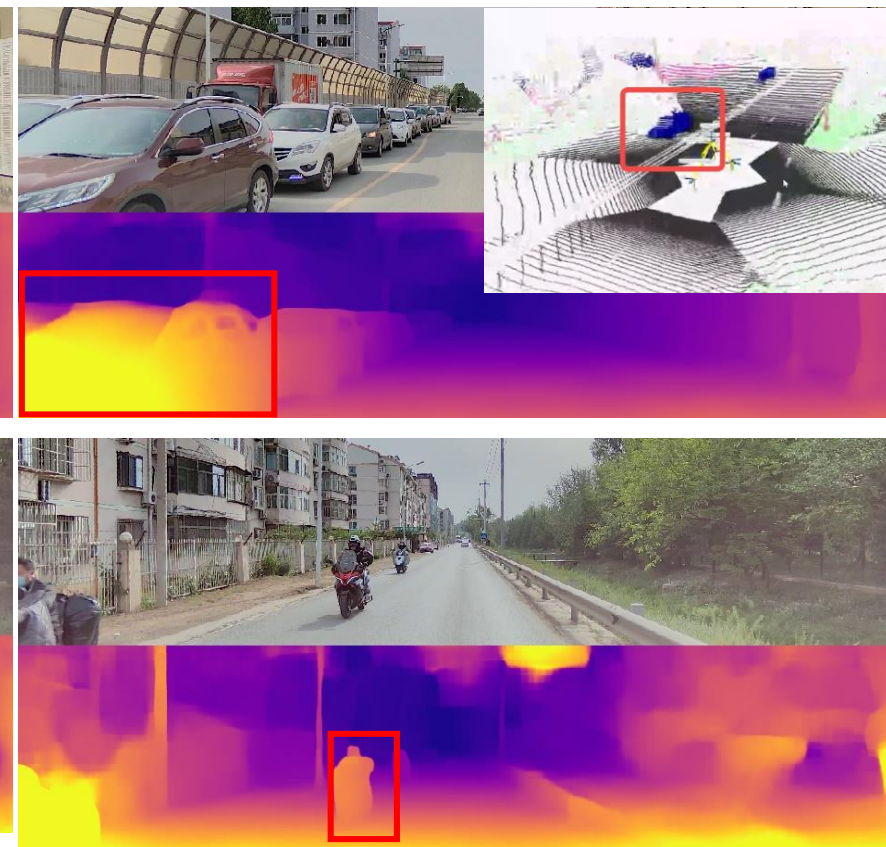
Carla提供的仿真数据



未使用仿真数据

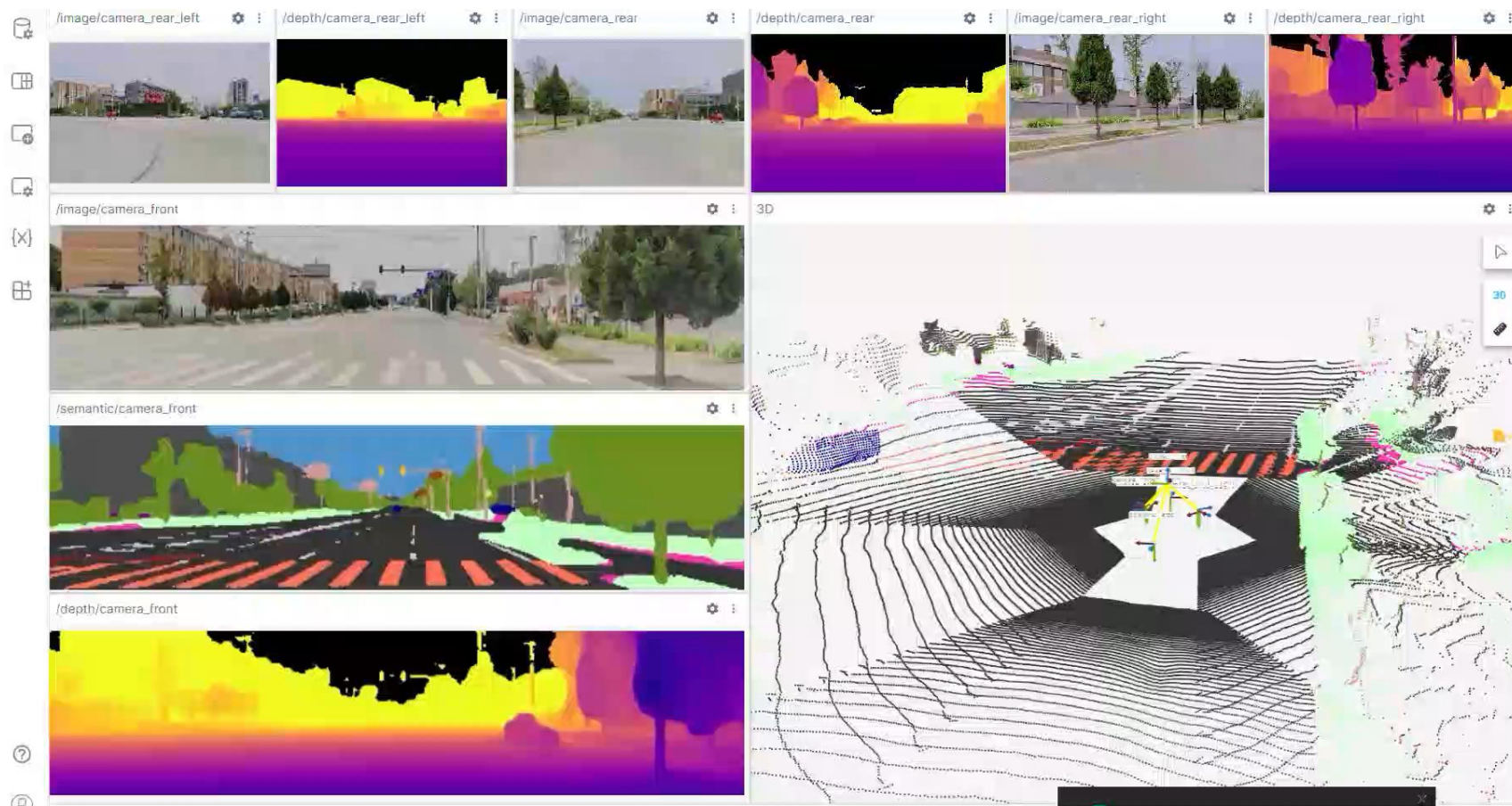


使用仿真数据



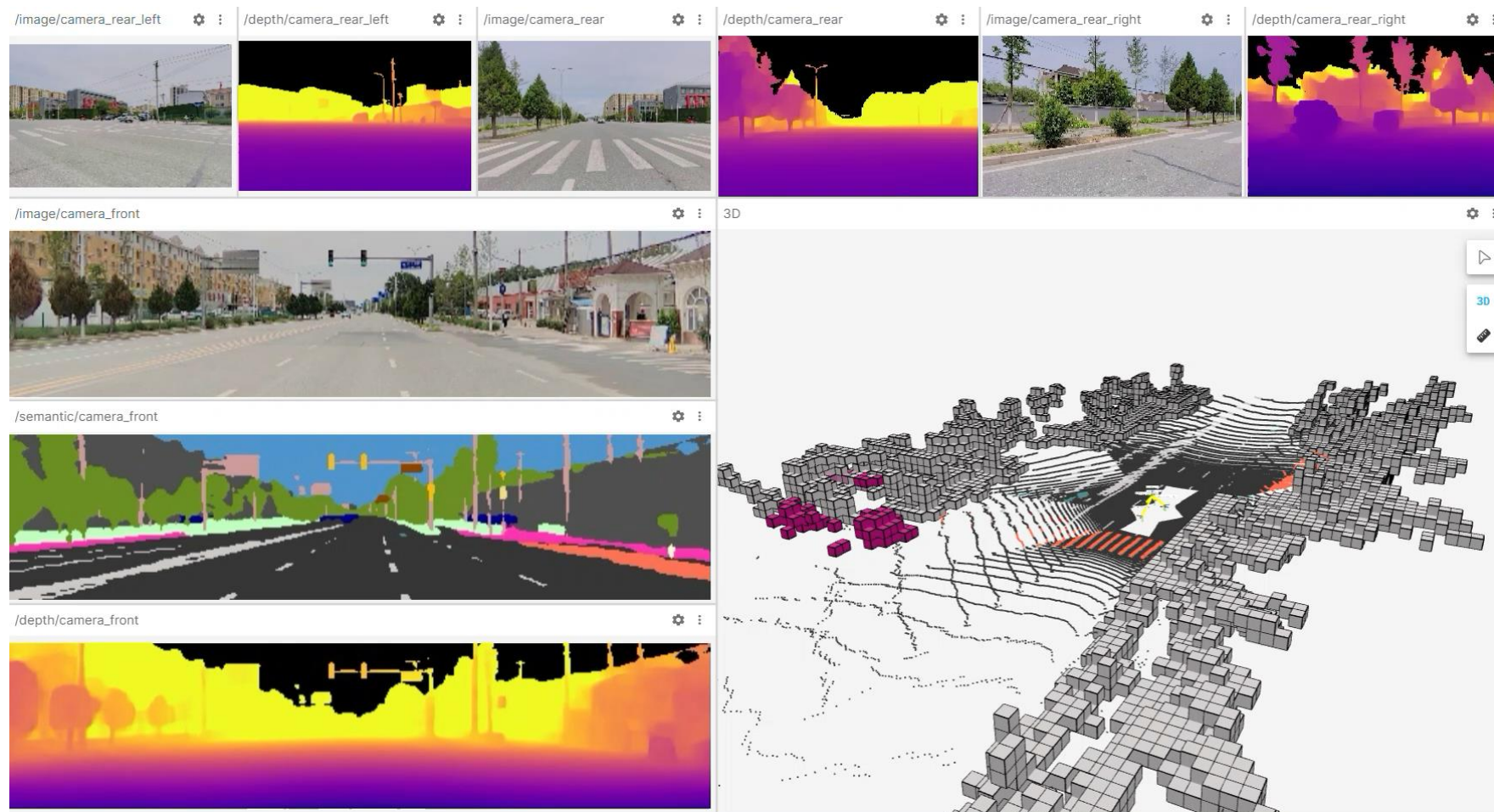
04. 云端自动标注-Visual LiDAR (VIDAR)

- 50m以内对较好的测距精度 $\text{abs.rel} < 8\%$
- 可用于对动态物体进行标注，包括车辆、行人等，也可以用于生成Occupancy/Occupancy flow的真值



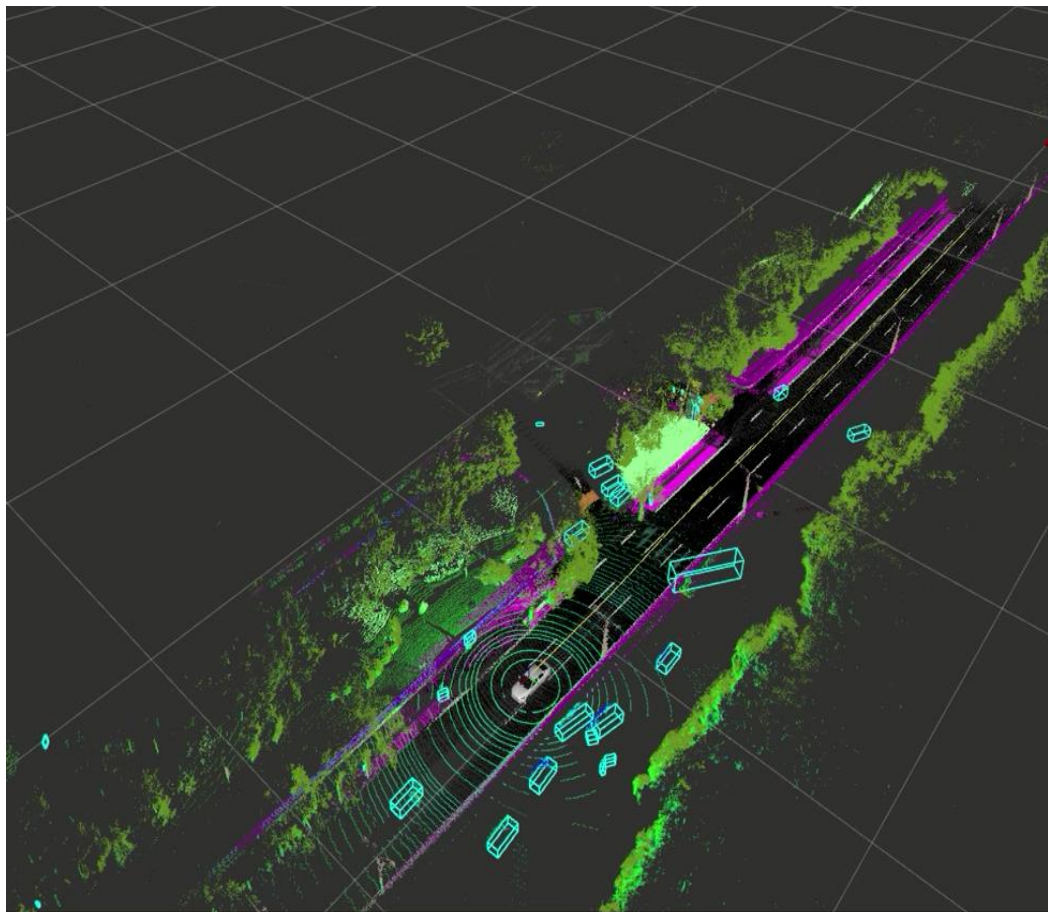
04. 云端自动标注-Visual LiDAR (VIDAR)

- Occupancy/Occupancy flow的真值可用于处理场景的中的通用障碍物
- Occupancy/Occupancy flow 可以看作是depth的端到端融合

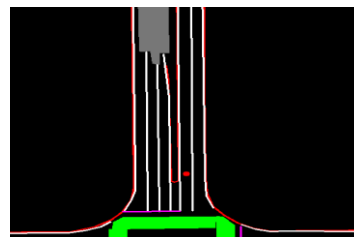


04. 云端自动标注-支持各种下游任务

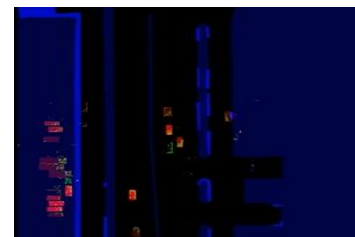
云端大模型提供GT预刷



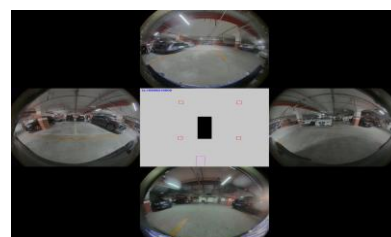
已经支持的GT任务



Bev-seg



Bev Freespace



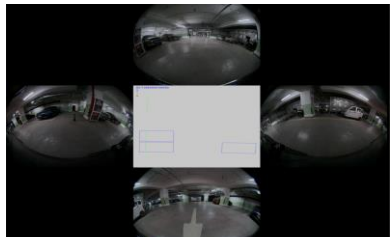
泊车-车位检测



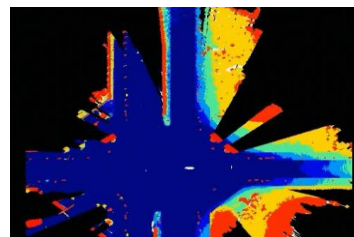
Bev-全要素



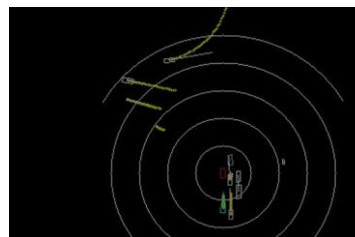
Bev 3D



泊车-可行驶区域



Bev-高程



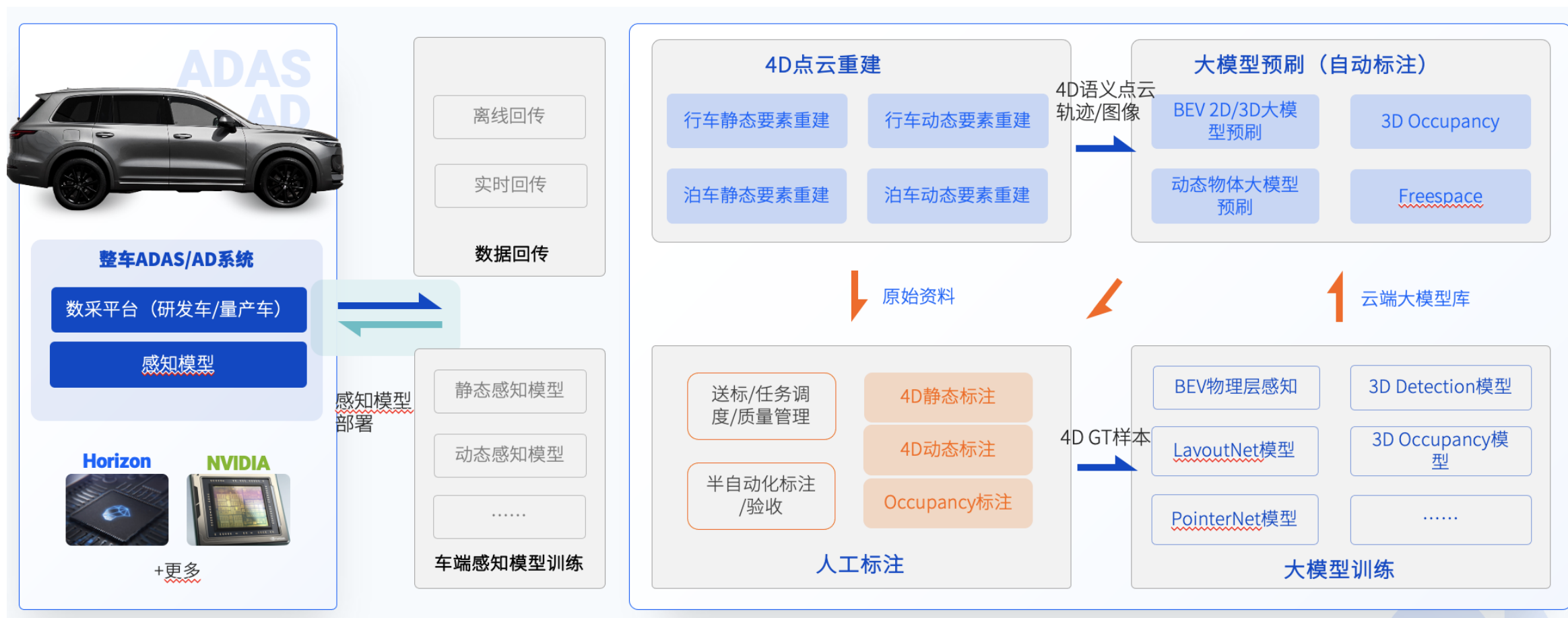
Bev-Tracking/prediction



Depth

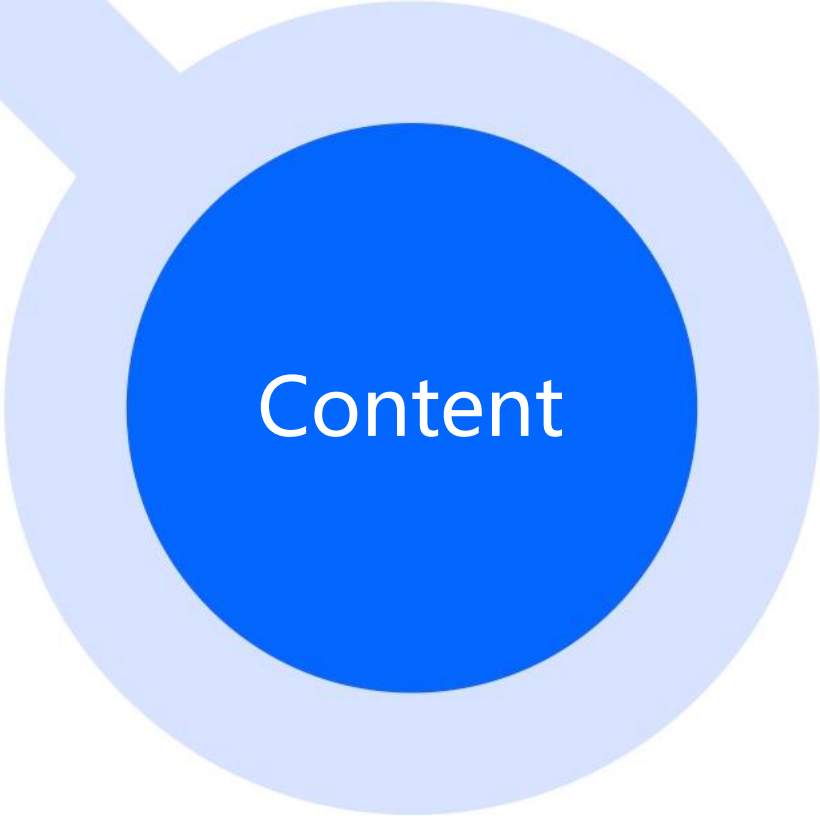
04. 云端自动标注-系统工程

低成本，可持续迭代的工程系统比算法本身还重要



04. 云端自动标注-总结

1. 模型的迭代效率和成本是AI产品的核心竞争力
2. 高精度高稳定性的定位和建图是4D Label的基础
3. 端到端是一个逐步的趋势，SW1.0->SW2.0->SW3.0
4. 车端后处理算法逐步从端上走到云上，成为真值标注方案
5. 数据合成和仿真系统在自动驾驶算法构建中作用逐渐凸显



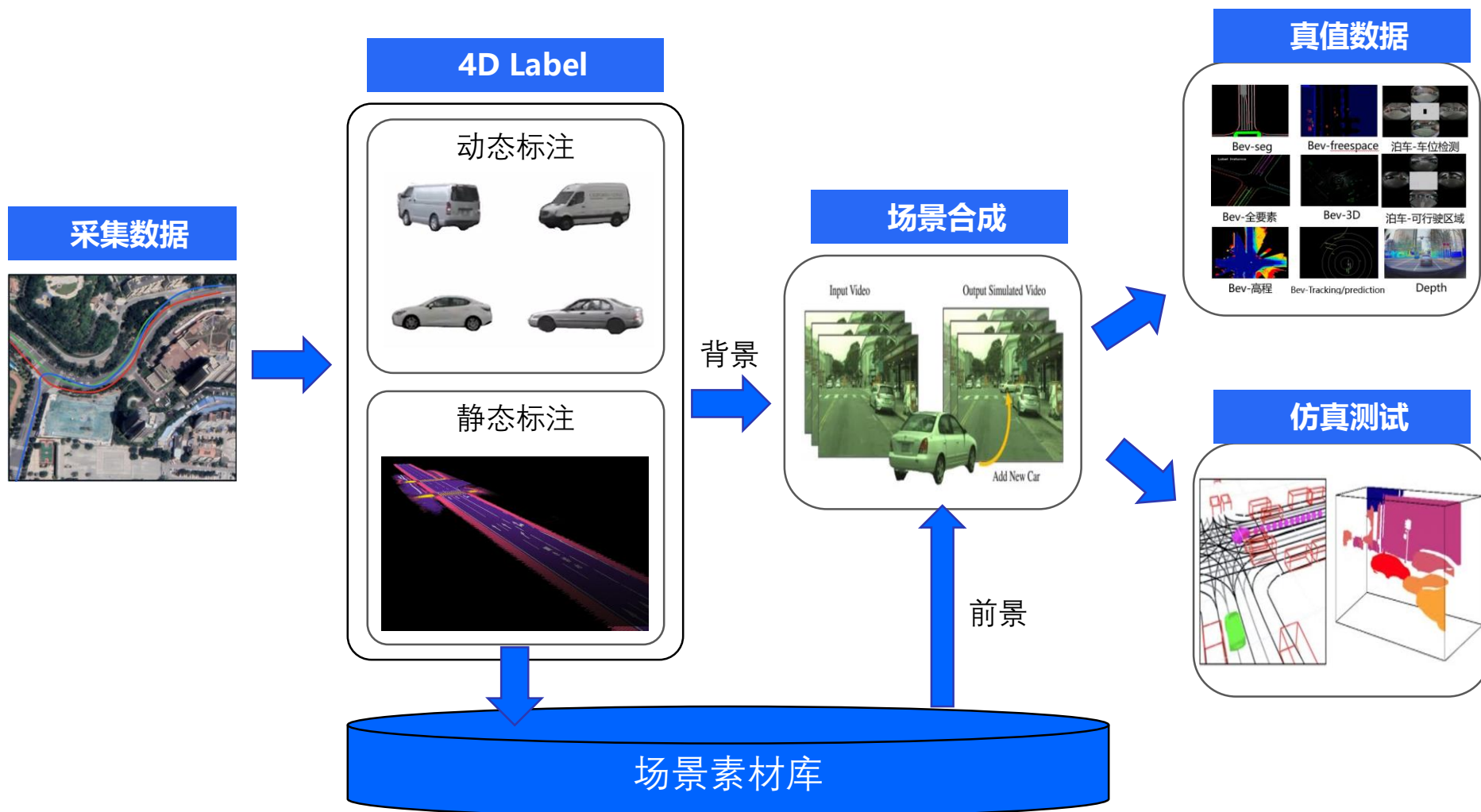
Content

05. 仿真系统

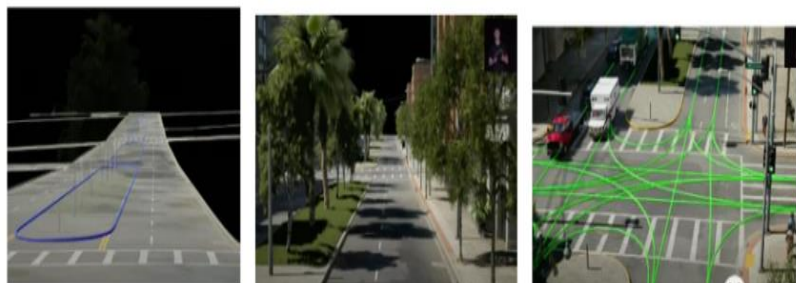
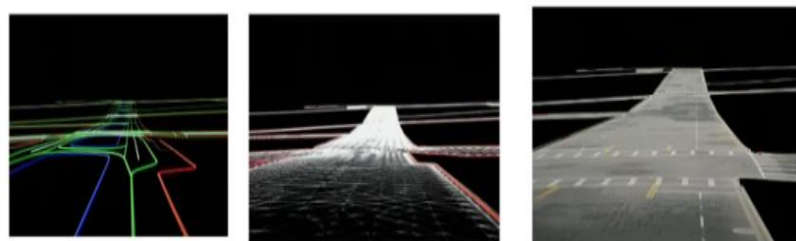
- (a) 传统仿真流程
- (b) 仿真系统作用
- (c) 基于NeRF仿真

05. 仿真系统-4DLabel的自然延续

- 数据仿真的主要难度在于场景库的构建，4D Label提供4D重建和结构信息提取，可以为仿真器提供丰富的场景库素材



05. 仿真系统-常见流程



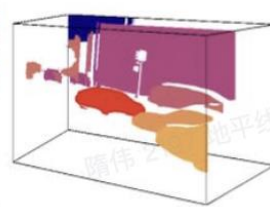
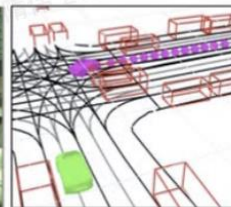
场景库构建



Realistic

Traffic-aware

Geometrically Consistent

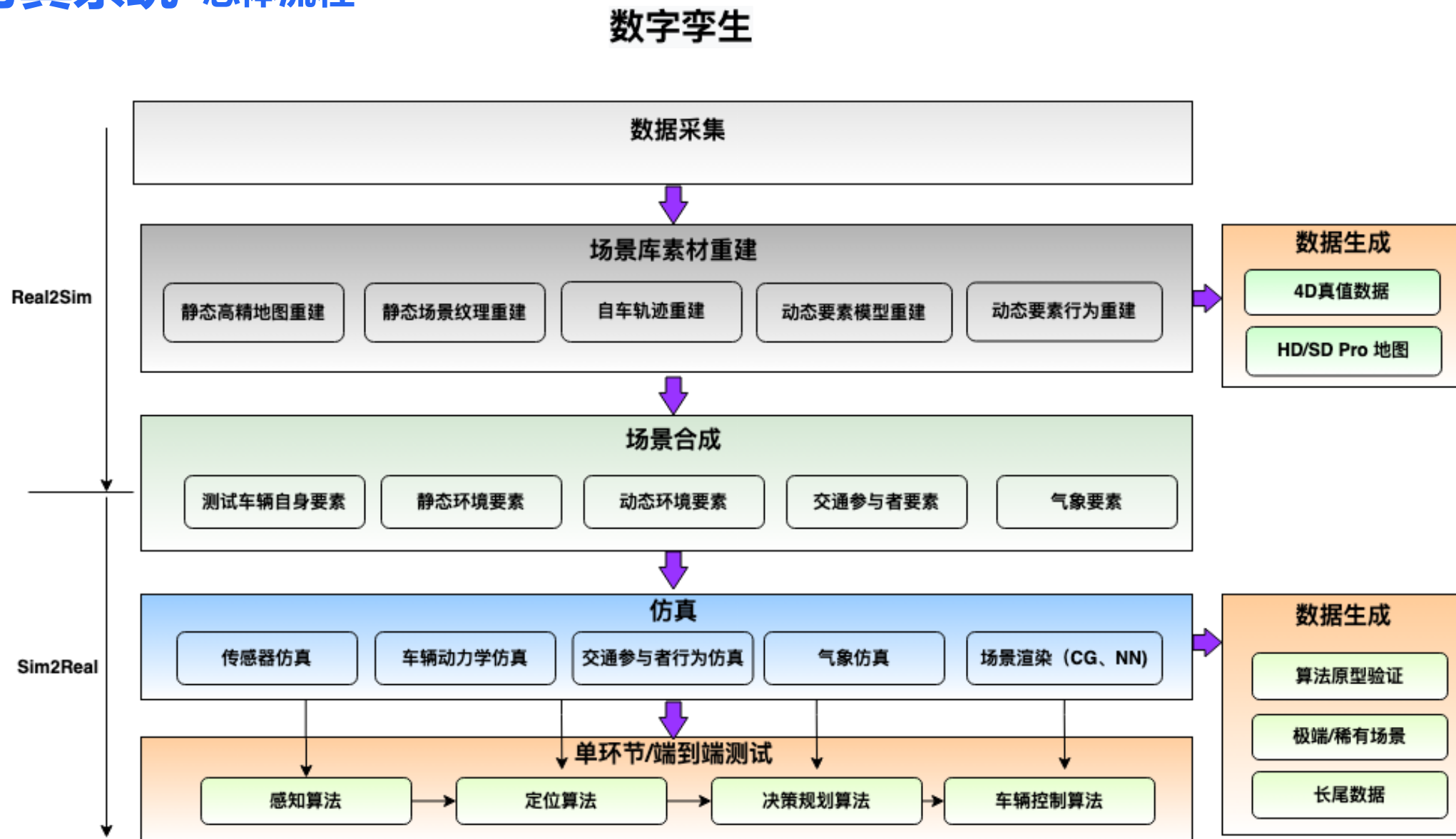


场景合成

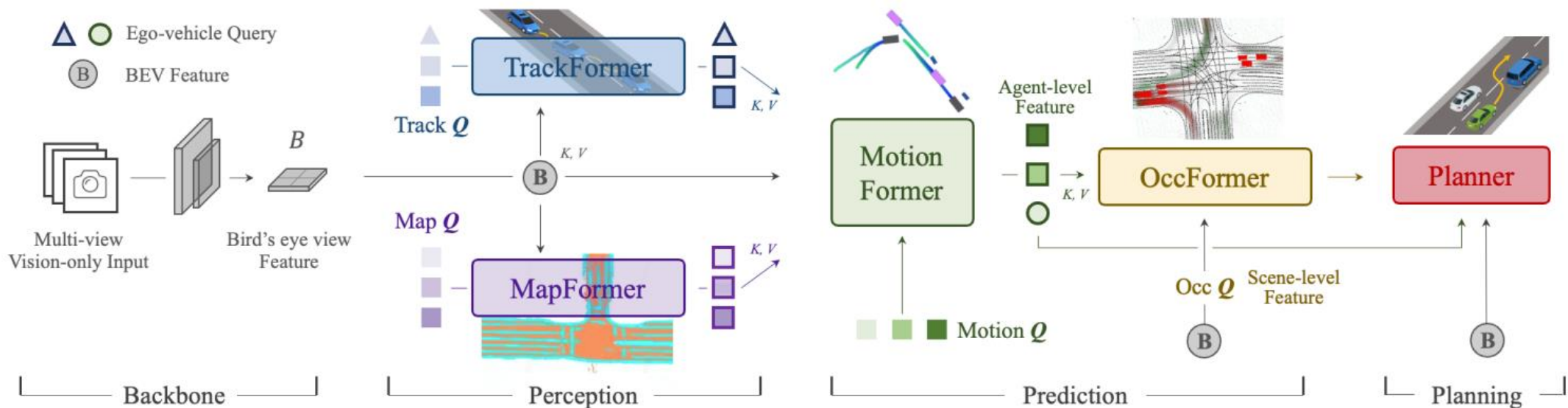


真实感渲染

05. 仿真系统-总体流程



05. 仿真系统-开环仿真



开环仿真： UniAD

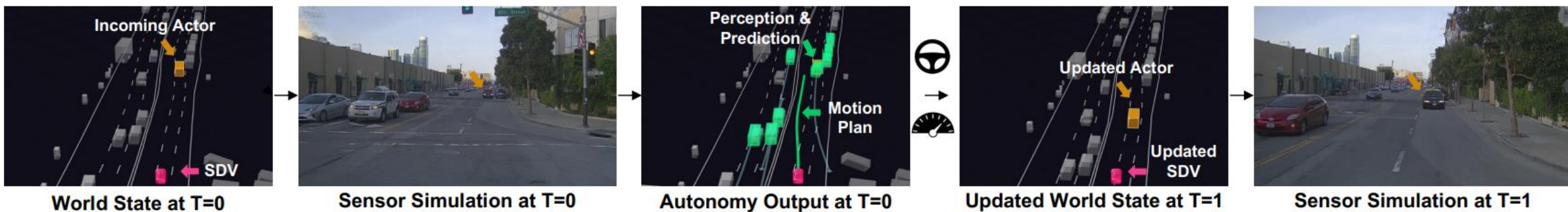
优点：

数据集固定，可以快速打通端到端自动驾驶

缺点：

训练出的Planner无法保证在真实场景取得同样的效果

05. 仿真系统-闭环仿真



闭环仿真：UniSim

优点：

真正的闭环仿真，planner的决策可以及时反馈

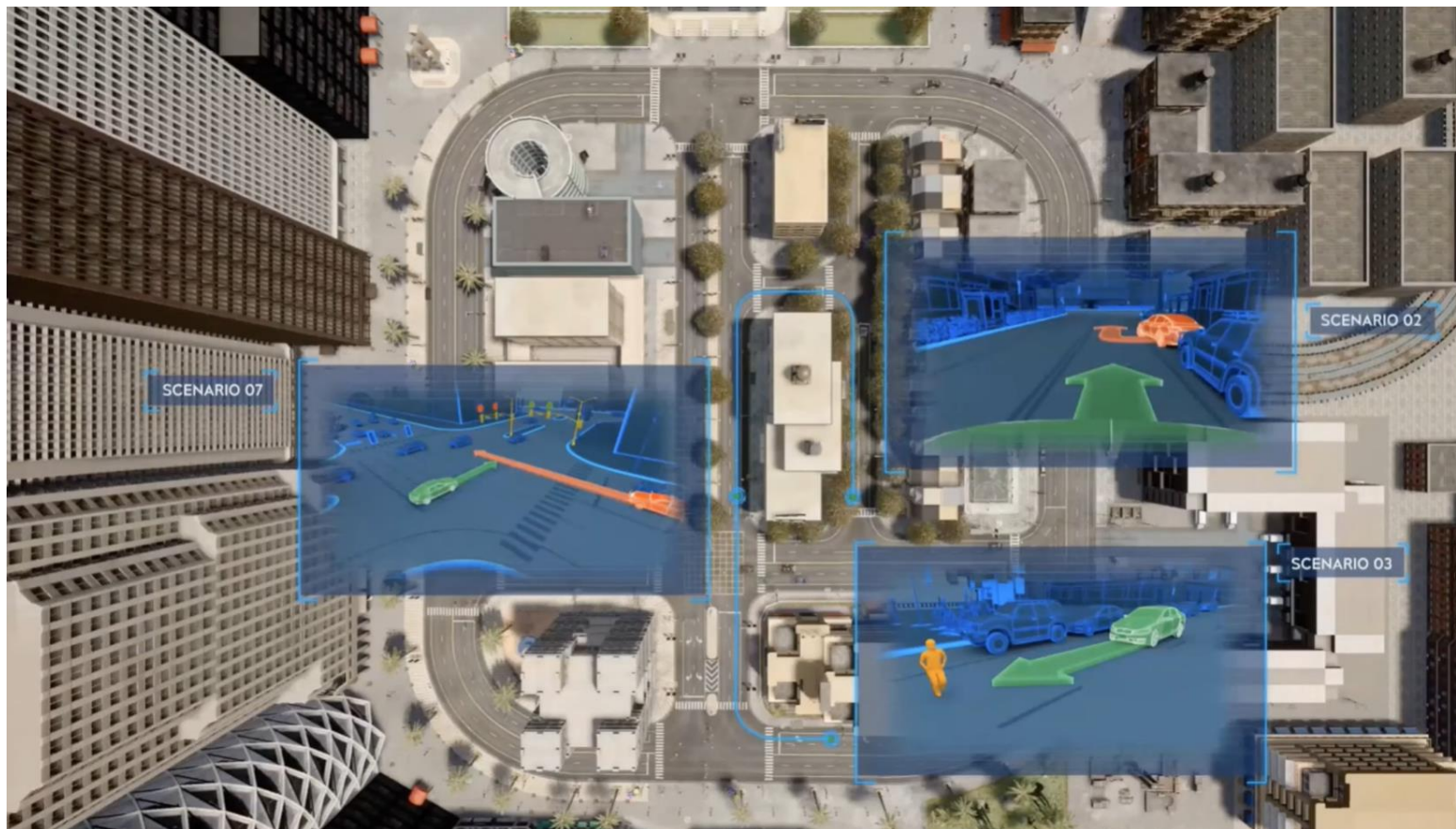
缺点：

目前基于NeRF的自动驾驶场景仿真是一项巨大的系统工程，目前还没有十分成熟的开源工作

05. 仿真系统-作用

提升

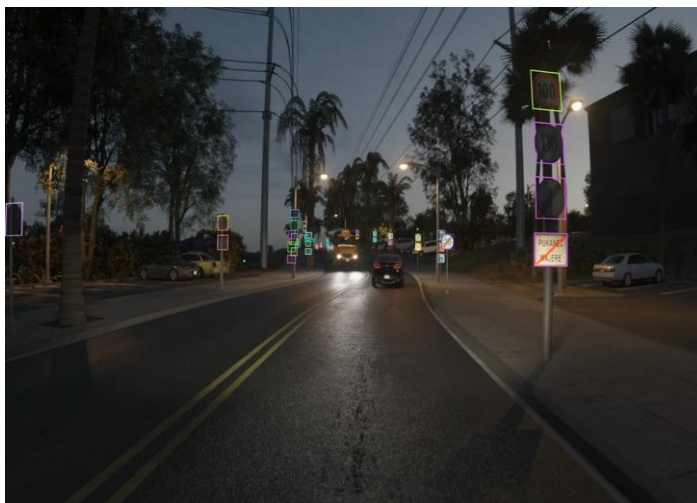
- ✓ 感知精度
- ✓ 传感器泛化性
- ✓ 场景泛化性
- ✓ 测试迭代效率



开源自动驾驶仿真器: CARLA

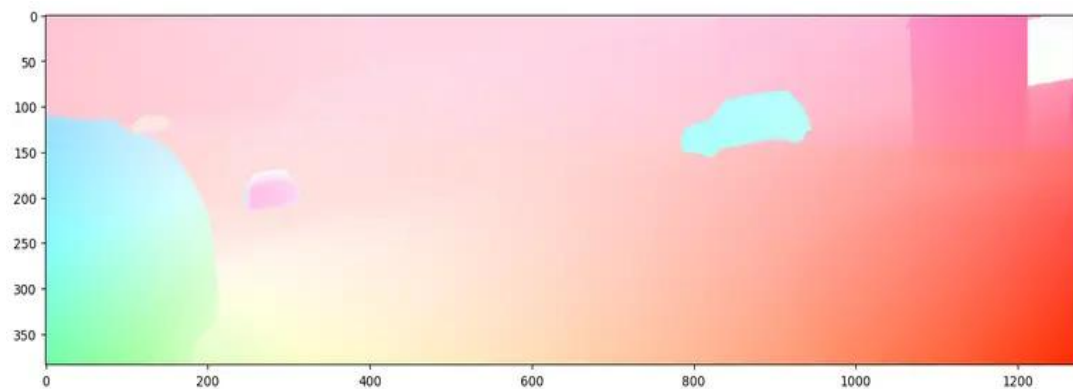
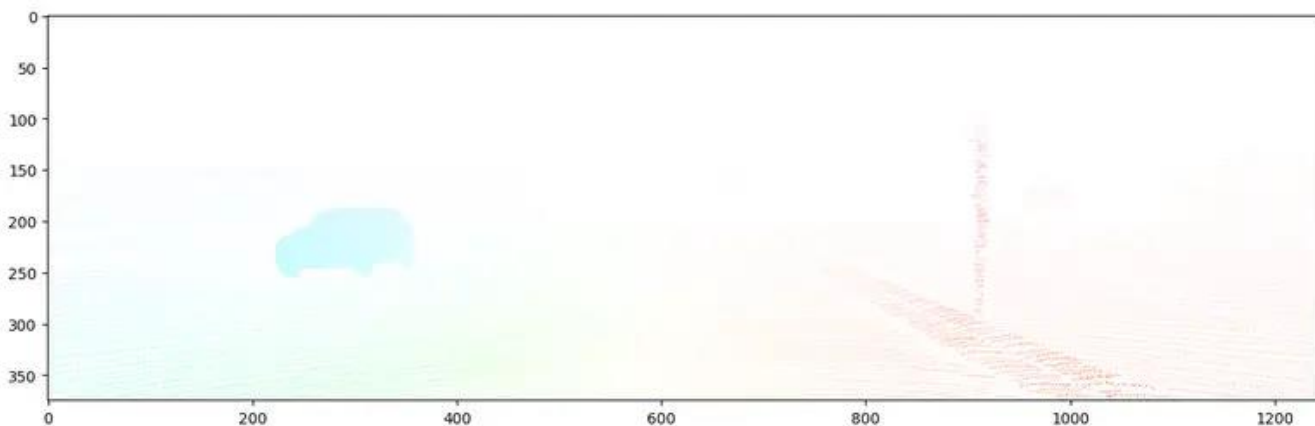
05. 仿真系统-迅速提升感知性能

PD可提供高度定制化的场景，交通标志，以及标志的形变，损毁等极端情况，模型精度平均提升 +15% mAP，部分数据较少的极端情况，精度提升可达 +50% mAP



05. 仿真系统-迅速提升感知性能

- 在已有的训练pipeline上，加入了PD的仿真数据（数据分布相对真实数据来说，分布差距比较明显），预测精度相对提升 +6%
- 调整PD的仿真数据分布，使其接近真实数据的分布，预测精度相对提升 +18.5%，会有一个明显的阶跃式提升



05. 仿真系统-提升感知性能

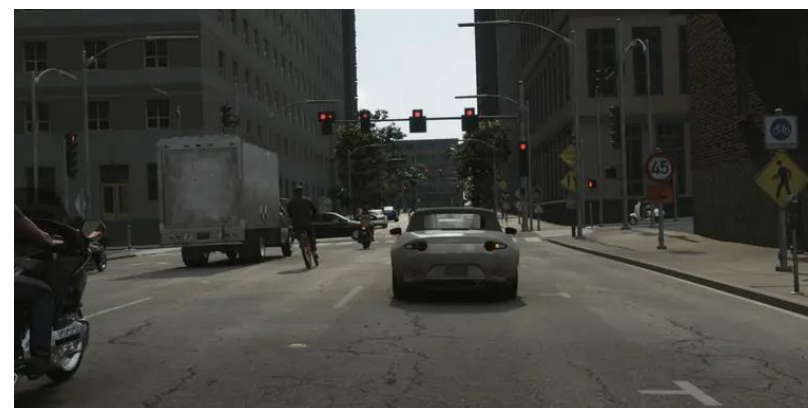
- 仿真数据从分布上越逼近真实数据，对模型的提升越明显



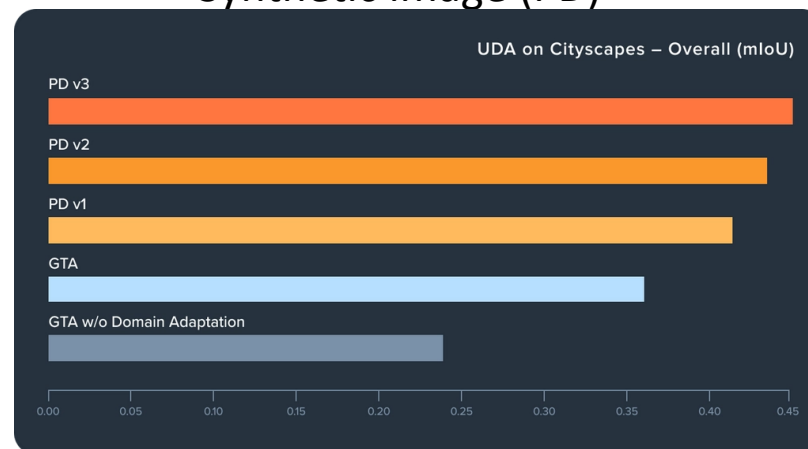
Real world image (Cityscapes)



GTA image



Synthetic image (PD)



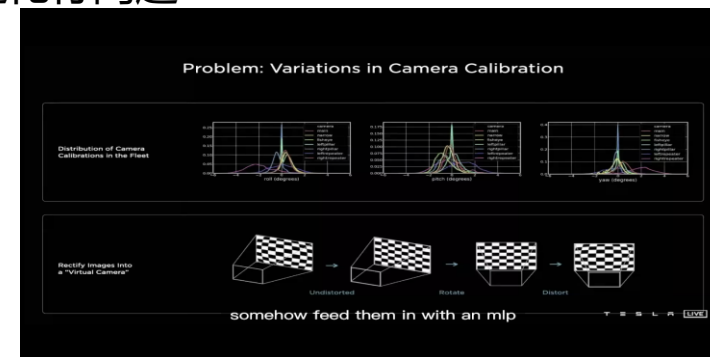
<https://paralleldomain.com/>

05. 仿真系统-提升传感器泛化性

- 对于3D感知任务，当传感器的安装位置发生变化时，会存在严重的模型泛化性的问题，是目前量产中的痛点问题
- Tesla采用rectify的方式事先对图像进行校正，但是仅仅能处理纯旋转的情况，无法处理相机的高度变化
- 采用vidar提供的depth可以对任意相机高度变化进行数据仿真合成，可以有效地解决泛化行问题



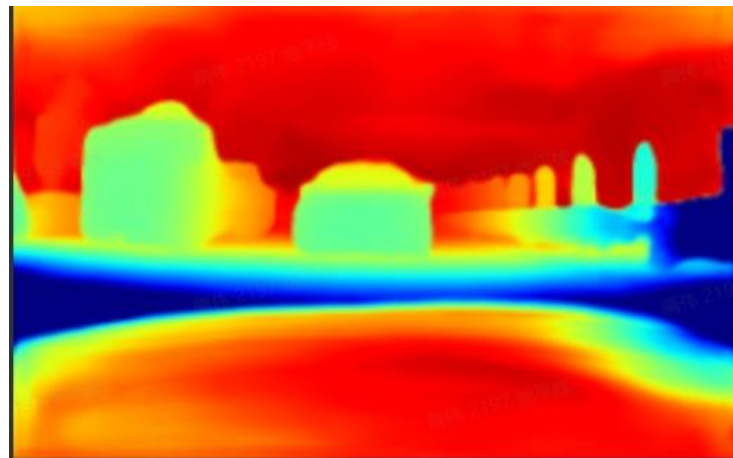
$$H = K * \left(R - \frac{N * T}{camH} \right) * K^{-1}$$



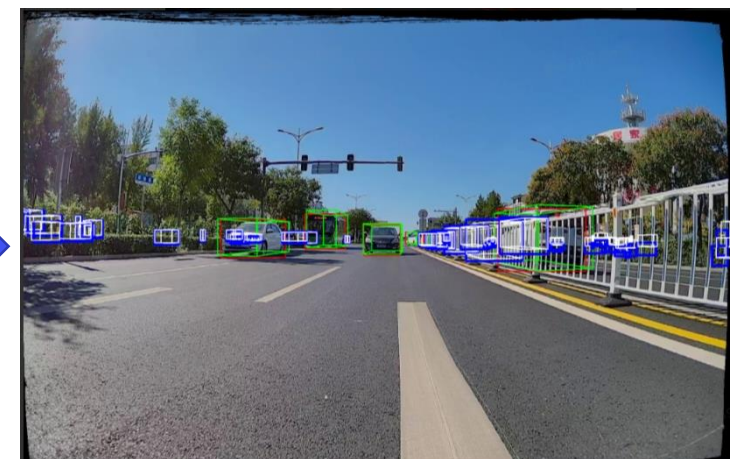
Tesla 的Rectify 层



已有真值数据



Depth数据



目标车辆真实数据

05. 仿真系统-提升场景泛化性

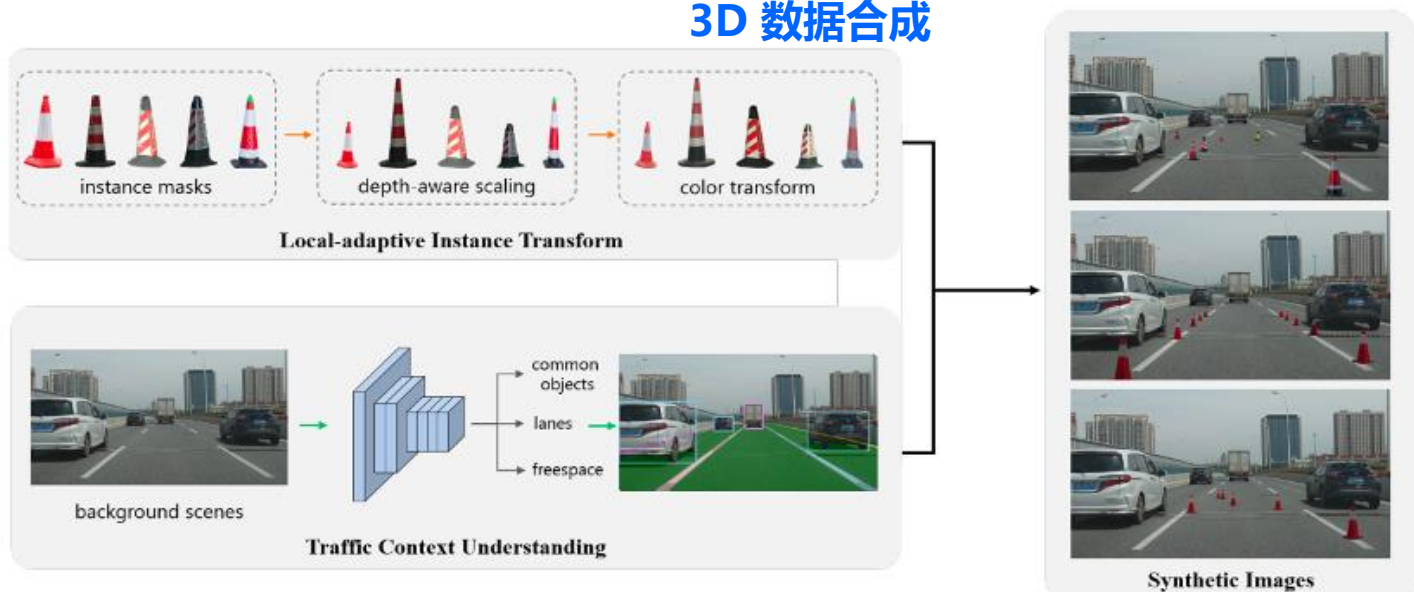
- 针对一些非常见的的数据，采用数据合成的方式可以显著提升模型的性能
- 在驾驶场景中翻车的2D检测任务中，采用数据合成的方式可以提升AP 10个点



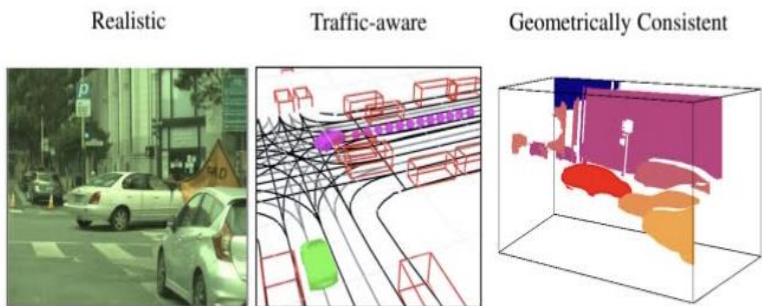
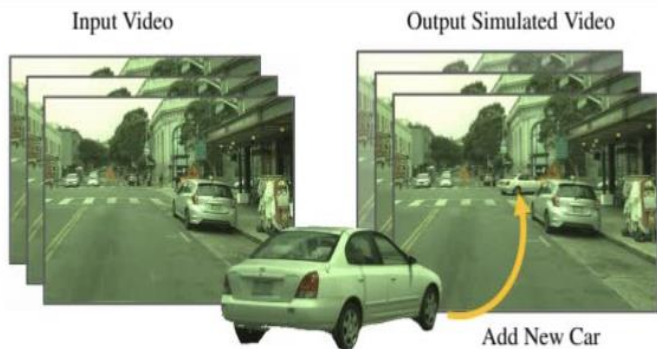
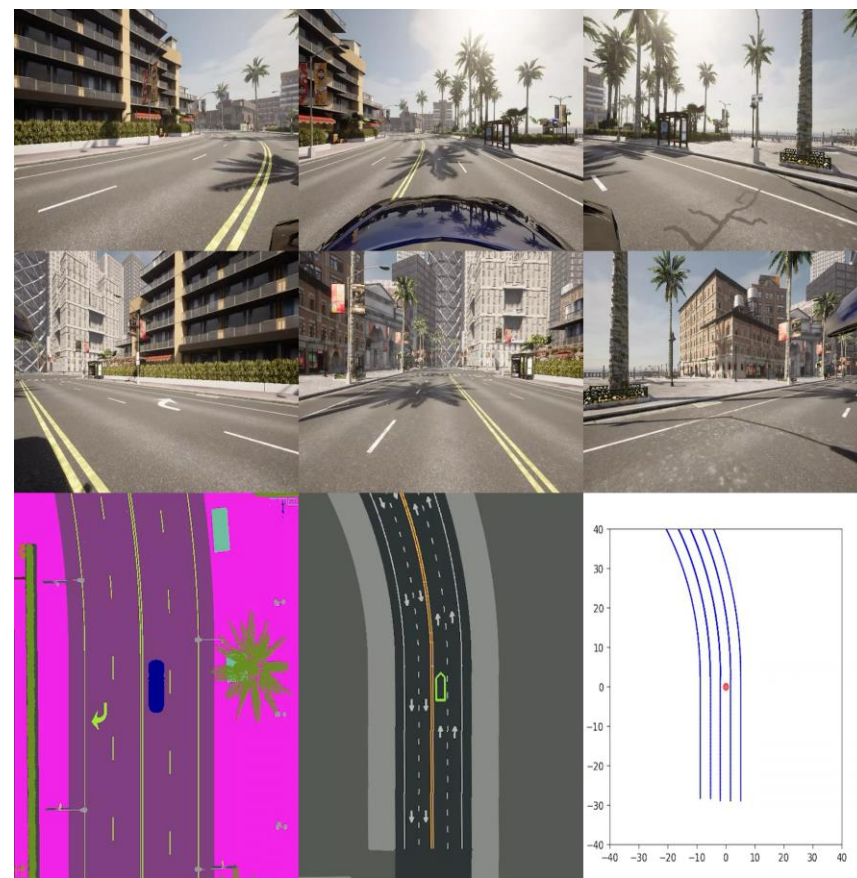
05. 仿真系统-提升场景泛化性

- 结合4D重建，可以再3D空间实现数据合成任务，从而为3D感知任务提供真值
- 相比较与全场景的仿真，数据合成的方式更容易实现，且保留了更多的真实感信息，从而能够保证较好的泛化性

3D 数据合成

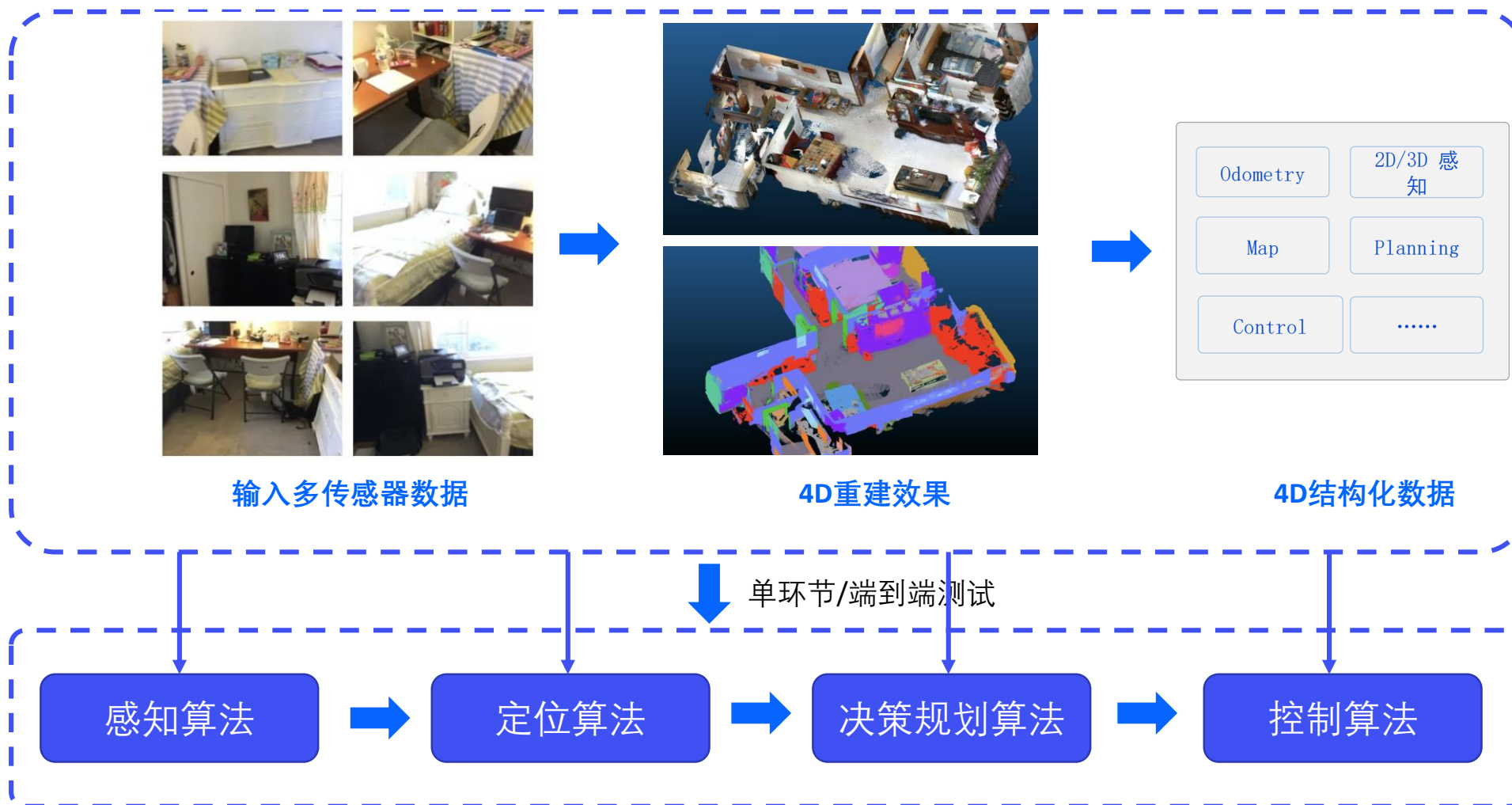


全场景仿真

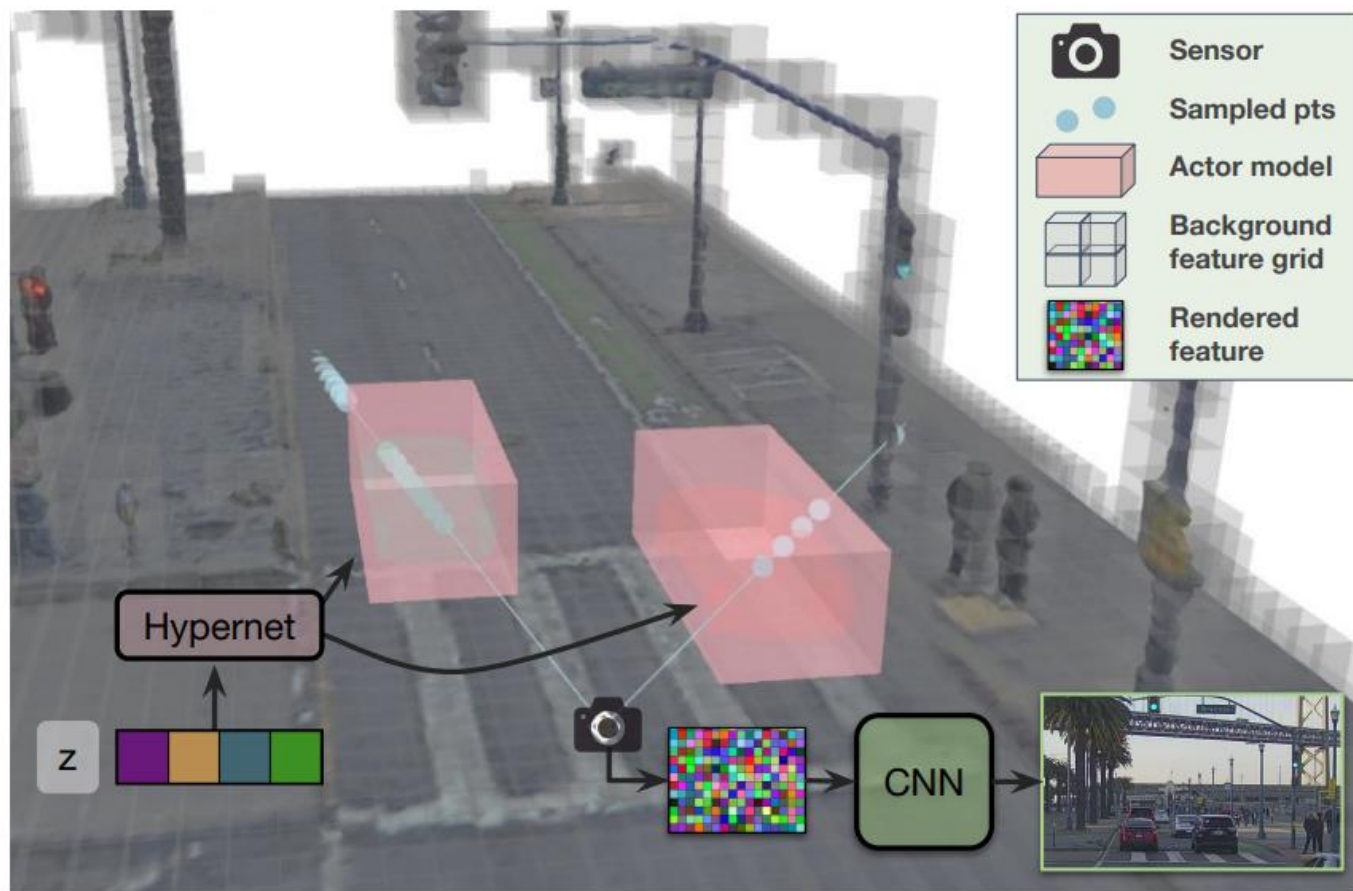


05. 仿真系统-提升测试迭代效率

- 通过对场景进行4D重建、4D标注，获取结构化的真值数据，快速用于单环节/端到端算法模块测试



05. 仿真系统-基于NeRF的自动驾驶仿真

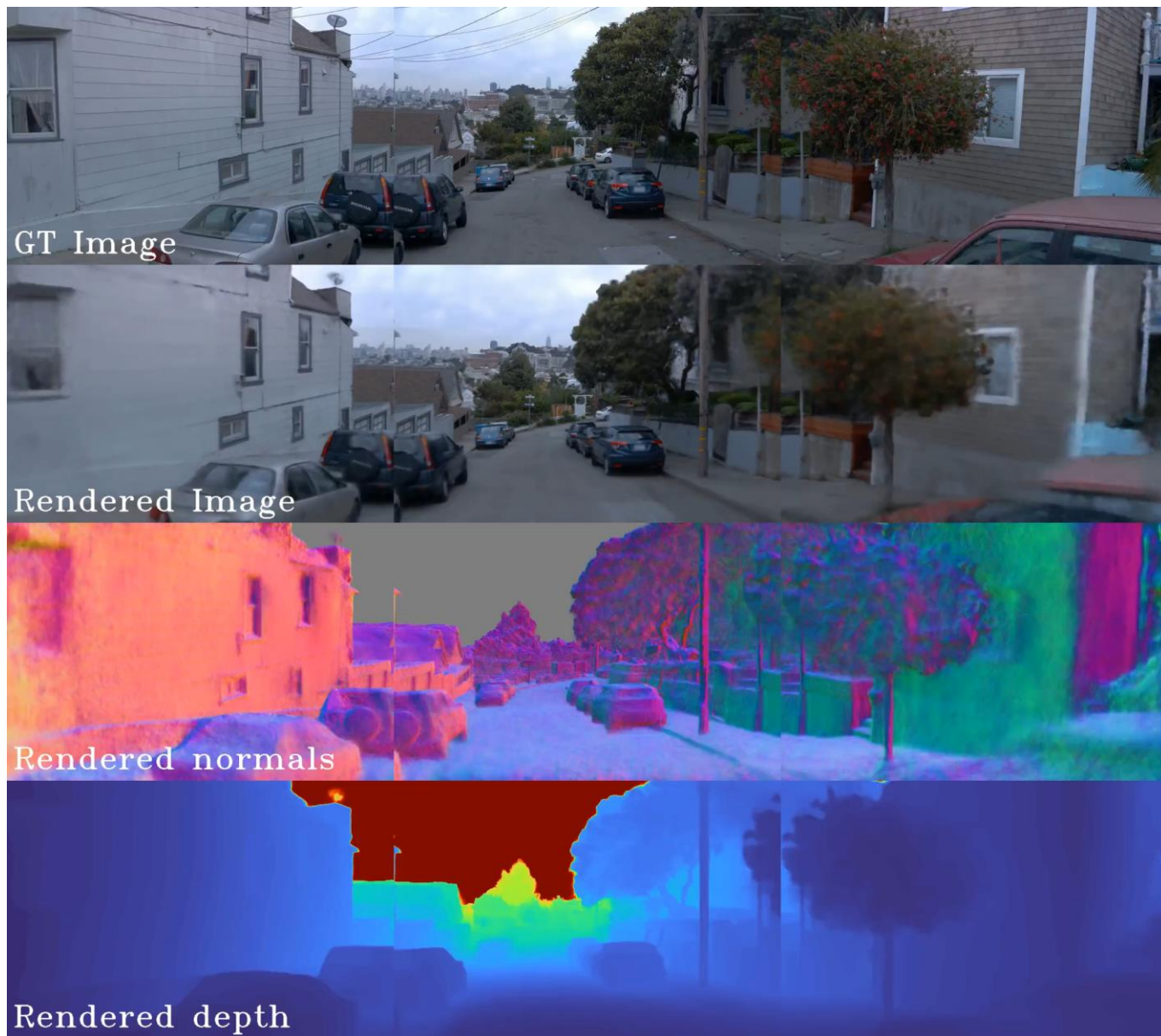


将NeRF技术应用到自动驾驶场景是一个复杂的系统工程，需要解决：

1. 采样视角单一且不够稠密
2. 不同相机光度一致性问题
3. 各种天气光照场景
4. 场景多样，包含远中近景
5. 动态物体

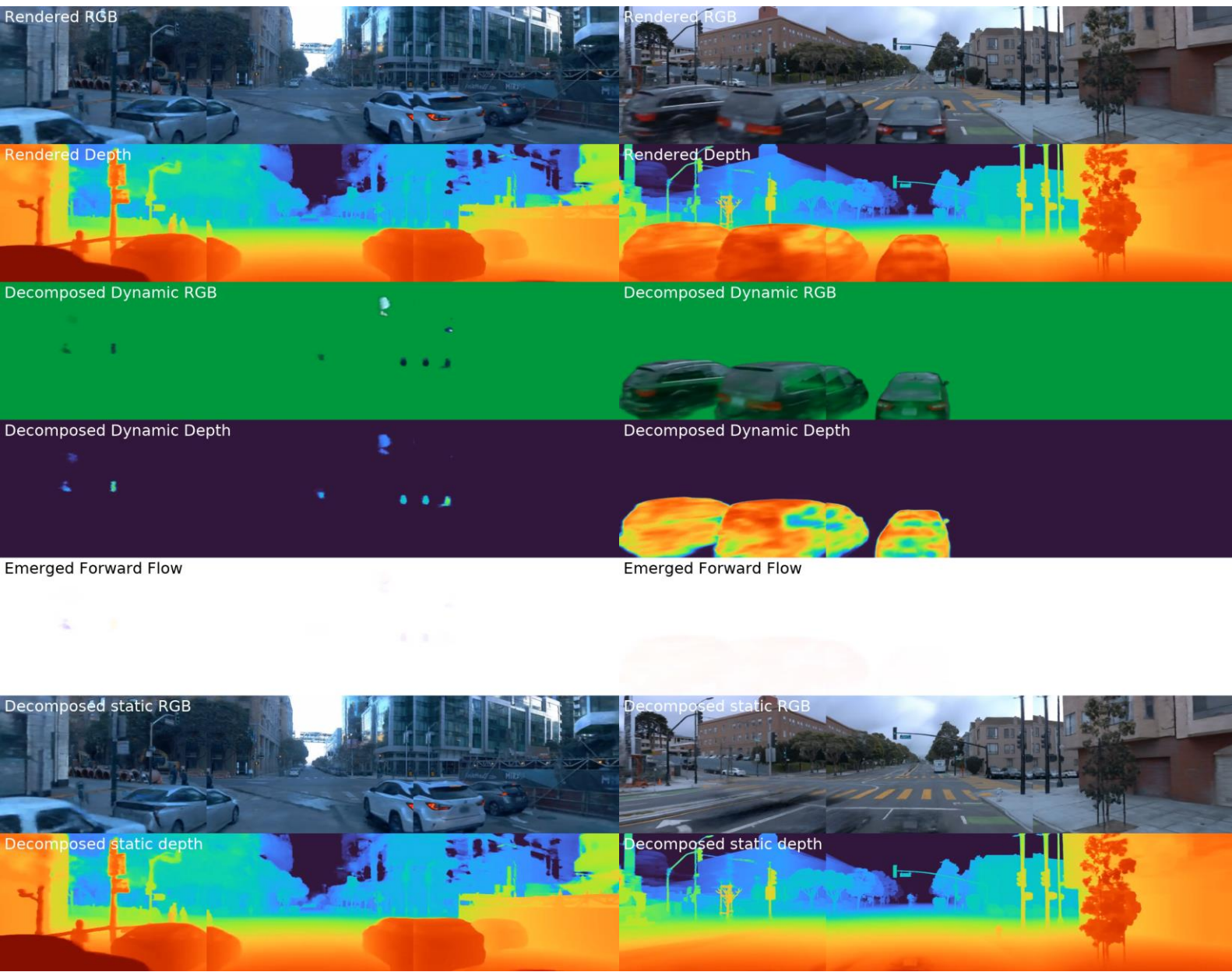
UniSim：对每一个动态物体进行单独建模

05. 仿真系统-基于NeRF的自动驾驶仿真



StreetSurf: 通过SDF(signed distant field)表达, 来添加更多的几何约束, 牺牲一点渲染精度的同时, 提高了NeRF在自动驾驶场景仿真重建的稳定性。

05. 仿真系统-基于NeRF的自动驾驶仿真



EmerNeRF 通过建模场景流
(sceneflow)

表达动态车辆，行人

05. 仿真系统-基于NeRF的自动驾驶仿真

基于NeRF技术的自动驾驶仿真技术优势

- 图像渲染真实，与原始数据接近
- 视角可以任意变化渲染
- 可以控制动态物体，天气，光照等

在自动驾驶场景的应用

- 通过新视角合成快速适应新的车型
- 快速合成大量特殊场景数据
- 端到端自动驾驶闭环模拟测试



06. 迈向端到端自动驾驶之路-视觉大模型



SORA



Tesla

06. 迈向端到端自动驾驶之路-视觉大模型

DriveDreamer: 驾驶场景可控视频生成



- 推理阶段: 文字+控制元素 生成视频
- 训练阶段: NeRF可以用来多视角一致性

06. 迈向端到端自动驾驶之路-Tesla V12.3



AI DRIVR ✓ There's a HUMAN inside Tesla FSD 12.3



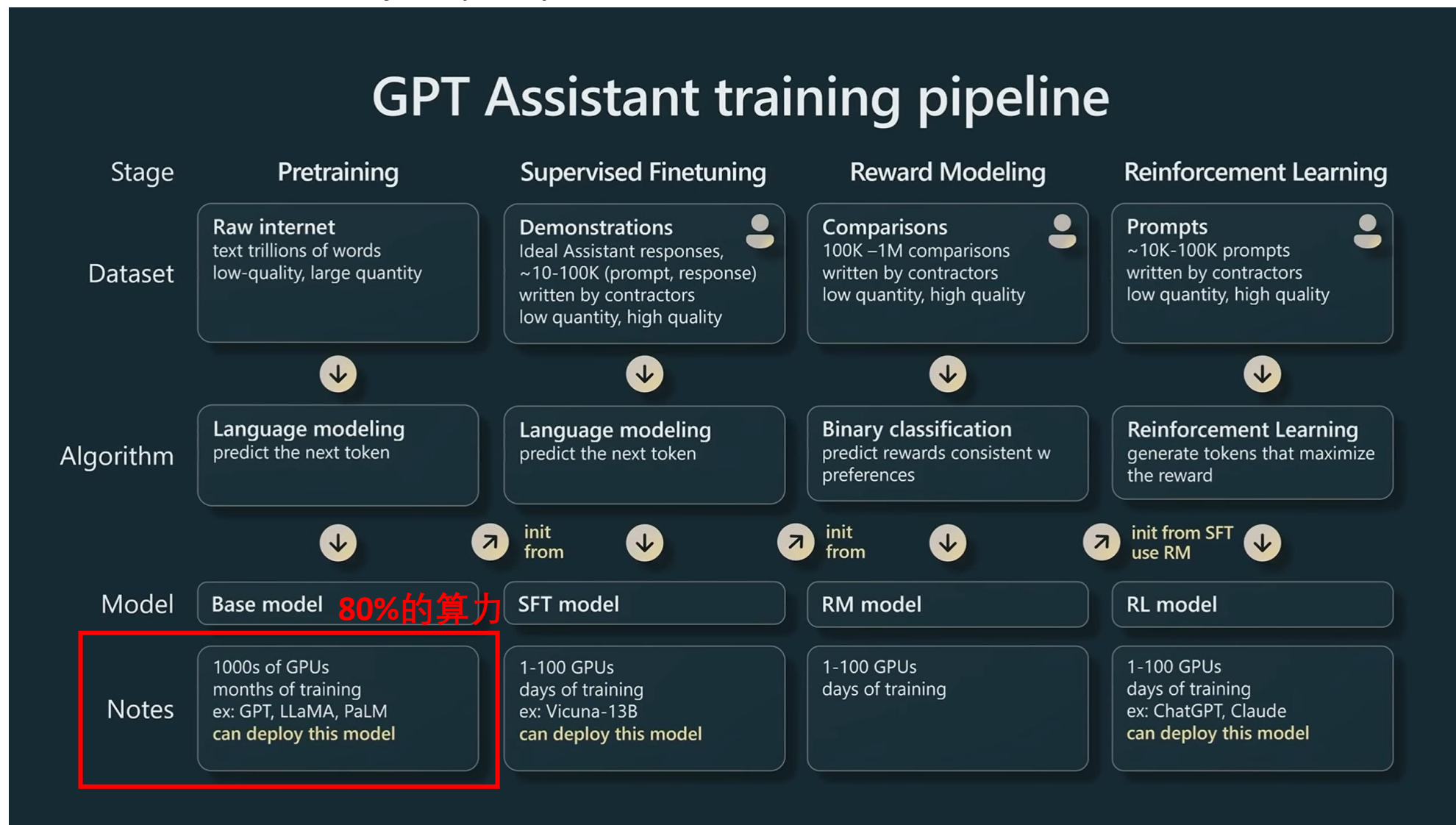
Ashok Elluswamy ✓ @aelluswamy · 14小时
FSD respects chicken that wanna cross the road



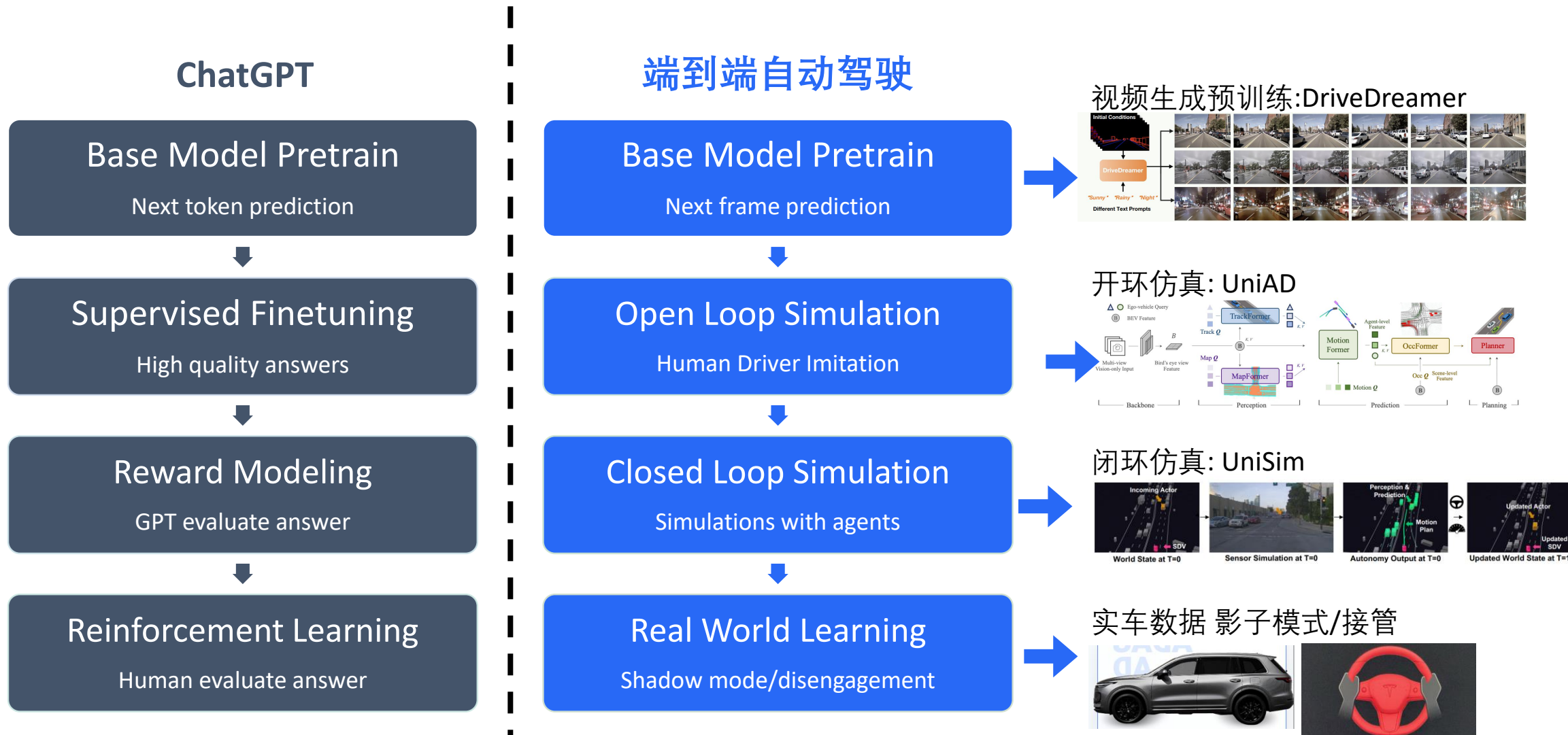
- 真实世界无法用规则化的代码穷举
- 高级别自动驾驶和AGI一样难

06. 迈向端到端自动驾驶之路-from GPT to driving

“State of GPT” Andrej Karpathy on Microsoft Build 2023 event.



06. 迈向端到端自动驾驶之路-from GPT to driving



07. 一点心灵鸡汤



Andrej Karpathy ✓ @karpathy · 2020年6月19日

I actually did spend about an hour playing around with the OpenAI GPT-3 API over the weekend trying to get the davinci model to recommend me the most interesting movies (e.g. asking it for more rare movies, etc). Little did I know I was 3.0 programming.

1

6

86



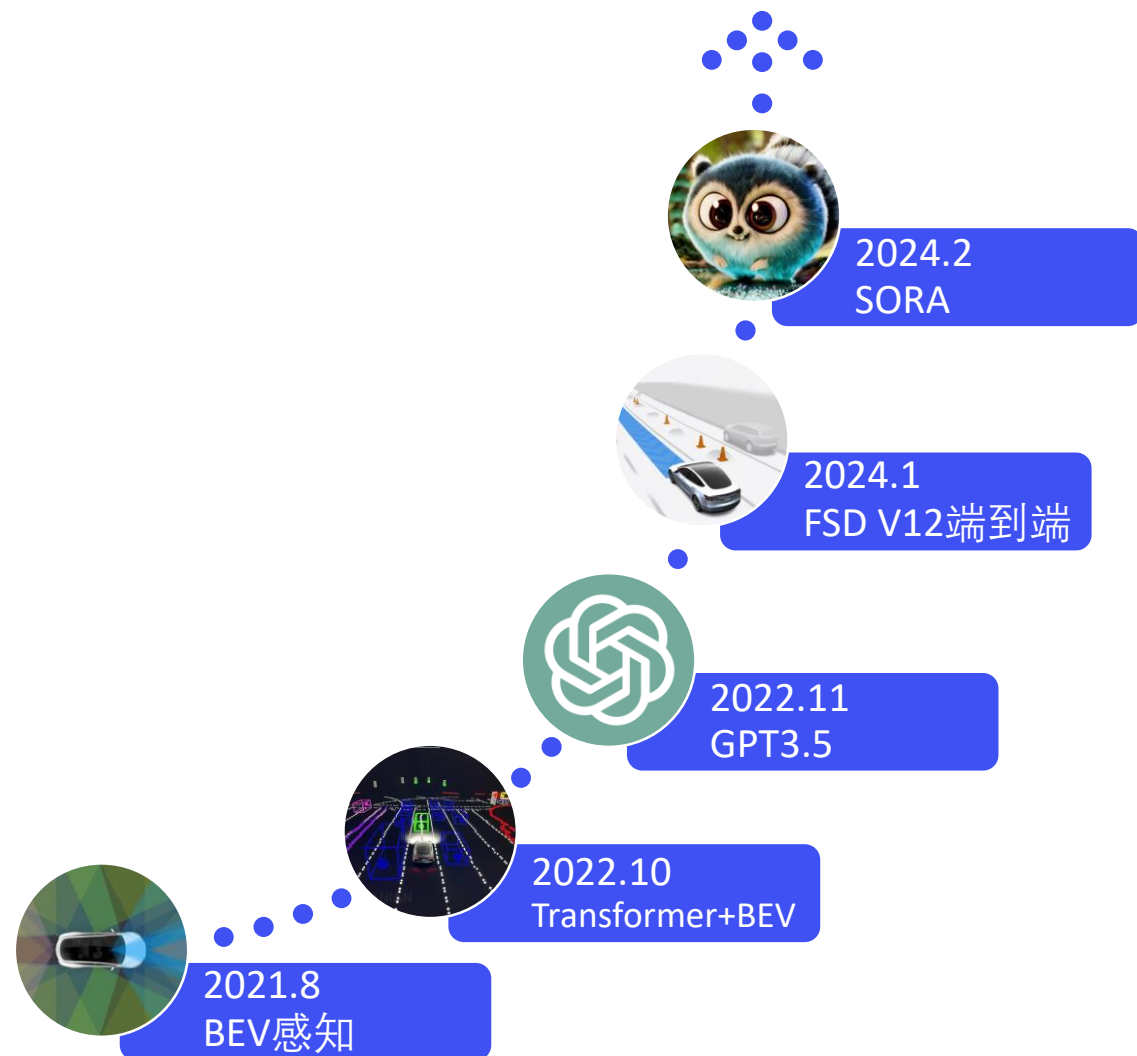
Andrej Karpathy ✓
@karpathy

I guess the ultimate Autopilot is giving a big multimodal GPT-10 the DMV handbook in the prompt concatenated with the last 10 seconds of sensor data and asking it very nicely to follow that 🤔

[翻译帖子](#)

上午9:34 2020年6月19日

仰望星空，脚踏实地



Thanks
&
QA

