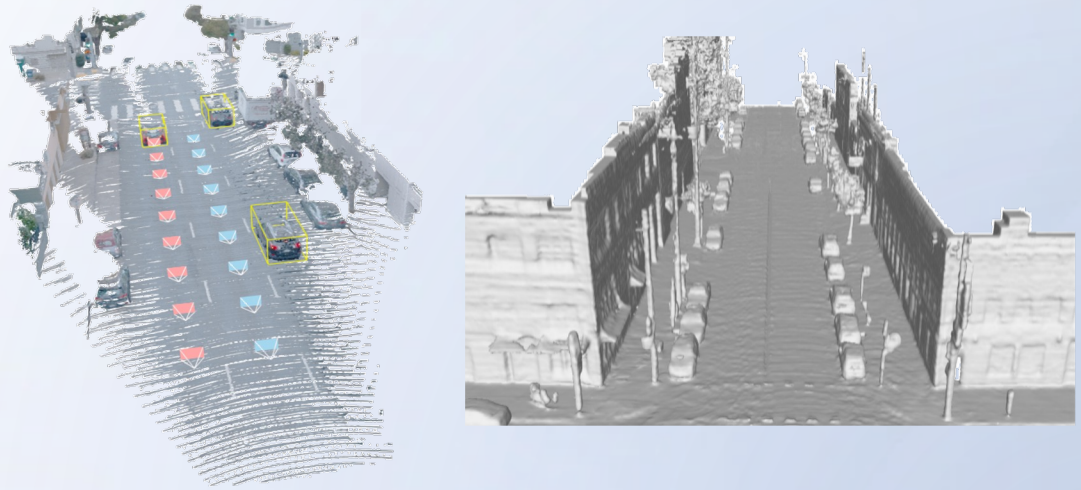
面向自动驾驶仿真的动态街景重建技术

彭思达
浙江大学CAD&CG全国重点实验室

研究目标

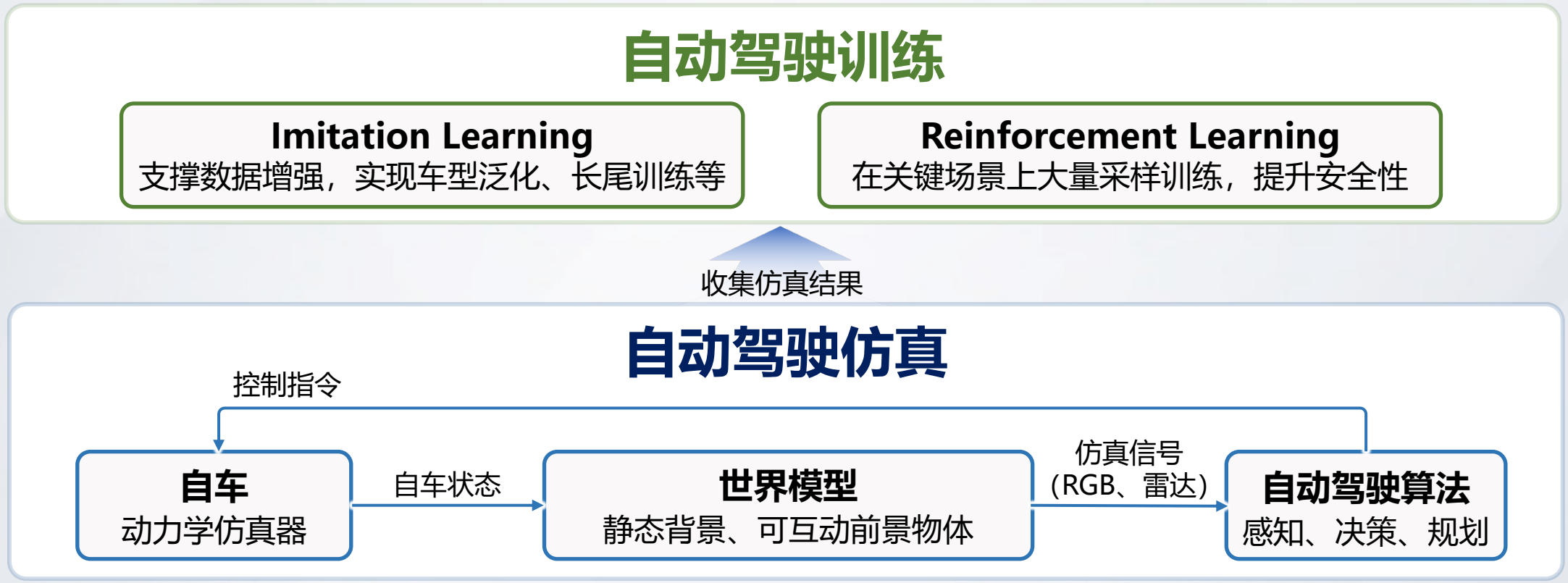
基于路采数据的动态街景高精建模与重渲染

路采数据

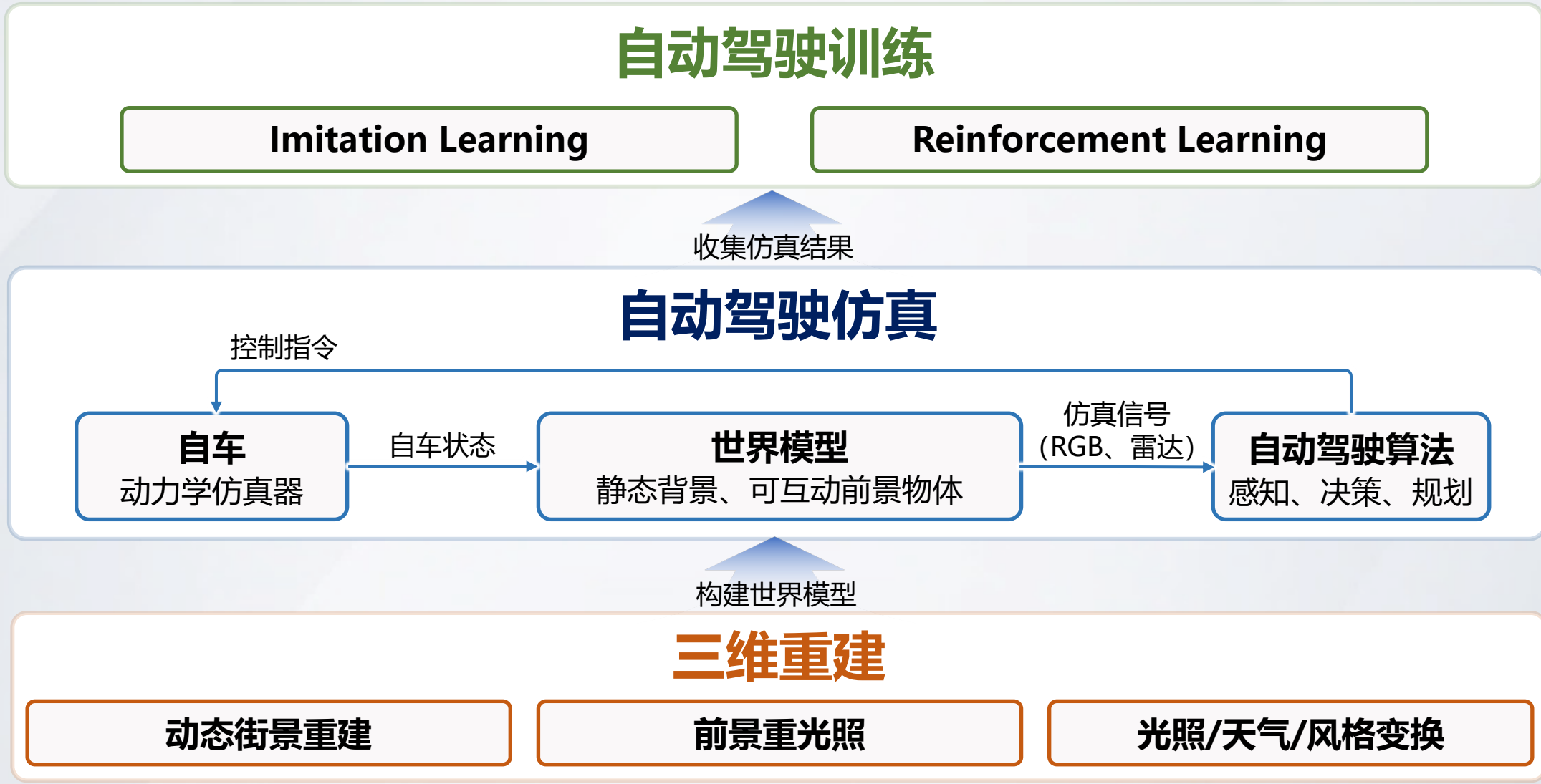


动态三维街景

自动驾驶仿真的价值

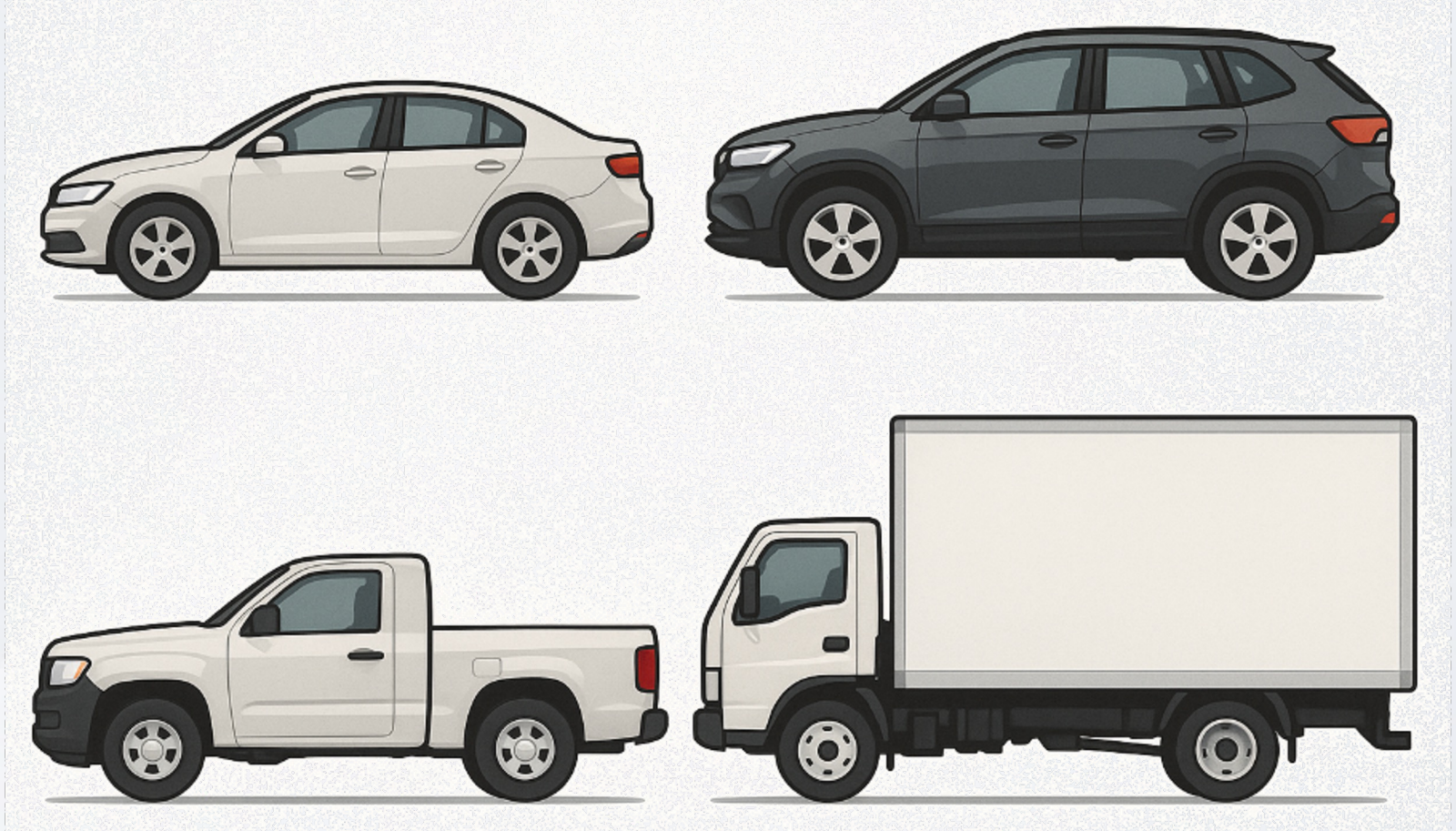

自动驾驶训练
Imitation Learning
支撑数据增强，实现车型泛化、长尾训练等
Reinforcement Learning
在关键场景上大量采样训练，提升安全性
收集仿真结果
自动驾驶仿真
控制指令
自车
动力学仿真器
自车状态
世界模型
静态背景、可互动前景物体
仿真信号
(RGB、雷达)
自动驾驶算法
感知、决策、规划

三维重建在自动驾驶仿真中的位置



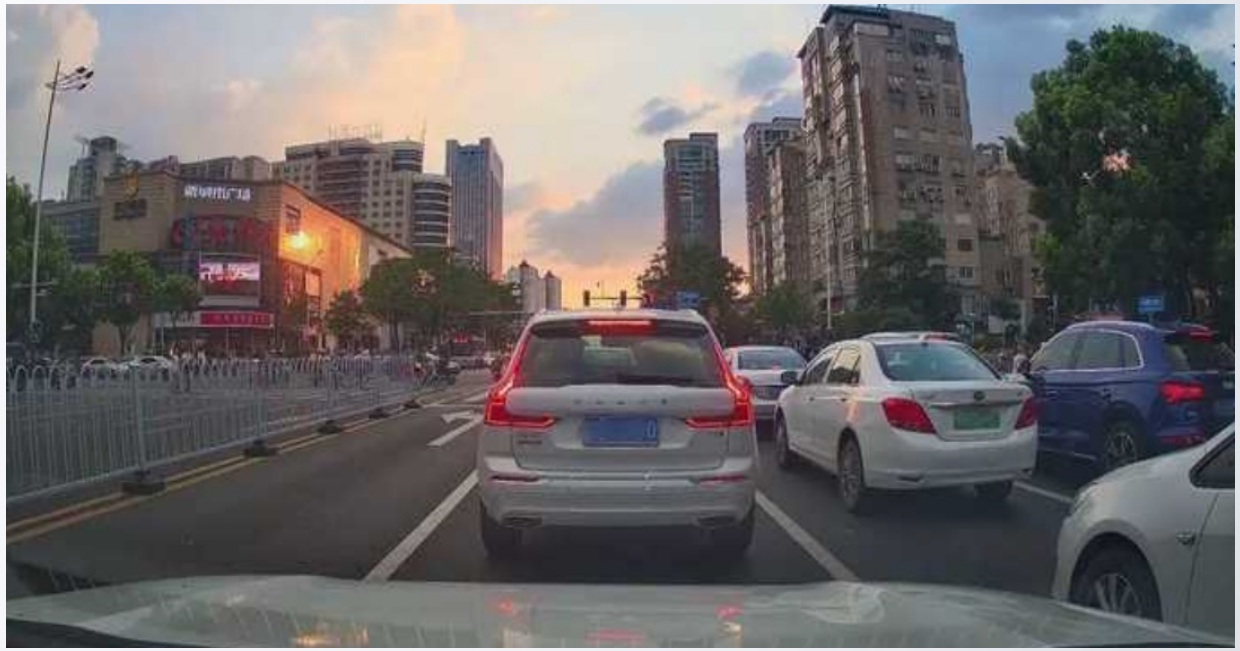
对于自动驾驶的价值：短期

- 车厂面临的问题：在小轿车训练的感知模型，无法泛化到其他车型。
- 目标：采集一次数据，实现多种车型的训练。



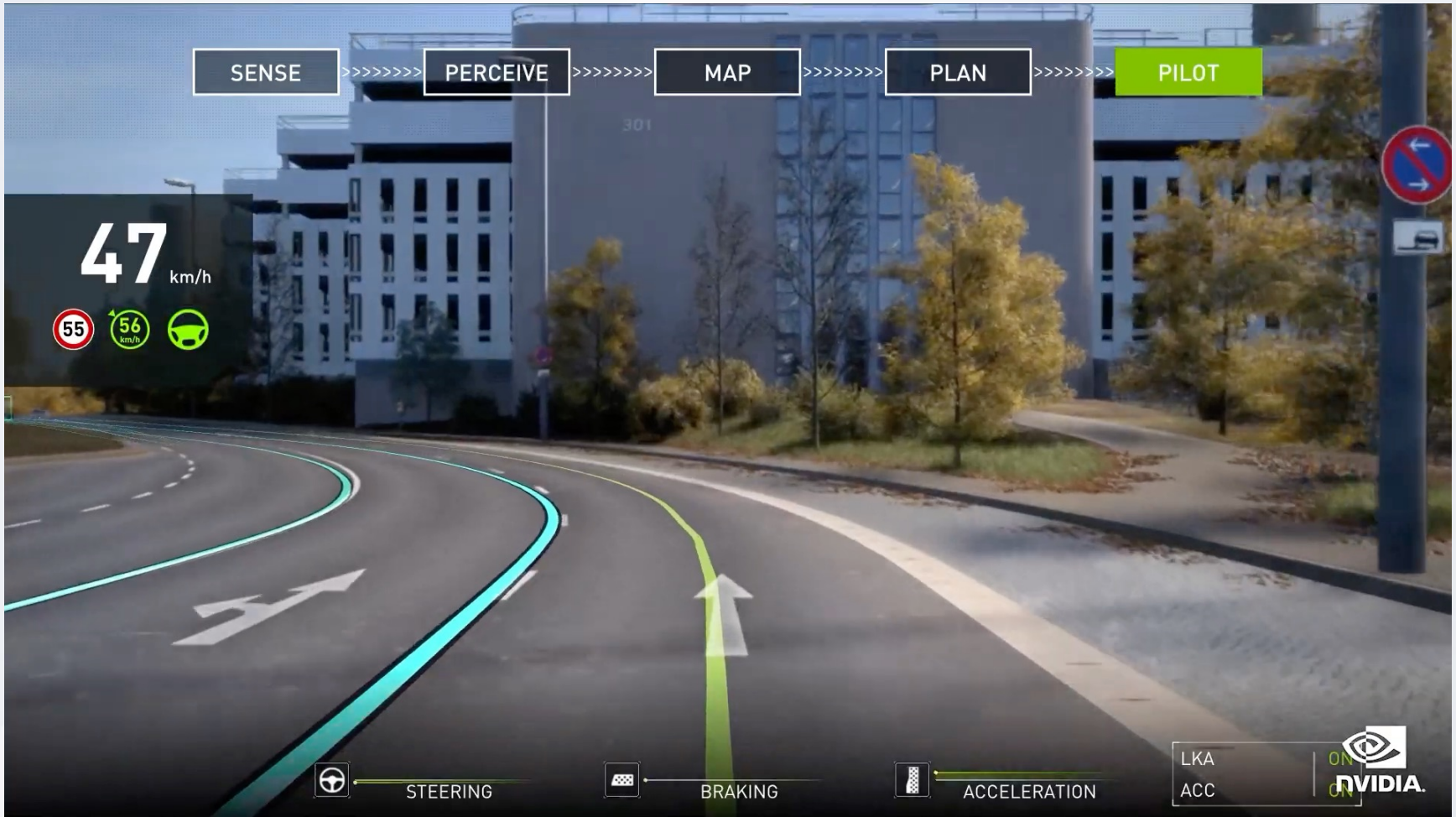
对于自动驾驶的价值：中期

- 车厂面临的问题：长尾数据难以收集。
- 目标：通过行车记录仪采集的视频构建自动驾驶训练数据。



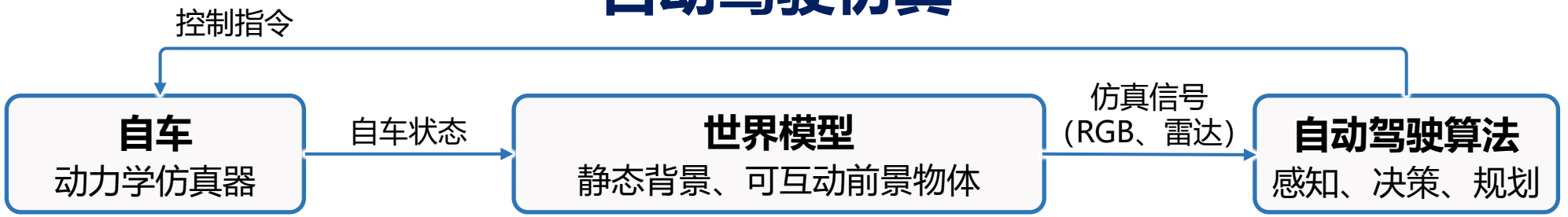
对于自动驾驶的价值：长期

- 车厂面临的问题：真实环境下训练和测试成本高。
- 目标：构建虚拟环境训练和测试自动驾驶系统。



自动驾驶仿真中的动态街景重建需要哪些能力

自动驾驶仿真

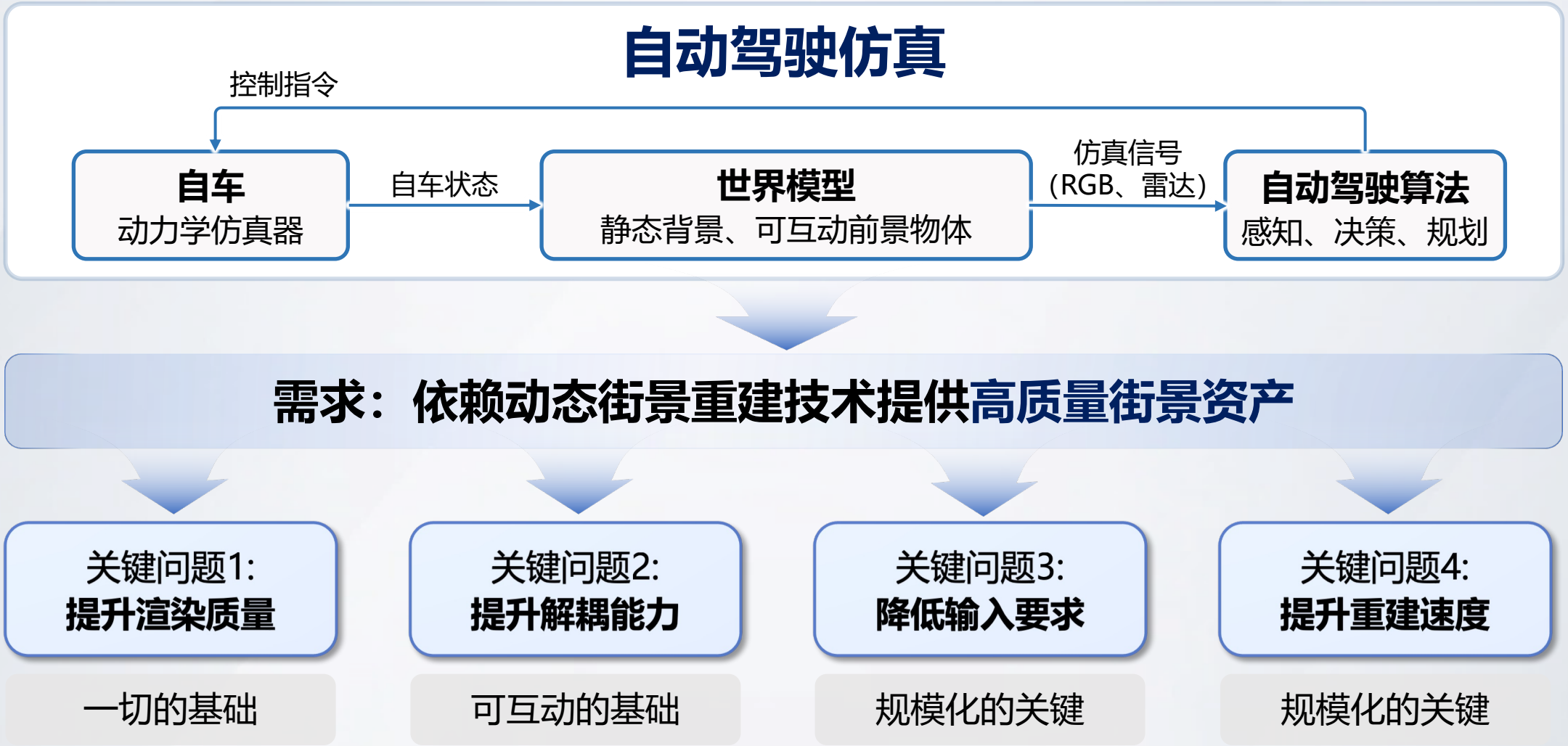


需求：依赖动态街景重建技术提供高质量街景资产

动态街景重建

- **静态背景**
- **可互动前景物体**

自动驾驶仿真中的动态街景重建需要哪些能力



实验室已有研究积累

自动驾驶仿真

需求：依赖动态街景重建技术提供高质量街景资产

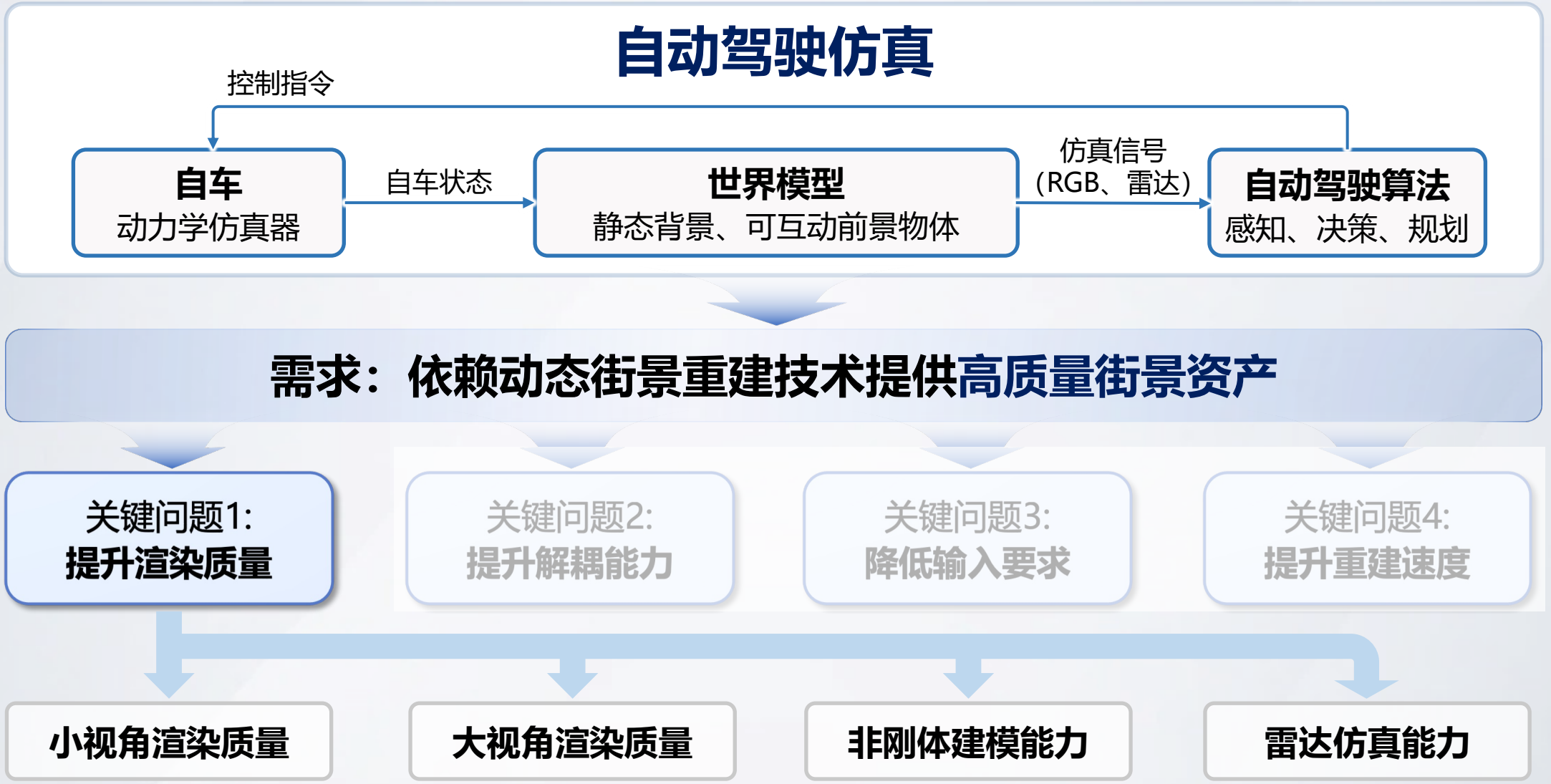
关键问题1: **提升渲染质量**	关键问题2: **提升解耦能力**	关键问题3: **降低输入要求**	关键问题4: **提升重建速度**
相关工作: **StreetGaussians、StreetCrafter、FreeTimeGS等**	相关工作: **StreetGaussians 、Split4D**	相关工作: **MatchAnything、D-SfM、Murre等**	相关工作: **PromptDA、ADGaussian**

论文发表：在CVPR、ICCV、ECCV等发表多篇Oral、Highlight论文，谷歌引用量数百次。

代码开源：论文GitHub stars累积数千次，数次被集成进知名算法库kornia和transformers。

竞赛获奖：获得谷歌举办的全球三维重建挑战赛冠军。

关键问题1: 提升仿真质量



关键问题1.1: 小视角渲染质量

原先方法存在的问题：小视角渲染质量低

Block-NeRF: Scalable Large Scene Neural View Synthesis. CVPR 2022.

关键问题1.1: 小视角渲染质量

为什么这个问题重要

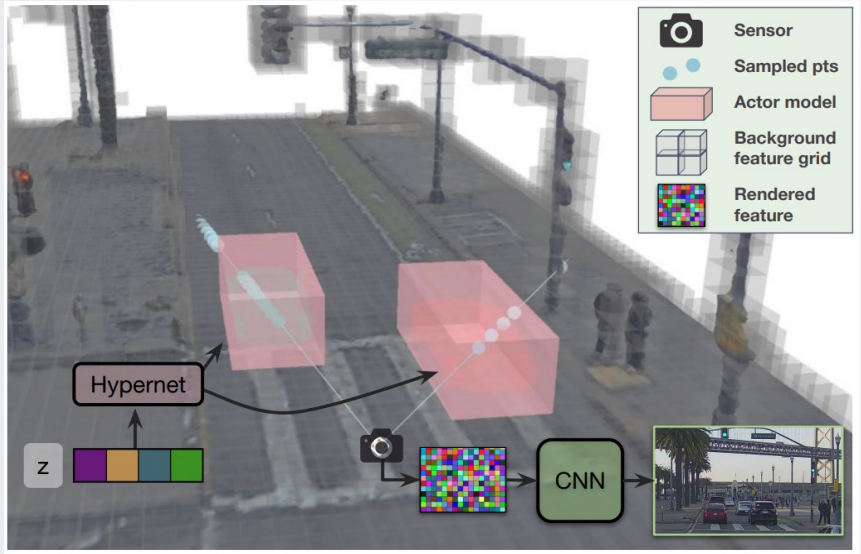
- 小视角可以满足自动驾驶仿真的短期价值 -> 车型泛化能力。
- 小视角质量先能做好，才有机会搞大视角渲染。

关键问题1.1: 小视角渲染质量

已有技术范式

范式1:
基于Feature Grids的场景表征

代表性工作：UniSim



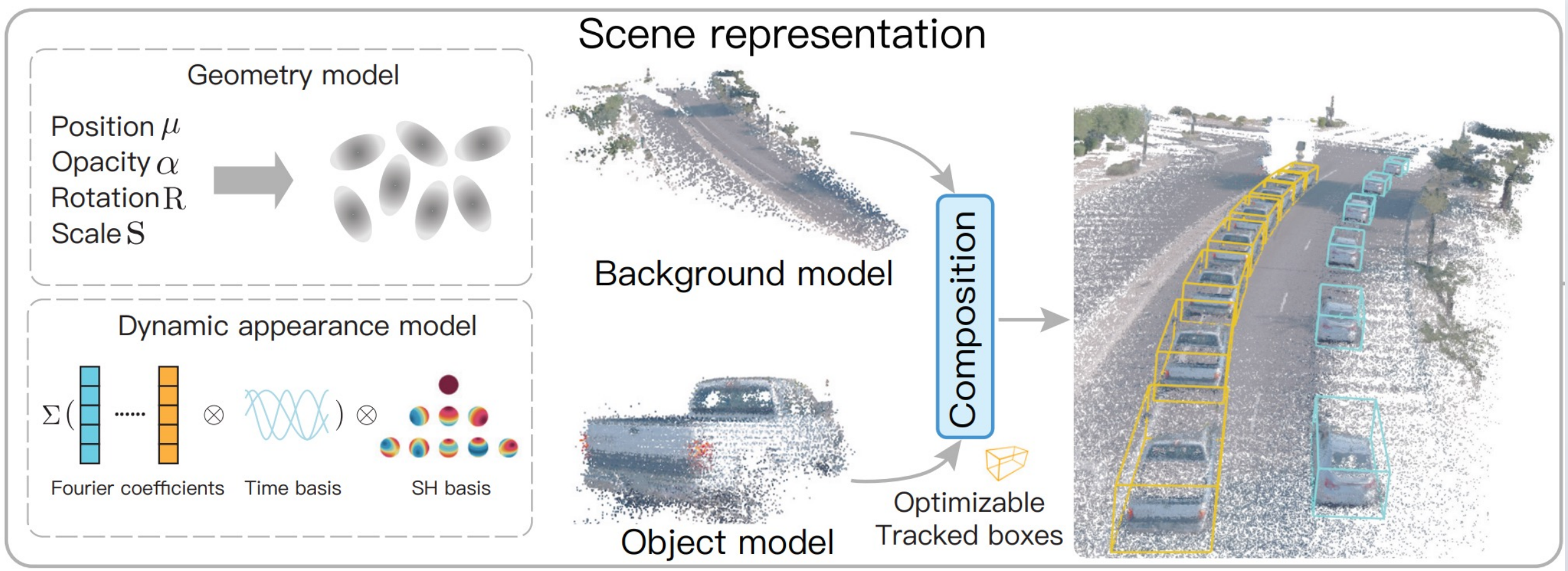
范式2:
基于3D Gaussians的场景表征

代表性工作：Street Gaussians

实验室成果

Street Gaussians（实验室成果）

- 基于**组合式的三维高斯**表示动态三维场景



Street Gaussians: Modeling Dynamic Urban Scenes with Gaussian Splatting. ECCV 2024.

与EmerNeRF的可视化比较

更多的渲染结果






关键问题1.2: 大视角渲染质量

原先方法存在的问题

相邻视角

变道视角

Street Gaussians: Modeling Dynamic Urban Scenes with Gaussian Splatting. ECCV 2024.

关键问题1.2: 大视角渲染质量

为什么这个问题重要

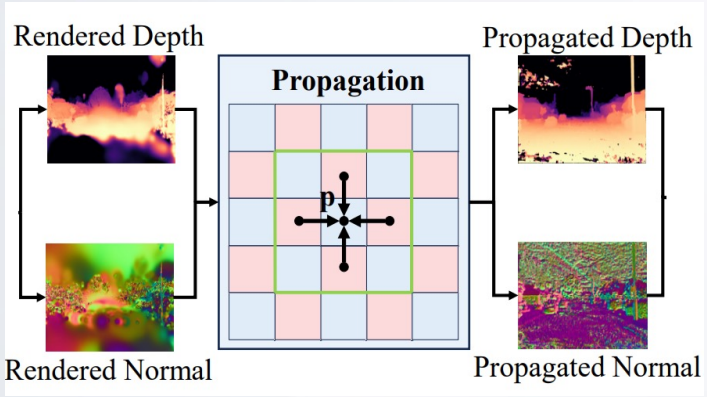
- 大视角渲染对自动驾驶仿真的短、中、长期目标都很有价值。
- 大视角渲染是闭环仿真的必要条件。

关键问题1.2: 大视角渲染质量

已有技术范式

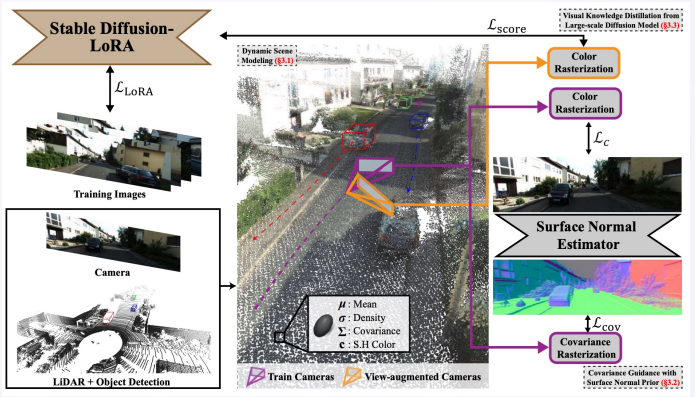
范式1:
基于逐场景优化

代表性工作：GaussianPro



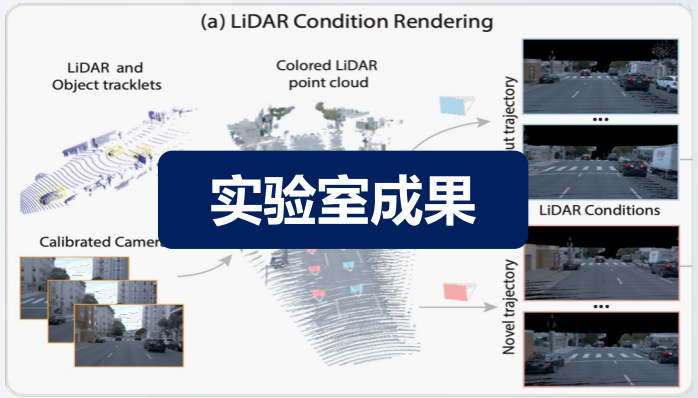
范式2:
引入SDS Loss约束

代表性工作：VEGS



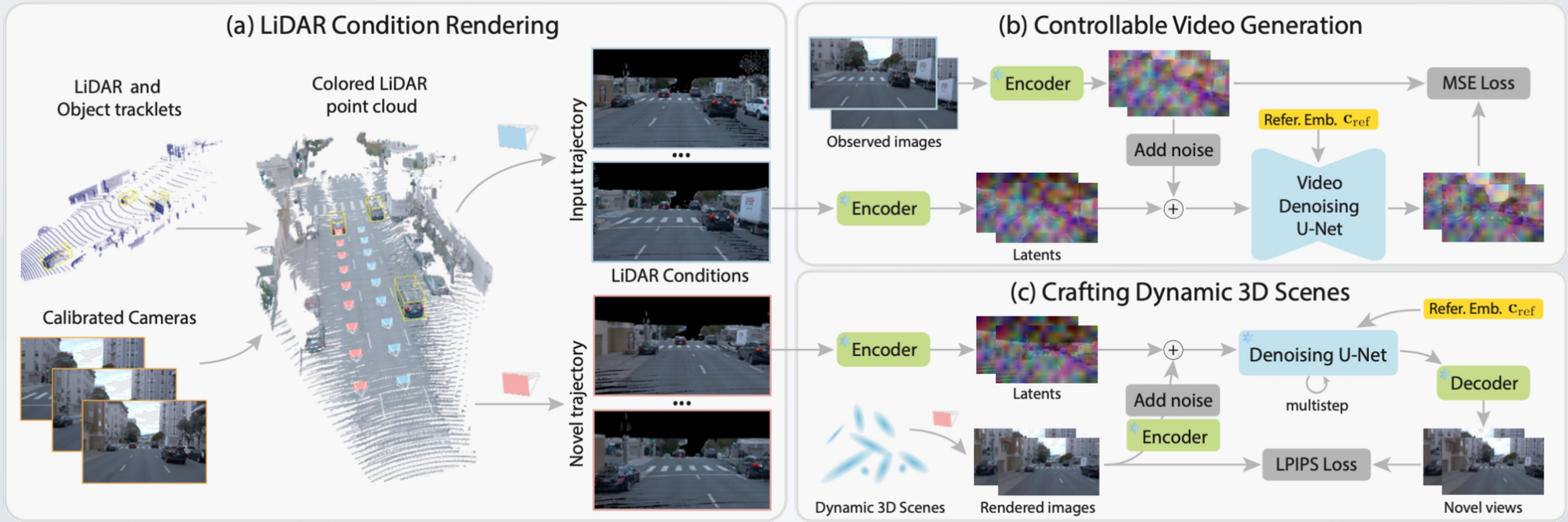
范式3:
Video model生成新视角

代表性工作：StreetCrafter



StreetCrafter （实验室成果）

利用雷达点云控制视频生成模型输出新视角的图片，作为StreetGS的监督。



StreetCrafter: Street View Synthesis with Controllable Video Diffusion Models. CVPR 2025.

实验结果：变车道新视角渲染

Ours-G

Street Gaussians

实验结果：变车道新视角渲染

Ours-G

Street Gaussians

关键问题1.3: 非刚性建模能力

原先方法存在的问题：无法高质量表征非刚性运动物体。

Real-time Photorealistic Dynamic Scene Representation and Rendering with 4D Gaussian Splatting. ICLR 2024.

关键问题1.3: 非刚性建模能力

为什么这个问题重要

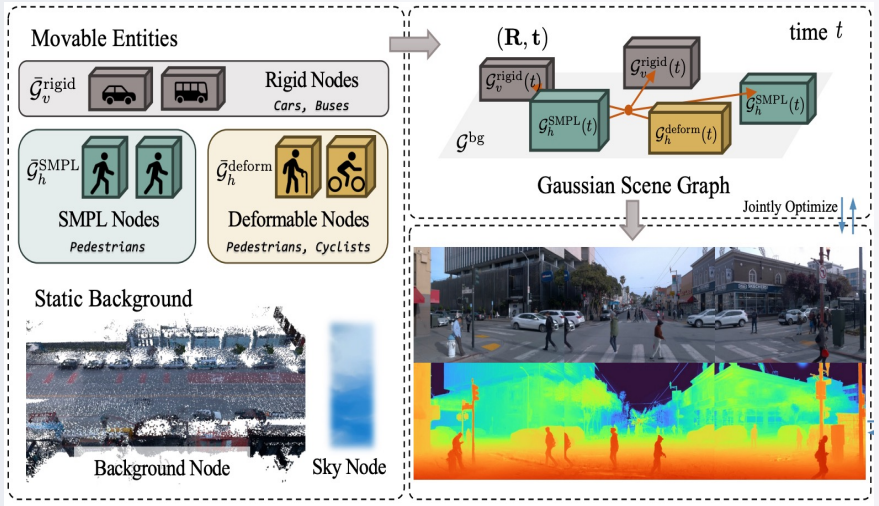
- 日常街道场景中，除了汽车，还有很多非刚性物体：行人、骑行人、动态灯光等。
- 扩展场景的丰富性必须建模好这些非刚性物体。

关键问题1.3: 非刚性建模能力

已有技术范式

范式1:
基于Neural Scene Graph

代表性工作：OmniRe



范式2:
基于native 4DGS

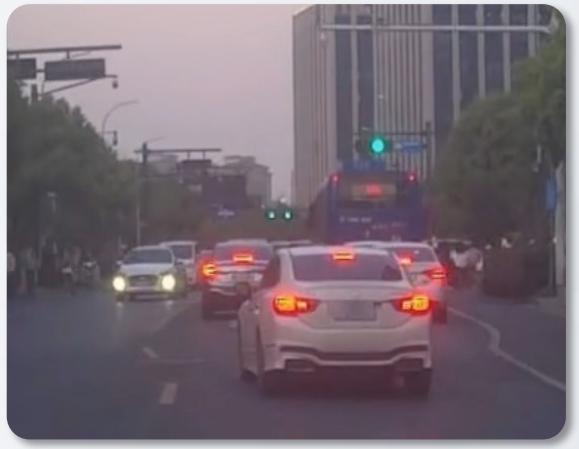
代表性工作：FreeTimeGS

FreeTimeGS （实验室成果）

不再区分前后背景，使用四维高斯建模整个动态场景。

静态背景

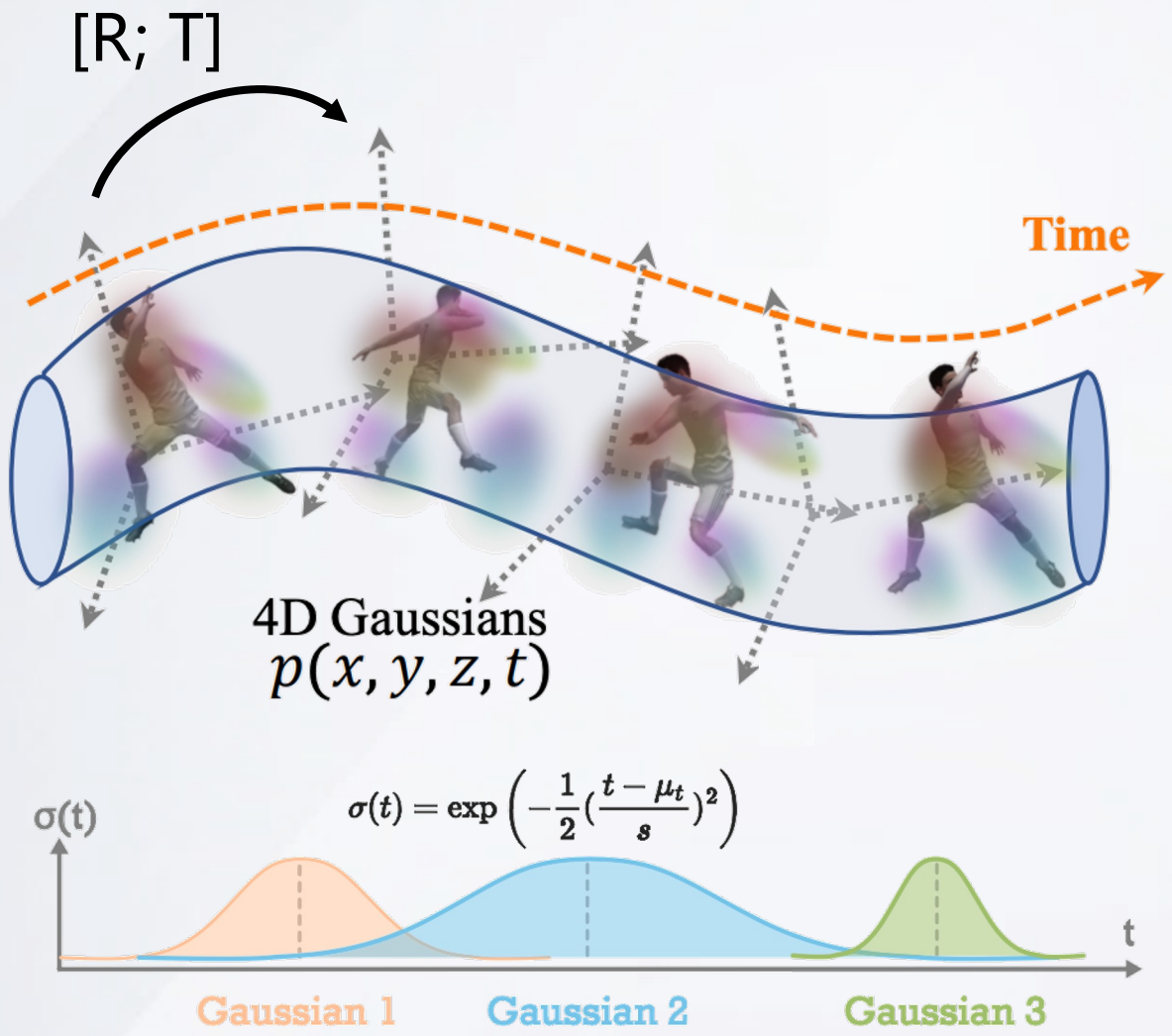
动态前景



动态光照

基于四维高斯的统一建模

FreeTimeGS：基于四维高斯的建模动态场景



三维高斯

+

随时间的运动

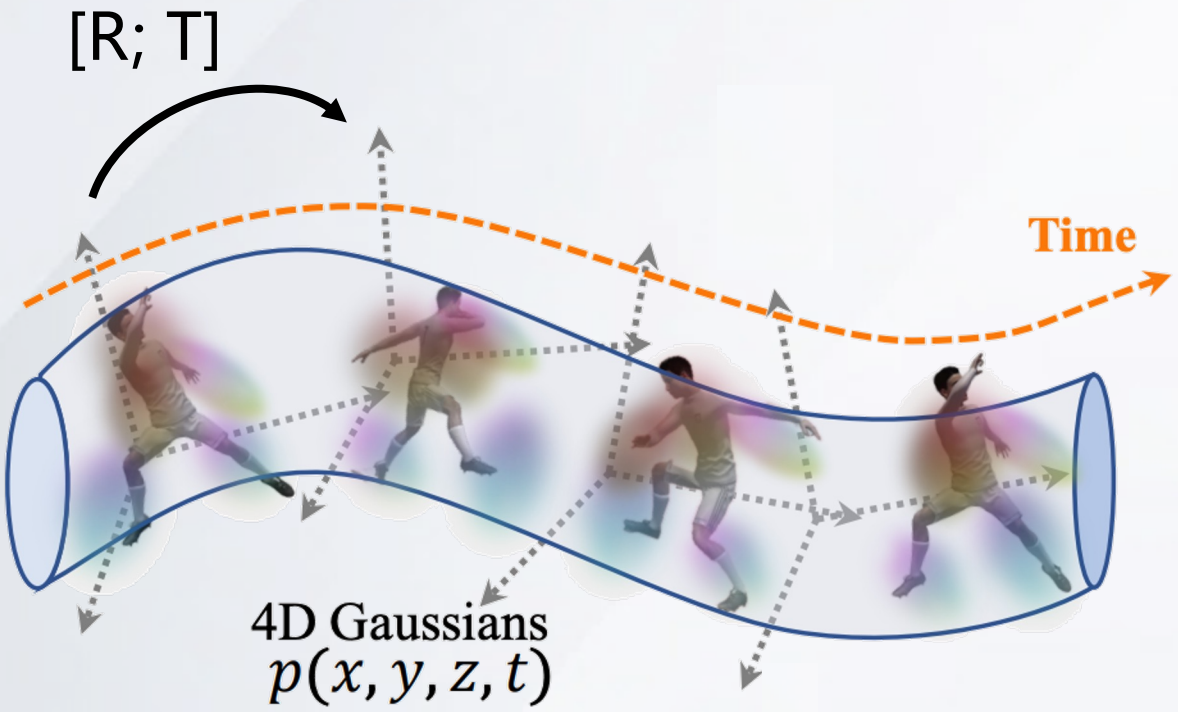
+

时序窗口

4D-Rotor Gaussian Splatting: Towards Efficient Novel View Synthesis for Dynamic Scenes. SIGGRAPH 2024.

遇到的难点：如何建模快速运动的物体

已有四维高斯使用[R; T]建模motion，难以优化，导致复杂运动物体效果不好。

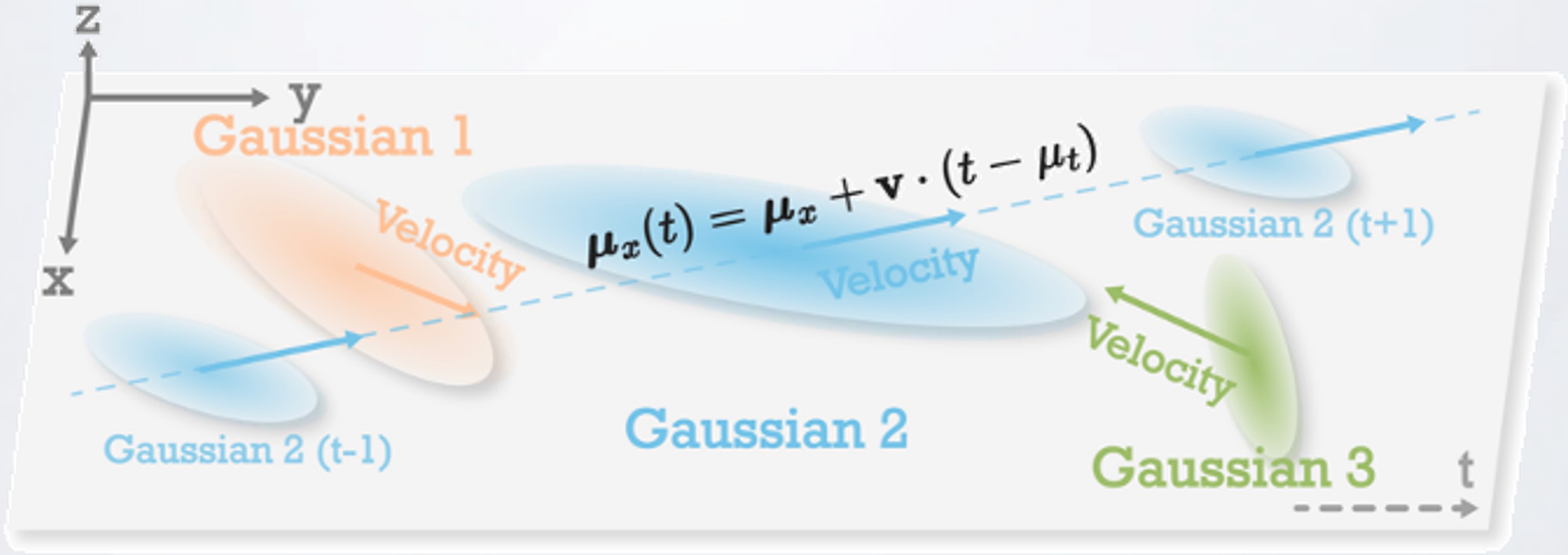


Real-time Photorealistic Dynamic Scene Representation and Rendering with 4D Gaussian Splatting. ICLR 2024.

解决思路：基于线性运动的四维高斯

在任意时刻任意位置定义高斯，为每个高斯赋予线性运动，用4DSH建模外观。

实验发现，FreeTimeGS可以很好地建模复杂的运动物体。



实验结果

实验结果

路边的人群

变换的灯光

实验结果

关键问题1.4: 雷达仿真能力

为什么这个问题重要

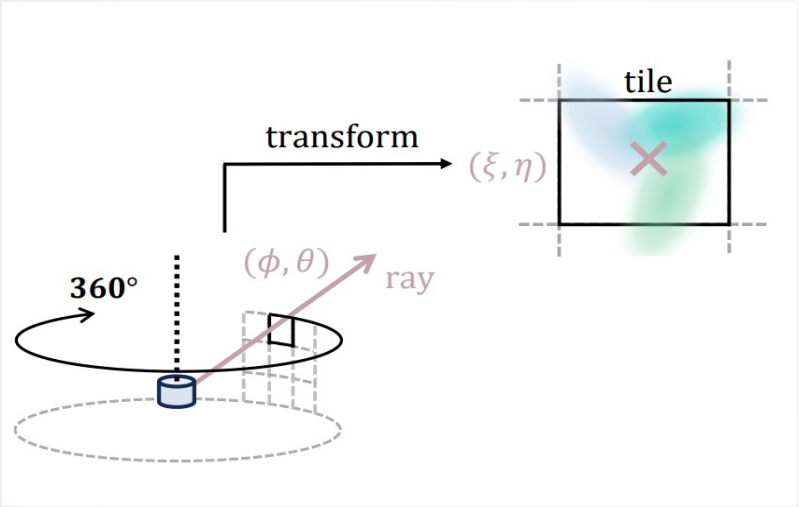
- 一些自动驾驶算法的输入包括雷达。
- 为了在闭环仿真器中训练这些自驾算法，必须实现雷达仿真。

关键问题1.4: 雷达仿真能力

已有技术范式

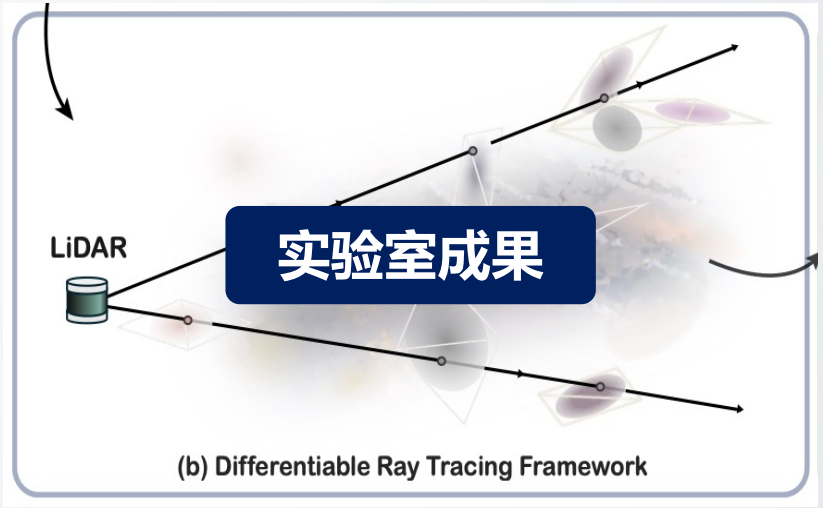
范式1:
基于Rasterization Rendering

代表性工作：GS-LiDAR



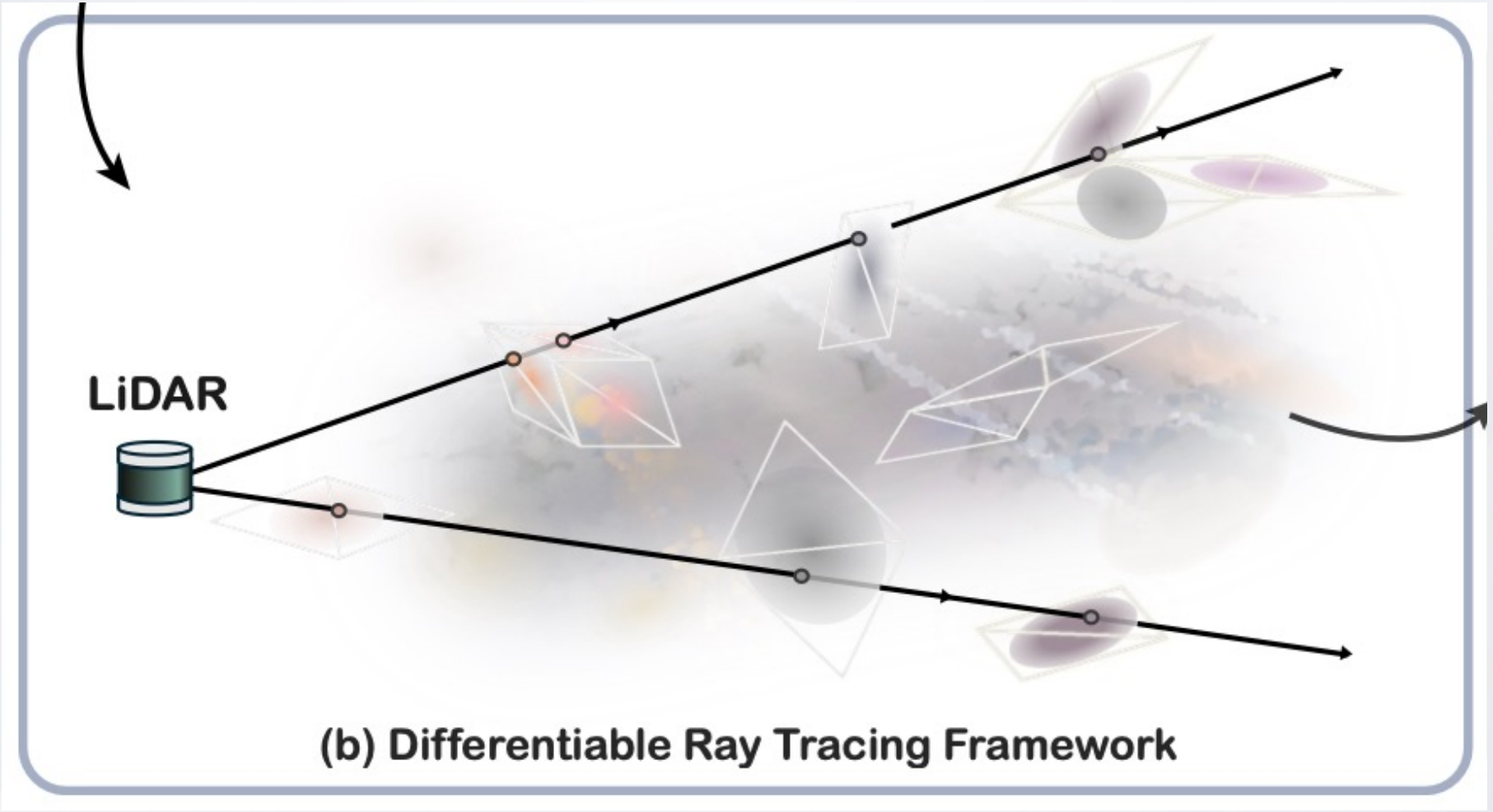
范式2:
基于Ray Tracing Rendering

代表性工作：LiDAR-RT



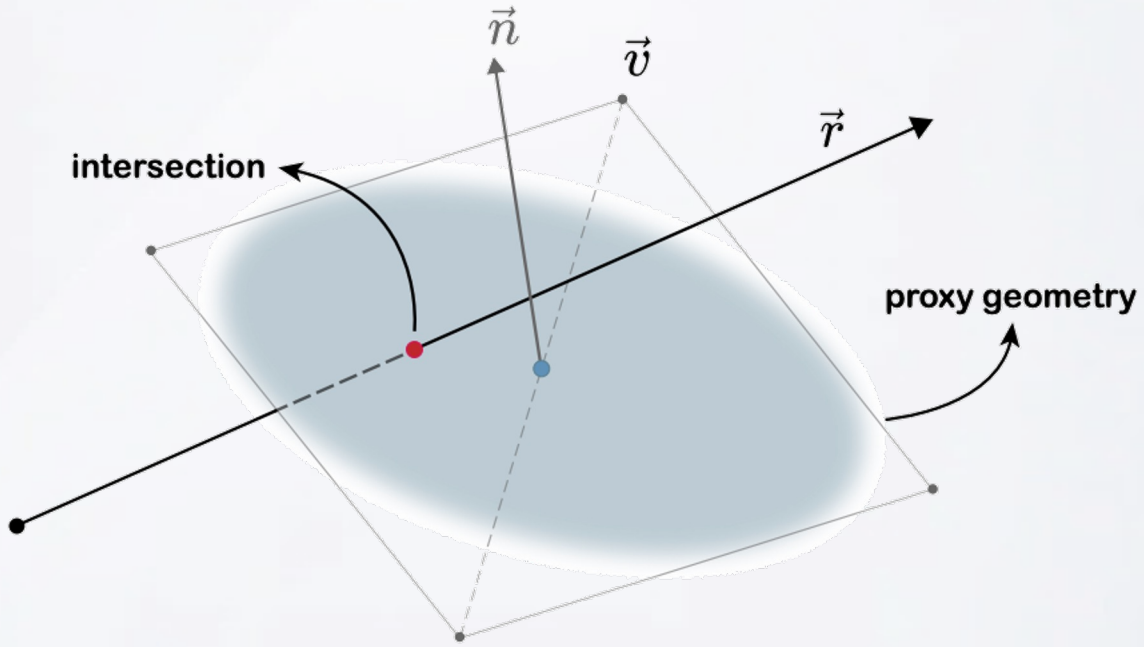
LiDAR-RT （实验室成果）

- Ray Tracing可以支持雷达柱面成像方式

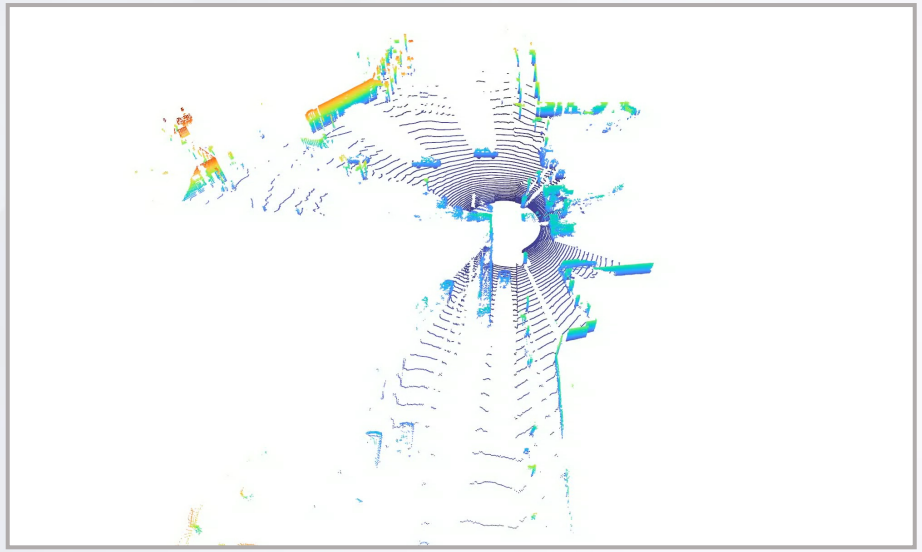


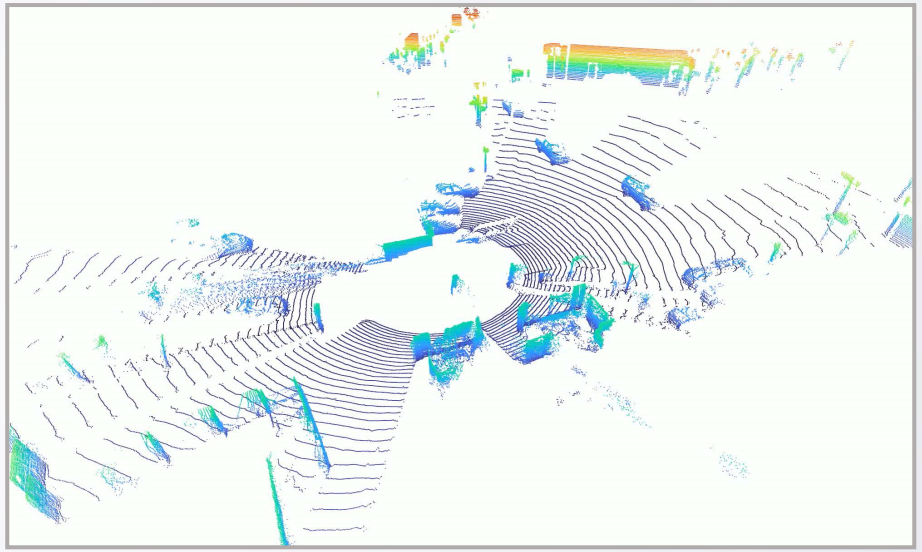
如何实现Ray Tracing

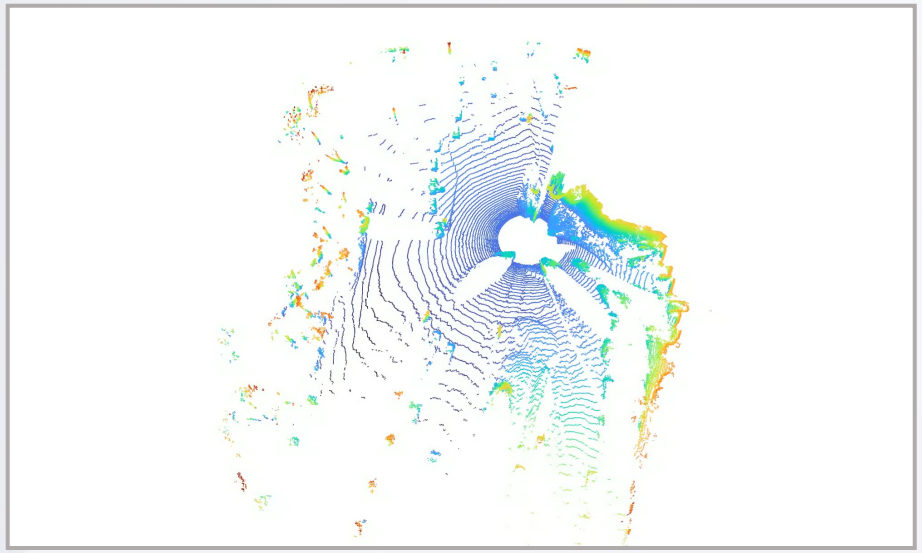
1. 构建加速结构以计算Ray Interaction。
2. 对于Intersection points，计算GS的值，再做Volume Rendering。

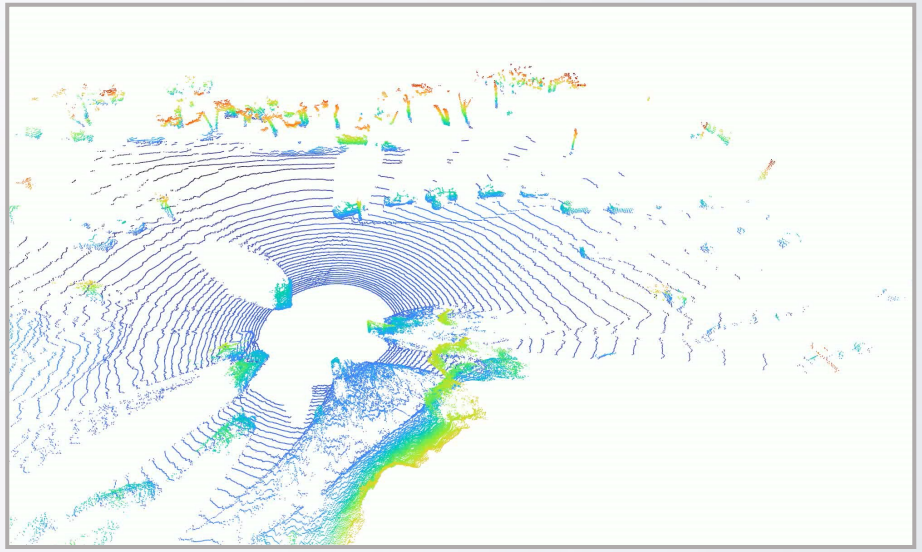


Novel view LiDAR point clouds



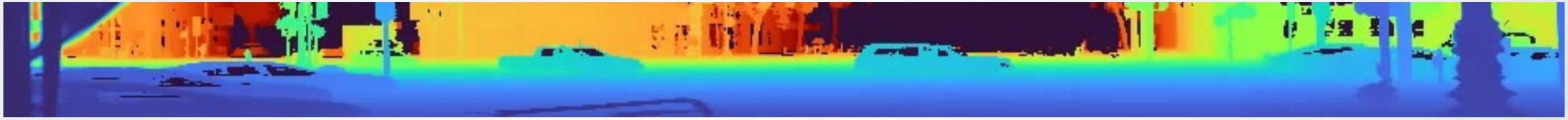






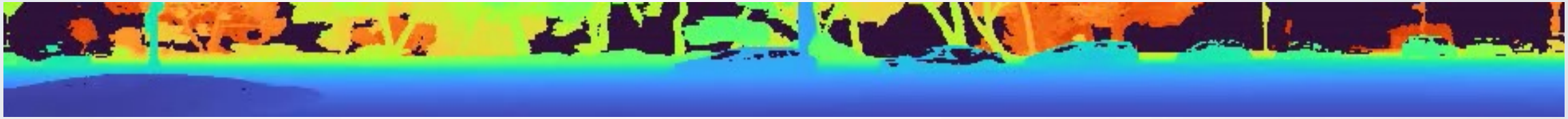
Novel view LiDAR range images

Depth



Intensity

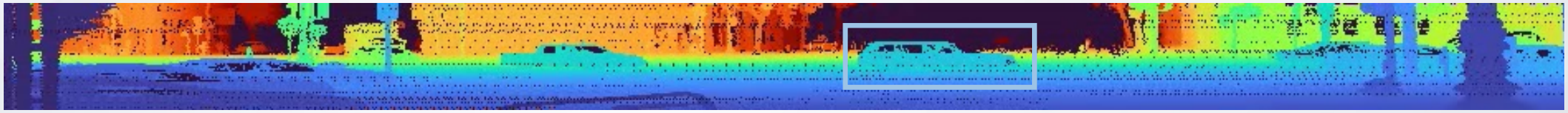
Depth



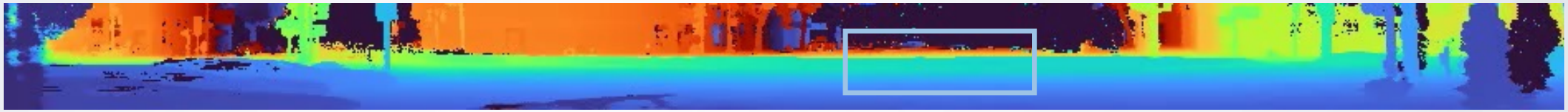
Intensity

Qualitative comparisons on Waymo dataset

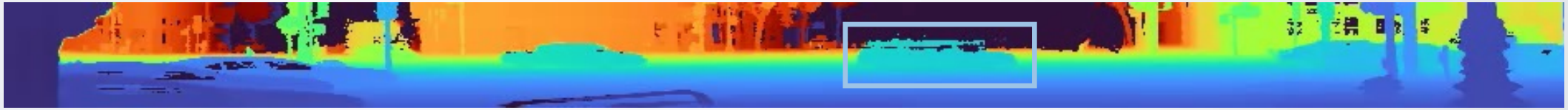
Ground Truth



LiDAR-NeRF

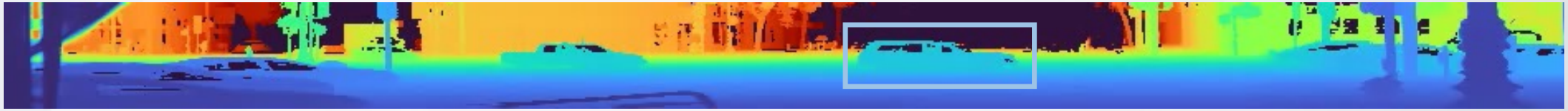


LiDAR4D

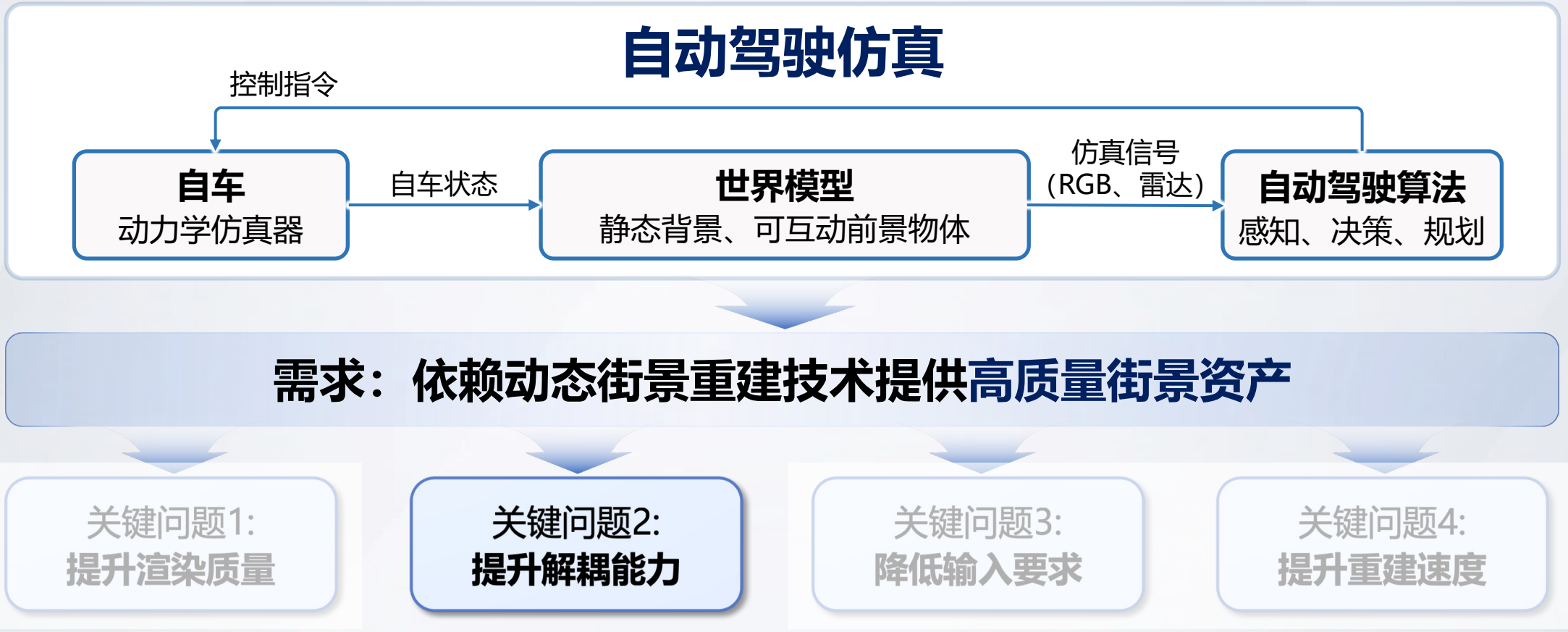


DyNFL

Ours



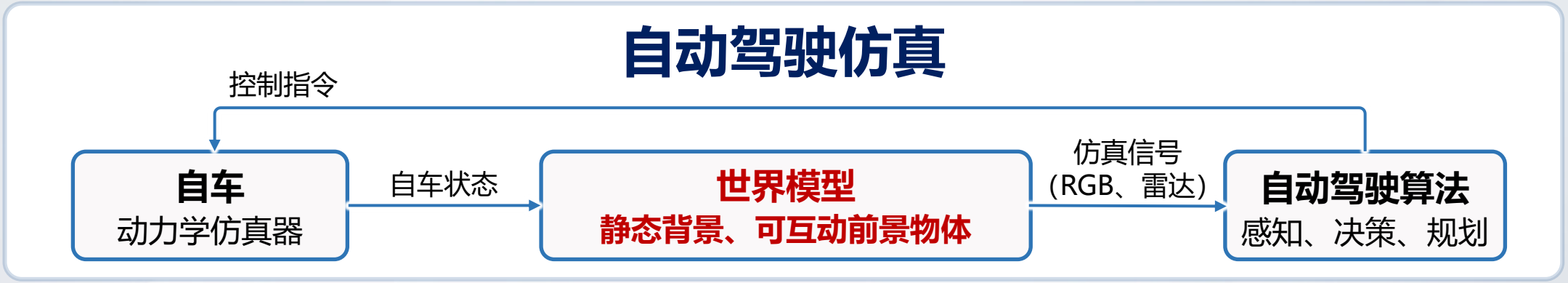
关键问题2: 提升解耦能力



关键问题2: 提升解耦能力

为什么这个问题重要

- 世界模型由**静态背景**和**可互动前景物体**组成。
- 因此，输入路采视频，我们需要将背景和前景分解开来，让前景物体可以自由移动或被编辑。



关键问题2: 提升解耦能力

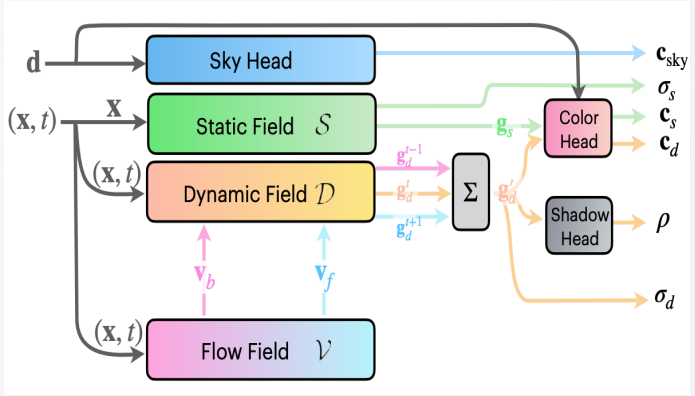
已有技术范式

范式1:
基于物体三维框

代表性工作：StreetGaussians

范式2:
基于自监督

代表性工作：EmerNeRF

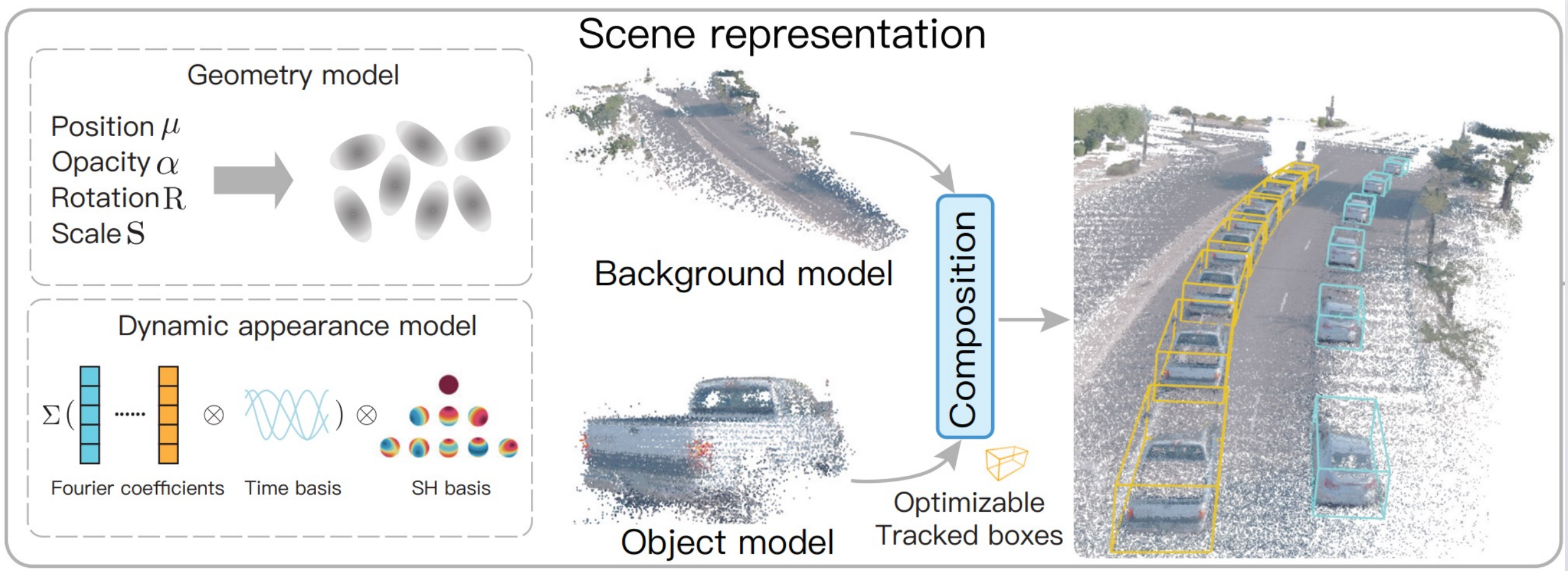


范式3:
基于逐帧实例分割

代表性工作：Split4D

范式1：Street Gaussians（实验室成果）

- 基于**组合式的三维高斯**表示动态三维场景

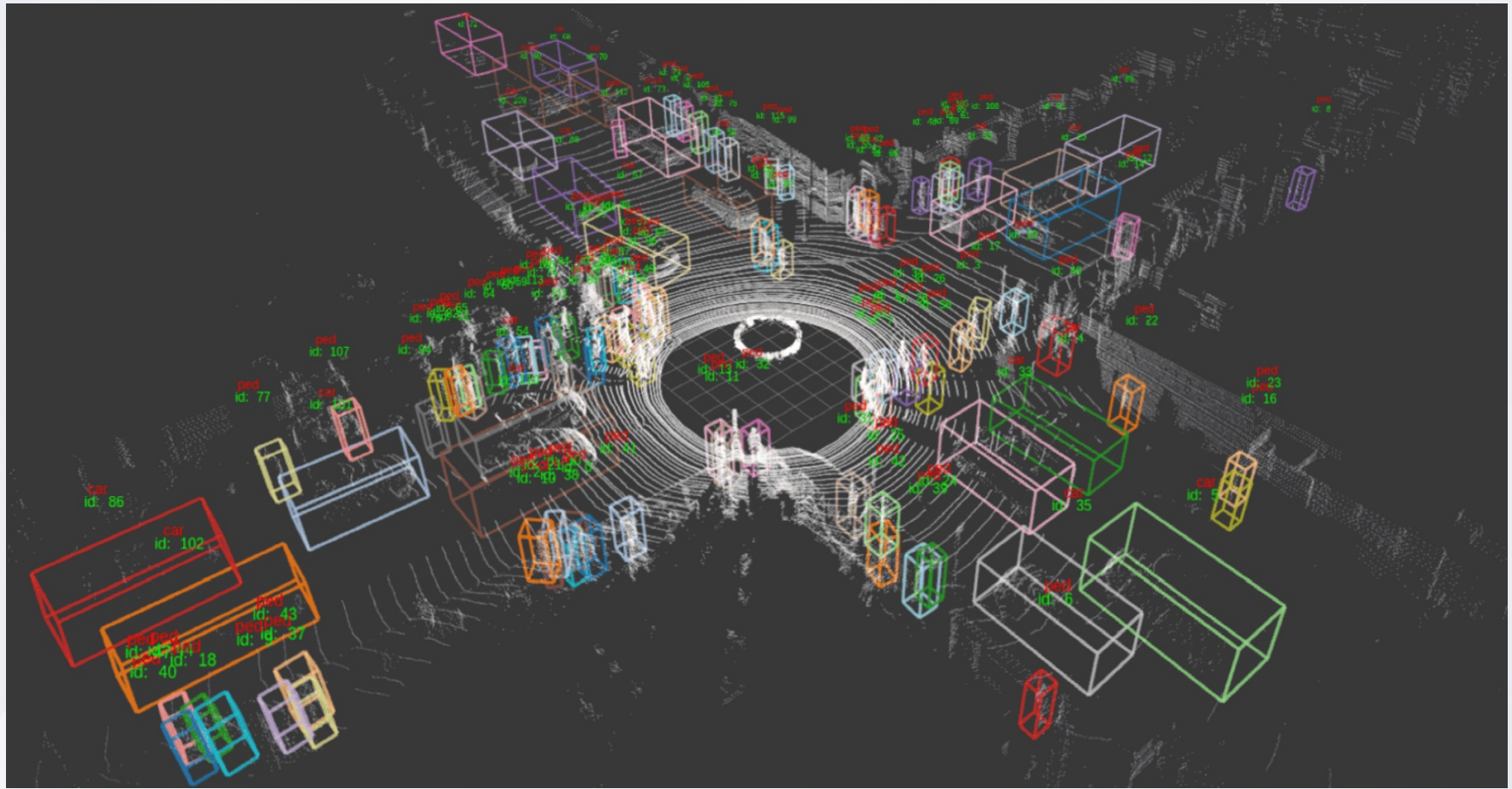


Street Gaussians: Modeling Dynamic Urban Scenes with Gaussian Splatting. ECCV 2024.

StreetGaussian的解耦效果

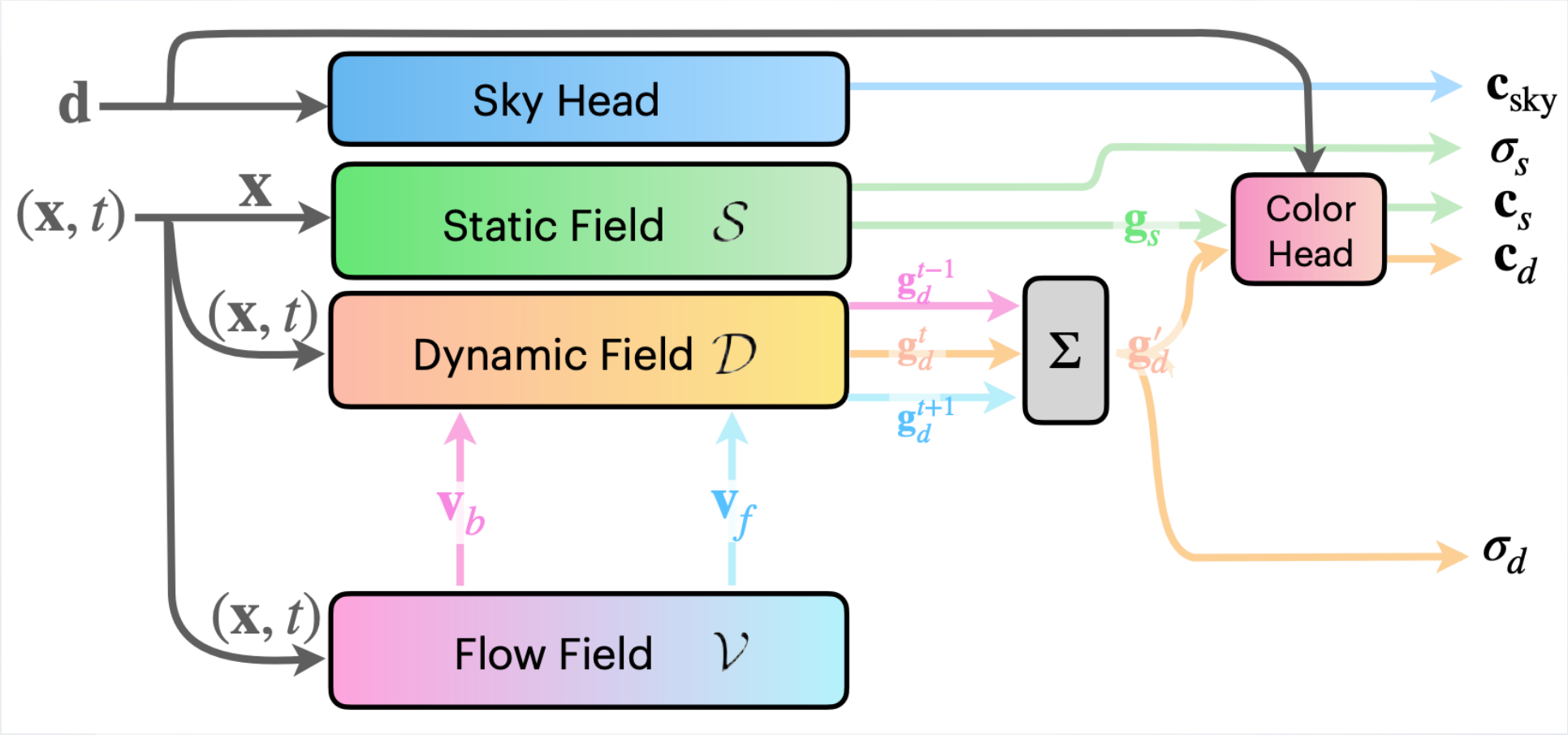
StreetGaussian的问题

- 日常路采数据中存在大量动态物体。难以准确标注全部物体的三维框，且标注成本高。



范式2：EmerNeRF

- 分别定义Dynamic Field和Static Field，再组合为最终图片。

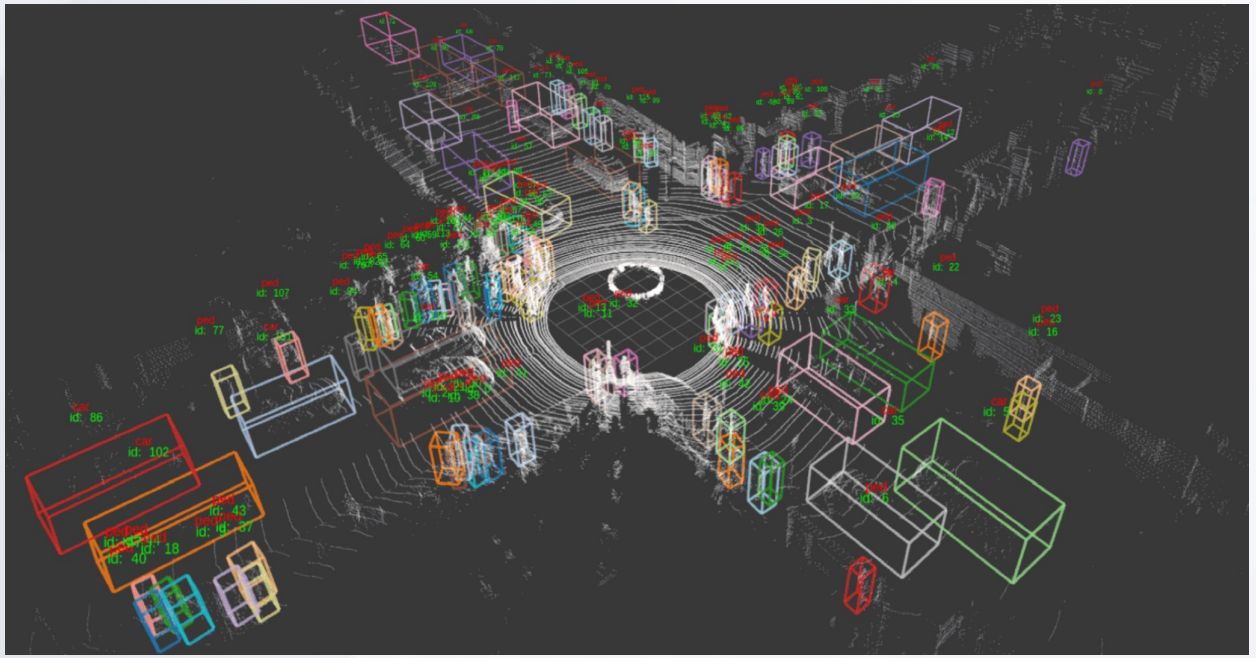


EmerNeRF: Emergent Spatial-Temporal Scene Decomposition via Self-Supervision. ICLR 2024.

仍存在的问题：解耦不够准确

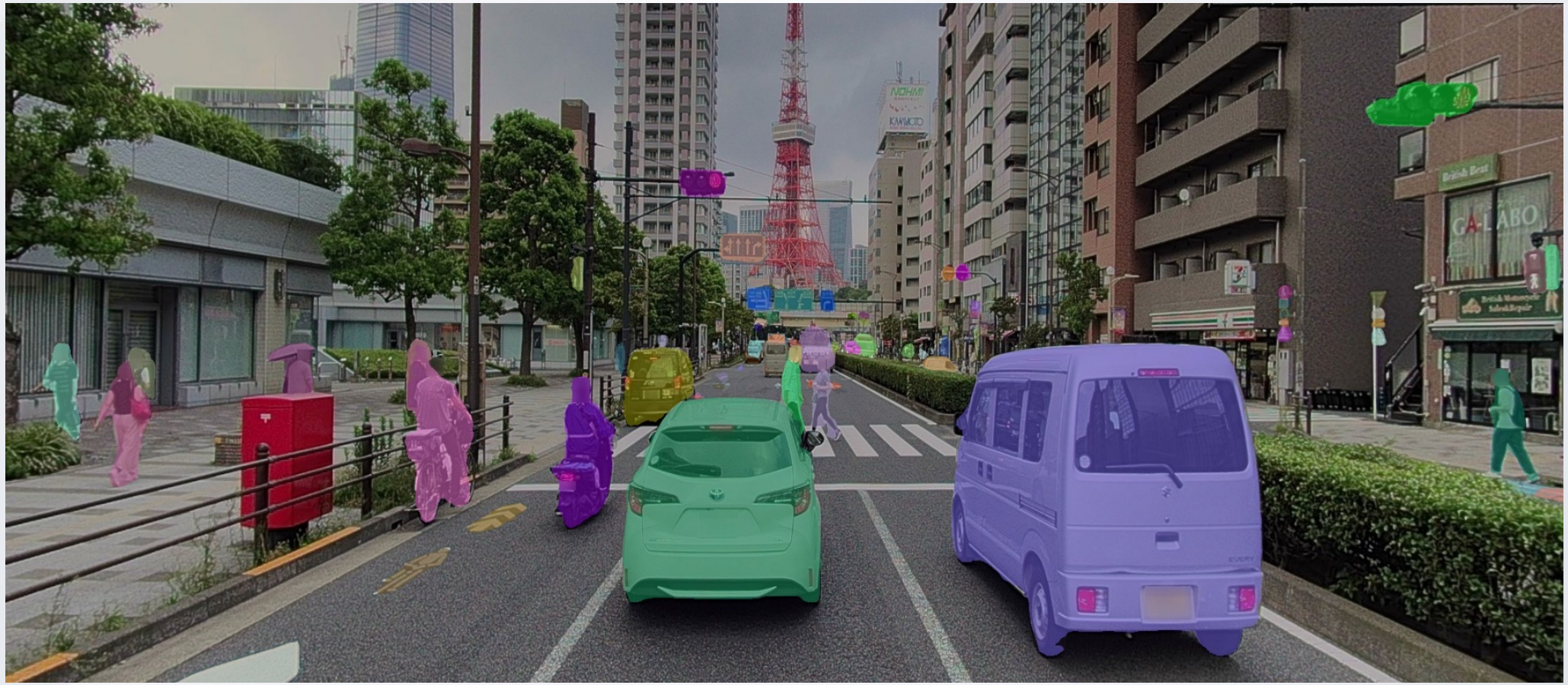
范式3：Split4D（实验室成果）

- 基于额外标注的方法面临的问题：
  - 输入的标注大多是三维框、Video segmentation，标注成本昂贵。
  - 如果采取自动化标注，很难准确。



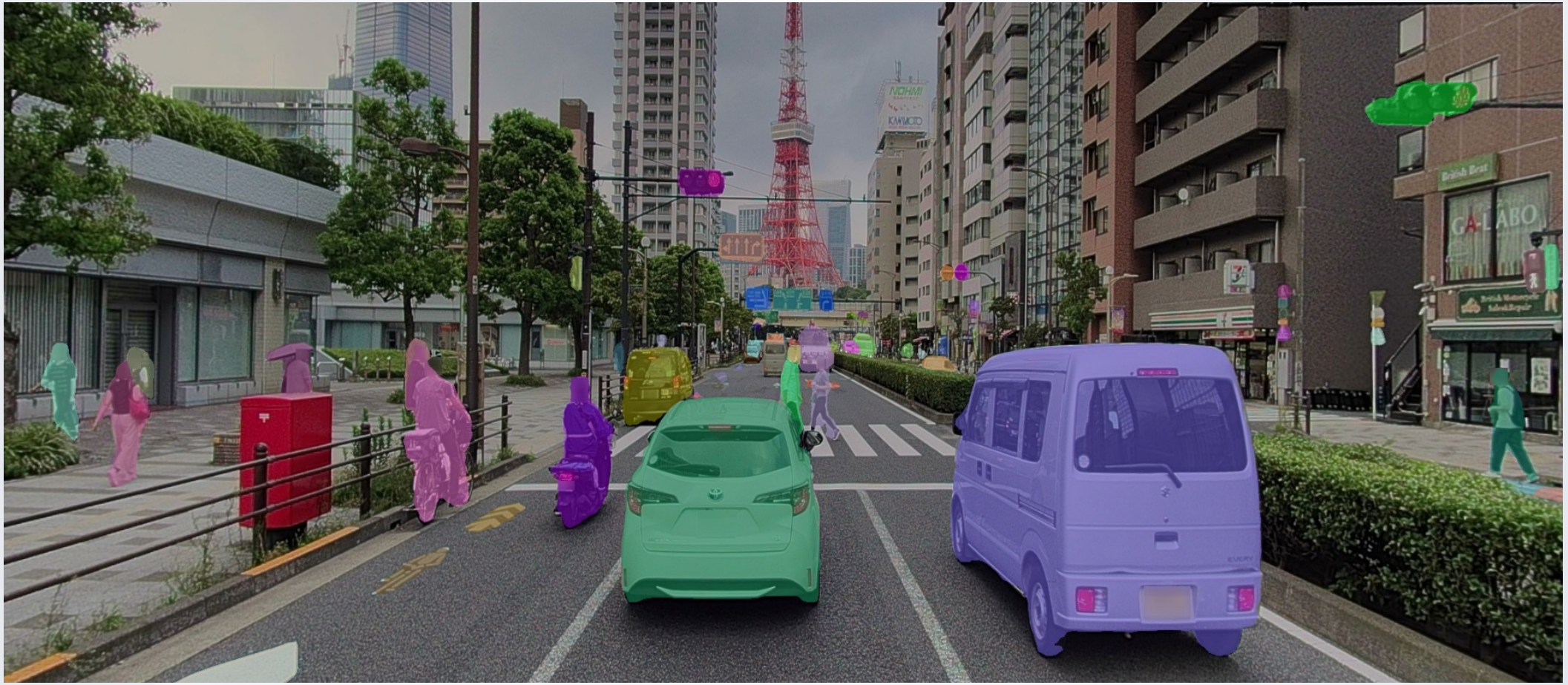
Split4D解决问题的思路

- 只使用**逐帧Segmentation**作为额外输入的标注：
  - 即使采取自动化标注，大部分情况下也足够准确。



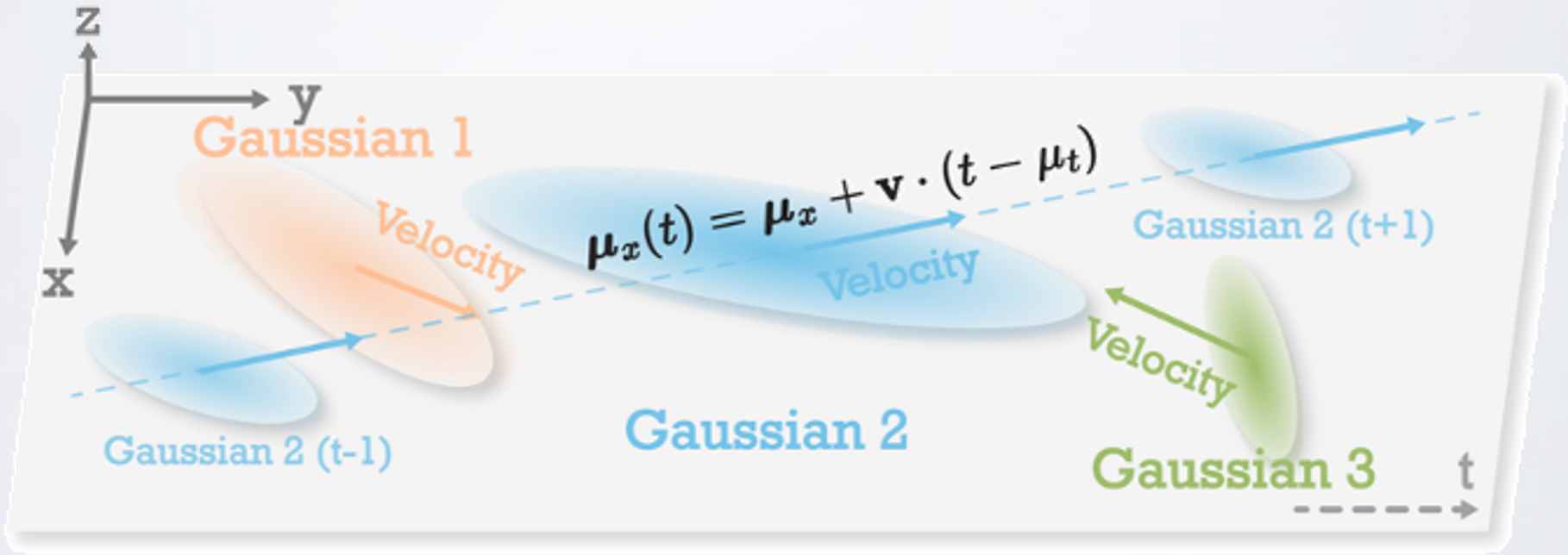
使用逐帧Segmentation面临的问题

- 如何将2D segmentation转变为4D segmentation?



如何将2D segmentation转变为4D segmentation?

- 解决问题的思路：
  - 借助FreeTimeGS的Short-range 3D correspondence。
  - 3D correspondence + 2D segmentation = 4D segmentation。



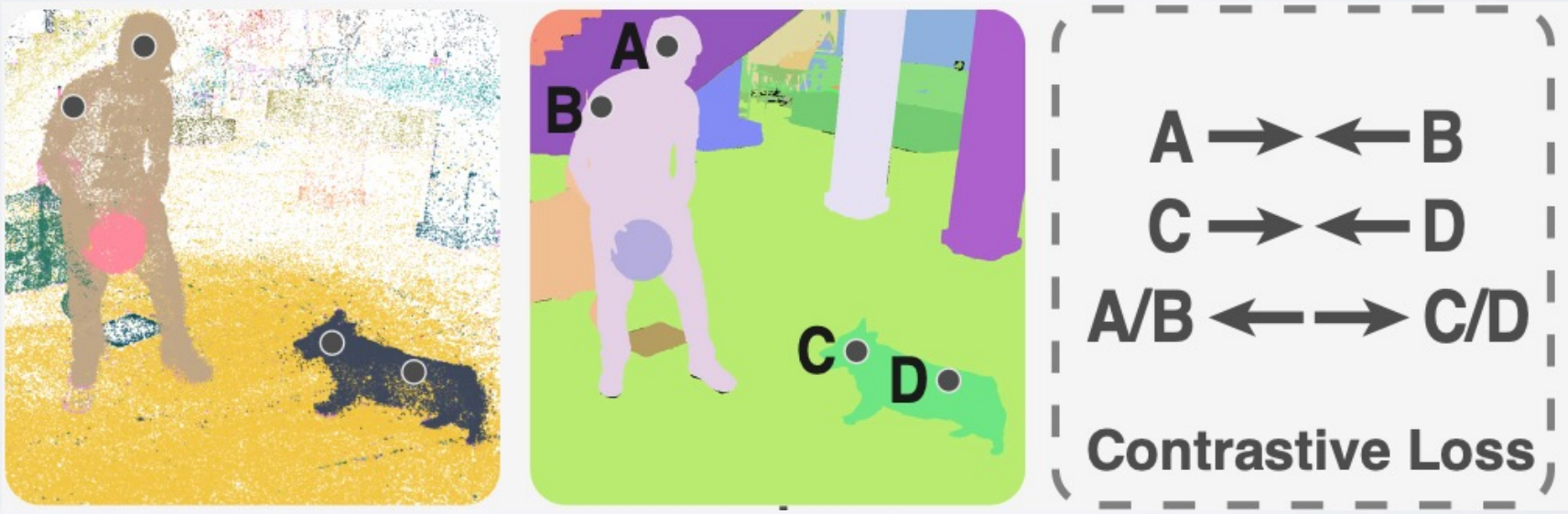
具体做法（1）：构建Freetime FeatureGS

Freetime FeatureGS：在FreeTimeGS上定义可学习特征。

具体做法（2）：对特征做contrastive loss

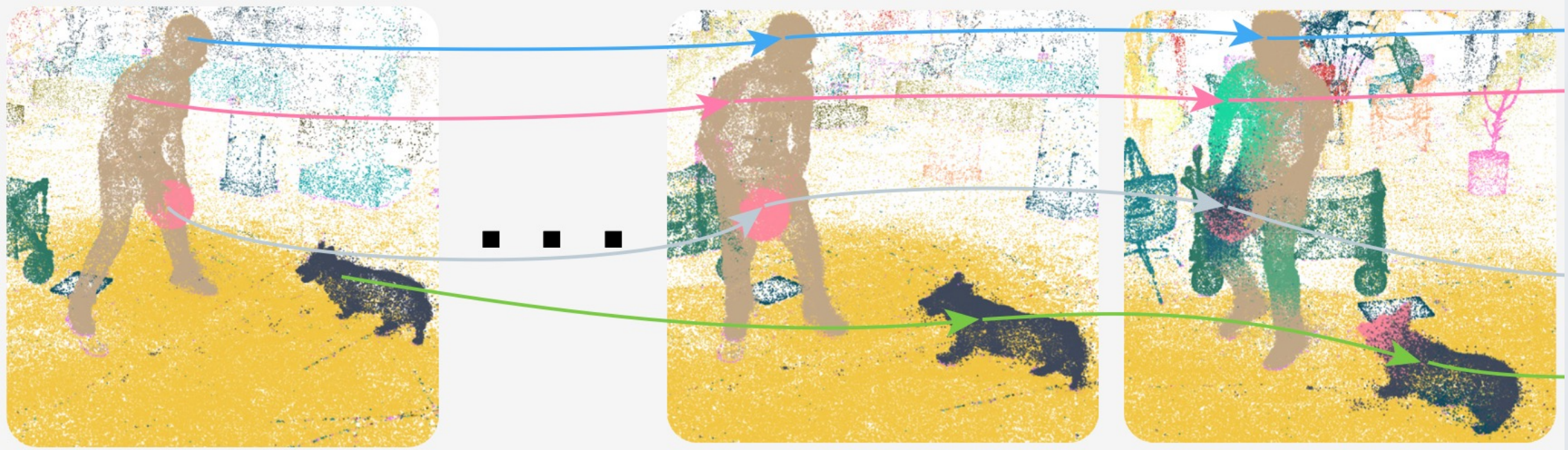
1. 渲染Freetime FeatureGS得到特征图。
2. 在特征图上采样像素点。

   属于同一个mask的特征相互拉近，否则相互变远。



具体做法（3）：利用3D correspondences传播特征

因为不同时刻的Gaussians相关联，因此可以得到时序一致的特征。



实验结果：整体重建效果

实验结果：前后背景分离

实验结果：场景编辑

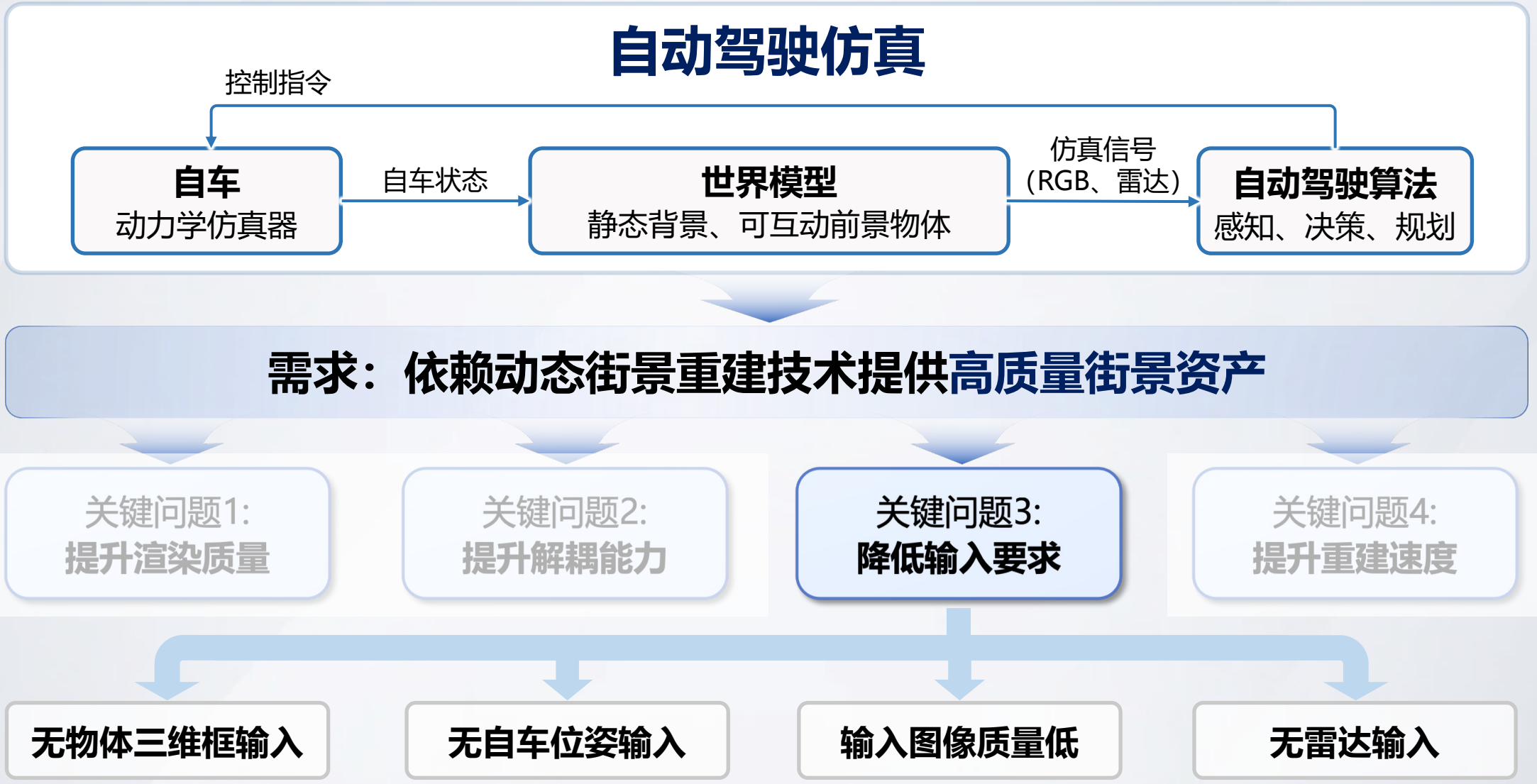
实验结果：整体重建效果

实验结果：场景编辑

实验结果：通用场景编辑

实验结果：通用场景编辑

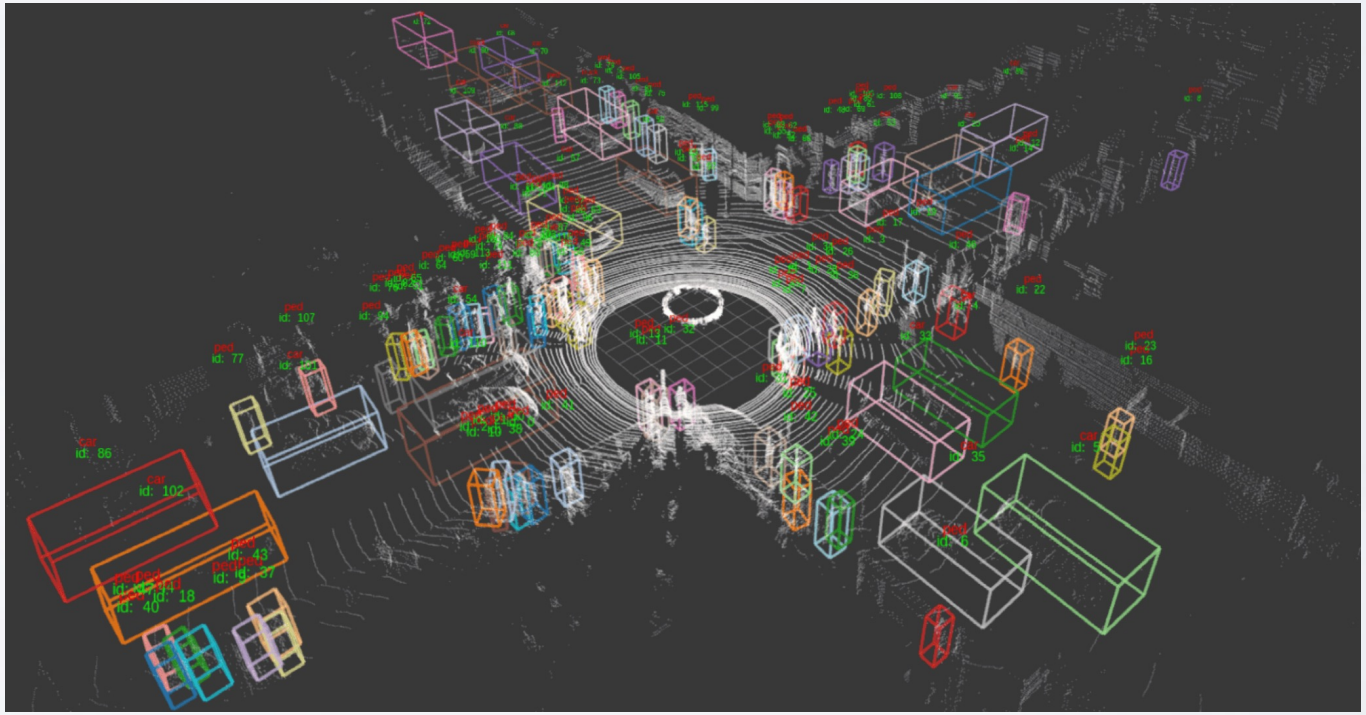
关键问题3: 降低输入要求



关键问题3.1: 无物体三维框输入

为什么这个问题重要

- 日常路采数据中存在大量动态物体。难以准确标注全部物体的三维框，且标注成本高。

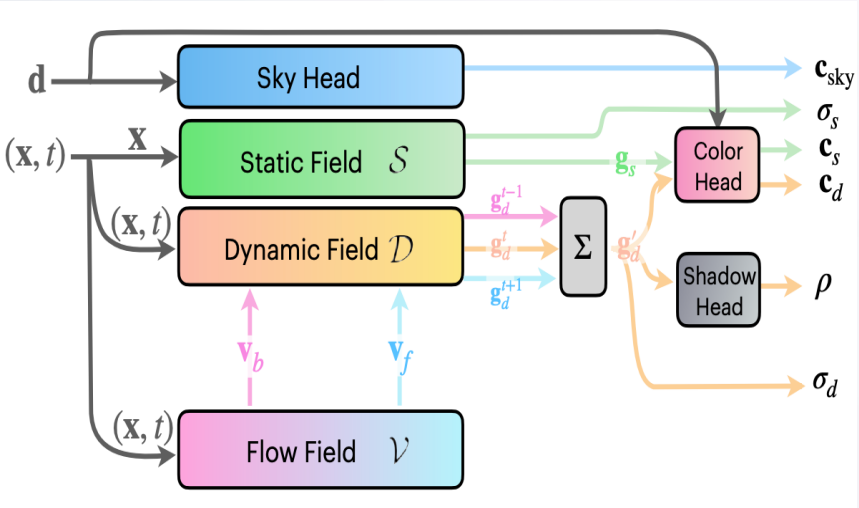


关键问题3.1: 无物体三维框输入

已有技术范式

范式1:
基于Dynamic NeRF

代表性工作：EmerNeRF



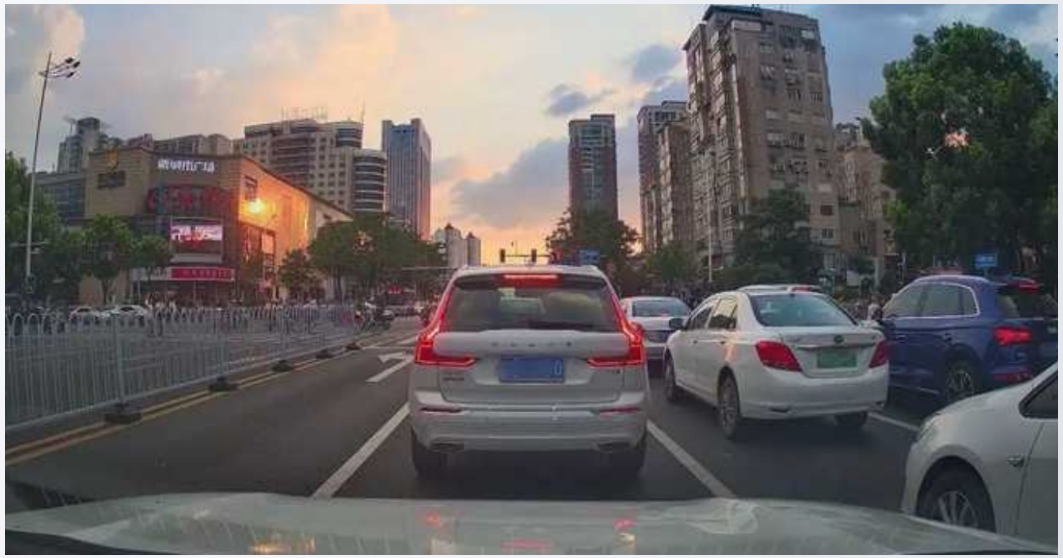
范式2:
基于native 4DGS

代表性工作：FreeTimeGS

关键问题3.2: 无自车位姿输入

为什么这个问题重要

- 行车记录仪采集的视频，不是由数据采集车传回，而是由量产车辆传回，不带有相机位姿高精标注设备。要重建行车记录仪视频，必须要有自车位姿。
- 即使是数据采集车，有时候传回的自车位姿也不够准确。

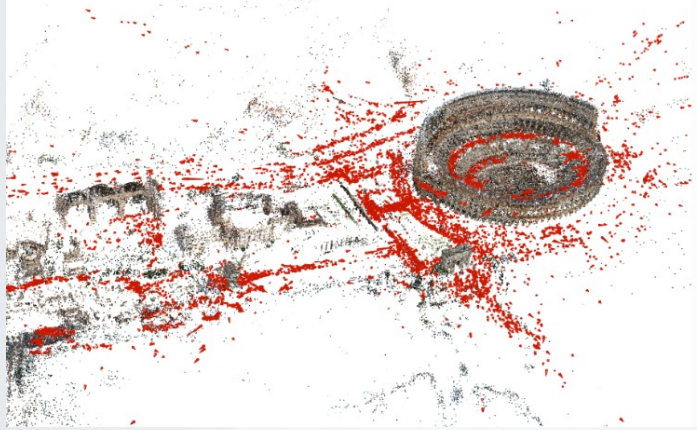


关键问题3.2: 无自车位姿输入

已有技术范式

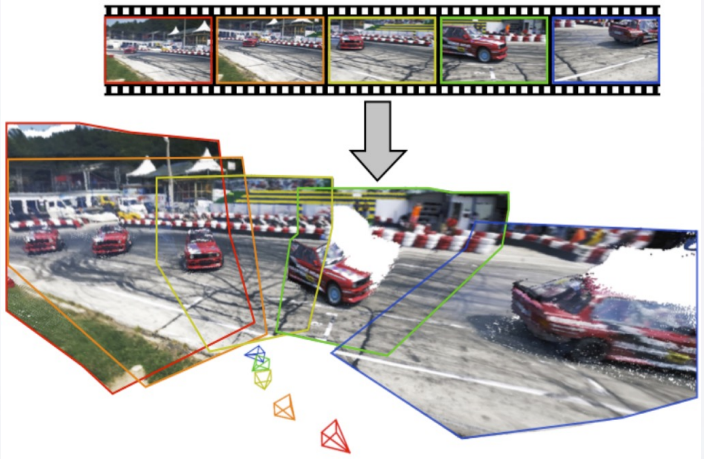
范式1:
Structure from Motion

代表性工作：COLMAP



范式2:
SLAM

代表性工作：MegaSaM

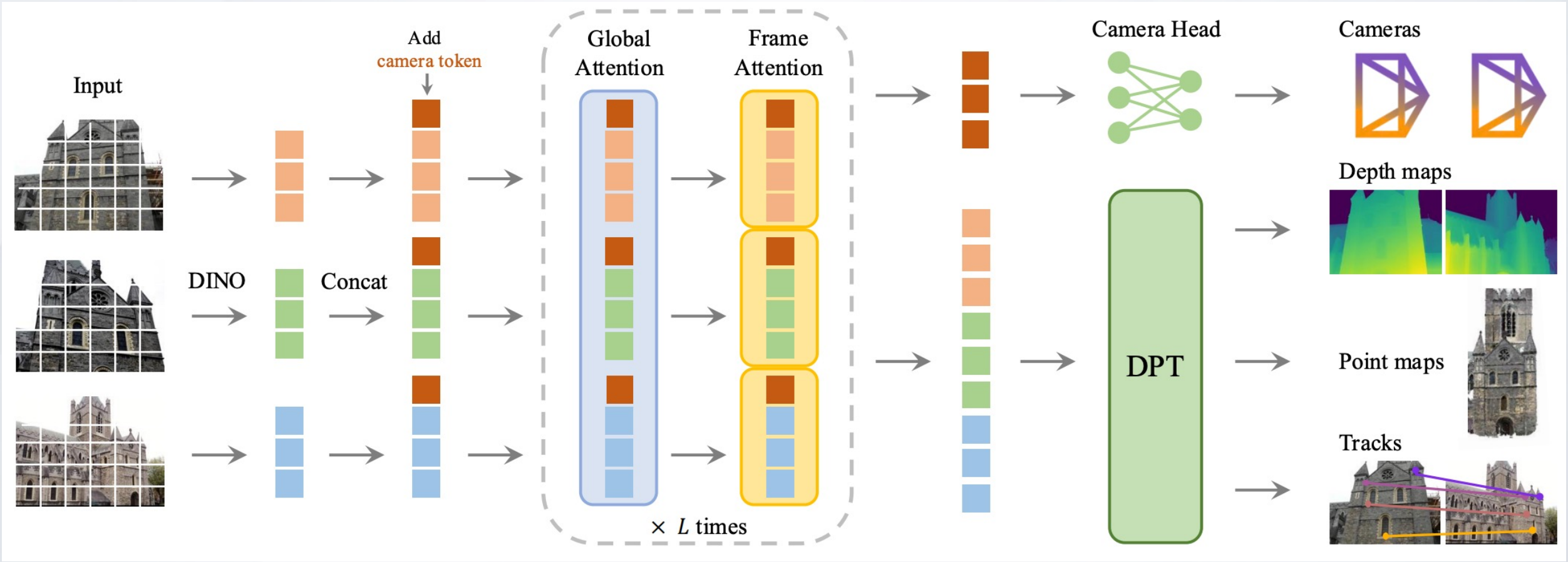


范式3:
Learning-based Methods

代表性工作：VGGT

VGGT

- 使用Transformer网络直接回归camera poses和depth maps。

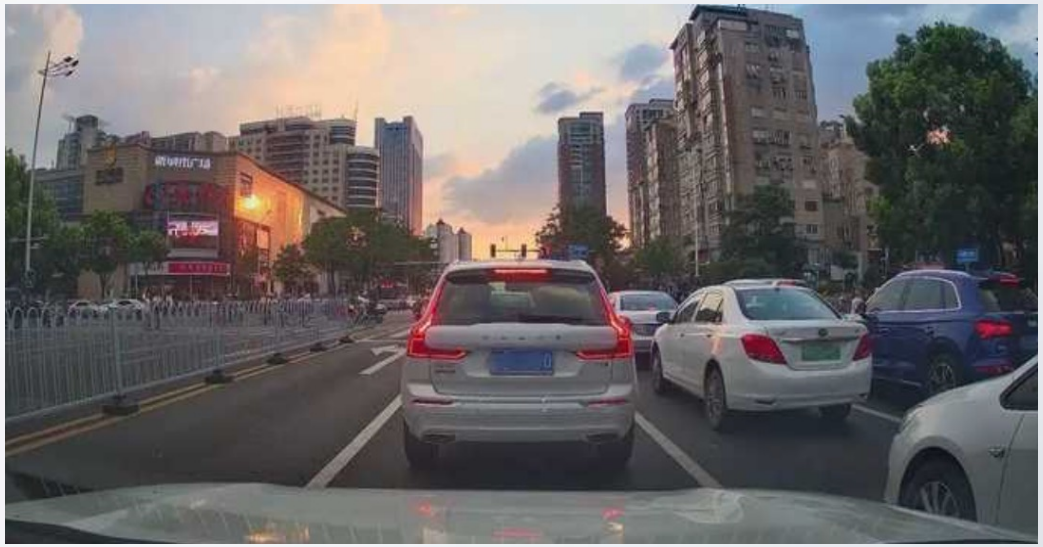


VGGT: Visual Geometry Grounded Transformer. CVPR 2025.

关键问题3.3: 输入图像质量低

为什么这个问题重要

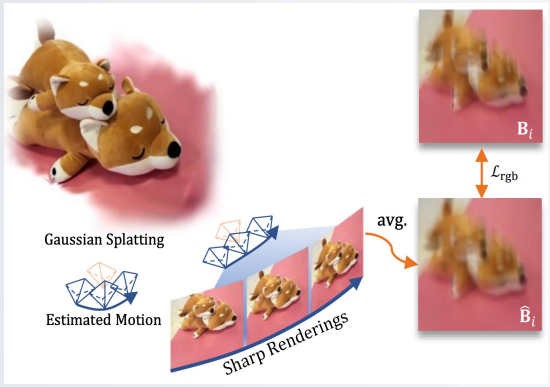
- 即使是数据采集车传回的数据，也经常面临各种图像质量问题，包括：炫光、动态模糊、黑夜、镜头有污渍。
- 环视相机之间经常存在色调不一致的情况。

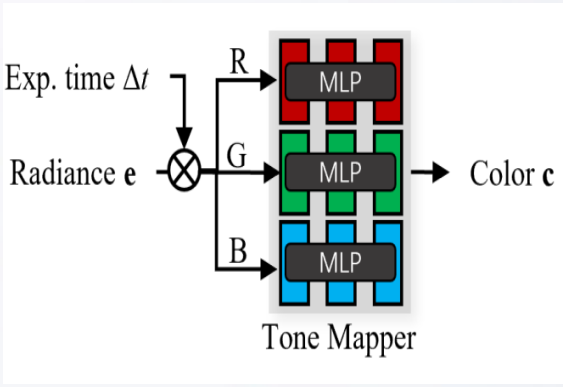


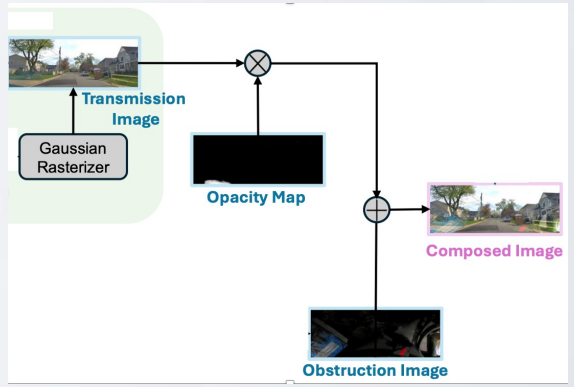
关键问题3.3: 输入图像质量低

常见低质量图像情况

图像模糊	多视角光照不一致	黑夜炫光	摄像头在前车窗内
相关工作：DeblurGS	相关工作： BilaRF	相关工作：RawNeRF	相关工作：DC-Gaussian







核心解决思路：在渲染过程中显式建模 Image Degradation Model

关键问题3.4: 无雷达输入

为什么这个问题重要

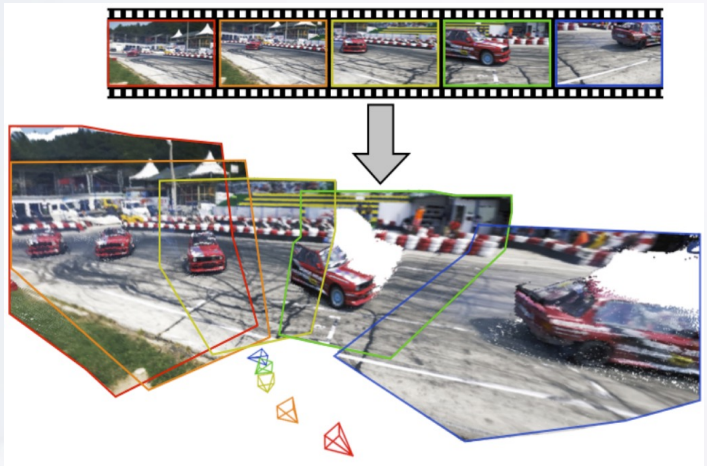
- 行车记录仪采集的视频，不是由数据采集车传回，而是由量产车辆传回，不带有高精度激光雷达。
- 已有重建算法通常依赖激光雷达提供三维高斯初始位置和几何约束。

关键问题3.4: 无雷达输入

已有技术范式：额外估计场景几何以初始化三维高斯

范式1:
SLAM

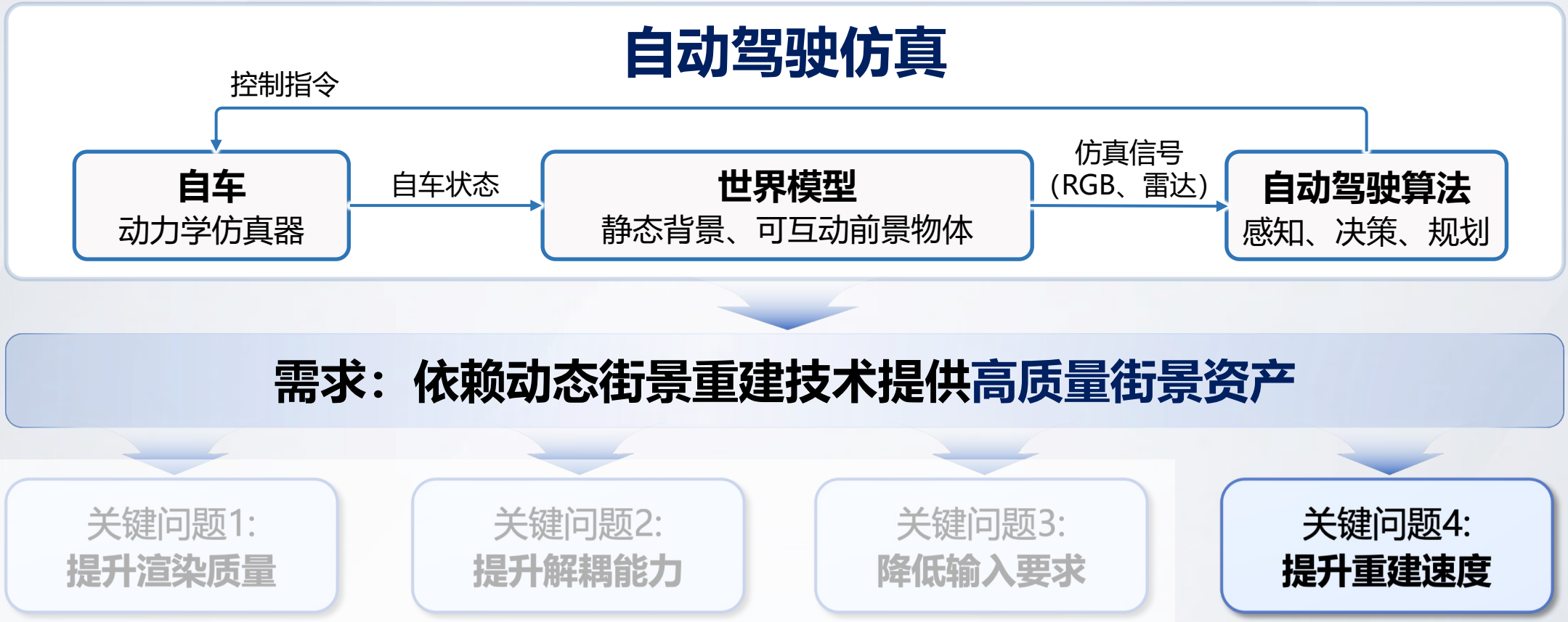
代表性工作：MegaSaM



范式2:
Learning-based MVS

代表性工作：VGGT

关键问题4: 提升重建速度



关键问题4: 提升重建速度

3DGS存在的问题：即使是小场景，重建也需5分钟

关键问题4: 提升重建速度

为什么这个问题重要

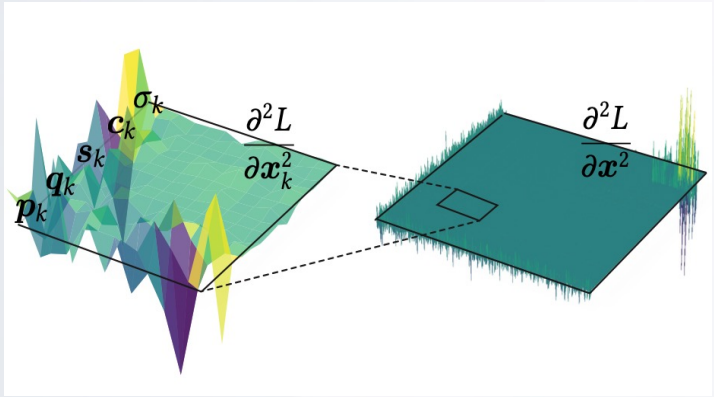
- 大多数车厂采集了几百万到几十亿公里的街道数据。
- 如果每公里重建需要2小时，那么所需时间成本和算力成本过大。

关键问题4: 提升重建速度

已有技术范式

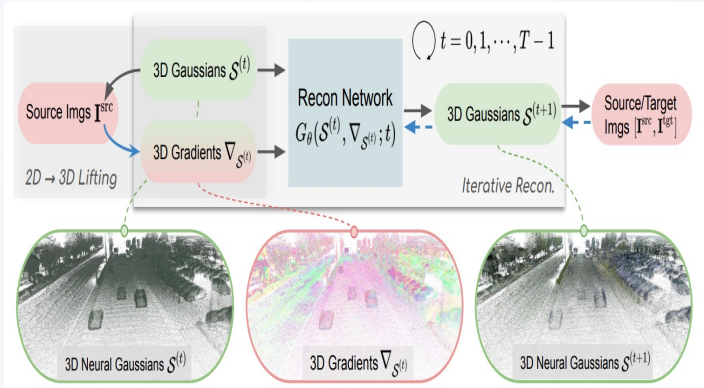
范式1:
二阶优化器

代表性工作：3DGS2



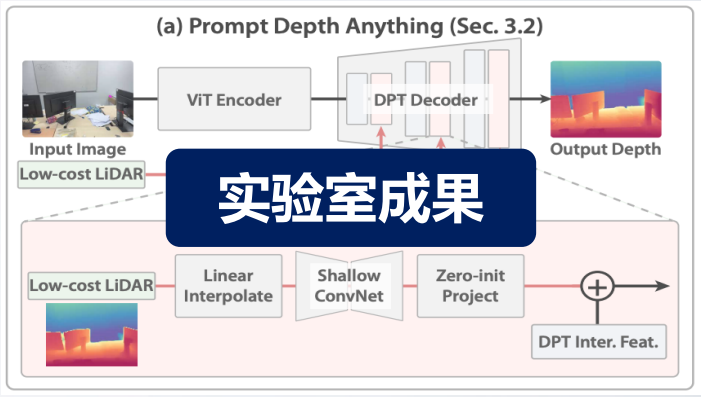
范式2:
基于学习的优化器

代表性工作：G3R



范式3:
前向推理模型

代表性工作：PromptDA



PromptDA （实验室成果）

- 根据Input LiDAR输出准确的绝对深度

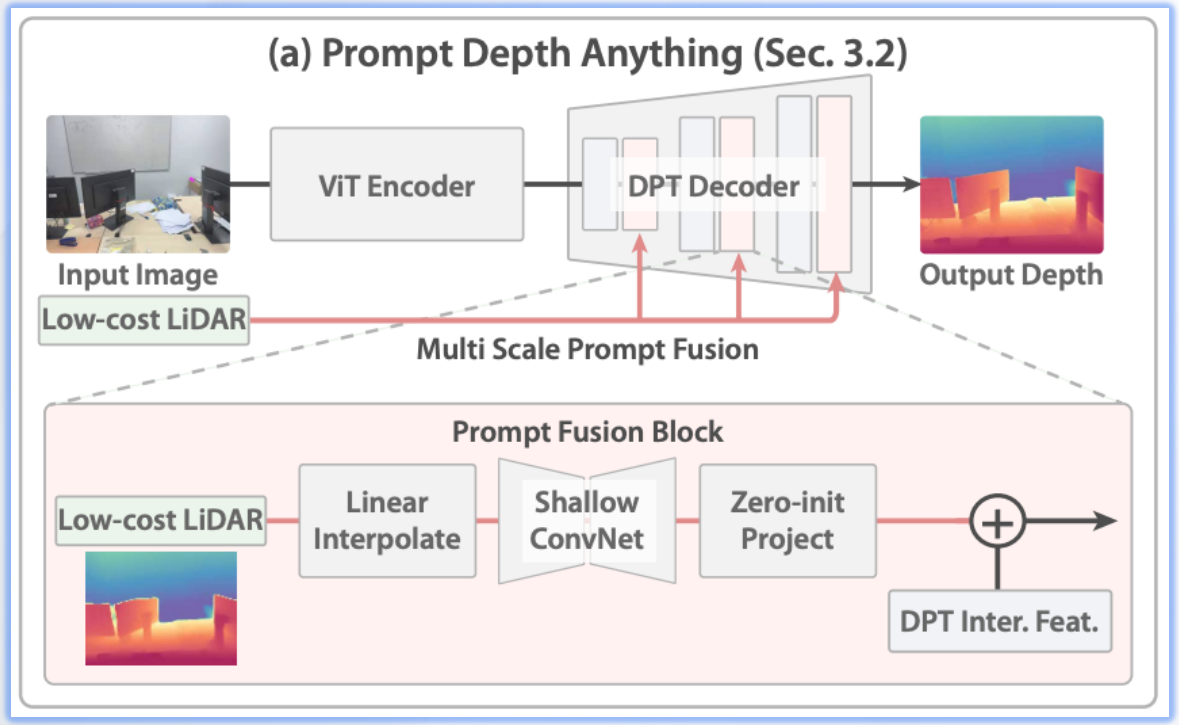
24x24 points

4K RGB &
Low-res Metric Depth

4K Depth

具体做法

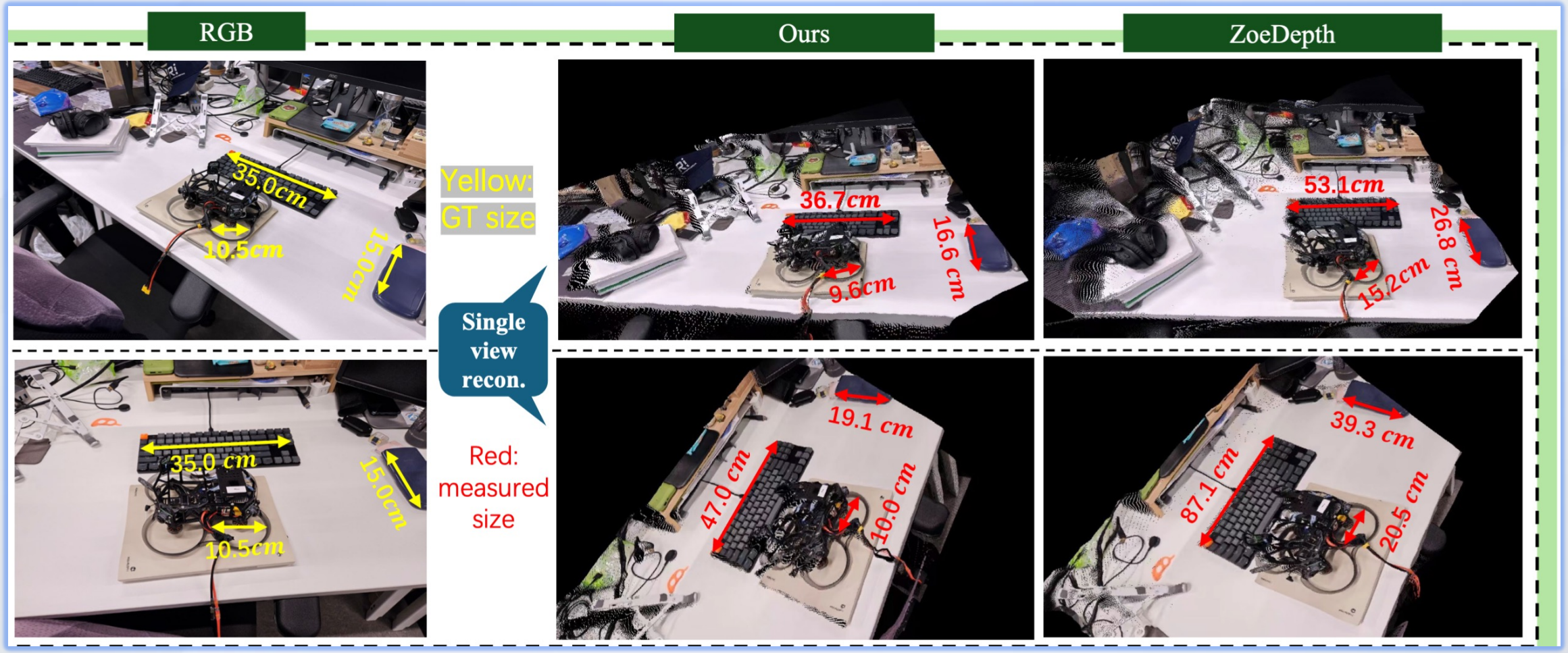
- 如何提示深度基础模型?



	ARKitScenes		ScanNet++		
	L1 ↓	AbsRel ↓	Acc ↓	Comp ↓	F-Score ↑
(a) Ours$_{syn}$ (synthetic data)	0.0163	0.0142	0.0746	0.0666	0.7307
(b) w/o prompting	0.0605	0.0505	0.0923	0.0801	0.5696
(c) w/o foundation model	0.0194	0.0169	0.0774	0.0713	0.7077
(d) AdaLN prompting	0.0197	0.0165	0.0795	0.0725	0.6943
(e) Cross-atten. prompting	0.0523	0.0443	0.0932	0.0819	0.5595
(f) Controlnet prompting	0.0239	0.0206	0.0785	0.0726	0.6899

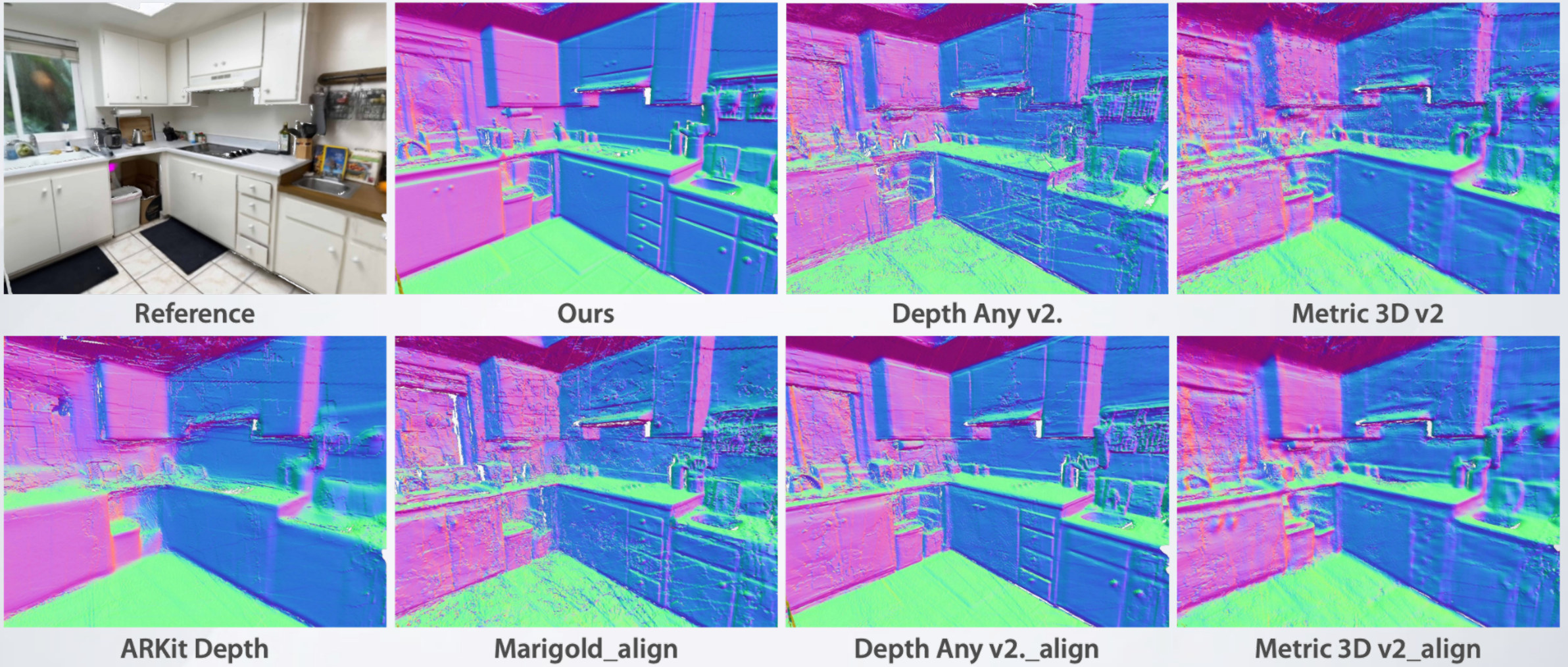
实验结果

- 厘米级误差的度量(Metric)深度估计



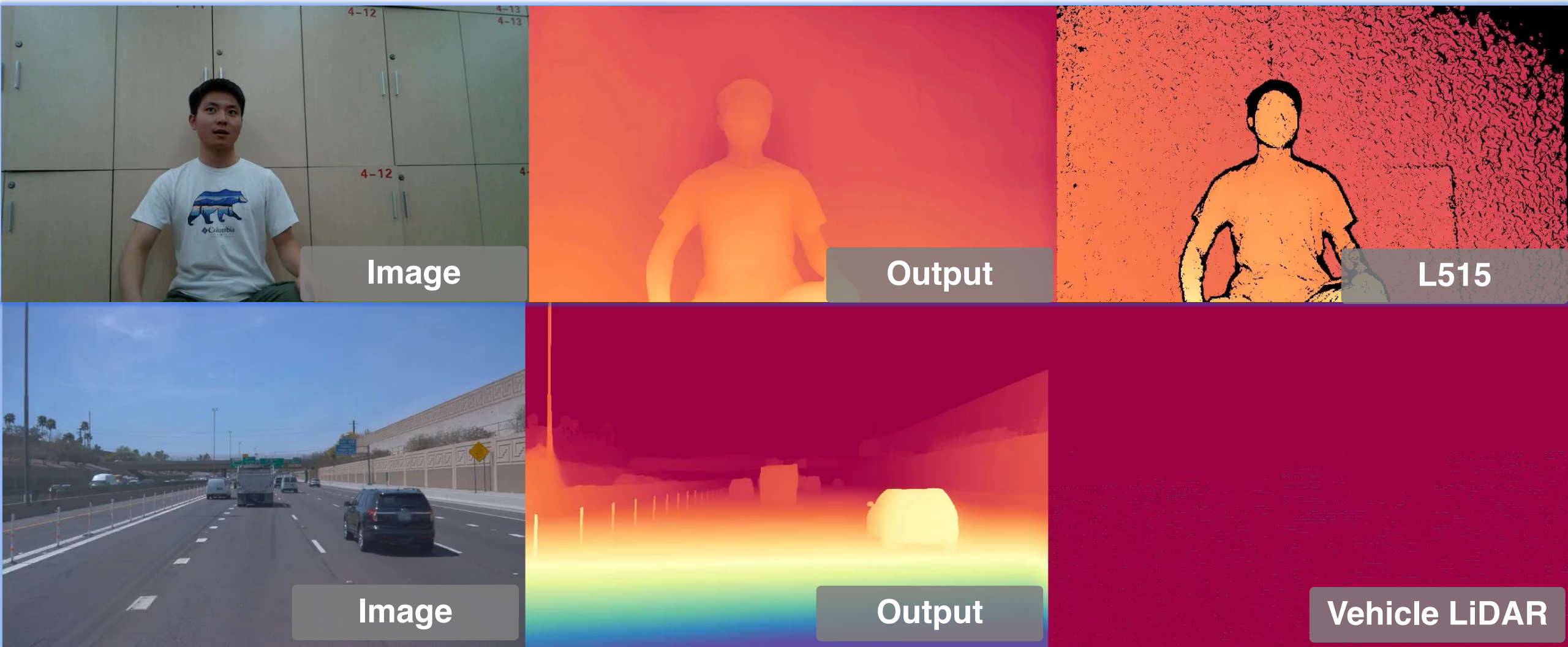
实验结果

- 大幅提升深度基础模型重建效果



实验结果

- 提示机制可泛化至Realsense L515、车载LiDAR等常用传感器

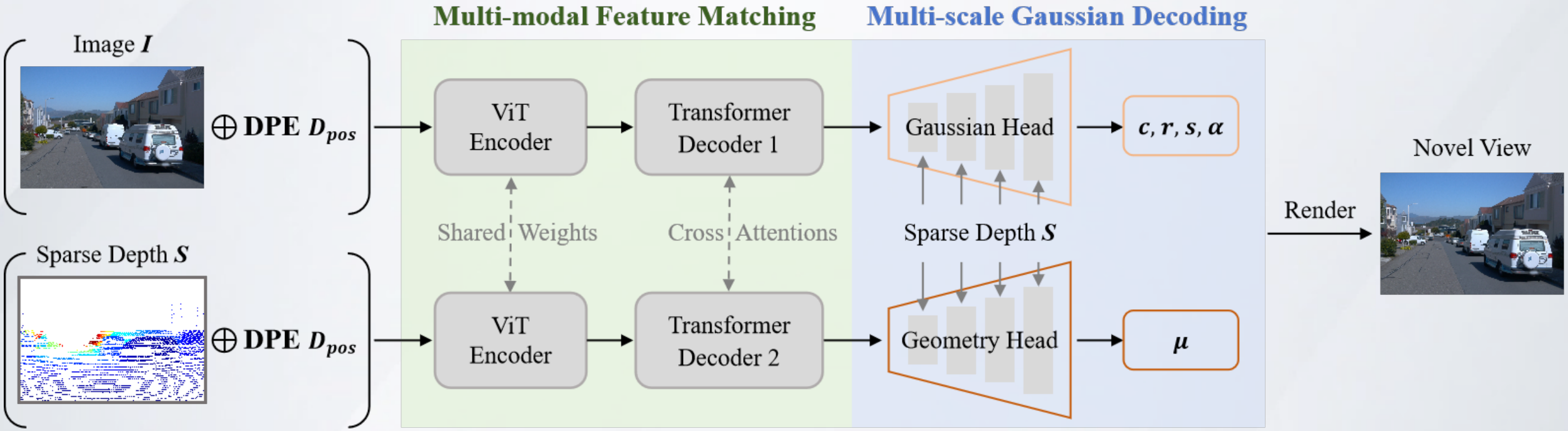


自动驾驶应用

- 基于街景雷达的室外大场景重建

ADGaussian （实验室成果）

- PromptDA预测高斯位置，Gaussian Head预测高斯外观



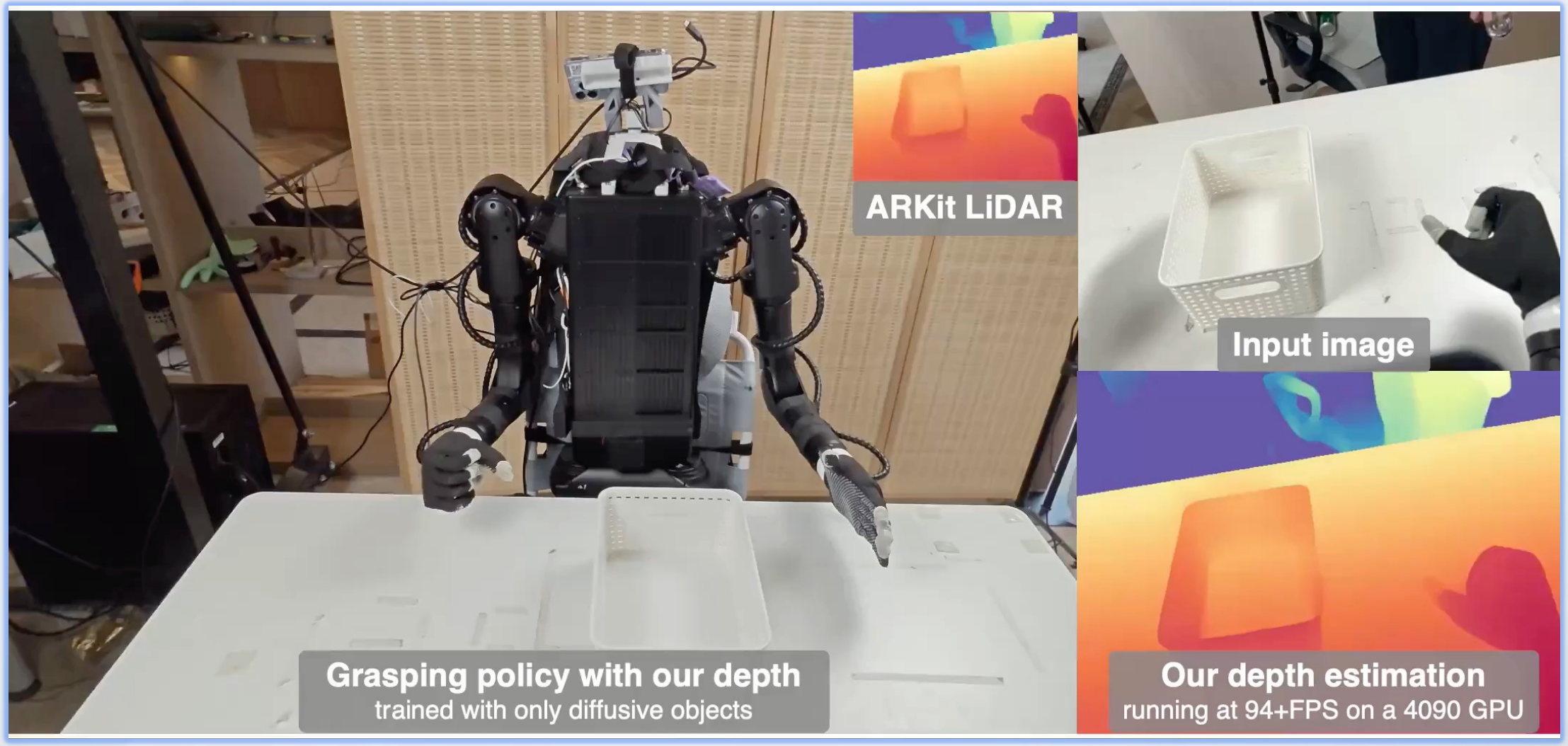
ADGaussian: Generalizable Gaussian Splatting for Autonomous Driving with Multi-modal Inputs. arXiv 2025.

ADGaussian的效果：换车道

ADGaussian的效果：时序预测

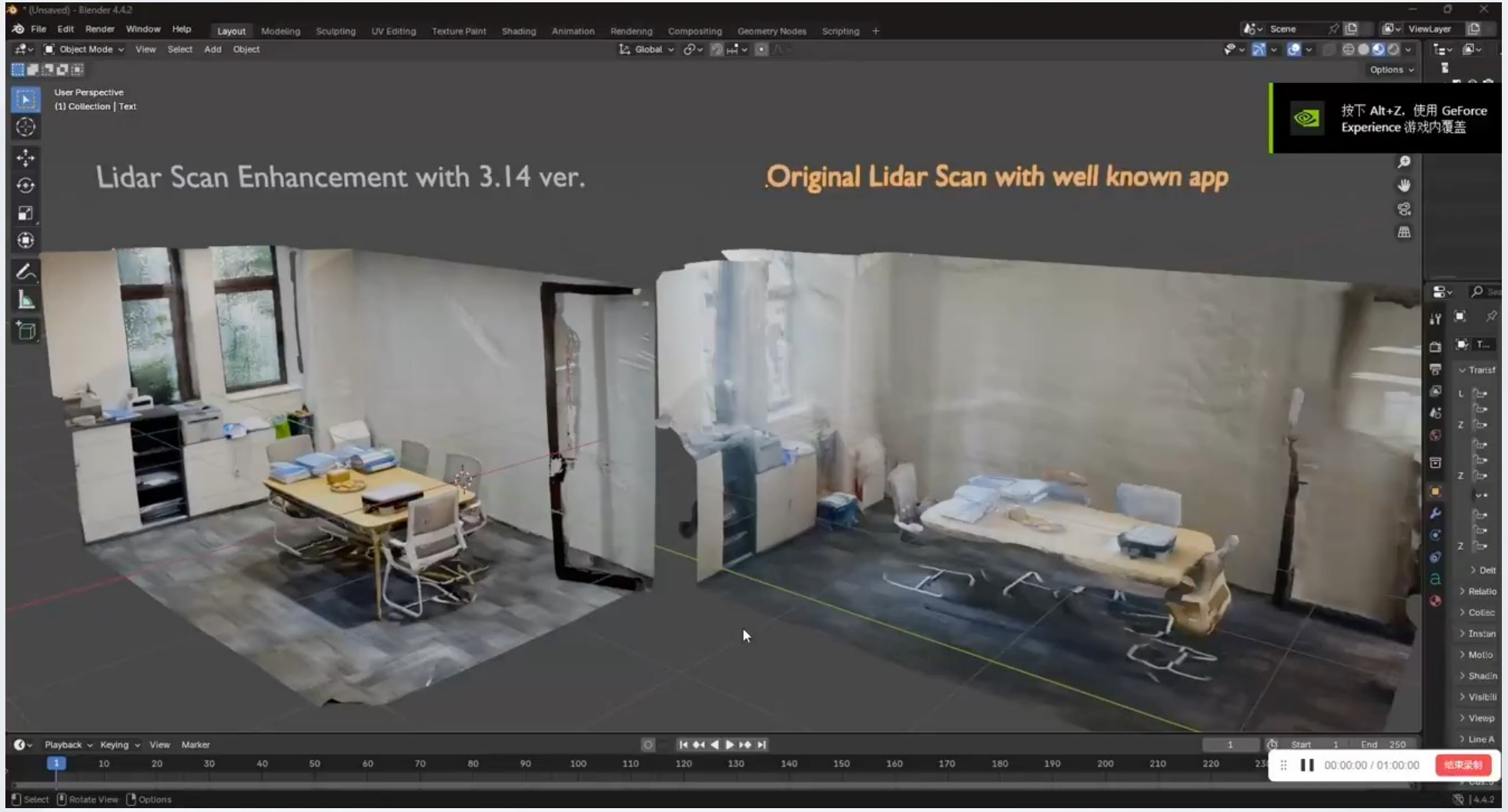
机器人应用

- 通用机器人抓取，抓取成功率超越Image及LiDAR作为输入

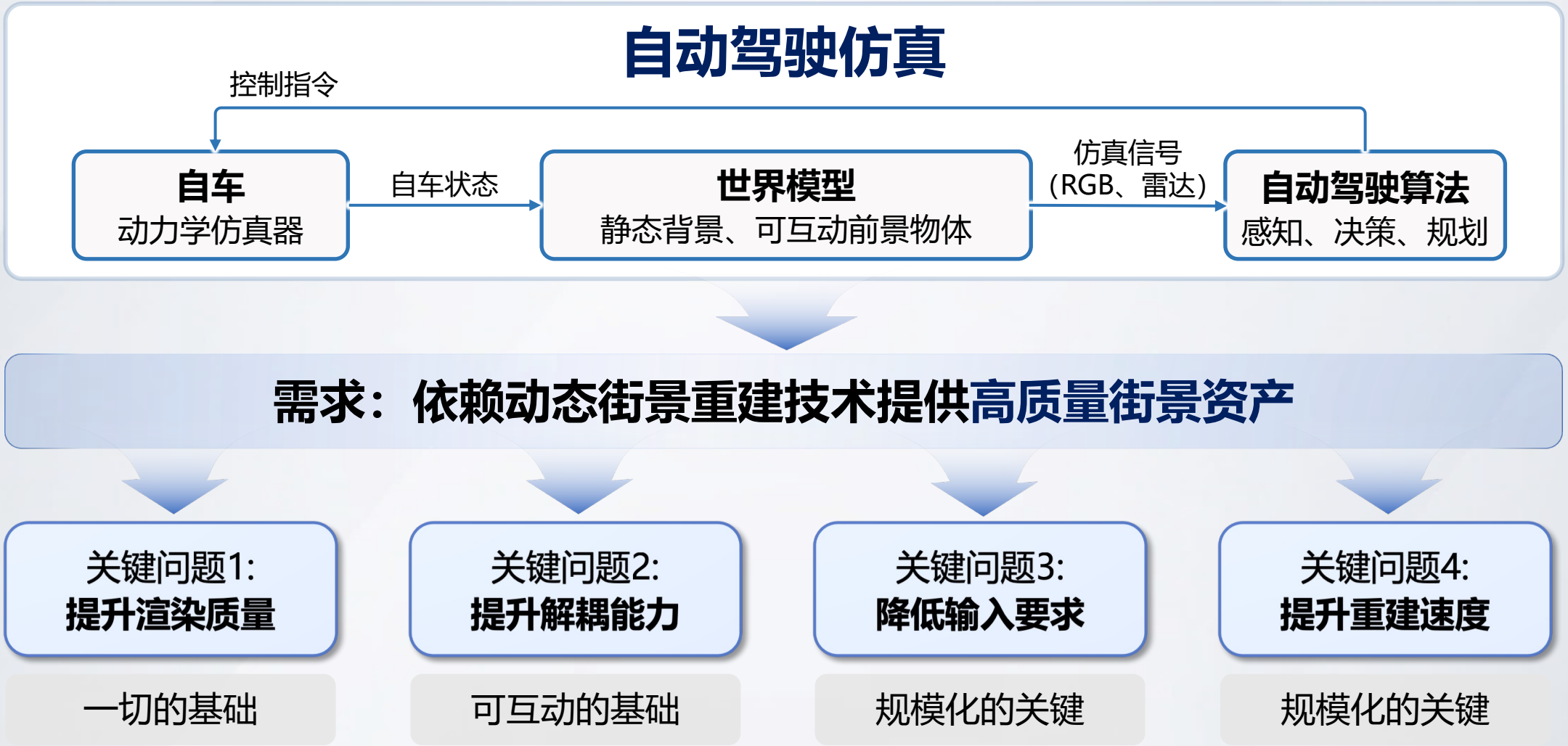


三维扫描应用

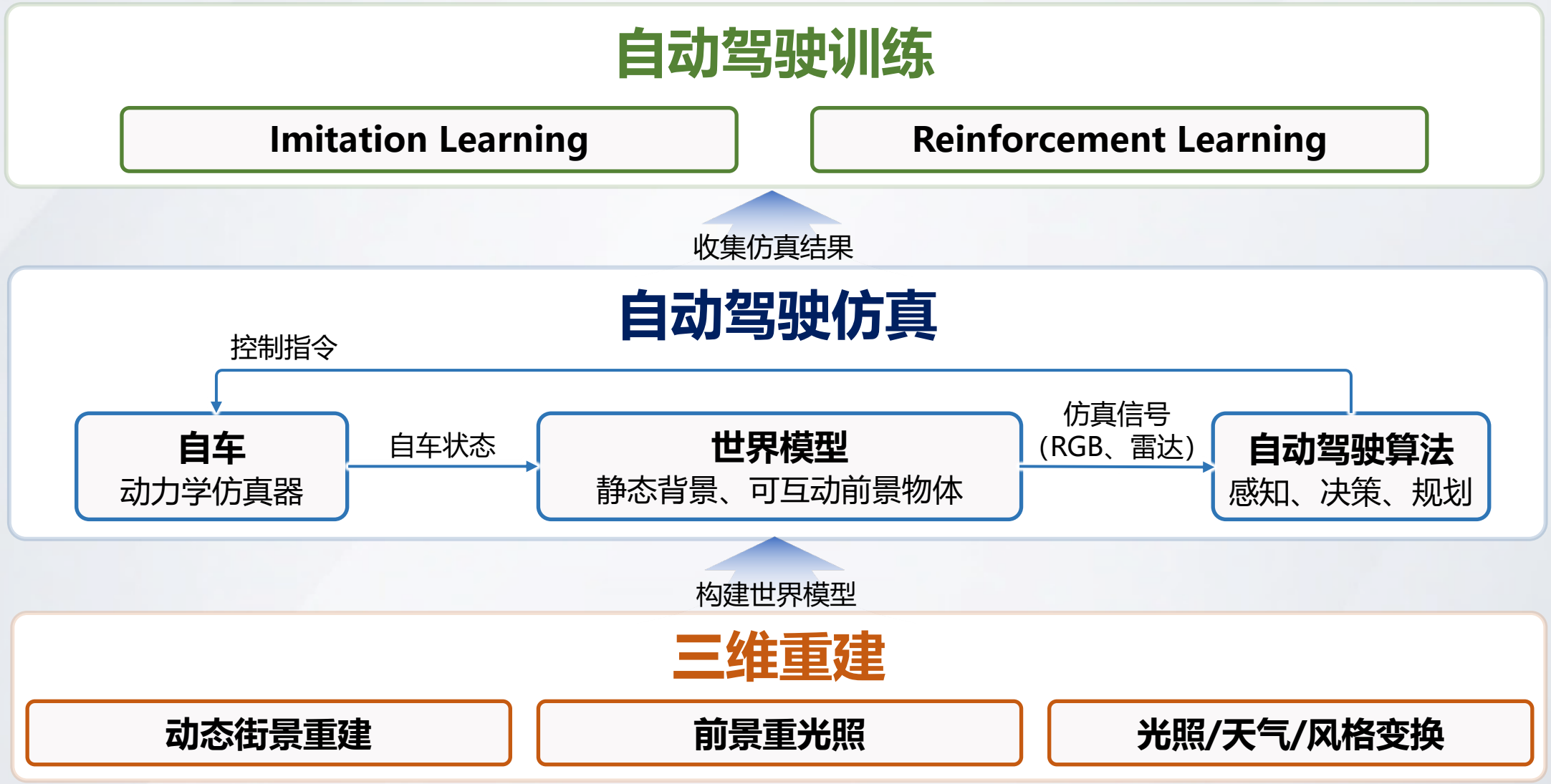
- PromptDA被集成进三维扫描App KIRI Engine



总结：自动驾驶仿真中的动态街景重建需要哪些能力



总结：三维重建在自动驾驶仿真中的位置



实验室已有研究积累

自动驾驶仿真

需求：依赖动态街景重建技术提供高质量街景资产

关键问题1: **提升渲染质量**	关键问题2: **提升解耦能力**	关键问题3: **降低输入要求**	关键问题4: **提升重建速度**
相关工作: **StreetGaussians、StreetCrafter、FreeTimeGS等**	相关工作: **StreetGaussians 、Split4D**	相关工作: **MatchAnything、D-SfM、Murre等**	相关工作: **PromptDA、ADGaussian**

论文发表：在CVPR、ICCV、ECCV等发表多篇Oral、Highlight论文，谷歌引用量数百次。

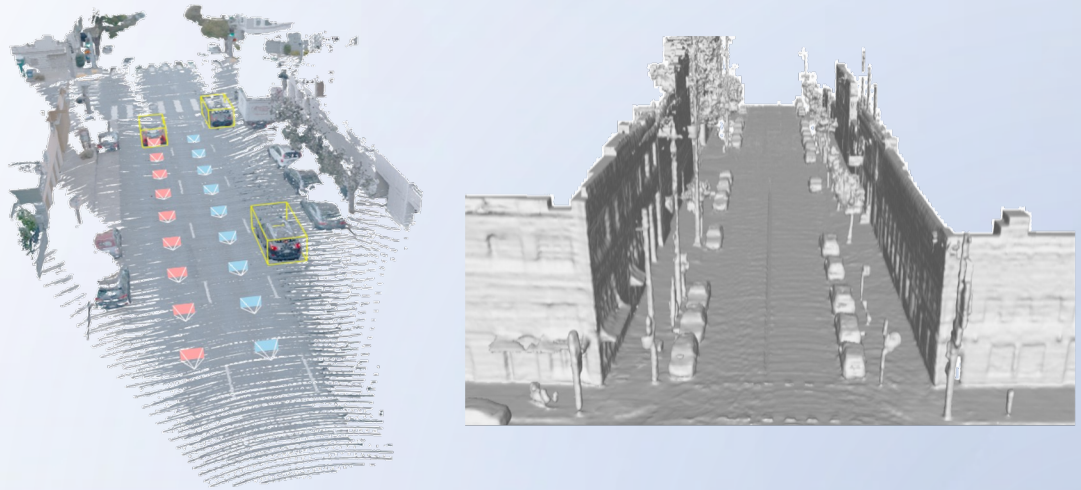
代码开源：论文GitHub stars累积数千次，数次被集成进知名算法库kornia和transformers。

竞赛获奖：获得谷歌举办的全球三维重建挑战赛冠军。

自动驾驶仿真，优先级靠前的5个问题

基于动态三维重建的自动驾驶闭环仿真

路采数据



动态三维街景

- 问题1: 闭环仿真强化学习效果
- 问题2: 大视角渲染质量
- 问题3: 前后景解耦质量
- 问题4: 动态场景表征重建速度
- 问题5: 街道级自车位姿标定

谢谢！

彭思达
浙江大学CAD&CG全国重点实验室