



第8届 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发

11月14-15日 | 深圳





2026 AI+ PRODUCT INNOVATION SUMMIT

01.16-17 · ShangHai

AI+产品创新峰会



Track 1: AI 产品战略与创新设计

从0到1的AI原生产品构建

论坛1: AI时代的用户洞察与需求发现

论坛2: AI原生产品战略与商业模式重构

论坛3: Agentic AI产品创新与交互设计

2-hour Speech: 回归本质



用户洞察的第一性

——2小时思维与方法论工作坊

在数字爆炸、AI迅速发展的时代，
仍然考验“看见”的“同理心”



Track 2: AI 产品开发与工程实践

从1到10的工程化落地实践

论坛1: 面向Agent智能体的产品开发

论坛2: 具身智能与AI硬件产品

论坛3: AI产品出海与本地化开发

Panel 1: 出海前瞻



“出海避坑地图”圆桌对话

——不止于翻译:

AI时代的出海新范式



Track 3: AI 产品运营与智能演化

从10到100的AI产品运营

论坛1: AI赋能产品运营与增长黑客

论坛2: AI产品的数据飞轮与智能演化

论坛3: 行业爆款AI产品案例拆解

Panel 2: 失败复盘



为什么很多AI产品“叫好不叫座”?

——从伪需求到真价值:

AI产品商业化落地的关键挑战

智能重构产品 数据驱动增长
Reinventing Products with Intelligence, Driven by Data

80+

业界大咖

汇聚产品大咖的真知灼见

40+

创新案例

开拓全新AI+产品视角

10+

分论坛

涵盖产品到商业增长的全链路

800+

听众规模

共同探索产品的智能重构



扫码查看会议详情

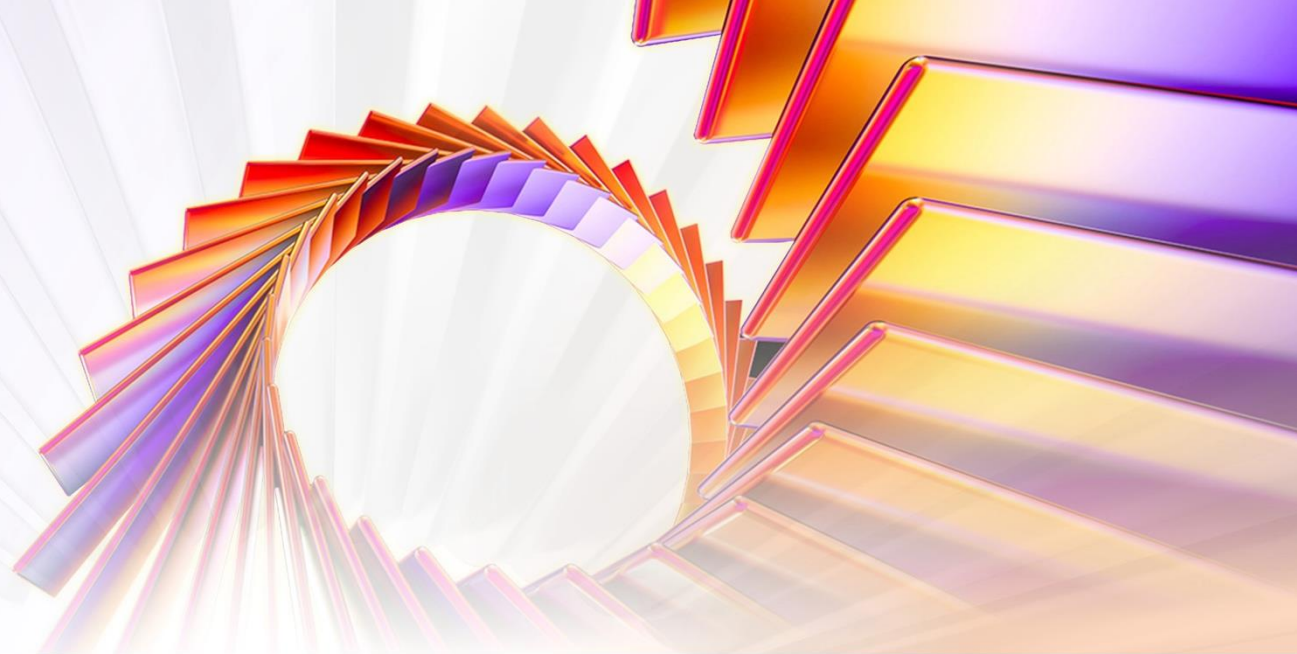


| 11月14-15日 | 深圳

第8届 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发



具身智能时代的数据集管理与生成

郝祁 | 南方科技大学



郝祁

IEEE 高级会员

南方科技大学计算机系教授、斯发基斯研究院执行院长、深圳市机器人视觉与导航重点实验室副主任，一直从事智能感知、无人驾驶、智能机器人方面的研究，研究项目获得了国自然、英特尔、华为等基金与企业的资助。目前已经在相关领域发表期刊与会议文章130+。多次担任国自然、科技部、美国科学基金、美国能源部先进项目评委。

目录

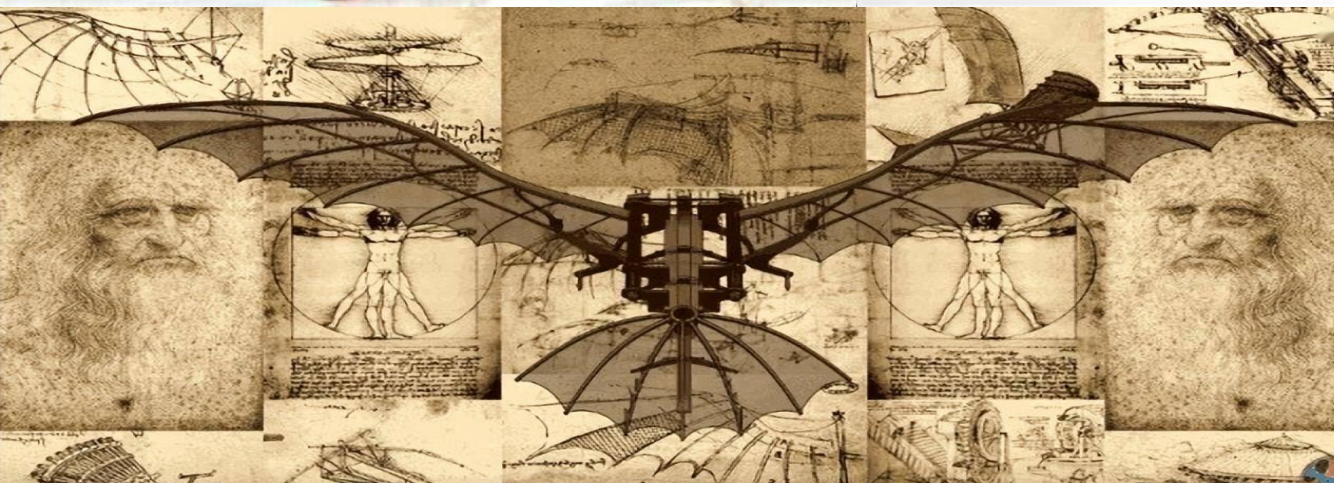
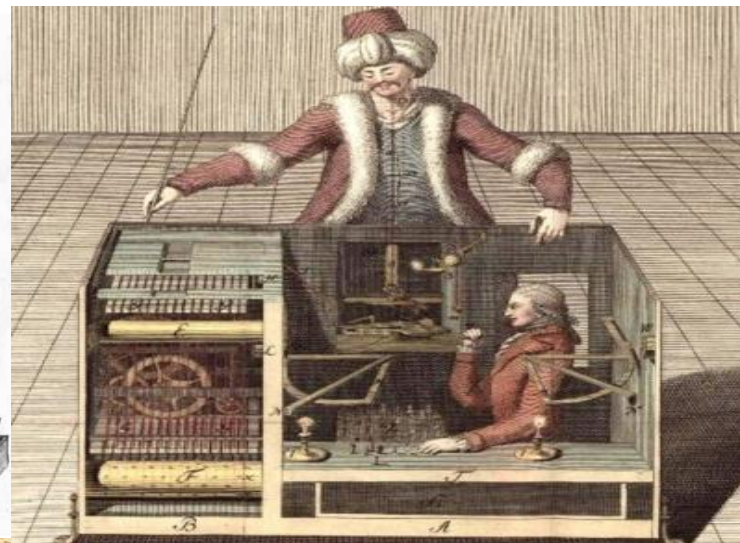
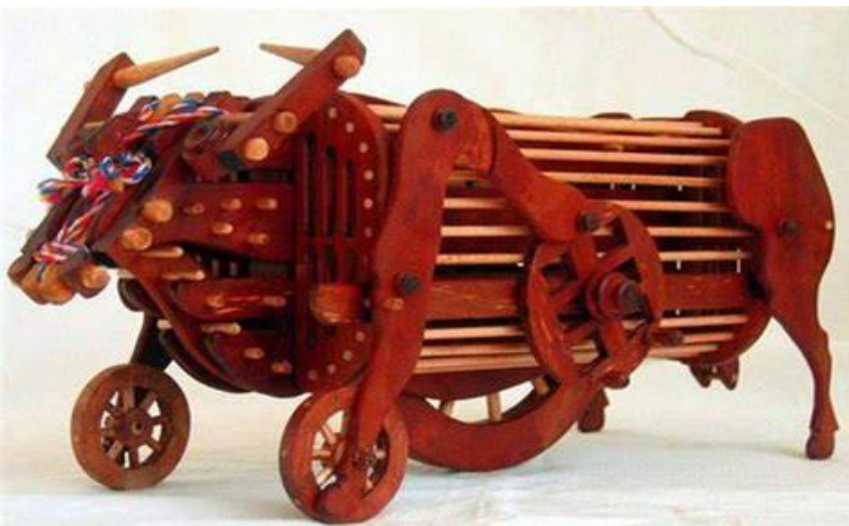
CONTENTS

- I. 具身智能技术发展
- II. 数据集面临的挑战
- III. 评估与优化
- IV. 标注与可视化
- V. 检索与生成
- VI. 总结与展望

PART 01

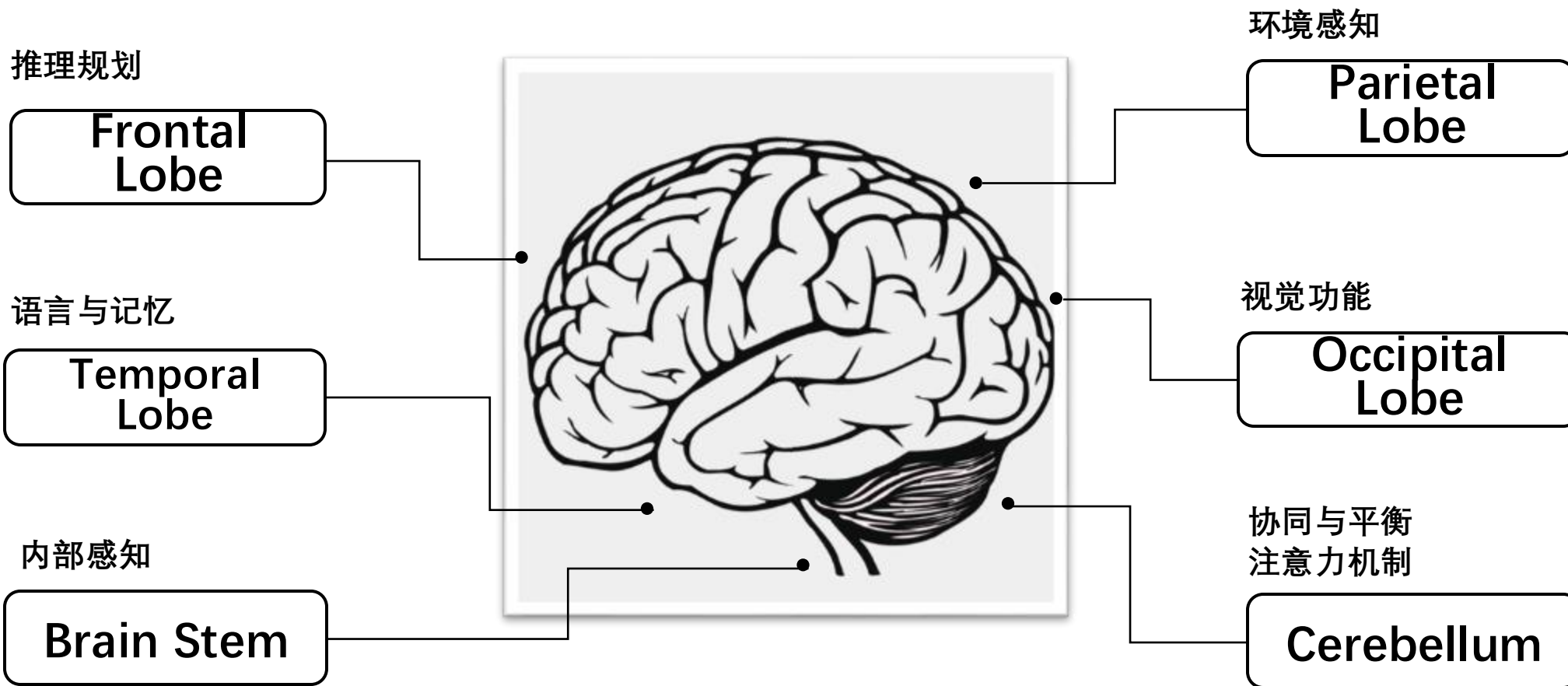
具身智能技术发展

▶▶ 人工智能技术起源



模仿生物与人类，实现有卓越能力的机械装置机器

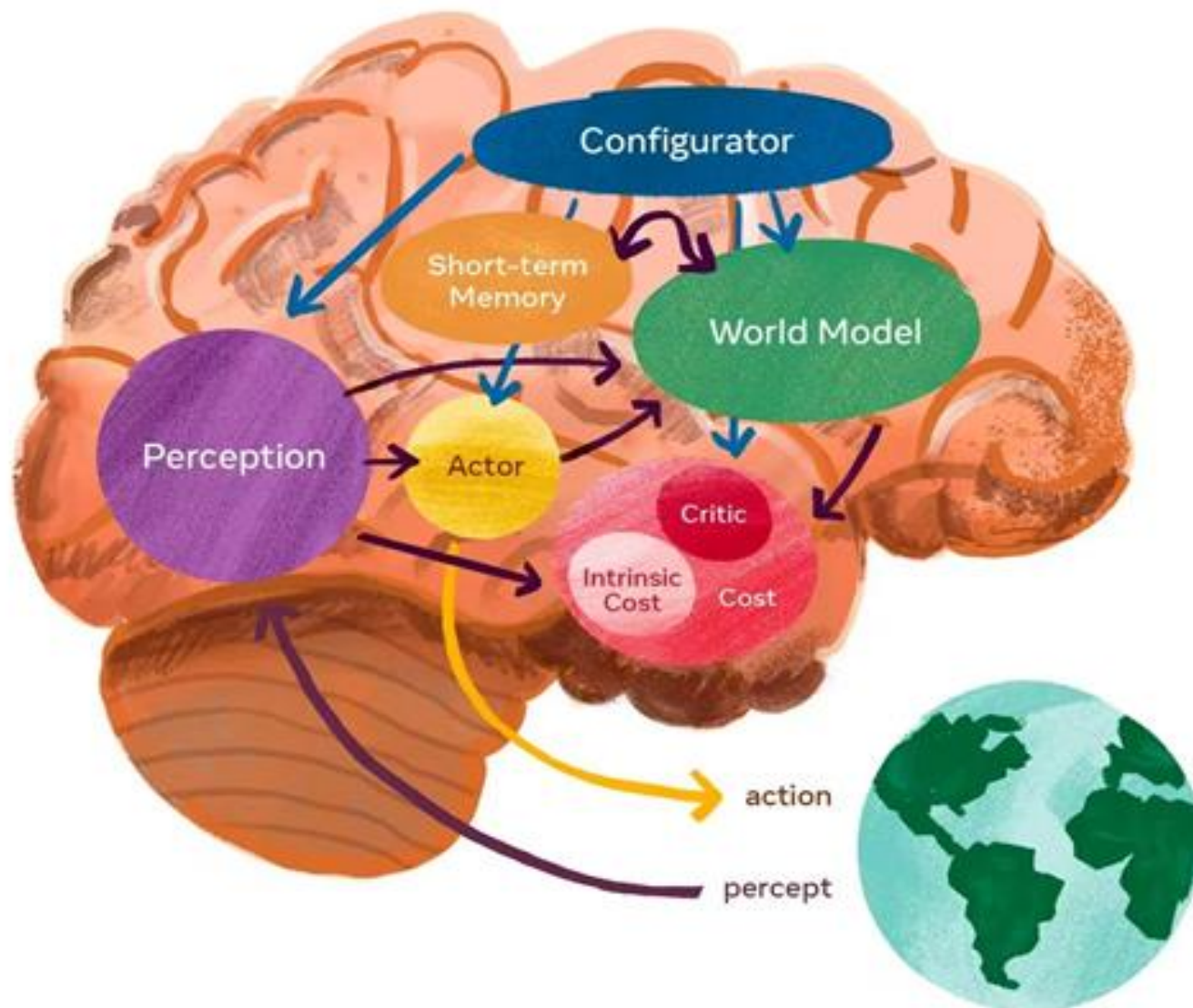




大语言模型(VLM)、视觉语言模型(VLM)、视觉语言导航模型(VLN)
视觉语言行动模型(VLA)、世界行动模型(WA)



世界行动模型



CONFIGURATOR

针对任务配置各个模块

PERCEPTION

估计世界状态

WORLD MODEL

预测世界未来状态

COST

“不舒适”程度成本

ACTOR

寻找最优行动序列

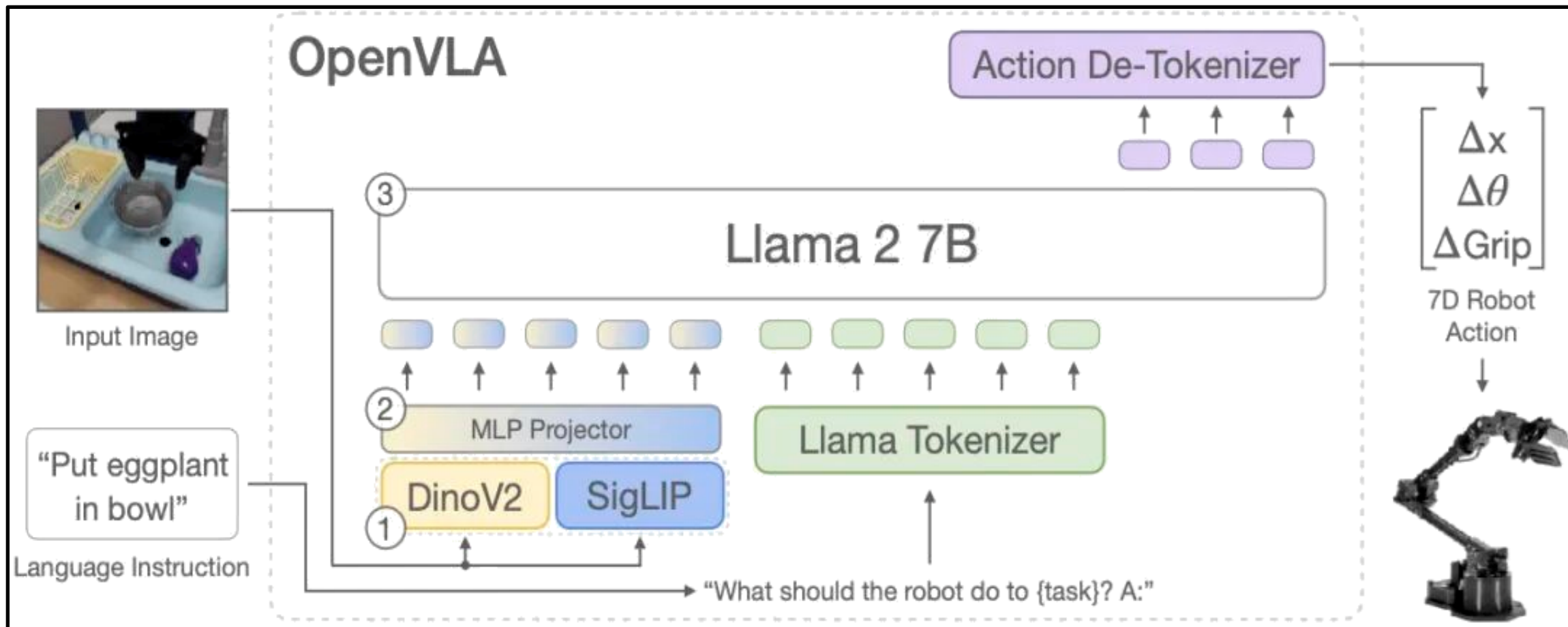
CRITIC

估算行动成本

SHORT TERM MEMORY

存储状态-成本事件记录

► 视觉语言行动模型



视觉语言行动模型 (VLA)：输入包括图像与文字
输出包括面向机器人的行动命令；基于大语言模型 (VLM)

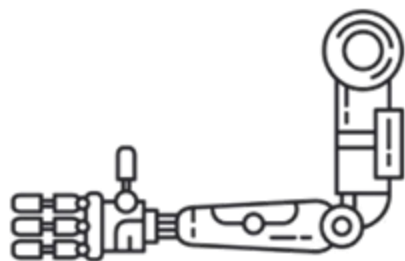
具身人工智能

- 在物理世界中运行并**交互**
- 与环境感知-运动的闭环**耦合**
- 持续学习与**进化**
- 可以通过直接设置目标来**控制**



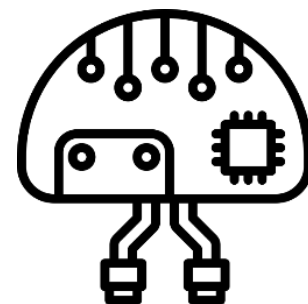
感知

环境感测的能力



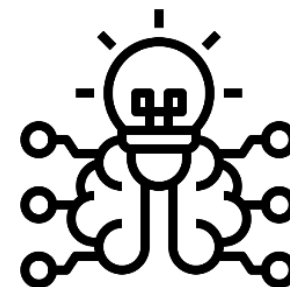
行动

交互与改变环境的能力



记忆

保留过往经验的能力



学习

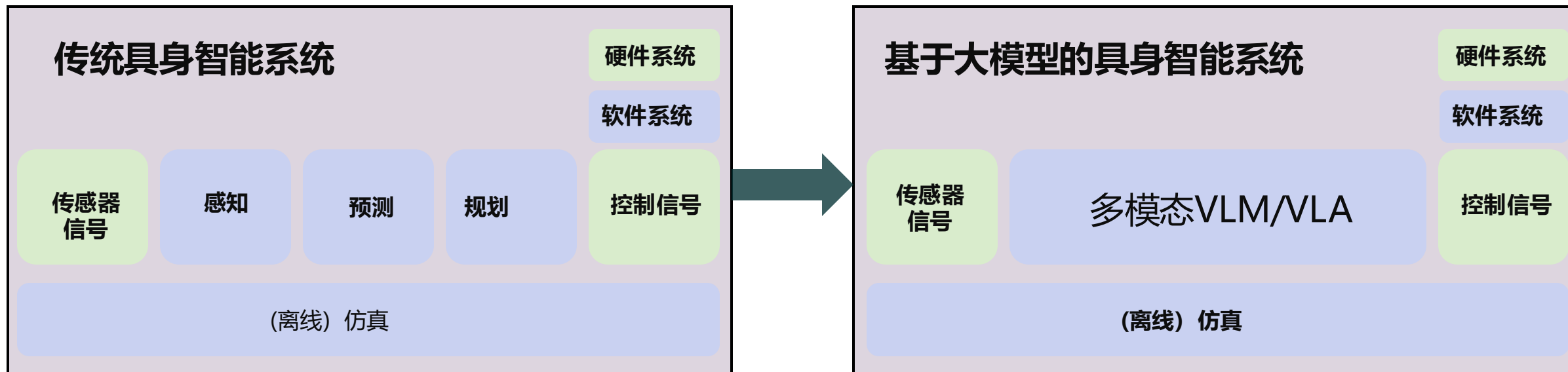
形成新知识与新能力的的能力

经典人工智能

- 在**静止**的数据集上训练
- 与物理世界完全**隔绝**/脱节
- 并非为理解**因果关系**而设计
- 仅能识别**浅显**的上下文和关联性
- 无法适应新的物理世界，需要**重新训练**



具身智能技术发展

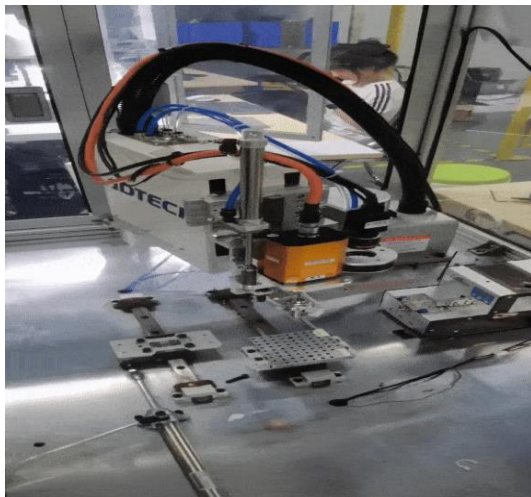


► 具身智能技术主要应用

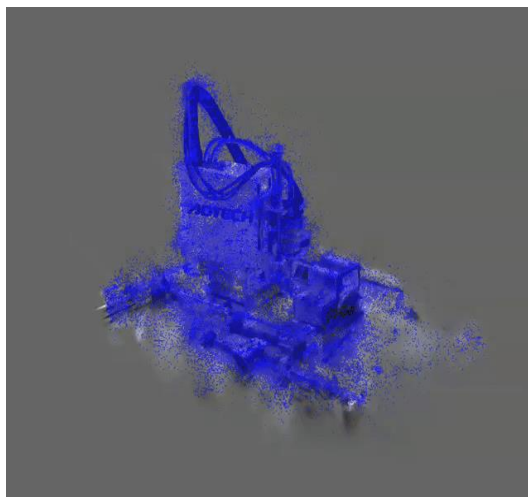


► 具身智能建立仿真环境 (real-to-sim)

真实场景视频采集



3D场景重建 (3DGS)



运动部分拆解：桌面基座/机械臂

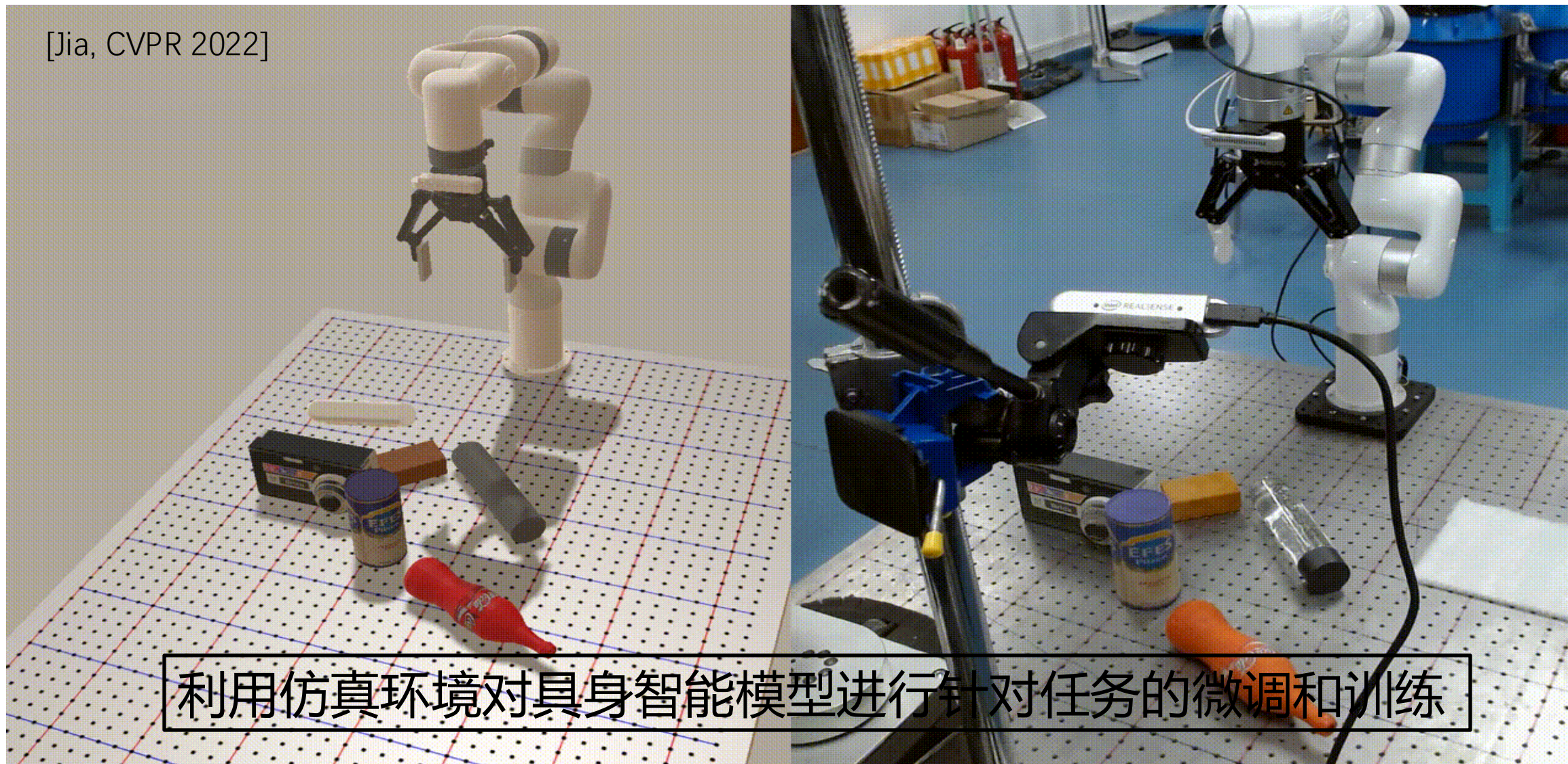


运动控制/多视角相机渲染



► 具身智能仿真训练(sim-to-real)

[Jia, CVPR 2022]

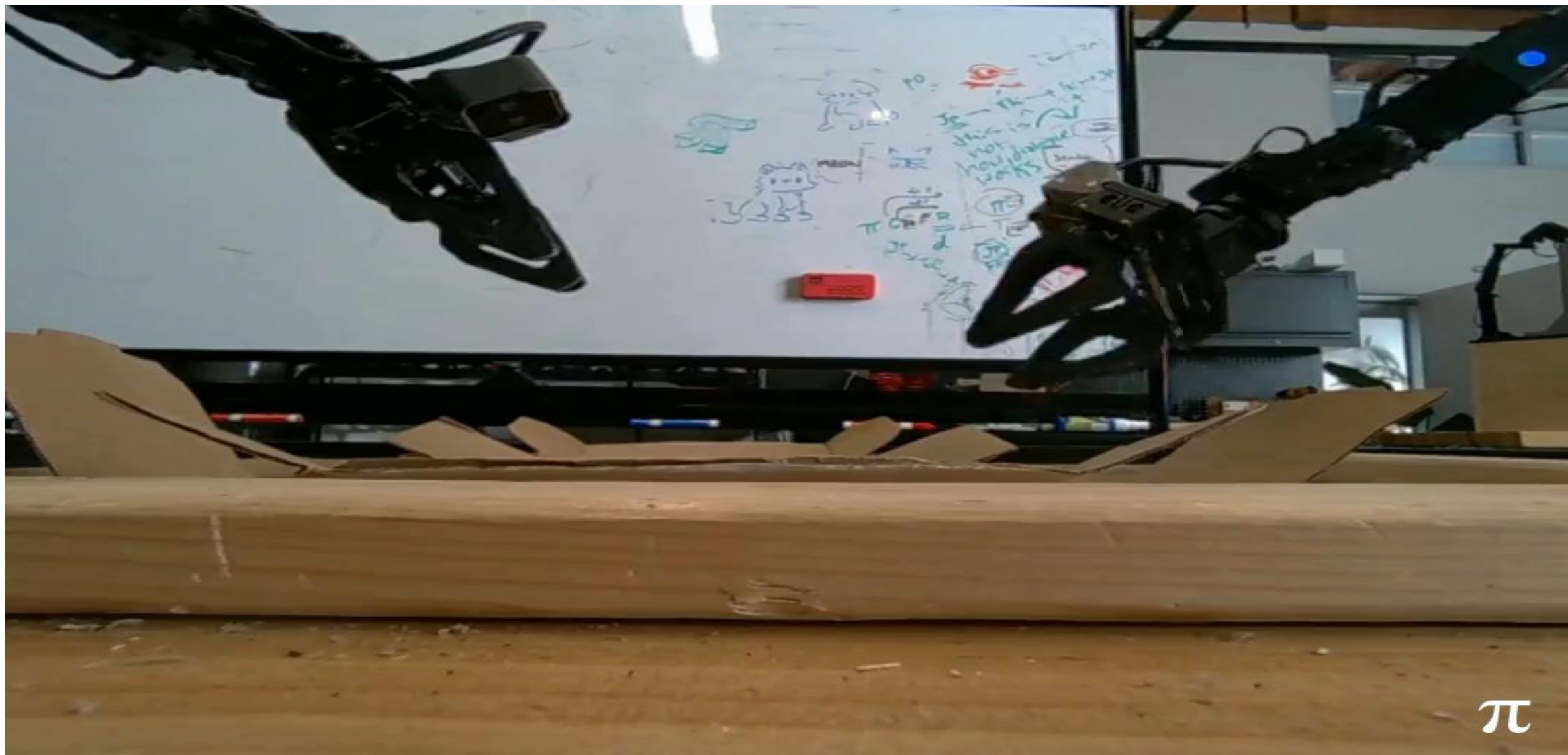




Intel Photo-realistic Simulator for Embodied AI



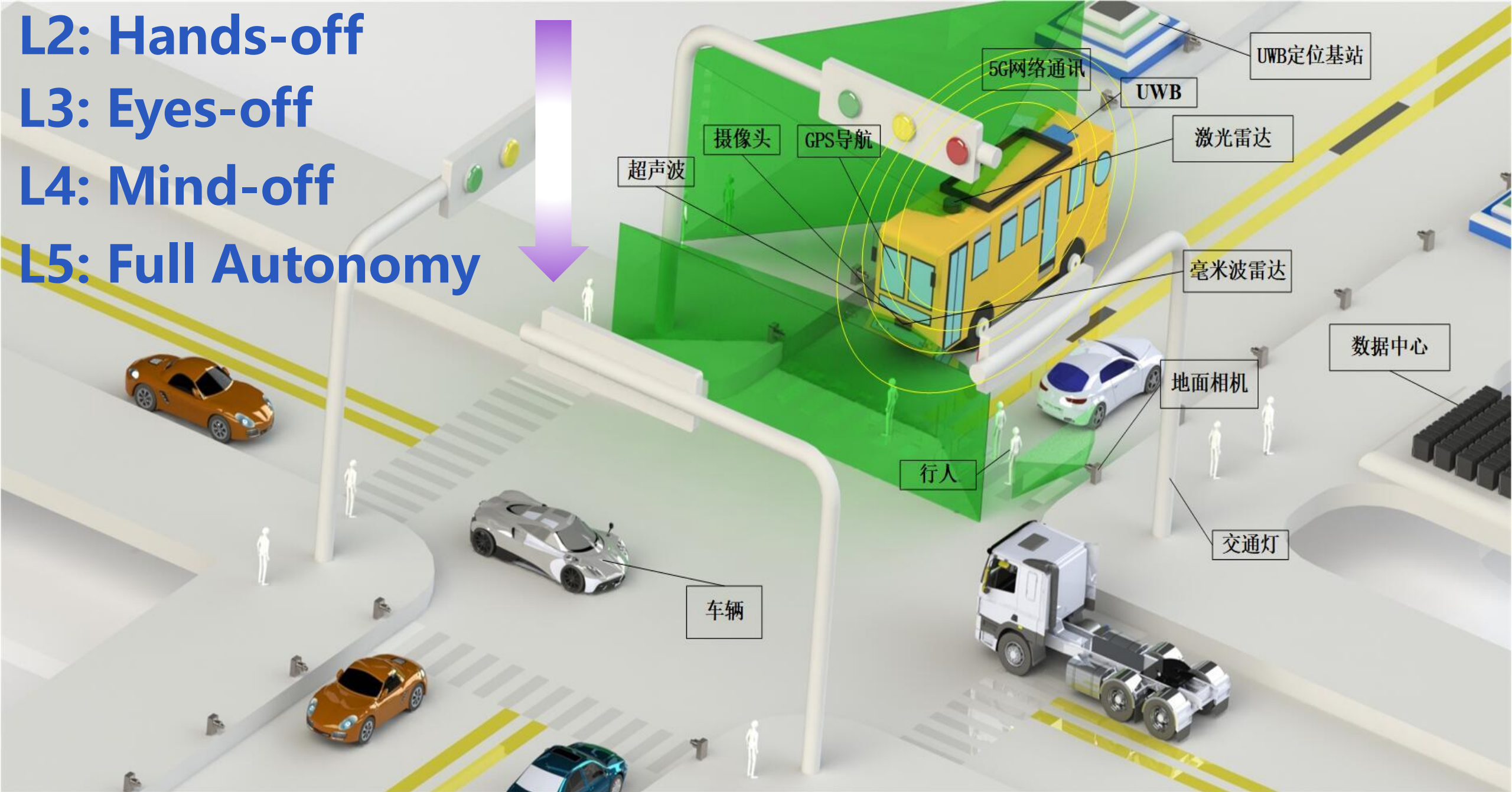
► 具身智能训练(data-to-real)



π



L2: Hands-off
L3: Eyes-off
L4: Mind-off
L5: Full Autonomy



L4: Public Autonomous Valet Parking 公共停车场无人代客泊车

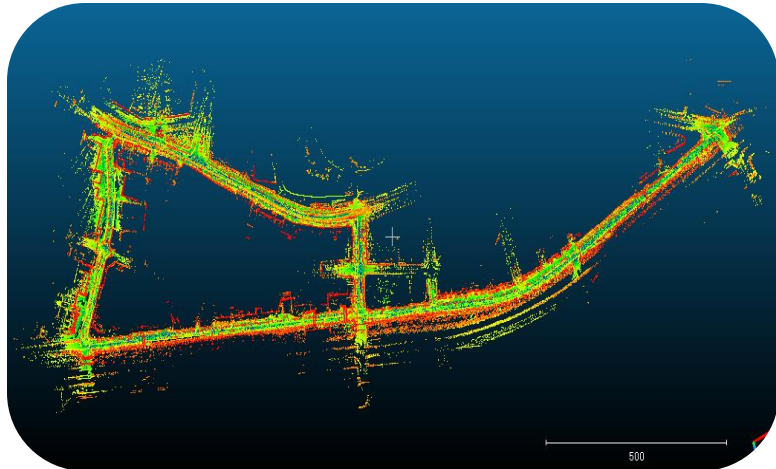


Last half a km
最后500米

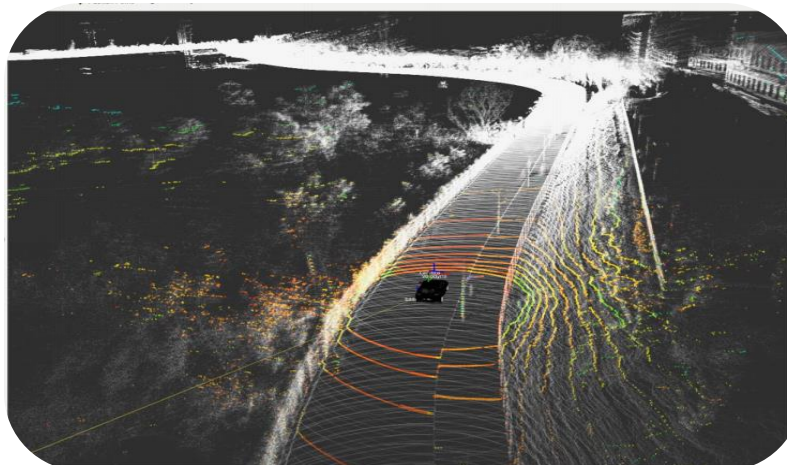


用户需求：
快速与安全自主泊车

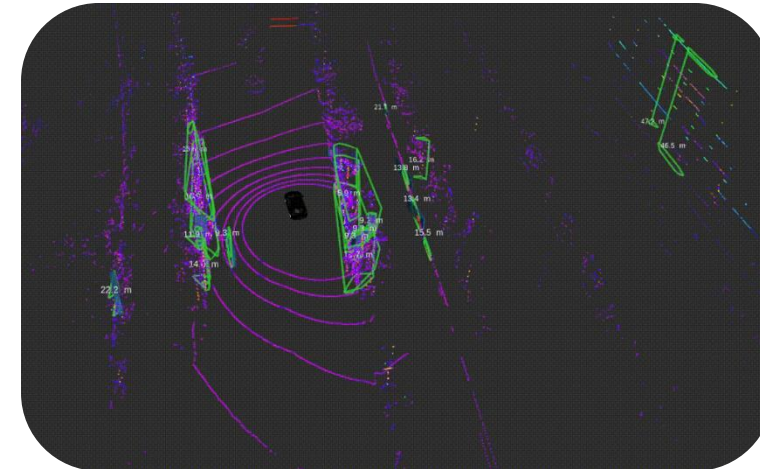
► 无人驾驶主要功能



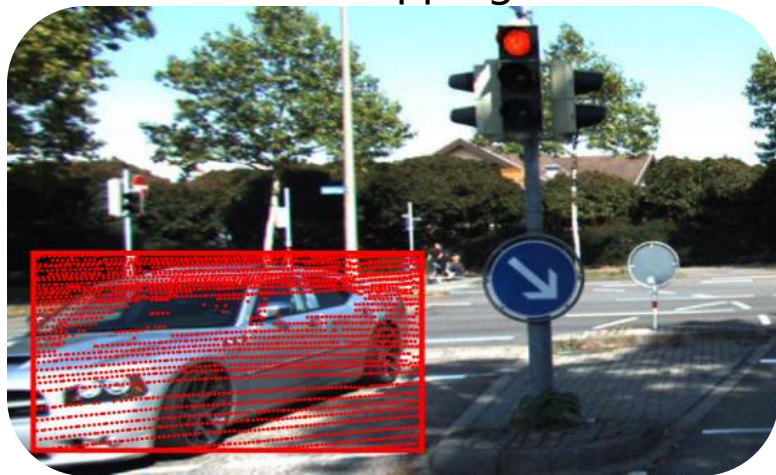
3D地图建图
Mapping



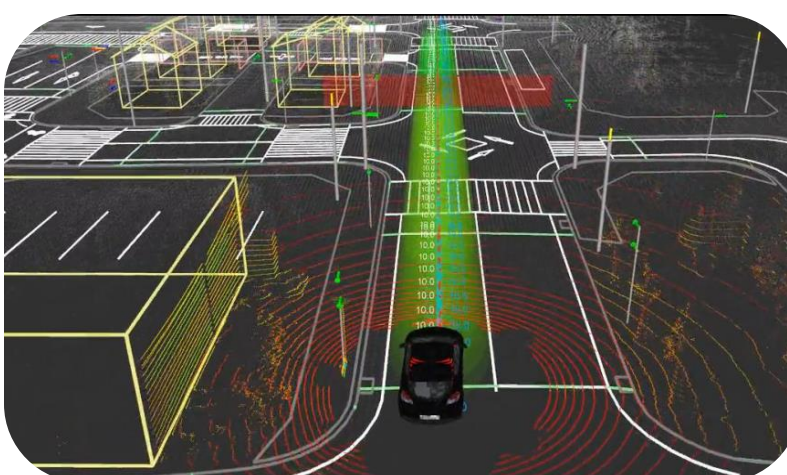
3D地图匹配定位
Localization



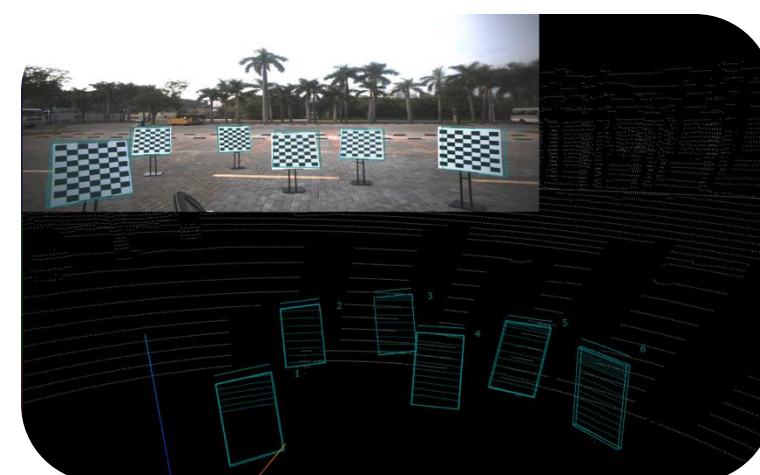
3D物体检测与运动预测
Detection & Prediction



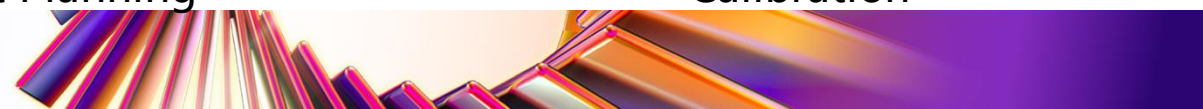
2D-3D融合目标识别
Recognition



运动决策与轨迹规划
Decision & Planning



多传感器标定与校准
Calibration



► 无人驾驶数据集训练与验证



1

传感阵列 Sensor Array

- 主动/被动
- 视觉/听觉
- 毫米波/激光雷达
- 器件选型
- 阵列位置
- 连接安装
- 同步测量
- 校准标定



2

数据采集 Data Collection

- 复杂度/多样性
- 异常事件
- 反常事件
- 复杂场景
- 极端天气
- 信息增益
- 数据采样
- 智能/随机采样



3

数据存储 Data Storage

- 数据压缩
- 数据冗余
- 数据分布
- 数据安全
- 数据隐私
- 标注/格式
- 云/雾服务
- 边缘计算



4

模型更新 Model Learning

- 参数优化
- 特征/模型融合
- 主动学习
- 半/弱监督学习
- 非监督学习
- 增强学习
- 高精建图
- 动态地图



5

模型应用 Applications

- 高精定位/里程计
- 实例/语义分割
- 目标检测与识别
- 目标追踪与管理
- 场景理解与预测
- 数据与算法评估
- 资源调度与分配
- 车辆遣调



► 无人驾驶仿真训练与验证



1

传感仿真 Sensing Sim.

- 可见光/红外相机
- 毫米波/激光雷达
- 声呐传感器
- 阵列位置
- 参数设置
- 天气光照
- 传感物理建模
- 各类感知算法



2

运动仿真 Motion Sim.

- 运动学模型
- 动力学模型
- 流体模型
- 震动模型
- 摩擦模型
- 延迟模型
- 运动系统建模
- 决策与控制



3

静态场景 Static Scene

- 语义矢量地图
- 建筑模型
- 植物模型
- 交通标志模型
- 车辆感知定位
- 高精地图构建



4

动态场景 Dynamic Scene

- 运动车流
- 行人单车
- 多车交互
- 反常事件
- 复杂场景
- 多智能体系统
- 动态地图构建

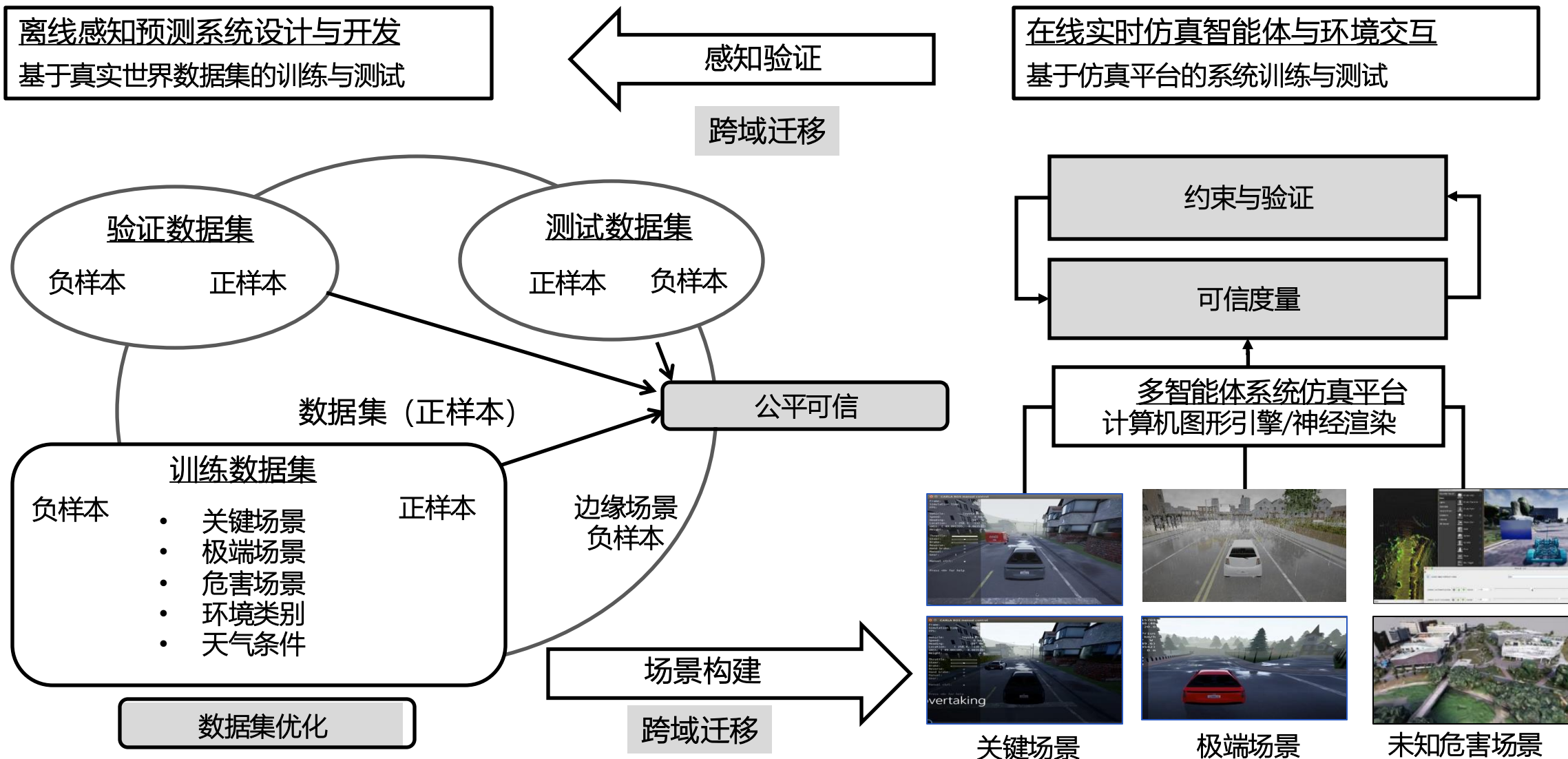


5

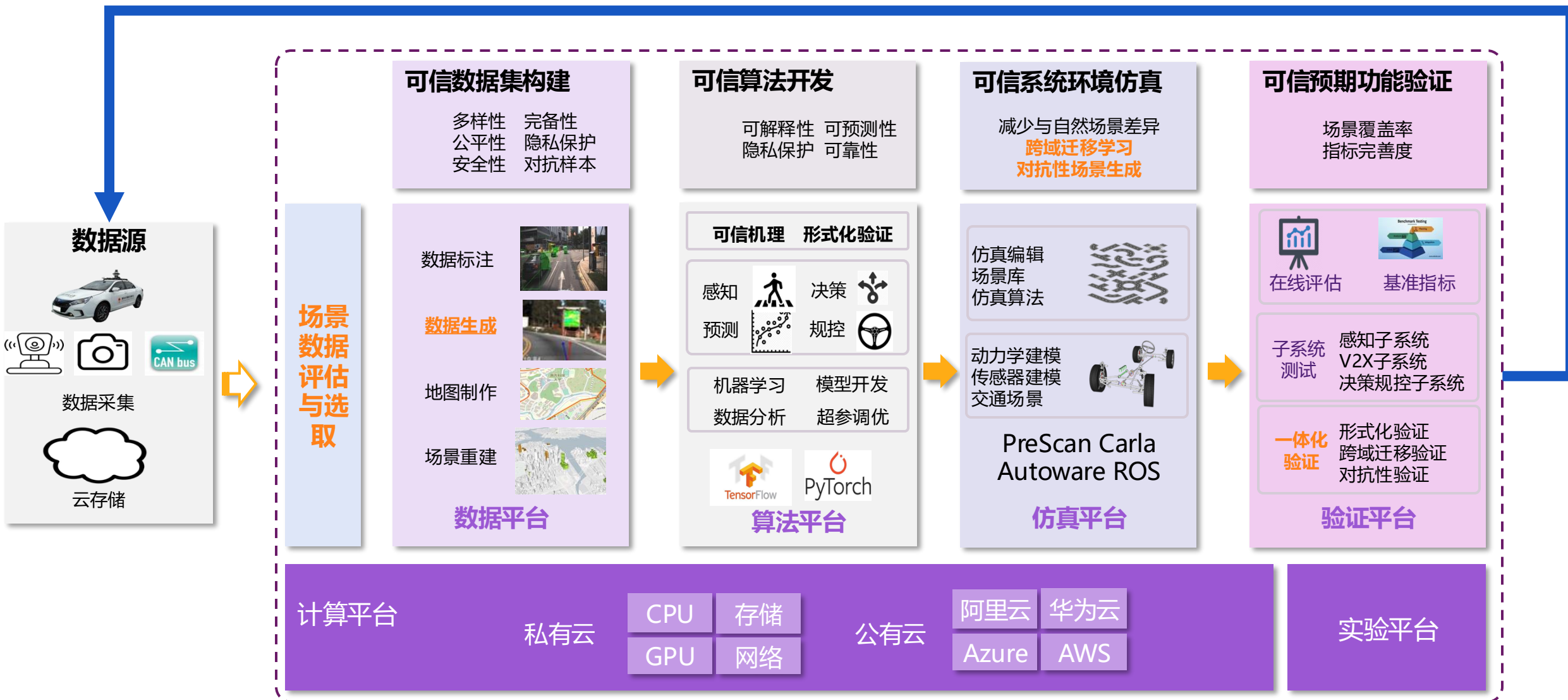
仿真应用 Applications

- 高精定位/里程计
- 实例/语义分割
- 目标检测与识别
- 目标追踪与管理
- 场景理解与预测
- 运动规划与避障
- 车联网车路协同
- 资源调度与分配
- 车辆遣调

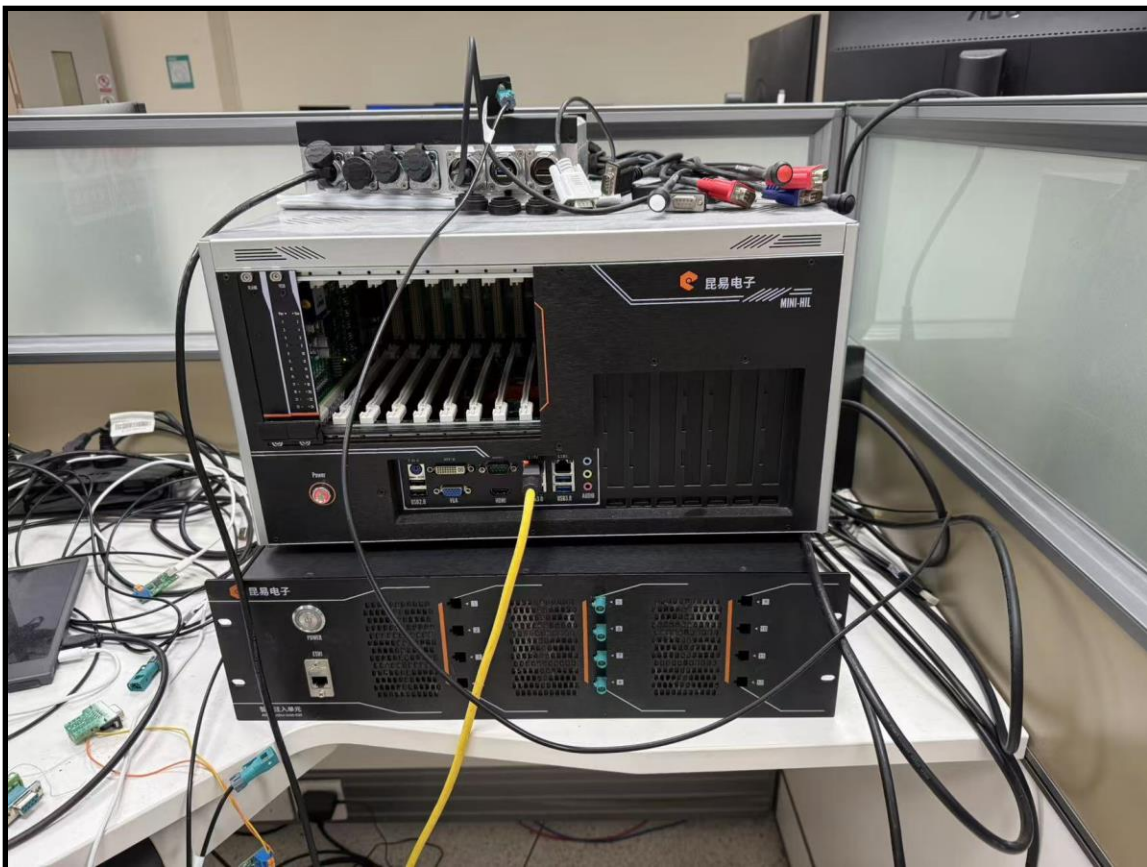
数据集与仿真一体化



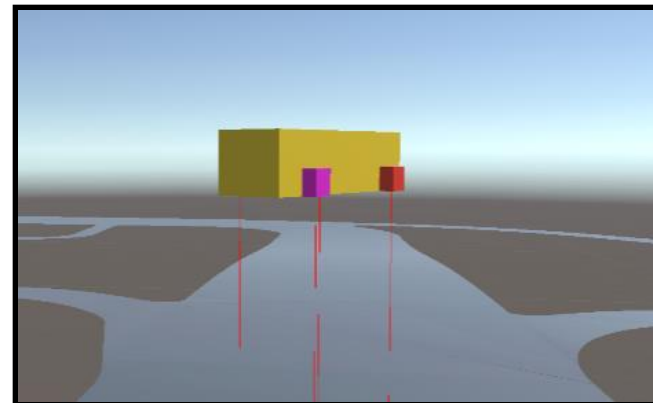
具身智能系统开发



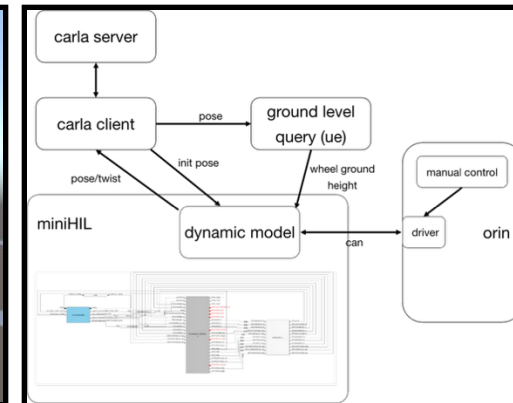
► 硬件在环训练与验证



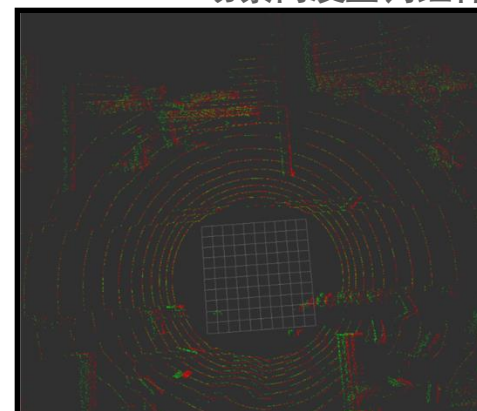
硬件在环仿真实体机



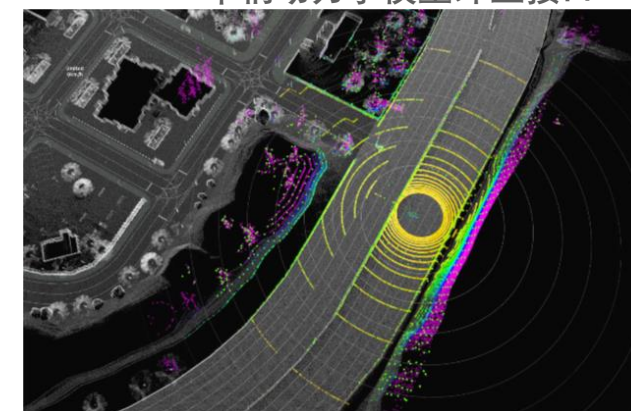
场景高度查询组件



车辆动力学模型外置接口



激光雷达仿真接口，速腾32



智能驾驶端实施效果

仿真硬件在环（Hardware In the Loop）将仿真器提供的虚拟环境接入真实的物理硬件，提供更加接近真实工况的训练与测试环境。



PART 02

建立具身智能数据集面临的挑战

具身智能数据采集



毫米波雷达



Tesla 人形机器人行动数据采集



无人驾驶数据集SUScape

- 1059 场景
- 每场景
 - 20s时长, 10Hz采集, 2Hz标注
 - 40帧标注数据
 - 40帧高密度点云, 160帧盲区雷达点云
 - 240张RGB图片
 - 240张红外图片
 - 240帧毫米波雷达点云
 - 360°全覆盖
- 1.26M 3D 框
- 60K 物体轨迹
- 36种目标类别
- 1.7M RGB图片2D框 (3D生成)
- 1.3M 红外图片2D框 (3D生成)
- 各种环境和道路类型: 快速路, 主干道, 街道, 住宅区, 隧道, 校园等
- 各种光照和天气: 晴/阴/雨(少量), 白天/夜间

[SUScape.net](https://suscape.net)



► SUScape物体3D样例

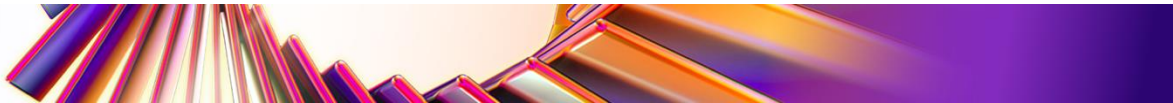


无人驾驶数据集比较

Dataset	Year	Scenes	Size(hr)	RGBimgs	IR imgs	PCslidar	PCs radar	ann.frames	3Dboxes	Night/Rain	Maplayers	Classes	Locations	Area(km^2)	LiDAR
KITTI	2012	22	1.5	15k		15k	0	15k	200k	No/No	0	8	Karlsruhe		1(64chan)
ASLidar (apolloscape)	2018	-	2	0		20k	0	20k	475k	-/-	0	6	China		
KAIST	2018	-	-	8.9k		8.9k	0	8.9k	0	Yes/No	0	3	Seoul		
H3D	2019	160	0.77	83k		27k	0	27k	1.1M	No/No	0	8	SF		
nuScenes	2019	1k	5.5	1.4M		400k (40k Key frm)	1.3M	40k	1.4M	Yes/Yes	11	23	Boston,SG	5	1(32chan)
Argoverse	2019	113	0.6†	490k†		44k	0	22k†	993k†	Yes/Yes	2	15	Miami,PT	1.6	
LyftL5 (perception)	2019	366 (60-90m)	2.5	323k		30K	0	55k	1.3M	No/No	7	9	Palo Alto		1+2(40chan)
WaymoOpen	2019	1k	6.4	1M		230k	0	230K	12M‡	Yes/Yes	0	4	3xUSA	76	1(64chan)+4
A*3D	2019	n/a	55	39k		39k	0	39k	230k	Yes/Yes	0	7	SG		
A2D2	2019	n/a	-	-		12k	0	12k	-	-/-	0	14	3xGermany		
BLVD (西安交大)	2019	654	3.3			?		?	249k	yes/no		3	China		
ONCE (华为)	2021	581	144	7M		1M(16k)		16k	417K	Yes/Yes		5	China	210	
IDD3D	2022			95k		15.5k		15.5k	223k			10(17)	印度		
SUScape	2023	1059	5.9	254k	254k	42.3k	254k	42.3k	1.26M	Yes/no	0	36	深圳	37	1(128chan)+4(32chan)

- 1. 根据带标注帧数指标，除Waymo外，规模与nuScenes/Lyft相当
- 2. 首个高密度(128线)的LiDAR数据集
- 3. 首个带红外相机数据

SUScape.net



场景数据标注与可视化

如何对场景进行自动标注，协助监督学习
如何对场景可视化，协助理解和评估

场景数据评估与优化

如何建立评价场景价值指标体系
如何优化数据集提升模型性能训练上限

价值场景数据检索

如何在大规模数据集中利用文字或者图像
或者视频搜索到场景多模态数据

价值场景数据生成

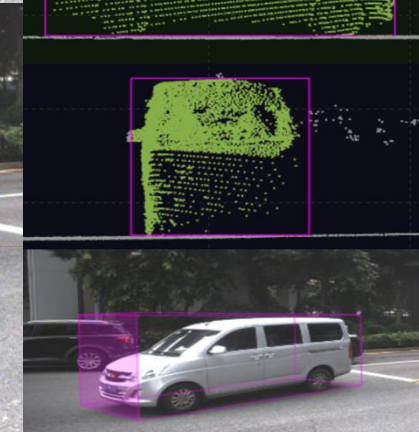
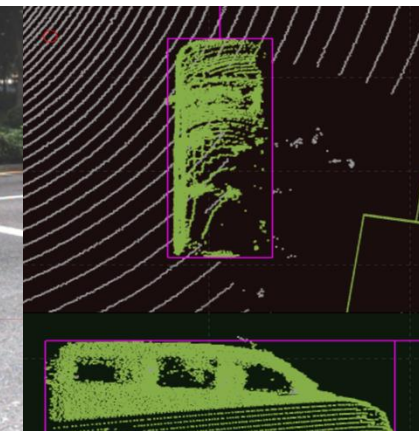
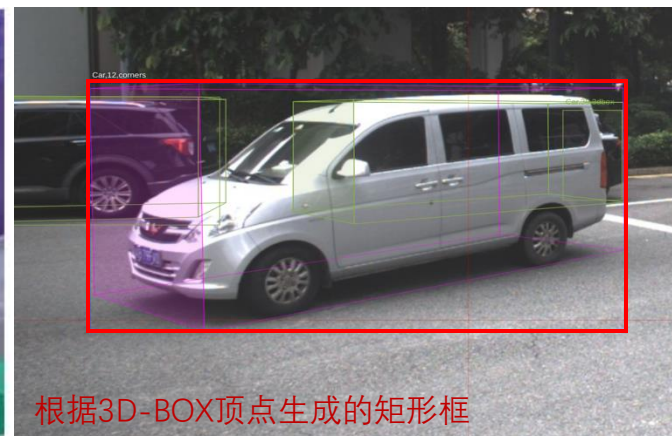
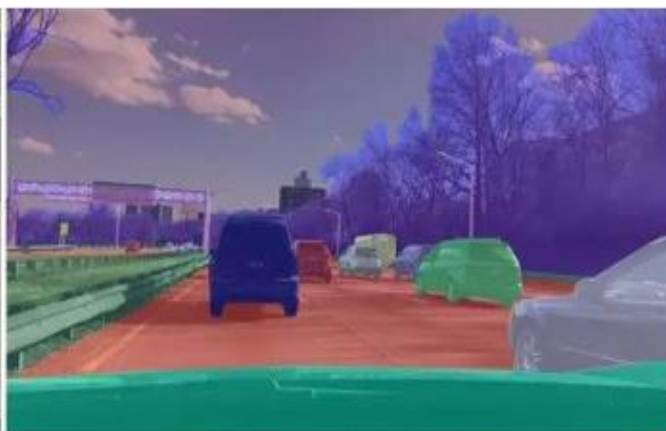
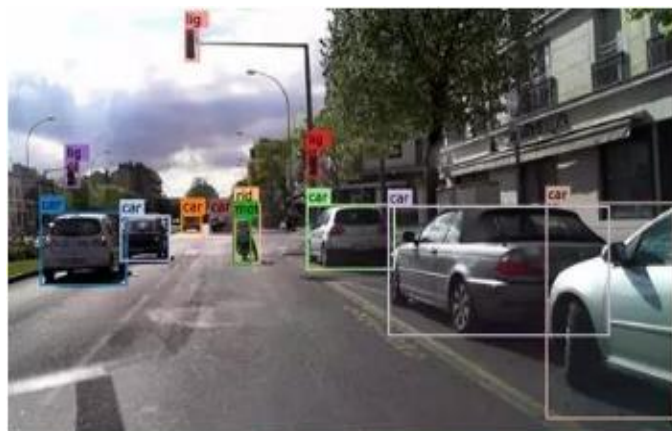
如何利用已有的场景生成各类极端、危害场景
提高对智能系统安全边界估计的可信性



PART 03

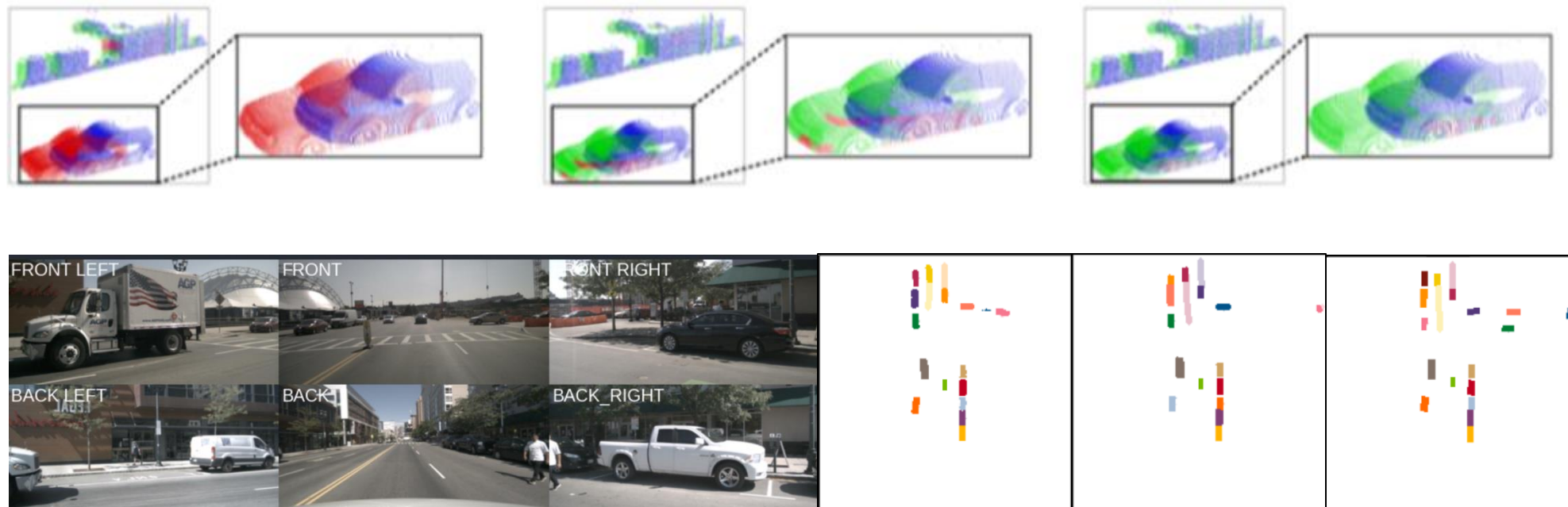
具身智能数据标注与可视化

▶ 传统数据集标注



面向交通标志识别，3D目标识别追踪、自车定位算法的数据标注





面向场景流，占用网格地图预测算法的数据标注

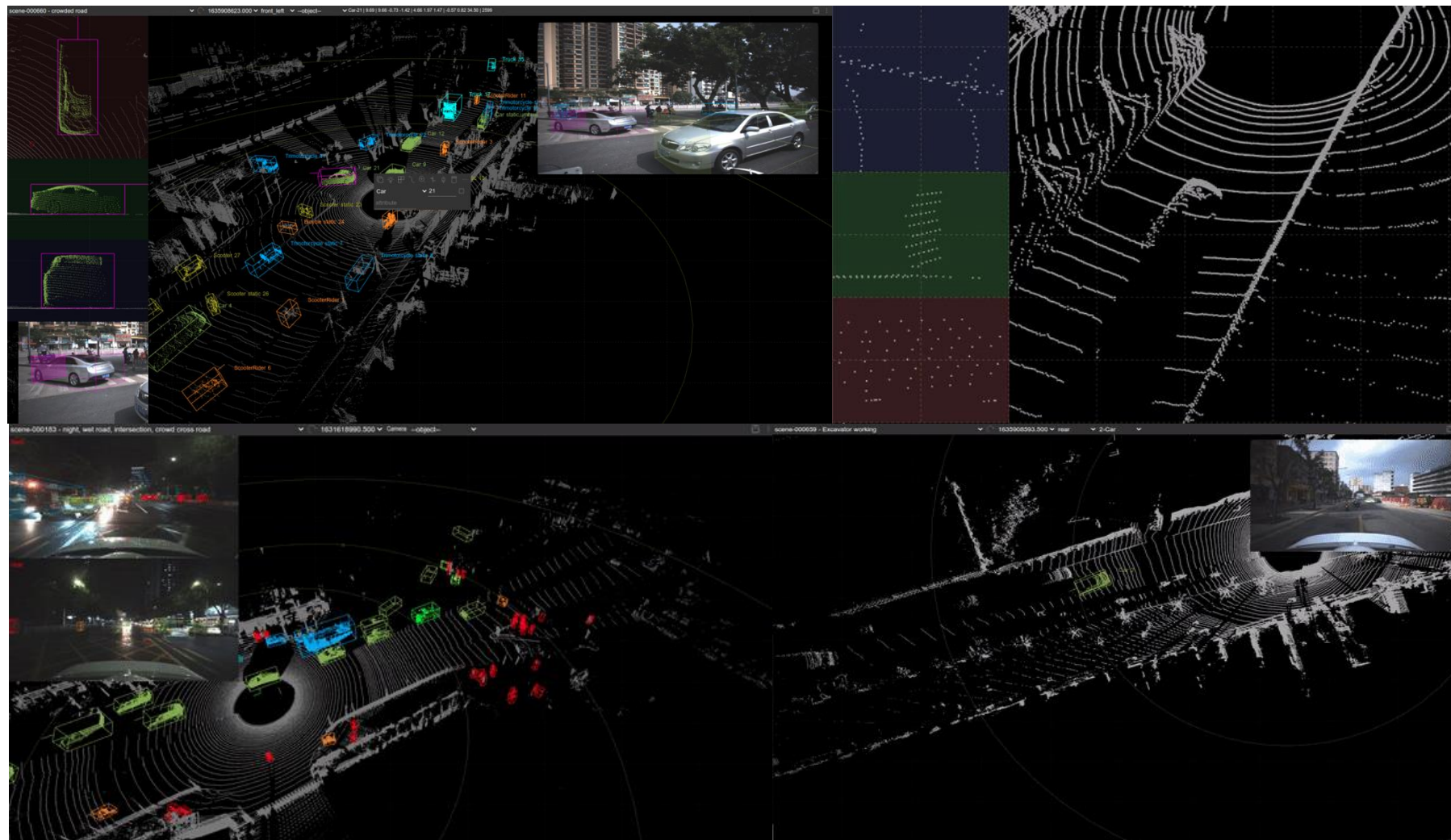


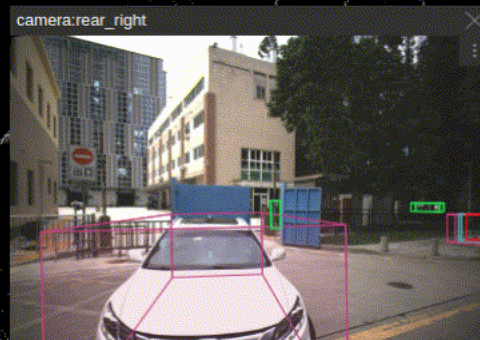
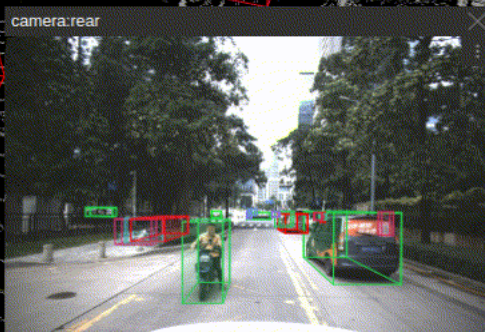
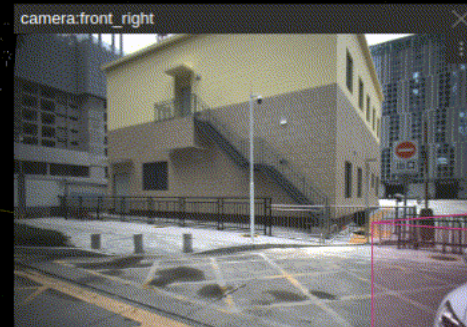
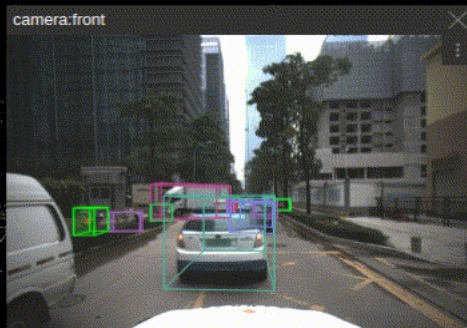
▶ 3D数据标注平台: POINTS

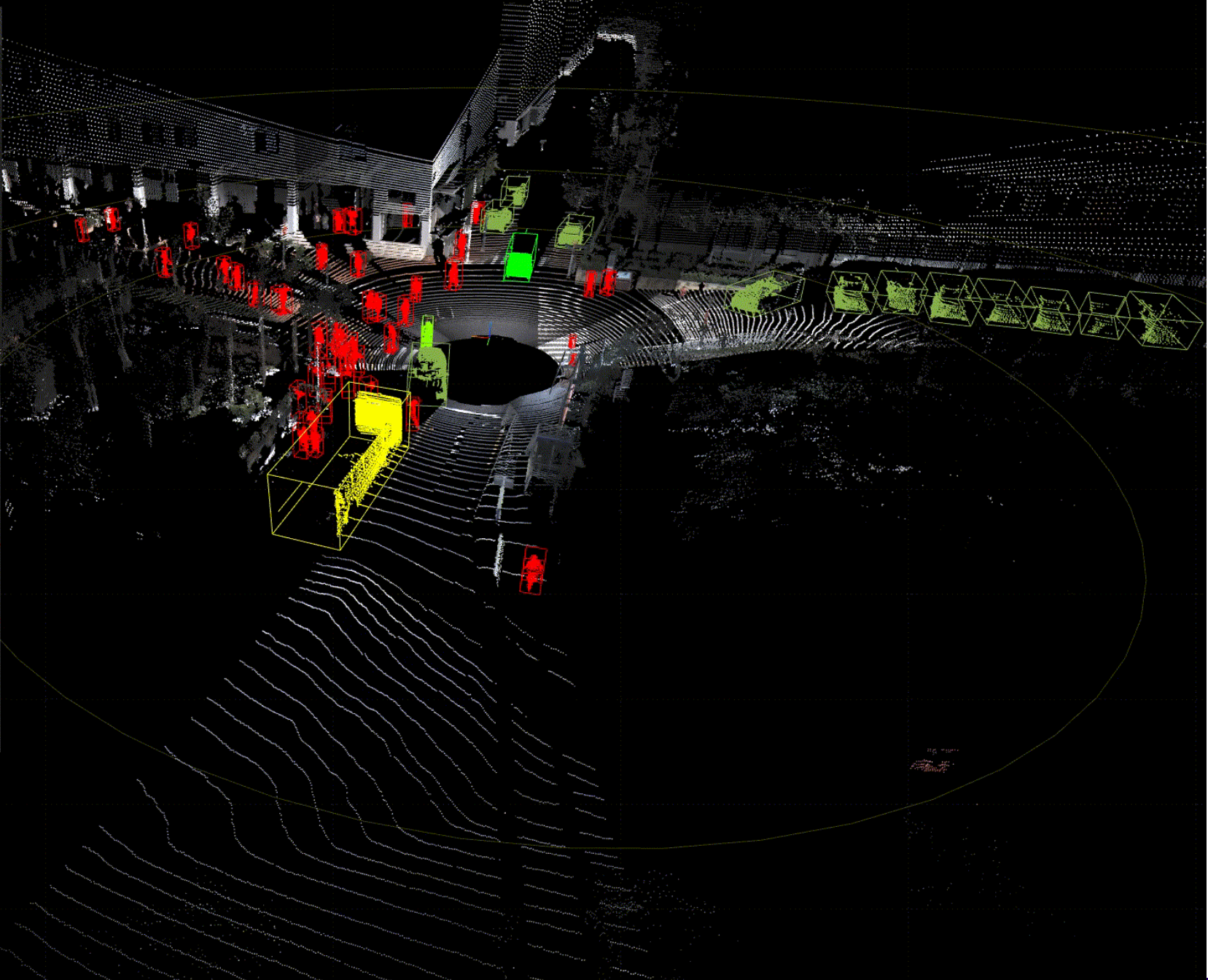
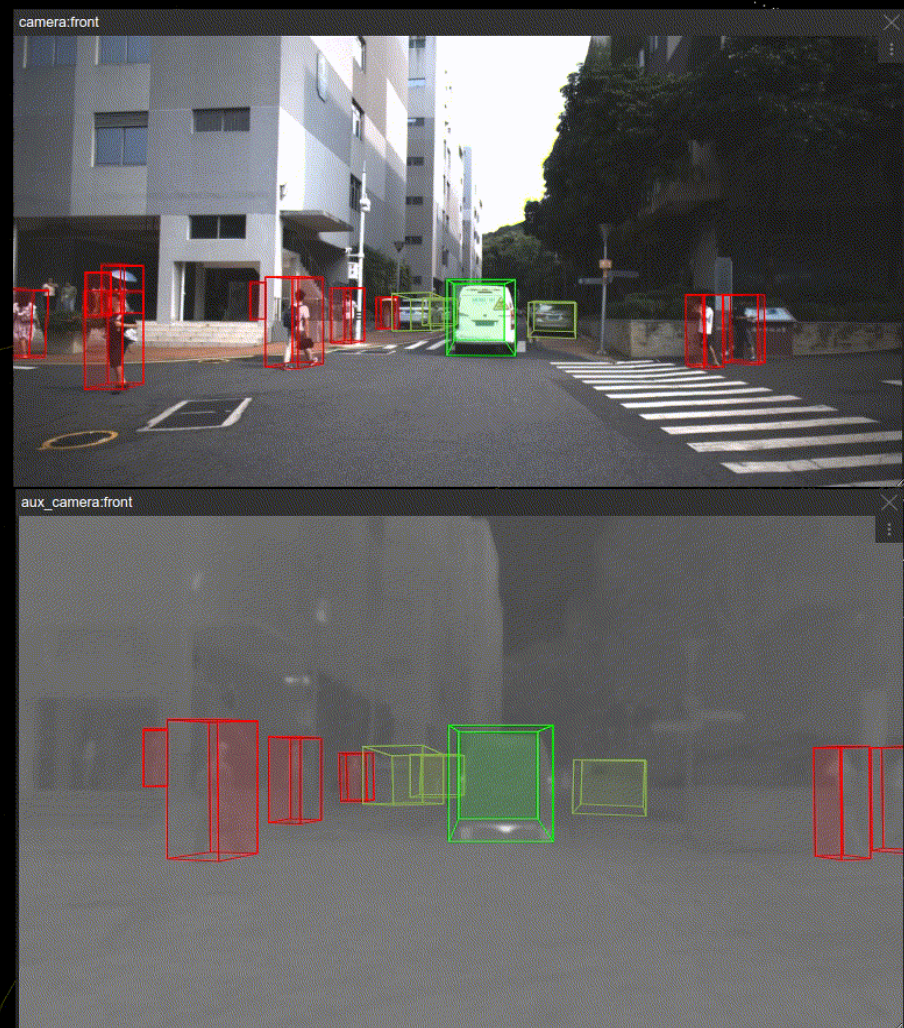
• SUSTech POINTS

- 对3D框进行9自由度编辑
 - ID/类别/属性编辑
 - 启发式/交互式的3D框自适应
 - AI辅助标注
 - 批量标注
 - 基于追踪的批量自动标注
 - 2D-3D融合标注
 - 错误检查
-
- IDD3D(印度), V2V4Read(v2v协作感知)等数据集使用本平台标注完成
 - 多个商业公司使用本平台进行点云标注
 - Github 600+ star
-
- 标注质量保证
 - 多轮交叉检查
 - 基于统计和帧间冗余信息的自动错误检查

<https://github.com/nauril/SUSTechPOINTS>







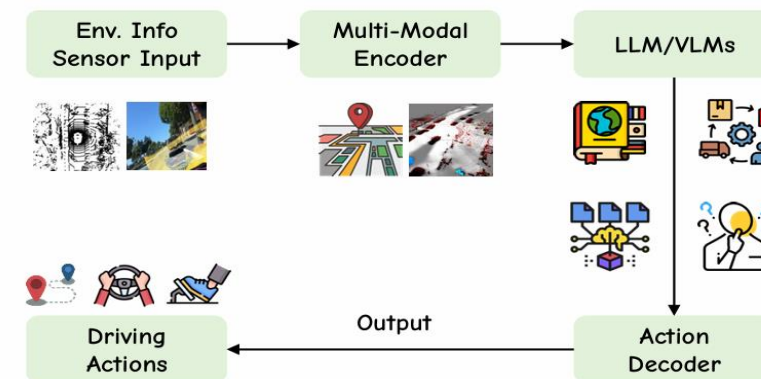
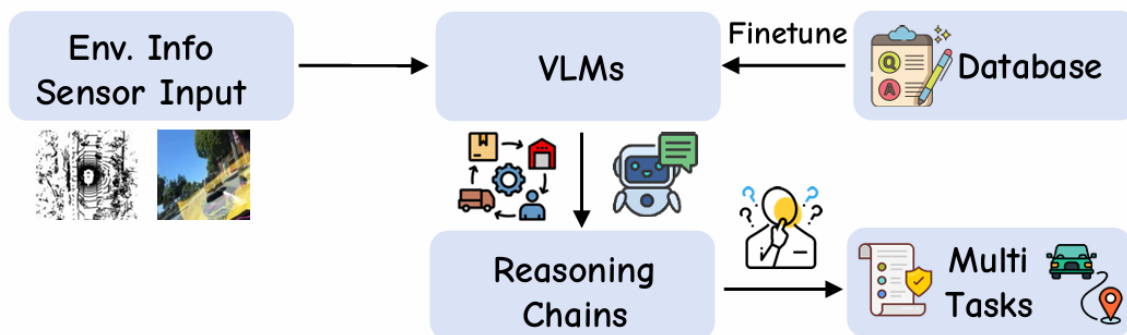
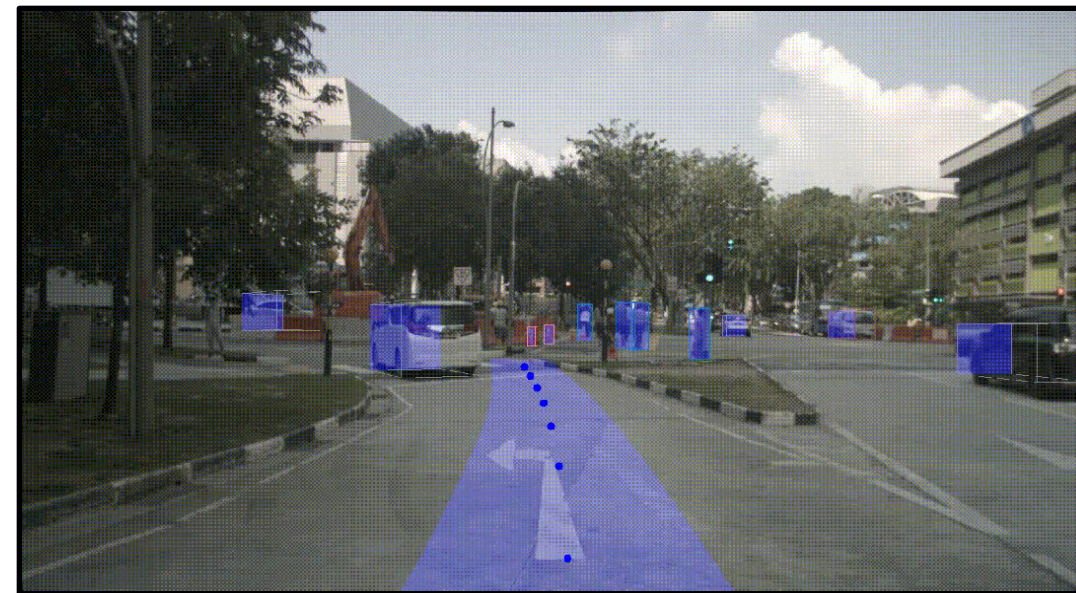
具身智能数据标注

How To Clean Snow off Your Car



[Input Instruction] Describe the video in detail.

The video starts with a car covered in snow, and a person is seen using a scraper to remove the snow from the car's windows. The person is seen scraping the snow off the windshield and the side windows of the car. The person continues to scrape the snow off the car, making sure that all the snow is removed. The video then shows the person wiping the snow off the car's windshield and side windows. The person is seen using a cloth to wipe the snow off the car's windows, making sure that the car is completely clean. The video ends with the person finishing the job and the car being shown off. Overall, the video shows a person removing snow from a car's windows and windshield, and then wiping the snow off the car's windows to make sure it is completely clean.



面向视觉语言模型(VLMs) 与视觉语言运动模型 (VLA) 的数据集标准



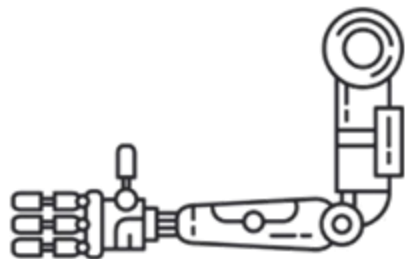
具身智能数据标注

- 人工标注
- 大模型辅助标注 (LLM/VLM)
- 群体标注 (多个大模型)
- RLHF标注 (强化学习人类反馈)



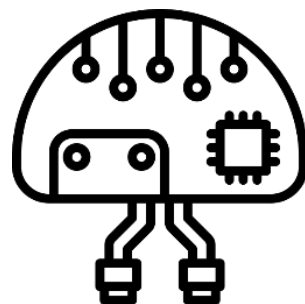
冷启动

人工标注小样本
微调初始大模型



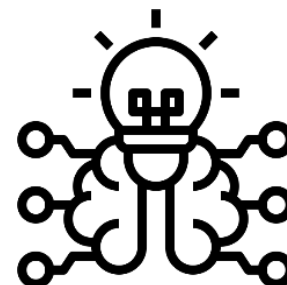
排序奖励

大模型对数据生成多个候选
人工排序训练奖励模型



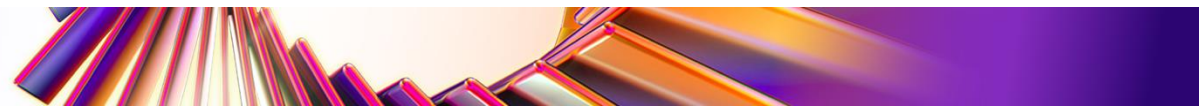
策略优化

用已有策略生成标签
利用奖励模型评分并优化策略



交叉检验

通过多个标注方法融合
提升标注质量

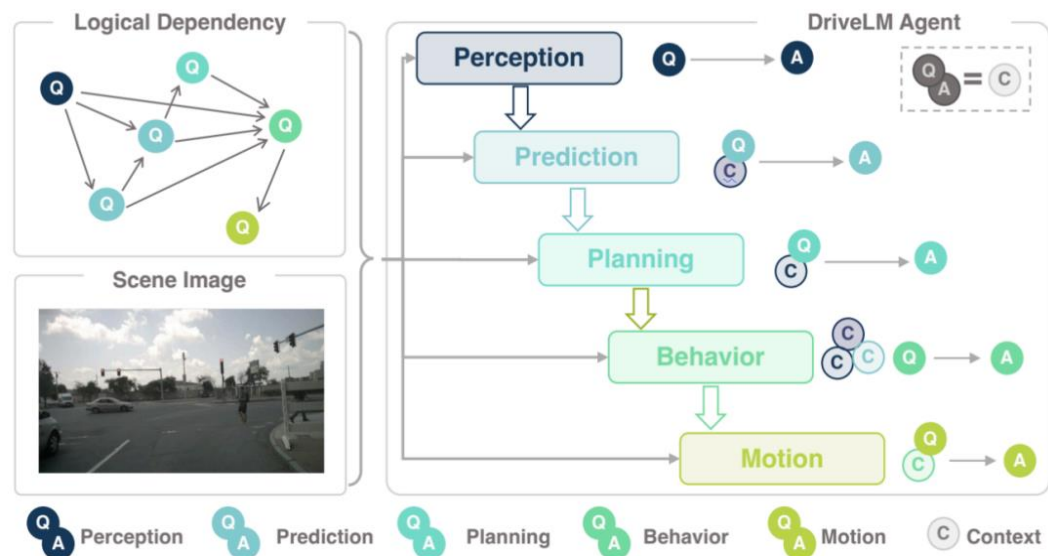


具身智能数据集标注：大模型工具

Feature Dimension	VAST (Tencent)	VideoLLaVA (PKU)	VideoLLaMA2 (Alibaba)	Video-ChatGPT (MSR)	GPT-4V (OpenAI)	Gemini (Google)
Core Idea	A unified video-language foundation model for various downstream tasks.	Extends LLaVA (Image) to the video domain, adding temporal understanding.	Audio-Video Fusion. Processes video frames and audio signals concurrently.	Generates temporally detailed, conversational descriptions	A powerful, general-purpose multimodal model that treats video as a sequence of frames .	Natively multimodal; built from the ground up to process multimodal inputs.
Input Modalities	Video, Image, Text	Video, Image, Text	Video, Audio , Image, Text	Video, Text	Image (Video as frame sequence), Text	Video, Audio , Image, Text
Temporal Reasoning	Good. Strong at generating coherent video captions.	Medium. Relies on the LLM to reason over frame sequences; may lack a specialized temporal module	Strong. Explicitly aligns and fuses visual and audio temporal information.	Strong. Specifically designed to generate temporally-aware descriptions.	Medium. Treats video as a collection of frames, relying on the LLM's context window to link events.	Very Strong. Natively supports long videos (e.g., 1 hour in 1.5 Pro) with high-density temporal retrieval.
Pros	Unified architecture , good support for multiple video tasks (captioning, QA).	Built on the strong LLaVA foundation; strong zero-shot video QA capabilities	Audio is a key advantage. Can understand honks, sirens, collisions , etc.	Generates highly detailed, human-like, and conversational descriptions.	Extremely strong general-world knowledge and zero-shot reasoning; "understands everything."	Long-context capability ; native audio-video fusion; powerful temporal localization.
Cons / Limitations	May be slightly weaker on "interaction" and "dialogue" compared to newer models.	Understanding of long videos and complex dynamics might be limited.	Adds complexity of audio processing; requires datasets with audio.	Its visual understanding may rely on external models; not fully end-to-end.	"Black-box" API; limited frame sampling rate for video; high cost.	"Black-box" API; researchers cannot modify its internal architecture.
Relevance to Autonomous Driving (AD)	Medium. A good starting point for generating baseline descriptions (captions).	High. Very suitable for your project (reproduction and migration) for fine-grained visual scene understanding.	Very High. In AD scenarios, sirens, horns, and tire sounds are crucial "interaction info" this model can capture.	High. Suitable for generating "human-readable" test reports and scene descriptions.	High. Can be used for complex zero-shot scene reasoning (e.g., "Judge pedestrian intent from this frame").	Very High. Can analyze long driving footage and precisely locate "interaction events" and their causes (with audio).



基于大模型的无人驾驶场景数据标注



human: <image> What objects are detected in front of the ego vehicle? (**Perception**)
gpt: There is a stopped vehicle and a pedestrian crossing ahead.

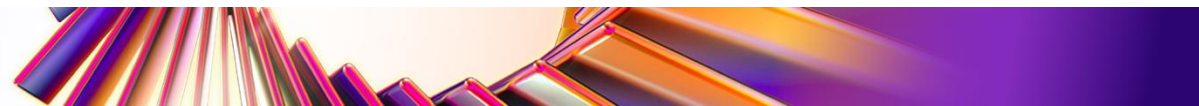
human: <image> Based on current speed and positions, will the pedestrian cross the road in front of the ego vehicle? (**Prediction**)
gpt: Yes, the pedestrian is moving towards the crosswalk and will cross soon.

human: <image> Is the ego vehicle planning to stop or slow down for the pedestrian? (**Planning**)
gpt: The ego vehicle is planning to slow down to yield to the pedestrian.

human: <image> Is the ego vehicle currently slowing down? (**Behavior**)
gpt: Yes, the ego vehicle is reducing speed gradually.

human: <image> Is the ego vehicle maintaining lane position during this maneuver? (**Motion**)
gpt: Yes, the vehicle remains centered in its lane while slowing down.

[OpenDriveLab, ECCV 2024]

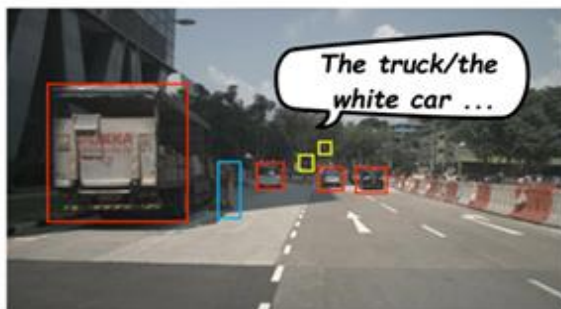


具身智能数据集标注: DriveLM

- Completeness in functionality (covering **Perception**, **Prediction**, and **Planning** QA pairs).

- 功能的完整性 (涵盖感知、预测和计划QA对)。

What objects should the ego-car notice?



Perception

What may these objects do in the future?



Prediction

What should the ego-car do in the future?



Planning

- Reasoning for **future events** that have not yet happened.

- 推理未来尚未发生的事件。

Many "**What If**"-style questions: imagine the future by language.

许多“如果”式的问题: 用语言想象未来。

If the direction of the ego-car leans to the left, it may cause collision with the pedestrian in an orange shirt.



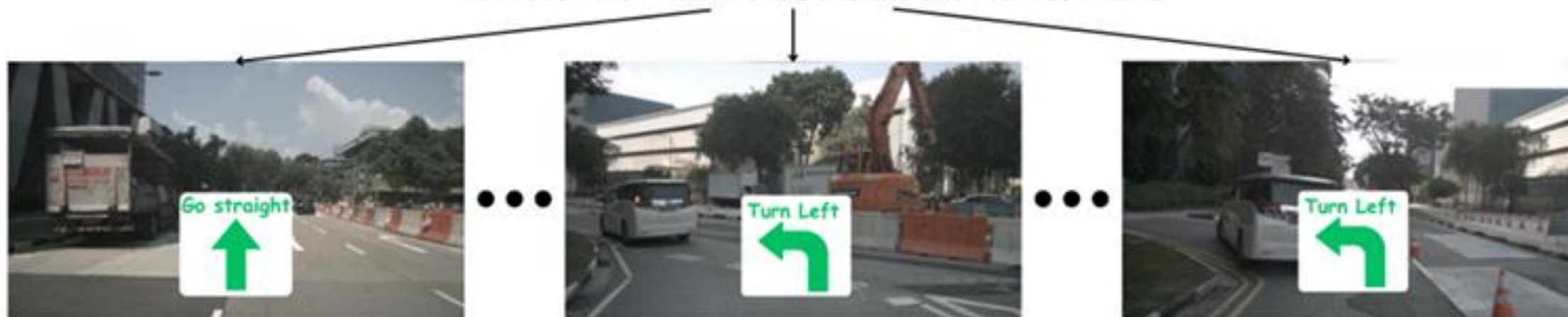
If the ego-car turns to the right lane, it will turn back to the straight lane, which is against the original goal.

[OpenDriveLab, ECCV 2024]

□ Task-driven decomposition.

One scene-level description into **many frame-level** trajectories & planning QA pairs.

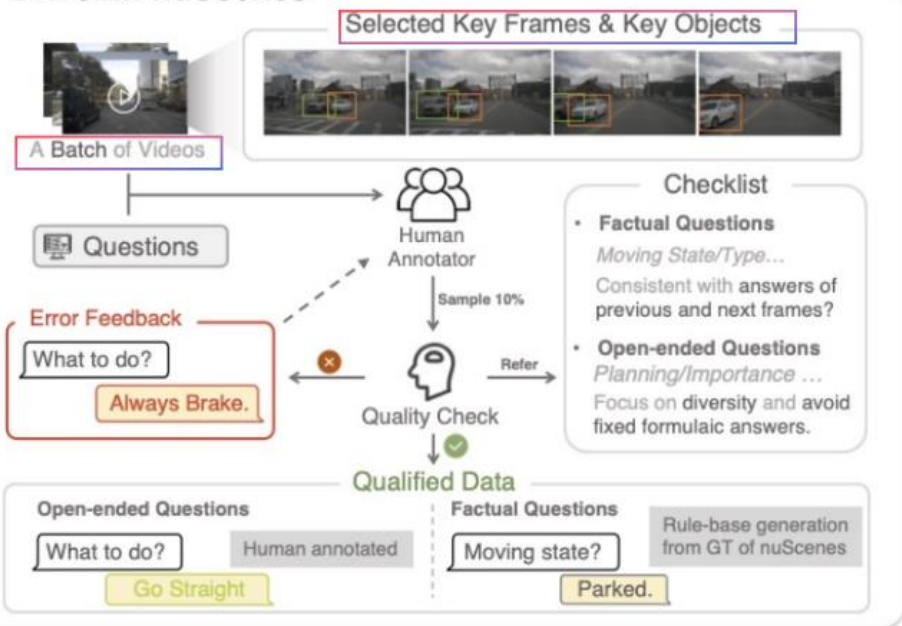
Scene-level goal: The ego-car should go ahead, turn left at the intersection. Then it should make another left turn.



[OpenDriveLab, ECCV 2024]



DriveLM-nuScenes

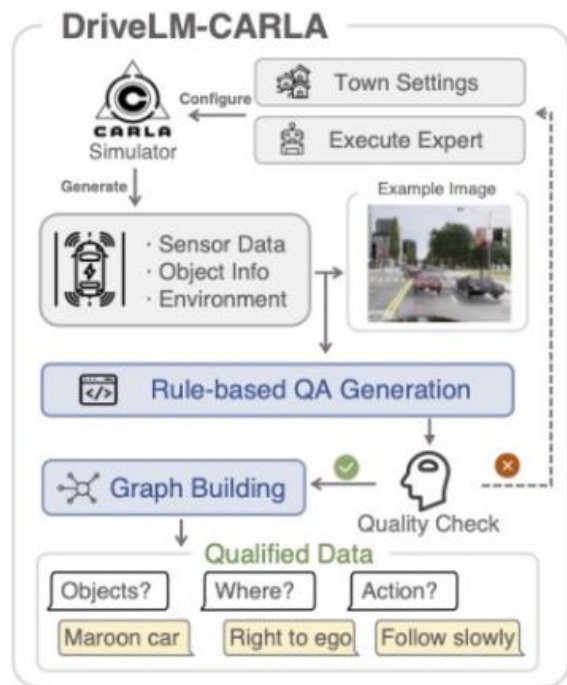


For DriveLM-nuScenes, we divide the annotation process into **three steps**:

- ❑ **Keyframe selection.** Given all frames in one clip, the annotator selects the keyframes that need annotation. **The criterion** is that those frames should involve **changes in ego-vehicle movement status** (lane changes, sudden stops, start after a stop, etc.).
 - ❑ 关键帧选择。给定一个剪辑中的所有帧，注释器选择需要注释的关键帧。标准是这些帧应该涉及自我车辆运动状态的变化（车道变化、突然停止、停止后启动等）。
- ❑ **Key objects selection.** Given keyframes, the annotator needs to pick up key objects in the **six surrounding images**. The criterion is that those objects should be able to affect the **action of the ego vehicle** (traffic signals, pedestrians crossing the road, other vehicles that move in the direction of the ego vehicle, etc.).
 - ❑ 关键点对象选择。给定关键帧，注释器需要在周围的六个图像中拾取关键对象。标准是这些物体应该能够影响本车的动作（交通信号、过马路的行人、向本车方向移动的其他车辆等）。
- ❑ **Question and answer annotation.** Given those key objects, we automatically generate questions regarding single or multiple objects about perception, prediction, and planning. More details can be found in our data.
 - ❑ 问答注释。给定这些关键对象，我们会自动生成关于感知、预测和规划的单个或多个对象的问题。更多细节可以在我们的数据中找到。

[OpenDriveLab, ECCV 2024]



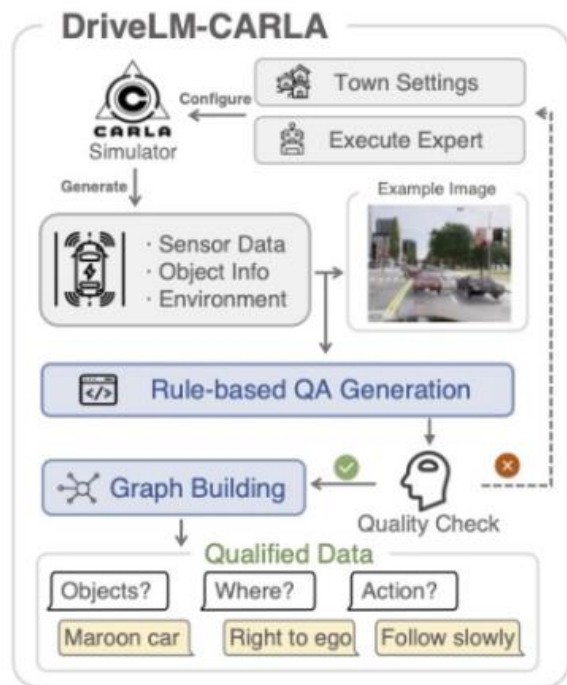


- We collect data using **CARLA 0.9.14** in the **Leaderboard 2.0** framework with a privileged rule-based **expert**. We **set up a series of routes** in urban, residential, and rural areas and execute the expert on these routes. During this process, we collect the **necessary sensor data**, **generate relevant QAs** based on privileged information about objects and the scene, and organize the logical relationships to connect this series of QAs into a graph.

- 对于DriveLM-CARLA，我们采用自动化注释方法：我们在Leaderboard 2.0框架中使用CARLA 0.9.14收集数据，并配备了一名特权规则专家。我们在城市，住宅和农村地区设置了一系列路线，并在这些路线上执行专家。在这个过程中，我们收集必要的传感器数据，根据有关对象和场景的特权信息生成相关的QA，并组织逻辑关系将这一系列QA连接到一个图中。

[OpenDriveLab, ECCV 2024]





- We collect data using **CARLA 0.9.14** in the **Leaderboard 2.0** framework with a privileged rule-based **expert**. We **set up a series of routes** in urban, residential, and rural areas and execute the expert on these routes. During this process, we collect the **necessary sensor data**, **generate relevant QAs** based on privileged information about objects and the scene, and organize the logical relationships to connect this series of QAs into a graph.

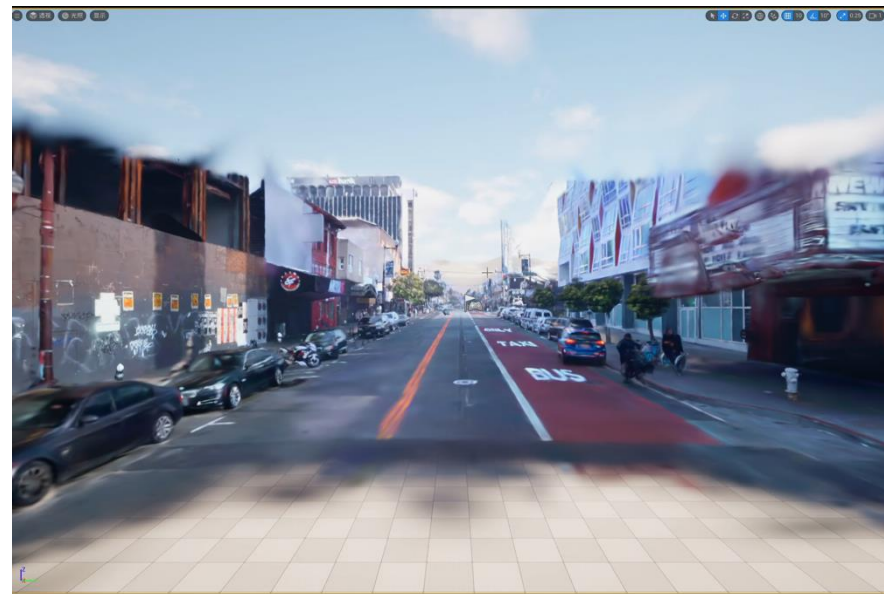
- 对于DriveLM-CARLA，我们采用自动化注释方法：我们在Leaderboard 2.0框架中使用CARLA 0.9.14收集数据，并配备了一名特权规则专家。我们在城市，住宅和农村地区设置了一系列路线，并在这些路线上执行专家。在这个过程中，我们收集必要的传感器数据，根据有关对象和场景的特权信息生成相关的QA，并组织逻辑关系将这一系列QA连接到一个图中。

[OpenDriveLab, ECCV 2024]

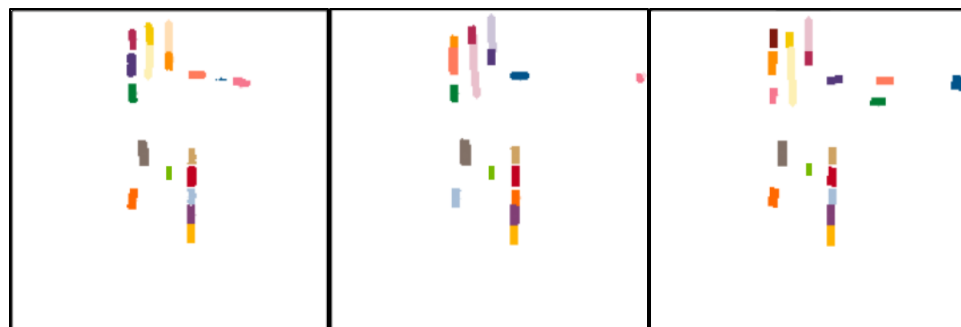




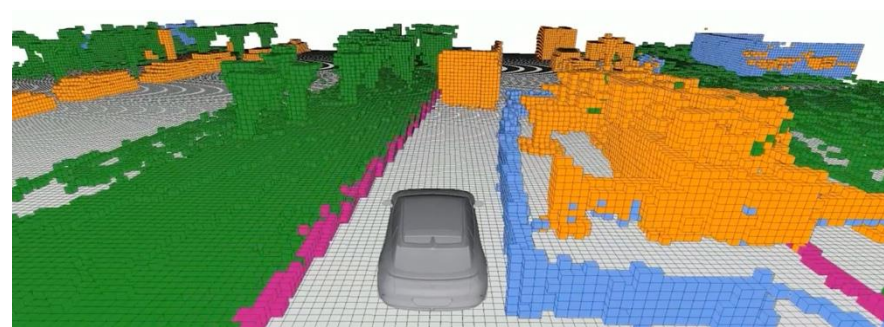
三维Mesh模型



三维高斯模型



二维占用网格地图



三维占用网格地图





Prompt: Current speed: -0.0 m/s. Target waypoint: <TARGET_POINT><TARGET_POINT>. What should the ego do next?
Answer: Follow the route. Remain stopped to stay behind the orange car that is to the front in 15.0 meters. Waypoints:



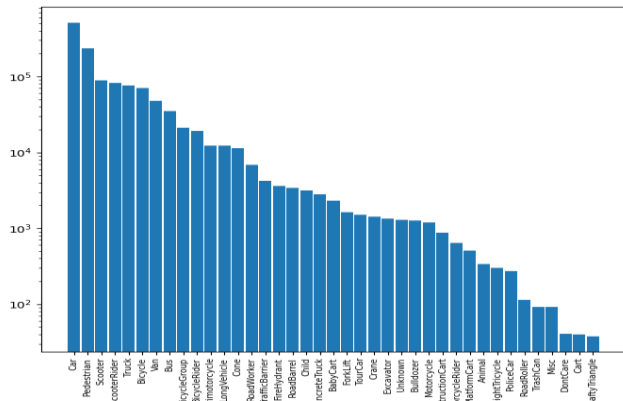
Prompt: Current speed: 0.0 m/s. Target waypoint: <TARGET_POINT><TARGET_POINT>. What should the ego do next?
Answer: Follow the route. Accelerate to follow the black car that is to the front. Waypoints:

CARLA中QA可视化

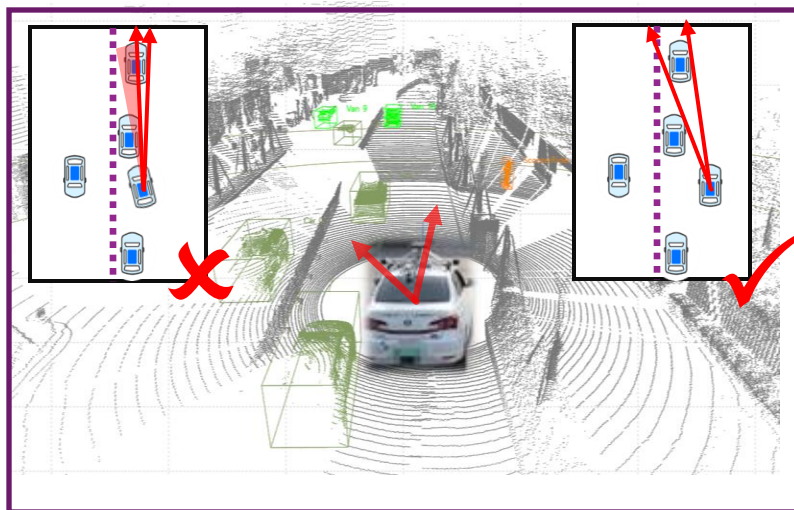


PART 04

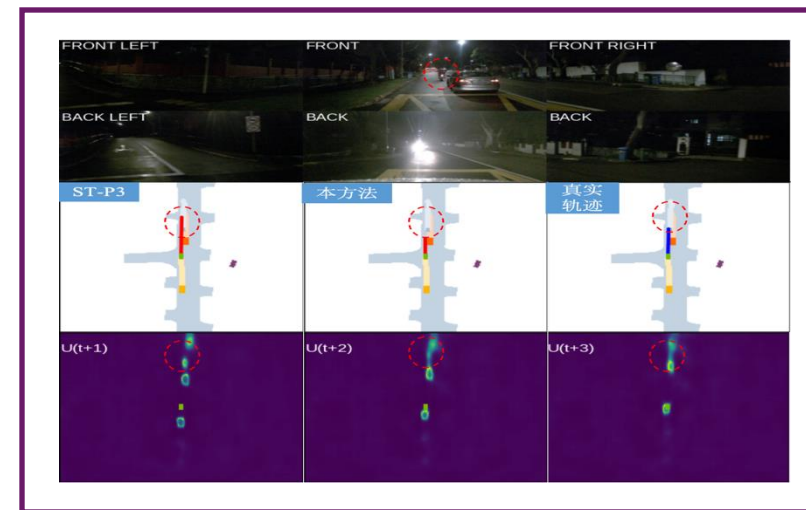
具身智能数据评估与优化



场景统计复杂性与多样性

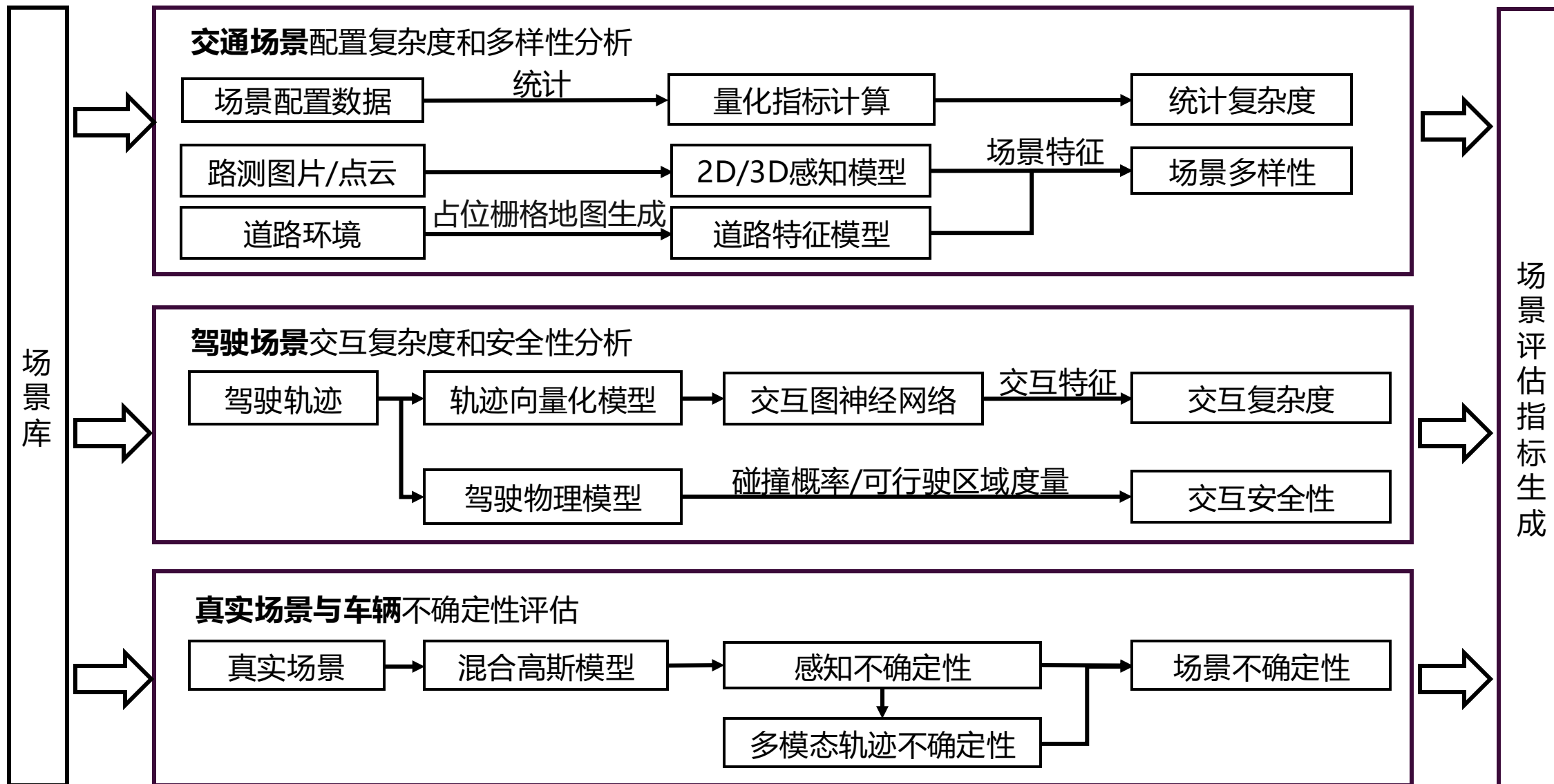


交互场景复杂性与安全性



真实场景不确定性



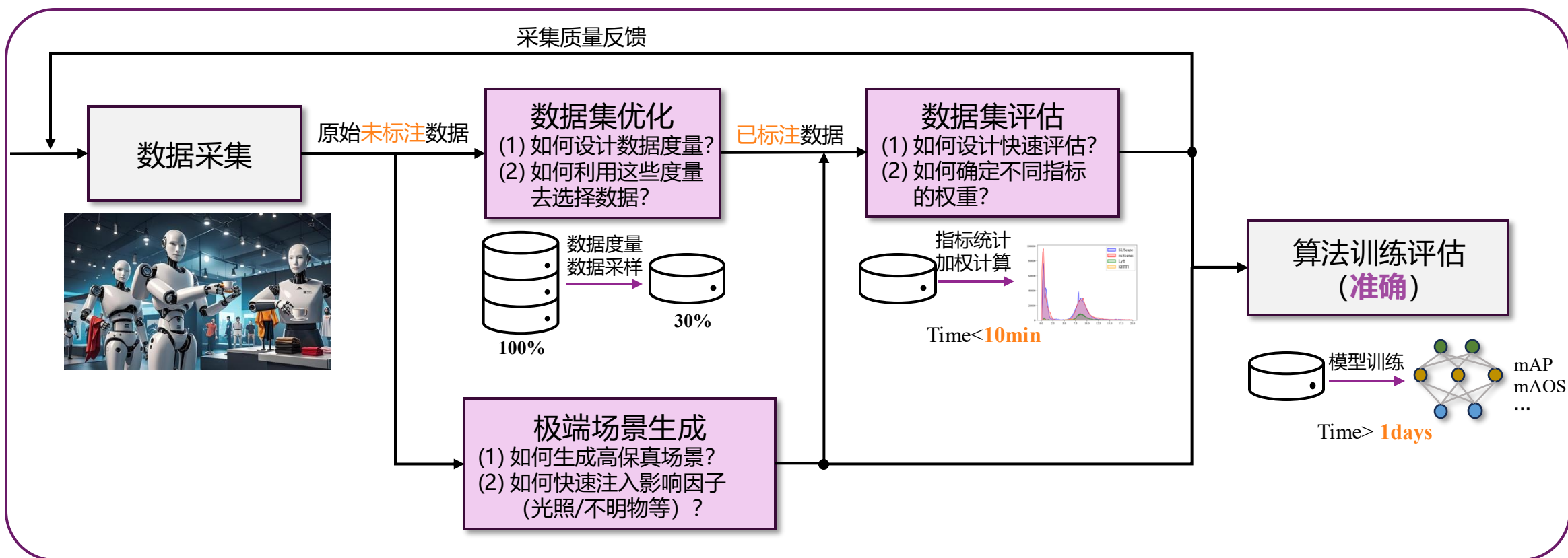


具身智能数据评估与优化

优化具身智能场景数据，针对具身智能场景特点设计数据度量，从**可信性**角度设计采样策略；

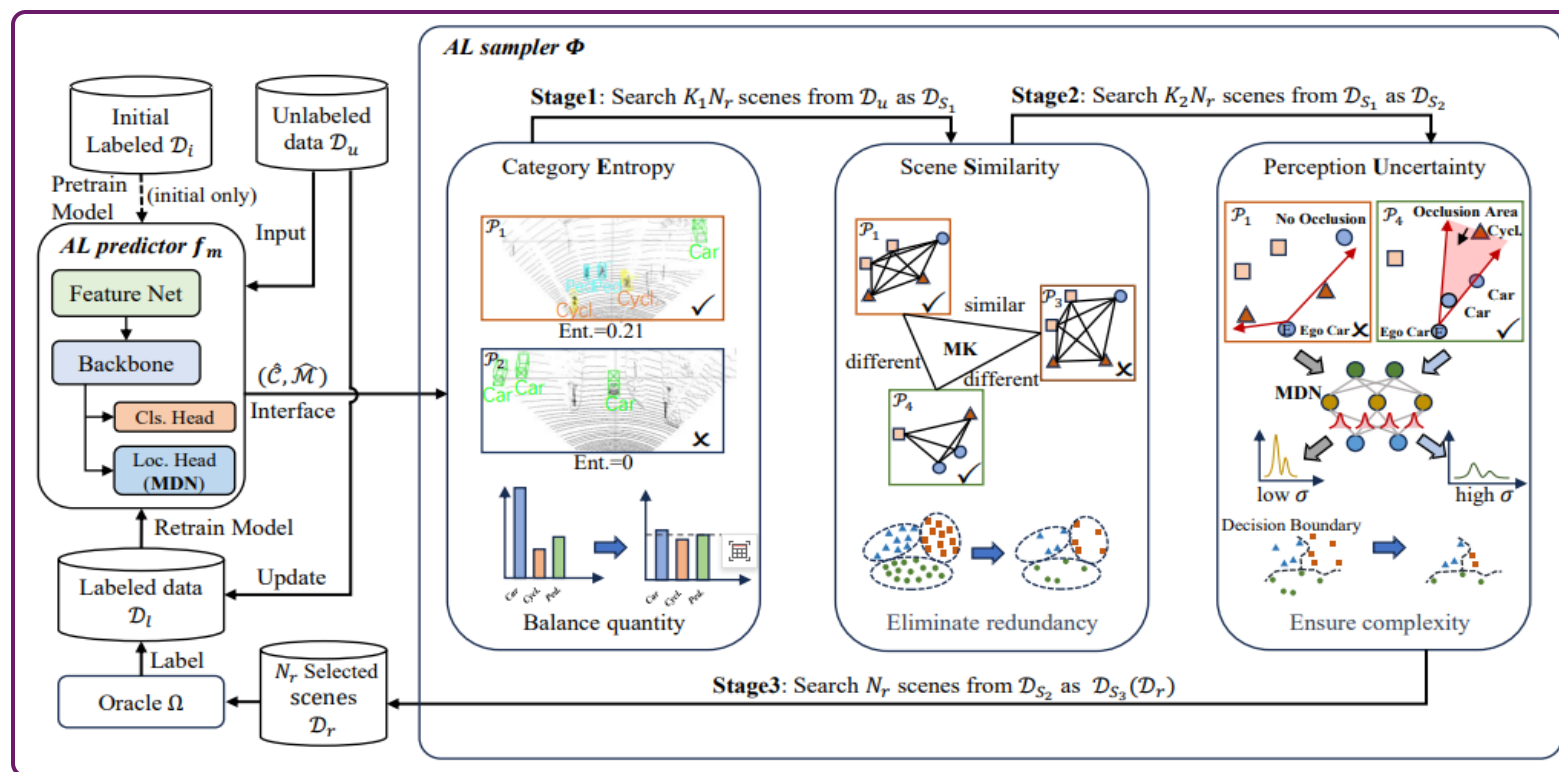
评估优化数据集，即从**平衡性**，**多样性**，**复杂性**角度设计统计指标，综合地对数据集进行评价；

极端场景生成，从数据**多样性**和**完备性**出发，可控地注入影响因子，快速生成多样的极端的测试场景。

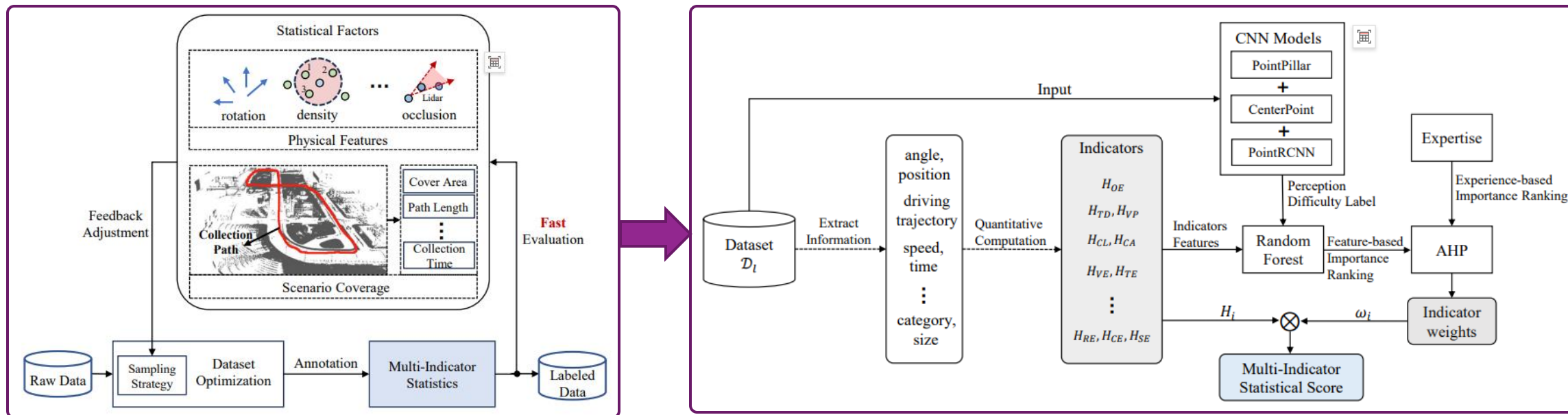


关键技术1：基于主动学习的多阶段场景智能采样

- 提出基于熵计算的类别熵度量 (增加平衡性)
- 提出基于图相似性的场景相似性度量 (减少冗余性)
- 提出基于混合密度网络的感知不确定性度量 (增加复杂性)
- 提出一种多阶段采样策略



关键技术2：综合性多指标智能数据评估指标



优化数据构建流程的快速数据集评估方法：

- 针对批量的未标注数据，采用数据集优化方法筛选出价值最高的数据样本。
- 运用多指标统计评估方法进行快速量化评价，包括对复杂程度、分布密度、遮挡程度等维度的统计分析。
- 根据评估得分反馈调整采样策略，从而实现数据集构建过程的动态优化。

具身智能数据极端场景

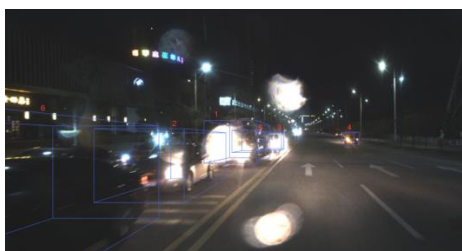
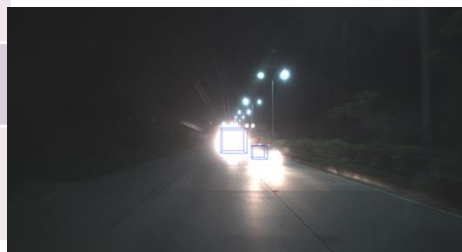
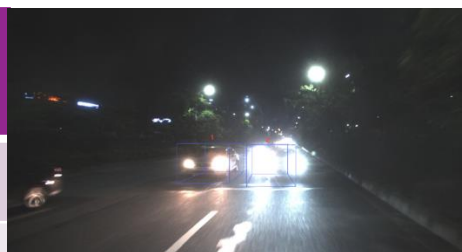
Night

Raining

Unstructured

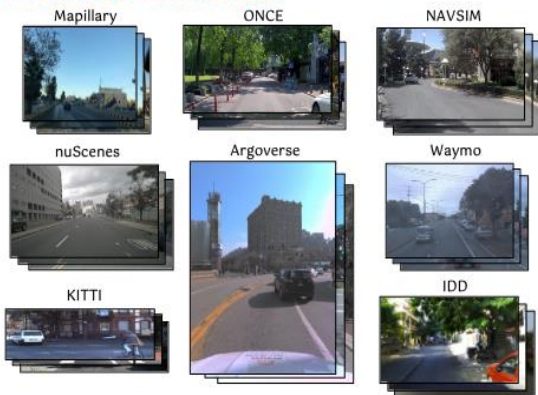
Complex

	场景类别 (class)	帧数 (frame)	时长 (hr)
1	Raining	~2200	6
2	Evening	~1100	3.5
3	Night	~3300	5.4
4	Unstructured Environment	~1800	10



具身智能数据评估与优化

1. Dataset Collection



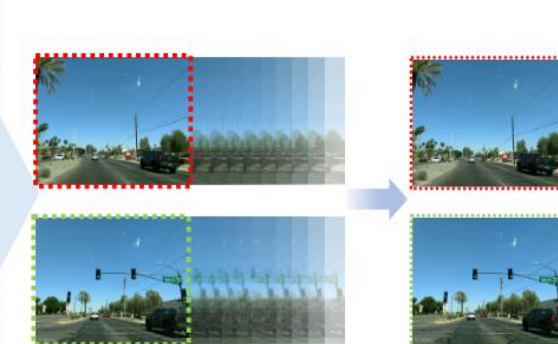
2. Taxonomy of Unstructured Scenarios



3. Frequency Alignment



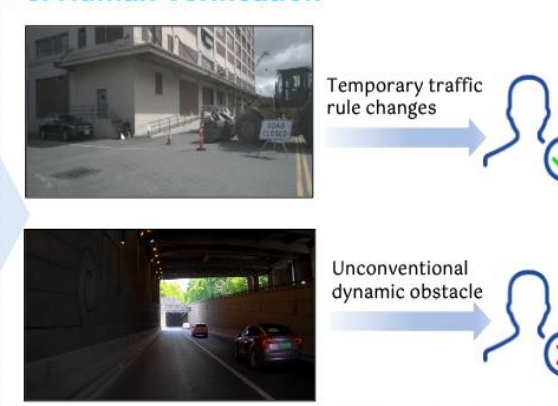
4. Clip Selection



5. Multi-Task Annotation Generation



6. Human Verification



Impromptu VLA 2025, 清华大学

<https://impromptu-vla.c7w.tech/>

原始数据集

- 8个开源大型自动驾驶数据集
- 200万个原始片段
- 2Hz采样频率

优化后数据集

- 1个挑战场景数据集
- 8万个视频片段
- 1.5秒过去数据、5秒预测轨迹，稳定性过滤
- 道路边界不清、临时交通规则变化
- 非常规动态障碍物、挑战性路况

标注类型	说明与描述	标注方法
场景描述	对整体环境（包括环境背景、时间、天气和交通状况）进行全面的自然语言描述	VLM (通过提示查询)
交通信号检测	识别和标注场景中活跃交通信号灯的存在、状态和类型	VLM (通过进一步查询)
脆弱道路使用者识别	识别和标注脆弱道路使用者，包括他们的存在、类型（例如，行人、骑车人）以及与自车的距离	依赖检测框真值数据
运动意图预测	预测场景中关键动态参与者未来的运动意图，包括速度决策((保持停加减速))和路径决策 (直行右/左变道/转向)	VLM (通过提示查询)
元动作规划	为自车制定高级别的计划 (例如，加速-左转、保持直行)，速度与路径决策	VLM (基于场景上下文提示)
规划解释	生成文本解释，合理化自车对场景做出的潜在或实际操作（为什么做出上述规划）	VLM (通过提示查询)
端到端轨迹预测	结构化地提供自车过去的状态（1.5 秒的状态，包括位移、速度和加速度）和对应的未来目标轨迹（5 秒）	依赖原始真值数据



PART 05

具身智能数据检索与生成

Query: *The ego car is about to go straight through the intersection, and there is a white car ahead that is turning right.*



ID: 001



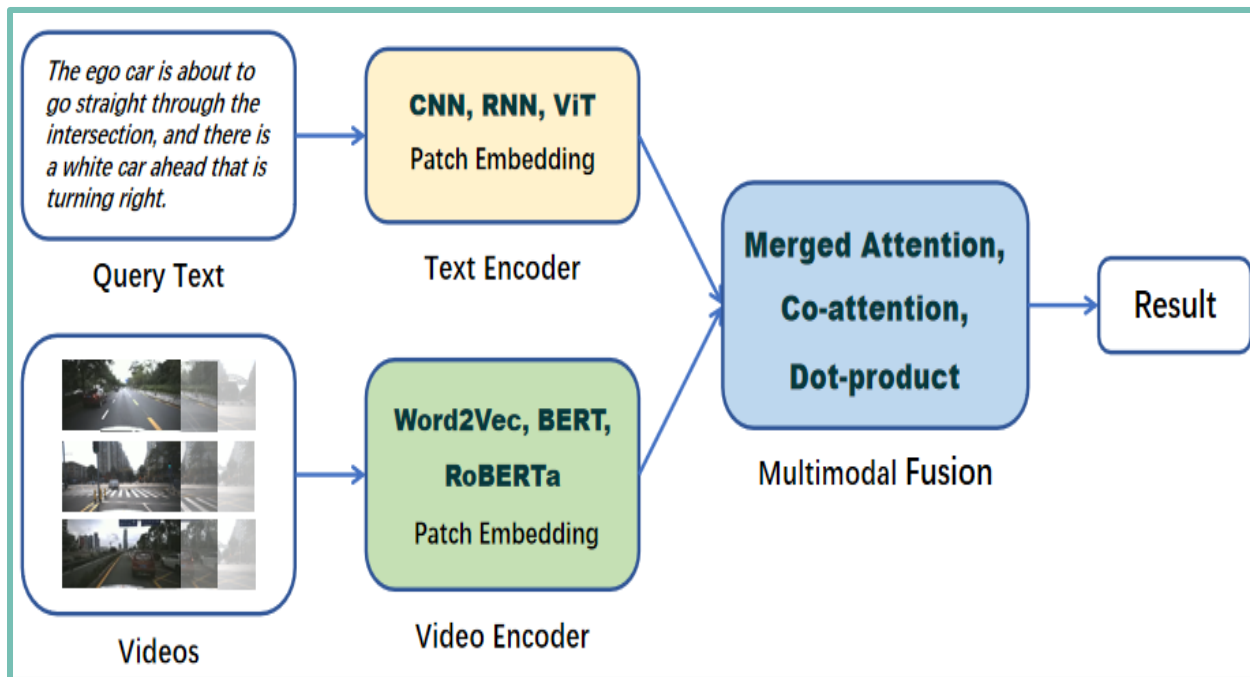
ID: 002 Ground truth



ID: 003

...

以文搜视频



以文/视频搜视频





Our labels

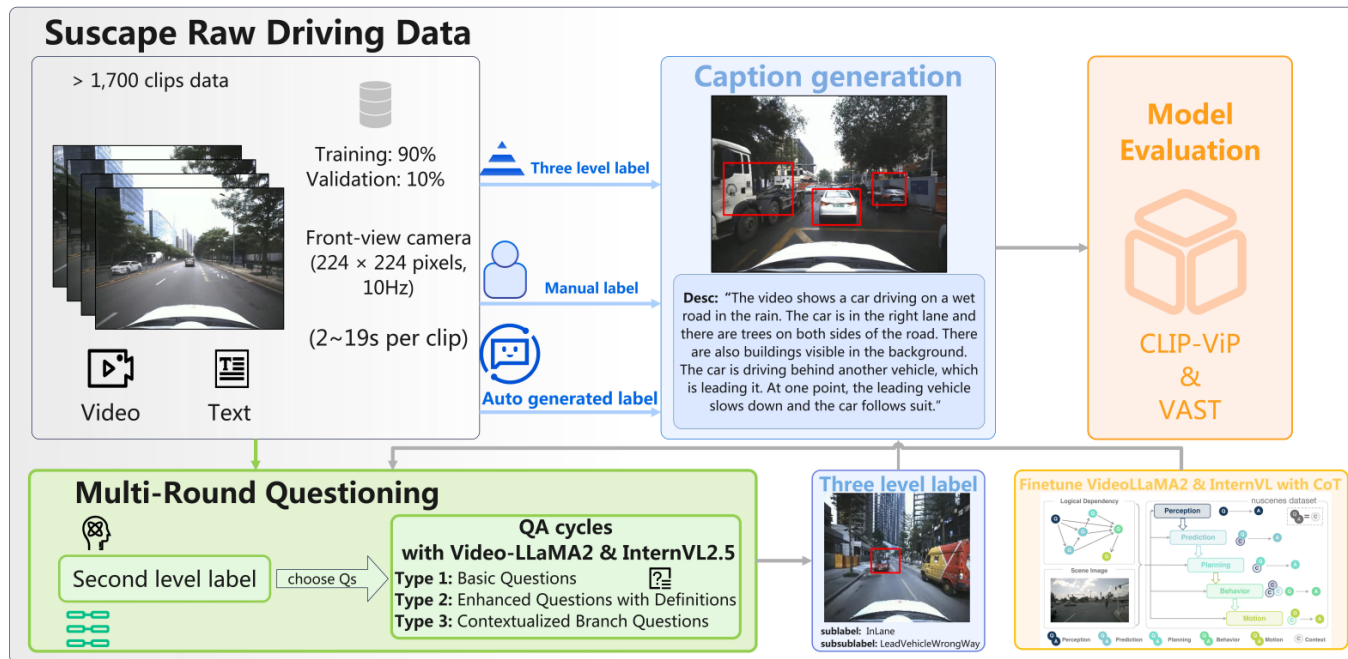
```
{
  "1.GoStraight": {
    "1.1 InLane": [
      "1.1.1 LeadVehicleConstant",
      "1.1.2 LeadVehicleDecelerating",
      "1.1.3 PedestrianCrossing"
    ],
    "1.2 ChangingLaneLeft": [],
    "1.3 ChangingLaneRight": []
  }
}
```

ChatGPT

You are approaching the scene of a car accident where a vehicle has overturned. Emergency personnel, including police and firefighters, are already present, along with an emergency vehicle further down the road. You should slow down significantly, be prepared to stop, and pay close attention to instructions from the emergency responders directing traffic. Give the emergency services ample space to work and proceed slowly and carefully when it is safe to do so.

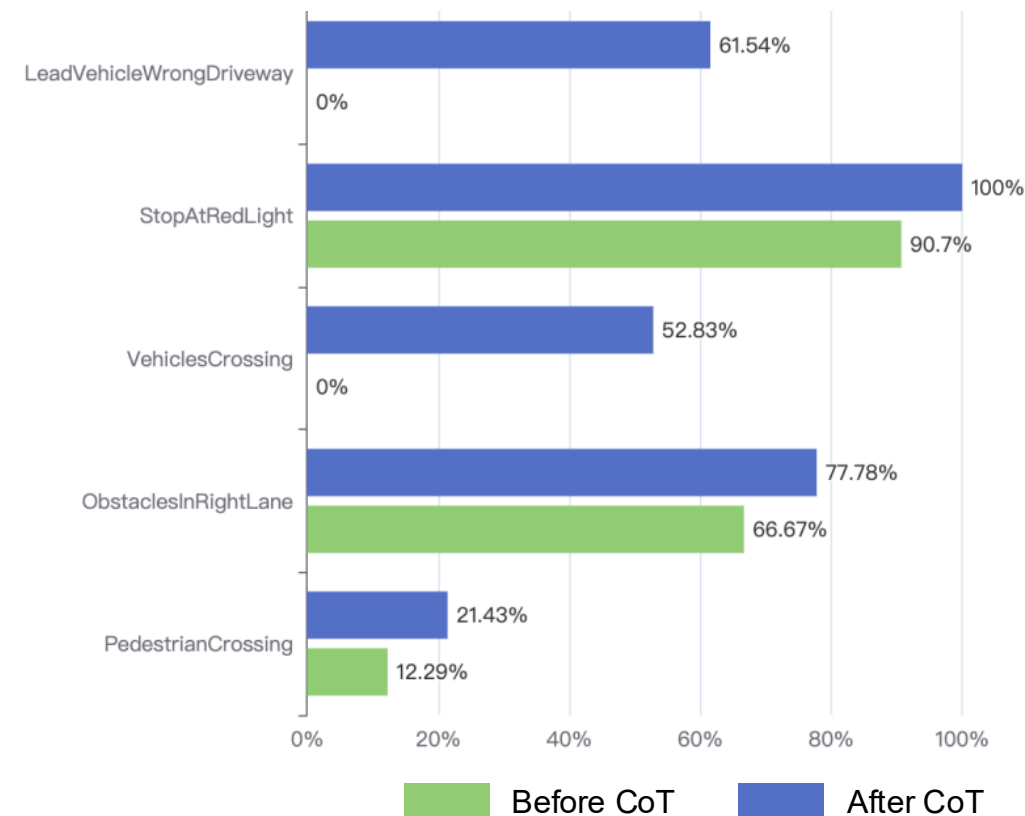


Generation of the scene description and evaluation of scene retrieval



Results of scene retrieval VAST for driving behavior

Model	Direction	Recall@1	Recall@5	Recall@10	Avg
itc_tv	text-to-video	23.7%	56.5%	73.4%	51.2
itm_tv	text-to-video	31.1%	65.0%	80.2%	58.8



Retrieval performance in driving behavior (After vs. Before CoT)

► 具身智能数据检索

ID: scene-000242-tag-5

Description:

"The video shows a car driving on the road and stopping at a red traffic light. There are other vehicles on the road, including a bus and a motorcycle. The car stops behind a bus and waits for the light to turn green before proceeding. There are also people walking on the sidewalk, and a man riding a bicycle. The weather appears to be sunny with clear skies, and there are trees visible in the background. Overall, the video depicts a typical day of driving in a city with various road users and traffic conditions."



scene-000242-tag-5

similarity: 0.5761



scene-000258-tag-5

similarity: 0.5447



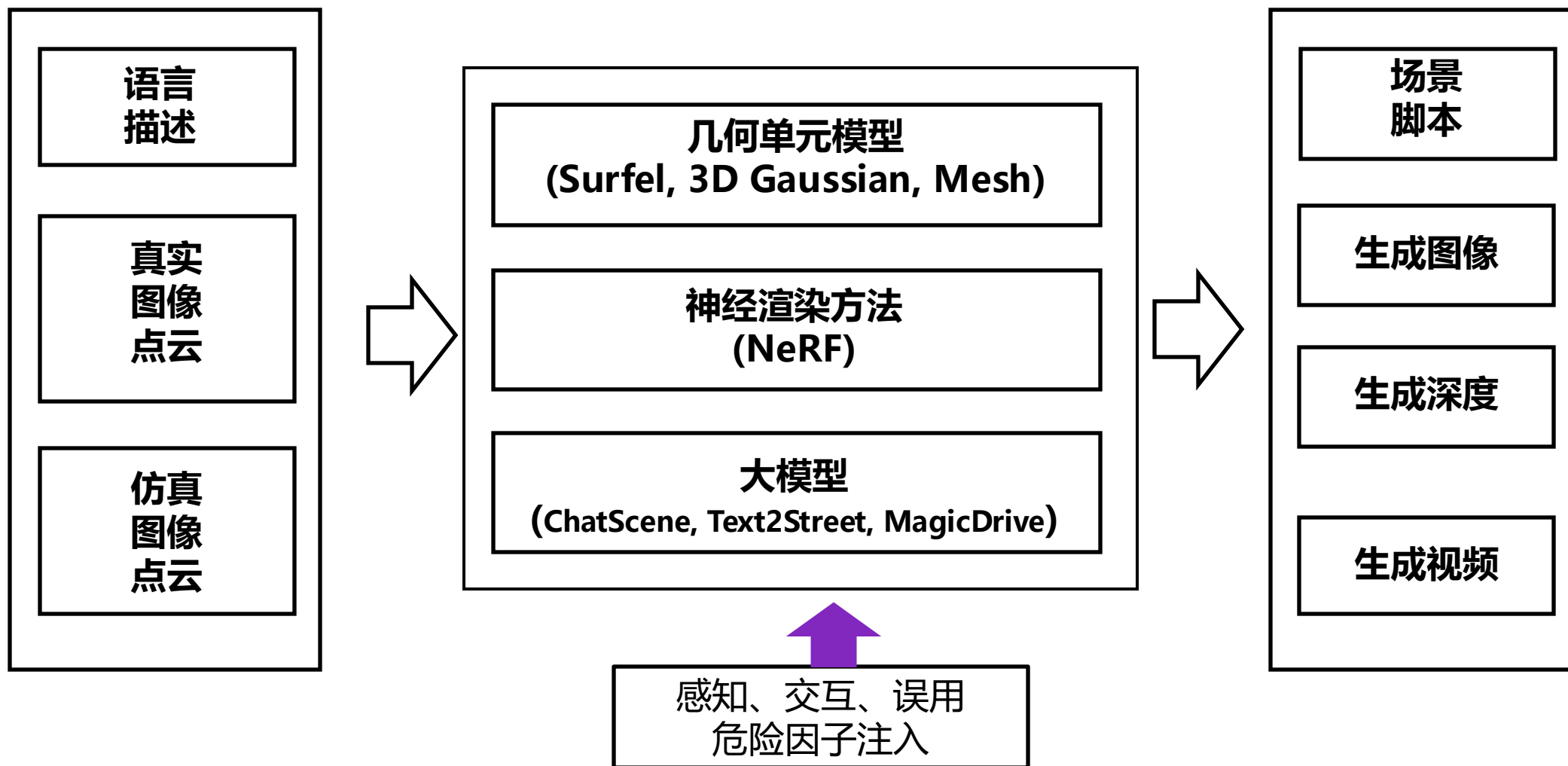
scene-000420-tag-8

similarity: 0.4504



scene-000779-tag-4

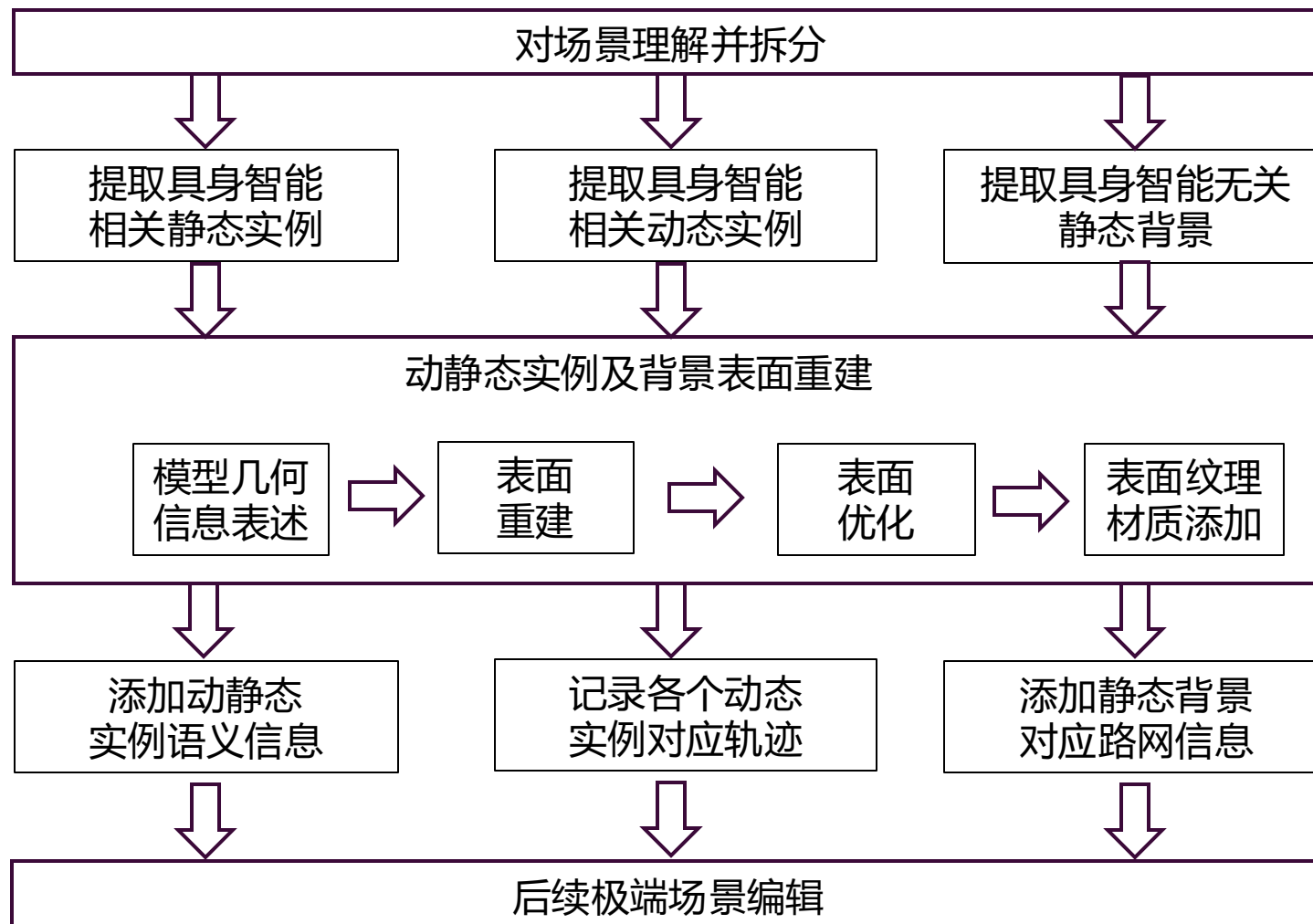
similarity: 0.0715



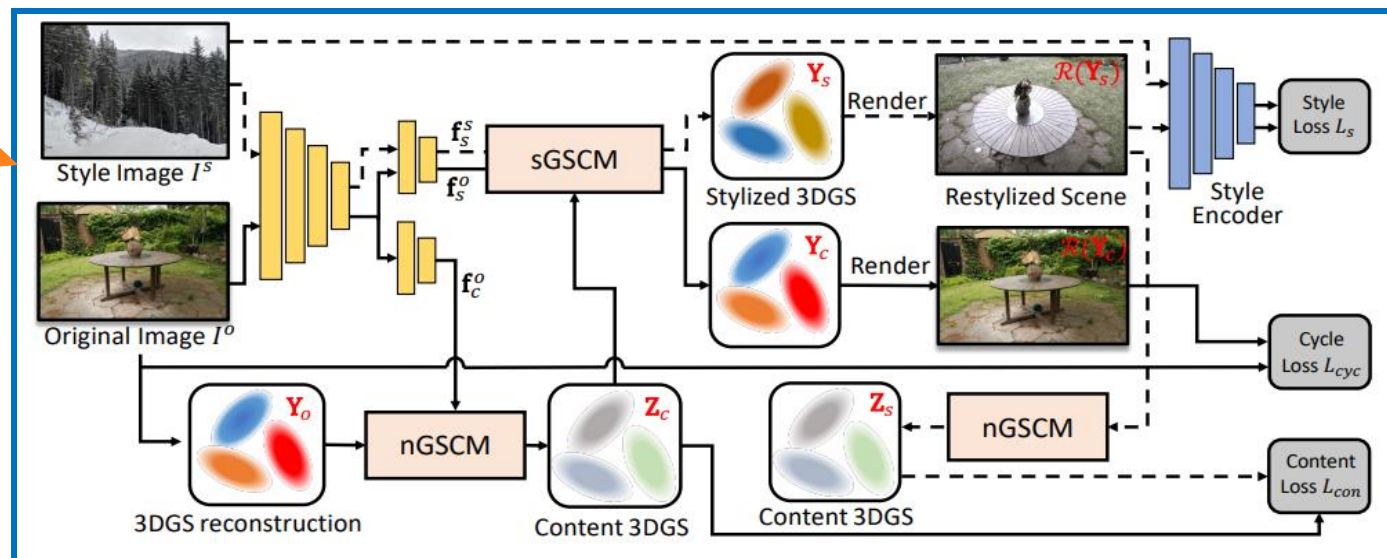
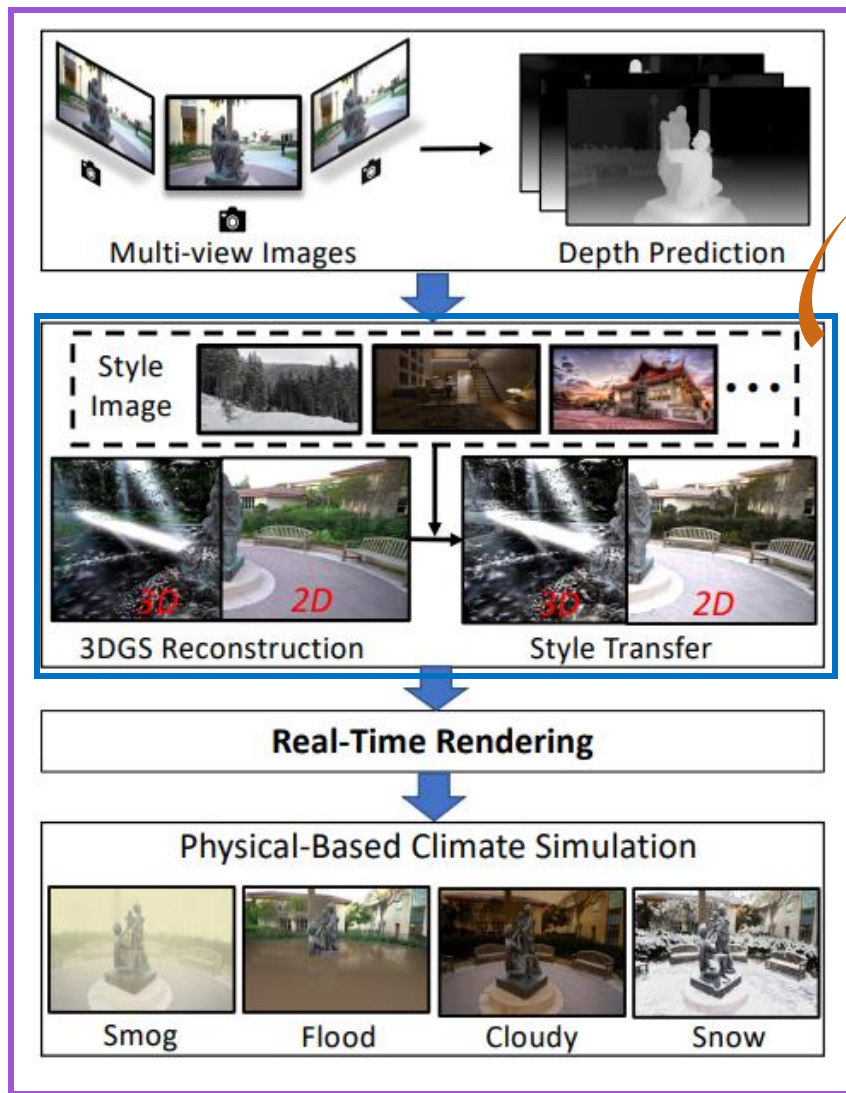
► 具身智能数据生成: 大模型生成



具身智能数据生成：可编辑场景生成



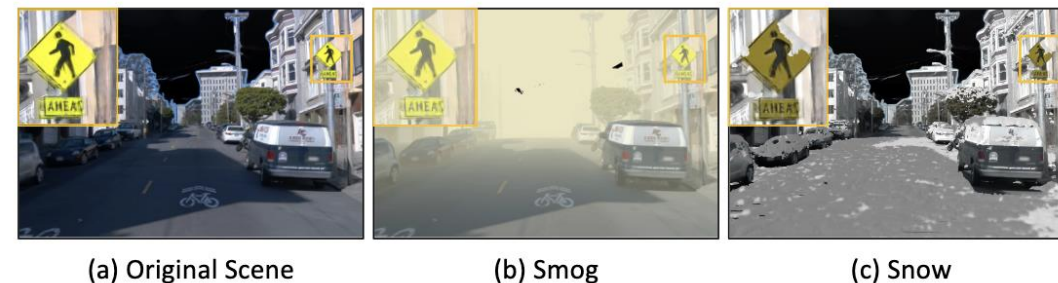
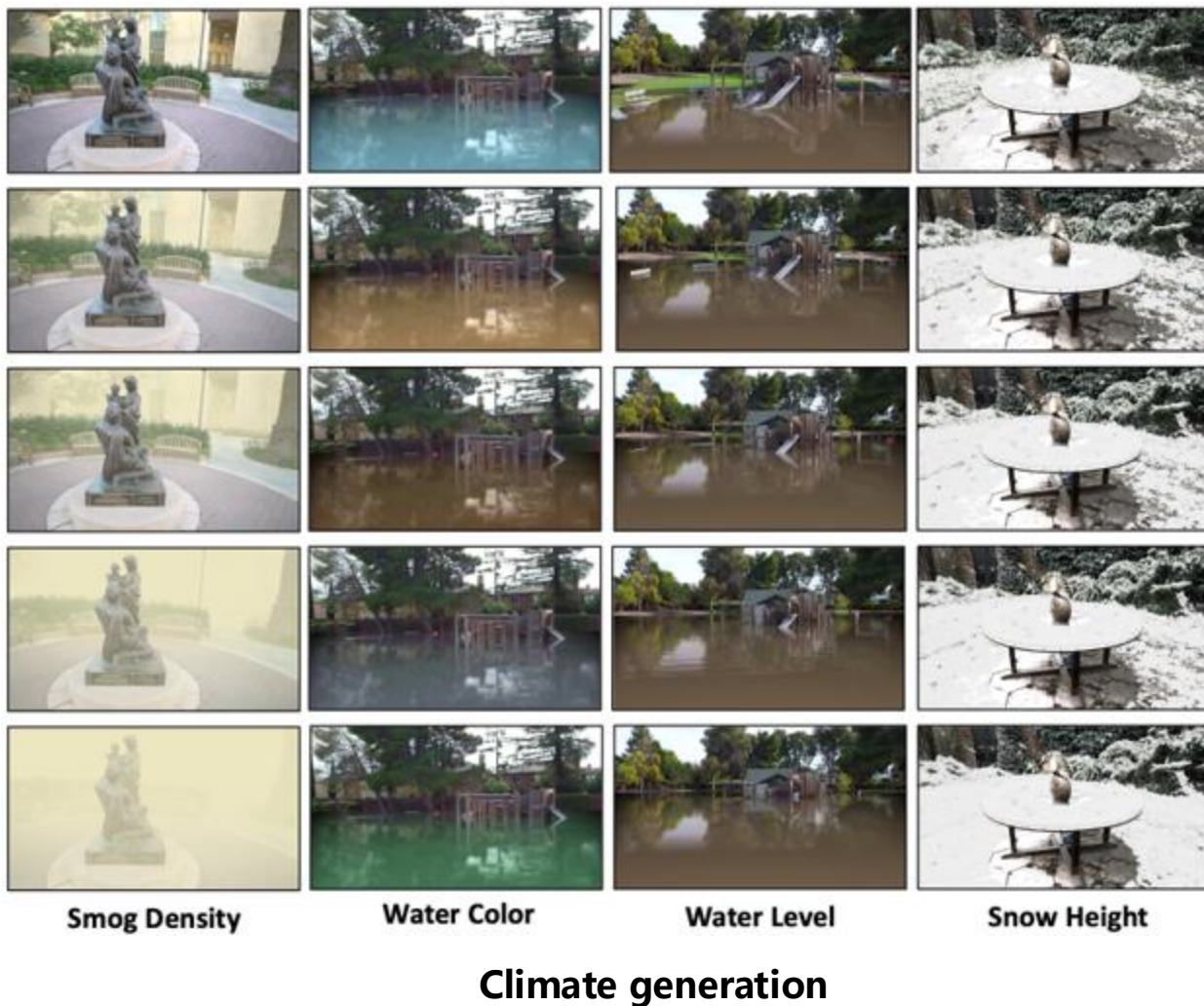
具身智能数据生成: 3D 高斯溅射方法



- 开发适用于3D高斯逼真风格迁移的线性变换方法，通过直接调整球谐函数各频带系数，实现高效且风格一致的场景适配；
- 提出联合训练策略，将监督学习与自监督学习相结合，在加速3D风格迁移收敛的同时保持原始场景细节；
- 实现气候模拟的实时渲染技术，将基于物理的模拟效果与3D高斯模型融合，达成高效逼真的可视化呈现。



具身智能数据生成 real-to-sim-to-real

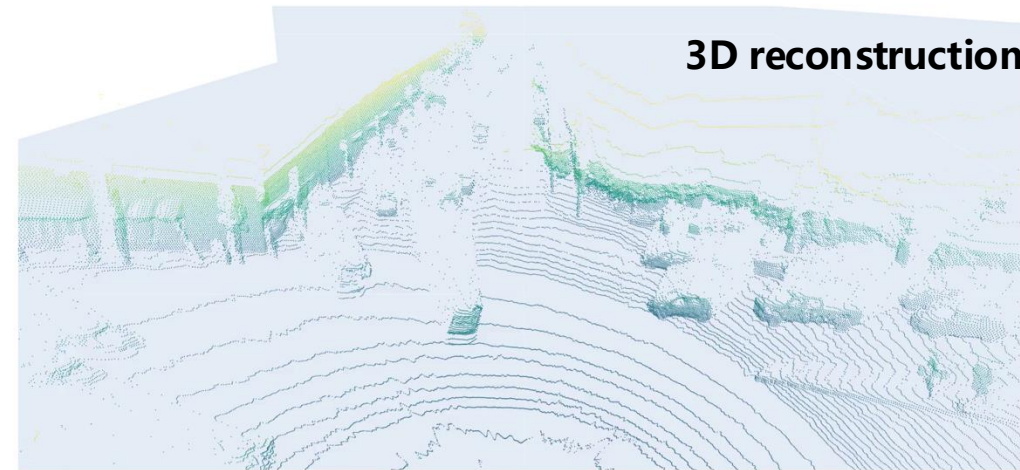
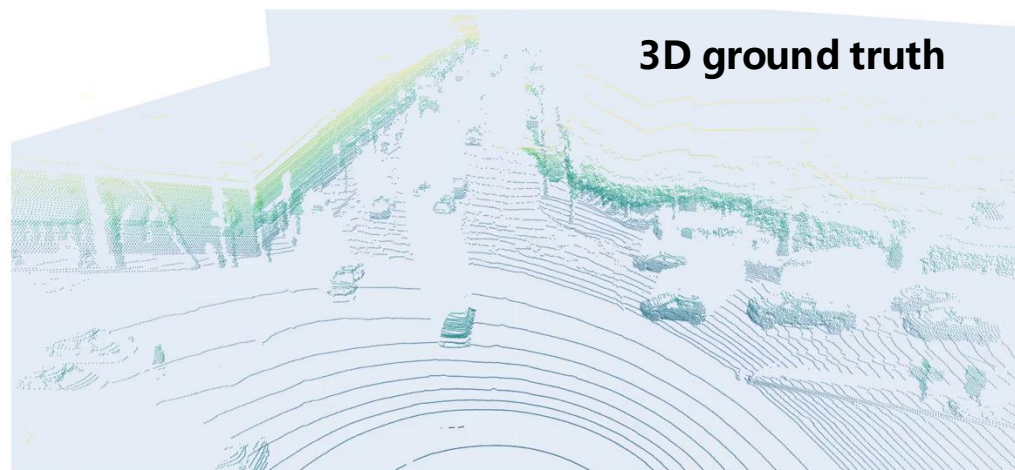


Method	Smog	Flood	Snow
ClimateNeRF [26]	1.687s	12.062s	1.5h / 4.374s
ours	0.020s	0.020s	1m / 0.025s

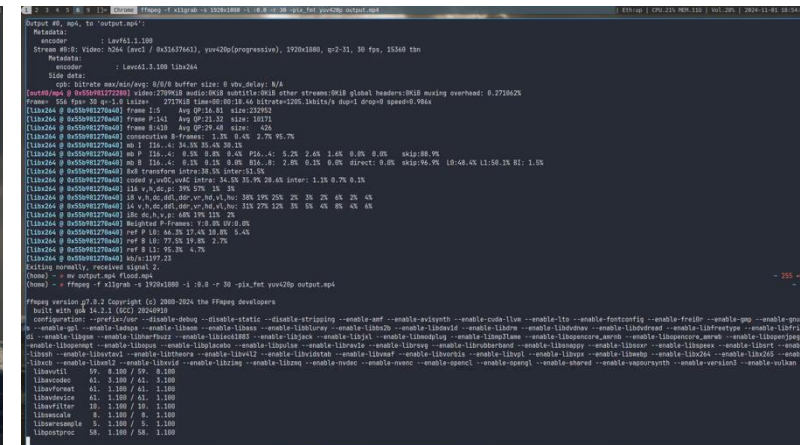
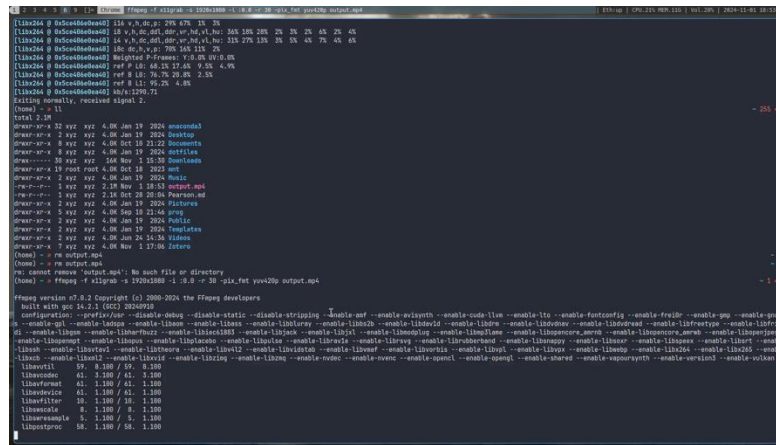
- 1.生成的场景可控制，能呈现不同严重程度灾害性场景。
- 2.可适用于大规模场景，不受限于小范围场景。
- 3.预处理时间更短，渲染过程更为迅速。



具身智能数据生成: real-to-sim-to-real



具身智能数据生成：real-to-sim-to-real

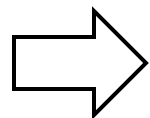


- 在场景中以粒子形式添加3D高斯球体，如局部雾气、雨滴、雪花等；
- 每个粒子具有速度、运动方向、尺寸与发光属性；
- 该技术将影响原始场景的深度信息与能见度，并确保不同模态数据的一致性。

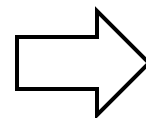


具身智能数据生成: sim-to-real 静态环境建模

Encoded Map
Encoding



2D Layout
Reinforcement
Learning



3D Model
Rendering

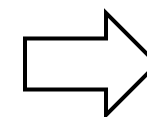
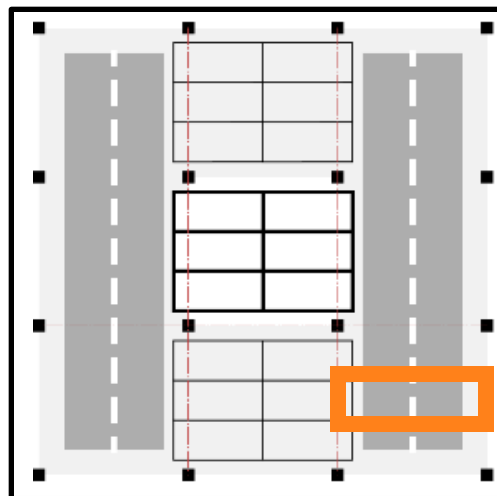
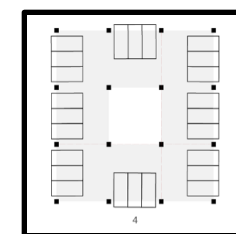
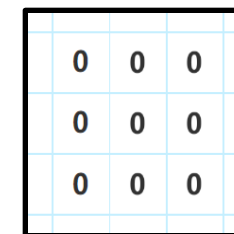
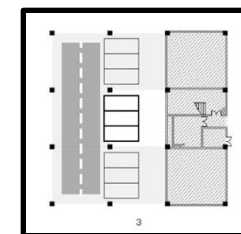
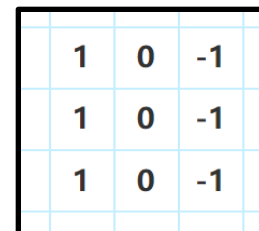
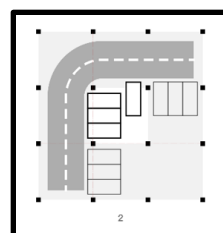
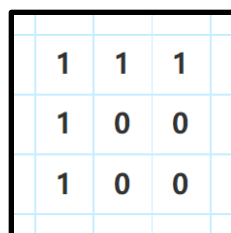
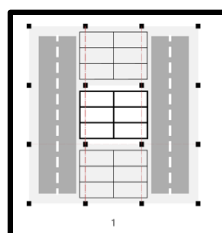
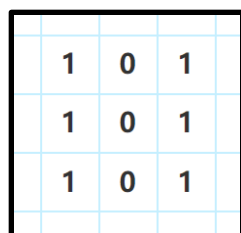
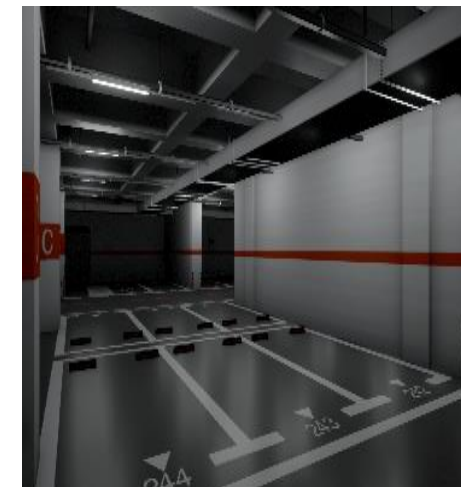
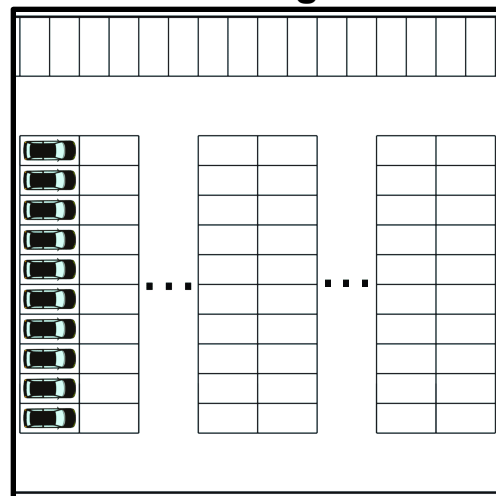


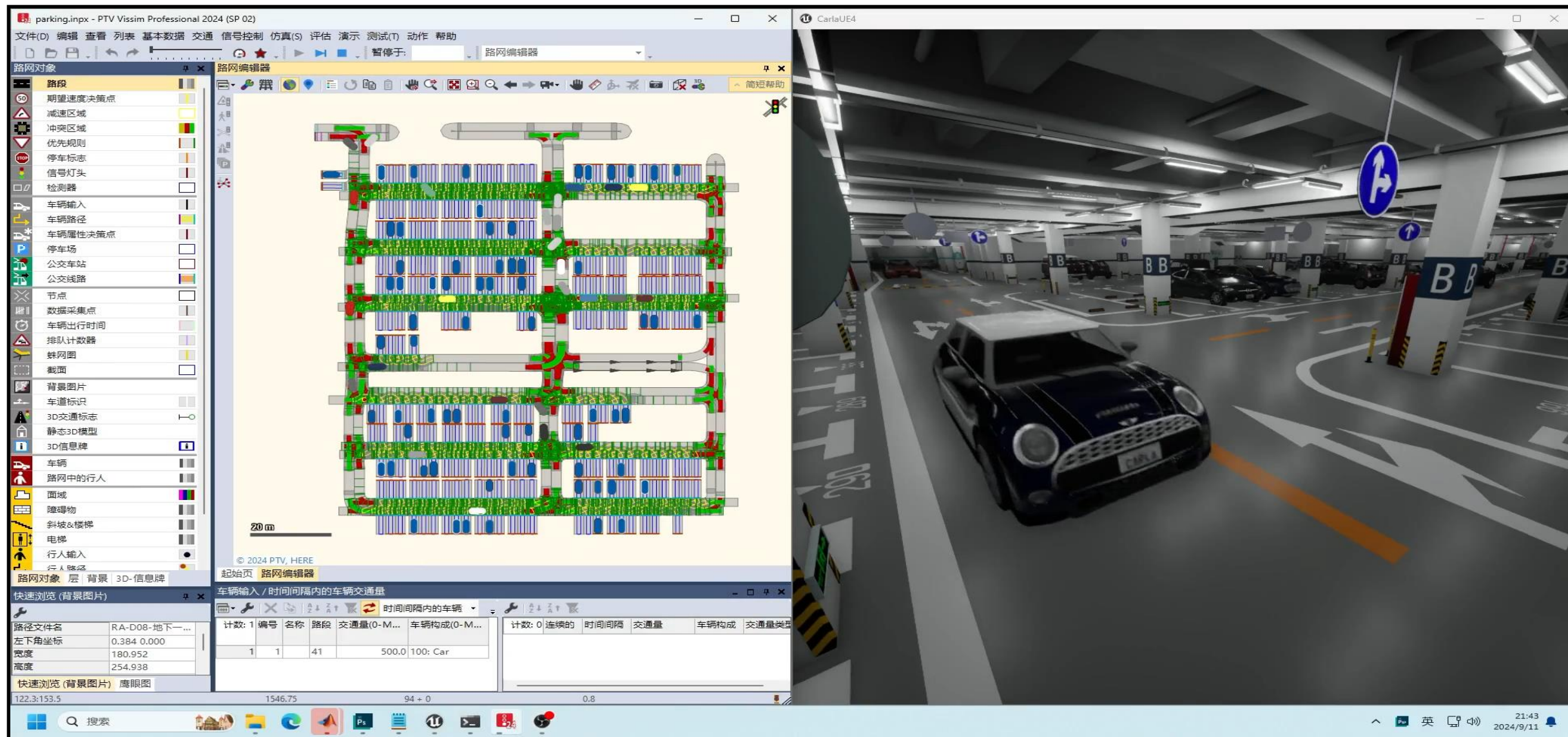
Photo-Realistic
Rendering



单个地块

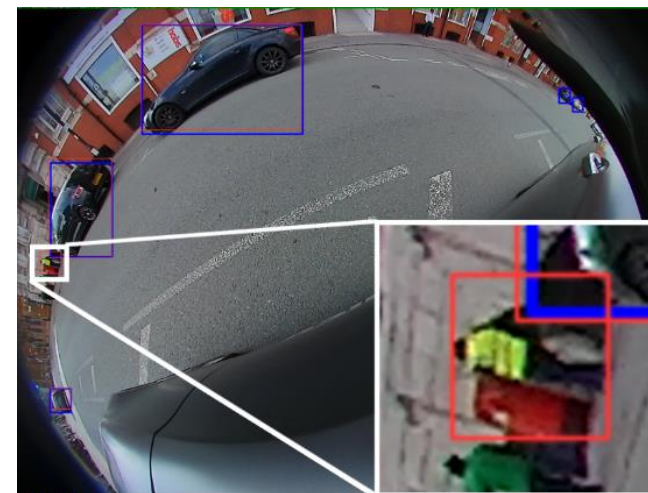
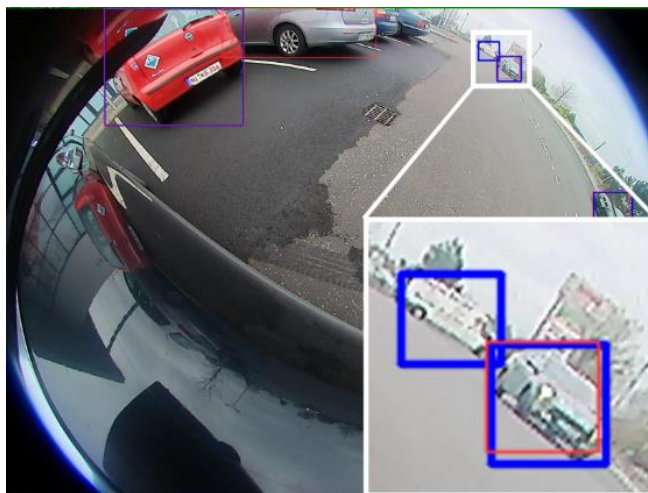


具身智能数据生成: 数字孪生仿真

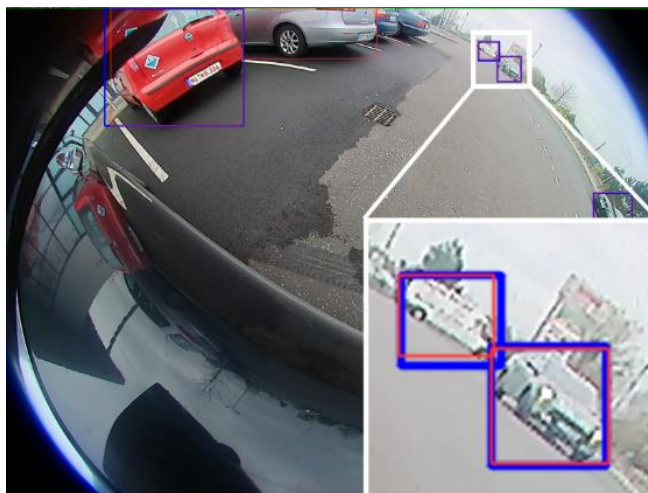


具身智能数据生成: sim-to-real相机数据迁移

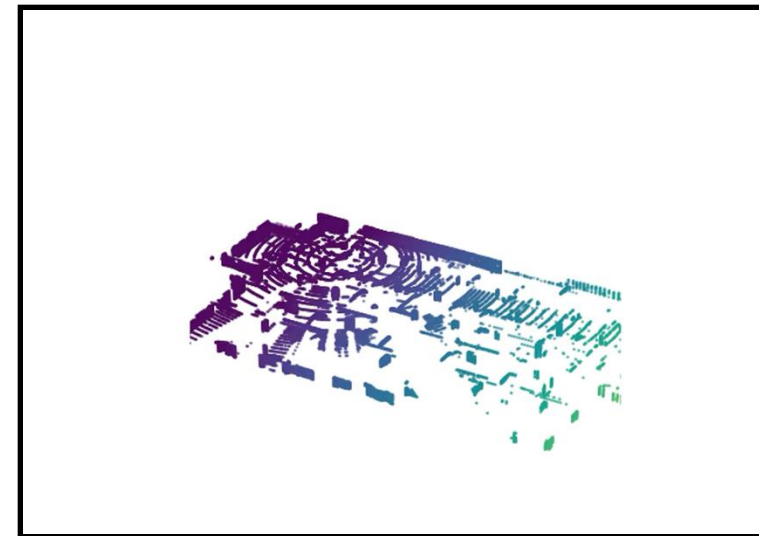
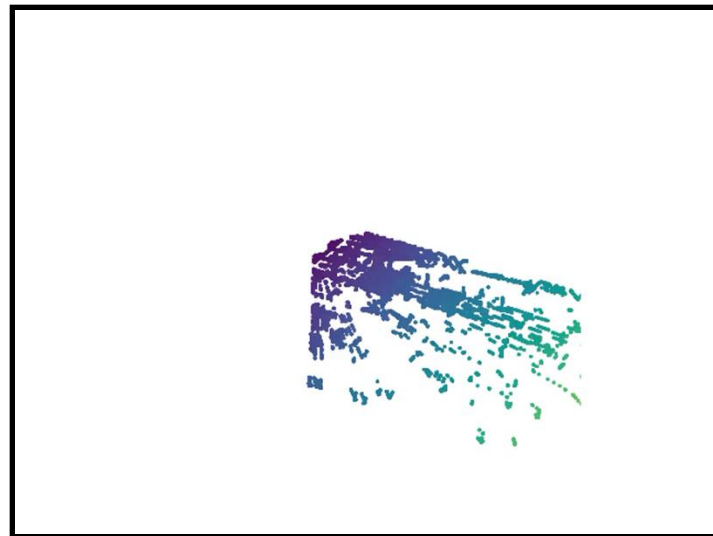
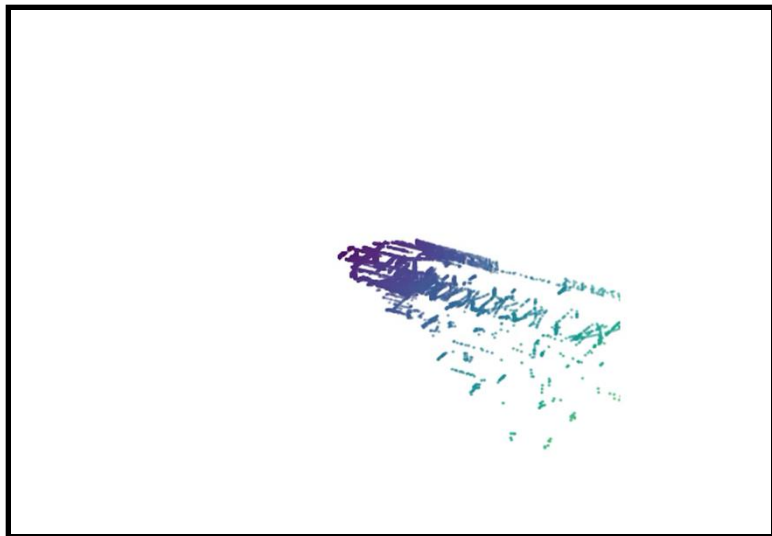
CARLA 模型
(仿真真实迁移)



我们的模型
(仿真真实迁移)



► 具身智能数据生成: sim-to-real 激光雷达数据迁移



Livox Horizon 固态激光雷达仿真 RS M1 固态激光雷达仿真 Pandar 64 机械激光雷达仿真

- 激光雷达波模型（物理特性）
- 物体表面反射率模型（环境交互）
- 天气介质传播模型（环境交互）



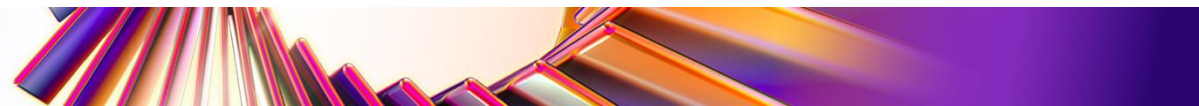
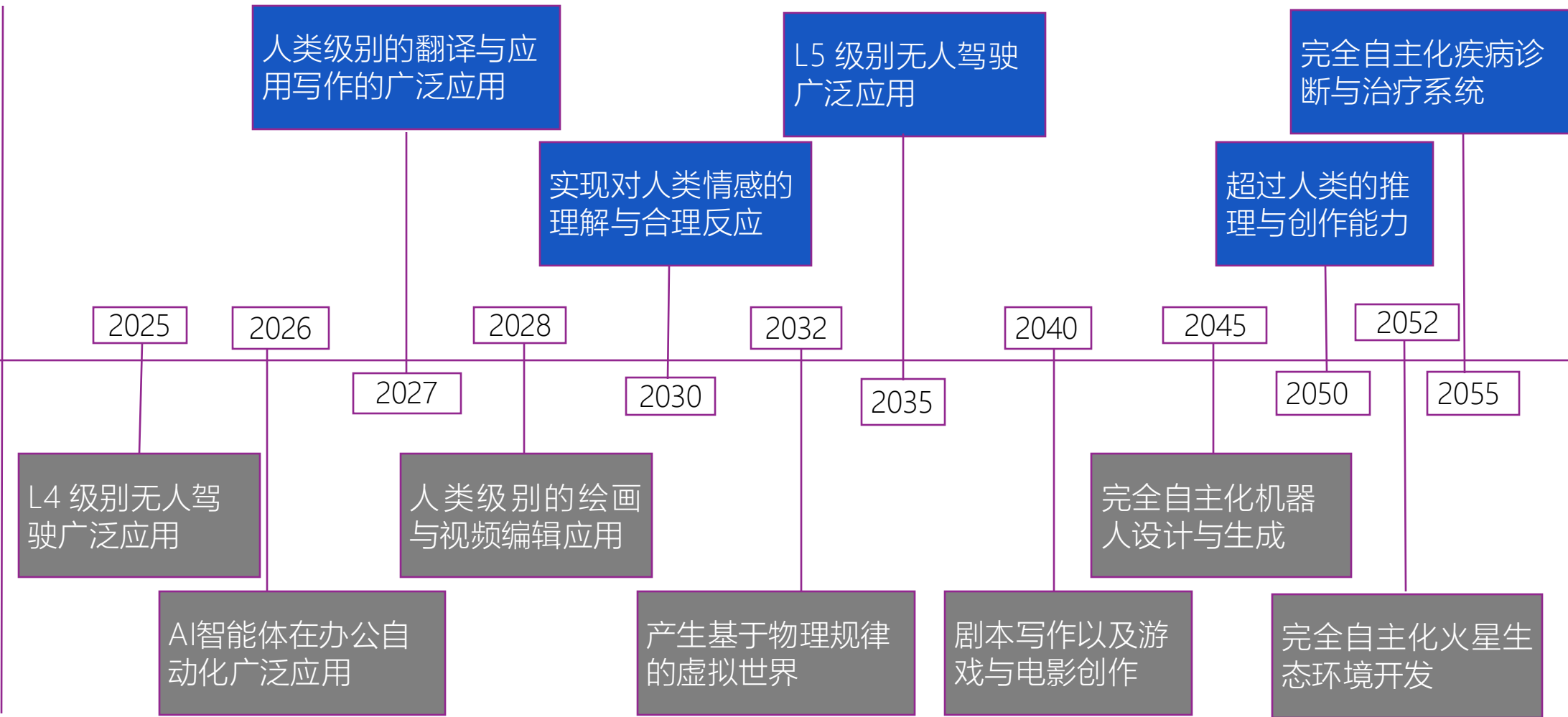
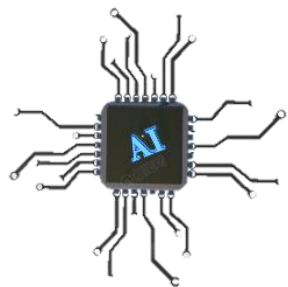
PART 05

未来与展望

具身智能未来展望



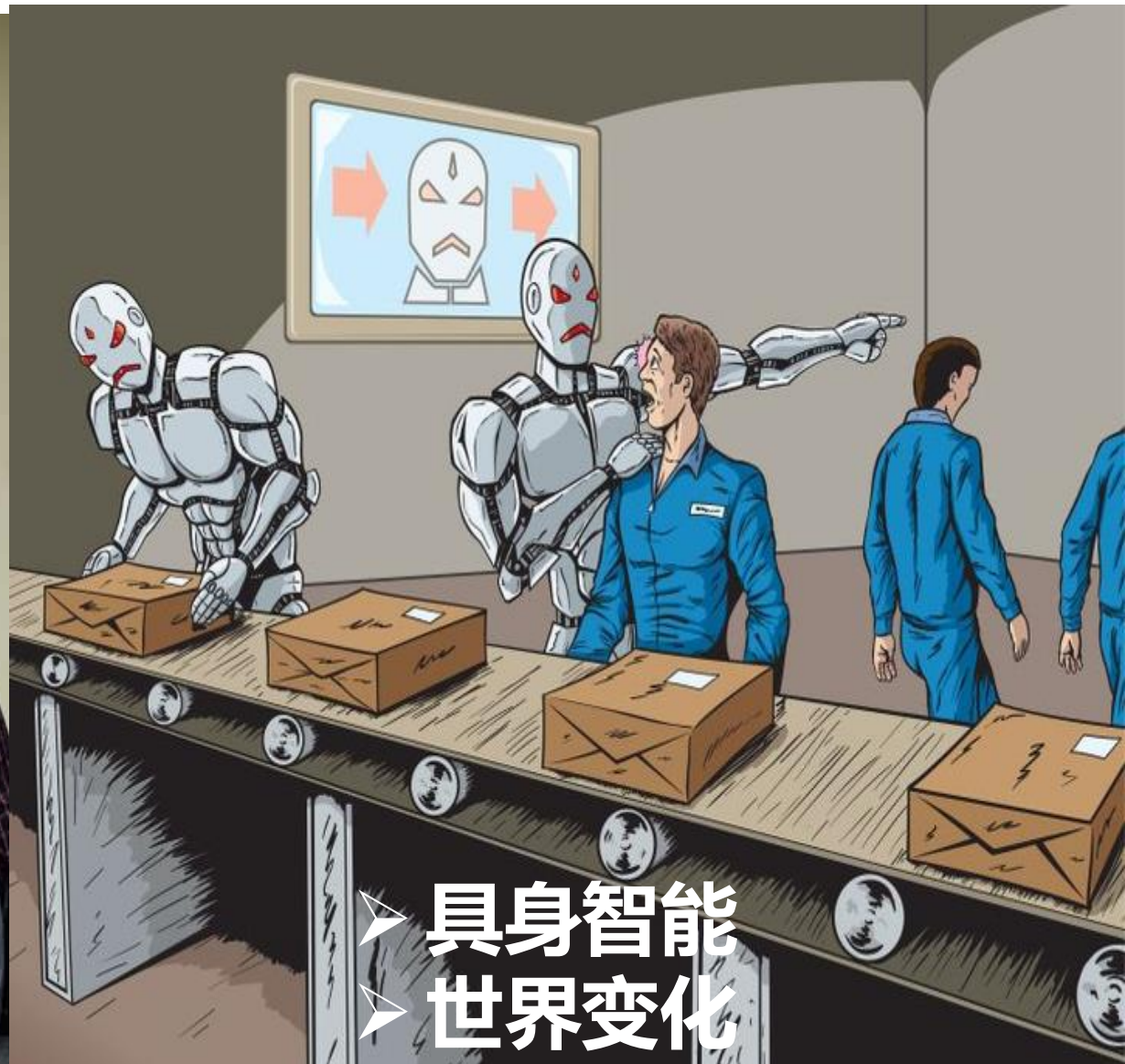
AI Technology



► 具身智能会给我们带来什么?



► 移动电话
► 生活变化



► 具身智能
► 世界变化



- 人机一体化
- 脑机芯片





IEEE 2021
ICRA
Xi'an, China



INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION

INTERNATIONAL CONFERENCE
ON ROBOTICS
AND AUTOMATION

June 1 2021

Plenary talk: Conference Hall

Keynote

Time	Session	Speaker	Time	Session
08:10 - 09:10	Behavior AI: Robot Cognitive Learning from Agent to Bcent	Fuchun Sun, Tsinghua	11:30 - 12:15	Perception and Intelligent performance Robotic Ma
09:15 - 10:15	Development of Intelligent Tunnel Boring Machine	Huayong Yang, Zhejiang	11:30 - 12:15	Perception System - from Driving to Autonomous
10:15 - 11:20	Full-Actuation: From Physical to Mathematical	Guangyuan, SUST	11:30 - 12:15	Bioinspired Design and Control of Highly Maneu Sw Robots

PLATINUM



GOLD



SLIVER



ACKNOWLEDGEMENT: WE HAVE A GREAT TEAM!

第8届 AI+研发数字峰会 | 拥抱 AI 重塑研发



科技生态圈峰会 + 深度研习

——1000+ 技术团队的选择



NiDD 峰会

NiDD 峰会 上海站

AI+研发数字峰会

时间: 2026.05.22-23

NiDD 峰会 北京站

AI+研发数字峰会

时间: 2026.08.21-22

NiDD 峰会 深圳站

AI+研发数字峰会

时间: 2026.11.20-21



AiDD峰会详情

NiDD × K AI+产品峰会

NiDD × K 上海站

AI+产品创新峰会

时间: 2026.01.16-17

NiDD × K 北京站

AI+产品创新峰会

时间: 2026.08.23-24



产品峰会详情



2026 AI+ PRODUCT INNOVATION SUMMIT

01.16-17 · ShangHai

AI+产品创新峰会



Track 1: AI 产品战略与创新设计

从0到1的AI原生产品构建

论坛1: AI时代的用户洞察与需求发现

论坛2: AI原生产品战略与商业模式重构

论坛3: Agentic AI产品创新与交互设计

2-hour Speech: 回归本质



用户洞察的第一性

——2小时思维与方法论工作坊

在数字爆炸、AI迅速发展的时代，
仍然考验“看见”的“同理心”



Track 2: AI 产品开发与工程实践

从1到10的工程化落地实践

论坛1: 面向Agent智能体的产品开发

论坛2: 具身智能与AI硬件产品

论坛3: AI产品出海与本地化开发

Panel 1: 出海前瞻



“出海避坑地图”圆桌对话

——不止于翻译:

AI时代的出海新范式



Track 3: AI 产品运营与智能演化

从10到100的AI产品运营

论坛1: AI赋能产品运营与增长黑客

论坛2: AI产品的数据飞轮与智能演化

论坛3: 行业爆款AI产品案例拆解

Panel 2: 失败复盘



为什么很多AI产品“叫好不叫座”?

——从伪需求到真价值:

AI产品商业化落地的关键挑战

智能重构产品 数据驱动增长
Reinventing Products with Intelligence, Driven by Data

80+

业界大咖

汇聚产品大咖的真知灼见

40+

创新案例

开拓全新AI+产品视角

10+

分论坛

涵盖产品到商业增长的全链路

800+

听众规模

共同探索产品的智能重构



扫码查看会议详情



第8届 AI+ 研发数字峰会
AI+ Development Digital Summit

感谢聆听!

扫码领取会议PPT资料

