



北京航空航天大學
BEIHANG UNIVERSITY

Beihang University

具身智能与智能机器人



北京航空航天大学
刘偲

个人介绍

刘偲，北航人工智能学院**副院长**，教授，博导，国家优青

- 研究方向是跨模态分析、具身智能
- 共发表CCF A类论文100余篇，其中包括**IEEE T-PAMI 14篇**。Google Scholar引用**15000+次**

所获奖项：

- 国家科技进步二等奖 (9/10)
- 中国图象图形学学会自然科学奖一等奖 (1/5)
- 中国图象图形学学会石青云女科学家奖
- ACM MM 2013、2021 最佳论文奖
- IJCAI 2021 最佳视频演示奖
- 10余项CCF A类会议竞赛冠军

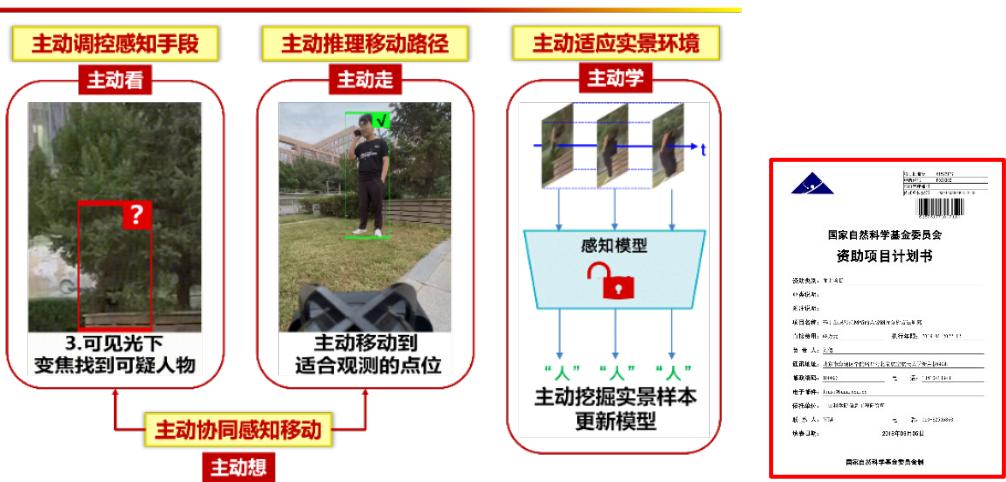


国内外学术任职：

- CCF-A类会议领域主席 (AC) : ICCV/CVPR/NeurIPS/MM
- 国际顶级期刊编委 (AE) : IEEE TCSVT/IEEE TMM/CVIU
- 中国图象图形学学会理事、副秘书长



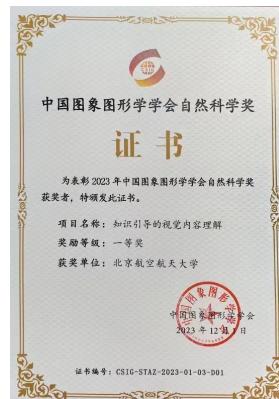
相关项目及获奖



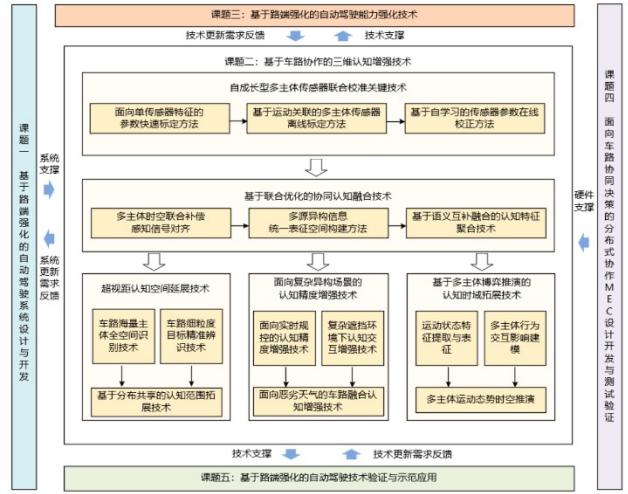
企业创新发展联合基金重点项目

《动态非结构场景海量异构信息主动感知技术研究》

知识引导的 视觉内容理解



中国图像图形学会自然科学奖一等奖



科技创新2030 — “新一代人工智能”重大项目-百度合作 《基于路端强化的自动驾驶决策关键技术》

面向公共安全的大规模 监控视频智能处理技术 及应用



中国图象图形学学会自然科学奖

视觉内容理解

底层：显著区域定位

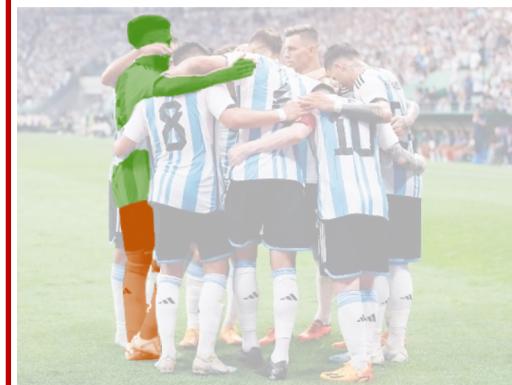
视觉信号**繁杂**
关键信息**稀疏**



找不准

中层：目标解析

视觉表观**相似**
目标相互**遮挡**



● 上身 ● 下身

分不清

高层：视觉推理

视觉信息**单一**
多模态数据**异构**



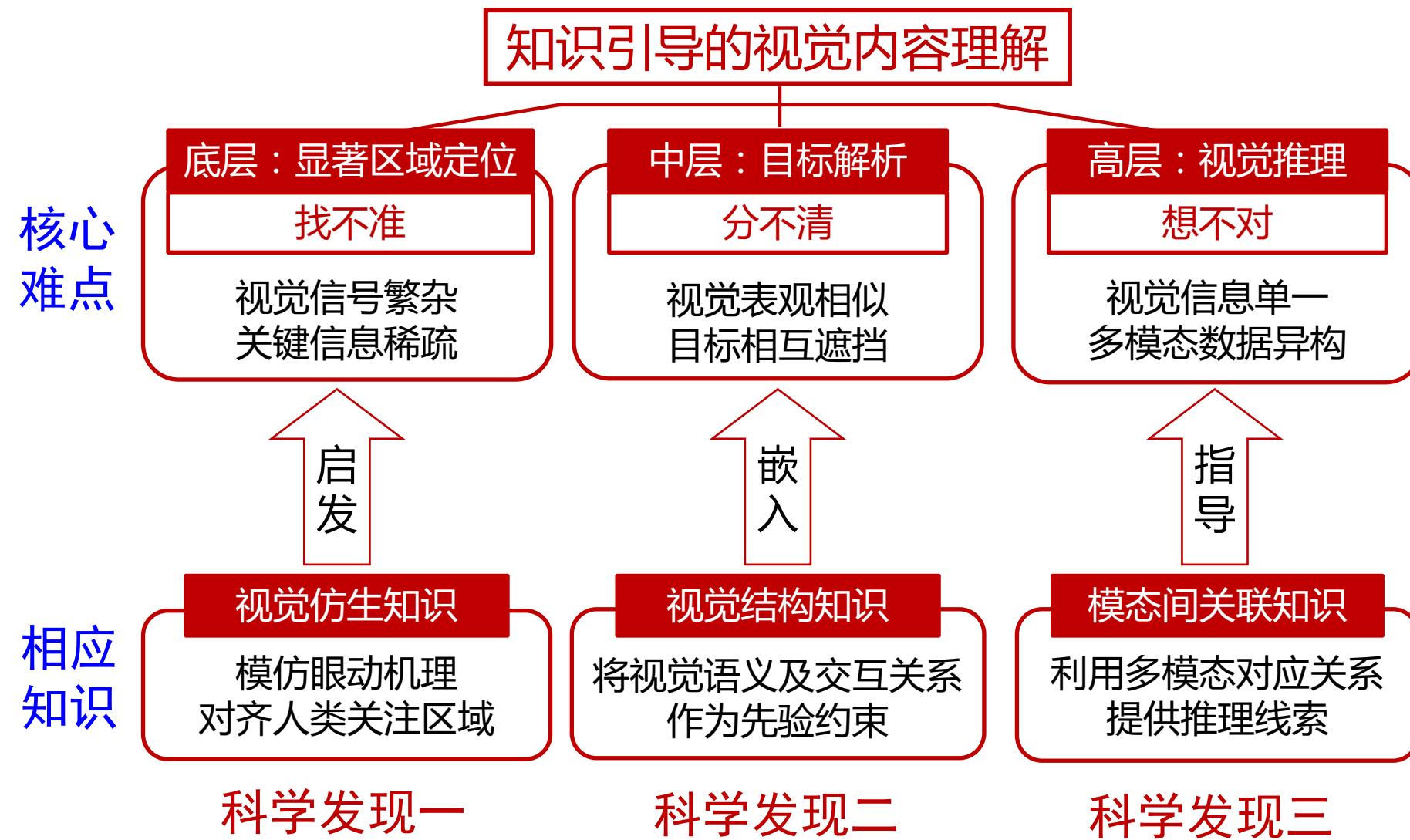
问题：栏杆为什么落下？

- A. 火车来了
- B. 停车场关门了
- C. 发生交通事故了
- D. 前方正在修路

想不对

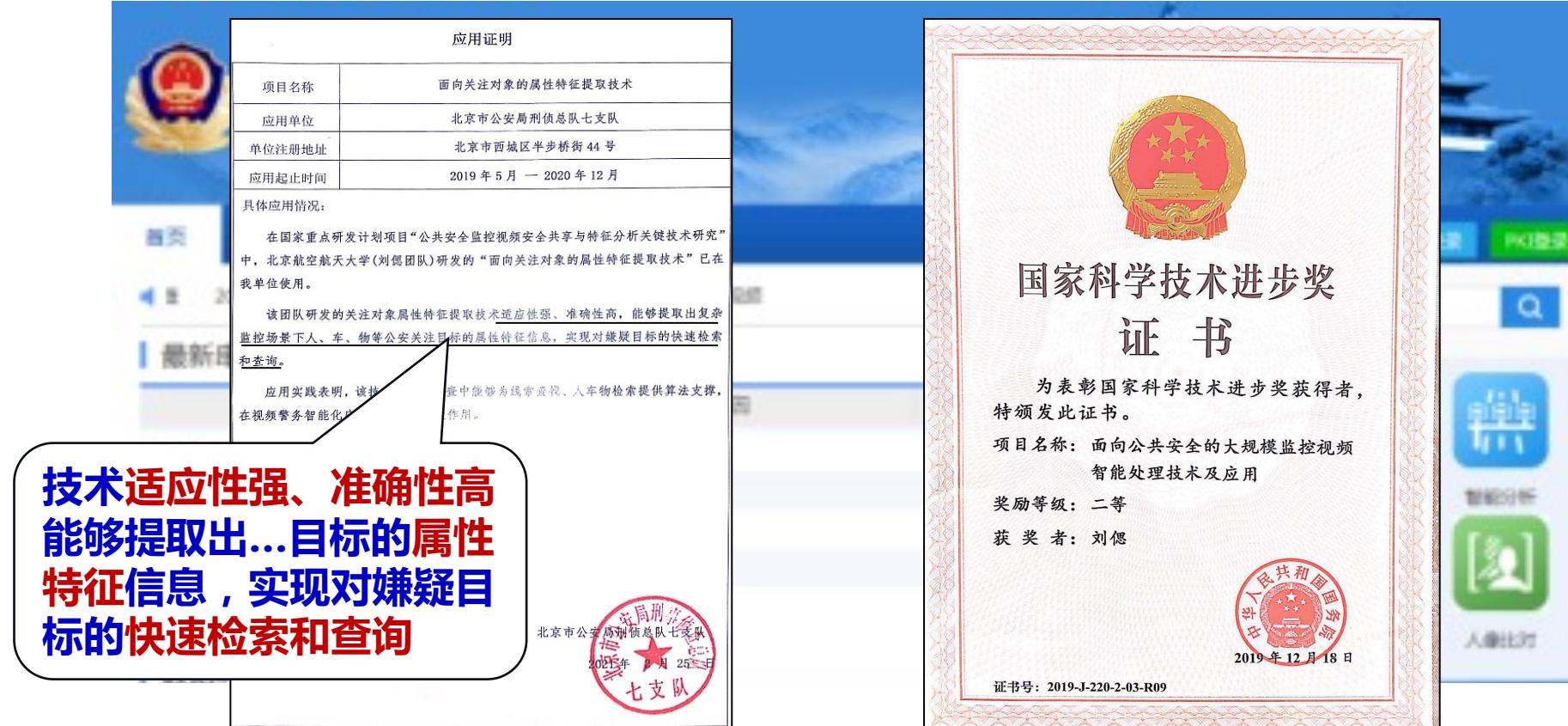
关键科学问题：如何利用任务适配的知识
针对性地引导不同层次的视觉内容理解

中国图象图形学学会自然科学奖



国家科技进步奖

□ 视频人像属性识别技术应用于视频侦查平台，提升案件侦破效率



**技术适应性强、准确性高
能够提取出...目标的属性
特征信息，实现对嫌疑目
标的快速检索和查询**



北京市公安局刑事侦查总队

**面向公共安全的大规模监控视频
智能处理技术及应用
2019年国家科学技术进步二等奖**

成果应用



科学技术应用证明

技术名称	人-物体交互定位技术
应用单位	北京市商汤科技开发有限公司
应用时间	2019年11月-至今

应用情况说明

我司与北京航空航天大学刘偲副教授团队开展人-物交互定位技术项目合作，针对“智能车舱”中的驾驶员危险动作检测问题，刘偲副教授团队负责完成了关键技术的研究。基于该技术的系统已经部署于我司“智能车舱”解决方案中，应用效果优秀。该解决方案已与超过30家国内外头部伙伴展开合作，定点量产项目覆盖车辆总数超过1300万辆。携手长城WEY、奇瑞捷途、哪吒，本田，上汽等汽车品牌共同打造了搭载该解决方案的多款汽车。

该算法是一种基于深度学习的单阶段的人-物交互定位算法。该技术具有很好的鲁棒性与稳定性，可以在多种硬件平台上高速、精准地检测出产生动作的人以及和人发生动作交互的物体。

本证明仅供刘偲副教授申请专利时使用。未经商汤公司允许禁止泄

本田、上汽、奇瑞等超过30家合作企业



应用于30余家汽车合作企业
车舱市场占有率全国第一



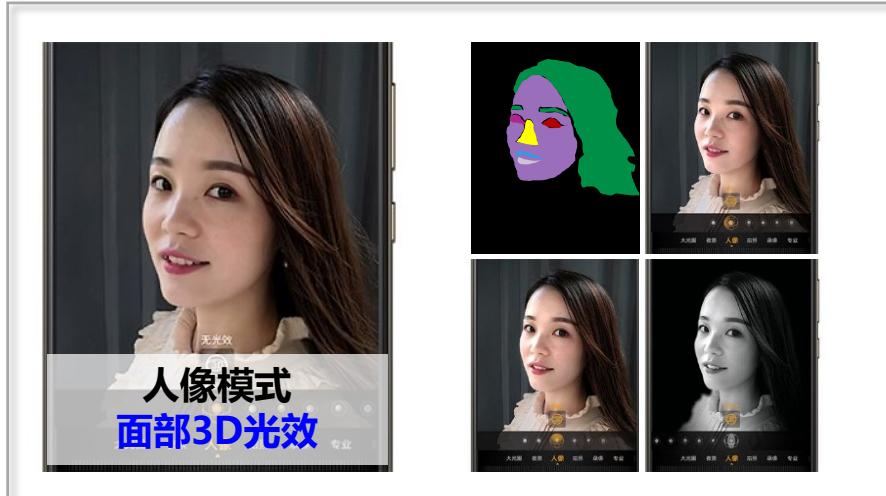
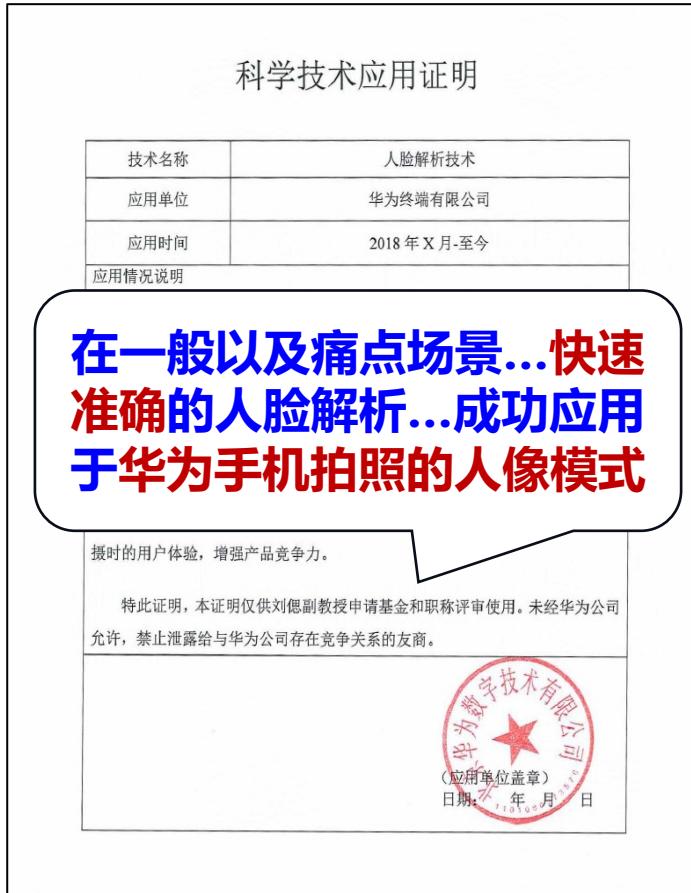
...



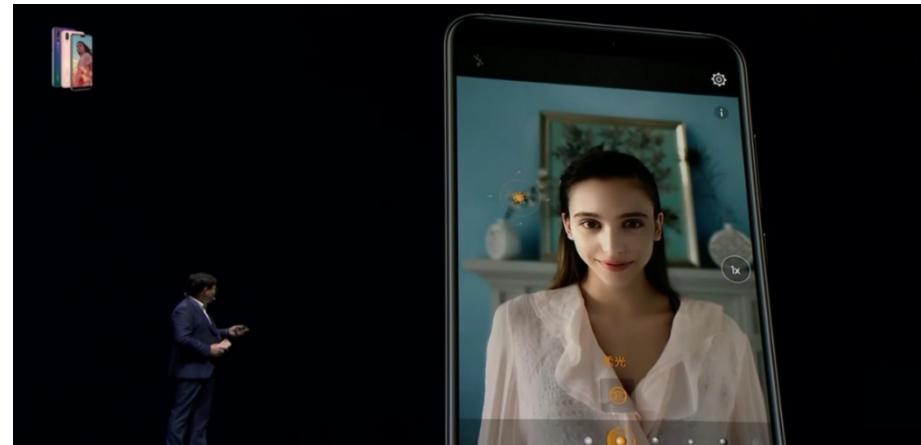
中国环球电视网报道
商汤科技智能车舱产品

成果应用

人脸解析技术应用于华为
P20/30/40、Mate30/40系
列旗舰手机，服务上亿用户



人脸解析辅助实现自然的面部打光效果

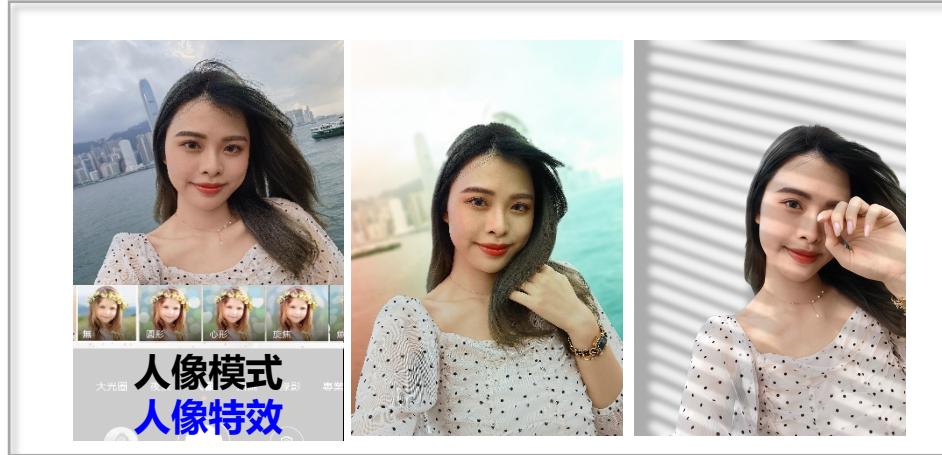


2018.3.27，华为消费者BG CEO余承东
在华为P20系列发布会重点介绍3D光效

成果应用

视频人体实例分割技术应用于华为P40和Mate40系列旗舰手机

表 2 应用情况说明	
(应用情况说明无格式要求, 此表只供参考)	
应用情况说明	拍照方式/分辨率
速度实时，精度高，成功应用于华为手机拍照功能的人像模式...提高终端产品在人像模式方面的竞争力	
技术的研究。该技术具有速度实时、精度高、稳定性好等特性，成功应用于华为手机拍照功能的人像模式，提高华为手机拍照功能在人像模式下的特性的效果，改善用户体验，提高终端产品在人像模式方面的竞争力。	
声明	我单位保证上述提供的应用情况真实无误。本说明仅作为申请职称等学术活动使用，不作为双方知识产权约定的依据。  法人单位盖章： 年 月 日 注：如表中所填内容不涉及经济效益情况，只需加盖应用单位法人公章。



通过人体实例分割实时定制拍照背景



2020.3.26，华为消费者BG CEO余承东在华为P40系列发布会重点介绍人像特效

01

什么是具身智能

具身智能-什么是具身智能-国内政策

具身智能契合国家战略需求

- 2023年1月，工信部等十七部门印发《“机器人+”应用行动实施方案》指出：到2025年，制造业机器人密度较2020年实现翻番，服务机器人、特种机器人行业应用深度和广度显著提升
- 2023年10月《人形机器人创新发展指导意见》指出：到2025年，人形机器人创新体系初步建立，“大脑、小脑、肢体”等一批关键技术取得突破，确保核心部组件安全有效供给
- 2024年3月《政府工作报告》指出：加快发展新质生产力。智能机器人作为新质产品，可能是进入千家万户的“四大件”之一

到2025年，制造业机器人密度较2020年实现翻番，服务机器人、特种机器人行业应用深度和广度显著提升，机器人促进经济社会高质量发展的能力明显增强。聚焦10大应

到2025年，人形机器人创新体系初步建立，“大脑、小脑、肢体”等一批关键技术取得突破，确保核心部组件安全有效供给。整机产品达到国际先进水平，并实现批量生产，在

(一)大力推进现代化产业体系建设，加快发展新质生产力。充分发挥创新主导作用，以科技创新推动产业创新，加快推进新型工业化，提高全要素生产率，不断塑造发展新动能新优势，促

五是新的产品和用途。每一个时代都有属于那个时代进入千家万户的“四大件”“五大件”，近几十年是家电、手机、汽车等等，未来可能是家用机器人、头戴式VR/AR设备、柔性显示、3D打印设备和智能汽车等等。

具身智能-什么是具身智能-国外政策

具身智能契合世界战略需求

- 美国：2024年4月28日，美国计算机社区联盟发布第五版《美国机器人路线图：机器人让明天更美好》，分析了机器人在人工智能（AI）**具身**、劳动力等方面的大趋势，提出了劳动力、精益物流（lean-logistics）等方面的挑战，最终映射到**智能具身**、操纵、感知、控制、规划等方面的研究机会。
- 日本：2023年1月12日统计指出，日本政府根据《机器人新战略》已在机器人领域投入超过了**9.305亿美元**，包括将成为下一代**人工智能和机器人核心**的集成技术。
- 欧盟：已在积极制定《人工智能驱动（AI-powered）》的机器人战略，指出**机器人市场快速发展**，日益受到制造、救援、检索、医疗保健、物流、农业等领域产品迭代的推动。同时，数百万个相关工作岗位将受到影响。

具身智能-什么是具身智能-国外政策

具身智能契合世界战略需求

- 美国：2018年，美国国防高级研究计划局（DARPA）分析了机器人的发展趋势，提出“具身智能”（embodied intelligence）概念，认为机器人应具备在真实世界中自主学习和适应的能力。
- 日本：2019年，日本政府发布了《日本未来战略》，提出将投资于“具身智能”领域的研究，以促进制造业、物流业等领域的智能化发展。
- 欧盟：已启动“地平线2020”计划，投入大量资金支持“具身智能”领域的研究，目标是通过与环境交互、感知、自主规划、决策、行动以及执行的能力，使机器人能够更好地服务于人类社会。

具身智能 (Embodied Artificial Intelligence)
是指智能体（包括具有实体的机器人等），通过与环境产生交互以及自身的学习，可以具备像人一样能与环境交互、感知、自主规划、决策、行动以及执行的能力

具身智能-什么是具身智能-具身智能演示



室内避障



灵活动作



精巧操作



动态抓取



自动灌溉



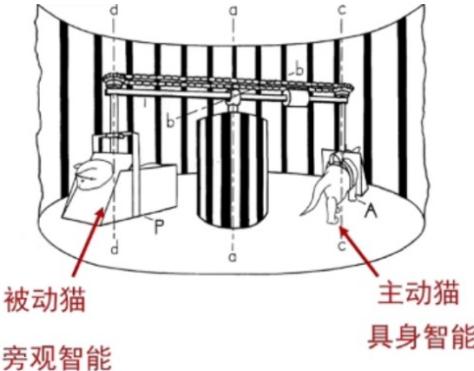
协同探索

具身智能的应用场景

具身智能-什么是具身智能-第一人称与第三人称智能

具身智能机器人需要以第一人称身份融入周边环境

➤ 第一人称与第三人称交互方式



实验中，主动猫学会了正常行走
但被动猫失去行走能力



进一步理解行为：机器人第一人称呼自主地感知世界，与世界交互，完成复杂行为

➤ 第一人称与第三人称智能

第三人称智能



第一人称智能

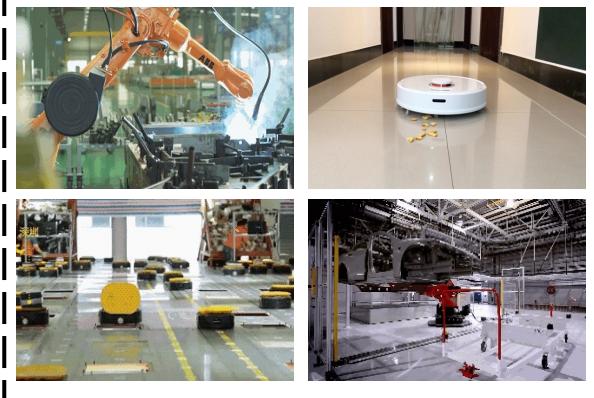


可以打开，可以装东西
我亲身体验盒子是什么

具身智能-什么是具身智能-具身智能VS传统机器人

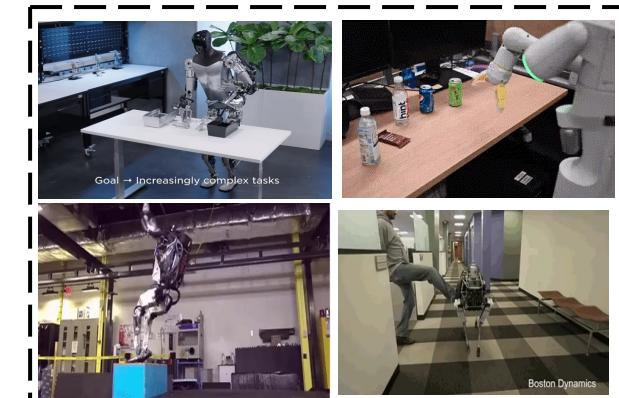
具身智能可以与环境**主动交互**、拥有**自主规划、推理、执行、学习能力**、
第一人称智能体

小脑
运动能力的关键



传统机器人研究聚焦于**小脑**层面
负责**底层运动控制**

大脑
思考能力的关键



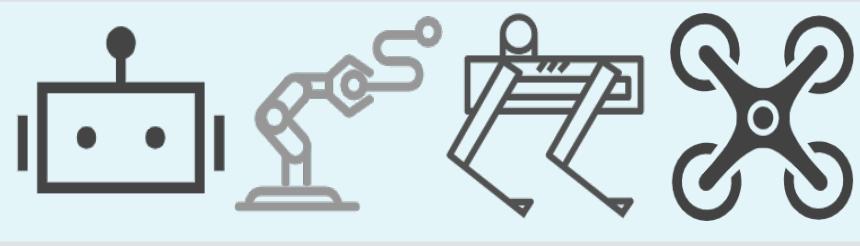
具身智能研究聚焦于**大脑**层面
负责**顶层规划决策**



VS

具身智能-什么是具身智能-核心要素

要素一：本体
要具有运动实体



要素二：智能体
要具有智能能力

决策

知识学习
思维推理



要素三：数据

行动

复杂操作
行动泛化

与环境的交互
呈现拟人化的交互

感知

开放环境
感知能力

要素四：学习和进化架构

要能与环境交互



物理环境



仿真环境

具身智能-什么是具身智能-本体

- 本体是具身智能的**物理载体**，负责在**物理或虚拟世界**中进行感知和任务执行
- 具备**环境感知能力、运动能力和操作执行能力**
- 本体形态日益多样化：四足，人形等

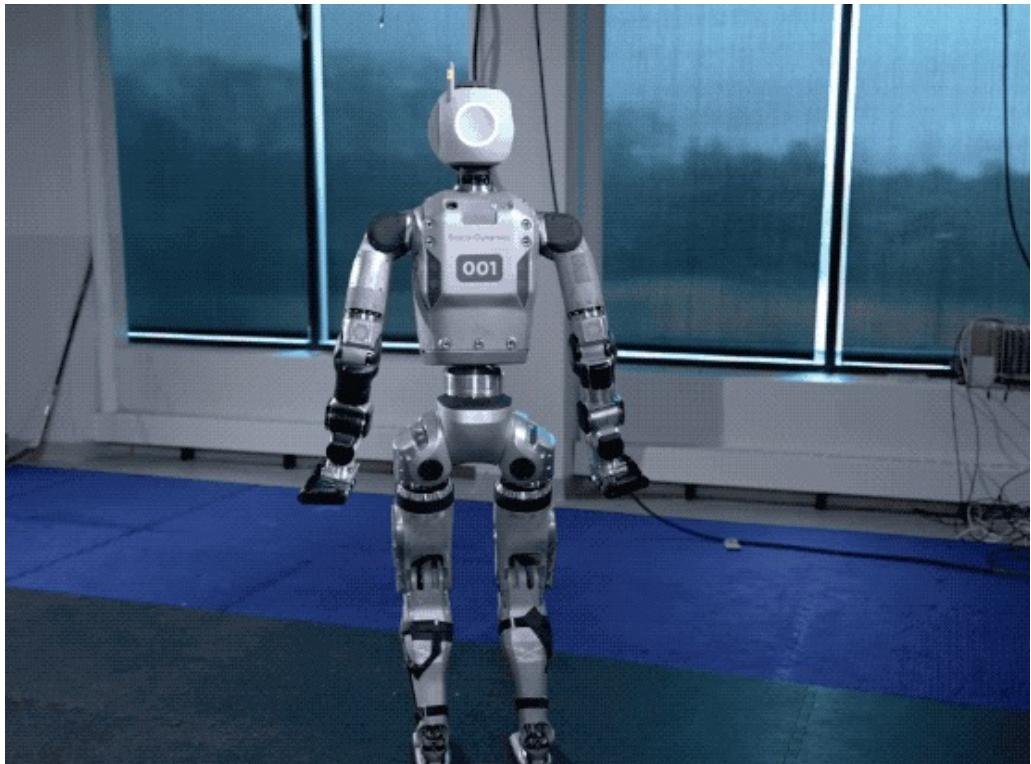
Unitree Go2

- 宇树开发的四足机器人
- 适用于多种应用场景，具有极强稳定性



Atlas 001

- 波士顿动力开发的人形机器人
- 关节可定制、高功率且非常灵活，使机器人获得了巨大的运动范围



具身智能-什么是具身智能-智能体

- 智能体是具身于本体之上的**智能核心**，负责**感知、理解、决策和控制**等核心工作
- 理解环境所包含的**语义信息**，并根据**环境变化和目标状态做出决策**，进而控制本体完成任务
- 通常由深度网络模型驱动，特别是**大语言模型（LLM）和视觉语言模型（VLM）**的结合

LLM

- 基于深度学习的自然语言处理模型
- 通过海量文本数据训练，能够理解和生成自然语言



GPT4



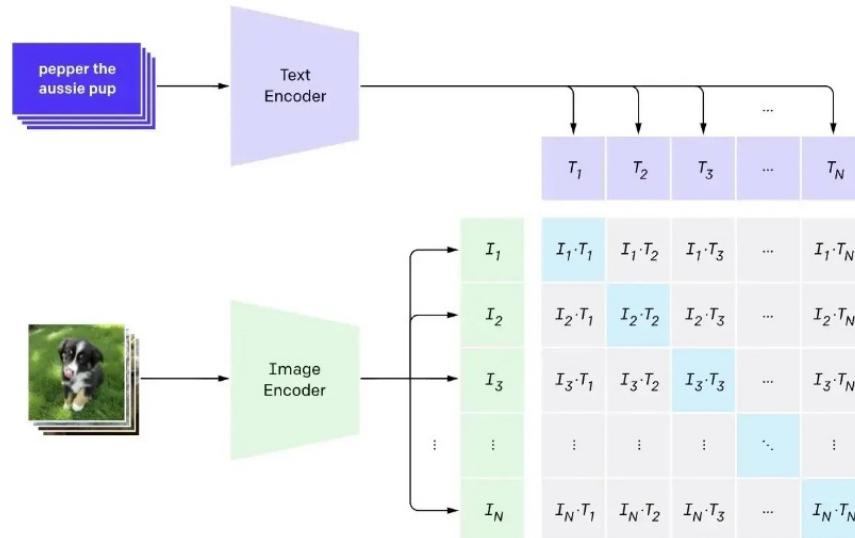
PaLM2



LLaMA3

VLM(CLIP)

- 通过对比学习将图像和文本进行关联学习

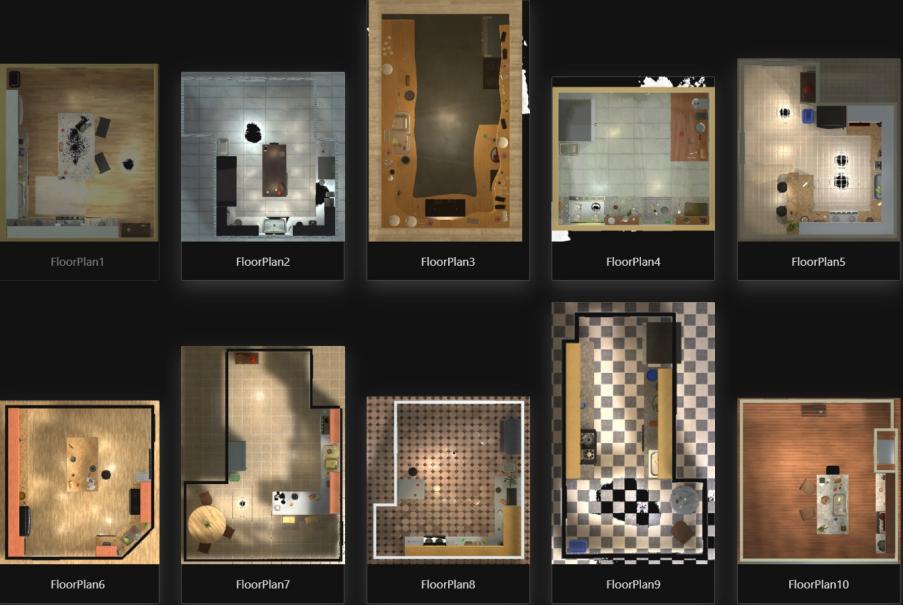


具身智能-什么是具身智能-数据

- 数据是智能体进行学习和进化的**基础**
- 系统需要**海量数据**来训练学习
- 高质量**数据稀缺且昂贵**

AI2-THOR

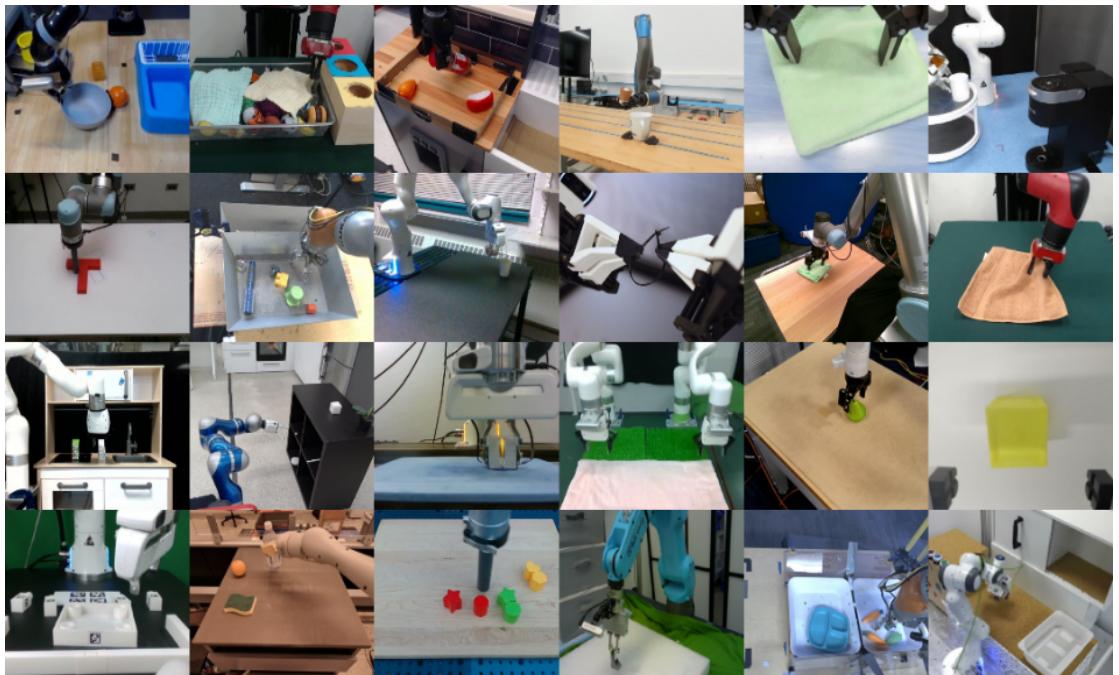
- 艾伦人工智能研究所提供了丰富的室内场景和交互任务
- 支持机器人导航、物体操作和视觉问答覆盖



FloorPlan1 FloorPlan2 FloorPlan3 FloorPlan4 FloorPlan5
FloorPlan6 FloorPlan7 FloorPlan8 FloorPlan9 FloorPlan10

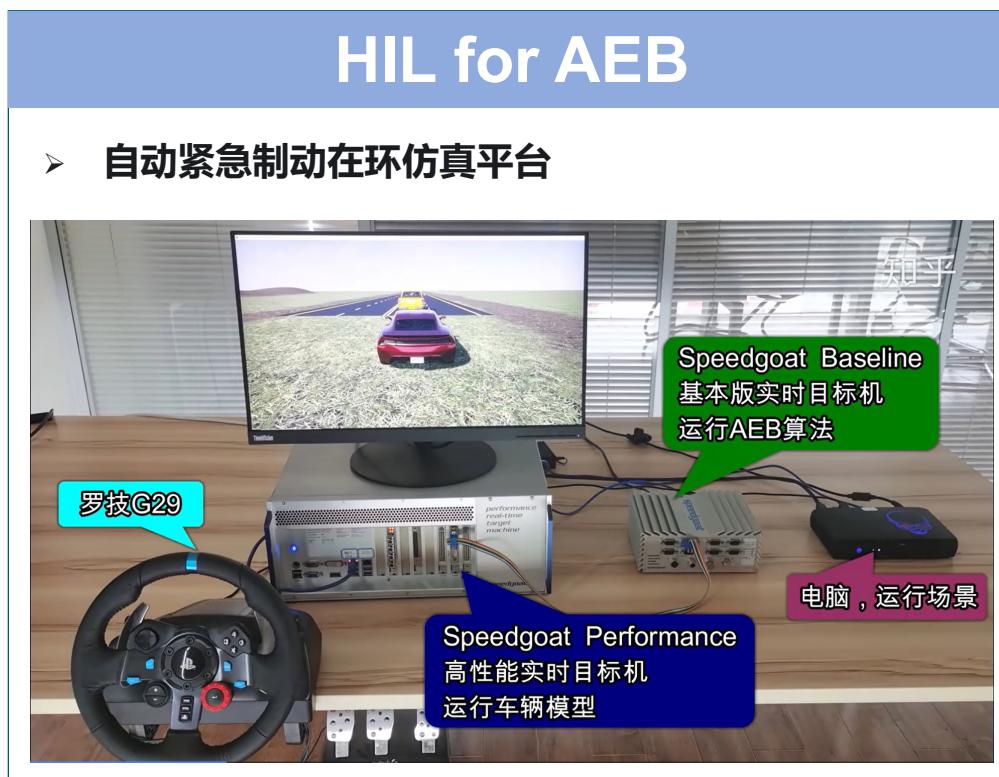
Open X-Embodiment

- 谷歌提供了用于机器人控制的**通用标准化数据集**
- 包含 **21 个机构**合作收集的 **22 个不同机器人的数据**
- 覆盖 **527 项技能，160266 项任务**



具身智能-什么是具身智能-学习和进化架构

- 学习和进化架构是智能体**适应新环境、学习新知识**并强化解决问题方法的关键
- 通过与**物理世界(虚拟的或真实的)**交互来不断学习和进化
- 真实环境，硬件在环仿真，**虚拟仿真**



Omniverse

- NVIDIA 提供的基于物理的实时仿真和协作平台
- 用于机器人导航、物体操作、路径规划等任务的仿真和测试

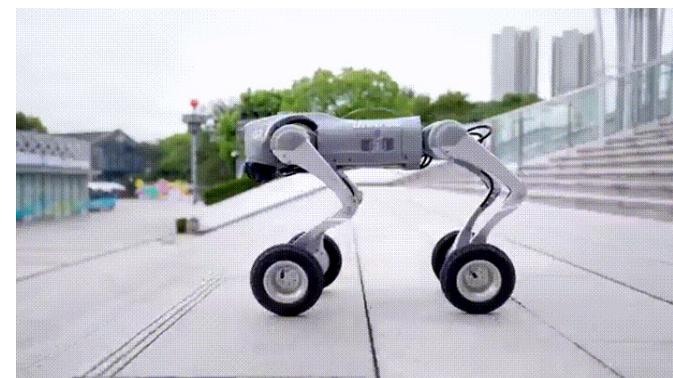
组件	名称	功能描述
	Omniverse Nucleus	NVIDIA Omniverse 的中央数据库和协作引擎，通过 Omniverse Nucleus，用户可以共享和修改虚拟世界的表现形式。
	Omniverse Connect	Omniverse Connect 是用于主流数字内容创建应用的插件。借助 Omniverse Connect，用户可以继续使用自己惯用的行业软件应用，如 SketchUp、Maya 和虚幻引擎，同时受益于其他 Omniverse 工具。
	Omniverse Kit	用于构建原生 Omniverse 应用、扩展程序和微服务的工具包。
	Omniverse RTX Renderer	基于 NVIDIA RTX 的高级、多 GPU 渲染器，支持实时光线追踪和路径追踪。
	Omniverse Simulation	一套功能强大的工具和 SDK，可模拟物理级精准的世界。

02

核心要素：本体

具身智能-本体-具身系统在物理世界的核心驱动力

具身智能-本体-具身系统
在物理世界的核心驱动力



强大的“小脑”释放了上层决策的压力，增强了任务执行的可靠性

具身智能-本体-无人机

浙江大学: 无人机集群

- 在两年多的研究中，浙大科研团队解决了**未知复杂环境下**无人机单机与群体的**智能导航与快速避障方法**等一系列核心技术
- 该团队研发的微型智能空中机器人集群可以在**密集的竹林间穿梭**。除了茂密垂直生长的竹子外，还有其他种类的障碍物，包括倾斜的竹子、树干、低矮的灌木、杂草沟、不平整的地面等，这些机器人集群都能**完美的通过**



具备智能-本体-机械臂

ABB: 机械臂 Yumi

- ABB 拥有当今**最多种类的机器人产品、技术和服务**，是全球**装机量最大的工业机器人供货商**
- ABB 的核心技术是**运动控制系统**，这也是对于机器人自身来说最大的难点。掌握了运动控制技术的ABB可以轻易实现**循径精度、运动速度、周期时间、可程序设计等机器人的性能**，大幅度提高生产的质量、效率以及可靠性



具身智能-本体-机械狗

波士顿动力: 机械狗 spot

- Spot 参数 : 身高 0.84 米 , 自重 30 公斤 , 最大负载 14 公斤 , 电池供电 , 电气驱动 , 采用 3D 视觉系统 , 17 个关节点
- 可用于办公室和家庭环境中。Spot 机器人加上了 5 自由度手臂 , 从而具备了移动抓取物体的功能。传感器系统包括深度相机 , 立体相机 , 惯导模块和位置 / 力传感器 , 最终实现机器人全自主导航功能



具身智能-本体-双足机器人

逐际动力: 双足机器人 P1

- 逐际动力深耕强化学习，将前沿技术转化为研发力，提出Real2Sim2Real闭环、神经网络架构设计、数据生成机制与训练算法设计等三大体系，完善管理验证，助力人形机器人功能开发
- P1 是逐际动力在中国率先推出的一款新颖的双足机器人，也是逐际动力强化学习系统化研发与模块化测试的重要平台，用于推进双足基础运动能力的研发和迭代



具身智能-本体-人形机器人

波士顿动力：Atlas机器人

- 波士顿动力的Atlas机器人通过先进的感知算法和模型预测控制技术，实现了复杂的跑酷动作。它利用深度相机和多平面分割算法识别环境中的障碍物，并根据高级地图和实时感知数据规划行动路径。Atlas 的行为库和 MPC 使其能够灵活应对环境变化，展现出接近人类的动态运动能力



具身智能-本体-人形机器人

Unitree机器人公司：人形机器人 H1

- 国内最早提出技术方案，最早将其商业化的四足机器人公司
- 多年来占据全球出货量至少60%以上，处于领先地位

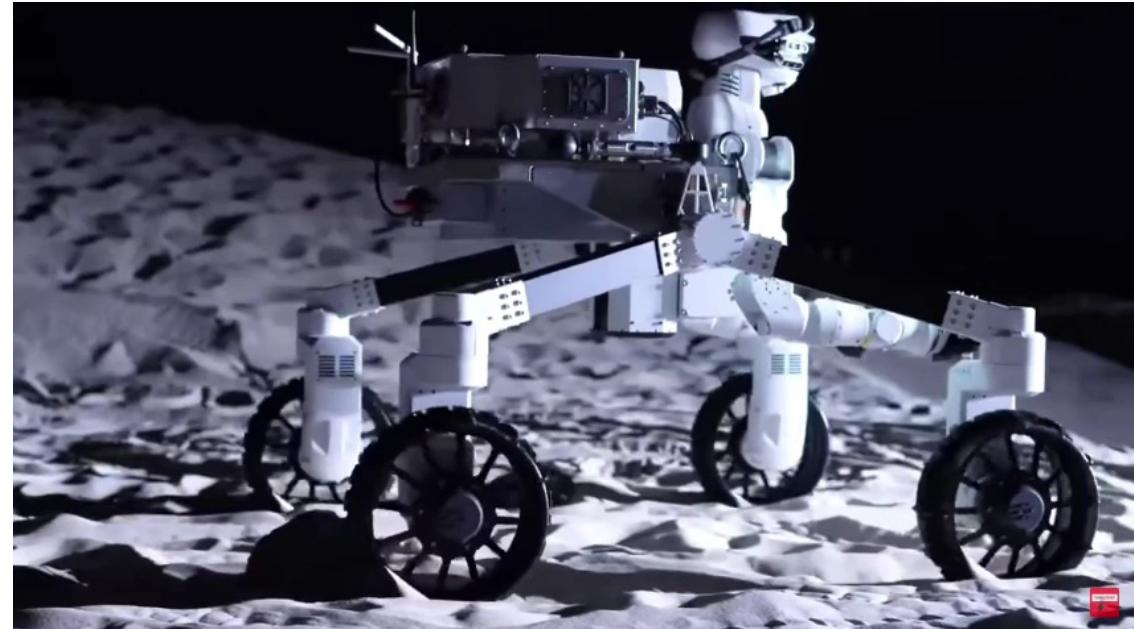
- 全尺寸纯电机驱动人形机器人H1，全球近似规格最高动力性能
- 采用先进的动力系统，在速度、力量、机动性和灵活性等方面具备最高水平



具身智能-本体-机器人

日本：太空通用人形机器人GITAI

- 日本研发的通用**人形GITAI机器人**在模拟太空环境中成功演示了**太空服务、组装和制造（ISAM）活动**
- 采用**人工智能和远程操作技术**
- GITAI 公司研制的**月球机器人R1**，在 JAXA的**模拟月面环境中**进行了操作演示。理论可以在月球上**执行勘探、采矿、检查、维护、组装等通用任务**



具身智能-本体-机器人

美国和法国：国际空间站CIMON机器人

- 美国和法国联合研发的 **CIMON** 机
器人参与国际空间站实验，首次任务
与两名欧洲宇航员合作

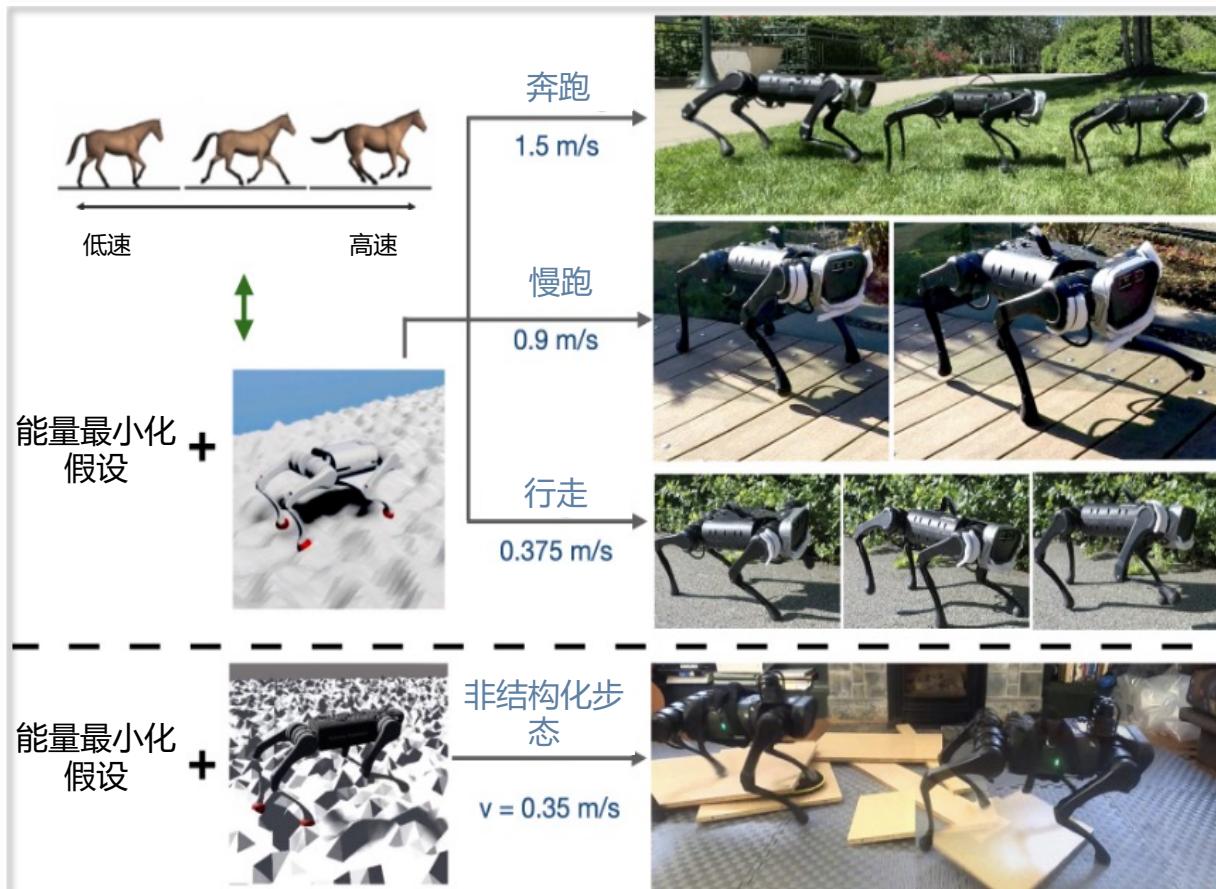
- 配备高精度传感器，能够**识别人脸、
声音并进行空间导航**

- 可以通过语音指令提供**程序指导，解
放宇航员双手**，进行空间导航

- 工具、控制板和科学设备的操作任务



具身智能-本体-机械狗（组内工作）



研究现状与算法动机

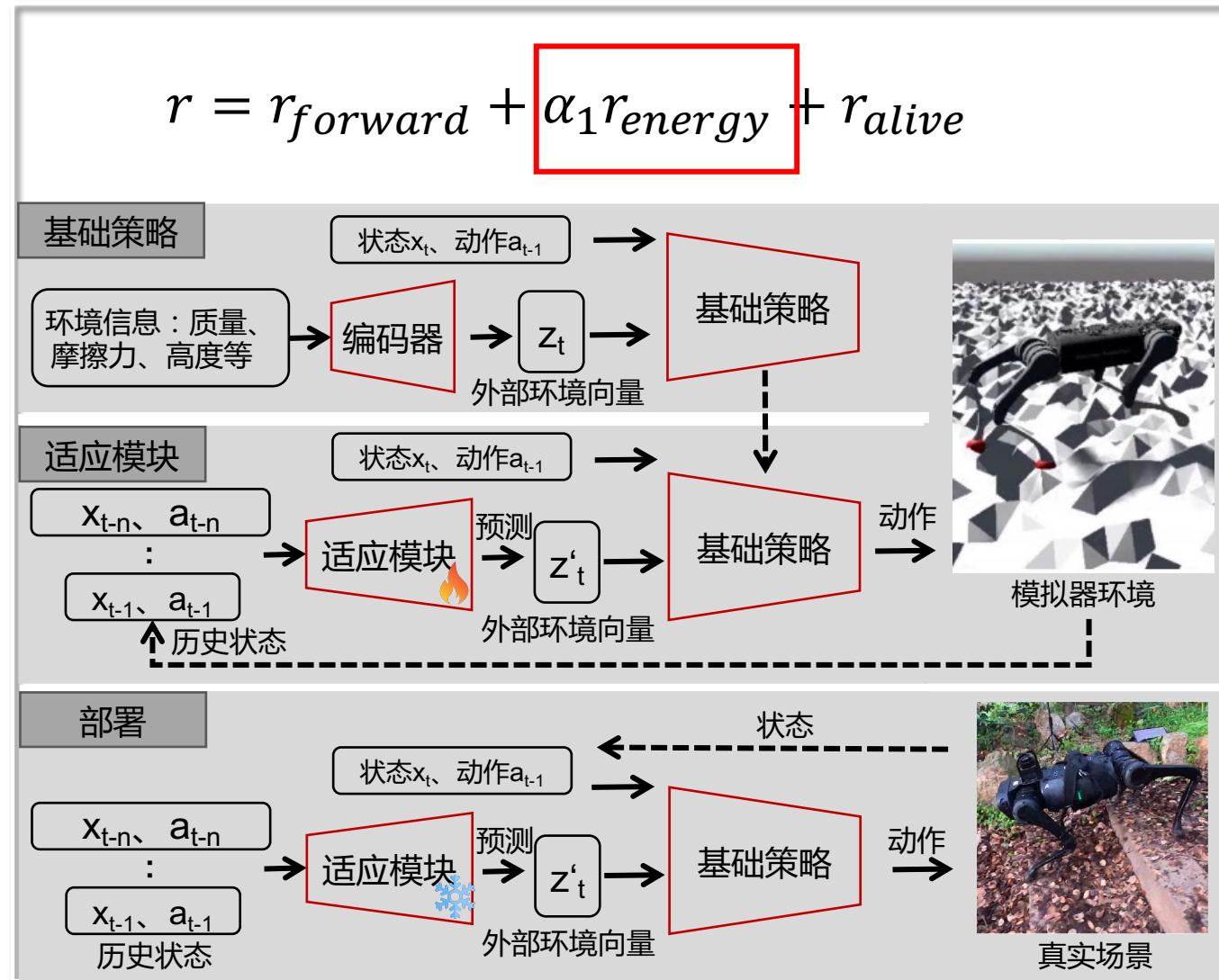
研究现状

- 传统的“**预编程步态**”算法限制了四足机器人在复杂环境下的通用性，无法满足具身系统的需要

算法动机

- 动物在不同速度和地形下会选用更加**节能**的步态
- 通过**最小化机械能的消耗**能够自发产生自然步态，即**最优步态 (能量最小化假设)**
- **最小化能量步态**可根据速度和地形及时调整步态，确保下游任务的**鲁棒性**

具身智能-本体-机械狗



技术策略

- 基础策略：基于强化学习，设计含速度、能量消耗惩罚、生存时间的奖励函数；引入**编码的环境信息**参与训练
- 适应模块：根据四足机器人状态和动作历史，预测**外部环境向量**，并最小化与真实环境信息的误差

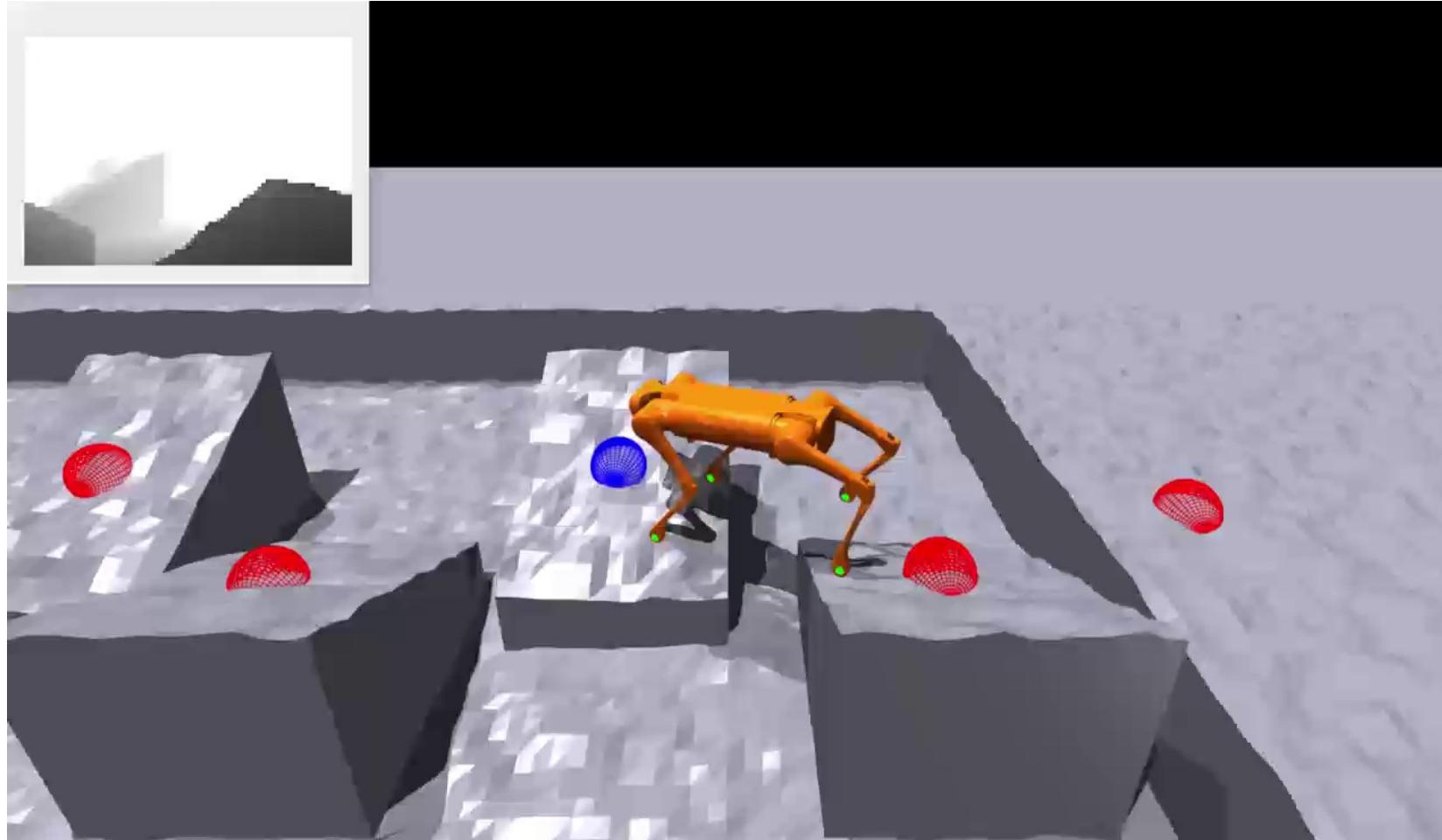
部署策略

- 每个步长生成一次**外部环境向量**，用于指导基础策略输出**关节位置**

具身智能-本体-机械狗

具身智能-本体-机械狗
自然步态生成算法
Colab实验室

Colab实验室：自然步态生成算法



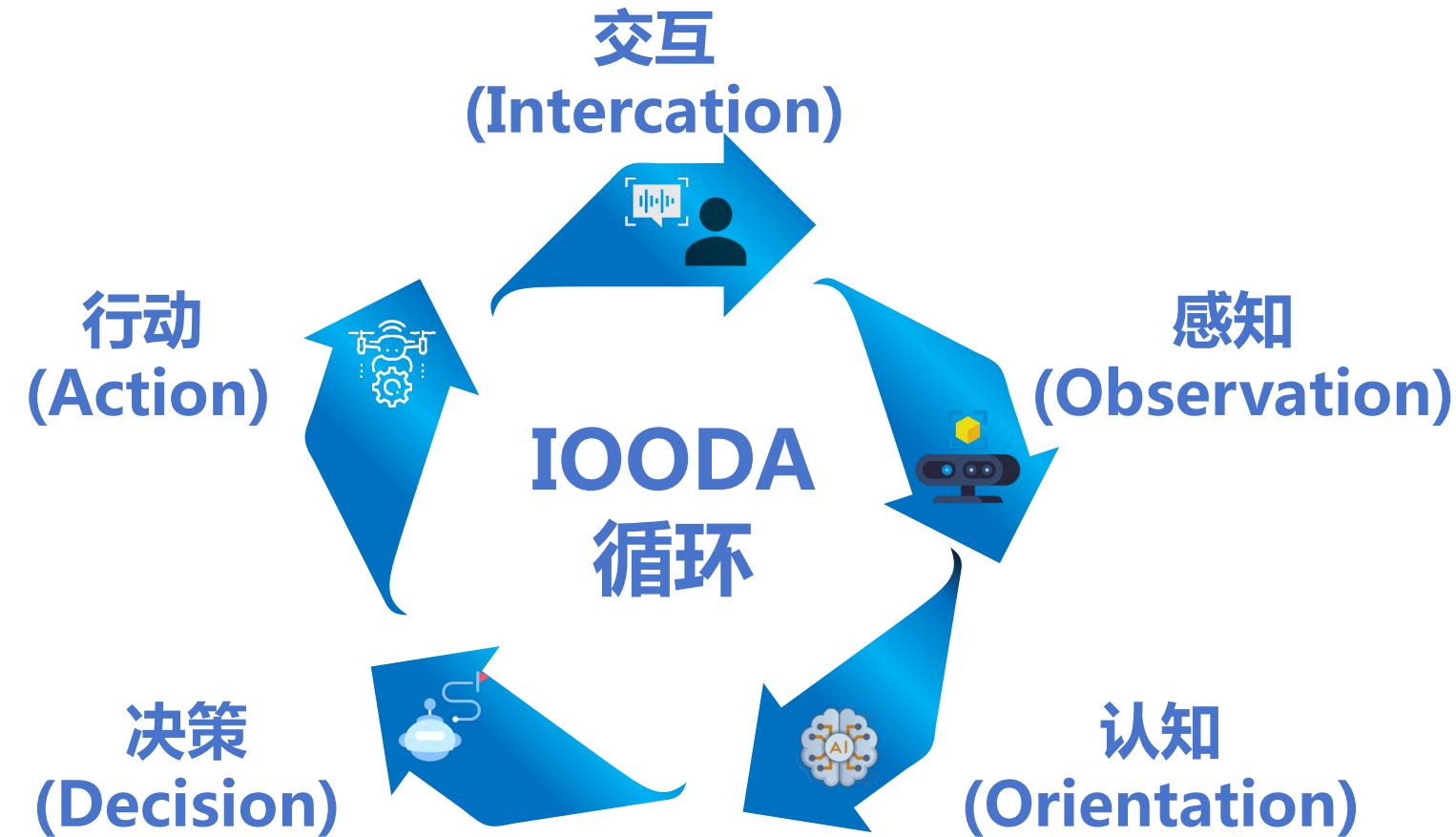
灵活适应地形，增强运动能力，显著提高下游任务的稳定性

03

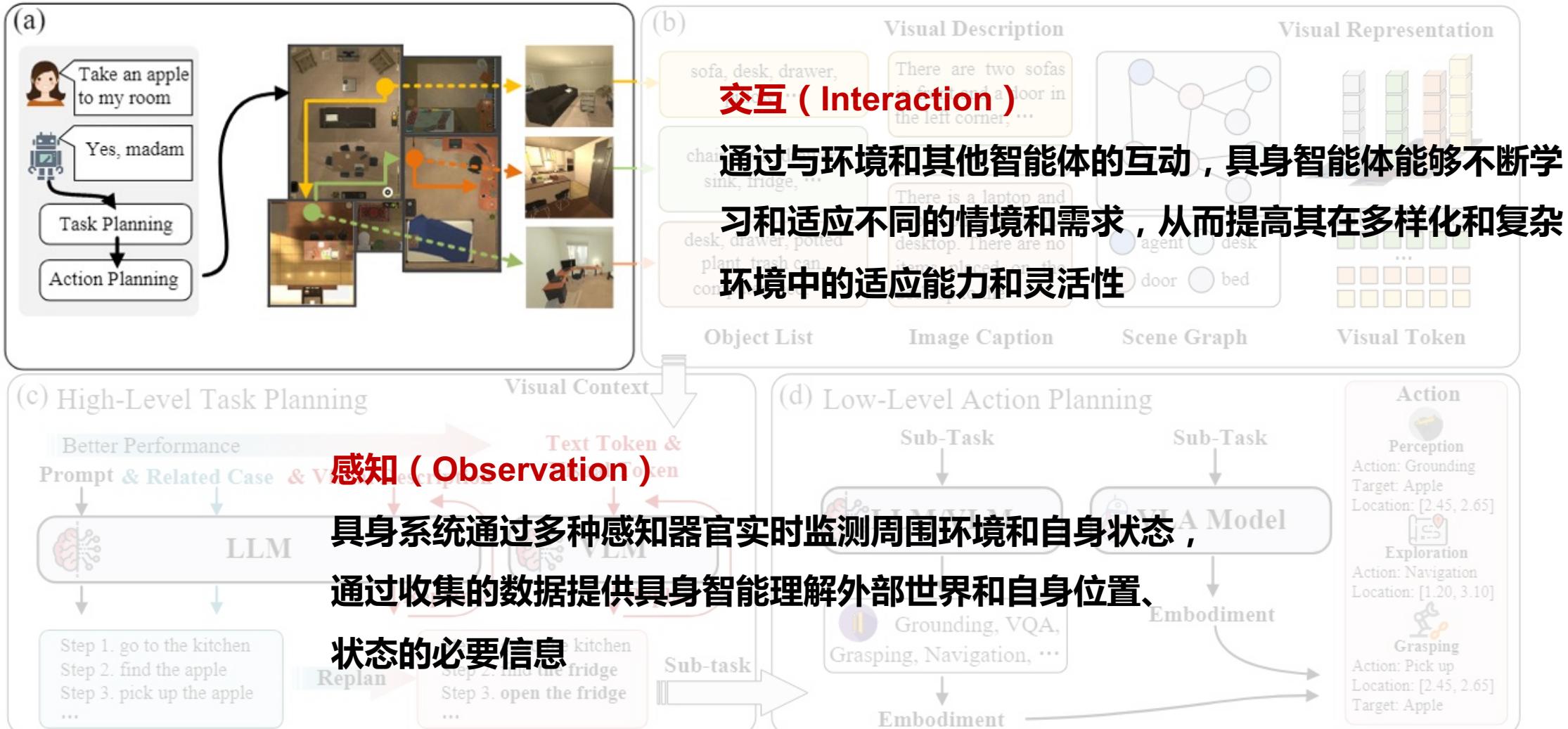
核心要素：智能体

具身智能-智能体-什么是“具身智能体”？

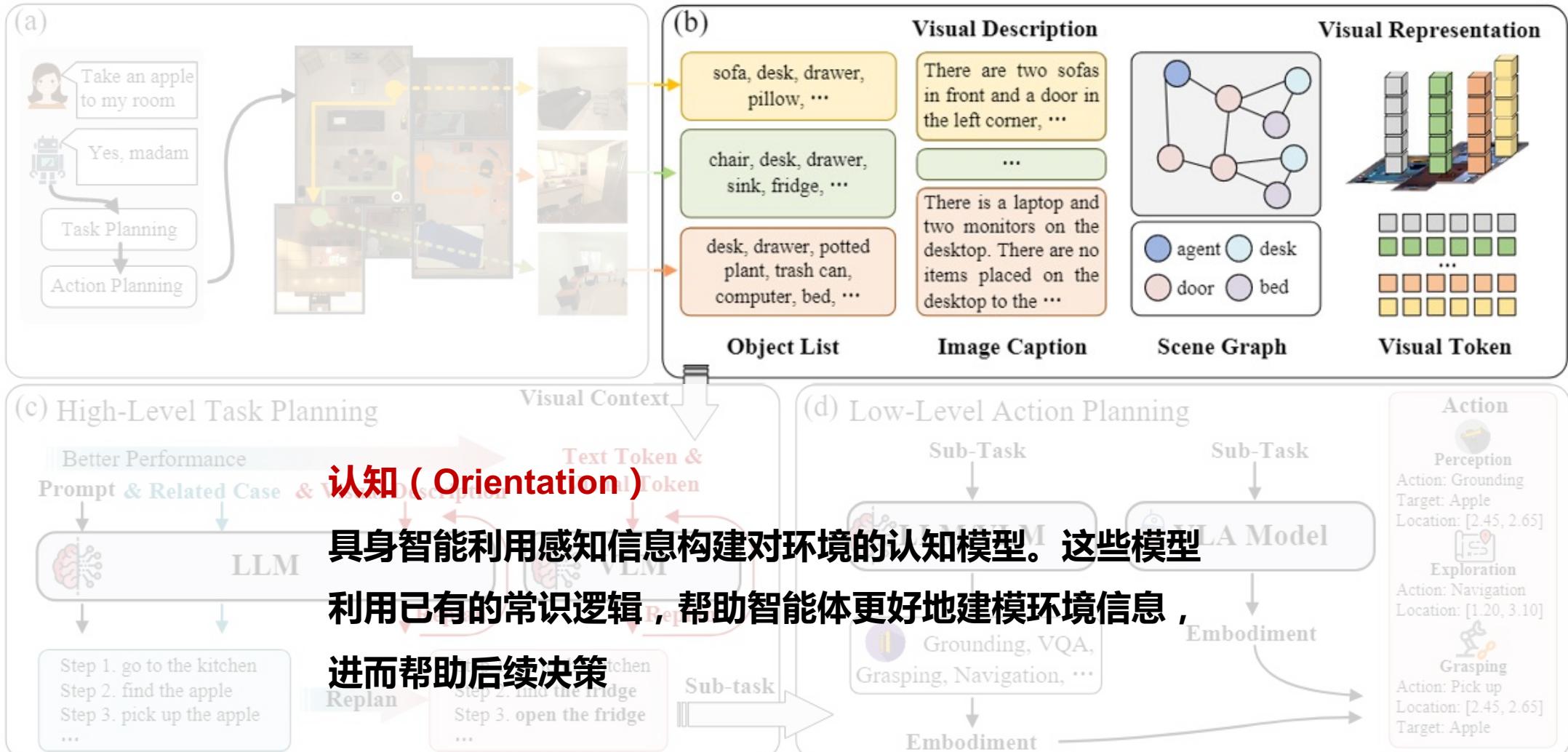
- 具身智能体需要具备**交互(I)、感知(O)、认知(O)、决策(D)、行动(A)**多种能力，以在现实世界中，实现交互式自主探索、持续学习
- **IOODA理论**用于指导智能体在复杂和快速变化的环境中进行决策和行动



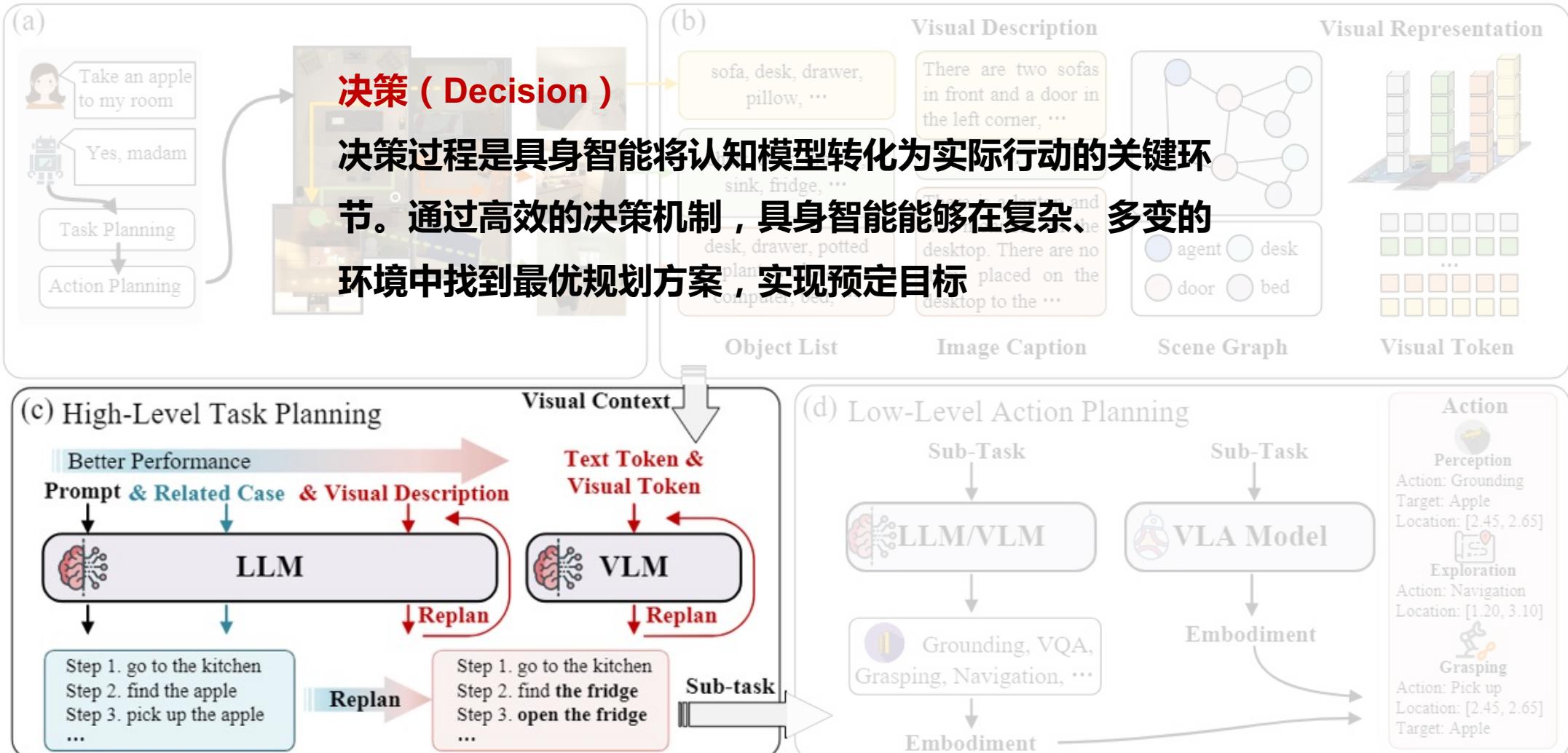
具身智能-智能体-什么是“具身智能体”？



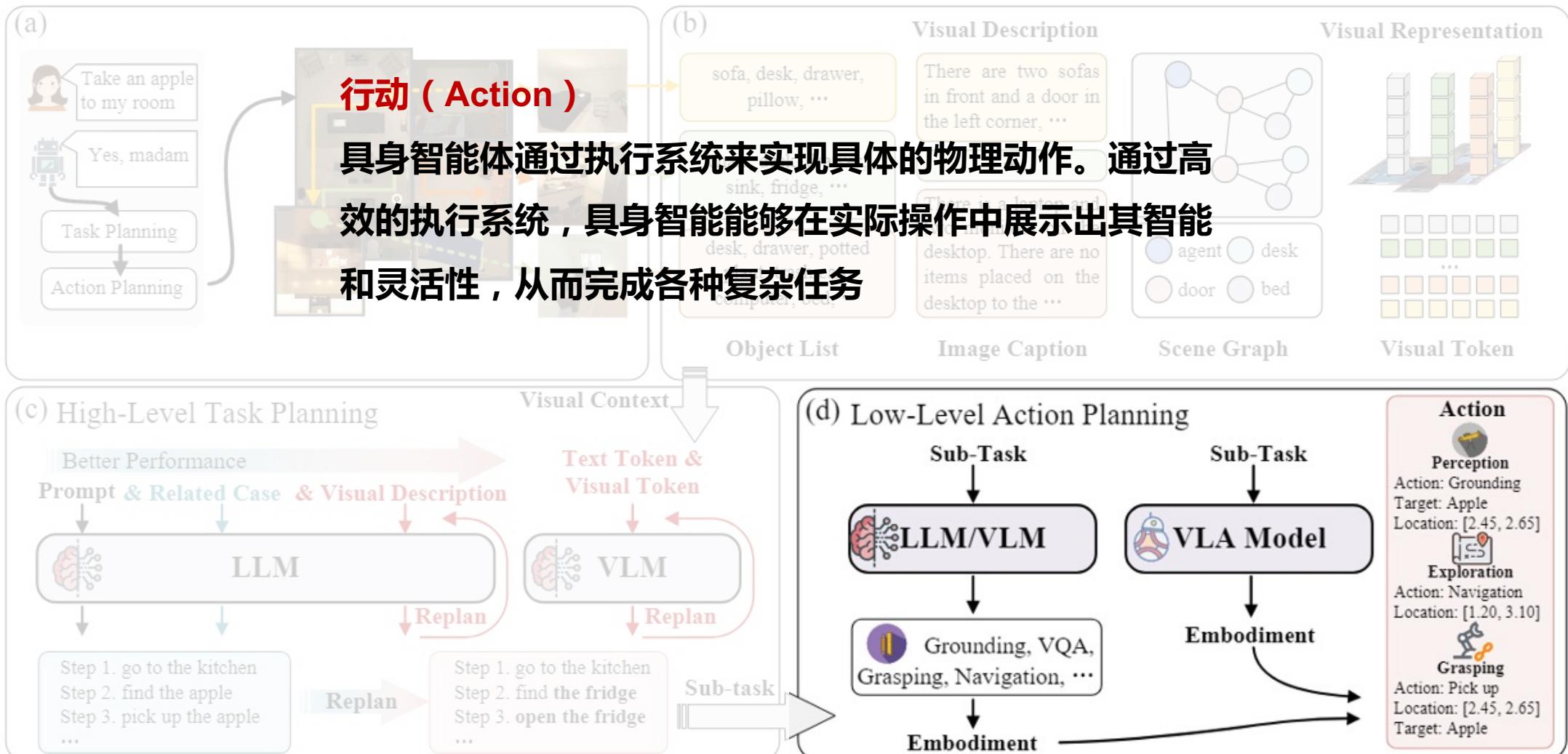
具身智能-智能体-什么是“具身智能体”？



具身智能-智能体-什么是“具身智能体”？



具身智能-智能体-什么是“具身智能体”？



具身智能-智能体-自动驾驶

- 自动驾驶网约车
 - 自动驾驶小巴车
 - 物流快递无人车
 - 智慧矿山无人车
 - 无人清扫车
 -



具身智能-智能体-自动驾驶

LLaDA (NVIDIA)

- 自动驾驶车辆在全球部署面临的主要挑战之一是**适应不同国家和地区的交通规则**
- LLaDA利用大型语言模型解读当地交通规则，帮助自动驾驶车辆（AVs）和人类驾驶员适应新环境



具身智能-智能体-自动驾驶

Tesla RoboTaxi

- 10月11日，马斯克乘坐Tesla Robotaxi 亮相。
- Tesla Robotaxi是一种无需人类司机操作、**完全依赖自动驾驶技术**的出租车服务。通过**先进的传感器、高精地图和人工智能技术**，Tesla Robotaxi能够在城市道路上自主导航，接送乘客，提供安全、高效、便捷的出行体验
- 完全FSD，没有后视镜、油门刹车、方向盘



具身智能-智能体-自动驾驶

自动驾驶网约车 (Robotaxi)

- 在北京、武汉等11个城市开始运营
- 已完成超过500万订单，安全行驶1亿公里
- 出险率只有人类司机的十四分之一
- 7*24小时运营，成本更低，好评率达94.19%



具身智能-智能体-自动驾驶

踏歌智行——矿山无人车

- 国内首个矿山无人驾驶运输项目
- 车-地-云一体化技术架构
- 云控平台调度，电铲、挖机等工程车协同作业
- 实现了安全员下车，安全运行时长**700万小时**
- 未来替代人类进行危险繁重岗位作业



美团——无人配送

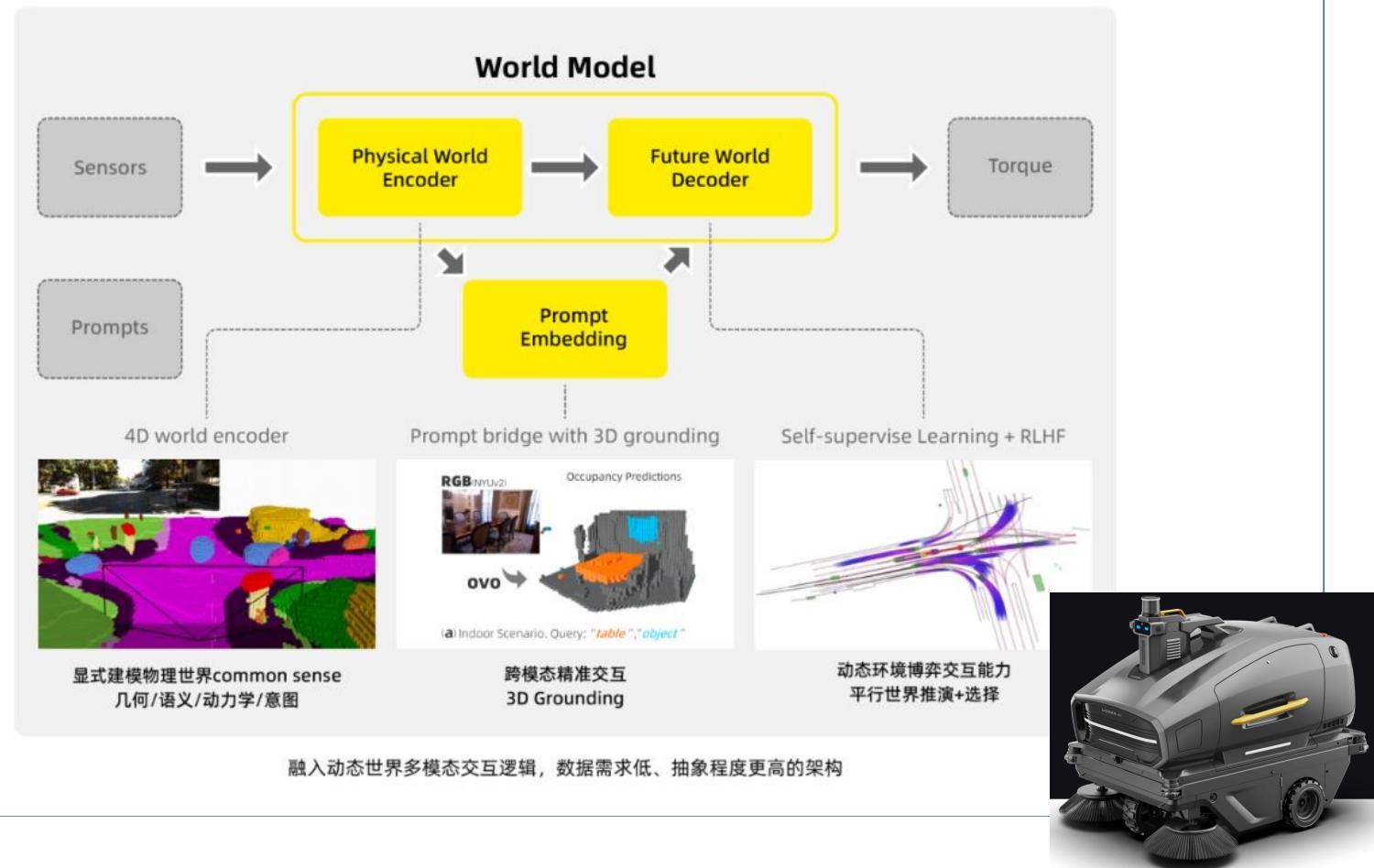
- 新一代无人配送车魔袋20，实现复杂城市路况下自主配送
- 在北京、深圳等城市测试
- 无人机、无人车配送已超过**400万单**



具身智能-智能体-自动驾驶

有鹿扫地车

- **智能清扫**：采用LPLM-10b大模型算法，实时分析地面垃圾量，智能调整清扫策略。
- **全覆盖清扫**：当垃圾量大时，进行全面清扫，确保清洁无遗漏。
- **节能清扫**：垃圾量少时，切换至低能耗清扫模式，节省电量和耗材。
- **接收指令**：允许自然语言指令，例如“去 1 号楼清扫一下落叶”



具身智能-智能体-无人机

西工大团队：无人机自主“聊天群”

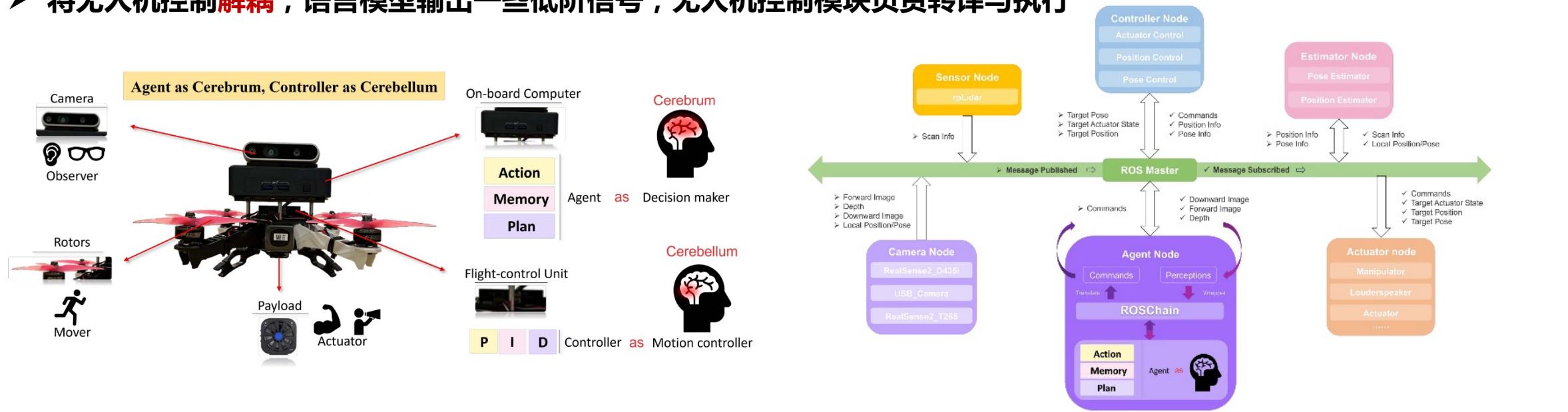
- 基于国产大模型，研发了“群聊式”无人机控制框架，实现了开放环境下“人机”和“多机”的对话交互



具身智能-智能体-无人机

北航航空学院团队：基于多模态大模型的具身智能体无人机控制

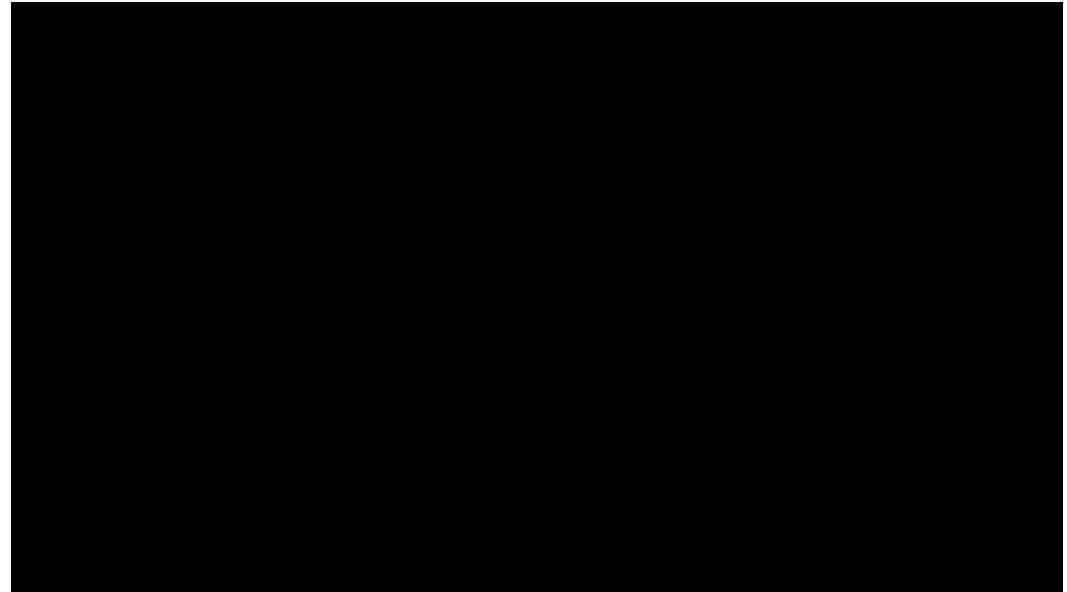
- 将大型多模态模型（LMMs）集成到无人机自主代理框架中的新方法，专注于搜索和救援等工业应用
- 将无人机控制解耦，语言模型输出一些低阶信号，无人机控制模块负责转译与执行



具身智能-智能体-机械臂与人形机器人

Google PaLM-E

- PaLM-E结合了**大型语言模型与机器人实体操作**，实现了对复杂任务的理解与执行
- PaLM-E的输入包括文本和连续的观测。与这些观察结果相对应的多模态表征与文本交织，形成**多模态句子**。PaLM-E的输出是由模型自动回归生成的文本，可以是一个问题的答案，也可以文本形式的机器人执行的决策
- 通过视觉和状态估计等多模态信息，PaLM-E能在外部干扰的情况下，有效规划并完成如“从抽屉取米”等精细任务



具身智能-智能体-机械臂与人形机器人

Google DeepMind: RT 系列多任务模型

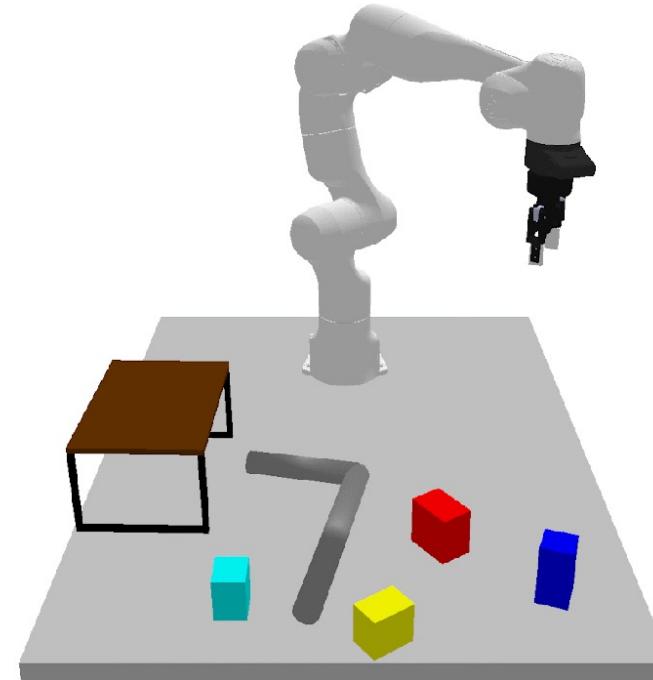
- RT-1 机器人在 700 条指令中达到 **97% 的成功率**
- RT-1 显著地改进了对新任务、环境和对象的零样本泛化，对未见过的指令达到 **76% 的成功率**
- RT-2 受益于互联网级别的训练数据，在 208 个评估任务中成功完成 **60%**，成功率是 RT-1 的 **3 倍以上**



具身智能-智能体-机械臂与人形机器人

Text2Motion (斯坦福大学)

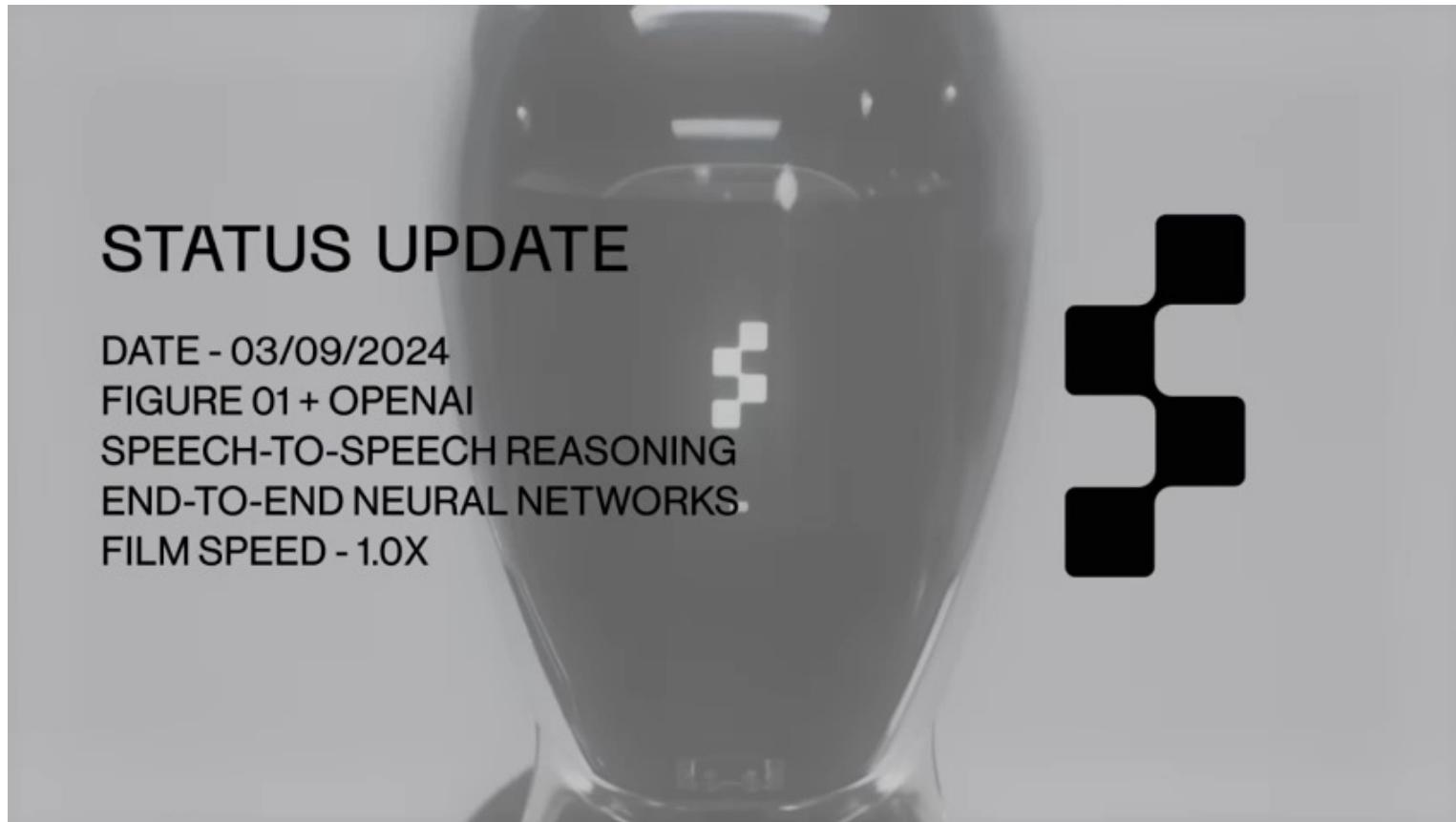
- Text2Motion 框架通过自然语言指令实现机器人的长期推理与顺序操作任务，它结合**大型语言模型、技能库及几何可行性计划器**，确保任务规划**既符合指令又动力学可行**
- 例如，在接收到“将两个原色物体放在架子上”的指令后，Text2Motion能有效规划机械臂先抓取红色物体放置于架子上，再利用钩子辅助抓取蓝色物体完成任务，整体成功率高达82%



具身智能-智能体-机械臂与人形机器人

Figure AI & OpenAI: 人形机器人 Figure 01

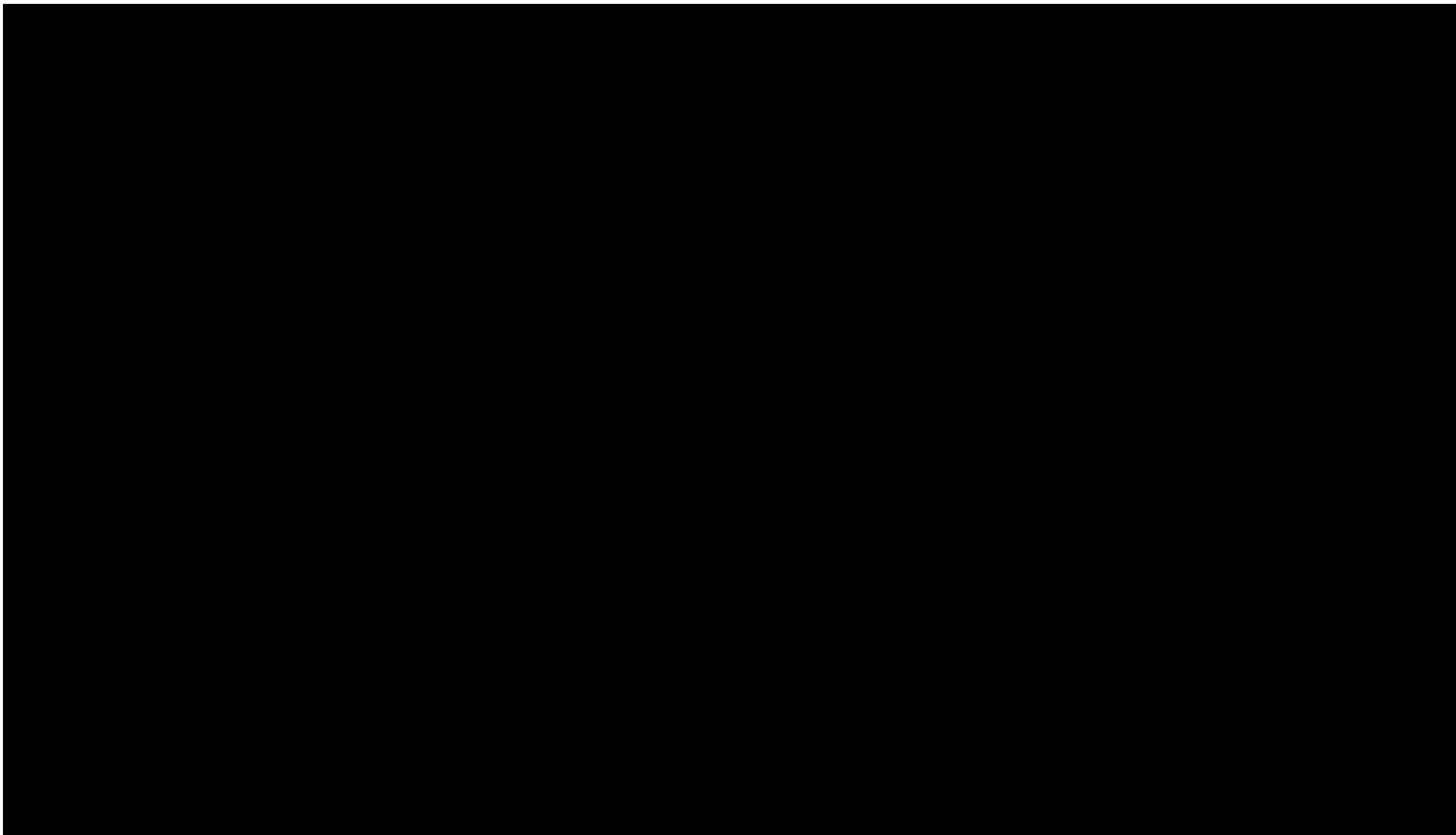
- Figure 01是 Figure 和 OpenAI 公司合作研发的第一个机器人产品，它能够实现自主行为，与人类进行对话，分类识别物品等功能。是世界首个**商用化的自主类人机器人**
- 截至 2024 年 2 月，Figure AI 共获得 **6.75 亿美元高额融资**，估值达到 **26 亿美元**
- 投资人包括**英伟达、英特尔等行业巨头**



具身智能-智能体-机械臂与人形机器人

Figure AI & OpenAI: 人形机器人 Figure 02

- Figure 02 较 Figure 01 在外观设计、结构、手部灵活性和视觉AI系统上都有显著改进。简而言之，Figure 02 更加智能、灵活且适应性更强
- Figure 02 已经在宝马的工厂进行了训练和数据收集。随着劳动力短缺问题在制造业、仓储业和医疗保健等多个市场日益严重，像 Figure 02 这样的人形机器人可能会成为解决这一问题的重要方案之一



具身智能-智能体-机械臂与人形机器人

8月21日，2024世界机器人大会在北京举行，27款人形机器人集中亮相



星尘智能新一代AI机器人
助理Astribot S1展示书法



Walker S系列人形机器人
演示分拣的过程



加速进化T1机器人能做到
全向行走和踢球

具身智能-智能体-机械臂与人形机器人

银河通用机器人：GALBOT

- GALBOT 是银河通用机器人公司的**多模态**通用**具身大模型**人形机器人，其采用移动双臂和轮式腿设计，可实现360°移动。
- 支持场景：
 - 制造业：对部件进行分类和包装，助装配生产线
 - 家庭：日常清洁、整理、寻找和取物
 - 药房：辅助药剂师进行药物分发和库存维护
 - 零售门店：全天候进行库存管理，提高运营效率



具身智能-智能体-机械臂与人形机器人

智元公司: 远征 A1

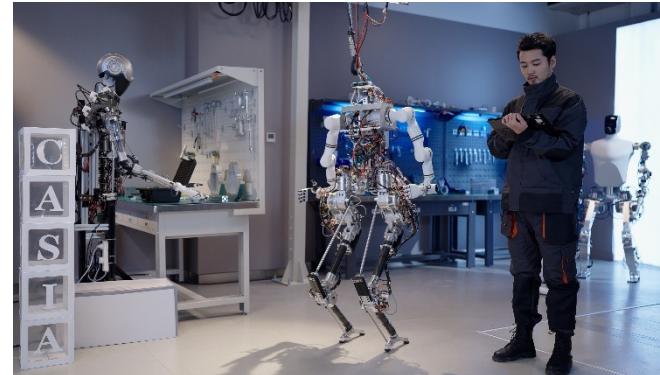
- 华为天才少年稚晖君创业智元机器人，并在 2023 年发布第一款能自主完成**复杂任务的通用具身智能机器人产品——智元机器人远征 A1**
- 支持场景：
 - 样本制备：帮助研究员样本制备、样本增扩
 - 外观检测：从事地盘装配、外观检测
 - 智能管家：支持做菜、给药、辅导功课



具身智能-智能体-机械臂与人形机器人

乔红：谱系化人型机器人

- 乔红院士提出“环境吸引域”：
 - 《基于环境约束和多空间分析的机器人操作理论研究》获得了国家自然科学奖二等奖
 - 在国际上受到高度赞誉，被誉为“乔的概念”
 - 通过构建高维环境吸引域，实现 0.0025 mm 的操作精度，已经应用到奇瑞汽车、秦川机床厂等公司中
- 她带领团队发布了谱系化人形机器人Q家族
 - 可根据应用场景快速设计构建人形机器人硬件和软件系统



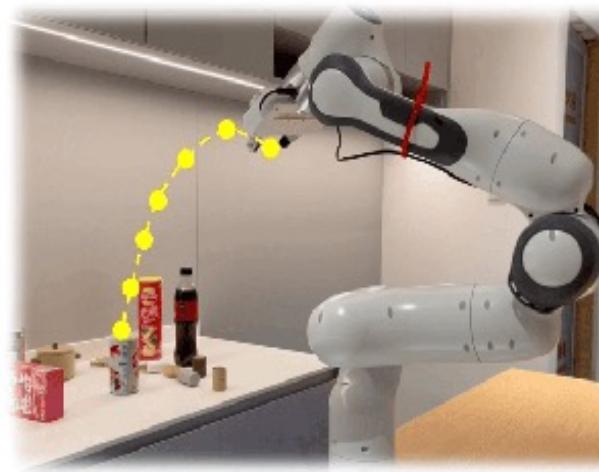
具身智能-智能体-异构多智能体

西工大团队：异构智能体自主协作

- 基于国产大模型，研发了**多智能体异构**控制框架，实现多个异构智能体协作的任务场景



无人机集群路径规划



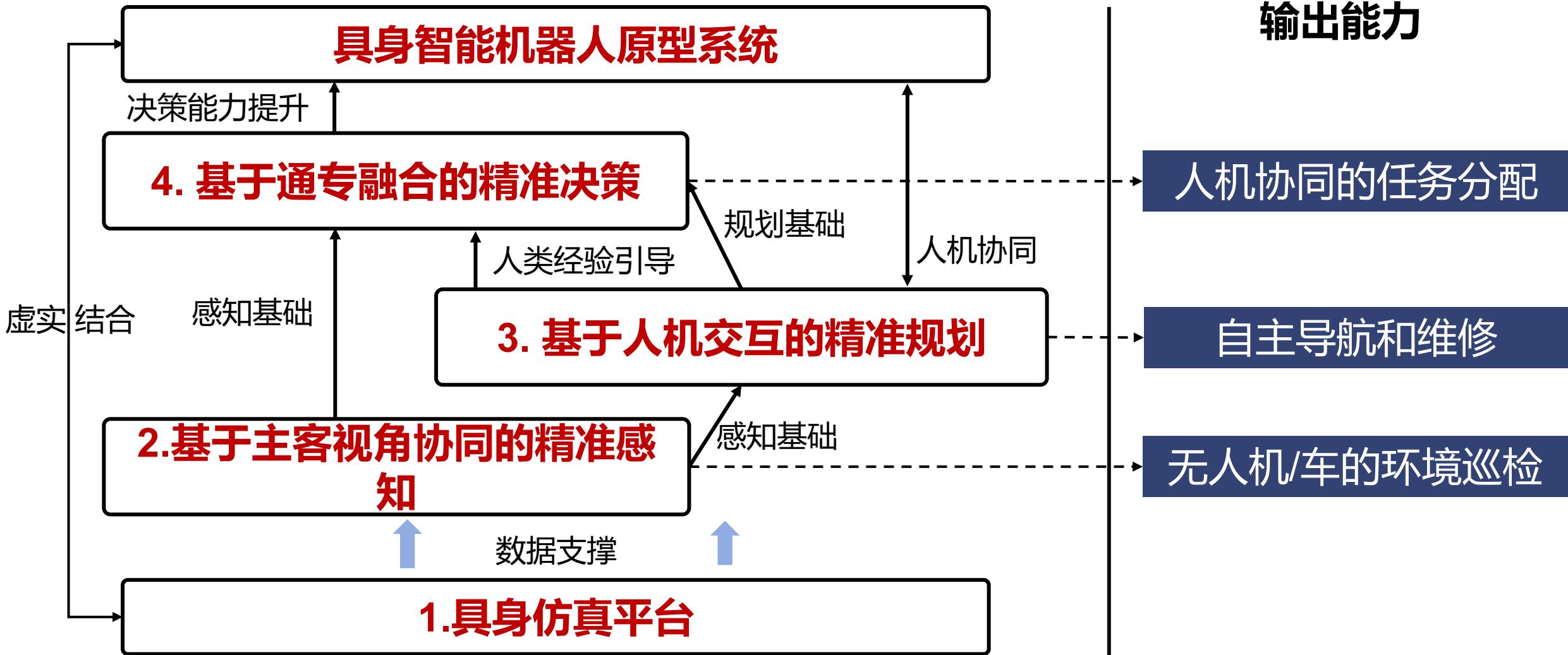
机械臂关节运动规划



机器狗操作运动规划

具身智能-智能体-具身大模型（组内工作）

面向新型工业化的精准具身智能



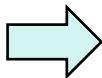
1. 具身仿真平台



构建物理高保真具身仿真平台，虚实结合支撑具身模型

基于虚拟环境到现实环境的模型迁移

虚拟环境：原型组构建



现实环境：在线学习更新



- 基于原型学习思想，充分挖掘仿真数据的易得优势
- 利用在线学习技术，持续提高真实场景的泛化能力

仿真验证

基于真实环境到虚拟环境的数据仿真

数据孪生

真实数据采集



复杂场景建模

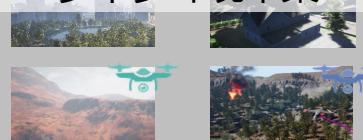


- 基于NeRF/3DGS的实景重建，真实感强、光照效果还原度高
- 支持Unity、Unreal，方便后续开发

数据支持

物理高保真的具身仿真引擎“哪吒”

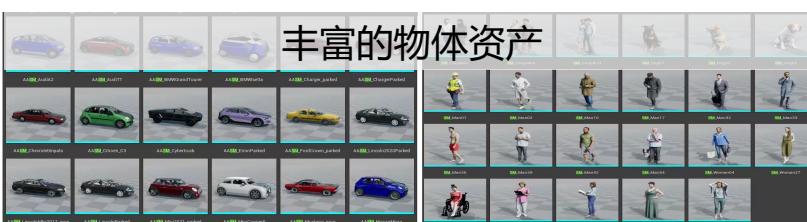
多卡多环境采集



控制和传感器信息



丰富的物体资产



无人机平台展示

1. 无人机、无人车等其他无人系统的仿真和实验搭建
2. 多传感器模拟，动力学模型仿真，具有多种支持具身智能开发的特性
3. 集成了丰富的**仿真场景**以及**物体资产**，具备无人机控制等**丰富的API**

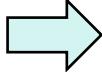
1. 具身仿真平台



构建物理高保真具身仿真平台，虚实结合支撑具身模型

基于虚拟环境到现实环境的模型迁移

虚拟环境：原型组构建



现实环境：在线学习更新



- 基于原型学习思想，充分挖掘仿真数据的易得优势
- 利用在线学习技术，持续提高真实场景的泛化能力

基于真实环境到虚拟环境的数据仿真

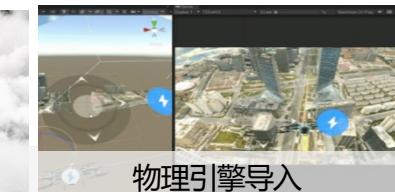
数据孪生



真实数据采集



复杂场景建模



物理引擎导入

- 基于NeRF/3DGS的实景重建，真实感强、光照效果还原度高
- 支持 Unity、Unreal，方便后续开发

仿真验证

数据

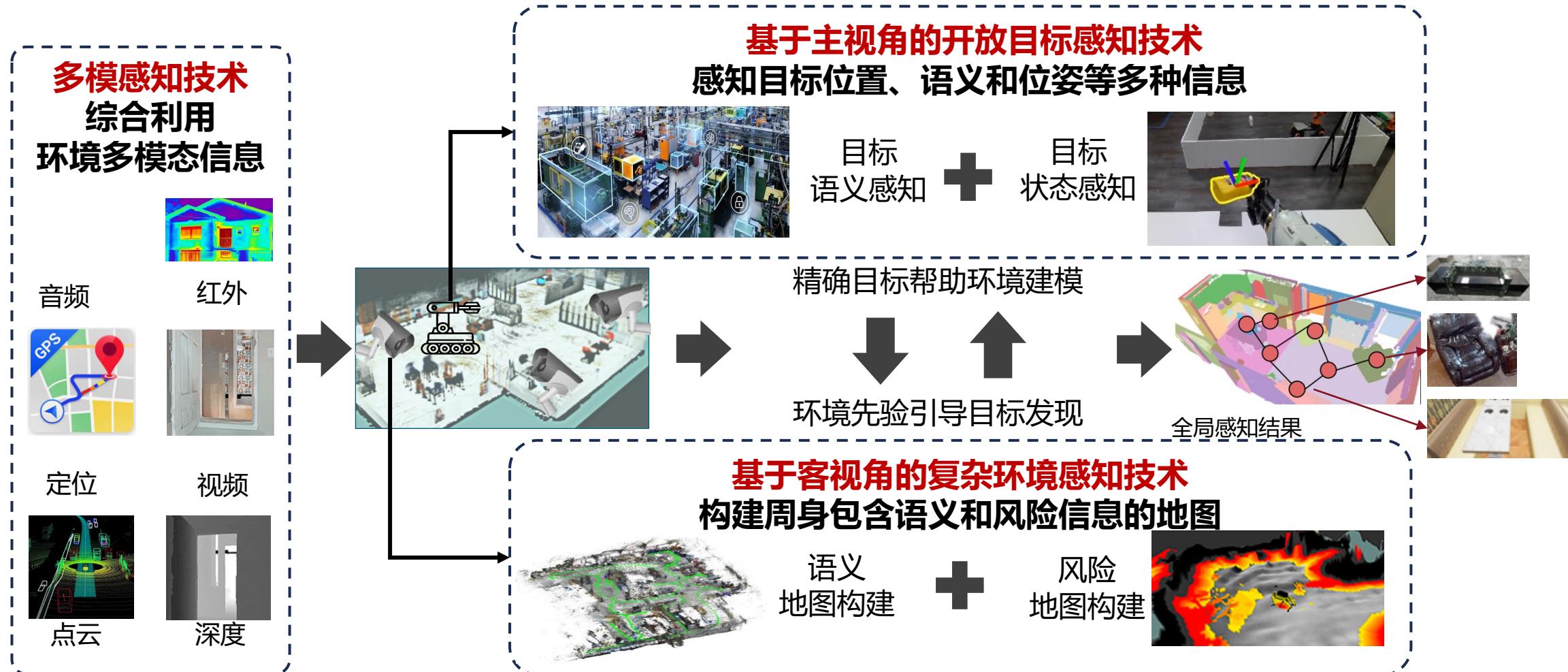
“哪吒”仿真引擎的优势

平台	地图	物体资产	智能体	支持的传感器配置	可拓展API	数据采集记录	定制控制器
哪吒	22	96(持续添加)	Drone, Vehicle	rgb, depth, lidar, radar, gps, imu	√	√	√
AirSim	2	--	Drone, Vehicle	rgb, depth, lidar, radar, gps, imu	√	✗	✗
Carla	15	143	Vehicle	rgb, depth, lidar, radar, gps, imu	√	√	✗
AirVLN	17	--	Drone	rgb, depth, gps	✗	✗	✗

2. 基于主客视角协同的精准感知



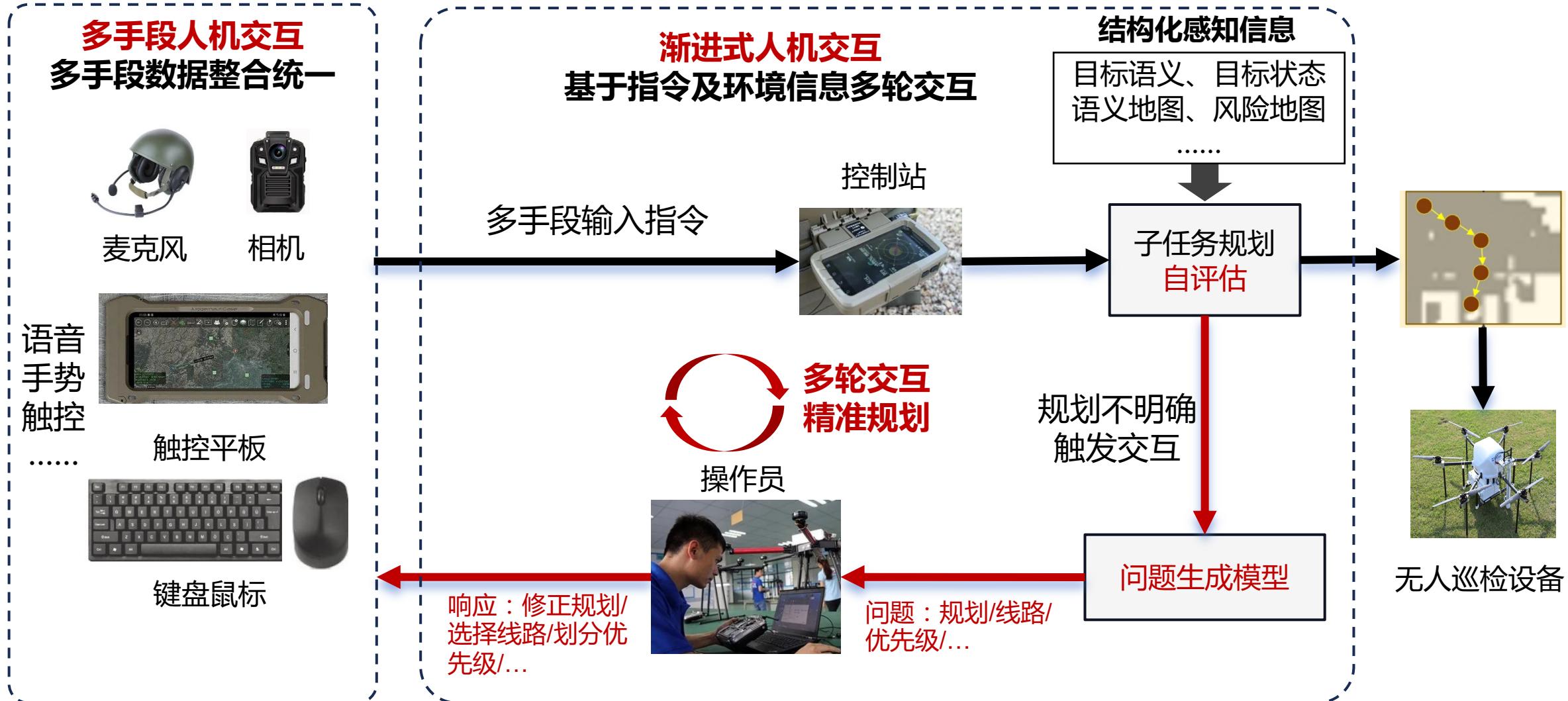
利用多模态信息，协同第一(主)及第三(客)人称视角信息，实现精准感知



3. 基于人机交互的精准规划



基于多手段渐进式交互，实现精准规划



4. 基于通专融合的精准决策

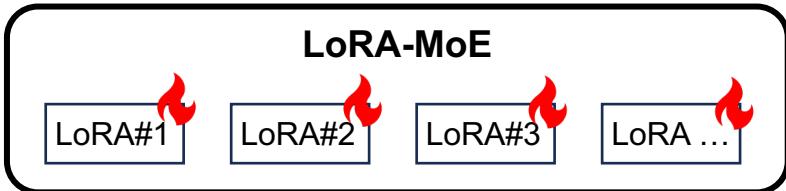


基于通专融合的高级任务规划，实现大小模型协同的精准决策

基于大模型混合微调的通用任务规划



+



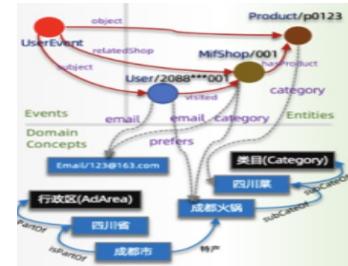
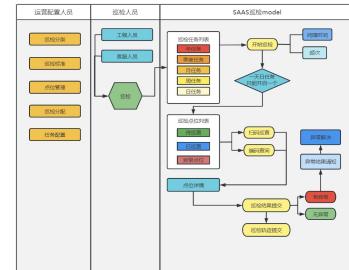
任务序列

- 步骤1：找到并拿起门口长方形桌子上的螺丝刀
- 步骤2：移动到检测出异常松动的设备处
- 步骤3：拧紧螺丝
- 步骤4：继续巡检，前往区域#2

约束

指导

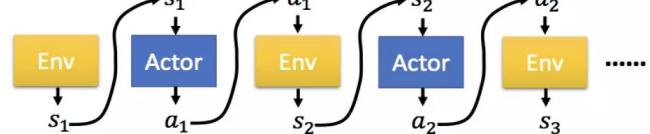
专家知识库



巡检流程

领域知识

专用运动决策小模型

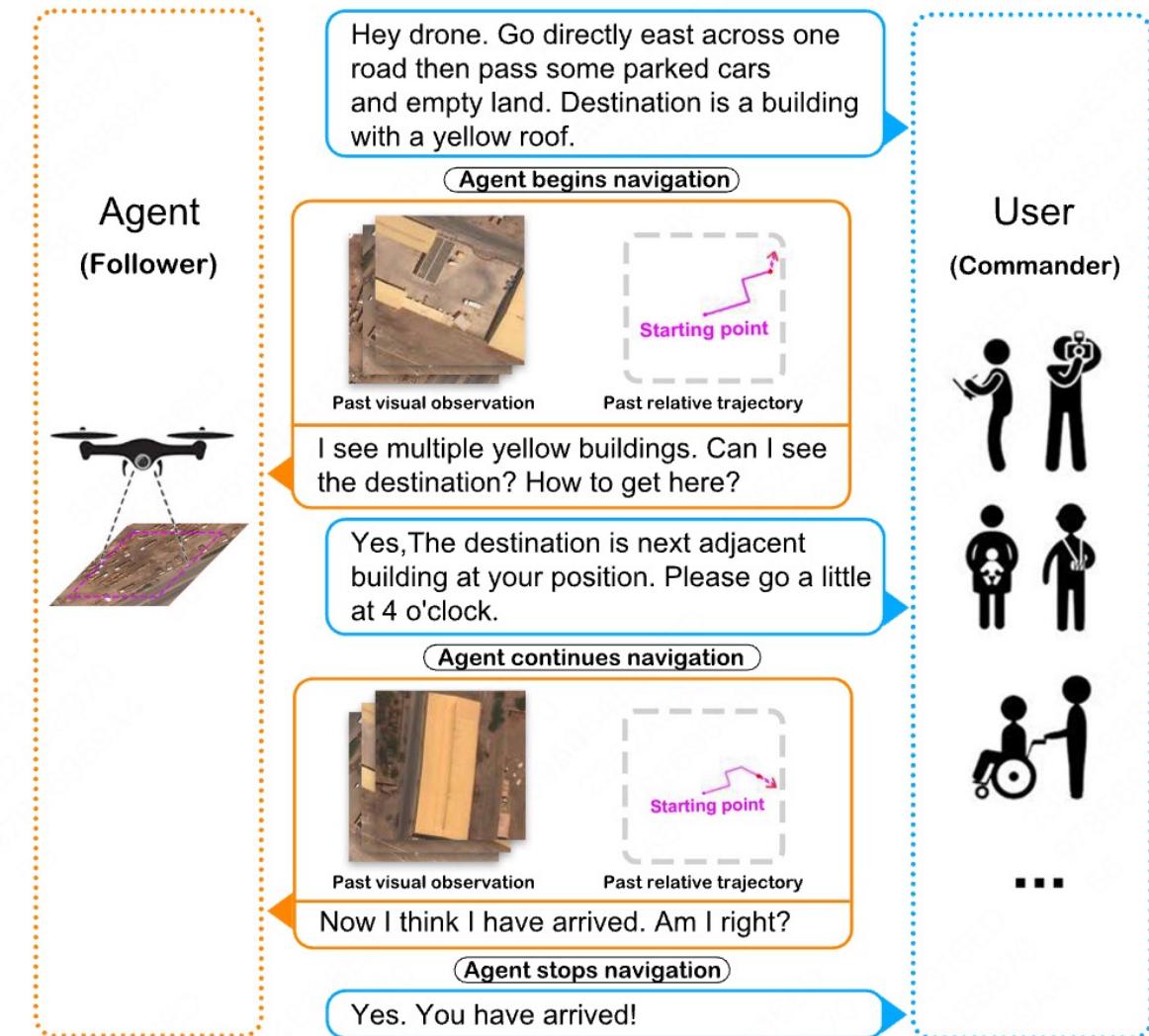


运动决策算法



具身智能-智能体-基于对话增强的无人机视觉语言导航（组内工作）

- **无人机对话导航任务**：用户向无人机提供语言指令以指示导航目标。当目标不明确时，无人机会主动提问，并基于用户的回答进行导航。
- **当前的无人机对话导航基准通常依赖于离线对话历史：**
 - 无人机缺乏根据环境**主动提问**的能力。
 - 对话历史中的内容并不总是与无人机当前的观察相一致。



具身智能-智能体-基于对话增强的无人机视觉语言导航（组内工作）

聚焦在线对话的无人机对话导航

- AVDN-OL : 一种新的基于大语言模型的**在线对话导航评测基准**
- AerialVLA : 构造了具有在线问答能力的无人机视觉-语言-动作大模型



follow the street, stop at the roof top
with blue at your eleven o'clock



I am at the cross of the road. Am I near
the target? ①



No, You are not near. Go down the
street at your one o'clock



I am on a white roof top. Am I at the
target position? ②

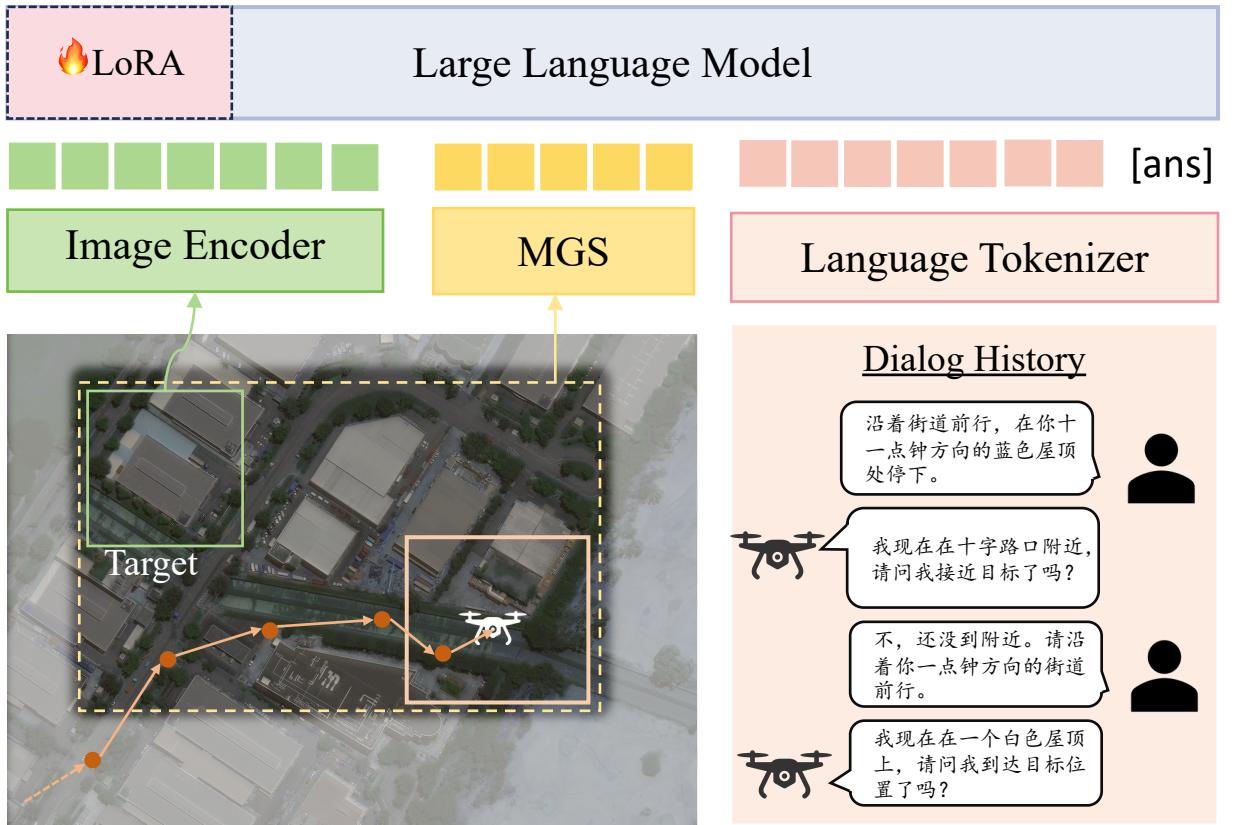


No, you should go to your left.



具身智能-智能体-基于对话增强的无人机视觉语言导航（组内工作）

聚焦在线对话的无人机对话导航

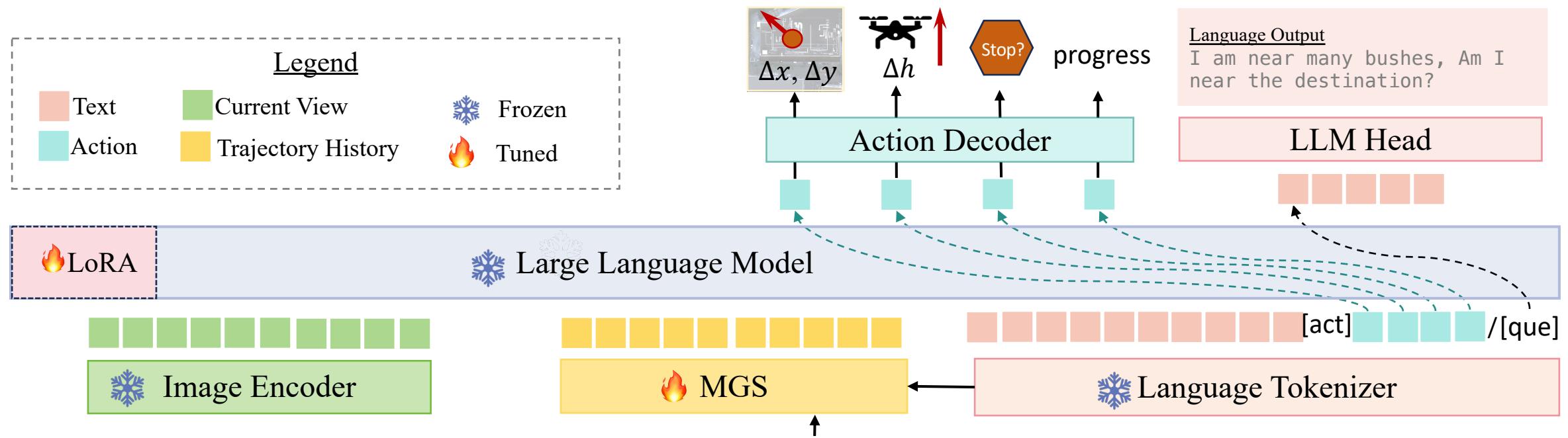


- **无人机指挥大语言模型** (Aerial Commander Large Language Model, 简称 AC-LLM)
- **无人机对话导航数据集和遥感问答数据训练**
- 根据无人机Agent的导航历史、当前状态、问题内容和目标位置生成导航指导，从而替代了对人工对话响应的需求。
- **新评估指标**：设计了基于对话触发频率和导航成功率的新评估指标，提供了对无人机Agent的视觉对话导航(VDN)能力的多维度分析。

具身智能-智能体-基于对话增强的无人机视觉语言导航（组内工作）

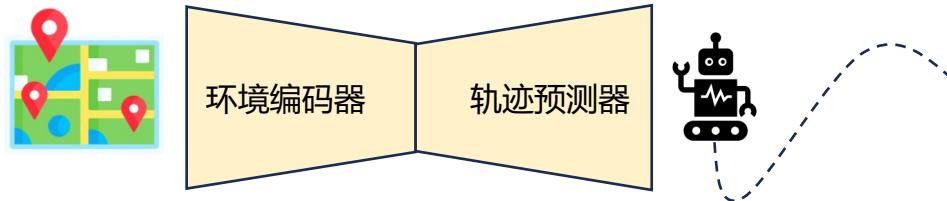
聚焦在线对话的无人机对话导航

- **基于超链接的语言-动作切换机制**：为使模型兼容导航决策和问题生成过程，将动作生成特征采用`<[act]+特殊token>`方式输入进行回归，判断是否需要提问，以及评估是否到达目的地。
- 判断需要提问后使用`[que]`引导模型提出问题。



具身智能-智能体-基于大小模型协同的无人设备控制（组内工作）

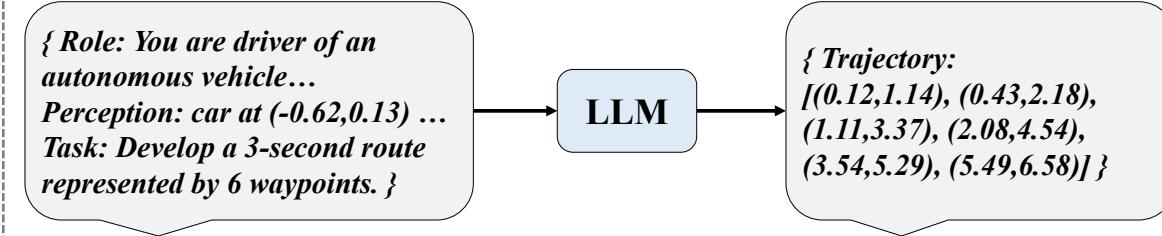
基于学习的小模型规划范式（主流）



基于学习的小模型规划算法，使用大量数据训练，直接基于简单视觉信息进行预测：

- 缺乏泛化性
- 缺乏交互性
- 缺乏可解释性

基于LLM的规划范式（主流）



基于大语言模型的规划算法，受限于语言模态：

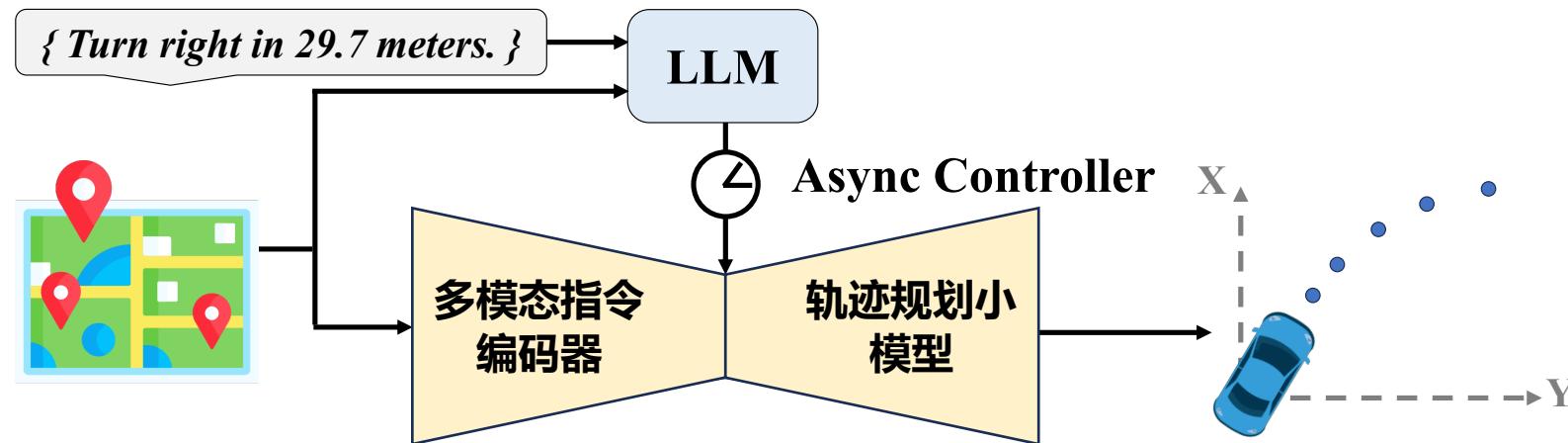
- 以语言做输入：难以对场景详尽描述
 - 以语言做输出：对浮点数回归能力不足
- 以LLM作为决策主体：
- 推理速度慢

具身智能-智能体-基于异步规划的无人车导航算法（组内工作）

AsyncDriver：基于异步规划的无人车导航算法

希望既能够利用LLM的优势**保留人机交互接口与交通常识嵌入的能力**；又能在保持精度的基础上实现实时的**推理速度**：

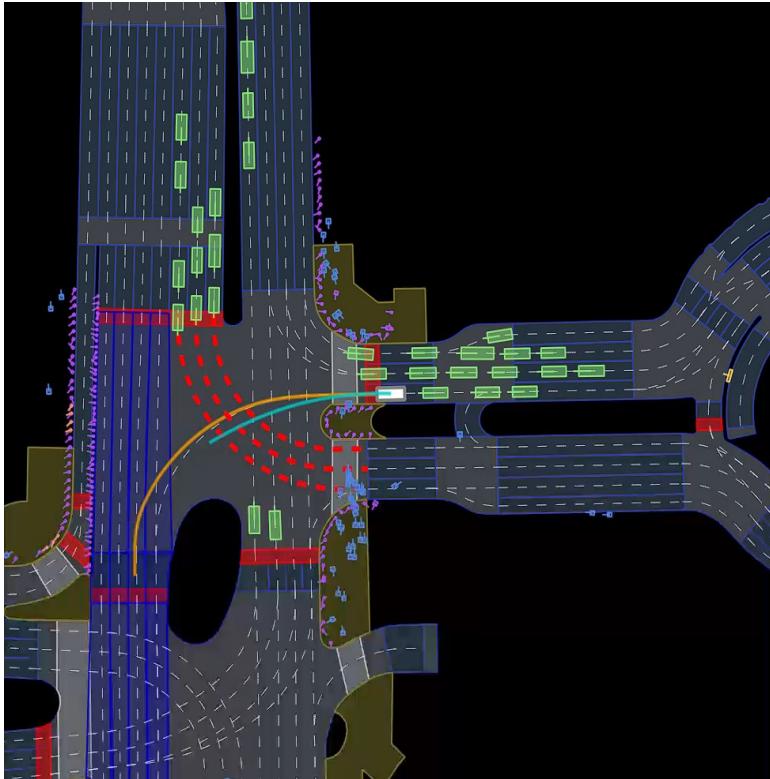
- 以LLM输出的场景相关-高阶指令特征作为引导，辅助小模型**实现指令可交互的规划预测**
- 通过将大小模型的推理过程解耦，控制大模型的推理频次，**实现推理速度与性能间的平衡**



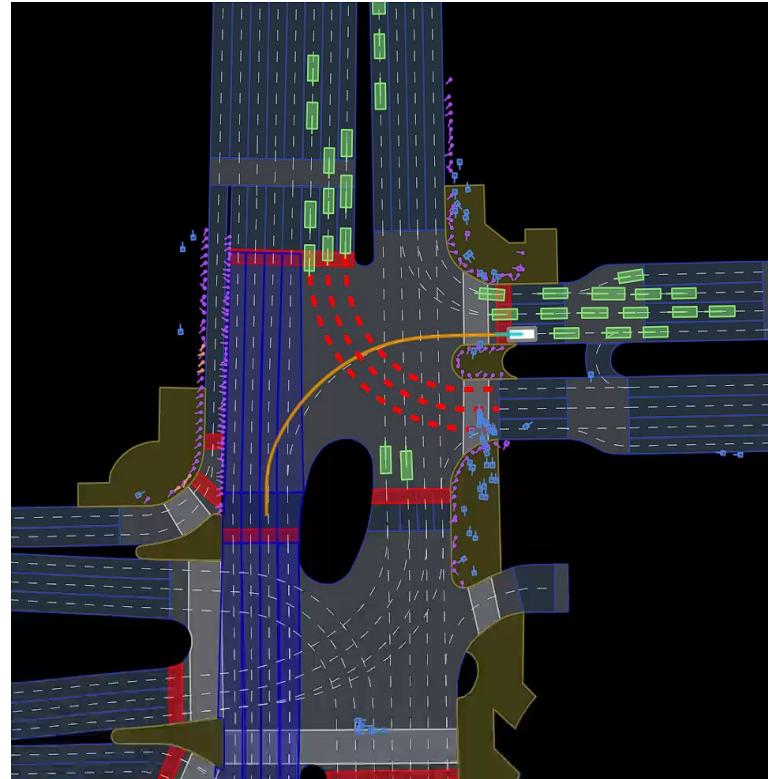
具身智能-智能体-基于异步规划的无人车导航算法（组内工作）

➤ 路口场景：我们的模型(右)成功避免了路口的车辆碰撞

Learning-Based 模型



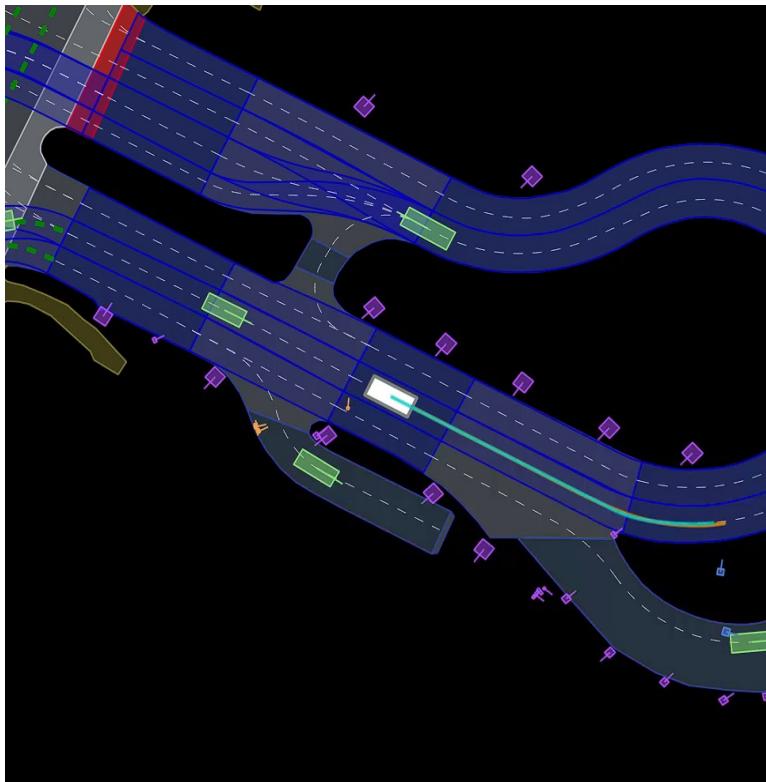
Ours



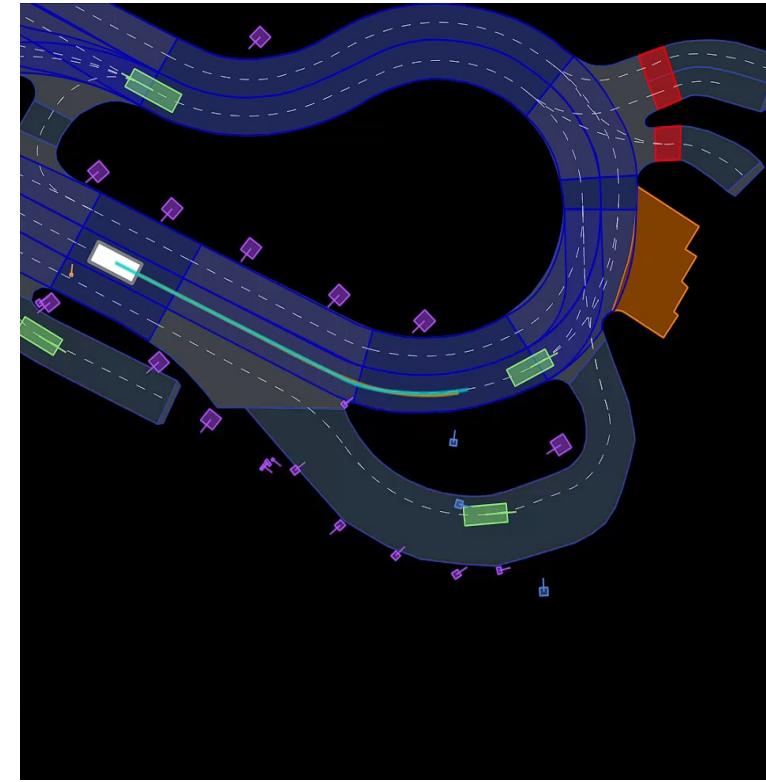
具身智能-智能体-基于异步规划的无人车导航算法（组内工作）

➤ 并道场景：我们的模型(右)成功避免了岔路口的信任和车辆碰撞

Learning-Based 模型



Ours



具身智能-智能体-基于多粒度轨迹规划的无人机视觉语言导航算法

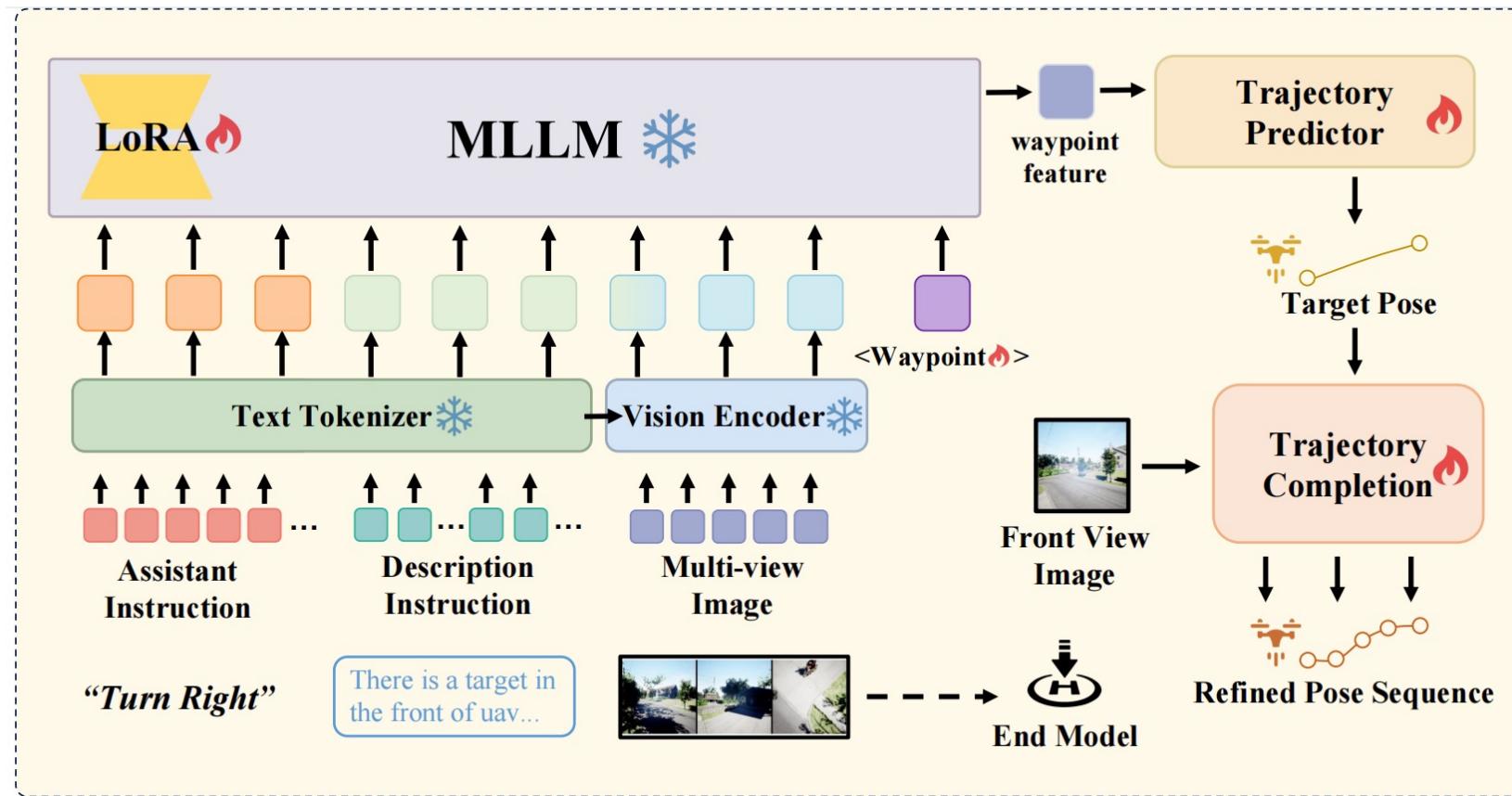
- 多模态大模型具有更强的**多模态信息理解能力**，适用于高层次信息指导
- 小模型**推理速度快**，适合加快推理速率，进行**细粒度的轨迹生成**



语言指令：目标位于无人机的前方。中心物体是一个身穿白色短袖和蓝色短裤的人。周围沿街高楼鳞次栉比，街道上设有车道和绿色自行车道，建筑物附近摆放着撑着彩色遮阳伞的街头小贩。

具身智能-智能体-基于多粒度轨迹规划的无人机视觉语言导航算法

算法框架



多模态大模型

输入物体描述指令、
多视图图片以及助
手的辅助信息，使
用可学习的查询提
取特征，并输出一
个远距离关键点

轨迹补全小模型

输入大模型给出的
关键点以及无人机
当前时刻前视图，
生成细粒度的轨迹
序列完成避障飞行

04

核心要素：数据

具身智能-数据

具身智能是实现通用人工智能的关键方向，受到美国等发达国家及OpenAI、英伟达等科技巨头的重视与投资。与此同时，中国也涌现出一批前沿科研机构和企业，积极参与具身智能的研究与应用。近年来，机器人技术迅速发展，广泛应用于工业、家庭、深海及太空等领域。作为具身智能的物理载体，机器人通过与环境的交互不断学习和提升。在大模型时代，增加数据量和模型规模可以提升模型性能。因此，构建一个高质量的开源机器人感知操作数据集是迫切需要的，类似于ImageNet对计算机视觉研究的推动。



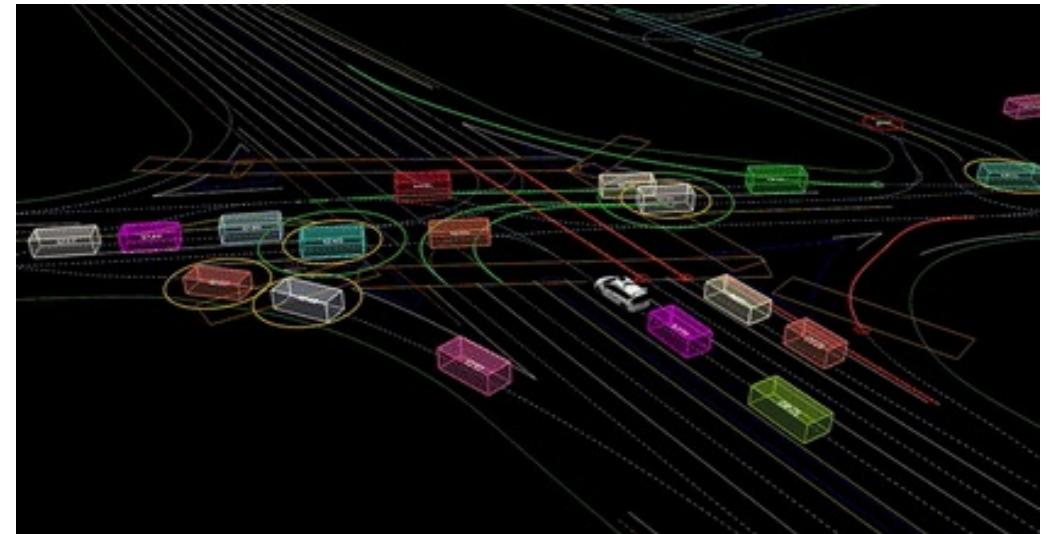
Image Classification on ImageNet



具身智能-数据-无人车

Waymo Open Dataset (Google子公司Waymo)

- 包含大量真实世界的驾驶场景，总数据量超过**10 TB**，涵盖了城市、乡村、晴天、雨天等多种驾驶环境。
- Waymo Open Dataset 提供了对**行人、车辆、自行车等丰富物体的3D标注**，每个物体的大小、位置、运动轨迹等均得到详细标注。
- Waymo 数据集支持多种自动驾驶研究任务，包括**物体检测、3D物体跟踪、运动预测、路径规划等任务**。



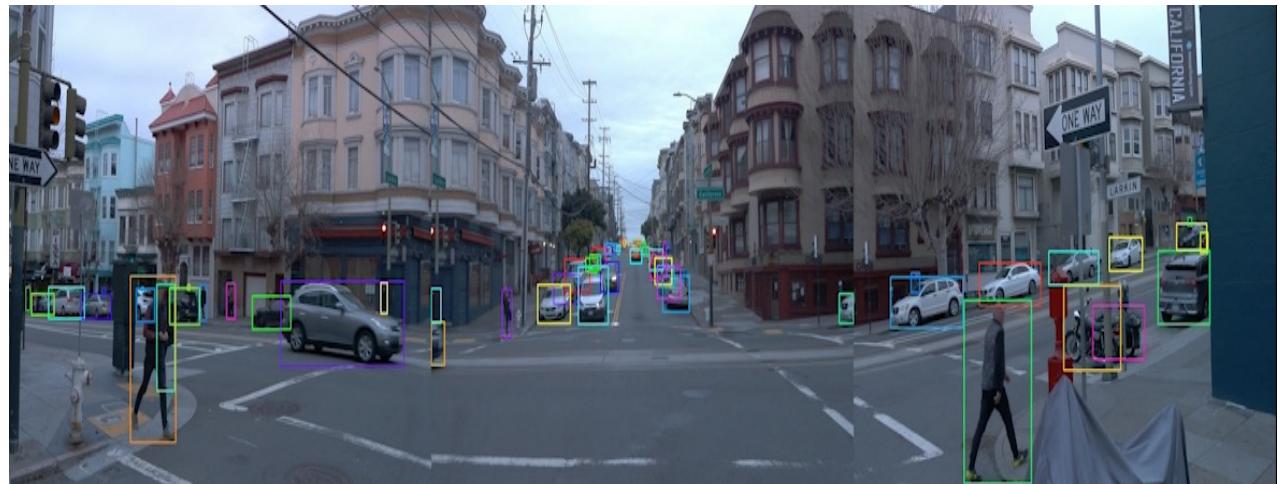
具身智能-数据-无人车

Waymo Open Dataset (Google子公司Waymo)

- 采集自各种不同的地点



- 包含多种不同的环境、物体和天气条件



具身智能-数据-无人车

Waymo Open Dataset (Google子公司Waymo)

- 包含市区、郊区、白天、晚上、行人、骑行、建筑、复杂天气等多种场景。



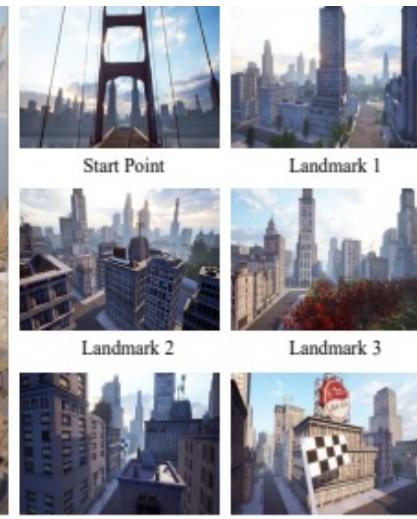
具身智能-数据-无人机

AerialVLN

- AerialVLN基于Unreal Engine 4和Microsoft AirSim插件实现，支持连续导航和近乎真实的渲染效果。
- 收集了25个市级环境，覆盖市中心、工厂、公园、村庄等多种场景，包含870多种不同的物体。
- 数据集包含由AOPA认证的无人机飞行员采集的8,446条飞行路径，每条路径配有3条由AMT标注员编写的指令。
- 每条指令平均包含83个词，涉及4,470个不同的词汇。

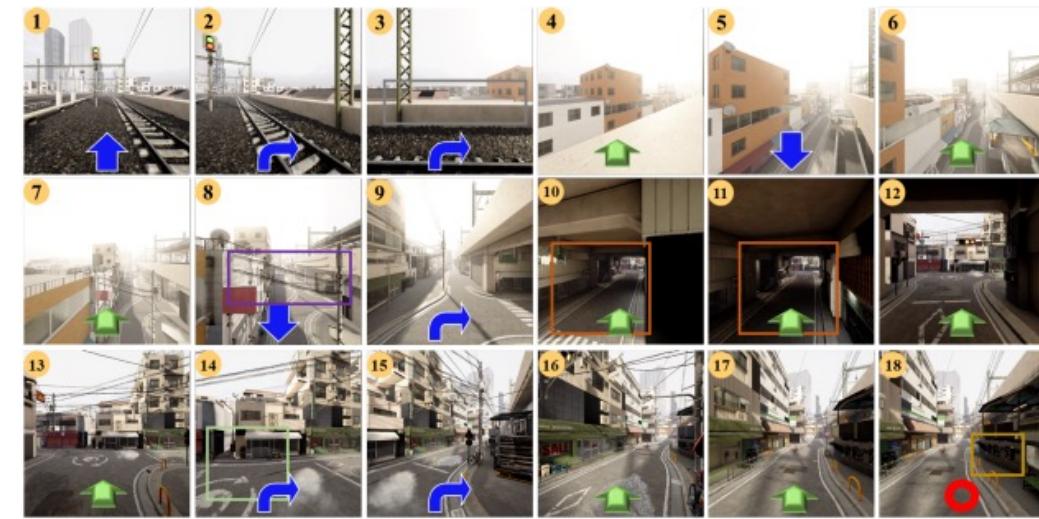


Route Overview



Landmark 4

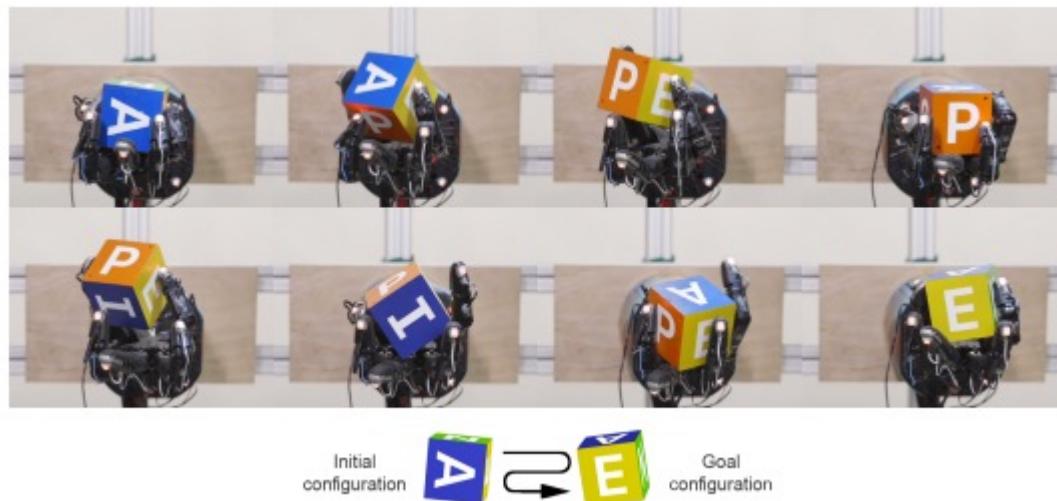
End Point



具身智能-数据-机械臂

Dactyl (OpenAI)

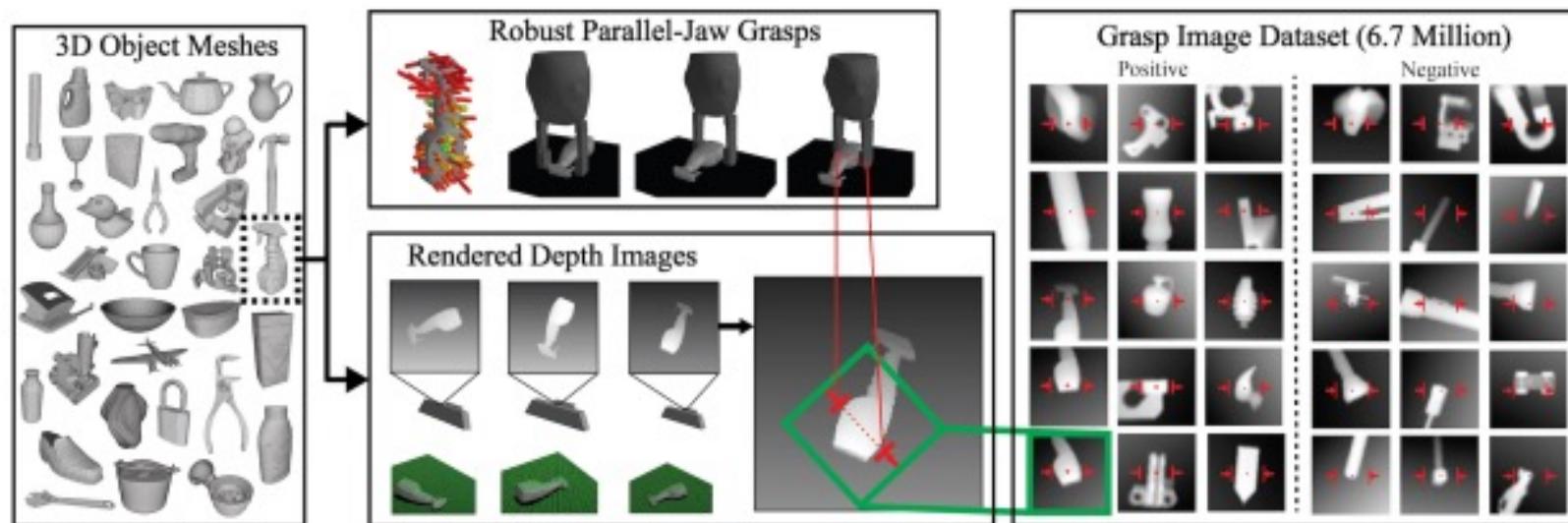
- Dactyl 数据集包含了机械手臂在**多种物体抓取和操作任务中的人类演示数据**，旨在支持机器人学习和操控研究。
- 数据集包括多种物体的抓取和操控示范，涉及**不同形状、大小和材质**的物品，为研究人员提供丰富的操控场景。
- 支持强化学习和模仿学习，包含机械手的**操作轨迹、传感器数据（如力传感器和视觉信息）等多模态数据流**。



具身智能-数据-机械臂

Dex-Net (UC Berkeley)

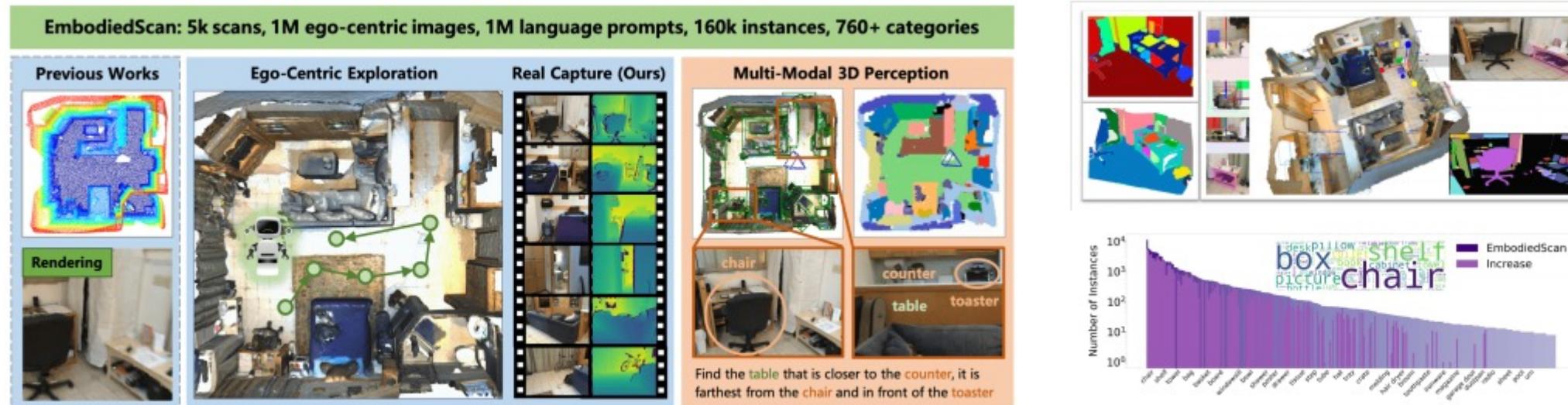
- 包含**1,500个3D物体网格模型**，涵盖了多种形状和尺寸的日常物品，这些物品通常在抓取和操作中会遇到。
- 包含超过**670万张抓取图像**，提供了丰富的训练样本，以提高机械臂的抓取能力和可靠性。
- 数据集中**涵盖了不同形状和表面的物体**，使得模型在处理各种现实情况时更加灵活和适应。



具身智能-数据-机器人

EmbodiedScan (上海AI Lab)

- 从现有的3D室内场景扫描数据集中，选取具备**自我中心视角数据及相机位姿**的部分，数据格式和采样频率统一处理，以适应多视角数据，并构建了**全局坐标系**以便整合多视角观测，便于输出参考。
- 包含**760多个类别**，覆盖日常物体数据集中大部分类别和实例，分析了不同类别的空间占据情况，利于导航和运动规划。



Socially Compliant Navigation Dataset (UT Austin)

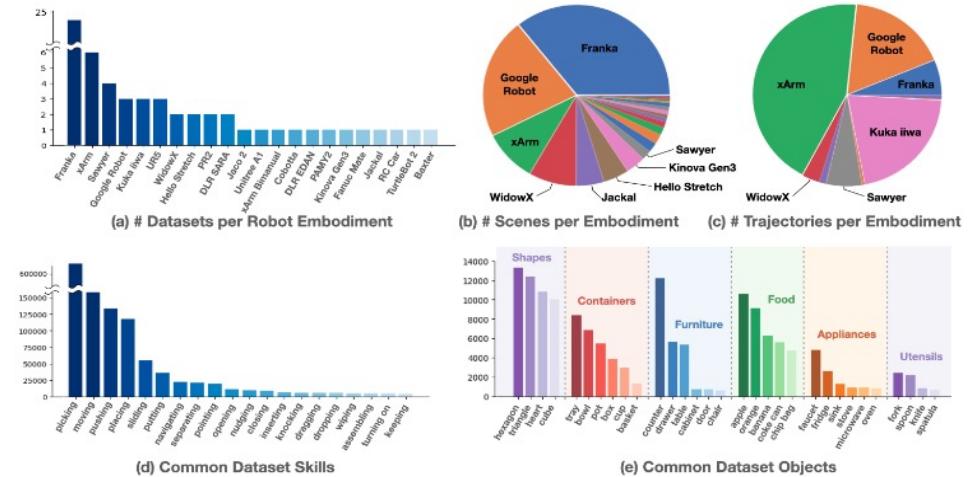
- 收集了多种模式的数据，包括**3D激光雷达、操纵杆命令、里程计、视觉信息和惯性信息**。
- 数据是在**两种形态不同的移动机器人**上收集的。
- 社会合规导航数据集（SCAND）是一个**大型的、第一人称视角**的数据集，专注于社会合规导航示范。
- 数据集包含**8.7小时的导航示范**，涵盖**138条轨迹**，总行驶距离为**25英里**。



具身智能-数据-机器人

Open X-Embodiment (Google)

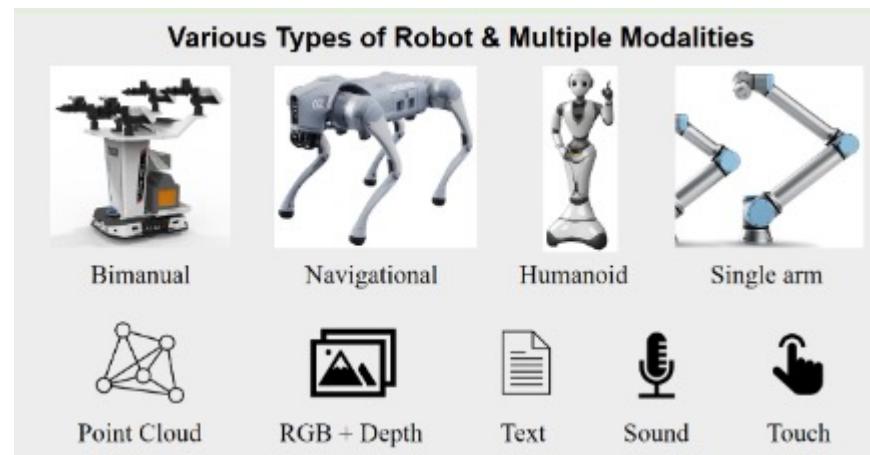
- Open X-Embodiment是大型的开源机器人真实数据集，包含超过100万条机器人轨迹。
- 涵盖22种不同的机器人形态，包括单臂机器人、双臂机器人以及四足机器人。
- 数据集由来自全球21个机器人研究实验室的60个现有数据集汇总而成。
- 数据集包含了多种常见行为和家庭物体。



具身智能-数据-机器人

ARIO (鹏城实验室)

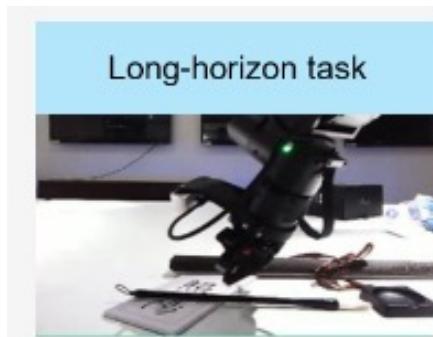
- ARIOD 数据集是为具身智能系统开发的，旨在提供一个统一的数据标准，以满足多样化和高效能的机器人训练需求。
- 数据集包含约 300 万个实验，涉及 258 个系列和 321,064 项任务，涵盖多种机器人操作和交互场景。
- ARIOD 数据集集成了多种感知模态，包括视觉、听觉和触觉等，提供全面的感知信息。
- 数据集结合了真实世界与模拟环境的数据，确保机器人能够在多样化的环境中学习和操作。



具身智能-数据-机器人

ARIO (鹏城实验室)

- 来自现实场景的一部分数据是在 Cobot Magic平台上收集的。研究人员设计了**超过 30 个任务**，主要集中在家庭环境中的**桌面操作**。这些任务不仅涵盖一般的**抓取和放置**技能，还包括更复杂的技能，如**扭转、插入、按压、切割等**。



Long-horizon task



Bimanual manipulation



Contact-rich task



Human-robot collaboration



Deformable object manipulation

"Pick up the rice paper, water dish, brush pen, and calligraphy copy, and place them in sequence at the top right, bottom right, bottom left, and top left corner of the table."

"Pick up the open book on the table with both hands for reading."

"Pick up the pencil on the table and write down the number '3' on the paper."

"Applying ointment to a person's skin."

"Fold the towel along both diagonals and place it at the center of the table"

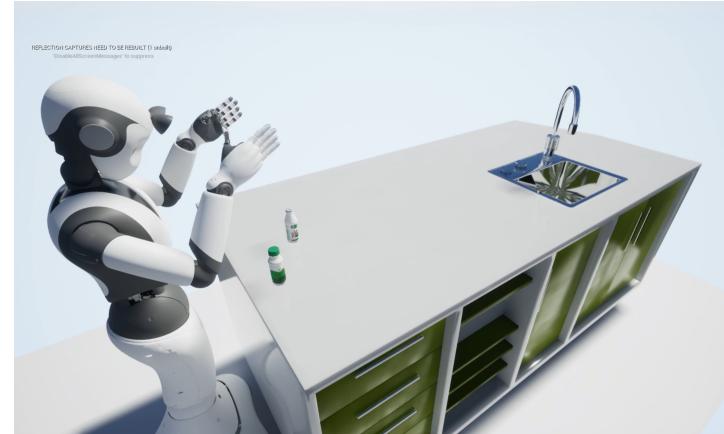
具身智能-数据-机器人

ARIO (鹏城实验室)

Cobot Magic 硬件的现实场景



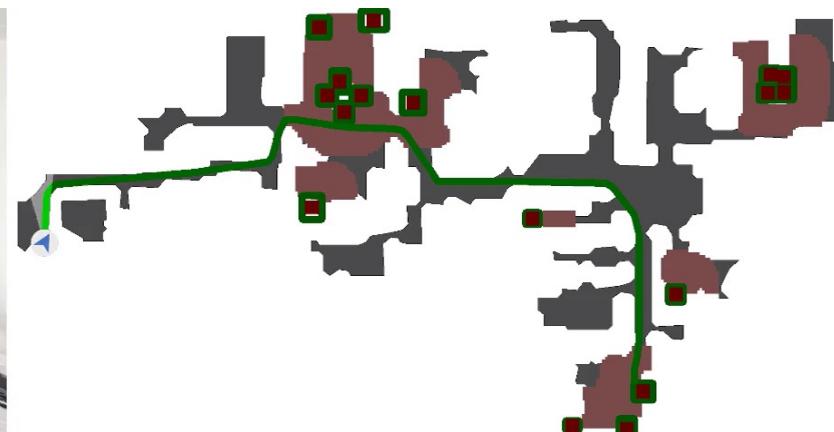
使用 Dataa SeaWave 软件进行仿真



使用 MuJoCo 引擎的抓取与放置仿真



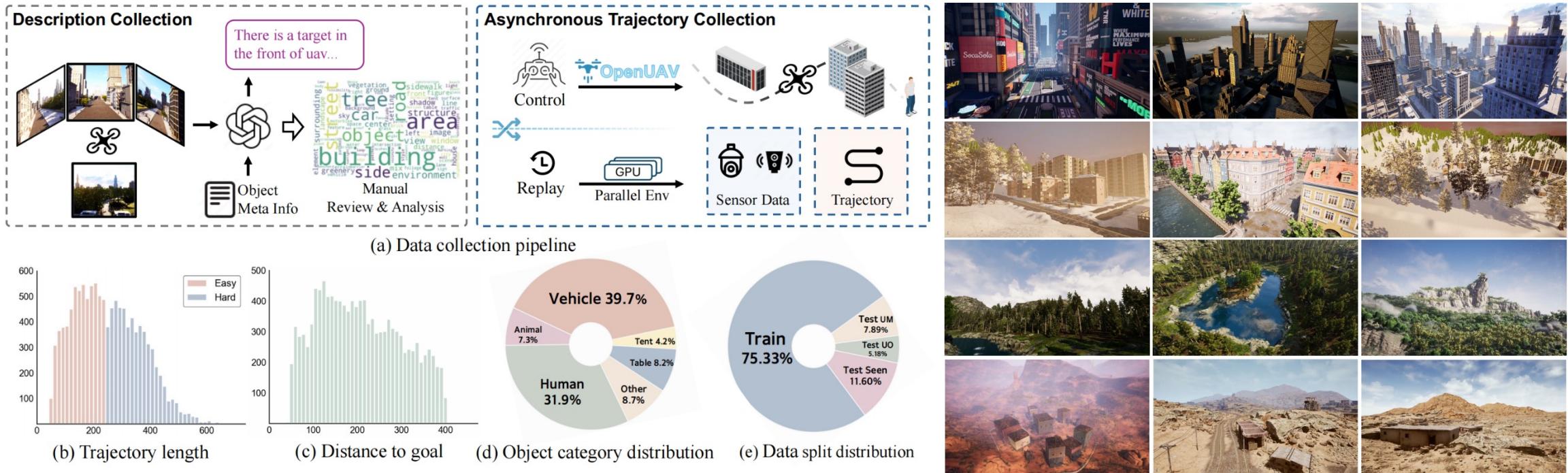
使用 Habita-sim 平台的导航仿真



具身智能-数据-无人机（组内工作）

UAV-Need-Help (可乐实验室)

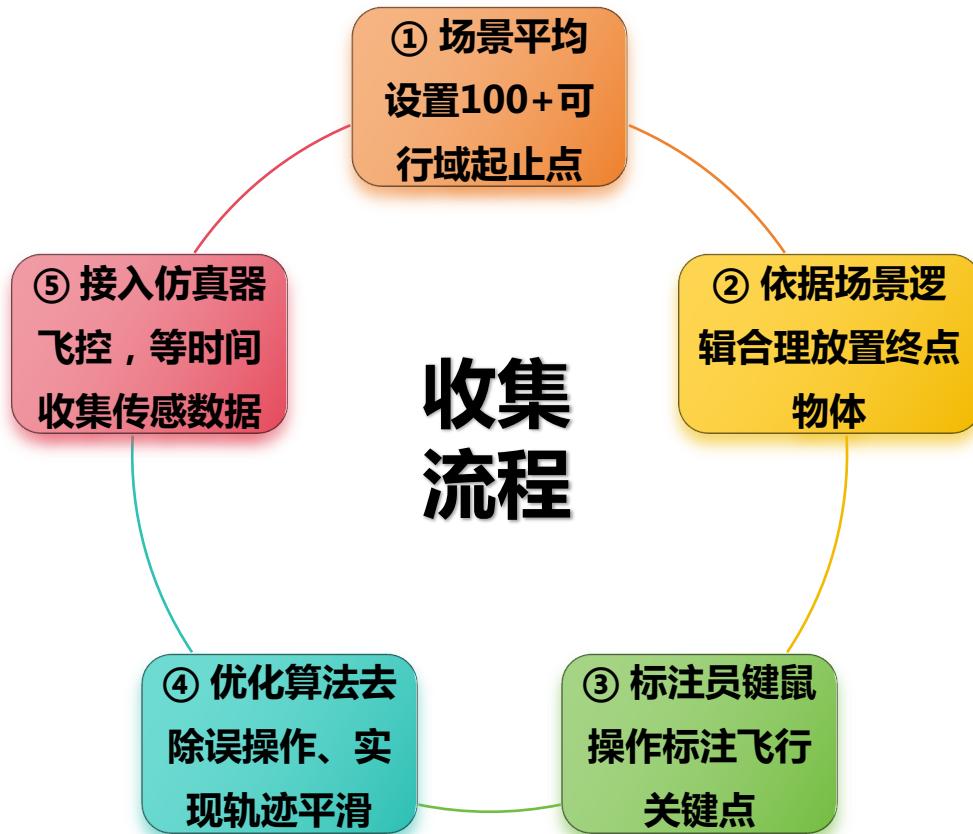
- 这个数据集是基于所提出的 OpenUAV 平台构建的，旨在为无人机（UAV）视觉语言导航（VLN）任务提供一个面向目标的真实数据集。



具身智能-数据-无人机（组内工作）

UAV-Need-Help (可乐实验室)

➤ 物体导向的无人机VLN数据集——轨迹收集



	AerialVLN	无人机平台
数据量	8446	12000
平均轨迹长度	661	295
平均每条轨迹的轨迹点/动作数量	动作数量 = 204	轨迹点 = 300
仿真类型	瞬移设置姿态	动力学飞行轨迹
语言指令	详细低阶动作描述	高阶任务+助手
数据格式	8个固定动作	轨迹点坐标
传感信息	前视图相机、深度相机、姿态信息	前后左右下视图相机、深度相机、激光雷达点云、IMU、姿态、速度、加速度信息

具身智能-数据-无人机（组内工作）

UAV-Need-Help (可乐实验室)

➤ 物体导向的无人机VLN数据集——物体描述收集



环境描述 Prompt：

#CONTEXT# 我需要你帮我描述图像中的物体及其周围环境。这是一个从底部、前、后、左、右视图拼接在一起的图像，中心对象的网格体名称为{name}。

#OBJECTIVE# 请从不同角度联系信息，描述中心物体的视觉特征，和周围的整体，优先描述周围物体>周围建筑>植被，描述优先描述颜色，形状。

#STYLE# 客观地描述，不要使用诸如网格体、骨架等术语。

#TONE# 客观地描述，不要用主观的评价词。例如:sunny, beautiful等。

#AUDIENCE# 无人机飞行员，描述目标及其周围环境，使无人机飞行员更容易找到目标。

RESPONSE# 请将以上任务写在一段话里，保持简洁，只是信息，不要有任何额外的输出。

最终得到的描述指令为 方位描述 + 物体描述 + 环境描述

具身智能-数据-无人机（组内工作）

UAV-Need-Help (可乐实验室)

➤ 物体导向的无人机VLN数据集——物体描述收集



环境描述 Prompt :

#CONTEXT# 我需要你帮我描述图像中的物体及其周围环境。这是一个从底部、前、后、左、右视图拼接在一起的图像，中心对象的网格体名称为 {name}。

#OBJECT# 请从不同角度获取信息，并描述中心物体的视觉特征，和周围的整体环境。优先描述周围物体>周围建筑>植被，描述为抽象或具体的形容词等。

描述指令：目标位于无人机的前方。中心物体是一个身穿白色短袖和蓝色短裤的人。周围沿街高楼鳞次栉比，街道上设有车道和绿色自行车道，建筑物附近摆放着撑着彩色遮阳伞的街头小贩。

#STYLE# 请使用抽象或具体的形容词等来评价。例如:sunny, beautiful等。

#AUDIENCE# 无人机飞行员，描述目标及其周围环境，使无人机飞行员更容易找到目标。

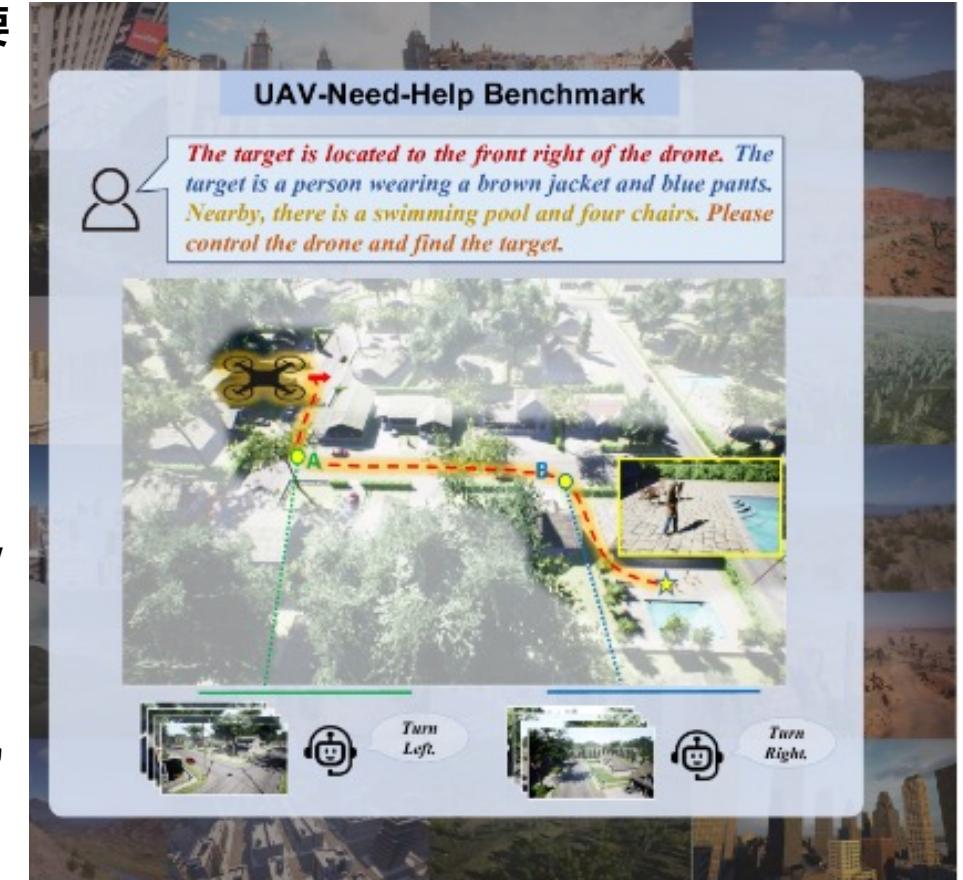
RESPONSE# 请将以上任务写在一段话里，保持简洁，只是信息，不要有任何额外的输出。

最终得到的描述指令为 方位描述 + 物体描述 + 环境描述

具身智能-数据-无人机（组内工作）

UAV-Need-Help (可乐实验室)

- 这是首个准确捕捉无人机复杂飞行动态的 UAV VLN 数据集，主要用于研究无人机在目标搜索任务中的表现。
- 利用 OpenUAV 平台的人工控制界面，由专家手动控制无人机搜索目标。记录无人机状态，并在无人机接近目标 5 米以内时终止轨迹记录，确保数据的有效性，最终获得 12,149 条有效轨迹。
- 数据集中包含的轨迹长度多样，短于 250 米的轨迹被分类为简单，超过 250 米的轨迹被分类为困难，确保任务的挑战性和复杂性。
- 最常见的目标描述包括建筑物、树木和汽车等，这些描述为无人机提供了上下文信息，增强了其通过视觉线索估计目标位置的能力。



05

核心要素：学习和进化架构

具身智能-学习和进化架构

智能体通过**和物理世界（虚拟的或真实的）的交互**，来适应新环境、学习新知识并强化出新的解决问题方法

真实环境

物理环境学习

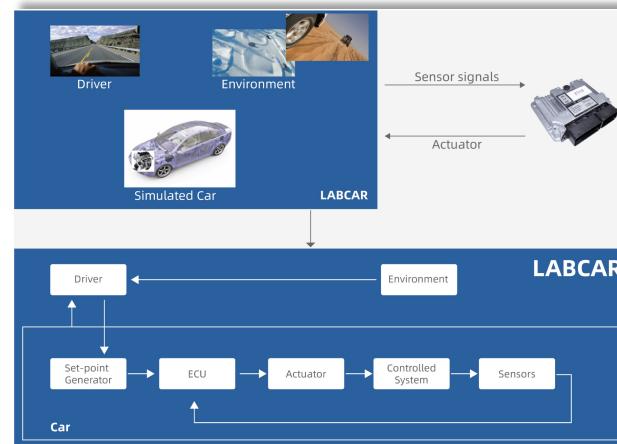
- 智能体必须在物理世界中移动并与环境互动



半实物仿真/硬件在环仿真 (Hardware-in-Loop)

Vehinfo LABCAR HiL

- 上海蔚赫提供了 HiL 硬件在环仿真测试设备



虚拟仿真

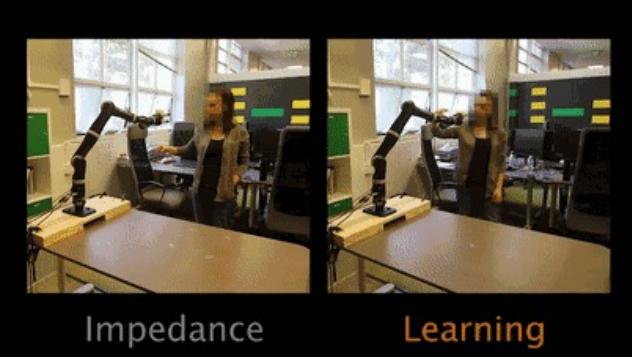
EmbodiedCity

- 清华团队发布基于虚幻引擎5的城市具身智能模拟环境



具身智能-学习和进化架构-真实环境学习

莱斯大学 Collab



- 莱斯大学的研究人员探索人与机器人互动的训练方法
- 能够让机器人**对于人类的碰触做出反应**，还能够根据推力输入改变它们的运动轨道

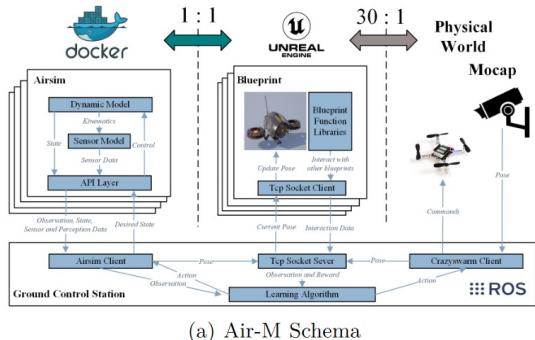
斯坦福大学 - Mobile ALOHA



- 低成本的**全身遥操作数据采集系统**
- 打造 Mobile ALOHA 的所有成本仅用了**3万美元**
- 通过与静态数据集联合训练可以将**成功率提高90%**，能够自主完成复杂的移动操作任务

具身智能-学习和进化架构-半实物仿真/硬件在环仿真学习

Air-M



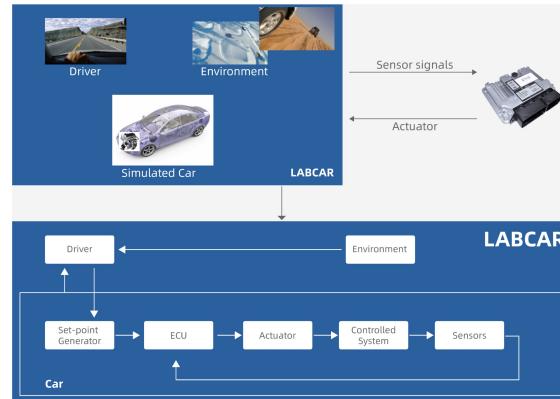
(a) Air-M Schema



➤ 传统机器人多智能体强化学习 (MARL) 平台面临困境：

- 已在各种模拟环境中部署，但**模拟环境外成功率有限**
- 支持训练的**代理数量有限**，甚至仅支持单个无人机
- Air-M 是一个利用 Docker 容器，实现**分布式大规模无人**机群体**仿真**，并支持**虚拟现实环境中部署**

上海蔚赫 - Vehinfo LABCAR HiL



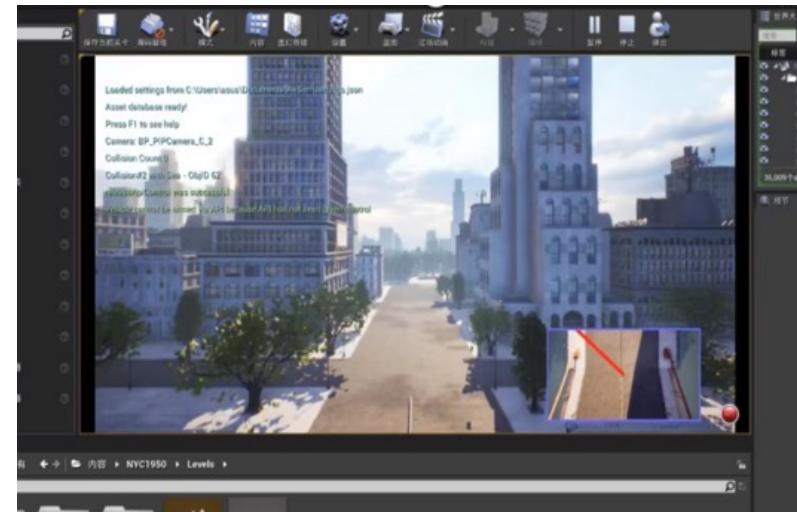
- **虚实融合仿真开发工具**
- 可用于船舶态势感知、汽车智能驾驶、低空飞行器、民用航空等**海陆空**多个领域

具身智能-学习和进化架构-仿真环境学习

- 在物理世界中直接训练智能体存在
速度慢、危险、资源受限、难以控制、不易复现等问题



- 仿真模拟三维环境，在复杂虚拟环境中训练和测试 AI 模型，**加速具身智能系统的研究**

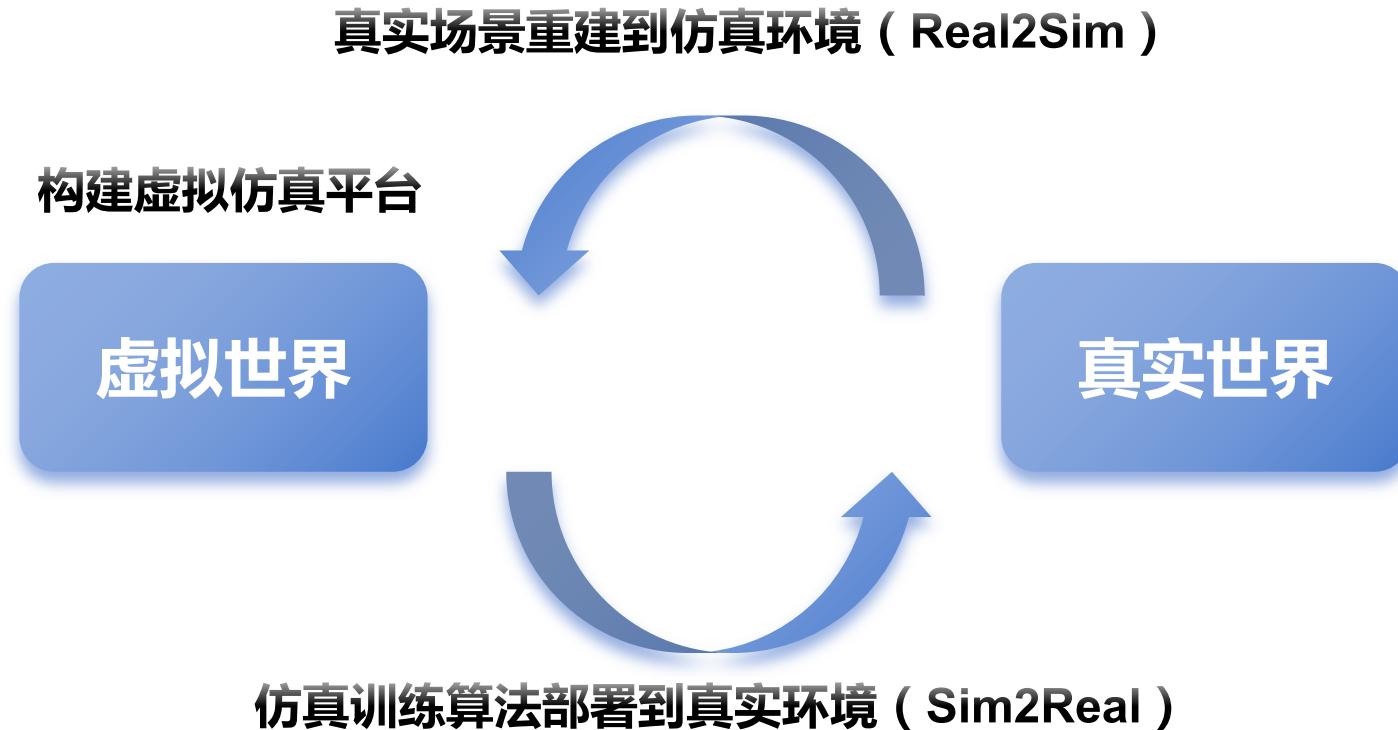


仿真验证

实机应用

具身智能-学习和进化架构-仿真环境学习

- 采用仿真环境进行部分学习是合理的设计，但真实环境的复杂度通常超过仿真环境，**如何耦合仿真和真实世界，进行高效率的迁移**是架构设计的关键



具身智能-学习和进化架构-仿真环境学习(Real2Sim)

商汤琼宇 SenseSpace

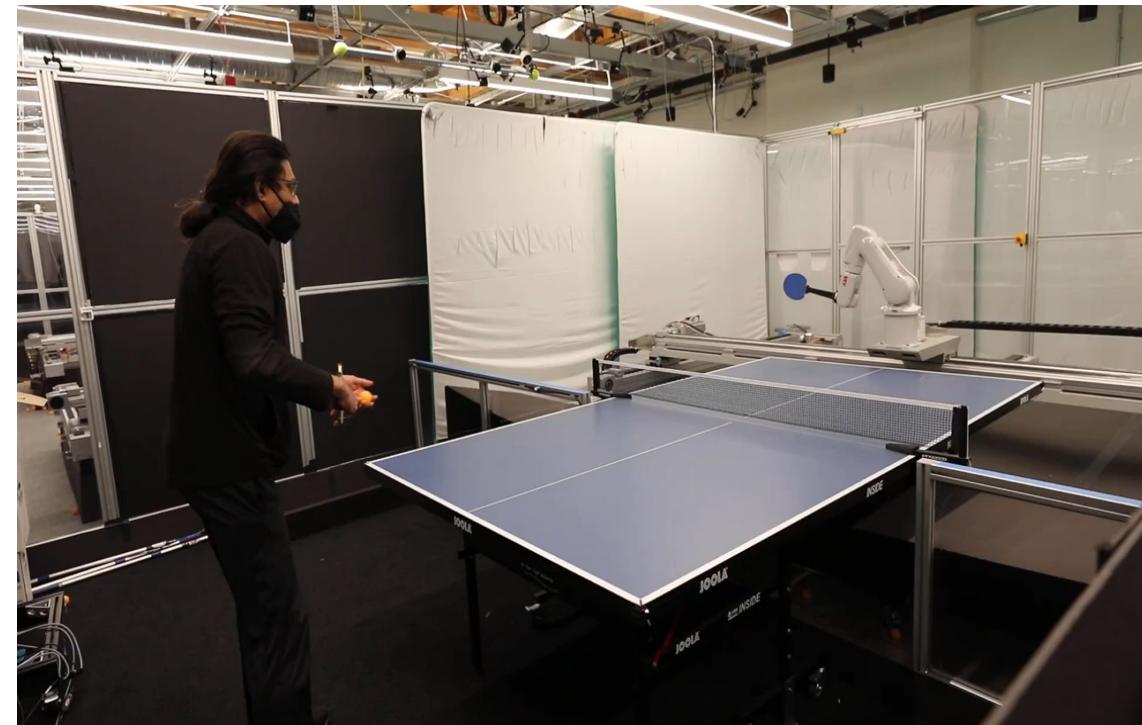
- 高精实景三维重建平台



具身智能-学习和进化架构-仿真环境学习(Sim2Real)

Google i-Sim2Real

- 从简单的人类行为模型引导，并在模拟训练和现实世界部署间交替进行



具身智能-学习和进化架构-仿真平台 (NVIDIA)

NVIDIA Isaac Sim



- 可扩展的机器人模拟器
- 可提供**真实物理模拟**和**可拓展的测试和验证**
- 可以使用**1000+**个3D资产，例如传送带、盒子、托盘等，以构建物理上精确的模拟

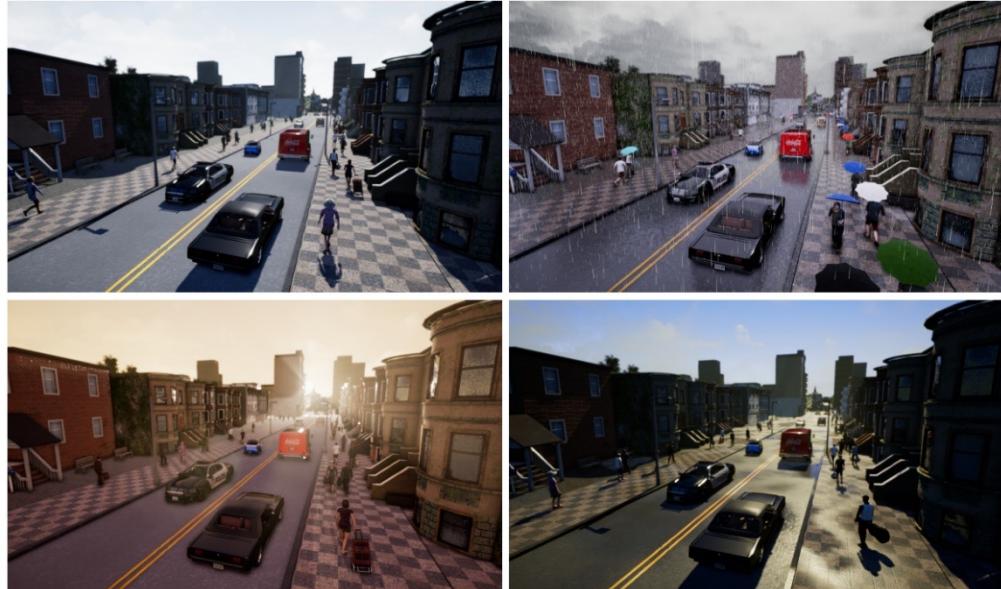
Orbit



- 基于**NVIDIA Isaac Sim**开发
- 提供了**交互式机器人学习环境的统一仿真框架**
- 提供了一系列**难度不同的基准任务**，从单阶段的柜门打开和衣物折叠，到多阶段的房间重组任务

具身智能-学习和进化架构-仿真平台 (Intel)

Carla



- 支持城市自动驾驶系统的**开发、训练和验证**
- Carla 支持传感器套件的灵活设置，并提供可用于训练驾驶策略的信号，例如GPS坐标、速度、加速度以及碰撞和其他违规行为的详细数据

SPEAR



- SPEAR 具有**300+**个虚拟室内环境，其中包含**2500+**个房间和**17000+**个可以单独操作的对象
- 构建了一系列评测任务，既包括**代表性的具身任务**，如场景理解、导航和规划等，也支持**传统任务**，如感知等

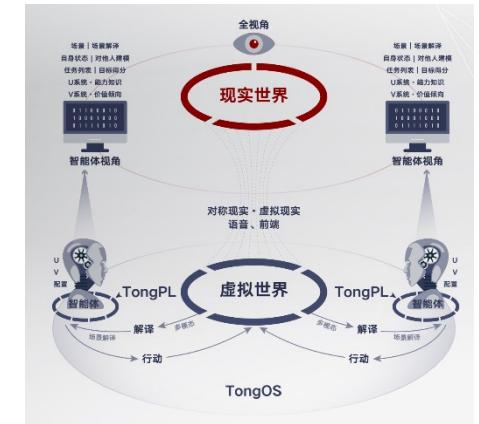
具身智能-学习和进化架构-仿真平台 (北京通用人工智能研究院)

通用智能体 “通通”



- 首个由**价值与因果驱动的AGI系统原型**，由**自研国产学习和推理框架**为底层支撑
- 拥有类人价值观，能**自主生成任务**，具备物理和社会常识，可保障**复杂任务高效执行**
- 国际领先实时**仿真与渲染引擎**，自研分布式组件架构，打造高度**物理逼真、可交互的训练场与软件系统平台**

“通境” (TongVerse)



- 集成了大规模场景生成能力，支持**多类型机器人技能训练**
- 提供自研**视觉-语言-运动联合解译架构**
- 可提供**10,000+**贴近工业生产和居家生活环境的**仿真场景**，同时支持动态开放环境下的机器人动力学仿真

具身智能-学习和进化架构-城市级仿真平台

上海AI Lab - GRUtopia (桃源)



- 首个**城市级**具身智能仿真平台，桃源仿真平台涵盖**89**种功能性场景、**10万**级别高质量可交互数据，构建起“**软硬虚实**”一体的机器人训练场，**数据、工具链、评测**三位一体

清华大学 - EmbodiedCity

Embodied City:

Embodied Agent in Urban Environment



- 以北京市国贸区域的**真实道路和建筑布局**为基础，结合了人流和车流的真实数据与模拟算法，构建了一系列评测任务，既包括**代表性的具身任务**，如场景理解、导航和规划等，也支持**传统任务**，如感知等

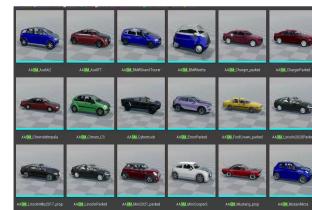
具身智能-学习和进化架构-“哪吒”具身仿真平台 (组内工作)

总体介绍

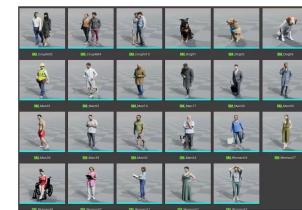
“哪吒”具身仿真平台是为无人机、无人车和其他无人系统的仿真和实验搭建的仿真平台，在开源框架**AirSim**的基础上开发，平台利用**UE4虚幻引擎**的图形渲染和物理模拟能力，支持多传感器模拟，动力学模型仿真，具有多种支持具身智能开发的特性。目前已集成了包含**城市、城镇、森林、沙漠等类型的22个仿真场景**，并具有**丰富的物体资产**，具备无人机控制和多种传感数据读取的**丰富API**，并具有**硬件在环仿真、多无人机控制、多卡多环境集群数据采集**等功能



控制和传感器信息



丰富的物体资产



多卡多环境采集

具身智能-学习和进化架构-“哪吒”具身仿真平台 (组内工作)

“哪吒”具身仿真平台具备强大的**自定义场景创建**并提供**丰富的场景素材**：

- 自定义场景创建**：可以导入社区或开源的地图、模型和其他资源，并可调整物体的样式、贴图、材质等细节
- 丰富的场景素材**：平台已包含22个场景，可分为城市、城镇、森林、沙漠等类型



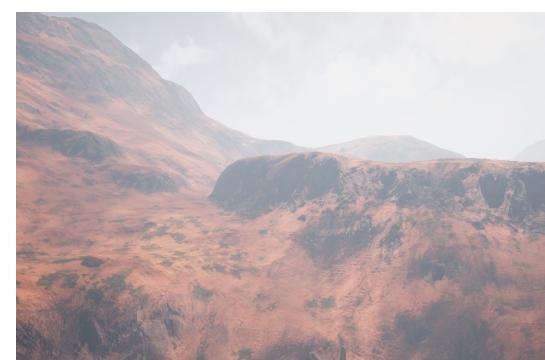
Carla城镇场景
场景包含有大量车辆、道路
以及城郊、野外场景



城市场景
具有丰富的建筑群
贴近现代城市风格



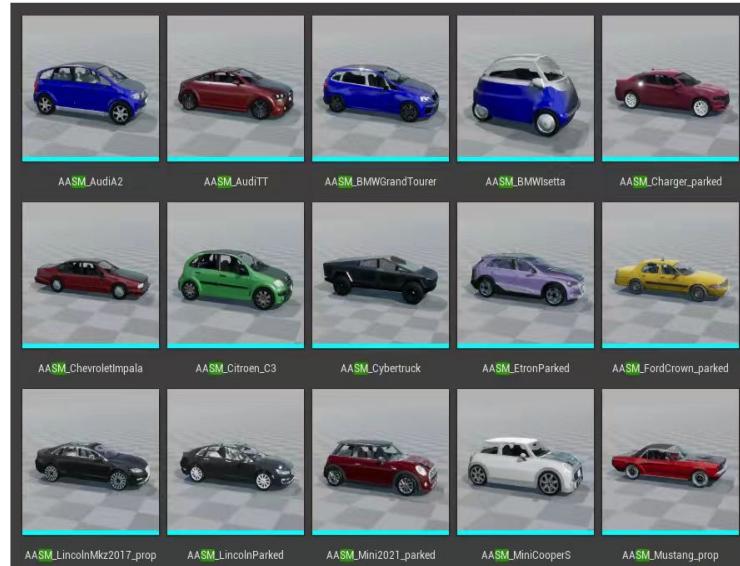
野外场景-森林
包含大量树木
地势起伏不平，山丘组成



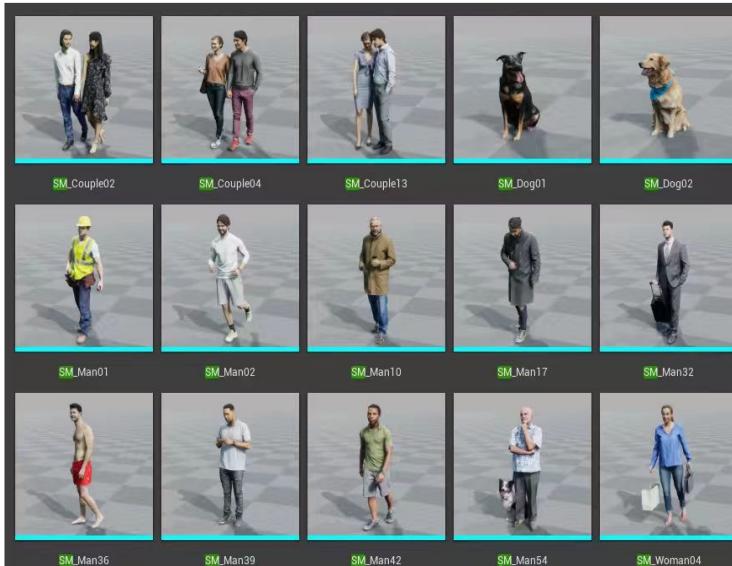
野外场景-沙漠
丘陵地貌
大部分由赤裸的沙丘组成

具身智能-学习和进化架构-“哪吒”具身仿真平台(组内工作)

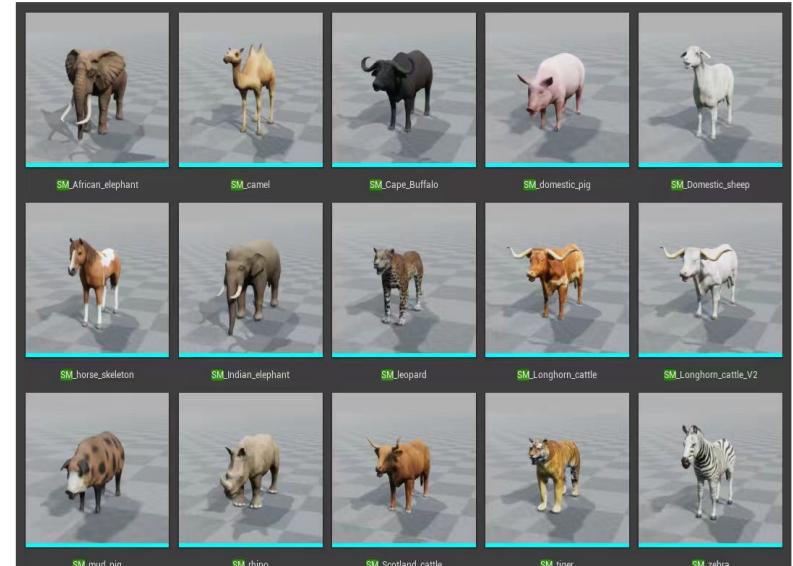
实体资产数量庞大：“哪吒”具身仿真平台现有**物体资产数量96个**，包括汽车类、自行车类、人类、动物、帐篷、桌子、椅子、路标、灭火器、垃圾桶等小物件



汽车



人



动物

物体资产可以通过API方式在仿真过程中动态创建、移动和销毁，以增加虚拟训练平台的仿真多样性

具身智能-学习和进化架构-“哪吒”具身仿真平台 (组内工作)

“哪吒”具身仿真平台在AirSim的基础上，可支持多种Python调用的无人机仿真API：

真实场景模拟：

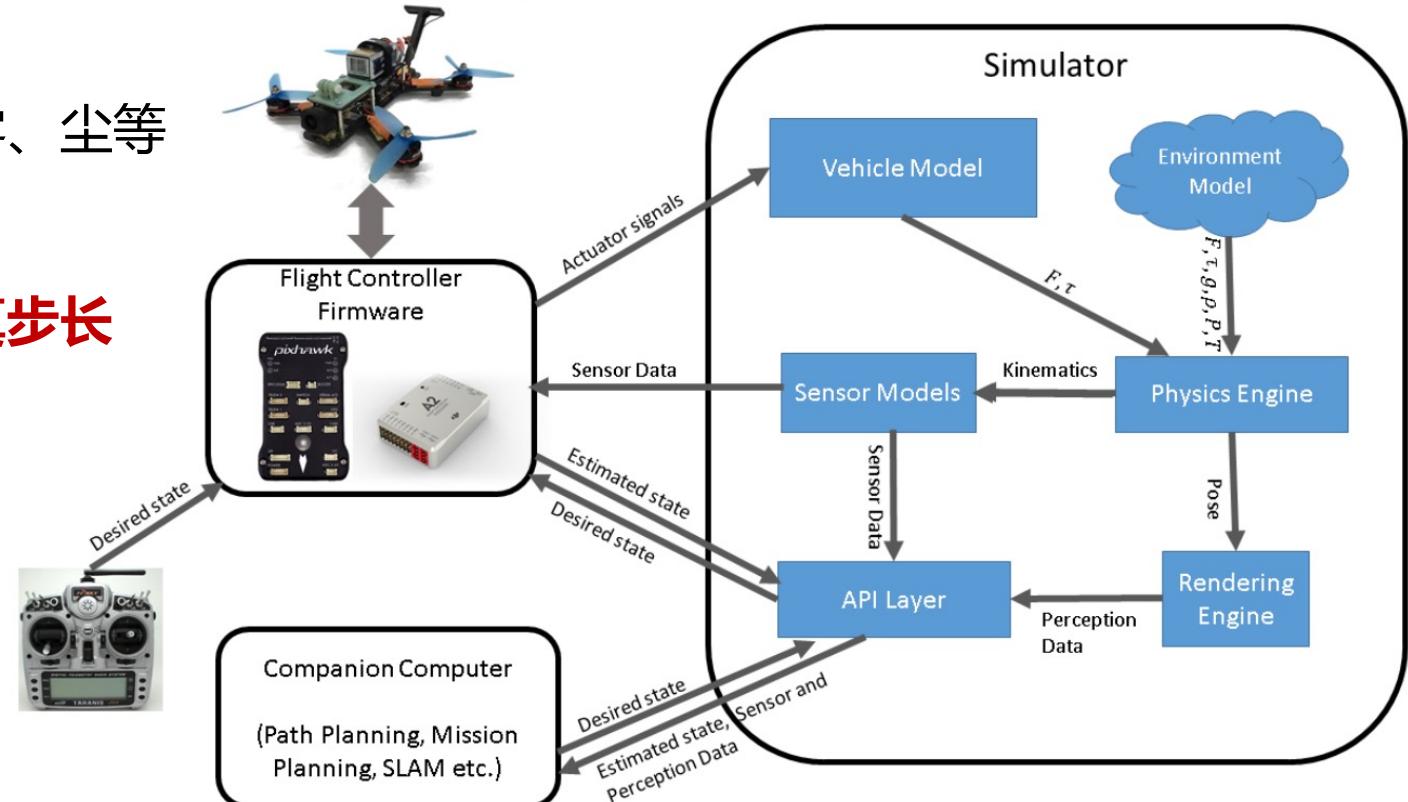
- 在场景中**生成物体并移动物体**
- 设置场景中的**天气**：雨、雪、雾、尘等

无人机控制：

- 控制仿真器**模拟速度**，设置**仿真步长**
- 根据速度飞行
- **根据航迹点飞行**

多种传感器的配置：

- IMU
- RGB-D 相机
- 激光雷达



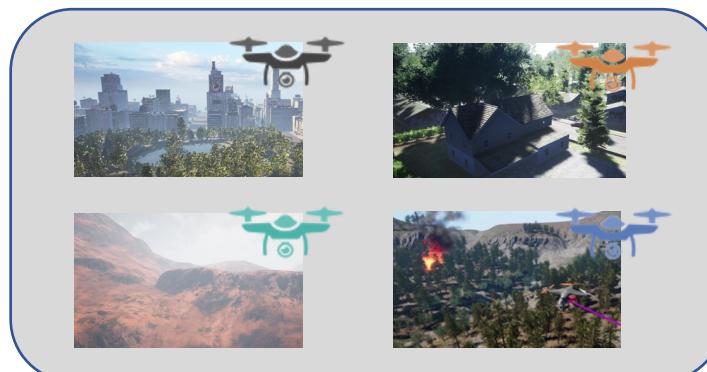
具身智能-学习和进化架构-“哪吒”具身仿真平台(组内工作)

“哪吒”具身仿真平台支持仿真和训练流程全覆盖：

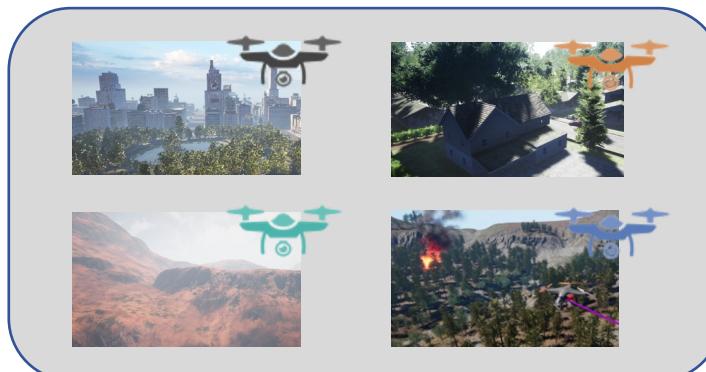
- **仿真控制**：支持PX4等飞行控制器的**硬件在环仿真**，支持通过遥控器控制无人机飞行，利用QGC-PX4通信链路，传输遥控器控制信号



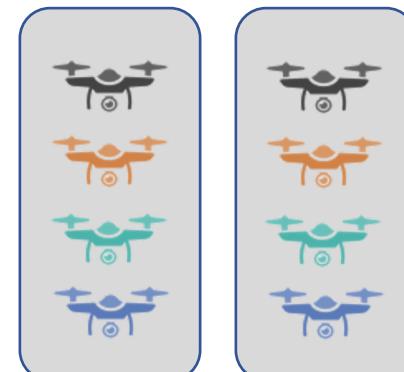
- **多机模拟**：支持**多无人机**在虚拟训练平台中同时控制飞行，具备开展多机任务的基础
- **并行仿真**：支持**多卡多环境集群式运行**，单台4卡服务器可同时运行16个仿真平台，大幅提高仿真效率



GPU0



GPU1



GPU2



GPU3

具身智能-学习和进化架构-“哪吒”具身仿真平台 (组内工作)

Demo展示

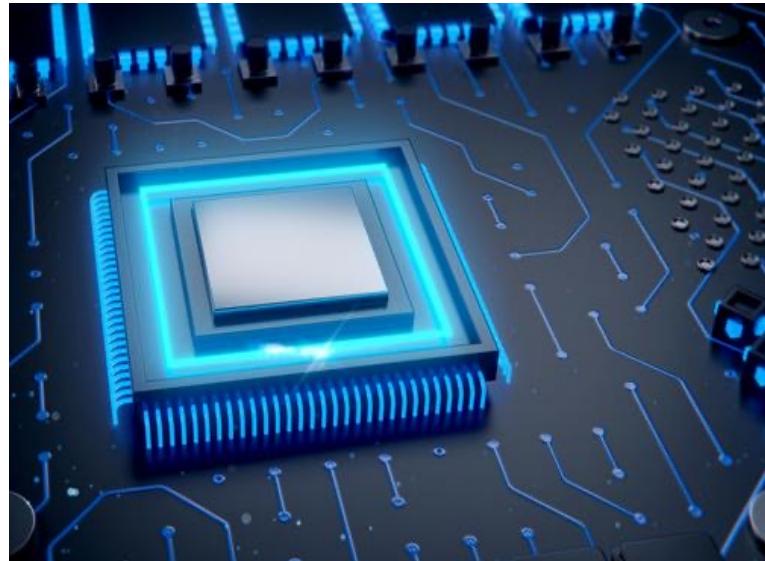
无人机平台展示

06

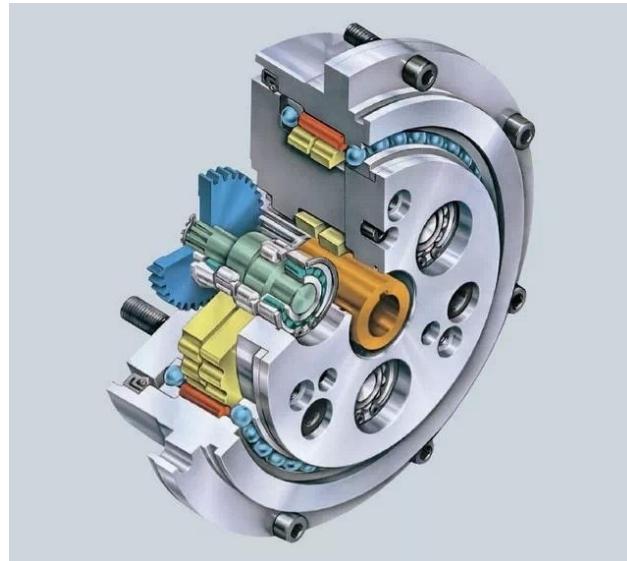
未来展望

具身智能-未来展望-高性能通用本体平台

- 形成具有**优秀运动能力和操作能力**的平台级**通用本体**
- 硬件与软件系统**深度集成**，以实现高效的数据处理和精确控制



- **存算一体大算力芯片**将大幅提升具身智能实时响应速度

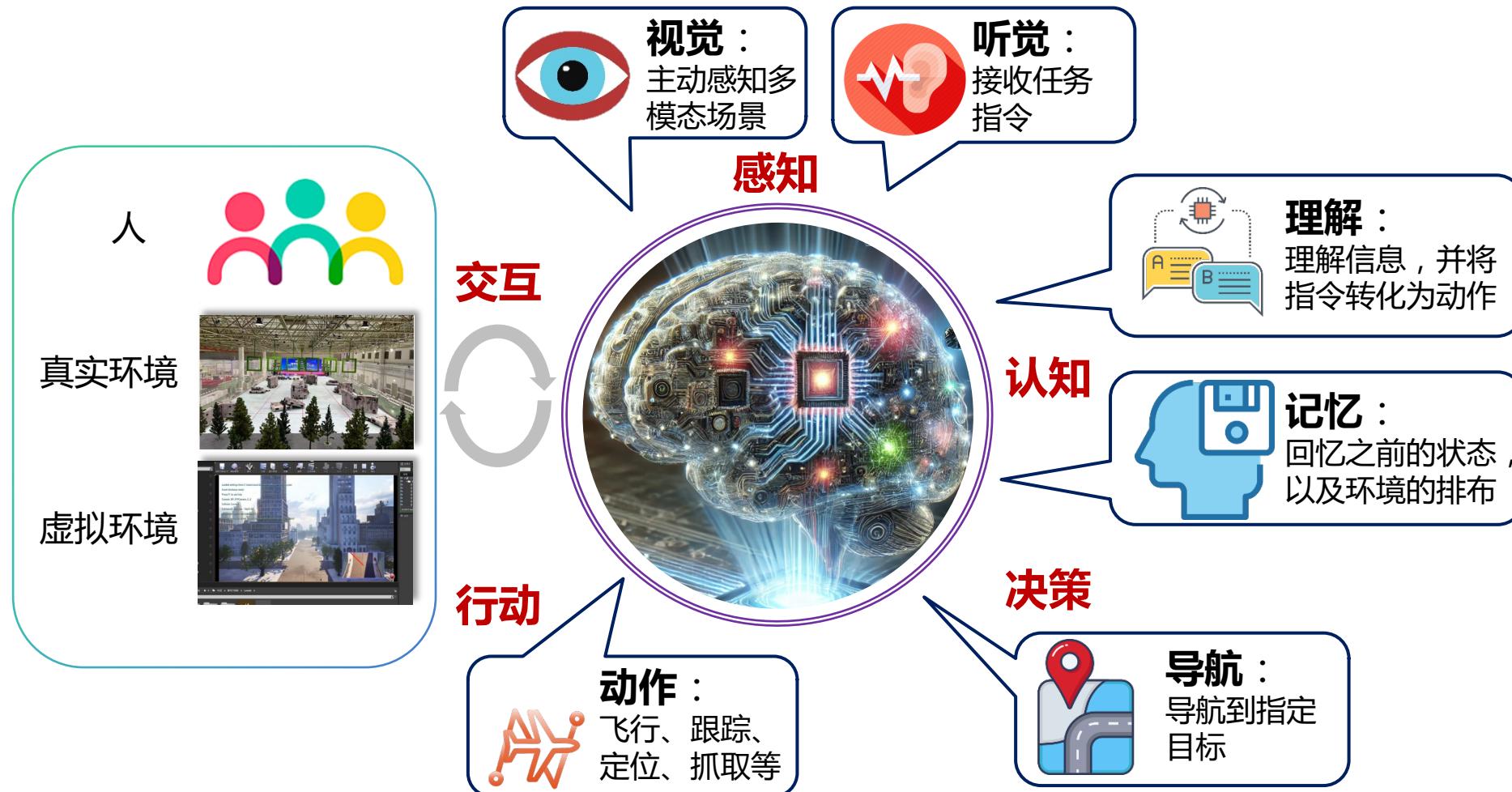


- **高精密零部件**将显著提升具身智能的行动精准性



- 具有**强大通用能力**的机器人将提升具身智能的泛用性，降低研发成本

具身智能-未来展望-数据驱动下的全链条贯通



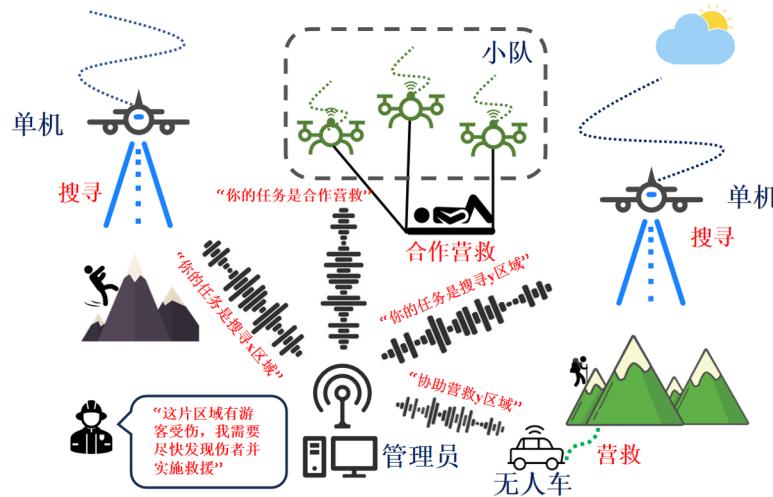
“交互-感知-认知-决策-行动” 贯通的具身智能系统将提高任务执行自动化水平，为智能化发展带来突破

具身智能-未来展望-异构多智能体协同

➤ 基于异构多智能体协同的控制算法

- 分配管理员、小队、单机及无人车角色，实现复杂环境任务分解、细化控制和协同操作
- 利用基于大小模型协同的算法，完成从自然语言到智能体具体控制的过程

□ 单机：收到管理员的任务之后，将其分解为函数级的指令，并提供相应的参数输入，以供轨迹预测模型执行



□ 小队：管理员向小队长机分发任务，队长机再发布相应队形指令给其他队员，完成整个过程

□ 管理员：给定人类的高阶指令，管理员决定调用智能体的数量和类别，并将高阶指令拆解为子任务，分发给相应智能体

□ 无人车：接收管理员指令和无人机的传感信息，规划路径到达任务目标点

具身智能-未来展望-具身智能大模型轻量化国产化

➤ 轻量化、国产化是具身智能大模型技术的主力突破方向



设计平衡模型的效果、功耗和推理速度的轻量化模型架构

- 通过**模型剪枝、量化、知识蒸馏、MoE**等技术，实现资源受限下设备部署
- 目前国内外的基座模型厂商、互联网公司等也在探索大模型轻量化技术
- 轻量化端侧部署在**隐私安全、可靠性**等方面有很大应用空间

端侧
部署



积极开展国产算力的模型移植，培育国产芯片大模型训练推理框架生态

- 国产大模型数量目前已**超过300个**，覆盖多个行业领域
- 各大国企积极布局大模型领域，**软硬件全栈国产化**，助力模型安全可信
- 持续建设人工智能软硬件支撑体系，在大模型技术及硬件算力自主创新

自主
创新

具身智能-未来展望-行业数据集的四个关键维度

高质量

高质量的数据能够确保模型学习的信息准确性和可靠性，数据的精准性直接影响模型的推理能力

多样性

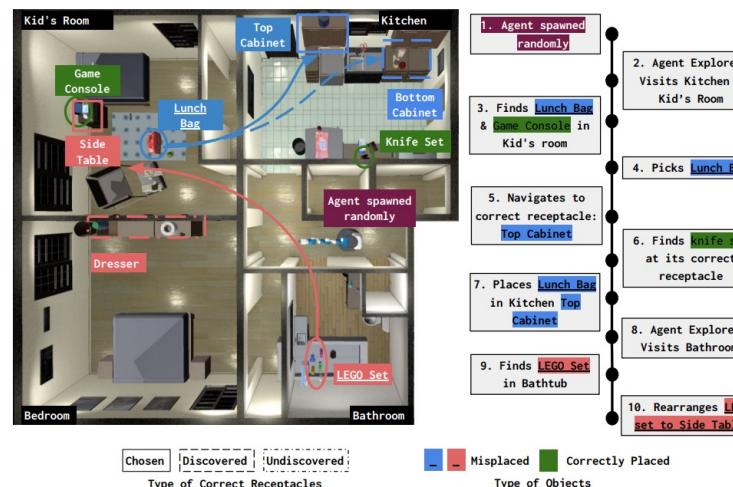
丰富的数据集能够增强智能体的泛化能力使其能在多变环境中表现良好

标准化

标准化的数据格式和处理方式能够提高训练效率，确保数据在不同模型和任务中都能被一致地使用

规模大

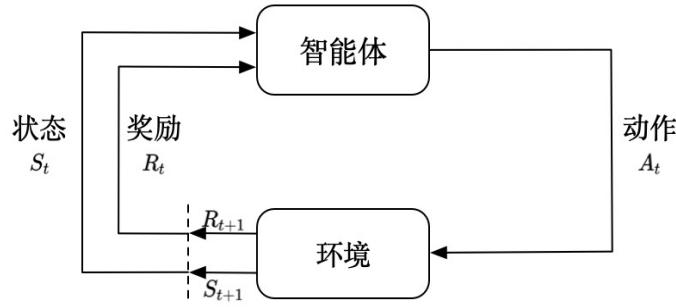
大数据量有助于提高模型的学习效果，更多的数据样本能够减少过拟合，提升模型的稳健性和表现



➤ 高质量、标准化、规模大并且多样化的数据集，将是未来具身智能大规模模型发展的基础和突破口

具身智能-未来展望-自适应学习和优化

强化学习



- 通过**强化学习算法**，具身智能可以在和环境交互中**学习最佳行为策略**，以最大化某种奖励

持续学习

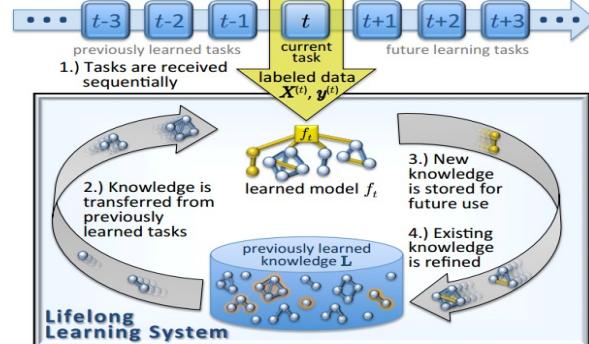
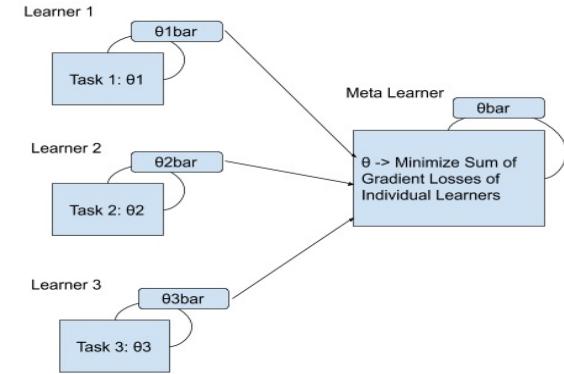


Figure 1. An illustration of the lifelong learning process.^[15]

- 具身智能将具备**持续学习能力**，即使在部署后也能不断从经验中学习，以**适应长期变化的任务和环境**

元学习



- 具身智能将采用元学习技术，快速适应新任务或环境，通过少量样本或经验快速调整自己的行为

- 探索如何减少人类干预，使控制系统更加自主成为重要发力点

- 实现具身智能形态和行为的自适应和优化，提升自主决策、能力和行为执行的精确性

具身智能-未来展望-产业跨界整合,开辟更广阔的市场空间



工业制造



家庭教育



自动驾驶



交通物流



航空航天



医疗健康

未来,具身智能将突破数据瓶颈和产品形态限制,以**经济,灵活且高效**的方式实现规模化应用

未来的具身智能应用将更加多样化,**个性化,智能化,跨界融合**成为机器人应用的新趋势



北京航空航天大學
BEIHANG UNIVERSITY

Beihang University

谢谢大家



北京航空航天大学

刘偲

liusi@buaa.edu.cn