



上海人工智能实验室  
Shanghai Artificial Intelligence Laboratory



Embedded OS in China  
国产嵌入式操作系统产业论坛

# Intern Robotics

## 『书生』具身全栈引擎及其关键技术

上海人工智能实验室

庞江淼

2025/08

# 机器人学习的发展



2019 年

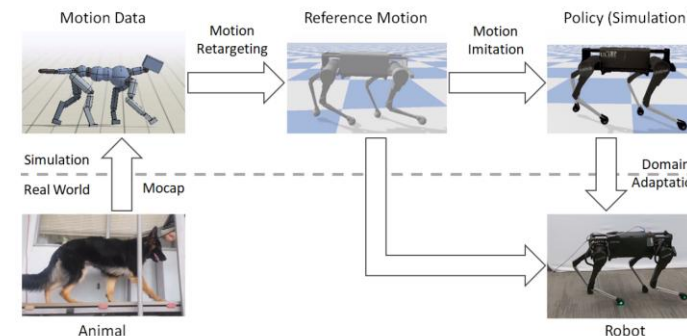
Learning agile of dynamic motor skills for legged robots

Science Robotics

首次强化学习 Sim2Real 控制机器人



Berkeley  
UNIVERSITY OF CALIFORNIA



2020 年

Learning Agile Robotic Locomotion Skills by Imitating Animals

RSS 2020 Best Paper

使用宇树科技机器人 Laikago



2021 年

Isaac Gym: High Performance GPU-Based  
Physics Simulation for Robot Learning

GPU 训练机器人控制策略, 效率  
提升 2-3 个数量级



2022 年



Unitree  
Robotics



傅利叶智能



智元机器人

X?

# 机器人学习的发展

机器人具身智能具备了成熟的基本条件: CV, NLP, etc...

- LLM -> Human-Robot Interaction
- CV/NLP -> Multi-modal Perception
- RL -> Generalizable Control
- 标准化的机器人: Unitree/Franka/etc...
- 仿真平台: Nvidia Omniverse/Isaac Sim
- 端侧计算: Nvidia AGX Orin

# 具身智能需要什么能力？

## 具身智能所需 三大核心能力

### 感知、认知

实时获取并理解环境中的多模态信息，构建对周围世界的准确表征。

### 推理、想象

基于已有信息进行内部模拟与逻辑推断，为后续决策生成提供支持。

### 移动、操作

自主规划运动路径并执行精细操作，实现与环境的物理交互。

## 关键难题

本体泛化

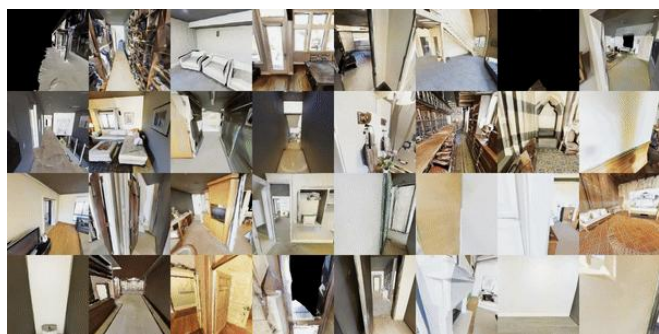
任务泛化

场景泛化

# 关键科学/技术问题

## 数据

合成数据：有助于本体泛化和场景泛化



真实数据：有助于任务泛化

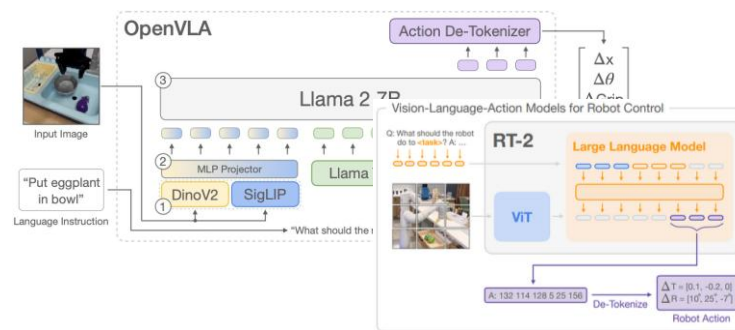


## 模型

双系统式VLA：大小脑协同

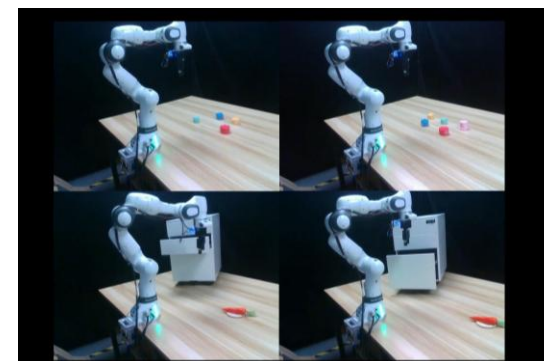


“A”式VLA

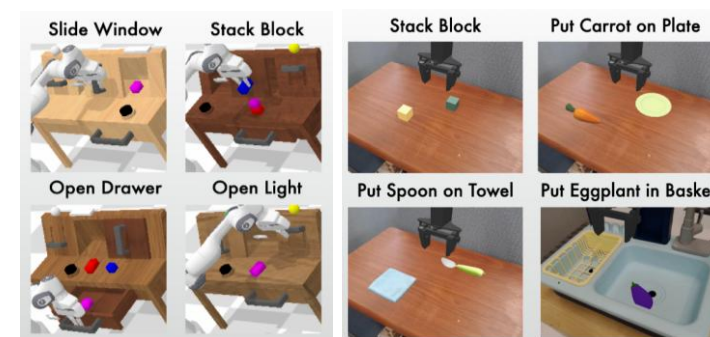


## 评测

真实评测：虚实一致但不可重复



仿真评测：可重复但虚实难以一致



# 一些过去的进展



“虚实贯通”技术体系



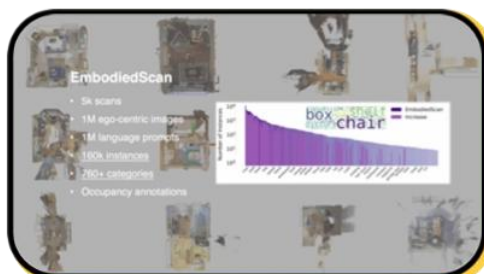
实景三维大模型书生·天际LandMark



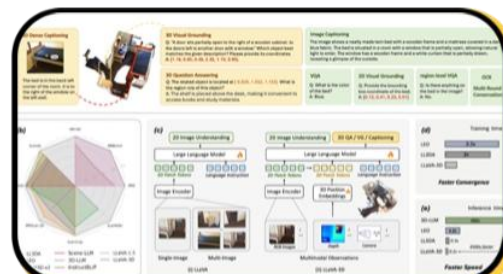
通用具身智能仿真平台GRUTopia2.0



端到端操作模型Seer



具身多模态感知: EmbodiedScan



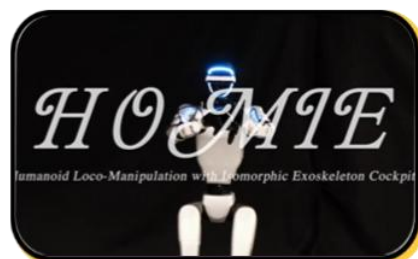
具身多模态感知: LLaVA-3D



4D世界模型Aether



可交互物体生成方法Infinite-Mobility



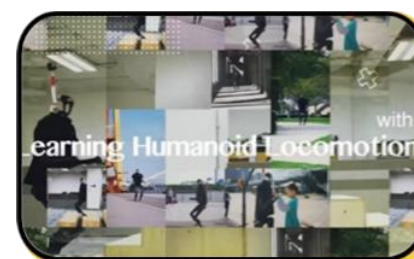
人形机器人驾驶舱HOMIE



人形机器人走梅花桩BeamDojo



人形机器人自主起立算法HoST



外感内模控制算法PIM



人形机器人全身控制算法HugWBC

# Intern Robotics

## 『书生』具身全栈引擎

# Intern Robotics

## 一脑多形、虚实贯通、训测一体

机器人实训场

赋能国地中心

具身作业场景

启动“光合计划”

开发者社区



Hugging Face



GitHub

具身AIGC

InternLandMark  
InternScenes  
InternWorldModel

数据引擎  
虚实贯通

具身数据集



打造高质量低成本的数据工厂

具身工具链

InternManip  
InternNav  
InternHumanoid  
InternSR

训测引擎  
一体协同

具身大模型



提供模块化可扩展的训练场

仿真引擎 InternUtopia

构建虚实交融的工作空间

标准化：任务定义、采集标注规范统一  
智能化：AIGC驱动多样、高效自动采集  
规模化：50000+条/台天，仅真机成本0.06%

一行代码实现跨本体部署  
三行代码定义任意任务  
五分钟上手实现机器人运动控制

易用：一键启动训测  
先进：多任务模型性能业界领先  
全面：六大任务、20+数据集、50+模型

# 通用具身仿真平台 InternUtopia

**通用模块化仿真框架：“三行代码”  
轻松定义任意具身任务**

**高精场景自动生成：十万级场景资产，  
百万级物体资产，“一键生成”复杂场景**

**数据的高效采集：提供多种遥操作工  
具，单机单日采集50000条高质量数据，  
一键完成分布式部署**

**虚实一致的仿真评测：一键评测导航  
和操作**

## 数据引擎 InternLandMark

# 场景数字化

《高精度、全场景通用重建》

高效低成本，采集与建模效率较传统方法提升数十倍，具备更强鲁棒性与适应性

高精度几何 + 照片级渲染，实现精准建模与高保真视觉效果的统一

全场景通用性，支持室内、室外、高空、地面等多类场景

## 数据引擎 InternScenes



高效无限生成，秒级生成速度，  
满足海量数据需求

类别丰富，结构多样，覆盖22  
个类别、6种运动关节，支持交  
互模拟

高质量模型，超越行业基准

## 数据引擎 InternWorldModel

# AETHER 4D世界模型

"重建—预测—规划"一体化框架

目标导向视觉规划，感知与决策协同，实现智能行动路径规划

4D 动态重建，实时建模动态环境时空变化

动作条件视频预测，支持零样本泛化，基于动作意图预测未来感知，迁移能力强

# 具身大模型 InternVLA-N1

上海人工智能实验室  
Shanghai Artificial Intelligence Laboratory

Intern Robotics



System2  
Grounded Waypoint Planning

穿过桌椅，左转进入咖啡厅，在咖啡柜台前停下

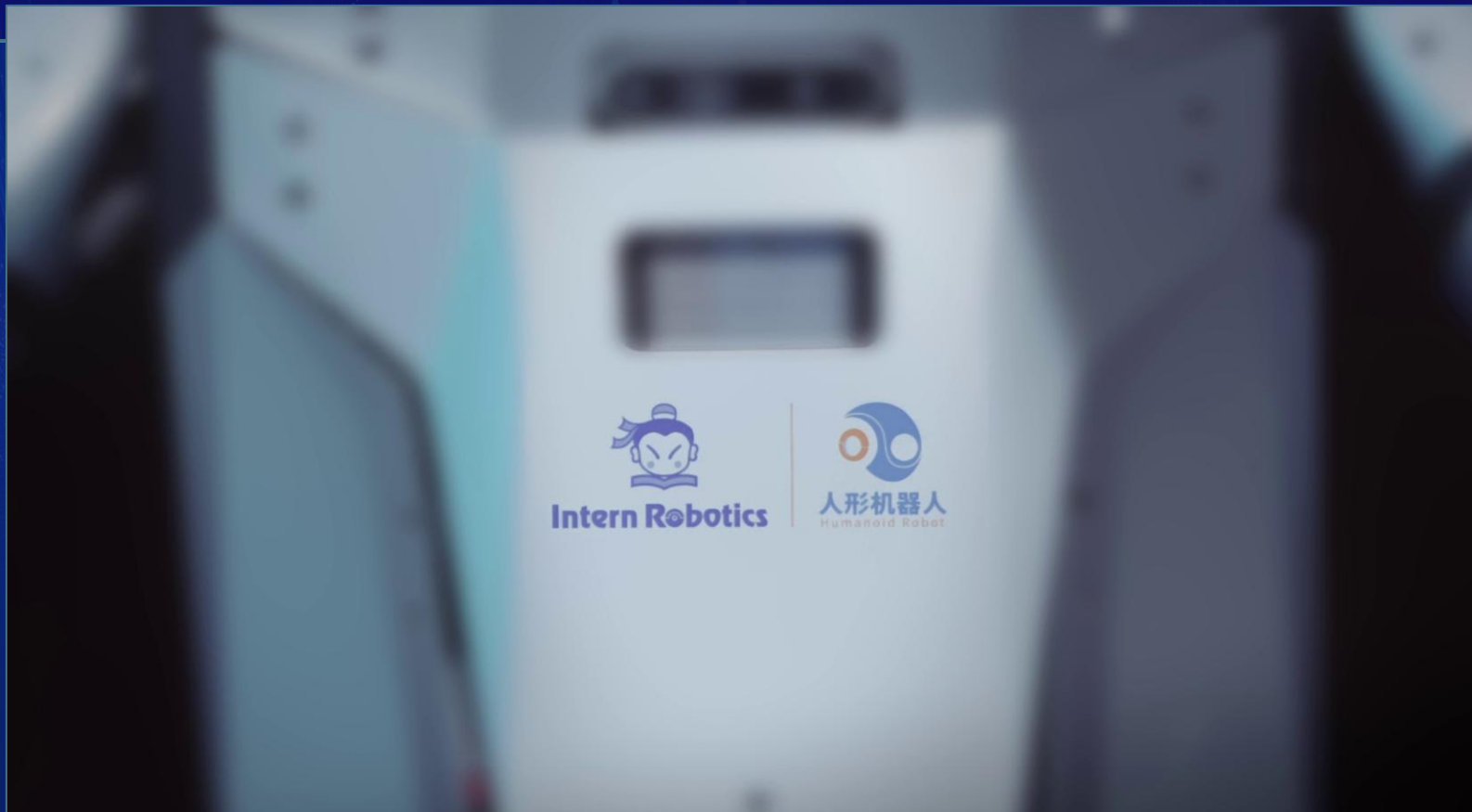
Walk through the tables and chairs, turn left into the café, and stop at the coffee counter.

纯仿真数据训练的首个双  
系统导航大模型

融合国际领先的超长程指令  
跟随和密集动态障碍物避障  
能力

跨本体跨场景零样本泛化

## 具身大模型 InternVLA-A1



统一的Transformer架构，集场景理解、任务想象与精准执行能力于一体

视觉预测指导动作执行，赋能稳定的动态交互

虚实数据混合训练，提升极端场景的泛化能力



## 具身大模型 InternHumanoid-多模态动作生成大模型

百万量级三维动作训练数据，  
实现文本零样本泛化

原生多模态自回归动作生成  
大模型，有效扩展至7B

可服务于一脑多形全身运动  
控制策略

## 具身大模型 InternHumanoid-一脑多形全身运动控制



统一机器人运动学语义抽象，提高跨本体泛化能力

通用全身运动控制策略，零样本泛化不同参考动作

模块化和可迁移的 Sim2Real 部署平台，轻量化跨本体运控算法部署和评估

## 具身大模型 InternVLA-RL-真实世界强化学习



评价-动作一体化通用模型，  
支持开放任务的进度判定

开放任务的大规模真实世界  
强化学习框架，支持自主进  
化和人机协同进化

# 运动智能 | 传统控制如何与强化学习交叉融合

The background is a solid blue gradient. Overlaid on this are faint, light blue circuit-like lines that form a network pattern, primarily concentrated in the center and right side. A large, semi-transparent 'AI' watermark is visible on the right side of the image.

# 运动智能：混合内模控制 (Hybrid Internal Model)



关节数量: 12

State Space:

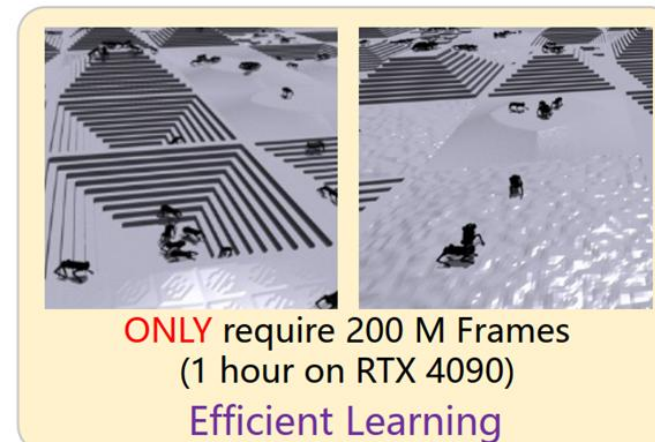
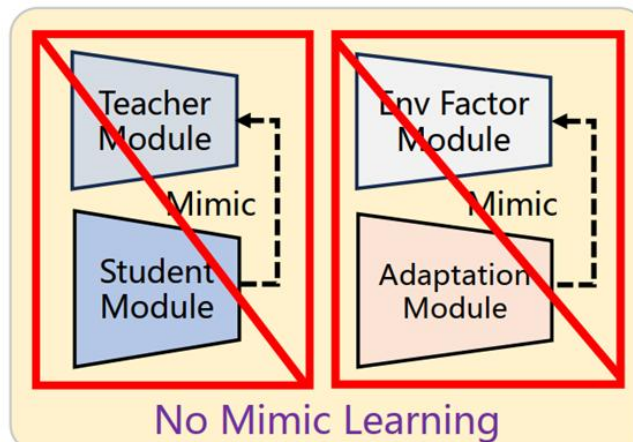
- 期望速度
- 关节信息: 位置、转速
- IMU: 角速度、重力方向
- 其他传感信息: 深度图等
- 之前几帧的 Actions

Action Space:

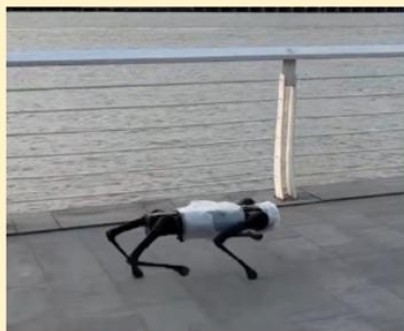
- 关节位置

关键: 准确的环境状态估计

# 运动智能：混合内模控制 (Hybrid Internal Model)



## Hybrid Internal Model



Sprinting



Stairs



Rough Slopes



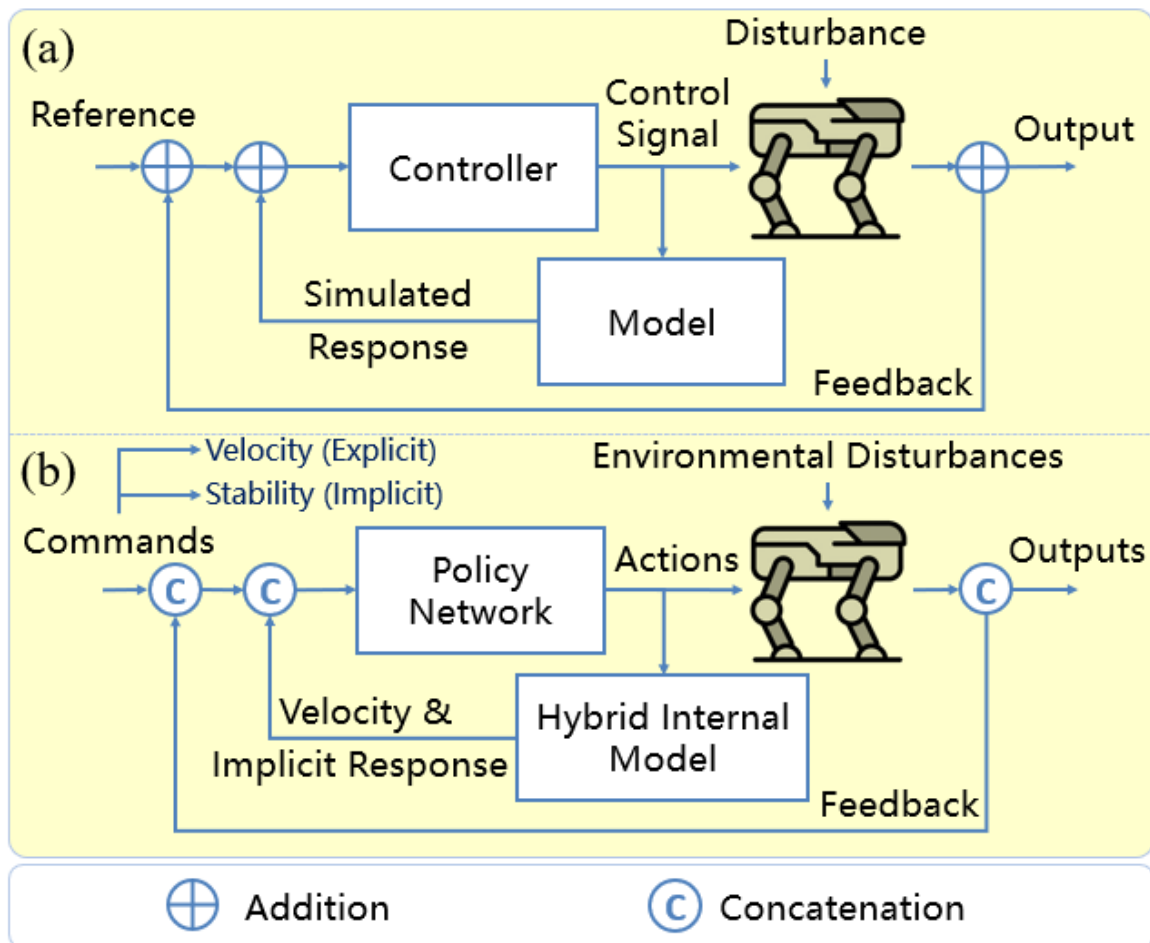
High Platform

Substantial Agility



# 运动智能：混合内模控制 (Hybrid Internal Model)

The Classical Internal Model Control



## Key Insight:

External States -> System Disturbances

System Disturbances -> Robot Responses

通过对机器人未来状态的估计

替代

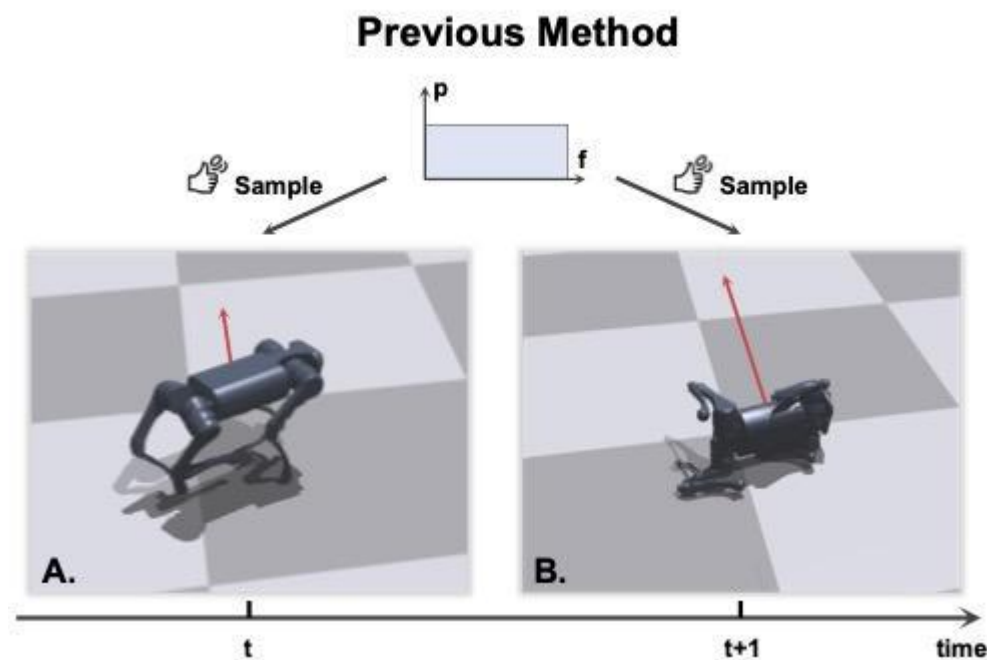
对外界环境扰动的估计

**Hybird = Velocity + Stability**

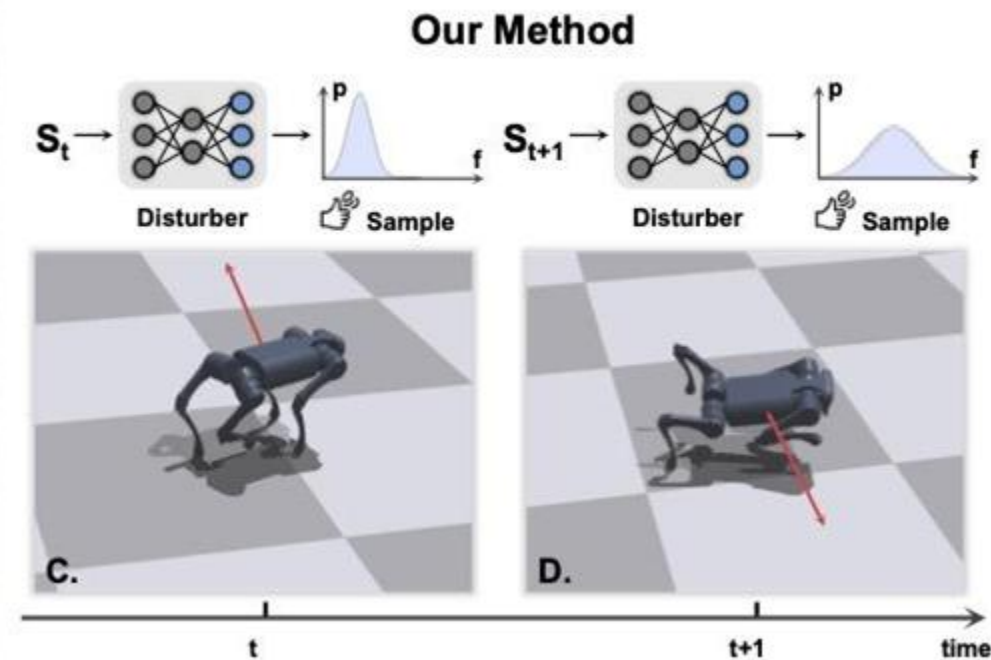
# | 运动智能：鲁棒控制 H-Infinity Control



# 运动智能：鲁棒控制 H-Infinity Control



Uniformly Sampled Disturbances



Adaptive Disturbances

Adversarial between the Robot and the Environment

Optimize with H-Infinity Constraint



原地起立可行走

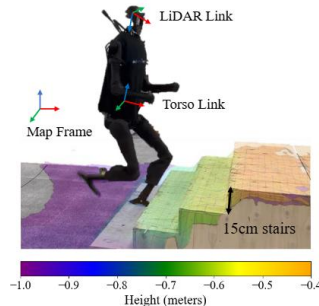
Learning H-Infinity Locomotion Control

OpenRobotLab

# 运动智能：从基础行走走到移动操作



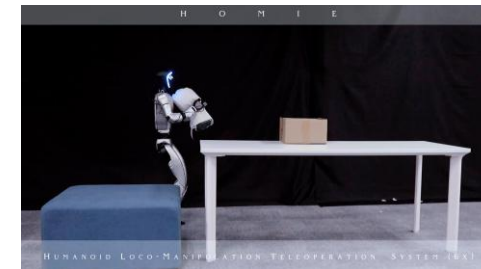
基础行走



感知融合



全身控制



移动操作

强化学习？传统控制？ Motion Priors？ 大规模数据？

**HIM**

传统内模控制

融合强化学习框架

仿真训练一小时

高效可泛化

**PIM**

复杂地形可泛化行走

**BeamDojo**

走梅花桩

**VC-Com**

处理动态场景

**HugWBC**

多步态全身控制方法

**HoST**

可泛化自主起立方法

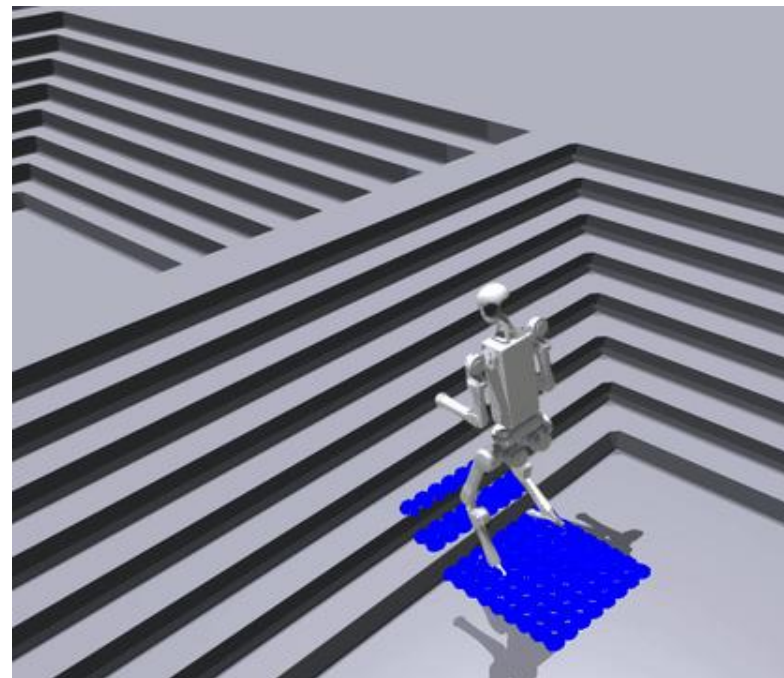
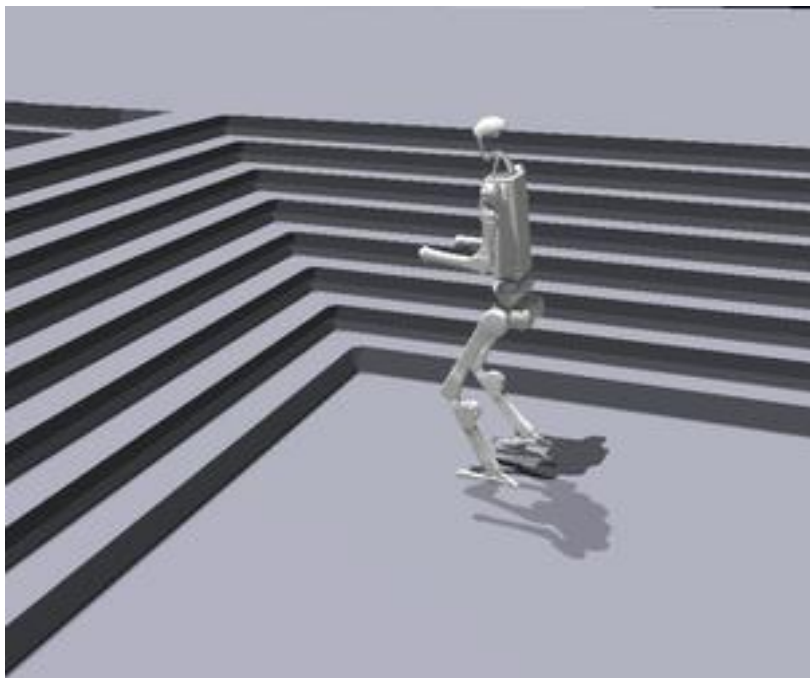
(RSS 2025 最佳系统论文提名)

**HOMIE**

人形机器人驾驶舱

# 运动智能：感知融合驱动复杂地形可泛化运动控制

**关键问题：**人形机器人本身结构“不稳定”，复杂地形可泛化运动控制必须融合感知信息



**感知融合运动控制的关键难点：**

**难点1：**感知信息很难完全精准，会不可避免的给控制系统带来扰动

**难点2：**感知信息 Sim-to-Real Gap 大，会增大虚实迁移的困难程度

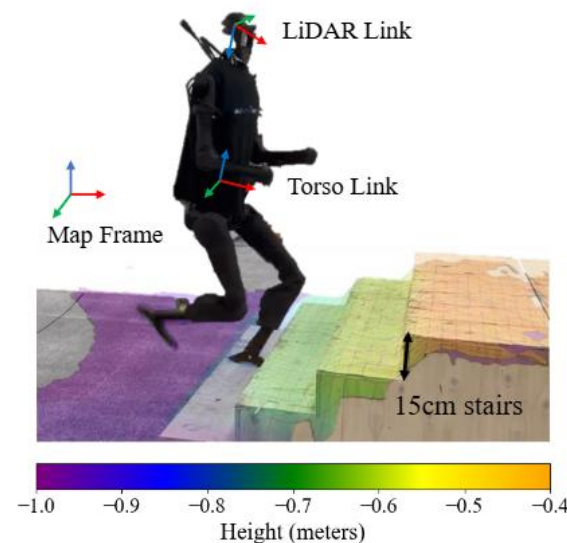
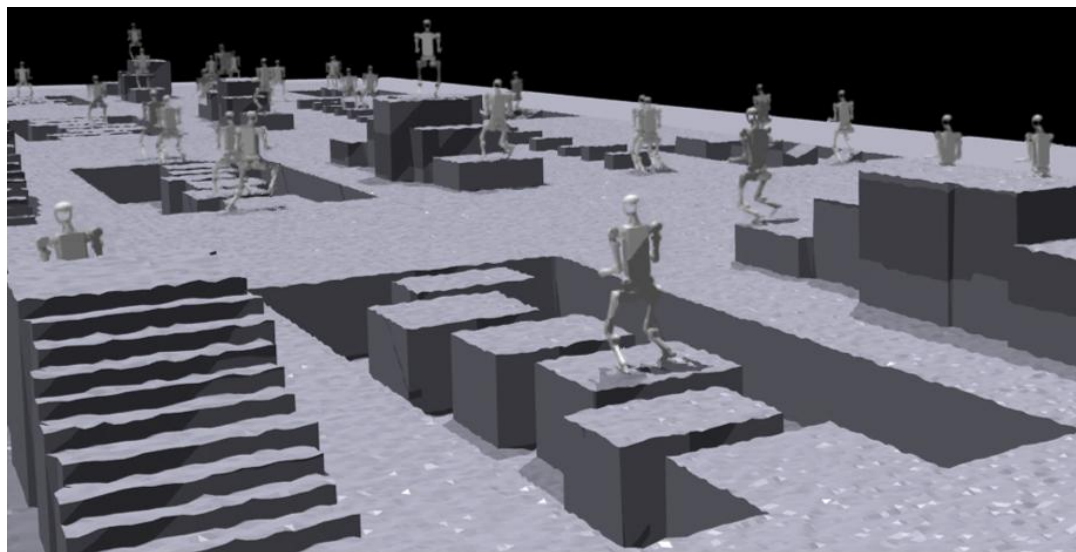
**难点3：**需要渲染感知图像，会影响控制策略在仿真中的学习效率

# 运动智能：感知融合驱动复杂地形可泛化运动控制

## 解决思路：

仿真训练直接使用障碍地形的绝对高度进行策略学习

真机推理采用使用 onboard, continuously updated elevation maps centered around the robot 进行感知



方法优势1: HIM 范式, Batch-level Contrastive Learning, 高效训练

方法优势2: 训练无需 encode/render RGBD 信息/原始点云, 仅需 RTX 4090 上 3 小时训练时间

方法优势3: 持续对机器人周围全局信息的感知, 能够抵抗传感器的移动/噪声

**系统优势：** 【多种人形机器人】 【室内外地形】 【上下台阶、跳 50 cm 高台、跳 Gap】 的策略

A wooden model of a staircase with a ramp, set in an office environment. The model is made of light-colored wood and consists of a series of steps on the left and a long, flat ramp on the right. The background shows office cubicles with grey partitions and fluorescent lights on the ceiling.

# Continuous up and down stairs

(single-shot view)



30cm

50cm

40cm

50cm

50cm



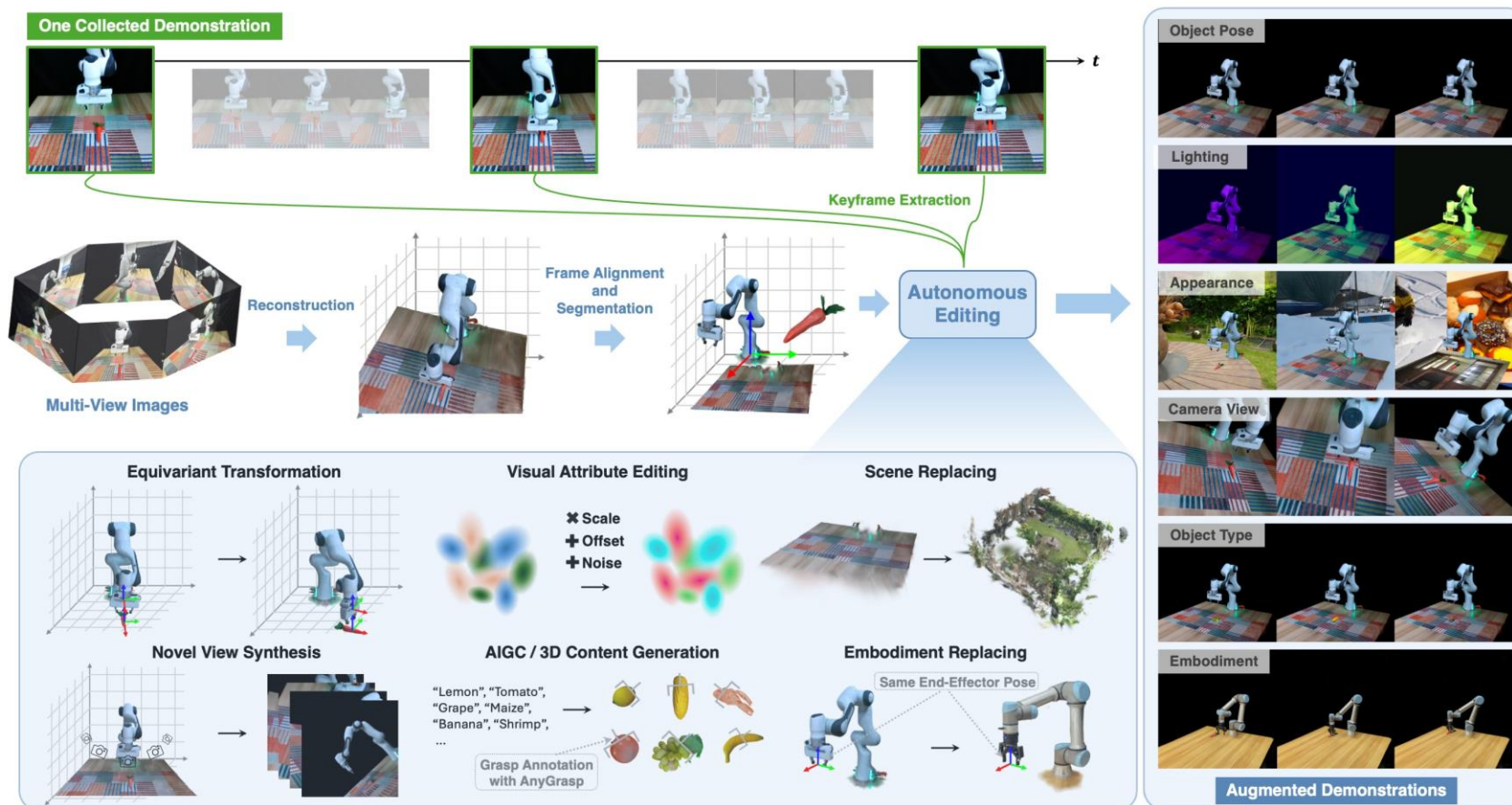


操作智能 | “Real2Sim2Real” 世界建模如何助益具身操作

The background features a large, light blue stylized letter 'A' on the right side. Overlaid on the entire background are intricate, light blue circuit-like patterns consisting of lines and small squares, resembling a network or data flow diagram.

# RoboSplat : 生成式世界建模赋能的数据扩增

借助**3DGS**, RoboSplat实现了高保真的操作场景重建。通过**可微分渲染**技术, 3DGS将重建后的场景与真实世界的尺寸和坐标系精准对齐, 并结合分割模型提取机械臂和物体, 构建结构化表示。



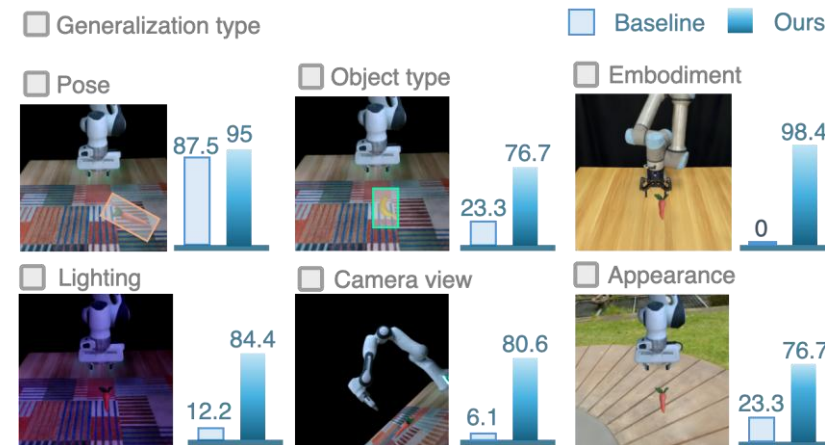
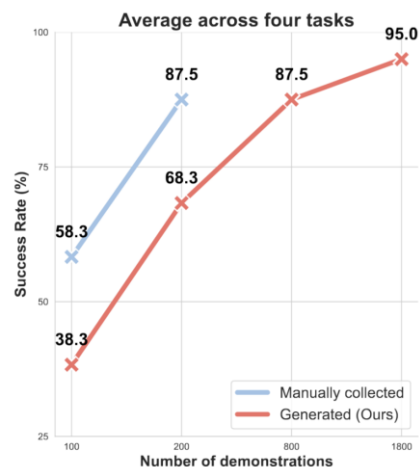
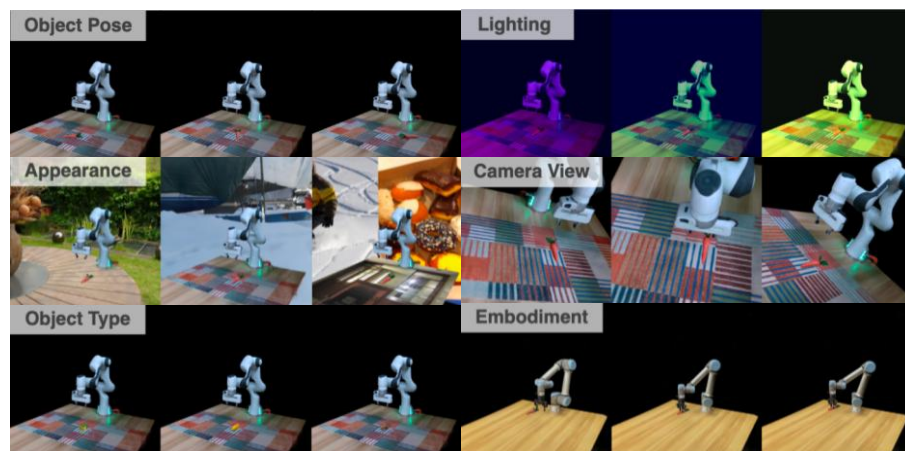
得益于3DGS的可编辑性与可渲染性, RoboSplat可以完成场景的编辑, 以完成训练数据的高效扩增。



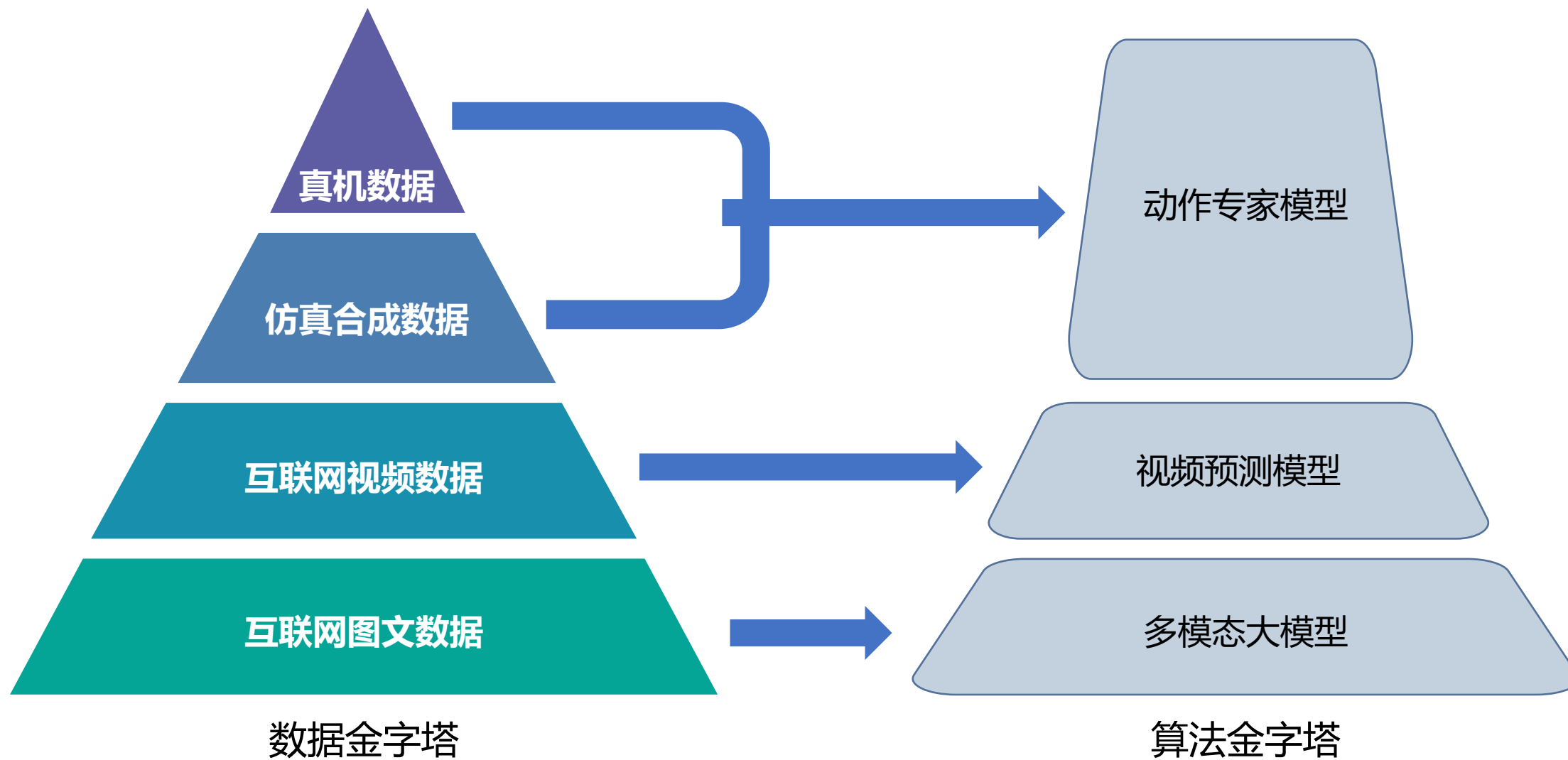
# RoboSplat : 生成式世界建模赋能的数据扩增

## 系统优势: AIGC数据扩增, 赋能可泛化具身操作

通过编辑3DGS所生成的**800**条示教数据, 与**200**条真机采集的示教数据效果相当, 且生成更多的数据, 能够实现更高的性能。  
借助基于3DGS的样本增广, 可实现**位置、物体类别、相机视角、背景、光线条件、本体**6种泛化。



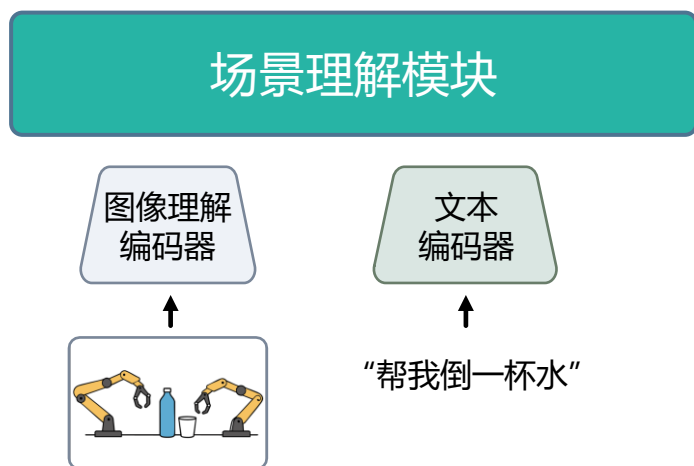
# 操作智能：InternVLA-A1



# InternVLA-A1 模型结构

## 理解、想象、执行一体化

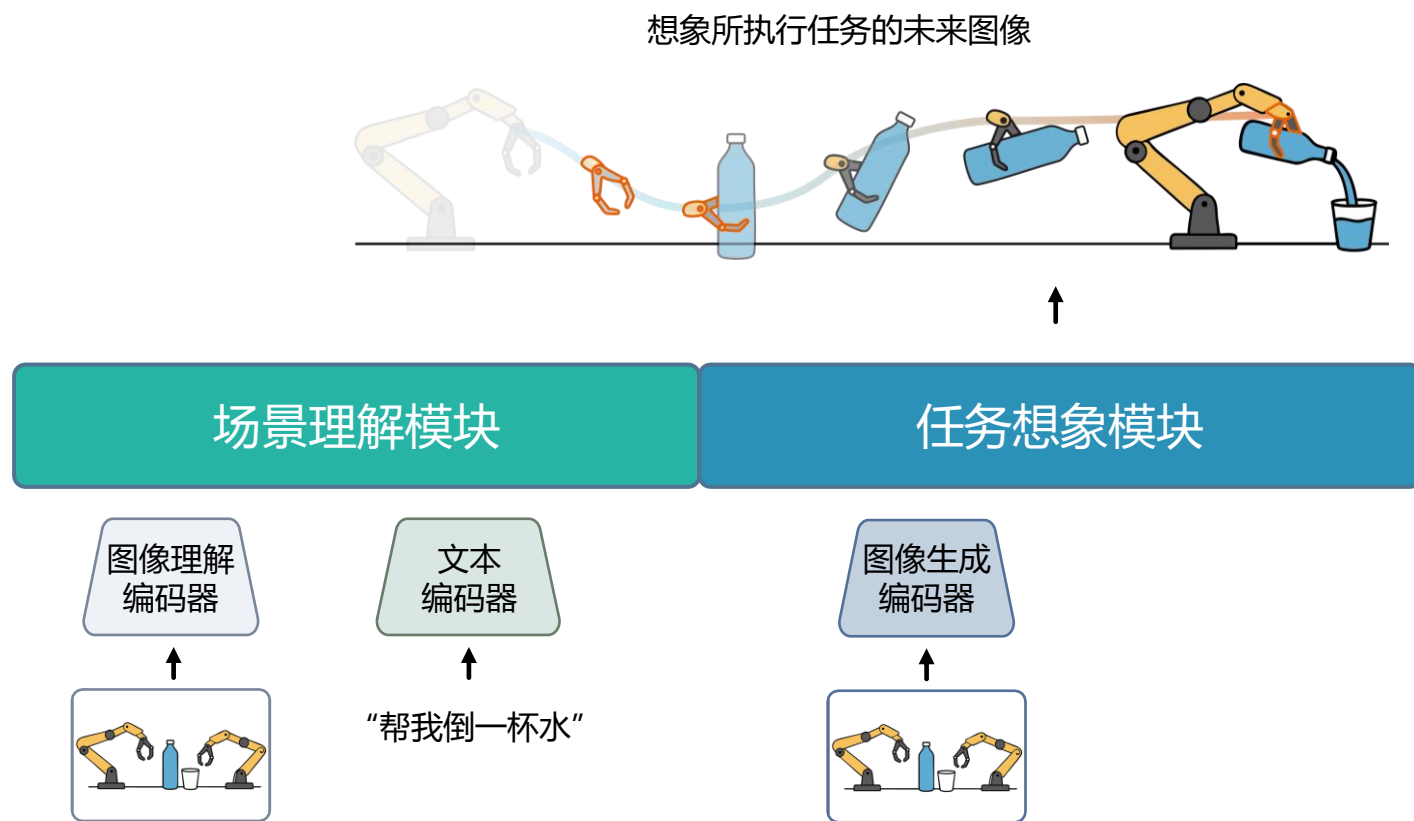
预训练的视觉-语言模型  
(如InternVL、QwenVL)



场景理解模块以图像和文本为输入，用于解析任务指令并理解任务场景

# InternVLA-A1 模型结构

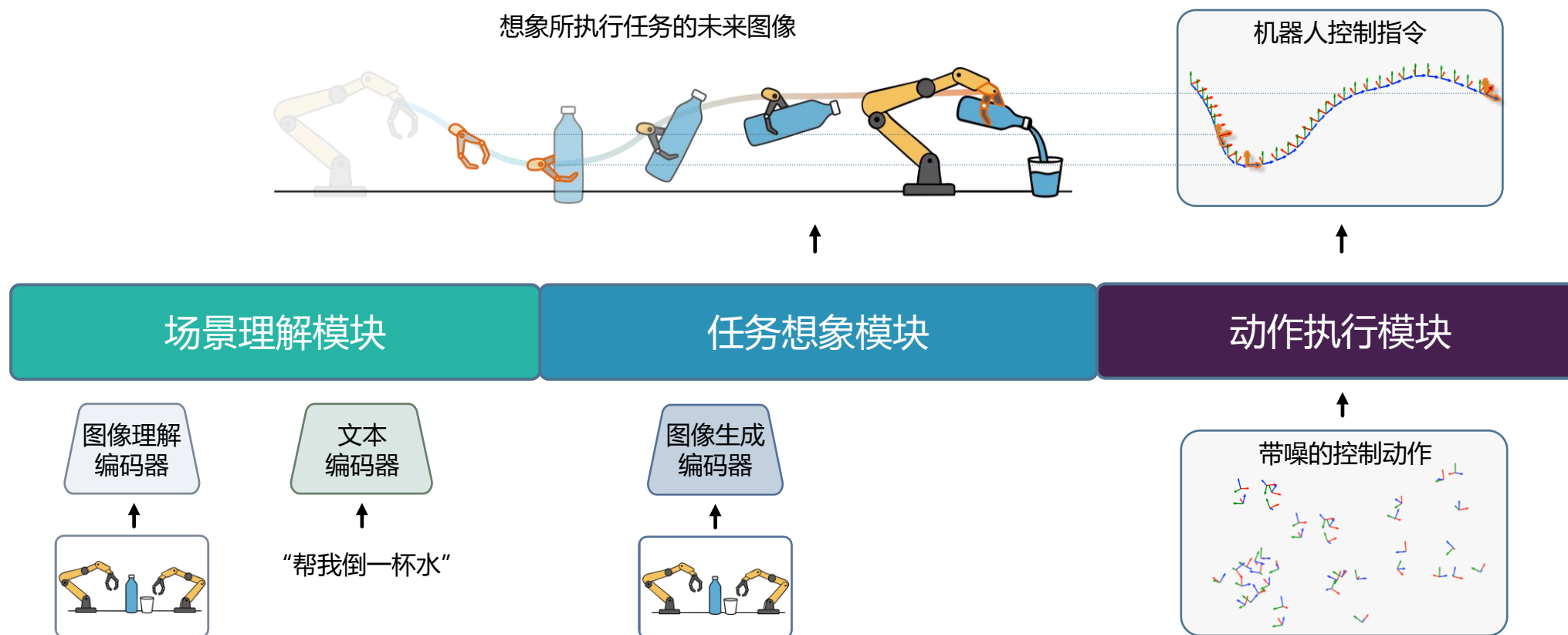
## 理解、想象、执行一体化



基于场景理解的解析结果，任务想象模块通过预测未来图像的形式，  
想象执行任务的未来演变

# InternVLA-A1 模型结构

## 理解、想象、执行一体化



最终，动作执行模块在任务想象的指导下，采用Flow Matching的方式输出机器人控制指令

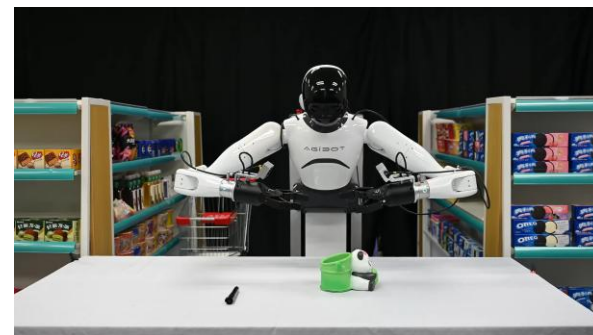
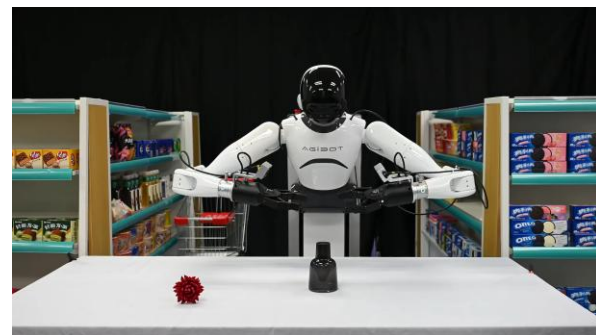
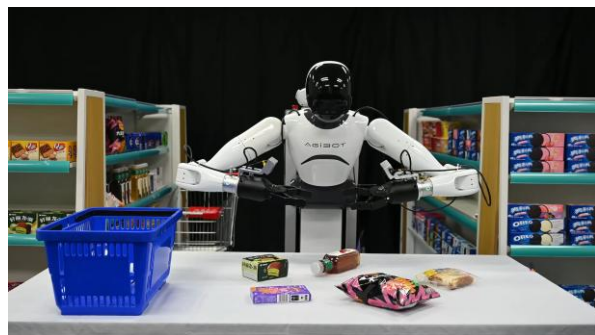
# InternVLA-A1 特点 | 一脑多形



方舟无限、松灵异构机器人协作



青龙人形机器人动态交互



智元A2D商超场景任务

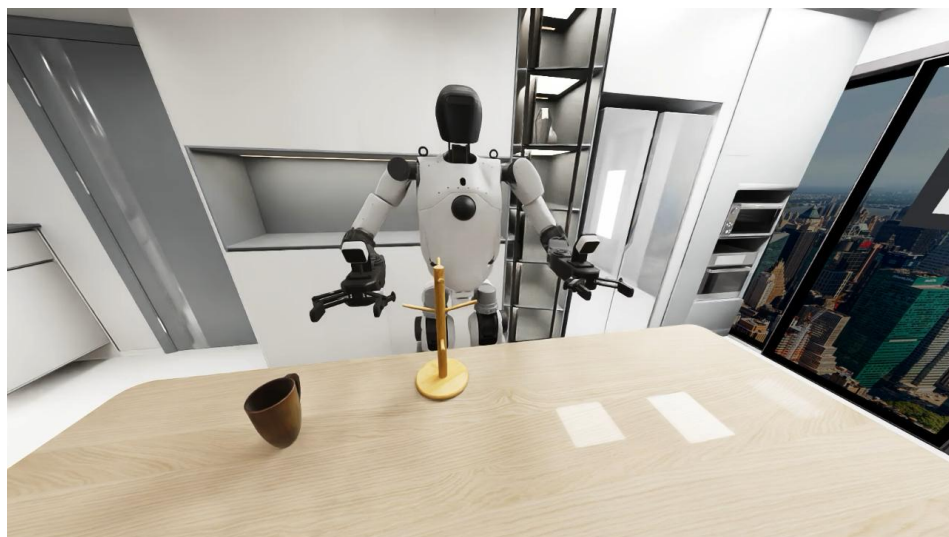
# InternVLA-A1特点 | 虚实贯通



传送带抓取



协作组装三明治

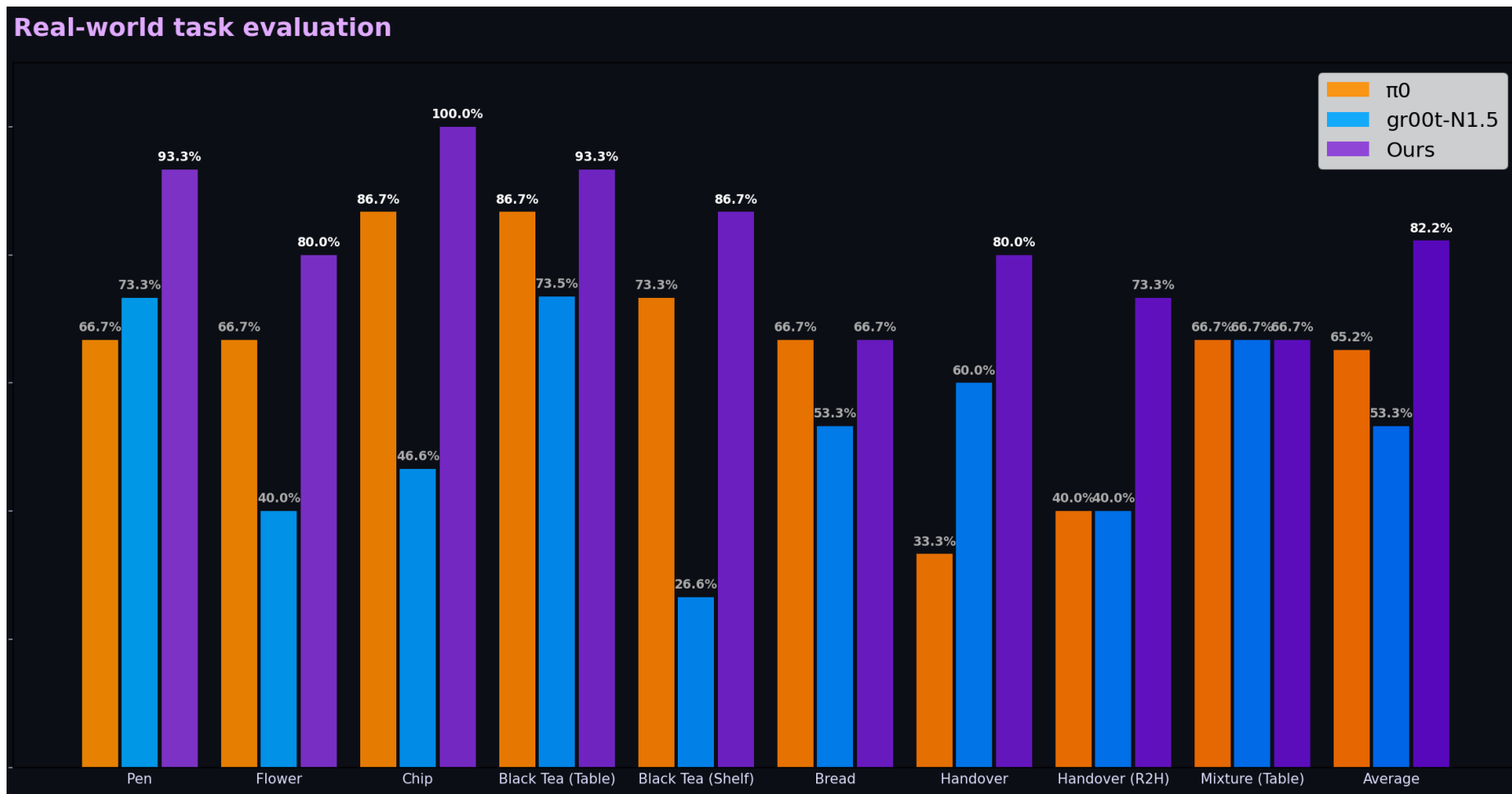


人形机器人挂马克杯



擦桌面污渍

# InternVLA-A1 | 真机与仿真评测

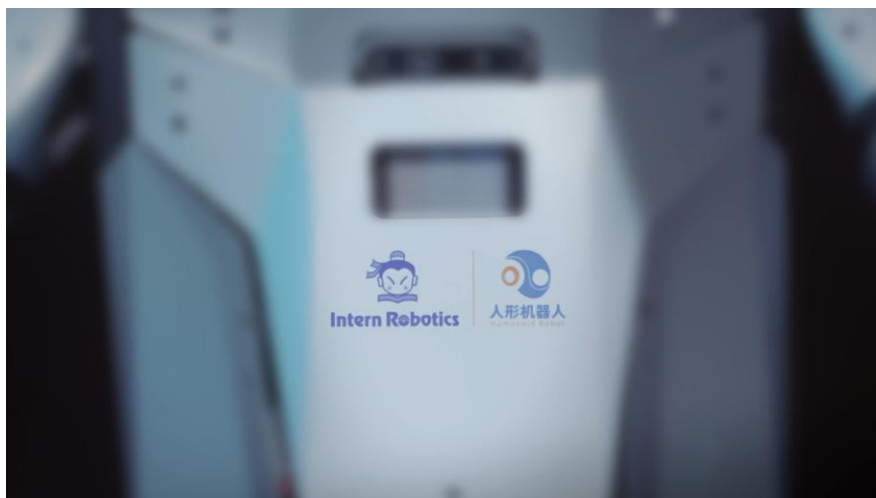


# Takeaway from InternVLA-A1

- 异构真机数据预训练，对下游任务有显著帮助

| Pretrain | LIBERO  |        |       |       | SimplerEnv Bridge |          |       |       | Avg   |
|----------|---------|--------|-------|-------|-------------------|----------|-------|-------|-------|
|          | Spatial | Object | Goal  | Long  | Carrot            | Eggplant | Spoon | Block |       |
| ✓        | 98.2%   | 97.8%  | 95.4% | 91.3% | 70.8%             | 66.7%    | 50.0% | 50.0% | 77.5% |
| ✗        | 98.0%   | 96.8%  | 94.4% | 88.4% | 0.0 %             | 16.6%    | 62.5% | 29.1% | 60.3% |

- 低成本、高效率、自动化的合成仿真数据，加速机器人训练

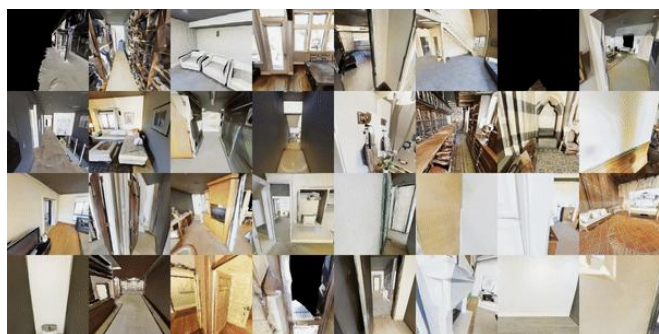


- VLA模型中嵌入一个视觉预测模块，可更有效利用互联网视频数据、仿真合成数据，实现视觉上的强泛化

# 关键科学/技术问题

## 数据

合成数据：有助于本体泛化和场景泛化



真实数据：有助于任务泛化

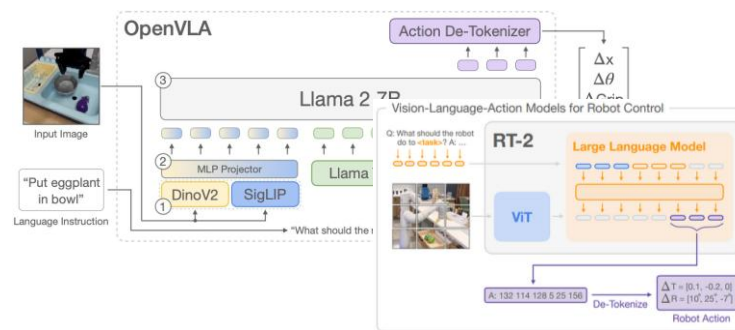


## 模型

双系统式VLA：大小脑协同

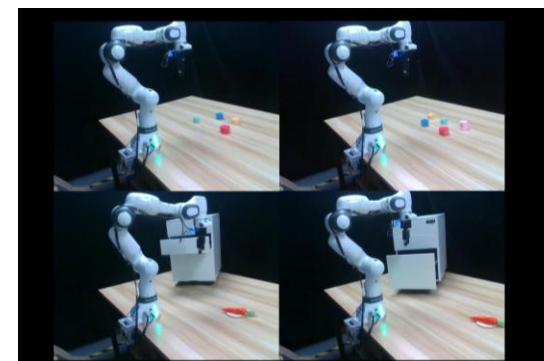


“A”式VLA

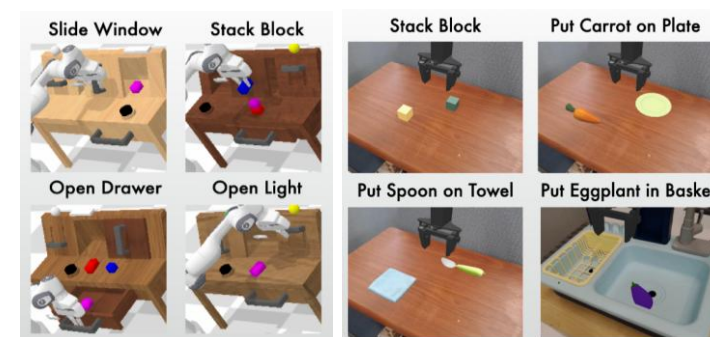


## 评测

真实评测：虚实一致但不可重复



仿真评测：可重复但虚实难以一致





上海人工智能实验室  
Shanghai Artificial Intelligence Laboratory

Intern Robotics

# Intern Robotics

—脑多形、虚实贯通、训测一体—



**GitHub**

<https://github.com/InternRobotics>



**官网**

<https://internrobotics.shlab.org.cn>



**Hugging Face**

<https://huggingface.co/InternRobotics>