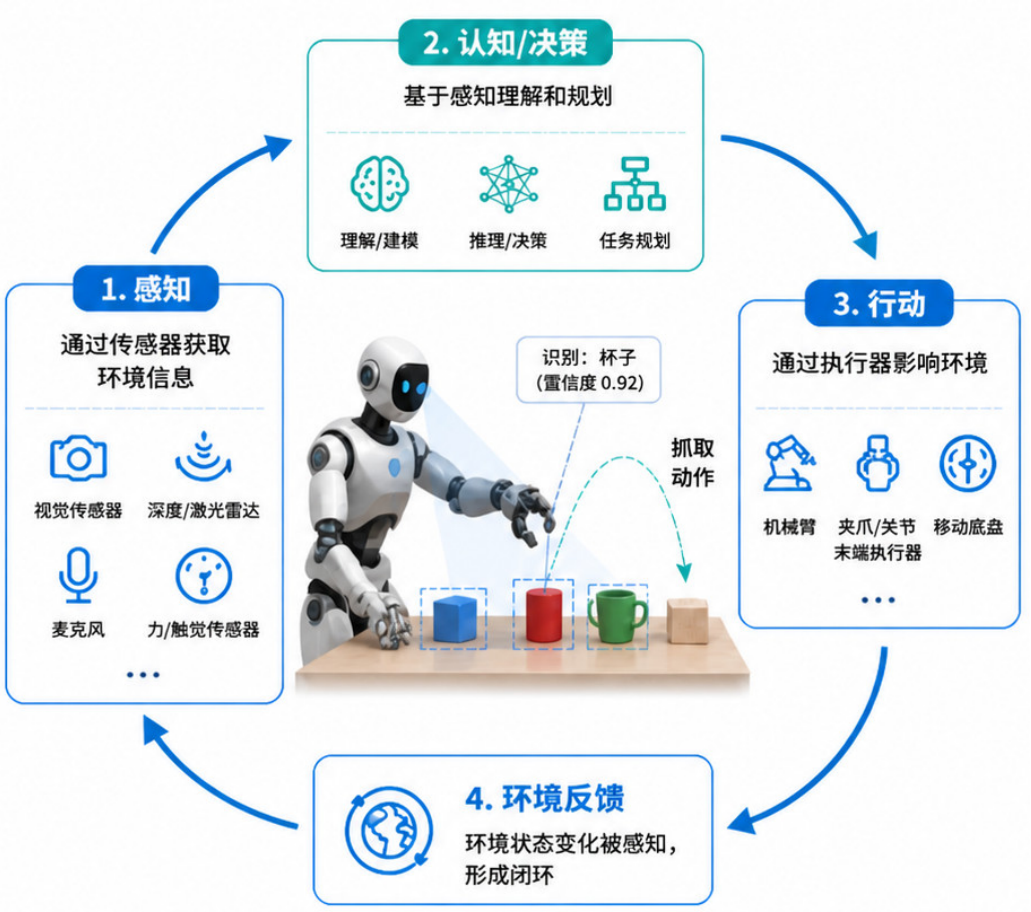
《多模态大模型原理与应用》

Lecture 10
具身智能基础理论与应用

刘阳
中山大学
人机物智能融合实验室（HCP Lab）
liuy856@mail.sysu.edu.cn

什么是具身智能



核心理念

智能源于身体与环境的交互

三大要素

感知：通过传感器获取环境信息

认知/决策：基于感知进行理解和规划

行动：通过执行器影响环境

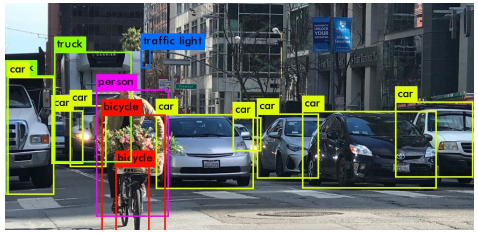
关键区分

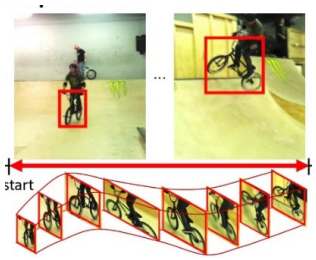
ChatGPT缺乏物理身体和环境交互闭环，

属于**离身智能**（Disembodied AI）

具身智能VS离身智能

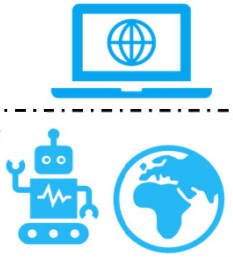
离身智能：模型**被动感知数字空间**，无法直接改变环境并作用于物理世界。





预测　跟踪　检测　分类

促进



具身智能：智能体**主动理解物理世界**，通过适应性行为和自主学习来完成任务。

视觉导航　移动操纵　具身问答　抓取　具身整理

具身智能 vs 传统AI：范式对比

对比维度	传统AI（非具身）	具身智能
信息来源	静态数据集	实时传感器流
交互模式	开环（感知→推理→输出）	闭环（感知→行动→反馈）
时间维度	离线/批处理	在线/实时流
物理约束	不考虑	核心约束（力/摩擦/碰撞）
学习机制	监督/无监督学习	交互式学习（RL/IL）
泛化来源	数据分布覆盖	物理规律与因果推断

具身智能的闭环范式意味着智能体必须持续与环境交互，从试错中学习

具身智能的研究范畴与技术体系

大脑

多模态传感器信息融合为紧凑、语义丰富的世界表征

小脑

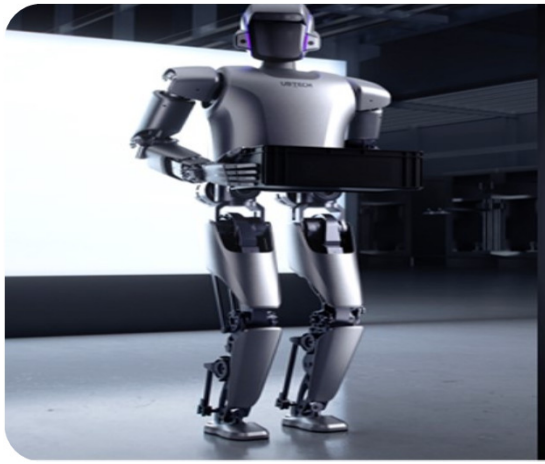
高维连续动作空间的多模态分布建模与生成

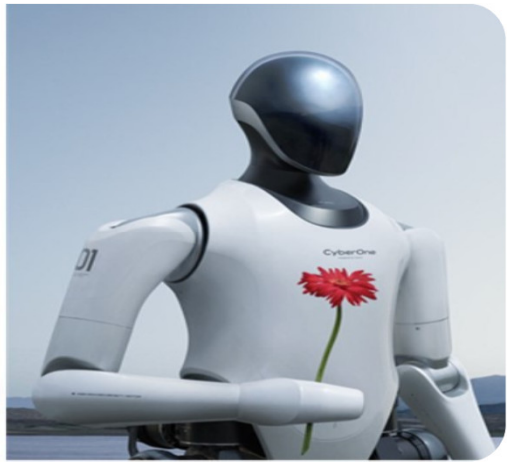
本体

仿真到现实的迁移学习与域适应方法

环境

大模型与机器人系统的深度融合架构





发展历程（上）：从哲学到机器人学

1991年《具身心智》

Varela、Thompson、Rosch合著，标志具身认知理论正式诞生

MIT Rodney Brooks

提出"基于行为的机器人学"

论文《Elephants Don't Play Chess》批评符号AI局限

核心理念

智能无需复杂内部表征

可通过身体与环境的直接交互涌现

"情境化行动"与"涌现智能"

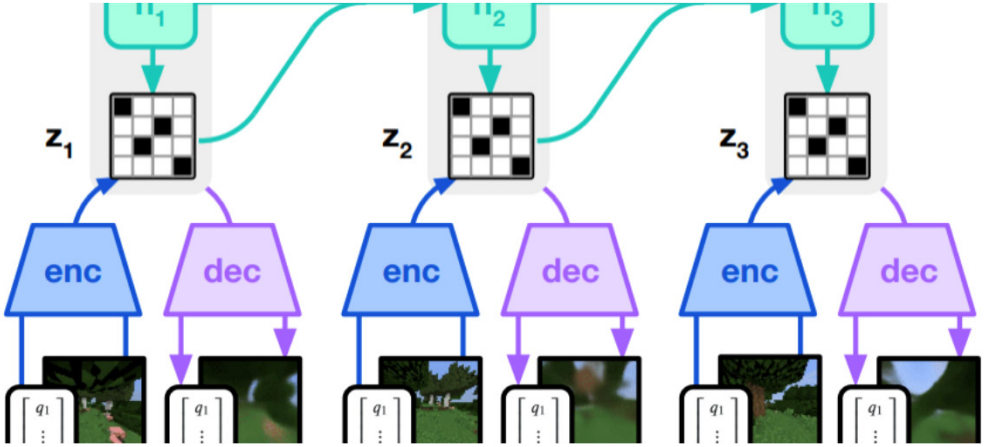
发展历程（中）：深度学习时代

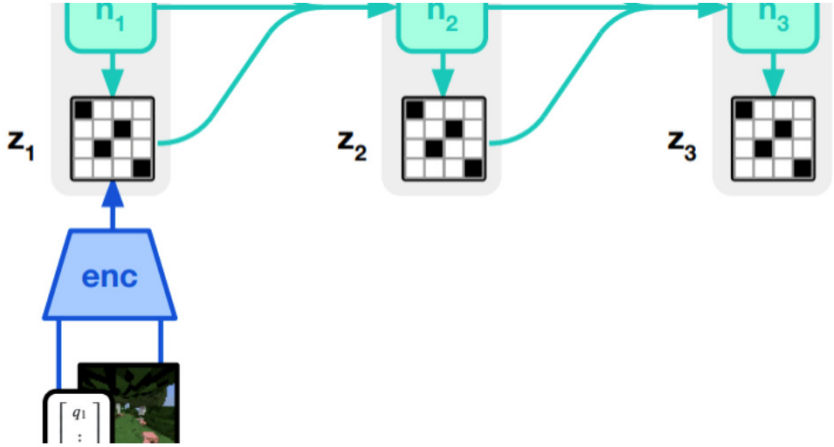
2016-2019 Sim2Real方法论奠定

1986年Brooks提出包容式架构，1990年代早期仿生机器人出现，受昆虫行为启发。

2018-2023 世界模型快速发展

2000年代ASIMO、BigDog等里程碑系统出现，强化学习开始应用于机器人控制。





发展历程（下）：大模型时代（2023-2026）

2023 基础模型元年 → 2024 VLA爆发年 → 2025-2026 产业落地加速

2023 基础模型元年

- RT-1：Transformer扩展到机器人控制
- PaLM-E：多模态推理与机器人控制统一

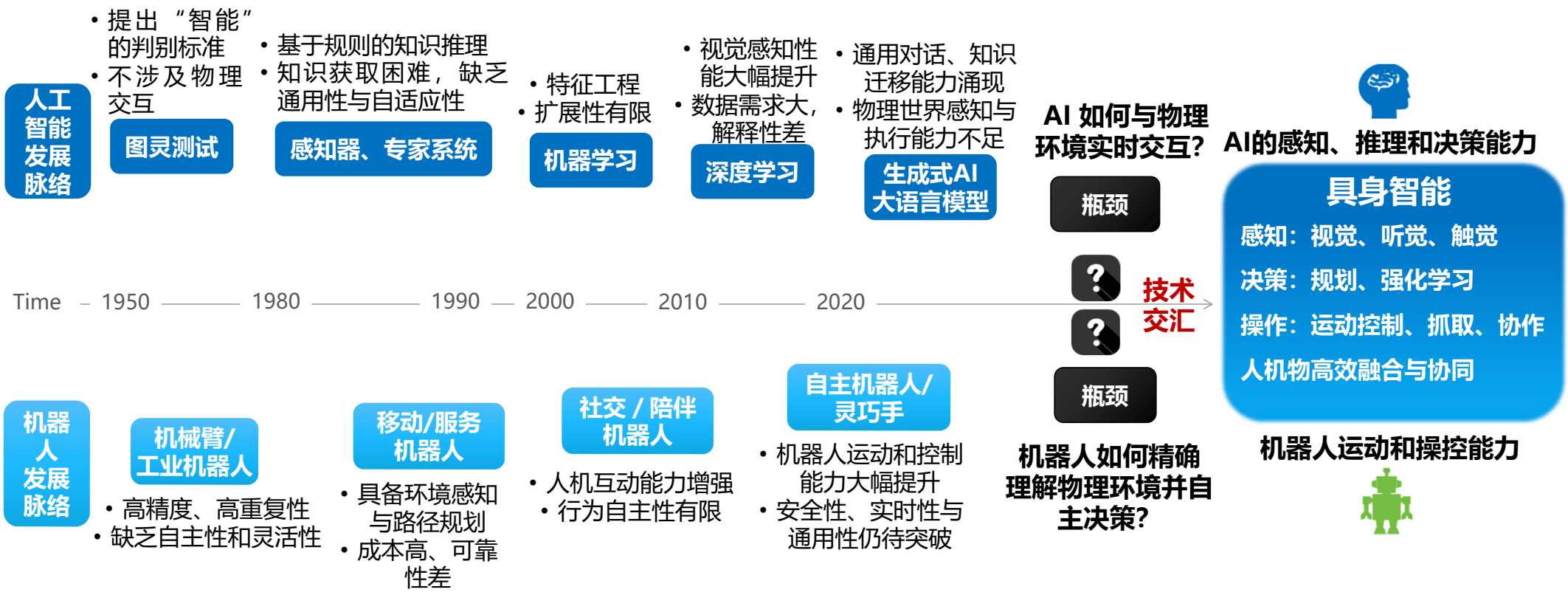
2024 VLA爆发年

2022+：GPT时代，大模型与机器人结合，VLA模型成为新范式。

2025-2026 产业落地

- Tesla Optimus 工厂部署
- Figure 02 宝马产线商用
- VLA+世界模型融合新前沿

具身智能：AI与机器人技术发展的交汇点



具身智能：AI 感知、推理、决策 与机器人物理操控执行能力的深度融合。

具身智能的六大核心科学问题

01 感知融合与表征学习

多模态传感器信息融合为紧凑、语义丰富的表征

02 动作表征与生成

高维连续动作空间的多模态分布建模

03 仿真到现实迁移

同一策略适用于不同形态的机器人

04 样本高效学习

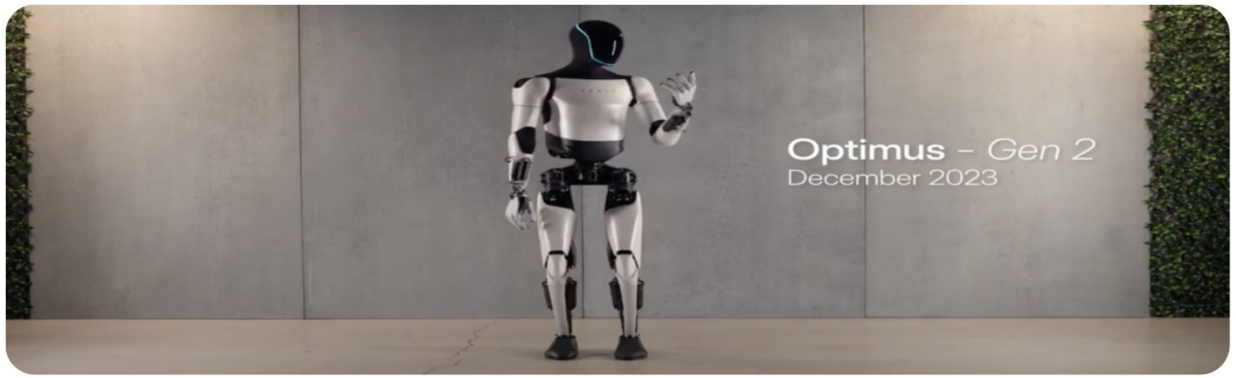
高效获取与利用机器人交互数据

05 安全与对齐

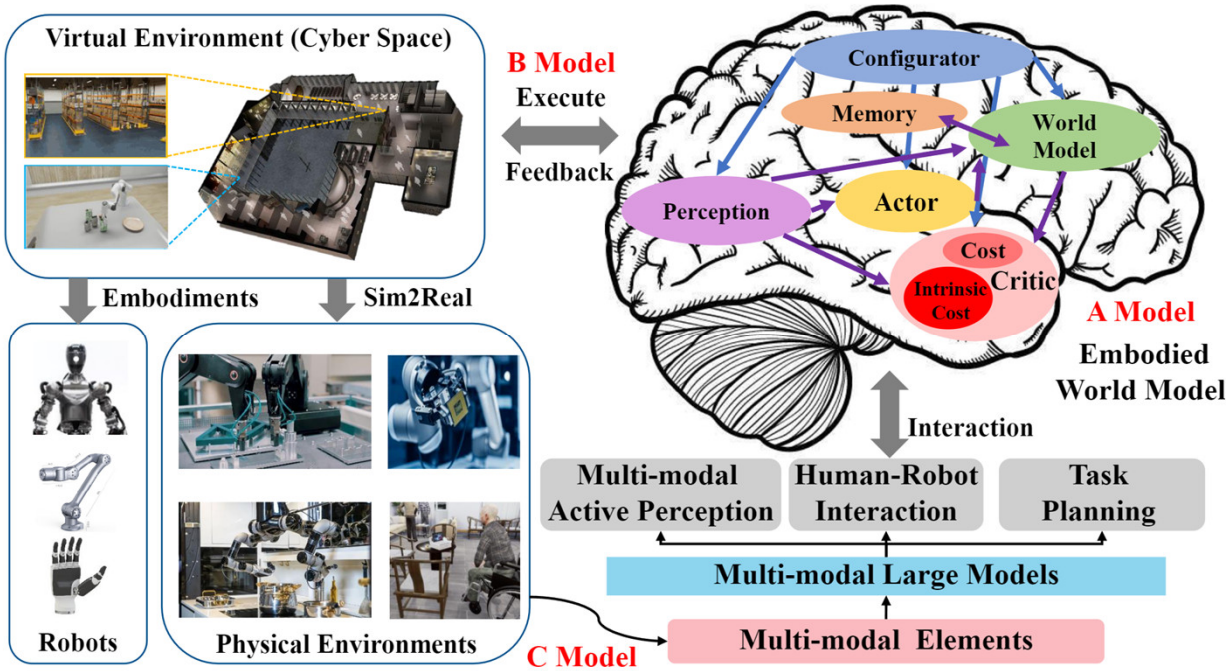
弥合仿真与现实的"现实鸿沟"

06 可解释性与泛化

力控安全、对抗鲁棒性与可解释性



具身智能的体系架构



具身世界模型架构：认知规划与物理模拟的高效协同

Sun Yat-sen University, Pengcheng Laboratory, China

Yang Liu, Weixing Chen, Yongjie Bai, Xiaodan Liang, Guanbin Li, Wen Gao, and Liang Lin. "Aligning cyber space with physical world: A comprehensive survey on embodied ai." IEEE/ASME Transactions on Mechatronics, 2025.

https://github.com/HCPLab-SYSU/Embodied_AI_Paper_List

本节内容

CONTENTS

认知革命的两次范式转移

第一次：认知革命（1956-1970s）

具身认知革命：智能不仅是大脑计算，更依赖身体与环境的动态交互。

第二次：具身转向（1980s-1990s）

- 《The Embodied Mind》(1991)集大成
- 认知是**身体行动与环境耦合**的动态过程
- 没有身体的感知运动经验，抽象概念失去根基

具身认知为AI研究提供了全新视角：智能源于交互。

《具身心智》：4E认知框架

E 具身的 Embodied

认知依赖身体的特定形态和感知运动能力

E 嵌入的 Embedded

认知活动发生在特定环境背景中

E 延展的 Extended

认知可延伸到外部工具和环境结构

E 生成的 Enactive

认知是行动生成意义的过程

实验证据

- Botvinick和Cohen（1998）的"**橡胶手错觉**"证明身体边界是可塑的认知建构
- Clark和Chalmers（1998）的"延展心智"论文以阿尔茨海默病患者使用笔记本记事为例
- 改变身体（如使用工具扩展可达范围）会改变认知本身

生成论：认知即行动

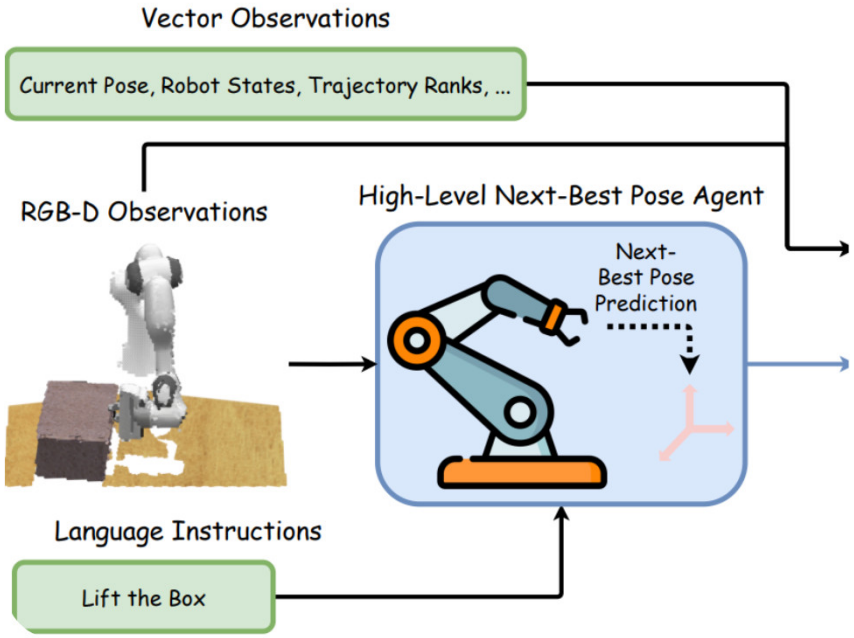
> *"认知不是表征预先给定的世界，而是生成一个世界"*
>
> — Varela

生成论核心命题

- 感知和行动是不可分割的同一过程的两面
- 知觉不是被动接收，而是主动的"知觉-行动耦合"

技术关联

- 端到端VLA模型直接从视觉+语言映射到动作
- 避免了显式中间表征的信息损失和延迟
- Diffusion Policy和π0是"生成论"的技术实现



Clark的"延展心智"与预测加工

"对等原则"（Parity Principle）

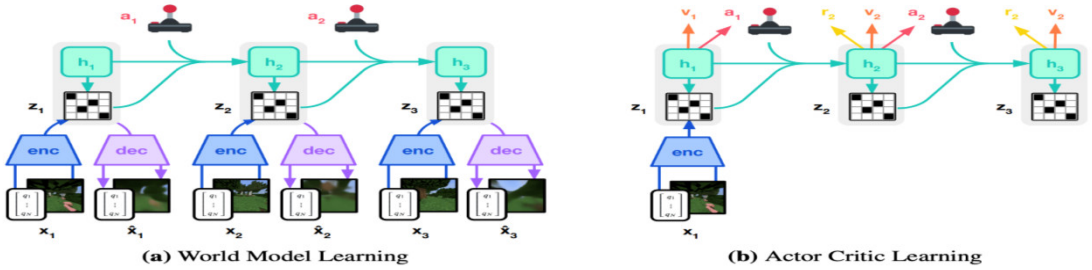
如果外部过程在功能上等同于内部认知过程，那么它就应当被视为认知的一部分。

挑战了"认知必须发生在大颅骨内部"的直觉。

预测加工框架（Predictive Processing）

- 大脑本质是**层级化的预测机器**
- 不断生成关于感官输入的预测
- 通过最小化预测误差更新内部模型

智能手机、笔记、计算工具都是认知过程的延伸。



(a) World Model Learning

(b) Actor Critic Learning

Brooks：情境机器人学与包容式架构

Brooks的包容式架构

- 分层控制：反射层→行为层→规划层
- 无需世界模型，直接感知-行动映射
- 实时响应环境变化
- 奠定了行为机器人学基础

反思与当代关联：包容式架构在简单任务上非常有效，但难以处理长期规划。2024年的VLA模型可以视为这种结合的当代形态——**端到端反应式动作生成 + 大模型驱动的语义理解。**

从哲学到工程：具身认知如何影响AI设计

1 哲学思想指导工程实践：交互式智能、环境耦合、多模态感知。

2 **从"通用推理"到"情境化智能"**
跨本体大规模训练+互联网规模预训练

3 **从"符号表征"到"感觉运动表征"**
动作token化和连续动作生成两种技术路径

4 **从"感知-认知-行动分离"到"耦合"**
端到端VLA模型直接从视觉+语言生成动作

具身智能的动物模型

昆虫级智能

从昆虫到哺乳动物，生物运动控制为机器人提供灵感。

哺乳动物级智能

- 海马体位置细胞
- 网格细胞
- 环境的内部"认知地图"
- 对应机器人SLAM

人类级智能

- 工具使用体现身体延展
- 琴键是手指的延伸
- 车辆边界是身体的边界
- 提示：工具融合设计

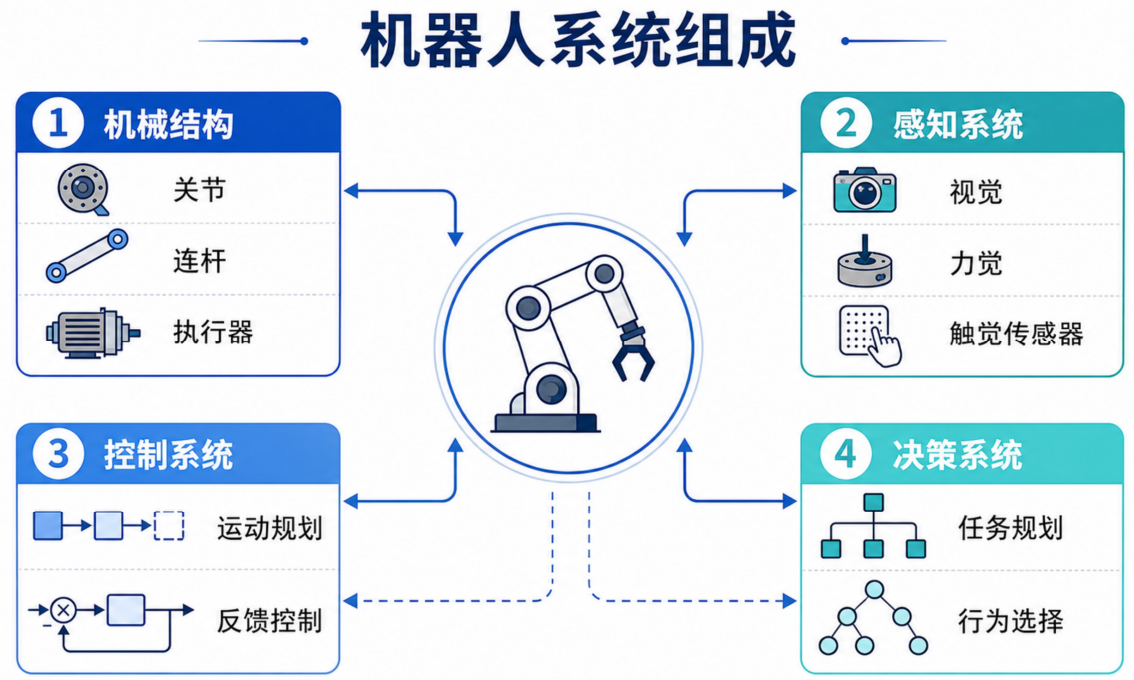
本节内容

CONTENTS

机器人系统概述

机器人系统组成

- 机械结构：关节、连杆、执行器
- 感知系统：视觉、力觉、触觉传感器
- 控制系统：运动规划、反馈控制
- 决策系统：任务规划、行为选择

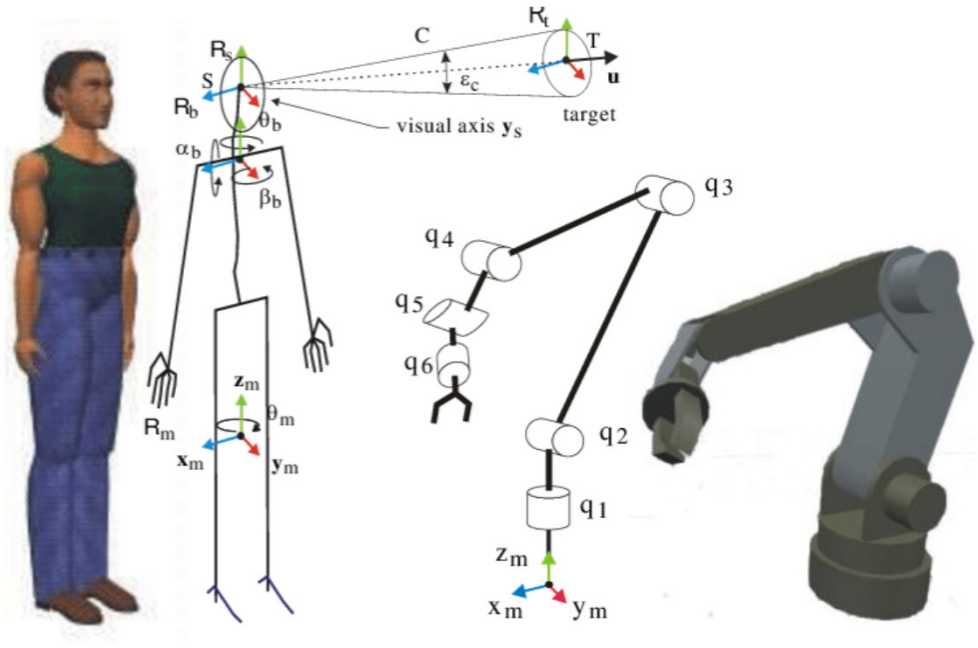


形态	代表产品	自由度	应用场景
操作型	Franka, UR5e	6-7 DoF	工业装配、精密操作
足式型	Spot, Atlas	12-30 DoF	复杂地形巡检
人形	Optimus, Figure 02	50+ DoF	通用服务、工业协作

正运动学

正运动学：已知关节角→计算末端位姿

- Denavit-Hartenberg参数法
- 齐次变换矩阵连乘
- 应用：路径规划、碰撞检测



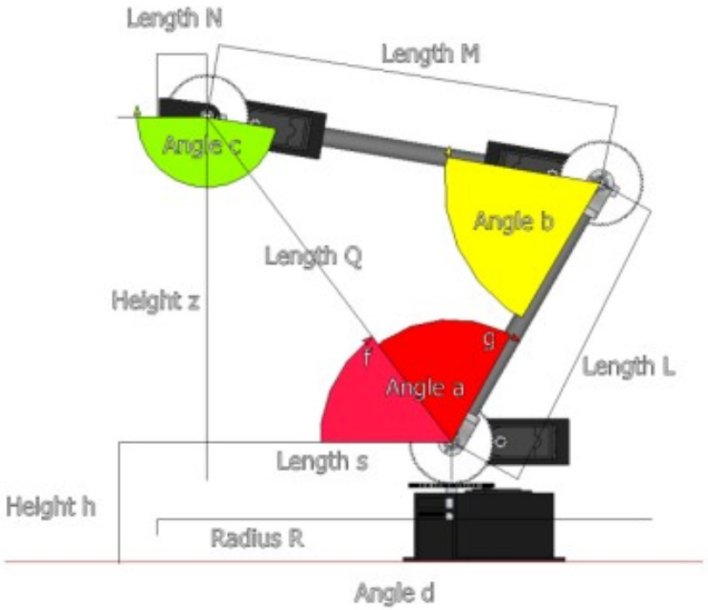
实践要点：现代仿真环境（MuJoCo、Isaac Sim）内置了自动的正逆运动学求解，研究者通常不需要手动推导D-H参数。但理解其原理对调试和自定义场景仍然重要。

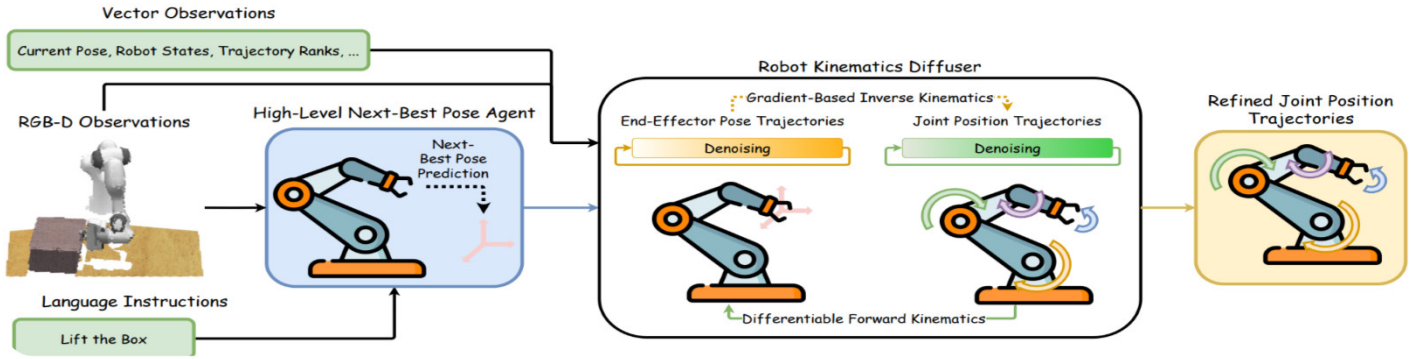
逆运动学

逆运动学：已知末端位姿→求解关节角

- 解析法：封闭解，速度快
- 数值法：迭代优化，通用性强
- 应用：抓取、操作、避障

可微分IK、学习-based IK是前沿方向。





机器人动力学与控制基础

$$M(\theta)\ddot{\theta} + C(\theta,\dot{\theta})\dot{\theta} + G(\theta) = \tau$$

PID控制

最广泛的工业控制器

比例-积分-微分反馈

简单但无法处理复杂约束

模型预测控制 MPC

有限时域优化问题

预测未来轨迹

足式机器人广泛使用

阻抗/导纳控制

控制力-位移关系

弹簧-阻尼模型

人机协作安全关键

与具身智能的关系

- 传统控制方法需要精确的机器人模型
- 具身智能通过**强化学习**学习策略，可以绕过精确建模
- 理解传统控制有助于设计奖励函数、动作空间和约束

传感器与感知系统

视觉传感器

RGB相机 · 深度相机（RealSense）

事件相机 · 全景相机

触觉传感器

力/力矩传感器

DIGIT / GelSight触觉阵列

本体感觉

关节编码器 · IMU

电流环（间接测量力矩）

其他传感器

LiDAR · GPS

麦克风（语音指令）

技术趋势：2024-2025年VLA模型主要依赖RGB视觉，但**多模态融合**（视觉+触觉+本体感觉）正在成为新方向。DreamTacVLA（2026）首次将触觉信号整合进VLA框架。

传感器融合与状态估计

卡尔曼滤波 KF

融合相机、激光雷达、触觉等多源传感器数据。

粒子滤波

卡尔曼滤波、粒子滤波处理不确定性感知。

端到端表征学习

SLAM技术实现同时定位与地图构建。

多模态感知是具身智能的核心能力，传感器融合质量直接决定系统性能上限。

动作空间设计

动作空间	描述	优点	缺点
关节空间	直接输出关节位置/速度/力矩	不需IK求解，可直接执行	不同机器人构型不同，难跨本体迁移
末端执行器空间	6-DoF位姿+夹爪开合	与本体解耦，便于跨本体泛化	需要下游IK求解器
全身体控制	基座移动+手臂操作	适用于移动操作和人形	20-50+ DoF，学习难度高

RT-2动作空间：8维（Δx, Δy, Δz, Δroll, Δpitch, Δyaw, 夹爪开合, 终止标志），每维均匀量化为256个离散bin，转化为"伪语言token"供VLM处理。

机器人任务的类型学

操作任务

抓取、放置、装配

工具使用、折叠

导航任务

点目标、物体目标

视觉语言导航VLN

移动操作

导航+操作结合

人形机器人核心场景

足式运动

双臂协调、人机协作等复杂交互任务。

接触丰富任务

移动操作、社交交互等高级任务。

机器人系统的数学抽象：MDP

马尔可夫决策过程五元组 (S, A, P, R, γ)

S 状态空间：视觉观测+本体感觉+语言指令

A 动作空间：关节/末端执行器控制

P 转移概率：物理动力学决定

R 奖励函数：评估动作好坏

γ 折扣因子：权衡即时与未来奖励

马尔可夫性质

下一状态只依赖当前状态和动作，与历史无关

POMDP：部分可观测环境需引入信念状态或历史观测

最优策略：$\pi^* = \arg\max E[\Sigma \gamma^t r_t]$ —— 最大化期望累积折扣奖励。策略 $\pi(a|s)$ 定义了在状态s下选择动作a的概率。

本节内容

CONTENTS

为什么需要仿真?

1 数据效率

Isaac Gym可并行4096个环境

训练从数月压缩到数小时

2 安全性

仿真提供"安全沙箱"

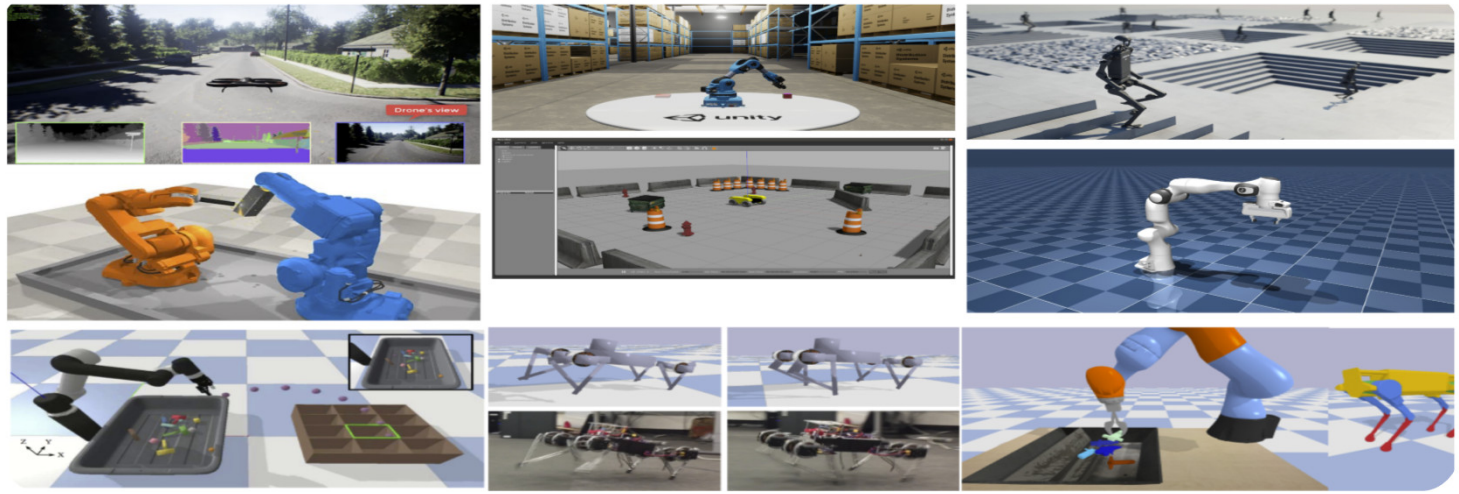
允许算法自由探索

3 可重复与可扩展

完全可重复（相同随机种子）

轻松创建多样化场景

真实机器人实验成本高、危险且耗时。



主流物理引擎概述

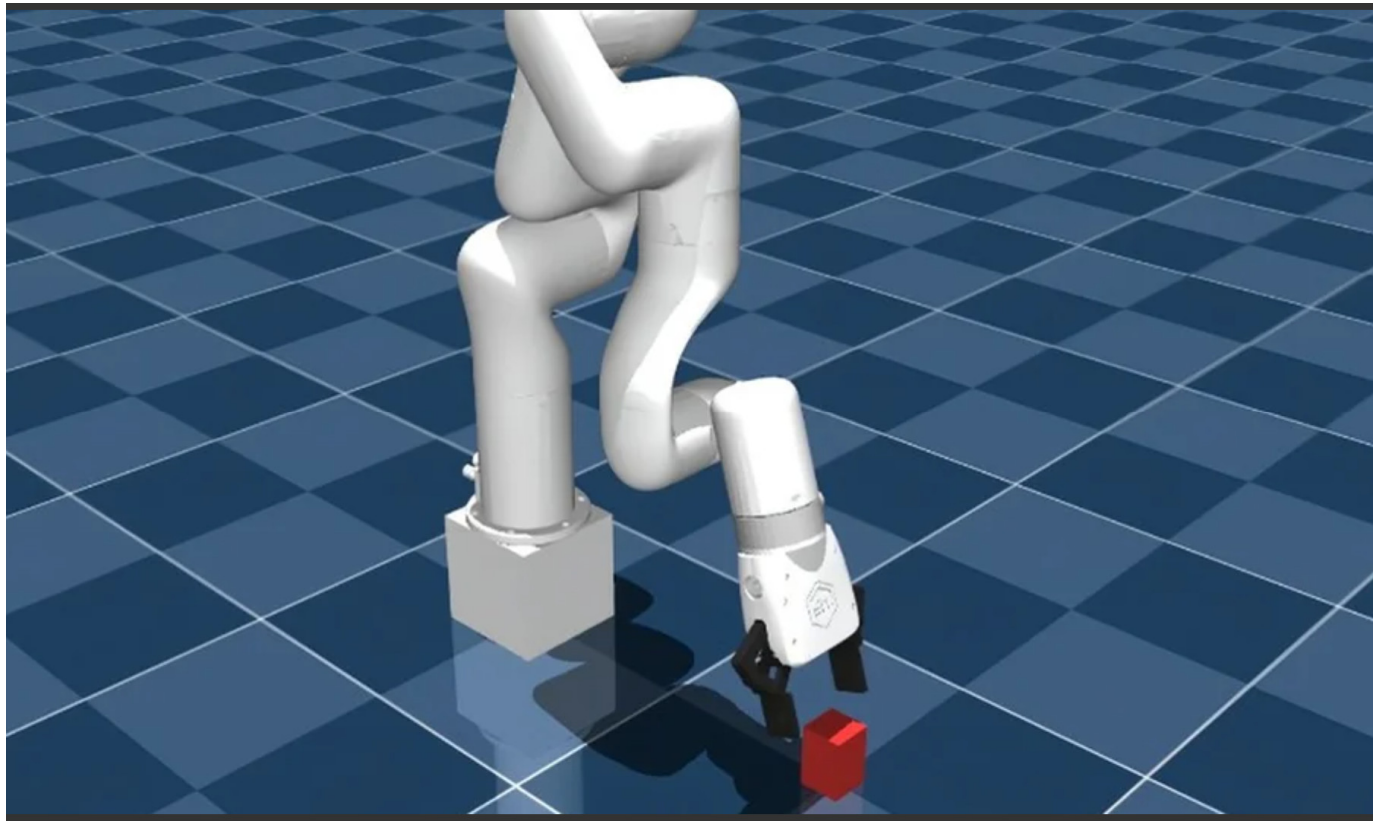
引擎	开发方	优势	适用场景
MuJoCo	DeepMind（开源）	计算精度和数值稳定性，复杂接触动力学	精度敏感任务、研究首选
PhysX	NVIDIA	GPU加速，大规模并行仿真	Isaac Sim/Gym/Lab基础
Bullet	开源	轻量易用，Python接口简单	快速原型开发、教育
其他	ODE/Havok/Chrono	各有专精领域	特定应用场景

选择原则：没有"最好的"物理引擎，只有"最适合的"。精度敏感→MuJoCo；大规模并行→PhysX；快速原型→Bullet；车辆动力学→Chrono。

MuJoCo与DeepMind Control Suite

MuJoCo：接触物理仿真引擎

- 由DeepMind开源，免费使用
- 精确接触动力学仿真
- 支持强化学习训练
- Python接口，易于集成



NVIDIA Isaac生态

Isaac Gym (2021)

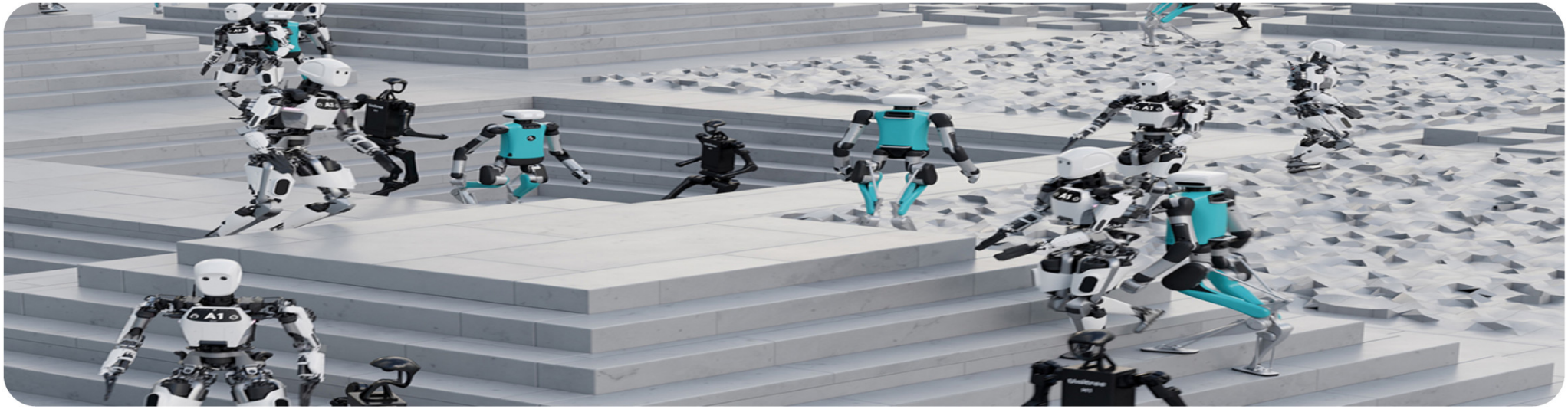
Isaac Sim：NVIDIA高保真GPU加速仿真平台。

Isaac Sim (2022+)

Isaac Gym：并行RL训练环境，支持数千个环境同时训练。

Isaac Lab (2024)

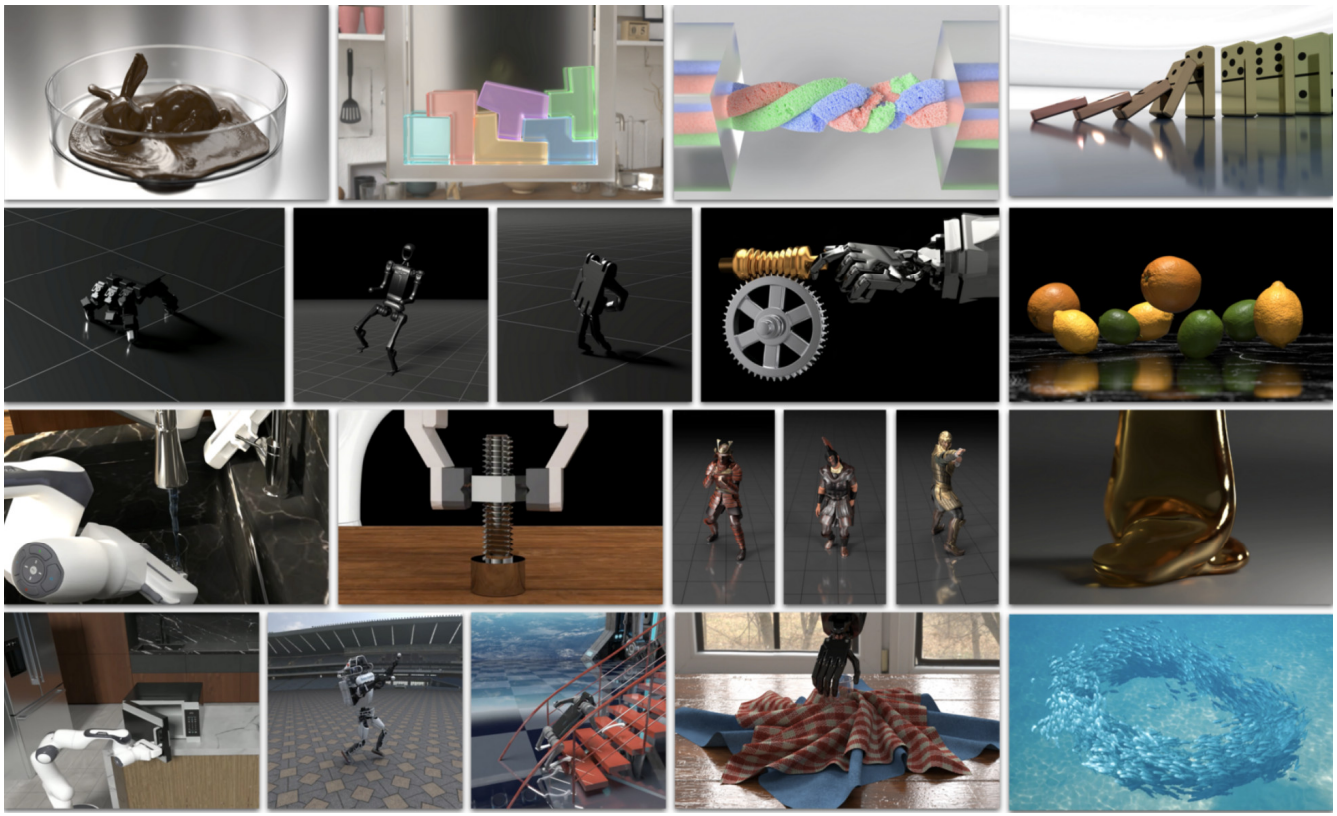
Isaac Lab：统一框架，整合Sim与Gym功能。



Genesis：通用物理引擎新势力

Genesis：新一代物理仿真平台

- 支持多种物理引擎后端
- 照片级真实感渲染
- 与生成式AI深度集成
- 开源且易于扩展



SAPIEN与ManiSkill：操作任务专用平台

SAPIEN

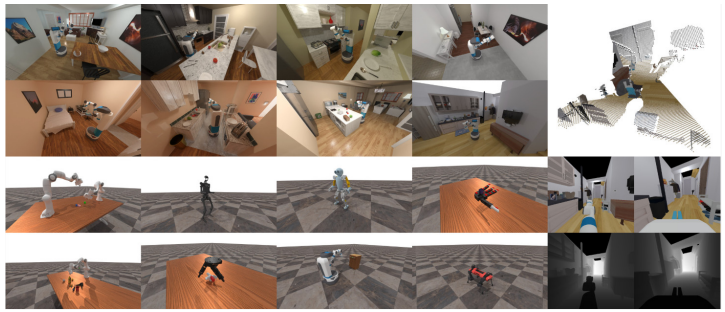
UC San Diego开发
零件级物理仿真
（抽屉、柜门等活动部件）
GPU并行渲染

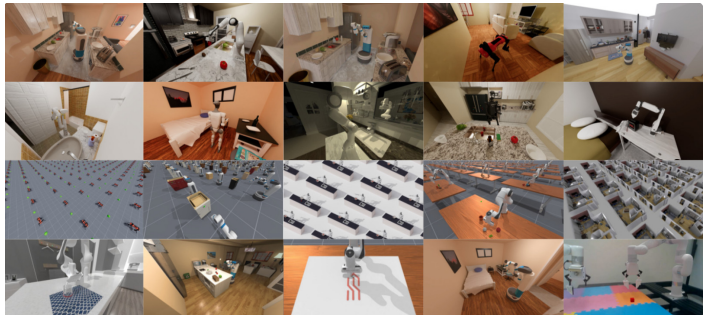
ManiSkill3 (2024)

数十个操作任务
30000+ FPS数据收集
GPU并行
视觉/点云观测

RLBench

Imperial College London
100+操作任务标准化
Franka机械臂
VLA模型标准评估平台





仿真环境对比与选择指南

特性	MuJoCo	Isaac Sim	Isaac Gym	Genesis	SAPIEN	RLBench
并行能力	中等	高	极高	极高	高	中
渲染质量	基础	照片级	基础	照片级	高	中
接触精度	极高	高	高	高	高	高
软体/流体	有限	支持	有限	全面	有限	有限
学习曲线	平缓	陡峭	中等	中等	平缓	平缓
开源	是	是	是	是	是	是

选择建议：入门→MuJoCo+DMC | 大规模并行→Isaac Gym/Genesis | 视觉策略→Isaac Sim | 操作任务→ManiSkill3/RLBench | 可微物理→Genesis

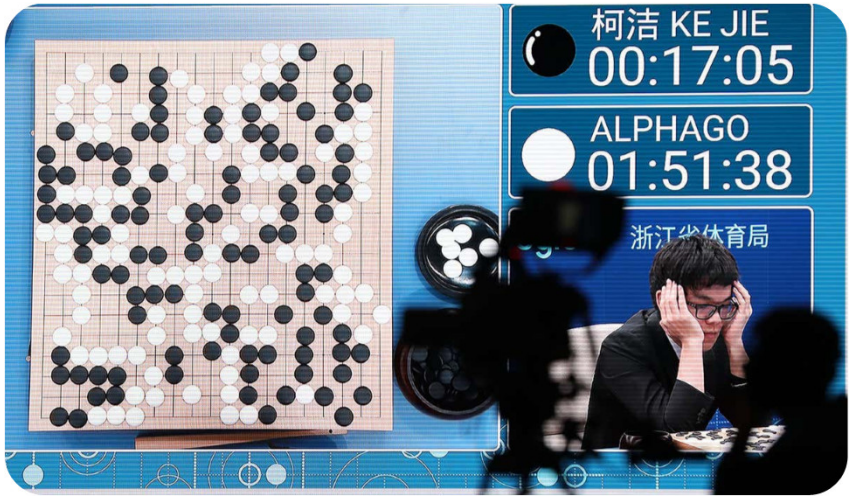
本节内容

CONTENTS

强化学习概述

强化学习核心框架

- 智能体在环境中通过试错学习
- 目标：最大化累积奖励
- 关键挑战：探索与利用平衡
- 在机器人控制中广泛应用



从Q-Learning(1989)到PPO(2017)再到SAC(2018)。

值函数与贝尔曼方程

状态值函数 $V^{\pi}(s)$

从状态s出发，遵循策略π的期望累积折扣奖励

动作值函数 $Q^{\pi}(s,a)$

从状态s执行动作a后再遵循π的期望累积折扣奖励

贝尔曼方程：$Q^{\pi}(s,a) = R(s,a) + \gamma E[V^{\pi}(s')]$

贝尔曼最优方程：$Q^{*}(s,a) = R(s,a) + \gamma E[\max Q^{*}(s',a')]$

核心洞察：贝尔曼方程的递归结构意味着——只要知道下一步的最优值，就能计算当前步的最优值。这为值迭代、Q-Learning等算法提供了理论基础。

策略梯度与Actor-Critic架构

策略梯度定理：$\nabla J(\theta) = E[\nabla \log \pi_\theta(a|s) \cdot A^\pi(s,a)]$

策略梯度定理

- 直接优化策略参数θ
- 目标：最大化期望累积奖励
- 梯度估计：$\nabla J = E[\nabla \log \pi(a|s) \cdot Q(s,a)]$

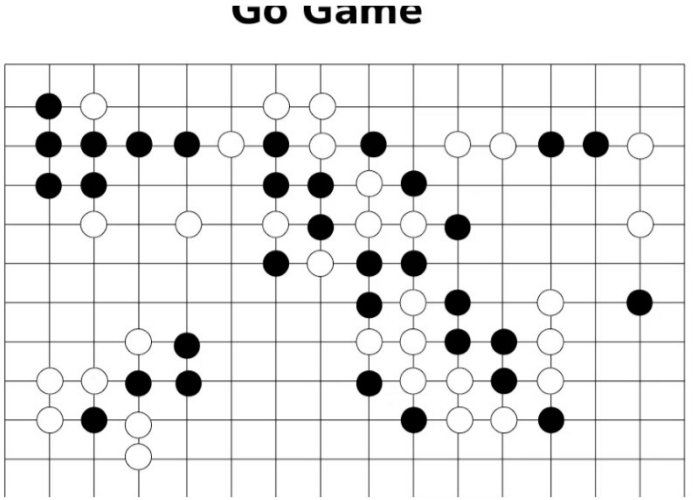
技术演进

- REINFORCE：蒙特卡洛梯度估计
- A2C/A3C：引入Critic减少方差
- TRU：信任区域更新稳定训练

PPO算法详解（上）

PPO：近端策略优化

- 裁剪目标函数限制策略更新幅度
- 平衡样本效率与训练稳定性
- OpenAI默认RL算法
- 实现简单，超参数鲁棒



$$L^{CLIP}(\theta) = E[\min(r_t \cdot \hat{A}_t, \text{clip}(r_t, 1-\varepsilon, 1+\varepsilon) \cdot \hat{A}_t)]$$

直观理解：当$r_t > 1+\varepsilon$时，如果$\hat{A}_t > 0$（好动作），目标取裁剪值阻止进一步增大；如果$\hat{A}_t < 0$（坏动作），不裁剪允许降低选择概率。不对称裁剪确保策略不会突变。

PPO算法详解（下）

训练流程

Step 1 — 数据采集：使用当前策略运行N个episode，收集轨迹

Step 2 — 优势估计：使用GAE计算每个时间步的优势值

Step 3 — 策略更新：K轮小批量梯度上升优化裁剪目标

Step 4 — 值函数更新：最小化值预测误差

优缺点

实现简单、超参数鲁棒、样本效率较高。

对连续动作空间支持有限，超参数敏感。

SAC算法详解

$$J(\pi) = \Sigma E[r(s_t,a_t) + \alpha H(\pi(\cdot|s_t))]$$

SAC：最大熵强化学习

- 同时学习Q函数和策略
- 引入熵正则化鼓励探索
- 自动调整温度参数α
- 连续动作空间的SOTA算法

维度	PPO	SAC
策略类型	On-policy	Off-policy
样本效率	较低	较高
探索机制	随机噪声	熵最大化
实现复杂度	简单	中等
典型应用	大规模并行	样本受限

实践建议：仿真环境充足→PPO；真实机器人/样本受限→SAC。SAC的回放缓冲区机制使其在真实机器人上更高效。

RL在机器人中的应用与挑战

1 **样本效率**

→ Sim2Real迁移 / 演示初始化

2 真实环境交互成本高、危险。

3 **安全性**

→ Safe RL约束优化（CPO、SDA）

4 仿真到现实迁移存在域差距。

从RL到模仿学习：为什么需要演示

模仿学习方法

- **行为克隆（BC）**：将演示视为监督学习
- **DAgger**：迭代收集专家标注
- **RL from Demonstrations**：演示初始化+RL微调

VLA模型本质上是**大规模模仿学习**

2023-2024 数据革命

Open X-Embodiment（100万+轨迹） · DROID · BridgeData V2

本节内容

CONTENTS

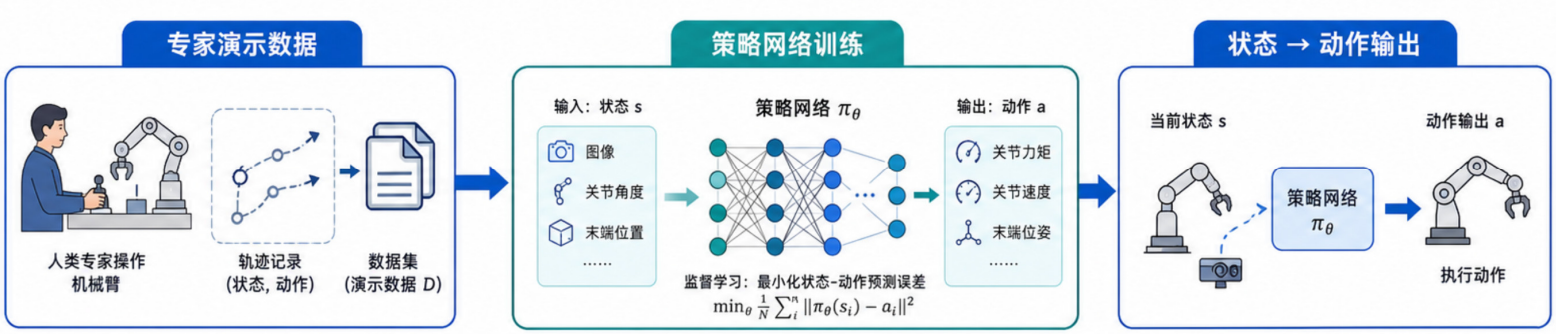
行为克隆：模仿学习的基础

$$L_{BC} = E_{(s,a)\sim D} \|\pi_\theta(s) - a\|^2$$

行为克隆：监督学习模仿专家

- 收集专家演示数据
- 训练策略网络拟合状态-动作映射
- 简单高效但存在复合误差
- DAgger通过迭代收集改善

DAgger迭代收集训练数据，缓解分布偏移。



大规模机器人数据集的崛起

数据集	规模	特点	用途
Open X-Embodiment	100万+轨迹	21机构60数据集，22种机器人	RT-X, Octo, OpenVLA训练
DROID	8万条	多样性'野外'操作数据	泛化能力评估
BridgeData V2	6万条	低成本双臂（<5000美元）	低成本数据收集示范
AgiBot-World	百万级	国产，人在回路验证	通用操作策略训练

意义：这些数据集使训练通用机器人策略成为可能。数据多样性（跨任务、跨环境、跨本体）可能比数据总量更重要——Figure AI的CEO称之为"数据飞轮"。

Diffusion Policy：动机与核心思想

行为克隆输出单峰分布，无法表达多模态行为。扩散模型可建模复杂多峰分布。

解决方案

将动作生成建模为条件扩散过程：从随机噪声出发，迭代去噪生成符合专家分布的动作。

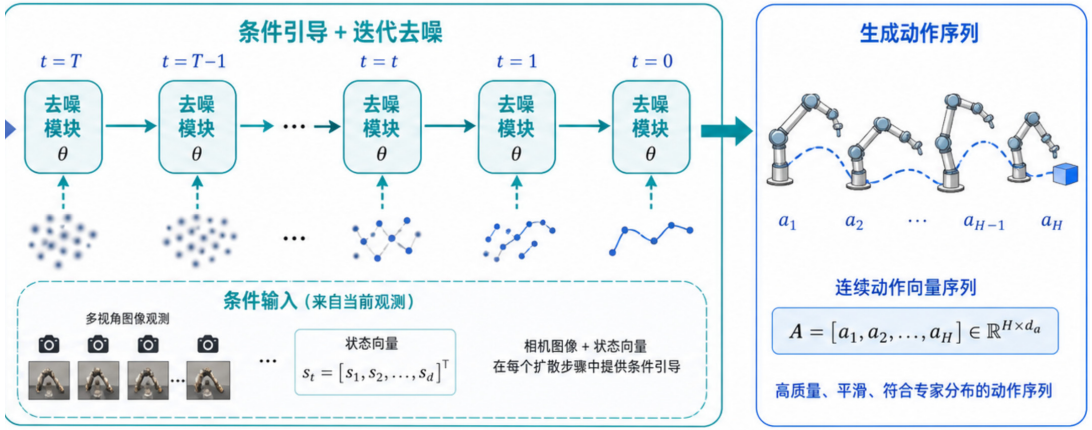
扩散模型数学基础

前向过程：逐步向动作添加高斯噪声

$$q(a^k|a^{k-1}) = N(a^k; \sqrt{(1-\beta_k)}a^{k-1}, \beta_k I)$$

反向过程：训练网络预测噪声，恢复动作

$$p_\theta(a^{k-1}|a^k,s) = N(a^{k-1}; \mu_\theta, \Sigma_k)$$



Diffusion Policy：架构与训练

观测编码器

CNN/ViT提取图像特征

MLP处理低维向量

融合为条件向量c

噪声预测网络

U-Net/Transformer架构

时间步嵌入注入

预测噪声ε_θ

动作采样

从纯噪声开始

迭代去噪

获得干净动作a^0

关键超参数

- 扩散步数K：10-100（步数越多质量越高但推理越慢）
- 噪声调度：cosine schedule最常用
- 动作序列长度（Action Horizon）：预测未来多步动作以提升时间一致性

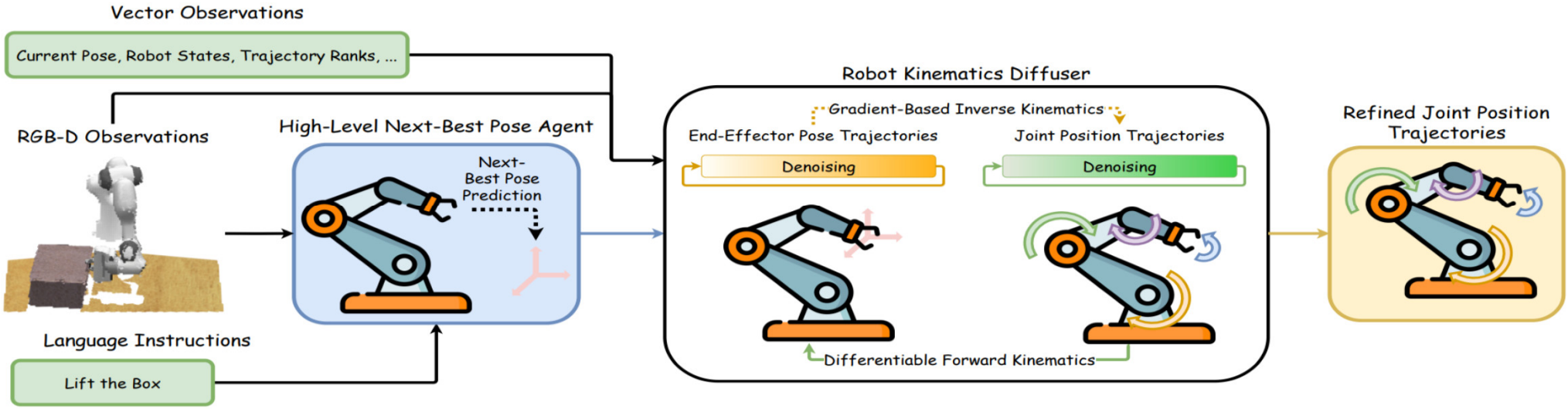
Diffusion Policy：优势与局限

优势

表达多模态行为、训练稳定、泛化能力强。

局限

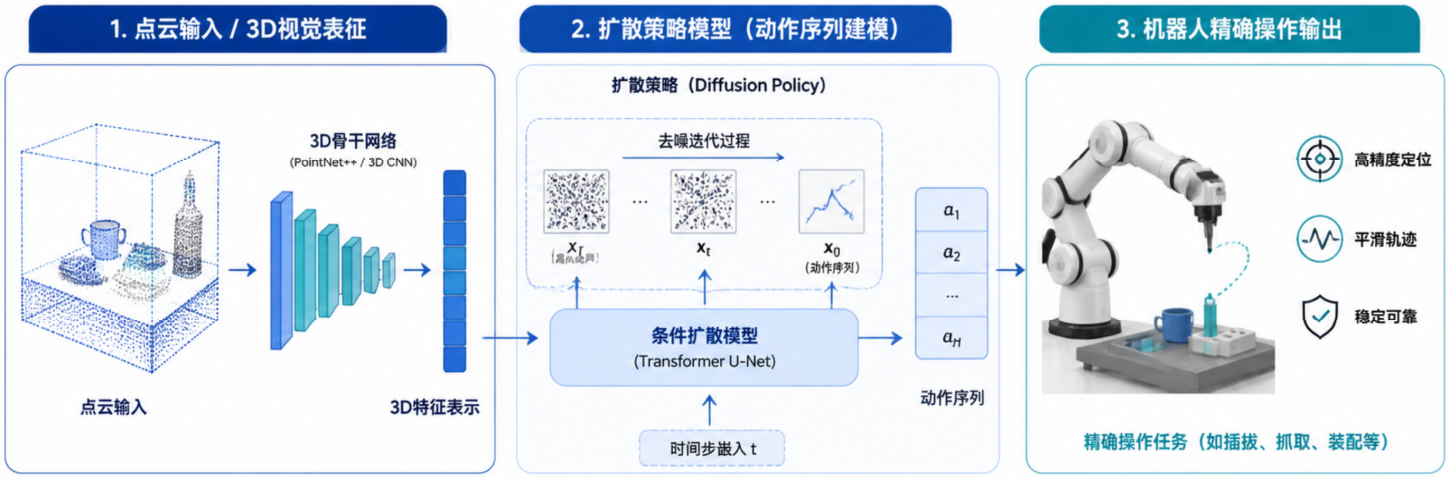
推理速度慢、计算成本高、需要大量数据。



3D Diffusion Policy（DP3）

DP3：3D视觉扩散策略

- 结合3D视觉表征与扩散模型
- 点云输入捕捉空间几何信息
- 在精确操作任务上表现优异
- 3D表征提升空间推理能力



DP3在精确操作任务上超越2D视觉方法。

Diffusion Transformer Policy

Octo（2024）

DiT：用Transformer替代扩散模型中的U-Net。

RDT-1B（2024）

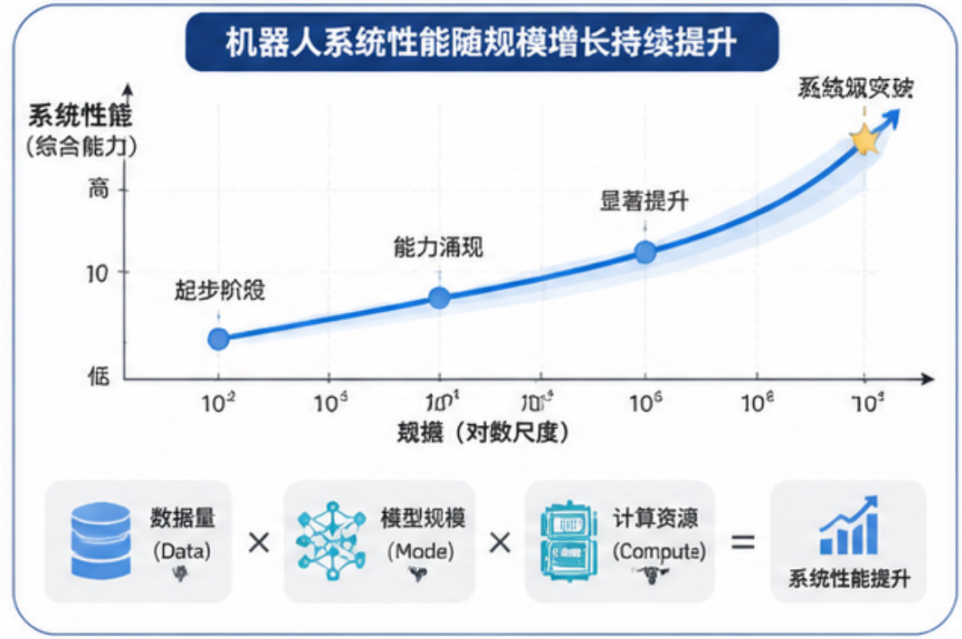
在图像生成和机器人控制中均表现优异。

DiT vs U-Net：DiT可扩展性更好，支持大规模训练；U-Net在小型任务上更高效。

从模仿到泛化：数据Scaling Law

Scaling Laws in Robotics

- 数据量增加→性能可预测提升
- 模型规模扩大→泛化能力增强
- 计算资源增长→训练更稳定
- 机器人领域验证GPT时刻



机器人领域的GPT时刻：规模带来质变。

动作分块与闭环控制

动作分块

每次预测T步动作序列

减少推理频率

提升时间一致性

默认16步

闭环重规划

每步重新观测推理

保证最新状态

要求100+ Hz频率

VLA大模型难以实现

折中方案

重叠分块（准闭环）

快慢分离

低频大脑~10Hz

高频小脑~100Hz

TriVLA三系统架构（2025）

① **感知系统**：高频视觉处理（~30Hz）

② **规划系统**：低频VLA推理（~10Hz），输出目标或策略参数

③ **控制系统**：高频传统控制器（~100Hz），执行闭环跟踪

这种"大脑-小脑"分离架构平衡了语义理解能力和实时控制需求。

本节内容

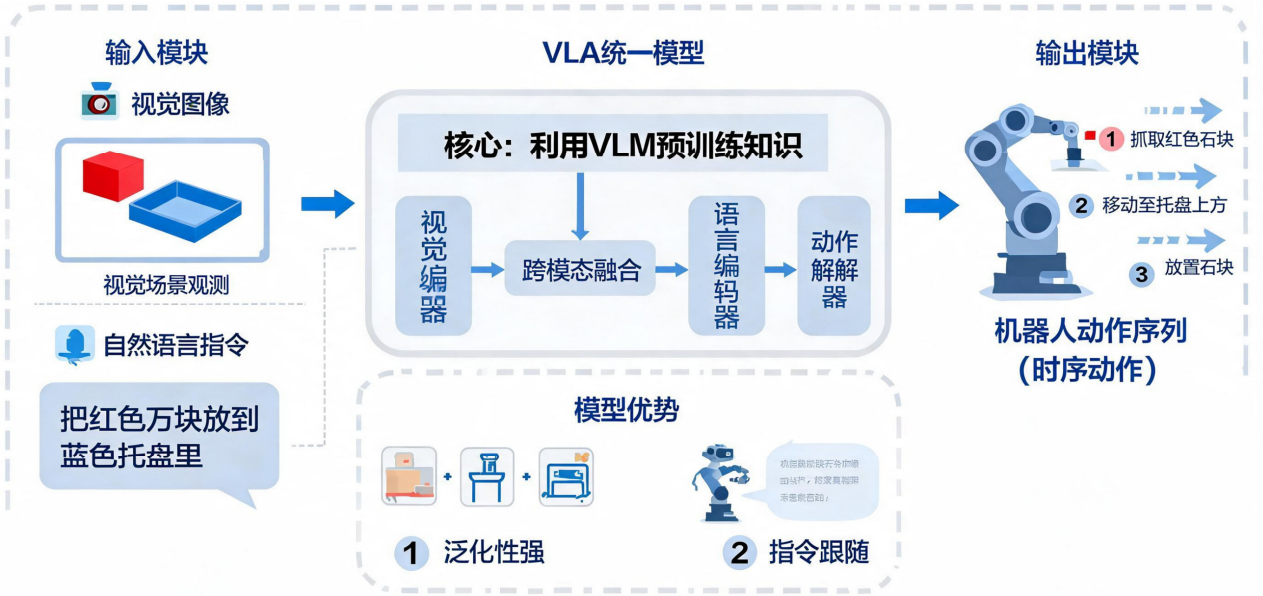
CONTENTS

一、具身认知的哲学与科学基础
二、机器人学基础
三、仿真环境与物理引擎
四、强化学习基础
五、模仿学习与Diffusion Policy
六、视觉-语言-动作（VLA）模型
七、世界模型与Sim2Real

VLA模型概述

VLA模型：视觉-语言-动作统一

- 输入：视觉图像 + 自然语言指令
- 输出：机器人动作序列
- 核心：利用VLM预训练知识
- 优势：泛化性强、指令跟随



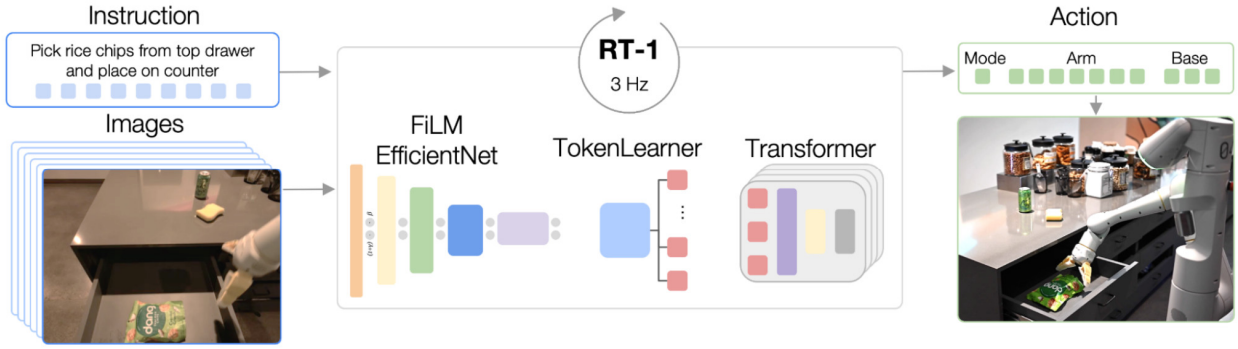
三阶段发展：萌芽期(2022-2023) → 爆发期(2023-2024) → 优化期(2024-2025)

革命性优势：VLA实现了"高层语义指令→低层物理行动"的直接映射，相比传统流水线方法（感知→语义理解→任务规划→运动规划→控制）大幅简化。

RT-1：VLA的奠基之作

首个展示Transformer在机器人控制中可扩展性的工作

- 17个月、13台机器人、约13万条轨迹
- EfficientNet-B3视觉编码 + FiLM条件
- Transformer解码器输出8维动作
- 700+任务，证明大规模数据→强泛化

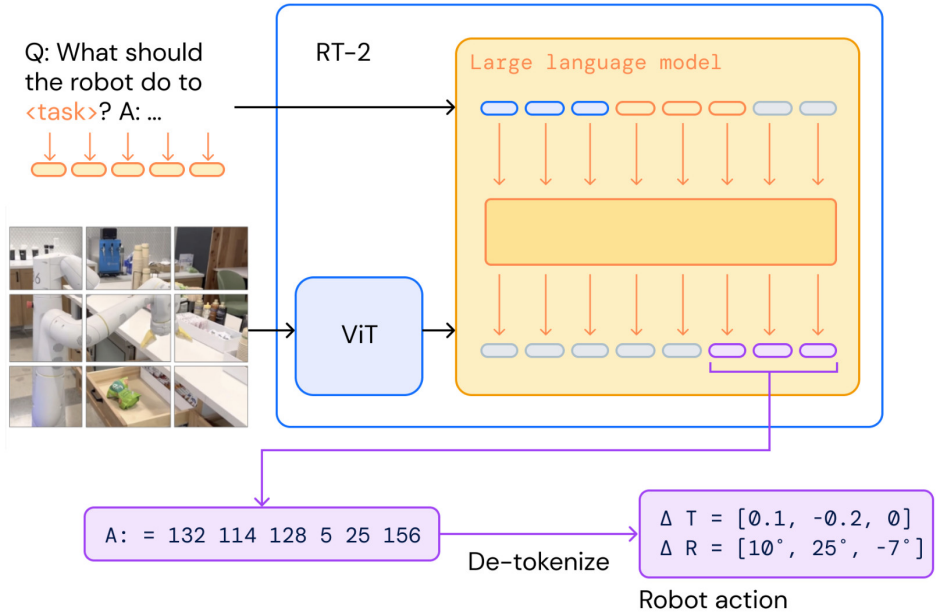


核心贡献：RT-1证明了Transformer+大规模数据可实现机器人策略的强泛化。

RT-2：将VLM知识迁移到机器人控制

RT-2：将VLM直接转化为机器人策略

- 使用PaLI-X/PaLM-E作为视觉-语言预训练骨干
- 端到端训练：视觉+语言指令→机器人动作
- 涌现能力：符号理解、推理、多语言跟随
- 泛化到未见过物体和抽象指令

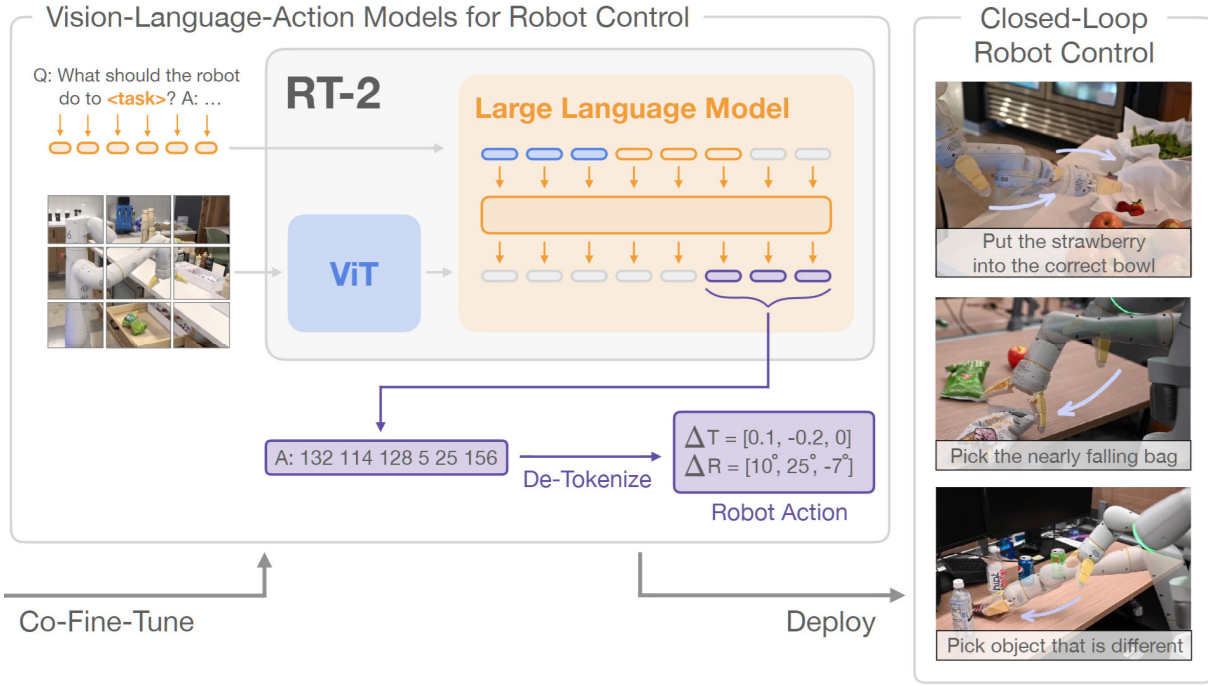


例如：对"捡起已灭绝动物"的指令，RT-2能从恐龙玩具中正确选择。

RT-2的架构细节与推理

RT-2架构：VLM + 动作微调

- 保留VLM的预训练权重和推理能力
- 将连续动作离散化为token序列
- 联合训练视觉-语言-动作目标
- Co-training比例约1:1防止灾难性遗忘

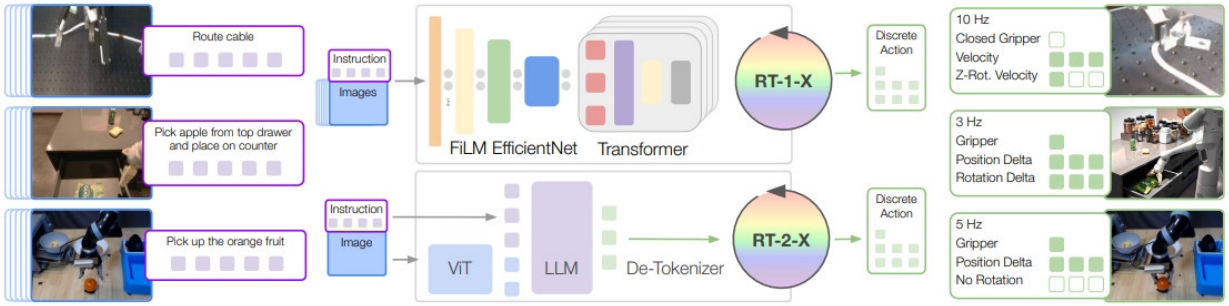


局限：计算成本高、推理延迟大、对精确操作任务表现有限。

RT-X：跨本体泛化的尝试

RT-X：大规模机器人协作训练

- 汇集22种机器人、约100万条轨迹
- 多机构协作的大规模实验
- 验证跨机器人迁移的可行性
- 开源Open X-Embodiment数据集

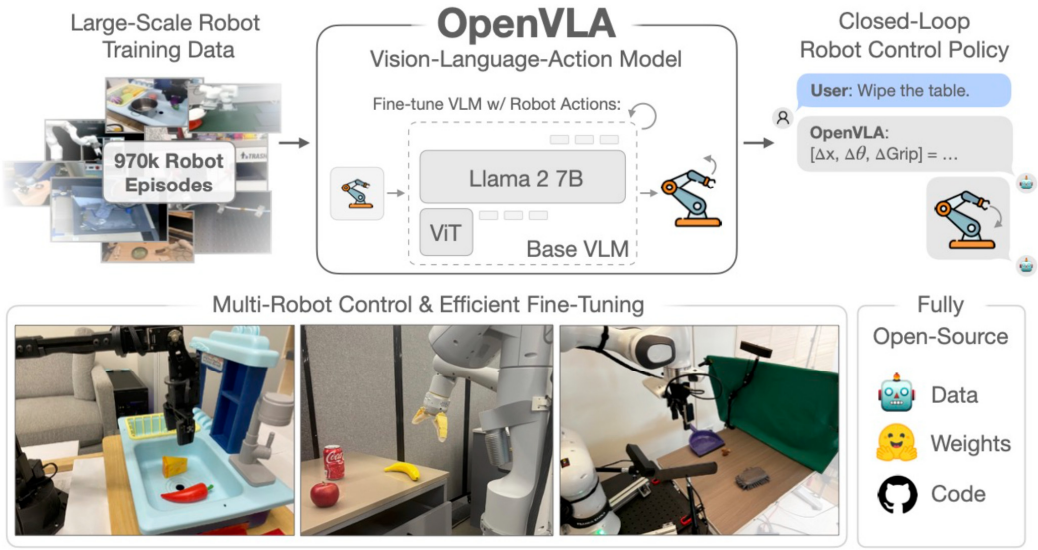


RT-X是机器人领域最大规模的协作研究项目。

OpenVLA：开源VLA的标杆

OpenVLA：全开源VLA方案

- 基于Llama 2 + DINOv2视觉编码
- 完全开源（模型、数据、代码）
- 支持多种机器人平台微调
- 7B参数，消费级GPU可运行

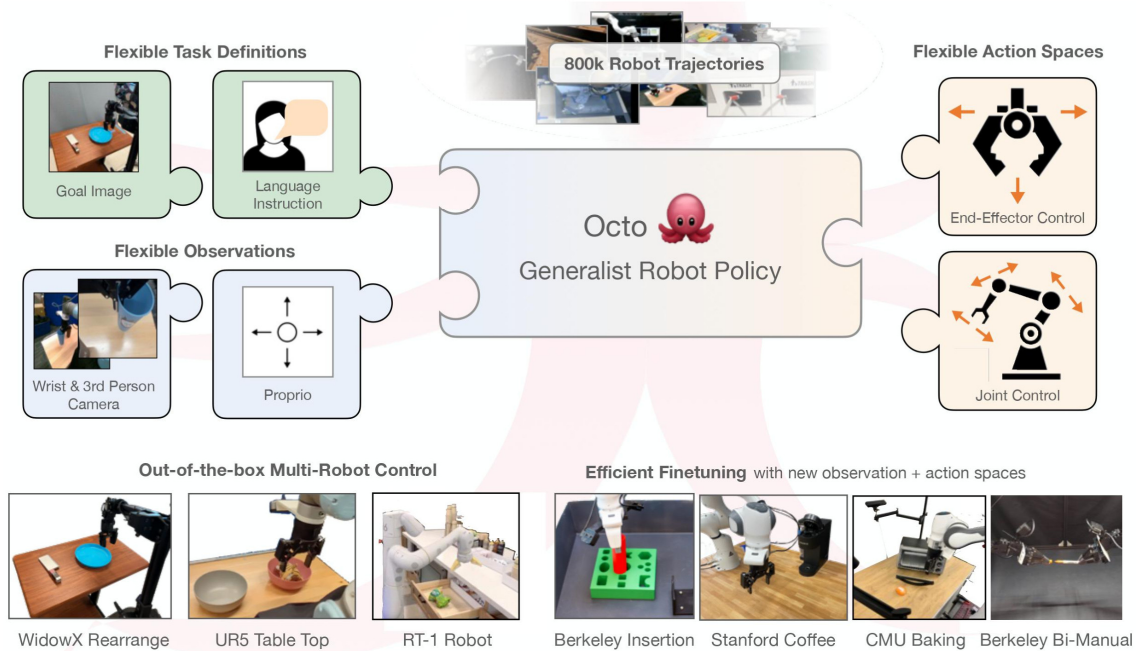


OpenVLA-OFT：引入并行解码减少推理延迟，在下游任务上的成功率比基础OpenVLA提升**15-30%**。

Octo：通用机器人策略

Octo：开源通才机器人策略

- 基于Transformer扩散模型的开源VLA
- 80万条轨迹、约100个数据集联合训练
- 支持多模态指令（语言、目标图像、动作）
- 模块化设计，易于微调和扩展

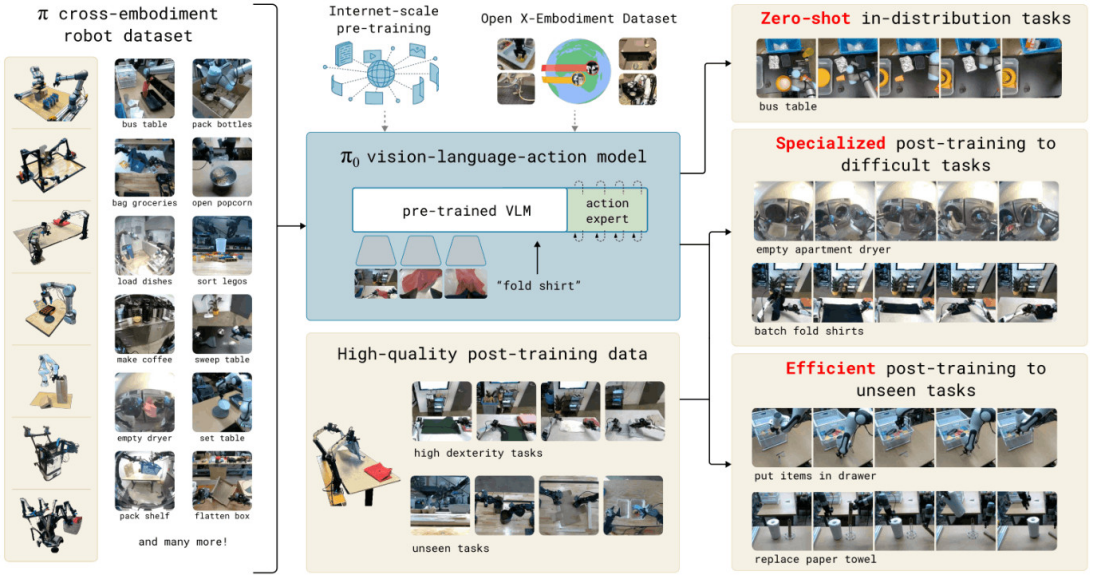


Octo vs RT-X：Octo完全开源（模型+数据+训练代码），RT-X仅开源部分组件。

π0与π0.5：流匹配革命

π0：流匹配VLA模型

- 采用流匹配(Flow Matching)替代扩散模型
- 更快的推理速度，更高的动作精度
- 支持高频控制（50Hz）和复杂操作
- 在真实机器人上验证的先进VLA系统



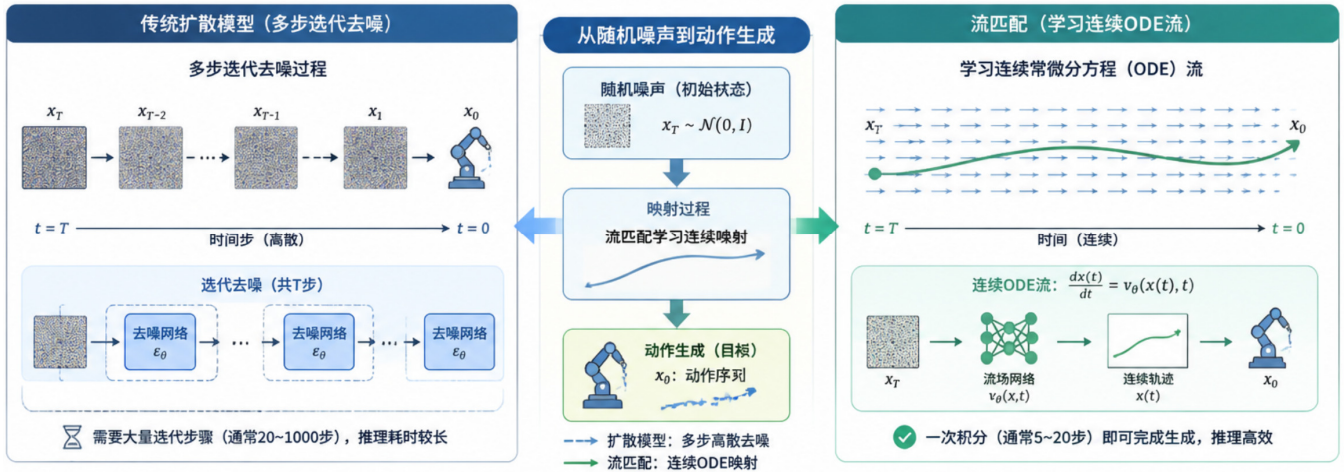
π0是首个在真实环境中展示流匹配优势的VLA模型。

π0的流匹配详解

CNF：dx/dt = v_t(x) | L_FM = E[||v_t(x_t) - u_t(x_t)||²]

流匹配：扩散模型的加速替代

- 直接学习ODE流，无需迭代去噪
- 推理速度提升10-100倍
- 动作质量与扩散模型相当
- 适合实时机器人控制场景

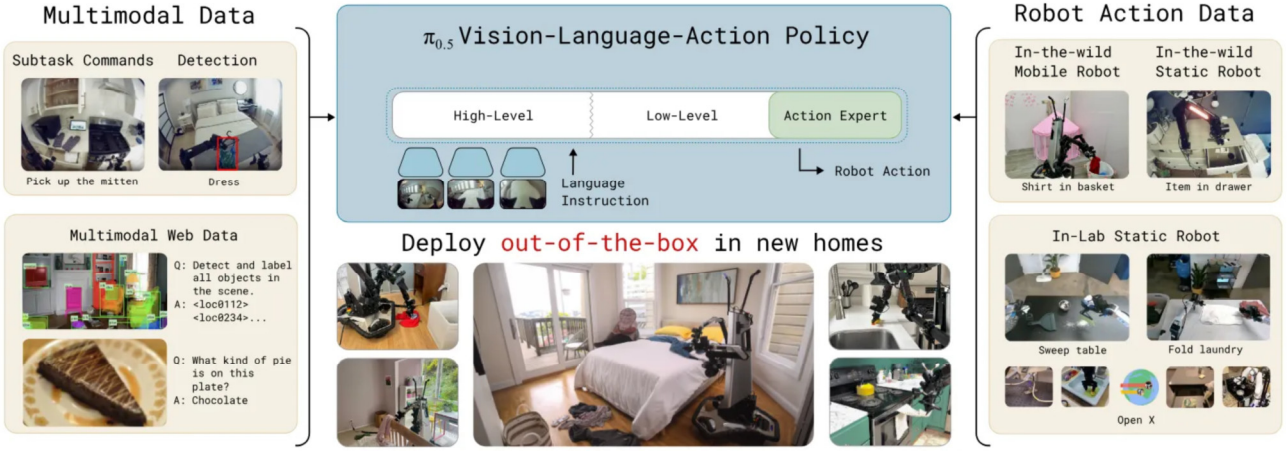


条件流匹配：将观测和语言指令注入向量场网络 v_t(x_t; s, l)，使生成的动作分布与当前情境匹配。推理时只需1-5步ODE积分即可获得动作。

π0.5：开放世界泛化

π0.5：高效VLA的演进

- 在π0基础上优化架构和训练策略
- 更高的数据效率和推理速度
- 支持更复杂的长期任务和推理
- 展示了VLA模型的持续改进路径

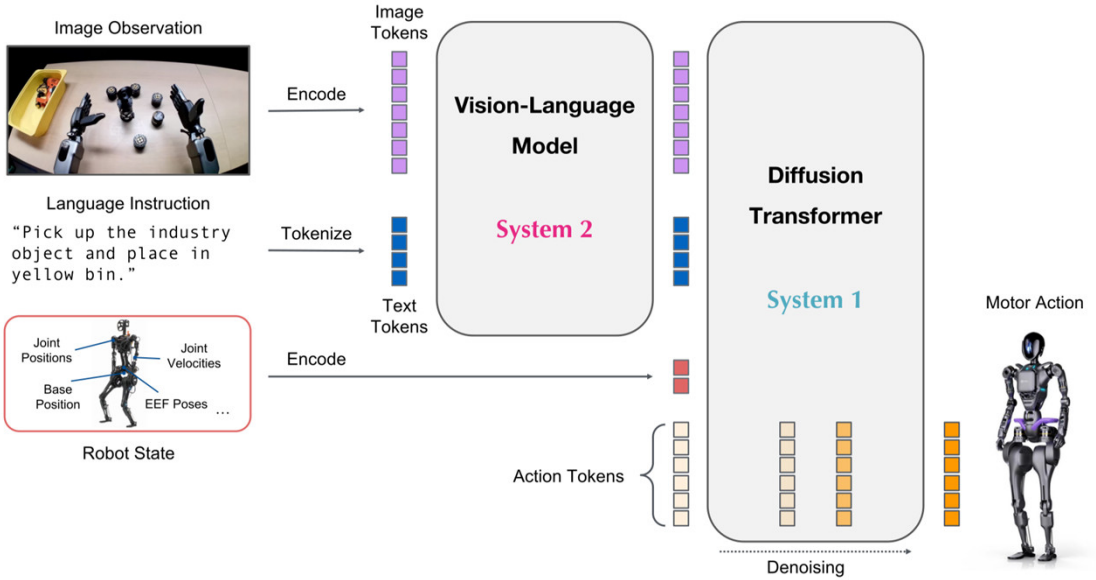


π0.5在保持性能的同时显著降低了训练和推理成本。

GR00T：NVIDIA的扩散Transformer策略

GR00T：NVIDIA人形机器人基础模型

- NVIDIA推出的通用人形机器人VLA模型
- 支持自然语言指令和视觉感知
- 与Isaac Sim仿真平台深度集成
- 旨在成为机器人领域的GPT时刻



GR00T代表了科技巨头对具身智能的战略布局。

VLA模型技术路线对比

维度	RT-2	OpenVLA	Octo	π0	GR00T
动作表示	离散token	离散token	扩散连续	流匹配	流匹配
参数规模	55B	7B	93M	~3.5B	~8B
开源	否	是	是	否	是
控制频率	1-3Hz	3-5Hz	10Hz	50Hz	20-50Hz
跨本体	部分	是	是	是	人形专精
核心创新	Web知识迁移	开源标杆	轻量扩散	流匹配	DiT+生态

VLA的推理效率优化

1 **动作分块**

一次预测16步，~5Hz执行

2 **投机验证**

轻量草稿模型+大VLA验证

3 **模型量化与蒸馏**

TinyVLA/EdgeVLA 4bit量化

4 **并行解码**

OpenVLA-OFT并行token解码

VLA的安全性与鲁棒性

对抗攻击

视觉和语言扰动敏感

轻微噪声导致危险动作

后门攻击

训练数据中注入特定模式

隐蔽后门触发恶意动作

安全机制

SAFE-Dict概念字典

推理时安全检测

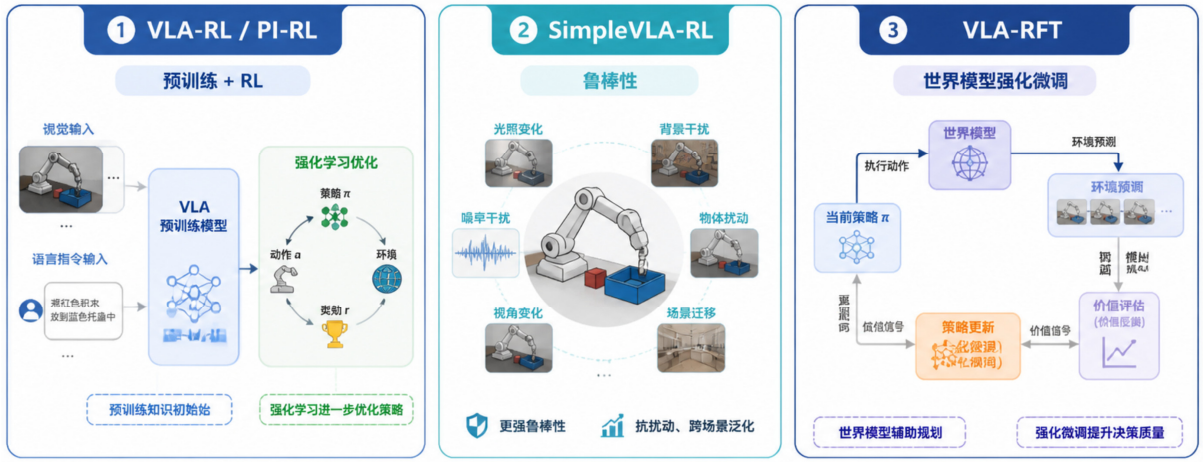
工业制造、物流仓储、家庭服务、医疗辅助、农业自动化等领域广泛应用前景。

VLA + RL：从模仿到超越

核心思想：从模仿中初始化，从探索中超越

方法：VLA-RL/PI-RL预训练+RL · SimpleVLA-RL鲁棒性 · VLA-RFT世界模型强化微调

挑战：参数量大 · 动作分布复杂 · 混合架构设计



关键洞察：VLA提供良好的初始化避免随机探索，RL通过环境反馈优化策略。VLA+RL融合是实现真正自主学习的关键路径。

本节内容

CONTENTS

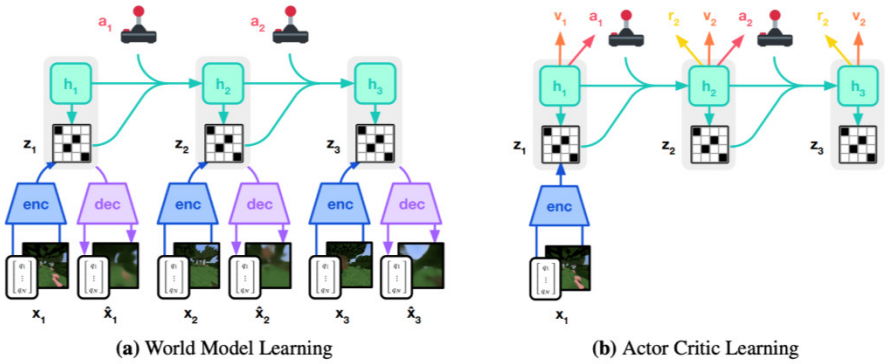
世界模型的概念与动机

核心思想：学习内部模拟器，预测动作后世界变化

三重价值：样本效率 · 规划推理 · 因果理解

思想渊源：

Craik(1943)→Schmidhuber(1990)→Ha(2018)



(a) World Model Learning (b) Actor Critic Learning

与具身智能的关系：世界模型让机器人在"想象中"学习，是解决样本效率问题的根本途径。与Clark的预测加工框架和LeCun的JEPA愿景一致。

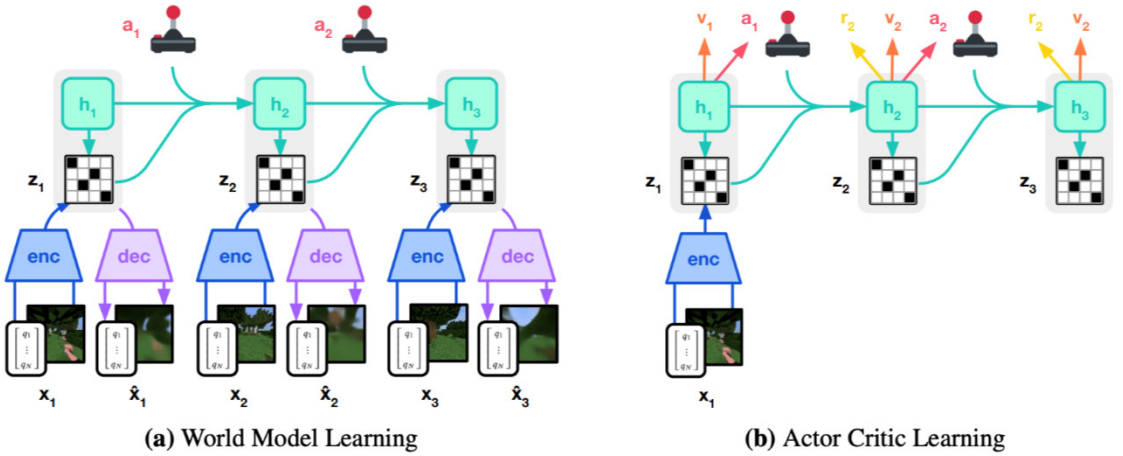
RSSM：循环状态空间模型

Dreamer系列核心架构

状态分解

- 确定性路径 h_t（编码历史）
- 随机状态 s_t（编码不确定性）

训练目标：重构损失 + KL散度约束



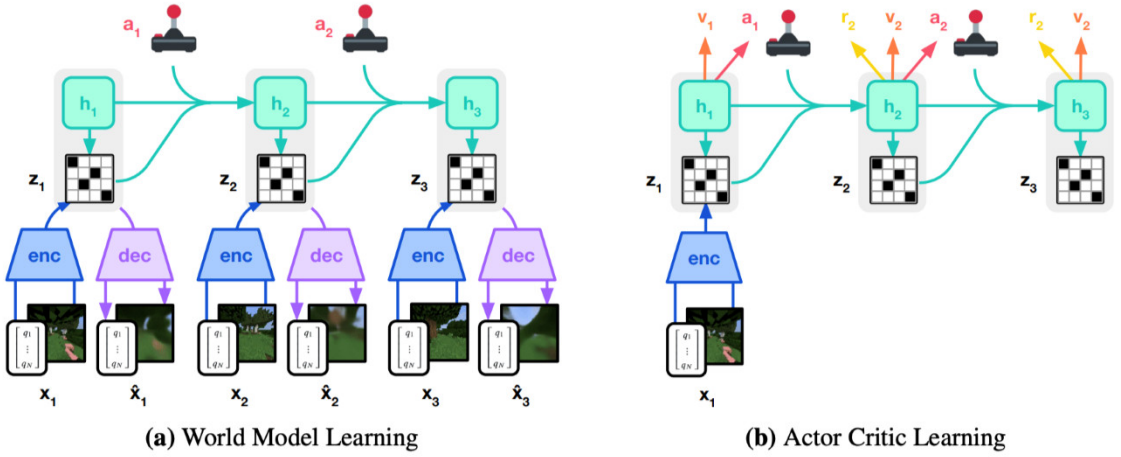
关键洞察：RSSM的"想象"能力来源于先验模型p(s_t|h_t)——在没有真实观测的情况下，可以通过链式预测生成未来轨迹。

DreamerV3：统一超参数的突破

发表于Nature：单一超参数跨越150+领域

关键创新：Symlog变换 · 自适应归一化 · 离散世界模型 · 两阶段策略

Minecraft：首个从零收集钻石的算法



意义：DreamerV3证明了世界模型方法的通用性和可扩展性——**无需针对每个任务调参**，单一配置即可跨越完全不同的领域。

DayDreamer与真实机器人

Unitree A1四足机器人：1小时内从零学会行走

核心流程：初始探索→世界模型学习→想象中训练→真实部署→迭代优化

关键洞察："模型数据比真实数据更廉价"

局限：在简单领域（平地行走）表现良好，但在复杂领域（崎岖地形、操作任务）中模型误差累积，导致想象中训练的策略在真实环境中性能下降。

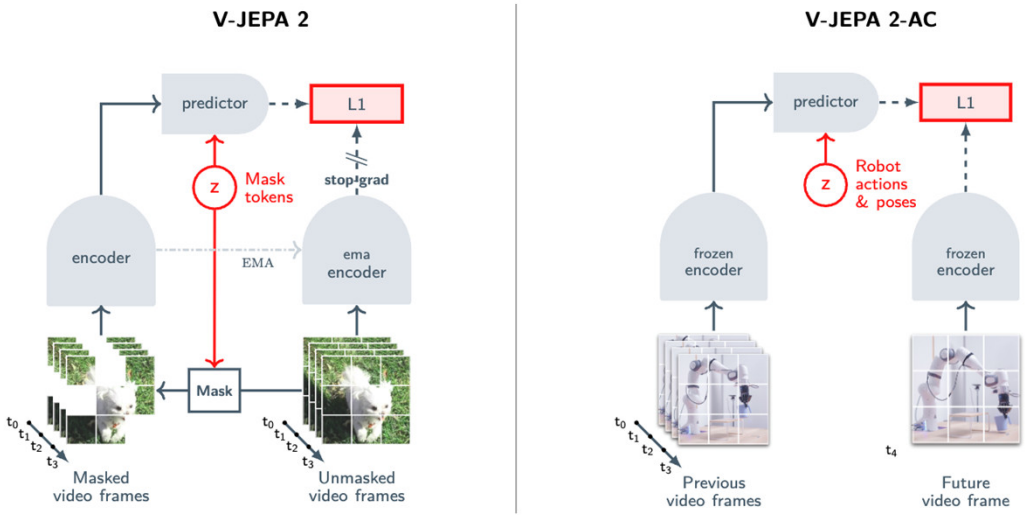
JEPA：LeCun的世界模型愿景

非生成式世界模型

批评生成式模型：预测原始数据是浪费的

三模块：感知编码器 · 预测器 · 目标编码器

训练目标：最小化预测表征与目标表征在隐空间的距离



核心思想：在**抽象表征空间**中预测，忽略无关细节（如风吹树叶的具体位置），关注语义结构和物理规律。

V-JEPA与V-JEPA 2

I-JEPA (2023)

图像版JEPA

遮挡区域表征预测

ImageNet性能

与MAE等方法相当

V-JEPA (2024)

视频版JEPA

3D多块掩码

遮挡约90%内容

200万视频训练

V-JEPA 2 (2025)

12亿参数

VLA视觉骨干

动作预测

世界建模能力

与生成式世界模型对比

优势：训练效率高 · 避免"幻觉" · 表征更抽象语义化

批评：尚未在复杂控制任务上展示与DreamerV3或VLA相媲美的性能

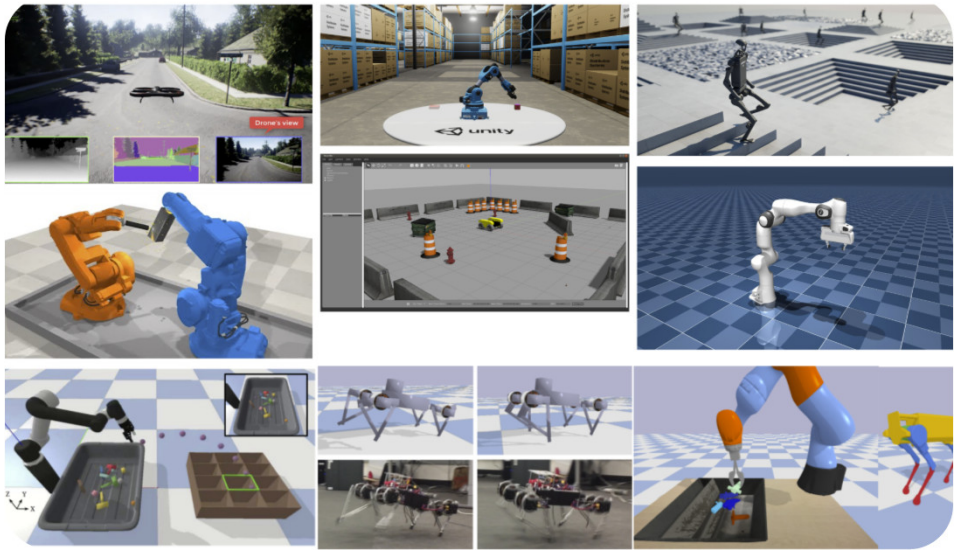
Sim2Real：现实鸿沟与域随机化

现实鸿沟（Reality Gap）

仿真与真实在物理/视觉/动力学上的差异

域随机化（DR）

- 随机化视觉属性（纹理、光照）
- 随机化物理属性（质量、摩擦）
- 使策略学到跨域鲁棒特征



自动域随机化：ADR（OpenAI, 2019）和DORAEMON（2023）尝试自动调整随机化分布，避免手动调参。

Sim2Real进阶：域适应与系统辨识

域适应

通过对抗训练减小仿真与现实的分布差异

系统辨识

精确测量物理参数构建高保真仿真SysID+DR组合

Real2Sim2Real

真实数据重建数字孪生训练迁移到真实

教师-学生

仿真教师策略真实学生策略模仿学习蒸馏

核心原则：没有单一完美的Sim2Real方法，通常需要组合多种技术

实践建议

- 先用系统辨识建立基础仿真
- 再用域随机化覆盖剩余不确定性
- 最后用少量真实数据微调

世界模型+VLA：融合趋势

代表性工作

• UniWorld · VLA-RFT · DreamVLA · World-Env

LeCun六模块：感知→世界模型→代价模块→记忆→动作→配置器

技术愿景

VLA提供语义理解和动作生成

世界模型提供未来预测和规划

两者的融合可能导向真正的"自主机器智能"

当前VLA覆盖感知-动作映射，世界模型填补"内部模拟"和"规划"的空白

人形机器人概览

美国阵营

- Tesla（Optimus）· Figure AI（Figure 02）
- Boston Dynamics · Agility Robotics（Digit）

中国阵营

- 宇树（Go2/H1）· 优必选（Walker）
- 智元（AimBot）· 智平方（Alpha Bot）

市场预测：Goldman Sachs预测2035年达**380亿美元**；IDTechEx估计**300亿美元**；McKinsey预测2025年全球部署约4200台，2028年达25000台。

Tesla Optimus：从汽车到机器人

核心优势：大规模制造 · Autopilot AI · Dojo超算 · 垂直整合

Optimus Gen-2：175cm / 63kg / 50+DoF / 20kg负载 / 8h续航

部署计划：2025年工厂5000台→2026年量产数万台

Figure AI：工业落地的先行者

Figure 02：BMW工厂 · 精密钣金装配 · 每天10h×5天

技术特点：OpenAI合作 · 端到端神经网络 · 力控±0.1N

融资：B轮6.75亿美元，估值26亿美元

2025年：Helix VLA系统实现多机器人协作

2025年：发布Helix VLA系统，实现多机器人协作和任务泛化。当前唯一经过真实工业产线验证的人形机器人。

国产具身智能：快速崛起的生态

本体厂商：宇树 · 优必选 · 智元 · 追觅 · 小鹏

具身大模型：智平方 · 银河通用 · 星动纪元

政策驱动：工信部《人形机器人创新发展指导意见》

竞争优势：制造供应链 · 成本优势 · 应用场景丰富

UBTECH WalkerS1

Unitree H1

AGIBOT Raise A2

Xiaomi CyberOne

问题和讨论