中关村智友

2024

具身智能科技前沿热点

联合发起单位

中关村智友研究院、青年科学家50人论坛

编者按

具身智能，作为人工智能领域的一颗璀璨新星，正以其独有的方式与深邃的内涵，在科技的浩瀚星空中勾勒出一幅幅壮丽的图景。它不仅仅是一种技术的革新，更是人类智慧探索未知边界的又一重要里程碑。通过模拟生物体的感知、认知与行动能力，具身智能实现了与环境的高度融合，这一过程涉及信息的精准捕捉、深度理解、快速决策与灵活执行，展现了强大的适应性和创造力。这一智能范式的崛起，不仅标志着人工智能技术的质的飞跃，更为全球科技竞争格局注入了新的活力与不确定性，预示着围绕具身智能技术的全球科技竞赛拉开帷幕。

从精密制造与智能工厂的自动化升级，到医疗健康领域的个性化治疗与辅助康复设备，再到智能家居与数字娱乐的深度融合，具身智能以其广泛的应用场景和深刻的行业影响力，正逐步重塑社会的运行逻辑与人们的生活方式，成为推动经济社会发展的新引擎。它不仅提升了生产效率，优化了服务体验，更是为人类解决复杂问题提供了前所未有的智能工具，加速了新质生产力的形成与发展。

为了全面剖析具身智能的发展现状，精准把握未来趋势，我们精心构建了一个连通产学研的智囊团，汇聚了来自顶尖高校、研究机构以及行业企业的专家学者。他们依托深厚的学术造诣与丰富的实战经验，紧密跟踪Nature、Science等国际顶级学术期刊的最新研究成果，结合产业数据分析，全方位、多层次地开展了深入研究与分析。在此基础上，精心编纂了《2024具身智能科技前沿热点》报告，旨在为行业内外人士提供一份权威、前沿的参考指南。

本报告精心筛选的具身智能科技热点，不仅覆盖了具身智能灵巧操作特点，还深入探讨了空间智能的拓展应用、人形机器人的商业化路径、大规模仿真训练平台的构建与优化、触感灵巧手的精密操控技术、以及具身机器人导航大模型的智能导航策略等。这些热点不仅代表了当前具身智能技术的最前沿，也预示着未来技术发展的可能方向。

作为持续关注并推动具身智能领域发展的年度系列报告，我们将持续跟踪行业动态，及时发布最新研究成果，与业界共享知识，共谋发展。同时，我们也深知，具身智能技术迭代速度之快、涉及领域之广，要求我们始终保持敬畏之心，严谨治学，科学预测。本报告中，所有分析与预测均基于编写团队在有限时间内的调研与数据整理，同时，我们的检索可能未覆盖所有相关领域，内容仅供参考，不构成任何投资建议或决策依据。我们鼓励读者结合各自领域的实际情况，审慎评估，科学决策。

在此，我们再次向所有参与本报告编纂工作的专家学者表示最诚挚的感谢，他们的智慧与汗水是这份报告得以问世的关键。同时，我们也深深感激每一位读者的关注与支持，正是你们的期待与鼓励，激励着我们不断前行，追求卓越。我们坚信，通过持续的探索与创新，具身智能必将为人类社会的可持续发展贡献更多力量，开启一个更加智能、更加美好的未来！

《2024具身智能科技前沿热点》编委会

2024年12月于北京

2024具身智能科技前沿热点
专家委员会

战略顾问

王田苗　北航机器人研究所名誉所长，中关村智友研究院院长

青年科学家专委会

董　豪　北京大学助理教授

方　斌　北京邮电大学教授

高　飞　浙江大学控制科学与工程学院长聘副教授

郭彦东　智平方创始人兼CEO

韩文娟　北京交通大学副教授

季　超　科大讯飞机器人首席科学家，科大讯飞-中国科学技术大学联合培养博士

李　淼　武汉大学特聘研究员、副教授，华中科技大学创业导师

刘华平　清华大学教授

马道林　上海交通大学副教授

苏　航　清华大学计算机系副研究员

陶　永　北京航空航天大学副教授，博导

王超群　山东大学控制科学与工程学院教授

王　越　浙江大学控制科学与工程学院教授

袁海辉　五八智能科技（杭州）有限公司副总经理

责任编委

方　斌　北京邮电大学教授

刘华平　清华大学教授

陶　永　北京航空航天大学副教授，博导

英语霏　中关村智友研究院副院长

（按姓氏拼音首字母排序）

目　录

一、具身智能灵巧操作大模型

近年来，具身智能领域发展迅猛，强调机器人在真实世界中与人类、环境及其他机器人之间的有效交互。然而，机器人所面临的实际环境通常是动态变化且充满不确定性的，其规划器和执行器难免出现误差。若这些误差未能及时纠正，将可能逐步累积，导致任务失败。因此，自我纠正技术在机器人和自动化领域的重要性日益凸显。这种技术不仅显著提升了机器人在复杂任务中的准确性和鲁棒性，还增强了机器人在变化环境中的适应能力，同时降低了对人工干预的依赖，从而大幅提高整体工作效率。

在这一背景下，端到端具身大模型作为具身智能领域的新技术范式，正通过统一架构实现从环境感知到任务执行的完整闭环。不同于传统模块化方法，具身大模型通过大规模数据驱动的端到端学习，直接优化整体性能，显著提升了任务执行的效率、鲁棒性和适应性。其核心是构建一个多模态、具有强推理能力的基础模型，融合视觉、语言、触觉等多种感知形式，同时整合规划、决策与控制功能，使机器人在动态和不确定的环境中能够高效完成复杂任务。这种架构通过消除中间人工设计步骤，简化了系统流程，具备整体优化、泛化能力强和可持续迭代的显著优势。尤其在具身智能灵巧操作这一研究难点上，2024年多项研究（如Aloha、OpenVLA、RDT等）表明，结合大模型预训练与强化学习的方式，使机器人操作的泛化能力和成功率有了显著提升。这种端到端架构也使机器人能够在多个领域实现更强的跨任务适应能力。

具身智能灵巧操作大模型不仅是具身智能技术发展的重要支柱，也是国家高科技发展水平和工业自动化程度的重要体现。通过对具身智能灵巧操作大模型的研究，为航天、工业制造等重大需求提供了核心技术支持，为机器人技术和人工智能的深度融合开辟了新的方向。

1.市场热点/行业前景

近年来，人工智能和机器学习的迅速发展推动了具身智能技术的突破，特别是在大模型驱动的机器人控制、操作和决策领域，展现出极大的技术潜力和市场前景。具身大模型通过统一的多模态架构，整合视觉、语音、触觉等信息，显著提升了机器人灵巧操作能力，推动机器人技术在多个行业中的广泛应用。

2024年3月，UC伯克利机器人领域的领军专家Sergey Levine创立了公司Pi（Physical Intelligence），核心团队汇集了硅谷机器人和人工智能领域的顶尖专家。Pi的目标是通过一个通用模型将AI带入物理世界，为各类机器人和物理设备提供动力，适用于广泛的应用场景。公司专注于开发纯软件的机器人基础模型，以VLA端到端具身操作大模型范式为基础，为多种硬件形态的机器人赋能。同年7月，Skild AI宣布完成3亿美元A轮融资，投资者包括杰夫·贝佐斯、日本软银集团、红杉资本和卡内基梅隆大学等，将公司估值推至15亿美元。Skild AI由卡内基梅隆大学教授Deepak Pathak和Abhinav Gupta于2023年创立，专注于开发基于物理世界的智能系统，致力于构建类似“机器人大脑”的机器人基础模型。其技术旨在赋能各类机器人应用，挑战“AGI只能来源于数字世界”的传统观念，展现了极大的行业潜力。

具身智能灵巧操作大模型在工业、医疗和家庭服务等领域落地应用，并取得显著成果：1）制造业：灵巧机器人承担精细装配、质量检测和智能决策任务，大幅提高生产效率和自动化水平；2）医疗领域：在手术辅助和康复训练中的应用提升了手术精确性和康复效果；3）家庭服务：灵巧机器人未来将成为家庭中的“伙伴”，提供更智能化和个性化的服务体验。

全球范围内，各类机构与企业积极布局具身智能灵巧操作大模型。清华大学TSAIL团队的RDT模型、Google DeepMind的RT系列等，不仅在任务执行的精确度和多样性上取得重大突破，还通过跨领域合作与开放共享，推动了机器人智能化的发展。这些技术创新为具身智能研究提供了新的方向，并缩小了机器人操作与人类操控之间的差距。

根据市场分析，具身智能领域已成为全球资本追逐的热点。2024年，中国具身智能领域记录了38起投融资事件，总金额达到51.1亿元人民币。随着技术进步和市场需求增长，具身大模型机器人市场预计将实现爆发式增长。例如，在智能生产线中，具身通用多模态大模型通过实时感知和智能操作，提升了自动化水平；在医疗与康复辅助领域，这些技术优化了个性化服务并提升了医疗质量。

展望未来，具身智能灵巧操作大模型不仅是人工智能和机器人领域技术进步的重要支柱，也是产业转型升级的核心动力。随着跨领域技术（如物联网、5G通信）的深度融合，智能灵巧操作具身系统将为社会提供更高效、更智能的生产和生活解决方案，推动社会全面向智能化方向发展。

2. 典型案例

2.1 谷歌RT系列：从传感到行动的全能模型

2024年1月，谷歌在RT-1、RT-2的基础上发布了RT-H，这一模型结合语言动作层级提升了机器人在多任务环境中的表现。通过将复杂任务分解为细粒度的语言动作，RT-H实现了任务间的数据共享和泛化能力，提高了机器人执行任务的准确性和适应性。与RT-1和RT-2相比，RT-H进一步优化了任务控制方式：RT-1依赖视觉和语言数据指导动作，RT-2引入视觉-语言-动作模型完成复杂任务，而RT-H通过语言动作层级提供更细粒度的控制，成功率比RT-2提高约15%，并展现出更强的灵活性和泛化能力。这标志着谷歌在具身智能领域迈出了关键一步，为机器人在多任务复杂环境中的应用带来了新突破。

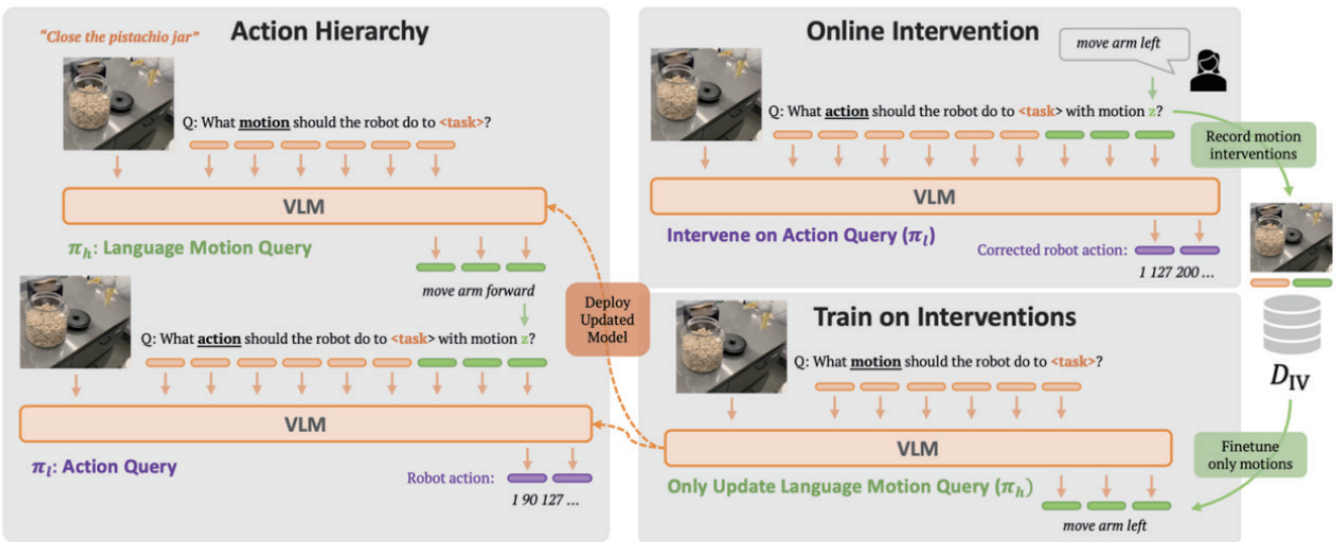

图 1.1 RT-H的总体流程

2.2 北京大学RoboMamba：高效的端到端VLA大模型-推理、操作一体化

RoboMamba是由北京大学与智平方团队联合推出的一款高效端到端视觉-语言-动作（VLA）具身大模型，专为机器人场景优化设计，旨在实现高效的推理与操作能力。2024年6月，这一成果以题为《RoboMamba：具备机器人推理与操控能力的高效视觉-语言-动作大模型》的论文，发表在全球顶级学术会议NeurIPS 2024上。

RoboMamba采用了先进的多模态设计，通过集成视觉编码器与线性复杂度的状态空间语言模型（SSM），显著提升了机器人在推理和操控中的表现。视觉编码器赋予模型强大的视觉常识理解能力，而SSM的高效计算能力则为模型提供了流畅的状态预测与任务规划能力。这种设计使RoboMamba能够在多任务场景中实现从高层次推理到低层次精细操控的端到端融合，同时大幅提高了模型的计算效率和任务执行效果。

该模型通过一种高效的微调策略，仅需调整模型参数的0.1%，就能在短短20分钟内完成微调。这种设计不仅提升了操作泛化能力，还使模型在适应多任务和多场景需求时更加灵活。与传统具身大模型相比，RoboMamba在推理速度上达到了现有模型的三倍，同时保持了卓越的鲁棒性与可靠性。在模拟与现实世界实验中，RoboMamba能够精准完成操控任务中的位姿预测，展现出对复杂机器人任务的高度适配性。

RoboMamba在机器人推理与操控领域实现了多项突破。在推理方面，模型具备精准的任务规划、长程任务规划、可操控性判断以及对过去与未来状态的预测能力，克服了传统方法的局限；在操控方面，RoboMamba通过高效的感知和推理，能够流畅完成复杂场景下的操控任务，为机器人“大脑”提供强大的推理思考能力，同时赋予其“小脑”精细的低层次操控技能。这样的能力组合使得RoboMamba在现实环境中的表现更加高效且可靠。

这一模型的显著优势还在于其以极低的训练成本实现高效性能的能力。通过生成精准的任务规划与位姿预测，RoboMamba有效

平衡了模型的泛化性、迁移性与运行速度，为具身智能的实际落地提供了强有力的技术支持。其快速适应能力和高效的运行机制，进一步降低了机器人在开发和应用中的时间成本，为推动智能机器人技术的广泛应用创造了更多可能性。

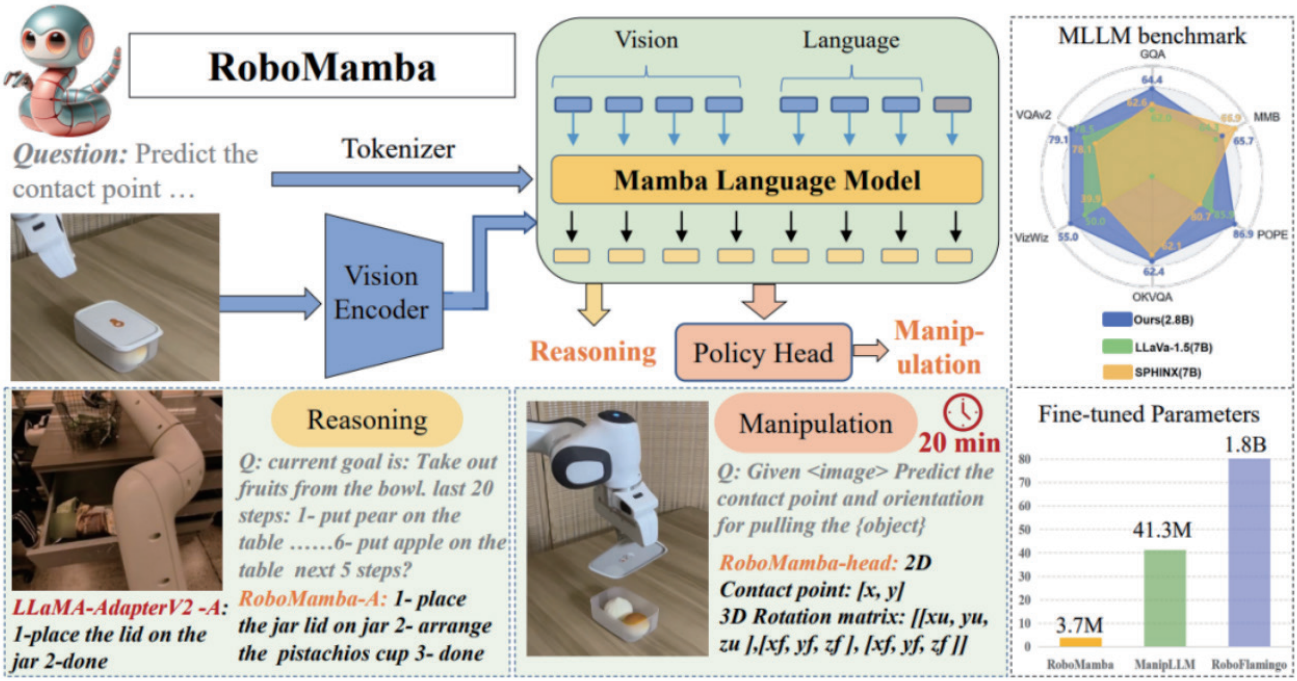

图 1.2 RoboMamba的总体流程

2.3 清华大学TSAIL团队：Robotics Diffusion Transformer（RDT）

清华大学人工智能研究院TSAIL团队于2024年10月推出了全球最大的双臂机器人操作任务扩散基础模型——Robotics Diffusion Transformer（RDT-1B）。这一创新模型通过基于扩散模型的设计与大规模预训练策略，为双臂操控任务的研究和应用带来了重要突破，成为运动控制领域最接近人类“小脑”的机器人控制模型之一。

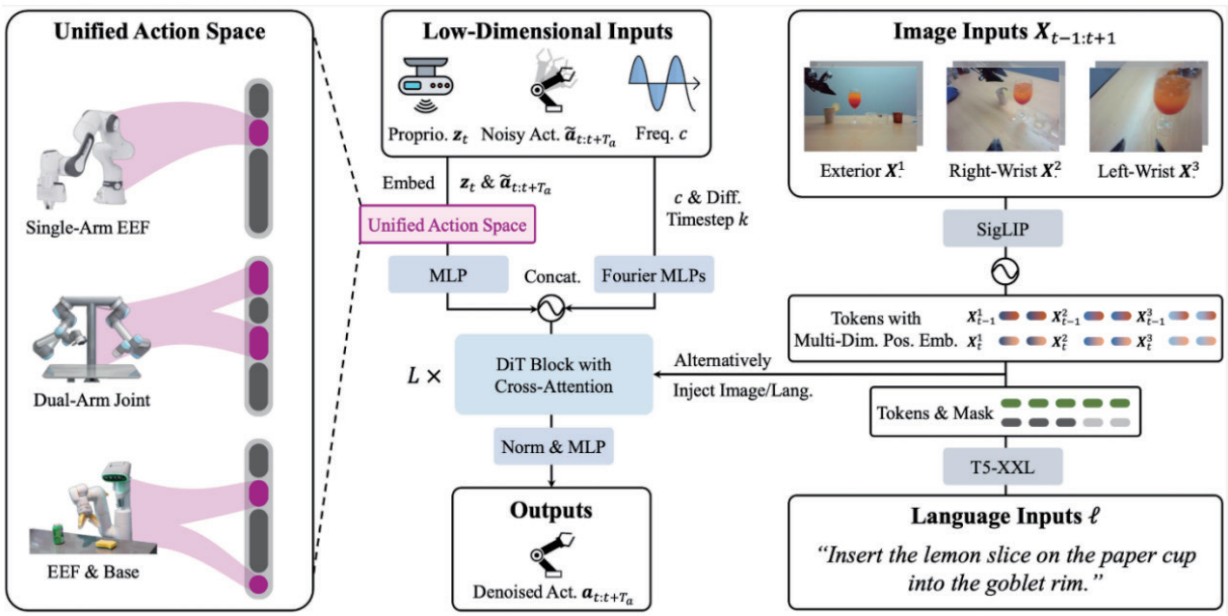

图 1.3 RDT的整体框架

RDT-1B具备1.2B参数量，采用了可扩展的Transformer架构，能够高效处理多模态输入的异质性，捕捉机器人数据中的非线性和高频特性。模型通过扩散模型的多模态行为分布表示，展现了卓越的动作预测与执行能力。在实际应用中，RDT-1B在ALOHA双臂机器人平台上得到了验证，其在家庭环境下的复杂任务中表现尤为出色。例如，在“洗杯子”任务中，RDT-1B能够精确完成一系列复杂操作，甚至在面对从未见过的新类型杯子时，也能展现出强大的零样本泛化能力。这种泛化能力使得模型能够快速适应全新的任务和物体，仅通过少量示范即可学习新技能。

RDT-1B在应对数据稀缺性问题上也取得了显著进展。模型引入了物理可解释的统一动作空间，使其能够统一不同机器人的动作表示，同时保留原始动作的物理意义。此设计极大提升了模型的跨平台知识迁移能力，使得RDT-1B能够在多个任务和物体场景中理解并执行复杂任务。这种能力不仅让模型具备出色的初始性能，也展现了强大的学习潜力和快速适应能力，为双臂操控领域的研究和优化奠定了坚实基础。

作为开源项目，RDT的发布将加速机器人技术的研发与产业化。凭借其多模态处理能力、高效的扩散模型架构和优越的泛化能力，RDT有望推动机器人在更多领域的应用，如家庭服务、工业自动化和医疗辅助等，成为推动机器人技术进步的重要驱动力。

2.4 基于多模态大模型的具身操作大模型RoboFlamingo

字节跳动联合清华大学开发的具身操作大模型RoboFlamingo，利用预训练的VLMs进行单步视觉语言理解，使用显式策略头对序列历史信息进行建模，并且仅在语言条件操作数据集上通过模仿学习进行微调。这种分解为RoboFlamingo提供了开环控制和在低性能平台上部署的灵活性。通过在测试基准上大大超过了最先进的性能，这表明RoboFlamingo可以成为使VLM适应机器人控制的有效和有竞争力的替代方案。广泛的实验结果还揭示了一些关于不同预训练VLM在操作任务上行为的有趣结论。RoboFlamingo有潜力成为机器人操作的具有成本效益的且易于使用的解决方案，使每个人都有能力微调自己的机器人策略。

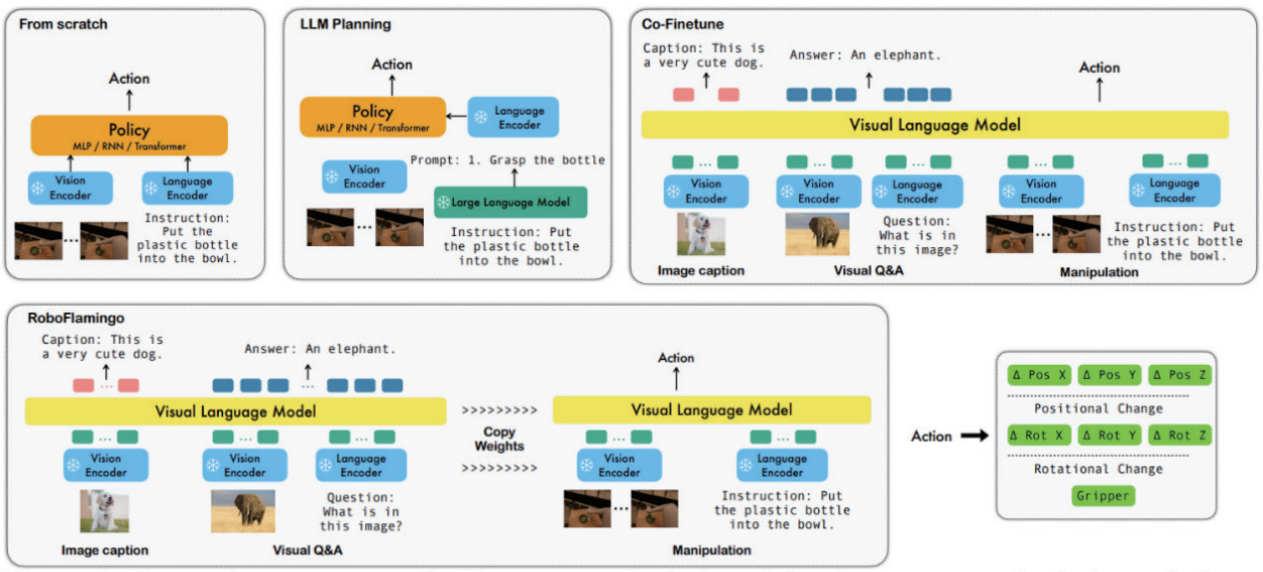

Figure 1: Comparison among *RoboFlamingo* and existing vision-language manipulation solutions.

图 1.4 RoboFlamingo的整体流程

2.5 基于大语言模型的机器人操作学习RobotGPT

三星电子中国研究院与中国工程院外籍院士张建伟教授、孙富春教授和方斌教授合作，提出了RobotGPT，一个创新的机器人操作决策框架，旨在推动ChatGPT在机器人操控应用中的实际应用。该框架的核心思想是将环境线索转换为自然语言，使得ChatGPT能够为智能体Q（Agent）生成动作代码，从而赋予机器人使用自然语言进行理性互动的能力，执行如拾取、放置等任务。

然而，ChatGPT生成的执行代码在稳定性和安全性方面存在一定的挑战。由于ChatGPT可能会对同一任务提供不同的答案，导致结果的不确定性，这种不稳定性使得直接将ChatGPT集成到机器人操作循环中成为一项困难。尽管将温度参数设定为0可以使输出更加一致，但这也可能牺牲多样性和创造力。

为了克服这些问题，RobotGPT引入了一种有效的提示结构，并结合强大的学习模型，以确保系统的可靠性和稳定性。框架中还加入了用于衡量任务难度的指标，以便更好地评估ChatGPT在机器人操作中的表现。通过在模拟和真实环境中的测试，RobotGPT显著提高了任务成功率，从38.5%提升至91.5%。这一结果表明，相比于直接使用ChatGPT作为任务规划者，利用ChatGPT训练RobotGPT能提供更加稳定和高效的解决方案。

尽管存在一定的限制和安全风险，RobotGPT框架为ChatGPT在机器人任务中的应用开辟了新的前景，并为相关研究提供了重要的启示，探索了ChatGPT在机器人操控中的潜力与能力边界。

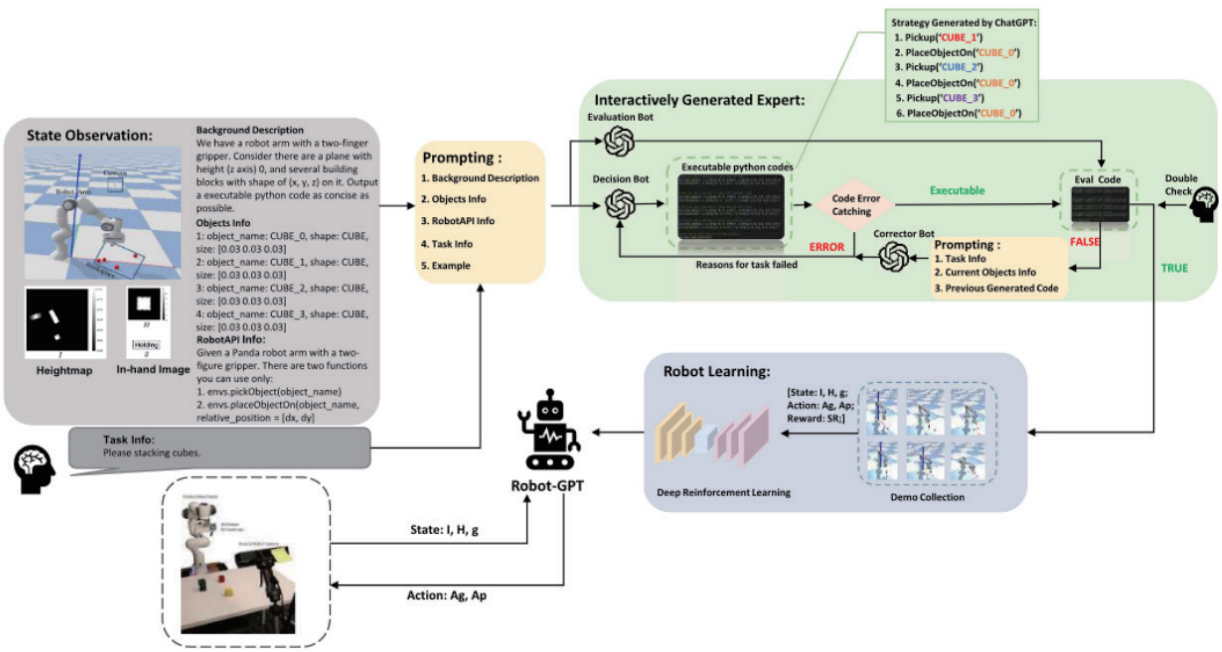

图 1.5 RobotGPT的整体流程

2.6 具身智能支气管镜机器人：提升医疗资源平等性与操作安全性

2024年1月，《Nature Communications》刊登了浙江大学团队关于AI辅助支气管镜机器人的研究。肺部疾病是全球健康负担，支气管镜检查在复杂气道导航中对医生技能要求高，导致其在欠发达地区普及率低。现有的机器人支气管镜虽有进展，但高成本和对经验的依赖限制了应用。研究提出了一种AI辅助的支气管镜机器人，结合AI-人类共享控制算法和创新硬件设计，旨在为新手医生提供专家级操作能力，提升检查安全性和效率，减少医疗资源不平等。系统包括可快速更换的导管、气管镜精确控制以及基于专家模仿的AI算法，能够实时接收医生指令并进行安全导航。实验结果表明，AI算法在模拟环境中的导航成功率达到93.3%，并且在体外和活体实验中，新手医生的导航精度超越专家，操作误差大幅降低。该系统不仅能减少误操作风险，提高诊疗质量，还能减轻医生的体力和认知负担。随着技术的进步和成本的降低，该系统有望广泛应用，促进医疗资源平等，提高全球健康水平。

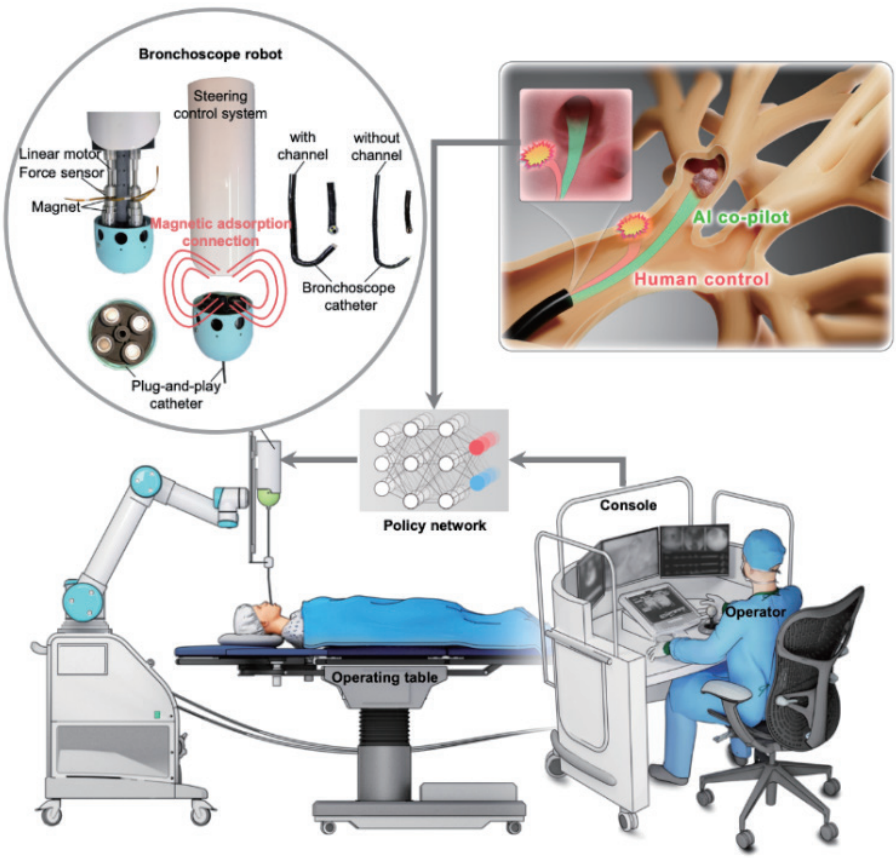

图 1.6 支气管镜机器人工作流程

二、空间智能

空间智能是人类智能的重要组成部分，不仅帮助人类理解并与周围世界交互，还赋予我们将内心想象转化为实际创造的能力。从求解问题到构建现实，无论是简单的沙堡还是宏伟的城市，空间智能的作用无处不在。同样，对于具身智能来说，空间智能是其发展的关键驱动力。以机器人为代表的智能终端，需要在物理世界中完成复杂任务，这要求其具备类似人类的能力，能够理解环境、进行交互并高效行动。

空间智能的核心在于通过对三维环境的精准理解和建模，生成动态的四维世界模型。通过这种能力，AI不仅可以识别开放环境中的物体和动态场景，还能够深入理解物理空间的动态变化关系，并进行空间推理。这些能力为具身智能的发展奠定了重要基础，使机器能够更深刻地理解人体与物理环境的关系，在复杂任务环境中实现自主学习与高效执行。同时，空间智能还进一步优化了人机交互及复杂场景中的运动能力，为人工智能技术的升级和在人类生活中更广泛的应用开辟了新的可能性。

未来，人工智能系统将以空间智能为核心，在推动技术发展的同时，为人类创造力的全面提升提供强有力的支持。空间智能通过理解3D环境信息，不仅能够生成3D空间，还可以深入理解物理空间并进行3D空间推理，逐步形成4D的世界模型，为具身智能的发展奠定重要基础。

1.市场热点/行业前景

在视觉大模型（VLM）和具身智能的领域，感知性能对整体性能的提升起到了至关重要的作用。然而，目前主流的VLM模型在空间智能方面表现仍有不足，特别是在精细空间推理能力上存在显著缺陷。FAIR团队由图灵奖得主Lecun Yan与Saining Xie教授领导，通过研究发现，感知模块的性能直接决定了VLM模型的整体表现。例如，他们通过简单混合CLIP和DINOv2这两个感知模型，就显著提高了VLM的空间推理能力。

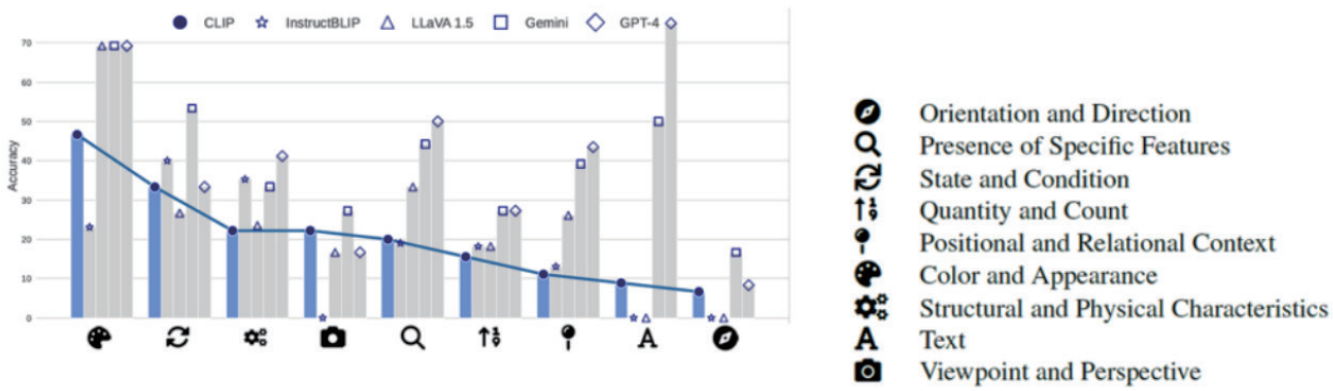

图 2.1 CLIP和MLLM在视觉模式上的表现

同样，硅谷初创公司Pi、斯坦福大学与伯克利大学的研究团队也通过类似的模型融合技术提升了机器人在端到端操作任务中的表现。然而，现有的空间感知模型依然无法完全满足具身智能对空间智能的高标准需求。

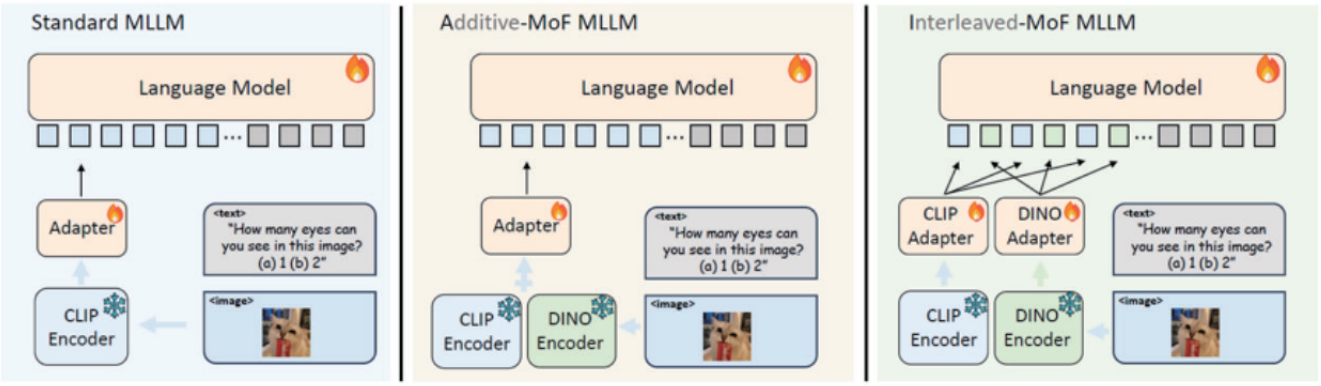

图 2.2 多模态语言模型的适配器设计对比

随着AI技术范式变革从数字世界向物理世界逐步扩展，感知性能的突破成为推动空间智能发展的关键技术支柱。空间智能基础模型的进步，不仅是技术层面的升级，更是实现具身智能的必经之路。这一趋势表明，具身智能需要从根本上提升感知能力，才能在复杂的物理世界中完成精准操作与推理。

随着技术进步，空间智能也成为投融资领域的热点方向。例如，由“AI教母”李飞飞教授创立的World Labs，在短短几个月内便以构建大型世界模型为目标，专注于生成、感知并交互3D世界。World Labs的定位是解决人工智能领域中最复杂且核心的问题——空间智能。其成立后迅速完成高额融资，公司估值超过10亿美元（约70亿人民币），投资方包括Andreessen Horowitz、英伟达旗下NVentures，以及DeepMind首席科学家Jeff Dean和AI教父Geoffrey Hinton等知名科学家。这表明，空间智能领域已吸引了全球顶尖资本与技术团队的关注。

根据Omdia的最新报告，全球空间计算市场预计在2024年达到45亿美元，并在2029年突破100亿美元，复合年均增长率（CAGR）高达18%。与此同时，泰伯智库预测，到2030年，中国元宇宙市场规模将达到8500亿元，其中与空间计算相关的市场规模将达到3400亿元，占元宇宙市场的40%。这些数据表明，空间智能不仅是人工智能发展的重要技术方向，更将成为推动元宇宙生态和相关产业发展的核心动力。

2. 典型案例

2.1 World Labs发布首个空间智能AI模型

2024年12月，World Labs推出了首个空间智能AI模型，可从单张图片一键生成3D世界。用户只需上传图片，模型便能围绕该图片生成对应的3D虚拟世界。这一技术显著提升了3D内容制作的效率和一致性，特别是在电影、游戏和VR等领域。3D世界生成仅是空间智能的第一步，未来将扩展至更全面的环境感知、理解与推理，最终打造大型世界模型（LWM）。

图 2.3 World Labs的生成实例

2.2 Genie 2：大型世界基础模型

2024年12月，谷歌DeepMind推出了大型基础世界模型Genie 2，在空间智能领域展现出卓越的应用能力，能够通过单张图片或文字描述生成3D场景。通过对大规模视频数据和生成模型的训练，Genie 2能够生成多样化、可交互的3D环境，并模拟物理现象（如重力、光照、反射等）以及长时间视频内容，体现出对空间和时间的综合理解。它支持对象交互、角色动画以及动作控制，即使未使用特定领域的数据也能实现精准模拟。这些功能使Genie 2广泛应用于AI代理的训练与测试、快速原型设计等场景，为AI系统在复杂空间任务中的理解和操作能力提供了创新平台，推动了人工智能的进一步发展。

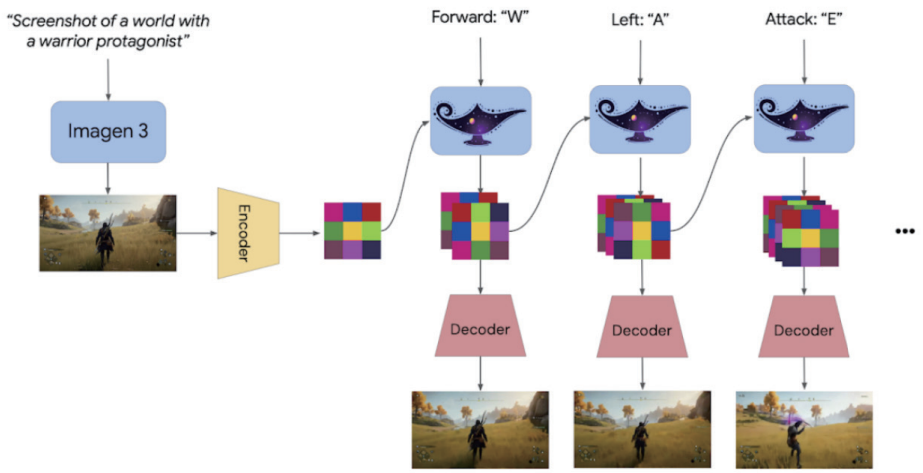

图 2.4 Genie 2的整体流程

2.3 NWM：导航世界模型

2024年12月，Meta的人工智能研究团队（FAIR）推出了导航世界模型（Navigation World Models，NWM），显著提升了AI在复杂环境中的空间智能和导航能力。NWM能够从单张图像生成连续视频，模拟智能体在环境中的移动过程，实现对空间和时间动态的深刻理解。它不仅在已知环境中沿指定轨迹移动表现出色，还能够在未知环境中自主探索路径，并通过结合外部导航策略评估多条潜在路径以选择最优路线。NWM展现了AI在动态和复杂空间中的适应性，为机器人导航、自动驾驶等领域的应用提供了强有力的支持，推动了空间智能的进一步发展。

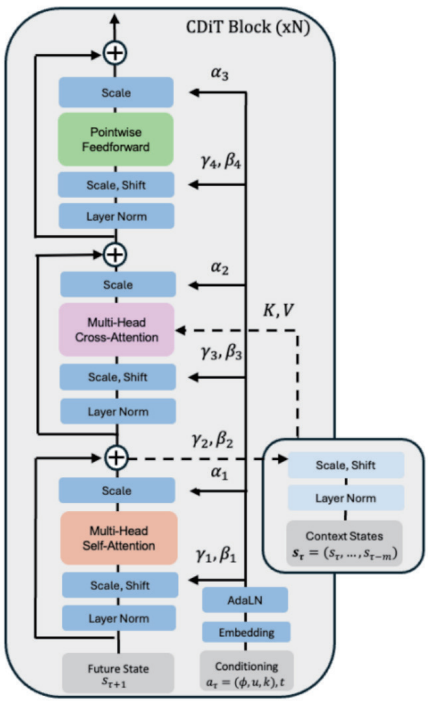

图2.5 扩散 Transformer中的高效的 CDiT 块

三、人形机器人

随着人工智能和自动化技术的飞速发展，政府的高度重视和政策支持为人形机器人行业创造了良好的发展环境。例如，工业和信息化部发布的《人形机器人创新发展指导意见》，明确了人形机器人产业的发展目标和重点任务，为行业快速发展提供了政策保障，并加速了技术的落地与推广。

人形机器人是模仿人类外形与功能的高智能机器人，具备双足行走、抓取物体、操作工具以及与环境自然交互的能力。凭借人工智能、机器学习、传感器技术和材料科学的持续突破，人形机器人已逐步从实验室迈向实际应用。它们能够适应人类生活与工作环境，灵活完成复杂任务，在医疗、养老、服务和制造等领域展现出巨大的应用潜力。此外，通过集成先进的传感器技术、人工智能算法以及柔性执行器的应用，人形机器人能够更加精准地感知和预测周围环境中的潜在风险，并实时调整行为，以确保与人类及环境交互时的安全性。例如，力控算法的进步显著提升了机器人在与人类物理交互中的柔顺性，从而降低误动作导致的冲击与摩擦。这种技术进步不仅提升了机器人的安全性，还加速了人机共融的实现。

随着人形机器人在社会各领域的广泛应用，其安全性和伦理问题也引发了广泛关注。关于机器人决策权的分配、责任划分以及隐私保护等问题，已经成为各国政府、国际组织和学术界讨论的重要议题。各方正积极制定相关的法律框架和伦理规范，确保机器人在各类应用场景中的行为可控、透明且符合伦理道德标准。

1.市场热点/行业前景

近年来，人工智能、传感器、三维仿真和大模型技术的突破，显著提升了人形机器人在复杂环境中的感知、自主性和交互能力。通过集成先进的语音识别、情感识别、自然语言处理等技术，人形机器人实现了更自然的人机交互，并具备了更高效的自主导航和任务执行能力。以特斯拉Optimus、优必选Walker等国内外知名品牌为代表的产品，已经展示了卓越的性能，标志着人形机器人从实验室迈向实际应用。与此同时，生成式AI技术的崛起进一步加速了人形机器人的商业化进程，使其在家庭服务、教育娱乐、智能导览等领域表现出巨大的市场潜力。

政府的重视和政策支持为人形机器人行业提供了坚实的发展环境。例如，《北京市促进通用人工智能创新发展的若干措施》提出，要推动具身智能系统的研究与应用，突破复杂环境中的关键技术；《人形机器人产业研究报告》预测，到2029年，中国人形机器人市场规模将达到750亿元，占全球市场的32.7%，位居世界第一。此外，2024年前十个月，全球人形机器人领域共记录了69起融资事件，融资总额超过110亿元人民币。资金的持续注入为行业研发提供了强劲动力，推动了更多高度灵活且具智能交互能力的机器人产品落地。

市场研究数据显示，全球人形机器人市场规模在未来几年内将快速增长，到2025年有望达到数十亿美元，2029年将进一步突破千亿美元。在中国，2024年机器人市场总规模预计达到4802亿元，其中人形机器人作为重要分支将成为高端制造领域的重要增长点。随着技术进步和生产成本的逐步降低，预计人形机器人将广泛应用于工业、物流、医疗、教育、娱乐等领域，推动相关行业的智能化转型升级。

2. 典型案例

特斯拉Optimus机器人手部已实现更加灵活和拟人，驱动技术方案已经基本定型。Boston Dynamics推出的新一代纯电动人形机器人具有比以往任何一代更强大的力量和更广泛的运动范围，能够执行更复杂的操作和任务。其四肢、躯干和头部都可以360度移动，给予了它极大的运动范围。国内优必选walker系列机器人历多次迭代，具备更快、更稳定的运动能力、更轻更安全的交互以及AI能力。宇树科技的G1人形机器人关节运动角度大，能实现多种复杂动作。

2.1 Figure 02 和特斯拉 Optimus ： 未来智能生活的“高效执行者”

Figure 02 是由人工智能机器人初创公司 Figure AI 发布的第二代人形机器人，部分媒体称其为“地表最强”人形机器人。该机器人采用外骨骼结构，外壳负责承载负载和压力，电源及算力布线集成于机体内部，提升了系统的可靠性和封装紧凑性。Figure 02 配备六个 RGB 摄像头，分别位于头部、胸前和后背，实现高效的视觉感知。其第四代手部装置具备 16 个自由度，拥有与人类相媲美的力量，能够承载高达 25 公斤的重量，灵活执行多种人类类似的任务。内部电池容量提升了 50%，达到 2.25 kWh，确保每日实际有效工作时间超过 20 小时。机器人集成了视觉语言模型（VLM），其计算和 AI 推理能力相比上一代产品提升了三倍，同时搭载了由 OpenAI 定制的语音推理模型，可以通过机载麦克风和扬声器实现与人类无障碍对话。

特斯拉公司开发的 Optimus是面向日常重复性任务的人形机器人项目，旨在推动机器人技术在工业和家用环境中的广泛应用。其最新版本 Optimus Gen 2 在近期进行了展示，表现出卓越的任务执行能力和广阔的应用前景。Optimus Gen 2高约 5 英尺 8 英寸，具备出色的负载能力，可举起 45 磅的物体，并能搬运高达 150 磅的重量。机器人配备 28 个关节驱动器，实现 11 至 22 个自由度，赋予其类人灵活性，能够执行诸如行走、物体分类以及精细操作（例如端茶送水）等复杂任务。通过集成特斯拉自主研发的神经网络与视觉感知系统，Optimus 能够进行自适应学习，而无需依赖逐步的编程指令。其学习方式包括观察人类示范或借助远程操控，实现任务的快速掌握。在近期的技术演示中，Optimus展示了完成家务任务的能力，包括折叠衣物、浇花以及精细操作（如轻柔地处理鸡蛋）。此外，特斯拉展示了Optimus在工厂环境中的应用实例，如完成电池搬运等简单工业任务，进一步验证了其在制造业中的潜力。

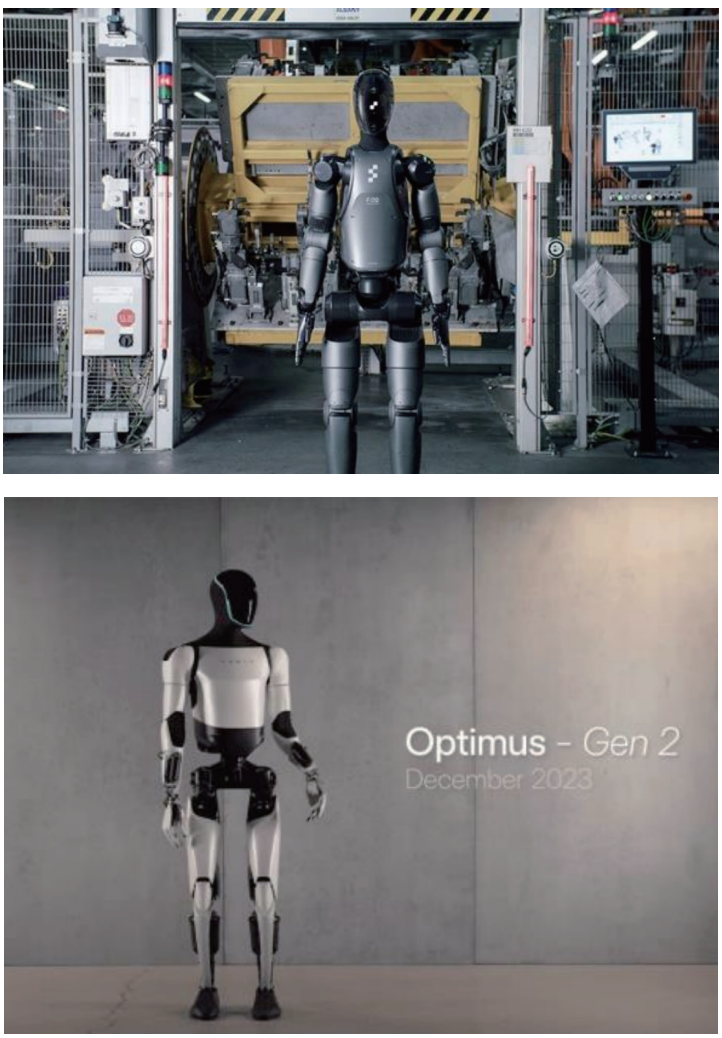

图 3.1 Figure 02机器人（上）；Optimus机器人（下）

2.2 Agility Robotics具身人形机器人在物流搬运的应用

2024年6月28日，Agility Robotics宣布其开发的双足机器人Digit已经在康涅狄格州的Spanx工厂投入使用。这标志着人形机器人首次在客户现场以“机器人即服务”（RaaS）的形式部署，开创了机器人商业应用的新纪元。Digit是一款高5英尺9英寸的双足机器人，能够搬运35磅（15.9千克）的负重。它的设计灵活，具有独特的“后退”腿，可以在各种环境中移动自如。Digit的主要任务是在Spanx工厂内搬运手提箱，具体工作包括从其他机器人那里接过手提箱并将其放置在传送带上。此次部署源于 Agility Robotics 和 GXO Logistics, Inc. 达成的多年期协议，旨在将 Digit 机器人引入 GXO 的多个仓库。根据 RaaS 模式，GXO 将使用一系列 Digit 机器人以及 Agility Arc—— 一个云端自动化平台，来管理和控制这些机器人。Agility Arc 提供完整的机器人控制功能，简化了设施映射、工作流定义、运营管理和故障排除等流程。

图 3.2 Agility Robotics的工作实例

2.3 优必选的Walker S1：人形机器人与无人物流车等协同作业

2024年10月，优必选发布的新一代工业人形机器人Walker S1，率先实现了与无人物流车、无人叉车和工业移动机器人等设备的协同作业，成为全球首个在工业场景中落地的综合解决方案。Walker S1通过软硬件全面升级，包括一体化关节技术、集成化头部设计和第三代仿人灵巧手，显著提升了其在复杂非结构化环境中的任务执行能力。同时，优必选自主研发的ROSA2.0操作系统和多模态规划大模型为机器人提供了高效的导航和任务规划能力。Walker S1已广泛应用于比亚迪等多家车厂，成功攻克工业场景中的关键难题，累计意向订单超过500台，展现了人形机器人在智能制造领域的巨大潜力，推动制造业高质量发展并缓解劳动力短缺问题。

图3.3 人形机器人与无人物流车等协同作业概念图

2.4 五八智能具身人形机器人在3C制造的应用

2024年7月，五八智能突破手身眼协同灵巧操作技术，完成自主物料搬运、开门、扫码贴签等任务操作，与长虹集团合作，在国内首次完成人形机器人在3C场景应用验证，受到央视《新闻联播》报道。

图 3.4 五八智能具身人形机器人的工作实例

2.5 拟人助老机器人

2024年9月24日，腾讯Robotics X实验室发布了最新研发的人居环境机器人“5号”（The Five，小五）。小五采用四腿轮足复合设计，结合自研双编码器大扭矩密度执行器和覆盖180个检测点的大面积触觉皮肤，具备行走、搬运物体等能力，并可通过自适应算法应对楼梯、斜坡、波浪坡等复杂地形。小五基于统一的控制框架，搭载激光雷达和IMU等传感器，结合高精度SLAM系统实现实时定位和环境建图，在养老院室内外场景中展现了精准的地形识别和路径规划能力。其负载能力显著提升，双臂可抱扶承重50千克，每条直线腿可单独伸缩，支持“上摸高、下摸地”的广阔作业空间，能够帮助用户取放高处物品或低矮空间操作。小五还具备多模态人机物理交互能力，可辅助完成抱扶老人等任务，在实验室环境下展现了强大的运动、感知和交互能力，为智能家居和人机共生发展提供了重要支持。

图 3.5 小五机器人的工作实例

四、大规模仿真训练平台

大规模仿真训练平台是人工智能与机器人技术发展的关键基础设施，旨在通过高精度的物理模拟和大规模数据生成，提升智能机器人的研发效率和性能表现。这些平台的核心功能包括支持物理环境模拟、生成高保真训练数据以及并行训练大规模模型，从而满足非结构化环境下复杂任务对机器人智能感知与控制的高要求。

随着大模型的兴起，十亿乃至百亿级参数模型在文本生成、图片生成、对话交流等领域展现出强大能力，而如何将这种智能能力迁移到机器人以完成现实世界中的复杂任务和交互操作，成为一项亟待解决的科学问题。高性能算力平台的发展为此提供了基础支持。例如，NVIDIA的A系列和H系列高性能显卡以及定制化算力平台，使得在三维物理环境中的大规模模型训练成为可能，大幅提升了仿真效率和智能机器人开发的可行性。

新一代仿真训练平台，例如ETHZ开发的RaiSim和NVIDIA推出的Isaac Sim，凭借高保真物理模拟与强大渲染能力，已成为现代智能机器人研发的核心工具。这些平台不仅能够为强化学习和深度学习模型生成难以通过现实采集获得的高质量训练数据，还支持虚实融合仿真技术，通过结合真实场地数据动态修正仿真模型，进一步提升了仿真精度和机器人适应复杂环境的能力。

1.市场热点/行业前景

大规模仿真训练平台市场正迎来前所未有的增长。智能机器人在非结构化环境中完成复杂任务的需求促使企业加快开发迭代，而仿真平台为此提供了高效、低成本的解决方案，成为行业竞争的重要工具。高性能算力平台的涌现为仿真平台的发展奠定了硬件基础。

新一代虚实融合仿真平台通过整合真实数据与仿真环境，大幅提升仿真精度，并缩短了智能机器人从研发到部署的周期。生成的高保真训练数据使机器人能够更好地适应复杂的真实环境，显著增强其实际操作的可靠性和性能。这类平台已成为现代智能机器人开发的重要趋势。仿真平台的应用已从传统机器人领域扩展至自动驾驶、智能制造、医疗、航空航天、智慧城市等多个行业。例如，虚实结合的仿真技术支持无人驾驶系统在极端天气等复杂环境中的测试，并为工业机器人优化精密装配流程提供了训练工具。这种多场景适应能力极大拓宽了仿真平台的市场潜力。

大规模仿真平台正与5G通信、云计算、数字孪生等前沿技术深度结合，推动了智能机器人性能的跨越式发展。这种技术融合不仅提升了仿真训练的实时性和扩展性，还为机器人在复杂环境中的自主学习与高效执行创造了更多可能性。未来，随着仿真训练平台技术的不断升级，其在智能机器人开发中的地位将更加突出。基于仿真的高效训练流程将进一步加快机器人迭代速度，并为复杂环境中的具身智能机器人提供坚实的技术基础。通过提升性能极限，大规模仿真技术将加速人工智能与机器人技术的深度融合，成为推动智能系统商业化应用的重要驱动力。

2. 典型案例

2.1 “通境”(TongVerse)平台：视觉-语言-运动联合解译架构

北京人工智能通研院（BIGAI）开发了一个名为TongVerse的仿真平台。这个平台是一个“AI+机器人”仿真训练场，为具身智能提供安全、可控的仿真环境。TongVerse平台支持多种类型的机器人（如人形机器人、复合协作机器人）的视觉-语言-运动联合解译。它还支持动态开放环境下的机器人动力学仿真，无论是机器人的全身运动控制、步态规划还是操作作业，都能得到精确的模拟。

这个平台在2024年CRAIC人形机器人创新挑战赛中首次实现了从模拟场景跨入真实家庭服务场景的创新突破。TongVerse平台为机器人提供了丰富的任务训练环境，并支持对具身智能体的智能能力进行全面测试。除了应用在科研、赛事等多场景中为人形机器人训练等提供支持，未来TongVerse平台还将面向智能制造、特种行业等不同需求，提供多场景、多任务的机器人应用解决方案。

图 4.1 “通境”仿真平台演示

2.2 NVIDIA Isaac Sim：加速机器人开发的一站式仿真平台

NVIDIA Isaac Sim 是 NVIDIA 推出的机器人仿真和合成数据生成平台，专为开发者设计，帮助他们高效完成基于 AI 的机器人及自主机器的设计、仿真、测试和训练任务。该平台建立在 NVIDIA Omniverse 平台之上，具备高度的可扩展性，并集成了丰富的功能和工具。通过 NVIDIA PhysX 引擎提供高保真的物理仿真，支持多种传感器（如摄像头、LiDAR 和接触传感器）的精准模拟。平台内置的 Replicator 工具可生成高质量的合成数据，为 AI 模型训练提供有力支持。此外，Isaac Sim 预置了多种第三方机器人模型和 SimReady 3D 资产，帮助开发者快速构建复杂的仿真场景。同时，平台提供了丰富的开发工具和 API，支持与 ROS 和 ROS2 的桥接功能，实现与实际机器人设备的无缝通信和集成，为开发者提供了一站式高效解决方案，加速机器人从研发到落地的全流程。

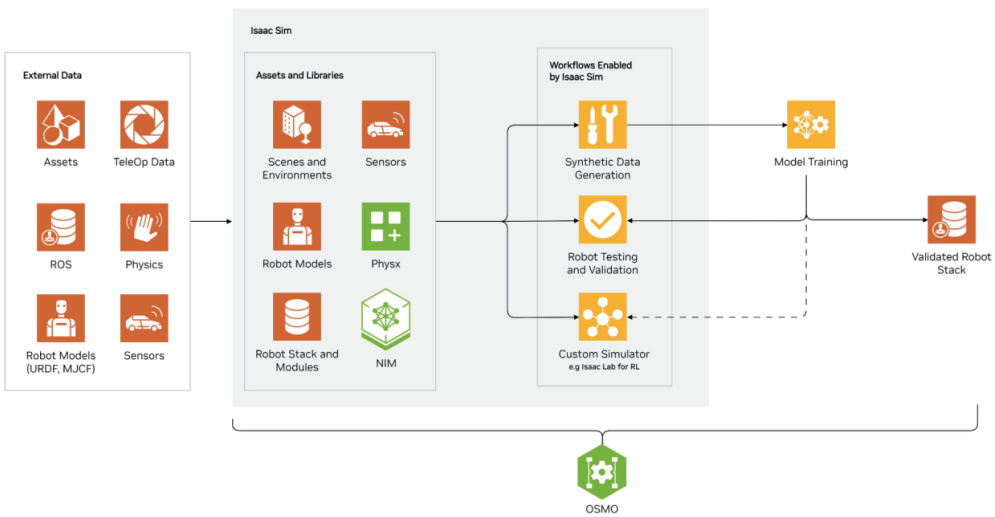

图 4.2 Isaac Sim工作流程

2.3 Taichi：高效仿真计算的核心驱动力

Taichi是一个高性能、开源的编程框架，专注于高效的数值计算和物理仿真，在仿真平台领域展现出巨大的应用潜力和技术优势。其设计灵活的编程模型使开发者能够以接近 Python 的语法编写高效代码，同时借助 GPU 加速和并行计算能力，大幅提升了计算性能，满足了大规模仿真任务对高效性的需求。此外，Taichi 对多种硬件架构（如 NVIDIA 和 AMD 的 GPU、Apple M1 芯片等）的广泛支持，使其能够适应多样化的部署场景。

在仿真平台领域，Taichi 已被广泛应用于机器人仿真、虚拟环境构建和强化学习数据生成等场景。它能够通过高保真的物理模拟为智能机器人提供关键技术支持，例如模拟机器人抓取物体的受力行为、触觉反馈及复杂交互，从而提高机器人在真实世界中的任务精度和适应性。在虚拟环境构建中，Taichi 高效生成高保真虚拟场景，例如流体、粒子和柔性体动力学模拟，为仿真训练和测试提供逼真的物理环境。在强化学习领域，Taichi 能够快速生成大规模高质量的仿真数据，为 AI 模型提供高效的训练支持，并显著提升其对复杂任务的适应能力。

此外，Taichi 的应用还覆盖粒子模拟和柔性体动力学等领域。它能够模拟沙土、液体等粒子物质的运动行为，以及布料、绳索等柔性材料的动力学特性，为仿真平台带来了丰富的场景支持。目前已有利用Taichi完成的扩展应用，如Tacchi系列视触觉仿真器。Tacchi是基于Taichi的一系列高效的视触觉仿真器，专注于刚性、弹性、塑性和弹塑性物体在视触觉传感器按压、旋转和滑移状态下触觉图像的生成。借助Taichi高效的运算效率，Tacchi可以高效的生成可靠的触觉数据，这些触觉数据有利于扩充和丰富多模态数据集，在具身机器人控制、多模态表征和触觉三维重建等方面具有应用潜力。

凭借其灵活性、性能和跨平台能力，Taichi 已成为推动仿真技术发展的重要工具。随着智能机器人、虚拟现实和智能制造等领域对高效仿真计算需求的不断增长，Taichi 在仿真平台领域的未来前景十分广阔，预计将在加速仿真平台创新与落地方面发挥更加重要的作用。

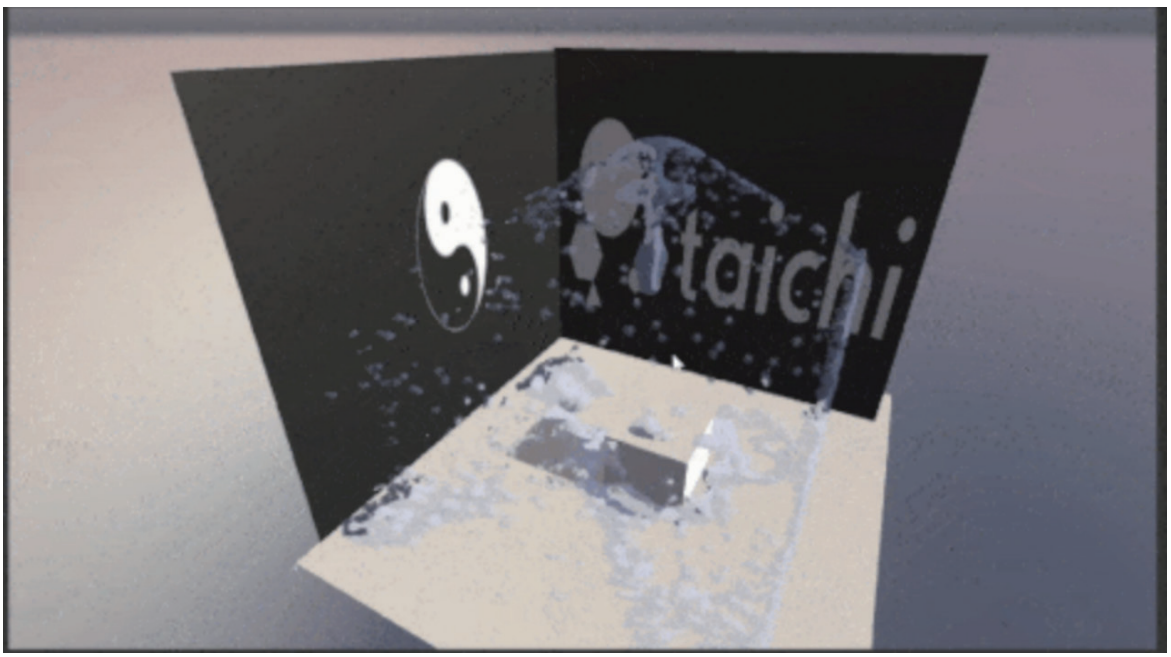

图 5.3 Taichi仿真平台实例

2.4 Genesis：机器人仿真新时代

Genesis是一个用于通用机器人学习的生成式和可微分的物理引擎，提供了一个统一的模拟平台，支持各种材料的模拟，能够模拟广泛的机器人任务，同时完全支持可微分特性。这将大幅度的提升model-based的机器人训练策略，大幅度的提升机器人的技能学习效率。与此同时，其具有语言交互性的特点也将成为新时代物理仿真器的一个重要特性。

Genesis平台不仅限于解决机器人领域的问题，还能惠及更多行业。虽然最初是为了提供机器人学所需的数据，但实际上，这些数据具有广泛的通用性。机器人学的数据包括静态数据（如任务描述、环境特征及其交互方式）和动态数据（如学习到的策略与环境交互）。借助物理引擎和前向模拟，可以生成各种动态数据，这些数据格式适用于多种应用场景。

例如，视频生成可以通过Genesis平台得到拓展应用。与现有的基于扩散模型的逐帧生成方法不同，Genesis平台通过构建三维场景、引入演员、设置摄像机参数和轨迹，再加上强大的物理引擎和先进的渲染技术，能够在虚拟世界中再现视频拍摄过程。这不仅能生成机器人数据，还能生成人物角色的动作、面部表情等，以及其他参数如光线强度、镜头焦距、运动轨迹等。这样就能生成高度物理精确的视频数据，提供了一种全新的视频生成方法。

通过这种方法，不仅可以直接生成视频，还可以用生成的数据来训练基于学习的模型。此外，平台能够通过自然语言描述控制场景中的各个元素，确保生成的数据具有高度一致性和细粒度对齐。这种高质量的数据生成能够为视频生成、互动场景以及4D视频数据等多个领域提供新机遇。

图　5.4 Genesis平台生成的机器人视频实例

五、触感灵巧手

灵巧手和具身触觉智能作为实现具身智能的关键技术，正在深刻改变机器人对物理世界的感知与交互方式，并展现出广泛的应用前景。灵巧手以其高自由度、灵活多变和高度仿生的特点，赋予机器人精细的操作能力，使其能够执行复杂任务，如抓取、操纵物体等。这种能力不仅是具身智能在物理世界中发挥作用的基础，更是推动智能体行动中思考、以实际操作解决问题的核心要素。通过灵巧手，机器人能够在生活服务、工业制造、医疗手术、特种排爆、抢险救援等多个领域展现出惊人的潜力。

与此同时，具身触觉智能进一步增强了机器人的感知和交互能力。作为人类智能核心体现的延伸，具身触觉智能使机器人具备感知物体形态、质地、温度等多种属性的能力，并通过力反馈实现对动态交互的精准控制。这种能力不仅模仿并超越了人类触觉的敏感与细腻，还弥补了视觉感知的不足，使机器人能够通过触觉探索物体的内部特性，优化交互方式，提高自主学习能力。

灵巧手与具身触觉智能相辅相成，共同推动了具身智能的发展。灵巧手为机器人提供了精细操作的硬件基础，而具身触觉智能则通过多模态信号（如压力、滑动、湿度、温度、震动等）的融合与解析，为机器人对物理环境的深度理解和高效适应提供了重要支持。具身触觉智能帮助机器人调整抓取力度以避免损坏物体，同时通过触觉信号的精准建模与解析，实现从表面感知到多维动态建模的跨越。

总体而言，灵巧手和具身触觉智能共同为机器人赋予了更加自然、高效的感知与操作能力，使其能够在动态环境中以智能、灵活和自适应的方式执行任务。这种结合不仅推动了人工智能技术从抽象认知向具身实践的转变，也为未来人机共融的触觉交互时代奠定了基础。随着相关技术的不断发展，机器人将在物理世界中展现出更广泛的应用潜力，开启一个更加智能、互联的新时代。

1.市场热点/行业前景

截至2024年上半年，全球机器人灵巧手市场容量达到66.69万只，市场规模达15.07亿美元，同比增长超过13%。与此同时，具身触觉领域也展现出强劲的发展势头，截至2024年前三季度，披露的融资事件达35起，累计融资金额高达31.5亿元。这些数据凸显了机器人灵巧手和具身触觉智能在全球范围内的快速发展以及市场对其的高度关注。

随着大语言模型、人工智能和具身智能技术的进步，人形机器人对手部精细操作的需求显著增加，推动了灵巧手市场的需求增量。国家和地方政府也高度重视这一领域的发展，出台多项支持政策，助力行业创新。在此背景下，多家企业推出了一系列灵巧手产品，推动其向智能化和拟人化方向演进。同时，具身触觉技术作为人工智能发展的重要方向，凭借触觉感知与交互能力的不断突破，进一步拓展了灵巧手和智能体的应用场景。

灵巧手和具身触觉技术相辅相成，共同赋能机器人更智能、更精细的操作能力。灵巧手作为机器人与环境交互的重要部件，通过拟人级别的高级触觉传感器与智能算法，实现对物体形态、质地、温度等属性的精准感知与操作。这不仅提高了工业制造中的效率和精准度，还在医疗康复、危险环境作业、太空探索等领域展现出广泛的应用潜力。而具身触觉技术则通过对多模态信息（如压力、湿度、震动等）的融合与解析，为机器人提供实时反馈，优化交互方式，进一步提升了复杂任务中的操作精度。例如，触觉感知能够在工业场景中优化精密装配流程，在医疗领域确保远程手术的安全性和可靠性。

近年来，触觉感知技术的进步体现在传感器设计、感知模型以及开发平台的创新上。例如，Meta推出的Digit Plexus平台将指尖与手掌传感器整合，为制造业和医疗领域的触觉控制提供更高精度；腾讯Robotics X实验室的人居环境机器人则结合触觉与视觉技术，为人类提供个性化的辅助服务。这些技术的迭代不仅增强了机器人的环境适应能力，还推动了从工业制造到智能服务、医疗康复等多领域的场景拓展。

投融资热潮进一步加速了灵巧手与具身触觉技术的商业化落地。众多行业巨头和技术团队积极布局这一领域，例如戴盟机器人专注于视触觉传感器与多模态操作模型的研发，迅速获得资本青睐。根据IDTechEx预测，触觉技术市场规模将在2024年至2035年间以复合年增长率增长，2035年将达到71亿美元。这表明，灵巧手与具身触觉技术不仅是未来人工智能系统的重要技术支柱，还将在解决劳动力短缺、提升生产效率、优化用户体验等方面发挥重要作用。

展望未来，随着人工智能、物联网、5G等技术的深度融合，灵巧手和具身触觉技术将不断突破技术边界，推动人机交互模式的变革。从工业制造到家庭服务，从医疗康复到特种作业，这些技术将在更多场景中实现落地应用，为智能制造与人性化服务注入新动能，引领全球人工智能迈向更加智能、精细化的新时代。

2. 典型案例

2.1 特斯拉 Optimus 触感灵巧手

特斯拉Optimus灵巧手拥有22个自由度，实现了接近人类手的灵活性，能够完成复杂的抓握和操作任务。采用绳驱的方式，使其关节运动更精准，同时操作速度显著提升（如最新一代操作响应时间提高了250ms）。Optimus触觉灵巧手通过集成力传感器和触觉传感器，能够全面感知物体的质地、硬度以及抓握力度，从而在处理复杂任务时表现得更加灵活可靠。例如，Optimus能够在演示中轻松拿起易碎物品如鸡蛋，展示了出色的精细操作能力。与视觉传感器相比，触觉传感器的优势在于更适合手部操作空间中可能存在的遮挡情况，能够准确感知和处理周围环境信息，从而提升其适应性和操作效率。此外，特斯拉还为Optimus灵巧手设计了先进的机器学习算法，使其能够通过模拟和学习人类手的动作不断提升操作技能。这种结合感知与学习的技术，不仅让Optimus在外观和功能上接近人类手，更使其能够以高精度完成制造业、医疗护理和家庭服务等领域的多样化任务。

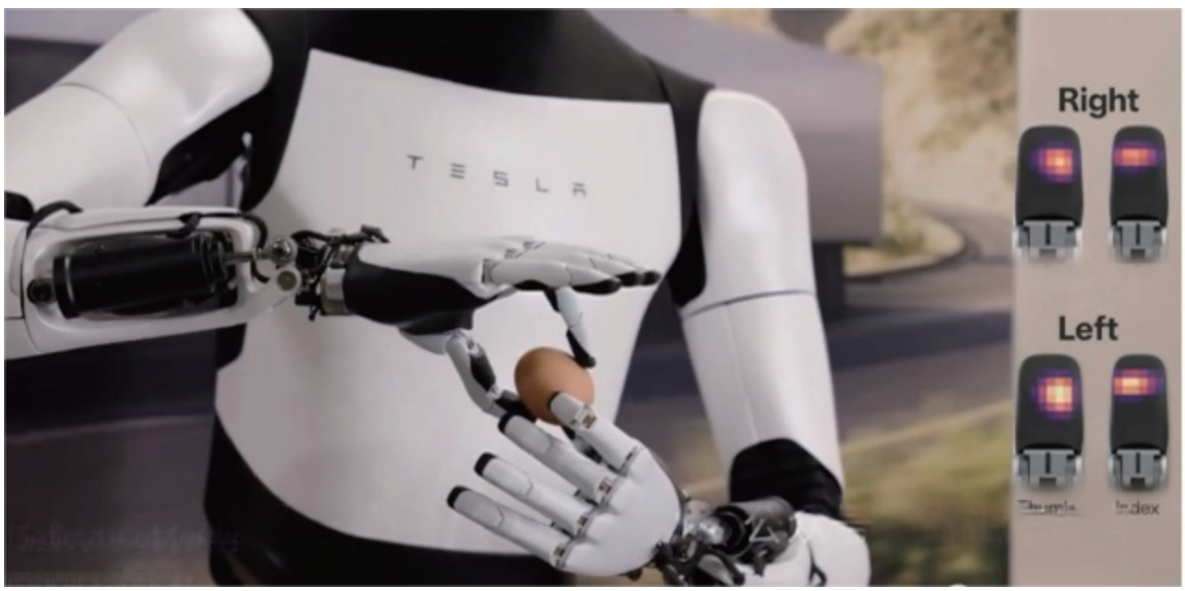

图 5.1 Optimus灵巧手实例

2.2 Linker hand

灵心巧手公司推出的Linker Hand灵巧手，是目前性价比最高的高自由度灵巧手产品，具备20个主动自由度。该产品配备了多传感器系统，包括柔性电子皮肤，实现精细触觉感知。技术创新方面，公司自主研发了微型谐波关节等关键技术，大幅提升性能，降低成本。此外，灵心巧手公司还在构建全球最大的灵巧操作数据集，包含了大量的人手操作数据，覆盖了各种复杂的抓取和操作任务，将为灵巧手的控制和应用提供重要的支持。

图 5.2 Linker hand灵巧手

2.3 因时RH56系列灵巧手

北京因时机器人科技有限公司的RH56系列仿人五指灵巧手是一款具有高度灵活性和实用性的机器人产品。它具备六个自由度和10N的指尖抓力，能抓取2-3公斤的物品，适应不同场景需求。通讯接口多样，支持RS232、RS485或CAN，且配备压力传感器以感知力度。编程软件兼容性强，适用于多种设备。这款灵巧手在服务机器人和医疗假肢领域有广泛应用，尤其在假肢应用中，通过肌电传感器实现智能控制。因时机器人公司凭借其专业技术，为多个行业提供了高性能的核心运动部件，展现了其在机器人领域的创新实力和应用潜力。

图 5.3 因时RH56系列灵巧手

2.4 Freedom仿人五指触感灵巧手

清华大学孙富春教授团队孵化的清瑞博源智能科技河北有限责任公司，开发的Freedom仿人五指触感灵巧手是一款专为人型机器人及机械臂设计的末端操作工具，具有整机重量轻、单指指尖抓取力强的特点。其指尖配备了多点阵列压力传感器，能高精度执行多种抓取操作，智能完成大部分人体手部动作。其触觉传感器通过复合传感结构和柔性技术实现，具有柔性好、抗疲劳性强、动态范围大等特点。在多次抓取测试中，各手指指尖输出力稳定，数据曲线平滑。Freedom灵巧手适用于工业生产特殊环境，能对复杂形状物体进行自适应抓取及复杂任务操作。

图 6.5 Freedom仿人五指触感灵巧手

2.5 灵巧手的视触觉传感技术

2024年，在灵巧手视触觉传感技术领域取得了一系列突破。首先，千觉机器人开发了高精度的多模态感知算法，包括触觉传感变形场感知、三维分布力感知、滑动感知等，并研发了闭环控制算法，为机器人细操作提供了算法支持。随后，这些算法与传感器技术和多种机器学习方法相结合，形成了包括RoboFusion和UniTouch在内的先进系统。其中，RoboFusion通过融合视觉、触觉、力觉等多模态传感器数据，结合自监督学习与强化学习，在工业生产和医疗康复领域完成了精密零件抓取、装配、导航等任务；UniTouch通过对比学习对齐触觉与视觉信号，在触觉抓取预测和触觉问答等任务中展现出色性能。近期北京交通大学联合北京邮电大学团队发布了首个大规模触觉、多粒度语言、视觉三模态数据集Touch100k，并提出TLV-Link预训练方法，为材料属性识别和抓取预测任务提供了高效的触觉表示能力，特别是在零样本触觉理解方面取得显著进展。这一系统化研究路径全面推动了灵巧手视触觉技术的发展，为其在复杂环境中的广泛应用奠定了坚实的基础。

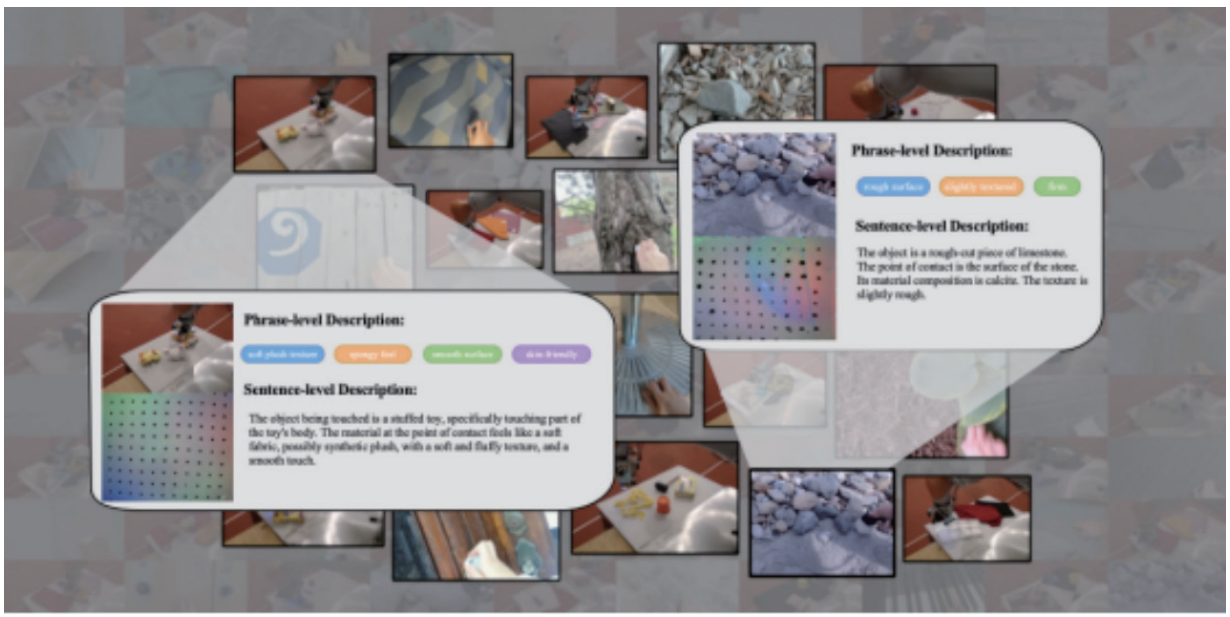

图 5.5 Touch100k数据集概述

六、具身智能导航大模型

具身智能导航大模型作为机器人智能化发展的重要里程碑，不仅实现了多模态感知、语言理解、路径规划与动作控制的协同，还通过深度学习与自我优化技术，进一步提升了机器人的认知能力和操作效率。依托先进的多模态传感器融合技术，机器人能够实时获取来自视觉、触觉、力觉、激光雷达等多种传感器的数据，构建具有精确语义标签的三维环境模型。这种多模态感知能力让机器人可以不仅识别物体和场景，还能理解其功能和关系，从而在动态变化的环境中做出更加精准的决策。

语言理解方面，大模型将自然语言指令解析为机器可执行的任务表示，并通过与环境模型的结合，使机器人能够准确理解任务目标背后的语境与意图。这种能力不仅增强了人机交互的自然性，还让机器人具备了更高的任务泛化能力。例如，机器人可以通过解析模糊指令（如"整理房间"）自动识别任务的多种子步骤，并在执行过程中动态调整策略。

路径规划和动作控制的高度结合，使机器人能够在动态环境中有效避开障碍、应对突发情况。大模型通过全局路径规划算法为机器人生成可行的宏观路径，并结合局部规划算法在复杂环境中优化微观动作。低级动作控制模块则依赖实时传感反馈，对机器人运动进行精细化调整，从而实现高效、精准的任务执行。

1.市场热点/行业前景

具身智能导航大模型作为人工智能与机器人技术的深度融合成果，凭借其多模态感知、自然语言交互和高效导航能力，正成为推动机器人行业智能化发展的核心技术之一。这类大模型结合视觉、触觉、力觉等多模态传感器技术，能够精准感知环境，同时通过自然语言理解指令并进行动态决策，显著提升了机器人在复杂场景中的操作效率和自主决策能力。

具身智能导航大模型在多个领域展现了广泛的应用前景。例如，在智能家居中，它可以帮助机器人高效完成清洁、物品分类和家庭安全监控；在无人配送中，它能够在动态城市环境中准确规划路径，完成精准配送；在服务机器人中，它可以辅助医院患者护理、商场客户服务等；而在工业自动化中，该技术能优化机器人在生产线上的协作操作，实现高精度的任务执行。此外，在物流、零售和辅助服务等快速发展的行业中，导航大模型通过降低部署适配成本，提升操作灵活性和环境适应能力，为企业节省了大量时间和资金成本。

随着人工智能技术的不断进步以及政策支持力度的加强，具身智能导航大模型的研发和应用进入了快速发展期。国家和地区层面陆续出台支持政策，加速智能制造与服务领域的技术升级。与此同时，大模型的普及正在推动机器人行业标准的统一化和生态建设，助力相关企业拓展全球市场。

展望未来，具身智能导航大模型有望成为机器人行业的重要驱动力。通过与5G、物联网、云计算等前沿技术的深度融合，导航大模型将进一步扩展机器人在教育、医疗、农业、建筑等领域的应用场景，为社会的智能化转型提供有力支持。尤其在面向高复杂度任务的环境中，这一技术将通过不断优化感知、交互和操作能力，开启机器人自主决策和高效执行的新纪元，其发展前景极为广阔。

2. 典型案例

2.1 InstructionNav:Unexplored环境下通用指令导航的零样本系统

2024年6月，北京大学前沿计算研究中心董豪课题组主导完成通用指令导航大模型系统InstructNav。由于在理解和遵循自然语言形式指令进行导航的过程中，机器人常常会执行错误的动作或步入错误的房间。此时，具有纠错能力的导航高层规划方法就显得十分必要。受到大语言模型思维链路机制的启发，研究提出Chain-of-Navigation机制，该机制准确把握指令导航过程中的要素"动作"和"地标"，引导大语言模型根据导航指令以及实时的场景视觉信息，对下一步导航动作和地标进行更新。这种闭环更新的策略，不仅能够纠正物体导航和人类需求导航过程中探索与目标物体弱相关区域的情况，还能大幅度减少视觉语言导航过程中执行错误动作或提前停止的问题，有效提升导航规划的正确率。基于Chain-of-Navigation规划和纠错机制，InstructNav无需任何训练，即可超越许多仅支持特定指令的导航方法。

图 6.1 InstructionNav整体流程

2.2 整体协调的移动操作：同时感知、交互和导航的新范式

2024年CVPR会议上，由CMU的Deepak Pathak教授团队发表的研究成果SPIN（Simultaneous Perception, Interaction and Navigation），为移动操作机器人在复杂环境中的自主导航和操控开辟了新路径。SPIN通过强化学习训练单一模型，实现了机器人底盘和手臂的低级控制，并预测机器人的自我中心相机在每个时间步应该看向何处，同时通过全身避障移动。该模型利用主动视觉系统有意识地感知并反应环境，类似于人类利用全身和手眼协调的方式，SPIN开发的移动操作机器人能够利用其移动和视觉的能力，即为了看到而移动，为了移动而看到。实验结果表明，SPIN在多种室内外场景中表现出色，能够在只有自我视觉的情况下，无需创建环境地图，灵活地协调全身动作，穿越复杂的杂乱环境。该研究不仅展示了移动操作机器人在动态环境中的适应性和灵活性，还证明了数据驱动方法在解决传统非反应式规划方法中的挑战方面的潜力。

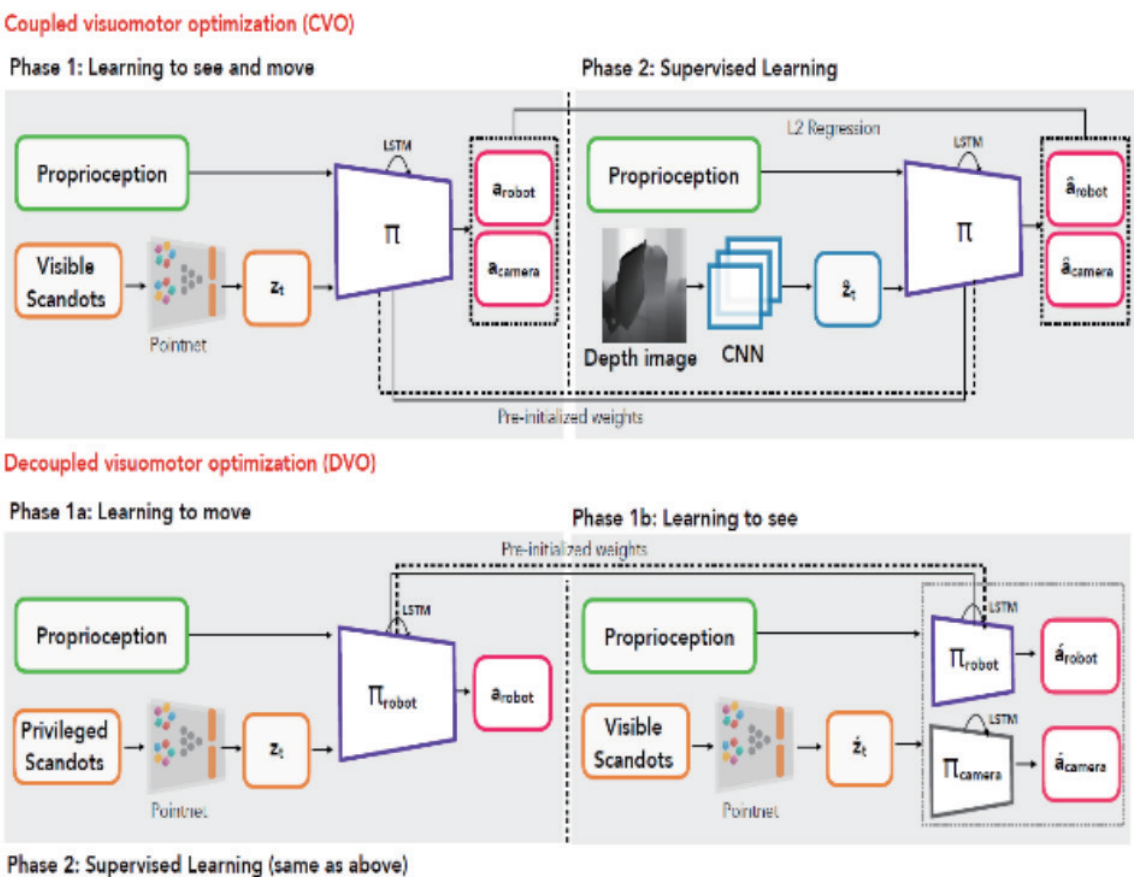

图 6.2 SPIN整体流程

2.3 Uni-NaVid: 一种用于统一具身导航任务的视频视觉-语言-动作模型

2024年12月，北京大学人工智能学院提出Uni-NaVid模型，这是首个基于视频的视觉-语言-动作（VLA）模型，旨在统一多种具身导航任务，并在未知的真实世界环境中实现对混合长时任务的无缝导航。Uni-NaVid 通过统一所有常用具身导航任务的输入和输出数据配置，将所有任务整合到单一模型中，从而实现这一目标。为训练 Uni-NaVid，研究从四个重要的导航子任务中共收集了360万条导航数据样本，促进跨任务的学习协同作用。在全面的导航基准上进行的大量实验清晰地展示了 Uni-NaVid 在统一建模方面的优势，并表明其达到了当前最先进的性能。此外，真实世界的实验验证了该模型的有效性和高效性，显示出其出色的泛化能力。

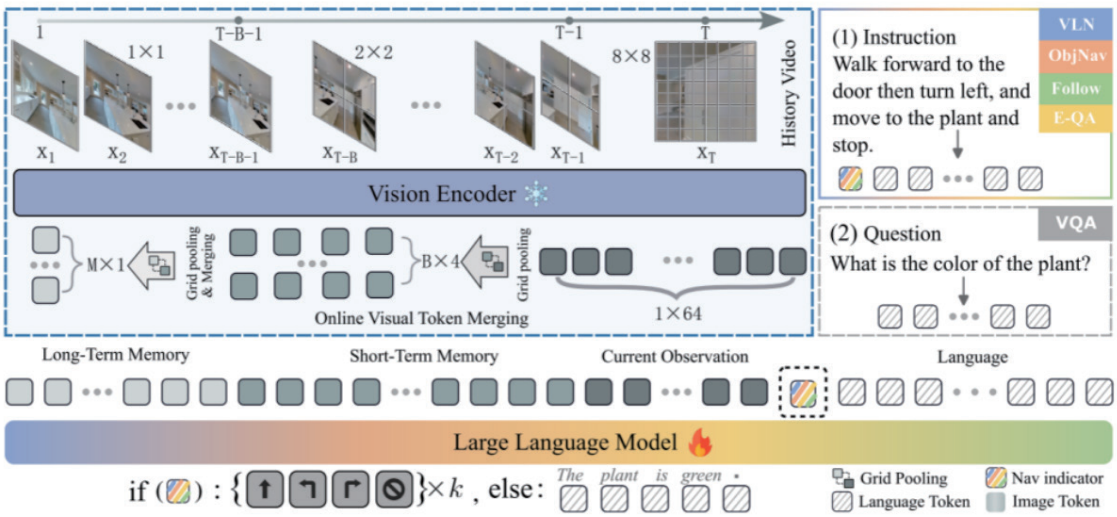

图 6.3 Uni-NaVid整体流程

参考文献

[1] Zitkovich B, Yu T, Xu S, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control[C]//Conference on Robot Learning. PMLR, 2023: 2165-2183.

[2] Liu, J., Liu, M., Wang, Z., An, P., Li, X., Zhou, K., ... & Zhang, S. RoboMamba: Efficient Vision-Language-Action Model for Robotic Reasoning and Manipulation. In The Thirty-eighth Annual Conference on Neural Information Processing Systems

[3] Brohan A, Brown N, Carbajal J, et al. Rt-1: Robotics transformer for real-world control at scale[J]. arXiv preprint arXiv:2212.06817, 2022.

[4] Li X, Liu M, Zhang H, et al. Vision-language foundation models as effective robot imitators[J]. arXiv preprint arXiv:2311.01378, 2023.

[5] Jin Y, Li D, Yong A, Shi J, Hao P, Sun F, ... & Fang B. Robotgpt: Robot manipulation learning from chatgpt. IEEE Robotics and Automation Letters.

[6] https://deepmind.google/discover/blog/genie-2-a-large-scale-foundation-world-model/

[7] https://arxiv.org/pdf/2412.03572v1

[8] Eyes Wide Shut? Exploring the Visual Shortcomings of Multimodal LLMs, FAIR, 2024

[9] https://www.worldlabs.ai/blog

[10] Liu M, Chen Z, Cheng X, et al. Visual whole-body control for legged loco-manipulation[J]. arXiv preprint arXiv:2403.16967, 2024.

[11] Uppal S, Agarwal A, Xiong H, et al. SPIN: Simultaneous Perception Interaction and Navigation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 18133-18142.

[12] Suresh S, Qi H, Wu T, et al. NeuralFeels with neural fields: Visuotactile perception for in-hand manipulation[J]. Science Robotics, 2024, 9(96): eadl0628.

[13] Carolina Higuera, et.al. Sparsh: Self-supervised touch representations for vision-based tactile sensing, CoRL, 2024

[14] https://www.youtube.com/watch?v=Y5zv0aJqMYE

[15] https://www.idtechex.com/en/research-article/haptics-technology-market-to-grow-to-us-7-1b-by-2035/31731

[16] F. Yang et al., "Binding Touch to Everything: Learning Unified Multimodal Tactile Representations," 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2024, pp. 26330-26343, doi: 10.1109/CVPR52733.2024.02488.

[17] Ning Cheng and Changhao Guan and Jing Gao and Weihao Wang and You Li and Fandong Meng and Jie Zhou and Bin Fang and Jinan Xu and Wenjuan Han. Touch100k: A Large-Scale Touch-Language-Vision Dataset for Touch-Centric Multimodal Representation, arXiv preprint arXiv:2406.03813, 2024.

[18] He H, Bai C, Pan L, et al. Learning an actionable discrete diffusion policy via large-scale actionless video pre-training [C]//The Thirty-eighth Annual Conference on Neural Information Processing Systems. 2024.

[19] Zhou, X., & Mu, Y. (2023). Tree-Structured Trajectory Encoding for Vision-and-Language Navigation. Proceedings of the AAAI Conference on Artificial Intelligence, 37(3), 3814-3824. https://doi.org/10.1609/aaai.v37i3.25494

联系我们

中关村智友研究院