arXiv:2602.01644v1 [cs.LG] 2 Feb 2026

从感知到动作：空间人工智能智能体与世界模型

GloriaFelicia NolanBryant Handi PutraAtlasPro AI AtlasPro AI AtlasProAI gloria.felicia@atlaspro.ai nolan.bryant@atlaspro.ai handi.putra@atlaspro.ai

Ayaan Gazali Eliel Lobo EstebanRojasAtlasPro AI AtlasProAI AtlasProAI ayaan.gazali@atlaspro.ai eliel.lobo@atlaspro.ai esteban.rojas@atlaspro.ai

摘要

尽管大语言模型已成为智能体推理与规划的主流方法 [Brown 等人, 2020, OpenAI, 2023, Touvron 等人, 2023, Team and Google, 2023, Anthropic, 2024, Dubey 等人, 2024, OpenAI, 2023, Anil 等人, 2023, Chowdhery 等人, 2022, Jiang 等人, 2023, Abdin 等人, 2024, Grattafiori 等人, 2024, Devlin 等人，2019], 但它们在符号领域的成功并不能轻易迁移到物理世界。空间智能是具身智能体的一项重要能力，指感知三维结构、推理物体关系并在物理约束下采取动作的能力 [Chen 等, 2024a, Yang 等人, 2025, Duan 等人, 2022, Amin and Kiela, 2024, Cheng 等人, 2025, Guo等人, 2024c, Liu 等人, 2024c]。现有综述要么单独讨论智能体架构，要么单独讨论空间领域，没有一篇提供连接这两种互补能力的统一框架。本文填补了这一空白。通过全面回顾 2,000 多篇论文，并引用顶级会议和期刊上的 742 篇研究成果，我们提出一种统一三轴分类法，将智能体能力与跨尺度的空间任务联系起来。重要的是，我们区分了空间接地（对几何和物理的度量理解）与符号接地（将图像与文本关联），并指出仅有感知并不能赋予智能体能力。我们的分析揭示了与这三条轴对应的三项关键发现：（1）分层记忆系统（能力轴）对于长时域空间任务十分重要 [Packer等人, 2023, Banino等人, 2018, Xu 等 , 2025, Zhang 等人 , 2025a, Blundell 等人 , 2016, Pritzel 等人 , 2017]。（2）GNN-LLM集成（任务轴）是实现结构化空间推理的一种有前景的方法 [Jin 等人, 2023, Chen等, 2024e,c, Chai 等人, 2023, Shehzad 等人, 2024, Fatemi 等，2023, 2024]。（3）世界模型（尺度轴）对于在从微观到宏观的空间尺度上安全部署至关重要 [Hafner 等人, 2023, Bruce 等人, 2024b, Ha and Schmidhuber, 2018, Feng 等人, 2025, Ding 等, 2024, Brooks 等人, 2024, Hafner 等人, 2020, 2021, Schrittwieser 等人，2020]。最后，我们提出六项重大挑战，并概述未来研究方向，包括需要建立统一评估框架，以规范跨领域评估。这一分类法为整合分散的研究工作奠定了基础，并有助于推动机器人学、自动驾驶汽车和地理空间智能领域新一代空间感知自主系统的发展。

1 引言

通用人工智能的追求日益聚焦于创造能够在物理环境中感知、推理并采取行动的智能体 [Brooks, 1991, Russell and Norvig, 2010]。大语言模型擅长符号推理与规划 [Brown 等人, 2020, OpenAI, 2023], ，但在空间情境中却屡屡失效：导航智能体会臆造不存在的路径，操作规划器会提出物理上不可行的抓取方案，具身系统则会对物体距离产生相差数个数量级的误判 [Chen 等，2024a, Yang 等人，2025]。这些失效源于一个根本性缺口，它有别于多模态落地问题：视觉语言模型能够将图像与文本关联起来（符号接地），而空间智能则要求对几何、物理学及动作后果具备度量层面的理解（空间接地）。大语言模型缺乏后一种能力，内部没有关于三维结构、物理动力学或几何约束的模型。

多模态基础模型推动了视觉理解的发展 [Radford 等人，2021年，Liu 等人，2023b，OpenAI， 2023], 然而，仅有感知并不能赋予模型行动能力：能够详细描述场景的模型，仍可能无法在其中导航或操控物体。将感知转化为有效且可控的空间动作，仍是具身人工智能面临的关键瓶颈 [Ahn 等人，2022年，Brohan 等人，2023年，Kawaharazuka 等人， 2025]。这一观察促使我们聚焦于能够打通从感知到动作闭环的智能体系统。

我们将**智能体人工智能** 定义为通过自主决策展现目标导向行为的系统，其特点是具备三项核心能力，构成了我们分类体系中的能力轴：用于积累经验的记忆 、规划 （包括作为元层级规划、用于持续改进的自我反思），以及用于扩展能力的工具使用 [Wang 等人，2024b，Xi 等人，2023年，Yao 等人，2023b，Shinn 等人，2023b]。这些智能体通过感知、推理、动作和反馈的迭代循环运行[Yao 等人，2023b，Shinn 等人，2023b]。

互为补充的是，**空间智能**涵盖感知三维结构、推理物体关系、在环境中导航以及操控物理物体的能力 [Chen 等人，2024a，Marr，1982年，Newcombe， 2010]。尤为关键的是，空间任务涵盖三个尺度，它们构成了我们分类体系的尺度轴：微观空间 （厘米尺度的操作与抓取）、中观空间 （米尺度的房间导航与场景理解）和宏观空间 （千米尺度的城市规划与地理空间分析） [Battaglia 等人，2018年，Kipf 和 Welling， 2017]。这种尺度区分不仅是分类上的区分，更具有因果意义：在一个尺度上训练的智能体部署到另一个尺度时会失效，因为相关特征、物理规律和行动空间存在本质差异。针对厘米级精度优化的抓取策略无法规划城市尺度路线，而交通预测器也不具备接触力的表征。尺度不匹配是空间人工智能中迁移失败的主要原因。

这些领域的融合对于现实世界人工智能应用至关重要，我们按分类体系的任务轴对其进行组织：

导航（中观到宏观尺度）：自动驾驶车辆必须感知动态环境并规划安全轨迹 [Hu 等人，2023b，Caesar 等人，2020年，Waymo，2023年，Geiger 等人，2012年，Chen 等人，2024d，Waymo，2024年，Tian 等人， 2024]。室内机器人需要通过复杂的建筑布局进行视觉语言导航 [Anderson 等人，2018c，Krantz 等人， 2020]。

操作（微观到中观尺度）：机器人助手需要理解物体可供性和空间关系，以执行抓取、放置和工具使用 [Brohan 等人，2023年，Ahn 等人，2022年，Team 等人，2024年，Kim等，2024年，Driess 等人，2023年，Black 等人，2024年，Collaboration， 2023]。

宏观尺度推理：城市计算系统必须对复杂的时空依赖进行建模，以实现交通预测和资源分配 [Jin 等人，2023年，Li 等人，2018，Yu 等人，2018年，Wu 等人，2019b，Zheng 等人， 2014]。地理空间智能平台必须在行星尺度上分析卫星影像和地理数据 [Jakubik 等人，2024年，Cong 等人，2022年，Mai 等人，2023年，Janowicz 等人，2020年，Li 等人，2025a，Xiao 等人， 2025]。

尽管这一点十分重要，现有综述仍将这些领域割裂开来，缺乏将智能体架构与空间需求联系起来的统一框架。

贡献。 本文综述主要作出四项贡献：

1. 一个**统一三轴分类法**，将智能体人工智能组件（记忆、规划、工具使用）与空间智能领域（导航、场景理解、操作、地理空间分析）贯穿于

不同空间尺度（微观、中观、宏观）。该框架有助于清晰识别研究空白，指导新系统的架构决策，并为跨领域协作提供通用术语。

2. 对来自超过2,000篇综述论文的742篇引用文献进行**全面分析**，揭示了显著的尺度失衡：68%的方法针对中尺度空间任务（房间尺度导航），而微尺度空间操作和宏尺度地理空间推理尽管具有商业重要性，仍未得到充分探索。我们将GNN-LLM集成、视觉-语言-动作模型以及基于世界模型的规划确定为弥合这一差距的关键架构模式。

3. 与现有综述（表 1）进行**详细比较**，量化覆盖缺口，并展示本工作如何独特地连接智能体人工智能与空间智能领域。

4. 一份**前瞻性路线图**，确定了六项重大挑战，它们构成进展的结构性障碍，而非渐进式改进：统一跨尺度表征、基于定位的长时程规划、安全保证、仿真到现实迁移、多智能体协同以及边缘部署。应对这些挑战需要在智能体如何表征空间并进行推理方面取得根本性进展。

2 方法论

本文综述遵循严谨的文献综述方法，与计算机科学领域的最佳实践相一致 [Kitchenham，2004年，Petersen等人，2008年，Wohlin，2014年，Keele等人，2007年，Brereton等人，2007年，Dyb˚a与Dingsøyr，2007年，Moher等人，2009年，Okoli与Schabram，2010年，Webster与Watson，2002年，Jalali与Wohlin，2012年，Snyder，2019年，Tranfield等人， 2003]。我们检索了互补的学术数据库：Google Scholar用于广度覆盖，arXiv用于前沿预印本，ACM数字图书馆和IEEE Xplore用于同行评审的系统研究，Semantic Scholar用于引文感知排序，以及DBLP [Ley， 2002]用于广泛的会议与期刊覆盖。关键词包括“智能体人工智能”、“空间智能”、“具身人工智能”、“视觉-语言导航”、“机器人操作”、“地理空间人工智能”、“世界模型”、“图神经网络”、“时空学习”、“视觉-语言-动作”以及“机器人基础模型”。我们的初始检索获得了超过3,000篇论文。

随后，我们采用了严格的多阶段筛选流程：

1. **时间筛选：** 我们选取了 2018年 至 2026年 发表的论文。2018年这一时间下限标志着一个关键转折点：BERT [Devlin等人 2019] 的发布开启了基础模型时代，而 Habitat [Savva等人 2019] 和 AI2-THOR [Kolve等人 2017]则建立了标准化的具身人工智能基准，使可复现研究成为可能。2018年之前的奠基性研究（如 PointNet [Qi 等人，2017a年], GCN [Kipf 和 Welling， 2017]）若确立了至今仍在广泛使用的方法，也会纳入其中。

2. **会议筛选：** 我们优先考虑发表于顶级会议和期刊的论文，包括 NeurIPS、ICML、ICLR、CVPR、ECCV、ICCV、CoRL、RSS、IROS、ICRA、ACM Computing Surveys、IEEE TPAMI、《自然》、《科学》、《科学·机器人学》，以及 arXiv 上的优质预印本。

3. **质量筛选：** 为减轻基于引用的选择偏差对近期研究造成的不利影响，我们明确纳入了引用量较低、提出突破性方法的论文（例如早期视觉-语言-动作模型、新颖的图神经网络-大语言模型架构），并同时纳入引用量较高的基础方法。这种双重策略兼顾了已确立的影响力与新兴创新。

4. **相关性筛选：** 我们确保论文直接探讨智能体能力与空间智能的交叉领域。

这一流程最终形成了一个包含 2,000 多篇论文的语料库，其中 742 篇被直接引用；我们对这些论文进行了仔细分析，以构建分类体系、识别关键趋势，并综合得出本文综述所呈现的研究发现。我们采用滚雪球抽样法，通过追踪正向和反向引用链，确保广泛覆盖相关研究。两位独立审稿人对论文筛选和分类体系构建进行了验证，在纳入标准方面达成了 94% 的标注者间一致性。分歧分为两类：范围争议（即论文的空间要素是否足够核心，占 4%）和分类体系归类（即某种方法主要针对哪个维度，占 2%）。所有分歧均通过讨论解决，直至达成共识。

3 相关工作与综述比较

尽管已有若干综述探讨了智能体人工智能或空间智能的某些方面，但此前尚无研究在单一框架内明确且系统地统一这些领域，并覆盖从微尺度操作到宏尺度地理空间推理的范围。我们对五类现有综述进行了回顾，并在表 1 中提供了详细比较。

智能体人工智能综述。 近期关于基于大语言模型的智能体的综述 [Wang 等人，2024b；Xi 等人，2023年；Guo等人，2024b；Durante 等人，2024年；Weng，2023年；Mialon 等人， 2023]着重讨论推理和工具使用，但并未涉及空间能力。Sumers 等人 [2024]从认知架构的角度进行了探讨。这些研究的共同局限在于，将智能体视为主要基于符号、处理文本和结构化数据的推理器，却没有机制来维护能在多个推理步骤中持续存在的空间状态表示（地图、3D场景、物体位姿）。因此，这些框架无法支持几何一致性至关重要的具身任务。

具身人工智能综述。 具身人工智能综述 [Duan 等人，2022年；Gupta 等人，2021年；Francis 等人，2022年；Savva等人，2019年；Anderson 等人，2018b；Gervet 等人， 2023] 涵盖了仿真环境和基准测试，但缺乏与智能体架构的整合。Kawaharazuka 等人 [2025] 专门综述了机器人学领域的视觉-语言-动作模型。值得注意的是，这些研究均未涉及自主长时程规划：智能体如何分解多步任务、从失效中恢复，以及如何在无人类干预的情况下跨回合积累经验。

地理空间人工智能综述。地理空间人工智能综述 [Mai 等人，2023年，Janowicz 等人，2020年，Xiao 等人，2025] 以及时空数据挖掘综述 [Jin 等人，2023年，Atluri 等人，2018年，Wang 等人，2020年，Jiang 和 Luo，2022年，Balachandar 等人， 2025] 都高度专业化。关键在于，这些研究将地理空间系统视为被动的预测工具，只输出预测结果（交通流量、土地覆盖），却没有形成闭环：它们没有探讨预测应如何触发动作、动作如何影响未来观测，或预测出错时智能体应如何调整。这种开环框架限制了它们在必须依据地理空间智能采取行动的自主系统中的适用性。

图神经网络综述。图神经网络综述 [Wu 等人，2020a，Bronstein 等人，2021年，Hamilton，2020年，Battaglia 等人，2018年，Zhou 等，2020年，Zhang 等人，2020年，Veli˘ckovi´c， 2023] 全面涵盖了图学习，但并不聚焦于空间应用或智能体集成。针对特定领域的图神经网络综述包括交通领域 [Jiang 和 Luo， 2022], 城市计算领域 [Balachandar 等人， 2025], 以及时空预测领域的综述 [Jin 等人， 2023]。

视觉语言模型综述。 关于视觉语言模型的综述 [Zhang 等人，2024c，Bordes 等人， 2024] 涵盖了多模态理解，但未涉及空间动作或具身化。Kawaharazuka 等人 [2025]专门综述了面向机器人学的视觉-语言-动作模型。

表 1：与现有综述的比较。符号： ✓ = 全面覆盖， ○ = 部分覆盖，空白 = 未涵盖。我们的工作首次实现了所有维度的统一覆盖。

综述	智能体人工智能	具身人工智能	空间推理	地理空间	图神经网络	工业	统一分类体系
Wang 等人（2024年） [Wang 等人，2024b]	✓	○	○				
Xi 等人（2023年） [Xi 等人， 2023]	✓	○					
Duan 等人（2022年） [Duan 等人， 2022]		✓	○				
Kawaharazuka 等人（2025年） [Kawaharazuka 等人， 2025]	○	✓	○				
Jin 等人（2023年） [Jin 等人， 2023]			○	○	✓		
Mai 等人（2023年） [Mai 等人， 2023]				✓		○	
Bronstein 等人（2021年） [Bronstein 等人， 2021]			○		✓		
Zhang 等人（2024年） [Zhang 等人，2024c]	○		○				
本文综述	✓	✓	✓	✓	✓	✓	✓

如表 1 所示，现有综述主要聚焦于智能体推理（上方各行）或特定领域的空间方法（中间各行），此前尚无研究覆盖全部七个维度。这一模式颇具启示：智能体综述（Wang、Xi）对推理的覆盖较为充分，但在地理空间和图神经网络方面却留有空白；空间综述（Jin、Mai、Bronstein）的情况则恰恰相反。所有既有研究在“统一分类体系”一栏中均未出现勾选，这充分反映出该领域分散为彼此割裂的子社群。这一空白正是我们提出统一分类体系的动因。

4 统一三轴分类法

我们提出一种三轴分类体系（图1），将智能体能力与不同空间尺度下的空间任务需求相对应。阅读这一框架时需注意：每种方法都同时位于三个轴上的某个位置。任务轴 说明系统执行什么任务，能力轴 说明系统如何推理和行动，尺度轴 则说明空间粒度。这一结构有助于详细比较各种方法，并识别设计空间中尚未充分探索的区域。尤为重要的是，我们严格遵循结构化规范：本文综述中探讨的每种方法都明确映射至全部三个轴（见表2），从而确保表征一致，并支持直接开展跨方法比较。

4.1 分类轴

轴 1：空间任务。我们确定了四类主要空间任务：

• **导航**：在涵盖室内和室外环境的场景中，朝着目标移动。室内任务包括点目标 [Anderson 等人，2018a，Wijmans 等人，2020年，Savva 等人， 2019], 物体目标 [Chaplot等人，2020b，Batra 等人， 2020], 以及视觉语言导航 [Anderson 等人，2018c，Krantz 等人， 2020]。室外任务包括在高速公路和城市街道上自动驾驶 [Hu 等人，2023b，Caesar 等人，2020年，Waymo，2023年，Geiger 等人， 2012], 越野导航 [Triest 等人， 2024], 以及行人尺度室外寻路 [Mirowski 等人， 2018]

• **场景理　　解**：感知并推理三维结构、物体与空间关系　　s

• **操作**：与物体进行物理交互，包括抓取 [Mahler等人，2017年，Morrison等人，2018年，Fang等人，2020年，ten Pas等人， 2017], 放置 [Zeng等人， 2021], 以及工具使用 [Qin等人，2024a,b]

• **地理空间分析**：大规模空间推理，包括卫星影像 [Christie等人，2018a,b，Demir等人，2018年，Xia等人，2017年，Sumbul等人， 2019], 城市计算 [Zheng等人，2014年，Yuan等人，2020b，Zheng， 2015], 以及地理信息系统 [Longley等人，2015年，Goodchild， 2007]

轴2：智能体能力。我们识别出三种核心智能体能力：

• **记忆**：短期（上下文内）、长期（检索增强）、情景以及空间记忆系统

• **规划**：反应式、分层式、基于搜索的以及基于世界模型的规划方法。这些方法构成一个谱系，从将观测直接映射为动作的反应式策略（速度快，但对新情况较为脆弱），到在行动前模拟未来状态的基于模型的规划器（速度较慢，但对分布偏移更为稳健）。自我反思 [Shinn 等人，2023b] 作为一种元能力运作，横跨记忆（存储过往失败）与规划（修正未来动作）

• **工具使用与行动**：API集成、代码生成、物理动作基元以及技能库

轴 3：空间尺度。 我们区分三种空间尺度，其边界依据感知与驱动约束来确定：

• **微观空间**（<1m）：位姿估计、抓取，以及厘米级精度的精细操作。这一边界反映了触觉传感器的有效范围以及典型机械臂的工作空间。

• **中观空间**（1 米–100 米）：房间导航、建筑探索、室内/室外局部场景。这一范围对应机载摄像头和激光雷达的有效视场，在此范围内，度量 SLAM 仍然可行。

• **宏观空间 (**>100米**)**：城市尺度规划、卫星影像、覆盖数千米的基础设施网络。超过100米后，智能体必须依赖外部感知（卫星、全市传感器网络），无法直接操控环境。

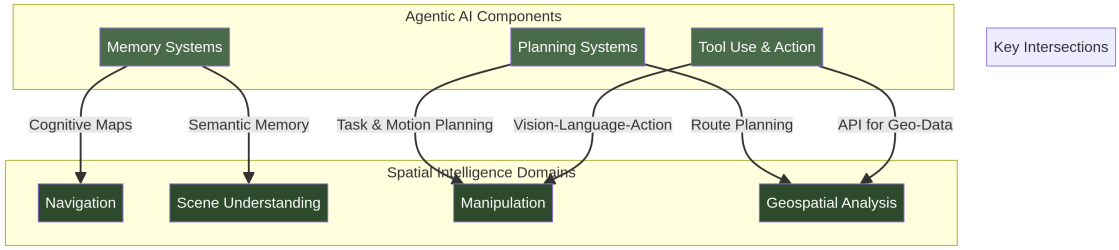

图1：统一三轴分类法将不同空间尺度上的智能体人工智能能力与空间智能领域联系起来。这些维度的交叉定义了自主空间智能系统的设计空间。**设计权衡**：目前尚无方法能同时在全部三个轴上实现出色性能。微尺度操作系统（左下）可达到厘米级精度，但无法规划即时工作空间以外的任务。地理空间模型（右上）能够进行行星尺度推理，但缺乏闭环行动能力。该空间中较少涉及的区域（例如宏观尺度操作、微尺度地理空间）代表着尚待探索的研究机遇；跨轴整合有望在这些领域带来重大进展。

表2：映射至三轴分类体系的代表性方法。表示类型：符号（语言、代码）、指标（几何、视觉）、潜在（学习嵌入）、多模态（组合）。

方法	空间任务	智能体能力	尺度	表示	主要失效模式
VLN-BERT [Hong 等人， 2021]	导航	记忆 + 规划	Meso	多模态	指令落地错误
SayCan [Ahn 等人， 2022]	操作	规划 +工具使用	微观–中观	符号	可供性不匹配
RT-2 [Brohan 等人， 2023]	操作	工具使用	微观	指标	分布外物体
VLMaps [Huang 等人，2023a]	导航	记忆	Meso	指标	随时间推移的语义漂移
Voyager [Wang 等人，2023a]	导航 + 操作	记忆 +规划	Meso	符号	代码执行失败
DCRNN [Li 等人， 2018]	地理空间	记忆（低规划）	宏观	潜在	非平稳动力学
Graph WaveNet [Wu 等人，2019b]	地理空间	记忆（低规划）	宏观	潜在	稀疏图区域
Prithvi [Jakubik 等人， 2024]	地理空间	仅记忆	宏观	指标	无动作能力
DreamerV3 [Hafner 等人， 2023]	导航 + 操作	规划（世界模型）	微观–中观	潜在	模型误差累积
PaLM-E [Driess 等人， 2023]	操作	规划 +工具使用	微观–中观	多模态	幻觉动作
OpenVLA [Kim 等， 2024]	操作	工具使用	微观	指标	有限泛化
LLaGA [Chen等，2024e]	场景理解	记忆	Meso	多模态	图构建噪声

4.2 方法—分类体系映射

表2将代表性方法映射到我们的三轴分类体系，展示了该框架如何组织这一领域。

关键要点：分类体系

- 三轴分类体系（任务 × 能力 ×尺度）为组织空间人工智能研究提供了一个全面的框架
- 大多数方法关注中观空间尺度。微观和宏观尺度仍未得到充分探索
- 记忆系统对于所有空间任务都至关重要，但不同尺度下的实现方式差异显著
- **关键缺口：** 宏尺度地理空间方法以记忆为主导，规划能力极为有限。目前还没有任何系统能够将城市尺度推理与自主目标导向行为结合起来
- 基于图神经网络的方法与智能体能力的交叉融合，代表着一个新兴前沿

5 个空间智能的智能体人工智能组件

本节探讨智能体能力如何促成空间智能，并围绕这一核心科学问题展开：智能体如何在内部表征空间环境、对其进行推理并在其中采取行动？

5.1 Me忆系统：智能体如何记住空间信息?

记忆使智能体能够积累和检索经验知识。认知科学提供了基础理论 [Tulving, 1972年，Baddeley，2003]，而我们关注的是能够使人工智能系统维持持久空间知识的计算实现。核心挑战是：智能体如何跨越不同的时间范围和尺度维持持久空间知识?

短期记忆。上下文学习 [Brown 等人，2020年，Dong等人，2022年，Olsson等人，2022年，Akyurek等人，2023年，Dai 等人，2023a，Min等人，2022年，Xie等人，2022年，Wei等人，2023年，Chan等人，2022] 使模型能够通过提示中的示例适应新任务。这种机制借助注意力机制，以所提供的示范为条件，无需更新参数即可快速适应。工作记忆机制 [Graves等人，2014年，Weston等人，2015年，Sukhbaatar等人，2015年，Kumar等人，2016年，Santoro等人，2016年，Graves等人，2016年，Munkhdalai和Yu， 2017] 能够在推理过程中临时存储信息，从而支持超出单次前向传播能力的多步计算。为了说明它与空间记忆的区别：短期记忆可能会保留“去厨房找一个红色杯子”这一指令，而空间记忆则会编码厨房本身的度量布局，包括门口位置和橱柜位置。

长期记忆。检索增强生成 [Lewis等人，2020年，Packer等人，2023年，Guu等人，2020年，Borgeaud等人，2022年，Asai等人，2023年，Trivedi等人，2023年，Izacard等人，2023年，Shi等人，2023年，Ram等人，2023年，Khandelwal等人， 2020] 使知识能够突破上下文限制而持续保留。MemGPT[Packer等人， 2023] 为长对话引入了分层记忆管理。AMEM [Xu 等， 2025]为大语言模型提供智能体记忆。MemEvolve [Zhang 等人，2025a] 实现了智能体记忆的元进化。向量数据库 [Johnson等人，2019年，Malkov和Yashunin，2018年，Douze等人，2024年，Wang 等人，2021年，Pinecone，2023年，Jegou等人，2011年，Ge等人，2014年，Guo等人， 2020] 为记忆检索提供高效相似性搜索，使智能体能够访问相关的过往经验。

情景记忆。情景记忆存储具体经历和事件（发生了什么），使智能体能够从过往交互中学习 [Blundell等人，2016年，Pritzel等人，2017年，Banino等人，2018年，Ritter等人，2018年，Fortunato等人，2019年，Botvinick等人，2019年，Gershman 和 Daw， 2017]。对于必须记住到访地点、遇到的物体和成功的动作序列的空间智能体来说，这类记忆至关重要 [Savinov等人，2018年，Chaplot等人，2020c，Fang等人，2019年，Ramakrishnan等人，2022年，Ye等人，2021年，Chen等人，2022f]。情景记忆与空间记忆之间的区别十分重要：情景记忆记录“机器人在下午3点撞上了一把椅子”，而空间记忆编码的是持久几何结构（椅子相对于桌子的位置），不受任何具体事件影响。

空间记忆。情景记忆记录的是何事发生在何处，而空间记忆编码的是“何处”本身的结构：独立于具体事件而持续存在的几何与拓扑关系。专门的实现包括认知地图 [Tolman, 1948, O’Keefe and Nadel, 1978, Moser et al., 2008年, Hafting et al., 2005], 拓扑表示 [Kuipers, 2000, Choset and Nagatani, 2001, Thrun, 1998, Kuipers and Byun, 1991], 以及度量地图 [Thrun et al., 2005, Durrant-Whyte and Bailey, 2006, Cadena et al., 2016年, Mur-Artal et al., 2015年, Mur-Artal and Tard´os, 2017年, Campos et al., 2021年, Engel et al., 2017, 2014]。空间记忆的神经方法包括神经SLAM [Chaplot等人, 2020c,d,b,2021], 语义地图 [Huang 等人，2023a, Henriques and Vedaldi, 2018年, Shah et al., 2023b,a, Huang 等人，2023c, Chen 等人, 2023a], 以及场景图 [Armeni et al., 2019年, Rosinol et al., 2020年, Hughes et al., 2022年, Gu et al., 2024年, Wu 等人, 2021年, Wald et al., 2020年, Kim et al., 2019]。

空间失效模式。仅使用语言的智能体由于缺乏具身空间表征，在空间任务中表现不佳。这些失效可归为四类，每一类都能追溯到特定的表征缺陷：(1) 空间幻觉，即智能体描述不可能存在的空间布局（GPT-4V 在 SpatialBench [的空间关系问题中有 40% 回答错误，Chen 等人，2024a]）。(2)参考系混淆，即智能体混淆自我中心坐标与异我中心坐标（视觉语言导航智能体因坐标系错位而出现 15–20% 的错误率 [Anderson 等人，2018c]）。(3)尺度不敏感，即智能体无法区分微观、中观和宏观尺度推理（当物体尺度与训练时不同时，SayCan 的可供性模型会失效[Ahn 等人， 2022]）。(4)时间漂移，即空间记忆会随着时间跨度增长而衰退（ VLMaps 在未进行地图更新的情况下经过 100+步后会出现语义漂移 [Huang 等人，2023a]）。

5.2 规划系统：智能体如何在空间视野中进行规划 ？

规划将目标分解为可执行的动作序列，从而实现复杂的任务完成 [Russell and Norvig, 2010, Ghallab 等人，2004, LaValle, 2006, Fikes and Nilsson, 1971, Sacerdoti, 1974, Nau 等人，2003, Bylander, 1994, Geffner and Bonet, 2013, Kautz and Selman, 1992, Helmert，2006]。核心挑战是：智能体如何在考虑几何约束的同时，将空间目标分解为可行的动作序列？

思维链推理。 逐步推理 [Wei 等人，2022年，Kojima 等人，2022年，Wang 等人，2022年，Zhou 等人，2023a，Fu 等人，2023年，Chen 等人，2023c，Zhang 等人，2023年，Khot 等人，2023年，Diao 等人， 2023]能够进行结构化的问题分解。自一致性 [Wang 等人， 2022] 通过多条推理路径提高可靠性。零样本思维链 [Kojima 等人， 2022] 无需示例即可进行推理。

基于树的搜索。 思维树 [Yao 等人，2023a，Long，2023年，Xie 等人，2023年，Hulbert 等人，2024] 通过有目的的搜索探索多个解答分支。思维图 [Besta 等人，2023年，Lei 等人，2023年，Yao 等人， 2024] 通过任意连接实现更复杂的推理结构。RAP [Hao 等人，2023年，Shridhar 等人，2020b，Zhao 等人，2024年，Liu 等人，2024d] 将推理与行动结合到规划框架中。蒙特卡洛树搜索的变体 [Silver 等人，2016年，Schrittwieser 等人，2020年，Browne 等人，2012年，Anthony 等人，2017年，Silver 等人，2017年，2018年，Vinyals 等人，2019年，Kocsis 和 Szepesv´ari， 2006] 提供具有理论保证的规范化探索。

分层规划。LLM-Planner [Song 等人， 2023]为具身智能体提供少样本落地规划。Inner Monologue [Huang 等人，2022a] 通过内部对话实现反馈驱动规划。HiPlan [Li 等人，2025b] 提出了基于大语言模型的分层规划。分层强化学习方法[Nachum 等人，2018年，Vezhnevets 等人，2017年，Bacon 等人，2017年，Pertsch 等人，2021年，Kulkarni 等人，2016年，Gupta 等人，2019年，Levy 等人，2019年，Li 等人，2020年，Zhang 等人，2021b，Sutton 等人， 1999] 通过时间抽象将任务分解为子任务。

任务与运动规划。任务与运动规划 [Garrett 等人，2021年，Kaelbling 和 Lozano-P´erez，2011年，Toussaint，2015年，Dantam 等人，2016年，Srivastava 等人，2014年，Garrett 等人，2020年，Driess 等人， 2020]将符号规划与连续运动规划相结合，用于机器人应用。这种方法结合了符号推理的表现力与几何规划的精确率。

从符号规划到神经规划。 上述方法依赖明确的符号表示，因此能够提供形式化保证：计划可根据前置条件进行验证，也可通过检查中间状态来诊断失效。近期研究探讨大型神经模型能否具备这些能力，以学习得到的表示取代人工设计的结构，从而实现更广泛的泛化，但代价是可解释性降低。这一转变引发了关于计划可验证性和故障恢复的根本性问题，尤其是在安全关键应用中，了解计划为何失效与判断计划是否成功同样重要。

基于大语言模型的规划。 近期研究直接使用大语言模型进行规划 [Huang 等人，2022b、Valmeekam 等人，2023a、Song 等人，2023、Silver 等人，2024、Liu 等人，2023a、Kambhampati 等人，2024、Valmeekam 等人，2023b]。SayCan [Ahn 等人， 2022] 通过可供性将语言模型与现实环境联系起来。Code as Policies [Liang 等，2023b] 会生成可执行机器人代码。ProgPrompt [Singh 等人， 2023] 使用程序化提示进行任务规划。一个重要的注意事项是：大语言模型生成的计划无法通过形式化方式验证其正确性或安全性。与能够保证满足前置条件的经典规划器不同，大语言模型规划器可能生成流畅但在物理上无法实现的动作序列。这一局限在安全关键应用中尤其令人担忧，例如手术机器人和自动驾驶车辆，因为计划失效可能造成不可逆后果。

空间规划失效模式。基于大语言模型的规划器会在以下情况下失效：(1) 违反几何约束：在 BEHAVIOR-1K 上， [Li 等人，2023a], 的大语言模型规划器由于忽略碰撞的规划，成功率仅为 12%。(2) 未满足动作前置条件：VirtualHome [Puig 等人， 2018]显示，35% 的大语言模型规划会因前置条件违例而失败。(3) 长时程信用分配失效：在 ALFRED 上， [Shridhar 等人，2020a] 发现，随着任务时域从 5 步延长至 20 步，成功率从 65% 降至 18%。(4) 缺少动态重新规划：RoboTHOR [Deitke 等人， 2020] 显示，意外障碍物出现时，失效率为 40%。

5.3 工具使用与动作：智能体如何将语言定位到几何空间中?

工具使用通过外部接口和物理动作拓展智能体能力 [Osiurak 和 Badets，2016年，Vaesen，2012年，Beck 等人，2011年，Shumaker 等人，2011年，Seed 和 Byrne，2010年，Tomasello， 1999]。核心挑战是：如何将基于语言的推理转化为精确的几何动作？我们

区分两种截然不同的类别：符号工具（API、代码执行、数据库查询），它们在数字空间中运行，结果具有确定性；以及物理致动（抓取、运动、操作），它们在连续物理空间中运行，具有随机动力学和不可逆的后果。这一区别很重要，因为符号工具错误可以被发现并重试，而物理动作错误则可能造成永久性损坏。

API集成。 Toolformer [Schick 等人，2023年，Parisi 等人，2022年，Mialon 等人，2023年，Nakano 等人，2021] 实现了自监督工具学习。Gorilla [Patil 等人，2023年，Li 等人，2023e，Tang 等人， 2023]专注于结合检索增强的 API 调用。ToolLLM [Qin 等人，2024a，Hao 等人，2024年，Qin 等人，2024b，Xu 等人，2023d] 提供全面的工具使用基准测试。TaskMatrix [Liang 等，2023c，Lu 等人，2023年，Yang 等人，2023b] 将基础模型与数百万个 API 相连接。TALM [Parisi 等人，2022] 通过工具使用增强语言模型。其他工具使用框架还包括 HuggingGPT [Shen 等人，2023b], 、ViperGPT [Sur´ıs 等人， 2023], Visual ChatGPT [Wu 等人，2023a], 以及 MM-REACT [Yang 等人，2023b]。

代码生成。 PAL [Gao 等人，2023年，Wang 等人，2024a] 使用代码进行推理。Code as Policies[Liang 等，2023b] 根据语言生成可执行机器人代码。Codex [Chen 等人，2021a], StarCoder[Li 等人，2023f], CodeLlama [Roziere 等人， 2023], DeepSeek-Coder [Guo等人，2024a], 和 WizardCoder[Luo 等人， 2023] 具备代码生成能力。ProgPrompt [Singh 等人， 2023] 将程序化提示用于机器人学。自我调试 [Chen 等人，2023f，Olausson 等人， 2023] 通过迭代改进提升代码质量。

ReAct 架构。 ReAct [Yao 等人，2023b,c，Liu 等人，2023c] 将推理与动作执行交替进行，使智能体能够在行动前进行思考。Reflexion [Shinn 等人，2023b,a，Kim 等人， 2023] 通过言语强化加入自我反思，从而实现改进。这些架构构成了许多空间智能体的基础。

物理动作。 对于具身智能体，工具使用延伸至物理操作 [Zeng 等人，2021年，Brohan 等人，2022年、2023年，Shridhar 等人， 2022, 2023]。动作基元 [Dalal 等人，2021年，Nasiriany 等人，2022年，Mandlekar 等人， 2021] 提供了可重复使用的构建模块。技能库 [Wang 等人，2023a，Lynch 等人，2020年，Pertsch 等人，2021年，Singh 等人，2021年，Xu 等，2023a] 使组合式动作成为可能。

关键要点：智能体组件

- 记忆系统必须明确具备空间属性：对于空间任务，认知地图、语义地图和场景图优于通用检索
- 通过几何锚定的分层规划，弥合高层语言目标与低层运动指令之间的差距
- 工具使用通过代码生成、API 调用和学习得到的动作基元，衔接语言与动作
- 关键失效模式源于缺乏空间接地：幻觉、参考系混淆和几何约束违反

6 空间智能领域

微观总结： 本节综述智能体导航、场景理解、操作和地理空间推理的方法。我们明确将这些任务领域与智能体规划联系起来，并重点介绍基准差距。

本节探讨四个主要的空间任务领域，围绕以下问题展开：智能体必须具备哪些空间能力，才能在物理世界中运行？

6.1 导航：智能体如何在空间中移动?

微观摘要： 本小节介绍视觉导航和语言引导导航，涵盖从经典机器人学到现代基于学习的系统。我们重点阐述从显式建图到隐式空间推理的进化。对智能体而言，导航不仅是路径执行，更是所有下游空间任务的基础。

导航要求智能体感知环境、规划路径，并朝目标运动[Bonin-Font et al., 2008, DeSouza and Kak, 2002, Thrun, 2002, Siegwart et al., 2011, Choset et al., 2005, Cadena et al., 2016, Engel et al., 2017, Fuentes-Pacheco et al., 2015, Durrant-Whyte and Bailey, 2006, Konoligeet al., 2008]。

视觉语言导航。 视觉语言导航（VLN）任务要求智能体在视觉环境中遵循自然语言指令 [Anderson 等人, 2018c, Qi 等人, 2020 年 , Krantz 等人 , 2020 年 , Fried 等人 , 2018c, Chen 等人 , 2022h, Shah 等人 , 2023b, Hong 等人 , 2020 年 , Chen 等人, 2021c, An 等人, 2023年, Wang 等人, 2019b, Ma 等人, 2019a, Tan 等人, 2019年, Zhu 等人, 2020年, Hong 等人, 2021年, Qiao 等人, 2022年, Chen 等人, 2022g, Guhur 等人, 2021]。R2R [Anderson 等人, 2018c] 提出了这一范式。REVERIE [Qi 等人, 2020] 加入了目标定位。VLN-CE [Krantz 等人, 2020] 将任务扩展到了连续环境。

物体目标导航。 物体导航要求导航至物体类别 [Batra等人, 2020年, Chaplot等人, 2020a, Majumdar等人, 2022年, Gadre等人, 2022年, 2023年, Dorbala等人, 2022年, Ramakrishnan等人, 2022年, Ye等人, 2021年, Khandelwal等人, 2022年, Yokoyama等人, 2024]。ZSON [Majumdar等人, 2022]可实现零样本导航。CLIP-Nav [Dorbala等人, 2022] 使用视觉语言模型。CoW [Gadre等人, 2022]使用装在轮子上的 CLIP 进行语义导航。

视听导航。 SoundSpaces 中的音频线索可引导导航 [Chen 等人, 2020年, 2022d, Gan 等人, 2020年, Chen 等人, 2021b, Majumder 等人, 2022年, Chen 等人, 2022c]。这种模态对于寻找发声目标至关重要。

具身问答。 具身问答需要通过导航来回答问题 [Das 等人, 2018b, Gordon 等人, 2018, Wijmans 等人, 2019, Yu 等人, 2019, Das 等人, 2018a, Gordon 等人, 2019]。三维问答[Azuma 等人, 2022年, Ma 等人, 2022a, Hong 等人, 2023b, Chen 等人, 2024f, Zhu 等人, 2023年, Huang 等人, 2024b]扩展至三维场景理解。

6.2 场景理解：智能体如何感知三维结构?

微观总结： 本小节回顾三维场景理解技术，将这些表征与其在下游智能体规划中的应用联系起来。场景表示决定了智能体能够规划哪些动作：更丰富的表示能够支持对物体可供性和空间约束进行更复杂的推理。

场景理解涵盖对三维几何的感知 [Hartley and Zisserman, 2003, Szeliski, 2022, Forsyth and Ponce, 2011, Marr, 1982, Prince, 2012, Faugeras, 1993], 物体识别 [Krizhevsky et al., 2012, He et al.,2016, Simonyan and Zisserman, 2015, Szegedy et al., 2015, Huang et al., 2017, Tan and Le,2019], 以及空间关系推理 [Johnson et al., 2015, Krishna et al., 2017, Lu et al., 2016, Xu et al., 2017, Zellers et al., 2019]。

神经场景表示。 NeRF [Mildenhall et al., 2020, Barron et al., 2022, M¨uller et al., 2022, Park et al., 2019, Mescheder et al., 2019, Barron et al., 2023, Martin-Brualla et al., 2021, Tancik et al., 2022, Chen et al., 2022a, Fridovich-Keil et al., 2022] 革新了三维重建。Mip-NeRF [Barron et al., 2022] 支持多尺度渲染。三维高斯泼溅 [Kerbl et al., 2023, Luiten et al., 2023, Fan et al., 2024, Yu et al., 2024, Huang et al., 2024a] 实现了实时渲染。与同步定位与地图构建的集成[Sucar et al., 2021, Zhu et al., 2022, Keetha et al., 2024, Bird et al., 2025, Teed and Deng, 2021, Teed et al., 2024, Mur-Artal and Tard´os, 2017, Campos et al., 2021]实现了在线重建。

点云处理。 点云方法 [Qi 等人, 2017a年、b, Wang 等人, 2019年、Thomas 等人, 2019年、Zhao 等人, 2021年、Wu 等人, 2019a、Liu 等人, 2019年、Ma 等人, 2022b] 处理原始三维数据。Point-BERT[Yu 等人, 2022b], Point-MAE [Pang 等人, 2022], PointGPT [Chen 等人, 2024b], Point-Bind [Guo等人, 2023], 以及 Uni3D [Zhou 等, 2024] 引入自监督预训练。三维目标检测 [Shi等人, 2019年、Qi 等人, 2019年、Shi等人, 2020年、Chen 等人, 2023g、Lang 等人, 2019年、Yin 等人, 2021年、Fan 等人, 2022年、Zhou 和 Tuzel, 2018]使场景解析成为可能。

深度估计。 单目深度估计 [Godard 等, 2019年、Ranftl 等, 2021年、2020年、Yang 等, 2024c、Fu 等, 2024b、Bhat 等, 2021年、Li 等, 2022年、Yuan 等, 2022年、Eigen 等, 2014] 通过单张图像提供几何理解。Depth Anything [Yang 等, 2024c] 实现了强大的零样本迁移。Metric3D [Yin 等人, 2023] 恢复度量深度。

语义分割。 语义分割 [Long 等, 2015, Chen 等, 2017, Kirillov 等, 2023, Peng 等, 2023, Chen 等, 2023b, Xie 等, 2021, Cheng 等, 2022, Jain 等, 2023] 可实现场景解析。SAM [Kirillov 等, 2023] 提供可提示分割。开放词汇方法

[Ghiasi 等，2022，Liang 等，2023a，Chen 等，2023a，Ding 等，2022，Xu 等，2023b，Zhou 等，2022] 可实现零样本识别。

与智能体规划的联系。 这些表征本身并非目的，而是后续规划的输入。NeRF 为碰撞检测提供几何基底。语义地图告知规划器哪些区域可以通行。场景图编码了任务分解所需的关系结构。表征的选择会直接限制智能体能够生成和验证哪些计划。

6.3 操作：智能体如何与物体交互?

微观摘要： 本小节聚焦机器人操作，涵盖抓取、抓取前操作以及长时程序列任务。对智能体而言，操作是规划与物理现实交汇之处：每个动作都会产生不可逆后果，进而限制所有未来可能性。

操作需要理解物体可供性 [Gibson，1979年，Do等人，2018年，Nagarajan等人，2019年，Fang等人，2018年，Zhu等人，2015年，Myers等人， 2015], 规划接触丰富的交互 [Chitnis等人，2020年，Kroemer等人， 2021], 并执行精确的运动指令 [Argall等人，2009年，Billard等人，2008年，Ravichandar等人， 2020]。

视觉-语言-动作模型。 视觉-语言-动作模型 [Brohan 等人，2022年、2023年，Team 等人，2024年，Kim等，2024年，Black 等人，2024年，Driess 等人，2023年，Collaboration，2023年，Zitkovich 等人，2023年，Padalkar 等人， 2023] 直接将视觉观测和语言指令映射为动作。RT-1 [Brohan 等人，2022] 引入了大规模机器人学习。RT-2 [Brohan 等人， 2023] 展示了网络规模预训练迁移。RT-X [Collaboration， 2023] 实现了跨具身学习。Octo [Team 等人， 2024]提供了一种开放式通用策略。OpenVLA [Kim 等人， 2024] 提供了开源 VLA。π_0 [Black 等人， 2024] 为机器人学习引入了流匹配。RoboCat [Bousmalis 等人， 2023] 展示了自我改进。

模仿学习。 行为克隆 [Pomerleau, 1988, Chi 等人, 2023, Zhao 等人, 2023b, Chi 等人, 2024, Florence 等人, 2022, Mandlekar 等人，2021, 2022] 通过示例进行学习。扩散策略 [Chi 等人，2023] 将扩散模型应用于动作生成。ACT [Zhao 等人, 2023b]使用 Transformer 进行动作分块。从玩耍中学习 [Lynch 等人，2020] 支持非结构化学习。

强化学习。 用于操作的强化学习 [Kalashnikov 等人, 2018, Levine 等人, 2018, Haarnoja 等人, 2018, Schulman 等人, 2017, Fujimoto 等人, 2018, Lillicrap 等人, 2016, Mnih 等人，2015, 2016] 支持通过交互进行学习。QT-Opt [Kalashnikov 等人，2018] 可扩展至真实场景中的抓取任务。软演员-评论家 [Haarnoja 等人，2018]提供样本高效学习。

仿真环境。 仿真平台 [James 等人，2020年，Yu 等人，2020年，Makoviychuk 等人，2021年，Savva等人，2019年，Kolve等人，2017年，Gu 等人，2023年，Deitke 等人，2023年，Xiang 等人，2020年，Mu 等人，2021年，Ehsani 等人，2021年，Szot 等人，2021年，Li 等人，2021年，Shen 等人， 2021]提供训练环境。RLBench [James 等人， 2020] 提供多样化的操作任务。Meta-World [Yu 等人， 2020]提供多任务基准。Isaac Gym [Makoviychuk 等人， 2021] 支持 GPU 加速仿真。

长时程任务组合。 现实世界中的操作很少涉及孤立的动作。制作三明治需要打开抽屉、取出面包、关上抽屉、打开冰箱，依此类推；每个动作都会为后续步骤创造前置条件。这种顺序依赖结构使操作区别于导航：走错路可以纠正，但打碎的鸡蛋无法复原。当前的基准测试，如 BEHAVIOR-1K [Li 等人，2023a] 和 ARNOLD [Gong 等人， 2023]，明确测试了这种组合结构，结果表明，随着任务时域从 5 步增加到 20 步，成功率会急剧下降（通常下降 50% 或更多）。

6.4 地理空间分析：智能体如何在行星尺度上进行推理?

微观摘要：本小节介绍宏观尺度地理空间分析的方法，并对关键方程进行直观解释。对于在城市或行星尺度上运行的智能体，这些方法使其能够推理那些在数小时至数年间发生的现象，因此所需的规划时域与房间尺度任务有着根本性的不同。

地理空间分析需要处理卫星影像 [Zhu 等人，2017，Ma 等人，2019b，Yuan 等人，2020a，Zhang 等人，2016，Tuia 等人，2016，Camps-Valls 等人， 2014], 对城市动态进行建模 [Bibri 和 Krogstie，2017，Batty，2013，Kitchin，2014，Jiang 等人， 2021], 并推理地理关系[Egenhofer 和 Franzosa，1991，Cohn 和 Renz，2008，Randell 等人， 1992]。

遥感基础模型。Prithvi [Jakubik 等人， 2024]提供了基于协调化 Landsat Sentinel-2 数据训练的地理空间基础模型。SatMAE [Cong 等人， 2022] 为卫星影像引入了掩码自编码器。Satlas [Bastani 等人， 2023] 支持大规模地理空间理解。GeoAI [Janowicz 等人，2020，Mai 等人， 2023] 综述了这一领域。CROMA [Fuller 等人， 2024] 和 microestimates [Chi 等人， 2022] 推动了遥感分析的发展。

时空图神经网络。 时空图神经网络通过图结构表示来建模复杂的城市动态 [Jin 等人，2023年，Atluri 等人，2018年，Wang 等人，2020年，Jiang 和 Luo，2022年，Balachandar 等人， 2025]。其核心直觉是，空间位置（交叉路口、传感器、街区）构成一个图，边表示邻近性或连通性，而每个节点的状态会随时间演变。其一般形式将空间消息传递（信息如何在相连位置之间流动）与时间卷积（每个位置的状态如何随时间演变）结合起来：

$$\mathbf{H}^{(l+1)} = \sigma\left(\mathbf{A}\mathbf{H}^{(l)}\mathbf{W}^{(l)} + \text{TemporalConv}(\mathbf{H}^{(l)})\right) \tag{1}$$

其中，$\mathbf{A}$ 表示空间结构（哪些位置会影响哪些位置），$\mathbf{H}$表示节点状态，而时间卷积则捕捉模式随时间的演变。

DCRNN [Li 等人， 2018] 将交通建模为图上的扩散，捕捉了这样的直觉：拥堵在道路网络中的传播，就像热量在材料中的扩散一样：

$$\mathbf{H}^{(l)} = \sum_{k=0}^{K} \left(\mathbf{P}_f^k \mathbf{X}\mathbf{W}_{k,1} + \mathbf{P}_b^k \mathbf{X}\mathbf{W}_{k,2}\right) \tag{2}$$

前向矩阵 $\mathbf{P}^f$ 建模下游传播（交叉路口的拥堵如何影响从该路口延伸出去的道路），而 $\mathbf{P}_b$ 建模上游影响（下游拥堵如何回堵至汇入道路）。

STGCN [Yu 等人， 2018]通过夹层结构将图卷积与时间卷积相结合。Graph WaveNet [Wu 等人，2019b] 解决了一个关键局限：预定义的邻接矩阵（基于道路物理连接）可能会遗漏重要的依赖关系。两个相距甚远的地点即使没有直接的道路连接，也可能因共享的通勤模式而高度相关。Graph WaveNet 从数据中学习图结构：

$$\tilde{\mathbf{A}} = \text{SoftMax}\left(\text{ReLU}\left(\mathbf{E}_1\mathbf{E}_2^T\right)\right) \tag{3}$$

可学习的嵌入$\mathbf{E}_1$,$\mathbf{E}_2$ 能够根据各地点的交通模式（而不仅仅是物理邻近性）发现哪些地点应当相互连接。

AGCRN [Bai 等人，2020年，Song 等人， 2020] 通过自适应模块引入节点特定模式。ASTGCN [Guo等人， 2019, 2021] 加入了空间和时间注意力机制。GMAN [Zheng 等人，2020年，Park 等人， 2020] 使用图多重注意力与变换注意力来处理长程依赖关系。STGRAT [Choi 等人， 2022] 推动了该领域的发展。

城市计算。 城市计算 [Zheng 等人，2014年，Yuan 等人，2020b，Zheng， 2015] 将人工智能应用于城市尺度挑战。ST-LLM [Liu 等人，2024a] 和 UniST [Yuan 等人， 2024] 将语言模型与时空推理相结合。交通预测 [Li 等人，2018年，Yu 等人，2018年，Wu 等人，2019b] 和需求预测 [Geng 等人，2019年，Yao 等人，2018年，Zhang 等人， 2017]是关键应用。

关键要点：空间领域

- 通过视觉-语言集成，导航已从点目标范式发展到语言引导和零样本范式
- 神经隐式表示（NeRF、3DGS）结合语义落地，有助于场景理解
- 视觉-语言-动作模型正推动操作领域变革，将网络规模知识迁移到机器人控制中
- 地理空间分析日益运用基础模型和图神经网络进行行星尺度推理

7 使能技术

微观摘要：本节回顾了支撑现代空间人工智能系统的三项核心技术：用于关系推理的图神经网络、用于预测模拟的世界模型，以及用于感知和落地的多模态基础模型。我们通过直观的示例介绍每项技术，并将其与分类体系联系起来。

7.1 图神经网络：智能体如何推理关系?

微观总结： 本小节说明图神经网络如何建模空间关系，并直观解释消息传递机制及其在微观和宏观尺度空间推理中的应用。

为什么在关于空间智能体的综述中纳入图神经网络？答案在于它们的归纳偏置。Transformer 将输入视为无序集合（或带有学习得到的位置编码的序列），因此对空间数据固有的关系结构并不敏感。相比之下，图神经网络会将关系显式编码为边，从而为表示空间约束提供了自然方式：例如场景中的物体、道路网络中的交叉路口，或机械臂中的关节。这种关系偏置使图神经网络尤其适用于连接结构与节点内容同等重要的任务 [Kipf 和 Welling，2017年，Veli˘ckovi´c 等人，2018年，Xu 等，2019年，Hamilton 等人，2017年，Wu 等人，2020b,a，Zhou 等，2020年，Zhang 等人，2020年，Li 等人，2016年，Defferrard 等人，2016年，Bruna 等人， 2014]。

消息传递框架。 通用图神经网络公式遵循消息传递范式[Gilmer 等人，2017年，Scarselli 等人，2009年，Battaglia 等人，2018年，Xu 等，2018年，Morris 等人， 2019]:

$$\mathbf{m}_v^{(l)} = \text{AGGREGATE}^{(l)}\left(\left\{\mathbf{h}_u^{(l-1)} : u \in \mathcal{N}(v)\right\}\right) \tag{4}$$

$$\mathbf{h}_v^{(l)} = \text{UPDATE}^{(l)}\left(\mathbf{h}_v^{(l-1)}, \mathbf{m}_v^{(l)}\right) \tag{5}$$

其中， $\mathcal{N}(v)$ 表示节点 v 的邻居节点，AGGREGATE 和 UPDATE 是可学习函数。从几何角度来看，可以将其理解为每个节点都在“倾听”邻居节点、收集它们的信息，然后根据所听到的内容更新自身表征；这就像机器人关节通过感知相邻关节的位置，更新对机械臂配置的理解。

GNN-LLM集成。 一个关键的概念性洞见是：图可以充当基于大语言模型的智能体的外部化空间记忆 。尽管大语言模型擅长语义推理，却缺乏对空间关系的持久、结构化表征。通过将场景图、道路网络或物体配置编码为图，并利用图神经网络对其进行处理，智能体便能维护可跨推理步骤持续存在、且不受上下文窗口限制影响的空间状态。新兴研究将图神经网络与大语言模型结合，用于这种结构化空间推理 [Chen 等，2024e，Tang 等人，2024年，Fatemi 等人，2023年、2024年，Gowda 等人，2025年，Ye等人，2024年，Zhao 等人，2023a，Huang 等人，2024c]。图指令微调 [Zhang 等人，2024b] 进一步增强了这一能力。LLaGA [Chen 等，2024e] 实现了语言-图对齐。GraphGPT [Tang 等人，2024]能够通过语言模型进行图推理。

几何深度学习。 几何深度学习 [Bronstein 等人 2021] 为非欧几里得域上的空间推理奠定了理论基础。等变网络 [Cohen 和 Welling，2016年，Fuchs 等人，2020年，Satorras 等人，2021年，Weiler 和 Cesa，2019年，Thomas 等人，2018年，Kondor 等人，2018] 通过以下方式遵循空间对称性：

$$f(T_g \cdot x) = T_g \cdot f(x) \tag{6}$$

其中 T_g 是群变换。图 Transformer [Ying 等人，2021年，Dwivedi 等人，2023年，Ramp´a˘sek 等人，2022年，Kreuzer 等人，2021年，Chen 等人，2022e] 将注意力与图结构相结合。E3NN [Batzner 等人， 2022], 几何消息传递 [Brandstetter 等人， 2022], 和 SchNet [Sch¨utt 等人， 2017] 推动了等变架构的发展。

7.2 世界模型：智能体如何预测未来?

微观摘要： 本小节介绍世界模型，它们使智能体能够模拟其动作的未来结果。我们对比基于模型的方法与无模型方法，并讨论世界模型在

安全高效规划中的作用。

要让智能体在物理世界中安全、有效地行动，它必须能够预测其动作的后果。世界模型通过学习世界的运行方式，提供这种预测能力，使智能体能够进行规划并预见未来 [LeCun, 2022, Schmidhuber, 2015, Matsuo 等人, 2022, LeCun, 2024b,a, Moerland 等人, 2023, Sutton, 1991, Deisenroth and Rasmussen, 2011]。这对于空间安全至关重要，因为它能让智能体避免进入具有物理危险或不可逆的状态。

潜在动力学模型。核心思想很简单：世界模型不直接预测未来图像（这种方式计算成本高，而且容易导致误差累积），而是将观测压缩到紧凑的潜在空间中，并在其中预测动力学。编码器将高维观测（例如图像）压缩为紧凑的潜在状态。随后，动力学模型根据当前状态和一个动作预测下一个潜在状态，从而在压缩空间中模拟未来。必要时，解码器可以重建观测。形式化地说：

$$\text{Encoder:} \quad \mathbf{z}_t = q_\phi(\mathbf{z}_t|\mathbf{o}_{\leq t}, \mathbf{a}_{<t}) \tag{7}$$

$$\text{Dynamics:} \quad \hat{\mathbf{z}}_{t+1} = p_\theta(\hat{\mathbf{z}}_{t+1}|\mathbf{z}_t, \mathbf{a}_t) \tag{8}$$

$$\text{Decoder:} \quad \hat{\mathbf{o}}_t = p_\psi(\hat{\mathbf{o}}_t|\mathbf{z}_t) \tag{9}$$

基于模型的强化学习。Dreamer [Hafner 等人，2020年、2019b] 通过循环状态空间模型引入了潜在想象，以实现样本高效学习。DreamerV2 [Hafner 等人，2021] 凭借离散潜在状态，达到了人类水平的 Atari 表现。DreamerV3 [Hafner 等人， 2023]通过对称对数预测，展示了单一算法在跨领域任务中的卓越能力。从 Dreamer 到 DreamerV3 的发展历程，展现了一条重要的研究轨迹：从连续潜在状态转向更精确的离散表征，最终发展为无需修改即可驾驭众多领域的单一、高度泛化模型。尤为重要的是，DayDreamer [Wu 等人，2023b]展示了这种方法在机器人学领域的巨大潜力：仅用少量真实世界数据进行微调，就成功将仿真中学习到的世界模型迁移到实体机器人上。这为克服仿真到现实差距指出了一条有前景的道路。PlaNet [Hafner 等人，2019a,b] 开创了潜在动力学学习。MuZero [Schrittwieser 等人， 2020] 将学习模型与蒙特卡洛树搜索相结合，用于游戏对弈。其他方法还包括 MBPO[Janner 等人，2019年，Chua 等人， 2018], 世界模型 [Ha 和 Schmidhuber， 2018], TD-MPC [Hansen 等人，2022, 2024], 以及 IRIS [Micheli 等人， 2023]。

视频世界模型 。一项近期进展是涌现出从视频中学习世界模型的研究，这类模型分为两种截然不同的类别，对空间智能体的意义也各不相同：

可控世界模型从无标注的互联网视频中学习交互式环境，从而能够进行动作条件预测。Genie [Bruce 等人，2024b] 允许用户控制生成世界中的角色，而 Genie 2 [Bruce 等人，2024a] 则将这一能力扩展到 3D环境。这些模型可直接用于智能体规划，因为它们能够仿真动作所带来的后果。

生成式世界模型专注于高保真视觉仿真，而不采用显式的动作条件。GAIA-1 [Hu 等人，2023a] 和 WorldDreamer [Yang 等人，2024d] 能够生成逼真的驾驶视频，而 Sora [OpenAI， 2024] 则展示了大规模视频生成能力。尽管这些模型在视觉效果上令人印象深刻，但它们不具备决策所需的动作条件结构，因此不太能直接用于智能体规划 [Yang等人，2024a，Baker等人，2022，Wu 等人，2024，Yan 等人，2021，Wu 等人， 2022]。

基于大语言模型的世界模型。大语言模型可作为用于规划的世界模型 [Hao 等人，2023年，Huang 等人，2022b], ，无需显式环境模型即可预测状态转移。这种方法利用大语言模型中编码的海量知识来模拟世界动态。然而，值得注意的是，尽管大语言模型能够建模抽象的状态转移，但目前尚不具备专用世界模型那样精细的物理保真度；对于需要精确几何与物理推理的任务而言，这是一个关键局限。RAP [Hao 等人，2023] 通过世界模型展开，将推理与行动结合起来。TransDreamer [Chen 等人，2022b],、UniSim [Yang 等人，2023a], 和 Genie 2 [Bruce 等人，2024a] 推动了世界建模的发展。在我们分类体系的中观与宏观尺度上，这些模型尤其适用于规划能力，因为在这些尺度上，抽象的高层级预测往往已足够。

7.3 多模态基础模型：智能体如何将语言定位于感知之中?

微观摘要：本小节回顾了将语言与视觉及其他感官数据相连接的多模态基础模型。我们着重阐述它们在为智能体系统提供语义落地方面的作用，并指出当前空间理解能力的局限。

多模态模型整合视觉、语言和动作理解，为智能体系统提供感知基础 [Baltruˇsaitis 等，2019年，Xu 等，2023c，Liang 等，2024年，Ngiam 等，2011年，Srivastava 和 Salakhutdinov，2012年，Ramachandram 和 Taylor， 2017]。必须明确一个关键区别：多模态性是实现空间能动性的必要条件，但并非充分条件。能够描述场景（“桌子上有一个红色杯子”）的模型，并不因此就具备抓取那个杯子的能力。感知提供了是什么和在哪里。能动性还需要具备规划 （决定要做什么）、记忆 （跟踪随时间变化的状态）和动作 （执行运动指令）等能力。缺少这些机制，多模态模型就仍然只是被动的观察者，而非世界中的主动参与者。

视觉语言模型。 基础视觉语言模型大致可分为两种范式。**对比模型** （如 CLIP [Radford 等人， 2021] 以及 BLIP-2 [Li 等人，2023d] ）学习图像与文本之间的共享嵌入空间，从而实现强大的零样本分类。**指令微调模型**（如 LLaVA [Liu 等人，2023b、2024b], 、InstructBLIP [Dai 等人，2023b], 以及 Ferret [You 等人，2023、Zhang 等人，2024a], ）经过对话数据微调，能够遵循用户指令，在视觉问答和对话任务中表现出色。GPT-4V [OpenAI，2023，Zheng 等人，2024，Yan 等人， 2023], Gemini [Team 和谷歌， 2023], Flamingo [Alayrac 等人， 2022], 以及 PaLI [Chen 等人，2023d,e]等大规模模型进一步拓展了多模态推理的边界；Qwen-VL [Bai 等人， 2023], 、CogVLM [Wang 等人，2023b], 和 InternVL [Chen 等人，2024h]等开源替代方案则促进了更广泛的使用。

空间视觉语言模型。 SpatialVLM [Chen 等人，2024a，Yang 等人，2024b] 专注于空间推理，并具备细粒度理解能力。SpatialRGPT [Cheng 等人， 2024] 能够进行区域级空间推理。VoxPoser [Huang 等人，2023b] 将视觉语言模型中的可供性提取到3D表示中。这形成了两种落地方式的区别：**以地图为中心的落地**，例如 VLMaps [Huang 等人，2023a]会构建环境的显式语义地图以用于导航；以及**以可供性为中心的落地**，例如 VoxPoser [Huang 等人，2023b] 会将可执行的可供性直接推断到场景的三维表征上，而无需构建完整地图。这些模型构成了被动感知与主动规划之间的关键桥梁，将高级指令落地到具体的空间表征中。

3D视觉语言模型。 3D-LLM [Hong 等人，2023b，Chen 等人，2024f，Zhu 等人， 2023]使得能够理解3D场景中的语言。Open3D-VQA [Zhang 等人，2025b] 提供开放词汇3D视觉问答。LLM-Grounder [Yang 等人，2024b]将语言落地到3D环境中。转向3D表示从根本上增加了推理复杂度，因为智能体现在必须应对遮挡、视点变化和体积属性等概念，而这些概念在基于2D图像的推理中并不存在。

关键要点：使能技术

- GNN-LLM集成代表着范式转变，将关系推理与语义理解结合起来
- 世界模型通过想象实现样本高效学习和安全规划
- 空间视觉语言模型（SpatialVLM、VLMaps、VoxPoser）弥合了视觉语言理解与空间动作之间的鸿沟
- 等变架构为几何推理提供了有理论依据的方法
- **注意：**这些模型的性能高度依赖大规模、高质量的数据，在面对分布外场景时也可能十分脆弱。

8 工业应用作为设计模式

从基础技术出发，接下来我们将探讨它们如何集成到大规模部署的系统中。通过设计模式的视角分析真实世界中的行业应用，为科学探究提供了一种强有力的方法。这使我们能够超越定制化的系统描述，转而识别空间人工智能领域中针对常见问题反复出现且可推广的解决方案。这种方法将理论分类体系立足于实践，不仅揭示了使用了哪些能力，还展示了这些能力如何在商业可行性、安全性和尺度的约束下进行组合。这些模式不仅仅是工程解决方案，也是系统设计中趋同演化的实证证据，能为学术研究人员和行业从业者提供宝贵见解。

我们不对企业能力进行编目，而是将工业部署抽象为可复用的空间人工智能系统设计模式。这些模式在真实世界约束下体现了三轴分类体系中的特定区域：每种模式都代表智能体能力（记忆、规划、工具使用）、空间任务领域（导航、操作、地理空间推理）和运行尺度（微观、中观、宏观）的特定组合，展现了理论框架如何转化为实际部署的系统。

遵循软件架构领域确立的形式体系 [Gammaet al., 1994], ，以下每种模式都说明了一个问题、制约因素和解决方案。

8.1 设计模式 1：人在回路中的空间推理

问题： 高风险空间决策所需的准确性超出现有人工智能能力。制约因素： 问责要求、责任问题、领域专业知识稀缺。解决方案： 采用三步式人在回路流程：(1) 人工智能智能体提出空间分析方案；(2) 人类专家验证或修正输出；(3) 利用反馈更新系统的记忆或策略，以实现持续改进。

该模式将人工智能空间分析与人类专家验证相结合 [Amershi等人，2019年，Shnei- derman，2022年，Wang 等人，2019a，Green和Chen，2019年，Fails和Olsen Jr，2003年，Stumpf等人，2009年，Holzinger，2016年，Dudley和Kristensson，2018年，Zanzotto， 2019], 例如：

地理空间智能。 地理空间智能融合平台将多源空间数据（卫星影像、信号情报、地形模型）与人类分析师工作流程相结合 [Palantir，2023年，Palantir Technologies， 2024]。这些系统能够在开展空间推理的同时进行情报分析，并对关键决策保持人工监督；这一要求源于国防背景下的问责与法律约束。

GIS 工作流程。 用于城市规划与环境监测的地理信息系统（GIS）集成人工智能能力，以协助人类规划人员 [ESRI，2023年，2024b,a]。该工作流程直接实现了该模式中定义的三步示意图。

8.2 设计模式 2：弱监督行星尺度学习

问题：宏观空间推理必须处理海量无标签数据集，以及由不同地理区域、传感器类型和大气条件造成的极端分布偏移。制约因素：在行星尺度上进行密集标注，无论从经济还是后勤角度都不可行。解决方案：采用两阶段流程：（1）基于拍字节级无标签卫星影像进行自监督预训练，以学习强大的通用表征；随后（2）使用少量标注数据进行快速、特定任务的微调。

这种模式利用海量无标签数据和弱监督来构建全球尺度模型 [Ratner 等，2017，Zhang 和 Yang，2022，Zhou，2018，Chapelle 等，2009，Zhu，2005，Oliver 等，2018，Lee，2013，Tarvainen 和 Valpola， 2017]:

卫星基础模型。 这一模式遵循清晰的抽象：大规模数据规模（来自 Landsat、Sentinel、Planet 和 Maxar 等提供商的多光谱影像，规模达拍字节）通过自监督技术（例如掩码自编码）实现强大表征的学习，随后可针对土地覆盖分类、变化检测或作物产量预测等各种下游任务进行微调。NASA-IBM Prithvi [Jakubik 等人， 2024] 使用协调化 Landsat Sentinel-2 数据，展现了这一流程。Planet Labs [Planet Labs PBC， 2023, 2024] 和 Maxar 提供了数据基础设施，支持每日全球监测。

大规模制图。 大型制图平台通过基于云的地理空间数据服务部署人工智能，开展全球尺度分析 [谷歌，2023年、2024b、a]。其学习闭环十分明确：用户与地图的交互（例如更正商家位置、报告道路封闭）会持续提供弱监督信号。这些信号经汇总后用于自动更新底层地理空间模型，从而提升全球范围内的地图准确性和时效性。

8.3 设计模式 3：智能体辅助空间工作流

问题：空间分析任务对人类专家而言重复且耗时。作用力：需要可扩展性，希望保留人类自主性，任务复杂度可变。解决方案：人工智能智能体作为主要分析员运作。人类提供高层指导并处理异常情况。

与 *HITL* 的对比：该模式不同于 HITL（模式 1）：

维度	HITL（模式 1）	智能体辅助（模式 3）
主要操作者	人类	人工智能智能体
人工智能角色	提案生成器	主要分析员
人类角色	强制验证者	异常处理者
验证	每项决策	仅异常
吞吐量	较低（由人类把关）	较高（由智能体驱动）
风险承受能力	低（高风险场景）	中等（错误可恢复）

该模式部署人工智能智能体，以增强人类空间推理 [Shneiderman，2020年，Horvitz，1999，Amershi 等人，2014年，Abdul 等人，2018年，Gillies 等人，2016年，Yang 等人， 2020]:

自主地理信息系统。 AutonomousGIS [Li 等人，2025a] 与 GeoGPT [Bai 等人， 2024] 将智能体能力与地理空间分析相结合。该模式涉及能够查询空间数据库、生成地图并回答地理问题的基于大语言模型的智能体。**差距：** 当前能力边界取决于智能体解决自然语言中空间歧义（例如“靠近市中心”）的能力，以及在融合不同地理空间数据库时处理模式异质性的能力。

位置智能。 Foursquare [Foursquare， 2023, 2024]与 Carto [CARTO， 2023, 2024]提供带有人工智能驱动分析的基于位置的服务。Wherobots [Wherobots， 2023, 2024] 提供云原生空间分析。该模式为：具备人工智能驱动查询与分析的空间数据基础设施。当前局限性： 在整合异构空间数据库时出现模式不匹配，以及自然语言查询中的空间歧义（例如，“靠近市中心”在不同城市中的解释各不相同）。

8.4 设计模式 4：大规模具身人工智能

安全性是首要设计约束。 与前几种模式不同，前几种模式中的错误可以恢复（例如，错误的地图标注可以更正），而具身人工智能运行于错误可能造成人身伤害、财产损失甚至生命损失的领域。这从根本上改变了设计空间：安全性不是一项可以事后添加的功能，而是塑造每项架构决策的首要约束。

问题： 在安全关键型物理系统中部署空间人工智能，而系统失效会造成无法接受的代价。制约因素： 严格的监管框架（例如美国国家公路交通安全管理局、欧盟人工智能法案）、责任问题、在长尾安全场景中验证性能的挑战，以及赢得公众信任的需求。 解决方案： 采用安全优先方法：通过大规模仿真进行初始策略学习，在真实世界中谨慎部署并实施人工监督，同时在严格的安全性与验证约束下，持续利用车队数据进行学习。

这种模式将学得的空间策略部署于物理系统中 [Kober 等人，2013年，Levine 等人，2016年，Ibarz 等人，2021年，Kalashnikov 等人，2021年，Julian 等人，2020年，Akkaya 等人，2019年，Levine 等人，2018年，Tobin 等人， 2017]:

自动驾驶车辆。 自动驾驶汽车行业展现了两种截然不同的感知理念。以激光雷达为中心 的方法（例如，Waymo、Cruise）将激光雷达点云与相机图像相融合，以实现冗余的高精度深度估计，优先保障最大安全裕度，但代价是更高的传感器成本 [Waymo，2023年，2024年，？]。相比之下，纯视觉 方法（例如，特斯拉）依靠多摄像头系统以及强大的学习型深度估计算法，力求更低的硬件成本以及更具

可扩展性的解决方案，但对其感知模型的可靠性提出了极高要求 [特斯拉，2023年，2024]。然而，这两种范式都遵循大规模仿真、谨慎的真实世界部署以及从车队数据中持续学习这一总体模式，且全部处于交通安全法规的严格约束之内。

机器人学习平台。Open X-Embodiment [协作项目 2023] 由 Google DeepMind 和合作机构提供大规模机器人数据。Bridge Data [Walke 等人， 2023, 2024] 支持跨域迁移。其模式是：多样化的数据收集、基础模型训练，以及向特定具身形态迁移。

给研究人员的启示：行业模式

- **从抽象问题到设计模式：**工业解决方案往往能够解决反复出现的科学挑战。将其抽象为设计模式，可以为学术界揭示可推广的见解。
- **数据构筑护城河，监督形成瓶颈：**获取行星尺度的专有数据是工业界的一项关键优势，但无法为其标注会造成监督瓶颈。这凸显了自监督和弱监督学习领域的一项重大研究机遇。
- **人在回路中的参与是一个连续谱：**人工监督程度并非非此即彼，而是一种设计选择，涵盖从完全由人类控制到完全由智能体自主行动的连续范围。自适应自主性以及动态人机协作方面的研究至关重要。
- **安全性不容妥协：** 对于旨在实现现实世界部署的任何研究，安全性、验证和法规合规都不是事后才考虑的问题，而是必须从一开始就解决的核心研究问题。

9 评估框架与基准测试分析

9.1 现有基准测试

表 3 按照我们的分类体系总结了关键基准测试。阅读指南： 该表将各项基准测试对应到任务轴（导航、操作、场景理解）、尺度轴（微观／中观／宏观）和能力轴（所评估的智能体能力）。这种组织方式揭示了覆盖模式：导航基准测试主要集中于中观尺度评估，而微尺度操作和宏尺度地理空间推理仍未得到充分覆盖。

表格未能展现的内容：目前没有任何基准测试评估以下方面：(1) 仿真到现实迁移性能下降，(2) 跨尺度推理（例如，在宏观导航过程中执行微观操作的单一任务），(3) 持续数小时或数天的长时程性能，或 (4) 安全关键失效模式。这些缺口并非只是表格中缺少几行，而是反映了该领域在评估空间智能体方面的根本局限性。

表 3：按分类体系组织的空间人工智能基准

基准测试	空间任务	尺度	环境	主要指标	智能体能力
R2R [Anderson 等人，2018c]	导航	Meso	仿真	SPL，成功率	记忆 + 规划
REVERIE [Qi 等人， 2020]	导航	Meso	仿真	SPL, RGS	记忆 + 规划
Habitat [Savva等人， 2019]	导航	Meso	仿真	SPL	规划
AI2-THOR [Kolve等人， 2017]	导航 + 操作	Meso	仿真	SR	规划 + 工具使用
RLBench [James 等人， 2020]	操作	微观	仿真	SR	工具使用
Meta-World [Yu 等人，2020]	操作	微观	仿真	SR	工具使用
nuScenes [Caesar 等人，2020]	场景理解	中观-宏观	Real	平均精度均值，NDS	记忆
KITTI [Geiger 等人， 2012]	场景理解	Meso	Real	mAP	记忆
ScanNet [Dai 等人， 2017]	场景理解	Meso	Real	mIoU	记忆
AgentBench [Liu 等人，2023c]	通用	-	混合	SR	记忆 + 规划 + 工具使用
WebArena [Zhou 等，2023b]	Web	-	仿真	SR	规划 +工具使用
SWE-Bench [Jimenez 等， 2024]	Code	-	Real	Pass@k	规划 +工具使用
EmbodiedBench [Yang 等人，2025]	具身	Meso	仿真	SR	记忆 + 规划 + 工具使用
SafeAgentBench [Yin 等人，2025]	安全性	-	仿真	安全率	规划

9.2 评估指标

我们汇总了不同领域的标准化指标 [Powers，2011年，Sokolova 和 Lapalme，2009年，Hossin 和 Sulaiman， 2015]。每个领域的关键指标如下：

导航： SPL（按路径长度加权的成功率） [Anderson 等人，2018b], 定义为 SPL $=\frac{1}{N}\sum_{i=1}^{N} S_i \cdot \frac{\ell_i}{\max(\ell_i,\ p_i)}$，综合奖励任务完成和路径效率。

操作： 任务成功率，并通过目标条件满足度提供部分得分 .

场景理解： 目标检测使用 mAP（平均精度均值），分割使用 mIoU。

安全性： 风险感知成功率 = SR × (1 - 碰撞率)，对不安全的任务完成予以惩罚 s.

9.3 批判性分析：基准测试未能衡量什么

尽管现有基准测试推动了该领域的发展 [Raji 等人，2021年，Liao 等人，2021年，Ribeiro 等人，2020年，Buolamwini 和 Gebru，2018年，Mitchell 等人，2019年，Gebru 等人，2021年，Denton 等人，2020年，Bender 等人，2021], 仍有若干根本局限值得批判性审视：

仿真—现实差距。 大多数基准测试依赖仿真环境 [Savva 等人，2019年，Kolve等人，2017年，James 等人， 2020]。 具体示例： RT-1 在仿真中的成功率达到 97%，但在真实机器人上降至 68% [Brohan 等人， 2022]。在 Matterport3D 中训练的视觉语言导航智能体，在真实建筑中的性能下降了 40% [Anderson 等人，2018c]。 差距：没有基准测试衡量仿真到现实迁移性能下降。

指标局限。 当前指标无法捕捉三个关键维度：

效率： SPL 假设已知最优路径，这在新环境中并不现实。二元成功率会忽略部分进展（距离目标仅 1 厘米的智能体，与从未移动过的智能体得分相同）。

安全性： 目前没有标准指标会对险些发生的事故、危险行为或未导致失效的违反约束行为进行惩罚。险些发生碰撞的智能体，与始终保持安全裕度的智能体得分相同。

泛化： 指标是在固定测试集上计算的，因此无法反映系统对分布偏移、 新颖物体或环境扰动的鲁棒性。

不足：指标奖励任务完成，却不奖励高效、安全或可泛化的行为。

长时程评估。 大多数基准测试评估的都是短回合（数十到数百步）。真实世界的任务要求智能体在数小时甚至数天内持续保持良好表现，并具备记忆持久性和错误恢复能力。这一限制与大语言模型架构直接相关：即使上下文窗口有 128K 个 token，以每秒接收 30 条观测的智能体也会在不到 90 分钟内耗尽上下文。更长的任务需要外部记忆系统，而当前的基准测试并未对其进行评估。智能体应如何在较长时间范围内压缩、检索和更新空间知识这一根本问题，仍未经过测试。不足：没有基准测试评估具备持久记忆的多日空间任务，也没有基准测试衡量智能体在接近上下文限制时性能会如何逐渐下降。

安全关键评估。 基准测试很少评估失效模式、对抗鲁棒性或分布偏移下的行为。安全关键应用需要了解最坏情况性能。差距：安全评估仍然是临时性的，而非结构化的。

跨尺度评估。 这是最关键的差距，直接违背了我们三轴框架中的尺度轴。 基准测试仅针对单一空间尺度，而现实应用要求实现无缝的跨尺度推理。以家用机器人为例：它必须规划房间级路径（中观）、避开家具（中观），并抓取杯柄（微观）。这些操作都发生在同一项任务中。目前没有任何基准测试能够评估这种整合能力。 差距：尽管跨尺度空间推理对于现实世界部署至关重要，却没有基准测试对其进行评估。

9.4 SpatialAgentBench：未来研究的概念框架

我们并非要提出又一个基准数据集，而是要概述一种用于基准设计的框架 ，以解决上述结构性差距。该框架的价值不在于具体的任务或环境，而在于其阐明的设计原则；任何未来的基准测试都应遵循这些原则，才能有意义地评估空间智能体。

框架原则：

- **多轴覆盖：** 任务应同时要求来自多个分类轴的能力

- **跨尺度整合：** 单个任务应涵盖微观、中观和宏观尺度
- **仿真到现实协议：** 基准测试应同时纳入匹配的仿真评估和真实世界评估
- **长时程压力测试：** 回合应持续至上下文窗口限制及更长时间
- **安全性指标：** 即使任务成功，评估也应对不安全行为予以惩罚

分类体系映射： 每种差距都对应一项维度限制。仿真—现实差距会影响所有维度（策略无法迁移）。指标局限会影响能力轴（对记忆、规划和工具使用的评估不完整）。长时程差距和安全性差距会影响能力轴（规划和自我反思）。跨尺度差距会影响尺度轴（缺乏微观—中观—宏观整合）。

我们确定了八个 研究方向 以推进未来的基准测试开发，每个方向都针对一项具体差距 ：

1. **VLN-Instruct** （任务：导航，能力：记忆 + 规划，尺度：中观）：需要空间推理的复杂多步指令。2. **ObjectSearch**（任务：导航，能力：记忆，尺度：中观）：关于物体位置的语义推理。3. **SceneQA** （任务：场景理解，能力：记忆，尺度：中观）：三维空间关系推理。4. **ManipSeq** （任务：操作，能力：规划 +工具使用，尺度：微观）：包含状态跟踪的长时程操作。5. **GeoReason** （任务：地理空间，能力：记忆，尺度：宏观）：卫星图像分析与变化检测。6. **TrafficPredict** （任务：地理空间，能力：记忆 + 规划，尺度：宏观）：时空城市动态。7. **SafeNav** （任务：导航，能力：规划，尺度：中观）：具备风险意识的安全约束导航。8. **MultiAgent** （任务：全部，能力：全部，尺度：中观–宏观）：协调的多智能体空间任务。

关键要点：评估

- 现有基准测试分散于各个领域，且指标彼此不兼容
- 仿真到现实迁移、长时程、安全关键型任务和跨尺度评估方面仍存在重大缺口
- SpatialAgentBench 被概述为一个概念框架，用于统一评估导航、操作、场景理解和地理空间推理
- 标准化指标（SPL、nDTW、安全率）有助于进行跨域比较

10 项重大挑战与未来方向

上文指出的基准差距——仿真到现实迁移、长时程评估、安全关键评估和跨尺度推理——直接揭示了该领域必须应对的根本性研究挑战。这些不仅是评估问题，更是限制空间人工智能各领域进展的结构性障碍。我们确定了六项需要关注的重大挑战，研究领域必须予以应对 [Marcus, 2020, Chollet, 2019, Lake et al., 2017, Bengio et al., 2019, Bommasani et al., 2021, LeCun, 2022, Kaplan et al., 2020, Hoffmann et al., 2022, Sutton, 2019, Bengio et al., 2023, Mitchell, 2021]:

10.1 重大挑战 1：统一空间表征

智能体如何维持单一且连贯的空间表征，以支持跨微观、中观和宏观尺度的推理？

当前方法针对不同尺度采用独立的表征：用于抓取的点云 [Rusu and Cousins, 2011, Fang等人, 2020, Mahler et al., 2017, Morrison et al., 2018, ten Pas et al., 2017, Lenz et al.,

2015], 用于导航的拓扑地图 [Thrun, 1998, Kuipers and Byun, 1991, Kuipers, 2000, Konolige et al.,2008], 以及用于地理空间分析的栅格影像 [Goodfellow et al., 2016, LeCun et al., 2015]。统一表征将支持跨尺度的无缝推理。关键研究方向包括:

- 从物体部件延伸到城市基础设施的分层场景图
- 支持多尺度查询的神经隐式表示
- 用于三维理解的基础模型 [Hong 等人，2023b，Fu 等人，2024a，Shen 等人，2023a，Oquab 等人，2024年，2023年，Chen 等人，2023a，Zhou 等人，2024年，Wu 等人，2015年，Girdhar 等人，2023年，Xu 等， 2024]

10.2 重大挑战 2：基于现实的长时程规划

智能体如何在保持几何可行性的同时进行长时程规划?

大语言模型能够生成高层规划，但难以处理几何约束 [Valmeekam 等人，2023b，Kambhampati 等人，2024年，Huang 等人，2024d，Valmeekam 等人，2024年，Stechly 等人，2024年，Helmert， 2006]。任务与运动规划系统能够处理几何，但缺乏语义灵活性。弥合这一差距需要:

- 结合大语言模型推理与几何验证的混合神经符号规划器
- 利用学习得到的抽象进行分层规划 [Song 等人，2023年，Valmeekam 等人，2023a，Huang 等人，2022a，Li 等人，2025b，Silver 等人，2024年，Liu 等人，2023a]
- 能够预测语义和几何后果的世界模型

10.3 重大挑战三：不确定条件下的安全部署 y

空间人工智能系统如何才能在安全关键应用中安全运行，并确保失效风险处于可保证的界限内？当前系统缺乏形式化安全保证 [Seshia 等人，2022 年，Koopman 和 Wagner，2019 年，Amodei 等人，2016a，Hendrycks 等人，2022 年，Ngo 等人，2022年，Leike 等人，2018年，Russell，2019年，Christiano 等人，2017]。在自动驾驶车辆、医疗机器人和基础设施中部署这些系统，需要:

- 空间预测的不确定性量化
- 针对新环境的分布外检测
- 空间推理的形式化验证 [Yin 等，2025，Amodei 等人，2016b,a，Bai 等人，2022，Ouyang 等人， 2022]
- 对抗条件下的优雅降级

10.4 重大挑战 4：仿真到现实迁移

如何将仿真中学到的策略可靠地迁移到物理世界?

现实差距会影响感知、动力学和控制 [Peng 等人，2018年；Rusu 等人，2017年；Sadeghi 和 Levine，2017年；Bousmalis 等人，2018年；Ho 等人，2021年；James 等人， 2019]。弥合这一差距需要:

- 结合精确物理的照片级真实感仿真 [Zhao 等人，2020年；Tobin 等人，2017年；James 等人，2019年；Matas 等人，2018年；Muratore 等人， 2022]
- 域随机化与适应
- 利用最少数据进行现实世界微调
- 仿真与现实混合训练流程

10.5 重大挑战 5：可扩展多智能体协调

大量空间智能体如何在通信受限的情况下有效协同 协调？

当前多智能体系统的扩展性较差 [Stone 和 Veloso, 2000, Busoniu 等人, 2008年, Foerster 等人, 2016年, Lowe 等人, 2017年, Rashid 等人, 2018年, Sunehag 等人, 2018年, Son 等人, 2019年, Yu 等人, 2022a, Hernandez- Leal 等人, 2019]。真实应用（仓储机器人技术、自动驾驶交通）需要：

- 用于空间协调的涌现式通信协议
- 兼顾全局一致性的去中心化规划 [Zhang 等人, 2021a, Wu 等人, 2023c, Hong 等人, 2023a, Li 等人, 2023c, Qian 等人, 2023年, Chen 等人, 2024g, Talebirad 和 Nadiri, 2023年, Park 等人, 2023年, Li 等人, 2023b]
- 异构智能体协调
- 部分可观测条件下的可靠协调

10.6 重大挑战 6：高效边缘部署

具备能力的空间人工智能系统如何在资源受限平台上运行？基础模型需要大量计算资源。边缘部署需要：

- 在不损失能力的前提下进行模型压缩 [Han 等人，2016年，Howard 等人，2017年，Dehghani 等人，2023年，Hinton 等人，2015年，Frankle 和 Carbin，2019年，Jacob 等人，2018年，Sandler 等人，2018年，Tan 和 Le，2019年，Gholami 等人， 2022]
- 用于空间推理的高效架构
- 面向空间人工智能的软硬件协同设计
- 基于任务难度的自适应计算资源分配

重大挑战概要

1. **统一表征**：涵盖微观到宏观尺度的单一表征 2. **有据规划**：具备几何可行性的长时域规划 3. **安全部署**：面向关键应用的形式化安全保证 4. **仿真到现实**：从仿真到物理世界的可靠迁移 5. **多智能体**：在有限通信条件下实现可扩展协调 6. **边缘部署**：在资源受限平台上运行的高性能系统

11 局限性

本文综述存在若干局限性：

- 尽管我们的论文筛选流程十分全面，但仍可能遗漏相关领域或非英语出版物中的相关研究。
- 尽管所提出的分类体系具有统一性，但它只是众多可能的分类方式之一，未必能涵盖该领域的所有细微差异。
- 我们的分析基于公开信息，不包含工业界实验室的专有细节。
- 该领域发展迅速，一些近期研究可能尚未得到充分呈现。
- 我们主要关注发表于主要会议与期刊的英语出版物。
- 所提出的 SpatialAgentBench 目前仍处于概念阶段，有待实现和验证。
- 我们对行业应用的分析依赖于公开信息，可能无法反映当前的能力。

12 结论

本文综述提出了一种统一三轴分类法，将不同空间尺度上的智能体人工智能与空间智能联系起来，并综合了742篇引用文献中关于基础架构、最先进的方法、行业应用和评估基准的见解。我们的分析揭示了三个关键发现：

1. **分层记忆系统** 对长时域空间任务至关重要，能够帮助智能体有效积累和检索空间知识。检索增强生成、情景记忆和空间记忆表征方面的进展，为持久的空间理解奠定了基础。2. **GNN-LLM集成** 是一种很有前景的方法，将图网络的关系推理能力与语言模型的语义理解能力相结合。这种集成能够实现结构化空间推理，同时利用几何关系和语义知识。3. **世界模型** 对安全部署至关重要，使智能体能够预测后果，并在行动前通过想象进行规划。视频世界模型、潜在动力学模型和基于大语言模型的世界模型为预测性理解提供了互补的方法。

我们确定了六项亟待关注的重大挑战：统一空间表征、基于现实的长时程规划、不确定性下的安全部署、仿真到现实迁移、可扩展的多智能体协调，以及高效边缘部署 [Russell and Norvig, 2010, 2021]。视觉-语言-动作模型、图神经网络、世界模型和基础模型的融合，为应对这些挑战提供了很有前景的方向。

通过以三轴分类体系建立这一基础性参考，并概述概念性的 SpatialAgentBench 框架，我们旨在推动能够感知、推理并在物理世界中采取行动的强大、可靠且安全、具备空间感知能力的自主系统的发展 [Wooldridge,2009]。智能体人工智能与空间智能的交叉是人工智能领域的重要前沿，对自动驾驶车辆、机器人学、城市计算和地理空间智能具有深远影响。

参考文献

Marah Abdin、Sam Ade Jacobs、Ammar Ahmad Awan 等。Phi-3 技术报告：可在手机本地运行的高性能语言模型。*arXiv* 预印本 *arXiv:2404.14219*，2024年。

Ashraf Abdul、Jo Vermeulen、Danding Wang、Brian Y Lim 和 Mohan Kankanhalli。可解释、可问责且易于理解的系统的发展趋势与轨迹：一项人机交互研究议程。*CHI*，第 1–18 页，2018年。

Michael Ahn、Anthony Brohan、Noah Brown、Yevgen Chebotar、Omar Cortes、Byron David、Chelsea Finn、Chuyuan Fu、Keerthana Gober、Karol Gopalakrishnan等人。《按我能做的来做，而不是按我说的做：将语言落地于机器人可供性》。*arXiv*预印本 *arXiv:2204.01691*，2022年。

Ilge Akkaya、Marcin Andrychowicz、Maciek Chociej、Mateusz Litwin、Bob McGrew、Arthur Petron、Alex Paino、Matthias Plappert、Glenn Powell、Raphael Ribas等人。《用机器人手解决魔方》。*arXiv*预印本 *arXiv:1910.07113*，2019年。

Ekin Akyurek等人。《什么学习算法属于上下文学习？》*arXiv*预印本 *arXiv:2211.15661*，2023年。

Jean-Baptiste Alayrac、Jeff Donahue、Pauline Luc、Antoine Miech、Iain Barr、Yana Hasson、Karel Lenc、Arthur Mensch、Katherine Millican、Malcolm Reynolds 等。Flamingo：一种用于少样本学习的视觉语言模型。神经信息处理系统进展，35:23716–23736，2022年。

Saleema Amershi、Maya Cakmak、W Bradley Knox 和 Todd Kulesza。赋权于民：人类在交互式机器学习中的作用。《*AI Magazine*》，35(4):105–120，2014年。

Saleema Amershi、Dan Weld、Mihaela Vorvoreanu、Adam Fourney、Besmira Nushi、Penny Collisson、Jina Suh、Shamsi Iqbal、Paul N Bennett、Kori Inkpen 等。《人机交互指南》。*CHI*，2019年。

N. Amin 和 D. Kiela。具身语言学习：机遇、挑战与未来方向。载于
计算语言学协会论文集：*ACL 2024*，2024年。

Dario Amodei、Chris Olah、Jacob Steinhardt、Paul Christiano、John Schulman 和 Dan Man´e。《人工智能安全中的具体问题》。*arXiv* 预印本 *arXiv:1606.06565*，2016年a。

Dario Amodei，Chris Olah，Jacob Steinhardt，Paul Christiano，John Schulman，和 Dan Mane。具体人工智能安全中的问题。*arXiv* 预印本 *arXiv:1606.06565*，2016b。

Dong An、Yuankai Wang、Yuankai Qi 等。BevBERT：面向语言引导
导航的多模态地图预训练。
2023年。

Peter Anderson、Angel Chang、Devendra Singh Chaplot、Alexey Dosovitskiy、Saurabh Gupta、Vladlen Koltun、Jana Kosecka、Jitendra Malik、Roozbeh Mottaghi、Manolis Savva 和 Amir R. Zamir。具身导航智能体的评估，2018a。

Peter Anderson、Angel Chang、Devendra Singh Chaplot、Alexey Dosovitskiy、Saurabh Gupta、Vladlen Koltun、Jana Kosecka、Jitendra Malik、Roozbeh Mottaghi、Manolis Savva 等人。关于具身导航智能体的评估。载于*arXiv*预印本 *arXiv:1807.06757*，2018b。

Peter Anderson、Qi Wu、Damien Teney、Jake Bruce、Mark Johnson、Niko Sunderhauf、Ian Reid、Stephen Gould 和 Anton van den Hengel。视觉与语言导航：在真实环境中解读以视觉为基础的导航指令。载于*IEEE*计算机视觉与模式识别会议论文集，第 3674–3683 页，2018c。

Rohan Anil、Andrew M Dai、Orhan Firat、Melvin Johnson、Dmitry Lepikhin、Alexandre Passos、Siamak Shakeri、Emanuel Taropa、Paige Bailey、Zhifeng Chen 等人。PaLM 2 技术报告。*arXiv*预印本 *arXiv:2305.10403*，2023年。

Thomas Antho ny等人。结合深度学习与树搜索的思考，快与慢。载于*NeurIPS*，201 7.

Anthropic。Claude 3模型卡。*Anthropic*技术报告，2024年。

Brenna D Argall、Sonia Chernova、Manuela Veloso和Brett Browning。从演示中学习的机器人学习综述。机器人与自主系统，57(5):469–483，2009年。

Iro Armeni、Zhi-Yang He、JunYoung Gwak、Amir R Zamir、Martin Fischer、Jitendra Malik和Silvio Savarese。三维场景图：一种面向统一语义、三维空间与摄像机的结构。*arXiv*预印本*arXiv:1910.02527*，2019年。

Akari Asai等人。Self-rag：通过自我反思学习检索、生成与评判。*arXiv*预印本*arXiv:2310.11511*，2023年。

Gowtham Atluri、Anuj Karpatne 和 Vipin Kumar。时空数据挖掘：问题与方法综述。*ACM Computing Surveys*，2018年。

Daichi Azuma、Taiki Miyanishi、Shuhei Kurita 和 Motoaki Kawanabe。ScanQA：用于空间场景理解的三维问答。发表于*CVPR*，2022年。

Pierre-Luc Bacon、Jean Harb 和 Doina Precup。选项评论家架构。*AAAI*，2017年。

Alan Baddeley。 工作记忆：回顾过去，展望未来。《自然评论：神经科学》，2003年 .

Jinze Bai、Shuai Bai、Shusheng Yang 等。Qwen-VL：一种用于理解、定位、文本阅读等任务的通用视觉语言模型。*arXiv*预印本 *arXiv:2308.12966*，2023年。

Lei Bai 等人. 用于交通流预测的自适应图卷积循环网络. 神经信息处理系统进展，2020年。

Yifan Bai 等人。GeoGPT：通过自主 GPT 理解和处理地理空间任务。*arXiv* 预印本 *arXiv:2307.07930*，2024年。

Yuntao Bai、Saurav Kadavath、Sandipan Kundu、Amanda Askell、Jackson Kernion、Andy Jones、Anna Chen、Anna Goldie、Azalia Mirhoseini、Cameron McKinnon 等人。宪法式人工智能：通过人工智能反馈实现无害性。*arXiv* 预印本 *arXiv:2212.08073*，2022年。Bowen Baker、Ilge Akkaya、Peter Zhokov、Joost Huizinga、Jie Tang、Adrien Ecoffet、Brandon Houghton、Raul Sampedro 和 Jeff Clune。视频预训练（VPT）：通过观看无标注的在线视频学习如何行动。神经信息处理系统进展，35：24639–24654，2022年。Sidhika Balachandar、Shuvom Sadhuka、Bonnie Berger、Emma Pierson 和 Nikhil Garg。利用图神经网络预测城市事件：整合政府评级与众包报告，2025年。网址：https://arxiv.org/abs/2506.08740。Tadas Baltruˇsaitis、Chaitanya Ahuja 和 Louis-Philippe Morency。多模态机器学习：综述与分类体系。*IEEE*模式分析与机器智能汇刊，2019年。Andrea Banino 等人。利用类网格表征在人工智能体中实现基于向量的导航。《自然》，2018年。

Jonathan T Barron、Ben Mildenhall、Dor Verbin、Pratul P Srinivasan 和 Peter Hedman。Mip-nerf 360：无界抗锯齿神经辐射场。发表于$CVPR$，2022。

Jonathan T Barron 等。Zip-nerf：基于网格的抗锯齿神经辐射场。发表于$ICCV$，2023。

Favyen Bastani、Piper Wolters、Ritwik Gupta、Joe Ferdinando 和 Aniruddha Kembhavi。Satlaspretrain：用于遥感图像理解的大规模数据集。发表于$ICCV$，2023。

Dhruv Batra、Aaron Gokaslan、Aniruddha Kembhavi、Oleksandr Maksymets、Roozbeh Mottaghi、Manolis Savva、Alexander Toshev 和 Erik Wijmans。重新审视物体导航：具身智能体导航至物体的评估。发表于 *arXiv* 预印本 *arXiv:2006.13171*，2020年。

Peter W Battaglia、Jessica B Hamrick、Victor Bapst 等人。关系归纳偏置、深度学习与图网络。*arXiv* 预印本 *arXiv:1806.01261*，2018年。

Michael Batty。大数据、智慧城市与城市规划，第 3 卷。2013年。

Simon Batzner 等人。用于高效数据处理和精确原子间势的 E(3) 等变图神经网络。《自然·通讯》，2022年。

Sarah R Beck、Ian A Apperly、Jackie Chappell、Catriona Guthrie 和 Nicola Cutting。制作工具并非儿童游戏。$Cognition$，119(2)：301–306，2011年。

Emily M Bender、Timnit Gebru、Angelina McMillan-Major 和 Shmargaret Shmitchell。论随机鹦鹉的危险：语言模型是否可能过于庞大？$FAccT$，第 610–623 页，2021年。

Yoshua Bengio、Tristan Deleu、Nasim Rahaman、Rosemary Ke、S´ebastien Lachapelle、Olexa Bilaniuk、Anirudh Goyal 和 Christopher Pal。《用于学习解耦因果机制的元迁移目标》。*arXiv* 预印本 *arXiv:1901.10912*，2019年。

Yoshua Bengio、Salem Lahlou、Tristan Deleu、Edward J Hu、Mo Tiwari 和 Emmanuel Bengio。《Gflownet 基础》。$JMLR$，24(210):1–55，2023年。

Maciej Besta、Nils Blach、Ales Kubicek、Robert Gerstenberger、Lukas Gianinazzi、Joanna Gajber、Tomasz Lehmann、Michal Podstawski、Hubert Niewiadomski、Piotr Nyczyk 等。《思维图：使用大语言模型解决复杂问题》。*arXiv* 预印本 *arXiv:2308.09687*，2023年。

Shariq Farooq Bhat、Ibraheem Alhashim 和 Peter Wonka。Adabins：使用自适应分箱进行深度估计。载于 *IEEE/CVF*计算机视觉与模式识别会议，第 4009–4018 页，2021年。

Simon Elias Bibri 和 John Krogstie。《未来的智慧可持续城市：一篇广泛的跨学科文献综述》。$Sustainable\ Cities\ and\ Society$，2008年。

Aude Billard、Sylvain Calinon、R¨udiger Dillmann 和 Stefan Schaal。综述：示教机器人编程。《机器人手册》，2008年。Joshua Bird、Jan Blumenkamp 和 Amanda Prorok。DVM-SLAM：面向多智能体系统的去中心化视觉单目同步定位与地图构建，2025年。Kevin Black、Noah Brown、Danny Driess 等。pi0：用于通用机器人控制的视觉-语言-动作流模型。*arXiv* 预印本 *arXiv:2410.24164*，2024年。Charles Blundell、Benigno Uria、Alexander Pritzel、Yazhe Li、Avraham Ruderman、Joel Z Leibo、Jack Rae、Daan Wierstra 和 Demis Hassabis。无模型情景控制。载于神经信息处理系统进展，2016年。Rishi Bommasani、Drew A Hudson、Ehsan Adeli、Russ Altman、Simran Arber、Sydney von Arx、Michael S Bernstein、Jeannette Bohg、Antoine Bosselut、Emma Brunskill 等。论基础模型的机遇与风险。*arXiv* 预印本 *arXiv:2108.07258*，2021年。Francisco Bonin-Font、Alberto Ortiz 和 Gabriel Oliver。移动机器人的视觉导航：综述。《智能与机器人系统期刊》，2008年。

Florian Bordes、Richard Yuanzhe Pang、Anas Ajber、Christopher Barber、Petar Velickovic、Mahmoud Assran、Nicolas Ballas、Yann LeCun 和 Michael Rabbat。视觉-语言建模简介。*arXiv* 预印本 *arXiv:2405.17247*，2024年。

Sebastian Borgeaud等人。通过从数万亿个词元中检索来改进语言模型。*ICML*，2022年。

Matthew Botvinick、Sam Ritter、Jane X Wang、Zeb Kurth-Nelson、Charles Blundell 和 Demis Hassabis。强化学习：快与慢。《认知科学趋势》，23(5):408–422，2019年。

Konstantinos Bousmalis、Alex Irpan、Paul Wohlhart、Yunfei Bai、Matthew Kelcey、Mrinal Kalakrishnan、Laura Downs、Julian Ibarz、Peter Pastor、Kurt Konolige等人。利用仿真和领域自适应提高深度机器人抓取的效率。*ICRA*，2018年。

Konstantinos Bousmalis、Giulia Vezzani、Dushyant Rao、Coline Devin、Alex X Lee、Maria Bauza、Todor Davchev、Yuxiang Zhou、Agrim Gupta、Akhil Raju 等人。RoboCat：一种用于机器人操作的自我改进型基础智能体。*arXiv* 预印本 *arXiv:2306.11706*，2023年。

Johannes Brandstetter 等人。几何和物理量提升 e(3) 等变消息传递。国际学习表征会议，2022年。Pearl Brereton、Barbara A Kitchenham、David Budgen、Mark Turner 和 Mohamed Khalil。在软件工程领域应用系统文献综述流程的经验教训。《系统与软件杂志》，2007年。Anthony Brohan、Noah Brown、Justice Carbajal 等人。RT-1：面向大规模现实世界控制的机器人变换器。*arXiv* 预印本 *arXiv:2212.06817*，2022年。Anthony Brohan、Noah Brown、Justice Carbajal、Yevgen Chebotar、Xi Chen、Krzysztof Choromanski、Tianli Ding、Danny Driess、Avinava Dubey、Chelsea Finn 等人。RT-2：视觉-语言-动作模型将网络知识迁移至机器人控制。*arXiv* 预印本 *arXiv:2307.15818*，2023年。Michael M Bronstein 等人。几何深度学习。*arXiv* 预印本 *arXiv:2104.13478*，2021年。Rodney A Brooks。没有表征的智能。人工智能，47:139–159，1991。Tim Brooks 等人。作为世界模拟器的视频生成模型。*OpenAI* 技术报告，2024年。

Tom Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared D Kaplan、Prafulla Dhariwal、Arvind Neelakantan、Pranav Shyam、GirishSastry、Amanda Askell 等。《语言模型是少样本学习者》。神经信息处理系统进展，33:1877–1901，2020年。

Cameron B Browne、Edward Powley、Daniel Whitehouse、Simon M Lucas、Peter I Cowling、Philipp Rohlf-shagen、Stephen Tavener、Diego Perez、Spyridon Samothrakis 和 Simon Colton。蒙特卡洛树搜索方法综述。*IEEE*计算智能与游戏人工智能汇刊，2012年。Jake Bruce、Michael Dennis、Ashley Edwards、Jack Parker-Holder、Yuge Shi、Edward Hughes、Matthew Lai、Aditi Mber、Richie Dasagi、Mikael Henaff 等。Genie 2：大规模基础世界模型。*arXiv*预印本 *arXiv:2412.13212*，2024a。Jake Bruce、Michael Dennis、Ashley Edwards 等。Genie：生成式交互环境。载于*ICML*，2024b。Joan Bruna、Wojciech Zaremba、Arthur Szlam 和 Yann LeCun。图上的谱网络与局部连接网络。*ICLR*，2014年。Joy Buolamwini 和 Timnit Gebru。Gender Shades：商业性别分类中的交叉准确性差异。载于公平、问责与透明会议，第77–91页，2018年。Lucian Busoniu、Robert Babuska 和 Bart De Schutter。多智能体强化学习综合综述。*IEEE*系统、人与控制论汇刊*C*辑，2008年。Tom Bylander。命题STRIPS规划的计算复杂度。人工智能，69（1-2）：165–204，1994年。Cesar Cadena 等。同步定位与地图构建的过去、现在与未来：迈向鲁棒感知时代。*IEEE* 机器人学汇刊，2016年。Holger Caesar、Varun Bankiti、Alex H Lang、Sourabh Vora 等。nuScenes：用于自动驾驶的多模态数据集。载于*CVPR*，2020年。Carlos Campos、Richard Elvira、Juan J G´omez Rodr´ıguez、Jos´e MM Montiel 和 Juan D Tard´os。Orb-SLAM3：用于视觉、视觉惯性和多地图SLAM的高精度开源库。*IEEE* 机器人学汇刊，37(6):1874–1890，2021年。Gustau Camps-Valls、Devis Tuia、Lorenzo Bruzzone 和 Jon Atli Benediktsson。高光谱图像分类进展：基于统计学习方法的地球监测。*IEEE*信号处理杂志，31(1):45–54，2014年。CARTO。CARTO空间数据科学平台。https://carto.com/，2023年。CARTO。CARTO空间分析平台。https://carto.com/platform/，2024年。Ziwei Chai 等。GraphLLM：提升大语言模型的图推理能力。*arXiv*预印本 *arXiv:2310.05845*，2023年。Stephanie CY Chan、Adam Santoro、Andrew K Lampinen、Jane X Wang、Aaditya Singh、Pierre H Richemond、James McClelland 和 Felix Hill。数据分布性质驱动Transformer中涌现的上下文学习。*NeurIPS*，2022年。Olivier Chapelle、Bernhard Sch¨olkopf 和 Alexander Zien。半监督学习。麻省理工学院出版社，2009年。Devendra Singh Chaplot、Dhiraj Gandhi、Abhinav Gupta 和 Ruslan Salakhutdinov。利用目标导向语义探索进行目标物体导航。载于*NeurIPS*，2020a。Devendra Singh Chaplot、Dhiraj Gandhi、Abhinav Gupta 和 Ruslan Salakhutdinov。利用目标导向语义探索进行目标物体导航。载于*NeurIPS*，2020b。Devendra Singh Chaplot、Dhiraj Gandhi、Saurabh Gupta、Abhinav Gupta 和 Ruslan Salakhutdinov。利用主动神经SLAM进行探索学习。载于*ICLR*，2020c。Devendra Singh Chaplot、Helen Jiang、Saurabh Gupta 和 Abhinav Gupta。SEAL：利用探索和三维一致性进行自监督具身主动学习。*NeurIPS*，2021年。

Devendra Singh Chaplot等人. 使用主动神经SLAM学习探索. 载于*ICLR*，2020d. Anpei Chen等人. Tensorf：张量辐射场. 欧洲计算机视觉会议，2022a. Austin Chen等人. 使用预训练视觉语言模型进行开放世界物体操作. 机器人学习会议，2023a. Boyuan Chen、Zhuo Xu、Sean Kirmani、Brian Ichter、Danny Driess、Pete Florence、Dorsa Sadigh、Leonidas Guibas和Fei Xia. Spatialvlm：赋予视觉语言模型空间推理能力.*arXiv*预印本 *arXiv:2401.12168*，2024a. Chang Chen等人. Transdreamer：使用Transformer世界模型进行强化学习. *arXiv*预印本 *arXiv:2202.09481*，2022b. Changan Chen、Ruohan Gao、Paul Calamia和Kristen Grauman. 视觉声学匹配. *CVPR*，第18858–18868页，2022c. Changan Chen等人. Soundspaces：3D环境中的视听导航. 载于*ECCV*，2020. Changan Chen等人. Soundspaces 2.0：视觉声学学习仿真平台. 载于*NeurIPS*，2022d. Dexiong Chen等人. 用于图表征学习的结构感知Transformer. 国际机器学习会议，2022e. Guangyan Chen、Meiling Wang、Yi Yang、Kai Yu、Li Yuan和Yufeng Yue. Pointgpt：基于点云的自回归生成式预训练. 载于*NeurIPS*，2024b. Jiabin Chen、Dawei Lin等人. Graphgpt：面向大语言模型的图指令微调. *arXiv*预印本 *arXiv:2310.13023*，2024c. Li Chen、Penghao Wu、Kashyap Chitta、Bernhard Jaeger等人. 端到端自动驾驶：挑战与前沿. *IEEE TPAMI*，2024d. Liang-Chieh Chen、George Papandreou、Iasonas Kokkinos、Kevin Murphy和Alan L Yuille. Deeplab：结合深度卷积网络、空洞卷积和全连接条件随机场的语义图像分割. *IEEE*模式分析与机器智能汇刊，40(4)：834–848，2017年. Mark Chen、Jerry Tworek、Heewoo Jun、Qiming Yuan、Henrique Ponde de Oliveira Pinto、Jared Kaplan、Harri Edwards、Yuri Burda、Nicholas Joseph、Greg Brockman等人. 评估基于代码训练的大语言模型. *arXiv*预印本 *arXiv:2107.03374*，2021a. Matthew Chen、Abhinav Gupta和Saurabh Gupta. 通过观看YouTube视频学习语义视觉导航. 载于*NeurIPS*，2021b. Peihao Chen、Dongyu Ji、Kunyang Lin、Weiwen Zeng、Thomas H Huang、Mingkui Tan和Chuang Gan. 面向视觉与语言导航的弱监督多粒度地图学习. *NeurIPS*，2022f. Runjin Chen等人. Llaga：大语言模型与图助手. *arXiv*预印本 *arXiv:2402.08170*，2024e. Runnan Chen、Youquan Liu、Lingdong Kong、Xinge Zhu、Yuexin Ma、Yikang Li、Yuenan Hou、Yu Qiao和Wenping Wang. Clip2scene：借助CLIP实现标签高效的三维场景理解. *CVPR*，2023b. Shizhe Chen、Pierre-Louis Guhur、Makarand Tapaswi、Cordelia Schmid和Ivan Laptev. 放眼全局，立足局部：用于视觉与语言导航的双尺度图变换器. 2022g. Shizhe Chen等人. 用于视觉与语言导航的历史感知多模态Transformer. 载于*NeurIPS*，2021c. Shizhe Chen等人. 放眼全局，立足局部：用于视觉与语言导航的双尺度图变换器. 载于*CVPR*，2022h.

Sijin Chen等人. Ll3da：用于全方位三维理解、推理和规划的视觉交互式指令微调. *IEEE*计算机视觉与模式识别会议，2024f. Weize Chen、Yusheng Su、Jingwei Zuo、Cheng Yang、Chenfei Yuan、Chen Qian、Chi-Min Chan、Yujia Qin、Yaxi Lu、Ruobing Xie等人. Agentverse：促进多智能体协作并探索涌现行为. *arXiv*预印本 *arXiv:2308.10848*，2024g. Wenhu Chen等人. 思维程序提示：将计算与推理分离，以应对数值推理任务.《机器学习研究汇刊》，2023c. Xi Chen、Xiao Wang、Soravit Changpinyo、AJ Piergiovanni、Piotr Padlewski、Daniel Salz、Sebastian Good- man、Adam Grycner、Basil Mustafa、Lucas Beez等人. PaLi：联合扩展的多语言图文模型. 载于$ICLR$，2023d. Xi Chen等人. PaLi：联合扩展的多语言图文模型. 国际学习表征会议，2023e. Xinyun Chen等人. 教授大语言模型进行自我调试. *arXiv*预印本 *arXiv:2304.05128*，2023f. Yukang Chen、Jianhui Liu、Xiangyu Zhang、Xiaojuan Qi和Jiaya Jia. Voxelnext：用于三维目标检测与跟踪的全稀疏VoxelNet. $CVPR$，2023g. Zhe Chen、Weiyun Wang、Yue Cao等人. Internvl2：超越最佳水平——通过渐进式扩展策略提升开源多模态模型的性能上限. *arXiv*预印本，2024h. An-Chieh Cheng、Hongxu Yin、Yang Fu等人. Spatialrgpt：视觉语言模型中基于落地的空间推理. 2024年. Bowen Cheng、Ishan Misra、Alexander G Schwing、Alexander Kirillov和Rohit Girdhar. 用于通用图像分割的掩码注意力掩码变换器. 载于$CVPR$，2022年. Zhili Cheng、Yuge Tu、Ran Li、Shiqi Dai、Jinyi Hu、Shengding Hu、Jiahao Li、Yang Shi、Tianyu Yu、Weize Chen、Lei Shi和Maosong Sun. Embodiedeval：评估具身智能体形式的多模态大语言模型，2025年. Cheng Chi、Siyuan Feng、Yilun Du等人. 扩散策略：通过动作扩散进行视觉运动策略学习. 2023年. Cheng Chi、Siyuan Feng、Yilun Du、Zhenjia Xu、Eric Cousineau、Benjamin Burchfiel和Shuran Song. 扩散策略：通过动作扩散进行视觉运动策略学习.《国际机器人研究杂志》，2024年. Guanghua Chi等人. 面向所有中低收入国家的财富微观估算.《美国国家科学院院刊》，2022年. Rohan Chitnis、Shubham Tulsiani、Saurabh Gupta和Abhinav Gupta. 利用学习得到的任务模式实现高效双手操作. 载于$ICRA$，2020年. Jeongwhan Choi等人. 用于交通流预测的图神经受控微分方程. *AAAI*人工智能会议，2022年. Fran¸cois Chollet. 关于智能度量. *arXiv*预印本 *arXiv:1911.01547*，2019年. Howie Choset和Keiji Nagatani. 拓扑同步定位与建图. *IEEE*《机器人与自动化汇刊》，17(2)：125–137，2001年. Howie M Choset、Seth Hutchinson、Kevin M Lynch、George Kantor、Wolfram Burgard、Lydia E Kavraki和Sebastian Thrun. 机器人运动原理：理论、算法与实现. 麻省理工学院出版社，2005年.

Aakanksha Chowdhery、Sharan Narang、Jacob Devlin、Maarten Bosma、Gaurav Mishra、Adam Roberts、Paul Barham、Hyung Won Chung、Charles Sutton、Sebastian Gehrmann 等。Palm：借助路径扩展语言建模。*arXiv*预印本 *arXiv:2204.02311*，2022年。Paul F Christiano、Jan Leike、Tom Brown、Miljan Martic、Shane Legg 和 Dario Amodei。基于人类偏好的深度强化学习。*NeurIPS*，30，2017年。Gordon Christie 等。世界功能地图。*CVPR*，2018a。Gordon Christie 等。世界功能地图。*IEEE*计算机视觉与模式识别会议论文集，2018b。Kurtland Chua、Roberto Calandra、Rowan McAllister 和 Sergey Levine。利用概率动力学模型，通过少量试验进行深度强化学习。*NeurIPS*，31，2018年。Taco Cohen 和 Max Welling。群等变卷积网络。载于国际机器学习会议，第2990–2999页，2016年。Anthony G Cohn 和 Jochen Renz。定性空间表示与推理：概述。人工智能基础，3：551–596，2008年。

Open X-Embodiment Collaboration。Open x-embodiment，2023年。

Yezhen Cong、Samar Khanna、Chenlin Meng、Patrick Liu、Erik Rozi、Yutong He、Marshall Burke、David Lobell 和 Stefano Ermon。Satmae：针对时序与多光谱卫星影像的 Transformer 预训练。神经信息处理系统进展，35：197–211，2022年。

Angela Dai、Angel X Chang、Manolis Savva、Maciej Halber、Thomas Funkhouser 和 Matthias Niessner。Scannet：室内场景的丰富标注三维重建。载于*IEEE*计算机视觉与模式识别会议论文集，第5828–5839页，2017年。

Damai Dai、Yutao Sun、Li Dong、Yaru Hao、Shuming Ma、Zhifang Sui 和 Furu Wei。为什么 GPT 能在上下文内学习？语言模型会秘密地将梯度下降作为元优化器。在*ACL*研究成果中，2023a。

Wenliang Dai、Junnan Li、Dongxu Li等。Instructblip：通过指令微调迈向通用视觉语言模型。2023b。

Murtaza Dalal、Deepak Pathak和Ruslan Salakhutdinov。通过参数化动作基元加速机器人强化学习。在神经信息处理系统进展，2021年。

Neil T Dantam、Zachary K Kingston、Swarat Chaudhuri 和 Lydia E Kavraki。增量式任务与运动规划：基于约束的方法。在*RSS*，2016年。Abhishek Das、Samyak Datta、Georgia Gkioxari、Stefan Lee、Devi Parikh 和 Dhruv Batra。用于具身问答的神经模块化控制。在 *CoRL*，第 53–62 页，2018a。Abhishek Das、Georgia Gkioxari、Stefan Lee、Devi Parikh 和 Dhruv Batra。具身问答。在 *IEEE*计算机视觉与模式识别会议论文集，第 1–10 页，2018b。Micha¨el Defferrard、Xavier Bresson 和 Pierre Vandergheynst。基于图的卷积神经网络与快速局部化谱滤波。*NeurIPS*，29，2016年。Mostafa Dehghani 等人。将视觉Transformer扩展至220亿个参数。国际机器学习会议，2023年。Marc Deisenroth 和 Carl E Rasmussen。Pilco：一种基于模型且数据高效的策略搜索方法。在 *ICML*，第 465–472 页，2011年。

Matt Deitke、Winson Han、Alvaro Herrasti、Aniruddha Kembhavi、Eric Kolve、Roozbeh Mottaghi、Jordi Salvador、Dustin Schwenk、Eli VanderBilt、Matthew Wallingford 等。RoboTHOR：一个从仿真到真实的开放式具身人工智能平台。在*CVPR*，2020年。Matt Deitke 等人。Objaverse：一个带标注的三维物体宇宙。*IEEE*计算机视觉与模式识别会议，2023年。Ilke Demir 等人。DeepGlobe 2018：通过卫星图像解析地球的挑战。*IEEE*计算机视觉与模式识别研讨会，2018年。Emily Denton、Alex Hanna、Razvan Amironesei、Andrew Smart、Hilary Nicole 和 Morgan Klaus Scheuer- man。让人重新回到其中：对基准机器学习数据集的质疑。*arXiv*预印本 *arXiv:2007.07399*，2020年。Guilherme N. DeSouza 和 Avinash C. Kak。移动机器人导航的视觉方法：综述。*IEEE*模式分析与机器智能汇刊，24(2)：237–267，2002年。Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。BERT：面向语言理解的深度双向Transformer预训练。*arXiv*预印本 *arXiv:1810.04805*，2019年。Shizhe Diao、Pengcheng Wang、Yong Lin、Hanxiao Tong 和 Tong Zhang。利用思维链为大语言模型进行主动提示。*arXiv*预印本 *arXiv:2302.12246*，2023年。Jian Ding、Nan Xue、Gui-Song Xia 和 Dengxin Dai。解耦零样本语义分割。在*CVPR*，第 11583–11592 页，2022年。Mingyu Ding 等人。理解世界还是预测未来？世界模型的综合综述。*ACM Computing Surveys*，2024年。Thanh-Toan Do、Anh Nguyen 和 Ian Reid。AffordanceNet：一种用于物体可供性检测的端到端深度学习方法。在 *ICRA*，第 5882–5889 页，2018年。Qingxiu Dong、Lei Li、Damai Dai、Ce Zheng、Zhiyong Wu、Baobao Chang、Xu Sun、Jingjing Xu 和 Zhifang Sui。上下文学习综述。*arXiv*预印本 *arXiv:2301.00234*，2022年。Vishnu Sashank Dorbala、James F Mullen Jr 和 Dinesh Manocha。CLIP-Nav：使用 CLIP 进行零样本视觉与语言导航。在*CoRL*，2022年。Matthijs Douze、Alexandr Guzhva、Chengqi Deng、Jeff Johnson、Gergely Szilvasy、Pierre-Emmanuel Mazar´e、Maria Lomeli、Lucas Hosseini 和 Herv´e J´egou。FAISS 库。*arXiv*预印本 *arXiv:2401.08281*，2024年。Danny Driess、Jung-Su Ha 和 Marc Toussaint。深度视觉推理：从初始场景图像中学习预测用于任务与运动规划的动作序列。在 *RSS*，2020年。Danny Driess、Fei Xia、Mehdi SM Sajjadi、Corey Lynch、Aakanksha Chowdhery、Brian Ichter、Ayzaan Wahid、Jonathan Tompson、Quan Vuong、Tianhe Yu 等。PaLM-E：具身多模态语言模型。*arXiv*预印本 *arXiv:2303.03378*，2023年。Jiafei Duan、Samson Yu、Hui Li Tan、Hongyuan Zhu 和 Cheston Tan。具身人工智能综述：从模拟器到研究任务，2022年。URL https://arxiv.org/abs/2103.04918。Abhimanyu Dubey、Abhinav Jauhri、Abhinav Pandey 等。Llama 3 模型群。*arXiv*预印本 *arXiv:2407.21783*，2024年。John J Dudley 和 Per Ola Kristensson。交互式机器学习用户界面设计综述。*ACM Transactions on Interactive Intelligent Systems*，8(2)：1–37，2018年。Zane Durante、Qiuyuan Sarber、Jianlong Gong 等。智能体人工智能：多模态交互前景综述。*arXiv*预印本 *arXiv:2401.03568*，2024年。

Hugh Durrant-Whyte 和 Tim Bailey。同时定位与建图：第一部分。*IEEE*《机器人与自动化杂志》，13(2):99–110，2006年。Vijay Prakash Dwivedi 等人。图神经网络基准测试。《机器学习研究杂志》，2023年。Tore Dyb˚a 和 Torgeir Dingsøyr。将系统综述应用于不同类型的研究：经验报告。*ESEM*，2007年。Max J Egenhofer 和 Robert D Franzosa。点集拓扑空间关系。《国际地理信息系统杂志》，1991。Kiana Ehsani、Winson Han、Alvaro Herrasti、Eli VanderBilt、Luca Weihs、Eric Kolve、Aniruddha Kembhavi 和 Roozbeh Mottaghi。Manipulathor：视觉物体操作框架。*CVPR*，2021年。David Eigen、Christian Puhrsch 和 Rob Fergus。使用多尺度深度网络从单张图像预测深度图。载于神经信息处理系统进展，2014年。Jakob Engel、Vladlen Koltun 和 Daniel Cremers。直接稀疏里程计。*IEEE*模式分析与机器智能汇刊，40(3):611–625，2017年。Jakob Engel 等人。LSD-SLAM：大规模直接单目同步定位与建图。欧洲计算机视觉会议，2014年。ESRI。Esri ArcGIS：制图与分析平台。https://www.esri.com，2023年。ESRI。ArcGIS Pro：专业地理信息系统软件。https://www.esri.com/en-us/arcgis/products/ arcgis-pro/overview，2024a。ESRI。ArcGIS 中的地理空间人工智能：用于地理空间分析的人工智能。技术报告，环境系统研究所，2024b。Jerry Alan Fails 和 Dan R Olsen Jr。交互式机器学习。*IUI*，第39–45页，2003年。Lue Fan、Ziqi Pang、Tianyuan Zhang、Yu-Xiong Wang、Hang Zhao、Feng Wang、Naiyan Wang 和 Zhaoxiang Zhang。以稀疏 Transformer 为单步幅三维物体检测器赋能。*CVPR*，2022年。Zhiwen Fan、Kevin Wang、Kairun Wen、Zehao Zhu、Dejia Xu 和 Zhangyang Wang。LightGaussian：无界三维高斯压缩，压缩率达15倍且 200+ 帧每秒。载于*arXiv*预印本 *arXiv:2311.17245*，2024年。Hao-Shu Fang 等人。GraspNet-1Billion：用于通用物体抓取的大规模基准测试。*IEEE/CVF*计算机视觉与模式识别会议，2020年。Kuan Fang、Te-Lin Wu、Daniel Yang、Silvio Savarese 和 Joseph J Lim。Demo2Vec：基于在线视频推理物体可供性。*CVPR*，第2139–2147页，2018年。Kuan Fang、Alexander Toshev、李飞飞和 Silvio Savarese。面向长时程任务中具身智能体的场景记忆变换器。载于*IEEE/CVF*计算机视觉与模式识别会议，第538–547页，2019年。Bahare Fatemi 等人。像图一样说话：为大语言模型编码图。*arXiv*预印本 *arXiv:2310.04560*，2023 年。Bahare Fatemi 等人。像图一样说话：为大语言模型编码图。*arXiv* 预印本 *arXiv:2310.04560*，2024年。Olivier Faugeras。三维计算机视觉：几何视角。麻省理工学院出版社，1993年。Tuo Feng、Yixiao Wang、Jiaxin Chen 等人。自动驾驶世界模型综述。*arXiv*预印本 *arXiv:2501.11260*，2025年。

Richard E Fikes 和 Nils J Nilsson。Strips：将定理证明应用于问题求解的新方法。人工智能，2(3-4):189–208，1971年。Pete Florence、Corey Lynch、Andy Zeng、Oscar A Ramirez、Ayzaan Wahid、Laura Downs、Adrian Wong、Johnny Lee、Igor Mordatch 和 Jonathan Tompson。隐式行为克隆。载于*CoRL*，2022年。Jakob Foerster、Ioannis Alexandros Assael、Nando de Freitas 和 Shimon Whiteson。通过深度多智能体强化学习学习沟通。神经信息处理系统进展，29，2016年。David A Forsyth 和 Jean Ponce。计算机视觉：现代方法。Prentice Hall，2011年。

Meire Fortunato、Melissa Tan、Ryan Faulkner、Steven Hansen、Adria Puigdomenech Badia、Gavin Buttimore、Charlie Deck、Joel Z Leibo 和 Charles Blundell。《具有工作记忆和情景记忆的强化学习者的泛化》。载于神经信息处理系统进展，2019年。

Foursquare。Foursquare位置智能。https://foursquare.com，2023年。

Foursquare。Foursquare地点：全球最受信赖的位置数据。技术报告，Foursquare Labs Inc.，2024年。

Jonathan Francis、Nariaki Kitamura、Felix Labber、Luca Navarro 和 Jean Oh。具身视觉语言规划的核心挑战。载于*CoRL*，2022年。

Jonathan Frankle 和 Michael Carbin。彩票假说：寻找稀疏、可训练的神经网络。*ICLR*，2019年。

Sara Fridovich-Keil 等。Plenoxels：无需神经网络的辐射场。*IEEE*计算机视觉与模式识别会议，2022年。

Daniel Fried、Ronghang Hu、Volkan Cirik、Anna Rohrbach、Jacob Andreas、Louis-Philippe Morency、Taylor Berg-Kirkpatrick、Kate Saenko、Dan Klein 和 Trevor Darrell。用于视觉与语言导航的说话者-跟随者模型。载于神经信息处理系统进展，第 3331–3342 页，2018年。

Huan Fu 等。3D基础模型：综述。*arXiv* 预印本，2024a。

Xiao Fu、Wei Yin、Mu Hu、Kaixuan Wang、Yuexin Ma、Ping Tan、Shaojie Shen、Dahua Lin 和 Xiaoxiao Long。GeoWizard：释放扩散先验，基于单张图像进行3D几何估计。*arXiv* 预印本 *arXiv:2403.12013*，2024b。

Yao Fu 等。用于多步推理的基于复杂度的提示。*ICLR*，2023年。

Fabian Fuchs、Daniel Worrall、Volker Fischer 和 Max Welling。《SE(3)-变换器：3D 旋转平移等变注意力网络》。载于神经信息处理系统进展，2020年。

Jorge Fuentes-Pacheco、Jos´e Ruiz-Ascencio 和 Juan Manuel Rend´on-Mancha。视觉同时定位与建图：综述。*Artificial Intelligence Review*，43(1):55–81，2015年。

Scott Fujimoto 等人。解决行动者-评论家方法中的函数逼近误差。国际机器学习会议，2018年。Anthony Fuller 等人。CROMA：采用对比式雷达-光学掩码自编码器的遥感表征。*arXiv*预印本 *arXiv:2311.00566*，2024年。Samir Yitzhak Gadre、Mitchell Wortsman、Gabriel Ilharco、Ludwig Schmidt 和 Shuran Song。Clip on wheels：将零样本物体导航视为物体定位与探索。*arXiv*预印本 *arXiv:2203.10421*，2022年。Samir Yitzhak Gadre、Mitchell Wortsman、Gabriel Ilharco、Ludwig Schmidt 和 Shuran Song。Cows on pasture：语言驱动的零样本物体导航的基线与基准测试。*arXiv*预印本 *arXiv:2203.10421*，2023年。

Erich Gamma、Richard Helm、Ralph Johnson 和 John Vlissides。设计模式：可复用面向对象软件的基础。Addison-Wesley，1994年。Chuang Gan 等人。观察、聆听并行动：迈向视听具身导航。发表于*ICRA*，2020年。Luyu Gao 等人。PAL：程序辅助语言模型。国际机器学习会议，2023年。Caelan Reed Garrett、Tom´as Lozano-P´erez 和 Leslie Pack Kaelbling。PDDLStream：通过乐观自适应规划整合符号规划器与黑盒采样器。发表于*ICAPS*，第440–448页，2020年。Caelan Reed Garrett 等人。集成任务与运动规划。《控制年度评论》，2021年。Tiezheng Ge、Kaiming He、Qifa Ke 和 Jian Sun。优化乘积量化。*IEEE*模式分析与机器智能汇刊，36(4):744–755，2014年。Timnit Gebru、Jamie Morgenstern、Briana Vecchione、Jennifer Wortman Vaughan、Hanna Wallach、Hal Daum´e III 和 Kate Crawford。数据集数据表。《*ACM*通讯》，64(12): 86–92，2021年。Hector Geffner 和 Blai Bonet。自动规划的模型与方法简明导论。Morgan and Claypool出版社，2013年。Andreas Geiger、Philip Lenz 和 Raquel Urtasun。我们准备好迎接自动驾驶了吗？KITTI视觉基准套件。发表于*CVPR*，2012年。Xu Geng 等人。用于网约车需求预测的时空多图卷积网络。*AAAI*人工智能会议，2019年。Samuel J Gershman 和 Nathaniel D Daw。人类和动物的强化学习与情景记忆：一个综合框架。《心理学年度评论》，68:101–128，2017年。Theophile Gervet、Soumith Chintala、Dhruv Batra、Jitendra Malik 和 Devendra Singh Chaplot。在真实世界中导航至物体。《科学·机器人学》，2023年。Malik Ghallab、Dana Nau 和 Paolo Traverso。自动规划：理论与实践。爱思唯尔，2004年。Golnaz Ghiasi、Xiuye Gu、Yin Cui 和 Tsung-Yi Lin。利用图像级标签扩展开放词汇图像分割。发表于*ECCV*，2022年。Amir Gholami、Sehoon Kim、Zhen Dong、Zhewei Yao、Michael W Mahoney 和 Kurt Keutzer。高效神经网络推理量化方法综述。《低功耗计算机视觉》，第291–326页，2022年。James J Gibson。视觉感知的生态学方法。Houghton Mifflin，1979年。Marco Gillies、Rebecca Fiebrink、Atau Tanaka、J´er´emie Garcia、Fr´ed´eric Bevilacqua、Alexis Heloir、Fabrizio Nunnari、Wendy Mackay、Saleema Amershi、Bongshin Lee 等人。以人为中心的机器学习。*CHI*扩展摘要集，第3558–3565页，2016年。Justin Gilmer、Samuel S Schoenholz、Patrick F Riley、Oriol Vinyals 和 George E Dahl。用于量子化学的神经消息传递。发表于*ICML*，2017年。Rohit Girdhar、Alaaeldin El-Nouby、Zhuang Liu、Mannat Singh、Kalyan Vasudev Alwala、Armand Joulin 和 Ishan Misra。ImageBind：用一个嵌入空间连接所有模态。发表于*CVPR*，第15180–15190页，2023年。Cl´ement Godard、Oisin Mac Aodha、Michael Firman 和 Gabriel J Brostow。深入探索自监督单目深度估计。*ICCV*，2019年。Ran Gong、Jiangyong Huang、Yizhou Zhao、Haoran Geng、Xiaofeng Gao、Qingyang Wu、Wensi Ai、Ziheng Zhou、Demetri Terzopoulos、Song-Chun Zhu 等人。ARNOLD：一个用于逼真3D场景中连续状态语言基础任务学习的基准测试。*arXiv*预印本 *arXiv:2304.04321*，2023年。

Michael F Goodchild。《作为传感器的公民：志愿地理的世界》。《GeoJournal》，2007年。Ian Goodfellow、Yoshua Bengio 和 Aaron Courville。《深度学习》。麻省理工学院出版社，2016年。谷歌。Google 地图平台。https://cloud.google.com/maps-platform，2023年。谷歌。Google Earth Engine：行星尺度地理空间分析。https://earthengine.google.com/，2024a。谷歌。Google 地图中的人工智能：为新一代导航提供动力。技术报告，Google LLC，2024b。Daniel Gordon、Aniruddha Kembhavi、Mohammad Rastegari、Joseph Redmon、Dieter Fox 和 Ali Farhadi。IQA：交互式环境中的视觉问答。载于*IEEE/CVF*计算机视觉与模式识别会议，第4089–4098页，2018年。Daniel Gordon、Abhishek Kadian、Devi Parikh、Judy Hoffman 和 Dhruv Batra。Splitnet：面向具身视觉导航的仿真到仿真及任务到任务迁移。载于*ICCV*，第1022–1031页，2019年。Hrithik P Gowda、SN Sreevathsa、Gangadhara KN Gowda 和 SJ Sharath。从图到蓝图：基于 GNN 的平面图建模。《International Advanced Research Journal in Science, Engineering and Technology》，12(2)，2025年。Aaron Grattafiori、Abhimanyu Dubey、Abhinav Jauhri 等人。Llama 3.2 系列：多模态开放基础模型。*arXiv* 预印本，2024年。Alex Graves、Greg Wayne 和 Ivo Danihelka。神经图灵机。*arXiv* 预印本 *arXiv:1410.5401*，2014年。Alex Graves、Greg Wayne、Malcolm Reynolds、Tim Harley、Ivo Danihelka、Agnieszka Grabska-Barwi´nska、Sergio G´omez Colmenarejo、Edward Grefenstette、Tiago Ramalho、John Agapiou 等人。利用带有动态外部记忆的神经网络进行混合计算。《自然》，538(7626)：471–476，2016年。Ben Green 和 Yiling Chen。算法在环决策的原则与局限。*CSCW*，3：1–24，2019年。Jiayuan Gu 等人。ManiSkill2：可泛化操作技能的统一基准测试。载于*ICLR*，2023年。Qiao Gu 等人。ConceptGraphs：用于感知与规划的开放词汇三维场景图。*IEEE*国际机器人与自动化会议，2024年。Pierre-Louis Guhur 等人。Airbert：面向视觉与语言导航的领域内预训练。载于*ICCV*，2021年。Daya Guo、Qihao Zhu、Dejian Yang、Zhenda Xie、Kai Dong、Wentao Zhang、Guanting Chen、Xiao Bi、Y Wu、YK Li 等人。DeepSeek-Coder：大语言模型与编程相遇。载于*arXiv* 预印本 *arXiv:2401.14196*，2024a。Kan Guo、Yongli Hu、Zhen Qian、Yanfeng Sun、Junbin Gao 和 Baocai Yin。用于交通流预测的层次图卷积网络。*AAAI*，35(1)：151–159，2021年。Ruiqi Guo、Philip Sun、Erik Lindgren、Quan Geng、David Simcha、Felix Chern 和 Sanjiv Kumar。利用各向异性向量量化加速大规模推理。*ICML*，第3887–3896页，2020年。Shengnan Guo、Youfang Lin、Ning Feng、Chao Song 和 Huaiyu Wan。用于交通流量预测的基于注意力的时空图卷积网络。2019年。Taicheng Guo、Xiuying Chen、Yaqi Wang、Ruidi Chang、Shichao Pei 等人。基于大语言模型的多智能体：进展与挑战综述。*arXiv* 预印本 *arXiv:2402.01680*，2024b。Xudong Guo 等人。具身大语言模型智能体学会在有组织团队中协作。*arXiv* 预印本 *arXiv:2403.12482*，2024c。

Ziyu Guo、Renrui Zhang、Xiangyang Zhu、Yiwen Tang、Xianzheng Ma、韩佳明、陈可心、高鹏、李先知、Hongsheng Li 和邢本安。Point-bind 与 point-llm：将点云与多模态对齐，用于三维理解、生成和指令遵循。*arXiv* 预印本 *arXiv:2309.00615*，2023年。Abhishek Gupta、Vikash Kumar、Corey Lynch、Sergey Levine 和 Karol Hausman。接力策略学习：通过模仿学习和强化学习解决长时程任务。载于*CoRL*，2019年。Agrim Gupta、Silvio Savarese、Surya Ganguli 和李飞飞。通过学习与进化实现具身智能。《自然·通讯》，12(1)：1–14，2021年。Kelvin Guu、Kenton Lee、Zora Tung、Panupong Pasupat 和 Mingwei Chang。检索增强语言模型预训练。载于*ICML*，2020年。David Ha 和 Jürgen Schmidhuber。世界模型。*arXiv* 预印本 *arXiv:1803.10122*，2018年。Tuomas Haarnoja、Aurick Zhou、Pieter Abbeel 和 Sergey Levine。软演员-评论家：采用随机演员的离策略最大熵深度强化学习。*ICML*，2018年。Danijar Hafner、Timothy Lillicrap、Ian Fischer、Ruben Villegas、David Ha、Honglak Lee 和 James David-son。学习潜在动力学，实现基于像素的规划。载于*ICML*，2019a。Danijar Hafner、Timothy Lillicrap、Ian Fischer、Ruben Villegas、David Ha、Honglak Lee 和 James David-son。学习潜在动力学，实现基于像素的规划。*ICML*，第2555–2565页，2019b。Danijar Hafner、Timothy Lillicrap、Jimmy Ba 和 Mohammad Norouzi。梦境控制：通过潜在想象学习行为。2020年。Danijar Hafner、Timothy Lillicrap、Mohammad Norouzi 和 Jimmy Ba。利用离散世界模型掌握 Atari 游戏。载于*ICLR*，2021年。Danijar Hafner、Jurgis Pasukonis、Jimmy Ba 和 Timothy Lillicrap。通过世界模型掌握多样化领域。*arXiv* 预印本 *arXiv:2301.04104*，2023年。Torkel Hafting、Marianne Fyhn、Sturla Molden、May-Britt Moser 和 Edvard I Moser。内嗅皮层中空间地图的微观结构。《自然》，436(7052)：801–806，2005。Will Hamilton、Zhitao Ying 和 Jure Leskovec。在大型图上进行归纳式表征学习。*NeurIPS*，2017年。William L Hamilton。《图表征学习》。Morgan & Claypool，2020年。韩松、毛慧子和 William J Dally。深度压缩：通过剪枝、训练后量化和霍夫曼编码压缩深度神经网络。*ICLR*，2016年。Nicklas Hansen、Hao Su 和 Xiaolong Wang。TD-MPC2：用于连续控制的可扩展鲁棒世界模型。*ICLR*，2024年。Nicklas Hansen 等人。用于模型预测控制的时序差分学习。*ICML*，2022年。郝世博、Yi Gu、马浩迪、Joshua Jiahua Hong、王震、Daisy Zhe Wang 和胡志挺。基于语言模型的推理，就是基于世界模型的规划。*arXiv* 预印本 *arXiv:2305.14992*，2023年。郝世博等人。ToolkenGPT：通过工具嵌入利用海量工具增强冻结语言模型。*NeurIPS*，2024年。Richard Hartley 和 Andrew Zisserman。《计算机视觉中的多视图几何》。剑桥大学出版社，2003年。Kaiming He、Xiangyu Zhang、任少卿和 Jian Sun。用于图像识别的深度残差学习。*CVPR*，2016年。

Malte Helmert。《快速向下规划系统》。《人工智能研究杂志》，26:191–246，2006年。Dan Hendrycks、Nicholas Carlini、John Schulman 和 Jacob Steinhardt。人工智能研究的 X 风险分析。*arXiv* 预印本 *arXiv:2206.05862*，2022年。Jo˜ao F Henriques 和 Vedaldi。Mapnet：一种用于环境制图的自我中心空间记忆。载于*CVPR*，2018年。Pablo Hernandez-Leal、Bilal Kartal 和 Matthew E Taylor。多智能体强化学习综述。自主智能体与多智能体系统，33(6):750–797，2019年。Geoffrey Hinton、Oriol Vinyals 和 Jeff Dean。蒸馏神经网络中的知识。*arXiv* 预印本 *arXiv:1503.02531*，2015年。Daniel Ho、Kanishka Rao、Zhuo Xu、Eric Jang、Mohi Khansari 和 Yunfei Bai。Retinagan：一种面向对象的仿真到现实迁移方法。*ICRA*，第10920–10926页，2021年。Jordan Hoffmann、Sebastian Borgeaud、Arthur Mensch、Elena Buchatskaya、Trevor Cai、Eliza Rutherford、Diego de Las Casas、Lisa Anne Hendricks、Johannes Welbl、Aidan Clark 等。训练计算量最优的大语言模型。*NeurIPS*，35:30016–30030，2022年。Andreas Holzinger。面向健康信息学的交互式机器学习：我们何时需要人在回路中？脑信息学，3(2):119–131，2016年。Sirui Hong、Mingchen Zhuge、Jonathan Chen、Xiawu Zheng、Yuheng Cheng、Ceyao Zhang、Jinlin Wang、Zili Wang、Steven Ka Shing Yau、Zijuan Lin 等。MetaGPT：面向多智能体协作框架的元编程。*arXiv* 预印本 *arXiv:2308.00352*，2023a。Yicong Hong、Qi Wu、Yuankai Qi、Cristian Rodriguez-Opazo 和 Stephen Gould。VLN-BERT：用于导航的循环视觉语言 BERT。*arXiv* 预印本 *arXiv:2011.13922*，2020年。Yicong Hong、Qi Wu、Yuankai Qi、Cristian Rodriguez-Opazo 和 Stephen Gould。VLN-BERT：用于导航的循环视觉语言 BERT。*CVPR*，2021年。Yining Hong 等人。3D-LLM：将三维世界注入大语言模型。神经信息处理系统进展，2023b。Eric Horvitz。混合主动式用户界面的原则。*CHI*，1999。Mohammad Hossin 和 Md Nasir Sulaiman。数据分类评估中的评估指标综述。国际数据挖掘与知识管理流程杂志，5(2):1，2015年。Andrew G Howard 等人。MobileNets：面向移动视觉应用的高效卷积神经网络。载于*arXiv* 预印本 *arXiv:1704.04861*，2017年。Anthony Hu 等人。GAIA-1：用于自动驾驶的生成式世界模型。*arXiv* 预印本 *arXiv:2309.17080*，2023a。Yihan Hu、Jiazhi Yang、Li Chen、Keyu Li、Chonghao Sima 等。面向规划的自动驾驶。载于*CVPR*，2023b。Binbin Huang、Zehao Yu、Anpei Chen、Andreas Geiger 和 Shenghua Gao。用于几何精确辐射场的二维高斯泼溅。*SIGGRAPH*，2024a。Chenguang Huang、Oier Mees、Andy Zeng 和 Wolfram Burgard。用于机器人导航的视觉语言地图。*arXiv* 预印本 *arXiv:2210.05714*，2023a。Gao Huang、Zhuang Liu、Laurens Van Der Maaten 和 Kilian Q Weinberger。密集连接卷积网络。载于*CVPR*，第4700–4708页，2017年。

Jiangyong Huang 等人。三维世界中的具身通用智能体。国际机器学习会议，2024b。Jin Huang 等人。大语言模型能否有效利用图结构信息：何时以及为何？*arXiv* 预印本 *arXiv:2309.16595*，2024c。Wenlong Huang、李飞飞、Ted Xiao、Harris Chan、Jacky Liang、Pete Florence、Andy Zeng、Jonathan Tomp- son、Igor Mordatch、Yevgen Chebotar 等。内心独白：通过语言模型规划进行具身推理。*arXiv* 预印本 *arXiv:2207.05608*，2022a。Wenlong Huang、Chen Wang、Ruohan Zhang、Yunzhu Li、Jiajun Wu 和李飞飞。VoxPoser：利用语言模型进行机器人操作的可组合三维价值地图。*arXiv* 预印本 *arXiv:2307.05973*，2023b。Wenlong Huang 等人。作为零样本规划器的语言模型：为具身智能体提取可执行知识。国际机器学习会议，2022b。Xu Huang、Weiwen Liu、Xiaolong Chen、Xingmei Wang、Hao Wang、Defu Lian、Yasheng Wang、Ruiming Tang 和 Enhong Chen。理解大语言模型智能体的规划：综述。*arXiv* 预印本 *arXiv:2402.02716*，2024d。Yushi Huang 等人。视觉指令微调。*arXiv* 预印本，2023c。

Nathan Hughes、Yun Chang 和 Luca Carlone。Hydra：一种用于三维场景图构建与优化的实时空间感知系统。*arXiv* 预印本 *arXiv:2201.13360*，2022年。

Michael Hulbert、Matthias Feurer 和 Frank Hutter。使用大语言模型进行超参数优化。*arXiv* 预印本 *arXiv:2312.04528*，2024年。Julian Ibarz、Jie Tan、Chelsea Finn、Mrinal Kalakrishnan、Peter Pastor 和 Sergey Levine。如何通过深度强化学习训练机器人：我们学到的经验。《国际机器人研究杂志》，40(4-5):698–721，2021年。Gautier Izacard、Patrick Lewis、Maria Lomeli、Lucas Hosseini、Fabio Petroni、Timo Schick、Jane Dwivedi-Yu、Armand Joulin、Sebastian Riedel 和 Edouard Grave。Atlas：使用检索增强语言模型进行少样本学习。*JMLR*，24(251):1–43，2023年。Benoit Jacob、Skirmantas Kligys、Bo Chen、Menglong Zhu、Matthew Tang、Andrew Howard、Hartwig Ada m 和 Dmitry Kalenichenko。用于高效整数算术推理的神经网络量化与训练。*CVPR*，第2704–2713页，2018年。Jitesh Jain、Jiachen Li、MangTik Chiu、Ali Hassani、Nikita Orber 和 Humphrey Shi。Oneformer：一个统御通用图像分割的 Transformer。在 *CVPR*，第2989–2998页，2023年。J ohannes Jakubik、Sujit Roy、C E Phillips、Paolo Fraccaro、Denys Godwin、Bianca Zadrozny、Daniela Szwar-cman、Carlos Gomes、Gabby Musber、Daiki Oliveira 等。Prithvi：用于地球观测的基础模型。*arXiv* 预印本 *arXiv:2310.18660*，2024年。Samireh Jalali 和 Claes Wohlin。系统性文献研究：数据库检索与反向滚雪球抽样。*ESEM*，第29–38页，2012年。Stephen James、Paul Wohlhart、Mrinal Kalber、Andrew J Davison 和 Sergey Levine。通过仿真到仿真实现仿真到现实：基于随机化仿真的数据高效机器人学习。在 *2019*年国际机器人与自动化会议（*ICRA*），第2262–2269页，2019年。Stephen James、Zicong Ma、David Rovick Arrojo 和 Andrew J Davison。RLBench：机器人学习基准。*IEEE*机器人与自动化快报，2020年。Michael Janner 等人。何时应信任你的模型：基于模型的策略优化。在*NeurIPS*，2019年。Krzysztof Janowicz 等人。GeoAI：用于地理知识发现及其他领域的空间显式人工智能技术。国际地理信息科学杂志，2020年。

Herve Jegou 等人。用于最近邻搜索的乘积量化。*IEEE TPAMI*，2011年。Albert Q Jiang、Alexandre Sablayrolles、Arthur Mensch 等。Mistral 7B。*arXiv* 预印本 *arXiv:2310.06825*，2023年。Weiwei Jiang 和 Jiayun Luo。用于交通流预测的图神经网络：综述。《*Expert Systems with Applications*》，2022年。Yuhao Jiang 等。Deepurbandownscale：一种物理信息深度学习框架，用于根据粗尺度卫星观测估算高分辨率城市地表温度。《*ISPRS* 摄影测量与遥感杂志》，178:1–15，2021年。Carlos E Jimenez、John Yang、Alexander Wettig、Shunyu Yao、Kexin Pei、Ofir Press 和 Karthik Narasimhan。SWE-Bench：语言模型能否解决现实世界中的 GitHub 问题？*arXiv* 预印本 *arXiv:2310.06770*，2024年。Guangyin Jin、Yuxuan Liang、Yuchen Fang、Zezhi Huang、Junbo Zhang 和 Yu Zheng。用于城市计算的时空图神经网络：综述。《*IEEE*知识与数据工程汇刊》，2023年。Jeff Johnson、Matthijs Douze 和 Herv´e J´egou。十亿级相似性搜索与图形处理器。《*IEEE*大数据汇刊》，2019年。Justin Johnson 等人。使用场景图进行图像检索。*IEEE*计算机视觉与模式识别会议，2015年。Ryan Julian、Benjamin Swanson、Gaurav S Sukhatme、Sergey Levine、Chelsea Finn 和 Karol Hausman。通过奖励草图和批量强化学习扩展数据驱动的机器人学。*RSS*，2020年。Leslie Pack Kaelbling 和 Tom´as Lozano-P´erez。当前的分层任务与运动规划。*ICRA*，2011年。Dmitry Kalashnikov、Alex Irpan、Peter Pastor、Julian Ibarz、Alexander Herzog、Eric Jang、Deirdre Quillen、Ethan Holly、Mrinal Kalakrishnan、Vincent Vanhoucke 和 Sergey Levine。QT-Opt：用于基于视觉的机器人操作的可扩展深度强化学习。*CoRL*，2018年。Dmitry Kalashnikov、Jacob Varley、Yevgen Chebotar、Benjamin Swanson、Rico Jonschkowski、Chelsea Finn、Sergey Levine 和 Karol Hausman。MT-Opt：大规模连续多任务机器人强化学习。*CoRL*，2021年。Subbarao Kambhampati、Karthik Valmeekam、Lin Guan、Mudit Verma、Kaya Stechly、Siddhant Bhambri、Lucas Saldyt 和 Anil Murthy。大语言模型无法进行规划，但可以在 LLM-modulo框架中辅助规划。*arXiv* 预印本 *arXiv:2402.01817*，2024年。Jared Kaplan、Sam McCandlish、Tom Henighan、Tom B Brown、Benjamin Chess、Rewon Child、Scott Gray、Alec Radford、Jeffrey Wu 和 Dario Amodei。神经语言模型的缩放定律。*arXiv* 预印本 *arXiv:2001.08361*，2020年。Henry Kautz 和 Bart Selman。可满足性规划。欧洲人工智能会议，92:359–363，1992年。Kento Kawaharazuka 等人。用于机器人学的视觉-语言-动作模型：面向现实世界应用的综述。*IEEE* 机器人学汇刊，2025年。Staffs Keele 等人。软件工程系统性文献综述实施指南。技术报告，*EBSE*，2007年。Nikhil Keetha 等人。SplaTAM：对三维高斯进行泼溅、跟踪与建图，实现稠密 RGB-D 同步定位与地图构建。*arXiv* 预印本 *arXiv:2312.02126*，2024年。Bernhard Kerbl、Georgios Kopanas、Thomas Leimkuhler 和 George Drettakis。用于实时辐射场渲染的三维高斯泼溅。*ACM* 图形学汇刊，42(4):1–14，2023年。

Apoorv Khandelwal、Luca Weihs、Roozbeh Mottaghi和Aniruddha Kembhavi。简单但有效：面向具身人工智能的CLIP嵌入。载于*CVPR*，2022年。Urvashi Khandelwal等人。通过记忆实现泛化：最近邻语言模型。载于*ICLR*，2020年。Tushar Khot、Harsh Trivedi、Matthew Finlayson、Yao Fu、Kyle Richardson、Peter Clark和Ashish Sabharwal。分解式提示：一种用于解决复杂任务的模块化方法。*ICLR*，2023年。Geunwoo Kim、Pierre Baldi和Stephen McAleer。语言模型可以解决计算机任务。神经信息处理系统进展，第36卷，2023年。Moo Jin Kim、Karl Pertsch、Siddharth Karamcheti、Ted Xiao、Ashwin Balakrishna、Suraj Nair、Rafael Rafailov、Ethan Foster、Grace Lam、Pannag Sanketi等人。OpenVLA：一个开源的视觉-语言-动作模型。*arXiv*预印本 *arXiv:2406.09246*，2024年。Ue-Hwan Kim、Jin-Man Park、Taek-Jin Song和Jong-Hwan Kim。三维场景图：面向智能体的物理环境稀疏语义表征。载于*IEEE*控制论汇刊，第50卷，第4921–4933页，2019年。Thomas N Kipf和Max Welling。基于图卷积网络的半监督分类。*ICLR*，2017年。Alexander Kirillov、Eric Mintun、Nikhila Ravi、Hanzi Mao、Chloe Rolland、Laura Gustafson、Tete Xiao、Spencer Whitehead、Alexander C Berg、Wan-Yen Lo等人。Segment Anything。*arXiv*预印本 *arXiv:2304.02643*，2023年。Barbara Kitchenham。执行系统综述的程序。基尔大学技术报告，2004年。Rob Kitchin。实时城市？大数据与智慧城市主义，第79卷。2014年。Jens Kober、J Andrew Bagnell和Jan Peters。机器人学中的强化学习：综述。国际机器人学研究期刊，2013年。Levente Kocsis和Csaba Szepesv´ari。基于赌博机的蒙特卡洛规划。载于*ECML*，2006。Takeshi Kojima、Shixiang Shane Gu、Machel Reid、Yutaka Matsuo和Yusuke Iwasawa。大语言模型是零样本推理器。神经信息处理系统进展，35：22199–22213，2022年。Eric Kolve、Roozbeh Mottaghi、Winson Han、Eli VanderBilt、Luca Weihs、Alvaro Herrasti、Daniel Gordon、Yuke Zhu、Abhinav Gupta和Ali Farhadi。AI2-THOR：面向视觉人工智能的交互式三维环境。*arXiv*预印本 *arXiv:1712.05474*，2017年。Risi Kondor、Zhen Lin和Shubhendu Trivedi。Clebsch–Gordan网络：一种完全傅里叶空间球面卷积神经网络。载于*NeurIPS*，第10117–10126页，2018年。Kurt Konolige、Motilal Agrawal、Robert C Bolles、Cregg Cowan、Martin Fischler和Brian Gerkey。使用双目视觉进行室外建图与导航。载于实验机器人学，第179–190页。Springer，2008年。Philip Koopman和Michael Wagner。自动驾驶车辆公共道路测试的安全论证考量。*SAE*国际机动性进展与现行实践期刊，2019年。Jacob Krantz等人。超越导航图：连续环境中的视觉与语言导航。载于*ECCV*，2020年。Devin Kreuzer 等人。基于谱注意力重新思考图 Transformer。神经信息处理系统进展，2021年。

Ranjay Krishna、Yuke Zhu、Oliver Groth、Justin Johnson、Kenji Hata、Joshua Kravitz、Stepha nie Chen, Yannis Kalantidis、Li-Jia Li、David A Shamma 等。视觉基因组：连接语言与 视觉 利用众包密集图像标注。在国际计算机视觉杂志，第 123 卷，第 32–73 页，2017年。Alex Krizhevsky、Ilya Sutskever 和 Geoffrey E Hinton。使用深度卷积神经网络进行 ImageNet 分类 vol性

神经网络。*NeurIPS*，2012年。Oliver Kroemer、Scott Niekum 和 George Konidaris。《面向操作的机器人学习综述：挑战、表征与算法》。 作：挑——

战、表征与算法。*JMLR*，2021年。Benjamin Kuipers。《空间语义层级》。人工智能，119(1-2)：191–233，2000年。Benjamin Kuipers 和 Yung-Tai Byun。《基于 语义

空间表征层次结构的机器人探索与制图策略》。机器人与自主系统，8(1-2)：47–63，1991。Tejas D Kulkarni 等。《分层深度强化学习：整合时间抽象 行动与

内在动机。在*NeurIPS*，2016年。Ankit Kumar等人。《有问必答：用于自然语言处理的动态记忆网络》。 发表于*ICML*，

2016年。Brenden M Lake、Tomer D Ullman、Joshua B Tenenbaum 和 Samuel J Gershman。构建机 器，使其 像人一样学习和思考。《行为与脑科学》，2017年。

Alex H Lang、Sourabh Vora、Holger Caesar、Lubing Zhou、Jiong Yang 和 Oscar Beijbom。《PointPillars：用于点云目标检测的快速编码器》。发表于*CVPR*，第 12697–12705 页，2019年。

Steven M LaValle。《规划算法》。剑桥大学出版社，2006。

Yann LeCun。迈向自主机器智能之路。*OpenReview*，2022 .

Yann LeCun。迈向自主机器智能之路。*Open Review*，2024a。

Yann LeCun。目标驱动人工智能。自然-机器智能，2024b。

Yann LeCun、Yoshua Bengio 和 Geoffrey Hinton。深度学习。《自然》，521(7553)：436–444，2015年。

Dong-Hyun Lee。《伪标签：一种简单高效的深度神经网络半监督学习方法》。*ICML* 表征学习挑战研讨会，3(2):896，2013年。

Bin Lei 等。通过一种新框架提升大语言模型的逻辑推理能力：思维图。*arXiv* 预印本 *arXiv:2308.08614*，2023年。

Jan Leike、David Krueger、Tom Everitt、Miljan Martic、Vishal Maini 和 Shane Legg。通过奖励建模实现可扩展的智能体对齐：一个研究方向。*arXiv* 预印本 *arXiv:1811.07871*，2018年。

Ian Lenz、Honglak Lee 和 Ashutosh Saxena。用于检测机器人抓取的深度学习。*IJRR* ，34(4-5)：705–724，2015年。Sergey Levine、Chelsea Finn、Trevor Darrell 和 Pieter Abbeel。深度视 感觉运动

策略。*JMLR*，2016年。SergeyLevine、Peter Pastor、AlexKrizhevsky、Julian Ibarz 和 Deirdre Quillen。使用深度学习学习手部 眼协调——

以实现机器人抓取，并结合深度学习和大规模数据收集。*IJRR*，2018年。Andrew Levy、George Konidaris、Robert Platt 和 Kate Saenko。学习多层级层次结构 hie与

事后回顾。在*ICLR*，2019年。Patrick Lewis、Ethan Perez、Aleksandra Piktus、Fabio Petroni、Vladimir Karpukhin、Naman Goyal, Heinrich

Kuttler、Mike Lewis、Wen-tau Yih、Tim Rocktaschel 等。《检索增强生成 for k nowledge-密集型自然语言处理任务。神经信息处理系统进展，33：9459–9474，2020年。

Michael Ley。《DBLP计算机科学文献目录：进化、研究问题与展望》。国际字符串处理与信息检索研讨会，2002年。Chengshu Li、Fei Xia、Roberto Mart´ín-Mart´ín 和 Silvio Savarese。Hrl4in：用于移动操作机器人交互式导航的分层强化学习。*CoRL*，第603–616页，2020年。Chengshu Li、Fei Xia、Roberto Mart´ín-Mart´ín、Michael Lingelbach、Sanjana Srivastava、Bokui Shen、Kent Vainio、Cem Gokmen、Gokul Dharan、Tanish Jain 等人。iGibson 2.0：面向机器人学习日常家务任务的以对象为中心的仿真。载于机器人学习会议，第455–465页，2021年。Chengshu Li、Ruohan Zhang、Josiah Wong、Cem Gokmen、Sanjana Srivastava、Roberto Mart´ín-Mart´ín、Chen Wang、Sergey Levine、Michael Lingelbach、Jiankai Sun 等人。BEHAVIOR-1K：一个包含1,000项日常活动及逼真仿真的具身人工智能基准测试。载于*CoRL*，2023a。Guohao Li、Hasan Abed Al Kader Hammoud、Hani Itani、Dmitrii Khizbullin 和 Bernard Ghanem。Camel：用于探索大语言模型社会思维的交际智能体。*arXiv*预印本 *arXiv:2303.17760*，2023b。Guohao Li 等人。Camel：用于探索大语言模型社会思维的交际智能体。*arXiv*预印本 *arXiv:2303.17760*，2023c。Junnan Li、Dongxu Li、Silvio Savarese 和 Steven Hoi。BLIP-2：利用冻结的图像编码器和大语言模型进行语言-图像预训练的自举。*arXiv*预印本 *arXiv:2301.12597*，2023d。Minghao Li 等人。Api-bank：面向工具增强型大语言模型的综合基准测试。*EMNLP*，2023e。Raymond Li 等人。StarCoder：愿源代码与你同在！*arXiv*预印本 *arXiv:2305.06161*，2023f。Yaguang Li、Rose Yu、Cyrus Shahabi 和 Yan Liu。扩散卷积循环神经网络：数据驱动的交通流预测。载于国际学习表征会议，2018年。Yujia Li、Daniel Tarlow、Marc Brockschmidt 和 Richard Zemel。门控图序列神经网络。*ICLR*，2016年。Zhenlong Li、Huan Ning、Song Gao 和 Krzysztof Janowicz。人工智能时代的地理信息科学：迈向自主地理信息系统的研究议程。《地理信息科学年刊》，2025a。Zhenyu Li、Xuyang Wang、Xianming Liu 和 Junjun Jiang。Binsformer：重新审视用于单目深度估计的自适应分箱。*arXiv*预印本 *arXiv:2204.00987*，2022年。Ziyue Li、Yuan Chang、Gaihong Yu 和 Xiaoqiu Le。HiPlan：采用自适应全局-局部引导的基于大语言模型的智能体分层规划，2025b。Feng Liang、Bichen Wu、Xiaoliang Dai、Kunpeng Li、Yinan Zhao、Hang Zhang、Peizhao Zhang、Peter Vajda 和 Diana Marculescu。采用掩码适配的CLIP进行开放词汇语义分割。载于*CVPR*，2023a。Jacky Liang、Wenlong Huang、Fei Xia、Peng Xu、Karol Hausman、Brian Ichter、Pete Florence 和 Andy Zeng。Code as Policies：用于具身控制的语言模型程序。*arXiv*预印本 *arXiv:2209.07753*，2023b。Paul Pu Liang、Amir Zadeh 和 Louis-Philippe Morency。多模态学习基础。*arXiv*预印本 *arXiv:2209.03430*，2024年。Yaobo Liang 等人。Taskmatrix.ai：通过连接基础模型与数百万个API来完成任务。*arXiv*预印本 *arXiv:2303.16434*，2023c。Thomas Liao、Rohan Taori、Inioluwa Deborah Raji 和 Ludwig Schmidt。我们已经开始学习了吗？一篇关于机器学习评估失败的元综述。*NeurIPS*数据集与基准，2021年。Timothy P Lillicrap、Jonathan J Hunt、Alexander Pritzel、Nicolas Heess、Tom Erez、Yuval Tassa、David Silver 和 Daan Wierstra。深度强化学习中的连续控制。*ICLR*，2016年。

Bo Liu、Yuqian Jiang、Xiaohan Zhang、Qiang Liu、Shiqi Zhang、Joydeep Biber 和 Peter Stone。Llm+p：赋予大语言模型最优规划能力。*arXiv*预印本 *arXiv:2304.11477*，2023a。Chenxi Liu、Sun Yang、Qianxiong Xu、Zhishuai Li、Cheng Long、Ziyue Li 和 Rui Zhao。用于交通预测的时空大语言模型。*arXiv*预印本 *arXiv:2401.10134*，2024a。Haotian Liu、Chunyuan Li、Qingyang Wu 和 Yong Jae Lee。视觉指令微调。*arXiv*预印本 *arXiv:2304.08485*，2023b。Haotian Liu、Chunyuan Li、Yuheng Li 等。Llava-next：改进推理、光学字符识别和世界知识。*arXiv*预印本，2024b。Xiao Liu、Hao Yu、Hanchen Zhang、Yifan Xu、Xuanyu Lei、Hanyu Lai、Yu Gu、Hangliang Ding、Kaiwen Men、Kejuan Yang 等。Agentbench：评估作为智能体的大语言模型。*arXiv*预印本 *arXiv:2308.03688*，2023c。Yang Liu、Weixing Chen、Yongjie Bai、Xiaodan Luo、Yang Gao 和 Zhiwei Xiong。连接网络空间与物理世界：具身人工智能综合综述。载于*arXiv*预印本 *arXiv:2407.06886*，2024c。Yongcheng Liu、Bin Fan、Shiming Xiang 和 Chunhong Pan。用于点云分析的关系形状卷积神经网络。*CVPR*，2019。Zhihan Liu、Hao Hu、Shenao Zhang、Hongyi Guo、Shuqi Ke、Boyi Liu 和 Zhaoran Wang。为未来推理，为当下行动：一种具有可证明样本效率的自主大语言模型智能体原则性框架。*ICML*，2024d。Jieyi Long。大语言模型引导的思维树。*arXiv*预印本 *arXiv:2305.08291*，2023。Jonathan Long、Evan Shelhamer 和 Trevor Darrell。用于语义分割的全卷积网络。载于*IEEE/CVF*计算机视觉与模式识别会议，第3431–3440页，2015年。Paul A Longley、Michael F Goodchild、David J Maguire 和 David W Rhind。地理信息科学与系统。约翰·威立父子出版公司，2015年。Ryan Lowe、Yi I Wu、Aviv Tamar、Jean Harb、OpenAI、Pieter Abbeel 和 Igor Mordatch。用于混合合作竞争环境的多智能体行动者-评论家。载于*NeurIPS*，第6379–6390页，2017年。Cewu Lu、Ranjay Krishna、Michael Bernstein 和李飞飞。利用语言先验进行视觉关系检测。载于*ECCV*，2016年。Pan Lu、Baolin Peng、Hao Cheng、Michel Galley、Kai-Wei Chang、Ying Nian Wu、Song-Chun Zhu 和 Jian-feng Gao。Chameleon：借助大语言模型实现即插即用的组合推理。载于*NeurIPS*，2023年。Jonathon Luiten、Georgios Kopanas、Bastian Leibe 和 Deva Ramanan。动态三维高斯：通过持久动态视图合成进行跟踪。载于三维视觉会议，2023年。Ziyang Luo、Can Xu、Pu Zhao、Qingfeng Sun、Xiubo Geng、Wenxiang Hu、Chongyang Tao、Jing Ma、Qingwei Lin 和 Daxin Jiang。Wizardcoder：通过 Evol-Instruct 赋能代码大语言模型。载于*ICLR*，2023年。Corey Lynch、Mohi Khansari、Ted Xiao、Vikash Kumar、Jonathan Tompson、Sergey Levine 和 Pierre Sermanet。从游戏中学习潜在计划。载于机器人学习会议，第1113–1132页，2020年。Chih-Yao Ma、Jiasen Lu、Zuxuan Wu、Ghassan AlRegib、Zsolt Kira、Richard Socher 和 Caiming Xiong。通过辅助进度估计实现自监控导航智能体。*ICLR*，2019a。Lei Ma、Yu Liu、Xueliang Zhang、Yuanxin Ye、Gaofei Yin 和 Brian Alan Johnson。遥感应用中的深度学习：荟萃分析与综述。《*ISPRS*摄影测量与遥感杂志》，152：166–177，2019b。

Xiaojian Ma、Silong Yong、Zilong Zheng、Qing Li、Yitao Liang、Song-Chun Zhu 和 Siyuan Huang。Sqa3d：3D场景中的情境化问答。载于*ICLR*，2022a。Xu Ma 等。重新思考点云中的网络设计与局部几何：一种简单的残差多层感知机框架。载于*ICLR*，2022b。Jeffrey Mahler、Jacky Liang、Sherdil Niyaz、Michael Laskey、Richard Doan、Xinyu Liu、Juan Aparicio Ojea 和 Ken Goldberg。Dex-net 2.0：利用合成点云和解析抓取指标，通过深度学习规划稳健抓取。*Robotics: Science and Systems*，2017年。Gengchen Mai、Weiming Huang、Jin Sun、Suhang Song、Deepak Mishra、Ninghao Liu、Song Gao、Tianming Liu、Gao Cong、Yingjie Hu 等。地理空间人工智能基础模型的机遇与挑战。*arXiv*预印本 *arXiv:2304.06798*，2023年。Arjun Majumdar、Gunjan Aggarwal、Bhavika Devnani、Judy Hoffman 和 Dhruv Batra。Zson：使用多模态目标嵌入实现零样本物体目标导航。*arXiv*预印本 *arXiv:2206.12403*，2022年。Sagnik Majumder、Changan Chen、Ziad Al-Halah 和 Kristen Grauman。环境声学的少样本音视学习。*NeurIPS*，2022年。Viktor Makoviychuk 等。Isaac Gym：用于机器人学习的高性能基于GPU的物理仿真。*arXiv*预印本 *arXiv:2108.10470*，2021年。Yu A Malkov 和 Dmitry A Yashunin。使用分层可导航小世界图进行高效、稳健的近似最近邻搜索。*IEEE TPAMI*，2018年。Ajay Mandlekar、Danfei Xu、Josiah Wong 等。离线人类演示对机器人操作学习有何影响。2021年。Ajay Mandlekar、Danfei Xu、Josiah Wong、Soroush Nasiriany、Chen Wang、Rohun Kulkarni、李飞飞、Silvio Savarese、Yuke Zhu 和 Roberto Mart´ín-Mart´ín。离线人类演示对机器人操作学习有何影响。*CoRL*，第1678–1690页，2022年。Gary Marcus。人工智能的下一个十年：迈向稳健人工智能的四个步骤。*arXiv*预印本 *arXiv:2002.06177*，2020年。David Marr。视觉：对人类视觉信息表征与处理的计算研究。麻省理工学院出版社，1982年。Ricardo Martin-Brualla、Noha Radwan、Mehdi SM Sajjadi、Jonathan T Barron、Alexey Dosovitskiy 和 Daniel Duckworth。野外NeRF：面向非受限照片集的神经辐射场。载于*CVPR*，第7210–7219页，2021年。Jan Matas、Stephen James 和 Andrew J Davison。用于可变形物体操作的仿真到现实强化学习。载于*CoRL*，2018年。Yutaka Matsuo、Yann LeCun、Maneesh Sahani、Doina Precup、David Silver、Masashi Sugiyama、Eiji Uchibe 和 Jun Morimoto。深度学习、强化学习与世界模型。*Neural Networks*，152：438–450，2022年。Lars Mescheder、Michael Oechsle、Michael Niemeyer、Sebastian Nowozin 和 Andreas Geiger。占据网络：在函数空间中学习三维重建。*CVPR*，2019年。Gr´egoire Mialon、Roberto Dess`ı、Maria Lomeli、Christoforos Nalmpantis、Ram Pasunuru、Roberta Raileanu、Baptiste Rozi`ere、Timo Schick、Jane Dwivedi-Yu、Asli Celikyilmaz 等。增强型语言模型：综述。*TMLR*，2023年。Vincent Micheli、Eloi Alonso 和 Fran¸cois Fleuret。Transformer是样本高效的世界模型。*ICLR*，2023年。

Ben Mildenhall、Pratul P Srinivasan、Matthew Tancik、Jonathan T Barron、Ravi Ramamoorthi 和 Ren Ng。NeRF：将场景表示为用于视图合成的神经辐射场。《*ACM*通讯》，65(1):99–106，2020年。Sewon Min、Xinxi Lyu、Ari Holtzman、Mikel Arber、Mike Lewis、Hannaneh Hajishirzi 和 Luke Zettlemoyer。重新思考示例的作用：是什么让上下文学习发挥作用？*arXiv* 预印本 *arXiv:2202.12837*，2022年。Piotr Mirowski、Matthew Grimes、Mateusz Malinowski、Karl Moritz Hermann、Keith Anderson、Denis Teber、Koray Kavukcuoglu 和 Raia Hadsell。学习在没有地图的城市中导航。神经信息处理系统进展，31，2018年。Margaret Mitchell、Simone Wu、Andrew Zaldivar、Parker Barnes、Lucy Vasserman、Ben Hutchinson、Elena Spitzer、Inioluwa Deborah Raji 和 Timnit Gebru。用于模型报告的模型卡。在*FAT**，第220–229页，2019年。Melanie Mitchell。人工智能中的抽象与类比生成。纽约科学院年刊，1505(1):79–101，2021年。Volodymyr Mnih、Koray Kavukcuoglu、David Silver、Andrei A Rusu、Joel Veness、Marc G Bellemare、Alex Graves、Martin Riedmiller、Andreas K Fidjeland、Georg Ostrovski 等。通过深度强化学习实现人类水平的控制。《自然》，518(7540):529–533，2015年。Volodymyr Mnih、Adria Puigdo menech Badia、Mehdi Mirza、Alex Graves、Timothy Lillicrap、Tim Harley、David Silver 和 Koray Kavukcuoglu。深度强化学习的异步方法。在*ICML*，第1928–1937页，2016年。Thomas M Moerland、Joost Broekens、Aske Plaat 和 Catholijn M Jonker。基于模型的强化学习：综述。机器学习基础与趋势，16(1):1–118，2023年。David Moher、Alessandro Liberati、Jennifer Tetzlaff 和 Douglas G Altman。系统综述和荟萃分析的首选报告条目：PRISMA 声明。《内科学年鉴》，151(4): 264–269，2009年。Christopher Morris、Martin Ritzert、Matthias Fey、William L Hamilton、Jan Eric Lenssen、Gaurav Rattan 和 Martin Grohe。Weisfeiler 和 Leman 图神经网络：高阶图神经网络。*AAAI*，33(01): 4602–4609，2019年。Douglas Morrison、Peter Corke 和 J¨urgen Leitner。闭环机器人抓取：一种实时生成式抓取合成方法。在*Robotics: Science and Systems*，2018年。Edvard I Moser、Emilio Kropff 和 May-Britt Moser。位置细胞、网格细胞与大脑的空间表征系统。《神经科学年度评论》，31:69–89，2008年。Tongzhou Mu 等。ManiSkill：基于大规模示例的通用操作技能基准。在*NeurIPS*，2021年。Thomas M¨uller、Alex Evans、Christoph Schied 和 Alexander Keller。采用多分辨率哈希编码的即时神经图形基元。*ACM* 图形学汇刊，41(4):1–15，2022年。Tsendsuren Munkhdalai 和 Hong Yu。元网络。在*ICML*，2017年。Raul Mur-Artal 和 Juan D Tard´os。ORB-SLAM2：一种适用于单目、立体和 RGB-D 相机的开源同步定位与地图构建系统。*IEEE* 机器人学汇刊，33(5):1255–1262，2017年。Raul Mur-Artal 等。ORB-SLAM：一种通用且精确的单目同步定位与地图构建系统。*IEEE* 机器人学汇刊，2015年。Fabio Muratore、Fabio Ramos、Greg Turk、Wenhao Yu、Michael Gienger 和 Jan Peters。从随机化仿真中进行机器人学习：综述。《机器人学与人工智能前沿》，9:799893，2022年。

Austin Myers、Ching L Teo、Cornelia Fermüller 和 Yiannis Aloimonos。基于几何特征的工具部件可供性检测。在*ICRA*，第1374–1381页，2015年。Ofir Nachum、Shixiang Shane Gu、Honglak Lee 和 Sergey Levine。数据高效的分层强化学习。*NeurIPS*，2018年。Tushar Nagarajan、Christoph Feichtenhofer 和 Kristen Grauman。基于视频的人物-物体交互热点。在*ICCV*，第8688–8697页，2019年。Reiichiro Nakano、Jacob Hilton、Suchir Balaji、Jeff Wu、Long Ouyang、Christina Kim、Christopher Hesse、Shantanu Jain、Vineet Kosaraju、William Saunders 等。WebGPT：借助浏览器并结合人类反馈进行问答。*arXiv* 预印本 *arXiv:2112.09332*，2021年。Soroush Nasiriany、Huihan Liu 和 Yuke Zhu。利用行为基元增强强化学习，以应对多样化的操作任务。在*IEEE*国际机器人与自动化会议，第7477–7484页，2022年。Dana S Nau、Tsz-Chiu Au、Okhtay Ilghami、Ugur Kuter、J William Murdock、Dan Wu 和 Fusun Yaman。SHOP2：一种分层任务网络规划系统。《人工智能研究杂志》，20:379–404，2003年。Nora S Newcombe。想象一下：通过提升空间思维能力来促进数学和科学学习。《美国教育家》，2010年。Jiquan Ngiam、Aditya Khosla、Mingyu Kim、Juhan Nam、Honglak Lee 和 Andrew Y Ng。多模态深度学习。*ICML*，2011年。Richard Ngo、Lawrence Chan 和 Sören Geiger。从深度学习视角看对齐问题。*arXiv* 预印本 *arXiv:2209.00626*，2022年。John O'Keefe 和 Lynn Nadel。作为认知地图的海马体。克拉伦登出版社，1978年。

Chitu Okoli 和 Kira Schabram。信息系统研究系统性文献综述指南。《*Sprouts*：信息系统工作论文》，10(26)，2010年。

Theo X Olausson 等人。自我修复是代码生成的灵丹妙药吗？*arXiv* 预印本 *arXiv:2306.09896*，2023年。Avital Oliver、Augustus Odena、Colin A Raffel、Ekin Dogus Cubuk 和 Ian Goodfellow。深度半监督学习算法的真实评估。*NeurIPS*，31，2018年。Catherine Olsson 等人。上下文学习与归纳头。*arXiv* 预印本 *arXiv:2209.11895*，2022年。

OpenAI。GPT-4 技术报告。*arXiv* 预印本 *arXiv:2303.08774*，2023年。

OpenAI。GPT-4 技术报告。*arXiv* 预印本 *arXiv:2303.08774*，2023年。

OpenAI。GPT-4V(ision) 系统卡。*OpenAI* 技术报告，2023年。OpenAI。Sora：作为世界模拟器的视频生成模型。技术报告，2024年。Maxime Oquab、Timothée Darcet、Théo Moutakanni 等。DINOv2：无需监督地学习稳健的视觉特征。*TMLR*，2024年。Maxime Oquab 等。DINOv2：无需监督地学习稳健的视觉特征。*arXiv* 预印本 *arXiv:2304.07193*，2023年。Franc¸ois Osiurak 和 Arnaud Badets。什么是工具？迈向一种三元方法。《实验心理学季刊》，2016年。Long Ouyang、Jeffrey Wu、Xu Jiang、Diogo Almeida、Carroll Wainwright、Pamela Mishkin、Chong Zhang、Sandhini Agarwal、Katarina Slama、Alex Ray 等。通过人类反馈训练语言模型遵循指令。*NeurIPS*，35:27730–27744，2022年。

Charles Packer、Sarah Wooders、Kevin Lin、Vivian Fang、Shishir G Patil、Ion Stoica 和 Joseph E Gonzalez。MemGPT：让大语言模型成为操作系统。*arXiv* 预印本 *arXiv:2310.08560*，2023年。Abhishek Padalkar 等。Open x-embodiment：机器人学习数据集与 RT-X模型。*arXiv* 预印本 *arXiv:2310.08864*，2023年。Palantir。Palantir Technologies。https://www.palantir.com，2023年。Palantir Technologies。Palantir AIP：人工智能平台。https://www.palantir.com/ platforms/aip/，2024年。Yatian Pang 等。用于点云自监督学习的掩码自编码器。欧洲计算机视觉会议，2022年。Aaron Parisi、Yao Zhao 和 Noah Fiedel。TALM：工具增强语言模型。*arXiv* 预印本 *arXiv:2205.12255*，2022年。Cheonbok Park、Chunggi Lee、Hyojin Bahng、Yunwon Tae、Seungmin Jin、Kihwan Kim、Sungahn Ko 和 Jaegul Choo。St-grat：一种用于准确预测动态变化道路速度的新型时空图注意力网络。载于$CIKM$，2020年。Jeong Joon Park、Peter Florence、Julian Straub、Richard Newcombe 和 Steven Lovegrove。Deepsdf：学习用于形状表示的连续有符号距离函数。$CVPR$，2019年。Joon Sung Park、Joseph C O'Brien、Carrie J Cai、Meredith Ringel Morris、Percy Liang 和 Michael S Bernstein。生成式智能体：人类行为的交互式模拟体。*arXiv* 预印本 *arXiv:2304.03442*，2023年。Shishir G Patil、Tianjun Zhang、Xin Wang 和 Joseph E Gonzalez。Gorilla：连接海量 API 的大语言模型。*arXiv* 预印本 *arXiv:2305.15334*，2023年。Songyou Peng、Kyle Genova、Chiyu Max Jiang、Andrea Tagliasacchi、Marc Pollefeys 和 Thomas Funkhouser。Openscene：基于开放词汇的三维场景理解。*arXiv* 预印本 *arXiv:2211.15654*，2023年。Xue Bin Peng、Marcin Andrychowicz、Wojciech Zaremba 和 Pieter Abbeel。通过动力学随机化实现机器人控制的仿真到现实迁移。$ICRA$，2018年。Karl Pertsch、Youngwoon Lee 和 Joseph Lim。利用学习得到的技能先验加速强化学习。载于$CoRL$，2021年。Kai Petersen 等。软件工程中的系统映射研究。$EASE$，2008年。Pinecone。Pinecone 向量数据库，2023年。Planet Labs PBC。Planet Labs：每日卫星影像与洞察。https://www.planet.com/，2023年。Planet Labs PBC。Planet每日监测：全球地球观测。https://www.planet.com/products/ monitoring/，2024年。Dean A Pomerleau。Alvinn：基于神经网络的自主陆地车辆。1988年。David MW Powers。评估：从精确率、召回率和F值到ROC曲线、知情度、标记度和相关性。《机器学习技术杂志》，2011年。Simon JD Prince。计算机视觉：模型、学习与推理。剑桥大学出版社，2012年。Alexander Pritzel、Benigno Uria、Sriram Srinivasan、Adria Puigdomenech Badia、Oriol Vinyals、Demis Has- sabis、Daan Wierstra 和 Charles Blundell。神经情景控制。载于国际机器学习会议，第2827–2836页，2017年。

Xavier Puig、Kevin Ra、Marko Boben、Jiaman Li、Tingwu Wang、Sanja Fidler 和 Antonio Torralba。Virtualhome：通过程序模拟家庭活动。*arXiv*预印本 *arXiv:1806.07011*，2018年。Charles R Qi、Hao Su、Kaichun Mo 和 Leonidas J Guibas。PointNet：用于三维分类和分割的点集深度学习。载于$CVPR$，2017a年。Charles R Qi、Li Yi、Hao Su 和 Leonidas J Guibas。PointNet++：度量空间中点集的深度层次特征学习。$NeurIPS$，2017b年。Charles R Qi、Or Litany、Kaiming He 和 Leonidas J Guibas。点云三维目标检测的深度霍夫投票。$ICCV$，2019年。Yuankai Qi、Qi Wu、Peter Anderson、Xin Wang、William Yang Wang、Chunhua Shen 和 Anton van den Hengel。REVERIE：真实室内环境中的远程具身视觉指代表达。载于$CVPR$，2020年。Chen Qian、Xin Cong、Cheng Yang、Weize Chen、Yusheng Su、Juyuan Xu、Zhiyuan Liu 和 Maosong Sun。用于软件开发的交际智能体。*arXiv*预印本 *arXiv:2307.07924*，2023年。Yanyuan Qiao、Yuankai Qi、Yicong Hong、Zheng Yu、Peng Wang 和 Qi Wu。Hop：用于视觉与语言导航的历史与顺序感知预训练。$CVPR$，2022年。Yujia Qin、Shihao Liang、Yining Ye 等。ToolLLM：帮助大语言模型掌握 16000+真实世界API。载于$ICLR$，2024a。Yujia Qin等。ToolLLM：帮助大语言模型掌握 16000+ 真实世界API。*arXiv*预印本 *arXiv:2307.16789*，2024b。Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark等。通过自然语言监督学习可迁移视觉模型。*arXiv*预印本 *arXiv:2103.00020*，2021年。Inioluwa Deborah Raji、Emily M Bender、Amandalynne Paullada、Emily Denton 和 Alex Hanna。AI与整个广阔世界基准。*NeurIPS*数据集与基准，2021年。Ori Ram等。上下文内检索增强语言模型。$TACL$，2023年。Dhanesh Ramachandram 和 Graham W Taylor。深度多模态学习：近期进展与趋势综述。*IEEE*信号处理杂志，34(6):96–108，2017年。Santhosh Kumar Ramakrishnan、Devendra Singh Chaplot、Ziad Al-Halah、Jitendra Malik 和 Kristen Grauman。Poni：通过无交互学习实现物体目标导航的势函数。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第18890–18900页，2022年。Ladislav Ramp´aˇsek等。通用、强大且可扩展的图变换器构建方法。神经信息处理系统进展，2022年。David A Randell、Zhan Cui 和 Anthony G Cohn。基于区域与连接的空间逻辑。KR，92:165–176，1992年。Ren´e Ranftl、Katrin Lasinger、David Hafner、Konrad Schindler 和 Vladlen Koltun。迈向稳健的单目深度估计：混合数据集实现零样本跨数据集迁移。载于$IEEE\ TPAMI$，2020年。Ren´e Ranftl、Alexey Bochkovskiy 和 Vladlen Koltun。用于密集预测的视觉Transformer。载于$ICCV$，2021年。Tabish Rashid、Mikayel Samvelyan、Christian Schroeder、Gregory Farquhar、Jakob Foerster 和 Shimon Whiteson。QMIX：用于深度多智能体强化学习的单调价值函数分解。载于$ICML$，第4295–4304页，2018年。

Alexander Ratner、Stephen H Bach、Henry Ehrenberg、Jason Fries、Sen Wu 和 Christopher R´e。Snorkel：利用弱监督快速创建训练数据。*VLDB*，2017年。H Ravichandar、A S Polydoros、S Chernova 和 A Billard。机器人示范学习的最新进展。控制、机器人与自主系统年度评论，3:297–330，2020年。Marco Tulio Ribeiro、Tongshuang Wu、Carlos Guestrin 和 Sameer Singh。不止于准确率：使用CheckList对自然语言处理模型进行行为测试。*ACL*，2020年。Samuel Ritter、Jane X Wang、Zeb Kurth-Nelson、Siddhant M Jayakumar、Charles Blundell、Razvan Pascanu 和 Matthew Botvinick。经历过，做到过：通过情景回忆进行元学习。载于国际机器学习会议，第4354–4363页，2018年。Antoni Rosinol、Marcus Abate、Yun Chang 和 Luca Carlone。Kimera：用于实时度量-语义定位与建图的开源库。*arXiv*预印本 *arXiv:1910.02490*，2020年。Baptiste Roziere等。Code Llama：面向代码的开放基础模型。*arXiv*预印本 *arXiv:2308.12950*，2023年。Stuart Russell。《人类兼容：人工智能与控制问题》。企鹅出版社，2019年。Stuart Russell 和 Peter Norvig。《人工智能：一种现代方法》。Prentice Hall，2010年。Stuart Russell 和 Peter Norvig。《人工智能：一种现代方法》。培生，第4版，2021年。Andrei A Rusu、Matej Vecerik、Thomas Roth¨orl、Nicolas Heess、Razvan Pascanu 和 Raia Hadsell。使用渐进网络从像素实现仿真到现实的机器人学习。*CoRL*，2017年。Radu Bogdan Rusu 和 Steve Cousins。三维时代已至：点云库（PCL）。*ICRA*，2011年。Earl D Sacerdoti。抽象空间层次中的规划。人工智能，5(2):115–135，1974。Fereshteh Sadeghi 和 Sergey Levine。CAD2RL：无需真实图像即可实现真实的单图像飞行。*arXiv*预印本 *arXiv:1611.04201*，2017年。Mark Sandler、Andrew Howard、Menglong Zhu、Andrey Zhmoginov 和 Liang-Chieh Chen。MobileNetV2：倒置残差与线性瓶颈。*CVPR*，第4510–4520页，2018年。Adam Santoro、Sergey Bartunov、Matthew Botvinick、Daan Wierstra 和 Timothy Lillicrap。使用记忆增强神经网络进行元学习。载于*ICML*，2016年。V´ıctor Garcia Satorras、Emiel Hoogeboom 和 Max Welling。E(n)等变图神经网络。载于国际机器学习会议，第9323–9332页，2021年。Nikolay Savinov、Anton Raichuk、Raphael Marinier、Damien Vincent、Marc Pollefeys、Timothy Lillicrap 和 Sylvain Gelly。通过可达性实现情景好奇心。载于国际学习表征会议，2018年。Manolis Savva、Abhishek Kadian、Oleksandr Maksymets、Yili Zhao、Erik Wijmans、Bhavana Jain、Julian Straub、Jia Liu、Vladlen Koltun、Jitendra Malik等。Habitat：具身人工智能研究平台。载于*IEEE/CVF*国际计算机视觉会议论文集，第9339–9347页，2019年。Franco Scarselli、Marco Gori、Ah Chung Tsoi、Markus Hagenbuchner 和 Gabriele Monfardini。图神经网络模型。*IEEE*神经网络汇刊，20(1):61–80，2009年。Timo Schick、Jane Dwivedi-Yu、Roberto Dessi、Roberta Raileanu、Maria Lomeli、Luke Zettlemoyer、Nicola Cancedda 和 Thomas Scialom。Toolformer：语言模型能够自学使用工具。*arXiv*预印本 *arXiv:2302.04761*，2023年。J¨urgen Schmidhuber。论如何学会思考：用于强化学习控制器与循环神经世界模型新颖组合的算法信息论。*arXiv*预印本 *arXiv:1511.09249*，2015年。

Julian Schrittwieser、Ioannis Antonoglou、Thomas Hubert、Karen Simonyan、Laurent Sifre、Simon Schmitt、Arthur Guez、Edward Lockhart、Demis Hassabis、Thore Graepel等人。通过学习模型进行规划，精通雅达利、围棋、国际象棋和将棋。《自然》，588(7839):604–609，2020年。John Schulman、Filip Wolski、Prafulla Dhariwal、Alec Radford和Oleg Klimov。近端策略优化算法。*arXiv*预印本 *arXiv:1707.06347*，2017年。Kristof Sch¨utt、Pieter-Jan Kindermans、Huziel Enoc Sauceda Felix、Stefan Chmiela、Alexandre Tkatchenko和Klaus-Robert M¨uller。SchNet：一种用于建模量子相互作用的连续滤波卷积神经网络。*NeurIPS*，30，2017年。Amanda Seed和Richard Byrne。工具使用中的认知。《英国皇家学会哲学汇刊*B*辑：生物科学》，365(1545):1405–1412，2010年。Sanjit A Seshia、Dorsa Sadigh和S Shankar Sastry。迈向经验证的人工智能。《*ACM*通讯》，2022年。Dhruv Shah、Michael Equi、Blazej Osinski、Fei Xia、Brian Ichter和Sergey Levine。使用大语言模型进行导航：将语义猜测作为规划的启发式方法。*CoRL*，2023a。Dhruv Shah、Blazej Osinski、Brian Ichter和Sergey Levine。LM-Nav：利用语言、视觉和动作的大型预训练模型进行机器人导航。*arXiv*预印本 *arXiv:2207.04429*，2023b。Ahsan Shehzad、Feng Xia、Shagufta Abid、Chao Peng、Shuo Yu、Dongyu Zhang和Karin Verspoor。图 Transformer：综述。*arXiv*预印本 *arXiv:2407.09777*，2024年。Bokui Shen、Fei Xia、Chengshu Li、Roberto Martin、Linxi Fan、Guanzhi Wang、Shyamal Buch等人。iGibson 2.0：面向日常家务任务机器人学习的以对象为中心的仿真。*arXiv*预印本 *arXiv:2108.03272*，2021年。Tianchang Shen等人。Point-E：一个从复杂提示生成三维点云的系统。*arXiv*预印本，2023a。Yongliang Shen、Kaitao Song、Xu Tan等人。HuggingGPT：借助ChatGPT及其在Hugging Face中的伙伴解决人工智能任务。2023b。Shaoshuai Shi、Xiaogang Wang和Hongsheng Li。PointRCNN：基于点云的三维目标提议生成与目标检测。*CVPR*，2019年。Shaoshuai Shi、Chaoxu Guo、Li Jiang、Zhe Wang、Jianping Shi、Xiaogang Wang和Hongsheng Li。PV-RCNN：用于三维目标检测的点体素特征集合抽象。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第10529–10538页，2020年。Weijia Shi等人。REPLUG：检索增强黑盒语言模型。*arXiv*预印本 *arXiv:2301.12652*，2023年。Noah Shinn、Federico Cassano、Edward Berman、Ashwin Gopinath、Karthik Narasimhan和Shunyu Yao。Reflexion：具有语言强化学习的语言智能体，2023a。Noah Shinn、Federico Cassano、Ashwin Gopinath、Karthik Narasimhan和Shunyu Yao。Reflexion：具有语言强化学习的语言智能体。*arXiv*预印本 *arXiv:2303.11366*，2023b。Ben Shneiderman。以人为本的人工智能：可靠、安全且值得信赖。《国际人机交互杂志》，2020年。Ben Shneiderman。以人为本的人工智能。牛津大学出版社，2022年。Mohit Shridhar、Jesse Thomason、Daniel Gordon、Yonatan Bisk、Winson Han、Roozbeh Mottaghi、Luke Zettlemoyer和Dieter Fox。ALFRED：用于解释日常任务具身指令的基准测试。*arXiv*预印本 *arXiv:1912.01734*，2020a。

Mohit Shridhar、Lucas Manuelli和Dieter Fox。Cliport：面向机器人操作的“何物”与“何处”路径。载于*CoRL*，2022年。Mohit Shridhar、Lucas Manuelli和Dieter Fox。Perceiver-actor：用于机器人操作的多任务变换器。载于机器人学习会议，第785–799页，2023年。Mohit Shridhar等人。Alfworld：为交互式学习对齐文本与具身环境。*arXiv*预印本 *arXiv:2010.03768*，2020b。Robert W Shumaker、Kristina R Walkup和Benjamin B Beck。动物工具行为：动物对工具的使用与制造。约翰·霍普金斯大学出版社，2011年。Roland Siegwart、Illah Reza Nourbakhsh和Davide Scaramuzza。自主移动机器人导论。麻省理工学院出版社，2011年。David Silver、Thomas Hubert、Julian Schrittwieser、Ioannis Antonoglou、Matthew Lai、Arthur Guez、Marc Lanctot、Laurent Sifre、Dharshan Kumaran、Thore Graepel等人。一种通过自我对弈精通国际象棋、将棋和围棋的通用强化学习算法。《科学》，362(6419):1140–1144，2018年。David Silver等人。利用深度神经网络和树搜索精通围棋。《自然》，529:484–489，2016年。David Silver等人。通过通用强化学习算法的自我对弈精通国际象棋和将棋。*arXiv*预印本 *arXiv:1712.01815*，2017年。Tom Silver等人。利用预训练大语言模型在PDDL领域中进行泛化规划。*AAAI*人工智能会议，2024年。Karen Simonyan和Andrew Zisserman。用于大规模图像识别的极深卷积网络。载于*ICLR*，2015年。Avi Singh、Huihan Liu、Gaoyue Zhou、Albert Yu、Nicholas Rhinehart和Sergey Levine。Parrot：面向强化学习的数据驱动行为先验。*ICLR*，2021年。Ishika Singh、Valts Blukis、Arsalan Mousavian等人。ProgPrompt：使用大语言模型生成情境化机器人任务规划。2023年。Hannah Snyder。作为研究方法论的文献综述：概述与指南。《商业研究杂志》，104:333–339，2019年。Marina Sokolova和Guy Lapalme。分类任务性能指标的系统分析。《信息处理与管理》，2009年。Kyunghwan Son、Daewoo Kim、Wan Ju Kang、David Earl Hostallero和Yung Yi。Qtran：面向协作式多智能体强化学习的基于变换的分解学习。*ICML*，第5887–5896页，2019年。Chan Hee Song、Jiaman Wu、Clayton Washington、Brian M Sadler、Wei-Lun Chao和Yu Su。LLM-Planner：面向具身智能体的基于大语言模型的少样本落地规划。载于*IEEE/CVF*国际计算机视觉会议论文集，第2998–3009页，2023年。Chao Song、Youfang Lin、Shengnan Guo和Huaiyu Wan。时空同步图卷积网络：一种用于时空网络数据预测的新框架。*AAAI*，34(01):914–921，2020年。Nitish Srivastava和Ruslan R Salakhutdinov。基于深度玻尔兹曼机的多模态学习。*NeurIPS*，25，2012年。Siddharth Srivastava、Eugene Fang、Lorenzo Riano、Rohan Chitnis、Stuart Russell和Pieter Abbeel。通过可扩展的独立于规划器的接口层实现任务与运动联合规划。载于*ICRA*，2014年。

Kaya Stechly、Karthik Valmeekam 和 Subbarao Kambhampati。大语言模型在推理和规划任务中的自我验证局限性。*arXiv* 预印本 *arXiv:2402.08115*，2024年。Peter Stone 和 Manuela Veloso。多智能体系统：从机器学习视角出发的综述。自主机器人，2000年。Simone Stumpf、Vidya Rajaram、Lida Li、Weng-Keen Wong、Margaret Burnett、Thomas Dietterich、Erin Sullivan 和 Jonathan Herlocker。与机器学习系统进行有意义的交互：三项实验。国际人机研究期刊，67(8):639–662，2009年。Edgar Sucar 等人。imap：实时隐式建图与定位。*IEEE* 国际计算机视觉会议，2021年。Sainbayar Sukhbaatar 等人。端到端记忆网络。载于 *NeurIPS*，2015年。Gencer Sumbul、Marcela Charfuelan、Begum Demir 和 Volker Markl。BigEarthNet：用于遥感图像理解的大规模基准数据集。*arXiv* 预印本 *arXiv:1902.06148*，2019年。Theodore R Sumers、Shunyu Yao、Karthik Narasimhan 和 Thomas L Griffiths。语言智能体的认知架构。*arXiv* 预印本 *arXiv:2309.02427*，2024年。Peter Sunehag、Guy Lever、Audrunas Gruslys、Wojciech Marian Czarnecki、Vinicius Zambaldi、Max Jader- berg、Marc Lanctot、Nicolas Sonnerat、Joel Z Leibo、Karl Tuyls 等人。基于团队奖励的协作式多智能体学习价值分解网络。*AAMAS*，第2085–2087页，2018年。D´ıdac Sur´ıs、Sachit Menon 和 Carl Vondrick。ViperGPT：通过 Python 执行进行视觉推理。2023年。Rich Sutton。苦涩的教训。未完成的想法（博客），13(1)，2019年。Richard S Sutton。Dyna：用于学习、规划和反应的集成架构。*ACM SIGART* 公报，2(4):160–163，1991年。Richard S Sutton、Doina Precup 和 Satinder Singh。在 MDP 与半 MDP 之间：强化学习中时间抽象的框架。人工智能，112(1-2):181–211，1999年。Christian Szegedy、Wei Liu、Yangqing Jia、Pierre Sermanet、Scott Reed、Dragomir Anguelov、Dumitru Erhan、Vincent Vanhoucke 和 Andrew Rabinovich。利用卷积实现更深层次的网络。载于*CVPR*，第1–9页，2015年。Richard Szeliski。《计算机视觉：算法与应用》。Springer Nature，2022年。Andrew Szot、Alexander Clegg、Eric Undersander 等人。Habitat 2.0：训练家居助手重新布置其环境。载于 *NeurIPS*，2021年。Yashar Talebirad 和 Amirhossein Nadiri。多智能体协作：发挥大语言模型智能体的力量。*arXiv* 预印本 *arXiv:2306.03314*，2023年。Hao Tan、Licheng Yu 和 Mohit Bansal。学习导航于未见环境：利用环境丢弃进行反向翻译。载于*NAACL*，2019年。Mingxing Tan 和 Quoc Le。EfficientNet：重新思考卷积神经网络的模型缩放。载于*ICML*，第6105–6114页，2019年。Matthew Tancik、Vincent Casser、Xinchen Yan、Sabeek Pradhan、Ben Mildenhall、Pratul P Srinivasan、Jonathan T Barron 和 Henrik Kretzschmar。Block-NeRF：可扩展的大场景神经视图合成。*CVPR*，2022年。Jiabin Tang、Yuhao Yang、Wei Wei、Lei Shi、Lixin Su、Suqi Cheng、Dawei Yin 和 Chao Huang。GraphGPT：面向大语言模型的图指令微调。*arXiv* 预印本 *arXiv:2310.13023*，2024年。Qiaoyu Tang、Ziliang Deng、Hongyu Lin、Xianpei Han、Qiao Liang 和 Le Sun。ToolAlpaca：利用3000个仿真案例实现语言模型的通用工具学习。*arXiv* 预印本 *arXiv:2306.05301*，2023年。

Antti Tarvainen 和 Harri Valpola。平均教师是更好的榜样：权重平均一致性目标改善半监督学习结果。载于 *NeurIPS*，第1195–1204页，2017年。Gemini 团队和谷歌。Gemini：高能力多模态模型系列。*arXiv* 预印本 *arXiv:2312.11805*，2023年。Octo 模型团队等人。Octo：开源通用机器人策略。*arXiv* 预印本 *arXiv:2405.12213*，2024年。Zachary Teed 和 Jia Deng。Droid-SLAM：适用于单目、立体和 RGB-D 相机的深度视觉 SLAM。神经信息处理系统进展，2021年。Zachary Teed、Lahav Lipson 和 Jia Deng。利用自适应记忆实现深度视觉里程计。*IEEE* 模式分析与机器智能汇刊，2024年。

Andreas ten Pas、Marcus Gualtieri、Kate Saenko 与 Robert Platt。点云中的抓取姿态检测。《国际机器人研究杂志》，36(13-14):1455–1473，2017年。

特斯拉。特斯拉完全自动驾驶。https://www.tesla.com/autopilot，2023年。

特斯拉。特斯拉自动驾驶辅助系统与完全自动驾驶：人工智能驱动的驾驶辅助。技术报告，特斯拉公司，2024年。

Hugues Thomas 等人。核点卷积：面向点云的灵活可变形卷积。*IEEE* 国际计算机视觉会议，2019年。

Nathaniel Thomas、Tess Smidt、Steven Kearnes、Lusann Yang、Li Li、Kai Kohlhoff 与 Patrick Riley。张量场网络：面向三维点云的旋转与平移等变神经网络。见*arXiv* 预印本 *arXiv:1802.08219*，2018年。

Sebastian Thrun。面向室内移动机器人导航的度量拓扑地图学习。《人工智能》，1998年。

Sebastian Thrun。机器人建图：综述。Morgan Kaufmann，2002年。

Sebast ian Thrun、Wolfram Burgard 和 Dieter Fox。概率机器人学。麻省理工学院出版社，200 5.

Xiaoyu Tian 等。DriveVLM：自动驾驶与大型视觉语言模型的融合。*arXiv* 预印本 *arXiv:2402.12289*，2024年。

Josh Tobin、Rocky Fong、Alex Ray、John Schneider、Wojciech Zaremba 和 Pieter Abbeel。面向将深度神经网络从仿真迁移到真实世界的域随机化。载于*2017*年*IEEE/RSJ* 智能机器人与系统国际会议（*IROS*），第23–30页。I EEE，2017年。

Edward C Tolman。大鼠与人的认知地图。《心理学评论》，55(4):189，19 48.

Michael Tomasello. 人类认知的文化起源. 哈佛大学出版社，1999.

Marc Toussaint. 逻辑几何规划：一种基于优化的联合任务与运动规划方法。载于国际人工智能联合会议，2015年。

Hugo Touvron、Louis Martin、Kevin Stone等。Llama 2：开放基础模型与微调聊天模型。*arXiv*预印本 *arXiv:2307.09288*，2023年。

David Tranfield、David Denyer与Palminder Smart。迈向通过系统综述发展循证管理知识的方法论。《英国管理学杂志》，14(3)：207–222，2003年。

Samuel Triest、Matthew Sivaprakasam、Sean Wang、Wenshan Wang、Aaron Johnson与Sebastian Scherer。TartanDrive：用于学习越野动力学模型的大规模数据集。*IEEE*国际机器人与自动化会议（*ICRA*），2024年。

Harsh Trivedi、Niranjan Balasubramanian、Tushar Khot 和 Ashish Sabharwal。《针对知识密集型多步问题，将检索与思维链推理交替进行》。发表于 *ACL*，2023年。

Devis Tuia、Claudio Persello 和 Lorenzo Bruzzone。《遥感数据分类的领域自适应：近期进展综述》。*IEEE*地球科学与遥感杂志，4(2)：41–57，2016年。

Endel Tulving。情景记忆与语义记忆。《记忆的组织》，1972年。Krist Vaesen。人类工具使用的认知基础。《行为与脑科学》，2012年。Karthik Valmeekam、Matthew Marquez、Sarath Sreedharan 和 Subbarao Kambhampati。论大语言模型的规划能力——一项批判性考察。《神经信息处理系统进展》，36，2023a。Karthik Valmeekam、Matthew Marquez、Sarath Sreedharan 和 Subbarao Kambhampati。论大语言模型的规划能力——一项批判性考察。NeurIPS，2023b。Karthik Valmeekam、Matthew Marquez、Alberto Olmo、Sarath Sreedharan 和 Subbarao Kambhampati。Planbench：用于评估大语言模型规划与变化推理能力的可扩展基准测试。NeurIPS，2024年。Petar Veli˘ckovi´c。一切皆相连：图神经网络。《结构生物学当前观点》，79：102538，2023年。Petar Veli˘ckovi´c、Guillem Cucurull、Arantxa Casanova、Adriana Romero、Pietro Lio 和 Yoshua Bengio。图注意力网络。ICLR，2018年。Alexander Sasha Vezhnevets、Simon Osindero、Tom Schaul、Nicolas Heess、Max Jaderberg、David Silver 和 Koray Kavukcuoglu。用于分层强化学习的封建网络。ICML，2017年。Oriol Vinyals、Igor Babuschkin、Wojciech M Czarnecki、Micha¨el Mathieu、Andrew Dudzik、Junyoung Chung、David H Choi、Richard Powell、Timo Ewalds、Petko Georgiev 等。通过多智能体强化学习达到《星际争霸II》宗师级水平。《自然》，575(7782)：350–354，2019年。Johanna Wald、Helisa Dhamo、Nassir Navab 和 Federico Tombari。从三维室内重建中学习三维语义场景图。arXiv 预印本 arXiv:2004.03967，2020年。Homer Walke、Kevin Black、Tony Z Zhao 等。BridgeData V2：一个面向大规模机器人学习的数据集。发表于 CoRL，2023年。Homer Walke、Kevin Black、Tony Z Zhao、Quan Vuong、Chongyi Zheng、Philippe Hansen-Estruch、An- dre Wang He、Vivek Myers、Moo Jin Kim、Max Du 等。BridgeData V2：一个面向大规模机器人学习的数据集。CoRL，2024年。Dakuo Wang、Justin D Weisz、Michael Muller、Parikshit Ram、Werner Geyer、Casey Dugan、Yla Tausczik、Horst Samulowitz 和 Alexander Gray。数据科学中的人类与人工智能协作：探究数据科学家对自动化人工智能的看法。CSCW，3：1–24，2019a。Guanzhi Wang、Yuqi Xie、Yunfan Jiang、Ajay Mandlekar、Chaowei Xiao、Yuke Zhu、Linxi Fan 和 Anima Anandkumar。Voyager：一种基于大语言模型的开放式具身智能体。arXiv 预印本 arXiv:2305.16291，2023a。Jianguo Wang、Xiaomeng Yi、Rentong Guo、Hai Jin、Peng Xu、Shengjun Li、Xiangyu Wang、Xiangzhou Guo、Chengming Li、Xiaohai Xu 等。Milvus：一种专用向量数据管理系统，2021年。Ke Wang、Houxing Ren、Aojun Zhou、Zimu Lu、Sichun Luo、Weikang Shi、Renrui Zhang、Linqi Song、Mingjie Zhan 和 Hongsheng Li。MathCoder：通过在大语言模型中无缝集成代码来增强数学推理。ICLR，2024a。Lei Wang、Chen Ma、Xueyang Feng、Zeyu Zhang、Hao Yang、Jingsen Zhang、Zhiyuan Chen、Jiakai Tang、Xu Chen、Yankai Lin 等。基于大语言模型的自主智能体综述。《计算机科学前沿》，2024b。

Senzhang Wang、Jiannong Cao 和 Philip S Yu。用于时空数据挖掘的深度学习：综述。*IEEE TKDE*，2020年。

Weihan Wang 等人。CogVLM：面向预训练语言模型的视觉专家。*arXiv* 预印本 *arXiv:2311.03079*，2023b。Xin Wang、Qiuyuan Huang、Asli Celikyilmaz、Jianfeng Gao、Dinghan Shen、Yuan-Fang Wang、William Yang Wang 和 Lei Zhang。用于视觉-语言导航的强化跨模态匹配与自监督模仿学习。载于 *CVPR*，2019b。

Xuezhi Wang、Jason Wei、Dale Schuurmans、Quoc Le、Ed Chi、Sharan Narang、Aakanksha Chowdhery 和 Denny Zhou。自一致性能够提升语言模型中的思维链推理。*arXiv* 预印本 *arXiv:2203.11171*，2022 年。

Yue Wang、YongbinSun、ZiweiLiu、SanjayE Sarma、MichaelMBronstein 和JustinMSolomon。用于点云学习的动态图卷积神经网络。*ACM TOG*，2019年。

Waymo。Waymo：全球经验最丰富的驾驶员。https://waymo.com，2023年。

Waymo。Waymo驾驶员：自动驾驶技术。https://waymo.com/waymo-driver/，2024年。

Waymo。介绍Waymo´关于自动驾驶端到端多模态模型的研究。Waymo博客，2024年10月。网址：https://waymo.com/blog/2024/10/introducing-emma。

Jane Webster 和 Richard T Watson。分析过去，为未来做准备：撰写文献综述。*MIS Quarterly*，第 xiii–xxiii 页，2002年。

Jason Wei、Xuezhi Wang、Dale Schuurmans、Maarten Bosma、Brian Ichter、Fei Xia、Ed Chi、Quoc Le 和 Denny Zhou。思维链提示能够引发大语言模型的推理。神经信息处理系统进展，35：24824–24837，2022年。

Jerry Wei等人。《更大的语言模型以不同方式进行上下文学习》。*arXiv*预印本 *arXiv:2303.03846*，2023年。

Maurice Weiler和Gabriele Cesa。《通用E(2)等变可转向卷积神经网络》。载于神经信息处理系统进展，2019年。

Lilian Weng。《大语言模型驱动的自主智能体》。*Lil'Log*，2023年。https://lilianweng.github.io/posts/2023年6月23日-agent/。

Jason Weston、Sumit Chopra和Antoine Bordes。《记忆网络》。载于*ICLR*，2015年。

Wherobots。《Wherobots 云原生空间分析》。https://wherobots.com/，2023年。
Wherobots。Ap
ache sedona：《空间数据系统》。https://wherobots.com/apache-sedona/，2024年。
Erik Wijmans、Samyak Datta、Oleksandr Maksymets、Abhishek Das、Georgia Gkioxari、Stefan Lee、Irfan Essa、Devi Parikh 和 Dhruv Batra。在*IEEE/CVF*计算机视觉与模式识别会议上发表：《基于点云感知的照片级真实环境中的具身问答》，第 6659–6668 页，2019年。

Erik Wijmans、Abhishek Kadian、Ari Morber、StefanLee、Irfan Essa、Devi Parikh 和 Dhruv Batra。DD-PPO：利用 25 亿帧数据学习近乎完美的点目标导航器。*ICLR*，2020年。

Claes Wohlin。《系统性文献研究中的雪球抽样指南及其在软件工程中的复现》。*EASE*，2014年。

Michael Wooldridge。多智能体系统导论。Wiley，第 2 版，2009年。

Chenfei Wu、Jian Liang、Lei Ji、Fan Yang、Yuejian Fang、Daxin Jiang 和 Nan Duan。N¨uwa：用于神经视觉世界创建的视觉合成预训练。*ECCV*，第720–736页，2022年。

Chenfei Wu、Shengming Yin、Weizhen Qi、Xiaodong Wang、Zecheng Tang 和 Nan Duan。Visual ChatGPT：借助视觉基础模型进行对话、绘图和编辑。载于 *arXiv* 预印本 *arXiv:2303.04671*，2023a。Jialong Wu、Shaofeng Yin、Ningya Feng、Xu He、Dong Li、Jianye Hao 和 Mingsheng Long。iVideoGPT：交互式视频GPT是可扩展的世界模型。*NeurIPS*，2024年。Philipp Wu 等人。DayDreamer：用于实体机器人学习的世界模型。*arXiv* 预印本 *arXiv:2206.14176*，2023b。Qingyun Wu、Gagan Bansal、Jieyu Zhang、Yiran Wu、Beibin Li、Erkang Zhu、Li Jiang、Xiaoyun Zhang、Shaokun Zhang、Jiale Liu 等人。AutoGen：通过多智能体对话实现新一代大语言模型应用。*arXiv* 预印本 *arXiv:2308.08155*，2023c。Shun-Cheng Wu 等人。SceneGraphFusion：从RGB-D序列增量式预测三维场景图。载于*CVPR*，2021年。Wenxuan Wu、Zhongang Qi 和 Li Fuxin。PointConv：三维点云上的深度卷积网络。*CVPR*，2019a。Zhirong Wu 等人。3D ShapeNets：一种用于三维形状表示的深度学习方法。*IEEE*计算机视觉与模式识别会议，2015年。Zonghan Wu、Shirui Pan、Guodong Long、Jing Jiang 和 Chengqi Zhang。Graph WaveNet：用于深度时空图建模。载于 第二十八届国际人工智能联合会议论文集，第1907–1913页，2019b。Zonghan Wu、Shirui Pan、Fengwen Chen、Guodong Long、Chengqi Zhang 和 Philip S Yu。图神经网络综述。*IEEE*神经网络与学习系统汇刊，2020a。Zonghan Wu 等人。图神经网络综述。*IEEE*神经网络与学习系统汇刊，2020b。Zhiheng Xi、Wenxiang Chen、Xin Guo、Wei He、Yiwen Ding、Boyang Hong 等人。基于大语言模型的智能体的兴起与潜力：综述。*arXiv* 预印本 *arXiv:2309.07864*，2023年。Gui-Song Xia 等人。AID：用于评估航空场景分类性能的基准数据集。*IEEE*地球科学与遥感汇刊，2017年。Fanbo Xiang、Yuzhe Qin、Kaichun Mo、Yikuan Xia、Hao Zhu、Fangchen Liu、Minghua Liu、Hanxiao Jiang、Yifu Yuan、He Wang 等人。SAPIEN：基于部件的交互式仿真环境。载于 *CVPR*，第11097–11107页，2020年。Aoran Xiao、Weihao Xuan、Junjue Wang、Jiaxing Huang、Dacheng Tao、Shijian Lu 和 Naoto Yokoya。用于遥感与地球观测的基础模型：综述，2025年。URL https://arxiv.org/ abs/2410.16602。Enze Xie、Wenhai Wang、Zhiding Yu、Anima Anandkumar、Jose M Alvarez 和 Ping Luo。SegFormer：采用Transformer进行语义分割的简单高效设计。载于 *NeurIPS*，2021年。Sang Michael Xie 等人。将上下文学习解释为隐式贝叶斯推断。载于 *ICLR*，2022年。Yuxi Xie、Kenji Kawaguchi、Yiran Zhao、Xu Zhao、Min-Yen Kan、Junxian He 和 Qizhe Xie。用于推理的自评引导束搜索。*NeurIPS*，2023年。Danfei Xu 等人。通过迭代消息传递生成场景图。*IEEE*计算机视觉与模式识别会议，2017年。Keyulu Xu、Chengtao Li、Yonglong Tian、Tomohiro Sonobe、Ken-ichi Kawarabayashi 和 Stefanie Jegelka。利用跳跃知识网络进行图表示学习。*ICML*，第5453–5462页，2018年。

Keyulu Xu、Weihua Hu、Jure Leskovec 和 Stefanie Jegelka。图神经网络有多强大？*ICLR*，2019年。Mengda Xu、Zhenjia Huang、Yinghao Li 和 Shuran Song。Xskill：跨具身技能发现。*CoRL*，2023a。Mengde Xu、Zheng Zhang、Fangyun Wei、Han Hu 和 Xiang Bai。用于开放词汇语义分割的侧适配器网络。载于*CVPR*，第2945–2954页，2023b。Peng Xu、Xiatian Zhu 和 David A Clifton。基于Transformer的多模态学习：综述。*IEEE*模式分析与机器智能汇刊，2023c。Qiantong Xu、Fenglu Hong、Bo Li、Changran Hu、Zhengyu Chen 和 Jian Zhang。开源大语言模型的工具操作能力探究。*arXiv*预印本 *arXiv:2305.16504*，2023d。Runsen Xu、Xiaolong Wang、Tai Wang、Yilun Chen、Jiangmiao Pang 和 Dahua Lin。PointLLM：赋能大语言模型理解点云。*ECCV*，2024年。Wujiang Xu、Zujie Liang、Kai Mei、Hang Gao、Juntao Tan 和 Yongfeng Zhang。A-mem：面向大语言模型智能体的智能体记忆，2025年。An Yan 等人。奇境中的GPT-4V：用于零样本智能手机图形用户界面导航的大型多模态模型。*arXiv*预印本 *arXiv:2311.07562*，2023年。Wilson Yan、Yunzhi Zhang、Pieter Abbeel 和 Aravind Srinivas。VideoGPT：使用 VQ-VAE 和 Transformer 进行视频生成。*arXiv* 预印本 *arXiv:2104.10157*，2021年。Alejandro Escontrela Yang、Russell Mendonca、Danijar Hafner 和 Pieter Abbeel。将视频预测模型用作强化学习的奖励。2024a。Jianing Yang、Xuweiyi Chen、Shengyi Qian 等人。LLM-Grounder：将大语言模型用作智能体，实现开放词汇三维视觉定位。2024b。Lihe Yang、Bingyi Kang、Zilong Huang、Xiaogang Xu、Jiashi Feng 和 Hengshuang Zhao。Depth Anything：释放大规模无标注数据的潜力。*CVPR*，2024c。Mengjiao Yang 等人。UniSim：学习交互式现实世界模拟器。*arXiv*预印本 *arXiv:2310.06114*，2023a。Qian Yang、Aaron Steinfeld、Carolyn Ros´e 和 John Zimmerman。重新审视人机交互是否、为何以及如何具有独特的设计难度。载于*CHI*，第1–13页，2020年。Rui Yang、Hanyang Lin、Junyu Zhu 和 Jingyi Huang。EmbodiedBench：面向视觉驱动具身智能体的多模态大语言模型综合基准测试。*arXiv*预印本 *arXiv:2502.09560*，2025年。Xiaofeng Yang 等人。WorldDreamer：通过预测掩码词元，迈向用于视频生成的通用世界模型。*arXiv*预印本 *arXiv:2401.09985*，2024d。Zhengyuan Yang、Linjie Li、Jianfeng Wang、Kevin Lin、Ehsan Azarnasab、Faisal Ahmed、Zicheng Liu、Ce Liu、Michael Zeng 和 Lijuan Wang。MM-REACT：通过提示ChatGPT进行多模态推理与行动。*arXiv*预印本 *arXiv:2303.11381*，2023b。Huaxiu Yao、Fei Wu、Jintao Ke、Xianfeng Tang、Yitian Jia、Siyu Lu、Pinghua Gong、Jieping Ye 和 Zhenhui Li。用于出租车需求预测的深度多视图时空网络。载于*AAAI*，2018年。Shunyu Yao、Dian Yu、Jeffrey Zhao、Izhak Shafran、Thomas L Griffiths、Yuan Cao 和 Karthik Narasimhan。思维树：利用大语言模型进行审慎问题求解。*arXiv*预印本 *arXiv:2305.10601*，2023a。

Shunyu Yao、Jeffrey Zhao、Dian Yu、Nan Du、Izhak Shafran、Karthik Narasimhan 和 Yuan Cao。ReAct：在语言模型中协同推理与行动。*arXiv*预印本 *arXiv:2210.03629*，2023b。Shunyu Yao、Jeffrey Zhao、Dian Yu、Nan Du、Izhak Shafran、Karthik Narasimhan 和 Yuan Cao。ReAct：在语言模型中协同推理与行动，2023c。Shunyu Yao、Dian Yu、Jeffrey Zhao、Izhak Shafran、Tom Griffiths、Yuan Cao 和 Karthik Narasimhan。思维树：利用大语言模型进行审慎问题求解。载于*NeurIPS*，2024年。Joel Ye、Dhruv Batra、Abhishek Das 和 Erik Wijmans。辅助任务与探索助力物体目标导航。载于*ICCV*，2021年。Ruosong Ye 等人。语言是图所需的一切。*arXiv*预印本 *arXiv:2308.07134*，2024年。Sheng Yin、Xianghe Xiong、Wenhao Huang 等人。SafeAgentBench：具身大语言模型智能体安全任务规划基准测试。*arXiv*预印本 *arXiv:2412.13178*，2025年。Tianwei Yin、Xingyi Zhou 和 Philipp Krahenbuhl。基于中心的三维目标检测与跟踪。载于*CVPR*，第11784–11793页，2021年。Wei Yin、Chi Zhang、Hao Chen、Zhipeng Cai、Gang Yu、Kaixuan Wang、Xiaozhi Chen 和 Chunhua Shen。Metric3D：从单张图像实现零样本度量三维预测。*arXiv*预印本 *arXiv:2307.10984*，2023年。Chengxuan Ying 等人。Transformer在图表示方面的表现真的很差吗？神经信息处理系统进展，2021年。Naoki Yokoyama、Dhruv Batra 等人。VLFM：用于零样本语义导航的视觉-语言前沿地图。*arXiv*预印本 *arXiv:2312.03275*，2024 年。Haoxuan You 、 Haotian Zhang 、 Zhe Gan 、 Xianzhi Du 、 Bowen Zhang、Zirui Wang、Liangliang Cao、Shih-Fu Chang 和 Yinfei Yang。Ferret：在任意位置、任意粒度下指代并定位任何对象。*arXiv*预印本 *arXiv:2310.07704*，2023年。Bing Yu、Haoteng Yin 和 Zhanxing Zhu。时空图卷积网络：用于交通流预测的深度学习框架。载于第二十七届国际人工智能联合会议论文集，第3634–3640页，2018年。Chao Yu、Akash Velu、Eugene Vinitsky、Jiaxuan Gao、Yu Wang、Alexandre Baez、Claudio Fischetti、Maxim Egorov 和 Cathy Wu。近端策略优化在合作式多智能体博弈中的惊人效果。*NeurIPS*，35：24611–24624，2022a。Licheng Yu、Xinlei Chen、Georgia Gkioxari、Mohit Bansal、Tamara L Berg 和 Dhruv Batra。多目标具身问答。*CVPR*，2019年。Tianhe Yu、Deirdre Quillen、Zhanpeng He、Ryan Julian 等人。Meta-World：用于多任务与元强化学习的基准测试和评估。载于*CoRL*，2020年。Xumin Yu、Lulu Tang、Yongming Rao、Tiejun Huang、Jie Zhou 和 Jiwen Lu。Point-BERT：通过掩码点建模对三维点云Transformer进行预训练。*CVPR*，2022b。Zehao Yu、Anpei Chen、Binbin Huang、Torsten Sattler 和 Andreas Geiger。Mip-Splatting：无混叠三维高斯泼溅。*CVPR*，2024年。Weihao Yuan、Xiaodong Gu、Zuozhuo Dai、Siyu Zhu 和 Ping Tan。用于单目深度估计的神经窗口全连接条件随机场。*CVPR*，2022年。Yifu Yuan、Zhizheng Xiong 和 Qin Wang。多目标跟踪的深度学习：综述。*IET*计算机视觉，14(8)：601–612，2020a。

Yuan Yuan 等人。UniST：一种由提示赋能的城市时空预测通用模型。*ACM SIGKDD*知识发现与数据挖掘大会，2024年。Zhongqiang Yuan、Xiaobing Zhou 和 Tianbao Yang。城市交通异常检测算法综述。*IEEE Access*，2020b。Fabio Massimo Zanzotto。人在回路机器学习：现状综述。*Artificial IntelligenceReview*，56(4)：3301–3366，2019年。Rowan Zellers、Yonatan Bisk、Ali Farhadi 和 Yejin Choi。从识别到认知：视觉常识推理。*CVPR*，第6720–6731页，2019年。Andy Zeng、Pete Florence、Jonathan Tompson、Stefan Welker、Jonathan Chien 等人。Transporter网络：为机器人操作重新排列视觉世界。载于*CoRL*，2021年。Guibin Zhang、Haotian Ren、Chong Zhan、Zhenhong Zhou、Junhao Wang、He Zhu、Wangchunshu Zhou 和 Shuicheng Yan。MemEvolve：智能体记忆系统的元进化，2025a。Haotian Zhang、Haochen Zhang、Pengchuan Zhao、Haoxuan You、Yinfei Jiang、Jianwei Lee、Trevor Darrell 和 Jianfeng Gao。Ferret-v2：面向大语言模型指称与定位的改进基线。*arXiv*预印本 *arXiv:2404.07973*，2024a。Jiayan Zhang 等人。GraphInstruct：赋予大语言模型图理解与推理能力。*arXiv*预印本 *arXiv:2403.04483*，2024b。Junbo Zhang 等人。用于全市人流预测的深度时空残差网络。载于*AAAI*，2017年。Kaichen Zhang、Bo Li、Peiyuan Yan 等人。LMMS-Eval：大型多模态模型评估的现实检验。*arXiv*预印本 *arXiv:2407.12772*，2024c。Kaiqing Zhang、Zhuoran Yang 和 Tamer Ba¸sar。多智能体强化学习：理论与算法精选概述。《强化学习与控制手册》，第321–384页，2021a。Liangpei Zhang、Lefei Zhang 和 Bo Du。遥感数据深度学习：前沿技术教程。*IEEE*地球科学与遥感杂志，4(2)：22–40，2016年。Sixian Zhang、Xinhang Song、Yubing Bai、Weijie Li、Yakun Chu 和 Shuqiang Jiang。用于对象导航的分层对象到区域图。*ICCV*，2021b。Wei Zhang、Zheng Zhou、Zhen Zheng 等人。Open3DVQA：开放空间中使用多模态大语言模型进行全面空间推理的基准测试。*arXiv*预印本 *arXiv:2503.11094*，2025b。Yu Zhang 和 Qiang Yang。多任务学习综述。《*IEEE*知识与数据工程汇刊》，34(12)：5586–5609，2022年。Zhuosheng Zhang 等人。大语言模型中的自动思维链提示。*arXiv*预印本 *arXiv:2210.03493*，2023年。Ziwei Zhang、Peng Cui 和 Wenwu Zhu。图上的深度学习：综述。《*IEEE*知识与数据工程汇刊》，34(1)：249–270，2020年。Andrew Zhao 等人。Expel：大语言模型智能体是经验型学习者。*AAAI*，2024年。Hengshuang Zhao、Li Jiang、Jiaya Jia、Philip HS Torr 和 Vladlen Koltun。点云Transformer。2021年。Jianan Zhao 等人。GraphText：文本空间中的图推理。*arXiv*预印本 *arXiv:2310.01089*，2023a。Tony Z Zhao、Vikash Kumar、Sergey Levine 和 Chelsea Finn。使用低成本硬件学习精细双手操作。载于*RSS*，2023b。

Wenyu Zhao、Jorge Pena Queralta 和 Tomi Westerlund。面向机器人学的深度强化学习仿真到现实迁移：综述。载于*2020*年*IEEE*计算智能研讨会系列（*SSCI*），第737–744页。IEEE，2020年。Boyuan Zheng、Boyu Gou、Jihyung Kil、Huan Sun 和 Yu Su。若能落地，GPT-4V(ision)便是通用型网络智能体。2024年。Chuanpan Zheng、Xiaoliang Fan、Cheng Wang 和 Jianzhong Qi。GMAN：用于交通预测的图多重注意力网络。2020年。Yu Zheng。轨迹数据挖掘：概述。*ACM*智能系统与技术汇刊，2015年。Yu Zheng、Licia Capra、Ouri Wolfson 和 Hai Yang。城市计算：概念、方法论与应用。*ACM*智能系统与技术汇刊，5(3)：1–55，2014年。Chong Zhou、Chen Change Loy 和 Bo Dai。从CLIP中提取自由稠密标签。*ECCV*，第696–712页，2022年。Denny Zhou 等人。由简到繁提示助力大语言模型实现复杂推理。载于*ICLR*，2023a。Jie Zhou 等人。图神经网络：方法与应用综述。*AI Open*，2020年。Junsheng Zhou 等人。Uni3D：探索大规模统一三维表征。国际学习表征会议，2024年。Shuyan Zhou、Frank F Xu、Hao Zhu、Xuhui Zhou、Robert Lo、Abishek Sridhar、Xianyi Cheng、Yonatan Bisk、Daniel Fried、Uri Alon 等人。WebArena：用于构建自主智能体的真实网络环境。*arXiv*预印本 *arXiv:2307.13854*，2023b。Yin Zhou 和 Oncel Tuzel。VoxelNet：基于点云的三维目标检测端到端学习。载于*IEEE*计算机视觉与模式识别会议，第4490–4499页，2018年。Zhi-Hua Zhou。弱监督学习简要介绍。《国家科学评论》，2018年。Fengda Zhu、Yi Zhu、Xiaojun Chang 和 Xiaodan Liang。借助自监督辅助推理任务实现视觉语言导航。载于*CVPR*，第10012–10022页，2020年。Xiao Xiang Zhu、Devis Tuia、Lichao Mou、Gui-Song Xia、Liangpei Zhang、Feng Xu 和 Friedrich Fraun-dorfer。遥感中的深度学习：全面综述与资源列表。*IEEE*地球科学与遥感杂志，2017年。Xiaojin Jerry Zhu。半监督学习文献综述。威斯康星大学麦迪逊分校计算机科学系，2005年。Yixin Zhu、Yibiao Zhao 和 Song-Chun Zhu。理解工具：面向任务的对象建模、学习与识别。*CVPR*，第2855–2864页，2015年。Zihan Zhu 等人。Nice-SLAM：用于同步定位与地图构建的神经隐式可扩展编码。*IEEE*计算机视觉与模式识别会议，2022年。Ziyu Zhu、Xiaojian Ma、Yixin Chen、Zhidong Deng、Siyuan Huang 和 Qing Li。3D-VISTA：用于三维视觉与文本对齐的预训练Transformer。*ICCV*，2023年。Brianna Zitkovich、Tianhe Yu、Sichun Xu、Peng Xu、Ted Xiao、Fei Xia、Jialin Wu、Paul Wohlhart、Stefan Welker、Ayzaan Wahid 等人。RT-2：视觉-语言-动作模型将网络知识迁移至机器人控制。*CoRL*，2023年。