具身智能入门指南 Embodied-AI-Guide

> Embodied AI(具身智能)入门的路径以及高质量信息的总结, 期望是按照路线走完后, 新手可以快速建立关于这个领域的认知, 希望能帮助到各位入门具身智能的朋友, 欢迎点Star、分享与提PR🌟 ~
>
> 【Embodied-AI-Guide, Latest Update: Mar. 15, 2025】 
>
> 

VITA具身智能社区 (筹备中)

Embodied-AI-Guide项目很快将会以网页版wiki的形式上传到VITA具身智能社区网站，敬请期待。如果你对合作构建VITA具身社区感兴趣（目前更倾向于机构、社区间合作），欢迎邮件联系Vita.Committee2025@gmail.com或联创微信TianxingChen_2002（请备注机构+姓名与来意）

Contents - 目录

1. Start From Here - 从这里开始

> 具身智能是指一种基于物理身体进行感知和行动的智能系统, 其通过智能体与环境的交互获取信息、理解问题、做出决策并实现行动, 从而产生智能行为和适应性。

How - 如何学习这份指南

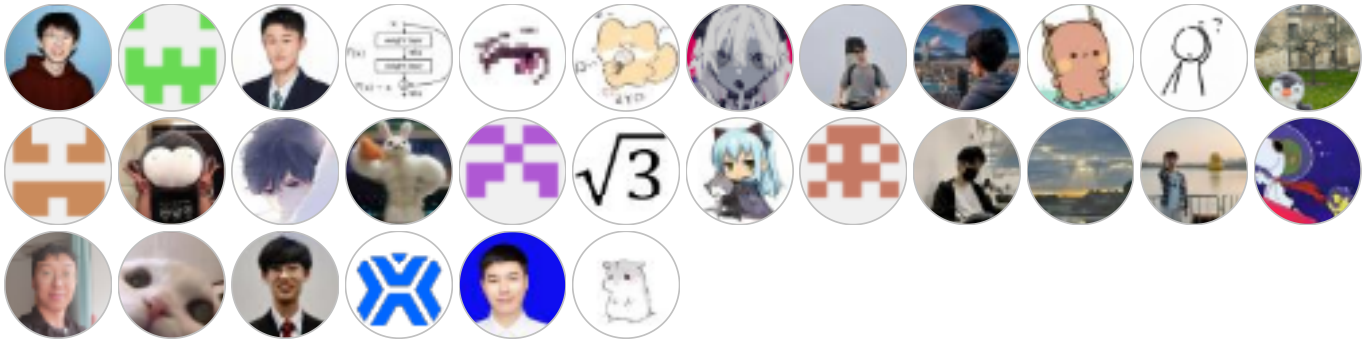
我们希望的是帮助新人快速建立领域认知, 所以设计理念是：**简要**介绍目前具身智能涉及到的主要技术, 让大家知道不同的技术能够解决什么问题, 未来想要深入发展的时候能够有头绪。

About us - 关于我们

我们是一个由具身初学者组成的团队, 希望能够通过我们自己的学习经验, 为后来者提供一些帮助, 加快具身智能的普及。欢迎更多朋友加入我们的项目, 也很欢迎交友、学术合作, 有任何问题, 可以联系邮箱 chentianxing2002@gmail.com。

🦉 **Contributors**: 陈天行 (深大BS), 王开炫 (25' 港大PhD), 贾越如 (北大Ms), 姚天亮 (25' 港中文PhD), 高焕昂 (清华PhD), 高宁 (西交BS), 郭常青 (清华Ms), 彭时佳 (深大BS), 邹誉德 (25' 上交AILab联培PhD), 陈思翔 (25' 北大PhD), 朱宇飞 (25' 上科大Ms), 韩翊飞 (清华Ms), 王文灏 (宾大Ms), 李卓恒 (港大PhD), 邱一航 (港大PhD), 梁升一 (港科广PhD), 林俊晓 (浙大Ms), 王冠锟 (港中文PhD), 吴志杰 (港中文PhD), 叶雯 (25' 中科院PhD), 陈

攒鑫 (深大BS), 侯博涵 (山大BS), 江恒乐 (25' 南科大PhD), 陈勇超 (MIT+哈佛PhD), 胡梦康 (港大PhD), 梁志烜 (港大PhD), 穆尧 (上交AP).



2. Useful Info - 有利于搭建认知的资料

- 具身智能基础技术路线-YunlongDong [2]: PDF, bilibili
- 社交媒体:
  - 可以关注的公众号: **石麻日记 (超高质量!!!)**, 机器之心, 新智元, 量子位, Xbot具身知识库, 具身智能之心, 自动驾驶之心, 3D视觉工坊, 将门创投, RLCN强化学习研究, CVHub
  - AI领域值得关注的博主列表 [3]: zhihu
- Robotics实验室总结 [4]: zhihu_1, zhihu_2
- 具身智能会投稿的较高质量会议与期刊：Science Robotics, TRO, IJRR, JFR, RSS, IROS, ICRA, ICCV, ECCV, ICML, CVPR, NIPS, ICLR, AAAI, ACL等。
- 斯坦福机器人学导论：website
- 共建全网最全具身智能知识库 [6]: website
- Awesome-Embodied-AI-Job (具身智能招贤榜): Repo
- 社区:
  - DeepTimber Robotics Innovations Community, 深木科研交流社区: website
  - 宇树具身智能社群: website
  - Simulately: Handy information and resources for physics simulators for robot learning research: website
  - DeepTimber-地瓜机器人社区: website
  - HuggingFace LeRobot (Europe, check the Discord): website
  - K-scale labs (US, check the Discord): website

3. Algorithm - 算法

3.1 Common Tools - 常用工具

> 这个部分是关于具身中常用技巧的分享

- 点云降采样: zhihu, 包括随机降采样、均匀降采样、最远点降采样、法线空间降采样等, 需要了解清楚每一种降采样的优劣, 这个技巧的选择对于3D应用来说是至关重要的。
- 手眼标定：github, 手眼标定用于确定相机和机械臂之间以及相机与相机之间的相对位置, 大部分Project的开始都需要做一次手眼标定, 分为眼在手上和眼在手外。

3.2 Vision Foundation Models - 视觉基础模型

> 以下是部分具身智能中常用的基础模型, 计算机视觉中发展的非常好的工具可以直接赋能具身智能的下游应用。

- CLIP: website, 来自OpenAI的研究, 最基本的应用是可以计算图像与语言描述的相似度, 中间层的视觉特征对各种下游应用非常有帮助。
- DINO: DINO repo, DINO-v2 repo, 来自Meta的研究, 可以提供图像的高层视觉特征, 对corresponding之类的信息提取非常有帮助, 比如不同个体之间的鼻子都有类似的几何特征, 这个时候不同图像中关于不同鼻子的视觉特征值可能是近似的。
- SAM: website, 来自Meta的研究, 可以基于提示点或者框, 对图像的物体进行分割。
- SAM2: website, 来自Meta的研究, SAM的升级版, 可以在视频层面持续对物体进行分割追踪。
- Grounding-DINO: repo, 在线尝试, **这个DINO与上面Meta的DINO没有关系**, 是一个由IDEA研究院(做了很多不错开源项目的机构)开发集成的图像目标检测的框架, 很多时候需要对目标物体进行检测的时候可以考虑使用。
- Grounded-SAM: repo, 比Grounding-DINO多了一个分割功能, 也就是支持检测后分割, 也有很多下游应用, 具体可以翻一下README。
- FoundationPose: website, 来自Nvidia的研究, 物体姿态追踪模型。
- Stable Diffusion: repo, website, 22年的文生图模型, 现在虽然不是SOTA了, 但是依然可以作为不错的应用, 例如中间层特征支持下游应用、生成Goal Image (目标状态) 等等。
- Depth Anything (v1 & v2): repo, repo, 港大和字节的研究工作, 单目深度估计模型。
- Point Transformer (v3): repo, 点云特征提取的工作。
- RDT-1B: website, 清华朱军老师团队的工作, 机器人双臂操作的基础模型, 具有强大的few-shot能力。
- SigLIP: huggingface, 类似CLIP。

3.3 Robot Learning - 机器人学习

机器人学习 Robot Learning 的发展: zhihu

3.3.1 Model Predictive Control (MPC) - 模型预测控制

3.3.2 Reinforcement Learning - 强化学习

- 强化学习的数学原理 - 西湖大学赵世钰: bilibili GitHub 这门课程作为强化学习的入门课程非常合适，适合只对机器学习略有了解，但没有了解过强化学习的初学者，可以了解强化学习的数学原理，其教材编写也十分用心。

Deep Reinforcement Learning - 深度强化学习

下面列出三门比较受欢迎的深度强化学习相关的课程，这几门课互有overlap，时间长短和授课风格也各有不同，读者可以选择适合自己的课程进行学习。此外，深度强化学习的经典算法相关的文章也在必读清单：如PPO, SAC, TRPO, A3C等。

- The Foundations of Deep RL in 6 Lectures YouTube 本门在线课程由在RL领域著名的Pieter Abbeel教授主讲，从MDP开始在六节课之内介绍了深度强化学习的主要知识。
- UC Berkeley CS285 深度强化学习: website | YouTube 本课程的主讲老师是在RL领域著名的Berkeley的Sergey Levine教授，DRL领域许多著名的工作如SAC就出自他之手。Sergey在授课方面非常用心，本课程对DRL提供了非常详细的介绍。
- 李宏毅老师也有一套关于强化学习的课程: bilibili上课+刷蘑菇书巩固+gymnasium动手实践, 重点了解一下PPO。
    - 台湾大学李宏毅公开课: bilibili
    - EasyRL - 蘑菇书: website, 基本是配套李宏毅老师的课程
    - 实践gymnasium, 可以尝试一下把玩一下登月着陆等经典强化学习场景, 思考+动手, 观察阶段agent的表现并分析, 有助于深入理解强化学习

然而，深度强化学习的Reward Tuning和参数调整非常依赖于经验，建议读者在对深度强化学习有相关经验之后，可以自己尝试训练一个policy并在机器人上部署，体会其中的Sim-to-Real Gap。常用的仿真平台有MuJoCo PlayGround, Isaac Lab, SAPIEN, Genesis等。

常用的Codebase有legged-gym(由ETH RSL开发，基于IsaacGym)等，也可以根据你想做的任务找到相近的codebase。

3.3.3 Imitation Learning - 模仿学习

- 《模仿学习简洁教程》 - 南京大学LAMDA: PDF
- Supervised Policy Learning for Real Robots, RSS 2024 Workshop 教程：真实机器人的监督策略学习, bilibili

3.4 LLM for Robotics - 大语言模型在机器人学中的应用

为了促使机器人更好的规划, 现代具身智能工作常常利用大语言模型强大的信息处理能力与泛化能力进行规划。

- Robotics+LLM系列通过大语言模型控制机器人 [2]: zhihu
- Embodied Agent wiki: website
- Lilian Weng 个人博客 - AI Agent 系统综述 [5]: 中文: website 英文: website
- 过去一系列工作常常仅使用LLM作为High-Level的策略生成器 用作High-Level 规划
    - 经典工作(1) PaLM-E: Arxiv
    - 经典工作(2) DO AS I CAN, NOT AS I SAY: Arxiv
    - 经典工作(3) Look Before You Leap: Arxiv
    - 经典工作(4) EmbodiedGPT: Arxiv
- 同时也有一些工作将High-Level的策略规划与 Low-Level的动作生成进行统一
    - 经典工作(1) RT-2: Arxiv

- 另一个代表性方向的工作将LLM与传统基于算法的Planner结合来做任务与移动规划
  - 经典工作(1) LLM+P: Arxiv
  - 经典工作(2) AutoTAMP: Arxiv
  - 经典工作(3) Text2Motion: Arxiv
- 利用LLM的code能力实现具身智能是一个有趣的想法
  - 经典工作(1) Code as Policy: Arxiv
  - 经典工作(2) Instruction2Act: Arxiv
- 有一些工作将三维视觉感知同LLM结合起来，共同促进具身智能规划
  - VoxPoser Arxiv
  - OmniManip Arxiv
- 还有一些工作试图把基于LLM的机器人规划扩展到多机器人协同合作的场景
  - 经典工作(1) RoCo: Arxiv
  - 经典工作(2) Scalable-Multi-Robot: Arxiv

3.5 Vision-Language-Action Models - VLA模型

Vision-Language-Action Models(VLA模型) 是一种结合VLM(Vision-Language Model)与机器人控制的模型，旨在将预训练的VLM直接用于生成机器人动作(RT-2中定义)。和以往利用VLM做planning以及build from strach的方法不同，VLA无需重新设计新的架构，将动作转化为token，微调VLM。

VLA的特点：端到端，使用LLM/VLM backbone，加载预训练模型, etc.

目前的VLA可以从以下几个方面进行区分：模型结构&大小(如action head的设计, tokenize的方法如FAST)，预训练与微调策略和数据集，输入和输出(2D vs. 3D | TraceVLA输入visual trace)，不同的应用场景等。

参考资料：

- Blog: 具身智能Vision-Language-Action的思考, zhihu
- Survey: A Survey on Vision-Language-Action Models for Embodied AI, 2024.11.28

经典工作：

- **Autoregressive Models**
  - **RT系列(Robotic Transformers)**:
    - **RT-1** (paper)
    - **RT-2** (page | paper, Google Deepmind, 2023.7)： 55B
    - **RT-Trajectory** (paper, Google Deepmind, UCSD, 斯坦福 2023.11)
    - **AUTORT** (paper, Google Deepmind, 2024.1)
  - **RoboFlamingo** (paper | code, 字节、清华, 2024.2)
  - **OpenVLA** (paper | code, OpenAI, 2024.6): 7B
  - **TinyVLA** (paper, 上海大学, 2024.11)
  - **TraceVLA** (paper | code, 微软，2024.12)
- **Diffusion Models for Action Head:**

    - **Octo** (paper | code, 斯坦福，伯克利, 2024.5): Octo-base (93M)
    - **π0** (paper | code, 斯坦福, physical intelligence,) : 3.3B; flow-based diffusion VLA; PaliGemma (3B VLM);
    - **CogACT** (paper | code, 清华，MSRA, 2024.11): 7B
    - **Diffusion-VLA** (paper | code, 华东师范，上海大学，美的, 2024.12)
- **3D Vision:**
    - **3D-VLA** (paper | code, UMass, 2024.3): 3D-based LLM
    - **SpatialVLA** (paper | code , 上海AI Lab, 2025.1): Adaptive Action Grid
- **VLA-related:**
    - **FAST (π0)** (paper, code, 斯坦福，伯克利, physical intelligence, 2025.1): autoregressive VLA
    - **RLDG** (paper | code, 伯克利, 2024.12): 强化学习(RL)生成高质量的训练数据进行微调
    - **BYO-VLA** (paper | code, 普林斯顿大学, 2024.10): 运行时图像干预，有效降低VLA模型对任务无关视觉干扰的敏感度
- **Different Locomotion:**
    - **RDT-1B (双臂)** (paper | code, 清华): 双臂控制的扩散模型
    - **QUAR-VLA (四足机器人)** (paper, 西湖大学，浙江大学, 2025.2.4)
    - **CoVLA (自动驾驶)** (paper | page, Turing, 2024.12)
    - **Mobility-VLA (导航)** (paper, Google Deepmind, 2024.7)
    - **NaVILA (腿式机器人导航)** (paper | code, USCD, 2024.12)

3.6 Computer Vision - 计算机视觉

CS231n (斯坦福计算机视觉课程): website, 该课程对深度学习在计算机视觉的应用有较为全面的介绍。因为已经在具体实现某个论文的算法了, 所以这个阶段可以不用做作业, 只需要看课程视频和课程讲义即可。

3.6.1 2D Vision - 二维视觉

- 2D Vision 领域的经典代表作
    - CNN (卷积神经网络): link
    - ResNet (深度残差网络): bilibili
    - ViT (第一个将Transformer用在视觉领域): bilibili
    - Swin Transformer (披着Transformer皮的CNN): bilibili
    - 对比学习论文综述: bilibili
- 以判别式模型为主的感知任务, 比如识别、分类、分割、检测等等, 看看即可, 现在继续刷点意义不大
- 生成式模型
    - 自回归综述: PDF
    - 扩散模型综述: PDF

    - 如果对扩散模型的理论推导感兴趣, 可以看苏剑林老师的博客 - 生成扩散模型漫谈(推导非常清楚): link

3.6.2 3D Vision - 三维视觉

- 三维视觉导论 - Andreas Geiger: website (重点关注课程作业)
- GAMES203 - 三维重建和理解: bilibili
- 三维生成的一些经典论文:
    - Diffusion Model for 2D/3D Generation 相关论文分类: link
    - 3D生成相关论文-2024: link

3.6.3 4D Vision - 四维视觉

- 视频理解
    - 开山之作: bilibili
    - 论文串讲: bilibili
    - LLM时代的视频理解综述: PDF
- 4D 生成
    - 视频生成博客(英文): link
    - 4D 生成的论文列表: website

3.6.4 Visual Prompting - 视觉提示

视觉提示是一种利用视觉输入引导大模型完成特定任务的方法，常用于具身智能领域。它通过提供示例图像、标注或视觉线索，让模型理解任务要求，而无需额外训练。例如，在机器人导航、操控等场景中，视觉提示可帮助模型适应新环境，提高泛化能力。相比传统方法，视觉提示具备更强的灵活性和可扩展性，使具身智能系统能够通过视觉信息快速适应复杂任务。

- 视觉提示综述：paper
- **PIVOT**, page: 通过将任务转化为迭代式视觉问答，实现在无需特定任务数据微调的情况下，zero-shot 控制机器人系统和进行空间推理。
- **Set-of-Mark Visual Prompting for GPT-4V**: page

3.7 Computer Graphics - 计算机图形学

如果说计算机视觉是考虑图像之间的变化以及从图像到三维模型(三维重建和生成)，那么计算机图形学主要研究的就是三维模型之间的变化以及从三维模型到图像的渲染过程。具身智能在开发和测试的时候离不开仿真器，而仿真也属于图形学的研究范畴。快速、高质量的渲染，并行化、准确的仿真一直是机器人仿真器追求的目标，而这一切通过计算机图形学来实现。

- 如果对传统图形学感兴趣, 可以看下面两门(闫令琪老师开的课, 讲得特别好):
    - **GAMES101 - 现代计算机图形学入门**: website
    - GAMES202 - 高质量实时渲染: website
- 如果对motion synthesis/computer animation感兴趣, 可以看:
    - GAMES105 - 计算机角色动画基础: website
- 如果对三维重建感兴趣, 可以看下面两门:
    - Nerf原理代码讲解: bilibili
    - 3DGS原理代码讲解: bilibili
- 三维预训练最新综述:

    - Advances in 3D pre-training and downstream tasks: a survey: PDF
- 3DGS在具身上的综述:
    - 3D Gaussian Splatting in Robotics: A Survey: PDF

3.8 Multimodal Models - 多模态模型

> 多模态旨在统一来自不同模态信息的表征, 在具身智能中由于面对着机器识别的视觉信息与人类自然语言的引导信息等不同模态的信息，多模态技术愈发重要。

- 最经典的工作CLIP: 知乎
- 多模态大语言模型的经典工作 LLaVA: website
- 多模态生成模型综述: pdf

3.9 Robot Navigation - 机器人巡航

机器人巡航（Robot Navigation）是一类要求智能体在未知场景中，通过获取并处理环境信息，实现达成某种目标的路径规划。机器人巡航是具身任务中的一个重要能力，是完成复杂任务不可缺少的基础技术。机器人巡航任务中，智能体一般接受传感器提供的RGB、深度、GPS等信息和相关目标指令，输出是一系列的动作指令。

按照任务类型分类，机器人巡航可以分为以下几个部分：

- **物体目标巡航（Object-Goal Navigation）**：最常见和最广泛的巡航任务。智能体接受的指令是对一个特定物体的描述，目标是找寻到这个物体。
- **图像目标巡航（Image-Goal Navigation）**：智能体接受的指令是一个图像，目标是找寻到这个图像所描述的场景。
- **视觉-语言巡航（Vision-Language Navigation，VLN）**：智能体接受的指令是一个自然语言指令描述，目标是跟随该指令行进。

按照模型架构分类，机器人巡航可以分为以下几个类别：

- **端到端模型（End-to-End Model）**：模型直接将传感器输入通过强化学习或模仿学习映射到动作指令。模型会先将传感器信息编码为视觉表征，结合历史动作作为输入，最后通过与环境交互获得reward实现动作决策的学习。端到端模型主要针对两方面进行优化：一是提升视觉表征能力，二是解决稀疏奖励等动作决策方面的问题。端到端模型的优势在于直截了当，但是面临着严重的过拟合和低泛化性问题，使得其在现实生活中的应用收到了挑战。
    - 经典工作：
        - Learning Object Relation Graph and Tentative Policy for Visual Navigation
        - VTNet: Visual Transformer Network for Object Goal Navigation
- **模块化模型（Modular Model）**：将传感器信息输入不同的模块，模块之间通过接口交互，输出动作指令。模块包括建图模块（Mapping，构建语义和占有地图），长期决策模块（Global Policy，决定长期的导航目标），短期决策模块（Local Policy，决定实现长期目标的具体操作）等。建图模块是模型的核心，包含有网格地图、包含预测的网格地图、图表示地图等多种形式。模块化模型的优势在于模块之间的解耦，大大加强了模型的可解释性。同时，独立的建图模块也使得模型更容易泛化到未知环境。但是模块化模型的建图模块仍然充斥着手动设计的规则，这一定程度上也限制了模型的通用性。
    - 经典工作：

    - Object Goal Navigation using Goal-Oriented Semantic Exploration： SemExp，最早提出语义地图的概念，学习区域和物体之间关联的语义先验，使智能体能够更好地判断目标物体可能在的方向。
    - PONI: Potential Functions for ObjectGoal Navigation with Interaction-free Learning：PONI，提出了基于potential functions的语义地图预测，即基于已有的语义地图学习"补全"的完整地图，想象物体最可能在整个房间的哪个位置，使智能体能够迁移在其他样本中观察到的知识。
    - 3D-Aware Object Goal Navigation via Simultaneous Exploration and Identification：把3D信息编码进巡航的经典工作，通过更精细的点云分割信息，避免了2D语义图在z轴上的信息损失，实现了更精确的语义地图构建。

- **零样本模型（Zero-shot Model）**：模型不接触训练数据，直接在测试阶段完成任务。零样本模型往往利用具有知识先验的大规模预训练模型（CLIP, LLM等）实现。零样本模型的提出旨在解决基于学习的方法面临的过拟合和低泛化性问题，同时也更适合迁移到现实场景。但是零样本模型的缺陷在于推理速度较慢，且性能受限，需要进一步微调以实现更好的性能。

    - 经典工作：

        - CoWs on Pasture: Baselines and Benchmarks for Language-Driven Zero-Shot Object Navigation：开放语义物体巡航的提出工作。思路很简单：用CLIP寻找目标物体，找到了就走过去。在不常见物体、复杂描述上取得了很好的效果，同时也有着对不同属性的同类别物体的区分能力。
        - L3MVN: Leveraging Large Language Models for Visual Target Navigation：利用LLM决定"我要向哪个边界前进"。利用LLM的人类知识先验，判断物体可能在的房间，以及与其他物体之间的相关关系，实现更快速更有效的巡航。
        - ESC: Exploration with Soft Commonsense Constraints for Zero-shot Object Navigation：显式提出了区域对于巡航的影响，在语义地图上标注出区域占有的位置，作为输入的一部分输入给LLM。结合了语义地图连续性和LLM知识丰富的优势。
        - SG-Nav: Online 3D Scene Graph Prompting for LLM-based Zero-shot Object Navigation：在线构建多层场景图(Scene Graph)并输入给LLM，利用CoT实现LLM对于物体位置的推理。

常用数据集：

- MatterPort3D(MP3D)：真实场景采集，场景复杂庞大，数据量大，难度高。
- Habitat-Matterport3D(HM3D)：同上
- RoboTHOR：仿真环境，场景小简单。

其他参考：

- 物体目标巡航综述
- awesome vision-language navigation
- Habitat Navigation Challenge(Habitat框架中整合了非常多常见的agent skill，例如语义地图构建，FBE和一些heuristic方法，非常适合模块化方法的开发)

3.10 Embodied AI for X - 具身智能+X

3.10.1 EAI for Healthcare - 具身医疗

> 具身智能技术的迅猛发展正在引领医疗服务模式迈向革命性的新纪元。作为人工智能算法、先进机器人技术与生物医学深度融合的前沿交叉学科, 具身智能+医疗这一研究领域不仅突破了传统医疗的边界, 更开创了智能化医疗的新范式。其多学科协同创新的特质, 正在重塑医疗服务的全流程, 为精准医疗、远程诊疗和个性化健康管理带来前所未有的发展机遇, 推动医疗行业向更智能、更人性化的方向转型升级。这一领域的突破性进展, 标志着医疗科技正迈向一个全新的智能化时代。

- 医疗具身智能综述: A Survey of Embodied AI in Healthcare: Techniques, Applications, and Opportunities

3.10.1.1 MLLM for Medical - 多模态大语言模型在医学中的应用

- 用于医学影像分析的通用人工智能综述: website
- 医学影像的通用分割模型-MedSAM： website
- 2024盘点：医学AI大模型, 从通用视觉到医疗影像: NEJM医学前沿
- 医疗领域基础模型的发展机遇与挑战: website
- SkinGPT-4 for dermatological diagnosis: website
- PneumoLLM for pneumoconiosis diagnosis: website
- BiomedGPT: website
- LLaVA-Med: website
- RoboNurse-VLA: website
- PathChat 哈佛医学院Faisal Mahmood教授团队的病理大模型。临床上, 病理被称为诊断的金标准: website
- DeepDR-LLM 糖尿病视网膜病变 (DR)的专科垂域多模态大模型: website
- VisionFM 通用眼科人工智能的多模式多任务视觉基础模型: website
- Medical-CXR-VQA 用于医学视觉问答任务的大规模胸部 X 光数据集: website

3.10.1.2 Medical Robotics - 医疗机器人

- 医疗机器人的五级自动化(医疗机器人领域行业共识), 杨广中教授于2017年在Science Robotics上的论著: Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy
- 医疗机器人的十年回顾(含医疗机器人的不同分类), 杨广中教授在Science Robotics上的综述文章： A decade retrospective of medical robotics research from 2010 to 2020
- 医疗具身智能的分级: A Survey of Embodied AI in Healthcare: Techniques, Applications, and Opportunities
- Artificial intelligence meets medical robotics, 2023年发表在Science正刊上的论著: website
- 医疗机器人的机器视觉
    - 3DGS在腔镜手术中的应用综述: website
- 达芬奇手术机器人是最为常用的外科手术机器人, 对于这类机器人自主技能操作的研究最为广泛
    - 达芬奇手术机器人研究套件dVRK介绍: website
    - 通过模仿学习在达芬奇机器人上学习外科手术操作任务 Surgical Robot Transformer (SRT): website
    - Domain-specific Simulators - 手术机器人技能学习领域的模拟器

    - SurRoL: RL-Centered and dVRK Compatible Platform for Surgical Robot Learning website
    - Surgical Gym: A high-performance GPU-based platform for surgical robot learning (ICRA 2024, work in progress, based on NVIDIA Omniverse): website
    - ORBIT-Surgical: An Open-Simulation Framework for Learning Surgical Augmented Dexterity (ICRA 2024, based on NVIDIA Omniverse): website
    - 缝合是手术机器人操作中的一个关键子任务，实现其自主化已有多项研究。关于自主缝合技能操作的综述可参考: website

- 连续体和软体手术机器人作为柔性医疗机器人的重要分支, 凭借其独特的结构设计和材料特性, 在微创介入诊疗领域展现出显著优势。它们能够灵活进入人体狭窄腔体, 实现精准操作, 同时最大限度地减小手术创口, 降低患者术后恢复时间及感染风险, 为现代微创手术提供了创新性的技术解决方案。

    - 连续体机器人在医疗领域的应用 (Nabil Simaan; Howie Choset等): Continuum Robots for Medical Interventions
    - 软体手术机器人在微创介入手术中的应用 (Ka-wai Kwok; Kaspar Althoefer等)： Soft Robot-Assisted Minimally Invasive Surgery and Interventions: Advances and Outlook

- 连续体和软体机器人因其超冗余自由度和高度非线性的结构特性, 采用传统的控制与传感方法构建正逆运动学方程时面临显著的计算复杂性和建模局限性。传统方法难以精确描述其多自由度耦合运动及环境交互中的动态响应。为此, 基于数据驱动的智能控制方法(如深度学习、强化学习及自适应控制算法)成为解决这一问题的前沿方向。这些方法能够通过大量数据训练, 高效学习系统的非线性映射关系, 显著提升运动控制的精度、自适应性和鲁棒性, 为复杂医疗场景下的机器人操作提供了更为可靠的技术支撑。

    - 什么是软体机器人？软体机器人的具身智能定义: 知乎, by Ke WU from MBUZAI
    - IROS 2024大会Program Chair新加坡国立大学Cecilia Laschi教授的论著: Learning-Based Control Strategies for Soft Robots: Theory, Achievements, and Future Challenges
    - 软体机器人中具身智能物理建模简明指南(也是出自NUS Cecilia教授团队): A concise guide to modelling the physics of embodied intelligence in soft robotics
    - 数据驱动方法在软体机器人建模与控制中的应用: Data-driven methods applied to soft robot modeling and control: A review

- 微纳机器人技术是一类集成了微纳米制造、生物工程和智能控制等多学科前沿技术的微型机器人系统。凭借其微纳米级的独特尺寸、优异的生物相容性和精准的操控性能，这一前沿技术为现代医学诊疗范式带来了突破性创新。在精准诊断方面，微纳机器人能够深入人体微观环境，实现细胞乃至分子水平的实时监测；在靶向治疗领域，其可作为智能药物载体，实现病灶部位的精准定位与可控释放；在微创手术应用中，微纳机器人系统为复杂外科手术提供了前所未有的精确操作平台。这些创新性应用不仅显著提升了诊疗效率，更为攻克重大疾病提供了全新的技术途径，推动着现代医学向更精准、更微创、更智能的方向发展。

    - 微纳机器人的机器学习(CUHK 张立教授团队在Nature Machine Intelligence上的论著): Machine learning for micro- and nanorobots

3.10.2 UAV - 无人机

无人机的发展来源于：

1. 从外部传感设备保护发展至机载传感与计算；

2. 从遥控/预先编程发展至自主。

不同于legged locomotion和manipulation，在无人机领域，data-driven的方法与model-based/modular的方法在不同任务中的优势不同，仍处于分庭抗礼的阶段。这主要是因为无人机的模型与驱动模式较为简单(如四旋翼的驱动机构只有四个电机)，且传统的无人机(即不具有操作设备)不会与环境产生交互，因此基于模型、优化和分层的方法，通过良好的状态机/规则设计和高效的局部优化技术，仍能够被赋予很强的性能。然而，无人机的难点在于其状态估计(通常需要)、感知和底层驱动充满噪声，这是因为小型化无人机的负载能力十分有限以及其成本被尽可能压低，因此在一些任务中data-driven/端到端的方法展现出了远超于传统方法的性能。因此，以下对无人机data-driven资料介绍的同时会穿插其与传统方法的对比，以便大家了解整个领域发展的动机。

总体而言，无人机的研究分为三个部分：

1. 技能实现/学习，例如避障、竞速、大机动飞行/特技等；
2. 任务实现/学习，例如探索、重建、跟踪等；
3. 飞行机器人本体设计。

无人机工作的开源代码并不多且良莠不齐，大部分需要通过论文学习。

3.10.2.1 技能实现/学习

- **支持RL的仿真器**

  无人机的仿真器普遍并不强大，并且几乎没有开源的RL sim2real项目。基于开源代码需要较大的内容改动才能实现理想的sim2real performance。

  - **AirSim** (https://microsoft.github.io/AirSim/)：基于UE4引擎，具有较为逼真动力学transition模拟。缺点是UE4底层功能较难修改并且运行速度较慢。
  - **Flightmare** (https://github.com/uzh-rpg/flightmare)：基于Unity渲染，CPU并行动力学。
  - **AerialGym** (https://github.com/ntnu-arl/aerial_gym_simulator)：基于IsaacSim，GPU并行动力学。

- **经典技能代表性工作**

  我们主要介绍一些data-driven方法在经典任务上的应用。值得一提的是，以下的工作中，出现了一些摆脱了对SLAM系统和里程计依赖的方法(而无人机最初的兴起正是依靠SLAM/里程计系统的日益成熟)，将成为无人机技能学习中有趣的进展方向。

  - **未知场景障碍物躲避**

    - Learning Monocular Reactive UAV Control in Cluttered Natural Environments. ICRA 2013, CMU. 受自动驾驶发展启发，第一个使用监督学习将图像映射为离散上游控制指令的系统。
    - CAD2RL: Real Single-Image Flight without a Single Real Image. RSS 2017，UCB. 第一个使用sim2real RL，对单目RGB图像进行大量domain randomization，在长廊中输出速度指令的系统。
    - DroNet: Learning to Fly by Driving. RAL 2018, UZH. 利用自动假设数据集让飞机输出速度指令，代码开源(https://github.com/uzh-rpg/rpg_public_dronet)。
    - Learning High-Speed Flight in the Wild. SciRob 2021, UZH. 使用dagger利用传统轨迹规划进行监督学习。文章claim网络推理的低延迟可以使未知环境中飞行速度更快。代码开

源(https://github.com/uzh-rpg/agile_autonomy)。
    - Back to Newton's Laws: Learning Vision-based Agile Flight via Differentiable Physics, Arxiv 2024, SJTU. 用differentiable physics提供的一阶梯度做策略优化，不需要显式的位置和速度估计。文章用低分辨率深度图，训练避障比RL更高效，实现高速飞行。
    - Flying on Point Clouds using Reinforcement Learning [Video].Arxiv 2025, ZJU. 使用机载雷达和sim2real RL实现自主避障。
    - 值得一提的是，作为无人机最常用的任务，避障现在最常用的还是传统方法的系统如开源的ego-planner(https://github.com/ZJU-FAST-Lab/ego-planner)，由于这样的方案已经足以胜任大部分场景(而不像四足的MPC)，因此在实际应用中比较少使用data-driven的方案。

  - **无人机竞速**

    - Champion-level drone racing using deep reinforcement learning. Nature 23, UZH. 用强化学习战胜人类冠军飞手, 近几年无人机领域影响力最高的文章，是UZH RPG实验室多年来深厚工程积累的结果，其中的RL方案较为简单直接。
    - Reaching the Limit in Autonomous Racing: Optimal Control versus Reinforcement Learning. SciRob 23, UZH. 强化学习与最优控制方法竞速飞行对比。
    - Demonstrating Agile Flight from Pixels without State Estimation. RSS 2024, UZH. 使用视觉，不需要显式状态估计的现实世界竞速demo。
    - UZH的Perception and Robotics Group (RPG) 使用最优控制和RL的方法在竞速上有诸多尝试，使得无人机在固定轨道上达到最快飞行速度。

  - **大机动/特技飞行**

    - Deep Drone Acrobatics. RSS 2020, UZH. 使用模仿学习，从视觉特征点中学习MPC的轨迹跟踪，实现姿态剧烈变化的特技飞行。
    - Whole-Body Control Through Narrow Gaps From Pixels to Action. ICRA 2025, ZJU. 使用强化学习实现视觉端到端窄缝穿越，不需要显式的位置和速度估计，超越传统方法性能。

- **经典任务实现代表性工作**

  - **追捕**
    - HOLA-Drone: Hypergraphic Open-ended Learning for Zero-Shot Multi-Drone Cooperative Pursuit. Arxiv 2024, University of Manchester.
    - Multi-UAV Pursuit-Evasion with Online Planning in Unknown Environments by Deep Reinforcement Learning. Arxiv 2024, THU.
  - **探索**
    - Deep Reinforcement Learning-based Large-scale Robot Exploration, Arxiv2024, National University of Singapore (NUS). 利用注意力机制学习不同空间尺度的依赖关系，对未知区域进行隐式预测，优化已知空间探索策略，提高探索效率。
    - ARiADNE: A Reinforcement learning approach using Attention-based Deep Networks for Exploration, Arxiv2023, National University of Singapore (NUS). 学习已知不同区域在多个空间尺度上的相互依赖关系，并隐式预测探索这些区域可能获得的潜在收益。这使得代理能够安排行动顺序，以平衡在已知区域对地图进行开发/细化与探索新区域之间的自然权衡。
    - DARE: Diffusion Policy for Autonomous Robot Exploration. Arxiv2024, National University of Singapore (NUS). DARE方法利用self-attention学习地图空间信息，并通过diffusion生

成通往未知区域的轨迹，以提高自主机器人的探索效率。

3.10.2.2 无人机硬件平台搭建

手搓一个遥控器操控的穿越机不是一个很难的事情，网上有很多爱好者分享教程。但想搭建一个具有自主导航功能的无人机并非易事，是一个系统工程，这里推荐浙大FAST-lab开源的教程：

- 从0制作自主空中机器人

3.10.2.3 新构型无人机设计

除了常规用于航拍，环境探索的四旋翼无人机，想让无人机具备更多能力，应用于更广泛的具身智能场景，除了算法上的创新外，也需要在硬件层面对无人机的构型进行创新设计。

- **空中机械臂(Aerial Manipulator)**

  空中机械臂，也叫空中操作无人机，兼具无人机的快速空间移动能力和机械臂的精确操纵能力，是具身智能的一种理想载体。西湖大学赵世钰老师组在知乎上有一系列文章介绍：

  - 空中作业机器人，下一代无人机技术?
  - 空中作业机器人—没那么简单!
  - 空中操作机器人：如何设计机械臂?
  - 空中作业机器人都有哪些应用?
  - 代表性工作
    - Past, Present, and Future of Aerial Robotic Manipulators. TRO 2022. 空中机械臂领域目前最全的综述文章，入门了解必备。
    - Millimeter-Level Pick and Peg-in-Hole Task Achieved by Aerial Manipulator. TRO 2023, BHU. 使用四旋翼加串联机械臂实现毫米精度peg-in-pole任务。
    - NDOB-Based Control of a UAV with Delta-Arm Considering Manipulator Dynamics [Video]. ICRA 2025, SYU. 使用四旋翼加并联机械臂实现毫米精度抓取。
    - A Compact Aerial Manipulator: Design and Control for Dexterous Operations [Video]. JIRS 2024, BHU. 用空中机械臂做一些有趣的应用，比如抓鸡蛋、开门等等。

- **全驱动无人机(Fully-Actuated UAV)**

  常见的四旋翼无人机具有欠驱动特性，即位置与姿态耦合。而具有位置姿态解耦控制的全驱动无人机，理论上更适合作为空中操作的飞行平台。

  - 代表性工作
    - Fully Actuated Multirotor UAVs: A Literature Review. RAM 2020. 全驱动无人机领域目前最全的综述文章，入门了解必备。
    - Design, modeling and control of an omni-directional aerial vehicle. ICRA 2016, ETH. 第一个实现全向飞行的固定倾角全驱动无人机。
    - The Voliro omniorientational hexacopter: An agile and maneuverable tiltable-rotor aerial vehicle. RAM 2018, ETH. 第一个实现全向飞行的可变倾角全驱动无人机
    - FLOAT Drone: A Fully-actuated Coaxial Aerial Robot for Close-Proximity Operations [Website]. Arxiv 2025, ZJU. 适合近端作业的小尺寸全驱动无人机。

- **可变形无人机(Deformable UAV)**

  除了通过往飞行平台上安装机械臂，让无人机本体可以变形，也是使其实现更多功能的一种方法。

    - 代表性工作
        - Design, Modeling, and Control of an Aerial Robot DRAGON: A Dual-Rotor-Embedded Multilink Robot With the Ability of Multi-Degree-of-Freedom Aerial Transformation. RAL 2018，东京大学. Best paper award on UAV in ICRA 2018，多关节可变形无人机。
        - The Foldable Drone: A Morphing Quadrotor That Can Squeeze and Fly. RAL 2019, Uzh. 四旋翼每个机臂上安装一个舵机，实现机体变形飞行。
        - Ring-Rotor: A Novel Retractable Ring-Shaped Quadrotor With Aerial Grasping and Transportation Capability [Video]. RAL 2023, ZJU. 一种可变形的环形四旋翼，可用于抓取、运输等任务。
        - Design and Control of a Passively Morphing Quadcopter [Video]. ICRA 2019, UCB. 一种被动变形的四旋翼无人机。

- **多模态无人机(Multi-Modal UAV)**

  无人机与地面机器人相比，其优势在于三维空间运动能力，劣势则是续航差。因此一些研究关注多模态无人机的构型设计、运动控制以及自主导航。多模态无人机具备空中、地面、水下等多域运动能力。这不仅能解决无人机的续航问题，也能让无人机具有更多应用潜力。

    - 代表性工作
        - A bipedal walking robot that can fly, slackline, and skateboard. SR 2021, Caltech. 多模态空地足式机器人。
        - Multi-Modal Mobility Morphobot (M4) with appendage repurposing for locomotion plasticity enhancement. NC 2023, Northeastern University. 具有很多种运动模式的多模态无人机。
        - Skater: A Novel Bi-Modal Bi-Copter Robot for Adaptive Locomotion in Air and Diverse Terrain [Video]. RAL 2024, ZJU. 适应多样地形的多模态空地双旋翼无人机。
        - Autonomous and Adaptive Navigation for Terrestrial-Aerial Bimodal Vehicles. RAL 2022, ZJU. 实现空地多模态无人机的自主导航。

3.10.3 Autonomous Driving - 自动驾驶

自动驾驶之心 (也有个微信公众号)

自动驾驶被称为“最小的具身智能验证场景”，这是因为它在具身智能的框架中，具备完整的感知、决策和行动闭环，但任务目标明确、物理交互简单、场景复杂性相对较低。作为一个技术验证场景，自动驾驶既能体现具身智能的核心特性，又为更复杂的具身智能任务提供了技术积累和理论支持。

Model：自动驾驶仿真

生成式仿真为具身智能释放无限灵感

自动驾驶仿真是自动驾驶技术开发中不可或缺的一部分。它通过提供安全、高效、可控的测试环境，不仅降低了研发成本和风险，还加速了技术的迭代和规模化部署。同时，仿真能够覆盖大量现实中难以复现的场景，为自动驾驶系统的安全性、可靠性和泛化能力提供了重要保障。

1. 3D/4D 场景重建

- 经典工作：NSG, MARS, StreetGaussians, OmniRe
    - **NSG**: CVPR 2021, github, arxiv, paper
    - **MARS**: github, arxiv

    - **StreetGaussians**: github, arxiv
    - **OmniRe**: ICLR 2025 Spotlight, demo page, github, arxiv

2. 场景可控生成(世界模型)

- 经典工作：GAIA-1, GenAD（OpenDV数据集）, Vista, SCP-Diff, MagicDrive -> MagicDriveDiT, UniScene, VaVAM
    - **GAIA-1**: demo page, arxiv
    - **GenAD**: CVPR 2024 Highlight, OpenDV数据集, github, arxiv
    - **Vista**: NeurIPS 2025, demo page, github, arxiv
    - **SCP-Diff**: demo page, github, arxiv
    - **MagicDrive** -> MagicDriveDiT: demo page, arxiv
    - **UniScene**: CVPR 2025, demo page, arxiv
    - **VaVAM**: github

Policy：自动驾驶策略

1. 从模块化到端到端

- 经典的模块化管线中，每个模型作为一个独立的组件，负责对应的特定任务(3D目标检测与跟踪 & BEV建图 -> 目标运动预测 -> 轨迹规划)，这种设计已逐渐被端到端模型所取代。

End-to-end Autonomous Driving: Challenges and Frontiers

2. 快系统与慢系统并行

理想端到端-VLM双系统

- 快系统经典论文：UniAD (CVPR 2023 Best Paper), VAD, SparseDrive, DiffusionDrive
    - **UniAD**: CVPR 2023 Best Paper, github, arxiv
    - **VAD**: ICCV 2023, github, arxiv
    - **SparseDrive**: github, arxiv
    - **DiffusionDrive**: CVPR 2025, github, arxiv
    - 快系统的 Scale up 特性探究：https://arxiv.org/pdf/2412.02689
- 慢系统经典论文：DriveVLM, EMMA
    - **DriveVLM**: CoRL 2024, arxiv
    - **EMMA**: arxiv
        - **Open-EMMA** 是EMMA的一个开源实现，提供了一个用于自动驾驶车辆运动规划的端到端框架。

未来发展方向

AIR ApolloFM技术全解读

4. Control and Robotics - 控制论与机器人学基础

4.1. 控制理论基础

4.1.1 经典控制原理

- 理解系统、反馈
- 时域与频域分析
- 传递函数
- 理解前馈控制、反馈控制
- **PID控制**：CSDN

4.1.2 现代控制理论(线性系统控制)

- Modern Control Systems (14th edition), Robert. H. Bishop, Richard. C, Dorf. z: Book
- 状态方程
- 状态反馈与最优控制
- **LQR控制**

4.1.3 先进控制技术

- 鲁棒控制
- 彻底搞懂阻抗控制、导纳控制、力位混合控制: CSDN
- **模型预测控制 MPC**
- 智能控制 (包含基于深度学习的控制)

4.2. 机器人学导论

4.2.1 推荐材料

- 现代机器人学(非常推荐！)video
- 经典教材
    - 《机构学与机器人学的几何基础与旋量代数》 戴建生院士 著
    - 《现代机器人学：机构、规划与控制》凯文·M. 林奇, 朴钟宇 著
    - 《机器人学的现代数学理论基础》丁希仑 著

4.2.2 机器人运动学 (Kinematics) 与动力学 (Dynamics)

1. 机器人运动学

> 想要快速了解什么是IK FK的同学可以看这个7分钟的短片, 可以对此建立一个粗略的认知：BiliBili
> 较为简单的过一遍IK和FK的原理可以看这个：CSDN

- IK (Inverse Kinematics) 逆运动学
    - 较为详细的视频课
        - BiliBili IK(1)
        - BiliBili IK(2)
    - 文字教学
        - Book, 较为详细的IK理论
- FK (Forward Kinematics) 正运动学
    - 较为详细的视频课
        - BiliBili FK(1)
        - BiliBili FK(2)

2. 机器人动力学(**重要！！！**)

- 理解斜对称矩阵, Twist和Exponential of a twist, 旋量代数

4.2.3 里程计和同步定位与建图 (Odometry&SLAM)

里程计(Odometry)用于为机器人实时提供定位，里程计常常基于扩展卡尔曼滤波(EKF)实现，融合来自IMU、相机、激光雷达、码盘、毫米波雷达、光流传感器等等各种常用于机器人位姿感知的传感器之中的多种观测，以较高的频率实现对机器人位姿的估计。

里程计中最常见的是视觉惯性里程计(VIO)和激光惯性里程计(LIO)，其中比较经典的工作包括VINS系列VINS-MonoVINS-Fusion，LOAM，FAST-LIO等等。此外还有融合了IMU、相机和激光传感器的里程计FAST-LIVO系列等。

SLAM(Simultaneous Locolization And Mapping)在定位的同时完成地图的构建，使得回环(Loop Closure)检测成为可能，回环检测的存在使得当机器人重新访问到某个位置时可以修正一部分的累计误差，提高在长时间作业时的定位精度。SLAM的实现主要有filter-based和optimization-based两种，实现中一般又分前端和后端，基于不同传感器的SLAM又各有其特点，在这里提供一些学习资源：

- SLAM Handboook
- Past, Present, and Future of Simultaneous Localization And Mapping: Towards the Robust-Perception Age: SLAM领域的经典综述
- 高翔老师的《视觉SLAM十四讲》
- 高翔老师的《激光SLAM十四讲》

此外，SLAM也有端到端的实现DROID-SLAM。

SLAM的经典工作有ORB-SLAM系列等。

4.2.4 杂项 Misc

- ROS基础:
  - 具身智能ROS1基础: website
  - 具身智能ROS2基础: website
- 常用的库
  - cuRobo：cuRobo, cuRobo是Nvidia的一个利用 CUDA 加速的机器人库, 提供了一套高效的机器人算法, 主要通过并行计算显著提升性能, 包括但不限于IK, 碰撞检测, 路径规划等。
  - IKFast：IKFast, 经典IK库。
  - mplib：mplib, Maniskill Benchmark以及Sapien仿真平台的IK库。
- ROS多传感器时间戳同步：website
- 动手实践LeRobot SO-100：website

5. Hardware - 硬件

> 具身智能硬件方面涵盖多个技术栈, 如嵌入式软硬件设计, 机械设计, 机器人系统设计, 这部分知识比较繁杂, 适合想要专注此方向的人 关于硬件部分的学习, 最好从实践出发!

5.1 Embedded - 嵌入式

- 嵌入式学习路线: CSDN
- 51单片机：BiliBili, 经典江科大自动协出品
- Stm32单片机：BiliBili, 经典江科大自动协出品
- Stm32电机驱动：BiliBili, 野火
- 野火Stm32标准库：BiliBili, 野火
- 正点原子Stm32：BiliBili, 正点原子
- 韦东山嵌入式Linux：BiliBili, 韦东山

5.2 Mechanical Design - 机械设计

- SoildWorks教学：BiliBili
- URDF生成：CSDN, 指导如何通过SolidWorks装配体出发生成机器人URDF文件。

5.3 Robot System Design - 机器人系统设计

- 《机器人学简介》, 来自[2]做的高质量教材: PDF
- 《机器人系统教材》: website

5.4 Sensors - 传感器

5.4.1深度相机

RealSense， RealSence Ros 开发套件

5.5 Tactile Sensing - 触觉感知

1. 视触觉传感器(Vision-Based Tactile Sensors)

视触觉传感器通过摄像头捕捉触觉信息，将触摸表面变形映射为视觉数据，以估计接触力、形变等信息。其设计涉及 **传感器形状**(影响接触范围与适应性)、**标记点设置**(追踪表面形变，提高分辨率)、**材料选择**(如硅胶或弹性体，提高灵敏度)以及 **光照与摄像系统**(增强视觉信号质量)。

- **优点**：提供高分辨率触觉信息、非侵入式感知、不影响物体表面特性，并且可与视觉系统集成，提高多模态感知能力。
- **缺点**：计算量大，依赖视觉处理和机器学习；易受环境光影响；光学设计复杂，封装和耐用性受限。

参考文献综述：写的非常详细，分别是算法和结构设计

- 算法：*When Vision Meets Touch: A Contemporary Review for Visuotactile Sensors From the Signal Processing Perspective*
- 结构：*On the Design and Development of Vision-Based Tactile Sensors*

2. 电子皮肤(Electronic Skin)

触觉感知的路径主要就是这两类。电子皮肤模拟人类皮肤的触觉能力，通常采用柔性电子材料(如压力传感薄膜、纳米传感器网络等)来感知外界压力、温度和形变，使机器人具备更接近生物的触觉感知能力。

- **优点**：电子皮肤可 **大面积覆盖** 机器人表面，实现全身触觉感知；具有 **高灵敏度**，能够检测微小的力变化，实现精准反馈；同时 **可伸缩性** 使其适应复杂表面，提高耐久性。

- **缺点**：电子皮肤的 **制造复杂**，材料和工艺要求高，成本较高；**数据处理挑战**，大规模触觉数据需要高效的计算与存储方案；此外，**稳定性问题** 可能导致长期使用后灵敏度下降，影响可靠性。

参考文献综述：*Toward an AI Era: Advances in Electronic Skins*

3. 触觉感知的应用和算法(视触觉)

- 3.1 姿态估计(Pose Estimation)
  - 估计in hand物体姿态
    - *3D Shape Perception from Monocular Vision, Touch, and Shape Priors*
  - in scene
    - *Fast Model-Based Contact Patch and Pose Estimation for Highly Deformable Dense-Geometry Tactile Sensors*
- 3.2 物体分类(Classification)
  - 区分不同液体、材料或透明物体。
    - *Understanding Dynamic Tactile Sensing for Liquid Property Estimation*
    - *Multimode Fusion Perception for Transparent Glass Recognition*
- 3.3 触觉操控(Manipulation)
  - 物体装配
    - *Active Extrinsic Contact Sensing: Application to General Peg-in-Hole Insertion*
    - *Building a Library of Tactile Skills Based on Fingervision*
  - 线缆整理
    - *Cable Manipulation with a Tactile-Reactive Gripper*
  - 精细手部操作
    - *Manipulation by Feel: Touch-Based Control with Deep Predictive Models*
    - *NeuralFeels with Neural Fields: Visuotactile Perception for In-Hand Manipulation*
- 3.4 触觉大模型(Large Tactile Models)
  - 以统一多模态触觉表示，提高通用性。
    - *Binding Touch to Everything: Learning Unified Multimodal Tactile Representations*

4. 传感器购买

市面上有一些成熟的视触觉传感器可供选择 🔗 **GelSight 官网**

5.6 Companies - 公司

公司	主营产品	Others
松灵AgileX	pipper机械臂 移动底盘	面向教育科研

公司	主营产品	Others
宇树Unitree	四足机器人开发指南 Go2机器狗 AlienGo机器狗 通用人形H1 通用人形G1	许多产出使用宇树的机器人作为硬件基础
方舟无限 ARX	X5机械臂 X7双臂平台 R5机械臂	适合复现很多经典的工作, eg. aloha RoboTwin松灵底盘+方舟臂
波士顿动力	spot机器狗 Atlas通用人形	具身智能本体制造商, 从液压驱动转向电机驱动
灵心巧手		
灵巧智能 DexRobot	Dexhand 021灵巧手	19自由度量产灵巧手
银河通用		已完成多轮融资
星海图 Galaxea	A1机械臂	
World Labs		专注于空间智能, 致力于打造大型世界模型(LWM), 以感知、生成并与 3D 世界进行交互。 相关介绍
星动纪元	Star1人形 XHAND1灵巧手	
加速进化	Booster T1人形	
青龙机器人		
科技云深处	绝影X30四足机器人 Dr.01人形机器人	
松应科技		具身智能仿真平台供应商
光轮智能		具身智能数据平台
智元机器人	A2人形机器人 A2-D数据采集机器人(轮式人形)	
Nvidia		具身智能基建公司
求之科技		
穹彻智能		
优必选		
具身风暴		落地具身智能通用按摩机器人

6. Software - 软件

6.1 Simulators 仿真器

常见仿真器wiki: wiki

仿真器	对应基准集
IsaacGym	legged gym parkour(包括蒸馏以及真机部署) extreme-parkour
IsaacSim	BEHAVIOR-1K(可跨平台)+omniGibson(工具链) ARNOID
MuJoCo	robosuite+robomimic(工具链) LIBERO MetaWorld Gymnasium-Robotics(Fetch; Shadow Dexterous Hand; Maze; Adroit Hand; Franka Kitchen; MaMuJoCo) RoboCasa RoboHive
Sapien	ManiSkill RoboTwin
CoppeliaSim	RLBench PerAct2 COLOSSEUM
PyBullet	Calvin Ravens VimaBench
Genesis	
SOFA	常用于软体机器人的仿真

教程:

- **Isaac 101**： Blog by Axi404.

6.2 Banchmarks 基准集

具身智能常用benchmark总结 [1]: zhihu

- **CALVIN**, github, website2022年, 第一个公开的结合了自然语言控制、高维多模态输入、7自由度的机械臂控制以及长视野的机器人操纵benchmark。支持不同的语言指令, 不同的摄像头输入, 不同的控制方式, 主要用来评估具身智能模型的多模态输入的能力和长程规划能力。
- **Meta-World**, webpage: 评估机器人在多任务和元强化学习场景下的表现。50个机器人操作任务(如抓取、推动物体、开门等), 组织成不同的基准测试集(如ML1、ML10、ML45、MT10、MT50等), 每个集

合都有明确的训练任务和测试任务。周边和文档比较全面, 基于mojoco, 有完整的API和工具, python import即可运行。
- **Embodied Agent Interface: Benchmarking LLMs for Embodied Decision Making**, website: 主要评估大型语言模型(LLMs)在具身决策中的表现, 重点在于决策过程, 包括目标解释、子目标分解、动作序列化和状态转换建模, 不涉及到具体的执行。
- **RoboGen**, repo, website: 不是生成policy, 而是生成任务、场景和带标记的数据, 能直接用来监督学习。
- **LIBERO**, repo, website: 用一个程序化生成管道来生成任务, 这个管道理论上可以生成无限数量的操作任务, 还提供了： 三种视觉运动策略网络架构(RNN、Transformer和ViLT) 和 三种终身学习算法, 以及顺序微调和多任务学习的基准。
- **RoboTwin**, repo: 使用程序生成双臂机器人无限操作任务数据, 并提供了所有任务的评测基准。

6.3 Datasets 数据集

- **Open X-Embodiment: Robotic Learning Datasets and RT-X Models**, website: 22种不同机器人平台的超过100万条真实机器人轨迹数据，覆盖了527种不同的技能和160,266项任务，主要集中在抓取和放置。
- **AgiBot World Datasets (智元机器人)**, website: 八十余种日常生活中的多样化技能，超过100万条轨迹数据，采集自**同构型机器人**, 多级质量把控和全程人工在环的策略，从采集员的专业培训，到采集过程中的严格管理，再到数据的筛选、审核和标注，每一个环节都经过了精心设计和严格把控。
- **RoboMIND**, website: 包含了在479种不同任务中涉及96类独特物体的10.7万条真实世界演示轨迹，来自四种不同协作臂，任务被分为基础技能、精准操作、场景理解、柜体操作和协作任务五大类。
- **All Robots in One,** website: ARIO 数据集，包含了 **2D、3D、文本、触觉、声音 5 种模态的感知数据**，涵盖**操作**和**导航**两大类任务，既有**仿真数据**，也有**真实场景数据**，并且包含多种机器人硬件，有很高的丰富度。在数据规模达到三百万的同时，还保证了数据的统一格式，是目前具身智能领域同时达到高质量、多样化和大规模的开源数据集。

7. Paper Lists - 论文列表

- Awesome Humanoid Robot Learning - Yanjie Ze: repo
- Paper Reading List - DeepTimber Community: repo
- Paper List - Yanjie Ze: repo
- Paper List For EmbodiedAI - Tianxing Chen: repo
- SOTA Paper Rating - Weiyang Jin: website
- Awesome-LLM-Robotics: A repo contains a curative list of papers using Large Language/Multi-Modal Models for Robotics/RL: website

8. Acknowledgement - 致谢

本文转载/引用了一些博主的文章, 我们对他们的知识分享表示感谢, 引用列表如下： [1] 知乎 穆尧, [2] 知乎 东林钟声, Github Yunlong Dong, [3] 知乎 强化学徒, [4] 知乎 Biang哥, [5] OpenAI Lilian Weng, [6] B站 木木具身, [7] Github Zhuoheng Li, [8] 知乎 Flood Sung, [9] Github Sida Peng

👍 Citation - 引用

If you find this repository helpful, please consider citing:

```
@misc{embodiedaiguide2025,
      title = {Embodied-AI-Guide},
      author = {Embodied-AI-Guide-Contributors, VITA-Robotics-Community},
      year = {2025},
      howpublished = {\url{https://github.com/tianxingchen/Embodied-AI-Guide}},
}
```

License - 许可证

This repository is released under the MIT license. See LICENSE for additional details.

Star History - Star历史