



中国科学技术大学
University of Science and Technology of China

丁鑫

手机：(+86) 185-8489-1264 · 政治面貌：中共党员
邮箱：xinding64@mail.ustc.edu.cn · 籍贯：山西省晋中市
研究方向：多模态大模型，具身 Agent · 性别：男



教育背景，实习经历

中国科学技术大学 | 信息科学与技术 | 博士 | 硕士 | 1985 2022.9 - 预计 2027.6
• 荣誉: 国家奖学金 (Top 1%)
华为 2012 中央研究院诺亚实验室 2023.4 - 2024.6 微软亚洲研究院 2024.6 - 至今
• 华为创新先锋 (中央研究院最高奖 Top 2%，唯一实习生)、华为 2012 之星、微软明日之星

论文发表

- 「Neurips 2026 在投 Corresponding Author」Ju R, Wang X, **Ding X**, et al. EmbodiSkill: Skill-Aware Reflection for Self-Evolving Embodied Agents[J]. 公众号链接 Link 获三千次以上转发.
- 「Neurips 2026 一作在投」**Ding X**, Wang X, Yang Y, et al. MemCompiler: Compile, Don't Inject-State-Conditioned Memory for Embodied Agents. 公众号链接 Link
- 「Neurips 2026 在投」Li X, Tang H, **Ding X**, Wang W, et al. Efficient Multi-Task VLA Inference via Unified KV Management for Mixture-of-Transformers
- 「Neurips 2026 共一在投」Chen F, **Ding X**, Zhu H, et al. FlowBridge: 3D Scene Flow as a Shared Dynamics Space for Cross-Embodiment VLA
- 「ECCV 2026 共一」Zheng Y, **Ding X**, Yang Y, et al. Em-Garde: A Propose-Match Framework for Proactive Streaming Video Understanding. 清华大学官方公众号宣传: 链接 Link
- 「ICML 2026」**Ding X**, Wei J, Yang Y, et al. AdaNav: Adaptive Reasoning with Uncertainty for Vision-Language Navigation, 获得视觉语言导航 (导航领域专业公众号) 的推荐: 链接 Link
- 「ICLR 2026」**Ding X**, Li X, Qin H, et al. Q&c: When quantization meets cache in efficient image generation.
- 「ICLR 2025 Spotlight」**Ding X**, Liu X, Tu Z, et al. Cbq: Cross-block quantization for large language models, 获得新智元公众号报道: 链接 Link
- 「ICCV 2025」**Ding X**, Wu H, Yang Y, et al. Streammind: Unlocking full frame rate streaming video dialogue through event-gated cognition, 被微软研究院官方报道: 链接 Link
- 「MobiCom 2025 Demo」**Ding X**, Wei J, Jia F, et al. EdgeMind-OS: A Plug-and-Play Embodied Intelligence System for Real-Time On-Device Deployment
- 「CVPR 2025」**Ding X**, Yu L, Li X, et al. Rass: Improving denoising diffusion samplers with reinforced active sampling scheduler
- 「ECCV 2024」**Ding X**, Liu X, Yu L, et al. Pq-sam: Post-training quantization for segment anything model

科研经历

面向自主进化具身智能体的世界动作模型与技能结晶研究 | 2026 在研 | 负责人 2026.04-至今

- **关键词**: 世界-动作因果建模, 技能结晶, 统一记忆, 长期决策, 自进化具身智能
- **关注重点**: 针对当前具身智能系统普遍存在的“刺激-反应式”范式局限, 探索从统计相关性建模走向因果结构建模的路径, 构建可持续进化的具身学习框架, 使智能体能够显式理解“动作如何改变世界”, 并将真实交互中的因果经验自动沉淀为可复用技能与记忆结构, 从而实现从一次性执行系统向持续学习型自主智能体的范式转变; 进一步在真实机器人系统中验证其自进化能力, 在非结构化环境下支持少样本快速适应 (<10 次自身交互 / <5 人类示范), 并实现新经验向统一记忆体系的无缝融合, 最终构建可演示系统原型, 面向具身智能产业落地应用。
- **方法设计**: 构建世界-动作因果建模机制, 利用历史交互中的动作-结果关系提升长程决策能力; 设计失败/成功驱动的技能结晶机制, 将经验结构化为可复用技能库; 统一情节记忆与语义技能记忆, 使智能体在单任务与跨任务层面同步进化。
- **实验进展**: 已在 GR00T N1.5 与 Cosmos policy 上初步验证因果建模有效性, 在 Robocasa365 上实现约 30-40% 性能提升; 进一步验证在低数据与跨场景设置下的显著适应优势。

面向多模态具身智能体的可微记忆管理框架 | Neurips 一作在投 2025.12-2026.05

- **关键词**: 记忆管理器 (Memory Manager), Soft-CoT, SFT+RL 训练;

- **关注重点：**针对具身智能中“何时以及如何将历史记忆作为经验参与决策”这一核心问题，提出可训练的 **Memory Manager** 框架，将多模态任务轨迹（3D / 空间 / 视频 / 文本）压缩为可供下游 LLM 调用的**连续潜在经验表征**。为避免纯文本序列带来的信息损失与高延迟问题，设计 **Soft-CoT 记忆传输机制**：在生成可解释明文推理链（Hard-CoT）的同时，通过若干可微 **soft thought tokens** 提取连续隐藏状态，经投影映射至下游 LLM 嵌入空间，实现“**明文可解释 + 潜变量高保真**”的双通道经验传递范式。该方法避免端到端大模型重训练的高计算成本与灾难性遗忘问题，支持 SFT + RL 分阶段优化，为多模态记忆与决策模型的模块化协同提供新的可扩展路径。

面向多任务 VLA 推理的统一 KV Cache 管理 | Neurips 2026 在投

2025.10-2026.02

- **关键词：**VLA、多任务并行、KV Cache 管理
- **关注重点：**针对 MoT-VLA 多任务推理中重复计算与资源竞争问题，提出统一 KV Cache 管理框架，通过跨任务 KV 共享与跨帧连续批处理实现多模态任务协同推理，大幅提升动作频率与语言生成吞吐率，实现最高 3.7x 系统推理加速。

基于跨具身数据训练的高效 VLA 模型 | Neurips 2026 在投 | 共一

2025.08-2026.1

- **关键词：**跨具身预训练，VLA, 3D 场景流；
- **关注重点：**针对预训练中跨具身数据在形态结构与控制空间上的**观测-动作分布不一致问题**，提出基于 **3D 交互场景流** 桥接的对齐框架。通过构建交互区域场景流 token，将多具身数据投影到共享动力学空间进行对齐，并结合**扩散式动作生成器**映射回各具身动作参数空间，实现“**对齐交互效果、解耦控制空间**”的跨平台学习范式。在 LIBERO 与 CALVIN 基准上取得显著性能提升，相比直接跨具身联合训练，任务成功率提升 **15%**，验证了跨具身数据对齐机制的稳定性与可扩展性。

面向实时主动响应的流视频理解模型 | ECCV 2026 | 共一

2025.10-2026.02

- **关键词：**主动流视频理解，触发决策，语义-感知解耦，强化学习；
- **关注重点：**针对流视频理解中“何时响应”的核心难题，提出**语义解析与流式感知解耦**的新范式，将高成本语义推理移出流式循环，通过结构化视觉 proposal 与轻量级匹配机制实现高效触发决策。在 StreamingBench 与 OVO-Bench 上显著优于现有 Streaming VideoLLMs，F1 提升 **10%+**，在保证性能提升的同时实现 **13fps** 实时推理且不随上下文增长退化，其所拓展等主动处理能力在具身机器人感知，AI 眼镜，家庭助手等众多智能设备上具有出色的应用潜力。

基于不确定度的自适应推理导航模型 | ICML 2026 | 一作

2025.04-2025.08

- **关键词：**自适应推理，多模态大模型，视觉-语言-导航；
- **关注重点：**提出基于不确定性的自适应推理框架了 AdaNav，通过引入不确定性自适应推理块（UAR Block）和启发式到强化学习的训练机制，使智能体能够在导航过程中根据需求**自适应地触发推理**，解决了固定步长推理导致的性能次优和计算开销问题，推理消耗降低 3 倍。在仅使用 **6K** 训练样本的情况下，AdaNav 在多个基准测试中取得显著的性能提升，**超过了使用百万级数据训练的闭源模型**。

面向高帧率的流视频理解大模型 | ICCV 2025 | 一作

2024.06-2025.01

- **关键词：**流视频理解，高帧率；
- **关注重点：**关注于高帧率流视频理解，能够持续感知和理解用户所处环境、发生的事件以及用户指令，并据此实时主动地给用户提供帮助，通过事件感知特征提取器（EPFE）和认知门控（cognition gate）两项核心技术，实现了感知与认知的解耦，突破了 **100fps** 高帧率视频处理的瓶颈。StreamMind 在多个基准测试中取得 SOTA 效果，在 **Microsoft xbox 游戏引擎**，和亚信监控异常检测项目转化。

大语言模型后训练量化 | ICLR Spotlight 2025 | 一作

2024.01-2024.06

- **关键词：**大语言模型, 量化, 极低比特；
- **关注重点：**提出了跨块依赖重建方案恢复量化模型精度损失；提出了对模型权重补偿矩阵运用低秩分解方法加速优化；对比现有的 LLM 量化方法，所提出方法在**多个自然语言处理任务中取得 SOTA 性能**。在 W4A16 和 W8A8 设置下，量化后的模型相比 FP16 模型的精度差距在 1% 以内。同时，所提出的方法第一次实现 W4A4 极低 bit 设置下对大模型的量化。该算法已被落地到**华为多项大模型加速项目中**，如盘古智子、ICT 数据存储大模型，云山大模型等。

SAM 模型的后训练量化 | ECCV 2024 | 一作

2026.06-2023.11

- **关键词：**SAM 模型，联合量化，Zero-Shot；
- **关注重点：**提出了基于梯度的特征重分布与量化步长联合优化方案使得可以充分考虑模型参数分布规律；提出了基于特征异常值检测的层次聚类方法，大幅度缩减可学习参数，提升优化精度和效率第一个针对 SAM 的量化算法，所提出方法在**多个视觉分割任务中取得 SOTA 性能**。在 W8A8 设置下，使用 PQ-SAM 量化后的模型相比浮点模型的精度差距在 1% 以内，达到无损量化。同时，所提出的方法可以做到更低 bit 数的量化，将 W6A6，W4A4 优化到可用水平，精度下降在 2%-9%。目前被应用于**华为海思端侧部署业务**。