秦正

电话 & 微信: 17737732966 | qinzheng@stu.xjtu.edu.cn | 出生日期: 2000-07-13

教育背景

特伦托大学, 人工智能专业, 访问博士 | 导师:Nicu Sebe 教授 2025.09—2026.07
西安交通大学, 控制科学与工程专业, 博士 | 导师:王乐教授 2021.09—2026.09
哈尔滨工业大学, 机器人工程专业, 学士 2017.09—2021.06

实习经历

蚂蚁集团 | 多模态交互组 | 多模态大模型科研实习生 |*2025.03—至今*

- **实习内容 1: HumanSense-OmniReasoning：大模型以人为中心的感知、交互** **AAAI26 (第一作者)**
  核心目的: 提升多模态大模型在复杂人机交互场景中的理解、推理和共情反馈能力，实现更自然、智能的人机互动。**技术方案**: 1. 定义并构建了一个覆盖从感知到交互策略的评测框架，用于评估模型理解人类意图与生成合理响应的能力，并在此基础上构造了训练数据; 2. 提出 Multi-stage, modality-progressive GRPO 框架，显著提升模型在复杂语境下的推理深度与响应准确性; 3. 提出 Prompt Enhancement 策略: 提取 GRPO 训练过程中模型生成的高质量思维链特质，将其整合为提示词，能够在无需额外训练的情况下，激发非推理模型产生类似的推理能力。
- **实习内容 2: 视频 MLLM 的 id-consistency 能力** ongoing
  核心目的: 提升大模型在视频中对目标身份保持（ID-consistency）的能力。**技术路径**: 1. 以视频多目标跟踪和计数任务作为研究大模型的 ID-consistency 能力的基座，构建详尽的评测体系，系统评估 MLLM 在多目标跟踪与计数任务中的能力边界；2. 对多目标跟踪任务进行子任务拆解，设计针对性子任务评测，以精准识别模型性能瓶颈；3. 针对发现的关键限制进行特定数据构造与训练，目前第三步骤仍在进行中。

蚂蚁集团 | 数字人算法组 | 数字人算法科研实习生 |*2024.05—2025.02*

- **实习内容 1: VersaAnimator: 具有多模态控制能力的说话人全身视频生成框架** **ACMMM25 (第一作者)**
  针对现有说话人视频生成方法中肢体动作单一且缺乏精细化动作编辑能力的问题；提出了业界首个同时支持音频驱动、文本控制与全身尺度生成的通用数字人视频生成方案; 实验结果表明, 该方法在口型同步精度以及肢体动作表现力等方面均取得优异表现,并进一步支持用户级个性化肢体动作定制。业界影响力: 用于支持医疗大模型数字人团队。
- **实习内容 2: Diverse-T2M: 引入不确定性的多样性 3D 人体运动生成方法 TCSVT under review (第一作者)**
  针对现有两阶段文本驱动运动生成方法生成的动作过于单一（同质化严重）的问题；显式将不确定性引入生成过程；显著提升了生成动作的多样性。业界影响力: 算法在内部被部署为服务，并供 Galacean 引擎调用，可用于驱动蚂蚁庄园小鸡和数字人 Luna，部署同学反馈效果很好。

伊利诺伊大学芝加哥分校 | *Wei Tang* 教授团队 | 计算机视觉科研实习生 |*2022.05—2024.03*

- **实习内容 1: MotionTrack: 基于长短期运动的多目标跟踪框架** **CVPR23 (第一作者)**
  针对人群密集和严重遮挡场景下多目标跟踪中的 ID Switch 问题; 提出融合了数据驱动社会力建模和历史轨迹回溯的多目标跟踪框架; 该方法摆脱复杂 Re-ID 依赖，在极低计算开销与硬件友好约束下显著提升了目标身份保持能力。业界影响力: 该工作已获得 Google Scholar 210+ 次引用，其核心技术因低算力、易部署的特性受到大疆（DJI）自动驾驶团队邀请，并被认为具备在低功耗视觉芯片上的落地潜力。
- **实习内容 2: GeneralTrack: 多应用场景统一跟踪框架** **CVPR24 (第一作者)**
  针对市面上同一款跟踪器无法适应多种跟踪应用场景的泛化性问题; 对不同应用场景的数据特性进行了系统定量分析，并提出了基于点-区域-目标的统一跟踪框架; 实验在 5 类不同类型的 benchmark 上同时取得 SOTA 性能。
- **实习内容 3: RSRNav: 基于空间关系推理的图像目标导航方法** **TCSVT under review (第一作者)**
  针对视觉导航 agent 在训练与应用视角不一致时性能大幅下降的问题; 提出了一种显式推理空间关系的新范式，将复杂的语义对齐简化为直观的方向指引信息; 实验表明，成功率和效率在训练与应用视角不一致时有大幅提高。

创业项目

AR 景区互动小程序 | 合作对象：洛阳市洛邑古城、丽景门 |*2023.11—2024.01*

具体内容: 开发了一个基于 AR 的景区互动小程序。用户打开后，摄像头识别景区大门或预设场景，就会触发 3D 凤凰特效，围着城墙飞舞；夜晚还可以在手机屏幕中看到我实现的烟花特效。更有趣的是，当扫描到特定贴图，图中的形象就会变成 3D 建模出现在屏幕里，游客可以和它合影。**结果**: 整个项目从与景区谈判、组建二人小型团队、打通技术

链路，产品迭代测试、到最终版权售卖 1.8 万元，虽然金额不大，但让我积累了宝贵的商业需求洞察、谈判技巧、团队协作以及产品落地经验。

个人荣誉

- 国家奖学金（博），2025
- 潍柴动力奖学金（博），2025
- 优秀研究生（博），2023，2024
- 一等新生奖学金，2021
- 哈尔滨工业大学优秀毕业生，2021
- 一等学业奖学金，2018，2019，2020

科研成果

谷歌学术

- MotionTrack: Learning Robust Short-term and Long-term Motions for Multi-Object Tracking. Qin Z, Zhou S, Wang L, Duan J, Hua G, Tang W. **CVPR2023**.
- Towards Generalizable Multi-Object Tracking. Qin Z, Wang L, Zhou S, Fu P, Hua G, Tang W. **CVPR2024**.
- Referencing Where to Focus: Improving Visual Grounding with Referential Query. Wang Y, Tian Z, Guo Q, Qin Z, Zhou S, Yang M, Wang L. **NIPS2024**
- RefDetector: A Simple yet Effective Matching-based Method for Referring Expression Comprehension. Wang Y, Tian Z, Qin Z, Zhou S, Wang L. **AAAI2025**
- Towards Precise Embodied Dialogue Localization via Causality Guided Diffusion. Wang H, Wang L, Qin Z, Wang Y, Hua G, Tang W. **CVPR2025**
- Versatile Multimodal Controls for Whole-Body Talking Human Animation. Qin Z, Zheng R, Wang Y, Li T, Zhu Z, Yang M, Yang M, Wang L. **ACM MM2025**
- HumanSense: From Multimodal Perception to Empathetic Context-Aware Responses through Reasoning MLLMs. Qin Z, Zheng R, Wang Y, Li T, Yuan Y, Chen J, Wang L. **AAAI2026**
- Single-Shot and Multi-Shot Feature Learning for Multi-Object Tracking. Li Y, Zhou S, Qin Z, Wang L, Wang J, Zheng N. **TMM2024**
- Robust Noisy Label Learning via Two-Stream Sample Distillation. Bai S, Zhou S, Qin Z, Wang L, Zheng N. **TMM2025**
- Semantic and Kinematics Guidance for RMOT. Li Y, Zhou S, Qin Z, Wang L. **TMM2025**
- Injecting Position and Relation Prior for Dense Video Captioning. Li Y, Zhou S, Qin Z, Lin J, Sun X, Wu K, Wang L. (Submitted for **TIP**)
- From Mapping to Composing: A Two-Stage Framework for Zero-shot Composed Image Retrieval. Wang Y, Tian Z, Guo Q, Qin Z, Zhou S, Yang M, Wang L. (Submitted for **TCSVT**)
- Embracing Aleatoric Uncertainty: Generating Diverse 3D Human Motion. Qin Z, Wang L, Wang Y, Yang M, Rong C, Yang M, Zheng N. (Submitted for **TCSVT**)
- RSRNav: Reasoning Spatial Relationship for Image-Goal Navigation. Qin Z, Wang L, Wang Y, Zhou S, Hua G, Tang W. (Submitted for **TCSVT**)
- Spatial Matters: Position-Guided 3D Referring Expression Segmentation. Wang Y, Tian Z, Wang L, Qin Z, Zhou S. (Submitted for **CVPR2026**)