个人简历

基本信息

王梓骅 | 男 | 软件工程博士在读| 17354024270| zihua@seu.edu.cn

教育背景

2016.09 – 2020.06	中国科学技术大学	少年班学院	电子信息工程	本科
2020.09 – 2023.03	东南大学	网络空间安全学院	电子信息	硕博连读
2023.03 – 2027（预计）	东南大学	计算机科学与工程学院	软件工程	博士

实习/项目经历

阿里巴巴-通义实验室-机器智能技术-自然语言处理 2022.05-2024.07 多模态图文模型

-研究图文学习的自适应聚类的注意力方法，提升图文语义理解能力以生成更优图像描述和问答
-研究基于核心集二步推荐的多模态上下文样本推荐方法，高效地实现高质量的上下文样本筛选
-研究标签选择、视觉信息预融合和自回归联合训练策略，提高非自回归视觉文本单步生成的质量

阿里巴巴-ATH 通义实验室-Qwen 模型训练（基模）2025.12-至今 GUI Agent、GUI Grounding

-研究面向 GUI Grounding 训练的数据合成链路，提升模型在复杂界面场景下的定位与理解能力，并应用于训练 Mobile-Agent v3.5 、Qwen 基模（Qwen3.6plus、Qwen3.7plus、Qwen3.8max）
-研究静态 Grounding 的强化学习优化方法，以提升模型在动态 GUI 任务中的准确度与鲁棒性
-研究 GUI 任务中的自适应思考机制，提高任务执行过程中的效率与稳定性

技能清单

编程语言：Python-torch（Verl、Sglang、Megatron 等） / 英语能力：CET-4/6，IELTS：7.0

论文/获奖情况

1. **Zihua Wang**, Xu Yang, Haiyang Xu, Hanwang Zhang, Ming Yan, Fei Huang, and Yu Zhang, Adaptively Clustering Neighbor Elements for Image-Text Generation, IEEE Transactions on Multimedia 2026. （CCF A 期刊，JCR Q1）
   - Proposes an unified adaptive neighbor-element clustering attention calculation that groups related visual and textual units to improve coherence and fidelity in image–text generation.
2. **Zihua Wang**, Jiarui Wang, Haiyang Xu, Ming Yan, Fei Huang, Xu Yang, Xiu-Shen Wei, Siya Mi, Yu Zhang. Efficient and Effective In-context Demonstration Selection with Coreset, AAAI 2026. （CCF A 会议）
   - Introduces a coreset-based dual retrieval algorithm that efficiently selects compact and high informative in-context demonstrations to boost few-shot performance with lower computation.
3. **Zihua Wang**, Ruibo Li, Haozhe Du, Joey Tianyi Zhou, Yu Zhang, Xu Yang. SpecFLASH: Latent-Aware Semi-Autoregressive Speculative Decoding for Multimodal Tasks, ACM MM2026 （CCF A 会议）
   - Presents a latent-aware semi-autoregressive speculative decoding framework to compress the redundant visual tokens and generate draft in parallel, which speeds up multimodal generation while retaining high-quality outputs.
4. **Zihua Wang**, Zhitao Lin, Ruibo Li, Yu Zhang, Xu Yang, Siya Mi, Xiu-Shen Wei. Open-Loop Planning, Closed-Loop Verification: Speculative Verification for VLA, ACM MM2026 （CCF A 会议）

- Proposes a speculative verification framework that uses a heavy VLA as a low-frequency macro-planner and a lightweight verifier for high-frequency closed-loop checking, enabling efficient yet robust embodied control via replanning.

5. **Zihua Wang**, Xu Yang, Xiu-Shen Wei, Siya Mi, Yu Zhang, Xin Geng. Efficient Multimodal In-Context Learning via Coreset Retrieval and Speculative Decoding, IEEE Transactions on Pattern Analysis and Machine Intelligence 2026.（CCF A 期刊，under review)
   - Introduce an efficient demonstration selection strategy that balances semantic diversity and query relevance. Motivated by a mutual-information analysis, we construct a diverse coreset via cluster-pruning, which serves as a set of compact anchors to efficiently extrapolate oracle scores and rank the support set for any unseen query.
   -Design a tailored speculative decoding scheme. It leverages a lightweight draft model operating on compressed visual demonstrations, while the target LVLM verifies the outputs using the uncompressed sequence, achieving substantial acceleration while preserving exact output fidelity.
6. **Zihua Wang**, Xu Yang, Haiyang Xu, Ming Yan, Ji Zhang, Fei Huang, Yu Zhang. Exploring Patterns and Semantics in Non-Autoregressive Image Captioning, Pattern Recognition 2025（CCF B 期刊，under review)
   - Deconstructing non autoregressive sentence generation into the sentence patterns and semantics and proposes label selection for sentence patterns and image pre-fusion for semantics. Besides, joint training is applied to close the quality gap with autoregressive models.
7. **Zihua Wang**, Siya Mi, Yu Zhang. DVC^2: Deep video cascade clustering from video structures. Neurocomputing 2025.（CCF C 期刊，JCR Q1）
   - Propose proposes a hierarchical deep clustering framework that leverages spatial–temporal dependencies inherent in video data. By learning the spatial features from frame-clustering and temporal features from long-short video clips, it achieves more interpretable and temporally stable video structure discovery.
8. Yihang Meng, Hao Cheng, **Zihua Wang**, Hongyuan Zhu, Xiuxian Lao, Yu Zhang. Efficient multi-modal human-centric contrastive pre-training with a pseudo body-structured prior. PRCV 2024.
9. Xu Yang, Jiawei Peng, **Zihua Wang**, Haiyang Xu, Qinghao Ye, Chenliang Li, Songfang Huang, Fei Huang, Zhangzikang Li, Yu Zhang. Transforming visual scene graphs to image captions. ACL 2023.
10. Yu Zhang, **Zihua Wang**, Jingyang Zhou, Siya Mi. Person video alignment with human pose registration. Frontiers of Computer Science 2023.

- 三好研究生、华为奖学金、唐仲英奖学金、东南大学二等奖学金、校庆杯创业大赛产业赛道银奖

- 项目论文：Haiyang Xu, Xi Zhang, Haowei Liu, Junyang Wang, Zhaozai Zhu, Shengjie Zhou, Xuhao Hu, Feiyu Gao, Junjie Cao, **Zihua Wang**, Zhiyuan Chen, Jitong Liao, Qi Zheng, Jiahui Zeng, Ze Xu, Shuai Bai, Junyang Lin, Jingren Zhou, Ming Yan. **Mobile-Agent-v3.5**: Multi-platform Fundamental GUI Agents. https://github.com/X-PLUG/MobileAgent.

研究方向

聚焦于高效和高质量视觉文本生成方向，包括视觉信息高效压缩、基于强化学习的高质量视觉文本生成、基于投机解码的多模态文本推理加速、基于投机解码的视觉语言动作任务加速、GUI grounding 与 RL 策略相关的 GUI Agent 技术等。

自我评价

具备良好沟通和团队合作能力；对新技术保持好奇心；具备扎实的算法和实践经验

Explore more on personal website：zihua.me