Qinchen Wu

Shanghai, China | qinchen.wu62@gmail.com | 18964321758 | Personal Website | GitHub

Education

National University of Singapore *Aug 2023 – Jan 2025*
MS in Computer Engineering
- GPA: 4.6/5.0

Nanjing University of Aeronautics and Astronautics *Aug 2018 – Jun 2022*
B.Eng in Electrical Engineering and Automation
- GPA: 88/100

Research Interest

Vision and Language, Memory Optimization, LLM Agents (GUI agents), LLM reasoning

Publications

[1] GUI-Narrator: Detecting and Captioning Computer GUI Actions ACMMM 2025
Qinchen Wu, Difei Gao, Lin Qinghong, Zhuoyu Wu, Mike Zheng Shou.
Proceeding ACMMM 25

[2] VideoGUI: A Benchmark for GUI Automation from Instructional Videos NeurIPS 2024 DB
Lin Qinghong, Linjie Li, Difei Gao, **Qinchen Wu**, Mingyi Yan, Zhengyuan Yang, Lijuan Wang, and Mike Zheng Shou.
VideoGUI

[3] Assistgui: Task-oriented desktop graphical user interface automation CVPR 2024
Gao D, Ji L, Bai Z, Ouyang M, Li P, Mao D, **Wu Q**, Zhang W, Wang P, Guo X, Wang H., Mike Z.
AssistGUI

[4] Harmonizing Unets: Attention Fusion module in cascaded-Unets for low-quality OCT image fluid segmentation Computers in Biology and Medicine
Zhuoyu Wu, **Qinchen Wu**, Wenqi Fang, Wenhui Ou, Quanjun Wang, Linde Zhang, Chao Chen, Zheng Wang
Harmonizing Unets

Research Experience

Detecting and Captioning GUI actions in Videos *Jan 2024 – Apr 2025*
NUS, ShowLab, Supervisor: Mike Zheng Shou
- Designed pipelines for collecting users' actions from GUI environments
- Integrated temporal grounding and ROI-aware mechanisms to enhance VLMs' action understanding and computational efficiency.
- Developed a benchmark standard for GUI-centric-action evaluation based on LLMs. (GPT-4o).
- Accepted by ACM MM 2025.

Developing AI Assistant For GUI Interface Automation *Aug 2023 – Mar 2024*
NUS, ShowLab, Supervisor: Mike Zheng Shou
- Proposed a new baseline for GUI-workflow automation. Including Agent planning, Agent acting, UI-Parsing, and Agent-Critic.
- Excavated the In Context Learning ability by Prompt engineering on LLMs (GPT, GLM, LLaMA) to generate concise and fine-grained plans from human demonstration videos.
- Empowered the Model with RAG to alleviate LLMs' hallucinations of Task planning, build a more robust planning method by LangChain and LLMs.

- Accepted as poster for CVPR 2024.

Attention Fusion-based Attention U-Net for OCT Image Segmentation. *Jan 2023 – March 2024*
Durham Univeristy

- Proposed a cascaded U-Net framework for low-resolution image segmentation, achieving improved Dice scores on three datasets.
- Introduced and experimentally validated the effectiveness of Adaptive Attention Fusion (Channel Fusion, Spatial Fusion).
- Accepted by Computers in Biology and Medicine.

Professional Experience

LLM algorithm Engineer *Apr 2025 – Present*
PDD, TEMU, Full Time

- Pretrained a 1.5B-parameter LLM from scratch using the Megatron-LM framework.
- Optimized the model performance via Multi-Token-Prediction (MTP) and partial RoPE (0.75) and Multi-latent attention for improved performance while maintaining high efficiency during inference.
- Enhanced reasoning capabilities and improved benchmark performance on ARC-Challenge and BBH by leveraging computation scaling and latent reasoning.
- Currently working on efficient attention mechanisms, specifically test-time training (TTT) and neural memory for LLMs.

Multimodal Algorithm Engineer *July 2024 – Oct 2024*
Tencent WeChat, Intern

- Improved the online performance of Matching model based on Sentence Bert. Including optimization in Data, Feature, Model Config, and Loss function. Increasing Precision from 75% to 81%.
- Upgraded the model to a multi-modal Mixture of Experts (MOE) framework, boosting offline precision performance by leveraging visual modalities. enhancing the model's ability to understand and describe dynamic visual content.
- Enabled the model to extract key elements from long-context inputs by finetuning Qwen2, boosting understanding performance in live-stream environments.

Skills

Languages & **Tools:** Python, C, Matlab, Spark, SQL

Computer Vision: OCT Segmentation, SAM, SIFT, Opencv

NLP & **LLM:** BERT, multimodal (llaVa, Qwen-VL, MMF), Retrieval Augmentation Generation

Machine Learning: K-means, Markov random field

Framework: Megatron LM, Deepspeed