

Zhekai Chen

陈哲恺

✉ zkchen66@outlook.com

☎ (+86) 18796807327

🌐 Aziily

🏠 2003-8

Research Interests: *Generative models, Unified Multimodal, Agent*

EDUCATION

University of Hong Kong , Electrical and Computer Engineering, PhD Student	2025.09 -
Supervised by Prof. Xihui Liu, HKU MMLab	
Zhejiang University , Software Engineering, Undergraduate	2021.09 - 2025.06
Major in Computer Science and Technology, with an average GPA of 3.94/4.0	
Suzhou Academy of Jiangsu Province , Key Class, High School	2018.09 - 2021.06

HONORS & AWARDS

HKU Presidential PhD Scholar Programme	2025.9
Ho Chi Kwan Education Scholarship (7 every year)	2024.12
3rd Scholarship of Zhejiang University	2024.11 / 2023.11 / 2022.11
3rd Prize of the Sixth Open Source Innovation Competition	2023.10

RESEARCH EXPERIENCE

Meituan	2025.11 -
Focused on Unified Multimodal Model. Conducted research on in-context generation and unified multi-modal, opensourced Macro , a 400K multi-reference long-context image generation dataset with up to 10 input images.	
Tongyi Laboratory, Alibaba Group	2024.11 - 2025.8
Focused on video generation models and auto-regressive generative models. Contributed to Wan-Mover (NIPS2025), a company-led trajectory control project. Conducted research on test-time scaling, developing the TTS-VAR (NIPS2025) scaling framework.	
National CAD & CG Laboratory , Prof. Chunhua Shen's Group	2022.10 - 2025.06
Developed foundational research skills. Led FreeCompose (ECCV2024), a zero-shot image composition method. Contributed to AutoStory (IJCV) and Framer (ICLR2025), focusing on generative models. Co-authored a patent.	
University of Illinois Urbana Champaign , Prof. Tong Zhang's Group	2024.5 - 2024.9
Focused on LLM advancements. Co-authored Image Textualization (NIPS2024 Dataset Track) for geometrically aware image captions. Explored RLHF in generative models.	

PUBLICATIONS

Papers

- FreeCompose: Generic Zero-Shot Image Composition with Diffusion Prior:** Zhekai Chen*, Wen Wang*, Zhen Yang, Zeqing Yuan, Hao Chen, Chunhua Shen. Published at ECCV 2024 Conference.
- TTS-VAR: A Test-Time Scaling Framework for Visual Auto-Regressive Generation:** Zhekai Chen, Ruihang Chu, Yukang Chen, Shiwei Zhang, Yujie Wei, Yingya Zhang, Xihui Liu. Published at NeurIPS 2025.
- MACRO: Advancing Multi-Reference Image Generation with Structured Long-Context Data:** Zhekai Chen, Yuqing Wang, Manyuan Zhang, Xihui Liu. Submitted to ECCV 2026.
- Mover: Motion-controllable Video Generation via Latent Trajectory Guidance:** Ruihang Chu*, Yefei He*, Zhekai Chen*, Shiwei Zhang, Xiaogang Xu, Bin Xia, Dingdong WANG, Hongwei Yi, Xihui Liu, Hengshuang Zhao, Yu Liu, Yingya Zhang, Yujiu Yang. Published at NeurIPS 2025.
- Framer: Interactive Frame Interpolation:** Wen Wang, Qiuyu Wang, Kecheng Zheng, Hao Ouyang, Zhekai Chen, Biao Gong, Hao Chen, Yujun Shen, Chunhua Shen. Published at ICLR 2025 Conference.
- Image Textualization: An Automatic Framework for Creating Accurate and Detailed Image Descriptions:** Renjie Pi*, Jianshu Zhang*, Jipeng Zhang, Rui Pan, Zhekai Chen, Tong Zhang. Published at NeurIPS 2024 Dataset Track.

7. **AutoStory: Generating Diverse Storytelling Images with Minimal Human Effort:** Wen Wang*, Canyu Zhao*, Hao Chen, **Zhekai Chen**, Kecheng Zheng, Chunhua Shen. Published in the IJCV Journal.
8. **Speculative Jacobi-Denoising Decoding for Accelerating Autoregressive Text-to-image Generation:** Yao Teng, Fu-Yun Wang, Xian Liu, **Zhekai Chen**, Han Shi, Yu Wang, Zhenguo Li, Weiyang Liu, Difan Zou, Xihui Liu. Published at NeurIPS 2025.
9. **Routing Matters in MoE: Scaling Diffusion Transformers with Explicit Routing Guidance:** Yujie Wei, Shiwei Zhang, Hangjie Yuan, Yujin Han, **Zhekai Chen**, Jiayu Wang, Difan Zou, Xihui Liu, Yingya Zhang, Yu Liu, Hongming Shan. Accepted at ICLR 2026.
10. **DreamVideo-Omni: Omni-Motion Controlled Multi-Subject Video Customization with Latent Identity Reinforcement Learning:** Yujie Wei, Xinyu Liu, Shiwei Zhang, Hangjie Yuan, Jinbo Xing, **Zhekai Chen**, Xiang Wang, Haonan Qiu, Rui Zhao, Yutong Feng, Ruihang Chu, Yingya Zhang, Yike Guo, Xihui Liu, Hongming Shan. Submitted to TPAMI.

Patents

1. A method of controllable generating video frame insertion based on diffusion model: Hao Chen, Wen Wang, **Zhekai Chen**, Chunhua Shen. Application Number 2024112552073.
2. A general zero-sample image synthesis method based on pre-training diffusion model: Hao Chen, **Zhekai Chen**, Wen Wang, Chunhua Shen. Application Number 2024108560819.
3. An image matting enhancement method and device based on dense annotation synthetic data: Kangyang Xie, **Zhekai Chen**, Hao Chen, Chunhua Shen. Application Number 2023115470403.

DEVELOPMENT PROJECT

QSC Mobile V4

This project is the fourth version of the Qiushichao mobile application, designed for convenient campus life. We built a brand new version under the Flutter framework. Specifically, we implemented full support for academic information, including class schedules, grade inquiries, and schedule arrangements, and through integration with the academic system, enabled automatic information updates. **There are 17k users at peak.**

OPEN SOURCE CONTRIBUTIONS

FaceChain (Github 9.1k stars)

A portrait synthesis tool based on diffusion models developed by Alibaba DAMO Academy. I made great contributions to the framework of the project and was recommended by the official for the competition.

LMFlow (Github 8.3k stars)

A convenient project for fine-tuning, developed by HKUST, I contributed to the diffusion part.

VOLUNTEERING

Reviewer, ICLR 2025/2026

Reviewer, ECCV 2026

Reviewer, TPAMI

SKILLS

Coding

- Languages: Python, C++, Go, C, Dart, HTML, CSS, JavaScript
- Frameworks: Pytorch, Django, Flask, Vue, Flutter
- Tools: Git, MySQL, SQLite
- Platforms: Linux, Windows, Mac, Raspberry Pi

Language

- Chinese: Native
- English: TOEFL iBT Best Score 96 (R 28, L 23, S 20, W 25)