Zhekai Chen
陈哲恺

✉ Chenzhekai@zju.edu.cn
✉ zkchen66@outlook.com
🗨 Azilly_
📞 (+86) 18796807327
⌂ Aziily
🎂 2003-8

EDUCATION

Zhejiang University, College of Computer Science and Technology   2021.09 - Present

  I am currently pursuing a bachelor's degree in Computer Science and Technology, with a primary focus on generative models.

  During my university years, I have achieved a GPA of **3.94/4.0** and have received the Zhejiang University Third Prize Scholarship in both my freshman and sophomore years.

Suzhou High School of Jiangsu Province, Key Class   2018.09 - 2021.06

  The comprehensive teaching methods of Suzhou High School promoted my overall development.

EXPERIENCE

Research: National Key Laboratory of CAD & CG, Zhejiang University, Prof. Chunhua Shen's Group   2022.10 - Present

- Paper1: ***FreeCompose: Generic Zero-Shot Image Composition with Diffusion Prior***

  Accepted by ECCV2024. First Author - arXiv:2407.04947

  In this research, we proposed a zero-shot image composition method based on diffusion models. By leveraging the priors of large-scale pre-trained diffusion models, and adjusting the self-attention layer and input conditions, we achieved multi-stage (object removal, object harmonization, and semantic object composition) image composition without training. Our method demonstrated that pre-trained diffusion model priors can be widely applied to downstream tasks, achieving excellent results on multiple datasets.

- Paper2: ***AutoStory: Generating Diverse Storytelling Images with Minimal Human Effort***

  Under review of IJCV. Fourth Author - arXiv:2311.11243

  In this research, we proposed a paradigm combining large language models and generative models (diffusion models). By inputting simple text from users, we use large language models to generate image descriptions and then generate images through diffusion models. This enables the entire pipeline to generate text-image pairs consistent with the user's required story. With customized processing, our method ensures consistency and diversity in generated images, with broad applications in downstream tasks like picture book generation.

- Patent1: ***A Portrait Matting Enhancement Method and Device Based on Densely Annotated Synthetic Data***

  Under review. Second Inventor - Application number 2023115470403

  This patent uses portrait images generated by generative models, densely annotated with methods like Ground-SAM, and through self-supervised learning, generates a large amount of portrait matting data for further training of portrait matting models. This patent's method effectively improves the generalization ability of portrait matting models and increases the accuracy of portrait matting.

- Patent2: ***A general zero-sample image synthesis method based on pre-training diffusion model***

  Under review. Second Inventor - Application number 2024108560819

  Corresponding patent for paper "FreeCompose".

Research: University of Illinois Urbana Champaign, Prof. Tong Zhang's Group   2024.5 - Present

- Paper1: ***Image Textualization: An Automatic Framework for Creating Accurate and Detailed Image Descriptions***

  Under review of NIPS2024. Fifth Author - arXiv:2406.07502

  In this research, we utilize information from expertise vision models like segmentation models and depth models to enhance the MLLM's ability of generating caption. We provides high-quality image-text pairs with stronger geometric information and the removal of hallucination, which achieves greater results in traditional benchmark and our new benchmark.

Development: Qiushichao, a Zhejiang University school-level student organization, Technical
Development Center   2022.3 - 2023.6

- Project1: *Qiushichao Mobile V4* Main Developer

  This project is the fourth version of the Qiushichao mobile application. In this project, we completely refactored the previous version using the Flutter framework, achieving cross-platform development, avoiding some information processing issues present in previous versions, and improving product stability and user experience. 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.

  As the main developer of this project, my contributions were designing the overall architecture, implementing most of the features, and identifying and fixing some potential issues during testing.

Development: Course Projects

- Project1: *IoT Middleware*

  In the BS Software Design course, I completed an IoT middleware project based on python+vue. This project receives simulated IoT device data, processes and saves it in the backend, and the frontend displays data by calling backend API interfaces. Specifically, I designed permissions, function layering, and visualization, packaged it with Docker, and ultimately implemented a complete IoT middleware. By simply replacing the backend simulated data with real data, this middleware can be directly used in practical tasks. I achieved a perfect score in this course.

- Project2: *miniOS*

  In the Operating Systems course, I implemented a miniOS system through a combination of assembly language and C++. In this system, I achieved kernel boot, clock interrupt, thread scheduling, virtual memory management, and user-mode programs, completing a relatively complete system. This gave me a better understanding of the design and implementation principles of the internal modules of an operating system. I also achieved a perfect score in this course.

OPEN SOURCE CONTRIBUTIONS

- LMFlow: A convenient project for LLM and Diffusion fine-tuning. Open source project of HKUST. I integrate the diffusion models into the original framework.
- FaceChain: A portrait synthesis tool based on diffusion models developed by Alibaba. I add functions to the gradio interface and optimize the design logic.
- Moore-AnimateAnyone: MooreThreads open-source project for generating character animations.

PROFESSIONAL SKILLS

Research Skills

- I have a thorough understanding and practical experience with generative models, especially diffusion models (both text-to-image and text-to-video models). I am familiar with several aspects of diffusion models, including training optimization, customized generation, sampling acceleration and so on.
- I have a certain understanding and practical experience with large language models. I am proficient in the basic principles of large language models and can apply them to specific tasks, such as image description generation and search problem-solving.
- I can skillfully use linux operation in server for environment configuration, code editing, process controlling and so on.

Programming Languages

Python, Flutter, Go, C++/C, Bash, HTML/CSS/JavaScript, etc. Listed in order of proficiency.

LANGUAGE SKILLS

- Chinese: Native.
- English: Fluent (TOEFL iBT: 94 total, R 28, L 24, S 18, W 24), capable of daily communication and literature reading.