Yuchao Gu

yuchaogu9710@gmail.com
http://guyuchao.github.io

EDUCATION

Ph.D. Electrical and Computer Engineering, National University of Singapore (exp.) Dec. 2025
Research Interests: Video Generation, Unified Understanding and Generation, Interactive World Model
Advisor: Prof. Mike Shou

M.Eng. Computer Technology, Nankai University 2022
Research Interests: Efficient Deep Learning, Low-Level Vision
Advisor: Prof. Ming-Ming Cheng

B.Eng. Computer Science and Technology, Beijing University of Chemical Technology 2019

WORKING EXPERIENCE

Research Intern, GenAI, Meta, US Jun 2023 – Oct 2024
Supervised by Yipin Zhou, Bichen Wu, Licheng Yu
- Worked on customized video editing and instance control.

Research Intern, PCG ARC Lab, Tencent, China Sep 2021 – Jun 2023
Supervised by Xintao Wang
- Worked on vector-quantized generative models for visual generation and face restoration.

INVITED TALKS

Toward Long-Context Vision Modeling, Kuaishou Technology (hosted by Xintao Wang) May 2025

AWARDS

Outstanding Graduates, Nankai University 2022

Nankai - SK Hynix Outstanding Research Scholarship, Nankai University 2022

National Scholarship, Ministry of Education, Government of China 2018

RESEARCH EXPERIENCE

Vision Generative Modeling Aug 2022 – present

Language generative modeling has demonstrated remarkable capabilities, whereas vision generative modeling still lags behind. One key reason is the high redundancy inherent in visual data, which reduces modeling efficiency. In this line of research, my work aims to explore effective vision compression to enable strong vision generative models. My work first reveals the reconstruction–generation dilemma in vision tokenizers and answers how to build an effective visual latent space in (*VQFR*, ECCV 2022 Oral and *RethinkVQ*, CVPR 2024). Additionally, I systematically investigate the elimination of context redundancy in video autoregression to enable long-context video modeling, as presented in (*FAR*, In submission). I also collaborate on developing the unified model for understanding and generation in (*Show-O*, ICLR 2025), as well as the video generation foundation model in (*Show-1*, IJCV 2025).

Controllable Visual Generation Aug 2022 – present

Language alone can provide only coarse and partial control in visual generation. For more precise and controllable results, users often require additional guidance from visual cues. In this line of research, my work has evolved from designing external modularity to investigating unified in-context learning for controllable visual generation. In the external modularity, I explore a wide range of user control requirements. This includes multi-instance identity control (*Mix-of-Show*, NeruIPS 2023), multi-instance spatial control (*ROICtrl*, CVPR 2025), and video instance identity control (*VideoSwap*, CVPR 2024). I also collaborate on developing video editing (*Tune-A-Video*, ICCV 2023) and video motion control (*Motion Director*, ECCV 2024 Oral). My research then evolve toward unifying multiple forms of visual control within an in-context learning framework by introducing visual context, as presented in my lead projects, (*Edit Transfer*, In submission).

PUBLICATIONS

[1] **Yuchao Gu**, Weijia Mao, Mike Zheng Shou. Long-Context Autoregressive Video Modeling with Next-Frame Prediction. *In submission*, 2025.

Highlight: Presented the first practical solution to reduce redundancy in long-context video modeling.

[2] **Yuchao Gu**, Yipin Zhou, Yunfan Ye, Yixin Nie, Licheng Yu, Pingchuan Ma, Kevin Qinghong Lin, Mike Zheng Shou. ROICtrl: Boosting Instance Control for Visual Generation. In: ***CVPR***, 2025.

[3] Jinheng Xie, Weijia Mao, Zechen Bai, David Junhao Zhang, Weihao Wang, Kevin Qinghong Lin, **Yuchao Gu**, Zhijie Chen, Zhenheng Yang, Mike Zheng Shou. Show-o: One Single Transformer to Unify Multimodal Understanding and Generation. In: ***ICLR***, 2025.

Highlight: Presented the first comprehensive approach to unify understanding and generation within a single model.

[4] Rui Zhao, **Yuchao Gu**, Jay Zhangjie Wu, David Junhao Zhang, Jia-Wei Liu, Weijia Wu, Jussi Keppo, Mike Zheng Shou. MotionDirector: Motion Customization of Text-to-Video Diffusion Models. In: ***ECCV (Oral)***, 2024.

Highlight: Introduced the first method for motion customization in video generative models.

[5] **Yuchao Gu**, Yipin Zhou, Bichen Wu, Licheng Yu, Jia-Wei Liu, Rui Zhao, Jay Zhangjie Wu, David Junhao Zhang, Mike Zheng Shou, Kevin Tang. VideoSwap: Customized Video Subject Swapping with Interactive Semantic Point Correspondence. In: ***CVPR***, 2024.

Highlight: Proposed the first framework for point trajectory control in video generative models.

[6] **Yuchao Gu**, Xintao Wang, Yixiao Ge, Ying Shan, Xiaohu Qie, Mike Zheng Shou. Rethinking the Objectives of Vector-Quantized Tokenizers for Image Synthesis. In: ***CVPR***, 2024.

Highlight: Provided the first in-depth analysis of the reconstruction-generation dilemma in vision tokenizers.

[7] Jay Zhangjie Wu, Yixiao Ge, Xintao Wang, Stan Weixian Lei, **Yuchao Gu**, Yufei Shi, Wynne Hsu, Ying Shan, Xiaohu Qie, Mike Zheng Shou. Tune-a-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation. In: ***ICCV***, 2023.

Highlight: Established the first diffusion-based framework for video editing.

[8] **Yuchao Gu**, Xintao Wang, Jay Zhangjie Wu, Yujun Shi, Yunpeng Chen, Zihan Fan, Wuyou Xiao, Rui Zhao, Shuning Chang, Weijia Wu, et al. Mix-of-Show: Decentralized Low-Rank Adaptation for Multi-Concept Customization of Diffusion Models. In: ***NeurIPS***, 2023.

Highlight: Introduced the first LoRA fusion framework for compatible multi-subject customization in diffusion models.

[9] **Yuchao Gu**, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, Ming-Ming Cheng. VQFR: Blind Face Restoration with Vector-Quantized Dictionary and Parallel Decoder. In: ***ECCV (Oral)***, 2022.

Highlight: Conducted the first study on VQ tokenizer properties for low-level image restoration.

[10] **Yuchao Gu**, Shang-Hua Gao, Xu-Sheng Cao, Shao-Ping Lu, Peng Du, Ming-Ming Cheng. iNAS: Integral NAS for Device-Aware Salient Object Detection. In: ***ICCV***, 2021.

[11] **Yuchao Gu**, Lijuan Wang, Yun Liu, Yi Yang, Yu-Huan Wu, Shao-Ping Lu, Ming-Ming Cheng. DOTS: Decoupling Operation and Topology in Differentiable Architecture Search. In: ***CVPR***, 2021.

[12] **Yuchao Gu**, Lijuan Wang, Zi-Qin Wang, Yun Liu, Ming-Ming Cheng, Shao-Ping Lu. Pyramid Constrained Self-Attention Network for Fast Video Salient Object Detection. In: ***AAAI***, 2020.