Jianhao Zeng

jh_zeng@tju.edu.cn | zengjianhao.github.io | Google Scholar

Education

Tianjin University Tianjin, China
M.S. in Electronic and Information Engineering 2021/09 – 2024/06
Advisor: Prof. Dan Song

Tianjin University Tianjin, China
B.Eng. in Mechanical Design & Manufacturing and Their Automation 2017/09 – 2021/06

Research Interests

I am broadly interested in computer vision and multi-modal learning, especially generative models and their application, including video generation, image generation and 3D content generation. I have extensively explored 2D virtual try-on and text-to-video generation. Additionally, automatic 3D content generation is crucial for building virtual worlds, so I am also interested in high-quality 3D content generation.

Publications and Manuscripts

[P.1] **Fashion Customization: Image Generation Based on Editing Clue**
Dan Song, **Jianhao Zeng**, Min Liu, Xuanya Li, Anan Liu#
IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)

[P.2] **CAT-DM: Controllable Accelerated Virtual Try-on with Diffusion Model**
Jianhao Zeng, Dan Song#, Weizhi Nie, Hongshuo Tian, Tongtong Wang, Anan Liu#
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024)

[P.3] **Better Fit: Accommodate Variations in Clothing Types for Virtual Try-on**
Dan Song, Xuanpu Zhang, **Jianhao Zeng**, Pengxin Zhan, Qingguo Chen, Weihua Luo, Anan Liu#
IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)

[M.1] **BooW-VTON: Boosting In-the-Wild Virtual Try-On via Mask-Free Pseudo Data Training**
Xuanpu Zhang, Dan Song, Pengxin Zhan, Tianyu Chang, **Jianhao Zeng**, Qingguo Chen, Weihua Luo, Anan Liu#
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2025) In submission

[M.2] **Robust-MVTON: Learning Cross-Pose Feature Alignment and Fusion for Robust Multi-View Virtual Try-On**
Nannan Zhang, Yijiang Li, Dong Du, Zheng Chong, Zhengwentai Sun, **Jianhao Zeng**, Yusheng Dai, Zhenyu Xie, Hairui Zhu, Xiaoguang Han#
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2025) In submission

[M.3] **FocusDiT: Masking Queries in Diffusion Transformers for Fine-grained Image Generation**
Xueji Fang, **Jianhao Zeng**, Zeyu Wu, Mingyuan Zhou, Liyuan Ma, Guojun Qi#
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2025) In submission

Research Experiences

Laboratory for MAchine Perception and LEarning (MAPLE), Westlake University Hangzhou, China
Research Assistant 2024/06 – Current
Advisor: Dr. Liyuan Ma, Dr. Zhiyang Chen and Prof. Guojun Qi (Fellow of IEEE, IAPR and AAIA)

- A text-to-image generation model called FocusDiT. It applies a Masking scheme to focus on critical query tokens that are exclusively fed into FFN, which were submitted to CVPR 2025. [M.3]
- The video generation model SnapVideo has been successfully replicated.

Institute of Television and Image Information, Tianjin University Tianjin, China
Graduate Student 2021/09 – 2024/06
Advisor: Prof. Dan Song and Prof. Anan Liu

- A novel framework for generating customized fashion images. This framework enables users to create tailored fashion visuals by providing multi-modal editing clues, which were accepted to TCSVT. [P.1]
- A model called CAT-DM based on ControNet and PBE for virtual try-on. This model utilizes the implicit distribution generated by a pre-trained GAN-based model to initiate the reverse denoising process. CAT-DM not only retains the pattern and texture details of the in-shop garment but also reduces the sampling steps without compromising generation quality, which were accepted to CVPR 2024. [P.2]
- An adaptive mask training paradigm that dynamically adjusts training masks for virtual try-on. It not only improves the alignment and fit of clothing but also significantly enhances the fidelity of virtual try on experience, which were submitted to TCSVT. [P.3]
- A mask-free virtual try-on diffusion model called BooW-VTON. It generates realistic try-on results without requiring any additional parser, which were submitted to CVPR 2025. [M.1]
- A Multi-View Try-On method called Robust-MVTON. It generates robust and high-quality multi-view ry-on results using front- and back-view clothing inputs, which were submitted to CVPR 2025. [M.2]

Competitions

- **Top 6.9%** in Jiangsu Meteorological AI Algorithm Challenge 2022/06
- **First Prize** in Tianjin University Undergraduate Physicists Tournament (TJUPT) 2019/08
- **Second Prize** in National College Students Mathematical Competition 2018/10
- **Third Prize** in Tianjin College Student Mathematics Competition 2018/05

Awards

- **CVPR Registration and Travel Support** 2024
- **Excellent Master's Degree Thesis of Tianjin University** (Top 5%) 2024
- **Tianjin University Academic Scholarship** 2021, 2022, 2023

Others

- **Reviewer**: ACM MM (2024), ICLR (2025), CVPR (2025)
- **Teaching Assistant**: Digital Logic Circuit, Tianjin University
- **Translation**: Physically Based Rendering: From Theory To Implementation, fourth edition
- **Patent**: A Fashion Image Editing Method and Device Based on Self-Attention Mechanism (CN115082295B)

Skills

- **Programming Languages** C, C++, Python, HTML, CSS, JavaScript
- **Frameworks** PyTorch, PyTorch Lightning, Accelerate
- **Tools** Linux, Git, LaTeX, Typst
- **Human Languages** Mandarin, English (TOEFL iBT: 94)