Chenyi Zhuang

chenyi.zhuang@nuaa.edu.cn (+86)19883840507 Homepage

EDUCATION

Nanjing University of Aeronautics and Astronautics **September 2022 - April 2025 (expected)**
Master of Engineering in Computer Science *Nanjing, CHN*

- Average Score: 90.66/100 GPA: 3.64/4.0
- Core Modules: Python For Data Science (100), Computer Vision and Artificial Intelligence (94), Mathematical Foundations in Information Security (93), Advanced Engineering Mathematics (92), etc.

Zhejiang Sci-Tech University **September 2018 - June 2022**
Bachelor of Engineering in Digital Media Technology *Hangzhou, CHN*

- Average Score: 90/100 GPA: 4.05/5
- Core Modules: Linear Algebra (99), Audio and Video Signal Processing (95), Digital Image Analysis and Artistic Processing (95), Computer Graphics (94), Discrete Mathematics (92), etc.

PUBLICATIONS

- ***Chenyi Zhuang***, Ying Hu, Pan Gao. Magnet: We Never Know How Text-to-Image Diffusion Models Work, Until We Learn How Vision-Language Models Function. (Accepted to NeurIPS 2024). [*arXiv*] [*code*]
- ***Chenyi Zhuang***, Pan Gao, Aljosa Smolic. (2023). StylePrompter: All Styles Need Is Attention. In Proceedings of the 31st ACM International Conference on Multimedia (pp. 2487-2497). [*arXiv*] [*code*]
- Qingguo Liu, ***Chenyi Zhuang***, Pan Gao, Jie Qin. (2024). CDFormer: When Degradation Prediction Embraces Diffusion Model for Blind Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7455-7464). [*arXiv*] [*code*]
- Ying Hu, ***Chenyi Zhuang***, Pan Gao. DiffuseST: Unleashing the Capability of the Diffusion Model for Style Transfer. (Accepted to ACM Multimedia Asia 2024)
- Ying Hu, ***Chenyi Zhuang***, Pan Gao. StyTips: Style Transfer via Transformer Filtering Prompts. (Under Review)

RESEARCH EXPERIENCES

Generative Models for Image Synthesis, Inversion, and Editing (Research Assistant) *Sep. 2022 - Present*

- Explored generative models (e.g., GANs and diffusion models) and identified their challenges in unconditional or conditional synthesis, and editing on real images, particularly for human faces.
- Investigated the compositional understanding of vision-language models and analyzed how the CLIP text encoder lacking compositional understanding affected text alignment for diffusion-based generation.
- Designed and implemented a novel framework that manipulates the text embedding of each object to enhance disentanglement between concepts in a training-free manner, addressing the attribute binding issue.

Diffusion-based Multi-modal Visual-guided Style Transfer Application (Research Leader) *Sep. 2023 - Present*
The Nanjing University of Aeronautics and Astronautics school-level innovation project

- Navigated comprehensive literature on style transfer and personalized image generation, including image-driven or text-based synthesis and editing, which build upon pre-trained text-to-image diffusion models.
- Developed an efficient network to perform controllable image generation for the style transfer task in a training-free manner, combining textual and spatial features, and separating the injection in different denoising steps.
- Interpreted the learning of designed injection modules through feature visualization in Fourier space, which verifies the enhancement of high-frequency information after injection, and advised extension to text-based editing.

AWARDS & HONORS

- The honor of research innovation individual, Nanjing University of Aeronautics and Astronautics, 2023.
- The honor of merit graduate student, Nanjing University of Aeronautics and Astronautics, 2023.
- The first price of academic scholarship, Nanjing University of Aeronautics and Astronautics, 2023.
- The provincial outstanding graduate student, Zhejiang Sci-Tech University, 2022.

SKILLS & INTERESTS

Technical Skills	Proficient in Python, C#, C/C++, Java, LaTex, Microsoft Office, Linux.
Languages	Mandarin (native), English (IELTS: 7), Japanese (basic).
Research Interests	Computer Vision, Machine Learning, Generative Model (e.g., GAN, Diffusion Model), Explainable AI, Vision-Language Model, Few-shot Learning.