Chenyi Zhuang

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

RESEARCH INTERESTS

Computer Vision, Generative Models (e.g., GANs and Diffusion Models), Vision-Language Models, Explainable AI, Multimodal Learning, Representation Learning from image, video, and 3D.

EDUCATION

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

- Average Score: 90.7/100 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).

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 (ranked #2/83)
- Core Modules: Linear Algebra (99), Audio and Video Signal Processing (95), Digital Image Analysis and Artistic Processing (95), Computer Graphics (94), Discrete Mathematics (92).
- Bachelor Thesis: Design and Implementation of Form Recognition and Reconstruction Algorithm Based on Image Processing (awarded as outstanding graduation thesis)

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). *equal contribution. *[arXiv] [code]*

Ying Hu, **Chenyi Zhuang**, Pan Gao. StyTips: Towards High-Quality, Efficient and Controllable Style Transfer via Transformer Filtering Prompts. (under review)

PATENTS

Training-free method for text-image generation based on diffusion model
Chenyi Zhuang, Ying Hu, Pan Gao. *Chinese Patent.* CN118485074A

RESEARCH EXPERIENCE

Research Assistant September 2022 - Present
Generative Models for Image Synthesis, Inversion, and Editing

- 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 with inaccurate concept representations can affect text alignment.
- Designed a novel framework to address the attribute binding issue that manipulates the text embedding of each object to enhance disentanglement between concepts in a training-free manner.

Principal Investigator September 2023 - October 2024
Postgraduate Research & Practice Innovation Program of NUAA

- Developed a diffusion-based multi-modal visual-guided style transfer approach, combining textual and spatial features of images, and separating the injection in different denoising steps.
- Visualized the intermediate representations of two injection modules in both feature space and Fourier space to verify the enhancement of high-frequency information after injection.
- Built a web application with Gardio that integrates the above style transfer pipeline with a well-designed user interface and straightforward controllable UI components for non-technical users.

Principal Investigator July 2020 - September 2021
Student Service Outsourcing Innovation and Entrepreneurship Competition

- Developed an end-to-end document recognition system for mobile and web applications, specifically targeting the conversion of form images into spreadsheet files.
- Adopted a Differentiable Binarization (DB) module for text detection, and a Convolutional Recurrent Neural Network (CRNN) for text recognition, both using a light-weight backbone MobileNetV3 which is fine-tuned with our hand-labeled data.
- Proposed a novel form reconstruction framework that utilized the Canny algorithm and morphological transformation to extract and operate the form edge information.

AWARDS & SCHOLARSHIPS

Honor of Innovation Research Advanced Individual 2023
Nanjing University of Aeronautics and Astronautics, Nanjing, China

Honor of Merit Graduate Student 2023
Nanjing University of Aeronautics and Astronautics, Nanjing, China

First Price of School Academic Scholarship 2022 - 2024
Nanjing University of Aeronautics and Astronautics, Nanjing, China

Provincial Outstanding Undergraduate Student June, 2022
Zhejiang Sci-Tech University, Hangzhou, China

First Prize in Provincial Undergraduate Competition May 2022
9th Haikang Cup Student Service Outsourcing Innovation and Application Competition, China

Third Prize in National Undergraduate Competition September 2021
12th Student Service Outsourcing Innovation and Entrepreneurship Competition, China

Second Prize in Provincial Undergraduate Competition May 2021
17th Challenge Cup Extracurricular Academic and Technological Competition, China

Chinese Government Scholarship 2019, 2021
Zhejiang Province Government, China

SKILLS

Technical skills: Proficient in Python, C#, C/C++, Java, LaTex, Unity, 3DMax, Office, Linux.
Language skills: Mandarin - native; English - fluent (IELTS 7/6); Japanese - basic.

REFERENCES

Prof. Dr. Pan Gao ✉ pan.gao@nuaa.edu.cn
Nanjing University of Aeronautics and Astronautics, China
Prof. Dr. Aljosa Smolic ✉ aljosa.smolic@hslu.ch
Lucerne University of Applied Sciences and Arts, Switzerland