

# WENYI MO

E-mail: mowenyi@ruc.edu.cn | Phone: (86)15815489499 | Homepage | Google Scholar

## EDUCATION

---

### Renmin University of China

*M.S. in Artificial Intelligence*

- Advisor: Prof. Bing Su.

Beijing, China

Sep. 2022 – Jun. 2025 (Expected)

### South China University of Technology

*B.E. in Computer Science*

- GPA: 3.92 / 4.0; Rank: 3 / 150

Canton, China

Sep. 2018 – Jun. 2022

## RESEARCH INTERESTS

---

My research interests center on multimodal learning, particularly improving image controllability under text-based conditions and enhancing user alignment in generative models. Recently, I have concentrated on generative models (such as autoregressive models and diffusion models) and their applications, including text-to-image generation and personalized content creation through visual-language understanding.

## PUBLICATIONS

---

- [1] **W. Mo**, T. Zhang, Y. Bai, B. Su, J.-R. Wen, Q. Yang, “Dynamic Prompt Optimizing for Text-to-Image Generation,” *CVPR*, 2024. [paper] [code]  
*TL;DR*: A reinforcement learning-based method for prompt optimization to improve text-to-image generation quality and user alignment.
- [2] **W. Mo**, T. Zhang, Y. Bai, J. Liu, B. Su, B. Li, J.-R. Wen, “Adaptive Preference Learning for Personalized Image Generation with Vision-Language Understanding,” *Under Review*, 2024. [paper]  
*TL;DR*: A Vision-Language Model framework for personalized image generation that uses latent preference prototypes to model shared and unique user preferences, improving accuracy in preference prediction.
- [3] **W. Mo**, T. Zhang, Y. Bai, B. Su, J.-R. Wen, “Uniform Attention Maps: Boosting Image Fidelity in Reconstruction and Editing,” *WACV*, 2025. [paper] [code]  
*TL;DR*: A tuning-free image editing technique that enhances fidelity in diffusion-based models using uniform attention maps.
- [4] J. Li\*, **W. Mo**\*, F. Song\*, C. Sun, W. Qiang, B. Su, C. Zheng, “Supporting Vision-Language Model Inference with Causality-pruning Knowledge Prompt,” *Neural Networks*, 2025. (\* equal contribution) [paper] [code]  
*TL;DR*: A confounder-pruned knowledge prompting method leveraging ontological knowledge graphs to improve vision-language model inference with enhanced semantic alignment and domain generalization.
- [5] J. Li\*, W. Qiang\*, Y. Zhang, **W. Mo**, C. Zheng, B. Su, H. Xiong, “MetaMask: Revisiting Dimensional Confounder for Self-Supervised Learning,” *NeurIPS*, 2022. **Spotlight** [paper] [code]  
*TL;DR*: A meta-learning-based approach that employs dimensional masks to address redundancy and confounders in self-supervised learning, enhancing downstream classification performance.

## RESEARCH EXPERIENCE

---

### Research Intern

Mar. 2024 – Present

*University of California, Santa Cruz*

*Remote*

- Supervisor: Prof. Cihang Xie
- Research focus: Vision-Language Learning

### Research Intern

Jan. 2024 – Mar. 2024

*ByteDance, Applied Machine Learning Group*

*Shanghai, China*

- Supervisor: Dr. Yongfei Liu and Dr. Quanzeng You
- Research focus: Controllable Video Generation with Conditional Inputs (e.g., Bounding Boxes)

### Research Intern

Sep. 2023 – Jan. 2024

*Du Xiaoman Technology*

*Beijing, China*

- Supervisor: Dr. Yalong Bai and Dr. Tianyu Zhang
- Research focus: Text-to-Image with Diffusion Model

## PROJECTS

---

### Prompt Optimizing for Text-to-Image Generation

Sep. 2023 – Jan. 2024

- Proposed the Prompt Auto-Editing (PAE) method to dynamically optimize text prompts in text-to-image generation using reinforcement learning.
- Introduced a two-stage training process: initial fine-tuning followed by reinforcement learning to automatically adjust prompt modifiers, effect ranges, and weights.
- Achieved an Aesthetic Score of 6.12 (0.05 above human performance) and a PickScore of 73.9% with PAE-generated images, outperforming human-written prompts by 1.4%.

### Image Reconstruction and Editing using Diffusion Model

Jan. 2024 - Sep. 2024

- Developed a tuning-free image editing method that enhances image reconstruction fidelity in diffusion-based models using uniform attention maps.
- Proposed an adaptive mask-guided editing technique to ensure consistency and precision during editing tasks.
- Achieved sota reconstruction performance with an SSIM of 0.839 and LPIPS of 0.041 on CelebA-HQ. Improved background consistency by 12.4% and editing accuracy by 1.4% compared to the baseline on the PIE benchmark.

## SELECTIVE SCHOLARSHIPS AND AWARDS

---

- China National Scholarship: Awarded to the top **1%** in the School of Computer Science. 2019
- China National Encouragement Scholarship: Awarded to the top **3%** in the School of Computer Science. 2021
- Renmin University of China Scholarship, 2024

## TEACHING EXPERIENCES

---

- Teaching Assistant: RUC, Comprehensive Artificial Intelligence Design, 2023 Fall
- Teaching Assistant: RUC, Artificial Intelligence and Python Programming, 2023 Summer

## PAPER REVIEWS

---

- **Conference Reviewer:** NeurIPS 2024, ICLR 2025, WACV 2025, AISTATS 2025, CVPR 2025, ICML 2025.

## TECHNICAL SKILLS

---

- Languages: Python, C/C++, LaTeX
- Frameworks: Pytorch