

Yuan Yao

yynobug@gmail.com ◊ 585-557-3646 ◊ [Google scholar](#)

EDUCATION

Shanghai Jiao Tong University (*ACM Honors Class*)

B.S., Computer Science, Zhiyuan Honorary Program, GPA: 3.9

Shanghai, China

09/2018 - 05/2022

University of Rochester

M.S., Computer Science

Rochester, NY

09/2022 - 05/2024

M.S., Data Science

08/2025 - 12/2025 *expected*

EXPERIENCES

PhD-level Researcher, University of Rochester

Research Assistant

Rochester, NY

09/2022 - 08/2025

- **Research Areas:** Diffusion models, Generative AI, Multi-modal learning, LLMs
- Foundation image/video generation models, including large-scale pretraining and cross-modal alignment with language models.
- Efficient pixel-space refinement methods to enhance boundary fidelity and reduce artifacts in generative editing.

Adobe Research

Research Scientist/Engineer Intern (2025)

San Jose, CA

06/2025 - Now

- Deployed **DiT-based multi-node video super-resolution** pipeline with optimized inference (tiling, localized shifted attention).
- Implemented **GRPO post-training** with reward models, improving perceptual quality and temporal stability of video outputs.

Research Scientist/Engineer Intern (2024)

06/2024 - 04/2025

- Built a novel generative model for efficient **4K image synthesis (published, patent-pending)**, achieving 2× faster inference.
- Proposed and deployed **multi-stage training pipeline** (teacher-forcing, distillation, SFT) on 200M+ image-text pairs **using FSDP**.

OPPO, US Research Center

Machine Learning Research Intern

Bellevue, WA

06/2023 - 12/2023

- Built **early multi-node video diffusion system** on top of Stable Diffusion 1.5, designing full pipeline from base model training to temporal interpolation and video upsampling, enabling temporally consistent, high-resolution outputs.
- Implemented scalable training and inference framework based on **PyTorch ddp**, supporting large-scale training.

Shanghai Artificial Intelligence Laboratory

Research Intern

Shanghai, China

06/2022 - 11/2023

- Developed **cross-modal pre-training framework** aligning **2D images, 3D point clouds, and language** via knowledge distillation from CLIP into 3D backbones, enabling unified multi-modal representations for 3D vision.
- Applied pretrained 3D foundation models to **diverse downstream 3D tasks** including classification, detection, segmentation, and retrieval, achieving **consistent +3–4% gains over strong baselines**, improving robustness of 3D real-world applications.

SELECTED PUBLICATIONS

■ PixPerfect: Seamless Latent Diffusion Local Editing with Discriminative Pixel-Space Refinement

Yuan Yao*, Haitian Zheng*, Yongsheng Yu, Yuqian Zhou, Zhe Lin, Jiebo Luo

NIPS 2025

Summary: Propose a **discriminative pixel-space refinement model** that removes artifacts in diffusion local editing, consistently **boosting quality across 11 leading models**, delivering state-of-the-art **artifact removal** without backbone modification.

■ Diffusion Transformer-to-Mamba Distillation for High-Resolution Image Generation ([link](#))

Yuan Yao, Yicong Hong, Difan Liu, Long Mai, Feng Liu, Jiebo Luo

BMVC 2025 **oral**

Summary: Propose the **teacher-forcing distillation** from DiT-based LDMs to state-space models with a multi-stage training strategy, achieving **efficient ultra-high-resolution image synthesis** with linear-time complexity.

■ 3D Point Cloud Pre-training with Knowledge Distillation from 2D Images ([link](#))

Yuan Yao, Yuanhan Zhang, Zhenfei Yin, Jiebo Luo, Wanli Ouyang, Xiaoshui Huang

ICME 2024

Summary: Propose a knowledge distillation based method for 3D point cloud backbones to directly learn from 2D pre-trained Vision-Language models through. Demonstrated consistent gains on downstream 3D tasks.

SELECTED TECHNICAL SKILLS

Systems and Engineering: Large-scale distributed training & optimization experience (multi-node clusters, DDP/FSDP); Cloud & DevOps (AWS, Docker, Git); Model compression & acceleration (AMP, distillation).

Programming Languages: Python, C++, Java, MATLAB, R, SQL.

Libraries and Tools: PyTorch, TensorFlow, CUDA, NumPy, SciPy, HuggingFace, W&B, valgrind.