Nanye (Willis) Ma

✉ Email   🌐 Website   in Linkedin   ⦿ Github

Education

New York University   **New York, NY**
Honors B.A. in Mathematics; B.A. in Computer Science with High Honors; 3.9/4.0 GPA   *September 2020 - May 2024*
Ph.D. in Computer Science   *September 2024 - Expected May 2029*

Publications / Preprints

- **N. Ma**, M. Goldstein, M. S. Albergo, N. M. Boffi, E. Vanden-Eijnden, S. Xie. SiT: Exploring Flow and Diffusion-based Generative Models with Scalable Interpolant Transformers. *ECCV 2024.*

Internship Experiences

Student Researcher, *Google*   **Jan 2024 - Present**
- Contributed to the `Instruct-Gen` team and led the modeling research.
- Extended the foundational `Flax` code of Diffusion Transformers to enable video inputs and identified a **8%** improvement in FVD score when benchmarking against the Masked Video Transformer.
- Conducted extensive ablation studies and identified significant design choices in video sampling process, enabled high-quality and temporal-consistent generation of 81-frames videos for under 10 seconds.
- Explored different training strategies to extend the video length to more than 200 frames.
- Coordinated with Senior Engineers and Researchers in validating and further improving the models with detailed report and performance analysis on a bi-daily basis.

Research Intern, *NYU*   **May 2023 - Jan 2024**
- Led the research on Scalable Interpolant Transformers(SiT), a novel generative framework built on dynamical transport.
- Engineered the framework single-handedly in `JAX` and `Flax` with over 120k lines of code.
- Conducted a thorough study on the SiT design space, collected and analyzed performance of different components, and identified specific configurations that yielded a **19%** improvement in FID score on ImageNet.
- Innovated a novel backward SDE construction that allows customized diffusion coefficients, enhancing flexibility and controllability in the sampling process while also leading to an extra **5%** performance boost.
- Organized weekly sessions to sync with the research team, present detailed update and findings, and coordinate the upcoming goals and action items.
- Submitted paper and accepted by *ECCV 2024* as the first author.

Projects

Text-to-Image Scalable Diffusion Models with Transformers   **Feb 2023 - May 2023**
- Integrated a pre-trained DistilBERT encoder with Diffusion Transformers (DiT) in `PyTorch` to enable rapid and high-quality text embedding.
- Achieved a **78%** reduction in memory consumption by gradient checkpointing and pre-extracted Variational Autoencoder(VAE) image features, enhancing training speed to 0.32 iters/sec on a single A100 GPU.
- Fine-tuned text-to-image DiT from pre-trained class-conditional DiT checkpoint by unfreezing `Embedders` and inject randomly initialized weights to `adaLN` modules.
- Benchmarking on MS-COCO, the fine-tuned model attained a competitive FID score of **15.49** with the LDM-KL-8-G model with only 200K training steps.

Dream Diffusion   **Jan 2023 - May 2023**
- Orchestrated the development of a full stack application using `Next.js`, designed specifically for recording dreams and writing dream journals with DallE backbone.
- Engineered session management from scratch to handle authentication and authorization, utilizing `Redis` as session store for efficient information retrieval.
- Designed robust authentication middleware using `Passport.js` and JWT-based strategies to bolster application security and safeguard user privacy.
- Optimized Serverless API architecture to support large-size (1024×1024) images transfers between frontend and backend, maintaining a low latency of up to 120ms for seamless user experience.

Technical Skills

Coding Languages: Python, JavaScript, C++, Java, C, Standard ML, Scheme, Prolog, Assembly, Julia
Tools/Frameworks & Libraries: Git, Linux, Mercurial / JAX, Flax, Optax, PyTorch, Numpy, Scipy, Pandas, Node.js, React, Express, MongoDB, Next.js