Hairui Yin

College Park, MD | yinhr@umd.edu | https://www.linkedin.com/in/hairui-yin-76b76831a/

Education

University of Maryland–College Park, MD   09/2024 - 05/2026
Master of Science, Data Science

Shanghaitech University, Shanghai   09/2020 - 06/2024
Bachelor of Engineering, Computer Science and Technology

Skills

Programming: Python, C++, C, MATLAB, SQL, JSON, Markdown
Common Tools: PyTorch, Cuda, OpenGL, Scikit Learn, OpenCV, Pillow, Hugging Face, Docker, Git, Visual Studio Code, Jupyter Notebook, Gitlab, Ubuntu Server

Professional Experience

Assistant Data Enginner | Glodon - Shanghai, China   01/2024 - 07/2024
- Enhanced the performance of camera-based construction site safety monitoring systems by utilizing **object detection models** (YOLO, Faster RCNN), achieving human-eye level accuracy and improving system efficiency.
- Conducted a data processing pipeline, including noise reduction, normalization, and augmentation (e.g., rotation, flipping, and cropping), resulting in a **20% improvement** in model accuracy during fine-tuning.
- Designed and implemented a scalable data generation pipeline leveraging **Blender**, **3D point cloud models**, and **OpenCV**, reducing data collection costs through the integration of data.

Security Engineer Intern | NSFOCUS – Shanghai, China   06/2022 - 08/2022
- Implemented robust data validation and preprocessing workflows to ensure data integrity and prevent **SQL injection** risks in **database interactions**.
- Developed secure data storage and transmission protocols, including **encryption** and **hashing techniques**, to safeguard sensitive information.

Research

Multi-modal data-driven extraction of genealogy images   09/2023 - 05/2024
Undergraduate Research Assistant | Advised by Prof. Haipeng Zhang at Shanghaitech University
- Constructed 2.8TB dataset and developed a multimodal model utilizing **OCR**, **ResNet**, and **LLMs** (from **Huggingface**) to extract valuable information (e.g., birth dates, locations, death dates) from book images.
- Conducted demographic and sociological analysis, visualizing insights to uncover historical trends and patterns.

Structural Causal model based Diffusion   02/2025 - Present
Advised by Prof. Abdirisak Mohamed at University of Maryland

Projects

CLIP-DB: A Database-Powered Semantic Image Search Engine
- Designed a semantic search engine using **CLIP** to map images / text to a latent space to allow cross-modal retrieval.
- Implemented database storage with **Redis** for efficient indexing and retrieval of high-dimensional embeddings, supporting semantic matching across large datasets like MS COCO.

CUDA-accelerated Fluid Simulation
- Built a fluid simulation system in **C++** using WCSPH, modeling weak compressibility, density, pressure, viscosity, surface tension, and external forces.
- Leveraged **CUDA parallel** computing to accelerate particle state updates, achieving **20x improvement** in simultaneous particle computations compared to CPU-based implements.
- Integrated **OpenGL** for real-time rendering and visualization of fluid interactions.

Diffusion and U-net Based Handwritten Digit Generation
- Developed diffusion models (**DDPM & DDIM**) with **U-Net** architecture on MNIST, designing distinct reverse processes leveraging DDPM's Markovian and DDIM's non-Markovian properties for controllable image synthesis.
- Implemented custom noise schedules and reverse diffusion functions, achieving FID scores ≤ 10, with DDIM showing **5x faster sampling** than DDPM while maintaining comparable generation quality.