Hairui Yin

College Park, MD | (202)841-6426 | 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 GPA: 4.0/4.0

University of Wisconsin–Madison, UW 01/2023 - 06/2023
Exchange student, Computer Science GPA: 3.8/4.0

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 03/2024 - 06/2024

- 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 by **40%** 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

Research Assistant | Shanghaitech University 09/2023 - 05/2024
Data Extracting and Mining on Genealogy Records | Advised by Prof. Haipeng Zhang

- Constructed 2.8TB dataset from library and other resource platforms using web scraping techniques.
- Employed a combination of **signal processing**, **deep learning models**, **OpenCV** for image processing to remove watermarks and reduce noise. Utilized regular expressions to accurately extract key metadata from textual data.
- Developed a multimodal model utilizing **OCR**, **ResNet**, and **Large Language Models** (Qwen, Llama) to extract valuable information (e.g., birth dates, locations, death dates) from book images.
- Automated structured entity extraction using **LLMs** from **Huggingface**, designed prompt to export structured outcomes, streamlining data processing workflows.
- Conducted advanced demographic and sociological analysis, visualizing insights to uncover historical trends and patterns.

Projects

CUDA-accelerated fluid simulation | Computer Graphics Course at Shanghaitech University

- 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, enabling instant, high-fidelity simulations.

CACoin Mining Hash Function Optimization | Computer Architecture Course at Shanghaitech University

- Utilized **SIMD instructions** to vectorize multiplication and addition, enhancing performance for batch operations.
- Applied **loop unrolling** in main loop to minimize control overhead.
- Optimized data access order in memory using **cashe blocking**, reducing cache misses and improving data locality.
- Ranked first 3 in the acceleration competition with **31x speedup**.