Jie Ren

jier@andrew.cmu.edu • (608) 867-7963 • https://www.linkedin.com/in/jie-ren-86b066212/

EDUCATION

Carnegie Mellon University — Pittsburgh
Master of Science in Information Networking — August 2025 - May 2027
- GPA: 3.7/4.0

University of Wisconsin-Madison — Wisconsin
Bachelor of Science in Computer Science — September 2023 - May 2025
- GPA: 4.0/4.0

PUBLICATIONS

[1] Jie Ren*, Pin-Yu Chen, Ren Wang. "Revisiting Mode Connectivity in Neural Networks with Bezier Surface," International Conference on Learning Representations (ICLR), 2025.

PROJECTS

Video Inpainting for Subtitles

University of Wisconsin-Madison Advisor: Prof. Yong Jae Lee
- Developed an end-to-end video inpainting pipeline leveraging the ProPainter framework to detect and remove hard-coded subtitles from Ultra-High-Definition video streams
- Engineered a synthetic data generation engine to create a large-scale dataset (50k+ images, 4 languages) of high-quality frames with diverse subtitle styles, solving data scarcity problem for supervised segmentation tasks
- Designed a two-stage vision detection framework, first performs high-precision semantic segmentation on text overlays, then utilizes flow-guided propagation to fill missing regions with temporally consistent textures

Dynamic Memory Allocator (Malloc CMU 15-213)
- Implemented a high-throughput 64-bit memory allocator (malloc, free, realloc) using segregated free lists and best-fit placement to maximize memory utilization and execution speed
- Engineered a custom heap consistency checker to enforce invariants and utilized **GDB** to debug complex pointer arithmetic and memory alignment issues in a simulated 64-bit environment.

Large Language Model Systems (CMU 11-868) | Spring 2026
- CUDA-Accelerated DL Framework: Built a deep learning framework ("MiniTorch") from scratch, implementing computational graph, reverse-mode automatic differentiation, and custom CUDA kernels (Map, Zip, Reduce, Matrix Multiplication) to train neural networks
- Transformer Optimization: Incorporating a GPT-2 style Transformer (Multi-Head Attention, LayerNorm) with custom C++ CUDA extensions for Softmax and LayerNorm operations, benchmarking speedups against standard PyTorch baselines
- Distributed Training Infrastructure: Built a distributed training system supporting Data Parallelism (DDP) and Pipeline Parallelism, analyzing throughput (tokens/sec) and memory scaling across multi-GPU setups

RELEVANT COURSEWORK

Large Language Model Systems, Distributed system, Machine learning, Deep learning, Computer vision, Big data system, Introduction to computer system

EXPERIENCE

Illinois Institute of Technology: Illinois Tech - Chicago — Illinois, Chicago
RESEARCH ASSISTANT INTERN — March 2024 - January 2025
Mode Connectivity in Neural Networks using Bezier Surface

Advisor: Ren Wang
Paper Accepted at ICLR 2025: "Revisiting Mode Connectivity in Neural Networks with Bézier Surface"
- Designed a novel Bézier surface-based optimization framework to connect independent neural network optima, extending mode connectivity from 1D curves to 2D surfaces to improve model generalization
- Developed a scalable "outer-to-inner" training algorithm optimizing high-dimensional loss landscapes, achieving 92.0% accuracy on CIFAR-10 with VGG16 via surface-based ensemble, outperforming traditional ensemble methods by 1.6%
- Architected and validated framework across diverse architectures (ResNet18, VGG16, ViT) and datasets (CIFAR-10/100, Tiny-ImageNet), identifying low-loss basins with high accuracy

SKILLS

Frameworks & Libraries: PyTorch, vLLM, HuggingFace Transformers, OpenCV, NumPy, Pandas, Scikit-learn
Machine Learning: LLM Inference & Serving, Computer Vision, Semantic Segmentation, Model Ensembling
Programming Languages: Python, C/C++, CUDA, SQL, Bash/Shell, Go