Zifei Dong

Nashville, TN 984-261-5818 zifei.dong@vanderbilt.edu Linkedin Google Scholar

Education

Vanderbilt University **May 2024 – May 2026**
Master Of Data Science, GPA: 3.8/4 *Nashville, TN*

University of North Carolina at Chapel Hill **May 2021 - May 2024**
B.S. in Computer Science double major in Statistics and Analytics, Dean's List, GPA: 3.625/4 *Chapel Hill, NC*

Technical Skills

Programming: Python, C, SQL, Bash, R, Java, MATLAB
Machine Learning: PyTorch, TensorFlow, Transformers, Representation Learning, Generative Modeling, Self-Supervised Learning
Computer Vision & Imaging: Image Reconstruction, Denoising, Segmentation, Synthetic Data Generation, Robustness Evaluation, Annotation Noise Modeling
Data & Systems: NumPy, Pandas, Linux, Multiprocessing, Spark, Hadoop, SGLang, QLoRA

Publication

- **Dong Z**, Wu W, Hao J, Chen T, Weng Z, Zhou B.
  AnyCXR: Human Anatomy Segmentation of Chest X-ray at Any Acquisition Position using Multi-stage Domain Randomized Synthetic Data with Imperfect Annotations and Conditional Joint Annotation Regularization Learning.
  Medical Image Analysis, under review. Available on arXiv.
  Contribution: Led project conception, end-to-end pipeline design, synthetic data generation strategy, model development, and experimental implementation.

Experience

Codebook Denoising Research, AIMP Lab, Northwestern University **December 2025 – Present**
Lead Researcher, Remote *Evanston, IL*

- Developed a **paired medical image denoising and reconstruction** framework for low-dose imaging, targeting motion artifacts, structural distortion, and anatomically implausible outputs under degraded conditions.
- Designed a **VAE-based 2.5D architecture** with **Spatio-Temporal Attention (STA)** and a structured latent representation combining **Dirichlet-constrained relaxation** with grouped **FSQ**.
- Engineered a dual-stage **Conditional Flow Matching (CFM)** pipeline in latent and pixel spaces, with **Optimal Transport (OT)**-based refinement for improved structural consistency and texture recovery.
- Achieved **31+ PSNR** in reconstruction and **36+ PSNR** in denoising while improving fidelity and suppressing anatomically implausible reconstructions with lightweight, modular components.

AnyCXR Project, AIMP Lab, Northwestern University **April 2024 – Present**
Lead Researcher, Remote *Evanston, IL*

- Led the design and development of **AnyCXR**, a robust anatomy segmentation framework for **chest X-rays from arbitrary viewpoints**, addressing severe acquisition and domain shift beyond standard clinical views.
- Built a **3D multi-stage synthetic data generation pipeline** over **8TB** of data from **50,188 CT scans** to simulate diverse projection geometries and improve cross-view generalization.
- Developed a **U-Net-based model** with **Conditional Joint Annotation Regularization** to train effectively on imperfect and partially labeled large-scale datasets.
- Achieved **94% mean Dice** across **54 anatomical structures** and improved downstream disease classification by **2%** through a segmentation-guided DenseNet pipeline.

Alliance Bernstein–Vanderbilt Joint Project: Factor-Aware Financial NLP **Jan 2025 – Present**
Research Collaborator, Part-Time *Nashville, TN*

- Developed an **LLM-based pipeline** for extracting structured signals from long-form SEC filings, leading training and evaluation of domain-adapted models based on **Qwen3-14B**.
- Fine-tuned and served models using **QLoRA**, **Unsloth**, and **SGLang**, improving scalable experimentation, inference efficiency, and output consistency through reasoning supervision and alignment-based refinement.

Ningbo Jinge Liangrui Asset Management Co. Ltd. **June 2025 – August 2025**
Quantitative Research Intern *Shanghai, China*

- Built and optimized a large-scale time-series modeling pipeline over **10 years** of daily market data, accelerating matrix-heavy preprocessing and neural network training by **10×** with **NumPy** and **Numba**.