Yipeng Sun

Fröbelstr. 19, 91058 Erlangen, Germany | +49-162-701-8659 | yipeng.sun@fau.de
Website | GitHub | Google Scholar | LinkedIn

Experience

Ph.D. Candidate — FAU Erlangen-Nürnberg, Erlangen **since 07/2023**
- Designed and implemented custom neural network architectures (CNNs, Transformers, Diffusion Models) for medical image reconstruction, outperforming conventional methods in image quality and reconstruction speed.
- Developed diffct, an open-source CUDA-accelerated differentiable CT operator library using Numba, enabling GPU-parallel forward/back projection for deep learning-based reconstruction pipelines.
- Developed self-supervised methods under extreme data constraints for motion estimation and artifact compensation in CT workflows.
- Applied AI techniques including LLM fine-tuning, agent-based systems, and autoregressive models to real-world applications.

Skills: Python, PyTorch, CUDA, Numba, NumPy, CNNs, Transformers, Diffusion Models, Medical Imaging, CT Reconstruction, LLMs

Project Leader (KI4D4E) — FAU Erlangen-Nürnberg, Erlangen **since 07/2023**
- Led AI software architecture design and implementation for large-scale 4D tomography reconstruction (BMBF Grant 05D23WE1).
- Developed and integrated deep learning solutions for artifact reduction and image reconstruction, optimizing for GPU performance.
- Coordinated technical workstreams across 14 partners in an agile, cross-functional international team.

Skills: Python, PyTorch, CUDA, Deep Learning, GPU Computing, Docker, Git, HPC/SLURM, Software Architecture, Agile

Researcher — Fraunhofer EZRT, Fürth **since 07/2023**
- Developed and optimized deep learning models for artifact compensation in high-throughput CT systems, achieving near real-time processing through GPU-accelerated inference.
- Integrated AI solutions into production imaging pipelines, ensuring model performance optimization across different hardware.
- Enhanced model robustness and generalizability across diverse X-ray imaging modalities (synchrotron, neutron, industrial CT).

Skills: Python, PyTorch, CUDA, ONNX, Deep Learning, NumPy, GPU Inference Optimization, Model Deployment, Docker

Master's Thesis — Anki Lab, FAU Erlangen-Nürnberg, Erlangen **12/2022 – 05/2023**
- Designed custom Transformer encoder-decoder architectures optimized for hardware-accelerated inference on edge devices.
- Implemented automated Neural Architecture Search (NAS) using genetic algorithms, balancing accuracy with efficiency.
- Achieved hardware-deployable models through network topology optimization, quantization, and compilation for Google Edge TPUs.

Skills: Python, TensorFlow, Keras, TensorFlow Lite, NAS, Genetic Algorithms, Edge TPU, Quantization, AutoML

Research Intern — Anki Lab, FAU Erlangen-Nürnberg, Erlangen **06/2022 – 12/2022**
- Achieved 7× efficiency improvement for real-time medical image processing through model optimization for edge devices.
- Designed and deployed Compact Convolutional Transformer (CCT) architectures for Google Edge TPUs via TensorFlow Lite quantization.
- Published a literature review on deep learning deployment strategies for resource-constrained edge devices.

Skills: Python, TensorFlow, Keras, TensorFlow Lite, Edge TPU, Model Quantization, Deployment, Technical Writing

Research Intern — Shanghai Jiao Tong University, Shanghai **05/2021 – 07/2021**
- Applied deep neural networks (Deep Belief Networks) for medical diagnosis using multi-modal physiological signals.
- Coordinated with clinical staff to recruit participants and manage data acquisition for multi-modal physiological experiments.

Skills: Python, NumPy, Scikit-Learn, Deep Learning, Deep Belief Networks, EEG Signal Processing, Data Acquisition

Technical Skills

Languages: Mandarin Chinese (Native), English (C1), German (B2)
Programming: Python (NumPy, SciPy, Pandas, Numba), C/C++, MATLAB
AI/ML Frameworks: PyTorch (Expert), TensorFlow/Keras (Expert), TensorFlow Lite, ONNX
AI/ML Techniques: Deep Learning (CNNs, Transformers, Diffusion Models), NAS, Self-Supervised Learning, Generative Models
GPU/HPC: CUDA, NVIDIA GPUs, Google Edge TPUs, Model Optimization for Real-Time Inference, HPC/SLURM
Tools: Git, Docker, Weights & Biases, Linux/Unix Shell

Open Source Projects

- **ConvNeXt Perceptual Loss** – A PyTorch perceptual loss implementation based on the modern ConvNeXt architecture.
- **AutoCitation** – An LLM agent that finds and verifies real citations for written content.
- **Diagnostic Devils Advocate** – A multi-agent AI system that challenges clinical diagnoses to prevent cognitive bias errors.
- **claude-skill-codex** – A Claude Code skill that integrates OpenAI Codex CLI into Claude Code.
- **Light Whisper** – A local offline Chinese speech-to-text desktop application built with Rust.

Awards & Honors

- **BVM 2026** – Paper received top review scores, invited to submit to IJCARS Special Issue (02/2026)
- **Honor Graduate**, Nanjing University of Science and Technology (06/2019)

Education

Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) since 07/2023
Dr.-Ing. in Computer Science (Focus: AI in Medical Imaging) *Erlangen, Germany*

Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) 09/2020 – 06/2023
M.Sc. in Medical Engineering *Erlangen, Germany*

Nanjing University of Science and Technology (NJUST) 09/2015 – 06/2019
B.Eng. in Measurement and Control Technology and Instrumentation (Honor Graduate) *Nanjing, China*

Selected Publications

1. Mei, S., Thies, M., Xia, Y., **Sun, Y.**, Wu, F., Fan, F., Gu, M., Ye, C., Huang, Y., Christlein, V., Maier, A. (2026). Vision transformer Hook for dense predictions. *Pattern Recognition*, 113818.

2. **Sun, Y.**, Schneider, L.S., Gu, M., Mei, S., Bayer, S., Maier, A. (2025). Learning Wavelet-Sparse FDK for 3D Cone-Beam CT Reconstruction. In *International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2025.

3. **Sun, Y.**, Schneider, L.S., Mei, S., Wang, J., Hu, G., Gu, M., Ye, C., Wagner, F., Song, L., Bayer, S., Maier, A. (2026). Filter2Noise: a framework for interpretable and zero-shot low-dose CT image denoising. *J. Med. Imag.*, 13(2), 024004.

4. **Sun, Y.**, Schneider, L.S., Gu, M., Mei, S., Bayer, S., Maier, A. (2025). Compensating CBCT Motion Artifacts with Any 2D Generative Model. In *International Conference on Industrial Computed Tomography (iCT)*, 2025.

5. **Sun, Y.**, Huang, Y., Yang, Z., Schneider, L.S., Thies, M., Gu, M., Mei, S., Bayer, S., Maier, A. (2025). EAGLE: An Edge-Aware Gradient Localization Enhanced Loss for CT Image Reconstruction. *Journal of Medical Imaging*, 12(1), 014001.

6. **Sun, Y.**, Schneider, L.S., Fan, F., Thies, M., Gu, M., Mei, S., Zhou, Y., Bayer, S., Maier, A. (2024). Data-Driven Filter Design in FBP: Transforming CT Reconstruction with Trainable Fourier Series. In *International Conference on Image Formation in X-Ray Computed Tomography (CT Meeting)*, 2024.

7. **Sun, Y.**, Kist, A. (2023). Compact Convolutional Transformers on Edge TPUs. In *Bildverarbeitung für die Medizin (BVM) Workshop*, pp. 141–146.

8. **Sun, Y.**, Kist, A. (2022). Deep Learning on Edge TPUs: A Review.