YANXIANG FU (FELIX)

e1582277@u.nus.edu | (+65) 8965 3746 | Singapore

PROFILE

MComp student at the NUS School of Computing, with a background in GPU systems engineering and five years of industrial embedded development.

Currently completing my capstone project under Prof. Bingsheng He, alongside research internships in Prof. Shuicheng Yan's LV-Robotics Lab and the G&M Lab.

My research interests center on multimodal (vision-touch) world-action models for robotics, and on-device inference optimization.

RESEARCH INTERESTS

- World models and Vision-Language-Action (VLA) models: systematic study of the literature from RT-1 through π0, including imitation-learning policies (ACT, Diffusion Policy).
- Robot systems: hands-on exploration of the Unitree G1 platform, its DDS-based SDK, and dexterous-hand control stack.
- Efficient inference for real-time policy execution on physical robots.

EDUCATION

National University of Singapore, School of Computing **Aug 2025 – Present**

Master of Computing (General Track), expected graduation 2027

- Coursework: CS5242 Deep Learning, IT5005 Artificial Intelligence, IT5003 Data Structures and Algorithms, CS5224 Cloud Computing.

Zhejiang University, College of Electrical Engineering **Sep 2016 – Jun 2020**

Bachelor of Science in Automation (Electrical)

PROJECT EXPERIENCE

G-Watch: a GPU Profiler with AI-Driven Automatic Optimization **Jan 2026 – Present**

NUS Capstone Project

- Designed and implemented a GPU profiling framework in C++/Python (bridged via pybind11), exposing a ProfileContext Python API for seamless integration into existing LLM inference pipelines.
- Instrumented NVIDIA GPU hardware counters via CUPTI to capture fine-grained kernel-level metrics (execution time, memory throughput, compute utilization) across SGLang and vLLM workloads.
- Diagnosed and resolved CUPTI subscriber registration failures caused by SGLang's child-process architecture, requiring process-level isolation and careful GPU resource lifecycle management.
- Built an AI-agent-driven profiling workflow (Gemini, Claude) to automate bottleneck analysis; validated profiler correctness and stability across concurrent multi-request inference sessions.

Marine Life Classification with Deep Learning **Mar 2026 – Apr 2026**

CS5242 Course Project

- Built an end-to-end classification pipeline on 20k+ real-world underwater species images, including data cleaning, augmentation, and handling of class imbalance.
- Fine-tuned a Transformer-based ViT to 97.4% test accuracy and a CNN-based EfficientNet-B0 to 87% with far fewer parameters; analyzed the accuracy–efficiency trade-off between the two.

WORK EXPERIENCE

Lead Intelligent Equipment Co., Institute of Technology **Jul 2022 – Dec 2024**

Intermediate Embedded Hardware Engineer

- Independently designed ARM-MCU-based industrial products: RS485 communication terminals, a proportional-valve PID controller, and detection/control terminals for lithium-battery formation and aging.
- Served as project manager, coordinating cross-functional resources; products passed CE/TUV certification.

SUPCON Technology Co., PLC Product Sub-Center **Jul 2020 – Jun 2022**

Project Manager & Hardware Engineer (promoted from Intern Hardware Engineer)

- Led the end-to-end design and productization of the minimum system control board for PLCs based on domestic Loongson CPUs.
- Developed analog voltage acquisition, RTD measurement, and digital I/O modules for PLC and DCS product lines.

SKILLS

- Programming: Python (PyTorch), C/C++, CUDA, SQL | Tools: Linux, ROS, RT-Thread, Git, AI coding agents.
- Hardware: Cadence Allegro, Altium Designer, SolidWorks | Languages: Mandarin (native), English (working proficiency).