Leon Lin 林敬翔

Machine Learning Engineer

l84328g@gmail.com | linleon1995 | jing-siang-lin-488608184

Work Experience

Machine Learning Engineer at ASUS AIoT *Jul. 2021 - Present*

- **Mission Inference Optimization** | **Pytorch, TensorRT**
  - Optimized the inference pipeline by leveraging TensorRT, adopting a dynamic batching strategy, switching to asynchronous inference, and offloading computations to the GPU using custom CUDA kernels to maximize throughput.
  - As a result, the number of supported real-time video streams per Kubernetes node increased from 8 to **14** on NVIDIA T4 (**75%** improvement), potentially saving clients over **$30,000 USD** annually in infrastructure costs.
- **EHS General Mission Framework** | **Python, System Design**
  - Built a flexible and extensible CCTV AI inference system supporting multiple AI capabilities—including object detection, action recognition, and other skills—that can be easily integrated and executed through the unified framework.
  - Closely collaborated with a team of 3 colleagues to enhance the EHS mission pipeline, focused on effective teamwork and clear coordination of responsibilities. Responsible for progress tracking and contributed to hands-on development.
- **EHS Scheduler** | **Python, Kubernetes, MongoDB**
  - Contributed to the development of the EHS Scheduler to handle mission scheduling requests from the EHS Portal, orchestrating the full lifecycle of CCTV AI task pods on Kubernetes.
- **AI Agent Service** | **Azure AI, Pydantic**
  - Designed and implemented a unified API service to register and forward AI task requests from the EHS Portal, integrating third-party AI services such as Azure AI and Copilot, with robust request validation handled through Pydantic.
- **Machine Learning Algorithm Development** | **PyTorch, Detection, Pose Estimation**
  - Designed and trained a new version of the Protective Personal Equipment (PPE) detection algorithm to help ensure workplace safety in the factory. Achieved a 34.22% improvement in accuracy, increasing mAP from 0.301 to 0.404.
- **Snoring Detection** | **C++, PyTorch, TFLite**
  - Developed a snoring detection algorithm using PyTorch, with spectrogram transformation implemented in C++, achieving 94.12% accuracy on the testing dataset. The model was exported to TFLite format for deployment on mobile devices.
- **3D Lung Nodule Segmentation** | **PyTorch, 3D MaskRCNN**
  - Developed lung nodule segmentation and classification algorithms by using 3D MaskRCNN. Achieved an average segmentation DSC of 0.7481 and classification accuracy of 91.84% on hospital test datasets.

Projects

- **Quantitative Trading and Backtesting System** | **Kafka, Reinforcement Learning, Time-series database**
  - Developed a quantitative trading and backtesting system using Apchie Kafka and a time-series database, supporting real-time data, multiple strategies, and data sources; integrated reinforcement learning to optimize trading decisions.
- **IoT Event RAG System** | **LangChain, Llama.cpp, RAG**
  - Developed a benchmarking framework to compare vector-based and graph-based RAG systems on IoT event data using the Gemma 3 LLM (12B), showing Graph RAG's superiority in handling complex, structured queries through knowledge graph reasoning.

Education

Master in Electrical Engineering at National Chung Cheng University *Sep. 2017 –Nov.2020*

- Thesis: Prior guiding based multiple organ segmentation
- 4.17/4.3 GPA

International Exchange at Deggendorf University of Applied Sciences *Sep. 2019 –Feb. 2020*

- Taiwan Ministry of Education International Education Exchange.

Bachelor of Electrical Engineering at National Yunlin University of Science and Technology *Sep. 2013 –Jun. 2017*

Skills

Languages	Python, C/C++
Software/Frameworks	Pytorch, TensorFlow, TensorRT, GStreamer, MLflow, DVC
Developer Tools	Git, GitLab, Docker, Kubernetes, Azure AI