Tran Quang Khai

Ho Chi Minh tquangkhai0302@gmail.com +84-938-514-528 tqk3002 TQKhai3002

Summary

With a strong foundation in Machine Learning, Deep Learning, and Computer Vision, I aim to pursue graduate studies to deepen my theoretical understanding and learn to deploy and maintain machine learning systems in production. I am particularly interested in building automated data and machine learning pipelines.

Education

University of Information Technology VNUHCM Sep 2021 – Jun 2025

- Bachelor of Science – Computer Science
- Current GPA: 8.43/10
- Coursework: Artificial Intelligence, Computer Vision, Deep Learning, Data Mining.

Experience

Lab assistant

Multimedia Communications Laboratory
May 2024 – Present

- Conduct research on aesthetic-aware image cropping using Large Language Models such as GPT and Gemini, contributing to a forthcoming publication in 2025.
- Developed and published two datasets for aesthetic-aware image cropping, evaluated on those datasets to set a benchmark.

Technical Specialist

Vietnam-Singapore Industry 4.0 Innovation Center (VSIIC), Eastern International University

October 2025 – Present

- Participating in developing AI for HVAC systems to control the AC of the school, achieving the fan speed accuracy of 91%
- Integrating the Yolo11x model for people counting on the CCTV server for building management using Deepstream. Achieving real-time human detection on multiple cameras.

Projects

Aesthetic-Aware Image Cropping May 2024 - present

- **Project Description**: Build a pipeline for prompt-based image cropping using aesthetic-awareness.
- **Methods**:
  - Generative AI models like GPT and Stable Diffusion to generate synthetic datasets.
  - Leveraged models such as CACNet, SAM2, and Grounding DINO for image cropping.
- **Results:**
  - Published 2 datasets: 1,000 manually labeled images and 4,000 automatically labeled.
  - mDSC: 0.522 using Gemini-2.0-Flash, and 0.453 using GPT-4o mini. This establishes a new benchmark for future work.
- **Tools:**
  - Python, GPT, Stable Diffusion, TensorFlow, Gemini, SAM2, Grounding DINO.

Using Dynamic Time Warping Stock Market Prediction October 2025 - present

- **Project Description**: Using Nvidia Deepstream for real-time human detection on CCTV
- **Methods**:
  - Fine-tuning Yolo11x for maximum accuracy on human detection.
  - Convert the model into TensorRT to boost FPS, and apply batch processing.
  - Integrated MediaMTX to transcode RTSP into low-latency WebRTC.
- **Results:**
  - Optimized YOLOv11x inference using TensorRT FP16 precision, enabling the system to scale from 1 to 10 concurrent video feeds on a single GPU while maintaining 20 FPS real-time.
- **Tools:**
  - Python, Deepstream, Mediamtx, Caddy.

Skills and Technologies

Languages: C++, Python.

AI & Computer Vision: NVIDIA DeepStream, TensorRT, PyTorch, OpenCV, Ultralytics YOLO, CUDA.

DevOps & Deployment: Docker, Docker Compose, Git, Linux (Ubuntu/WSL).

Cloud Platforms: Azure (Familiar with compute & basic networking).

IELTS: 8.0 (L: 8.5, R: 9.0, W: 7.0, S: 6.5)