Nguyen Thanh Trung

✉: trungnt240602@gmail.com, ☎: (+84)-982666066

https://www.linkedin.com/in/nguyen-trung-7aa3a0221/

https://github.com/nttrung2406

Experience

AI Software Engineer, Viettel Solution   05/2023 – 01/2025
- **Project**: Tessel.AI (team size: 06) - Automatically scoring market images following display structure and reward structure of the business from sale-mans by utilizing Computer Vision and Image Processing.
  Key responsibility:
    - Developed and implemented backend APIs for computer vision functionalities.
    - Deployed and tested the YOLOv8 model to detect products, ensuring adaptability to multiple display rules.
    - Designed a robust training pipeline using Celery and RabbitMQ for efficient and reliable model training.
    - Set up Prometheus and Grafana to monitor the model training process, ensuring real-time tracking and performance analysis.
    - Researched, analyzed, and implemented solutions to resolve image uploading challenges.
    - Preprocessed images to enhance training data quality and improve model performance.

AI Engineer, Viettel High Tech   01/2025 - Now
- **Project**: An intelligent traffic devices that tracks violations and automates penalty issuance.
  Key responsibility:
    - Develop, train, and optimize deep learning models for license plate detection on edge devices.
    - Enhance model accuracy and efficiency for real-time processing in embedded camera systems.
    - Adapt models for household edge devices to support person re-identification tasks.

AI Software Engineer, AI4LI - Artificial Intelligence for Life Inclusion   12/2024 - Now
- **Project**: AI4li.org (team size: 10) - A non-profit initiative led by Dr.Pham Dinh Lam (Kyonggi University, Korea), leveraging AI to assist individuals with disabilities.
  Key responsibility:
    - Design and implement a microservice-based architecture for computer vision processing.
    - Optimize object detection and tracking models for real-time responsiveness to voice commands.
    - Improve system scalability and efficiency for seamless user interaction.

Research

Image Segmentation using text prompts
github.com/nttrung2406/Image_segmentation_using_text-prompts   02/2024–08/2024
- Expanded the CLIP model by integrating a transformer decoder and incorporating text constraints to generate segmentation masks.
- Fine-tuned the model to enhance performance, especially for complex object segmentation.

Achievements

Encrypted text classification (Viettel AI Challenge): Top 6 Leaderboard
github.com/nttrung2406/Encrypted_text_classification   November 2024
- Designed a solution for a text classification competition where data consists of encrypted numerical sequences.
- Applied NLP techniques and fine-tuned large language models (BERT variants, XLNet) to improve classification accuracy.

Projects

Chat-with-Paper RAG Application

github.com/nttrung2406/chat_with_paper   February 2025

An end-to-end Retrieval-Augmented Generation (RAG) application enabling interactive chatting with PDF documents, enhancing document comprehension and user engagement with Llama-3-8B.

AI Generated Detection

github.com/nttrung2406/AI_generated_detection   January 2025

– Developed a multi-modal detection system to identify AI-generated content, including images (ResNet18), text (BERT), and audio (self-implemented CNN).

COVID19 disease detection

github.com/nttrung2406/COVID19_disease_detection   July 2024

– Built a lightweight CNN model for COVID-19 detection using X-ray images.
– Optimized model performance by converting it to ONNX and TensorRT for faster inference.

Skills

Programming & Scripting: Python, C++, Bash, Javascripts, HTML, CSS
AI/ML: Tensorflow, Pytorch, Keras, CUDA, ONNX, TensorRT, NLTK, Triton Inference Server, Langchain
Software Engineer: OOP, SOLID principles, Flask, FastAPI, React, Selenium
Tools: Git, Docker, RabbitMQ, Prometheus, Postman, Figma, Canva
Database: MongoDB, Postgres, SQL Server, Milvus
Data Analytics: Pandas, Sklearn, Pyspark, Grafana, Tableau, D3, PowerBI

Education

University of Science, (HCMUS, Ho Chi Minh City, Vietnam)   2020–2024

– Department of Information Technology
– Specialization: Computer Science

Certification

– AWS Cloud Practitioner Essentials
– Advanced LLMs with Retrieval Augmented Generation (LinkedIn)
– DevOps Foundations: Continuous Delivery/Continuous Integration (LinkedIn)
– LLMOps in Practice: A Deep Dive (LinkedIn)
– Cloud and Configuration Management by Google (Coursera)
– Google Advanced Data Analytics (Coursera)
– Software Engineering Virtual Experience Program of Goldman Sachs
– English (Intermediate, IELTS 6.5)

References

– **Vo Thanh Cuong**   AI Team Leader of Viettel Solution
  Contact: cuongvt13@viettel.com.vn
– **Dr.Pham Dinh Lam**   Research Professor at Kyonggi University, Korea
  Contact: phamdinhlam@kyonggi.ac.kr