Binh Phan

Undergraduate Student, Department of Information Technology,
The University of Da Nang – University of Science and Technology (DUT), Viet Nam
Email: binhphan77373@gmail.com | Phone: +84 333971481

LinkedIn | Github

ABOUT ME

- As an Undergraduate Student at DUT, I have developed and optimized multimodal deep learning and computer vision models, including attention-based architectures, for applications such as product categorization and manure identification.
- These models were designed to achieve high accuracy and efficiency, effectively handling imbalanced data and performing well in resource-constrained environments, ensuring robust and real-time performance.
- Specialize in real-time change detection, object detection, single-camera tracking, multi-camera tracking, generative ai, and crowd estimation.

WORKING EXPERIENCE

AI Team, VJ TECHNOLOGIES CO.LTD Da Nang, Viet Nam
Research Engineer | AI, Computer Vision *Jul. 2025 – Present*

- **Automated Document Processing System for Japanese Real Estate Using Vision-Language Models**
  - Developed an automated document processing system for Japanese real estate using Vision-Language Models (QwenVL) to extract and populate data from PDFs into Excel templates.
  - Implemented end-to-end pipeline for document digitization: detecting empty cells in templates, extracting structured data from PDF documents, mapping extracted information to predefined schemas, and generating formatted output files.
  - Built REST API with FastAPI, asynchronous processing, callback support, and error handling.
  - Integrated LLM for table detection, OCR, and form filling with prompt engineering.
  - **Results:** automated document processing workflow, reduced processing time from manual to automated, high accuracy with Japanese text and complex tables.
  - **Technologies:** Python, FastAPI, QwenVL, Vision-Language Models, OCR, PDF Processing, Excel Automation, RESTful APIs, Docker, Pydantic.

- **PV Module Layout Optimization & Rooftop Detection System**
  - Designed a 2D bin-packing optimizer for PV module layouts on single- to four-slope roofs, honoring code setbacks, obstructions, azimuth/tilt, and spacing constraints to maximize energy yield and roof utilization.
  - Reconstructed roof geometry from plan dimensions (CAD/PDF) and auto-placed panels.
  - Built a computer-vision pipeline to detect, reconstruct, and vectorize rooftop facets from Google Maps/aerial imagery (ridge lines, facets, slopes).
  - **Technologies**: Python, OpenCV, NumPy, Shapely, scikit-image, PyTorch; models: YOLOv11x-seg; methods: illumination-invariant preprocessing, adaptive thresholding, K-means, RANSAC, polygonization/simplification.

Center for Advance Robotics Innovation Technology, Nanyang Technological University Nanyang, Singapore
Research Associate | AI, Computer Vision & Embedded System *Mar. 2025 – Apr. 2025*

- **AI-powered smart glasses for the visually impaired**
  - Integrated object detection, depth estimation, text to speech, speech to text, and vision-language models on embedded platforms like NVIDIA Jetson Orin NX, Intel NUC, and Aria smart glasses to assist visually impaired users, achieving real-time performance up to 14 FPS on Jetson Orin NX.
  - Enabled real-time assistance for visually impaired users by guiding them to locate specific objects upon request and navigate safely by detecting and avoiding obstacles along their path.

AI Team, EyeCode company Da Nang, Viet Nam

Research Engineer | AI, Computer Vision *May. 2024 – Oct. 2024*

- **YOLOv5s Pruning Optimization:** Pruning YOLOv5s by 10% reduced inference time by around 5%, with a minor mAP drop (0.812 to 0.797). At 50% pruning, inference time decreased further, but mAP dropped to 0.654. Beyond 50%, performance declined significantly. Tests were conducted on an Odroid N2+.
- **YOLOv5s Transfer Learning**: The YOLOv5s model, enhanced using Transfer Learning with pre-trained VOC weights, improved object detection on a smaller dataset. The mAP increased from 0.695 to 0.812, then further to 0.846, showing significant accuracy gains with limited data. The slight increase in inference time on the Odroid N2+ (up by 0.88%) was negligible compared to the accuracy improvement.
- **Quantization YOLOv5s:** The project applied Quantization techniques to compress the YOLOv5s model, using both Post-training Quantization and Quantization Aware Training. Post-training Quantization reduced inference time by 3.57% with a slight mAP drop from 0.824 to 0.806, while Quantization Aware Training further enhanced speed (reduced by 9.45% compared to the original model), maintaining mAP at 0.819 with minimal accuracy loss. These results, tested on the Odroid N2+, achieved times faster inference with high accuracy retention.

AI Team, D-soft jsc Da Nang, Viet Nam

Research Engineer | AI, Computer Vision *Jan. 2024 – Apr. 2024*

- **Couting people in the ROI area of many CCTVs**
  - **Locating and Analyzing Areas of Interest:** Users can seamlessly select and analyze specific areas within the video stream by clicking on optional points that define the corners of the Region of Interest (ROI).
  - **Dynamic People Detection and Counting:** The system leverages the YOLOv8n model to detect and count individuals within the ROI, delivering real-time updates with approximately 98% accuracy, a processing time of about 40ms per frame, and a performance of around 25 fps on an NVIDIA GTX 1650 (4GB).
  - **Advanced Analytics:** By integrating cutting-edge technologies such as OpenCV and Streamlit, the platform delivers robust data visualization and real-time analytical capabilities.

IT department at DUT

Research Assistant | AI, Computer Vision *Dec. 2023 – Feb. 2024*

- **Enhanced Attention-based Multimodal Deep Learning for Product Categorization on E-commerce Platform**
  - Designed and implemented an attention-based multimodal deep learning model to improve the accuracy of product classification on e-commerce platforms. The model integrates image and text data, leveraging a robust fusion module with attention mechanisms to classify 16 product categories.
  - The proposed model achieved a significant accuracy of 91.18%, outperforming traditional multimodal and unimodal deep learning models, which reached a maximum of 77.21%. This improvement enhances product searchability and the overall customer experience on e-commerce platforms.
  - My role involved building and optimizing the deep learning architecture, focusing on fusing multimodal data for better product categorization. This model contributes to solving the challenge of automatic product classification, with practical implications for e-commerce platform management.

Research Assistant | AI, Signal Processing *Sep. 2023 – Nov. 2023*

- **The modified Vision Transformer for imbalanced NIRs data classification**
  - Led the design and implementation of NIRsViT, a deep learning model leveraging the Vision Transformer architecture. The model was specifically tailored for manure classification using near-infrared spectroscopy (NIRS) data, aimed to improve identification accuracy across various manure types.
  - Introduced innovative methods such as Focal Loss and Upsampling to tackle imbalanced datasets, which are common in agricultural data. This approach significantly enhanced the model's classification performance, achieving an F1-Score of 93.03% and an accuracy of 97.96%, outperforming existing models.

- My responsibilities encompassed developing the deep learning model, preprocessing data, and optimizing classification algorithms. This project established a new benchmark in NIRS-based manure identification, setting a foundation for future research in this field.

SKILLS

- **Software and Hardware**
  - **Language and framework**: Python, Pytorch, Tensorflow, ONNX, OpenCV, ROS2.
  - **Hardware**: sensors (humidity, light, infrared, temperature, accelerator, cameras), microcontrollers (Arduino, Renesas, Odroid N2+, Jetson Orin NX).
  - **Software tools:** Github, Visual Studio Code.
  - **Deep learning skills:** Data preparation, learning techniques (transfer learning, supervised, self-supervised learning, and unsupervised learning), compression techniques (quantization, pruning), deploying deep learning models on embedded devices.
  - **Computer vision skills:** Used to work with change detection, image classification, object detection, image segmentation, crowd counting, genarative ai, and single and multi-camera tracking.
- **Language Proficiency:** Vietnamese: Native language. English: TOEIC 595 (Tested in Nov 2021).

EDUCATION

The University of Da Nang - University of Science and Technology (DUT) Da Nang City, Viet Nam
Engineer's degree in Data Science and Artificial Intelligence *Aug. 2020 – Aug. 2025*

- **Awards**:
  - Top 5 Unihack Contest (Provinces and cities in the Central region of Vietnam) 2022

PUBLICATION

- Le Viet Hung, **Phan Binh**, Phan Minh Nhat, and Nguyen Van Hieu. "Enhanced Attention-based Multimodal Deep Learning for Product Categorization on E-commerce Platform", 13th International Conference on Information Technology and Its Applications 2024 (Springer).

- Phan Minh Nhat, Ngo Le Huy Hien, Dinh Minh Toan, Viet Hung Le, **Phan Binh**, Phung Thi Anh, Nguyen Thi Hoang Phuong, Hieu Nguyen Van. "EBAR: A Novel Machine Learning Model for Quantifying Chemical Concentrations using NIR Spectroscopy", The Journal of Universal Computer Science.

- Nguyen Van Hieu, Ngo Le Huy Hien, Minh Toan Dinh, **Phan Binh**, Minh Nhat Phan, Phung Thi Anh, Le Viet Hung, Nguyen Huy Tuong. "NIRsViT: a novel deep learning model for manure identification using near-infrared-spectroscopy and imbalanced data handling", International Physics and Control Society (IPACS).