

WONG Yi Jie

+60***** yjwong1999@gmail.com <https://yjwong1999.github.io/> <https://github.com/yjwong1999>

EDUCATION

Doctor of Philosophy, Digital Technology, Universiti Tunku Abdul Rahman July 2022 – Ongoing (Left Viva Voce)

- Awarded UTAR Education Fund 100% Scholarship for PhD students.
- Research Scholarship Scheme under **industry funding** (Greotech Integration Sdn Bhd).
- Research Member of an ASEAN IVO project (supported by **NICT**).
- Research Member of a UK-ASEAN Institutional Links project (supported by **British Council**).
- One of the **invited workshop speakers** for **UAS and NTN towards 6G**, hosted in London, UK.
- Co-authored several top journals and conferences: [Yi Jie Wong - Google Scholar](#)

Visiting PhD Student, Université Sorbonne Paris Nord

October 2023 – December 2023

- Funded by **2023 Doctoral Research Mobility Grant, Campus France**.
- Worked on QoE-based Video Streaming under supervision of Prof. Anissa Mokraoui, Head of Institut Galilée.

Bachelor Degree, Biomedical Engineering, Universiti Tunku Abdul Rahman

2018 – 2022

- Graduated with First class honours with CGPA of 3.76/4.00, and awarded Book Prize.

Foundation in Science, Universiti Tunku Abdul Rahman

2017 – 2018

- CGPA of 4.00/4.00 (Top 1 in Universiti Tunku Abdul Rahman, Kampar Branch).

PROFESSIONAL EXPERIENCE

MIMOS Bhd - Freelance AI Engineer

July 2023 – Present

- **Overview:** Provide research consultancy for Malaysia's leading national applied R&D centre, MIMOS Bhd, focusing on cutting-edge technology solutions. Led two consultancy projects: (1) Industrial 3D Object Detection using Few Shot Learning, and (2) AI-based Machine Vision for Multidimensional Quality Control.
- **GUI Design:** Designed an interactive end-to-end Graphical User Interface (GUI) for annotations, augmentations, training, and inference, tailored for **MIMOS' semiconductor industry partner**. The GUI is based on Python, PyTorch, NumPy, and OpenPCDet.
- **Custom Data Preprocessing:** Engineered a custom data preprocessing method to remove noise from the client's raw point cloud data, achieving up to a 50% reduction in unwanted artefacts.
- **Custom Data Augmentation:** Developed a specialised data augmentation pipeline to create a synthetic dataset for the client's imbalanced data, with a pass-to-defect ratio of 24:1.
- **Performance Improvement:** Integrated the custom Data Preprocessing and Augmentation pipeline with state-of-the-art models from OpenPCDet module. Our proposed method improved the client's defect detection rate (mAP) by 54%, from 61.64% to 95%; and reduced training time by 86%, from 720 epochs to just 100 epochs.
- **C++ Inference Implementation:** Implemented a C++ inference on CUDA using TensorRT and ONNX GraphSurgeon.

Universiti Tunku Abdul Rahman - Research Assistant (GenAI)

December 2024 – Present

- **Retrieval-Augmented Generation (RAG):** Designed and implemented a RAG prototype to enhance the efficiency of knowledge retrieval and response generation. Utilized Ollama for local large language model (LLM) hosting, ensuring low-latency inference while maintaining data privacy and eliminating cloud dependency. Developed and optimized the backend infrastructure using LlamaIndex to enable efficient document parsing, indexing, and retrieval.
- **Vision-Language Model (VLM) Integration:** Integrated a Vision-Language Model (VLM) to enhance the chatbot's ability to process and understand PowerPoint slides, enabling multimodal document interpretation for improved contextual retrieval. This integration significantly expanded the chatbot's capabilities beyond text-based queries.
- **Embedding Fine-Tuning with NUDGE:** Designed and implemented a fine-tuning module leveraging NUDGE to optimize embedding representations for domain-specific data. Achieved up to 57% improvement in retrieval quality on a custom dataset by refining vector representations and enhancing semantic similarity scoring.

Université Sorbonne Paris Nord - Research Attachment

October 2023 – December 2023

- **QoE-centric Adaptive Video Streaming:** Optimised the Quality-of-Experience (QoE) for adaptive bitrate video streaming (i.e. Youtube and Netflix). Specifically, a deep reinforcement learning algorithm was developed to replace existing algorithms to download the next video chunk for video playback.
- **Parallel Training:** The algorithm is trained using the Federated Learning (FL) paradigm, a privacy-centric distributed training methodology. Instead of collecting user data, this approach involves collecting models trained on the user's data. This allows for parallel training on multiple user devices simultaneously, followed by model aggregation, all while preserving user privacy.
- **Performance Improvement:** The proposed algorithm achieved the best user's mean opinion score in 8 out of 10 test videos across various network simulation settings. It was also stress tested under extreme low bandwidth conditions, where it delivered state-of-the-art results. The output of the project has been submitted to reputable IEEE journals for the review process.

UK-ASEAN SYRACUSE - Research Member

October 2022 – September 2023

- **Algorithm Design:** Developed a custom deep reinforcement learning algorithm for the joint optimization of UAV trajectory, UAV active beamforming, and RIS passive beamforming. The primary objective was to maximise Secrecy Energy Efficiency (SEE) by optimising signal transmission while ensuring privacy and energy efficiency.
- **Performance Improvement:** Achieved an 18.63% improvement in SEE, increasing it from 40.8 to 48.4 bits/s/Hz/kJ, while simultaneously reducing energy consumption by 9.68%, from 12.4 kJ to 11.2 kJ.
- **Publication and Recognition:** The project's output was published at **IEEE WCNC 2023**, a core conference of the IEEE Communication Society [\[Click me\]](#). Mr. Wong was also invited to present the project findings as a workshop speaker for UAV and NTN towards 6G, hosted in London, UK, showcasing expertise in cutting-edge communication technologies.

ASEAN IVO - Research Member

July 2021 – December 2022

- **Custom Model Design:** Designed a custom YOLO-based multi-task model for joint victim detection and disaster classification. This approach involved attaching a pruned head from a MobileNet-based disaster classification model onto the backbone of YOLO, creating an innovative multi-head, multi-task architecture.
- **Model Optimization:** Improved model inference speed using a series of OpenVINO optimization techniques, including the OpenVINO Model Optimizer, OpenVINO DL Workbench, and INT8 Quantization with the Post-Training Optimization Tool. The optimised model was tested on the following devices:
- **Intel Core i7-10875H (CPU):** Achieved a 2.51X inference speedup, increasing from 6.55 FPS to 16.46 FPS, comparable to the 20.31 FPS inference speed on an NVIDIA RTX 2070 Super.
- **Intel Neural Compute Stick 2 (NCS2) on Raspberry Pi 4:** Achieved an 18% inference speedup, increasing from 1.52 FPS to 1.80 FPS. For comparison, the inference speed on an NCS2 with an Intel CPU is 2.50 FPS.
- **Award Recognition:** The project received the Best Paper Presentation award at IEEE UEMCON 2021 in New York, highlighting its significance and impact. In addition, the complete project was also awarded the Certificate of Excellence by the ASEAN IVO committee.

RELEVANT PROJECTS

Cross-City Generalizability of Instance Segmentation Model

June 2024 – Aug 2024

- **2024 IEEE BigData Cup Challenge:** The challenge was hosted by the reputable IEEE International Conference on BigData, ranked 6th in Data Mining & Analysis in Google Scholar Metrics [\[Click me\]](#). This approach aims to inspire the development of models with robust generalisation capabilities, capable of overcoming the hurdles of automatic building footprint detection and extraction across various landscapes.
- **Achievement:** Led the team to secure 1st place in the challenge. [\[Click me\]](#)
- **Domain Data:** We leveraged the Microsoft Building Footprint Dataset to increase the training dataset size and improve the generalisation capability of our YOLOv8 model beyond the provided training set. We conducted several ablation studies to identify the most suitable region data that resembles the domain data for the challenge scope.
- **Synthetic Data:** We proposed a synthetic data generation technique that replaces the background of satellite images with a synthetic background while keeping the actual building footprint. Our team also explored end-to-end synthetic data generation using stable diffusion, which maps the label mask to image generation.

Coastal Building Damage Assessment using Satellite Imagery

Feb 2024 – May 2024

- **2024 EY Open Science Data Challenge:** The challenge was hosted by Ernst & Young, focusing on machine learning applications for assessing storm damage using high-resolution satellite data in data-poor coastal environments.
- **Achievement:** Led the team to secure 8th place among 10 global semi-finalists from over 11,000 registrants, demonstrating our innovative approach and technical expertise. [\[Click me\]](#)
- **Model-Assisted Labelling Technique:** Proposed a model-assisted labelling technique to accelerate the annotation process by 45%. Utilised our fine-tuned YOLOv8 model to predict bounding boxes for objects in satellite imagery, with minimal human intervention to correct faulty label predictions.
- **OOD Problem Mitigation:** Split the annotated dataset into subsets to address the Out-of-Distribution (OOD) problem. Fine-tuned YOLOv8 on each subset, evaluated them on the validation dataset, and rejected poor-performing subsets. This strategy ensured the final model achieved 0.51 mAP, 27.5% higher than the competition threshold of 0.40.

Automated Pill and Empty Blisters Detections in Microwave Images

April 2022 – May 2022

- **2022 Rohde & Schwarz Engineering Competition:** The engineering competition is hosted by Rohde & Schwarz, focusing on the development and optimization of a YOLOv5 model for detecting pills and empty blisters in images.
- **Achievement:** Led the team to achieve 11th place among 13 global finalists, demonstrating our commitment to innovation and technical expertise. [\[Click me\]](#)
- **Model Development:** Developed and optimised a YOLOv5 model specifically designed to detect pills and empty blisters in the provided images, showcasing proficiency in object detection tasks.
- **Hyperparameter Tuning:** Experimented with various hyperparameter configurations to enhance YOLOv5's performance, achieving optimal results for the given task.
- **Custom Anchor Box Design:** Customised the anchor box sizes to improve detection accuracy for small objects, ensuring precise identification of pills and empty blisters.

Multi-Camera Face Detection and Recognition

Sep 2022 – Current

- **Project Overview:** While face detection and recognition technologies have exhibited high accuracy rates in controlled conditions, recognizing individuals in open environments remains challenging due to factors such as changes in illumination, movement, and occlusions. In this paper, we propose a multi-camera face detection and recognition (MCFDR) pipeline, which consists of three main parts: face detection, face recognition, and tracking.
- **Model Development:** YOLOv8m and OpenSphere's SphereFace are used in this project as the face detection and face recognition modules. The two models are integrated with BoT-SORT tracking algorithms to enable real-time detection and tracking.
- **Deployment:** To facilitate real-time tracking, we integrated a multithreading methodology into the pipeline to run face tracking on multiple video streams simultaneously. The pipeline was tested on four real-time RTSP streams, all running on a single NVIDIA RTX 4070 GPU.

Smart Car Plate Recognition System using Multi-Task Learning

Sep 2023 – April 2024

- **Project Overview:** Co-supervised a Final Year Project (FYP) student to develop an advanced one-stage multi-head multi-task YOLO model integrating Automatic License Plate Recognition (ALPR) and Vehicle Recognition.
- **Model Development:** Presented a novel multi-task YOLO model capable of handling both ALPR and vehicle recognition tasks simultaneously, showcasing innovative approaches to object detection and recognition.
- **Synthetic Data Generation:** Since there is a limited amount of license plate data curated for Malaysia, we designed a custom synthetic data generation pipeline to generate license plates and improve the diversity of our training dataset. This pipeline increased our license plate recognition performance by 15.1%, from 0.696 to 0.801 mAP.
- **Custom Loss Function:** Engineered a custom multi-task loss function using a weighted sum approach to optimise the performance of the multi-task YOLO model.
- **Performance Improvement:** Improved the overall mAP by 5.56% and reduced inference time by 78.90% compared to conventional methods, demonstrating significant advancements in efficiency and accuracy.
- **Publication and Recognition:** The project was submitted and accepted as a conference paper at IEEE EAIS 2024 and received the gold prize in the LKCFES FYP Poster Competition 2024, highlighting its impact and success.

Mixture-of-Experts for Zero Shot Generalisation in UAV-RIS Networks

Sep 2023 – June 2024

- **Project Overview:** This study introduces a framework where Deep Reinforcement Learning (DRL) agents are organised as specialised "experts" for groups of similar tasks, significantly reducing the number of agents needed. A mixture of experts (MoE) model combines these specialised agents into a single, unified model, which autonomously selects the best expert for each new task. This approach enhances the DRL system's adaptability, allowing it to generalise to new tasks without additional training.
- **Federated Learning (FL):** By using FL, tasks with similar reward structures are grouped quickly and efficiently, enabling faster training of shared multi-task agents. This FL-driven approach reduces the number of required agents and boosts training speed by 10 times compared to conventional single-task DRL models.

COMPETITIONS

1. **Global Finalist (Champion)**, IEEE BigData Cup 2024: Building Extraction Generalisation Challenge
2. **Global Semi Finalist** (Top 10), EY Open Science Data Challenge 2024
3. **Global Finalist**, Rohde & Schwarz Engineering Competition 2022
4. **Global Finalist**, OpenCV AI Competition 2021
5. **Region Qualifiers**, Huawei ICT Competition Asia-Pacific 2021/22
6. **National 1st Runner Up**, Huawei ICT Competition Malaysia 2021/22
7. **National 1st Runner-Up**, National Intel-CREST Industry University Challenge 2021
8. **National Champion**, 2022 IEEE Malaysia Section Final Year Project Competition

AWARDS

1. Best Paper Presentation, Session 13: Session File and Storage Systems, Artificial Intelligence (Session 13), 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (IEEE UEMCON 2021), New York, USA, 1-4 Dec, 2021. [🔗\[Click me\]](#)
2. Certificate of Excellence, Context-Aware Disaster Mitigation using Mobile Edge Computing and Wireless Mesh Network, Highest Project Evaluation Score among 17 ASEAN IVO Projects, 15 Dec 2023. Mr Wong contributed 2 first-author papers and the best paper presentation for the project. [🔗\[Click me\]](#)

SKILLS

Technical skills : Artificial Intelligence, Deep Learning, PyTorch, TensorFlow, OpenVINO, TensorRT, Python, C++
Language : English, Mandarin, Malay

📁 Feel free to check out my [homepage](#) to see the complete list of research works, competitions and achievements!