
EDUCATION

Southern University of Science and Technology

Shenzhen, Sep 2022 – Jun 2026

B.E. in Computer Science and Technology

- GPA: 3.6 / 4.0
- Core Courses: Artificial Intelligence, Intelligent Robotics, Operating Systems, Software Engineering, Computer Vision

The Chinese University of Hong Kong (Incoming)

M.SC. in Computer Science

Hong Kong, Sep 2026 - Jul 2027

- Admitted, expected to enroll in Fall 2026.

The Hong Kong Polytechnic University

Hong Kong, Sep 2025 – Dec 2025

Semester Exchange in Department of EEE

- Program: Artificial Intelligence Engineering
- Core Courses: Deep Learning, Multimedia Communication, Web System and Technology

National University of Singapore

Singapore, Jul 2024 – Aug 2024

Summer Research Program in School of Computing

- Project: Focused on Visual Computing studies, with research on traffic sign recognition in extreme scenarios
-

PUBLICATIONS

G. Lan*, **J. Ma***, Z. Zhao, Q. Qi, Y. Liu & Q. Hao. RoGaussianOCC: A Gaussian Fusion Method for Accurate and Robust Occupancy Prediction in Autonomous Driving. *Information Fusion*. (Under Review, Jan. 2026) (Co-first Author)

X. Chen, Z. Liu, **J. Ma**, B. Du, T. Zhang, X. Wang & B. Zhou. AirHunt: Bridging VLM Semantics and Continuous Planning for Efficient Aerial Object Navigation. *IEEE Transactions on Automation Science and Engineering*. (Under Review, Jan. 2026)

RESEARCH EXPERIENCE

AirHunt: Bridging VLM Semantics and Continuous Planning for Efficient Aerial Object Navigation.

Research Intern | Supervisor: Prof. Boyu Zhou

Apr 2025 – Oct 2025

- **OBJECTIVE** – To develop a real-time aerial navigation system capable of objects navigation in large-scale, unknown outdoor environments based on natural language commands. The core goal was to bridge the gap between high-latency Vision-Language Models reasoning and the low-latency, continuous control required for aerial agent, enabling efficient zero-shot objects navigation
- **METHOD** – 1) *Dual-pathway Asynchronous Architecture*: designed a novel framework where the agent performs continuous, real-time replanning, with a VLM concurrently reasoning about high-level strategy, which decouples slow VLM inference from fast UAV control; 2) *Active Dual-Task Reasoning*: implemented a module to select and retain keyframes with rich semantic value, creating a compact and informative scene memory for VLM reasoning; 3) *Semantic-Geometric Coherent Planning*: developed a planning algorithm that integrates semantic guidance from VLMs with the environmental geometric map, enabling continuous navigation
- **ACHIEVEMENT** – 1) Achieved a 78% navigation success rate in AirSim Simulation, and reaching 98% of human performance; 2) Completed navigation tasks with 37% of flight time compared to existing methods by leveraging the *Dual-pathway Asynchronous Architecture* to reduce reasoning latency and flight detours; 3) Validated effectiveness and feasibility of *AirHunt* on a custom-built quadrotor through over 10 hours of outdoor objects navigation experiments in diverse real-world scenarios
- **CONTRIBUTION** – 1) Conducted research on the latest work about Vision Language Navigation and proposed research question; 2) Designed and implemented a high-frequency geometric navigation path planner, a low-frequency VLM semantic reasoning module and a ROS based cross-modules communication system; 3) Implemented baseline models and conducted ablation studies in AirSim simulation; 4) Designed and developed custom AirHunt quadrotor agent and conducted real-world experiment and demonstration

RoGaussianOCC: An Accurate and Robust Gaussian Occupancy Prediction in Autonomous Driving

Research Leader | Supervisor: Prof. Qi Hao & Dr. Gongjin Lan

Sep 2024 – Apr 2025

- **OBJECTIVE** – To develop a robust 3D Gaussian occupancy prediction model for vision-based autonomous driving in extreme environment by tackling the following limitations: 1) Inefficient Gaussian representation due to random initialization; 2) Imprecise prediction under occlusion caused by a single-frame framework; 3) Sparse supervision signals hindering early convergence
- **METHOD** – 1) *Historical Gaussian Fusion*: utilized the prior historical Gaussian properties with temporal information to update the current Gaussians; 2) *Gaussian Feature Aggregation*: proposed 4D Gaussian Sampling to incorporate both historical and current frames, thereby enhancing scene perception and reducing occlusion; 3) *Perspective Supervision*: leveraged auxiliary supervision

signal to improve accuracy and facilitate the convergence process of the training backbone

- **ACHIEVEMENT** – 1) Achieved a 32.20% IoU and 20.65% mIoU in occupancy prediction on nuScenes datasets, with a 7.95% and 8.12% relative improvement over baseline models; 2) Validated sub-modules effectiveness and contribution to overall performance by ablation study; 3) Validated robustness by quantitative and qualitative experiments on an extreme-task dataset
- **CONTRIBUTION** – 1) Conducted related work research and identified limitations of existing work through experiments; 2) Proposed and implemented RoGaussianOCC framework to tackle proposed limitations; 3) Performed model training, parameter tuning, and effectiveness evaluation through ablation studies and visualization analysis; 4) Led project management, including task allocation, milestone tracking, and reporting progress to supervisors; 5) Authored a research paper as a co-first author and submitted it to Information Fusion (under review), with demo and code released at Github

Natural Language Instructions Based 3D Cross-Modal Perception System for UAVs

Core Member | Supervisor: Prof. Feng Zheng

Dec 2023 – Sep 2024

- **OBJECTIVE** – To develop a generalized embodied UAV agent, leveraging VLM to enable natural language instruction and interaction
- **METHOD** – 1) *Perception & Mapping*: Processed RGB-D input with SLAM algorithm to generate 3D point clouds; 2) *Interaction*: Employed 3D-LLMs to process cross-modal inputs of environmental point clouds and user prompt, enabling user-agent interaction through question-answering about environmental information and object detection; 3) *Cognition & Decision-making*: Utilized tuned VLMs to generate waypoints from natural language inputs; 4) *Action*: Generated exploration path from discrete waypoints
- **CONTRIBUTION** – 1) Conducted research on 3DLLM, trained and finetuned models for UAV agent interaction tasks; 2) Finetuned VLM to generate waypoints for exploration tasks; 3) Developed the UAV platform's ROS system and deployed the model on UAVs

SCHOOL PROJECTS

Autonomous Limo Robot

Mar 2025 – May 2025

- **OBJECTIVE** – To develop a Limo Robot with visual detection capabilities, and auto-exploration based on Visual Language Navigation
- **TOOL** – CV detection models, Visual Language Navigation, ROS, Nvidia Jetson AGX
- **SOLUTION** – 1) Deployed computer vision models to detect crosswalks, traffic lights, and road signs to aid in environmental decision-making; 2) Used Fast-SLAM for simultaneous mapping and localization during exploration; 3) Deployed a Visual Language Frontier Map model as the decision-making module for semantic navigation, which provided waypoints and feasible paths
- **HIGHLIGHT** – Our robot can efficiently search for unknown objects with deployed VLN models based on user's instruction.

Smart Album Software Development

Mar 2025 – May 2025

- **OBJECTIVE** – To develop a web-based smart album application with features including image searching, clustering, and captioning
- **TOOL** – Flask, Spring Boot, Vue, SQL, object detection model, VLM, Git
- **SOLUTION** – 1) Built the backend using Spring Boot and the frontend with Vue and deployed models on a Flask server; 2) Implemented core functionalities such as object and face detection, CLIP-based image search and clustering, and VLM-powered Q&A and caption generation; 3) Utilized DevOps principles for collaboration, and automated testing and deployment using Git

COMPETITIONS & AWARDS

Excellent Creativity Award (Second Place), Meituan 3rd Low-Altitude Economy Intelligent Flight Management Challenge *Sep 2025*

Third Prize, Guangdong Contemporary Undergraduate Mathematical Contest in Modelling 2023

Dec 2023

SKILLS

Languages – Fluent in English (TOEFL 102), Cantonese; Native in Mandarin

Programming – Python, Java, C/C++, MATLAB, SQL

Frameworks & Tools – OpenCV, PyTorch, ROS, MDCV, TensorRT, AirSim, Unreal Engine, Nvidia Jetson, Git

Datasets & Benchmarks – nuScenes, KITTI, AirSim Simulation Environment

Areas of Expertise – Vision-Language Model Training, Fine-tuning and Deployment; 3D Scene Perception & Occupancy Prediction; Gaussian-based Representation & Fusion; Embodied Intelligence Multimodal Perception, Navigation & Motion Planning; Embodied Agent System Development