

Chi Lung Ma

kevinma20000215@gmail.com | (206) 889-0169 | linkedin.com/in/clm-kevin-robotics | github.com/Kevinma0215

EDUCATION

University of Washington <i>M.S. in Mechanical Engineering (Data Science Option)</i> , GPA: 3.83/4.0 Coursework: Advanced Machine Learning, AI Robotics, High Performance Computing	Seattle, WA, US Expected Jun. 2027
National Yang Ming Chiao Tung University (NYCU) <i>B.S. in Mechanical Engineering</i>	Hsinchu, Taiwan Jun. 2022

SKILLS

Programming: Python, C++ **Frameworks:** OpenCV, PyTorch, ROS, Gazebo, MuJoCo, TensorRT **Tools:** Git, CUDA
Areas: Multimodal LLM, Computer Vision, Model Optimization, Robot Learning

RESEARCH PROJECTS

Multimodal LLM Quantization and Inference Optimization - Edge GPU Deployment May. 2026

- Benchmarked 7 TensorRT-LLM quantization configs (BF16 through W4A8 NVFP4) on Qwen2-VL-2B-Instruct on an edge GPU; identified SmoothQuant as the Pareto-optimal tier — 2.13× decode speedup while matching or exceeding BF16 baseline accuracy on VQAv2, POPE, and MME.
- Diagnosed a severe FP8 accuracy regression (VQAv2 -8.8 pp, MME -189) through targeted ablation: isolated FP8 KV cache quantization as root cause, tracing it to visual token embeddings exceeding FP8 E4M3 dynamic range - a failure mode when applying text-LLM deployment recipes to multimodal LLMs without adaptation.
- Identified task-specific collapse in INT4-AWQ text translation (-27.5 pp) absent in plain INT4, pointing to multilingual calibration data as a structural dependency of activation-aware quantization methods.

VLA Manipulation Spatial Reasoning Robustness under Distribution Shift Mar. 2026

- Discovered a sharp behavioral collapse: success rate dropped from 95% to 20% under just 5 cm of spatial perturbation to the initial end-effector position — out-of-distribution configurations unseen during training.
- Designed perturbation experiments to probe spatial reasoning robustness independently of trajectory memorization: by varying only the initial end-effector position outside the training distribution, isolated failures attributable to 2D-to-3D grounding rather than policy recall.
- Built a hierarchical VLA pipeline in MuJoCo combining vision-language planning, depth-based 3D waypoint grounding, and trajectory execution via classical control — providing the diagnostic substrate for tracing root causes to specific pipeline stages.

PROFESSIONAL EXPERIENCE

Multi-Agent Unmanned Aerial Vehicles (WAAR) - Autonomy Team Jan. 2026 - Apr. 2026

Autonomy Team Member

- Designed mission-level behavior planning framework for multi-drone autonomous exploration: state-based task execution under partial observability, resilient to worst-case agent failures.
- Implemented frontier-based exploration with target-biased prioritization and confidence-weighted grid representation for uncertainty-driven region coverage.

Autonomous Underwater Vehicle (ORCA) - Perception & Avionics Team Jun. 2020 - Sep. 2022

Perception & Avionics Group Lead

- Improved robustness of underwater object perception using a ROS-based YOLO detection pipeline.
- Mitigated underwater visual degradation effects by addressing light attenuation and color shift through data preprocessing and WaterGAN-based domain adaptation, resulting in more stable detection performance across varying water conditions.

RESEARCH EXPERIENCE

Aerospace Systems and Aerodynamics Research Lab, NYCU Hsinchu, Taiwan

Undergraduate Researcher Feb. 2021 - Aug. 2021

- Designed a synchronized multi-camera system and embedded capture pipeline on Jetson Nano for 3D point cloud reconstruction of high-altitude balloon deformation dynamics.
- Contributed to a peer-reviewed publication: Lin, M. Y., Ma, C. L., et al. *Reconnect 2020: Collegiate High-Altitude Balloon Mission to 25,000 m*. International Conference on Astronautics and Space Exploration(2021).