Hao Vo

Website | Gmail | GitHub | Google Scholar

Research Interests

My research focuses on spatial and physical understanding - injecting geometric and physical priors into robots systems. My goal is to enable robots not only to *"see"* the world but also to *"sense"* things. Longer term, I aim to help these systems act reliably, advancing from task-specific solutions toward general-purpose, collaborative autonomy.

Education

University of Arkansas — Arkansas, USA
Master's, Computer Science - advised by Dr. Ngan Le — *May 2025 – May 2027 (Expected)*

University of Information Technology - VNUHCM — HCM, Vietnam
B.Sc., Computer Science — *Aug. 2021 – Dec. 2024*

Selected Publications

SemLT3D: Semantic-Guided Expert Distillation for Camera-only Long-Tailed 3D Object Detection.
***Hao Vo**, Khoa Vo, Thinh Phan, Ngo Xuan Cuong, Gianfranco Doretto, Hien Nguyen, Anh Nguyen, Ngan Le*
CVPR 2026

DriveSpatial: A Benchmark for Spatiotemporal Intelligence in VLMs for Autonomous Driving.
***Hao Vo**, Khoa Vo, Phu Loc Nguyen, Sieu Tran, Duc Minh Nguyen, Ngo Xuan Cuong, Gladys Gawugah, Sreevenkata Anjani Tishita Godavarthi, Chase Rainwater, Nghi D. Q. Bui, Anh Nguyen, Duy Minh Ho Nguyen, Ngan Le*
NeurIPS 2026

OmniSpace: Efficient Geometry Awareness for Autonomous Vehicles MLLMs.
Hao Vo, Phu Loc Nguyen*, Khoa Vo, Sieu Tran, Duc Minh Nguyen, Ngo Xuan Cuong, Nghi D. Q. Bui, Anh Nguyen, Duy Minh Ho Nguyen, Ngan Le*
NeurIPS 2026

Robust Motorcycle Helmet Detection in Real-World Scenarios: Using Co-DETR and Minority Class Enhancement.
***Hao Vo**, Sieu Tran, Duc Minh Nguyen, Thua Nguyen, Tien Do, Duy-Dinh Le*
CVPRW 2024

MVDGC: Joint 3D and 2D Multi-view Pedestrian Detection via Dual Geometric Constraints.
*Thinh Phan, **Hao Vo**, Khoa Vo, Thanh Ngo, Cuong Pham, Ngan Le*
TMLR 2026

Dual-State Slot Attention: Decoupling Appearance and Identity for Video Object-Centric Learning.
*Sieu Tran, Duc Nguyen, **Hao Vo**, Khoa Vo, Ngan Le*
ACM MM 2026

Experience

Graduate Research Assistant — May 2025 - Present
University of Arkansas — *AR, USA*

Research on multi-view understanding and spatial intelligence for autonomous systems, spanning self-driving cars and robotics, with papers published and under submission at venues such as CVPR, ACMMM, and TMLR.

- **3D camera-based perception.** Developed a semantic-guided expert distillation framework that transfers the generalizable knowledge of VLMs into a mixture-of-experts detector specialized for minority classes, achieving state-of-the-art performance with 2× higher mAP on tail classes over the baseline. Published at **CVPR 2026**.
- **Spatial intelligence in multi-view embodied systems.** Built a comprehensive evaluation framework and benchmark to probe the spatial intelligence of VLMs in this setting, revealing that constructing a coherent cognitive map is the primary bottleneck of current VLM systems. First-author paper under review.
- **Geometry-aware MLLMs for end-to-end driving.** Building on the above insight, designed a method that injects geometric knowledge into MLLMs from a 3D foundation model, combined with geometric priors such as camera pose and epipolar constraints, without adding a significant number of parameters. First-author paper under review.
- **Physical understanding for robot manipulation.** Building a method and benchmark that teach models to reason about physical laws such as gravity, force, friction, and inertia. The method fuses high-frequency tactile and force feedback, while the benchmark is constructed in Isaac Sim by mining publicly released assets. Ongoing paper.

AI Engineer Sep. 2023 – Apr. 2025
viAct *Hong Kong*

Built computer vision for construction safety and led AI deployment across Singapore and KSA, covering 20+ projects and 300+ cameras.

- **3D computer vision R&D.** Designed a 3D-aware method to detect people inside suspended-load danger zones and optimized it to run 10 cameras at real-time FPS on a single Jetson NX.
- **MLOps system.** Designed an end-to-end pipeline for data collection, labeling, training, and deployment, cutting data-processing time by 30%.

Undergraduate Research Assistant Sep. 2021 – May 2024
University of Information Technology – VNUHCM *HCM, Vietnam*

Research on 2D perception and multimodal video retrieval systems for real-world deployment.

- **2D perception.** Proposed a fusion method combining multiple foundation detectors (e.g., Co-DETR) for robustness under extreme conditions, paired with an adaptive confidence-scoring scheme for long-tailed distributions. Placed 1st in the NVIDIA AI City Challenge; published at CVPRW 2024.
- **Video retrieval system.** Designed and implemented an end-to-end system retrieving ∼10-second events from 1.3 TB of video in 0.61 s via speech, text, or image queries. Introduced search-space reduction techniques, including timestamp filtering and CLIP-based temporal features, to disambiguate temporally similar events. The system won Vietnam's largest AI competition and was later extended to secure three additional first-place finishes.
- **Face detection.** Surveyed facial recognition methods and built the analysis UI for a face-detection pipeline, later deployed for student attendance tracking at UIT.

Awards

First Prize, AI City Challenge (Track 5), **NVIDIA**. Highest performance at the CVPR Workshop. (2024)
First Prize, Ho Chi Minh City AI Challenge, Vietnam. Best-performing event retrieval system. (2022)
Second Prize, Vietnam Ministerial-level Scientific Research. Helmet-violation detection for motorbike riders. (2025)
Second Prize, FPT Vietnam AI Challenge. Runner-up Lost-and-Found system. (2023)
Third Prize, BKAI–Naver Challenge. Our method recognized as top-3 for scene text recognition. (2021)
Consolation Prize, UIT Game Hackathon. Computer vision system for environmental protection. (2022)
Finalist, Zalo AI Challenge. Top-5 in adversarial personalized image generation. (2024)

Mentorship & Leadership

Research mentoring: Remotely advised an undergraduate student at UIT on problem formulation and experimental design; the resulting work earned an *Oral at ACM MM 2026*.
Peer mentoring: Mentored five junior students at UIT to a 3rd-place finish at a national video retrieval competition and to their first publication at a regional conference.
Academic leadership: Led the majority of research projects undertaken during my Master's, as well as all seven AI competition teams competed on during my Bachelor's.
Industrial leadership: Led a team of 8 AI engineers at viAct on algorithm design and research direction, serving as the primary point of contact for the product and software teams and working directly with the CEO.

Academic Services

Conference Reviewer: International Conference on Learning Representations (ICLR) 2026.

Technical Skills

Languages: Python, JavaScript, C/C++, Go.
Frameworks: Isaac Sim, MuJoCo, Torch, PyTorch Lightning, React, Node.js, Next.js, FastAPI.
Developer Tools: Git, Docker, Airflow, Ansible, Vault, Signoiz, MLflow, Celery, Redis, RabbitMQ, Nginx, Slurm, AWS, GitLab CI, MongoDB, SQL, DVC.