

HAOKUN ZHU

☎ +1 (412) 954-8608 ✉ haokunz@andrew.cmu.edu 🌐 haokunzhu.com 🏠 [Haokun Zhu](#) 🌀 [zwandering](#)

EDUCATION

- **Carnegie Mellon University** Pittsburgh, United States
Ph.D. in Robotics; Advisor: Ji Zhang Aug. 2026 – Present
- **Carnegie Mellon University** Pittsburgh, United States
Master of Science in Robotics; **GPA: 4.25/4.3** Aug. 2024 – Jun. 2026
- **Shanghai Jiao Tong University** Shanghai, China
B.Eng. in Computer Science and Technology; **GPA: 89.14/100; 3.81/4.3** Sep. 2020 – Jun. 2024

PUBLICATIONS (* MEANS EQUAL CONTRIBUTION, # MEANS CORRESPONDING AUTHOR)

- **SysNav: Multi-Level Systematic Cooperation Enables Real-World, Cross-Embodiment Object Navigation**
Haokun Zhu, Zongtai Li, Zihan Liu, Kevin Guo, Zhengzhi Lin, Yuxin Cai, Guofei Chen, Chen Lv, Wenshan Wang, Jean Oh, Ji Zhang
IEEE RA-L 2026. [Website] [ArXiv:2603.06914]
- **STRIVE: Structured Representation Integrating VLM Reasoning for Efficient Object Navigation**
Haokun Zhu*, Zongtai Li*, Zhixuan Liu, Wenshan Wang, Ji Zhang, Jonathan Francis, Jean Oh
ICRA 2026; RSS 2025 SemRob Workshop. [Website] [ArXiv:2505.06729]
- **MOSAIC: Generating Consistent, Privacy-Preserving Scenes from Multiple Depth Views in Multi-Room Environments**
Zhixuan Liu, Haokun Zhu, Rui Chen, Jonathan Francis, Soonmin Hwang, Ji Zhang, Jean Oh
ICCV 2025. [Website] [ArXiv:2503.13816]
- **AesStyler: Aesthetic Guided Universal Style Transfer**
Ran Yi*#, Haokun Zhu*, Teng Hu, Yu-Kun Lai, Paul L. Rosin
ACM MM 2024. [Website]
- **SAMVG: A Multi-stage Image Vectorization Model with the Segment-Anything Model**
Haokun Zhu*, Juang Ian Chong*, Teng Hu, Ran Yi#, Yu-Kun Lai, Paul L. Rosin
ICASSP 2024. [ArXiv:2311.05276]
- **M3DM-NR: RGB-3D Noisy-Resistant Industrial Anomaly Detection via Multimodal Denoising**
Chengjie Wang, Haokun Zhu, Jinlong Peng, Yue Wang, Ran Yi, Yunsheng Wu, Lizhuang Ma, Jiangning Zhang
TPAMI 2025. [ArXiv:2406.02263]
- **Stroke-based Neural Painting and Stylization with Dynamically Predicted Painting Region**
Teng Hu, Ran Yi#, Haokun Zhu, Liang Liu, Jinlong Peng, Yabiao Wang, Chengjie Wang, Lizhuang Ma
ACM MM 2023. [Code] [ArXiv:2309.03504]
- **Phasic Content Fusing Diffusion Model with Directional Distribution Consistency for Few-Shot Model Adaption**
Teng Hu, Jiangning Zhang, Liang Liu, Ran Yi#, Siqi Kou, Haokun Zhu, Xu Chen, Yabiao Wang, Chengjie Wang, Lizhuang Ma
ICCV 2023. [Code] [ArXiv:2309.03729]

EXPERIENCE & INTERNSHIP

- **Zhang Lab in CMU** Pittsburgh, United States
Ph.D. Student advised by Dr. Ji Zhang Aug. 2026 – Present
 - **JEPA World Model for Navigation**: exploring how an action-conditioned V-JEPA world model can serve navigation, especially tasks that require understanding other agents' intentions.
- **roBot Intelligence Group (BIG) in CMU** Pittsburgh, United States
Graduate Research Assistant co-advised by Prof. Jean Oh and Dr. Ji Zhang Sep. 2024 – Jun. 2026
 - **VLM-Guided Object Navigation in Simulation**: designed structured multi-layer scene representation (viewpoint, object, and room nodes) and two-stage navigation policy combining VLM reasoning with classical exploration methods, achieving state-of-the-art performance across HM3D, RoboTHOR, and MP3D benchmarks (ICRA 2026).
 - **SysNav: Cross-Embodiment Navigation System in Real-World**: built a three-level system for building-scale object navigation, deployed on wheeled, quadruped, and humanoid robots across 190 real-world experiments with 4-5x higher efficiency than baselines (RA-L 2026).
- **Digital Media Computer Vision Laboratory (DMCV) in SJTU** Shanghai, China
Undergraduate Research Assistant advised by Prof. Ran Yi Oct. 2022 – Jun. 2024

- **Image Generation & Stylization:** few-shot diffusion model adaption (ICCV 2023), aesthetic guided style transfer (ACM MM 2024), and SAM-based image vectorization (ICASSP 2024).

• **Youtu Lab at Tencent Technology (Shanghai) Co., Ltd.**

Shanghai, China

Research Intern advised by Jinlong Peng

Dec. 2023 – Jun. 2024

- **Multimodal Industrial Anomaly Detection:** addressed ineffective feature integration issues in 3D point cloud and RGB images by applying multimodality to enhance industrial anomaly detection.

RESEARCH PROJECTS

JEPA World Model for Navigation (Robotics, World Models)

Jul. 2026 – Present

Advisor: Ji Zhang

- We study how an action-conditioned JEPA world model can serve navigation, especially in dynamic scenes where the robot must anticipate the intentions of other agents (e.g., pedestrians). In this project,
 - I built action-conditioned world models in the frozen V-JEPA 2.1 latent space that imagine how a scene, including other agents, evolves under candidate robot actions.
 - I compared regression and flow-matching predictors with camera-ray conditioning, trained on 710 hours of real-world robot data.
 - I am developing world-model-based planning that reasons about other agents' intentions.

SysNav: Cross-Embodiment Object Navigation (Robotics)

May 2025 – Jun. 2026

Advisor: Ji Zhang, Jean Oh

- We propose SysNav, a three-level object navigation system that decouples semantic reasoning, navigation planning, and motion control, treating rooms as minimal decision-making units and supporting both standard and conditional object navigation in building-scale real-world environments. In this project,
 - I integrated VLM reasoning into the ROS2-based TARE Planner and deployed the system on a wheeled robot, Unitree Go2 quadruped, and Unitree G1 humanoid, validating the embodiment-agnostic design.
 - I evaluated the system through 190 real-world experiments, achieving 4-5x higher navigation efficiency than existing baselines.
 - This work is accepted by IEEE Robotics and Automation Letters (RA-L) 2026.

STRIVE: VLM-Guided Object Navigation in Simulation (Robotics)

Sep. 2024 – May 2025

Advisor: Jean Oh, Ji Zhang, Jonathan Francis

- We propose STRIVE, a framework that enables robots to incrementally build a multi-layer scene representation during exploration and leverages VLM reasoning for efficient object navigation. We designed a two-stage navigation policy combining high-level VLM-guided planning with low-level exploration methods. In this project,
 - I led the design of the structured scene representation (viewpoint, object, and room nodes) and implemented the two-stage VLM-guided navigation policy.
 - I conducted extensive benchmark evaluations across HM3D v1/v2, RoboTHOR, and MP3D, achieving state-of-the-art performance in both success rate ($\uparrow 7.1\%$) and efficiency ($\uparrow 12.5\%$).
 - This work is accepted by ICRA 2026 and the RSS 2025 SemRob Workshop.

AesStyler: Aesthetic Guided Universal Style Transfer (CV)

Jun. 2023 – Dec. 2023

Advisor: Ran Yi, Yu-Kun Lai, Paul L. Rosin

- We propose AesStyler, a novel Aesthetic Guided Universal Style Transfer method, utilizing pre-trained aesthetic assessment model, a novel Universal Aesthetic Codebook and a novel Universal and Specific Aesthetic-Guided Attention module. Extensive experiments and user-studies have demonstrated that our approach generates aesthetically more harmonious and pleasing results than the state-of-the-art methods. In this project,
 - I proposed and independently implemented three novel contributions, and drafted the initial paper including methodology, figures, and result analysis.
 - I single-handedly conducted all qualitative and quantitative experiments and 2 user-studies.
 - This work is accepted by ACM MM 2024.

SAMVG: An Image Vectorization Model with SAM (CV, CG)

Aug. 2023 – Dec. 2023

Advisor: Ran Yi, Yu-Kun Lai, Paul L. Rosin

- We propose SAMVG, a multi-stage model to vectorize raster images into Scalable Vector Graphics. Extensive experiments demonstrate that SAMVG can produce high quality SVGs in any domain with less computation time and complexity compared to previous SOTA methods. In this project,
 - I co-proposed three innovative aspects with the co-author and finished the final draft of the paper, including the methodology and result analysis.
 - This work is accepted by ICASSP 2024.

M3DM-NR: RGB-3D Noisy Industrial Anomaly Detection (CV)

Dec. 2023 – Jun. 2024

Advisor: Jinlong Peng

- We propose M3DM-NR, a noise-resistant framework leveraging the multi-modal (image and point cloud) discriminative capabilities of CLIP for noisy 3D-RGB anomaly detection. In this project,
 - I proposed the 3-stage noise-resistant framework, designed and conducted all qualitative and quantitative experiments, and drafted the initial paper.
 - This work is accepted by TPAMI 2025.

TECHNICAL SKILLS

- **Languages** Python, C/C++, MATLAB, Mandarin (native), English (fluent, TOEFL: 111)
- **Tools** ROS, PyTorch, TensorFlow, OpenCV, OpenGL, L^AT_EX, Markdown, Git