

# HAOKUN ZHU

☎ +1 4129548608 ✉ [haokunz@andrew.cmu.edu](mailto:haokunz@andrew.cmu.edu) 🌐 [zwandering.github.io](https://zwandering.github.io) 📄 [Haokun Zhu](#) 🌐 [zwandering](#)

## EDUCATION

- **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)

- **Cross-embodiment Object Navigation System**  
• [Haokun Zhu](#), Zongtai Li, Zihan Liu, Kevin Guo, Yuxin Cai, Guofei Chen, Chen Lv, Wenshan Wang, Jean Oh, Ji Zhang  
On Going Project. [Website]
- **STRIVE: Structured Representation Integrating VLM Reasoning for Efficient Object Navigation**  
• [Haokun Zhu](#)<sup>\*</sup>, Zongtai Li<sup>\*</sup>, Zhixuan Liu, Wenshan Wang, Ji Zhang, Jonathan Francis, Jean Oh  
RSS 2025 SemRob Workshop; Under Review(ICRA 2026). [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<sup>\*#</sup>, [Haokun Zhu](#)<sup>\*</sup>, Teng Hu, Yu-Kun Lai, Paul L. Rosin  
ACM MM 2024. [PDF]
- **SAMVG: A Multi-stage Image Vectorization Model with the Segment-Anything Model**  
• [Haokun Zhu](#)<sup>\*</sup>, Juang Ian Chong<sup>\*</sup>, Teng Hu, Ran Yi<sup>#</sup>, Yu-Kun Lai, Paul L. Rosin  
ICASSP 2024. [ArXiv:2311.05276]
- **M3DM-NR: RGB-3D Noisy-Resistant Industrial Anomaly Detection via Multimodal Denoise**  
• Chengjie Wang, [Haokun Zhu](#), Jinlong Peng, Yue Wang, Ran Yi, Yunsheng Wu, Lizhuang Ma, Jiangning Zhang  
TPAMI [ArXiv:2406.02263]
- **Stroke-based Neural Painting and Stylization with Dynamically Predicted Painting Region**  
• Teng Hu, Ran Yi<sup>#</sup>, [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<sup>#</sup>, Siqu Kou, [Haokun Zhu](#), Xu Chen, Yabiao Wang, Chengjie Wang, Lizhuang Ma  
ICCV 2023. [Code] [ArXiv:2309.03729]

## EXPERIENCE & INTERNSHIP

- **roBot Intelligence Group (BIG) in CMU** Pittsburgh, United States  
• *Graduate Research Assistant 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.
  - **Cross-Embodiment Navigation System in Real-World**: integrated VLM reasoning into ROS2-based TARE Planner to enable building-scale object navigation with conditional queries, successfully deploying across wheeled robot, Unitree Go2 quadruped, and Unitree G1 humanoid with 75 real-world demonstrations.
- **Digital Media Computer Vision Laboratory(DMCV) in SJTU** Shanghai, China  
• *Undergraduate Research Assistant advised by Prof Ran Yi* Oct. 2022 – Jun. 2024
  - **Few-shot Image Generation with Diffusion Model**: employed diffusion models to produce high-quality and diverse images in new domains with limited training data.
  - **Aesthetic Guided Universal Style Transfer**: transferred arbitrary image styles to content images while balancing aesthetic qualities, style transformation and content preservation.
  - **Image Vectorization**: transformed raster images into scalable vector graphics with superior adaptability.
- **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.

**Cross-Embodiment Navigation System in Real-World (Robotics)**

May. 2025 - Present

Advisor: Ji Zhang, Jean Oh

- We developed a unified, cross-embodiment object navigation framework that integrates VLM reasoning into ROS2-based TARE Planner, capable of operating across building-scale real environments. The system constructs multi-layer structured scene representations to enable long-horizon, semantically informed decision-making and supports conditional navigation tasks based on object attributes and spatial relations. In this project,
  - I integrated VLM reasoning capabilities into the ROS2-based TARE Planner to enable building-scale autonomous navigation with semantic understanding.
  - I developed and deployed the system across three embodiments: wheeled robot, Unitree Go2 quadruped, and Unitree G1 humanoid, validating the embodiment-agnostic design.
  - I conducted 75 real-world demonstrations (51 on wheeled robot, 18 on quadruped, 6 on humanoid) spanning an entire building floor, demonstrating system scalability and robustness.

**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 module including viewpoint nodes, object nodes, and room nodes to ground VLM understanding in robotic navigation.
  - I implemented core components of the navigation framework and the two-stage policy that integrates VLM reasoning with classical exploration methods.
  - 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 RSS 2025 SemRob Workshop and is under review at ICRA 2026.

**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 three novel contributions and implemented them independently, which empowers AesStyler to produce results aesthetically more harmonious and pleasing compared to current SOTA methods.
  - I completed the initial drafting of the research paper by myself, meticulously explaining the methodology, crafting figures to illustrate the pipelines and analyzing the experimental results in detail.
  - I single-handedly designed and conducted all the experiments, both qualitative and quantitative and 2 user-studies, demonstrating the superiority of AesStyler over current state-of-the-art methods.
  - 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 collaborated with the co-author to propose three innovative aspects.
  - I finished the final drafting of the paper, elaborating in detail on the methodology and conducted a thorough analysis of the experimental results in the paper.
  - 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 novel noise-resistant framework to leverage the strong multi-modal(image and point cloud) discriminative capabilities of CLIP. Extensive experiments show that M3DM-NR outperforms state-of-the-art methods in 3D-RGB multi-modal noisy anomaly detection. In this project,
  - I proposed the 3-stage noise-resistant framework for noisy industrial anomaly detection.
  - I single-handedly designed and conducted all the experiments, both qualitative and quantitative, demonstrating the superiority of M3DM-NR over current state-of-the-art methods.
  - I completed the initial drafting of the research paper by myself, meticulously explaining the methodology, crafting figures to illustrate the pipelines and analyzing the experimental results in detail.
  - This work is accepted by TPAMI.

## TECHNICAL SKILLS

- **Languages** Python, C/C++, Matlab, LaTeX, Mandarin(native), English(fluent, TOEFL: 111)
- **Tools** ROS, PyTorch, TensorFlow, OpenCV, OpenGL, LaTeX, Markdown, git