

Muyi Bao

Phone: +1 412-512-8085 | muyib@andrew.cmu.edu | Personal Website | Google Scholar

Education

Carnegie Mellon University (CMU)

Sep. 2025 – Jun. 2027 (Expected)

M.S. in Electrical and Computer Engineering (Expected), GPA: 4.0/4.0

Xi'an Jiaotong-Liverpool University (XJTLU)

Sep. 2021 – Jun. 2025

B.E. in Computer Science and Technology, GPA: 3.92/4.0 (Top 5%), First Class Honors

Selected Publications ^{*}Equal contribution.

- [1] **Map2Route: Benchmarking Compositional Language-Grounded Route Planning over Semantic Maps.**
Muyi Bao, Hang Xu, Jingfan Tang, Zihan Liu, Yuxin Cai, Chen Lv, Wenshan Wang, and Ji Zhang. Submitted to IEEE International Conference on Robotics and Automation (ICRA), 2027.
- [2] **Goal2Pixel: Grounding Goals to Pixels for Vision-Language Navigation.**
Muyi Bao^{*}, Yuxin Cai^{*}, Hang Xu, Zongtai Li, Jinxi He, Jingfan Tang, Chen Lv, Ji Zhang, Yaqi Xie, and Wenshan Wang. Submitted to IEEE International Conference on Robotics and Automation (ICRA), 2027.
- [3] **IntentNav: Learning Spatial-Visual Object Navigation from Human Demonstrations.**
Yuxin Cai^{*}, Zongtai Li^{*}, Maonan Wang, Muyi Bao, Haokun Zhu, Ruofei Bai, Ding Zhao, Zirui Li, Wenshan Wang, Wei-Yun Yau, Ji Zhang, and Chen Lv. Conference on Robot Learning (CoRL), 2026.
- [4] **Vision Mamba in Remote Sensing: A Comprehensive Survey of Techniques, Applications and Outlook.**
Muyi Bao, Shuchang Lyu, Zhaoyang Xu, Huiyu Zhou, Jinchang Ren, Shiming Xiang, Xiangtai Li, and Guangliang Cheng. Remote Sensing, 2026.
- [5] **FTCFFormer: Fuzzy Token Clustering Transformer for Image Classification.**
Muyi Bao^{*}, Changyu Zeng^{*}, Yifan Wang, Zhengni Yang, Zimu Wang, Guangliang Cheng, Jun Qi, and Wei Wang. European Conference on Artificial Intelligence (ECAI), 2025.
- [6] **ASP-VMUNet: Atrous Shifted Parallel Vision Mamba U-Net for Skin Lesion Segmentation.**
Muyi Bao, Shuchang Lyu, Zhaoyang Xu, Qi Zhao, Changyu Zeng, Wenpei Bai, Guangliang Cheng. arXiv, 2025.

Research Experience

Mobile Manipulation, Graduate Researcher, Carnegie Mellon University, Sep. 2026 – Now

Project: **Agentic Skill Library for Mobile Manipulation (Ongoing)**

Advisor: Prof. Ji Zhang

Sep 2026-Now

- Developing an agentic mobile manipulation framework that uses VLMs to compose hierarchical skill libraries into executable skill graphs, enabling long-horizon task planning, failure recovery, and dynamic skill selection.

Embodied Navigation, Graduate Researcher, Carnegie Mellon University, Dec 2025 – Now

Project: **Universal Map-Based Navigation (Ongoing)**

Advisor: Prof. Ji Zhang

Sep 2026-Now

- Developing a generalist map-conditioned navigation framework that unifies heterogeneous navigation tasks under a shared trajectory-generation interface, e.g., goal-directed, object-centric, and language-conditioned navigation, producing complete routes from semantic maps and task specifications.
- Designing a unified architecture with a semantic-spatial map encoder, a VLM-based reasoning module for instruction reasoning, and a diffusion-based trajectory generator for long-horizon planning across heterogeneous navigation tasks.

Project: **Map2Route: Benchmarking Compositional Language-Grounded Route Planning over Semantic Maps [1]**

Advisor: Prof. Ji Zhang

Jun. 2026-Sep. 2026

- Developed **Map2Route**, a human-curated benchmark of 1,000 episodes across 40 ProcTHOR scenes for evaluating compositional, nested, comparative language grounding and soft-preference-aware route planning over semantic maps.
- Proposed **Grounding2Route**, a Code-as-Policy-style framework that uses LLM-generated executable programs for language grounding, with structured verification and repair followed by sequential A* route planning.
- Conducted a comprehensive empirical study, showing substantial gains over adapted baselines (+28.5% in hard-requirement satisfaction), identifying key components and major sources of task difficulty, and conducting 48 real-robot trials.

Project: **Goal2Pixel: Grounding Goals to Pixels for Vision-Language Navigation** [2]

Advisor: Prof. Ji Zhang and Prof. Wenshan Wang

Dec. 2025–Jun. 2026

- Proposed **Goal2Pixel**, a pure pixel-based navigation paradigm that predicts navigable image pixels instead of low-level actions, achieving 54.1% SR on R2R-CE, a 21.2% improvement over action-based paradigm.
- Developed Visibility-Aware Keyframe Memory (ViKeyMem) for long-horizon navigation, which selects keyframes from historical observations based on visibility, improving SR by 9% on average while reducing training/inference time by nearly 50%.
- Introduced semantic embeddings on special tokens and coordinate-aware auxiliary losses to adapt pretrained vision- language models to image-space navigation and improve sensitivity to action regions and pixel coordinates.

Project: **IntentNav: Learning Spatial-Visual Object Navigation from Human Demonstrations** [3]

Advisor: Prof. Ji Zhang and Prof. Chen Lv

Dec. 2025-Jun. 2026

- Co-developed IntentNav, a BEV-grounded spatial-visual imitation framework for ObjectNav, achieving 53.8%, 70.5%, and 82.2% SR on MP3D, HM3D-v1, and HM3D-v2, respectively.
- Contributed to Frontier-based Human-Intent Labeling, a rollout-data generation algorithm that recovers high-level search intent from low-level human demonstrations to produce large-scale candidate-level waypoint supervision.
- Contributed to evaluation and cross-embodiment validation on wheeled, quadruped, and humanoid robots.

Computer Vision, Undergraduate Research, XJTLU and UoL

Sep. 2023 – May. 2025

Focused on architectural innovation for performance improvement in computer vision, with experience designing and adapting CNN-, Transformer-, and Mamba-based models across image classification, segmentation, and detection tasks.

Project: **A Survey of Vision Mamba in Remote Sensing** [4]

Advisor: Prof. Guangliang Cheng

- Conducted a systematic review of **120+ papers** and developed a taxonomy covering Vision Mamba backbones, micro- and macro-architecture advances, downstream applications, and open challenges and future directions.

Project: **FTCFormer: Fuzzy Token Clustering Transformer for Image Classification** [5]

Advisor: Prof. Wei Wang

- Designed fuzzy-token-clustering-based downsampling for Vision Transformers; and this method outperformed the TCFormer baseline across 32 classification datasets, with up to a 1.43% average accuracy improvement.

Project: **ASP-VMUNet: Atrous Shifted Parallel Vision Mamba U-Net** [6]

Advisor: Prof. Guangliang Cheng

- Designed a CNN–Mamba hybrid architecture with atrous scanning (Mamba-specific component) and local-global feature fusion for skin lesion segmentation, achieving an average 0.92% mIoU gain over the next-best methods.

Skills

- **Programming:** Python, C++ , C, MATLAB
- **Machine Learning & Robotics:** PyTorch, Hugging Face Transformers, LoRA/PEFT, VLM fine-tuning, CUDA, OpenCV, ROS 2
- **Systems & Tools:** Linux, Git, SLURM, Weights & Biases
- **Languages:** Mandarin (Native), English (TOEFL iBT: 100/120)