

Yifei Yang

Email: yifei05.yang@gmail.com | Tel: (+86) 13326268486

Education

Beijing Jiaotong University, China

Sep 2023 – Jun 2027 (exp)

Lancaster University, UK

Sep 2023 – Jun 2027 (exp)

Bachelor of Engineering in Communication Engineering

Publications

Yang, Y., Fan, Z., Yan, Z., Li, Z., & Wang, Y. (2026). SkipWorld: Latent Skip Planning with Distilled World Models for Efficient Visual Navigation. *arXiv preprint arXiv:2610.18135*. [<https://arxiv.org/abs/2610.18135>] [Under Review]

Yang, Y., Fan, Z., Li, H., Wang, A., Huang, L., Yu, H., ... & Wang, Y. (2026). AR Forcing: Towards Long-Horizon Robot Navigation World Model. *arXiv preprint arXiv:2605.31314*. [<https://arxiv.org/abs/2605.31314>]

Fan, Z., Lyu, W., Song, W., Zhao, L., **Yang, Y.**, Wang, X., ... & Gao, F. (2026). PROSPECT: Unified Streaming Vision-Language Navigation via Semantic--Spatial Fusion and Latent Predictive Representation. *arXiv preprint arXiv:2603.03739*. [<https://arxiv.org/abs/2603.03739>] [Accepted by *IEEE Robotics and Automation Letters (RA-L)*]

You, Z., Fan, Z., Lyu, W., Wang, A., Wang, N., **Yang, Y.**, ... & Wang, Y. (2026). Language2Drive: Do Vision-Language-Action Models Really Follow the Instructions in Autonomous Driving? *arXiv preprint arXiv:2610.16634*. [<https://arxiv.org/abs/2610.16634>] [Under Review]

Fan, Z., He, J., Song, W., Wang, X., Lyu, W., Zhao, L., Li, Fu., You, Z., **Yang, Y.**, ... & Wang, Y. (2026). MobileWAM: Bridging World Action Models to Mobile Manipulation with Chain-of-Foresight. *arXiv preprint arXiv:2608.04657*. [<https://arxiv.org/abs/2608.04657>] [Under Review]

Research Experiences

Latent Skip Planning for Efficient Visual Robot Navigation (First Author)

May 2026 – Oct 2026

Research Intern, The Chinese University of Hong Kong (CUHK)

Advisor: Prof. Zhongyu Li (Assistant Professor of the Chinese University of Hong Kong)

- Investigated the computational bottleneck of long-horizon world-model-based robot navigation and developed a latent skip planning framework for efficient trajectory evaluation.
- Designed a planning strategy combining fast skip future prediction with accurate autoregressive trajectory refinement, reducing redundant rollout computation.
- Evaluated the proposed method on four visual navigation benchmarks (RECON, HuRoN, TartanDrive, and SCAND), achieving an average **ATE of 3.21** and **RPE of 0.71** while reducing planning time from **2110 seconds to 840 seconds** compared with NWM.

AR Forcing: Towards Long-Horizon Robot Navigation World Model (First Author)

Sep 2025 - Mar 2026

Research Intern, Institute for AI Industry Research (AIR), Tsinghua University

Advisor: Prof. Yan Wang (Assistant Professor of the Institute for AI Industry Research, Tsinghua University)

- Identified the train-test mismatch problem in diffusion-based robot navigation world models and proposed AR Forcing, an autoregressive training strategy that exposed models to their inference distributions.
- Implemented AR Forcing training pipeline by integrating diffusion objectives into autoregressive rollout, enabling models to recursively update contexts with self-generated predictions.
- Conducted extensive experiments on four visual navigation benchmarks (RECON, SCAND, HuRoN, and TartanDrive), improving long-horizon prediction consistency and trajectory accuracy and reducing the **16-second LPIPS metric from 0.533 to 0.463** on RECON compared with NWM.

PROSPECT: Unified Streaming Vision-Language Navigation via Semantic-Spatial Fusion and Latent Predictive Representation [Accepted by IEEE Robotics and Automation Letters (RA-L)]

Apr 2025 - Mar 2026

Research Intern, Institute for AI Industry Research (AIR), Tsinghua University

Advisor: Prof. Yan Wang (Assistant Professor of Institute for AI Industry Research, Tsinghua University)

- Developed a unified streaming vision-language navigation framework by integrating a streaming Vision-Language-Action policy with latent predictive representation learning.
- Implemented semantic-spatial fusion modules combining 3D spatial features from CUT3R and semantic representations from SigLIP through cross-attention for long-horizon navigation.
- Evaluated the framework on VLN-CE benchmarks and real-robot platforms, achieving state-of-the-art navigation performance with **58.9% success and 54.0% SPL on R2R val-unseen** and demonstrating improved robustness across diverse environments.

MobileWAM: Bridging World Action Models to Mobile Manipulation with Chain-of-Foresight Mar 2026 - Aug 2026
Research Intern, Institute for AI Industry Research (AIR), Tsinghua University

Advisor: Prof. Yan Wang (Assistant Professor of Institute for AI Industry Research, Tsinghua University)

- Investigated the adaptation of pretrained video world models for mobile manipulation by integrating video diffusion models with lightweight action experts.
- Implemented the Chain-of-Foresight training strategy to provide dense future supervision through sequential latent prediction without additional inference cost.
- Conducted experiments on ManiSkill-HAB and real-world mobile manipulation platforms, achieving a **73.0% success on ManiSkill-HAB** and improving generalization across diverse whole-body manipulation tasks.

A Low and High-Frequency Fusion Framework for Anti-UAV Detection and Tracking May 2024 - May 2025
Project Leader, Undergraduate Innovation Project, Beijing Jiaotong University

Advisor: Prof. Hang Zhou (Associate Professor of School of Electronic and Information Engineering, Beijing Jiaotong University)

- Led research project on anti-UAV detection under challenging scenarios including low illumination and dynamic backgrounds.
- Designed a low-high frequency feature fusion framework with Decoupling Light Reconstruction (DLR) and Null Space Super-Resolution Model (NSRM) modules to recover motion details.
- Built the complete training and evaluation pipeline and validated the proposed method on anti-UAV detection benchmarks, and received the Beijing Municipal Undergraduate Innovation Project Award.

Work Experiences

Algorithm Intern, Wuxi Research Institute of Applied Technologies, Tsinghua University Jul 2025-Aug 2025
VLA and World Models for Robot Manipulation and Navigation

- Integrated a robotic arm and multi-sensor stack (camera, force and torque) in ROS2 and debugged end-to-end communication for stable runtime operation.
- Implemented model-to-control interfaces for VLA models (π 0-series and world-model inference), enabling direct execution of predicted actions on the robot.
- Validated grasping, placement, and container-operation demos with sim-to-real consistency by unifying observation and action formats and task scripts.

Honors and Awards

Award of S, Mathematical Contest in Modeling (MCM) May 2024

Beijing Municipal Undergraduate Innovation Project May 2025

Skills

Programming: Python, PyTorch, Multi-GPU distributed training and debugging.

Foundations: Computer Vision and VLA, diffusion models and recent VLA (Vision-Language-Action), WAM (World Action Model), and WMA (World Model Action) methods.

Hardware: Schematic and multi-layer PCB design.

English: IELTS overall band score 6.5