

Yangcheng Yu

+86 19870223130 | yuyc23@mails.tsinghua.edu.cn | <https://cocoyawn.github.io/>

EDUCATION

Tsinghua University, Beijing, China

09/2023 - 07/2027

Bachelor of Engineering in Electronic Engineering | GPA: 3.75/4.0 (rank 46/266, top 20%)

SELECTED PUBLICATIONS

1. **Yangcheng Yu**, et al. (2025). MoWM: **Mixture-of-World-Models** for embodied planning via latent-to-pixel feature modulation. arXiv. <https://doi.org/10.48550/arXiv.2509.21797>. *ICML 2026 Conference Submission*.
2. Lixuan He, Haoyu Dong, **Yangcheng Yu**, et al. (2025). Learning to Navigate in Open Urban Environments Using a Simple Sim2Real Strategy. *ICLR 2025 Workshop EmbodiedAI*.
3. Haohan Chi, Huan-ang Gao, Ziming Liu, Jianing Liu, Chenyu Liu, Jinwei Li, Kaisen Yang, **Yangcheng Yu**, et al. (2025). Impromptu VLA: Open Weights and Open Data for Driving **Vision-Language-Action Models**. *NeurIPS 2025 (Datasets and Benchmarks Track)*.
4. Mingju Gao, Kaisen Yang, Huan-ang Gao, Bohan Li, Ao Ding, Wenyi Li, **Yangcheng Yu**, et al. (2025). PAM: A Pose–Appearance–Motion Engine for Sim-to-Real HOI Video Generation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2026*.

INTERNSHIPS

Manifold AI, Beijing, China

06/2025 - 01/2026

MoWM: Mixture-of-World-Models for Embodied Planning via Latent-to-Pixel Feature Modulation

- Proposed a **hybrid world model architecture for embodied action planning** that fuses latent motion-aware representations with fine-grained pixel features to addressing limitations for both pixel-level and latent only world models and significantly improve embodied action planning performance in **robotics** tasks.
- Introduced latent-to-pixel feature modulation, leveraging high level latent motion priors to guild extraction of actionable fine-grained visual details and enabling more precise action decoding and robust generalization with average **+11.2%** average task success improvement over no-fusion baselines.
- Achieved **state-of-the-art results on the CALVIN benchmark**, improving average task success rate by **+5.7%** over competitive VLA-based methods and **+13.4%** over prior world model-based baselines across multi-stage tasks.

WMA 3.0 Model

- Developed a unified Reasoner-Dreamer-Actor architecture to synchronize language, video generation and robot control by using the unified **Action VAE model**.
- Curated 7 high quality heterogeneous robotic and human arm datasets by processing over **10 million action trajectories**, enabling effective large-scale training and robust model performance.
- Developed a universal action representation bridging the gap between diverse robot embodiments and human demonstrations and facilitating stable and scalable training for the Action VAE model across 5+ heterogeneous robotic platforms.
- Validated the model’s ability to generalize across multiple high-dimensional datasets while maintaining semantic consistency in the latent space.

RESEARCH EXPERIENCES

Research Assistant, Institute for AI Industry Research, Tsinghua University

06/2025 - 09/2025

PAM: A Pose-Appearance-Motion Engine for Sim-to-Real PAM Video Generation

06/2025 - 09/2025

- Co-designed a decoupled three-stage framework for **video generation**, disentangling pose, appearance, and motion through sequential synthesis of motion trajectories, photorealistic initial frames, and video sequences

that help to bridge the gap between simulation and real-world scenarios, providing a robust and flexible solution for sim-to-real transfer in embodied AI.

- Designed multi-conditional geometric control mechanisms, incorporating depth maps, segmentation masks, and hand key points as intermediate signals to improve geometric consistency by **35.5%** (DexYCB) and **51.6%** (OAKINK2) in terms of MPJPE reduction.
- **Led comprehensive ablation studies on input conditions and experimental configurations**, demonstrating that multi-condition combinations consistently achieved best performance across 6 evaluation metrics, including **FVD** and **MPJPE**.
- Validated downstream performance on **SimpleHand**, demonstrating that augmenting training with PAM-generated synthetic videos allows models trained on **only 50% real data to match the performance of a full 100% real-data baseline**.

Impromptu VLA

05/2025 - 06/2025

- **Curated and distilled a large-scale multimodal driving dataset (80K + clips)** concentrated on unstructured and long-tail urban driving scenarios by filtering over 2M raw video clips from 8 open-source driving corpora.
- **Enabled advanced semantic annotation workflows** for visual observations via the combination of LLM/VLM and human-in-the-loop pipelines, generating planning-oriented Q&A and textual description that capture complex traffic logic and temporary rule changes.
- **Conducted multimodal feature analysis and outlier mining** by encoding video clips into high-dimensional embedding spaces using CLIP model and applying **clustering/nearest-neighbor screening** atypical non-structured driving scenarios.
- **Refined and validated over 80K video clips**, correcting noisy annotations and enforcing **temporal consistency** including boundaries, state transitions, to ensure coherent decision-making trajectories for VLA training and evaluation.

Future Intelligence Lab, Tsinghua University

11/2024 - 05/2025

Learning to Navigate in Open Urban Environments Using a Simple Sim2Real Strategy

- **Designed, assembled, and debugged an embodied robotic platform for a Sim2Real vision and language navigation (VLN)**, enabling stable real-world execution of language based navigation policies under open urban environments.
- **Built and maintained a full stack robotic system based on ROS2 and Xiaomi Cyberdog**, supporting multimodal sensor integration, system calibration, and runtime robustness for memory augmented navigation frameworks on-robot evaluation.
- **Implemented and tested remote teleoperation and communication pipelines**, facilitating real-world data collection and system-level validation of memory-enhanced VLN models.

HONORS & AWARDS

Innovation Excellence Scholarship (National Endeavor Scholarship), Tsinghua University, 2025

Innovation Excellence Scholarship (National Endeavor Scholarship), Tsinghua University, 2024

1st Prize, National Undergraduate Physics Competition (Regional Level), Beijing Physics Society, 2024

2nd Prize, Tsinghua AI Challenge (Team Leader), Tsinghua University, 2024

2nd Prize, Tsinghua Hardware Design Competition (Team Leader), Tsinghua University, 2024

2nd Prize, National Mathematical Olympiad (Provincial Level), China Mathematical Society, 2022

2nd Prize, National Physics Olympiad (Provincial Level), China Physics Society, 2022

SKILLS

Programming: Proficient in Python, C/C++, C#, R, MATLAB, and experienced with deep learning frameworks

Frameworks/Tools: Linux, Git, CUDA, PyTorch

Mathematics: Capable of solving problems in linear algebra and probability theory encountered in research