

Xiangyu Zeng

Tel: +12177213642 · Email: xzeng28@illinois.edu · Web: zengxy0624.com

RESEARCH INTERESTS

Vision-Language-Action models for robot manipulation: adaptive dual-system architectures, preference-aligned policy transfer, and real-world imitation learning infrastructure.

EDUCATION

University of Illinois Urbana Champaign 09/2025-Present

- MEng in Mechanical Engineering, concentration in Control & Robotics

Wuhan University of Technology 09/2021-06/2025

- Bachelor of Engineering in Intelligent Manufacturing Engineering GPA: 4.071/5.0

HONORS & AWARDS

- Seichi Konzo Memorial Award, recognizing an outstanding student in Controls and Robotics, Department of Mechanical Science and Engineering, University of Illinois Urbana-Champaign, 2026
- First Prize of Yanchang Petroleum Scholarship (Top 0.5%), Wuhan University of Technology, 2023
- University-level First Scholarship (Top 3%), Wuhan University of Technology, 2022-2024
- Outstanding Student Leader, Wuhan University of Technology, 2023
- Independent Innovation Research Fund (Undergraduate Program), Wuhan University of Technology, 2022

PUBLICATIONS

[1] **Xiangyu Zeng***, Yiyang Wei*(co-first author), Xirui Chen. HTSA-LSTM: Leveraging Driving Habits for Enhanced Long-Term Urban Traffic Trajectory Prediction. Accepted by *Applied Sciences*.

[2] Haoran Yuan, Weigang Yi, Zhenyu Zhang, Wendi Chen, Yuchen Mo, Jiashi Yin, Xinzhuo Li, **Xiangyu Zeng**, Chuan Wen, Cewu Lu, Katherine Driggs-Campbell, Ismini Lourentzou. VTAM: Video-Tactile-Action Models for Complex Physical Interaction Beyond VLAs. Submitted to *ECCV 2026*.

[3] Xu F, Zong W, Zeng X, et al. Construct a health indicator for bearing based on unsupervised SDAE with Euclidean distance[J]. *Journal of Vibration and Control*, 2025: 10775463251342614

RESEARCH EXPERIENCES

Preference-Aware Generalization in Vision-Language-Action Models for Robot Manipulation
12/2025–Present

Lab Member, UIUC ECE, advised by Prof. Katherine Driggs-Campbell CoRL 2026 (2nd author)

- Studied cross-task generalization of preference-conditioned behaviors (e.g., obstacle avoidance style) in VLA models, contributing to experimental design, large-scale data collection, and evaluation infrastructure—addressing the gap that existing imitation learning benchmarks evaluate only task completion, not execution style
- Implemented and iterated the full experimental pipeline—including parameterized waypoint trajectory generation, multi-configuration data collection (9,600 episodes, 250 GB across 83 task configurations), and obstacle-proximity evaluation metrics—on 50 bimanual tasks on RoboTwin 2.0 (CVPR 2025 Highlight) with 259 language instruction templates
- Fine-tuned π_0 VLA flow model with LoRA and evaluated preference generalization across seen/unseen task splits, with Diffusion Policy as baseline; contributed to the skill-decoupling analysis via visual and textual randomization experiments

End-to-End Robot Imitation Learning Infrastructure: From Demonstration to Deployment
11/2025–Present

Lab Member, UIUC ECE, advised by Prof. Katherine Driggs-Campbell

- Independently built a general-purpose imitation learning infrastructure integrating FastUMI collection,

Diffusion Policy control framework, and LeRobot data standard into an end-to-end pipeline from human demonstration to closed-loop deployment; supports dual-arm UR5e and single-arm Kinova, reused as the unified platform for all lab real-robot experiments

- Designed online multi-source sensor fusion for heterogeneous-frequency streams (SLAM 500Hz, Vive 100Hz, gripper 200Hz, RGB 60Hz) with a fault-latching state machine—BAD-state latch with 50-frame hysteresis release models sensor failures as persistent events, ensuring temporal consistency of fused trajectories
- Implemented full-chain coordinate frame alignment from collection to deployment (SLAM device → world → robot base → end-effector), including device-grip rotation correction and cross-platform extrinsic calibration, enabling direct trajectory replay across robot platforms
- Built automated data quality pipeline with full-order kinematic analysis, adaptive curvature-based thresholds, and cumulative-displacement evaluation masks; standardized conversion to LeRobot v0.3 format for downstream policy training
- Achieved accurate trajectory replay on both UR5e and Kinova with 10D state-action representation (3D position + 6D rotation + gripper) across 53 collected sessions—validating end-to-end correctness of the full pipeline

Object-Centric Feature Fusion for Diffusion-Based Manipulation Policies 01/2026–Present
Course Research Project, UIUC CS 444 (Deep Learning for Computer Vision)

- Designed and implemented a two-stage pipeline: a visual grounding module localizes the instruction-specified target via cross-attention between visual and language tokens, outputting a bounding box as an interpretable attention bottleneck; a diffusion policy generates 7-DoF action chunks conditioned on fused global scene, ROI-pooled object-centric, and language features. Independently implemented all components in PyTorch
- Built a zero-annotation grounding data pipeline that exploits simulator instance segmentation to automatically generate (image, instruction, bbox) supervision, eliminating manual labeling and enabling scalable dataset creation (10k+ samples). Designed three-way ablation (no grounding / frozen CLIP / task-specific trained grounding) to isolate whether grounding improvement stems from vision-language pretraining or task-specific fine-tuning; extending to RoboTwin 2.0 preference-conditioned scenarios

Enhanced Long-Term Urban Traffic Trajectory Prediction (HTSA-LSTM) 06/2024–06/2025
Supervised by Dr. Yiyang Wei, Wuhan University of Technology

- Designed HTSA-LSTM integrating dual spatiotemporal attention with unsupervised driving style analysis (SICC-SC) to extract driving primitives and cluster trajectory patterns without predefined labels
- Achieved 20.72% RMSE reduction and 24.98% NLL reduction on NGSIM for 5-second predictions over benchmark models

TECHNICAL SKILLS

Programming Languages: Python, MATLAB

ML/DL Frameworks: PyTorch, HuggingFace LeRobot

Robotics Platforms: UR5e, Kinova Gen3, FastUMI Pro

Simulation & Tools: ROS/ROS2, SAPIEN, MuJoCo/robosuite, LIBERO, Isaac Sim/Isaac Gym

Methods: Imitation Learning, Diffusion Policy, Vision-Language-Action Models, Multi-Sensor Fusion, Sim-to-Real Transfer

REFERENCES

Prof. Katherine Driggs-Campbell (Research Advisor)

krdc@illinois.edu

Associate Professor, Department of Electrical and Computer Engineering, UIUC