

Yichi Zhang

each0521@gmail.com — +1 9342464246 — Website — GitHub — Google Scholar — LinkedIn

Education

Stony Brook University

Ph.D. in Computer Science

Fall 2024 – Present

GPA: 3.81 / 4.0

Shanghai Jiao Tong University

B.Eng. in Computer Science and Technology

Sep 2018 – Jun 2022

GPA: 3.4 / 4.3

Publications

E-Ro Nguyen*, **Yichi Zhang***, Kanchana Ranasinghe, Xiang Li, Michael S. Ryoo. “Pixel Motion Diffusion is What We Need for Robot Control.” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Mengping Yang*, Ceyuan Yang*, **Yichi Zhang**, Qingyan Bai, Yujun Shen, Bo Dai. “Revisiting the Evaluation of Image Synthesis with GANs.” *NeurIPS Datasets and Benchmarks*, 2023.

*Equal contribution

Research Projects

Pixel Motion Diffusion for Robot Control

Jun 2025 – Dec 2025

PhD Research Assistant — Advisor: Michael S. Ryoo

- Investigated pixel-motion representations as intermediate representations. Designed a robot control framework conditioned on motion representations generated by a high-level motion director model.
- Built real-world benchmarking environments on single and bi-manual robot arms. Implemented inference pipelines for **3** policies and evaluated them on **7** real-world manipulation tasks.
- Reproduced and benchmarked representative baselines including Diffusion Policy, VPP, and π_0 in both simulation and real-world settings. Achieved near-SOTA on CALVIN, MetaWorld, and real-world benchmarks.

Evaluation of Image Synthesis Models

Jul 2022 – Feb 2023

Researcher Engineer — Advisors: Ceyuan Yang, Bo Dai

- Conducted large-scale empirical analysis of GAN evaluation metrics across **16** generative models and **3** widely used datasets: ImageNet, FFHQ, and LSUN Church.
- Identified systematic weaknesses in commonly used metrics through statistical analysis, Grad-CAM-based visualization, and human evaluation.
- Proposed alternative evaluation metrics with improved reliability and stronger correlation with human judgment. Validated through quantitative experiments and human studies.

Video-to-Sound-Effect Generation

Mar 2023 – Sep 2023

Researcher Engineer — Advisor: Bo Dai

- Studied cross-modal generation for synthesizing sound effects conditioned on visual video inputs.
- Constructed audio-video training pipelines and evaluated candidate datasets including UnAV100, Car Passing By, and Greatest Hits.
- Applied diffusion-based audio generation models to improve high-frequency detail and timbre realism.
- Implemented conditioning mechanisms using text and video labels to improve multimodal alignment.

Robot Chassis Control and Localization

Sep 2019 – Sep 2020

Algorithm Developer

- Developed low-level robot control algorithms using PID control and encoder feedback; integrated gyroscope signals for improved motion precision.
- Implemented trajectory reconstruction methods for field localization using encoder-derived motion data.

Professional Experience

JAKA Robotics

Engineer Intern, Computer Vision Group

Apr 2024 – Jul 2024

Advisor: Cheng Ding

- Developed a speech-driven robot control system for a bimanual humanoid robot to execute natural-language manipulation instructions.
- Built the speech input pipeline using an iFLYTEK microphone and speech-to-text system for real-time command recognition.
- Integrated a Qwen-based large language model to convert spoken instructions into structured motion plans for robot execution.
- Incorporated YOLO-based object detection for target localization and bounding-box estimation to ground language commands in the visual scene.
- Connected perception outputs with end-effector control to enable object manipulation based on spoken instructions. Maintained accuracy via camera calibration.

Shanghai Artificial Intelligence Laboratory

Research Engineer

Jul 2022 – Oct 2023

Advisor: Bo Dai

- Conducted research on the comprehensive evaluation of image generative models, including VAEs, GANs, and diffusion models.
- Explored and developed model architectures and training pipelines for sound-effect generation with video alignment tasks using diffusion models.
- Contributed to internal research infrastructure, technical documentation, and codebases supporting multiple research projects.

Skills

Programming: Python, C/C++, JavaScript

Machine Learning: PyTorch, Diffusion Models, Large Language Model, Vision Language Model

Robotics: Policy Learning, Manipulation Systems, Visual Servoing, Hardware Build-up

Systems/Tools: Linux, ROS-1&2, Git, TensorRT, OpenCV, Jetson Nano, Intel RealSense

AI Agents: Claude, GPT & Codex, Gemini, Copilot

Honors

World Champion, VEX Robot World Championship Skills Challenge (2023)

All-round Award, World Robot Contest (2020)