Zhutian (Skye) Yang

ztyang@mit.edu | https://www.zt-yang.com

Motivation

I am a final-year PhD candidate in robotics and embodied intelligence at MIT CSAIL. My research focuses on developing algorithms for solving ***multi-step manipulation problems*** in ***geometrically complex environments***, such as packing objects in boxes or organizing the office. I use a ***combination of learning*** (e.g., vision language models, diffusion models, RL) and ***model-based planning methods***. I am excited to apply my skills in deep learning and simulation in a full-time research/engineering position in robotics/ML.

Education

Massachusetts Institute of Technology (MIT) Cambridge, MA
Ph.D. in Electrical Engineering & Computer Science (EECS); GPA: 4.9/5 *Expected Feb 2025*

- **Minor Studies**: Minor in Computer Vision.
- **Selected Technical Courses**: 6.843 Robotic Manipulation. 6.863J Natural Language Processing. 16.485 Visual Navigation for Autonomous Vehicles. 6.246 Dynamic Programming & Reinforcement Learning. 9.357 Touching and Grasping with Soft Fingers and Hands. 6.438 Algorithms for Inference.

Nanyang Technological University (NTU) Singapore
B.Eng. in Electrical Engineering, major in Information Engineering and Media; GPA: 4.92/5 *Jul 2019*

- **Award**: Lee Kuan Yew Gold Medal, awarded as the top student graduated in the major from NTU.

Research Experience

Large Behaviors Team, Robotics, Toyota Research Institute Cambridge, MA
Research Intern (Full-time & Part-time), advised by Russ Tedrake *June 2024 - Present*

- Developing ***language-conditioned multi-skill policies*** to solve long-horizon mobile manipulation problems and chaining them by ***fine-tuning vision language models*** for skill planning.

Learning and Intelligent Systems Group, CSAIL, MIT Cambridge, MA
Graduate Research Assistant, co-advised by Leslie Kaelbling and Tomás Lozano-Pérez *Jan 2021 - Present*

- Developed a general-purpose solver for ***continuous constraint satisfaction problems*** (CCSP) in multi-step robot manipulation by composing ***diffusion models***. See: *https://diffusion-ccsp.github.io/*
  - Constraints include geometric collision-free, physical stability, and data-defined spatial constraints.
  - The method, Diffusion-CCSP, finds solutions to continuous variables that satisfy all constraints by composing the scores from multiple diffusion models trained for individual constraint types.
  - It has been applied to robotic domains such as solving geometric fitting puzzles (success rate 80-100%), stacking objects between shelves (success rate 40-80%), and packing objects in a box (success rate 85-100%). It is able to generalize to problems with more objects than being trained on.
- Developed a ***hierarchical mobile manipulation*** policy for handling objects with unknown dynamics:
  - The high-level motion planner proposes way points that a local diffusion policy follows. It enabled a Boston Dynamics Spot robot to rearrange office chairs in cluttered spaces (success rate 8/10 compared to 0 and 5/10 for baseline motion planning and diffusion policy). It generalizes to new environments (7/10), new chairs (5/10), and new initial conditions (5/10). Paper is in submission.

Seattle Robotics Lab, NVIDIA Cambridge, MA
Research Intern (Part-time), advised by Dieter Fox *Sep 2023 - May 2024*

- Developed an general algorithm for solving multi-step mobile manipulation problems combining the commonsense capabilities of ***large pre-trained vision-language models*** and the geometric soundness guarantee of ***task and motion planners***. See: *https://zt-yang.github.io/vlm-tamp-robot/*
- It solves cooking motion planning problems that require performing 30-50 actions in sequence and interacting with up to 21 objects. It achieves 50 to 100% success rate (given various world initial state and robot embodiments), versus 0% by baselines that doesn't use TAMP to solve geometric infeasibility.

Seattle Robotics Lab, NVIDIA Seattle, WA
Research Intern, mentored by Caelan Reed Garrett, advised by Dieter Fox *May - Aug 2022*

- Developed a ***novel Transformer-based architecture***, PIGINet, that predicts plan feasibility based on the initial state, goal, and candidate plans, fusing image and text embeddings with state features. PIGINet ***reduced runtime by 50%-80%*** on pick-and-place kitchen problems with articulated and movable obstacles, after training on only 300-600 problems. It also achieves ***zero-shot generalization to unseen object geometry*** thanks to its visual encoding of objects. See: *https://piginet.github.io/*

Adaptive Computing Lab, National University of Singapore Singapore
Research Intern, advised by Professor David Hsu *Sep 2018 - Dec 2018*

- Developed an ***interactive task learning system*** that generates hierarchical task networks through natural language conversations with human users, which was successful in 7 out of 11 human tests.
- Integrated the task learning system with controllers and visual grounding modules for a ***Kinova Jaco arm*** using ROS to demonstrate learning and making simple cuisines, such as breakfast cereal and chopped fruit salad. Video demo: *https://groups.csail.mit.edu/genesis/Archive/FruitSalad.mp4*
- Constructed a 3D table-setting simulation in Unity. Collected a dataset of natural language instructions and corresponding human motions through a 2D table-setting web interface.

Selected Publications

Yang, Z., Garrett, C., Kaelbling, L., Lozáno-Pérez, T., & Fox, D.. Guiding Long-Horizon Task and Motion Planning with Vision Language Models. *arXiv.2410.02193.* https://zt-yang.github.io/vlm-tamp-robot/

Ravan, Y., **Yang, Z.**, Chen, T., Lozáno-Pérez, T., & Kaelbling, L.. Combining Planning and Diffusion for Mobility with Unknown Dynamics. *in submission; arXiv coming soon.*

Yang, Z., Mao, J., Du, Y., Wu, J., Tenenbaum, J., Lozáno-Pérez, T., & Kaelbling, L.. Compositional Diffusion-Based Continuous Constraint Solvers. *The Conference of Robot Learning 2023.* https://diffusion-ccsp.github.io/

Yang, Z., Garrett, C., Kaelbling, L., Lozáno-Pérez, T., & Fox, D.. Sequence-Based Plan Feasibility Prediction for Efficient Task and Motion Planning. *Robotics: Science and Systems 2023.* https://piginet.github.io/

Yang, Z., Curtis A. Lets Handle It: Generalizable Manipulation of Articulated Objects. *The International Conference on Learning Representations 2022, Workshop on Generalizable Policy Learning in the Physical World.* Won **2nd place in the ManiSkill Challenge 2022 Robotics Track**.

Yang, Z., Kryven, M., Shrobe, H., & Tenenbaum, J. Modeling human planning in a life-like search-and-rescue mission (Poster). *In Proceedings of the Annual Meeting of the Cognitive Science Society, 2021.*

Yang, Z., Winston, P. H. Learning by Asking Questions and Learning by Aligning Stories: How a Story-Grounded Problem Solver can Acquire Knowledge. Technical Report in *DSpace@MIT, 2018.*

Awards & Scholarships

EECS David S Y Wong Fellowship (*2019*): Granted for outstanding graduate application.

Lee Kuan Yew Gold Medal (*2019*): Awarded as the top student graduate in the major from NTU.

SM2 Scholarship (*2014 - 2019*): Granted a full-tuition scholarship by the Singapore Ministry of Education.

Skills

Programming Languages: Python, C++; JavaScript, HTML, CSS, MySQL; MatLab.

Technical Skills: ROS; PyTorch, CUDA; Isaac Gym, Isaac Sim, Drake, PyBullet; Game development using Unity; motion video production using Adobe Premiere and AfterEffects; music production using Logic Pro.

Hobbies: Triathlon; Kickboxing (AFAA certified group exercise instructor); Singing; Improv comedy.