Zhutian Yang

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

Motivation

I am a final-year PhD candidate in robotics at MIT CSAIL. My research focuses on developing algorithms for solving ***multi-step manipulation problems*** in ***complex and diverse environments***. I combine deep learning and planning-based methods to achieve policy generalization to diverse objects, scenes, and tasks. I am excited to apply my experience in ***multi-modal deep learning***, training and deploying large models, ***designing algorithms for solving complex manipulation problems***, and generating long-horizon visual-language-action data to a Research Scientist position in Robotics or Machine Learning.

Education

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

- **Thesis**: Solving Long-Horizon Mobile Manipulation with Generative Planning and Action Models
- **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 Behavior Models Team, Robotics, Toyota Research Institute Cambridge, MA
Research Intern (Full-time & Part-time), advised by Russ Tedrake *June 2024 - Expected Mar 2025*

- Developed a hierarchical multi-task policy architecture where a Robot Visual Planning and Reasoning Network generates language goals and guides a low-level ***language-conditioned multi-skill policy***.

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

- Focused on ***multi-step robot manipulation*** by ***composing and chaining generative models***.
  - Developed a general-purpose solver for ***continuous constraint satisfaction problems*** (CCSP) in multi-step robot manipulation by ***composing diffusion models***. The method, Diffusion-CCSP, finds solutions to continuous variables (e.g. SE3 poses) that satisfy all constraints by composing the scores from multiple diffusion models trained for individual constraint types (e.g. ***geometric collision-free, physical stability, and data-defined spatial constraints***). It has been applied to robotic domains such as stacking 5+ objects between shelves (success rate 40-80%), and ***packing objects in a box*** (success rate 85-100%). See: *Project Page*
  - Developed a ***hierarchical policy*** for ***long-horizon mobile manipulation of objects with unknown dynamics***, named PoPi. The high-level motion planner proposes ***key poses that a local diffusion policy follows***. PoPi 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). See: *Project Page*
- Mentoring MIT undergraduate and master students on research projects in ***imitation learning*** and ***reinforcement learning*** for ***humanoid whole-body manipulation*** (e.g. Unitree G1).
  - Aiming to achieve long-horizon humanoid mobile manipulation by training ***generalizable x-conditioned policies*** (where x can be language, goal bounding box, end-effector or base poses, hidden states, etc.) on sim and real robot data, alongside training ***high-level x-generation policies*** on diverse data sources, including VR teleop data, augmented synthetic data, TAMP generated data, and open-source human retargeting data.

Seattle Robotics Lab, NVIDIA Seattle, WA

Research Intern (Full-time & Part-time), advised by Dieter Fox *May - Aug 2022, Sep 2023 - May 2024*

- Focused on developing learning-based manipulation planning algorithms for efficient generation of robot data in visually diverse and geometrically complex environments. See open source project *kitchen-world*
  - Developed a general algorithm for solving long-horizon mobile manipulation problems combining the commonsense capabilities of ***large pre-trained vision-language models*** (e.g. GPT-4v & Claude 3) with the geometric soundness guarantee of ***task and motion planners***. It generates full trajectories for a mobile bimanual robot in procedurally generated kitchen problems requiring 30 - 50 actions in sequence and interacting with up to 21 movable and articulated objects. It achieves a 50 - 100% success rate across various world initial states and robot embodiments. See: *Project Page*
  - Developed a ***multimodal Transformer-based architecture***, PIGINet, that ***predicts task plan feasibility*** based on the initial state, goal, and candidate plans, fusing image and text embeddings with state features. PIGINet ***reduces runtime by 50 - 80%*** on kitchen problems with articulated and movable obstacles after training on only 300 - 600 problems. It achieves ***zero-shot generalization to unseen object geometry***. See: *Project Page*

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. Integrated the task learning system with controllers and visual grounding modules for a ***Kinova Jaco arm*** using ROS. Demonstrated learning and making simple cuisines, such as breakfast cereal and fruit salad. *Video Demo*

Selected Publications

Yang, Z., Garrett, C., Kaelbling, L., Lozáno-Pérez, T., & Fox, D.. Guiding Long-Horizon Task and Motion Planning with Vision Language Models. *CoRL 2024 LangRob Workshop (Spotlight). Project Page*

Ravan, Y., **Yang, Z.**, Chen, T., Lozáno-Pérez, T., & Kaelbling, L.. Combining Planning and Diffusion for Mobility with Unknown Dynamics. *arXiv:2410.06911. (Submitted to ICRA 2025). Project Page*

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. Project Page*

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. Project Page*

Yang, Z., Curtis A. Lets Handle It: Generalizable Manipulation of Articulated Objects. *ICLR 2022 Workshop on Generalizable Policy Learning in the Physical World (Spotlight).* **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, Java, R, MATLAB.

Technical Skills: *ML*: PyTorch, multi-node multi-GPU training, AWS SageMaker; *Robotics*: ROS, Isaac Gym, Drake, MuJoCo, PyBullet; *Communications*: Video production using Adobe Premiere and AfterEffects.

Non-technical Skills: Kickboxing (AFAA certified group exercise instructor), singing, improv comedy.