

SILONG YONG

silongy@andrew.cmu.edu

+1 (412)-589-5111

[Google Scholar](#)

[Personal Homepage](#)

EDUCATION

Carnegie Mellon University, Robotics Institute

August 2025 - December 2027 (Expected)

Ph.D. in Robotics

Carnegie Mellon University, Robotics Institute

August 2023 - May 2025

Master of Science in Robotics

Overall GPA: 4.0/4.0

Tsinghua University, Department of Automation

August 2019 - July 2023

Bachelor of Engineering in Electrical and Computer Engineering

Overall GPA: 3.8/4.0

SELECTED PUBLICATIONS

1. **Silong Yong**, Anji Liu, Cunxi Dai, Carl Busart, Guanya Shi, Yilun Du, Katia Sycara, Yaqi Xie. Trajectory Planning without Trajectory Data: A Manifold-Guided Approach. NeurIPS 2026.
2. **Silong Yong**, Stephen Sheng, Carl Qi, Xiaojie Wang, Evan Sheehan, Anurag Shivaprasad, Yaqi Xie, Katia Sycara, Yesh Dattatreya. Generalizable Dense Reward for Long-Horizon Robotic Tasks. IROS 2026.
3. **Silong Yong**, Venkata Nagarjun Pudureddyur Manivannan, Bernhard Kerbl, Zifu Wan, Simon Stepputtis, Katia P. Sycara, Yaqi Xie. OMG: Opacity Matters in Material Modeling with Gaussian Splatting. ICLR 2025.
4. **Silong Yong**, Yaqi Xie, Simon Stepputtis, Katia P. Sycara. GL-NeRF: Gauss-Laguerre Quadrature Enables Training-Free NeRF Acceleration. NeurIPS 2024.
5. Jiangyong Huang*, **Silong Yong***, Xiaojian Ma*, Xiongkun Linghu*, Puhao Li, Yan Wang, Qing Li, Song-Chun Zhu, Baoxiong Jia, Siyuan Huang. An Embodied Generalist Agent in 3D World. ICML 2024. (* equal contribution).
6. Xiaojian Ma*, **Silong Yong***, Zilong Zheng, Qing Li, Yitao Liang, Song-Chun Zhu, Siyuan Huang. SQA3D: Situated Question Answering in 3D Scenes. ICLR 2023. (* equal contribution).

RESEARCH EXPERIENCES

Carnegie Mellon University

Research Assistant / PhD Student

August 2023 – May 2025 / August 2025 – Present

Advisor: Prof. Katia Sycara

• Controllable World Action Model

- * Developing controllable and steerable World Action Models (WAMs) through data-centric post-training and test-time adaptation.
- * Studying controllability through the alignment of visual imaginations, language instructions, and robot actions.
- * Improving vision-action alignment for reliable long-horizon action generation, enabling effective utilization of predicted action chunks rather than frequent replanning.
- * Improving language-vision-action alignment for precise, quantitative instruction following, such that actions and predicted visual outcomes faithfully satisfy commands such as “move 5 cm.”

• Agentic Verification Framework for Precise Physical Changes

- * Developing an agentic, tool-augmented framework for verifying precise physical changes in world-model predictions and embodied interactions.
- * Constructing verifiable signals for quantitative spatial correctness, including displacement, alignment, contact, and insertion accuracy.
- * Exploring verification-driven evaluation and learning beyond semantic task success, with the goal of improving physical accuracy and controllability of embodied models.

- **Sequence Modeling without Sequence Data**

- * Developed a manifold-guided approach for using diffusion models to generate trajectories without training on trajectory data.
- * Evaluated the approach on classical planning benchmarks, including maze navigation and robotic planning domains.
- * Extending the framework toward leveraging non-trajectory state data for data-efficient trajectory and policy learning across embodiments and downstream tasks.
- * Studying how non-trajectory data can provide transferable priors for downstream trajectory and policy learning.

- **Neural 3D Rendering and Reconstruction**

- * Developed new mathematical and physics-based formulations for neural rendering and inverse rendering with NeRF and 3D Gaussian Splatting.
- * Designed efficient GPU implementations, including CUDA-based forward/backward operators and a WebGL-based NeRF viewer, to accelerate rendering and visualization.
- * Conducted extensive analytical and empirical evaluations across datasets and baselines. The works were accepted by NeurIPS 2024 and ICLR 2025.

Adobe Research Scientist Intern

May 2026 - August 2026

Mentor: Xiaohang Zhan, Mang Tik Chiu, Haoran You, Ayush Shrivastava, Yilin Wang, Lingzhi Zhang, Zhixin Shu, Yuqian Zhou, Eli Shechtman, Sohrab Amirghodsi

- **Precise Verification for Visual Physical Changes**

- * Designed verifiable reward functions for image editing tasks involving precise physical changes, grounding correctness in measurable visual outcomes rather than human preference alone.
- * Demonstrated that reinforcement learning with verifiable rewards can teach new image-editing skills without paired supervision, enabling learning from automatically checkable feedback.
- * Showed that combining supervised fine-tuning and reinforcement learning improves fine-grained physical accuracy and controllability in image editing, especially for quantitatively specified edits. In submission.

Amazon Robotics Applied Scientist Intern

May 2025 - August 2025

Mentor: Stephen Sheng, Xiaojie Wang, Evan Sheehan, Anurag Shivaprasad, Yesh Dattatreya

- **Foundation Model Learning for Robotics**

- * Studied the problem of using Reinforcement Learning to finetune robotic foundation models on long-horizon tasks.
- * Developed a generalizable reward model for RL-based adaptation of robotic foundation models across long-horizon tasks, combining signals from external foundation models with the robot policy itself.
- * Conducted extensive evaluations across in-distribution and out-of-distribution tasks, demonstrating the effectiveness and generalization of the proposed reward model. Accepted by IROS 2026.

Beijing Institute for General Artificial Intelligence (BIGAI)

Research Intern

August 2021 - July 2023

Director: Prof. Song-Chun Zhu / Advisor: Dr. Xiaojian Ma and Dr. Siyuan Huang

- **Embodied 3D Vision-Language Learning**

- * Developed SQA3D, a large-scale benchmark connecting 3D scene understanding with embodied reasoning, including an interactive data collection pipeline for human-annotated situated question answering.
- * Developed LEO, a generalist multimodal embodied agent that jointly processes natural language, 3D point clouds, and 2D visual observations across diverse 3D vision-language and embodied tasks.
- * Designed multimodal alignment and representation learning methods to unify language, 2D, and 3D signals, and evaluated them on scene understanding, reasoning, and embodied navigation tasks. The works were accepted by ICLR 2023 and ICML 2024.