The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)

Building Compositional Robot Autonomy with Modularity and Abstraction

Yuke Zhu[1]

[1]Department of Computer Science, The University of Texas at Austin

My research investigates two intimately connected areas: Robotics and Embodied Artificial Intelligence (AI). While tremendous progress has been made in designing low-cost and reliable collaborative robot (cobot) hardware in recent years, a grand challenge is programming these general-purpose robots to perceive natural environments, make intelligent decisions, and adapt continually. My ongoing and past research has played a vital role in developing fundamental algorithms and principles of learning and perception for embodied agents, with the goal of accelerating the widespread deployment of adept and reliable robots from research settings to the real world.

Specifically, my work seeks to dovetail the *modular structure of the classical autonomy stack* and the *representational power of deep learning* to build the next generation of robot autonomy. My methodology draws inspiration from the principles of *compositionality* and *abstraction*, which have been widely used in science and engineering disciplines for developing complex systems. I adopt the modular design philosophy of the classical robot autonomy stack. Yet, unlike the staged pipelines in the classical stack, I design functional modules of perception, planning, and control, each modeled as modern neural networks of rich inductive biases. Further, I tightly couple these modules through new forms of *perceptual-motor abstractions*, making them amenable to full-stack optimization. To realize this modular yet highly coupled autonomy stack, I organize my research efforts into the following three research themes.

The first research theme centers around learning actionable visual representations from an agent's situated and skillful interactions in the open world. Robots are embodied and active perceivers. Perception guides actions, and actions, in turn, emit novel stimuli otherwise unavailable. Departing from the convention of building vision models from hand-labeled examples, I develop algorithms that form *perceptual abstractions* in tight coordination with an agent's self-directed actions. These algorithms dramatically reduce the costs of human supervision and improve the efficacy of the learned representations.

The second research theme concerns constructing a rich repertoire of sensorimotor skills. These skills (such as grasping, pouring, and stacking) are temporally extended building blocks for realizing purposeful behaviors. They serve as *motor abstractions* to scaffold long-term task executions. I develop sensorimotor learning algorithms that synergize the agent's self-exploration and human feedback. These algorithms can generate intelligent robotic skills that are infeasible to hand-engineer while retaining high learning efficiency. I further investigate new mechanisms of learning skills in a multi-task, multi-modal setting, which improve the robustness and versatility of the learned skills to synthesize general-purpose robot behaviors.

The third theme is building compositional algorithms that leverage the perceptual-motor abstractions introduced above to solve real-world decision-making tasks. I develop algorithms exploiting compositional task structures for high-level reasoning and planning. A unified theme for this line of research is neuro-symbolic methods, where I aim to harness the complementary strengths of neural networks and symbolic planners. My algorithms have shown great promise to systematically generalize to novel situations where today's state-of-the-art deep learning methods struggle.

While continuing my ongoing research on these three themes, my next research plans include exploring continual learning algorithms for robot autonomy during real-world deployment. To deploy state-of-the-art AI-powered robots with confidence, I aspire to enable them to collaborate with human teammates for safe and reliable task execution. I am developing a novel human-in-the-loop framework that ensures robust manipulation behaviors through human-robot collaboration while enabling the robots to improve their autonomy continually. The crux is a suite of memory-based lifelong learning algorithms that make the robots learn, adapt, and cooperate with humans in long-term deployments. I am also developing a multi-modal user interface for human-robot communication, allowing non-expert users to partner and work with robots in human-centered environments.

In summary, my past research has developed fundamental algorithms, open-source tools, and practical systems for building robust and generalizable robot behaviors. My ongoing and future work will lay the foundations for transformative robotics technologies in practical applications, including manufacturing, healthcare, and home assistance. For more details about my lab's research, please visit our group website: https://rpl.cs.utexas.edu/.
