Learning Structured World Models From and For Physical Interactions

Yunzhu Li

Department of Computer Science, Columbia University
500 W. 120th Street, New York, NY 10027
yunzhu.li@columbia.edu

Humans have a strong intuitive understanding of the physical world. Through observations and interactions with the environment, we build mental models that predict how the world would change if we applied a specific action (i.e., intuitive physics). My research draws on these human insights to develop model-based RL agents that learn from their interactions and build predictive models that generalize widely across a range of objects made with different materials. The core idea behind my research is to introduce novel representations and integrate structural priors into learning systems to model dynamics at different levels of abstraction. I will discuss how such structures can make model-based planning algorithms more effective, helping robots accomplish complex manipulation tasks (e.g., manipulating an object pile, shaping deformable foam into a target configuration, and making a dumpling from dough using various tools).

Specifically, I will first discuss Dynamic Particle Interaction Networks (DPI-Net) (Li et al. 2019), in which we propose learning a particle-based simulator for complex control tasks. Combining learning with particle-based systems offers two major benefits: first, the learned simulator, like other particle-based systems, applies broadly across objects of different materials; second, the particle-based representation introduces a strong inductive bias, as particles of the same type follow the same dynamics. This enables the model to quickly adapt to new environments with unknown dynamics, laying the foundation for robot learning in dynamic scenes with particle-based representations.

Next, I will cover the integration of DPI-Net into a model-based RL framework, enabling perception, modeling, and long-horizon manipulation of elasto-plastic objects with various tools. The resulting system, RoboCook, won the **Best Systems Paper Award** at **CoRL-23** (Shi et al. 2023). We demonstrate that with just 20 minutes of real-world interaction data per tool, a general-purpose robot arm can learn complex long-horizon soft object manipulation tasks, such as making dumplings and alphabet-shaped cookies.

Towards the end of the talk, I will also discuss our recent advances in the fine-grained modeling of a broader range of materials, including both rigid and deformable objects (Zhang et al. 2024) (Figure 1a), as well as our approach to determining the optimal abstraction level for constructing graphs to balance efficiency and effectiveness (Wang et al. 2023) (Figure 1c). These two papers won the **Best Paper Awards** at an **ICRA-24 Workshop** and an **IROS-23 Workshop**, respectively.

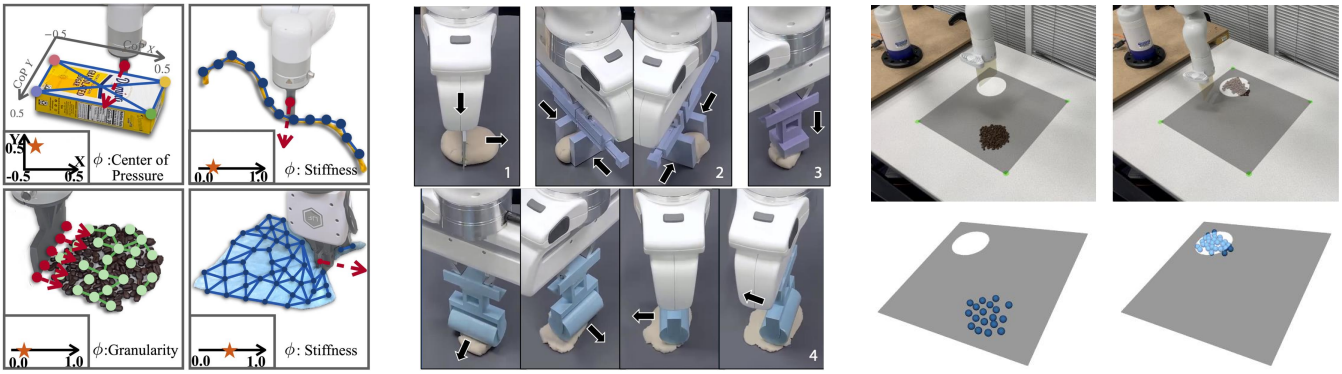
References

Li, Y.; Wu, J.; Tedrake, R.; Tenenbaum, J. B.; and Torralba, A. 2019. Learning Particle Dynamics for Manipulating Rigid Bodies, Deformable Objects, and Fluids. In *International Conference on Learning Representations*.

Shi, H.; Xu, H.; Clarke, S.; Li, Y.; and Wu, J. 2023. RoboCook: Long-Horizon Elasto-Plastic Object Manipulation with Diverse Tools. In *7th Annual Conference on Robot Learning*.

Wang, Y.; Li, Y.; Driggs-Campbell, K.; Fei-Fei, L.; and Wu, J. 2023. Dynamic-Resolution Model Learning for Object Pile Manipulation. In *Robotics: Science and Systems*.

Zhang, K.; Li, B.; Hauser, K.; and Li, Y. 2024. AdaptiGraph: Material-Adaptive Graph-Based Neural Dynamics for Robotic Manipulation. In *Robotics: Science and Systems*.



(a) Graph-based neural dynamics (b) Dough manipulation using diverse tools (c) Granular object manipulation

Figure 1: This talk will focus on learning **structured world models** from and for physical interactions. I will survey work that (a) represents robot-object interactions using **graph-based neural dynamics models**, applicable to both rigid and a diverse range of deformable objects, (b) applies these models within a model-based reinforcement learning framework for **long-horizon manipulation of elastoplastic objects using diverse tools**, and (c) explores the balance between efficiency and effectiveness to determine the **optimal level of abstraction for graph construction**, particularly in applications involving granular object manipulation.