RoboCraft: Learning to See, Simulate, and Shape Elasto-Plastic Objects with Graph Networks

Haochen Shi* Stanford University hshi74@stanford.edu

Huazhe Xu* Stanford University huazhexu@stanford.edu

Zhiao Huang UC San Diego z2huang@eng.ucsd.edu

Yunzhu Li MIT liyunzhu@mit.edu

Jiajun Wu Stanford University jiajunwu@cs.stanford.edu

***Abstract*—Modeling and manipulating elasto-plastic objects are essential capabilities for robots to perform complex industrial and household interaction tasks (e.g., stuffing dumplings, rolling sushi, and making pottery). However, due to the high degree of freedom of elasto-plastic objects, significant challenges exist in virtually every aspect of the robotic manipulation pipeline, e.g., representing the states, modeling the dynamics, and synthesizing the control signals. We propose to tackle these challenges by employing a particle-based representation for elasto-plastic objects in a model-based planning framework. Our system, RoboCraft, only assumes access to raw RGBD visual observations. It transforms the sensing data into particles and learns a particle-based dynamics model using graph neural networks (GNNs) to capture the structure of the underlying system. The learned model can then be coupled with model-predictive control (MPC) algorithms to plan the robot's behavior. We show through experiments that with just 10 minutes of real-world robotic interaction data, our robot can learn a dynamics model that can be used to synthesize control signals to deform elasto-plastic objects into various target shapes, including shapes that the robot has never encountered before. We perform systematic evaluations in both simulation and the real world to demonstrate the robot's manipulation capabilities and ability to generalize to a more complex action space, different tool shapes, and a mixture of motion modes. We also conduct comparisons between RoboCraft and untrained human subjects controlling the gripper to manipulate deformable objects in both simulation and the real world. Our learned model-based planning framework is comparable to and sometimes better than human subjects on the tested tasks.** [1]

I. Introduction

Effective manipulation of deformable objects is an essential skill for robots deployed in real-world industrial and household environments. However, due to deformable objects' high degrees of freedom (DoF) and consequent challenges in state estimation and dynamics modeling, manipulating deformable objects requires significant innovations beyond the typical robotic paradigm that focuses only on rigid objects. Recent advances show promising results in manipulating clothes [10, 8, 17, 2, 19, 3] and ropes [18, 16], yet the manipulation of objects with high plasticity, such as dough or plasticine, poses a unique set of challenges and is currently underexplored [1, 9], despite the ubiquity of such objects in household and industrial settings. In this paper, we investigate how to empower robots to model and manipulate elasto-plastic objects based on raw RGBD visual observations.

[1]Project page: http://hxu.rocks/robocraft/.

*Denotes equal contribution, random order.

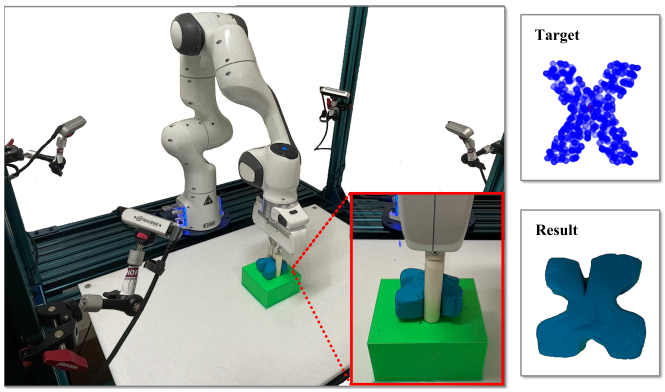
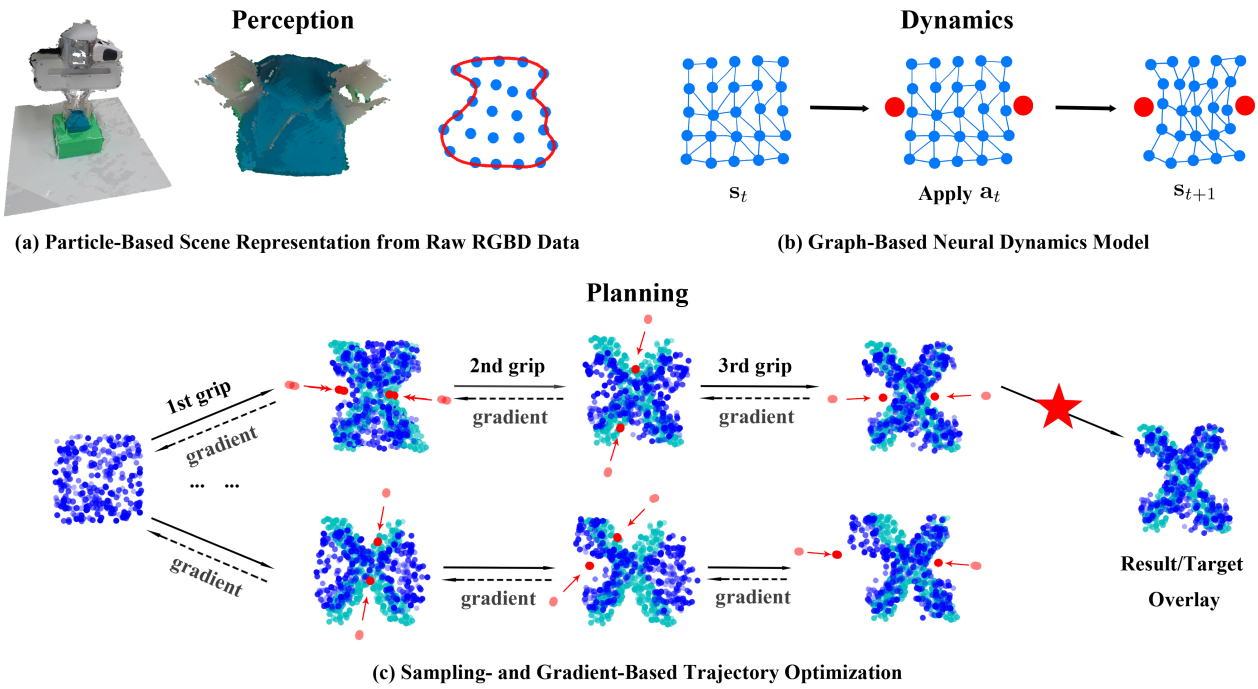


Fig. 1: **RoboCraft**. The robot uses a parallel 2-finger gripper to shape an 'X' conditioned on the target shape at the top right corner. The result is shown at the bottom right corner.

The primary challenges of manipulating deformable objects stem from their high DoFs, partial observability, and complex non-linear local interactions. Learning dynamics models directly from high-dimensional sensory data offers a promising data-driven avenue for us to perform effective planning. For example, model-based reinforcement learning (RL) algorithms have achieved great success in various planning and control tasks [14, 12, 7]. However, when faced with elasto-plastic objects, these prior methods may fail due to a lack of explicit exploitation of the objects' structure. Another thread of works represents deformable objects using particles and employs graph neural networks (GNNs) to model their dynamics [11, 5, 6, 13, 15]. They have shown great generalization results, demonstrating the benefits of explicit structured modeling. However, most of them require full-state information and a particle-based simulator to provide particle-to-particle correspondence between frames. Such strong supervision is difficult to obtain from raw sensory data, limiting their use in real-world applications. Hence, the natural question to ask here is: would it be possible to model the dynamics and manipulate elasto-plastic objects in the real world solely based on RGBD visual observations, without needing particle-to-particle temporal correspondence?

To tackle this problem, we propose RoboCraft, a model-based planning framework that represents elasto-plastic objects using particles, but employs distribution-based loss functions



(a) Particle-Based Scene Representation from Raw RGBD Data

(b) Graph-Based Neural Dynamics Model

(c) Sampling- and Gradient-Based Trajectory Optimization

Fig. 2: **Overview of RoboCraft.** (a) The perception module obtains the particle representation from RGBD cameras. The algorithm first crops out the object point cloud, then samples particles to represent the object. (b) The dynamics model predicts the object's deformation based on graph neural networks (GNN). (c) After obtaining the learned dynamics model, we apply a combination of sampling- and gradient-based trajectory optimization techniques to solve the model-predictive planning problem.

and makes novel improvements over recently-developed GNNs to model the objects' dynamics. The learned dynamics model is then coupled with gradient-based trajectory optimization techniques to plan the robot's behaviors. The proposed approach closes the perception and control loops, which allows accurate modeling and manipulation of the elasto-plastic objects in both simulated and real-world settings. Specifically, our framework consists of (1) a perception module that constructs the particle representation of the object by sampling from the reconstructed object mesh, (2) a dynamics model that models the particle interactions using GNNs, and (3) a planning module that uses model-predictive control (MPC) and solves the trajectory optimization problem using gradients from the learned model. Unlike prior learning-based particle dynamics works which assume temporal correspondence [11, 5, 6, 13, 15], we train the dynamics model directly from raw visual data using loss functions that measure the distance between predicted and observed particle distributions.

II. Method

A. Problem Statement

The objective of this work is to use a parallel 2-finger robot gripper to shape an elasto-plastic object to match a target shape $\mathbf{g}$. Specifically, we focus on using a sequence of pinching actions $\mathbf{a}_{0,...,T-1} \in \mathcal{A}$, given an observation of the initial state $\mathbf{s}_0$ of the plasticine. At time step t, the robot applies action $\mathbf{a}_t \in \mathcal{A}$ upon the plasticine, and the state of the plasticine transitions from $\mathbf{s}_t$ to $\mathbf{s}_{t+1}$ in response.

To predict the complex dynamics of the deformable plasticine, we propose to use a graph neural network (GNN) Φ to learn the transition function $\Phi : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$. This dynamics model takes as input environment observations $\mathbf{s}_{t-h,...,t} \in \mathcal{S}$ and actions $\mathbf{a}_{t-h,...,t} \in \mathcal{A}$ and predicts a future observation $\hat{\mathbf{s}}_{t+1}$, where h is the length of history and t is the current time step. With the learned dynamics model in hand, we can naturally formulate the manipulation task as a model-predictive control (MPC) problem. The cost function $\mathcal{J}$ of the MPC problem measures the distance between the state of the plasticine at the last time step T and the target shape $\mathbf{g}$. And a sequence of actions of length T can be selected by minimizing the cost function:

$$(\mathbf{a}_0, ..., \mathbf{a}_{T-1}) = \underset{\mathbf{a}_{0,...,T-1} \in \mathcal{A}}{\arg\min} \mathcal{J}(\Phi(\mathbf{s}_0, (\mathbf{a}_0, ..., \mathbf{a}_{T-1})), \mathbf{g}) \quad (1)$$

Figure 2 shows the overall framework of RoboCraft.

III. Experimental Results

A. Results in Simulation

We use a physics simulator based on Material Point Method (MPM) from previous work [4]. In Figure 3, we visualize the procedure of manipulating the object towards the target shape using a gradient-based method. We find that the agent can handle various challenges such as small grooves in the letter 'E' and asymmetry in the letter 'Z'. This demonstrates that the method can leverage the GNN-based dynamics model for effective manipulation under the MPC framework.

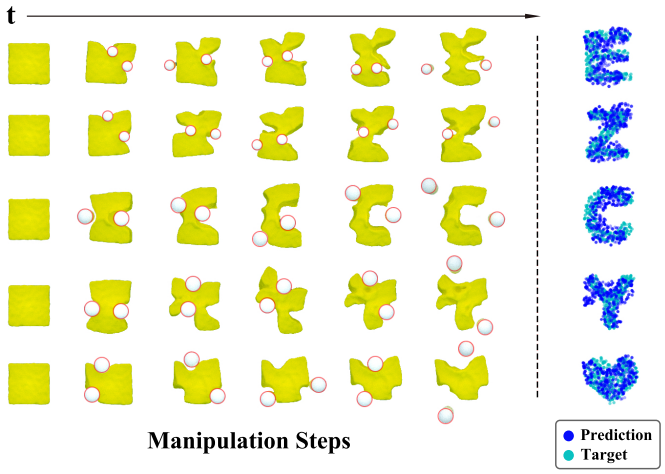


Fig. 3: **Manipulation result in the simulation.** On the left are the manipulation steps. On the right are the result and its overlay with the target point cloud. The cyan point cloud is the target, blue the result.

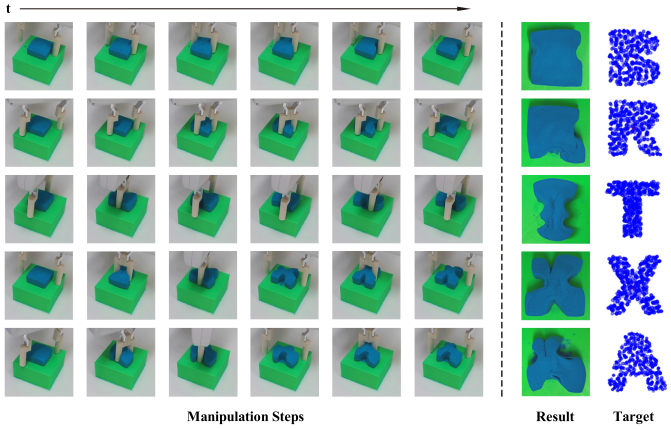


Fig. 4: **Manipulation results on a real robot.** On the left are the shaping steps. On the right are the results and corresponding target point clouds. We want to emphasize that our model is learned purely from offline data collected via random interactions (10 minutes); thus, the target shapes have never been seen during training. Yet our pipeline can still achieve these targets with reasonable accuracy.

B. Learning Real-World Manipulation of Deformable Objects

The proposed method is able to manipulate the plasticine to shapes that are unseen in the training data. Example trajectories of the robot manipulating the plasticine are shown in Figure 4. The method successfully identifies the asymmetry in the target shape 'B' by putting the finger closer to one side of the plasticine at the beginning of the grip. For more complex shapes such as the letter 'A', the method also seems to creatively discover a solution that roughly achieves the target shape. These results illustrate that, although the task is very challenging, our method is able to perform well with a small amount of training data.

TABLE I: Results of human subjects and the robot in the simulator. Numbers are averaged over all the tested shapes.

Methods	CD↓	EMD↓
Human Subjects	0.0655 ± 0.025	0.0661 ± 0.023
RoboCraft (ours)	**0.0359** ± 0.007	**0.0340** ± 0.005

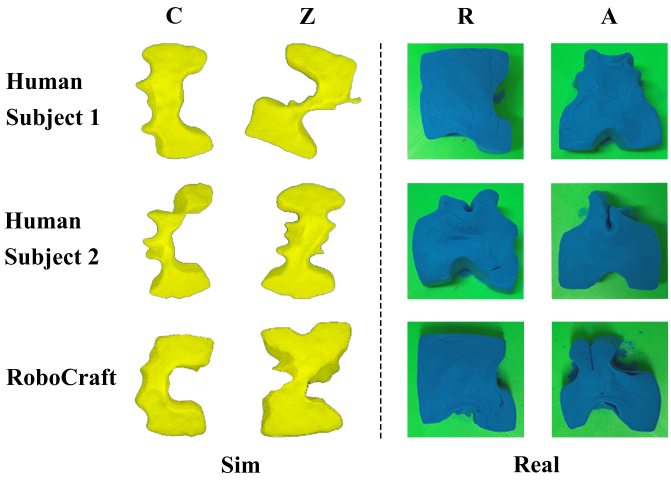


Fig. 5: **Shaping results by amateur humans and RoboCraft.** Results in the first two rows are from human subjects. The results in the third row are from the robot. The left two columns are results in the simulation. The right two columns are the results in the real world.

C. Results on Comparing with Human Performance

We invited four amateur humans to perform the same task with the robot gripper in both simulation and real-world settings. While humans are not trained to manipulate plasticine, they usually have strong intuitive understandings of the dynamics of plasticine. Each user was asked to shape three pieces of plasticine in each domain. For the simulation experiments, we provided a successful manipulation video as an example for the users and allowed two trials for each shape since the dynamics in the simulation were also new to each human. In the real world, we allowed the users to play with the plasticine for one minute before resetting it to the starting shape for the experiment. In Table I, we show the average distance over all the target shapes from four users, in comparison with that of our agent. Empirical evidence suggests that the proposed tasks are challenging for both manipulation algorithms and untrained human subjects alike. We also find that RoboCraft is comparable to or stronger than amateur humans on the tested tasks. One observation is that RoboCraft outperforms humans in the distance metrics. However, the visualized human results are recognizable even when the distances are high, suggesting that better evaluation metrics are desired. In Figure 5, we show the outcome from both human users and the robot for comparison.

References

[1] Andrea Cherubini, Valerio Ortenzi, Akansel Cosgun, Robert Lee, and Peter Corke. Model-free vision-based shaping of deformable plastic materials. *The International Journal of Robotics Research*, 39(14):1739–1759, 2020.

[2] Huy Ha and Shuran Song. Flingbot: The unreasonable effectiveness of dynamic manipulation for cloth unfolding. In *Conference on Robot Learning (CoRL)*, pages 24–33. PMLR, 2022.

[3] Ryan Hoque, Daniel Seita, Ashwin Balakrishna, Aditya Ganapathi, Ajay Kumar Tanwani, Nawid Jamali, Katsu Yamane, Soshi Iba, and Ken Goldberg. Visuospatial foresight for physical sequential fabric manipulation. *Autonomous Robots*, 46(1):175–199, 2022.

[4] Zhiao Huang, Yuanming Hu, Tao Du, Siyuan Zhou, Hao Su, Joshua B Tenenbaum, and Chuang Gan. Plasticinelab: A soft-body manipulation benchmark with differentiable physics. In *International Conference on Learning Representations (ICLR)*, 2020.

[5] Yunzhu Li, Jiajun Wu, Russ Tedrake, Joshua B Tenenbaum, and Antonio Torralba. Learning particle dynamics for manipulating rigid bodies, deformable objects, and fluids. In *International Conference on Learning Representations (ICLR)*, 2018.

[6] Yunzhu Li, Toru Lin, Kexin Yi, Daniel Bear, Daniel Yamins, Jiajun Wu, Joshua Tenenbaum, and Antonio Torralba. Visual grounding of learned physical models. In *International Conference on Machine Learning (ICML)*, pages 5927–5936. PMLR, 2020.

[7] Lucas Manuelli, Yunzhu Li, Pete Florence, and Russ Tedrake. Keypoints into the future: Self-supervised correspondence in model-based reinforcement learning. In *Conference on Robot Learning (CoRL)*, pages 693–710. PMLR, 2021.

[8] Jan Matas, Stephen James, and Andrew J Davison. Sim-to-real reinforcement learning for deformable object manipulation. In *Conference on Robot Learning (CoRL)*, pages 734–743. PMLR, 2018.

[9] Carolyn Matl and Ruzena Bajcsy. Deformable elasto-plastic object shaping using an elastic hand and model-based reinforcement learning. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3955–3962. IEEE, 2021.

[10] Stephen Miller, Jur Van Den Berg, Mario Fritz, Trevor Darrell, Ken Goldberg, and Pieter Abbeel. A geometric approach to robotic laundry folding. *The International Journal of Robotics Research (IJRR)*, 31(2):249–267, 2012.

[11] Damian Mrowca, Chengxu Zhuang, Elias Wang, Nick Haber, Li F Fei-Fei, Josh Tenenbaum, and Daniel L Yamins. Flexible neural representation for physics prediction. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018.

[12] Anusha Nagabandi, Kurt Konolige, Sergey Levine, and Vikash Kumar. Deep dynamics models for learning dexterous manipulation. In *Conference on Robot Learning (CoRL)*, pages 1101–1112. PMLR, 2020.

[13] Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *International Conference on Machine Learning (ICML)*, pages 8459–8468. PMLR, 2020.

[14] Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.

[15] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2(2):021001, 2020.

[16] Priya Sundaresan, Jennifer Grannen, Brijen Thananjeyan, Ashwin Balakrishna, Michael Laskey, Kevin Stone, Joseph E Gonzalez, and Ken Goldberg. Learning rope manipulation policies using dense object descriptors trained on synthetic depth data. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9411–9418. IEEE, 2020.

[17] Yilin Wu, Wilson Yan, Thanard Kurutach, Lerrel Pinto, and Pieter Abbeel. Learning to Manipulate Deformable Objects without Demonstrations. In *Robotics: Science and Systems (RSS)*, July 2020. doi: 10.15607/RSS.2020.XVI.065.

[18] Wilson Yan, Ashwin Vangipuram, Pieter Abbeel, and Lerrel Pinto. Learning predictive representations for deformable objects using contrastive estimation. In *Conference on Robot Learning (CoRL)*, pages 564–574. PMLR, 2021.

[19] Hang Yin, Anastasia Varava, and Danica Kragic. Modeling, learning, perception, and control methods for deformable object manipulation. *Science Robotics*, 6(54), 2021.