Efficient End-effector Co-Design by Demonstration for Deformable Fragile Object Manipulation

Kei Ikemura[1], Yifei Dong[1], David Blanco-Mulero[2], Alberta Longhini[1], Li Chen[1], Florian T. Pokorny[1]

***Abstract*—Manipulating deformable and fragile objects remains a critical challenge in robotics due to their complex dynamics and susceptibility to damage. Existing approaches typically address either hardware design or control policies in isolation. In this work, we present the first co-design framework that simultaneously optimizes both end-effector design and control for deformable and fragile object manipulation. Our key insight is incorporating human priors through demonstrations, guiding the search for high-performance designs and control policies while maintaining sample efficiency. Our approach integrates a compact design space using cage-based deformation, a differentiable inverse design process, and a reinforcement learning algorithm (RLPD) to efficiently explore the joint design-control space. We evaluate our approach in the challenging task of grasping silk tofu. Preliminary experiment results demonstrate that our co-designed end-effector significantly reduces damage compared to the original parallel-jaw gripper. This work highlights the potential of co-adaptive design and control for deformable fragile object manipulation tasks.**

I. Introduction

Deformable fragile object manipulation [8, 28] finds its application in a variety of fields, including food handling, surgical operations, and caregiving. It has been a challenging problem due to several factors: (1) deformable object manipulation is inherently difficult due to high dimensional state space and complex dynamics, (2) variation of physical properties such as stiffness, fracture stress and surface friction pose difficulties to generalization; (3) precise and dynamic response is required during the manipulation process to maintain the integrity of the object.

Existing approaches to deformable object manipulation typically follow one of two lines: hardware-centric or control-centric. The former involves the mechanical design of rigid [1], soft [8, 27], or hybrid grippers [10]. For example, Wang et al. [27] propose a dual-mode soft gripper made of rubber material that can grasp different types of objects via prehensile or suction-based interaction. These tailored grippers are the result of expert human designers and have superior performance over simple parallel jaw grippers. However, many of the designs are not optimized in a data-driven manner against empirical manipulation results, and the control policy usually remains simple. On the other hand, control-oriented methods rely on predefined standard end-effectors (such as parallel jaw grippers). These approaches design advanced control schemes by leveraging additional sensing to enable gentle object handling [5, 11, 16, 28, 30]. However, these methods usually assume fixed hardware and thus cannot exploit the full potential of design-control that co-adaptation offers.

Despite the success of hardware- or control-oriented methods, joint hardware and control optimization remains underexplored for deformable fragile object manipulation. This disconnect motivates our interest in co-design [24] — the simultaneous optimization of a robot's physical design and its control policy — as a promising solution. Co-design has been applied to several fields in robotics, including locomotion [3, 31], manipulation [7, 25, 29], and modular robots [6, 23]. A widely adopted co-design framework is bi-level optimization: the lower level module typically is an Reinforcement Learning (RL) policy that learns the optimal control given a pre-defined design, while the upper level, e.g. done via Bayesian Optimization (BO) [4, 7, 20, 22], systematically selects different designs for evaluation. Recent works also formulate the co-design process as a two-stage Markov Decision Process (MDP) consisting of a design phase and a control phase [12, 17, 31], resulting in higher sample efficiency. By leveraging co-design methods, optimal design and control can be acquired through a data-driven manner. This has the potential to outperform both design-oriented and control-oriented methods for deformable fragile object manipulation.

To this end, this work takes an initial step towards leveraging co-design approaches for deformable fragile object manipulation. Here, we develop a framework that integrates (i) a compact yet expressive design space via cage-based deformation [21], (ii) an inverse design method for incorporating expert-provided shape demonstrations to improve sample efficiency, and (iii) a reinforcement learning algorithm (RLPD [2]) capable of leveraging smobjectall amount of expert demonstrations to efficiently explore the joint design-control space. We evaluate the proposed method in both simulated and real-world experiments on a silk tofu pick-up task. The results demonstrate that our co-designed gripper and control policy can successfully manipulate objects while avoiding damage. object

To summarize, our contributions are as follows:

1) This is the first work, to the best of our knowledge, that introduces a co-design framework tailored to deformable fragile object manipulation.
2) We incorporate few human demonstrations into the co-design policy learning process, improving performance and sample efficiency.

The authors are with: [1]the division of Robotics, Perception and Learning, KTH Royal Institute of Technology, Stockholm, Sweden; [2]Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain. Funded by the European Commission under the Horizon Europe Framework Programme project SoftEnable, grant number 101070600. Contact: {ikemura, yifeid}@kth.se.

Experiment videos and other supplementary materials are available at: https://sites.google.com/view/efficient-end-effector-and-con/home

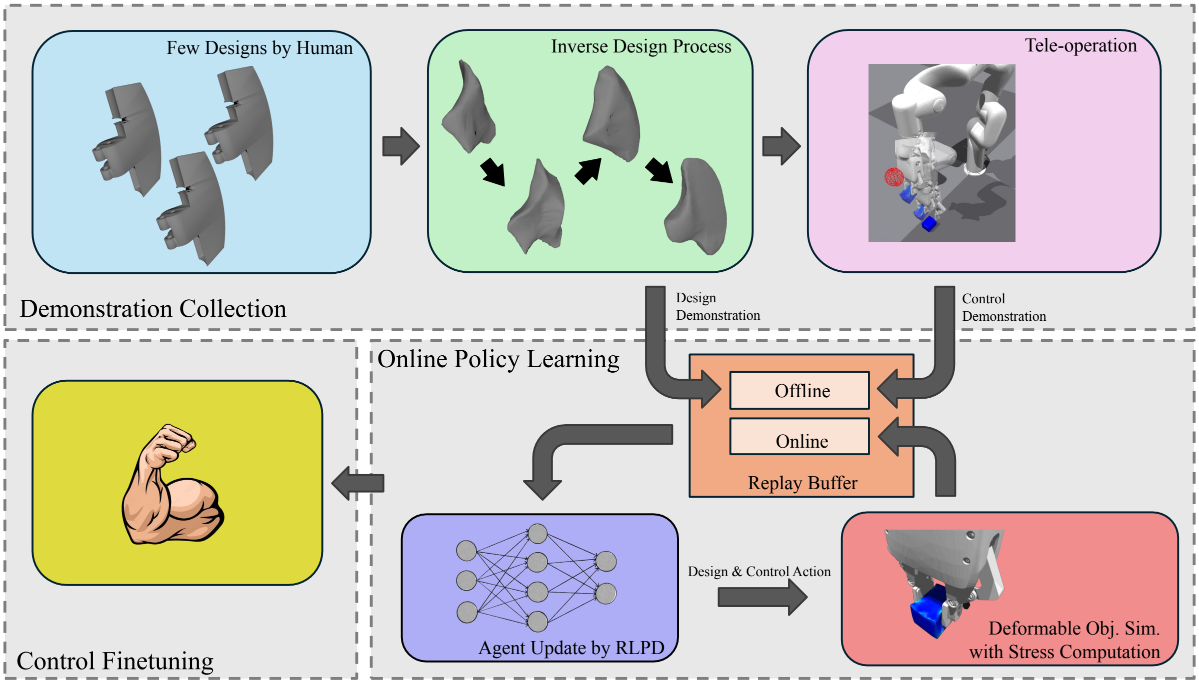


Fig. 1: Overview of the proposed co-design pipeline. The co-design process starts by gathering human demonstrations of hardware designs, an inverse design process, and control demonstrations of the resulting inverse designs. This is followed by online sample-efficient co-design policy learning with RLPD, which leverages both offline demonstrations and online exploration experience. Lastly, we extract the design chosen by the converged co-design policy and perform control finetuning.

3) Preliminary experiments on a silk tofu grasping task in simulation and real world validate the effectiveness and efficiency of our proposed pipeline.

II. Problem Formulation

We tackle the problem of deformable object manipulation and formulate it as a co-design process. The process optimizes the manipulation policy as well as the physical design of a robot end-effector. Following [12, 17], we formulate it as a two-stage MDP, comprising a design stage and a control stage.

We aim to optimize end-effector design and control policy, enabling manipulation of deformable fragile objects without inducing damage. For this purpose, we formulate it as a stress-minimization manipulation problem. That is, apart from maximizing the manipulation task success rate, we also minimize the stress applied to the target object during handling. Previous works have formulated this as a manipulation problem where the object deformation is minimized by measuring its change in shape [5], the applied force [16], or inferring the applied force via tactile sensors [11]. By contrast, we decide to measure the stress applied to the object. This correlates directly to any potential damage suffered by the object and enables reducing its deformation while, e.g., grasping.

Here, we assume that the stress applied to the object cannot be measured in the real world. Therefore, the co-design process is a two-stage Partially Observable MDP (POMDP). The co-design process starts with a design policy $\pi_D(\mathbf{a}_D|\mathbf{o}_D)$, where $\mathbf{a}_D \in A_D$ and $\mathbf{o}_D \in O_D$. The design $\mathbf{a}_D$ remains constant during the control phase, with a control policy $\pi_C(\mathbf{a}_C|\mathbf{o}_C)$, where $\mathbf{a}_C \in A_C$ and $\mathbf{o}_C \in O_C$. To handle the different observation spaces, we define a common observation space that includes both co-design stages. Additionally, we augment the observation with a binary flag to identify the MDP phase. That is, $\mathbf{o} = [\mathbf{o}_D, \mathbf{o}_C, f] \in O$, where the flag f is zero for the design stage and one for the control stage.

The design action $\mathbf{a}_D$ defines the geometry of the fingers of a standard parallel jaw gripper. The parameterization of the action space is described further in Section III-A. The control action space is defined as $\mathbf{a}_C = [\delta\mathbf{p}_{\text{ee}}, \delta\mathbf{q}_{\text{ee}}, \delta w_{\text{ee}}] \in \mathbb{R}^7$, where $\delta\mathbf{p}_{\text{ee}} \in \mathbb{R}^3$ is the displacement of the end-effector in the Cartesian frame, $\delta\mathbf{q}_{\text{ee}} \in \mathbb{R}^3$ is the delta rotation vector, and $\delta w_{\text{ee}} \in \mathbb{R}$ is the change in the gripper closing width. In practice, we concatenate the design action space and the control action space $\mathbf{a} = [\mathbf{a}_D, \mathbf{a}_C]$. During transition, only one part of the concatenated action vector is active depending on the phase, while the other is ignored. The observation for both phases is defined as $\mathbf{o} = [\mathbf{p}_{\text{ee}}, \mathbf{p}_{\text{obj}}, w_{\text{ee}}, f, d]$, where $\mathbf{p}_{\text{ee}}$ and $\mathbf{p}_{\text{obj}}$ are the positions of the end-effector and the target object in Cartesian space, respectively, w_{ee} is the current finger width, and $d \in \mathbb{R}$ is a scalar that informs about the current design used to manipulate the object.

III. Co-Design by Demonstration

Our co-design process is composed of three modules depicted in Figure 1. Firstly, to initiate the co-design process, a small set of initial designs from a human expert is provided based on domain expertise. These potentially sub-optimal designs are utilized in an inverse design process that fits the

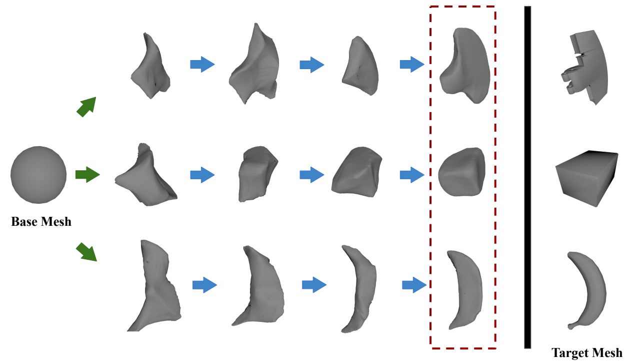


Fig. 2: We propose an inverse design process that leverages the differentiability of cage-based deformation. Given a target design provided by an expert, we iteratively find the corresponding design parameters that best match the target mesh, which is used as the design demonstration.

Algorithm 1 Inverse Design with CBD.

Require: Target object mesh (V_t, F_t), cage mesh (H, F_c), base mesh $(\hat{V}_b, F_b)$, maximum number of trials N, epoch E

```
for i ← 1 : N do
    Initialize H_0 ← H
    Sample an initial design parameter δH_0 randomly
    for j ← 1 : E do
        H_j ← H_{j−1} + δH_{j−1}
        Deform the base mesh using new cage H_j
        l_j ← Chamfer((V̂_b, F_b), (V_t, F_t))
        Compute gradient w.r.t. design parameters ∇_{δH_j} l_j
        δH_{j+1} ← δH_j − λ∇_{δH_j} l_j
        Save δH* ← δH_j if l_j is minimum so far.
    end for
end for
```

corresponding action design parameters, $\mathbf{a}_D$, which serve as the initial design demonstration. With the fitted designs, a human operator controls the robot in simulation to generate a single control demonstration per design. Both the design and control demonstrations are stored in an offline replay buffer. Secondly, we train a co-design agent online using the offline data with the Reinforcement Learning using Prior Data (RLPD) [2] algorithm. A physics simulation provides information about the stress induced on the object, which is used to reward the agent during training. Finally, after the co-design policy converges, we finetune the control for the resulting design to further improve the performance. The rest of this section provides the details for each sub-module.

A. End-Effector Design: Cage-Based Deformation

In this work, we adopt a design action space that offers a favorable trade-off between dimensionality, diversity, and smoothness: cage-based deformation (CBD). Although Xu et al. [29] proposed the use of CBD for co-design, the deformations to the base mesh explored in that context were relatively modest—typically involving only slight alterations, such as elongation. In contrast, we demonstrate that CBD is capable of producing significantly more dramatic shape variations, as showcased in Figure 2. This flexibility is particularly valuable for the manipulation of deformable objects while inducing minimal stress. More specifically, by enabling designs to conform more closely to the surface curvature of the target object, a greater contact area is provided and localized stress concentrations are minimized.

We parameterize the design action space as the displacement vector applied to the cage handle points. Given a cage with handles at rest position $H = \{h_1, h_2, ..., h_n\}, h_i \in \mathbb{R}^3$, we define a set of displacement vectors $\delta H = \{\delta h_i\}$ for each handle such that the deformation result is determined by a new cage with handles $\hat{h}_i = h_i + \delta h_i$. As a result, our design space consists of $3N$ parameters, where N is the number of cage handles. To address potential issues such as self-intersecting mesh faces resulting from aggressive deformations in CBD, we apply mesh correction and smoothing as post-deformation steps to ensure the physical validity of the design. Specifically, we sample a point cloud from the surface of the initial deformed mesh and use it to reconstruct a new, valid mesh. In this work, we use a simple unit sphere as the base mesh.

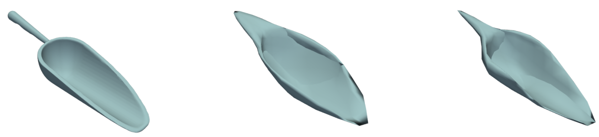
B. Learning from Demonstration

To ensure sample efficiency, we incorporate demonstrations, including both design demonstrations and associated control demonstrations, into the learning process. We choose RLPD, a Soft-Actor Critic (SAC) [13] variant, as our learning algorithm. RLPD incorporates offline data during online training, leading to higher sample efficiency and performance compared to SAC.

To provide design demonstrations, we manually design M end-effectors using CAD software, and subsequently infer the corresponding design parameters, δH, that would result in an end-effector shape as close to the manual design as possible. Since the cage-based deformation process is differentiable, we employ gradient descent based on the Adam optimizer. We measure shape similarity using the Chamfer distance, Chamfer$((\hat{V}_b, F_b), (V_t, F_t))$, where ($\hat{V}_b$, F_b) and (V_t, F_t) are the vertices and faces of the current deformed mesh and the target mesh, respectively. Algorithm 1 illustrates the inverse design process. Figure 2 shows examples of target designs as well as the deformation result of the inferred design parameters. Note that when the number of handles is increased, the inverse design produces deformations that are closer to the target shape, as depicted in Figure 3.

C. Stress Measurement

To simulate the manipulation of deformable fragile objects we use NVIDIA Isaac Gym [19]. The deformable objects are modeled as tetrahedral meshes and simulated through a 3D co-rotational FEM framework, with the resulting equations solved using a GPU-accelerated Newton method [18]. The simulation provides ready-to-use Von Misses stress tensors for the tetrahedral. Prior studies validated Isaac Gym's accuracy in deformation modeling as well as stress simulation [14, 15].



(a) (b) (c)

Fig. 3: An example of inverse design results with different numbers of cage handles. (a): the target design made by a human, (b): inverse design with a 12-handle cage, (c): inverse design with a 30-handle cage. For all cage-based deformation results, we use a simple unit sphere as the base mesh.

Since material fracture is dictated by the maximum stress applied, we focus our reward on the top 10 percent of the stress value across the entire object tetrahedral. In addition, we wish to penalize high stress values more aggressively, thus we employ a quadratic function over the stress values. As a result, we formulate our stress penalty as follows:

$$r_{\text{stress}} = -\frac{(\alpha\sigma_{\text{top10}} + (1-\alpha)\sigma_{\text{mean})})^2}{\beta}, \tag{1}$$

where $\alpha \in [0, 1]$ balances the penalty from peak stress σ_{top10} as well as average stress σ_{mean}. We additionally include a division term β to indicate aggressive penalization after stress values over a stress value near β.

D. Finetuning

After the policy converges to a suitable design-control pair, we further finetune the control of the end-effector optimized for the design. This is achieved by fixing the design and trimming the design stage in the dual POMDP. Specifically, we first generate some "pseudo control demonstrations" from the previous policy to fill the offline replay buffer. Then, we train a control policy given the design using RLPD online, with the offline data. We multiply the previously used stress penalty by 5 times for this second stage of training, to further push the performance in terms of stress-minimization. Note that a more aggressive stress term was infeasible in the previous co-design process, as it would compromise the design exploration efficiency and task completion rate.

IV. Experiment

We evaluate the effectiveness of our method in a silk tofu pick-up task both in simulation and real-world experiments. Silk tofu has low fracture stress, thus requiring minimized stress during grasping.

In our experiments, we analyze and answer the following questions:

- Can the co-design framework successfully manipulate deformable fragile objects while applying minimal stress?
- Is it more sample efficient than the baselines?
- Can the co-design policies learned in simulation by the proposed framework be transferred to a real robot?

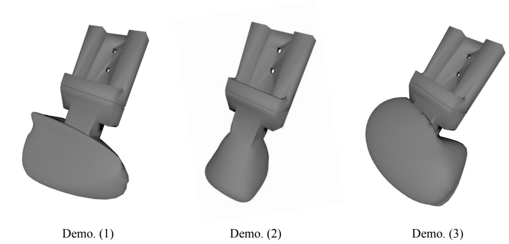


Fig. 4: The three human demonstrations of design utilized for the silk tofu pick-up task.

A. Experiment Setup and Evaluation Metrics

Task: We define a silk tofu block of size 30mm x 50mm x 40mm (width, length, and height, respectively). At the start of a new training episode, we apply domain randomization on the tofu's deformation properties, i.e., the Young's modulus (from 8 kPa to 10 kPa) and the Poisson ratio (from 0.3 to 0.325) [26]. The value range is selected based on the material properties of common silk tofu. In addition, we utilize a Ufactory Xarm7 for the manipulation.

Metrics: For quantitative evaluation of the task outcome, we adopt three metrics:

(1) *Success rate:* We measure the success rate of the policy as the average of the binary success flag over multiple evaluation episodes.

(2) *Top 10% stress mean:* Since the fracture of a deformable fragile object is dictated by the highest local stress values exerted, we define a metric focusing on the mean of the top 10% stress value across the entire object tetrahedrons.

(3) *Median stress:* This metric computes the median value over the stress values across the object tetrahedrons. It focuses on the "average" level of the stress, insensitive to outliers.

Note that the latter two stress metrics are taken at each time step in an episode, and the final values are computed as an average over the interaction period (i.e. excludes times when the robot is not in contact with the object).

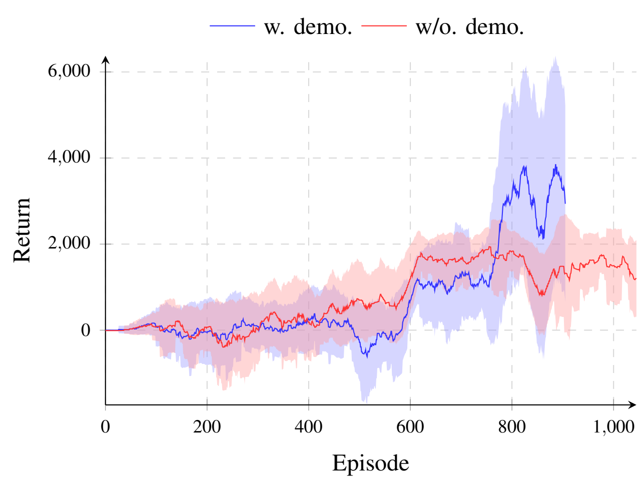


Fig. 5: The episodic return of the proposed framework with expert demonstrations (blue) and without demonstrations (red).

Baselines: To evaluate the effectiveness and efficiency of the proposed framework, we compare it with the following baselines:

(1) *Original Gripper + RLPD*: Given the original parallel jaw gripper of Ufactory XArm7, we train a control policy to pick up the tofu. Control demonstration is provided.

(2) *BO + RLPD*: We follow the common bi-level optimization paradigm for co-design, where the outer loop continues to sample candidate designs, and the inner loop (with RLPD) learns a control policy given the design. Here, we use Bayesian Optimization in the outer loop. We terminate the algorithm after evaluating 12 candidate designs, and take the best among the 12.

(3) *Ours w/o demonstration*: An ablation of our framework where design and control demonstrations are not provided, i.e., a pure online version of RLPD.

(4) *Ours w/o control finetuning*: An ablation study without the control finetuning.

Note that for BO + RLPD, we choose 12 designs due to practical limitations: each design's control policy requires about 4 hours of training, thus 2 days (4×12=48) for the training of a single bi-level method.

Implementation details: We use a N=7 handle cage, totaling 21 design parameters. The left and right finger designs are defined as a mirrored pair. We provide M=3 distinct episodes of demonstration with different expert designs, as shown in Figure 4. This results in a demonstration dataset with only 600 transitions. For stress penalty, we use α=0.8 and β=2000.

Method	Success Rate	Top 10% Stress Mean (Pa)	Median Stress (Pa)
RLPD w. original gripper	**1.0**	3440.04 (±216.04)	743.05 (±120.05)
BO + RLPD (pure online)	**1.0**	8634.69 (±1115.44)	1197.04 (±107.72)
Ours, w/o. demo.	0.0	N/A	N/A
Ours, w/o. control finetune	**1.0**	3298.41 (±491.62)	609.50 (±96.95)
Ours	**1.0**	**2542.21 (±168.01)**	**407.78 (±29.67)**

TABLE I: Quantitative results on tofu pick-up task in simulation, evaluation is taken as the mean over 10 episodes. The value in bracket (±) indicates the standard deviation over 10 evaluations.

B. Framework Effectiveness and Efficiency

Effectiveness: As shown in Table I, compared to the original parallel jaw gripper, our co-design framework produces a design-control pair that reduces the exerted stress during manipulation, while maintaining a high task success rate. This also validates the necessity of co-design for deformable fragile object manipulation. Compared to our framework, *BO + RLPD* achieves lower converged performance. A potential reason for the failure of the bi-level approach is the relatively high dimension of the design space (21-dimensional in this case), where 12 samples could not provide a good coverage for BO. However, sampling more designs significantly increases the computational cost of BO due to the computational complexity of the Gaussian Process. In addition, without expert guidance or priors, it is difficult to explore meaningful designs in such a high-dimensional space.

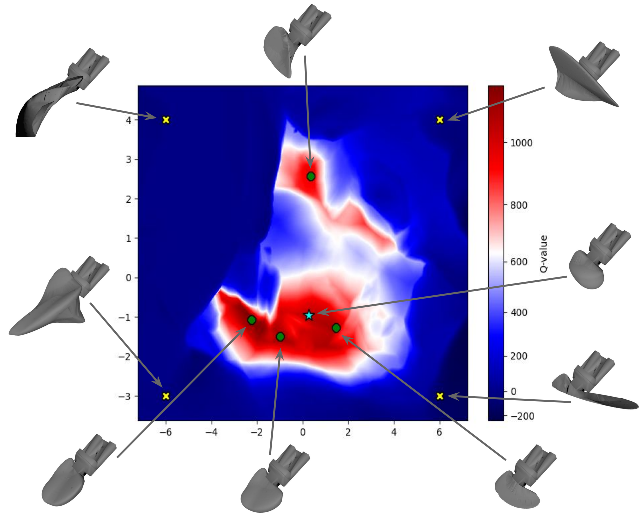


Fig. 6: Visualization of the converged Q-function values. We show the first two principle components of the 21-dimensional CBD design space obtained using all the explored designs during RL training. The optimal design selected by the final policy is marked by a cyan star. The designs with low Q-values are shown by yellow crosses, while those with high values are marked by green circles. From the corresponding mesh visualization, we observe a clear distinction in the learned design policy Q-function.

Additionally, we observe that without demonstrations, the policy fails to converge during training and the tofu cannot be picked up. As shown in Figure 5, the return stops increasing after 600 episodes. This shows the difficulties of learning a co-design policy for deformable fragile object manipulation: (1) the policy has to explore the joint design-control space, and (2) successful grasp and stress minimization are conflicting objectives. This highlights the need for human prior knowledge, which can significantly improve convergence even with a small number of demonstrations. Furthermore, we validate

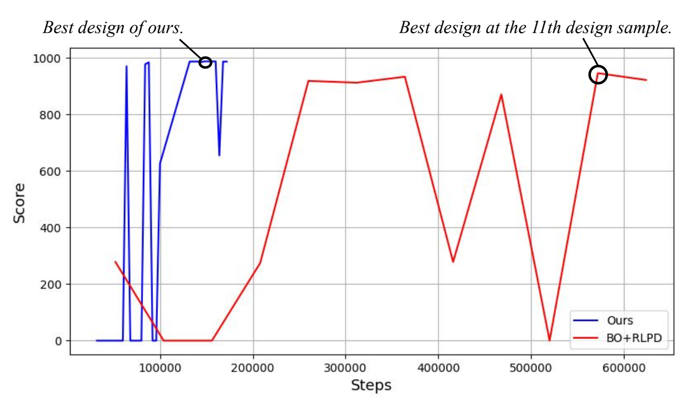
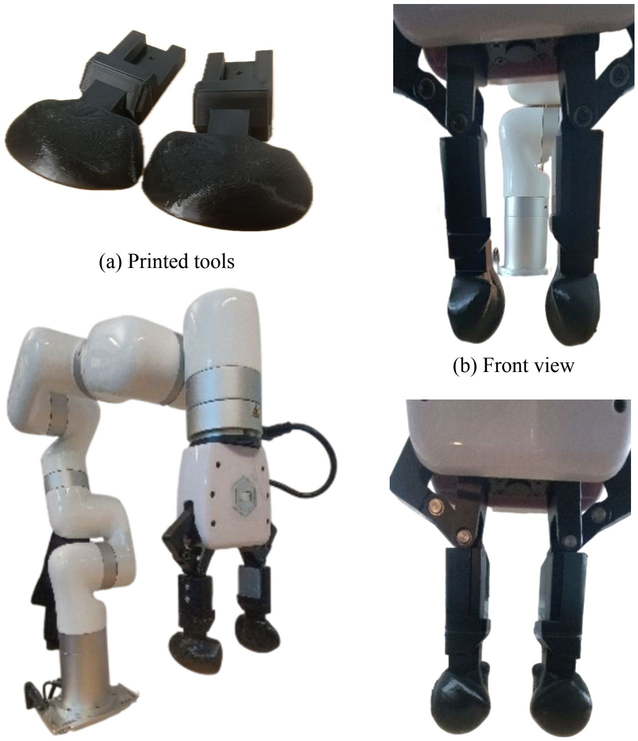


Fig. 7: Visualization of the sample efficiency of our method against *BO + RLPD*. We define a single score (Equation 2) to capture both success rate and stress minimization, for visualization convenience. Note that in our method, the samples collected during the finetune stage is also included in the plot.



(a) Printed tools

(b) Front view

(c) Installation on XArm7

(d) Back view

Fig. 8: We print out the best design selected by the converged co-design policy and install it on an XArm7 robot arm to perform real-world experiments.

the choice of control finetuning by an ablation study without it. The results show that the stress applied to the tofu is further minimized through refining the control policy.

We evaluate the converged Q-value function over a range of design actions. The Q-function serves as a surrogate model for design quality, revealing how design parameters influence task performance. Specifically, we apply Principal Component Analysis (PCA) [9] to the design space using all explored designs collected during training, and project them onto the first two principal components. As shown in Figure 6, the resulting Q-value landscape exhibits two distinct clusters of high-performance designs near the center, while designs in peripheral regions exhibit significantly lower Q-values.

To better understand these patterns, we sample representative designs from different regions of the latent space and visualize their corresponding meshes. Designs sampled from high Q-value regions generally feature relatively flat contact surfaces, which maximizes contact area with the manipulated object and reduces stress concentration. In contrast, those from low Q-value regions often exhibit irregular geometries that may compromise task success and lead to higher stress.

Efficiency: By constructing an integrated MDP and providing expert demonstrations, we learn a co-design policy with high sample efficiency using our framework. We visualize the sample efficiency in Figure 7. Since we are interested in both task success rate and stress minimization, we capture two aspects into a single score, s, for sample efficiency visualization given by:

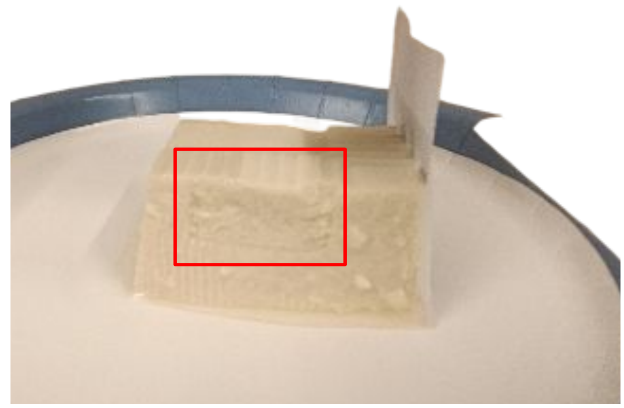


Fig. 9: When we deploy the control policy with the original parallel jaw gripper, a visible indentation on the tofu is created, as marked by the red rectangle.

$$s = \max(0, 1000 * \text{success\_rate} - \frac{\sigma_{\text{top10}}}{200}). \tag{2}$$

The best result of *BO + RLPD* appeared at the 11th iteration of the design. This required over 500k training steps in total. In contrast, our method converges in about 120k steps and significantly outperforms *BO + RLPD* in terms of stress minimization. These results demonstrate both the effectiveness of our approach and the efficiency gains from using limited demonstrations to seed and accelerate the co-design process. The performance advantage over classical hierarchical co-design validates our integrated learning paradigm. Note that due to time constraint, we only compare a single run for each method, primarily validating our efficiency.

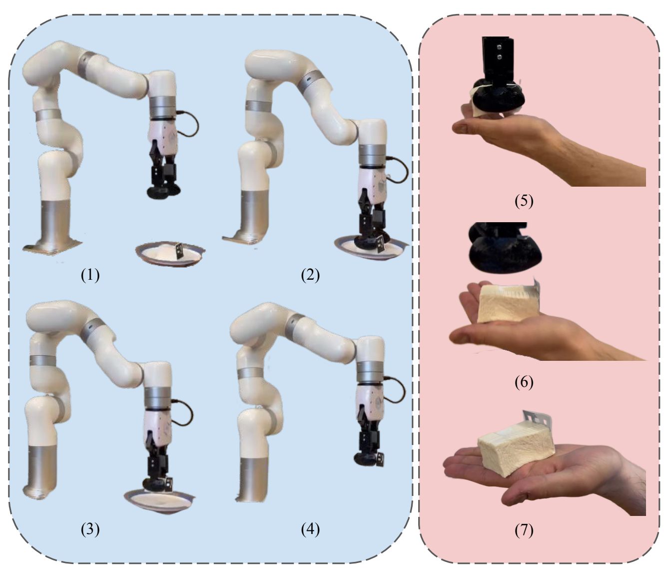


Fig. 10: Roll-out of our co-design policy in the real world. The silk tofu is successfully picked up without any visible damage.

C. Real World Evaluation

We evaluate the policies learned in simulation by transferring them to the real world in a zero-shot manner. We compare the grasp outcome of a silk tofu block with the same dimensions as in simulation, using policies from *Original Gripper + RLPD* and *Ours*. We 3D print and install the learned design as shown in Figure 8. In addition, we utilize an external camera to identify the object pose.

As shown in Figure 9, by using the original parallel jaw gripper the robot damages the tofu while attempting to lift it, preventing successful task completion. In contrast, with the learned design the tofu can be held tightly in hand without noticeable damage after picking, as presented in Figure 10. These qualitative results demonstrate that the co-designed robot gripper significantly reduces the damage to deformable fragile objects during manipulation.

V. Conclusion

In this workshop contribution, we presented ongoing work towards a data-driven co-design framework for deformable and fragile object manipulation, with a particular focus on stress minimization during contact. Our framework incorporates human demonstrations into both the design and control processes via an inverse design module and an offline-online reinforcement learning pipeline using RLPD. This facilitates sample-efficient policy learning in a high-dimensional, coupled design-control space. The cage-based deformation method provides a compact yet expressive design space that generates diverse end-effector geometries specifically tailored to the manipulation task.

Through simulated and real-world experiments on a silk tofu grasping task, we demonstrate that our co-design framework produces design-control pairs that outperform baselines in terms of both task success and stress minimization. Notably, we find that expert demonstrations significantly improve convergence and performance, highlighting the importance of incorporating human priors into co-design.

In future, we identify several promising research directions. First, extending our framework to incorporate high-dimensional point cloud observations could enhance performance for contact-rich manipulation tasks involving large object deformations, such as food item pickup and scooping operations. Second, while the current CBD formulation uses spherical base meshes suitable for simple topologies, exploring alternative base meshes (e.g., tori) could expand the approach to objects with more complex topological structures. Finally, future work could investigate the co-design of shape, material softness, and control for soft end-effectors, as well as the integration of additional sensing modalities like tactile sensors to further improve robustness and performance.

References

[1] Austin Allison, Nathaniel Hanson, Sebastian Wicke, and Taşkın Padır. Hashi: Highly adaptable seafood handling instrument for manipulation in industrial settings. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4191–4197. IEEE, 2024.

[2] Philip J Ball, Laura Smith, Ilya Kostrikov, and Sergey Levine. Efficient online reinforcement learning with offline data. In *International Conference on Machine Learning*, pages 1577–1594. PMLR, 2023.

[3] Jagdeep Bhatia, Holly Jackson, Yunsheng Tian, Jie Xu, and Wojciech Matusik. Evolution gym: A large-scale benchmark for evolving soft robots. *Advances in Neural Information Processing Systems*, 34:2201–2214, 2021.

[4] Filip Bjelonic, Joonho Lee, Philip Arm, Dhionis Sako, Davide Tateo, Jan Peters, and Marco Hutter. Learning-based design and control for quadrupedal robots with parallel-elastic actuators. *IEEE Robotics and Automation Letters*, 8(3):1611–1618, 2023.

[5] Ignacio Cuiral-Zueco, Yiannis Karayiannidis, and Gonzalo López-Nicolás. Contour based object-compliant shape control. *IEEE Robotics and Automation Letters*, 8(8):5164–5171, 2023.

[6] Ruta Desai, Ye Yuan, and Stelian Coros. Computational abstractions for interactive design of robotic devices. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1196–1203. IEEE, 2017.

[7] Yifei Dong, Shaohang Han, Xianyi Cheng, Werner Friedl, Rafael I. Cabral Muchacho, Máximo A. Roa, Jana Tumova, and Florian T. Pokorny. Cagecoopt: Enhancing manipulation robustness through caging-guided morphology and policy co-optimization, 2025. URL https://arxiv.org/abs/2409.11113.

[8] Ho Enomoto, Matthew Ishige, Takuya Umedachi, Mitsuhiro Kamezaki, and Yoshihiro Kawahara. Delicate jamming grasp: Detecting deformation of fragile objects using permanent magnet elastomer membrane. *IEEE Robotics and Automation Letters*, 9(2):979–986, 2023.

[9] Karl Pearson F.R.S. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[10] A Gafer, D Heymans, Domenico Prattichizzo, and Gionata Salvietti. The quad-spatula gripper: A novel soft-rigid gripper for food handling. In *2020 3rd IEEE International Conference on Soft Robotics (RoboSoft)*, pages 39–45. IEEE, 2020.

[11] Yuru Gong, Yuxuan Wang, Jianhua Wu, and Zhenhua Xiong. Deformability-aware stiffness-guided robotic grasping strategy based on tactile sensing. In *2024 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pages 31–37. IEEE, 2024.

[12] Michelle Guo, Ziang Liu, Stephen Tian, Zhaoming Xie, Jiajun Wu, and C Karen Liu. Learning to design 3d printable adaptations on everyday objects for robot manipulation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 824–830. IEEE, 2024.

[13] Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel, et al. Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*, 2018.

[14] Isabella Huang, Yashraj Narang, Clemens Eppner, Balakumar Sundaralingam, Miles Macklin, Ruzena Bajcsy, Tucker Hermans, and Dieter Fox. Defgraspsim: Physics-based simulation of grasp outcomes for 3d deformable objects. *IEEE Robotics and Automation Letters*, 7(3): 6274–6281, 2022.

[15] Isabella Huang, Yashraj Narang, Ruzena Bajcsy, Fabio Ramos, Tucker Hermans, and Dieter Fox. Defgraspnets: Grasp planning on 3d fields with graph neural nets. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5894–5901. IEEE, 2023.

[16] Joshua D Langsfeld, Ariyan M Kabir, Krishnanand N Kaipa, and Satyandra K Gupta. Robotic bimanual cleaning of deformable objects with online learning of part and tool models. In *2016 IEEE international conference on automation science and engineering (CASE)*, pages 626–632. IEEE, 2016.

[17] Ziang Liu, Stephen Tian, Michelle Guo, Karen Liu, and Jiajun Wu. Learning to design and use tools for robotic manipulation. In Jie Tan, Marc Toussaint, and Kourosh Darvish, editors, *Proceedings of The 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pages 887–905. PMLR, 2023.

[18] Miles Macklin, Kenny Erleben, Matthias Müller, Nuttapong Chentanez, Stefan Jeschke, and Viktor Makoviychuk. Non-smooth newton methods for deformable multi-body dynamics. *ACM Transactions on Graphics (TOG)*, 38(5):1–20, 2019.

[19] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv preprint arXiv:2108.10470*, 2021.

[20] Jonas Mockus. The bayesian approach to global optimization. In *System Modeling and Optimization: Proceedings of the 10th IFIP Conference New York City, USA, August 31–September 4, 1981*, pages 473–481. Springer, 2005.

[21] Jesús R Nieto and Antonio Susín. Cage based deformations: a survey. In *Deformation Models: Tracking, Animation and Applications*, pages 75–99. Springer, 2012.

[22] Xinlei Pan, Animesh Garg, Animashree Anandkumar, and Yuke Zhu. Emergent hand morphology and control from optimizing robust grasps of diverse objects. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7540–7547. IEEE, 2021.

[23] Deepak Pathak, Christopher Lu, Trevor Darrell, Phillip Isola, and Alexei A Efros. Learning to control self-assembling morphologies: a study of generalization via modularity. *Advances in Neural Information Processing Systems*, 32, 2019.

[24] Chang Rajani, Karol Arndt, David Blanco-Mulero, Kevin Sebastian Luck, and Ville Kyrki. Co-imitation: learning design and behaviour by imitation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 6200–6208, 2023.

[25] Marc Toussaint, Jung-Su Ha, and Ozgur S Oguz. Co-optimizing robot, environment, and tool design via joint manipulation planning. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6600–6606. IEEE, 2021.

[26] Li-Jun Wang, Dong Li, Eizo Tatsumi, Zhi-Sheng Liu, Xiao Dong Chen, and Li-Te Li. Application of two-stage ohmic heating to tofu processing. *Chemical Engineering and Processing: Process Intensification*, 46(5):486–490, 2007.

[27] Zhongkui Wang, Keung Or, and Shinichi Hirai. A dual-mode soft gripper for food packaging. *Robotics and Autonomous Systems*, 125:103427, 2020.

[28] Ruoshi Wen, Kai Yuan, Qiang Wang, Shuai Heng, and Zhibin Li. Force-guided high-precision grasping control of fragile and deformable objects using semg-based force prediction. *IEEE Robotics and Automation Letters*, 5(2): 2762–2769, 2020.

[29] Jie Xu, Tao Chen, Lara Zlokapa, Michael Foshey, Wojciech Matusik, Shinjiro Sueda, and Pulkit Agrawal. An end-to-end differentiable framework for contact-aware robot design. *arXiv preprint arXiv:2107.07501*, 2021.

[30] Rinto Yagawa, Reina Ishikawa, Masashi Hamaya, Kazutoshi Tanaka, Atsushi Hashimoto, and Hideo Saito. Learning food picking without food: Fracture anticipation by breaking reusable fragile objects. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 917–923. IEEE, 2023.

[31] Ye Yuan, Yuda Song, Zhengyi Luo, Wen Sun, and Kris Kitani. Transform2act: Learning a transform-and-control policy for efficient agent design. In *International Conference on Learning Representations*, 2022.