Initial Results on Grasping and Lifting Physical Deformable Bags with a Bimanual Robot

Daniel Seita[1], Justin Kerr[1], John Canny[1], Ken Goldberg[1]

***Abstract*— Rearranging and manipulating highly deformable bags is common in daily life, such as when we put food in grocery bags or papers in backpacks or garbage in trash bags. In contrast to rigid object manipulation, manipulating deformable objects is challenging for robots due to complex object configurations and dynamics. Furthermore, the task of manipulating highly deformable bags introduces extra complexities due to reasoning about 3-dimensional space. In this work, we consider a novel task: grasping and lifting physical bags to contain items. Using a bimanual ABB YuMi robot we test three grasping methods, where positions are determined via a human, a random baseline, and a baseline based on grasping the leftmost and rightmost points of the bag ("Maximum Width"). Across experiments where the YuMi grasps and lift bags to contain a fabric and a cable, we perform 15 trials for each method. Results demonstrate that the human has the best success rates (14/15), followed by Maximum Width (10/15) and then random (5/15). Supplementary material is available at `https://sites.google.com/view/physical-bags/home`.**

Fig. 1: **Setup**. We use a bimanual ABB YuMi robot with a PhotoNeo RGBD camera mounted 0.8m above a cardboard surface. Before each trial, we cast a highly deformable bag on the surface, then cast a wrapped charger cable and a 5x5 inch fabric on top of the bag. Given the camera depth image, the robot performs algorithmic or human policies to determine where to grasp the bag. After grasping, the robot lifts the bag upwards, and shakes its grippers to test for robustness. We consider the trial a success if both items are contained in the bag and do not make contact with the surface.

I. Introduction

From putting school items in a backpack to arranging groceries into bags, humans use bags, sacks, and other similar 3D structures on a regular basis to efficiently contain and transport multiple items. Translating this type of behavior to robots, however, remains challenging. In contrast to rigid object manipulation, deformable object manipulation is challenging for robots due to infinite-dimensional object configuration spaces, complex dynamics, and occlusions. While prior work has advanced robot manipulation of deformables for 1D tasks (*e.g.,* configuring rope to targets [23], [34], [37], [38]) and 2D tasks (*e.g.,* fabric smoothing and folding [7], [11], [21], [30], [36], [38], [19], [1]), manipulation of bags with items has extra complexities due to its fundamental 3D nature. For example, bags might not be able to easily stretch or smooth on a flat surface, and lifting bags with items requires understanding if items will remain contained.

The main contribution of this paper is the formulation of bimanual bag grasping and lifting. We assume that a highly deformable bag is strewn on a workspace with most of its opening visible from a top-down view, and where two items are cast on top of the bag and have to be contained (see Figure 1). Using a bimanual ABB YuMi robot, we test three methods for determining the bag grasping points. Results over 45 total trials, 15 per method, suggest that grasping the leftmost and rightmost points is better than a random baseline, but does not match human performance. This motivates exciting avenues for future research.

[1]University of California, Berkeley, USA.
Correspondence to `seita@berkeley.edu`

II. Related Work

In this work, a "deformable" refers to an object whose configuration cannot be adequately expressed as a 6-DoF pose. Examples include ropes, strings, cables, clothing, blankets, pillows, dough, sand, and liquids. While difficult, researchers have designed methods for deformable manipulation, with applications including knot-tying [10], [22], [28], surgical suturing [32], dressing assistance [5], [6], [8], ironing [17], laundry folding [18], [39], [20], bed-making [31], handling granular media [27], [3], perceiving clothing texture [40], and fabric upholstery manufacturing [25], [33]. We refer the reader to [2], [26], [41] for surveys.

While there has been significant prior work in manipulation of 1D structures such as ropes and 2D structures such as fabrics [2], there has been little prior work on manipulation of complex 3D structures such as bags and sacks. We use the term "bags" to refer to any deformable object meant to contain items. Some of the earliest research on robots manipulating bags include robots grasping and lifting bags with pre-filled hazardous waste [12] or deformable material [13].

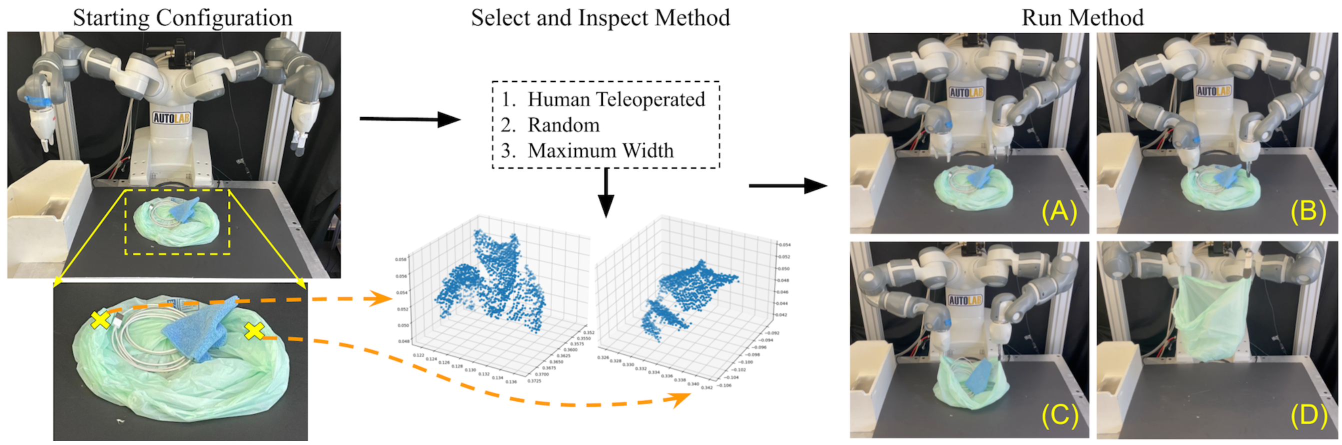

Fig. 2: **Procedure for 1 Trial**. The robot begins in a standardized home position. Given the bag and items on the surface (left), we randomly choose a grasping method: Human Teleoperated, Random, or Maximum Width; for this figure, we show Human Teleoperated. The method provides two grasping points (visualized with yellow "X"s), and we compute the principal grasp axes of the point clouds near each grasping point to determine rotations. The robot moves its grippers above the grasping point with the grippers rotated (A), lowers and closes them (B), raises grippers by 0.4 meters (C) and finally performs a shaking motion to test robustness of item containment (D).

Other work with robots and bags constrains the setup so that bags experience limited deformations [9], [16]. A separate research focus is on the mechanical design of robots for grasping [14] or unloading [15] bags. Recently, researchers have simulated bags and modeled their interaction with rigid items [35], which can be used to predict a bag's future configuration. In this paper, we use a bimanual robot without specialized end-effectors for an unstructured physical bag that has be manipulated to contain items.

In among the most relevant prior work, Seita et al. [29] propose a suite of deformable tasks using PyBullet [4] simulation and among the tasks include several that require a robot to open a bag, to insert item(s) in the bag, then to lift and transport items to a target zone. Unlike this prior work, we use real-world bags with physical robots. To the best of our knowledge, this work is the first which addresses physical robotic bag manipulation for item containment.

III. Problem Statement

A bimanual robot is fixed with two parallel-jaw end-effectors, and uses a top-down RGBD camera to observe a flat surface supporting one deformable bag with most of its opening facing upwards, along with items strewn on top of the bag's exterior surface. We assume all points on the surface are reachable by at least one of the two robot grippers. The robot utilizes a 4-DoF action space[1] for each arm, and the action space is defined as the choice of two 4-DoF actions (one per arm):

$$\mathbf{a} = (\mathbf{a}^{(0)}, \mathbf{a}^{(1)}) \tag{1}$$

where $\mathbf{a}^{(0)}$ and $\mathbf{a}^{(1)}$ operate under the same action space, so without loss of generality, hereafter we describe $\mathbf{a}^{(0)}$. We parameterize the single-arm action as:

$$\mathbf{a}^{(0)} = (x, y, z, \gamma) \tag{2}$$

[1]We also attempted full 6-DoF grasping per gripper, but the additional degrees of freedom did not show benefits for the experimental setup and made enforcing safety requirements more complex.

where x, y, z represent the end-effector tip position and γ represents the angle of rotation about the z-axis for a gripper pointing downwards. The task is then posed as determining a function π that, given a visual observation $\mathbf{o} \in \mathcal{O}$ from a top-down RGBD camera, determines an action $\mathbf{a}$ to take.

We define a *trial* as an instance of the task. Each trial consists of one robot action, where it moves both grippers simultaneously to grasp the bag, then lifts the grippers by a fixed height. The robot finally shakes the grippers to test whether the bag is robustly grasped, and whether the items are robustly contained. If, after shaking, all items have been transported off the surface and stay contained in the bag or remain in midair, the trial is considered a success.

IV. Methods

We use the bimanual ABB YuMi robot and test the following 3 methods for grasping the bag:

1) **Human Teleoperated**: a human decides the grasping points by selecting pixels on a click interface.
2) **Random**: randomly sample two points of the bag.
3) **Maximum Width**: set the grasping points to be the leftmost and rightmost regions of the bag.

To avoid collisions between the arms, all 3 methods enforce a minimum distance threshold for the two grasping points. We test Human Teleoperated to get an upper bound on performance, Random to obtain a lower bound, and Maximum Width to investigate the reliability of grasping at the left and right ends to contain items. The surface is longer in the left-right direction with respect to the robot's base; thus, the bag tends to be stretched more in this direction.

For all methods, the robot executes top-down grasps for both grippers by first obtaining a depth image of the bag and items and extracting the resulting point cloud. Given the desired real world position for the left and right grippers, we perform a flood fill algorithm [24] to detect a set of *local* points about both grasping points, and compute the principal axes of these points. We then project these axes to

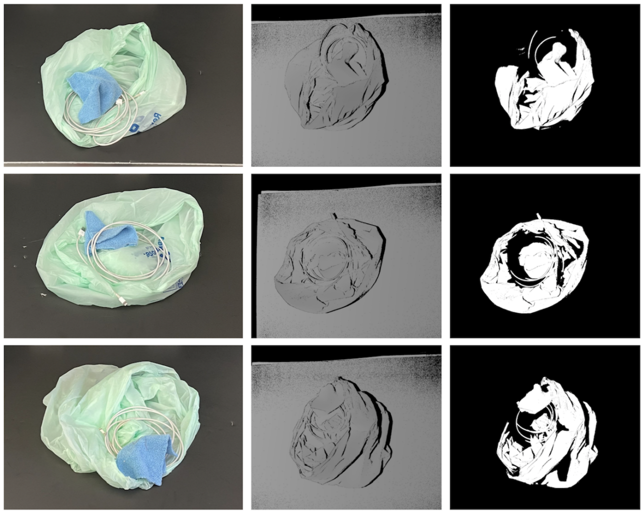

Fig. 3: Three examples of starting configurations for a trial (one per row). Each row has three types of images, from left to right: photo of the scene taken from an external third-person view, the depth image from the robot's camera, and the "bag mask" used to determine bag and item pixels for Maximum Width and Random methods. The depth and bag mask images are from the same top-down angle. All photos are cropped and zoomed-in for clarity.

the xy-plane to ignore any z-coordinate. At the end of this process, the robot lifts and then shakes the bag with both of its grippers, one after another, by rotating about the wrist axis pointing to the robot's left. See Figure 2 for a visualization.

V. Experiments

We use a double-layer Repurpose 3-Gallon Compostable trash bag, due to high deformability, suitable size and weight, and ease of point cloud detection. The double layer helps to increase thickness to reduce the chances of the bag slipping from the gripper. We test with a 2 meter iPhone charger cable and a 5x5 inch fabric (used in [11], [30]). For each trial, we engage in the following protocol:

1) We initialize the bag with starting configurations that have most of the bag opening facing upwards
2) We drop the cable and fabric from 0.4 meters at a point above the bag opening. See Figure 3 for representative examples of starting configurations after this step.
3) We randomly choose one of the 3 methods from Section IV and execute it. This is done after setting the starting configuration to avoid bias.
4) We judge success or failure based on whether the bag and both items are entirely above the surface and do not make contact after the shaking motion.

We run for 45 total trials, with 15 for each method. The project website has videos of the trials.

VI. Results

Table I reports success rates, where Human Teleoperated has the best performance with a 14/15 success rate, while Maximum Width is next at 10/15, and Random is last with 5/15. The results suggest that the choice of grasping method has a significant impact on performance.

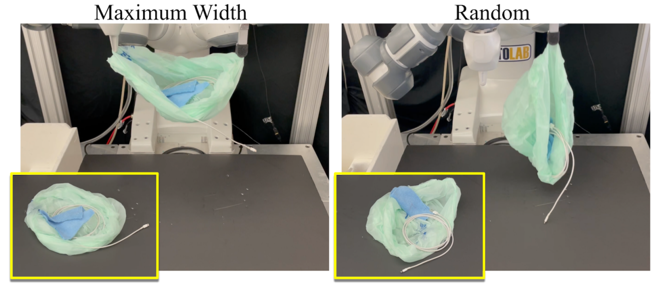

Fig. 4: **Qualitative Observations**. Left: Maximum Width can lift the bag to support items, but the bag's cavity might not be as large compared to when the robot grasps at the bag opening edges, in contrast to Figure 2's example. Right: the bag can slip from grippers, typically when grasping at a thin layer. Here, using Random, the cable touches the surface after shaking. For Table I purposes, the left shows a success and the right shows a failure. In both cases, we overlay the starting configuration to the lower left.

TABLE I: **Bag grasping and lifting results**. We run 3 methods for determining the grasp points (Section IV). A success is counted as whether at least one gripper holds the bag in midair and both items (the cable and fabric) also remain in midair and do not touch the surface after the shaking motion of grippers.

Algorithm	**Success**
Human Teleoperated	14/15
Random	5/15
Maximum Width	10/15

Qualitatively, the human is effective at picking grasp points that lead to item containment, which may stem from strong priors on how gravity works and how items interact with bag edges. See Figure 2 for an example frame-by-frame overview of the Human Teleoperated method. Maximum Width is less effective than Human Teleoperated; it picks at the leftmost and rightmost points of the bag, which is often not the same as picking at the bag opening's edges. This can lead to a smaller cavity to contain items, though this can still sometimes lift items off the workspace (see Figure 4). In other cases, the robot was only able to lift the bag with one of its grippers. This typically happens when a gripper grasps a thinly-layered area of the bag, making it difficult to experience a firm grip. This can result in either a success or a failure. For example, Figure 4 shows that the Random method was only able to lift with one gripper, which caused the cable to make contact with the surface (*i.e.,* failure).

VII. Conclusion and Future Work

We formulate a system and report initial experiments for a one-step physical bag grasping and lifting motion with a bimanual ABB YuMi robot. We plan on expanding subtasks to include bag opening and item insertion, while widening the range of bag and item types. In future work, we will use a learning-based approach to increase generalization and address observed failure cases. We will also incorporate the use of human-in-the-loop policies to assist physical bag manipulation and apply curriculum learning to guide and stabilize learning. We look forward to discussing this with the workshop participants.

References

[1] Anonymous, "FabricFlowNet: Bimanual Cloth Manipulation with a Flow-based Policy," in *Submitted to 5th Annual Conference on Robot Learning*, 2021, under review. [Online]. Available: https://openreview.net/forum?id=TsqkJJMgHkk

[2] J. Borras, G. Alenya, and C. Torras, "A Grasping-centered Analysis for Cloth Manipulation," *arXiv preprint arXiv:1906.08202*, 2019.

[3] S. Clarke, T. Rhodes, C. Atkeson, and O. Kroemer, "Learning Audio Feedback for Estimating Amount and Flow of Granular Material," in *Conference on Robot Learning (CoRL)*, 2018.

[4] E. Coumans and Y. Bai, "PyBullet, a Python Module for Physics Simulation for Games, Robotics and Machine Learning," http://pybullet.org, 2016–2020.

[5] Z. Erickson, H. Clever, G. Turk, C. K. Liu, and C. Kemp, "Deep Haptic Model Predictive Control for Robot-Assisted Dressing," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[6] Z. Erickson, M. Collier, A. Kapusta, and C. Kemp, "Tracking Human Pose During Robot-Assisted Dressing using Single-Axis Capacitive Proximity Sensing," in *IEEE Robotics and Automation Letters (RA-L)*, 2018.

[7] A. Ganapathi, P. Sundaresan, B. Thananjeyan, A. Balakrishna, D. Seita, J. Grannen, M. Hwang, R. Hoque, J. Gonzalez, N. Jamali, K. Yamane, S. Iba, and K. Goldberg, "Learning Dense Visual Correspondences in Simulation to Smooth and Fold Real Fabrics," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[8] Y. Gao, H. J. Chang, and Y. Demiris, "Iterative Path Optimisation for Personalised Dressing Assistance using Vision and Force Information," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[9] R. B. Hellman, C. Tekin, M. van der Schaar, and V. J. Santos, "Functional Contour-following via Haptic Perception and Reinforcement Learning," in *IEEE Transactions on Haptics*, 2018.

[10] J. Hopcroft, J. Kearney, and D. Krafft, "A Case Study of Flexible Object Manipulation," in *International Journal of Robotics Research (IJRR)*, 1991.

[11] R. Hoque, D. Seita, A. Balakrishna, A. Ganapathi, A. Tanwani, N. Jamali, K. Yamane, S. Iba, and K. Goldberg, "VisuoSpatial Foresight for Multi-Step, Multi-Task Fabric Manipulation," in *Robotics: Science and Systems (RSS)*, 2020.

[12] A. Howard and G. Bekey, "Prototype System for Automated Sorting and Removal of Bags of Hazardous Waste," in *SPIE, Intelligent Robots and Computer Vision*, 1996.

[13] ——, "Intelligent Learning for Deformable Object Manipulation," in *Autonomous Robtoics*, 2000.

[14] H. Kazerooni and C. Foley, "A Robot Mechanism for Grapsing Sacks," in *IEEE Transactions on Automation Science and Engineering*, 2005.

[15] A. Kirchheim, M. Burwinkel, and W. Echelmeyer, "Automatic Unloading of Heavy Sacks From Containers," in *IEEE International Conference on Automation and Logistics (ICAL)*, 2008.

[16] E. Klingbeil, D. Rao, B. Carpenter, V. Ganapathi, A. Y. Ng, and O. Khatib, "Grasping With Application to an Autonomous Checkout Robot," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2011.

[17] Y. Li, X. Hu, D. Xu, Y. Yue, E. Grinspun, and P. K. Allen, "Multi-Sensor Surface Analysis for Robotic Ironing," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[18] Y. Li, Y. Yue, D. Xu, E. Grinspun, and P. K. Allen, "Folding Deformable Objects using Predictive Simulation and Trajectory Optimization," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[19] X. Lin, Y. Wang, and D. Held, "Learning Visible Connectivity Dynamics for Cloth Smoothing," *arXiv preprint arXiv:2105.10389*, 2021.

[20] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel, "Cloth Grasp Point Detection Based on Multiple-View Geometric Cues with Application to Robotic Towel Folding," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.

[21] J. Matas, S. James, and A. J. Davison, "Sim-to-Real Reinforcement Learning for Deformable Object Manipulation," in *Conference on Robot Learning (CoRL)*, 2018.

[22] T. Morita, J. Takamatsu, K. Ogawara, H. Kimura, and K. Ikeuchi, "Knot Planning from Observation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2003.

[23] A. Nair, D. Chen, P. Agrawal, P. Isola, P. Abbeel, J. Malik, and S. Levine, "Combining Self-Supervised Learning and Imitation for Vision-Based Rope Manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[24] W. M. Newman and R. F. Sproull, Eds., *Principles of Interactive Computer Graphics (2nd Ed.)*. USA: McGraw-Hill, Inc., 1979.

[25] J. K. Parker, R. Dubey, F. W. Paul, and R. J. Becker, "Robotic Fabric Handling for Automating Garment Manufacturing," *Journal of Manufacturing Science and Engineering*, vol. 105, 1983.

[26] J. Sanchez, J.-A. Corrales, B.-C. Bouzgarrou, and Y. Mezouar, "Robotic Manipulation and Sensing of Deformable Objects in Domestic and Industrial Applications: a Survey," in *International Journal of Robotics Research (IJRR)*, 2018.

[27] C. Schenck, J. Tompson, D. Fox, and S. Levine, "Learning Robotic Manipulation of Granular Media," in *Conference on Robot Learning (CoRL)*, 2017.

[28] J. Schulman, J. Ho, C. Lee, and P. Abbeel, "Learning from Demonstrations Through the Use of Non-Rigid Registration," in *International Symposium on Robotics Research (ISRR)*, 2013.

[29] D. Seita, P. Florence, J. Tompson, E. Coumans, V. Sindhwani, K. Goldberg, and A. Zeng, "Learning to Rearrange Deformable Cables, Fabrics, and Bags with Goal-Conditioned Transporter Networks," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[30] D. Seita, A. Ganapathi, R. Hoque, M. Hwang, E. Cen, A. K. Tanwani, A. Balakrishna, B. Thananjeyan, J. Ichnowski, N. Jamali, K. Yamane, S. Iba, J. Canny, and K. Goldberg, "Deep Imitation Learning of Sequential Fabric Smoothing From an Algorithmic Supervisor," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[31] D. Seita, N. Jamali, M. Laskey, R. Berenstein, A. K. Tanwani, P. Baskaran, S. Iba, J. Canny, and K. Goldberg, "Deep Transfer Learning of Pick Points on Fabric for Robot Bed-Making," in *International Symposium on Robotics Research (ISRR)*, 2019.

[32] S. Sen, A. Garg, D. V. Gealy, S. McKinley, Y. Jen, and K. Goldberg, "Automating Multiple-Throw Multilateral Surgical Suturing with a Mechanical Needle Guide and Sequential Convex Optimization," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[33] E. Torgerson and F. Paul, "Vision Guided Robotic Fabric Manipulation for Apparel Manufacturing," in *IEEE International Conference on Robotics and Automation (ICRA)*, 1987.

[34] A. Wang, T. Kurutach, K. Liu, P. Abbeel, and A. Tamar, "Learning Robotic Manipulation through Visual Planning and Acting," in *Robotics: Science and Systems (RSS)*, 2019.

[35] Z. Weng, F. Paus, A. Varava, H. Yin, T. Asfour, and D. Kragic, "Graph-based Task-specific Prediction Models for Interactions between Deformable and Rigid Objects," *arXiv preprint arXiv:2103.02932*, 2021.

[36] Y. Wu, W. Yan, T. Kurutach, L. Pinto, and P. Abbeel, "Learning to Manipulate Deformable Objects without Demonstrations," in *Robotics: Science and Systems (RSS)*, 2020.

[37] M. Yan, Y. Zhu, N. Jin, and J. Bohg, "Self-Supervised Learning of State Estimation for Manipulating Deformable Linear Objects," in *IEEE Robotics and Automation Letters (RA-L)*, 2020.

[38] W. Yan, A. Vangipuram, P. Abbeel, and L. Pinto, "Learning Predictive Representations for Deformable Objects Using Contrastive Estimation," in *Conference on Robot Learning (CoRL)*, 2020.

[39] P.-C. Yang, K. Sasaki, K. Suzuki, K. Kase, S. Sugano, and T. Ogata, "Repeatable Folding Task by Humanoid Robot Worker Using Deep Learning," in *IEEE Robotics and Automation Letters (RA-L)*, 2017.

[40] W. Yuan, Y. Mo, S. Wang, and E. Adelson, "Active Clothing Material Perception using Tactile Sensing and Deep Learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[41] J. Zhu, A. Cherubini, C. Dune, D. Navarro-Alarcon, F. Alambeigi, D. Berenson, F. Ficuciello, K. Harada, X. Li, J. Pan, and W. Yuan, "Challenges and Outlook in Robotic Manipulation of Deformable Objects," *arXiv preprint arXiv:2105.01767*, 2021.