Bagging by Learning to Singulate Layers Using Interactive Perception

Lawrence Yunliang Chen[1], Baiyu Shi[1], Roy Lin[1], Daniel Seita[2], Ayah Ahmad[1], Richard Cheng[3], Thomas Kollar[3], David Held[2], Ken Goldberg[1]

***Abstract*— Many fabric handling and 2D deformable material tasks in homes and industries require singulating layers of material such as opening a bag or arranging garments for sewing. In contrast to methods requiring specialized sensing or end effectors, we use only visual observations with ordinary parallel jaw grippers. We propose SLIP: Singulating Layers using Interactive Perception, and apply SLIP to the task of autonomous bagging. We develop SLIP-Bagging, a bagging algorithm that manipulates a plastic or fabric bag from an unstructured state and uses SLIP to grasp the top layer of the bag to open it for object insertion. In physical experiments, a YuMi robot achieves a success rate of 67% to 81% across bags of a variety of materials, shapes, and sizes, significantly improving in success rate and generality over prior work. Experiments also suggest that SLIP can be applied to tasks such as singulating layers of folded cloth and garments. Supplementary material is available at `https://sites.google.com/view/slip-bagging/`.**

I. Introduction

Many tasks in homes and factories require grasping a single layer of 2D deformable objects. Examples include taking one napkin from a stack of napkins, grasping the top layer of a folded towel to unfold it, grasping a single layer of a T-shirt to insert into a hanger, and grasping a single layer of a bag to hold it open while placing items inside. Humans manipulate such deformable objects with great dexterity using touch and vision. Such tasks are very challenging for robots, as a 1 mm change in the gripper height can lead to the difference between a missed (0-layer) grasp, a 1-layer grasp, and a 2-layer grasp. On the other hand, enabling touch sensing may require equipping the robot end effector with compliant grippers or special tactile sensors.

In this work, we achieve single-layer grasping with a high success rate using a bimanual robot with an ordinary parallel-jaw gripper. We use self-supervised learning to identify where to grasp, and we use interactive perception [1] to determine the number of layers grasped. The robot iteratively adjusts its grasp until it successfully grasps a single layer.

We propose SLIP: Singulating Layers using Interactive Perception and apply SLIP to bags and fabrics. We use the task of autonomous bagging as a motivating example—opening a deformable bag from an unstructured initial state and putting objects into it—and demonstrate in physical experiments how SLIP can significantly improve the success rate of robot bagging by 5× from prior work and is effective across a variety of bag materials. Moreover, we conduct physical experiments to evaluate the applicability of SLIP to singulating layers for various fabrics and garments.

[1]The AUTOLab at UC Berkeley (autolab.berkeley.edu).
[2]The Robotics Institute at Carnegie Mellon University.
[3]Toyota Research Institute, Los Altos, USA.
Correspondence to: `yunliang.chen@berkeley.edu`

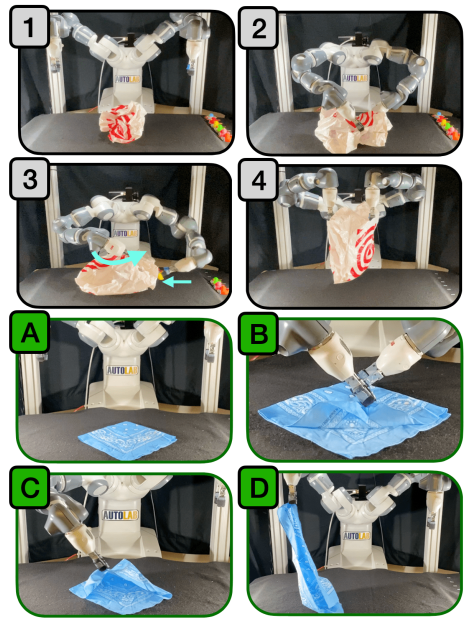

Fig. 1: SLIP-Bagging. **Top 2 rows:** (1) Initial unstructured and deformed bag. (2) The robot flattens the bag, and then (3) uses SLIP to grasp the top layer of the bag, rotates it by 90°, and inserts objects. (4) The robot lifts the bag filled with the inserted items. **Bottom 2 rows:** (A) Initial configuration of a piece of folded cloth. (B) The robot uses SLIP to grasp the top layer of a folded square cloth. (C) After grasping the top layer, the robot lifts the cloth up. (D) After shaking, the cloth is successfully expanded.

II. Related Work

A. Deformable Bags and Single-Layer Grasping

There is a rich literature on deformable object manipulation; see [2–4] for representative surveys. Among deformable object manipulation, fabric manipulation is one of the most widely-studied areas [5–14]. While some prior work has studied singulating a single sheet or fabric layer from a stack, most use tactile sensing or specialized end effectors. Tirumala et al. [15] use a ReSkin sensor [16] and Manabe et al. [17] design a rolling hand mechanism to separate a single sheet of fabrics. Guo et al. [18] use a XELA uSkin tactile sensor combined with visual inputs to turn a single book page. In this work, we propose to singulate layers with standard end effectors purely from visual feedback. Demura et al. [19] study grasping the top folded towel from a stack using visual feedback with a scooping action, using towels which are each several millimeters thick. In contrast, we study tasks where layers can be thinner than 1 mm.

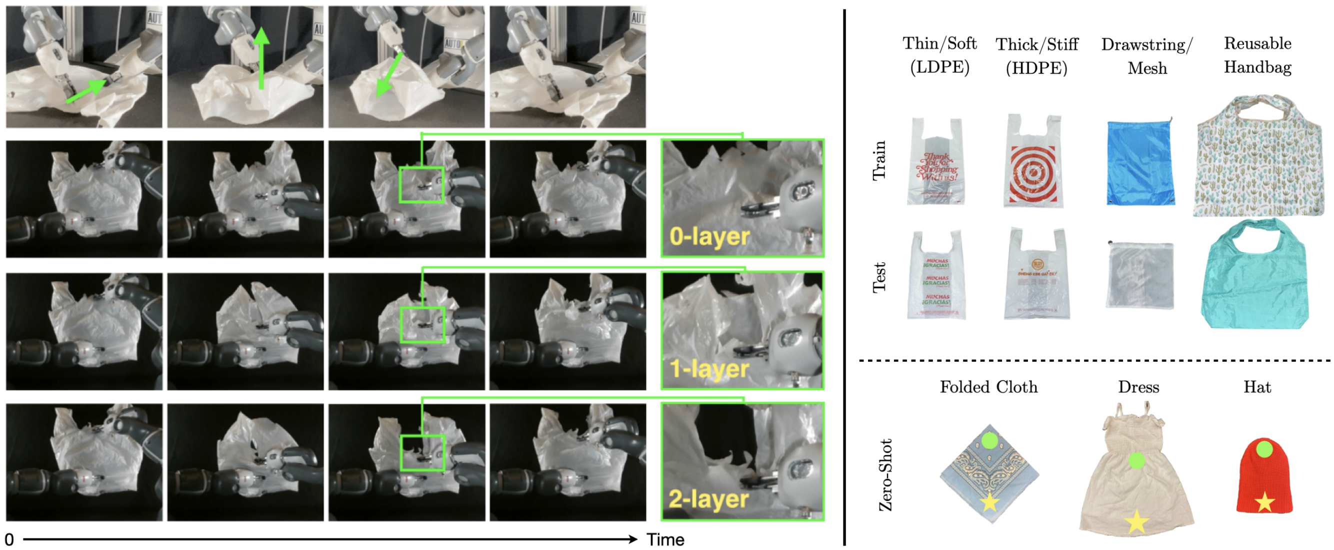

Fig. 2: **Left:** Examples of SLIP in action. Each row shows an example of one cyclic, triangular trajectory T in an iteration where one gripper moves the bag while the other one pins the bag. **Top row:** a third-person view of the robot. **Next three rows**: top-down RGB camera views of one cyclic trajectory for different trials. They show, respectively, a 0-layer, 1-layer, and 2-layer grasp on a plastic bag. We provide zoomed-in versions of images in the third column to see the layers in more detail. See Sec. III. **Right: Top panel:** Training and test bags with various materials and shapes used in experiments (Tab. I). **Bottom panel:** Fabric and garments (Tab. II). See Sec. IV.

Autonomous bagging has wide applications in retail, food handling, home cleaning, and packing. Much of the prior work on physical experiments with deformable bags assumes a semi-structured bag state [20–25]. Recently, Chen et al. [26] propose the AutoBag algorithm for manipulating a thin plastic bag from an unstructured state. However, AutoBag frequently fails when attempting to orient the bag upward since it is not a stable pose for deformable bags. In this work, we avoid orienting the bag upward. Instead, we flatten the bag and singulate the top layer of the bag to open it, which results in much higher success rates. Moreover, while AutoBag is designed specifically for opening thin plastic bags, we show evidence that SLIP-Bagging is effective on other bag materials and shapes.

III. SLIP: Singulating Layers using Interactive Perception

In this section, we describe SLIP in the context of grasping a single layer of a deformable bag, but the algorithm applies to other fabric materials (see Sec. IV). SLIP is motivated by how, after the robot performs a grasp, it cannot easily tell how many layers it grasps from visual inputs of a static scene, as the top layer occludes the layers underneath. However, by moving the gripper and observing how the object's top surface is moving, the robot can infer how many layers it has grasped. Formally, SLIP consists of 3 components: a cyclic trajectory T of the gripper, a video classification model, and an iterative height adjustment algorithm.

The trajectory T needs to satisfy two properties: (1) The movement should reveal enough information for the robot camera to infer how many layers are grasped, and (2) the trajectory should be cyclic, so the bag roughly recovers its original state after executing T, allowing the robot to retry the grasp at the same location but with a different height. For (1), we tilt the gripper at an angle $\theta = 50°$ so the grasp point is visible in the camera. For (2), we use a triangular trajectory, where the robot gripper first moves backward, then upward, and finally forward and downward back to the original position. To prevent the deformable object from translating as a whole, we use the robot's second gripper to pin the other side. See Figure 2 for a visualization. In our implementation, the trajectory takes about 5 secs.

While the robot executes the trajectory T, the camera takes an RGB video stream of the bag. A video classification model M takes the video and classifies the grasp into 3 categories: 0 layer, 1 layer, and 2 layers. We use a SlowFast network [27], which takes in 32 images of size 224×224 sampled with a uniform interval from the video stream. Fig. 2 illustrates the visual differences among different layers grasped on a plastic bag.

Given the model classification, SLIP adjusts the gripper height and retries the grasp if it does not successfully grasp a single layer. We choose to use a fixed height adjustment each time similar to the strategy in [15], with height deltas Δh_- and Δh_+. One could also choose to let the adjustment height decay over time or use the bisection method, but we empirically find that a fixed height adjustment is more robust.

IV. SLIP-Bagging and Physical Experiments

A. *SLIP-Bagging*

We present the SLIP-Bagging algorithm. It consists of 5 steps: (1) flatten the bag, (2) grasp the top layer of the bag near the bag opening using SLIP, (3) rotate the bag sideways, (4) use the other gripper to insert objects, and (5) lift the bag. In the first step, the robot leverages parameterized primitives such as Shake, Rotate, Dilate, and Fling actions, where the grasp points are learned similar to Chen et al. [26]. Algorithm and implementation can be found on the project website.

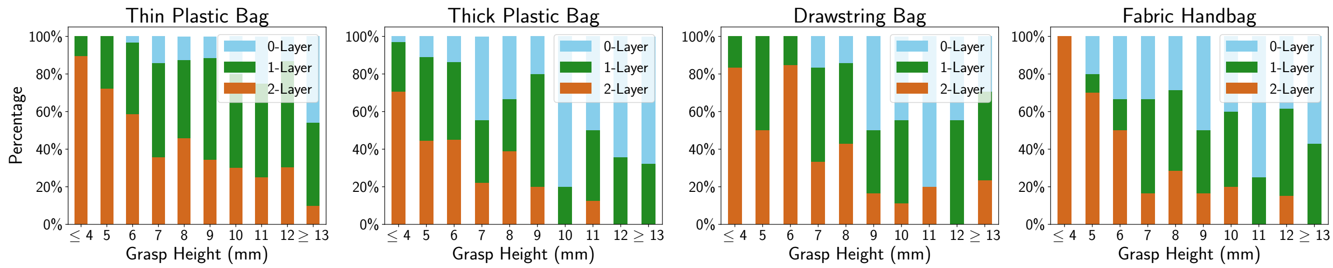

Fig. 3: Distribution of the number of layers grasped for different grasp heights for 4 different bags.

Category	Bag	Open/Flatten		Sing. Grasp		Full Success			% Objects Inserted		
		AB	SB	PD	SB	PD	AB	SB	PD	AB	SB
Thin Plastic	Train	3/6	**6/6**	1/6	**6/6**	1/6	1/6	**5/6**	17%	39%	**94%**
	Test	3/6	**6/6**	1/6	**6/6**	1/6	1/6	**4/6**	17%	36%	**75%**
Thick Plastic	Train	3/6	**5/6**	1/6	**5/5***	1/6	1/6	**3/6**	17%	36%	**56%**
	Test	2/6	**5/6**	0/6	**4/5***	0/6	2/6	**4/6**	0%	33%	**67%**
Draw-string	Train	0/6	**6/6**	0/6	**5/6**	0/6	0/6	**5/6**	0%	0%	**83%**
	Test	0/6	**6/6**	1/6	**5/6**	0/6	0/6	**4/6**	0%	0%	**67%**
Reusable Handbag	Train	0/6	**5/6**	0/6	**3/5***	0/6	0/6	**3/6**	0%	0%	**50%**
	Test	0/6	**6/6**	1/6	**6/6**	1/6	0/6	**4/6**	17%	0%	**81%**

TABLE I: Physical experiment results of SLIP-Bagging compared with baselines. 6 trials were run on each of the 8 bags (Fig. 2) for each method. Each trial attempts to insert 6 rubber ducks, and "% Objects Inserted" is the average percentage of objects inserted and contained after bag lifting. "Sing. Grasp" is the success rate of SLIP. **PD**: Perceived-Depth baseline. **AB**: AutoBag. **SB**: SLIP-Bagging. *Denominator is the number of successful flattened trials that proceed to the SLIP stage.

Objects	0-Shot Recall			SLIP Succ. Rate
	0-layer	1-layer	2-layer	
Cloth	100%	100%	62%	4/6
Dress	100%	75%	75%	4/6
Hat	100%	83%	25%	5/6

TABLE II: Non-bag experiments. The middle 3 columns show the (multi-class) recall of the video classification model trained on bags and tested on garments without finetuning. The last column shows the success rate of grasping a single layer using SLIP with the classification model.

B. *Bagging Experiments Setup*

We use a bimanual ABB YuMi robot with an overhead RealSense D435 camera. We evaluate SLIP-Bagging on 8 bags, shown in Figure 2. For each trial, we randomly initialize the bag state by taking the bag, compressing and deforming it with our hands, and the goal is to insert 6 rubber ducks. We compare SLIP-Bagging to 2 baselines: (1) Perceived-Depth (PD), which ablates the SLIP algorithm in SLIP-Bagging and grasps at the perceived depth of the grasp point measured by the depth camera. (2) AutoBag (AB), as proposed in [26]. For each bag, we conduct 6 trials.

C. *Experiments Results*

Figure 3 shows the distribution of the number of layers grasped at various grasp heights (measured from the surface height) for each of the 4 training bags in the training data. As expected, as the grasp height decreases, it is less likely to grasp 0 layers and more likely to grasp 2 layers. However, there is no single grasp height that always works, as the success depends highly on the bags' specific configuration.

Results in Table I demonstrate that SLIP-Bagging achieves a higher success rate than baselines. Perceived-Depth has a low success rate of grasping a single layer. This is because, for the mesh bag, the perceived depth is often too deep due to holes, resulting in 2-layer grasps, while for other bags, the perceived depth is often not deep enough and leads to 0-layer grasps. AutoBag, designed for thin plastic bags, is not effective on drawstring bags and fabric handbags. In contrast, SLIP-Bagging is effective on various bag materials. Check out the project website for videos and failure mode analysis.

D. *Single-Layer Grasping on Fabrics*

We test SLIP on other materials to evaluate its applicability to general single-layer grasping tasks. We consider 3 deformable objects: a blue piece of cloth folded twice into a square, a white dress, and a red hat (Fig. 2). The task goal is to grasp their top layer only. We apply our video classification model trained on bags to these objects without any finetuning. Table II shows the results. In each case, the model predicts accurately on a 0-layer grasp and 1-layer grasp, but less accurately on a 2-layer grasp, for which there are greater visual differences across objects. A failure mode associated with grasping a folded cloth is that the cloth has 4 layers. Grasping 1, 2, and 3 layers look visually similar, so the model would mistaken those 2- and 3-layer grasps as a 1-layer grasp. While the model accuracy is lower than that of bags the model is trained on, the SLIP success rate is fairly high. The robot starts from a grasp height higher than the surface height and gradually decreases its height, so it suffices for the model to accurately recognize a 1-layer grasp. We believe with some finetuning of the video classification model to bridge the gap between the visuals of the fabrics and bags, the success rate of SLIP can be further improved.

V. Conclusion and Future Work

In this paper, we propose SLIP: singulating layers using interactive perception. Experiments show that SLIP is successful on many materials and that SLIP-Bagging achieves significantly higher success rates over baselines for autonomous bagging. Future work can apply SLIP to related tasks such as packing and wrapping.

References

[1] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. S. Sukhatme, "Interactive perception: Leveraging action in perception and perception in action," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1273–1291, 2017.

[2] 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 *Int. Journal of Robotics Research (IJRR)*, 2018.

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

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

[5] L. Y. Chen, H. Huang, E. Novoseller, D. Seita, J. Ichnowski, M. Laskey, R. Cheng, T. Kollar, and K. Goldberg, "Efficiently Learning Single-Arm Fling Motions to Smooth Garments," in *Int. S. Robotics Research (ISRR)*, 2022.

[6] Y. Avigal, L. Berscheid, T. Asfour, T. Kröger, and K. Goldberg, "SpeedFolding: Learning Efficient Bimanual Folding of Garments," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2022.

[7] 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 *Proc. Robotics: Science and Systems (RSS)*, 2020.

[8] J. Qian, T. Weng, L. Zhang, B. Okorn, and D. Held, "Cloth Region Segmentation for Robust Grasp Selection," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2020.

[9] 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 *Int. S. Robotics Research (ISRR)*, 2019.

[10] H. Ha and S. Song, "FlingBot: The Unreasonable Effectiveness of Dynamic Manipulation for Cloth Unfolding," in *Conf. on Robot Learning (CoRL)*, 2021.

[11] T. Weng, S. Bajracharya, Y. Wang, K. Agrawal, and D. Held, "FabricFlowNet: Bimanual Cloth Manipulation with a Flow-based Policy," in *Conf. on Robot Learning (CoRL)*, 2021.

[12] X. Lin, Y. Wang, Z. Huang, and D. Held, "Learning Visible Connectivity Dynamics for Cloth Smoothing," in *Conf. on Robot Learning (CoRL)*, 2021.

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

[14] X. Ma, D. Hsu, and W. S. Lee, "Learning Latent Graph Dynamics for Deformable Object Manipulation," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2022.

[15] S. Tirumala, T. Weng, D. Seita, O. Kroemer, Z. Temel, and D. Held, "Learning to Singulate Layers of Cloth using Tactile Feedback," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2022.

[16] R. Bhirangi, T. Hellebrekers, C. Majidi, and A. Gupta, "ReSkin: versatile, replaceable, lasting tactile skins," in *Conf. on Robot Learning (CoRL)*, 2022.

[17] K. Manabe, X. Tong, and Y. Aiyama, "Single sheet separation method from piled fabrics using roller hand mechanism," in *2021 IEEE International Conference on Intelligence and Safety for Robotics (ISR)*, 2021, pp. 359–362.

[18] Y. Guo, X. Jiang, and Y. Liu, "Deformation Control of a Deformable Object Based on Visual and Tactile Feedback," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2021.

[19] S. Demura, K. Sano, W. Nakajima, K. Nagahama, K. Takeshita, and K. Yamazaki, "Picking up one of the folded and stacked towels by a single arm robot," in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2018, pp. 1551–1556.

[20] 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.

[21] H. Ichiwara, H. Ito, K. Yamamoto, H. Mori, and T. Ogata, "Contact-rich manipulation of a flexible object based on deep predictive learning using vision and tactility," in *2022 International Conference on Robotics and Automation (ICRA)*, IEEE, 2022, pp. 5375–5381.

[22] Z. Xu, C. Chi, B. Burchfiel, E. Cousineau, S. Feng, and S. Song, "DextAIRity: Deformable Manipulation Can be a Breeze," in *Proc. Robotics: Science and Systems (RSS)*, 2022.

[23] C. Gao, Z. Li, H. Gao, and F. Chen, "Iterative Interactive Modeling for Knotting Plastic Bags," in *Conf. on Robot Learning (CoRL)*, 2022.

[24] D. Seita, J. Kerr, J. Canny, and K. Goldberg, "Initial Results on Grasping and Lifting Physical Deformable Bags with a Bimanual Robot," in *IROS Workshop on Robotic Manipulation of Deformable Objects in Real-world Applications*, 2021.

[25] A. Bahety, S. Jain, H. Ha, N. Hager, B. Burchfiel, E. Cousineau, S. Feng, and S. Song, "Bag all you need: Learning a generalizable bagging strategy for heterogeneous objects," *arXiv preprint arXiv:2210.09997*, 2022.

[26] L. Y. Chen, B. Shi, D. Seita, R. Cheng, T. Kollar, D. Held, and K. Goldberg, "AutoBag: Learning to Open Plastic Bags and Insert Objects," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2023.

[27] C. Feichtenhofer, H. Fan, J. Malik, and K. He, "Slowfast networks for video recognition," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6202–6211.