Learning to Singulate Layers of Cloth using Tactile Feedback

Sashank Tirumala[1,*], Thomas Weng[1,*], Daniel Seita[1,*], Oliver Kroemer[1], Zeynep Temel[1], David Held[1]

Abstract— **Robotic manipulation of cloth has applications ranging from fabrics manufacturing to handling blankets and laundry. Cloth manipulation is challenging for robots largely due to their high degrees of freedom, complex dynamics, and severe self-occlusions when in folded or crumpled configurations. Prior work on robotic manipulation of cloth relies primarily on vision sensors alone, which may pose challenges for fine-grained manipulation tasks such as grasping a desired number of cloth layers from a stack of cloth. In this paper, we propose to use tactile sensing for cloth manipulation; we attach a tactile sensor (ReSkin) to one of the two fingertips of a Franka robot and train a classifier to determine whether the robot is grasping a specific number of cloth layers. During test-time experiments, the robot uses this classifier as part of its policy to grasp one or two cloth layers using tactile feedback to determine suitable grasping points. Experimental results over 180 physical trials suggest that the proposed method outperforms baselines that do not use tactile feedback and has better generalization to unseen cloth compared to methods that use image classifiers. Code, data, and videos are available at `https://sites.google.com/view/reskin-cloth`.**

I. INTRODUCTION

Cloth manipulation remains an active research area in robotics with significant real world applications, including folding laundry [8, 31], assistive dressing [10–12, 55], bed-making [41], and manufacturing fabrics [48]. Cloth manipulation is challenging because it is difficult to infer the complete configuration of the cloth from robot observations when the cloth is in a crumpled or folded state, due to the high degrees of freedom and self-occlusions [3, 39].

In light of these challenges, researchers have recently proposed numerous data-driven methods for canonical cloth manipulation tasks such as smoothing [42, 52] and folding [25, 33, 51]. While showing promising results, many prior works focus on top-down grasping of one cloth. Such grasping may be ineffective for manipulation tasks involving multiple cloths, such as picking a desired number of layers of a stack of cloth, because performance is extremely sensitive to the height of the gripper when it grasps. Indeed, a common failure case reported in prior work [13, 51] is picking the wrong number of layers. Yet, manipulating a specific number of cloth layers is common in daily life, such as in folding and unfolding tasks, or handling piles of stacked clothing in stores. How, then, can robots achieve accurate grasping of multiple layers of cloth?

Incorporating tactile sensing is an under-explored direction for deformable object manipulation. While there has

*Equal contribution.

[1]The Robotics Institute at Carnegie Mellon University

Correspondence: {`stirumal,tweng,dseita`}`@andrew.cmu.edu`

Fig. 1: We present a tactile-based cloth manipulation system. The robot utilizes a ReSkin [2] sensor attached to the lower one of its two fingertips, which is visualized in more detail in the upper right inset. We train a classifier to distinguish among grasping different numbers of cloth layers from tactile feedback (no images are provided as input). The robot then uses this classifier at test time to determine suitable grasping points for obtaining a desired number of cloth layers.

been recent work on optical-based tactile sensors such as GelSight [56] and DIGIT [24], these sensors have primarily been applied to cloth *perception* [28, 57] instead of cloth *manipulation*. Recent work on magnetometer-based sensors such as ReSkin [2] have benefits over optical sensors, such as lower-dimensional sensor readings, more direct measurements of normal and shear forces, and a compact form factor. However, research into the applications of *magnetometer-based* sensors for deformable object manipulation is currently limited.

In this paper, we study the application of magnetometer-based tactile sensing for deformable cloth manipulation. We focus on precisely grasping and lifting layers of stacked cloth; due to the flexibility of cloth and unpredictable crumpling behavior, this task is challenging while being a well-defined manipulation problem. Furthermore, precise grasping of layers of cloth is a prerequisite for many downstream manipulation tasks (e.g., folding cloth in half twice).

We present a robotic system consisting of a 7-DOF Franka arm, a mini-Delta gripper [32], and a Reskin [2] sensor on the gripper finger to perform precise cloth grasping (see Fig. 1). The system uses a tactile classifier as feedback for a grasping policy. We show that simple approaches to both classifying tactile data and incorporating feedback into the policy (e.g., as a termination condition) work surprisingly well.

This paper makes the following contributions:

1) A robot hardware system which incorporates ReSkin tactile sensors for cloth manipulation.
2) A training procedure for developing a classifier based on this hardware to use in a grasping policy.
3) Experiments showing success on the task of grasping a desired number of cloth layers.

II. RELATED WORK

Manipulation of deformable objects such as cloth has a long history in robotics; see Yin *et al.* [54] and Zhu *et al.* [59] for representative surveys.

A. Cloth Manipulation Policies

In early research on cloth manipulation, a common strategy was to utilize a bimanual robot to grip cloth in midair to smooth it using gravity. This standardizes the configuration of cloth and exposes its corners, which can then facilitate planning subsequent manipulation tasks such as smoothing and folding [8, 31]. Other researchers have relied on using geometric features of cloth, such as by fitting polygon contours to clothing [34]. While these works showed impressive results, such approaches may be time-consuming or require strong assumptions on cloth configurations.

With the rise of deep learning, researchers have recently employed data driven techniques to obtain large amounts of interaction data with cloth to learn manipulation policies using powerful function approximators, often with the help of simulators [27, 47]. These works tend to learn quasi-static pick-and-place policies, which allow the cloth to settle between robot actions [13, 17, 26, 29, 36, 42, 43, 51–53]. Other researchers have learned continuous servoing policies [33], dynamic policies [15] or have explored learning cloth manipulation from purely real world interaction [25].

In contrast to these works which employ vision-manipulation policies, we focus on tactile sensing for cloth grasping.

B. Grasping for Cloth Manipulation

Perhaps the most important part of cloth *manipulation* tasks is cloth *grasping*, since a suitable grasp is necessary for subsequent actions such as dragging or lifting. Defining and identifying ideal cloth grasps remains challenging and is the subject of extensive research [3]. Early cloth manipulation research focused on vertically smoothing via gravity. A common such grasping strategy to reliably standardize cloth was to hold it with one gripper while iteratively grasping the lowest hanging corner with the other gripper [4, 21, 22, 31].

Other cloth grasping techniques do not require assuming that the cloth is lifted in midair. For example Ramisa *et al.* [38] and Sun *et al.* [45] determine suitable grasping points for cloth on a flat table by detecting wrinkles and edges using depth and classical computer vision techniques. Other applications of cloth manipulation may utilize specialized gripper designs [23] or may simplify the process by assuming that cloth is gripped in advance of the task [18].

Recently, Qian *et al.* [37] study how to robustly grasp cloth using dense segmentation of images to distinguish between edges and interior creases. Their method involves a self-supervised labeling procedure and a sliding grasp. Nonetheless, robustly grasping cloth remains challenging, particularly when the goal is to generalize to a wide variety of types and configurations of cloth. Prior work has reported that a typical failure cases is grasping the wrong number of cloth layers, particularly when unfolding [33, 42, 51]. Furthermore, many works employ heuristics such as hand-tuning the gripper design and grasp depth [13].

Prior work has also investigated learning to grasp one cloth from a stack using grasping and scooping actions from vision input only [5], as well as designing a robot system to turn a single book page using vision and force sensing [14]. In this work, we consider the novel task of grasping more than one cloth layer, and show the benefits of tactile sensing without requiring vision.

C. Tactile Sensor Hardware

The robotics community has developed numerous tactile sensors. Examples of sensors include the class of optical-based sensors such as GelSight [56], GelSlim [7], and DIGIT [24], which have been used for cloth perception. For example, Yuan *et al.* [57] demonstrate how to use active learning to identify where to grasp a garment to classify it among several categories of clothing, and Yuan *et al.* [58] and Luo *et al.* [28] study how to combine tactile information with vision to infer properties of cloth. Similarly, Khan *et al.* [19, 20] use tactile information to classify single fabrics into different material and clothing types using piezoelectric pressure sensors [6]. In addition, Drigalski *et al.* [9] use tactile sensing and rubbing behavior to classify textiles into one of 18 categories, and can distinguish between 1 or 2 layers. In contrast to these works, we focus on cloth manipulation instead of pure perception, and additionally focus on fine-grained manipulation which may be challenging with sensors such as the GelSight due to their relatively large size.

Other types of robotic tactile sensors include BioTac [46] and stretchable piezoresistive [1] sensors. These sensors are durable, but remain expensive and may not be easily replaceable. Research teams have also explored tactile sensing using a customized force-torque sensor [35] for manipulating deformable blocks [40]. To our knowledge, none of these sensors have been used for cloth manipulation tasks.

Recently, Bhirangi *et al.* [2] proposed the ReSkin, a class of magnetic sensors which is well suited to machine learning due to its low cost, durability, form factor, ability to cover a large area, and ease of replacement. The researchers demonstrate ReSkin on robotic grasping tasks that involve handling delicate objects such as blueberries and grapes. Due to these advantages and existing applications, we use the ReSkin for novel cloth manipulation tasks that involve fine-grained manipulation of cloth layers.
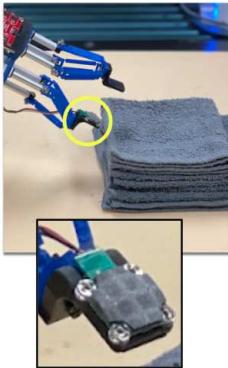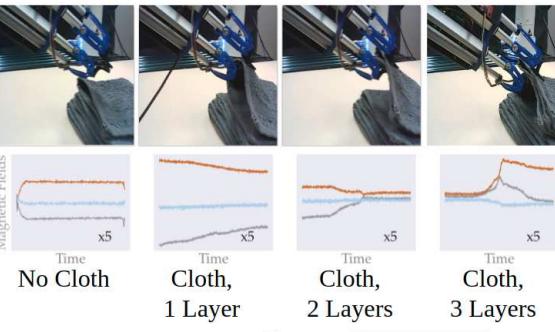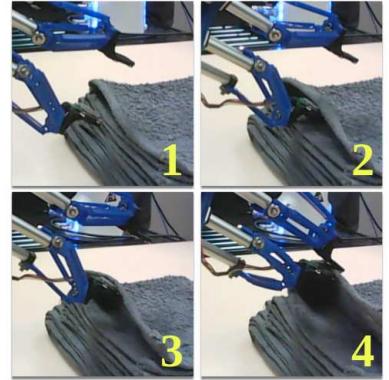
Fig. 2: The proposed tactile-based cloth manipulation pipeline. A 7-DOF Franka robot uses a mini-Delta [32] gripper with two finger tips, the lower one of which has a ReSkin [2] sensor (see yellow circle and zoomed-in inset). Using this gripper, we collect tactile data from the ReSkin by performing grasps of different categories: grasping nothing, or pinching 1, 2, or 3 cloth layers (see Fig. 3 for more examples). The graphs above visualize the tactile time series data. At test time, the robot uses the trained tactile-based classifier to grasp a desired number of cloth layers.

III. PROBLEM STATEMENT

We study the task of grasping a desired number of layers from a stack of cloths. Given a set of at least 3 cloth layers stacked on each other, the goal is to grasp the top $k \in \{1, 2\}$ cloth layers. For each *trial* (a given instance of the task), we specify a target value for k. We assume a robot has a two-finger gripper where one of the gripper tips is equipped with a *tactile sensor*. We assume each trial begins with the robot's tactile sensor facing a set of edges from a stack of cloth layers, as shown in Fig. 1. A trial is a *success* when the robot grasps exactly k cloth layers and is able to lift its gripper upwards by 4 cm while preserving its grasp of the k layers.

IV. METHOD

This proposed system for tactile sensing involves designing hardware with tactile data (Sec. IV-A), training a classifier to distinguish grasping cloth layers (Sec. IV-B), then using this classifier for a grasping policy (Sec. IV-C). See Fig. 2 for the overall pipeline.

A. Hardware

The proposed system uses a ReSkin [2] sensor, which comprises of a soft magnetized skin and a circuit board with a 5-magnetometer array (see bottom-left inset of Fig. 2). The board sits beneath the skin, and any deformations caused by normal/shear forces are read via distortions in magnetic fields. For each of the 5 magnetometers, 3 magnetic flux values $\langle B_X, B_Y, B_Z \rangle$ are reported, corresponding to flux in the X-, Y-, and Z- magnetometer coordinate axes. Concatenating these values for a single time step t results in a 15-dim vector $\mathbf{B}^{(t)} \in \mathbb{R}^{15}$. ReSkin publishes these values at up to 400 Hz.

We attach ReSkin to a finger on a mini-Delta gripper [32]. We use the mini-Delta largely due to its length and form factor, since it facilitates grasping a layered stack of cloth

folds by approaching it from the side, instead of top-down. The mini-Delta has 3 DOFs for each finger, and is compliant due to the 3D-printed soft links (blue component in Fig. 1), though in this work, we do not rely on the additional DOFs or on compliance. Our contribution centers on tactile sensing for grasping of cloth layers, and we leave investigation of exploiting additional DOFs and compliance for future work. The gripper and attached sensor are mounted on a 7-DOF Franka robot.

B. Grasp Classifier Training

We train a classifier to predict the number of cloth layers grasped to use as part of the grasp policy (Sec. IV-C). The classifier takes as input a tactile reading from a single time step $\mathbf{B}^{(t)}$. While analyzing sensory data across a time series seems natural for the tactile modality, we find that predictions based on point estimates are surprisingly effective, as we later show in Sec. VI. We do not take proprioceptive data as input, as this modality is not currently available with the mini-Delta gripper: the compliant links can bend from their commanded position given sufficiently high external force, and estimating proprioception for these types of compliant links is an area of active research.

The classifier uses the tactile readings to predict how the gripper is interacting with the cloth, among 4 classes: (1) pinching with no cloth between the fingers, (2) pinching 1 cloth layer, (3) pinching 2 cloth layers, and (4) pinching 3 cloth layers. We limit the number of cloth layers under consideration to 3 to make classification tractable, while also allowing feedback-based policies to recover if they overshoot when grasping two layers. We leave classifying an arbitrary number of layers to future work.

We collect training data in the real world for the classifier due to the lack of a suitable simulator.[1] We define a single

---

[1] While there has been progress in developing high-fidelity simulators for tactile sensors [44, 50] and for deformables [27, 30], simulating both is challenging and to our knowledge has not yet been shown.

| No Cloth Layers	Cloth, 1 Layer	Cloth, 2 Layers	Cloth, 3 Layers
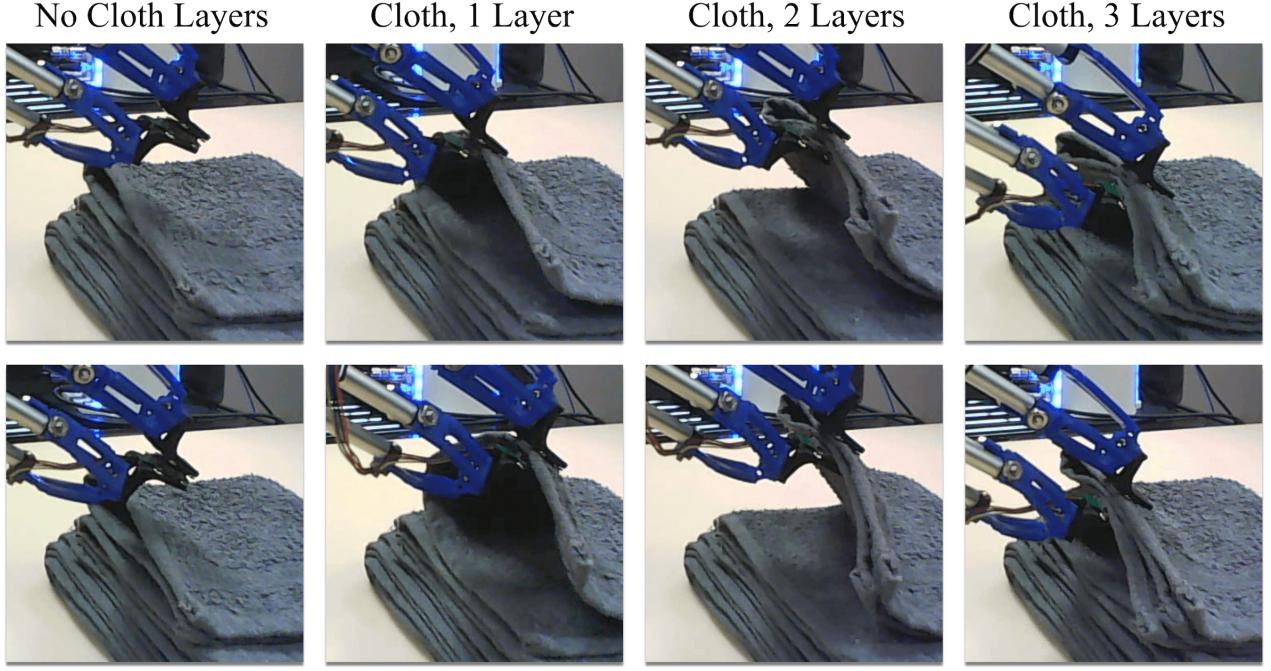
Fig. 3: Examples of collecting data for tactile-based classification, with the ReSkin attached to the bottom gripper finger tip. From left to right, we show two examples each of collecting data with (1) contact, but without cloth, (2) 1 cloth layer, (3) 2 cloth layers, and (4) 3 cloth layers. The classifier only takes as input the data collected from the ReSkin sensor $\mathbf{B}^{(t)}$ at any give time step. The images above are collected with a webcam and are used both to visualize the tactile data collection, and also are the RGB inputs to the image classifiers that we train as baselines for comparison. See Sec. IV and Sec. V for further details.

"training episode" as the process of getting a set of tactile data from one grasp. First, a human stacks several layers of cloth with edges facing the gripper. The height at which the robot approaches the cloth is uniformly sampled per attempt within a $\pm 2\,\mathrm{mm}$ range to collect a variety of grasps. The robot then approaches the cloth, closes its fingertips to grasp firmly, records ReSkin data during the grasp, then releases. Each training episode lasts roughly 5 seconds and produces approximately 350 sensor readings of 15 values each (3 per magnetometer). We visually inspect videos from the recorded data to determine the number of grasped cloth layers, and we label all points from a training episode with the same label, speeding up annotation time and effort. See Fig. 3 for example visualizations of training episodes for all classes.

We then use this collected data to train a classifier to distinguish the numbers of layers grasped from the tactile readings. We experimented with various types of classifiers, including k-Nearest Neighbor (kNN), SVM, Logistic Regression, and Random Forests, and we found the performance to be fairly similar across classifiers. For simplicity, we use a k-Nearest Neighbor (kNN) classifier with $k = 10$ neighbors.

C. Proposed Grasp Policy

Next we describe how we use the above trained classifier to enable the robot to grasp the desired number of layers. We divide the robot trajectory into three parts. First, the gripper moves vertically down by a distance d_{vert}, then horizontally towards the cloth stack by a distance d_{slide}, then lifts up by a distance d_{lift}, then closes its gripper tips (see

Fig. 4 for a visualization). At this point, we record tactile data and classify the number of layers that are grasped. If the predicted number of grasped layers (according to the classifier) matches the target number of grasped layers, it lifts the gripper further by 4 cm to indicate the end of the trial; otherwise, it resets the gripper back to the starting position and tries again (see below for details). The values of d_{slide} and d_{lift} are tuned and fixed ahead of time by a human operator, while d_{vert} is determined by the policy, as explained below.

The grasping policy uses the output of the grasp classifier (Sec. IV-B) to determine the vertical distance that the gripper lowers before grasping, d_{vert}. For a target number of layers k to grasp, the robot begins at some height with the grippers open, moves towards the cloth stack, and attempts a grasp. If the grasp classifier determines that it has not grasped the correct number of layers, then the robot releases, moves back, and adjusts the gripper height (d_{vert}). If the classifier predicts that it has grasped too many layers, d_{vert} is decremented by a small value to decrease the grasp height; if it has grasped too few, d_{vert} is incremented by a small value. The policy continues until either the classifier determines that it has grasped the desired number of layers and ends the trial, or until the maximum number of grasp attempts is reached.

During each grasp attempt on the physical system, the classifier starts predicting the class once the gripper closes, and stops predicting after the robot lifts by d_{vert}. This results in a set of about 160 separate predictions. We use the mode of all the predictions as the final prediction to determine
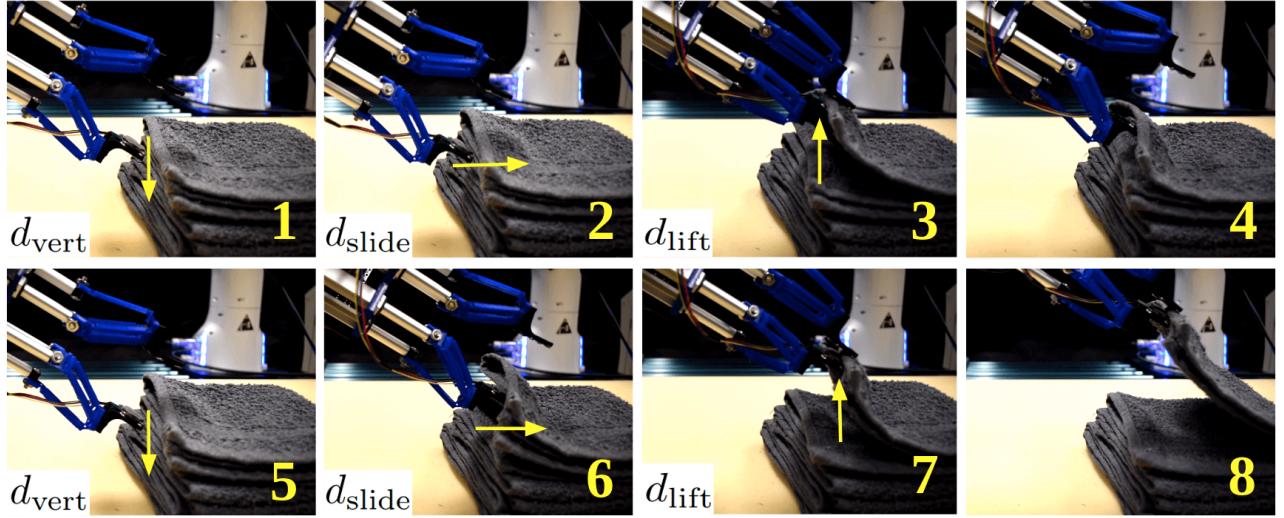
Fig. 4: The proposed grasp policy parameterization (described in Sec. IV-C), visualized with a frame-by-frame overview of an example trial from the experiments. Each row, consisting of four frames, shows one action. The first part of an action (shown in frames 1 and 5) adjusts the initial gripper height by d_{vert}, possibly from prior tactile feedback. The second part of an action (shown in frames 2 and 6) moves towards the cloth stack by some distance d_{slide}. Then, the third part (frames 3 and 7) lifts upwards by d_{lift} and closes the grippers. At this point, the robot queries the classifier and may decide to release and re-attempt the grasp (frames 4 and 5) or the robot concludes that it has grasped the correct number of layers and further lifts the cloth to end the trial (frame 8).

whether to raise or lower the grasp height.

V. PHYSICAL EXPERIMENTS

We evaluate the methods using the physical system described in Sec. IV-A. The experiments are designed to answer the following questions:

- Can magnetometer-based tactile sensing with ReSkin sensors provide sufficient information about grasping a target number of cloth layers?
- What are the benefits of the proposed method that uses tactile-based feedback to adjust the gripper height?
- Can a classifier trained on tactile feedback generalize to different cloths?

A. Experiment Protocol

We train our tactile classifier on a gray cloth; we then evaluate our system on the gray training cloth and on two other unseen cloths to measure the generalization of our method to new cloths (see Fig. 6). We use the same training data from the gray cloth for all of the tactile-based method variations described in Sec. V-B. The tactile data collection results in a total of 18,838 such $\mathbf{B}^{(t)}$ readings. We train a classifier on 95 % of the training episodes (to allow for a small validation set). We normalize the tactile data so that each of the 15 features has mean 0 and variance 1 in the training set.

We perform two sets of experiments, in which we set the desired number of cloth layers to grasp as one layer and two layers, respectively. Each *trial* begins with a human arranging folded cloths on the workspace with edges exposed and facing the robot gripper. The number of folded cloths is the same across trials, but variations in the depth of the layers up to 1.5 mm can occur due to slight differences in the initial cloth configuration. We initialize the robot's gripper at an

angle (30°) which increases the likelihood that a horizontal motion can slide the robot finger tips in between layers of cloth.

Each experiment set consists of comparing several grasping methods (see Sec. V-B). When running experiments, we randomly sample the method to run in the given trial *after* the cloths have been set, to reduce potential human bias in the data initialization. The robot employs the selected method to grasp the appropriate number of cloth layers. The robot is allowed up to $T = 10$ actions per trial, though it can terminate earlier if the classifier estimates that it has grasped the appropriate number of layers. Upon termination, the robot lifts the gripper by 4 cm and a human measures this as a success if the correct number of layers are still grasped. All other cases result in the trial as a failure.

We categorize failures into two types, *prediction* and *grasping* failures. Prediction failures are a result of mispredictions by the trained classifier, where it either: (1) incorrectly predicts that the robot has grasped the desired number of layers and terminates the trial prematurely, or (2) the classifier incorrectly predicts that the robot has grasped the wrong number of layers, causing unnecessary regrasps and leading to the robot reaching the max number of attempts for the trial. Grasping failures are due to either failing to grasp the desired number of layers at the last time step in a trial, or failing to robustly grasp the cloth, such that cloth layers slip out of the robot's control when lifting (see Fig. 5).

B. Methods and Baselines

We evaluate the following methods for grasping 1 and 2 cloth layers:

1) **Fixed-Open-Loop**: Initialize the gripper at a fixed height, manually tuned for grasping 1 or 2 cloth layers:
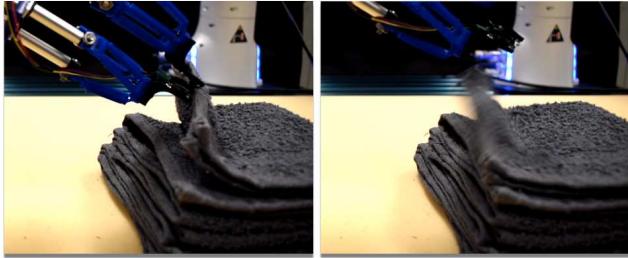
Fig. 5: An example grasping failure case of the task. Due to an insufficiently robust grasp when lifting (left), the layers may slip out of the robot's control during the lifting portion (right).

Gray Towel   White Towel   Patterned Cloth

Fig. 6: The cloths we use for experiments. We use the gray towel (left) for training, and test on all 3 cloths for evaluation. The white towel and patterned cloth test generalizing to novel cloths. The cloths have thicknesses between 3-5 mm and variation in surface texture and stiffness.

$d_{\text{vert}}^{(1)}$ and $d_{\text{vert}}^{(2)}$ respectively. This method terminates after a single trial as it has no access to feedback.

2) **Random-Tactile**: Randomly try different gripper heights within the range $\left[d_{\text{vert}}^{(2)} - 2\,\text{mm}, d_{\text{vert}}^{(1)} + 2\,\text{mm}\right]$ until the tactile classifier determines that the correct number of layers have been grasped.

3) **Random-Image**: Same as Random-Tactile, but uses an image classifier (instead of a tactile classifier) to determine when the correct number of layers has been grasped. The image classifier is an 18-layer ResNet [16] pre-trained on ImageNet and finetuned on the images (see Figure 3) from the same training episodes used to train the tactile classifier.

4) **Feedback-Image**: Same as Feedback-Tactile (our method, below) except with the image classifier.

5) **(Ours) Feedback-Tactile**: Initialize the gripper height to $d_{\text{vert}}^{(1)} + 2\,\text{mm}$; use the grasp policy described in Sec. IV-C to adjust the height per grasp ($\pm 2\,\text{mm}$) based on the tactile classifier predictions.

VI. RESULTS

We first present results from training a classifier on ReSkin data followed by physical experiment results in which we run 10 trials for each method and condition.

A. The Tactile Classifier

To better understand the kNN performance, we perform 100 folds of cross-validation and average the validation performance. Each entire training episode is assigned to either the training or validation set.

Table I demonstrates the average normalized confusion matrix obtained from these 100 cross-validation runs, and

Class \ Prediction	0	1	2	3
0 (0 Layers)	1.000	0.000	0.000	0.000
1 (1 Layer)	0.000	0.999	0.000	0.001
2 (2 Layers)	0.030	0.003	0.866	0.100
3 (3 Layers)	0.128	0.256	0.138	0.478

TABLE I: The average normalized confusion matrix from the cross-validation training results for the k-nearest neighbor classifier we use for tactile-based experiments.

also reports the average per-class accuracy. We also computed the average balanced accuracy metric [49] to consider the data imbalance and obtain 0.84 ± 0.06. Inspecting the confusion matrix, we find that the tactile classifier can classify classes 0 (i.e., pinching with no cloth between the fingers) and 1 (i.e., pinching 1 cloth layer) with extremely high effectiveness. Results for classes 2 and 3 suggest that identifying 2 and 3 cloth layers is more challenging.

B. Grasping 1 Cloth Layer

In the first set of physical experiments, we report the success and failures of methods on grasping and lifting the top layer of cloth from a stack. See Table II for results. Our method, Feedback-Tactile, succeeds at grasping one layer of cloth in all 10 trials, whereas all competing ablations have lower success rates. Methods with the tactile classifier outperform those using the image classifier, with most failures attributed to mis-prediction rather than poor grasping.

The fixed-height open loop method (Fixed-Open-Loop) poorly handles variations in the initial cloth configuration. There can be up to 1.5 mm variation in the height of the cloth stack based on how they are placed at the start of the trial, which can lead to failures in the open loop grasping method. Both random grasping approaches, Random-Image (5/10) and Random-Tactile (6/10) have lower success rates compared to using feedback-based height adjustment with Feedback-Image (8/10) and Feedback-Tactile (10/10).

For testing generalization, Feedback-Tactile significantly outperforms Feedback-Image on the white towel and patterned cloth. Feedback-Tactile obtains 8/10 and 7/10 success rates for the white towel and patterned cloth, respectively, while Feedback-Image only succeeds in 3/10 and 2/10 trials.

We have analyzed the failure types of each method in Table II. Grasping failures are rare for most methods on 1-layer grasping; grasping failures can occur if the robot does not robustly grip the cloth, and cloth slips out of the grasp when the robot lifts it (see Fig. 5). Our method (Feedback-Tactile), also has few prediction failures when generalizing to unseen cloths compared to Feedback-Image.

C. Grasping 2 Cloth Layers

In the next set of experiments, we evaluate grasping and lifting the top two layers of cloth. The results in Table III suggest that the methods achieve success rates similar to 1-layer grasping (Table II) for the gray towel, but performance is lower on the unseen cloths. While Feedback-Image performs slightly better than Feedback-Tactile on the gray towel, Feedback-Tactile performs slightly better on unseen cloths.

Cloth Type	Method	Success Rate ↑	Prediction Failure	Grasp Failure	Attempts ↓
Gray Towel (Train)	Fixed-Open-Loop	6/10	-	4/10	1 (fixed)
	Random-Image	5/10	5/10	0/10	1.8±0.7
	Random-Tactile	6/10	3/10	1/10	4.8±2.8
	Feedback-Image	8/10	2/10	0/10	2.3±0.8
	Feedback-Tactile	**10/10**	0/10	0/10	3.1±1.0
White Towel (Generalization)	Feedback-Image	3/10	5/10	2/10	1.6±0.5
	Feedback-Tactile	**8/10**	0/10	2/10	2.3±0.8
Patterned Cloth (Generalization)	Feedback-Image	2/10	8/10	0/10	5.1±4.3
	Feedback-Tactile	**7/10**	2/10	1/10	4.6±3.2

TABLE II: Results for the first set of physical experiments described in Sec. VI-B with grasping at 1 cloth layer. We run all methods for 10 trials each and report the success rate, the failure types (grasping and prediction), and the average number of grasp attempts per trial.

Cloth Type	Method	Success Rate ↑	Prediction Failure	Grasp Failure	Attempts ↓
Gray Towel (Train)	Fixed-Open-Loop	7/10	-	3/10	1 (fixed)
	Random-Image	6/10	1/10	3/10	5.3±3.0
	Random-Tactile	4/10	4/10	2/10	6.0±3.0
	Feedback-Image	**9/10**	0/10	1/10	4.7±0.9
	Feedback-Tactile	7/10	1/10	2/10	6.4±2.6
White Towel (Generalization)	Feedback-Image	0/10	8/10	2/10	9.2±1.8
	Feedback-Tactile	**4/10**	2/10	4/10	5.0±3.4
Patterned Cloth (Generalization)	Feedback-Image	0/10	10/10	0/10	10.0±0.0
	Feedback-Tactile	**1/10**	3/10	6/10	6.4±3.6

TABLE III: Experimental results for grasping at the top 2 cloth layers as described in Sec. VI-C. Besides the change of 1 to 2 layers, the results are formatted in the same way as in Table II.
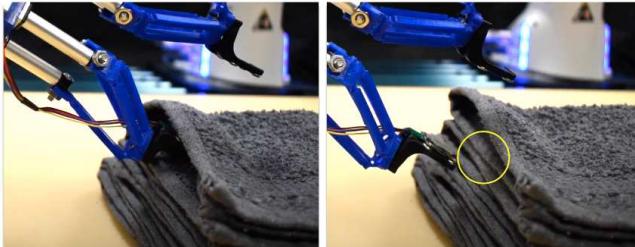
Fig. 7: A qualitative example of how the task is challenging, particularly with grasping two layers. Because of the horizontal motion of the gripper, layers of cloth can be pushed apart (left), creating air pockets between the top and second layer after the action has finished (right). This gap makes it easier to grasp the top layer but harder to grasp the top *two* layers, due to a smaller gap between the second and third layers (see overlaid yellow circle).

Table III shows that both prediction and grasping failures lead to errors for our method (Feedback-Tactile), though grasping failures are more common (accounting for 2/3 of our total failures). The higher incidence of grasp failures by our method in this experiment suggests that 2-layer grasping is more difficult than 1-layer grasping. Fig. 7 highlights some challenges with grasping two layers; for example, we observe that failures tend to occur due to crumpling the fabric when attempting to grasp 2 layers. Furthermore, the top layer of cloth can push downwards on the layer below it, which reduces the gap between the second and third layers; this reduced gap can make it difficult to grasp 2 layers. These observations and results suggest that further innovation on grasp policies may be necessary to improve 2-layer grasp performance on unseen cloths.

VII. CONCLUSION

In this paper, we present a robotic system that uses magnetometer-based tactile sensing for precisely grasping layers of cloth. We train a classifier on tactile sensor readings from a ReSkin sensor. At test time, the classifier determines the number of layers of cloth grasped, which informs the policy to adjust the height of the gripper for subsequent grasp attempts. The system obtains strong results with grasping the top 1 or 2 cloth layers out of a stack of cloth, and generalizes to unseen cloth. We hope this work motivates future research on tactile-based robotic policies that can manipulate a wide variety of complex objects.

REFERENCES

[1] T. Bhattacharjee, A. Jain, S. Vaish, M. D. Killpack, and C. C. Kemp, "Tactile Sensing over Articulated Joints With Stretchable Sensors," in *World Haptics Conference (WHC)*, 2013.

[2] R. Bhirangi, T. Hellebrekers, C. Majidi, and A. Gupta, "ReSkin: versatile, replaceable, lasting tactile skins," in *Conference on Robot Learning (CoRL)*, 2021.

[3] J. Borràs, G. Alenyà, and C. Torras, "A grasping-centered analysis for cloth manipulation," *IEEE Transactions on Robotics*, vol. 36, no. 3, pp. 924–936, 2020.

[4] M. Cusumano-Towner, A. Singh, S. Miller, J. F. O'Brien, and P. Abbeel, "Bringing Clothing Into Desired Configurations with Limited Perception," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2011.

[5] 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 *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2018.

[6] S. Denei, P. Maiolino, E. Baglini, and G. Cannata, "On the Development of a Tactile Sensor for Fabric Manipulation and Classification for Industrial Applications," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[7] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez, "GelSlim: A High-Resolution, Compact, Robust, and Calibrated Tactile-sensing Finger," in *IEEE Robotics and Automation Letters (RA-L)*, 2018.

[8] A. Doumanoglou, A. Kargakos, T.-K. Kim, and S. Malassiotis, "Autonomous Active Recognition and Unfolding of Clothes Using Random Decision Forests and Probabilistic Planning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

[9] F. von Drigalski, M. Gall, S.-G. Cho, M. Ding, J. Takamatsu, T. Ogasawara, and T. Asfour, "Textile Identification Using Fingertip Motion and 3D Force Sensors in an Open-source Gripper," in *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2017.

[10] 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.

[11] 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.

[12] Z. Erickson, V. Gangaram, A. Kapusta, C. K. Liu, and C. C. Kemp, "Assistive Gym: A Physics Simulation Framework for Assistive Robotics," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[13] 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.

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

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

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[17] 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.

[18] B. Jia, Z. Hu, J. Pan, and D. Manocha, "Manipulating Highly Deformable Materials Using a Visual Feedback Dictionary," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[19] A. A. Khan, M. H. Tanveer, T. Rasheed, and A. A. Ajmal, "Tactile Discrimination of Fabrics Using Machine Learning Techniques," in *IEEE 3rd International Conference on Engineering Technologies and Social Sciences (ICETSS)*, 2017.

[20] A. Khan, M. Khosravi, S. Denei, P. Maiolino, W. Kasprzak, F. Mastrogiovanni, and G. Cannata, "A Tactile-based Fabric Learning and Classification Architecture," in *IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, 2016.

[21] Y. Kita, T. Ueshiba, E. S. Neo, and N. Kita, "A Method For Handling a Specific Part of Clothing by Dual Arms," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2009.

[22] ——, "Clothes State Recognition Using 3D Observed Data," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2009.

[23] P. N. Koustoumpardis, K. X. Nastos, and N. A. Aspragathos, "Underactuated 3-finger Robotic Gripper for Grasping Fabrics," in *International Conference on Robotics in Alpe-Adria-Danube Region (RAAD)*, 2014.

[24] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, D. Jayaraman, and R. Calandra, "DIGIT: A Novel Design for a Low-Cost Compact High-Resolution Tactile Sensor with Application to In-Hand Manipulation," in *IEEE Robotics and Automation Letters (RA-L)*, 2020.

[25] R. Lee, D. Ward, A. Cosgun, V. Dasagi, P. Corke, and J. Leitner, "Learning Arbitrary-Goal Fabric Folding with One Hour of Real Robot Experience," in *Conference on Robot Learning (CoRL)*, 2020.

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

[27] X. Lin, Y. Wang, J. Olkin, and D. Held, "SoftGym: Benchmarking Deep Reinforcement Learning for Deformable Object Manipulation," in *Conference on Robot Learning (CoRL)*, 2020.

[28] S. Luo, W. Yuan, E. Adelson, A. G. Cohn, and R. Fuentes, "Vitac: Feature Sharing Between Vision and Tactile Sensing for Cloth Texture Recognition," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[29] X. Ma, D. Hsu, and W. S. Lee, "Learning Latent Graph Dynamics for Deformable Object Manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

[30] M. Macklin, M. Muller, N. Chentanez, and T.-Y. Kim, "Unified Particle Physics for Real-Time Applications," *ACM Trans. Graph.*, vol. 33, no. 4, Jul. 2014.

[31] 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.

[32] P. Mannam, A. Rudich, K. Zhang, M. Veloso, O. Kroemer, and F. Z. Temel, "A Low-Cost Compliant Gripper Using Cooperative Mini-Delta Robots for Dexterous Manipulation," in *Robotics: Science and Systems (RSS)*, 2021.

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

[34] S. Miller, J. van den Berg, M. Fritz, T. Darrell, K. Goldberg, and P. Abbeel, "A Geometric Approach to Robotic Laundry Folding," in *International Journal of Robotics Research (IJRR)*, 2012.

[35] K. Mohy el Dine, J. Sanchez, J. A. Corrales, Y. Mezouar, and J.-C. Fauroux, "Force-torque Sensor Disturbance Observer Using Deep Learning," in *International Symposium on Experimental Robotics (ISER)*, 2018.

[36] K. Puthuveetil, C. C. Kemp, and Z. Erickson, "Bodies Uncovered: Learning to Manipulate Real Blankets Around People via Physics Simulations," in *IEEE Robotics and Automation Letters (RA-L)*, 2022.

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

[38] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras, "Using Depth and Appearance Features for Informed Robot Grasping of Highly Wrinkled Clothes," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012.

[39] 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.

[40] J. Sanchez, K. M. E. Dine, J. A. Corrales, B.-C. Bouzgarrou, and Y. Mezouar, "Blind Manipulation of Deformable Objects Based on Force Sensing and Finite Element Modeling," in *Frontiers in Robotics and AI*, 2020.

[41] 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.

[42] 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.

[43] 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.

[44] Z. Si and W. Yuan, "Taxim: An example-based Simulation Model for GelSight Tactile Sensors," in *IEEE Robotics and Automation Letters (RA-L)*, 2022.

[45] L. Sun, G. Aragon-Camarasa, S. Rogers, and J. P. Siebert, "Accurate Garment Surface Analysis using an Active Stereo Robot Head with Application to Dual-Arm Flattening," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[46] B. Sundaralingam, A. Lambert, A. Handa, B. Boots, T. Hermans, S. Birchfield, N. Ratliff, and D. Fox, "Robust Learning of Tactile Force Estimation through Robot Interaction," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.

[47] E. Todorov, T. Erez, and Y. Tassa, "MuJoCo: A Physics Engine for Model-Based Control," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012.

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

[49] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[50] S. Wang, M. Lambeta, P.-W. Chou, and R. Calandra, "TACTO: A Fast, Flexible, and Open-source Simulator for High-Resolution Vision-based Tactile Sensors," in *IEEE Robotics and Automation Letters (RA-L)*, 2022.

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

[52] 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.

[53] 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.

[54] H. Yin, A. Varava, and D. Kragic, "Modeling, Learning, Perception, and Control Methods for Deformable Object Manipulation," *Science Robotics*, vol. 6, no. 54, 2021.

[55] W. Yu, A. Kapusta, J. Tan, C. C. Kemp, G. Turk, and C. K. Liu, "Haptic Simulation for Robot-Assisted Dressing," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[56] W. Yuan, S. Dong, and E. H. Adelson, "GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force," *Sensors*, vol. 17, no. 12, 2017.

[57] 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.

[58] W. Yuan, S. Wang, S. Dong, and E. Adelson, "Connecting Look and Feel: Associating the Visual and Tactile Properties of Physical Materials," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[59] J. Zhu, A. Cherubini, C. Dune, D. Navarro-Alarcon, F. Alambeigi, D. Berenson, F. Ficuciello, K. Harada, J. Kober, X. Li, *et al.*, "Challenges and outlook in robotic manipulation of deformable objects," *IEEE Robotics and Automation Magazine*, 2021.

In Section IV-B, we describe the real-world procedure for collecting a training dataset of ReSkin sensor readings. Each training episode lasts for about 5 seconds to complete the approach, grasp, and release cycle. The grasp stage lasts for 1 second, resulting in approximately 350 sensor readings of 15 values each (3 per magnetometer). We filter out training episodes where no cloth was grasped. Combining all episodes into a dataset results in a total of 18,838 ReSkin tactile sensor readings, each of which is 15-D. The readings in the dataset are separated into training and validation sets, grouped by episode to prevent data leakage. Thus, all time steps in one episode are either all in training or all in validation. We train classifiers taking individual readings as input and found that this input outperformed our baselines on the cloth singulation task; classification on time-series inputs are a promising area of future work.

In Section V-B, we introduce the **Random-Image** and **Feedback-Image** baseline methods. These use images instead of tactile data, but the objective is the same: to predict the number of cloth layers grasped. When collecting tactile-based data, we simultaneously collect image data, so the data collection time for image data is the same as the time that was spent on collecting the Reskin data. We mount a webcam approximately $30\,\text{cm}$ away from the starting position of the ReSkin, which queries images of the robot when it is collecting data. See Fig. 3 for examples of what the RGB images look like.

The image classifier is an 18-layer ResNet [16]. We use an off-the-shelf ResNet-18 from PyTorch which has been pre-trained on ImageNet, and change the last layer to output 4 dimensions instead of 1000. This results in a total of 11,178,564 trainable parameters. The RGB input images are first cropped to 360×360 such that the ReSkin sensor is located roughly in the middle, then images are further resized to 224×224 before being passed as input to the ResNet.

There is about 7-8X more tactile data compared to image-based data because the ReSkin can query data much faster than the webcam. This may be an inherent advantage of the ReSkin sensor over most commercial webcams. To strengthen the image-based baseline, we use data augmentation (which the tactile classifiers do not use). In training, we employ random crops by selecting random 224×224 crops within the 360×360 image. We also use random horizontal flips. At test time, we only use center crops for consistency. We train for 30 epochs, use a batch size of 64, and optimize using Adam with learning rate 0.001.