Tracking cloth deformation: a novel dataset for closing the sim-to-real gap for robotic cloth manipulation learning

Franco Coltraro[1,2], Júlia Borràs[1], Maria Alberich-Carramiñana[1,2] and Carme Torras[1]

Abstract

Robotic learning for deformable object manipulation –such as textiles– is often done in simulation due to the current limitation of perception methods to understand cloth's deformation. For this reason, the robotics community is always on the search for more realistic simulators to reduce as much as possible the sim-to-real gap, which is still quite large specially when dynamic motions are applied. We present a cloth dataset consisting of 120 high-quality recordings of several textiles during dynamic motions. Using a *Motion Capture System*, we record the location of key-points on the cloth surface of four types of fabrics (cotton, denim, wool and polyester) of two sizes and at different speeds. The scenarios considered are all dynamic and involve rapid shaking and twisting of the textiles, collisions with frictional objects, strong hits with a long and thin rigid object and even self-collisions. We explain in detail the scenarios considered, the collected data and how to read it and use it. In addition, we propose a metric to use the dataset as a benchmark to quantify the sim-to-real gap of any cloth simulator. Finally, we show that the recorded trajectories can be directly executed by a robotic arm, enabling learning by demonstration and other imitation learning techniques.

Dataset: https://doi.org/10.5281/zenodo.14644526

Video: https://fcoltraro.github.io/projects/dataset/

Keywords

cloth manipulation, real datasets, robotic learning, motion capture, cloth simulation, sim-to-real gap

Introduction

The recent surge in the successful use of large foundation models across various application domains is primarily due to the availability of vasts amounts of data to train general enough systems. However, this success has not yet been replicated in the robotic manipulation domain, mainly because of the challenges in gathering enough data. One possible way of obtaining enough data is through the use of simulators if they become sufficiently good. Bridging the sim-to-real gap, which can be narrowed to the development of better simulators is therefore a crucial step. Indeed, successful tasks have been learned using physics simulations for rigid interactions such as feasible grasps Eppner et al. (2019) or in-hand manipulation Handa et al. (2023). However, most simulators for highly-deformable objects like textiles are still not realistic enough, specially when recreating dynamic interactions. This is partly because most simulators originate from the graphics domain which prioritizes visual appearance over physical realism

Blanco-Mulero et al. (2024). Adding to the complexity, the possibility of learning from real interactions is hindered by the limited ability of current perception methods to understand cloth deformation, due to large self-occlusions and complex shape estimators even using depth cameras or 3D scanners. As a result, there are not many cloth datasets where real tracking of the cloth deformation is captured, specially during dynamic motions. Meaning that, with current data, an accurate measure of the sim-to-real gap of existing simulators is quite limited, hindering their progress.

The dataset presented in this publication was collected originally to test a simulator designed for

[1]Institut de Robòtica i Informàtica Industrial, CSIC-UPC.
[2]Departament de Matemàtiques and Institut de Matemàtiques de la UPC-BarcelonaTech (IMTech), Universitat Politècnica de Catalunya · BarcelonaTech.

Corresponding author:
Franco Coltraro, Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas, 4-6, 08028, Barcelona, Spain.
Email: franco.coltraro@upc.edu

robotic cloth manipulation Coltraro et al. (2022, 2024) and it has real information of the depth of several keypoints of the cloth using Opti-Track small markers. To our knowledge this is the first time that such a comprehensive dataset of highly dynamic motions with non-trivial deformations has been collected and publicly released. Together with the dataset, we propose a measure to benchmark the sim-to-real gap of any existing or novel cloth simulator using our data. We also show how the recorded trajectories of the 2 upper corners for one of the scenarios considered can be directly executed by a robotic arm with success.

Related work

In the field of computer vision, there have been many released datasets of real images of cloth addressing different perception problems, e.g. cloth recognition/classification or landmark points detection Liu et al. (2016), surface reconstruction for cloth objects Bednarik et al. (2018) or segmentation of cloth parts when worn by humans Zhao et al. (2018). Datasets related to fashion as Liu et al. (2016) include a large amount of images with labeling indicating cloth type and landmarks to facilitate localization of the object, e.g. where the sleeves are or the location of the end of the trousers. Clothes are either completely flat or worn by humans. Datasets for surface reconstruction like Bednarik et al. (2018) consist of the RGB-D data the system needs in order to reconstruct depth and normals from a single-view image.

Robotic manipulation of cloth has very different perception requirements, as clothes may appear in many more unstructured configurations as crumpled, hanged or folded. Cloth classification is needed to define different folding strategies, so there are datasets to learn to identify clothes from different crumpled states Sun et al. (2017), or to classify them and estimate their pose Mariolis et al. (2015) or to segment and identify wrinkles Wagner et al. (2013). The problem of identifying landmarks on cloth has been adapted from the literature of computer vision to apply it during robotic manipulation in Gustavsson et al. (2022), using combinations of existing datasets and adding some more images from robotic manipulation. In the dataset Verleysen et al. (2020) they use real recordings of people folding clothes, identifying the skeleton of people, and recording RGB-D from 3 perspectives, but without a clear way of obtaining the depth of only the pieces of cloth. Another relevant problem in robotics is to identify a set of very particular landmarks: corners and edges. Datasets to recognize corners provide RGB-D data where color Qian et al. (2020) or UV-light Thananjeyan et al. (2022) are used

to segment regions of interest in pieces of cloth that are later identified using only depth. A more complex issue is that of tracking deformation while manipulating the cloth. Few datasets exist of point-clouds with a labeling of what points correspond to important parts of the cloth like corners or edges during a manipulation Schulman et al. (2013). Another important trend in robotics is using Deep Neural Networks to predict actions from images. An example in this field is the dataset in Avigal et al. (2022) containing RGB-D images where the action is annotated as a pick up point and a direction of motion in the image, and a different image before and after the action. More complex actions have been tackled lately with Visual Language Models where a sequence of images is linked to a sequence of positions of the end-effector Chi et al. (2024).

Recent large-scale robot manipulation datasets like Padalkar et al. (2023); Khazatsky et al. (2024) include images and robot trajectories for different robot environments, scenes and objects, some clothes among them, but without ground truth on cloth deformation data. Due to the complexity of understanding deformations in cloth, a lot of the literature on learning manipulation policies tailored to textiles use cloth simulators. However, for most of the cloth simulators widely used to learn dynamic actions there is a large sim-to-real gap Blanco-Mulero et al. (2024) that needs to be closed by training with real data. This is where comprehensive datasets such as ours can become of critical importance, since they can be used for the calibration of the physical parameters of the cloth models, which is one of the main difficulties in using cloth simulators for planning and manipulation purposes.

The most closely related datasets to ours are those meant to test cloth simulators. In computer graphics, cloth simulations have been looking increasingly realistic, although not reflecting the real behaviour of cloth that robots would require to predict motions. Real image datasets of cloth in this field are meant to estimate the parameters of the simulation, focusing on local properties of cloth such as elasticity or rigidity Wang et al. (2011); Miguel et al. (2012) where specially designed machines are used to measure elasticity parameters, and datasets contain static images of the fabrics before and after certain deformations Clyde et al. (2017). Other works that study friction such as Rasheed et al. (2021) use video recordings of clothes in motion with very simple friction and collision scenarios, but without any depth information.

The dataset presented in this work falls into the previous category. To the best of our knowledge,
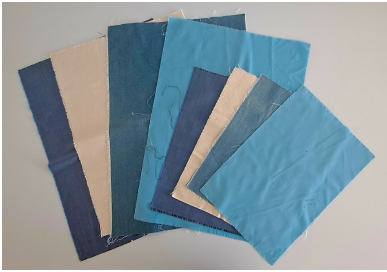
	Density (kg/m²)	Stiffness (%)	Friction Coeff.	Elasticity (%)
Polyester	0.1042	7,6	0.48	(3,3%, 3,5%, 11,3%)
Cotton	0.3046	43,7	0.68	(0,8%, 0,6%, 1,0%)
Wool	0.1804	12,9	0.59	(5,0%, 2,4%, 9,6%)
Denim	0.3046	20,3	0.8	(10,0%, 3,8%, 11,3%)

Elasticity given for the short, long edges and the diagonal, respectively.

Figure 1. All the fabrics (size A2 and A3) recorded for the dataset. From left to right we have: wool, (stiff) cotton, denim and polyester (first A2 sizes and then A3). On the table we provide the mechanical parameters of the fabrics following the standarization rules for measuring them proposed in Garcia-Camacho et al. (2024).

it is the first to include real depth information recorded during highly dynamic motions with non-trivial deformations including collisions and self-collisions of the textiles. Recently, a dataset has been published as the byproduct of the benchmark done in Blanco-Mulero et al. (2024), where they compared several cloth simulators to evaluate the sim-to-real gap using depth images collected with an RGB-D camera. However, that dataset includes only one dynamic task (placing the cloth flat on the table, also included in our dataset) and one quasi-static task of a very similar nature, with 3 rectangular clothes. The dataset presented in this work is more complete in terms of types of dynamic motions, cloth materials and cloth sizes, comprising 120 recordings involving rapid shaking and twisting of the textiles, collisions with frictional objects, strong hits with a long and thin rigid object and even self-collisions. Therefore, it is our hope, that with the provided dataset, a much throughout comparison between simulators can be done. In addition, the Opti-track recordings, although having less recorded points, provide much better accuracy than the RGB-D cameras, which can be very noisy around the boundaries of the cloths.

Data collection

We recorded the motion of real pieces of cloth under several dynamic conditions, including self-collision, collisions with a table and a rigid stick. In total, we recorded 120 motions with a *Motion Capture System* that captured the deformations of the cloth. In the following we give more details about the cloths used and the recording setting.

Cloth's materials and sizes

For the recording in this work, we employ four cloth materials described in Figure 1 and two different sizes: A3 (0.297 x 0.420 m with area 0.1247 m²) and

A2 (0.42 x 0.594 m with area 0.2495 m²). Before performing the experiments they were ironed to remove all considerations of plasticity from the validation process. All the textiles are used in all recordings, except in the collision scenarios, where we only record the A2 textiles. In Table in Fig. 1 we can see the density of all the fabrics and we also report the mechanical characteristics of all the fabrics using the standardization measures proposed in Garcia-Camacho et al. (2024). We can see how the chosen clothes present a variety of mechanical properties, the cotton being the stiffest and less elastic and with large friction, while the polyester is the less stiff, but showing elasticity mostly in the diagonal direction (i.e. shearing). The denim cloth, made of cotton and elastane is the second most stiff, but presents the highest friction and elasticity.

Motion capture system for cloth

To record the motion of the textiles we use a system of cameras that detects and tracks reflective markers that are hooked on the cloth. This technology has been extensively used to track the motion of rigid and articulated bodies. Nevertheless, its use for deformable objects has been less common since the weight of the markers could affect the dynamics of the object. To avoid this, we used very small markers, with a diameter of 3 mm and a weight of 0.013 g, and therefore account for less than 1.25% of cloth's weight even for the lightest materials. Depending on the size of the cloth we used different number of markers. For the A2 size, we used 20 reflective markers, whereas for the A3 ones 12 were used. In both cases, the makers are placed equidistantly in order to obtain a faithful representation of the dynamics of the fabrics. An example can be seen in Figure 2 - right. Notice that from this configuration of the markers we can easily obtain a mesh for the recorded cloths.

The setup used for data collection is shown in Figure 2 - left. We used five *Optitrack Flex 13*

Figure 2. *Left*: setup used to record the motion of the textiles. Five cameras surround the scene so that every marker (encircled in red on the right) is visible to at least 2 cameras at the same time. *Right*: reflective markers attached to the denim sample, with a diameter of 3 mm and a weight of 0.013 g.

cameras surrounding a scene from the manufacturer *NaturalPoint Inc*. We found that 5 was enough to record a varied set of fast movements without losing track of the textiles. The cameras cannot face each other, since this causes blind spots that make markers become invisible. At the beginning of each recording session the cameras were calibrated automatically with respect to a user-defined reference system. We defined the plane $z = 0$ to be either at the floor or at a table for the collision scenario. The post-process of the data was done with the provided software *Motive*. This combination of software and hardware offers sub-millimeter marker precision, in most applications less than 0.10 mm according to the manufacturers.

Free-hanging motions

These recordings allow us to study how the characteristics, speed and size of the textiles affect their motion without collisions. To this end, we execute two different trajectories:

(a) Shaking: the cloth is held by two corners and shaken back and forwards following the motion in Figure 3-(a).

(b) Twisting: the cloth is held by two corners. The line formed by the two grasps is rotated multiple times (approximately 30 degrees) with respect to the z-axis, as shown in Figure 3-(b).

We recorded these trajectories for the 8 different textiles listed in Figure 1. Each motion is repeated at two different speeds: *slow* and *fast* and in two different grasping modes. In the grasping mode I, the human moves the cloth with a hanger that has the
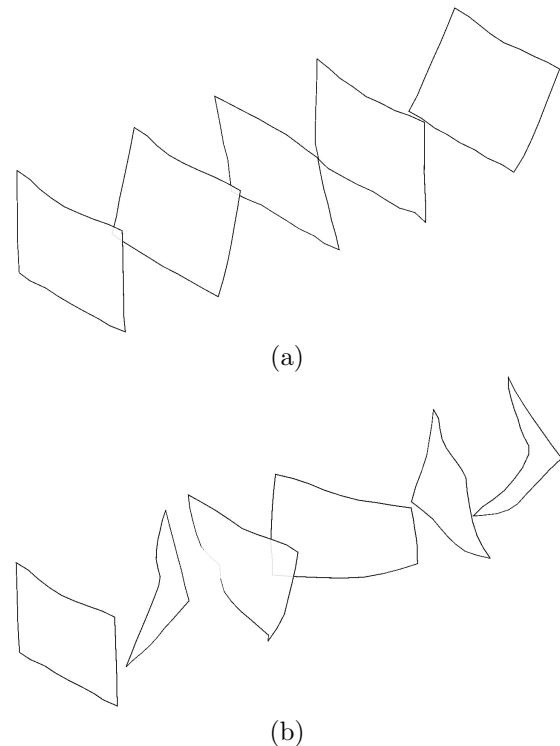
Figure 3. (a) Shaking motion sequence (left to right): the cloth is shaken back and forwards. (b) Twisting motion sequence (left to right): the cloth is rotated with respect to the z-axis back and forth several times.

cloth grasped, as shown in Figure 2, right. In the grasping mode II, the human holds the cloth by the two upper corners with bare hands. The combinations of

Table 1. Average velocities ($\mathrm{m \cdot s^{-1}}$) for the twisting motion of the A2 textiles. We display the speeds for the first repetition I (with a hanger) and for the second II (with bare hands).

| Material	Slow I	Fast I	Slow II	Fast II
Polyester	0.081	0.250	0.090	0.247
Wool	0.093	0.228	0.085	0.256
Denim	0.092	0.308	0.090	0.292
Stiff-cotton	0.091	0.218	0.107	0.246

all the materials, sizes and modes leads to 64 different recordings.

Each motion lasts approximately 15 seconds with a frame-rate of 120 Hz. As the motions are performed by a human, every movement has its own unique variability. As an example, in Table 1 we can see the average speeds of the twisting motion comparing the fast and slow speeds of the A2 textiles. Despite the variability natural in human repetitions, the table shows that overall the speeds are maintained consistently.

Motions interacting with the environment

This set of recordings are meant to study the interactions of cloth with the environment, that is, when it collides with the objects of the environment with different frictions and shapes. Moreover, we also recorded (to our knowledge for the first time in literature) self-collisions of real cloths. Notice that self-collisions are intrinsically more difficult to record with the motion capture system than other motions, since in many cases they involve self-occlusions (think about folding a cloth in four, three quarters of the markers would be completely self-occluded by the time the fold is completed). Therefore, we designed a scenario in which every part of the cloth could be fully recorded, and the self-collisions were non-trivial. We recorded all these collision tests only with the A2 size cloths. We performed the following motions:

(a) Collision with a table: The textile starts grasped by two corners at about 10 cm of height and is afterwards laid dynamically on the table as shown in Figure 4-(a). Each motion lasts approximately 5 seconds and is performed with two different table surfaces: one with *low* friction consisting in a raw polished table and one with *high* friction, a table with a tablecloth. Moreover we consider two additional sub-cases depending on if the lay is complete or partial:
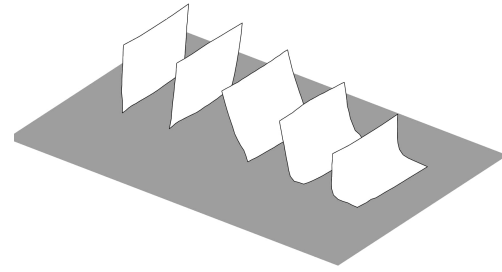
(a.1)
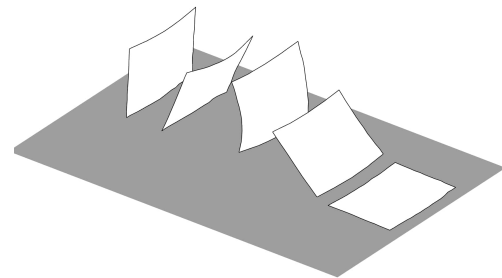
(a.2)
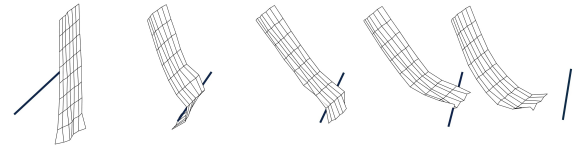
(b)

Figure 4. (a.1): the cloth starts suspended and is afterwards dynamically laid partially onto the table. (a.2): the cloth starts suspended and is afterwards dynamically laid fully onto the table. (b): the cloth is held by its two upper corners and then is hit repeatedly with a long thin stick.

1. Half lay with and without friction: the cloth is laid only partially, so that half of the cloth is still suspended (see Figure 4-(a.1)).
2. Full lay with and without friction: the cloth is laid fully, so that the the cloth is fully flat on the table (Figure 4-(a.2)).

For these motions we did a total of 16 recordings.

(b) Hitting scenario: the cloths were grasped by two corners, and held suspended in the air with the long sides perpendicular to the floor. Then, they were hit repeatedly with a long thin stick. Each

Table 2. Summary of all the motions recorded in this work. We report the size of the textile used (DIN A2 or A3), the amount of markers attached, how many recordings we have, the typical duration of one recording, whether there are collisions involved and a link to the relevant figure.

Scenario	Sizes	Markers	Recordings	Duration	Collisions	Details
Free-hanging	A3, A2	12 or 20	64	∼15 s	No	Shake or twist (Figure 3).
Collision with table	A2	20	16	∼5 s	Yes	Half or full lay (Figure 4-(a))
Hitting	A2	20	4	∼40 s	Yes	With stick (Figure 4-(b))
Self-collisions	A2	20	36	∼7 s	Yes	3 sub-cases (Figure 5)

textile is hit four times at various locations of the cloth with varied strengths and speeds as shown in Figure 4-(b). The recordings lasted around 40 seconds. The stick has a length of 75 cm and a diameter of 1.5 cm. Two markers are put at both ends of the stick to record its trajectory. For this scenario there are a total of 4 recordings.

(c) Self-collisions: the cloths were grasped by four or two corners, and held suspended in the air with their long sides parallel to the floor and the middle of the textile resting on top of a metallic rod, see Figure 5. Then, the corners were released so that the cloths collide with themselves. We consider 3 different sub-cases:

 1. The cloth is held by its four corners with its long side perpendicular (or normal) to the rod (see Figure 5-(c.1)). Then the 4 corners are released at the same time.

 2. The cloth is held by its four corners with its long side parallel to the rod (see Figure 5-(c.2)). Then the 4 corners are again released at the same time.

 3. The cloth is held by two of its corners with its long side perpendicular to the rod (see Figure 5-(c.3)). In this case approximately half of the cloth is already down (perpendicular to the floor). Then the 2 corners are released to cause the self-collision.

The recordings lasted between 6 and 8 seconds. The rod has a length of 164 cm and a diameter of 7 mm. Two markers are put at both ends of the rod to record its static position. For this scenario each sub-case is recorded 3 times and thus there are a total of 36 recordings.

A summary of all the scenarios considered can be seen in Table 2.

Data format and processing

Each cloth motion is stored in a comma-separated values (CSV) text file containing the trajectory in space of all reflective markers. We show an example of one of the files in Figure 6. The relevant parts of each file are:

- *Header*: contains general information about the recording equipment and other miscellaneous data (e.g. the frame rate of the recording).

- *3rd row*: unique IDs for each marker. This is only relevant for the hitting scenario where the ends of the stick are labeled with ID = 21 and ID = 22.

- *6th row and onwards*: here is where the most important data is stored, each row consists of a frame identifier, a time-stamp and the concatenation of the x,y,z coordinates of each marker at the given time-stamp.

If we denote by $(x_k(t), y_k(t), z_k(t)) \in \mathbb{R}^3$ the position in space of marker $k = 1, \ldots, N$ (e.g. $N = 12$ in the case of the A3 textiles) at time t, then (omitting the frame identifier column) from the 6th row and onwards each CSV file can be represented by a matrix A where the $j-$th row $a_j \in \mathbb{R}^{3N+1}$ of A is given by:

$$a_j = (t_j, x_1(t_j), y_1(t_j), z_1(t_j), \ldots, x_N(t_j), y_N(t_j), z_N(t_j)).$$

Finally, let us mention that since in the hitting scenario we are also recording the motion of the rigid stick colliding with the cloth, we also get six more real numbers per row, which correspond to the position in space of the ends of the stick (these markers have labels in the 3rd row of the CSV file with ID = 21 and ID = 22 as previously mentioned). This also happens for the self-collision scenario, but that case is even simpler because the rod is static. The coordinates of the rod are
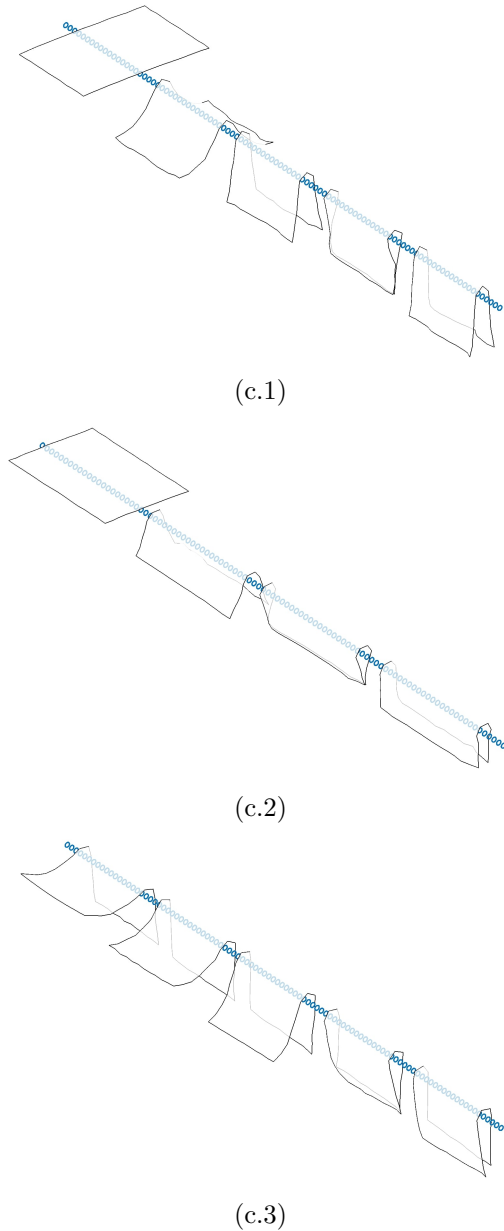
(c.1)

(c.2)

(c.3)

Figure 5. (c.1): the cloth is held by its four corners with its long side perpendicular to the rod and then the 4 corners are released. (c.2): the cloth is held by its four corners with its long side parallel to the rod and then the corners are released. (c.3): the cloth is held by 2 of its corners with its long side perpendicular to the rod and then these 2 corners are released.

given in this case by the first 6 columns of A (omitting the first column with the time-stamps)

The software *Motive* does an automatic labeling and tracking of the markers, nevertheless when the movements are too rapid it loses some of them and creates new labels. We have post-processed manually

the data to identify and merge the different labels that correspond to the same markers. Inevitably some markers are lost some of the time (especially with fast or abrupt movements), for instance when the textiles deform so much that the corners are no longer visible to the cameras. We have taken care that in our recordings these disappearances only happen for short periods of time. Nevertheless, in the rare cases when a marker is lost for some frames, an empty value is stored in the CSV file for each time-stamp in which the marker is missing.

In order to quantify this phenomenon, in Table 3 we compute the total amount of missing markers in each frame for all the scenarios. This means that in each frame of every recording we count how many markers are missing and add them all up.

Table 3. Amount of missing markers for all the scenarios.

Scenario	Missing	Total	%
Free-hanging	7,397	1,953,576	0.38
Collision table	1,961	291,060	0.67
Hitting	3,195	261,800	1.22
Self-collisions	341	478,654	0.07
Overall	12,894	2,985,120	0.43

As we can see in the table, the amount of missing data is very small (less than 1%) for all scenarios with the exception of the hitting scenario, where the speed and strength of the hits cause some markers go missing for small amounts of time.

Reading the data

As explained in the previous section, extracting the data from each file sums up to reading a rectangular matrix A of dimensions $T \times (3N + 1)$ where T is the number of time-stamps and N is the number of markers (i.e. $N = 12$ or $N = 20$ depending on the size of the cloth). For convenience of the user we provide a MATLAB file `read_data.m` that automates the reading process. The naming convention we have used for the files and how they are distributed is as follows. We provide 4 separate folders:

- **Free-hanging**: these correspond to the free-hanging motions shaking and twisting of all the textiles. Each file has the generic name:

  material_size_type_speed_grasp.csv

  where

  $material \in \{\text{denim, cotton, wool, polyester}\},$

```
Format Version,1.2,Take Name,LANA P2,Capture Frame Rate,120.000000,Export Frame Rate,120.00

,,Marker,Marker,Marker,Marker,Marker,Marker,Marker,Marker,Marker,Marker,Marker,Marker,Marke
,,1,1,1,9,9,9,6,6,6,2,2,2,5,5,5,3,3,3,10,10,10,11,11,11,12,12,12,8,8,8,7,7,7,4,4,4
,,"8392B20CF49E357047E","8392B20CF49E357047E","8392B20CF49E357047E","8392B20CF49E357047C","
,,Position,Position,Position,Position,Position,Position,Position,Position,Position,Position
Frame,Time,X,Y,Z,X,Y,Z,X,Y,Z,X,Y,Z,X,Y,Z,X,Y,Z,X,Y,Z,X,Y,Z,X,Y,Z,X,Y,Z,X,Y,Z,X,Y,Z
3788,31.566667,-0.066073,0.484585,-0.073028,-0.049170,0.198841,-0.075579,-0.064458,0.338040
3789,31.575000,-0.066090,0.484598,-0.073036,-0.049022,0.198936,-0.075624,-0.064261,0.338118
3790,31.583333,-0.066100,0.484613,-0.073035,-0.048929,0.198952,-0.075640,-0.064243,0.338121
3791,31.591667,-0.066113,0.484625,-0.073045,-0.048844,0.198971,-0.075658,-0.064214,0.338128
3792,31.600000,-0.066123,0.484631,-0.073040,-0.048723,0.199008,-0.075827,-0.063802,0.338122
3793,31.608333,-0.066150,0.484644,-0.073037,-0.048647,0.199021,-0.075847,-0.063775,0.338126
3794,31.616667,-0.066154,0.484659,-0.073041,-0.048567,0.199038,-0.075861,-0.063643,0.338058
3795,31.625000,-0.066162,0.484668,-0.073036,-0.048514,0.199053,-0.075882,-0.063904,0.338179
3796,31.633333,-0.066165,0.484675,-0.073025,-0.048430,0.199074,-0.075887,-0.063843,0.338174
3797,31.641667,-0.066180,0.484687,-0.073019,-0.048014,0.199081,-0.075971,-0.063801,0.338171
3798,31.650000,-0.066169,0.484693,-0.073007,-0.047933,0.199089,-0.075975,-0.063822,0.338179
3799,31.658333,-0.066163,0.484692,-0.073007,-0.047748,0.199124,-0.076075,-0.063802,0.338183
```

Figure 6. CSV text file: each row gives a time-stamp and the x,y,z coordinates of every marker at the given time-stamp.

$size \in \{$A2, A3$\}$,

$type \in \{$shake, twist$\}$,

$speed \in \{$slow, fast$\}$ and

$grasp \in \{$hands, hanger$\}$.

Then, for instance, the fast shaking of A3 cotton with grasping mode hanger is named cotton_A3_shake_fast_hanger.csv.

• **Tablecloth**: these correspond to the collision with a table scenario as described before. Each file has the generic name:

$$material\_size\_type\_friction.\text{csv}$$

where

$material \in \{$denim, cotton, wool, polyester$\}$,

$size \in \{$A2$\}$,

$type \in \{$half_lay, full_lay$\}$ and

$friction \in \{$low_friction, high_friction$\}$.

So for instance the full lay of A2 denim using a high friction table is named:

$$denim\_A2\_full\_lay\_high\_friction.\text{csv}$$

Moreover, in the MATLAB file `read_data.m` for these motions, we provide code to extract the trajectories of the two upper corners (these can be directly executed by a robot as we highlight later).

• **Hitting**: these correspond to the second collision scenario where the textiles are hit with a rigid stick. Each file has the generic name:

$$material\_size\_type.\text{csv}$$

where

$material \in \{$denim, cotton, wool, polyester$\}$,

$size \in \{$A2$\}$ and

$type \in \{$hitting$\}$.

So for instance the hitting of A2 polyester is named polyester_A2_hitting.csv.

• **Self-collision**: these correspond to the final collision scenario where the textiles collide with a rod and themselves. Each file has the generic name:

$$material\_size\_number\_corners\_position\_rep.\text{csv}$$

where

$material \in \{$denim, cotton, wool, polyester$\}$,

$size \in \{$A2$\}$,

$number \in \{$four, two$\}$,

$position \in \{$normal, parallel$\}$ and

$rep \in \{$rep1, rep2, rep3$\}$.

Then, for instance, the second repetition of the self-collision recording for wool, when held by 4 corners and with its long edge normal to the rod is named

wool_A2_four_corners_normal_rep2.csv.

Therefore we have a total of $4 \times 2 \times 2 \times 2 \times 2 = 64$ free-hanging files, $4 \times 1 \times 2 \times 2 = 16$ tablecloth collision motions, 4 hitting recordings and $4 \times 3 \times 3 = 36$ self-collision cases. As already mentioned, we provide for each case a specialized MATLAB file `read_data.m` that automates the reading process.

Dataset use-cases

In the following we present two possible applications of the previous dataset: extracting the trajectory of the two upper corners of the textiles to execute them with a robotic arm (possibly enabling imitation learning methods) and its use for the empirical validation of cloth models.

Empirical validation of cloth simulators

A direct application of the dataset presented in this work is the measurement of the sim-to-real gap of a cloth simulator. For that, one needs to compare the recorded cloths with the simulated ones. In order to do that, denote the sequence of positions of the recorded nodes given by the motion capture system every $\Delta t > 0$ seconds by $\{\phi^0, \phi^1, \ldots, \phi^m\}$ and the simulated sequence obtained from a cloth model by $\{\varphi^0, \varphi^1, \ldots, \varphi^m\}$. This sequence is obtained by taking $\varphi^0 = \phi^0$ (the cloth is meshed by the canonical meshing given by the markers: see Figure 2, right) and simulating the cloth using the recorded trajectories of the two upper corners with the same $\Delta t > 0$. Then, we can compute the mean absolute error:

$$\bar{e} = \frac{1}{m} \sum_{n=1}^{m} \sqrt{||\varphi^n - \phi^n||_{\mathbf{M}}^2}, \tag{1}$$

where $|| \cdot ||_{\mathbf{M}}$ is the norm induced by the mass matrix $\mathbf{M}$ of the mesh discretization of the recorded cloth. This means that $||\mathbf{x}||_{\mathbf{M}}^2 = \mathbf{x}^{\mathsf{T}} \cdot \mathbf{M} \cdot \mathbf{x}$ and $\mathbf{M}$ is a diagonal matrix with diagonal elements $m_k > 0$ equal to one third of the sum of the areas of all incident triangles (one fourth in the case of quadrilaterals) to the kth node of the discrete mesh. The use of the mass matrix $\mathbf{M}$ ensures that we are taking a sound approximation of the integral $\frac{1}{m} \sum_{n=1}^{m} \int ||\varphi^n - \phi^n|| dA$. This is appropriate for an error measure that is global in the domain, and tries to be as independent of the mesh as possible. In the case that a marker k was missing for a given frame n, we just omit its corresponding coordinate from the computation of the norm $|| \cdot ||_{\mathbf{M}}$ for that given frame, i.e. we use a reduced diagonal sub-matrix of $\mathbf{M}$ by removing the kth row and column. Notice that in the case that the simulated cloth has a finer resolution than the recording, we must only use the subsample of the simulated nodes that coincide with the recorded ones to compute $\bar{e}$.

In sum, the proposed error measure (1) is very adequate to compare simulations to recordings for any cloth model since it is robust with respect to the number of nodes of the mesh, it has physical units and moreover it compares the same regions of the simulated and recorded cloths, e.g. the two lower corners of the recording are compared to the two lower corners of the simulation, as opposed to other error measures (such as the Chamfer and Hausdorff distances) used e.g. in Blanco-Mulero et al. (2024).

Generalisability: notice that in our dataset we have enough data to even test the *out of sample* accuracy of cloth simulators after their parameters have been estimated (e.g. minimizing (1)). For instance: using only the shake scenario to 'train' the simulator, one could validate its realism with the twist scenario. Or using only the half lay collision with a table case, one could test the accuracy with the full lay motion. Furthermore, as the physical parameters of most cloth models are intrinsic and only depend on the physical characteristics of the textile in question, any estimation made for rectangular cloths would work for garments with non-trivial topology made of the same material. In fact, many cloth models simulate garments with non-trivial topologies as ensembles of flat patches identified along their boundaries, which would correspond to real-life seams.

Inextensible cloth simulator. Part of this dataset has been used for the empirical validation of the cloth model presented in Coltraro et al. (2022, 2024). It was shown that this model was able to simulate properly friction and to model the dynamics of fast and strong hits with a rigid object by using the *collision with a table* and *hitting* recordings described in this work, with simulations running two times faster than real-time, e.g. 12 seconds of real time are simulated in only 6. To carry out this empirical validation, three physical parameters of the cloth model were fitted: α (Rayleigh damping), δ (virtual aerodynamics mass) and μ (friction coefficient) by minimizing the absolute error (1) with respect to α, δ, μ. Then, optimal values of the parameters for both the high and a low friction *collision with a table* cases were found, with absolute errors (1) under 1 cm for all the DIN A2 textiles (for a video of the comparison, see https://youtu.be/sWJcxfTwKHE). Furthermore, using only two physical parameters (μ was set to 0), the model was able to simulate faithfully the *hitting scenario* with average errors around 1 cm (see https://youtu.be/U7-p_1E09L8) for all A2 textiles.
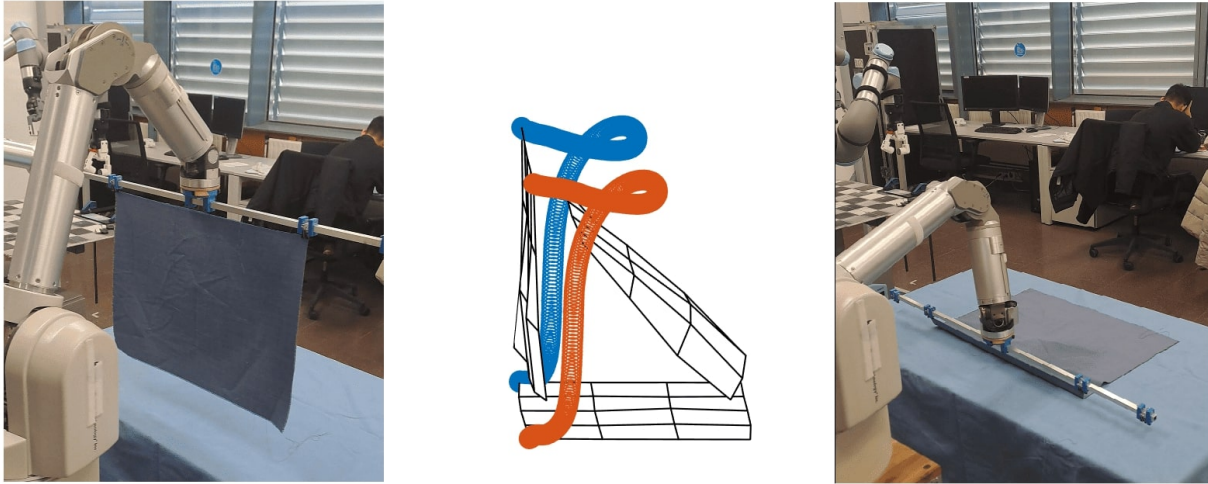
Figure 7. The robotic arm Barrett WAM with a hanger (left) executes the recorded upper corner trajectories for wool_A2_full_lay_low_friction.csv (middle) by computing the inverse kinematics so that the end-effector of the WAM follows precisely the average of the 2 curves. The result is a successful and smooth dynamic laying of the A2 wool textile onto a table (right). For the full motion, see the supplementary video.

Robot execution of recorded trajectories

As previously mentioned, the recorded trajectories of the Opti-Track markers are very precise and do not suffer from noise issues, as opposed to other recording methods, e.g. depth cameras. This opens the door to using these trajectories as demonstrations of successful tasks performed by a human that a robot can imitate or learn from. For instance, in the *collision with a table* scenario, we have recordings of a human laying a textile flat onto a table dynamically. Hence, the trajectory of the two upper corners of the cloths can be used as a demonstration from which the robot can learn from. Developing such learning methods is bond to be complex and it is out of the scope of this work, but as a proof of concept we execute with a robotic arm Barrett WAM with a hanger the recorded upper corner trajectories for wool_A2_full_lay_low_friction.csv (see Figure 7 and the supplementary video attached to this article). In order to do so, we simply extract the recorded trajectory of the 2 upper corner markers (Figure 7, middle) we average them, and then we compute the inverse kinematics so that the end-effector of the WAM follows precisely this averaged curve. The result is a successful and smooth dynamic laying of the A2 wool textile onto the table.

Influence of the markers or the grasping points. In order to study to which extent the grasping points or the markers that we attach to the cloths influence their dynamics, we will use the WAM robot and the trajectory shown in Figure 7 as described in the previous section. The robot reproducibility was first
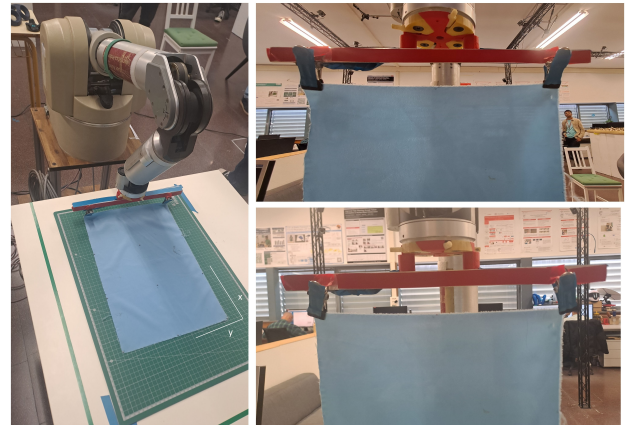

Figure 8. To study the influence of the grasping points or the markers in the cloths' dynamics, we consider 3 scenarios with varying conditions: a reference grasp (upper right), the reference grasp plus the markers and an alternative grasp without markers (lower right). The same trajectory is executed several times by the robot and we annotate the final position of the 2 lower corners using a metric board (left).

tested by executing the same trajectory many times, showing a millimetric margin of error with respect to the final position and orientation of the end-effector. We performed a quantitative comparison of 3 different scenarios using the lightest of all textiles: the size A3 polyester sample (see the right most textile in Figure 1). This sample weights 13 g and we attach 12 markers weighting 0.013 g each, so they amount to 1.2% of the cloth's weight. The 3 scenarios considered are:

Reference : the trajectories are executed with a standard grasp and without markers (see Figure 8, upper right).

With markers : the trajectories are executed with the same grasp as in the reference scenario but with the markers attached.

Alternative grasp : the trajectories are executed with a different grasp and without markers (see Figure 8, lower right).

In order to measure the differences, we use a metric board which allows us to measure the final in-plane position of the lower two corners of the cloth (see Figure 8, left). For each of the described scenarios we execute the same trajectory 14 times and carefully (to a margin of error of 0.5 cm) annotate the final position of the 2 lower corners with respect to the axis shown in Figure 8.

The results of all the trials can be seen in Figure 9. On the top we plot the 'left corner' and on the bottom the 'right corner' (as seen in Figure 8, left). We also plot *confidence ellipses* corresponding to 2 standard deviations.

At first glance, the distributions of the corner's positions for the 3 scenarios considered seem to be very similar (the confidence ellipse for the 'alternative grasp' case might be somehow smaller because there are less outliers for that scenario). In order to have a more quantitative measure, in Table 4 we compute the mean of each group for each corner separately.

Table 4. Mean of each scenario for the left and right corners separately. The coordinates are given in cm by the metric board (Figure 8, left).

Scenario	Left corner	Right corner
Reference	(12.0 , 38.2)	(8.9 , 10.5)
With markers	(11.7 , 37.9)	(8.5 , 10.6)
Alternative grasp	(11.9 , 38.5)	(9.2 , 10.5)

Discussion: the goal of this set of experiments was to study the influence of the markers or the grasping points in the dynamics of the textiles. Since we needed to measure the influence of the markers, we could not use the motion capture system as before. As a result, we performed a very dynamic motion (Figure 7) and recorded the final position of the 2 lower corners of a cloth (A3 polyester, the lightest of all) as given by a metric board (Figure 8, left). It is interesting to note that even within the same scenario there can be quite a lot of variability (see Figure 9). This might be because the cloth is extremely light and many aerodynamics effects influence the motion. To the best
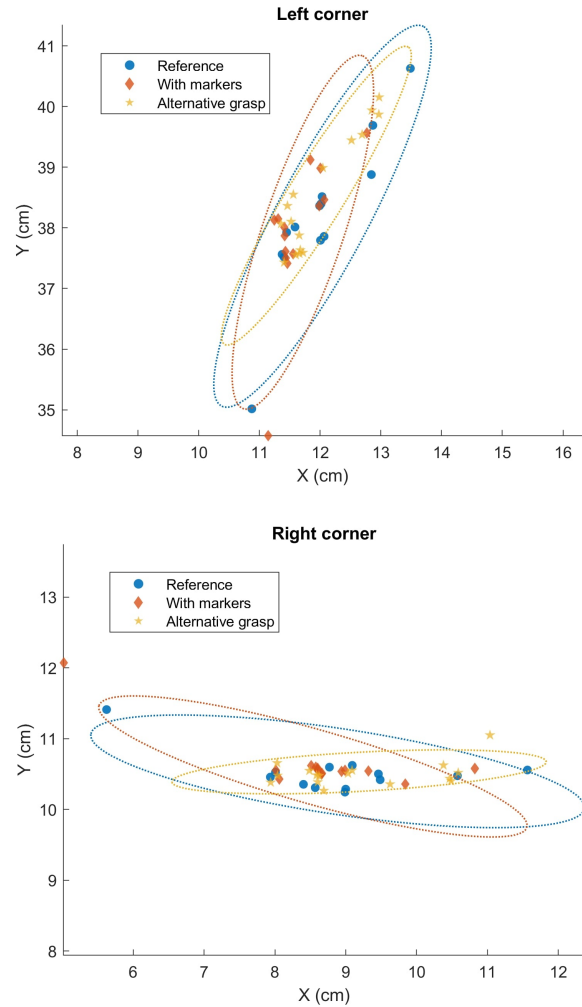

Figure 9. Final positions given by the metric board of the 'left corner' and the 'right corner' (as seen in Figure 8, left) for the 3 scenarios considered (reference, with markers and alternative grasp). In dotted lines we also plot confidence ellipses corresponding to 2 standard deviations.

of our knowledge, this is the first time this has been measured. But as can be seen in Figure 9 and Table 4 no noticeable bias is apparent when we attach the markers to the cloth or change its grasping points. This shows the reproducibility of the recorded motion, even under slightly different conditions (e.g. with a different grasp).

Conclusions and further work

We have presented a comprehensive dataset of highly dynamic motions of real clothes tracked with a *Motion Capture System*. To the best of our knowledge, it is the first time that such a dataset with recordings

of the real deformations of cloth in 3D has been compiled for complex cloth motions. We have a total of 120 motions with a variety of rectangular clothes of different stiffness, elasticity and friction properties, and with very different dynamic motions, including interactions with the environment and self-collisions. This dataset has a direct application to fine-tune cloth simulators to minimize the sim-to-real gap, and to benchmark existing simulators offering a clear ground-truth to compare to. This can ultimately increase the usability of simulators to train methods with a minimal sim-to-real gap. Moreover since the recorded trajectories for the upper corners of the textiles are of very high quality with very little noise, they can be directly executed by a robotic arm, and therefore are perfect candidates for applying learning algorithms to them, in order to generalize the recorded motions to other fabrics of different materials or sizes.

The methodology we have used to track and record the 3D position of key points on the cloth surface using Opti-Track markers is also novel and with great potential. Further work would involve re-recording this data jointly with synchronized RGB or RGB-D images of the clothes in motion to obtain deformation ground truth, opening the door to train perception methods and state-estimation learning algorithms simultaneously.

Acknowledgements

References

Avigal Y, Berscheid L, Asfour T, Kröger T and Goldberg K (2022) Speedfolding: Learning efficient bimanual folding of garments. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1–8.

Bednarik J, Fua P and Salzmann M (2018) Learning to reconstruct texture-less deformable surfaces from a single view. In: *Int. Conf. on 3d vision (3DV)*. pp. 606–615.

Blanco-Mulero D, Barbany O, Alcan G, Colomé A, Torras C and Kyrki V (2024) Benchmarking the sim-to-real gap in cloth manipulation. *IEEE Robotics and Automation Letters* 9: 2981–2988.

Chi C, Xu Z, Pan C, Cousineau E, Burchfiel B, Feng S, Tedrake R and Song S (2024) Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots. *arXiv pre-print* URL https://arxiv.org/abs/2402.10329.

Clyde D, Teran J and Tamstorf R (2017) Modeling and data-driven parameter estimation for woven fabrics. In: *Proceedings of the ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, SCA '17. New York, NY, USA: Association for Computing Machinery.

Coltraro F, Amorós J, Alberich-Carramiñana M and Torras C (2022) An inextensible model for the robotic manipulation of textiles. *Applied Mathematical Modelling* 101: 832–858. DOI:https://doi.org/10.1016/j.apm.2021.09.013.

Coltraro F, Amorós J, Alberich-Carramiñana M and Torras C (2024) A novel collision model for inextensible textiles and its experimental validation. *Applied Mathematical Modelling* 128: 287–308. DOI:https://doi.org/10.1016/j.apm.2024.01.030.

Eppner C, Mousavian A and Fox D (2019) A billion ways to grasps - an evaluation of grasp sampling schemes on a dense, physics-based grasp data set. In: *Proceedings of the International Symposium on Robotics Research (ISRR)*. Hanoi, Vietnam.

Garcia-Camacho I, Loghnini A, Welle MC, Alenyà G, Kragic D and Borras J (2024) Standardization of cloth objects and its relevance in robotic manipulation. In: *IEEE International Conference on Robotics and Automation (ICRA)*.

Gustavsson O, Ziegler T, Welle MC, Bütepage J, Varava A and Kragic D (2022) Cloth manipulation based on category classification and landmark detection. *International Journal of Advanced Robotic Systems* 19(4): 17298806221110445.

Handa A, Allshire A, Makoviychuk V, Petrenko A, Singh R, Liu J, Makoviichuk D, Wyk KV, Zhurkevich A, Sundaralingam B, Narang Y, Lafleche JF, Fox D and State G (2023) Dextreme: Transfer of agile in-hand manipulation from simulation to reality. In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 5977–5984.

Khazatsky A, Pertsch K, Nair S, Balakrishna A, Dasari S, Karamcheti S, Nasiriany S, Srirama MK, Chen LY, Ellis K, Fagan PD, Hejna J, Itkina M, Lepert M, Ma YJ, Miller PT, Wu J, Belkhale S, Dass S, Ha H, Jain A, Lee A, Lee Y, Memmel M, Park S, Radosavovic I, Wang K, Zhan A, Black K, Chi C, Hatch KB, Lin S, Lu J, Mercat J, Rehman A, Sanketi PR, Sharma A, Simpson

C, Vuong Q, Walke HR, Wulfe B, Xiao T, Yang JH, Yavary A, Zhao TZ, Agia C, Baijal R, Castro MG, Chen D, Chen Q, Chung T, Drake J, Foster EP, Gao J, Herrera DA, Heo M, Hsu K, Hu J, Jackson D, Le C, Li Y, Lin K, Lin R, Ma Z, Maddukuri A, Mirchandani S, Morton D, Nguyen T, O'Neill A, Scalise R, Seale D, Son V, Tian S, Tran E, Wang AE, Wu Y, Xie A, Yang J, Yin P, Zhang Y, Bastani O, Berseth G, Bohg J, Goldberg K, Gupta A, Gupta A, Jayaraman D, Lim JJ, Malik J, Martín-Martín R, Ramamoorthy S, Sadigh D, Song S, Wu J, Yip MC, Zhu Y, Kollar T, Levine S and Finn C (2024) Droid: A large-scale in-the-wild robot manipulation dataset. *IEEE International Conference on Robotics and Automation (ICRA)* .

Liu Z, Luo P, Qiu S, Wang X and Tang X (2016) Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In: *IEEE Conf. on Computer Vision and Pattern Recognition*. pp. 1096–1104.

Mariolis I, Peleka G, Kargakos A and Malassiotis S (2015) Pose and category recognition of highly deformable objects using deep learning. In: *Int. Conf. on advanced robotics*. pp. 655–662.

Miguel E, Bradley D, Thomaszewski B, Bickel B, Matusik W, Otaduy M and Marschner S (2012) Data-driven estimation of cloth simulation models. *Computer Graphics Forum* 31: 519–528. DOI:10.1111/j.1467-8 659.2012.03031.x.

Padalkar A, Pooley A, Jain A, Bewley A, Herzog A, Irpan A, Khazatsky A, Rai A, Singh A, Brohan A, , O'Neill A, Rehman A, Maddukuri A, Gupta A, Padalkar A, Lee A, Pooley A, Gupta A, Mandlekar A, Jain A, Tung A, Bewley A, Herzog A, Irpan A, Khazatsky A, Rai A, Gupta A, Wang A, Singh A, Garg A, Kembhavi A, Xie A, Brohan A, Raffin A, Sharma A, Yavary A, Jain A, Balakrishna A, Wahid A, Burgess-Limerick B, Kim B, Schölkopf B, Wulfe B, Ichter B, Lu C, Xu C, Le C, Finn C, Wang C, Xu C, Chi C, Huang C, Chan C, Agia C, Pan C, Fu C, Devin C, Xu D, Morton D, Driess D, Chen D, Pathak D, Shah D, Büchler D, Jayaraman D, Kalashnikov D, Sadigh D, Johns E, Foster E, Liu F, Ceola F, Xia F, Zhao F, Stulp F, Zhou G, Sukhatme GS, Salhotra G, Yan G, Feng G, Schiavi G, Berseth G, Kahn G, Wang G, Su H, Fang HS, Shi H, Bao H, Amor HB, Christensen HI, Furuta H, Walke H, Fang H, Ha H, Mordatch I, Radosavovic I, Leal I, Liang J, Abou-Chakra J, Kim J, Drake J, Peters J, Schneider J, Hsu J, Bohg J, Bingham J, Wu J, Gao J, Hu J, Wu J, Wu J, Sun J, Luo J, Gu J, Tan J, Oh J, Wu J, Lu J, Yang J, Malik J, Silvério J, Hejna J, Booher J, Tompson J, Yang J, Salvador J, Lim JJ, Han J, Wang K, Rao K, Pertsch K, Hausman K, Go K, Gopalakrishnan K, Goldberg K, Byrne K, Oslund K, Kawaharazuka K, Black K, Lin K, Zhang K, Ehsani K, Lekkala K, Ellis K, Rana K, Srinivasan K, Fang K, Singh KP, Zeng KH, Hatch K, Hsu K, Itti L, Chen LY, Pinto L, Fei-Fei L, Tan L, Fan LJ, Ott L, Lee L, Weihs L, Chen M, Lepert M, Memmel M, Tomizuka M, Itkina M, Castro MG, Spero M, Du M, Ahn M, Yip MC, Zhang M, Ding M, Heo M, Srirama MK, Sharma M, Kim MJ, Kanazawa N, Hansen N, Heess N, Joshi NJ, Suenderhauf N, Liu N, Palo ND, Shafiullah NMM, Mees O, Kroemer O, Bastani O, Sanketi PR, Miller PT, Yin P, Wohlhart P, Xu P, Fagan PD, Mitrano P, Sermanet P, Abbeel P, Sundaresan P, Chen Q, Vuong Q, Rafailov R, Tian R, Doshi R, Mart'in-Mart'in R, Baijal R, Scalise R, Hendrix R, Lin R, Qian R, Zhang R, Mendonca R, Shah R, Hoque R, Julian R, Bustamante S, Kirmani S, Levine S, Lin S, Moore S, Bahl S, Dass S, Sonawani S, Song S, Xu S, Haldar S, Karamcheti S, Adebola S, Guist S, Nasiriany S, Schaal S, Welker S, Tian S, Ramamoorthy S, Dasari S, Belkhale S, Park S, Nair S, Mirchandani S, Osa T, Gupta T, Harada T, Matsushima T, Xiao T, Kollar T, Yu T, Ding T, Davchev T, Zhao TZ, Armstrong T, Darrell T, Chung T, Jain V, Vanhoucke V, Zhan W, Zhou W, Burgard W, Chen X, Wang X, Zhu X, Geng X, Liu X, Liangwei X, Li X, Lu Y, Ma YJ, Kim Y, Chebotar Y, Zhou Y, Zhu Y, Wu Y, Xu Y, Wang Y, Bisk Y, Cho Y, Lee Y, Cui Y, Cao Y, Wu YH, Tang Y, Zhu Y, Zhang Y, Jiang Y, Li Y, Li Y, Iwasawa Y, Matsuo Y, Ma Z, Xu Z, Cui ZJ, Zhang Z and Lin Z (2023) Open X-Embodiment: Robotic learning datasets and rt-x models. *arXiv preprint* URL https://arxiv.org/abs/2310.08864.

Qian J, Weng T, Zhang L, Okorn B and Held D (2020) Cloth region segmentation for robust grasp selection. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 9553–9560.

Rasheed AH, Romero V, Bertails-Descoubes F, Wuhrer S, Franco JS and Lazarus A (2021) A visual approach to measure cloth-body and cloth-cloth friction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP.

Schulman J, Lee A, Ho J and Abbeel P (2013) Tracking deformable objects with point clouds. In: *IEEE Int. Conf. on Robotics and Automation*. pp. 1130–1137.

Sun L, Aragon-Camarasa G, Rogers S, Stolkin R and Siebert JP (2017) Single-shot clothing category recognition in free-configurations with application to autonomous clothes sorting. In: *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. pp. 6699–6706.

Thananjeyan B, Kerr J, Huang H, Gonzalez JE and Goldberg K (2022) All you need is LUV: Unsupervised collection of labeled images using uv-fluorescent markings. In: *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. pp. 3241–3248.

Verleysen A, Biondina M and Wyffels F (2020) Video dataset of human demonstrations of folding clothing for robotic folding. *The Int. Journal of Robotics Research* 39(9): 1031–1036.

Wagner L, Krejcová D and Smutnỳ V (2013) Ctu color and depth image dataset of spread garments. *Center for Machine Perception, Czech Technical University, Tech. Rep. CTU-CMP-2013-25* .

Wang H, Ramamoorthi R and O'Brien J (2011) Data-driven elastic models for cloth: Modeling and measurement. *ACM Transactions on Graphics* 30(4): 71:1–11. Proceedings of ACM SIGGRAPH 2011, Vancouver, BC Canada.

Zhao J, Li J, Cheng Y, Sim T, Yan S and Feng J (2018) Understanding humans in crowded scenes: Deep nested adversarial learning and a new benchmark for multi-human parsing. In: *26th ACM Int. Conf. on Multimedia*. pp. 792–800.