PREPRINT-EUROGRAPHICS 2023 / K. Myszkowski and M. Nießner
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 42 (2023), Number 2

How Will It Drape Like?
Capturing Fabric Mechanics from Depth Images

Carlos Rodriguez-Pardo[1,2,3] and Melania Prieto-Martin[2] and Dan Casas[2] and Elena Garces[1,2]

[1]SEDDI, Madrid, Spain
[2]Universidad Rey Juan Carlos, Madrid, Spain
[3]Universidad Carlos III de Madrid, Spain

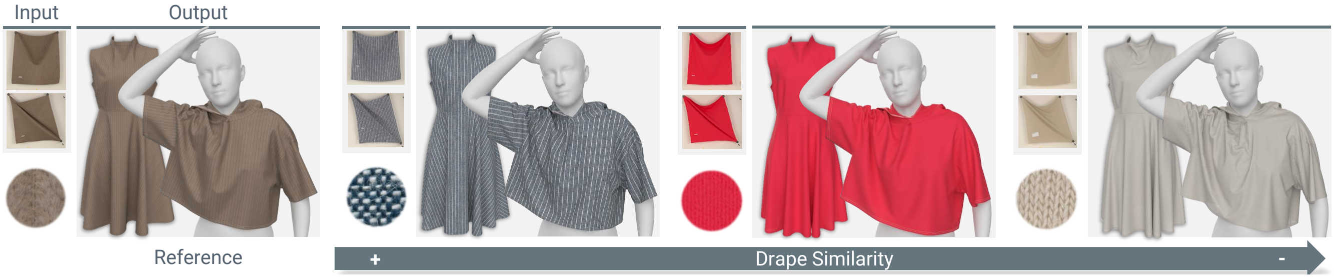


Figure 1: From just two depth images of a fabric sample casually placed in two specific configurations (input), we accurately infer the corresponding set of mechanical parameters of the material. Estimated parameters can be used in a cloth simulator, enabling to visualize the overall drape of any garment (output). Our method further introduces a novel perceptually-validated drape similarity metric, which enables sorting materials based on their final drape.

Abstract

We propose a method to estimate the mechanical parameters of fabrics using a casual capture setup with a depth camera. Our approach enables to create mechanically-correct digital representations of real-world textile materials, which is a fundamental step for many interactive design and engineering applications. As opposed to existing capture methods, which typically require expensive setups, video sequences, or manual intervention, our solution can capture at scale, is agnostic to the optical appearance of the textile, and facilitates fabric arrangement by non-expert operators. To this end, we propose a sim-to-real strategy to train a learning-based framework that can take as input one or multiple images and outputs a full set of mechanical parameters. Thanks to carefully designed data augmentation and transfer learning protocols, our solution generalizes to real images despite being trained only on synthetic data, hence successfully closing the sim-to-real loop. Key in our work is to demonstrate that evaluating the regression accuracy based on the similarity at parameter space leads to an inaccurate distances that do not match the human perception. To overcome this, we propose a novel metric for fabric drape similarity that operates on the image domain instead on the parameter space, allowing us to evaluate our estimation within the context of a similarity rank. We show that out metric correlates with human judgments about the perception of drape similarity, and that our model predictions produce perceptually accurate results compared to the ground truth parameters.

CCS Concepts

• ***Computing methodologies*** → ***Computer vision;*** *Neural networks; Computer graphics;*

1. Introduction

Creating accurate digital representations of real-world materials, or *Digital Twins*, is crucial for enabling realistic 3D visualizations suitable for interactive design and predictive engineering. Some industries, like fashion or textile manufacturing, further require these methods to work at a scale to cope with the fast pace of the current production workflows. However, digitalizing cloth is challenging due to the high variability and type of fabric samples, where the fabric composition, the microstructure, or the finishing play crucial roles in the perceived appearance.

While casual systems which obtain optical appearance have long been a focus of research, comparably less attention has been paid to estimating *mechanical* properties. Indeed, capturing and simulating the mechanical behavior of cloth is challenging due to the

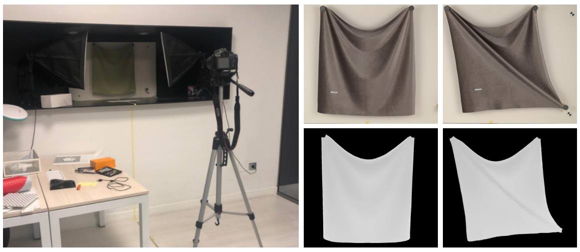


Figure 2: Capture setup, RGB (top), and depth (bottom) images for *hanging* (left) and *stretch* (right) scenes in rest position. Each scene conveys a different mechanical appearance of the fabric: *hanging* exhibits the overall drape; *stretch* exhibits an extra diagonal tension, which is key to understand the stretching properties.

complex interplay between the internal and external forces occurring in this type of physical system, which is highly sensitive to the environmental conditions. Nevertheless, with the current need to create virtual copies instantly, a casual setup able to produce automatic and accurate estimates –beyond having a set of *presets* from which to manually choose the closest one– of mechanical parameters could prove very valuable.

Previous methods are impractical for scalable and customizable workflows. Accurate fabric parameter acquisition systems require specialized and expensive devices [Kaw80, Min95, CPGE90], which are often slow and need skilled operators. Existing casual capture setups use input video sequences [BTH*03, BXBF13, YLL17] or, even if they take a single image, might require manual user input [JC20]. In this paper, we present a casual capture system that only requires taking two depth images of the textile posed in a static drape. Our capture setup does not require complex calibration, can be easily manipulated by non-expert operators, and is agnostic to the optical properties of the fabric thanks to leveraging depth images instead of RGB data. Figure 2 illustrates our capture setup. It involves capturing the fabric with a depth camera in two relaxed positions: the *hanging* scene, that conveys the drape when no force other than gravity is applied to it, and the *stretch* scene, which provides cues on the stretching properties of the fabric.

We propose a learning-based method to instantly return the mechanical parameters given static depth images and the fabric density as input. Our method relies on a sim-to-real strategy [ZGO*21], leveraging transfer learning and building on a dataset of physical fabrics digitalized with precision equipment. Our model is trained solely with synthetic data, and thanks to carefully designed policies of data augmentation and neural features, it can generalize to real-world scenarios. Under the hood, our approach leverages a custom architecture that enables a flexible design, that can take one or multiple images as input, enhancing its performance when more data is available. We perform an extensive evaluation by means of ablation studies and by measuring aggregated neural network saliency maps, which show that some scenes are more informative than others for predicting the mechanical properties of the fabric. Furthermore, we demonstrate the performance of our model on real-world captured samples, showcasing our system's generalization capabilities.

Key to our work is to demonstrate that evaluating the prediction accuracy of mechanical properties using typical error metrics, such as the Mean Absolute Error (MAE) on the parameter space, leads to inaccurate distances that do not match the human perception. We identify that such mismatches occur due to two factors: first, the parameter space is not bijective –i.e., different set of parameters might convey the same drape–; and second, a numerical error in a parameter does not necessarily correlate with what we perceive as an error. To address this shortcoming, common in all existing works, and inspired by previous work on similarity metrics for material appearance [LMS*19], natural images [ZIE*18] or illustration [GAGH14], we propose an image-based metric that measures differences on the mechanical behavior of textiles taking into account the overall drape. We validate that our metric agrees with human perception, and it and can be used to sort materials by drape similarity with respect to a reference fabric.

Using our similarity metric, we finally validate that the estimations of our method correlate with human judgments about drape similarity, and that our model predictions produce perceptually accurate results compared to the ground truth parameters. All in all, our approach makes an important step towards solving sim-to-real problems for mechanical estimation since it shows that simulated cloth using inferred parameters maximize the similarity with respect to real-world target fabrics.

2. Related Work

2.1. Parameter Estimation Methods

Estimating the mechanical properties of real fabric samples is a highly challenging problem for several reasons: the number of uncontrollable extrinsic factors (e.g., wind forces, initial state, collisions, etc.) which affect the predictiveness of the physical simulation; the lack of a standard deformation model and parameter spaces; and the use of computation-intensive simulation methods. Accordingly, a wide range of strategies exist, aimed at overcoming these challenges.

Measurement Devices. It is common to combine optimization techniques with the output of testing devices to find the optimal set of parameters which best explain the observations [MTLVL07, SB08, VMTF09, WOR11, MBT*12, CTT17]. Existing technologies of this type are diverse and, as discussed by Kuijpers *et al.* [KLBG20], lack of a clear standard. The Kawabata Evaluation System (KES) [Kaw80] is perhaps one of the most well known, measuring 16 coefficients including bending, shearing, and tensile among others. Despite its precision, this method was not widely adopted by the industry due to its lengthy processes and the need of expensive equipment. Consequently, several other methods tried to simplify and unify the methodology with partial success according to some studies [LMT08, Pow13]: the Fabric Assurance by Simple Testing (FAST) [Min95], the Fabric Touch Tester (FTT), the CLO Fabric Kit 2.0, the Fabric Analyser by Browzwear (FAB), the Optitex Mark 10, and the cantilever principle [CPGE90].

Reconstruction-Optimization Methods. Another set of techniques jointly tackles the reconstruction and parameter optimization problems. By taking as input data from arbitrary real simulations (e.g., the cloth deforming on an avatar [YPA*18]), they iteratively reconstruct and simulate the scene which better explains the observation. Bhat *et al.* [BTH*03] takes as input a video sequence

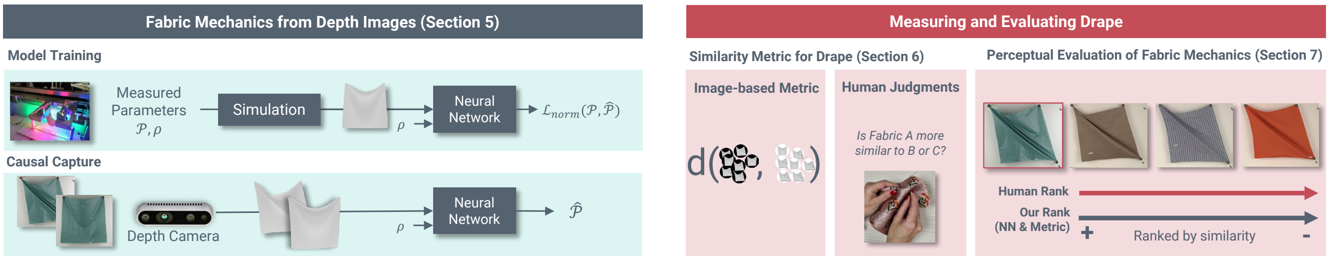

Fabric Mechanics from Depth Images (Section 5)
Model Training
Measured Parameters $\mathcal{P}, \rho$
Simulation
ρ
Neural Network
$\mathcal{L}_{norm}(\mathcal{P}, \hat{\mathcal{P}})$
Causal Capture
Depth Camera
ρ
Neural Network
$\hat{\mathcal{P}}$
Measuring and Evaluating Drape
Similarity Metric for Drape (Section 6)
Image-based Metric
d(,)
Human Judgments
Is Fabric A more similar to B or C?
Perceptual Evaluation of Fabric Mechanics (Section 7)
Human Rank
Our Rank (NN & Metric)
+
Ranked by similarity
-

Figure 3: An overview of the main components of our method. We propose a technique to estimate fabric mechanics using depth images of *hanging* and *stretch* scenes as input. To validate the error of our estimations in a perceptual manner –accounting for the global drape–, we propose an image-based drape similarity metric which we validate with human judgments and can be used to sort fabrics by similarity. We show through several metrics that the estimations provided by our method using our similarity metric agree with those given by humans.

of the cloth and use simulated annealing to optimize the parameters by measuring its folds. Yang *et al.* [YL15] use multi-view stereo reconstruction to initialize the 3D shape. Runia *et al.* [RGSS20] also introduces simulation steps to explain observed phenomena of cloth in the wind. They rely on similarity metrics computed on deep latent spaces to supervise the optimization of the parameters. These methods require simulation steps embedded into the fitting processes, making them computationally expensive due to the high dimensionality of the parameter spaces. Recent differentiable simulation techniques [LLK19, JBH19, HAL*19, MMG*20, LDW*22] have proven to be efficient ways to reduce the fitting burden by enabling the computation of gradients with respect to these parameters within latent spaces of neural networks, taking into account dynamics, self-collisions, and contacts.

Data-Driven and Regression Methods. The third set of methods avoids reconstructing the original 3D scene by working on an estimated feature space and leveraging previously simulated data and machine learning techniques. Our approach falls in this category. Taking videos as input has been explored by Bouman *et al.* [BXBF13] to recover stiffness and area weight using a descriptor of the image based on PCA and optical flow, and later by Yang *et al.* [YLL17], who leverage neural networks to extract image features used for regression. Davis *et al.* [DBC*15] estimated the same simulation parameters by exploiting imperceptible vibrations in high-speed video recordings. Bi *et al.* [BJNX18] further evaluated that humans also need fabric motion to understand its stiffness. Friction coefficients have been estimated using reflectance values [ZDN16] or dynamic videos of cloth sliding through a surface [RRBD*20] Instead of regressing the parameters, Huber *et al.* [HEW17] find the most similar cloth in a database using motion descriptors. A different approach only using a single image of the *Cusick drape* was followed by Ju *et al.* [JC20], but it requires a 360° scan to reconstruct the target cloth, and a manually fitted Bezier curve to obtain the feature vector. In contrast, we just require a depth map that can be captured easily. Concurrent work [FHXW22] uses multiple-view depth images as input to a trained regressor.

Our approach is inspired by these ideas; however, we do not require optimization –providing instant estimation of the parameters– and leverage neural features to understand and model fabric behavior in a semi-controlled setup. We demonstrate that our approach works with two images as input to predict bending and stretching coefficients without requiring a full video of the piece of fabric.

2.2. Pre-Trained Models and Transfer Learning

Deep learning models typically require vast amounts of data for generalizing to unseen examples. When this amount of data is not possible to acquire, *tranfer learning* techniques helps by reusing model parameters trained on a related task [KSL19, RZKB19, BHA*21]. These techniques include: *Fine-tuning* the weights of a pre-trained classification model [WKW16, RD17, CZW*18, WLZ*18, RJ19, GSK*19, KBZ*20]. Pre-training an image descriptor model on contrastive or self-supervised learning tasks, and use the activations of its last layer as input to the downstream task (*Linear Probing*) [CKNH20, RKH*21, CXH21, HCX*22, KRJ*22]. For domain adaptation problems, it is common to *adapt* the internal representations of pre-trained CNNs so as to efficiency [RBV17, RBV18, RPB19, PCYS20, LLB22]. Inspired by these approaches, we design a model that leverages fine-tuning of a pre-trained image CNN classifier as a feature extractor, capable of processing depth images, and extend it to account for additional input variables, and handling multiple images at the same time during test.

2.3. Similarity Metrics

Full-Reference Image Quality Assessment (IQA) aims to provide a single score which measures the amount of distortion between two images. Traditionally, these metrics leveraged low-level image statistics. *PSNR* is commonly used for measuring image degradation, but correlates poorly with human perception [ZIE*18]. More sophisticated alternatives have been developed, including *SSIM*, *mSSIM* [WBSS04], and others [ZZMZ11, NSHC16, ZSL14, ZL12, RBKW18]. Algorithms based on latent spaces of CNNs [GEB16] have been extended to better approximate human perception, for example, by training on a large pool of human evaluations *LPIPS* [ZIE*18], or by other means [DMWS20, PCMS18].

Besides, similarity metrics that measure abstract or complex concepts like style have been proposed for 3D furniture [LKS15], illustration [GAGH14], icons [LGG19], product design [LKS15], or material appearance [LMS*19]. Unlike ours, these metrics require to be trained with human ratings, thus incurring a considerable cost to collect such information via user studies. Instead, our metric does not require specific training, leveraging an off-the-shelf image-based metric. Despite this, we show that our metric correlates with human judgments on the perception of fabric drape similarity and that can be used to evaluate the overall drape.

3. Overview

Figure 3 presents an overview of our work. First, in Section 5, we introduce our novel solution to infer fabric mechanics directly from depth images. As input, our approach only requires static depth images in two specific configurations, shown in Figure 2, as well as the fabric density which can be easily obtained with conventional equipment. In Section 4 we describe the datasets of *synthetic* and *real* samples, with mechanical ground truth parameters, used train and evaluate our regressor.

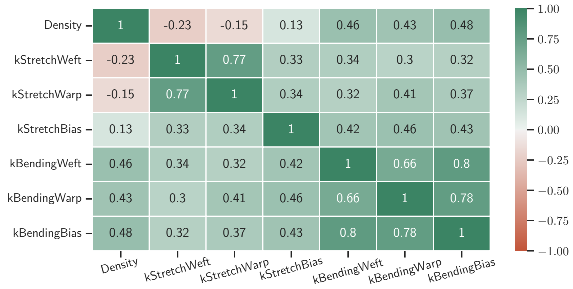
The quantitative evaluation suggests that our method is able to estimate the mechanics within a certain error. However, since a direct interpretation of that error is not human-friendly, we propose a method to evaluate the overall drape in the context of a real scene. In Section 6, we introduce our image-based similarity metric for drape, which takes as input renders of the chosen scenes and provides a relative value that is useful to compare the drape of different fabrics. With a user study, explained in Section 7.1, we validate that our metric agree with human preferences on the global perception of fabric drape. Also, in Section 7.2, we evaluate our capture method using our drape similarity metric and compare it with human judgments. We effectively validate that our estimations agree with human assessments and provide several qualitative examples in Section 7.3,

4. Datasets

We develop two different datasets of depth images, which we use at different steps of the pipeline to train and evaluate our models: a *synthetic* dataset, generated using physics-based cloth simulation; and *real* dataset, generated using images of real fabric samples. In both datasets, a sample consists of a depth image of a fabric simulated in a scene for which we have the corresponding mechanical parameters, namely: bending [GHDS03] and stretch [VMTF09] in the warp, weft, and bias directions and the fabric density, $\{kStretchWarp, kStretchWeft, kStretchBias, kBendinghWarp, kBendingWeft, kBendingBias, \rho\} \in \mathbb{R}^7$. We support two different static configurations for the scenes: *hanging*, which exhibits the overall drape; and *stretch*, which exhibits diagonal tension and is key to understanding the stretching properties. Figure 2 depicts examples of each configuration.

Simulated Dataset. To train our model, we generate a synthetic dataset by simulating fabrics in a virtual scenario, replicating the *hanging* and *stretch* configurations. We use a standard simulator, similar to ARCSim [NSO12], with a quadratic strain, a linear strain/stress relationship, and standard definitions for bending and stretch [GHDS03, VMTF09]. To model highly anisotropic fabrics, we use three parameters for warp, weft, and bias. Note that the model used for stretch already has some nonlinear behavior (quadratic strain), but more parameters (one per direction) are required to control the nonlinearity of the forces. Thickness is not included as it is implicitly accounted for in the other parameters. Contrarily, density is required as the simulator is dynamic. In a static one, it could be dropped after normalizing the other parameters.

After simulation, we render the resulting mesh (discretized at 5mm/edge) with a white Lambertian material and extract the depth buffer. To ensure that our synthetic dataset covers a wide range of materials, we densely sample the parameter space using a distribution of common mechanical parameters of real-world fabrics. Figure 5 shows a sweep of parameters showcasing the variability of resulting drapes in the hanging and stretch scenes. To better understand potential relationships between the parameters, we compute the Spearman correlation r, shown in Figure 4, where we observe the higher correlation between the three kStretch coefficients and some correlation between the kBendingBias and the density.



	Density	kStretchWeft	kStretchWarp	kStretchBias	kBendingWeft	kBendingWarp	kBendingBias
Density	1	-0.23	-0.15	0.13	0.46	0.43	0.48
kStretchWeft	-0.23	1	0.77	0.33	0.34	0.3	0.32
kStretchWarp	-0.15	0.77	1	0.34	0.32	0.41	0.37
kStretchBias	0.13	0.33	0.34	1	0.42	0.46	0.43
kBendingWeft	0.46	0.34	0.32	0.42	1	0.66	0.8
kBendingWarp	0.43	0.3	0.41	0.46	0.66	1	0.78
kBendingBias	0.48	0.32	0.37	0.43	0.8	0.78	1

Figure 4: Spearman correlation matrix between parameters of our synthetic dataset.

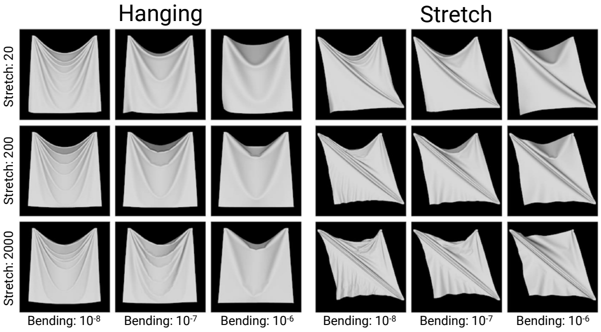


Figure 5: Sweep of simulation parameters for hanging and stretch scenes. For *kBending*: warp, weft, and bias have the same value, while for *kStretch*, bias changes as 100, 144, 1000.

Real Dataset. To evaluate our model, we test it with real data from images captured by the Intel RealSense SR300. To this end, we casually hang 50 × 50 cm fabric samples using several magnets into a metallic panel, which requires little to no expertise and can be done very fast. Pin location does not need to be centimeter-accurate. Because we use depth images, no special lighting is necessary. We measure fabric area density by weighing a 10 × 10 cm sample and dividing by its area. See Figure 2 for a visualization of our capture setup and the accompanying video for an illustration of the process. We capture ten fabrics of diverse compositions and structures, for which we have ground truth mechanical parameters previously obtained with specific equipment and methods [SSBL*22].

In the supplementary material, we include further details of this dataset, including each material's composition, structure, and closeups.

4.1. Data Augmentation

To make our model robust to potential noise and scenario variations common in uncontrolled capture setups (e.g., slightly different camera viewpoint or fabric configuration), we apply several data augmentation strategies to our *synthetic* dataset.

Simulation-Space Data Augmentation. To enforce robustness to different camera viewpoints, simulated meshes are rendered within a range of different inclinations with respect to the vertical plane.

This range covers ±5 degrees from the rest position orientation, creating 11 depth maps per material and scene.

Image-Space Data Augmentation. Real images are largely different to synthetic images, due to distortions, noise, perspective changes, unknown illumination, lens and sensor characteristics, blurs, etc. To enforce the robustness of our models to those distortions, we design an extensive image data augmentation policy consisting of random individual deformations, performed in a particular order. These not only include random noise, blurs, perspective changes and rescales, but also more complex policies such as thin-plate deformations, posterization and erasing. This data augmentation policy bridges the gap between synthetic renders and real depth images, which are typically more noisy. See supplementary material for more details.

5. Fabric Mechanics from Depth Images

In this section, we present our learning-based approach to estimate fabric mechanical parameters $\hat{\mathcal{P}}$ from depth images. Given a set of depth images, $\mathcal{I} = \{I_{\text{hanging}} \mid I_{\text{stretch}}\} \geq 1$, which depicts the *hanging* and *stretch* scenes, we train a model $\mathcal{M}$ which maps $\mathcal{I}$, along with the material *density* ρ, to mechanical parameters: $\mathcal{M}(\mathcal{I}, \rho) = \hat{\mathcal{P}} \in \mathbb{R}^6$. To this end, at train time we learn to extract relevant features from depth images, which are then fed into a regressor to learn to predict mechanical features. Importantly, our architecture enables to input sets of images at test time by fusing their respective features. Figure 6 illustrates the train and test pipelines. See the supplementary material for implementation details.

5.1. Neural Network Architecture

Feature Extractor The first part of the model is a *feature extractor* $\mathcal{F}$, which receives as input a single depth image $I_{\text{sc}} \in \mathcal{I}$ and outputs a feature vector $\mathbf{f}_{\text{sc}} = \mathcal{F}(I_{\text{sc}})$ that describes it. This feature extractor is composed of three different components, shown in Figure 6. First, the image is processed by a Convolutional Neural Network (CNN) that outputs a dense latent representation. We use a ResNet-18 [HZRS16] pre-trained on ImageNet [DDS*09], which we finetune with our data. To reduce the gap between real and synthetic images, we equalize the image before feeding it to the feature extractor, during both training and evaluation. Then, the output of this CNN is passed through a *Self-Attention* [ZGMO19] module, which helps the model learn non-local dependencies, enlarging the model receptive field, so it accounts for distant information in the input images. Self-Attention mechanisms were originally designed for language models [VSP*17] but have recently demonstrated significant efficacy for computer vision tasks [RPV*19,ZGMO19,ZJK20,HWC*22]. We add a single Self-Attention layer, as they are expensive to train and evaluate.

		Baseline	Data Augmentation		Architecture		
Metric	Parameter		w/ sim	w/ image	w/ pre-Train	w/ attention	w/ pooling
$\ell_1 \downarrow$	kStretchWeft	**0.122**	0.118	0.112	0.087	**0.071**	**0.071**
	kStretchWarp	**0.109**	0.101	0.102	0.074	0.066	**0.061**
	kStretchBias	**0.050**	0.043	0.046	0.043	0.042	**0.038**
	Avg. Stretch	**0.094**	0.087	0.087	0.068	0.060	**0.057**
	kBendingWeft	**0.095**	0.093	0.091	0.084	0.0.82	**0.072**
	kBendingWarp	**0.124**	0.119	0.112	0.110	0.101	**0.094**
	kBendingBias	**0.086**	0.081	**0.086**	0.081	0.068	**0.060**
	Avg. Bending	**0.102**	0.098	0.097	0.092	0.084	**0.075**
$r \uparrow$	kStretchWeft	**0.445**	0.475	0.453	0.643	0.788	**0.798**
	kStretchWarp	0.543	**0.503**	0.510	0.645	0.715	**0.771**
	kStretchBias	**0.477**	0.508	0.608	0.701	0.733	**0.781**
	Avg. Stretch	**0.488**	0.495	0.523	0.660	0.745	**0.783**
	kBendingWeft	**0.611**	0.684	0.781	0.783	0.798	**0.863**
	kBendingWarp	**0.614**	0.654	0.747	0.772	0.806	**0.921**
	kBendingBias	**0.624**	0.828	0.837	0.872	0.893	**0.942**
	Avg. Bending	**0.616**	0.722	0.788	0.809	0.832	**0.909**

Table 1: Ablation study of the neural architecture and data augmentation. From left to right, we build upon our baseline and progressively add: simulation-space data augmentation, image-space data augmentation, pre-training, self-attention, and average-pooling. On both MAE (ℓ_1) and correlation (r) metrics, we observe increased performance on the validation set in every added component. Using a pre-trained network for feature extraction yields the largest gains. We use a color code to highlight **best** and **worst** cases.

Finally, we perform pooling operations to transform the output of the Self-Attention layer to a feature vector, $\mathbf{f}_{\text{sc}}$, of a fixed size. In addition to the commonly used max-pooling [SZ14], we further concatenate it with the output of average-pooling, which has shown to improve performance of attention modules [ZKL*16, HSS18, WPLK18]. The feature vector $\mathbf{f}_{\text{sc}}$ is thus a concatenation of max-pooled features and average-pooled features: $\mathbf{f}_{\text{sc}} = \{\mathbf{f}_{\text{sc}}^{max} \oplus \mathbf{f}_{\text{sc}}^{avg}\}$.

Fusion Our design allows us to combine features $\mathbf{f}_{\text{sc}}$ from more than one scene into a single feature vector $\mathbf{f}$. For every image in I, we compute $\mathbf{f}_{\text{sc}}$. We then *fuse* those feature vectors into a single vector by performing pooling across $\mathbf{f}$. Similarly to $\mathbf{f}_{\text{sc}}$, $\mathbf{f}$ is composed of two types of features: $\mathbf{f} = \{max_{sc}\{\mathbf{f}_{sc}^{max}\} \oplus \{avg_{sc}\{\mathbf{f}_{sc}^{avg}\}\}$. The max-pooled features are fused using the maximum value across scenes, while the average-pooled features are fused using the mean across scenes. We illustrate this procedure in Figure 6.

Parameter Regressor Our last component is a fully-connected Multi-Layer Perceptron (MLP), which takes as input the feature vector $\mathbf{f}$ and the material density, ρ, and outputs the simulation parameters $\hat{\mathcal{P}}$. As our loss function, we compare the real parameters $\mathcal{P}$ with the model estimations $\mathcal{M}(\mathcal{I}, \rho) = \hat{\mathcal{P}}$ using an ℓ_2 norm.

5.2. Quantitative Evaluation

In this section we quantitatively evaluate the performance of the method for estimating the mechanical parameters. We validate the design choices of the model, and evaluate our results depending on the type of input used.

5.2.1. Ablation Study of the Model Design

We aim to understand the effective contribution of the data augmentation strategy, and the network architecture design. For these

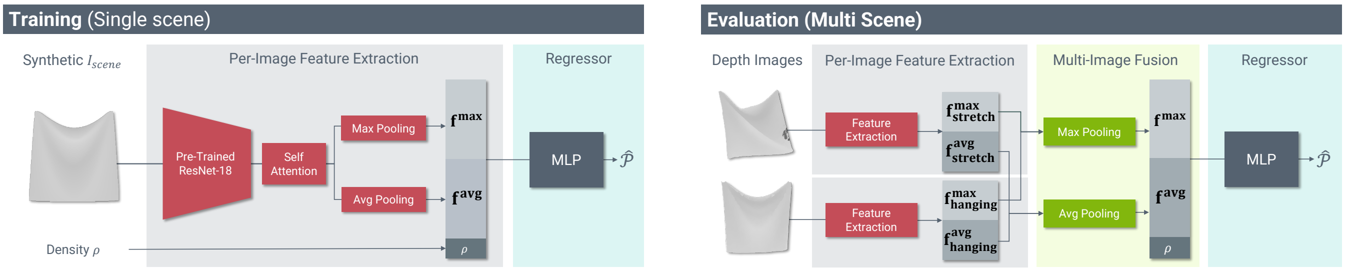


Figure 6: Diagram of our training and evaluation pipelines. For **training**, we use a single image of the material, along with its density. The image is processed by our *Feature Extractor*, followed by an MLP, which computes the parameter estimation $\hat{\mathcal{P}}$. For **evaluation**, we process each available image with our trained feature extractor and use a *fusion* operator before feeding it to the trained regressor. We use the same Feature Extractor and MLP for both scenes.

experiments, we randomly split the synthetic data in 90% for training and 10% for validation, using the same split for every experiment. Results are shown in Table 1. As our baseline, we use a simple model without *self-attention*, without *average-pooling* features, where the CNN backbone is randomly initialized, and without data augmentation. From this baseline, we progressively add different components and measure the performance of the validation data using Mean Absolute Error (MAE) and Spearman correlation (r). The parameters are normalized using the minimum and maximum values of the training set.

Given the training configuration with all the data augmentation –which provides a small increase in performance most likely because the validation dataset is synthetic data– we evaluate the neural architecture.

Using a CNN backbone pre-trained on ImageNet [DDS*09] instead of a randomly initialized one, we observe a significant increase in model performance across every parameter and metric. Training a feature extractor that receives images from both scenes at the same time would not allow us to leverage pre-training, which would negatively impact generalization. Then, we add a *self-attention* [ZGMO19] layer after the CNN backbone, which allows the model to integrate information that is present on distant areas of the images. Interestingly, this module significantly helps predict the *kStretch* parameters while having a more minor influence on the *kBending*. Finally, adding *average-pooling* in addition to the commonly used *max-pooling* have a highly positive impact in the error rates. We use this last configuration with all the components for all the results shown in the paper. It is worth noting that the *kBendingBias* is easily predicted by the model, even in its most basic configuration. This is likely because this parameter correlates most strongly with the density of the material (shown in Figure 4), so the model can leverage this information for the predictions.

5.2.2. Evaluation of Input Influence

The design of our method supports taking as input one or multiple images. In this experiment, presented in Table 2, we evaluate the error testing different configurations of the input. Note that we train a different model for each configuration. In every case, using the *density* as input helps the model to generalize. This is particularly relevant for *kBending* parameters, for which the *density* alone provides more information than depth images. The *stretch* scene is typically more informative than the *bending* one, as both MAE and correlations are usually better when it is provided. When using both scenes simultaneously, the model provides more accurate estimations than any of the scenes individually, showing that the two scenes provide complimentary information.

		Density	Only Depth			Density & Depth		
Metric	Parameter		Stretch	Hanging	Both	Stretch	Hanging	Both
$\ell_1 \downarrow$	kStretchWeft	**0.113**	0.102	0.107	0.062	0.054	0.056	**0.051**
	kStretchWarp	**0.091**	0.067	0.081	0.056	0.055	0.059	**0.052**
	kStretchBias	0.034	0.039	**0.054**	0.034	0.033	0.036	**0.031**
	Mean Stretch	0.079	0.069	**0.081**	0.051	0.047	0.050	**0.045**
	kBendingWeft	0.142	**0.233**	0.213	0.145	0.128	0.139	**0.125**
	kBendingWarp	0.126	**0.184**	0.126	0.074	0.037	0.063	**0.035**
	kBendingBias	0.094	**0.166**	0.081	0.054	0.055	0.069	**0.046**
	Mean Bending	0.121	**0.194**	0.140	0.091	0.073	0.090	**0.069**
$r \uparrow$	kStretchWeft	**0.184**	0.407	0.403	0.418	0.712	0.469	**0.728**
	kStretchWarp	**0.002**	0.502	0.407	0.503	0.520	0.433	**0.533**
	kStretchBias	0.289	**-0.141**	-0.068	-0.007	0.367	0.383	**0.550**
	Mean Stretch	**0.158**	0.256	**0.247**	0.305	0.533	0.428	**0.604**
	kBendingWeft	0.483	**0.052**	0.267	0.625	0.673	0.683	**0.717**
	kBendingWarp	0.317	**0.048**	0.156	0.407	0.467	0.433	**0.533**
	kBendingBias	0.357	**-0.044**	0.108	0.250	0.333	0.417	**0.546**
	Mean Bending	0.386	**0.019**	0.177	0.427	0.491	0.511	**0.599**

Table 2: Results for real depth images varying the input. From left to right, the input is: only density, only depth images, and both density and depth. The best results are obtained using every data source as input. For *kBending*, the *density* alone provides more information than only depth images. We use a color code to highlight **best** and **worst** cases.

5.2.3. Neural Saliency Maps

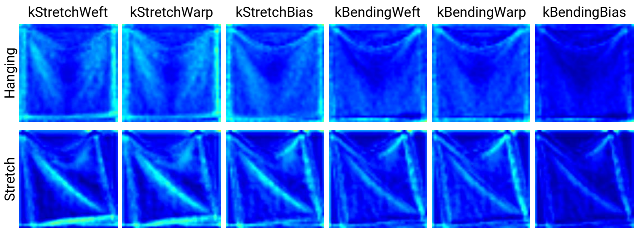


Figure 7: Saliency maps [FHD*20] aggregated per parameter. The model relies on the central areas of the fabric samples for predicting the stretch parameters. For bending, it is most sensitive to areas on the borders of the samples.

We aim to understand which part of the scenes are most relevant for the model when making its predictions. To do so, we create saliency maps using *Axion-Based Class-Activation Mappings* [FHD*20], that averages the activations of the deep layers weighted by their importance with respect to each target parameter, and aggregate them over our real dataset. In Figure 7 we observe that each scene provides the model with different cues. For *kStretch* parameters, the model is sensitive to the central wrinkle of the *stretch* scene and the central fold of the *hanging*. For *kBending* parameters, the model is sensitive to areas in the borders of the fabric where small but noticeable wrinkles are present, in both scenes. As in other experiments, we observe that the model can find more relevant features on the *stretch* scene.

6. A Similarity Metric for Drape

The regressor introduced in Section 5 allows to infer the mechanical parameters of target fabric. However, since a direct interpretation of such parametric space is not human friendly, it is nearly impossible to understand the residual errors shown in Tables 1 and 2. Do the regressed parameters produce a drape similar to the target image? Notice that, since the mechanical parameter space is non-orthogonal, small parameter changes may produce unexpected deformations. Therefore, we hypothesize that a *perceptual* similarity metric for drape is needed to interpret our quantitative results. We describe this metric next.

6.1. Image-based Similarity of Drape

Motivated by our hypothesis that the *hanging* and *stretch* scenes are sufficient to convey the fabric mechanics, we propose an image-based similarity metric using renders of such scenes. Let $\mathcal{P} \subseteq \mathbb{R}^7$ be the parameter space of our simulator, $P_a \in \mathcal{P}$ and $P_b \in \mathcal{P}$ two different parameter sets, and $a \sim \mathcal{R}(P_a, \text{scene})$ and $b \sim \mathcal{R}(P_b, \text{scene})$ two rendered simulations obtained for a certain scene configuration. We define a distance metric for a particular scene as

$$d_{\text{scene}}(P_a, P_b) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \text{IM}(a_i, b_j)}{N^2} \tag{1}$$

where IM is an image-space distance metric, and N is the number of different simulations we run. Since real cloth is very sensitive to parameters such as initial state or initial shape, in order to learn a metric that is robust to real-world conditions, we perturb the initial state and boundary conditions in a set of simulations using random jittering to the initial forces. We empirically found that averaging over multiple simulations for the same set of (P_a, P_b) gives us a more informative metric.

Further more, we take into account both *hanging* and *stretch* scenes, hence our final metric is defined by averaging their distances across both scenarios, resulting in our final metric:

$$d(P_a, P_b) = \frac{d_{\text{hanging}}(P_a, P_b) + d_{\text{stretch}}(P_a, P_b)}{2} \tag{2}$$

Note that we propose a similarity metric, which, as opposed to real distance metrics, does not necessarily have to meet the metric axioms [TK74]: it can produce asymmetric values, violate the triangle inequality, and does not need to define what the identity is. According to the Equation 1, the distance of one fabric with itself is not necessarily zero; it just needs to satisfy a minimum requirement where the distance of every material with itself should be smaller than the distance of any material with any other material,

$$d(P_a, P_a) < d(P_a, P_b), \forall P_a \neq P_b \tag{3}$$

In order to remove the possible influence of optical properties, scene illumination, and camera parameters, we use the same scene configuration for every render, with grayscale albedo and a lambertian BRDF.

For IM we use LPIPS [ZIE*18] which we empirically found to perform better than other alternative image metrics. We find that metrics based on pre-trained neural networks work better than lower-level alternatives, while content-aware distances are more powerful for this purpose than style-aware metrics. This suggests that the size, position and shape of the wrinkles and deformations of the fabrics are important factors that explain differences between materials. We empirically found that $N = 5$ simulations is typically enough, as more samples provide very marginal improvements. See the supplementary material for more details about the proposed metric.

7. Evaluation

We propose to evaluate our method from Section 5 and the metric from Section 6 by comparing our estimations with human ratings (recall that we provide quantitative errors per parameter in Section 5.2). To this end, in Section 7.1, we first collect a large number of ground truth human judgments about the similarity of triplets of fabrics. Then, in Section 7.2, we demonstrate that our image-based metric using ground truth parameters, as well as the estimated ones, encode the same preferences. Finally, in Section 7.3, we show qualitative comparisons and demonstrate the usefulness of our approach in a downstream task consisting of ‘search by similarity’.

7.1. Human Judgment Perceptual Similarity of Drape

We then use ten samples from our real dataset with known ground truth mechanical parameters and setup the user study as follows. Participants are presented with a triplet of fabrics and, using one fabric as a reference, they are asked which of the two remaining fabrics is most similar [ZIE*18, GAGH14] to the reference fabric. Participants are encouraged to manipulate the samples and focus only on the mechanical similarity and overall drape, and to ignore properties like material reflectance. Each participant rated 20 triplets that were pseudo-randomly sampled, ensuring that at least each of the ten test fabrics is used twice as reference. Given the same triplet, we observe an average of 86.68% agreement between our participants across all experiments and materials, suggesting that there is a perceptual understanding of fabrics mechanics that humans share. We did not observe any significant differences in agreement depending on the volunteer demographics or level of expertise in fabric handling or simulation.

Leveraging the user study described above, we can compute an embedding that captures the relative distance between real materials according to human perception (i.e., ground truth perceptual similarity). Figure 8 depicts such embedding in 2D, computed using tSTE [TLB*11], which allows us to calculate perceptual distances between materials using the Euclidean norm. The embed-

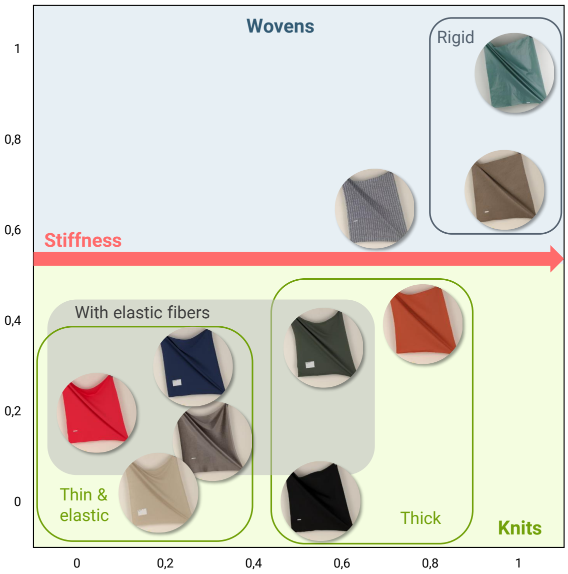


Figure 8: Human Judgments 2D tSTE embedding [TLB*11] computed from human perceptual judgments about real fabric similarity (i.e., ground truth). We observe interesting patterns: woven and knits are separated; elastic materials are clustered; thick and thin materials are separated. Neither axes directly correspond to any material property, instead they emerge from the embedding.

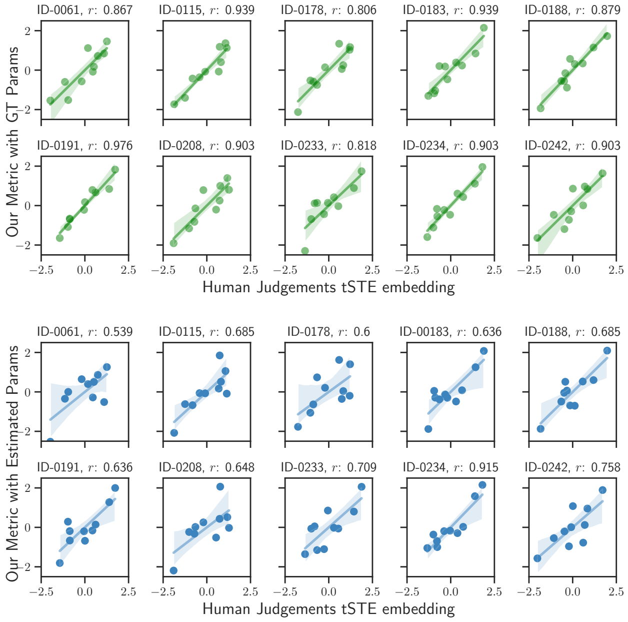


Figure 9: Correlation between the ordering provided by the Human Judgments (x-axis) and our drape similarity metric with (top) the Ground Truth simulation parameters (y-axis), and (bottom) the estimations of our model. We plot z-scores instead of the raw distances to help visualization.

	Parameter Distance	Similarity Metric
GT	0.431 ± 0.17	**0.893**±0.05
Estimated	0.377 ± 0.19	**0.680**±0.09

Table 3: Average (± std.) correlation with rankings obtained through the Human Judgments tSTE Embedding, depending on the parameter source (ground truth or predicted), and metric used to compute similarity (parameter distance or our drape similarity).

ding depicts many interesting patterns, including perfectly separated woven and knitted fabrics; elastic and thin fabrics are clustered together; thick and thin knits are well separated. These patterns suggest that fabric structure, composition, and density play an important, non-linear role in the overall perception of the mechanical properties of fabrics.

7.2. Image-based vs. Human Perceptual Similarity

We evaluate the agreement between humans and our estimations within the context of a similarity rank. For each fabric of our real dataset, we compute the distance to the rest of the materials using different metrics: 1) the Euclidean distances on the Human Judgments tSTE embedding shown Section 7.1; 2) the z-score distances in the parametric space of the mechanical simulation, $\mathcal{P}$; 3) the distance using our similarity metric for drape explained in Section 6. We compare ground truth parameters and estimated ones for the second and third cases. The summary of results is shown in Table 3, and the complete analysis is presented in the supplementary material.

First, we demonstrate that our similarity metric using ground truth parameters correlates with human judgments. Figure 9 (top) illustrates the outcome using Spearman correlation (r). We observe that the correlation for each fabric is higher than 0.8, with an average of 0.893, showcasing a strong correlation. These results suggest that our metric, which only takes images of the *hanging* and *stretch* scenes as input, can measure distances between materials as humans would. Then, as shown in Figure 9 (bottom), we use our estimated parameters instead of the ground truth, reaching an average correlation of 0.680. Even though this value is slightly smaller than the ground truth, it is still significant to conclude the estimations of our model agree with human judgments. Note that the materials with higher correlation are those lying on the extreme areas of the embedding obtained in Figure 8 that have very distinct characteristics. Likewise, we also compute the distance for each material using merely the parameter spaces of the simulation. As can be seen, using this space does not produce correlated outputs with human judgments, reaching correlations below 0.44 in any case tested.

7.3. Qualitative Results

Figure 10 compares the simulations obtained with the parameters of our method with the ground truth parameters for a few examples. We can observe that our estimations are very close to the ground truth in this cases. The full results are contained in the supplementary material. Finally, our metric can be used to search between materials of similar drape. We illustrate this in Figure 12. As shown, a naive ranking using directly the parameter space does not provide any meaningful ordering. On the contrary, using our similarity metric, we obtain ranks that agree with those given by the human

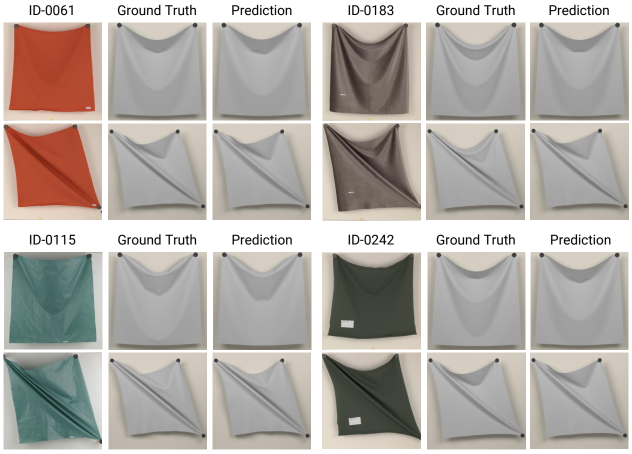


Figure 10: A comparison between the simulations obtained through the ground truth parameters, and those obtained using the predictions of our model, from a representative set of fabrics of our test set. As shown, the estimations of the model yield similar drapes to those of their ground truth counterparts, which we also evaluate quantitatively.

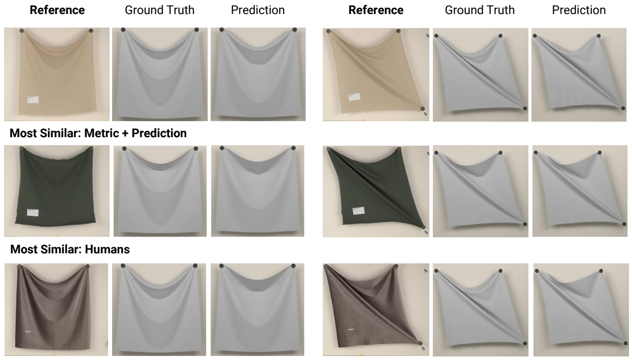


Figure 11: A failure case of our method. Given the reference material (first row) as input, the model predicts fewer bends than the ground truth. According to our metric, this prediction is closer to a thicker material (middle row), than to what humans perceive as most similar to the reference fabric (bottom row).

embedding, showcasing the potential of our automatic metric to explore fabric collections.

Limitations Even if our model provides accurate predictions, its estimations are not always truthful to the real materials. We illustrate this in Figure 11, where the model predicts fewer bends on the final drape than what the ground truth generates. According to our metric, this prediction is closer to thicker materials than to what humans perceive as most similar to the reference material.

8. Conclusions

In this work, we have presented a casual method to estimate mechanical parameters of fabrics from depth images of fabric samples placed at two specific configurations. We have validated our architecture and inputs numerically, proving that all the components of our method are necessary to provide accurate estimations. While our quantitative analysis helped us understand the importance of each component, we found that these errors are not interpretable, nor do they help us understand the overall appearance of the predicted drape. Therefore, we have presented the first metric, which, by purely working on the image space, can capture differences in fabric mechanics like humans do. We have used such metric for two purposes: first, to validate the accuracy of our estimated parameters perceptually, and second, to showcase a novel application of search by drape similarity.

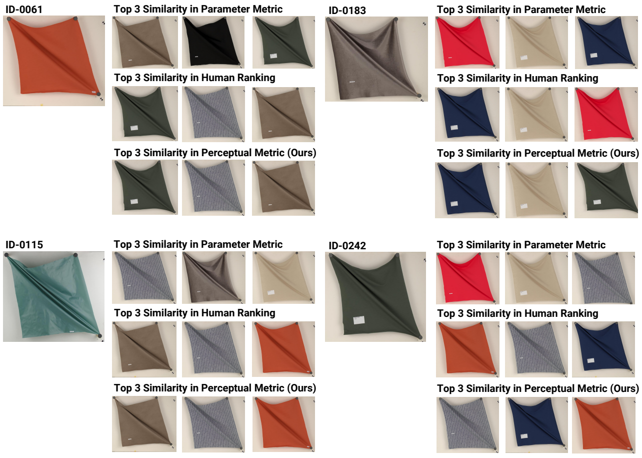

ID-0061
Top 3 Similarity in Parameter Metric
Top 3 Similarity in Human Ranking
Top 3 Similarity in Perceptual Metric (Ours)
ID-0183
Top 3 Similarity in Parameter Metric
Top 3 Similarity in Human Ranking
Top 3 Similarity in Perceptual Metric (Ours)
ID-0115
Top 3 Similarity in Parameter Metric
Top 3 Similarity in Human Ranking
Top 3 Similarity in Perceptual Metric (Ours)
ID-0242
Top 3 Similarity in Parameter Metric
Top 3 Similarity in Human Ranking
Top 3 Similarity in Perceptual Metric (Ours)

Figure 12: Search by drape similarity. The ordering provided by the parameter space (first row) does not match human judgments (second row), while the arrangement obtained by our metric matches humans with high consensus (third row).

Our work could be improved in several ways. Our neural network is trained using a purely regression loss. Training the network using differentiable simulation could improve training, and help generalization and error interpretation. We could incorporate our perceptual metric as a loss function. However, it requires multiple differentiable simulations and a deep feature extractor, which will result in a significant computational overhead. In addition, less expressive simulation engines may correlate less with human perception. Similarly, we would like to scale our training dataset and user study to handle more and more diverse samples, to cover a broader variety of fabric families. Interesting possible extensions would include taking as input RGB images instead of depth maps, training with real samples, incorporating symmetry consistency losses, or learning a similarity metric that can work by using captured images as input (instead of simulations). Finally, we hope our work might inspire future work in the long-standing problem of validating fabric mechanics in a way that is agnostic to the simulator parameters.

Acknowledgments We would like to thank Gabriel Cirio and Alejandro Rodríguez for valuable discussions, Alicia Nicas for her help with the simulations, and Sofía Domínguez for her help capturing data. Elena Garces was partially supported by a Juan de la Cierva - Incorporacion Fellowship (IJC2020-044192-I).

References

[BHA*21] Bommasani R., Hudson D. A., Adeli E., Altman R., Arora S., von Arx S., Bernstein M. S., Bohg J., Bosselut A., Brunskill E., et al.: On the opportunities and risks of foundation models. 3

[BJNX18] Bi W., Jin P., Nienborg H., Xiao B.: Estimating mechanical properties of cloth from videos using dense motion trajectories: Human psychophysics and machine learning. *Journal of vision 18*, 5 (2018), 12–12. 3

[BTH*03] Bhat K. S., Twigg C. D., Hodgins J. K., Khosla P. K., Popovic Z., Seitz S. M.: Estimating Cloth Simulation Parameters from Video. In *Symposium on Computer Animation* (2003), The Eurographics Association. 2

[BXBF13] Bouman K. L., Xiao B., Battaglia P., Freeman W. T.: Estimating the material properties of fabric from video. In *Proc. IEEE International Conference on Computer Vision* (2013), pp. 1984–1991. 2, 3

[CKNH20] Chen T., Kornblith S., Norouzi M., Hinton G.: A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (2020), PMLR, pp. 1597–1607. 3

[CPGE90] Clapp T. G., Peng H., Ghosh T. K., Eischen J. W.: Indirect measurement of the moment-curvature relationship for fabrics. *Textile Research Journal 60*, 9 (1990), 525–533. 2

[CTT17] Clyde D., Teran J., Tamstorf R.: Modeling and data-driven parameter estimation for woven fabrics. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2017), pp. 1–11. 2

[CXH21] Chen X., Xie S., He K.: An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 9640–9649. 3

[CZW*18] Chen Y., Zhang Q., Wu Y., Liu B., Wang M., Lin Y.: Fine-tuning resnet for breast cancer classification from mammography. In *The International Conference on Healthcare Science and Engineering* (2018), Springer, pp. 83–96. 3

[DBC*15] Davis A., Bouman K. L., Chen J. G., Rubinstein M., Durand F., Freeman W. T.: Visual vibrometry: Estimating material properties from small motion in video. In *Proceedings of the ieee conference on computer vision and pattern recognition* (2015), pp. 5335–5343. 3

[DDS*09] Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L.: Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (2009), IEEE, pp. 248–255. 5, 6

[DMWS20] Ding K., Ma K., Wang S., Simoncelli E. P.: Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence* (2020). 3

[FHD*20] Fu R., Hu Q., Dong X., Guo Y., Gao Y., Li B.: Axiom-based grad-cam: Towards accurate visualization and explanation of cnns. In *31st British Machine Vision Conference 2020, BMVC 2020,* (2020), BMVA Press. 6, 7

[FHXW22] Feng X., Huang W., Xu W., Wang H.: Learning-based bending stiffness parameter estimation by a drape tester. *ACM Transactions on Graphics (TOG) 41*, 6 (2022), 1–16. 3

[GAGH14] Garces E., Agarwala A., Gutierrez D., Hertzmann A.: A similarity measure for illustration style. *ACM Transactions on Graphics (TOG) 33*, 4 (2014), 1–9. 2, 3, 7

[GEB16] Gatys L. A., Ecker A. S., Bethge M.: Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 2414–2423. 3

[GHDS03] Grinspun E., Hirani A. N., Desbrun M., Schröder P.: Discrete shells. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation* (2003), Citeseer, pp. 62–67. 4

[GSK*19] Guo Y., Shi H., Kumar A., Grauman K., Rosing T., Feris R.: Spottune: transfer learning through adaptive fine-tuning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), pp. 4805–4814. 3

[HAL*19] Hu Y., Anderson L., Li T.-M., Sun Q., Carr N., Ragan-Kelley J., Durand F.: Difftaichi: Differentiable programming for physical simulation. In *International Conference on Learning Representations* (2019). 3

[HCX*22] He K., Chen X., Xie S., Li Y., Dollár P., Girshick R.: Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 16000–16009. 3

[HEW17] Huber M., Eberhardt B., Weiskopf D.: Cloth animation retrieval using a motion-shape signature. *IEEE Computer Graphics and Applications 37*, 6 (2017), 52–64. 3

[HSS18] Hu J., Shen L., Sun G.: Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 7132–7141. 5

[HWC*22] Han K., Wang Y., Chen H., Chen X., Guo J., Liu Z., Tang Y., Xiao A., Xu C., Xu Y., et al.: A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence* (2022). 5

[HZRS16] He K., Zhang X., Ren S., Sun J.: Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778. 5

[JBH19] Jaques M., Burke M., Hospedales T.: Physics-as-inverse-graphics: Unsupervised physical parameter estimation from video. In *International Conference on Learning Representations* (2019). 3

[JC20] Ju E., Choi M. G.: Estimating cloth simulation parameters from a static drape using neural networks. *IEEE Access 8* (2020), 195113–195121. 2, 3

[Kaw80] Kawabata S.: The standardization and analysis of hand evaluation. *The Textile Machinery Society of Japan* (1980). 2

[KBZ*20] Kolesnikov A., Beyer L., Zhai X., Puigcerver J., Yung J., Gelly S., Houlsby N.: Big transfer (bit): General visual representation learning. In *European conference on computer vision* (2020), Springer, pp. 491–507. 3

[KLBG20] Kuijpers S., Luible-Bär C., Gong R. H.: The measurement of fabric properties for virtual simulation—a critical review. *IEEE SA INDUSTRY CONNECTIONS* (2020), 1–43. 2

[KRJ*22] Kumar A., Raghunathan A., Jones R., Ma T., Liang P.: Fine-tuning can distort pretrained features and underperform out-of-distribution. *arXiv preprint arXiv:2202.10054* (2022). 3

[KSL19] Kornblith S., Shlens J., Le Q. V.: Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), pp. 2661–2671. 3

[LDW*22] Li Y., Du T., Wu K., Xu J., Matusik W.: Diffcloth: Differentiable cloth simulation with dry frictional contact. *ACM Transactions on Graphics (TOG)* (2022). 3

[LGG19] Lagunas M., Garces E., Gutierrez D.: Learning icons appearance similarity. *Multimedia tools and applications 78*, 8 (2019), 10733–10751. 3

[LKS15] Lun Z., Kalogerakis E., Sheffer A.: Elements of style: learning perceptual shape style similarity. *ACM Transactions on graphics (TOG) 34*, 4 (2015), 1–14. 3

[LLB22] Li W.-H., Liu X., Bilen H.: Cross-domain few-shot learning with task-specific adapters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 7161–7170. 3

[LLK19] LIANG J., LIN M., KOLTUN V.: Differentiable cloth simulation for inverse problems. *Advances in Neural Information Processing Systems 32* (2019). 3

[LMS*19] LAGUNAS M., MALPICA S., SERRANO A., GARCES E., GUTIERREZ D., MASIA B.: A similarity measure for material appearance. *ACM Transactions on Graphics (TOG) 38*, 4 (2019), 1–12. 2, 3

[LMT08] LUIBLE C., MAGNENAT-THALMANN N.: The simulation of cloth using accurate physical parameters. In *Proceedings of the Tenth IASTED International Conference on Computer Graphics and Imaging (CGIM '08)* (2008). 2

[MBT*12] MIGUEL E., BRADLEY D., THOMASZEWSKI B., BICKEL B., MATUSIK W., OTADUY M. A., MARSCHNER S.: Data-driven estimation of cloth simulation models. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 519–528. 2

[Min95] MINAZIO P. G.: Fast–fabric assurance by simple testing. *International Journal of Clothing Science and Technology* (1995). 2

[MMG*20] MURTHY J. K., MACKLIN M., GOLEMO F., VOLETI V., PETRINI L., WEISS M., CONSIDINE B., PARENT-LÉVESQUE J., XIE K., ERLEBEN K., ET AL.: gradsim: Differentiable simulation for system identification and visuomotor control. In *International Conference on Learning Representations* (2020). 3

[MTLVL07] MAGNENAT-THALMANN N., LUIBLE C., VOLINO P., LYARD E.: From measured fabric to the simulation of cloth. In *2007 10th IEEE International Conference on Computer-Aided Design and Computer Graphics* (2007), IEEE, pp. 7–18. 2

[NSHC16] NAFCHI H. Z., SHAHKOLAEI A., HEDJAM R., CHERIET M.: Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator. *Ieee Access 4* (2016), 5579–5590. 3

[NSO12] NARAIN R., SAMII A., O'BRIEN J. F.: Adaptive anisotropic remeshing for cloth simulation. *ACM transactions on graphics (TOG) 31*, 6 (2012), 1–10. 4

[PCMS18] PRASHNANI E., CAI H., MOSTOFI Y., SEN P.: Pieapp: Perceptual image-error assessment through pairwise preference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 1808–1817. 3

[PCYS20] PHAM M. Q., CREGO J.-M., YVON F., SENELLART J.: A study of residual adapters for multi-domain neural machine translation. In *Conference on Machine Translation* (2020). 3

[Pow13] POWER J.: Fabric objective measurements for commercial 3d virtual garment simulation. *International Journal of Clothing Science and Technology* (2013). 2

[RBKW18] REISENHOFER R., BOSSE S., KUTYNIOK G., WIEGAND T.: A haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication 61* (2018), 33–43. 3

[RBV17] REBUFFI S.-A., BILEN H., VEDALDI A.: Learning multiple visual domains with residual adapters. *Advances in neural information processing systems 30* (2017). 3

[RBV18] REBUFFI S.-A., BILEN H., VEDALDI A.: Efficient parametrization of multi-domain deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 8119–8127. 3

[RD17] RATTANI A., DERAKHSHANI R.: On fine-tuning convolutional neural networks for smartphone based ocular recognition. In *2017 IEEE international joint conference on biometrics (IJCB)* (2017), IEEE, pp. 762–767. 3

[RGSS20] RUNIA T. F., GAVRILYUK K., SNOEK C. G., SMEULDERS A. W.: Cloth in the wind: A case study of physical measurement through simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 10498–10507. 3

[RJ19] REDDY A. S. B., JULIET D. S.: Transfer learning with resnet-50 for malaria cell-image classification. In *2019 International Conference on Communication and Signal Processing (ICCSP)* (2019), IEEE, pp. 0945–0949. 3

[RKH*21] RADFORD A., KIM J. W., HALLACY C., RAMESH A., GOH G., AGARWAL S., SASTRY G., ASKELL A., MISHKIN P., CLARK J., ET AL.: Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning* (2021), PMLR, pp. 8748–8763. 3

[RPB19] RODRÍGUEZ-PARDO C., BILEN H.: Personalised aesthetics with residual adapters. In *Iberian Conference on Pattern Recognition and Image Analysis* (2019), Springer, pp. 508–520. 3

[RPV*19] RAMACHANDRAN P., PARMAR N., VASWANI A., BELLO I., LEVSKAYA A., SHLENS J.: Stand-alone self-attention in vision models. *Advances in Neural Information Processing Systems 32* (2019). 5

[RRBD*20] RASHEED A. H., ROMERO V., BERTAILS-DESCOUBES F., WUHRER S., FRANCO J.-S., LAZARUS A.: Learning to measure the static friction coefficient in cloth contact. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 9912–9921. 3

[RZKB19] RAGHU M., ZHANG C., KLEINBERG J., BENGIO S.: Transfusion: Understanding transfer learning for medical imaging. *Advances in neural information processing systems 32* (2019). 3

[SB08] SYLLEBRANQUE C., BOIVIN S.: Estimation of mechanical parameters of deformable solids from videos. *The Visual Computer 24*, 11 (2008), 963–972. 2

[SSBL*22] SPERL G., SÁNCHEZ-BANDERAS R. M., LI M., WOJTAN C., OTADUY M. A.: Estimation of yarn-level simulation models for production fabrics. *ACM Transactions on Graphics (TOG) 41*, 4 (2022), 1–15. 4

[SZ14] SIMONYAN K., ZISSERMAN A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014). 5

[TK74] TVERSKY A., KAHNEMAN D.: Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science 185*, 4157 (1974), 1124–1131. 7

[TLB*11] TAMUZ O., LIU C., BELONGIE S., SHAMIR O., KALAI A. T.: Adaptively learning the crowd kernel. In *Proceedings of the 28th International Conference on International Conference on Machine Learning* (2011), pp. 673–680. 7, 8

[VMTF09] VOLINO P., MAGNENAT-THALMANN N., FAURE F.: A simple approach to nonlinear tensile stiffness for accurate cloth simulation. *ACM Transactions on Graphics 28*, 4 (2009), Article–No. 2, 4

[VSP*17] VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER Ł., POLOSUKHIN I.: Attention is all you need. *Advances in neural information processing systems 30* (2017). 5

[WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing 13*, 4 (2004), 600–612. 3

[WKW16] WEISS K., KHOSHGOFTAAR T. M., WANG D.: A survey of transfer learning. *Journal of Big data 3*, 1 (2016), 1–40. 3

[WLZ*18] WANG G., LI W., ZULUAGA M. A., PRATT R., PATEL P. A., AERTSEN M., DOEL T., DAVID A. L., DEPREST J., OURSELIN S., ET AL.: Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE transactions on medical imaging 37*, 7 (2018), 1562–1573. 3

[WOR11] WANG H., O'BRIEN J. F., RAMAMOORTHI R.: Data-driven elastic models for cloth: modeling and measurement. *ACM transactions on graphics (TOG) 30*, 4 (2011), 1–12. 2

[WPLK18] WOO S., PARK J., LEE J.-Y., KWEON I. S.: Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)* (2018), pp. 3–19. 5

[YL15] YANG S., LIN M. C.: Materialcloning: Acquiring elasticity parameters from images for medical applications. *IEEE transactions on visualization and computer graphics 22*, 9 (2015), 2122–2135. 3

[YLL17] Yang S., Liang J., Lin M. C.: Learning-based cloth material recovery from video. In *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 4383–4393. 2, 3

[YPA*18] Yang S., Pan Z., Amert T., Wang K., Yu L., Berg T., Lin M. C.: Physics-inspired garment recovery from a single-view image. *ACM Transactions on Graphics (TOG) 37*, 5 (2018), 1–14. 2

[ZDN16] Zhang H., Dana K., Nishino K.: Friction from reflectance: Deep reflectance codes for predicting physical surface properties from one-shot in-field reflectance. In *European Conference on Computer Vision* (2016), Springer, pp. 808–824. 3

[ZGMO19] Zhang H., Goodfellow I., Metaxas D., Odena A.: Self-attention generative adversarial networks. In *International conference on machine learning* (2019), PMLR, pp. 7354–7363. 5, 6

[ZGO*21] Zhu W., Guo X., Owaki D., Kutsuzawa K., Hayashibe M.: A survey of sim-to-real transfer techniques applied to reinforcement learning for bioinspired robots. *IEEE Transactions on Neural Networks and Learning Systems* (2021). 2

[ZIE*18] Zhang R., Isola P., Efros A. A., Shechtman E., Wang O.: The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 586–595. 2, 3, 7

[ZJK20] Zhao H., Jia J., Koltun V.: Exploring self-attention for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 10076–10085. 5

[ZKL*16] Zhou B., Khosla A., Lapedriza A., Oliva A., Torralba A.: Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 2921–2929. 5

[ZL12] Zhang L., Li H.: Sr-sim: A fast and high performance iqa index based on spectral residual. In *2012 19th IEEE international conference on image processing* (2012), IEEE, pp. 1473–1476. 3

[ZSL14] Zhang L., Shen Y., Li H.: Vsi: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image processing 23*, 10 (2014), 4270–4281. 3

[ZZMZ11] Zhang L., Zhang L., Mou X., Zhang D.: Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing 20*, 8 (2011), 2378–2386. 3