Learning-Based Animation of Clothing for Virtual Try-On

Igor Santesteban   Miguel A. Otaduy   Dan Casas

Universidad Rey Juan Carlos, Madrid, Spain

Figure 1: *Given a garment (left), we learn a deformation model that enables virtual try-on by bodies with different shapes and poses (middle). Our model produces cloth animations with realistic dynamic drape and wrinkles at* 250 *fps (right).*

Abstract

This paper presents a learning-based clothing animation method for highly efficient virtual try-on simulation. Given a garment, we preprocess a rich database of physically-based dressed character simulations, for multiple body shapes and animations. Then, using this database, we train a learning-based model of cloth drape and wrinkles, as a function of body shape and dynamics. We propose a model that separates global garment fit, due to body shape, from local garment wrinkles, due to both pose dynamics and body shape. We use a recurrent neural network to regress garment wrinkles, and we achieve highly plausible nonlinear effects, in contrast to the blending artifacts suffered by previous methods. At runtime, dynamic virtual try-on animations are produced in just a few milliseconds for garments with thousands of triangles. We show qualitative and quantitative analysis of results.

CCS Concepts

• *Computing methodologies → Physical simulation; Neural networks;*

1. Introduction

Clothing plays a fundamental role in our everyday lives. When we choose clothing to buy or wear, we guide our decisions based on a combination of fit and style. For this reason, the majority of clothing is purchased at brick-and-mortar retail stores, after physical try-on to test the fit and style of several garments on our own bodies. Computer graphics technology promises an opportunity to support online shopping through virtual try-on animation, but to date virtual try-on solutions lack the responsiveness of a physical try-on experience. Beyond online shopping, responsive animation of clothing has an impact on fashion design, video games, and interactive graphics applications as a whole.

One approach to produce animations of clothing is to simulate the physics of garments in contact with the body. While this

approach has proven capable of generating highly detailed results [KJM08, SSIF09, NSO12, CLMMO14], it comes at the expense of significant runtime computational cost. On the other hand, it bears no or little preprocessing cost, hence it can be quickly deployed on almost arbitrary combinations of garments and body shapes and motions. To fight the high computational cost, interactive solutions sacrifice accuracy in the form of coarse cloth discretizations, simplified cloth mechanics, or approximate integration methods. Continued progress on the performance of solvers is bringing the approach closer to the performance needs of virtual try-on [TWL*18].

An alternative approach for cloth animation is to train a data-driven model that computes cloth deformation as a function of body motion [WHRO10, dASTH10]. This approach succeeds to produce

plausible cloth folds and wrinkles when there is a strong correlation between body pose and cloth deformation. However, it struggles to represent the nonlinear behavior of cloth deformation and contact in general. Most data-driven methods rely to a certain extent on linear techniques, hence the resulting wrinkles deform in a seemingly linear manner (*e.g.*, with blending artifacts) and therefore lack realism.

Most previous data-driven cloth animation methods work for a given garment-avatar pair, and are limited to representing the influence of body pose on cloth deformation. In virtual try-on, however, a garment may be worn by a diverse set of people, with corresponding avatar models covering a range of body shapes. In this paper, we propose a learning-based method for cloth animation that meets the needs of virtual try-on, as it models the deformation of a given garment as a function of body motion *and* shape. Other methods that account for changes in body shape do not deform the garment in a realistic way, and either resize the garment while preserving its style [GRH*12, BSBC12], or retarget cloth wrinkles to bodies of different shapes [PMPHB17, LCT18].

We propose a two-level strategy to learn the complex nonlinear deformations of clothing. On one hand, we learn a model of garment fit as a function of body shape. And on the other hand, we learn a model of local garment wrinkles as a function of body shape and motion. Our two-level strategy allows us to disentangle the different sources of cloth deformation.

We compute both the garment fit and the garment wrinkles using nonlinear regression models, *i.e.*, artificial neural networks, and hence we avoid the problems of linear data-driven models. Furthermore, we propose the use of recurrent neural networks to capture the dynamics of wrinkles. Thanks to this strategy, we avoid adding an external feedback loop to the network, which typically requires a dimensionality reduction step for efficiency reasons [CO18].

Our learning-based cloth animation method is formulated as a pose-space deformation, which can be easily integrated into skeletal animation pipelines with little computational overhead. We demonstrate example animations such as the ones in Figure 1, with a runtime cost of just 4ms per frame (more than 1000x speed-up over a full simulation) for cloth meshes with thousands of triangles, including collision postprocessing.

To train our learning-based model, we leverage state-of-the-art physics-based cloth simulation techniques [NSO12], together with a parametric human model [LMR*15] and publicly available motion capture data [CMU, VRM*17]. In addition to the cloth animation model, we have created a new large dataset of dressed human animations of varying shapes and motions.

2. Related Work

Fast Cloth Simulation. Physics-based simulation of clothing entails three major processes: computation of internal cloth forces, collision detection, and collision response; and the total simulation cost results from the combined influence of the three processes. One attempt to limit the cost of simulation has been to approximate dynamics, such as in the case of position-based dynamics [BMO*14]. While approximate methods produce plausible and

expressive results for video game applications, they cannot transmit the realistic cloth behavior needed for virtual try-on.

Another line of work, which tries to retain simulation accuracy, is to handle efficiently both internal forces and collision constraints during time integration. One example is a fast GPU-based Gauss-Seidel solver of constrained dynamics [FTP16]. Another example is the efficient handling of nonlinearities and dynamically changing constraints as a superset of projective dynamics [OBLN17]. Very recently, Tang *et al.* [TWL*18] have designed a GPU-based solver of cloth dynamics with impact zones, efficiently integrated with GPU-based continuous collision detection.

A different approach to speed up cloth simulation is to apply adaptive remeshing, focusing simulation complexity where needed [NSO12]. Similar in spirit, Eulerian-on-Lagrangian cloth simulation applies remeshing with Eulerian coordinates to efficiently resolve the geometry of sharp sliding contacts [WPLS18].

Data-Driven Models. Multiple works, both well established [LCF00, SRIC01] and recent [BODO18, CO18], propose to model surface deformations as a function of pose. Similar to them, some existing data-driven methods for clothing animation also use the underlying kinematic skeletal model to drive the garment deformation [KV08, WHRO10, GRH*12, XUC*14, HTC*14]. Kim and Vendrovsky [KV08] first introduced a pose-space deformation approach that uses a skeletal pose as subspace domain. Hahn *et al.* [HTC*14] went one step further and performed cloth simulation in pose-dependent dynamic low-dimensional subspaces constructed from precomputed data. Wang *et al.* [WHRO10] used a precomputed database to locally enhance a low-resolution clothing simulation based on joint proximity.

Other methods produce detailed cloth animations by augmenting coarse simulations with example-based wrinkle data. Rohmer *et al.* [RPC*10] used the stretch tensor of a coarse animation output as a guide for wrinkle placement. Kavan *et al.* [KGBS11] used example data to learn an upsampling operator that adds fine details to a coarse cloth mesh. Zurdo *et al.* [ZBO13] proposed a mapping between low and high-resolution simulations, employing tracking constraints [BMWG07] to establish a correspondence between both resolutions. Saito *et al.* [SUM14] proposed an upsampling technique that adds physically feasible microscopic detail to coarsened meshes by considering the internal strain at runtime. More recently, Oh *et al.* [OLL18] have shown how to train a deep neural network to upsample low-resolution cloth simulations.

A different approach for cloth animation is to approximate full-space simulation models with coarse data-driven models. James and Fatahalian [JF03] used efficient precomputed low-rank approximations of physically-based simulations to achieve interactive deformable scenes. De Aguiar *et al.* [dASTH10] learned a low-dimensional linear model to characterize the dynamic behavior of clothing, including an approximation to resolve body-cloth collisions. Kim *et al.* [KKN*13] performed a near-exhaustive precomputation of the state of a cloth throughout the motion of a character. At run-time a secondary motion graph was explored to find the closest cloth state for the current pose. Despite its efficient implementation, the method cannot generalize to new motions. Xu *et*

Data-generation pipeline

Figure 2: *Overview of our preprocessing and runtime pipelines. As a preprocess, we generate physics-based simulations of multiple animated bodies wearing the same garment. At runtime, our data-driven cloth deformation model works by computing two corrective displacements on the unposed garment: global fit displacements dependent on the body's shape, and dynamic wrinkle displacements dependent on the body's shape and pose. Then, the deformed cloth is skinned on the body to produce the final result.*

al. [XUC*14] used a precomputed dataset to mix and match parts of different samples to synthesize a garment mesh that matches the current pose.

As discussed in the introduction, virtual try-on requires cloth models that respond to changes in body pose and shape. However, this is a scarce feature in data-driven cloth animation methods. Guan *et al.* [GRH*12] dressed a parametric character and independently modeled cloth deformations due to shape and pose. However, they relied on a linear model that struggles to generate realistic wrinkles, specially under fast motions. Moreover, they accounted for body shape by resizing the cloth model. Other works also apply a scaling factor to the garment to fit a given shape, without realistic deformation [YFHWW18,PMPHB17,LCT18].

The variability of the human body has also been addressed in garment design methods. Wang [Wan18] customized sewing patterns for different body shapes by means of an optimization process. Wang *et al.* [WCPM18] learned a shared shape space that allows the user to directly provide a sketch of the desired look, while the system automatically generates the corresponding sewing patterns and draped garment for different *static* bodies. In contrast, our method aims to estimate the fit of a specific garment (*i.e.*, without altering the underlying sewing patterns) for a wide range of animated bodies.

Performance Capture Re-Animation. Taking advantage of the recent improvements on performance capture methods [BPS*08, ZPBPM17, PMPHB17], virtual animation of real cloth that has been previously captured (and not simulated) has become an alternative. Initial attempts fit a parametric human model to the captured 3D scan to enable the re-animation of the captured data, without any explicit cloth layer [JTST10, FCS15]. More elaborate meth-

ods extract a cloth layer from the captured 3D scan and fit a parametric model to the actor [YFHWW18,NH13,PMPHB17,LCT18]. This allows editing the actor's shape and pose parameters while keeping the same captured garment or even changing it. However, re-animated motions lack realism since they cannot predict the nonrigid behavior of clothing under unseen poses or shapes, and are usually limited to copying wrinkles across bodies of different shapes [PMPHB17,LCT18].

Image-Based Methods. Cloth animation and virtual try-on methods have also been explored from an image-based point of view [SM06,ZSZ*12,HSR13,HFE13,HWW*18]. These methods aim to generate compelling 2D images of dressed characters, without dealing with any 3D model or simulation of any form. Hilsmann *et al.* [HFE13] proposed a pose-dependent image-based method that interpolates between images of clothes. More recently, Han *et al.* [HWW*18] have shown impressive photorealistic results using convolutional neural networks. However, these image-based methods are limited to 2D static images and fixed camera positions, and cannot fully convey the 3D fit and style of a garment. Other image-based works aim at recovering the full 3D model of both the underlying body shape and garment from monocular video input [YPA*18, RKS*14]. This allows garment replacement and transfer, but requires highly complex and expensive pipelines. Alternatively, other methods seek to recover the cloth material properties from video [YLL17], which also opens the door to cloth re-animation.

3. Clothing Animation

In this section, we describe our learning-based data-driven method to animate the clothing of a virtual character. Figure 2 shows an

overview of the method, separating the preprocessing and runtime stages.

In Section 3.1 we overview the components of our shape-and-pose-dependent cloth deformation model. The two key novel ingredients of our model are: (i) a Garment Fit Regressor (Section 3.2), which allows us to apply global body-shape-dependent deformations to the garment, and (ii) a Garment Wrinkle Regressor (Section 3.3), which predicts dynamic wrinkle deformations as a function of body shape and pose.

3.1. Clothing Model

We denote as M_b a deformed human body mesh, determined by shape parameters β (*e.g.*, the principal components of a database of body scans) and pose parameters θ (*e.g.*, joint angles). We also denote as M_c a deformed garment mesh worn by the human body mesh. A physics-based simulation would produce a cloth mesh $S_c(β, θ)$ as the result of simulating the deformation and contact mechanics of the garment on a body mesh with shape β and pose θ. Instead, we approximate S_c using a data-driven model.

Based on the observation that most garments closely follow the deformations of the body, we design our clothing model inspired by the Pose Space Deformation (PSD) literature [LCF00] and subsequent human body models [ASK*05, FCS15, LMR*15]. We assume that the body mesh is deformed according to a rigged parametric human body model,

$$M_b(β, θ) = W(T_b(β, θ), β, θ, \mathcal{W}_b), \qquad (1)$$

where $W(\cdot)$ is a skinning function, which deforms an unposed body mesh $T_b(β, θ) \in \mathbb{R}^{3 \times V_b}$ with V_b vertices based on: first, the shape parameters $β \in \mathbb{R}^{|β|}$, which define joint locations of an underlying skeleton; and second, the pose parameters $θ \in \mathbb{R}^{|θ|}$, which are the joint angles to articulate the mesh according to a skinning weight matrix $\mathcal{W}_b$. The unposed body mesh may be obtained additionally by deforming a template body mesh $\bar{\mathbf{T}}_b$ to account for body shape and pose-based surface corrections (See, *e.g.*, [LMR*15]).

We propose to model cloth deformations following a similar overall pipeline. For a given garment, we start from a template cloth mesh $\bar{\mathbf{T}}_c \in \mathbb{R}^{3 \times V_c}$ with V_c vertices, and we deform it in two steps. First, we compute an unposed cloth mesh $T_c(β, θ)$, and then we deform it using the skinning function $W(\cdot)$ to produce the full cloth deformation. A key insight in our model is to compute body-shape-dependent garment fit and shape-and-pose-dependent garment wrinkles as corrective displacements to the template cloth mesh, to produce the unposed cloth mesh:

$$T_c(β, θ) = \bar{\mathbf{T}}_c + R_G(β) + R_L(β, θ), \qquad (2)$$

where $R_G()$ and $R_L()$ represent two nonlinear regressors, which take as input body shape parameters and shape and pose parameters, respectively.

The final cloth skinning step can be formally expressed as

$$M_c(β, θ) = W(T_c(β, θ), β, θ, \mathcal{W}_c). \qquad (3)$$

Assuming that most garments closely follow the body, we define the skinning weight matrix $\mathcal{W}_c$ by projecting each vertex of the

template cloth mesh onto the closest triangle of the template body mesh, and interpolating the body skinning weights $\mathcal{W}_b$.

The pipeline Figure 2 shows the template body mesh $\bar{\mathbf{T}}_b$ wearing the template cloth mesh $\bar{\mathbf{T}}_c$ (Figure 2-a), and then the template cloth mesh in isolation (Figure 2-b), with the addition of garment fit (Figure 2-c), with the addition of garment wrinkles (Figure 2-d), and the final deformation after the skinning step (Figure 2-e).

By training regressors with collision-free data, our data-driven model learns naturally to approximate contact interactions, but it does not guarantee collision-free cloth outputs. In particular, when the garments are tight, interpenetrations with the body can become apparent. After the skinning step, we apply a postprocessing step to cloth vertices that collide with the body, by pushing them outside their closest body primitive. An example of collision postprocessing is shown in Figure 3.

3.2. Garment Fit Regressor

Our learning-based cloth deformation model represents corrective displacements on the unposed cloth state, as discussed above. We observe that such displacements are produced by two distinct sources. On one hand, the shape of the body produces an overall deformation in the form of stretch or relaxation, caused by tight or oversized garments, respectively. As we show in this section, we capture this deformation as a static global fit, determined by body shape alone. On the other hand, body dynamics produce additional global deformation and small-scale wrinkles. We capture this deformation as time-dependent displacements, determined by both body shape and motion, as discussed later in Section 3.3. We reach higher accuracy by training garment fit and garment wrinkles separately, in particular due to their static vs. dynamic nature.

We characterize static garment fit as a vector of per-vertex displacements $\Delta_G \in \mathbb{R}^{3 \times V_c}$. These displacements represent the deviation between the cloth template mesh $\bar{\mathbf{T}}_c$ and a smoothed version of the simulated cloth worn by the unposed body. Formally, we define the ground-truth garment fit displacements as

$$\Delta_G^{GT} = ρ(S_c(β, \mathbf{0})) - \bar{\mathbf{T}}_c, \qquad (4)$$

(a) *Before*   **(b)** *After*

Figure 3: *For tight clothing, data-driven cloth deformations may suffer from apparent collisions with the body (left). We apply a simple postprocessing step to push colliding cloth vertices outside the body (right).*

(a) *Garment Fit Regression*

(b) *Garment Wrinkle Regression*

Figure 4: *Results of Garment Fit Regression (top) and Garment Wrinkle Regression (bottom), for different bodies and poses.*

where $S_c(\beta, \mathbf{0}))$ represents a simulation of the garment on a body with shape β and pose $\theta = \mathbf{0}$, and ρ represents a smoothing operator.

To compute garment fit displacements in our data-driven model, we use a nonlinear regressor $R_G : \mathbb{R}^{|\beta|} \to \mathbb{R}^{3 \times V_c}$, which takes as input the shape of the body β. In particular, we implement the regressor $\Delta_G = R_G(\beta)$ using a single-hidden-layer multilayer perceptron (MLP) neural network. We train the MLP network by minimizing the mean squared error between predicted displacements Δ_G and ground-truth displacements Δ_G^{GT}.

See Figure 4a for a visualization of the garment fit regression. Notice how the original template mesh is globally deformed but lacks pose-dependent wrinkles.

3.3. Garment Wrinkle Regressor

We characterize dynamic cloth deformations (*e.g.*, wrinkles) as a vector of per-vertex displacements $\Delta_L \in \mathbb{R}^{3 \times V_c}$. These displacements represent the deviation between the simulated cloth worn by the moving body, $S_c(\beta, \theta)$, and the template cloth mesh $\bar{\mathbf{T}}_c$ corrected with the global garment fit Δ_G. We express this deviation in the body's rest pose, by applying the inverse skinning transformation $W^{-1}(\cdot)$ to the simulated cloth. Formally, we define the ground-

truth garment wrinkle displacements as

$$\Delta_L^{GT} = W^{-1}(S_c(\beta, \theta), \beta, \theta, \mathcal{W}_c) - \bar{\mathbf{T}}_c - \Delta_G. \qquad (5)$$

To compute garment wrinkle displacements in our data-driven model, we use a nonlinear regressor $R_L : \mathbb{R}^{|\beta|+|\theta|} \to \mathbb{R}^{3 \times V_c}$, which takes as input the shape β and pose θ of the body. In contrast to the static garment fit, garment wrinkles exhibit dynamic, history-dependent deformations. We account for such dynamic effects by introducing recursion within the regressor. In particular, we implement the regressor $\Delta_L = R_L(\beta, \theta)$ using a Recurrent Neural Network (RNN) based on Gated Recurrent Units (GRU) [CVMG*14], which has proven successful in modeling dynamic systems such as human pose prediction [MBR17]. Importantly, GRU networks do not suffer from the well-known vanishing and exploding gradients common in vanilla RNNs [PMB13]. Analogous to the MLP network in the garment fit regressor, we train the GRU network by minimizing the mean squared error between predicted displacements Δ_L and ground-truth displacements Δ_L^{GT}.

See Figure 4b for a visualization of the garment wrinkle regression. Notice how the garment obtained in the first step of our pipeline is further deformed and enriched with pose-dependent dynamic wrinkles.

4. Training Data and Regressor Settings

In this section, we give details on the generation of synthetic training sequences and the extraction of ground-truth data to train the regressor networks. In addition, we discuss the network settings and the hyperparameters used in our results.

4.1. Dressed Character Animation Dataset

To produce ground-truth data for the training of the Garment Fit Regressor and the Garment Wrinkle Regressor, we have created a novel dataset of dressed character animations with diverse motions and body shapes. Our prototype dataset has been created using only one garment, but it can be applied to other garments or their combinations.

As explained in Section 3.1, our approach relies on the use of a parametric human model. In our implementation, we have used SMPL [LMR*15]. We have selected 17 training body shapes, as follows. For each of the 4 principal components of the shape parameters β, we generate 4 samples, leaving the rest of the parameters in β as 0. To these 16 body shapes, we add the nominal shape with $\beta = 0$.

As animations, we have selected character motions from the CMU dataset [CMU], applied to the SMPL body model [VRM*17]. Specifically, we have used 56 sequences containing $7,117$ frames in total (at 30 fps, downsampled from the original CMU dataset of 120 fps). We have simulated each of the 56 sequences for each of the 17 body shapes, wearing the same garment mesh (*i.e.*, the T-shirt shown throughout the paper, which consists of $8,710$ triangles).

All simulations have been produced using the ARCSim physics-based cloth simulation engine [NSO12, NPO13], with remeshing

Figure 5: *Our nonlinear regression method succeeds to retain the rich and history-dependent wrinkles of the physics-based simulation. Linear regression, on the other hand, suffers blending and smoothing artifacts even on the training sequence shown in the figure.*

turned off to preserve the topology of the garment mesh. ARC-Sim requires setting several material parameters. In our case, since we are simulating a T-shirt, we have chosen an interlock knit with 60% cotton and 40% polyester, from a set of measured materials [WRO11]. We have executed all simulations using a fixed time step of 3.33ms, with the character animations running at 30 fps and interpolated to each time step. We have stored in the output database the simulation results from 1 out of every 10 time steps, to match the frame rate of the character animations. This produces a total of 120,989 output frames of cloth deformation.

ARCSim requires a valid collision-free initial state. To this end, we manually pre-position the garment mesh *once* on the template body mesh $\bar{\mathbf{T}}_b$. We run the simulation to let the cloth relax, and thus define the initial state for all subsequent simulations. In addition, we apply a smoothing operator $\rho(\cdot)$ to this initial state to obtain the template cloth mesh $\bar{\mathbf{T}}_c$.

The generation of ground-truth garment fit data requires the simulation of the garment worn by unposed bodies of various shapes. We do this by incrementally interpolating the shape parameters from the template body mesh to the target shape, while simulating the garment from its collision-free initial state. Once the body reaches its target shape, we let the cloth rest, and we compute the ground-truth garment fit displacements Δ_G^{GT} according to Equation 4.

Similarly, to simulate the garment on animations with arbitrary pose and shape, we incrementally interpolate both shape and pose parameters from the template body mesh to the shape and initial pose of the animation. Then, we let the cloth rest before starting the actual animation. The simulations produce cloth meshes $S_c(\beta, \theta)$,

and from these we compute the ground-truth garment wrinkle displacements Δ_L^{GT} according to Equation 5.

4.2. Network Implementation and Training

We have implemented the neural networks presented in Sections 3.2 and 3.3 using Tensorflow [AAB*15]. The MLP network for garment fit regression contains a single hidden layer with 20 hidden neurons, which we found enough to predict the global fit of the garment. The GRU network for garment wrinkle regression also contains a single hidden layer, but in this case we obtained the best fit of the test data using 1500 hidden neurons. In both networks, we have applied dropout regularization to avoid overfitting the training data. Specifically, we randomly disable 20% of the hidden neurons on each optimization step. Moreover, we shuffle the training data at the beginning of each training epoch.

During training, we use the Adam optimization algorithm [KB14] for 2000 epochs with an initial learning rate of 0.0001. For the garment fit MLP network, we use for training the ground-truth data from all 17 body shapes. For the garment wrinkle GRU network, we use for training the ground-truth data from 52 animation sequences, leaving 4 sequences for testing purposes. When training the GRU network, we use a batch size of 128. Furthermore, to speed-up the training process of the GRU network, we compute the error gradient using Truncated Backpropagation Through Time (TBPTT), with a limit of 90 time steps.

Figure 6: *Quantitative evaluation of generalization to new shapes, comparing our method to retargeting techniques [LCT18, PM-PHB17]. The top plot shows the error as we increase the body shape to values not used for training, and back, on a static pose (See Figure 8). The bottom plot shows the error as we change both the body shape and pose during a test sequence not used for training.*

5. Evaluation

In this section, we discuss quantitative and qualitative evaluation of the results obtained with our method. We compare our results with other state-of-the-art methods, and we demonstrate the benefits of our method for virtual try-on, in terms of both visual fidelity and runtime performance.

5.1. Runtime Performance

We have implemented our method on an Intel Core i7-6700 CPU, with a Nvidia Titan X GPU with 32GB of RAM. Table 1 shows average per-frame execution times of our implementation for different garment resolutions, including garment fit regression, garment wrinkle regression, and skinning, with and without collision postprocessing. For reference, we also include simulation timings of a CPU-based implementation of full physics-based simulation using ARCSim [NSO12].

The low computational cost of our method makes it suitable for interactive applications. Its memory footprint is as follows: 1.1MB for the Garment Fit Regressor MLP, and 108.1MB for the Garment Wrinkle Regressor GRU, both without any compression.

Figure 7: *Quantitative evaluation of generalization to new poses, comparing our method to Linear Blend Skinning (LBS) and Linear Regression (LR). Differences with Linear Regression, although visible in the plot, are most evident in the accompanying video.*

Triangles	ARCSim [NSO12]		Our method (w/o postprocess)		Our method (w/ postprocess)	
	mean	std	mean	std	mean	std
4,210	2909.4	1756.8	1.47	0.31	3.39	0.30
8,710	5635.4	2488.5	1.51	0.28	4.01	0.27
17,710	10119.5	5849.0	2.12	0.32	5.47	0.32
26,066	15964.4	4049.3	2.40	0.33	6.87	0.30

Table 1: *Per-frame execution times (in milliseconds) of our method for garments of different resolutions, with and without collision postprocessing. Full physics-based simulation times are also provided for reference.*

5.2. Quantitative Evaluation

Linear vs. nonlinear regression. In Figure 5, we compare the fitting quality of our nonlinear regression method vs. linear regression (implemented using a single-layer MLP neural network without nonlinear activation function), on a training sequence. While our method retains the rich and history-dependent wrinkles, linear regression suffers smoothing and blending artifacts.

Generalization to new body shapes. In Figure 6, we quantitatively evaluate the generalization of our method to new shapes (*i.e.*,

Figure 8: *Our method matches qualitatively the deformations of the ground-truth physics-based simulation when changing the body shape beyond training values. In particular, notice how the T-shirt achieves the same overall drape and mid-scale wrinkles. Retargeting techniques [LCT18, PMPHB17], on the other hand, scale the garment, and suffer noticeable artifacts away from the base shape.*

not in the training set). We depict the per-vertex mean error on a static pose (top) and a dynamic sequence (bottom), as we change the body shape over time. To provide a quantitative comparison to existing methods, we additionally show the error suffered by our implementation of cloth retargeting [LCT18, PMPHB17]. As discussed in Section 2, such retargeting methods scale the garment in a way analogous to the body to retain the garment's style. As we show in the accompanying video, even if retargeting produces appealing results, it does not suit the purpose of virtual try-on, and produces larger error w.r.t. a physics-based simulation of the garment. This is clearly visible in Figure 6, where the error with retargeting increases as the shape deviates from the nominal shape, while it remains stable with our method.

Generalization to new body poses. In Figure 7, we depict the per-vertex mean error of our method in 2 test motion sequences with constant body shape but varying pose. In particular, we validate our cloth animation results on the CMU sequences 01_01 and 55_27 [CMU], which were excluded from the training set, and exhibit complex motions including jumping, dancing and highly dynamic arm motions. Additionally, we show the error suffered by two baseline methods for cloth animation. On one hand, Linear Blend Skinning (LBS), which consists of applying the kinematic transformations of the underlying skeleton directly to the garment

template mesh. On the other hand, a Linear Regressor (LR) that predicts cloth deformation directly as a function of pose, implemented using a single-layer MLP neural network without nonlinear activation function. The results demonstrate that our two-step approach, with separate nonlinear regression of garment fit and garment wrinkles, outperforms the linear approach. This is particularly evident in the accompanying video, where the linear regressor exhibits blending artifacts.

5.3. Qualitative Evaluation

Generalization to new shapes. In Figure 8, we show the clothing deformations produced by our approach on a static pose while changing the body shape over time. We compare results with a physics-based simulation and with our implementation of retargeting techniques [LCT18, PMPHB17]. Notice how our method successfully reproduces ground-truth deformations, including the overall drape (*i.e.*, how the T-shirt slides up the belly due to the stretch caused by the increasingly fat character) and mid-scale wrinkles.

We also compare our method to state-of-the-art data-driven methods that account for changes in both body shape and pose. Figure 9 shows the result of DRAPE [GRH*12] when the same

Figure 9: *Comparison between DRAPE [GRH*12] (top) and our method (bottom). DRAPE cannot realistically cope with shape variations, and it is limited to scaling the garment to fit the target shape. In contrast, our method predicts realistically how a garment fits avatars with very diverse body shapes.*

garment is worn by two avatars with significantly different body shapes. DRAPE approximates the deformation of the garment by scaling it such that it fits the target shape, which produces plausible but unrealistic results. In contrast, our method deforms the garment in a realistic manner.

In Figure 10, we compare our model to ClothCap [PMPHB17]. Admittedly, the virtual try-on scenario considered by Pons-Moll and colleagues differs from ours: while we assume that the virtual garment is provided by a brand and the customer tests the fit of a certain sized garment on a virtual avatar of his/her body, Pons-Moll *et al.* reconstruct the clothing and body shape from 4D scans, and transfer the captured garment to different shapes. However, their retargeting lacks realism because cloth deformations are simply copied across different shapes. In contrast, our method produces realistic pose- *and* shape-dependent deformations.

Generalization to new poses. We visually evaluate the quality of our model in Figure 11, where we compare ground-truth physics-based simulation and our data-driven cloth deformations on a test sequence. The overall fit and mid-scale wrinkles are successfully predicted using our data-driven model, with a performance gain of three orders of magnitude. Similarly, in Figure 12, we show more frames of a test sequence. Notice the realistic wrinkles in the belly area that appear when the avatar crouches. Please see the accompanying video for animated results and further comparisons.

Figure 10: *Comparison between ClothCap [PMPHB17] (top) and our method (bottom). In ClothCap, the original T-shirt (top-left) is obtained using performance capture, and then scaled to fit a bigger avatar. While the result appears plausible for certain applications, it is not suited for virtual try-on. In contrast, our method produces pose- and shape-dependent drape and wrinkles, thus enabling a virtual try-on experience. The skeletal animation used in this comparison and the meshes shown for ClothCap were both provided by the original authors.*

Figure 11: *Comparison between a ground-truth physics-based simulation (top) and our data-driven method (bottom), on a test sequence not used for training (01_01 from [CMU]). Even though our method runs three orders of magnitude faster, it succeeds to predict the overall fit and mid-scale wrinkles of the garment.*

Figure 12: *Cloth animation produced by our data-driven method on a test sequence not used for training. Notice how our model successfully deforms the T-shirt and exhibits realistic wrinkles and dynamics.*

6. Conclusions and Future Work

We have presented a novel data-driven method for animation of clothing that enables efficient virtual try-on applications at over 250 fps. Given a garment template worn by a human model, our two-level regression scheme independently models two distinct sources of deformation: garment fit, due to body shape; and garment wrinkles, due to shape and pose. We have shown that this strategy, in combination with the ability of the regressors to represent nonlinearities and dynamics, allows our method to overcome the limitations of previous data-driven approaches.

We believe our approach makes an important step towards bridging the gap between the accuracy and flexibility of physics-based simulation methods and the computational efficiency of data-driven methods. Nevertheless, there are a number of limitations that remain open for future work.

First, our method requires independent training per garment size and material. In particular, the dataset used in this work consists of a single one-size garment with specific fabric material. Since the material specification can greatly affect the behavior of the cloth, including it as an input to our model would make the wrinkle regression task significantly more challenging. Although this particular scenario remains to be tested, using the proposed network architecture would likely result in over-smoothed results. Moreover, multiple independent garment animations would not capture correctly the interactions between garments. Mix-and-match virtual try-on requires training each possible combination of test garments.

Second, collisions are not fully handled by our method. Our regressors are trained with collision-free data, and therefore our model implicitly learns to approximate contact, but it is not guaranteed to be collision-free. Future work could address this limitation by imposing low-level collision constraints as an explicit objective for the regressor.

Our results show that our method succeeds to predict the overall drape and mid-scale wrinkles of garments, but it smooths excessively high-frequency wrinkles, both spatially and temporally. We wish to investigate alternative methods of recursion to handle accurately both history-dependent draping and highly dynamic wrinkles.

Finally, our model is rooted in the assumption that most garments follow closely the body. This assumption may not be valid for loose clothing, and the decomposition of the deformation into a static fit and dynamic wrinkles would not lead to accurate results. It remains to test our method under such conditions.

References

[AAB*15] ABADI M., AGARWAL A., BARHAM P., BREVDO E., CHEN Z., CITRO C., CORRADO G. S., DAVIS A., DEAN J., DEVIN M., GHEMAWAT S., GOODFELLOW I., HARP A., IRVING G., ISARD M., JIA Y., JOZEFOWICZ R., KAISER L., KUDLUR M., LEVENBERG J., MANÉ D., MONGA R., MOORE S., MURRAY D., OLAH C., SCHUSTER M., SHLENS J., STEINER B., SUTSKEVER I., TALWAR K., TUCKER P., VANHOUCKE V., VASUDEVAN V., VIÉGAS F., VINYALS O., WARDEN P., WATTENBERG M., WICKE M., YU Y., ZHENG X.: TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. URL: https://www.tensorflow.org/. 6

[ASK*05] ANGUELOV D., SRINIVASAN P., KOLLER D., THRUN S., RODGERS J., DAVIS J.: SCAPE: Shape Completion and Animation of People. *ACM Transactions on Graphics 24*, 3 (2005), 408–416. doi:10.1145/1073204.1073207. 4

[BMO*14] BENDER J., MÜLLER M., OTADUY M. A., TESCHNER M., MACKLIN M.: A Survey on Position-Based Simulation Methods in Computer Graphics. *Computer Graphics Forum 33*, 6 (2014). doi:10.1111/cgf.12346. 2

[BMWG07] BERGOU M., MATHUR S., WARDETZKY M., GRINSPUN E.: TRACKS: Toward Directable Thin Shells. *ACM Transactions on Graphics (SIGGRAPH) 26*, 3 (jul 2007), 50:1–50:10. doi:10.1145/1276377.1276439. 2

[BODO18] BAILEY S. W., OTTE D., DILORENZO P., O'BRIEN J. F.:

Fast and deep deformation approximations. *ACM Trans. Graph. 37*, 4 (July 2018), 119:1–119:12. `doi:10.1145/3197517.3201300`. 2

[BPS*08] BRADLEY D., POPA T., SHEFFER A., HEIDRICH W., BOUBEKEUR T.: Markerless Garment Capture. *ACM Trans. Graphics (Proc. SIGGRAPH) 27*, 3 (2008), 99. `doi:10.1145/1360612.1360698`. 3

[BSBC12] BROUET R., SHEFFER A., BOISSIEUX L., CANI M.-P.: Design Preserving Garment Transfer. *ACM Trans. Graph. 31*, 4 (2012), 36:1–36:11. `doi:10.1145/2185520.2185532`. 2

[CLMMO14] CIRIO G., LOPEZ-MORENO J., MIRAUT D., OTADUY M. A.: Yarn-level simulation of woven cloth. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH Asia) 33*, 6 (2014). `doi:10.1145/2661229.2661279`. 1

[CMU] CMU Graphics Lab Motion Capture Database . `http://mocap.cs.cmu.edu/`. 2, 5, 8, 9

[CO18] CASAS D., OTADUY M. A.: Learning nonlinear soft-tissue dynamics for interactive avatars. *Proceedings of the ACM on Computer Graphics and Interactive Techniques 1*, 1 (may 2018). `doi:10.1145/3203187`. 2

[CVMG*14] CHO K., VAN MERRIËNBOER B., GULCEHRE C., BAHDANAU D., BOUGARES F., SCHWENK H., BENGIO Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. `arXiv:1406.1078`. 5

[dASTH10] DE AGUIAR E., SIGAL L., TREUILLE A., HODGINS J. K.: Stable spaces for real-time clothing. *ACM Transactions on Graphics 29*, 4 (July 2010), 106:1–106:9. `doi:10.1145/1778765.1778843`. 1, 2

[FCS15] FENG A., CASAS D., SHAPIRO A.: Avatar reshaping and automatic rigging using a deformable model. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games* (2015), pp. 57–64. `doi:10.1145/2822013.2822017`. 3, 4

[FTP16] FRATARCANGELI M., TIBALDO V., PELLACINI F.: Vivace: A practical gauss-seidel method for stable soft body dynamics. *ACM Trans. Graph. 35*, 6 (Nov. 2016), 214:1–214:9. `doi:10.1145/2980179.2982437`. 2

[GRH*12] GUAN P., REISS L., HIRSHBERG D. A., WEISS A., BLACK M. J.: Drape: Dressing any person. *ACM Trans. Graph. 31*, 4 (July 2012), 35:1–35:10. `doi:10.1145/2185520.2185531`. 2, 3, 8, 9

[HFE13] HILSMANN A., FECHTELER P., EISERT P.: Pose space image based rendering. *Computer Graphics Forum 32*, 2 (2013), 265–274. `doi:10.1111/cgf.12046`. 3

[HSR13] HAUSWIESNER S., STRAKA M., REITMAYR G.: Virtual try-on through image-based rendering. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 19*, 9 (2013), 1552–1565. `doi:10.1109/TVCG.2013.67`. 3

[HTC*14] HAHN F., THOMASZEWSKI B., COROS S., SUMNER R. W., COLE F., MEYER M., DEROSE T., GROSS M.: Subspace Clothing Simulation Using Adaptive Bases. *ACM Transactions on Graphics 33*, 4 (jul 2014), 105:1–105:9. `doi:10.1145/2601097.2601160`. 2

[HWW*18] HAN X., WU Z., WU Z., YU R., DAVIS L. S.: VITON: An image-based virtual try-on network. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018). 3

[JF03] JAMES D. L., FATAHALIAN K.: Precomputing Interactive Dynamic Deformable Scenes. *ACM Transactions on Graphics (Proc. SIGGRAPH) 22*, 3 (jul 2003), 879–887. `doi:10.1145/882262.882359`. 2

[JTST10] JAIN A., THORMÄHLEN T., SEIDEL H.-P., THEOBALT C.: MovieReshape: Tracking and Reshaping of Humans in Videos. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2010) 29*, 5 (2010). `doi:10.1145/1882261.1866174`. 3

[KB14] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. `arXiv:1412.6980`. 6

[KGBS11] KAVAN L., GERSZEWSKI D., BARGTEIL A. W., SLOAN P.-P.: Physics-inspired upsampling for cloth simulation in games. *ACM Transactions on Graphics 30*, 4 (2011), 1. `doi:10.1145/2010324.1964988`. 2

[KJM08] KALDOR J. M., JAMES D. L., MARSCHNER S.: Simulating knitted cloth at the yarn level. *ACM Trans. Graph. 27*, 3 (2008), 65:1–65:9. `doi:10.1145/1360612.1360664`. 1

[KKN*13] KIM D., KOH W., NARAIN R., FATAHALIAN K., TREUILLE A., O'BRIEN J. F.: Near-exhaustive precomputation of secondary cloth effects. *ACM Transactions on Graphics 32*, 4 (2013), 1. `doi:10.1145/2461912.2462020`. 2

[KV08] KIM T.-Y., VENDROVSKY E.: Drivenshape: a data-driven approach for shape deformation. In *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2008), Eurographics Association, pp. 49–55. `doi:10.2312/SCA/SCA08/049-055`. 2

[LCF00] LEWIS J. P., CORDNER M., FONG N.: Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In *Annual Conference on Computer Graphics and Interactive Techniques* (2000), pp. 165–172. 2, 4

[LCT18] LÄHNER Z., CREMERS D., TUNG T.: Deepwrinkles: Accurate and realistic clothing modeling. In *European Conference on Computer Vision (ECCV)* (2018). 2, 3, 7, 8

[LMR*15] LOPER M., MAHMOOD N., ROMERO J., PONS-MOLL G., BLACK M. J.: SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 34*, 6 (Oct. 2015), 248:1–248:16. `doi:10.1145/2816795.2818013`. 2, 4, 5

[MBR17] MARTINEZ J., BLACK M. J., ROMERO J.: On human motion prediction using recurrent neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), IEEE, pp. 4674–4683. `doi:10.1109/CVPR.2017.497`. 5

[NH13] NEOPHYTOU A., HILTON A.: Shape and pose space deformation for subject specific animation. In *3DV* (2013), pp. 334–341. `doi:10.1109/3DV.2013.51`. 3

[NPO13] NARAIN R., PFAFF T., O'BRIEN J. F.: Folding and crumpling adaptive sheets. *ACM Transactions on Graphics 32*, 4 (July 2013), 51:1–51:8. `doi:10.1145/2461912.2462010`. 5

[NSO12] NARAIN R., SAMII A., O'BRIEN J. F.: Adaptive anisotropic remeshing for cloth simulation. *ACM Transactions on Graphics 31*, 6 (Nov. 2012), 152:1–152:10. `doi:10.1145/2366145.2366171`. 1, 2, 5, 7

[OBLN17] OVERBY M., BROWN G. E., LI J., NARAIN R.: ADMM ⊇ Projective Dynamics: Fast Simulation of Hyperelastic Models with Dynamic Constraints. *IEEE Transactions on Visualization and Computer Graphics 23*, 10 (Oct 2017), 2222–2234. `doi:10.1109/TVCG.2017.2730875`. 2

[OLL18] OH Y. J., LEE T. M., LEE I.-K.: Hierarchical cloth simulation using deep neural networks. In *Proceedings of Computer Graphics International 2018* (2018), CGI 2018, pp. 139–146. `doi:10.1145/3208159.3208162`. 2

[PMB13] PASCANU R., MIKOLOV T., BENGIO Y.: On the difficulty of training recurrent neural networks. In *International Conference on Machine Learning (ICML)* (2013), pp. 1310–1318. 5

[PMPHB17] PONS-MOLL G., PUJADES S., HU S., BLACK M.: ClothCap: Seamless 4D Clothing Capture and Retargeting. *ACM Transactions on Graphics, (Proc. SIGGRAPH) 36*, 4 (2017). `doi:10.1145/3072959.3073711`. 2, 3, 7, 8, 9, 10

[RKS*14] ROGGE L., KLOSE F., STENGEL M., EISEMANN M., MAGNOR M.: Garment replacement in monocular video sequences. *ACM Transactions on Graphics (TOG) 34*, 1 (2014), 6. 3

[RPC*10] ROHMER D., POPA T., CANI M.-P., HAHMANN S., SHEFFER A.: Animation wrinkling: augmenting coarse cloth simulations with realistic-looking wrinkles. *ACM Transactions on Graphics (TOG) 29*, 6 (2010), 157. `doi:10.1145/1882261.1866183`. 2

[SM06]   SCHOLZ V., MAGNOR M.: Texture replacement of garments in monocular video sequences. In *Eurographics Conference on Rendering Techniques* (2006), pp. 305–312. 3

[SRIC01]   SLOAN P.-P. J., ROSE III C. F., COHEN M. F.: Shape by example. In *Proceedings of Symposium on Interactive 3D graphics* (2001), ACM, pp. 135–143. 2

[SSIF09]   SELLE A., SU J., IRVING G., FEDKIW R.: Robust high-resolution cloth using parallelism, history-based collisions, and accurate friction. *IEEE Transactions on Visualization and Computer Graphics 15*, 2 (Mar. 2009), 339–350. `doi:10.1109/TVCG.2008.79`. 1

[SUM14]   SAITO S., UMETANI N., MORISHIMA S.: Macroscopic and microscopic deformation coupling in up-sampled cloth simulation. *Computer Animation and Virtual Worlds 25*, 3-4 (2014), 435–444. `doi:10.1002/cav.1589`. 2

[TWL*18]   TANG M., WANG T., LIU Z., TONG R., MANOCHA D.: I-Cloth: Incremental collision handling for GPU-based interactive cloth simulation. *ACM Transaction on Graphics (Proc. of SIGGRAPH Asia) 37*, 6 (2018), 204:1–10. `doi:10.1145/3272127.3275005`. 1, 2

[VRM*17]   VAROL G., ROMERO J., MARTIN X., MAHMOOD N., BLACK M. J., LAPTEV I., SCHMID C.: Learning from synthetic humans. In *Conference on Computer Vision and Pattern Recognition* (2017). `doi:10.1109/CVPR.2017.492`. 2, 5

[Wan18]   WANG H.: Rule-free Sewing Pattern Adjustment with Precision and Efficiency. *ACM Transactions on Graphics 37*, 4 (2018), 53:1–53:13. `doi:10.1145/3197517.3201320`. 3

[WCPM18]   WANG T. Y., CEYLAN D., POPOVIC J., MITRA N. J.: Learning a Shared Shape Space for Multimodal Garment Design. *ACM Trans. Graph. 37*, 6 (2018), 1:1–1:14. `doi:10.1145/3272127.3275074`. 3

[WHRO10]   WANG H., HECHT F., RAMAMOORTHI R., O'BRIEN J.: Example-based wrinkle synthesis for clothing animation. *ACM Transactions on Graphics 29*, 4 (2010), 1. `doi:10.1145/1833351.1778844`. 1, 2

[WPLS18]   WEIDNER N. J., PIDDINGTON K., LEVIN D. I. W., SUEDA S.: Eulerian-on-lagrangian cloth simulation. *ACM Transactions on Graphincs 37*, 4 (2018), 50:1–50:11. `doi:10.1145/3197517.3201281`. 2

[WRO11]   WANG H., RAMAMOORTHI R., O'BRIEN J. F.: Data-Driven Elastic Models for Cloth: Modeling and Measurement. *ACM Transactions on Graphics (Proc. SIGGRAPH) 30*, 4 (July 2011), 71:1–11. `doi:10.1145/2010324.1964966`. 6

[XUC*14]   XU W., UMETANI N., CHAO Q., MAO J., JIN X., TONG X.: Sensitivity-optimized Rigging for Example-based Real-time Clothing Synthesis. *ACM Transactions on Graphics 33*, 4 (2014), 107:1–107:11. `doi:10.1145/2601097.2601136`. 2, 3

[YFHWW18]   YANG J., FRANCO J.-S., HÉTROY-WHEELER F., WUHRER S.: Analyzing Clothing Layer Deformation Statistics of 3D Human Motions. In *European Conference on Computer Vision (ECCV)* (Sept. 2018). `doi:10.1007/978-3-030-01234-2_15`. 3

[YLL17]   YANG S., LIANG J., LIN M. C.: Learning-based cloth material recovery from video. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 4383–4393. `doi:10.1109/ICCV.2017.470`. 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 37*, 5 (2018), 170:1–170:14. `doi:10.1145/3026479`. 3

[ZBO13]   ZURDO J. S., BRITO J. P., OTADUY M. A.: Animating wrinkles by example on non-skinned cloth. *IEEE Transactions on Visualization and Computer Graphics 19*, 1 (2013), 149–158. `doi:10.1109/TVCG.2012.79`. 2

[ZPBPM17]   ZHANG C., PUJADES S., BLACK M. J., PONS-MOLL G.: Detailed, Accurate, Human Shape Estimation From Clothed 3D Scan Sequences. In *The IEEE Conference on Computer Vision and Pattern*

Recognition (CVPR) (July 2017). `doi:10.1109/CVPR.2017.582`. 3

[ZSZ*12]   ZHOU Z., SHU B., ZHUO S., DENG X., TAN P., LIN S.: Image-based clothes animation for virtual fitting. In *SIGGRAPH Asia 2012 Technical Briefs* (2012), ACM, p. 33. `doi:10.1145/2407746.2407779`. 3