Self-Supervised Collision Handling via Generative 3D Garment Models for Virtual Try-On

Igor Santesteban[1]   Nils Thuerey[2]   Miguel A. Otaduy[1]   Dan Casas[1]

[1]Universidad Rey Juan Carlos, Spain
[2]Technical University of Munich, Germany

first.last@{urjc.es}{tum.de}

Figure 1: Our data-driven method regresses deformed garments via a generative model that is trained to avoid collisions.

Abstract

We propose a new generative model for 3D garment deformations that enables us to learn, for the first time, a data-driven method for virtual try-on that effectively addresses garment-body collisions. In contrast to existing methods that require an undesirable postprocessing step to fix garment-body interpenetrations at test time, our approach directly outputs 3D garment configurations that do not collide with the underlying body. Key to our success is a new canonical space for garments that removes pose-and-shape deformations already captured by a new diffused human body model, which extrapolates body surface properties such as skinning weights and blendshapes to any 3D point. We leverage this representation to train a generative model with a novel self-supervised collision term that learns to reliably solve garment-body interpenetrations. We extensively evaluate and compare our results with recently proposed data-driven methods, and show that our method is the first to successfully address garment-body contact in unseen body shapes and motions, without compromising realism and detail.

1. Introduction

The digitalization of 3D garments has important applications in many areas of our everyday lives such as online shopping, video games, visual effects, and fashion design, and it has traditionally been addressed with physics-based methods [37, 38]. However, even if these methods offer solutions that generalize well to any type of garment, produce physically-accurate results, and solve body-garment contact, they require computationally expensive runtime evaluations. Consequently, they do not meet the combined robustness and performance needed for real-time applications such as virtual try-on. Furthermore, they are not easily differentiable and cannot be integrated into computer vision pipelines that, for example, fit deformable models into images to extract information about the scene.

Data-driven methods have emerged as a popular alternative to physics-based methods. The core idea is to *learn* a function that mimics the garment behavior observed in a large dataset. To this end, recent methods leverage the capability of neural networks to learn nonlinear functions, and propose differentiable models that output 3D deformed garments as a function of the target shape, motion, style, size, and other design parameters [62, 53, 41, 61, 64, 16, 31]. These methods showcase great realism and robustness, however, we identify a fundamental limitation in all existing works: despite using a loss term that penalizes unphysical body-garment interpenetrations at training time [16, 4], predicted garments commonly suffer from body-garment interpenetrations in test sequences. This is usually addressed with an added postprocessing step that pushes the problem-

atic regions of the garment, identified by exhaustive search, outside of the body [41, 53].

The undesired interpenetrations arise from natural residual errors in test samples when optimizing neural networks which, combined with the extremely narrow gap between body surface and garment, can produce artifacts even if the predicted 3D mesh closely matches the ground truth deformed garment. In this work, we address this inherent limitation and propose, to the best of our knowledge, the first data-driven method to reliably solve garment-body interpenetrations *without* requiring any postprocessing step. We achieve this through three main contributions.

First, we propose to enhance existing human body models [29] by learning to smoothly expand the surface parameters to any 3D point. Intuitively, this allows us to model the deformation at any 3D point, *e.g.*, a vertex of a deformed loose garment, leveraging the deformation capabilities of existing human body models. This expanded human body represents a fundamental building block for our method.

Our second contribution addresses the common assumption, made by existing data-driven models, that garment deformations closely follow the underlying body deformations. This popular simplification is often used to define garment models that use skinning parameters based on the closest body vertex in *rest pose*, and subsequently articulate the garment using a standard linear blend skinning (LBS) approach. We show that simplified transformations to bring ground truth data into a normalized representation, *e.g.* via inverse LBS [53, 41], cannot correctly represent the complex deformations that garments exhibit, and often introduce undesirable artifacts. Instead, we propose a garment model that represents deformations in a novel *unposed* and *deshaped* canonical space by removing deformations already captured via our expanded human body model. Since it yields correct skinning attributes for any 3D point, our garment model is designed to not to introduce collisions during projection operations between the canonical space and the posed space.

Our third and most important contribution is to leverage the novel canonical representation of garments to learn a generative subspace of deformations. Garments in this canonical space are encoded with respect to a *constant* reference body configuration. This not only gives an improved representation of garment deformations, but also allows us to reliably learn to solve collisions via self-supervision, by exhaustively sampling the generative space. We then learn a regressor that outputs mesh deformations encoded in this subspace, and use our garment model to project them to the final deformed state. Since both the deformation subspace and the projection step are designed to avoid collisions, our final deformed 3D garments do not interpenetrate the underlying body mesh, regardless the shape and pose parameter.

2. Related Work

Current approaches to animate 3D clothing can be classified into: physics-based simulation and data-driven models.

Physics-Based Cloth Simulation. Physics-based methods use discretizations of classical mechanics to model how cloth deforms, and typically comprise three steps: computation of internal forces, collision detection, and collision response [38]. These methods produce highly-realistic simulations, generalize to different garments, and can handle body-garment collisions, however, fail to meet the combined robustness and performance needed for real-time applications such as virtual try-on.

A wide range of strategies have been proposed to address the computational bottleneck in physics-based methods. Recent attempts include approximations of the dynamics to trade physical accuracy for speed [3, 6, 30], adaptive remeshing to refine surface discretization [26, 37], upsampling details to enrich coarse simulations [22, 73], and GPU-based solvers [58, 12, 59]. Another challenge in physics-based simulation is the estimation of the model parameters. To this end, mechanical approaches have been proposed [63, 34], but expensive studio settings are required. Alternative methods attempt to recover physical parameters directly from videos by a model fitting process [5, 57, 36] or learn this task directly from data [7, 66, 69, 46, 50]. Despite the impressive progress towards addressing the critical points in physics-based models, virtual try-on applications require faster and easier to set up methods.

Data-Driven Cloth Models. The deformation of 3D clothing can also be modeled from a data-driven perspective. Inspired by pioneering works that model pose-dependent surface deformations [27], many methods propose to *learn* from examples how 3D garments deform as a function of the underlying human body. This has been demonstrated for a variety of goals, including design [54, 64], animation [19, 8, 65, 68], and virtual try-on [15, 62, 31, 41, 16, 61]. Training data in form of 3D meshes can be obtained using physics-based simulations [4, 37, 28], using 3D reconstruction methods for from either multiview [71, 42, 48, 47] or single image [1, 13, 17, 39, 72], or using hybrid methods [70].

To encode garment deformations a variety of representations have been used, including normal maps [24], displacement maps [20], and geometry images [44], but closest to our work are the methods that use 3D positions or offsets to deform a template [41, 62, 53, 64]. To parameterize the deformation of clothing, a common strategy is to borrow the parametric space from a human body model [29, 21, 67]. However, some methods use geometric features such as PointNet [45, 16] or sketches [64] to describe the target body or garment style instead.
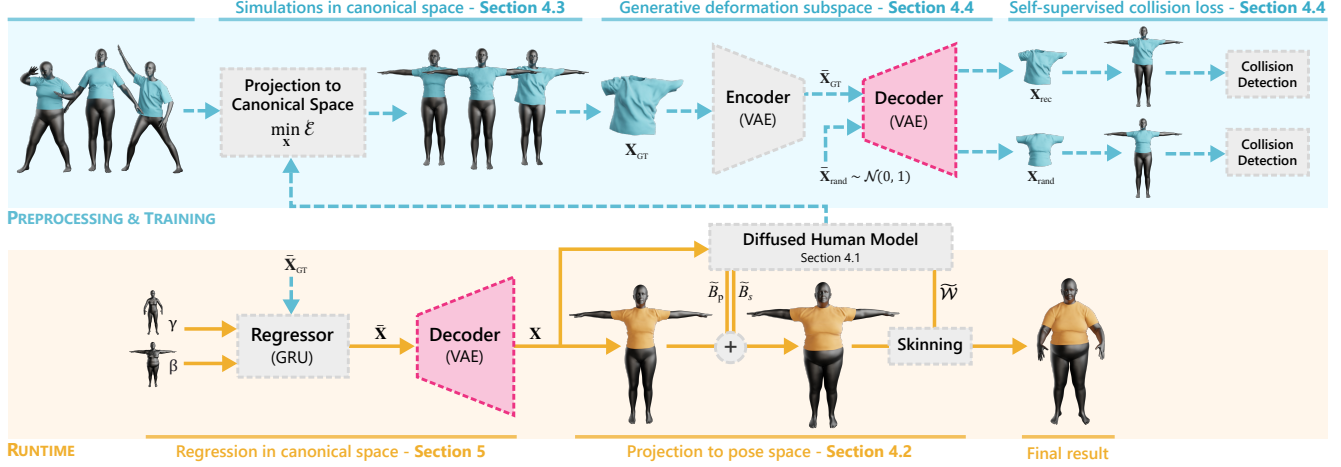
Figure 2: Overview of our preprocessing (top) and runtime pipelines (bottom). The decoder network is trained to avoid collisions in a self-supervised fashion, and then employed by the regressor network to reproduce these states at runtime.

Related to our data-driven approach for implementing the regressor, the early work of Guan *et al.* [15] use a linear model to produce garment deformations due to body pose and shape. The method outputs plausible pose-dependent wrinkles, but shape-dependent deformations are limited to resizing the cloth model. More recent methods use deep learning to train more sophisticated nonlinear regressors that deform a garment template. Santesteban *et al.* [53] use a fully-connected architecture with GRU [10] modules to learn pose and shape garment deformations, including dynamics. Similarly, Patel *et al.* [41] learn deformations to produce wrinkles of different scales with a desired style, but are limited to static deformations. Vidaurre *et al.* [62] use fully convolutional graph architecture to learn shape-dependent deformations for any garment mesh topology, but do not model pose or style. Instead of garment templates, Bertiche *et al.* [4] are capable of learning deformations for a much wider variety of garments using a representation based on 3D offsets of the body vertices. Despite the success of these methods in modeling how cloth deforms as a function of different factors including body shape, pose, motion, design, and material, a common critical weakness in all data-driven approaches is the handling of body-garment collisions. Most methods [41, 16, 64] use losses that penalize geometric garment-body penetrations at training time, but all methods require exhaustive postprocess steps to fix collisions in test sequences. We address this remaining challenge with a self-supervised loss enabled by our generative space of garment deformations.

Our work is also related to the recent trend of implicit function learning to predict binary occupancy [33, 51, 14], signed distance to the surface [40, 2, 9] for any 3D point, or screen space RGB color given camera parameters [56, 35]. To the best of our knowledge, these methods have only been used to encode 3D surfaces, learn generative spaces for shapes, or neural rendering applications. In contrast, we propose two new uses for implicit function learning: to efficiently solve garment-body collisions with a novel fully differentiable loss term; and to smoothly diffuse 3D shape correctives and rigging weights to any 3D point.

3. Method Overview

Our goal is to learn a function to predict how a 3D garment dynamically deforms given a target human body pose and shape. In contrast to existing methods [16, 41, 53, 4], we put special emphasis in learning a model that directly outputs garment geometry that does not interpenetrate with the underlying human body, *i.e.,* it is physically correct after inference without requiring any post-processing. Hence, the final state is not compromised in terms of the regressed garment details such as wrinkles and dynamics.

To this end, in Section 4.1 we introduce an extension of standard statistical human body models [29] that learns to smoothly diffuse skinning surface parameters, such as rigging weights and blendshape correctives, to any point in 3D space. In Section 4.2, we leverage these learned diffused skinning parameters to define a novel garment deformation model. The key idea is to remove the deformations already captured by our diffused body model to built an *unposed* and *deshaped* canonical space of garments. In this space, garments appear in rest pose and mean shape but pose- and shape-dependent wrinkle details are preserved. In Section 4.3, we introduce a novel optimization-based strategy to project physics-based simulations to our canonical space of garments. Importantly, we show that the use of the learned diffuse skinning parameters is fundamental for this task, since they enable the correct representation of complex phenomena such as garment-body sliding or loose clothing.

Using projected physics-based simulations as ground-truth data, in Section 4.4 we describe how we learn a generative space of garment deformations. Key to our success is a novel self-supervised loss enabled by the canonical space of garments, which allows us to exhaustively sample *random* instances of garment deformations (*i.e.*, arbitrary shape, pose, and dynamics for which ground truth data is unavailable) and test collisions against a *constant* body mesh. Finally, in Section 5 we describe a neural network-based regressor that outputs deformed garments with dynamics, that do not interpenetrate the body, as a function of body shape and motion.

For detailed information about network architectures, training data, parameters, and other implementation details, please refer to the supplementary material.

4. Canonical Space of Garment Deformations

The central aim of our method is to obtain a regressor R that infers the deformation of the garment via

$$\mathbf{X} = R(\beta, \gamma), \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{N_\mathrm{G} \times 3}$ is the garment deformation in canonical space computed as a function of body shape β and motion descriptor γ. We will first describe how to obtain the canonical space into which the garment data is transformed, before detailing how the regressor R is trained in Section 5.

4.1. Diffused Human Model

Our garment model, defined later in Section 4.2, is driven by a new *diffused* human body model that extends current approaches in order to generalize to vertices beyond the body surface. More specifically, current body models [11, 29, 21] deform a rigged parametric human template

$$M_\mathrm{B}(\beta, \theta) = W(T_\mathrm{b}(\beta, \theta), J(\beta), \theta, \mathcal{W}) \tag{2}$$

where W is a skinning function (*e.g.*, linear blend skinning, or dual quaternion) with skinning weights $\mathcal{W}$ and pose parameters θ that deforms an unposed parametric body mesh $T_\mathrm{b}(\beta, \theta)$. The nowadays standard SMPL model [29] defines the unposed body mesh as

$$T_\mathrm{B}(\beta, \theta) = \mathbf{T}_\mathrm{b} + B_\mathrm{s}(\beta) + B_\mathrm{p}(\theta) \tag{3}$$

where $\mathbf{T}_\mathrm{b} \in \mathbb{R}^{N_\mathrm{B} \times 3}$ is a body mesh template with N_B vertices that is deformed using two blendshapes that output *per-vertex* 3D displacements: $B_\mathrm{s}(\beta) \in \mathbb{R}^{N_\mathrm{B} \times 3}$ models deformations to change the body shape; and $B_\mathrm{p}(\theta) \in \mathbb{R}^{N_\mathrm{B} \times 3}$ models deformations to correct skinning artifacts. Follow-up works propose additional blendshapes to model soft-tissue [43, 52] and garments [31, 1, 42].

We observe that existing data-driven garment models [53, 41] leverage the human body models defined in Equation 2 assuming that clothing closely follows the deformations of the body. Consequently, a common approach is to

borrow the skinning weights $\mathcal{W}$ to model the articulation of garments, usually by exhaustively searching the nearest body vertex for each garment vertex in rest pose. Our key observation is that such naive static assignment cannot correctly model complex nonrigid clothing effects. The reason is twofold: first, the garment-body nearest vertex assignment must be dynamically updated, for example, when a garment slides over the skin surface; and second, the garment-body vertex assignment cannot be driven only by the closest vertex since this causes undesirable discontinuities in medial-axis areas.

To address these weaknesses, we propose to extend existing body models formulated in Equation 2 by smoothly *diffusing* skinning parameters to any 3D point around the body. It is worth mentioning that we are not the first to diffuse surface parameters, but previous works are limited to interpolate inwards to create a volumetric mesh [23, 49] in a less smooth strategy. In Section 4.2 we show how our generalization of skinning parameters beyond the body surface is a fundamental piece for our novel garment model.

More formally, we define the functions $\widetilde{\mathcal{W}}(\mathbf{p})$, $\widetilde{B}_\mathrm{s}(\mathbf{p}, \theta)$, and $\widetilde{B}_\mathrm{p}(\mathbf{p}, \theta)$ that generalize skinning weights, shape blend-shape offset, and pose blendshape offset, respectively, to any point $\mathbf{p} \in \mathbb{R}^3$ by smoothly diffusing the surface values

$$\widetilde{\mathcal{W}}(\mathbf{p}) = \frac{1}{N} \sum_{\mathbf{q}_n \sim \mathcal{N}(\mathbf{p}, \mathbf{d})} \mathcal{W}(\phi(\mathbf{q}_n)) \tag{4}$$

$$\widetilde{B}_\mathrm{s}(\mathbf{p}, \beta) = \frac{1}{N} \sum_{\mathbf{q}_n \sim \mathcal{N}(\mathbf{p}, \mathbf{d})} B_\mathrm{s}(\phi(\mathbf{q}_n), \beta) \tag{5}$$

$$\widetilde{B}_\mathrm{p}(\mathbf{p}, \theta) = \frac{1}{N} \sum_{\mathbf{q}_n \sim \mathcal{N}(\mathbf{p}, \mathbf{d})} B_\mathrm{p}(\phi(\mathbf{q}_n), \theta) \tag{6}$$

where $\phi(\mathbf{p})$ computes the closest surface point to $\mathbf{p} \in \mathbb{R}^3$, $\mathbf{d}$ the distance from $\mathbf{p}$ to the surface body, and $B_\mathrm{p}(\mathbf{p}, \theta)$ a function that returns the 3D offset of the vertex $\mathbf{p}$ computed by the blendshape B_p. Notice that, for each point, we average the values of N neighbors and therefore mitigate potential discontinuities in areas around a medial-axis.

In order to obtain differentiable functions that seamlessly integrate into an optimization or learning process, we employ recent works on learning implicit functions and learn $\widetilde{\mathcal{W}}(\mathbf{p})$, $\widetilde{B}_\mathrm{s}(\mathbf{p}, \beta)$, and $\widetilde{B}_\mathrm{p}(\mathbf{p}, \theta)$ with fully-connected neural networks. This additionally yields a very efficient evaluation on modern GPUs.

4.2. Garment Model

Our next goal is to define a garment model that is capable of representing the deformations naturally present in real garments, including dynamics, high-frequency wrinkles, and garment-skin sliding. To this end, a common approach to ease this task is to decouple the deformations

caused by different sources, and model each case independently. For example, Santesteban *et al.* [53] decouple deformations due to shape and pose, and Patel *et al.* [41] due to shape, pose, and style. More specifically, as discussed in Section 4.1, both works model pose-dependent deformations leveraging the skinning weights associated with the body in the unposed state and a linear blend skinning technique. This disentanglement removes many nonlinear deformations and enables to efficiently represent (and learn) deformations due to other sources directly in an *unpose* (*i.e.*, normalized) state.

We propose to go one step further and remove the shape-dependent deformations already captured by the underlying human body model. This effectively constructs a canonical *unposed* and *deshaped* representation of garments, improving the disentanglement proposed by earlier works. As we show later in Section 4.4, this is a fundamental step towards learning a generative space of garment deformations that do not interpenetrate the underlying body.

To formulate our unposed and deshaped garment model we leverage the diffused skinning functions proposed in Section 4.1

$$M_G(\mathbf{X}, \beta, \theta) = W(T_G(\mathbf{X}, \beta, \theta), J(\beta), \theta, \widetilde{\mathcal{W}}(\mathbf{X})) \quad (7)$$

$$T_G(\mathbf{X}, \beta, \theta) = \mathbf{X} + \widetilde{B}_s(\mathbf{X}, \beta) + \widetilde{B}_p(\mathbf{X}, \theta), \quad (8)$$

where $T_G()$ is the deformed garment after diffused blend-shapes correctives are applied, and $\mathbf{X}$ are the garment deformations in canonical space. Notice that our garment model is well-defined for any garment with any topology, thanks to the generalized diffused skinning functions. (*i.e.,* no need to retrain $\widetilde{\mathcal{W}}(), \widetilde{B}_s(), \widetilde{B}_p()$ for each garment). The key property of this model is that skinning parameters used to articulate the garment (Equations 7 and 8) are defined as a function of the unposed and deshaped *deformed* garment $\mathbf{X}$. This is in contrast to existing methods [53, 41] that use a *fixed* weight assignment, usually defined in a relaxed state or template, and cannot guarantee that the rigging step of the regressed deformed garment does not introduce collisions.

4.3. Projecting the Ground-truth Data

Our ultimate goal is to learn the function $R()$ from Equation 1, which predicts garment deformations in canonical space, in a data-driven manner. However, obtaining ground truth data is not trivial since we need to project deformed 3D garments –computed with a physics-based simulator [37]– to the unposed and deshaped space. Previous methods formulate the projection to the unposed state as the inverse of the linear blend skinning operation [41, 53, 42]. Due to their static rigging weights assignment, this operation can introduce body-garment collisions in the unposed state for frames where the garment has deformed significantly or slid in the tangential direction of the body (see Figure 3b). Even
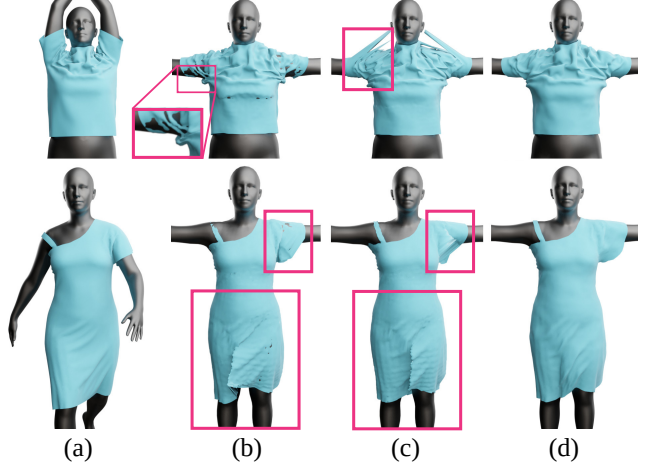

Figure 3: Unposing of a T-shirt and a dress in challenging poses: (a) input mesh; (b) unposing with constant weights [41, 53], notice the collisions; (c) unposing with variable weights assigned with nearest vertex, it avoids collisions but introduces skinning artifacts and is not temporally stable, better seen in video; (d) unposing with our optimization.

if a data-driven method can potentially learn to fix these artifacts to output collision-free *posed* deformations, our key contribution discussed in detail in Section 4.4 is to show that if a collision-free projection-and-unprojection operation exists, then the learning can be defined entirely in the unposed and deshaped state. This carries many positive properties that we discuss later.

We therefore need an strategy to project ground-truth garments to our canonical space, without introducing collisions. Notice that we cannot use the inverse of Equation 7 because the diffused skinning $\widetilde{\mathcal{W}}(\mathbf{X})$ are only defined for unposed shapes. Furthermore, exhaustive search of garment-body nearest vertices for each frame is highly expensive and introduces discontinuities in medial axis areas (see Figure 3c). Therefore, we propose a new optimization-based strategy to find the optimal vertex positions of the garment in the canonical space. Formally, given a ground-truth deformed garment mesh $\mathbf{M}_G$ (*i.e.,* generated with physics-based simulation) with known pose θ and shape β, we find its unposed and deshaped representation $\mathbf{X}$ by minimizing

$$\min_{\mathbf{X}} \quad \mathcal{E}_{rec} + \omega_1 \mathcal{E}_{strain} + \omega_2 \mathcal{E}_{collision}. \quad (9)$$

In the minimization objective, the data term

$$\mathcal{E}_{rec} = \left\| \mathbf{M}_G - M_G(\mathbf{X}, \beta, \theta) \right\|_2^2 \quad (10)$$

aims at reducing the difference between the simulated garment, and the unposed and deshaped representation projected back to the original state. Notice that $M_G(\mathbf{X}, \beta, \theta)$,

defined in Equation 7, is well defined for any set of 3D vertices $\mathbf{X}$, and it is fully differentiable thanks to the diffused skinning parameters.

The regularization term

$$\mathcal{E}_{\text{strain}} = \left\| \tfrac{1}{2} (F(\mathbf{T_G}(\mathbf{X},\beta,\theta))^\top F(\mathbf{T_G}(\mathbf{X},\beta,\theta)) - I) \right\|_2^2 \quad (11)$$

penalizes unrealistic deformations. To measure the amount of deformation of each triangle we use the Green-Lagrange strain tensor, which is rotation and translation invariant. F denotes the deformation gradient of each triangle.

Lastly, we include a term to prevent optimized vertex positions $\mathbf{X}$ to interpenetrate with the underlying body:

$$\mathcal{E}_{\text{collision}} = \max(\epsilon - SDF(\mathbf{X}), 0) \quad (12)$$

This term requires to compute the distance to the body surface for all vertices of the deformed garment, which is usually modeled with a Signed Distance Field (SDF). We leverage the fact that bodies in our canonical space are represented with a *constant body mesh*, and therefore the SDF is static and can be precomputed. In practice, and inspired by recent works on implicit function learning [40, 2, 9, 55], we learn the SDF with a shallow fully-connected network that naturally provides a fully differentiable formulation.

To optimize a sequence, we initialize the optimization with the result of the previous frame. This not only accelerates convergence, but also contributes to stabilize the projection over time. For the first frame, we initialize the optimization with the garment template, which is obtained by simulating the garment with the average body model (*i.e.*, θ and β set to zero).

4.4. Generative Garment Deformation Subspace

With the garment model defined in Section 4.2, and the strategy to project ground truth data into our canonical space defined in Section 4.3, we could train a data-driven method (*e.g.* a neural network) to learn the garment deformation regressor $R()$ defined in Equation 1. However, even though our garment model is designed in such a way that the (un)projection operation between canonical space and posed space does not introduce collisions, residual errors in the optimization of the regressor $R()$ could lead to regressed deformed garments $\mathbf{X}$ with body-garment collisions in the canonical space, which would inevitably propagate to the posed space. In fact, this is a common source of collisions in all data-driven methods [16, 41, 53, 64].

Our key contribution to address this challenge is to learn a compact subspace for garment deformations that *reliably solves* garment-body interpretations. To do so, we leverage the fact that in our unposed and deshaped canonical representation of garments, the underlying body shape is *constant*, namely, it is a body shape with $\beta = \mathbf{0}$ and $\theta = \mathbf{0}$. This property enables us to train a variational autoencoder

(VAE) to learn a *generative* space of garment deformations with a novel self-supervised collision loss term that is independent of the underlaying body and shape, and therefore naturally generalizes to arbitrary bodies. More specifically, we train the VAE with a loss

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{rec}} + \lambda_1 \mathcal{L}_{\text{laplacian}} + \lambda_2 \mathcal{L}_{\text{collision}} + \lambda_3 \mathcal{L}_{\text{KL}}. \quad (13)$$

We define the standard VAE reconstruction term as

$$\mathcal{L}_{\text{rec}} = \left\| \mathbf{X} - D(E(\mathbf{X})) \right\|_1, \quad (14)$$

where $E()$ and $D()$ are the encoder and decoder networks, respectively. Since $\mathcal{L}_{\text{rec}}$ does not take into account the neighborhood of the vertex, we add an additional loss term that penalizes error between the mesh laplacians [60, 65]

$$\mathcal{L}_{\text{laplacian}} = \left\| \Delta\mathbf{x} - \Delta D(E(\mathbf{X})) \right\|_1 \quad (15)$$

To enforce a subspace free of garment-body collisions, we propose the collision term

$$\mathcal{L}_{\text{collision}} = max(\epsilon - SDF(D(E(\mathbf{X})), 0) \\ + max(\epsilon - SDF(D(\bar{\mathbf{X}}_{\text{rand}})), 0) \quad (16)$$

where $\bar{\mathbf{X}}_{\text{rand}} \sim \mathcal{N}(0, 1)$. The first term penalizes collisions in the reconstruction of train data. Our fundamental contribution is the second term, $max(\epsilon - SDF(D(\bar{\mathbf{X}}_{\text{rand}})), 0)$, that samples the latent space and, enabled by the deshaped and unpose canonical representation, checks collisions against a *constant body mesh* with a self-supervised strategy (*i.e.*, we do not need ground truth garments for this term). This key ingredient allows us to exhaustively sample the latent space and learn a compact garment representation that reliably solves garment-body interpenetrations. As already highlighted, since our garment model is designed to not to introduce body-garment collisions in both the projection and unprojection operations, garment deformations regressed in the generative subspace do not suffer from collisions even in unseen (*i.e.,* test) sequences.

The self-supervised loss is only useful if the values are sampled from the same distribution as the data. For this purpose, we include an additional term $\mathcal{L}_{\text{KL}}$ to enforce a normal distribution in our latent space.

5. Regressing Garment Deformations

Once we have built our generative garment subspace, we encode the ground-truth data and use it to train the recurrent regressor $R(\beta, \gamma)$ from Equation 1, which predicts garment deformations as a function of body shape β and motion γ.

Our motion descriptor γ carries information of the current pose as well as its global movement. The off-the-shelf encoding for pose information is to use the joint rotations $\theta \in \mathbb{R}^{72}$ of the underlying human model, but this representation suffers from several problems such as discontinuities,
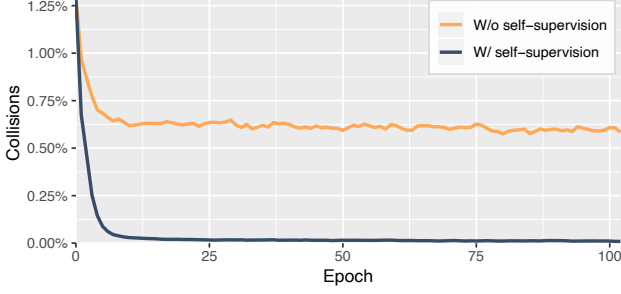
Figure 4: Number of body-garment collisions, evaluated in a test set, during the training of the generative subspace. Our novel self-supervised term, described in Equation 16, is the key term to reduce collisions in unseen sequences.

redundant joints, and unnecessary degrees of freedom. Instead, we adopt a more compact, learned pose descriptor $\bar{\theta} \in \mathbb{R}^{10}$ [52] which we found to generalize better. We build the motion vector γ for a given frame by concatenating the descriptor to the velocities and accelerations (computed with finite differences) of the pose, the global rotation $\mathbf{K}$ (represented as Euler angles) and translation $\mathbf{H}$

$$\gamma = \{\bar{\theta}, \frac{d\bar{\theta}}{dt}, \frac{d^2\bar{\theta}}{dt^2}, \frac{d\mathbf{H}}{dt}, \frac{d^2\mathbf{H}}{dt^2}, \frac{d\mathbf{K}}{dt}, \frac{d^2\mathbf{K}}{dt^2}\} \qquad (17)$$

The regressor takes as input the motion descriptor $\gamma \in \mathbb{R}^{42}$ and the shape coefficients $\beta \in \mathbb{R}^{10}$ and predicts the encoded garment deformation $\bar{\mathbf{X}}_{\text{pred}} \in \mathbb{R}^{25}$. To learn dynamic effects that depend on previous frames, we use Gated Recurrent Units [10] as the building blocks of our model. We train using the L1-error of encoded canonical space positions, velocities, and accelerations, which we find improves dynamics compared to optimizing positions alone.

$$\mathcal{L}_{\text{R}} = \mathcal{L}_{\text{pos}} + \rho_1 \mathcal{L}_{\text{vel}} + \rho_2 \mathcal{L}_{\text{acc}} \qquad (18)$$

6. Evaluation

6.1. Quantitative Evaluation

To quantitatively evaluate the ability of our compact generative subspace to solve body-garment collisions, we show the number of collisions during the training in Figure 4, evaluated on a *test set* that includes 4 unseen sequences and 17 different shapes. Specifically, we plot an ablation study that shows, in orange, the collisions remaining at each epoch when using only the supervised collision loss (*i.e.*, 1st term of Equation 16), and, in black, when also using our self-supervision with the 2nd term of Equation 16. The latter dramatically improves the collision handling, and it shows the generalization capabilities of our approach by reaching values close to 0 collisions in unseen sequences.

In Figure 5 we show a quantitative evaluation of the collisions in a test sequence from the AMASS dataset [32], and
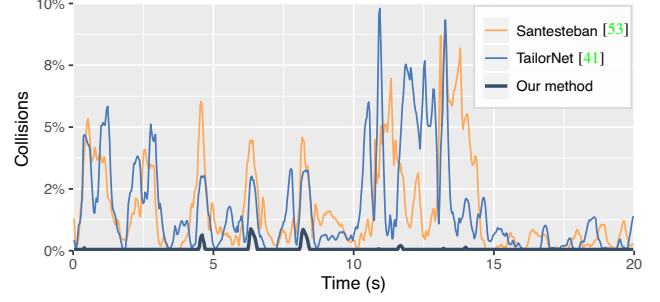

Figure 5: Quantitative evaluation of collisions per frame in test sequence 86_07. See supplementary video for a qualitative visualization of this comparison.

	TailorNet [41] (w/o postprocess)	Santesteban [53] (w/o postprocess)	Our method (w/o collision loss)	Our method (w/o self-supervision)	Our method
Collisions	5.70%	8.80%	0.62%	0.24%	**0.09%**

Table 1: Average number of collisions in 105 test motions from the AMASS dataset [32].

compare our results with the recent works of Santesteban *et al.* [53] and TailorNet [41]. These previous methods, without the postprocessing step, generate garment deformations that consistently collide with the underlying body mesh. In contrast, our method directly regresses garment deformations with almost no collisions. Importantly, the primary source of the remaining collisions for our method are self-intersections in the body mesh already present AMASS dataset (*e.g.*, a hand interpenetrates the torso).

Furthermore, in Table 1 we present an ablation study evaluated on test sequences from AMASS [32] with 53,998 frames and 20 body shapes. We report the number of collisions for 3 configurations of our method: without the full collision loss, without the self-supervised term, and the full model. All components of our model contribute, leading to a residual of 0.09% with our full model. In contrast, competitive methods suffer from a significantly higher number.

6.2. Qualitative Evaluation

We qualitatively evaluate the output of our method and compare to recent approaches. Please notice that the visual quality of our results is much better demonstrated in the supplementary video, which contains an exhaustive qualitative analysis, including a real-time demo in test sequences with unseen shapes from the AMASS dataset [32].

In Figure 6 we qualitatively evaluate the generalization capabilities our method to *unseen body shapes*. Specifically,

Figure 6: Generalization to new shapes. Interpolation between two unseen body shapes (left and right) from the AMASS dataset [32]. Our deshaped canonical space avoids collisions even in shapes far from the training data.

(a) TailorNet [41]
(w/o postprocess)

(b) TailorNet [41]
(w/ postprocess)

(c) Our method

Figure 7: Fixing collisions in a postprocess step can introduce undesired bulges, see chest area in (b).

we interpolate between 2 extremely different real shapes from AMASS [32], and compare to state-of-the-art data-driven garment models. Importantly, the input shapes are far beyond the range of our training data, therefore here we are also evaluating the *extrapolation* capabilities of the methods. Our method handles such extremely challenging cases very well and does not show visible garment-body collision, while existing methods [53, 41] suffer from very noticeable interpenetrations. In Figure 7 we show that, although a postprocessing step can effectively mitigate this issue, it can also introduce additional problems.

In Figure 8 we evaluate the generalization capabilities of our approach to *unseen motions*, and we compare our results against those of a physics-based simulator. Notice that our method is the first to showcase such a highly-challenging scenario featuring a dress sequence with dynamics.

6.3. Runtime Performance

In Table 2 we show our runtime performance in a regular desktop PC (AMD Ryzen 7 2700 CPU, Nvidia GTX 1080 Ti GPU, and 32GB of RAM). We produce detailed meshes at high frame rates, even for garments with many triangles.

Figure 8: Generalization to new motions. Qualitative comparison with physical simulation [37] (top) in sequence 01_01. Our model (bottom) synthesizes highly realistic dynamics and wrinkles even for challenging unseen motions.

	Triangles	Regressor	Decoder	Projection
T-shirt	8,710	1.7 ms	1.6 ms	1.4 ms
Dress	23,949	1.7 ms	3.5 ms	2.9 ms

Table 2: Execution time of each step of our model.

7. Conclusions and Future Work

We have presented a first algorithm to learn garment deformations such that they are essentially collision free. We achieve this by a diffused, volumetric representation of the underlying body together with the construction of a subspace for the garment that yields a differentiable, canonical configuration. This subspace is crucial for the regression of the garment deformation and its dynamics. Our algorithm not only allows for avoiding collisions, it also reduces complexity for inference, such that a learned representation yields higher quality than previously achievable. Generated garments exhibit a large amount of spatial and temporal detail, and can be inferred extremely quickly.

Additionally, the high quality of the garments generated by our method and the differentiable nature of our algorithm point to a large number of highly interesting avenues for follow up work: from virtual try-on applications [18], to inverse problems in computer vision [25]. Extending our method to handle collisions between multiple layers of clothing is also a promising line of future research.

References

[1] Thiemo Alldieck, Marcus Magnor, Bharat Lal Bhatnagar, Christian Theobalt, and Gerard Pons-Moll. Learning to Reconstruct People in Clothing from a Single RGB Camera. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2019. 2, 4

[2] Matan Atzmon and Yaron Lipman. SAL: Sign Agnostic Learning of Shapes from Raw Data. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2020. 3, 6

[3] Jan Bender, Matthias Müller, Miguel A Otaduy, Matthias Teschner, and Miles Macklin. A Survey on Position–Based Simulation Methods in Computer Graphics. *Computer Graphics Forum*, 33(6):228–251, 2014. 2

[4] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. CLOTH3D: Clothed 3D Humans. In *Proc. of European Conference on Computer Vision (ECCV)*, 2020. 1, 2, 3

[5] Kiran S. Bhat, Christopher D. Twigg, Jessica K. Hodgins, Pradeep K. Khosla, Zoran Popović, and Steven M. Seitz. Estimating Cloth Simulation Parameters from Video. In *Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, page 37–51, 2003. 2

[6] Sofien Bouaziz, Sebastian Martin, Tiantian Liu, Ladislav Kavan, and Mark Pauly. Projective Dynamics: Fusing Constraint Projections for Fast Simulation. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 33(4):1–11, 2014. 2

[7] Katherine L Bouman, Bei Xiao, Peter Battaglia, and William T Freeman. Estimating the Material Properties of Fabric from Video. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 1984–1991, 2013. 2

[8] Dan Casas, Marco Volino, John Collomosse, and Adrian Hilton. 4D Video Textures for Interactive Character Appearance. *Computer Graphics Forum (Proc. Eurographics)*, 33(2):371–380, 2014. 2

[9] Zhiqin Chen and Hao Zhang. Learning Implicit Fields for Generative Shape Modeling. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2019. 3, 6

[10] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. 3, 7

[11] Andrew Feng, Dan Casas, and Ari Shapiro. Avatar Reshaping and Automatic Rigging Using a Deformable Model. In *Proc. of ACM SIGGRAPH Conference on Motion in Games (MIG)*, page 57–64, 2015. 4

[12] Marco Fratarcangeli, Valentina Tibaldo, and Fabio Pellacini. Vivace: a Practical Gauss-Seidel Method for Stable Soft Body Dynamics. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 35(6):1–9, 2016. 2

[13] Valentin Gabeur, Jean-Sébastien Franco, Xavier Martin, Cordelia Schmid, and Gregory Rogez. Moulding Humans: Non-parametric 3D Human Shape Estimation from Single Images. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 2232–2241, 2019. 2

[14] Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T Freeman, and Thomas Funkhouser. Learning Shape Templates With Structured Implicit Functions. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[15] Peng Guan, Loretta Reiss, David A Hirshberg, Alexander Weiss, and Michael J Black. DRAPE: DRessing Any PErson. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 31(4), 2012. 2, 3

[16] Erhan Gundogdu, Victor Constantin, Amrollah Seifoddini, Minh Dang, Mathieu Salzmann, and Pascal Fua. GarNet: A two-stream network for fast and accurate 3D cloth draping. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2019. 1, 2, 3, 6

[17] Marc Habermann, Weipeng Xu, Michael Zollhofer, Gerard Pons-Moll, and Christian Theobalt. DeepCap: Monocular Human Performance Capture Using Weak Supervision. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[18] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S Davis. Viton: An image-based virtual try-on network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7543–7552, 2018. 8

[19] Zeng Huang, Yuanlu Xu, Christoph Lassner, Hao Li, and Tony Tung. ARCH: Animatable Reconstruction of Clothed Humans. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3093–3102, 2020. 2

[20] Ning Jin, Yilin Zhu, Zhenglin Geng, and Ron Fedkiw. A Pixel-Based Framework for Data-Driven Clothing. *Computer Graphics Forum (Proc. of SCA)*, 2020. 2

[21] Hanbyul Joo, Tomas Simon, and Yaser Sheikh. Total Capture: A 3D Deformation Model for Tracking Faces, Hands, and Bodies. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 4

[22] Ladislav Kavan, Dan Gerszewski, Adam W. Bargteil, and Peter-Pike Sloan. Physics-Inspired Upsampling for Cloth Simulation in Games. In *Proc. of ACM SIGGRAPH*, 2011. 2

[23] Meekyoung Kim, Gerard Pons-Moll, Sergi Pujades, Seungbae Bang, Jinwook Kim, Michael J. Black, and Sung-Hee Lee. Data-Driven Physics for Human Soft Tissue Animation. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 36(4), 2017. 4

[24] Zorah Lahner, Daniel Cremers, and Tony Tung. Deepwrinkles: Accurate and realistic clothing modeling. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018. 2

[25] Christoph Lassner, Gerard Pons-Moll, and Peter V Gehler. A Generative Model of People in Clothing. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pages 853–862, 2017. 8

[26] Yongjoon Lee, Sung-eui Yoon, Seungwoo Oh, Duksu Kim, and Sunghee Choi. Multi-Resolution Cloth Simulation. 29(7):2225–2232, 2010. 2

[27] John P Lewis, Matt Cordner, and Nickson Fong. Pose Space Deformation: A Unified Approach to Shape Interpolation and Skeleton-Driven Deformation. In *Proc. of the Annual Conference on Computer Graphics and Interactive Techniques*, pages 165–172, 2000. 2

[28] Jie Li, Gilles Daviet, Rahul Narain, Florence Bertails-Descoubes, Matthew Overby, George E Brown, and Laurence Boissieux. An Implicit Frictional Contact Solver for

Adaptive Cloth Simulation. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 37(4):1–15, 2018. 2

[29] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 34(6):1–16, 2015. 2, 3, 4

[30] Mickaël Ly, Jean Jouve, Laurence Boissieux, and Florence Bertails-Descoubes. Projective Dynamics with Dry Frictional Contact. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 39(4), 2020. 2

[31] Qianli Ma, Jinlong Yang, Anurag Ranjan, Sergi Pujades, Gerard Pons-Moll, Siyu Tang, and Michael J. Black. Learning to Dress 3D People in Generative Clothing. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 2, 4

[32] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of Motion Capture as Surface Shapes. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pages 5442–5451, Oct. 2019. 7, 8

[33] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy Networks: Learning 3D Reconstruction in Function Space. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[34] Eder Miguel, Derek Bradley, Bernhard Thomaszewski, Bernd Bickel, Wojciech Matusik, Miguel A Otaduy, and Steve Marschner. Data-driven estimation of cloth simulation models. *Computer Graphics Forum (Proc. Eurographics)*, 31:519–528, 2012. 2

[35] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Proc. of European Conference on Computer Vision (ECCV)*, 2020. 3

[36] Domen Mongus, Blaz Repnik, Marjan Mernik, and B Žalik. A hybrid evolutionary algorithm for tuning a cloth-simulation model. *Applied Soft Computing*, 12(1):266 – 273, 2012. 2

[37] Rahul Narain, Armin Samii, and James F O'brien. Adaptive Anisotropic Remeshing for Cloth Simulation. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 31(6):1–10, 2012. 1, 2, 5, 8

[38] Andrew Nealen, Matthias Müller, Richard Keiser, Eddy Boxerman, and Mark Carlson. Physically Based Deformable Models in Computer Graphics. *Computer Graphics Forum*, 25(4):809–836, 2006. 1, 2

[39] Hayato Onizuka, Zehra Hayirci, Diego Thomas, Akihiro Sugimoto, Hideaki Uchiyama, and Rin-ichiro Taniguchi. TetraTSDF: 3D Human Reconstruction From a Single Image With a Tetrahedral Outer Shell . In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6011–6020, 2020. 2

[40] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2019. 3, 6

[41] Chaitanya Patel, Zhouyingcheng Liao, and Gerard Pons-Moll. The Virtual Tailor: Predicting Clothing in 3D as a Function of Human Pose, Shape and Garment Style. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 2, 3, 4, 5, 6, 7, 8

[42] Gerard Pons-Moll, Sergi Pujades, Sonny Hu, and Michael J. Black. ClothCap: Seamless 4D clothing capture and retargeting. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 36(4), 2017. 2, 4, 5

[43] Gerard Pons-Moll, Javier Romero, Naureen Mahmood, and Michael J. Black. Dyna: A Model of Dynamic Human Shape in Motion. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 34(4), 2015. 4

[44] Albert Pumarola, Jordi Sanchez-Riera, Gary Choi, Alberto Sanfeliu, and Francesc Moreno-Noguer. 3DPeople: Modeling the Geometry of Dressed Humans. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 2242–2251, 2019. 2

[45] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 652–660, 2017. 2

[46] Abdullah-Haroon Rasheed, Victor Romero, Florence Bertails-Descoubes, Stefanie Wuhrer, Jean-Sébastien Franco, and Arnaud Lazarus. Learning to Measure the Static Friction Coefficient in Cloth Contact. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[47] Nadia Robertini, Dan Casas, Edilson De Aguiar, and Christian Theobalt. Multi-view performance capture of surface details. *International Journal of Computer Vision*, 124(1):96–113, 2017. 2

[48] Nadia Robertini, Dan Casas, Helge Rhodin, Hans-Peter Seidel, and Christian Theobalt. Model-Based Outdoor Performance Capture. In *Proc. of International Conference on 3D Vision (3DV)*, pages 166–175, 2016. 2

[49] Cristian Romero, Miguel A. Otaduy, Dan Casas, and Jesus Perez. Modeling and Estimation of Nonlinear Skin Mechanics for Animated Avatars. *Computer Graphics Forum (Proc. Eurographics)*, 39(2), 2020. 4

[50] Tom F. H. Runia, Kirill Gavrilyuk, Cees G. M. Snoek, and Arnold W. M. Smeulders. Cloth in the Wind: A Case Study of Physical Measurement Through Simulation. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[51] Shunsuke Saito, , Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2019. 3

[52] Igor Santesteban, Elena Garces, Miguel A. Otaduy, and Dan Casas. SoftSMPL: Data-driven Modeling of Nonlinear Soft-tissue Dynamics for Parametric Humans. *Computer Graphics Forum (Proc. Eurographics)*, 39(2), 2020. 4, 7

[53] Igor Santesteban, Miguel A. Otaduy, and Dan Casas. Learning-Based Animation of Clothing for Virtual Try-On. *Computer Graphics Forum (Proc. Eurographics)*, 38(2), 2019. 1, 2, 3, 4, 5, 6, 7, 8

[54] Yu Shen, Junbang Liang, and Ming C. Lin. GAN-based Garment Generation Using Sewing Pattern Images. In *Proc. of European Conference on Computer Vision (ECCV)*, 2020. 2

[55] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit Neural Representations with Periodic Activation Functions. In *Proc. NeurIPS*, 2020. 6

[56] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations. In *Advances in Neural Information Processing Systems*, pages 1121–1132, 2019. 3

[57] Carsten Stoll, Juergen Gall, Edilson De Aguiar, Sebastian Thrun, and Christian Theobalt. Video-based reconstruction of animatable human characters. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 29(6), 2010. 2

[58] Min Tang, Huamin Wang, Le Tang, Ruofeng Tong, and Dinesh Manocha. CAMA: Contact-Aware Matrix Assembly with Unified Collision Handling for GPU-based Cloth Simulation. *Computer Graphics Forum*, 35(2), 2016. 2

[59] Min Tang, Tongtong Wang, Zhongyuan Liu, Ruofeng Tong, and Dinesh Manocha. I-Cloth: Incremental Collision Handling for GPU-Based Interactive Cloth Simulation. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 37(6), 2018. 2

[60] Gabriel Taubin. A signal processing approach to fair surface design. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '95, page 351–358, New York, NY, USA, 1995. Association for Computing Machinery. 6

[61] Garvita Tiwari, Bharat Lal Bhatnagar, Tony Tung, and Gerard Pons-Moll. SIZER: A Dataset and Model for Parsing 3D Clothing and Learning Size Sensitive 3D Clothing. In *Proc. of European Conference on Computer Vision (ECCV)*, 2020. 1, 2

[62] Raquel Vidaurre, Igor Santesteban, Elena Garces, and Dan Casas. Fully Convolutional Graph Neural Networks for Parametric Virtual Try-On. *Computer Graphics Forum (Proc. SCA)*, 39(8), 2020. 1, 2, 3

[63] Huamin Wang, James F O'Brien, and Ravi Ramamoorthi. Data-Driven Elastic Models for Cloth: Modeling and Measurement. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 30(4):1–12, 2011. 2

[64] Tuanfeng Y Wang, Duygu Ceylan, Jovan Popović, and Niloy J Mitra. Learning a Shared Shape Space for Multimodal Garment Design. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 37(6), 2018. 1, 2, 3, 6

[65] Tuanfeng Y Wang, Tianjia Shao, Kai Fu, and Niloy J Mitra. Learning an Intrinsic Garment Space for Interactive Authoring of Garment Animation. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 38(6), 2019. 2, 6

[66] Jiajun Wu, Joseph J Lim, Hongyi Zhang, Joshua B Tenenbaum, and William T Freeman. Physics 101: Learning Physical Object Properties from Unlabeled Videos. In *The British Machine Vision Conference (BMVC)*, 2016. 2

[67] Hongyi Xu, Eduard Gabriel Bazavan, Andrei Zanfir, William T Freeman, Rahul Sukthankar, and Cristian Sminchisescu. GHUM & GHUML: Generative 3D Human Shape and Articulated Pose Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6184–6193, 2020. 2

[68] Jinlong Yang, Jean-Sébastien Franco, Franck Hétroy-Wheeler, and Stefanie Wuhrer. Analyzing Clothing Layer Deformation Statistics of 3D Human Motions. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018. 2

[69] Shan Yang, Junbang Liang, and Ming C. Lin. Learning-Based Cloth Material Recovery From Video. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2017. 2

[70] Tao Yu, Zerong Zheng, Yuan Zhong, Jianhui Zhao, Qionghai Dai, Gerard Pons-Moll, and Yebin Liu. Simulcap: Single-view human performance capture with cloth simulation. In *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[71] Chao Zhang, Sergi Pujades, Michael J Black, and Gerard Pons-Moll. Detailed, accurate, human shape estimation from clothed 3d scan sequences. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 4191–4200, 2017. 2

[72] Heming Zhu, Yu Cao, Hang Jin, Weikai Chen, Dong Du, Zhangye Wang, Shuguang Cui, and Xiaoguang Han. Deep Fashion3D: A Dataset and Benchmark for 3D Garment Reconstruction from Single Images. In *Proc. of European Conference on Computer Vision (ECCV)*, 2020. 2

[73] Javier S Zurdo, Juan P Brito, and Miguel A Otaduy. Animating Wrinkles by Example on Non-Skinned Cloth. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 19(1):149–158, 2012. 2