

# Fully Convolutional Graph Neural Networks for Parametric Virtual Try-On

Raquel Vidaurre<sup>1</sup>

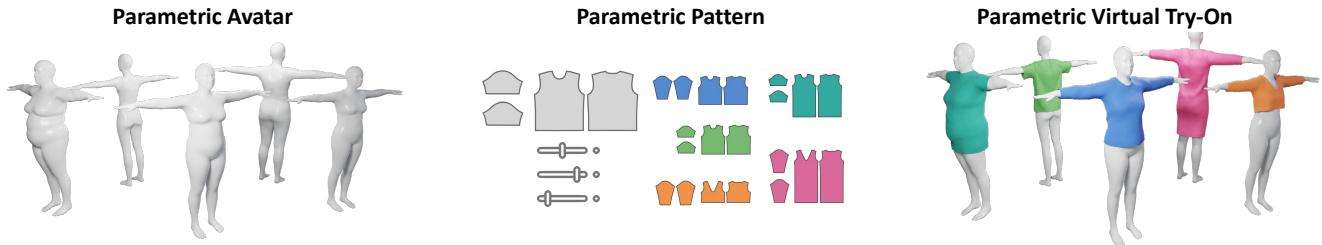
Igor Santesteban<sup>1</sup>

Elena Garces<sup>2</sup>

Dan Casas<sup>1</sup>

<sup>1</sup>Universidad Rey Juan Carlos, Madrid, Spain.

<sup>2</sup>SEDDI Labs, Madrid, Spain.



**Figure 1:** Our method predicts the 3D draping for an arbitrary body shape and garment parameters at interactive rates. From left to right, a variety of body shapes obtained from a parametric avatar model, different 2D panel configurations of our parameterized garment types, and corresponding dressed 3D bodies generated with our novel fully convolutional approach.

## Abstract

We present a learning-based approach for virtual try-on applications based on a fully convolutional graph neural network. In contrast to existing data-driven models, which are trained for a specific garment or mesh topology, our fully convolutional model can cope with a large family of garments, represented as parametric predefined 2D panels with arbitrary mesh topology, including long dresses, shirts, and tight tops. Under the hood, our novel geometric deep learning approach learns to drape 3D garments by decoupling the three different sources of deformations that condition the fit of clothing: garment type, target body shape, and material. Specifically, we first learn a regressor that predicts the 3D drape of the input parametric garment when worn by a mean body shape. Then, after a mesh topology optimization step where we generate a sufficient level of detail for the input garment type, we further deform the mesh to reproduce deformations caused by the target body shape. Finally, we predict fine-scale details such as wrinkles that depend mostly on the garment material. We qualitatively and quantitatively demonstrate that our fully convolutional approach outperforms existing methods in terms of generalization capabilities and memory requirements, and therefore it opens the door to more general learning-based models for virtual try-on applications.

## CCS Concepts

- Computing methodologies → Animation; Machine learning;

## 1. Introduction

The digitization of clothing is a long-standing goal in Computer Graphics and Animation, with important applications in many areas including garment design, virtual try-on, film visual effects, and video games. The classic –and still nowadays prevalent– approach to tackle the digitization of clothing is based on physics-based methods that simulate the deformations of garments in contact with the body [NMK\*06, NSO12, CLMMO14]. Despite the tremendous realism achieved with these methods, their high computational cost and potential instabilities hinder their deployment in

everyday applications. Many methods have been proposed to limit such computational cost, including coarse discretizations of the cloth [KGBS11, ZBO12], simplified simulation models [MC10], subspaces [DASTH10, FMD\*19], and tailored GPU-based solvers [TWL\*18].

As an alternative to physics-based methods, data-driven solutions aim at learning a function that approximates the ground truth deformations. Initial attempts used linear models [GRH\*12], which struggle to reproduce the complexity of wrinkles and dynamics inherent in clothing. More recently, deep learning methods

[WCPM18, SOC19, PLPM20, MYR<sup>\*</sup>20] have demonstrated that it is possible to learn realistic and highly efficient models for cloth animation. Most of these methods leverage existing parametric human models [LMR<sup>\*</sup>15] to use as input the shape and/or pose body parameters and output the corresponding deformed garment. Despite the realism of the predicted animations, a common underlying limitation of existing data-driven methods is the dramatic inability to generalize to new garments and mesh topologies. While physics-based methods naturally generalize to almost arbitrary combinations of garment meshes, most data-driven solutions are trained for a specific garment or mesh topology.

The reason for such limitation are the architectures used in existing solutions, which typically comprise one or more fully-connected layers (also known as multilayer perceptrons, MLP) [WCPM18, SOC19]. While this type of layers are known to be easy to use in any domain (images, meshes, etc.) since input values (*e.g.*, pixels, vertices, etc.) are just flattened into a vector, they carry a number of limitations. First, it constrains the size of the input vector to a fixed number, which enforces input meshes to have always the same topology. Second, it disregards spatial information due to the flattening vectors as input, which causes the loss of important local or neighboring information. And third, it requires a large number of parameters, since all input nodes (or neurons) are densely connected to each other.

In this work, we address such fundamental limitation in data-driven cloth by using a *fully convolutional graph neural network* (FCGNN). Fully convolutional architectures have shown to be successful in Euclidean domains, where the input data is regularly distributed over 2D or 3D grid. For example in image segmentation methods that take as input an image or volume of arbitrary size and produce a correspondingly-sized output with efficient per-pixel inference [SLD17, MNA16]. For non-Euclidian domains such as 3D meshes, the use of fully convolutional architectures is more challenging due to the non-trivial definition of a convolution operators in such irregular domain. To this end, we leverage recent research that formally defines the required operators for graph-like structures [DBV16, BBL<sup>\*</sup>17] and propose a FCGNN that, given a 3D *parametric garment (represented with known 2D panels)* with *arbitrary mesh topology* (*i.e.*, random) and a target body shape, outputs the accurate 3D draped garment.

Under the hood, our novel geometric deep learning approach learns to drape 3D garments by decoupling the three different sources of deformations that condition the fit of clothing: garment type, target body shape, and material. To this end, we initially build a parametric space for garment design that is capable of representing a large number of garment types, including loose dresses, shirts, t-shirts, and tight tops. Using this design space, we create a dataset of 3D garments and use physics-based simulation [NSO12] to dress a wide range of body shapes. We use this data to train three different networks, each of which serves for a different purpose in the virtual try-on pipeline. First, we learn a regressor that efficiently predicts, given the garment parameters, the coarse 3D drape of a garment onto the *mean* body shape. Then, to provide sufficient surface detail to each garment type, we apply a mesh topology optimization step that generates a regular and homogeneous size triangular mesh. Deformations caused by target body shape are modeled in our sec-

ond deformation step, which consists of a regressor that deforms the topology-optimized *mean shape* fitted garment as a function of the body shape. Our final step further deforms the garment to account for material-specific deformations, which mostly produce fine-scale wrinkles. Furthermore, we fine-tune our regressors with a novel *self-supervised* (*i.e.*, does not require on physically-based simulated data) strategy that penalizes body-cloth collisions.

All in all, our main contribution is a novel geometric deep learning framework that is able to cope with parametric garments represented as predefined 2D panels, arbitrary mesh topology, and any target body shape. We discuss and evaluate the advantages of the proposed architecture, and compare with existing methods and other baselines. To the best of our knowledge, our approach is the first fully convolutional method (*i.e.*, no fully connected layers are used) for data-driven cloth.

## 2. Related Work

The modeling of clothing has been approached in different ways. Here we discuss existing works by grouping them into simulation, 3D reconstruction, and data-driven models.

**Cloth Simulation.** Physics-based simulation methods use discretizations of classical mechanics to deform cloth by solving an ordinary differential equation [NMK<sup>\*</sup>06]. Based on this strategy many approaches have been proposed, with differences in the underlying representation, numerical solution methods, collision detection, and constraints. Despite the high level of realism shown with these approaches, capable of modeling even yarn mechanics [KJM10, CLMO15], they usually have a significant runtime computational costs that hinders the use in interactive applications.

A wide variety of attempts exist to limit the computational cost. For example, using position-based dynamics [MHHR07, KCMF12, MCKM14], which produce approximated but plausible results, but may lack the realism needed for real-world applications such as virtual try-on. Other methods use subspaces or model reduction techniques [DASTH10, SB12, HDDN19, FMD<sup>\*</sup>19] to perform simulation in a reduced space. Projecting the equations of motion into the subspace is simple, but adding constraints or problem-specific details is challenging. Alternatively, some methods speed up physics-based simulation by adding details to low-resolution simulated meshes [KGBS11, ZBO12, GPV<sup>\*</sup>15] or simplified physical models [RPC<sup>\*</sup>10, MC10].

Physics-based approaches have also been proposed from a design perspective, where the user specifies garment parameters and simulation is used to compute the 3D drape [UKIG11, BGK<sup>\*</sup>13]. We also predict the 3D drape given garment parameters, but we seek to skip the computationally-expensive simulation step while being capable of handling any garment designs, mesh topology, and target body shape. Very recently, physics-based differentiable methods have been proposed [LLK19], which also enable the efficient optimization of design parameters to produce the desired 3D drape.

**Cloth Reconstruction.** As an alternative to simulation, 3D reconstruction methods aim at recovering the surface of real clothing. Reconstructed garments can potentially be used later to dress new

subjects or train data-driven models, as we discuss in the next subsection.

Early attempts required customized clothing or special patterns to capture the deforming surface of a worn garment [SSK\*05, WCF07]. Subsequent research by Bradley *et al.* [BPS\*08] succeed at reconstructing sequences of markerless garments, with relatively low wrinkle level detail, using a multi-view stereo approach. Notably, their method even achieves temporally-coherent geometry by enforcing as-isometric-as-possible inter-frame constraints. Detail upsampling techniques, similar to those used to enhance low-resolution simulated meshes, can be used in this context too [PZB\*09]. Zhou *et al.* [ZCF\*13] reduced the input requirements and reconstruct garments from a single image. They automatically detect the garment outline to roughly create a smooth 3D mesh that is further refined with shape-from-shading cues. A common limitation of these garment capture methods is their inability or significant difficulty to manipulate the reconstructed clothing, for example, as a function of the underlying body shape for virtual try-on applications.

A different trend in 3D reconstruction, usually referred to as *performance capture*, aims at recovering the full body of a dressed actor while moving [SH07, DAST\*08, VBMP08, XCZ\*18]. Pioneering methods use a full-body 3D template of the actor that it is deformed using an optimization scheme such that it matches images captured using a multi-camera studio [DAST\*08, VBMP08]. Follow up methods reduced the input requirements by employing just a single depth camera [ZFYY14, BBLR15], or even a monocular video [XCZ\*18, YPA\*18]. Recently, template-based reconstruction methods have also been proposed for outdoor settings [RCR\*16, XCZ\*18], and for animals on the wild [ZKB18]. Alternatively, template-free methods [SH07] combine visual hull-based techniques with stereo reconstruction to extract per-frame surface of the actor. Even if re-animation of captured performances is possible [CVCH14, PKC\*16], the main limitation of these approaches is the single mesh output used to represent both the human body and clothing, which hinders the digital manipulation of the captured garment.

There exist works that address the problem of segmenting reconstructed 3D meshes into body and clothing layers. Neophytou and Hilton [NH14] estimate the underlying body shape by fitting a parametric human model, and learn a clothing deformation model with the residual of the fit. The learned model can be then used to dress different body shapes. Similarly, Pons-Moll *et al.* [PMH17] present a remarkable multi-camera system that is capable of reconstructing the underlying human shape and multiple garment layers, with fine wrinkle detail, at 60fps. Reconstructed garments can be transferred to new body shapes, but the dynamics in new sequences may look unrealistic since they are just a copy of the captured deformations. Yang *et al.* [YFHW18] go one step beyond and enrich the captured dataset with simulated data that exhibits variations in clothing size and physical materials. This enables the learning of a richer garment deformation model, capable of representing semantic parameters such as material properties.

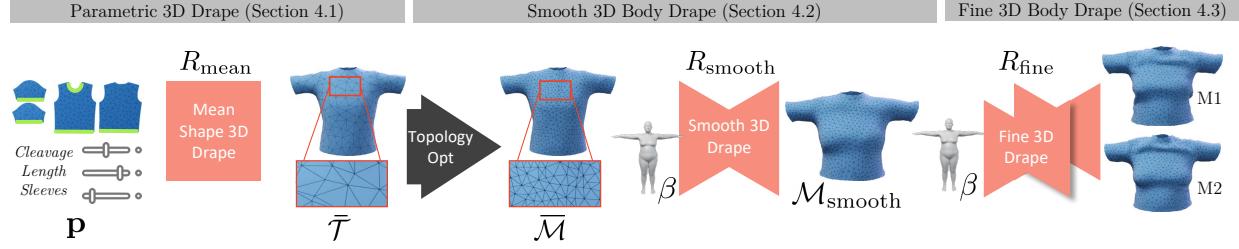
Deep learning techniques have also been proposed to address the *3D reconstruction* of garments. These methods circumvent the need for the error-prone model-fitting or optimization step in pre-

vious methods, and achieve a faster performance. Daněček *et al.* [DDÖ\*17] use synthetic data to train a CNN that regresses 3D vertex offsets to reconstruct a single garment from images. They require a known 3D template, and a tight crop of the image. Alldieck *et al.* [AMB\*19] and Bhatnagar *et al.* [BTTPM19] learn to reconstruct clothing and hair from video as displacements on top of SMPL human model [LMR\*15]. DeepWrinkles [LCT18] learns a 3D clothing deformation model from scans that is subsequently used to regress pose-dependent deformations for a specific garment. Even if our approach is not aimed for reconstructing tasks, we also use deep learning to regress 3D garment deformations. In contrast to these methods, our model generalizes parametric garments, does not rely on *a priori* known topology, and generalizes to different body shapes.

**Data-driven 3D Models.** Inspired by the success of the extensive literature in statistical 3D human body models learned from scans [ASK\*05, FCS15, LMR\*15, CCM\*18], with recent works even capable of learning highly dynamic soft skin deformations [SGOC20], many methods have proposed to learn 3D clothing models from data. Guan *et al.* [GRH\*12] use simulated data to learn to deform a garment as a function of the shape and pose of the underlying body. However, they rely on a linear model that struggles to learn fine details. Similarly, Xu *et al.* [XUC\*14] retrieve garment parts from a simulated dataset to synthesize pose-dependent clothing meshes, but shape deformations are not modeled. More recent methods use machine learning to predict garment deformations as a function of body pose alone [GCS\*19], or pose and shape [SOC19], some even capable of learning style [PLPM20] or animation dynamics as a function of fabric parameters [WSFM19]. A common limitation of these methods is the need to train a regressor for each garment, which hinders their deployment to massive use. In contrast, our method is able to learn to deform a large variety of garments using the same model.

The *design* of garments have been also tackled with data-driven models. Particularly relevant for us is the work of Wang *et al* [WCPM18], who learn a multi-modal subspace that allows editing the design of a specific garment using both 2D panel size and a sketch of the desired drape. Given a target body shape, the method outputs the 3D draped garment according to different input modalities. In contrast, we focus on the virtual try-on scenario instead of the design aspect. The fully-convolutional machinery of our method is able to cope with a wider range of clothing variability, ranging from tight tops to long loose dresses, using the *same* trained model.

Recently, following the success of graph CNNs (GCNNs) in non-euclidean domains such as triangular surfaces [DBV16, BBL\*17], some methods have explored the use of GCNN to learn 3D mesh deformations from data. Tan *et al.* [TGLX18] propose a mesh variational autoencoder (VAE) to learn an efficient latent space for 3D surfaces. The generative nature of the subspace enables to sample, interpolate, and explore deformable shapes. In a similar spirit, Ranjan *et al.* [RBSB18] use a graph convolutional autoencoder to learn a generative model for 3D faces. Closer to ours are the works that use GCNNs to learn garment deformations. Ma *et al.* [MYR\*20] employ a conditional mesh VAE-GAN [LSLW16] to compute additive displacements that are applied onto a full-body



**Figure 2:** Our pipeline for parametric virtual try-on. First, we estimate a parametric 3D mesh  $\bar{T}$  for the mean human shape given a set of design parameters  $\mathbf{p}$ . Then, after a topology optimization step that allows us to deal with any input topology and fine geometric details, a fully convolutional regressor  $R_{\text{smooth}}$  estimates smooth 3D deformations  $\mathcal{M}_{\text{smooth}}$  for a target human shape  $\beta$ . Finally, this mesh is further refined using another fully convolutional regressor  $R_{\text{fine}}$  to take into account fine wrinkles and material-specific deformations, represented as  $M_1$  and  $M_2$  in the figure.

human mesh. Despite being able to reproduce global and local cloth deformations, they require a fully-connected layer and a fix size input mesh. Similarly, Bertiche *et al.* [BME20] propose a model that is able to learn deformations for a large number of garments, but cannot cope with varying topology. In contrast, our method is fully convolutional and, assuming a known garment design space with a predefined parametric 2D panel layout (see Figure 2 left), we are able to cope with *any* mesh topology with arbitrary number of vertices.

### 3. Overview

Our goal is to predict the accurate 3D draping of garments, worn by any body shape, for virtual try-on purposes. We put special emphasis on the ability to cope with a large variety of garments, a feature mostly ignored by existing works since it requires a model that can deal with varying topology input. To this end, we propose the three-stage approach depicted in Figure 2 that effectively decouples (and therefore, eases learning tasks) the different sources of deformations (*e.g.*, due to garment type, body shape, or material) that condition the fit of clothing.

Following the traditional garment design workflow, our first step (Section 4.1) uses a set of parameters  $\mathbf{p}$  to define the 2D sewing patterns of a garment (*e.g.*, sleeve length, chest circumference, etc.), and learns a regressor  $R_{\text{mean}}(\mathbf{p}) = \bar{T}$  to estimate the corresponding 3D mesh  $\bar{T}$  draped into the *mean* human shape. Then, in order to accurately represent all garments potentially designed with the parameters  $\mathbf{p}$  (*e.g.*, from tight sleeveless tops to loose dresses), our second step (Section 4.2) computes an optimized mesh topology  $\bar{\mathcal{M}}$  from the regressed 3D drape  $\bar{T}$ , and learns a regressor  $R_{\text{smooth}}$  to predict a smooth (*i.e.*, lacking fine wrinkles) fit  $\mathcal{M}_{\text{smooth}}$  onto the target body shape  $\beta$ . Lastly, the third step (Section 4.3) learns a regressor  $R_{\text{fine}}$  that predicts a deformed mesh  $\mathcal{M}_{\text{fine}}$  with the realistic draping of the garment into the *target* body shape  $\beta$ . Importantly, regressors  $R_{\text{smooth}}$  and  $R_{\text{fine}}$  are implemented in a novel fully convolutional graph neural network (FCGNN) that is able to cope with *any* combination of garment, topology, and target body. Furthermore, in order to resolve potential body-garment collisions due to small inaccuracies when predicting surface deformations (a common issue in learning-based garment deformation methods,

*e.g.* [WCPM18, SOC19]), in Section 4.4 we propose a novel self-supervised strategy to fine-tune the regressor  $R_{\text{fine}}$ . The proposed supervision is based on a geometric definition of the distance between garment vertices and body faces, and does not require ground truth data (*i.e.*, avoids the need for expensive cloth simulation).

## 4. Garment Parametric Virtual Try-On

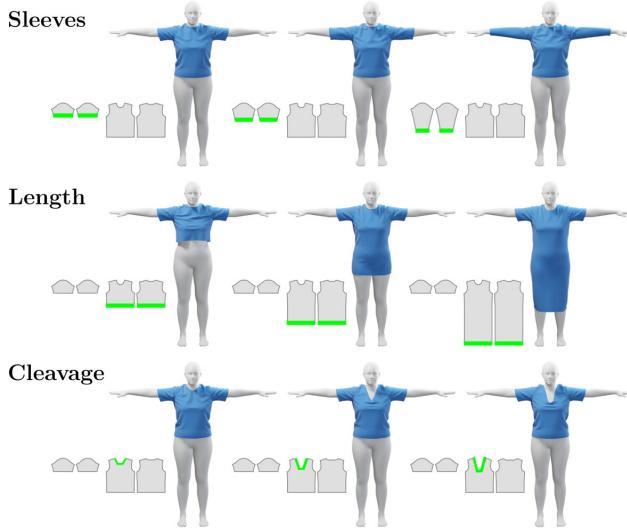
### 4.1. Parametric 3D Drape

In order to predict the 3D draping of garments for virtual try-on applications, we first need to define the actual garment type. Inspired by the traditional clothing manufacturing workflow, and similar to existing works [UKIG11, WCPM18], we characterize garment design properties using 2D sewing patterns. However, our observation is that we can use a single 2D layout to model a large family of garments by simply editing the length of some specific parts. For example, a tight sleeveless top and a long dress can be represented with the same 2D layout contour, with differences just in terms of size of each layout part. This is in contrast to the common use of 2D layouts, which are usually edited only to model size or small style changes. For example, Wang *et al.* [WCPM18] only allow minimal edits in 2D to change the style of the garment, and require independent models for dresses and t-shirts.

Based on this observation, our first step is learning to predict a *coarse* 3D drape of a garment given a specific 2D sewing pattern. In particular, we encode the parameters of the 2D layout (*e.g.*, sleeve length, chest circumference, etc.) in a vector  $\mathbf{p}$  that is fed into a non-linear regressor

$$R_{\text{mean}}(\mathbf{p}) = \bar{T} \quad (1)$$

that outputs the drape of the garment onto a *mean* human shape (note that in the rest of the paper we will use the overline symbol to refer to mean-shape-related variables). The motivation of this initial step is twofold: first, it roughly fits the garment on a generic human subject, which we use later in Section 4.2 to parameterize garment vertices using their closest body skinning weights; and second, it allows us to disentangle garment type-dependent deformations (*i.e.*, that depend on  $\mathbf{p}$ ) from material-dependent and body shape-dependent deformations.



**Figure 3:** Sewing pattern parameters (rows) used to build our dataset of garments. Each column shows the effect of the minimum, mean, and maximum values for each parameter.

To train our regressor  $R_{\text{mean}}(\mathbf{p})$  we build a dataset of 3D garments by manipulating a single 2D layout. Specifically, as shown in Figure 3, we manually edit parts of the 2D panels to design a family of garments including tops, t-shirts, sweaters, and short and long dresses. We then label each sample according to a set of measurements  $\mathbf{p}$  in the corresponding 2D representation, and simulate the sample worn by a mean human shape using a state-of-the-art physics-based cloth simulator [NSO12], with remeshing option turned off, until it reaches equilibrium to obtain a 3D mesh  $\bar{\mathcal{T}}$  of the draped garment. We implement the regressor  $R_{\text{mean}}: \mathbb{R}^P \rightarrow \mathbb{R}^{3 \times V^{\bar{\mathcal{T}}}}$  using a fully connected neural network that outputs the vertices positions of the mesh  $\bar{\mathcal{T}}$  with a predefined topology.

#### 4.2. Smooth 3D Body Drape

**Mesh Topology Optimization.** To accurately represent the draping of 3D garments with fine-scale detail it is necessary to use a topology with sufficient resolution (*i.e.*, number of triangles) for each garment type. Since one of our goals is to build a model that can predict the deformations for a large family of garments, we need to adapt the topology of the mesh  $\bar{\mathcal{T}}$  depending on the type of garment. To give a more practical example, we assume that the number of triangles required to represent high-quality draping of a t-shirt is smaller than those required for a long dress.

We model such garment type-dependent topology requirement by applying a remeshing operation to the coarse mean draped garment  $\bar{\mathcal{T}}$ . Specifically, we generate a new mesh

$$\bar{\mathcal{M}} = \phi(\bar{\mathcal{T}}, \mathbf{p}, T_{\text{dist}}, T_{\text{area}}), \quad (2)$$

where  $\phi()$  is a remeshing operation that, given an input mesh  $\bar{\mathcal{T}}$  and the 2D design parameters  $\mathbf{p}$ , aims at maintaining a (manually specified) average triangle distortion  $T_{\text{dist}}$  and surface area  $T_{\text{area}}$ .



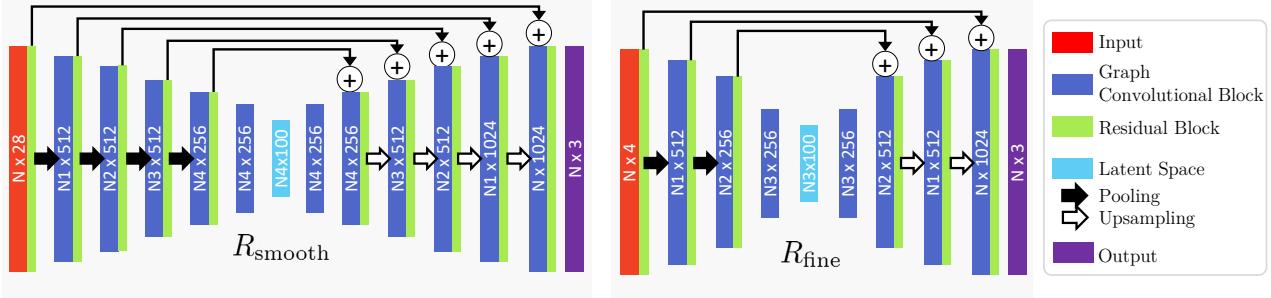
**Figure 4:** Garment type-dependent topology optimization, here shown in 2D panel space. Left: the 2D layout of a long dress design, with the template topology  $\bar{\mathcal{T}}$ . Right: the same design after the topology optimization step, resulting in the mesh  $\bar{\mathcal{M}}$  with homogeneous triangle size and without degenerated geometry.

Notice that these parameters are constant for all garments, therefore we only need to set them once. We implement  $\phi()$  based on the method proposed by Narain *et al.* [NSO12]. We write the optimized mesh as  $\bar{\mathcal{M}} = \{\mathbf{V}^{\bar{\mathcal{M}}}, \mathbf{E}^{\bar{\mathcal{M}}}\}$ , where  $\mathbf{V}^{\bar{\mathcal{M}}} \in \mathbb{R}^{3 \times V^{\bar{\mathcal{M}}}}$  are the vertices of the optimized surface, and  $\mathbf{E}^{\bar{\mathcal{M}}}$  the edges of the mesh. Figure 4 shows an example of the template topology  $\bar{\mathcal{T}}$  for a long dress design, which result in many degenerated triangles, and the optimized topology  $\bar{\mathcal{M}}$ . In practice,  $\phi()$  works in the UV-space of the 2D panels, which are automatically sew together to obtain  $\bar{\mathcal{M}}$ . We have simplified the notation for the sake of clarity. Notice that the *surface* of  $\bar{\mathcal{M}}$  and  $\bar{\mathcal{T}}$  is analogous, but their topology is different.

**Shape-Dependent Smooth Garment Deformation.** Having the optimized mesh topology  $\bar{\mathcal{M}}$  computed, in this second step we address the modeling of garment deformations caused by the target body shape. To represent parametric bodies, we use the popular model SMPL [LMR<sup>\*</sup>15], which provides a PCA-based representation of human bodies in T-pose, parameterized by  $\beta \in \mathbb{R}^{10}$ . We use the first component throughout the paper, since it encapsulates the largest variance in body shape. Importantly, SMPL also provides per-vertex rigging weights  $\mathbf{w}_i$ , which we use later in this section as a descriptor for garment vertices.

We therefore seek to learn a regressor  $R_{\text{smooth}}$  that deforms the mean shape garment  $\bar{\mathcal{M}}$  and outputs a mesh that reproduces a smoothed drape of the garment onto the target body shape  $\beta$ . We design  $R_{\text{smooth}}$  such that it learns global and smooth deformations, which has two main advantages: first, it eases the learning task since it reduces the variance in data and second, it decouples target body-dependent deformations (*i.e.*, global stretching and draping effects) from material-dependent (*i.e.*, fine wrinkles) deformations, which we will learn on a subsequent step. However, formulating such regression task is not trivial: the topology of the input mesh  $\bar{\mathcal{M}}$  is unknown at train time since we generate it at run time depending on the design parameters  $\mathbf{p}$ . Therefore, we cannot employ a fully connected network, where the input is a fix-size vector corresponding to the number of vertices of the mesh (a strategy commonly used in most of recent learning-based garment deformation methods [WSFM19, WCPM18, SOC19]) and, instead, we propose to use a graph-based fully convolutional architecture.

Two key ingredients are required to design the regressor  $R_{\text{smooth}}$  as a graph-based fully convolutional neural network: first, a convolution operator that is able to deal with graph input and, second, an



**Figure 5:** UNet-based architecture for regressors  $R_{\text{smooth}}$  (left) and  $R_{\text{fine}}$  (right). Each pooling or upsampling pass reduces or augments the number of nodes to half or double size. The input number of nodes is the same for both regressors, they differ in the number of intermediate layers, which is bigger for  $R_{\text{smooth}}$  as it has to learn a broader range of deformations.

efficient graph pooling operator that is able to coarsen the mesh by clustering together similar vertices. Specifically for this work, for graph convolutions we use the operator based on truncated Chebyshев polynomial proposed by Defferrard *et al.* [DBV16], which has shown to be very efficient given its linear computational complexity and constant learning complexity, like classical convolutional neural networks (*e.g.*, for images or other Euclidean domains). For mesh coarsening we use the approach proposed by Ranjan *et al.* [RBSB18], which consists of precomputing down- and upsampling matrices using a traditional method for surface simplification by Garland and Heckbert [GH97].

Having the operators defined, we now explain how we design our fully convolutional regressor  $R_{\text{smooth}}$ . Starting from the mean shape 3D drape mesh  $\bar{\mathcal{M}} = \{\mathbf{V}^{\bar{\mathcal{M}}}, \mathbf{E}^{\bar{\mathcal{M}}}\}$ , we first build an analogous undirected graph  $\bar{\mathcal{G}} = (\mathbf{N}, \mathbf{C})$ , with as many nodes and edges, as vertices and edges in the mesh,  $\mathbf{N} = \mathbf{V}^{\bar{\mathcal{M}}} \in \mathbb{R}^{3 \times V^{\bar{\mathcal{M}}}}$  and  $\mathbf{C} = \mathbf{E}^{\bar{\mathcal{M}}} \in \mathbb{R}^{3 \times E^{\bar{\mathcal{M}}}}$ , which we wish to use as input to the graph neural network. However, using vertices position as a descriptor for the graph nodes does not leverage all the information available in this context. Our key observation is that we can also append *semantic body part* information into the graph. To this end, for each garment vertex  $\mathbf{v}_i^{\bar{\mathcal{M}}}$  we find the closest body vertex  $\mathbf{v}_k^B$ , and append its associated rigging weights  $\mathbf{w}_k$  into each graph node descriptor. Additionally, we also append the shape descriptor  $\beta$  to each node. Therefore, the  $i^{\text{th}}$  node of the graph  $\bar{\mathcal{G}}$  is defined as  $\mathbf{n}_i = \{\mathbf{v}_i^{\bar{\mathcal{M}}}, \mathbf{w}_k, \beta\} \in \mathbb{R}^{3+J+|\beta|}$ , where  $J$  is the number of body joints (24 for SMPL [LMR<sup>\*</sup>15]), and  $|\beta|$  the number of shape coefficients (1 for the results shown in this paper).

We then input the graph  $\bar{\mathcal{G}}$  into our fully convolutional regressor

$$R_{\text{smooth}}(\bar{\mathcal{G}}) = \Delta_{\text{smooth}} \quad (3)$$

to predict a vector of 3D displacements  $\Delta_{\text{smooth}} \in \mathbb{R}^{3 \times V^{\bar{\mathcal{M}}}}$ . The architecture of the network, inspired by the success of fully convolutional U-Net [RFB15] for image segmentation, is depicted in Figure 5. The final deformed mesh of this second stage is then computed by adding the predicted 3D offsets to the mean shape 3D

drape

$$\mathcal{M}_{\text{smooth}} = \bar{\mathcal{M}} + \Delta_{\text{smooth}}. \quad (4)$$

To train the regressor  $R_{\text{smooth}}$  we create a dataset of ground-truth deformations of two different materials and a range of body shapes using the physics-based cloth simulation [NSO12]. We leverage the whole set of training data without introducing bias due to material-dependent deformations by first applying a Laplacian smoothing operator to each generated mesh, and then computing the average of each corresponding sample (*i.e.*, those with same topology, garment type, and target shape) before subtracting it from the mean shape to obtain the displacements  $\Delta_{\text{smooth}}^{\text{GT}}$ . As a loss function we use the  $\ell^2$ -norm of the error between ground truth displacements and predictions, in addition to the  $\ell^2$  regularization of the network weights

#### 4.3. Fine 3D Body Drape

The garment mesh  $\mathcal{M}_{\text{smooth}}$  successfully reproduces the global garment deformations due to target body shape, but lacks fine details that depend largely on the material. We address such source of deformations in this third and last step by further deforming the garment mesh. To this end, we learn to regress a new set of 3D displacements  $\Delta_{\text{fine}}$  using a fully convolutional network that takes as input a graph  $\mathcal{G}$  built from the vertices positions  $\mathbf{v}_i^{\mathcal{M}_{\text{smooth}}}$  and its associated rigging weights, analogous to the graph  $\bar{\mathcal{G}}$  described in Section 4.2

$$R_{\text{fine}}(\mathcal{G}) = \Delta_{\text{fine}}. \quad (5)$$

Our final predicted 3D drape  $\mathcal{M}_{\text{fine}}$  is then computed by adding the fine displacements onto the mesh  $\mathcal{M}_{\text{smooth}}$

$$\mathcal{M}_{\text{fine}} = \mathcal{M}_{\text{smooth}} + \Delta_{\text{fine}}. \quad (6)$$

To train the regressor  $R_{\text{fine}}$  we use the same simulated fits as in Section 4.2. However, in this case, we take advantage of the material-dependent deformations and train one regressor per material type. We generate the ground truth offsets  $\Delta_{\text{fine}, m}^{\text{GT}}$  per each material  $m$  by subtracting the smoothed fits from the simulated fits. As loss function for  $R_{\text{fine}}$  we use the same loss as  $R_{\text{smooth}}$ , with the ground truth fine-scale displacements  $\Delta_{\text{fine}, m}^{\text{GT}}$  instead.

#### 4.4. Self-Supervised Learning of Body-Garment Collisions

The objective losses used to train regressors  $R_{\text{smooth}}$  and  $R_{\text{fine}}$  minimize the reconstruction error but, due to expected residual errors in unseen shapes and topologies, this term alone does not guarantee predicted deformations to be free of body-garment collisions. This is a common issue in learning based solutions, which has been addressed with rendering tricks [DASTH10], postprocessing steps [SOC19], or explicit collision loss terms [GCS\*19] using supervised training. Inspired by the latter, we propose a collision loss term that we can train in a *self-supervised strategy*, and therefore does not require to generate expensive ground truth simulations. This is a major advantage over previous explicit collision losses.

Specifically, for each vertex of the garment  $\mathbf{v}_i^{\mathcal{M}}$  we find the closest body vertex  $\mathbf{v}_k^{\mathcal{B}}$  and compute the collision loss as

$$\mathcal{L}_{\text{collision}} = \max(-\mathbf{n}_k^{\mathcal{B}} \cdot (\mathbf{v}_i^{\mathcal{M}} - \mathbf{v}_k^{\mathcal{B}}), 0), \quad (7)$$

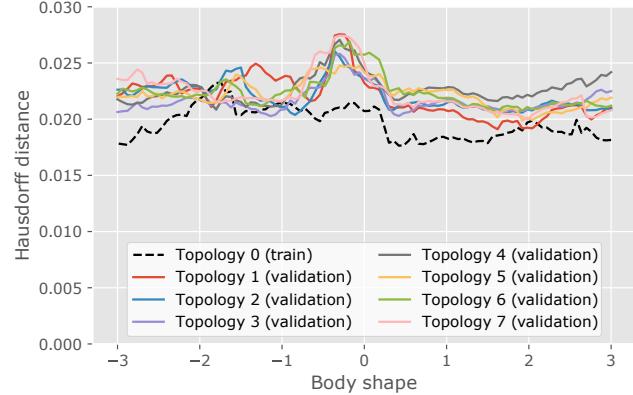
where  $\mathbf{n}_k^{\mathcal{B}}$  is the normal vector of the body vertex. The work of Gundogdu *et al.* [GCS\*19] uses this loss to penalize collisions during training, but unless the train dataset is exhaustive enough, this approach does not guarantee collision-free results for unseen inputs. In our particular case this is particularly bad, since creating an exhaustive dataset of cloth simulations is not feasible due to the arbitrary topology input of our method.

Therefore, starting from network weights trained for  $R_{\text{smooth}}$  and  $R_{\text{fine}}$ , we propose a novel strategy to fine-tune our networks using Equation 7 to produce collision-free results for arbitrary inputs. The key insight of our approach is that evaluating the collision loss does *not* require ground-truth data. Therefore, we can feed the network with random inputs and train on the collision loss only until it converges to a value near zero. To this end, during the self-supervised step we sample random body shapes  $\beta$  and garment topologies  $\mathcal{M}$ , feed them into our pipeline, and use the predicted mesh to fine-tune  $R_{\text{fine}}$  with Equation 7. Thanks to this strategy the number of collisions has been reduced by 70% during training, and 20% in validation.

### 5. Evaluation and Results

In this section we quantitatively and qualitatively evaluate our results in different scenarios. Specifically, we demonstrate our generalization capabilities, compare with the state-of-the-art method of Santesteban *et al.* [SOC19], and with a newly proposed brute force baseline for parametric virtual try-on.

**Dataset and Implementation Details.** Our ground truth dataset has been generated from 19 different garment pattern designs, two different topologies per design, and 201 values for the body shape  $\beta$  from the SMPL body model [LMR\*15], uniformly sampled within the range -3 and 3 (from which 100 have been exclusively used for test). The resulting meshes have between 1,414 (for the simpler case) and 3,581 (for long dresses) vertices.  $\mathcal{T}$  has a fixed size of 403 vertices, value which dynamically changes for  $\mathcal{M}$  depending on the garment complexity after the topology optimization step. To generate our data for the first step described in Section 4.1, in order to avoid potential topology-related problems (*e.g.*, highly distorted triangles, irregular vertex positions, etc.) at simulation time, we



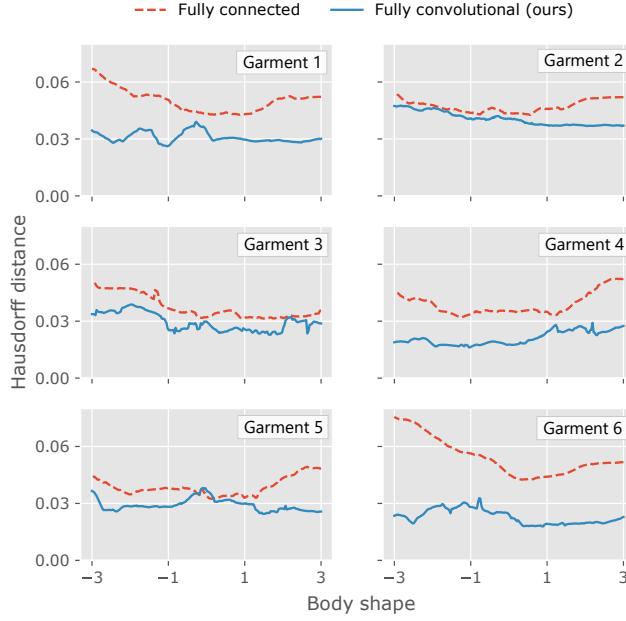
**Figure 6:** Generalization to new topologies. Hausdorff distance between the predicted and the ground truth meshes for a range of body shapes and 7 validation topologies. Errors in test topologies are consistent, demonstrating the generalization capabilities of our method, and on par to topologies used for training (dashed black).

first use a high-resolution mesh of 17,246 vertices, and then consistently downsample the simulated meshes to 403 vertices. 2D panel meshes are manually generated on a 3D modeling software, and the design parameters interpolate between these hand-made panels.

We have implemented our pipeline in TensorFlow for an efficient GPU training and execution. The parametric 3D draping is a fully connected layer with 3 input neurons (one per design parameter) and a single hidden layer (of ten neurons) trained for less than a minute. Training the fully convolutional networks  $R_{\text{smooth}}$ , and  $R_{\text{fine}}$  took approximately 20, and 14 hours respectively. Fine-tuning the self-supervised collisions took around one day. Everything was executed on a NVIDIA Titan X with 12GB.

**Evaluation of Generalization to New Topologies.** In Figure 6 we quantitatively evaluate the generalization capabilities of the regressors  $R_{\text{smooth}}$  and  $R_{\text{fine}}$  to new topologies. Specifically, for a given garment parameters  $\mathbf{p}$  and material for which we have ground truth simulated data, we randomized the topology (keeping the mean triangle area constant) of the mean shape predicted mesh  $\bar{\mathcal{M}}$ , and feed each topology to the regressors  $R_{\text{smooth}}$  and  $R_{\text{fine}}$  for a range of target shapes  $\beta$ . For each predicted mesh, we then compute the Hausdorff distance to the ground truth simulations. Results demonstrate that our method predictions are quantitatively consistent, regardless the topology and target body shape. Importantly, we also show that the error of the topologies unseen at train time (*i.e.*, validation set) is on par with the error of topologies used to train (in dash black).

**Comparison with Parametric Fully Connected Baseline.** Despite the lack of methods that can cope with parametric garments due to the need for different topologies, an alternative *brute-force* approach could be to use a highly-dense topology in  $\mathcal{M}$  to represent *all* garments, followed by a fully-connected end-to-end network that predicts displacements over such mesh. This high dense topology would provide an over-discretized mesh which, although unnecessarily complex for small garments such as a t-shirt, would provide sufficient details for large garments such as dresses, tech-



**Figure 7:** Quantitative evaluation of our fully convolutional (solid blue) approach and the fully-connected baseline (i.e., using the same highly-dense topology for all garments and a fully-connected architecture, dashed red), for 6 garment designs not present in the training set. Our approach consistently outperforms the fully connected baseline since the latter cannot generalize well to unseen garment types.

nically enabling the use of fully-connected pipelines [SOC19]. We implemented such solution, which can be considered a baseline for data-driven parametric garments, and compared it with our fully convolutional approach.

In Figure 7 we present a quantitative evaluation of the precision accuracy of our method, and the fully connected baseline. Specifically, for a given garment design (unseen at training time) we compute the Hausdorff error for a range of target body shapes, and demonstrate that our predictions  $M_{\text{fine}}$  are consistently more accurate. Our hypothesis is that the fully-connected approach cannot generalize to garment types outside the training set due to the *global* nature of the densely connected neurons, that are unable to learn local features. In contrast, the convolutional nature of our approach is able to capture local features, and therefore correctly predicts deformations of garment types unseen at train time but locally present in train examples.

Furthermore, we also evaluate the memory footprint of each method, which also results favorable for us. The fully connected network size is 167 MB, while ours ( $R_{\text{smooth}} + R_{\text{fine}}$ ) is 71MB. This is also expected, since the number of parameters for a fully connected network is significantly higher in comparison to the parameters used in the convolutional kernels. Note also that the fully connected approach needs to be fully trained for any new material while our approach enables easier generalization and transfer learning for new materials through fine-tuning  $R_{\text{fine}}$ .

**Comparison with Santesteban et al. 2019.** In Figure 8 we qualitatively compare our results with the state-of-the-art method of Santesteban *et al.* [SOC19], which is limited to a single garment. For a garment design analogous to the t-shirt used to train their method, we demonstrate that the predictions of both methods are on par (rows 1 and 2), while we are capable to predict the draping of a much larger number of garments (rows 3 and 4). This demonstrates the generalization capabilities of our method to arbitrary parametric garment design (and therefore, arbitrary topology).

**Qualitative Results.** In Figure 9 we show qualitative results of our method, for a variety of body shapes, garment types and topologies, all of them *unseen at train time*. Notice how the wrinkles predicted with our approach naturally match the expected behavior of the garment, and change for each shape-garment pair. This demonstrates that our method generalizes well to new garment types, topologies, and shapes. Check the supplementary video for more qualitative results.

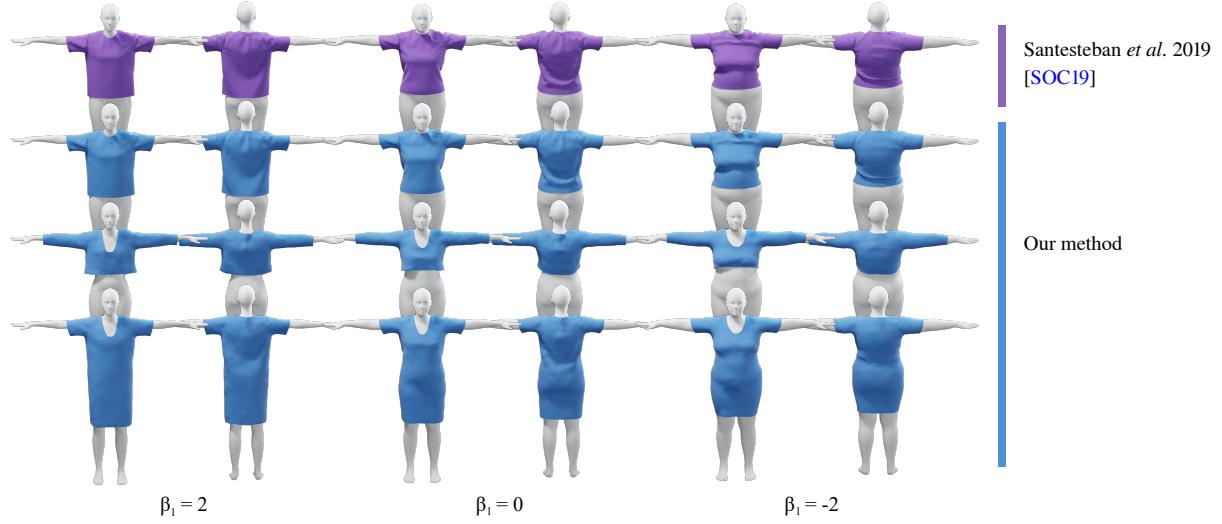
In Figure 10 we show qualitative predictions of our method, for two different materials, but the same target body shape and garment type (both unseen at train time). We demonstrate how our final step  $R_{\text{fine}}$  is able to learn material-specific deformations, resulting in visually different folds and wrinkles. Specifically for this comparison, the blue t-shirt is train on gray-interlock (60% Cotton, 40% Polyester) material and the pink on white-dots-on-black (100% Polyester) from ARCSim materials [NSO12]. See [WOR11] for additional material details.

## 6. Conclusions

We have presented a method to predict the drape of a predefined parametric space of garments onto an arbitrary target body shape. To achieve this, we propose a novel fully convolutional graph neural network that, in contrast to existing methods, it is not limited to a single garment or topology. Our novel pipeline, based on U-Net architecture and efficient graph convolutions, generalizes to unseen mesh topologies, garment parameters, and body shapes. To the best of our knowledge, ours is the first fully convolutional approach for virtual try-on purposes, which opens the door to more general data-driven cloth animation methods based on geometric deep learning.

Despite our step forward in geometric learning-based solutions for cloth animation, our approach still suffers from the following weaknesses that could be addressed by follow up works. Pose-dependent and material-dependent *input* parameters are not considered to our approach, and you need to retrain the model to consider these configurations. Multi-layer garments and contact with external forces are not considered either. Additionally, commercial garment design probably requires more than 3 parameters. The analysis of the scalability of the proposed method to a larger garment space remains open for future research.

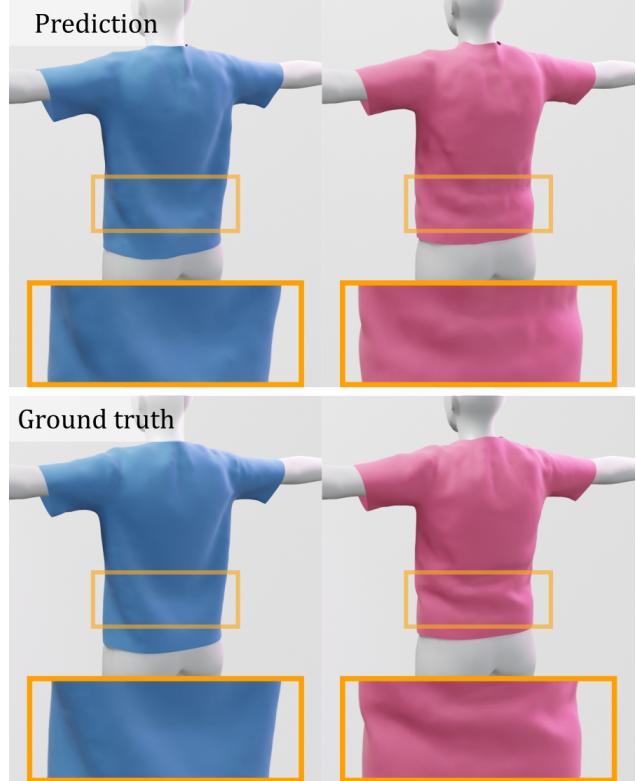
**Acknowledgments.** Igor Santesteban was supported by the Pre-doctoral Training Programme of the Department of Education of the Basque Government (PRE\_2019\_2\_0104), and Elena Garces was supported by a Torres Quevedo Fellowship (PTQ2018-009868). The work was also funded in part by the Spanish Ministry of Science (project RTI2018-098694-B-I00 VizLearning).



**Figure 8:** Qualitative comparison with the single-garment and fix topology method of Santesteban et al. [SOC19] and ours. When sampling the same garment type used to train their method, our results are on par with Santesteban’s (rows 1, 2), while our approach allows for a much richer space of garment types and topologies (rows 3, 4).



**Figure 9:** Virtual try-on results with our method, for a variety of garments (rows), fitted into a range of shapes (columns), both unseen at train time. Our method successfully predicts the drape of the garment, with natural folds and wrinkles at different scales that depend both on the input garment type and the target body shape.



**Figure 10:** Deformations regressed by our method for two different materials, presented in blue and pink. We demonstrate that, given the same target shape and input garment type, our method (top) is able to learn material-specific details that produce distinctive folds and wrinkles, closely matching the ground truth deformations (down).

## References

- [AMB\*19] ALLDIECK T., MAGNOR M., BHATNAGAR B. L., THEOBALT C., PONS-MOLL G.: Learning to Reconstruct People in Clothing from a Single RGB Camera. In *Proc. of Computer Vision and Pattern Recognition (CVPR)* (2019). 3
- [ASK\*05] ANGUELOV D., SRINIVASAN P., KOLLER D., THRUN S., RODGERS J., DAVIS J.: SCAPE: Shape Completion and Animation for PEople. In *Proc. of ACM SIGGRAPH* (2005), pp. 408–416. doi:[10.1145/1186822.1073207](https://doi.org/10.1145/1186822.1073207). 3
- [BBL\*17] BRONSTEIN M. M., BRUNA J., LECUN Y., SZLAM A., VANDERHEYNST P.: Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine* 34, 4 (2017), 18–42. doi:[10.1109/MSP.2017.2693418](https://doi.org/10.1109/MSP.2017.2693418). 2, 3
- [BBLR15] BOGO F., BLACK M. J., LOPER M., ROMERO J.: Detailed Full-Body Reconstructions of Moving People From Monocular RGB-D Sequences. In *Proc. of Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 2300–2308. doi:[10.1109/ICCV.2015.265](https://doi.org/10.1109/ICCV.2015.265). 3
- [BGK\*13] BERTHOUZOZ F., GARG A., KAUFMAN D. M., GRINSPUN E., AGRAWALA M.: Parsing Sewing Patterns into 3D Garments. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 32, 4 (2013), 1–12. doi:[10.1145/2461912.2461975](https://doi.org/10.1145/2461912.2461975). 2
- [BME20] BERTICHE H., MADADI M., ESCALERA S.: CLOTH3D: Clothed 3D Humans. In *Proc. of European Conference on Computer Vision (ECCV)* (2020). 4
- [BPS\*08] BRADLEY D., POPA T., SHEFFER A., HEIDRICH W., BOUBEKEUR T.: Markerless Garment Capture. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 27, 3 (2008), 99. doi:[10.1145/1360612.1360698](https://doi.org/10.1145/1360612.1360698). 3
- [BTTPM19] BHATNAGAR B. L., TIWARI G., THEOBALT C., PONS-MOLL G.: Multi-garment net: Learning to Dress 3D People from Images. In *Proc. of IEEE International Conference on Computer Vision (ICCV)* (2019), pp. 5420–5430. doi:[10.1109/ICCV.2019.00552](https://doi.org/10.1109/ICCV.2019.00552). 3
- [CCM\*18] CHENG Z.-Q., CHEN Y., MARTIN R. R., WU T., SONG Z.: Parametric modeling of 3D human body shape – A survey. *Computers & Graphics* 71 (2018), 88–100. doi:[10.1016/j.cag.2017.11.008](https://doi.org/10.1016/j.cag.2017.11.008). 3
- [CLMMO14] CIRIO G., LOPEZ-MORENO J., MIRAUT D., OTADUY M. A.: Yarn-level simulation of woven cloth. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)* 33, 6 (2014), 1–11. doi:[10.1145/2661229.2661279](https://doi.org/10.1145/2661229.2661279). 1
- [CLMO15] CIRIO G., LOPEZ-MORENO J., OTADUY M. A.: Efficient simulation of knitted cloth using persistent contacts. In *Proc. of ACM SIGGRAPH / Eurographics Symposium on Computer Animation (SCA)* (2015), pp. 55–61. doi:[10.1145/2786784.2786801](https://doi.org/10.1145/2786784.2786801). 2
- [CVCH14] CASAS D., VOLINO M., COLLOMOSSE J., HILTON A.: 4D Video Textures for Interactive Character Appearance. *Computer Graphics Forum (Proc. Eurographics)* 33, 2 (2014), 371–380. doi:[10.1111/cgf.12296](https://doi.org/10.1111/cgf.12296). 3
- [DAST\*08] DE AGUIAR E., STOLL C., THEOBALT C., AHMED N., SEIDEL H.-P., THRUN S.: Performance capture from sparse multi-view video. In *Proc. ACM SIGGRAPH* (2008). doi:[10.1145/1399504.1360697](https://doi.org/10.1145/1399504.1360697)
- [DASTH10] DE AGUIAR E., SIGAL L., TREUILLE A., HODGINS J. K.: Stable spaces for real-time clothing. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 29, 4 (2010). doi:[10.1145/1778843](https://doi.org/10.1145/1778843). 1, 2, 7
- [DBV16] DEFFERRARD M., BRESSON X., VANDERHEYNST P.: Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In *Proc. of International Conference on Neural Information Processing Systems (NIPS)* (2016), pp. 3844–3852. doi:[10.5555/3157382.3157527](https://doi.org/10.5555/3157382.3157527). 2, 3, 6
- [DDÖ\*17] DANÉREK R., DIBRA E., ÖZTIRELI C., ZIEGLER R.,
- GROSS M.: DeepGarment: 3D Garment Shape Estimation from a Single Image. *Computer Graphics Forum (Proc. Eurographics)* 36, 2 (2017), 269–280. doi:[10.1111/cgf.13125](https://doi.org/10.1111/cgf.13125). 3
- [FCS15] FENG A., CASAS D., SHAPIRO A.: Avatar Reshaping and Automatic Rigging Using a Deformable Model. In *Proc. of ACM SIGGRAPH Conference on Motion in Games* (2015), pp. 57–64. doi:[10.1145/2822013.2822017](https://doi.org/10.1145/2822013.2822017). 3
- [FMD\*19] FULTON L., MODI V., DUVENAUD D., LEVIN D. I., JACOBSON A.: Latent-space Dynamics for Reduced Deformable Simulation. *Computer Graphics Forum (Proc. Eurographics)* 38, 2 (2019), 379–391. doi:[10.1111/cgf.13645](https://doi.org/10.1111/cgf.13645). 1, 2
- [GCS\*19] GUNDOGDU E., CONSTANTIN V., SEIFODDINI A., DANG M., SALZMANN M., FUÀ P.: GarNet: A two-stream network for fast and accurate 3D cloth draping. In *Proc. of IEEE International Conference on Computer Vision (ICCV)* (2019). doi:[10.1109/ICCV.2019.00883](https://doi.org/10.1109/ICCV.2019.00883). 3, 7
- [GH97] GARLAND M., HECKBERT P. S.: Surface simplification using quadric error metrics. In *Proc. of the Annual conference on Computer Graphics and interactive techniques* (1997), pp. 209–216. doi:[10.1145/258734.258849](https://doi.org/10.1145/258734.258849). 6
- [GPV\*15] GILLETT R., PETERS C., VINING N., EDWARDS E., SHEFFER A.: Real-Time Dynamic Wrinkling of Coarse Animated Cloth. In *Proc. of ACM SIGGRAPH / Eurographics Symposium on Computer Animation (SCA)* (2015). doi:[10.1145/2786784.2786789](https://doi.org/10.1145/2786784.2786789). 2
- [GRH\*12] GUAN P., REISS L., HIRSHBERG D. A., WEISS A., BLACK M. J.: DRAPE: DRessing Any PErsom. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 31, 4 (2012). doi:[10.1145/2185520.2185531](https://doi.org/10.1145/2185520.2185531). 1, 3
- [HDDN19] HOLDEN D., DUONG B. C., DATTA S., NOWROUZEZHARAI D.: Subspace Neural Physics: Fast Data-Driven Interactive Simulation. In *Proc. of ACM SIGGRAPH / Eurographics Symposium on Computer Animation (SCA)* (2019). doi:[10.1145/3309486.3340245](https://doi.org/10.1145/3309486.3340245). 2
- [KCMF12] KIM T.-Y., CHENTANEZ N., MÜLLER-FISCHER M.: Long range attachments – A method to simulate inextensible clothing in computer games. In *Proc. of ACM SIGGRAPH / Eurographics Symposium on Computer Animation (SCA)* (2012), pp. 305–310. doi:[10.5555/2422356.2422399](https://doi.org/10.5555/2422356.2422399). 2
- [KGBS11] KAVAN L., GERSZEWSKI D., BARGTEIL A. W., SLOAN P.-P.: Physics-Inspired Upsampling for Cloth Simulation in Games. In *Proc. of ACM SIGGRAPH* (2011). doi:[10.1145/1964921.1964988](https://doi.org/10.1145/1964921.1964988). 1, 2
- [KJM10] KALDOR J. M., JAMES D. L., MARSCHNER S.: Efficient yarn-based cloth with adaptive contact linearization. In *Proc. of ACM SIGGRAPH* (2010). doi:[10.1145/1833349.1778842](https://doi.org/10.1145/1833349.1778842). 2
- [LCT18] LAHNER Z., CREMERS D., TUNG T.: Deepwrinkles: Accurate and realistic clothing modeling. In *Proc. of European Conference on Computer Vision (ECCV)* (2018). doi:[10.1007/978-3-030-01225-0\\_41](https://doi.org/10.1007/978-3-030-01225-0_41). 3
- [LLK19] LIANG J., LIN M., KOLTUN V.: Differentiable Cloth Simulation for Inverse Problems. In *Advances in Neural Information Processing Systems (NeurIPS)* (2019), pp. 771–780. 2
- [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 (2015), 1–16. doi:[10.1145/2816795.2818013](https://doi.org/10.1145/2816795.2818013). 2, 3, 5, 6, 7
- [LSLW16] LARSEN A. B. L., SØNDERBY S. K., LAROCHELLE H., WINTHER O.: Autoencoding beyond Pixels Using a Learned Similarity Metric. In *Proc. of International Conference on International Conference on Machine Learning (ICML)* (2016), pp. 1558–1566. 3
- [MC10] MÜLLER M., CHENTANEZ N.: Wrinkle meshes. In *Proc. of ACM SIGGRAPH / Eurographics Symposium on Computer Animation (SCA)* (2010), pp. 85–92. 1, 2

- [MCKM14] MÜLLER M., CHENTANEZ N., KIM T.-Y., MACKLIN M.: Strain Based Dynamics. In *Proc. of ACM SIGGRAPH / Eurographics Symposium on Computer Animation (SCA)* (2014), pp. 149–157. [doi:10.1145/2343483.2343501](https://doi.org/10.1145/2343483.2343501). 2
- [MHHR07] MÜLLER M., HEIDELBERGER B., HENNIX M., RATCLIFF J.: Position Based Dynamics. *Journal of Visual Communication and Image Representation* 18, 2 (2007). [doi:10.1016/j.jvcir.2007.01.005](https://doi.org/10.1016/j.jvcir.2007.01.005). 2
- [MNA16] MILLETARI F., NAVAB N., AHMADI S.-A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *International Conference on 3D Vision (3DV)* (2016), pp. 565–571. [doi:10.1109/3DV.2016.79](https://doi.org/10.1109/3DV.2016.79). 2
- [MYR\*20] MA Q., YANG J., RANJAN A., PUJADES S., PONS-MOLL G., TANG S., BLACK M. J.: Learning to Dress 3D People in Generative Clothing. In *Proc. of Computer Vision and Pattern Recognition (CVPR)* (2020). 2, 3
- [NH14] NEOPHYTOU A., HILTON A.: A layered model of human body and garment deformation. In *Proc. of International Conference on 3D Vision (3DV)* (2014), pp. 171–178. [doi:10.1109/3DV.2014.52](https://doi.org/10.1109/3DV.2014.52). 3
- [NMK\*06] NEALEN A., MÜLLER M., KEISER R., BOXERMAN E., CARLSON M.: Physically Based Deformable Models in Computer Graphics. *Computer Graphics Forum* 25, 4 (2006), 809–836. 1, 2
- [NSO12] NARAIN R., SAMII A., O'BRIEN J. F.: Adaptive Anisotropic Remeshing for Cloth Simulation. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)* 31, 6 (2012), 1–10. [doi:10.1145/2366145.2366171](https://doi.org/10.1145/2366145.2366171). 1, 2, 5, 6, 8
- [PKC\*16] PRADA F., KAZHDAN M., CHUANG M., COLLET A., HOPPE H.: Motion Graphs for Unstructured Textured Meshes. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 35, 4 (2016). [doi:10.1145/2897824.2925967](https://doi.org/10.1145/2897824.2925967). 3
- [PLPM20] PATEL C., LIAO Z., PONS-MOLL G.: 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). 2, 3
- [PMPH17] PONS-MOLL G., PUJADES S., HU S., M. J.: Cloth-Cap: Seamless 4D clothing capture and retargeting. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 36, 4 (2017). [doi:10.1145/3072959.3073711](https://doi.org/10.1145/3072959.3073711). 3
- [PZB\*09] POPA T., ZHOU Q., BRADLEY D., KRAEVOY V., FU H., SHEFFER A., HEIDRICH W.: Wrinkling Captured Garments Using Space-Time Data-Driven Deformation. *Computer Graphics Forum (Proc. Eurographics)* 28, 2 (2009), 427–435. [doi:10.1111/j.1467-8659.2009.01382.x](https://doi.org/10.1111/j.1467-8659.2009.01382.x). 3
- [RBSB18] RANJAN A., BOLKART T., SANYAL S., BLACK M. J.: Generating 3D Faces Using Convolutional Mesh Autoencoders. In *Proc. of European Conference on Computer Vision (ECCV)* (2018), pp. 725–741. [doi:10.1007/978-3-030-01219-9\\_43](https://doi.org/10.1007/978-3-030-01219-9_43). 3, 6
- [RCR\*16] ROBERTINI N., CASAS D., RHODIN H., SEIDEL H.-P., THEOBALT C.: Model-Based Outdoor Performance Capture. In *Proc. of International Conference on 3D Vision (3DV)* (2016), pp. 166–175. [doi:10.1109/3DV.2016.25](https://doi.org/10.1109/3DV.2016.25). 3
- [RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical image computing and computer-assisted intervention (MICCAI)* (2015), pp. 234–241. 6
- [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 (Proc. SIGGRAPH Asia)* 29, 6 (2010), 1–8. [doi:10.1145/1882261.1866183](https://doi.org/10.1145/1882261.1866183). 2
- [SB12] SIFAKIS E., BARBIC J.: FEM simulation of 3D deformable solids: a practitioner's guide to theory, discretization and model reduction. In *SIGGRAPH 2012 Courses*. ACM, 2012, pp. 1–50. [doi:10.1145/2343483.2343501](https://doi.org/10.1145/2343483.2343501). 2
- [SGOC20] SANTESTEBAN I., GARCES E., OTADUY M. A., CASAS D.: SoftSMPL: Data-driven Modeling of Nonlinear Soft-tissue Dynamics for Parametric Humans. *Computer Graphics Forum (Proc. Eurographics)* 39, 2 (2020). [doi:10.1111/cgf.13912](https://doi.org/10.1111/cgf.13912). 3
- [SH07] STARCK J., HILTON A.: Surface Capture for Performance-Based Animation. *IEEE Computer Graphics and Applications* 27, 3 (2007), 21–31. [doi:10.1109/MCG.2007.68](https://doi.org/10.1109/MCG.2007.68). 3
- [SLD17] SHELHAMER E., LONG J., DARRELL T.: Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 4 (2017), 640–651. [doi:10.1109/TPAMI.2016.2572683](https://doi.org/10.1109/TPAMI.2016.2572683). 2
- [SOC19] SANTESTEBAN I., OTADUY M. A., CASAS D.: Learning-Based Animation of Clothing for Virtual Try-On. *Computer Graphics Forum (Proc. Eurographics)* 38, 2 (2019). [doi:10.1111/cgf.13643](https://doi.org/10.1111/cgf.13643). 2, 3, 4, 5, 7, 8, 9
- [SSK\*05] SCHOLZ V., STICH T., KECKEISEN M., WACKER M., MAGNOR M.: Garment Motion Capture Using Color-Coded Patterns. *Computer Graphics Forum* 24, 3 (2005), 439–447. [doi:10.1111/j.1467-8659.2005.00869.x](https://doi.org/10.1111/j.1467-8659.2005.00869.x). 3
- [TGLX18] TAN Q., GAO L., LAI Y.-K., XIA S.: Variational Autoencoders for Deforming 3D Mesh Models. In *Proc. of Computer Vision and Pattern Recognition (CVPR)* (2018), pp. 5841–5850. [doi:10.1109/CVPR.2018.00612](https://doi.org/10.1109/CVPR.2018.00612). 3
- [TWL\*18] TANG M., WANG T., LIU Z., TONG R., MANOCHA D.: I-Cloth: Incremental Collision Handling for GPU-Based Interactive Cloth Simulation. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)* 37, 6 (2018). [doi:10.1145/3272127.3275005](https://doi.org/10.1145/3272127.3275005). 1
- [UKI11] UMETANI N., KAUFMAN D. M., IGARASHI T., GRINSPUN E.: Sensitive couture for interactive garment modeling and editing. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 30, 4 (2011). [doi:10.1145/2010324.1964985](https://doi.org/10.1145/2010324.1964985). 2, 4
- [VBMP08] VLASIC D., BARAN I., MATSIK W., POPOVIĆ J.: Articulated mesh animation from multi-view silhouettes. In *Proc. of ACM SIGGRAPH* (2008), pp. 1–9. [doi:10.1145/1399504.1360696](https://doi.org/10.1145/1399504.1360696). 3
- [WCF07] WHITE R., CRANE K., FORSYTH D. A.: Capturing and animating occluded cloth. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 26 (2007). [doi:10.1145/1275808.1276420](https://doi.org/10.1145/1275808.1276420). 3
- [WCPM18] WANG T. Y., CEYLAN D., POPOVIĆ J., MITRA N. J.: Learning a Shared Shape Space for Multimodal Garment Design. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)* 37, 6 (2018). [doi:10.1145/3272127.3275074](https://doi.org/10.1145/3272127.3275074). 2, 3, 4, 5
- [WOR11] WANG H., O'BRIEN J. F., RAMAMOORTHI R.: Data-Driven Elastic Models for Cloth: Modeling and Measurement. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 30, 4 (2011), 1–12. [doi:10.1145/2010324.1964966](https://doi.org/10.1145/2010324.1964966). 8
- [WSFM19] WANG T. Y., SHAO T., FU K., MITRA N. J.: Learning an Intrinsic Garment Space for Interactive Authoring of Garment Animation. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)* 38, 6 (2019). [doi:10.1145/3355089.3356512](https://doi.org/10.1145/3355089.3356512). 3, 5
- [XCZ\*18] XU W., CHATTERJEE A., ZOLLHÖFER M., RHODIN H., MEHTA D., SEIDEL H.-P., THEOBALT C.: MonoPerfCap: Human Performance Capture From Monocular Video. *ACM Transactions on Graphics* 37, 2 (2018). [doi:10.1145/3181973](https://doi.org/10.1145/3181973). 3
- [XUC\*14] XU W., UMENTANI N., CHAO Q., MAO J., JIN X., TONG X.: Sensitivity-optimized rigging for example-based real-time clothing synthesis. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 33, 4 (2014). [doi:10.1145/2601097.2601136](https://doi.org/10.1145/2601097.2601136). 3
- [YFHW18] YANG J., FRANCO J.-S., HÉTROY-WHEELER F., WUHRER S.: Analyzing Clothing Layer Deformation Statistics of 3D Human Motions. In *Proc. of European Conference on Computer Vision (ECCV)* (2018). [doi:10.1007/978-3-030-01234-2\\_15](https://doi.org/10.1007/978-3-030-01234-2_15). 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). doi:[10.1145/3026479](https://doi.org/10.1145/3026479). 3

[ZBO12] ZURDO J. S., BRITO J. P., OTADUY M. A.: Animating Wrinkles by Example on Non-Skinned Cloth. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 19, 1 (2012), 149–158. doi:[10.1109/TVCG.2012.79](https://doi.org/10.1109/TVCG.2012.79). 1, 2

[ZCF\*13] ZHOU B., CHEN X., FU Q., GUO K., TAN P.: Garment modeling from a single image. *Computer Graphics Forum* 32, 7 (2013), 85–91. doi:[10.1111/cgf.12215](https://doi.org/10.1111/cgf.12215). 3

[ZFYY14] ZHANG Q., FU B., YE M., YANG R.: Quality Dynamic Human Body Modeling Using a Single Low-cost Depth Camera. In *Proc. of Computer Vision and Pattern Recognition (CVPR)* (2014). doi:[10.1109/CVPR.2014.92](https://doi.org/10.1109/CVPR.2014.92). 3

[ZKB18] ZUFFI S., KANAZAWA A., BLACK M. J.: Lions and tigers and bears: Capturing non-rigid, 3d, articulated shape from images. In *Proc. of Computer Vision and Pattern Recognition (CVPR)* (2018), pp. 3955–3963. doi:[10.1109/CVPR.2018.00416](https://doi.org/10.1109/CVPR.2018.00416). 3