High-Fidelity Modeling of Generalizable Wrinkle Deformation

Jingfan Guo[1], Jae Shin Yoon[2], Shunsuke Saito[3], Takaaki Shiratori[3], and Hyun Soo Park[1]

[1] University of Minnesota
[2] Adobe Research
[3] Codec Avatars Lab, Meta

Abstract. This paper proposes a generalizable model to synthesize high-fidelity clothing wrinkle deformation in 3D by learning from real data. Given the complex deformation behaviors of real-world clothing, this task presents significant challenges, primarily due to the lack of accurate ground-truth data. Obtaining high-fidelity 3D deformations requires special equipment like a multi-camera system, which is not easily scalable. To address this challenge, we decompose the clothing into a base surface and fine wrinkles; and introduce a new method that can generate wrinkles as high-frequency 3D displacement from coarse clothing deformation. Our method is conditioned by Green-Lagrange strain field—a local rotation-invariant measurement that is independent of body and clothing topology, enhancing its generalizability. Using limited real data (*e.g.*, 3K) of garment meshes, we train a diffusion model that can generate high-fidelity wrinkles from a coarse clothing mesh, conditioned on its strain field. Practically, we obtain the coarse clothing mesh using a body-conditioned VAE, ensuring compatibility of the deformation with the body pose. In our experiments, we demonstrate that our generative wrinkle model outperforms existing methods by synthesizing high-fidelity wrinkle deformation from novel body poses and clothing while preserving the quality comparable to the one from training data.

1 Introduction

In the digital frontier, modeling virtual clothing has become increasingly important in many domains, ranging from fashion design to online shopping, and from animation to virtual reality. In immersive VR applications like telepresence, clothing plays a pivotal role in conveying the user identity, where the realism of clothing deformation is crucial for the authenticity of such an experience.

Simulating realistic clothing deformations based on the law of physics has been successful [5,7,28,33–36,54]. Yet the inverse problem, digitalizing real-world clothing and replicating its deformation behavior, is still challenging. Instead of solving this problem in a physical way, data-driven approaches present viable solutions. However, these methods also face significant challenges due to the lack of ground-truth data: capturing high-fidelity 3D clothing deformation [16,

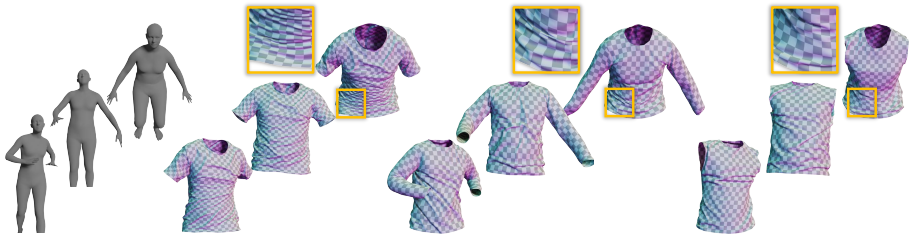


Fig. 1: Generalizable wrinkles: Given unseen SMPL body poses, we generate clothing deformation with realistic wrinkles. The high-frequency wrinkle deformation is learned from real captures of a short-sleeve T-shirt, which is generalizable to long-sleeve T-shirts and tank tops.

21, 39, 40, 51, 52] often requires expensive equipment such as a multi-camera system, which has poor scalability. The models learned from limited data often fail to produce realistic deformation due to the prediction ambiguity from unseen poses and clothing shapes, which often erases the high-frequency components, *e.g.*, wrinkles.

To take advantage of limited real clothing data capture and animate them with unseen body shapes and poses, existing methods [4, 31, 40, 43] rely on the prior knowledge of human body deformation. Specifically, they extend the linear blend skinning (LBS) from a parametric body [30] to clothing, so it can be animated with the body skeleton with the generalizability to unseen body poses. However, they have difficulty in modeling loose clothing like skirt, due to the lack of accurate body estimation under loose clothing, as well as the topological gap between human body and loose clothing. Other methods [26, 37, 56] propose to model wrinkle-level clothing deformation independently of body, aiming for learning clothing deformations that can be generalized to unseen body poses without requiring large-scale body-clothing pairwise data. However, they generate wrinkles based on the normal map [26, 56] or the vertex positions [37] of the coarsely deformed clothing. These global representations are still inherently coupled with the underlying body pose, limiting their generalizability when trained on small-scale data. There is no trivial conversion of these representations from global to local. For example, a tangent-space normal map of the coarse clothing surface contains no information to the geometry [26]. For the vertex positions, transforming them to a pose-invariant canonical space can make it generalizable, but finding this canonicalization for arbitrary clothing is non-trivial. Resorting to the body LBS for the canonicalization results in drawbacks as discussed above.

In this paper, we address this challenge by making the wrinkle generation dependent on a local representation of the coarse clothing state—Green-Lagrange strain. Since the strain measures how the clothing is locally deformed in comparison to its rest state in a rotation-invariant space, it is highly generalizable. Using this strain field representation, we propose to learn the correlation between the strain and the wrinkling from real clothing captures. Since the strain measurement is local, and its relationship with wrinkling is an intrinsic property

of the fabric material, this representation is completely independent of the body poses, and therefore, has the ability to generalize to unseen body poses.

Since clothing wrinkling is extremely complex and can be affected by many factors other than the strain measurement, we model the cloth wrinkling as a generative task conditioned on the strain. Specifically, we use a denoising diffusion model [18], which is shown to be effective on geometry modeling [15,24,46], for realistic wrinkling. We represent the wrinkling as vertex displacement in local coordinate frame on the base mesh, and train the diffusion model on real clothing with accurate registrations [16]. In particular, we generate base meshes by removing wrinkles from the ground-truth meshes, then compute vertex displacement between the base mesh and the ground-truth mesh, and transform it to the local coordinate frame on the base mesh. We also compute Green-Lagrange stain on the base mesh, and use it as conditioning to the diffusion model.

Based on the strain-conditioned diffusion model for wrinkle generation, we propose a practical pipeline to model realistic body-dependent clothing deformations for virtual avatars. In particular, we rig the clothing mesh with a virtual skeleton, which controls the coarse clothing deformation using LBS. We use a conditional variational auto-encoder (CVAE) [25, 47] to model the latent space of virtual skeleton transformations conditioned on body shape and pose. At inference time, the CVAE samples feasible virtual skeleton transformations given body shape and pose, which produces coarse clothing deformation. We compute Green-Lagrange strain on the coarsely deformed cloth, and use our diffusion model to generate realistic wrinkles on top of it. Our experiments on the test set of real captures and synthetic data demonstrate that our method can generate realistic wrinkles with strong generalizability to arbitrary body poses.

We summarize our contributions as follows:

- A strain-conditioned diffusion model for realistic clothing wrinkle generation, which is learned from high-fidelity real 3D clothing captures, with strong generalizability to unseen body poses.
- A practical pipeline for realistic body-dependent clothing deformations, where we learn diverse body-dependent coarse clothing deformation from synthetic simulation data using a VAE based network.
- Applications of our method to wrinkle transfer, material modeling, and down-stream tasks like image-based wrinkle fitting.

2 Related work

Cloth modeling has attracted significant attention in computer graphics and computer vision research communities. Recent efforts on modeling clothing deformation can be categorized into two groups: physics-based approaches and data-driven approaches.

Physics-based cloth simulation. Based on the established knowledge of elastic material mechanics, researchers have proposed numerous physics-based cloth simulation methods to model physically plausible clothing deformation [5, 7, 28, 33–36, 54]. These methods usually study the force exerted on the clothing by

its interaction with the body and environment, which is integrated over time to gradually deform the clothing geometry. Due to the intricate nature of clothing dynamics, some simplifications are often made to make the simulation tractable, sacrificing the realism of clothing deformation.

Among the physics-based cloth simulation methods, the coarse-to-fine strategy has been explored. Wrinkle Meshes [33] can run in parallel with a dynamic simulation of coarse clothing mesh, and add wrinkles to it. Chen et al. [7, 8] decompose a clothing geometry into a smooth base surface and fine wrinkles, where the wrinkles are represented as sinusoidal waves in the local frame of the base surface. Zhang et al. [54] propose an interactive tool for real-time quasi-static cloth simulation, where the user can control the coarse deformation of the cloth, and fine wrinkles will be progressively added to the coarsely deformed surface in a physically plausible way. These methods have shown promising results on realistic clothing deformation by only studying the fine wrinkles.

Cloth simulations usually generate open-loop clothing deformations, while using them to replicate a real-world clothing is non-trivial. Wang et al. [50] and Feng et al. [11] design special equipment to measure real-world fabrics' stretching and bending properties, representing them as material parameters in a computational elastic model that can be integrated into physics-based cloth simulators. However, they measure the material properties of the fabric in a local manner, which may not faithfully reflect the property of an actual clothing, where multiple pieces of fabrics are sewed together. Efforts have also been made on differentiable cloth simulation [10, 14, 20, 29, 41], which can be used in gradient-based simulation parameter estimation. However, these gradient-based methods require reasonable initial guess to start with, making it difficult to solve real-world problems. Moreover, Zhong et al. [57] have empirically shown that the gradient computation in these differentiable simulators may not match the analytical results even for a simple collision task, making it questionable to use them for real-world cloth modeling.

Data-driven cloth modeling. Researchers have been trying to model clothing deformation in a data-driven manner, which can be more flexible and scalable than physics-based methods. This is especially appealing when the clothing deformation can be learned from high-fidelity clothing captures [12, 16, 21, 31, 39, 40, 51, 52]. Since complex data processing is usually necessary in order to acquire the data, it is difficult to scale up the data collection. As a result, these data captures show limited diversity of body variations associated with each clothing. In addition, ground-truth body geometry is hard to acquire in the high-fidelity data captures, because the body is heavily occluded by the cloth. Only pseudo ground-truth body is available by fitting the shape and pose of a parametric body to the data capture.

Considering the fact that each clothing is captured with limited body variations, it requires some prior knowledge in order to deform the captured clothing beyond the body shape and poses seen during capture. Existing methods [4,31,40,43] take advantage of the parametric body, and extend the LBS from body to cloth, so the clothing can be animated by controlling the body skeleton.

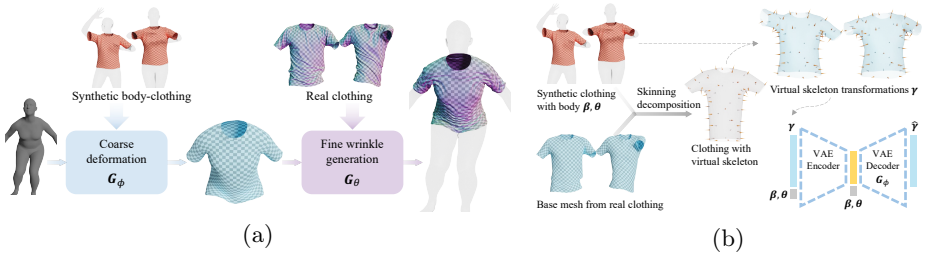
Although these methods can successfully model tight-fitting clothing, deformation artifacts usually present. More importantly, they have difficulty in modeling loose clothing like skirt. Pan et al. [37] have shown that clothing-specific LBS independent of body skeleton is promising for modeling loose clothing.

Another line of research is to combine data-driven techniques with physics-based simulation. Physics-based simulation is a useful tool for generating large-scale training data, so the body-dependent clothing deformation can be effectively learned [37,38,44,55]. Although the ideas directly apply to real-world data in theory, how to acquire large-scale real-world data that meet the training requirement is an open problem. Physics-inspired methods [2,3,13,45] use physics heuristics to enable unsupervised learning of clothing deformations, so realism of the result, especially for high-frequency wrinkles, is bounded by the underlying physics model. Our work is parallel to them, exploring an alternative direction to build a fully data-driven model that learns to reproduce real physical behaviors, thus pushing the boundaries of quality.

Among the above works, the idea of coarse-to-fine decomposition of clothing geometry has been extensively explored [6, 23, 26, 37, 38, 56]. Specifically, Kavan et al. [23] propose a simple data-driven framework accompanied with strong regularization to model wrinkle-level clothing deformation on top of a coarse simulation. Lahner et al. [26] and Zhang et al. [56] argue that the wrinkle-level cloth geometry details can be baked in normal maps, and propose methods that can learn to upsample low-resolution normal maps. Although the normal map can be lifted to 3D [56], it shows limited contribution to the improvement of surface geometry. Patel et al. [38] and Pan et al. [37] decompose the clothing geometry into low-frequency and high-frequency components, showing promising capability of modeling complex clothing deformation. However, their methods requires body-cloth data pairs that can only be synthesized, and hard to acquire in real-world as discussed above.

3 Method

The folding and wrinkling of clothing contribute most to our perception on its deformation, so we decompose the clothing surface into a smooth base surface and fine wrinkles, and model them separately as illustrated in Figure 2 (a). The base surface depends on the global body-clothing interaction, while the wrinkle model captures the local correlation between the state of base surface and fine wrinkles. This decomposition has a practical benefit that relaxes the requirement for body-clothing pairwise data from real capture. The body-dependent coarse clothing deformation can be learnt from large-scale synthetic body-clothing pairwise data. Although the synthetic data usually show low-quality clothing deformation, they only contribute to the coarse deformation component, which does not dominate our sense of realism on clothing deformation. In contrast, the realism of clothing deformation comes from the fine wrinkle component, which is learned from real clothing capture.



(a) (b)

Fig. 2: (a) At inference time, our method first generate coarse clothing deformation driven by parametric body, then generate realistic wrinkles on the base clothing surface. We train the coarse clothing deformation component using large-scale synthetic body-clothing pairwise data. The fine wrinkle generation component, which dominate the realism of clothing deformation, is learned from limited real clothing capture. (b) CVAE for body-dependent coarse clothing deformation. A clothing is rigged with a virtual skeleton for LBS. The virtual skeleton is optimized on the combination of the synthetic clothing and the base surface derived from the real clothing capture. The CVAE is trained to reconstruct the virtual skeleton transformations conditioned on SMPL parameters.

3.1 Problem Formulation

We represent the clothing as a thin elastic surface $\boldsymbol{f}(\boldsymbol{u}) : \Omega \to \mathbb{R}^3$, where the surface in the planar parameter space Ω is embedded in the 3D deformed space. Inspired by how wrinkles develop on real thin sheets [1, 7], we decompose the clothing surface into a smooth base surface $\boldsymbol{f}_b$ and a fine wrinkle field $\boldsymbol{f}_w$ as

$$\boldsymbol{f}(\boldsymbol{u}) = \boldsymbol{f}_b(\boldsymbol{u}) + \boldsymbol{f}_w(\boldsymbol{u}), \tag{1}$$

For a point $\boldsymbol{u} \in \Omega$, we can define a local coordinate frame as $\{\frac{\partial \boldsymbol{f}_b}{\partial u_1}, \frac{\partial \boldsymbol{f}_b}{\partial u_2}, \hat{\boldsymbol{n}}_b\}$, where the first two are tangential components, and $\hat{\boldsymbol{n}}_b$ is the unit normal on the base surface. We encode the wrinkles as a vector field $\tilde{\boldsymbol{w}}(\boldsymbol{u}) : \Omega \to \mathbb{R}^3$ in the local frame, which contributes to the wrinkle in the deformed space as

$$\boldsymbol{f}_w = \tilde{w}_1 \frac{\partial \boldsymbol{f}_b}{\partial u_1} + \tilde{w}_2 \frac{\partial \boldsymbol{f}_b}{\partial u_2} + \tilde{w}_3 \hat{\boldsymbol{n}}_b \tag{2}$$

where $\tilde{w}_i$ represents a scalar field in Ω, and the other terms are vector fields in Ω.

Existing works [7, 36] have explored how the stretching measurement like Green-Lagrange strain can affect the clothing deformation under physics-based settings. Considering a point $\boldsymbol{u} \in \Omega$ on the reference surface that is mapped to $\boldsymbol{x}_b \in \mathbb{R}^3$ on the deformed base surface, the stretching of the surface at the point can be measured by the Green-Lagrange strain tensor $\mathbf{E} \in \mathbb{R}^{2\times 2}$

$$\mathbf{E} = \frac{1}{2}\left(\mathbf{F}^\intercal \mathbf{F} - \mathbf{I}\right), \tag{3}$$

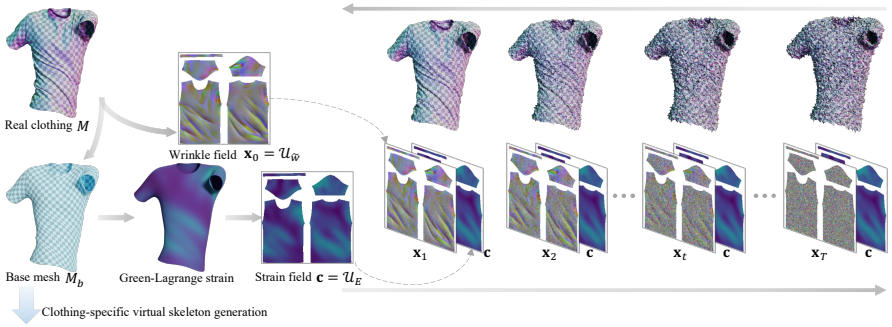


Fig. 3: Strain-conditioned diffusion model for fine wrinkles. We acquire base mesh by smoothing out the wrinkles on the real clothing capture, then construct a wrinkle field $\mathbf{x}_0$ as the vertex displacement from base mesh to wrinkled mesh in base mesh local coordinate frame unwrapped to UV space. Meanwhile, we compute Green-Lagrange strain on the base mesh, and parameterize it in UV space as the strain field $\mathbf{c}$. During training, the diffusion model takes $\mathbf{x}_0$ and $\mathbf{c}$ as inputs, adds Gaussian noise to $\mathbf{x}_0$, and learns to recover it conditioned on $\mathbf{c}$. At inference time, the diffusion model takes $\mathbf{c}$ as input, and generate $\mathbf{x}_0$ from Gaussian noise. The base mesh is used to compute clothing-specific virtual skeleton together with the synthetic cloth.

where $\mathbf{F} = \frac{\partial \boldsymbol{x}_b}{\partial \boldsymbol{u}} \in \mathbb{R}^{3\times 2}$ is the deformation gradient at point $\boldsymbol{x}_b$, and $\mathbf{I}$ is an identity matrix. We compute the magnitude of Green-Lagrange strain by taking its Frobenius norm $\|\mathbf{E}\|_F$, and use it to form a scalar field $E(\boldsymbol{u}) : \Omega \to \mathbb{R}$ on the entire surface, measuring the stretching on the base surface comparing to its rest state. We pursue a data-driven approach, and model the cloth wrinkling behavior based on the stretching of the base surface. Specifically, we aim to learn a generative model for the wrinkle field $\tilde{\boldsymbol{w}}$ conditioned on the stretching state of the base surface measured by E

$$\hat{\tilde{\boldsymbol{w}}} = \boldsymbol{G}_\theta(\boldsymbol{\epsilon}; E), \tag{4}$$

where $\boldsymbol{G}_\theta$ is a neural network parameterized by θ.

Discretization. In practice, we discretize the surface in Ω as a triangle mesh $\bar{M} = (\bar{V}, F)$ in UV space, where $\bar{V} \in \mathbb{R}^2$ is the set of vertex positions in the UV space, and F is the set of triangle faces. The canonical base surface, deformed base surface, and wrinkled surface embedded in 3D are discretized similarly as triangle meshes $\bar{M}_b = (\bar{V}_b, F)$, $M_b = (V_b, F)$ and $M = (V, F)$, respectively, where $\bar{V}_b, V_b, V \in \mathbb{R}^3$.

The vector field $\tilde{\boldsymbol{w}}$ and the scalar field E defined on the base surface in Ω are discretized as UV images $\mathcal{U}_{\tilde{\boldsymbol{w}}} \in \mathbb{R}^{H\times W\times 3}$ and $\mathcal{U}_E \in \mathbb{R}^{H\times W}$. They can be mapped to per-vertex and per-face quantities by the UV mapping given by $\bar{M}$.

3.2 Strain-Conditioned Fine Wrinkle Field

We learn $\boldsymbol{G}_\theta$ in Equation (4) from real 3D clothing data based on a conditional diffusion model [18], which consists of a forward process and a reverse process as shown in Figure 3.

In the forward process, we learn a transition probability from the complete wrinkle signal to a random noise $\mathbf{x}_T$ by adding noise

$$\mathbf{x}_t = \sqrt{1-\beta_t}\mathbf{x}_{t-1} + \beta_t\boldsymbol{\epsilon}, \qquad \mathbf{x}_0 = \mathcal{U}_{\tilde{\boldsymbol{w}}}, \tag{5}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a sample from the Gaussian distribution. We gradually increase the variance schedule β_t as increasing t, which reduces the impact of $\mathbf{x}_{t-1}$ while increasing that of Gaussian noise.

The reverse process gradually reconstructs the wrinkle signal from random noise by denoising:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t; \mathbf{c})\right) + \sigma_t\mathbf{z} \tag{6}$$

where $\mathbf{c} = \mathcal{U}_E$ is the strain field conditioning. Based on the variance schedule β_t, we define $\alpha_t = 1-\beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t}\alpha_i$, and $\sigma_t = \sqrt{\tilde{\beta}_t} = \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t}$. The learnable neural network $\boldsymbol{\epsilon}_\theta$ parameterized by θ aims to predict the noise $\boldsymbol{\epsilon}$ from corrupted data $\mathbf{x}_t$, conditioned on the strain field $\mathbf{c} = \mathcal{U}_E$. The generator $\boldsymbol{G}_\theta$ in Equation (4) is effectively the iterative reverse process that produces $\mathbf{x}_0$ from $\mathbf{x}_T$.

We train $\boldsymbol{\epsilon}_\theta$ with a weighted variational bound [18] as the objective:

$$L = \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t; \mathbf{c})\|^2\right] \tag{7}$$

3.3 Body-Dependent Base Surface

The wrinkle field generated in the above step is conditioned on, and will be added to, a base surface that shows coarse clothing deformation. The base clothing surface can deform in a variety of ways depending on its interaction with other entities, where the underlying body is a major source of such interaction. In pursue of the goal on modeling realistic clothed avatar, we seek to build the connection between the clothing deformation and the underlying body.

We represent the human body using SMPL [30], parameterizing a body geometry $\mathcal{M}(\boldsymbol{\theta}, \boldsymbol{\beta})$ as a function of body pose $\boldsymbol{\theta}$ and body shape $\boldsymbol{\beta}$. The base surface in 3D deformed space is controllable by clothing-specific virtual skeleton $\boldsymbol{\gamma} \in \mathbb{R}^{k\times 6}$, which is optimized [27] on a combination of real and synthetic cloth meshes. Each virtual bone's movement has 6 degrees-of-freedom, including translation $\boldsymbol{t} \in \mathbb{R}^3$ and rotation $\boldsymbol{r} \in SO(3)$.

The virtual skeleton transformations are decoded from a latent vector $\boldsymbol{z}$ conditioned on SMPL pose $\boldsymbol{\theta}$ and shape $\boldsymbol{\beta}$ parameters

$$M_b = \boldsymbol{W}\left(\bar{M}_b; \boldsymbol{\gamma}, \mathcal{W}\right), \tag{8}$$

$$\boldsymbol{\gamma} = \boldsymbol{G}_\phi(\boldsymbol{z}; \boldsymbol{\theta}, \boldsymbol{\beta}), \tag{9}$$

where $\boldsymbol{W}(\cdot)$ is a function that deforms the base mesh M_b from its rest pose $\bar{M}_b$ based on virtual skeleton transformations $\boldsymbol{\gamma}$ and skinning weight $\mathcal{W} \in \mathbb{R}^{|V_b| \times 6k}$, $\boldsymbol{G}_\phi$ is a neural network parameterized by ϕ. We learn $\boldsymbol{G}_\phi$ in Equation (9) from synthetic body-clothing pairwise data as the decoder in a conditional variational auto-encoder (CVAE) [25, 47] as shown in Figure 2 (b).

4 Experiments

4.1 Dataset and Settings

In order to train the coarse clothing deformation component, we take 49 body motion sequences from the AMASS [32] dataset, totaling 27415 frames, and simulate each clothing with the body. We post-process the simulated clothing meshes by curved Hessian smoothing to remove unnecessary high-frequency details.

We learn the fine wrinkles from the patterned clothing dataset [16], which provides accurate 3D registrations for the clothing geometry. To create smooth base meshes, we remove wrinkles from the ground-truth clothing meshes by smoothing based on curved Hessian energy [49]. We use two sequences T-shirt on "subject_00" (*T-shirt*) and skirt on "subject_04" (*Skirt*) in our experiments. The frames in each sequence are split into training set and in-distribution test set. Meanwhile, we take 6 smoothed simulation sequences, and form an out-of-distribution test set. The in-distribution test set only contains coarse clothing and GT wrinkled clothing, but no body pose, while the out-of-distribution test set has body pose and synthetic clothing. The pose distribution in the out-of-distribution test set highly deviates from the training set as Figure 4 shows. Please note that the body pose is not available in the real-world dataset, and we approximate it using a rigged T-shirt.

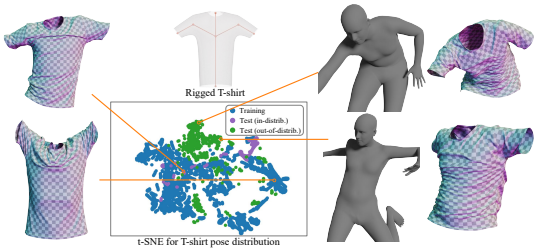


Fig. 4: (Left) Sample clothing from the training set and the in-distribution test set. (Middle) Pose approximation with rigging, and t-SNE for pose distribution of the dataset. (Right) Sample SMPL body and clothing from the out-of-distribution test set.

Implementation details. For the coarse deformation component, each clothing is rigged with 64 joints in the virtual skeleton. For the wrinkle generation component, we train the diffusion model with $T = 1,000$ steps, and inference using DDIM [48] with $S = 50$ steps and $\eta = 0$ in the reverse process. We use a linear variance schedule that increases from $\beta_1 = 10^{-4}$ to $\beta_T = 0.02$. The network $\boldsymbol{\epsilon}_\theta$ is implemented as a U-Net [42] that takes in a 256×256 UV image, consisting of both strain field and noisy wrinkle field.

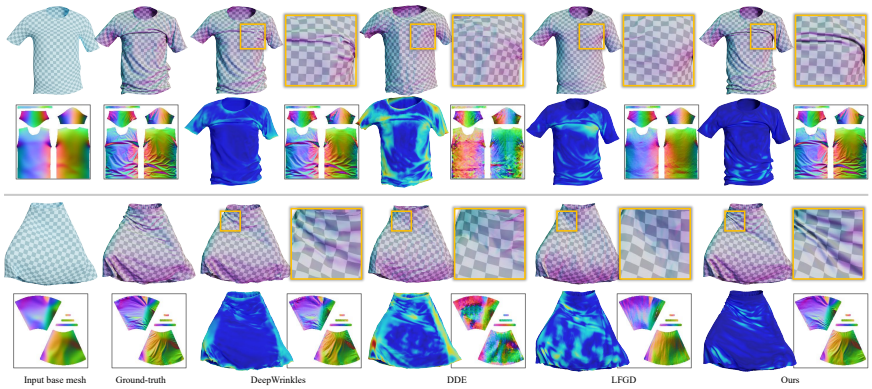


Fig. 5: Comparison of 3D wrinkle reconstruction with baseline methods. In each example, the top row shows the geometry, as well as a zoom-in for the result of each method. The bottom row shows the normal map, as well as the vertex error E_v in color ($0m$  $> 0.05m$). Please note that DeepWrinkles and DDE only generate normal maps, which are used to deform the base mesh. LFGD and our method predict 3D displacement, and the normal maps are derived from the deformed mesh.

Metrics. To quantitatively evaluate the geometric quality of the generated wrinkles on the in-distribution test set, we compute the $L2$ vertex distance E_v and the $L2$ Chamfer distance E_{cd} between the ground-truth mesh and the predicted mesh from each method:

$$E_v = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{v}_i - \hat{\mathbf{v}}_i\|_2\,, \quad E_{cd} = \frac{1}{|V|}\sum_{\mathbf{v}\in V}\min_{\hat{\mathbf{v}}\in\hat{V}}\|\mathbf{v}-\hat{\mathbf{v}}\|_2^2 + \frac{1}{|\hat{V}|}\sum_{\hat{\mathbf{v}}\in\hat{V}}\min_{\mathbf{v}\in V}\|\hat{\mathbf{v}}-\mathbf{v}\|_2^2$$

To evaluate the coarse deformation component on synthetic data, we report the $L2$ vertex distance E_v for the base mesh, as well as $L2$ joint distance E_j for the clothing-specific virtual skeleton. Moreover, as a perceptual evaluation for the quality of clothing wrinkles, we compute FID score [17] on UV-space normal maps between the real clothing captures and the results for each method.

4.2 Comparison with Baseline Methods

Comparison of 3D wrinkle reconstruction. We compare the wrinkle generation component of our method with DeepWrinkles [26], DDE [56], and LFGD [37] on the in-distribution test set. Please note that both DeepWrinkles and DDE aim at generating normal maps for rendering purpose, which is different from our method that focuses on generating detailed 3D geometry. We apply DDE's technique to deform the base mesh using normal map guidance. Specifically, the deformed vertex position is found by minimizing an energy function that consists of normal matching, Laplacian smoothing, and a regularization term:

$$\mathbf{v}^* = \arg\min_{\mathbf{v}} \sum_{\mathbf{v}\in V}\sum_{\mathbf{p}\in\mathbb{N}(\mathbf{v})} \|\mathbf{n}_{\mathbf{v}}\cdot\frac{\mathbf{v}-\mathbf{p}}{\|\mathbf{v}-\mathbf{p}\|}\|^2 + \lambda_l\sum_{\mathbf{v}\in V}\|\Delta\mathbf{v}\|^2 + \lambda_r\sum_{\mathbf{v}\in V}\|\mathbf{v}-\mathbf{v}_0\|^2 \quad (10)$$

This is applied to both DeepWrinkles and DDE, then the geometric quality of their deformed meshes are compared with our method. For DDE, we remove the material classifier, as our experiment is done on each clothing separately. Since we focus on static wrinkle prediction, we discard temporal information when training and testing DeepWrinkles and LFGD. We only use LFGD's high-frequency component, and discard its low-frequency component in this experiment.

We report the quantitative evaluation on the in-distribution test sets in Table 1. Because our method is generative, we run it 5 times with different Gaussian noise initialization and denoising trajectories, and report the average evaluation results. Our method achieves the lowest vertex distance and Chamfer distance on both sequences, which demonstrates the effectiveness of our method in generating realistic 3D wrinkles. Examples of qualitative results can be found in Figure 5 and Supp. Mat.. DeepWrinkles can generate normal maps in high quality, which is applicable to rendering, but it does not directly produce detailed clothing geometry. Following DDE, we apply Equation 10 to deform the base surface with the guidance of DeepWrinkles's normal map prediction. Since the normal map does not provide information on the geometric scale of the wrinkle, the deformed mesh may have misalignment with the ground-truth. In contrast, our method directly deforms the clothing geometry, where the surface normal can be derived trivially. DDE is good at enhancing existing wrinkles as shown in the original paper. In our experiment, each method starts from a smooth base mesh without wrinkles, and DDE fails to generate meaningful wrinkle patterns even as normal maps. LFGD's high-frequency component represents the wrinkle details as vertex displacement in the global coordinate frame, making it difficult to generalize to unseen coarse deformation.

Table 1: Comparison of 3D wrinkle reconstruction with baseline methods. E_v and E_{cd} are measured in $10^{-2}m$ and $10^{-4}m^2$, respectively. Ours is the average of 5 runs.

	T-shirt		*Skirt*	
	E_v	E_{cd}	E_v	E_{cd}
DeepWrinkles [26]	0.835	0.815	1.012	1.157
DDE [56]	1.507	3.405	1.457	3.552
LFGD [37]	0.586	0.391	0.941	1.225
Ours	**0.347**	**0.205**	**0.526**	**0.666**

Comparison of wrinkle quality. Besides DeepWrinkles, DDE, and LFGD, we compare our method with body-dependent clothing deformation methods HOOD [13], NCS [3], and PBNS [2] on the out-of-distribution test set. HOOD, NCS, PBNS are physics-inspired methods that can be trained in an unsupervised way. Instead of explicitly decomposing clothing deformation into coarse and fine components, they directly predict the final clothing geometry given body poses. Similar methods include TailorNet [38] and SNUG [45], which have been surpassed by state-of-the-art methods HOOD, NCS, and PBNS.

In this experiment, HOOD, NCS, PBNS, and our method take the same SMPL body as input, and generate wrinkled clothing, which may show different coarse deformation, especially for loose clothing. As a mid-product, our method

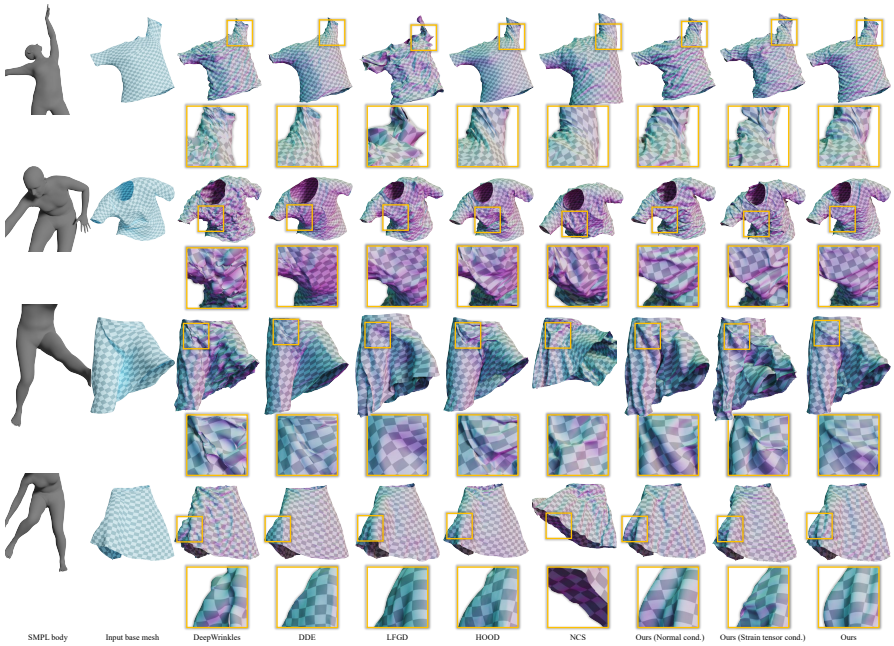


Fig. 6: Comparison of wrinkle quality with baseline methods. Given SMPL body as input, HOOD, NCS, PBNS and our method generate wrinkled clothing, which may show different coarse deformation. The base mesh is generated by our method as a mid-product, which is taken as input by DeepWrinkles, DDE, and LFGD for wrinkle generation. Normal conditioning and strain tensor conditioning are variants of our method for ablation study. See Supp. Mat. for more results.

generates base clothing mesh, which is used by DeepWrinkles, DDE, and LFGD as input for wrinkle generation. We report FID scores for this experiment in Table 2. Qualitative results in Figure 6 illustrate that our method generates realistic wrinkles comparing with other methods.

4.3 Ablation study

Using the norm of Green-Lagrange strain as the conditioning signal is crucial for our method to generate generalizable realistic wrinkles. Alternative choices for conditioning signals include the surface normal [26, 56] and the Green-Lagrange strain tensor. We do an ablation study to justify our choice of the norm of Green-Lagrange strain by comparing it with surface normal and Green-Lagrange strain tensor as conditioning signals. In order to compare the generalizability, all variants are trained on the same training set, and tested on the out-of-distribution test set. Quantitative evaluation in Table 2 and qualitative results in Figure 6 show that our method outperforms variants conditioned on surface

normal and Green-Lagrange strain tensor. When using surface normal as the conditioning signal, the generated wrinkles are not always realistic, *e.g.*, wrinkles in different directions may intersect with each other, which is not physically plausible. In contrast, using the norm Green-Lagrange strain as the conditioning signal leads to more plausible wrinkles, because it captures the stretching state of the base mesh and makes the model physics-aware. The Green-Lagrange strain tensor captures unnecessary directional information, which can lead to overfitting with limited training data. We take its norm to compress the feature and reduce learnable parameters, leading to better generalization.

Table 2: FID score on UV-space normal maps as a perceptual evaluation for wrinkle quality. Ours is the average of 5 runs.

	T-shirt	*Skirt*
Base mesh	212.08	131.29
DeepWrinkles [26]	231.40	92.46
DDE [56]	218.29	126.77
LFGD [37]	305.89	104.63
HOOD [13]	163.22	79.79
NCS [3]	142.87	90.50
PBNS [2]	143.58	129.87
Ours (Normal)	129.19	60.24
Ours (Strain tensor)	134.52	116.05
Ours	**75.05**	**45.09**

4.4 Applications

Wrinkle transfer to other clothing styles. As shown in Figure 1, our method can be applied to other clothing styles with compatible UV parameterization. The high-frequency wrinkle deformation is learned from real captures of a short-sleeve T-shirt, which can be successfully applied to long-sleeve T-shirts and tank tops. Although the learned wrinkling is only for the short-sleeve T-shirt, and may not faithfully reflect how actual long-sleeve T-shirts and tank tops deform, it enables artistic applications like wrinkle style transfer.

Modeling fabric material variations. By changing Equation 6 and Equation 7 following classifier-free guidance [19], we can train a material-aware diffusion model for strain-conditioned wrinkle generation. Once trained, our method can generate slightly different wrinkling by altering the fabric material label for the same base mesh as illustrated in Figure 7 (a). To enable the training, we combine the two sequences T-shirt on "subject_00" and T-shirt on "subject_04" from the patterned clothing dataset [16], and align their UV parameterization.

Image-based wrinkle fitting. Our method can be applied to down-stream tasks like image-based wrinkle fitting, where the 3D clothing can be fitted to a 2D image by matching the surface normal. This is possible by guiding [9] the wrinkle generation from our wrinkle diffusion model. In Figure 7 (b), we demonstrate the wrinkle fitting results where we first align the coarse mesh with the input image based on the body pose prediction [53]; and guide the denoising process by comparing the differentiably rendered surface normal from our model with the normal prediction [22]. Examples are from the TikTok [22] and HUMBI [51] datasets.

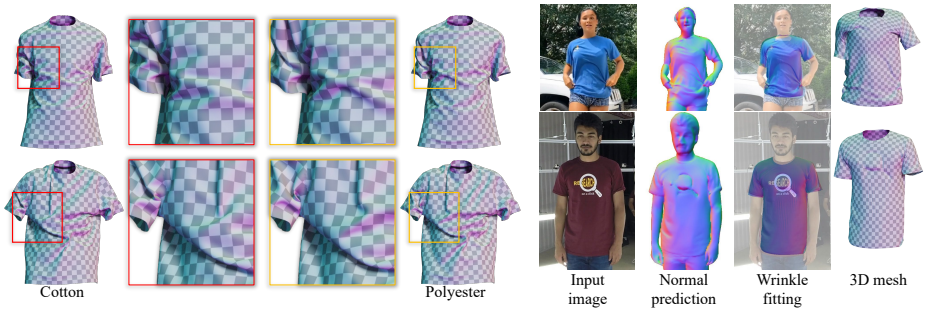


Fig. 7: (Left) Material-aware wrinkle generation. The strain-conditioned diffusion model is trained on a dataset of two T-shirts made with cotton and polyester, respectively. By specifying the material label for the same base mesh, our method can generate slightly different wrinkling to reflect different fabric property. (Right) Image-based wrinkle fitting. The coarse mesh is first aligned with the input image based on the body pose prediction, then the denoising process for wrinkle generation is guided by matching the differentiably rendered surface normal to the normal prediction.

5 Conclusion

We have presented a generalizable method for generating high-fidelity clothing wrinkle using strain-conditioned diffusion model. It learns the correlation between the strain and the wrinkling from real clothing captures. The strain measurement captures local rotation-invariant geometric information independently of body and clothing topology, enhancing the generalizability of our method.

Based on the strain-conditioned diffusion model for wrinkle generation, we have also presented a practical pipeline to model realistic body-dependent 3D clothing deformation for virtual avatars. Specifically, a virtual skeleton controls the coarse clothing deformation, while a CVAE captures the body-dependent latent space of this deformation. Since the coarse deformation contributes little to the realism of clothing deformation, it can be learned from low-quality synthetic body-clothing pairwise data, not to be bounded by the limited diversity of body shape and pose variations in real capture.

Limitations and future work. Our method only models static clothing deformation, and ignores dynamics. Since dynamics is an important factor that contributes to the realism of clothing deformation, future work can extend our method to model dynamic clothing deformation. Our method does not incorporate physics-based constraints, so physical plausibility is not guaranteed, e.g., the body-clothing collision and clothing self-collision may exist in the result. Although this can be fixed by post-processing, it would be interesting to incorporate physics heuristics into our method. We use LBS for simplicity to deform the base mesh based on virtual skeletons, resulting in our coarse deformation inheriting the artifacts of LBS. Future work can explore more advanced methods to improve the realism of coarse clothing deformation.

References

1. Aharoni, H., Todorova, D.V., Albarrán, O., Goehring, L., Kamien, R.D., Katifori, E.: The smectic order of wrinkles. Nature communications p. 15809 (2017)
2. Bertiche, H., Madadi, M., Escalera, S.: Pbns: Physically based neural simulation for unsupervised garment pose space deformation. ACM TOG pp. 1–14 (2021)
3. Bertiche, H., Madadi, M., Escalera, S.: Neural cloth simulation. ACM TOG (2022)
4. Bhatnagar, B.L., Sminchisescu, C., Theobalt, C., Pons-Moll, G.: Loopreg: Self-supervised learning of implicit surface correspondences, pose and shape for 3d human mesh registration. In: NeurIPS. pp. 12909–12922 (2020)
5. Bouaziz, S., Martin, S., Liu, T., Kavan, L., Pauly, M.: Projective dynamics: Fusing constraint projections for fast simulation. ACM TOG (2014)
6. Chen, L., Gao, L., Yang, J., Xu, S., Ye, J., Zhang, X., Lai, Y.K.: Deep deformation detail synthesis for thin shell models. In: Computer Graphics Forum (2023)
7. Chen, Z., Chen, H.Y., Kaufman, D.M., Skouras, M., Vouga, E.: Fine wrinkling on coarsely meshed thin shells. ACM TOG pp. 1–32 (2021)
8. Chen, Z., Kaufman, D., Skouras, M., Vouga, E.: Complex wrinkle field evolution. ACM TOG pp. 1–19 (2023)
9. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. In: ICLR (2022)
10. Du, T., Wu, K., Ma, P., Wah, S., Spielberg, A., Rus, D., Matusik, W.: Diffpd: Differentiable projective dynamics. ACM TOG pp. 1–21 (2021)
11. Feng, X., Huang, W., Xu, W., Wang, H.: Learning-based bending stiffness parameter estimation by a drape tester. ACM TOG pp. 1–16 (2022)
12. Furukawa, Y., Ponce, J.: Dense 3d motion capture from synchronized video streams. In: CVPR. pp. 1–8 (2008)
13. Grigorev, A., Black, M.J., Hilliges, O.: Hood: Hierarchical graphs for generalized modelling of clothing dynamics. In: CVPR. pp. 16965–16974 (2023)
14. Guo, J., Li, J., Narain, R., Park, H.S.: Inverse simulation: Reconstructing dynamic geometry of clothed humans via optimal control. In: CVPR. pp. 14698–14707 (2021)
15. Guo, J., Prada, F., Xiang, D., Romero, J., Wu, C., Park, H.S., Shiratori, T., Saito, S.: Diffusion shape prior for wrinkle-accurate cloth registration. arXiv preprint arXiv:2311.05828 (2023)
16. Halimi, O., Stuyck, T., Xiang, D., Bagautdinov, T., Wen, H., Kimmel, R., Shiratori, T., Wu, C., Sheikh, Y., Prada, F.: Pattern-based cloth registration and sparse-view animation. ACM TOG pp. 1–17 (2022)
17. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: NeurIPS (2017)
18. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS. pp. 6840–6851 (2020)
19. Ho, J., Salimans, T.: Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598 (2022)
20. Hu, Y., Anderson, L., Li, T.M., Sun, Q., Carr, N., Ragan-Kelley, J., Durand, F.: Difftaichi: Differentiable programming for physical simulation. In: ICLR (2020)
21. Işık, M., Rünz, M., Georgopoulos, M., Khakhulin, T., Starck, J., Agapito, L., Nießner, M.: Humanrf: High-fidelity neural radiance fields for humans in motion. ACM TOG pp. 1–12 (2023)

22. Jafarian, Y., Park, H.S.: Learning high fidelity depths of dressed humans by watching social media dance videos. In: CVPR. pp. 12753–12762 (2021)
23. Kavan, L., Gerszewski, D., Bargteil, A.W., Sloan, P.P.: Physics-inspired upsampling for cloth simulation in games. ACM TOG pp. 1–10 (2011)
24. Kim, B., Kwon, P., Lee, K., Lee, M., Han, S., Kim, D., Joo, H.: Chupa: Carving 3d clothed humans from skinned shape priors using 2d diffusion probabilistic models. arXiv preprint arXiv:2305.11870 (2023)
25. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: ICLR (2014)
26. Lahner, Z., Cremers, D., Tung, T.: Deepwrinkles: Accurate and realistic clothing modeling. In: ECCV. pp. 667–684 (2018)
27. Le, B.H., Deng, Z.: Smooth skinning decomposition with rigid bones. ACM TOG pp. 1–10 (2012)
28. Li, J., Daviet, G., Narain, R., Bertails-Descoubes, F., Overby, M., Brown, G.E., Boissieux, L.: An implicit frictional contact solver for adaptive cloth simulation. ACM TOG pp. 1–15 (2018)
29. Li, Y., Du, T., Wu, K., Xu, J., Matusik, W.: Diffcloth: Differentiable cloth simulation with dry frictional contact. ACM TOG pp. 1–20 (2022)
30. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: Smpl: A skinned multi-person linear model. ACM TOG pp. 1–16 (2015)
31. 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: CVPR. pp. 6469–6478 (2020)
32. Mahmood, N., Ghorbani, N., Troje, N.F., Pons-Moll, G., Black, M.J.: Amass: Archive of motion capture as surface shapes. In: ICCV. pp. 5442–5451 (2019)
33. Müller, M., Chentanez, N.: Wrinkle meshes. In: Symposium on Computer Animation. pp. 85–91 (2010)
34. Müller, M., Heidelberger, B., Hennix, M., Ratcliff, J.: Position based dynamics. Journal of Visual Communication and Image Representation pp. 109–118 (2007)
35. Narain, R., Pfaff, T., O'Brien, J.F.: Folding and crumpling adaptive sheets. ACM TOG pp. 1–8 (2013)
36. Narain, R., Samii, A., O'brien, J.F.: Adaptive anisotropic remeshing for cloth simulation. ACM TOG pp. 1–10 (2012)
37. Pan, X., Mai, J., Jiang, X., Tang, D., Li, J., Shao, T., Zhou, K., Jin, X., Manocha, D.: Predicting loose-fitting garment deformations using bone-driven motion networks. In: ACM SIGGRAPH 2022 Conference Proceedings. pp. 1–10 (2022)
38. Patel, C., Liao, Z., Pons-Moll, G.: Tailornet: Predicting clothing in 3d as a function of human pose, shape and garment style. In: CVPR. pp. 7365–7375 (2020)
39. Peng, S., Zhang, Y., Xu, Y., Wang, Q., Shuai, Q., Bao, H., Zhou, X.: Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In: CVPR (2021)
40. Pons-Moll, G., Pujades, S., Hu, S., Black, M.J.: Clothcap: Seamless 4d clothing capture and retargeting. ACM TOG pp. 1–15 (2017)
41. Qiao, Y.L., Liang, J., Koltun, V., Lin, M.C.: Scalable differentiable physics for learning and control. In: ICML. pp. 7847–7856 (2020)
42. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015)
43. Saito, S., Yang, J., Ma, Q., Black, M.J.: Scanimate: Weakly supervised learning of skinned clothed avatar networks. In: CVPR. pp. 2886–2897 (2021)
44. Santesteban, I., Otaduy, M.A., Casas, D.: Learning-based animation of clothing for virtual try-on. In: Comput. Graph. Forum. pp. 355–366 (2019)
45. Santesteban, I., Otaduy, M.A., Casas, D.: Snug: Self-supervised neural dynamic garments. In: CVPR. pp. 8140–8150 (2022)

46. Shao, R., Zheng, Z., Zhang, H., Sun, J., Liu, Y.: Diffustereo: High quality human reconstruction via diffusion-based stereo using sparse cameras. In: ECCV. pp. 702–720 (2022)
47. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. In: NeurIPS (2015)
48. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: ICLR (2021)
49. Stein, O., Jacobson, A., Wardetzky, M., Grinspun, E.: A smoothness energy without boundary distortion for curved surfaces. ACM TOG pp. 1–17 (2020)
50. Wang, H., O'Brien, J.F., Ramamoorthi, R.: Data-driven elastic models for cloth: modeling and measurement. ACM TOG pp. 1–12 (2011)
51. Yoon, J.S., Yu, Z., Park, J., Park, H.S.: Humbi: A large multiview dataset of human body expressions and benchmark challenge. IEEE TPAMI pp. 623–640 (2021)
52. Zhang, C., Pujades, S., Black, M.J., Pons-Moll, G.: Detailed, accurate, human shape estimation from clothed 3d scan sequences. In: CVPR. pp. 4191–4200 (2017)
53. Zhang, H., Tian, Y., Zhou, X., Ouyang, W., Liu, Y., Wang, L., Sun, Z.: Pymaf: 3d human pose and shape regression with pyramidal mesh alignment feedback loop. In: ICCV. pp. 11446–11456 (2021)
54. Zhang, J.E., Dumas, J., Fei, Y., Jacobson, A., James, D.L., Kaufman, D.M.: Progressive simulation for cloth quasistatics. ACM TOG pp. 1–16 (2022)
55. Zhang, M., Ceylan, D., Mitra, N.J.: Motion guided deep dynamic 3d garments. ACM TOG pp. 1–12 (2022)
56. Zhang, M., Wang, T., Ceylan, D., Mitra, N.J.: Deep detail enhancement for any garment. In: Computer Graphics Forum. pp. 399–411 (2021)
57. Zhong, Y.D., Han, J., Brikis, G.O.: Differentiable physics simulations with contacts: Do they have correct gradients wrt position, velocity and control? arXiv preprint arXiv:2207.05060 (2022)