

DiffAvatar: Simulation-Ready Garment Optimization with Differentiable Simulation

Yifei Li*² Hsiao-yu Chen¹ Egor Larionov¹ Nikolaos Sarafianos¹ Wojciech Matusik² Tuur Stuyck¹

¹Meta Reality Labs ²MIT CSAIL

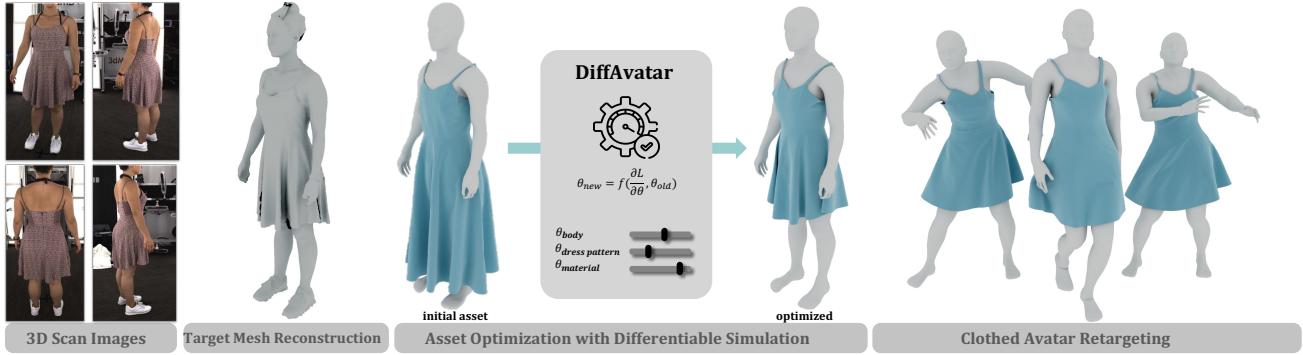


Figure 1. We present **DiffAvatar**, an automated computational method to recover simulation-ready garment and body assets. Starting from a multi-view capture, we reconstruct a semantically segmented 3D mesh. The segmented clothing geometry acts as a target shape for our optimization pipeline. Our method recovers body shape and pose, clothing pattern and clothing material parameters from a single scan. We optimize a clothing template in 2D pattern space to reproduce the captured clothing in 3D in a physical way. We compute gradients of required parameters using a differentiable simulation approach.

Abstract

The realism of digital avatars is crucial in enabling telepresence applications with self-expression and customization. While physical simulations can produce realistic motions for clothed humans, they require high-quality garment assets with associated physical parameters for cloth simulations. However, manually creating these assets and calibrating their parameters is labor-intensive and requires specialized expertise. Current methods focus on reconstructing geometry, but don't generate complete assets for physics-based applications. To address this gap, we propose **DiffAvatar**, a novel approach that performs body and garment co-optimization using differentiable simulation. By integrating physical simulation into the optimization loop and accounting for the complex nonlinear behavior of cloth and its intricate interaction with the body, our framework recovers body and garment geometry and extracts important material parameters in a physically plausible way. Our experiments demonstrate that our approach generates realistic clothing and body shape suitable for downstream applications. We provide additional insights and results on our webpage: people.csail.mit.edu/liyifei/publication/diffavatar.

1. Introduction

Virtual avatars are increasingly gaining importance as they serve as a digital extensions of users, enabling novel social and professional interactions. The physical realism of avatars, including realistic clothing and accurate body shape, is crucial for such applications. This need for realism extends beyond visual aesthetics but also includes dynamic interactions and motion obtained by accurate physical simulation of clothing and body dynamics. Physical simulation and rendering techniques can be used as tools to achieve physical realism in the virtual world. However, this requires the creation of high-quality clothing assets for individual users, which presents a substantial challenge. The conventional approach requires meticulous manual design by artists, a process that is exceedingly time-consuming. This manual approach is fundamentally unfeasible for individualized avatar clothing, especially considering the continuously growing user base of telepresence applications. The notion of having an artist create a unique virtual outfit for every user is simply impractical. This scenario underscores the pressing need for automated solutions for scalable and personalized avatar asset creation and optimization. Recent advancements in computer vision and graphics have accelerated the automation of avatar asset creation from user images or scans. However, the predominant focus has been on geometry reconstruction, with limited focus on generating

*This work was conducted during an internship at Meta Reality Labs

complete assets that can be used in physics-based applications.

DiffAvatar endeavors to bridge this gap by introducing a body and garment co-optimization pipeline using differentiable simulation. By entwining physical simulation within the optimization loop, we ensure that the dynamics of the clothing are considered in the optimization process. We optimize for all assets required for physics-based simulation and other downstream applications in a physically plausible way by leveraging differentiable cloth simulation for body shape recovery and extending it to optimize for garment shape directly in the rest shape pattern space. Specifically, we recover garment patterns, body pose and shape, as well as retrieving the crucial physical material parameters leveraging only a minimal garment template library. We believe that our work is the first to leverage high-resolution differentiable simulation for asset recovery from real scans which often contain holes and compromised boundaries. In summary, our key contributions are as follows:

- A novel approach that utilizes differentiable simulation for co-optimizing garment shape and materials, and body shape and pose, while taking into account cloth deformations and collisions in the context of avatar asset recovery.
- A unified method for body shape, pose and garment assets recovery from one *real noisy 3D scan of a clothed person*.
- For the first time in a differentiable cloth simulation algorithm, we incorporate optimization through the cloth rest shape. Additionally, we develop a differentiable control cage representation for garment shape optimization to regularize the 2D garment pattern space and produce effective optimization results.

2. Related Work

Pose and Shape Estimation precedes the garment shape estimation and its properties since the underlying body directly impacts how cloth drapes and behaves when in motion. Prior works focus on reconstructing the body from minimally clothed [2, 36] captures or focus on complex captures of clothed humans [3, 10, 33, 34, 65] to extract a representation for body and garments. These methods build on SMPL [36], and work in conjunction with simulated [3, 33, 34, 65], or trained [10] garment models.

Cloth Simulation methods have been widely used for digitally modeling the behavior of fabrics in visual effects and movie productions since the pioneering work on implicit cloth simulation [4] and follow-up works [12, 17, 20] allowing for stable and efficient simulations. Position Based Dynamics (PBD) [41] updates positions directly to project constraints in a highly parallel fashion resulting in high performance simulations. eXtended Position Based Dynamics (XPBD) [38] overcome the limitation of iteration dependent behavior of PBD while Projective Dynamics [7] connects

nodal Finite Element methods and Position Based methods, leading to an efficient and accurate solver. Stuyck [52] provides an overview on cloth simulation techniques.

Garment Pattern Estimation is essential as the 2D sewing pattern influences the fit and the formation of wrinkles on the 3D body. One approach involves flattening the 3D shape into several developable [51] 2D pieces. However, these methods require manual cutting input to generate pieces with minimal distortion [3, 42]. Other works use neural networks to learn the seams [18], but the patterns obtained through direct flattening of the 3D shape are only suitable for nearly undeformed cloth, which is rarely the case in real-world draped garments. Follow-up works employed neural networks to learn the 2D rest shape and yield more accurate patterns, but their generality is limited to the training data and specific garments [11, 65]. Parameterized garment patterns [25, 26] address these limitations and can adapt to a wide range of shapes but struggle to generalize to real garments and lack control over symmetry and matching seam lines. Alternatively 2D patterns can be optimized using a physics simulator in an iterative manner [5, 56, 62].

Differentiable Simulation allows for gradient computation with respect to simulation parameters, enabling the use of gradient-based optimization algorithms to find solutions for inverse design and system identification. Early works applied the adjoint method to fluid [31, 40] and cloth [61] simulation models to analytically compute gradients. Recent techniques differentiate through complex simulations such as Projective Dynamics [14] and XPBD [53]. Differentiable simulation methods have successfully been applied to cloth simulation with frictional contact [32], material estimation [9, 27] and shape and pose estimation [21].

Learning-based approaches have focused on garment draping, modeling of cloth dynamics, or handling collisions and contact [6, 8, 19, 22, 24, 28, 30, 35, 45, 46, 48, 49, 54, 55, 63, 64] and hair [59, 60] and the introduction of new large-scale datasets [67] albeit synthetic, will further accelerate this progress. DrapeNet [13] predicts a 3D deformation field conditioned on the latent codes of a generative network, which models garments as unsigned distance fields allowing it to handle and edit unseen clothes. Qiu *et al.* [44] reconstruct 3D clothes from monocular videos using SDFs and deformation fields. Qi *et al.* [43] proposed a personalized 2D pattern design method using synthetic data, where the user can input specific constraints for personal 2D pattern design from 3D point clouds. Li *et al.* [29] proposed a parametric garment representation model for garment draping using SDFs.

3. Preliminaries

3.1. Body and Garment Models

Body Shape and Pose are represented using a parameterized statistical body model similar to SMPL [36] in our

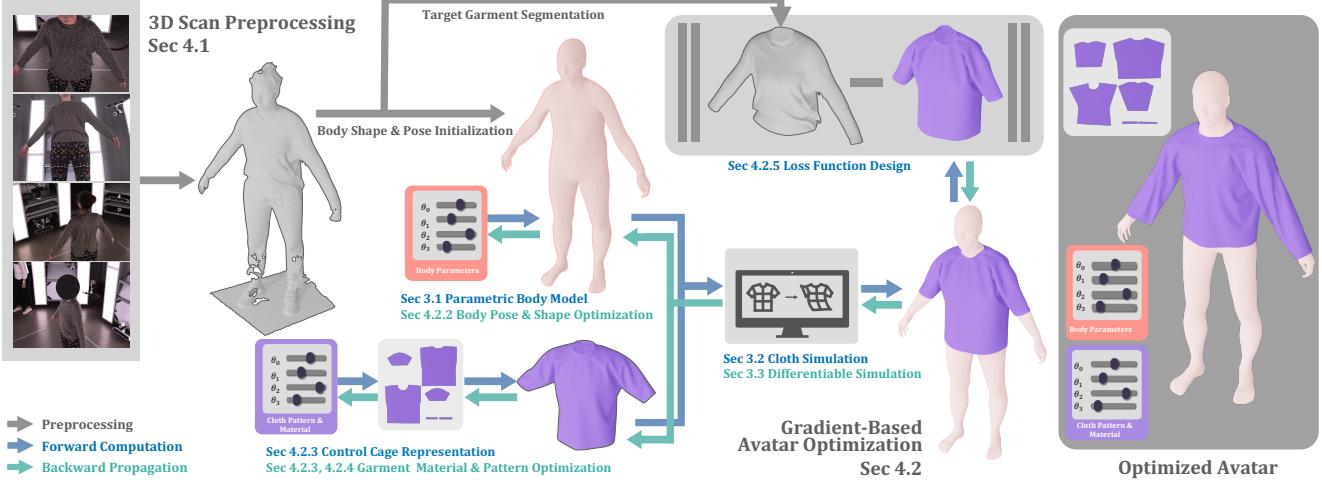


Figure 2. **DiffAvatar** generates simulation-ready avatar assets from inputs obtained through a multi-view capture. Our pipeline initially preprocesses the 3D scan to segment the target garment and establish the initial pose and shape of the parametric body model. We employ a differentiable simulation framework to align our simulated garment with the segmented garment by jointly optimizing the garment’s design and material parameters, along with the body shape.

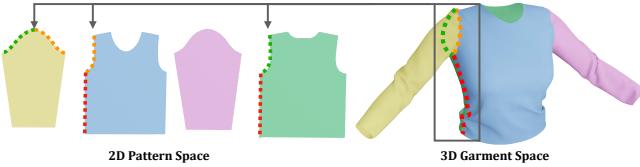


Figure 3. 3D garments (right) can be compacted represented as their 2D panels (left). Seams are visualized as dotted-lines.

method. The skeleton is defined by joints which are described by P parameters encoding local transformations through joint angles ψ and bone lengths. The shape is encoded by the statistical shape coefficients ν as $\mathcal{V}_0 + \nu\mathcal{V}$ where \mathcal{V}_0 and \mathcal{V} encode the average body shape and the shape basis functions respectively. The body shape with V_b vertices is posed with the skeleton using a linear blend skinning function $\mathcal{S} : \mathbb{R}^{3 \times V_b} \times \mathbb{R}^P \rightarrow \mathbb{R}^{3 \times V_b}$ [39].

Garments can take on a wide range of 3D shapes when draped onto a body, due to factors such as changing pose and dynamics or wearer manipulations. Despite this large variation in configurations, garments are compactly represented by their 2D patterns (Fig. 3), which consist of the individual pieces of fabric that are sewn together to create the 3D clothing. Therefore, we represent clothing in 2D pattern space, which ensures developable [51] meshes and manufacturable clothing. Virtual garments are modeled as triangle meshes, with their rest shape encoded in these 2D patterns. The rest shape is crucial for modeling the in-plane stretching and shearing behavior of different fabrics.

3.2. Cloth Simulation

We compute the deformation of a garment mesh consisting of V vertices which is draped on a posed body us-

ing dynamic physics-based simulation. The simulator effectively solves Newton’s equations of motion given by $\mathbf{M}\dot{\mathbf{v}} = -\nabla U(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^{3V}$ and $\mathbf{v} \in \mathbb{R}^{3V}$ are the vertex positions and velocities, $U(\mathbf{x})$ is the energy potential and \mathbf{M} is the mass matrix. The simulator advances the garment state $\mathbf{q}_n = (\mathbf{x}_n, \mathbf{v}_n)$ at time step n forward in time at discrete time steps Δt . \mathbf{Q} consists of states over all time steps N . Although any simulation model can be used, in this work, we make use of XPBD [38] due to its excellent performance characteristics. The energy potential $U(\mathbf{x})$ is formulated in terms of a vector of all constraint functions $\mathbf{C}(\mathbf{x})$ and an inverse compliance matrix $\boldsymbol{\alpha}^{-1}$ as $U(\mathbf{x}) = \frac{1}{2}\mathbf{C}(\mathbf{x})^\top \boldsymbol{\alpha}^{-1} \mathbf{C}(\mathbf{x})$. The constraints include triangle constraints, dihedral bending and collision constraints, modelling in-plane stretching and shearing, out-of-plane bending and collisions respectively. At each time step, a position update $\Delta\mathbf{x}$ is computed using a Gauss-Seidel-like iterative solver indexed by i of the following system:

$$(\nabla \mathbf{C}(\mathbf{x}_i)^\top \mathbf{M}^{-1} \nabla \mathbf{C}(\mathbf{x}_i) + \tilde{\boldsymbol{\alpha}}) \Delta \boldsymbol{\lambda} = -\mathbf{C}(\mathbf{x}_i) - \tilde{\boldsymbol{\alpha}} \boldsymbol{\lambda}_i \\ \Delta \mathbf{x} = \mathbf{M}^{-1} \nabla \mathbf{C}(\mathbf{x}_i) \Delta \boldsymbol{\lambda}, \quad (1)$$

where $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}/\Delta t^2$ and $\boldsymbol{\lambda}$ is the constraint multiplier. Due to the decoupled nature of the solve, the position update $\Delta\mathbf{x}$ can be computed separately for each constraint type. We compute vertex positions as

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \Delta\mathbf{x} + \Delta t (\mathbf{v}_n + \Delta t \mathbf{M}^{-1} \mathbf{f}_{\text{ext}}) \quad (2)$$

and velocities $\mathbf{v}_{n+1} = \frac{1}{\Delta t} (\mathbf{x}_{n+1} - \mathbf{x}_n)$ where \mathbf{f}_{ext} denote the external forces acting on the system.

3.3. Differentiable Cloth Simulation

Given a minimizing goal function ϕ computed through complex dynamic simulations, differentiable simulation enables gradient-based optimization methods by computing its gradient ϕ with respect to the control parameters θ as

$$\frac{d\phi}{d\theta} = \frac{\partial\phi}{\partial\mathbf{Q}} \frac{d\mathbf{Q}}{d\theta} + \frac{\partial\phi}{\partial\theta} \quad (3)$$

However, due to the intractability of computing $d\mathbf{Q}/d\theta$ directly, adjoint method is used to replace the vector-matrix product with an equivalent, more efficient computation involving the adjoint of \mathbf{Q} , denoted by $\hat{\mathbf{Q}}$ which contains all adjoint states $\hat{\mathbf{q}}_n = (\hat{\mathbf{x}}_n \in \mathbb{R}^{3V}, \hat{\mathbf{v}}_n \in \mathbb{R}^{3V})$ over all N steps. We use prior work DiffXPBD [53] to compute gradients through the XPBD simulation model. Using the adjoint states, the full derivative $d\phi/d\theta$ is obtained using

$$\frac{d\phi}{d\theta} = \hat{\mathbf{Q}}^\top \frac{\partial\Delta\mathbf{x}}{\partial\theta} + \frac{\partial\phi}{\partial\theta} \quad (4)$$

where $\Delta\mathbf{x}$ refers to the position updates computed in the XPBD framework in Eq. 1. We refer the readers to [53] for detailed derivations of the adjoint states $\hat{\mathbf{Q}}$. The quantities $\partial\Delta\mathbf{x}/\partial\theta$ and $\partial\phi/\partial\theta$ are problem-specific, which we detail in the next sections.

4. Methodology

We introduce our computational method for extracting garment and body assets from real 3D scans of clothed humans. Our method uses a differentiable simulator for simultaneous co-optimization of garment 2D pattern shape, cloth material, body pose and shape. See Fig. 2 for a visual overview. Starting from an automatically selected template, our goal is to optimize garment patterns and materials that replicate the overall style and fit of the scan. Note that the drape of a given garment, including the wrinkles and surface details, can be different on different body shape and state, and may be adjusted by the wearer, therefore we do not aim to perfectly recreate the garment shapes exactly as they appear in the scan. Additionally, we aim to recover the overall body shape and pose but do not intend to recover other appearance aspects such as the face, since it does not influence the simulated behavior of the clothing.

4.1. Extracting 3D Garments and Parametric Body

We process multi-view images to reconstruct and segment a 3D scan and use the resulting geometry to initialize the shape and pose of the parametric body model.

3D Scan Semantic Garment Segmentation From multi-view images of a clothed person, we reconstruct a noisy 3D scan using the 3dMD system [1]. These scans tend to be noisy, contain holes and might not capture regions such as

hair or loose clothes accurately. We extract the 3D geometry of the isolated garment(s) of interest using a cloth segmentation algorithm [16] on each of the 18 camera views to obtain per-pixel class predictions. We enforce multi-view class consistency by selecting the majority garment classes.

Body Shape and Pose Initialization To fit our parametric body model to the scan, we optimize the body shape ν , pose ψ and joint lengths by minimizing the Chamfer distance between the vertices and those of the full person scan. We use a Gauss-Newton solver that takes joint limits into account and penalizes self-penetrations of the body mesh.

4.2. Avatar Optimization

We leverage differentiable simulation (Sec. 3.3) to simultaneously recover garment pattern and material, as well as body pose and shape. Starting from an initial pattern, our method automatically adjusts the size and shape of each panel in the pattern. To achieve this, we require a minimal garment library that defines the pattern structure for each type of garment. We use the semantic information (Sec. 4.1) to automatically identify the garment types. With the estimated body shape and pose, we drape the garment through physical simulation to obtain the initial 3D garment state.

4.2.1 Optimization Problem Statement

Once initialized, we aim to find the parameters θ that minimize a loss function $\phi(\theta, \mathbf{Q})$. The loss function (Sec. 4.2.5) encodes how close the geometry is to the segmented scan. The control variables θ include statistical body shape ν and pose ψ coefficients to model the body shape under the clothing, material parameters λ to model the fabric properties and most importantly, the control cage handles ζ that deform the 2D pattern space coordinates \mathbf{p} of the garment (Sec. 4.2.2).

We use gradient descent to optimize these variables over multiple iterations. In each iteration, we run a dynamic differentiable simulation until the garments reach quasi-equilibrium state with the current set of parameters θ to obtain a draped garment. This draped garment is then used to compute a loss, and the gradient information $d\phi/d\theta$ is obtained through back-propagation using the differentiable simulator. We determine the full gradient by computing the jacobian $\partial\Delta\mathbf{x}/\partial\theta$ and $\partial\phi/\partial\theta$ to evaluate Eq. 4. In the following subsections, we explain how to compute gradients with respect to θ . Note that our pipeline is not limited to the specific implementation of DiffXPBD and can be applied to any differentiable cloth simulation framework.

4.2.2 Garment Pattern Optimization

We propose a regularized differentiable cage formulation to effectively and robustly optimize for the 2D patterns of gar-

ments such that the simulated and draped 3D representation of the garment closely aligns with the scan.

Control Cage Pattern Representation: The 3D positions of each garment is controlled by its corresponding 2D pattern (Fig. 3). While it is possible to directly optimize for the 2D pattern vertices \mathbf{p} directly, this approach is highly non-regularized and can produce ill-shaped or even non-physical inverted rest shape geometries that cause simulators to fail. A high number of optimization variables can also cause the optimization to get stuck in a local minimum (See our ablation study in Sec. 5.4). Additionally, directly optimizing for the 2D coordinates does not respect design constraints that are better represented in a limited subspace of reasonable designs. Therefore, we further regulate the optimization problem by selecting and optimizing a set of 2D control vertices ζ on the boundaries of the individual panels of the 2D pattern that directly deform and manipulate the underlying 2D patterns through control cages instead.



Control Cage Handle Selection: We use the geometric information of the 2D garment patterns to automatically identify control cage points, see the inset figure above. Our algorithm first extracts the boundary loop of the underlying mesh for each connected component representing a garment panel in the 2D garment pattern, then processes the boundary loop and marks a vertex as a control point if it lies on the convex hull of the pattern or when its local curvature exceeds a threshold (10° in our implementation).

Differentiable Control Cage Optimization: We use the control handles to deform the underlying 2D pattern via Mean Value Coordinates [23]. During initialization, we compute a generalized barycentric coordinate for each vertex in the 2D pattern with respect to each vertex on the control cage, expressed as $\bar{\mathbf{x}} = \mathbf{W}\zeta$. We compute the required derivatives to evaluate Eq. 4 following the chain rule as:

$$\frac{\partial \Delta \mathbf{x}}{\partial \zeta} = \frac{\partial \Delta \mathbf{x}}{\partial \bar{\mathbf{x}}} \frac{\partial \bar{\mathbf{x}}}{\partial \zeta} = \frac{\partial \Delta \mathbf{x}}{\partial \bar{\mathbf{x}}} \mathbf{W} \quad (5)$$

We detail the derivation for $\frac{\partial \Delta \mathbf{x}}{\partial \bar{\mathbf{x}}}$ in the Supplementary Material.

4.2.3 Body Shape and Pose Optimization

The initial body shape obtained from the geometry-based optimization (Sec. 4.1) only uses the geometric information from the scan. We improve accuracy for the body shape and pose through differentiable simulation to explicitly account for the separate cloth geometry layer on top of the body. The body interacts with the garments during simulation solely through the collisions. Therefore we only need to compute the derivatives for the collision response updates

$\partial \Delta \mathbf{x}_{\text{cloth-body collision}}$ to evaluate Eq. 4. Using chain rule, we compute

$$\frac{\partial \Delta \mathbf{x}_{\text{cloth-body collision}}}{\partial \alpha} = \frac{\partial \Delta \mathbf{x}_{\text{cloth-body collision}}}{\partial \mathbf{x}_{\text{body}}} \frac{\partial \mathbf{x}_{\text{body}}}{\partial \alpha} \quad (6)$$

where α is either body shape ν or pose ψ . The first term $\partial \Delta \mathbf{x}_{\text{cloth-body collision}} / \partial \mathbf{x}_{\text{body}}$ measures how the cloth position updates changes with a change in position update of the body vertices. For the body shape parameters, the final term is computed by back-propagating the gradients through the body shape model described in Sec. 3.1. To obtain joint angle gradients, we differentiate through the linear blend skinning operation.

4.2.4 Material Property Estimation

We optimize for cloth material properties to better match the shape of the scanned garments. See the supp. material for details. The bending parameter has the most significant effect [15] on the wrinkling of the cloth and can be inferred from draped garments. Since the bending material parameter λ only enters the computational graph when computing the dihedral constraint, computing the jacobian reduces to $\partial \Delta \mathbf{x} / \partial \lambda = \partial \Delta \mathbf{x}_{\text{Dihedral}} / \partial \lambda$.

4.2.5 Loss Function Design

Our loss function is designed with two main components: the *feature matching* term $\mathcal{L}_{\text{features}}$ and the *regularization* term $\mathcal{L}_{\text{regularization}}$. The feature matching term encourages the optimization to converge to the scan in the 3D world coordinates after simulations, while the regularization terms operate on the 2D patterns to maintain desired features. The loss function is thus given by $\phi = \mathcal{L}_{\text{features}} + \mathcal{L}_{\text{regularization}}$.

Feature Matching is used to ensure that the simulated garment matches the scan. We use two distinct terms with individual weights ρ and σ to achieve this goal. The *boundary loss* term measures how well the boundaries overlap and serves to drive correct lengths of the pattern to match size, whereas the *interior loss* is designed to match the looseness of the fit: $\mathcal{L}_{\text{features}} = \rho \mathcal{L}_{\text{boundary}} + \sigma \mathcal{L}_{\text{interior}}$.

Boundary Feature Matching We segment boundary points on the scan and align and match with those of the simulated garment and minimize the L2 distance to boundaries such as sleeve lengths and hems.

Interior Point Feature Matching We measure and minimize the Chamfer distance between the interior points of the simulated garment and target segmented garment scan.

Regularization To improve our loss formulation, we include regularization terms that act on the pattern space to maintain desired features in the design. We enable the optimizer to adapt individual patterns in the garment design to

match features in the 3D scan. However, this process could result in designs where seams that are to be sewn together have different edge lengths, leading to undesired artifacts such as gathering, which produces a ruffled effect. To prevent this, we add two regularizers with weights α and β , giving $\mathcal{L}_{\text{regularization}} = \alpha \mathcal{L}_{\text{seam length}} + \beta \mathcal{L}_{\text{curvature}}$. We penalize seam length differences for edges on the individual patterns that are to be sewn together, as shown in Fig. 3. The color coded seams that are to be sewn together should have the same length. We color-coded a subset of the seams since the right half is symmetric. Mathematically, we express this as follows

$$\mathcal{L}_{\text{seam length}} = \sum_{i \in \text{Seam edges}} \|\mathbf{p}_i - \mathbf{p}_{i+1}\|^2 - \|\mathbf{p}'_i - \mathbf{p}'_{i+1}\|^2 \quad (7)$$

Additionally, to prevent noisy and undesired designs, we penalize the changes in boundary curvature of the 2D pattern with respect to the original garment template similar to the work of Wang [56]. We seek a scaled rotation matrix $\mathbf{T}_i = s\mathbf{R}_i \in \mathbb{R}^{2 \times 2}$ at each point \mathbf{p}_i with least curvature distortion to its connected boundary edges, $\mathbf{T}_i = \arg \min_{\mathbf{T}} \|\mathbf{e}_{i1} - \mathbf{T}\bar{\mathbf{e}}_{i1}\|^2 + \|\mathbf{e}_{i2} - \mathbf{T}\bar{\mathbf{e}}_{i2}\|^2$ with $\mathbf{e}_{i1} = \mathbf{p}_{i+1} - \mathbf{p}_i$ and $\mathbf{e}_{i2} = \mathbf{p}_{i-1} - \mathbf{p}_i$. The loss is defined as the accumulation of the curvature distortion as

$$\mathcal{L}_{\text{curvature}} = \sum_{i \in \partial\Omega} w_i \|(\mathbf{e}_{i1} - \mathbf{T}_i \bar{\mathbf{e}}_{i1}) + (\mathbf{e}_{i2} - \mathbf{T}_i \bar{\mathbf{e}}_{i2})\|^2, \quad (8)$$

where the quantities denoted by $\bar{\cdot}$ refer to the UV coordinates in the original garment pattern.

5. Experiments

Data: We evaluate our method on a variety of 3D scans of humans captured with a 3dMD [1] system. We select 4 subjects wearing different garments (dress, long-sleeve, polo, shirt) and obtain their corresponding 3D reconstructions. Note that these scans tend to be noisy and contain holes. Nevertheless, they can still serve as 3D targets during the optimization process. Since there are no perfect or clean “ground-truth” 3D scans available for evaluation, we have asked a skilled professional artist to create virtual garments that match the scans to the best of their ability. This process serves as our upper quality bar. Therefore, we provide quantitative comparisons using both the 3D scan input and the artist-made garments as ground-truth to demonstrate the clear improvements DiffAvatar offers over previous work.

Baselines: We evaluate the output geometry of our method against two methods that employ 2D images as an input and one that uses 3D scans. Specifically, we run PiFu-HD [47] on the clothed human scan and segment the garment in 3D to use it for evaluations. We also employed a recent diffusion-based approach that given a single garment image, synthesizes six multi-view consistent novel views,

we then use NeuS [57] to extract the 3D geometry. We also compare our method against PoP [37] which is the closest to our work as it uses 3D scans as an input and outputs a point cloud of the clothed human which we pass to Poisson reconstruction to obtain a mesh. Finally, we provide a comparison of the 2D patterns against NeuralTailor [25] which estimates 2D garment patterns from 3D point clouds of a draped garment in Fig. 4.

Evaluation Metrics: Similar to [54] we use the Chamfer Distance (CD) for comparisons directly in 3D, and LPIPS [66] and SSIM [58] for perceptual metrics in the 2D space using the exact same rendering conditions for all methods. We evaluate the mesh quality of the results using the triangle conditioning metric [50]. All metrics are reported in Tab. 1.

Implementation Details: Our method is implemented in C++ using open-source libraries such as Eigen and libigl. All experiments were conducted on a machine with 14-core i7 CPU with 32GB RAM. Note that there is no limitation to implement our method on GPU, and it would benefit from our implementation since the expensive Jacobian computations map well to highly parallel GPU code. The linear solve can be further accelerated with cuSPARSE.

Performance: DiffAvatar incorporates simulation within an optimization process to yield high-quality outcomes at the expense of increased computational requirements. Optimization takes about one minute per iteration and total times vary between 20 to 200 minutes depending on the garment and iteration count. Baseline methods range between several seconds to 10 minutes for complete inference results. Note that our method can be executed in batch mode automatically. Our method runs on CPU, whereas the baselines are run on an NVIDIA GeForce RTX 4080 GPU.

5.1. Garment Pattern Optimization

The corresponding 2D optimized patterns for the dress and long sleeve shirt are shown in Fig. 4. Starting from the initial panel in the first column, DiffAvatar (last column) generates patterns closely resembling the one designed manually by an artist (second last column). In contrast, although NeuralTailor [25] (second column) does not need an initial template, the result can be far from the target and can even be missing key features like the shirt sleeves.

5.2. Body and Cloth Material Optimization

We visualize different stages of our body shape estimation in Fig. 5 (left) and demonstrate how our physics-aware method improves the estimated body shape. The right of the figure demonstrates our ability to recover cloth material properties to closely match the garment drape in the scan.

5.3. Method Evaluations

We evaluate the reconstructed 3D geometry of our approach against three prior works (Fig. 6) and report quan-

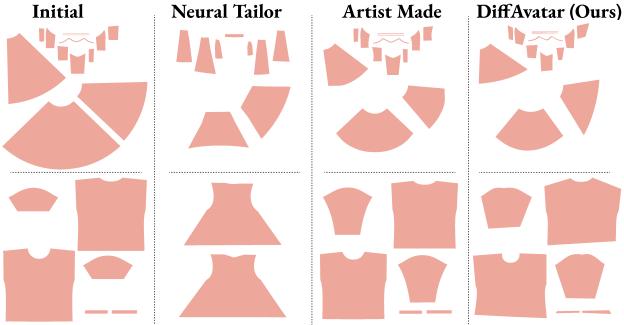


Figure 4. **2D pattern comparison.** The automatically optimized 2D patterns of the dress (first row) and long sleeve shirt (second row) by DiffAvatar closely match the manually created artist ones. However, those generated by NeuralTailor [25] do not resemble the artist-made patterns closely and miss important details.

Method	Against GT Scan			Against GT Artist			Mesh Quality ↑ min(avg)
	CD ↓	LPIPS ↓	SSIM ↑	CD ↓	LPIPS ↓	SSIM ↑	
Scan	-	-	-	1.045	0.127	0.852	0.099(0.422)
Artist	1.045	0.127	0.852	-	-	-	0.171(0.389)
Initial	3.071	0.165	0.815	3.396	0.123	0.859	0.188(0.391)
PiFU-HD [47]	1.930	0.145	0.836	2.009	0.129	0.836	0.000(0.305)
Diffusion+NeuS [57]	3.410	0.171	0.799	3.362	0.177	0.797	0.000(0.266)
PoP [37]	1.695	0.140	0.831	1.866	0.092	0.842	0.000(0.316)
DiffAvatar (Ours)	1.311	0.133	0.842	1.688	0.085	0.893	0.143(0.373)

Table 1. **Quantitative Comparisons.** We used the 3D scan and artist-made mesh as ground truth to evaluate our method on the dress example. Our results show that we achieve the closest CD fit, best perceptual metrics, and produce good mesh quality. In contrast, all competing methods produce a minimum mesh quality of 0 (or near 0), making their output unsuitable for simulation.

itative results in Tab. 1. Regardless of the ground-truth considered (scan or artist-made), DiffAvatar outperforms all prior works across both 2D and 3D metrics. PiFU-HD [47] and Diffusion+NeuS reconstruct the frontal part of the geometry fairly well but the back side is smooth and all results lack fine-level details (wrinkles and folds). PoP works better for tight-fit garments but fails to accurately reconstruct the details of a dress, producing closed-surface meshes without arm holes that are unsuitable for simulation. It is evident from these results that our approach is the only one which faithfully captures the garment with simulation-ready topology. To evaluate mesh quality, we apply a conditioning quality metric [50] to the 3D meshes for a fair comparison with the baseline methods that do not produce rest shape geometry. Prior methods produce a near 0 minimum mesh quality, which indicates the presence of poorly-conditioned or zero-area triangles that are unsuitable for simulation. Our results show favorable quality compared to all past works.

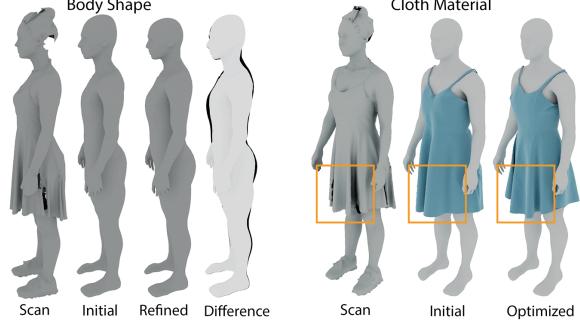


Figure 5. **Body shape and cloth material estimation.** *Left:* We fit a statistical body model to the 3D scan and refine this estimate using our differentiable simulation pipeline and show the difference in shape between initial and refined in black. *Right:* Our initial material estimate produces large folds that do not match the scan as well as our optimized result shown rightmost.

Method Variant	CD ↓	LPIPS ↓	SSIM ↑
w/o Control Cage	3.866	0.168	0.792
w/o Seam Length Term	1.409	0.115	0.843
w/o Boundary Curvature Term	2.249	0.124	0.838
DiffAvatar (Complete)	1.122	0.102	0.863

Table 2. **Ablation Study.** By removing each of the proposed components of DiffAvatar we showcase their impact in 3D with Chamfer Distance and 2D with perceptual metrics to the final result for the *dress* against the ground-truth scan.

5.4. Ablation Studies

We perform an ablation study on the importance of individual components in our design, see Fig. 7 and Tab. 2. We demonstrate that our control cage formulation is crucial for producing physically correct results that can be simulated. The seam length regularization is necessary to prevent seam length mismatches, which leads to excessive gathering of the fabric. The boundary curvature regularization is required to preserve the design intent of the garment.

5.5. Novel Simulated Sequences

In contrast to baseline methods, DiffAvatar is the only one that generates high-quality simulation-ready geometry with associated 2D rest shape and cloth material properties enabling us to create new simulations that are faithful to the original garment with ease. Fig. 1 shows select frames from novel simulated sequences using the optimized dress. See the supplemental material for additional animations.

5.6. Limitations and Future Work

Our method optimizes through a continuum of pattern variations starting from a template based on the garment category. Although we do not address discrete changes in

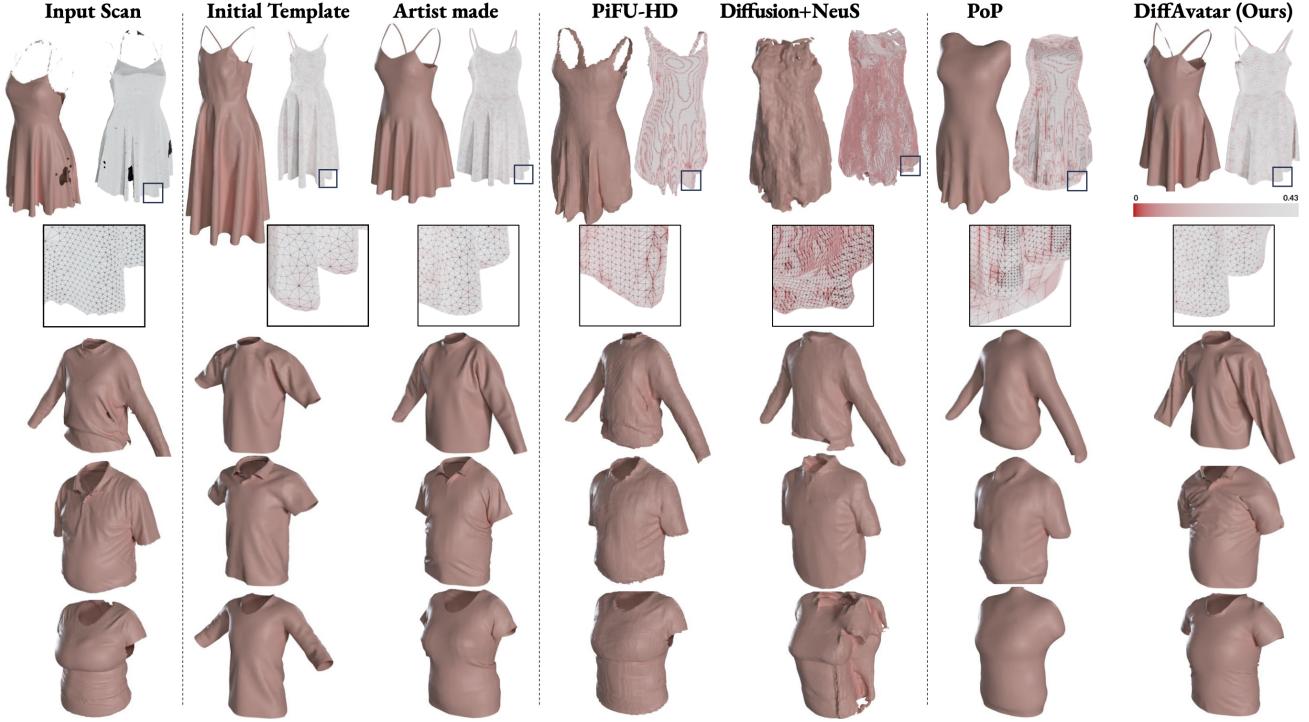


Figure 6. **Qualitative Comparisons.** DiffAvatar faithfully captures garments with natural draping behavior and wrinkle details where all prior works fail to reconstruct simulation-ready meshes. **Mesh quality (Top row).** The generated mesh quality of the mesh visualized with red-to-white gradient representing lowest to highest quality. DiffAvatar generates simulation-ready meshes of high quality, comparable to artist-made meshes where 2D prior works such as PiFu-HD [47] and Diffusion+NeuS [57], or 3D works such as PoP [37] come short.

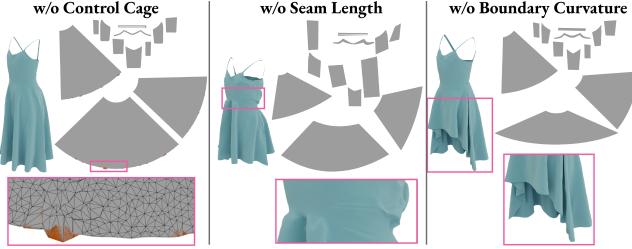


Figure 7. **Ablation Study.** *Left:* w/o control cage, the optimizer quickly produces inverted non-physical triangle elements (highlighted in orange) in the rest shape which causes any simulator to fail. *Middle:* w/o seam length regularization, the seam lines do not match leading to excessive amount of fabric. *Right:* w/o boundary curvature regularization, the pattern distorts into unwanted shapes.

the number of pattern pieces or mesh topology, such a system can be incorporated into the pipeline retroactively. We use dynamic simulation but rely on states close to quasi-equilibrium. This implies that draped garments with strong friction or dynamic effects can be challenging to estimate and can produce a different final aesthetic. Since we are already using a dynamic simulator, a straightforward extension is to match dynamic sequences and recover additional parameters. For multi-layered clothing, garments can be

occluded, making recovery a fundamentally difficult goal. Our method is well suited to handle occlusions due to its strong physical priors about the behavior fabric.

6. Conclusion

We introduced DiffAvatar a new approach that utilizes differentiable simulation for scene recovery to generate high-quality, physically plausible assets that can be used for simulation applications. Our method considers the complex non-linear behavior of cloth and its intricate interaction with the underlying body when optimizing for scene parameters in a unified and coupled manner that takes into account the interplay of all components. We showcased that DiffAvatar outperforms prior works across different metrics, producing high-quality garment results in both 3D and the 2D pattern space and generates simulation-ready assets close to those that are manually designed by a trained artist.

References

- [1] 3dmd applications. from healthcare to artificial intelligence. <https://3dmd.com/>. Accessed on November 2023. **4, 6**
- [2] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: shape

completion and animation of people. In *ACM SIGGRAPH 2005 Papers*, pages 408–416. 2005. 2

[3] Seungbae Bang, Maria Korosteleva, and Sung-Hee Lee. Estimating garment patterns from static scan data. *Computer Graphics Forum*, 40(6):273–287, 2021. 2

[4] David Baraff and Andrew Witkin. Large steps in cloth simulation. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 43–54, 1998. 2

[5] Aric Bartle, Alla Sheffer, Vladimir G. Kim, Danny M. Kaufman, Nicholas Vining, and Floraine Berthouzoz. Physics-driven pattern adjustment for direct 3d garment editing. *ACM Trans. Graph.*, 35(4), jul 2016. 2

[6] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. Neural cloth simulation. *ACM Transactions on Graphics (TOG)*, 41(6):1–14, 2022. 2

[7] Sofien Bouaziz, Sebastian Martin, Tiantian Liu, Ladislav Kavan, and Mark Pauly. Projective dynamics: Fusing constraint projections for fast simulation. *ACM transactions on graphics (TOG)*, 33(4):1–11, 2014. 2

[8] Andrés Casado-Elvira, Marc Comino Trinidad, and Dan Casas. Pergamo: Personalized 3d garments from monocular video. In *Computer Graphics Forum*, volume 41, pages 293–304. Wiley Online Library, 2022. 2

[9] Hsiao-yu Chen, Edith Tretschk, Tuur Stuyck, Petr Kadlecak, Ladislav Kavan, Etienne Vouga, and Christoph Lassner. Virtual elastic objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15827–15837, 2022. 2

[10] Xin Chen, Anqi Pang, Wei Yang, Peihao Wang, Lan Xu, and Jingyi Yu. Tightcap: 3d human shape capture with clothing tightness field. *ACM Transactions on Graphics (TOG)*, 41(1):1–17, 2021. 2

[11] Xipeng Chen, Guangrun Wang, Dizhong Zhu, Xiaodan Liang, Philip H. S. Torr, and Liang Lin. Structure-preserving 3d garment modeling with neural sewing machines. In *NeurIPS*, 2022. 2

[12] Kwang-Jin Choi and Hyeong-Seok Ko. Stable but responsive cloth. In *ACM SIGGRAPH 2005 Courses*, 2005. 2

[13] Luca De Luig, Ren Li, Benoit Guillard, Mathieu Salzmann, and Pascal Fua. DrapeNet: Garment Generation and Self-Supervised Draping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2

[14] Tao Du, Kui Wu, Pingchuan Ma, Sebastien Wah, Andrew Spielberg, Daniela Rus, and Wojciech Matusik. Diffpd: Differentiable projective dynamics. *ACM Transactions on Graphics (TOG)*, 41(2):1–21, 2021. 2

[15] Xudong Feng, Wenchao Huang, Weiwei Xu, and Huamin Wang. Learning-based bending stiffness parameter estimation by a drape tester. *ACM Trans. Graph.*, 41(6), nov 2022. 5

[16] Cheng-Yang Fu, Tamara L Berg, and Alexander C Berg. Imp: Instance mask projection for high accuracy semantic segmentation of things. In *ICCV*, 2019. 4

[17] Yotam Gingold, Adrian Secord, Jefferson Y Han, Eitan Grinspun, and Denis Zorin. A discrete model for inelastic deformation of thin shells. In *ACM SIGGRAPH/Eurographics symposium on computer animation*. Citeseer, 2004. 2

[18] Chihiro Goto and Nobuyuki Umetani. Data-driven garment pattern estimation from 3d geometries. In *Eurographics*, 2021. 2

[19] Artur Grigorev, Michael J Black, and Otmar Hilliges. Hood: Hierarchical graphs for generalized modelling of clothing dynamics. In *CVPR*, 2023. 2

[20] Eitan Grinspun, Anil N Hirani, Mathieu Desbrun, and Peter Schröder. Discrete shells. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 62–67. Citeseer, 2003. 2

[21] Jingfan Guo, Jie Li, Rahul Narain, and Hyun Soo Park. Inverse simulation: Reconstructing dynamic geometry of clothed humans via optimal control. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14698–14707, 2021. 2

[22] Oshri Halimi, Tuur Stuyck, Donglai Xiang, Timur Bagautdinov, He Wen, Ron Kimmel, Takaaki Shiratori, Chenglei Wu, Yaser Sheikh, and Fabian Prada. Pattern-based cloth registration and sparse-view animation. *ACM Transactions on Graphics (TOG)*, 41(6):1–17, 2022. 2

[23] Tao Ju, Scott Schaefer, and Joe Warren. Mean value coordinates for closed triangular meshes. *ACM Transactions on Graphics*, 24(3):561–566, 2005. 5

[24] Navami Kairanda, Marc Habermann, Christian Theobalt, and Vladislav Golyanik. Neuralclothsim: Neural deformation fields meet the kirchhoff-love thin shell theory. *arXiv preprint arXiv:2308.12970*, 2023. 2

[25] Maria Korosteleva and Sung-Hee Lee. Neuraltailor: Reconstructing sewing pattern structures from 3d point clouds of garments. *ACM Transactions on Graphics (TOG)*, 41(4):1–16, 2022. 2, 6, 7

[26] Maria Korosteleva and Olga Sorkine-Hornung. Garment-Code: Programming parametric sewing patterns. *ACM Transaction on Graphics*, 42(6), 2023. SIGGRAPH ASIA. 2

[27] Egor Larionov, Marie-Lena Eckert, Katja Wolff, and Tuur Stuyck. Estimating cloth elasticity parameters using position-based simulation of compliant constrained dynamics. *arXiv preprint arXiv:2212.08790*, 2022. 2

[28] Dohae Lee and In-Kwon Lee. Multi-layered unseen garments draping network. *arXiv preprint arXiv:2304.03492*, 2023. 2

[29] Ren Li, Benoît Guillard, and Pascal Fua. Isp: Multi-layered garment draping with implicit sewing patterns. In *NeurIPS*, 2023. 2

[30] Ren Li, Benoît Guillard, Edoardo Remelli, and Pascal Fua. Dig: Draping implicit garment over the human body. In *ACCV*, 2022. 2

[31] Yifei Li, Tao Du, Sangeetha Gramma Srinivasan, Kui Wu, Bo Zhu, Eftychios Sifakis, and Wojciech Matusik. Fluidic topology optimization with an anisotropic mixture model. *ACM Trans. Graph.*, nov 2022. 2

[32] Yifei Li, Tao Du, Kui Wu, Jie Xu, and Wojciech Matusik. Diffcloth: Differentiable cloth simulation with dry frictional contact. *ACM Transactions on Graphics (TOG)*, 42(1):1–20, 2022. 2

[33] Yue Li, Marc Habermann, Bernhard Thomaszewski, Stelian Coros, Thabo Beeler, and Christian Theobalt. Deep physics-aware inference of cloth deformation for monocular human performance capture. In *2021 International Conference on 3D Vision (3DV)*, pages 373–384, Los Alamitos, CA, USA, dec 2021. IEEE Computer Society. 2

[34] Junbang Liang and Ming Lin. Fabric material recovery from video using multi-scale geometric auto-encoder. In *European Conference on Computer Vision*, pages 695–714. Springer, 2022. 2

[35] Lijuan Liu, Xiangyu Xu, Zhijie Lin, Jiabin Liang, and Shuicheng Yan. Towards garment sewing pattern reconstruction from a single image. *ACM Transactions on Graphics (SIGGRAPH Asia)*, 2023. 2

[36] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, Oct. 2015. 2

[37] Qianli Ma, Jinlong Yang, Siyu Tang, and Michael J Black. The power of points for modeling humans in clothing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10974–10984, 2021. 6, 7, 8

[38] Miles Macklin, Matthias Müller, and Nuttapong Chentanez. Xpbd: position-based simulation of compliant constrained dynamics. In *Proceedings of the 9th International Conference on Motion in Games*, pages 49–54, 2016. 2, 3

[39] Thalmann Magnenat, Richard Laperrière, and Daniel Thalmann. Joint-dependent local deformations for hand animation and object grasping. Technical report, Canadian Inf. Process. Soc, 1988. 3

[40] Antoine McNamara, Adrien Treuille, Zoran Popović, and Jos Stam. Fluid control using the adjoint method. *ACM Transactions On Graphics (TOG)*, 23(3):449–456, 2004. 2

[41] Matthias Müller, Bruno Heidelberger, Marcus Hennix, and John Ratcliff. Position based dynamics. *Journal of Visual Communication and Image Representation*, 18(2):109–118, 2007. 2

[42] Nico Pietroni, Corentin Dumery, Raphael Falque, Mark Liu, Teresa Vidal-Calleja, and Olga Sorkine-Hornung. Computational pattern making from 3d garment models. *ACM Trans. Graph.*, 41(4), jul 2022. 2

[43] Anran Qi, Sauradip Nag, Xiatian Zhu, and Ariel Shamir. Personal tailor: Personalizing 2d pattern design from 3d garment point clouds. *arXiv preprint arXiv:2303.09695*, 2023. 2

[44] Lingteng Qiu, Guanying Chen, Jiapeng Zhou, Mutian Xu, Junle Wang, and Xiaoguang Han. Rec-mv: Reconstructing 3d dynamic cloth from monocular videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4637–4646, 2023. 2

[45] Carlos Rodriguez-Pardo, Melania Prieto-Martin, Dan Casas, and Elena Garces. How will it drape like? capturing fabric mechanics from depth images. In *Computer Graphics Forum*, volume 42, pages 149–160. Wiley Online Library, 2023. 2

[46] Cristian Romero, Dan Casas, Maurizio M Chiaramonte, and Miguel A Otaduy. Contact-centric deformation learning. *ACM Transactions on Graphics (TOG)*, 41(4):1–11, 2022. 2

[47] Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 84–93, 2020. 6, 7, 8

[48] Igor Santesteban, Miguel A Otaduy, and Dan Casas. Snug: Self-supervised neural dynamic garments. In *CVPR*, 2022. 2

[49] Igor Santesteban, Nils Thuerey, Miguel A Otaduy, and Dan Casas. Self-supervised collision handling via generative 3d garment models for virtual try-on. In *CVPR*, 2021. 2

[50] Jonathan Richard Shewchuk. What is a good linear element? interpolation, conditioning, and quality measures. In *International Meshing Roundtable Conference*, 2002. 6, 7

[51] Oded Stein, Eitan Grinspun, and Keenan Crane. Developability of triangle meshes. *ACM Transactions on Graphics (TOG)*, 37(4):1–14, 2018. 2, 3

[52] Tuur Stuyck. *Cloth simulation for computer graphics*. Springer Nature, 2022. 2

[53] Tuur Stuyck and Hsiao-yu Chen. Diffxpbd: Differentiable position-based simulation of compliant constraint dynamics. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 6(3):1–14, 2023. 2, 4

[54] Zhaoqi Su, Liangxiao Hu, Siyou Lin, Hongwen Zhang, Shengping Zhang, Justus Thies, and Yebin Liu. Caphy: Capturing physical properties for animatable human avatars. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14150–14160, 2023. 2, 6

[55] Garvita Tiwari, Nikolaos Sarafianos, Tony Tung, and Gerard Pons-Moll. Neural-gif: Neural generalized implicit functions for animating people in clothing. In *ICCV*, 2021. 2

[56] Huamin Wang. Rule-free sewing pattern adjustment with precision and efficiency. *ACM Transactions on Graphics (TOG)*, 37(4):1–13, 2018. 2, 6

[57] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021. 6, 7, 8

[58] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[59] Ziyan Wang, Giljoo Nam, Tuur Stuyck, Stephen Lombardi, Chen Cao, Jason Saragih, Michael Zollhöfer, Jessica Hodgins, and Christoph Lassner. Neuwigs: A neural dynamic model for volumetric hair capture and animation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8641–8651, June 2023. 2

[60] Ziyan Wang, Giljoo Nam, Tuur Stuyck, Stephen Lombardi, Michael Zollhöfer, Jessica Hodgins, and Christoph Lassner. Hvh: Learning a hybrid neural volumetric representation for dynamic hair performance capture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6143–6154, 2022. 2

[61] Chris Wojtan, Peter J Mucha, and Greg Turk. Keyframe control of complex particle systems using the adjoint method. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 15–23, 2006. 2

- [62] Katja Wolff, Philipp Herholz, Verena Ziegler, Frauke Link, Nico Brügel, and Olga Sorkine-Hornung. Designing personalized garments with body movement. *Computer Graphics Forum*, 42(1):180–194, 2023. [2](#)
- [63] Donglai Xiang, Timur Bagautdinov, Tuur Stuyck, Fabian Prada, Javier Romero, Weipeng Xu, Shunsuke Saito, Jing-fan Guo, Breannan Smith, Takaaki Shiratori, et al. Dressing avatars: Deep photorealistic appearance for physically simulated clothing. *ACM Transactions on Graphics (TOG)*, 41(6):1–15, 2022. [2](#)
- [64] Yuxuan Xue, Bharat Lal Bhatnagar, Riccardo Marin, Niko-laos Sarafianos, Yuanlu Xu, Gerard Pons-Moll, and Tony Tung. Nsf: Neural surface fields for human modeling from monocular depth. In *ICCV*, 2023. [2](#)
- [65] Shan Yang, Zherong Pan, Tanya Amert, Ke Wang, Licheng Yu, Tamara Berg, and Ming C. Lin. Physics-inspired garment recovery from a single-view image. *ACM Trans. Graph.*, 37(5), nov 2018. [2](#)
- [66] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. [6](#)
- [67] Xingxing Zou, Xintong Han, and Waikeung Wong. Cloth4d: A dataset for clothed human reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12847–12857, 2023. [2](#)