MGN: Multi-Layered Garment Animation Generation Neural Network

Min Shi, Guoqing Han, Tianlu Mao, Xinru Zhuo, ZhenYu Li, Xinran Wang, Lin Gao, Yukun Lai and Dengming Zhu

Abstract

This paper presents a multi-layered garment animation generation method. Generating realistic dynamics in 3D garment animations is a challenging task due to the complex nature of multi-layered garments and the variety of outer forces involved. Existing data-driven approaches have mainly focused on the study of static draping deformation of multi-layer garments, with less consideration for the temporal deformation of garments, such as the time-varying motion behaviors of individual layers and their continuous interactions during motion. Additionally, these methods require a substantial amount of high-quality paired garment datasets for network training, leading to a costly data acquisition and annotation process. To address these challenges, we propose a multi-layered garment animation generation method that explicitly models different garment layers as separate meshes, and employs a combination of unsupervised and temporally supervised learning strategies to analyze and model the behavior of individual garment layers and their interactions. Our primary contribution lies in introducing a two-stage network architecture for layered garment processing, which decomposes multi-layer garment deformation prediction into single-layer garment generation and inter-layer garment interaction deformation. We focus more on generating two-layered clothing animations. Of course, our two-layered approach can be used iteratively to support more layers by using the current outer layer as the inner layer for the next iteration. This approach achieves dynamic simulation of multi-layer garments, and experimental results demonstrate that our method can generate realistic multi-layer garment deformation effects, outperforming existing methods both visually and in terms of evaluation metrics.

KEYWORDS： Multi-layer garments animation, Computer Graphics, Unsupervised Learning, Garment deformation, Transformer

Introduction

Animating 3D garments plays a significant role in various domains such as computer animation, gaming, and virtual reality. The quality of garment animations often profoundly impacts the visual experience of an entire animation scene. To achieve realistic garment animation effects, creators typically invest a substantial amount of time in performing physical simulations[1, 2] and rendering garment animations using offline baking methods[3]. However, such methods are costly and not easily reusable.

In recent years, data-driven approaches[4, 5] have gained widespread popularity in garment animation due to their speed and ability to generate visually appealing deformations[6]. Existing learning-based methods have primarily focused on single-layer garments driven by human body motion and struggle with the complexities of multi-layer garment simulations in typical scenarios. In the process of simulating deformations in multi-layer garments, the complex interactions between layers during motion need to be considered, and an increase in the number of garment layers leads to an increase in sample data sets and network computational costs. Existing multi-layer garment deformation methods[7–9] have primarily concentrated on complex multi-layer garment hanging effects, neglecting the temporal aspect of garment deformations, making them unsuitable for handling multi-layer garment deformation animations. While Shao et al. [10] proposed a method that can handle dynamic deformations in multi-layer garments, it divides the garment model into a two-tier structure based on UV patches and converts them into particle representations for deformation simulation, reducing network computational load but sacrificing some garment fold details.

It is highly challenging to predict and generate physically plausible garment deformations in the presence of complex interactions among multi-layer garments during motion when using learning-based approaches. Inspired by the independent layer calculations in physical simulations, we present a method for generating multi-layer garment animations, breaking down the prediction of multi-layer garment deformations into single-layer garment deformations and inter-layer garment deformation interactions. Through a combination of unsupervised learning based on layer interactions and temporally supervised training strategies, individual garment layers and interactions between different layers are separately analyzed and modeled, achieving dynamic simulation of multi-layer garments, while substantially reducing the demands for ground truth training data. Our two-layered approach can be used iteratively to support more layers, by using the current outer layer as the inner layer for the next iteration. A series of experiments demonstrate that our network outperforms state-of-the-art garment deformation methods, both qualitatively and quantitatively. Some examples of our method are presented in Fig. 1.

FIGURE 1 3D garment deformation predicted by the proposed MGN following given body motions. Leveraging garment layer calculations, MGN is capable of realistically deforming multi-layer garment meshes.

The main contributions of our work are as follows:

(1) We introduce a two-stage network architecture for layered garment processing. This architecture can be based on unsupervised single-layer clothing generation methods, reducing dependency on clothing sample data and providing a data foundation for multi-layer clothing deformation.

(2) We propose a new multi-layer garment deformation prediction network based on Transformer. This network integrates deformations generated by the unsupervised method and those of another garment layer, learning the deformation patterns between multiple garment layers, suitable for multi-layer garment deformations under different postures and animation generation.

Related Work

Physics-based simulations

As a conventional approach in garment deformation, physics-based simulation[11–14] has been an active research topic in the field of computer graphics due to its realistic and detailed simulation results. By establishing dynamic models[1, 15] and solving a large number of dynamic equations, it can achieve realistic and detailed garment simulation effects.

Physics-based simulation constantly analyzes the forces acting on the garments[16, 17] to generate garment deformations. However, any change in the model's conditions necessitates the recalculation of the equation system, which incurs a significant time and computational cost.

Learning-based models

Learning-based methods have received increasing attention in recent years due to their fast computational speed and ability to generate visually appealing deformations[18–22]. They have found widespread application in garment animation.

Xu et al.[23], by studying garment model instances and human motion characteristics, treated garment deformation as a function of body shape and pose. They adapted the learning methods from SMPL[24] to learn garment deformation functions. They proposed a garment animation synthesis method that, based on a given human pose and garment instance data, synthesizes garment deformations to create realistic garment animation effects. Patel et al.[6] introduced TailorNet, which decouples garment deformation into high-frequency and low-frequency components. By recombining these components, rich garment fold deformations can be synthesized.

Most research works have only considered specific single garment deformations and are not suitable for handling complex multi-layer scenarios. Furthermore, they require a large dataset of garment deformations obtained through physical simulations to extract deformation patterns. Bertiche et al.[25] and Santesteban et al.[26] employed unsupervised or self-supervised training strategies to generate garment deformation effects without the need for constructing extensive garment sample datasets. However, these methods are limited to single-layer garment deformations. The PGN-Cloth method[27], NCS method[28] and HOOD methods[29] also encounter the same challenges. The DrapeNet method[30] and the GenSim method[31], while eliminating dependency on datasets by transforming physical constraints, can only generate static garment drape deformation.

Bhatnagar et al.[32] can handle multiple garments, but there are very few overlapping areas among different pieces of clothing. Zhang et al.'s method[33] can generate multi-layer clothing animations iteratively but requires extensive preparation of clothing sample data in the initial stages. Santesteban et al.[7] were among the first to explore decoupling multi-layer garments, but their approach is limited to standard postures and requires learning for each garment in a preprocessing stage, making it challenging to apply in dynamic garments deformation. Li et al.[8]introduced ISP, which represents garment models using the UV map of 2D patches and implicitly represents garment shapes with 2D SDF (Signed Distance Function) values, addressing the hanging issue in multi-layer garments by avoiding the limitations of SDF on open 3D surfaces. Lee et al.[9] proposed MLU-Net, which maps garments of arbitrary topology onto the UV map of the human body, converting them into a uniform representation. They used a graph neural network (GNN) to simulate interactions between garments, appropriately deforming garments to prevent interpenetration and effectively simulating the hanging effects of complex multi-layer garments. However, the aforementioned methods do not consider the temporal aspect of garment deformation and are primarily suitable for generating single-frame multi-layer garment deformations. Bertiche et al.[34] employ Graph Neural Networks to animate garment. They supplement supervised learning with implicit constraints to handle collision penetration, generating dynamic garment deformation effects through a combination of ground truth and implicit constraints. However, their network training still requires a large amount of sample data. By utilizing the rotational invariance and additivity of physical systems, Shao et al.[10] captured and handled interactions between garment components, different garments and driver factors.

The method transformed garment UV patches into particle representations for deformation simulation, reducing network computational load but sacrificing some garment fold details.

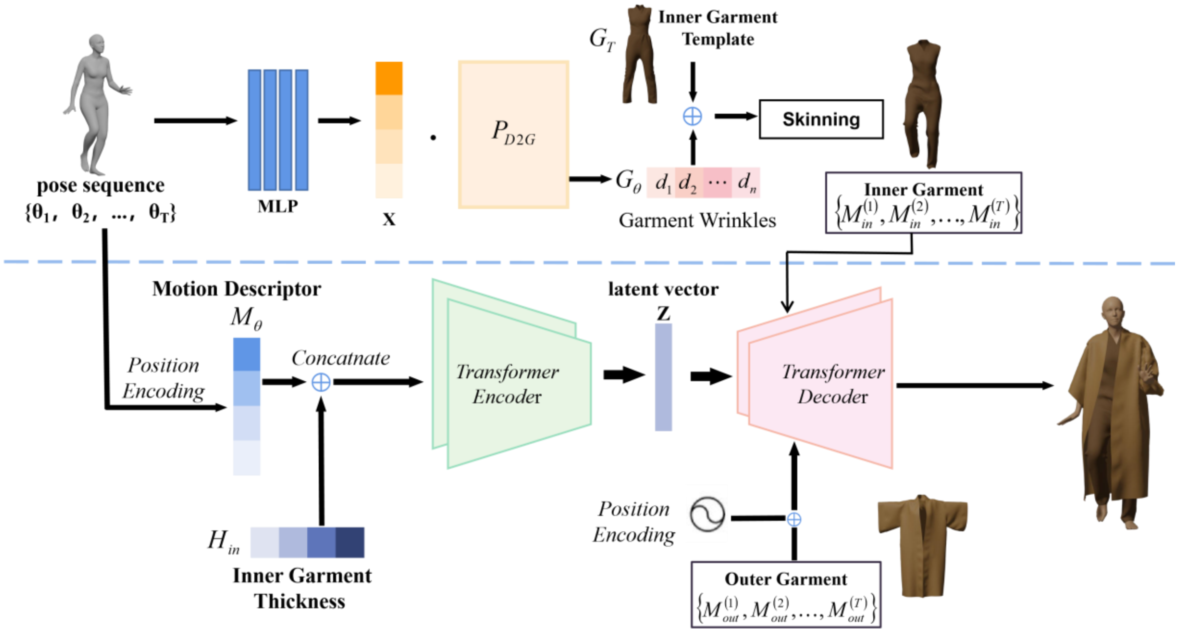

FIGURE 2 Overview of the proposed MGN. Given a body pose sequence with T frames $\theta_1, \theta_2, \ldots, \theta_T$, inner garment thickness H_{in} and an inner garment mesh template G_T, the garment deformation is predicted by MGN. The unsupervised generation network primarily consists of MLP layers, which construct target constraint functions to constrain garment deformations to adhere to physical laws. The layer-interaction network is based on a Transformer encoder-decoder architecture and combines deformations generated through the unsupervised strategy for single-layer garments with outer layer garment deformation features. This fusion and training process allows the model to learn the deformation patterns between multiple garment layers.

Method

Overview

This method is designed to predict and generate multi-layer garment deformations based on a given human motion sequence. An overview of our method is illustrated in Fig. 2.

We employ a mesh representation to define the garment model, which includes the garment's vertices and their corresponding connectivity structure. This representation is specifically denoted as $M = (V, F)$, where V represents the set of garment mesh vertices, and F represents the set of faces describing vertex connectivity relationships. We take a multi-layer approach and describe the whole garment as inner garment and outer garment, using meshes M_{in} and M_{out}, respectively. We use an unsupervised single-layer garment generation network to generate inner layer garment mesh M_{in}. We then propose a layer-interaction network to handle inter-layer interactions and generate outer layer garment mesh M_{out}. To produce a garment animation sequence, we take a pose sequence $\theta_1, \theta_2, \ldots, \theta_T$ as input, where the pose parameters $\theta_i \in R^{24\times3}$, and produce the inner and outer garment meshes $M_{in}^{(t)}, M_{out}^{(t)}$, where $t = 1, 2, \ldots, T$ is the frame index. $^{(t)}$ is omitted for simplicity when it does not cause confusion. For meshes in the sequence, the connectivity remains unchanged and only vertex positions change over the frames.

Single-layer Unsupervised Generative Network

We use single-layer unsupervised network to predict the inner garment deformation, which reduces training data demands as the inner garment is directly influenced by the human body deformation. The following description in this subsection only applies to inner garment meshes.3D animated models are often constructed using skinning and/or blend shapes. We utilize a representation based on the parameterized human body model SMPL[24] to represent the garments, which allows us to use an unsupervised training strategy to generate clothing deformations. We decompose a tight-fitting garment model into pose-driven garment fold deformations and skin-driven rigid deformations. The specific representation is as follows:

$$M_{in} = W(G_T + G_\theta, \theta, W_G) \tag{1}$$

Where $W(\cdot)$ is the skinning function, G_T is the garment template in the standard pose, G_θ represents the pose-driven garment fold deformations. W_G is the garment skin weight generated based on the human body model.

Our single-layer garment generation network primarily consists of MLP perception layers. The network takes the pose parameters $\theta_i \in R^{24\times3}$ as input and generates the pose feature representation X through the MLP layers. G_θ represents the garment fold deformations influenced by the pose features. It establishes a mapping relationship $f(\cdot): X \to G_\theta$ between the pose features and garment deformations through the P_{D2G} (Pose-to-Garment) transformation module. The specific representation is as follows:

$$G_\theta = f_{MLP}(\theta) \cdot P_{D2G} \tag{2}$$

Where $G_\theta = \{d_1, d_2, \ldots, d_n\}$, represents the garment fold deformations influenced by the pose features, which d_i represents the offset of the i-th vertex under the influence of pose. After adding it to the garment template G_T in the standard pose, the garment deformations are generated using the skinning function in Equation 1.

To train our single-layer unsupervised generative network, we transform the kinematic equations complying with the laws of physics and the geometric constraints conforming to physical principles into optimization objectives for the network. The loss function is as follows:

$$L = \lambda_e L_{edge} + \lambda_b L_{bend} + \lambda_c L_{collision} + \lambda_g L_{gravity} \tag{3}$$

Where L_{edge} is the edge error term, L_{bend} is the curvature error term, $L_{collision}$ is the penetration loss term, and $L_{gravity}$ is the gravity error term. λ_e , λ_b , λ_c , and λ_g are the corresponding weights for these terms. L_{edge} and L_{bend} represent geometric loss constraints in accordance with physical principles. L_{edge} enforces constraints on edge lengths between garment vertices, while L_{bend} enforces constraints on the curvature of fabric patches. Their specific formulas are as follows:

$$L_{edge} = \sum_{j=1}^{n_E} \left(\widehat{E_J} - E_j\right)^2 \tag{4}$$

$$L_{bend} = \sum_{k=1}^{n_F} || \vartriangle (N_k)||^2, \tag{5}$$

Where $\widehat{E_J}$ represents the j-th edge length of the garment template in the standard pose, E_j represents the length of the j-th edge generated by the prediction, n_E is the number of edges.

The L_{edge} loss function prevents excessive stretching of the garment mesh. $\Delta(\cdot)$ is the Laplace operator, and N_k represents the normal vector of the k-th garment mesh face, and n_F is the number of faces.

The L_{bend} loss function constrains unnatural bending deformations of the garment mesh faces.

The gravity loss constraint $L_{gravity}$ and the collision force loss $L_{collision}$ are mechanical constraints designed to conform to physical laws. Their specific definitions are as follows:

$$L_{gravity} = \lambda \sum_{i=1}^{n} \cdot V_i^{(z)} \tag{6}$$

$$L_{collision} = \sum_i \min\left(\left(V_{g,i} - V_{b,i'}\right) \cdot N_{b,i'} - \varepsilon, 0\right)^2 \tag{7}$$

Where $\lambda = m \cdot g$ is a preset fixed term in the gravity loss constraint. $V_i^{(z)}$ represents the height in the world space for the i-th vertex location and is used to calculate the gravitational potential energy of garment mesh points. In the collision force loss constraint, $V_{g,i}$ stands for the i-th vertex of the garment, $V_{b,i'}$ represents the i'-th vertex of the nearest point on the human body mesh to the i-th garment vertex, $N_{b,i'}$ represents the normal vector of the i'-th point on the human body, and ε is a predefined threshold representing the penetration tolerance term.

Garment Deformation

The inter-layer interaction network takes the given outer layer garment template and the inner layer garment predicted by the single-layer garment generation network as inputs. It is built on the Transformer architecture and is responsible for learning the deformation patterns of garments influenced by temporal features and inter-layer interaction features. It then predicts and generates the corresponding outer layer garment deformation.

$$M_{out} = Transformer\left(\Theta \,\middle|\, H_{in}\right) \tag{8}$$

In detail, the inter-layer interaction network takes the motion pose parameters $\Theta = \{\theta_1, \theta_2, \ldots, \theta_T\}$ as input and encodes them into pose vectors through an Embedding layer, where T is the number of frames. In the positional encoding layer, temporal information is combined with pose vectors to generate temporal pose vectors D_θ. To generate multi-layered garment animations with different thickness combinations, we concatenate the inner garment thickness H_{in} and the temporal pose vectors D_θ to serve as the input for the Transformer encoder. This allows us to generate outer garment deformations under the influence of varying thicknesses. When predicting the outer layer garment deformation in the decoder, besides the temporal-pose features provided by the encoder, the inner layer garment deformations generated by the single-layer generation module are also used as input. This additional input data is essential for learning inter-layer interaction deformations.

We utilize a temporally supervised training strategy to learn collision contact and temporal deformation features between multi-layer garments. The construction of the loss function is as follows:

$$L_{ML} = L_{cloth} + \lambda_{lap} L_{lap} + \lambda_{ML} L_{repulsive} \tag{9}$$

Where L_{cloth} is the garment prediction loss, measuring the deviation between predicted values and ground truth. L_{lap} is the Laplace loss, which prevents excessive offset deformations in generated garments. $L_{repulsive}$ is the loss related to the repulsive forces between different layers of garments. λ_{lap} and λ_{ML} are the corresponding weights. Their specific definitions are as follows:

$$L_{cloth} = \sum_{t=1}^{T}\left(V_t - \widehat{V_t}\right)^2 \tag{10}$$

$$L_{lap} = \sum_{t=1}^{T} || \Delta\left(V_t\right) - \Delta\left(\widehat{V_t}\right)||_2^2 \tag{11}$$

$$L_{repulsive} = \sum_{t=1}^{T} \sum_{(i,i')\in\mathcal{M}_t} \min\left(\left(v_{out,t}^{i'} - v_{in,t}^{i}\right)\cdot N_{in,t}^{i}, 0\right) \tag{12}$$

Where $\widehat{V_t}$ represents the ground truth outer garment vertices in the L_{cloth} loss term, while V_t represents the network-generated predictions. $\Delta(\cdot)$ in the L_{lap} loss term is the Laplace operator, which maintains local details as invariant as possible. In the $L_{repulsive}$ loss term, $(i, i') \in \mathcal{M}_t$ represents a matching vertex pair, where the i-th vertex of the inner garment mesh is matched to the i'-the vertex of the outer garment mesh. $V_{in,t}^{i}$ stands for the i-th vertex of the inner garment, and $V_{out,t}^{i'}$ represents the position the nearest neighboring vertex (the i'-th vertex) of the outer garment. $N_{in,t}^{i}$ is the i-th vertex normal on the inner garment mesh.

Experiment

Garment Model Data. We used the Marvelous Designer software to model and generate our garment sample data. We have designed multiple garment combinations, taking a tight-fitting Chinese-style sleeveless shirt as the inner layer garment and a loose-fitting wide-sleeved shirt as the outer layer garment as an example, and using Marvelous Designer for modeling. Based on the AMASS[35] human motion data, we selected 8,100 frames of motion data from the AMASS dataset and divided them into 1,350 motion sequences, each with a length of 6 frames, for training and testing.

Implementation Details. The generation network comprises 4 MLP layers, each with a dimension of 32, and the P_{D2G} transformation space dimension is set to $[32, v_{num}, 3]$, Where v_{num} represents the number of vertices in the inner garment. In the inter-layer interaction network, the number of Transformer[36] sub-layers N is 2, the number of attention heads is set to 4, and the dimensions of the feedforward networks in both encoder and decoder are set to $[256,512,256]$, using ReLU activation functions. The model uses the Adam optimizer with a learning rate set to $lr = 1 \times 10^{-3}$ and a batch size of 20. The loss function settings include a weight of $\lambda_e = 15$ for the edge loss in the generation network, $\lambda_b = 5 \times 10^5$ for the curvature loss, and $\lambda_c = 25$ for the penetration loss. In the inter-layer interaction network, the Laplace loss weight is set to $\lambda_{lap} = 0.1$, and the repulsive force loss weight is $\lambda_{ML} = 0.9$.

Reconstruction Experiment

To evaluate the network's deformation prediction capability, 150 frames of human motion data were randomly selected from the test set to drive garment deformation predictions. These predictions were then compared to the ground truth data. The experimental results are shown in Table 1, and the garment deformation effects can be seen in Figure 3.

Table 1. Quantitative Comparison Results With Ground truth.

Metrics	STED	Hausdorff	RMSE
Ours	2.16	2.49	3.25

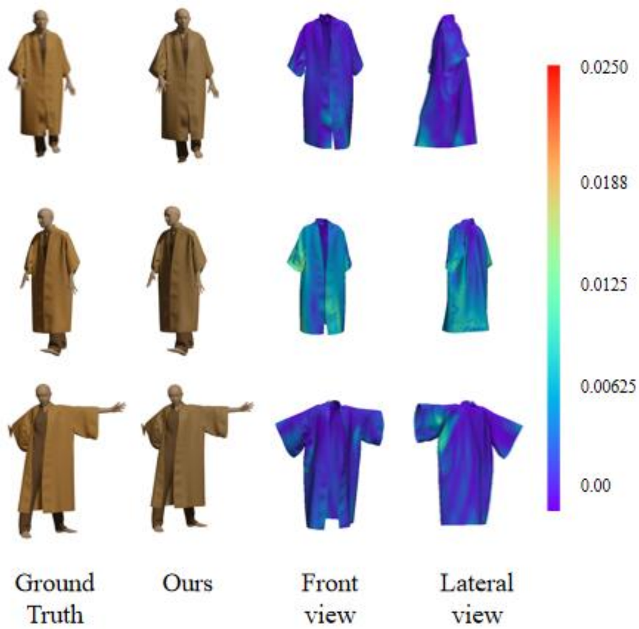

FIGURE 3 Examples of our MGN for garment deformation in different pose. The first column is the ground truth generated by physics-based simulation. The second column is the results obtained by MGN. Our method can produce natural and realistic clothing dynamics. The third column and the fourth column represent the vertex deviations from the front view and the side view, respectively. Colder colors indicate smaller errors.

For quantitative analysis, this paper utilized evaluation metrics such as STED distance, Hausdorff distance, and RMSE. As seen from Table 1 and the garment deformation effects in Figure 3, the network's generated garment deformations exhibit relatively small error values in the metrics, and the model is capable of generating visually reasonable garment deformation effects.

Comparison with Related Methods

To verify the superiority of our method in terms of generated results compared to existing methods, the selected test data is used to drive garment deformation predictions. Our method is compared to the state-of-the-art DeePSD method[34] and the LayersNet method[10] for quantitative and qualitative analysis. We utilize evaluation metrics such as STED distance, Hausdorff distance, RMSE[37], L-Coll and H-Coll. L-Coll represents the degree of penetration between different garment layers, while H-Coll represents the degree of penetration between garments and the human body model.

The evaluation metrics L-Coll and H-Coll[10] are shown in Table 2. The qualitative comparison results among DeePSD, LayersNet, and our method are shown in Fig. 4.

From Table 2, it is evident that our method outperforms DeePSD in all the evaluation metrics, including the error between predictions and ground truth, penetration between multiple garment layers, and penetration loss between garments and the human body model. Our method demonstrates superiority in the metrics compared to the LayersNet method. This is further supported by the visual error plot in Fig. 4, where our method exhibits smaller prediction errors. In the comparison images of garment deformation effects, it can be observed that DeePSD exhibits unnatural wrinkles and penetrations between garments and the human body model. In contrast, the Laplacian smoothing loss and curvature loss terms in our method effectively constrain the curvature between fabric patches, preventing unnatural bending deformations of garment mesh faces. The penetration loss and inter-layer garment repulsive force loss also reduce penetration between inner layer garments and the human body model, as well as penetration between different garment layers. The LayersNet method induces increased errors through UV patch-constrained garment deformation, and in contrast our approach utilizes vertex deformation constraints, resulting in more accurate predictions.

FIGURE 4 Qualitative comparison of 3D garment deformation methods for different types of garments, including LayersNet, DeePSD and Our MGN. With the loss constraints constructed by the inter-layer interaction network, our method is capable of producing smoother and more accurate garment deformations, with fewer instances of garment penetrations. The last three columns visualize the error distributions DeePSD view and Ours view display the error quantities between the predicted garment and the ground truth for DeePSD, LayersNet, and our method. Colder colors indicate smaller errors.

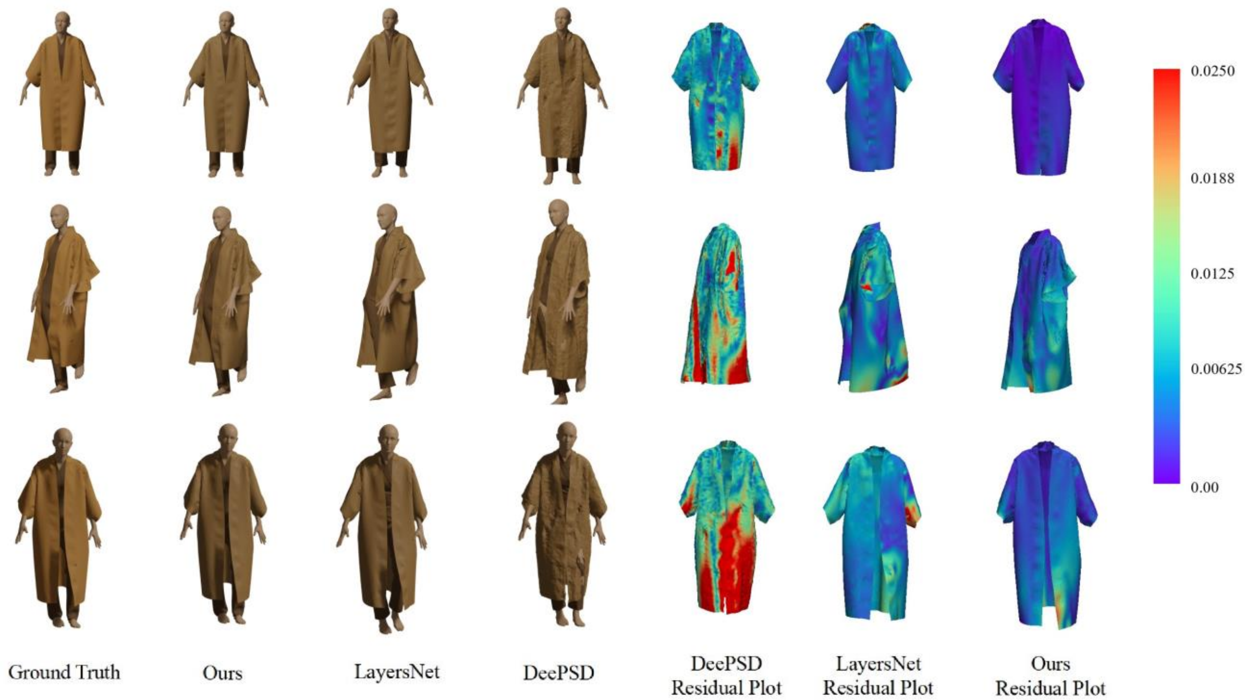

Table 2. Quantitative Comparison Results.

Metrics	STED ↓	Hausdorff↓	RMSE↓	L-Coll ↓	H-Coll ↓
DeePSD	10.68	7.41	11.65	2.43	6.33
LayersNet	**1.25**	4.31	5.66	1.23	4.91
Ours	2.38	**1.68**	**3.42**	**1.08**	**0.12**

Different Combinations of Multi-layer Garments

Our method proposed in this paper is applicable for predicting multi-layer garment deformations and generating animations. We selected various combinations of multi-layer garments and used the method described in this paper for deformation simulation and animation generation. The results, as shown in Fig. 5, depict animations of virtual characters in different pose sequences.

To validate the applicability of our method for different garment combinations, we conducted corresponding quantitative experiments. We used STED, Hausdorff, and RMSE to evaluate the garment reconstruction effects, and L-coll and H-coll to assess the degree of garment penetration.

FIGURE 5 The deformation effects of different types of multi-layer garment combinations.

The results in Table 3 show that our method is capable of handling multi-layer garment deformations for various combinations, generating stable deformations that satisfy visual requirements.

Table 3. Quantitative Comparison Results With Different garment.

Metrics	STED ↓	Hausdorff↓	RMSE↓	L-Coll ↓	H-Coll ↓
Garment 1	2.38	1.68	3.42	1.08	0.12
Garment 2	2.22	1.58	2.15	1.01	0.05
Garment 3	2.23	1.88	3.09	0.07	0.01

Generalizability Experiments

Although our method requires retraining the network for different styles of multi-layer garment combinations, it enables us to generate deformations of multi-layer garments with varying thickness combinations for the same garment style.

We assessed the generalization of the model based on the repulsive force loss between multi-layer garments. We randomly selected 200 frames of human motion data from the test dataset and six variations in clothing thickness [2.5,5,7.5,10,12.5,15] for testing. Using the network proposed in this paper, we predicted and generated the deformation of the outer garment under the influence of randomly selected inner garments of different thicknesses.

The results in Table 4 demonstrate that the proposed model shows consistent performance across different inner garment thicknesses, providing accurate predictions and effectively mitigating garment penetration issues.

Inner Garment Thickness	STED Distortion (10^{-2})	Hausdorff dist (10^{-2})	RMSE	L-Coll (%)	H-Coll (%)
2.5	3.42	4.32	5.81	2.52	1.77
5	3.01	4.17	2.11	1.79	1.41
10	3.41	3.96	3.76	1.26	0.87
15	2.84	2.92	2.87	0.74	0.67
7.5	2.82	4.35	5.46	2.62	1.81
12.5	2.63	3.36	5.09	1.75	1.43

The qualitative results in Fig. 6 further support the model's ability to generate realistic multi-layer garment deformations under varying conditions. The visual results presented in Fig. 6 demonstrate the multi-layer garment deformations for different combinations of inner garment thicknesses ([2.5,5,7.5,10,12.5,15]). The two rows represent the deformations of inner and outer garments under different thicknesses, and each column shows the deformations for the same thickness of both inner and outer garments. Both Table 4 and Fig. 6 confirm that the proposed method is capable of generating garment deformations with small errors and realistic visual effects for various combinations of multi-layer garments.

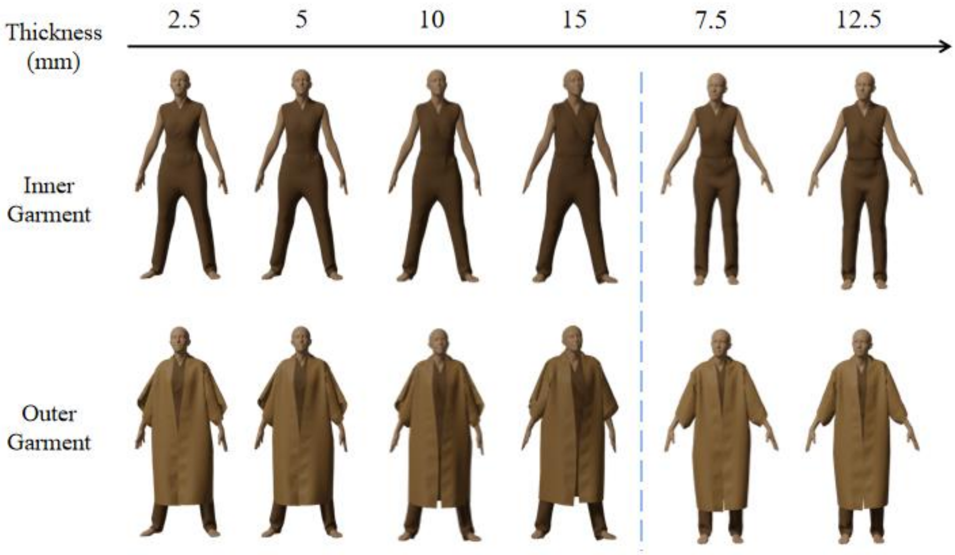

FIGURE 6 The result of garment deformation under different thickness combinations.

In Fig. 7, we present the garment deformation effects under motion postures, observing how changes in inner layer thickness from different views influence the deformation of the outer layer. On the left side of Fig. 7, we show the garment deformation effects from a side view, while on the right side, we depict the effects from a front view. Our method demonstrates the capability to generate corresponding outer layer garment deformations in response to varying thicknesses of the inner layer. As the thickness of the inner layer increases, we observe a more pronounced expansion effect in the outer layer garment.

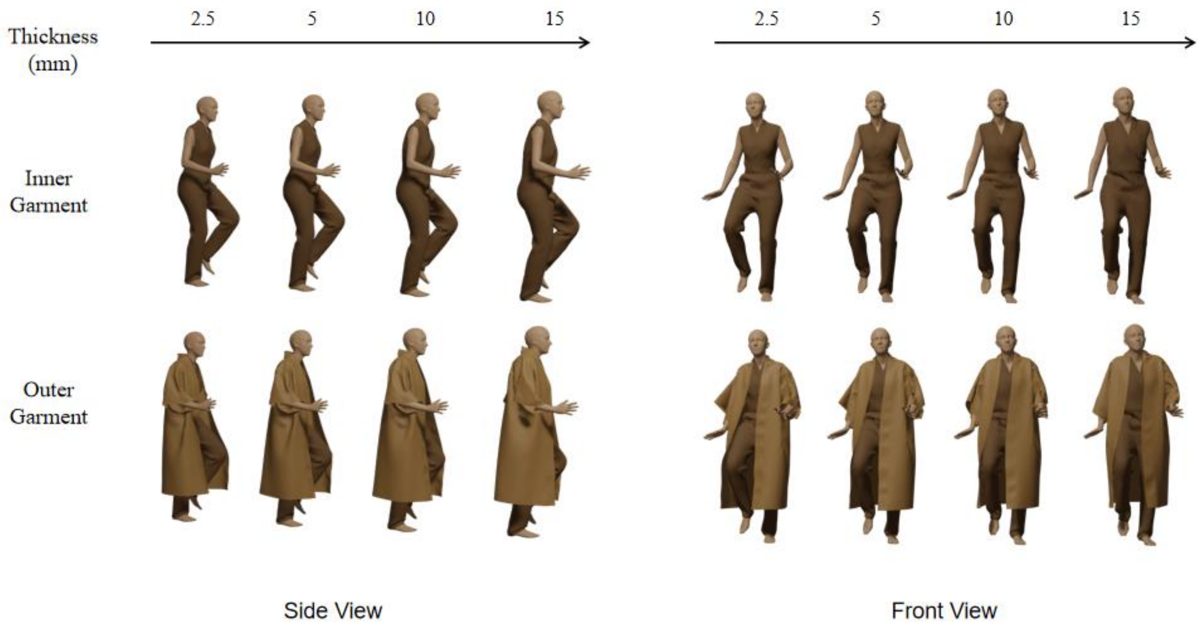

FIGURE 7 Results of thickness variation from different views under motion pose.

The detailed comparison results presented in Fig. 8 show close-up views of the inner and outer garments for the same thickness (each row) and different thicknesses (each column) under the same pose. It is evident that our method is capable of generating outer garment deformations corresponding to changes in inner garment thickness (including unseen cases), making it suitable for multi-layer garment animation generation for various combinations.

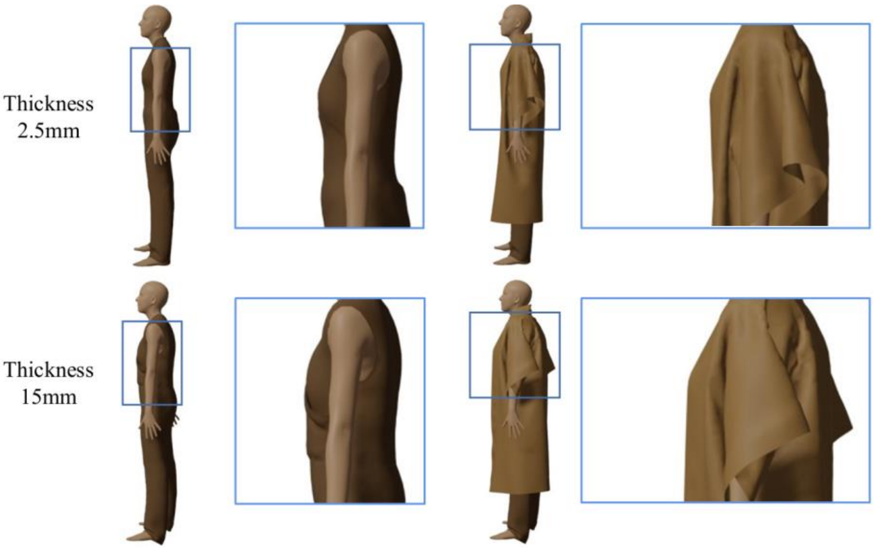

FIGURE 8 Results of Detailed Comparisons

Ablation Study

To validate the effectiveness of the inter-layer garment repulsive force loss between multiple garment layers, ablation experiments were conducted, and quantitative and qualitative analyses were performed.

Table 4. Quantitative results for different thicknesses.

Metrics	STED ↓	L-Coll ↓	H-Coll ↓
w/o $L_{repulsive}$	9.96	3.08	1.33
ours	**3.25**	**1.56**	**0.14**

The comparative results depicted in Table 5 and Fig. 10 reveal that the inclusion of the repulsive force loss term effectively mitigates unrealistic garment deformations. This results in a decrease in errors between the network model's predictions and the ground truth. Furthermore, there are notable enhancements in terms of L-Coll (garment layer penetration) and H-Coll (penetration between garments and the human body model).

In Fig. 10, the left side shows the results without the repulsion loss term, while the right side shows the results with the repulsion loss term added. From the figure, we can see that the repulsion loss term, by constraining the vector distance between adjacent layer vertices, significantly reduces the penetration loss and garment penetration rate between garment layers. From the magnified details in the figure, we can also observe that our method generates stable deformations that satisfy visual requirements for garment.

FIGURE 9 Examples of our MGN for dynamic garment deformation in motion.

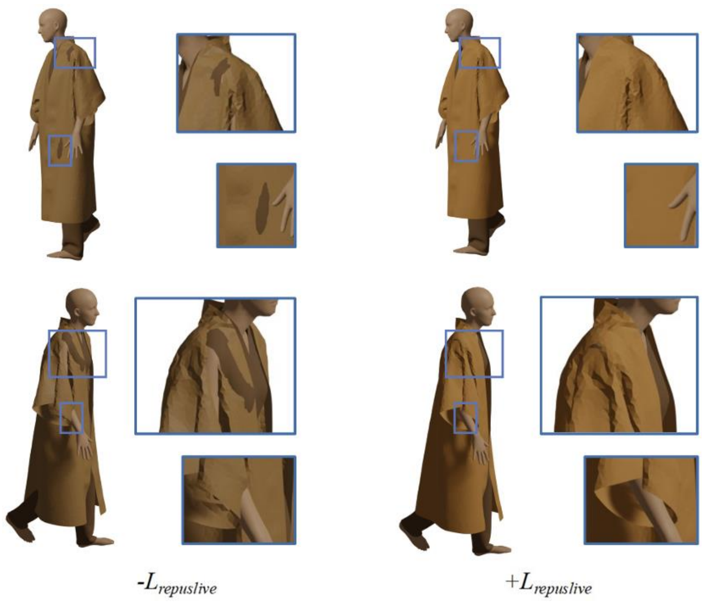

FIGURE 10 Deformation results of the ablation study for garment deformation. The leftmost column presents the garment generation results without the inter-layer garment repulsive force loss, and the second column shows the garment generation results with the repulsive force loss added.

Sample dress animation

The method presented in this paper is well-suited for predicting multi-layer garment deformation and generating animations. We have selected various types of human motion postures and used the method described in this paper for deformation simulation and animation generation. The results are demonstrated in Fig. 9, which displays animations of virtual characters dressed in different pose sequences. Please refer to the supplementary video for visual comparisons of dynamic deformation results.

After generating the garment deformation sequences using our method, the results can be imported into professional software like Blender for texture adjustments, producing more visually appealing garment animations.

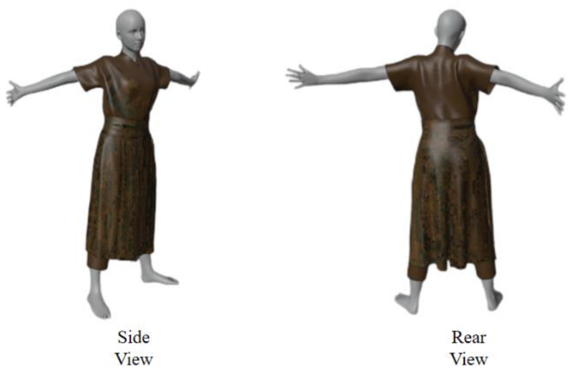

FIGURE 11 Garment animation with textures added using Blender.

As shown in Fig. 11, the garment deformations generated by our method are compatible with the existing animation production pipeline in Blender for rendering. From the figure, it can be seen that the garments with added textures offer a more visually appealing experience.

Conclusions

This paper conducts research on the generation of multi-layer garment animations by decomposing the prediction of multi-layer garment deformations into single-layer garment deformations and inter-layer garment interactions. Through an unsupervised training strategy, it formulates target loss functions for the network modules based on kinematic equations in accordance with physical laws and geometric constraints, thereby reducing the preparation time for garment deformation data. By employing a temporally supervised training strategy, the model learns collision contacts and temporal deformation features between multi-layer garments, fusing unsupervised single-layer garment deformations and outer garment deformation features for training, leading to the learning of deformation patterns among multi-layer garments.

As a result, the model generates multi-layer garment animations that satisfy temporal pose constraints. The experiments conducted in the paper validate the effectiveness of the proposed method and showcase its superiority in terms of garment animation quality, penetration loss, and other aspects when compared to other existing state-of-the-art methods.

Limitations: Our Transformer-based two-stage network can rely on our unsupervised single-layer generation module or be compatible with existing single-layer clothing works to reduce data dependency, but it still cannot completely eliminate the need for clothing sample data. Exploring a self-supervised multi-layer clothing animation generation method to completely eliminate dependency on clothing deformation datasets is part of our future work plan.

Acknowledgements

Acknowledgements will be added after the completion of the review process to maintain anonymity.

References

[1] Provot X, others. Deformation constraints in a mass-spring model to describe rigid cloth behaviour. In: *Graphics interface*. Canadian Information Processing Society, 1995, pp. 147–147.

[2] Rohmer D, Popa T, Cani M-P, et al. Animation wrinkling: augmenting coarse cloth simulations with realistic-looking wrinkles. *ACM Transactions on Graphics (ToG)* 2010; 29: 1–8.

[3] Wu L, Wu B, Yang Y, et al. A safe and fast repulsion method for GPU-based cloth self collisions. *ACM Transactions on Graphics (TOG)* 2020; 40: 1–18.

[4] Wang TY, Shao T, Fu K, et al. Learning an intrinsic garment space for interactive authoring of garment animation. *ACM Transactions on Graphics (TOG)* 2019; 38: 1–12.

[5] Santesteban I, Otaduy MA, Casas D. Learning-based animation of clothing for virtual try-on. In: *Computer graphics forum*. 2, Wiley Online Library, 2019, pp. 355–366.

[6] Patel C, Liao Z, Pons-Moll G. Tailornet: Predicting clothing in 3d as a function of human pose, shape and garment style. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 7365–7375.

[7] Santesteban I, Otaduy M, Thuerey N, et al. ULNeF: untangled layered neural fields for mix-and-match virtual try-on. *Adv Neur In* 2022; 35: 12110–12125.

[8] Li R, Guillard B, Fua P. Isp: Multi-layered garment draping with implicit sewing patterns. *Adv Neur In*; 36.

[9] Lee D, Kang H, Lee I-K. ClothCombo: Modeling inter-cloth interaction for draping multi-layered clothes. *ACM Transactions on Graphics (TOG)* 2023; 42: 1–13.

[10] Shao Y, Loy CC, Dai B. Towards multi-layered 3D garments animation. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2023, pp. 14361–14370.

[11] Li C, Tang M, Tong R, et al. P-cloth: interactive complex cloth simulation on multi-GPU systems using dynamic matrix assembly and pipelined implicit integrators. *ACM Transactions on Graphics (TOG)* 2020; 39: 1–15.

[12] Müller M. Hierarchical position based dynamics.

[13] Müller M, Chentanez N, Kim T-Y, et al. Air meshes for robust collision handling. *ACM Transactions on Graphics (TOG)* 2015; 34: 1–9.

[14] Müller M, Heidelberger B, Hennix M, et al. Position based dynamics. *J Visual Commun Image Represent* 2007; 18: 109–118.

[15] Vassilev T, Spanlang B, Chrysanthou Y. Fast cloth animation on walking avatars. In: *Computer graphics forum*. 3, Wiley Online Library, 2001, pp. 260–267.

[16] Tang M, Wang H, Tang L, et al. CAMA: Contact-aware matrix assembly with unified collision handling for GPU-based cloth simulation. In: *Computer graphics forum*. 2, Wiley Online Library, 2016, pp. 511–521.

[17] Daviet G, Bertails-Descoubes F, Casati R. Fast cloth simulation with implicit contact and exact coulomb friction. In: *Proceedings of the 14th ACM SIGGRAPH/eurographics symposium on computer animation*. 2015, pp. 197–197.

[18] Wang TY, Ceylan D, Popovic J, et al. Learning a shared shape space for multimodal garment design. *arXiv preprint arXiv:180611335*.

[19] Gundogdu E, Constantin V, Seifoddini A, et al. Garnet: A two-stream network for fast and accurate 3d cloth draping. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 8739–8748.

[20] Shen Y, Liang J, Lin MC. Gan-based garment generation using sewing pattern images. In: *Computer vision–ECCV 2020: 16th european conference, glasgow, UK, august 23–28, 2020, proceedings, part XVIII 16*. Springer, 2020, pp. 225–247.

[21] Tiwari G, Bhatnagar BL, Tung T, et al. Sizer: A dataset and model for parsing 3d clothing and learning size sensitive 3d clothing. In: *Computer vision–ECCV 2020: 16th european conference, glasgow, UK, august 23–28, 2020, proceedings, part III 16*. Springer, 2020, pp. 1–18.

[22] D. Li Y, Tang M, Yang Y, et al. N-cloth: Predicting 3D cloth deformation with mesh-based networks. In: *Computer graphics forum*. 2, Wiley Online Library, 2022, pp. 547–558.

[23] Xu W, Umetani N, Chao Q, et al. Sensitivity-optimized rigging for example-based real-time clothing synthesis. *ACM Trans Graph* 2014; 33: 107–1.

[24] Loper M, Mahmood N, Romero J, et al. SMPL: A skinned multi-person linear model. In: *Seminal graphics papers: Pushing the boundaries, volume 2*. 2023, pp. 851–866.

[25] Bertiche H, Madadi M, Escalera S. Pbns: Physically based neural simulator for unsupervised garment pose space deformation. *arXiv preprint arXiv:201211310*.

[26] Santesteban I, Otaduy MA, Casas D. Snug: Self-supervised neural dynamic garments. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 8140–8150.

[27] Peng T, Wu W, Liu J, et al. PGN-Cloth: Physics-based graph network model for 3D cloth animation. *Displays* 2023; 80: 102534.

[28] Bertiche H, Madadi M, Escalera S. Neural cloth simulation. *ACM Transactions on Graphics (TOG)* 2022; 41: 1–14.

[29] Grigorev A, Black MJ, Hilliges O. HOOD: Hierarchical graphs for generalized modelling of clothing dynamics. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023, pp. 16965–16974.

[30] Luigi LD, Li R, Guillard B, et al. DrapeNet: Garment generation and self-supervised draping. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 2022; 1451–1460.

[31] Tiwari L, Bhowmick B, Sinha S. GenSim: Unsupervised generic garment simulator. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 2023; 4169–4178.

[32] Bhatnagar BL, Tiwari G, Theobalt C, et al. Multi-garment net: Learning to dress 3d people from images. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 5420–5430.

[33] Zhang M, Ceylan D, Mitra NJ. Motion guided deep dynamic 3D garments. *ACM Transactions on Graphics (TOG)* 2022; 41: 1–12.

[34] Bertiche H, Madadi M, Tylson E, et al. DeePSD: Automatic deep skinning and pose space deformation for 3D garment animation. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 5471–5480.

[35] Mahmood N, Ghorbani N, Troje NF, et al. AMASS: Archive of motion capture as surface shapes. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 5442–5451.

[36] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advances in neural information processing systems*; 30.

[37] Zhao F, Li Z, Huang S, et al. Learning anchor transformations for 3d garment animation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023, pp. 491–500.