Learning Anchor Transformations for 3D Garment Animation

Fang Zhao[1], Zekun Li[1], Shaoli Huang[1][*], Junwu Weng[1], Tianfei Zhou[2],

Guo-Sen Xie[3], Jue Wang[1], Ying Shan[1]

[1]Tencent AI Lab [2]ETH Zurich [3]Nanjing University of Science and Technology

Abstract

*This paper proposes an **anchor**-based **def**ormation model, namely **AnchorDEF**, to predict 3D garment animation from a body motion sequence. It deforms a garment mesh template by a mixture of rigid transformations with extra nonlinear displacements. A set of anchors around the mesh surface is introduced to guide the learning of rigid transformation matrices. Once the anchor transformations are found, per-vertex nonlinear displacements of the garment template can be regressed in a canonical space, which reduces the complexity of deformation space learning. By explicitly constraining the transformed anchors to satisfy the consistencies of position, normal and direction, the physical meaning of learned anchor transformations in space is guaranteed for better generalization. Furthermore, an adaptive anchor updating is proposed to optimize the anchor position by being aware of local mesh topology for learning representative anchor transformations. Qualitative and quantitative experiments on different types of garments demonstrate that AnchorDEF achieves the state-of-the-art performance on 3D garment deformation prediction in motion, especially for loose-fitting garments.*

1. Introduction

Animating 3D garments has a wide range of applications in 3D content generation, digital humans, virtual try-on, video games, and so on. Existing pipelines of 3D garment animation usually rely on physics based simulation (PBS), which requires a large amount of computational resources and time costs, particularly for high-quality PBS methods.

Some data-driven or learning-based methods have been proposed to quickly produce 3D garment deformation from static poses or motion sequences with low computational complexity [4,5,9,11,12,16,25–28,31,40]. However, many of them attach garment templates to the skeleton of human body for modeling the articulation of garments, which

[*]Corresponding author.

Figure 1. 3D garment deformation predicted by the proposed AnchorDEF with body motions. Leveraging the anchor transformations, AnchorDEF is able to realistically deform the garment mesh, especially for loose-fitting garments, *e.g.*, dresses.

only work with tight garments, *e.g.*, T-shirts and pants, and poorly address loose-fitting ones, *e.g.*, dresses and skirts. In this case, the topology of garments is different from the human body. Therefore, using the skinning blend weights of body yields discontinuities on deformed garment mesh. Several methods [26, 34] smooth out the blend weights to alleviate the discontinuities but may lose the shape details for large deformations of loose-fitting garments, which do not closely follow body movements.

To this end, we propose an **anchor**-based **def**ormation model, namely **AnchorDEF**, for predicting 3D garment deformation from a body motion sequence. It leverages a mixture of rigid anchor transformations to represent principle rotations and translations of garments to detach the

garment articulation from the body skeleton while preserving the body movement prior, then nonlinear displacements can be regressed relatively easily in a canonical space. As shown in Fig. 1, our method can exploit anchor transformations to realistically deform the garment mesh, especially for loose-fitting garments, *e.g.*, dresses.

Specifically, given a sequence of body motions including poses and translations, we first estimate rigid transformations of a set of anchors around the garment mesh. Using the linear blending skinning (LBS), the garment mesh template is deformed by a weighted combination of the anchor transformations meanwhile per-vertex displacements of the mesh template are regressed to correct artifacts of the blended rigid transformations. To learn physically meaningful anchor transformations for better generalization, we enforce the transformed anchors to maintain consistency with the target's position and normal. In addition, a relative direction constraint is employed to reduce garment-body interpenetrations, which is efficient due to the sparseness of anchors. To make the learned anchor transformations effectively represent the garment deformation, an adaptive anchor updating is further introduced to utilize mesh simplification as supervision to optimize the anchor position. It parameterizes the position with a local attention mask on adjacent mesh vertices and pushes the anchors towards folds and boundaries of garment mesh which usually determine the way of deformation.

The main contributions of our work can be summarized as follows: 1) We propose an anchor-based deformation model which learns a set of anchor transformations and blend weights in a unified framework to represent the deformation of 3D garments, especially for loose-fitting ones. 2) We propose to learn anchor transformations by position and normal consistencies as well as relative direction constraint for better generalization and fewer garment-body interpenetrations. 3) We introduce an adaptive anchor updating with the mesh simplification as supervision to optimize the anchor position for learning representative anchor transformations.

2. Related Work

Most garment animation and deformation works can be categorized into physics-based simulations (PBS) and learning-based models.

Physics-based simulations. PBS methods have been developed for decades and achieved impressive accuracy, leveraging laws of physics and stable accurate solvers [1,13,18], collision detection [23, 29, 35], and dry frictional contact [19, 21]. But the expensive computational costs, *i.e.*, long simulation time and high-performance devices often are unaffordable in our daily lives.

Learning-based models. In contrast, learning-based models, which utilize garment animation data attained from PBS

or digital scanning to learn the deformation, can often give satisfactory enough animation visual results with less computational costs [26, 31], regardless of no explicit physical guarantee on accuracy. Previously, researchers have proposed many techniques depending on the motions of human bodies with linear conditional models [8, 11] or secondary motion graphs [14, 15], with a constructed database of high-resolution meshes. Recent works [4, 26, 34] assume that the garment is a submesh of the skinned multi-person linear model (SMPL) and use linear blend skinning (LBS) to transfer body shape to cloth shape. Beyond learning static results within a single frame, adopting recurrent neural networks to handle the dynamic wrinkles of clothes is widely used [25, 31, 34]. For example, [31] prepares a database of PBS for dressed characters with some body shapes and motion and uses the database to learn a function of body shape and dynamics for modeling cloth drape and wrinkles. Although learning-based models can be efficient, their performance heavily relies on large amounts of high-quality PBS results or scanning data, which are usually costly [2,30,34]. Thus, some works [3, 32] propose self-supervised training methods to constrain the model with some sort of physical laws to alleviate this problem. Additionally, to ensure the generalization of simulation, combining several pretrained pivot motion networks can handle different unseen garment styles [26] and simulation parameters [25].

But most learning-based methods are limited to tight garments and not suitable for dresses or clothes with more freedom, due to their assumption that garments are complementary to SMPL, *i.e.*, the deformation is closely related to body shape, not the motion. To tackle the dynamics, some works propose a neural field to retain high-frequency details [39], handle challenging areas [22], tackle multi-layer garments [33], and warp garments from the canonical pose to the deformation shape [6]. Without utilizing body shape parameters to establish a garment-body aligned model, [37] learns an intrinsic cloth shape descriptor embedding space and using different motion parameters to guide garment latent can attain different shape representation. To eliminate the alignment between clothes dynamics and body vertices, VirtualBones [25] extracts a set of rigid bones from the garment animation and uses body motion to predict the bones' transformation to interpolate the garment shape. Instead of learning based on mesh, [38] uses a point transformer to predict point clouds of various outfits, for various human poses and shapes, which attains more topological flexibility and the ability to model clothing separately from the body.

VirtualBones [25] is the most related work to our method, which requires garment skeleton extraction with SSDR [17] as data preprocessing. Different from [25], we decompose the garment deformation by a set of adaptive anchors whose rigid transformations and blend weights are learned in a unified framework.
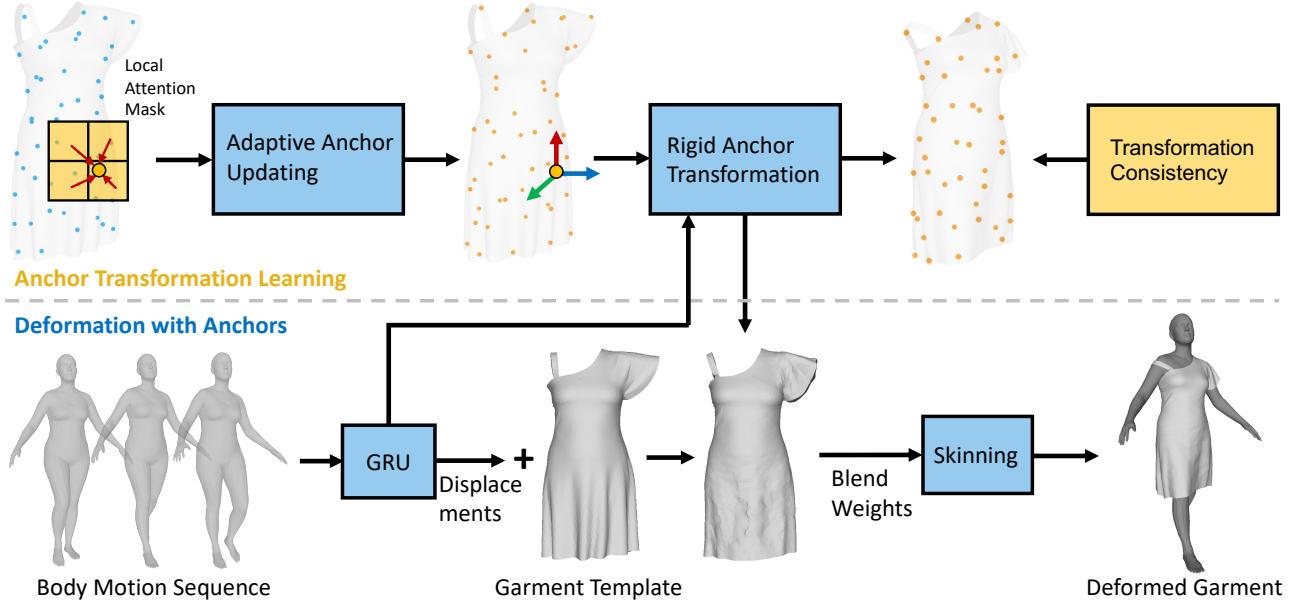
Figure 2. Overview of the proposed AnchorDEF. Given a body motion sequence and a garment mesh template, the garment deformation is predicted by a weighted average of rigid anchor transformations with per-vertex displacements in a canonical space. We explicitly constrain the transformed anchors to satisfy the consistencies of position, normal and direction to guarantee the physical meaning of learned anchor transformations in space for better generalization. Starting from the cluster centers of mesh vertices, the anchor position is updated by a local attention map on adjacent mesh vertices with mesh simplification as supervision. All parameters in our model can be jointly optimized in a unified framework.

3. Our Method

We aim to learn an animated model which is able to predict dynamic 3D garment deformations given a body motion sequence. In this section, we introduce **AnchorDEF**, an **anchor**-based **def**ormation model, which leverages s set of adaptive anchors around mesh surface to deform a garment mesh template by a mixture of rigid anchor transformations with additional per-vertex displacements in a canonical space, where the anchor transformations are learned by the consistencies of position, normal and direction. The overview of AnchorDEF is illustrated in Fig. 2.

3.1. Deformation with Anchors

3D animated models of human body, *e.g.*, SMPL [20], usually deform a body mesh template $\mathbf{T}$ with blendshapes and skinning, which can be defined as

$$V = W(\mathbf{T} + B(\theta), \mathbf{J}, \theta, \mathbf{w}), \tag{1}$$

where W is a skinning function, *e.g.*, linear blend skinning, with a skeleton $\mathbf{J}$, pose parameters θ and blend weights $\mathbf{w}$, V is the posed vertices and B is the blendshape corresponding to the body pose θ. Since the garment is a kind of soft body and it is hard to directly predict the coordinates of each vertex from a body motion sequence, here we also consider to decompose the garment deformation into two parts including a weighed average of rigid transformations and

nonlinear displacements.

Most data-driven garment models directly associate garments with the skeleton of human body [4, 26, 31], *i.e.*, they apply the skinning function to animate the garment by directly using the blend weights of body mesh vertices in the neighborhood, where it is assumed that garments closely follow the body movement. However, for loose-fitting garments, the clothing deformation usually contains more complex nonrigid effects compared to human body which can be seen as a rigid body. In this case, garments behave significantly differently from the body. Thus, it is unreasonable to attach the garment to the body skeleton.

To address this issue, we adopt a set of anchors $\{p_n\}$ around the mesh surface to guide the garment deformation, which detaches the garment from the body skeleton while preserving the body movement prior. Given a body motion sequence consisting of poses and translations, rigid anchor transformations including rotations $\{R_n\}$ and translations $\{T_n\}$ are estimated firstly by using Gated Recurrent Units (GRU) [7] (see the supplementary material for the network structure). Then, the m-th deformed garment vertex v_m can be represented by a weighted average of anchor transformations:

$$v_m = \sum_{n=1}^{N} w_{mn}(R_n v_m^c + T_n), w_{mn} \geq 0 \text{ and } \sum_{n=1}^{N} w_{mn} = 1, \tag{2}$$

where v_m^c is the m-th vertex of garment mesh template, w_{mn} is the blend weight of v_m regarding to the n-th anchor and N is the number of anchors. To determine the anchor-vertex relationship and maintain the sparsity of $\{w_n\}$, we initialize the anchors as the clustering centers of vertices of the mesh template. Then, each vertex is only influenced by the blend weights of its N' ($N' \ll N$) nearest anchors and the others remain zeros during training. Compared to the direct prediction of deformed mesh vertex coordinates, the rigid anchor transformations provide more generalized representations of complex deformations due to fewer parameters required to be learned for the prediction.

Two strategies of anchor transformation configuration are designed for loose and tight types of garments. For the loose type, to utilize the prior of the global human body orientation R_{global}, we predict transformations relative to the root joint J_{root} of the body and the anchor transformation can be defined as:

$$R_n^{loose} = R_{global}R_n([\theta, t]),$$
$$T_n^{loose} = R_{global}(T_n([\theta, t]) - R_n([\theta, t])J_{root}) + J_{root}, \tag{3}$$

where θ and t are the body pose and translation, respectively. For the tight type, the garment mostly follows the body movement. Thus, we predict an incremental change on the basis of body vertex rotation R_{body} and offset T_{body}:

$$R_n^{tight} = R_{body}R_n([\theta, t]),$$
$$T_n^{tight} = R_{body}T_n([\theta, t]) + T_{body}, \tag{4}$$

where R_{body} and T_{body} is the transformation of the body mesh vertex closest to the anchor p_n. Furthermore, we can combine these two strategies for garments with both tight and loose parts, e.g., dresses, by classifying which part the anchor belongs to according to its distance to body joints.

Once the primary rotations and translations in the garment deformation caused by the articulation are expressed in Eq. (2), the nonlinear residual deformation can be disentangled with the garment articulation and defined as per-vertex displacements $D([\theta, t])$ in an unposed canonical space. Utilizing the skinning function in Eq. (1), the overall garment deformation is given by

$$V_{pred} = W(\mathbf{T}_g + D([\theta, t]), G([\theta, t]), \mathbf{w}_p), \tag{5}$$

where $\mathbf{T}_g$ is the garment mesh template, $\mathbf{w}_p$ is the anchor-vertex blend weights and $G([\theta, t])$ is the anchor transformations predicted with the body motion. We argue that such decomposition by anchors extremely reduces the complexity of pose-dependent garment deformation prediction because large deformations have been handled by transformations of sparse anchors and only small displacements need to be regressed in the canonical space.
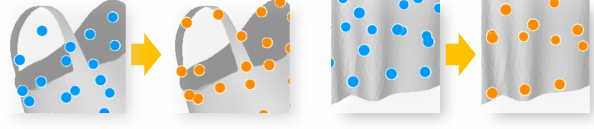

Figure 3. Examples of adaptive anchor updating by using the local attention mask with the mesh simplification as supervision. The blue points are anchors initialized by k-means and the orange ones are the updated anchors.

3.2. Learning Anchor Transformations

Different from [25] which resorts to the external skinning decomposition algorithm, i.e., SSDR [17], to extract the garment skeleton from smoothed mesh sequences, here we jointly learn rigid transformations, blend weights and positions of anchors in a unified framework.

We first apply an L2 loss term on the predicted mesh vertices to recover the shape of deformed mesh:

$$\mathcal{L}_{rec} = ||V_{pred} - V_{gt}||_2^2, \tag{6}$$

which minimizes the Euclidean distance between the predicted vertices and the ground-truth deformed vertices. Then a loss term is used to keep the similarity of the graph Laplacians between two meshes:

$$\mathcal{L}_{lap} = || \triangle V_{pred} - \triangle V_{gt}||_2^2. \tag{7}$$

A collision term [25] is also added to avoid garment-body collisions. Thus, the loss on the vertex can be given by

$$\mathcal{L}_{vert} = \mathcal{L}_{rec} + \beta_1 \mathcal{L}_{lap} + \beta_2 \mathcal{L}_{collision}, \tag{8}$$

where β_1 and β_2 are trade-off coefficients for controlling the degree of smoothing and collision, respectively.

However, only optimizing the vertex position cannot guarantee the physical meaning of learned anchor transformations in space, which leads to poor generalization. Thus, we explicitly constrain the transformed anchors to be located around the deformed mesh surface. To this end, the position of anchor p is recalculated by its adjacent mesh vertices:

$$p = \frac{1}{K} \sum_{v_k^c \in K(p, \mathbf{T}_g)} v_k^c, \tag{9}$$

where $K(p, \mathbf{T}_g)$ is the K nearest vertices of p on the mesh template. Then, the target point p_{tgt} of transformed anchor p_G can be defined with the same set of vertices $K(p, \mathbf{T}_g)$ on the deformed mesh and a transformation consistency loss is defined as

$$\mathcal{L}_{consis} = ||p_G - p_{tgt}||_2^2 + \gamma||\mathbf{n}(p_G) - \mathbf{n}(p_{tgt})||_2^2, \tag{10}$$

where $\mathbf{n}(.)$ is the unit normal of point and γ is a weight factor. By encouraging the transformed anchor p_G to be similar to p_{tgt} in position and normal, the learned anchor transformation indicates the mean transformation of K nearest

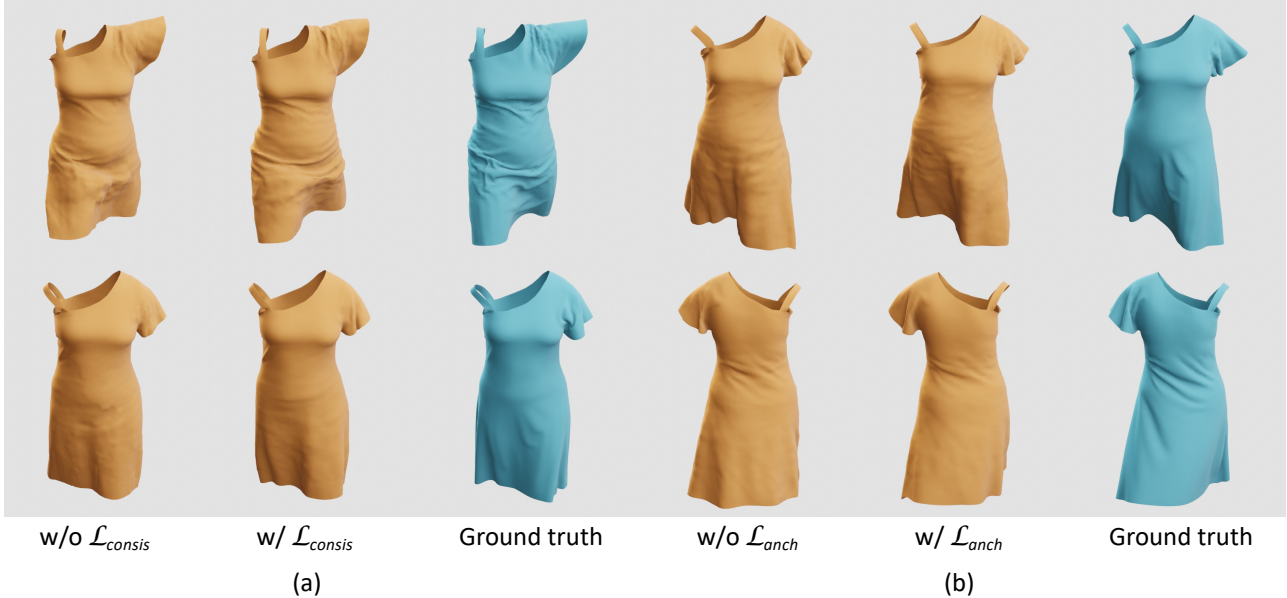| w/o $\mathcal{L}_{consis}$ | w/ $\mathcal{L}_{consis}$ | Ground truth | w/o $\mathcal{L}_{anch}$ | w/ $\mathcal{L}_{anch}$ | Ground truth |

(a)   (b)

Figure 4. Qualitative comparison with and without different components of AnchorDEF including (a) anchor transformation consistency $\mathcal{L}_{consis}$ and (b) adaptive anchor updating $\mathcal{L}_{anch}$. Without $\mathcal{L}_{consis}$, the model cannot accurately predict the shape of deformed mesh and loses some surface geometry details. Without $\mathcal{L}_{anch}$, the model is unable to produce realistic garment deformations due to the under-expressed anchor transformations.

vertices of p. It can be viewed as a regularization for the transformation magnitude and direction to discover essential transformations in deformation.

We further enforce the direction from p_{tgt} to p_G to be consistent with the direction of $\mathbf{n}(p_{tgt})$ by minimizing the following penalty term:

$$\mathcal{L}_{dir} = 1 - \cos(p_G - p_{tgt}, \mathbf{n}(p_{tgt})), \qquad (11)$$

where $\cos(\cdot, \cdot)$ is the cosine similarity. Intuitively, this penalty term makes p_G tend to be outside of the plane formed by p_{tgt} to reduce the garment-body interpenetration.

Adaptive Anchor Updating. It is still not enough to have a physically meaningful anchor transformation, we also need an expressive transformation in geometry to better represent the garment deformation. We observe that mesh regions with folds and boundaries usually play an important role on showing garment shape characteristics and contain more deformation information compared with smooth regions.

To make the anchors reflect such intrinsic structure of garment shape, we propose to employ mesh simplification to provide a signal of supervision for finding such representative anchors. Here Quadric Error Metric (QEM) [10] is used to obtain simplified mesh vertices by collapsing edges while preserving the garment topology. Because there is no one-to-one correspondence between the anchors $\{p_n\}$ and the simplified vertices $\{\tilde{v}_m\}$, the Chamfer distance is adopted as the optimization objective:

$$\mathcal{L}_{anch} = \sum_p \min_{\tilde{v}} ||p - \tilde{v}||_2^2 + \sum_{\tilde{v}} \min_p ||p - \tilde{v}||_2^2. \quad (12)$$

To be able to update the anchor position, we rewrite Eq. (9) using learnable weights α as

$$p = \frac{1}{\sum_{j=1}^K e^{\alpha_j}} \sum_{k=1}^K e^{\alpha_k} v_k^c. \qquad (13)$$

Actually, this equals to learning a local attention mask for the anchor to determine how the adjacent vertices contribute to its position. Fig. 3 shows that by optimizing the weights α with supervision of the simplified vertices, the anchors are pushed to inflection points of folds and boundaries of the mesh and thus able to better represent the deformed mesh vertices by interpolating their transformation matrices. Note that we optimize the anchor position on the entire dataset not only on the mesh template because some smooth regions of the template may create new folds after posing.

On the other hand, if we reformulate the transformed anchor p_G with Eq. (13) as:

$$p_G = Rp + T = \frac{1}{\sum_{j=1}^K e^{\alpha_j}} \sum_{k=1}^K e^{\alpha_k} (Rv_k^c + T), \quad (14)$$

the process of transformation learning for the anchor, *i.e.*, Eq. (10), can be interpreted as a way to find an appropriate transformation for its adjacent mesh vertices by the importance which is reflected in the weights learned from the mesh simplification. The larger the weight, the greater the impact of the vertex transformation on the anchor transformation learning.
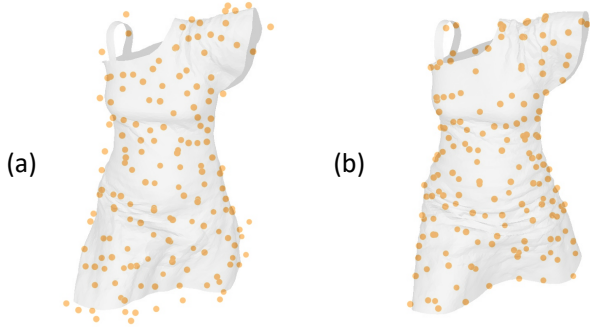
Figure 5. Visualization of anchors transformed by (a) only optimizing the mesh vertices during training and (b) AnchorDEF.

Lastly, combining Eq. (8), Eqs. (10) to (12), the overall objective function for our model is given by:

$$\mathcal{L} = \mathcal{L}_{vert} + \lambda_1 \mathcal{L}_{consis} + \lambda_2 \mathcal{L}_{dir} + \lambda_3 \mathcal{L}_{anch}, \quad (15)$$

where λs are loss weights.

4. Experiments

We evaluate the proposed AnchorDEF on the VTO [31] dataset containing two types of garments, *i.e.*, dress and T-shirt. Following [25], RMSE (Root Mean Squared Error), Hausdorff distance and STED (Spatio-Temporal Edge Difference) [36] are adopted to assess the quality of deformed meshes. The first two measure the distance between predicted and ground truth meshes and STED reflects the difference between their dynamics. We also compare AnchorDEF against other 3D garment deformation methods [25, 26].

Implementation Details. To capture the motion dynamics, a network with 2 GRU layers of size 256 is used as the body motion encoder. Two 2-layered MLPs of size 256 are used to regress the anchor transformations and per-vertex displacements, respectively. The number of anchors associated with each mesh vertex is set to 8, which can obtain the desired smooth surface, and the number of neighboring vertices for computing the anchor position is set to 128. For the garment sequences of VTO, 4 clips are randomly selected for testing and the rest is used as the training set, and the length of training sequence is 30. Please refer to the supplemental materials for more details on training and inference.

4.1. Ablation Study

We validate the effectiveness of main components in our proposed AnchorDEF. We firstly assess the importance of anchor transformation consistency, *i.e.*, $\mathcal{L}_{consis}$ in Eq. (10). As reported in Table 1, keeping the consistency of learned anchor transformations is able to improve all metrics due

Methods	w/o L_{consis}	w/o L_{anch}	AnchorDEF
RMSE ↓	17.21	16.49	**16.05**
Hausdorff ↓	75.01	74.53	**74.20**
STED ↓	0.0595	0.0526	**0.0493**

Table 1. Quantitative results obtained by different components of AnchorDEF, *i.e.*, without anchor transformation consistency (w/o $\mathcal{L}_{cosis}$), without adaptive anchor updating (w/o $\mathcal{L}_{anch}$) and the full model (AnchorDEF).

| w/o $\mathcal{L}_{dir}$ | w/ $\mathcal{L}_{dir}$ | w/o $\mathcal{L}_{dir}$ | w/ $\mathcal{L}_{dir}$ |

Figure 6. Qualitative comparison with and without the penalty $\mathcal{L}_{dir}$ on the direction from the target anchor to the transformed one.

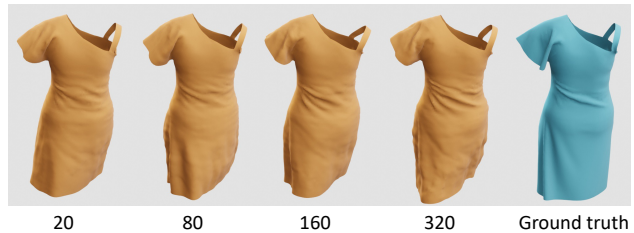| 20 | 80 | 160 | 320 | Ground truth |

Figure 7. Qualitative comparison of using different numbers of anchors. Using 160 anchors produces finer and more accurate deformation compared with other anchor numbers.

Num. Anchor	20	80	160	320
RMSE ↓	17.78	16.82	**16.05**	17.52
Hausdorff ↓	75.30	74.71	**74.20**	75.13
STED ↓	0.0620	0.0539	**0.0493**	0.0625

Table 2. Quantitative results of using different numbers of anchors.

to better generalization. Fig. 4 (a) illustrates the qualitative comparison. One can observe that without $\mathcal{L}_{consis}$ the model cannot accurately predict the shape of deformed mesh and loses some surface geometry details.

To evaluate the influence of adaptive anchor updating, Table 1 also reports the results without using $\mathcal{L}_{anch}$ in Eq. (12). The performance drops only using the clustering centers of k-means as the anchors. The reason is that compared with the clustering, the supervision with mesh simplification can provide key topology information, *e.g.*, folds and boundaries, which is useful for learning representative

Figure 8. Examples of our AnchorDEF for dynamic garment deformation in motion. The first row is the ground truth generated by physics-based simulation [24]. The second row is the results obtained by AnchorDEF. Our method can produce natural and realistic clothing dynamics, *e.g.*, wrinkles from jumping and the swing of the dress caused by running.

anchor transformations. The qualitative results are shown in Fig. 4 (b). It can be observed that without $\mathcal{L}_{anch}$ the model is unable to produce realistic garment deformations due to the under-expressed anchor transformations.

Fig. 5 (a) and (b) visualizes the anchors transformed by only optimizing the mesh vertices during training and AnchorDEF, respectively. For Fig. 5 (a), without the consistency constraint for anchor transformation, many of the transformed anchors are not on the deformed mesh surface, which is lack of clear physical meaning in space and easy to overfit. In contrast, AnchorDEF enforces the anchors to be transformed onto the deformed mesh surface, which prevents overfitting to a certain extent, thus has better generalization when testing. Furthermore, as shown in Fig. 5 (b), the transformed anchors are mainly distributed at the folds and boundaries of the mesh because of our adaptive anchor updating according to simplified mesh vertices, and thus capable of better expressing the deformation of mesh vertices, in particular those vertices far from them. In this way, the complexity of garment deformation space is reduced more effectively and more mesh shape details can be preserved.

Then we investigate the impact of the penalty $\mathcal{L}_{dir}$ (Eq. (11)) on the direction from the target anchor to the transformed one. As illustrated in Fig. 6, with $\mathcal{L}_{dir}$ the model can reduce the garment-body interpenetration. By enforcing the transformed anchors to be on the positive direction of target normals, the garment mesh vertices influenced by these anchors can be effectively pushed to the outside of the body surface while keeping the deformation accuracy due to moving along the direction of target normals.

We further assess the influence of the anchor number. Table 2 reports the results obtained by using 20, 80, 160 and 320 anchors, respectively. One can see that using more anchors to guide the mesh deformation obtains better results on all three metrics because the decomposition for the mesh deformation by using the rigid anchor transformations makes the nonlinear deformation can be more easily predicted as the small displacements in the canonical space without considering large rotations and translations caused by the body pose which has been handled by the anchor transformations. However, the performance drops as the number of anchors continues to increases because too many transformation parameters lead to unstable forecasts during testing. Fig. 7 shows the qualitative comparison. Using 160 anchors produces finer and more accurate deformation compared with other anchor numbers. Using 20 anchors only recover the contour of the deformed mesh, and using 80 and 320 anchors fail to obtain realistic deformation.

Fig. 8 shows some examples of our AnchorDEF for dynamic garment deformation in motion. Given a body motion sequence, our method can produce natural and realistic clothing dynamics, *e.g.*, wrinkles from jumping and the swing of the dress caused by running.

4.2. Comparison with Related Methods

We compare our method against related 3D garment deformation methods, including TailorNet [26] and Virtual-Bones [25] on different types of garments. As reported in Table 3, our AnchorDEF achieves the best results in all cases. Fig. 9 presents the qualitative comparison. Tailor-

Methods	Dress			T-shirt		
	RMSE ↓	Hausdorff ↓	STED ↓	RMSE ↓	Hausdorff ↓	STED ↓
TailorNet [26]	22.95	76.80	0.0757	9.90	27.02	0.0418
VirtualBones [25]	19.91	83.39	0.0722	10.52	31.51	0.0452
AnchorDEF	**16.05**	**74.20**	**0.0493**	**6.25**	**26.31**	**0.0262**

Table 3. Quantitative results of 3D garment deformation methods for different types of garments.

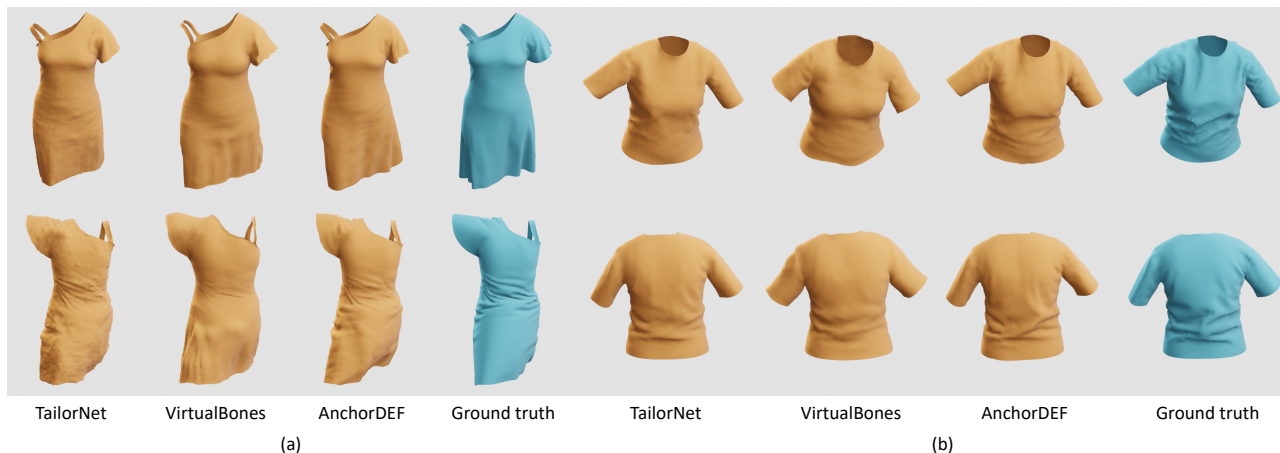| TailorNet | VirtualBones | AnchorDEF | Ground truth | TailorNet | VirtualBones | AnchorDEF | Ground truth |

(a)   (b)

Figure 9. Qualitative comparison of 3D garment deformation methods for different types of garments, including TailorNet [26], Virtual-Bones [25] and Our AnchorDEF. With the learned anchor transformations, our method not only recovers large deformations of the dress but also produces fine surface details on the T-shirt.

Net [26] recovers folds on the T-shirt, but it fails to address the deformation of the dress which has different topology from human body. By interpolating the skinning blend weights, it alleviates the discontinuity of blend weights borrowed from the SMPL [20] body. However, this makes TailorNet only produce the outline of shapes with some folds missing and artifacts for large deformations of loose-fitting garments. VirtualBones [25] can estimate large deformations of the dress by using SSDR [17] to extract virtual bones of garments. However, it tends to keep the original folds of the dress template on the deformed mesh and cannot estimate well new folds generated by body motions on the T-shirt. VirtualBones needs SSDR to extract the garment skeleton and gets blend weights of bones in advance and fixes the bone position and blend weights during network training. Moreover, SSDR is performed on the low-frequency mesh sequences processed by Laplacian smoothing because it fails to address high-frequency garment deformations. Thus, the learned virtual bones are not necessarily optimal. In contrast, by jointly learning the rigid transformations, blend weights and positions of anchors, our AnchorDEF not only recovers large deformations of the dress but also produces fine surface details on the T-shirt. More qualitative results can be found in the supplemental materials.

5. Conclusions

We present AnchorDEF, an anchor based deformation model, to predict 3D garment animation given a body motion sequence. With a set of anchors around the mesh surface, the garment deformation can be decomposed into a mixture of rigid anchor transformations and extra per-vertex displacements in a canonical space. The transformation consistency learning is introduced to guarantee the physical meaning of learned anchor transformations in space. Taking mesh simplification as supervision, We further optimize the anchor position for learning representative anchor transformations by making the anchor aware of local mesh topology with an attention mask. Extensive experiments on different types of garments demonstrate that AnchorDEF can effectively predict 3D garment deformation in motion for both tight and loose-fitting garments.

Limitations. Compared with the PBS methods, our method still lacks fine-level dynamics. Better exploiting the temporal information for the network architecture and supervision signals is a good direction for future work.

References

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

[2] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. Cloth3d: clothed 3d humans. In *Proceedings of the European conference on computer vision (ECCV)*, 2020. 2

[3] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. Pbns: Physically based neural simulator for unsupervised garment pose space deformation. *arXiv preprint arXiv:2012.11310*, 2020. 2

[4] Hugo Bertiche, Meysam Madadi, Emilio Tylson, and Sergio Escalera. Deepsd: Automatic deep skinning and pose space deformation for 3d garment animation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 1, 2, 3

[5] Lan Chen, Juntao Ye, Liguo Jiang, Chengcheng Ma, Zhanglin Cheng, and Xiaopeng Zhang. Synthesizing cloth wrinkles by cnn-based geometry image superresolution. *Computer Animation and Virtual Worlds*, 2018. 1

[6] Cheng Chi and Shuran Song. Garmentnets: Category-level pose estimation for garments via canonical space shape completion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 2

[7] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014. 3

[8] Edilson De Aguiar, Leonid Sigal, Adrien Treuille, and Jessica K Hodgins. Stable spaces for real-time clothing. *ACM Transactions on Graphics (TOG)*, 2010. 2

[9] Wei-Wen Feng, Yizhou Yu, and Byung-Uck Kim. A deformation transformer for real-time cloth animation. *ACM transactions on graphics (TOG)*, 2010. 1

[10] Michael Garland and Paul S Heckbert. Surface simplification using quadric error metrics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 1997. 5

[11] Peng Guan, Loretta Reiss, David A Hirshberg, Alexander Weiss, and Michael J Black. Drape: Dressing any person. *ACM Transactions on Graphics (ToG)*, 2012. 1, 2

[12] Erhan Gundogdu, Victor Constantin, Amrollah Seifoddini, Minh Dang, Mathieu Salzmann, and Pascal Fua. Garnet: A two-stream network for fast and accurate 3d cloth draping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 1

[13] David Harmon, Etienne Vouga, Breannan Smith, Rasmus Tamstorf, and Eitan Grinspun. Asynchronous contact mechanics. In *ACM SIGGRAPH 2009 papers*. 2009. 2

[14] Doyub Kim, Woojong Koh, Rahul Narain, Kayvon Fatahalian, Adrien Treuille, and James F O'Brien. Near-exhaustive precomputation of secondary cloth effects. *ACM Transactions on Graphics (TOG)*, 2013. 2

[15] Tae-Yong Kim and Eugene Vendrovsky. Drivenshape: a data-driven approach for shape deformation. In *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 49–55, 2008. 2

[16] Zorah Lahner, Daniel Cremers, and Tony Tung. Deepwrinkles: Accurate and realistic clothing modeling. In *Proceedings of the European conference on computer vision (ECCV)*, 2018. 1

[17] Binh Huy Le and Zhigang Deng. Smooth skinning decomposition with rigid bones. *ACM Transactions on Graphics (TOG)*, 2012. 2, 4, 8

[18] Cheng Li, Min Tang, Ruofeng Tong, Ming Cai, Jieyi Zhao, and Dinesh Manocha. P-Cloth: Interactive cloth simulation on multi-GPU systems using dynamic matrix assembly and pipelined implicit integrators. *ACM Transaction on Graphics (Proceedings of SIGGRAPH Asia)*, 2020. 2

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

[20] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 2015. 3, 8

[21] Mickaël Ly, Jean Jouve, Laurence Boissieux, and Florence Bertails-Descoubes. Projective dynamics with dry frictional contact. *ACM Transactions on Graphics (TOG)*, 2020. 2

[22] Qianli Ma, Jinlong Yang, Michael J. Black, and Siyu Tang. Neural point-based shape modeling of humans in challenging clothing. In *2022 International Conference on 3D Vision (3DV)*, 2022. 2

[23] Matthias Müller, Nuttapong Chentanez, Tae-Yong Kim, and Miles Macklin. Air meshes for robust collision handling. *ACM Transactions on Graphics (TOG)*, 2015. 2

[24] Rahul Narain, Armin Samii, and James F O'brien. Adaptive anisotropic remeshing for cloth simulation. *ACM transactions on graphics (TOG)*, 2012. 7

[25] Xiaoyu Pan, Jiaming Mai, Xinwei Jiang, Dongxue Tang, Jingxiang Li, Tianjia Shao, Kun Zhou, Xiaogang Jin, and Dinesh Manocha. Predicting loose-fitting garment deformations using bone-driven motion networks. In *ACM SIGGRAPH Conference Proceedings*, 2022. 1, 2, 4, 6, 7, 8

[26] Chaitanya Patel, Zhouyingcheng Liao, and Gerard Pons-Moll. 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 (CVPR)*, 2020. 1, 2, 3, 6, 7, 8

[27] Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations*, 2021. 1

[28] Gerard Pons-Moll, Sergi Pujades, Sonny Hu, and Michael J Black. Clothcap: Seamless 4d clothing capture and retargeting. *ACM Transactions on Graphics (ToG)*, 2017. 1

[29] Xavier Provot. Collision and self-collision handling in cloth model dedicated to design garments. In *Computer Animation and Simulation '97*, pages 177–189. Springer, 1997. 2

[30] Albert Pumarola, Jordi Sanchez-Riera, Gary Choi, Alberto Sanfeliu, and Francesc Moreno-Noguer. 3dpeople: Modeling the geometry of dressed humans. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2242–2251, 2019. 2

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

[32] Igor Santesteban, Miguel A Otaduy, and Dan Casas. SNUG: Self-Supervised Neural Dynamic Garments. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

[33] Igor Santesteban, Miguel A. Otaduy, Nils Thuerey, and Dan Casas. ULNeF: Untangled layered neural fields for mix-and-match virtual try-on. In *Advances in Neural Information Processing Systems, (NeurIPS)*, 2022. 2

[34] Igor Santesteban, Nils Thuerey, Miguel A Otaduy, and Dan Casas. Self-supervised collision handling via generative 3d garment models for virtual try-on. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1, 2

[35] Min Tang, Huamin Wang, Le Tang, Ruofeng Tong, and Dinesh Manocha. CAMA: Contact-aware matrix assembly with unified collision handling for GPU-based cloth simulation. *Computer Graphics Forum (Proceedings of Eurographics 2016)*, 35(2):511–521, 2016. 2

[36] Libor Vasa and Vaclav Skala. A perception correlated comparison method for dynamic meshes. *IEEE transactions on visualization and computer graphics (TVCG)*, 2010. 6

[37] Tuanfeng Y Wang, Tianjia Shao, Kai Fu, and Niloy J Mitra. Learning an intrinsic garment space for interactive authoring of garment animation. *ACM Transactions on Graphics (TOG)*, 2019. 2

[38] Ilya Zakharkin, Kirill Mazur, Artur Grigorev, and Victor Lempitsky. Point-based modeling of human clothing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 2

[39] Meng Zhang, Tuanfeng Y Wang, Duygu Ceylan, and Niloy J Mitra. Dynamic neural garments. *ACM Transactions on Graphics (TOG)*, 2021. 2

[40] Fang Zhao, Wenhao Wang, Shengcai Liao, and Ling Shao. Learning anchored unsigned distance functions with gradient direction alignment for single-view garment reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 1