EUROGRAPHICS 2026 / B. Masia and J. Thies
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 45 (2026), Number 2

Skeletal-Driven Animation of Anatomical Humans via Neural Deformation Gradients

G. Nolte[1,3†] F. Kemper[1,3†] U. Schwanecke[2] M. Botsch[1,3]

[1]Computer Graphics Group, TU Dortmund University, Germany
[2]Computer Vision and Mixed Reality Group, RheinMain University of Applied Sciences, Wiesbaden, Germany
[3]Lamarr Institute for Machine Learning and Artificial Intelligence, Dortmund, Germany

Figure 1: *Five frames of a volumetric anatomical human animated with our neural deformation gradients method. The left and right models include all the anatomical details (skeleton in yellow, muscles in red, and skin in blue), but only the surface of the skin is visible because no protrusions occur. The second models to the left and right show only the skeleton and muscles, while the center shows only the skeleton.*

Abstract

Most real-time animation techniques for digital humans are limited to deforming the outer skin surface. Geometric skinning methods are highly efficient but struggle with artifacts such as collapsing joints or self-intersections when animating inner anatomy along with the outer skin. Volumetric physics-based simulations, on the other hand, naturally resolve these issues by coordinating bones, muscles, and skin, but are far too slow for interactive use.

We solve this problem by training a neural network to predict deformation gradients. Learning deformation gradients instead of vertex displacements makes our method naturally robust to artifacts such as element inversion or volume deviation. Our model, trained on high-quality finite element simulations, generalizes well across diverse body shapes and poses. This enables anatomically consistent and physically grounded animation of bones, muscles, and skin at interactive frame rates.

CCS Concepts

• ***Computing methodologies*** → ***Physical simulation; Neural networks; Volumetric models; Mesh models;***

1. Introduction

Plausible animation of virtual humans is a central topic for computer games, movie production, and social VR applications, to name just a few. Interactive animation techniques tend to show only the outer skin surface and either ignore the inner anatomy or approximate its effects on the skin using learned models. However, access to the underlying anatomy, such as bones and muscles, offers many benefits, for example, anatomical visualization for educational purposes, sports science, or entertainment. Moreover, it provides physically grounded volumetric information that can be

† The first two authors contributed equally

used to compute internal stresses, more accurately handle collisions, or infer necessary muscle activations [HCO*24].

Geometric skinning methods are unfit to handle anatomical deformation: Errors that are acceptable in a surface deformation can easily lead to obvious artifacts where anatomy protrudes the outer skin. Existing anatomical models such as SKEL [KWS*23] and HIT [KAD*24] rely on such computationally efficient but simplistic posing models. As a result, they frequently produce penetration artifacts, as shown in Section 5. Physics-based simulation methods can handle the anatomical structure and produce impressive results, but typically require multiple seconds or even minutes of optimization per frame [KE20], making them unsuitable for interactive use.

Targeting this gap of interactive, anatomical animation, we present *Neural Deformation Gradients* (NDG), a physics-inspired neural network approach for quasi-static animation of volumetric human bodies. NDG is trained to reproduce FEM-simulated deformations computed on a large dataset of various bodies and poses. By embedding detailed anatomical meshes into a volumetric mesh, NDG enables interactive, coordinated animation of bones, muscles, and outer skin while mitigating interpenetration artifacts. Unlike vertex-based neural methods, NDG predicts per-element deformation gradients, a more physically meaningful representation. Our experiments show that this representation leads to improved volume preservation, reduced element inversions, and robust generalization, allowing NDG to animate novel body shapes and poses without retraining. By choosing an efficient prism-based approximation of deformation gradients, NDG can run at 30 frames per second on consumer-level hardware while closely approximating FEM results. This enables anatomically consistent human animation suitable for interactive applications.

Some existing approaches learn deformations of surfaces represented by deformation gradients [AGK*22; QSA*23; LWK*24], but target deformations of individual surfaces only. To our knowledge, our approach is the first that learns deformation gradients in a volumetric setting, targeting consistent multi-layered deformations.

The full source code and model can be downloaded at github.com/fabiankemper/NeuralDeformationGradients.

2. Related Work

We aim at interactive, rig-based volumetric character animations that produce physically plausible and consistent deformations. We specifically target stable, global volumetric deformations that are applicable to a variety of body shapes and poses.

Surface-Based Skinning. Classic rig-based techniques such as Linear Blend Skinning (LBS) [MLT88] and Dual Quaternion Skinning (DQS) [KCŽO07; KCŽO08] are computationally efficient but ignore volumetric structure, producing artifacts like joint collapse and bulging. Several extensions reduce these issues, e.g., Delta Mush Skinning [MDRW14; LL19] or optimized centers of rotation [LH16], but they remain heuristic and limited to surface deformation. Lacking awareness of the underlying anatomy, these methods cannot enforce consistency between bones, muscles, and skin, producing volume loss and anatomy that penetrates the skin.

Data-Driven Surface Deformation. Data-driven extensions improve geometric skinning by predicting per-vertex corrective displacements. Pose-Space Deformation linearly interpolates example-based offsets [LCF00], while more recently, neural methods are used to learn nonlinear correctives for LBS [CO18; HCO*24; LAH*21; BODO18; XBZ*20; SGOC20; CPH*25]. The widely used SMPL model [LMR*15] augments LBS with learned pose-correctives. These works remain focused on surface deformation: Even works that animate multiple layered surfaces, such as Han et al. [HCO*24], rely on non-volumetric, surface-based learning methods and hence are prone to anatomical artifacts, especially when generalizing to new body shapes and poses (see Section 5).

Gradient-Based Neural Surface Deformations. Only a few approaches learn surface deformations in deformation-gradient space. Gradient-based deformation representations have been used to explore the deformation space [TGLX18; GYQ*18] or for facial animation [WBS23]. Aigerman et al. [AGK*22] introduced Neural Jacobian Fields (NJF) and showed that diverse functions, represented by their gradient fields, can be learned by neural networks, including surface deformations of humans. Li et al. [LWK*24] use NJFs to learn garment deformations, while Qin et al. [QSA*23] do so for facial animation. The aim of these methods is to predict surface deformations that are agnostic to the underlying triangulation. Our method shares the mathematical foundation of these methods, i.e., deformation gradients as a learnable representation of deformation and Poisson solves to obtain concrete meshes. Instead of focusing on mesh-agnostic surface deformations, our method explicitly targets physically plausible, consistent deformation of volumes.

Volumetric Modeling of Humans. Physically meaningful animation of human bodies requires a volumetric model of the human body. As the inner structure of humans cannot usually be observed, different body models have been designed to represent the shape space of human anatomy [KZBP22; KAD*24; KWS*23; WKSB24; KWB21; KIL*16; SZK15]. By fitting such a model to a given surface scan, the position of inner anatomy can be estimated. Most of these works focus on estimating the position of the skeleton. Only some explicitly model the muscle structure [KAD*24; KWB21; KIL*16; SZK15]. SKEL and HIT incorporate pose models based on SMPL [LMR*15], allowing the body shapes to be deformed in a rig-based manner. This method is simple and fast, but frequently produces anatomy that protrudes through the outer skin, highlighting the need for an efficient method for consistent volumetric deformation.

Volumetric Simulation. Physics-based methods model volumetric deformations with (hyper-)elastic energies using the Finite Element Method (FEM) [KC71]. Such simulations yield highly accurate and consistent deformations of human tissue [MZS*11; SGK18] and can accurately model highly detailed anatomy [KIL*16]. This comes at the cost of solving highly nonlinear optimization problems that are prohibitively expensive for real-time use. To meet the requirements of interactive applications, simplified physics-based methods have been employed. Capell et al. [CBC*05] use a co-rotated form of linear elasticity; projective skinning [KB18; KB19] utilizes a projective dynamics [BML*14] approach, while others [DB13; AF15] use position-based dynam-

ics [MHHR07]. Hybrid approaches have also been explored, utilizing data driven methods to deform inner layers while applying simplified physics-based deformation methods for the remaining outer layers [KPP*17; ROCP20; TRPO21]. These methods are fast but rely on a twofold simplification: They operate on non-anatomical models and either rely on simplified energies or heuristic, position-based formulations. NDG avoids these simplifications: Training on FEM-simulated data allows it to produce deformations of similar quality as full FEM simulations at interactive frame rates.

Neural Physics Simulation. Neural methods have been used to accelerate physics simulations of elastic materials or to infer material parameters from data [ZZG19; ZZCB21; DHG23; CG23]. Inferring material parameters and rest shape with neural methods can allow for realistic animations of body and faces [KIL*16; KK19; YZC*24]. Zheng et al. [ZZCB21] employ a neural network to approximate implicit Euler updates in dynamic simulations of volumetric human flesh. Neural methods in a simulation context typically target dynamic simulation with small time steps, enabling the use of neural network models that process only local patches of the mesh at a time. However, the local nature of these models does not translate well to global quasi-static deformations that we target.

Our Method. To obtain a consistent representation of different human body shapes, we employ a volumetric body model derived from InsideHumans [KWB21], enabling us to obtain compatibly tesselated, volumetric body models of different bodies. This allows our neural method to extend to a variety of body shapes without retraining (see Subsection 3.1). Training data is generated from high-quality FEM simulation of volumetric muscle and fat layers, while the pose is controlled by rig-based geometric skinning of the bones (Subsection 3.2). Deformations are transferred from a volumetric simulation mesh to high-resolution visualization meshes of the inner anatomy by an embedded deformation approach (Subsection 3.3). To approximate the FEM-based results at interactive speeds, we train a neural network, which we discuss in Section 4. To ensure volumetrically consistent deformations, our model combines two ideas from surface-based approaches: our network learns correctives for an efficient deformation method, which is standard in vertex-based neural approaches [HCO*24; LAH*21], but does so in deformation gradient space, extending related surface-based approaches to the volumetric setting [AGK*22]. We show in Section 5 that this physically meaningful representation enables NDG to outperform vertex-based neural approaches and interactive, physics-based methods in a volumetric setting, considerably improving w.r.t. volume preservation, inversion prevention, and generalization to novel body shapes. In addition, we perform ablation studies to evaluate architecture choices.

3. Volumetric Anatomical Simulation

Our approach is built on a volumetric body model that we use to represent human anatomy, an FEM-based deformation that provides target training data for our neural network, and an embedded deformation for transforming the detailed anatomy.

Figure 2: *Overview of the template meshes used in our method. The left (male) illustrates the wrap surface meshes: skin* $\mathbf{S}$ *(blue), muscle wrap* $\mathbf{M}$ *(red), and bone wrap* $\mathbf{B}$ *(yellow). The right (female) shows the skin* $\mathbf{S}$ *(blue) and the corresponding high-resolution anatomical meshes: muscles* $\mathcal{M}$ *(red) and bones* $\mathcal{B}$ *(yellow). The skin, muscle wrap, and bone wrap surfaces serve as boundaries enclosing the high-resolution meshes in volumetric layers.*

3.1. Layered Anatomical Human Model

We base our animation method on a volumetric body model consisting of volumetrically connected layers, a concept that has been successfully applied to animation [KWB21; DB13]. Our model is built from a skin surface $\mathbf{S}$ together with high-resolution bone and muscle meshes, $\mathcal{B}$ and $\mathcal{M}$. These are enclosed by wrap meshes, $\mathbf{B}$ for the bones and $\mathbf{M}$ for muscles (see Figure 2), which share the tessellation of $\mathbf{S}$. The skin and both wraps contain 24 k vertices each, while $\mathcal{B}$ and $\mathcal{M}$ contain 37 k and 97 k vertices, respectively.

The skin, muscle wrap, and bone wrap surfaces define a layered volume by connecting corresponding triangles across adjacent surfaces. This volume consists of two layers: the muscle layer $\mathbb{M}$ connects the bone wrap and muscle wrap, and the subcutaneous fat layer $\mathbb{S}$ connects the muscle wrap and skin. By construction, the elements in these volumetric layers are sheared triangular prisms. We refer to these simply as *prisms*. We tetrahedralize the enclosed volume of the bone wrap $\mathbf{B}$ using TetGen [Si06], ensuring that the entire volume is covered, yielding a tetrahedral bone mesh $\mathbb{B}$. The muscle layer $\mathbb{M}$ and fat layer $\mathbb{S}$ will be animated using soft-body simulation with the bone wrap $\mathbf{B}$ controlling the pose. Because the head, hands, and toes contain little volume, we model them as surfaces rather than volumetric regions and animate them using LBS.

After excluding the head, hands, and toes, the volumetric layers $\mathbb{B}$, $\mathbb{M}$, and $\mathbb{S}$ contain $N_p \approx 53\,\text{k}$ prisms and $N_v \approx 40\,\text{k}$ vertices. This template model can be fitted to human skin surfaces using a similar approach to Komaritzan et al. [KWB21], providing us with access to anatomical models of a variety of human bodies. We denote the combined vertices of the simulation mesh, including the bone wrap $\mathbf{B}$, the muscle wrap $\mathbf{M}$, and the skin $\mathbf{S}$ as

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\mathsf{T} \in \mathbb{R}^{N_v \times 3}. \tag{1}$$

and use $\bar{\mathbf{X}}$ to explicitly refer to its undeformed state.

3.2. FEM-Based Simulation

To produce physically plausible ground-truth deformations, we employ an FEM-based simulation of quasi-static elastic deformation.

The volumetric muscle and fat layers, $\mathbb{M}$, and $\mathbb{S}$, are modeled using the stable Neo-Hookean energy without the barrier term [SGK18], as recommended by Kim and Eberle [KE20]:

$$\Psi_{\text{flesh}}(\mathbf{X}) = \sum_{p=1}^{N_p} \bar{V}_p \psi_{\text{flesh}}\left(\mathbf{F}_p(\mathbf{X})\right), \tag{2}$$

$$\psi_{\text{flesh}}(\mathbf{F}) = \frac{\mu}{2}\left(\|\mathbf{F}\|^2 - 3\right) + \frac{\lambda}{2}\left(\det\mathbf{F} - \alpha\right)^2. \tag{3}$$

The variable $\alpha = 1 + \frac{\mu}{\lambda}$ is a constant chosen to ensure stability in volume-preserving configurations as proposed in [SGK18], $\|\cdot\|$ denotes the Frobenius matrix norm, μ and λ are Lamé material parameters, and $\bar{V}_p$ is the volume of the prism p in the rest pose.

The deformation gradient $\mathbf{F}_p$ is a per-prism approximation of the continuous deformation gradient

$$\mathbf{F}(\mathbf{x}) = \nabla\phi(\mathbf{x}), \tag{4}$$

where $\phi\colon \mathbb{R}^3 \to \mathbb{R}^3$ is the deformation function mapping undeformed to deformed body shapes. We approximate the deformation gradient as the linear part of the affine mapping that best maps each undeformed prism to its deformed state, integrated over the prism interior (see supplementary material for its construction). Splitting each prism into (at least) three tetrahedra would allow us to use standard linear elements with constant gradients. However, as a representation for neural network learning, the prism-based representation is advantageous: While the deformation gradients are approximations within individual prisms, they collectively still act as an exact descriptor of global deformation, allowing for reconstruction of meshes with almost machine precision. At the same time, using per-tetrahedron gradients would triple the amount of information to be learned and potentially introduce poorly shaped tetrahedra. Similar approximations of the deformation gradient have been successfully used for efficient elastic simulations [MZS*11] and related deep learning methods [TGLX18; GYQ*18].

For the outer skin $\mathbf{S}$, we apply the discrete shell energies presented by Grinspun et al. [GHDS03] to the N_e edges of the skin:

$$\Psi_{\text{bend}}(\mathbf{X}) = \frac{1}{2}\sum_{e=1}^{N_e} \frac{\bar{l}_e^2}{\bar{A}_e}\left(\theta_e - \bar{\theta}_e\right)^2, \tag{5}$$

$$\Psi_{\text{stretch}}(\mathbf{X}) = \frac{1}{2}\sum_{e=1}^{N_e} \frac{1}{\bar{l}_e^2}\left(l_e - \bar{l}_e\right)^2. \tag{6}$$

Here, $\theta_e, \bar{\theta}_e$, and $l_e, \bar{l}_e$ represent the dihedral angles and lengths of edge e in the deformed and undeformed mesh, respectively, while $\bar{A}_e$ is a third of the rest area of the two triangles incident to e. These energies yield smooth, inextensible skin surfaces and realistic skin folds. Combining the volumetric and surface energies using stiffness parameters k_f, k_b, and k_s leads to the total energy

$$\Psi(\mathbf{X}) = k_f\Psi_{\text{flesh}}(\mathbf{X}) + k_b\Psi_{\text{bend}}(\mathbf{X}) + k_s\Psi_{\text{stretch}}(\mathbf{X}). \tag{7}$$

The simulation is driven by temporally varying pose, represented as joint angles of the animation rig. We use LBS to deform the bone wrap $\mathbf{B}$. The vertices of the muscle wrap $\mathbf{M}$ and skin $\mathbf{S}$ are determined by solving for a minimum energy state of $\Psi(\mathbf{X})$, with the vertices on $\mathbf{B}$ acting as hard Dirichlet constraints. This is similar to the layered animation method of Bender et al. [DB13]. We solve for quasi-static equilibria using Newton optimization following the principles outlined by Teran et al. [TSIF05].

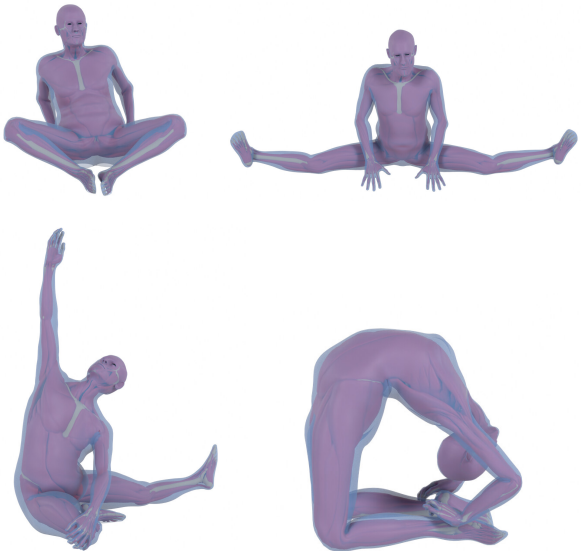


Figure 3: *Four subjects generated by our FEM-based simulation, which drives the deformation of the high-resolution musculature and skeletal anatomy, $\mathcal{B}$ and $\mathcal{M}$. The images show the resulting deformed anatomy, where the inner structures remain enclosed by the outer skin mesh,* $\mathbf{S}$.

We let $k_f = 1$ and choose $k_s = 3 \cdot 10^{-2}$ and $k_b = 2 \cdot 10^{-3}$ for the outer skin energies, letting the flesh energy dominate, while the skin energies act as visual regularizers. We choose Lamé parameters μ and λ from Poisson's ratios and Young's moduli, using the conversion formulas detailed in [LLAF20]: We use Young's modulus $E_{\text{muscle}} = 45\,\text{kPa}$ for the muscle layer and $E_{\text{fat}} = 15\,\text{kPa}$ for the subcutaneous fat layer, which is in the range of human tissue behavior [AMG*13; CAD*12]. We choose a high Poisson's ratio of 0.47 for both muscle and fat. This reflects the highly volume-preserving nature of human tissue [SGK18] and, with $\lambda \approx 16\mu$, lets the volume-preserving term $(\det\mathbf{F} - \alpha)^2$ dominate the simulation.

3.3. Embedding High-Resolution Anatomy

Given a deformation of the volumetric layers, $\mathbb{M}$ and $\mathbb{S}$, we aim to transfer that deformation to the high-resolution anatomical meshes $\mathcal{B}$ and $\mathcal{M}$. For this problem, Komaritzan et al. [KWB21] propose an RBF warp. However, we found this RBF warp to be inefficient: with few kernels, the warp produced poor deformations, causing volume changes and protrusions in the anatomical meshes, as discussed in more detail in the supplemental material.

Instead, we embed the high-resolution meshes into the volumetric layers and interpolate their positions at runtime [MTG04; ZZCB21; SCSG18]. Our volumetric muscle layer $\mathbb{M}$ and fat layer

$\mathbb{S}$ consist of prisms, while the interior of the bone wrap $\mathbf{B}$ was tetrahedralized as $\mathbb{B}$. For each high-resolution vertex of $\mathcal{B}$ and $\mathcal{M}$, we find either its enclosing prism (if contained within $\mathbb{M}$ or $\mathbb{S}$) or enclosing tetrahedron (if contained within $\mathbb{B}$). Then, we express that vertex as a barycentric combination of the vertices of its enclosing element. For tetrahedra, we use standard barycentric coordinates. For prisms, we interpolate based on the FEM shape functions for triangular prisms [ZTZ05] (see supplementary material). This embedding strategy enables inference of the high-resolution anatomy in approximately 10 ms on an Intel Core i7-12700K CPU, which is 60 times faster than an RBF warp with 5000 centers. We found the high number of centers necessary to achieve comparable performance to the barycentric warp (see supplementary material). We show results of our FEM-based animation with embedded deformation of anatomical details in Figure 3.

4. Neural Deformation Gradients

The FEM-based animation produces physically plausible results that preserve volume well and mitigate inversions. However, runtimes are dominated by nonlinear optimization, requiring multiple seconds per frame, as is typical for FEM-based animation [KE20]. We aim to approximate the high-quality FEM-based animation results with a neural method.

The typical neural approach is to predict pose-corrective vertex offsets [HCO*24; BODO18; LAH*21]. Applying this method directly to layered volumetric models often causes inversions of the volumetric layers (see Section 5). We therefore employ a neural network that predicts correctives as deformation gradients rather than as vertex displacements. Deformation gradients provide a more robust representation: in thin elements, a small displacement in vertex space can induce a non-physical inversion, leading to elements with negative volume and anatomy protruding the outer skin. In deformation gradient space, the same configuration would correctly incur a large error. This makes deformation gradients a robust basis for physically meaningful, inversion-resistant deformations [KE20], motivating them as the representation of choice for our Neural Deformation Gradients (NDG).

Our pipeline begins by simulating a dataset of diverse poses, which provides the training data for our model. Body shapes are encoded using principal component analysis (PCA) that was trained on volumetric body models fitted to the CAESAR dataset [RDP99]. Poses are represented by joint rotations. A neural network is then trained to predict residual deformation gradients from those inputs. Finally, vertex positions are reconstructed by solving a Poisson system that enforces consistency with the predicted gradients. An overview of this process is shown in Figure 4.

4.1. Training Data

To train NDG, we generate a dataset of diverse volumetric human bodies in diverse poses using the FEM-based simulation described in Subsection 3.2.

Volumetric Body Shapes. We use a set of about 1500 different male and female subjects from the European subset of the CAESAR dataset [RDP99], to which we fit our template surface model

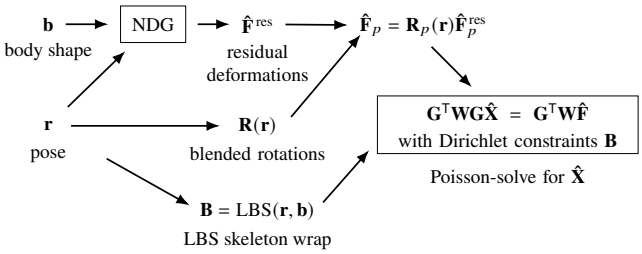


Figure 4: *Computation pipeline for our NDG deformation model. The network takes as input the body shape parameters* $\mathbf{b}$ *and joint angles* $\mathbf{r}$*, and predicts residual deformations* $\hat{\mathbf{F}}^{\text{res}}$ *batched over all volumetric elements. The final deformation gradient for each prism* p *is then given by* $\hat{\mathbf{F}}_p = \mathbf{R}_p(\mathbf{r})\hat{\mathbf{F}}_p^{\text{res}}$*, where* $\mathbf{R}_p(\mathbf{r})$ *is the quaternion-blended pose-induced rotation. The muscle wrap* $\mathbf{M}$ *and skin* $\mathbf{S}$ *are subsequently reconstructed by solving a Poisson system, using the vertex positions of the linear blend skinned bone wrap* $\mathbf{B}$ *as Dirichlet boundary conditions.*

to obtain compatibly tessellated skin surfaces using the approach of Achenbach et al. [AWLB17]. From these skin surfaces, we obtain compatible anatomical meshes $\mathcal{B}$ and $\mathcal{M}$ as well as wrap surfaces $\mathbf{B}$ and $\mathbf{M}$ using a modified version of InsideHumans [KWB21].

If the rest shape contains degenerate or inverted elements, deformation gradients become unstable and lose physical meaning, posing substantial hurdles for neural network training. We observed that naively applying both the non-rigid surface fitting as well as the InsideHumans methodology for inner anatomy can lead to degenerate prisms, especially around the armpits in A-pose, where scan data is often unreliable and drift on one layer can lead to prisms with excessive shear.

In the surface-fitting pipeline of Achenbach et al. [AWLB17] we omit the PCA step from the optimization and increase the weight of the Laplacian regularizer by a factor of 10^4 around the armpits, to avoid degenerated elements. We modify InsideHumans to treat the skin vertices as degrees of freedom, allowing it to better control prism quality. The skin vertices are attached to their target positions by a quadratic spring energy and the skin surface is regularized using the Laplacian Term also used in surface fitting.

Pose Data. To obtain diverse and realistic human poses for training, we re-targeted existing motion capture sequences from the PosePrior [AB15], EyeDatasetJapan [Ltd], MOYO [TMH*23], and Human4D [CSB*20] subsets of the AMASS dataset [MGT*19]. These animation sequences were re-targeted to our custom animation skeleton using the damped least-squares inverse kinematics approach [Wam07], ensuring compatibility with the template of our multi-layered volumetric body model. To curate a compact yet diverse set of poses, we apply farthest-point sampling [Gon85] in pose space using as distance metric the sum, over joints, of the angular differences between different poses. This discourages clusters of near-duplicate poses and improves coverage of the pose space; in practice, it reduces redundancy and allows us to construct a final dataset of 15 000 diverse and representative poses.

Dataset Generation. To avoid data leakage, we split both poses and bodies into strictly distinct training, validation, and test sets (80 %, 10 %, 10 %). We randomly sample unique pairs of bodies and poses. For each pair, we run our FEM-based animation, producing a ground-truth deformation. We represent each ground-truth deformation of an undeformed mesh $\bar{\mathbf{X}}$ into $\mathbf{X}$ as the stacked deformation gradients $\mathbf{F} \in \mathbb{R}^{3N_p \times 3}$ of all N_p prisms in the muscle and fat layers, $\mathbb{M}$ and $\mathbb{S}$. As we use a linear gradient operator $\mathbf{G}$ that is constant within each prism (see supplementary material for more details), the stacked deformation gradients can be obtained as:

$$\mathbf{F} = \left[\mathbf{F}_1, \ldots, \mathbf{F}_{N_p}\right]^\mathsf{T} = \mathbf{G}\mathbf{X}. \tag{8}$$

Through this data generation process, we obtain training, validation, and test sets with strictly disjoint body shapes and poses and 20 k, 2.5 k, and 2.5 k samples each. Data generation took approximately 40 h on an Intel Core i7-12700K CPU.

4.2. Mesh Reconstruction from Deformation Gradients

To reconstruct mesh vertex positions from predicted deformation gradients $\hat{\mathbf{F}}$, we find the deformed mesh $\hat{\mathbf{X}}$ that best matches these gradients across all elements. This requires solving a weighted least-squares problem [SP04; BSPG06]:

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \left\| \mathbf{W}^{\frac{1}{2}} \left(\hat{\mathbf{F}} - \mathbf{G}\mathbf{X}\right) \right\|^2, \tag{9}$$

where $\mathbf{W}$ is a diagonal weighting matrix, containing the prism rest volumes. The normal equations for this optimization problem yield the Poisson system:

$$\left(\mathbf{G}^\mathsf{T}\mathbf{W}\mathbf{G}\right)\hat{\mathbf{X}} = \mathbf{G}^\mathsf{T}\mathbf{W}\hat{\mathbf{F}}. \tag{10}$$

The gradient operator $\mathbf{G}$ is highly sparse and depends on the undeformed mesh only, allowing the matrix to be pre-factored, requiring only right-hand side construction and back-substitution during animation, which is viable for interactive animation. As done in the FEM simulation, we use LBS to deform the bone wrap $\mathbf{B}$ and use its vertices as hard Dirichlet constraints for the system in Equation 10, solving for the vertices in the muscle wrap $\mathbf{M}$ and skin $\mathbf{S}$, which span the volumetric layers $\mathbb{M}$ and $\mathbb{S}$.

4.3. Residual Deformations

The rigid motion induced by bones is easy to model analytically, but imposes additional information to learn for a neural network. For this reason, vertex-based neural methods typically model them with LBS and learn only correctives, usually in the rest pose [LAH*21; HCO*24]. We follow this approach and factor out the bone-induced rigid rotation, such that NDG learns non-rigid corrective deformations in the rest pose.

We estimate the bone-induced rotation at each prism p by blending up to eight adjacent bone rotations, represented as unit quaternions $\bar{\mathbf{q}}_b$, using skinning weights $w_{p,b}$ as in DQS [KCŽO07]:

$$\bar{\mathbf{q}}_p = \sum_{b=1}^{8} w_{p,b}\,\bar{\mathbf{q}}_b\,, \qquad \mathbf{q}_p = \frac{\bar{\mathbf{q}}_p}{\|\bar{\mathbf{q}}_p\|}\,. \tag{11}$$

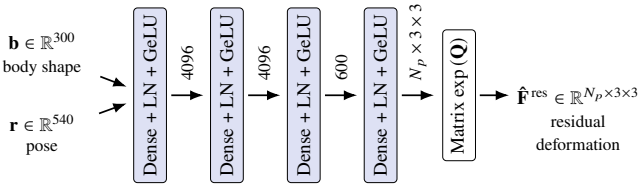


Figure 5: *Architecture of our NDG model. Vectorized rotation matrices for each joint (60 joints) and PCA encoding of the rest body shape act as input for the first block. Each block consists of a dense linear layer, and uses layer normalization and the GeLU as activation function. After the final block, the matrix exponential is applied to each deformation matrix to get the final residual deformations.*

Quaternion antipodality is resolved by aligning signs before blending, as is standard in DQS. From $\mathbf{q}_p$ we obtain the equivalent rotation matrix $\mathbf{R}_p$. Similar to SCAPE [ASK*05], we decompose each deformation gradient as

$$\mathbf{F}_p = \mathbf{R}_p \mathbf{F}_p^{\text{res}}\,, \tag{12}$$

where $\mathbf{F}_p^{\text{res}} = \mathbf{R}_p^\mathsf{T}\mathbf{F}_p$ is the residual deformation defined in the rest-pose. We use $\mathbf{F}^{\text{res}}$ to refer to the stacked matrix of all residual deformations. Intuitively, $\mathbf{R}_p$ describes the rigid bone motion, while $\mathbf{F}_p^{\text{res}}$ captures the remaining deformation in the rest pose. Note that $\mathbf{F}_p^{\text{res}}$ is not generally rotation free. Extracting the exact rotational part of $\mathbf{F}_p$ is possible with polar decomposition, but problematic: Unlike our approximate choice of $\mathbf{R}_p$, the exact rotational part of $\mathbf{F}_p$ cannot be easily computed just from the pose at inference time. Our choice is akin to the co-rotational method of Capell et al. [CBC*05], which also estimates bone motion through quaternion blending and then factors out the rotation during force computation.

4.4. The NDG Model

We now introduce the architecture of our neural model, NDG, and the chosen representations of its inputs and outputs.

Pose and Body Shape Representations. Poses are represented by a vector $\mathbf{r} \in \mathbb{R}^{9N_j}$ obtained by flattening and stacking the local 3×3 joint rotation matrices of all $N_j = 60$ skeletal joints. Representing body shapes as vectors of stacked vertex positions would increase the parameter count drastically. To reduce dimensionality, we compute a PCA over all training body shapes, defining a linear mapping

$$\text{PCA} : \mathbb{R}^{N_v \times 3} \to \mathbb{R}^d\,, \tag{13}$$

with $d = 300$. This dimensionality retains 99.992 % of the variance in the training data, losing almost no information while keeping dimensionality small. Each rest body shape $\bar{\mathbf{X}}$ is then represented by its PCA encoding $\text{PCA}(\bar{\mathbf{X}})$. Formally, NDG is then a function

$$\text{NDG} : \mathbb{R}^{N_j \times 3 \times 3} \times \mathbb{R}^d \to \mathbb{R}^{N_p \times 3 \times 3}\,, \tag{14}$$

mapping a pose $\mathbf{r}$ and a body shape $\mathbf{b}$ to predicted residual deformation gradients $\hat{\mathbf{F}}^{\text{res}} \in \mathbb{R}^{N_p \times 3 \times 3}$.

Architecture. NDG is implemented as a multilayer perceptron (MLP) with three hidden layers of sizes 4096, 4096, and 600, each

using LayerNorm and GeLU activations (see Figure 5). The input pose and body shape vectors are concatenated and passed through the network.

We use the matrix exponential as a final activation function to ensure that the determinants of predicted matrices are positive, thereby preventing our network from predicting physically implausible elements with negative volume, which would lead to anatomical protrusion artifacts. The predicted deformation gradient for prism p is therefore given as

$$\hat{\mathbf{F}}_p^{\text{res}} = \exp\left(\mathbf{Q}_p\right), \tag{15}$$

where $\mathbf{Q}_p$ is a 3×3 matrix predicted by the last trainable layer of the neural network. The positive determinant of $\hat{\mathbf{F}}_p^{\text{res}}$ follows from properties of the matrix exponential [Hab18], since

$$\det\left(\exp(\mathbf{Q}_p)\right) = \exp\left(\operatorname{tr}\left(\mathbf{Q}_p\right)\right) > 0\,. \tag{16}$$

We experimented with QR- and LU-factorizations (using positive diagonals in the triangular matrices) as alternative parameterizations of matrices with positive determinants. While all methods performed similarly, the exponential proved most robust to learn, required the minimal parameter count of 9 per 3×3 matrix, and benefits from an efficient polynomial approximation available in PyTorch [BBC19]. This leaves the Poisson solve in Subsection 4.2 as the only part of our pipeline that can produce inverted elements. However, inversions induce a large change in the deformation gradient, which is highly penalized by the objective function (see Equation 17).

The described network configuration was found to be most effective in our experiments: larger models with more parameters per layer overfit more quickly without improving generalization, and deeper networks consistently showed reduced performance.

4.5. Training

The training loss is the mean squared error between the predicted and simulated residuals, summed over all N_t training examples:

$$\sum_{i=1}^{N_t} \frac{1}{9N_p} \left\| \mathbf{F}^{\text{res}}\left(\mathbf{r}_i, \bar{\mathbf{X}}_i\right) - \text{NDG}\left(\mathbf{r}_i, \text{PCA}\left(\bar{\mathbf{X}}_i\right)\right)\right\|^2. \tag{17}$$

We train the network using batches of 32 examples for 102 epochs, at which point the validation loss plateaus. The optimization is performed with AdamW [KB15; LH19] and a learning rate of $1.3 \cdot 10^{-3}$. The final model achieves a training loss of $6.472 \cdot 10^{-4}$, showing no signs of overfitting. Training was completed in approximately 2 h on an NVIDIA RTX 6000. We show some animation results produced by NDG on diverse body shapes in Figure 6.

5. Evaluation

We evaluate our method in terms of accuracy, physical plausibility, and generalization to diverse body shapes and poses. We first introduce the quantitative metrics used for comparison and then benchmark NDG against both neural and physically-based baselines. Our experiments indicate that NDG achieves improved volume preservation and inversion avoidance, and, when compared to a vertex-based neural approach, achieves higher fidelity, while generalizing more robustly to out-of-distribution body shapes.

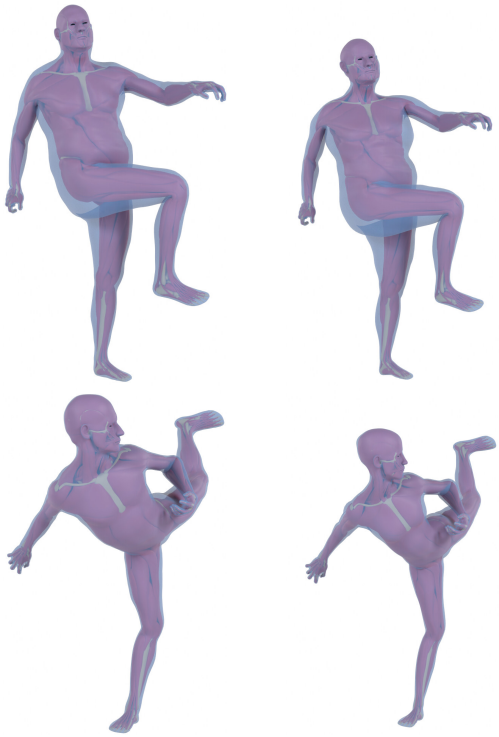


Figure 6: *Animation using NDG for diverse bodies. Top row: higher BMI (> 30); Bottom row: lower BMI (< 23). Left column: male; Right column: female. The outer skin layer is rendered in transparent blue, the musculature in red, and the skeleton in yellow.*

5.1. Evaluation Metrics

To assess the performance of our method, we use four key metrics. The first two metrics quantify how close a mesh $\hat{\mathbf{X}}$ is to the FEM reference $\mathbf{X}$, while the last two assess physical plausibility.

The first metric measures geometric accuracy in vertex space. We compute the average Euclidean distance (in millimeters) of each vertex to its target on the FEM-simulated mesh (excluding bone wrap vertices, which act as hard Dirichlet constraints)

$$E_{\text{pos}}\left(\mathbf{X}, \hat{\mathbf{X}}\right) = \frac{1}{N_v} \sum_{v=1}^{N_v} \left\|\mathbf{x}_v - \hat{\mathbf{x}}_v\right\|. \tag{18}$$

The second metric evaluates gradient-based accuracy as the mean squared error of deformation gradients

$$E_{\text{grad}}\left(\mathbf{X}, \hat{\mathbf{X}}\right) = \frac{1}{9N_p} \left\|\mathbf{G}\mathbf{X} - \mathbf{G}\hat{\mathbf{X}}\right\|^2. \tag{19}$$

For physical plausibility, we focus on volume preservation and inversion-freeness. As human tissue is nearly incompressible, inner volume should be preserved during motion. We measure the relative loss of volume

$$E_{\text{vol}}\left(\bar{\mathbf{X}}, \hat{\mathbf{X}}\right) = \sum_{p=1}^{N_p} \left(\frac{V_p}{\bar{V}_p} - 1\right)^2, \tag{20}$$

Method	E_{pos} [mm]	E_{grad} [$\cdot 10^{-3}$]	E_{inv}	E_{vol} [$\cdot 10^{-3}$]
FEM	0.00	0.00	3.10	2.05
NDG	**3.19**	**1.11**	**9.22**	**3.75**
VertexNet	3.41	3.01	25.63	13.96
Gradient Skinning	6.69	9.89	93.99	33.32
LBS	7.56	15.04	167.63	40.67
Projective Skinning	–	–	11.31	28.05
CoRot LinFEM	–	–	108.18	21.99

Table 1: *Quantitative comparison of our method (NDG) against its baseline (Gradient Skinning), VertexNet, Linear Blend Skinning (LBS), Projective Skinning, and Co-Rotated Linear Elasticity (CoRot LinFEM) on test set bodies and poses. The target FEM simulation (FEM) is shown for comparison. Lower values indicate better performance; the best values are highlighted in bold.*

where V_p and $\bar{V}_p$ are the volume of the prism p in the deformed and undeformed state, respectively. This measurement is an exact version of the approximate volume term $(\det \mathbf{F}_p - 1)^2$ that is dominant in our FEM simulation. A derivation of the analytic prism volume is provided in the supplementary material. Inversions are measured as the number of elements with negative volume after deformation ($V_p < 0$). We call this metric E_{inv}. Each metric evaluates to a single scalar per posed body. Unless noted otherwise, we report averages over all 2500 body/pose pairs in the test set.

5.2. Comparison to Reference Animation Methods

We evaluate NDG against other viable alternatives including a vertex-based neural deformation approach and other physics-based approaches that can produce quasi-static solutions in real time. A summary of the results is shown in Table 1. All results are computed on our test set consisting of 2500 body/pose pairs. We observe that the FEM-based animation method performs well with respect to volume preservation and inversion avoidance, underlining that it properly respects physical behavior, despite utilizing an approximate deformation gradient operator.

5.2.1. Vertex-Based Neural Pose Correctives

We compare to a vertex-based neural method that predicts pose-corrective vertex offsets for LBS instead of deformation gradients. We implement a vertex-based neural network $\text{MLP}_{\text{vertex}}$ and train it on our data. We choose the concrete neural network architecture proposed by Han et al. [HCO*24] to model musculoskeletal deformations, as it is a recent approach designed for quasi-static, pose-driven simulations of a layered model. We condition the network on the body rest shape and pose in the same way as we do for our NDG network. The deformed mesh is given as

$$\mathbf{X} = \text{LBS}\left(\bar{\mathbf{X}} + \text{MLP}_{\text{vertex}}\left(\mathbf{r}, \text{PCA}\left(\bar{\mathbf{X}}\right)\right), \mathbf{r}\right). \tag{21}$$

$\text{MLP}_{\text{vertex}}$ was trained on pose-correctives that were obtained from our FEM-simulated targets. It follows the same architecture – including the number and size of layers – as NDG, except for the last layer which is replaced with a PCA decoder, following Han et al. [HCO*24]. To ensure that the network is not limited by the PCA, we allow 680 dimensions for the PCA. The PCA explains 99.5 % of the variance in correctives. This PCA is much larger than the one used by Han et al., but encodes correctives for a variety of body shapes instead of a single shape only. We train on the same data as NDG until convergence is reached after 212 epochs. Hyperparameters are tuned via automatic sweeps. For simplicity, we will call this network *VertexNet* throughout this section.

For a complete analysis, we evaluate both the individual networks as well as their baselines without neural predictions. In the case of VertexNet, this collapses to LBS. In the case of NDG, this is equivalent to solving for gradients that closely match the quaternion blended bone rotations $\mathbf{R}_p$. This network-free baseline, which we call *Gradient Skinning* for brevity, can be considered a volumetric version of the method of Weber et al. [WSLG07].

Figure 7: *A comparison of VertexNet and NDG with respect to per-vertex accuracy. For each skin vertex, the average distance of both the skin vertex and its corresponding muscle vertex to their respective target position is visualized. NDG attains $E_{\text{pos}} = 1.81\,\text{mm}$, while VertexNet attains $E_{\text{pos}} = 2.27\,\text{mm}$.*

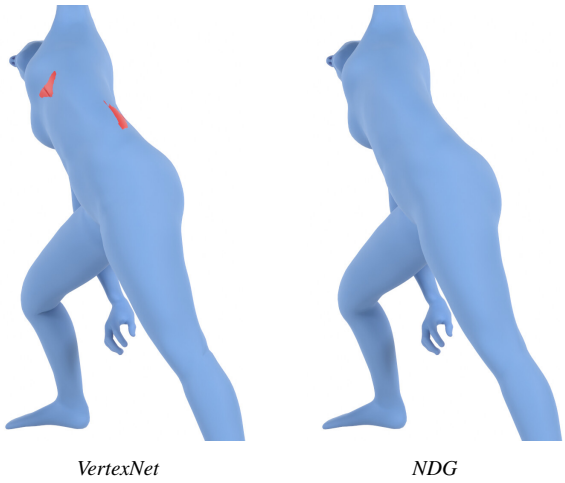
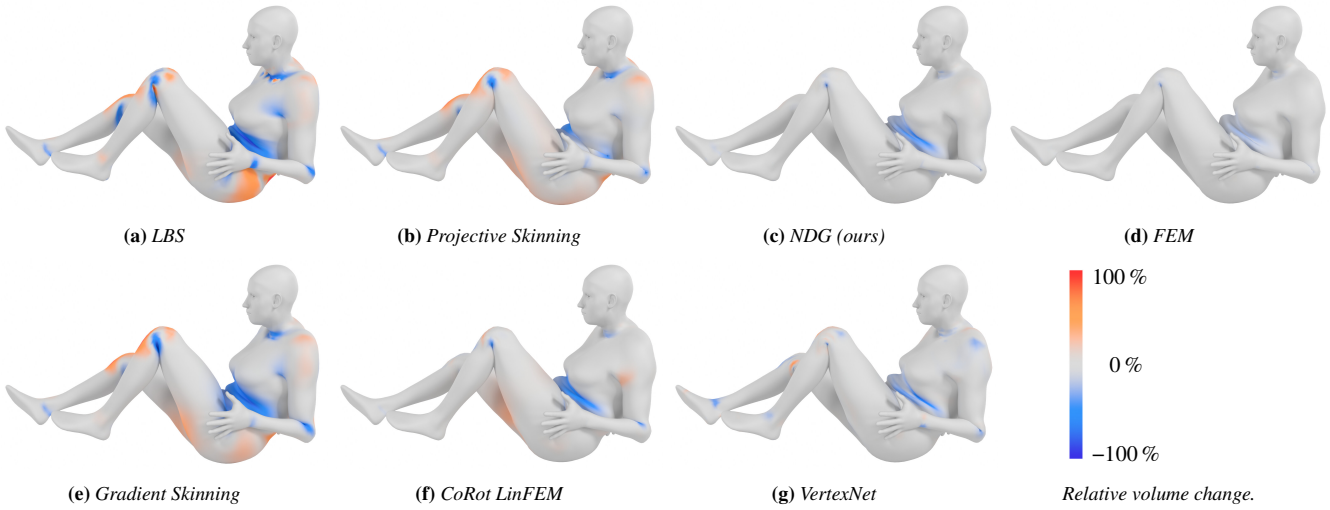


Figure 8: *Under stretching, the muscles predicted by the VertexNet (left) frequently intersect the skin surface. In contrast, the deformation gradients employed by NDG avoids intersections even for this challenging pose (right).*



(a) *LBS* **(b)** *Projective Skinning* **(c)** *NDG (ours)* **(d)** *FEM*

(e) *Gradient Skinning* **(f)** *CoRot LinFEM* **(g)** *VertexNet* *Relative volume change.*

Figure 9: *Comparison of animation methods. We visualize the change in local volume* $\left(\frac{V_i}{\bar{V}_i} - 1\right)$. *FEM is best at maintaining volume, closely followed by the neural methods. NDG outperforms VertexNet. The interactive physics based methods incur higher volume deviation, caused by their lack of a proper volume term. The network-free baseline methods, LBS and Gradient Skinning perform worst.*

Quantitative Evaluation. As shown in Table 1, NDG achieves slightly better vertex accuracy and better performance in volumetric metrics: It more than halves the number of inverted elements and reduces deformation-gradient and volume errors by a factor of three. This confirms that predicting in gradient space is better suited for volumetrically consistent deformation than vertex offsets. The per-vertex error is visualized in Figure 7. NDG's robustness to inversions is particularly noticeable in regions of strong stretching. One such stretching pose is shown in Figure 8. While NDG manages to animate this challenging pose without inversions, the VertexNet produces obvious inversions.

The performance of NDG, VertexNet, and the FEM target with respect to volume preservation is illustrated in Figure 9. As expected, the FEM simulation excels at preserving volume. NDG also learns to preserve volume well, losing only small amounts of volume around the upper abdominal region. VertexNet exhibits artifacts around joint areas, where the pose correctives fail to compensate for the LBS-induced distortions.

Gradient Skinning, the baseline of NDG, is noticeably stronger than LBS, especially with respect to volumetric measurements. It produces considerably fewer inversions and preserves volume better. While it does not produce convincing deformations on its own, this improved baseline reduces the amount of information the network has to learn.

Out of Distribution Evaluation. As a stress test of generalization ability, we examine the behavior of VertexNet and NDG when they are confronted with a body shape that is outside of their training distribution. We use an artist-generated body shape that represents a stylized yet realistic human, as opposed to the scanned body shapes the networks were trained on. We test the performance of both neural methods on the 2500 test poses. Quantitative results are shown in Table 2 and a qualitative comparison is shown in Figure 10. Since this body shape lies outside the training distribution, both networks incur a loss in accuracy. Notably, while its prediction quality remains relatively stable in vertex space, VertexNet experiences a steep drop in all volumetric measurements, producing many element inversions. NDG generalizes significantly better to this unseen body shape, achieving results much closer to in-distribution data. This highlights the robustness of NDG, which allows it to handle even entirely novel body shapes without requiring expensive retraining.

Method	E_{pos} [mm]	E_{grad} $[\cdot 10^{-3}]$	E_{inv}	E_{vol} $[\cdot 10^{-3}]$
NDG	**1.06**	**1.94**	**12.89**	**4.43**
VertexNet	1.54	14.77	149.15	44.35

Table 2: *Quantitative evaluation of NDG and VertexNet on the task of animating an out-of-distribution body shape. Lower values indicate better performance, best values are bold.*

5.2.2. Comparison to Interactive Animation Methods

Instead of neural methods, some physics-based methods can be used to meet interactive requirements. These methods simplify the physical energies, whereas NDG approximates higher quality physical deformations. We compare two interactive, physics-based animation methods that are well suited for quasi-static simulation.

Co-Rotated Linear Elasticity. An early approach to interactive, physically-based skinning uses linear elasticity [CBC*05], using the energy

$$\frac{\lambda}{2}\,\mathrm{tr}\,(\epsilon)^2 + \mu\,\mathrm{tr}\left(\epsilon^2\right) \tag{22}$$

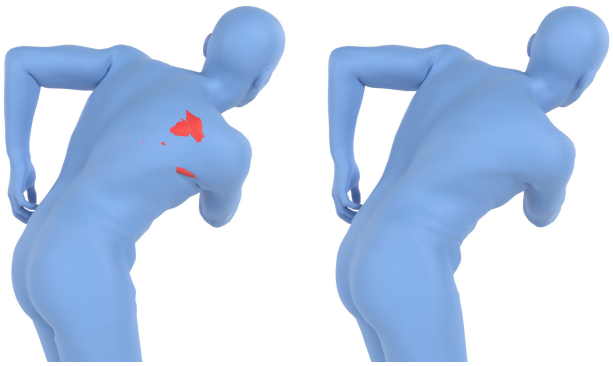


Figure 10: *When confronted with a body shape outside of the training distribution, VertexNet (left) produces noticeable protrusion artifacts much more frequently. In contrast, the deformation gradient based NDG remains robust.*

based on the linear small strain tensor ϵ

$$\epsilon(\mathbf{F}) = \frac{1}{2}\left(\mathbf{F} + \mathbf{F}^{\mathsf{T}}\right) - \mathbf{I}. \tag{23}$$

As the small strain tensor is inaccurate under large rotations, a rotation $\mathbf{R}_v$ is estimated per vertex and factored out during force computation. Given these rotations, a quasi-static solution can be found solving a linear system.

We compare NDG to animation computed using this co-rotated linear energy. We drive the pose using the same skeletal constraints as NDG and compute rotations $\mathbf{R}_v$ using quaternion blending for up to eight surrounding bones, matching Capell et al. [CBC*05].

Projective Skinning. Projective dynamics [BML*14] can be used to animate volumetric humans [KB18; KB19]. This approach models flesh with the elastic rotation strain energy proposed by [CPSS10]:

$$\sum_{p=1}^{N_p} \min_{\mathbf{R}_p \in \mathcal{SO}(3)} \|\mathbf{F}_p - \mathbf{R}_p\|^2 . \tag{24}$$

While designed for dynamics, this energy can be solved for a quasi-static solution with minimal changes to the solver.

We compare NDG with deformations produced using the energy and a CPU version of the local-global solver as used in projective skinning. The pose is driven in the same way as for NDG. We run the iterative local-global solver of projective skinning for a maximum of 30 iterations, allocating three times as much computational budget as the original, interactive implementation [KB18].

Quantitative Evaluation. As shown in Table 1, NDG substantially outperforms both projective skinning and co-rotated linear elasticity with respect to volume preservation. Both co-rotated elasticity and projective skinning show pronounced difficulties in preserving volume, caused by the lack of a proper volume term in their formulations [KE20]. NDG also reduces the number of inversions, achieving slightly fewer inversions than projective skinning, while at the same time approximating more complex deformations at lower computational cost.

Method	Frames w/ intersection [%]	Frames w/ protrusion [%]	Avg. protrusion magnitude [mm]
NDG	**17.2**	**0.0**	**0.0**
HIT	100.0	100.0	3.4
SKEL	100.0	100.0	4.1

Table 3: *Percentage of frames in which the muscles and the skeleton intersect the outer skin and protrude through the outer skin together with the average amount of protrusion. NDG significantly reduces interpenetration and the average protrusion amount compared to SKEL and HIT.*

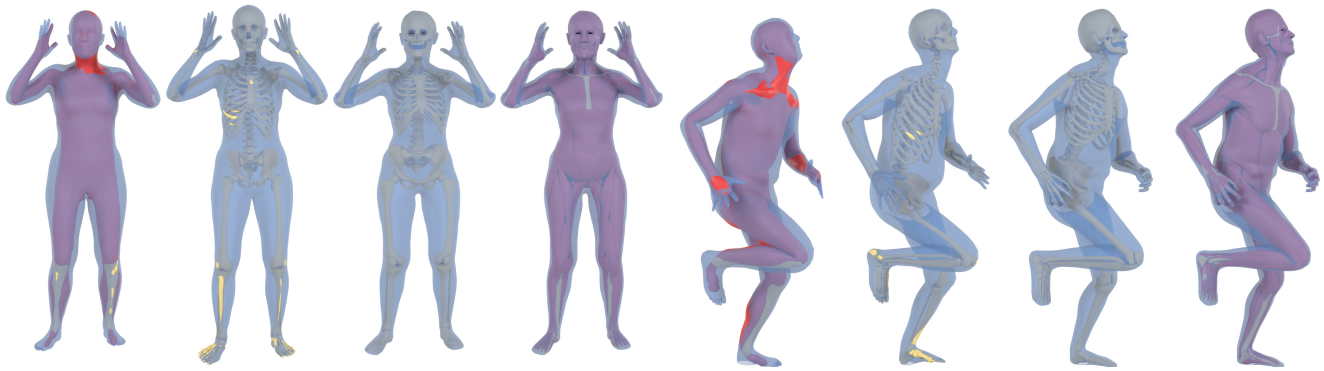
The lack of volume preservation is visualized in Figure 9, where both methods can be seen to incur relatively large changes in volume while NDG keeps the volume well preserved.

5.3. Comparison to SKEL and HIT

We compare our approach to SKEL [KWS*23], a joint model of both outer skin and an anatomical skeleton that derives a biomechanically plausible skeleton from an OpenSim-based simulation. SKEL shares the animation model of SMPL, which combines LBS with pose corrective offsets that depend linearly on pose parameters. This is structurally similar to our VertexNet but uses a linear model instead of a neural network. This model is not volumetric. As a result, SKEL's bones frequently protrude through the skin surface, even in relatively neutral poses.

We also compare to HIT [KAD*24], which learns an implicit representation of human tissues from the data and is conditioned on the space of the SMPL parameters. As HIT does not explicitly enforce non-penetration constraints for the internal tissue and the outer skin, internal iso-surfaces can intersect the outer skin mesh. NDG manages to properly fit both the bones and the surrounding muscles within the skin mesh. The comparison to SKEL and HIT is illustrated in Figure 11.

Quantitative Evaluation. We evaluate anatomical layer separation of NDG, SKEL, and HIT using $N = 500$ body-pose pairs uniformly sampled from AMASS (Human4D, Eyes Dataset Japan, PosePrior, and MOYO), pre-filtered for shape-pose mismatches via collision checks across five coarse body partitions (arms, legs, torso with head). We exclude the head, hands, and toes from the analysis, as InsideHumans lacks volumetric modeling for the head, hands, and toes. We benchmark NDG (high-resolution skeleton and musculature) against SKEL (skeleton) and HIT (bone and lean tissue) using a naive intersection count derived from triangle-triangle intersection test between the skin and internal tissues. We report the frequency of these intersections as *frames with intersection*. Because this metric accumulates internal joint overlaps that may exist within the skin volume, we differentiate surface protrusions using a winding number test [JKS13] to isolate internal vertices located outside the skin surface boundary. We report the frequency of these



(a) *HIT* **(b)** *SKEL* **(c)** *NDG* **(d)** *NDG* **(e)** *HIT* **(f)** *SKEL* **(g)** *NDG* **(h)** *NDG*

Figure 11: *Qualitative comparison of anatomical reconstructions: HIT (a, e), SKEL (b, f), NDG with hidden high-resolution muscles (c, g), and NDG with high-resolution skeleton and muscles (d, h). In these examples, HIT's reconstructed implicit meshes of the bone tissue (yellow) and the muscle (red) tissue intersect the skin (transparent blue), while SKEL's skeleton (yellow) penetrates the skin surface (transparent blue) at the rib cage and the lower legs. In contrast, NDG keeps the inner structures within the skin, thereby avoiding penetration artifacts.*

outliers as *frames with protrusion* and quantify the severity as *protrusion magnitude* in Table 3. Notably, NDG skin intersections occur in joint regions only, are internal to the skin surface and hence do not protrude. Thus, we report a protrusion magnitude of 0.0.

5.4. Runtime Evaluation

While NDG enables more robust volumetric animation than VertexNet, it requires a additional Poisson solve (see Equation 10) at each step. In contrast, VertexNet relies only on LBS.

The left-hand side of the Poisson solve is constant; each frame performs only right-hand side assembly and back-substitution. This takes 20–25 ms per frame on a consumer-level Intel Core i7-12700K CPU. Inference of the neural network runs in 5–7 ms on an NVIDIA 3070 GPU. On the same hardware, NDG is faster than both Projective Skinning and corotational linear elasticity. Projective Skinning takes 20 ms per iteration, while corotational linear elasticity takes 130 ms per frame. To find optimal rotations in Projective Skinning, we use the AVX-accelerated rotation-fitting proposed by Zhang et al. [ZJA21]. Despite the Poisson overhead, NDG can run at 25–30 fps on consumer-level hardware and yields fewer artifacts (inversions, volume preservation) than vertex-based neural methods – an attractive accuracy/robustness trade-off for anatomical animation.

5.5. Ablation Studies

To assess the contribution of the key components in our method, we conducted a series of ablation studies by altering three aspects of the model and comparing their effects to the full version. Specifically, we examined the effects of (i) enforcing a positive determinant on residual deformations, (ii) removing the rotation offset provided by the basic underlying skinning, and (iii) removing the body shape information in the input data. In all cases, the ablated models underperformed relative to our model, highlighting the necessity of each component. The results are summarized in Table 4.

Enforcing Positive Determinants NDG explicitly enforces that all predicted deformation gradients have positive determinants by applying a matrix exponential to its output (see Equation 16). For comparison, we also trained a neural network with the same architecture as NDG, but without the matrix exponential, predicting 3×3 deformation gradients directly for each prism. This neural network reaches a slightly lower training loss than NDG, which is a consequence of the network being less constrained. However, the increase in training performance does not translate well to test performance. Accuracy on the test set is equivalent for both networks and average inversions increased by over 30 % compared to the original NDG (measured on the test data). This underlines that enforcing positive determinants leads to more robust deformations.

Learning Residual Deformations NDG is trained to predict residual deformations $\mathbf{F}_p^{\text{res}}$ per prism, such that $\mathbf{F}_p = \mathbf{R}_p \mathbf{F}_p^{\text{res}}$. Alternatively, the network could be trained to predict deformation gradients $\hat{\mathbf{F}}_p$ directly. Both methods are equally expressive; they differ only in how well the representation is suited for neural network training. We trained a neural network with the same architecture as NDG, but configured to predict deformation gradients directly. This variant proved less stable and more difficult to optimize, converging to a 50% higher training loss. These findings support the expectation that forcing the neural network to also learn pose-induced joint rotations introduces unnecessary complexity and hinders effective training. Consequently, all evaluation metrics worsened (see Table 4).

Incorporating Body Shape Information NDG is conditioned on the rest-pose body shape, represented as a low-dimensional PCA encoding $\mathbf{b}$ of the vertex positions of the body's rest shape (see Figure 4). To examine the effect of this conditioning, we trained a variant of NDG without access to any information about the body shape. This model performs noticeably worse, reaching a training loss almost triple that of NDG. Consequently, almost all evaluation metrics worsen: The number of inverted elements increases by 50%

Method	E_{pos} [mm]	E_{grad} [$\cdot 10^{-3}$]	E_{inv}	E_{vol} [$\cdot 10^{-3}$]
NDG	**3.19**	1.11	**9.22**	**3.75**
NDG no matrix exp	**3.19**	**1.10**	12.45	4.04
NDG no rotation	3.39	1.38	11.23	4.42
NDG no body shape	3.69	1.93	14.02	5.23

Table 4: *Quantitative evaluation of NDG variants. Lower values indicate better performance, the best value is highlighted in bold.*

and the deformation gradient error almost doubles. This underlines the importance of conditioning the network on the body shape.

6. Conclusion

We presented a neural network-driven method for the skeletal-driven animation of human models. By embedding anatomical details in a volumetric mesh and animating using our proposed Neural Deformation Gradients (NDG) method, outer skin and detailed anatomy can jointly and consistently be animated.

Our NDG model is trained on FEM simulated deformations of a variety of different poses and body shapes, approximating them with high fidelity at a fraction of the simulation's cost. We showed that, by learning in deformation gradient space, NDG outperforms vertex-based deep-learning approaches in the particular setting of volumetric animation: It produces fewer inverting elements, preserves volume better, and generalizes more robustly to novel body shapes. This enables robust volumetric animation of different bodies without retraining. By building on a body model similar to InsideHumans, volumetric bodies can be automatically fitted to scanned surfaces and then animated using our method. These benefits come at the cost of solving a constant Poisson system at each frame. By employing an efficient per-prism approximation of the deformation gradient field, our animation method can reach at 25–30 frames per second on consumer hardware.

Despite these advantages, our method also has limitations that we want to address in future work. One core limitation of our model lies in the deformation of the skeleton. Currently, LBS is used to deform the bone wrap, from which the high-resolution skeleton is interpolated. While simple and efficient, this heuristic is not physically grounded and leads to bones that are not perfectly rigid. As NDG is fundamentally compatible with any method that deforms the bone wrap, finding a more suitable method to deform the bone wrap will improve bone rigidity and overall deformation results. Furthermore, since the head, hands, and toes are not volumetrically modeled, the embedded deformation combined with LBS may sometimes produce minor artifacts around these areas. Modeling these areas volumetrically would address these shortcomings.

Another limitation lies in the lack of anatomical ground truth data. Accurate data of human anatomy is hard to obtain, especially when diverse poses are required. Incorporating medical data, similar to [KAD*24], could enhance biological plausibility and provide a realistic baseline to evaluate against. In the absence of medical ground truth, improvements could be made to our simulated dataset. Notably, it lacks self-collision handling. Collision constraints would add realism [HCO*24], but likely require additional treatment as most current neural deformation methods, including ours, do not naturally handle collisions.

Moreover, NDG is unable to change, e.g., material parameters at inference time. Changing animation settings requires re-training which takes considerable efforts, in particular with respect to data generation.

Finally, our current body shape space is based on a simple PCA model. A richer shape representation could better capture anatomical variability. In particular, a jointly trained body and pose model could unlock synergistic effects.

7. Acknowledgments

This research has been funded by the Federal Ministry of Education and Research of Germany and the state of North Rhine-Westphalia as part of the Lamarr Institute for Machine Learning and Artificial Intelligence.

References

[AB15] AKHTER, IJAZ and BLACK, MICHAEL J. "Pose-Conditioned Joint Angle Limits for 3D Human Pose Reconstruction". *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2015.

[AF15] ABU RUMMAN, NADINE and FRATARCANGELI, MARCO. "Position-based skinning for soft articulated characters". *Computer Graphics Forum* 34.6 (2015), 240–250.

[AGK*22] AIGERMAN, NOAM, GUPTA, KUNAL, KIM, VLADIMIR G., et al. "Neural jacobian fields: learning intrinsic mappings of arbitrary meshes". *ACM Transactions on Graphics* 41.4 (2022).

[AMG*13] ALKHOULI, NADIA, MANSFIELD, JESSICA, GREEN, ELLEN, et al. "The mechanical properties of human adipose tissues and their relationships to the structure and composition of the extracellular matrix". *American Journal of Physiology-Endocrinology and Metabolism* 305.12 (2013), E1427–E1435.

[ASK*05] ANGUELOV, DRAGOMIR, SRINIVASAN, PRAVEEN, KOLLER, DAPHNE, et al. "SCAPE: Shape Completion and Animation of People". *ACM Transactions on Graphics* 24.3 (2005), 408–416.

[AWLB17] ACHENBACH, JASCHA, WALTEMATE, THOMAS, LATOSCHIK, MARC ERICH, and BOTSCH, MARIO. "Fast generation of realistic virtual humans". *Proc. of ACM Symposium on Virtual Reality Software and Technology*. 2017, 1–10.

[BBC19] BADER, PHILIPP, BLANES, SERGIO, and CASAS, FERNANDO. "Computing the matrix exponential with an optimized Taylor polynomial approximation". *Mathematics* 7.12 (2019), 1174.

[BML*14] BOUAZIZ, SOFIEN, MARTIN, SEBASTIAN, LIU, TIANTIAN, et al. "Projective Dynamics: Fusing Constraint Projections for Fast Simulation". *ACM Transactions on Graphics* 33.4 (2014), 154:1–154:11.

[BODO18] BAILEY, STEPHEN W., OTTE, DAVE, DILORENZO, PAUL, and O'BRIEN, JAMES F. "Fast and deep deformation approximations". *ACM Transactions on Graphics* 37.4 (2018).

[BSPG06] BOTSCH, MARIO, SUMNER, ROBERT W., PAULY, MARK, and GROSS, MARKUS. "Deformation Transfer for Detail-Preserving Surface Editing". *Vision, Modeling and Visualization*. 2006, 357–364.

[CAD*12] CHINO, KAZUKI, AKAGI, RYOTA, DOHI, MASAHIRO, et al. "Reliability and Validity of Quantifying Absolute Muscle Hardness Using Ultrasound Elastography". *PLoS ONE* 7.9 (2012).

[CBC*05] CAPELL, STEVE, BURKHART, MATTHEW, CURLESS, BRIAN, et al. "Physically Based Rigging for Deformable Characters". *Proc. of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. 2005, 301–310.

[CG23] CHEN, CHUN-TEH and GU, GRACE X. "Physics-informed deep-learning for elasticity: forward, inverse, and mixed problems". *Advanced Science* 10.18 (2023).

[CO18] CASAS, DAN and OTADUY, MIGUEL A. "Learning nonlinear soft-tissue dynamics for interactive avatars". *Proc. of the ACM on Computer Graphics and Interactive Techniques* 1.1 (2018), 1–15.

[CPH*25] CORIGLIANO, DAVIDE, PETER, DANIEL, HUBER, NIKO BENJAMIN, et al. "NeuRiPhy: Neural Baking of Physics-Based Deformations for Facial Rigs". *Proc. of the ACM on Computer Graphics and Interactive Techniques* 8.4 (2025), 1–18.

[CPSS10] CHAO, ISAAC, PINKALL, ULRICH, SANAN, PATRICK, and SCHRÖDER, PETER. "A simple geometric model for elastic deformations". *ACM Transactions on Graphics* 29.4 (2010), 1–6.

[CSB*20] CHATZITOFIS, ANARGYROS, SAROGLOU, LEONIDAS, BOUTIS, PRODROMOS, et al. *HUMAN4D: A Human-Centric Multimodal Dataset for Motions & Immersive Media*. 2020.

[DB13] DEUL, CRISPIN and BENDER, JAN. "Physically-Based Character Skinning". *Virtual Reality Interactions and Physical Simulations (VRIPhys)*. Eurographics Association, 2013.

[DHG23] DALTON, DAVID, HUSMEIER, DIRK, and GAO, HAO. "Physics-informed graph neural network emulation of soft-tissue mechanics". *Computer Methods in Applied Mechanics and Engineering* 417 (2023).

[GHDS03] GRINSPUN, EITAN, HIRANI, ANIL N, DESBRUN, MATHIEU, and SCHRÖDER, PETER. "Discrete shells". *Proc. of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. 2003, 62–67.

[Gon85] GONZALEZ, TEOFILO F. "Clustering to minimize the maximum intercluster distance". *Theoretical computer science* 38 (1985), 293–306.

[GYQ*18] GAO, LIN, YANG, JIE, QIAO, YI-LING, et al. "Automatic unpaired shape deformation transfer". *ACM Transactions on Graphics* 37.6 (2018), 1–15.

[Hab18] HABER, HOWARD E. "Notes on the matrix exponential and logarithm". *Santa Cruz Institute for Particle Physics, University of California: Santa Cruz, CA, USA* (2018).

[HCO*24] HAN, YUSHAN, CHEN, YIZHOU, ONG, CARMICHAEL, et al. "A Neural Network Model for Efficient Musculoskeletal-Driven Skin Deformation". *ACM Transactions on Graphics* 43.4 (2024).

[JKS13] JACOBSON, ALEC, KAVAN, LADISLAV, and SORKINE-HORNUNG, OLGA. "Robust inside-outside segmentation using generalized winding numbers". *ACM Transactions on Graphics* 32.4 (2013), 1–12.

[KAD*24] KELLER, MARILYN, ARORA, VAIBHAV, DAKRI, ABDELMOUTTALEB, et al. "HIT: Estimating Internal Human Implicit Tissues from the Body Surface". *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2024, 3480–3490.

[KB15] KINGMA, DIEDERIK P. and BA, JIMMY. "Adam: A Method for Stochastic Optimization". *International Conference on Learning Representations*. 2015.

[KB18] KOMARITZAN, MARTIN and BOTSCH, MARIO. "Projective skinning". *Proc. of the ACM on Computer Graphics and Interactive Techniques* 1.1 (2018), 1–19.

[KB19] KOMARITZAN, MARTIN and BOTSCH, MARIO. "Fast Projective Skinning". *Proc. of the ACM Symposium on Interactive 3D Graphics and Games*. ACM, 2019, 1–10.

[KC71] KAVANAGH, KENNETH T and CLOUGH, RAY W. "Finite element applications in the characterization of elastic solids". *International Journal of Solids and Structures* 7.1 (1971), 11–23.

[KCŽO07] KAVAN, LADISLAV, COLLINS, STEVEN, ŽÁRA, JIŘÍ, and O'SULLIVAN, CAROL. "Skinning with dual quaternions". *Proc. of the ACM Symposium on Interactive 3D Graphics and Games*. 2007, 39–46.

[KCŽO08] KAVAN, LADISLAV, COLLINS, STEVEN, ŽÁRA, JIŘÍ, and O'SULLIVAN, CAROL. "Geometric skinning with approximate dual quaternion blending". *ACM Transactions on Graphics* 27.4 (2008), 1–23.

[KE20] KIM, THEODORE and EBERLE, DAVID. "Dynamic deformables: implementation and production practicalities". *ACM SIGGRAPH 2020 courses*. 2020, 1–182.

[KIL*16] KADLEČEK, PETR, ICHIM, ALEXANDRU-EUGEN, LIU, TIANTIAN, et al. "Reconstructing personalized anatomical models for physics-based body animation". *ACM Transactions on Graphics* 35.6 (2016), 1–13.

[KK19] KADLEČEK, PETR and KAVAN, LADISLAV. "Building accurate physics-based face models from data". *Proc. of the ACM on Computer Graphics and Interactive Techniques* 2.2 (2019), 1–16.

[KPP*17] KIM, MEEKYOUNG, PONS-MOLL, GERARD, PUJADES, SERGI, et al. "Data-driven physics for human soft tissue animation". *ACM Transactions on Graphics* 36.4 (2017), 1–12.

[KWB21] KOMARITZAN, MARTIN, WENNINGER, STEPHAN, and BOTSCH, MARIO. "Inside Humans: Creating a Simple Layered Anatomical Model from Human Surface Scans". *Frontiers in Virtual Reality* 2 (2021).

[KWS*23] KELLER, MARILYN, WERLING, KEENON, SHIN, SOYONG, et al. "From Skin to Skeleton: Towards Biomechanically Accurate 3D Digital Humans". *ACM Transactions on Graphics* 42.6 (2023).

[KZBP22] KELLER, MARILYN, ZUFFI, SILVIA, BLACK, MICHAEL J., and PUJADES, SERGI. "OSSO: Obtaining Skeletal Shape from Outside". *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York, NY, USA: IEEE, 2022, 20492–20501.

[LAH*21] LI, PEIZHUO, ABERMAN, KFIR, HANOCKA, RANA, et al. "Learning skeletal articulations with neural blend shapes". *ACM Transactions on Graphics* 40.4 (2021).

[LCF00] LEWIS, JOHN P, CORDNER, MATT, and FONG, NICKSON. "Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation". *Proc. of the ACM Annual Conference on Computer Graphics and Interactive Techniques*. 2000, 165–172.

[LH16] LE, BINH HUY and HODGINS, JESSICA K. "Real-time skeletal skinning with optimized centers of rotation." *ACM Transactions on Graphics* 35.4 (2016), 37–1.

[LH19] LOSHCHILOV, ILYA and HUTTER, FRANK. "Decoupled Weight Decay Regularization". *International Conference on Learning Representations* (2019).

[LL19] LE, BINH HUY and LEWIS, JOHN-PETER. "Direct delta mush skinning and variants." *ACM Transactions on Graphics* 38.4 (2019).

[LLAF20] LANDAU, LEV D, LIFSHITZ, EVGENII M, ATKIN, RJ, and FOX, N. "The theory of elasticity". *Physics of Continuous Media*. 2020, 167–178.

[LMR*15] LOPER, MATTHEW, MAHMOOD, NAUREEN, ROMERO, JAVIER, et al. "SMPL: a skinned multi-person linear model". *ACM Transactions on Graphics* 34.6 (2015).

[Ltd] LTD., EYES JAPAN CO. *Eyes Japan MoCap Dataset*. URL: http://mocapdata.com.

[LWK*24] LI, PEIZHUO, WANG, TUANFENG Y, KESDOGAN, TIMUR LEVENT, et al. "Neural Garment Dynamics via Manifold-Aware Transformers". *Computer Graphics Forum* 43.2 (2024), e15028.

[MDRW14] MANCEWICZ, JOE, DERKSEN, MATT L., RIJPKEMA, HANS, and WILSON, CYRUS A. "Delta Mush: Smoothing Deformations While Preserving Detail". *Proc. of the ACM Symposium on Digital Production*. 2014, 7–11.

[MGT*19] MAHMOOD, NAUREEN, GHORBANI, NIMA, TROJE, NIKOLAUS F., et al. "AMASS: Archive of Motion Capture As Surface Shapes". *Proc. of the IEEE/CVF International Conference on Computer Vision*. 2019, 5442–5451.

[MHHR07] MÜLLER, MATTHIAS, HEIDELBERGER, BRUNO, HENNIX, MARCUS, and RATCLIFF, JOHN. "Position based dynamics". *Journal of Visual Communication and Image Representation* 18.2 (2007), 109–118.

[MLT88] MAGNENAT-THALMANN, NADIA, LAPERRIÈRE, RICHARD, and THALMANN, DANIEL. "Joint-Dependent Local Deformations for Hand Animation and Object Grasping". *Proc. of Graphics Interface*. 1988, 26–33.

[MTG04] MULLER, MATTHIAS, TESCHNER, MATTHIAS, and GROSS, MARKUS. "Physically-based simulation of objects represented by surface meshes". *Proc. IEEE Computer Graphics International*. 2004, 26–33.

[MZS*11] MCADAMS, ALEKA, ZHU, YONGNING, SELLE, ANDREW, et al. "Efficient Elasticity for Character Skinning with Contact and Collisions". *ACM Transactions on Graphics* 30.4 (2011), 37:1–37:12.

[QSA*23] QIN, DAFEI, SAITO, JUN, AIGERMAN, NOAM, et al. "Neural face rigging for animating and retargeting facial meshes in the wild". *ACM SIGGRAPH Conference Proceedings*. 2023, 1–11.

[RDP99] ROBINETTE, KATHLEEN M., DAANEN, HANS, and PAQUET, ERIC. "The CAESAR Project: A 3-D Surface Anthropometry Survey". *Proc. of the IEEE International Conference on 3-D Digital Imaging and Modeling*. 1999, 380–386.

[ROCP20] ROMERO, CRISTIAN, OTADUY, MIGUEL A., CASAS, DAN, and PEREZ, JESUS. "Modeling and Estimation of Nonlinear Skin Mechanics for Animated Avatars". *Computer Graphics Forum* 39.2 (2020), 77–88.

[SCSG18] SUJAR, AARON, CASAFRANCA, JUAN JOSE, SERRURIER, ANTOINE, and GARCIA, MARCOS. "Real-time animation of human characters' anatomy". *Computers & Graphics* (2018), 268–277.

[SGK18] SMITH, BREANNAN, GOES, FERNANDO DE, and KIM, THEODORE. "Stable neo-hookean flesh simulation". *ACM Transactions on Graphics* 37.2 (2018), 1–15.

[SGOC20] SANTESTEBAN, IGOR, GARCES, ELENA, OTADUY, MIGUEL A., and CASAS, DAN. "SoftSMPL: Data-driven Modeling of Nonlinear Soft-tissue Dynamics for Parametric Humans". *Computer Graphics Forum* 39.2 (2020), 65–75.

[Si06] SI, HANG. "TetGen, A Quality Tetrahedral Mesh Generator and Three-Dimensional Delaunay Triangulator". *Weierstrass Institute for Applied Analysis and Stochastic, Berlin, Germany* 81 (2006), 12.

[SP04] SUMNER, ROBERT W and POPOVIĆ, JOVAN. "Deformation transfer for triangle meshes". *ACM Transactions on Graphics* 23.3 (2004), 399–405.

[SZK15] SAITO, SHUNSUKE, ZHOU, ZI-YE, and KAVAN, LADISLAV. "Computational bodybuilding: Anatomically-based modeling of human bodies". *ACM Transactions on Graphics* 34.4 (2015), 1–12.

[TGLX18] TAN, QINGYANG, GAO, LIN, LAI, YU-KUN, and XIA, SHIHONG. "Variational Autoencoders for Deforming 3D Mesh Models". *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, 5841–5850.

[TMH*23] TRIPATHI, SHASHANK, MÜLLER, LEA, HUANG, CHUN-HAO P., et al. "3D Human Pose Estimation via Intuitive Physics". *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

[TRPO21] TAPIA, JAVIER, ROMERO, CRISTIAN, PÉREZ, JESÚS, and OTADUY, MIGUEL A. "Parametric Skeletons with Reduced Soft-Tissue Deformations". *Computer Graphics Forum* 40.6 (2021), 34–46.

[TSIF05] TERAN, JOSEPH, SIFAKIS, EFTYCHIOS, IRVING, GEOFFREY, and FEDKIW, RONALD. "Robust Quasistatic Finite Elements and Flesh Simulation". *Proc. of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. Association for Computing Machinery, 2005, 181–190.

[Wam07] WAMPLER, CHARLES W. "Manipulator inverse kinematic solutions based on vector formulations and damped least-squares methods". *IEEE Transactions on Systems, Man, and Cybernetics* 16.1 (2007), 93–101.

[WBS23] WAGNER, NICOLAS, BOTSCH, MARIO, and SCHWANECKE, ULRICH. "SoftDECA: Computationally efficient physics-based facial animations". *Proc. of the ACM Symposium on Interactive 3D Graphics and Games*. 2023, 1–11.

[WKSB24] WENNINGER, STEPHAN, KEMPER, FABIAN, SCHWANECKE, ULRICH, and BOTSCH, MARIO. "TailorMe: Self-Supervised Learning of an Anatomically Constrained Volumetric Human Shape Model". *Computer Graphics Forum* 43.2 (2024).

[WSLG07] WEBER, OFIR, SORKINE, OLGA, LIPMAN, YARON, and GOTSMAN, CRAIG. "Context-aware skeletal shape deformation". *Computer Graphics Forum* (2007), 265–274.

[XBZ*20] XU, HONGYI, BAZAVAN, EDUARD GABRIEL, ZANFIR, ANDREI, et al. "GHUM & GHUML: Generative 3D Human Shape and Articulated Pose Models". *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, 6184–6193.

[YZC*24] YANG, LINGCHEN, ZOSS, GASPARD, CHANDRAN, PRASHANTH, et al. "Learning a Generalized Physical Face Model From Data". *ACM Transactions on Graphics* 43.4 (2024).

[ZJA21] ZHANG, JIAYI ERIS, JACOBSON, ALEC, and ALEXA, MARC. "Fast Updates for Least-Squares Rotational Alignment". *Computer Graphics Forum* 40.2 (2021), 13–22.

[ZTZ05] ZIENKIEWICZ, O.C., TAYLOR, R.L., and ZHU, J.Z. *The Finite Element Method: Its Basis and Fundamentals*. 6th ed. Butterworth-Heinemann, 2005.

[ZZCB21] ZHENG, MIANLUN, ZHOU, YI, CEYLAN, DUYGU, and BARBIČ, JERNEJ. "A Deep Emulator for Secondary Motion of 3D Characters". *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, 5928–5936.

[ZZG19] ZHANG, JINAO, ZHONG, YONGMIN, and GU, CHENGFAN. "Neural network modelling of soft tissue deformation for surgical simulation". *Artificial intelligence in medicine* 97 (2019), 61–70.