

PhysGA: Physics-Guided Garment Dynamics for Topology-Aware Gaussian Avatars

Yong Deng¹, Baoxing Li¹, and Xu Zhao¹

Shanghai Jiao Tong University, Shanghai, China
{dengyong2016, lbx11sjtu, zhaoxu}@sjtu.edu.cn

Abstract. Monocular human avatar reconstruction remains challenging, particularly in modeling physically plausible and structurally stable garment dynamics under complex motions. To address this, we present a physics-guided Gaussian avatar framework (PhysGA) with explicit topology-aware garment modeling. Our approach is built upon two coupled priors: a structural prior for coherent garment structure and a physical prior for plausible garment dynamics. Specifically, the structural prior is derived from temporally reconstructed garment geometry by explicitly encoding inherited connectivity. This connectivity is further reinforced through a topology-preserving loss, preventing structural degradation under large and complex motions. Built on this topology-aware structure, the physical prior is encoded by an elastic graph network that incorporates motion-dependent states and material properties to approximate underlying elastic garment dynamics. The predicted deformations are further refined by a differentiable physics correction module grounded in a classical mass-spring formulation, compensating for inertial effects that are poorly captured by purely data-driven models. By coupling topology-aware structure with physics-guided dynamics in a unified Gaussian representation, PhysGA demonstrates competitive performance compared with existing monocular avatar reconstruction methods, particularly in producing realistic and structurally stable garment dynamics.

Keywords: Human Avatars · 3D Gaussian Splatting · Physical Priors

1 Introduction

Creating high-fidelity 3D human avatars from monocular video remains a fundamental problem in computer vision and graphics, with broad implications for virtual reality, telepresence, and digital content creation. Recent advances in neural rendering [5, 7, 46, 56, 57], particularly 3D Gaussian Splatting (3DGS) [9, 21, 43, 61], have demonstrated impressive rendering quality and efficiency for static scenes. However, extending such representations to dynamic human avatars remains far from trivial. Unlike static scenes, dynamic humans exhibit complex, highly non-rigid, and pose-dependent deformations, whose physically plausible and structurally stable modeling from sparse monocular observations is severely ill-posed. These challenges motivate the need for physics-aware modeling beyond purely

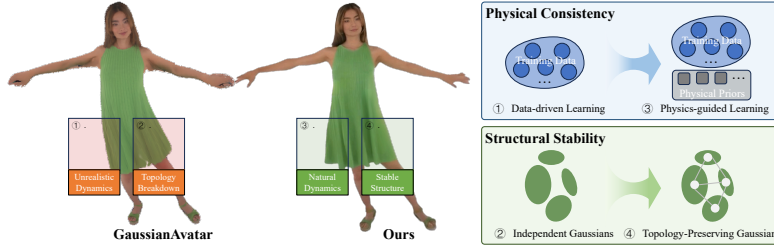


Fig. 1: 3D Gaussian avatars with improved physical consistency and structural stability. Compared with GaussianAvatar [15], our method achieves improved results (left) via physics-guided learning and topology-preserving Gaussians (right).

data-driven representations. In this work, we introduce a physics-guided framework for monocular human avatar reconstruction.

Early approaches primarily relied on implicit neural representations, such as Neural Radiance Fields (NeRF) [19, 27, 58, 63] and Signed Distance Functions (SDF) [12, 20, 50, 54], to model human geometry and appearance. While capable of producing high-quality novel views, these methods often suffer from slow rendering and inefficient training, and struggle to model dynamic deformations under novel poses. Recently, 3DGS [16, 25, 35, 50, 59] has emerged as a promising alternative, enabling high-quality rendering with significantly improved efficiency. By representing human avatars as a set of explicit, anisotropic 3D Gaussians and employing a differentiable rasterizer [24], 3DGS enables real-time rendering while maintaining high visual fidelity. Existing 3DGS-based methods either attach Gaussians to a parametric mesh (e.g., SMPL [30]) or learn a continuous deformation field to drive human avatars, and then rely on image data of the training subject to learn the Gaussian parameters.

However, despite their impressive rendering performance, these unconstrained, data-driven Gaussian avatar learning strategies suffer from two key limitations affecting realism and robustness: 1) *Physical inconsistency*. Most existing methods [15, 35, 52] typically parameterize non-rigid deformations in a purely data-driven manner, such as learning time-varying offsets of Gaussian centers [38], predicting pose-dependent skinning weight offsets [16], or introducing latent bones to implicitly encode deformation patterns [26]. In the absence of explicit dynamical modeling, these approaches struggle to capture inertial effects and material responses, particularly for garment regions that are weakly constrained by the underlying body. Consequently, the predicted deformations fail to accurately capture physically consistent dynamic behavior, resulting in unrealistic motion patterns. 2) *Structural instability*. The Gaussian representation consists of a collection of independent Gaussian primitives [29] without explicit topological coupling. As a result, deformation models operating on individual Gaussians or unconstrained latent representations become fragile under large or complex motions. This issue is particularly severe for loose garments, such as flowing skirts or

sleeves, where the spatial distribution of Gaussians becomes poorly regularized, resulting in structural disorganization and surface discontinuities.

To overcome the limitations of existing methods, we integrate dynamic physical priors and topology-aware geometric priors into the Gaussian representation, yielding human avatar reconstructions that are physically plausible and structurally stable (Fig. 1). *For physical consistency*, we introduce an Elastic Graph Network (EGN) that goes beyond pose-driven deformation by learning subject-specific garment dynamics from motion states and stiffness features. EGN predicts an initial non-rigid deformation field, which is subsequently refined by a differentiable physics correction module (PCM) based on a classical mass-spring energy formulation [6]. This module complements EGN by imposing physics-inspired constraints on its predicted deformation field, encouraging physically plausible inertial cloth motion while suppressing unrealistic dynamic artifacts. *For structural stability*, we explicitly enforce topology-aware structural preservation through a Gaussian adjacency graph. By leveraging temporally coherent geometry across frames, we initialize the Gaussian representation and inherit garment connectivity to construct a topology-aware graph. This graph serves as a shared structural scaffold: it preserves local garment coherence and provides the propagation paths for physics-guided deformation. Based on this graph, we further impose a topology-preserving loss to penalize structural violations. Additionally, to reduce computational overhead, we assign a semantic label to each Gaussian and apply physics-guided deformation only to garment Gaussians. This semantic decomposition also enables straightforward garment transfer across avatars.

In summary, our contributions are threefold: 1) We present a physics-guided elastic graph network for initial non-rigid deformation prediction, incorporating core physical cues to produce plausible deformations. 2) We propose a lightweight differentiable physics correction module, serving as a physics-inspired regularizer to enhance the modeling of inertial effects and suppress unrealistic dynamic artifacts. 3) We construct a Gaussian adjacency graph from temporally coherent geometry, which preserves local garment connectivity and supports a topology-preserving loss for stable garment structures. Together, these contributions enable physically plausible and structurally stable Gaussian avatars.

2 Related Work

2.1 Dynamic Human Representations

Research on dynamic human representation has undergone significant paradigm shifts. Parametric mesh models, such as SMPL [30], SMPL-X [39], and STAR [36], are widely used for human reconstruction. Their explicit kinematic structure and compact shape space enable robust pose-driven motion, but their body-centric fixed topology limits the modeling of garments with independent dynamics. To address these geometric constraints, the field shifted toward neural implicit representations, including occupancy fields [10, 33], signed-distance

fields [37, 47], and radiance fields [34, 42]. Implicit fields excel at modeling continuous surfaces without fixed topology and provide smooth spatial gradients for optimization. However, implicit fields are globally coupled and require whole-field recomputation for dynamics, lack explicit topology for stable structure, and suffer from slow volumetric rendering. In contrast, 3DGS [21] offers an explicit point-based representation that naturally supports dynamic modeling while enabling real-time rendering. Nevertheless, Gaussian primitives lack inherent topological constraints, often leading to structural drift or breakage under large motions. We therefore construct a Gaussian adjacency graph from temporally reconstructed geometry and introduce topology-preserving constraints to maintain coherent garment structure during deformation.

2.2 Monocular Avatar Reconstruction

Monocular human avatar reconstruction [11] remains challenging due to the difficulty of jointly achieving accurate geometry and photorealistic rendering from sparse observations. Early methods [4, 12, 20] learn a canonical SDF under photometric supervision, enabling coherent surface reconstruction but requiring expensive per-subject optimization and struggling with highly dynamic garments. Radiance-field-based approaches [18, 53] model a canonical radiance field with learned deformation fields to integrate monocular observations, enabling flexible motion representation but incurring slow volumetric rendering and high training costs. More recently, 3D Gaussian Splatting (3DGS) has emerged as an efficient alternative with explicit primitives and real-time differentiable rendering. 3DGS-based avatar methods [15, 16, 23, 26, 35] optimize dynamic Gaussian attributes to match observations, but often exhibit clothing distortion under large motions due to the absence of explicit geometric constraints. Vid2Avatar-Pro [13] mitigates this issue with a universal prior learned from multi-view data. Recent methods bind Gaussians to articulated parametric body models to enhance structural consistency across poses [38, 45, 49, 52]. However, the fixed topology of the underlying body model limits their ability to represent loose and non-body-conforming garments. To preserve geometric constraints without restricting topology, we replace parametric templates with a temporally reconstructed surface mesh with explicit garment geometry.

2.3 Garment Dynamics Modeling

Realistic avatar animation depends critically on modeling non-rigid garment deformations beyond skeleton-driven body motion. Existing avatar methods commonly model non-rigid deformation through pose-conditioned residual representations, such as per-point offsets [41, 53, 55], refined skinning weights [16], auxiliary deformation structures [26], or dynamic Gaussian parameters [15]. As these methods are largely data-driven and lack explicit physical constraints, they may produce physically implausible dynamics under unseen poses. PhysAvatar [62] estimates garment parameters from multi-view video through inverse physics, while PICA [40], Gaussian Garments [44], and PGC [14] construct simulation-oriented

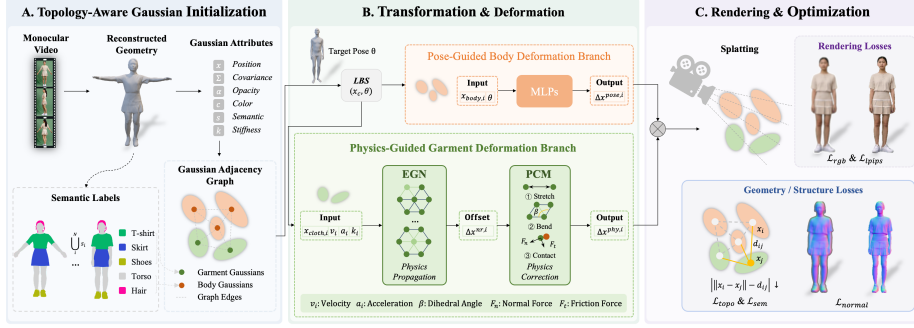


Fig. 2: Method overview. Our approach initializes Gaussians using a temporally consistent human mesh reconstructed across frames and assigns an initial semantic label s to each Gaussian based on semantic segmentation results. After applying pose transformations (Linear Blend Skinning, LBS), a dual-branch network with a pose-guided branch (PoseDeform) and a physics-guided branch (PhyDeform) predicts the non-rigid deformations of the inner body and garments, respectively. Finally, the Gaussian model is jointly optimized using image-, geometry-, and semantic-related loss terms.

clothed avatars or garments. These approaches often rely on multi-view captures or simulation-ready garment assets, which may introduce additional acquisition and computational costs. In contrast, our method targets monocular Gaussian avatar reconstruction without requiring full standalone cloth simulation. We instead design a garment-specific deformation branch that couples physics-guided graph prediction with lightweight differentiable physics correction, producing physically plausible garment dynamics with limited extra computation.

3 Method

Given a monocular video of a dynamic human, our goal is to reconstruct an animatable 3D Gaussian avatar with stable structure and physically consistent dynamics, as illustrated in Fig. 2. We initialize Gaussian parameters from temporally reconstructed meshes (Sec. 3.1). We then introduce a baseline deformation framework combining Linear Blend Skinning for articulated motion with a non-rigid deformation predictor for dynamics (Sec. 3.2). To enhance the realism of clothing dynamics, we further design a differentiable physics-guided deformation branch that enforces physically plausible motion (Sec. 3.3). Finally, we present the training objectives and the two-stage optimization strategy (Sec. 3.4).

3.1 3D Gaussian Initialization

3D Gaussian Splatting. A 3D Gaussian’s spatial distribution can be represented by its density function:

$$D(y) = \exp\left(-\frac{1}{2}(y - x)^\top \Sigma^{-1}(y - x)\right), \quad (1)$$

where $x \in \mathbb{R}^3$ denotes the Gaussian 3D center position and $\Sigma \in \mathbb{R}^{3 \times 3}$ is the covariance matrix that determines the Gaussian’s scale, anisotropy, and orientation in space. In addition to its spatial parameters, each 3D Gaussian also carries a color parameter c , and an opacity parameter α , which jointly determine its appearance in the rendering process. We further assign each Gaussian a semantic label s to indicate the body or clothing region it belongs to, enabling region-aware deformation and regularization. Moreover, we introduce a stiffness coefficient k that characterizes the local resistance to non-rigid deformation and serves as a physics-related attribute in our deformation branch. Therefore, the full parameter set of a Gaussian can be written as:

$$\mathcal{G}_i = \{x_i, \Sigma_i, c_i, \alpha_i, s_i, k_i\}. \quad (2)$$

The final pixel color is computed by alpha-compositing all projected Gaussians along the camera ray:

$$C = \sum_{i \in \mathcal{N}} T_i \alpha_i c_i, \quad T_i = \prod_{j < i} (1 - \alpha_j), \quad (3)$$

where $\mathcal{N}$ denotes the depth-ordered Gaussians along the ray and T_i is the accumulated transmittance.

Initialization. We use TCR [8] to obtain a temporally consistent human mesh in the T-pose. Following [38], Gaussians are bound to the mesh triangles, and their centers x_c are initialized accordingly. Next, we employ the Sapiens [22] model to obtain 2D semantic maps, which are aggregated across frames using visibility-aware weights ρ to assign the semantic label to each Gaussian:

$$s = \arg \max_{n \in \mathcal{C}} \sum_{t \in \mathcal{T}} \rho_t I_t^n(\pi(x_p^t)), \quad (4)$$

where $\mathcal{C}$ denotes the set of semantic classes, $\mathcal{T}$ denotes the set of frames, and $I_t^n(\pi(x_p^t))$ denotes the class probability at the projection $\pi(\cdot)$ of the posed Gaussian center x_p in frame t . The visibility of each Gaussian is estimated based on the angle between surface normal and the viewing direction. After obtaining the semantic labels of the Gaussians, we initialize their stiffness coefficients based on empirical values. In addition, we construct a Gaussian adjacency graph that preserves the adjacency relationships of points on the reconstructed mesh, denoted as E_{ori} . Therefore, the topological adjacency graph is represented as:

$$G = (\mathcal{X}, E_{ori}), \quad (5)$$

where $\mathcal{X}$ denotes the set of Gaussians. Garment dynamics can introduce self-collisions, folds, and layering, but they do not change the intrinsic adjacency of a continuous garment surface under normal motion. The preserved topology therefore provides a stable structural prior, helping maintain local shape and material coherence during deformation.

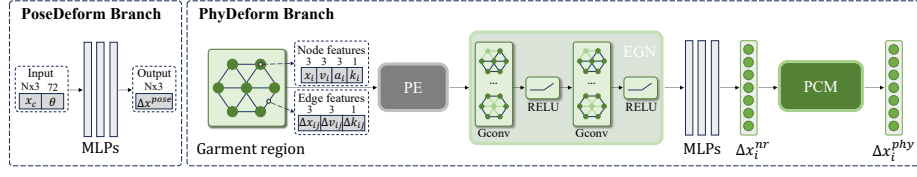


Fig. 3: Dual-branch network. The PhyDeform branch comprises a Physics Encoder (PE), an Elastic Graph Network (EGN), and a Physics Correction Module (PCM).

3.2 Articulated Gaussian Deformation

Linear Blend Skinning. Given a pose θ , we transform Gaussian centers from the canonical space x_c to the posed space x_p via linear blend skinning (LBS) [32] using skinning weights ω and joint transformations $\mathcal{B}(\theta)$:

$$x_p = \sum_{i=1}^{N_b} \omega_i \mathcal{B}(\theta)_i x_c, \quad (6)$$

where N_b denotes the number of joints. We initialize the Gaussian skinning weights using the SMPL [30] model by computing the n nearest vertices and applying distance-based weighting γ :

$$\omega_i = \sum_{j=1}^n \gamma_j \omega_{ij}^{SMPL}. \quad (7)$$

The resulting skinning weights are then treated as learnable parameters and are updated automatically during training.

Pose-Dependent Gaussian Deformation. To model non-rigid deformations, we adopt a dual-branch architecture that separately handles the deformations of the inner body and clothing. The pose-guided branch operates on body Gaussians to model relatively small pose-dependent deformations, while the physics-guided branch is applied exclusively to garment Gaussians to capture large-scale dynamic motions and inertial effects. Details of the physics-guided branch are provided in Sec. 3.3. The pose-guided deformation branch predicts offsets of Gaussian centers Δx^{pose} associated with inner body semantic labels. These offsets are inferred by an MLP conditioned on the current pose θ :

$$\Delta x^{pose} = \text{MLP}(\phi(x_c, \theta)), \quad (8)$$

where ϕ denotes the positional encoding [34].

3.3 Differentiable Physics-Guided Deformation

From fundamental physical laws such as Newton’s second law and Hooke’s law, the dynamic deformation of garments is mainly governed by two factors: body-induced accelerations that act as external driving forces, and the internal elastic

properties of the fabric. Specifically, local accelerations a generated during human movement drive garment motion, while the elastic coefficients k regulate the cloth’s deformation and recovery. Motivated by this physical insight, we propose a physics-guided Elastic Graph Network (EGN) to estimate the initial dynamic deformations of garments. While the EGN provides an initial estimate of garment deformations, it cannot fully capture complex inertial effects and may produce physically implausible predictions. To address this limitation, we introduce a physics correction module (PCM), derived from differentiable physical simulation, to refine the predicted dynamics and enforce physically plausible behavior. The overall network architecture is illustrated in Fig. 3. We describe the core design of EGN and PCM here, while deferring detailed architectures and implementation details to the supplementary material.

Physics-Guided Graph Network. We use EGN to obtain physically plausible initial predictions. The graph’s vertices $\mathcal{X}_c$ correspond to the Gaussian centers labeled with garment semantics, and its edges E_c preserve the intrinsic garment connectivity obtained from the temporally consistent garment surface:

$$G_{cloth} = (\mathcal{X}_c, E_c). \quad (9)$$

For a garment Gaussian point x_i , we define its node feature as:

$$f_i = \Phi(x_i, a_i, v_i, k_i), \quad (10)$$

where a_i and v_i are the local acceleration and velocity, and k_i is the stiffness, initialized from material priors and optimized during training. These physical quantities are passed through a lightweight Physics Encoder (PE) $\Phi(\cdot)$ implemented as an MLP to normalize the scale differences among heterogeneous inputs and to project them into a unified high-dimensional embedding space suitable for graph message passing. Since LBS provides the posed-space positions of Gaussian points at each time step, we estimate their kinematic quantities directly from these trajectories. Under the causal setting, we use backward finite differences to compute velocities and accelerations. Specifically, given the LBS-derived positions $x_i^{LBS}(t)$, the velocity and acceleration at time t are approximated as:

$$v_i(t) = \frac{x_i^{LBS}(t) - x_i^{LBS}(t-1)}{\Delta t}, \quad (11)$$

$$a_i(t) = \frac{v_i(t) - v_i(t-1)}{\Delta t}. \quad (12)$$

Following prior work on edge-conditioned message passing [48], we incorporate edge features to model relative motion and stiffness-aware interactions between neighboring Gaussians. Formally, the edge feature is defined as:

$$e_{ij} = [\Delta x_{ij}, \Delta v_{ij}, k_{ij}], \quad (13)$$

where Δx_{ij} and Δv_{ij} denote the relative position and velocity between neighboring nodes, and k_{ij} is their average stiffness.

Given the physics-embedded node features $\{f_i\}$ and edge features $\{e_{ij}\}$, we perform elastic deformation propagation on the garment graph G_{cloth} using two layers of edge-conditioned graph convolution (Edge-GConv) to model local kinematic and material interactions among neighboring Gaussians. Each Edge-GConv layer is followed by a ReLU non-linearity. At each message passing layer, the message from node j to node i is computed as

$$m_{ij} = \varphi(f_i, f_j, e_{ij}), \quad m_i = \sum_{j \in \mathcal{N}(i)} m_{ij}, \quad (14)$$

where $\varphi(\cdot)$ is implemented as an MLP encoding relative motion and stiffness-aware interactions between neighboring nodes. The messages from all neighbors of node i are aggregated using a permutation-invariant operator. The node feature is then updated as

$$f'_i = \psi(f_i, m_i), \quad (15)$$

where $\psi(\cdot)$ denotes the node update function of Edge-GConv, implemented as an MLP with residual connections. Based on the refined node embeddings after message passing, we predict a per-Gaussian non-rigid deformation offset as

$$\Delta x_i^{nr} = \mathcal{F}(f'_i), \quad (16)$$

where $\mathcal{F}(\cdot)$ is an MLP regressor. The predicted offsets represent elastic deformations and are further refined by the differentiable physics correction module.

Differentiable Physics Correction Module. Purely data-driven non-rigid deformation prediction is prone to producing physically implausible deformations, primarily due to the lack of explicit modeling of physical constraints such as inertia, elasticity, and collisions. To address this issue, we introduce a differentiable physics correction module (PCM) that refines the initial non-rigid deformation offsets predicted by the EGN by projecting them onto a physically plausible solution space. Given the predicted non-rigid offset Δx^{nr} at each frame, the physics correction module takes it as an initial estimate and outputs a corrected result Δx^{phy} that satisfies physical constraints. Traditional cloth simulation [3] models cloth as a mechanical system, which can be equivalently formulated as an energy minimization problem:

$$x^{t+1} = \arg \min_x \frac{1}{2h^2} \|x - \tilde{x}\|_M^2 + E_{phys}(x), \quad (17)$$

where h denotes the time step size, M is the mass matrix, and $\tilde{x} = x^t + hv^t + h^2 M^{-1} \mathbf{f}_{ext}$ represents the explicit prediction under external forces $\mathbf{f}_{ext}$. Gravity is the dominant external force in garment animation. The external force is therefore simplified to $\mathbf{f}_{ext} = M\mathbf{g}$. The first term enforces inertial consistency with the previous state. The physical terms $E_{phys}(x)$ encode physical energies such as elastic deformations and collision constraints, as formulated below:

$$E_{phys}(x) = E_{stretch}(x) + E_{bend}(x) + E_{contact}(x), \quad (18)$$

where $E_{stretch}(x)$ and $E_{bend}(x)$ penalize in-plane stretching and out-of-plane bending [2], respectively, while $E_{contact}(x)$ is our designed contact energy term, which prevents interpenetration while modeling frictional effects.

In cloth simulation, particularly for loose garments, accurate modeling of contact and friction is crucial for visual realism. Inspired by [28], we design a contact term consisting of a penetration penalty and frictional dissipation:

$$E_{contact}(x) = \sum_{j=1}^N w_c(d_j) \left[\frac{k_n}{2} \max(0, \epsilon - d_j)^2 + \mu k_t \sqrt{\|\Delta x_{j|T}\|^2 + \delta^2} \right], \quad (19)$$

where d_j is the signed distance from the j -th garment point to the body surface, k_n and k_t are normal and tangential penalty stiffnesses, μ is the friction coefficient, $\Delta x_{j|T}$ is the tangential displacement increment on the contact surface, and δ is a small constant ensuring differentiability. To enable end-to-end differentiability, we replace discrete contact indicators with continuous contact weights, softened using a sigmoid function, defined as follows:

$$w_c(d_j) = \text{sigmoid}\left(\frac{\epsilon - d_j}{\tau}\right), \quad (20)$$

where ϵ is the contact threshold and τ controls the degree of softening. This formulation ensures stable gradient-based optimization while approximating the Signorini–Coulomb contact law [31]. Meanwhile, we fix the number of iterations to enable efficient automatic differentiation. We minimize the total energy E_{total} using gradient descent with a fixed iteration scheme to ensure a static computational graph for efficient automatic differentiation. Starting from the neural network prediction $\Delta x^{(0)} = \Delta x^{nr}$, we perform K optimization iterations:

$$\Delta x^{(k+1)} = \Delta x^{(k)} - \eta^{(k)} \nabla E_{total}(\Delta x^{(k)}), \quad k = 0, \dots, K-1 \quad (21)$$

where $\eta^{(k)}$ is the learning rate at iteration k . The final output of the physics-guided deformation branch is $\Delta x^{phy} = \Delta x^{(K)}$.

3.4 Training

Given the training images I and the corresponding pose inputs θ , we optimize a Gaussian avatar representation by minimizing the following objective function:

$$\mathcal{L} = \mathcal{L}_{rgb} + \lambda_{lips} \mathcal{L}_{lips} + \lambda_{normal} \mathcal{L}_{normal} + \lambda_{topo} \mathcal{L}_{topo} + \lambda_{sem} \mathcal{L}_{sem}, \quad (22)$$

where λ denotes the corresponding loss weights. Here, $\mathcal{L}_{rgb}$ denotes the L1 photometric loss between rendered and ground-truth images, $\mathcal{L}_{lips}$ [60] measures perceptual similarity, and $\mathcal{L}_{normal}$ enforces consistency of surface normal orientations. To prevent the destruction of the underlying topological structure during training, we introduce a topology-preserving loss that leverages the pre-computed adjacency graph obtained during initialization. The formulation is:

$$\mathcal{L}_{topo} = \sum_{(i,j) \in E_c} \exp(\beta \|x_i - x_j\|_2 - d_{ij}), \quad (23)$$

Table 1: Quantitative evaluation. Quantitative comparison with state-of-the-art methods on three public datasets, including ablation studies on individual modules. **Best** and **second-best** results are highlighted in green and orange, respectively.

Method	ZJU-MoCap [41]			PeopleSnapshot [1]			ActorsHQ (01) [17]			
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	CD $\downarrow$
Vid2Avatar [12]	28.70	0.969	0.0521	28.52	0.964	0.0546	27.50	0.956	0.0580	2.450
InstantAvatar [18]	29.47	0.926	0.0488	28.65	0.968	0.0331	27.30	0.925	0.0624	2.956
GauHuman [16]	31.78	0.966	0.0335	30.52	0.956	0.0338	29.10	0.955	0.0358	1.980
iHuman [38]	30.32	0.955	0.0332	28.75	0.963	0.0342	29.87	0.955	0.0380	1.575
GART [26]	29.99	0.970	0.0327	28.40	0.967	0.0377	29.40	0.955	0.0372	2.254
GomAvatar [52]	30.22	0.958	0.0325	30.12	0.973	0.0295	28.75	0.952	0.0368	1.923
GaussianAvatar [15]	31.57	0.973	0.0321	30.54	0.976	0.0345	29.30	0.958	0.0364	1.852
Baseline	30.53	0.959	0.0345	30.13	0.963	0.0362	30.15	0.955	0.0365	1.493
Baseline+EGN	31.76	0.967	0.0319	31.44	0.975	0.0271	31.13	0.959	0.0352	1.465
Baseline+EGN+PCM	32.21	0.972	0.0302	31.87	0.985	0.0285	31.54	0.967	0.0328	1.398

where d_{ij} denotes the rest-pose distance between Gaussian points and β controls the strength of the exponential penalty, suppressing excessive topological changes without overly constraining normal deformations. The semantic loss $\mathcal{L}_{sem}$ is computed by volume-rendering the predicted semantic logits into a 2D map and applying a cross-entropy loss against the ground-truth segmentation. To improve training stability, we adopt a two-stage avatar fitting strategy. In the first stage, the PCM module is disabled, and the Gaussian avatar together with the EGN is optimized to obtain a stable geometry and appearance initialization. In the second stage, we freeze the parameters of the EGN and introduce the PCM module for further training, enabling dynamic refinement.

4 Experiments

In this section, we present a comprehensive evaluation of our method on three public benchmark datasets. All experiments are conducted on an Intel Xeon Gold 6226R CPU and an NVIDIA GeForce RTX 3090 GPU.

4.1 Datasets and Metrics

Datasets. We evaluate our approach on the ZJU-MoCap [41], PeopleSnapshot [1], and ActorsHQ [17] datasets. The ZJU-MoCap dataset consists of dynamic human videos captured by more than 20 synchronized cameras. It provides foreground masks, camera poses, and SMPL-X parameters. For each subject, we use two-thirds of the frames from camera 0 for training, while the remaining frames are reserved for evaluation. The PeopleSnapshot dataset provides monocular videos in which the subject slowly rotates through 360 degrees. Similarly, we use the first two-thirds of the frames for training, and reserve the remaining frames for evaluation. The ActorsHQ dataset consists of multi-view videos of dynamic human performances from eight subjects. The camera system comprises 160 cameras. The dataset also provides per-frame mesh reconstructions

Table 2: Ablation study. Ablation results of the loss function. Values in parentheses show relative percentage changes compared with the model using all loss terms. Red and green indicate performance **degradation** and **improvement**, respectively.

$\mathcal{L}_{rgb}$	$\mathcal{L}_{ssim}$	$\mathcal{L}_{lips}$	$\mathcal{L}_{normal}$	$\mathcal{L}_{topo}$	$\mathcal{L}_{sem}$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	CD $\downarrow$
✓	✗	✓	✓	✓	✓	32.13 (-0.16%)	0.970 (-0.10%)	0.0294 (-1.34%)	1.084 (-0.37%)
✓	✓	✗	✓	✓	✓	31.72 (-1.43%)	0.964 (-0.72%)	0.0568 (+90.60%)	1.197 (+10.02%)
✓	✓	✓	✗	✓	✓	30.98 (-3.73%)	0.952 (-1.96%)	0.0415 (+39.26%)	1.348 (+23.90%)
✓	✓	✓	✓	✗	✓	31.15 (-3.20%)	0.956 (-1.54%)	0.0402 (+34.90%)	1.315 (+20.86%)
✓	✓	✓	✓	✓	✗	31.40 (-2.42%)	0.960 (-1.13%)	0.0349 (+17.11%)	1.124 (+3.31%)
✓	✓	✓	✓	✓	✓	32.18	0.971	0.0298	1.088

for each actor, which can be used to evaluate the geometric accuracy of our reconstructions. We use the first two-thirds of the frames captured by camera 127 for training, and reserve the remaining frames for evaluation.

Metrics. We evaluate the quality of synthesized human images under novel poses using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) [51], and Learned Perceptual Image Patch Similarity (LPIPS) [60], which respectively measure pixel-wise fidelity, structural consistency, and perceptual similarity. To evaluate geometric accuracy, we compute the Chamfer Distance (CD) between Gaussian centers and the ground-truth mesh surface.

4.2 Experimental Evaluation

Quantitative Evaluation. Quantitative comparisons with existing monocular approaches are presented in Tab. 1. Rendering quality is evaluated on ZJU-MoCap [41] and PeopleSnapshot [1] using PSNR, SSIM, and LPIPS, while geometric accuracy is measured on subject 01 of ActorsHQ [17] using Chamfer Distance (CD). Subject 01 features a female performer wearing a skirt and is reported separately to reflect performance under loose-clothing conditions. Our method achieves competitive rendering quality over prior approaches, with more pronounced gains in loose-garment scenarios due to physics-guided dynamic modeling. The CD results further suggest that temporally consistent initialization and topology preservation improve geometric accuracy.

Qualitative Evaluation. Fig. 4 shows qualitative results on four subjects from the PeopleSnapshot [1] and ActorsHQ [17] datasets, compared with recent monocular Gaussian avatar methods. In comparison, GaussianAvatar [15] and iHuman [38] tend to produce fragmented and structurally inconsistent garment regions under large pose changes, especially around skirt boundaries and loose cloth regions. The orange annotations in Fig. 4 highlight representative failure cases. Thanks to topological preservation, our method maintains structural integrity across a wide range of poses. Fig. 5 shows a skirted subject from an in-the-wild video performing a rotation, where our method captures plausible inertial garment dynamics, demonstrating its generalization.

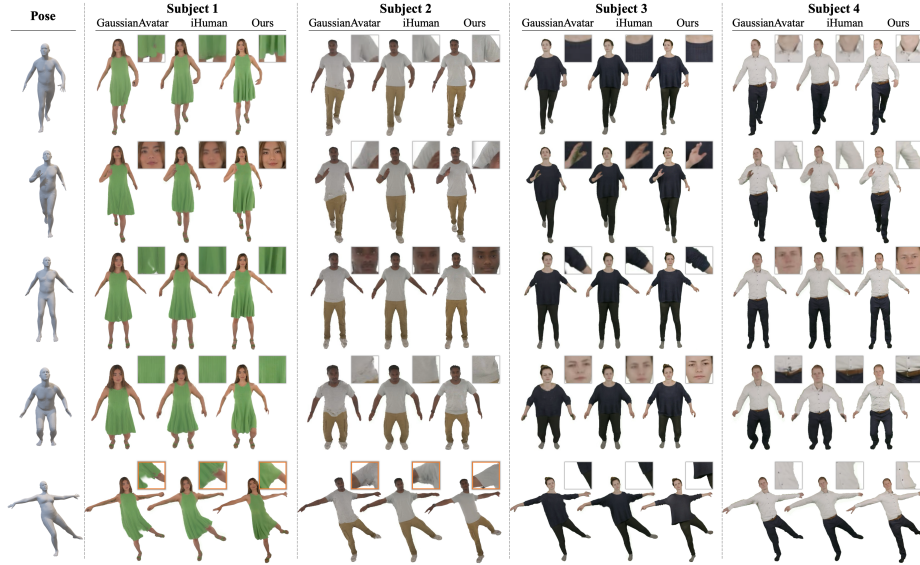


Fig. 4: Qualitative comparison with state-of-the-art methods. Our approach was qualitatively compared with GaussianAvatar [15] and iHuman [38].

4.3 Ablation Study

Module Ablation. To demonstrate the effectiveness of our physics-guided non-rigid deformation modules, we report ablation results in Tab. 1. We adopt a reconstruction framework with temporally consistent reconstruction initialization as the baseline, and evaluate the performance gains obtained by incorporating the Elastic Graph Network (EGN) and the Physics Correction Module (PCM). The results indicate that both modules play an important role in improving the overall performance. EGN leverages physical parameters to guide garment dynamics, providing more realistic motion, while PCM employs simulation principles to model inertial effects, correcting unrealistic garment responses.

Loss Function Ablation. To demonstrate the contribution of each loss term, we present the ablation results of individual loss components in Tab. 2. The ablation experiments are conducted on subjects 01, 02, and 05 of the ActorsHQ dataset [17], and we report the average performance across these subjects. As shown by the results, the proposed topological preservation loss plays a crucial role in improving overall performance. $\mathcal{L}_{ssim}$ has a marginal influence on the overall performance. Removing this term slightly decreases PSNR and SSIM, while yielding minor improvements in LPIPS and CD. Considering its limited contribution, we exclude it from the final loss function. Moreover, the Chamfer Distance (CD) results indicate that the normal consistency loss $\mathcal{L}_{normal}$ plays an important role in improving geometric fidelity.



Fig. 5: Qualitative results. We showcase challenging long-skirt examples. To highlight the physics-guided dynamic behavior, the Gaussian avatar is driven by a single rotational motion.



Fig. 6: Semantic garment transfer. Our method enables cross-subject replacement and adaptation of garment Gaussians with aligned clothing semantics, resulting in realistic animations.

4.4 Applications

Garment Retargeting Application. We assign semantic labels to Gaussians, which allows us to disentangle garment-related ones. As a result, our framework can be naturally extended to garment transfer applications. For each garment Gaussian, we identify its corresponding body-part label in the parametric model and perform position transformation based on the label, enabling the garment Gaussian template to be aligned with a new subject. Furthermore, we compensate for body shape differences estimated by the parametric model to adjust the Gaussian positions, thereby adapting the transferred garments to body-shape variations. As shown in Fig. 6, the avatar with transferred garments generates natural and realistic animations. The implementation details of garment retargeting are provided in the supplementary material.

5 Conclusion and Discussion

Summary. In this paper, we presented PhysGA, a physics-guided avatar reconstruction framework. We model dynamic garment behavior under coupled physical and topological constraints, enabling non-rigid Gaussian deformation to exhibit physically plausible and structurally coherent dynamics. The semantic-aware Gaussian representation naturally supports applications such as garment transfer. In addition, the framework maintains practical training and inference efficiency, with detailed runtime results provided in the supplementary material.

Limitations. Our framework uses semantic parsing and mesh extraction as upstream modules, whose accuracy may affect reconstruction quality. Nevertheless, they are modular and replaceable by stronger alternatives. In addition, the texture representation heavily depends on the human observations in the input video sequence, causing regions that are rarely observed or entirely missing to suffer from texture degradation or incompleteness. Addressing these limitations remains an important direction for future research.

References

1. Alldieck, T., Magnor, M., Xu, W., Theobalt, C., Pons-Moll, G.: Video based reconstruction of 3d people models. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8387–8397 (Jun 2018), CVPR Spotlight Paper
2. Baraff, D., Witkin, A.: Large steps in cloth simulation. In: Seminal Graphics Papers: Pushing the Boundaries, Volume 2, pp. 767–778 (2023)
3. Bouaziz, S., Martin, S., Liu, T., Kavan, L., Pauly, M.: Projective dynamics: Fusing constraint projections for fast simulation. In: Seminal Graphics Papers: Pushing the Boundaries, Volume 2, pp. 787–797 (2023)
4. Budria, A., Lopez-Rodriguez, A., Lorente, Ó., Moreno-Noguer, F.: Instantgeoavatar: effective geometry and appearance modeling of animatable avatars from monocular video. In: Proceedings of the Asian Conference on Computer Vision. pp. 1827–1848 (2024)
5. Chen, X., Jiang, T., Song, J., Rietmann, M., Geiger, A., Black, M.J., Hilliges, O.: Fast-snarf: A fast deformer for articulated neural fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(10), 11796–11809 (2023)
6. Choi, K.J., Ko, H.S.: Stable but responsive cloth. In: ACM SIGGRAPH 2005 Courses, pp. 1–es (2005)
7. Chung, J., Lee, K., Baik, S., Lee, K.M.: Meil-nerf: Memory-efficient incremental learning of neural radiance fields. *IEEE Access* (2025)
8. Deng, Y., Li, B., Yang, Y., Zhao, X.: Temporally consistent reconstruction of 3d clothed human surface with warp field. *Image and Vision Computing* **137**, 104782 (2023)
9. Feng, G., Chen, S., Fu, R., Liao, Z., Wang, Y., Liu, T., Hu, B., Xu, L., Pei, Z., Li, H., et al.: Flashgs: Efficient 3d gaussian splatting for large-scale and high-resolution rendering. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 26652–26662 (2025)
10. Feng, Q., Liu, Y., Lai, Y.K., Yang, J., Li, K.: Fof: Learning fourier occupancy field for monocular real-time human reconstruction. *Advances in Neural Information Processing Systems* **35**, 7397–7409 (2022)
11. Gu, M., Li, J., Wu, Y., Luo, H., Zheng, J., Bai, X.: 3d human avatar reconstruction with neural fields: A recent survey. *Image and Vision Computing* **154**, 105341 (2025)
12. Guo, C., Jiang, T., Chen, X., Song, J., Hilliges, O.: Vid2avatar: 3d avatar reconstruction from videos in the wild via self-supervised scene decomposition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12858–12868 (2023)
13. Guo, C., Li, J., Kant, Y., Sheikh, Y., Saito, S., Cao, C.: Vid2avatar-pro: Authentic avatar from videos in the wild via universal prior. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5559–5570 (2025)
14. Guo, M., Chiang, M.J.Y., Santesteban, I., Sarafianos, N., Chen, H.y., Halimi, O., Božič, A., Saito, S., Wu, J., Liu, C.K., Stuyck, T., Larionov, E.: Pgc: Physics-based gaussian cloth from a single pose. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21215–21225 (2025)
15. Hu, L., Zhang, H., Zhang, Y., Zhou, B., Liu, B., Zhang, S., Nie, L.: Gaussianavatar: Towards realistic human avatar modeling from a single video via animatable 3d gaussians. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 634–644 (2024)

16. Hu, S., Hu, T., Liu, Z.: Gauhuman: Articulated gaussian splatting from monocular human videos. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20418–20431 (2024)
17. Işık, M., Rünz, M., Georgopoulos, M., Khakhulin, T., Starck, J., Agapito, L., Nießner, M.: Humanrf: High-fidelity neural radiance fields for humans in motion. *ACM Transactions on Graphics (TOG)* **42**(4), 1–12 (2023)
18. Jiang, T., Chen, X., Song, J., Hilliges, O.: Instantavatar: Learning avatars from monocular video in 60 seconds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16922–16932 (2023)
19. Jiang, W., Yi, K.M., Samei, G., Tuzel, O., Ranjan, A.: Neuman: Neural human radiance field from a single video. In: European Conference on Computer Vision. pp. 402–418. Springer (2022)
20. Karthikeyan, A., Ren, R., Kant, Y., Gilitschenski, I.: Avatarone: Monocular 3d human animation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3647–3657 (2024)
21. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**(4), 139–1 (2023)
22. Khirodkar, R., Bagautdinov, T., Martinez, J., Zhaoen, S., James, A., Selednik, P., Anderson, S., Saito, S.: Sapiens: Foundation for human vision models. In: European Conference on Computer Vision. pp. 206–228. Springer (2024)
23. Kocabas, M., Chang, J.H.R., Gabriel, J., Tuzel, O., Ranjan, A.: Hugs: Human gaussian splats. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 505–515 (2024)
24. Lassner, C., Zollhofer, M.: Pulsar: Efficient sphere-based neural rendering. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1440–1449 (2021)
25. Lee, C., Lee, J., Kim, T.K.: Mpmavatar: Learning 3d gaussian avatars with accurate and robust physics-based dynamics. *arXiv preprint arXiv:2510.01619* (2025)
26. Lei, J., Wang, Y., Pavlakos, G., Liu, L., Daniilidis, K.: Gart: Gaussian articulated template models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 19876–19887 (2024)
27. Li, R., Tanke, J., Vo, M., Zollhöfer, M., Gall, J., Kanazawa, A., Lassner, C.: Tava: Template-free animatable volumetric actors. In: European Conference on Computer Vision. pp. 419–436. Springer (2022)
28. Li, Y., Du, T., Wu, K., Xu, J., Matusik, W.: Diffcloth: Differentiable cloth simulation with dry frictional contact. *ACM Transactions on Graphics (TOG)* **42**(1), 1–20 (2022)
29. Lombardi, S., Simon, T., Schwartz, G., Zollhoefer, M., Sheikh, Y., Saragih, J.: Mixture of volumetric primitives for efficient neural rendering. *ACM Transactions on Graphics (ToG)* **40**(4), 1–13 (2021)
30. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: Smpl: A skinned multi-person linear model. In: *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pp. 851–866 (2023)
31. Ly, M., Jouve, J., Boissieux, L., Bertails-Descoubes, F.: Projective dynamics with dry frictional contact. *ACM Transactions on Graphics (TOG)* **39**(4), 57–1 (2020)
32. Magnenat-Thalmann, N., Laperrière, R., Thalmann, D.: Joint-dependent local deformations for hand animation and object grasping. In: *Proceedings on Graphics interface’88*. pp. 26–33 (1989)
33. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. In: *Proceedings of the*

- IEEE/CVF conference on computer vision and pattern recognition. pp. 4460–4470 (2019)
34. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65**(1), 99–106 (2021)
 35. Moon, G., Shiratori, T., Saito, S.: Expressive whole-body 3d gaussian avatar. In: *European Conference on Computer Vision*. pp. 19–35. Springer (2024)
 36. Osman, A.A., Bolkart, T., Black, M.J.: Star: Sparse trained articulated human body regressor. In: *European Conference on Computer Vision*. pp. 598–613. Springer (2020)
 37. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: Learning continuous signed distance functions for shape representation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 165–174 (2019)
 38. Paudel, P., Khanal, A., Paudel, D.P., Tandukar, J., Chhatkuli, A.: ihuman: Instant animatable digital humans from monocular videos. In: *European Conference on Computer Vision*. pp. 304–323. Springer (2024)
 39. Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A.A., Tzionas, D., Black, M.J.: Expressive body capture: 3d hands, face, and body from a single image. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 10975–10985 (2019)
 40. Peng, B., Tao, Y., Zhan, H., Guo, Y., Zhang, J.: Pica: Physics-integrated clothed avatar. *IEEE Transactions on Visualization and Computer Graphics* (2025)
 41. Peng, S., Zhang, Y., Xu, Y., Wang, Q., Shuai, Q., Bao, H., Zhou, X.: Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 9054–9063 (2021)
 42. Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: D-nerf: Neural radiance fields for dynamic scenes. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 10318–10327 (2021)
 43. Rai, A., Wang, D., Jain, M., Sarafianos, N., Chen, K., Sridhar, S., Prakash, A.: Uvgs: Reimagining unstructured 3d gaussian splatting using uv mapping. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 5927–5937 (2025)
 44. Rong, B., Grigorev, A., Wang, W., Black, M.J., Thomaszewski, B., Tsalioglou, C., Hilliges, O.: Gaussian garments: Reconstructing simulation-ready clothing with photorealistic appearance from multi-view video. In: *2025 International Conference on 3D Vision (3DV)*. pp. 1054–1063. IEEE (2025)
 45. Shao, Z., Wang, Z., Li, Z., Wang, D., Lin, X., Zhang, Y., Fan, M., Wang, Z.: Splattingavatar: Realistic real-time human avatars with mesh-embedded gaussian splatting. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 1606–1616 (2024)
 46. Shentu, Y., Zhang, S., Xu, M., Shuai, Q., Chen, T., Peng, S., Zhou, X.: Dyn-e: Local appearance editing of dynamic neural radiance fields. *Computers & Graphics* **126**, 104140 (2025)
 47. Shim, J., Kang, C., Joo, K.: Diffusion-based signed distance fields for 3d shape generation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 20887–20897 (2023)
 48. Simonovsky, M., Komodakis, N.: Dynamic edge-conditioned filters in convolutional neural networks on graphs. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3693–3702 (2017)

49. Svitov, D., Morerio, P., Agapito, L., Del Bue, A.: Haha: Highly articulated gaussian human avatars with textured mesh prior. In: *Proceedings of the Asian Conference on Computer Vision*. pp. 4051–4068 (2024)
50. Wang, R., Prada, F., Wang, Z., Jiang, Z., Yin, C., Li, J., Saito, S., Santesteban, I., Romero, J., Joshi, R., et al.: Fresa: Feedforward reconstruction of personalized skinned avatars from few images. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 281–291 (2025)
51. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)
52. Wen, J., Zhao, X., Ren, Z., Schwing, A.G., Wang, S.: Gomavatar: Efficient animatable human modeling from monocular video using gaussians-on-mesh. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 2059–2069 (2024)
53. Weng, C.Y., Curless, B., Srinivasan, P.P., Barron, J.T., Kemelmacher-Shlizerman, I.: Humannerf: Free-viewpoint rendering of moving people from monocular video. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern Recognition*. pp. 16210–16220 (2022)
54. Xiu, Y., Yang, J., Cao, X., Tzionas, D., Black, M.J.: ECON: Explicit Clothed humans Optimized via Normal integration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2023)
55. Xu, H., Alldieck, T., Sminchisescu, C.: H-nerf: Neural radiance fields for rendering and temporal reconstruction of humans in motion. *Advances in Neural Information Processing Systems* **34**, 14955–14966 (2021)
56. Yin, Z.X., Jiao, P.Y., Qiu, J., Cheng, M.M., Ren, B.: Ms-nerf: Multi-space neural radiance fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025)
57. Yu, X., Tian, X., Chen, J., Wang, Y.: Freqspace-nerf: A fourier-enhanced neural radiance fields method via dual-domain contrastive learning for novel view synthesis. *Computers & Graphics* **127**, 104171 (2025)
58. Yu, Z., Cheng, W., Liu, X., Wu, W., Lin, K.Y.: Monohuman: Animatable human neural field from monocular video. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 16943–16953 (2023)
59. Zhan, Y., Shao, T., Yang, Y., Zhou, K.: Real-time high-fidelity gaussian human avatars with position-based interpolation of spatially distributed mlps. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 26297–26307 (2025)
60. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 586–595 (2018)
61. Zhang, Y., Jia, W., Niu, W., Yin, M.: Gaussianspa: An "optimizing-sparsifying" simplification framework for compact and high-quality 3d gaussian splatting. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 26673–26682 (2025)
62. Zheng, Y., Zhao, Q., Yang, G., Wang, Y., Xiang, D., Dubost, F., Lagun, D., Beeler, T., Tombari, F., Guibas, L., Wetzstein, G.: Physavatar: Learning the physics of dressed 3d avatars from visual observations. In: *European Conference on Computer Vision*. pp. 262–284. Springer (2024)
63. Zheng, Z., Huang, H., Yu, T., Zhang, H., Guo, Y., Liu, Y.: Structured local radiance fields for human avatar modeling. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 15893–15903 (2022)