

# 基于3D高斯溅射的三维人体重建算法研究

王文龙

浙江师范大学计算机科学与技术学院, 浙江 金华

收稿日期: 2026年3月9日; 录用日期: 2026年4月9日; 发布日期: 2026年4月20日

## 摘要

三维人体重建在数字人、动作捕捉与沉浸式交互等应用中具有重要价值, 但真实场景下衣物褶皱、软组织位移及快速运动带来的非刚性形变, 使得仅依赖参数化人体模型的模板驱动方法难以获得稳定且细节贴合的重建结果。本文提出形变修正的高斯人体模型DCGaussianAvatar, 结合3D高斯溅射的高效渲染能力与SMPL结构先验, 实现可驱动的人体重建与新视角合成。该方法首先在SMPL网格表面初始化高斯基元, 并通过姿态编码器提取全局姿态嵌入; 在此基础上, 设计姿态条件的非刚性形变矫正模块, 以高斯中心位置嵌入与姿态嵌入构成的高级特征为输入, 分别对蒙皮权重与外观参数进行残差式更新, 从而补偿衣物与软组织等非刚性细节带来的残差误差。在ZJU-MoCap数据集上与多种代表性方法进行对比, 结果表明DCGaussianAvatar在PSNR上取得最高值30.93, 同时SSIM达到0.964、LPIPS为0.032, 并在衣物褶皱与关节区域呈现更稳定的对齐效果与更清晰的细节纹理。

## 关键词

三维人体重建, 3D高斯溅射, SMPL, 非刚性形变矫正

# Research on 3D Human Reconstruction Algorithms Based on 3D Gaussian Splatting

Wenlong Wang

School of Computer Science and Technology, Zhejiang Normal University, Jinhua Zhejiang

Received: March 9, 2026; accepted: April 9, 2026; published: April 20, 2026

## Abstract

3D human reconstruction is of great importance for applications such as digital humans, motion capture, and immersive interaction. However, non-rigid deformations caused by clothing wrinkles, soft-tissue motion, and fast body movements in real-world scenarios make it difficult for template-driven methods that rely solely on parametric human models to achieve stable reconstructions with

well-aligned fine details. To address this issue, we propose DCGaussianAvatar, a deformation-corrected Gaussian avatar model that combines the efficient rendering capability of 3D Gaussian Splatting with the structural prior of SMPL for animatable human reconstruction and novel-view synthesis. Our method initializes Gaussian primitives on the SMPL mesh surface and extracts a global pose embedding via a pose encoder. Building on this, we design a pose-conditioned non-rigid deformation correction module, which takes high-level features formed by concatenating the embedded Gaussian centers and the pose embedding, and performs residual updates for both skinning weights and appearance parameters to compensate for residual errors introduced by non-rigid details such as clothing and soft tissues. Experiments on the ZJU-MoCap dataset against multiple representative baselines show that DCGaussianAvatar achieves the best PSNR of 30.93, along with an SSIM of 0.964 and an LPIPS of 0.032, and delivers more stable alignment and sharper fine-grained details, particularly around clothing wrinkles and articulated joints.

## Keywords

3D Human Reconstruction, 3D Gaussian Splatting, SMPL, Non-Rigid Deformation Correction

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

三维人体重建是计算机视觉、图形学与虚拟现实领域的重要基础问题，在数字人[1]、动作捕捉[2]、影视游戏[3]、沉浸式交互[4]以及虚拟试衣[5]等场景中具有广泛的应用价值。然而，真实场景下人体存在姿态变化频繁、遮挡严重、衣物非刚性形变显著以及复杂光照变化等因素，使得重建过程难以恢复人体的几何与外观细节。因此，如何在有限观测条件下重建出高质量、稳定且具有可解释性的三维人体模型，仍然具有重要研究意义。

传统的人体重建方法大多采用显式网格或点云表示，并结合多视角几何或模板拟合完成重建。该方法通常具有较强的可解释性，但在复杂衣物、细节纹理与快速运动条件下，往往需要引入形变建模或额外的先验约束才能获得高保真的人体模型。近年来，隐式表示与可微渲染方法的兴起推动了人体重建在细节保真度方面的提升，例如基于神经辐射场(Neural Radiance Fields, NeRF) [6]的人体渲染框架能够通过端到端优化实现视角一致的新视角合成，但其训练与渲染开销较大，往往需要较长的优化时间，难以满足实时应用需求。与之相比，3D 高斯溅射(3D Gaussian Splatting, 3DGS) [7]以各向异性的高斯点作为基本渲染基元，并采用高效的栅格化渲染流程，在保证较高视觉质量的同时显著提升了训练与渲染效率，因而逐渐成为静态场景下三维重建与新视角合成的重要方向。

尽管 3DGS 在静态场景中表现突出，但将其直接用于动态人体仍面临关键挑战：一方面，人体的骨骼运动可通过参数化人体模型 SMPL [8]进行描述，但 SMPL 的线性混合蒙皮(Linear Blend Skinning, LBS)在宽松衣物、软组织形变与局部褶皱等非刚性因素下表达能力受限；另一方面，SMPL 姿态参数的初始误差会导致渲染对齐偏差，从而出现伪影与细节模糊现象，使得模型在渲染误差的优化过程中更难收敛。此外，固定的 LBS 权重难以适配不同个体与服饰类型，也限制了模型重建精度的进一步提升。因此，如何在引入 SMPL 先验的同时，兼顾 3DGS 的高效渲染优势，并增强对非刚性形变的鲁棒建模，是当前基于 3DGS 的三维人体重建研究中亟需解决的问题。

针对上述问题，本文结合 3DGS 与 SMPL 参数化人体先验，提出一种形变修正的高斯人体模型

**DCGaussianAvatar**: 首先引入姿态参数编码器对 SMPL 姿态进行特征化表达, 以提供更稳定的优化方向; 随后设计非刚性形变矫正模块, 在 LBS 变形的基础上学习残差形变以补偿衣物与软组织等非线性位移, 并进一步联合优化姿态参数与蒙皮权重, 以提升跨视角对齐精度与局部细节表达能力。本文在 ZJU-MoCap 数据集的多个序列数据上进行实验验证, 结果表明所提方法在渲染质量与训练速度方面均取得了更优表现。

## 2. 相关工作

### 2.1. 基于隐式表示的三维人体重建算法

基于隐式表示的三维人体重建方法通常利用隐式函数对三维人体的几何与外观进行统一编码, 例如符号距离场(Signed Distance Function, SDF) [9] [10]、占据场(Occupancy Field) [11] [12]和神经辐射场 NeRF 等。在众多隐式表示中, NeRF 相关方法已成为当前最主流的技术路线之一, 它通过可微体渲染直接从视频或多视图图像中端到端学习人体的辐射场表示, 能够在无需显式网格拓扑的情况下实现高质量、跨视角一致的新视角合成, 并在细节保真与视角一致性方面展现出显著优势。

为了刻画人体的运动, 许多 NeRF 相关的方法使用 SMPL 提供的几何先验, 将不同姿态映射到共享的规范空间, 并结合逆向 LBS 变换与残差形变场构建可驱动的人体模型。该类方法通常将形变分解为骨骼驱动的刚性运动与额外的非刚性形变两部分: 例如 HumanNeRF [13]通过骨骼形变与非刚性形变联合建模人体运动, 并融合相邻视角的颜色与深度信息以增强纹理细节; Anim-NeRF [14]则学习姿态条件的逆向 LBS 形变场, 以进一步捕获更精细的局部几何与外观细节。尽管上述方法在新视角合成任务上取得了良好的渲染质量, 但需要通过多层感知机(Multilayer Perceptron, MLP)对采样点进行密集查询, 导致人体表面的高频细节易被过度平滑而出现模糊现象, 同时渲染效率受限, 难以满足高效推理与实时交互的需求。

### 2.2. 基于 3D 高斯溅射的三维人体重建算法

3D 高斯溅射以一组可学习的高斯基元表示场景, 每个基元携带位置、尺度与方向等几何属性, 并以不透明度与颜色刻画外观。相较于 NeRF 依赖体渲染的密集采样与积分计算, 3DGS 采用基于栅格化的前向渲染与屏幕空间的加权合成, 在保持视角一致的高质量渲染效果的同时显著降低计算开销, 从而成为 NeRF 在实时渲染与静态场景下的一种有效替代方案。

近年来, 研究者逐步将 3DGS 从静态场景扩展为可驱动人体建模。在相关研究中, 一个主流思路是引入参数化人体模型或模板网格作为结构先验, 以约束高斯基元的空间分布与运动方式, 从而在动态场景中获得更稳定的对齐与更可控的姿态驱动。例如, ASH [15]将高斯点附着于可变形的人体模板, 并在纹理空间学习高斯参数以获得更高效的训练与实时渲染能力, 从而实现姿态可控的运动形变。SplattingAvatar [16]则将高斯点嵌入网格表面, 通过网格显式控制人体的低频运动, 并由高斯点表示高频几何和外观细节, 实现了高帧率的人体模型渲染。

尽管模板先验能够显著提升模型驱动稳定性和跨帧一致性, 但由于其形变主要受骨骼与蒙皮约束, 难以建模人体衣物、头发和软组织位移等非刚性形变。为增强对非刚性形变的表达能力, GART [17]在关节化模板的基础上设计了一个可学习的驱动机制, 通过引入未观测的潜在骨骼和可学习的前向蒙皮来驱动高斯点的非刚性形变。3DGS-Avatar [18]则构建一个姿势相关的非刚性形变网络, 通过编码姿态参数使形变网络能感知全局姿态, 有效建模衣物随姿态变化的褶皱细节。然而, 上述方法在复杂环境遮挡和极端姿态变化条件下仍可能出现局部细节模糊或像素对齐偏差等问题, 模型鲁棒性和泛化能力仍有提升空间。总体而言, 如何更准确、稳定地建模非刚性形变仍是三维人体重建领域的研究热点。

与现有基于 3DGS 的人体建模方法相比, 本文提出的 DCGaussianAvatar 在模型设计与残差矫正机制上具有明确优势。首先, SplattingAvatar [16]通过“网格嵌入高斯”的混合表示将低频运动交由模板网格

控制,但其外观与运动主要受网格驱动约束,对衣物褶皱与软组织位移等非刚性细节的自适应修正能力相对有限。其次,3DGS-Avatar [18]采用姿态相关的非刚性形变网络在高斯空间中学习形变场,以增强跨姿态的几何和外观适配能力,但该方法需要较长的训练时间,难以满足实时应用场景,难以。不同于上述两类思路,DCGaussianAvatar以SMPL提供的可解释结构驱动为基础,引入姿态编码器获得全局姿态嵌入,并通过残差更新的方式协同矫正:一方面通过学习蒙皮权重残差在关节维度上自适应调整高斯的运动分配,另一方面通过外观残差更新对姿态变化引起的局部纹理与明暗细节进行补偿,从而在保持模板先验稳定性的同时提升非刚性细节贴合与像素对齐一致性。该差异化设计使本文方法更强调在先验约束附近进行可控修正,兼顾训练效率与重建质量。

### 3. 方法

#### 3.1. 方法概述

本文提出了一种形变修正的高斯人体模型DCGaussianAvatar,旨在结合3D高斯喷射的高效渲染能力与SMPL参数化人体先验,重建出高保真、可驱动的三维人体模型。该方法在SMPL提供的可解释运动驱动基础上,引入姿态参数编码器对全局姿态进行特征化表达,并设计姿态条件的非刚性形变矫正模块,用于联合更新LBS蒙皮权重与高斯外观属性,以补偿衣物褶皱、软组织位移等非刚性细节带来的残差误差。最后,通过可微高斯渲染与重建损失的端到端优化,模型能够在动态姿态变化下保持稳定对齐与细节贴合。DCGaussianAvatar模型框架如图1所示。

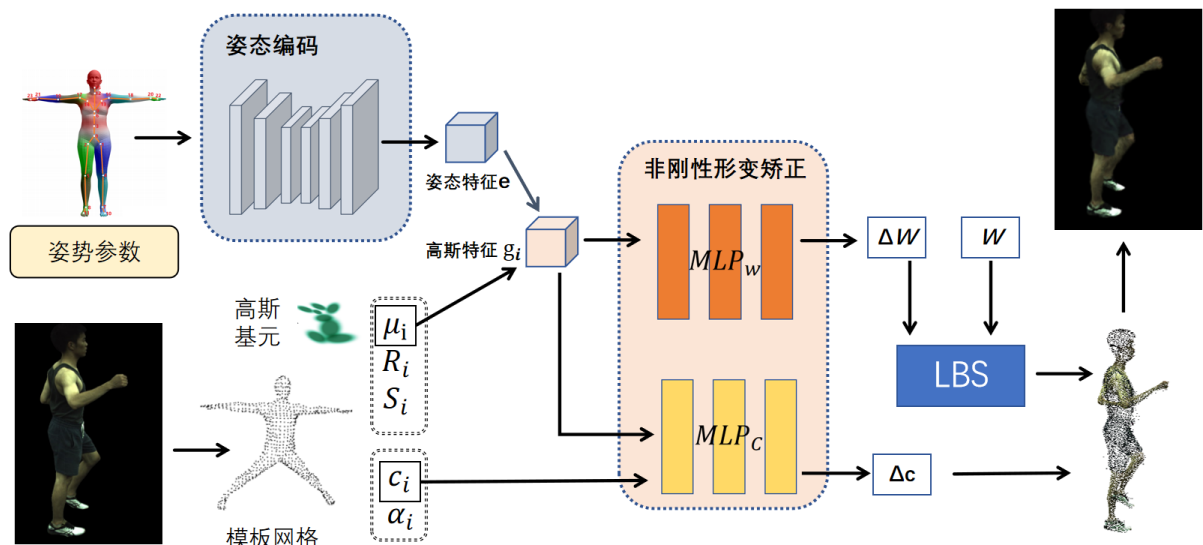


Figure 1. Diagram of the DCGaussianAvatar model framework  
图 1. DCGaussianAvatar 模型框架图

#### 3.2. SMPL 先验约束下的高斯参数初始化

DCGaussianAvatar以SMPL提供的人体结构先验作为高斯表示构建的基础。给定视频序列和相机参数,首先利用参数回归器获得每帧的SMPL参数,包括人体的形状参数 $\beta$ 和姿态参数 $\theta$ ,并选取初始帧生成模板网格 $M(\beta, \theta)$ 。为了获得与人体表面一致的初始几何分布,我们在模板网格表面进行采样,得到一组表面点 $V = \{v_i\}_{i=1}^n$ ,并将其作为高斯基元的初始中心。相较于在空间中随机初始化,采用SMPL表面约束能够显著缩小搜索空间,使后续优化更易保持人体拓扑结构与整体对齐稳定性。

每个高斯基元  $G_i$  都有一组可学习的属性，用于表示局部几何覆盖和外观贡献，包括几何中心  $\mu_i$ 、局部旋转  $R_i$ 、缩放尺度  $S_i$ 、颜色参数  $c_i$  以及不透明度  $\alpha_i$ 。其中，几何中心  $\mu_i$ 、局部旋转  $R_i$  和缩放尺度  $S_i$  是高斯基元的几何属性。 $\mu_i$  为网格表面采样点的三维坐标。 $R_i$  是根据网格法线建立的正交矩阵，为各向异性的高斯基元提供合理的初始朝向。 $S_i$  表示高斯基元对局部表面的覆盖范围，可根据三角面尺度进行自适应设置。颜色参数  $c_i$  和不透明度  $\alpha_i$  为高斯基元的外观属性，需要设置较小的初始值以避免早期渲染饱和，并提升优化稳定性。

为使相邻高斯基元在姿态变化时保持一致的运动结构，我们进一步利用 SMPL 的线性混合蒙皮为高斯基元分配初始蒙皮权重  $W_i$ 。具体而言，我们将采样点  $v_i$  对应的网格顶点的蒙皮权重作为该高斯基元的初始权重；当采样点位于三角面片内部时，我们对该三角面片三个顶点的权重进行重心插值得到初始权重。通过上述初始化，我们在高斯基元和骨骼关节之间建立稳定的关联，为后续高斯基元的形变提供一致的运动参考。

### 3.3. 姿态参数编码

为使后续非刚性形变矫正模块能够显式感知全局姿态信息，本文对 SMPL 姿态参数  $\theta$  进行序列化编码，并采用 Transformer [19] 架构提取骨骼关节间的长程依赖关系，得到紧凑的姿态嵌入向量。考虑到 SMPL 原始轴角在旋转空间中可能存在不连续与跳变问题，我们将每个关节的旋转统一转换为更稳定的 6D 旋转表示，以获得连续、可微且更利于优化的姿态表征。转换后的姿态参数可表示为：

$$\theta = \{r_1, r_2, \dots, r_{24}\}, \quad r_j \in \mathbb{R}^6 \quad (1)$$

其中， $r_j$  表示第  $j$  个关节的 6D 旋转向量。为使 Transformer 能够接收并处理关节序列特征，我们将  $r_j$  通过线性映射得到关节级嵌入，接着引入一个可学习的位置编码  $p_j$  以区分不同关节在序列中的位置信息，两者相加得到最终的输入序列  $Z = \{z_j\}_{j=1}^{24}$ 。该过程可描述为：

$$z_j = W_{\text{emb}} r_j + p_j \quad (2)$$

最后，我们将  $Z$  输入 Transformer 编码器，通过多头自注意力在关节维度建模全局姿态的关联性，接着采用均值池化聚合关节特征，得到全局姿势特征  $e$ 。该过程描述为：

$$e = \frac{1}{24} \sum_{j=1}^{24} [\text{TransEnc}(Z)]_j \quad (3)$$

全局姿势特征  $e$  将作为条件向量输入非刚性形变矫正模块，用于驱动后续对蒙皮权重与高斯外观属性的更新。

### 3.4. 非刚性形变矫正模块

尽管 SMPL 先验能为高斯基元的形变提供了稳定的结构驱动，但受限于 LBS 的线性表示，高斯基元难以捕获人体的非刚性形变和动态纹理细节。因此，本文设计了非刚性形变矫正模块，在模板驱动的粗略运动基础上学习残差修正，并将这种修正同时作用于蒙皮权重和高斯基元的外观属性。

我们首先通过位置嵌入函数  $Embed(\cdot)$  对高斯基元的初始几何中心  $\mu_i$  进行编码，得到特定频率的位置特征，接着与全局姿势特征  $e$  拼接得到高斯基元的高级特征  $g_i$ ，可表示为：

$$g_i = \text{Concat}(Embed(\mu_i), e) \quad (4)$$

该特征用于同时驱动蒙皮权重微调与外观校正：权重分支直接以  $g_i$  预测蒙皮权重残差，而外观分支则将  $g_i$  和高斯基元初始的颜色参数  $c_i$  共同输入预测颜色残差，在保持原始外观的基础上进行自适应修正。

在权重分支中，我们使用轻量的权重网络  $MLP_w(\cdot)$  预测高斯基元的权重残差分量：

$$\Delta W(i, j) = MLP_w(g_i)[j] \quad (5)$$

其中， $MLP_w(g_i)[j]$  为网络输出向量的第  $j$  个分量，表示第  $j$  个关节对高斯基元非刚性形变的影响权重。接着，我们对先验权重进行偏移，并采用 softmax 归一化得到最终蒙皮权重  $\hat{W}_i$ ：

$$\hat{W}_i(j) = \frac{\exp(\log(W_i(j) + \varepsilon) + \Delta W(i, j))}{\sum_{k=1}^{24} \exp(\log(W_i(k) + \varepsilon) + \Delta W(i, k))} \quad (6)$$

通过上述残差式更新，模型能够在保留 SMPL 权重先验约束的同时，自适应地调整各关节对高斯基元运动的贡献比例，从而提升关节区域及衣物随动区域的运动贴合与优化稳定性。

在外观分支中，我们使用轻量的外观网络  $MLP_c(\cdot)$  预测颜色残差，并得到更新后的外观参数  $\hat{c}_i$ 。该过程可表示为：

$$\hat{c}_i = c_i + MLP_c(\text{Concat}(g_i, c_i)) \quad (7)$$

残差更新能够在保持原始外观一致性的同时，仅对受姿态变化影响显著的区域进行有针对性的修正，从而降低跨帧漂色并提升细节稳定性。

## 4. 实验与结果

### 4.1. 数据集设置

本文在公开人体动态数据集 ZJU-MoCap 上对 DCGaussianAvatar 的有效性进行验证。ZJU-MoCap 提供了多视角同步采集的人体序列数据，包含不同主体在连续动作过程中的姿态变化，可用于训练与评估动态人体重建与新视角渲染任务。在 ZJU-MoCap 数据集中，我们选取了 377、386、387、392、393、394 这 6 个视频序列进行实验。对于每个视频序列，我们选取一个相机视角并以 5 帧为间隔进行采样，共选取 100 帧用于训练，其余相机视角采用相同的采样策略用于测试评估。

### 4.2. 实验结果与分析

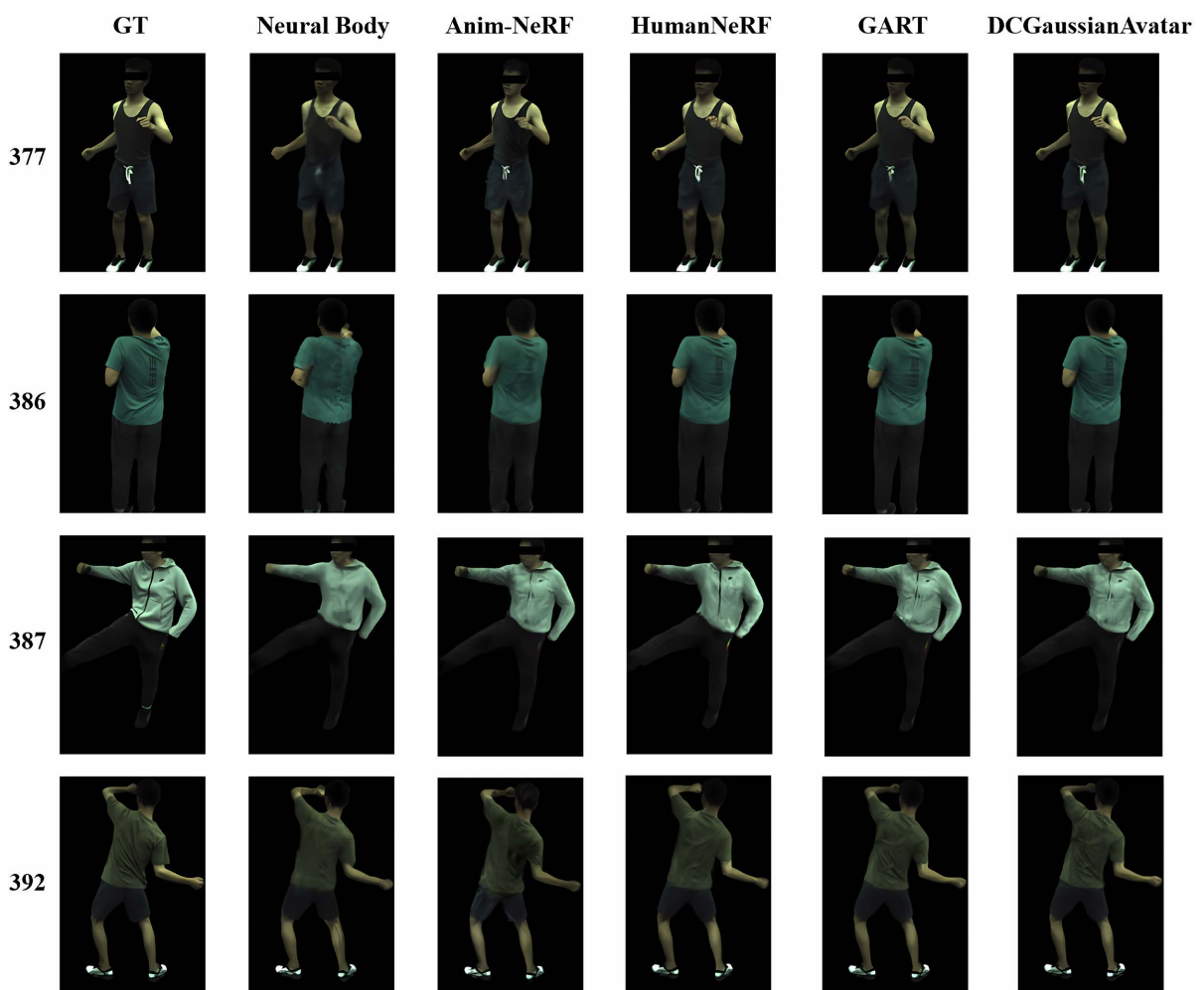
为全面评估 DCGaussianAvatar 在动态人体重建与新视角渲染任务中的性能，本文在 ZJU-MoCap 数据集上与 Neural Body [20]、HumanNeRF [13]、Anim-NeRF [14] 和 GART [17] 进行对比。考虑到不同方法的输出形式均可渲染得到目标视角图像，本文采用 PSNR、SSIM 和 LPIPS 对实验结果进行定量评估，前两个指标的数值越高代表重建效果越好，最后一个指标数值越低代表感知效果越好。

表 1 和图 2 给出了各方法在 ZJU-MoCap 上的对比结果和渲染结果。可以看出，隐式表示的方法训练开销较大，训练时间约为 10 小时，且在拳头、衣服边缘出现明显的伪影和不自然形变。相比之下，基于 3D 高斯溅射的 GART 在训练效率上具有显著优势，仅需 3 分钟就能完成训练，并在 LPIPS 上取得最优结果，表明其感知质量较好。然而，GART 的 SSIM 相对偏低，说明在结构一致性与细节对齐方面仍存在一定不足。

本文提出的 DCGaussianAvatar 在 PSNR 上取得最高值 30.93，同时 SSIM 达到 0.964、LPIPS 为 0.032，表现出较为均衡的重建质量与感知效果。值得注意的是，DCGaussianAvatar 的训练时间仅为 5 分钟，显著低于 Neural Body、HumanNeRF 与 Anim-NeRF 的小时级训练成本，并接近 GART 的分钟级效率。这表明本文方法在保持高效训练与渲染优势的同时，通过姿态条件的非刚性形变矫正与权重/外观残差更新，有效提升了动态姿态下的细节贴合与整体重建质量。总体而言，DCGaussianAvatar 在质量与效率之间取得了更优的综合权衡，更适合用于需要快速建模与交互式渲染的应用场景。

**Table 1.** Comparative experiments with baseline models  
**表 1.** 模型对比实验

模型	PSNR	SSIM	LPIPS	训练时间
Neural Body	28.93	0.964	0.042	10 小时
HumanNeRF	30.46	0.969	0.033	10 小时
Anim-NeRF	29.74	0.965	0.045	10 小时
GART	30.86	0.961	0.031	3 分钟
DCGaussianAvatar	30.93	0.964	0.032	5 分钟



**Figure 2.** Comparison of rendered results  
**图 2.** 模型渲染结果对比图

### 4.3. 消融实验

为验证 DCGaussianAvatar 各关键组件的有效性，本文在 ZJU-MoCap 数据集上进行消融实验。消融设置以完整模型为基准，并依次移除核心模块，考察其对重建质量与感知效果的影响。消融实验设置如下：

(1) w/o Pose Encoding: 移除姿态参数编码器, 令非刚性矫正模块不使用全局姿态嵌入, 仅基于高斯中心位置嵌入进行预测。

(2) w/o LBS Weight Refinement: 移除蒙皮权重残差更新分支, 固定使用 SMPL 插值得到的初始权重, 仅保留外观残差更新。

(3) w/o Appearance Residual: 移除外观残差更新分支, 仅进行权重残差更新。

如表 2 所示, 完整模型在三项指标上取得最优表现。首先, 移除姿态编码器后, 模型的 PSNR 下降至 30.48, SSIM 降至 0.961, LPIPS 上升至 0.036, 说明缺少全局姿态条件会削弱模型对“不同姿态下非刚性形变模式”的区分能力。其次, 移除蒙皮权重残差更新模块后, SSIM 下降最明显, 表明权重微调对关节附近的运动分配与结构对齐具有关键作用, 当仅依赖 SMPL 初始权重时, 局部区域更易产生轻微错位, 从而影响结构一致性。最后, 移除外观残差更新模块对感知指标影响最大, LPIPS 从 0.032 上升至 0.039, 尽管 SSIM 仍保持在 0.962, 但细节纹理与局部明暗变化难以得到有效补偿, 导致视觉真实感下降, 这说明外观残差分支主要贡献于高频纹理与感知质量的提升。

**Table 2.** Ablation study

**表 2.** 消融实验

模型	PSNR	SSIM	LPIPS
w/o Pose Encoding	30.48	0.961	0.036
w/o LBS Weight Refinement	30.62	0.959	0.034
w/o Appearance Residual	30.55	0.962	0.039
DCGaussianAvatar	30.93	0.964	0.032

## 5. 局限性

尽管 DCGaussianAvatar 在重建质量与训练效率方面取得了较好的综合表现, 但仍存在一定局限性: 一方面, 方法依赖 SMPL 的初始化精度与前景分割质量, 当初始姿态和形状估计误差较大或存在严重遮挡时, 可能导致高斯基元的初始对齐偏差, 从而影响后续残差矫正的上限; 另一方面, 本文采用逐场景优化的训练范式, 虽为分钟级但仍需针对每个序列单独优化, 尚难直接满足“即插即用”的跨场景实时建模需求, 尤其在宽松衣物和大幅度非刚性形变条件下仍有进一步提升空间。

## 6. 结论

本文提出形变修正的高斯人体模型 DCGaussianAvatar, 结合 3D 高斯溅射与 SMPL 结构先验实现可驱动的人体重建与新视角合成。方法通过 Transformer 姿态编码器提取全局姿态特征, 并在姿态条件下对蒙皮权重与外观参数进行残差式更新, 以补偿衣物与软组织等非刚性细节误差。实验结果表明, DCGaussianAvatar 在 ZJU-MoCap 数据集上相较多种代表方法取得更优的综合表现, 并在效率与重建质量之间实现了更好的平衡。

未来工作将从两方面展开: 一是引入更强的视角相关外观建模与遮挡鲁棒机制, 以进一步提升复杂光照与强遮挡场景下的渲染一致性; 二是探索更精细的非刚性几何校正与时序一致性约束, 以增强对大幅度动作与宽松衣物形变的泛化能力, 并推动方法在实时交互与数字人应用中的落地。

## 参考文献

- [1] Fieraru, M., Zanfir, M., Oneata, E., Popa, A., Olaru, V. and Sminchisescu, C. (2020) Three-Dimensional Reconstruction

- of Human Interactions. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 7212-7221. <https://doi.org/10.1109/cvpr42600.2020.00724>
- [2] Shi, M., Aberman, K., Aristidou, A., Komura, T., Lischinski, D., Cohen-Or, D., *et al.* (2020) MoTioNet: 3D Human Motion Reconstruction from Monocular Video with Skeleton Consistency. *ACM Transactions on Graphics*, **40**, 1-15. <https://doi.org/10.1145/3407659>
- [3] Zackariasson, P. and Wilson, T.L. (2012) *The Video Game Industry: Formation, Present State, and Future*. Routledge.
- [4] Kachach, R., Perez, P., Villegas, A. and Gonzalez-Sosa, E. (2020) Virtual Tour: An Immersive Low Cost Telepresence System. 2020 *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, Atlanta, 22-26 March 2020, 504-506. <https://doi.org/10.1109/vrw50115.2020.00104>
- [5] Zhao, F., Xie, Z., Kampffmeyer, M., Dong, H., Han, S., Zheng, T., *et al.* (2021) M3D-VTON: A Monocular-to-3d Virtual Try-On Network. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 13219-13229. <https://doi.org/10.1109/iccv48922.2021.01299>
- [6] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R. and Ng, R. (2021) NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *Communications of the ACM*, **65**, 99-106. <https://doi.org/10.1145/3503250>
- [7] Kerbl, B., Kopanas, G., Leimkuehler, T. and Drettakis, G. (2023) 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, **42**, 1-14. <https://doi.org/10.1145/3592433>
- [8] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G. and Black, M.J. (2015) SMPL: A Skinned Multi-Person Linear Model. *ACM Transactions on Graphics*, **34**, 1-16. <https://doi.org/10.1145/2816795.2818013>
- [9] Jiang, B., Hong, Y., Bao, H. and Zhang, J. (2022) SelfRecon: Self Reconstruction Your Digital Avatar from Monocular Video. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 19-23 June 2022, 5605-5615.
- [10] Xu, Z., Peng, S., Geng, C., Mou, L., Yan, Z., Sun, J., *et al.* (2024) Relightable and Animatable Neural Avatar from Sparse-View Video. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 990-1000. <https://doi.org/10.1109/cvpr52733.2024.00100>
- [11] Mihajlovic, M., Zhang, Y., Black, M.J. and Tang, S. (2021) LEAP: Learning Articulated Occupancy of People. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 10456-10466. <https://doi.org/10.1109/cvpr46437.2021.01032>
- [12] Li, Z., Zheng, Z., Zhang, H., Ji, C. and Liu, Y. (2022) AvatarCap: Animatable Avatar Conditioned Monocular Human Volumetric Capture. In: *Lecture Notes in Computer Science*, Springer, 322-341. [https://doi.org/10.1007/978-3-031-19769-7\\_19](https://doi.org/10.1007/978-3-031-19769-7_19)
- [13] Weng, C., Curless, B., Srinivasan, P.P., Barron, J.T. and Kemelmacher-Shlizerman, I. (2022) HumanNeRF: Free-Viewpoint Rendering of Moving People from Monocular Video. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 8-24 June 2022, 16189-16199. <https://doi.org/10.1109/cvpr52688.2022.01573>
- [14] Chen, J., Zhang, Y., Kang, D., Zhe, X., Bao, L., Jia, X. and Lu, H. (2021) Animatable Neural Radiance Fields from Monocular RGB Videos.
- [15] Pang, H., Zhu, H., Kortylewski, A., Theobalt, C. and Habermann, M. (2024) ASH: Animatable Gaussian Splats for Efficient and Photoreal Human Rendering. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 1165-1175. <https://doi.org/10.1109/cvpr52733.2024.00117>
- [16] Hu, L., Zhang, H., Zhang, Y., Zhou, B., Liu, B., Zhang, S., *et al.* (2024) GaussianAvatar: Towards Realistic Human Avatar Modeling from a Single Video via Animatable 3D Gaussians. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 634-644. <https://doi.org/10.1109/cvpr52733.2024.00067>
- [17] Lei, J., Wang, Y., Pavlakos, G., Liu, L. and Daniilidis, K. (2024) GART: Gaussian Articulated Template Models. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 19876-19887. <https://doi.org/10.1109/cvpr52733.2024.01879>
- [18] Qian, Z., Wang, S., Mihajlovic, M., Geiger, A. and Tang, S. (2024) 3DGS-Avatar: Animatable Avatars via Deformable 3D Gaussian Splatting. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 5020-5030. <https://doi.org/10.1109/cvpr52733.2024.00480>
- [19] Zheng, C., Zhu, S., Mendieta, M., Yang, T., Chen, C. and Ding, Z. (2021) 3D Human Pose Estimation with Spatial and Temporal Transformers. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 11636-11645. <https://doi.org/10.1109/iccv48922.2021.01145>
- [20] Peng, S., Zhang, Y., Xu, Y., Wang, Q., Shuai, Q., Bao, H., *et al.* (2021) Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 9050-9059. <https://doi.org/10.1109/cvpr46437.2021.00894>