三维人体重建技术调研报告

22451322 吴晟涛

摘要

近年来，3D 人类虚拟形象建模技术在计算机视觉和图形学领域蓬勃发展，作为游戏和动画等众多实际应用的核心，三维人体建模吸引了广泛关注，围绕三维人体角色创建的大量研究工作涌现，为三维人体建模构建了一个丰富的新知识体系。这一研究方向涉及从多种输入源（如单视图图像、多视图图像及视频）重建高度拟真的人类三维模型。特别是，基于隐式表示的方法，例如像素对齐隐式函数（PIFu）和神经辐射场（NeRF）等技术，通过深度学习显著提升了重建的精度和细节表现。其中，3D 隐式重建方法利用神经网络直接预测点的几何与纹理信息，已成功解决传统建模中面对复杂姿态和细节还原不足的问题。但这些方法也面临数据需求高、效率不足及通用性受限的挑战。为了应对这些问题，近年来提出了各种优化方案，包括引入 3D 几何先验、使用多视图信息融合，以及结合新型的表示方法（例如 3D Gaussian Splatting）。本文调研了当前 3D 人体重建技术的主要进展，分析了不同方法的技术细节与优缺点，并在此基础上探讨了未来的研究方向，为学术界和工业界从事 3D 建模的研究者提供参考。

关键词: 三维人体建模、隐式表示、NeRF、3D Gaussian Splatting

目录

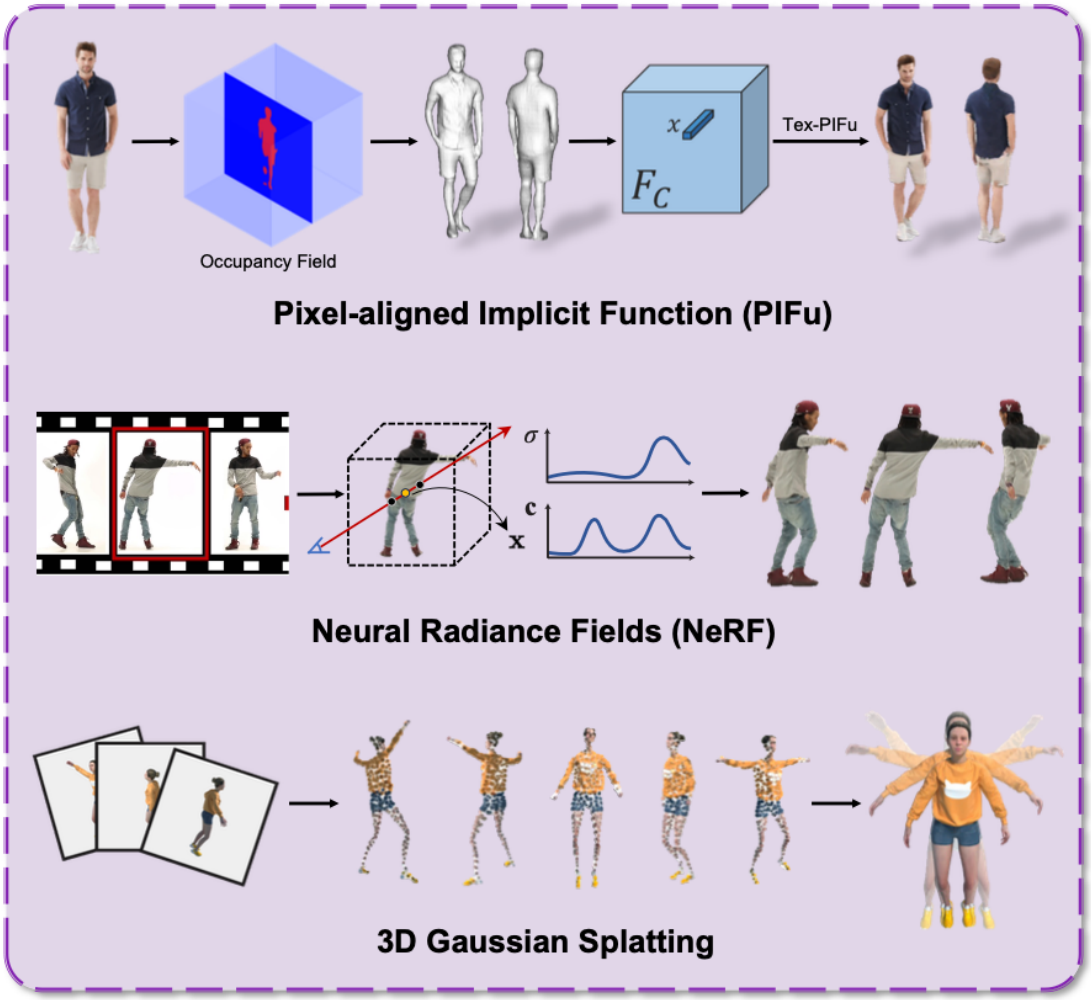
一、 引言

近年来，三维人体建模技术取得了显著进展，在计算机图形学、游戏、虚拟现实以及医学成像等领域拥有广泛的应用。早期方法依赖于昂贵的捕捉硬件和耗时的校准过程以生成高质量模型。随着技术发展，现在可以通过多种输入形式（如图像、视频或文本提示）更方便地重建和生成三维人体模型。

三维人体网格重建方法大致可分为基于模型的方法和无模型的方法。基于模型的方法通常通过调整显式参数化的人体模型（如 SMPL[1]）以拟合输入图像，但难以捕捉衣服、头发等精细细节。而无模型方法则通过预测体积空间中的占据值克服了这些局限。例如，PIFu[2]使用多层感知机（MLP）构建隐式函数，利用从输入图像中提取的像素对齐特征预测特定点的占据值。然而，PIFu 未充分利用人体结构信息，在处理复杂姿势、自遮挡和深度模糊方面表现有限。后续工作通过结合法线图、SMPL 模型及深度信息等先验知识，改进了这些问题。然而，这些方法在处理宽松衣物时受限于拓扑约束，随后通过显式方法（如 ECON[3]）加以解决。此外，基于多视角输入的场景能提供丰富的视角信息，从而改善重建效果。

然而，基于 PIFu 的方法在很大程度上受到稀缺的高质量三维训练数据集的限制。NeRF[4]通过输入有限图像集合合成新视角，实现了从每个三维点获取 RGB 颜色和密度值的目标。在此基础上，研究者提出了众多三维人体重建方法，通过将三维人体表示为神经辐射场，摆脱了对先验知识和预训练模型的依赖。除重建外，基于用户控制的新姿态序列探索任意视角动画生成也是一个有前景的研究方向。此外，一些方法结合表面场和辐射场，以实现高保真三维人体视角合成。

尽管神经辐射场方法表现优秀，但其高质量重建仍需昂贵的神经网络训练和渲染成本。3D Gaussian Splatting（3DGS）[5]提供了一种新的方法，以三维高斯表示和渲染复杂场景，在较短训练时间内保持了质量与速度的平衡。Kerbl 等人提出了这种显式且面向对象的方法，有别于隐式表示（如 NeRF 和 DMTET[6]）。许多后续工作基于其核心原则进一步提升三维人体重建性能，生成高度可动画化和真实感的人体模型。



图表 1 典型三维人体建模方法

二、 3D 隐式人体重建

随着隐式表示的兴起，利用深度神经网络对人体形态进行建模和重建的方法快速发展。这些方法旨在从单视图或多视图输入中高效生成细致的人体三维模型，同时克服传统基于网格或点云的显式表示面临的拓扑限制问题。在众多隐式方法中，像素对齐隐式函数（PIFu）[2]为领域奠定了技术基础。基于 PIFu 的改进方法通过引入更强的特征提取技术和三维先验知识，显著提高了复杂姿态和动态人体的重建质量。

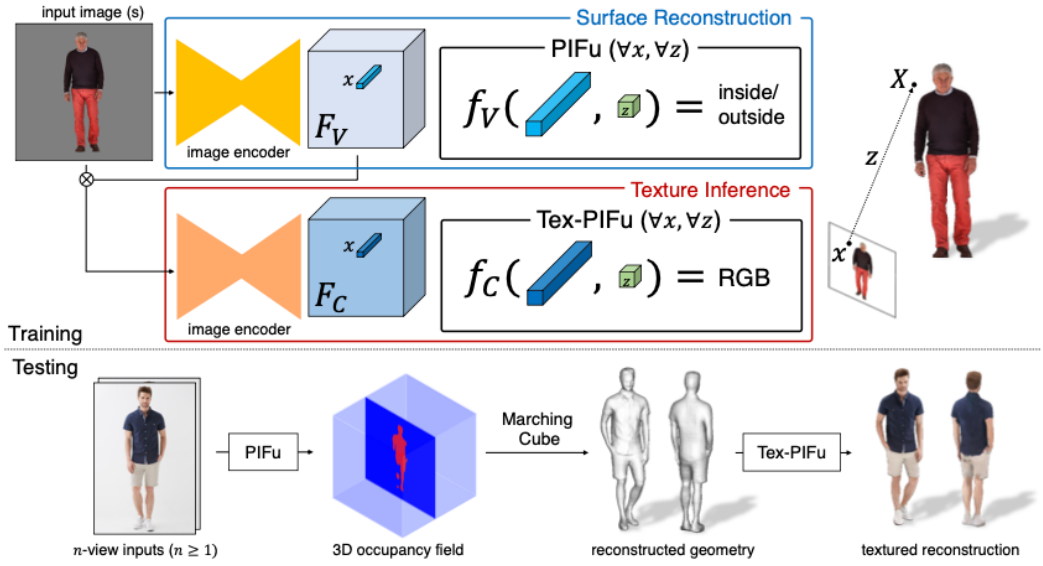
2.1 像素对齐隐式函数（PIFu）

随着单视图三维建模需求的增加，传统显式表示方法（如基于网格或点云的建模技术）在处理复杂拓扑和姿态变化时逐渐显现出局限性。显式方法通常需要通过离散网格

直接表示物体表面，这种方式对分辨率和几何复杂度敏感，很难应对动态变化和非刚体表面。同时，生成的网格模型通常受限于固定的拓扑结构（如封闭表面），无法直接处理松散衣物或复杂纹理形变等特性。

针对这些挑战，Saito 等人在 2019 年提出了像素对齐隐式函数（PIFu），这一方法通过引入隐式表示和像素对齐特征，突破了传统建模的拓扑限制。其核心思想是利用神经网络直接学习一个连续的隐式函数，将任意三维点的几何属性（如占据值）映射为 0 到 1 之间的值，以表示该点是否属于物体表面。这种表示方式摒弃了网格或点云的离散化表征，允许模型在任意分辨率下生成更细腻的表面几何。

PIFu 的关键创新是“像素对齐”特征提取策略。具体而言，它将输入图像的二维像素特征与三维空间中的查询点投影对齐，通过提取该点的局部像素特征，结合一个多层感知机（MLP）模型，预测三维点的占据值。这种策略不仅高效地利用了输入图像的信息，还显著简化了传统三维重建中视角变换和特征对齐的复杂过程。



图表 2 PIFu **方法架构图**

尽管 PIFu 在精细建模上展现了卓越能力，但其在处理复杂姿态、深度歧义和自遮挡时仍面临挑战。这些问题主要源自单视图输入中缺乏足够的三维信息，例如：对于松散衣物和非刚体表面，由于深度模糊或特征不足，重建的表面细节可能出现混叠。此外，在遮挡严重或动态复杂的情况下，输入图像提供的局部信息不足以支持模型精确建模。

为解决上述问题，研究者对 PIFu 进行了改进：首先是引入了高分辨率特征，PIFuHD[7]

是 PIFu 的关键后续工作，通过结合高分辨率图像特征提取模块，以及从图像生成的法向图和深度图，大幅提升了对几何细节的捕捉能力。PIFuHD 采用粗到精的处理流程，首先从低分辨率输入图像生成初步的隐式场，再通过高分辨率信息校准表面细节。这一方法在衣物细节和纹理复杂度较高的场景（如松散服装、褶皱等）中表现尤为出色。

为了解决单视图重建中的深度歧义问题，StereoPIFu[8]采用立体视觉增强模型，通过结合双视图的几何信息生成高质量的深度图。其独特的深度修正机制使模型在复杂环境下具有更高的准确性，尤其适合多人互动和复杂动态情景。

Geo-PIFu[9]进一步改进了 PIFu 的特征提取方式，将二维像素对齐扩展到三维空间的体素对齐特征。通过在模型中引入多尺度体素特征，Geo-PIFu 成功捕捉了复杂非刚体的形态，例如宽松服装、弯曲肢体以及复杂拓扑的精细部分。这一扩展使隐式模型能处理更广泛的应用场景。

PIFu 的提出不仅推动了隐式建模领域的快速发展，还为诸如影视制作、虚拟试衣、游戏动画等应用场景提供了技术支撑。它的核心思想（像素对齐与隐式表示结合）为后续的大量改进方法奠定了基础。在改进技术的支持下，PIFu 系列方法逐步解决了复杂形态、多视角结合以及动态环境下的建模难题，使隐式人体建模逐步接近真实世界应用的需求。

2.2 引入 3D 几何先验

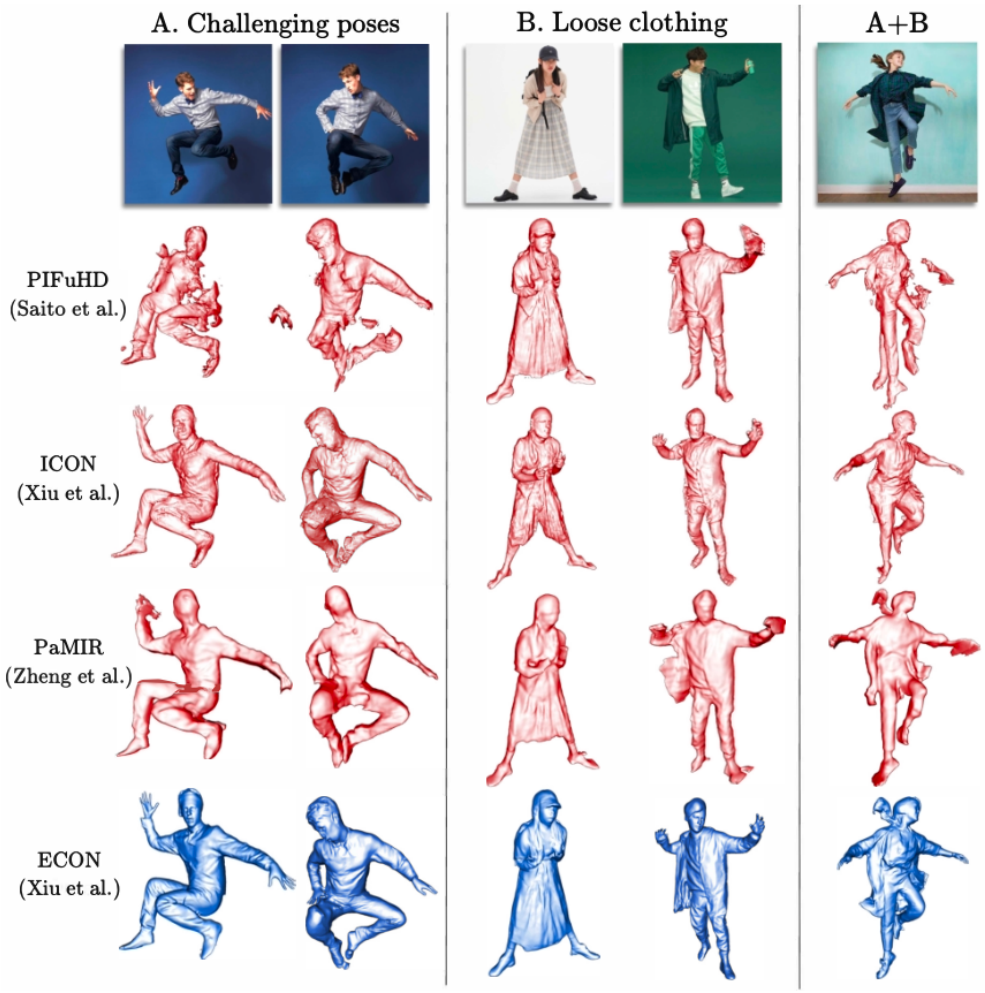
单视图 3D 人体建模因输入信息的稀缺性，始终面临深度模糊和遮挡歧义等问题。单一角度无法完整捕捉目标的空间几何特性，使得隐式表示方法在预测不易观察到的部位（如背部或被遮挡区域）时，容易出现形态扭曲或错误重建。此外，对于人体复杂姿态和动作变化，由单一视图推断其体态构造也是一个技术挑战。为了应对这些问题，研究者开始探索将显式几何先验与隐式表示结合的方法，通过将事先已知的标准化三维人体模型嵌入神经网络，提供关于人体拓扑、姿态和形状的基本信息。例如，基于 SMPL（Skinned Multi-Person Linear）模型的几何先验，利用标准化的关节骨骼结构和人体表面形态，为深度学习模型提供了更准确和一致的参考框架，从而在大幅减少训练样本依赖的同时，有效提高建模质量。

PaMIR[10]是此方向的一个代表性方法，它首先从输入图像中估计初始 SMPL 模型，

随后将模型转化为占据值体素编码，使得神经网络能够从图像特征和体素特征中推断目标点的占据值。这种结合二维与三维信息的方式，使 PaMIR 能够在复杂姿态下重建人体，同时大幅减少遮挡问题。

ARCH 通过定义语义空间（Semantic Space）和语义变形场（Semantic Deformation Field），将三维点从观察空间映射到标准化空间。该方法利用空间一致性减少姿态变化带来的几何重建误差，在不同动作和形变情境下表现出较高的鲁棒性。ARCH++[11]进一步扩展了这一思路，通过引入观测空间与标准化空间的双重一致性约束，进一步提升了对跨场景重建任务的适应性。

相比于 PaMIR 和 ARCH，ICON[12]采用了一种更加高效的局部特征提取机制，能够以更低的计算成本实现对衣物纹理和几何细节的高保真还原。ICON 还借助法向图信息，优化了深度预测结果，从而提高了松散衣物的细节还原能力。



图表 3 单视角场景下人体建模方法比较

2.3 多视图融合技术

为了弥补单视图信息的不足，多视图方法通过从不同视角捕获的图像中提取互补信息，进一步提升了人体建模的效果。PIFu 最初采用简单的平均池化策略整合多视图特征，这种方法虽然易于实现，但忽略了不同视角的质量差异。为解决这一问题，SeSDF[13]提出了一种遮挡感知特征融合方法，该方法通过动态调整来自不同视角的权重，在遮挡区域显著提升预测的准确性。此外，DiffuStereo[14]通过利用多层扩散网络生成深度图，并采用高效的多视图融合机制实现了高质量的 3D 人体建模。

DeepMultiCap[15]是另一种高效多视图技术，它引入了自注意力机制，能够根据视角重要性分配更精确的特征权重，从而大幅优化动态姿态下的表现。相比之下，StereoPIFu 针对立体视觉优化设计了特征提取网络，对双视图的深度预测表现更具优势。

2.4 显式隐式方法的结合

隐式表示方法在处理人体表面复杂性时具有显著优势，但神经网络推断的过程往往需要较大的计算成本。为应对这一问题，一些研究尝试结合显式表示的优势，例如 ECON[3]。在 ECON 中，通过显式生成前后法向图和深度图，并通过轮廓一致性优化几何外形，从而显著减少了计算成本，同时保持了对细节的优异捕获能力。与 ICON 相比，ECON 在处理大范围动态形变时表现更加稳定，尤其适合包含大量非刚体结构的任务场景。

三、 基于 NeRF 技术的三维人体重建

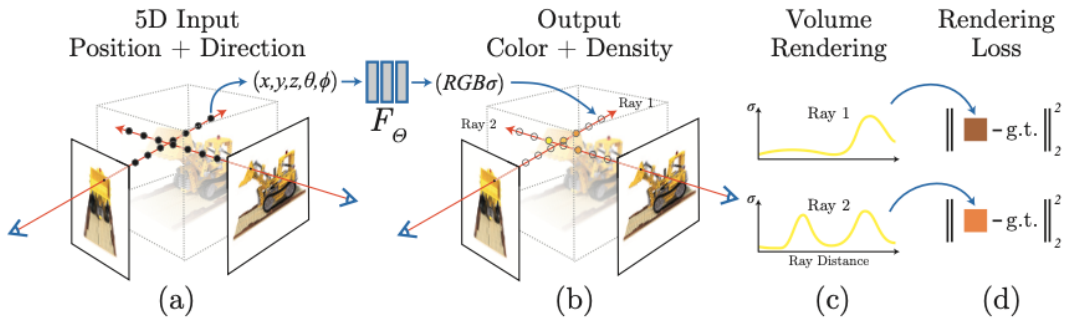
3.1 NeRF 的技术背景与核心原理

神经辐射场（Neural Radiance Fields, NeRF）[4]通过将三维场景表示为体素中的密度和颜色场，利用神经网络实现从多视角图像中合成高质量的新视点图像。NeRF 的核心是利用多层感知机（MLP）将三维点的空间坐标和方向编码映射到颜色和密度值：

$$F_\Theta(\gamma(\mathrm{x})) \rightarrow (\mathrm{c}, \sigma)$$

其中，$\gamma(x)$是高频位置编码，c 和 σ 分别表示颜色和密度值。具体来说，NeRF 会利用一个频率编码模块（Positional Encoding），将三维点坐标和观察方向提升到高维空

间作为网络输入。网络输出的密度和颜色值进一步通过可微分体积渲染模块将三维场景投影到二维图像。相比传统基于显式网格或体素的建模方法，NeRF 避免了大量的空间离散化开销，不仅更高效，也允许对复杂细节进行光滑的连续建模。NeRF 对训练数据的要求是多个固定视角的静态场景图像，每幅图像需提供准确的相机位置信息。这些特性使 NeRF 成为 3D 重建领域的里程碑，为更复杂的动态场景建模奠定了基础。



图表 4 NeRF **技术架构图**

3.2 静态摄像头下的 3D 人体重建

对于静态摄像头捕获的多视角图像，NeRF 被扩展用于动态人体的三维重建。例如，HumanNeRF[16]是一种针对动态人体建模的技术，通过结合 SMPL（Skinned Multi-Person Linear）模型，设计了一种基于变形场的映射策略，将观察空间中的点映射到标准化空间中，然后利用辐射场预测点的密度和颜色值。这样不仅解决了动态姿态下身体不同部位位置的不一致性，也显著提高了重建的细节质量。然而，由于输入的 SMPL 姿态估计可能存在误差，HumanNeRF 引入了姿态优化模块，在训练期间进一步细化这些估计结果以提升精度。

Neural Body 在 HumanNeRF 的基础上引入了“结构化潜码”（Structured Latent Code）的概念。这些潜码被固定到 SMPL 网格的顶点，并在训练过程中进行优化，使得网格点的几何特征能够被精确捕获。通过稀疏卷积网络将潜码生成一个三维体素特征体，这种方式有效地整合了多视图输入信息，实现更自然的三维重建。这使得 Neural Body 不仅能够表现细腻的人物纹理，还在处理复杂动作场景时表现出了强大的建模能力。

Neural Human Performer[17] 则直接在观察空间中建模，首次引入时间增强骨骼特征库（Time-augmented Skeletal Feature Bank）。通过从不同时刻的图像中索引特征，再结合

多视图融合策略，该方法能更精确地捕捉复杂的动态信息。

3.3 动态摄像头下的 3D 人体重建

当摄像头处于动态移动的场景下，人体建模的难点在于如何平衡相机运动与人体运动的双重动态特征，同时避免因复杂场景带来的遮挡和模糊影响。为解决这些问题，HOSNeRF[18]引入了一种针对人-物体交互的建模方案。通过扩展 SMPL 骨架为“物体骨架”，HOSNeRF 可将携带包袋等动态物体纳入模型，生成更具连贯性的动态人体表现。此外，HOSNeRF 使用状态条件表示（State-conditional Representations），通过学习一组可调整的状态嵌入向量，对物体在不同时间点的变化状态进行建模，并结合逆向线性混合蒙皮（LBS）映射回标准化空间。最终，该方法在使用 Mip-NeRF 360 建模背景的同时，还能够同步对人体与背景之间的交互进行精确建模。

在动态摄像头下，另一个方法是 NeuMan[19]。该方法通过独立训练两组 NeRF 模型——一个专注于背景场景的建模，另一个专注于人体本身——实现了动态背景与人物动作的清晰分离。同时，它通过多个时间帧的特征聚合来解决遮挡现象，并有效提升了建模精度。

3.4 动态人体动画与自由视角生成

NeRF 不仅可以用于单帧图像的三维重建，在动画和动态姿态建模中也具有很大的应用潜力。例如，Animatable NeRF[20] 提出了基于神经混合权重场（Neural Blend Weight Field）的动态身体变形模型，该方法不仅处理动态人物的骨骼变形，还捕获复杂的衣物和柔性表面变化。通过对人体动态场景中每个点的姿态、位置信息和运动特征进行多任务建模，该方法在生成动画时能够表现出更强的运动一致性。

此外，PersonNeRF [21]专注于静态图像集合中的人物自由视角生成任务。其核心设计是引入一个共享运动场（Shared Motion Field），通过对不同服饰和姿态的联合训练，在有限的输入样本条件下学习到一种全局一致的姿态表示。PersonNeRF 支持用户探索未见过的姿态与外观组合，适合个性化的虚拟角色创建。

3.5 网格与辐射场的结合

NeRF 技术的另一扩展方向是结合传统网格表示与隐式辐射场建模以弥补彼此短板。

例如，DoubleField[22]同时在隐式场中建模表面几何和辐射场，通过共享的特征嵌入向量实现两者的联合优化。这种方法不仅可以显著提高细节还原和纹理渲染的质量，同时通过共享特征有效减少了网络的学习开销。

另一个例子是 Function4D[23]，该方法引入滑动窗口动态融合（Dynamic Sliding Fusion）技术，在多视角时间序列上完成细致的人体几何重建。这种设计避免了长时间追踪带来的冗余问题，同时提升了几何一致性和纹理的真实感。

图表 5 基于隐式表达的人体重建技术结果比较

四、 基于 3DGS 技术的三维人体重建

4.1 3DGS 的技术背景与核心原理

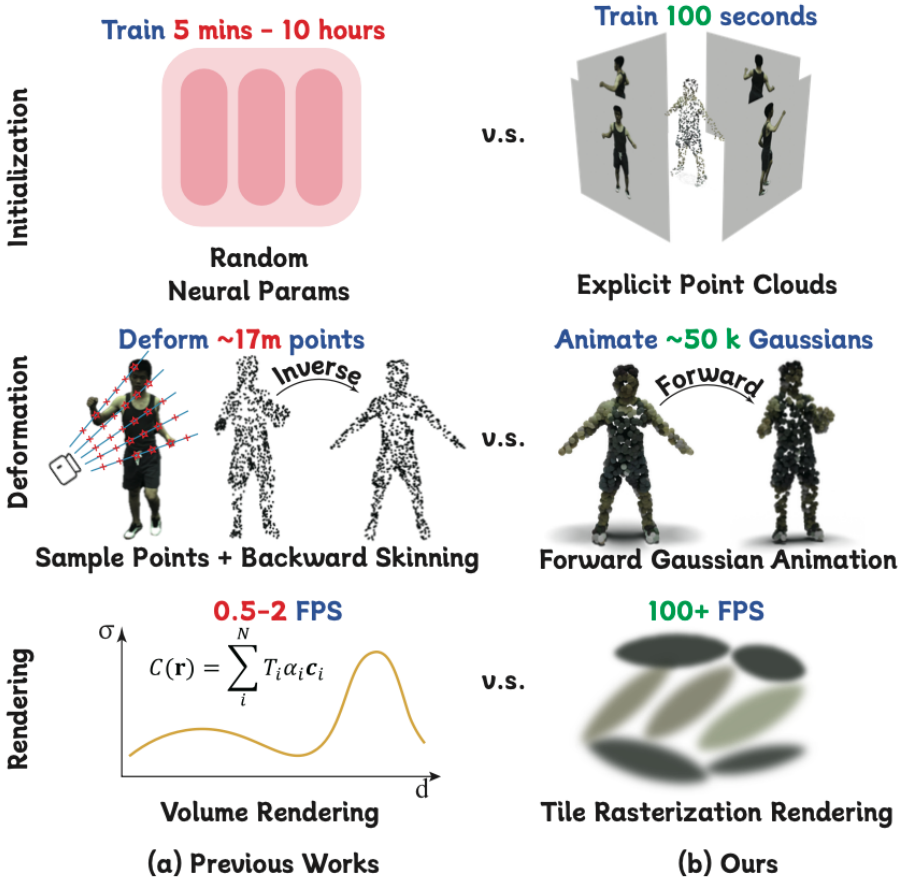
3D Gaussian Splatting（3DGS）[5]是一种新型的 3D 表示方法，致力于突破 NeRF 在训练效率和计算成本上的限制。与 NeRF 依赖多层感知机（MLP）建模不同，3DGS 通过优化场景中三维高斯分布的参数直接进行建模。这种方法以 3D 高斯分布的显式几何为核心，简化了神经网络的复杂度，显著提升了训练效率，同时保留了细腻的场景细节和高分辨率渲染能力。

在 3DGS 中，每个三维点以一个高斯分布描述，其关键参数包括位置、透明度、各向异性协方差矩阵及球谐函数（SH）系数。高斯分布的使用能有效捕获点云数据和复杂

场景特征，通过优化这些参数来描述密度场和颜色场。此外，与 NeRF 依赖隐式网络预测密度和颜色不同，3DGS 直接使用显式点云定义场景，具有更直观的几何意义和更高效的优化路径。

3DGS 通过“点分布渲染”（Point Splattering）实现场景的 2D 投影。这种方法基于光线投射，对图像平面上的像素进行插值，整合了高斯分布的密度与颜色信息，生成逼真的图像。同时，为保证模型训练的精确性和效率，3DGS 引入了随机梯度下降（SGD）优化算法，结合 L1 损失和 D-SSIM 损失，逐步优化每个高斯分布的参数。此外，采用“密度强化与筛选”的策略，通过定期增加高斯点密度和移除透明点，确保模型对场景细节的描述更加准确。

与 NeRF 等隐式方法相比，3DGS 有以下显著优势：首先，3DGS 显式优化场景点云的几何和颜色参数，无需依赖复杂的网络结构，大幅缩短了训练时间；其次，通过高斯点的各向异性建模，3DGS 在渲染动态物体或细节复杂的场景时表现更优；最后，相比于 NeRF 的高计算成本，3DGS 的实时性能显著增强，为动态交互应用铺平了道路。



图表 6 基于 NeRF 技术和基于 3DGS 技术重建人体方法比较

4.2 动态人体建模

在动态人体建模中，3DGS 能够通过骨骼驱动的高斯点云变形场捕捉人体及衣物在不同动作间的几何变化。这种方法首先通过逆线性混合蒙皮（Inverse Linear Blend Skinning）将三维点从观察空间映射到标准化空间，然后优化高斯点的旋转、位移和比例以模拟人体各部位的动态行为。例如，Animatable-3D-Gaussian[24]的方法结合高斯点动态变形与时间编码技术，捕获复杂的人体姿态变化，同时保持高效的渲染速度。

在更复杂的动态情境中，如衣物与人体的分离式建模，可驱动化身（Drivable 3D Gaussian Avatars）[25]采用笼形变形（Cage Deformation）的技术，使独立的服饰和人体骨架能够实现协调变形。通过设计专门的网络对高斯参数进行优化，模型能够准确还原松散衣物的细节运动和遮挡关系，适应多样化的用户输入视角视频驱动的建模

在单视角视频输入的场景中，3DGS 还可以高效生成可驱动的人体模型，例如 Human101[26]。这类方法通过解析单视角视频中包含的多视角特性，对人体多个姿态进行建模，并利用网格高斯化策略将生成的显式模型转换为高斯点云。在实时渲染中，模型利用时间编码预测高斯点残差修正参数，从而保障动态动画的自然流畅。此外，这些方法也大幅减少了模型训练和渲染的资源开销，为实际应用场景提供了更具可行性的解决方案。

4.3 面部与局部动态细节建模

3DGS 技术在面部与局部动态细节建模方面展现出显著优势，特别是在处理复杂表情变化、头部运动和局部几何细节时，能够生成高度细腻且可控的动态效果。传统的面部建模方法，如基于 3DMM 的手动标注技术，通常只能描述静态表情或标准化形态，缺乏对动态变化和细节捕捉的支持。3DGS 突破了这一限制，通过显式建模策略将复杂的面部运动和动态表情变化嵌入到 3D 高斯点分布中，同时通过球谐函数（SH）参数实现更逼真的视角相关纹理渲染。

一个关键技术是将高斯分布与表情参数空间关联，如 Gaussian Head Avatar[27]利用 3DMM 生成初始头部网格模型，然后通过高斯化转换每个点的位置、旋转和透明度，捕捉其在面部表情和姿态中的动态变化。模型不仅对头部旋转、倾斜等大尺度运动有较强适应性，还能针对微表情、肌肉运动等细节建模。这种方式尤其适用于头像表情生成、

虚拟主播等高需求场景。此外，引入三平面特征（Triplane Features）作为辅助，在纹理表现上进一步提升了脸部细节的真实感。这些方法极大地提高了表情动态模拟的精度，为实时生成具有高保真度的虚拟角色打下了坚实基础。

然而，3DGS 在面部建模中的局限性也不容忽视。虽然其在单一的头部对象上表现优异，但多头部、多视角或光照复杂的动态场景下仍存在适应能力不足的问题。此外，动态场景中的遮挡和透视一致性问题未完全解决，使得表情交叠或者过于快速变化的场景难以被充分表达。因此，对局部区域的特定优化和多视角增强技术正逐渐成为研究的重点方向。

五、 反思与未来挑战

随着 3D 人体建模技术的快速发展，神经表示方法（如 NeRF 和 3DGS）在高精度建模和复杂动态捕捉方面已取得显著进展。然而，回顾现有方法的技术特点，我们可以发现其仍面临一些核心挑战。首先，尽管 NeRF 和 3DGS 显著降低了对高质量 3D 数据集的需求，但真实场景中的训练数据不足仍然限制了模型性能的上限。当前方法依赖稀疏、单一分辨率的数据来源，难以充分捕获复杂材质和多尺度动态的特性。此外，3D 人体重建技术中的通用性和迁移能力仍需提升。大多数方法针对单一场景或人体形态设计，在多材质复杂背景、极端动态动作或遮挡严重的环境中仍表现不足。

在性能优化方面，计算效率和实时性是主要瓶颈。尽管 3DGS 通过显式建模提升了效率，但动态场景中高分辨率的点云存储和渲染仍需占用大量硬件资源，而 NeRF 则在体积渲染计算中表现出高昂的资源消耗。进一步优化实时渲染性能以支持工业级应用，如虚拟现实、游戏动画和实时交互，已成为研究的关键目标。

未来的发展方向主要集中在几个关键领域。首先是构建大规模的 3D 基础模型（Foundation Models）。通过融合大范围、多分辨率的 3D 语义和几何数据，可使建模算法具备更强的通用性和适应性。其次，结合显式表示与隐式表示的优点，有望在细节还原和计算效率之间取得平衡。例如，可将显式的几何骨架与隐式的辐射场或高斯场联合使用，实现灵活可控的人体动态动画生成。再者，多视角一致性优化将帮助解决遮挡和场景歧义的问题，从而支持多场景、多设备间的无缝数据协作。最后，智能驱动的建模方法，如基于文本、音频或手势输入的交互式建模，将推动虚拟数字人的便捷生成，满

足个性化和多样化应用需求。

总体而言，虽然当前的 3D 人体建模技术已在学术领域和特定应用场景中取得不俗成绩，但其与大规模实际应用间仍存在一段距离。结合优化算法、增强硬件支持以及高效数据采集与融合，将是实现这一领域全面突破的核心方向。

参考文献

[1] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: a skinned multi-person linear model. ACM Trans. Graph. 34, 6, Article 248 (November 2015), 16 pages. https://doi.org/10.1145/2816795.2818013

[2] Saito, Shunsuke et al. “PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization.” 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (2019): 2304-2314.

[3] XIU Y., YANG J., CAO X., TZIONAS D., BLACK M. J.: ECON: Explicit Clothed humans Obtained from Normals. In IEEE-Learning Transferable Visual Models From Natural Language Supervision Conference on Computer Vision and Pattern Recognition (2023).

[4] MILDENHALL B., SRINIVASAN P. P., TANCIK M., BARRONJ. T., RAMAMOORTHI R., NG R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM (2021).

[5] Kerbl, Bernhard et al. “3D Gaussian Splatting for Real-Time Radiance Field Rendering.” ACM Transactions on Graphics (TOG) 42 (2023): 1 - 14.

[6] Shen, Tianchang et al. “Deep Marching Tetrahedra: a Hybrid Representation for High-Resolution 3D Shape Synthesis.” Neural Information Processing Systems (2021).

[7] Saito, Shunsuke et al. “PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization.” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020): 81-90.

[8] Hong, Yang et al. “StereoPIFu: Depth Aware Clothed Human Digitization via Stereo Vision.” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 535-545.

[9] He, Tong et al. “Geo-PIFu: Geometry and Pixel Aligned Implicit Functions for Single-view Human Reconstruction.” ArXiv abs/2006.08072 (2020): n. pag.

[10] Zheng, Zerong et al. “PaMIR: Parametric Model-Conditioned Implicit Representation for Image-Based Human Reconstruction.” IEEE Transactions on Pattern Analysis and Machine

Intelligence 44 (2020): 3170-3184.

[11] He, Tong et al. “ARCH++: Animation-Ready Clothed Human Reconstruction Revisited.” 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021): 11026-11036.

[12] Xiu, Yuliang et al. “ICON: Implicit Clothed humans Obtained from Normals.” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 13286-13296.

[13] Cao, Yukang et al. “SeSDF: Self-Evolved Signed Distance Field for Implicit 3D Clothed Human Reconstruction.” 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023): 4647-4657.

[14] Shao, Ruizhi et al. “DiffuStereo: High Quality Human Reconstruction via Diffusion-based Stereo Using Sparse Cameras.” ArXiv abs/2207.08000 (2022): n. pag.

[15] Zheng, Yang et al. “DeepMultiCap: Performance Capture of Multiple Characters Using Sparse Multiview Cameras.” 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021): 6219-6229.

[16] Weng, Chung-Yi et al. “HumanNeRF: Free-viewpoint Rendering of Moving People from Monocular Video.” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022): 16189-16199.

[17] Kwon, Youngjoon et al. “Neural Human Performer: Learning Generalizable Radiance Fields for Human Performance Rendering.” Neural Information Processing Systems (2021).

[18] LIU J.-W., CAO Y.-P., YANG T., XU E. Z., KEPPO J., SHANY., QIE X., SHOU M. Z.: Hosnerf: Dynamic human-object-scene neural radiance fields from a single video. arXiv preprint arXiv:2304.12281 (2023).

[19] Jiang, Wei et al. “NeuMan: Neural Human Radiance Field from a Single Video.” European Conference on Computer Vision (2022).

[20] [PDW∗21] PENG S., DONG J., WANG Q., ZHANG S., SHUAI Q., ZHOUX., BAO H.: Animatable neural radiance fields for modeling dynamic human bodies. In Proceedings of the IEEE/CVF International Conference on Computer Vision (2021), pp. 14314–14323.

[21] Weng, Chung-Yi et al. “PersonNeRF : Personalized Reconstruction from Photo Collections.”

2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023): 524-533.

[22] Shao, Ruizhi et al. “DoubleField: Bridging the Neural Surface and Radiance Fields for High-fidelity Human Reconstruction and Rendering.” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 15851-15861.

[23] Yu, Tao et al. “Function4D: Real-time Human Volumetric Capture from Very Sparse Consumer RGBD Sensors.” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 5742-5752.

[24] Liu, Yang et al. “Animatable 3D Gaussian: Fast and High-Quality Reconstruction of Multiple Human Avatars.” ArXiv abs/2311.16482 (2023): n. pag.

[25] Zielonka, Wojciech et al. “Drivable 3D Gaussian Avatars.” ArXiv abs/2311.08581 (2023): n. pag.

[26] Li, Mingwei et al. “Human101: Training 100+FPS Human Gaussians in 100s from 1 View.” ArXiv abs/2312.15258 (2023): n. pag.

[27] Xu, Yuelang et al. “HHAvatar: Gaussian Head Avatar with Dynamic Hairs.” (2023).