arXiv:2304.02013v2 [cs.CV] 1 Sep 2023

NPC：来自视频的神经点角色

Shih-Yang Su[1] Timur Bagautdinov[2] Helge Rhodin[1]
不列颠哥伦比亚大学 [2]Reality Labs Research



图 1：**神经点角色（NPC）**是一种可动画的基于点的身体模型，可提升保真度和通用性。NPC 可从单个或多个视频中学习，能够很好地泛化到新姿态，且无需预先构建表面。**从后向前：**驱动运动、估计的点云和神经角色模型。为保护匿名性，所有人脸均已模糊处理。

摘要

如今，可以直接从视频中学习高保真人类三维模型，通常是将基于模板的表面模型与神经表示相结合。然而，获取模板表面需要昂贵的多视角采集系统、激光扫描，或严格受控的条件。以往方法虽不使用模板，却依赖从观测空间映射到规范空间的高成本或病态映射。我们提出一种混合式基于点的表示，用于重建可动画角色；该方法无需显式表面模型，同时能够泛化到新姿态。对于给定的视频，我们的方法会自动生成一组显式三维点，用以表示近似的规范几何，并学习一种关节形变模型，以生成依赖姿态的点变换。这些点既作为高频神经特征的支架，也作为在观测空间与规范空间之间高效映射的锚点。我们在成熟的基准测试中证明，与在规范空间或观测空间中运行的先前方法相比，我们的表示克服了它们的局限。此外，我们的自动点提取方法使模型能够学习人类和动物角色；尽管通用性更强，其性能仍可媲美采用绑定表面模板的方法。项目网站：*https://lemonatsu.github.io/npc/*。

1. 引言

如今，已能够从单目视频中重建出照片级真实感的人物，但要实现高保真重建，仍需受控条件和专用采集硬件，因此难以大规模应用。静态场景可以通过单台移动摄像机拍摄的多视角画面进行重建，而捕捉动态人体运动则需要受控的摄影棚条件 [2, 22,23, 39], ，通常还需要大量同步摄像机。解决单目情形的一种方法，是利用这样一种直觉：摄像机相对于身体某个部位的运动，大致等同于身体该部位相对于摄像机的运动 [43, 44]。然而，在不受约束的条件下，这仍然很困难，因为这需要建立跨帧对应关系，并处理动态形变。

一项重要的研究方向沿用了传统动画流程：将表面网格绑定到底层骨架，并为网格配备神经纹理 [2, 22] 或可学习的顶点特征 [16, 34, 35]。这种方法效率很高，因为基于前向运动学的前向映射能够以闭合形式稳健地估计底层几何，同时也能实现高质量重建，因为神经纹理能够表示高频细节 [22, 46]。然而，这种方法确实需要非常准确的三维姿势和身体形状，而这些通常

通过昂贵的激光扫描获得 [50] ，或通过离线拟合多视角捕捉数据中的个性化参数化人体模型获得 [2, 55]。此外，现有的大多数参数化模型都依赖线性混合蒙皮 [18], ，而这种技术容易产生伪影。在这项工作中，我们旨在从单段单目视频构建一个可动画化的全身模型，无需依赖预定义模板或复杂的捕捉系统。

另一类研究将问题转化为在规范空间中构建人体模型，例如学习处于 T 姿势的目标人物的神经辐射场 [14, 19, 50]。实际上，这类方法需要找到从观测空间到规范空间的反向映射，可通过高保真扫描学习 [49], ，例如在刚体运动的基础上学习形变 [31, 52],，也可通过求根 [19, 50]实现。这些方法通常泛化能力有限，需要高质量的训练数据，并且在测试时计算开销较大。

我们的目标是兼具无表面规范模型的通用性和正向映射的速度，同时不依赖会限制适用范围的预定义模板。为此，我们采用一种隐式、无表面的人体表示，它既不依赖精确的三维姿势，也不依赖骨架 [43, 44]。具体而言，我们使用 A-NeRF [44]，结合现成的姿势估计器 [15] ，获得合理的三维姿势估计，并使用 DANBO [43] 在规范空间中学习初始表面估计。

给定这样的初始估计，我们的方法——神经点角色（NPC）——能够重建高质量的神经角色，使其可以采用新的姿态进行动画制作，并渲染出新视角（图 1）。难点在于，初始形状非常粗略且带有噪声，不足以用于建模高质量几何。因此，NPC 的核心是一种新颖的点表示，旨在改进含噪的形状估计，随后学习表示精细的纹理细节和姿态相关形变。我们的两项主要贡献如下：首先，我们提出通过最近邻查找对前向映射求逆，从而找到规范空间与观测空间之间的对应关系，其中包括由图神经网络（GNN）建模的非线性姿态相关形变场。表面点充当锚点，使查询点能够通过高效的摊销查找，从规范空间中的神经场推断其反向映射。与求根方法相比，我们的方法效率更高；与假设分片刚性的模型相比，我们的方法精度更高。其次，我们提出将非线性形变后的点用作观测空间中的支架，以表示高频姿势相关细节。我们还将规范空间中的点特征和神经场与骨骼相对几何编码相结合，以促进更好的泛化。我们对该方法进行了全面评估，并在既有基准测试中取得了单目和多视角人体重建任务上的最先进结果。此外，我们还展示了该方法的通用性，能够重建人类和动物角色。

2. 相关工作

神经渲染 近年来已成为一类强大的方法，可用于构建可控的三维场景神经表示 [26, 29, 45, 53]。

基于点的表示。 NPBG [1]使用预先计算的静态点云来表示场景的大致几何形状，并学习一组附着在各点上的神经描述符；随后，这些描述符会以多个分辨率进行光栅化，并由可学习的渲染网络进一步处理。ADOP [40]采用类似策略，但还会优化相机参数和点云位置，从而实现更高保真度的重建。Pulsar [17] 将场景表示为一组球体，这些球体的位置和半径参数均可优化，随后通过定制的可微渲染器进行渲染，并由神经着色卷积网络进行后处理。SMPLpix [37]和基于点的人体服装 [59] 同样利用二维卷积神经网络，对光栅化后的二维点特征图进行修复以完成渲染。Power of Points [24] 学习一个自动解码器，以预测点位移来建模不同的服装及其动态变化。同时开展的工作 PointAvatar [61] 使用可变形点云表示人脸化身，并通过逐步扩展点云来捕捉精细细节，再使用可微点光栅化器进行渲染。我们的方法同样基于点的表示构建，但配备了关节模型，并依赖体渲染。

神经场表示。 神经辐射场（NeRF） [29, 42]以隐式方式表示场景，即使用坐标条件化的多层感知机，并实现静态场景的高保真新视角合成。NeRF 类方法也通过添加形变场 [9, 21, 32, 38, 54]应用于动态场景。与我们的方法尤其相关的是 Point-NeRF [56]。Point-NeRF 使用通过多视图立体视觉获得的点云作为几何代理，并在其上使用局部条件化 NeRF 对外观进行建模，从而加快学习速度，并实现静态场景的高保真新视角合成。类似地，NPC 使用基于点的神经场，但通过姿态驱动的关节骨架模型变换点云，以实现动态人体建模，并动态解码每个点的姿态相关载荷，以处理姿态相关形变。

用于人体建模的神经场。 近期研究将 NeRF 应用于从视频序列重建动态人体外观 [16, 27,35, 38, 52] ，并学习可动画化人体模型 [19, 22,30, 34, 43, 44, 50, 62]。对于静态场景，NeRF 会将查询点坐标直接映射到不透明度和颜色值。对可动画化NeRF 模型而言，关键挑战是将逐帧三维查询映射到规范人体模型空间。

应对这一挑战的一种常见方法是学习**反向映射**，将查询点从人物处于摆好姿态的观测空间映射到人物处于静止姿势（例如 T 姿势）的规范空间 [7, 14, 22, 28, 34, 41, 47]。随后在规范空间中定义神经场进行渲染。虽然学习反向映射能够高质量地重建动态人体几何和外观，但这些方法通常需要使用三维扫描进行训练 [7, 28, 41, 47], 需要纹理图 [22], 或依赖表面先验 [14, 34]。此外，反向映射难以学习，因为它需要求解多对一映射 [5, 19]因此，这些方法在渲染未见姿态时容易产生伪影。

此外，一些现有方法依赖**前向映射**，将特征从规范空间映射到观测空间。早期方法将神经特征锚定到 SMPL 顶点 [16, 35]，并通过线性混合蒙皮（LBS）[18] 将人体网格变形到观测空间。随后，利用三维卷积神经网络在姿态化网格周围扩散逐顶点特征，并在生成的体素特征网格上定义神经场。这些方法仍受到底层 LBS 模型质量的限制，且无法捕捉非线性形变，因此往往会产生模糊的重建结果。

另一种近期流行起来的反向映射方法，是对学习得到的前向映射应用可微求根算法 [19, 50]。TAVA [19] 是一种无模板方法，通过迭代求根来求解规范空间中的点。ARAH [50]提出了一种联合根查找方法，在使用预训练超网络 [11, 49] 初始化的 NeRF 人体模型上寻找射线交点，以进行渲染。尽管 TAVA 和 ARAH 都是无模板方法，并且在未见姿态上展现出最先进的合成质量，但两者在训练和测试时的计算开销都很大；此外，ARAH 还需要利用三维扫描构建表面先验。与这些方法相比，NPC 使用稀疏点云，在规范空间与观测空间之间建立高效的前向映射，从而以更低的计算成本实现更准确的查询特征定位，且无需依赖预训练表面先验。

与我们的方法在概念上相关的有 SLRF [62], KeypointNeRF [27], 和 AutoAvatar [3]。SLRF 采用了相似的理念，即利用表面点锚定独立的辐射场，以表示局部外观。相比之下，NPC 使用密集得多的特征点云显式表示局部结构，并在不同人体姿势之间建立密集对应关系，而无需依赖预定义的参数化网格。KeypointNeRF 使用稀疏关键点，对从多视角图像中提取的二维像素对齐特征进行三角测量，以实现三维人脸重建。与 KeypointNeRF 不同，NPC 将点特征直接存储在三维空间中，能够捕捉姿态相关形变，并可由外部驱动。AutoAvatar 使用基于 K-NN 的编码来表示动态神经人体形变场，但不对纹理建模。

3. 方法

NPC 从单段视频 $\{\mathbf{I}_1, \cdots, \mathbf{I}_N\}$ 中重建出一个可进行动画制作的神经角色，该视频包含 N 帧。图 2 展示了整体概况。该表示包括规范空间中的神经场 [29]（角色处于静止姿势），以及一组表面点 $\{\mathbf{p}_1, \cdots, \mathbf{p}_P\}$，这些点会动态变形，并在规范空间与观测空间之间进行映射。下文将说明如何获取这些表示，以及如何将它们结合起来构建 NPC 的几何与外观。

3.1. 骨架与点表示。

姿势与形状初始化。 输入视频中的每一帧 $\mathbf{I}_t$ 均使用现成的姿态估计器进行处理[1]，随后通过 A-NeRF 进行细化 [44]；如果条件允许，也可采用多视图三角测量。所得姿态 $\theta_t = \{\mathbf{R}_{t,0}, \cdots, \mathbf{R}_{t,J-1}\} \in \mathbb{R}^{J\times 3}$表示具有 J 个关节的预定义骨架的关节角旋转，随后被用作训练现有神经人体模型 DANBO [43]的输入。该人体模型被设为 T 姿势，并使用行进立方体提取底层隐式模型的等值面。请注意，我们训练 DANBO 时大幅减少了光线采样数量，并将迭代次数设为较少的 10k 次，相当于完整训练集的 ∼2.5%；在单个 NVIDIA RTX 3080 GPU 上，这一过程耗时不到 30 分钟。

分部件点云采样。 我们通过在等值面上进行最远点采样，获取稀疏表面点的位置 $\mathbf{p}$，直到得到 200 个点 $\{\mathbf{p}_{i,j}\}_{i=1}^{200}$ 对于骨架的每个部件 j。为简洁起见，当部件归属无关紧要时，我们省略部件下标。如图 2所示，生成的点云只是对底层几何的非常粗略的近似。接下来，我们将介绍主要贡献：在这一粗略估计的基础上学习高质量细节；与现有基于模型的方法相比，无需对干净表面作严格假设；与 DANBO 及其他无表面模型相比，能够呈现更清晰的细节并实现更好的纹理一致性。

从规范空间到观测空间（前向映射）。 提取的点 $\mathbf{p}_i \in \mathbb{R}^3$ 位于规范空间中，并在整个视频中保持固定。帧 t 中的形变通过一个可学习的映射建模，该映射将$\mathbf{p}_{i,j}$ 映射到姿态化的观测空间，

$$\mathbf{p}^{o}_{i,j,t} = \mathrm{LBS}(\theta_t, \mathbf{w}, \mathbf{p}_{i,j}) + R_{t,j}(\theta_t)\Delta\mathbf{p}_{i,j}(\theta_t), \quad (1)$$

包含用于建模刚性形变的线性混合蒙皮（LBS）[18] 项和非线性项 $\Delta\mathbf{p}$（θ_t），它

[1]我们沿用近期无模板人体 NeRF [19, 43] 的方法，使用基于 SMPL 的姿势估计器获取关节角旋转；其他姿势估计器也同样适用。

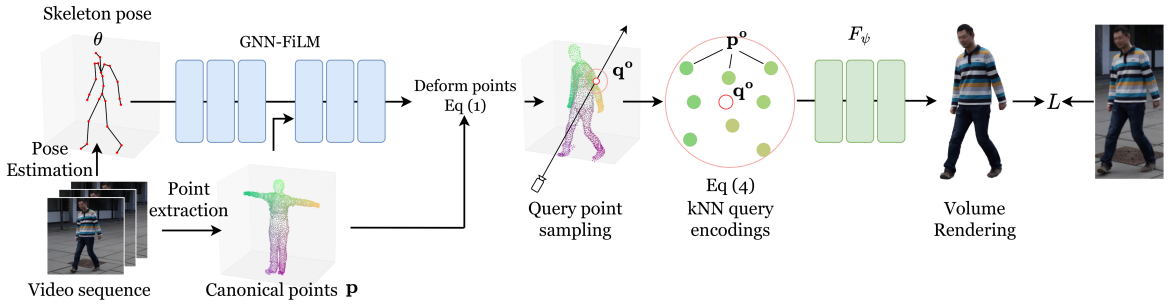


图2： **概述**。 NPC 使用 NeRF 对角色进行体渲染 F_ψ ，并以从动态变形点云聚合的特征进行局部条件化。给定原始视频，我们首先使用隐式人体模型（第 **3.1** 节）估计规范点云 p。随后，图神经网络根据骨架姿态 $\mathbf{p}$ 对规范点进行形变 θ，并生成一组姿态相关的逐点特征（第 3.2 节、第 3.3 节）。观测空间中的每个查询点 $\mathbf{q^o}$ 都会从姿态化点云中的 $\mathbf{k}$ 近邻聚合特征。聚合后的特征会传入 F_ψ 进行体渲染。我们的模型直接使用输入视频进行监督（第 3.4 节）。

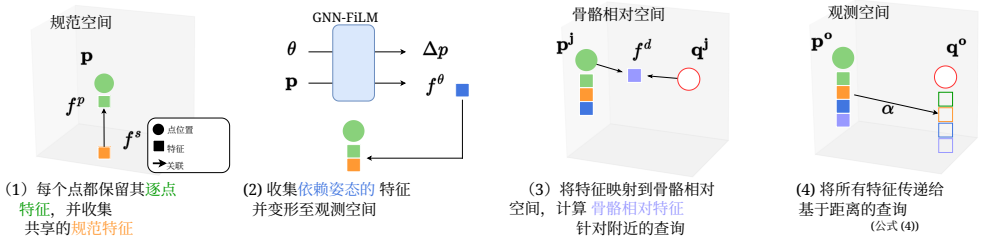


图3： **点特征编码**。 我们的核心思路是使用点云 $\mathbf{p}$ 作为锚点，将特征从规范空间传递到观测空间，从而在两者之间建立高效映射。(1) 每个 $\mathbf{p}$ 都携带一个可学习特征 f^p，其位置用于从规范场查询特征 f^s 。(2) 图神经网络添加姿态相关特征 f^θ 并使点云发生形变$\Delta\mathbf{p}$。(3) 在骨骼相对空间中加入视线方向和距离。(4) 使用 $q^\mathbf{o}$ 在 $\{\mathbf{p}_i^\mathbf{o}\}_i^N$ 中的 $\mathbf{k}$ 近邻，建立从查询点到姿态化点和规范点的前向和反向映射。

是相对于底层骨架的方向 $R_{t,\ j}(\theta_t)$ 学习得到的。

我们方法的关键在于学习依赖于姿态的变形 $\Delta\mathbf{p}_{i,j}(\theta_t)$ 以及针对角色的 LBS 权重 $\mathbf{w}$ ，它们定义了每个关节对每个点的影响。这与先前使用带有艺术家创建的 LBS 权重的模板网格的方法不同 [22, 34, 35, 50]。我们将点$\mathbf{p}$的可学习 LBS 权重初始化为

$$\mathbf{w}^0 = \{w_1, \cdots, w_j\}, \text{ where } w_j = \frac{1}{\sqrt{d_j(\mathbf{p})}}, \quad (2)$$

其中 $d_j(\cdot)$ 表示点 $\mathbf{p}$ 到骨骼 j 的距离。下文中，若帧索引 t 无关紧要，我们将省略它。

3.2. 基于点支架的神经渲染

˜角色图像$\mathbf{I}$ 通过对隐式辐射场和有符号距离场进行体积渲染生成($\mathbf{c}$ s)，如 [58]所示。渲染图像用于优化 NPC，具体内容见第3.4节。我们方法的特点在于，如何将稀疏的各种点特征组合起来，形成一个连续场 $f(\mathbf{q^o})$。给定一个人体姿态 θ 以及观测空间中的一个三维查询点 $\mathbf{q^o}$ ，我们利用神经场预测相应的颜色$\mathbf{c}$ 和有符号距离 s 。

F_ψ ，其可学习参数 ψ 来自 $f(\mathbf{q^o})$，并结合已有的帧编码 $f(t)$ [25, 35] 和几何特征 $g(\mathbf{q^o})$，

$$(\mathbf{c}, s) = F_\psi(f(\mathbf{q^o}), g(\mathbf{q^o}), f(t), \mathbf{w}) . \quad (3)$$

各个特征将在后续章节和补充文档中介绍。接下来，我们介绍从稀疏到连续的变换，以及观测空间与规范空间之间的映射。

点特征到观测空间（前向映射）。我们的逐点特征编码了局部外观和形状。它们与点$\mathbf{p}_i$相关，这些点位于规范空间中。为了将特征从规范空间映射到观测空间，我们首先应用式1，得到三维位置$\mathbf{p^o}$。为了创建连续场，我们查询 $K = 8$ 的最近邻$\mathbf{q^o}$ ，并使用高斯径向基函数核 f_i 聚合特征，

$$f(\mathbf{q^o}) = \sum_{i \in \mathcal{N}(\mathbf{q^o})} f_i a_i, \quad a_i = \exp\left(-\frac{(\mathbf{q^o} - \mathbf{p}_i^\mathbf{o})^2}{\beta_i}\right) , \quad (4)$$

其中 β_j 是一个可学习尺度参数，用于确定

特征 i 在最近邻$\mathcal{N}$中的影响范围（$\mathbf{q}^o$）。

观测空间到骨骼空间（刚性反向映射）。骨架在姿态 θ 下的运动学链为每个关节 j 定义了一个刚性变换$\mathbf{T}^j$ 及其逆变换，

$$\mathbf{q}^j = \mathbf{T}_j^{-1}(\theta)\mathbf{q}^o, \tag{5}$$

将姿态空间中的查询点$\mathbf{q}^o$ 映射到相对于关节的空间。这种方法高效且应用广泛，但忽略了非刚性运动。

观测空间到规范空间（非刚性反向映射）。对非线性的姿态相关形变进行求逆并非易事。以往工作尝试用另一个网络 [30, 34, 47], 来近似这一过程，但这需要以姿态空间中的位置为条件，导致泛化能力较差。另一种方法是迭代求根 [19, 50], ，但无论在训练还是推理时，速度都极其缓慢。

相反，我们利用如下事实：通过公式1 对表面点进行前向映射后，我们可以将这些点作为锚点，逐点求得该函数的逆映射。前向映射会产生有限数量的点对($\mathbf{p}^o$，$\mathbf{p}$)，它们分别位于观测空间和规范空间中，且一一精确对应。此外，形状重建所需的重要特征都位于表面上，因此我们发现，只需在最近邻 f^s 的规范位置处查询规范空间特征$\mathcal{N}(\mathbf{q}^o)$即可。直观地说，我们的表面点$\mathbf{p}$ 充当锚点，将外观信息从规范空间传递到位于观测空间中的点$\mathbf{q}^o$ 。

高效的最近邻查找 我们的前向和反向映射都依赖最近邻搜索。如果对每次查询都直接在整个点云中搜索，速度会很慢。为了使这一方法切实可行，我们采用分而治之的策略，仅在与距离最近的三个关节相关联的点中进行搜索。此外，由于我们的点构成一个表面，我们发现：i）查询点的最近邻在观测空间和规范空间中基本保持不变；ii）最近邻决定了其余的 $K-1$。因此，可以为每个规范点预先记录其 $K-1$ 个最近邻。这将运行时计算简化为在姿态空间中查找一个最近邻，再查表获取其余邻点。因此，即使使用大量邻点，也不会增加运行时开销。请注意，为避免接缝伪影，我们允许预先记录的最近邻跨越关节边界。

3.3. 神经特征与场表示

通过点特征建立不同空间之间的映射后，我们现在转而讨论各个空间中的神经特征结构及其优化。

点特征。我们为每个点 i 关联一个特征向量 f^p，以显式表示对应身体区域的局部信息。这有助于提升不同姿态下纹理的一致性，并捕捉局部高频细-节和几何信息。特征向量 f^p 和位置 $\mathbf{p}_i$均为学习得到的，分别初始化为 $\mathcal{N}(0，0。1)$ 和来自 DANBO 的规范点云。

姿态相关的点特征。我们使用姿态条件化的图神经网络来预测每个点的姿态相关特征 f^θ 和形变 Δp,

$$f^\theta, \Delta p = \text{GNN-FiLM}(\mathbf{p}, \theta)\ . \tag{6}$$

该架构受 DANBO 启发 [43], 并扩展为结合 FiLM 条件化 [36]。

规范特征。我们使用逐部件的因式分解体积 $\mathbf{v}_j \in \mathbf{R}^{12\times 3V}$ [43]来表示 f^s。为了检索点 $\mathbf{p}$的特征，我们将其投影到每个轴对应的因式分解向量上，并进行插值，

$$\mathbf{v}_j(\mathbf{p}) = (v^x[x(\mathbf{p})], v^y[y(\mathbf{p})], v^z[z(\mathbf{p})]) \tag{7}$$

$$f^s = \text{MLP}_{spatial}\left(\mathbf{v}_j(\mathbf{p})\right) \in \mathbf{R}^{3V}\ , \tag{8}$$

其中 $v^{(\cdot)}[\cdot]$ 是插值得到的特征，$\text{MLP}_{spatial}$ 是一个结合所有轴特征的两层多层感知机。

骨骼相对特征。为了使 NeRF 能够推理非线性反向映射中的近似误差，我们以 f^d 为条件输入，即查询点相对于每个最近邻的偏移量；这些偏移量在最近关节 j的局部骨骼坐标系中定义。该距离在学习特征空间$\mathbf{c}_j \in \mathbf{R}^{16\times 3C}$中计算，

$$f_i^d = \mathbf{c}_j(\mathbf{R}_j^{-1}\mathbf{p}_{i,j}^o) - \mathbf{c}_j(\mathbf{R}_j^{-1}\mathbf{q}_j^o), \tag{9}$$

使用$\mathbf{c}j$ 逐部件因式分解特征体积，如$\mathbf{v}$。

最近邻查找后，所有引入的特征都与点相关联$\mathbf{p}_i$ (见图3）。它们共同构成 $f_i = (f_i^p，\ f_i^\theta，\ f_i^s，\ f_i^d)$，用于公式4，它同时包含共享特征和独立特征，并能够适应变形引起的外观变化。补充文档中还介绍了其他常用特征以及 $g(\mathbf{q})$。

3.4. 训练

我们使用真实图像$\mathbf{I}$上的光度损失来训练NPC,

$$L_p = \sum_{(u,v)\in I} |I(u,v,\psi) - \hat{I}(u,v)|. \tag{10}$$

为提高鲁棒性，我们在真实世界数据集上使用L1损失，否则使用L2损失。为鼓励网络预测合适的水平集，我们采用Eikonal损失 [10, 58]

$$L_{eik} = \sum_{\tilde{\mathbf{p}}} (||\nabla s|| - 1)^2\ , \tag{11}$$

其中 $\tilde{\mathbf{p}}$ 是围绕姿态表面点 $\mathbf{p}^o$随机选取的稀疏点。我们通过以下方式对变形网络的输出进行正则化：

$$L_{\Delta p} = \max(\Delta p - \delta, 0)^2, \tag{12}$$

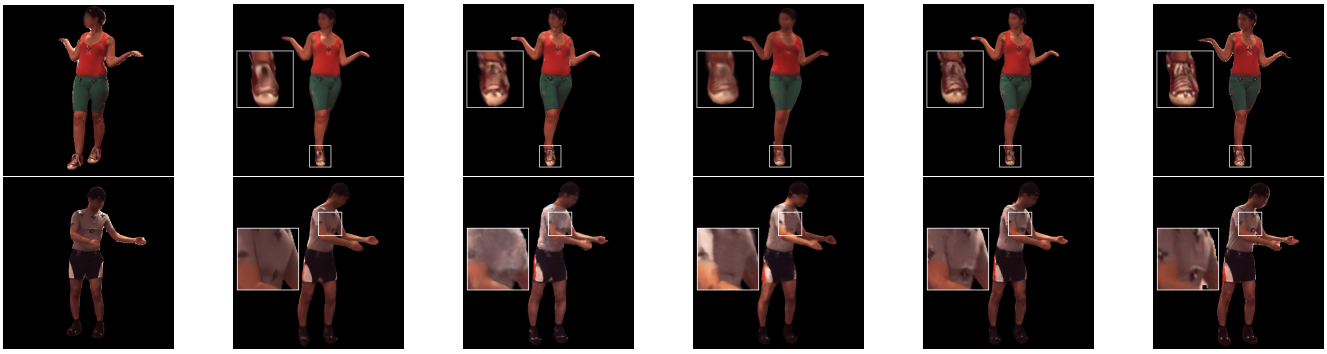


图4： **Anim-NeRF Human3.6M 上的新视角合成 [12, 13, 34]测试划分**。我们的基于点的表示确保不同训练姿态之间具有更好的特征一致性，因此能够为新视角合成出更精细的细节。

当 Δp 超过阈值 $\delta = 0.04$时，该损失会惩罚非线性变形。第二个损失则会抑制网络将点移离其邻居，

$$L_N = \sum_{\mathbf{p^o}} \sum_{\mathbf{p}_i \in \mathcal{N}(\mathbf{p})} (||\mathbf{p^o} - \hat{\mathbf{p}}_i^{\mathbf{o}}|| - ||\mathbf{p} - \hat{\mathbf{p}}_i||)^2. \quad (13)$$

我们加入第三个损失，以鼓励预测的表面点有符号距离位于人体表面上或其内部，

$$L_s = \sum_{\tilde{\mathbf{p}}} \max(s, 0). \quad (14)$$

总而言之，我们完整的训练目标为

$$L = L_p + \lambda_{eik} L_{eik} + \lambda_{\Delta p} L_{\Delta p} + \lambda_N L_N + \lambda_S L_S, \quad (15)$$

其中 $\lambda_{eik} = 0.01$、$\lambda_{\Delta p} = 1.0$、$\lambda_N = 10.0$ 和 $\lambda_S = 0$。1。我们训练 15 万次迭代，耗时 12 小时，使用单块 NVIDIA RTX 3080 GPU，包括点初始化过程。

4. 实验

我们量化了 NPC 相较于最新的无表面方法 TAVA [19], DANBO [43]和 A-NeRF [44], ，以及利用模板或基于扫描的先验的最新且成熟方法（包括 ARAH [50], Anim-NeRF [34] 和 NeuralBody [35]）所取得的改进。此外，我们通过消融研究验证了所提出的逐点编码、姿态相关特征和坐标特征如何帮助提升感知质量。补充视频和文档提供了更多定性结果和实现细节。代码将在论文发表后公开，以促进人体建模研究[2]。

[2]所有数据来源、建模代码和实验均由不列颠哥伦比亚大学开发。Meta 未获取本研究中的数据或代码，也未开展任何实验。

评估指标。我们采用标准图像指标，包括逐像素峰值信噪比（PSNR）和结构相似性（SSIM） [51], ，以及感知指标学习感知图像块相似度（LPIPS） [60] 和 KID [4, 33] ，这些指标能够更准确地衡量轻微错位时纹理细节方面的改进。在仅使用稀疏姿态驱动精细角色时，这种错位不可避免。所有分数均在留出测试集上计算。具体而言，我们通过从未见过的测试视角渲染已见姿态，评估多视角数据集上的新视角合成性能；并通过合成训练期间未见过的人体姿势的外观，评估新姿态合成性能。

数据集。 我们在已确立的人体建模基准测试上评估 NPC，涵盖室内、室外、单目视频采集以及合成动物数据集。

- Human3.6M [12, 13]: 我们采用 Anim-NeRF [34], 的评估方案，在总共 7 名受试者上进行评估，并使用作者提供的前景图。

- MonoPerfCap [57]: 该数据集由多个单目户外视频组成。我们使用与 [44] 相同的 2 名受试者。训练姿态和相机参数是 SPIN 的估计结果 [15]，并由 A-NeRF 进一步优化。

此外，我们在 ZJU-Mocap [8, 35]，中一名身穿宽松服装的受试者上评估 NPC，并使用具有挑战性的运动序列（包括来自 AIST++ [20] 和 SURREAL+ CMU-Mocap [6, 48] 数据集的舞蹈和体操姿态）来驱动我们学习到的角色。最后，我们使用来自 TAVA [19] 的合成狼数据集作为概念验证，展示如何像其他无模板方法 [19, 30, 43] 一样将 NPC 应用于非人类对象。

4.1. 新视角合成

我们的基于点的表示能够锚定所有训练视角共享的高频局部细节，因此即使从训练数据中未出现的新视角进行渲染，也能呈现更丰富的视觉细节。与 ARAH [50] 采用隐式表面先验 [49] 并通过求根进行规范映射相比，NPC 仅使用提取出的含噪表面点，也能合成更清晰的结果

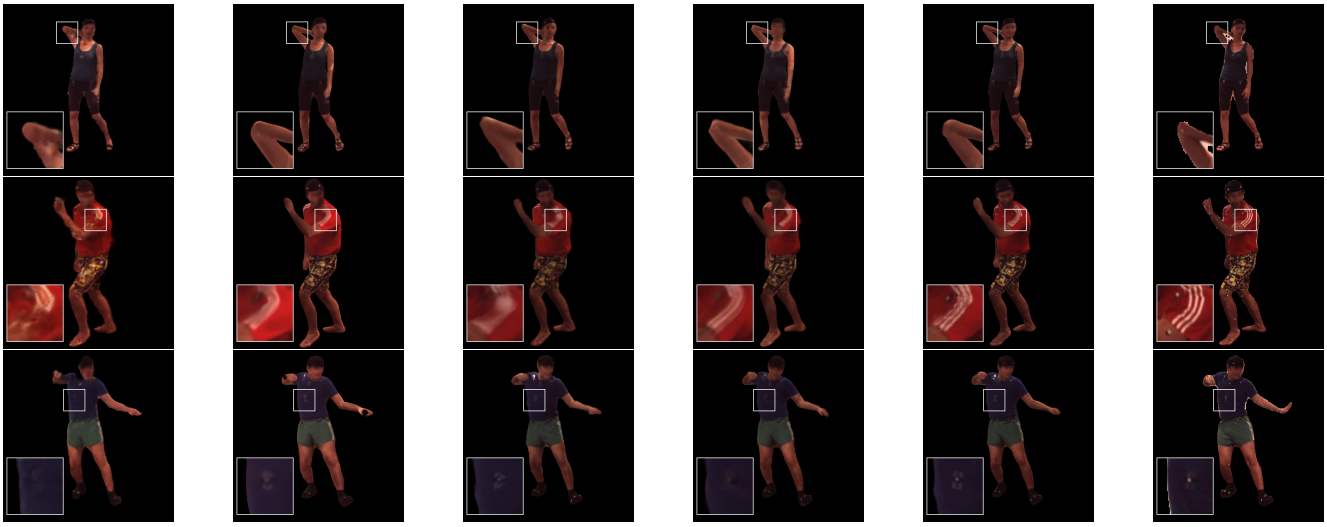


图5： **Anim-NeRF Human3.6M 上的未见姿态合成 [12, 13, 34] 测试划分**。我们提出的 NPC 在未见姿态下展现出更好的外观一致性，关节运动也得到改善，纹理更加清晰，并能呈现运动追踪器和条纹等精细细节。

表1：**Anim-NeRF Human3.6M 上的新视角合成对比 [12, 13, 34]测试划分**。我们提出的 NPC 得益于基于点的表示，整体感知质量更佳。

	S1			S5			S6			S7			S8			S9			S11			Avg		
	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
基于模板/扫描的先验																								
NeuralBody	22.88	0.897	0.139	24.61	0.917	0.128	22.83	0.888	0.155	23.17	0.915	0.132	21.72	0.894	0.151	24.29	0.911	0.122	23.70	0.896	0.168	23.36	0.905	0.140
Anim-NeRF				23.40			22.85			21.97						24.86			24.76			23.34		
ARAH†	24.53	0.921	0.103	24.67	0.921	0.115	24.37	0.904	0.133	24.41	0.922	0.115	**24.15**	**0.924**	**0.104**	25.43	0.924	0.112	24.76	0.918	0.128	24.63	0.920	0.115
无模板																								
A-NeRF	23.93	0.912	0.118	24.67	0.919	0.114	23.78	0.887	0.147	24.40	0.917	0.125	22.70	0.907	0.130	25.58	0.916	0.126	24.38	0.905	0.152	24.26	0.911	0.129
DANBO	23.95			24.86			24.54			24.45						26.15			25.58			24.73		
TAVA	**25.28**	**0.928**	0.108	24.00	0.916	0.122	23.44	0.894	0.138	24.25	0.916	0.130	23.71	0.921	0.116	26.20	0.923	0.119	**26.17**	**0.928**	0.133	24.72	0.919	0.124
NPC（我们的方法）																								

....：我们使用官方发布的 ARAH 进行评估，该版本经过了重构，因此数值与 [50] 中的略有不同。

，无需监督。图 4 展示了我们恢复的外观细节更好，例如标记物和鞋带。我们观察到，TAVA 虽然像 ARAH 一样使用求根进行映射，但模糊伪影更为明显。我们推测，对于动作复杂的视频序列，求根算法需要良好的初始化。表1量化了相较于近期无模板方法以及基于模板/扫描的神经人体场方法所取得的改进。

4.2. 未见姿态与 NPC 动画

未见姿态合成要求该方法在分布外的姿态配置下保持外观一致。如图 5 所示，我们的 NPC 整体上能生成更精细且一致的渲染结果。与 Neural Body [35], 基于模板人体网格锚定人体表示，并使用三维卷积神经网络扩散特征不同，我们的 NPC 能生成更清晰的细节，例如动作捕捉人体追踪器和皱纹，同时保持更好的纹理一致性，例如条纹更加清晰。我们将此归因于我们基于点的特征编码，以及对表面点显式施加的非形变约束；两者结合，使人体特征能够得到更精确的定位。表 2 和表 3 验证了我们的 NPC 在感知质量方面持续提升，不仅适用于多视角数据集 Hu-man3.6M [12, 13], ，也适用于户外单目视频序列 MonoPerfCap [57]。请注意，表 3 未列出峰值信噪比和结构相似性，因为这些指标容易受到户外序列中光照条件变化的影响。

我们进一步使用图 6 中分布外人体姿态的运动序列，为训练得到的 NPC 模型生成动画，以展示我们基于点的角色在动画中的潜在应用。请注意，由于没有可用于评估的真实图像，我们无法量化这些示例的性能。

几何对比。即使输入是单目视频，NPC 也能重建细节丰富的三维几何，并能很好地泛化到未见过的姿势，同时生成整体上更加完整的人体轮廓，如图 7所示。与

表2： **Anim-NeRF Human3.6M 上未见姿态渲染的比较 [12, 13, 34]测试划分**。我们基于点的方法所实现的高效规范特征映射，使其在未见姿态下具有更好的泛化能力。

	S1	S5	S6	S7	S8	S9	S11	Avg
	峰值信噪比↑结构相似性↑学习感知图像块相似度↓峰值信噪比 结构相似性 学习感知图像块相似度	峰值信噪比 结构相似性 学习感知图像块相似度	峰值信噪比 结构相似性 学习感知图像块相似度	峰值信噪比 结构相似性 学习感知图像块相似度	峰值信噪比↑结构相似性↑学习感知图像块相似度↓峰值信噪比 结构相似性	峰值信噪比 结构相似性 学习感知图像块相似度	峰值信噪比 结构相似性 学习感知图像块相似度	峰值信噪比 结构相似性 学习感知图像块相似度
基于模板/扫描的先验								
NeuralBody	22.10 0.878 0.143	23.52 0.897 0.144	23.42 0.892 0.146	22.59 0.893 0.163	20.94 0.876 0.172	23.05 0.885 0.150	23.72 0.884 0.179	22.81 0.888 0.157
Anim-NeRF		22.29	22.59	22.22		23.73	23.92	22.61
ARAH†	23.18 0.903 0.116	22.91 0.894 0.133	23.91 0.901 0.125	22.72 0.896 0.143	22.50 0.899 0.128	24.15 0.896 0.135	23.93 0.899 0.143	23.27 0.897 0.134
无模板								
A-NeRF	22.67 0.883 0.159	22.96 0.888 0.155	22.77 0.869 0.170	22.80 0.880 0.182	21.95 0.886 0.170	24.16 0.889 0.164	23.40 0.880 0.190	23.02 0.883 0.171
DANBO	23.03	**23.66**	24.57	23.08		24.79	24.57	23.74
TAVA	**23.83 0.908** 0.120	22.89 0.898 0.135	24.54 0.906 0.122	22.33 0.882 0.163	22.50 0.906 0.130	24.80 0.901 0.138	**25.22 0.913**0.145	23.52 0.899 0.141
NPC（我们的		23.63	**24.59**	**23.46**		**24.86**	25.13	**23.96**

†: 我们使用官方发布的 ARAH 进行评估。该版本经过了重构，因此结果与 [50] 中的数字略有不同。

表3：**MonoPerfCap 上的未见姿态合成 [57]**。在利用单目视频学习泛化模型方面，NPC 的整体感知质量优于 DANBO。

	ND		WP		Avg	
	KIDx100 ↓ 学习感知图像块相似度 ↓		KIDx100 ↓ 学习感知图像块相似度 ↓		KIDx100 ↓ 学习感知图像块相似度 ↓	
A-NeRF	4.97	0.197	6.53	0.223	5.75	0.208
DANBO	4.83	**0.194**	4.66	0.214	4.74	0.202
NPC（我们的方法）	**2.57**	0.198	**3.56**	**0.207**	**3.07**	**0.202**



训练示例　参考 pose　DANBO　NPC（我们的方法）

图6：**在多个受试对象上，将分布外姿态进行运动重定向**。NPC 保持了更好的外观一致性和纹理细节。

DANBO 和其他隐式方法不同，我们的点支架能够提供跨帧对应关系，因此适用于稠密三维姿态估计。最后，我们展示了 NPC 能够恢复不完美几何，例如修正图 8中的缺失的脚部和裤子形状。

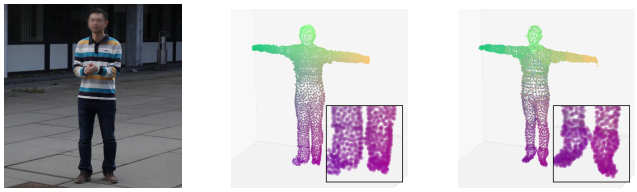
4.3. NPC 在宽松服装上的形变

为进一步比较依赖姿态的形变，我们在 ZJU-Mocap 的受试对象 387 上进行了测试 [8, 35],，该对象包含宽松服装和长程形变依赖性。我们按照既定协议报告了表4中的结果 [35]。NPC 匹配或超出



GT　DANBO　NPC（我们的方法）　GT　DANBO NPC（我们的方法）

图7：**未见姿态下的身体几何**。NPC 的形状质量与 DANBO 相当。



参考　初始点　优化后的 NPC

图 8：NPC 根据 [43] 获得的噪声初始化恢复精细几何结构。

表 4：**ZJU-Mocap [35] 受试对象 387 的新视角与新姿态合成**。NPC 的 FID 和 KID 指标更优。

	ARAH			NPC（我们的方法）		
	峰值信噪比 学习感知图像块相似度	弗雷歇起始距离 KID×100		峰值信噪比 ↑ 学习感知图像块相似度 ↓	弗雷歇起始距离 ↓	KID×100 ↓
新视角	**25.83** 0.096	38.2	1.17	25.10	**0.093**	**36.1**	**0.94**
新姿态	**22.9**		2.74	21.88	0.134	**49.2**	**2.59**

在感知指标 [50]（学习感知图像块相似度、KID、弗雷歇起始距离）上，针对新视角的表现优于 ARAH；在针对新姿态的 KID 和弗雷歇起始距离指标上也优于 ARAH，尽管 ARAH 使用了在大规模扫描数据集上预训练的先验 [49]。进一步分析（见图9）发现，纹理细节有所改善，而姿态相关的皱纹细节与之相当，甚至更好。ARAH 生成的结果过于平滑，而 NPC 的结果略显颗粒感，因此峰值信噪比较低，但弗雷歇起始距离和 KID 指标有所改善。



训练示例　　ARAH　　NPC（我们的方法）

图 9：**ZJU-Mocap 受试对象 387 的新视角合成结果**。与 ARAH 相比，NPC 的结果更清晰，但略有颗粒感。



真实值　　初始点　　NPC

图 10：NPC 可用于学习非人类对象，因为它不需要预训练表面先验。

4.4. 动物上的 NPC

我们验证了 NPC 在 TAVA 幼狼对象上的适用性，该对象没有经过预训练的表面先验 [19]。我们对 NPC 进行了 1 万次迭代的轻量训练，并在图10中展示了一个分布外姿态的渲染结果。

4.5. 消融研究

我们使用 Human3.6M S9 进行消融研究。在表5中，我们展示了所提出的所有特征都有助于提升最终的感知质量。若没有 f^{θ}，渲染结果会变得嘈杂。若没有 f^{p}，模型生成的纹理会前后不一致且模糊。禁用 Δp 会降低具有更极端形变的姿态的感知质量，因为此时最近邻映射会变得不准确。在表6中，我们发现 NPC 对点的数量并不敏感，使用超过200 个点（我们的默认值）带来的收益会逐渐减少。我们还通过运行训练框架 4 次，进一步验证 NPC 能够以稳定的质量学习角色模型；其中包括从头开始重新训练 DANBO，以初始化规范点云。如表7所示，在 Human3.6M S9 和 MonoPerfCap weipeng 上，不同运行之间的性能波动都很小。

5. 局限性与讨论

尽管 NPC 的运行速度优于大多数现有的隐式神经人体模型，但它仍使用由中等规模网络编码的神经场，

表 5：**对我们提出的各项设计进行消融研究**。它们均有助于提升最终的感知质量。

	KIDx100 ↓	学习感知图像块相似度 ↓
无 f^{d}	5.54	0.122
无 f^{p}	4.51	0.122
无 f^{θ}	4.50	0.120
无 Δp	4.61	0.120
NPC（我们的方法）	**4.43**	**0.115**

表 6：**不同身体部位点数的消融研究**。在 Human 3.6M 上，NPC 在不同点数设置下表现一致 [12]。

40×19		**200×19**		120×19		280×19		360×19	
峰值信噪比 ↑	学习感知图像块相似度 ↓	峰值信噪比 ↑	学习感知图像块相似度 ↓	峰值信噪比 ↑	学习感知图像块相似度 ↓	峰值信噪比 ↑	学习感知图像块相似度 ↓	峰值信噪比 ↑	学习感知图像块相似度 ↓
24.83	0.123	**24.88**	0.117	24.86	0.115	24.87	**0.114**	24.84	0.117

表 7：**使用 4 种不同规范点云初始化进行训练的消融研究**。括号中报告的是标准差。这表明我们的点初始化策略可靠，且 NPC 在不同训练运行中表现一致。

S9			WP		
KIDx100 ↓	LPIPS（VGG）↓	LPIPS（Alex）↓	KIDx100 ↓	LPIPS（VGG）↓	LPIPS（Alex）↓
4.34 (± 0.12)					

这使其实时应用成为不可能。当稀疏点无法完全覆盖人体表面时，我们的方法有时会产生小尺度的球状伪影。根据 Point-NeRF [56] 中提出的方法，这一问题有望通过点增长得到解决。详情请参阅补充文档。最后，NPC 能够从原始视频序列中学习具有精细外观的人物专属模型，为公众创建 3D 角色提供了一种更便捷的方式。另一方面，它也可能被用于深度伪造，在被恶意使用时引发伦理问题。

6. 结论

NPC 表明，无需使用激光扫描技术，也无需对底层模板模型作出严格限制，便可学习到细节丰富的人物模型。由于整个过程完全自动化，数字化全身虚拟形象得以惠及更广泛的人群，包括在现有数据集中代表性不足的少数群体。

致谢。感谢 Shaofei Wang 和 Rui- long Li 就 ARAH 和 TAVA 相关问题展开的有益讨论。感谢 Luis A. Bolaños 提供的帮助和参与的讨论，也感谢 Frank Yu、Chunjin Song、Xingzhe He 和 Eric Hedlin 提出的宝贵意见。我们还感谢不列颠哥伦比亚大学的 ARC 和 Compute Canada 提供计算资源。

参考文献

[1] Kara-Ali Aliev、Artem Sevastopolsky、Maria Kolos、Dmitry Ulyanov 和 Victor Lempitsky。基于神经点的图形学。载于 ECCV，2020 年。2[2] Timur Bagautdinov、Chenglei Wu、Tomas Simon、Fabian Prada、Takaaki Shiratori、Shih-En Wei、Weipeng Xu、Yaser Sheikh 和 Jason Saragih。驾驶信号感知的全身虚拟形象。ACM TOG（SIGGRAPH 论文集），2021 年。1, 2[3] Ziqian Bai、Timur Bagautdinov、Javier Romero、Michael Zollhöfer、Ping Tan 和 Shunsuke Saito。AutoAvatar：用于动态虚拟形象建模的自回归神经场。载于 ECCV，2022 年。3[4] Mikołaj Bi´nkowski、Danica J Sutherland、Michael Arbel 和 Arthur Gretton。揭秘 MMD 生成对抗网络。载于 ICLR，2018 年。6[5] Xu Chen、Yufeng Zheng、Michael J Black、Otmar Hilliges 和 Andreas Geiger。SNARF：用于驱动非刚性神经隐式形状的可微前向蒙皮。载于 ICCV，2021 年。3[6] CMU 图形实验室动作捕捉数据库。http://mocap.cs.cmu.edu。6[7] Boyang Deng、John P Lewis、Timothy Jeruzalski、Gerard Pons-Moll、Geoffrey Hinton、Mohammad Norouzi 和 An- drea Tagliasacchi。NASA 神经关节形状近似。载于 ECCV，2020 年。3[8] Qi Fang、Qing Shuai、Junting Dong、Hujun Bao 和 Xiaowei Zhou。通过观看镜中的人体来重建三维人体姿态。载于 CVPR，2021 年。6, 8[9] Chen Gao、Ayush Saraf、Johannes Kopf 和 Jia-Bin Huang。从动态单目视频合成动态视图。载于 ICCV，2021 年。2[10] Amos Gropp、Lior Yariv、Niv Haim、Matan Atzmon 和 Yaron Lipman。用于学习形状的隐式几何正则化。载于 ICML，2020 年。5[11] David Ha、Andrew M. Dai 和 Quoc V. Le。超网络。载于 ICLR，2017 年。3[12] C. Io nescu、F. Li 和 C. Sminchisescu。用于人体姿态估计的潜在结构模型。载于 ICCV，2011 年。6, 7, 8, 9[13] C. Ionescu、I. Papava、V. Olaru 和 C. Sminchisescu。Human3.6M：自然环境中三维人体感知的大规模数据集与预测方法。TPAMI，2014 年。6, 7, 8[14] Wei Jiang、Kwang Moo Yi、Golnoosh Samei、Oncel Tuzel 和 Anurag Ranjan。Neuman：从单段视频重建神经人体辐射场。载于 ECCV，2022 年。2, 3[15] Nikos Kolotouros、Georgios Pavlakos、Michael J Black 和 Kostas Daniilidis。通过在环模型拟合学习重建三维人体姿态和形状。载于 ICCV，2019 年。2, 6[16] Youngjoong Kwon、Dahun Kim、Duygu Ceylan 和 Henry Fuchs。Neural Human Performer：学习用于人体表演渲染的可泛化辐射场。NeurIPS，2021 年。1, 2, 3[17] Christoph Lassner 和 Michael Zollhöfer。Pulsar：高效的基于球体的神经渲染。载于 CVPR，2021 年。2[18] John P Lewis、Matt Cordner 和 Nickson Fong。姿态空间变形：一种统一的形状插值方法，以及骨架驱动变形。发表于 ACM TOG（SIGGRAPH 论文集），2000 年。2, 3[19] Ruilong Li、Julian Tanke、Minh Vo、Michael Zollhofer、Jurgen Gall、Angjoo Kanazawa 和 Christoph Lassner。TAVA：无模板的可动画体积演员。发表于 ECCV，2022 年。2, 3, 5, 6, 9[20] Ruilong Li、Shan Yang、David A. Ross 和 Angjoo Kanazawa。AI 编舞师：音乐条件驱动的 AIST 三维舞蹈生成。发表于 ICCV，2021 年。6[21] Zhengqi Li、Simon Niklaus、Noah Snavely 和 Oliver Wang。用于动态场景时空视图合成的神经场景流场。发表于 CVPR，2021 年。2[22] Lingjie Liu、Marc Habermann、Viktor Rudnev、Kripasindhu Sarkar、Jiatao Gu 和 Christian Theobalt。Neural Actor：通过姿态控制对真人演员进行神经自由视点合成。ACM TOG（SIGGRAPH Asia 论文集），2021 年。1, 2, 3, 4[23] Stephen Lombardi、Tomas Simon、Gabriel Schwartz、Michael Zollhoefer、Yaser Sheikh 和 Jason Saragih。用于高效神经渲染的体积基元混合。ACM TOG（SIGGRAPH 论文集），2021 年。1[24] Qianli Ma、Jinlong Yang、Siyu Tang 和 Michael J. Black。点在着装人体建模中的力量。发表于 ICCV，2021 年。2[25] Ricardo Martin-Brualla、Noha Radwan、Mehdi S. M. Sajjadi、Jonathan T. Barron、Alexey Dosovitskiy 和 Daniel Duckworth。野外 NeRF：用于无约束照片集的神经辐射场。发表于 CVPR，2021 年。4[26] Moustafa Meshry、Dan B Goldman、Sameh Khamis、Hugues Hoppe、Rohit Pandey、Noah Snavely 和 Ricardo Martin-Brualla。野外神经重渲染。发表于 CVPR，2019 年。2[27] Marko Mihajlovic、Aayush Bansal、Michael Zollhoefer、Siyu Tang 和 Shunsuke Saito。KeypointNeRF：使用关键点相对空间编码实现基于图像的体积化虚拟形象泛化。发表于 ECCV，2022 年。2, 3[28] Marko Mihajlovic、Shunsuke Saito、Aayush Bansal、Michael Zollhoefer 和 Siyu Tang。COAP：人物的组合式关节占据表示。发表于 CVPR，2022 年 6 月。3[29]Ben Mildenhall、Pratul P Srinivasan、Matthew Tancik、Jonathan T. Barron、Ravi Ramamoorthi 和 Ren Ng。NeRF：将场景表示为神经辐射场以进行视图合成。发表于 ECCV，2020 年。2, 3[30] Atsuhiro Noguchi、Xiao Sun、Stephen Lin 和 Tatsuya Harada。神经关节辐射场。发表于 ICCV，2021 年。2, 5, 6[31] Atsuhiro Noguchi、Xiao Sun、Stephen Lin 和 Tatsuya Harada。通过无监督学习实现高效的几何感知神经关节表示。发表于 ECCV，2022 年。2[32] Keunhong Park、Utkarsh Sinha、Peter Hedman、Jonathan T. Barron、Sofien Bouaziz、Dan B Goldman、Ricardo Martin-Brualla 和 Steven M. Seitz。HyperNeRF：用于拓扑变化神经辐射场的高维表示。ACM TOG（SIGGRAPH 论文集），2021 年。2[33] Gaurav Parmar、Richard Zhang 和 Jun-Yan Zhu。关于混叠缩放及 GAN 评估中的意外细节。发表于 CVPR，2022 年。6

[34] Sida Peng、Junting Dong、Qianqian Wang、Shangzhan Zhang、Qing Shuai、Xiaowei Zhou 和 Hujun Bao。用于动态人体建模的可动画神经辐射场。发表于 ICCV，2021 年。1, 2, 3, 4, 5, 6, 7, 8[35] Sida Peng、Yuanqing Zhang、Yinghao Xu、Qianqian Wang、Qing Shuai、Hujun Bao 和 Xiaowei Zhou。Neural Body：通过结构化潜代码实现动态人类新视角合成的隐式神经表示。发表于 CVPR，2021 年。1, 2, 3, 4, 6, 7, 8[36] Ethan Perez、Florian Strub、Harm de Vries、Vincent Dumoulin 和 Aaron C. Courville。FiLM：使用通用条件层进行视觉推理。发表于 AAAI，2018 年。5[37] Sergey Prokudin、Michael J Black 和 Javier Romero。SMPLpix：基于三维人体模型的神经化身。发表于 WACV，2021 年。2[38] Albert Pumarola、Enric Corona、Gerard Pons-Moll 和 Francesc Moreno-Noguer。D-NeRF：用于动态场景的神经辐射场。发表于 CVPR，2020 年。2[39]Edoardo Remelli、Timur Bagautdinov、Shunsuke Saito、Chenglei Wu、Tomas Simon、Shih-En Wei、Kaiwen Guo、Zhe Cao、Fabian Prada、Jason Saragih 等。使用纹素对齐特征的可驱动体积化虚拟形象。发表于 ACM TOG（SIGGRAPH 论文集），2022 年。1[40] Darius Rückert、Linus Franke 和 Marc Stamminger。ADOP：可微单像素点渲染的近似方法。ACM TOG（SIGGRAPH 论文集），2022 年。2[41] S hunsuke Saito、Jinlong Yang、Qianli Ma 和 Michael J. Black。SCANimate：蒙皮服装化身网络的弱监督学习。发表于 CVPR，2021 年。3[42] Vincent Sitzmann 、 Julien Martel 、 Alexander Bergman、David Lindell 和 Gordon Wetzstein。采用周期性激活函数的隐式神经表示。发表于 NeurIPS，2020 年。2[43]Shih-Yang Su、Timur Bagautdinov 和 Helge Rhodin。DANBO：通过图神经网络实现解耦关节式神经人体表示。发表于 ECCV，2022 年。1, 2, 3, 5, 6, 8[44] Shih-Yang Su、Frank Yu、Michael Zollhöfer 和 Helge Rhodin。A-NeRF：用于学习人体形状、外观和姿态的关节式神经辐射场。发表于 NeurIPS，2021 年。1, 2, 3, 6[45] A. Tewari、O. Fried、J. Thies、V. Sitzmann、S. Lo mbardi、K. Sunkavalli、R. Martin-Brualla、T. Simon、J. Saragih、M. Nießner、R. Pandey、S. Fanello、G. Wetzstein、J.-Y. Zhu、C. Theobalt、M. Agrawala、E. Shechtman、D. B Goldman 和 M. Zollhöfer。《神经渲染的最新进展》。Computer Graphics Forum（EG STAR 2020），2020 年。2[46] Justus Thies、Michael Zollhöfer 和 Matthias Nießner。延迟神经渲染：使用神经纹理进行图像合成。ACM TOG（SIGGRAPH 论文集），2019 年。1[47] Garvita Tiwari、Nikolaos Sarafianos、Tony Tung 和 Gerard Pons-Moll。Neural-GIF：用于动画化着装人物的神经广义隐式函数。发表于 ICCV，2021 年。3, 5[48] Gül Varol、Javier Romero、Xavier Martin、Naureen Mahmood、Michael J. Black、Ivan Laptev 和 Cordelia Schmid。从合成人体中学习。发表于 CVPR，2017 年。6[49] Shaofei Wang、Marko Mihajlovic、Qianli Ma、Andreas Geiger 和 Siyu Tang。MetaAvatar：学习可动画化的

通过少量深度图像生成着装人体模型。在 NeurIPS，2021。2, 3, 6, 8[50] Shaofei Wang、Katja Schwarz、Andreas Geiger 和 Siyu Tang。Arah：可动人体有符号距离场的可动画体渲染。发表于欧洲计算机视觉会议，2022。2, 3, 4, 5, 6, 7, 8[51] Zhou Wang、Alan C Bovik、Hamid R Sheikh 和 Eero P Simoncelli。图像质量评估：从误差可见性到结构相似性。TIP，2004。6[52] Chung-Yi Weng、Brian Curless、Pratul P. Srinivasan、Jonathan T. Barron 和 Ira Kemelmacher-Shlizerman。Hu-manNeRF：从单目视频渲染运动人物的自由视点视频。发表于 CVPR，2022。2[53] Olivia Wiles、Georgia Gkioxari、Richard Szeliski 和 Justin Johnson。Synsin：从单张图像端到端进行视图合成。发表于 CVPR，2020。2[54] Wenqi Xian、Jia-Bin Huang、Johannes Kopf 和 Changil Kim。用于自由视点视频的时空神经辐照度场。发表于 CVPR，2021。2[55] Hongyi Xu、Thiemo Alldieck 和 Cristian Sminchisescu。H-nerf：用于渲染运动人体及其时序重建的神经辐射场。NeurIPS，2021。2[56] Qiangeng Xu、Zexiang Xu、Julien Philip、Sai Bi、Zhixin Shu、Kalyan Sunkavalli 和 Ulrich Neumann。Point-nerf：基于点的神经辐射场。发表于 CVPR，2022。2, 9[57] W. Xu、A. Chatterjee、M. Zollhöfer、H. Rhodin、D. Mehta、H.-P. Seidel 和 C. Theobalt。Monoperfcap：基于单目视频的人体表演捕捉。ACM TOG（SIGGRAPH 论文集），2018。6, 7, 8[58] Lior Yariv、Jiatao Gu、Yoni Kasten 和 Yaron Lipman。神经隐式表面的体渲染。发表于 NeurIPS，2021。4, 5[59] Ilya Zakharkin、Kirill Mazur、Artur Grigorev 和 Victor Lempitsky。基于点的人体服装建模。发表于 ICCV，2021。2[60] Richard Zhang、Phillip Isola、Alexei A Efros、Eli Shechtman 和 Oliver Wang。深度特征作为感知度量的惊人效能。发表于 CVPR，2018。6[61] Yufeng Zheng、Yifan Wang、Gordon Wetzstein、Michael J Black 和 Otmar Hilliges。Pointavatar：基于视频的可变形基于点的头部数字人。发表于 CVPR，2023。2[62] Zerong Zheng、Han Huang、Tao Yu、Hongwen Zhang、Yandong Guo 和 Yebin Liu。用于人体数字人建模的结构化局部辐射场。发表于 CVPR，2022。2, 3