计算机科学
COMPUTER SCIENCE

基于多尺度层次网络的人体重建神经辐射场

王洋，王国栋，赵俊莉，盛筱萌

引用本文

王洋，王国栋，赵俊莉，盛筱萌. 基于多尺度层次网络的人体重建神经辐射场[J]. 计算机科学，2025，52(11)：175-183.

WANG Yang, WANG Guodong, ZHAO Junli, SHENG Xiaomeng. Neural Radiance Field for Human Reconstruction Based on Multi-scale Hierarchical Network [J]. Computer Science, 2025, 52(11): 175-183.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240900141

基于多尺度层次网络的人体重建神经辐射场

王 洋 王国栋 赵俊莉 盛筱萌
青岛大学计算机科学技术学院 山东 青岛 266071
(wangyang0689@qdu.edu.cn)

摘 要 单目RGB视频中的三维人体重建面临着准确捕捉人体姿态的挑战,尤其在使用诸如SMPL人体先验模型时,其刚性假设限制,难以描述姿态的细微变化,导致重建效果不理想。此外,现有的基于神经辐射场的人体建模方法在处理未见过的姿态时,容易在局部区域产生不自然的阴影或漂浮现象,且在纹理细节的呈现上有所不足。为了解决这些问题,提出了一种基于三平面多尺度分解网络,旨在通过神经辐射场方法增强三维人体的纹理细节,并提高模型对不同姿态的泛化能力。在方法上,使用多分辨率哈希编码技术替代传统的三角频率编码函数,能够更高效地捕获人体的高频特征,并加快模型的收敛速度。三平面多尺度学习策略被应用于人体姿态的细节捕捉,从而有效提高了三维重建的精度与视觉质量。在实验中,所提出的改进方法显著提升了人体三维模型的重建效果,尤其在处理复杂的姿态变化时表现突出。该方法在训练速度、渲染质量以及姿态泛化能力上均优于传统方法,展示出较大的优势。应用该模型生成的三维人体模型在细节上更加逼真,且在新颖姿态下的合成结果表现良好,进一步推动了单目视频中的三维人体重建技术的发展。

关键词:神经网络辐射场;蒙皮多人线性模型;人体重建;深度学习;多层感知机
中图分类号 TP391.4

Neural Radiance Field for Human Reconstruction Based on Multi-scale Hierarchical Network

WANG Yang,WANG Guodong,ZHAO Junli and SHENG Xiaomeng
College of Computer Science and Technology,Qingdao University,Qingdao,Shandong 266071,China

Abstract The reconstruction of 3D human models from monocular RGB video faces challenges in accurately capturing human poses,especially when using prior models like SMPL. Due to its rigid assumptions,such models struggle to depict subtle pose variations,leading to suboptimal reconstruction results. Additionally,existing NeRF-based human modeling methods often generate unnatural shadows or floating artifacts around certain body parts when rendering unseen poses,and their representation of texture details tends to be insufficient. To address these issues,this paper proposes a hierarchical network based on the Triplane Multiscale learning,aims at enhancing the texture details of 3D human models through NeRF techniques and improving the model's generalization capability across different poses. In terms of methodology,multi-resolution hash encoding is employed to replace the traditional sinusoidal frequency encoding function,allowing for more efficient capture of high-frequency human features and speeding up model convergence. The Triplane Multiscale learning strategy is applied to capture pose details,effectively improving the accuracy and visual quality of 3D reconstructions. Experiments demonstrate that the proposed improvements significantly enhance the reconstruction of 3D human models,especially when handling complex pose variations. The method shows notable advantages in terms of training speed,rendering quality,and pose generalization capabilities. By applying this model,the resulting 3D human models exhibit more realistic details,and the synthesized results for novel poses are of high quality,further advancing the development of 3D human reconstruction technology from monocular video.

Keywords Neural radiance field,SMPL,Human reconstruction,Deep learning,MLP

1 引言

创建和渲染高保真的数字人在许多应用中至关重要,包括沉浸式远程呈现,新兴的元宇宙虚拟世界、游戏或电影制作,以及支持远程写作、教育和娱乐。然而,在现有的获取高质量三维人体重建的方法中,需要大量研究人员的努力和昂贵的扫描设备,这极大地阻碍了该技术在其他应用中的发展。本文旨在通过从日常生活中最容易获得的单目RGB视频中

到稿日期:2024-09-23 返修日期:2025-02-06
基金项目:国家自然科学基金(62172247);青岛市自然科学基金(23-2-1-163-zyyd-jch)
This work was supported by the National Natural Science Foundation of China(62172247) and Qingdao Natural Science Foundation(23-2-1-163-zyyd-jch).
通信作者:王国栋(doctorwgd@gmail.com)

获取高质量的 3D 数字人来解决这一问题。为此，本研究引入了 Neural Radiance Field(NeRF)[1-3]来隐式地重建人体的几何和外观，并通过体绘制生成逼真的图像。这种方法通过训练一个大型的多层感知器(Multilayer Perceptron，MLP)来实现，在姿势无关的规范空间中模拟人体形状和外观。MLP 接受规范空间位置上的点作为输入，输出对应点在空间中的颜色和密度，最后进行体渲染以获得最终的像素颜色。但是这类方法通常需要昂贵的计算资源和较长的时间成本，且在未见过的人体姿势上表现不佳。本文目标是设计一种可广泛应用于单目视频中、能够快速高效地重建三维人体的方法。为了更好地学习人体的规范形状和外观，将神经网络辐射场与参数化人体模型 SMPL[4]结合，通过对骨骼点附近的参数化模型进行离散刚性变换来转换 3D 点。然而，使用这种简单的变化可能导致关节处出现不自然的形变或缝隙。为了提升从观测空间到规范空间变换的泛化能力，本文基于最新高效的衔接模块 fast-snarf[5]，并在此基础上设计了一种创新的多层次网络结构，该结构旨在从空间中捕获多尺度的点特征，并将点特征有效地映射到 3 个不同的平面上。通过在这 3 个平面上进一步划分不同尺寸的特征窗口，模型能够提取出具有不同细节层次的纹理特征。该方法不仅丰富了特征表示，也有效减少了由衔接模块快速映射过程所引起的不自然形变等缺陷。最终，经过细致处理的特征数据被集成到经过设计的人体神经辐射场中，以此实现对三维人体模型的自然重建。

为了更好地学习到人体的高频特征，本研究总结了先前工作[6]的优缺点，设计将多分辨率哈希编码策略应用到神经网络辐射场的重建技术中。该策略的核心是用一个精简的多层感知机网络替代传统的神经辐射场模型，并辅以额外的编码参数训练，这些参数被存储于网格顶点，用于学习并编码人体的多级细节信息。这种方法允许在保持质量的情况下使用较小的网络，从而减少了浮点运算和内存访问操作的数量，加快了网络的训练速度。通过这一方法，能够更有效地利用计算资源，在提高神经网络在辐射场重建任务中的训练速度的同时保持了较高的重建质量。实验结果表明，本文方法在 PeopleSnapshot[7]数据集上取得了不错的成绩，并显著减少了神经网络所需的训练时间。

2 相关工作

2.1 三维人体重建

近年来，研究人员越来越关注如何重建出高质量 3D 人体外观和形状的问题[8-9]。传统的 3D 数字人体重建方法通过融合密集的相机阵列[10-11]或基于深度传感器[12-13]的观察结果实现，但昂贵的硬件要求限制了这些方法在非专业环境中的使用。随着深度学习技术的发展，以数据驱动[10-11,14-15]的方式重建人体引起了人们的广泛关注。有些方法尝试从单视图图像[16-17]、多视图图像[18-19]、RGB 视频[20-23]和 RGB-D 视频[24-26]中重建数字人体。基于 PIFU[27]和 PIFuHD[28]的方法在使用隐式表示重建三维人体表面方面取得了令人兴奋的进展。在处理具有复杂姿势和服装的人物时，这些方法取得了令人印象深刻的结果。然而，由于其昂贵的计算成本，生成的 3D 人体外观相对模糊，效果并不理想。

另一方面，一些研究[29-31]通过采用参数化人体模板网络模型 SMPL[4]实现了单目视频中高质量纹理的 3D 人体的重建，该方法通过将模板根据 2D 关节和轮廓进行变形来实现 3D 人体的重建，但在处理复杂的几何形状，如头发和衣服褶皱时，表现能力有限，难以模拟高保真细节和不同服装的拓扑。为了处理更加复杂的姿态输入和建模更复杂的纹理信息，一些方法[32-35]结合了隐式函数学习和参数化模型的方法将参数化人体模型与隐式表示结合，取得了更加鲁棒的结果。

2.2 神经场景表征

最近，隐式神经网络表示已经成为建模 3D 人体的强大工具[36]。使用隐式神经网络，可以表示具有任意拓扑结构的 3D 形状。许多研究展示了可以仅从单目视频[37-40]或稀疏人体视图[41-42]中重建 3D 人体的方法。Saito 等[14]引入的像素对齐隐式函数利用 MLP 确定给定三维位置的体占有值。PIFuHD[15]在此基础上，利用更高精度的特征和预测的法向信息，实现了更多几何细节的穿衣人体重建。然而，上述方法缺乏整体几何的限制和人体先验，可能会产生一些破碎的结构。基于 MLP 表达的神经辐射场在新视角合成方面取得了极大的成功，激发了三维人体重建的新研究方向。Neural body 团队[43]使用稀疏卷积对辐射体积进行建模。Lombardi 等使用相同网络回归 3D 人体，通过不同潜在代码去引导。而 Chen 等[44]通过引入显式的姿态引导变形将神经网络辐射场扩展到动态场景和人体运动中。然而，单一的神经网络辐射场很难隐式控制人体运动这种复杂的非刚性变形。一些方法将隐式神经网络和参数化人体模型 SMPL 结合[34,45-48]，以重建 3D 人类化身。尽管这些方法能够学习到 3D 人类化身，但在训练和渲染速度以及表示细节方面仍存在缺陷。

2.3 加速神经辐射场

为了加速 NeRF 的推理和训练速度[49-50]，许多研究[3-5,51-56]采用显式数据结构替代多层感知机，以加速静态神经网络辐射场。NSF 团队[33]使用了稀疏的体素八叉树结构，通过由粗到细的方式加速视图的重建。Plenoxels 团队[45]和 Planoctree 团队[48]利用球面谐波函数进行重建。DVGO 团队[46]利用体素网格实现了快速收敛。Instant-NGP 团队[5]和 Tensor4D 团队[57]进一步采用多分辨率哈希表替代体素，以记录高频细节并显著提高了训练速度。部分研究[58-59]通过占用网格中的空区域来提高渲染效率，从而进一步提高训练和推理速度。

3 方法描述

在给定的 T 张人体图像以及其相机位姿下，本文设计的 TriRF(Triplane Neural Radiance Fields)模型首先使用预训练的 SMPL 人体模型回归估计每一帧中人体图像的 SMPL 姿态参数 $\theta_1,\cdots,\theta_T$ 以及共享在图像之间的身体形状参数 β。通过 TriRF 模型的快速姿势变形模块(Fast Posture Deformation，FPD)建立姿势空间(Pose Space)和规范空间(Canonical Space)的三维位置的映射关系，将参数变化为规范空间下的表示形式。在此基础上，利用三平面多尺度分解模块(Tri-

plane Multiscale Decomposition，TMD）捕捉空间中不同尺度的人体结构及纹理特征信息，将不同尺度的特征信息拼接融合后，送入即时规范神经辐射场来重建更加准确和生动的人体表达。在训练期间，针对辐射场渲染所得的图像结果与真实图像，采用特定的损失函数 magic loss，以两个像素集合之间的相似性作为训练损失，以此促使渲染结果能够更为精准地逼近真实图像。图 1 展示了本文设计的 TriRF 模型。对于给定单目视频的每一帧，TriRF 模型展示了获取观测空间中光线上某点的密度 σ 和颜色值 c。图 1 中，快速姿势变形模块（Fast Posture Deformation，FPD）将观测空间中的点映射到规范空间中的对应位置，使得模型能够对不同姿势的人体进行统一处理；三平面多尺度分解模块（Triplane Multiscale Decomposition，TMD）对不同尺度的特征进行学习，并将这些特征连接成混合特征，以此增强模型的表达能力和泛化能力；即时规范神经辐射场（Instant Canonical Neural Rdiance Filed，ICNRF）利用由较小的多层感知机组成的高效渲染网络，基于混合特征来预测点的密度 σ 和颜色值 c，并通过特定的损失函数（magic loss）对网络进行优化。TriRF 模型能够捕获人体姿势的细微变化以及纹理信息，为数字人物建模和虚拟现实应用提供了更优秀的解决方案。

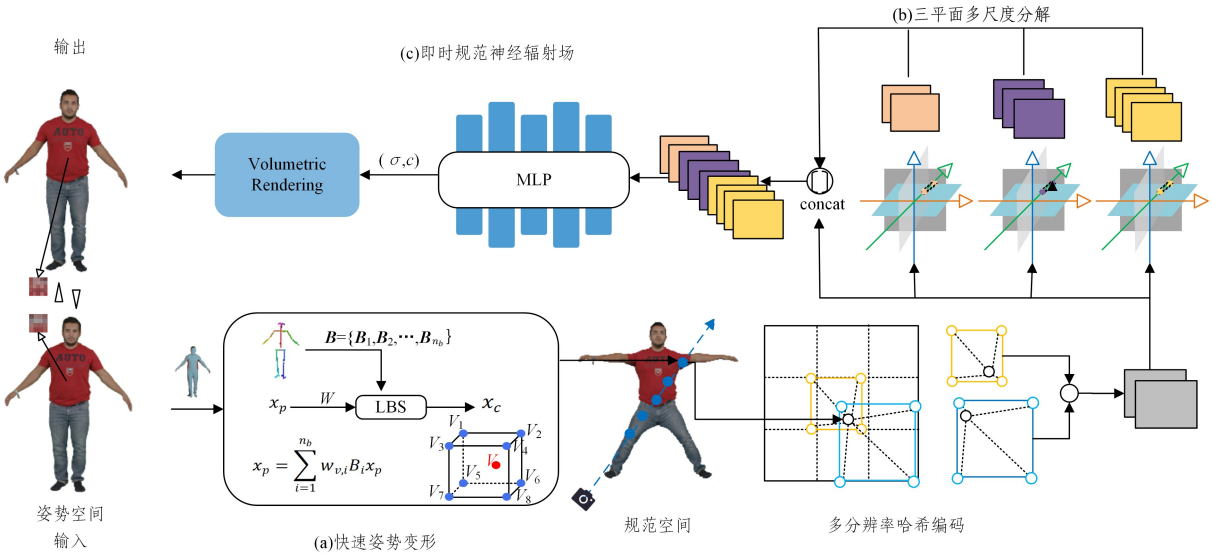


图 1　TriRF 模型结构

Fig. 1　Structure of TriRF model

3.1　前期工作

3.1.1　神经辐射场

本研究工作基于最近取得显著进展的神经网络辐射场模型。该模型利用密度场 σ 和与视图相关的颜色场 $c(x,d)$ 实现高质量的体渲染，从而在三维视觉领域开辟了新的研究方向。在给定射线 $r(\tau)=o+td$ 的情况下，颜色值的计算式为：

$$C(r)=\int_{T_n}^{T_f}(T_\tau)\sigma(r(\tau))c(r(\tau),d)\mathrm{d}\tau \tag{1}$$

$C(r)$ 是沿射线从近平面 τ_n 到远平面 τ_f 的积分得到，其中 $T(\tau)=\exp(-\int_{\tau_n}^{\tau}\sigma(r(s))\mathrm{d}s)$ 表示得到的累计透过率。

3.1.2　三投影分解

近期的一些工作[4,60-61]采用了三投影分解的策略来加速神经辐射场网络的训练和渲染过程。三投影分解是基于网格的三维数据结构，将 n 维张量 $\boldsymbol{V}_h$ 分别沿 x，y 和 z 轴投影到 3 个彼此正交的特征平面 $T=(t_{xy},t_{xz},t_{yz})$ 上，这些特征平面的分辨率为 $T_1\in\mathbb{R}^{R\times R\times n_f}$，其中 n_f 表示特征维数。因此，只需将查询点 $x\in\mathbb{R}^3$ 投影到 3 个特征平面 x 上，再通过双线性插值即可获取对应的特征向量（$\boldsymbol{F}_{xy}$，$\boldsymbol{F}_{xz}$，$\boldsymbol{F}_{yz}$）。这一策略不仅优化了数据的处理效率，还提升了特征提取的精确度。

3.2　即时规范神经辐射场

神经网络辐射场在场景渲染等领域具有显著优势，例如可高效生成高质量的三维场景视觉效果[62-63]。然而，在对人体重建过程中，它难以精确地模拟人体在不同姿态和形状下的复杂变化。鉴于此，本研究设计了一个隐式神经辐射场 F_c，旨在模拟人体在规范空间的形态和外观。该模型首先通过快速姿势变形模块，将姿势空间中的三维点 x_p 以及与之对应的人体姿势 p 的骨骼点变换矩阵 $\boldsymbol{B}$ 转换为规范空间中的对应点，其中骨骼变换矩阵 $\boldsymbol{B}$ 是由一系列变换矩阵 $\boldsymbol{B}_1,\cdots,\boldsymbol{B}_{n_b}$ 组成，它能够反映人体骨骼关节运动和位置变换的矩阵集合。接着，将规范空间的点 x_c 送入专门设计的基于人体的即时规范神经辐射场中，用于预测密度值 σ 和颜色值 c。通过该设计，能够实现对人体姿态和形状的精确捕捉。

$$F_c(T_{p2c}(x_p,\boldsymbol{B}))=(\sigma,c) \tag{2}$$

其中，$T_{p2c}(x_p,\boldsymbol{B})$ 的作用是将姿势空间中的三维点 x_p 转换为规范空间中的对应点 x_c。辐射场 F_c 接受规范空间中人体三维点 x_c，并通过多层感知机网络的学习与计算，根据输入的规范空间的点 x_c 预测出该点处的密度和颜色。然而，在实际操作中发现，设计的模型所需要的时间和计算成本极高。传统的神经辐射场方法通过大量的特征去表示空间信息，并且 NeRF 的网络结构主要由多层感知器（MLP）组成，而 MLP 在学习高频信息时存在一定的局限。为了降低计算成本，加快模型的训练速度，以及让网络模型能够更好地学习到高频信息，本文创新性地对传统神经网络辐射场进行修改，设计了一

种基于多分辨率哈希编码的网络，在不同分辨率下将空间点映射到哈希表，提取出高频信息并对辐射场进行参数化。通过采用不同分辨率的哈希表，使得模型能够在低分辨率层次获取全局信息，并在高分辨率层次来捕捉更精细的纹理细节。

$$F_c'(x_p)=\sum_{i=1}^{n}H_i(x_p) \tag{3}$$

其中，F_c 为辐射场，H 为可学习的多分辨率哈希编码函数，x_p 为规范空间中的空间点。具体来说：在第 i 个分辨率下，通过哈希表对空间点 x_p 进行映射，以获得高频信息来对辐射场 F_c 进行参数化，通过不同分辨率的哈希表来模拟多分辨率的体素网格，使网络既能分辨低分辨率信息，也能分辨高分辨率信息；接着，利用三线性插值获取输入的空间点的特征，并将不同层次级别的多分辨率特征串联后送入一个更小的 MLP 网络中。这样能够增强神经辐射场作为 3D 人类化身的表征，还降低了模型复杂度，实现了快速的训练和推理速度，提高了模型的性能。

3.3 快速姿势变形

在对具有复杂非刚性变形的人体外观与几何形状进行模拟时，推断姿势空间和规范空间之间的 3D 位置变换关系极具挑战性，其核心在于把姿势空间中的三维位置转换为规范空间中对应的位置，即将姿势空间中的三维位置 x_p 转换为规范空间中对应的位置 x_c。参数化人体模型 SMPL(Skinned Multi-Person Linear model)可用于表示及生成人体姿态和形状，它能对空间点在姿势空间和规范空间之间的映射起到显式引导作用。基于蒙皮的线性混合(LBS)算法虽能利用离散的骨骼关节点将姿态应用于三维几何网格上，但该算法仅能处理离散数据。传统方法直接将 LBS 算法应用到 SMPL 人体模型中，但此过程效率不高。本研究创新性地引入参数化的 LBS 权重场来实现快速变形[3,64]，其思想是，在规范空间下定义一个 LBS 权重场来进行建模铰接，将从规范空间到骨骼权重空间的映射函数参数化为基于低分辨率体素网格 ω_v 的 LBS 权重场。具体而言，每个体素网格 x_v 的值被确定为其在 SMPL 模型上最近定点的权重值。对于非网格对齐点的蒙皮权重值，则采用三线性插值的方式在体素网格上相邻的 8 个网格点查询其蒙皮权重值，以此来定义和使用权重场，有效避免了传统基于 MLP 的构建方式[52]在每个根查找迭代中消耗大量计算资源的问题。

$$x_c=\sum_{i=1}^{n_b}\omega_{v_i}\cdot B_i\cdot x_p \tag{4}$$

其中，$v_i(i=0,1,\cdots,7)$ 是点 x 的 8 个相邻网格点。这样，任何规范点所需的映射在少量网格点上即可得到，从而实现了快速反向蒙皮根查找，无需遍历所有查询点，如 Fast Posture Deformation 模块所示(见图 1)，该策略能够非常快速地建立姿势空间和规范空间之间的映射关系。

3.4 三平面多尺度分解

快速建立映射关系虽然提高了效率，但插值会导致精度的损失，且低分辨率体素网格难以完全准确捕捉人体连续变化的细节。为了更好地感知人体纹理细节，本研究基于三平面[65]设计了一种高效的分层特征提取模块。该模块将投影后的特征信息划分成具有不同窗口大小的特征块，并利用单层自注意力网络(Single-Layer Self-Attention，SLWA)来学习不同尺度的特征。SLWA 能够自动关注特征信息中的重要部分，并重新组织信息以实现更有效的信息传递。

$$T=concat[T',SLWA^j(T')] \tag{5}$$

$$SLWA^j(T')=\{T'^j_{xy},T'^j_{xz},T'^j_{yz}\} \tag{6}$$

其中，T' 为投影到三平面的原始特征，j 表示将要进行多少个不同尺度的窗口划分，在本文中 j 取 3，即进行 3 种不同尺度的窗口自注意力操作，这样能够在计算效率和特征表达丰富度之间取得较好的平衡，有效地提取出不同尺度下的关键特征信息。$concat[\cdot]$ 是通道拼接操作，用于将 T' 和单层窗口注意力操作 $SLWA(T')$ 得到的特征在通道维度上进行拼接。通过这样的方式，获得了不同尺度自适应特征的三平面表示，从而使模型能够更丰富、更准确地表达人体几何纹理细节。具体来说，针对变形至规范空间中的三维点 x_i，模型首先提取其特征并将其投影到 3 个二维平面 $T'\{T'_{xy},T'_{xz},T'_{yz}\}$ 上。然后，通过单层自注意力网络(SLWA)，将其划分成不同窗口大小的特征块。第 j 次单层窗口注意力(SLWA)产生的特征表示为 $SLWA^j(T')$。之后，将所有 SLWA 模块的输出与原始特征 T' 相连接，并送入辐射场 F_c'。

3.5 渲染方式

本文使用与神经网络辐射场[1]相同的体积渲染技术渲染出神经辐射场 F'_c 的新视角视图。给定一个像素，投射一条光线 $r=o+td$，其中 o 是相机中心，d 是光线方向。在实践中，沿着相应的相机光线 r 在近边界和远边界之间对 N 个点 $\sum_{i=1}^{N}x_i$ 进行采样，并从辐射场 F_c' 中累计每个点的颜色和密度，得到最终的像素颜色 C：

$$C(r)=\sum_{i=1}^{N}T_i(1-\exp(\sigma(x_i)\delta_i))c(x_i) \tag{7}$$

$$T_i=\exp(-\sum_{j=1}^{i-1}\sigma(x_j)\delta_j) \tag{8}$$

其中，$\sigma(x_i)$ 表示点 x_i 处的密度，$c(x_i)$ 表示点 x_i 处的颜色，$\delta_j=\|x_{i+1}-x_i\|_2$ 表示相邻采样点之间的距离。

3.6 损失定义

本文以像素为单位计算损失，具体与神经网络辐射场的 MSE loss 计算方式类似，即将通过最小化相机光线渲染得到的图像像素值 C 和真实像素值 C_{gt} 进行比较。对每个像素 i，计算预测颜色 C_i 和真实颜色 C_{gt_i} 之间的平方差，并求平均值。通过最小化这个损失，模型的颜色预测可以更好地接近真实颜色，从而提高图像渲染质量。

$$L_{mse}(c,c_{gt})=\frac{1}{N}\sum_{i=1}^{N}\|C_i-C_{gt_i}\|^2 \tag{9}$$

然而，这种均方误差(MSE)损失函数只能捕捉相邻像素的局部信息，无法捕捉到更远像素间的结构信息。在真实场景中，即使在远处，也可能包含重要的结构信息[66]。本研究针对性地设计了一种名为“magic loss”的损失函数，它以两个像素集合之间的相似性作为训练损失。在训练过程中，每个小批量数据的像素被随机组合成一个“随机补丁”(Stochastic Patch)，然后利用基于卷积核的相似性度量来处理这些随机补丁。这些像素集合一般包含数千个像素，它们共同贡献

全局、相互关联的结构信息。通过最小化这个损失函数，使模型能够更好地理解人体特征的全局结构完整性和空间关联性，从而提高重建的准确性。

$$L_{\text{magic}}(\theta)=\frac{1}{\|P\|}\sum_{p\in P}M(\theta,L_{\text{mse}}(p(\hat{c}),p(\widehat{c_{\text{gt}}})) \tag{10}$$

其中，$p(\hat{c})$表示从渲染后的图像中提取的像素集合，$p(\widehat{c_{\text{gt}}})$表示对应的真实图像中的像素集合，$\widehat{c_{\text{gt}}}=\{\widehat{c_{\text{gt}}}(r)\mid r\in p\}$，$c=\{c(r)\mid r\in p\}$，$M$函数基于卷积核的相似性度量来处理像素集合，$L_{\text{MSE}}(p(\hat{c}),p(\widehat{c_{\text{gt}}}))$是基于MSE的像素级损失函数，$\theta$表示模型参数，$P$表示像素集合，$\|P\|$表示像素集合$P$的大小。通过最小化损失函数，使模型能够更好地理解人体特征的全局结构完整性和空间关联性，从而提高模型重建的效果。

4　实验论证

本章在单目人体视频上全面评估了本文方法的准确性，表1列出了本文方法与最近提出的一些方法的详尽比较。图2展示了本文方法与Anim-NeRF，NeuralBody和Human-NeRF方法的实验比较结果。图中每列展示了不同方法在相应数据上的效果表现。与HumanNeRF相比，本文方法能够有效减少人体重建中的伪影和毛刺现象，显著提高重建效果。与Anim-NeRF相比，本文方法不仅大幅缩短了训练时间，还在服装纹理等细节重建上表现更优。与NeuralBody相比，本文方法同样在显著缩短训练时间的同时，保持了高质量的重建效果。总体而言，本文方法在重建效果与训练效率上均表现出较大的优势。

表1　在PeopleSnapshot[7]数据集上不同方法在各项指标上的比较

Table 1　Comparison of different methods on various metrics using the PeopleSnapshot dataset[[7]]

Method	GPU	male-3-casual			male-4-casual			female-3-casual			female-4-casual		
	Time	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeuralBody[44]	14 h	24.94	0.9428	0.0326	24.71	0.9649	0.04230	23.87	0.9504	0.0346	24.37	0.9451	0.0382
HumanNeRF[46]	20 h	26.90	0.9605	0.0181	25.50	0.9397	0.03573	24.46	0.9516	0.0269	27.07	0.9615	0.0152
Anim-NeRF[67]	13 h	29.37	0.9703	**0.0168**	**28.37**	0.9605	**0.02678**	**28.91**	0.9743	**0.0215**	28.90	0.9678	0.0174
Our method	**5 min**	**29.63**	**0.9738**	0.0172	27.82	**0.9710**	0.02910	28.14	**0.9754**	0.0243	**29.23**	**0.9697**	**0.0153**
Anim-NeRF[67]	5 min	23.17	0.9266	0.0784	22.30	0.9235	0.09110	22.37	0.9311	0.0784	23.18	0.9292	0.0687
InstantAvatar[68]	5 min	29.53	0.9716	**0.0155**	27.67	0.9626	0.03070	27.66	0.9709	**0.0210**	29.11	0.9683	0.0167
Our method	5 min	**29.63**	**0.9738**	0.0172	**27.82**	**0.9710**	**0.02910**	**28.14**	**0.9754**	0.0243	**29.23**	**0.9697**	**0.0153**

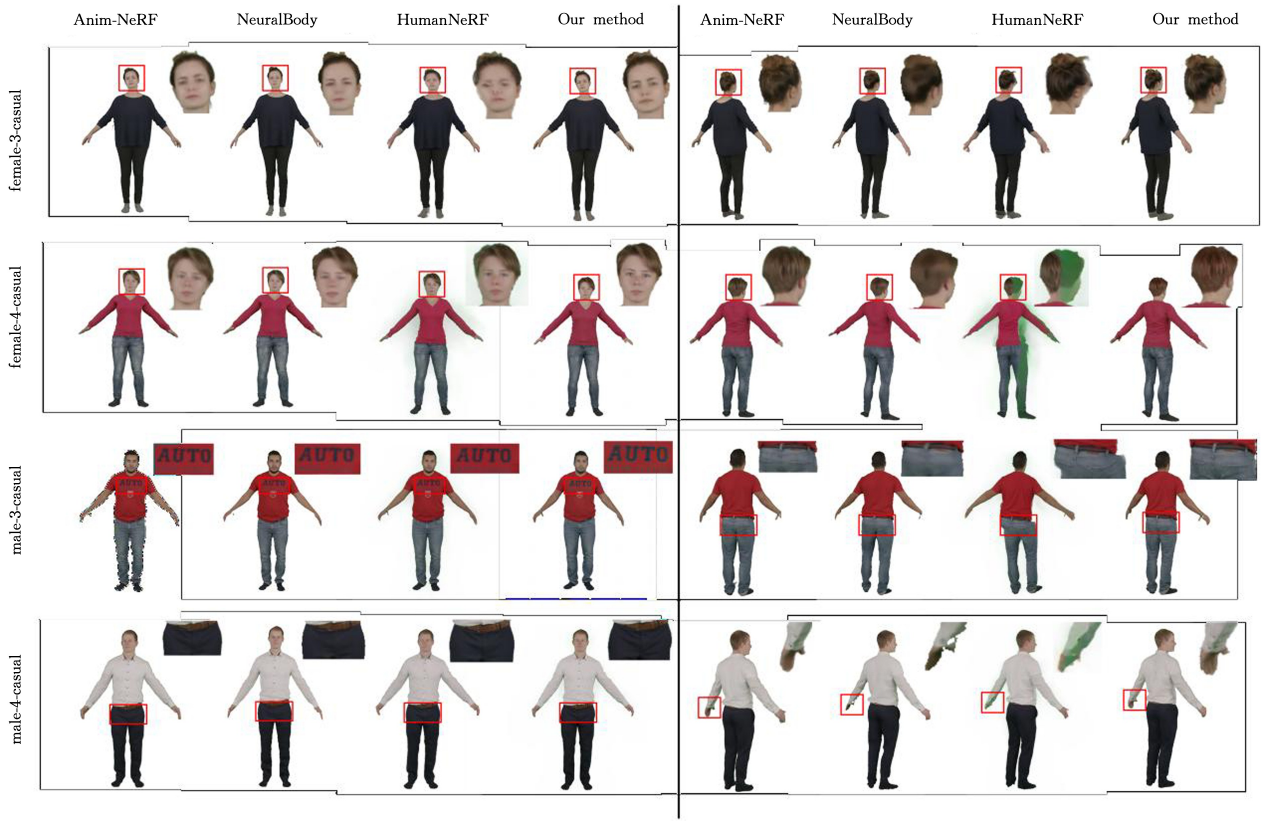
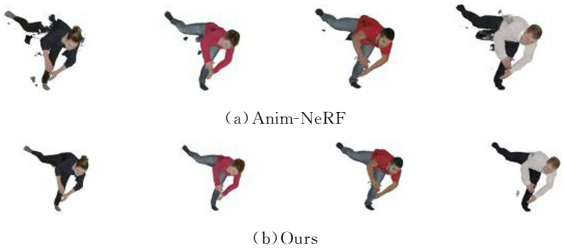


图2　在PeopleSnapshot[7]数据集上的定性结果

Fig. 2　Qualitative results on the PeopleSnapshot dataset

针对新颖的视图合成，采用以下指标来评估新视图合成和新姿态合成方法：峰值信噪比(PSNR)、结构相似指数(SSIM)[47]以及学习感知图像斑块相似度(LPIPS)[69]。由于真实场景数据集缺乏相应的地面真值几何信息，目前本文仅提供了定性结果。测试结果验证了本文方法相对于近期提出的一些方法的优越性和创新性。图3展示了本文方法和

Anim-NeRF[64]在新姿态合成任务上的比较，本文方法的新姿态的重建效果更好，在与训练姿态有很大不同的新姿态上，具有更好的泛化能力。此外，本文系统地消融了所提出的TriRF模型的每个组成部分，显示了它们在更好的渲染质量方面的有效性。



(a) Anim-NeRF

(b) Ours

图3 Anim-NeRF[64]和本文方法在新姿态合成任务上的比较

Fig. 3 Comparison of Anim-NeRF and the proposed method on the task of novel pose synthesis

4.1 评估数据集

本文对PeopleSnapshot[7]数据集中的4个序列进行了实验。该数据集[16]包含在现实场景中拍摄的视频，描绘了24个人在固定相机前以A形姿势旋转。除了提供人体的遮罩图像外，该数据集还标注了24个关键点。为确保公平性，严格遵循Anim-NeRF[64]中定义的评估协议，包括利用文献[54]提供的姿势，并采用与Anim-NeRF中所提到的针对该数据集相同的优化的方法来处理该数据集中姿势参数与图像之间的不对齐问题。本文遵循InstantAvatar[65]中的数据分割方法，并在新姿态合成方面将本文方法与InstantAvatar进行了比较。此外，将这些优化后的姿态参数纳入本文方法的训练中，以确保公平比较。

本文采用的基线方法如下。

1) HumanNeRF[46]：该方法通过单目视频输入实现人体任意视角重建，可在任何暂停时刻从任何角度渲染360度人体姿势。其算法核心在于优化T形姿势人体表示和运动场，以实现规范空间与视频帧的映射。相较于该方法，本模型在训练时间上具有明显优势，本文方法在一张RTX3090上仅需十几分钟，而该方法需要20小时。

2) Anim-NeRF[64]：该方法将静态神经辐射场扩展到人体运动的动态场景，模拟规范空间中人体的形状和外观。给定一个视频序列，它首先生成与视频帧的姿势相对应的SMPL人体。通过利用SMPL模型的姿势引导采样点到NeRF中的规范空间，然后使用标准的NeRF流程进行训练。本文方法在训练时间和计算成本上优于该方法。本文方法在一张RTX 3090上，十几分钟即可取得显著成果，而该方法需要两张RTX 3090且耗时13小时。在减少相邻身体部位伪影方面，本文方法表现也更为出色。

3) NeuralBody[44]：该方法对给定视频的每一帧使用相同的人体模板顶点集，即定义一组潜在编码。它通过这组潜在编码为不同的帧生成场景，并将在不同帧中观察到的信息与一组潜在编码相关联。因此，它将单目人体视频的所有帧的信息整合到这组潜在编码中。

4) Instantavatar[65]：该方法利用基于Instant-NGP的神经辐射场来模拟人体的形状和外观。此外，它利用占用网格来过滤网格中的空白空间，以加速网络训练。在短时间训练下，本文方法的图像重建结果在细节上比该方法的图像重建结果更清晰，人体表征细节更丰富。

4.2 新视角合成

本文方法在训练时间和计算成本上展现出突出优势。从硬件资源利用角度来看，本文方法仅需一张RTX 3090，而部分基线方法需要两张。从时间成本来看，本文方法在十几分钟内即可取得显著成果，相较于其他方法，大大节省了时间。

在PeopleSnapshot数据集中，本文方法在多个场景中指标均有提升。在PSNR(峰值信噪比)和SSIM(结构相似性指数)上有所提升，面对训练集中未出现的姿势也能生成清晰合理的结果。对于姿态参数与图像不完美对齐的复杂情况，采用定性评价。对比发现，模型在减少相邻身体部位伪影方面优于基于多层感知器的Anim-NeRF；与基于Instant-NGP架构的InstantAvatar相比，本模型在合成新颖姿态图像时，图像更清晰、细节更丰富。从实际应用场景来看，本模型在图像清晰度和细节丰富度上的提升，对于虚拟现实、影视制作等领域具有积极意义。在虚拟现实中，更清晰和细节丰富的人体图像能够增强用户的沉浸感；在影视制作中，则可以提高后期制作的效率和质量。

综上所述，本文方法在新颖视图合成方面取得了一定的进步，无论是在计算效率、模型性能还是实际应用方面，都具有独特的优势，为姿态合成领域提供了新的思路。

4.3 对比实验

为了验证本文方法的有效性及优越性，在PeopleSnapshot数据集上进行了广泛的消融实验，通过添加不同的设计组件来深入分析各组件对模型性能的影响，实验结果如表2和图4所示。图4展示了完整模型(Full Model)、不含快速姿势变形(Fast Posture Deformation，FPD)和三平面多尺度分解(Triplane Multiscale Decomposition，TMD)组件，以及不含设计的魔法损失L_{magic}的模型之间的性能差异。

表2 在PeopleSnapshot上的消融研究

Table 2 Ablation study on PeopleSnapshot

Metric	PSNR ↑	SSIM ↑	LPIPS ↓
Full Model	28.69	0.9710	0.0275
W/O FPD	28.56	0.9641	0.0308
W/O TMD	28.52	0.9628	0.0299
W/O TMD, FPD	28.31	0.9600	0.0305
W/O L_{magic}	28.45	0.9672	0.0297

基准模型通常采用基本的离散刚性变换的方法建立姿势空间和规范空间之间的映射。在这种传统方法中，由于变换方式的局限性，容易在关节处出现不自然变形或者不连续的现象，缺乏对关节复杂结构的适应性。为了缓解这一问题，引入了“+FPD(Fast Posture Deformation)”设计。该设计基于体素网格实现人体参数在姿势空间和规范空间之间的快速变换映射，有效增强空间转换的通用性，不仅避免了复杂计算，显著缩短了网络训练时间，还有效地解决人体在姿势变换过程中关节处的过度不连续问题。另一方面，在基准模型的基础上进一步引入了“+TMD(Triplane Multiscale Decomposition)”设计。该设计通过减少模型参数数量和增强特征表示能力，使模型能够更好地捕获输入数据中的局部结构和空间

关系。同时，针对损失函数展开了进一步的消融对比，通过将含有损失函数L_{magic}的完整模型（Full Model）与去除该损失函数的模型（w/o L_{magic}）进行对比，可以观察到，损失函数L_{magic}有助于模型更加准确地还原人体特征的纹理细节，对于衣服的颜色和褶皱细节（图 4 中第一和第二列）有着更保真的效果，有效增强了模型对全局特征关系的理解和重建能力。鉴于人体不同部位之间存在复杂的几何关系，该设计的引入对提升模型性能起到了积极作用。将“＋FPD”和“＋TMD”两种设计相结合，使得本文模型能够学习到三维人体的多层次特征，更精确地捕捉全局范围内人体姿势的细微变化，并呈现出更丰富的纹理细节。这为数字人体建模和虚拟现实应用提供了更为优秀的解决方案。

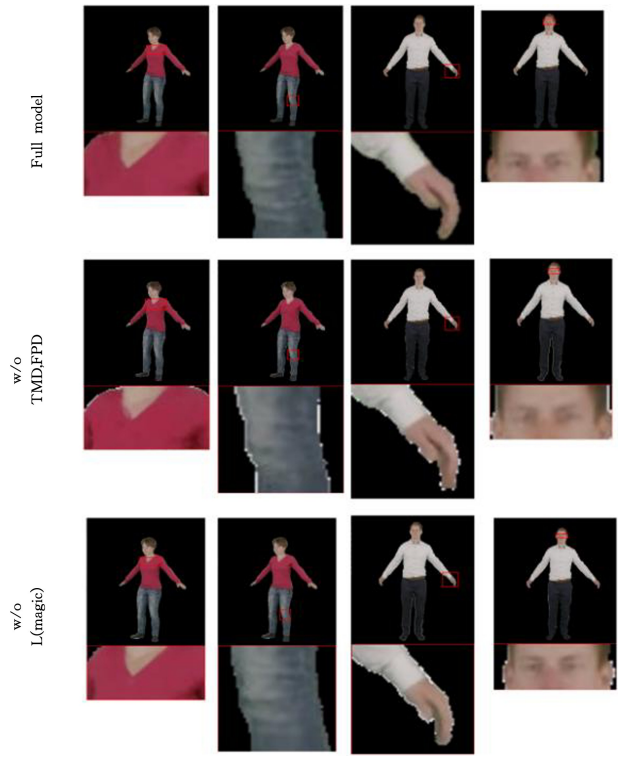


图 4　在 PeopleSnapshot 上进行的消融实验的对比结果

Fig. 4　Comparative results of ablation experiments conducted on PeopleSnapshot

4.4　方法局限性讨论

尽管本文提出了一种高效的单目视频生成 3D 数字人的方法，但该方法仍存在若干局限性，限制了其在实际应用中的效果与普适性。首先，单目视频输入缺乏深度信息，导致身体细节重建的精度不足，尤其在复杂动态动作中，可能出现不自然或失真的表现。其次，在处理快速变化的动作时，生成的人体模型可能无法保持稳定的表现，如动作流畅性较差。此外，模型在一些细节，如人物的精细纹理、复杂的服装材质等方面，可能难以准确地重建。最后，该方法对训练数据的依赖较大，若训练集缺乏足够的多样性，可能影响模型在不同场景下的适应性和表现。

结束语　本文提出了一种基于神经网络辐射场的三平面多尺度分解网络，旨在从单目 RGB 视频中重建高质量的 3D 人体模型。集成神经辐射场（NeRF）与参数化人体模型 SMPL，借助提出的三平面多尺度分解模块，有效地学习了更多的人体特征信息。同时，多分辨率哈希编码的引入不仅提升了训练效率与模型收敛性，还确保了高质量重建效果的达成。在此过程中，本文设计了快速姿势变形模块，它能够快速地建立姿势空间和规范空间之间的映射，但在提高效率的同时也损失了人体细节。为了弥补这一不足，本文进一步设计了三平面多尺度分解模块来学习丰富的特征信息。实验结果充分证明了本文方法在实现高质量 3D 人体重建和新姿态合成方面的有效性。综上所述，本研究在单目 RGB 视频的 3D 人体建模领域取得了切实的进展，提供了极具价值的见解。本文的创新之处在于融合多种先进技术，从不同角度对人体建模进行优化，为后续研究开辟了新的方向，相信能对相关领域的研究探索起到有力的推动作用。

参 考 文 献

[1] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. Representing scenes as neural radiance fields for view synthesis [J]. Communications of the ACM, 2021, 65(1): 99-106.

[2] HE G X, ZHU B, XIE B, et al. Progress in Novel View Synthesis Using Neural Radiance Fields[J]. Laser & Optoelectronics Progress, 2024, 61(12): 71-83.

[3] LI J Y, CHENG L C, HE J X, et al. Research Status and Prospects of Neural Radiance Fields [J]. Journal of Computer-Aided Design & Computer Graphics, 2024, 36(7): 995-1013.

[4] LOPER M, MAHMOOD N, ROMERO J, et al. Skinned multi-person linear model [C]// Seminal Graphics Papers: Pushing the Boundaries, Volume 2. 2023: 851-866.

[5] CHEN X, JIANG T, SONG J, et al. Fast-snarf: A fast deformer for articulated neural fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(10): 11796-11809.

[6] MULLER T, EVANS A, SCHIED C, et al. Instant neural graphics primitives with a multiresolution hash encoding [J]. ACM Transactions on Graphics, 2022, 41(4): 1-15.

[7] ALLDIECK T, MAGNOR M, XU W, et al. Detailed human avatars from monocular video [C]// 2018 International Conference on 3D Vision(3DV). IEEE, 2018: 98-109.

[8] HAN K, XU J. Research on 3D Scene Rendering Technology-Neural Radiance Field[J]. Application Research of Computers, 2024, 41(8): 2252-2260.

[9] WANG Z R, CHANG Y, LU P, et al. A Review of Acceleration Algorithms for Neural Radiance Fields[J]. Journal of Graphics, 2024, 45(1): 1-13.

[10] COLLET A, CHUANG M, SWEENEY P, et al. High-quality streamable free-viewpoint video [J]. ACM Transactions on Graphics, 2015, 34(4): 1-13.

[11] DOU M, KHAMIS S, DEGTYAREV Y, et al. Fusion4D: Real-time performance capture of challenging scenes [J]. ACM Transactions on Graphics, 2016, 35(4): 1-13.

[12] GUO K, LINCOLN P, DAVIDSON P, et al. The Relightables: Volumetric performance capture of humans with realistic relighting [J]. ACM Transactions on Graphics, 2019, 38(6): 1-19.

[13] MATUSIK W, BUEHLER C, RASKAR R, et al. Image-based visual hulls [C]// Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. 2000: 369-374.

[14] SAITO S, HUANG Z, NATSUME R, et al. Pifu: Pixel-aligned

implicit function for high-resolution clothed human digitization [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:2304-2314.

[15] SAITO S, SIMON T, SARAGIH J, et al. PifuHD: Multi-level pixel-aligned implicit function for high-resolution 3D human digitization [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:84-93.

[16] LAZOVA V, INSAFUTDINOV E, PONS-MOLL G. 360-degree textures of people in clothing from a single image [C]// 2019 International Conference on 3D Vision (3DV). IEEE, 2019: 643-653.

[17] ALLDIECK T, PONS-MOLL G, THEOBAL T, et al. Tex2Shape: Detailed full human body geometry from a single image [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:2293-2303.

[18] ZHENG Z, YU T, LIU Y, et al. Pamir: Parametric model-conditioned implicit representation for image-based human reconstruction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(6):3170-3184.

[19] LIU X N, CHEN C Y, HU X J, et al. Virtual View-point Image Synthesis of Neural Radiance Field with Depth Information Supervision [J]. Journal of Image and Graphics, 2024, 29(7):2035-2045.

[20] PESAVENTO M, XU Y, SARAFIANOS N, et al. ANIM: accurate neural implicit model for human reconstruction from a single RGB-D image [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024:5448-5458.

[21] ALLDIECK T, MAGNOR M, XU W, et al. Video-based reconstruction of 3D people models [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8387-8397.

[22] ALLDIECK T, MAGNOR M, BHATNAGAR BL, et al. Learning to reconstruct people in clothing from a single RGB camera [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:1175-1186.

[23] SONG C, WANDT B, RHODIN H. Pose modulated avatars from video[J]. arXiv:2308.11951, 2023.

[24] ALLDIECK T, MAGNOR M, BHATNAGAR B L, et al. Learning to reconstruct people in clothing from a single RGB camera [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:1175-1186.

[25] LING S, NGUYEN K, ROUX-LANGLOIS A, et al. A lattice-based group signature scheme with verifier-local revocation [J]. Theoretical Computer Science, 2018, 730(19):1-20.

[26] VAMBOL A, KHARCHENKO V, POTII O, et al. McEliece and Niederreiter Cryptosystems Analysis in the Context of Post-Quantum Network Security [C]// International Conference on Mathematics & Computers in Sciences & in Industry. IEEE Computer Society, 2017:134-137.

[27] SAITO S, HUANG Z, NATSUME R, et al. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:2304-2314.

[28] SAITO S, SIMON T, SARAGIH J, et al. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:84-93.

[29] DONG Z, CHEN X, YANG J, et al. Ag3d: Learning to generate 3d avatars from 2d image collections [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 14916-14927.

[30] ZHI T, LASSNER C, TUNG T, et al. Texmesh: Reconstructing detailed human texture and geometry from rgb-d video [C]// Computer Vision-ECCV 2020: 16th European Conference. Springer, 2020:492-509.

[31] ZHAO X, WANG L, SUN J, et al. Havatar: High-fidelity head avatar via facial model conditioned neural radiance field [J]. ACM Transactions on Graphics, 2023, 43(1):1-16.

[32] XIANG D, PRADA F, WU C, et al. Monoclothcap: Towards temporally coherent clothing capture from monocular rgb video [C]// 2020 International Conference on 3D Vision (3DV). IEEE, 2020:322-332.

[33] HABERMANN M, XU W, ZOLLHOEFER M, et al. Livecap: Real-time human performance capture from monocular video [J]. ACM Transactions On Graphics, 2019, 38(2):1-17.

[34] HABERMANN M, XU W, ZOLLHOFER M, et al. Deepcap: Monocular human performance capture using weak supervision [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:5052-5063.

[35] ZHANG H, FENG Y, KULITS P, et al. Text-guided generation and editing of compositional 3D avatars[J]. arXiv:2309.07125, 2023.

[36] SUN C, QIU J, WU L N, et al. Dynamic human body neural radiance field reconstruction based on monocular vision[J]. Acta Optica Sinica, 2024, 44(19):256-266.

[37] PENG S, DONG J, WANG Q, et al. Animatable neural radiance fields for modeling dynamic human bodies[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:14314-14323.

[38] GUO C, CHEN X, SONG J, et al. Human performance capture from monocular video in the wild[C]// 2021 International Conference on 3D Vision(3DV). IEEE, 2021:889-898.

[39] XIU Y, YANG J, TZIONAS D, et al. Icon: Implicit clothed humans obtained from normals[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022:13286-13296.

[40] XIU Y, YANG J, CAO X, et al. Econ: Explicit clothed humans optimized via normal integration[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023:512-523.

[41] WANG S, SCHWARZ K, GEIGER A, et al. Arah: Animatable volume rendering of articulated human SDFs [C]// European Conference on Computer Vision. Springer, 2022:1-19.

[42] JIANG B, HONG Y, BAO H, et al. Selfrecon: Self-reconstruction your digital avatar from monocular video[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:5605-5615.

[43] PENG S, ZHANG Y, XU Y, et al. Neural body: Implicit neural representations with structured latent codes for novel view syn-

thesis of dynamic humans[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 9054-9063.

[44] CHEN M, ZHANG J, XU X, et al. Geometry-guided progressive nerf for generalizable and efficient neural human rendering[C]// European Conference on Computer Vision. Cham: Springer, 2022: 222-239.

[45] PENG S, DONG J, WANG Q, et al. Animatable neural radiance fields for modeling dynamic human bodies[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 14314-14323.

[46] WENG C Y, CURLESS B, SRINIVASAN P P, et al. Human-NeRF: Free-viewpoint rendering of moving people from monocular video[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and pattern Recognition. 2022: 16210-16220.

[47] XU H, ALLDIECK T, SMINCHISESCU C. H-NeRF: Neural radiance fields for rendering and temporal reconstruction of humans in motion[J]. Advances in Neural Information Processing Systems, 2021, 34: 14955-14966.

[48] WANG Z, WU S, XIE W, et al. NeRF-: Neural radiance fields without known camera parameters[J]. arXiv: 2102. 07064, 2021.

[49] XIAO Y L, DENG Y Q, CHEN Z G. Accelerating Method of Neural Radiance Fields for Dynamic 3D Human Reconstruction [J/OL]. https://doi. org/10. 19678/j. issn. 1000-3428. 0069317.

[50] JING W P, WANG Y F, LI C. NeRF 3D Reconstruction Method Based on Cone Tracing and Network Decomposition[J]. Computer Engineering, 2024, 50(10): 334-341.

[51] HU S, HONG F, PAN L, et al. Sherf: Generalizable human NeRF from a single image[J]. arXiv: 2303. 12791, 2023.

[52] GAFNI G, THIES J, ZOLLHOFER M, et al. Dynamic neural radiance fields for monocular 4D facial avatar reconstruction[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 8649-8658.

[53] SU S Y, YU F, ZOLLHOEFER M, et al. A-NeRF: Surface-free human 3D pose refinement via neural rendering[J]. arXiv: 2102. 06199, 2021.

[54] SUN C, SUN M, CHEN H T. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5459-5469.

[55] TAKIKAWA T, LITALIEN J, YIN K, et al. Neural geometric level of detail: Real-time rendering with implicit 3d shapes[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 11358-11367.

[56] YU A, LI R, TANCIK M, et al. Plenoctrees for real-time rendering of neural radiance fields[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5752-5761.

[57] SHAO R, ZHENG Z, TU H, et al. Tensor4d: Efficient neural 4d decomposition for high-fidelity dynamic reconstruction and rendering[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 16632-16642.

[58] MARTIN-BRUALLA R, RADWAN N, SAJJADI M S, et al. Nerf in the wild: Neural radiance fields for unconstrained photo collections[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 7210-7219.

[59] PUMAROLA A, CORONA E, PONS-MOLL G, et al. D-nerf: Neural radiance fields for dynamic scenes[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 10318-10327.

[60] CHAN E R, LIN C Z, CHAN M A, et al. Efficient geometry-aware 3d generative adversarial networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 16123-16133.

[61] ZHANG J W, ZHANG H X, LI S H, et al. 3D Reconstruction of Human Head Based on TE-NeuS[J]. Software Engineering, 2024, 27(7): 56-60.

[62] WU S P, MA J S, SHE J F. An Implicit Representation-Based Method for Instant Real-Scene 3D Reconstruction and Neural Rendering[J]. Science of Surveying and Mapping, 2024, 49(4): 147-158.

[63] CHEN Q, QIN Z B, CAI X Y, et al. Dynamic 3D reconstruction of soft tissue with neural radiation field for robotic surgery simulator[J]. Acta Optica Sinica, 2024, 44(7): 279-291.

[64] CHEN X, ZHENG Y, BLACK M J, et al. Snarf: Differentiable forward skinning for animating non-rigid neural implicit shapes [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 11594-11604.

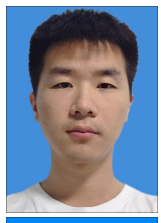
[65] FAN T, YANG H, YIN W, et al. Multi-scale view synthesis based on neural radiance fields[J]. Journal of Graphics, 2023, 44(6): 1140-1148.

[66] XIE Z, YANG X, YANG Y, et al. S3IM: Stochastic Structural SIMilarity and Its Unreasonable Effectiveness for Neural Fields [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 18024-18034.

[67] CHEN J, ZHANG Y, KANG D, et al. Animatable neural radiance fields from monocular rgb videos[J]. arXiv: 2106. 13629, 2021.

[68] JIANG T, CHEN X, SONG J, et al. Instantavatar: Learning avatars from monocular video in 60 seconds[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 16922-16932.

[69] TIWARI G, SARAFIANOS N, TUNG T, et al. Neural-gif: Neural generalized implicit functions for animating people in clothing [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 11708-11718.



WANG Yang, born in 1998, postgraduate. His main research interests include neural radiance fields and 3D human body reconstruction.

WANG Guodong, born in 1980, Ph. D, professor, is a member of CCF (No. 16234M). His main research interests include computer graphics and artificial intelligence.

（责任编辑：何杨）