计算机科学
COMPUTER SCIENCE

基于神经辐射场的即时高保真人脸生成算法

盛筱萌，赵俊莉，王国栋，王洋

引用本文

盛筱萌，赵俊莉，王国栋，王洋. 基于神经辐射场的即时高保真人脸生成算法[J]. 计算机科学，2025，52(10)：159-167.

SHENG Xiaomeng, ZHAO Junli, WANG Guodong, WANG Yang. Immediate Generation Algorithm of High-fidelity Head Avatars Based on NeRF [J]. Computer Science, 2025, 52(10): 159-167.

相似文章推荐（请使用火狐或 IE 浏览器查看文章）

Similar articles recommended (Please use Firefox or IE to view the article)

基于双流深度学习的Dockerfile安全误配置检测方法

Dual-stream Feature Fusion Approach for Dockerfile Security Misconfiguration Detection

计算机科学, 2025, 52(10): 395-403. https://doi.org/10.11896/jsjkx.241000014

基于DS理论的多模态信息抽取方法

Multimodal Information Extraction Fusion Method Based on Dempster-Shafer Theory

计算机科学, 2025, 52(10): 208-216. https://doi.org/10.11896/jsjkx.240200081

基于Retinex理论的低照度图像自适应增强算法

Low Light Image Adaptive Enhancement Algorithm Based on Retinex Theory

计算机科学, 2025, 52(10): 168-175. https://doi.org/10.11896/jsjkx.240800057

基于雷达和视觉融合的多模态空中手写体识别

Multimodal Air-writing Gesture Recognition Based on Radar-Vision Fusion

计算机科学, 2025, 52(9): 259-268. https://doi.org/10.11896/jsjkx.240400143

数据分类分级技术研究综述

Survey of Data Classification and Grading Studies

计算机科学, 2025, 52(9): 195-211. https://doi.org/10.11896/jsjkx.240800149

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.241000066

基于神经辐射场的即时高保真人脸生成算法

盛筱萌　赵俊莉　王国栋　王　洋
青岛大学计算机科学技术学院　山东 青岛 266071
(2018204306@qdu.edu.cn)

摘　要　针对数字娱乐、虚拟现实和元宇宙等领域对快速生成高精度数字人需求的日益增长问题,提出了一种基于单目 RGB 视频来快速生成高精度人脸模型的新方法,同时构建了一个专注于面部和颈部区域精确建模的新框架。具体来说,该框架将神经图元融入面部和颈部的参数化模型,同时采用 Head-And-neCK(HACK)模型替代常用的 Face Latent Animated Mesh Estimator(FLAME)模型,从而显著提升了 3D 面部重建的精度和效率。此外,针对性设计了一种实时自适应神经辐射场,有效加快了训练和重建过程。通过引入多分辨率哈希网格,并在变形空间内使用最近三角形搜索计算变形梯度,该方法能够在几分钟内快速重建高保真面部和颈部模型。通过广泛的定量和定性实验结果表明,相较于现有的先进方法,所提模型在渲染质量和训练时间方面表现出显著优势。

关键词:人脸重建;神经辐射场;高仿真模型;HACK 模型;深度学习
中图分类号　TP391.9

Immediate Generation Algorithm of High-fidelity Head Avatars Based on NeRF

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

Abstract　To resolve the escalating need for quickly generating high-precision digital humans in fields such as digital entertainment, virtual reality, and the metaverse, this paper proposes a novel method for rapidly generating a high-precision face model based on monocular RGB videos. Meanwhile, a new framework dedicated to precise modeling of the facial and neck regions is constructed. In particular, the proposed framework integrates neural primitives into a parameterized model of the head and neck, utilizing Head-And-neCK(Hereinafter referred to as HACK) as a superior alternative to the widely adopted Face Latent Animated Mesh Estimator(Hereinafter referred to as FLAME). This substitution markedly enhances the precision and efficiency of 3D facial reconstruction. Additionally, the proposed method has designed a real-time adaptive neural radiance field that significantly accelerates the training and reconstruction processes. By introducing a multi-resolution hash grid and employing the nearest triangle search for deformation gradient calculation within the deformation space, the proposed method achieves rapid reconstruction of high-fidelity head and neck models within minutes. Extensive quantitative and qualitative evaluations demonstrate that the proposed model exhibits notable improvements in both rendering quality and training time compared to existing state-of-the-art methods.

Keywords　Head avatar, Neural radiance field, High-fidelity model, HACK model, Deep learning

1　引言

在计算机视觉领域,模型重建技术的快速发展推动了诸多细分方向的显著进步。三维人脸重建技术凭借其广泛的影响力和潜在的应用价值,已成为计算机视觉领域研究的核心焦点之一[1]。该技术致力于通过二维图像或视频生成逼真的三维人脸模型,这为元宇宙、虚拟现实、数字娱乐等多个领域带来了广阔的发展前景[2-3]。众多基于学习的方法[4-9]借助多样化的数据集来提升模型的鲁棒性,进而使其能够适应不同的光照条件、运动姿态以及几何外观特征的变化。

现有方法在重建面部区域方面表现出色,但往往忽视了颈部区域的重建,导致生成模型在颈部区域与现实存在显著差异。从解剖学角度来看,颈部由骨骼、肌肉组织和其他复杂组织组成,这一特点使得对其精确建模颇具挑战。目前,全面

到稿日期:2024-10-14　返修日期:2024-12-09
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)

捕捉面部和颈部运动细节的数据集十分匮乏。尽管部分研究[10-14]已经开始关注颈部建模，但这些方法所采用的模型较为简单，导致模型在变形时与实际情况存在偏差，模型精度欠佳。另外，还有一些研究[15-18]提出通过可微分体积渲染合成复杂静态场景的新颖视图，但此类方法的训练时间较长，通常需要 1 至 5 天。

针对上述问题，本文提出了一种快速生成高精度人脸模型的新方法。首先，以 HACK 模型[19]替代 FLAME 模型进行面部重建。与 FLAME 模型相比，HACK 模型在功能层上独具优势，它能够解耦颈部和喉部运动、面部表情以及外观变化，基于人体解剖学实现对头部的全方位精确控制，并能提供诸如表情、姿势、喉部位置等更多控制参数，极大地提升了模型的个性化与表现力。其次，设计了一种基于神经图形基元[20]实现的神经辐射场(Neural Radiance Fields，NeRF)，并将基元集成到参数化面部模型[21]中。在此基础上，考虑到 HACK 模型包含了更多的参数，会使得模型训练的复杂度有所增加，为了有效加快辐射场的训练速度，设计引入多分辨率哈希编码技术。多分辨率哈希编码技术是一种先进的编码方式，它能够依据数据的不同分辨率特性进行高效编码处理，大幅减少了训练过程中数据处理的计算量，从而实现加快辐射场训练速度的目的。为了提高面对未曾见过的视角或视点时能预测并生成相应的合理图像或数据的能力，集成了基于头颈参数模型的面部重建，在神经辐射场训练期间提供渲染深度图的几何先验[22]。所提方法不仅可以提高渲染质量，而且能够显著加快训练和推理速度。

2 相关工作

2.1 参数化头部和颈部建模

Blanz 等[23]首次提出了通用面部表示概念，借助面部特征的线性组合来生成真实的面部形态，并实现面部形状与颜色的分离。此后，多种面部重建的模型[24-28]相继涌现。尽管这些模型在高保真重建面部区域方面成效显著，但是颈部区域重建常被忽略，致使构建的颈部模型与现实存在偏差。一些研究方法[29-31]虽然考虑了颈部建模，但采用的模型相对简易，在变形过程中仍可能与现实不符。

在头部建模领域，仅有部分方法结合了解剖学约束条件，且生成逼真的头颈模型往往需要大量的人工干预[32-38]。许多基于学习的方法利用多样化的数据集来提升模型的鲁棒性，促使通用模型能够有效适应不同光照条件、动作或个体独特的几何和外观细节[28,32-34]。部分研究采用生成对抗网络开展头部建模工作，旨在解决数据多样性问题[35-36]。但此类方法普遍缺乏明确的控制机制，在实际生产应用场景中部署时面临诸多挑战。综上所述，当前头部建模领域在颈部处理、解剖学结合、控制机制等方面仍存在诸多有待改进之处，为后续研究提供了探索方向。

2.2 静态神经辐射场

许多研究[37-39]聚焦于借助可微体积渲染技术，实现复杂静态场景生成新视图的合成方法。然而，这种方法通常存在训练时间较长的问题。为攻克这一难题，一系列加速技术应运而生，旨在提升训练速度。Gafin 等[39]使用稀疏体素网格来存储每个节点的密度和球谐系数，实现了 100 倍的加速效果。TensorRF[40]将 4D NeRF 场景分解为多个紧凑的低秩张量组件，提升了模型的性能和紧凑性。同时，该方法提出以体素特征网格替代基于坐标的 MLP 的方法，在一定程度上改变了模型的数据处理流程。Müller 等[41]提出的微型 MLP，将空间划分为独立的多级网格，在网格顶点处存储特征向量。空间哈希、点通过插值对特征向量进行编码，然后将这些编码传递给神经网络，合成最终的颜色。此外，一些静态 NeRF 方法[42-45]利用了额外的深度图来优化静态场景的对齐和保真度。

2.3 动态神经辐射场

在将 NeRF[22]用于静态场景重建后，研究者们开始探索将这种高效的模型重建技术扩展到动态变化的场景应用中[46-48]。此过程中，面部重建的方式呈现出多元化发展趋势，其中多数方法把面部重建工作划分于变形空间与规范空间这两个不同的空间范畴内，并运用神经网络来构建二者之间的映射关联。Gafni 等[39]使用 3DMM 跟踪获取的全局表情代码来调整 NeRF 中的多层感知机(Multi-Layer Perceptron，MLP)[49]，实现面部表情的隐式建模。IMAvatar[50]通过学习特定 ID 的纹理、表情混合形状和线性混合蒙皮(Linear Blending Skinning，LBS)权重的隐式表示来进行建模，但该方法的训练时间较长，并且可能会产生不稳定结果。此外，部分研究[51-52]提出了姿态条件逆 LBS 场，从姿态条件的视角对 LBS 场进行逆向构建；还有一些研究[53-54]采用多重初始化寻根循环的优化技术，借助多次初始化与寻根循环操作来提升模型的精确性与稳定性。但体积渲染带来的计算开销，限制了这些方法在实时应用中的可行性。

3 基本方法

为实现快速生成高保真的面部和颈部数字模型的目标，本文设计了一种新方法，以 HACK 模型[19]取代广泛应用的 FLAME 模型。HACK 模型整合了脊椎关节、面部网格、喉部几何形状等多种关键信息，极大地丰富了模型对于不同人种的面部和颈部特征的覆盖度，有助于细致地勾勒出面部的轮廓、五官的比例与形状等；喉部几何形状信息可准确呈现喉部的隆起、凹陷等细节，这很大程度上弥补了 FLAME 模型处理这些复杂特征时的不足。此外，依托神经辐射场，本文引入嵌套于多分辨率哈希网格中的几何信息进行引导[20]，并结合可微分体积渲染技术[55]，实现对面部和颈部区域更为精确的 3D 建模。技术上的融合充分发挥了 HACK 模型的优势，而且在整体建模精度和效果上与现有技术形成了鲜明对比，为快速生成高保真的面部和颈部数字模型提供了一种全新的、更具优势的解决方案。

本文的模型结构如图 1 所示。对于给定图像集 $I=\{I_i\}$ 和经过优化的内参矩阵 $\boldsymbol{K}\in\mathbb{R}^{3\times3}$ 的单目视频[56]，通过跟踪 HACK 网格参数 $M=\{M_i\}$、对应的面部表情系数 $E=\{E_i\}$ 和姿势参数 $P=\{P_i\}$，生成一个基于 NeRF 中规范空间的可操控面部模型。

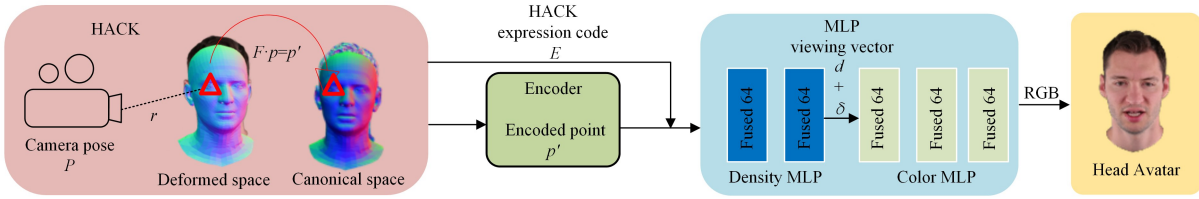

图 1 即时高精度人脸重建模型的结构流程图

Fig. 1 Structure flow digram of immediate generation of high-fidelity facial avatars model

3.1 高保真动画头颈参数模型

为了实现特定的面部表情的渲染，借助先进的神经辐射场技术，对处于变形空间中的光线样本进行标准化处理，并查询规范空间中的辐射场信息，实现面部细节的精确渲染。为了快速重建高精度面部模型，创新性地引入参数化模型 HACK[19] 来替换掉传统 3D 面部重建中广泛使用的 FLAME 模型，以对面部和颈部区域进行一体化建模。在模型构建中，设计了一种映射函数 ϕ 将点 p 从体积渲染的时变变形空间投影至规范空间。该映射函数基于时变表面近似和规范空间中的预定网格构建，确保在不同空间转换过程中的面部和颈部的细节得以正确处理。本文模型包含了各种形状、姿势、表情和喉部混合形状参数，实现了对面部表情以及面部和颈部的骨骼姿态的封装。这些参数与瞬时可变性神经辐射场技术相结合，在可微分体积渲染技术的辅助下，进一步提高了整体建模的精度和效果。该模型可以描述为：

$$IGFA(\beta,\varphi,\theta,\eta,\rho,\alpha)=\{H(\beta,\varphi,\theta,\eta,\rho),A(\alpha)\} \tag{1}$$

其中，$A(\cdot)$ 用于生成基于物理的外观表现；β,φ,θ 和 α 分别为用于控制形状、表情、姿势和外观的参数；η 表示喉部形状参数，以实现颈部尺寸的调节；ρ 为喉部切片控制参数，以实现面部和颈部运动与面部表情动画的分离建模。具体来说，首先利用喉部形状参数 η 来调整颈部尺寸，以预测颈部的几何形状；接着使用喉部切片控制参数 ρ 控制 UV 空间中沿垂直方向的喉部切片运动，从而实现更加精细且逼真的重建效果。$H(\cdot)$ 为面部和颈部的几何参数配置函数，其具体定义如下：

$$H(\beta,\varphi,\theta,\eta,\rho)=LBS(G(\beta,\varphi,\theta,\eta,\rho),B(\beta),\theta,W) \tag{2}$$

其中，$LBS(\cdot)$ 是线性混合蒙皮函数；$G(\cdot)$ 表示包含依赖于身份、姿势、表情和喉部参数的校正变形的网格；$B(\beta)$ 用于指定联合位置；W 表示 $LBS(\cdot)$ 的学习蒙皮权重参数。

为了实现更加个性化的表情和姿势混合形状，使用正交主成分分析(Orthogonal Principal Component Analysis，PCA)对混合形状进行处理，将颈部尺寸参数 η 与丰富的身份信息关联。对于表情和姿势混合形状，本方法没有使用通用参数集，而是选择融入个性化的属性，使混合形状能够更好地反映个体特征。本文模型学习了从颈部形状参数 η 到表情混合形状基础 E_β 和姿势混合形状基础 P_β 的附加映射 M_E 和 M_P，实现了个性化身份的动态生成，从而能在确保面部表情和姿态建模的过程中对个体独特特征的准确捕捉。映射计算式如下：

$$M_E(\beta)\mapsto E_\beta \tag{3}$$

$$M_P(\beta)\mapsto E_\beta \tag{4}$$

为了实现与面部表情相关变形的建模，将表情混合形状与面部动作编码系统(Facial Action Coding System，FACS)[56] 中描述的动作单元相关联。对于每个具有形状参数 η 的对象，从静态捕获的 FACS 表情扫描中提取对应参数，从而增强了模型对不同面部表情的泛化能力。为了有效地学习表情映射 M_E，首先使用包含多种不同身份的数据集 $\{E(\eta)\}$ 来定义表情混合形状的 PCA 空间，然后通过训练一个浅层 MLP 预测 PCA 权重，生成个性化的表情混合形状。这种设计能够更精准地捕捉个体独特的面部表情特征。

对于与姿势相关的变形，首先采用姿势混合形状 P_η，并通过喉部切片姿态参数 θ 推导出的旋转矩阵来捕捉其动态顺序扫描。随后，进一步细化了姿态混合形状 $\{P_\eta\}$ 在 PCA 空间中的表达。与表情映射网络 M_E 类似，采用相同的浅层 MLP 映射网络，基于身份参数估计 PCA 权重。这种设计显著提升了本文提出的参数化模型在个性化控制方面的能力。

在喉部建模方面，将喉部描述为附加在静态姿态网格上的顶点位移，并对颈部以外的区域进行排除。该网格不仅能够表现喉部形状的变化，还在颈部垂直方向上实现了运动约束。喉部形状函数定义如下：

$$N(\beta,\eta,\rho,l)=\eta\cdot\sum_{i=1}^{|\beta|}\beta_i N_i(\rho) \tag{5}$$

其中，$l=[N_1(\rho),N_2(\rho),\cdots,N_{|\eta|}(\rho)]\in\mathbb{R}^{3N\times|\eta|}$ 表示喉部混合形状基础，它能够根据喉部位置参数 ρ 实现垂直位移。该计算式将喉部形状参数 η、喉部尺寸 υ 和喉部位置 ρ 与顶点位移关联，以模拟颈部运动时皮肤下骨骼的滑动效果。为了简化在实际实现中的计算，本文在 UV 空间中对喉部的几何位移进行建模。其中 $N_i(\rho)\in\mathbb{R}^{3\times H\times N}$ 表示喉部混合形状基础，并通过纹理将顶点位移编码为相应的 UV 坐标，呈现为二维图像。因此，喉部函数可以重新表述为：

$$N(\beta,\eta,\rho,l)=\{\upsilon(\mu,\upsilon)\} \tag{6}$$

$$\upsilon(\mu,\upsilon)=\eta\cdot\sum_{i=1}^{|\beta|}N_i(\mu,\nu+\rho) \tag{7}$$

其中，$\upsilon(\mu,\upsilon)$ 表示与喉部相关的顶点 υ 的位移，并映射到点 n 中位置 (μ,υ) 处的值。这种设计能够确保喉部在颈部的垂直方向上移动，提高模型的稳定性和精确性。

3.2 瞬时可变形神经辐射场

HACK 模型的引入虽然为面部与颈部重建工作带来了精度层面的显著提升，但也增加了计算的复杂度并占用了大量的内存空间。为了显著提升模型效率，设计了一种瞬时可变形神经辐射场，其能够在几分钟内完成快速训练与面部重建。其核心优化策略在于将几何引导的可变形神经辐射场集成到多分辨率哈希网格中，借助哈希函数把空间点映射至哈希表，提取高频信息，并结合可微分的体积渲染技术。其中，多分辨率哈希网格凭借其独特的结构特性，能够有效地组织和索引几何信息，为神经辐射场提供精准的几何引导。可微

分体积渲染技术确保了在渲染过程中能够进行精确的梯度计算，使得模型能够在训练过程中快速收敛。

网络以图像序列 $I=\{I_i\}$ 和经过优化的相机内参矩阵 $\boldsymbol{K}\in\mathbb{R}^{3\times3}$ 构成的单目视频为输入，对网格参数 $M=\{M_i\}$ 对应的面部表情系数 $E=\{E_i\}$ 和姿势系数 $P=\{P_i\}$ 进行精确跟踪。为实现特定面部表情的渲染，使用体积渲染技术对来自变形空间的射线上的样本点进行规范化，实现规范空间中的辐射场的查询。传统神经辐射场因基于 MLP 网络架构，在学习高频信息方面存在不足，难以捕捉面部精细纹理与表情微妙变化等细节。为提升性能，本文以神经图形基元[20]为依托，采用多分辨率哈希编码实现高效表示。先对输入的单目视频数据进行预处理，提取关键特征信息，并依据这些信息初始化神经辐射场的参数，训练时借助多分辨率哈希网格的几何引导，使辐射场依据面部、颈部结构与表情姿态变化动态调整参数。该框架中的可微分体积渲染方程可表示为：

$$V=\int_0^D R(t)\cdot\sigma(t)\cdot c(t)\mathrm{d}t+R(D)\cdot v_{\mathrm{bg}} \tag{8}$$

其中，$R(t_n)=\exp\left(-\int_0^{t_n}\sigma(t)\mathrm{d}t\right)$ 表示透射率，即光线在区间 $(0,t_n]$ 内未与任何粒子发生相互作用的概率[33]；$\sigma(t)$ 表示密度；$c(t)$ 表示点 p_t 处的辐射率。

在面部表情编码调节方面，本研究将视频第 i 帧的面部表情编码 $E_i\in\mathbb{R}^{16}$ 和沿光线上的各个采样点 p_t 进行条件化处理。与 NeRFace[39] 和 IMAvatar[50] 有所不同，本文摒弃了额外的可学习帧编码方式，将球谐函数投影到 4 个基函数上[20,57]，以此对观察向量 $\boldsymbol{d}\in\mathbb{R}^{16}$ 进行编码操作，进而生成最终的观察向量编码。随后，把该向量编码与密度 MLP 的输出相结合，以提高面部表情的精确度。值得注意的是，面部表情的调整范围主要限定在动态变化的嘴部区域，而其他区域的编码则设置为常量向量 $E_i=1$。这样的处理策略简化了计算流程，在确保面部表情精准表达的同时提高了整体计算效率。

此外，本文还设计了一个映射函数 $\phi(p,M_i)$，将点 $p\in\mathbb{R}^4$ 从体积渲染的时变变形空间投影至规范空间。该映射函数基于时变表面近似 M_i 和规范空间 M_{sta} 中的预定义网格。在变形空间中，通过最近三角形搜索[58]计算变形梯度 $F\in\mathbb{R}^{4\times4}$，从而将点 p 映射到对应的规范空间点 p'。变形梯度 F 的计算依赖于变形三角形 $T_{\mathrm{def}}\in M_i$ 和规范三角形 $T_{\mathrm{sta}}\in M_{\mathrm{sta}}$ 的对应关系。具体而言，根据三角形相应的切线向量、双切线向量和法线向量计算得出旋转矩阵 $\{\boldsymbol{R}_{\mathrm{sta}},\boldsymbol{R}_{\mathrm{def}}\}\in\mathbb{R}^{3\times3}$，将三角形顶点的平移定义为 $\{t_{\mathrm{sta}},t_{\mathrm{def}}\}\in\mathbb{R}^3$，这些元素共同构成的坐标系框架可以表示为：

$$\boldsymbol{A}_{\mathrm{def}}=\begin{pmatrix}R_{\mathrm{def}} & t_{\mathrm{def}}\\ 0^T & 1\end{pmatrix} \tag{9}$$

$$\boldsymbol{A}_{\mathrm{sta}}=\begin{pmatrix}R_{\mathrm{sta}} & t_{\mathrm{sta}}\\ 0^T & 1\end{pmatrix} \tag{10}$$

其中，$\boldsymbol{A}_{\mathrm{def}}$ 表示变形空间坐标系，$\boldsymbol{A}_{\mathrm{sta}}$ 表示规范空间中的坐标系。

由于变形空间与规范空间之间存在潜在的三角形尺寸变化问题，因此通过三角形相对表面积的变化对三角形的各向同性缩放因子 $\gamma=\frac{a_{\mathrm{def}}}{a_{\mathrm{sta}}}$ 进行计算。变形梯度 F 的定义为：

$$F=A_{\mathrm{sta}}\cdot\begin{pmatrix}\gamma\mathbf{I} & 0\\ 0^T & 1\end{pmatrix}\cdot A_{\mathrm{def}}^{-1} \tag{11}$$

$$C=\begin{pmatrix}\gamma\mathbf{I} & 0\\ 0^T & 1\end{pmatrix} \tag{12}$$

为了解决各三角形局部坐标系引起的变换不连续性问题，对三角形边界相邻面的变换应用指数加权平均，以实现平滑过渡。

$$\hat{F}=\frac{1}{\sum_{f\in A_t}\omega_t}\sum_{f\in A_t}\omega_f F_f \tag{13}$$

其中，$\omega_f=\exp(-\beta\parallel c_f-p\parallel_2)$，$A_t$ 为三角形边界相邻面的集合，$\beta=4$，c_f 为对应的质心。

为了加速神经辐射场的交互式渲染并实现瞬时优化，采用经典的包围体层次结构（Bounding Volume Hierarchy，BVH）[59]，其显著提升了射线上采样点 p_t 的最近三角形的查找速度[33]。本文方法基于第 i 帧的变形网格 M_i 构建 BVH，从而实现与规范网格的映射。

3.3 训练损失

神经辐射场的优化是通过色彩重建目标和 HACK 几何先验的协同作用来实现的。本文设计具有分段恒定密度和颜色的体积渲染方法[60]，该方法表示为：

$$\hat{C}(t_{N+1})=\sum_{n=1}^{N}\tau_n\cdot a_n\cdot c_n \tag{14}$$

其中，$\tau_n=\prod_{n=1}^{N-1}(1-a_n)$ 表示透射率；权重 $a_n=\mathbb{I}-\exp(-\sigma_n\delta_n)$，$\delta_n$ 表示步长。

关于颜色损失，采用 Huber 损失函数[61]，其中参数 $\rho=1$。

$$L_{\mathrm{color}}(r)=\begin{cases}\frac{1}{2}(C-\hat{C})^2, & |C-\hat{C}|<\rho\\ \rho(|C-\hat{C}|-\frac{1}{2}\rho), & \text{其他}\end{cases} \tag{15}$$

对于深度损失，用新重建模型生成面部模型的几何先验结构对跟踪网络进行光栅化。深度损失的计算式为：

$$L_{\mathrm{dep}}(r)=\sum_r|\mathbb{I}_{\mathrm{face}}(z(r)-\hat{z}(r))| \tag{16}$$

其中，$\hat{z}=\sum_{n=1}^{N}\tau_n\cdot a_n\cdot t_n$，$t_n$ 为当前采样点的位置；$\mathbb{I}_{\mathrm{face}}\cdot$ 为面部区域损失计算中的分割指示函数[62]。

最终的总损失计算式为：

$$L_{\mathrm{total}}=\sum_r\lambda_{\mathrm{color}}(r)L_{\mathrm{color}}(r)+\lambda_{\mathrm{dep}}(r)L_{\mathrm{dep}}(r) \tag{17}$$

其中，$\lambda_{\mathrm{dep}}(r)$ 控制几何先验的影响，$\lambda_{\mathrm{color}}(r)$ 控制面部掩模对颜色损失进行加权。

对于颜色和密度预测，设计了两个完全集成的 MLPs[63]，每个 MLP 包含 64 个神经元。密度 MLP 输出一个特征向量 $\boldsymbol{\sigma}\in\mathbb{R}^{16}$，其中第一个元素表示对数空间密度。随后将向量 $\boldsymbol{\sigma}$ 与编码的相机视点向量 $\boldsymbol{d}$ 连接，并作为颜色 MLP 的输入。

4 实验论证

4.1 实验数据

本文方法以单目视频作为输入，使用了一个包含 12 名不同人物的多样化数据集[60]。该数据集充分考虑了多方面因素以确保多样性和有效性，旨在全面涵盖可能影响面部和

头部重建的各种变量，增强使用结果的普适性。数据来源包括相机录制的演员视频以及 YouTube 视频序列。首先将视频进行裁剪处理，并通过子采样将帧率调整至 25 fps，视频的分辨率标准化为 512×512 像素。在数据预处理阶段，利用先进的抠图技术[64]，结合现有的面部解析框架[62]，实现对背景和人物前景的精确分割，将人物从复杂背景中分离出来，并去除其衣物，使模型更专注于面部及头部区域的重建。为了确保对单目输入视频数据面部跟踪的时序稳定性，使用 MICA[65]提供的基于合成分析的面部跟踪器。该跟踪器以 Face2Face[52]为基础构建。Face2Face 是一种经典的面部处理技术，在面部表情迁移等方面有着重要应用，而 MICA 则在其基础上结合了基于采样的可微分渲染技术，以便更精准地捕捉面部跟踪过程中的细节变化。此外，本文还使用了 Mediapipe[66]，它是一种广泛应用于多媒体处理的工具库，提供了丰富的图像处理功能。在本实验中，通过借助两个额外的混合形状来优化眼睑和虹膜的跟踪，提升了跟踪的细腻度和准确性。

4.2 评价指标

在本文的定量图像评估过程中，采用了多种像素级指标以全面衡量各方法生成图像的质量。使用的评估指标包括结构相似性指数（Structural Similarity Index Measure，SSIM）、峰值信噪比（Peak Signal-to-Noise Ratio，PSNR）、感知图像补偿相似度（Learned Perceptual Image Patch Similarity，LPIPS）[67]。其中，SSIM 依据图像亮度、对比度与结构信息评估生成图像与真实图像的结构相似性，数值越接近 1，则结构越相似；PSNR 计算像素点灰度值差异，PSNR 值越高，像素差异越小；LPIPS 利用深度学习模型特征从视觉感知评估差异，值越低则与人眼感知的真实图像越相似。同时，统计生成单帧图像平均渲染时间（Times）以评估图像生成速度。

为深入探究模型在实际应用中的表现及受众接受度，本文采用李克特测试的方法进一步对模型的重建效果开展了主观心理学测试。此次测试共召集了 28 位同学，每位同学分别对生成结果的图像质量、模型自然性、模型可信度和重建效果是否符合期望值进行了测评，每个问题的回答均采用 1－5 分的评分系统，1 分表示完全不同意，2 分表示不同意，3 分表示中立或无意见，4 分表示同意，5 分表示完全同意。对于图像质量，96.4%的受试者（27 人）评价为 4 分或 5 分；对于模型自然性，96.4%的受试者（27 人）评价为 4 分或 5 分；对于模型可信度，89.3%的受试者（25 人）评价为 4 分或 5 分；对于重建效果是否符合期望，92.8%的受试者（26 人）评价为 4 分或 5 分。由结果可看出，模型生成效果较为真实，虽然在部分细节的构建上还有进一步提升空间，但模型总体评价较好，符合大众的心理预期。

4.3 实验对比

将本文方法与当前的先进方法进行了对比。为了评估所提方法生成的图像质量和视图外推方面的性能，将其与 NeRFace[39]，IMAvatar[50]，Neural Head Avatars（NHA）[51]，INSTA[60]进行了比较，如表 1 所列。

表 1 本文方法与不同方法的对比结果

Table 1 Comparison of results of this paper's method with different methods

Method	*PSNR*	*SSIM*	*LPIPS*	*Times*
NHA	27.71	0.95	0.04	0.63
IMAvatar	27.62	0.94	0.06	12.34
NeRFace	29.28	0.95	0.07	9.68
INSTA	28.97	0.95	0.05	0.05
Ours	29.11	0.97	0.02	0.05

从表 1 中可以看出，本文方法在图像质量和生成速度方面均表现优秀。具体来说，本文方法不仅能够生成与真实图像高度相似的图像，还能够有效降低视觉感知上的差异。

基于以上方法的原始代码，在相同的数据集上进行对比实验，结果对比如图 2 所示。

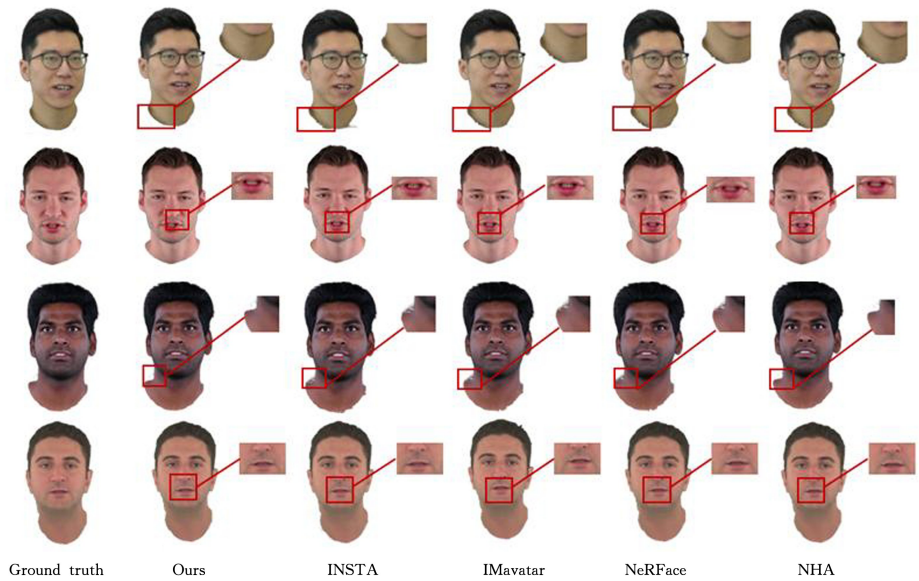

图 2 本文方法与其他最先进的方法生成的结果图对比

Fig. 2 Comparison of result plots generated by the proposed method with other state-of-the-art methods

实验结果显示，NHA[51]在耳朵、头发、牙齿和脖颈等特定区域呈现出较明显的伪影现象，影响了整体的图像真实性。

从像素层面的误差分析来看，这些区域的像素误差相较于本文方法较大。NeRFace[39]在眼部区域，特别是瞳孔周围出现了一定程度的模糊和伪影，削弱了图像细节的清晰度；而本文方法基于神经元表征的辐射场，能很好地学习到眼部的细微结构和光影变化等高频信息，在瞳孔的纹理、光泽以及周围组织的过滤等方面进行细致的重建与渲染，有效避免了模糊和伪影现象。INSTA[60]在颈部边缘位置出现了特定程度的伪影瑕疵；而本文方法渲染的结果在骨骼结构到皮肤表面的过渡都能自然呈现，使颈部区域的重建效果更为精准。IMAvatar[50]在训练和运行过程中表现出了一定的不稳定性，存在收敛性问题，会导致训练过程异常终止，影响了模型的正常运行。相比之下，本文方法表现出了显著的优势。在不同表情和姿势条件下，本文方法能够更加精准地生成面部模型，尤其在重建颈部结构时表现出色。相较于其他使用 FLAME 模型的方法，本文方法采用更精确的 HACK 模型，对于颈部的重建能够充分利用 HACK 模型所包含的颈部、喉部几何形状等关键信息，避免了 FLAME 模型在处理颈部细节时的不确定性；通过对动态面部表情的细致处理，以及将辐射场和多分辨率哈希网格相结合，使得在辐射场训练过程中能够有效捕捉和处理面部表情变化过程中产生的高频信息，特别是嘴角、眼角等部位的细微变化。这一策略使得模型在渲染表情时更细腻，生成更真实、自然的效果。此外，本文方法在训练和生成速度上也具备优势，HACK 模型提供的初始参数能够加速辐射场的收敛过程，减少了不必要的计算量，因此能快速完成面部模型的训练和渲染过程。

为了全面评估本文方法的性能，分别在不同的表情（见图 3）、姿态条件（见图 4）和光照（见图 5）下进行了鲁棒性测试。对于表情变化，选择了具有典型表情的人物样本，测试了模型在大笑、皱眉、张嘴等常见表情下的生成效果。对于姿态变化，考查了模型在侧脸、仰头等不同姿态下的表现。针对光照条件，模拟了正常光照、强光照和弱光照 3 种情况，并在这些条件下进行了模型生成测试。实验结果表明，本文方法在各类表情、姿态和光照条件下均能保持良好的性能，能够精准地重建面部和颈部模型。

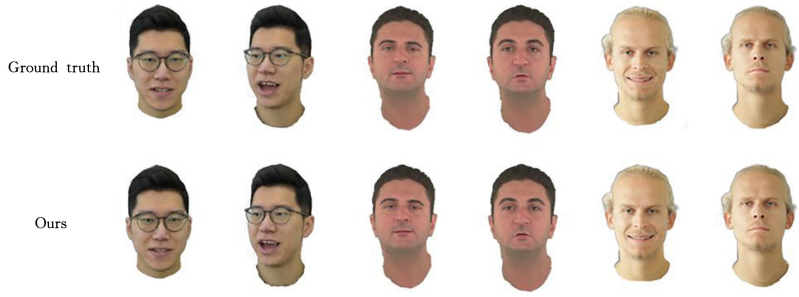

图 3 不同表情条件生成效果的对比

Fig. 3 Comparison of generation effect of different expression conditions

图 4 不同姿势条件生成效果的对比

Fig. 4 Comparison of generation effect of different pose conditions

图 5 不同光照条件生成效果的对比

Fig. 5 Comparison of generation effect of different light conditions

4.4 消融实验

为了进一步评估本文方法，进行了多轮消融实验，分析了网络中 HACK 模型[19]对模型性能的贡献，特别是对视图合成效果的影响；比较了本文提出的动态神经辐射场与不同变形场在生成效果上的差异。本文方法通过对喉部尺寸、位置等关键参数的精确调节，不仅提高了图像的真实性，还增强了模型在复杂姿势下的表现能力。如图 6 所示，与现有的主流方法相比，本文方法在重建颈部区域时展现了更高的精确度，能够更自然地处理颈部的细节和复杂形变，使得模型在不同角度和姿势下都能保持较高的重建质量，有效解决了传统方法在该区域容易出现的伪影和失真问题，进一步提升了整体生成图像的精细程度。

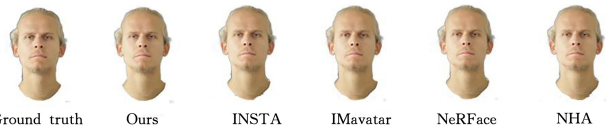

图 6 颈部重建的对比

Fig. 6 Comparison of neck reconstruction

为了验证所提出的基于神经图形基元的动态神经辐射场中变形场对渲染质量的显著影响，进行了两组实验，结果如图 7 所示。具体来说，进行了以下两种设置的对比实验：1）使用全局条件替代局部条件；2）采用具有每帧可学习代码但没有变形场的全局调节[39]，旨在分析变形场在图像重建中的关键作用。

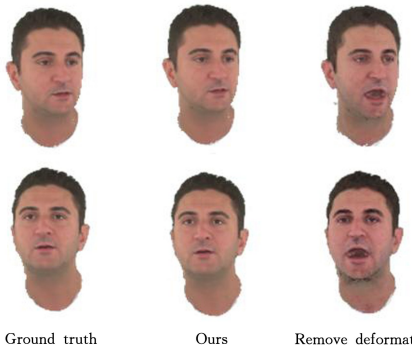

图 7 变形场消融研究对比

Fig. 7 Comparison of deformation field ablation research

实验结果表明，移除变形场给渲染质量带来了明显的负面影响。当采用全局条件替代局部条件时，模型生成的图像更容易出现重影之类的伪影问题。这种伪影现象会干扰模型对细节的捕捉，在复杂的动态场景中，这种影响尤为突出，会导致生成图像的质量明显下降。在第二组实验中，运用全局调节且不含变形场的方式，模型虽然具备一定的灵活性，但缺乏精准的局部调节能力，因此图像重建的精度仍难以达到理想水平。相比之下，本文方法通过结合局部条件与基于网格的变形场，极大地减少了过拟合现象，在短训练序列的情况下表现尤为出色。局部条件使得模型在动态变化的场景中能够更精确地调整每个区域；变形场则提供了精确的形变处理，确保了生成的图像在不同视角和复杂姿势下都能维持高质量水准。这一设计有效弥补了传统方法在复杂场景下过度依赖全局信息而产生伪影的缺陷，进一步提升了模型的重建能力和渲染效果。

结束语 本文方法尽管在质量和效率上优于当前最先进的 3D 面部参数化建模技术，但在未来发展中仍存在挑战。首先，尽管模型在面部和颈部区域的重建效果令人满意，但模型的整体细节水平仍有进一步优化的空间，例如对更夸张的表情的建模不够生动自然。其次，在捕捉头发的动态变化时，仍存在一定局限性，对于复杂发型和发丝的细微运动的重建精度仍需要进一步提升，采用合理的算法提高头发重建的精度将是一个提高重建模型真实度的重要方向。此外，面部跟踪的准确性也是一个关键挑战，直接影响着生成模型的质量，面部跟踪精确度不高会导致面部细节丢失或模型不够逼真，对最终的建模效果产生不利影响。因此，在未来的工作中，提升面部跟踪算法的鲁棒性，确保更高的对齐精度，是提高模型重建质量的重点。

本文提出一种以单目 RGB 视频为输入，引入 HACK 模型并结合基于神经图形基元的动态神经辐射场，快速生成高精度三维面部和颈部重建模型的新方法。该方法兼顾人类身份特征、面部表情与颈部喉部运动，使用瞬时可变性神经网络显著提升重建效率，在虚拟现实、影视制作及医疗仿真等领域有卓越表现，对比实验与消融研究也证实其在速度、真实感、精度及表现力等方面具有显著优势。所提模型虽优于现有技术，但仍面临整体细节待优化、头发动态捕捉有局限、面部跟踪准确性待提升等挑战，后续将重点关注面部跟踪算法鲁棒性与对齐精度的提升，以提高重建质量。

参 考 文 献

[1] 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): 1200005.

[2] LIU X N, CHEN C Y, HU X J, et al. Neural radiation field virtual viewpoint picture synthesis with depth information supervision[J]. Chinese Journal of Image and Graphics, 2024, 29(7): 2035-2045.

[3] LI J Y, CHENG L C, HE J X, et al. Research status and prospects of neural radiation fields[J]. Journal of Computer-Aided Design and Graphics, 2024, 36(7): 995-1013.

[4] BOUAZIZ S, WANG Y, PAULY M. Online modeling for realtime facial animation[J]. ACM Transactions on Graphics, 2013, 32(4): 1-10.

[5] CAO C, SIMON T, KIM J K, et al. Authentic volumetric avatars from a phone scan[J]. ACM Transaction on Graphics, 2022, 41(4): 1-19.

[6] CHEN A, XU Z, GEIGER A, et al. Tensorf: Tensorial radiance fields[C]// European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 333-350.

[7] CHEN W Z, GAO J, LING H, et al. Learning to predict 3d objects with an interpolation-based differentiable renderer[C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019: 9609-9619.

[8] JI Y, YU Y Q. Optimization algorithm for speech facial video generation based on dense convolutional generative adversarial networks and keyframes[J]. Journal of Jilin University (Engineering and Technology Edition), 2025, 55(3): 986-992.

[9] ZHANG X, KU S P. Facial super-resolution reconstruction method based on generative adversarial networks[J]. Journal of Jilin University (Engineering and Technology Edition), 2025, 55(1): 333-338.

[10] CHOI B, EOM H, MOUSCADET B, et al. Animatomy: An animator-centric, anatomically inspired system for 3d facial modeling, animation and transfer[C]// SIGGRAPH Asia 2022 Conference Papers. 2022: 1-9.

[11] DANĚČEK R, BLACK M J, BOLKART T. Emoca: Emotion driven monocular face capture and animation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

Recognition. 2022:20311-20322.
[12] LATTAS A, MOSCHOGLOU S, PLOUMPIS S, et al. Avatarme++: Facial shape and brdf inference with photorealistic rende-ring-aware gans[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(12): 9269-9284.
[13] LIU L, GU J, ZAW LIN K, et al. Neural sparse voxel fields[J]. Advances in Neural Information Processing Systems, 2020, 33: 15651-15663.
[14] LOPER M, MAHMOOD N, ROMERO J, et al. SMPL: A skinned multi-person linear model[J]. ACM Transaction on Graphics, 2015, 34(6): 1-16.
[15] BARRON J T, MILDENHALL B, TANCIK M, et al. MipneRF: A multiscale representation for anti-aliasing neural radiance fields[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5855-5864.
[16] BARRON J T, MILDENHALL B, VERBIN D, et al. Mip-nerf 360: Unbounded anti-aliased neural radiance fields[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5470-5479.
[17] ZHANG K, RIEGLER G, SNAVELY N, et al. NeRF++: Analyzing and improving neural radiance fields[J]. arXiv: 2010.07492, 2020.
[18] MILDENHALL B, HEDMAN P, MARTIN-BRUALLA R, et al. NeRF in the dark: High dynamic range view synthesis from noisy raw images[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 16190-16199.
[19] ZHANG L, ZHAO Z, CONG X, et al. Hack: Learning a parametric head and neck model for high-fidelity animation[J]. ACM Transactions on Graphics, 2023, 42(4): 1-20.
[20] MÜLLER 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.
[21] LI T, BOLKART T, BLACK M J, et al. Learning a model of facial shape and expression from 4D scans[J]. ACM Transactions on Graphics, 2017, 36(6): 1-17.
[22] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis[J]. Communications of the ACM, 2021, 65(1): 99-106.
[23] BLANZ V, VETTER T. A morphable model for the synthesis of 3D faces[C]// Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'99). 1999: 187-194.
[24] KIM H, GARRIDO P, TEWARI A, et al. Deep video portraits [J]. ACM transactions on graphics, 2018, 37(4): 1-14.
[25] 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.
[26] WANG Y J, LI S Z, HAN J Y, et al. Single-image three-dimensional face reconstruction based on convolutional neural network [J]. Sensors and Microsystems, 2021, 40(6): 52-56.
[27] DONG J Z, ZUO W M. Research on self-encoding voxel network for 3D face reconstruction [J]. Intelligent Computers and Applications, 2020, 10(6): 303-309.
[28] LI T, BOLKART T, BLACK M J, et al. Learning a model of facial shape and expression from4D scans[J]. ACM Transaction on Graphics, 2017, 36(6): 1-17.
[29] LOPER M, MAHMOOD N, ROMERO J, et al. SMPL: A skinned multi-person linear model[J]. Transaction on Graphics, 2015, 34(6): 1-16.
[30] CHOI B, EOM H, MOUSCADET B, et al. Animatomy: An animator-centric, anatomically inspired system for 3d facial modeling, animation and transfer[C]// SIGGRAPH Asia 2022 Conference Papers. 2022: 1-9.
[31] 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.
[32] LI R, BLADIN K, ZHAO Y, et al. Learning formation of physically-based face attributes[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3410-3419.
[33] LAINE S, KARRAS T, AILA T, et al. Production-level facial performance capture using deep convolutional neural networks [C]// Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation. 2017: 1-10.
[34] CALISKAN A, KICANAOGLU B, KIM H. PAV: Personalized Head Avatar from Unstructured Video Collection[J]. arXiv: 2407.21047, 2024.
[35] KABADAYI B, ZIELONKA W, BHATNAGAR B L, et al. Ganavatar: Controllable personalized gan-based human head avatar [C]// 2024 International Conference on 3D Vision (3DV). IEEE, 2024: 882-892.
[36] RAI A, GUPTA H, PANDEY A, et al. Towards realistic generative 3d face models[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 3738-3748.
[37] BARRON J T, MILDENHALL B, TANCIK M, et al. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5855-5864.
[38] LIU L, GU J, ZAW LIN K, et al. Neural sparse voxel fields[J]. Advances in Neural Information Processing Systems, 2020, 33: 15651-15663.
[39] 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.
[40] CHEN A, XU Z, GEIGER A, et al. Tensorf: Tensorial radiance fields[C]// European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 333-350.
[41] MÜLLER 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.
[42] TESCHNER M, HEIDELBERGER B, MÜLLER M, et al. Optimized spatial hashing for collision detection of deformable objects[C]// VMV. 2003: 47-54.
[43] WEI Y, LIU S, RAO Y, et al. Nerfingmvs: Guided optimization

of neural radiance fields for indoor multi-view stereo[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:5610-5619.

[44] ROESSLE B,BARRON J T,MILDENHALL B,et al. Dense depth priors for neural radiance fields from sparse input views [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:12892-12901.

[45] DENG K,LIU A,ZHU J Y,et al. Depth-supervised nerf:Fewer views and faster training for free[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:12882-12891.

[46] 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.

[47] XU Y,CHEN B,LI Z,et al. Gaussian head avatar:Ultra high-fidelity head avatar via dynamic gaussians[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024:1931-1941.

[48] DENG K,LIU A,ZHU J Y,et al. Depth-supervised nerf:Fewer views and faster training for free[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:12882-12891.

[49] THIES J,ZOLLHOFER M,STAMMINGER M,et al. Face2face:Real-time face capture and reenactment of RGB videos[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:2387-2395.

[50] ZHENG Y,ABREVAYA V F,BÜHLER M C,et al. Im avatar:Implicit morphable head avatars from videos[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:13545-13555.

[51] GRASSAL P W,PRINZLER M,LEISTNER T,et al. Neural head avatars from monocular rgb videos[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:18653-18664.

[52] 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.

[53] 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.

[54] 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.

[55] KIM H,GARRIDO P,TEWARI A,et al. Deep video portraits [J]. ACM transactions on graphics,2018,37(4):1-14.

[56] PARK K,SINHA U,BARRON J T,et al. Nerfies:Deformable neural radiance fields[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:5865-5874.

[57] AXLER S,BOURDON P,WADE R. Harmonic function theory [M]. Springer Science & Business Media,2013.

[58] EKMAN P,FRIESEN W V. Facial action coding system[J/OL]. http://doi.org/10.1037/t27734-000.

[59] CLARK J H. Hierarchical geometric models for visible surface algorithms[J]. Communications of the ACM,1976,19(10):547-554.

[60] ZIELONKA W,BOLKART T,THIES J. Instant volumetric head avatars[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023:4574-4584.

[61] HUBER P J. Robust estimation of a location parameter[M]// Breakthroughs in statistics:Methodology and distribution. New York:Springer,1992:492-518.

[62] YU C,GAO C,WANG J,et al. Bisenet v2:Bilateral network with guided aggregation for real-time semantic segmentation [J]. International Journal of Computer Vision,2021,129:3051-3068.

[63] MÜLLER T. tiny-cuda-nn [J/OL]. https://github.com/NVlabs/tiny-cuda-nn.

[64] LIN S,YANG L,SALEEMI I,et al. Robust high-resolution video matting with temporal guidance[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022:238-247.

[65] ZIELONKA W,BOLKART T,THIES J. Towards metrical reconstruction of human faces[C]// European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 250-269.

[66] LUGARESI C,TANG J,NASH H,et al. Mediapipe:A framework for building perception pipelines[J]. arXiv:1906.08172,2019.

[67] ZHANG R,ISOLA P,EFROS A A,et al. The unreasonable effectiveness of deep features as a perceptual metric[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:586-595.

SHENG Xiaomeng, born in 2000, postgraduate. Her main research interests include neural radiance fields and 3D human face 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：柯颖）