

## 基于法线贴图增强隐式函数的单图像三维头部重建

刘乐元<sup>1,2)</sup>, 刘旭<sup>1)</sup>, 孙见弛<sup>1)</sup>, 陈靓影<sup>1,2)\*</sup>

<sup>1)</sup> (华中师范大学国家数字化学习工程技术研究中心 武汉 430079)

<sup>2)</sup> (华中师范大学教育大数据应用技术国家工程研究中心 武汉 430079)  
(chenjy@ccnu.edu.cn)

**摘要:** 三维头部重建是建构元宇宙的基础技术之一, 在影视制作、游戏娱乐、智能教育等领域也具备广阔的应用前景. 从单图像重建三维头部模型可以最大程度地节省成本并大幅提高操作的便捷性. 然而, 单图像三维头部重建是一个非适定问题, 现有算法普遍存在重建的三维头部模型保真度低、细节少以及算法泛化能力差等问题. 为此, 文中提出了一种基于法线贴图增强隐式函数的单图像三维头部重建算法. 首先, 设计了一个法线贴图估计子网络从单幅图像估计头部法线贴图; 其次, 将三维头部模型表面看作由隐式函数描述的水平集, 建立了端到端的深度神经网络, 从输入图像及法线贴图中提取视觉特征并判别三维空间中各点位于该水平集等值面的概率. 文中算法在 FCH(FaceScape, CoMA, HeadSpace) 和 FaceVerse 公共数据集上所重建三维头部模型的平均倒角距离分别为 0.7696 mm 和 1.3080 mm, 大幅优于对比的单图像三维头部重建算法; 在 2 个从互联网采集的数据集上的实验结果也表明, 文中算法能从包含不同年龄、人种、性别、面部表情的单幅肖像图像中重建出具备高保真度和丰富细节的三维头部模型, 且具备强大泛化能力.

**关键词:** 三维头部重建; 单图像; 深度神经网络; 隐式函数; 法线贴图  
**中图分类号:** TP391.41 **DOI:** 10.3724/SP.J.1089.2024.20166

## Single-Image 3D Head Reconstruction Using Normals Enhanced Implicit Function

Liu Leyuan<sup>1,2)</sup>, Liu Xu<sup>1)</sup>, Sun Jianchi<sup>1)</sup>, and Chen Jingying<sup>1,2)\*</sup>

<sup>1)</sup> (National Engineering Research Center for E-Learning, Central China Normal University, Wuhan 430079)

<sup>2)</sup> (National Engineering Laboratory for Educational Big Data, Central China Normal University, Wuhan 430079)

**Abstract:** 3D head reconstruction is one of the fundamental techniques for building the metaverse, and it also has broad applications in the fields of film and cartoon creation, game design, and intelligent education. Compared to modeling 3D heads manually by artists or capturing head scans using 3D scanners or stereo imaging systems, reconstructing 3D heads from a single image is much more economical and practical. However, single-image 3D head reconstruction is an ill-posed problem, and existing methods often suffer from low fidelity, fewer details, and poor generalization ability. To this end, this paper proposes a single-image 3D head reconstruction method using normals enhanced implicit function. First, a deep network for estimating normal maps of heads from a single image is designed. Second, the surface of a 3D head model is defined as a level set described by an implicit function, and an end-to-end deep neural network is built to extract visual features from the input image and the estimated normal map and predict whether a 3D point lies on the head surface or not. Our method achieves Chamfer distances of 0.7696 mm and 1.3080 mm

收稿日期: 2023-02-27; 修回日期: 2023-11-03. **基金项目:** 国家自然科学基金面上项目(62077026); 中央高校基本科研业务费优秀青年团队项目(CCN22QN012). 刘乐元(1982—), 男, 博士, 副教授, 博士生导师, CCF 会员, 主要研究方向为计算机视觉、计算机图形学、人机交互技术; 刘旭(1999—), 男, 硕士研究生, 主要研究方向为计算机视觉、计算机图形学; 孙见弛(1999—), 男, 硕士研究生, 主要研究方向为计算机视觉; 陈靓影(1973—), 女, 博士, 教授, 博士生导师, 论文通信作者, 主要研究方向为计算机视觉、模式识别.

respectively on the FCH (FaceScape, CoMA, HeadSpace) and FaceVerse datasets, which outperform other single-image 3D head reconstruction methods by a large margin; experimental results on images collected from the Internet also show that our method can reconstruct 3D head models with high fidelity and rich details from single images with different ages, human races, genders, and facial expressions, and has a strong generalization capability.

**Key words:** 3D head reconstruction; single-image; deep neural network; implicit function; normal map

## 1 相关工作

单图像三维头部重建是指利用深度学习算法从单幅肖像图像自动生成由网格(mesh)表示的三维头部模型。随着元宇宙的兴起, 三维头部重建展现出了巨大的应用前景<sup>[1]</sup>。除此以外, 三维头部重建在影视制作<sup>[2]</sup>、游戏娱乐<sup>[3]</sup>和智能教育<sup>[4]</sup>等领域都具备广阔的应用前景。与手工建模或基于三维扫描仪和立体视觉系统<sup>[5]</sup>重建三维头部模型相比, 单图像三维头部重建算法可以最大程度地节省成本, 且大幅提高操作的便捷性<sup>[6]</sup>。因此, 单图像三维头部重建成为了近年来一个研究热点。

作为一个典型的非适定(ill-posed)问题<sup>[7-8]</sup>, 从单幅图像中重建三维头部模型是一项具有挑战的任务。为应对这一挑战, 研究者陆续提出了一系列基于深度学习的方法, 以从数据中学习先验知识, 它们可以粗略地分为参数化方法<sup>[9-18]</sup>和非参数化方法<sup>[19-25]</sup>2 大类。

参数化方法一般采用三维可形变模型(如 BFM<sup>[26]</sup>, Flame<sup>[27]</sup>和 UH3D<sup>[28]</sup>等)表示面部或头部, 从图像中提取特征并估计出少量模型参数<sup>[29]</sup>, 即可从初始统计模型形变出特定的三维头部或面部。Deng 等<sup>[15]</sup>通过残差网络估计出 BFM 模型参数, 并采用弱监督学习在小规模数据上提升重建效果。Shang 等<sup>[16]</sup>在训练时输入同一个体的多视角图像以估计多个模型参数, 并利用多视角几何一致性进一步约束模型重建。Bao 等<sup>[17]</sup>利用微分渲染扩展数据量, 并生成扰动来提升三维可形变模型的表达能力, 进而提高重建头部模型几何形状的准确度。Feng 等<sup>[18]</sup>首先利用回归网络预测出 Flame 模型参数, 用于构建粗糙三维头部模型, 然后采用细节捕获模块生成细节丰富的 UV 位移图, 最后将粗糙模型与 UV 位移图结合, 以增强重建模型的纹理细节。综上所述, 参数化方法虽然简便, 但是受限于参数化模型的表达能力, 其重建出的三维头部或面部模型容易趋近于原始统计模型, 难以恢复出

高保真度、细节丰富的三维头部/面部模型。

与参数化方法不同, 非参数化方法直接从单幅图像回归/判别三维头部模型上各个顶点的位置。Jackson 等<sup>[23]</sup>将三维模型网格体素化成易于卷积神经网络操作的形式, 之后利用卷积操作直接学习像素到三维坐标的映射关系, 以重建三维头部模型。虽然该方法可以恢复中性表情的三维头部模型, 但受制于体素模型的分辨率<sup>[30]</sup>, 其重建模型的精度有限。Feng 等<sup>[24]</sup>使用二维 UV 位移图来记录三维人脸点云的位置, 并利用级联卷积神经网络从输入图像中回归出顶点位移图, 进而将其转换成重建的三维人脸模型。然而, UV 位移图仅能表达有限的三维人脸模型顶点语义信息, 因此其重建精度低且对细节恢复能力有限。近年来, 一些研究者尝试将三维网格和图卷积神经网络相结合进行三维头部重建。Lin 等<sup>[25]</sup>采用回归器预测出三维可形变模型参数后, 通过图卷积神经网络生成高质量的纹理, 但该方法并没有对三维头部模型的几何形状进行任何调整。Cheng 等<sup>[22]</sup>采用卷积神经网络提取图像特征, 并利用图卷积解码器回归三维模型的顶点坐标, 但其受限于图卷积对于空间信息的捕捉能力<sup>[31]</sup>, 其重建的三维模型的几何结构准确度欠佳。

综上, 虽然单图像三维头部重建算法在近年来取得了一定的进展, 但仍然存在以下问题: (1) 重建出的三维模型保真度较低。虽然一些算法所重建的三维头部模型在纹理的掩饰下看起来较为逼真, 但其几何结构实际上与真实头部存在较大差异, 在去除纹理之后甚至难以与输入图像中的人物身份相匹配。(2) 重建的三维模型表面细节较少或恢复的局部几何结构精度不佳。大部分算法只能重建出面部区域, 对耳朵、头发等部分几乎无能为力; 虽然一些算法可以恢复出一般面部表情的大致形状, 但对于极端面部表情和微观面部表情的表达能力较弱。(3) 算法的泛化能力较差, 难以适应年龄、人种、性别等变化。例如, 通过成年人数据集

训练的模型在对儿童肖像进行重建时,重建的结果看起来仍然像成年人.

针对现有算法重建的三维头部模型保真度低、细节少以及算法泛化能力差的问题,本文提出了一种基于法线贴图增强隐式函数的单图像三维头部重建算法.首先,设计了一个法线贴图估计子网络从单幅图像估计头部法线贴图.其次,将三维头部模型表面看作由隐式函数描述的水平集,建立了端到端的深度神经网络,从输入图像及法线贴图中提取视觉特征,并判别三维空间中各点位于该水平集等值面的概率.图 1 所示为本文算法从单幅肖像图像重建的三维头部模型.可以看出,本文算法对不同年龄、人种、性别、面部表情的肖像图像都能重建出高保真度、细节丰富的三维头部模型.

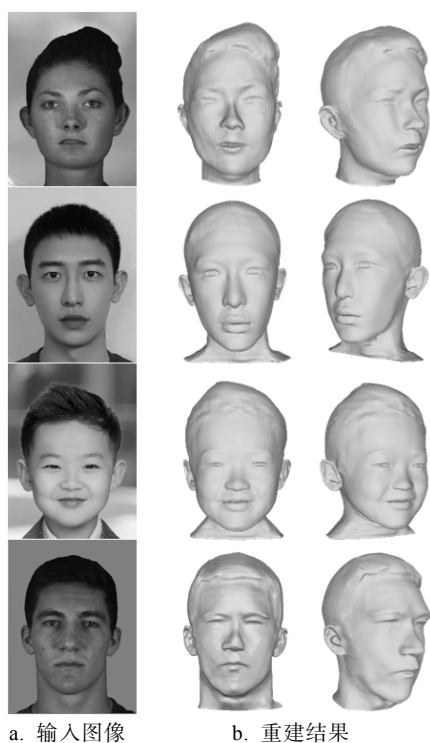


图 1 本文方法从单图像重建的三维头部模型

本文的主要贡献为: (1) 提出了基于深度隐式函数的单图像三维头部重建算法.得益于隐式函数对三维物体表面强大的表达能力与灵活性,本文算法可以有效地解决现有算法重建的三维头部模型保真度低和算法泛化能力差的问题. (2) 从单图像中估计了法线贴图,并从中提取视觉特征用于三维头部重建.借助法线贴图的局部几何结构描述能力,丰富了重建三维头部模型的表面细节并提高了局部微小几何结构的精准度. (3) 在公共

数据集和从互联网采集的数据集上进行了充分的实验验证.实验结果表明,本文方法对年龄、人种、性别、面部表情等因素的变化都具备强大的适应能力;与现有单图像三维头部重建算法相比,其在客观指标和主观感受上均取得了最佳效果.

## 2 本文方法

为从单幅图像中重建出高保真度的三维头部模型,本文提出基于法线信息辅助隐式函数的单图像三维头部重建算法.具体而言,将三维头部模型表面看作由隐式函数描述的水平集,并从输入图像及估计的法线贴图中提取视觉特征训练隐式函数.图 2 所示为本文方法的总体框架.本文方法步骤如下.

Step1. 通过一个深度神经网络  $\mathcal{G}$  从输入肖像图像  $I$  估计头部的法线贴图  $\hat{\mathcal{N}}$ .

Step2. 将法线贴图  $\hat{\mathcal{N}}$  与输入图像  $I$  输入编码器  $\mathcal{E}$  提取视觉特征.

Step3. 将三维空间中各点进行坐标映射后选取对应的局部视觉特征,并通过深度隐式函数  $\mathcal{F}(\cdot)$  判别各点位于三维头部模型表面的概率.

Step4. 将位于三维头部模型表面的三维点组成一个等值面,通过 Marching Cubes 算法<sup>[32]</sup>将该等值面转换为网格表示,再经过网格光顺处理后得到最终的三维头部重建结果.

### 2.1 法线贴图估计

法线贴图通过三通道彩色图像像素值表示三维物体表面每个顶点的法线方向,仅消耗少量资源即可直观地描述三维物体的局部细节.因此,在三维头部重建过程中,本文利用从法线贴图中提取的特征来增强局部几何约束并丰富重建结果的表面细节.对于已有三维头部模型,可以通过光栅渲染得到其法线贴图

$$\mathcal{N}_v = \mathcal{R}(\mathbf{M}_p \mathbf{M}_w \mathbf{M}_m \mathbf{V}, \mathbf{N}_v) \quad (1)$$

其中,  $\mathbf{V}$  为三维网格模型的一个顶点;  $v$  为顶点  $\mathbf{V}$  在法线贴图上对应的二维点;  $\mathbf{N}_v$  是在三维模型上计算的顶点  $\mathbf{V}$  的法线方向;  $\mathbf{M}_p$ ,  $\mathbf{M}_w$  和  $\mathbf{M}_m$  分别为透视投影变换、视角变换和归一化矩阵;  $\mathcal{R}(\cdot)$  为光栅渲染函数.然而,三维头部模型是需要重建的对象,需要从训练数据中学习先验知识,以从单幅图像中估计对应三维头部模型的法线贴图.

如图 3 所示,本文设计了一个深度卷积子网络  $\mathcal{G}$  来估计法线贴图.

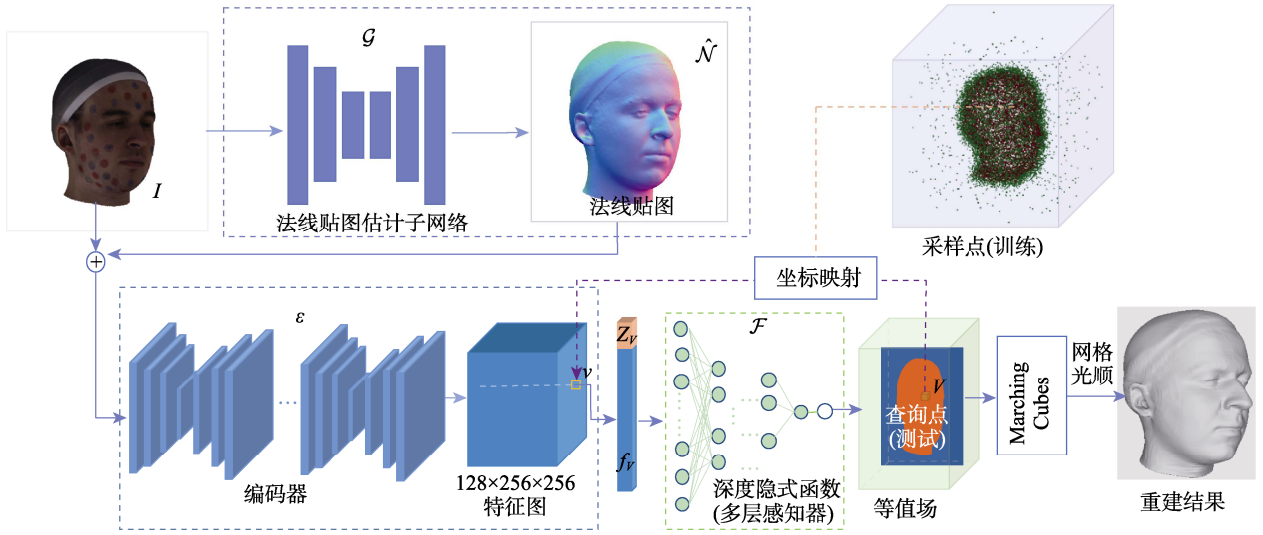


图 2 本文方法总体框架图

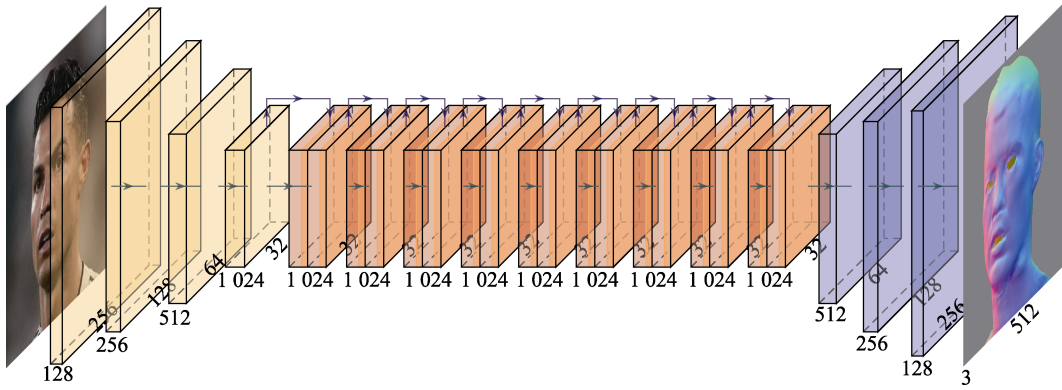


图 3 法线贴图估计子网络结构

法线贴图估计子网络包含 3 个单元: 由 4 个卷积层组成的初级特征提取单元, 由 9 个残差层组成的特征信息增强单元, 以及由 4 个转置卷积层组成的特征图分辨率提升单元. 法线贴图估计子网络采用损失函数  $\mathcal{L}_N = \lambda_N \mathcal{L}_e(\mathcal{N}, \hat{\mathcal{N}}) + \mathcal{L}_p(\mathcal{N}, \hat{\mathcal{N}})$  进行训练. 其中,  $\mathcal{L}_e$  为估计误差损失;  $\mathcal{L}_p$  为感知损失;  $\lambda_N$  为平衡 2 种损失的权重因子;  $\mathcal{N}$  是通过式(1)从标签(ground-truth)三维头部模型渲染的法线贴图;  $\hat{\mathcal{N}}$  为估计的法线贴图. 估计误差损失具体定义为  $\mathcal{L}_e = \|\mathcal{N} - \hat{\mathcal{N}}\|_1$ .

受文献[33]启发, 感知损失  $\mathcal{L}_p$  采用 VGG 网络提取的特征图来描述, 定义为

$$\mathcal{L}_p = \sum_{j=1}^5 \varphi_j \frac{1}{C_j H_j W_j} \|\Phi_j(\mathcal{N}) - \Phi_j(\hat{\mathcal{N}})\|_2^2.$$

其中,  $\Phi_j$  为将法线贴图输入 VGG16 网络<sup>[34]</sup>后第  $j$  个卷积层输出的特征图;  $\varphi_j$  为赋予第  $j$  层特征图的权重;  $C_j$ ,  $W_j$  和  $H_j$  分别为特征图的通道数、宽

度和高度. 仅使用估计误差损失会导致头部高频细微法线部分(如皱纹)平滑和模糊<sup>[35]</sup>, 而使用感知损失有利于提高对这些高频细微法线特征的捕获能力, 进而帮助重建的三维头部模型具备更多的表面细节. 在图 4 所示的例子中, 仅采用估计误差损失训练的法线贴图估计子网, 估计的法线贴图缺少了额头、嘴唇、面颊等区域的褶皱, 而增加了感知损失之后预测的法线贴图具备更多的细节.

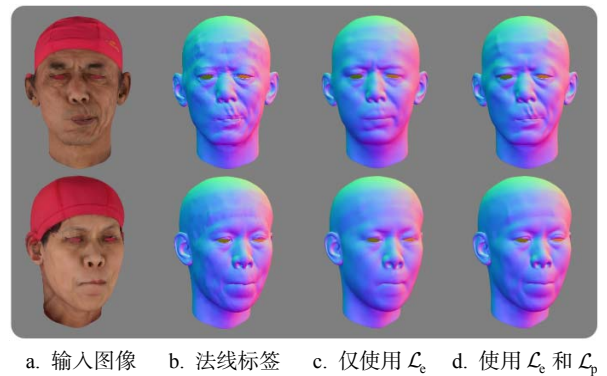


图 4 采用不同损失估计的法线贴图

## 2.2 基于隐式函数的三维头部重建

### 2.2.1 隐式函数表示

基于隐式函数<sup>[36-37]</sup>的三维物体重建算法<sup>[38-41]</sup>将三维物体的表面看作一个  $o^*$  水平集, 即  $\{V | \mathcal{F}(V) = o^*\}$ . 其中,  $\mathcal{F}(\cdot)$  称为隐式函数. 换言之, 三维模型表面可由函数隐式描述的点集来表示, 采用隐式函数表示可以以极小的内存消耗代价表示任意复杂拓扑结构的三维物体表面. 通过隐式函数  $\mathcal{F}(\cdot)$  可以判定三维空间中的任意点  $V$  与三维物体表面(即  $o^*$  等值面)的关系, 即

$$\begin{cases} \text{如果 } \mathcal{F}(V) > o^*, \text{ 则 } V \text{ 在物体表面外} \\ \text{如果 } \mathcal{F}(V) = o^*, \text{ 则 } V \text{ 在物体表面上} \\ \text{如果 } \mathcal{F}(V) < o^*, \text{ 则 } V \text{ 在物体表面内} \end{cases}$$

在实现算法时, 将  $\mathcal{F}(\cdot)$  设计为符号距离函数 (sign distance function, SDF), 且取  $o^* = 0$ . 为提高三维头部重建的鲁棒性, 本文方法基于隐式函数进一步描述三维空间中的顶点  $V$  位于三维头部模型表面的概率

$$P_V = \sigma \left( \frac{1}{|\mathcal{F}(V) - o^*| + \epsilon} \right).$$

其中,  $\sigma(\cdot)$  表示 Logistic 函数;  $\epsilon$  为防止除零的极小正值.

### 2.2.2 特征提取及重建

本文基于隐式函数表示构建一个深度神经网络, 以从单幅图像重建三维头部模型. 为此, 从输入图像  $I$  中提取的特征, 将其作为隐式函数  $\mathcal{F}$  的输入. 此外, 三维模型的顶点  $V$  的深度值也将为三维头部重建提供重要信息. 因此, 本文将隐式函数  $\mathcal{F}$  拓展为  $\alpha_V = \mathcal{F}(V | f(I, V) \oplus Z_V)$ . 其中,  $f(I, V)$  表示特征提取函数;  $Z_V$  为  $V$  的  $Z$  坐标值(深度值). 如图 2 所示, 本文方法利用法线贴图估计子网  $\mathcal{G}$  估计法线贴图, 并将其与输入图像联合后输入编码器  $\mathcal{E}$  提取图像的特征, 提取的特征图为多通道二维图像. 为获得三维模型的顶点对应的特征, 本文方法利用弱投影变换将三维模型的顶点  $V$  映射到二维特征图上. 此外, 邻域特征信息将会提高鲁棒性, 为此, 本文方法还从特征图上提取邻域信息. 于是, 将特征提取函数详写为

$$f(I, V) = \mathcal{U}(\mathcal{E}(I \oplus \mathcal{G}(I)), \chi(\pi(V))).$$

其中,  $\pi(\cdot)$  表示弱投影变换;  $\chi(\cdot)$  表示邻域函数; 本文算法采用二维正态分布在特征图上生成邻域点坐标集;  $\mathcal{U}(\cdot, \cdot)$  表示特征穿刺函数, 它从特征图上选取指定坐标集的多通道特征, 并将选取的特征值重塑为一个向量.

本文算法中, 编码器  $\mathcal{E}$  由 4 层堆叠式沙漏网络<sup>[42]</sup>构成, 深度隐式函数  $\mathcal{F}$  采用多层感知器实现. 在  $\mathcal{E}$  的内部, 每层子沙漏网络输出的特征都作为下一层的输入, 并且每层输出的图像特征都参与到总损失计算中. 与只考虑一个输出层的传统监督模式相比, 这种带中间监督的模式有利于提高特征的精度. 对于编码器  $\mathcal{E}$  与深度隐式函数  $\mathcal{F}$  连接成的深度神经网络, 通过损失函数

$$\mathcal{L} = -\frac{1}{|V|} \sum_{V \in V} \lambda g(V) \ln(P_V) + (1 - \lambda)(1 - g(V)) \ln(1 - P_V)$$

进行训练. 其中,  $V$  为采样点集;  $g(V)$  函数计算采样点  $V$  是否位于标签模型的表面之上;  $\lambda$  为权重因子.

网络训练时, 采样点集  $V$  的空间分布对重建模型的精细度和准确度都会产生影响. 如果在三维空间中均匀采样, 落在三维头部模型表面及其附近的采样点比较少, 会导致训练出的深度神经网络偏向于将本来位于模型表面的查询点错误地预测为不在表面上; 如果只在三维头部模型表面及其附近采样, 又会导致过拟合. 为此, 在训练时, 本文采用混合采样策略<sup>[43]</sup>采集采样点. 首先, 在标签三维头部模型表面随机选取  $K$  个采样点, 并以三维正态分布在各个采样点周围再采集  $C$  个采样点; 然后, 在三维空间中均匀地采集  $rK(C+1)$  个采样点, 其中  $r$  为混合采样比例. 测试时, 将标准三维空间离散为  $300 \times 300 \times 300$  的体素块, 并逐个查询各个体素是否位于三维头部模型的表面之上; 最后, 使用 Marching Cubes 算法<sup>[32]</sup>将体素模型网格化.

### 2.3 网格模型光顺

采用 Marching Cubes 算法<sup>[32]</sup>进行网格化后, 网格模型表面仍然会存在颗粒感. 为进一步提高重建三维头部模型的保真度, 本文采用基于曲率的网格光顺算法对网格模型进行优化. 对于三维模型的顶点  $V$ , 通过光顺函数移动其位置, 移动后的顶点位置为

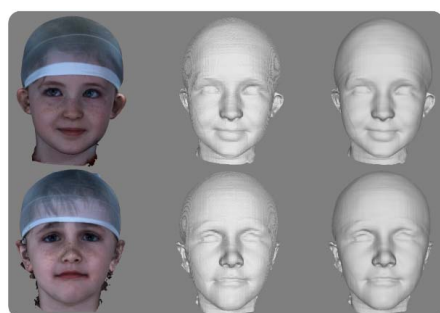
$$\tilde{V} = V + \eta \sum_{k=1}^K \frac{\omega_k}{\sum_{k=1}^K \omega_k} (V_i - V).$$

其中,  $\eta$  用于控制光顺速率;  $\{V_i, i=1, 2, \dots, K\}$  为三维模型的顶点  $V$  在网格模型上所有的邻接顶点;  $\omega_k$  为切角权重, 其计算公式为

$$\omega_k = \frac{1}{2} (\cot \alpha_k + \cot \beta_k).$$

其中,  $\alpha_k$  和  $\beta_k$  是顶点  $V$  与其邻接顶点  $V_k$  所构成的边分别在該边所在的 2 个三角片面中的对角. 本文方法对 Marching Cubes 网格化后的三维头部模

型执行 3 次光顺优化操作. 如图 5 所示, 基于曲率的光顺算法, 使每个顶点移动的位置与其法向量方向相反, 可以在保持模型原有大致几何形状的前提下, 有效地消除网格模型表面的颗粒感.



a. 输入图像 b. 光顺前 c. 光顺后

图 5 采用光顺算法优化后的三维模型

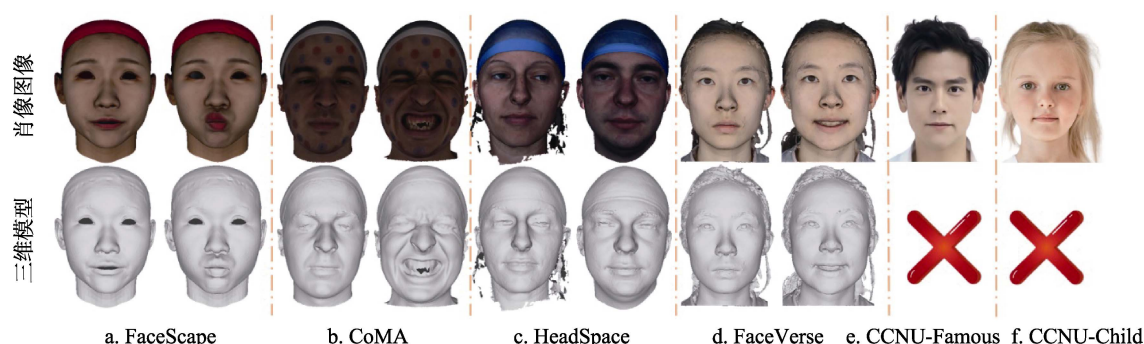


图 6 数据集样本示例

### (2) CoMA 数据集

CoMA 数据集采用连续记录的方式采集了 12 名个体在缓慢变化表情下的三维头部模型. 虽然该数据集共计包含 20 466 个三维头部模型, 但相邻帧模型间的差异性较小. 由于该数据集中的大量头部模型表面存在破洞, 因此, 本文在进行实验前采用补洞算法<sup>[48]</sup>对模型进行了修补.

### (3) HeadSpace 数据集

HeadSpace 数据集使用静态三维扫描系统捕捉了 1 519 名个体在中性表情下的高清三维头部模型. 与 CoMA 数据集类似, 该数据集中的三维头部模型也存在大量破洞.

### (4) FaceVerse 数据集

FaceVerse 数据集使用结构光深度传感器对 18 名个体(每个人表现出 21 种特定的表情)进行采集, 共提供了 378 个三维头部模型. 与 FaceScape, CoMA 和 HeadSpace 刻意约束发型不同, FaceVerse 数据集采集了各个个体的自然发型.

### (5) CCNU-Famous 和 CCNU-Child 数据集

CCNU-Famous 和 CCNU-Child 数据集中的肖像

## 3 实验

### 3.1 数据集

在 FaceScape<sup>[44]</sup>, CoMA<sup>[45]</sup>, HeadSpace<sup>[46]</sup>, FaceVerse<sup>[47]</sup>, 以及从互联网上收集的名人(CCNU-Famous)和儿童(CCNU-Child)数据集上, 将本文与其他算法进行了对比实验. 各数据集的样本示例如图 6 所示, 各数据集的简介如下.

#### (1) FaceScape 数据集

FaceScape 数据集采集了年龄分布为 16~70 岁的 938 名个体, 其中每名个体表现 20 种特定的表情, 因此该数据集共包含 18 760 个三维头部模型. 受限于原始数据采集设备, FaceScape 数据集中的三维头部模型分辨率较低.

图像从互联网收集, 收集时考虑了人种、肤色、性别、面部表情和光照等因素的多样性. CCNU-Famous 中包含 1 900 幅成年人图像, CCNU-Child 包含 780 幅儿童图像. 由于 CCNU-Famous 和 CCNU-Child 数据集中不包含三维头部模型, 因此本文仅将它们用于定性实验.

为了获得大规模且具备多样性的训练数据, 实验时将 FaceScape, CoMA 和 HeadSpace 合并为一个混合数据集(记为 FCH), 并将其按个体随机划分为训练集和测试集. 具体而言, FCH 训练集包含 8 828 个三维头部模型, FCH 测试集包含 249 个三维头部模型. 为进一步扩展训练数据的数量, 将每个三维头部模型在偏航角为 $[-50^{\circ} \sim +50^{\circ}]$ 的区间内每间隔  $10^{\circ}$  渲染一幅肖像图像, 共得到 88 280 幅图像, 并将所有渲染图像与其对应的三维头部模型组成 88 280 对〈肖像图像, 三维头部模型〉作为训练样本集. 同时, 将 FCH 测试集中的三维头部模型以随机角度渲染为 2 490 幅测试图像. 除此以外, 还在 FaceVerse 及 CCNU-Famous 和 CCNU-Child 进行了跨数据集实验. 同样, 将 FaceVerse 数据集

中的 378 个三维头部模型按随机视角渲染得到 3 780 幅测试图像。

### 3.2 实现细节

本文算法采用两阶段训练法, 即第 1 阶段训练法线信息估计模块, 第 2 阶段训练隐式等值面估计器模块。2 个阶段的训练均采用 RMSprop<sup>[49]</sup>作为优化器。在第 1 阶段训练时, 学习率固定为 0.000 1, 总计训练 10 个 epoch。在第 2 阶段训练时, 采样点总数设置为 10 000, 混合采样比例设置为 0.062 5; batch-size 设置为 3, 学习率的初始值设置为 0.001, 并在 epoch 为 2, 4, 6 时以速率降低到 10%进行衰减, 共训练 8

个 epoch。此外, 本文采用组归一化(GroupNormlization)<sup>[50]</sup>代替批归一化(BatchNormalization)<sup>[51]</sup>, 以实现小批量数据训练的稳定性。本文算法采用 Python 语言并基于 PyTorch 框架实现。在一台配置了单张 NVIDIA GeForce RTX 3080 GPU 的计算机上, 训练本文算法总计耗时 27 d, 其中第 1 阶段和第 2 阶段训练分别耗时 11 d 和 16 d。测试时, 本文算法从单幅 RGB 图像生成三维头部模型的平均耗时为 5.17 s。

### 3.3 本文算法的结果

图 7 所示为本文算法在实验所用到的 6 个数据集上的三维头部重建结果。

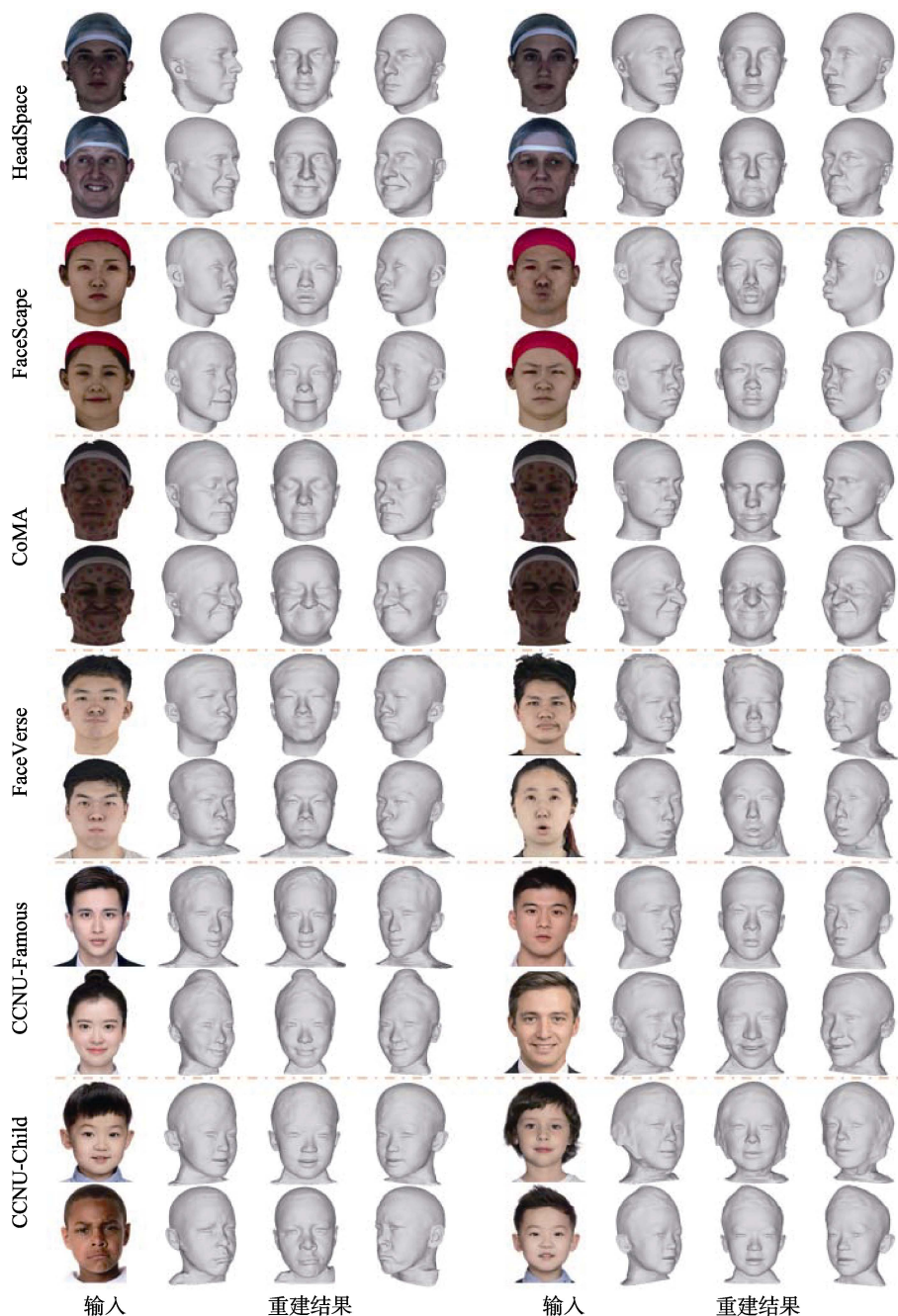


图 7 本文算法在 6 个数据集上不同视角的三维头部重建结果

由图 7 可见, 对于不同人种、性别、年龄和面部表情的输入肖像, 本文算法在所有的数据集上均能重建出从各个角度观察都逼真的三维头部模型。值得注意的是, 训练所用到的数据集中的个体均进行了束发, 但本文算法仍然可以较为精确地重建出测试图像中人物的自然发型。

### 3.4 定量评价指标

本文使用倒角距离(chamfer distance, CD)和点曲面距离(point surface distance, PSD)作为各种三维重建算法的定量评价指标, 其计算公式分别为

$$\varepsilon_{CD}(S_1, S_2) = \frac{1}{S_1} \sum_{m \in S_1} \min_{n \in S_2} \|m - n\|_2^2,$$

$$\varepsilon_{PSD}(S_1, S_2) = \frac{1}{S_1} \sum_{f \in S_1} \min_{p \in S_2} \|f - p\|_2^2.$$

其中,  $S_1$  和  $S_2$  分别为标签模型和重建模型上的顶点集。由于实验所采用的几个数据集单位和尺度并不完全一致, 本文将各个数据集中的模型统一采用 mm 作为单位, 并归一化到瞳距为 65 mm。此外, 由于部分参与比较的算法只能重建出面部, 为了能顺利、公平地进行比较, 本文以三维头部模型鼻尖为中心进行半径为 95 mm 的球形裁剪得到面部区域, 并只在裁剪后的区域上计算客观评价指标<sup>[21]</sup>。由于本文算法能比其他算法更好地重建头发、耳朵等区域, 如果直接在整个头部模型上计算客观评价指标, 本文算法相对于其他算法将具备

更大的优势。在计算定量评价指标时, 同样将重建的三维头部模型进行上述的归一化和裁剪处理, 并利用 ICP<sup>[52]</sup>算法将其与标签模型对齐。

### 3.5 消融实验

为了证实法线贴图有助于加强三维头部重建时的局部几何约束并增强表面细节, 本文去除法线贴图估计子网络后, 重新训练了网络并进行消融实验。在分别使用(w)和去除(w/o)法线贴图估计子网络时, 本文算法在 FCH 和 FaceVerse 数据集上的平均 CD 和 PSD 如表 1 所示。

表 1 使用和去除法线贴图估计子网络时在 FCH 和 FaceVerse 数据集上的 CD 和 PSD mm

| 是否使用<br>法线贴图 | CD      |           | PSD     |           |
|--------------|---------|-----------|---------|-----------|
|              | FCH     | FaceVerse | FCH     | FaceVerse |
| w/o          | 1.010 3 | 1.635 3   | 1.418 5 | 3.235 0   |
| w            | 0.769 6 | 1.308 0   | 0.804 7 | 2.677 0   |

在去除法线贴图子网络之后, 本文算法在 2 个数据集上的平均 CD 分别增加了 0.240 7 mm 和 0.327 3 mm; 其使用(w)和去除(w/o)法线贴图估计子网络时, 在 CCNU-Famous 和 CCNU-Child 数据集上的三维头部重建结果如图 8 所示。可见, 在不使用法线信息时, 重建结果上的表面细节减少(尤其嘴部、脸颊等区域), 局部几何结构也会出现错误。

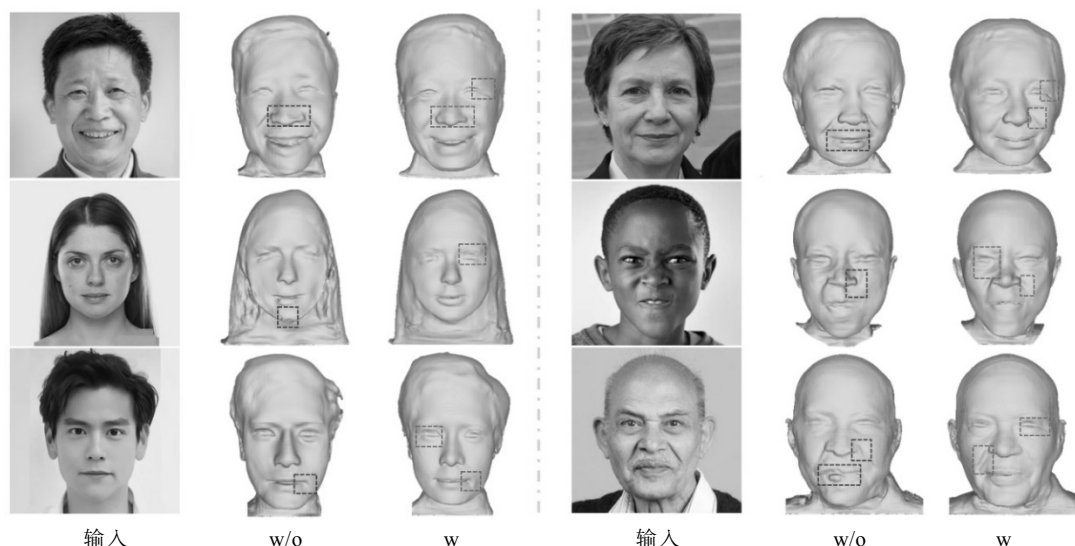


图 8 使用和去除法线贴图估计子网络时在 CCNU-Famous 和 CCNU-Child 数据集上的重建结果

### 3.6 与其他算法的比较

在 FCH 测试集, 以及 FaceVerse, CCNU-Famous 和 CCNU-Child 数据集上, 将本文算法与 Deep3D<sup>[15]</sup>, MGCNet<sup>[16]</sup>, Hifi3dface<sup>[17]</sup>, DECA<sup>[18]</sup>, PRNet<sup>[24]</sup> 和

3DDFA\_V2<sup>[29]</sup> 6 种三维头部/面部重建算法进行对比实验。其中, DECA, Hifi3dface, MGCNet, 3DDFA\_V2 和 Deep3D 属于参数化方法, PRNet 为非参数化方法。对比的 6 种算法都开源了代码, 为

了比较的公平性,实验时在 FCH 训练集上对各算法采用作者提供的代码和最佳参数进行了重新训练.由于所有算法均未在 FaceVerse, CCNU-Famous, CCNU-Child 上进行过训练,在这 3 个数据集上的实验能够充分地检验各算法的跨数据集推广性能.由于添加纹理信息后容易掩饰重建结果中的几何结构误差,因此,在展示结果时本文隐去了所有重建结果的纹理,只保留几何形状.

表 2 所示为不同算法在 FCH 测试集上的 CD 与 PSD 定量评价指标.本文算法在 FCH 测试集上的平均 CD 和 PSD 分别为 0.7696 mm 和 0.8047 mm,优于所有参与比较的算法.

表 2 各种算法在 FCH 数据集上的定量比较 mm

| 算法                         | CD             | PSD            |
|----------------------------|----------------|----------------|
| DECA <sup>[18]</sup>       | 1.741 3        | 4.003 8        |
| Hifi3dface <sup>[17]</sup> | 2.510 3        | 5.915 3        |
| MGCNet <sup>[16]</sup>     | 1.688 3        | 3.879 2        |
| 3DDFA_V2 <sup>[29]</sup>   | 1.751 8        | 4.069 2        |
| Deep3D <sup>[15]</sup>     | 1.662 5        | 3.761 1        |
| PRNet <sup>[24]</sup>      | 1.740 2        | 4.041 0        |
| 本文                         | <b>0.769 6</b> | <b>0.804 7</b> |

注:粗体表示最优值.

图 9 所示为各算法在 FCH 数据集上的重建结果.为了分析重建误差,将重建模型上的各个顶点与标签模型顶点的 CD 以热力图的方式予以展示.



图 9 各种三维头部重建算法在 FCH 数据集上的重建结果及误差热力图

不难看出:(1) 本文算法重建的三维头部模型的误差明显低于所有参与比较的 6 种算法.(2) 本文算法重建的三维头部模型的少量误差主要出现在对重建模型的逼真性影响较小的发际线边缘和

下巴区域;而对比算法的重建误差大多分布在对重建模型的身份、性别、表情、年龄等属性影响较大的鼻子、眼睛、嘴巴和脸颊等部位.通过误差在人脸部位的分布差异,能够很好地解释本文算法

与其他算法重建的三维头部模型在表 2 中的 CD 相差不大(例如, 本文算法与 Deep3D 输出的重建模型的平均 CD 差值只有 0.8929 mm), 而在视觉感知上的逼真度却差异甚大的原因。

表 3 展示了不同算法在 FaceVerse 数据集上的定量评价指标。本文算法所重建的三维头部模型与标签模型之间的 CD 和 PSD 的平均值分别为 1.3080 mm 和 2.6770 mm, 仍然优于对比算法。

图 10 所示为各种算法在 FaceVerse 数据集上的重建结果和误差热力图。在视觉感受上的逼真度方面, 本文算法重建的三维头部模型大幅优于

其他算法的结果; 在定量误差方面, 本文算法的误差明显低于其他算法。

表 3 各种算法在 FaceVerse 数据集上的定量比较 mm

| 算法                         | CD/mm          | PSD/mm         |
|----------------------------|----------------|----------------|
| DECA <sup>[18]</sup>       | 1.961 8        | 5.084 6        |
| Hifi3dface <sup>[17]</sup> | 2.135 1        | 5.443 6        |
| MGCNet <sup>[16]</sup>     | 1.857 8        | 4.869 0        |
| 3DDFA_V2 <sup>[29]</sup>   | 2.065 6        | 5.807 5        |
| Deep3D <sup>[15]</sup>     | 1.682 3        | 4.018 7        |
| PRNet <sup>[24]</sup>      | 2.136 7        | 6.250 4        |
| 本文                         | <b>1.308 0</b> | <b>2.677 0</b> |

注: 粗体表示最优值。

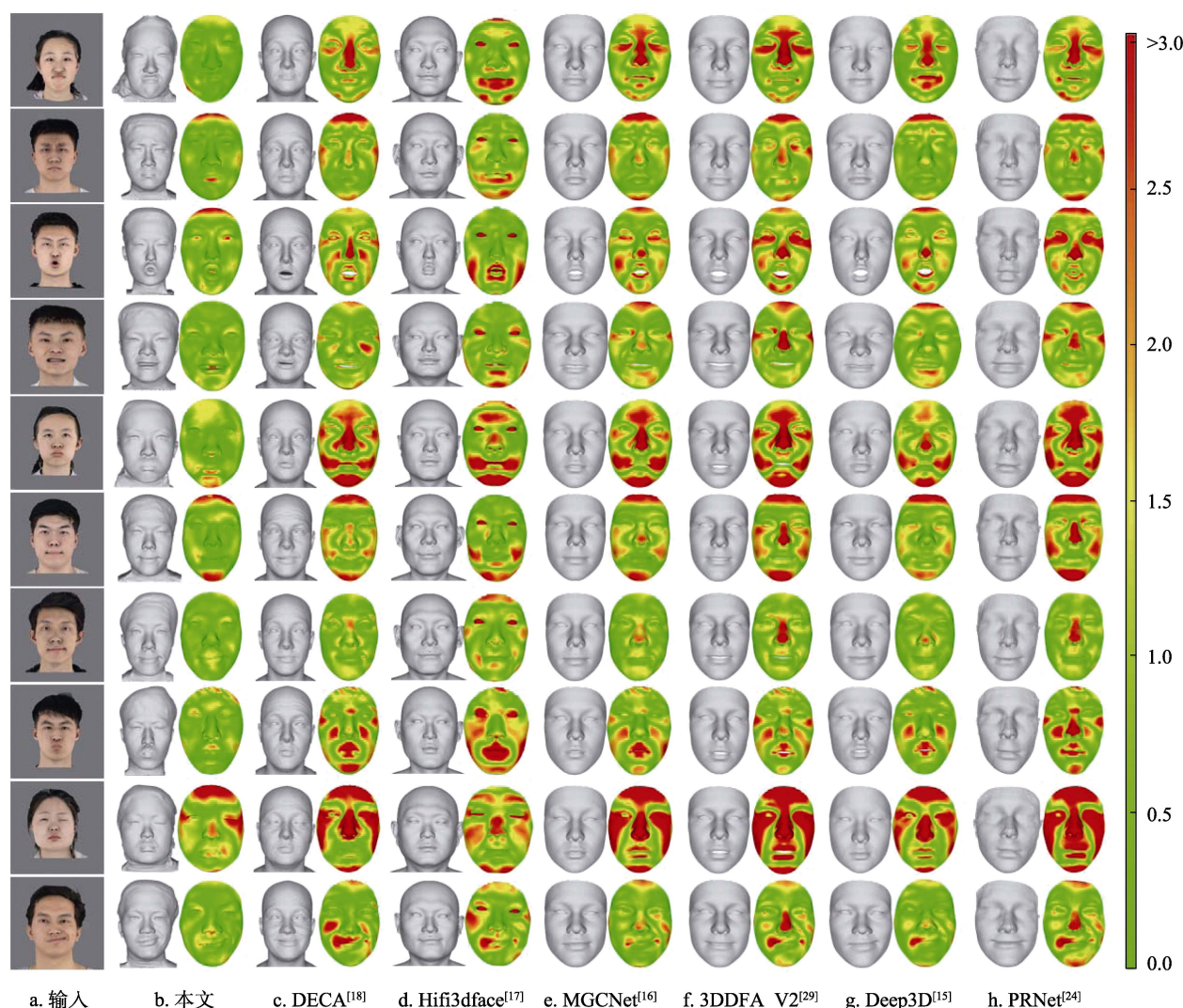


图 10 各种三维头部重建算法在 FaceVerse 数据集上的重建结果及误差热力图

图 11 所示为不同算法在 CCNU-Famous 数据集上的三维头部重建结果。该数据集中的肖像图像均为从互联网收集的“野图(in-the-wild images)”。在视觉感知上, 本文算法在野图上重建的三维头部仍然具备优良的逼真度, 重建质量与在 FCH 和 FaceVerse 数据集上并无明显差异。反观其他算法,

在隐去了纹理信息后, 重建的三维头部/脸部模型大多趋向于平均脸, 且几乎无法从重建模型中辨识出身份和性别等信息, 大多数算法也没有恢复出输入肖像的逼真表情。这说明本文算法在野图上的三维头部重建结果大幅优于参与比较的其他算法, 具备优异的推广性能。

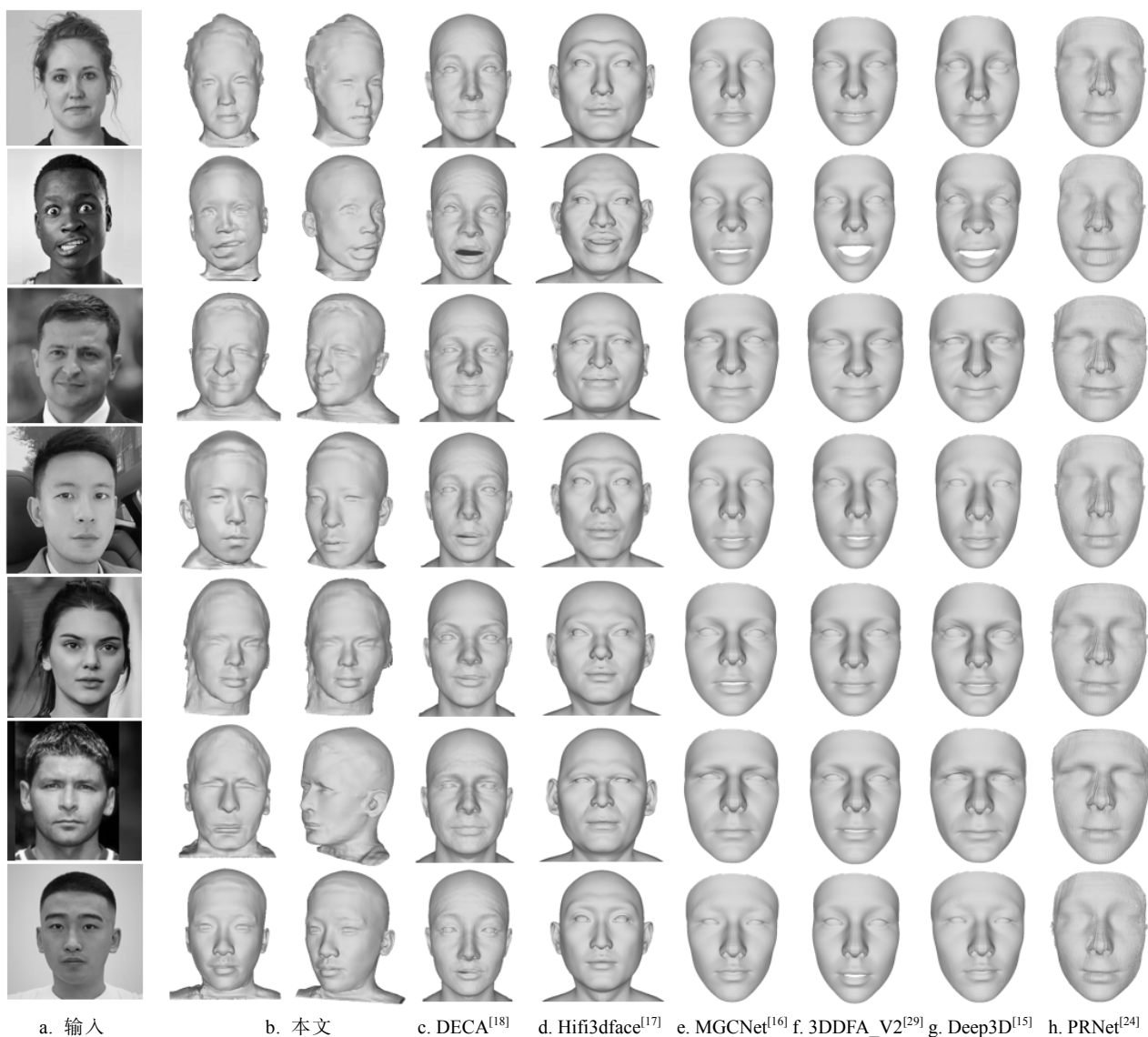


图 11 各种三维头部重建算法在 CCNU-Famous 数据集上的重建结果

图 12 展示了不同算法在 CCNU-Child 数据集上的三维头部重建结果。众所周知，儿童与成年人在五官比例等面部特征上存在明显的差异，因而将在成年人数据集上训练的三维头部重建模型直接泛化到儿童肖像上存在挑战。虽然本文算法从未在任何包含儿童的数据集上训练过，但采用该算法从儿童肖像图像重建出的三维头部模型非常逼真，身份、年龄、性别、表情和发型都可从重建模型上一目了然地感知出来。而对比的其他算法对儿童肖像的重建结果看起来仍然像成年人，其身份、年龄、性别和表情等属性几乎无法与输入图像中的儿童肖像形成对应。可见，本文算法针对儿童的三维头部重建结果大幅优于所有参与比较的算法，具备强大的泛化能力。

## 4 结 语

本文提出了一种基于法线贴图增强隐式函数的单图像三维头部重建算法。本文算法融合了隐式函数对三维物体表面强大的表达能力以及法线贴图的局部几何结构描述能力，有效地解决了现有算法重建的三维头部模型保真度低、细节少和泛化能力差的问题。在公共数据和从互联网采集的图像的实验结果表明，本文算法能从包含不同年龄、人种、肤色、性别、面部表情的单幅肖像图像中重建具备高保真度和丰富细节的三维头部模型。

本文的工作仍存在一些局限：(1) 虽然本文算法能重建出人物的大致发型，但由于隐式函数不适合描述头发的拓扑结构，因此其重建的头发精

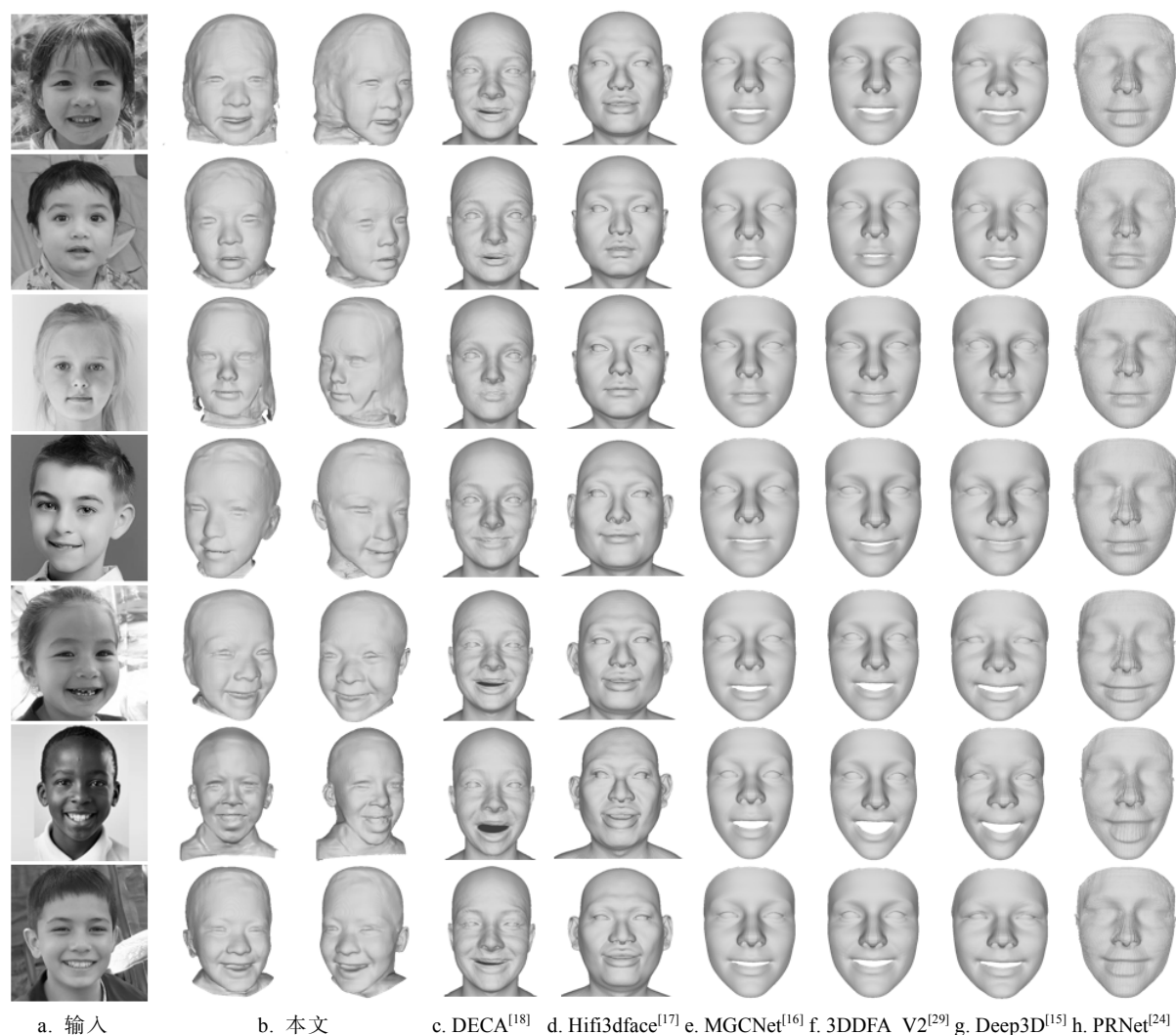


图 12 各种三维头部重建算法在 CCNU-Child 数据集上的重建结果

度欠佳. 在未来的工作中, 将考虑采用“管线”模型来描述头发, 进一步提升本文算法对发型重建的精细度. (2) 相对于参数化算法, 虽然本文算法重建的三维模型在保真度、细节以及泛化能力算法具备巨大优势, 但由于其顶点数和顶点顺序不固定, 难以像参数化模型那样便捷、灵活地驱动本文算法重建的三维头部模型. 在未来的工作中, 将考虑采用语义信息来建立本文重建模型与参数化模型顶点之间映射关系, 以通过参数化模型(如 FLAME<sup>[27]</sup>)来驱动本文算法重建的三维头部模型.

## 参考文献(References):

- [1] Zhao Y H, Jiang J J, Chen Y, *et al.* Metaverse: perspectives from graphics, interactions and visualization[J]. Visual Informatics, 2022, 6(1): 56-67
- [2] Song S L, Shi W Q, Reed M. Accurate face rig approximation with deep differential subspace reconstruction[J]. ACM Transactions on Graphics, 2020, 39(4): Article No.34
- [3] Orts-Escalano S, Rhemann C, Fanello S, *et al.* Holoportation: virtual 3D teleportation in real-time[C] //Proceedings of the 29th Annual Symposium on User Interface Software and Technology. New York: ACM Press, 2016: 741-754
- [4] Reisoğlu I, Topu B, Yılmaz R, *et al.* 3D virtual learning environments in education: a meta-review[J]. Asia Pacific Education Review, 2017, 18(1): 81-100
- [5] Zhang T, Liu J H, Liu S L, *et al.* A 3D reconstruction method for pipeline inspection based on multi-vision[J]. Measurement, 2017, 98: 35-48
- [6] Zheng Taixiong, Huang Shuai, Li Yongfu, *et al.* Key techniques for vision based 3D reconstruction: a review[J]. Acta Automatica Sinica, 2020, 46(4): 631-652(in Chinese)  
(郑太雄, 黄帅, 李永福, 等. 基于视觉的三维重建关键技术研究综述[J]. 自动化学报, 2020, 46(4): 631-652)
- [7] Chen Jia, Zhang Yuqi, Song Peng, *et al.* Application of deep learning to 3D object reconstruction from a single image[J]. Acta Automatica Sinica, 2019, 45(4): 657-668(in Chinese)  
(陈加, 张玉麒, 宋鹏, 等. 深度学习在基于单幅图像的物体三维重建中的应用[J]. 自动化学报, 2019, 45(4): 657-668)
- [8] Bas A, Smith W A P. What does 2D geometric information re-

- ally tell us about 3D face shape?[J]. *International Journal of Computer Vision*, 2019, 127(10): 1455-1473
- [9] Yenamandra T, Tewari A, Bernard F, *et al.* i3DMM: deep implicit 3D morphable model of human heads[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2021: 12798-12808
  - [10] Zielonka W, Bolkart T, Thies J. Towards metrical reconstruction of human faces[C] // *Proceedings of the 17th European Conference on Computer Vision*. Heidelberg: Springer, 2022: 250-269
  - [11] Wood E, Baltrušaitis T, Hewitt C, *et al.* 3D face reconstruction with dense landmarks[C] // *Proceedings of the 17th European Conference on Computer Vision*. Heidelberg: Springer, 2022: 160-177
  - [12] Sanyal S, Bolkart T, Feng H W, *et al.* Learning to regress 3D face shape and expression from an image without 3D supervision[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2019: 7755-7764
  - [13] He H Y, Li G Q, Ye Z H, *et al.* Data-driven 3D human head reconstruction[J]. *Computers & Graphics*, 2019, 80: 85-96
  - [14] Jiang Z H, Wu Q Y, Chen K Y, *et al.* Disentangled representation learning for 3D face shape[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2019: 11949-11958
  - [15] Deng Y, Yang J L, Xu S C, *et al.* Accurate 3D face reconstruction with weakly-supervised learning: from single image to image set[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Los Alamitos: IEEE Computer Society Press, 2019: 285-295
  - [16] Shang J X, Shen T W, Li S W, *et al.* Self-supervised monocular 3D face reconstruction by occlusion-aware multi-view geometry consistency[C] // *Proceedings of the 16th European Conference on Computer Vision*. Heidelberg: Springer, 2020: 53-70
  - [17] Bao L C, Lin X K, Chen Y J, *et al.* High-fidelity 3D digital human head creation from RGB-D selfies[J]. *ACM Transactions on Graphics*, 2021, 41(1): Article No.3
  - [18] Feng Y, Feng H W, Black M J, *et al.* Learning an animatable detailed 3D face model from in-the-wild images[J]. *ACM Transactions on Graphics*, 2021, 40(4): Article No.88
  - [19] Burkov E, Rakhimov R, Safin A, *et al.* Multi-NeuS: 3D head portraits from single image with neural implicit functions[J]. *IEEE Access*, 2023, 11: 95681-95691
  - [20] Wang X Y, Guo Y D, Yang Z Q, *et al.* Prior-guided multi-view 3D head reconstruction[J]. *IEEE Transactions on Multimedia*, 2022, 24: 4028-4040
  - [21] Ramon E, Triginer G, Ecur J, *et al.* H3D-Net: few-shot high-fidelity 3D head reconstruction[C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2021: 5600-5609
  - [22] Cheng S Y, Tzimiropoulos G, Shen J, *et al.* Faster, better and more detailed: 3D face reconstruction with graph convolutional networks[C] // *Proceedings of the 15th Asian Conference on Computer Vision*. Heidelberg: Springer, 2020: 188-205
  - [23] Jackson A S, Bulat A, Argyriou V, *et al.* Large pose 3D face reconstruction from a single image via direct volumetric CNN regression[C] // *Proceedings of the IEEE International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2017: 1031-1039
  - [24] Feng Y, Wu F, Shao X H, *et al.* Joint 3D face reconstruction and dense alignment with position map regression network[C] // *Proceedings of the 15th European Conference on Computer Vision*. Heidelberg: Springer, 2018: 557-574
  - [25] Lin J K, Yuan Y, Shao T J, *et al.* Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2020: 5890-5899
  - [26] 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*. New York: ACM Press/Addison-Wesley Publishing Co., 1999: 187-194
  - [27] Li T Y, 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): Article No.194
  - [28] Ploumpis S, Ververas E, O'Sullivan E, *et al.* Towards a complete 3D morphable model of the human head[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(11): 4142-4160
  - [29] Guo J Z, Zhu X Y, Yang Y, *et al.* Towards fast, accurate and stable 3D dense face alignment[C] // *Proceedings of the 16th European Conference on Computer Vision*. Heidelberg: Springer, 2020: 152-168
  - [30] Liu S C, Saito S, Chen W K, *et al.* Learning to infer implicit surfaces without 3D supervision[C] // *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. New York: ACM Press, 2019: 8295-8306
  - [31] Li Q M, Han Z C, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning[C] // *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI Press, 2018: 3538-3545
  - [32] Lorensenwilliam E, Clineharvey E. Marching cubes: a high resolution 3D surface construction algorithm[C] // *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*. New York: ACM Press, 1987: 163-169
  - [33] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C] // *Proceedings of 14th European Conference on Computer Vision*. Heidelberg: Springer, 2016: 694-711
  - [34] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[OL]. [2023-02-27]. <https://arxiv.org/abs/1409.1556>
  - [35] Xiu Y L, Yang J L, Tzionas D, *et al.* Icon: implicit clothed humans obtained from normals[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2022: 13286-13296
  - [36] Mescheder L, Oechsle M, Niemeyer M, *et al.* Occupancy networks: learning 3D reconstruction in function space[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2019: 4455-4465

- [37] Chen Z Q, Zhang H. Learning implicit fields for generative shape modeling[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 5932-5941
- [38] Duggal S, Pathak D. Topologically-aware deformation fields for single-view 3D reconstruction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 1526-1536
- [39] Park J J, Florence P, Straub J, *et al.* DeepSDF: learning continuous signed distance functions for shape representation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 165-174
- [40] Cai H R, Feng W Q, Feng X T, *et al.* Neural surface reconstruction of dynamic scenes with monocular RGB-D camera[OL].[2023-02-27]. <https://arxiv.org/abs/2206.15258>
- [41] Li Lei, Xu Hao, Wu Suping. Fuzzy probability points reasoning for 3D reconstruction via deep deterministic policy gradient[J]. *Acta Automatica Sinica*, 2022, 48(4): 1105-1118(in Chinese)  
(李雷, 徐浩, 吴素萍. 基于DDPG的三维重建模糊概率点推理[J]. *自动化学报*, 2022, 48(4): 1105-1118)
- [42] Newell A, Yang K Y, Deng J. Stacked hourglass networks for human pose estimation[C] //Proceedings of the 14th European Conference on Computer Vision. Heidelberg: Springer, 2016: 483-499
- [43] 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. Los Alamitos: IEEE Computer Society Press, 2019: 2304-2314
- [44] Yang H T, Zhu H, Wang Y R, *et al.* FaceScape: a large-scale high quality 3D face dataset and detailed riggable 3D face prediction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 598-607
- [45] Ranjan A, Bolkart T, Sanyal S, *et al.* Generating 3D faces using convolutional mesh autoencoders[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 725-741
- [46] Dai H, Pears N, Smith W, *et al.* A 3D morphable model of craniofacial shape and texture variation[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 3104-3112
- [47] Wang L Z, Chen Z Y, Yu T, *et al.* FaceVerse: a fine-grained and detail-controllable 3D face morphable model from a hybrid dataset[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 20301-20310
- [48] Zhao W, Gao S M, Lin H W. A robust hole-filling algorithm for triangular mesh[J]. *The Visual Computer*, 2007, 23(12): 987-997
- [49] Tieleman T, Hinton G. Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude[J]. *COURSERA: Neural Networks for Machine Learning*, 2012, 4(2): 26-31
- [50] Wu Y X, He K M. Group normalization[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 3-19
- [51] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C] //Proceedings of the 32nd International Conference on Machine Learning. New York: PMLR Press, 2015: 448-456
- [52] Besl P J, McKay N D. Method for registration of 3-D shapes[C] //Proceedings of the SPIE 1611, Sensor fusion IV: Control Paradigms and Data Structures. Bellingham: Society of Photo-Optical Instrumentation Engineers, 1992: 586-606