基于特征核相关的单张图像三维人脸重建方法

王阿龙,戴家树,倪斌帆,曹 韬
安徽工程大学 计算机与信息学院,安徽 芜湖 241000

摘 要:目的 解决现有基于弱监督学习的单张图像三维人脸重建方法聚焦于全局监督信息,缺少对局部特征的约束,导致人脸局部区域重建效果较差的问题。**方法** 提出一种基于特征核相关的单张图像三维人脸重建方法,在由低层级感知损失、中层级人脸关键点损失、高层级身份损失和形状一致性损失等构成的全局多层级损失函数基础上,构造特征核相关损失函数,对局部进行深层次约束,旨在提高三维人脸的重建精度。**结果** 将所提损失函数用于对嘴部形状的约束,在 REALY 公开基准上进行定量对比实验,整体重建的误差均值为 2.102 mm,嘴部重建误差为 2.206 mm;在 FFHQ 公开数据集上进行定性对比实验;最后在 FaceScape 数据集和 LYHM 数据集上进行消融实验和可视化展示,以验证所提损失函数的有效性和模型鲁棒性。**结论** 实验表明:在不同视角中,所提的重建方法整体和局部重建效果均优于其他方法,面对不同姿态、表情和光照环境,所提模型均有较好的性能表现。

关键词:三维人脸重建;特征核相关;弱监督学习;卷积神经网络;3D 形变模型
中图分类号:TP391.41 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2026.0004.003

3D Face Reconstruction from a Single Image Based on Feature Kernel Correlation

WANG Along, DAI Jiashu, NI Binfan, CAO Tao
School of Computer and Information, Anhui Polytechnic University, Wuhu 241000, Anhui, China

Abstract: Objective This paper aims to address the issue that existing single-image 3D face reconstruction methods based on weakly supervised learning focus on global supervision information and lack constraints on local features, which leads to poor reconstruction results in local facial regions. **Methods** A single-image 3D face reconstruction method based on feature kernel correlation is proposed. On the basis of a global multi-level loss function composed of low-level perceptual loss, mid-level facial key-point loss, high-level identity loss, and shape consistency loss, a feature kernel correlation loss function was constructed to better constrain local parts, aiming to improve the reconstruction accuracy of 3D faces. **Results** The proposed loss function was used to constrain the shape of the mouth. Quantitative comparison experiments were conducted on the REALY public benchmark, with the mean error of the overall reconstruction being 2.102 mm and the mouth reconstruction error being 2.206 mm. Qualitative comparison experiments were carried out on the FFHQ public dataset. Finally, ablation experiments and visual demonstrations were performed on the FaceScape dataset and the LYHM dataset to verify the effectiveness of the proposed loss function and the robustness of the model. **Conclusion** Experimental results demonstrate that the proposed reconstruction method outperforms other methods from different viewpoints in both

收稿日期:2024-04-08 **修回日期:**2024-08-18 **文章编号:**1672-058X(2026)04-0018-10
基金项目:安徽省教育厅高校科学研究重点项目资助(KZ22020130);产学研合作项目资助(KH10002119).
作者简介:王阿龙(2000—),男,安徽霍邱人,硕士研究生,从事三维人脸重建研究.
通信作者:戴家树(1987—),男,安徽和县人,博士,副教授,从事机器视觉、智能控制相关研究. Email:daijiashudai@ahpu.edu.cn.
引用格式:王阿龙,戴家树,倪斌帆,等. 基于特征核相关的单张图像三维人脸重建方法[J]. 重庆工商大学学报(自然科学版),2026,43(4):18-27.
Wang Along, Dai Jiashu, Ni Binfan, et al. 3D face reconstruction from a single image based on feature kernel correlation[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(4): 18-27.

overall and local reconstruction. Moreover, the proposed model also exhibits outstanding performance in various poses, expressions, and lighting environments.

Keywords: 3D face reconstruction; feature kernel correlation; weakly supervised learning; convolutional neural network; 3D morphable model

单张图像的三维人脸重建技术被广泛应用于人脸识别、虚拟现实和医学美容等领域,是计算机视觉和人工智能领域的一项关键任务。传统重建方法主要分为两种,第一种方法通常对人脸图像中的光照、纹理等信息进行建模,然后通过数学算法推断出人脸的三维结构,该类方法在处理复杂场景时往往受到遮挡、光照等因素影响,重建结果不准确;另一种方法是借助三维人脸形变模型(3D Morphable Model, 3DMM)[1]将人脸表示为一个低维的参数空间,通过迭代优化这些参数来拟合给定的图像,然而,这种方法在处理复杂表情和姿态变化时受限,且需要大量的计算资源和时间进行优化。

近年来,随着深度学习的发展,基于神经网络的单张图像三维人脸重建方法被广泛提出,在三维人脸重建任务上取得了突破性进展。一些三维人脸重建方法[2-3]通过回归出 UV 空间中的位置图来表示三维人脸模型,从而降低了任务复杂度并提升了运行效率。但该类方法重建模型的投影结果受拍摄距离制约,且存在局部纹理拉伸、轮廓形状模糊等问题。另一些方法[4-8]通过卷积神经网络(Convolutional Neural Network, CNN)估计人脸图像对应的 3DMM 系数,将复杂的重建任务转换为 3DMM 系数回归任务,显著提高了重建的精度和效率,并且重建出的模型更加完整。基于有监督学习的方法需要大规模的标注数据进行训练,然而,这些 3D 标注的数据集制作成本较高且格式不一致,难以满足网络训练需要。一些方法使用合成人脸数据集进行训练,但缺乏多样性,导致训练后模型泛化性能较差,难以用于真实环境。

一些研究[6-7,9-13]使用弱监督或自监督学习方法来应对真实三维数据集采集和制作的困难。基于弱监督的方法仅需利用人脸关键点和面部掩膜图作为监督信号。该方法的核心是通过设计的神经网络,重建出人脸模型,经可微分渲染器投影至平面后,计算其与输入图像之间的损失,通常对像素颜色[14]、身份[15-16]和轮廓[17-18]等差异建立损失函数,最后,优化器对网络模型参数进行优化迭代,完成模型的训练。

然而,现有的弱监督重建方法在设计损失函数时,聚焦于对全局特征进行约束,而较少关注局部特征。它通常简单地约束网络模型学习局部轮廓形状,如计算各部位关键点加权后的 L_1 或 L_2 损失,因此导致模型难以高效地学习局部形状细节信息。为了缓解这一问题,本文创新地将核相关(Kernel Correlation, KC)[19]引入计算局部投影后关键点与真实关键点之间的相似度,并设计一个新的损失函数——特征核相关损失函数。由于嘴巴部位在表情中扮演着重要的角色,可以传达情绪和意图,相较于其他部位,嘴部的形状、张合程度以及嘴唇的动作灵活多样,因此,重建该部位具有挑战性。本文将所提出的损失函数应用于对嘴巴部位的约束,旨在促使模型高效地学习嘴部形状和纹理信息,提升重建精度。

在 FFHQ 数据集[20]和 REALY 基准[21]上分别进行定性和定量对比实验。实验表明,在不影响整体重建精度的前提下,嘴部重建质量相较于其他方法有更好的表现。在 FaceScape 数据集[22]上进行可视化展示,验证了模型对同一个体不同表情和姿态下的鲁棒性。通过消融实验,验证了本文提出的特征核相关损失函数的有效性。

1　相关工作

在三维人脸重建领域的早期发展阶段,为了解决数据不足导致的重建效果不佳问题,研究人员进行了一系列工作来改进构建人脸基向量的质量、种类和数量,以建立更好的先验统计 3DMM。其中,Paysan 等[23]采集 200 组三维人脸扫描数据,用光流算法(Optical Flow, OF)[24]稠密对应采集数据,构建了巴塞尔人脸模型(Basel Face Model, BFM),保证了人脸顶点一一对应且三维人脸网格拓扑结构一致。然而,由于缺乏表情参数,无法重建带有表情的人脸模型。因此,Cao 等[25]在 BFM 模型的基础上构建了 FaceWarehouse 模型,由于新增表情基,因此重建后的人脸可以用于表情驱动、迁移等任务。2017 年,Booth 等[26]提出大规模人脸模型(Large Scale Face Model, LSFM),该模型在构建时使用固定拓扑结构的三维人脸网格作为模板,利用非刚性配准(Non-rigid ICP, NICP)和主成分分析(Principal Component Analysis, PCA)建立人脸形状和纹理的参数模型,包含了 9 663 名个体。上述工作仅关注人脸面部区域的模型构建,并未涉及整个头部区域的构建。Li 等[27]利用 3 个不同的数据集,共计 33 000 个真实三维人脸扫描数据,成功构建了更加完整的头部模型 FLAME。该模型能够更准确地建模头部姿势和眼球旋转。本文使用 FLAME 中的 2023 版本作为人脸模型,该版本有着更加丰富的表现力,是目前主流人脸形变模型。

近年来,深度学习模型在计算机视觉等领域的成功应用也推动了三维人脸重建领域的研究进展。Richardson 等[4]将 CNN 引入三维人脸重建,以提高人脸特征提取和表达能力,提高了对 3DMM 人脸基系数求解的准确性。然而,获取单张图像对应的真实三维人脸数据仍然是一项挑战。为解决这一问题,一些工作[6,9-11]开始探索弱监督学习方法,以在没有任何三维标签数据的情况下实现良好的重建效果。在这些方法中,损失函数的设计至关重要,它能够约束重建过程,引导模型学习到不同环境下的人脸图像特征,从而得到更真实、更丰富的三维人脸模型。如 Zielonka 等[15]提出的 MICA 方法,它使用人脸识别网络 Arcface[28]保留重建后模型的身份信息;Feng 等[17]提出 DECA 重建方法,用于捕捉静态和动态细节,该方法使用一致性损失函数约束模型对同一个体不同表情下的形状参数一致,并且在低维空间重建模型基础上,增加静态和动态的置换贴图,有着更丰富的形状细节信息以及较好的模型鲁棒性;Deng 等[29]提出一种混合损失函数,结合视觉感知相似度、人脸身份相似度、关键点投影准确度以及光照一致性等方面约束,考虑人脸的特征、表情、姿势,以及照明等因素,该方法提升了模型的性能。

为了更加生动地重建出局部细节,本文提出一种基于特征核相关的单张图像三维人脸重建方法,同时关注全局和局部形状细节。使用弱监督学习方法,构建多层级损失函数用于约束网络学习全局面部特征;对于面部局部特征,构建基于特征核相关的损失函数,促使模型学习更加深层的局部细节。大量实验结果显示,使用所提出的方法训练网络,得到的模型在鲁棒性和重建质量上均有较好表现。

2 系统模型与方法设计

2.1 人脸形变模型

FLAME 是一个先验统计参数化模型,通过身份参数 $\boldsymbol{\beta}\in\mathbf{R}^{|\beta|}$、姿态参数 $\boldsymbol{\theta}\in\mathbf{R}^{|\theta|}$、表情参数 $\boldsymbol{\psi}\in\mathbf{R}^{|\psi|}$得到一个完整的头部网格(Mesh)模型,FLAME 模型 $\boldsymbol{M}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\psi})$描述如下:

$$\boldsymbol{M}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\psi})=\boldsymbol{W}(T_P(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\psi}),\boldsymbol{J}(\boldsymbol{\beta}),\boldsymbol{\theta},\boldsymbol{W}) \tag{1}$$

式(1)中,线性混合蒙皮函数 $\boldsymbol{W}(T_p,\boldsymbol{J},\boldsymbol{\theta},\boldsymbol{W})$,$\boldsymbol{\beta}$ 为身份参数、$\boldsymbol{\theta}$ 为姿态参数、$\boldsymbol{\psi}$ 是表情参数,关节位置 $\boldsymbol{J}$ 由身份参数 $\boldsymbol{\beta}$ 控制,线性平滑由混合权重 $\boldsymbol{W}$ 控制。模型 $\boldsymbol{M}$ 包含 5 023 个顶点,9 976 个三角面。T_p 定义为

$$T_P(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\psi})=T+B_S(\boldsymbol{\beta};S)+B_P(\boldsymbol{\theta};P)+B_E(\boldsymbol{\psi};E) \tag{2}$$

式(2)中,T 表示平均人脸模板;T_p 表示添加了形状、位姿和表情偏移的模型,S、P、E 分别为形状、位姿和表情混合变形函数的 PCA 基;B_s、B_p、B_E 混合形状(Blend shape)函数分别计算由形状、位姿和表情参数引起的顶点偏移量。

2.2 外观模型

FLAME 缺少纹理基,在重建过程中,需要计算顶点的反照率(Albedo)用于模型约束。本文将 BFM 纹理模型作为解码器 D_A,可由反照率参数 $\boldsymbol{\alpha}\in\mathbf{R}^{|\alpha|}$生成 UV 反照率贴图 $\boldsymbol{A}(\boldsymbol{\alpha})$,通过 UV 空间,将反照率贴图映射到 FLAME 模型上。

2.3 相机模型

本文使用正交投影将 3D 模型投影到二维图像空间中。定义如下:

$$\boldsymbol{v}_i=s\boldsymbol{\Pi}(\boldsymbol{M}_i)+\boldsymbol{t} \tag{3}$$

式(3)中,$\boldsymbol{\Pi}\in\mathbf{R}^{2\times3}$ 是投影矩阵,$\boldsymbol{M}_i\in\mathbf{R}^3$ 是模型 $\boldsymbol{M}$ 中的第 i 个顶点,其中 $\boldsymbol{v}_i\in\mathbf{R}^2$ 是二维图像第 i 个顶点。相机参数 $\boldsymbol{c}=\{s,\boldsymbol{t}\}$,其中,$s\in\mathbf{R}^1$ 为缩放系数,$\boldsymbol{t}\in\mathbf{R}^2$ 为平移量。

2.4 光照模型

为了更真实模拟光照情况,假设人脸表面为 Lambertian 面,采用球谐(Spherical Harmonics, SH)光照函数。纹理贴图定义为

$$\boldsymbol{B}(\boldsymbol{\alpha},\boldsymbol{l},\boldsymbol{N}_{uv})_{i,j}=\boldsymbol{A}(\boldsymbol{\alpha})_{i,j}\odot\sum_{k=1}^{9}\boldsymbol{l}_kH_k(\boldsymbol{N}_{i,j}) \tag{4}$$

式(4)中,$\boldsymbol{l}\in\mathbf{R}^{3\times9}$ 表示球谐光照参数,$\boldsymbol{N}_{i,j}\in\mathbf{R}^3$ 为 UV 空间坐标(i,j)点的表面法向量,同理,$\boldsymbol{B}_{i,j}\in\mathbf{R}^3$ 为该点的纹理值,⊙表示 Hadamard 积,H_k 表示球谐函数的基。

2.5 特征核相关

KC 是一种相似性测量方法,对于两个点 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 之间的核相关性定义为

$$\mathrm{KC}(\boldsymbol{x}_i,\boldsymbol{x}_j)=\int K(\boldsymbol{x},\boldsymbol{x}_i)\times K(\boldsymbol{x},\boldsymbol{x}_j)\,\mathrm{d}x \tag{5}$$

式(5)中,$K(\boldsymbol{x},\boldsymbol{x}_i)$是一个以点 $\boldsymbol{x}_i$ 为中心的核函数。高斯核函数定义为

$$K_{\mathrm{G}}(\boldsymbol{x},\boldsymbol{x}_i)=(\pi\sigma^2)^{-D/2}\exp(-\|\boldsymbol{x}-\boldsymbol{x}_i\|^2/\sigma^2) \tag{6}$$

式(6)中,D 为 $\boldsymbol{x}$ 向量的维数,σ 为高斯函数方差。高斯核函数对应的核相关定义为

$$\mathrm{KC}_{\mathrm{G}}(\boldsymbol{x}_i,\boldsymbol{x}_j)=(2\pi\sigma^2)^{-D/2}\exp\{-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2/2\sigma^2\} \tag{7}$$

对于两个点集 $\boldsymbol{S}$ 和 $\boldsymbol{M}$,KC 算法定义点集相似性损失函数为

$$\mathrm{cost}(\boldsymbol{S},\boldsymbol{M})=-\sum_{m\in\boldsymbol{M}}\sum_{s\in\boldsymbol{S}}\mathrm{KC}(s,\boldsymbol{m}) \tag{8}$$

2.6 方法流程

本文方法整体流程如图 1 所示。首先,编码器 E 回归出输入单张人脸图像 $\boldsymbol{I}$ 的各项参数 $\boldsymbol{e}=\{\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{\alpha},\boldsymbol{c},\boldsymbol{l}\}$。其次,将全连接层输出的编码 $\boldsymbol{e}$ 分解为形状参数($\boldsymbol{\beta}$)、表情参数($\boldsymbol{\psi}$)和位姿参数($\boldsymbol{\theta}$)后,经 3DMM 得到三维人脸形状模型 M;同时,反照率参数($\boldsymbol{\alpha}$)经外观模

型生成 UV 反照率贴图 $\boldsymbol{A}(\boldsymbol{\alpha})$。然后,利用 UV 反照率贴图和光照参数,通过光照模型生成带有阴影的纹理贴图;相机参数用于控制模型的观察角度和距离;可微分渲染器渲染后投影到二维平面得到二维图像 $\boldsymbol{I}_r$。最后,计算 $\boldsymbol{I}$ 和 $\boldsymbol{I}_r$ 之间的特征差异,得到损失值,使用优化器对模型参数进行迭代优化。

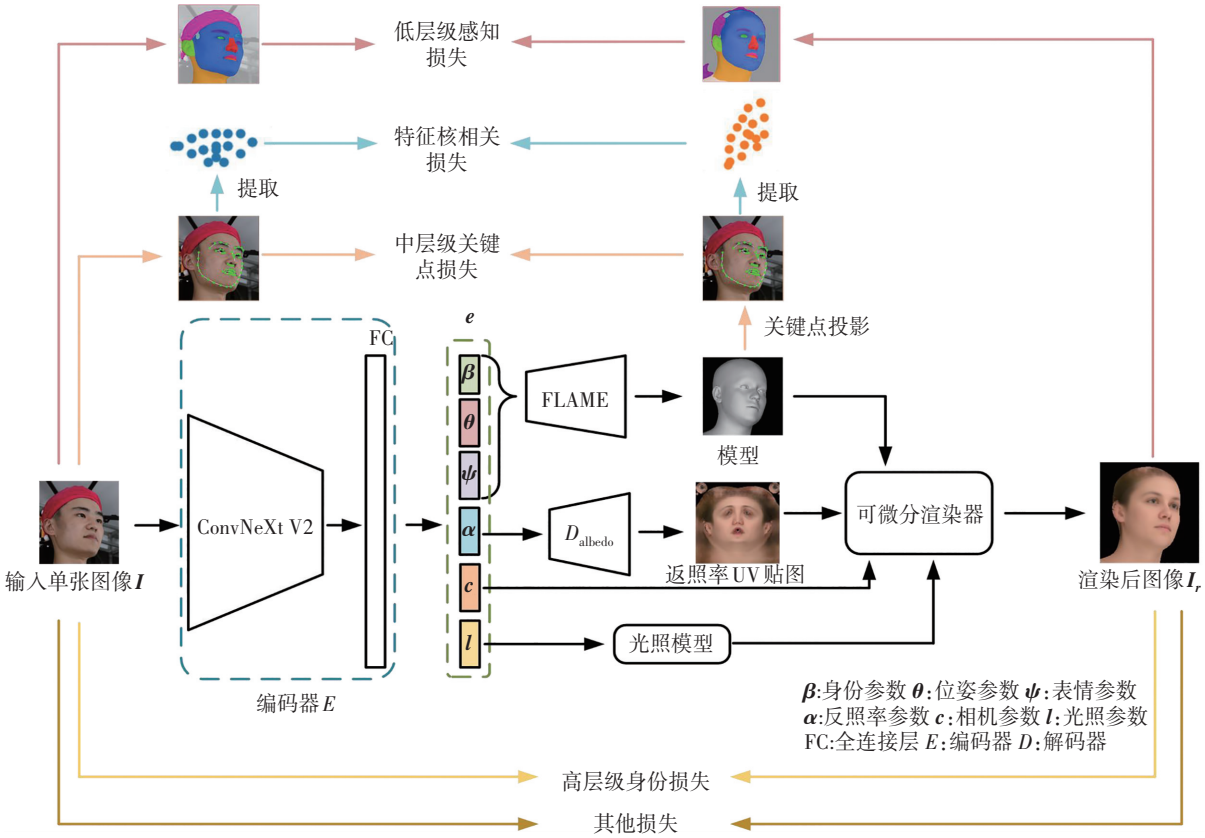


图 1　本文方法整体流程

Fig. 1　The overall process of the proposed method

2.7　损失函数

本文通过弱监督学习方法训练网络模型,设计一种混合损失函数用于模型训练,促使模型学习到全局和局部细节特征。总损失函数由局部形状特征损失和全局多层级特征损失构成,定义为

$$L=L_{\text{pho}}+L_{\text{lmk}}+L_{\text{eye}}+L_{\text{lip}}+L_{\text{kc}}+L_{\text{id}}+L_{\text{sc}}+L_{\text{reg}} \tag{9}$$

式(9)中,L 表示重建总损失;为了保留图像的细节和纹理,考虑底层特征,设计了低层级感知损失 L_{pho};为了约束模型学习中频轮廓信息,设计了中层级人脸关键点损失 L_{lmk};为了捕捉闭眼信息,设计了眼皮距离损失 L_{eye};为了捕捉嘴巴张合信息,设计了嘴唇距离损失 L_{lip};为了捕捉局部形状信息,设计了特征核相关损失 L_{kc};为了确保重建后模型身份与输入图像一致,设计了高层级身份损失 L_{id};为了约束同一个体形状不受视角表情的干扰,设计了形状一致性损失 L_{sc};为了防止模型过拟合,加入重建正则化项 L_{reg}。

(1) 低层级感知损失。用于计算渲染后图像与输入图像像素间颜色的差异。由于不同部位的重建关注度不同,本文使用面部掩膜方式对不同部位加以权重,面部掩膜图可以通过面部语义分割方法获得,如图 2 所示,不同颜色代表着不同部位权重。

$$L_{\text{pho}}=\frac{\sum_{(i,j)\in I} G_{(i,j)}\boldsymbol{W}_{(i,j)}\left\|\boldsymbol{I}_{(i,j)}-\boldsymbol{I}_{r(i,j)}\right\|_2}{\sum_{(i,j)\in I} G_{(i,j)}} \tag{10}$$

式(10)中,$\boldsymbol{W}_{(i,j)}$ 表示输入图像 (i,j) 位置像素点对应的面部掩膜权重;在可见面部皮肤区域中的像素点 G 值为 1,在其他地方 G 值为 0。由于仅对面部区域计算误差,一定程度上提高了模型对头发、衣服、太阳镜等常见遮挡的鲁棒性。

图 2　掩膜示意图

Fig. 2　Illustration of the facial mask

(2) 中层级关键点损失。人脸关键点主要包含面部各部位轮廓,它可以有效地帮助网络校正人脸的姿态,使其更加自然和符合真实世界中的人脸姿势。本文使用68个关键点描述,人脸关键点损失定义如下:

$$L_{lmk}=\sum_{i=1}^{68}\boldsymbol{w}_i\parallel\boldsymbol{K}_i-\boldsymbol{s}\prod(\boldsymbol{M}_i)+\boldsymbol{t}\parallel_1 \tag{11}$$

式(11)中,$\boldsymbol{w}_i$ 为第 i 个关键点的权值,$\boldsymbol{K}_i\in\mathbf{R}^2$ 为第 i 个关键点真实坐标。

(3) 眼皮距离损失与嘴唇距离损失。计算投影图像与输入图像特定关键点对相对偏移量之间的差异。定义如下:

$$L_{eye(lip)}=\sum_{(i,j)\in\boldsymbol{Y}}\parallel\boldsymbol{K}_i-\boldsymbol{K}_j-\boldsymbol{s}\prod(\boldsymbol{M}_i-\boldsymbol{M}_j)\parallel_1 \tag{12}$$

式(12)中,在计算上下眼皮距离时,$\boldsymbol{Y}$ 为上下眼皮关键点索引对的集合;在计算嘴唇距离损失时,$\boldsymbol{Y}$ 则为上下嘴唇关键点索引对的集合。

(4) 特征核相关损失。计算投影后局部特征点与真实特征点之间的相似度。投影后特征点集的KC损失函数值应趋近于真实特征点集的KC损失函数值,即

$$\mathrm{cost}(\boldsymbol{Z},\boldsymbol{Z}')\rightarrow\mathrm{cost}(\boldsymbol{Z},\boldsymbol{Z}) \tag{13}$$

式(13)中,$\boldsymbol{Z}$ 为真实特征点集合,$\boldsymbol{Z}'$为投影后的特征点集合,cost为式(8)中KC定义的点集相似性损失函数。本文使用加权高斯核函数作为KC的核函数,定义为

$$\mathrm{KC}_G(\boldsymbol{x}_i,\boldsymbol{x}_j)=\boldsymbol{w}_i\boldsymbol{w}_j(2\pi\sigma^2)^{-D/2}\exp\{-\parallel\boldsymbol{x}_i-\boldsymbol{x}_j\parallel^2/2\sigma^2\} \tag{14}$$

式(14)中,$\boldsymbol{w}_i$ 为第 i 个特征点权值。由于不同表情下,局部特征点分布和距离不同,即 $\mathrm{cost}(\boldsymbol{Z},\boldsymbol{Z})$ 值不固定,因此,特征核相关损失函数定义为

$$L_{kc}=\frac{\mathrm{cost}(\boldsymbol{Z},\boldsymbol{Z}')-\mathrm{cost}(\boldsymbol{Z},\boldsymbol{Z})}{\mathrm{cost}(\boldsymbol{Z},\boldsymbol{Z})}=\frac{\mathrm{cost}(\boldsymbol{Z},\boldsymbol{Z}')}{\mathrm{cost}(\boldsymbol{Z},\boldsymbol{Z})}-1 \tag{15}$$

最后,代入 KC_G 得

$$L_{kc}=\frac{\sum_{i\in\boldsymbol{Z}}\sum_{j\in\boldsymbol{Z}'}\boldsymbol{w}_i\boldsymbol{w}_j\exp\{-\parallel\boldsymbol{x}_i-\boldsymbol{x}_j\parallel^2/2\sigma^2\}}{\sum_{i\in\boldsymbol{Z}}\sum_{j\in\boldsymbol{Z}}\boldsymbol{w}_i\boldsymbol{w}_j\exp\{-\parallel\boldsymbol{x}_i-\boldsymbol{x}_j\parallel^2/2\sigma^2\}}-1 \tag{16}$$

(5) 高层级身份损失。计算两图像身份特征向量的余弦相似度,定义为

$$L_{id}=1-\frac{f(\boldsymbol{I})f(\boldsymbol{I}_r)}{\parallel f(\boldsymbol{I})\parallel_2\cdot\parallel f(\boldsymbol{I}_r)\parallel_2} \tag{17}$$

式(17)中,f 为训练好的人脸识别网络。

形状一致性损失:对于来自同一个体的不同图像($\boldsymbol{I}_i$ 和 $\boldsymbol{I}_j$),理论上,网络回归出形状参数 $\boldsymbol{\beta}$ 一致。仅交换 $\boldsymbol{\beta}_i$ 和 $\boldsymbol{\beta}_j$,保持其他参数不变,交换形状参数后计算总损失。定义如下:

$$L_{sc}=L(\boldsymbol{I}_i,R(M(\boldsymbol{\beta}_j,\boldsymbol{\theta}_i,\boldsymbol{\psi}_i),B(\boldsymbol{\alpha}_i,\boldsymbol{I}_i,\boldsymbol{N}_{uv,i}),\boldsymbol{c}_i)) \tag{18}$$

式(18)中,R 是可微分渲染函数。

正则化项定义如下:

$$L_{reg}=\boldsymbol{w}_{id}\parallel\boldsymbol{\beta}\parallel_2^2+\boldsymbol{w}_{exp}\parallel\boldsymbol{\psi}\parallel_2^2+\boldsymbol{w}_{albedo}\parallel\boldsymbol{\alpha}\parallel_2^2 \tag{19}$$

式(19)中,$\boldsymbol{w}$ 表示各项权重。

3 仿真实验与结果分析

3.1 实验环境

深度学习框架为PyTorch,系统为Linux,处理器为Intel Xeon Platinum 8352V,显卡为NVIDIA RTX 4090。

3.2 网络结构与参数设置

为了提高网络的性能和泛化能力,编码器使用先进的CNN网络结构ConvNeXt V2[30]。为保证输出层编码维数与参数 $\boldsymbol{e}$ 维数相同,在输出层后连接一个全连接层,三维人脸形变模型使用FLAME,反照率贴图译码器使用BFM纹理基,光照模型使用球谐光照函数,可微分渲染器使用PyTorch3D,优化器使用Adam。

由于本文使用加权高斯核函数作为KC核函数,根据加权高斯核函数特性,在两点集距离较大时,L_{kc} 值变化量较小,即对局部特征点位置大变化不敏感。为了保证模型的训练速度和最终性能,本文首先对模型进行预训练。初始学习率设置为 3×10^{-4},模型经迭代120万次预训练后,对局部部位(嘴巴)加入 L_{kc},学习率改设为 1×10^{-4} 进行模型精训练。输入图像大小为224×224,UV图大小为256×256,眉毛、眼睛、鼻子、嘴巴、皮肤掩膜权重设置为1∶1.2∶1.5∶1.5∶1。

3.3 数据集

使用VGGFace2[31]、BUPT-Balancedface[32]和300W-LP[33]公开数据集作为训练集。总计挑选出200余万张图像,包括亚洲、欧洲和非洲人种,小孩、成人的人脸图像。为了准确检测关键点在图像中的坐标,先采用MediaPipe方法裁剪出人脸部位,然后使用最先进的关键点检测算法DADNet[34]检测关键点。人脸掩膜检测器使用基于BiSeNet V2[35]的人脸语义分割算法获得。将预处理后的关键点及分割掩膜作为弱监督信号,与

单张人脸图像一起送入网络中训练。

3.4　定性分析

为验证面部整体重建的质量,在 FFHQ 数据集上,与 MICA、PRNet 和 FaceVerse[36] 重建方法进行定性对比实验,结果如图 3 所示。与其他先进重建方法相比,本文方法在大姿态、多表情下,均取得了较好的结果,兼顾了面部整体和局部形状的重建质量。与 FaceVerse 和 PRNet 相比,本文方法借助 FLAME 模型,构建了一个完整的头部模型,重建出的面部表情更加自然,更具表现力。此外,本文方法在眼部重建方面表现出较高的准确性和细节性。

虽然 FaceVerse 在嘴部重建方面更富有张力,但是该方法使用 BFM 作为人脸形变模型,并预测细节贴图,最后得到的网格模型具有 451 840 个顶点,901 553 个三角面,远多于 FLAME 顶点数和三角面数,因此导致模型推理时间较长。

PRNet 未借助 3DMM,直接通过神经网络预测 UV 位置图(Position Map, PM)得到三维人脸模型。可以看出,该方法对于局部轮廓信息描述较为模糊,如嘴巴和眼睛部位。由于 MICA 方法注重全局信息,即注重投影渲染后的人脸图像与原输入图像的年龄、性别和身份信息保持一致,因此,难以重建出不同表情下的嘴部形状。相较而言,本文在不同表情下嘴部的细节重建效果更加真实。

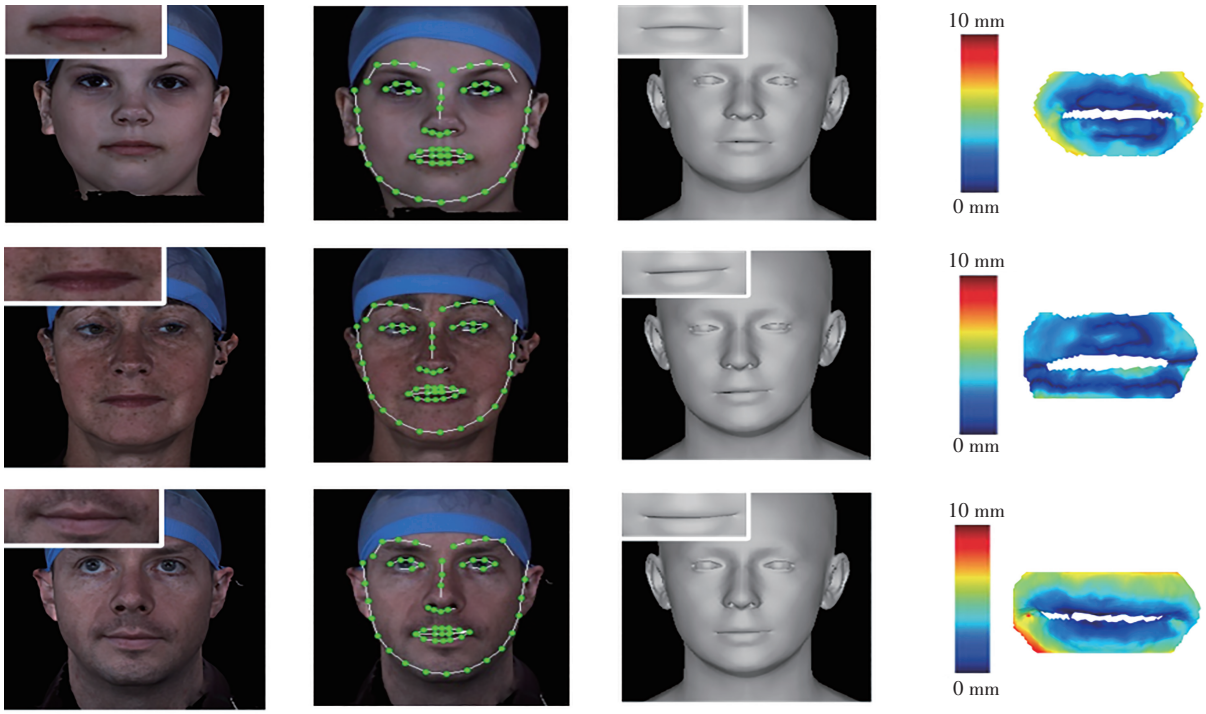
(a) 输入图像　(b) MICA　(c) FaceVerse　(d) PRNet　(e) 本文方法

图 3　重建效果对比

Fig. 3　Comparison of reconstruction results

3.5　可视化展示

为进一步验证局部重建质量,图 4 展示了本文方法在 LYHM[37] 数据集上,对不同个体嘴部的重建效果。图 4(b)展示模型中 68 个 3D 特征点投影到 2D 输入图像上的位置,可以看出模型可以精准地对齐面部各部位轮廓;图 4(c)为本文方法重建后的头部模型以及嘴部模型,可以看出模型完整、自然且生动;图 4(d)展示嘴部模型与真实扫描模型进行局部刚性配准后,得到的顶点真实误差(mm),可以看出重建后嘴部形状有着较小的误差。

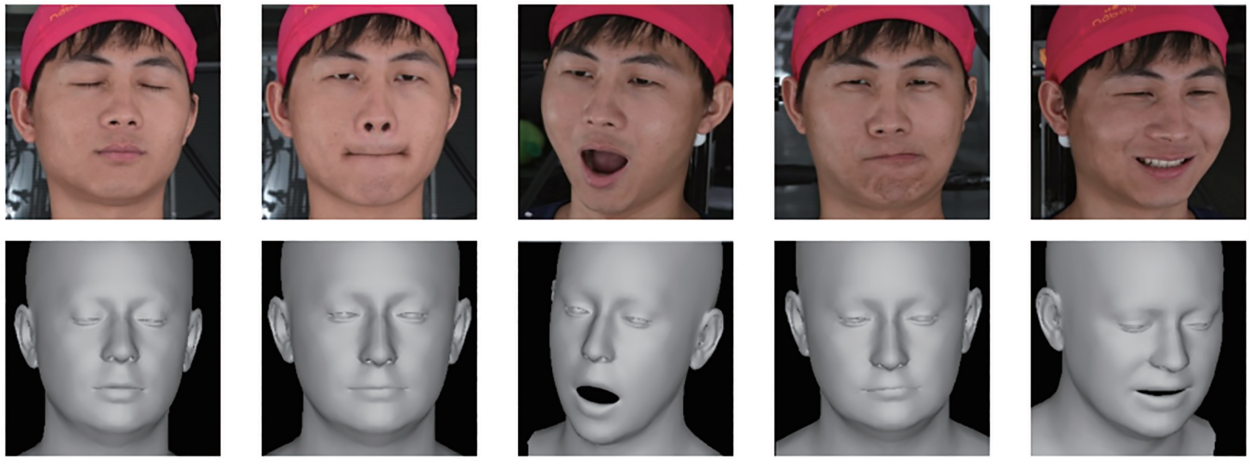


(a) 输入图像 (b) 68关键点 (c) 重建后模型 (d) 嘴巴部位误差

图 4 嘴巴部位重建效果

Fig. 4 Reconstruction results in the mouth region

为了验证本文方法的鲁棒性,从 FaceScape 数据集上选取同一个体在不同视角和表情下的图像,分别对这些图像进行人脸重建,如图 5 所示。结果显示,在不同视角下,该个体做出了复杂的表情,如嘴巴大张合、闭眼、抿嘴等。本文方法可以重建出完整且生动的面部表情,在姿态方面也能够准确地描述。



(a) 闭眼 (b) 抿嘴 (c) 张嘴 (d) 向左抿嘴 (e) 露齿笑

图 5 同一个体不同表情的重建效果

Fig. 5 Reconstruction results for the same individual with different expressions

3.6 定量分析

为了定量分析重建结果,将本文方法与其他先进的重建方法在 REALY 公开基准上进行量化评估对比。REALY 基准包含 100 张不同个体在不同光照环境下采集的人脸图像及其全局对齐的面部扫描,具有准确的面部关键点、高质量的区域掩模和拓扑一致性网格。该评估方法执行区域形状对齐,能够计算准确的双向对齐形状误差。REALY 基准计算出鼻子、嘴巴、额头、脸颊和整个面部对齐后的归一化均方误差(Normalized Mean Square Error, NMSE)后,将误差重新缩放为扫描的原始大小。通过这种方式,形状差异以适当的公制单位进行测量,反映了现实世界的差异。

从表 1 可以看出,在嘴巴部位误差测试中,本文方法的误差均值和中值最小,并且整体上误差均值低于

大部分目前先进的方法。在鼻子部位和脸颊部位误差测试中,由于 MICA 模型在受表情影响小的部位有着出色的表现,因此均取得了最低的误差。DECA 和 EMOCA 的精细重建表现反而低于其粗略重建,可能由于置换贴图预测不准确、位置偏移或者鲁棒性较差等原因。

表 1　在 REALY 基准上的各重建方法对比

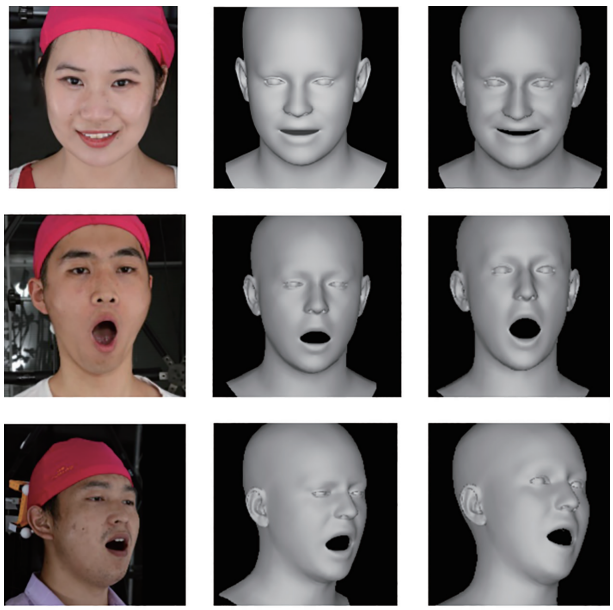
Table 1　Comparison of different face reconstruction methods on the REALY benchmark　/mm

方　法	鼻子误差			嘴巴误差			额头误差			脸颊误差			整体误差
	均值	中值	标准差	均值	中值	标准差	均值	中值	标准差	均值	中值	标准差	均值
FaceVerse-c[36]	3.136	3.132	0.803	3.532	3.364	0.683	5.026	4.974	0.836	2.221	2.191	0.544	3.479
MICA[15]	**1.585**	**1.542**	**0.325**	3.478	3.439	1.204	2.374	2.251	0.683	**1.099**	**1.003**	**0.324**	2.134
FaceVerse-f[36]	2.837	2.877	0.702	3.312	3.327	0.712	4.324	4.342	0.763	2.369	2.345	0.581	3.211
EMOCA-f[18]	2.532	2.563	0.539	2.929	2.676	1.106	2.595	2.505	0.631	1.495	1.360	0.469	2.388
DECA-f[17]	2.138	2.137	0.461	2.802	2.699	0.868	2.457	2.341	**0.559**	1.443	1.353	0.498	2.210
EMOCA-c[18]	1.868	1.821	0.387	2.679	2.419	1.112	2.426	2.383	0.641	1.438	1.294	0.501	2.103
FS-Opt-c[22]	2.516	2.382	0.643	2.569	2.540	**0.597**	2.350	2.184	0.636	2.261	2.205	0.529	2.424
FS-Opt-f[22]	2.474	2.352	0.636	2.539	2.495	0.600	**2.341**	**2.175**	0.632	2.259	2.205	0.526	2.403
N-3DMM[38]	2.936	2.857	0.810	2.375	2.390	0.599	4.582	4.452	1.488	1.918	1.717	0.801	2.953
本文方法	1.845	1.786	0.409	**2.206**	**2.187**	0.738	2.563	2.514	0.645	1.796	1.606	0.718	**2.102**

注:“-c”为粗略重建方法,“-f”为精细重建方法,“黑体”为最小值,“细线”为次小值。

3.7　消融实验

为验证本文提出的特征核相关损失函数的有效性,首先在 FaceScape 数据集上进行了消融实验,如图 6 所示。为了有效验证,本次实验采用同样的学习率、批次大小和各项权重对模型进行训练,模型迭代相同次数。实验结果可以看出,在不同的嘴部表情和姿态下,模型在新增嘴部特征核相关损失函数后,对于局部细节有了更好的刻画。



(a) 原图　(b) 未使用特征核相关损失　(c) 使用特征核相关损失

图 6　FaceScape 数据集上的消融实验

Fig. 6　Ablation experiments on the FaceScape dataset

另一方面,在 LYHM 公开数据集上进行消融实验,通过比较是否使用特征核相关损失得到对嘴部的约束。函数重建后的模型在嘴部的误差值如表 2 所示。本次选取 5 组不同个体的单张图像及其真实扫描模型,使用 REALY 基准中区域形状对齐方法,计算重建后的模型与真实扫描模型在嘴部的 NMSE。为了反映真实世界的差异,对 NMSE 值缩放到原扫描模型大小(mm)。

表 2　使用特征核相关损失函数前后的嘴部误差(LYHM 数据集)

Table 2　NMSE in the mouth region with and without feature kernel correlation loss (LYHM dataset)　/mm

序　号	未使用特征核相关损失	使用特征核相关损失
1	4.300	1.377
2	3.144	1.569
3	2.718	1.667
4	2.987	1.854
5	3.561	2.261

从表 2 可以看出,使用特征核相关损失函数对局部(嘴巴)进行约束后,局部形状误差有较大幅度减小。这说明特征核相关损失约束起到了至关重要的作用。

4　结　论

本文提出了一种基于特征核相关的人脸重建方法,使用弱监督学习方式训练卷积神经网络,无需大量带有三维标签的数据,训练后的模型可以对单张人脸图像重建出三维人脸模型。将提出的特征核相关损失

用于具有挑战的嘴部约束。在 FaceScape、FFHQ 和 LYHM 公开数据集上的实验结果表明,加入新的损失函数后,可以有效提升模型对嘴部形状特征的捕捉能力,重建出的嘴部形状更加生动逼真,优于目前优秀的方法,并且在不同的姿态和表情中展现了良好的鲁棒性。

本文将所提的损失函数仅用于嘴部约束,未来还可以尝试将其应用于其他部位或者整个面部。另外,目前的先进方法是在 3DMM 基础上添加置换贴图,以进一步刻画面部细节,如皱纹、伤疤、酒窝等,实现更精细的重建。未来工作将尝试在本文重建的模型中加入细节贴图和纹理贴图,重建出更加逼真的人脸模型。

参考文献(References):

[1] 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. New York: ACM, 1999: 187-194.

[2] Feng Y, Wu F, Shao X, et al. Joint 3D face reconstruction and dense alignment with position map regression network [C]// Computer Vision-ECCV 2018, Cham: Springer, 2018: 557-574.

[3] 沈钺潇, 钱丽萍, 俞宁宁. 融合 UV 位置图与 CGAN 的单图像大视角三维彩色人脸重建[J]. 计算机辅助设计与图形学学报, 2022, 34(4): 614-622.
Shen Chengxiao, Qian Liping, Yu Ningning. Large-view 3D color face reconstruction from single image via UV location map and CGAN[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(4): 614-622.

[4] Richardson E, Sela M, Or-el R, et al. Learning detailed face reconstruction from a single image[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 5553-5562.

[5] Fan X, Cheng S, Kang H, et al. Dual neural networks coupling data regression with explicit priors for monocular 3D face reconstruction[J]. IEEE Transactions on Multimedia, 2021, 23: 1252-1263.

[6] Lin J, Yuan Y, Shao T, 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. Piscataway: IEEE Press, 2020: 5890-5899.

[7] 朱磊, 王善敏, 刘青山. 基于人脸部件掩膜的自监督三维人脸重建[J]. 计算机科学, 2023, 50(2): 214-220.
Zhu Lei, Wang Shanmin, Liu Qingshan. Self-supervised 3D face reconstruction based on detailed face mask[J]. Computer Science, 2023, 50(2): 214-220.

[8] 章毅, 吕嘉仪, 兰星, 等. 结合面部动作单元感知的三维人脸重建算法[J]. 软件学报, 2024, 35(5): 2176-2191.
Zhang Yi, Lyu Jiayi, Lan Xing, et al. AU-aware algorithm for 3D facial reconstruction[J]. Journal of Software, 2024, 35(5): 2176-2191.

[9] Deng Y, Yang J, Xu S, 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. Piscataway: IEEE Press, 2019: 285-295.

[10] Gecer B, Ploumpis S, Kotsia I, et al. GANFIT: generative adversarial network fitting for high fidelity 3D face reconstruction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 1155-1164.

[11] 申冲, 刘川, 张满囤, 等. 基于弱监督学习的细节三维人脸重建[J]. 燕山大学学报, 2023, 47(2): 144-151, 163.
Shen Chong, Liu Chuan, Zhang Mandun, et al. Detailed 3D face reconstruction based on weakly supervised learning[J]. Journal of Yanshan University, 2023, 47(2): 144-151, 163.

[12] 周大可, 张超, 杨欣. 基于多尺度特征融合及双重注意力机制的自监督三维人脸重建[J]. 吉林大学学报(工学版), 2022, 52(10): 2428-2437.
Zhou Dake, Zhang Chao, Yang Xin. Self-supervised 3D face reconstruction based on multi-scale feature fusion and dual attention mechanism[J]. Journal of Jilin University (Engineering and Technology Edition), 2022, 52(10): 2428-2437.

[13] 周健, 黄章进. 基于改进三维形变模型的三维人脸重建和密集人脸对齐方法[J]. 计算机应用, 2020, 40(11): 3306-3313.
Zhou Jian, Huang Zhangjin. 3D face reconstruction and dense face alignment method based on improved 3D morphable model [J]. Journal of Computer Applications, 2020, 40(11): 3306-3313.

[14] Tewari A, Zollhofer M, Kim H, et al. MoFA: model-based deep convolutional face autoencoder for unsupervised monocular reconstruction[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2017: 1274-1283.

[15] Zielonka W, Bolkart T, Thies J. Towards metrical reconstruction of Human faces[C]//Computer Vision: ECCV 2022. Cham: Springer, 2022: 250-269.

[16] Genova K, Cole F, Maschinot A, et al. Unsupervised training for 3D morphable model regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8377-8386.

[17] Feng Y, Feng H, 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): 1-13.

[18] Danecek R, Black M, Bolkart T. EMOCA: emotion driven monocular face capture and animation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 20279-20290.

[19] Tsin Y, Kanade T. A correlation-based approach to robust point set registration[M]//Computer Vision: ECCV 2004. Berlin, Heidelberg: Springer, 2004: 558-569.

[20] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4396-4405.

[21] Chai Z, Zhang H, Ren J, et al. REALY: Rethinking the Evaluation of 3D Face Reconstruction[M]//Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022: 74-92.

[22] Zhu H, Yang H, Guo L, et al. FaceScape: 3D facial dataset and benchmark for single-view 3D face reconstruction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(12): 14528-14545.

[23] Paysan P, Knothe R, Amberg B, et al. A 3D face model for pose and illumination invariant face recognition[C]//Proceedings of the Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway: IEEE Press, 2009: 296-301.

[24] Fyffe G, Jones A, Alexander O, et al. Driving high-resolution facial scans with video performance capture[J]. ACM Transactions on Graphics, 2014, 34(1): 1-14.

[25] Cao C, Weng Y, Zhou S, et al. FaceWarehouse: a 3D facial expression database for visual computing[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(3): 413-425.

[26] Booth J, Roussos A, Ponniah A, et al. Large scale 3D morphable models[J]. International Journal of Computer Vision, 2018, 126(2): 233-254.

[27] 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.

[28] Deng J, Guo J, Xue N, et al. ArcFace: additive angular margin loss for deep face recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4685-4694.

[29] Deng Q, Le B H, Jin A, et al. End-to-end 3D face reconstruction with expressions and specular albedos from single In-the-wild images[C]//Proceedings of the 30th ACM International Conference on Multimedia. New York: ACM, 2022: 4694-4703.

[30] Woo S, Debnath S, Hu R, et al. ConvNeXt V2: co-designing and scaling ConvNets with masked autoencoders[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 16133-16142.

[31] Cao Q, Shen L, Xie W, et al. VGGFace2: a dataset for recognising faces across pose and age[C]//Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition. Piscataway: IEEE Press, 2018: 67-74.

[32] Wang M, Deng W, Hu J, et al. Racial faces in the wild: reducing racial bias by information maximization adaptation network[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 692-702.

[33] Zhu X, Lei Z, Liu X, et al. Face alignment across large poses: a 3D solution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 146-155.

[34] Martyniuk T, Kupyn O, Kurlyak Y, et al. DAD-3DHeads: a large-scale dense, accurate and diverse dataset for 3D head alignment from a single image[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 20910-20920.

[35] 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(11): 3051-3068.

[36] Wang L, Chen Z, 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. Piscataway: IEEE Press, 2022: 20301-20310.

[37] Dai H, Pears N, Smith W, et al. Statistical modeling of craniofacial shape and texture[J]. International Journal of Computer Vision, 2020, 128(2): 547-571.

[38] Tran L, Liu X. Nonlinear 3D face morphable model[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7346-7355.

责任编辑:李翠薇