

基于弱监督学习的三维人脸形状与纹理重建^①



吴越, 董兰芳

(中国科学技术大学 计算机科学与技术学院, 合肥 230027)

通讯作者: 董兰芳, E-mail: lfdong@ustc.edu.cn

摘要: 三维人脸相较于二维人脸包含了更多特征信息, 可应用于如人脸识别、影视娱乐、医疗美容等更多实际应用场景, 因此三维人脸重建技术一直是计算机视觉领域的研究热点. 由于真实三维人脸数据较难获取, 很多基于深度学习的重建算法首先利用传统重建方法为大量二维人脸图像构建三维标签, 作为训练数据, 这些数据可能并不精准, 从而导致算法的重建精度受到影响. 为此, 本文提出一种基于 multi-level 损失函数的弱监督学习模型, 结合传统三维人脸形变模型 3DMM 与深度学习方法, 直接从大量无三维标签的二维人脸图像中学习三维人脸特征信息, 从而实现基于单张二维人脸图像的三维人脸重建算法. 此外, 为解决二维人脸图像中常存在遮挡或大姿态情况而影响人脸纹理重建的问题, 本文使用基于 CelebAMask-HQ 数据集的人脸解析分割算法对图像进行预处理去除遮挡区域. 实验结果表明, 基于本文方法的三维人脸重建质量与重建精度均实现了一定的提升.

关键词: 三维人脸重建; 深度学习; 弱监督学习; 三维形变模型; 人脸解析分割; 纹理重建

引用格式: 吴越, 董兰芳. 基于弱监督学习的三维人脸形状与纹理重建. 计算机系统应用, 2020, 29(11): 183-189. <http://www.c-s-a.org.cn/1003-3254/7654.html>

3D Face Shape and Texture Reconstruction Based on Weakly Supervised Learning

WU Yue, DONG Lan-Fang

(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)

Abstract: Compared with two-dimensional faces, three-dimensional faces contain more feature information and can be applied to more practical application scenarios, such as face recognition, film and television entertainment, medical beauty, etc. Therefore, 3D face reconstruction technology has become a research focus in the field of computer vision. Due to real 3D face data is difficult to obtain, many deep learning-based reconstruction algorithms first use traditional reconstruction methods to construct 3D labels for a large number of 2D face images. These training data may not be accurate which will affect the reconstructive accuracy of these algorithm. To this end, this study proposes a weakly supervised learning model based on a multi-level loss function, which combines traditional 3D morphable model 3DMM and deep learning methods to directly learn 3D face feature from a large number of 2D face images without 3D labels to implement a 3D face reconstruction algorithm based on a single 2D face image. In addition, in order to solve the problem that occlusion or large poses in 2D face images often affect the reconstruction of face texture, this paper uses a face parse segmentation algorithm based on the CelebAMask-HQ dataset to preprocess the images to remove the occlusion areas. The experimental results show that the quality and accuracy of 3D face reconstruction based on the proposed method have been improved greatly.

Key words: 3D face reconstruction; deep learning; weakly supervised learning; 3D morphable model; face parse segmentation; texture reconstruction

① 收稿时间: 2020-04-09; 采用时间: 2020-05-10; csa 在线出版时间: 2020-10-29

1 引言

随着计算机技术的发展,三维人脸重建成为计算机视觉领域热门的研究课题.相较于二维人脸,三维人脸包含更多特征信息,能够应用于更丰富的实际生活场景,如人脸识别^[1]、影视娱乐^[2]、医疗美容^[3]等.

近年来,基于单张二维人脸图像重建三维人脸的工作取得了巨大的成功^[4],本文将现有三维人脸重建技术大致分为两大类:传统重建方法和基于深度学习的重建方法.目前自动化程度最高、应用最广泛的传统重建方法是基于统计模型的三维人脸形变模型(3D Morphable Model, 3DMM)^[5],但该方法主要问题是其迭代拟合过程中的计算代价过大、重建效率偏低.

随着深度学习在计算机视觉领域取得重大突破,很多学者研究了如何利用深度学习对二维人脸图像进行3DMM三维人脸形状和纹理系数估计.深层神经网络提供了从图像像素到三维模型参数的回归学习能力,然而利用深度学习进行三维人脸重建的首要问题是获取三维人脸数据的代价过高,训练数据不足.

一种常用的解决训练数据不足的方案是先通过传统的三维人脸重建方法生成大量三维人脸合成数据作为训练数据. Tran等^[6]针对训练数据不足的问题,利用一种多图像3DMM估计方法生成足量带标签的三维人脸数据,并在ResNet-101上训练回归3DMM模型系数,实现基于单张二维人脸图像的三维人脸重建. Richardson等^[7]通过随机采样3DMM模型,修改模型参数后生成新的三维人脸后进行渲染得到二维图像,将生成的新的三维人脸和渲染投影后的二维图像作为一组训练对并作为CNN网络模型的输入.然而,当面对二维人脸图像中存在遮挡、不寻常的光照时,这些纯粹由合成数据训练的网络表现的并不好.除此以外,大量的实验结果表明合成的图像往往缺乏真实特征,对重建精度易产生负面影响.

另一种解决训练数据不足的方法是直接从大量无三维标注的二维人脸图像中学习3DMM系数. Tewari等^[8]首先提出了从无标签的二维人脸图像中直接学习3DMM系数的方法,他们训练了一个自动编解码网络以同时预测3DMM人脸形状、纹理、姿势、光线等参数,该方法不需要三维标注与二维人脸图像的监督数据对.然而,由于该方法的训练损失函数是基于图像像素的,网络容易混淆相关变量之间的变化. Genova

等^[9]进一步改进了这一方法,通过使用预先训练的人脸识别网络,如VGG-Face或谷歌的FaceNet,在损失函数中添加了高层特征的约束.

此外,目前大多数基于单张图像的三维人脸重建算法只重建三维人脸几何形状信息,后续为增强真实感直接使用纹理贴图方法将图像中人脸纹理映射到三维人脸.该方法最大的问题是由于人脸图像存在不可见区域,侧脸、耳朵以及脖子处的纹理将无法获取;对于存在遮挡或大姿态情况下的二维人脸图像,大面积不可见区域的人脸纹理更是无法直接从二维图像中提取.

针对以上问题,本文提出一种可同时重建三维人脸形状和纹理信息的弱监督学习算法.首先针对训练数据不足问题,本文设计了一种multi-level的弱监督损失函数,可直接从大量二维人脸图像中学习多种弱监督信息来监督网络模型学习.对于人脸纹理重建,本文通过可微分渲染器SoftRas^[10]将重建后的三维人脸渲染至二维图像平面与输入人脸图像建立起像素级别的弱监督信息,从而实现对面脸纹理信息的约束与重建.此外为避免遮挡对面脸纹理重建造成影响,本文首先利用基于CelebAMask-HQ数据集^[11]训练的人脸解析分割算法去除遮挡.实验结果显示,相较于其它三维人脸重建算法,本文算法的重建质量与重建精度均有显著提升.

2 网络模型与关键技术

本文提出的重建算法的整体框架如图1所示.给定无三维标签的二维人脸图像 I ,先用人脸解析分割算法计算人脸分割掩码mask,利用该mask可去除人脸面部的遮挡并使模型更关注人脸中如眼、眉毛等重要部位.此外利用目前最先进的人脸特征点检测算法^[12]计算68个人脸特征点lmk作为弱监督信息.随后将去遮挡后的图像作为输入,通过CNN模型回归输出代表各关键系数的特征向量,包括3DMM、位姿以及光照参数,完成三维人脸重建.为实现重建三维人脸纹理信息,给定位姿和光照参数,利用可微分渲染器SoftRas将重建三维人脸映射到二维图像平面,得到渲染图像 I' ,从而实现与输入图像 I 之间的像素级别的纹理监督信号.将输入图像和渲染图像输入一个人脸识别网络计算高维度的身份识别损失,实现一种基于multi-level损失函数的CNN网络模型,以取得更好的重建结果.

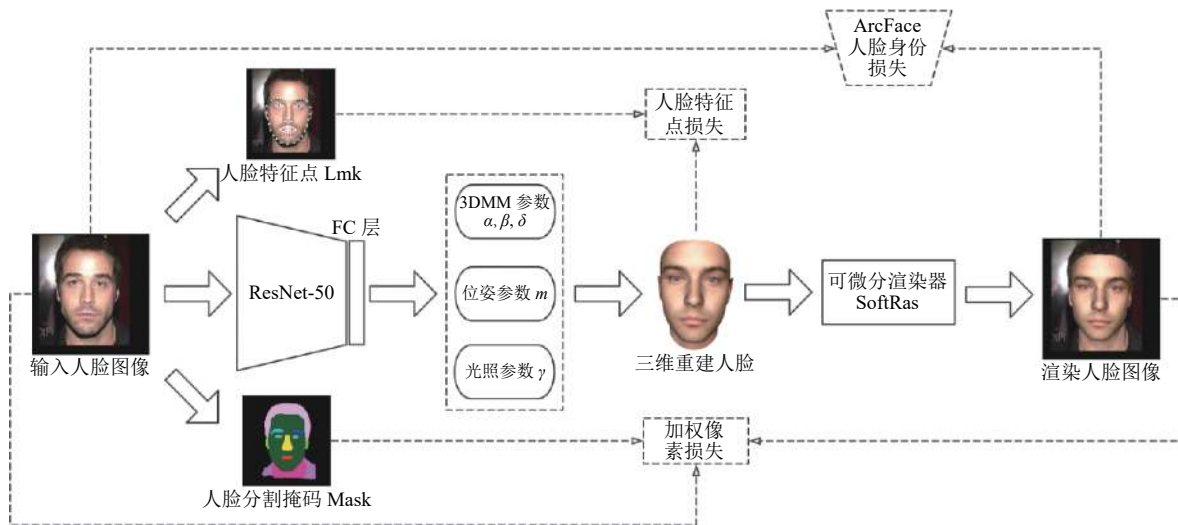


图1 模型训练流程图

2.1 模型参数

为实现同时重建三维人脸形状和纹理信息,需要借助 CNN 卷积神经网络强大的特征提取能力,从二维人脸图像中提取 3DMM 人脸参数、位姿参数以及光照参数以实现基于单张二维人脸图像的三维人脸重建.

(1) 三维人脸形变模型 3DMM

传统三维人脸形变模型 3DMM 将三维人脸统一用两个向量表示:

$$S = (x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_n, y_n, z_n)^T \in R^{3n} \quad (1)$$

$$T = (r_1, g_1, b_1, r_2, b_2, \dots, r_n, g_n, b_n)^T \in R^{3n} \quad (2)$$

其中,形状向量 S 表示三维人脸形状,纹理向量 T 表示三维人脸纹理.由于人脸空间被假设为一个线性空间,任何一张三维人脸都可以由线性子空间内的其它三维人脸线性组合表示,利用主成分分析法 PCA 进行降维与去相关后,新的三维人脸模型可表示为:

$$S_{\text{new}} = \bar{S} + A_{\text{id}}\alpha + A_{\text{exp}}\beta \quad (3)$$

$$T_{\text{new}} = \bar{T} + A_{\text{tex}}\delta \quad (4)$$

其中, S 和 T 分别代表平均人脸形状和纹理向量, $A_{\text{id}}, A_{\text{exp}}$ 和 A_{tex} 分别表示人脸形状、表情和纹理 PCA 主成分.本文中的三维人脸形状和纹理信息来源于 Basel Face Mode 模型^[13],三维人脸表情信息来源于 FaceWarehouse 三维人脸表情库^[14].通过改变 3DMM 人脸形状参数 α 、表情参数 β 以及纹理参数 δ ,即可生成新的三维人脸.

(2) 相机模型

假设三维人脸中顶点为 V_{3d} ,通过弱透视投影将三维人脸顶点投影至二维平面顶点 v 的过程可表示为:

$$v = f * P * R * V_{3d} + t \quad (5)$$

其中, f 为缩放因子, P 是相机矩阵,因本文使用弱透视投影模型, P 为正交矩阵 $[[1, 0, 0], [0, 1, 0]]$, R 为旋转矩阵, t 为平移向量.三维人脸的位姿参数即可用 $[f, R, t]$ 表示.

(3) 光照模型

光照模型,也称明暗模型,用于计算物体某点处的光强.光照估计就是从图片中获取光照信息,从而降低光照对纹理的影响.本文采用球谐函数 (Spherical Harmonics, SH) 来估计二维人脸图像中的光照信息.

对于三维人脸模型中的一点 p ,点 p 处的像素纹理信息即为 p 处的反照率 (去除光照和阴影后的皮肤纹理) 乘以球谐基函数的线性组合:

$$I(p) = t \cdot \sum_{j=1}^9 \gamma_j h_j(n) \quad (6)$$

其中, t 代表顶点 p 处的皮肤纹理, $n = (n_x, n_y, n_z)$ 代表法线方向, h_j 代表 9 个球谐基函数, γ_j 表示球谐基函数的系数,即光照系数.

综上,为实现基于单张二维人脸图像的三维人脸形状和纹理重建任务,需准确预测输入人脸图像的各关键参数,包括 3DMM 人脸形状参数 $\alpha \in R^{100}$ 、人脸表情参数 $\beta \in R^{80}$ 、人脸纹理参数 $\delta \in R^{100}$ 、人脸位姿参数 $m \in R^7$ 以及光照条件参数 $\gamma \in R^{27}$.

2.2 可微分渲染器 SoftRas

为了在三维人脸模型和二维渲染人脸图像间建立起像素级别的监督,使用可微分渲染器 (Soft-Rasterizer,

SoftRas)对重建后的三维人脸渲染生成二维人脸图像,从而重建出精度更高,更有真实感的三维人脸。

可微分渲染器 SoftRas 将渲染过程看作是一个联合函数,不同于传统渲染器只选择最近的三角形的颜色的离散采样操作,SoftRas 假设所有的网格三角形对每个渲染的像素都有概率贡献,融合所有三角形对渲染像素的概率贡献,即可获得最终的渲染结果。作为一个真正可微分的渲染框架,SoftRas 能够生成从像素到三维网格顶点、纹理、法向量等属性的有效梯度,因此可以将 SoftRas 作为一个通用框架应用在深度神经网络模型中,从而为基于单张无三维标签图像的三维重建任务提供一种像素级别的监督信号。

2.3 人脸解析分割算法

由于训练数据中很可能会因为人脸区域内存在如头发、帽子、眼镜等遮挡物,造成重建后相应遮挡区域内的人脸纹理出现问题。因此需要对训练数据进行预处理以避免遮挡物对最终的重建效果造成影响。

在实际数据中,最常见的遮挡是头发对人脸的遮挡,而传统的肤色检测算法并不能很好区分与肤色像素相近的头发。因此本文采用基于 CelebAMask-HQ 数据集的人脸解析分割算法,为输入图像去除遮挡。

CelebAMask-HQ 数据集是一个高分辨率的人脸图像集,每张图像都定义了 19 种面部属性的分割掩码,包括人脸器官如皮肤、眼睛、眉毛、头发等,以及一些外界遮挡如眼镜、帽子等物体。本文在 CelebAMask-HQ 数据集上训练语义分割网络 BiseNet^[15],实现一个人脸解析分割算法,为训练数据计算人脸图像的分割掩码 mask,从而在计算像素级损失时仅计算去除遮挡的人脸区域,该人脸解析分割算法效果如图 2 所示。

2.4 损失函数

本文仅通过无三维标签的二维人脸数据重建三维人脸,因此,如何设计有效的弱监督信号对最终的重建效果至关重要。为了实现输入人脸图像与渲染人脸图像间的对应关系,本算法充分考虑了图像的多个维度层面的特征,包括低维度的加权像素级光度损失,中层的人脸特征点损失以及作为高维度的人脸身份识别损失,最终实现一种 multi-level 的弱监督损失函数。因此,整个网络模型的目的是最小化如下损失函数:

$$L = \lambda_{\text{pixel}}L_{\text{pixel}} + \lambda_{\text{lmk}}L_{\text{lmk}} + \lambda_{\text{id}}L_{\text{id}} + \lambda_{\text{reg}}L_{\text{reg}} \quad (7)$$

其中, λ_{pixel} , λ_{lmk} , λ_{id} , λ_{reg} 是控制损失项重要性的常数。

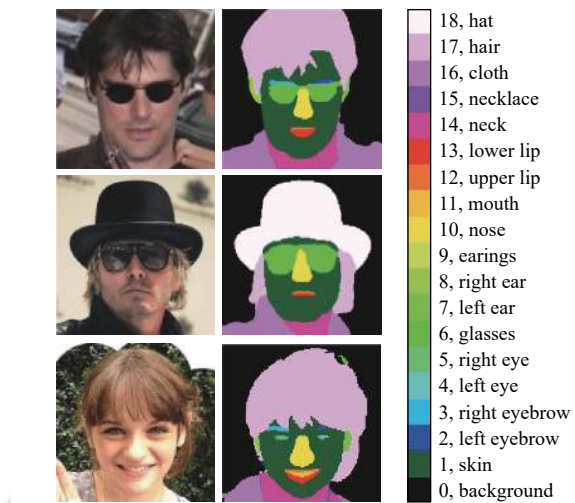


图 2 人脸解析分割算法示例

(1) 低层加权像素级损失

为使输入图像和重建渲染图像尽可能相似,最简单直观的方法是直接计算两张图像间低维度的像素差异。得益于人脸解析分割算法,不仅能检测出如头发、帽子等遮挡物,还可以根据分割掩码 mask 准确判断各人脸器官如眉毛、眼睛等在图像中的位置。因此,在计算这部分损失时,添加一定的权重让模型更加关注这些人脸特征部位的纹理,使重建三维人脸纹理更具真实感。该部分低维度的加权像素损失可定义为:

$$L_{\text{pixel}} = \frac{M}{|V|} \sum_{(i,j) \in V} \|I_{(i,j)} - I'_{(i,j)}\|_2 \quad (8)$$

其中, V 代表了渲染后人脸图像的可见区域, (i, j) 代表可见区域内的像素位置,因此,只需要取输入图像内相同区域进行比较即可。此外, M 代表掩码权重,主要作用是通过分割掩码 mask 判断该可见区域内是否存在如头发、帽子等遮挡物,若存在则将该区域权重设为 0,以避免遮挡对重建造成影响。此外,将眼睛、眉毛等人脸重要器官区域的权重设为 2,其它人脸区域如皮肤等权重仍设为 1。

(2) 中层人脸特征点损失

本文利用目前最先进的人脸特征点检测算法来检测无标签二维图像的 68 个人脸特征点 $\{q_n\}$ 作为一种弱监督信号,并将重建后的三维人脸投影至二维平面得到重建后的 68 个人脸特征点 $\{q'_n\}$,使用专门为人脸特征点检测任务设计的损失函数 Wing Loss^[16],计算一个中层特征的人脸特征点损失:

$$L_{lmk} = \begin{cases} w \ln \left(1 + \frac{|\Delta Q_{lmk}|}{\epsilon} \right), & \text{if } |\Delta Q_{lmk}| < w \\ |\Delta Q_{lmk}| - c, & \text{otherwise} \end{cases} \quad (9)$$

其中, $\Delta Q_{lmk} = q'_n - q_n$, 代表了根据检测得到的人脸特征点与重建得到的人脸特征点间的误差. 通过特征点损失, 能够有效引导网络模型对人脸形状、表情以及姿态参数的学习.

(3) 高层人脸身份损失

仅使用低维的像素级损失很可能会导致局部极小值问题, 因此考虑图像的高维度特征, 对于一个人脸识别网络, 从输入图像和重建图像中提取的人脸特征信息是有一定关联的. ArcFace^[17]是目前最优秀的人脸识别网络, 给定一张二维人脸图像, ArcFace可输出一个512维的人脸特征向量. 因此, 本算法在训练过程中利用预训练的 ArcFace 人脸识别网络, 同时提取输入二维人脸图像 I 和重建渲染图像 I' 的512维人脸特征向量 $f(I)$ 和 $f(I')$, 并计算两者之间的余弦距离:

$$L_{id} = f(I) \cdot f(I') \quad (10)$$

通过该损失, 可以避免模型出现局部极小值问题. 此外, 对于人脸识别网络, 尽管同一个人在多张图像中呈现不同的位姿、表情, 最终识别的结果仍是同一个人. 基于人脸识别网络这一特性, 本算法在计算高维度的人脸身份损失时, 通过对重建后的三维人脸的位姿与表情参数随机添加一个偏置, 变换后的人脸渲染图像 I'' 与输入图像 I 中的人脸身份应仍是一致的. 同理, 两张渲染图像的人脸身份也应是一致的, 即:

$$L'_{id} = f(I) \cdot f(I'') \quad (11)$$

$$L''_{id} = f(I') \cdot f(I'') \quad (12)$$

因此, 本算法在计算人脸身份损失时, 对重建得到的人脸表情 β 和位姿参数 m 进行小范围的随机变换, 实现多视角多表情的人脸身份损失, 从而能够训练出更鲁棒的网络模型.

(4) 正则化损失

由于3DMM参数是服从正态分布的, 若从二维无标签人脸图像中回归的3DMM参数值偏离零点太多, 则重建的三维人脸可能是扭曲的, 因此为了防止人脸形状和纹理退化, 为回归的3DMM人脸形状、表情和纹理参数添加了正则化损失项作为一种弱监督损失:

$$L_{reg} = \omega_{id} \|\alpha\|^2 + \omega_{exp} \|\beta\|^2 + \omega_{tex} \|\delta\|^2 \quad (13)$$

综上, 通过上述介绍的 multi-level 的多种弱监督损失函数, 可以在不同程度上约束待回归的人脸形状、表情、纹理、位姿以及光照参数, 从而能够训练出一个鲁棒且具有真实感的三维人脸重建网络模型.

3 实验分析

3.1 实现细节

本文算法基于 Pytorch 深度学习框架, 采用 ResNet-50 网络模型, 将最后一层全连接层修改为314维, 并采用 Adam 方法优化参数, 设置初始学习率为 1×10^{-4} , batch size 设为16, 共训练101340次迭代, 10个epoch. 训练过程中, 各损失函数的权重 λ_{pixel} 、 λ_{lmk} 、 λ_{id} 、 λ_{reg} 分别设置为3.0、 2.4×10^{-3} 、0.24和 4×10^{-4} , 正则化损失项中的 ω_{id} 、 ω_{exp} 和 ω_{tex} 分别设置为1.2、0.8和0.001.

3.2 重建质量

为能够定性定量地评估本章算法的重建结果, 将从重建质量以及重建精度两方面来验证模型的有效性. 重建质量部分首先将本算法在特定图像上的重建结果与目前一些先进的三维人脸重建算法如 Tran 等^[6]、MoFA^[8]以及 Genova 等^[9]算法的重建结果在不同的人脸数据集如 MoFA-test 以及 LFW 数据集上选择若干人脸图像进行视觉上的对比比较, 如图3与图4所示, 部分实验结果图来源于文献[9].

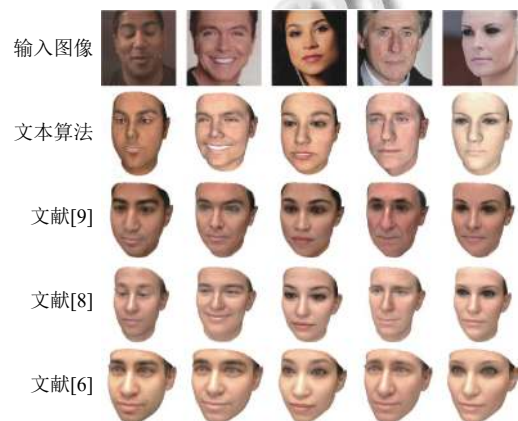


图3 部分 MoFA-test 数据上的重建质量对比

由图3与图4可以看出, 本文算法对于人脸形状和纹理的重建都优于另外几种算法, 本文算法重建得到的三维人脸的面部轮廓以及五官形状与输入图像更为一致, 更能体现人脸的身份特征. 此外, 其它几种算法在面对不同肤色不同光照的人脸图像时重建得到的人脸纹理信息并没有较大差别, 相较而言本文算法能

够更加真实地反映人脸纹理信息. 除了人脸形状和纹理信息外, 本文算法还可以准确预测输入人脸图像的表情.

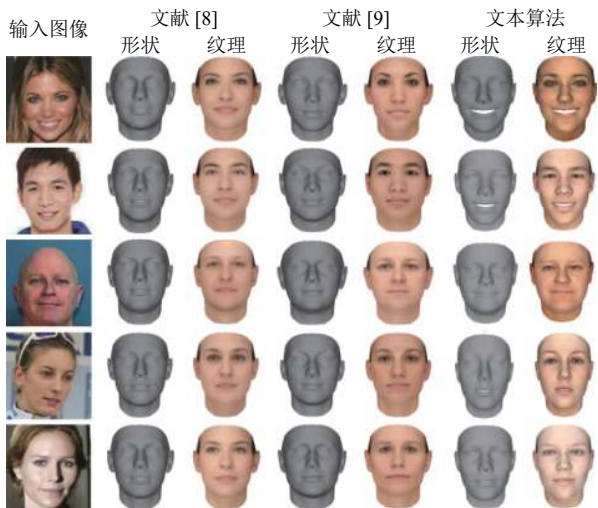


图4 部分 LFW 数据上的重建质量对比

为了验证本文算法在大姿态或遮挡情况下仍能较好完成三维人脸形状和纹理重建, 在二维人脸图像中存在大姿态或遮挡的情况与普通的纹理贴图方法进行了对比实验, 如图5所示. 可以看出, 对于存在遮挡或大姿态的人脸图像, 被遮挡部分的纹理信息并不能够从输入二维人脸图像中提取, 因此传统的纹理贴图方法将无法提取不可见区域的纹理信息, 得益于人脸解析分割算法, 在面对存在遮挡或大姿态的人脸图像时, 本文算法仍能重建出具有真实身份特征的真实感三维人脸.



图5 遮挡或大姿态情况下人脸纹理贴图与重建结果对比

最后, 为验证本文人脸重建算法的鲁棒性, 对同一身份处于不同姿态、表情、光照条件下的多张二维人脸图像进行重建, 结果如图6所示. 实验结果表明, 本

算法提出的多视角多表情的高维度人脸身份损失能够有效提升重建模型的鲁棒性, 对于同一身份在不同表情、姿势、光照条件下仍能输出较为一致的三维人脸模型.



图6 同一身份在不同表情、姿态、光照条件下的三维人脸

3.3 重建精度

为了评估本算法的重建精度, 将本算法重建得到的三维人脸模型与 MICC 数据集中的 ground truth 三维真值人脸模型进行定量的对比评估, 并与目前一些优秀的三维人脸重建算法进行了比较.

MICC 数据集中除了利用专业设备扫描得到的作为 ground truth 的真实三维人脸扫描模型外, 还包括 53 个受试者的 3 段短视频, 包括交互、室内以及室外 3 种场景. 与其它基于单张图像重建的算法类似, 本文对视频中的每一帧图像都进行三维人脸重建, 并取平均作为最终的三维人脸模型.

为公平比较本算法与文献 [6]、文献 [9] 等算法之间的重建精度, 首先采用文献 [6] 中的方法, 用球心在鼻尖, 半径为 95 mm 的球体对真实三维人脸网络进行裁剪, 从而只考虑人脸面部区域的重建精度. 随后, 采用文献 [9] 中的对齐方法, 将重建三维人脸网格与真实三维人脸网格进行密集对齐, 随后计算 point-to-plane 点对面距离误差. 表 1 展示了不同算法在 MICC 数据集中载 3 种不同视频片段下的重建三维人脸与真实三维人脸间的点对面距离的平均误差 (以 mm 为单位).

表 1 不同重建算法在 MICC 数据集上的平均误差

算法	交互场景		室内场景		室外场景	
	Mean	Std	Mean	Std	Mean	Std
文献[6]	1.93	0.27	2.02	0.25	1.86	0.23
文献[9]	1.50	0.13	1.50	0.11	1.48	0.11
本章算法	1.36	0.12	1.37	0.11	1.36	0.11

由上表结果表明, 与其它两种先进的重建算法相比, 本算法最终的重建精度取得了约 10% 的提升, 其中一个主要原因是文献 [9] 算法虽然考虑了高维度的身份损失, 但没有考虑低维度的像素级损失, 本算法不

仅通过可微分渲染器实现了一种加权光度像素损失,还实现了多姿势多表情的身份识别损失,在一定程度上提升了模型参数回归的能力。

4 结论与展望

本文提出了一种基于单张二维人脸图像的同时重建三维人脸形状和纹理信息的算法。为解决训练数据不足问题,本文设计了一种 multi-level 的弱监督损失函数,直接从大量二维人脸图像中学习多种弱监督信息。为避免遮挡对人脸纹理重建造成影响,本文利用一种人脸解析分割算法进行预处理。实验结果显示,相较于其它算法,本文算法的重建质量与重建精度均有显著提升,且本算法对二维图像中存在遮挡或大姿态等问题有较强的鲁棒性。然而,目前本方法受限于三维人脸形变模型 3DMM,重建出的三维人脸表面有些平滑,并不能很好地凸显人脸皮肤纹理细节特征,因此在接下来的工作中,将考虑研究一种更为细节的三维人脸重建算法,能够在重建过程中考虑如人脸皱纹等细节重建,进一步提升重建三维人脸的真实感。

参考文献

- 1 Abate AF, Nappi M, Riccio D, *et al.* 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 2007, 28(14): 1885–1906. [doi: [10.1016/j.patrec.2006.12.018](https://doi.org/10.1016/j.patrec.2006.12.018)]
- 2 Dornaika F, Davoine F. Head and facial animation tracking using appearance-adaptive models and particle filters. *Proceedings of 2004 Conference on Computer Vision and Pattern Recognition Workshop*. Washington, DC, USA. 2004. 153.
- 3 Scherbaum K, Ritschel T, Hullin M, *et al.* Computer-suggested facial makeup. *Computer Graphics Forum*, 2011, 30(2): 485–492. [doi: [10.1111/j.1467-8659.2011.01874.x](https://doi.org/10.1111/j.1467-8659.2011.01874.x)]
- 4 Zollhöfer M, Thies J, Garrido P, *et al.* State of the art on monocular 3D face reconstruction, tracking, and applications. *Computer Graphics Forum*, 2018, 37(2): 523–550. [doi: [10.1111/cgf.13382](https://doi.org/10.1111/cgf.13382)]
- 5 Blanz V, Vetter T. A morphable model for the synthesis of 3D faces. *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*. Los Angeles, CA, USA. 2002. 187–194.
- 6 Tran AT, Hassner T, Masi I, *et al.* Regressing robust and discriminative 3D morphable models with a very deep neural network. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2017. 1493–1502.
- 7 Richardson E, Sela M, Kimmel R. 3D face reconstruction by learning from synthetic data. *Proceedings of the 2016 4th International Conference on 3D Vision (3DV)*. Stanford, CA, USA. 2016. 460–469.
- 8 Tewari A, Zollhöfer M, Kim H, *et al.* Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. *Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy. 2017. 3735–3744.
- 9 Genova K, Cole F, Maschinot A, *et al.* Unsupervised training for 3d morphable model regression. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. 2018. 8377–8386.
- 10 Liu SC, Chen WK, Li TY, *et al.* Soft rasterizer: A differentiable renderer for image-based 3D reasoning. *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Republic of Korea. 2019. 7707–7716.
- 11 Lee CH, Liu ZW, Wu LY, *et al.* MaskGAN: Towards diverse and interactive facial image manipulation. *arXiv: 1907.11922*, 2019.
- 12 Bulat A, Tzimiropoulos G. How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230, 000 3D facial landmarks). *Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy. 2017. 1021–1030.
- 13 Paysan P, Knothe R, Amberg B, *et al.* A 3D face model for pose and illumination invariant face recognition. *Proceedings of the 2009 6th IEEE International Conference on Advanced Video and Signal Based Surveillance*. Genova, Italy. 2009. 296–301.
- 14 Cao C, Weng YL, Zhou S, *et al.* FaceWarehouse: A 3D facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 2014, 20(3): 413–425. [doi: [10.1109/TVCG.2013.249](https://doi.org/10.1109/TVCG.2013.249)]
- 15 Yu CQ, Wang JB, Peng C, *et al.* BiSeNet: Bilateral segmentation network for real-time semantic segmentation. *Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany. 2018. 334–349.
- 16 Feng ZH, Kittler J, Awais M, *et al.* Wing loss for robust facial landmark localisation with convolutional neural networks. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. 2018. 2235–2245.
- 17 Deng JK, Guo J, Xue NN, *et al.* ArcFace: Additive angular margin loss for deep face recognition. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA, USA. 2019. 4685–4694.