

基于人脸部件掩膜的自监督三维人脸重建

朱磊, 王善敏, 刘青山

引用本文

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.

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)

[基于数据增强的自监督飞行航迹预测](#)

Self-supervised Flight Trajectory Prediction Based on Data Augmentation

计算机科学, 2023, 50(2): 130-137. <https://doi.org/10.11896/jsjcx.211200016>

[一种鲁棒的双教师自监督蒸馏哈希学习方法](#)

Robust Hash Learning Method Based on Dual-teacher Self-supervised Distillation

计算机科学, 2022, 49(10): 159-168. <https://doi.org/10.11896/jsjcx.210800050>

[一种面向电商网络的异常用户检测方法](#)

Method for Abnormal Users Detection Oriented to E-commerce Network

计算机科学, 2022, 49(7): 170-178. <https://doi.org/10.11896/jsjcx.210600092>

[基于深度学习的单幅图像三维人脸重建研究综述](#)

Review of 3D Face Reconstruction Based on Single Image

计算机科学, 2022, 49(2): 40-50. <https://doi.org/10.11896/jsjcx.210500215>

[基于多判别器的多波段图像自监督融合方法](#)

Multi-band Image Self-supervised Fusion Method Based on Multi-discriminator

计算机科学, 2021, 48(8): 185-190. <https://doi.org/10.11896/jsjcx.200600132>

基于人脸部件掩膜的自监督三维人脸重建

朱磊¹ 王善敏² 刘青山¹

1 南京信息工程大学数字取证教育部研究中心 南京 210044

2 南京航空航天大学计算机学院 南京 210016

(leizhu9702@163.com)

摘要 三维人脸重建旨在从二维人脸图片中恢复出三维人脸模型。自监督三维人脸重建能够缓解三维人脸数据缺乏的问题,因此成为了近年来的研究热点。现有的自监督方法通常聚焦于使用全局监督信号,对人脸的局部细节关注不足。为了更好地恢复出细节生动的精细化三维人脸,提出了一种基于人脸部件掩膜的精细化三维人脸重建方法,该方法在不需要任何三维人脸标注的情况下,可以重建出精细化三维人脸。其主要思想是在二维图片一致性损失、图片深层感知损失等基本损失函数上,通过人脸部件掩膜,给予人脸区域精细化约束,并对人脸部件掩膜进行自监督约束,从而提高重建的三维人脸局部的准确性。在 AFLW2000-3D 和 MICC Florence 数据集上进行了定性以及定量实验,验证了所提方法的有效性和优越性。

关键词: 三维人脸重建;人脸对齐;人脸建模;自监督学习;人脸渲染

中图分类号 TP391

Self-supervised 3D Face Reconstruction Based on Detailed Face Mask

ZHU Lei¹, WANG Shanmin² and LIU Qingshan¹

1 Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044, China

2 College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Abstract Self-supervised 3D face reconstruction can alleviate the problem of lack of 3D face data, and has therefore become a hot research topic in recent years. Existing self-supervised methods usually focus on using globally supervised signals and do not pay enough attention to the local details of faces. In order to better recover fine-grained 3D faces with vivid details, this paper proposes a fine-grained 3D face reconstruction method based on face part masks, which can reconstruct fine-grained 3D faces without any 3D face annotation. The main idea is to improve the local accuracy of the reconstructed 3D face by giving refinement constraints on the face region through the face part mask and self-supervised constraints on the face part mask on top of the basic loss functions such as 2D image consistency loss, image deep perception loss, etc. Qualitative and quantitative experiments on AFLW2000-3D and MICC Florence datasets demonstrate the effectiveness and superiority of the proposed method.

Keywords 3D face reconstruction, Face alignment, Face modeling, Self-supervised learning, Face rendering

1 引言

三维人脸重建是计算机视觉领域的热点研究方向,其旨在从单张或多张二维人脸图片中重建出三维人脸。相对于二维人脸,三维人脸具有更多的信息,因此在人脸识别^[1-3]、安防安保^[4-6]、人脸编辑^[7]、VR/AR 等方面有着广泛的应用,吸引了大量研究者的关注。

基于形变模型的三维人脸重建是早期常用的方法。1999年,Blanz 等提出了一种三维人脸可形变模型(3D Morphable Models, 3DMM),此模型通过统计学分析,对三维人脸数据集进行主成分分析,将三维人脸表征为线性人脸模型,通过调节

线性形状和纹理参数即可得到对应的三维人脸,大量的三维人脸研究工作基于此展开。常用的三维人脸可形变模型有 BFM^[7],LSFM^[8]等。研究者通常在上述形变模型的基础上,不断迭代拟合 3DMM 的形变参数,从而重建出目标三维人脸的纹理和形状^[2,7]。然而,基于拟合的三维人脸重建方法迭代过程复杂,具有较高的时间复杂度,且迭代过程容易陷入局部最优,得到的三维人脸形状与纹理往往效果较差。

随着深度学习的兴起,卷积神经网络(Convolutional Neural Network, CNN)在三维人脸重建任务上取得了突破性的进展^[9-12]。一些方法^[9-12]将 CNN 与 3DMM 相结合,利用 CNN 强大的特征提取能力来直接提取人脸图像的深层特征,

到稿日期:2022-06-03 返修日期:2022-09-28

基金项目:国家杰出青年基金(61825601)

This work was supported by the National Science Fund for Distinguished Young Scholars of China(61825601).

通信作者:刘青山(qsliu@nuist.edu.cn)

并回归出相应的形状和纹理参数。除此之外,为了解决3DMM模型的线性空间约束问题,研究者们也提出基于非线性三维人脸模型的方法^[13-15]来扩充三维人脸模型表征空间。在三维人脸重建中,除了使用3DMM形变模型来表示三维人脸之外,基于UV图^[16]和体素^[17]的三维人脸表征方式也被用于三维人脸重建。

尽管上述三维人脸重建的方法取得了一定的进展,但其需要三维人脸标注作为监督信号。然而,现有三维人脸数据集的构建非常昂贵,包括场景搭建、激光扫描、后期处理等。为了避免对三维人脸标签的依赖,基于自监督的三维人脸重建方法成为了研究热点,这些方法^[18-20]通常将重建的三维人脸通过渲染器渲染至二维图像,并利用二维图像的信息与特征进行监督,从而保证了重建的三维人脸的正确性。为了提升自监督三维人脸重建的性能,现有的重建方法在二维图像一致性约束的基础上,还提出了人脸深层特征约束^[20]、人脸关键点约束^[18]等。然而,这些方法主要关注全局的纹理特征,忽略了人脸局部信息。

为了解决上述问题,本文提出了人脸部件掩膜约束来提升三维人脸重建的细节纹理。具体地,部件掩膜将人脸分成左眼、右眼、鼻子、上嘴唇和下嘴唇等部件区域,这样能够根据人脸部件不同类别的重要性,在局部人脸区域施加重约束。

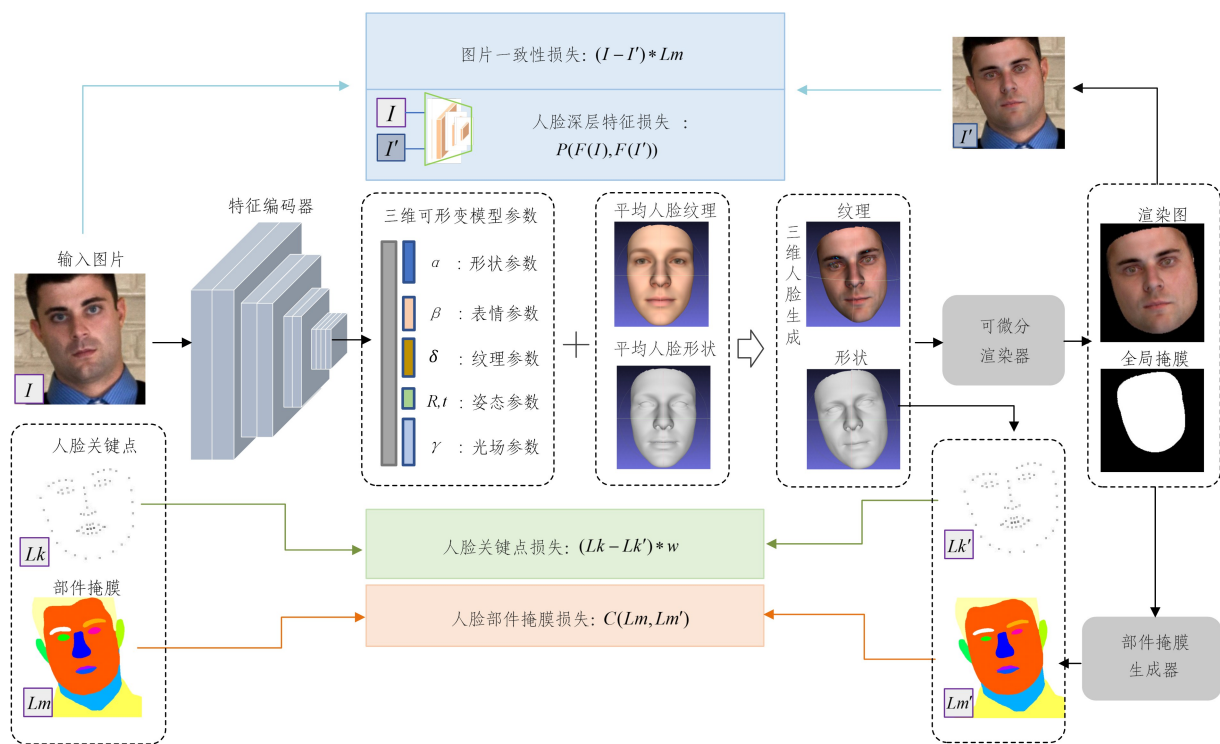


图1 基于人脸部件掩膜的自监督三维人脸重建网络框架图

Fig. 1 Framework of self-supervised 3D face reconstruction based on detailed face mask

2.1 三维人脸生成

Resnet50由4个卷积块和全连接层组成。我们采用Resnet50作为特征编码器来对人脸输入图片进行特征编码,进而回归出3DMM参数,3DMM参数包括形状参数、表情参数、纹理参数、姿态参数、光场参数等。3DMM参数和3DMM模型结合可以生成三维人脸。

三维人脸可形变模型3DMM^[7]是由Blaiz和Vetter于

此外,为了保证原始人脸部件掩膜和渲染图片人脸部件掩膜的一致性,对其进行了损失函数监督。为了验证方法的有效性,我们在AFLW2000-3D和MICC Florence两个数据集上进行了实验。定量和定性实验的结果均表明,本文方法能够有效地提高三维人脸重建细节的准确度。

2 基于人脸部件掩膜的自监督三维人脸重建

如图1所示,本文构建了一个基于人脸部件掩膜的自监督三维人脸重建网络。该网络的输入为单幅人脸图像,输出为对应的3DMM参数,并由3DMM参数生成三维人脸并渲染得到2D人脸图像。为了重建出准确、高质量的人脸,在二维输入和渲染图片上施加了一系列自监督约束损失函数。具体地,特征编码器网络主要基于Resnet50^[21]。首先,由特征编码器回归出相关的3DMM参数,即 $(\alpha, \beta, \delta, \gamma, R, t, \Pr) \in \mathbb{R}^{259}$,再将3DMM参数与模型相结合重建出三维人脸。为了在二维人脸层面对重建网络进行自监督训练,本文利用可微渲染器将得到的三维人脸渲染得到二维人脸,并分别得到二维人脸的全局掩膜和局部部件掩膜。在掩膜的权重约束下,在关键点位置、图像纹理一致性约束、部件掩膜分割和人脸深层特征感知等方面构建了损失函数,以约束三维人脸重建网络。

1999年提出的,此方法基于三维人脸扫描样本,进行PCA主成分分析,并建立起人脸向量空间,可以用线性参数在归一化的标准基人脸上进行人脸生成。3DMM人脸模型可以表示为:

$$\begin{aligned} S &= S(\alpha, \beta) = \bar{S} + B_{id}\alpha + B_{exp}\beta \\ T &= T(\delta) = \bar{T} + B_s\delta \end{aligned} \quad (1)$$

其中, \bar{S} 和 \bar{T} 为平均人脸形状和纹理; B_{id} , B_{exp} 和 B_s 为

形状、表情和纹理的主成分基;而 α, β 和 δ 为相对应的形状、表情和纹理 3DMM 参数。

本文采用了主流的 3DMM 人脸模型 BFM2009^[3] 中的平均人脸形状 \bar{S} 、人脸身份主成分基 B_{id} 、平均人脸纹理 \bar{T} 和人脸纹理主成分基 B_t , 采用了 FacewareHouse^[22] 中的人脸主成分表情基 B_{exp} , 选择的 3DMM 人脸形状、表情和纹理参数的维度分别为 $\alpha \in \mathbb{R}^{80}, \beta \in \mathbb{R}^{64}, \delta \in \mathbb{R}^{80}$ 。最终, 三维人脸模型包含 35 709 个顶点。因此, 只需要得到输入图片对应的 3DMM 参数, 由式(1)即可生成三维人脸。

2.2 人脸形状约束

三维人脸重建追求人脸形状重建的精准度, 人脸关键点刻画人脸几何形状的重要方式, 因此我们引入人脸关键点位置来作为人脸形状约束。大量的实验表明, 三维人脸对齐网络 3DFAN^[23] 可以高精度地定位人脸关键点。本文也采用该网络来自动定位人脸关键点。

由 3DMM 参数生成三维人脸后, 人脸关键点的坐标空间为三维空间, 而损失函数是在二维空间对人脸关键点进行监督。因此, 需要采用透视投影相机模型将三维人脸投影至二维平面, 再获取其二维关键点。透视投影模型的计算式如下:

$$\begin{aligned} V' &= R \times (\bar{S} + B_{id}\alpha + B_{exp}\beta) + t \\ V_p &= Pr \times V' \end{aligned} \quad (2)$$

其中, $R \in SO(3)$ 为由人脸姿态参数得到的旋转矩阵, t 为三维人脸在空间中的位移, V' 为在三维空间中经过旋转平移后的三维人脸点集, Pr 为透视投影矩阵, 而 V_p 为投影到二维平面的二维点集。得到二维投影点 V_p 后, 基于人脸关键点损失函数的定义如下:

$$L_{lmk} = \sum_{i=1}^{68} \|k_i - K(V_p)\|_1 \cdot \omega_i$$

其中, K 为从二维投影点集中得到的 68 个人脸关键点, k_i 为三维人脸对齐网络对输入图像所得到的人脸关键点, ω_i 为对应的人脸关键点权重。

2.3 图像纹理一致性约束

重建出高质量纹理是三维人脸重建性能评估的一个重要标准。因此, 利用重建所生成的人脸图像与原始图像需要保持一致性, 构造图像纹理一致性约束损失函数, 在生成三维人脸之后, 将其渲染并得到二维人脸图像, 并与原始输入图像进行度量比较。在纹理重建的过程中, 离不开光照模型和纹理渲染器。本文采用的光照模型和渲染器如下。

(1) 光照模型。三维人脸重建过程中, 基于球谐函数来估计场景光照是常见的做法, 此做法假设人脸表面为 Lambertian 表面。首先, 由三维人脸形状 V_i 计算得到人脸表面法向量 n_i , 加上三维人脸纹理 t_i , 三维人脸的每个点的光线反射纹理可由式(4)可得:

$$C(n_i, t_i | \gamma) = t_i \times \sum_{b=1}^{B^2} \gamma_b \Phi_b(n_i) \quad (4)$$

其中, 函数 $\Phi_b: \mathbb{R}^3 \rightarrow \mathbb{R}$ 为球谐基函数, γ_b 是相应的球谐基函数的参数, 与文献[18]的设置相同, 设置 B 为 3, 并设置白光为 $\gamma \in \mathbb{R}^3$ 。

(2) 渲染器。为了使梯度在网络中能够反向传播, 并与文献[24]保持一致, 使用基于 Pytorch3d^[25] 的可微分渲染器。在给定相机参数和光场参数后, 每个顶点的纹理和法向量

参数值可以通过邻近点的数值进行重心插值得到。

之前的三维人脸重建算法^[18, 20] 通常使用人脸全局掩膜, 即只分割出人脸区域和背景区域。而在此基础之上, 如图 2 所示, 人脸部件掩膜能够更加精细化地表征出人脸部件区域, 如鼻子、上嘴唇、下嘴唇、眼睛等。

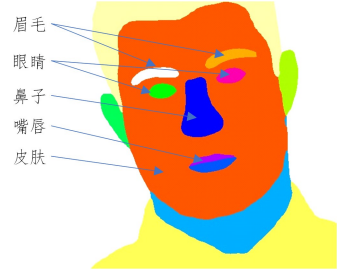


图 2 人脸部件掩膜示意图

Fig. 2 Example of detailed facial mask

Yu 等^[26] 提出的人脸分割网络 Bisenet 能够精准分割人脸区域, 本文也采用该网络来生成人脸部件掩膜。精细化的部件掩膜能够使网络在计算损失函数时根据人脸不同区域给予不同权重的约束。基于局部人脸部件掩膜的图像纹理一致性损失函数的定义如下:

$$L_{photo} = \frac{\sum_{i \in R} Lm_i \cdot A_i \cdot \|I_i - I_i'\|_2}{\sum_{i \in R} A_i} \quad (5)$$

其中, I_i 和 I_i' 分别为真实输入图片和渲染得到的图片, i 为图片像素的索引, R 为人脸区域, A 为人脸全局掩膜, $A_i = \begin{cases} 1, & \text{像素点在人脸区域内} \\ 0, & \text{像素点不在人脸区域内} \end{cases}$ 。在损失函数计算时, 眉毛、眼睛、鼻子、嘴唇、皮肤的权重设置为 1:1.2:0.8:1.5:1。人脸全局掩膜和部件掩膜联合使用, 首先全局人脸掩膜让损失函数只关注人脸区域, 避免将与人脸无关的遮挡信息带入损失计算中; 其次部件掩膜以不同比例权重更好地约束人脸特征。

2.4 人脸深层特征感知损失约束

文献[27-28]表明, 基于图像深层特征的损失函数约束能够更好地优化网络, 因此我们采用在 VGGFace2^[29] 数据集上预训练好的人脸识别网络 VGG 网络^[29] 对人脸图片进行深层特征提取。将输入图片和渲染图片输入人脸识别网络即 VGG 网络, 得到两张图片的深层特征编码, 通过度量这两者的深层特征编码的余弦距离, 来构成人脸深层特征感知损失函数, 定义如下:

$$L_{id} = 1 - \frac{P(I)P(I')}{\|P(I)\|_2 \times \|P(I')\|_2} \quad (6)$$

其中, P 为利用人脸识别网络 VGG 得到人脸深层特征编码的过程。该损失函数通过计算人脸深层特征编码的余弦距离, 来监督渲染得到的图片以捕捉人脸深层特征, 从而使渲染图片看起来更像输入图片。

2.5 人脸部件掩膜一致性约束

对于渲染图片所得到的人脸部件掩膜, 其部件分割区域的准确性反映了人脸重建的准确性。输入人脸图片和渲染图片所得到的人脸部件掩膜应该保持一致, 因此我们创新地提出人脸部件掩膜一致性约束, 损失函数的定义如下:

$$L_{seg} = - \sum_{i=0}^{H \times W \times C} Lm_i \log Lm_i' \quad (7)$$

其中, $L_m, L_m' \in \mathbb{R}^{H \times W \times C}$ 分别为输入人脸图片和渲染图片所得到的人脸部件掩膜, H, W 和 C 分别为分割图的高度、宽度以及人脸解析的类别数(如眼睛、鼻子等)。部件人脸掩膜一致性损失函数 L_{seg} 通过计算交叉熵损失来使渲染图片的掩膜分割在局部区域的分割更加准确, 更加接近输入图片的掩膜分割。

2.6 正则化约束

为了防止三维人脸重建时出现形状或者纹理退化, 3DMM 参数需要正则化约束。正则化损失函数如下:

$$L_{reg} = \omega_\alpha \|\alpha\|^2 + \omega_\beta \|\beta\|^2 + \omega_\delta \|\delta\|^2 + \omega_\gamma \|\gamma\|^2 \quad (8)$$

其中, 各项参数分别设置为: $\omega_\alpha = 10^{-4}$, $\omega_\beta = 0.8 \times 10^{-4}$, $\omega_\delta = 10^{-1}$, $\omega_\gamma = 10^{-2}$ 。

2.7 总损失函数

综上, 本文所采用的损失函数为:

$$L_{sum} = \omega_{photo} \cdot L_{photo} + \omega_{lmk} \cdot L_{lmk} + \omega_{id} \cdot L_{id} + \omega_{seg} \cdot L_{seg} + \omega_{reg} \cdot L_{reg} \quad (9)$$

3 实验

3.1 数据集介绍及实验细节

由于本文采用的是自监督学习, 因此不需要三维人脸标注的数据集来进行训练。本文采用的是 VGGFace2^[29] 和 300W-LP^[30] 数据集。VGGFace2 是一个大规模人脸识别数据集, 包括 331 万张图片和 9131 个 ID, 本文随机挑选了其中 90 万张图片作为训练数据集。300W-LP 是由 300W 人脸关键点检测数据集扩充而来, 包含 12 万张图片, 注意这里的 300W-LP 为有三维人脸标注的数据集, 但是网络训练时没有使用数据集中的三维人脸标注。

本文在 AFLW2000-3D 和 MICC Florence^[31] 上进行定量和定性实验。AFLW2000-3D 数据集是在 AFLW^[32] 数据集中挑选前 2000 张, 并估计 3DMM 参数和 68 点人脸关键点作为真实标签。MICC Florence 数据集包含 53 个人的三维激光扫描, 而每个人分别在非自发、室内和室外的情况下被采集数据。

训练网络时, 输入的图片尺寸为 224×224 , 特征编码器 Resnet50 在 ImageNet 上进行预训练。人脸部件掩膜生成器所生成的掩膜维度为 $224 \times 224 \times 17$, 采用的网络框架为 PyTorch, 优化迭代器为 Adam。网络训练时, 批大小设为 16, 学习率设定为 10^{-6} , 最终网络总共在 VGGFace2 和 300W-LP 联合完成 50 万次迭代训练。总损失函数各项系数分别为: $\omega_{photo} = 1.6$, $\omega_{lmk} = 100$, $\omega_{id} = 0.5$, $\omega_{seg} = 0.25$, $\omega_{reg} = 1$ 。

3.2 消融实验

为验证本文提出的各损失函数及提出的人脸掩膜分割模块的作用, 首先在 AFLW2000-3D 数据集上验证所提出的各个损失函数的有效性。

在 AFLW2000-3D 上, 和文献[16, 33]保持一致, 采用归一化平均误差(NME)来作为重建误差指标, NME 的具体公式如下:

$$NME = \frac{1}{N} \sum_{k=1}^N \frac{\|\mathbf{x}_k - \mathbf{y}_k\|_2}{d} \quad (10)$$

其中, \mathbf{x}_k 和 \mathbf{y}_k 分别为真实人脸关键点和预测得到的三维人脸

关键点; d 为真人人脸方框的大小; NME 值越低, 代表重建精度越高。

如表 1 所列, 在主要的 4 种损失函数逐渐叠加的情况下, 本文方法在 AFLW2000-3D 数据集上的平均 NME 值呈现逐渐变优的趋势。一般采用自监督三维人脸重建方法所采用的损失函数为表格的前三项。在此基础上, 加上提出的人脸部件掩膜损失函数之后, 数据集的平均 NME 值为 3.004%。

表 1 各个损失函数的消融实验结果

Table 1 Ablation study on each module

人脸关键点 损失	图片真实性 损失	图片深层 特征损失	掩膜分割 损失	数据集平均 NME 值/%
✓				3.272
✓	✓			3.151
✓	✓	✓		3.052
✓	✓	✓	✓	3.004

为了验证所提出人脸部件掩膜模块的作用, 我们将模型的最终结果和基线方法(没有将人脸部件掩膜加入网络)在可视化实验结果上进行实验对比。

图 3 为在两种方法下的渲染图和渲染贴图。可以明显看出, baseline 方法所生成的渲染图片缺乏纹理细节, 而由于本文方法增加了人脸部件掩膜的约束, 对人脸重点区域增加约束, 并且对人脸掩膜加以自监督掩膜分割约束, 促使网络更加关注人脸特征丰富区域, 从而使得渲染出的人脸细节更加逼真。比如第一行中少年的脸部纹理更加逼真, 第二行中的男子嘴巴附近的胡须纹理更加明显。

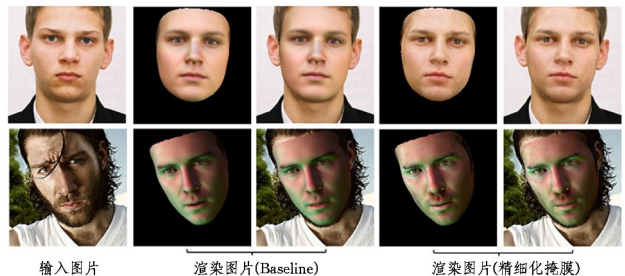


图 3 定性可视化实验结果

Fig. 3 Visualization of results

3.3 与其他方法的对比实验结果

首先在 AFLW2000-3D 和 MICC Florence 数据集上与其他方法进行定量实验对比。

如表 2 所列, 与文献[16]的工作相同, 根据数据集姿态标签, 给出了 $[0^\circ \sim 30^\circ]$, $[30^\circ \sim 60^\circ]$, $[60^\circ \sim 90^\circ]$ 这 3 个区间平均值和在数据集上的平均值结果。其中, 带有 * 的对比方法的训练方式为有三维数据标签训练, 而没有标注 * 的方法为自监督学习的方式。在对比方法中, 本文的结果在 $[0^\circ \sim 30^\circ]$ 区间的 NME 值已经取得最好的结果, 在 $[30^\circ \sim 60^\circ]$ 区间的结果也取得了次优的结果。综合看来, 本文方法在 3 个区间的平均值已经超过大部分采用有真实三维标签方式训练的方法。区间平均值取得最好结果的 MGCnet 则采用的是多视角图片输入, 因此在中等姿态和大姿态的区间, MGCNet 的表现优异。另外, SADRNet 则是采用带有真实三维标签的数据集进行训练。

表2 各方法在 AFLW2000-3D 数据集上的实验结果

Table 2 Performance comparison on AFLW2000-3D database

方法	[0°~30°]	[30°~60°]	[60°~90°]	区间平均值/%	数据集平均值/%
SDM ^[34] [ECCV16]*	3.67	4.94	9.67	6.12	—
3DDFA ^[35] [TPAMI17]*	3.43	4.24	7.17	4.94	6.03
3DTSN ^[36] [ICCV17]*	3.15	4.33	5.98	4.49	4.36
DeFA ^[37] [ICCV17]*	—	—	—	4.50	3.26
PRNet ^[16] [ECCV18]*	2.75	3.51	4.61	3.62	—
MGCNet ^[ECCV20]	2.72	3.12	3.76	3.20	—
SADRNet ^[TIP21] *	2.66	3.30	4.42	3.46	3.05
Ours	2.44	3.27	5.10	3.60	3.00

由于数据集在 3 个区间的分布数量是不平衡的,因此本文还给出了在数据集上整体的均值 NME 结果,本文方法取得了最好的结果。综合看来,所提方法虽然在大姿态角度的重建精度不是很好,但是在小姿态的情况下还是取得了比较优异的重建精度。

在 MICC Florence 数据集上,采取了与文献[38]一致的评测方式,随机从数据集视频中挑选 5 帧图片,并送入到网络用于生成对应的三维人脸结果(mesh)。与文献[39]一致,对生成得到的三维人脸结果在鼻尖处以 95ms 为半径进行裁剪,然后以最近邻迭代^[40](Iterative Closest Point, ICP)的方式将得到的三维人脸 mesh 再与真实人脸 mesh 进行对齐,最终以归一化点到面的误差来衡量重建的精准度。

表 3 列出了在 MICC Florence 数据集上的实验结果,本文方法取得了次优的结果,排名第一的为 Gecer 所提出的 GANFit^[38]网络。而 GANfit 网络是以基于生成对抗网络(GAN)的拟合方式来进行训练的,不具备鲁棒性。

表3 各方法在 MICC Florence 数据集上的实验结果

Table 3 Performance comparison on MICC Florence database (单位:mm)

方法	Cooperate	Indoor	Outdoor
Tran ^[39]	1.93±0.270	2.02±0.250	1.86±0.230
Booth ^[41]	1.82±0.290	1.85±0.220	1.63±0.160
Genova ^[24]	1.50±0.130	1.50±0.110	1.48±0.110
Gecer ^[38]	0.95±0.107	0.94±0.106	0.94±0.106
Ours	1.18±0.270	1.17±0.280	1.19±0.290

3.4 可视化实验结果

此外,本文还在 AFLW2000-3D 上进行了可视化实验,并与 PRNet^[16]和 3DDFAv2^[42]进行了对比。实验结果如图 4 所示,在第一行的输入图片的结果中,3DDFAv2 的重建结果比较逼真,这是由于 3DDFAv2 在重建三维人脸时,纹理的生成是直接由输入图片采样得到的,因此可以看到,3DDFAv2 的生成结果有头发遮挡的效果,而本文的实验结果没有出现这类现象。在第二行的输入图片具有大姿态和遮挡的情况下,3DDFAv2 由于从输入图片中获取纹理信息,因此重建的效果较差。综合来看,相比前两者,在大姿态和遮挡的情况下,本文方法仍然能够重建出较高质量的结果。



图4 三维重建可视化实验结果

Fig. 4 Visualization comparison of reconstructed 3D faces

如图 5 所示,本文分别提供了渲染图片、渲染贴图 and 三维人脸纹理图。从图中可以看出,本文重建出的人脸与原始输入图片在不同性别、不同姿态下的场景,均能够达到十分相似的程度。

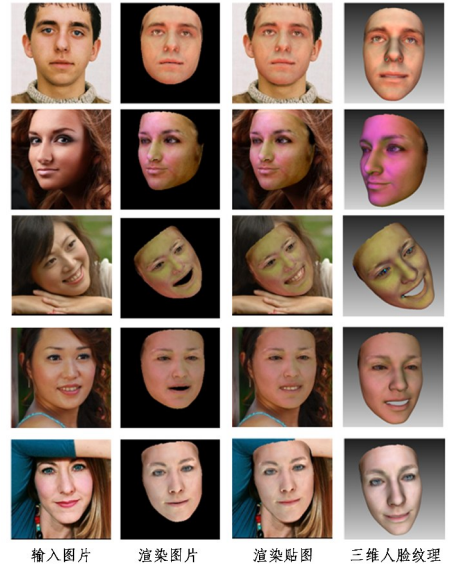


图5 三维重建可视化实验结果

Fig. 5 Visualization of reconstructed 3D face

结束语 自监督三维人脸重建是当下研究的热点之一,其能够缓解三维人脸数据缺乏的问题。现有的自监督方法通常聚焦于使用全局监督信号,对人脸的局部细节关注不足。为了恢复出细节生动的精细化三维人脸,本文提出了一种基于人脸部件掩膜的精细化三维人脸重建方法,该方法在不需要任何三维人脸标注的情况下,可以重建出精细化三维人脸。其主要思想是在二维图片一致性损失、图片深层感知损失等基本损失函数上,通过人脸部件掩膜,给予人脸区域精细化约束,并对人脸部件掩膜进行自监督约束,从而提高重建的三维人脸局部的准确性。在 AFLW2000-3D 和 MICC Florence 数据集上进行了定性以及定量实验,验证了所提方法的有效性和优越性。

目前本文使用的三维人脸模型为 BFM 模型,仅仅包括人脸部分、缺乏脖子部分以及重建难度更大的头发部分。未来的研究方向为如何对人的头部进行整体重建,另外就是如何对人的头发进行精确重建。

参考文献

- [1] BLANZ V, ROMDHANI S, VETTER T. Face identification across different poses and illuminations with a 3d morphable

- model[C]//Proceedings of fifth IEEE International Conference on Automatic Face Gesture Recognition. IEEE,2002;202-207.
- [2] BLANZ V,VETTER T. Face recognition based on fitting a 3d morphable model[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2003,25(9):1063-1074.
- [3] PAYSAN P,KNOTHE R,AMBERG B, et al. A 3D face model for pose and illumination invariant face recognition[C]//2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE,2009;296-301.
- [4] XIONG X,DE LA TORRE F. Global supervised descent method [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
- [5] YU Z,LI X,NIU X, et al. Face anti-spoofing with human material perception[C]//European Conference on Computer Vision. Springer,2020;557-575.
- [6] QIN Y,ZHAO C,ZHU X, et al. Learning meta model for zero-and few-shot face anti-spoofing[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019.
- [7] 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. 1999.
- [8] BOOTH J,ROUSSOS A,PONNIAH A, et al. Large scale 3d morphable models[J]. International Journal of Computer Vision,2018,126(2):233-254.
- [9] YI H,LI C,CAO Q, et al. Mmface:A multi-metric regression network for unconstrained face reconstruction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- [10] TRẦN A T,HASSNER T,MASI I, et al. Extreme 3d face reconstruction:Seeing through occlusions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [11] JIANG L,WU X J,KITTLER J. Dual attention MobDenseNet (DAMDNet) for robust 3D face alignment[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.
- [12] 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. 2020.
- [13] TRAN L,LIU X. Nonlinear 3d face morphable model[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [14] BAGAUTDINOV T,WU C,SARAGIH J, et al. Modeling facial geometry using compositional vaes [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [15] TRAN L,LIU F,LIU X. Towards high-fidelity nonlinear 3D face morphable model[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- [16] FENG Y,WU F,SHAO X, et al. Joint 3d face reconstruction and dense alignment with position map regression network [C]//Proceedings of the European Conference on Computer Vision(ECCV). 2018.
- [17] 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. 2017.
- [18] 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. 2019.
- [19] TEWARI A,ZOLLHÖFER M,GARRIDO P, et al. Self-supervised multi-level face model learning for monocular reconstruction at over 250 HZ[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [20] 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(TOG),2021,40(4):1-13.
- [21] HE K,ZHANG X,REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
- [22] 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,2013,20(3):413-425.
- [23] 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) [C] // Proceedings of the IEEE International Conference on Computer Vision. 2017.
- [24] GENOVA K,COLE F,MASCHINOT A, et al. Unsupervised training for 3d morphable model regression[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [25] RAVI N,REIZENSTEIN J,NOVOTNY D, et al. Accelerating 3d deep learning with pytorch3d[J]. arXiv:2007.08501,2020.
- [26] YU C,WANG J,PENG C, et al. Bisenet:Bilateral segmentation network for real-time semantic segmentation[C]//Proceedings of the European Conference on Computer Vision(ECCV). 2018.
- [27] JOHNSON J,ALAHY A,FEI FEI L. Perceptual losses for real-time style transfer and super-resolution[C]//European Conference on Computer Vision. Springer,2016;694-711.
- [28] LEDIG C,THEIS L,HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
- [29] CAO Q,SHEN L,XIE W, et al. Vggface2:A dataset for recognising faces across pose and age[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE,2018;67-74.
- [30] 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. 2016.
- [31] BAGDANOV A D,DEL BIMBO A,MASI I. The florence 2d/3d hybrid face dataset[C] // Proceedings of the 2011 Joint ACM

- Workshop on Human Gesture and Behavior Understanding, 2011.
- [32] KOESTINGER M, WOHLHART P, ROTH P M, et al. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization[C]//2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). IEEE, 2011: 2144-2151.
- [33] TU X, ZHAO J, JIANG Z, et al. 3D Face Reconstruction from A Single Image Assisted by 2D Face Images in the Wild[J]. IEEE Transactions on Multimedia, 2021, 23: 1160-1172.
- [34] MCDONAGH J, TZIMIROPOULOS G. Joint face detection and alignment with a deformable hough transform model[C]//European Conference on Computer Vision. Springer, 2016: 569-580.
- [35] ZHU X, LIU X, LEI Z, et al. Face alignment in full pose range: A 3d total solution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 41(1): 78-92.
- [36] BHAGAVATULA C, ZHU C, LUU K, et al. Faster than real-time facial alignment: A 3d spatial transformer network approach in unconstrained poses[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017.
- [37] LIU Y, JOURABLOO A, REN W, et al. Dense face alignment [C] // Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017.
- [38] 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, 2019.
- [39] TRAN A T, HASSNER T, MASI I, et al. Regressing robust and discriminative 3D morphable models with a very deep neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [40] BESL P J, MCKAY N D. Method for registration of 3-D shapes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(2): 239-256.
- [41] BOOTH J, ANTONAKOS E, PLOUMPIS S, et al. 3d face morphable models "in-the-wild"[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [42] GUO J, ZHU X, YANG Y, et al. Towards fast, accurate and stable 3d dense face alignment[C]//European Conference on Computer Vision. Springer, 2020: 152-168.



ZHU Lei, born in 1997, postgraduate. His main research interest is 3D face reconstruction.



LIU Qingshan, born in 1975, Ph.D., professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include image and vision analysis, pattern recognition, and machine learning.

(责任编辑:喻藜)