

基于深度学习的单幅图像三维人脸重建研究综述

何嘉玉¹ 黄宏博¹ 张红艳¹ 孙牧野¹ 刘亚辉² 周哲海³

1 北京信息科技大学计算机学院 北京 100192
2 北京信息科技大学信息管理学院 北京 100192
3 北京信息科技大学仪器科学与光电工程学院 北京 100192
(hejiayu_97@163.com)

摘要 在计算机视觉领域中,三维人脸重建是一个具有研究价值的方向,高质量地重建出三维人脸在人脸识别、防伪、游戏娱乐、影视动画和美容医疗等领域具有重要的意义。近二十年来,虽然基于单幅图像的三维人脸重建领域已经取得很大的进展,但使用传统算法进行重建的结果仍会受到人脸表情、遮挡、环境光的影响,并且会出现重建效果精度不佳和鲁棒性不足等问题。随着深度学习进入三维人脸重建领域,各种优于传统重建算法的方法相继出现。文中首先重点介绍了基于深度学习的单幅图像三维人脸重建算法,将算法按不同的网络架构分为4类,并对各类最具有代表性的方法进行了详细阐述。然后汇总了基于单幅图像的三维人脸重建算法常用的三维人脸数据集,并在数据集上对具有代表性的方法进行了性能评估。最后对基于单幅图像的三维人脸重建领域进行了总结与展望。

关键词: 三维人脸重建;深度学习;卷积神经网络;三维可变形模型;人脸识别
中图法分类号 TP391

Review of 3D Face Reconstruction Based on Single Image

HE Jia-yu¹, HUANG Hong-bo¹, ZHANG Hong-yan¹, SUN Mu-ye¹, LIU Ya-hui² and ZHOU Zhe-hai³
1 School of Computing, Beijing Information Science and Technology University, Beijing 100192, China
2 School of Information Management, Beijing Information Science and Technology University, Beijing 100192, China
3 School of Instrument Science and Optoelectronic Engineering, Beijing Information Science and Technology University, Beijing 100192, China

Abstract In the field of computer vision, 3D face reconstruction is a valuable research direction. High quality reconstruction of 3D faces can find applications in face recognition, anti-proofing, animation and medical cosmetology. In the last two decades, although great progress has been made 3D face reconstruction based on a single image, the results of reconstruction using traditional algorithms are still facing the challenge of facial expression, occlusion and ambient light, and there will be problems such as poor reconstruction accuracy and robustness. With the rapid development of deep learning in 3D face reconstruction, various methods which are superior to traditional reconstruction algorithms have emerged. Firstly, this paper focuses on deep-learning-based reconstruction algorithms. The algorithms are divided into four categories according to different network architecture, and the most popular methods are described in detail. Then commonly used 3D face data sets are introduced, and performance of representative methods are evaluated. Finally, conclusions and prospects of the single-image-based 3D face reconstruction are given.
Keywords 3D face reconstruction, Deep learning, Convolutional neural networks, 3D morphable model, Face recognition

1 引言

近年来,三维人脸重建技术发展迅速,一方面是由于二维图像信息的局限性,使其包含的身份特征信息相比三维数据不够全面,因此很多学者尝试使用三维数据进行人脸的研究;另一方面,三维人脸重建也具有广泛的实际应用价值,在人脸

识别、游戏娱乐以及医疗美容等领域都有很好的应用前景。
目前三维人脸数据在实际应用中还存在采集不便、数据量大、易受外界条件干扰等不足,考虑到二维图像采集的普遍性和便捷性,基于二维人脸图像进行三维重建,构建出三维人脸数据成了为一个值得探索的方向。
三维人脸重建的研究方法可以分为基于多幅图像的三维

人脸重建与基于单幅图像的三维人脸重建,基于多幅图像的三维人脸重建能够获取人脸多个角度的信息,具有更好的准确性和鲁棒性,而基于单幅图像的三维人脸重建更便捷、快速、易于实现,因此具有更广阔的前景。

与基于多幅图像的人脸重建相比,基于单幅图像的人脸重建可获取的信息量受到较多限制,这也是其面临的主要挑战。具体来说,主要体现在以下几个方面:

(1)当使用多幅图像进行三维人脸重建时,可以从不同的视角获取人脸的形状信息,因此该模型建立的三维人脸的精度较高;而使用单幅图像重建时,仅能够从一个视角提取人脸的三维信息,因此其难度要远大于多幅图像。

(2)使用多幅图像进行三维人脸重建时可以通过三角测量来估计人脸特征的深度信息,而使用单幅图像进行三维人脸重建时求解空间大,是一个病态问题,因此难以准确学习到人脸的深度信息。

(3)基于单幅图像进行三维人脸重建时,人脸的解集存在因投影带来的多义性,因此该算法需要先验知识来辅助其获取度量重建,如何设定模型的假设,以恰当的方式引入先验知识也是研究的主要难点。

传统的三维人脸重建算法主要依赖于通用人脸模型,其通过关键点进行三维重建,这种方法虽然能够快速重建出三维人脸,但是在计算过程中会丢弃大量有用信息,导致重建结

果整体上比较粗糙。由于卷积神经网络能够很好地学习二维图像中多层次、多尺度的特征,因此基于深度学习的单幅图像三维人脸重建效果在重建精度和鲁棒性上均有显著提升。

近年来,基于深度学习的单幅图像三维人脸重建吸引了大量研究,并取得了丰硕的成果。本文在较全面地分析这些工作的基础上,系统地梳理了相关算法,将基于深度学习的单幅图像三维人脸重建算法按照神经网络架构的不同分为基于卷积神经网络(Convolutional Neural Networks,CNN)的三维人脸重建算法、基于图卷积网络(Graph Convolutional Network,GCN)的三维人脸重建算法、基于自编码器(Auto Encoders,AE)的三维人脸重建算法和基于生成对抗网络(Generative Adversarial Network,GAN)的三维人脸重建算法4个类别;详细地阐述了每种方法的思想、步骤及特点;在主流评测数据集和平台上对具有代表性的算法进行了性能评估;对基于深度学习的单幅图像三维人脸重建领域未来面临的挑战和发展方向进行了分析和展望。

2 相关工作

2.1 三维人脸重建算法的发展历程

基于单幅图像的三维人脸重建领域发展至今已存在大量的算法,图1以时间顺序给出了各个时间节点上具有代表性的算法。

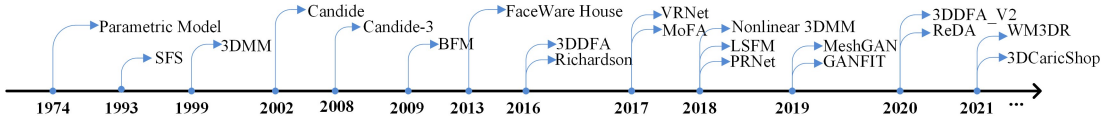


图1 单幅图像的三维人脸重建算法在各时间节点的典型算法

Fig. 1 Representative algorithm of single image 3D face reconstruction algorithm at each time node

基于单幅图像的三维人脸重建算法主要有3个类别:第一个类别是基于三维可变形模型(3D Morphable Model, 3DMM)的三维人脸重建方法;第二个类别是基于深度学习的三维人脸重建方法;第三个类别是融合三维可变形模型与深度学习的三维人脸重建方法。本文主要从使用深度学习模型的角度来介绍基于单幅图像的三维人脸重建方法。

2.2 传统的三维人脸重建算法

传统的三维人脸重建算法主要是基于三维可变形模型^[1]进行的,该模型由 Blanz 和 Vetter 于 1999 年提出,是一个基于统计学的三维人脸模型。该模型的核心思想是将人脸的面部空间视为线性空间,将在数据集中计算得到的人脸的平均形状与平均纹理作为人脸空间的线性基常量,把三维人脸重建问题转化为输入图像在线性基下的系数求解问题。该算法将人脸的形状和纹理分别用向量 S_i 和 T_i 表示,平均形状和平均纹理分别用 \bar{S} 和 \bar{T} 表示,然后用以下的方式来重建输入图像对应的三维人脸:

$$S_{\text{model}} = \bar{S} + \sum_{i=1}^{m-1} \alpha_i S_i \quad (1)$$

$$T_{\text{model}} = \bar{T} + \sum_{i=1}^{m-1} \beta_i T_i \quad (2)$$

其中, S_i 和 T_i 分别表示人脸形状和纹理的主成分分析向量, α_i 和 β_i 分别表示人脸形状和纹理的主成分分析向量所对应的系数。三维可变形模型的目的是求解出最优的形状系数和纹理

系数,生成与输入图像相对应的三维人脸。三维可变形模型重建的流程如图2所示。

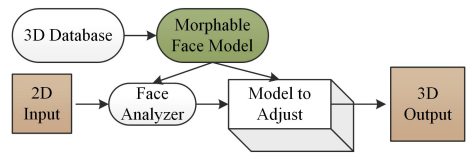


图2 3DMM模型的流程图^[1]

Fig. 2 Flow chart of 3DMM^[1]

若要提高三维人脸重建算法的精确度,则需要将3D扫描仪扫描出的三维人脸数据作为算法训练的数据集。在三维人脸重建领域发展之初,为了解决因数据的空白而导致的三维人脸重建效果不佳的问题,Paysan等^[2]于2009年发布了公开可用的巴塞尔人脸模型(BFM),其基本原理仍然是三维可变形模型。该模型的数据集包括200个自然面孔的3D扫描人脸,通过为BFM提供更高精度的形状和纹理来重建三维人脸,实验结果表明,使用巴塞尔人脸模型重建出的三维人脸的质量超出了当时的其他模型。与BFM解决人脸数据量不足的问题的初衷类似,大规模人脸模型(Large Scale Facial Model,LSFM)^[3]的提出也是为了解决此问题。LSFM模型的数据集包含了9663条三维人脸的数据,其中不仅有普通常见的人脸,还存在着大量的长相独特的人脸。该模型从特殊

面孔的重建性能方面,对 3D 人脸重建做出了新的改进。

三维可变形模型的提出极大地降低了人脸重建的复杂度,对于多数情况下的人脸都可以重建出很好的效果,但是该算法对输入图像中人脸的姿态和表情有较强的约束,降低了算法的适用性。Faceware House 实验室^[4]在 3DMM 的基础上提出了一种具有表情建模能力的双线性人脸模型,该模型的表现力相比 3DMM 得到了很大的提高。该模型在训练时使用了具有表情的人脸作为训练数据,尽管数据集中的人脸仍是在有约束条件下采集的,但是其引入人脸表情的思想促进了三维人脸重建算法的发展。

当面对无配合采集的人脸图像时,大多数三维重建算法对人脸形状和纹理的重建仍有较大困难,Booth 等^[5]首次提出了在无约束条件下进行三维人脸重建的算法,该算法在 3DMM 中学习了改进的人脸纹理特征,简化了算法的拟合策略。

另外,尽管三维可变形模型将人脸的面部空间转换为形状和纹理的线性表示简化了人脸的重建过程,但是有限的线性解空间会导致模型表达力较弱以及重建出的人脸存在面部局部区域失真和模糊等问题。针对上述问题,Vlasic 等^[6]开创性地提出了双线性模型,在身份和表情两个维度上对人脸模型进行参数化;Yang 等^[7]将身份和表情的双线性模型应用于人脸形状的拟合。上述两种算法均提高了重建模型的表现力,有效地弥补了三维可变形模型的缺陷。

3 基于深度学习的三维人脸重建算法

近年来,深度学习在计算机视觉等领域取得了令人印象深刻的效果,在三维人脸重建工作中也出现了该领域的大量研究工作,使基于深度学习的三维人脸重建成为了目前的热点研究方向。本文将基于深度学习的三维人脸重建算法按照不同的网络架构划分为 4 个类别,然后分别阐述了每个类别中代表性算法的思想及其实现过程。

3.1 基于卷积神经网络的三维人脸重建算法

经典的三维可变形模型在进行基于单幅图像的三维人脸重建时,能够很好地恢复其全局几何信息,具有快速、计算量小的特点,但该模型在人脸局部细节方面描述得还不够准确,导致重建出的三维人脸的精度相对较低。而卷积神经网络因具有较好的特征表达和学习能力而被引入到单幅图像的三维人脸重建领域^[8]。

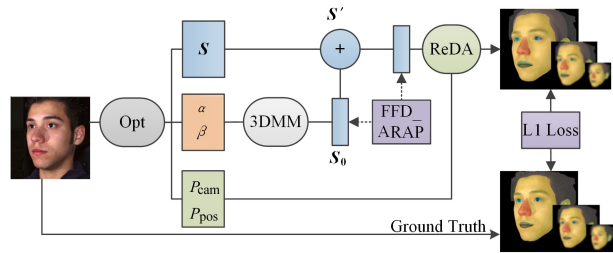
但是,相比传统的三维可变形模型,卷积神经网络存在算法复杂度高、缺乏先验知识等不足。如何把卷积神经网络和三维可变形模型相结合,重建出较好的三维人脸成为了研究者们比较关注的问题。Richardson 等^[9]提出了一种端到端的三维人脸重建算法,该算法由两个网络模块组成,分别是 CoarseNet 和 FineNet。其中,CoarseNet 模块通过 3DMM 计算得到三维人脸的系数,其作用是初步恢复出粗糙的人脸几何,FineNet 模块的作用是将 CoarseNet 模块得到的粗糙人脸几何进行细化,两个模块通过连接层来连接。实验结果证明,该算法不仅重建出了高精度的三维人脸,还加快了算法的运行速度。类似地,Fan 等^[10]提出的双神经网络结构算法、SC-NN^[11]算法以及 UH-E2FAR^[12]算法等均结合了卷积神经网络

与三维可变形模型的优点,使重建出的三维人脸具有较好的表现力。

在三维人脸重建的过程中,若训练数据具备精确的标注信息,则会降低模型重建的复杂度。而许多算法根据标注信息进行三维重建时会存在或多或少的误差,其原因是在进行图像对齐时容易存在差错。当对无约束或者低分辨率以及有遮挡的图像进行人脸特征点检测时,即使是专业的人工标注员也难以准确地对其进行定位。另外,当图像中的人脸处于不同姿态时,对特征点标记的难度也会提高,这会对重建效果产生较大的影响。Zhu 等^[13]提出了 3DDFA 算法来进行密集人脸对齐,该算法将特征点标注问题转化为图像与模型拟合的问题,通过将三维人脸模型拟合到图像上,计算出输入图像特征点的位置。该算法可以解决由三维变换引起的外观变换和自遮挡问题,对大姿态人脸和模糊图片等情况均表现出很强的鲁棒性。

为了在准确性和速度方面对 3DDFA 算法进行进一步优化,MT-CNN(Multi-Task CNN)^[14]算法使用级联多任务的卷积神经网络来联合回归三维人脸形状。该算法在第一阶段运用 MT-CNN 联合估计出三维人脸形状、二维特征点位置和人脸的姿态,然后将二维特征点位置信息提取到的 MSIFs (Modified Shape Indexed Features)作为第二阶段 MT-CNN 的输入,最终回归出目标三维人脸。实验结果表明,该算法具有精确的面部特征点检索功能和高效的计算率。除此之外,研究人员还做了很多其他的尝试^[15],如不借助二维图像特征点信息来回归出三维人脸的形状和纹理参数^[16]、预测出人脸的姿态参数^[17]和表情参数^[18]等。Wu 等^[19]提出了一种直接从大量无三维标签的二维人脸数据中重建三维人脸的算法,针对训练数据的不足,该算法设计了一种弱监督损失函数,使其可以仅根据二维人脸图像来进行模型的学习。另外,该算法通过缩小渲染人脸与输入人脸间的差距,来达到重建出具有真实感纹理的目的。

在进行图像与模型的对齐时,有很多工作严重地依赖先验知识来减少深度模糊,这种方法虽然取得了不错的效果^[20],但由于 3DMM 有限的解空间限制了其表达能力,因此该算法仍有很大的改进空间^[21]。另外,在进行图像渲染时,很多工作通常是直接使用 Z-buffer 进行渲染的,但是由于其并不一定是可微的,因此会使其产生一定的误差。Zhu 等^[22]设计了 ReDA 图像拟合算法来解决上述问题,该算法的整体框架如图 3 所示。该算法通过优化网络 Opt 回归出形状残差 S 以及 3DMM 所需的形状、表情参数 α 和 β 、 P_{cam} 和 P_{pos} 分别表示摄像机的投影矩阵和位姿。通过初始化形状和纹理参数生成初始人脸 S_0 , S' 表示修正后的人脸,其中 $S' = S_0 + S$,该过程称为自由变形,其作用是减小初始三维人脸 S_0 和最终三维人脸 S' 间的距离,设计该过程的目的是解决 3DMM 容量的限制问题,以确保网格有足够的空间来处理输入图像。FFD-ARAP 表示自由变形的过程中产生的损失。ReDA 图像拟合算法通过计算渲染图像和相应的真值之间的差异损失来进行优化。

图3 ReDA算法的框架图^[22]Fig. 3 ReDA algorithm framework^[22]

为了实现增强的可微渲染,ReDA 算法可将微属性从单一的颜色属性扩展到深度以及人脸解析掩码等更广泛的属性,使其能够进行更加有效的密集学习。另外,ReDA 借助软栅格化的思想,将网格沿 Z 方向分割成若干块,每一块都采用传统的方法进行栅格化,然后使用一些具有不同核尺寸的二维卷积运算沿空间和跨层进行聚合。实验结果表明,ReDA 算法能够建立较好的密集对应关系,重建出高质量的三维人脸。

精确的三维人脸重建往往需要采用迭代的算法不断进行优化,这导致模型的计算成本很高,不利于实时或移动设备上的应用。针对这个问题,研究人员做了很多尝试^[23]。MobileFace^[24]是一个能在移动设备上实时进行三维人脸重建的算法,该算法基于 MobileNet CNN 等轻量级网络来回归三维人脸的形状,并将人脸的纹理引入到能量函数中,以达到提高算法准确率的目的。实验结果表明,该算法在保持重建精度的同时,模型的速度和占内存空间的大小均有显著改善。同样,为了加快算法的运行速度,保持计算的实时性,3DDFA_V2^[25]也引入了轻量级网络结构。3DDFA_V2 是 3DDFA^[13]算法的改进,将其二维人脸关键点的标记问题转化为图像与三维模型间的拟合问题。与 3DDFA 不同的是,3DDFA_V2 算法引入了轻量级网络,并设计了元联合优化策略进行三维可变形模型参数的回归,相比 3DDFA 算法,该算法在精确度与运行速度方面有进一步的提升。

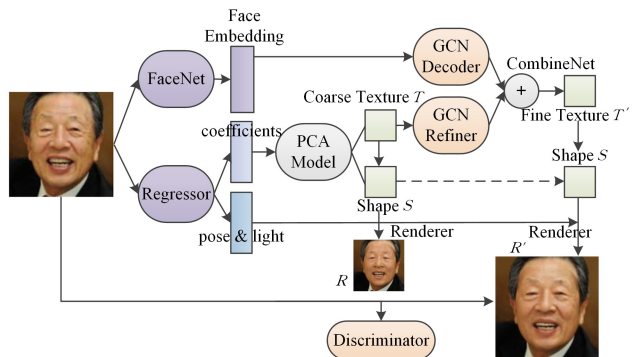
VRNet^[26]使用体素的方式来表示重建出的三维人脸,该算法不需要人脸对齐和建立密集对应关系,它采用了端到端的训练方式直接回归出单张二维图像所对应的三维人脸,解决了以往算法中存在的只有约束人脸的姿态和光照条件等因素,才能重建出高质量的三维人脸的问题。

基于卷积神经网络的算法通常会依赖大量的标记数据,但事实上公开可用的带有标注的三维人脸数据较少,这也是基于深度学习的算法进行三维人脸重建面临的主要障碍之一。针对这个问题,研究人员设计了很多增广数据的方法,把合成的人脸数据和真实人脸数据组合进行模型训练。实践中用于训练网络的数据分为以下 3 类:真实的三维人脸数据^[27]、合成的三维人脸数据^[28]以及二者按不同比例的组合数据^[20]。Deng 等^[29]认为由于训练标签不完整等的影响,合成数据的准确性可能会有所下降。因此,他们提出了利用弱监督学习来进行三维人脸重建的算法。该算法引入了一个将像素色彩信息和感知损失信息组合的损失函数,利用混合的图像信息进行弱监督学习,在进行重建时不需要三维人脸的真实值数据。基于弱监督学习的三维人脸重建算法对数据的要求较低,对模型训练友好,有望成为未来发展的一个重要方向。

3.2 基于图卷积网络的三维人脸重建算法

卷积神经网络在计算机视觉领域展示了其特有的优势,可以有效地提取图像的特征进行学习。但适用于卷积神经网络处理的视觉数据中的像素点排列通常为整齐的矩阵形式,而人脸特征点的排列呈非规则式网络结构,为此研究人员考虑引入适用于处理网状拓扑结构的图卷积网络来进行三维人脸重建。

CoMA(Convolutional Mesh Autoencoder)^[30]是一种网格卷积自编码器,能够在神经网络中保存不同尺度网格特征的拓扑结构。该自编码器的目的是将卷积神经网络在结构化数据上的应用成功推广到图结构或网格结构的数据上。受 CoMA 的启发,Lin 等提出了一种基于图卷积网络的三维人脸重建算法^[31],该算法的框架如图 4 所示,其目的是重建出具有高保真纹理的三维人脸。该算法通过将输入图像中的面部细节引入到面部网格中的方法,来进一步细化人脸的初始纹理。其中,Regression 模块根据输入图像 I 回归出 3DMM 的形状和纹理系数 coefficients 以及人脸的姿态和光照参数 pose&light。将 FaceNet 提取到的特征向量和纹理系数生成的人脸纹理分别输入到 GCN Decoder 和 GCN Refiner 中,二者的输出沿着通道轴输入到 Combine Net 中,最终得到一个精细的人脸纹理 T' 。Discriminator 的作用是通过对抗训练来提高纹理细化模块的输出效果。

图4 基于图卷积网络的三维人脸重建算法的框架示例^[31]Fig. 4 Example of 3D face reconstruction algorithm framework based on graph convolution network^[31]

在卷积神经网络领域,很多研究人员致力于设计算法,用于漫画和动画领域的三维人脸重建,同样,在图卷积神经网络领域也有此类算法存在。与真实世界相比,漫画通常运用抽象或夸张的面部特征来描绘特定的个体,因此其重建难度也相对较高。为了保证漫画人脸重建的精确度与鲁棒性,Qiu 等^[32]结合了隐式函数存在的对任意拓扑的对象都有较好性能参数化网格模型能够高效重建出三维人脸的优势,提出了一种将隐式函数重建出的高保真几何人脸转移到一个具有拓扑结构的模板网格中的方法。在转移过程中,为实现精确的网格对齐,他们提出了一种视图协同图卷积网络(VC-GCN),用于从隐式网格中提取关键点。该部分首先由 3 个渲染视图预测的二维关键点信息来获取初始的三维关键点位置,然后根据级联的 VC-GCN 将 3 个视图中的三维关键点位置进行融合,进而将融合后的三维关键点信息发送到 GCN 中进行解耦,重新得到 3 个视图的关键点信息,新的三维关键点信息在不断的融合与解耦过程中得到优化,

进而实现精确的网格对齐。

3.3 基于自编码器结构的三维人脸重建算法

三维人脸的生成模型可以引入先验信息,通过提供人脸模型的参数表示,使任务可以区分由身份、表情或年龄等自然因素造成的变化。一个理想的生成模型应该具有能利用所有可用的训练数据、能处理被噪声破坏的数据和很少或不需要进行预处理等特点。Abrevaya 等^[33]提出了一个多线性自编码器框架来实现上述目标。该算法将基于卷积神经网络的编码器和多线性解码器结合,使其具有对有损数据和不完整数据的鲁棒性以及多线性模型建模与解耦人脸形状变化的优势。该算法还能够从不同的数据集中泛化学习生成模型。实验结果表明,该弱监督学习算法优于其他类型的弱监督重建算法。

对于单幅图像中包含多张人脸的情形,传统的人脸重建算法通常是先对图像中的人脸进行检测和对齐,然后实现人脸的重建,但这会导致大量的重复计算。另外,传统方法通常会对图像中的每个人脸分别估计相机参数,但这种独立的估计会忽略全局人脸的头部位置和方向。WM3DR 算法^[34]为了解决上述问题,提出了一种端到端的弱监督单图像多人脸重建框架。该框架可以通过神经网络的单次前向传递来联合恢复出多人脸,通过自编码器网络结构同时预测出图像中人脸的位置和每个人脸对应的 3DMM 参数。该框架以 ResNet-50 为编码器的骨干网络,将人脸细节信息存储到特征图中,然后对存储的信息进行解码,分别得到人脸的姿势、表情、身份和反照率等参数。

在监督学习中,3DMM 的参数往往通过优化拟合或者合成数据得到,这会导致生成的参数不够精确、重建出的人脸缺乏真实感等,为此,研究人员考虑应用无监督学习的方法^[35],其目的是直接从未经标记的人脸图像中提取所需的人脸信息。MoFA^[20]算法是一种基于模型的无监督学习算法,该算法的解码器的输入是一个具有精确语义定义的编码向量,该向量可编码详细的面部姿势、形状、表情、皮肤反射和场景照明,从而使 MoFA 算法能够从输入图像中提取到有意义的参数。在进行三维人脸重建时,该算法的提出使其在大型无监督数据上训练成为可能。

除了无监督学习算法之外,Chen 等提出了一种通过自监督学习回归 3DMM 参数和重构面部细节的算法^[36]。该算法由两阶段组成:第一阶段运用编码器联合多级损失项的组合来训练回归 3DMM 参数,使网络能够保留人脸的外观信息和鉴别信息;第二阶段将输入图像与通过参数得到的纹理信息展开到 UV 空间中来预测人脸的深度信息,进而将该信息添加到第一阶段回归得到的模型中,得到一个精确的三维人脸模型。该算法是以端到端的方式进行训练的,在学习过程中利用输入图像本身作为监督,既保证了重建过程中存在的约束,又降低了粗糙数据对算法性能的影响。此外,还有很多研究人员应用基于自监督学习的三维人脸重建算法来提高人脸重建的精确度和保真度等^[37-39]。

当使用三维可变形模型进行三维人脸重建时,因训练数据的类型、数量以及线性基的存在,限制了模型的表达能力。为了解决上述问题,Tran 等^[40]提出了一种非线性回归 3DMM 系数算法,直接运用算法的编码器生成了 3DMM 的

投影参数 m 、人脸的形状参数 f_s 和纹理参数 f_T ,解码器将生成的 3DMM 系数映射为三维人脸的形状和纹理,该算法的流程如图 5 所示。另外,通过 300W 数据集拟合 3DMM 得到了新的投影参数 m' 、人脸形状 S' 和人脸纹理 T' ,并通过计算其与编码器得到的对应参数间的损失 L_m 、 L_s 和 L_T 、生成模型与真实的三维人脸关键点集 U 之间的损失 L_L 以及模型生成的渲染图像与真实人脸图像间的损失 L_{adv} 来优化算法。实验结果表明,该算法降低了三维人脸重建的复杂性,并提高了三维人脸重建的精度。

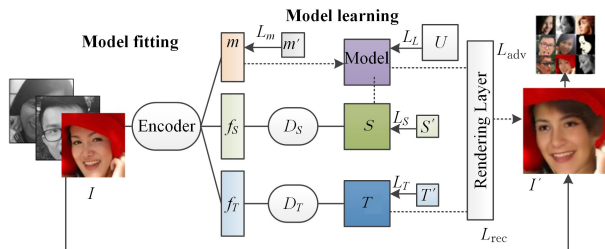


图 5 Nonlinear 3DMM 算法流程^[40]

Fig. 5 Nonlinear 3D face morphable model algorithm flow^[40]

为了能从图像中准确地学习到三维可变形模型的参数信息,需要引入大量的正则化来避免歧义,减小模型误差,这阻碍了算法对高保真模型的学习,为此 Tran 等提出了一种避免运用大量正则化手段来进行三维人脸重建的算法^[41]。该算法的网络结构类似于 Nonlinear 3DMM,运用了编码器网络回归出 3DMM 的参数,然后分别通过形状和反照率解码器重建出对应的结果。不同的是,该算法的形状解码器和反照率解码器的输出都是由两个分支组成的,除了重建出的形状和反照率外,还分别有正则化后形状和反照率结果,进而通过形状和反照率的两两组合来渲染出不同的图像,重建出具有高保真度的三维人脸。该算法不仅解决了强正则化对算法学习过程的阻碍,还保留了正则化的约束优势。实验结果表明,该算法重建出的人脸具有高保真的效果。

在使用自编码器网络结构进行三维人脸重建时,很多算法^[42]都使用单一的编码器对人脸的不同属性进行编码,但人脸不同属性特征的提取是单独的任务,使用相同的编码器会影响各属性的鉴别力。为了改善这个问题,Li 等^[43]提出了 2PEDN 算法。该算法用两个不同的编码器分别进行编码,以得到人脸身份和表情参数,再运用局部解码器回归出三维人脸的表情,运用全局解码器回归出三维人脸的身份。该算法明确区分了人脸的表情和身份特征,重建出的人脸模型也具有较高的精度。

另外,当通过卷积神经网络从图像中回归三维可变形模型的参数时,大多数情况下是在参数化空间上使用全连接层或二维卷积神经网络,这导致其产生了具有许多参数的大型网络,这些网络占用了计算机大量的存储空间并降低了算法的计算效率。为了解决该问题,Zhou 等^[44]提出了一种基于 CNN 的网格卷积网络和自编码器相结合的算法来联合学习人脸的纹理和形状。该算法可以直接在网格域上解码人脸的形状和纹理,具有较高的运行效率。使用网格卷积的原因不仅是因为它可以更加直观地对非线性 3DMM 进行定义,还因为网格卷积是由参数非常少的网络定义的,这使得算法具有相对较低的计算成本。

尽管在进行三维人脸重建时,三维可变形模型具有快速、简便等优点,但是由于模型存在上述缺陷,且深度学习在很多领域都取得了较好的成果,因此有诸多研究人员选择仅使用深度学习的方法来进行三维人脸重建^[45]。

PRNet^[46]使用一个二维的UV位置图来记录三维人脸的3D点在UV空间的位置,然后根据一个将损失函数进行加权的神经网络将人脸从二维的UV平面映射到三维空间中。在该算法中,计算输入图像与三维人脸间的密集对应关系和三维人脸重建这两个步骤是同时进行的,这加快了算法的运行速度。

进一步,Huang等^[47]在PRNet的基础上设计了一种结合多尺度特征融合模块与残差注意力机制的算法。该算法从整体上分为特征提取模块和编码器-解码器模块。对于特征提取模块来说,本文分别应用了密集连接方式与空洞卷积来获得输入图像的多尺度特征,进而在编码器-解码器模块中引入残差注意力机制来提升网络对特征的筛选能力。实验结果证明,不仅算法的运算效率得到了优化,其重建出的人脸模型的精度也得到了提升。

鉴于三维人脸重建可应用场景的广泛性,除了对真实人脸的重建,很多学者也在研究应用于漫画和动画领域的算法。对漫画人脸进行三维重建的挑战主要来自缺乏足够的几何和纹理变化多样的漫画人脸数据,Cai等^[48]首先建立了具有不同风格的2D漫画和3D形状的数据集,然后建立了基于特征点的三维漫画人脸变形空间的参数化模型,在神经网络上回归出了三维漫画人脸。在动画领域,Feng等首次提出学习动画细节并重建出对应的三维人脸的方法^[49],该算法引入了一致性损失,使其能够从特定的细节中解开依赖于表情的皱纹,并通过一致性损失实现了较好的形状重建性能,低维的细节空间使重建具有抗噪声和抗遮挡的能力。

3.4 基于生成对抗网络的三维人脸重建算法

生成对抗网络的思想源自于博弈论中的二人零和博弈,引申到深度学习中则体现在网络中的生成模型(generative model)和判别模型(discriminative model)间的相互博弈学习,其中生成模型的目标是不断学习并生成与真实数据分布相似的样本,判别模型的目标是尽可能地真实数据与生成模型生成的伪数据区分开。生成对抗网络最终的理想结果是生成模型和鉴别模型达到纳什均衡,这时可以认为生成模型学习到了真实数据的近似分布。

三维可变形模型与深度神经网络相结合的算法重建出的三维人脸相比仅使用三维可变形模型重建出的人脸拥有更多的纹理细节^[50-51],但是这类算法仍然存在对人脸纹理学习不足的问题。GANFIT算法^[52]的提出有效地改善了上述情况,该算法将生成对抗网络生成的纹理贴图作为3DMM中的纹理先验信息,利用非线性优化算法来寻找最佳的3DMM系数,从而进行三维人脸重建。该算法的优点是利用生成对抗网络提高纹理重建的真实性,但其在训练时仍需要大量的UV数据的支持。

为了重建出具有高频细节的三维人脸,Lattas等提出了AvatarMe算法^[53]。该算法首先对3DMM与输入图像相匹配形成的UV纹理进行8次上采样,得到可信的高频人脸细节,然后通过去除UV图上的纹理信息来获得具有高频细节

的漫反射率,最后根据漫反射率和3DMM的法线来计算出图像的镜面反照率、漫反射光线等。该算法将上述信息分别映射回UV图像上,然后将该UV图像投影到重建的三维人脸模型,以重建出具有高频细节的三维人脸。

在三维人脸重建算法中,有很多基于优化和基于学习的算法都进行可微渲染来重建目标人脸,但是一方面,由于可微渲染的规则较简单,使其产生的渲染图像与真实图像间的距离较大;另一方面,由于可微渲染的性质,使其优化较为困难,上述两个原因均导致算法无法产生具有真实感的图像。为了改进上述问题,Piao等^[54]提出了一种生成对抗渲染器(Generative Adversarial Render,GAR)。该渲染器以自监督的方式进行训练,将人脸法线映射作为几何条件,将潜在的编码向量作为其他的影响因素进行建模,并通过最小化渲染图像与输入图像间的距离来优化算法,从而提高三维人脸重建的精确度。

基于深度学习的方法在单幅图像的三维人脸重建领域取得了很大的进步,然而,由于缺乏三维数据以及带有标注信息的三维数据,因此限制了算法重建的性能^[53,55]。为了解决三维数据较少的情况,Zhang等^[56]提出了一种基于GANs无监督回归三维参数化人脸模型的方法,该方法通过减小三维人脸投影后的二维人脸与输入的人脸之间的差异、重建后的三维人脸与三维人脸数据关键点间的差异,来优化重建出的人脸模型,使其具有真实感。另外,为了解决缺乏带有标注信息的三维数据的问题,Tu等提出了2DASL算法^[57]。该算法能够仅根据无配合采集图像中稀疏的特征点来预测三维人脸的特征点,并引入了一致性损耗来约束二维和三维空间中人脸的匹配,并运用自我批评的学习方法缩小3DMM重建出的三维人脸与真实人脸的差异,有效地改进了三维人脸学习的过程。该算法能够较好地解决缺乏标注数据的问题,从而达到精确地重建三维人脸的目的。

与2DASL不同,Piao等^[58]提出的算法没有引入三维可变形模型,而是直接基于深度学习、端到端的训练方式回归出目标三维人脸。Piao等提出了一个改进的CycleGAN,将原本的样本生成器的作用改变为将输入图像转化为渲染图像的作用,以减小合成图像与真实图像间的差距,以便进行精确的人脸重建。另外,为尽可能地保留人脸的形状信息,其还提出了一个一致性损失函数来监督生成模型的训练,该损失函数能够为人脸形状提供更多的约束,从而提高三维人脸重建的精确度。

在三维人脸重建时,常会出现人脸重建精度低的情况。生成对抗网络在生成图像数据等方面表现出了卓越的性能,但其在三维视觉领域却表现得一般,其中一个主要的原因是由于生成对抗网络对三维形状通常使用离散的体素来表示,使其易生成粗糙、低质量的三维形状^[30],而MeshGAN算法^[59]的提出有效改善了上述问题。MeshGAN算法提出了一种改进的生成对抗网络结构,其网络结构如图6所示。其运用切比雪夫卷积滤波器来开发MeshGAN,分别进行了4次下采样和4次上采样来提取人脸特征,并通过引入内在的生成模型网络来生成三维人脸网格,该网络的设计使算法能够直接在网格上进行卷积来生成三维人脸。

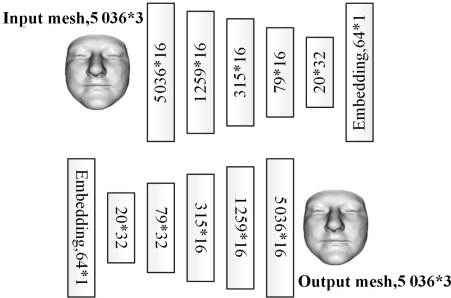


图 6 MeshGAN 的网络结构^[59]

Fig. 6 Network structure of MeshGAN^[59]

该算法的体系结构通过多次上采样和下采样来学习人脸的细节,使其能够生成具有不同身份、不同表情的三维人脸网

格和可以较好地模拟人脸的分布,但该结构的不足之处在于网络训练比较复杂,导致算法的复杂度较高。

4 常见的三维人脸数据集

在基于深度学习的三维人脸重建领域中,人脸数据是至关重要的,数据的数量和质量会对重建性能产生举足轻重的影响。由于三维人脸数据采集不如二维数据简便,因此从整体上看,三维人脸数据库比二维人脸数据库更少,而且在数据精确度、个体的数量、每类个体中样本的数量、公开获取的便利性等方面都有较大的限制。尽管如此,但仍然有一些开放的三维人脸数据库用于进行三维人脸重建方法的训练和评测。表 1 列出了常用的三维人脸重建数据库。

表 1 常见的人脸数据库
Table 1 Common face database

数据集	对象数目	对象特征	其他内容
AFLW2000-3D ^[60]	2000 张图像	每张人脸图像都包含对应的 3DMM 系数与 68 个人脸关键点	常用于测试集
BJUT-3D ^[61]	1200 名	中性表情	对数据进行去噪处理、面部信息分离预处理
BU-3DFE ^[62]	100 名	1 个中性表情和 6 个其他表情,分别为幸福、厌恶、恐惧、愤怒、惊讶和悲伤	每种表情各 4 个强度等级
CASIA 3D FACE ^[63]	123 名	每个个体包含 37 或 38 次扫描	共 4246 张图像;考虑了姿态、表情和光照等因素的单一变化情况和组合变化
Faceware House ^[4]	150 名	中性表情和其他 19 个表情;对象年龄分布为 7~80 岁	在面部标注 74 个人脸特征点
Florence ^[64]	50 名	两个正面模型、一个左侧模型和一个右侧模型	在多个图像分辨率下捕获被摄体面部,使用 5 个级别的缩放
FRGC-v2.0 ^[65]	943 名	中性和微笑表情以及 2~3 种灯光的组合、无控制图像为不同光照下的中性表情和微笑、1 张三维人脸	4 张受控图像、2 张无控制条件图像和 1 张受控条件下拍摄的三维人脸图像
GavabDB ^[66]	61 名	中性表情正面和左右侧面各两张,抬头和低头各一张图片以及微笑、大笑和随意表情姿态的图像各一张	共 549 张图像
3DCaricShop ^[32]	2000 张漫画图像与模型对	由艺术家根据二维漫画数据集制作三维漫画人脸模型数据集	每张人脸都标注了面部关键点,并进行了网格拓扑,为网格提供了姿态标签

AFLW2000-3D^[60]数据集包含 2000 条三维人脸信息,其三维人脸数据由 AFLW 二维人脸数据集中的前 2000 张图像通过三维可变形模型重建得到的。由于 AFLW 中的人脸图像是在无配合的情况下采集得到的,为此其对应重建的 AFLW2000-3D 数据集中的人脸位姿也具有多样性,因此该数据集也适用于无配合三维人脸算法的训练。

BJUT-3D^[61]由北京工业大学多媒体与智能软件技术实验室提供,该研究小组采用了 Cyberware 3030RGB/PS 激光扫描仪扫描来获取三维人脸数据。数据集中所有的数据都进行了对齐处理,保证了三维人脸统一的拓扑结构和人脸形状与纹理一致的表述形式。该数据集丰富的三维人脸数据可以为人脸跟踪、识别、三维人脸重建等计算机视觉任务提供数据支持。

BU-3DFE^[62]是一个包含面部表情的三维人脸数据集,其中每个人脸分别由中性表情和其他 6 个表情组成,该数据集包含 100 个种族不同的对象,共有 606 个面部表情。该数据集首次在三维人脸数据中引入了表情数据,促进了情感计算的研究和对人类面部表情行为的理解。

CASIA 3D FACE^[63]由中国科学院自动化研究所发布,该数据库由非接触式的三维扫描仪美能达 Vivid 910 采集。该数据库中共有 123 个对象的 4246 张三维人脸,每个对象包含 37 或 38 次扫描。在建立数据库时,不仅考虑了

姿态、表情和光照等因素的单一变化,还考虑了多因素的组合变化。

Faceware House^[4]由浙江大学发布,研究人员使用 Kinect 相机收集了 150 名不同种族的个体,对每个个体收集 20 个不同的表情,其中分别为 1 个中性表情和 19 个其他的表情,这些表情均为常见的面部动作。通过对数据集进行计算与组合,文献[4]还提出了一个双线性人脸模型,该模型可用于人脸图像操作、实时性能人脸图像动画等人脸操作。

Florence^[64]数据集由具有高分辨率的三维人脸和多个具有不同分辨率的视频序列组成。每个受测试者都分别扫描了两个头部姿势一致的正面三维人脸模型,一个从鼻子中部覆盖到头后部的左侧三维模型和具有同样属性的右侧三维模型。该数据集中的三维人脸均由精细化的网格表示,该模型的精确度高,适用于高精度面部分析任务。

FRGC-v2.0^[65]的数据由美达能 900/910 传感器收集。该数据集包含了 943 名对象,其中每名对象都有 4 张受控条件下拍摄的图像、2 张无控制条件下拍摄的图像和 1 张受控条件下拍摄的三维人脸图像。该数据集具有多个单独的测试集,便于对算法从一个数据集到另一个数据集的泛化能力进行度量。由于多测试集的存在,该算法无法针对单测试集进行调优,但有助于对算法进行无偏估计。

GavabDB^[66]由胡安卡洛斯国王大学提供。61 名受测试

者都采集了 6 个无表情的数据和 3 个有表情的数据。在无表情数据中,人脸分别展示了两个正面姿态、抬头和低头的姿态(角度均为 90 度),以及左转和右转的姿态(角度均为 30 度)。在有表情的数据中,人脸的表情分别为微笑、大笑和随机表情。该数据集对人脸信息的描述较为详细,有助于高要求的面部分析任务。

3DCaricShop^[32]是在名为 WedCaricature 的二维漫画数据集集中选取的 2 000 张图片、4 个专业艺术家手工制作的高度多样化和高质量的三维漫画模型数据集,其为三维漫画人脸重建提供了具有高保真度的形状先验信息。

5 典型三维人脸重建方法的性能对比

在对三维人脸重建算法的优劣进行评估时,应该选择一种统一的、三维人脸重建领域公认的评价指标来进行判别。当前,三维人脸重建领域常用的判别指标为精度、内存空间

占用量和算法运行的时间。

对精度这一评价指标来说,常见的为归一化平均误差(NME)、累积误差分布(CED)、曲线下面积(AUC)、标准差(STD)、均方根误差(MSE)、峰值信噪比(PSNR)和结构相似度(SSIM)等。在实验评估阶段,一般会选择 NME,STD 和 MSE 这 3 种指标进行综合对比分析;对于内存空间占用量这一评价指标来说,常见的为模型所占内存空间大小(model size)和网络参数量(params);对于算法运行时间这一评价指标来说,通常使用毫秒(ms)、浮点运算数(FLOPs)和定点乘累加次数(MACs)来计算重建一个三维人脸所消耗的时间。其中浮点运算数(FLOPs)和定点乘累加次数(MACs)也可以用来衡量算法的复杂度。

表 2 列出了选择 NME,params 和 FLOPs 这 3 个指标来对三维人脸重建算法进行评价的结果,并列举了各种算法的优势及算法所涉及的数据集。

表 2 深度学习三维人脸重建算法的对比
Table 2 Deep learning 3D face reconstruction algorithm comparison

算法类别	模型/作者名称	优势	测试集	训练集	NME	Params	FLOPS
基于卷积神经网络的三维人脸重建算法	3DDFA ^[13]	大姿态人脸	AFLW, AFLW2000-3D, 300W-LP	300W-LP	5.42	3.27 MB	1.9GB
	MobileFace ^[24]	移动设备上实时进行三维人脸重建	BU4DFE	BU4DFE	—	3.21MB	47.1MB
	3DDFA_V2 ^[25]	轻量、高精度、鲁棒	AFLW, AFLW2000-3D, Florence, Menpo-3D	300W-LP	3.59	3.27 MB	91.5 MB
	VRNet ^[26]	高精度,可处理面部表情和遮挡	3DMM 拟合	3DMM 拟合	—	—	—
基于图卷积网络的三维人脸重建算法	Lin et al. ^[31]	高保真纹理、不需要大规模人脸纹理数据库	CelebA	CelebA	—	—	—
	VC-GCN ^[32]	重建出具有高精度和鲁棒性的三维漫画人脸	3D CaricShop 数据集中的 1600 张图像	3D CaricShop 数据集中的 400 张图像	—	—	—
基于自编码器的三维人脸重建算法	Nonlinear 3DMM ^[40]	不需要 3D 数据集	CelebA	300W-LP	4.7	5.8 MB	—
	MoFA ^[20]	无监督,不需要标记数据	CelebA, Facewarehouse, LFW, 300-V	CelebA, Facewarehouse, LFW, 300-V	—	—	—
	WM3DR ^[34]	重建出单幅图像内的多张人脸、考虑全局相机信息和人脸的相对位置	CelebA, 300VW, 300W LP	CelebA, 300VW, 300W LP	3.62	—	—
	PRNet ^[46]	对姿态、光照和遮挡有较好的鲁棒性	AFLW-LFPA, AFLW2000-3D, Florence	3DMM 生成	3.62	13.4 MB	—
基于生成对抗网络的三维人脸重建算法	Piao et al. ^[54]	减少在对大规模训练数据的需求	AFLW, AFLW-LFPA, Florence	100 000 Rendered Images Based on the BFM	3.88	—	—
	GANFIT ^[52]	高精度、真实感	MICC	LSFM	—	—	—
	AvatarMe ^[53]	高频细节	Pix2pixHD 模型生成的高分辨率图像	Pix2pixHD 模型生成的高分辨率图像	3.53	—	—
	2DASL ^[57]	缓解数据集中缺乏 2D-3D 成对标注的问题	AFLW-LFPA, AFLW2000-3D, Florence	UMDFaces, 300W-LP	3.53	23.6MB	4.7GB

注:“—”表示未开源或无法从文献中查到的相关数据

(1)精度

精度是三维人脸重建算法评价的重要指标,其中 NME 用于评估人脸对齐的精确度,如式(3)所示:

$$NME=\frac{\sum_{i=1}^N\left\|\mathbf{x}_i-\mathbf{x}_i^*\right\|^2}{N\times d}$$

(3)

其中, \mathbf{x}_i 为预测的第 i 个坐标, \mathbf{x}_i^* 为第 i 个坐标的真实值, d 表示两眼瞳孔间的距离或者外眼角间的距离, N 表示关键点的个数。

STD 表示的是三维人脸重建误差的标准差,该指标用于

评估重建误差的分散程度:

$$STD=\sqrt{\frac{\sum_{i=1}^N\left(\mathbf{x}_i-\bar{\mathbf{x}}\right)^2}{N-1}}$$

(4)

其中, $\bar{\mathbf{x}}$ 表示真实三维人脸关键点到重建人脸的平均距离。

MSE 表示的是真实的三维人脸与重建人脸之间未进行归一化的距离:

$$MSE=\frac{\sum_{i=1}^N\left(\mathbf{x}_i-\mathbf{x}_i^*\right)^2}{N}$$

(5)

(2)内存空间占用量

当算法的精确度达到一定水准时,需要考虑其占用内存空间大小的问题,因此内存空间占用大小也是评价算法的重要指标之一。

(3)算法的运行时间

在三维人脸重建领域,算法的运行时间也是评价算法优劣的重要指标之一。该指标能够反映算法的运行速度,表示为算法重建一个三维人脸所消耗的时间。

在三维人脸重建精度方面,通过表 2 中的 NME 数据可以发现:AvatarMe 和 2DASL 两种算法的 NME 均为 3.53,都展现了较好的对齐精度。其中,AvatarMe 算法设计的初衷是重建出具有高频细节的三维人脸,因此进行了多次上采样来提取输入人脸的高频细节;设计 2DASL 算法的目的是改进图像-模型间对应的数据标注短缺的问题,适用于缺乏标注数据的算法。

在内存占用空间方面,MobileFace 算法表现出了良好的效果。相比 PRNet 算法 13.4 MB 的网络参数量,3.21 MB 参数量的 MobileFace 算法在内存占用空间方面具有极大的优势。该算法的初衷是在移动设备上实时地进行三维人脸重建,因此在算法中引入了轻量级的网络框架来减小算法的规模。

在算法的运行时间方面,MobileFace 算法仍占一席之地。其算法的运算量为 47.1 MB,是 3DDFA_V2 算法运算量的 1/2。另外,3DDFA_V2 算法的运算量也较优异,相比 3DDFA 和 2DASL 算法来说,其运算量提升了一个数量级。

结束语 自从深度学习算法进入基于单幅图像的三维人脸重建领域,越来越多的三维人脸重建算法改进和提升了重建的效果。本文主要对当前基于深度学习的三维人脸重建算法进行了分类,然后根据不同的类别对具有代表性的算法进行了评估。由于数据集相对缺乏、算法计算依赖性高、重建鲁棒性不足等问题依然存在,三维人脸重建依然有很大的提升空间,许多方向仍然是开放的。本文认为以下几点可能是未来研究需要重点关注的问题。

(1)充足的三维人脸数据集

由于三维数据的采集需要大量精准快速的设备,而当前经过特征点标注的无约束表情人脸图像和对应的三维人脸数据较少,因此很多研究人员致力于构建出可用的三维人脸数据库。

(2)降低重建算法对标注数据的需求

由于带有标注的三维人脸数据较少,而且很多带标注的数据的精确度不高,影响了算法重建的结果,因此研究人员致力于研究自监督或弱监督算法,使重建算法降低对标注数据的需求。

(3)提高不同场景下人脸重建的鲁棒性

在进行三维人脸重建时,经常会出现同一对象的二维图像由于光照、姿态等的不同而导致生成不同的三维人脸的情况,因此研究人员希望能够做到同一个重建对象在任何姿态、任何表情下的三维人脸重建都可以重建出同样的人脸结果。而目前基于单幅图像的三维人脸重建算法均未达到理想状态。

(4)实时的三维人脸重建

基于单幅图像的实时三维人脸重建的需求日益增长,特别是在移动端对输入人脸的快速重建方面。而目前大多数算法均是在 PC 端来实现的,尽管很多研究人员在设计算法时选择轻量级的网络结构,使算法能够实现快速、简便的效果,但是在移动端实时重建三维人脸是人脸重建推广应用的一个重要方向。

参 考 文 献

[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. ACM Press, 1999:187-194.

[2] 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.

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

[4] 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.

[5] 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. IEEE,2017:48-57.

[6] VLASIC D,BRAND M,PFISTER H,et al. Face transfer with multilinear models[J]. ACM Transactions on Graphics (TOG), 2005,24(3):426-433.

[7] YANG H,ZHU H,WANG Y,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. IEEE, 2020: 601-610.

[8] KIM H,ZOLLHOFER M,TEWARI A,et al. Inversefacenet: Deep Monocular Inverse Face Rendering[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2018:4625-4634.

[9] 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. IEEE,2017:1259-1268.

[10] FAN X,CHENG S,KANG H,et al. Dual neural networks coupling data regression with explicit priors for monocular 3D face reconstruction[J/OL]. IEEE Transactions on Multimedia,2020. https://www.researchgate.net/publication/341364310_Dual_Neural_Networks_Coupling_Data_Regression_With_Explicit_Priors_for_Monocular_3D_Face_Reconstruction.

[11] LUO Y,TU X,XIE M. Learning Robust 3D Face Reconstruction and Discriminative Identity Representation[C]//2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP). IEEE,2019:317-321.

[12] DOU P,SHAH S K,KAKADIARIS I A. End-to-end 3D Face

- Reconstruction with Deep Neural Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2017;5908-5917.
- [13] 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. IEEE,2016;146-155.
- [14] ZHANG G,HAN H,SHAN S,et al.Face Alignment Across Large Pose Via MT-CNN Based 3D Shape Reconstruction [C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE,2018;210-217.
- [15] XING Y,TEWARI R,MENDONCA P. A Self-Supervised Bootstrap Method for Single-Image 3D Face Reconstruction[C]// 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE,2019;1014-1023.
- [16] 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. IEEE,2017;5163-5172.
- [17] CHANG F J,TRAN A T,HASSNER T,et al.ExpNet:Landmark-Free,Deep,3D Facial Expressions[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE,2018;122-129.
- [18] CHANG F,TRAN A T,HASSNER T,et al.FacePoseNet: Making a Case for Landmark-Free Face Alignment[C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW). IEEE,2017;1599-1608.
- [19] WU Y,DONG L F. 3D Face Shape and Texture Reconstruction Based on Weak Supervised Learning [J]. Application of Computer System,2020,29(11):183-189.
- [20] 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. IEEE,2017;1274-1283.
- [21] TEWARI A,BERNARD F,GARRIDO P,et al.Fml:Face Model Learning from Videos[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,2019;10812-10822.
- [22] ZHU W,WU H T,CHEN Z,et al.Red:Reinforced Differentiable Attribute for 3D Face Reconstruction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,2020;4958-4967.
- [23] CHAUDHURI B,VESDAPUNT N,WANG B.Joint Face Detection and Facial Motion Retargeting for Multiple Faces[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,2019;9719-9728.
- [24] CHINAEV N,CHIGORIN A,LAPTEV I.MobileFace:3D Face Reconstruction with Efficient CNN Regression[C]//European Conference on Computer Vision. Cham:Springer,2018;15-30.
- [25] GUO J Z,ZHU X Y,YANG Y,et al.Towards Fast, Accurate and Stable 3D Dense Face Alignment[C]//European Conference on Computer Vision. ECCV,2020;152-168.
- [26] 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. IEEE,2017;1031-1039.
- [27] RICHARDSON E,SELA M,KIMMEL R.3D Face Reconstruction by Learning from Synthetic Data[C]//2016 Fourth International Conference on 3D Vision (3DV). IEEE,2016;460-469.
- [28] MCDONAGH S,KLAUDINY M,BRADLEY D,et al.Synthetic Prior Design for Real-Time Face Tracking[C]//2016 Fourth International Conference on 3D Vision (3DV). IEEE,2016;639-648.
- [29] 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. IEEE,2019;285-295.
- [30] RANJAN A,BOLKART T,SANYAL S,et al.Generating 3D Faces Using Convolutional Mesh Autoencoders[C]//Proceedings of the European Conference on Computer Vision. ECCV,2018;704-720.
- [31] 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. IEEE,2020;5891-5900.
- [32] QIU Y,XU X,QIU L,et al.3DCaricShop:A Dataset and A Baseline Method for Single-View 3D Caricature Face Reconstruction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR,2021;10236-10245.
- [33] ABREVAYA V F,WUHRER S,BOYER E.Multilinear Autoencoder for 3D Face Model Learning[C]//2018 IEEE Winter Conference on Applications of Computer Vision. IEEE,2018;1-9.
- [34] ZHANG J,LIN L,ZHU J,et al.Weakly-Supervised Multi-Face 3D Reconstruction[J]. arXiv;2101.02000,2021.
- [35] 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. IEEE,2018;8377-8386.
- [36] CHEN Y,WU F,WANG Z,et al.Self-supervised learning of detailed 3d face reconstruction[J]. IEEE Transactions on Image Processing,2020,29;8696-8705.
- [37] YANG M,GUO J,YE J,et al.Detailed 3D face reconstruction from single images via self-supervised attribute learning[M]//SIGGRAPH Asia 2020 Posters. 2020;1-2.
- [38] YOON J S,SHIRATORI T,YU S I,et al.Self-Supervised Adaptation of High-Fidelity Face Models for Monocular Performance Tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,2019;4601-4609.
- [39] DEVI P R S,BASKARAN R,SL2E-AFRE:Personalized 3D face reconstruction using autoencoder with simultaneous subspace learning and landmark estimation [J]. Applied Intelligence,2021,51(4):2253-2268.
- [40] TRAN L,LIU X.Nonlinear 3D Face Morphable Model[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2018;7346-7355.
- [41] 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. IEEE, 2019;1126-1135.
- [42] LIU F, ZHU R, ZENG D, et al. Disentangling Features in 3D Face Shapes for Joint Face Reconstruction and Recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018;5216-5225.
- [43] LI X, WENG Z, LIANG J, et al. A Novel Two-Pathway Encoder-Decoder Network for 3D Face Reconstruction [C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020). IEEE, 2020;3682-3686.
- [44] ZHOU Y, DENG J, KOTSIA I, et al. Dense 3D Face Decoding Over 2500fps: Joint Texture & Shape Convolutional Mesh Decoders [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2019;1097-1106.
- [45] BAI Z, CUI Z, LIU X, et al. Riggable 3D Face Reconstruction via In-Network Optimization [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021;6216-6225.
- [46] 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;534-551.
- [47] HUANG Y D, ZHOU D K, YANG X. Joint 3D Face Reconstruction and Dense Alignment Algorithm Based on Multiscale Fusion Features and Residual Attention Mechanism [J]. Computer Application Research, 2021, 23(5): 1-5.
- [48] CAI H, GUO Y, PENG Z, et al. Landmark detection and 3D face reconstruction for caricature using a nonlinear parametric model [J]. Graphical Models, 2021, 115: 101-103.
- [49] 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.
- [50] SLOSSBERG R, SHAMAI G, KIMMEL R. High Quality Facial Surface and Texture Synthesis via Generative Adversarial Networks [C]//Proceedings of the European Conference on Computer Vision Workshops (ECCV). 2018;498-513.
- [51] ABREVAYA V F, BOUKHAYMA A, WUHRER S, et al. A generative 3D Facial Model By Adversarial Training [J]. arXiv: 1902. 03619v1, 2019.
- [52] 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. IEEE, 2019; 1155-1164.
- [53] LATTAS A, MOSCHOLOU S, GECER B, et al. AvatarMe: Realistically Renderable 3D Facial Reconstruction “ In-the-Wild” [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020;760-769.
- [54] PIAO J, SUN K, WANG Q, et al. Inverting Generative Adversarial Renderer for Face Reconstruction [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021;15619-15628.
- [55] TEWARI A, ZOLLHOFER 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. IEEE, 2018; 2549-2559.
- [56] ZHANG X X, LI J L. 3D parametric face model based on Gans unsupervised regression [J]. Information Technology and Network Security, 2020, 39(11): 50-55.
- [57] TU X, ZHAO J, XIE M, et al. 3D Face Reconstruction From a Single Image Assisted by 2D Face Images in The Wild [J]. IEEE Transactions on Multimedia, 2020, 23: 1160-1172.
- [58] PIAO J, QIAN C, LI H. Semi-Supervised Monocular 3D Face Reconstruction with End-to-End Shape-Preserved Domain Transfer [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2019;9398-9407.
- [59] CHENG S, BRONSTEIN M, ZHOU Y, et al. Meshgan: Non-linear 3d morphable models of faces [J]. arXiv:1903. 10384, 2019.
- [60] 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). IEEE, 2011;2144-2151.
- [61] YI B C, SUN Y F, WANG C Z, et al. BJUT-3D 3D face database and processing technology [J]. Computer Research and Development, 2009(6): 1009-1018.
- [62] YIN L, WEI X, SUN Y, et al. A 3D Facial Expression Database for Facial Behavior Research [C]//7th International Conference on Automatic Face and Gesture Recognition (FGR'06). IEEE, 2006;211-216.
- [63] ALMABDY S, ELREFAEI L. Feature extraction and fusion for face recognition systems using pre-trained convolutional neural networks [J]. International Journal of Computing and Digital Systems, 2021, 9: 1-7.
- [64] BAGDANOV A D, BIMBO A D, MASI I. The Florence 2D/3D Hybrid Face Dataset [C]//Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding. ACM, 2011;79-80.
- [65] PHILLIPS P J, FLYNN P J, SCRUGGS T, et al. Overview of the Face Recognition Grand Challenge [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005;947-954.
- [66] MORENO A. GavabDB: A 3D Face Database [C]//Proc. 2nd COST275 Workshop on Biometrics on the Internet. 2004;75-80.



HE Jia-yu, born in 1997, postgraduate. Her main research interests include 3D face reconstruction and face recognition.



HUANG Hong-bo, born in 1976, Ph.D., associate professor. His main research interests include computer vision, deep learning and optimization theory.