数字人面部表情迁移: 从人脸表示到表情迁移综述

包仪华[1,2], 涂子奇[3], 骆乐[1,2], 刘越[1,2], 翁冬冬[1,2]*

[1] (北京理工大学光电学院 北京 100081)
[2] (北京市混合现实与新型显示工程技术研究中心 北京 100081)
[3] (中国电子科学研究院共性技术产品中心 北京 100041)
(crgj@bit.edu.cn)

摘 要: 数字人技术通过数字化手段将真实世界的人类与虚拟世界连接, 具有广阔应用潜力. 在数字人的构建中, 面部构建尤其具有挑战性, 为了实现数字人的逼真面部表情, 通常采用真实演员的面部表情进行驱动, 然后将这些表情迁移到数字人上, 即面部表情迁移. 文中综述了数字人面部表情迁移技术的研究进展, 包括三维人脸表示、面部成分提取及表情迁移方法 3 个关键技术环节. 回顾了从显式三维可变形模型到隐式神经网络表达的技术演进, 分析了表情与身份信息分离、跨维度表情映射等关键挑战. 文中指出当前技术主要面临三维模型表现力有限、特征提取易受环境干扰、跨主体适应性不足等问题. 基于对大量文献的分析评估, 提出应发展具备更强局部表达能力的可变形人脸建模方法, 探索基于卷积隐式神经网络的精确表情分离技术, 以及融合隐式神经网络与非线性人脸表示的表情迁移框架; 这些创新方向有助于简化迁移流程, 增强系统泛化能力, 最终实现更自然、高效的数字人面部表情生成, 推动数字人技术在军事训练、教育教学、心理舒缓与影视娱乐等领域的广泛应用.

关键词: 数字人; 人脸描述; 面部成分提取; 表情迁移; 隐式表达
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2024-00296

Digital Human Facial Expression Transfer: A Review from Face Representation to Expression Transfer

Bao Yihua[1,2], Tu Ziqi[3], Luo Le[1,2], Liu Yue[1,2], and Weng Dongdong[1,2]*

[1] (*School of Optics and Photonics, Beijing Institute of Technology, Beijing* 100081)
[2] (*Beijing Engineering Research Center of Mixed Reality and Advanced Display, Beijing* 100081)
[3] (*Generic Technology Product Center, China Academy of Electronics and Information Technology, Beijing* 100041)

Abstract: Digital human technology achieves the mapping of human entities from physical to virtual space through digital means, demonstrating vast potential and promising prospects across diverse fields. Developing realistic facial expressions for digital humans presents practical applications and significant challenges. To achieve realistic facial expressions, a common strategy involves using real actors as references, transferring their facial expressions to digital humans. This review systematically examines the research progress in digital human facial expression transfer technologies, focusing on three key components: 3D face representation, facial feature extraction, and expression transfer methods. The paper traces technological evolution from explicit 3D morphable models to implicit neural network representations, analyzing critical challenges including expression-identity disentanglement and cross-dimensional expression mapping. Current technical limitations

收稿日期: 2024-06-17; 修回日期: 2025-05-14. **基金项目**: 国家重点研发计划(2022YFF0902303); 2022 年度长沙市重大科技专项(kh2301019). **包仪华**(1988—), 女, 博士研究生, CCF 会员, 主要研究方向为虚拟现实、增强现实、人机交互; **涂子奇**(1991—), 男, 博士, 工程师, 主要研究方向为虚拟人面部表情驱动、智能人机交互; **骆乐**(1989—), 男, 博士, 主要研究方向为虚拟现实、混合现实、数字人、人机交互; **刘越**(1968—), 男, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为混合现实、人机交互; **翁冬冬**(1979—), 男, 博士, 研究员, 博士生导师, CCF 会员, 论文通信作者, 主要研究方向为虚拟现实、增强现实、人机交互.

are identified as restricted expressiveness of 3D face models, susceptibility of feature extraction to environmental interferences, and insufficient cross-subject adaptability. Based on comprehensive literature analysis, the review proposes developing deformable face modeling methods with enhanced local expressiveness, exploring precise expression disentanglement techniques based on convolutional implicit neural networks, and creating expression transfer frameworks that integrate implicit neural networks with nonlinear face representations. These innovative directions will help simplify transfer processes, strengthen system generalization capabilities, and ultimately achieve more natural and efficient digital human facial expression generation, advancing digital human technology applications across military training, education, psychological relief, and entertainment industries.

Key words: digital human; face description; facial component extraction; expression transfer; implicit representation

数字人技术实现了人类从物理空间向虚拟空间的映射, 其核心目标是通过数字化手段将现实世界的人类与虚拟世界进行有效地连接, 该技术在军事训练、教育教学、心理舒缓与影视娱乐等诸多领域具有广泛的应用潜力和发展前景. 与真实的人类相同, 数字人的面部表情也是非语言沟通的重要渠道, 不仅能传递丰富的情感信息, 也在用户交互中起到增强临场感与沉浸体验的关键作用, 是提升交互质量与数字人可信度的核心要素. 为数字人创造逼真的面部表情不仅具有实际应用价值, 同时也面临多项挑战. 目前, 为数字人呈现面部表情的常用策略是通过真实演员进行驱动, 即将演员的面部表情迁移到数字人上. 为了降低实现门槛, 本文关注单摄像头场景下的面部表情迁移, 这种从二维图像到三维数字人的面部动态迁移被称作跨维度面部表情迁移. 该迁移流程分为 2 个方面: (1) 检测并提取真实演员的面部表情; (2) 实现数字人对应面部表情的生成.

目前, 数字人的相关研究中仍然存在以下问题: (1) 三维可变形人脸模型(3D morphable models, 3DMM)的表现力有限. 在表情迁移的过程中, 通常采用 3DMM 描述表情的变化, 通过扫描获取演员在不同表情下的三维面部形状作为基础模型, 并通过基础模型之间的线性插值生成需要的表情形状. 因此, 3DMM 表示数字人面部表情的准确性与丰富度在很大程度上取决于基础模型的数量与质量, 而获取大量高精度基础模型的难度明显制约了该模型的表现能力. 为此, 需要对更灵活的三维人脸表示方法进行探讨, 简化人脸模型的构建流程并减少对基础模型的依赖性. (2) 从人脸图像中提取面部成分容易受到多种因素的干扰. 人脸图像不仅记录丰富的面部成分细节信息, 还伴随着如环境光照等额外信息, 影响面部形态的因素、表情和身份信息是核心. 从人脸图像中提取面部形状时, 需要尽量剔除年龄、视角、环境光照等因素的影响, 但在此过程中, 表情与身份信息的提取容易相互干扰, 对于准确地执行面部表情迁移不利. 为此, 必须探索从人脸图像中单独提取表情和身份信息的方法, 并在此过程中施加约束, 提高表情迁移的精确性. (3) 面对不同的用户以及三维人脸, 现有的表情迁移技术的适应性有限. 为了确保面部表情迁移的准确性, 需要对用户的表情进行标定校准, 该步骤不仅需要录制大量演员的表情序列, 也要建立用户表情与数字人面部形状之间的映射关系. 如果演员或数字人有所变化, 则需要重新标定校准, 提升了表情迁移方法的复杂度, 从而限制了该方法的广度. 因此, 需要探索基于隐式表达的表情迁移技术和相应的非线性人脸描述方法, 提高表情迁移技术的通用性.

上述问题在三维建模、图像理解和迁移机制等层面对系统性能形成制约, 影响数字人在多场景下的可用性与拓展性. 为此, 本文系统梳理国内外相关研究进展, 针对三维人脸表示、面部成分获取以及面部表情迁移等主题进行详细的文献综述. 本文研究为表情迁移方法的持续优化与创新提供了新的方向, 有助于实现更准确、更便捷的跨维度面部表情迁移, 为提高数字人的表现能力及其多元化的应用开辟了道路.

1 本文概述

数字人面部表情迁移技术涵盖从人脸表示到

表情生成的完整技术链条. 本文系统地综述三维人脸表示、面部成分提取和表情迁移三大关键技术. 其中, 三维人脸表示为数字人表情迁移提供形状建模和表情描述的几何基础; 面部成分提取从图像中分离出表情和身份特征, 是表情映射的核心数据来源; 表情迁移技术则通过整合表示模型和提取特征, 完成表情从源主体到目标主体的动态映射. 为了更加清晰地阐述本文的研究内容, 给出如图 1 所示数字人面部表情迁移技术整体结构及逻辑关系. 其中, 三维人脸表示为表情迁移技术提供形状建模的支撑, 面部成分提取通过特征分离为表情迁移提供高质量输入, 而表情迁移方法涵盖从二维图像驱动到基于隐式表达的迁移等多种技术方向. 这三者相辅相成, 共同推动数字人表情迁移技术的发展.

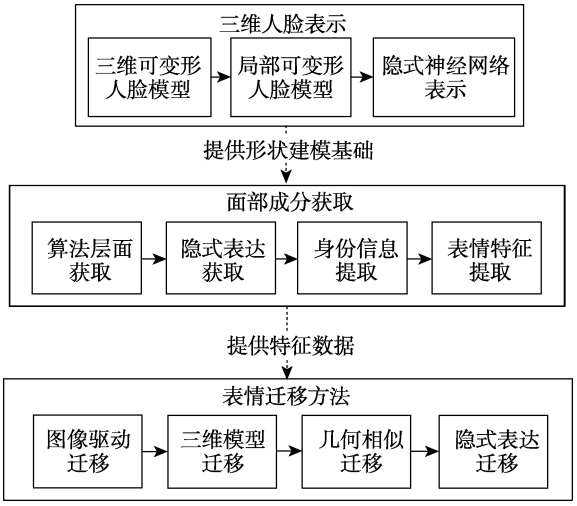


图 1　数字人面部表情迁移技术框架

2　三维人脸形状表示方法

三维人脸表示是数字人面部表情迁移的基础技术. 高质量的人脸表示模型能够精确地描述面部几何形状和动态表情变化, 为后续表情提取与迁移提供稳定的形状支持. 在表情迁移中, 三维人脸表示不仅决定了目标表情的自然性, 还直接影响表情的精确映射效果, 因此, 深入探讨三维人脸表示技术对理解表情迁移至关重要.

2.1　3DMM

在表情迁移过程中, 通常采用由顶点、边线与面组成的网格模型描述三维人脸. 为减少三维人脸建模中的人工干预, Blanz 等[1]提出 3DMM 的构建方法, 成为三维人脸研究的重要基础. 该模型通过线性插值分别控制人脸形状与纹理, 可生成多样化的三维人脸. 为了简化创建 3DMM 的构建流程, Paysan 等[2]提出公开的模型 BFM(Basel face model), 得益于三维扫描设备的进步和标定算法的愈加成熟, 该模型具有更准确的面部形状和更精细的纹理贴图; 但是, BFM 仅包含不同的人脸形状, 没有考虑面部表情的变化. Cao 等[3]借助深度相机创建了三维人脸数据集 FaceWarehouse, 该数据集包含 150 人的三维人脸数据, 每人拥有 47 个不同的面部表情, 其中的表情构成参考面部表情编码系统[4]的标准, 涵盖了大部分常见的面部表情. 基于该数据集构建的可变形人脸模型同时具备身份与表情 2 个线性维度, 可生成更丰富的面部形状; 然而, 该模型缺乏贴图信息, 成为其主要局限. 为弥补上述不足, 许多研究将 FaceWarehouse 的表情维度与 BFM 结合使用, 形成兼具形状与表情控制的人脸模型. 为降低对设备的依赖并提升三维建模质量, Li 等[5]提出 FLAME(faces learned with an articulated model and expressions)模型. 该模型基于大量三维人脸扫描数据, 学习形状与表情的变化规律, 并融合线性变形空间与骨骼运动机制, 能够表达更加丰富的面部细节及人脸姿态变化, 如图 2 所示.

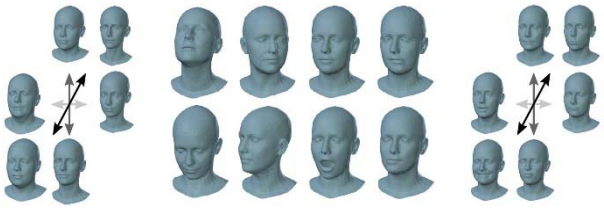


图 2　FLAME 人脸模型[5]

近年来, FLAME 模型受到越来越多的关注, 并被用于许多三维人脸相关的研究中. 不满足于已有人脸模型的质量, Yang 等[6]通过 68 台相机组成的阵列获取大规模的人脸数据集 FaceScape, 其中的人脸数据包含面部皱纹以及毛孔级别的面部结构. 如图 3 所示, 由 FaceScape 数据集构建的 3DMM 进一步提升了面部形状的精细程度.

2.2　局部可变形人脸模型

通过扩充 3DMM 的表情与身份维度能够提升模型的变形能力, 但是扩充维度带来的提升逐渐减少, 而构建模型的工作量则急剧增加. 局部可变形人脸模型为进一步提升三维人脸的表现力提供了新的思路, 该模型指面部不同区域可以独立变形的 3DMM. Zollhöfer 等[7]指出, 局部可变形模型能够提供更多的自由度, 因此具有更加灵活且能

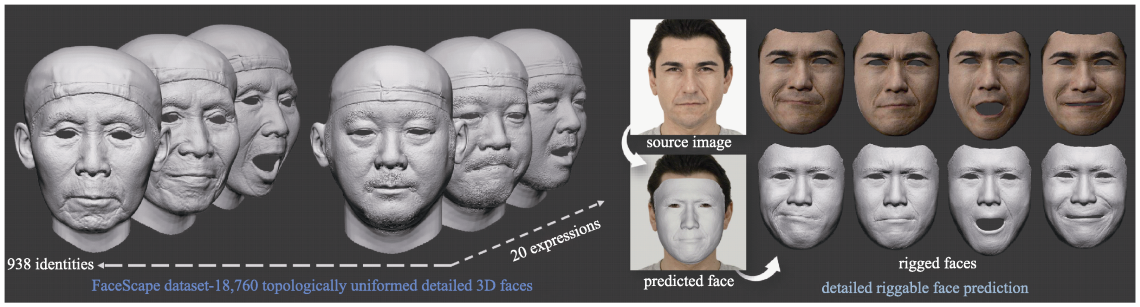
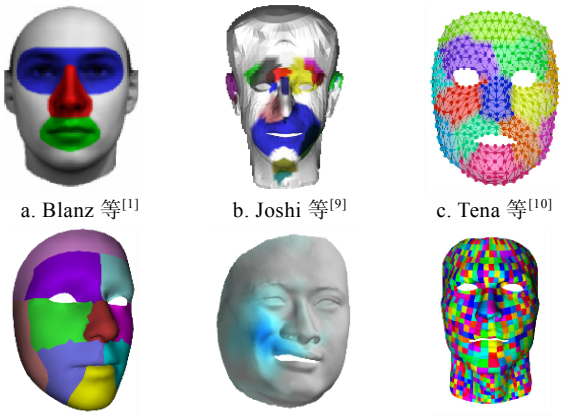


图 3 FaceScape 人脸模型[6]

发掘数据中隐藏信息的优势；Black 等[8]对图像中的局部参数化人脸模型进行探索，提出的局部参数化模型能够有效地识别与恢复人脸的非刚性运动. 在三维人脸方面，Blanz 等[1]指出，将人脸划分为独立变形的区域能够增加模型的自由度，并在构建可变形人脸模型时将面部划分为如图 4a 所示眼睛、鼻子、嘴，以及其他区域，但是并没有说明划分面部区域的依据. Joshi 等[9]认为，模型的分割应该反映人脸的特质，便于三维人脸编辑以及提供不同级别的细节，并于 2006 年提出一种基于面部运动的自动分割方法，能够根据面部变化的丰富程度实现模型分割，其中一种分割方式如图 4b 所示. Tena 等[10]根据动作捕捉数据中的人脸运动构建局部可变形模型，对大量的面部运动数据进行分析，并根据模型顶点的空间相关性进行聚类，将人脸模型划分为 13 个独立变形的区域，如图 4c 所示；并通过实验证明，局部可变形模型对于生成人脸形状具有更好的泛化能力. Cao 等[11]采用类似 Tena 等[10]的方法，从真实的面部数据集中构建局部可变形人脸模型，如图 4d 所示；与 Tena 等[10]的方法不同的是，Cao 等[11]对于构建完成的可变形人脸模型进行划分，而不是在已分割的模型上构建可变形模型，保存了基础表情的整体性. Neumann 等[12]从动画序列中获取局部变形成分，构建的模型便于直观地控制形状变化并组合生成新的形状，该模型嘴角的局部变形如图 4e 所示；该方法不仅容易实现并且很通用，能够用于面部表情、身体动作、服装变化以及肌肉的变形等. Wu 等[13]将人脸模型分割为 1 000 个可独立变形的面片，如图 4f 所示，模型通过基于解剖结构的骨骼划分对面片姿态与形状施加约束. 该分割方式结合面部解剖与区域变形特性，可在确保局部运动表达的同时保持结构连续性，仅用 10 个极限表情即可生成准确多样的形变. 面片数量过多反而会导致区域过小，约束困难.

此外，局部可变形人脸模型也不仅局限于面部区域分割. Brunton 等[14]通过小波变换将面部形状分解为许多独立的局部线性模型，能够从干扰与遮挡的数据中恢复良好的面部细节，如图 5 所示.



a. Blanz 等[1] b. Joshi 等[9] c. Tena 等[10]

d. Cao 等[11] e. Neumann 等[12] f. Wu 等[13]

图 4 6 种面部区域划分与局部可变形人脸模型

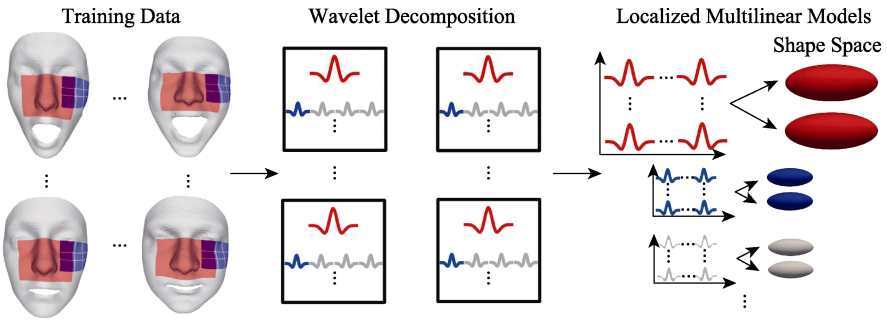


图 5 基于小波变换的局部可变形人脸模型[14]

2.3　隐式神经网络中的三维人脸表示方法

近年来, 隐式神经网络在三维数据处理领域展现了强大的潜力, 通过参数化函数对三维形状进行连续建模, 不仅可避免显式表示中对分辨率的依赖, 还能高效地处理复杂几何形态. 与传统的显式三维表示(如网格、体素或点云)相比, 隐式方法通过神经网络函数直接描述三维形状, 克服了传统方法中对离散数据结构的依赖问题.

基于参数化模型的三维人脸表示是隐式方法的重要研究方向之一. 早期的线性三维人脸模型通过对三维形状和纹理进行线性参数化, 能够实现对人脸形状和表情的简单控制[1]. Tran 等[15]提出非线性三维人脸变形模型, 通过非线性低维空间中的参数化建模, 显著地提升了三维人脸的形变灵活性; Jiang 等[16]提出结合解耦表征学习的三维人脸模型, 通过分离身份与表情参数, 进一步提升了人脸生成的灵活性与精确性.

在深度隐式建模方面, Jackson 等[17]利用隐式网络对三维空间中的人脸形状进行参数化建模, 并通过卷积神经网络从单幅图像中生成高质量的三维人脸形状; Yenamandra 等[18]提出 i3DMM, 通过隐式神经网络直接建模三维人脸几何结构, 在紧凑性与表达能力之间取得了良好平衡, 并且支持动态表情和姿态的高效表示; Giebenhain 等[19]进一步优化隐式参数化头部模型的学习能力, 增强了表情和头部形态的细粒度建模能力; Raj 等[20]将神经辐射场(neural radiance fields, NeRF)引入三维人脸建模, 利用隐式方法对几何结构与表面细节进行建模, 显著提升了模型的表达能力. 该方法进一步拓展了隐式表达在三维建模中的适用范围, 能够从多个视角生成逼真的三维人脸图像. Wang 等[21]和 Hong 等[22]分别通过采样策略优化和集成二维渲染技术, 显著地提升了 NeRF 在三维人脸建模中的效率. 然而, 传统的 NeRF 主要用于静态三维形状, 对于动态表情的描述能力有限. Gafni 等[23]和 Athar 等[24]通过参数化建模, 提出一种结合动态表情和姿态变化的隐式方法, 显著地提升了动态三维人脸的建模能力, 如图 6 所示.

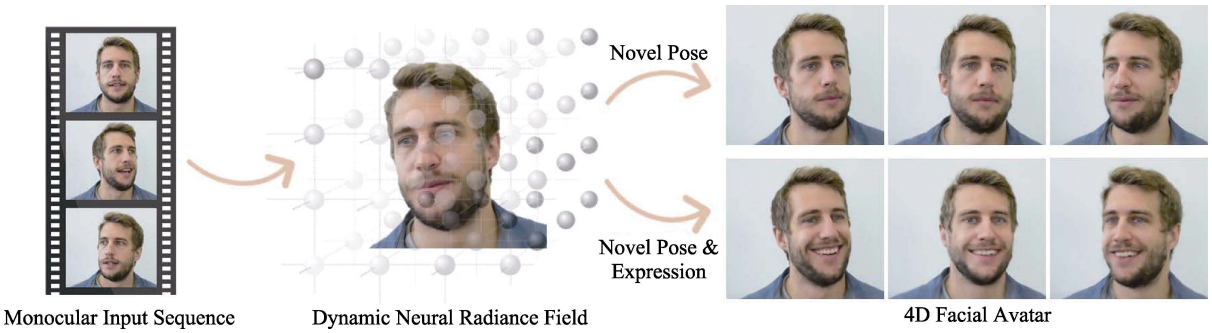


图 6　动态 NeRF[23]

隐式神经网络的优势还体现在点云数据的处理上. 点云数据作为一种显式表示方法, 通过一组三维坐标点直接描述物体的几何结构. Qi 等[25]提出 PointNet, 利用隐式神经网络对点云数据进行特征提取, 并应用于物体分类、分割与场景识别等任务. 在 PointNet 的基础上, Bhople 等[26]扩展对三维人脸点云的识别能力; Li 等[27]进一步设计了点云超分辨率网络, 用于从低分辨率点云恢复高分辨率的三维人脸, 如图 7 所示.

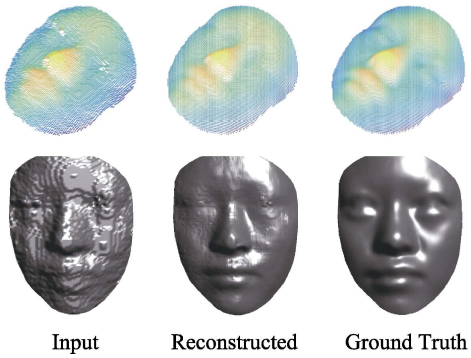


图 7　点云人脸模型的超分辨率采样[27]

此外, 设计特殊的卷积操作与网络结构是隐式网络处理三维网格模型的重要手段. Ranjan 等[28]提出基于隐式方法的网格采样网络, 能够从三维人脸模型中提取多尺度的非线性变化信息, 并通过自编码器生成新的三维人脸形状; Hanocka 等[29]和 Feng 等[30]则开发了直接从网格模型中提取特征的隐式神经网络. 这些方法通过将传统网格数据的几何信息嵌入隐式网络中, 增强了三维表示的灵活性和表达能力, 如图 8 所示.

隐式神经网络的另一个重要应用是三维人脸

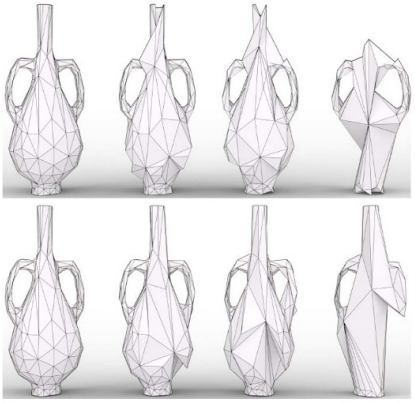


图 8 网格模型的卷积示意图[29]

数据与二维图像的结合. Feng 等[31]提出二维纹理坐标(uniform texture coordinate system, UV)位置映射的三维人脸模型描述方法, 将三维人脸顶点映射到UV空间对应的位置, 并通过轻量化卷积神经网络从单幅图像中生成高精度三维人脸模型; Moschoglou 等[32]采用与 Feng 等[31]类似的三维人脸表示方法, 将三维人脸的形状映射到 UV 空间中, 并结合生成式对抗网络(generative adversarial networks, GAN)[33]实现了保留面部细节的三维人脸生成, 如图 9 所示; Bailey 等[34]和 Li 等[35]在描述面部形状的变化过程中也都采用了类似的三维人脸表示方法.

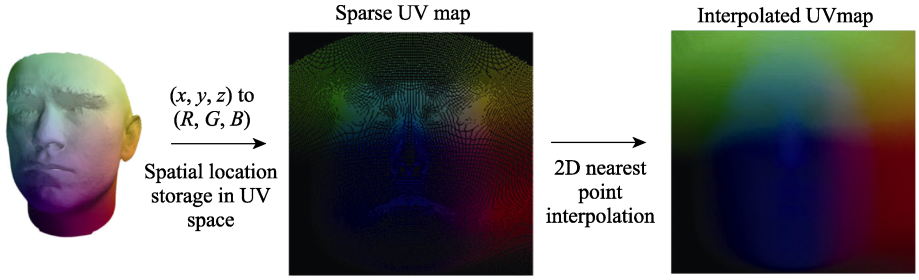


图 9 在 UV 空间中映射三维人脸形状[32]

综上所述, 为了准确地呈现三维人脸形状, 已有许多构建 3DMM 的工作. 随着获取三维人脸数据技术的日趋成熟, 通过增加可变形人脸模型的维度能够提升三维人脸的变形能力, 而构建可局部变形的人脸模型也为提升三维人脸的灵活度提供了新的思路. 然而, 仅靠三维人脸表示并不足以实现面部表情的动态迁移. 在跨主体表情迁移过程中, 还需要从二维图像中提取与表情相关的特征信息, 包括表情分离与身份去耦等关键任务. 下节重点讨论面部成分提取技术及其在表情迁移中的核心作用.

3 二维图像的面部成分获取方法

面部成分提取技术是连接二维图像与三维表情映射的关键环节. 通过提取表情、身份和其他面部特征, 成分提取技术为表情迁移提供了高质量的数据输入; 特别是在动态表情迁移中, 准确的表情分离与身份去耦是实现跨主体映射的重要前提.

3.1 算法层面获取面部成分

为了从人脸图像中获取描述用户身份的信息, Edwards 等[36]通过人脸统计模型估算图像中的面部形状参数作为用户的身份信息; 随后, 他们基于主动外观模型将人脸图像中的身份信息与其他信息解耦[37], 实现了对用户身份的稳定估计, 避免了角度、照明以及表情变化对于结果的影响, 如图 10 所示.

a. 身份信息变化

b. 其他信息变化

图 10 身份信息与其他信息解耦[37]

Zhou 等[38]提出一种面向图像中用户身份建模的通用框架, 支持从单幅或多幅图像中提取身份信息, 并根据任务需求输出离散或连续的身份表示. 鉴于人类外貌随年龄变化, 衰老过程的个体差异对身份建模构成挑战. Zhou 等[39]提出一个基于年龄外貌的身份推理模型, 通过同时对身份和年龄变量进行建模确定身份子空间, 从而提高对衰老面孔的识别准确率.

由于面部表情与情绪高度相关, 许多研究以情绪识别为切入点, 将愤怒、厌恶、恐惧、快乐、悲伤和惊喜 6 种基本情绪的强度作为描述面部表

情的依据. Gritti 等[40]提出一种基于方向梯度直方图的局部特征方法，用于提取情绪分量并实现人脸表情识别; Happy 等[41]提出一种基于面部外观特征的表情识别框架，通过眼睛和嘴附近的特征点，获取在表情变化过程中变化明显的面部区域，如图 11 所示，并根据这些区域的特征获得描述面部表情的情绪类别. 然而，6 种基本情绪及其组合仍难以全面刻画面部表情，且二者之间并不具备一一对应的关系. 面部表情编码系统通过运动单元(action units, AU)定义面部肌肉的运动，更适合作为面部表情的描述. Pantic 等[42]提出一种面部表情自动识别系统，通过静态人脸图像中的轮廓特征推理面部 AU 数据，并通过 AU 数据描述面部表情; Jiang 等[43]在局部相位量化(local phase quantization, LPQ)方法的基础上扩展动态纹理检测，实现对面部 AU 的实时检测，并提升了获取面部表情信息的准确度.

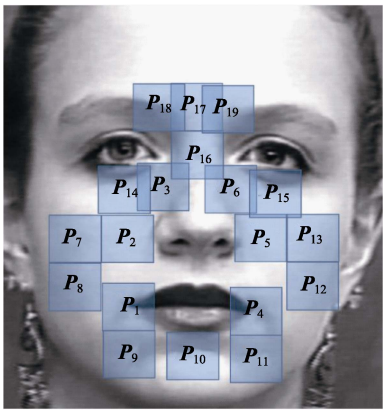


图 11　表情变化明显的面部区域[41]

3.2　隐式表达方法中的面部成分获取

隐式神经网络表达方法可通过大规模数据学习潜在规律，其中卷积神经网络在面部特征提取任务中展现出显著优势. Nakada 等[44]基于卷积神经网络提出一种主动人脸识别系统，通过卷积神经网络提取面部特征，并采用最近邻算法识别用户身份信息. 为了提升人脸识别的效果，Kang 等[45]提出对称关系网络，并通过该网络从特征图中寻找不同身份之间的对称关系，获得图 12 所示描述人脸身份信息的有效特征.

Li 等[46]提出一种用于跨年龄人脸识别的隐式卷积网络，结合身份识别网络与共享特征层的年龄识别网络，能够有效地分离身份特征和年龄特征，具备了在年龄变化的情况下准确识别身份信息的能力; 在此基础上，Huang 等[47]提出一种具有

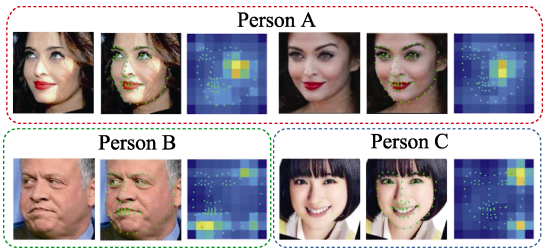


图 12　描述身份信息的特征可视化[45]

并行网络结构的年龄对抗卷积神经网络，通过特征融合的金字塔结构辅助网络训练，获得了对年龄变化保持不变的身份信息.

在从人脸图像中提取表情信息时，身份因素往往构成干扰. Liu 等[48]认为，分离身份相关因素有助于提升表情识别效果，并提出一种身份分离的表情识别框架; Wang 等[49]在表情特征提取中引入对抗训练，抑制身份与面部朝向对特征的干扰. 为进一步减弱身份属性的影响，Meng 等[50]提出一种对身份信息敏感的对比损失函数，实现了不受身份影响的面部表情识别; Cai 等[51]采用条件 GAN 进行面部表情识别，在不改变表情的情况下将输入的人脸图像转换为特定身份，如图 13 所示，减少了身份信息对于表情识别的干扰. 然而，Zhang 等[52]认为，人脸图像中的身份信息能够促进表情识别，并提出双线性聚合模块，能够合理地将身份信息与表情信息融合，提升表情识别的准确率.

图 13　人脸图像的身份信息转换[51]

综上所述，从人脸图像中提取身份信息需综合考虑姿态角度、环境光照与年龄变化等多种干扰因素. 表情识别则多以基本情绪的分类为基础，面部表情编码系统通过动作单元精确描述肌肉运动过程，为表情建模提供了更具可操作性的参数表达. 卷积神经网络不仅有效地提升了面部结构与纹理特征的提取效率，也增强了身份与表情信息的表征能力与鲁棒性.

4　面部表情迁移方法

面部表情迁移是融合三维人脸建模与面部成

分提取成果的关键技术, 旨在将源主体(如图像或三维模型)中的表情状态映射至目标主体, 实现自然逼真的表情合成效果, 是构建高真实感数字人的核心能力之一. 作为本文的重点内容, 本节系统梳理了当前表情迁移领域的主流技术路径, 涵盖图像驱动的表情映射、三维模型之间的表情传递以及基于隐式表达的生成方法. 根据实现方式与应用需求的差异, 现有表情迁移方法可大致划分为 4 类: (1) 基于图像的表情迁移. 主要针对二维图像, 通过提取面部特征实现表情映射, 该类方法利用表情特征点或光流等技术, 能较为直观地完成二维图像的表情变化[53-54]; (2) 基于三维可变形模型的表情迁移. 通过三维人脸模型的参数化表示进行表情驱动, 该类方法能够跨越视角限制, 为动态和多视角场景提供高效的表情生成方案[55]; (3) 基于几何相似的表情迁移. 利用网格形状或特征点的几何匹配关系完成表情的跨主体传递, 该类方法在三维模型间实现精准的表情映射方面表现出色[56]; (4) 基于隐式表达的表情迁移. 采用隐式神经网络学习表情映射函数或表情向量, 进一步提升表情迁移的灵活性和适用性[57]. 上述分类不仅体现了表情迁移技术的发展脉络, 也为不同应用场景提供了技术支撑.

4.1 针对图像的表情迁移

图像层面的表情迁移虽然无法直接应用于为虚拟数字人生成面部表情, 但是在描述图像中的面部表情以及建立表情映射关系等方面, 也能为三维情形的表情迁移提供参考.

为了在远程会议或者多人虚拟世界中高效率地传递面部表情, Buck 等[58]提出一种通过用户的面部表情驱动手绘卡通人脸的方法, 采用的卡通人脸由一个面部背景和数个面部组件构成, 面部组件如图 14 所示.

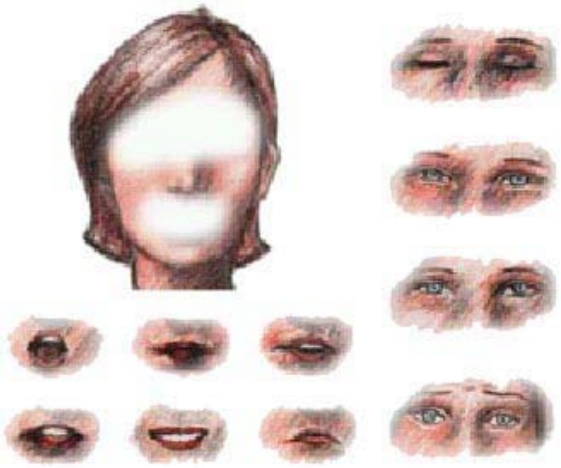


图 14 卡通人脸的面部组件[58]

该组件包含 6 种嘴部形状以及 4 种眼睛状态; 该方法根据用户图像中眼睛和嘴的位置与形状定义一组面部特征, 通过从图像中检测该面部特征估算面部组件的形变系数, 获得与用户图像对应的面部表情. 卡通风格人脸简化了面部特征, 支持低硬件配置下的实时表情驱动, 但该方法表达能力有限, 且流程复杂、通用性差. 为了呈现自然的面部表情和面部细节, 光照与阴影是不可或缺的因素. Liu 等[59]从面部表情的光照变化中获取表情比率图像(expression ratio image, ERI), 能够反映表情变化中的皱纹等面部细节, 如图 15 所示. 设 ERI 为 R 可以通过中性表情图像 I_{n} 和表情图像 I_{e} 计算得到, 即

$$R(x,y)=\frac{I_{\mathrm{e}}(x,y)}{I_{\mathrm{n}}(x,y)}.$$

其中, (x,y) 表示像素坐标. 利用 ERI 进行表情迁移时, 目标图像可以表示为

$$I_{\mathrm{t}}(x,y)=I'_{\mathrm{n}}(x,y)\cdot R(x,y).$$

其中, I'_{n} 表示目标对象的中性表情图像. 通过这种方式, ERI 能够有效地捕捉和迁移面部细节的变化. 通过在表情迁移的过程中同步映射 ERI, 该方法能够获得更加生动的面部表情. ERI 广泛地应用于针对图像的表情迁移方法中, 但该方法对不同光照环境的适应性比较弱.

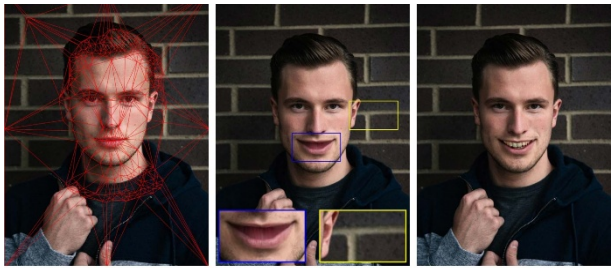
a. 中立人脸图像 b. 表情人脸图像 c. 表情比率图像

图 15 表情迁移及比率图像[59]

在面部表情迁移的过程中, 表情的映射关系对于生成相似的面部表情具有决定性的作用. Li 等[53,60]结合针对图像的表情相似性估计与表情映射方法实现图像层面的表情迁移, 通过光流衡量面部表情的相似性, 并在时间和空间上对表情映射进行约束, 生成相似且自然的面部表情. Li 等[60]通过最小化能量函数

$$E_{\mathrm{opt}}=\arg\min_{E}\sum_{t}\left\|E_t-\hat{E}_t\right\|_2^2+\lambda\sum_{t}\left\|\nabla E_t\right\|_2^2$$

保证时空一致性. 其中, λ 是平衡空间和时间一致性的权重系数. Averbuch-Elor 等[54]提出一种用于单幅人脸图像的表情驱动方法, 如图 16 所示, 该方法首先通过自动检测的面部特征点对齐人脸区域, 再根据置信度向量场对图像进行变形, 并通过填补嘴部与添加局部细节完成表情迁移. 该方法仅依赖单幅图像即可实现驱动, 显著地降低数据需求, 拓展了表情迁移的应用范围; 但无法调整人脸朝向, 且当原图表情非中性时易产生不自然结果.



a. 输入的图像　　b. 图像变形　　c. 添加面部细节

图 16　单幅图像的表情驱动[54]

总体而言, 图像层面的表情迁移方法具备实现简单、无需三维建模和适应低资源条件的优势, 在视频表情增强、卡通化合成及前处理阶段具有广泛应用潜力. 但该类方法通常难以建模面部姿态变化及三维结构细节, 生成结果在表情连续性和空间一致性等方面仍存在一定局限, 更多作为三维迁移方法的重要补充路径存在.

4.2　基于 3DMM 的面部表情迁移

相较于图像层面的表情迁移易受固定视角限制, 基于 3DMM 的表情迁移方法可有效地突破视角约束, 实现更灵活的表情生成. 3DMM 通过低维参数建模复杂人脸形状, 在表情提取与映射中具有效率高和可控性强的优势. Thies 等[55]提出一种针对视频的实时面部表情迁移方法, 通过深度相机获取用户的面部形状, 并由此估算可变形人脸模型的参数, 将替换了表情系数的模型渲染并拼接入视频中实现表情迁移, 该方法如图 17 所示.

对于直接替换表情系数的表情迁移方法, 从图像中提取表情参数的准确性直接影响最终合成表情的质量[61]. Cao 等[62-63]基于 FaceWarehouse 模型公开人脸图像数据集, 训练了一个无需标定的动态表情回归器(displaced dynamic expression, DDE), 可实时重建视频帧中的三维人脸与表情, 并通过表情系数替换为卡通角色生成面部动画. 为增强单目跟踪的稳定性, Cao 等[11]进一步提出局部区域自适

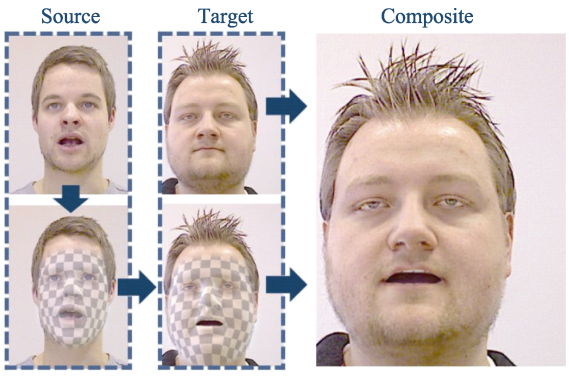


图 17　替换人脸模型的表情系数实现表情迁移[55]

应加权的刚性稳定方法, 在可变形人脸模型上提高姿态估计与表情重建精度, 但其实现前提是三维人脸模型具有一致的基础表情构成, 限制了其在异构模型间的迁移适用性. 针对该问题, Li 等[64]提出基于少量样例生成目标模型基础表情的方法, 使表情系数替换可跨模型进行; Thies 等[65]提出 Face2Face 系统, 通过引入形状约束, 建立不同模型间的表情系数映射关系. 该方法仅依赖单目彩色视频即可实现高质量驱动, 但也带来了更大的计算开销.

在基于标志点跟踪的表情迁移方法中, 通过在演员面部贴附特征标记, 可为表情参数估计提供更直接的几何对应关系[66-68]. Seol 等[69]将演员面部特征点在一定时间内的运动轨迹与三维模型上对应点进行匹配, 从而估算表情系数并还原面部表情. 在此基础上, Ribera 等[70]改进表情迁移的过程, 将三维人脸的运动空间自适应地转换为演员的面部运动范围, 同时通过增加形状约束提升了表情迁移的效果, 面部运动空间的映射如图 18 所示; 然而, 该方法对特征标记精度要求高, 标定流程复杂, 限制了其通用性.

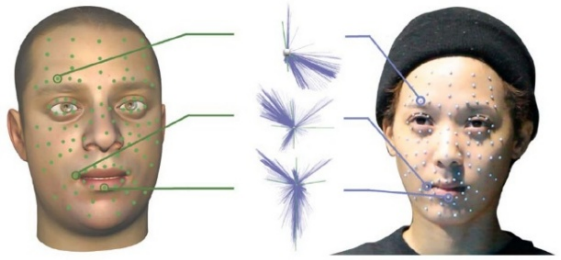


图 18　面部特征点构成运动空间[70]

4.3　基于几何相似的表情迁移

在基于模型驱动的表情迁移方法中, 除了通过 3DMM 的参数映射直接生成目标表情外, 也有方法从几何结构出发, 通过网格形变实现表情传

递. Sumner 等[56]提出一种在三维网格模型之间传递形变的方法, 通过手工标注的对应点估算 2 个网格模型的稠密对应关系, 并将模型中三角面的仿射变换转移给另外模型上对应的三角面, 经过平滑约束后即可得到具有相同形变的网格模型, 如图 19 所示.

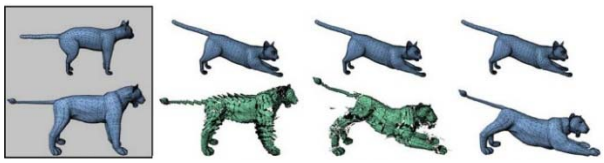


图 19 三维网格的形变传递[56]

对于源模型中的三角面片 $\boldsymbol{S}$ 及目标模型中的对应三角面片 $\boldsymbol{T}$, 仿射变换可以表示为

$$\boldsymbol{T}=\boldsymbol{RS}+\boldsymbol{t}.$$

其中, $\boldsymbol{R}$ 表示旋转矩阵; $\boldsymbol{t}$ 表示平移向量. 对于三角面的顶点, 仿射变换可以表示为

$$\begin{bmatrix}V_x'\\V_y'\\V_z'\end{bmatrix}=\begin{bmatrix}r_{11}&r_{12}&r_{13}\\r_{21}&r_{22}&r_{23}\\r_{31}&r_{32}&r_{33}\end{bmatrix}\begin{bmatrix}V_x\\V_y\\V_z\end{bmatrix}+\begin{bmatrix}t_x\\t_y\\t_z\end{bmatrix}.$$

该类几何变换能够在表情迁移过程中保持局部面部特征的一致性, 具有较好的形变保真性. 当应用于人脸模型时, 同样可实现有效的表情迁移; 但若源目标模型之间存在显著脸型差异, 则可能导致表情失真. 此外, 该方法也常被作为几何形状约束, 嵌入至其他表情迁移算法中, 以增强形变的结构一致性约束[64-65,70]. Suwajanakorn 等[71]和 Rotger 等[72]提出通过将三维网格模型上对应点在表情变化过程中的空间偏移传递至另一模型, 实现跨个体的表情迁移. 此类方法基于三维人脸几何结构的相似性, 不依赖于线性表情空间建模, 且不要求参与模型具备一致的拓扑结构, 因此其在不同拓扑的人脸模型之间具有较好的适应性. 然而, 该类方法仍依赖于准确的面部特征点对应关系, 并且由于其相似性建立在三维几何结构基础上, 因此难以直接应用于基于演员图像的表情驱动任务.

4.4 基于隐式表达的面部表情迁移

随着技术不断迭代, 隐式神经表达已广泛应用于面部表情迁移等任务中, 提供了更高效的图像层建模方式. Wiles 等[57]提出一个隐式神经网络 X2Face, 通过图像或者音频等方式驱动目标视频中人脸的表情和姿态, 基于 pix2pix 网络[73]获取像素流实现隐式映射. Bansal 等[74]进一步结合时空一致性约束, 实现表情变化下的面部外观保持. 尽管基于图像映射的方法训练数据要求较低, 但训练过程复杂. 为提升表情表达效果, Wu 等[75]提出一种多对一的二维面部表情迁移方法, 将面部特征点连接起来构成边界轮廓线用于描述面部表情, 并通过 GAN 将轮廓线转换为具有对应表情的人脸图像. 面部特征点包含了足够的表情、脸型等信息, 有助于隐式神经网络对表情进行提炼, 在表情迁移方法中广泛地用于构建面部表情的描述[76-77]. 将面部特征点与面部细节生成相结合, 能够实现更真实的面部表情迁移[78-79], 有助于提升表情迁移效果.

基于隐式表达的表情迁移方法很少涉及三维人脸模型. 参数化人脸模型可以通过替换表情系数实现表情迁移, 但是, 由不同基础表情构成的三维人脸则无法直接替换表情系数. 为此, Zhang 等[80]提出一种用于表情域转换的隐式神经网络, 能够将三维人脸模型的表情成分映射为表情向量, 根据人工标注的数据建立 2 个表情向量的对应关系, 并通过转换表情向量实现表情迁移, 如图 20 所示; 该网络为表情迁移提供了一个易用的表情系数转换方法, 但是对于新的模型需要重新训练, 通用性较低.

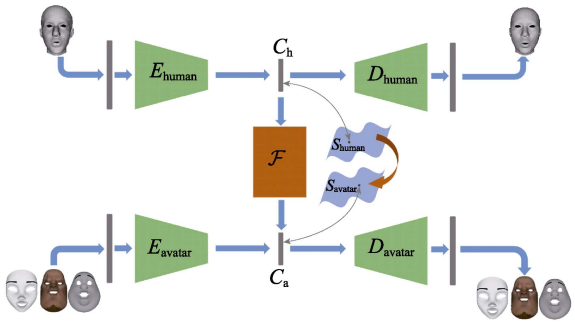


图 20 表情向量映射网络[80]

同样地, 为了映射 2 个不同模型的表情系数, Kim 等[81]提出一种基于图像映射的三维表情迁移网络, 将三维人脸模型渲染为图像, 并在图像层面进行表情迁移, 然后根据得到的表情图像估算三维模型的表情系数; 由于该网络从图像层面获得表情迁移, 因此可以借助成熟的图像转换方法获得较好的效果, 但是表情图像可能丢失面部表情的部分细节, 降低表情迁移的准确性. Li 等[82]提出一种基于高斯场的动态表情迁移技术, 通过对面部和口腔区域分解建模, 实现高保真度的动态表情生成. 该方法中, 面部区域的高斯场分布可以表示为

$$G(x)=\sum_{i=1}^{N} w_i \exp\left(-\frac{\|x-\mu_i\|^2}{2\sigma_i^2}\right).$$

其中，u_i 和 σ_i^2 分别表示第 i 个高斯核的中心和标准差；w_i 表示权重系数. 这种表示方法能够有效地捕捉面部的局部变形特征，通过结合隐式神经网络的灵活性和高斯分布的高效性，能够在保持细节精确性的同时支持复杂的动态表情和姿态，如图 21 所示.

在隐式表达方法中，三维人脸重建的研究不仅专注于高精度建模，还为动态表情迁移提供了几何和特征支持. 大部分隐式表达的三维人脸重建方法采用 3DMM，其参数化表示能够有效地描述复杂的面部形状与表情变化[83]. 这类模型为动态表情迁移提供了重要的几何基础和灵活的表达能力. Chaudhuri 等[84]提出一个同步检测图像中人脸位置和估计 3DMM 参数的隐式网络，能够同时重建图像中的多个三维人脸模型，如图 22 所示. 训练用于三维人脸重建的隐式神经网络需要人脸图像和对应的人脸模型参数，但当前包含这类数据的公开数据集有限. 为此，Tran 等[85]提出一种生成大量训练数据的方法，并在此基础上搭建并训练了能够进行高质量三维人脸重建的隐式神经网络.

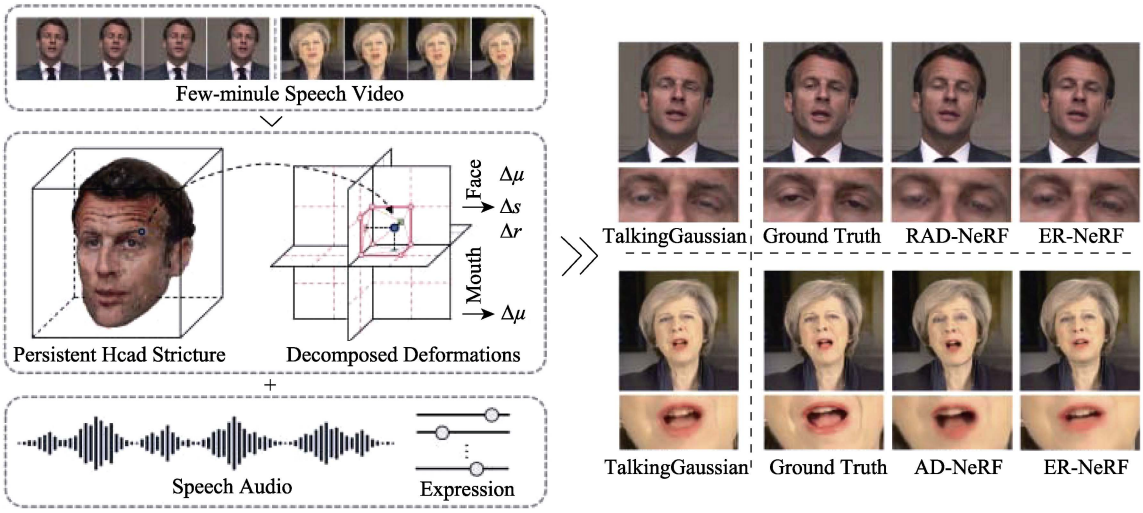


图 21　基于高斯分布的面部表情建模与迁移方法[82]

图 22　表情向量映射网络[84]

Tewari 等[86]提出一种不依赖三维数据的人脸重建方法，在估算人脸形状参数的同时，对人脸姿态、皮肤反射和光照进行预测，并根据这些参数将三维人脸渲染为二维图像，通过约束渲染的人脸图像与输入图像一致实现训练；该方法仅需要人脸图像数据即可进行训练，因此能够采用丰富的公开数据集，并且也能获得相当不错的三维人脸重建效果. 在此基础上，Genova 等[87]提出一种通用的可微分渲染器，能够按照任意角度和光照将三维人脸模型渲染为二维图像，避免因渲染图像质量低导致的训练误差. 采用可微分渲染器的三维人脸重建方法能够仅通过人脸图像完成隐式神经网络的训练[88-89]. Feng 等[90]通过可微分渲染器实现无监督的学习，其流程如图 23 所示. 可微分渲染器不仅提升了训练网络的效率，而且也为改善人脸重建效果提供了有力的工具.

Chaudhuri 等[91]认为，线性模型不足以表达丰富多变的面部表情，因此在从人脸图像中估计可变形人脸模型参数的同时，也预测了个性化的表情形状，并采用动态反照率贴图对三维人脸的形状和外观进行修正，获得了更准确的三维人脸重建. 类似地，Zhu 等[92]提出一种捕获个性化人脸形状的方法，通过模拟渲染多个角度的人脸图像获得模型顶点的偏移量，使得重建的三维形状与输入的人脸图像具有更高的视觉相似度. 基于上述可变形建模方法，预测 3DMM 的各项参数可以重建三维人脸模型，同时也成为实现表情迁移的重

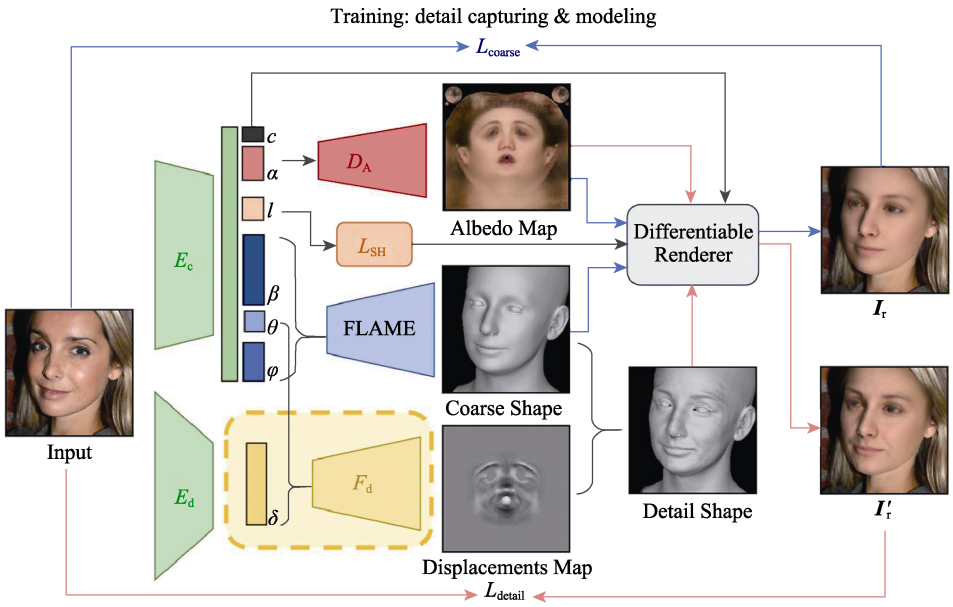


图 23 基于可微分渲染器的无监督训练[90]

要手段. 但是在实际人脸模型重建过程中, 表情系数与身份系数通常是同时回归的, 二者存在耦合, 易导致表情迁移偏差.

当前, 面部表情迁移是为虚拟数字人创造面部表情的主流方式, 国内外研究人员围绕不同表现形式的面部表情迁移展开研究, 在面部形状重建与表情映射方面取得了一系列成果. 隐式表达方法的广泛应用为准确地描述面部表情提供了支持, 尽管现有方法在不同场景下表现出显著优势, 但仍存在如数据需求、实时性和复杂性等方面的挑战, 仍需进一步探索.

为比较当前主流建模范式在三维人脸表情迁移任务中的性能表现, 本文选取具有代表性的方法, 从训练数据规模、表情种类、生成质量与实时性能等维度进行归纳对比, 如表 1 所示. 从表中数据可以直观观察到不同技术路径的性能差异, 验

表 1 10 种方法的三维人脸表情迁移性能对比

方法	出版年	训练数据集	测试数据集	数据集规模	表情类别	评价指标	实时性/(帧·s^{-1})
DDE[63]	2014	FacWarehouse, LFW, GTAV	无明确提及	14 460 帧	无明确类别标注	无具体数值	28
Face2Face[65]	2016	YouTube 视频片段, 网络摄像头采集的视频	YouTube 视频片段	测试序列: 965, 1 436 和 1 791 帧	全范围面部表情	无具体数值	27.6~28.4
FLAME[5]	2017	CAESAR 3 800 个头部扫描数据, D3DFACS, 自采集序列	BU-3DFE, 自采集序列, Beeler 等人序列	训练数据约 33 000 帧(采样 21 000 帧用于模型训练)	基于正交表达空间的全范围面部表情	训练数据: 90%顶点误差<0.5 mm; 测试数据: 67%~75%顶点误差<1.0 mm	60
X2Face[57]	2018	VoxCeleb	VoxCeleb, AFLW, Beeler 等人序列	900 764 帧(训练), 125 131 帧(测试)	无明确类别标注	平均重建误差: 0.052 1~0.063 2 mm	实时, 未给出具体帧率
ReenactGAN[75]	2018	CelebV, WFLW, Helen	CelebV, 自采集序列	200 000 帧	无明确类别标注	AU 相关性: 84.7%~92.5%	30
FaceScape[6]	2020	FaceScape(888 人, 17 760 幅置换图)	FaceScape (50 人)	18 760 帧	20 种	平均重建误差: 1.39 mm (全部)/1.22 mm (源)	实时, 未给出具体帧率
FLNet[79]	2020	TCD-TIMIT, FaceForensics	TCD-TIMIT, FaceForensics	TCD-TIMIT: 62 位说话者 6 913 个视频; FaceForensics: 1 004 个视频	无明确类别标注	L_1 误差: TIMIT 7.99, FaceForensics 10.20; FID: TIMIT 17.07, FaceForensics 20.62	实时, 未给出具体帧率

续表

方法	出版年	训练数据集	测试数据集	数据集规模	表情类别	评价指标	实时性/(帧·s^{-1})
i3DMM[18]	2021	64 人头部扫描数据	6 人头部扫描数据	64 人头部扫描数据(训练), 6 人头部扫描数据(测试)	10 种	Chamfer 距离: 3.31 mm↓, *F*-score: 63.38%↑	未明确说明
NPHM[19]	2023	203 人, 3 720 个头部扫描数据	6 位女性 12 位男性	3 720 帧(训练), 414 帧(测试)	23 种	Chamfer 距离: 0.001 82 mm↓, 法向量一致性: 0.978↑, *F*-score: 0.954↑	未明确说明
TalkingGaussian[82]	2025	主流公开说话头像数据集: Macron, Lieu, Obama, May	视频: 原视频分割验证; 音频: NVP/SynObama 音频、跨语言音频、跨性别音频	26 000 帧(4 段视频, 每段 6 500 帧)	9 种	PSNR: 33.61 dB, LPIPS: 0.025 9, SSIM: 0.910, Sync-C: 6.516	108

证了前文对各类表情迁移方法的分析. 可以看出, 各类方法在技术实现路径与应用适配方向上呈现明显差异. 隐式表达方法, 如 NPHM[19], TalkingGaussian[82]等)通过表格中的质量评估指标(如峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性指数(structural similarity index measure, SSIM)、倒角距离(chamfer distance)等, 显示在细节还原与自然度方面表现优越, 适用于高保真动态表情生成. 3DMM 方法依托其精确的平均重建误差数据, 具备良好结构稳定性与视角通用性, 适用于多视角动态重建. 而结构驱动方法(如 DDE[63], Face2Face[65]等)则从帧率数据可见其推理速度优势, 适合实时交互与灵活部署场景. 这种基于量化指标的技术分化为选择合适的表情迁移方法以满足不同应用需求提供了数据支持.

5　结论与展望

自然逼真的面部表情是提升数字人交互真实感与沉浸感的关键. 作为当前主流路径, 表情迁移通过捕捉演员表情并映射至虚拟角色, 广泛应用于虚拟社交与远程协作等场景, 展现出其良好的适应性与应用价值. 本文系统地梳理了三维人脸表示、面部成分提取与表情迁移的相关研究, 回顾了国内外技术进展, 旨在为表情迁移方法的优化提供理论支持, 推动其在精度与效率上的进一步提升.

对未来工作的展望如下:

(1) 在构建 3DMM 的过程中, 为了准确地刻画三维人脸形状, 需要采集大量的基础表情模型, 并且生成的面部形状也受限于所采用的基础模型. 可针对具备更强局部表达能力的可变形人脸建模方法开展研究, 结合表情系数映射策略, 提升模型在面部表情迁移任务中的适应性与准确性.

(2) 人脸图像中的身份信息和表情信息都会反映到图像的呈现效果中, 准确地分离身份和表情信息是实现面部表情迁移的关键问题. 可通过卷积隐式神经网络提取面部成分特征, 并引入变分自编码器对其进行分布约束, 从而实现更精确的表情表达建模, 增强表情迁移系统的泛化与可控能力.

(3) 当前三维表情迁移在模型间的适应性仍有限, 受制于线性建模方式和烦琐的标定流程, 难以推广至更广泛的应用场景. 可探索基于隐式神经网络与非线性三维人脸表示的融合框架, 简化表情迁移流程, 并通过引入数据增强策略提升模型的表达丰富度与跨主体泛化能力, 从而增强方法的实用性与普适性.

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. New York: 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] //Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance. Los Alamitos: IEEE Computer Society Press, 2009: 296-301

[3] Cao C, Weng Y L, 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

[4] Ekman P, Friesen W V. Facial action coding system: a technique for the measurement of facial movement[M]. Palo Alto: Consulting Psychologists Press, 1978

[5] Li T Y, Bolkart T, Black M J, *et al*. Learning a model of facial shape and expression from 4D scans[J]. ACM Transactions on Graphics, 2017, 36(6): Article No.194

[6] Yang H T, Zhu H, Wang Y R, *et al*. Facescape: a large-scale high quality 3D face dataset and detailed riggable 3D face prediction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 598-607

[7] Zollhöfer M, Thies J, Garrido P, *et al*. State of the art on monocular 3D face reconstruction, tracking, and applications[J]. Computer Graphics Forum, 2018, 37(2): 523-550

[8] Black M J, Yacoob Y. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 1995: 374-381

[9] Joshi P, Tien W C, Desbrun M, *et al*. Learning controls for blend shape based realistic facial animation[C] //Proceedings of the ACM SIGGRAPH 2006 Courses. New York: ACM Press, 2006: 17-es

[10] Tena J R, De la Torre F, Matthews I. Interactive region-based linear 3D face models[J]. ACM Transactions on Graphics, 2011, 30(4): Article No.76

[11] Cao C, Chai M L, Woodford O, *et al*. Stabilized real-time face tracking via a learned dynamic rigidity prior[J]. ACM Transactions on Graphics, 2018, 37(6): Article No.233

[12] Neumann T, Varanasi K, Wenger S, *et al*. Sparse localized deformation components[J]. ACM Transactions on Graphics, 2013, 32(6): Article No.179

[13] Wu C L, Bradley D, Gross M, *et al*. An anatomically-constrained local deformation model for monocular face capture[J]. ACM Transactions on Graphics, 2016, 35(4): Article No.115

[14] Brunton A, Bolkart T, Wuhrer S. Multilinear wavelets: a statistical shape space for human faces[C] //Proceedings of the 13th European Conference on Computer Vision. Heidelberg: Springer, 2014: 297-312

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

[16] Jiang Z H, Wu Q Y, Chen K Y, *et al*. Disentangled representation learning for 3D face shape[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 11949-11958

[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. Los Alamitos: IEEE Computer Society Press, 2017: 1031-1039

[18] Yenamandra T, Tewari A, Bernard F, *et al*. i3DMM: deep implicit 3D morphable model of human heads[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 12798-12808

[19] Giebenhain S, Kirschstein T, Georgopoulos M, *et al*. Learning neural parametric head models[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 21003-21012

[20] Raj A, Zollhoefer M, Simon T, *et al*. PVA: pixel-aligned volumetric avatars[OL]. [2024-06-17]. https://arxiv.org/abs/2101.02697

[21] Wang Z Y, Bagautdinov T, Lombardi S, *et al*. Learning compositional radiance fields of dynamic human heads[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 5700-5709

[22] Hong Y, Peng B, Xiao H Y, *et al*. HeadNeRF: a realtime NeRF-based parametric head model[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 20342-20352

[23] Gafni G, Thies J, Zollhöfer M, *et al*. Dynamic neural radiance fields for monocular 4D facial avatar reconstruction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 8645-8654

[24] Athar S, Shu Z X, Samaras D. FLAME-in-NeRF: neural control of radiance fields for free view face animation[C] //Proceedings of the 17th IEEE International Conference on Automatic Face and Gesture Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 1-8

[25] Qi Charles R, Su H, Kaichun M, *et al*. PointNet: deep learning on point sets for 3D classification and segmentation[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 77-85

[26] Bhople A R, Shrivastava A M, Prakash S. Point cloud based deep convolutional neural network for 3D face recognition[J]. Multimedia Tools and Applications, 2021, 80(20): 30237-30259

[27] Li J X, Zhu F Y, Yang X, *et al*. 3D face point cloud super-resolution network[C] //Proceedings of the IEEE International Joint Conference on Biometrics. Los Alamitos: IEEE Computer Society Press, 2021: 1-8

[28] Ranjan A, Bolkart T, Sanyal S, *et al*. Generating 3D faces using convolutional mesh autoencoders[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 725-741

[29] Hanocka R, Hertz A, Fish N, *et al*. MeshCNN: a network with an edge[J]. ACM Transactions on Graphics, 2019, 38(4): Article No.90

[30] Feng Y T, Feng Y F, You H X, *et al*. MeshNet: mesh neural network for 3D shape representation[C] //Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto:

AAAI Press, 2019: 8279-8286

[31] Feng Y, Wu F, Shao X H, *et al*. Joint 3D face reconstruction and dense alignment with position map regression network[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 557-574

[32] Moschoglou S, Ploumpis S, Nicolaou M A, *et al*. 3DFaceGAN: adversarial nets for 3D face representation, generation, and translation[J]. International Journal of Computer Vision, 2020, 128(10): 2534-2551

[33] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets[C] //Proceedings of the 28th International Conference on Neural Information Processing Systems. New York: ACM Press, 2014: 2672-2680

[34] Bailey S W, Omens D, Dilorenzo P, *et al*. Fast and deep facial deformations[J]. ACM Transactions on Graphics, 2020, 39(4): Article No.94

[35] Li R L, Bladin K, Zhao Y J, *et al*. Learning formation of physically-based face attributes[C] //Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 3407-3416

[36] Edwards G J, Taylor C J, Cootes T F. Learning to identify and track faces in image sequences[C] //Proceedings of the 6th International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 1998: 317-322

[37] Edwards G J, Cootes T F, Taylor C J. Face recognition using active appearance models[C] //Proceedings of the 5th European Conference on Computer Vision. Heidelberg: Springer, 1998: 581-595

[38] Zhou S K, Chellappa R. Probabilistic identity characterization for face recognition[C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: ACM Press, 2004: 805-812

[39] Zhou H L, Lam K M. Age-invariant face recognition based on identity inference from appearance age[J]. Pattern Recognition, 2018, 76: 191-202

[40] Gritti T, Shan C F, Jeanne V, *et al*. Local features based facial expression recognition with face registration errors[C] //Proceedings of the 8th IEEE International Conference on Automatic Face & Gesture Recognition. Los Alamitos: IEEE Computer Society Press, 2008: 1-8

[41] Happy S L, Routray A. Automatic facial expression recognition using features of salient facial patches[J]. IEEE transactions on Affective Computing, 2015, 6(1): 1-12

[42] Pantic M, Rothkrantz L J M. Facial action recognition for facial expression analysis from static face images[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2004, 34(3): 1449-1461

[43] Jiang B H, Valstar M F, Pantic M. Action unit detection using sparse appearance descriptors in space-time video volumes[C] //Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition. Los Alamitos: IEEE Computer Society Press, 2011: 314-321

[44] Nakada M, Wang H, Terzopoulos D. AcFR: active face recognition using convolutional neural networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2017: 35-40

[45] Kang B N, Kim Y, Kim D. Pairwise relational networks for face recognition[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 646-663

[46] Li H X, Hu H F, Yip C. Age-related factor guided joint task modeling convolutional neural network for cross-age face recognition[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(9): 2383-2392

[47] Huang Y J, Hu H F. A parallel architecture of age adversarial convolutional neural network for cross-age face recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(1): 148-159

[48] Liu X F, Vijaya Kumar B V K, Jia P, *et al*. Hard negative generation for identity-disentangled facial expression recognition[J]. Pattern Recognition, 2019, 88: 1-12

[49] Wang C, Wang S F, Liang G. Identity-and pose-robust facial expression recognition through adversarial feature learning[C] //Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM Press, 2019: 238-246

[50] Meng Z B, Liu P, Cai J, *et al*. Identity-aware convolutional neural network for facial expression recognition[C] //Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 558-565

[51] Cai J, Meng Z B, Khan A S, *et al*. Identity-free facial expression recognition using conditional generative adversarial network[C] //Proceedings of the IEEE International Conference on Image Processing. Los Alamitos: IEEE Computer Society Press, 2021: 1344-1348

[52] Zhang H F, Su W, Yu J, *et al*. Identity–expression dual branch network for facial expression recognition[J]. IEEE Transactions on Cognitive and Developmental Systems, 2021, 13(4): 898-911

[53] Li K, Xu F, Wang J, *et al*. A data-driven approach for facial expression synthesis in video[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 57-64

[54] Averbuch-Elor H, Cohen-Or D, Kopf J, *et al*. Bringing portraits to life[J]. ACM Transactions on Graphics, 2017, 36(6): Article No.196

[55] Thies J, Zollhöfer M, Nießner M, *et al*. Real-time expression transfer for facial reenactment[J]. ACM Transactions on Graphics, 2015, 34(6): Article No.183

[56] Sumner R W, Popović J. Deformation transfer for triangle meshes[J]. ACM Transactions on Graphics, 2004, 23(3): 399-405

[57] Wiles O, Koepke A S, Zisserman A. X2Face: a network for controlling face generation using images, audio, and pose codes[C] //Proceedings of the 15th European Conference on

Computer Vision. Heidelberg: Springer, 2018: 690-706

[58] Buck I, Finkelstein A, Jacobs C, *et al*. Performance-driven hand-drawn animation[C] //Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference. New York: ACM Press, 2006: 25-es

[59] Liu Z C, Shan Y, Zhang Z Y. Expressive expression mapping with ratio images[C] //Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 2001: 271-276

[60] Li K, Dai Q H, Wang R P, *et al*. A data-driven approach for facial expression retargeting in video[J]. IEEE Transactions on Multimedia, 2014, 16(2): 299-310

[61] Weise T, Bouaziz S, Li H, *et al*. Realtime performance-based facial animation[J]. ACM Transactions on Graphics, 2011, 30(4): Article No.77

[62] Cao C, Weng Y L, Lin S, *et al*. 3D shape regression for real-time facial animation[J]. ACM Transactions on Graphics, 2013, 32(4): Article No.41

[63] Cao C, Hou Q M, Zhou K. Displaced dynamic expression regression for real-time facial tracking and animation[J]. ACM Transactions on Graphics, 2014, 33(4): Article No.43

[64] Li H, Weise T, Pauly M. Example-based facial rigging[J]. ACM Transactions on Graphics, 2010, 29(4): Article No.32

[65] Thies J, Zollhöfer M, Stamminger M, *et al*. Face2Face: real-time face capture and reenactment of RGB videos[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 2387-2395

[66] Deng Z G, Chiang P Y, Fox P, *et al*. Animating blendshape faces by cross-mapping motion capture data[C] //Proceedings of the Symposium on INTERACTIVE 3D Graphics and Games. New York: ACM Press, 2006: 43-48

[67] Kholgade N, Matthews I, Sheikh Y. Content retargeting using parameter-parallel facial layers[C] //Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation. New York: ACM Press, 2011: 195-204

[68] Huang H D, Chai J X, Tong X, *et al*. Leveraging motion capture and 3D scanning for high-fidelity facial performance acquisition[J]. ACM Transactions on Graphics, 2011, 30(4): Article No.74

[69] Seol Y, Lewis J P, Seo J, *et al*. Spacetime expression cloning for blendshapes[J]. ACM Transactions on Graphics, 2012, 31(2): Article No.14

[70] Ribera R B I, Zell E, Lewis J P, *et al*. Facial retargeting with automatic range of motion alignment[J]. ACM Transactions on graphics, 2017, 36(4): Article No.154

[71] Suwajanakorn S, Seitz S M, Kemelmacher-Shlizerman I. What makes tom hanks look like tom hanks[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 3952-3960

[72] Rotger G, Lumbreras F, Moreno-Noguer F, *et al*. 2D-to-3D facial expression transfer[C] //Proceedings of the 24th International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 2008-2013

[73] Isola P, Zhu J Y, Zhou T H, *et al*. Image-to-image translation with conditional adversarial networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 5967-5976

[74] Bansal A, Ma S G, Ramanan D, *et al*. Recycle-GAN: unsupervised video retargeting[C] //Proceedings of the 15th European Conference on Computer Vision . Heidelberg: Springer, 2018: 122-138

[75] Wu W, Zhang Y X, Li C, *et al*. ReenactGAN: learning to reenact faces via boundary transfer[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 622-638

[76] Zakharov E, Shysheya A, Burkov E, *et al*. Few-shot adversarial learning of realistic neural talking head models[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 9458-9467

[77] Zhao R Q, Wu T Y, Guo G D. Sparse to dense motion transfer for face image animation[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Los Alamitos: IEEE Computer Society Press, 2021: 1991-2000

[78] Geng J H, Shao T J, Zheng Y Y, *et al*. Warp-guided GANs for single-photo facial animation[J]. ACM Transactions on Graphics, 2018, 37(6): Article No.231

[79] Gu K X, Zhou Y Q, Huang T. FLNet: landmark driven fetching and learning network for faithful talking facial animation synthesis[C] //Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 10861-10868

[80] Zhang J Y, Chen K Y, Zheng J M. Facial expression retargeting from human to avatar made easy[J]. IEEE Transactions on Visualization and Computer Graphics, 2022, 28(2): 1274-1287

[81] Kim S, Jung S, Seo K, I, *et al*. Deep learning-based unsupervised human facial retargeting[J]. Computer Graphics Forum, 2021, 40(7): 45-55

[82] Li J H, Zhang J W, Bai X, *et al*. TalkingGaussian: structure-persistent 3D talking head synthesis via Gaussian splatting[C] //Proceedings of the 18th European Conference on Computer Vision. Heidelberg: Springer, 2025: 127-145

[83] Dou P F, 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. Los Alamitos: IEEE Computer Society Press, 2017: 1503-1512

[84] Chaudhuri B, Vesdapunt N, Wang B Y. Joint face detection and facial motion retargeting for multiple faces[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 9711-9720

[85] 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. Los Alamitos: IEEE Computer Society Press, 2017: 1493-1502

[86] Tewari A, Zollhöfer 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. Los Alamitos: IEEE Computer Society Press, 2017: 3735-3744

[87] 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. Los Alamitos: IEEE Computer Society Press, 2018: 8377-8386

[88] Deng Y, Yang J L, Xu S C, *et al*. Accurate 3D face reconstruction with weakly-supervised learning: from single image to image set[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2019: 285-295

[89] Lee G H, Lee S W. Uncertainty-aware mesh decoder for high fidelity 3D face reconstruction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 6099-6108

[90] Feng Y, Feng H W, Black M J, *et al*. Learning an animatable detailed 3D face model from in-the-wild images[J]. ACM Transactions on Graphics, 2021, 40(4): Article No.88

[91] Chaudhuri B, Vesdapunt N, Shapiro L, *et al*. Personalized face modeling for improved face reconstruction and motion retargeting[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 142-160

[92] Zhu X Y, Yu C, Huang D, *et al*. Beyond 3DMM: learning to capture high-fidelity 3D face shape[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 1442-1457