

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2024)09-2441-30

论文引用格式: Liu F, Zhang K B, Yang Q, Zhou S B, Wang Y L and Sun Z N. 2024. 3D face imaging and reconstruction technology: a review. Journal of Image and Graphics, 29(09):2441-2470(刘菲, 张堃博, 杨青, 周树波, 王云龙, 孙哲南. 2024. 三维人脸成像及重建技术综述. 中国图象图形学报, 29(09):2441-2470)[DOI:10.11834/jig.230697]

三维人脸成像及重建技术综述

刘菲¹, 张堃博², 杨青¹, 周树波³, 王云龙², 孙哲南^{2*}

1. 首都经济贸易大学管理工程学院, 北京 100070; 2. 中国科学院自动化研究所模式识别实验室, 北京 100190;

3. 东华大学信息科学与技术学院, 上海 201620

摘要: 得益于新型三维视觉测量技术及深度学习模型的飞速发展, 三维视觉成为人工智能、虚拟现实等领域的重要支撑技术, 三维人脸成像及重建技术取得了突破性进展, 不仅能够更好地应对光照、遮挡、表情和姿态等变化, 同时增大了伪造攻击难度, 大大推动了真实感“虚拟数字人”的重建与渲染, 有效提升了人脸系统的安全性。本文对三维人脸成像技术和重建模型进行了全面综述, 尤其对基于深度学习的三维人脸重建进行系统深入地分析。首先, 对三维人脸成像设备及采集系统进行详细梳理及对比归纳, 并介绍了基于新传感技术的人脸成像系统; 然后, 对基于深度学习的三维人脸重建模型进行系统分析, 从输入数据源角度分为基于单目图像、基于多目图像、基于视频和基于语音的三维人脸重建算法4类。通过深入分析, 总结三维人脸成像的研究现状及面临的难点与挑战, 对未来发展方向及应用进行积极探讨与展望。本文涵盖了近5年经典的三维人脸成像及重建相关的技术与研究, 为人脸研究、发展和应用提供了很好的参考。

关键词: 三维人脸成像; 三维人脸重建; 深度学习 (DL); 生成对抗网络 (GAN); 隐式神经表示 (INR)

3D face imaging and reconstruction technology: a review

Liu Fei¹, Zhang Kunbo², Yang Qing¹, Zhou Shubo³, Wang Yunlong², Sun Zhenan^{2*}

1. School of Management and Engineering, Capital University of Economics and Business, Beijing 100070, China;

2. National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;

3. College of Information Science and Technology, Donghua University, Shanghai 201620, China

Abstract: As the breakthrough technology of artificial intelligence (AI) in the big data era, deep learning (DL) has prompted the renewed upsurge of face technology. Powered by rapid developments of new technologies, such as three-dimensional (3D) vision measurement, image processing chips, and DL models, 3D vision transformed into a key supporting technology in AI, visual reality, etc. The studies and applications of 3D facial imaging and reconstruction technologies have achieved important breakthroughs. 3D face data represent exact multidimensional facial attributes on account of rich visual information, such as texture, shape, space, etc. Moreover, 3D face data shows robust changes in large occlusions, expressions, and poses and increases the difficulty of forgery attack. Therefore, 3D face imaging and reconstruction effectively promote realistic “virtual digital human” reconstruction and rendering. In addition, these processes contribute to the improved security of the face system. In this paper, we comprehensively study the 3D face imaging technology and reconstruction models. The 3D face reconstruction methods based on DL are systematically and deeply analyzed. First, the

收稿日期: 2023-09-28; 修回日期: 2024-03-20; 预印本日期: 2024-03-27

* 通信作者: 孙哲南 zhenan.sun@ia.ac.cn

基金项目: 国家自然科学基金项目 (61806197, 61803372, 62071468, U23B2054, 62276263)

Supported by: National Natural Science Foundation of China (61806197, 61803372, 62071468, U23B2054, 62276263)

development and innovation of 3D face imaging devices and capturing systems are discussed through a summary of public 3D face datasets. The devices and systems include consumer imaging devices (such as Kinect) and complex hybrid systems that fuse active and passive 3D imaging technologies to achieve precise geometry and appearance. Moreover, 3D face imaging based on new sensing technologies are introduced. Then, from the perspective of input resources, 3D face reconstruction methods based on DL are categorized into monocular, multiview, video and audio reconstruction methods. 3D face imaging technology introduces public classic 3D face datasets, popular 3D face imaging devices, and capturing systems. Most high-quality 3D face datasets, such as BU-3DFE, FaceScape and FaceVerse, are captured through a large imaging volume with a certain number of high-resolution cameras and controlled lighting conditions. They play key roles in applications of realistic rendering, driven animation, retargeting, etc. On the other hand, novel optical devices and imaging modules with small size and lightweight algorithm must be innovated for tiny AI as intelligent mobile devices. For 3D face reconstruction based on DL, monocular reconstruction has become the most popular technology. The state-of-the-art 3D face reconstruction method is generally self-supervised training on large-scale 2D face databases. The difficulties encountered in 3D face reconstruction include the lack of large-scale 3D face datasets, occlusions and poses of in-the-wild 2D face images, continuous expression deformations, etc. The DL network structure is categorized into general deep convolutional neural network (such as ResNet, U-Net, and Autoencoder), generative adversarial networks (GANs), implicit neural representation (INR) (such as neural radiance field (NeRF) and signed distance functions (SDF)), and Transformer. 3DMM and FLAME are widely used 3D face representation models. The StyleGAN model gives excellent performance in recovering high-quality face texture. INR has achieved remarkable results in 3D scene reconstruction, and the NeRF model plays an important role in the reconstruction of accurate head avatars. The combination of NeRF with GAN shows great potential in the reconstruction of high-fidelity 3D face geometry and realistic rendering appearances. Moreover, the Transformer model, which greatly improves the breakthrough of accuracy and speed, is mainly used in audio-driven 3D face reconstruction. Through in-depth analyses, the research difficulties accompanying 3D face are summarized, and future developments are actively being discussed and explored. Although recent research has made amazing progresses, challenges on how to improve the robustness and generalization to real-world lighting, extreme expressions/poses, and how to effectively disentangle facial attributes (such as identity, expression, albedo, and specular reflectance) and recover accurate detailed geometry of facial motions (such as wrinkles). In this study, we proposed a comprehensive and systematic review and covered classical technologies and studies on 3D face imaging and reconstruction in the last five years to provide a good reference for face studies, developments, and applications.

Key words: 3D face imaging; 3D face reconstruction; deep learning (DL); generative adversarial network (GAN); implicit neural representation (INR)

0 引言

生物特征技术是指智能机器通过获取和分析人体的生理和行为特征(如人脸、虹膜、指纹和步态等),实现个体自动身份鉴别、状态分析、属性估计的科学和技术(孙哲南等,2021)。不同于其他生物特征模态,人脸具有自然性、非强制性、非接触性和并发性的独特优势,而深度学习作为大数据时代人工智能的突破性技术,将人脸成像、人脸识别等相关技术的研究再次推向了新高潮。近10年来,三维人脸成像及重建技术不断突破创新,在国防、金融、教育和医疗等各个领域得到广泛应用,尤其是“虚拟数

字人”及智能移动终端轻量化人工智能(tiny artificial intelligence, Tiny AI)技术的飞速发展,已然成为三维人脸成像及重建技术发展与应用的核心驱动力。

相比二维人脸数据,三维人脸具有独特优势和更大潜力。首先,三维人脸数据包含纹理、形状和空间等更丰富的视觉信息,能够多维度地表征人脸属性;其次,三维人脸能够更好地应对光照、遮挡、表情以及姿态等关键因素的变化,具有更强的鲁棒性,同时增大了伪造攻击难度,有效提升了系统精度和安全性;再者,得益于新型光电成像、三维视觉测量(张宗华等,2021)等技术的发展,三维视觉成为人工智能、虚拟现实等领域的重要支撑技术,三维人脸

数据获取和模型创新不再是制约因素。因此,三维人脸成像及重建技术成为计算机视觉领域非常重要的研究方向。

三维人脸成像及重建技术,从成像过程、数据预处理、数据配准,到特征提取、网络模型,目前依然面临诸多难题与挑战。1)数据集的规模和多样性是基于学习的深度模型的基础,直接决定了精度、鲁棒性和泛化性。但是,大规模高质量三维人脸数据库相对比较缺乏,现有公开数据库在规模、表情、姿态和遮挡等方面仍具有一定局限性。2)表情和遮挡是三维人脸研究的两大难题,尤其是基于大规模野外(“in-the-wild”)二维人脸图像的三维人脸重建,复杂的表情运动及遮挡异常区域,是影响精度的重要因素。3)基于智能移动成像设备进行三维人脸采集和重建,将深度模型计算过程轻量化,将大大促进三维人脸的深入发展与推广应用(罗常伟等,2021),也是三维人脸研究面临的更大挑战。

近年来,本领域研究综述聚焦在三维人脸识别(罗常伟等,2021;Jing等,2021)和三维人脸重建

(Zollhöfer等,2018;Egger等,2020;Morales等,2021;Fu等,2021;Sharma和Kumar,2022;何嘉玉等,2022;王静婷和李慧斌,2023),其中,Zollhöfer等人(2018)总结分析了基于单目或RGB-D的人脸重建、跟踪和重演;Egger等人(2020)介绍了经典的三维可变形人脸模型(3D morphable face model, 3DMM);Sharma和Kumar(2022)将三维人脸重建技术分为3DMM、对极几何、明暗恢复形状、单样本学习、深度学习和混合学习6类;何嘉玉等人(2022)将单目三维人脸重建技术分为基于卷积神经网络、图卷积网络、自编码器结构和生成对抗网络4类;王静婷和李慧斌(2023)将单目三维人脸重建技术分为基于隐空间编码和基于显空间回归两大类。

本综述旨在让读者更全面、更深入地了解三维人脸数据获取技术(测量硬件和计算算法),从三维人脸成像技术和重建模型算法两个角度,进行系统梳理和对比分析,涵盖了主要的三维人脸成像系统及近年基于深度学习的经典算法,并对三维人脸相关研究的挑战与未来发展进行了探讨,如图1所示。

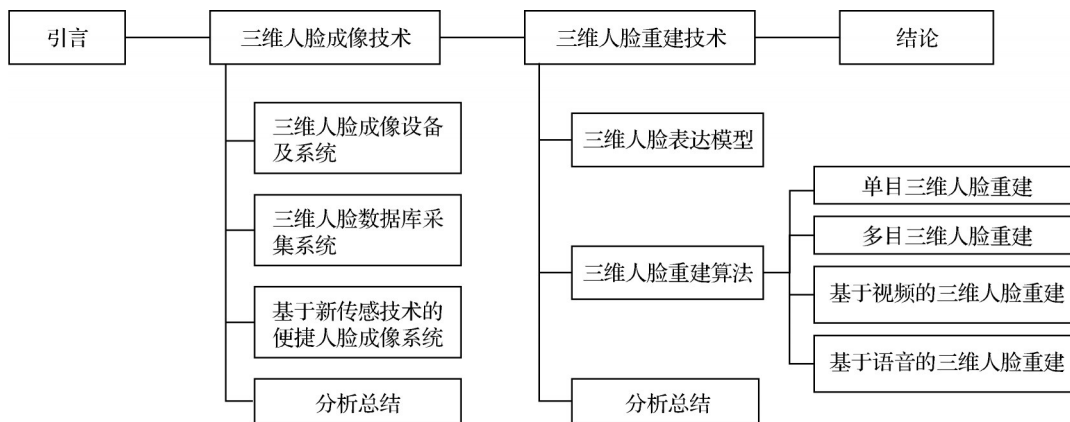


图1 本文框架结构

Fig. 1 The framework of the paper

1 三维人脸成像技术

近年来,随着三维视觉测量及深度学习技术的飞速发展,从三维传感、底层硬件到计算算法、系统方案,三维人脸研究取得了重大突破。但是,人脸的弱纹理性、复杂的光照特性及表情变化属性,都大大增加了高精度三维人脸数据获取的难度。

三维视觉测量技术按照是否投射光源分为主动视觉测量和被动视觉测量(张宗华等,2021)。主动

视觉测量需要向被测目标投射光源,相机拍摄包含光源的反射光或透射光的图像,分为结构光(如点扫描、线扫描、面扫描)、飞行时间法和相移法等。被动视觉测量硬件简单,通过相机拍摄的二维图像恢复三维形状信息,分为单目(monocular)、双目立体视觉(stereo vision)和多目立体视觉(multi-view stereo, MVS)。

三维人脸成像技术包括成熟的消费级成像设备和混合成像系统,根据视觉测量方式分为主动式、被动式和混合式。混合三维人脸成像通过融合主动、

被动成像及设计照明环境,搭建高质量采集系统。本节从三维人脸成像设备、数据库采集系统、基于新传感技术的便捷人脸成像系统3方面进行分析。

1.1 三维人脸成像设备及系统

表1对三维人脸成像设备及系统进行了详细总结归纳。目前,市场上高精度主动三维成像设备价格昂贵、测距范围有限,且对成像环境要求严苛,采集对象需要友好配合,获取的三维数据可以作为真实值(ground truth),如Vivid 910、3dMD。低精度主动三维成像设备成本低、速度快、操作便捷,但初始深度信息含较大的噪声,如Kinect。相比之下,混合式三维人脸成像系统,为三维人脸的研究与应用发挥了重要作用,如图2所示。

美国南加州大学创意技术研究所(Institute for Creative Technologies, USC)Paul Debevec团队建设了首个经典的三维人脸成像系统Light Stage。瑞士苏黎世迪士尼研究中心(disney research studios)Thabo Beeler、Derek Bradley团队提出多个应用性强的人脸采集及重建方案,建设了Medusa Facial Capture System。上海科技大学虞晶怡团队及影眸科技公司搭建亚洲唯一的微米级人脸扫描系统一穹顶光场Light Field Stage,创造更高精度的可驱动数字化身,引领了三维生成技术的突破创新,如Zhang等人(2023a)提出基于文本指导的渐进式三维人脸资产生成DreamFace,包括由粗到精的几何生成、基于物理的材质扩散模型和基于视频的动画生成,最终生成细粒度三维人脸、PBR(physically-based rendering)纹理及个性化人脸动画。

但是,如此大型的超高精度的混合三维人脸成像系统需要投入巨大的人力和算力,成本非常高。目前工业界通过提供相应的扫描数据,为相关领域学术研究的深度学习网络模型的训练服务。

1.2 三维人脸数据库采集系统

三维人脸数据存储格式有点云(point cloud)、深度图(depth map)、网格(mesh)和体素(voxel)。表2列出了常用的三维人脸数据库,大部分可通过申请获得使用许可。

FaceScape(Yang等,2020)是南京大学曹汛团队建设的三维人脸数据库,由68个单反相机构成采集系统,并提出双线性人脸表达模型,实现可操控表情的三维人脸重建。网络包括基础模型拟合、位移图预测(displacement map)和动态细节合成3个模块,

通过对齐到模板人脸生成拓扑一致的基础网格,基于pix2pixHD(Wang等,2018)学习动态表情,生成高分辨率位移图表征的几何细节,因此,精细三维人脸几何模型表示为拓扑一致的基础模型及不同表情的位移图。

另外,该团队还提出了MoFaNeRF(Zhuang等,2022)、MVFR(Xiao等,2022)、RAFaRe(Guo等,2023b)和Describe3d(Wu等,2023b)。Describe3d建设了一个大规模三维人脸文本数据集,包含1 627个来自FaceScape和HeadSpace(Dai等,2020)的高精度三维人脸及人工标注的文本描述,来学习从文本嵌入空间到三维人脸参数空间的映射并优化。

FaceVerse(Wang等,2022)是清华大学刘焯斌团队建设的细粒度混合三维人脸数据库,包括两部分:1)消费级深度传感器(kinect)捕获的大规模低质量人脸数据(60K);2)由128个单反相机构成的多目采集系统捕获的高质量人脸数据。其设计了由粗到精的参数化模型,基于大规模低质人脸数据构建基础模型,使用StyleGANv1(Karras等,2019)构建细节模型,输出高泛化性、高保真度的三维人脸。

LYHM(Dai等,2020)是William A. P. Smith、Nick Pears团队建设的头部空间数据库HeadSpace和表达模型。HIFI3D(Bao等,2021)是腾讯人工智能实验室张正友、暴林超团队建设的三维头部数据库和表达模型。输入是便携式深度相机(如iPhoneX)采集的RGB-D视频帧,并基于Beeler等人(2010)方法采集200个东亚人扫描数据,快速重建高质量完整头部形状、反射图和法线贴图。融合HeadSpace、FaceScape和HIFI3D,REALY(Chai等,2022)对三维人脸重建的评价进行了重新探索和定义,是目前最新的三维人脸评价基准数据库。通过全局缩放对齐、语义一致关键点定位、区域分割和视角渲染,基于HIFI3D生成拓扑一致的三维人脸模型,整合2 000个三维人脸扫描构建了高质量头部表达模型HIFI3D++,具有更强的表达能力和更高的网格质量。

另外,还有人脸纹理数据库FFHQ-UV(Bai等,2023)和语音驱动多模态数据库MMFace4D(Wu等,2023a)等。FFHQ-UV基于StyleGANv1进行UV纹理校正补全,包含50 000个多样化高质量UV纹理图,应用于三维人脸拟合、光照建模和真实感渲染。MMFace4D使用3个Kinect采集人脸运动视频并捕

表 1 经典的三维人脸成像设备及系统总结归纳
Table 1 Summary of 3D face imaging devices and systems

成像设备或系统	类型	测量方式	精度	特点	应用
RGB-D 相机	结构光或飞行时间	主动	>2 mm	成本低、速度快、操作便捷,但分辨率低、深度信息噪声大,需要结合 RGB 图像及重建算法(如 KinectFusion (Richard 等, 2011))来恢复高精度三维人脸数据。	用于采集目标的深度信息,在体感识别、手势识别和场景建模等应用广泛,相关设备如基于结构光的 Kinect、RealSense、华为手机和 iPhoneX 等,基于飞行时间的 Kinect2、PMD 等。
Konica Minolta Vivid900/910 (Konica, 2024)	激光扫描仪	主动	0.1 mm	基于激光三角测距原理的无接触高精度扫描仪,3 款可更换镜头进行全尺寸测量,最佳测距范围 0.6 ~ 1.2 m,最快扫描时间 0.3 s。	用于高分辨率的人体、物体等 3D 测量,应用广泛,如 CT 数据、文物扫描等计算机图形学任务。
3dMDface 系统 (3dMD, 2024)	结构光	主动	0.2 mm	由结构光投射器及两个同步模块化相机单元构成,每个模块由 3 台机器视觉摄像机和一个工业级 LED 照明系统组成,最快扫描时间 0.1 s。	用于采集 180°人脸表情动作的高精度 4D 数据。
DI4D 系统 (DI4D, 2024)	立体视觉	被动	高精度	基于立体视觉技术的 3D/4D 动态人脸捕捉处理系统,DI3D 由 4 台机器视觉摄像机组成,DI4D Pro 由九台组成,无需结构光投影和人脸标记点,每个相机分辨率为 4 112 × 3 008 像素。	用于采集高分辨率精细 3D/4D 动态人脸数据及视频游戏的逼真渲染。
Light Stage 系统 (Light Stage, 2024)	立体视觉	混合	亚毫米级	Light Stage1 (Debevec 等, 2000)使用一个照明光源、2 个相机加投影仪,绕对象螺旋下降拍摄;Light Stage5 (Wenger 等, 2005)由 156 个可控白光光源形成稠密的球形照明环境,基于 5 个单反相机采集静态或动态视频人脸数据,实现快速采集、高质量重建和重打光(Ghosh 等, 2011; Fyffe 等, 2014);Light Stage X (Debevec, 2012)向小体积可移动的设计发展,使用高速摄像机进行更快速的 3D 扫描、反射场建模和面部行为捕捉。	有效捕获任意视角和照明下的人脸亮度、纹理、反射场和法线等渲染属性,实现快速、高精度的三维人脸形状与反射场建模,对三维人脸、人体成像技术的研究与发展产生了重要影响,广泛应用于游戏开发、电影渲染等领域。
Medusa (Medusa, 2024)	立体视觉	混合	高精度	多个应用性强的人脸采集及重建方案(Beeler 等, 2010, 2011; Gotardo 等, 2018; Riviere 等, 2020)。Gotardo 等人(2018)由 8 台单色工业相机排列成 4 组立体对,插入 4 个彩色相机及 16 个 LED 白光灯带照明;Riviere 等人(2020)由 12 个单反相机和 4 个闪光灯构成,配置简单灵活,自建多视角头部数据集进行监督训练,包含 15 个人的三维扫描及对应 12 个视角图像。	实现在完全运动中采集并重建高精度高分辨率的人脸属性(漫反射、镜面反射、法线)和三维几何、逆渲染,应用于电影特效、医疗成像等领域。
Light Field Stage (Light Field Stage, 2024)	立体视觉	混合	微米级	实现高速多通道动态扫描 4D 面部几何与材质,100 多个光源,捕捉表情变化的细微肌肉运动,输出精细到皮肤微结构的逐帧三维人脸几何与基于物理的渲染材质效果(Zhang 等, 2021a, 2022a, 2023a)。基于 Light Field Stage, Zhang 等人(2022a)提出 FaStage 采集系统及视频驱动的面部资产重建算法,由 3 个高速相机、24 个工业相机和 146 个 LED 组成实现微表情的几何纹理建模及渲染、编辑、重定向等应用。	对当下元宇宙“数字人”前沿技术的研究发展具有重要意义,在 AR/VR 场景、影视特效、虚拟社交等相关行业得到广泛应用,具有更高准确性、更快速度和更好 CG 兼容性。

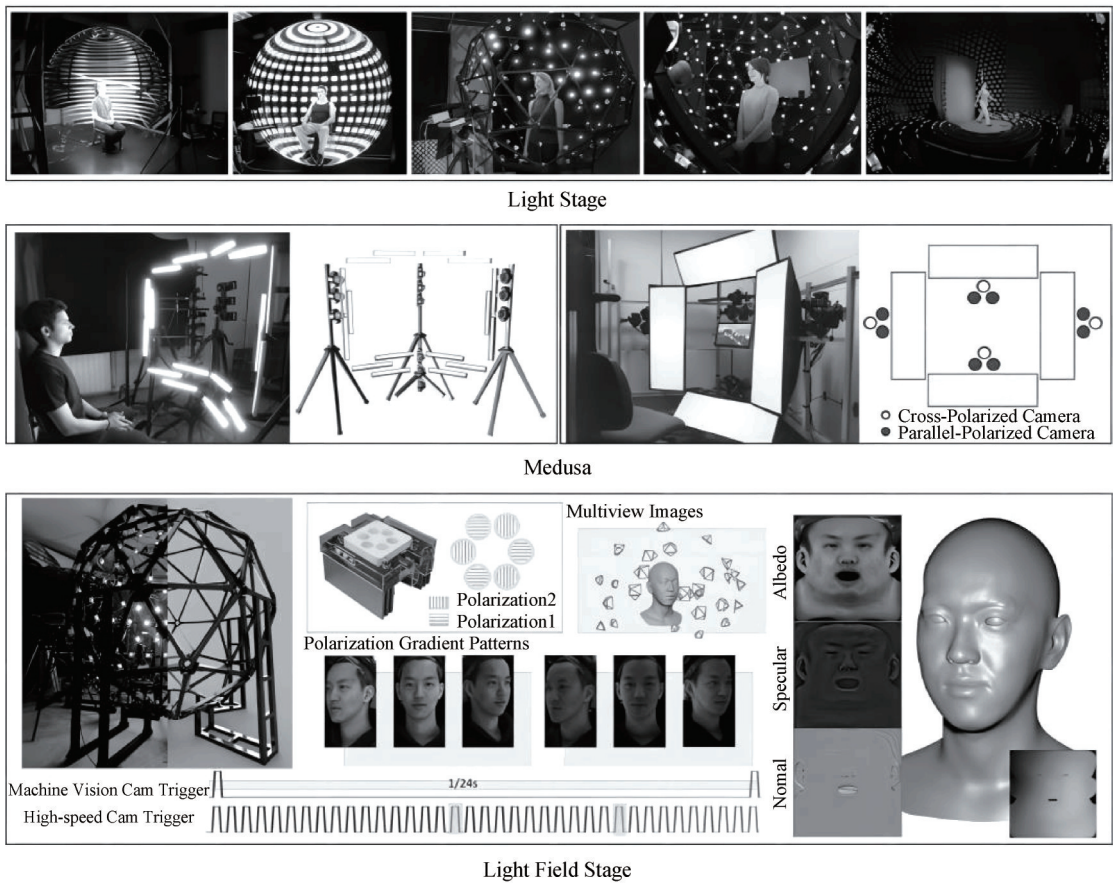


图2 高精度三维人脸成像系统
Fig. 2 High-quality 3D face capturing system

获语音音频,包含431个对象,结合3DMM人脸模型,重建拓扑统一的三维网格序列并进行压缩存储,为音频驱动人脸动画提供数据支撑。

相比工业界大型高精度混合成像系统,科研团队自主搭建的三维人脸数据库采集系统是面向不同的研究任务,在系统复杂度、数据规模等方面都更精简,通过强大的后处理及重建算法得到更高质量的三维人脸数据,推动了三维人脸研究的发展。

1.3 基于新传感技术的便捷人脸成像系统

新型图像传感器、光学成像技术与人工智能技术的协同发展,推动了三维人脸的研究。Bae等人(2022)研发微型近红外光场成像相机,实现高对比度、照明不变的三维人脸重建和表情理解。Kang等人(2022)搭建全像素双核相机系统,由8个单反相机、6个LED(light-emitting diode)照明和结构光投影仪构成,并发布DualPixelFace数据库,包括人脸图像、深度图和法线贴图。

Lattas等人(2022a)提出两种实用性强、易扩展

的便捷桌面人脸采集方案:1)5个智能手机从不同角度采集高分辨率人脸图像,用于反射和法线建模;8个平板电脑控制连续照明,平板摄像头捕捉随照明变化的低分辨率人脸图像,重建三维人脸。2)4个桌面液晶显示器,配合4对高分辨率单反相机,实现静态和动态人脸成像;通过调制二元照明模式进行漫反射/镜面反射分离、法线建模,采集高质量三维人脸形状和表观。

Cao等人(2022)只使用1台带深度传感器的智能手机进行头部多姿态多表情扫描,并使用40个彩色和50个单色摄像机搭建人脸成像系统,采集255个对象的高分辨多视角视频人脸用于网络训练。Azinović等人(2023)基于1台手机搭配偏振滤光片实现人脸捕捉和重建,在不同偏振光的黑暗环境下记录两个短序列来分离镜面反射和漫反射,再结合SFM(structure from motion)、MVS和FLAME人脸模型(Li等,2017),恢复三维人脸几何和表观。

表 2 常见的公开三维人脸数据库
Table 2 Popular public 3D face datasets

数据库	人数	样本总数	表情	姿态	遮挡	成像设备/系统	课题组
FRGCv2(Phillips 等,2005)	466	4 007	√	√	–	Minolta Vivid 910 激光扫描仪	美国国家标准技术研究所(NIST)
BU-3DFE(Yin 等,2006)	100	2 500	√	–	–	3dMD	美国纽约州立宾汉姆顿大学尹立军团队
CASIA3D(Xu 等,2006)	123	4 624	√	√	–	Minolta Vivid 910 激光扫描仪	中国科学院自动化研究所 CRIPAC 谭铁牛团队
BU-4DFE(Yin 等,2008)	101	60 600 v	√	√	–	DI3D 动态采集系统	美国纽约州立宾汉姆顿大学尹立军团队
Bosphorus(Savran 等,2008)	105	4 666	√	√	√	Inspeck Mega Capturor II 3D	土耳其海峡大学
BJUT-3D(尹宝才 等,2009)	500	1 200	√	–	–	Cyberware3030 激光扫描仪	北京工业大学计算机学院孙艳丰团队
Florence(Bagdanov 等,2011)	53	11 395 v	–	√	–	3dMD	意大利佛罗伦萨大学媒体整合与传播中心
FaceWarehouse(Cao 等,2014)	150	9 900	√	–	–	Kinect	浙江大学 CAD&CG 重点实验室周昆团队
LYHM(Dai 等,2017,2020)	1 212	1 212	–	–	–	3dMDhead 5-view 扫描系统	英国约克大学 William A. P. S、Nick Pears 团队
4DFAB(Cheng 等,2018)	180	1 800 000 v	√	–	–	DI4D 动态采集系统+Kinect	英国帝国理工学院智能行为理解组(i•bug)Stefanos Zafeiriou 团队
FaceScape(Yang 等,2020)	938	18 760	√	–	–	MVS (68 DSLR Cameras)	南京大学电子科学与工程学院曹汛团队
SIAT-3DFE(Ye 等,2020)	500	8 000	√	–	√	CASZM-3D (结构光)	中国科学院深圳先进技术研究院集成所宋展团队
FaceVerse(Wang 等,2022)	128	2 688	√	–	–	MVS(128 DSLR Cameras)	清华大学自动化系刘烨斌团队
REALY(Chai 等,2022)	100	100	–	–	–	3dMDhead system	清华大学深圳国际研究生院袁春和 腾讯人工智能实验室暴林超团队
MimicME(Papaioannou 等,2022)	4 700	280 000 v	√	–	–	3dMDface system	英国帝国理工学院智能行为理解组(i•bug)Stefanos Zafeiriou 团队
NPHM(Giebenhain 等,2023)	124	2 200	√	–	–	Artec Eva 结构光三维扫描仪	德国慕尼黑工业大学 Matthias NieBner 团队
Florence 4D(Principi 等,2023)	95	205 200 v	√	–	–	3dMD HR 扫描仪	意大利佛罗伦萨大学媒体整合与传播中心

注:“v”代表视频帧数据,“√”代表包含,“–”代表不包含。

1.4 分析总结

数据是基于学习的深度模型的关键因素,而大规模三维人脸数据的成像复杂度高且数据处理难度大,学术界和工业界一直在努力解决数据问题。

三维人脸成像技术是三维人脸研究的基础与支撑,在精度、速度和应用方面取得了巨大进步,体现

在:1)搭建精密的大体积、复杂照明和协同控制的三维人脸成像系统,融合先进的深度学习模型,实现更高精度的人脸表观属性采集、光照建模、三维几何重建、逼真渲染和动态重演等;2)不同成像设备或系统捕获的三维人脸数据差异性大,融合多个公开数据库,构建大规模、高质量和通用型三维人脸数据库,能够显著提升深度模型的精度、泛化性和鲁棒性;

3)创新光学器件、成像模组和三维传感技术,研发面向智能移动终端的小尺寸、高精度和快速便捷三维人脸成像设备,具有非常重要的意义。

然而,每一类三维成像设备和系统都有各自的优缺点和特定的应用条件,如照明环境、最佳成像距离、视角范围和分辨率等。因此,面向现实世界复杂场景,需要综合考虑应用需求、成像场景、预算成本和人力等,从测量硬件和计算算法两方面联合优化,推动三维人脸的研究发展与落地应用。

2 三维人脸重建技术

本节首先介绍三维人脸表达模型,然后从输入数据源角度,对基于深度学习的三维人脸重建算法进行分析、对比与总结,包括基于单目图像、基于多目图像、基于视频和基于语音的重建算法。

2.1 三维人脸表达模型

三维人脸表达模型是三维人脸重建的重要组成部分,分为线性和非线性模型两类,如图3所示。

三维可变形人脸模型 3DMM (Blaⁿz 和 Vetter, 1999)是基于统计的参数模型,发展至今一直受到重点关注,是引用率最高的基础表达模型,对三维人脸研究具有不可替代的意义(Egger等,2020)。FaceWarehouse(Cao等,2014)提出了身份和表情双线性人脸表达模型,广泛应用于人脸表情的表征。

AlbedoMM(Smith等,2020)是首个三维可变形人脸反照率表达模型。随后,清华大学徐枫团队提出 ReflectanceMM(Han等,2023),基于低成本公开数据训练网络,人脸反射率表达为空间变化的双向反射分布函数,即朗伯 BRDF(bidirectional reflectance distribution function)与多个 Blinn-Phong BRDFs的线性组合。

FLAME(Li等,2017)和 COMA(Ranjan等,2018)是 Timo Bolkart、Michael Black 团队提出的三维头部表达模型,是继 3DMM 应用非常广泛的表达模型。FLAME 能够精确表征头部形状、表情和姿态属性,构建个性化新表情及表情迁移;COMA 引入深度卷积网格自编码器学习人脸形状和表情的多尺度非线性表达。Tran 和 Liu(2018)提出首个非线性表达模型 Nonlinear3DMM,输入大规模野外二维人脸,通过编码器—解码器网络端到端学习三维人脸形状和纹理表达。

i3DMM(Yenamandra等,2021)、ImFace(Zheng等,2022a)和 NPHM(Giebenhain等,2023)都是基于隐式神经表示的模型,采用符号距离函数(signed distance functions, SDF)学习非线性连续空间隐式表示。i3DMM(Yenamandra等,2021)是首个隐式完整头部表达模型,解耦为隐式参考形状和形状变形,包括参考形状、几何变形和颜色子网络。ImFace(Zheng等,2022a)包括表情形变、身份形变和模板空间3个子网,将人脸特征分解为语义局部并用局部隐式函数编码,实现多样化人脸表达。

MoFaNeRF(Zhuang等,2022)是基于神经辐射场 NeRF(neural radiance field)(Mildenhall等,2020)的三维可变形人脸模型,使用 FaceScape 双线性模型,引入新的纹理编码器和特定身份调制模块,将不同个体的表情用形状和表情参数的调制关系表达,实现逼真视角渲染及形状、表情、表观的解耦表达和编辑。

2.2 三维人脸重建算法

结合前文对三维人脸成像技术的综述分析,本节将三维人脸重建算法分为:基于单目图像、基于多目图像、基于视频和基于语音4大类。另外,根据深度学习模型进行统计分类,如图4所示。

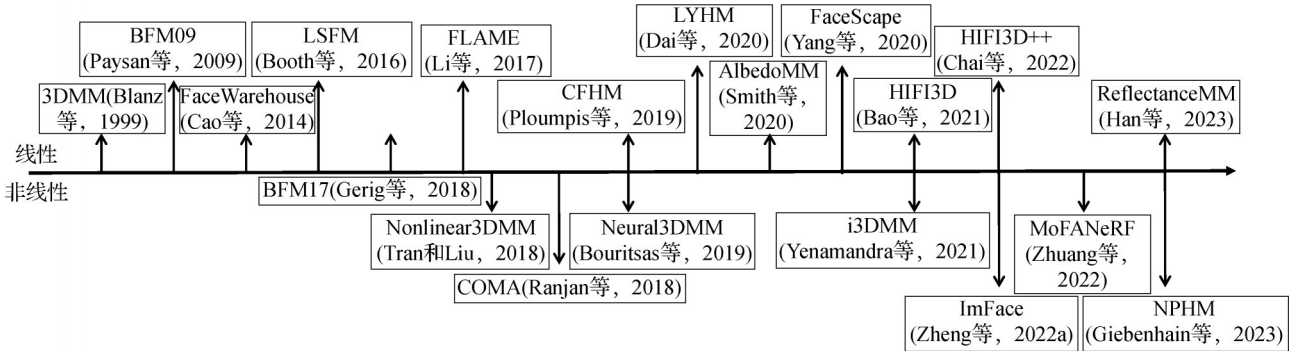


图3 三维人脸表达模型

Fig. 3 3D face representation model

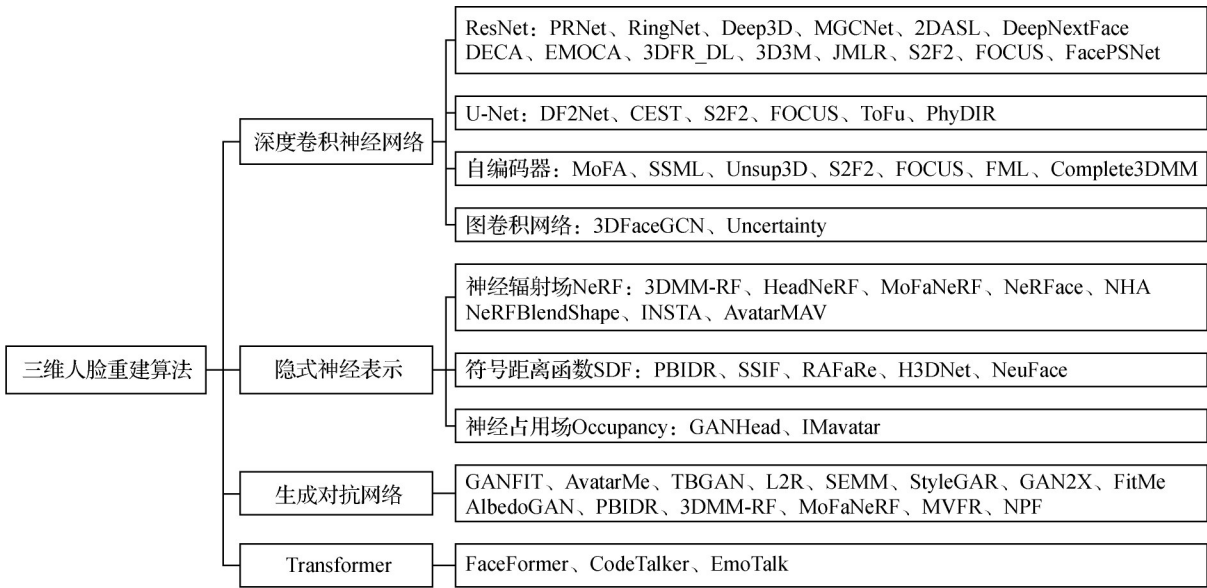


图 4 基于深度学习模型的重建算法分类
Fig. 4 A taxonomy for deep learning reconstruction algorithms

2. 2. 1 单目三维人脸重建

基于单幅二维人脸图像重建三维人脸,从成像角度而言是个具有挑战性的病态问题,但由于输入数据简单,而且基于学习的深度神经网络模型飞速发展,不断涌现的新模型具有强大的特征提取与表达能力,单目三维人脸重建已经成为当前最流行的技术,在人脸研究创新与应用中表现出巨大潜力。

根据算法主体框架使用的深度模型,主要分为基于深度卷积神经网络、基于生成对抗网络和基于隐式神经表示,表3—表5对算法、数据库、输出和表达模型等进行了详细的统计分析。基于深度卷积神经网络的算法框架通常融合多个神经网络,表中明确标示了自编码器和图卷积网络结构,其他 ResNet (He 等, 2016)、U-Net (Ronneberger 等, 2015) 等网络使用情况在详细分析一栏中明确说明。

1) 基于深度卷积神经网络的方法。人脸几何表征方法直接决定了深度学习重建模型的网络结构,包括深度图、法线贴图及 UV 空间纹理图、位移图和位置图等。VRN (Jackson 等, 2017) 提出首个体素表示三维人脸模型,无需配准直接学习从二维到三维形状的体表征映射。PRNet (Feng 等, 2018a) 提出首个 UV 位置图人脸表示模型,UV 位置图记录三维人脸形状及 UV 空间的对齐信息,实现三维人脸重建及对齐。HRN (Lei 等, 2023) 提出层次化表征网络,充分利用 UV 空间人脸表征,将几何细节解耦为低频、中频 (轮廓、局部形状) 及高频 (皱纹),设计人脸

级基形状、顶点级变形图、像素级位移图表征,目前 REALY (Chai 等, 2022) 数据库取得 SOTA (state-of-the-art) 精度。

基于大规模二维人脸数据集,结合人脸参数化统计表达模型与深度学习网络模型,重建高质量三维人脸形状、纹理及语义属性,是目前主流的重建框架。Deep3D (Deng 等, 2019b) 提出图像级和感知级 (学习的内在特征一致) 的新混合损失,融合多图互补信息进行形状聚合,一经提出得到了广泛引用,目前 REALY 数据库精度排名第 2。周昆团队提出 3DFaceGCN (Lin 等, 2020), 首个基于图卷积神经网络 (graph convolutional networks, GCN) 对 3DMM 模型纹理进行优化,包括基于 GCN 的解码、细化和组合 3 个网络,重建高保真纹理的三维人脸。

NextFace 是 Abdallah Dib 等人 (NextFace, 2024) 提出的单目重建项目,输出三维人脸几何、反照率、阴影、法线和照明。DeepNextFace (Dib 等, 2021) 结合编码器与可微分光线追踪,编码器回归基础人脸形状、表情等,两个解码器估计漫反射和镜面反射增量,捕获个性化反射细节。S2F2 (Dib 等, 2023) 增加了精阶段分支,在 UV 空间预测几何细节的法线和漫反射增量,更快速准确地重建三维人脸。

高精度表情细节重建和动画生成,是三维人脸研究的核心任务。基于 FLAME 表达模型, Timo Bolkart、Michael Black 等提出 RingNet (Sanyal 等, 2019)、DECA (Feng 等, 2021)、MICA (Zielonka 等, 2022) 和

表3 基于深度卷积神经网络的单目重建方法

Table 3 Monocular reconstruction algorithms based on deep convolutional neural network					
算法	模型	详细分析	使用数据库	输出	表达模型
VRN (Jackson 等, 2017)	深度卷积神经网络	首次提出体素表示 (volumetric representation) 三维人脸, 不需密集配准, 通过基于沙漏网络 (Newell 等, 2016) 的体回归网络, 端到端地直接学习从任意姿态和表情的二维人脸到三维人脸形状的体表征映射。	300W-LP (Zhu 等, 2016)	形状	
MoFA (Tewari 等, 2017)	自编码器	基于模型的深度卷积人脸自编码器算法, 编码器使用 AlexNet (Krizhevsky 等, 2017) 或 VGG-Face (Parkhi 等, 2015), 从输入图像学习形状、表情、纹理和照明等语义特征参数, 通过基于图像成像的可微分解码器生成图像, 并引入密集光度损失进行端到端网络训练。	CelebA (Liu 等, 2015) FaceWarehouse LFW (Huang 等, 2007) 300VW (Chrysos 等, 2015)	形状、姿态、表情 反射、照明	3DMM
SSML (Tewari 等, 2018)	自编码器	多层级人脸模型学习算法, 基于 MoFA 网络, 但都在多层级上学习, 更好地表征人脸的多样性, 结合 3DMM 的正则化和学习的校正空间增强模型泛化性, 快速重建三维人脸。	FaceWarehouse CelebA LFW 300 VW	形状、姿态、表情 反射	3DMM
PRNet (Feng 等, 2018a)	深度卷积神经网络	首次提出 UV 位置图表示三维人脸, UV 位置图回归网络使用 ResNet (He 等, 2016), 并基于语义分割生成权重掩膜图, 集成到损失函数, 实现端到端三维人脸重建以及密集对齐。	300W-LP	形状、姿态、纹理 关键点	
RingNet (Sanyal 等, 2019)	深度卷积神经网络	多编码器—解码器环形结构重建, 编码器基于 ResNet50, 解码器使用 FLAME, 输入同一个体的多张人脸和另一个体的一张人脸, 学习 FLAME 模型参数、解耦形状、表情和姿态, 设计新的人脸形状一致性损失, 构建的 NoW 数据库广泛应用于定性定量评价。	VGGFace2 (Cao 等, 2018); 发布 NoW 数据库, iPhone X 采集 100 人 2 054 幅图像及 3dMD 采集三维头部扫描	形状、姿态、表情	FLAME
Deep3D (Deng 等, 2019b)	深度卷积神经网络	提出单图和多图重建两个框架。引入图像级和感知级的新混合损失函数, 感知级损失是渲染图像和输入人脸图像学习的内在特征一致, 使用 FaceNet (Schroff 等, 2015) 提取特征。基于 ResNet50, 结合新混合损失进行单图重建; 通过轻量级置信度子网学习单图重建的置信度测量, 融合多图互补信息进行形状聚合。	CelebA 300W-LP LFW Multi-PIE (Gross 等, 2010)	形状、纹理	BFM09 FaceWare-house
DF ² Net (Zeng 等, 2019)	深度卷积神经网络	通过回归深度图由粗到精重建三维人脸, 包括深度图生成网络及两个优化网络, 采用不同的数据集进行网络训练, 并在优化网络引入多分辨率超列块结构, 学习中间层和不同域的多分辨率特征。深度图生成网络: 基于 U-Net (Ronneberger 等, 2015), 使用合成数据和重建数据训练, 优化网络: 基于 SFS 细化方法, 分别使用二维重建数据和真实数据训练, 输出高质量深度图。	三维合成数据, 基于 3DMM 模型; 二维重建数据, 基于 PRNet (Feng 等, 2018a)、Extreme3D (Tran 等, 2018); CACD (Chen 等, 2014)	形状、纹理	
MMFace (Yi 等, 2019)	深度卷积神经网络	多度量回归网络, 重建三维人脸几何和三维关键点; 体表征子网: 基于两个沙漏网络, 估计人脸几何的体表征; 参数子网: 基于体表征回归 3DMM 参数。	LS3D-W (Bulat 和 Tzimiropoulos, 2017)	形状、3D 关键点	BFM09 FaceWare-house
ReDA (Zhu 等, 2020a)	深度卷积神经网络	针对 3DMM 提出增强的可微属性, 将颜色、深度和解析掩膜等多种可微属性集成到光栅化, 通过一组具有多尺度核的卷积操作增强渲染可微分性, 增加在顶点空间建模的校正形状残差, 并提出刚性自由变形损失。	BU-3DFE Florence	形状、纹理	3DMM
MGCNet (Shang 等, 2020)	深度卷积神经网络	基于 ResNet50 提出遮挡感知的视角合成与新的多视角几何一致性损失, 针对大姿态变化引起的自遮挡构建共视图。网络训练需要多视角输入, 测试只需单张人脸, 对每个视图重建姿态一致的 3DMM 模型, 利用可微分渲染器生成渲染图像、深度图和共视图。	CelebA 300W-LP Multi-PIE LS3D-W	形状、姿态、纹理 照明	BFM09 FaceWare-house
3DDFA_V2 (Guo 等, 2020)	深度卷积神经网络	基于大姿态对齐模型 3DDFA, 引入轻量化网络, 如 MobileNets (Howard 等, 2017), 回归 3DMM 参数, 设计元联合优化策略及关键点回归正则化, 有效提升拟合速度和精度; 同时提出三维辅助短视频合成方法, 模拟平面内外旋转, 由一张静态人脸生成短视频训练网络, 实现更快速、精确、稳定的三维人脸对齐和重建。	300W-LP AFLW2000-3D (Zhu 等, 2016)	形状、3D 关键点	BFM09 FaceWare-house
3DFaceGCN (Lin 等, 2020)	图卷积网络 GCN	首个基于图卷积神经网络对 3DMM 模型纹理优化。FaceNet 提取人脸特征, Deep3D 回归 3DMM 模型参数和姿态、光照, 生成人脸基础形状和纹理; 基于 Ranjan 等 (2018) 的纹理优化生成人脸网格顶点颜色, 由解码、细化和组合 3 个网络组成, 解码特征并细化 3DMM 基础纹理, 组合网络生成精细纹理, 基于可微分渲染器并设计新的损失, 重建高保真纹理的三维人脸。	CelebA CelebA-HQ (Karras 等, 2018)	形状、纹理	BFM09
Uncertainty (Lee 等, 2020b)	图卷积网络 GCN	不确定性特征感知的重建。编码器: 估计人脸不同属性特征的高斯分布表示不确定性, 提升对遮挡、姿态的鲁棒性; 解码器: 包括基于 GCN (Kipf and Welling, 2017) 的形状解码器和基于 BigGANs (Brock 等, 2019) 的纹理解码器, 设计新的不确定性感知、多视图身份损失等。	COMA 300W-LP CelebA CASIA-Webface (Yi 等, 2014)	形状、纹理	

表 3 基于深度卷积神经网络的单目重建方法

Table 3 Monocular reconstruction algorithms based on deep convolutional neural network

(续表)

算法	模型	详细分析	使用数据库	输出	表达模型
Unsup3D (Wu 等, 2020)	自编码器	利用目标的对称性作为几何约束线索,将输入图像分解为深度、反照率、照明、视点及对称概率置信图,通过翻转内部表示实现对称,显式建模照明的非对称性,扩展模型来推理目标潜在的非对称性,预测每个像素的置信度分数来表征像素具有对称点的概率,使用预测的深度和光照估计阴影,重建解耦三维人脸。	BFM09 CelebA 3DFAW (Pillai 等, 2019)	形状 姿势 纹理、 反照率 照明	-
2DASL (Tu 等, 2021)	深度卷积神经网络	利用野外人脸图像增强三维建模的自监督学习,两个训练通道:有 3DMM 模型标注的监督训练;野外人脸图像及其关键点的自监督训练。基于 ResNet50 回归 3DMM 参数,将二维和三维关键点预测看做自我映射过程,引入新的一致性损失及潜在表征与 3DMM 系数的自我评价学习,实现三维人脸重建和对齐。	300W-LP UMDFaces (Bansal 等, 2017)	形状 关键点	BFM09 Face- Ware- house
SADNet (Ruan 等, 2021)	深度卷积神经网络	针对遮挡和大姿态的自对齐双回归网络,包括编码器、注意分支和解码器,人脸几何分解为 UV 空间的姿态、平均脸和形状变形,注意分支子网生成注意掩膜以增强可见特征抑制遮挡,由遮挡感知的自对齐融合,使用可见稀疏关键点对齐,实现三维人脸对齐与重建。	AFLW2000-3D	形状 姿态 3D 关键点	BFM09
SynergyNet (Wu 等, 2021)	深度卷积神经网络	首个 3DMM 模型与三维关键点的协同网络,实现朝向估计、对齐和建模。正向表征:基于 3DDFA_V2 回归 3DMM 模型参数,提取三维关键点,通过多属性特征整合细化关键点;反向过程:从稀疏三维关键点回归嵌入的 3DMM 参数,利用自约束损失增强重建精度。	300W-LP	形状 纹理 3D 关键点	BFM09
DeepNextFace (Dib 等, 2021)	深度卷积神经网络	结合编码器与可微分光线追踪的重建,包括编码器、两个解码器及可微分光线追踪器,使用 BFM17 及 AlbedoMM,编码器基于 ResNet-152 回归基础人脸属性,两个解码器估计漫反射和镜面反射增量,捕获个性化反射率,光线追踪建模自阴影。	CelebA	朝向 形状 姿态 纹理 反射	BFM17
DECA (Feng 等, 2021)	深度卷积神经网络	首个学习表情细节和动画的框架,基于 FLAME,从身份和通用表情参数组成的低维空间表示生成 UV 位移图,编码器基于 ResNet50 回归 FLAME 参数及反照率、光照和相机参数,通过 UV 位移图增强基础形状,引入同一个体多张人脸的细节一致性损失,分离个体身份与动态表情,操控表情参数生成几何细节,实现遮挡、姿态、照明鲁棒的重建、表情迁移和动画生成。	VGGFace2 VoxCeleb2 (Chung 等, 2018); BUPT (Wang 等, 2019)	形状 姿态 表情	FLAME
EMOCA (Daněček 等, 2022)	深度卷积神经网络	情感驱动的表情捕捉和动画框架,基于 DECA 引入深度感知情感一致性损失,确保重建表情与情绪匹配,使用情感数据训练表情部分,设计情感一致性损失、眼睛和嘴部闭合损失、表情正则化损失等,实现高保真情绪内容感知的三维人脸表情重建。	AffectNet (Mollahosseini 等, 2019)	形状 姿态 表情	FLAME
MICA (Zielonka 等, 2022)	深度卷积神经网络	新的度量人脸重建,对中小型数据库标注,统一到 FLAME 模型拓扑,生成三维数据库 (2 300 个个体),基于预训练 Sub-center ArcFace 特征微调,重建对表情、姿态不变的度量人脸形状,用于人脸表情动态跟踪。	融合 LYHM、FaceWarehouse、Stirling/ESRC (Feng 等, 2018b) 等多个数据库	形状	FLAME
3DFR_DL (Wood 等, 2022)	深度卷积神经网络	首次仅使用密集的关键点进行 3DMM 拟合,基于精确标注的大规模合成数据训练网络,使用简单网络预测密集关键点,对每个点回归包含位置和不确定性参数的高斯分布,设计新拟合损失。	自建合成数据库 (100 k)	形状 3D 关键点	3DMM
3D3M (Li 等, 2023b)	深度卷积神经网络	可调制三维人脸形变模型,通过随机重组 3DMM 人脸属性编码进行调制。输入一组人脸图像,基于 ResNet50 对 3DMM 形状、纹理和光照编码;随机组合人脸属性编码,再重新编码和渲染,生成新的人脸,利用网格顶点形状对应和重建像素一致性约束自监督学习。	CelebA 300W-LP Multi-PIE LS3D-W	形状 纹理 3D 关键点	BFM09 Face- Ware- house
JMLR (Guo 等, 2023a)	深度卷积神经网络	联合回归三维人脸网格和二维人脸关键点,基于 RetinaFace 人脸检测及归一化,设计搜索 ResNet 网络结构进行联合回归,估计人脸姿态,增强对表情和姿态的鲁棒性。	竞赛数据库:250 个人,33 种表情, iPhone 11 采集, 包括三维人脸网格、姿态及 68 个关键点	形状 关键点	-
S2F2 (Dib 等, 2023)	自编码器	结合自编码器与可微分光线追踪的由粗到精重建,基于 DeepNextFace 增加精阶段分支,使用两个 U-Net 在 UV 空间预测法线贴图 and 漫反射增量,结合可微分光线追踪与基于顶点的渲染器,重建形状、反射率和照明。	CelebA	形状 纹理 法线 反射	BFM17
FOCUS (Li 等, 2023a)	自编码器	联合人脸自编码器与异常值分割的网络。人脸自编码器由 ResNet50、BFM17 及可微分渲染器构成,异常值分割 U-Net,分割网络指导人脸编码器只拟合正常值,增强的人脸模型又能够优化分割网络。	CelebA-HQ	形状 分割图	BFM17
HRN (Lei 等, 2023)	深度卷积神经网络	层次化表征网络,人脸几何细节解耦为低频、中频和高频,引入三级表征:人脸级基形状、顶点级变形图和像素级位移图。基于 Deep3D 预测人脸网格和反照率,生成 UV 空间位置图和纹理图;细节建模网络引入先验引导分层细节的建模,并扩展实现稀疏多视重建。	FaceScape Stirling/ESRC 自建 FaceHD-100 数据库, 100 个人, 20 种表情	形状 纹理	BFM09

注:“-”表示未使用表达模型。

EMOCA(Daněček等,2022)。RingNet提出多编码器—解码器的环形结构,发布基准数据库 NoW,广泛应用于人脸重建的定性和定量评价。DECA提出首个学习表情细节和动画的框架,从身份参数和通用表情参数组成的低维空间生成UV位移图,分离个体身份与表情细节,实现表情迁移及动画。EMOCA引入深度感知情感一致性损失,增加情感识别模型分支进行情感回归,实现情绪内容感知的表情重建。同时,为了建立度量空间上的正确重建和评价,MICA对公开数据库进行标注和统一拓扑,用于有监督学习,重建对表情、姿态和照明不变的度量三维人脸形状,NoW数据库获得SOTA精度。

三维人脸重建和人脸对齐是密切相关的研究任务。二维人脸关键点是单目重建的重要输入信息和优化损失项,有效表达人脸特征,如2DASL(Tu等,2021)、MGCNet(Shang等,2020)、3DDFA_V2(Guo等,2020)和3D3M(Li等,2023b)等。SynergyNet(Wu等,2021)提出参数模型与三维关键点协同作用的网络,包括正向和反向过程。正向过程回归3DMM参数并提取、优化关键点;反向过程从三维关键点回归3DMM参数,实现人脸朝向估计、对齐和重建。3DFR_DL(Wood等,2022)首次仅使用非常密集的关键点进行3DMM拟合,构建精确标注的大规模合成数据训练网络,对关键点回归包含位置和不确定性参数的高斯分布,NoW上精度排名第3。

中国科学院自动化研究所提出3DDFA(Zhu等,2016)、3DDFA_V2(Guo等,2020)和Beyond3DMM(Zhu等,2020b)等三维人脸对齐和重建算法。3DDFA是针对大姿态人脸对齐的自遮挡问题的模型,3DDFA_V2引入轻量化网络模型回归3DMM参数,设计新的动态组合改进损失的优化策略及关键点回归正则化,提出三维辅助短视频合成方法,模拟运动模糊和平面内外旋转,生成短视频数据训练网络,自提出后得到了广泛关注和引用。

遮挡一直是三维人脸重建研究的难题,包括姿态、照明引起的自遮挡和口罩、眼镜等外物遮挡。MGCNet(Shang等,2020)提出遮挡感知的视角合成与多视角几何一致性损失,针对姿态变化自遮挡构建共视图。SADRNet(Ruan等,2021)提出针对遮挡和大姿态的自对齐双回归网络,解耦人脸姿态和形状估计,引入注意力机制处理遮挡,实现三维人脸密集对齐与重建。FOCUS(Li等,2023a)针对遮挡、化

妆等异常区域联合学习模型拟合与异常值分割,分割网络指导人脸编码器只拟合正常值,增强的人脸模型又反过来优化分割网络,提升遮挡鲁棒性。

InsightFace是郭佳、邓健康等人(InsightFace,2024)提出的二维/三维人脸分析与识别项目,实现人脸检测、对齐和识别,包括ArcFace、Sub-center Arcface、RetinaFace(Deng等,2019a,2020a,2020b)、JMLR(Guo等,2023a)和PBIDR(Ren等,2023)。JMLR提出联合回归三维人脸网格和关键点,估计6个自由度人脸姿态,针对三维重建中透视投影问题提出有效的解决方案,对大的表情、姿态变化鲁棒性强。

另外,无需人脸统计模型先验,Uncertainty(Lee等,2020b)提出不确定性特征感知的重建,通过估计人脸不同属性特征的高斯分布表示不确定性,提升对遮挡、姿态的鲁棒性。Unsup3D(Wu等,2020)利用目标对称性作为几何约束线索来建模,预测新的对称概率置信图,重建解耦的三维人脸。

2)基于生成对抗网络的方法。基于生成对抗网络(generative adversarial network,GAN)的单目重建是一个重要分支,近年来GAN应用于高质量人脸图像的生成,为解决深度模型的数据难题提供了新思路,如表4所示。

Stefanos Zafeiriou团队提出GANFIT、TBGAN(Gecer等,2019,2020,2021)、AvatarMe(Lattas等,2020,2022b)和FitMe(Lattas等,2023)。GANFIT是首个基于GAN的重建框架,结合GAN纹理生成、3DMM模型及可微分渲染器,利用预训练身份特征设计新的损失。AvatarMe通过超分辨率网络生成高频细节UV纹理图,使用去光照网络结合基础形状预测漫反射/镜面反射。TBGAN提出主干—分支GAN架构学习UV空间的形状、纹理和法线,并构建大规模三维表情数据集,实现表情驱动的重建。FitMe融合风格GAN和3DMM拟合的人脸反射模型,设计多模式BRDF-GAN网络,构建大规模人脸反射数据库,生成可渲染人脸几何和反射属性。

SEMM(Ling等,2022)是结构感知可编辑三维人脸形变模型,其联合UV位移图及无符号距离场图表征皱纹线条,回归表情和年龄变化,实现人脸细节操控和动画。AlbedoGAN(Rai等,2023)提出生成式三维可控人脸模型,利用阴影图解耦照明生成高质量反照率,结合FLAME模型与顶点位移图得到精

表 4 基于生成对抗网络的单目重建方法
Table 4 Monocular reconstruction algorithms based on generative adversarial network

算法	详细分析	数据库	输出	表达模型
GANFIT (Gecer等,2019,2021)	首个基于 GAN 的重建算法,结合 PGGAN (Karras 等,2018)与 3DMM 及可微分渲染器,使用大量高精度 UV 纹理图训练生成网络,作为人脸纹理的统计先验,再与 3DMM 模型参数共同渲染,设计新的损失,利用预训练深度身份特征优化参数,重建三维人脸形状和纹理。	LSFM	形状 纹理	LSFM
AvatarMe (Lattas等,2020,2022b)	基于 GANFIT,通过超分辨率网络 RCAN (Zhang 等,2018) 对人脸 UV 纹理图上采样,生成高频细节,再基于 pix2pixHD 网络获得高分辨漫反射属性,然后结合 3DMM 拟合人脸形状预测镜面反射率、漫反射法线和镜面反射法线,重建高分辨率三维人脸并进行逼真渲染。	使用 Light Stage 采集系统自建数据库 RealFaceDB,200 人,7 种表情。	形状 纹理 漫反射 镜面反射	LSFM
TBGAN (Gecer等,2020)	基于主干—分支 GAN 学习 UV 空间的形状、纹理和法线,基于 PGGAN,分支网络确保不同模式的独特性,主干网络学习局部对应相关性,构建表情数据集并通过表情识别网络自动标注,将表情编码与其他模式的 UV 映射耦合,联合生成高质量形状、纹理、法线并生成不同表情。	MeIn3D (Booth 等,2016); 自建大规模三维人脸表情数据库, 3dMD 采集 5 000 人。	形状 纹理 法线 表情	LSFM
L2R (Zhang 等,2022c)	基于野外退化人脸图像,挖掘生成人脸先验,包括反照率恢复网络和几何细化网络,通过代理网络 Unsup3D 提取人脸风格和空间先验编码,指导预训练 StyleGANv2 (Karras 等,2020) 生成器恢复高质量人脸纹理,学习建模 UV 位移图,设计三维一致性对抗损失及相互学习策略优化。	CelebA CelebAMask-HQ (Lee 等,2020a); FFHQ (Karras 等,2019)	形状 纹理 法线	-
SEMM (Ling 等,2022)	结构感知的可编辑三维人脸形变模型,是对 FaceScape 模型的增强,从 UV 位移图提取无符号距离场图,联合表征人脸细节结构(皱纹线条),基于 StyleGANv2 对细节结构建模,基于一致性约束构建隐式对应,回归表情和年龄变化,实现高质量重建、细节编辑和动画。	FaceScape	形状	FaceScape
StyleGAR (Piao 等,2021)	新的生成对抗渲染器(generative adversarial renderer,GAR),代替可微分渲染器。基于 StyleGANv2,渲染器的输入包括几何因素(法线贴图)和非几何因素(潜在编码、噪声),利用法线注入模块并设计新的损失进行法线估计。基于训练好的 GAR,设计对称的反演网络,预测可靠潜在编码初始值,结合 3DMM 模型拟合,重建三维人脸。	FFHQ CelebA Florence	形状 纹理	3DMM
GAN2X (Pan 等,2022)	新的非朗伯逆渲染算法,基于 StyleGANv2 设计镜面反射感知的隐式神经表示,利用伪配对生成图像预测形状、反照率和镜面反射属性,基于可微分体渲染和 Phong 阴影渲染图像,采用基于阴影的细化技术优化细节,重建三维形状、非朗伯材质表现及新视角和重光照渲染。	CelebA CelebA-HQ (Karras 等,2018)	形状 反照率 法线 反射	-
FitMe (Lattas 等,2023)	融合风格 GAN 和 3DMM 拟合的人脸反射模型及可微分渲染优化,从单个/多个野外人脸,生成可渲染详细几何和反射属性。基于 StyleGANv2 设计多模式 BRDF-GAN 网络,生成漫反射率、镜面反射率和法线,并使用 BRDF-GAN 纹理和 GAN 反演优化取代 3DMM 的纹理和拟合,包括基于渲染的反演和微调,重建三维人脸数字化身。	自建数据库,在 MimicMe 上微调 AvatarMe++模型,生成包含 5 000 个人的 人脸反射数据库	形状 漫反射 镜面反射 法线	UHM (Ploumpis 等,2021)
AlbedoGAN (Rai 等,2023)	生成式三维可控人脸模型,基于 StyleGANv2,利用阴影图解耦照明,生成反照率图像,提出位移图生成器,结合 ArcFace、FLAME 与位移图得到精细人脸形状,可微分渲染器对反照率网络微调,生成高分辨率反照率,重建几何细节、新视角渲染及基于文本的三维表情控制。	FFHQ	形状 纹理 表情 反照率	FLAME

注:“-”表示未使用表达模型。

细网格,实现细节重建、新视角渲染及基于文本的表情控制。

另外,GAN 还应用于解决基于退化人脸的三维人脸重建(Zhang 等,2022c)、设计新的生成对抗渲染器融合到基于优化的重建(Piao 等,2021)、非朗伯反射逆渲染模型(Pan 等,2022)等。

3)基于隐式神经表示的方法。随着神经辐射场 NeRF 的提出,隐式神经表示为新视角合成、场景建模带来了新突破,再结合 GAN,将隐式表示应用于高质量三维人脸重建和渲染收获了新进展,如表 5 所示。

PBIDR(Ren 等,2023)提出遮挡鲁棒的纹理引导重建,结合 GANFIT 人脸形状拟合、风格 GAN 纹理补全及隐式表示 SDF,通过隐式可微渲染优化几何并解耦表观属性。3DMM-RF(Galanakis 等,2023)提出融合 GAN 与 NeRF 的隐式三维人脸形变模型,创建了大规模标记生成人脸数据集,联合基于深度值

的采样策略,实现解耦重建和渲染。

SSIF(Lin 等,2023)提出隐式三维人脸形变模型,融合隐式表示与显式 UV 纹理,人脸模型参数转化为身份和表情潜在编码,通过隐式表示几何分支和 UV 纹理参数化分支,实现大姿态新视角生成、人脸动画和纹理编辑。RAFaRe(Guo 等,2023b)提出非参数方法,利用大量野外人脸和有限的高精度三维数据,创建大规模严格对齐的伪 2D&3D 数据库,提出分层隐式表示网络,学习对姿态、表情和照明鲁棒泛化的人脸模型。

2.2.2 多目三维人脸重建

基于多目图像的三维人脸重建,能够有效利用多视角一致性,如表 6 所示。

融合几何和学习的方法,DFNRMVS(Bai 等,2020)提出首个多层级网络优化非刚性多目重建,包括特征提取、自适应人脸模型生成和非刚性多视角立体三级顺序优化。ToFu(Li 等,2021)提出基于体

表 5 基于隐式神经表示的单目重建方法
Table 5 Monocular reconstruction algorithms based on implicit neural representation

算法	模型	详细分析	数据库	输出
PBIDR (Ren 等,2023)	隐式神经表示 SDF 生成对抗网络 GAN	遮挡鲁棒的纹理引导重建框架,结合 GAN 纹理补全与隐式表示 SDF,重建精细几何并解耦法线、反照率等。高质量 UV 纹理生成:基于 OSTeC(Gecer 等,2021),使用 GANFIT 和 StyleGANv2,处理遮挡进行纹理补全,再结合初拟合的人脸形状构建准确的多视图集合;纹理引导的几何增强;构建隐式表示,设计新的损失,通过基于物理的隐式可微渲染优化几何细节,解耦漫反射法线和镜面法线,预测顶点的位移。	NoW 3DFAW Florence	形状 纹理 法线 反照率 阴影
3DMM-RF (Galanakis 等,2023)	隐式神经表示 NeRF 生成对抗网络 GAN	融合 GAN 与 NeRF 的隐式三维人脸形变模型。首先联合 LSFM 人脸模型、TBGAN 纹理生成及 Avatarme++ 反射材质网络,自建大规模标记生成数据库用于网络训练;基于 StyleGANv2,由生成器、体渲染和鉴别器组成,并提出基于深度值的采样策略提升渲染效率,训练好的模型能够进行三维人脸解耦重建和任意光照渲染。	自建大规模(10 k)逼真渲染生成人脸数据库,包括身份、姿态、表情和照明属性变化	形状 纹理 表情 姿态 照明
SSIF (Lin 等,2023)	隐式神经表示 SDF	隐式三维人脸形变模型,融合隐式表示的高质量拓扑几何与显式 UV 参数化纹理映射的可编辑性,将人脸模型解耦为身份(几何、颜色)和表情潜在编码,框架包括几何分支 SDF MLP 和 UV 纹理参数化分支,纹理分支网络包括参数化映射 MLP、正则化逆过程 MLP 和颜色网络,设计新的损失,学习连续一致的纹理图,重建高质量三维数字化身。	Triplegangers 高精度数据库,515 个人,20 种表情,每个样本包括三维头部网格、UV 漫反射纹理和漫反射照明图	形状 纹理 表情 姿态
RAFaRe (Guo 等,2023b)	隐式神经表示 SDF	非参数方法从伪 2D&3D 人脸数据库重建精确三维人脸。构建数据库:基于 FaceScape、LYHM 及 pix2pixHD 网络,渲染详细三维人脸模型再将原始图像人脸与渲染人脸交换,创建大规模(100 k)严格对齐的伪 2D&3D 人脸数据库,涵盖不同姿态、表观和照明。提出分层隐式表示网络,学习像素对齐的 SDF 表示人脸几何,恢复不同尺度的形状细节。	FaceScape LYHM、FFHQ CelebA IMDB-WIKI (Rothe 等,2015)	形状 法线 姿态

表 6 多目三维人脸重建方法
Table 6 3D face reconstruction algorithms from multi-view images

算法	模型	详细分析	数据库	输出	表达模型
MVFNet (Wu 等, 2019)	深度卷积 神经网络	输入表情形状不变的 3 个视图, 基于每个视角特征回归 3DMM 姿态参数, 拼接所有特征回归 3DMM 身份和表情参数, 可微分渲染器预测纹理, 得到不同视角三维人脸; 利用多视角相互投影建立约束, 基于非遮挡部分在纹理、光流和关键点特征上的不变性进行优化。	300W-LP Multi-PIE	形状	BFM09 Face- Ware- house
DFNRMVS (Bai 等, 2020)	深度卷积 神经网络	首个多层级网络优化非刚性多目重建。输入不同表情的多视角人脸, 包括特征提取 (Lin 等, 2017)、自适应人脸模型生成、非刚性 MVS 优化, 人脸模型包括共用 BFM 基和 UV 空间基 (UV 特征图和位置图)。	Stirling/ESRC	形状 姿态 表情	BFM09 Face- Ware- house
FacePSNet (Wang 等, 2020)	深度卷积 神经网络	结合光度立体和 3DMM 模型的轻量化网络, 使用 ResNet18、PS-FCN (Chen 等, 2018), 基于稀疏输入由粗到精重建。首先构建数据库, 包括真实和合成人脸数据; 预测 3DMM 形状和姿态参数, 再进行法线估计, 然后利用快速优化从法线恢复精细几何表面。	采集真实人脸数据 (84 人、29 种表情), 并基于 Light Stage (Ma 等, 2007) 数据库 (23 人、15 种表情) 构建合成数据集。	形状 法线	BFM09 Face- Ware- house
ToFu (Li 等, 2021)	深度卷积 神经网络	基于体素采样和分层网格表示的一致拓扑三维人脸重建。特征提取网络使用预训练 U-Net, 嵌入人脸拓扑结构到体素网格, 体素特征采样有效聚合不同尺度的视图特征, 网络包括全局网格生成、局部网格细化和表观属性生成, 快速生成高分辨率表情微观细节及操控动画。	利用 Light Stage (Debevec 等, 2000) 自建数据库, 64 人, 26 种表情, 包括多视角图像、三维扫描和皮肤属性真实值。	形状 姿态 表情 反照率 反射	-
H3DNet (Ramon 等, 2021)	隐式神经 表示 SDF	结合三维头部隐式表示先验与可微分渲染器, 基于 3 幅输入图像用几分钟重建头部 (包括头发和肩膀)。学习 SDF 分布, 将预训练的先验集成到可微分渲染器, 指导优化两个基于坐标的神经网络, 分别建模几何和表观。	利用结构光扫描仪自建数据集 H3DS, 10 人, 包含高分辨率头部扫描、图像和相机姿态等。	形状 纹理	-
LAP (Zhang 等, 2021c)	深度卷积 神经网络	基于 Unsup3D 学习聚合和个性化信息。聚合网络: 自适应学习聚合身份一致的几何和纹理, 设计松弛多图一致性损失的课程学习方法训练网络。属性细化网络: 过滤连接模块进行属性注入和特征融合, 学习人脸细节。	CelebA CASIA-WebFace 使用 Interfacegan (Shen 等, 2020) 生成数据	形状 纹理 法线	-
PhyDIR (Zhang 等, 2022b)	深度卷积 神经网络	基于物理引导解耦的隐式神经渲染, 包括隐式纹理建模网络和神经表观渲染器, 三维物理引导模块连接这两部分 (基于 Unsup3D 和 LAP), 然后作为可微分渲染器进行形状建模。显式物理引导包括多图像阴影、光栅化和纹理映射, 对形状、光照和视点准确解耦控制。	CelebA CASIA-WebFace CelebAMask-HQ	形状 纹理 法线	-
HeadNeRF (Hong 等, 2022)	神经辐射 场 NeRF	基于 NeRF 的参数化头部模型。基于 Nonlinear3DMM 及 StyleNeRF (Gu 等, 2021) 网络, 结合 NeRF 体渲染与新的神经渲染, 生成低分辨率特征图再上采样, 实现快速高保真头部渲染, 并直接操控姿态、身份和表情等属性。	FaceSEIP FaceScape FFHQ	形状 纹理 表情 姿态	Nonlinear 3DMM
MoFaNeRF (Zhuang 等, 2022)	神经辐射 场 NeRF 生成对抗 网络 GAN	可变形人脸 NeRF 模型, 由形变人脸 NeRF 网络和基于 Pix2PixHD 的细化网络组成。引入新的纹理编码器和特定身份调制模块, 将不同个体的表情用形状和表情参数的调制关系表达, 实现视角渲染及形状、表情解耦表达。	FaceScape	形状 姿态 纹理 表情	FaceS- cape
MVFR (Xiao 等, 2022)	隐式神经 表示 生成对抗 网络 GAN	学习隐式函数, 来正则化匹配代价和多视图融合, 包括基网格拟合、隐式函数学习和精细微观预测 (基于 Pix2PixHD)。基于多视图关键点和 FaceScape 模型, 得到拓扑一致的网格; 学习隐式函数, 转换到 UVD 空间 (UV + Displacement), 减少隐式函数学习的解空间。	FaceScape	形状 法线	FaceS- cape
NPF (Zhang 等, 2023b)	生成对抗 网络 GAN	神经人脸原型场学习框架, 基于三平面表征的 3DGAN 网络 EG3D (Chan 等, 2022), 挖掘多图身份一致性先验, 解耦身份、表情和细节。身份感知聚合: 基于 LAP, 利用不确定性建模方法学习人脸原型; 变形建模: 基于 Deep3D (Deng 等, 2019b) 表情姿态参数, 利用表情一致性损失约束变形, 解耦表情和身份。	CelebA CASIA-WebFace CelebAMask-HQ	形状 纹理 法线	-
NeuFace (Zheng 等, 2023)	隐式神经 表示 SDF	三维人脸渲染模型, 整合神经 BRDFs 表示到 PBR 渲染框架, 建模皮肤光学特性和几何形状。引入新的低秩先验减小解空间, 学习人脸反射模型 BRDFs; 使用基于 SDF 的模型 ImFace 作为形状先验, 在逆渲染中同步优化人脸表观和几何场, 协同捕捉人脸几何和表观属性。	FaceScape	形状 纹理 法线 反射	ImFace
GANHead (Wu 等, 2023c)	隐式神经 表示 占用场	生成式神经头部模型, 挖掘显式表情参数与隐式神经占用场表示, 实现高保真头部化身及动画生成, 包括生成、变形和渲染模块。生成模块由 3 个 MLP 网络建模形状、法线、纹理和变形模块引入线性混合蒙皮技术 LBS 定义变形场, 使用学习的连续姿态、表情基和 LBS 权值实现网格变形, 保证几何一致性和姿态/表情泛化性。	FaceVerse	形状 纹理 表情 姿态 法线	FLAME

注:“-”表示未使用表达模型。

素采样和分层网格表示的一致拓扑三维人脸重建,渐进式网格生成网络嵌入人脸拓扑结构,体素特征采样有效聚合不同尺度的视图特征。

可解耦重建算法 LAP(Zhang 等, 2021c)、PhyDIR(Zhang 等, 2022b)和 NPF(Zhang 等, 2023b)被研究者提出。LAP 先学习聚合信息再学习个性信息,解耦身份一致性结构和场景个性化属性,包括深度/反照率聚合网络和属性细化网络。PhyDIR 提出基于物理引导解耦的隐式神经渲染,包括隐式纹理建模网络和神经表观渲染器,以及显式三维物理引导模块。NPF 提出学习人脸原型场,基于三平面表征的 3DGAN 网络,从开放人脸集合挖掘多视图一致性先验,解耦身份、表情和细节,利用表情一致性损失约束变形,重建个性化表情。

为了提升模型速度,轻量化网络 FacePSNet(Wang 等, 2020)和基于 NeRF 的参数化模型 HeadNeRF(Hong 等, 2022)开始涌现。FacePSNet 利用 3 个近距离点光源采集人脸,基于 Light Stage 数据集及数据增强方法合成人脸数据,包括 3DMM 参数预测、法线估计和快速优化从法线恢复人脸几何。HeadNeRF 结合 NeRF 体渲染与新的神经渲染,收集大规模数据训练网络,实现快速高保真(25 ms/帧)头部渲染及语义编辑,自提出后引起了广泛关注。

MVFR(Xiao 等, 2022)是整合隐式函数学习的重建算法,学习一个隐式函数来正则化匹配代价和多视图融合,实现几十秒重建精细三维人脸。NeuFace(Zheng 等, 2023)是三维人脸神经渲染模型,通过神经渲染学习精确的、有物理意义的三维表示,将神经 BRDFs 表示整合到 PBR 渲染,并引入新的低秩先验,建模复杂的皮肤光学特性和几何,实现逼真渲染、重光照并能够泛化到通用物体渲染。

2.2.3 基于视频的三维人脸重建

基于视频的三维人脸重建是一项具有挑战性的任务,如表 7 所示。可以发现,基于隐式神经表示尤其是神经辐射场 NeRF 的重建技术飞速发展,取得了 SOTA 精度和速度。

结合 NeRF 与 3DMM,NeRFace(Gafni 等, 2021)提出动态神经辐射场模型,引入每帧可学习的潜在编码,利用体渲染合成三维一致的新表情和姿态,实现重建、表情编辑和重演。结合隐式占用场与 FLAME,IMavatar(Zheng 等, 2022b)提出隐式形变头部化身模型,利用 3 个连续的隐式占用场进行几何、

变形和纹理建模,提升重建几何精度和对未见新表情新姿态的泛化能力。

为了克服 NeRF 的计算瓶颈,优化模型效率,INGP(Müller 等, 2022)通过建立多分辨率体素网格存储,并利用哈希编码进行体素网格压缩,大大提升了渲染速度。NeRFBlendShape(Gao 等, 2022)设计结合多层级体素场与表情系数的新表征,使用轻量级 MLP 进行体渲染,能够在 10~20 min 重建头部化身。INSTA(Zielonka 等, 2023)融合多分辨率哈希与动态神经辐射场,建立几何引导的变形神经辐射场,10 min 完成重建并以交互帧率驱动渲染。

AvatarMAV(Xu 等, 2023)用 5 min 重建真实感头部化身,用体素网格表达表情相关的动态变形,通过解耦表情相关运动与表观,建立运动体素网格,每个表情基表示为一个运动体素网格基,加权连接构成整个运动感知神经体素,再连接轻量级 MLP 建模表情运动场和纹理。

2.2.4 基于语音的三维人脸重建

基于语音驱动的三维人脸研究是创建“数字人”的重要支撑技术,主要解决两个关键问题:1)高质量数据库;2)跨模态映射模型。

由于标准数据库的缺乏,目前大多数模型都是在自主构建的有限数据集上进行训练和评价。VOCA(Cudeiro 等, 2019)提出 VOCASET 数据库,得到了广泛应用。MeshTalk(Richard 等, 2021)解耦音频相关(嘴型)和音频无关(上面部)信息,重建精确的嘴唇运动、眨眼和眉毛运动。上述都是有监督学习,CMP(Wu 等, 2022)通过有监督学习和无监督知识蒸馏两种方案来探讨跨模态人脸几何重建,构建数据库用于监督学习,无监督框架使用 GAN 网络合成二维人脸再估计 3DMM 参数,联合训练语音到图像和图像到三维的模块。

近年基于 Transformer(Vaswani 等, 2017)的语音驱动三维人脸动画发展迅速。FaceFormer(Fan 等, 2022)提出基于 Transformer 的自回归模型,对长时音频上下文和人脸运动的历史序列进行编码,预测一组三维人脸网格。为了减少跨模态映射的不确定性,CodeTalker(Xing 等, 2023)基于 FaceFormer 提出基于离散运动先验的时间自回归模型,将人脸运动空间建模为一个离散编码本,三维人脸动画表示为编码查询任务,能够重建协同的嘴唇运动和表情。为了捕捉语音中的情感,EmoTalk(Peng 等, 2023)构

表 7 基于视频的三维人脸重建方法
Table 7 3D face reconstruction algorithms based on videos

算法	模型	详细分析	数据库	输出	表达模型
FML (Tewari 等, 2019)	自编码器	基于互联网视频的多帧一致性人脸模型学习,学习帧间共享的身份参数(形状和表观)和帧间独立的姿态、照明和表情参数,通过可微分网格形变层和可微分渲染器构成的自编码器,引入多帧一致性损失,实现人脸模型学习和重建。	VoxCeleb2	形状 姿态 表情 反照率	3DMM Face- Ware- house
Complete3DMM (Mallikarjun 等, 2021)	自编码器	基于 FML,设计新的损失,融合表观和人脸分割损失,引入嘴唇分割一致性损失,学习包含嘴部和嘴唇动作的表情基,预测 3DMM 模型的身份几何、反照率和表情参数,提高泛化性。	VoxCeleb2 EmotionNet (Benitez-Quiroz 等, 2016)	形状 姿态 表情 反照率	3DMM Face- Ware- house
CEST (Wen 等, 2021)	深度卷积神经网络	条件估计自监督学习,显式挖掘三维人脸参数的统计相关性,按顺序和条件分别估计视角、形状和照明,基于视角和形状生成 UV 空间表征的反射率(U-Net),融合帧间反射对称性和一致性,使用可微分渲染器及随机优化策略。	Voxceleb (Nagrani 等, 2017) 300W-LP	形状 纹理 反射 照明	BFM09 3DDFA
NeRFace (Gafni 等, 2021)	隐式神经表示 NeRF	动态神经辐射场,首次将 NeRF 与 3DMM 结合,输入视频片段及其背景图像,使用 Face2Face (Thies 等, 2016) 重建动态表情和姿态参数,引入每帧可学习的潜在编码,利用体渲染合成三维一致的新表情和姿态,实现重建、表情编辑和重演。	自建数据集	形状 姿态 表情 纹理	3DMM
IMavatar (Zheng 等, 2022b)	隐式神经表示 占用场	隐式形变头部化身模型,通过 3 个连续的隐式占用场进行几何、变形、纹理建模,变形由表情基、姿态基和蒙皮权重组成,提出新的解析梯度公式实现端到端网络学习,重建精确几何并生成未见的新表情和姿态。	自建数据集(合成和真实视频)	形状 姿态 表情 纹理	FLAME
NHA (Grassal 等, 2022)	隐式神经表示 NeRF	完整头部化身模型(包括头发),基于 Face2Face、FLAME、SIREN MLP (Sitzmann 等, 2020),设计几何和纹理网络,进行三维网格细化和动态纹理合成(皱纹、反射),引入颜色相关和无关的优化项来解耦,重建大视角的几何和细节纹理。	自建数据集,包括合成视频(4 人, 200 帧),真实视频(4 人, 750 帧)	形状 姿态 表情 纹理	FLAME
NeRFBlend - Shape (Gao 等, 2022)	隐式神经表示 NeRF	定义在多层级体素场上的人脸模型重建。追踪每帧的表情和姿态参数,基于 INGP (Müller 等, 2022),提出结合多层级体素场和表情系数的新表征方法,每个基都是一个特定表情的多级体素场表示,任意表情表示为体素基与表情系数的加权组合;使用轻量级 MLP 进行体渲染,设计表情感知的密度网格更新策略,快速(10~20 min)高质量重建和渲染。	采集不同表情和姿态的视频数据; SSP-NeRF (Liu 等, 2022) NVPuppetry (Thies 等, 2020)	形状 纹理 表情 姿态	Face- Ware- house
INSTA (Zielonka 等, 2023)	隐式神经表示 NeRF	融合多分辨率哈希网格 INGP 与动态神经辐射场 NeRFace,使用 FLAME 人脸参数模型和 MICA 人脸跟踪,建立几何引导的可变形神经辐射场,更快速重建逼真的神经头部化身(10 min)并以交互式帧率渲染新表情新姿态。	自建数据集; Youtube 视频	形状 姿态 表情 纹理	FLAME
AvatarMAV (Xu 等, 2023)	隐式神经表示 NeRF	基于运动感知神经体素的快速三维头部化身重建(5 min),用体素网格表达表情相关的动态变形。解耦表情相关运动与人脸表观,建立运动体素网格(motion voxel grid, MVG),每个表情基表示为一个 MVG 基,以 3DMM 表情系数为权重加权连接构成整个 MVG,人脸表观使用单个表观体素网格表示,再连接轻量级 MLP 建模表情运动场和纹理。	9 个视频数据, 手机采集 4 个视频、HDTF (Zhang 等, 2021b) 4 个视频、NeRFBlend-Shape 的 1 个视频	形状 姿态 表情 纹理	BFM17
AVFace (Chatziagapi 和 Samaras, 2023)	隐式神经表示 SIREN Transformer	基于视频和语音的由粗到精的混合重建。粗阶段编码器基于 DECA 和 Transformer,使用 DeepSpeech (Hannun 等, 2014) 提取语音特征,预测每一帧人脸形状,通过 SIREN MLP 学习嘴唇形状的隐式神经表示;精阶段采用 UNet-ResNet 网络 (Abrevaya 等, 2020),由伪人脸法线真实值引导,重建几何细节,使用合成遮挡数据微调,提高网络对单一模态的鲁棒性。	Voxceleb2	形状	FLAME

建大规模伪三维情感谈话人脸数据集,解耦情感和内容,生成准确表达情感的人脸动画并操控风格。

2.3 分析总结

综上所述,受益于深度学习新模型、大模型的不
断创新,三维人脸重建技术取得了令人印象深刻的
突破性进展,无论精度、速度还是鲁棒性、泛化性,都
发生了飞跃式提升,三维人脸的研究与应用正迎来
新的时代,如表8所示。

2.3.1 单目三维人脸重建

1)数据源角度。(1)基于单幅图像的三维人脸重
建研究通常是在大规模二维人脸数据库上进行自监
督学习。一方面,大规模二维人脸数据库为深度模
型提供了丰富的数据资源,且数据库的通用性强,有
效提升了模型精度和泛化性;另一方面,为了解决二

维图像与三维模型的对齐问题,基于经典的通用三
维人脸表达模型,通过模型拟合生成合成二维、三维
人脸数据库,对深度模型的训练有重要意义,但也依
赖于三维人脸模型的表达能力。(2)基于高质量三维
人脸数据库,构建大规模基准评价数据库,用于深度
模型性能的统一评估,对三维人脸重建的研究有指
导意义。构建更大规模、更丰富多样性的基准评价
数据库,仍然是重要的任务,是三维人脸研究与应用
的重要支撑。

2)模型角度。(1)人脸统计先验在单目三维人脸
重建中发挥着非常重要的作用,从人脸几何表征、深
度网络、损失函数和可微分渲染等角度进行创新,解
耦人脸形状、纹理、表情和反照率等属性,由粗到精
实现重建、渲染、编辑和动画等。但是,受输入图像

表8 三维人脸重建算法总结归纳
Table 8 Summary of 3D face reconstruction algorithms

方法	特点	缺点	难点
单目三维人脸重建	1)输入数据获取便捷,通常基于大规模野外人脸数据进行自监督学习,或者使用大规模合成数据进行监督学习。2)根据是否使用三维人脸基础参数模型,可以分为两种方案:(1)结合经典的预训练网络(如ResNet)并设计新的网络结构对参数模型(如3DMM、FLAME)进行几何、纹理和反射等属性的解耦及优化,重建高精度个性化细节;(2)挖掘新的人脸几何表征(如体素、位移图)、新的属性(如目标对称性)及生成人脸先验等。3)基于生成对抗网络GAN的方法,通常利用生成网络(如StyleGANv2)解决纹理数据(如退化人脸)和渲染难题,重建细节属性、新视角渲染及重光照等。4)基于隐式神经表示或融合隐式表示与GAN的方法,通过构建数据库、隐式表示几何及可微渲染实现解耦重建和优化,用于新视角生成、人脸动画、纹理编辑等。	1)输入数据简单,通常需要借助其他人脸先验特征或假设先验,因此容易受数据和先验知识的限制,算法的精度和鲁棒性还需要更多的研究;2)真实场景中,极端表情、姿态、遮挡、照明和运动等条件变化都会造成重建精度的严重下降。	1)高精度表情细节重建、表情迁移及人脸动画生成,是研究难点;2)遮挡,包括姿态、照明引起的自遮挡和口罩、眼镜等外物遮挡,是研究难点,三维人脸重建及对齐的协同多任务模型,是重要研究方向;3)实现速度与精度的突破提升是研究难点,通过挖掘强大的新的深度神经网络,解决数据、参数、渲染和驱动等难点,将大大推动单目重建的落地应用。
多目三维人脸重建	1)输入多幅不同视角的二维人脸图像,能够有效利用多视角一致性,解耦身份一致性结构和场景个性化属性,一定程度上解决姿态造成的遮挡问题;2)多目重建是三维视觉测量的经典方法,有很多成熟的高性能模型可以使用,深度融合几何与学习的方法,对提升重建精度有重要意义。	1)输入数据获取更复杂,而数据库的规模和多样性不及单目;2)算法复杂度较高,深度神经网络结构设计及训练更困难。	如何挖掘多视角一致性先验,如何对表情、光照变化等进行表征,如何融合多通道网络,提高算法时间复杂度,是研究难点。
基于视频的三维人脸重建	1)输入数据获取简单且数据量大,大多来自互联网或使用便携设备无约束采集;2)基于隐式神经表示的方法是当前主流方法,设计新的表征和轻量化网络,并进行网格压缩,能够大大提升渲染速度,同时解耦表情运动与人脸表观。	1)输入数据预处理复杂,帧间差异性大;2)提升模型的效率,实现实时高质量数字化身创造是研究重点。	如何挖掘视频帧嵌入特征,进行模型压缩提升渲染速度,如何设计新的表征和网络结构,是研究难点。
基于语音的三维人脸重建	1)输入数据获取相对复杂,目前研究大都基于自主构建的小规模数据集;2)正处在迅速发展阶段,尤其是基于Transformer的模型不断出现,为解决跨模态映射问题提供了新方法。	1)缺少大规模基准数据库;2)精度还有很大提升空间。	如何跨模态映射建模,从语音信号提取人脸视觉特征,提升模型特征表达和重建精度,是研究难点。

的分辨率、人脸统计先验的几何拓扑、训练数据库及预训练模型的限制,重建准确度量的三维人脸并进行细粒度三维控制,仍然是一项挑战性任务。(2)深度模型直接从训练数据回归学习图像到参数或图像到几何的映射,不借助人脸统计先验前提下,面对较大的身份差异、极端姿态和表情变化时,容易出现大偏差而无法获得准确重建结果。近两年基于生成对抗网络,或者结合隐式神经表示与生成对抗网络,能够生成高质量的纹理图,建模更复杂的几何特征,且受分辨率和拓扑限制更小,为单目三维人脸重建开辟了新思路,表现出了巨大潜力。但是,重建高精度微观尺度的面部几何细节,实现真实感人脸渲染、可驱动动画、表情编辑迁移等,提高模型对未见的新表情新姿态的泛化性,还需要解决很多难题。表 9 列出了部分单目三维人脸重建在 REALY 数据库上的定量评价结果。

表 9 单目三维人脸重建方法在 REALY 数据库上
定量评价结果

Table 9 Quantitative evaluation results of monocular 3D
face reconstruction methods on REALY dataset

/mm

方法	标准均方差		
	@nose	@mouth	@all
HRN(Lei 等, 2023)	1.722	1.357	1.537
Deep3D(Deng 等, 2019b)	1.719	1.368	1.657
AlbedoGAN(Rai 等, 2023)	1.656	2.087	1.746
MGCNet(Shang 等, 2020)	1.771	1.417	1.774
GANFIT(Gecer 等, 2019)	1.928	1.812	1.868
SADNet(Ruan 等, 2021)	1.791	1.591	1.913
3DDFA_V2(Guo 等, 2020)	1.903	1.597	1.926
DECA(Feng 等, 2021)	1.697	2.516	2.010
PRNet(Feng 等, 2018a)	1.923	1.838	2.013
CEST(Wen 等, 2021)	2.779	1.448	2.017
SynergyNet(Wu 等, 2021)	2.026	1.731	2.021

注:加粗字体表示各列最优结果。

2.3.2 多目三维人脸重建

1)数据源角度。(1)基于大规模野外人脸图像,虽然有助于提升学习模型的泛化性,但模型复杂度高,重建精度和速度很难实现质的突破;(2)基于良好配准的高分辨率多目人脸图像,也存在很多困难,

如数据库建设成本高、数据处理难度大,对大部分研究者来说并不容易实现。而通过自主构建限制条件下的小规模数据集,或者基于高质量数据库合成数据集,进行网络训练和性能评价,无法对算法进行统一评价,模型对未见的新表情新姿态的泛化性更有待提升。

2)模型角度。首先,结合人脸统计先验,融合传统几何方法与深度神经网络,充分挖掘多视一致性,能够有效提升重建精度;其次,借助可学习的三维人脸表示,引入最先进的深度模型,尤其是神经辐射场和生成对抗网络的联合应用,并嵌入三维人脸先验信息,取得了非常显著的效果,实现了身份、表情、姿态及表观属性的解耦重建和逼真渲染。但是,现有模型对野外人脸数据的复杂照明、极端表情、大姿态和遮挡等,尚无法得到满意的结果,提出轻量化、快速、鲁棒和稳健的系统,还需更多的努力。

2.3.3 基于视频的三维人脸重建

1)数据源角度。输入或是来自互联网的野外视频,或是使用便携设备无约束自主采集的视频,数据获取难度不大,但是存在分辨率低、噪声和运动模糊等缺点,并且数据量激增及成像条件的未知,都导致计算代价和网络训练的负荷,限制了重建精度的提升。

2)模型角度。基于神经辐射场的高保真头部化身重建,大量工作聚焦在模型速度提升上,已经从数十小时提升到几分钟,但是人脸的肌肉运动、表情变化非常复杂,现有人脸跟踪模型难以捕捉连续的动态表情细节,同时受制于人脸表达模型、深度推理过程等因素,实现协同提升重建精度、渲染质量和训练效率,仍然面临很多挑战。如何挖掘视频帧丰富的嵌入信息,如何与新模型结合,实现逼真实时重建渲染,在“数字人”、VR(virtual reality)/AR(augmented reality)等领域有重要的应用前景。

2.3.4 基于语音的三维人脸重建

1)数据源角度。构建大规模基准数据库。相比基于图像的三维人脸重建,该研究方向的数据明显偏少,目前大都是自主构建小规模数据集,缺乏公开的大规模基准数据库,限制了深度模型的训练及统一评价。

2)模型角度。重建精度和渲染质量还有很大提升空间。目前主要关注人脸几何形状及同步嘴唇运动的重建,但是人脸还包含纹理、姿态和表情等复杂

属性,将跨模态映射建模为回归任务可能导致过于平滑的人脸几何,还会受限于通用人脸模型,无法生成细粒度的人脸细节。如何从语音信号中分析提取人脸相关视觉特征,提升模型特征表达能力和重建精度,是未来研究的主要任务。

3 结 语

无约束/非受控场景下,光照、遮挡、表情、姿态、分辨率和运动模糊等诸多环境和成像因素,都会对人脸系统的精度产生严重影响,还无法满足许多特定目标任务的应用需求,尤其是涉及安全防控的场景。尽管深度学习技术一定程度上解决了部分难题,模型的精度、速度取得了一定成果,但是依然存在很多亟待解决的挑战,尤其是如何保证系统的鲁棒性、稳定性和安全性,是人脸相关技术发展与应用巨大挑战。

1)数据。基于数据驱动的学习方法,用于深度神经网络训练的二维和三维人脸数据库,应该是大规模的、多样性的以及对齐配准的,能够覆盖不同的面部形状、表情、姿态、阴影、遮挡和运动等,表达现实场景的光照环境及变化情况。虽然面向具体的应用任务,各类新设备、新系统和新模型不断涌现,但是依然存在硬件成本高、计算代价大、算法复杂度高和数据差异性大等问题,三维人脸成像与重建技术的不断革新与协同发展,是解决三维人脸数据问题的关键。

2)提升对大的姿态和表情变化、极端成像环境、复杂遮挡的精度和鲁棒性。虽然针对姿态、表情、光照和遮挡,提出了一系列成果,但是还远达不到实际应用的需求,这些挑战性问题很容易造成系统异常,提升成像设备和重建模型的鲁棒性迫在眉睫。如何实现人脸运动和表情变化的精确捕捉和连续协同控制,赋予“数字人”准确的情绪、基于物理的渲染性能,给用户带来良好的沉浸式交互体验,是未来研究的重要目标。

3)多模态整合。个性化“数字人”的创造,是当下计算机视觉、人工智能的新兴技术,正处在蓬勃发展时期,对各个领域的发展都有重要作用。基于图像视觉特征的三维人脸重建技术,是最直接、最主要的三维建模方法。融合图像、文本和语音等多种模态,借助不同模态最新深度模型强大的表达能力,能

够实现更高精度的三维人脸重建、真实感渲染和面部资产生成,并对人脸形状、表情、表观和头发等属性编辑、可操控动画和特殊定制。

三维人脸成像及重建技术是人工智能领域的研究热点,尤其面向小尺寸、快速便捷的智能移动终端的微型人工智能,是三维人脸技术的重点研究方向,更是当下和未来人工智能突破发展与普及应用的主要场景。本文为三维人脸相关研究提供了较完整的参考资料,期望面向应用不断提出新方法、解决新问题,推动三维人脸的技术突破与更深更广的落地应用。

参考文献(References)

- 3dMD Inc. [EB/OL]. [2024-06-06]. <https://3dmd.com/>
- Abrevaya V F, Boukhayma A, Torr P H S and Boyer E. 2020. Cross-modal deep face normals with deactivable skip connections//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4978-4988 [DOI: 10.1109/CVPR42600.2020.00503]
- Azinović D, Maury O, Hery C, Nießner M and Thies J. 2023. High-res facial appearance capture from polarized smartphone images//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 16836-16846 [DOI: 10.1109/CVPR52729.2023.01615]
- Bae S I, Lee S, Kwon J M, Kim H K, Jang K W, Lee D and Jeong K H. 2022. Machine-learned light-field camera that reads facial expression from high-contrast and illumination invariant 3D facial images. *Advanced Intelligent Systems*, 4(4): #2100182 [DOI: 10.1002/aisy.202100182]
- Bagdanov A D, Del Bimbo A and Masi I. 2011. The florence 2D/3D hybrid face dataset//Proceedings of 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding. Scottsdale, USA: ACM: 79-80 [DOI: 10.1145/2072572.2072597]
- Bai H R, Kang D, Zhang H X, Pan J S and Bao L C. 2023. FFHQ-UV: normalized facial UV-texture dataset for 3D face reconstruction//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 362-371 [DOI: 10.1109/CVPR52729.2023.00043]
- Bai Z Q, Cui Z P, Rahim J A, Liu X M and Tan P. 2020. Deep facial non-rigid multi-view stereo//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5849-5859 [DOI: 10.1109/cvpr42600.2020.00589]
- Bansal A, Nanduri A, Castillo C D, Ranjan R and Chellappa R. 2017. UMDFaces: an annotated face dataset for training deep networks//Proceedings of 2017 IEEE International Joint Conference on Bio-

- metrics. Denver, USA: IEEE: 464-473 [DOI: 10.1109/BTAS.2017.8272731]
- Bao L C, Lin X K, Chen Y J, Zhang H X, Wang S, Zhe X F, Kang D, Huang H Z, Jiang X W, Wang J, Yu D and Zhang Z Y. 2021. High-fidelity 3D digital human head creation from RGB-D selfies. *ACM Transactions on Graphics*, 41(1): #3 [DOI: 10.1145/3472954]
- Beeler T, Bickel B, Beardsley P, Sumner B and Gross M. 2010. High-quality single-shot capture of facial geometry. *ACM Transactions on Graphics*, 29(4): #40 [DOI: 10.1145/1778765.1778777]
- Beeler T, Hahn F, Bradley D, Bickel B, Beardsley P, Gotsman C, Sumner R W and Gross M. 2011. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics*, 30(4): #75 [DOI: 10.1145/2010324.1964970]
- Benitez-Quiroz C F, Srinivasan R and Martinez A M. 2016. EmotioNet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 5562-5570 [DOI: 10.1109/CVPR.2016.600]
- Blanz V and Vetter T. 1999. A morphable model for the synthesis of 3D faces//*Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*. [s.l.]: ACM Press/Addison-Wesley Publishing Co.: 187-194 [DOI: 10.1145/311535.311556]
- Booth J, Roussos A, Zafeiriou S, Ponniah A and Dunaway D. 2016. A 3D morphable model learnt from 10 000 faces//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 5543-5552 [DOI: 10.1109/CVPR.2016.598]
- Bouritsas G, Bokhnyak S, Ploumpis S, Zafeiriou S and Bronstein M. 2019. Neural 3D morphable models: spiral convolutional networks for 3D shape representation learning and generation//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 7212-7221 [DOI: 10.1109/ICCV.2019.00731]
- Brock A, Donahue J and Simonyan K. 2019. Large scale GAN training for high fidelity natural image synthesis [EB/OL]. [2024-06-06]. <https://arxiv.org/pdf/1809.11096.pdf>
- Bulat A and Tzimiropoulos G. 2017. How far are we from solving the 2D and 3D face alignment problem? (and a dataset of 230 000 3D facial landmarks)//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 1021-1030 [DOI: 10.1109/ICCV.2017.116]
- Cao C, Simon T, Kim J K, Schwartz G, Zollhoefer M, Saito S S, Lombardi S, Wei S E, Belko D, Yu S I, Sheikh Y and Saragih J. 2022. Authentic volumetric avatars from a phone scan. *ACM Transactions on Graphics*, 41(4): #163 [DOI: 10.1145/3528223.3530143]
- Cao C, Weng Y L, Zhou S, Tong Y Y and Zhou K. 2014. FaceWarehouse: a 3D facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 20(3): 413-425 [DOI: 10.1109/TVCG.2013.249]
- Cao Q, Shen L, Xie W D, Parkhi O M and Zisserman A. 2018. VGG-Face2: a dataset for recognising faces across pose and age//*Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition*. Xi'an, China: ACM: 67-74 [DOI: 10.1109/FG.2018.00020]
- Chai Z H, Zhang H X, Ren J, Kang D, Xu Z Z, Zhe X F, Yuan C and Bao L C. 2022. REALY: rethinking the evaluation of 3D face reconstruction//*Proceedings of the 17th European Conference on Computer Vision*. Tel Aviv, Israel: Springer: 74-92 [DOI: 10.1007/978-3-031-20074-8_5]
- Chan E R, Lin C Z, Chan M A, Nagano K, Pan B X, de Mello S, Gallo O, Guibas L, Tremblay J, Khamis S, Karras T and Wetzstein G. 2022. Efficient geometry-aware 3D generative adversarial networks//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 16102-16112 [DOI: 10.1109/CVPR52688.2022.01565]
- Chatziagapi A and Samaras D. 2023. AVFace: towards detailed audio-visual 4D face reconstruction//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 16878-16889 [DOI: 10.1109/CVPR52729.2023.01619]
- Chen B C, Chen C S and Hsu W H. 2014. Cross-age reference coding for age-invariant face recognition and retrieval//*Proceedings of the 13th European Conference on Computer Vision*. Zurich, Switzerland: Springer: 768-783 [DOI: 10.1007/978-3-319-10599-4_49]
- Chen G Y, Han K and Wong K Y K. 2018. PS-FCN: a flexible learning framework for photometric stereo//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01240-3_1]
- Cheng S Y, Kotsia I, Pantic M and Zafeiriou S. 2018. 4DFAB: a large scale 4D database for facial expression analysis and biometric applications//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 5117-5126 [DOI: 10.1109/CVPR.2018.00537]
- Chrysos G G, Antonakos E, Zafeiriou S and Snape P. 2015. Offline deformable face tracking in arbitrary videos//*Proceedings of 2015 IEEE International Conference on Computer Vision Workshop*. Santiago, Chile: IEEE: 954-962 [DOI: 10.1109/ICCVW.2015.126]
- Chung J S, Nagrani A and Zisserman A. 2018. VoxCeleb2: deep speaker recognition//*Interspeech* 2018. Hyderabad, India: [s.n.]: 1086-1090 [DOI: 10.21437/Interspeech.2018-1929]
- Cudeiro D, Bolkart T, Laidlaw C, Ranjan A and Black M J. 2019. Capture, learning, and synthesis of 3D speaking styles//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 10093-10103 [DOI: 10.1109/CVPR.2019.01034]
- Dai H, Pears N, Smith W and Duncan C. 2017. A 3D morphable model of craniofacial shape and texture variation//*Proceedings of 2017*

- IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 3104-3112 [DOI: 10.1109/ICCV.2017.335]
- Dai H, Pears N, Smith W and Duncan C. 2020. Statistical modeling of craniofacial shape and texture. *International Journal of Computer Vision*, 128(2): 547-571 [DOI: 10.1007/s11263-019-01260-7]
- Daněček R, Black M and Bolkart T. 2022. EMOCA: emotion driven monocular face capture and animation//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 20279-20290 [DOI: 10.1109/CVPR52688.2022.01967]
- Debevec P, Hawkins T, Tchou C, Duiker H P, Sarokin W and Sagar M. 2000. Acquiring the reflectance field of a human face//*Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. [s.l.]: ACM Press/Addison-Wesley Publishing Co.: 145-156 [DOI: 10.1145/344779.344855]
- Debevec P. 2012. The light stages and their applications to photoreal digital actors. *SIGGRAPH Asia*, 2(4): 1-4
- Deng J K, Guo J, Liu T L, Gong M M and Zafeiriou S. 2020a. Sub-center ArcFace: boosting face recognition by large-scale noisy web faces//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 741-757 [DOI: 10.1007/978-3-030-58621-8_43]
- Deng J K, Guo J, Ververas E, Kotsia I and Zafeiriou S. 2020b. RetinaFace: single-shot multi-level face localisation in the wild//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 5202-5211 [DOI: 10.1109/CVPR42600.2020.00525]
- Deng J K, Guo J, Xue N N and Zafeiriou S. 2019a. ArcFace: additive angular margin loss for deep face recognition//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 4685-4694 [DOI: 10.1109/CVPR.2019.00482]
- Deng Y, Yang J L, Xu S C, Chen D, Jia Y D and Tong X. 2019b. Accurate 3D face reconstruction with weakly-supervised learning: from single image to image set//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Long Beach, USA: IEEE: 285-295 [DOI: 10.1109/CVPRW.2019.00038]
- DI4D Inc. [EB/OL]. [2024-01-24]. <https://di4d.com/>
- Dib A, Ahn J, Thébaud C, Gosselin P H and Chevallier L. 2023. S2F2: self-supervised high fidelity face reconstruction from monocular image//*Proceedings of the 17th IEEE International Conference on Automatic Face and Gesture Recognition*. Waikoloa Beach, USA: ACM: 1-8 [DOI: 10.1109/FG57933.2023.10042713]
- Dib A, Thébaud C, Ahn J, Gosselin P H, Theobalt C and Chevallier L. 2021. Towards high fidelity monocular face reconstruction with rich reflectance using self-supervised learning and ray tracing//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 12799-12809 [DOI: 10.1109/ICCV48922.2021.01258]
- Egger B, Smith W A P, Tewari A, Wuhler S, Zollhoefer M, Beeler T, Bernard F, Bolkart T, Kortylewski A, Romdhani S, Theobalt C, Blanz V and Vetter T. 2020. 3D morphable face models-past, present, and future. *ACM Transactions on Graphics*, 39(5): #157 [DOI: 10.1145/3395208]
- Fan Y R, Lin Z J, Saito J, Wang W P and Komura T. 2022. FaceFormer: speech-driven 3D facial animation with transformers//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 18749-18758 [DOI: 10.1109/CVPR52688.2022.01821]
- Feng Y, Feng H W, Black M J and Bolkart T. 2021. Learning an animatable detailed 3D face model from in-the-wild images. *ACM Transactions on Graphics*, 40(4): #88 [DOI: 10.1145/3450626.3459936]
- Feng Y, Wu F, Shao X H, Wang Y F and Zhou X. 2018a. Joint 3D face reconstruction and dense alignment with position map regression network//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 557-574 [DOI: 10.1007/978-3-030-01264-9_33]
- Feng Z H, Huber P, Kittler J, Hancock P, Wu X J, Zhao Q J, Koppen P and Raetsch M. 2018b. Evaluation of dense 3D reconstruction from 2D face images in the wild//*Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition*. Xi'an, China: IEEE: 780-786 [DOI: 10.1109/FG.2018.00123]
- Fu H B, Bian S J, Chaudhry E, Iglesias A, You L H and Zhang J J. 2021. State-of-the-Art in 3D face reconstruction from a single RGB image//*Proceedings of the 21st International Conference on Computational Science*. Krakow, Poland: Springer: 31-44 [DOI: 10.1007/978-3-030-77977-1_3]
- Fyffe G, Jones A, Alexander O, Ichikari R and Debevec P. 2014. Driving high-resolution facial scans with video performance capture. *ACM Transactions on Graphics*, 34(1): #8 [DOI: 10.1145/2638549]
- Gafni G, Thies J, Zollhöfer M and Nießner M. 2021. Dynamic neural radiance fields for monocular 4D facial avatar reconstruction//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 8645-8654 [DOI: 10.1109/CVPR46437.2021.00854]
- Galanakis S, Gecer B, Lattas A and Zafeiriou S. 2023. 3DMM-RF: convolutional radiance fields for 3D face modeling//*Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE: 3525-3536 [DOI: 10.1109/WACV56688.2023.00353]
- Gao X, Zhong C L, Xiang J, Hong Y, Guo Y D and Zhang J Y. 2022. Reconstructing personalized semantic facial NeRF models from monocular video. *ACM Transactions on Graphics*, 41(6): #200

- [DOI: 10.1145/3550454.3555501]
- Gecer B, Deng J K and Zafeiriou S. 2021. OSTeC: one-shot texture completion//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 7624-7634 [DOI: 10.1109/CVPR46437.2021.00754]
- Gecer B, Lattas A, Ploumpis S, Deng J K, Papaioannou A, Moschoglou S and Zafeiriou S. 2020. Synthesizing coupled 3D face modalities by trunk-branch generative adversarial networks//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 415-433 [DOI: 10.1007/978-3-030-58526-6_25]
- Gecer B, Ploumpis S, Kotsia I and Zafeiriou S. 2019. GANFIT: generative adversarial network fitting for high fidelity 3D face reconstruction//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1155-1164 [DOI: 10.1109/CVPR.2019.00125]
- Gerig T, Forster A M, Blumer C, Egger B, Luthi M, Schoenborn S and Vetter T. 2018. Morphable face models——an open framework//Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition. Xi'an, China: ACM: 75-82 [DOI: 10.1109/FG.2018.00021]
- Ghosh A, Fyffe G, Tunwattanapong B, Busch J, Yu X M and Debevec P. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Transactions on Graphics*, 30(6): 1-10 [DOI: 10.1145/2070781.2024163]
- Giebenhain S, Kirschstein T, Georgopoulos M, Rünz M, Agapito L and Nießner M. 2023. Learning neural parametric head models//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 21003-21012 [DOI: 10.1109/CVPR52729.2023.02012]
- Gotardo P, Riviere J, Bradley D, Ghosh A and Beeler T. 2018. Practical dynamic facial appearance modeling and acquisition. *ACM Transactions on Graphics*, 37(6): #232 [DOI: 10.1145/3272127.3275073]
- Grassal P W, Prinzer M, Leistner T, Rother C, Nießner M and Thies J. 2022. Neural head avatars from monocular RGB videos//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 18632-18643 [DOI: 10.1109/CVPR52688.2022.01810]
- Gross R, Matthews I, Cohn J, Kanade T and Baker S. 2010. Multi-PIE. *Image and Vision Computing*, 28(5): 807-813 [DOI: 10.1016/j.imavis.2009.08.002]
- Gu J T, Liu L J, Wang P and Theobalt C. 2021. StyleNeRF: a style-based 3D-aware generator for high-resolution image synthesis [EB/OL]. [2024-06-06]. <https://arxiv.org/pdf/2110.08985.pdf>
- Guo J, Yu J K, Lattas A and Deng J K. 2023a. Perspective reconstruction of human faces by joint mesh and landmark regression//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 350-365 [DOI: 10.1007/978-3-031-25072-9_23]
- Guo J Z, Zhu X Y, Yang Y, Yang F, Lei Z and Li S Z. 2020. Towards fast, accurate and stable 3D dense face alignment//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 152-168 [DOI: 10.1007/978-3-030-58529-7_10]
- Guo L W, Zhu H, Lu Y X, Wu M H and Cao X. 2023b. RAFaRe: learning robust and accurate non-parametric 3D face reconstruction from pseudo 2D and 3D pairs//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI: 719-727 [DOI: 10.1609/aaai.v37i1.25149]
- Han Y X, Wang Z B and Xu F. 2023. Learning a 3D morphable face reflectance model from low-cost data//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 8598-8608 [DOI: 10.1109/CVPR52729.2023.00831]
- Hannun A, Case C, Casper J, Catanzaro B, Diamos G, Elsen E, Prenger R, Satheesh S, Sengupta S, Coates A and Ng A Y. 2014. Deep speech: scaling up end-to-end speech recognition [EB/OL]. [2024-06-06]. <https://arxiv.org/pdf/1412.5567.pdf>
- He J Y, Huang H B, Zhang H Y, Sun M Y, Liu Y H and Zhou Z H. 2022. Review of 3D face reconstruction based on single image. *Computer Science*, 49(2): 40-50 (何嘉玉, 黄宏博, 张红艳, 孙牧野, 刘亚辉, 周哲海. 2022. 基于深度学习的单幅图像三维人脸重建研究综述. *计算机科学*, 49(2): 40-50) [DOI: 10.11896/jsjx.210500215]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hong Y, Peng B, Xiao H Y, Liu L G and Zhang J Y. 2022. HeadNeRF: a realtime NeRF-based parametric head model//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 20342-20352 [DOI: 10.1109/CVPR52688.2022.01973]
- Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2024-06-06]. <https://arxiv.org/pdf/1704.04861.pdf>
- Huang G B, Ramesh M A, Berg T L and Learned-Miller E. 2007. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts InsightFace. [EB/OL]. [2024-06-06]. <https://github.com/deepinsight/insightface>
- Jackson A S, Bulat A, Argyriou V and Tzimiropoulos G. 2017. Large pose 3D face reconstruction from a single Image via direct volumetric CNN regression//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 1031-1039 [DOI: 10.1109/ICCV.2017.117]

- Jing Y P, Lu X Q and Gao S. 2021. 3D face recognition: a survey [EB/OL]. [2024-06-06]. <https://arxiv.org/pdf/2108.11082.pdf>
- Kang M J, Choe J, Ha H, Jeon H G, Im S, Kweon I S and Yoon K J. 2022. Facial depth and normal estimation using single dual-pixel camera//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 181-200 [DOI: 10.1007/978-3-031-20074-8_11]
- Karras T, Aila T, Laine S and Lehtinen J. 2018. Progressive growing of GANs for improved quality, stability, and variation [EB/OL]. [2024-06-06]. <https://arxiv.org/pdf/1710.10196.pdf>
- Karras T, Laine S and Aila T. 2019. A style-based generator architecture for generative adversarial networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4396-4405 [DOI: 10.1109/CVPR.2019.00453]
- Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J and Aila T. 2020. Analyzing and improving the image quality of styleGAN//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 8107-8116 [DOI: 10.1109/CVPR42600.2020.00813]
- Kipf T N and Welling M. 2017. Semi-supervised classification with graph convolutional networks [EB/OL]. [2024-06-06]. <https://arxiv.org/pdf/1609.02907.pdf>
- Konica Minolta Vivid 910. 2024. [EB/OL]. [2024-06-06]. https://www.konicaminolta.com.cn/instruments/download/catalog/pdf/vivid910_C.pdf
- Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84-90 [DOI: 10.1145/3065386]
- Lattas A, Lin Y M, Kannan J, Ozturk E, Filipi L, Guarnera G C, Chawla G and Ghosh A. 2022a. Practical and scalable desktop-based high-quality facial capture//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 522-537 [DOI: 10.1007/978-3-031-20068-7_30]
- Lattas A, Moschoglou S, Gecer B, Ploumpis S, Triantafyllou V, Ghosh A and Zafeiriou S. 2020. AvatarMe: realistically renderable 3D facial reconstruction "in-the-wild"//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 757-766 [DOI: 10.1109/CVPR42600.2020.00084]
- Lattas A, Moschoglou S, Ploumpis S, Gecer B, Deng J K and Zafeiriou S. 2023. FitMe: deep photorealistic 3D morphable model avatars//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 8629-8640 [DOI: 10.1109/CVPR52729.2023.00834]
- Lattas A, Moschoglou S, Ploumpis S, Gecer B, Ghosh A and Zafeiriou S. 2022b. AvatarMe⁺⁺: facial shape and BRDF inference with photorealistic rendering-aware GANs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12): 9269-9284 [DOI: 10.1109/TPAMI.2021.3125598]
- Lee C H, Liu Z W, Wu L Y and Luo P. 2020a. MaskGAN: towards diverse and interactive facial image manipulation//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5548-5557 [DOI: 10.1109/CVPR42600.2020.00559]
- Lee G H and Lee S W. 2020b. Uncertainty-aware mesh decoder for high fidelity 3D face reconstruction//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 6099-6108 [DOI: 10.1109/CVPR42600.2020.00614]
- Lei B W, Ren J Q, Feng M Y, Cui M M and Xie X S. 2023. A hierarchical representation network for accurate and detailed face reconstruction from in-the-wild images//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 394-403 [DOI: 10.1109/CVPR52729.2023.00046]
- Li C L, Morel-Forster A, Vetter T, Egger B and Kortylewski A. 2023a. Robust model-based face reconstruction through weakly-supervised outlier segmentation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 372-381 [DOI: 10.1109/CVPR52729.2023.00044]
- Li T Y, Bolkart T, Black M J, Li H and Romero J. 2017. Learning a model of facial shape and expression from 4D scans. ACM Transactions on Graphics, 36(6): #194 [DOI: 10.1145/3130800.3130813]
- Li T Y, Liu S C, Bolkart T, Liu J Y, Li H and Zhao Y J. 2021. Topologically consistent multi-view face inference using volumetric sampling//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 3804-3814 [DOI: 10.1109/ICCV48922.2021.00380]
- Li Y, Hao Q, Hu J G, Pan X M, Li Z C and Cui Z. 2023b. 3D3M: 3D modulated morphable model for monocular face reconstruction. IEEE Transactions on Multimedia, 25: 6642-6652 [DOI: 10.1109/TMM.2022.3212282]
- Light Stage. [EB/OL]. [2024-06-06]. <https://vgl.ict.usc.edu/LightStages/>
- Light Field Stage. [EB/OL]. [2024-01-24]. <https://deemos.com/>
- Lin C, Nagano K, Kautz J, Chan E, Iqbal U, Guibas L, Wetzstein G and Khamis S. 2023. Single-shot implicit morphable faces with consistent texture parameterization//Proceedings of 2023 ACM SIGGRAPH Conference Proceedings. Los Angeles, USA: ACM: 83 [DOI: 10.1145/3588432.3591494]
- Lin J K, Yuan Y, Shao T J and Zhou K. 2020. Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5890-5899 [DOI: 10.1109/CVPR42600.2020.00593]
- Lin T Y, Dollar P, Girshick R, He K M, Hariharan B and Belongie S.

2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Ling J W, Wang Z B, Lu M, Wang Q, Qian C and Xu F. 2022. Structure-aware editable morphable model for 3D facial detail animation and manipulation//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 249-267 [DOI: 10.1007/978-3-031-20062-5_15]
- Liu X, Xu Y H, Wu Q Y, Zhou H, Wu W and Zhou B L. 2022. Semantic-aware implicit neural audio-driven video portrait generation//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 106-125 [DOI: 10.1007/978-3-031-19836-6_7]
- Liu Z W, Luo P, Wang X G and Tang X O. 2015. Deep learning face attributes in the wild//Proceedings of 2015 International Conference on Computer Vision. Santiago, Chile: IEEE: 3730-3738 [DOI: 10.1109/ICCV.2015.425]
- Luo C W, Yu J, Yu L Y, Li Y L and Wang S J. 2021. Overview of research progress on 3-D face recognition. Journal of Tsinghua University (Science and Technology), 61(1): 77-88 (罗常伟, 於俊, 于灵云, 李亚利, 王生进. 2021. 三维人脸识别研究进展综述. 清华大学学报(自然科学版), 61(1): 77-88). [DOI: 10.16511/j.cnki.qhdxsb.2020.21.016]
- Ma W C, Hawkins T, Peers P, Chabert C F, Weiss M and Debevec P. 2007. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination//Proceedings of the 18th Eurographics Conference on Rendering Techniques. Grenoble, France: Eurographics Association: 183-194
- Mallikarjun B R, Tewari A, Seidel H P, Elgharib M and Theobalt C. 2021. Learning complete 3D morphable face models from images and videos//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 3360-3370 [DOI: 10.1109/CVPR46437.2021.00337]
- Medusa Facial Capture System. 2024 [EB/OL]. [2024-06-06].
<https://la.disneyresearch.com/medusa/>
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2020. NeRF: representing scenes as neural radiance fields for view synthesis//Proceedings of the 16th European Conference on Computer Vision. Dublin, Ireland: Springer: 405-421 [DOI: 10.1007/978-3-030-58452-8_24]
- Mollahosseini A, Hasani B and Mahoor M H. 2019. Affectnet: a database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing, 10(1): 18-31 [DOI: 10.1109/TAFFC.2017.2740923]
- Morales A, Piella G and Sukno F M. 2021. Survey on 3D face reconstruction from uncalibrated images. Computer Science Review, 40: #100400 [DOI: 10.1016/j.cosrev.2021.100400]
- Müller T, Evans A, Schied C and Keller A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics, 41(4): #102 [DOI: 10.1145/3528223.3530127]
- Nagrani A, Chung J S and Zisserman A. 2017. VoxCeleb: a large-scale speaker identification dataset//Interspeech 2017. Stockholm, Sweden: [s.n.]: 2616-2620 [DOI: 10.21437/Interspeech.2017-950]
- Newell A, Yang K Y and Deng J. 2016. Stacked hourglass networks for human pose estimation//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 483-499 [DOI: 10.1007/978-3-319-46484-8_29]
- NextFace. [EB/OL]. [2024-06-06].
<https://github.com/abdallahdib/NextFace>
- Pan X G, Tewari A, Liu L J and Theobalt C. 2022. GAN2X: non-lambertian inverse rendering of image GANs//Proceedings of 2022 International Conference on 3D Vision. Prague, Czech Republic: IEEE: 711-721 [DOI: 10.1109/3DV57658.2022.00081]
- Papaioannou A, Gecer B, Cheng S Y, Chrysos G, Deng J K, Fotiadou E, Kampouris C, Kollias D, Moschoglou S, Songsri-In K, Ploumpis S, Trigeorgis G, Tzirakis P, Ververas E, Zhou Y X, Ponniah A, Roussos A and Zafeiriou S. 2022. MimicME: a large scale diverse 4D database for facial expression analysis//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 467-484 [DOI: 10.1007/978-3-031-20074-8_27]
- Parkhi O M, Vedaldi A and Zisserman A. 2015. Deep face recognition//Proceedings of 2015 British Machine Vision Conference. IEEE: 41: 1-12
- Paysan P, Knothe R, Amberg B, Romdhani S and Vetter T. 2009. A 3D face model for pose and illumination invariant face recognition//Proceedings of the 6th IEEE International Conference on Advanced Video and Signal based Surveillance. Genova, Italy: IEEE: 296-301 [DOI: 10.1109/AVSS.2009.58]
- Peng Z Q, Wu H Y, Song Z B, Xu H, Zhu X Y, He J, Liu H Y and Fan Z X. 2023. EmoTalk: speech-driven emotional disentanglement for 3D face animation [EB/OL]. [2024-06-06].
<https://arxiv.org/pdf/2303.11089.pdf>
- Phillips P J, Flynn P J, Scruggs T, Bowyer K W, Chang J, Hoffman K, Marques J, Min J and Worek W. 2005. Overview of the face recognition grand challenge//Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE: 947-954 [DOI: 10.1109/CVPR.2005.268]
- Piao J T, Sun K Q, Wang Q, Lin K Y and Li H S. 2021. Inverting generative adversarial renderer for face reconstruction//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 15614-15623 [DOI: 10.1109/CVPR46437.2021.01536]
- Pillai R K, Jeni L A, Yang H Y, Zhang Z, Yin L J and Cohn J F. 2019.

- The 2nd 3D face alignment in the wild challenge (3DFAW-video): dense reconstruction from video//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul, Korea (South): IEEE: 3082-3089 [DOI: 10.1109/ICCVW.2019.00371]
- Ploumpis S, Ververas E, O'Sullivan E, Moschoglou S, Wang H Y, Pears N, Smith W A P, Gecer B and Zafeiriou S. 2021. Towards a complete 3D morphable model of the human head. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(11): 4142-4160 [DOI: 10.1109/TPAMI.2020.2991150]
- Ploumpis S, Wang H Y, Pears N, Smith W A P and Zafeiriou S. 2019. Combining 3D morphable models: a large scale face-and-head model//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 10926-10935 [DOI: 10.1109/CVPR.2019.01119]
- Principi F, Berretti S, Ferrari C, Oterdout N, Daoudi M and Del Bimbo A. 2023. The florence 4D facial expression dataset//Proceedings of the 17th IEEE International Conference on Automatic Face and Gesture Recognition. Waikoloa Beach, USA: ACM: 1-6 [DOI: 10.1109/FG57933.2023.10042606]
- Rai A, Gupta H, Pandey A, Carrasco F V, Takagi S J, Aubel A, Kim D, Prakash A and de la Torre F. 2023. Towards realistic generative 3D face models [EB/OL]. [2024-06-06].
<https://arxiv.org/pdf/2304.12483.pdf>
- Ramon E, Triginer G, Escur J, Pumarola A, Garcia J, Giró-i-Nieto X and Moreno-Noguer F. 2021. H3D-net: few-shot high-fidelity 3D head reconstruction//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 5600-5609 [DOI: 10.1109/ICCV48922.2021.00557]
- Ranjan A, Bolkart T, Sanyal S and Black M J. 2018. Generating 3D faces using convolutional mesh autoencoders//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 725-741 [DOI: 10.1007/978-3-030-01219-9_43]
- Ren X Y, Lattas A, Gecer B, Deng J K, Ma C and Yang X K. 2023. Facial geometric detail recovery via implicit representation//Proceedings of the 17th IEEE International Conference on Automatic Face and Gesture Recognition. Waikoloa Beach, USA: IEEE: 1-8 [DOI: 10.1109/FG57933.2023.10042505]
- Richard A, Zollhöfer M, Wen Y D, de la Torre F and Sheikh Y. 2021. MeshTalk: 3D face animation from speech using cross-modality disentanglement//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 1153-1162 [DOI: 10.1109/ICCV48922.2021.00121]
- Richard A. N, Shahram I, Otmar H, David M, David, K, Andrew J D, Pushmeet K, Jamie S, Steve H and Andrew F. 2011. KinectFusion: real-time dense surface mapping and tracking//Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality. Basel, Switzerland: IEEE: 127-136 [DOI: 10.1109/ISMAR.2011.6092378]
- Riviere J, Gotardo P, Bradley D, Ghosh A and Beeler T. 2020. Single-shot high-quality facial geometry and skin appearance capture. ACM Transactions on Graphics, 39(4): #81 [DOI: 10.1145/3386569.3392464]
- Ronneberger O, Fischer P and Brox T. 2015. U-net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Rothe R, Timofte R and Van Gool L. 2015. DEX: deep expectation of apparent age from a single image//Proceedings of 2015 IEEE International Conference on Computer Vision Workshop. Santiago, Chile: IEEE: 252-257 [DOI: 10.1109/ICCVW.2015.41]
- Ruan Z Y, Zou C Q, Wu L H, Wu G S and Wang L M. 2021. SADRNet: self-aligned dual face regression networks for robust 3D dense face alignment and reconstruction. IEEE Transactions on Image Processing, 30: 5793-5806 [DOI: 10.1109/tip. 2021.3087397]
- Sanyal S, Bolkart T, Feng H W and Black M J. 2019. Learning to regress 3D face shape and expression from an image without 3D supervision//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 7755-7764 [DOI: 10.1109/CVPR.2019.00795]
- Savran A, Alyüz N, Dibeklioglu H, Çeliktutan O, Gökberk B, Sankur B and Akarun L. 2008. Bosphorus database for 3D face analysis//Proceedings of the 1st European Workshop on Biometrics and Identity Management. Roskilde, Denmark: Springer: 47-56 [DOI: 10.1007/978-3-540-89991-4_6]
- Schroff F, Kalenichenko D and Philbin J. 2015. FaceNet: a unified embedding for face recognition and clustering//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 815-823 [DOI: 10.1109/CVPR. 2015.7298682]
- Shang J X, Shen T W, Li S W, Zhou L, Zhen M M, Fang T and Quan L. 2020. Self-supervised monocular 3D face reconstruction by occlusion-aware multi-view geometry consistency//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 53-70 [DOI: 10.1007/978-3-030-58555-6_4]
- Sharma S and Kumar V. 2022. 3D face reconstruction in deep learning era: a survey. Archives of Computational Methods in Engineering, 29(5): 3475-3507 [DOI: 10.1007/s11831-021-09705-4]
- Shen Y J, Gu J J, Tang X O and Zhou B L. 2020. Interpreting the latent space of GANs for semantic face editing//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 9240-9249 [DOI: 10.1109/CVPR42600.2020.00926]
- Sitzmann V, Martel J N P, Bergman A W, Lindell D B and Wetzstein

- G. 2020. Implicit neural representations with periodic activation functions//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 7462-7473
- Smith W A P, Seck A, Dee H, Tiddeman B, Tenenbaum J B and Egger B. 2020. A morphable face albedo model//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5010-5019 [DOI: 10.1109/CVPR42600.2020.00506]
- Sun Z N, He R, Wang L, Kan M N, Feng J J, Zheng F, Zheng W S, Zuo W M, Kang W X, Deng H W, Zhang J, Han H, Shan S G, Wang Y L, Ru Y W, Zhu Y H, Liu Y F and He Y. 2021. Overview of biometrics research. *Journal of Image and Graphics*, 26(6): 1254-1329 (孙哲南, 赫然, 王亮, 阚美娜, 冯建江, 郑方, 郑伟诗, 左旺孟, 康文雄, 邓伟洪, 张杰, 韩晓, 山世光, 王云龙, 茹一伟, 朱宇豪, 刘云帆, 何勇. 2021. 生物特征识别学科发展报告. *中国图象图形学报*, 26(6): 1254-1329) [DOI: 10.11834/jig.210078]
- Tewari A, Bernard F, Garrido P, Bharaj G, Elgharib M, Seidel H P, Pérez P, Zollhöfer M and Theobalt C. 2019. FML: face model learning from videos//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 10804-10814 [DOI: 10.1109/CVPR.2019.01107]
- Tewari A, Zollhöfer M, Garrido P, Bernard F, Kim H, Perez P and Theobalt C. 2018. Self-supervised multi-level face model learning for monocular reconstruction at over 250 Hz//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2549-2559 [DOI: 10.1109/CVPR.2018.00270]
- Tewari A, Zollhofer M, Kim H, Garrido P, Bernard F, Perez P and Theobalt C. 2017. MoFA: model-based deep convolutional face autoencoder for unsupervised monocular reconstruction//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 3735-3744 [DOI: 10.1109/ICCV.2017.401]
- Thies J, Elgharib M, Tewari A, Theobalt C and Nießner M. 2020. Neural voice puppetry: audio-driven facial reenactment//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 716-731 [DOI: 10.1007/978-3-030-58517-4_42]
- Thies J, Zollhofer M, Stamminger M, Theobalt C and Nießner M. 2016. Face2Face: real-time face capture and reenactment of RGB videos//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2387-2395 [DOI: 10.1109/CVPR.2016.262]
- Tran A T, Hassner T, Masi I, Paz E, Nirkin Y and Medioni G. 2018. Extreme 3D face reconstruction: seeing through occlusions//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3935-3944 [DOI: 10.1109/CVPR.2018.00414]
- Tran L and Liu X M. 2018. Nonlinear 3D face morphable model//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7346-7355 [DOI: 10.1109/CVPR.2018.00767]
- Tu X G, Zhao J, Xie M, Jiang Z H, Balamurugan A, Luo Y, Zhao Y, He L X, Ma Z and Feng J S. 2021. 3D face reconstruction from a single image assisted by 2D face images in the wild. *IEEE Transactions on Multimedia*, 23: 1160-1172 [DOI: 10.1109/TMM.2020.2993962]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6000-6010
- Wang J T and Li H B. 2023. Review of single-image 3D face reconstruction methods. *Computer Engineering and Applications*, 59(17): 1-21 (王静婷, 李慧斌. 2023. 单幅图像三维人脸重建方法综述. *计算机工程与应用*, 59(17): 1-21) [DOI: 10.3778/j.issn.1002-8331.2210-0041]
- Wang L Z, Chen Z Y, Yu T, Ma C G, Li L and Liu Y B. 2022. FaceVerse: a fine-grained and detail-controllable 3D face morphable model from a hybrid dataset//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 20301-20310 [DOI: 10.1109/CVPR52688.2022.01969]
- Wang M, Deng W H, Hu J N, Tao X Q and Huang Y H. 2019. Racial faces in the wild: reducing racial bias by information maximization adaptation network//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 692-702 [DOI: 10.1109/ICCV.2019.00078]
- Wang T C, Liu M Y, Zhu J Y, Tao A, Kautz J and Catanzaro B. 2018. High-resolution image synthesis and semantic manipulation with conditional GANs//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8798-8807 [DOI: 10.1109/CVPR.2018.00917]
- Wang X Y, Guo Y D, Deng B L and Zhang J Y. 2020. Lightweight photometric stereo for facial details recovery//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 737-746 [DOI: 10.1109/CVPR42600.2020.00082]
- Wen Y D, Liu W Y, Raj B and Singh R. 2021. Self-supervised 3D face reconstruction via conditional estimation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 13269-13278 [DOI: 10.1109/ICCV48922.2021.01304]
- Wenger A, Gardner A, Tchou C, Unger J, Hawkins T and Debevec P. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics*,

- 24(3): 756-764 [DOI: 10.1145/1073204.1073258]
- Wood E, Baltrušaitis T, Hewitt C, Johnson M, Shen J J, Milosavljević N, Wilde D, Garbin S, Sharp T, Stojiljković I, Cashman T and Valentin J. 2022. 3D face reconstruction with dense landmarks//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 160-177 [DOI: 10.1007/978-3-031-19778-9_10]
- Wu C Y, Hsu C C and Neumann U. 2022. Cross-modal perceptionist: can face geometry be gleaned from voices?//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 10442-10451 [DOI: 10.1109/CVPR52688.2022.01020]
- Wu C Y, Xu Q G and Neumann U. 2021. Synergy between 3DMM and 3D landmarks for accurate 3D facial geometry//Proceedings of 2021 International Conference on 3D Vision. London, UK: IEEE: 453-463 [DOI: 10.1109/3DV53792.2021.00055]
- Wu F Z, Bao L C, Chen Y J, Ling Y G, Song Y B, Li S N, Ngan K N and Liu W. 2019. MVF-net: multi-view 3D face morphable model regression//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 959-968 [DOI: 10.1109/CVPR.2019.00105]
- Wu H Z, Jia J, Xing J L, Xu H W, Wang X Y and Wang J. 2023a. MMFace4D: a large-scale multi-modal 4D face dataset for audio-driven 3D face animation [EB/OL]. [2024-06-06]. <https://arxiv.org/pdf/2303.09797.pdf>
- Wu M H, Zhu H, Huang L J, Zhuang Y Y, Lu Y X and Cao X. 2023b. High-fidelity 3D face generation from natural language descriptions//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 4521-4530 [DOI: 10.1109/CVPR52729.2023.00439]
- Wu S J, Yan Y C, Li Y H, Cheng Y H, Zhu W H, Gao K, Li X B and Zhai G T. 2023c. GANHead: towards generative animatable neural head avatars//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 437-447 [DOI: 10.1109/CVPR52729.2023.00050]
- Wu S Z, Rupprecht C and Vedaldi A. 2020. Unsupervised learning of probably symmetric deformable 3D objects from images in the wild//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1-10 [DOI: 10.1109/CVPR42600.2020.00008]
- Xiao Y Z, Zhu H, Yang H T, Diao Z Y, Lu X J and Cao X. 2022. Detailed facial geometry recovery from multi-view images by learning an implicit function//Proceedings of the 36th AAAI Conference on Artificial Intelligence. Virtually: AAAI: 2839-2847 [DOI: 10.1609/aaai.v36i3.20188]
- Xing J B, Xia M H, Zhang Y C, Cun X, Wang J and Wong T T. 2023. CodeTalker: speech-driven 3D facial animation with discrete motion prior//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 12780-12790 [DOI: 10.1109/CVPR52729.2023.01229]
- Xu C H, Tan T N, Li S, Wang Y H and Zhong C. 2006. Learning effective intrinsic features to boost 3D-based face recognition//Proceedings of the 9th European Conference on Computer Vision. Graz, Austria: Springer: 416-427 [DOI: 10.1007/11744047_32]
- Xu Y L, Wang L Z, Zhao X C, Zhang H W and Liu Y B. 2023. AvatarMAV: fast 3D head avatar reconstruction using motion-aware neural voxels//Proceedings of 2023 ACM SIGGRAPH Conference Proceedings. Los Angeles, USA: ACM: #47 [DOI: 10.1145/3588432.3591567]
- Yang H T, Zhu H, Wang Y R, Huang M K, Shen Q, Yang R G and Cao X. 2020. FaceScape: a large-scale high quality 3D face dataset and detailed riggable 3D face prediction//Proceedings of 2010 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 598-607 [DOI: 10.1109/CVPR42600.2020.00068]
- Ye Y P, Song Z, Guo J G and Qiao Y. 2020. SIAT-3DFE: a high-resolution 3D facial expression dataset. IEEE Access, 8: 48205-48211 [DOI: 10.1109/ACCESS.2020.2979518]
- Yenamandra T, Tewari A, Bernard F, Seidel H P, Elgharib M, Cremers D and Theobalt C. 2021. i3DMM: deep implicit 3D morphable model of human heads//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 12798-12808 [DOI: 10.1109/CVPR46437.2021.01261]
- Yi D, Lei Z, Liao S C and Li S Z. 2014. Learning face representation from scratch [EB/OL]. [2024-06-06]. <https://arxiv.org/pdf/1411.7923.pdf>
- Yi H W, Li C, Cao Q, Shen X Y, Li S, Wang G P and Tai Y W. 2019. MMFace: a multi-metric regression network for unconstrained face reconstruction//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 7655-7664 [DOI: 10.1109/CVPR.2019.00785]
- Yin B C, Sun Y F, Wang C Z and Gai Y. 2009. BJUT-3D large scale 3D face database and information processing. Journal of Computer Research and Development, 46(6): 1009-1018 (尹宝才, 孙艳丰, 王成章, 盖赞. 2009. BJUT-3D 三维人脸数据库及其处理技术. 计算机研究与发展, 46(6): 1009-1018)
- Yin L J, Chen X C, Sun Y, Worm T and Reale M. 2008. A high-resolution 3D dynamic facial expression database//Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition. Amsterdam, Netherlands: IEEE: 1-6 [DOI: 10.1109/AFGR.2008.4813324]
- Yin L J, Wei X Z, Sun Y, Wang J and Rosato M J. 2006. A 3D facial expression database for facial behavior research//Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition. Southampton, UK: IEEE: 211-216 [DOI: 10.1109/FGR.2006.6]
- Zeng X X, Peng X J and Qiao Y. 2019. DF²Net: a dense-fine-finer net-

- work for detailed 3D face reconstruction//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 2315-2324 [DOI: 10.1109/ICCV.2019.00240]
- Zhang L W, Qiu Q W, Lin H Y, Zhang Q X, Shi C, Yang W, Shi Y, Yang S B, Xu L and Yu J Y. 2023a. DreamFace: progressive generation of animatable 3D faces under text guidance. *ACM Transactions on Graphics*, 42(4): #138 [DOI: 10.1145/3592094]
- Zhang L W, Zeng C X, Zhang Q X, Lin H Y, Cao R X, Yang W, Xu L and Yu J Y. 2022a. Video-driven neural physically-based facial asset for production. *ACM Transactions on Graphics*, 41(6): #208 [DOI: 10.1145/3550454.3555445]
- Zhang L W, Zhang Q X, Wu M Y, Yu J Y and Xu L. 2021a. Neural video portrait relighting in real-time via consistency modeling//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 782-792 [DOI: 10.1109/ICCV48922.2021.00084]
- Zhang Y L, Li K P, Li K, Wang L C, Zhong B N and Fu Y. 2018. Image super-resolution using very deep residual channel attention networks//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 294-310 [DOI: 10.1007/978-3-030-01234-2_18]
- Zhang Z H, Liu W, Liu G D, Song L M, Qu Y F, Li X D and Wei Z Z. 2021. Overview of the development and application of 3D vision measurement technology. *Journal of Image and Graphics*, 26(6): 1483-1502 (张宗华, 刘巍, 刘国栋, 宋丽梅, 屈玉福, 李旭东, 魏振忠. 2021. 三维视觉测量技术及应用进展. *中国图象图形学报*, 26(6): 1483-1502) [DOI: 10.11834/jig.200841]
- Zhang Z M, Li L C, Ding Y and Fan C J. 2021b. Flow-guided one-shot talking face generation with a high-resolution audio-visual dataset//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 3660-3669 [DOI: 10.1109/CVPR46437.2021.00366]
- Zhang Z Y, Chen R W, Cao W J, Tai Y and Wang C J. 2023b. Learning neural proto-face field for disentangled 3D face modeling in the wild//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 382-393 [DOI: 10.1109/CVPR52729.2023.00045]
- Zhang Z Y, Ge Y H, Chen R W, Tai Y, Yan Y, Yang J, Wang C J, Li J L and Huang F Y. 2021c. Learning to aggregate and personalize 3D face from in-the-wild photo collection//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 14209-14219 [DOI: 10.1109/CVPR46437.2021.01399]
- Zhang Z Y, Ge Y H, Tai Y, Cao W J, Chen R W, Liu K L, Tang H, Huang X M, Wang C J, Xie Z F and Huang D J. 2022b. Physically-guided disentangled implicit rendering for 3D face modeling//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 20321-20331 [DOI: 10.1109/CVPR52688.2022.01971]
- Zhang Z Y, Ge Y H, Tai Y, Huang X M, Wang C J, Tang H, Huang D J and Xie Z F. 2022c. Learning to restore 3D face from in-the-wild degraded images//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 4227-4237 [DOI: 10.1109/CVPR52688.2022.00420]
- Zheng M W, Yang H Y, Huang D and Chen L M. 2022a. ImFace: a nonlinear 3D morphable face model with implicit neural representations//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 20311-20320 [DOI: 10.1109/CVPR52688.2022.01970]
- Zheng M W, Zhang H Y, Yang H Y and Huang D. 2023. NeuFace: realistic 3D neural face rendering from multi-view images//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 16868-16877 [DOI: 10.1109/CVPR52729.2023.01618]
- Zheng Y F, Abrevaya V F, Buhler M C, Chen X, Black M J and Hilliges O. 2022b. I M avatar: implicit morphable head avatars from videos//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 13535-13545 [DOI: 10.1109/CVPR52688.2022.01318]
- Zhu W B, Wu H T, Chen Z Y, Vedapunt N and Wang B Y. 2020a. ReDA: reinforced differentiable attribute for 3D face reconstruction//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4957-4966 [DOI: 10.1109/CVPR42600.2020.0050]
- Zhu X Y, Lei Z, Liu X M, Shi H L and Li S Z. 2016. Face alignment across large poses: a 3D solution//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 146-155 [DOI: 10.1109/CVPR.2016.23]
- Zhu X Y, Yang F, Huang D, Yu C, Wang H, Guo J Z, Lei Z and Li S Z. 2020b. Beyond 3DMM space: towards fine-grained 3D face reconstruction//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 343-358 [DOI: 10.1007/978-3-030-58598-3_21]
- Zhuang Y Y, Zhu H, Sun X S and Cao X. 2022. MoFaNeRF: morphable facial neural radiance field//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 268-285 [DOI: 10.1007/978-3-031-20062-5_16]
- Zielonka W, Bolkart T and Thies J. 2022. Towards metrical reconstruction of human faces//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 250-269 [DOI: 10.1007/978-3-031-19778-9_15]
- Zielonka W, Bolkart T and Thies J. 2023. Instant volumetric head avatars//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 4574-4584 [DOI: 10.1109/CVPR52729.2023.00444]
- Zollhöfer M, Thies J, Garrido P, Bradley D, Beeler T, Pérez P, Stamminger M, Nießner M and Theobalt C. 2018. State of the art on

monocular 3D face reconstruction, tracking, and applications.
Computer Graphics Forum, 37(2): 523-550 [DOI: 10.1111/cgf.13382]

作者简介

刘菲,女,讲师,主要研究方向为计算机视觉和模式识别。
E-mail: fei.liu@cueb.edu.cn
孙哲南,通信作者,男,研究员,主要研究方向为计算机视觉和模式识别。E-mail: zhenan.sun@ia.ac.cn

张堃博,男,副研究员,主要研究方向为生物特征识别和计算成像。E-mail: kunbo.zhang@ia.ac.cn
杨青,女,讲师,主要研究方向为隐私保护和模式识别。
E-mail: yangqing@cueb.edu.cn
周树波,男,讲师,主要研究方向为计算机视觉和光场成像。
E-mail: zhoushubo@dhu.edu.cn
王云龙,男,副研究员,主要研究方向为生物特征识别和计算机视觉。E-mail: yunlong.wang@ia.ac.cn