软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2024,35(5):2176−2191 [doi: 10.13328/j.cnki.jos.007029]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

结合面部动作单元感知的三维人脸重建算法*

章 毅, 吕嘉仪, 兰 星, 薛 健

(中国科学院大学 工程科学学院, 北京 100049)

通信作者: 薛健, E-mail: xuejian@ucas.ac.cn

摘 要: 三维人脸重建在计算机视觉及动画领域是一项重要任务, 它可以为人脸多模态应用提供三维模型结构和丰富的语义信息. 然而, 单目二维人脸图像缺乏深度信息, 预测的三维模型参数不够可靠, 从而导致重建效果不佳. 提出采用与模型参数高度相关的面部动作单元和人脸关键点作为桥梁, 引导模型相关参数回归, 以解决单目人脸重建的不适定问题. 基于人脸重建的现有数据集, 提供一套完整的面部动作单元半自动标注方案, 并构建 300W-LP-AU 数据集. 进而提出一种结合动作单元感知的三维人脸重建算法. 该算法实现端到端的多任务学习, 有效降低了整体训练难度. 实验结果表明, 该算法能有效地提升三维人脸重建性能, 重建的人脸模型具有更高的保真度.

关键词: 面部动作单元; 人脸关键点; 三维人脸重建

中图法分类号: TP18

中文引用格式: 章毅, 吕嘉仪, 兰星, 薛健. 结合面部动作单元感知的三维人脸重建算法. 软件学报, 2024, 35(5): 2176–2191. http://www.jos.org.cn/1000-9825/7029.htm

英文引用格式: Zhang Y, Lü JY, Lan X, Xue J. AU-aware Algorithm for 3D Facial Reconstruction. Ruan Jian Xue Bao/Journal of Software, 2024, 35(5): 2176–2191 (in Chinese). http://www.jos.org.cn/1000-9825/7029.htm

AU-aware Algorithm for 3D Facial Reconstruction

ZHANG Yi, LÜ Jia-Yi, LAN Xing, XUE Jian

(School of Engineering Science, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: As a critical task in computer vision and animation, facial reconstruction can provide 3D model structures and rich semantic information for multi-modal facial applications. However, monocular 2D facial images lack depth information and the parameters of the predicted facial model are not reliable, which causes poor reconstruction results. This study proposes to employ facial action unit (AU) and facial keypoints which are highly correlated with model parameters as a bridge to guide the regression of model-related parameters and thus solve the ill-posed monocular facial reconstruction. Based on existing facial reconstruction datasets, this study provides a complete semi-automatic labeling scheme for facial AUs and constructs a 300W-LP-AU dataset. Furthermore, a 3D facial reconstruction algorithm based on AU awareness is put forward to realize end-to-end multi-tasking learning and reduce the overall training difficulty. Experimental results show that it improves the facial reconstruction performance, with high fidelity of the reconstructed facial model.

Key words: facial action unit; facial key point; 3D facial reconstruction

三维人脸重建是计算机视觉领域的重要任务, 它的目标是从人脸图像中推断出对应的三维人脸模型, 以实现对人脸更全面的分析和处理. 它可以应用在人脸多模态任务中, 例如: 人脸语音驱动、虚拟主播、虚拟化妆等. 相较于二维人脸相关任务, 它可以提供更为丰富的信息, 也可以支持更为广泛的应用场景. 但相对地, 其算法设计也更具有挑战性. 随着深度学习的发展, 三维人脸重建算法取得了持续突破, 仅依靠单张或多张二维 RGB 图像就能

* 基金项目: 国家自然科学基金 (62236006, 62032022, 61972375); 北京市自然科学基金 (4222040)
章毅和吕嘉仪为共同第一作者.
本文由“多模态协同感知与融合技术”专题特约编辑孙立峰教授、宋新航副研究员、蒋树强教授、王莉莉教授、申恒涛教授推荐.
收稿时间: 2023-04-10; 修改时间: 2023-06-08, 2023-08-16; 采用时间: 2023-08-23; jos 在线出版时间: 2023-09-11
CNKI 网络首发时间: 2023-12-14

获取较为可靠、稳定的三维人脸模型.

目前, 三维人脸重建的研究路线主要分为基于参数化模型和基于非参数化模型两大类, 基于参数化模型的方法通过解决非线性优化问题或使用卷积神经网络即可直接回归模型参数[1], 实现了高保真的面部形状重建和准确的头部姿态估计, 其性能和实用性优于非参数模型的方法, 所以本文主要对参数化模型的方法进行研究. 现阶段人脸重建任务可以分为多目图像重建和单目图像重建两类, 多目图像重建可以从多幅二维图像中提取特征点的多视角信息, 然后通过匹配特征点来进行约束, 最后使用三角测量的方法推理出所有特征点的三维信息. 而对于单目图像重建, 由于从单目二维图形重建三维人脸模型缺少充足的深度信息, 其本身是病态且不可靠. 之前的单目三维重建方法主要是通过编码器直接推理出相应的三维特征点, 然后通过数据集的三维标注信息进行监督, 但该类方法很难对三维信息进行准确预测, 其精度有限.

人脸参数化模型 3DMM (3D morphable model)[1] 由形状 (shape)、姿态 (pose)、表情 (expression) 这 3 类系数组成, 本文认为通过利用与这些参数强相关且能够准确预测的任务, 可以有效地解决上述基于单目重建方法的不适定问题, 并进一步指导模型参数的回归. 人脸在拓扑结构上存在相似性, 都是由双眼、鼻子、嘴巴、双耳、眉毛组成的, 这些特征对重建人脸模型提供可靠的依据. 具体来说, 本文认为人脸特征点与参数化模型中姿态和形状系数是强相关的, 因为它们都高度反映了头部的旋转角度与脸部轮廓; 而面部动作单元信息与表情系数存在强相关关系, 因为它们都是对面部表情的描述. 因此, 本文将人脸特征点、面部动作单元任务作为辅助, 隐式约束参数模型的系数回归, 缓解单目重建不适定问题, 从而实现更准确的头部重建任务.

本文首先对人脸重建数据集的人脸数据进行面部动作单元标注, 提供了一套完整的半自动标注方案并构建了 300W-LP 的面部动作单元数据集 300W-LP-AU. 进而, 本文提出了一种基于动作单元感知的人脸重建算法 (AU-aware face reconstruction, AUFR), 该算法包括动作单元感知模块和人脸重建模块. 动作单元感知模块能够感知面部动作单元激活的概率, 并编码提供显式特征输入以辅助 3DMM 参数的回归. 人脸重建模块基于骨干网络的输出特征和面部动作单元编码特征回归 3DMM 参数. 此外, 由于 3DMM 的形状参数和姿态参数与三维人脸关键点位置密切相关, 该模块还将通过回归得到的 3DMM 参数重建出人脸关键点, 采用真实 3DMM 标签重建出的人脸关键点进行监督进而约束 3DMM 参数. 整个算法基于面部动作单元和生成的人脸关键点实现了端到端的三维人脸重建. 值得注意的是, 该算法是第 1 个将面部动作单元与人脸重建结合的工作, 为该方向的研究提供了一条新的思路. 此外, 基于预测的 3DMM 参数能够准确完成 3D 关键点检测、头部姿态估计等任务.

本文通过在 AFLW2000-3D 评测集上进行一系列实验验证了所提出的模块对性能和精度的提升, 实验包括人脸对齐、头部姿态旋转、三维人脸重建等任务. 实验结果证明, 与其他相关工作相比, 本文提出的算法在各项任务上都有出色的表现. 图 1 是 AUFR 可视化的效果图, 原图来源: https://www.pexels.com/zh-cn/photo/8190816/

(a) 原图 (b) 关键点 (c) 头部姿态 (d) 人脸重建

图 1 算法可视化效果图

本文的主要贡献如下.

(1) 本文指出采用面部动作单元和人脸关键点作为桥接二维人脸图像与 3DMM 参数的显式特征, 可以缓解基于单目图像进行三维人脸重建的不适定问题. 同时, 本文还提出一个端到端的多任务学习算法, 采用面部动作单元和人脸关键点检测作为辅助任务引导 3DMM 参数的回归, 可以有效降低整体训练难度. 该方法是第一个将面部动作单元与三维人脸重建相结合的工作.

(2) 本文提供了一套 AU 的半自动标注方案, 并构建了 300W-LP-AU 数据集, 该数据集囊括 63 万张人脸图像, 并对 41 个面部动作单元进行标注. 该数据集将有助于人脸重建方向研究新的算法和模型.

(3) 本文提出的算法在各项任务上都有出色的表现, 同时具备实时性能. 经实验测试, 该算法在 AMD 3950X CPU 设备上的输出帧率可达 100 f/s. 极低的输出延迟使得该算法可以在各类边缘计算或者是移动端设备中快速运行, 为各类人脸重建相关系统的落地应用提供可能.

1 人脸检测及三维人脸重建相关算法

本节将详细的介绍面部动作单元检测、三维人脸重建、三维人脸关键点检测和头部姿态估计相关算法. 这些算法对于提高人脸特征点检测精度、头部姿态回归精度以及提升三维人脸重建效果具有重要意义. 本文的研究旨在利用三维形状、表情和旋转特征来重建三维人脸, 同时本文认为表情特征与面部动作单元检测任务高度相关, 形状特征和旋转特征与三维特征点检测任务密切相关, 而旋转特征与头部姿态回归任务有密切联系. 因此, 本文将探索与上述任务相关的算法, 提供一种综合性的解决方案.

1.1 面部动作单元检测

在研究面部动作单元 (action unit, AU) 相关任务时, 科研工作者会把人脸对称性、AU 之间的协同和互斥关系等先验信息融入深度模型中, 通常采用图卷积神经网络或协作特征学习框架来实现, 以提高 AU 的检测效果和强度回归精度. 近年来, Luo 等人[2]提出了一种基于多维边缘特征的图神经网络方法, 描述面部表示和 AU 之间的关系. 该方法首先将每个 AU 的激活状态与其他 AU 的关联关系编码为节点特征, 然后, 学习一对多维度的边缘特征以描述 AU 之间的多种特定任务关系. 在节点特征学习和边缘特征学习过程中, 该方法将完整的人脸表示作为输入, 以考虑面部表情对 AU 关系的影响. 根据实验结果显示, Luo 等人[2]提出的节点边缘特征学习模块对于 BP4D 和 DISFA 数据集来说非常有效, 能够显著提升基于 CNN 和 Transformer 骨干网络的表现, 达到了当前领先水平. 在第 3 节中, 本文引入 Luo 等人[2]的工作为面部重建提供 AU 监督信号.

1.2 三维人脸重建

近年来, 针对三维人脸重建, 基于参数化模型的方法和基于非参数化模型方法都得到了广泛的研究和应用. 明暗恢复形状算法 (shape from shading, SFS)[3]是一种利用图像中的明暗变化来重建形状的计算机视觉方法. 该方法能够从给定的光照系数和反射率图中恢复出精细的形状细节. 但是, 该方法存在着凸凹歧义性的问题[4], 即同一个三维物体在图像上表现出不同的凸凹性, 这是由于光照方向的变化所导致的. 为了处理这种歧义性, 需先给定一个初始化的三维形状[5]. 人脸参数化模型 3DMM (3D morphable model) 是一种基于模型的三维人脸重建方法, 它通过建立人脸形状和纹理之间的映射关系, 将人脸的三维形状和纹理参数分别表示为低维空间的线性组合, 从而实现了对人脸的高效重建. 相比之下, SFS 方法存在着凹凸歧义性的问题, 对于光照条件的要求也比较严格, 对于复杂的光照环境和阴影效果很难进行准确的恢复, 计算成本相对较高. 而 3DMM 的优势在于它可以对人脸的表情、姿态等细节进行建模, 而且建模过程相对简单, 仅需要少量的训练数据即可得到一个较为准确的人脸模型. 因此, 3DMM 相对于 SFS 在人脸重建领域具有更广泛的适用性和更高的精度.

早期, Blanz 等人[6]提出 3DMM 可以通过使用 PCA (principal component analysis) 得到的正交基的线性组合来表示人脸模型, 用这种方式进行人脸重建将问题转化为 3DMM 的参数回归问题. 随着深度学习的发展, 人们将重心转移至使用深度卷积神经网络来学习 3DMM 参数[7–13], 这些方法比求解非线性优化函数的传统方法[14–18]更快更准确.

即便如此, 由于训练数据分布不平衡等原因, 3DMM 模型在面部表情表达方面能力有限. 为解决这个问题, Zhao 等人[19]设计了一种以网格编辑方法为基础的表情合成器, 并提出了通过利用面部形状和表情合成的方法, 从面部表情多样性不足的不平衡数据集中学习线性 3DMM 模型. 为了提升重建性能, Chen 等人[20]将 3DMM 回归和 3D 集合回归相结合, 提出使用基于 3DMM 的粗略模型和 UV 空间中的位移贴图来表示三维人脸的框架, 用于从单幅图像中重建细粒度三维人脸. Wu 等人[21]认为从单目二维图像直接提取三维信息是不鲁棒的, 他们提出多

属性特征融合提取细粒度的三维人脸特征点, 然后基于三维特征点反推一维 3DMM 信息的方法. Jung 等人[22]认为 3DMM 参数模型的表达能力在一定程度上受限于固定数量的三维扫描点以及其全局的线性计算, 他们提出了一种通过自由形变 (free-form deformation, FFD) 重建三维人脸的方法. 该方法解决了部分参数化模型表达受限的问题, 但从结果上看, 该工作在舍弃 3DMM 这种模型表达方式时, 也同时失去了部分三维模型处理病态问题的能力. 为了补充数据集的样本数量, 上述大多数有监督的方法会通过数据增广来扩大数据集以提高模型的训练强度, Chen 等[23]则提出了一种条件生成网络 (cGAN) 和三维人脸重建网络相结合的模式来解决样本少的问题, 该工作可以有效利用含标签以及不含标签的人脸图像数据进行训练, 有效提高模型的精度及泛化能力. 本文则引入了 AU 信息和人脸关键点, 以解决面部表情表达受限和难以提高人脸重建能力的问题.

1.3 三维人脸关键点检测

三维人脸关键点检测任务是指通过计算机视觉技术或深度学习算法对图像中的人脸进行关键点检测和三维重建, 以获得准确的人脸形状和姿态信息. 该任务需要从输入的人脸图像中提取包括眼睛、鼻尖、嘴巴和下巴中心等部分的关键点三维位置以及头部角度的估计. 三维人脸关键点可用于人脸识别、面部表情分析和虚拟现实等应用领域. 传统方法通过建立二维信息和三维特征点之间的映射关系, 使用线性回归计算三维特征点, 但这需要大量人为输入二维的人脸信息. 基于深度学习的方法使用 ResNet、HRNet 等深度学习模型将人脸图像作为输入, 输出人脸关键点的坐标或人脸姿态的参数. 而三维人脸关键点回归任务需要输出三维坐标, 因此需要采用特殊的技术进行预测, 如: 从二维人脸关键点开始进行三维人脸形状的估计或使用多视角图像进行三维人脸重建等. 3DDFA[24,25]是一种有代表性的人脸重建方法, 它使用 BFM (basel face model) 和 3DMM 从单目图像重建人脸网格. 但是, 该方法对于各个 3DMM 参数的波动非常敏感, 难以达到高精度. 相比之下, PRNet[26]通过预测编码三维点的二维 UV 位置图, 并使用 BFM 网格连接建立人脸模型, 由于其三维点不是来自 3DMM 参数, 因此具有更高的网格变形能力, 但是更难获得平滑且可靠的网格. 3DDFA-V2[27]在 3DDFA 的基础上进一步提出了一种元联合优化策略和一种短视频合成方法, 通过顶点索引提取出三维人脸和粗糙的三维关键点. 与此不同的是, 本文的方法先聚合多属性特征得到粗糙的三维关键点, 再进行细化, 最后从三维关键点和其他特征回归出 3DMM 参数.

1.4 头部姿态估计

头部姿态估计在人机交互方面可以应用在人机对话、用户认证、增强现实、驾驶员辅助等领域. Deep Head Pose[28]使用深度网络同时预测二维人脸特征点和面部方向. HopeNet[29]训练一个多损失卷积神经网络来确定姿势, 通过联合分档的姿势分类和回归, 直接从图像强度预测内在欧拉角, Hsu 等人提出的 QuatNet[30]设计了 L_2 回归损失与序数回归损失相结合的多回归损失函数, 用于训练卷积神经网络, 从没有深度信息的 RGB 图像中估计头部姿势. Yang 等人提出的 FSA-Net[31]设计了一个基于回归和特征聚合的方法, 为聚合特征构建了一个细粒度的结构映射, 并在聚合前对特征进行空间分组. Cao 等人提出了 TriNet[32]模型用矢量的表示进行姿态估计, 并使用矢量的平均绝对误差 (MAEV) 来评估性能. Hempel 等人[33]提出了一种无需关键点的头部姿态估计的方法, 他们使用旋转矩阵来回归准确的头部姿态, 这样的方法可以在一定程度上解决关键点歧义导致的头部姿态不准确的问题. 本工作与其相似之处是使用参数化的模型来避免直接回归关键点带来的歧义.

2 基础知识

本文所提方法主要基于人脸特征点和 AU 表情系数的显式特征, 将其作为监督信号实现人脸头部的 3DMM 模型重建. 下面介绍相关概念和基本理论.

2.1 面部动作单元 (AU)

1978 年, 美国心理学家 Paul Ekman 将解剖学中的面部肌肉运动和人脸表情联系起来, 提出了人脸运动编码系统 (facial action coding system, FACS)[34], 该系统是当今最全面、应用最广泛的面部表情描述系统. 如图 2 所示, FACS 定义了 44 个面部动作单元 (AU), 并具体定义了每个 AU 的运动区域, 运动特征及各种表情的 AU 构成[35].

人脸的各种表情最终都能对应分解到各个 AU 所表示的运动上, 也就是说, 通过 AU 间的相互编码组合可以

生成任意表情. 不做任何面部动作的人脸表情, 称为中立表情. AU 可以被单独激活, 也可以被组合激活, 例如: 只做闭眼的表情, 就只有 AU43 被激活; 微笑表情可以被描述为嘴角拉升 (AU12) 和脸颊提升 (AU6) 的组合. 因此, 可以通过面部动作单元序列表示表情和面部运动. 近年来有关 AU 检测和强度回归的研究也日益增多[36–39], 本文引用了有关 AU 检测的工作以制作 AU 标签和辅助三维头部重建.

图 2 动作编码系统定义的部分 AU[37]

2.2 三维可变形模型

三维可变形模型 3DMM[1]是一种可以用参数表达人脸形状和纹理特征的三维模型, 该模型通过不同的参数组合, 可以重建出对应的人脸. 目前, 基于深度学习的 3DMM 方法被广泛用于三维人脸重建任务, 同时 3DMM 相关数据集也在不断扩充. 其中, 由 Parscal 等人[1]于 2009 年提出的 BFM 数据集是最为经典的数据集. 他们使用激光扫描仪采集了 200 人的精确三维人脸数据, 每个人脸模型约包含 5.3 万个点. 随后, BFM 模型于 2017 年得到了扩充, 并增加了表情参数以提高其表达能力.

同一种 3DMM 模型输入的参数数量和种类是一致的, 所以同一种 3DMM 模型拥有相同数量的顶点和三角面片, 而且相同序列号表示的顶点也具有相同的语义信息. 现在 3DMM 模型常用于通过二维人脸重建三维头部外观的工作. 3DMM 模型根据不同的数据集或者是应用场景会有些许的区别, 包括但不限于模型参数的类型和数量、顶点数量、重建三维人脸的计算方式等. 这里以 BFM 模型为例, 其计算方式如公式 (1) 所示.

$$S = \bar{S} + A_s\Psi_s + A_e\Psi_e \tag{1}$$

其中, A_s 是三维形状 (shape) 的基础常量, Ψ_s 是形状估计参数, 也就是输入的形状参数; A_e 和 Ψ_e 则是表情 (expression) 的基础常量及估计参数; $\bar{S}$ 是该模型定义的平均人脸, S 则是通过前面所有参数计算得到的三维人脸模型. 本文也采用该模型, 但与之不同的是, 本文依照论文 3DDFA-V2[27]的方法将原 199 维的形状参数减少到 40 维, 29 维的表情参数减少到 10 维, 通过上述参数即可算出所需的三维人脸特征点. 此外, 本文还计算了头部姿态参数 Ψ_p 来估计头部的旋转偏移量.

3 本文方法

3.1 300W-LP-AU 数据集

面部动作单元 (AU) 作为一种显式的二维面部特征与 3DMM 中表情参数密切相关, 能够反映面部表情的相关信息. 因此, 本文将 AU 检测作为辅助任务引入三维人脸重建中. 近年来, ME-GraphAU[2]方法在 AU 检测和强度回归任务中性能较佳. 该方法中的 OpenGraphAU 模块可以检测出 41 种 AU 是否被激活, 并且具有良好的泛化性. 因此, 本文将其引入以生成 AU 标签. 基于 300W-LP 数据集的 AU 标签整体制作流程如图 3 所示, 具体分为以下几个步骤.

第 1 步, 输入图像准备: 本文参考 Guo 等人[27]的方法对 300W-LP 数据集进行预处理. 他们通过旋转、遮挡和翻转等数据增强方法将原始数据集扩充到 68 万张图片. 因此, 需找到数据增强前的图片作为输入图像, 共计 3 837 张. 一般 AU 标签与数据增强方式不会产生冲突, 因为无论图像如何变化, AU 的语义信息保持不变. 所以, 每张图

片只需进行一次 AU 检测任务. 但是水平翻转的数据增强会导致 AU 的语义发生改变, 因此对该情况需要单独处理.

第 2 步, 图像预处理: 本文使用与 OpenGraphAU 相同的人脸检测器 MTCNN[40]进行人脸定位, 以确保 AU 标签的准确性. 首先, 使用 MTCNN 检测器定位人脸存在两种可能. 如果无法找到人脸, 则进行手工标注. 如果检测到多个人脸, 则需要选择最佳有效人脸框. 通常情况下被检人脸靠近图像中心位置, 且人脸框面积较大, 具有较高置信度. 如图 4 所示, 为防止误检, 本文采取以下处理: 首先, 计算所有人脸框的面积并归一化处理, 计算人脸框中心点坐标与图像中心点之间的欧氏距离, 并获取人脸框置信度. 然后, 根据中心点和面积对人脸框进行排序. 进而, 判断面积最大的人脸框是否靠近图像中心. 如果靠近中心位置的人脸框占比大于整个图像的 15% 且置信度大于 95%, 则选择靠近图像中心的人脸框; 否则, 进行人工核实并标注. 选定人脸框后, 根据最大边长的 1.2 倍重新划定一个正方形的检测框, 以获得更完整的人脸框.

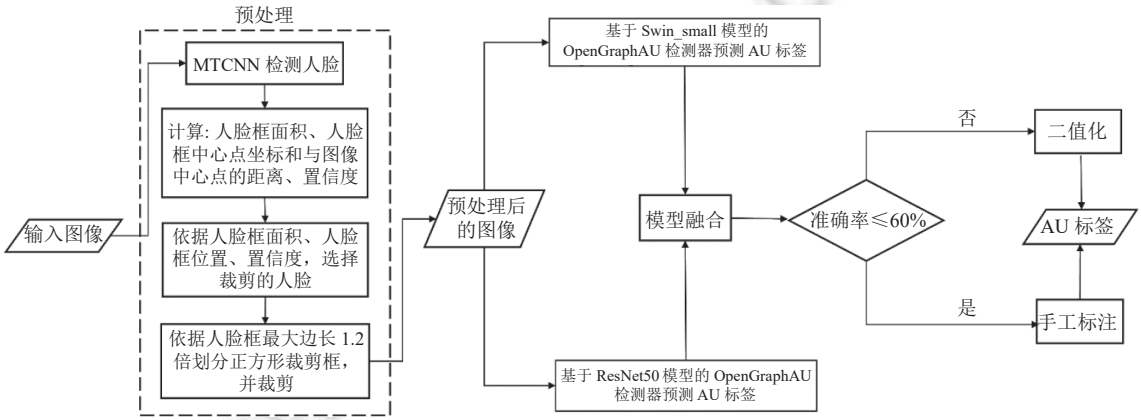

图 3 AU 标签制作流程

图 4 图像预处理过程

第 3 步, AU 标签预测: 本文采用 Swin-Small 和 ResNet50 的基线模型, 分别把经过预处理的图像送入 OpenGraphAU 模块进行 AU 标签检测, 而后, 根据每个 AU 的激活概率制作 41 个 AU 标签. 考虑到图像中被激活的 AU 应与 3DMM 中表情参数密切相关, 所以本文把 AU 的激活概率视为监督信号.

第 4 步, 模型预测融合: 由于 300W-LP 数据集中的部分图像存在光照、角度和遮挡等问题, 导致少数 AU 标

签的预测结果存在误差. 因此, 本文采用投票法将两个模型的预测结果进行融合, 以得到更鲁棒的结果. 若两个模型对于同一个 AU 的激活概率预测结果之间的差距在 40% 以内, 则认为该 AU 的预测准确, 并将两个模型预测结果的平均值作为最终的融合结果. 如果差距较大, 则人工核实并标注. 进而, 计算融合结果与预测标签之间的准确率. 如果图像中有 25 个以上 AU 预测结果准确 (准确率达到 60%), 则认为该图像的预测准确. 相反, 则需要进行手工标注.

第 5 步, AU 标签生成: 在获得最终的预测结果后, 将 AU 的激活概率进行二值化处理, 其中 0 表示 AU 未被激活, 1 表示 AU 被激活. 同时, 将经过数据增强的图像与原始图像的标签进行匹配.

图 4 展示了第 2 步图像预处理的效果. 输入数据可以分为两类: 正常图片和水平翻转后的图片. 观察可发现, 水平翻转后的图像会导致左右对称的 AU 的语义信息发生变化. 图 4 中第 1 行是原始输入图像, 第 2 行展示了人脸框分割的可视化效果. 其中, 浅蓝色表示人脸裁剪框, 红色表示有效人脸, 深蓝色表示 MTCNN 模型检测到的其他人脸. 第 3 行展示了裁剪后的结果, 该步骤能够正确裁剪出多人照片中的有效人脸, 避免生成无效样本.

本文随机抽取 10% 标准结果并进行验证, 正确率达到 98.3%. 图 5 展示了 AU 标签生成的效果, 其结果为: 上眼睑提升、脸颊向上、眼睑收缩、上嘴唇向上、嘴角向上和嘴唇分开的 AU 被激活了, 通过观察预测结果与图片所示结果相符.

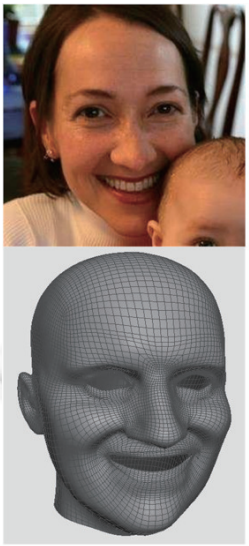

图 5 AU 标签生成效果图

3.2 基于动作单元感知的三维人脸重建算法

本节提出一种基于动作单元感知的人脸重建算法 (AUFR), 其整体流程如图 6 所示. 该算法由动作单元感知模块和人脸重建模块构成. 动作单元感知模块能够感知面部动作单元激活的概率, 并编码提供显式特征以辅助 3DMM 参数的回归. 人脸重建模块基于骨干网络的输出特征和面部动作单元编码特征回归 3DMM 参数. 此外, 该模块在生成 3DMM 参数后重建出三维人脸关键点, 通过真实的人脸关键点与 3DMM 参数进行监督约束最终的 3DMM 参数回归. 整个算法基于面部动作单元和生成的人脸关键点实现了端到端的三维人脸重建.

动作单元感知模块如图 6 的 AU aware module 所示, 输入单目人脸图像到主干网络得到网络特征, 并采用多层感知机网络作为 AU 部分的解码器, 输出面部动作单元参数 $\Upsilon \in R^{41}$, 包含 41 种面部动作单元的激活概率. 此处采用 BCE 损失函数监督面部动作单元参数的回归, 如公式 (2) 所示. 进而, 面部动作单元序列在可学习的 82 维状

态表 (每个单元是否激活, 两种状态) 进行嵌入式编码得到 AU 特征 $F_{\mathrm{AU}} \in R^{41\times 16}$, 并与之前的网络特征融合, 输入之后的人脸重建模块.

$$L_{\mathrm{AU}} = -\frac{1}{n}\sum_{i=1}^{n}(\Upsilon_i^*\log(\Upsilon_i)) + (1-\Upsilon_i^*)\log(1-\Upsilon_i)) \tag{2}$$

其中, n 代表每次迭代的批量大小, Υ_i, Υ_i^* 分别代表第 i 个人脸图片对应 AU 的预测值和真实参数.

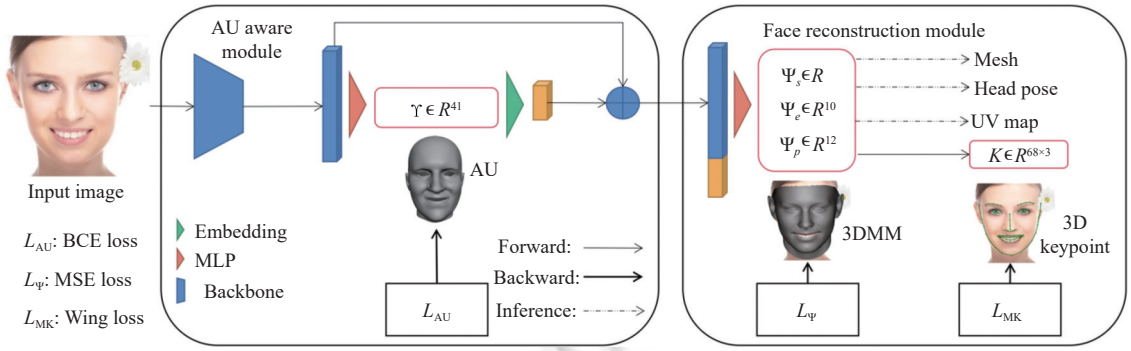

图 6 基于动作单元感知的人脸重建算法的整体流程

人脸重建模块如图 6 的 face reconstruction module 所示, 在得到融合的 AU 特征后, 将其输入 3DMM 参数的解码器中, 预测 62 维的 3DMM 参数 $\Psi \in \{\Psi_s, \Psi_e, \Psi_p\}$, 分别用来表示形状特征 (shape, 40 维)、表情特征 (exp, 10 维) 以及姿态特征 (pose, 12 维). 由于 3DMM 参数在连续空间下自由分布, 此处采用 MSE 损失函数监督参数的回归.

$$L_{\Psi} = \frac{1}{n}\sum_{i=1}^{n}(\|\Psi_{s_i} - \Psi_{s_i}^*\|_2^2 + \|\Psi_{e_i} - \Psi_{e_i}^*\|_2^2 + \|\Psi_{p_i} - \Psi_{p_i}^*\|_2^2) \tag{3}$$

其中, Ψ_{s_i}, $\Psi_{s_i}^*$ 代表第 i 张图片对应的 3DMM 形状参数的预测值和真实值, Ψ_{e_i}, $\Psi_{e_i}^*$ 和 Ψ_{p_i}, $\Psi_{p_i}^*$ 同理. 在回归 3DMM 参数之后重建三维人脸模型, 再计算出旋转矩阵, 可以重建出稠密的三维人脸关键点. 本文对其中几何特征明显的 68 个点进行监督, 其表示为 $K \in R^{68\times 3}$. 对关键点进行监督如公式 (4) 所示.

$$L_{\mathrm{MK}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} L_{\mathrm{Wing}}(K_{ij}, K_{ij}^*) \tag{4}$$

其中, k 指代的是关键点的数量, K_{ij}, K_{ij}^* 分别由 Ψ, Ψ^* 生成, 对应第 i 张图片中第 j 个三维关键点的预测坐标和真实坐标. W 表示的是 Wing loss 损失函数, 它是由 Feng 等人[41]在 2018 年提出的, 是深度学习中一种针对回归问题的损失函数, 常用于关键点检测. 该损失函数通过平衡误差的上下限减缓训练过程中的过拟合, 如公式 (5) 所示.

$$W(x,y) = \begin{cases} \omega \times \log\left(1+\dfrac{d}{\varepsilon}\right), & \text{if } d < \omega \\ d - C, & \text{otherwise} \end{cases} \tag{5}$$

其中, d 表示预测值 x 和真实值 y 之间的欧氏距离, ε 是较小的正数, 用于保证分母不为零, 避免梯度爆炸问题. ω 和 C 是分别用于控制两个分段函数的平滑度和偏移量的超参数. 在预测值和真实值之间的误差较小时, Wing loss 的第 1 个分段函数将使误差的惩罚更加平滑. 当误差超过阈值时, 第 2 个分段函数将加重误差的惩罚. 通过使用这种损失函数, 可以使模型在训练过程中更加稳定并减缓过拟合.

最后, 本文使用下面的总损失公式来监督整个训练过程.

$$L = \lambda_1 L_{\Psi} + \lambda_2 L_{\mathrm{AU}} + \lambda_3 L_{\mathrm{MK}} \tag{6}$$

其中, λ_1 对应 3DMM 参数的损失权重, λ_2 是面部动作单元的损失权重, λ_3 代表三维关键点的损失权重. 各参数之间是自由无约束的, 其中 3DMM 参数回归作为主要任务, 固定 $\lambda_1 = 1.0$, 调整其他超参, 详细设置参见第 4.5 节.

4 实验结果与分析

4.1 实验数据

本文在公开的人脸数据集上进行实验, 即有 AU 标签的 300W-LP 数据集和 AFLW2000-3D, 以完成训练和效果评估. 表 1 给出了所用数据集的详细信息.

表 1 实验数据集参数

数据集	分类	标签	样本数量 (张)	数据增广
300W-LP	训练集	面部动作单元 3DMM参数 (BMF) 三维人脸特征点	636252	旋转 (离线) 翻转 (离线) 色彩变换 (亮度0.4, 对比度0.4, 饱和度0.4)
AFLW2000-3D	测试集	三维人脸特征点 头部旋转角 3DMM参数 (BMF)	2000	无

本文遵循之前的工作[42]将 300W-LP 数据集[25]作为训练集, 该数据集被广泛应用于各类三维人脸任务中. 300W-LP 收集大量自然人脸图片并对其做了平面旋转的数据增强操作, 同时对人脸轮廓进行分析以生成 BFM 模型的 3DMM 参数, 该数据集共有 122450 张面部图片和相应的 3DMM 参数. 在此基础上, 本文提取 300W-LP 作为数据增强的原始数据, 参考 Guo 等人[27]数据处理的方式, 对数据集进一步采用随机颜色抖动、随机翻转等数据增强操作, 同时融入 41 个用于监督的 AU 标签, 共获得 687854 组训练数据. AFLW2000-3D[25]数据集作为测试集, 它包括 AFLW[43]的前 2000 张图片以及相应的 3DMM 标签和 68 个人脸特征点的三维标注. AFLW2000-3D 有两套特征点标注方案, 分别为原始标注版本和 LS3D-W 重新标注的高精度版本[44]. 此外, AFLW2000-3D 还包含了头部姿态的标签.

4.2 评价指标

本文使用标准化平均误差 (*NME*) 评价人脸关键点的准确度. 定义如公式 (7) 所示, *NME* 为关键点预测值和真实值之间欧氏距离的平均值, 除以标准因子得到的结果. *NME* 的值越小表示误差越小, 算法的准确度越高.

$$NME=\frac{1}{N}\sum_{n=1}^{N}\frac{\left\|k_n-k_n^*\right\|_2}{d} \tag{7}$$

其中, k_n 和 k_n^* 是三维人脸特征点的预测值和真实值, N 是样本的数量, d 作为每张人脸图像的归一化项是人脸边界框大小. 在三维人脸重建任务中, 本文对大于 45K 的稠密顶点进行评估. 在三维人脸关键点检测任务中, 对 68 个关键点进行评估.

在头部姿态估计任务中, 本文在 AFLW2000-3D 数据集使用每张人脸的头部姿态真实角度作为标准计算预测角度的平均绝对误差 (*MAE*), 如公式 (8) 所示.

$$MAE=\frac{1}{N}\sum_{n=1}^{N}|v_n-v_n^*| \tag{8}$$

其中, v_n 和 v_n^* 是预测的人脸头部姿态角度的预测值和真实值, N 表示的是样本总量.

4.3 实现细节

该实验在 Ubuntu 18.04 采用 NVIDIA RTX 3090 GPU 对模型进行训练. 在训练过程中, 将批处理大小设置为 768, 初始学习率为 0.001, 并选择了动量为 0.9 的 Adam 优化器. 其中, 学习率策略采用在第 20、35、50 epoch 处下降学习率, 每次缩减至原来的 20%. 总共训练 60 个 epoch. 在测试阶段, 将经过主干网络预测的 3DMM 参数作为最终输出, 然后基于 3DMM 参数恢复出的三维关键点、头部姿态参数、人脸重建模型进行评估. 为验证算法的鲁棒性, 本文分别采用 MobileNet-V2[45]和 ResNet50[46]作为主干网络. 本文算法在采用 MobileNet-V2 作为骨干网络情况下对三维人脸重建的速度在 NVIDIA RTX 3090 GPU 上达到了 300 f/s, 在 AMD 3950X CPU 上达到了 100 f/s. 通过广泛的实验验证, 本文将损失函数的权重 $[\lambda_1,\lambda_2,\lambda_3]$ 设置为 $[1.0,0.01,0.2]$.

4.4 消融实验

本节介绍了在 AFLW2000-3D 数据集上针对人脸对齐的消融实验, 旨在验证本文方法动作单元感知模块和人脸重建模块的作用及超参权重的影响. 此处使用 *NME* 分析每个模块对最终性能的提升效果, 具体结果如表 2 示. 其中, "√"表示该模块使用对应的损失函数, 第 1 行展示仅使用骨干网络 MobileNet-V2 的简单基线模型结果, 只通过图像回归 3DMM 参数来进行预测. 本文固定 $\lambda_1 = 1.0$ 调整其他超参数大小, 其中 λ_2, λ_3 分别为动作单元感知模块和人脸重建模块的权重. 表 2 关于 AU 损失权重参数与三维关键点损失权重参数的消融实验.

表 2 消融实验结果

(λ_2, λ_3)	AU	3D Landmarks	*NME*
(0, 0)	—	—	3.912
(0.01, 0)	√	—	3.751
(0.1, 0)			3.632
(0, 0.1)	—	√	3.576
(0, 0.2)			3.589
(0, 0.3)			3.628
(0.01, 0.1)	√	√	3.524
(0.01, 0.2)			3.507
(0.1, 0.1)			3.621
(0.1, 0.2)			3.686

表 2 中显示了动作单元感知模块和人脸重建模块对最终性能的贡献. 动作单元感知模块可用于表示人脸面部表情信息, 并向人脸重建模块提供监督信号. 人脸重建模块可将多属性的二维和三维信息融合, 并回归人脸特征点的坐标. 人脸特征点和 AU 表示了旋转、形状和表情信息等 3DMM 的要素. 在表 2 中可以发现, 单独使用面部表情单元或人脸关键点对人脸重建都有帮助. 而同时 3DMM 与之相关的要素并不一样, 因此同时使用这两个模块作为辅助, 总体性能可以进一步提升.

另外, 本文考虑了 3DMM、AU 和三维关键点 3 种损失的权重, 本文固定 3DMM 的权重为 1.0 不变, 然后大批量进行实验发现当两个模块损失权重取 (0.01, 0.2) 时效果最好, 同时本文将这组参数应用到其他骨干网络的实验上. 因为 AU 检测不直接使用它的概率值, 而是根据最终的 AU 状态标签使用嵌入的状态表, 所以不需太强的监督. 而关键点是从 3DMM 中根据世界坐标系变化恢复出来, 与 3DMM 参数直接相关, 因此权重比 AU 的权重要大.

4.5 对比实验结果与分析

4.5.1 三维人脸关键点检测

在三维人脸关键点检测任务的评测中, 本文使用了 AFLW2000-3D 数据集作为测试集. 原始版本的评价结果列在后文表 3 中, 并与其他相关工作进行了比较. 本文方法基于 ResNet50 版本的 *NME* 为 3.46, 优于基于 MobileNet-V2 的版本. 同时, 与其他方法相比, 本文方法的表现更为出色, 尤其是航向角 (yaw) 在 [0, 30°) 和 [60°, 90°] 区间的样本在 *NME* 分别达到 2.60, 4.37. 此外, 本文还在重新标注的版本上进行了测试和比较, 其结果如后文表 4 所示. 在重新标注版本上基于 MobileNet-V2 版本的 *NME* 为 2.83, 优于 ResNet50 版本的 2.94.

4.5.2 头部姿态估计

本工作在测试集 AFLW2000-3D 进行了头部姿态性能的评测, 评价指标是预测欧拉角与真值之间的平均绝对误差 (*MAE*). 后文表 5 展示了本文工作与其他工作在头部姿态估计任务中的比较结果. 本文在头部姿态估计任务的 *MAE* 为 3.70, 而目前性能领先的方法 SADRNet[42]也仅有 3.82.

4.5.3 三维人脸重建

在本节中, 本文对人脸重建的结果进行了量化比较. 如表 6 所示, 本研究在 AFLW2000-3D 数据集上对超过 45K 个顶点进行误差估计. 评价指标为经过边界框尺寸归一化后的 *NME*. 基于 MobileNet-V2 为主干网络的模型 *NME* 达到了 4.04. 本文还与 3DDFA-V2 进行了可视化结果的比较, 如图 7 所示, 中间一行展示了 3DDFA-V2 的结果, 最底部一行展示了本文提出算法的结果. 3DDFA-V2 代码局限于纯脸部结果, 而本工作重建了整个头部并进行可视化, 其中可以发现本文的算法效果更为逼真.

表 3 原始版本 AFLW2000-3D 数据集上人脸关键点检测的对比实验结果

方法	*NME*			
	yaw ∈ [0, 30°)	yaw ∈ [30°, 60°)	yaw ∈ [60°, 90°]	所有情况
ESR[47]	4.60	6.70	12.67	7.99
3DDFA[25]	3.43	4.24	7.17	4.94
Dense Corr[48]	3.62	6.06	9.56	6.41
3DSTN[49]	3.15	4.33	5.98	4.49
3DFAN[44]	3.16	3.53	4.60	3.76
3DFFA-PAMI[24]	2.84	3.57	4.96	3.79
PRNET[26]	2.75	3.51	4.61	3.62
2DASL[50]	2.75	3.46	4.45	3.55
3DDFA-V2[27]	2.75	3.49	4.53	3.59
SADRNet[42]	2.66	**3.30**	4.42	3.64
B-spline FFD[22]	**2.60**	3.44	4.50	3.51
Chen等人[23]	2.64	3.41	4.49	3.53
AUFR (基于MobileNet-V2)	2.64	3.43	4.45	3.51
AUFR (基于ResNet-50)	**2.60**	3.41	**4.37**	**3.46**

表 4 重新标注版本 AFLW2000-3D 数据集上人脸关键点检测的对比实验结果

方法	*NME*			
	yaw ∈ [0, 30°)	yaw ∈ [30°, 60°)	yaw ∈ [60°, 90°]	所有情况
DHM[51]	2.28	3.10	6, 95	4.11
3DDFA[25]	2.84	3.52	5.15	3.83
PRNet[26]	2.35	2.78	4.22	3.11
MGCNet[52]	2.82	3.12	**3.76**	3.20
Deng等人[53]	2.56	3.11	4.45	3.37
3DDFA-V2[27]	2.84	3.03	4.13	3.33
AUFR (基于MobileNet-V2)	**2.20**	**2.55**	**3.76**	**2.83**
AUFR (基于ResNet50)	2.31	2.70	3.82	2.94

表 5 AFLW2000-3D 数据集头部姿态工作的对比实验结果

方法	*MAE*			
	航向角 (yaw)	俯仰角 (pitch)	翻滚角 (roll)	所有情况 (all)
SSRNet-MD[54]	5.14	7.09	5.89	6.01
FSANet[31]	4.50	6.08	4.64	5.07
QuatNet[30]	3.97	5.62	3.92	4.15
TriNet[32]	4.20	5.77	4.04	3.97
3DDFA-PAMI[24]	4.33	5.98	4.3	4.87
2DASL[50]	3.85	5.06	3.5	4.13
3DDFA-V2[27]	4.06	5.26	3.48	4.27
GLDL[55]	3.02	5.06	3.68	3.92
FDN[56]	3.78	5.61	3.88	4.42
MNN[57]	3.34	4.69	3.48	3.83
SADRNet[42]	**2.93**	5.00	3.54	3.82
6DRepNet[33]	3.63	4.91	3.37	3.91
AUFR (基于MobileNet-V2)	3.88	4.72	3.00	3.87
AUFR (基于ResNet50)	3.63	**4.49**	**2.98**	**3.70**

表 6 AFLW2000-3D 数据集三维人脸重建方法的对比实验结果

方法	DeFA[58]	3DDFA-V2[58]	PRNet[26]	SADRNet[42]	AUFR (基于ResNet50)	AUFR (基于MobileNet-V2)
NME	6.04	4.18	4.40	**4.02**	4.07	4.04

图 8 所示为各任务的可视化效果图, 从上往下依次是原图, 头部位姿, 人脸关键点, 三维人脸模型, 纹理图. 图 8 结果表明 AUFR 在人脸重建、关键点检测和位姿估计任务中表现良好且面对大角度头部偏转和遮挡等复杂场景性能较鲁棒.

(a) 原图

(b) 3DDFA-V2结果图

(c) 本文算法结果图

图 7　人脸重建可视化结果对比

(a) 原图

(b) 头部姿态可视化

(c) 三维关键点可视化

(d) 三维人脸模型可视化

(e) 纹理图可视化

图 8　本文方法各任务可视化效果图

5 总　结

本文提出了一种结合面部动作单元感知的三维人脸重建算法, 辅助使用面部动作单元信息和面部关键点检测来

指导 3DMM 参数回归, 有效降低了整体的训练难度. 该算法是第 1 个将面部动作单元和面部重建相结合的算法, 在多项任务中表现出了出色的性能, 并克服了单视角图像的限制, 提高了重建人脸模型的保真度. 此外, 该算法实时性较好, 为解决各种人脸任务的实施提供了基础方案, 并为单视角图像重建三维人脸提供了新的思路. 本文还提出了一套半自动的动作单元标注方案, 并构建了 300W-LP-AU 数据集. 该数据集包括 63 万张人脸图像, 每张图像含有 41 个具有检测信息的面部动作单元, 为新的面部动作单元辅助三维人脸重建算法或模型的研究提供了数据支撑.

本文中仅使用了是否被激活的 AU 先验就对人脸重建性能有所增幅, 这证明面部动作单元对人脸重建工作的确有良性的引导. 在下一步工作中将继续研究更加准确、分类更加细致的 AU 标签 (例如 AU 强度) 对人脸重建或是其他人脸姿态相关任务的影响.

References:

[1] Paysan P, Knothe R, Amberg B, Romdhani S, Vetter T. A 3D face model for pose and illumination invariant face recognition. In: Proc. of the 6th IEEE Int'l Conf. on Advanced Video and Signal Based Surveillance. Genova: IEEE, 2009. 296–301. [doi: 10.1109/AVSS.2009.58]

[2] Luo C, Song SY, Xie WC, Shen LL, Gunes H. Learning multi-dimensional edge feature-based au relation graph for facial action unit recognition. In: Proc. of the 31st Int'l Joint Conf. on Artificial Intelligence. Vienna: IJCAI, 2022. 1239–1246.

[3] Zhang R, Tsai PS, Cryer JE, Shah M. Shape-from-shading: A survey. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1999, 21(8): 690–706. [doi: 10.1109/34.784284]

[4] Quéau Y, Mélou J, Castan F, Cremers D, Durou JD. A variational approach to shape-from-shading under natural illumination. In: Proc. of the 11th Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition. Venice: Springer, 2018. 342–357. [doi: 10.1007/978-3-319-78199-0_23]

[5] Zeng XX. Deep learning methods for face detection and 3D reconstruction [Ph.D. Thesis]. Shenzheng: Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, 2021. [doi: 10.27822/d.cnki.gszxj.2021.000009]

[6] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces. In: Proc. of the 26th Annual Conf. on Computer Graphics and Interactive Techniques. Los Angeles: ACM, 1999. 187–194. [doi: 10.1145/311535.311556]

[7] Jourabloo A, Liu XM. Large-pose face alignment via CNN-based dense 3D model fitting. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 4188–4196. [doi: 10.1109/CVPR.2016.454]

[8] Sela M, Richardson E, Kimmel R. Unrestricted facial geometry reconstruction using image-to-image translation. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision (ICCV). Venice: IEEE, 2017. 1585–1594. [doi: 10.1109/ICCV.2017.175]

[9] Tran AT, Hassner T, Masi I, Medioni G. Regressing robust and discriminative 3D morphable models with a very deep neural network. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 1493–1502. [doi: 10.1109/CVPR.2017.163]

[10] Yi HW, Li C, Cao Q, Shen XY, Li S, Wang GP, Tai YW. MMFace: A multi-metric regression network for unconstrained face reconstruction. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 7655–7664. [doi: 10.1109/CVPR.2019.00785]

[11] Wu F, Bao L, Chen Y, *et al*. MVF-net: Multi-view 3D face morphable model regression. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 959–968. [doi: 10.1109/CVPR.2019.00105]

[12] Chen AP, Chen Z, Zhang GL, Mitchell K, Yu JY. Photo-realistic facial details synthesis from single image. In: Proc. of the 2019 IEEE/CVF Int'l Conf. on Computer Vision (ICCV). Seoul: IEEE, 2019. 9428–9438. [doi: 10.1109/ICCV.2019.00952]

[13] Lattas A, Moschoglou S, Gecer B, Ploumpis S, Triantafyllou V, Ghosh A, Zafeiriou S. AvatarMe: Realistically renderable 3D facial reconstruction "in-the-wild". In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020. 757–766. [doi: 10.1109/CVPR42600.2020.00084]

[14] van Rootseler RTA, Spreeuwers LJ, Veldhuis RNJ. Using 3D morphable models for face recognition in video. In: Proc. of the 33rd Symp. on Information Theory in the Benelux and the 2nd Joint WIC/IEEE Symp. on Information Theory and Signal Processing in the Benelux. Boekelo: WIC, 2012. 235–242. http://purl.utwente.nl/publications/80462

[15] Lee YJ, Lee SJ, Park KR, Jo J, Kim J. Single view-based 3D face reconstruction robust to self-occlusion. EURASIP Journal on Advances in Signal Processing, 2012, 2012: 176. [doi: 10.1186/1687-6180-2012-176]

[16] Huber P, Feng ZH, Christmas W, Kittler J, Rätsch M. Fitting 3D morphable face models using local features. In: Proc. of the 2015 IEEE Int'l Conf. on Image Processing (ICIP). Quebec City: IEEE, 2015. 1195–1199. [doi: 10.1109/ICIP.2015.7350989]

[17] Romdhani S, Vetter T. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In: Proc. of the 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005. 986–993. [doi: 10.1109/CVPR.2005.145]

[18] Zhu XY, Lei Z, Yan JJ, Yi D, Li SZ. High-fidelity pose and expression normalization for face recognition in the wild. In: Proc. of the 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015. 787–796. [doi: 10.1109/CVPR.2015.7298679]

[19] Zhao SW, Wang XM, Zhang DH, Zhang GY, Wang ZY, Liu HH. FM-3DFR: Facial manipulation-based 3-D face reconstruction. IEEE Trans. on Cybernetics, 2024, 54(1): 209–218. [doi: 10.1109/TCYB.2023.3242368]

[20] Chen YJ, Wu FZ, Wang ZY, Song YB, Ling YG, Bao LC. Self-supervised learning of detailed 3D face reconstruction. IEEE Trans. on Image Processing, 2020, 29: 8696–8705. [doi: 10.1109/TIP.2020.3017347]

[21] Wu CY, Xu QG, Neumann U. Synergy between 3DMM and 3D landmarks for accurate 3D facial geometry. In: Proc. of the 2021 Int'l Conf. on 3D Vision (3DV). London: IEEE, 2021. 453–463. [doi: 10.1109/3DV53792.2021.00055]

[22] Jung H, Oh MS, Lee SW. Learning free-form deformation for 3D face reconstruction from in-the-wild images. In: Proc. of the 2021 IEEE Int'l Conf. on Systems, Man, and Cybernetics (SMC). Melbourne: IEEE, 2021. 2737–2742. [doi: 10.1109/SMC52423.2021.9659124]

[23] Chen Z, Guan T, Wang YS, Luo YW, Xu LY, Liu WK. Learning 3-D face shape from diverse sources with cross-domain face synthesis. IEEE MultiMedia, 2023, 30(1): 7–16. [doi: 10.1109/MMUL.2022.3195091]

[24] Zhu XY, Liu XM, Lei Z, Li SZ. Face alignment in full pose range: A 3D total solution. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2019, 41(1): 78–92. [doi: 10.1109/TPAMI.2017.2778152]

[25] Zhu XY, Lei Z, Liu XM, Shi HL, Li SZ. Face alignment across large poses: A 3D solution. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 146–155. [doi: 10.1109/CVPR.2016.23]

[26] Feng Y, Wu F, Shao XH, Wang YF, Zhou X. Joint 3D face reconstruction and dense alignment with position map regression network. In: Proc. of the 15th European Conf. on Computer Vision (ECCV). Munich: Springer, 2018. 557–574. [doi: 10.1007/978-3-030-01264-9_33]

[27] Guo JZ, Zhu XY, Yang Y, Yang F, Lei Z, Li SZ. Towards fast, accurate and stable 3D dense face alignment. In: Proc. of the 16th European Conf. on Computer Vision. Glasgow: Springer, 2020. 152–168. [doi: 10.1007/978-3-030-58529-7_10]

[28] Mukherjee SS, Robertson NM. Deep head pose: Gaze-direction estimation in multimodal video. IEEE Trans. on Multimedia, 2015, 17(11): 2094–2107. [doi: 10.1109/TMM.2015.2482819]

[29] Ruiz N, Chong E, Rehg JM. Fine-grained head pose estimation without keypoints. In: Proc. of the 2018 IEEE Conf. on Computer Vision and Pattern Recognition Workshops. Salt Lake City: IEEE, 2018. 2074–2083. [doi: 10.1109/CVPRW.2018.00281]

[30] Hsu HW, Wu TY, Wan S, Wong WH, Lee CY. QuatNet: Quaternion-based head pose estimation with multiregression loss. IEEE Trans. on Multimedia, 2019, 21(4): 1035–1046. [doi: 10.1109/TMM.2018.2866770]

[31] Yang TY, Chen YT, Lin YY, Chuang YY. FSA-NET: Learning fine-grained structure aggregation for head pose estimation from a single image. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 1087–1096. [doi: 10.1109/CVPR.2019.00118]

[32] Cao ZW, Chu ZC, Liu DF, Chen YJ. A vector-based representation to enhance head pose estimation. In: Proc. of the 2021 IEEE Winter Conf. on Applications of Computer Vision. Waikoloa: IEEE, 2021. 1187–1196. [doi: 10.1109/WACV48630.2021.00123]

[33] Hempel T, Abdelrahman AA, Al-Hamadi A. 6d rotation representation for unconstrained head pose estimation. In: Proc. of the 2022 IEEE Int'l Conf. on Image Processing (ICIP). Bordeaux: IEEE, 2022. 2496–2500. [doi: 10.1109/ICIP46576.2022.9897219]

[34] Rosenberg EL, Ekman P. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). 3rd ed., Oxford: Oxford University Press, 2020.

[35] Li Y, Zeng JB, Liu X, Shan SG. Progress and challenges in facial action unit detection. Journal of Image and Graphics, 2020, 25(11): 2293–2305. [doi: 10.11834/jig.200343]

[36] Yan YF, Lyu K, Xue J, Wang C, Gan W. Facial animation method based on deep learning and expression AU parameters. Journal of Computer-aided Design & Computer Graphics, 2019, 31(11): 1973–1980. [doi: 10.3724/SP.J.1089.2019.17682]

[37] Cohn JF, Ambadar Z, Ekman P. Observer-based measurement of facial expression with the Facial Action Coding System. In: Coan JA, Allen JJB, eds. Handbook of Emotion Elicitation and Assessment. Oxford: Oxford University Press, 2007. 203–221.

[38] Gan W, Xue J, Lu K, Yan YF, Gao PC, Lyu JY. FEAFA+: An extended well-annotated dataset for facial expression analysis and 3D facial animation. In: Proc. of the 14th Int'l Conf. on Digital Image Processing (ICDIP 2022). Wuhan: SPIE, 2022. 1234211. [doi: 10.1117/12.2643588]

[39] Yan YF, Lu K, Xue J, Gao PC, Lyu JY. FEAFA: A well-annotated dataset for facial expression analysis and 3D facial animation. In: Proc. the 2019 IEEE Int'l Conf. on Multimedia & Expo Workshops (ICMEW). Shanghai: IEEE, 2019. 96–101. [doi: 10.1109/ICMEW.

2019.0-104]

[40] Yin X, Liu XM. Multi-task convolutional neural network for pose-invariant face recognition. IEEE Trans. on Image Processing, 2018, 27(2): 964–975. [doi: 10.1109/TIP.2017.2765830]

[41] Feng ZH, Kittler J, Awais M, Huber P, Wu XJ. Wing loss for robust facial landmark localisation with convolutional neural networks. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 2235–2245. [doi: 10.1109/CVPR.2018.00238]

[42] Ruan ZY, Zou CQ, Wu LH, Wu GS, Wang LM. SADRNet: Self-aligned dual face regression networks for robust 3D dense face alignment and reconstruction. IEEE Trans. on Image Processing, 2021, 30: 5793–5806. [doi: 10.1109/TIP.2021.3087397]

[43] Köstinger M, Wohlhart P, Roth PM, Bischof H. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In: Proc. of the 2011 IEEE Int'l Conf. on Computer Vision Workshops (ICCV Workshops). Barcelona: IEEE, 2011. 2144–2151. [doi: 10.1109/ICCVW.2011.6130513]

[44] Bulat A, Tzimiropoulos G. How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks). In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision. Venice: IEEE, 2017. 1021–1030. [doi: 10.1109/ICCV.2017.116]

[45] Sandler M, Howard A, Zhu ML, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proc. of the 2018 IEEE Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510–4520. [doi: 10.1109/CVPR.2018.00474]

[46] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: 10.1109/CVPR.2016.90]

[47] Cao XD, Wei YC, Wen F, Sun J. Face alignment by explicit shape regression. Int'l Journal of Computer Vision, 2014, 107(2): 177–190. [doi: 10.1007/s11263-013-0667-3]

[48] Yu R, Saito S, Li HX, Ceylan D, Li H. Learning dense facial correspondences in unconstrained images. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision. Venice: IEEE, 2017. 4733–4742. [doi: 10.1109/ICCV.2017.506]

[49] Bhagavatula C, Zhu CC, Luu K, Savvides M. Faster than real-time facial alignment: A 3D spatial transformer network approach in unconstrained poses. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision. Venice: IEEE, 2017. 4000–4009. [doi: 10.1109/ICCV.2017.429]

[50] Tu XG, Zhao J, Xie M, Jiang ZH, Balamurugan A, Luo Y, Zhao Y, He LX, Ma Z, Feng JS. 3D face reconstruction from a single image assisted by 2D face images in the wild. IEEE Trans. on Multimedia, 2021, 23(99): 1160–1172. [doi: 10.1109/tmm.2020.2993962]

[51] Sun B, Shao M, Xia SY, Fu Y. Deep evolutionary 3D diffusion heat maps for large-pose face alignment. In: Proc. of the 2018 British Machine Vision Conf. 2018. Newcastle: BMVC, 2018. 256.

[52] Shang JX, Shen TW, Li SW, Zhou L, Zhen MM, Fang T, Quan L. Self-supervised monocular 3D face reconstruction by occlusion-aware multi-view geometry consistency. In: Proc. of the 16th European Conf. on Computer Vision. Glasgow: Springer, 2020. 53–70. [doi: 10.1007/978-3-030-58555-6_4]

[53] Deng Y, Yang JL, Xu SC, Chen D, Jia YD, Tong X. Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops. Long Beach: IEEE, 2019. 285–295. [doi: 10.1109/CVPRW.2019.00038]

[54] Yang TY, Huang YH, Lin YY, Hsiu PC, Chuang YY. SSR-Net: A compact soft stagewise regression network for age estimation. In: Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence. Stockholm: IJCAI, 2018. 1078–1084.

[55] Liu ZX, Chen ZZ, Bai JQ, Li SH, Lian SG. Facial pose estimation by deep learning from label distributions. In: Proc. of the 2019 IEEE/CVF Int'l Conf. on Computer Vision Workshop. Seoul: IEEE, 2019. 1232–1240. [doi: 10.1109/ICCVW.2019.00156]

[56] Zhang H, Wang MM, Liu Y, Yuan Y. FDN: Feature decoupling network for head pose estimation. In: Proc. of the 2020 AAAI Conf. on Artificial Intelligence. New York: AAAI, 2020. 12789–12796. [doi: 10.1609/aaai.v34i07.6974]

[57] Valle R, Buenaposada JM, Baumela L. Multi-task head pose estimation in-the-wild. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2021, 43(8): 2874–2881. [doi: 10.1109/TPAMI.2020.3046323]

[58] Liu YJ, Jourabloo A, Ren W, Liu XM. Dense face alignment. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision Workshops. Venice: IEEE, 2017. 1619–1628. [doi: 10.1109/ICCVW.2017.190]

附中文参考文献:

[5] 曾小星. 深度人脸检测与三维重建方法研究 [博士学位论文]. 深圳: 中国科学院大学 (中国科学院深圳先进技术研究院), 2021. [doi: 10.27822/d.cnki.gszxj.2021.000009]

[35] 李勇, 曾加贝, 刘昕, 山世光. 面部动作单元检测方法进展与挑战. 中国图象图形学报, 2020, 25(11): 2293–2305. [doi: 10.11834/jig.

200343]

[36] 闫衍芙, 吕科, 薛健, 王聪, 甘玮. 基于深度学习和表情 AU 参数的人脸动画方法. 计算机辅助设计与图形学学报, 2019, 31(11): 1973–1980. [doi: 10.3724/SP.J.1089.2019.17682]

章毅(1999－), 男, 硕士生, CCF 学生会员, 主要研究领域为人体姿态估计, 头部重建.

兰星(1997－), 男, 博士生, CCF 学生会员, 主要研究领域为模式识别, 计算机视觉.

吕嘉仪(1999－), 女, 博士生, CCF 学生会员, 主要研究领域为面部动作单元检测, 强度回归.

薛健(1979－), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为数字图像分析与处理, 计算机图形学, 科学计算可视化.