

中图分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2025)12-3804-20

论文引用格式: Zhang K, Liang L P, Guo Y R and Wang Z N. 2025. Open-set semi-supervised multi-task learning method for facial age estimation. Journal of Image and Graphics, 30(12):3804-3823(张珂, 梁龙萍, 郭玉荣, 王子念. 2025. 面向人脸年龄估计的开集半监督多任务学习方法. 中国图象图形学报, 30(12):3804-3823)[DOI:10.11834/jig.250023]

面向人脸年龄估计的开集半监督多任务学习方法

张珂^{1,2,3,4}, 梁龙萍^{2,5}, 郭玉荣^{1,2,3,4*}, 王子念²

1. 华北电力大学 燕赵电力实验室, 保定 071003; 2. 华北电力大学电子与通信工程系, 保定 071003;
3. 河北省电力物联网技术重点实验室, 保定 071003; 4. 电力物联智能化技术河北省工程研究中心, 保定 071003;
5. 贵州电网有限责任公司安顺供电局, 安顺 561000

摘要: 目的 人脸图像年龄估计在数字营销和人机交互等领域具有重要应用价值。然而, 实现精确人脸年龄估计面临缺乏大规模有标签数据集的挑战。半监督学习方法能利用无标签数据集缓解此问题, 但现有方法易引入错误伪标签, 对年龄估计性能产生负面影响。因此, 提出一种面向人脸年龄估计的开集半监督多任务学习方法。方法 首先, 为了增强模型对局部和全局特征的处理能力, 提出 SwinLEDF 模型, 该模型以 Swin Transformer 作为主干网络, 用于提取全局特征, 并通过融合 LEFF(local enhanced feed-forward)模块和 DFN(dynamic filter networks)模块, 进一步提升模型对局部特征的提取能力。其次, 为了有效利用有标签数据和无标签数据中的有效信息, 设计开集半监督多任务学习框架。在此框架中, 模型通过标准闭集分类器和多类二元分类器的协同工作有效排除异常数据的干扰, 采用自适应阈值方法确定性别、种族和年龄的伪标签, 并引入负学习策略, 以提高对无标签数据的利用率。结果 在 MORPH 数据集上, 仅使用有标签数据集时, 模型的平均绝对误差为 1.908; 同时使用有标签数据集和无标签数据集时, MAE(mean absolute error)降至 1.885。在 UTKface 数据集上, 仅使用有标签数据集时, MAE 为 4.343; 而结合有标签数据集和无标签数据集时, MAE 降至 4.246。与现有的人脸年龄估计方法相比, 本文方法提高年龄估计的性能, 能够有效利用无标签数据集进一步优化年龄估计性能。结论 本文提出一种面向人脸年龄估计的开集半监督多任务学习方法, 能够从有标签数据集和无标签数据集中有效提取人脸图像的性别、种族和年龄特征, 从而提升人脸年龄估计的精度。这为实现更加精准的人脸年龄估计提供了新的思路和解决方案。

关键词: 人脸年龄估计; 开集半监督学习; 多任务学习; SwinLEDF 模型; 伪标签

Open-set semi-supervised multi-task learning method for facial age estimation

Zhang Ke^{1,2,3,4}, Liang Longping^{2,5}, Guo Yurong^{1,2,3,4*}, Wang Zinian²

1. Yanzhao Electric Power Laboratory of North China Electric Power University, Baoding 071003, Hebei, China;
2. Department of Electronic and Communication Engineering, North China Electric Power University, Baoding 071003, Hebei, China;
3. Hebei Key Laboratory of Power Internet of Things Technology, Baoding 071003, Hebei, China;
4. Hebei Engineering Research Center of Intelligent Technology for Power Internet of Things, Baoding 071003, Hebei, China;
5. An Shun Power Supply Bureau of Guizhou Grid Co. Ltd., Anshun 561000, Guizhou, China

收稿日期: 2025-01-24; 修回日期: 2025-04-11; 预印本日期: 2025-04-18

* 通信作者: 郭玉荣 guoyurong@ncepu.edu.cn

基金项目: 国家自然科学基金项目(62506129, 62076093, 61871182, 62206095); 中央高校基本科研业务费专项资金资助(2022MS078, 2023JG002, 2023JC006, 2025MS117); 河北省自然科学基金项目(F2024502017)

Supported by: National Natural Science Foundation of China(62506129, 62076093, 61871182, 62206095); Fundamental Research Funds for the Central Universities(2022MS078, 2023JG002, 2023JC006, 2025MS117); Natural Science Foundation of Hebei Province, China(F2024502017)

Abstract: Objective Facial age estimation from images constitutes a prominent area of research within the field of computer vision, offering extensive potential applications in fields such as biometrics, digital marketing, healthcare, and human-computer interaction. Despite substantial efforts by numerous researchers in this field, achieving accurate facial age estimation remains a formidable challenge, primarily due to the lack of high-quality, large-scale labeled datasets for facial age estimation. The manual annotation of facial datasets necessitates considerable time and financial costs. Semi-supervised learning has emerged as a promising strategy for solving this problem because it enables the simultaneous utilization of labeled and unlabeled data. However, achieving satisfactory results in the domain of facial age estimation using semi-supervised learning methods is difficult. This difficulty arises from the limited accuracy of the pseudo-labels produced by these methods, as well as their susceptibility to the influence of outlier data. These factors hinder the effective utilization of unlabeled data, consequently limiting overall performance. Aiming to address these challenges, optimizing the capability of the model to extract features is essential. Such improvements will facilitate the effective acquisition of valuable representations from unlabeled data, thereby yielding highly precise pseudo-labels. Additionally, establishing a semi-supervised learning framework that can adeptly manage the challenges associated with outlier data while optimizing the utilization of the unlabeled dataset is crucial. Consequently, this study presents an open-set semi-supervised multi-task approach for facial age estimation. **Method** This research presents the SwinLEDF model to optimize the capability of the model to extract local and global features from facial images. This model is based on the Swin Transformer architecture and integrates local enhanced feedforward (LEFF) modules along with dynamic filter networks (DFNs). The Swin Transformer demonstrates proficient capabilities in capturing long-range dependencies and global characteristics, particularly in the analysis of age-related trends and the overall morphology of facial structures. The LEFF module incorporates non-linear transformations at the feature level, facilitating the identification of local patterns within images or feature representations. This capability is essential for differentiating age-related attributes, including intricate details such as wrinkles and skin texture. The DFN module implements a dynamic filtering operation within the spatial dimension of the model's output, thereby enhancing model flexibility and adaptability. Furthermore, this research presents an open-set semi-supervised multi-task learning algorithm to optimize the use of labeled and unlabeled data. In this algorithm, the model assesses the probability of unlabeled data being classified as outliers by integrating the outcomes of a closed-set classifier and a multi-class binary classifier. Subsequently, the model generates pseudo-labels for non-outlier data that meet a specified confidence threshold. Additionally, the model simultaneously learns to estimate sex, race, and age using labeled and unlabeled data. Through this process, the model learns not only the unique characteristics associated with each specific task but also the interrelationships among gender, race, and age, thereby enhancing the capability of the model to process diverse data and increases its expressive power and robustness. Furthermore, the process enables the effective utilization of unlabeled datasets, addressing the challenge of limited labeled data in the field of age estimation. This study employs an adaptive threshold mechanism and a negative learning strategy to optimize the use of unlabeled data. The adaptive threshold mechanism dynamically adjusts the confidence threshold for pseudo-labels based on the model's training performance across different categories, effectively addressing category imbalance and improving the precision of pseudo-label production. The negative learning strategy enhances the handling of unlabeled data by identifying categories to which the input data does not belong, thereby mitigating the adverse effects of false pseudo-labels on model performance. **Result** This study assesses the proposed methodology using the MORPH and UTKface datasets. On the MORPH dataset, the model exhibits a mean absolute error (MAE) of 1.908 when trained solely on labeled data. This error is further reduced to 1.885 with the inclusion of labeled and unlabeled datasets. Similarly, for the UTKface dataset, the initial MAE is recorded at 4.343 using only labeled datasets, which subsequently reduces to 4.246 following the integration of labeled and unlabeled datasets. Compared to current facial age estimation methods, the proposed approach exhibits superior performance and further optimizes its accuracy by leveraging unlabeled facial datasets. **Conclusion** This study introduces an open-set semi-supervised multi-task learning method for facial age estimation. The proposed method effectively extracts gender, race, and age attributes from facial images while leveraging unlabeled data and appropriately handling potential outliers. This approach addresses the challenges associated with limited labeled data, thereby enhancing the accuracy of facial age estimation. Furthermore, the methodology presents innovative strategies for achieving precise results and holds strong potential for practical applications. **Key words:** facial age estimation; open-set semi-supervised learning; multi-task learning; SwinLEDF model; pseudo-label

0 引言

人脸图像年龄估计作为计算机视觉领域的热门研究方向,在个性化服务、市场研究、安全监控和人机交互等多个领域具有广泛的应用前景。提升人脸图像年龄估计的准确性和鲁棒性,将为社会研究和商业应用带来巨大的便利。然而,面部年龄估计面临诸多挑战,如年龄变化具有随机性和复杂性、年龄估计受多种因素影响且难以控制,以及年龄估计领域缺乏可靠标签的大规模数据集等。

研究者针对人脸图像年龄估计领域中的挑战开展了大量的研究。首先,针对年龄与面部特征的非线性导致年龄准确估计困难的问题,提出通过引入年龄顺序属性(Cao等,2020; Zhao等,2021)、采用移动窗口回归(Shin等,2022)、结合回归和标签分布(Gao等,2018; Li等,2020; Berg等,2021)等策略来提升年龄预测模型的建模精度。针对年龄特征高效准确表征困难的问题,提出通过引入深度随机森林(Guehairia等,2020)或注意力模块(Zhang等,2020; Lin等,2022; Wang等,2022)、结合全局和局部特征以及集成特征(Tan等,2019)以及采用对比学习网络(Kong等,2022)等方法改进模型结构,以增强特征提取能力。针对不同年龄样本量不均衡导致低样本量年龄估计精度差的问题,提出诸如引入暹罗图学习(Liu等,2023)、自适应方差(Wen等,2020)、渐进式边缘损失函数(Deng等,2021b)以及均值一方差损失函数(Pan等,2018)等策略,以平衡不同年龄段样本在训练中的影响。最后,针对有标签样本不足导致模型的泛化性、准确性和鲁棒性差,采用主动学习(Singh和Chakraborty,2021; Bhattacharya和Chakraborty,2022)、半监督学习(Bao等,2023; Akbari等,2024; Liu等,2023)等方法,通过利用无标签人脸图像数据集进行有效信息提取,进一步推动了该领域的研究发展。

半监督学习方法是解决年龄估计领域缺乏可靠标签的大规模数据集问题的有效方法。通过充分利用少量有标签数据和大量无标签数据,半监督学习能够实现比仅使用有标签数据集时更优的性能。然而,人脸图像年龄估计任务本身具有较高的复杂性,应用半监督学习方法时,往往更容易引入错误标签。这些错误标签主要源于两方面:1)预测错误,由于年

龄估计任务的高复杂度,模型的预测准确度较低,从而产生错误标签;2)异常数据问题,即无标签数据集中可能包含有标签数据集中没有的类别,这些数据被错误地归类为已有类别,进而导致错误标签的生成。针对此问题,OpenMatch(Saito等,2021)、SSB(simple but strong baseline)(Fan等,2023)以及IOMatch(Li等,2023a)等研究采用开集半监督学习方法。与传统的半监督学习方法不同,开集半监督学习假设无标签数据集中可能存在有标签数据集中未出现的异常类别。在年龄估计任务中,这种异常类别表现为无标签数据集中可能包含有标签数据集中未覆盖的年龄范围或异常人脸图像。但这些方法在人脸年龄估计任务中的表现仍存在局限性。

针对以上问题,本文提出一种面向人脸年龄估计的开集半监督多任务学习方法。首先,针对人脸年龄估计中既需捕捉全局特征识别人脸的年龄趋势和整体结构,又需关注局部特征以捕捉与年龄相关的细节差异这一问题,本文提出SwinLEDF人脸年龄特征提取网络,该网络使用Swin Transformer(Liu等,2021)作为主干网络以提取全局特征,并融合LEFF(local enhanced feed-forward)网络(Yuan等,2021)和DFN(dynamic filter networks)模块(De Brabandere等,2016)以增强局部特征提取能力;其次,针对半监督学习在人脸年龄估计领域可能出现的异常数据干扰和伪标签生成精度不足问题,提出开集半监督多任务学习方法,充分利用有标签数据集和无标签数据集中的性别、种族和年龄信息,在无标签数据集中排除异常数据的干扰并提升伪标签质量,进而提高年龄估计任务精度。具体而言,在SwinLEDF模型中,Swin Transformer通过引入层次化特征结构和窗口注意力机制,能够有效捕捉长距离依赖关系和全局特征;LEFF与DFN两个卷积模块的结合提升模型提取局部特征的能力。其中,LEFF模块通过在特征层面上引入非线性变换,能有效捕捉图像或特征的局部模式,DFN模块则在模型输出的空间维度上应用动态滤波操作,增强模型的灵活性和适应性。这使得SwinLEDF能够同时提取局部特征和全局特征,从而在复杂任务中展现出更优的性能。开集半监督多任务学习算法借助标准闭集分类器和多类二元分类器共同确定无标签数据为异常数据的概率,然后为符合置信度阈值条件的非异常数据生成伪标签,并利用有标签数据集和无标签数据集进行性别、种

族和年龄3种图像估计任务。在此过程中,模型不仅能学习每个任务的专有属性,还通过任务间的关联提升对不同类型数据的处理能力,增强模型的表达能力和鲁棒性,进而有效利用无标签数据集,缓解年龄估计领域数据集不足的问题。此外,为进一步提高无标签数据的利用率,本文采用自适应阈值机制和负学习策略。其中,自适应阈值机制根据模型在不同类别上的学习表现,动态调整伪标签的判定阈值,以应对类别不平衡问题并提高标签预测准确性;负学习策略通过识别输入数据不属于哪些类别,充分利用无标签数据,减轻错误伪标签对模型的负面影响。

本文的主要贡献有:1)提出SwinLEDF网络。该网络在Swin Transformer结构上融合LEFF模块和DFN模块,使模型能够同时有效地提取局部特征和全局特征,从而增强模型提取特征的能力。2)提出开集半监督多任务学习优化算法。增加多类二元分类器以有效排除异常数据的干扰;综合考虑性别和种族对年龄估计的影响,提高模型的表达能力;并采用自适应阈值机制和负学习策略进一步增加无标签数据的利用率。该方法能充分利用性别和种族信息以及无标签数据,有效缓解数据集不足的问题,提升年龄估计的准确性。3)最后,在两个基准数据集上进行了广泛的实验评估,实验表明,本文方法表现出更好的性能,能有效提升年龄估计的准确性。

1 相关工作

1.1 人脸年龄估计

Kwon和Lobo(1994)首次使用面部图像进行年龄估计研究,将年龄分为婴儿、年轻人和老年人3组,引发了对人脸图像年龄估计的广泛研究。早期的年龄估计方法主要依赖于手工制作的特征,例如Gabor滤波器(Gabor, 1946)、局部二进制模式(local binary patterns, LBP)(Ojala等, 1994)、尺度不变特征变换(scale-invariant feature transform, SIFT)(Lowe, 1999)方法和广泛使用的“生物启发特征”(bio-inspired feature, BIF)方法(Guo等, 2009)。这些基于手工制作特征的方法在受控条件下表现良好,但当面临真实环境中的照明和位姿变化时,其精度会显著下降,使得其很难成为年龄估计问题的最佳解决方案。随着深度学习的快速发展,面部年龄估

计逐渐趋向于采用深度学习方法,从而取得了准确度和鲁棒性更高的年龄估计结果。

基于深度学习的人脸图像年龄估计方法分为基于排序、基于分类、基于回归、基于标签分布和混合方法。基于回归的方法(Gustafsson等, 2019;Cao等, 2020)将年龄视为连续变量,将面部老化视为一个回归问题,并通过建立回归模型直接从面部图像预测真实年龄,但容易导致过拟合;基于排序的方法(Shin等, 2022)检查老化过程中嵌入的顺序属性,将年龄值视为秩序数据,并使用多个二值分类器确定人脸图像中年龄的秩;基于分类的方法(Guehairia等, 2020)则将不同的年龄视为独立的类别,将年龄估计转换为多类分类问题;基于分类的方法便于神经网络的训练,但可能会忽视不同类别之间的相关性;基于标签分布的方法(Lin等, 2022; Wang等, 2022)充分考虑相邻年龄之间的相关性,将年龄估计建模为所有潜在年龄值的概率分布,可以有效地解决数据不足的问题;混合方法则通过并行或更加灵活的方式将两种或两种以上的方法进行结合,以获得更好的模型性能。例如, Gao等人(2018)将年龄分布和回归单一年龄统一到DLDL-V2框架中,可以缓解训练阶段和验证阶段的不一致性; Li等人(2020)提出标签细化网络(label refinery network, LRN),该网络包括标签分布细化和松弛回归细化两个并行过程; Zhao等人(2021)在面部年龄估计任务中提出一种基于两个teachers的知识蒸馏方法,将排序模型中的有序知识和多类分类模型中的暗知识转移到紧凑的年龄估计模型中; Berg等人(2021)通过引入标签分布多样性改进传统的顺序回归模型,以提升模型的学习能力和检测准确性。

在基于深度学习的人脸图像年龄估计中,研究者提出多种方法来提升模型预测性能。设计模型架构,进而提高模型的特征提取能力是该领域常用的方法。Tan等人(2019)设计了深度混合对齐架构(deep hybrid-aligned architecture, DHAA),该架构能联合学习全局信息、局部细节和集成特征,并引入了集成对齐区域池化和循环融合等新组件。Zhang等人(2020)提出一种基于注意力长短期记忆(long-short term memory, LSTM)网络的人脸年龄估计的AL-ResNets(attention long short-term memory residual networks)和AL-RoR(attention long short-term memory residual network of residual network models)结构,注

注意力 LSTM 单元可以获得年龄敏感区域的鉴别局部特征。Kong 等人(2022)提出一种新的深度对比网络(deep contrastive network, DCN),可以通过对比深度特征图来学习输入的未知年龄人脸图像与参考图像之间的年龄距离,并提出一种循环迭代近似算法(cyclic iterative approximation algorithm, CIAA)进一步提高年龄估计准确度。秦瑾等人(2025)在时间和空间序列上针对人脸图像的细粒度特征,提出一种基于注意力 ConvLSTM(convolution long-short term memory)模型的人脸图像年龄估计方法,提升预测精度。许多方法设计更加符合人脸年龄估计任务特点的损失函数,从而提升任务的准确性。Wen 等人(2020)提出基于方差的自适应分布学习(adaptive variance based distribution learning, AVDL)年龄估计方法,引入元学习,在单次迭代中自适应调整每幅图像的方差,在方差上执行元梯度下降。Deng 等人(2021b)提出一种无约束面部年龄分类的渐进边缘损失(progressive margin loss, PML)方法,通过强制执行序数边缘和变分边缘来逐步优化年龄标签模式。Pan 等人(2018)设计了一种结合均值损失和方差损失的均值一方差损失函数,与 softmax 损失函数一起嵌入到神经网络中以实现稳健的年龄估计。此外,多任务学习也广泛应用于人脸年龄估计领域,例如 Deng 等人(2021a)的研究、EGroupNet 方法(Duan 等, 2020)以及多任务多尺度软注意机制(multi-task multi-scale soft-attention, MMSA)(Shi 等, 2023)都采用多任务学习的方法同时学习多个人脸属性,旨在挖掘年龄相关属性与年龄属性之间的相关性,利用性别属性和种族属性辅助年龄预测。

1.2 半监督学习

半监督学习通过利用有标签数据和大量无标签数据进行联合学习来提升学习性能。近年来,深度半监督学习快速发展,在目标检测、图像分类、语义分割和姿态估计等领域都有着广泛应用。半监督学习的主流方法是一致性正则化和伪标签方法。一致性正则化方法通过保持模型在不同扰动下的稳定输出,帮助模型从无标签数据中学习,以最小化输入空间或权重空间中变化的预测方差;伪标签方法则使用预测模型或其变体为无标签数据生成伪标签,并将这些伪标签与有标签的数据混合用于训练,为模型提供一些额外的训练信息,但伪标签方法依赖于伪标签的高置信度。FixMatch(Sohn 等, 2020)将一

致性正则化和伪标签方法结合,在使用一致性正则化时使用单独的弱增强和强增强,实现简单但高效的半监督学习。在此基础上,研究者提出多种改进方法,其中不少方法专注于提升伪标签的质量和数量,采用动态调整阈值、负学习和缩小类空间等方法。针对非随机缺失标签的类感知伪标签方法(class-aware pseudo labeling, CAPL)(Gui 等, 2022)通过动态调整不同类别的伪标签阈值来充分利用无标签数据,有效缓解伪标签的不平衡性。FullMatch(Chen 等, 2023b)将熵意义损失(entropy meaning loss, EML)和自适应负学习(adaptive negative learning, ANL)集成到 FixMatch 中, EML 约束非目标类的输出分布,产生更高置信度的预测分布, ANL 为所有无标签数据引入负伪标签,利用低置信度的示例。ShrinkMatch(Yang 等, 2023)手动检测并删除混淆类,自适应地缩小类空间,在此类空间重新计算 top-1 置信度,以充分利用无标签数据集。设计更高效的半监督训练策略也是许多研究的重点方向,例如一些方法通过动态交互或为伪标签动态加权,从无标签数据集中提取到更多有效信息。SimMatch(Zheng 等, 2022)则同时考虑语义级和实例级的一致性正则化,允许语义级和实例级的伪标签通过聚合和展开技术在标签的内存缓冲区进行交互。交叉标签监督学习方法(cross labeling supervision, CLS)(Yao 等, 2022)根据预测置信度重新加权伪标签和互补标签,并在两个不同初始化的网络之间交换这些标签进行共同训练。DWPC 算法(Li 等, 2023a)是一种动态加权算法,为每个无标签数据动态提高权重,并提出了名为“对损失”的损失函数来加强有标签和无标签数据之间的联系。Chen 等人(2023a)提出 SoftMatch 方法,结合截断高斯加权函数和均匀对齐策略,为伪标签生成动态权重,从而实现数量与质量的均衡。朱徽等人(2024)提出一种基于样本动态权重的课程式半监督学习方法,使模型由简单至困难地利用样本,有效缓解噪声干扰问题,增强模型泛化能力。此外,还有一些方法在模型中集成新颖的技术或模块,以增强对无标签数据的处理能力。Kang 等人(2023)提出近邻持续半监督学习,利用最近邻分类器对特征空间进行非线性划分,并利用其非参数特性灵活地对底层数据分布进行建模。Relimatch(Jiang 等, 2023)通过集成课程标签、特征过滤模块和伪标签过滤模块来提高半监督学习分类的准确性

和可靠性。

上述方法在半监督学习领域取得显著进展,但大多基于闭集假设,即假设有标签数据集和无标签数据集的分布相同。然而,在人脸年龄估计等复杂问题中,很难确保无标签数据集中不包含新类别,且无法通过手动检查来排除所有离群值。因此,开集半监督学习得以提出,它要求模型不仅能够学习已知类别的特征表示,还具备更强的泛化能力和异常值检测能力,以便在存在异常类别的情况下,仍能保持准确高效的性能。在开集半监督学习中,一些研究采取直接滤除异常数据的处理方法。OpenMatch (Saito等,2021)将FixMatch与基于OVA(one-vs-all)分类器的新颖性检测相结合,根据OVA分类器输出的样本置信度分数来滤除异常值,但这样可能会错误地过滤掉有用信息。SSB(Fan等,2023)通过纳入高置信度的伪标签、利用非线性变换分离多任务学习框架中用于初始类和离群检测的特征,引入负伪挖掘方法,实现开集下简单但强大的半监督方法。权重感知蒸馏(weight-aware distillation, WAD)(Du等,2023)是一种鲁棒的半监督学习框架,通过探索表示空间中的点互信息(point mutual information, PMI)捕获目标实例的自适应权重和高质量伪标签,有选择性地将对目标任务有益的知识从无监督的对比表示转移到目标分类器,过滤掉异常类别。He等人(2023)提出一种安全深度半监督学习方法 safe-student,通过使用新的ED(energy-discrepancy)评分函数、未见类标签分布学习模块和迭代优化策略来保证半监督学习的安全性。另一些研究则将异常类别视为一个新类,进而生成新的开集分布,以实现开集下的半监督学习。IOMatch(Li等,2023a)通过将多二元分类器和闭集分类器相结合来产生统一的开集分类目标,可以同时有效利用类内信息和离群值;Ma等人(2023)将开集下的 K 类分类转换为闭集下的 $K+1$ 类分类问题,并进一步引入迭代负学习挖掘更广泛的类的更多知识来充分利用低置信度的伪标签。

近年来,半监督学习也广泛应用于人脸年龄估计领域。Akbari等人(2024)提出RAgE,通过保持相似度的伪标签算法和新的耐噪声一致性正则化项,利用无标签数据提高鲁棒性并减少年龄估计的不确定性。Bao等人(2023)提出一种发散驱动的一致性训练(divergence-driven consistency training, DCT)方法,该方法在基于伪标签和一致性正则化思想的半

监督方法中引入高效样本选择(efficient sample selection, ESS)策略和身份一致性(identity consistency, IC)正则化附加损失函数。Liu等人(2023)通过暹罗图生成、节点幻觉和图对比正则化3个模块来密集地建模连续老化模式,在半监督学习中实现轻监督和高精度的状态。这些基于半监督学习的人脸图像年龄估计方法为该领域的发展提供了新的思路和可能性。然而,这些方法生成的伪标签准确度仍需提高,且未能充分考虑到无标签数据集中可能存在的异常干扰。

2 本文方法

本文方法的具体实现流程如图1所示。首先,为了增强模型的特征提取能力,采用Swin Transformer作为主干网络,并结合LEFF和DFN模块,构建SwinLEDF模型;其次,为了有效利用无标签数据,解决人脸年龄估计领域中标注数据稀缺、标注成本高且费时的问题,采用开集半监督多任务学习算法进行模型优化。具体训练流程如下:1)使用有性别、种族和年龄的有标签数据集对SwinLEDF模型进行预训练;2)预训练完成后,SwinLEDF模型同时利用有标签数据和无标签数据进行微调,在充分利用有标签数据的同时,能够筛选合适的无标签数据并生成性别、种族和年龄的伪标签,利用带有伪标签的无标签数据进行微调,并使用自适应阈值方法和针对年龄估计任务的负学习方法,充分挖掘无标签数据的潜力,从而提升年龄估计性能。

2.1 SwinLEDF模型

Swin Transformer在捕捉长距离依赖关系和全局特征方面表现出色,但在细致的局部特征提取上存在一定不足。相比之下,卷积神经网络在捕捉图像的局部纹理、边缘等细节方面具有独特优势。为了提升模型的局部和全局特征提取能力,提高年龄预测的准确性,本文提出SwinLEDF模型,该模型是一个集成了Swin Transformer、LEFF和DFN 3个模块的统一框架。

SwinLEDF模型整体架构如图1所示。首先,输入图像通过Swin Transformer进行全局特征提取。在Swin Transformer中,输入图像先被转换为序列嵌入,并被切分为特征维度为 $4 \times 4 \times 3$ 的patch tokens,投影到任意维度;这些patch tokens通过多个Swin

Transformer块和patch融合层进行分层处理,逐步生成高层次特征。然后,通过局部特征提取模块增强模型对局部特征的捕捉能力。在该模块中,首先对特征进行层归一化处理,以减少内部协变量偏移并提高模型的稳定性;随后,LEFF模块通过非线性变换进一步增强模型对图像局部模式的捕捉能力,DFN模块根据样本的不同动态生成滤波参数和滤波器,从而提升模型的表现能力和泛化能力;最后,经过平均池化处理得到最终的特征。

2.1.1 全局特征提取

本文选取Swin Transformer来捕捉人脸图像的整体年龄趋势和结构特征。Swin Transformer从较小的patch开始,逐步在较深层中合并相邻patch,形成层次化的结构。这种分层设计不仅增强模型在不同尺度上建模的灵活性,还确保了计算复杂度与图像大小成线性关系。Swin Transformer通过基于窗口的多头自注意力机制(window-based multi-head self-attention, W-MSA)和基于移位窗口的多头自注意力机制(shifted window-based multi-head self-attention, SW-MSA)计算这种层次化特征表示,将自注意力计算限制在非重叠的局部窗口内,通过窗口间的交叠

和移动,有效捕捉全局信息,并且允许不同窗口之间的信息交换,从而提高全局建模能力。

Swin Transformer块是Swin Transformer的关键模块。在两个连续的Swin Transformer块中,首先对输入特征 z^{l-1} 进行层归一化(layer normalization, LN)处理,并通过W-MSA模块进行处理,然后将处理后的特征与输入特征 z^{l-1} 进行残差连接,得到输出特征 z^l 。具体为

$$z^l = W-MSA(LN(z^{l-1})) + z^{l-1} \quad (1)$$

接着,对输出特征 z^l 进行层归一化处理和多层感知器(multilayer perceptron, MLP)处理,并将处理后的特征与 z^l 进行残差连接,得到第1个Swin Transformer块的输出特征 z^l ,具体为

$$z^l = MLP(LN(z^l)) + z^l \quad (2)$$

输出特征 z^l 被传递到下一个Swin Transformer块中。首先,对 z^l 进行层归一化处理,并通过SW-MSA模块进行处理,然后将处理后的特征与 z^l 进行残差连接得到输出特征 z^{l+1} ,具体为

$$z^{l+1} = SW-MSA(LN(z^l)) + z^l \quad (3)$$

接着,对 z^{l+1} 进行层归一化处理和多层感知器处理,并将处理后的特征与 z^{l+1} 进行残差连接得到

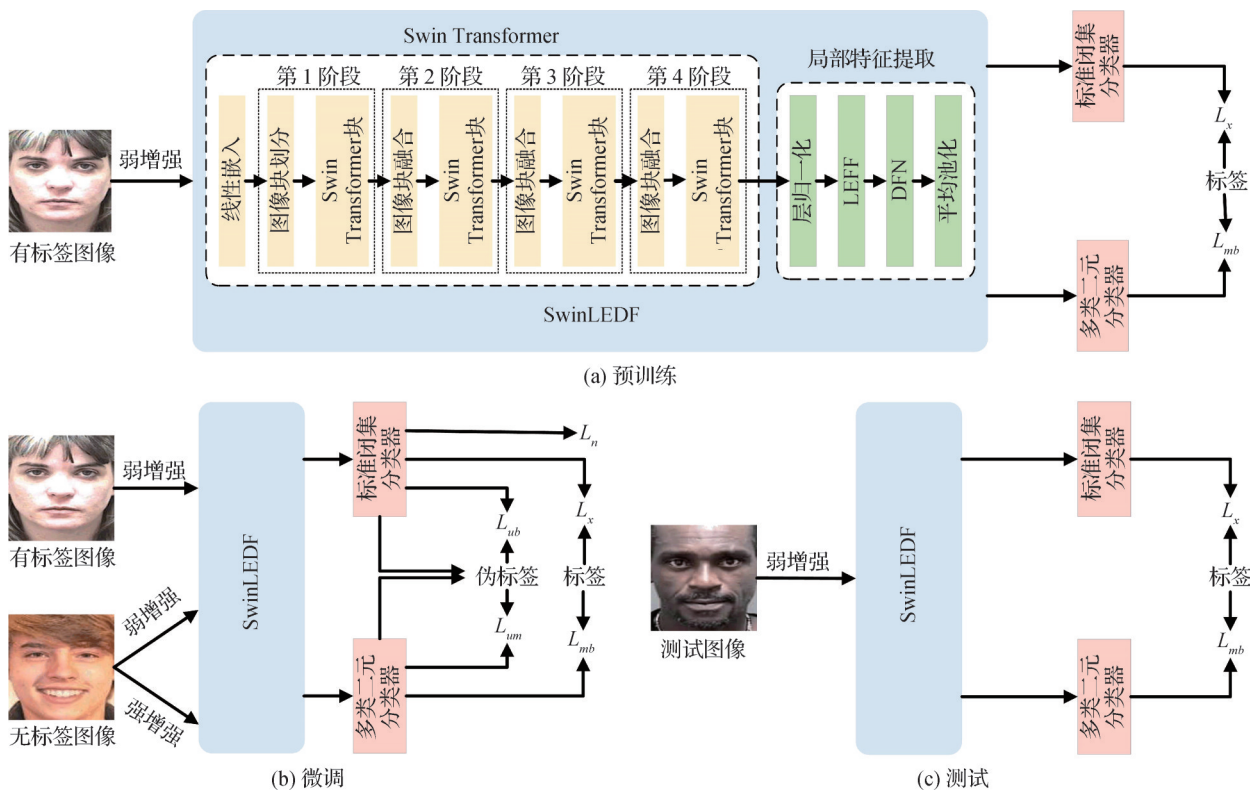


图1 本文方法的实现流程

Fig. 1 Implementation process of the method used in this paper ((a) pre-training; (b) fine-tuning; (c) testing)

第2个Swin Transformer块的输出特征 z^{l+1} ,具体为

$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \quad (4)$$

2.1.2 局部特征提取

本文采用层归一化、LEFF、DFN和平均池化的组合作为局部特征提取模块,用于处理经过Swin Transformer提取的特征表示,增强模型对局部特征信息的提取能力。

其中,LEFF模块结构如图2所示,主要由线性层和深度卷积组成,并在每个线性投影层和卷积层后加入GELU(Gaussian error linear unit)激活函数层。其核心思想是通过高维投影和深度卷积增强每个patch token的局部特征表达,从而提升模型性能。鉴于LEFF模块最初是为Vision Transformer设计的,在应用于Swin Transformer时,针对两者的架构差异,对LEFF模块进行相应的改进和调整。在改进后的LEFF中,输入特征序列 z^h 首先经过线性层被映射到更高的维度空间,得到输出的特征序列 z^l ,具体为

$$z^l = GELU(Linear1(z^h)) \quad (5)$$

然后,在这一高维空间中, z^l 通过reshape操作恢复为类似“图像”的结构 z^s ,具体为

$$z^s = SpatialRestore(z^l) \quad (6)$$

接着,对 z^s 执行深度卷积操作,以增强局部区域内特征的表示能力,得到特征表示 z^d ,具体为

$$z^d = GELU(DWConv(z^s)) \quad (7)$$

最后, z^d 被重新展平为特征序列 z^f ,并通过线性层投影回到初始的低维度空间,形成输出的特征序列 z^2 ,具体为

$$z^f = Flatten(z^d) \quad (8)$$

$$z^2 = GELU(Linear2(z^f)) \quad (9)$$

式(5)一式(9)中, $Linear1$ 代表从低维映射到高维的线性层, $GELU$ 代表GELU激活函数层, $SpatialRestore$ 代表空间恢复层, $DWConv$ 代表深度卷积操作层, $Flatten$ 代表展平操作层, $Linear2$ 代表从高维映射到低维的线性层。

传统的过滤器在训练过程中学习到的权重是固定的,不随输入数据变化。而DFN网络能够根据输入数据动态生成权重,为输入样本生成最适合的过滤器,从而更精准地捕捉特征。这种动态调整机制提升了模型的灵活性和适应性,同时不会显著增加额外的参数量,尤其在处理复杂多变的任务时具有显著优势。动态过滤网络由过滤器生成网络和动态过滤层两部分组成。在此模块中,经过LEFF局部增强处理后的特征被输入到动态过滤网络中。过滤器生成网络根据输入数据动态生成适合当前样本的滤波器。随后,动态过滤层将这些为特定样本生成的过滤器应用于输入特征,得到过滤结果。

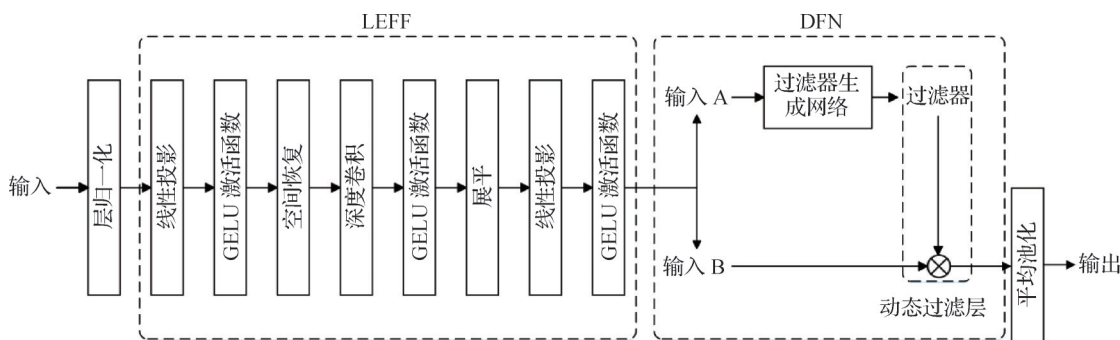


图2 局部特征提取模块

Fig. 2 Local feature extraction module

2.2 开集半监督多任务学习

为了更有效地利用有标签数据和无标签数据中的年龄、性别和种族信息以提升年龄估计性能,并避免可能存在的异常类别对模型的干扰,本文提出开集半监督多任务学习算法进行模型优化。该算法通过有标签数据和无标签数据的联合学习来提升模型的鲁棒性和精度。

在有标签数据学习中,模型通过同时进行性别、

种族和年龄的预测任务,充分利用标签信息指导模型学习到更准确的特征表示。

在无标签数据学习中,首先通过标准闭集分类器与多类二元分类器协同工作,筛选出类内数据。接着,采用自适应阈值方法,根据模型当前的学习状态为每个类别生成动态阈值,当类内数据的预测置信度超过此阈值时,将此预测结果作为伪标签。通过这一过程,模型为无标签数据生成性别、种族和年

龄的伪标签。然后,结合生成的伪标签和一致性正则化方法,进一步优化标准闭集分类器和多类二元分类器性能。

针对复杂的年龄估计问题,本文还引入负学习方法以识别和筛选伪样本,并将其纳入到学习过程中,增强模型在不同类别间的区分能力,提高无标签数据的利用率,进而提升预测准确度。

2.2.1 有标签数据学习

设 $L = \{(x_i, y_{a,i}, y_{g,i}, y_{r,i}) : i \in (1, \dots, N)\}$ 为一批包含性别、种族和年龄标签的 N 个人脸样本, x_i 为个人脸样例, $y_{a,i}$, $y_{g,i}$ 和 $y_{r,i}$ 为对应的年龄、性别和种族标签。根据性别、种族和年龄标签,可以确定对应的真实分布。由于年龄的真实分布可视为高斯分布,样本 x_i 属于 k_1 类年龄的真实分布概率 $q_{a,i}^{k_1}$ 为

$$q_{a,i}^{k_1} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(k_1 - y_{a,i})^2}{2\sigma^2}\right) \quad (10)$$

式中, σ 为标准差。

而性别和种族的真实分布被视为单点分布,则样本 x_i 属于 k_2 类性别的真实分布概率 $q_{g,i}^{k_2}$ 为

$$q_{g,i}^{k_2} = \begin{cases} 1 & k_2 = y_{g,i} \\ 0 & k_2 \neq y_{g,i} \end{cases} \quad (11)$$

样本 x_i 属于 k_3 类种族的真实分布概率 $q_{r,i}^{k_3}$ 为

$$q_{r,i}^{k_3} = \begin{cases} 1 & k_3 = y_{r,i} \\ 0 & k_3 \neq y_{r,i} \end{cases} \quad (12)$$

人脸样本通过 SwinLEDF 模型输出标准闭集分类器上的预测概率分布 $p_{a,i}$, $p_{g,i}$ 和 $p_{r,i}$, 则有标签数据集在年龄、性别和种族标准闭集分类器上的损失 L_{x-a} , L_{x-g} 和 L_{x-r} 为

$$L_{x-a} = \frac{1}{N} \sum_{i=1}^N \sum_{k_1=1}^{K_1} p_{a,i}^{k_1} \times \log\left(\frac{p_{a,i}^{k_1}}{q_{a,i}^{k_1}}\right) \quad (13)$$

$$L_{x-g} = \frac{1}{N} \sum_{i=1}^N \sum_{k_2=1}^{K_2} p_{g,i}^{k_2} \times \log(q_{g,i}^{k_2}) \quad (14)$$

$$L_{x-r} = \frac{1}{N} \sum_{i=1}^N \sum_{k_3=1}^{K_3} p_{r,i}^{k_3} \times \log(q_{r,i}^{k_3}) \quad (15)$$

式中, K_1 , K_2 , K_3 为年龄、性别和种族的类别数, $p_{a,i}^{k_1}$ 为样本 x_i 属于 k_1 类年龄的标准闭集预测分布概率, $p_{g,i}^{k_2}$ 为样本 x_i 属于 k_2 类性别的标准闭集预测分布概率, $p_{r,i}^{k_3}$ 为样本 x_i 属于 k_3 类种族的的标准闭集预测分布概率。

标准闭集分类器的总损失 L_x 为

$$L_x = L_{x-a} + L_{x-g} + L_{x-r} \quad (16)$$

此外,多类二元分类器对于样本 x_i 的第 k 类的二元分类输出为 (o_i^k, \bar{o}_i^k) , o_i^k 表示样本属于第 k 类的概率, \bar{o}_i^k 表示样本不属于第 k 类的概率,且 $o_i^k + \bar{o}_i^k = 1$ 。采用硬负分类器采样策略(Saito等,2021)对年龄、性别和种族的多类二元分类器进行优化,可以得到对应的损失 L_{mb-a} , L_{mb-g} 和 L_{mb-r} , 具体为

$$L_{mb-a} = \frac{1}{N} \sum_{i=1}^N (-\log(o_{a,i}^{y_{a,i}}) - \min_{k_1 \neq y_{a,i}} \log(\bar{o}_{a,i}^{k_1})) \quad (17)$$

$$L_{mb-g} = \frac{1}{N} \sum_{i=1}^N (-\log(o_{g,i}^{y_{g,i}}) - \min_{k_2 \neq y_{g,i}} \log(\bar{o}_{g,i}^{k_2})) \quad (18)$$

$$L_{mb-r} = \frac{1}{N} \sum_{i=1}^N (-\log(o_{r,i}^{y_{r,i}}) - \min_{k_3 \neq y_{r,i}} \log(\bar{o}_{r,i}^{k_3})) \quad (19)$$

式中, $o_{a,i}^{y_{a,i}}$ 表示样本 x_i 在年龄多类二元分类器上输出的属于 $y_{a,i}$ 类的预测概率; $\bar{o}_{a,i}^{k_1}$ 表示样本 x_i 在年龄多类二元分类器上输出的不属于 k_1 类的预测概率; $o_{g,i}^{y_{g,i}}$ 表示样本 x_i 在性别多类二元分类器上输出的属于 $y_{g,i}$ 的预测概率; $\bar{o}_{g,i}^{k_2}$ 表示样本 x_i 在性别多类二元分类器上输出的不属于 k_2 类的预测概率; $o_{r,i}^{y_{r,i}}$ 表示样本 x_i 在种族多类二元分类器上输出的属于 $y_{r,i}$ 类的预测概率; $\bar{o}_{r,i}^{k_3}$ 表示样本 x_i 在种族多类二元分类器上输出的不属于 k_3 类的预测概率。

最后,多类二元分类器的总损失 L_{mb} 为

$$L_{mb} = L_{mb-a} + L_{mb-g} + L_{mb-r} \quad (20)$$

因此,有标签数据学习的损失 L_s 为

$$L_s = L_x + \beta L_{mb} \quad (21)$$

式中, β 为超参数,代表 L_{mb} 的权重系数。

2.2.2 无标签数据学习

设 $U = \{u_i : i = (1, \dots, \mu N)\}$ 为 μN 个无标签人脸样本, u_i 为无标签样本,样本分别经过弱增强和强增强处理后,输入到模型中进行学习。

1) 异常数据判断。如图3所示,为了保证伪标签的准确性,首先需要判断该样本是否属于异常数据,这就需要计算样本属于异常数据的概率。

设 $p_{a,i}^{k_1,w}$ 和 $o_{a,i}^{k_1,w}$ 是模型输出的 u_i 的弱增强版本属于 k_1 类年龄的标准闭集预测概率和多类二元预测概率; $p_{g,i}^{k_2,w}$ 和 $o_{g,i}^{k_2,w}$ 是模型输出的 u_i 的弱增强版本属于 k_2 类性别的标准闭集预测概率和多类二元预测概率; $p_{r,i}^{k_3,w}$ 和 $o_{r,i}^{k_3,w}$ 是模型输出的 u_i 的弱增强版本属于 k_3 类种族的的标准闭集预测概率和多类二元预测概率。则 u_i 属于 k_1 类年龄的概率 $\tilde{p}_{a,i}^{k_1}$, u_i 属于 k_2 类性别的概率 $\tilde{p}_{g,i}^{k_2}$ 和 u_i 属于 k_3 类种族的概率 $\tilde{p}_{r,i}^{k_3}$ 表示为

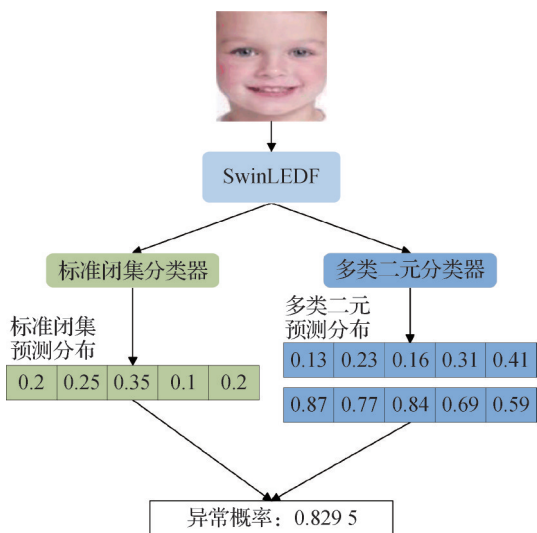


图3 开集半监督学习

Fig. 3 Open-set semi-supervised learning

$$\tilde{p}_{a,i}^{k_1} = p_{a,i}^{k_1,w} \times p_{a,i}^{k_1,u} \quad (22)$$

$$\tilde{p}_{g,i}^{k_2} = p_{g,i}^{k_2,w} \times o_{g,i}^{k_2,u} \quad (23)$$

$$\tilde{p}_{r,i}^{k_3} = p_{r,i}^{k_3,w} \times o_{r,i}^{k_3,u} \quad (24)$$

那么,可以得到 u_i 的属于年龄、性别和种族的异常类别的概率 $S_{a,i}$ 、 $S_{g,i}$ 、 $S_{r,i}$,具体为

$$S_{a,i} = 1 - \sum_{k_1=1}^{K_1} \tilde{p}_{a,i}^{k_1} \quad (25)$$

$$S_{g,i} = 1 - \sum_{k_2=1}^{K_2} \tilde{p}_{g,i}^{k_2} \quad (26)$$

$$S_{r,i} = 1 - \sum_{k_3=1}^{K_3} \tilde{p}_{r,i}^{k_3} \quad (27)$$

2)自适应阈值和伪标签生成。当确认样本为类内数据后,根据置信度阈值判断其是否满足伪标签条件。高置信度阈值可以确保伪标签质量,但可能丢弃一些正确伪标签;低置信度阈值能利用更多伪标签,但增加了引入错误伪标签的风险。而手工调整阈值既复杂又费时。为此,本文提出一种动态调整置信度阈值的方法,对出现频率较高的类别,使用较大的置信度阈值以防止过拟合;对出现频率较低的类别,使用较小的置信度阈值以提高模型对这些类别的敏感性,并降低将这些类别错误分类的风险。这种方法使模型能够充分考虑所有类别,有效降低了调整置信度阈值的复杂度和时间成本,同时避免模型对特定类别的过拟合风险,从而提升模型的性能。

在本文中,首先给定年龄、性别和种族估计的初始置信度阈值 τ_1 、 τ_2 和 τ_3 ,然后计算在 t 时满足初始

阈值条件的性别类别 k_2 和种族类别 k_3 的出现频率 $B_{g,t}(k_2)$ 和 $B_{r,t}(k_3)$ 。具体为

$$B_{g,t}(k_2) = \sum_{i=1}^{\mu N} \mathbb{I}(\max(p_{g,i}^w) > \tau_2) \times \mathbb{I}(\operatorname{argmax}(p_{g,i}^w)) \quad (28)$$

$$B_{r,t}(k_3) = \sum_{i=1}^{\mu N} \mathbb{I}(\max(p_{r,i}^w) > \tau_3) \times \mathbb{I}(\operatorname{argmax}(p_{r,i}^w)) \quad (29)$$

式中, $\mathbb{I}(\cdot)$ 为指示函数,满足条件时,其值为1,不满足条件时,其值为0。

针对年龄估计问题,由于相邻的年龄类别特征相似,所以不再统计单个类别的出现频率,而是考虑 k_1 类别及与其年龄差距不超过 m 的类别,计算出现频率 $B_{a,t}(k_1)$,具体为

$$B_{a,t}(k_1) = \sum_{i=1}^{\mu N} \mathbb{I}(\max(p_{a,i}^w) > \tau_1) \times \mathbb{I}(|\operatorname{argmax}(p_{a,i}^w) - k_1| \leq m) \quad (30)$$

然后通过以下方式进行归一化,使其范围在0~1之内。具体为

$$X_{a,t}(k_1) = a + \frac{B_{a,t}(k_1)}{2\max(B_{a,t} + B_{a,t}(k_1))} \quad (31)$$

$$X_{g,t}(k_2) = a + \frac{B_{g,t}(k_2)}{2\max(B_{g,t} + B_{g,t}(k_2))} \quad (32)$$

$$X_{r,t}(k_3) = a + \frac{B_{r,t}(k_3)}{2\max(B_{r,t} + B_{r,t}(k_3))} \quad (33)$$

最后,使用其缩放初始置信度阈值,得到年龄、性别和种族的自适应置信度阈值 $T_{a,t}(k_1)$ 、 $T_{g,t}(k_2)$ 和 $T_{r,t}(k_3)$ 。具体为

$$T_{a,t}(k_1) = X_{a,t}(k_1) \times \tau_1 \quad (34)$$

$$T_{g,t}(k_2) = X_{g,t}(k_2) \times \tau_2 \quad (35)$$

$$T_{r,t}(k_3) = X_{r,t}(k_3) \times \tau_3 \quad (36)$$

式中, a 为超参数,可避免在准确率较低时阈值为0。

当样本 u_i 在弱增强下的预测不属于异常值,即异常概率小于异常数据阈值 δ ,并且置信度超过自适应阈值时,可将此预测作为伪标签。即此时年龄的预测值 $\hat{p}_{a,i} = \operatorname{argmax}(p_{a,i}^w)$ 满足条件

$$M_{a,i} = \mathbb{I}(\max(p_{a,i}^w) > T_{a,t}(\hat{p}_{a,i})) \times \mathbb{I}(S_{a,i} < \delta) \quad (37)$$

当 $M_{a,i}$ 值为1时, $\hat{p}_{a,i}$ 为年龄伪标签。同理,当性别的预测值 $\hat{p}_{g,i}$ 满足条件 $M_{g,i}$ 时,为性别伪标签;当种族的预测值 $\hat{p}_{r,i}$ 满足条件 $M_{r,i}$ 时,为种族伪标签。

3)标准闭集分类器和多类二元分类器优化。在确定伪标签后,通过一致性正则化损失函数对标准闭集分类器进行优化,可以得到年龄、性别和种族的标准闭集分类器上的一致性正则化损失 L_{uc-a} 、 L_{uc-g} 和

L_{uc-r} 具体为

$$L_{uc-a} = -\frac{1}{\mu N} \sum_{i=1}^{\mu N} M_{a,i} \left(\sum_{k_1=1}^{K_1} p_{a,i}^{k_1,s} \times \log\left(\frac{p_{a,i}^{k_1,s}}{\hat{p}_{a,i}^{k_1,s}}\right) \right) \quad (38)$$

$$L_{uc-g} = -\frac{1}{\mu N} \sum_{i=1}^{\mu N} M_{g,i} \left(\sum_{k_2=1}^{K_2} p_{g,i}^{k_2,s} \times \log\left(\frac{p_{g,i}^{k_2,s}}{\hat{p}_{g,i}^{k_2,s}}\right) \right) \quad (39)$$

$$L_{uc-r} = -\frac{1}{\mu N} \sum_{i=1}^{\mu N} M_{r,i} \left(\sum_{k_3=1}^{K_3} p_{r,i}^{k_3,s} \times \log\left(\frac{p_{r,i}^{k_3,s}}{\hat{p}_{r,i}^{k_3,s}}\right) \right) \quad (40)$$

式中, $p_{a,i}^{k_1,s}$ 、 $p_{g,i}^{k_2,s}$ 和 $p_{r,i}^{k_3,s}$ 分别为模型的标准闭集分类器输出的 u_i 的强增强版本属于 k_1 类年龄、 k_2 类性别和 k_3 类种族的概率。

标准闭集分类器的总一致性正则化损失 L_{uc} 为

$$L_{uc} = L_{uc-a} + L_{uc-g} + L_{uc-r} \quad (41)$$

在年龄估计这类复杂问题中, 预测分布往往存在模糊不清的情况, 导致某些类别的预测置信度远低于阈值, 无标签数据无法得到有效利用。为充分利用这部分数据, 采用负学习方法, 自适应地为无标签数据生成负标签, 指示当前输入不属于的类别。这些负标签有助于校准错误预测并提升模型性能。其中, 当第 k_1 类的预测概率低于负标签阈值 θ , 或排在预测概率最低的 c 个类别中时, 将其视为负标签; 此外, 考虑到年龄估计符合高斯分布, 若某一类别与 k_1 类的年龄差距不超过 b , 其预测概率最低且低于负标签阈值 θ , 那么 k_1 类也被视为负标签。则年龄标准闭集分类器上的负学习损失 L_n 为

$$L_n = -\frac{1}{\mu N} \sum_{i=1}^{\mu N} \sum_{k_1=1}^{K_1} (\mathbb{I}(p_{a,i}^{k_1,w} < \theta) \cup (\mathbb{I}(\min(p_{a,i}^w) < \theta) \times \mathbb{I}(|\operatorname{argmin}(p_{a,i}^w) - k_1| \leq b)) \cup (\operatorname{Rank}^\uparrow(p_{a,i}^{k_1,w}) < c)) \times \log(1 - p_{a,i}^{k_1,s}) \quad (42)$$

式中, \cup 表示并集操作, $\operatorname{Rank}^\uparrow$ 表示按照升序排序。

标准闭集分类器的总损失 $L_{u-close}$ 为

$$L_{u-close} = L_{uc} + \lambda_n L_n \quad (43)$$

式中, λ_n 为超参数, 代表 L_n 的权重系数。

年龄、性别和种族多类二元分类器上的一致性正则化损失 L_{um-a} 、 L_{um-g} 和 L_{um-r} 为

$$L_{um-a} = \frac{1}{\mu N} \sum_{i=1}^{\mu N} \sum_{k_1=1}^{K_1} |o_{a,i}^{k_1,w} - o_{a,i}^{k_1,s}|^2 \quad (44)$$

$$L_{um-g} = \frac{1}{\mu N} \sum_{i=1}^{\mu N} \sum_{k_2=1}^{K_2} |o_{g,i}^{k_2,w} - o_{g,i}^{k_2,s}|^2 \quad (45)$$

$$L_{um-r} = \frac{1}{\mu N} \sum_{i=1}^{\mu N} \sum_{k_3=1}^{K_3} |o_{r,i}^{k_3,w} - o_{r,i}^{k_3,s}|^2 \quad (46)$$

式中, $o_{a,i}^{k_1,s}$ 、 $o_{g,i}^{k_2,s}$ 和 $o_{r,i}^{k_3,s}$ 分别为模型的多类二元分类器

输出的 u_i 的强增强版本属于 k_1 类年龄、 k_2 类性别和 k_3 类种族的概率。

多类二元分类器的总一致性正则化损失 L_{um} 为

$$L_{um} = L_{um-a} + L_{um-g} + L_{um-r} \quad (47)$$

最终, 总损失 L_{train} 包含有标签数据损失、标准闭集分类器损失和多类二元分类器损失, 具体为

$$L_{\text{train}} = L_s + \lambda_c L_{u-close} + \lambda_m L_{um} \quad (48)$$

式中, λ_c 和 λ_m 为超参数, 表示 $L_{u-close}$ 和 L_{um} 的权重系数。

2.2.3 开集半监督多任务学习算法

开集半监督多任务学习算法具体流程步骤如下:

预训练:

1) 输入: 有标签数据集 $L = \{(x_i, y_{a,i}, y_{g,i}, y_{r,i}) : i \in (1, \dots, N)\}$ 。

2) 计算有标签样本 x_i 的年龄、性别和种族标准闭集预测分布 $p_{a,i}$ 、 $p_{g,i}$ 和 $p_{r,i}$ 以及年龄、性别和种族多类二元预测分布 $o_{a,i}$ 、 $o_{g,i}$ 和 $o_{r,i}$ 。

3) 计算标准闭集分类器损失 $L_x = L_{x-a} + L_{x-g} + L_{x-r}$ 。

4) 计算多类二元分类器损失 $L_{mb} = L_{mb-a} + L_{mb-g} + L_{mb-r}$ 。

5) 计算预训练的总损失 $L_s = L_x + \lambda L_{mb}$ 。

微调:

6) 输入: 有标签数据集 $L = \{(x_i, y_{a,i}, y_{g,i}, y_{r,i}) : i \in (1, \dots, N)\}$ 、无标签数据集 $U = \{u_i : i = (1, \dots, \mu N)\}$; 年龄、种族、性别伪标签初始阈值 τ_1 、 τ_2 、 τ_3 ; 异常数据阈值 δ 、负标签阈值 θ 、年龄距离 m 、伪标签年龄距离 b 、伪标签排序类别数 c 。

有标签学习:

7) 计算有标签样本 x_i 的年龄、性别和种族标准闭集预测分布 $p_{a,i}$ 、 $p_{g,i}$ 和 $p_{r,i}$ 以及年龄、性别和种族多类二元预测分布 $o_{a,i}$ 、 $o_{g,i}$ 和 $o_{r,i}$ 。

8) 计算有标签数据集的闭集分类器损失 $L_x = L_{x-a} + L_{x-g} + L_{x-r}$ 。

9) 计算有标签数据集的多类二元分类器损失 $L_{mb} = L_{mb-a} + L_{mb-g} + L_{mb-r}$ 。

10) 计算有标签数据集上总损失 $L_s = L_x + \beta L_{mb}$ 。

无标签学习:

11) 计算无标签样本 u_i 经过弱增强处理后的年龄、性别和种族标准闭集预测分布 $p_{a,i}^w$ 、 $p_{g,i}^w$ 和 $p_{r,i}^w$ 以及年龄、性别、种族多类二元预测分布 $o_{a,i}^w$ 、 $o_{g,i}^w$ 和 $o_{r,i}^w$; 计

算无标签数据 u_i 经过增强处理后的年龄、性别和种族标准闭集预测分布 $p_{a,i}^s$ 、 $p_{g,i}^s$ 和 $p_{r,i}^s$ 以及年龄、性别和种族多类二元预测分布 $o_{a,i}^s$ 、 $o_{g,i}^s$ 和 $o_{r,i}^s$ 。

12) 计算 u_i 属于年龄、性别和种族异常数据概率 $S_{a,i}$ 、 $S_{g,i}$ 和 $S_{r,i}$ 。

13) 计算年龄、性别和种族自适应置信度阈值 $T_{a,t}(k_1)$ 、 $T_{g,t}(k_2)$ 和 $T_{r,t}(k_3)$ 。

14) 选择确定伪标签。

15) 将伪标签纳入到学习中, 计算标准闭集分类器的一致性正则化损失 $L_{uc} = L_{uc-a} + L_{uc-g} + L_{uc-r}$ 。

16) 计算年龄标准闭集分类器的负学习损失 L_n 。

17) 计算标准闭集分类器的总损失函数 $L_{u-close} = L_{uc} + \lambda_n L_n$ 。

18) 计算多类二元分类器的损失 $L_{um} = L_{um-a} + L_{um-g} + L_{um-r}$ 和微调总损失 $L_{train} = L_s + \lambda_c L_{u-close} + \lambda_m L_{um}$ 。

3 实验

3.1 数据集和评价指标

3.1.1 数据集

本文使用 CelebA 数据集作为无标签数据集, 分别使用 MORPH 数据集和 UTKface 数据集作为有标签数据集进行实验。

1) MORPH 数据集 (Ricanek 和 Tesafaye, 2006)。MORPH 数据集是一个纵向人脸数据库, 包含来自 13 617 人的 55 134 幅图像, 涵盖年龄、性别、种族、身高和体重等信息, 年龄范围为 16~77 岁。该数据集涉及 5 个种族: 非洲裔、亚洲裔、拉美裔 (或西班牙裔)、高加索人种和其他。本文将数据集随机分为两个非重叠集, 即训练集 (80%) 和验证集 (20%)。

2) UTKface 数据集 (Zhang 等, 2017)。UTKface 数据集包含 23 708 幅面部图像, 涵盖年龄、性别和种族信息, 年龄范围为 0~116 岁。种族包括 5 个类别: 白人、黑人、亚洲人、印度人和其他人。这些图像拍摄于自然环境中, 涵盖不同的拍照姿势、照明条件、遮挡程度、分辨率和面部表情。本文将数据集随机分为两个非重叠集, 即训练集 (80%) 和验证集 (20%)。

3) CelebA 数据集 (Liu 等, 2015)。CelebA 数据集是由香港中文大学多媒体实验室发布的大规模人脸属性开放数据集, 包含来自 10 177 个名人的 202 599 幅图像, 每幅图像都有着 40 个属性标签, 涵盖年龄、性别、表情等多种特征。其中年龄属性以二

分类形式标注为“年轻”和“非年轻”。CelebA 数据集不仅包括了多样化的背景、姿势以及光照条件, 还涵盖了广泛的种族分布, 包含白人、黑人、印度人、亚洲人等种族 (Kärkkäinen 和 Joo, 2021), 可为人脸属性分析、生成和识别等任务提供丰富的实验数据。

此外, 如图 4—图 6 所示, 在实验开始前需对数据集进行预处理, 去除原始图像中不利于进行年龄估计的背景信息。本文采用公开可用的 MTCNN (multitask cascaded convolutional network) (Zhang 等, 2016) 模型检测每幅图像中的 5 个关键人脸标志: 两个眼睛中心、鼻尖和两个嘴角。随后, 根据检测到的人脸标志点进行仿射变换, 将每个人脸对齐并调整为 230×230 像素的标准直立姿势人脸图像。



图4 MORPH数据集预处理

Fig. 4 MORPH dataset preprocessing

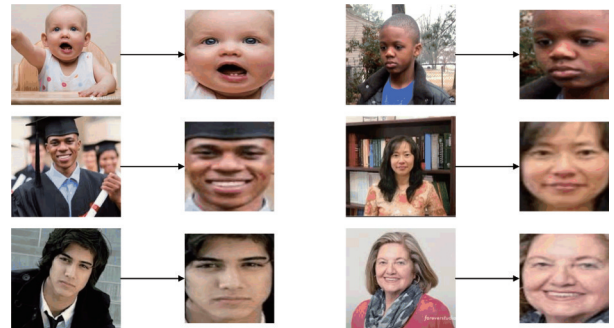


图5 UTKface数据集预处理

Fig. 5 UTKface dataset preprocessing

3.1.2 评价指标

本文采用年龄累计分数和平均绝对年龄估计作为评价指标。

年龄累计分数 (cumulative score, CS) 定义为在测试数据集中, 模型预测年龄与真实年龄标签的绝对误差 e 小于给定范围 j 的样本数量 $N_{e \leq j}$ 占测试示例样本总数 N 的百分比。即

$$CS(j) = \frac{N_{e \leq j}}{N} \times 100\% \quad (49)$$

CS 值越大, 表示模型性能越优。本文选

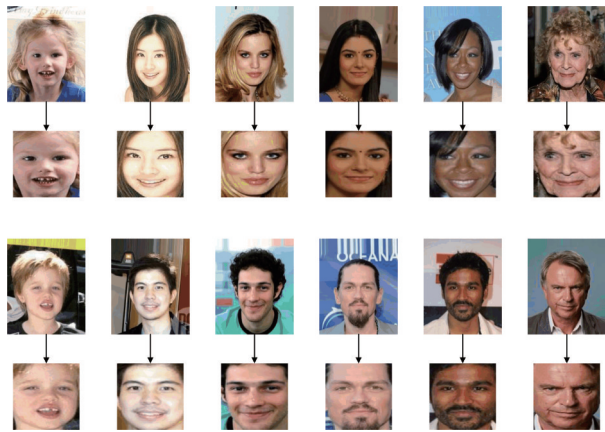


图6 CelebA数据集预处理

Fig. 6 CelebA dataset preprocessing

取 $j = 5$ 。

平均绝对误差 (mean absolute error, MAE) 定义为测试集上模型预测年龄 $\hat{y}_{a,i}$ 与真实年龄标签 $y_{a,i}$ 之间绝对误差的平均值。即

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_{a,i} - y_{a,i}| \quad (50)$$

式中, N 为测试样例总数。MAE 是评估人脸图像年龄估计算法性能的常用指标, MAE 值越小, 表示模型性能越好。

3.2 实验设置

本研究采用数据增强处理, 具体如下: 1) 在有标签数据和无标签数据的弱增强处理中, 所有训练图像随机裁剪为 224×224 像素, 并在水平方向进行随机翻转。2) 在无标签数据的强增强处理中, 图像随机裁剪为 224×224 像素, 并引入图像失真处理, 即每幅图像以 50% 的概率随机从调整图像亮度、对比度、饱和度、清晰度、减少调数量、随机翻转和随机裁剪等 10 种增强方法中选择 2 种操作执行, 然后对图像进行 Cutout 操作, 遮挡一个 16×16 的区域。3) 验证集的弱增强处理则是将图像先调整为 256×256 像素后进行中心裁剪, 以生成 224×224 像素的图像。

本文所有实验均在基于 Pytorch 2.1.0 框架的 NVIDIA GeForce RTX 4090 上进行, 选用 AdamW 优化器。

1) 针对不同数据集的预训练, 本文采用不同的训练策略。在 MORPH 数据集上, 初始学习率设置为 0.000 1, batch-size 设置为 64, 每训练 120 个 epoch, 学习率衰减为当前值的 1/10。训练在模型性

能不再提升或达到 360 个 epoch 时停止。在 UTKface 数据集上, 初始学习率设置为 0.000 1, batch-size 设置为 8, 每训练 20 个 epoch, 学习率衰减为当前值的 1/10。训练在模型性能不再提升或达到 120 个 epoch 时停止。

2) 在模型的微调过程中, 初始学习率设置为 0.000 1, 每 4 个 epoch 将学习率衰减为当前值的 1/10, 直至模型性能不再提升时停止训练。

3.3 对比实验

3.3.1 与其他人脸年龄估计方法的对比

如表 1 所示, 在 MORPH 数据集上, 本文方法在只进行有标签数据学习时, MAE 为 1.908, 这一性能仅次于 LRN 和 MCGRL; 同时进行有标签数据学习和无标签数据学习后, MAE 降至 1.885, 达到了最优性能。

如表 2 所示, 在 UTKface 数据集中, 只进行有标签数据学习时, MAE 为 4.343, 这一性能仅次于 GroupFace; 同时进行有标签数据学习和无标签数据学习后, MAE 降低至 4.246, 表现优于其他对比方法。这表明, 本文提出的方法在 MORPH 和 UTKface 数据集上取得了较为先进的性能, 且展现了从无标签数据集中提取有效信息的能力。

3.3.2 与其他半监督方法的对比

在相同的训练模型和参数设置下, 将本文提出的开集半监督多任务学习方法与闭集半监督学习方法 FixMatch、FullMatch、Fullflex、SoC4SS-FGVC 以及开集半监督学习方法 OpenMatch、IOMatch 在 MORPH 和 UTKface 数据集数据集进行对比, 结果如表 3 和表 4 所示。可以看出, 在 MORPH 数据集中, 其他半监督学习方法的模型性能反而有所下降, 而本文方法仍能有效提取无标签数据信息, 提升模型性能。在 UTKface 数据集中, 所有半监督学习方法均能有效提升模型性能, 但本文方法表现更为优越。这些结果表明, 针对人脸年龄估计任务, 本文的开集半监督多任务学习方法优于其他对比方法, 验证了本文方法的有效性。

3.4 消融实验

3.4.1 LEFF 和 DFN 模块在单任务和多任务学习中的有效性

如表 5 和表 6 所示, 无论是在单任务学习还是在多任务学习中, 相较于基准的 Swin Transformer 模型, 单独引入 DFN 模块、LEFF 模块或两者联合使用, 均

表1 在MORPH数据集上不同人脸年龄估计方法的对比
Table 1 Comparison with different facial age estimation methods on the MORPH dataset

方法	MAE
Zhao等人(2021)	2.73
CORAL(Cao等,2020)	2.64
ADPF(Wang等,2022)	2.54
Shi等人(2023)	2.45
Mean-Variance Loss+softmax Loss(Pan等,2018)	2.41/2.16*
AL-ROR-34(Zhang等,2020)	2.36*
MDL(Pan等,2018)	2.31
DCT(Bao等,2023)	2.28/2.17*
PML(Deng等,2021b)	2.15
EgroupNet(Duan等,2020)	2.13
GroupFace(Zhang等,2025)	2.09
FP-Age(Lin等,2022)	2.04/1.90‡
SGL(Liu等,2023)	2.01
DLDL-V2(Gao等,2018)	1.969*
DCN(Kong等,2022)	1.946 2
AVDL(Wen等,2020)	1.94*
DHAA(Tan等,2019)	1.908
LRN(Li等,2020)	1.905*
MCGRL(Shou等,2025)	1.89
本文(有标签数据学习)	1.908
本文(有标签数据学习+无标签数据学习)	1.885

注:加粗字体表示最优结果。*表示在IMDB-WIKI(Rothe等,2015)上预训练;‡表示在IMDB-Clean(Lin等,2022)上预训练;#表示在MS-Celeb-1M上预训练(Guo等,2016)。

能有效提升年龄估计的性能,降低平均绝对年龄误差MAE,提升年龄累积分数CS(5)。尤其是DFN和LEFF模块联合使用时,模型性能的提升超过单独使用其中一个模块的性能,验证了两者协同作用对于增强年龄估计精度的贡献。此外,在相同模型架构下,多任务学习的表现明显优于单任务学习,这可能得益于多任务学习能综合考虑性别和种族信息以优化年龄估计。实验结果表明,DFN模块、LEFF模块以及多任务学习均能提升年龄估计模型性能。

3.4.2 多任务学习的有效性

为了验证性别和种族属性对年龄估计结果的影响,并证明多任务学习方法的有效性,本文设计了多

表2 在UTKface数据集上不同人脸年龄估计方法的对比

Table 2 Comparison with different facial age estimation methods on the UTKface dataset

方法	MAE
CORAL(Cao等,2020)	5.47
Lin等人(2024)	4.82
DCDT(Gustafsson等,2019)	4.65
Equal Width(Berg等,2021)	4.58*
Randomized Bins(Berg等,2021)	4.55*
Moving Window Regression(Shin等,2022)	4.37
GroupFace(Zhang等,2025)	4.32*
本文(有标签数据学习)	4.343
本文(有标签数据学习+无标签数据学习)	4.246

注:加粗字体表示最优结果。*表示在IMDB-WIKI(Rothe等,2015)上预训练。

表3 在MORPH数据集上不同半监督学习方法的对比
Table 3 Comparison with different semi-supervised learning methods on the MORPH dataset

半监督学习方法	MAE	CS(5)/%
FixMatch(Sohn等,2020)	2.026	90.60
FullMatch(Chen等,2023b)	2.022	90.80
Fullflex(Chen等,2023b)	2.020	90.62
SoC4SS-FGVC(Duan等,2024)	2.024	90.60
OpenMatch(Saito等,2021)	2.019	90.82
IOMatch(Li等,2023a)	2.018	90.83
本文	1.885	92.06

注:加粗字体表示各列最优结果。

表4 UTKface数据集上不同半监督学习方法的对比
Table 4 Comparison with different semi-supervised learning methods on the UTKface dataset

半监督方法	MAE	CS(5)/%
FixMatch(Sohn等,2020)	4.285	68.98
FullMatch(Chen等,2023b)	4.271	69.09
Fullflex(Chen等,2023b)	4.261	69.40
SoC4SS-FGVC(Duan等,2024)	4.286	69.01
OpenMatch(Saito等,2021)	4.276	69.05
IOMatch(Li等,2023a)	4.271	69.11
本文	4.246	69.75

注:加粗字体表示各列最优结果。

表5 LEFF和DFN模块在单任务学习中的有效性

Table 5 Effectiveness of LEFF and DFN modules in single-task learning

模块		MORPH		UTKface	
DFN	LEFF	MAE	CS(5)/%	MAE	CS(5)/%
-	-	2.080	88.90	4.440	68.53
√	-	2.069	90.12	4.423	69.00
-	√	2.059	90.49	4.416	69.07
√	√	2.049	90.90	4.381	69.20

注:加粗字体表示各列最优结果,“√”表示采用,“-”表示未采用。

表6 LEFF和DFN模块在多任务学习中的有效性

Table 6 Effectiveness of LEFF and DFN modules in multi-task learning

模块		MORPH		UTKface	
DFN	LEFF	MAE	CS(5)/%	MAE	CS(5)/%
-	-	1.942	90.42	4.406	69.50
√	-	1.919	91.30	4.403	69.53
-	√	1.912	91.43	4.391	69.70
√	√	1.908	92.04	4.343	69.74

注:加粗字体表示各列最优结果,“√”表示采用,“-”表示未采用。

组实验。具体而言,在 MORPH 和 UTKface 数据集上分别测试 4 种情况:1)仅进行年龄估计;2)进行年龄估计和性别估计;3)进行年龄估计和种族估计;4)进行年龄、性别和种族估计。实验结果如表 7 所示。可以看出,仅进行年龄估计时性能最差;引入性别或种族属性均能提升年龄估计性能,且种族属性的引入对性能的提升优于性别属性;而同时引入性别和种族属性时,模型达到了最优性能。实验结果表明,性别和种族属性会对年龄估计的结果造成影响,多任务学习方法能够利用属性的相关性有效提升年龄估计的精确度。

3.4.3 权重系数 β 、 λ_c 和 λ_m 对年龄估计结果的影响

本文在预训练阶段在 MORPH 和 UTKface 数据集上对参数 β 进行系统实验,在 0~1 范围内以 0.1 为间隔进行取值,以平均绝对误差(MAE)作为评价指标,以评估其对模型性能的影响。

如表 8 所示,当 β 取 0.1 时,模型性能略优于取 0 时的情况,实现最佳性能,之后,随着 β 取值的进一

表7 多任务学习的有效性

Table 7 Effectiveness of multi-task learning

方法	MORPH		UTKface	
	MAE	CS(5)/%	MAE	CS(5)/%
年龄估计	2.049	90.9	4.381	69.2
年龄估计+性别估计	2.026	91.02	4.373	69.42
年龄估计+种族估计	1.979	91.45	4.361	69.51
年龄估计+性别估计+种族估计	1.908	92.06	4.343	69.74

注:加粗字体表示各列最优结果。

步增加,模型性能呈现逐渐下降的趋势。值得注意的是,当 β 取 0 时,模型未对多类二元分类器进行优化,而多类二元分类器在后续微调过程中对异常数据过滤和伪标签生成具有重要作用。因此,基于性能表现和模型功能需求的综合考虑,将 β 值设定为 0.1,并在后续实验和微调过程中保持固定。

表8 系数 β 取值对年龄估计结果的影响Table 8 Influence of the value of coefficient β on age estimation results

β	MORPH	UTKface
0	1.910	4.344
0.1	1.908	4.343
0.2	1.919	4.362
0.3	1.934	4.386
0.4	1.949	4.394
0.5	1.969	4.401
0.6	1.996	4.417
0.7	2.014	4.428
0.8	2.031	4.443
0.9	2.048	4.456
1	2.065	4.489

注:加粗字体表示各列最优结果。

为研究损失函数中权重系数 λ_c 和 λ_m 对年龄估计结果的影响及其相互作用,分别对 λ_c 和 λ_m 取值为 0、0.5、1、1.5 和 2,构建了 25 种参数组合,并在 MORPH 数据集和 UTKface 数据集上进行实验。实验以平均绝对误差(MAE)作为评价指标,结果如表 9 和表 10 所示。实验结果表明,当 λ_c 和 λ_m 均取 1 时,模型取得最低的 MAE 值,性能达到最优。其中,

λ_c 和 λ_m 分别表示在微调过程中无标签数据集上闭集分类器和多类二元分类器的损失权重。在本文中,有标签部分损失权重固定为1,当 λ_c 和 λ_m 都为1时,有标签数据的损失、无标签数据集上闭集分类器和多类二元分类器的损失贡献达到均衡。这种均衡使得模型能够充分利用有标签数据中的标签信息,有效挖掘无标签数据中的信息,促进多类二元分类器和闭集分类器的协同优化,从而提升模型的年龄估计性能。这一结果表明,在微调过程中,通过合理的参数设置同时优化闭集分类器和多类二元分类器,能够有效利用闭集分类器和多类二元分类器的协同作用,进而提升模型的年龄估计性能。

表9 在MORPH数据集上系数 λ_c 和 λ_m 取值对年龄估计结果的影响

Table 9 Influence of the values of coefficients λ_c and λ_m on age estimation results on the MORPH dataset

λ_m	λ_c				
	0	0.5	1	1.5	2
0	2.016	1.956	1.911	1.921	1.933
0.5	2.010	1.932	1.899	1.916	1.925
1	2.003	1.921	1.885	1.911	1.917
1.5	2.011	1.934	1.909	1.914	1.923
2	2.014	1.952	1.913	1.924	1.941

注:加粗字体表示最优结果。

表10 在UTKface数据集上系数 λ_c 和 λ_m 取值对年龄估计结果的影响

Table 10 Influence of the values of coefficients λ_c and λ_m on age estimation results on the UTKface dataset

λ_m	λ_c				
	0	0.5	1	1.5	2
0	4.305	4.283	4.274	4.277	4.285
0.5	4.293	4.272	4.257	4.265	4.278
1	4.285	4.265	4.246	4.253	4.269
1.5	4.296	4.274	4.255	4.271	4.273
2	4.309	4.281	4.269	4.279	4.283

注:加粗字体表示最优结果。

3.4.4 自适应阈值和负学习方法的有效性

如表11所示,在UTKface数据集上,基准开集半监督多任务学习微调后,平均绝对年龄误差MAE下

降,但年龄累积分数CS(5)变差。引入自适应阈值方法和负学习方法后,MAE进一步降低,CS(5)逐步提升,两者同时使用时,MAE和CS(5)均达到最优值,超过未微调时的结果。对于MORPH数据集,微调后的模型性能反而变差。引入自适应阈值和负学习方法后,能够缓解模型性能变差的情况,二者同时使用时,性能优于未微调模型。这可能是由于MORPH数据集与CelebA数据集差异较大,半监督学习会引入一些错误信息,但自适应阈值方法和负学习方法能有效减小这些负面影响,使得模型在相差较大的无标签数据集中依然能提取到有效信息。综上所述,自适应阈值和负学习方法在开集半监督多任务学习中的应用,有助于从无标签数据集中提取有效信息,提升模型的人脸年龄估计性能。

表11 自适应阈值方法和负学习方法在开集半监督多任务学习中的有效性

Table 11 Effectiveness of adaptive threshold method and negative learning method in semi-supervised learning

方法		MORPH		UTKface	
自适应阈值	负学习	MAE	CS(5)/%	MAE	CS(5)/%
-	-	2.000	90.70	4.273	69.30
√	-	1.960	91.22	4.261	69.42
-	√	1.934	91.43	4.254	69.51
√	√	1.885	92.06	4.246	69.75

注:加粗字体表示各列最优结果,“√”表示采用,“-”表示未采用。

3.5 可视化结果

为了更直观地展示模型的性能,本文在MORPH和UTKface数据集上进行实验,将模型预测的年龄与真实年龄标签进行对比。如图7所示,第1行的黑色数字代表人脸图像对应的真实标签,第2行展示了图像的年龄预测结果,其中蓝色数字表示正确预测,红色数字表示错误预测。实验结果表明,本文方法在MORPH数据集上的年龄预测存在一定偏差,但大多数预测误差均小于2岁。在UTKface数据集上,本文方法在部分图像上存在错误预测,但错误预测主要集中在年龄较大的类别,这可能是由极端年龄段的面部特征较为复杂、个体差异大,以及该数据集在这些年龄段的样本数量较少造成的。

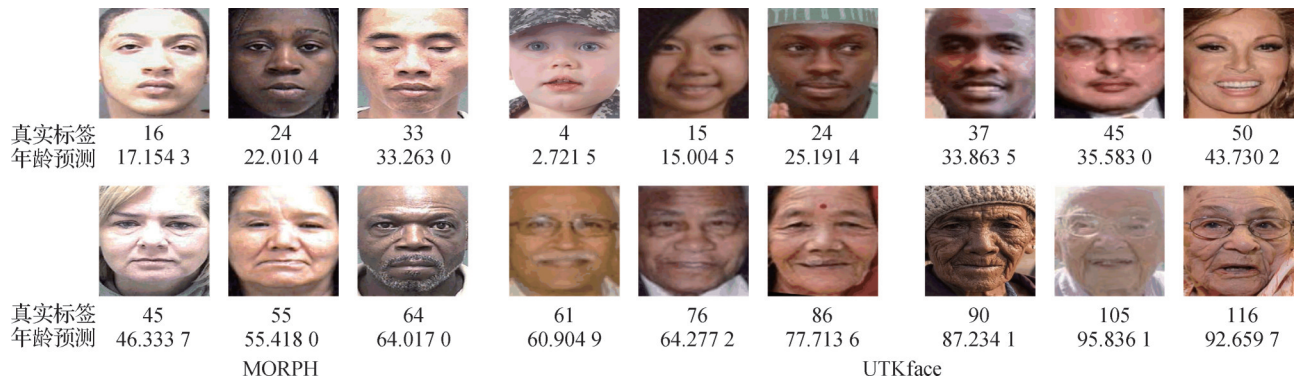


图7 可视化结果图

Fig. 7 Visualization of the results

4 结 论

针对人脸年龄估计领域有标签数据集稀缺,现有半监督学习方法无法有效利用无标签数据集的问题,提出一种面向人脸年龄估计的开集半监督多任务学习方法。首先,为增强模型的特征提取能力,提出SwinLEDF模型,通过Swin Transformer从人脸图像中学习整体年龄趋势和面部架构信息,同时引入LEFF模块和DFN模块有效捕捉与年龄相关的面部细微差异。其次,提出开集半监督多任务学习方法,模型同时进行性别、种族和年龄估计任务,利用标准闭集分类器和多类二元分类器联合排除异常数据干扰,并根据模型在各类别上的表现动态生成自适应阈值,以此筛选并生成性别、种族和年龄的伪标签,同时引入负学习方法生成负标签,然后将这些伪标签和负标签融入学习过程中,从而有效利用无标签数据集来提升年龄估计性能。

本文方法在MORPH和UTKface人脸年龄估计基准数据集上取得先进的性能,并展现了从无标签数据中提取有效信息的能力。本文方法体现了半监督学习处理复杂问题的优势,为实现精确年龄估计提供了新方案。但半监督学习通常需要大量计算资源和时间,且在无标签数据的有效利用方面仍有提升空间。因此,未来研究可进一步探索如何优化计算效率并提高无标签数据集的利用率。

参考文献(References)

Akbari A, Awais M, Fatemifar S, Khalid S S and Kittler J. 2024. RAaE: robust age estimation through subject anchoring with consis-

tency regularisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(3): 1603-1617 [DOI: 10.1109/TPAMI.2022.3187079]

Bao Z H, Tan Z C, Wan J, Ma X B, Guo G D and Lei Z. 2023. Divergence-driven consistency training for semi-supervised facial age estimation. *IEEE Transactions on Information Forensics and Security*, 18: 221-232 [DOI: 10.1109/TIFS.2022.3218431]

Berg A, Oskarsson M and O'Connor M. 2021. Deep ordinal regression with label diversity//*Proceedings of the 25th International Conference on Pattern Recognition (ICPR)*. Milan, Italy: IEEE: 2740-2747 [DOI: 10.1109/ICPR48806.2021.9412608]

Bhattacharya A R and Chakraborty S. 2022. Deep active learning with range feedback for facial age estimation//*Proceedings of 2022 International Joint Conference on Neural Networks (IJCNN)*. Padua, Italy: IEEE: 1-9 [DOI: 10.1109/IJCNN55064.2022.9892113]

Cao W Z, Mirjalili V and Raschka S. 2020. Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognition Letters*, 140: 325-331 [DOI: 10.1016/j.patrec.2020.11.008]

Chen H, Tao R, Fan Y, Wang Y D, Wang J D, Schiele B, Xie X, Raj B and Savvides M. 2023a. SoftMatch: addressing the quantity-quality trade-off in semi-supervised learning [EB/OL]. [2025-01-10]. <https://arxiv.org/pdf/2301.10921.pdf>

Chen Y H, Tan X, Zhao B R, Chen Z W, Song R J, Liang J J and Lu X Q. 2023b. Boosting semi-supervised learning by exploiting all unlabeled data//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, Canada: IEEE: 7548-7557 [DOI: 10.1109/CVPR52729.2023.00729]

De Brabandere B, Jia X, Tuytelaars T and Van Gool L. 2016. Dynamic filter networks [EB/OL]. [2025-01-10]. <https://arxiv.org/pdf/1605.09673.pdf>

Deng Y L, Teng S H, Fei L K, Zhang W and Rida I. 2021a. A multifeature learning and fusion network for facial age estimation. *Sensors*, 21(13): #4597 [DOI: 10.3390/s21134597]

Deng Z Y, Liu H, Wang Y X, Wang C Y, Yu Z K and Sun X H. 2021b.

- PML: progressive margin loss for long-tailed age classification//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 10498-10507 [DOI: 10.1109/CVPR46437.2021.01036]
- Du P, Zhao S Y, Sheng Z S, Li C P and Chen H. 2023. Semi-supervised learning via weight-aware distillation under class distribution mismatch//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 16364-16374 [DOI: 10.1109/ICCV51070.2023.01504]
- Duan M X, Li K L, Ouyang A J, Win K N, Li K Q and Tian Q. 2020. EGroupNet: a feature-enhanced network for age estimation with novel age group schemes. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 16(2): #42 [DOI: 10.1145/3379449]
- Duan Y, Zhao Z, Qi L, Zhou L P, Wang L and Shi Y H. 2024. Roll with the punches: expansion and shrinkage of soft label selection for semi-supervised fine-grained learning//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 11829-11837 [DOI: 10.1609/aaai.v38i10.29068]
- Fan Y, Kukleva A, Dai D X and Schiele B. 2023. SSB: simple but strong baseline for boosting performance of open-set semi-supervised learning//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 16022-16032 [DOI: 10.1109/ICCV51070.2023.01472]
- Gabor D. 1946. Theory of communication. Part I: the analysis of information. *Journal of the Institution of Electrical Engineers—Part III: Radio and Communication Engineering*, 93(26): 429-441 [DOI: 10.1049/ji-3-2.1946.0074]
- Gao B B, Zhou H Y, Wu J X and Geng X. 2018. Age estimation using expectation of label distribution learning//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: AAAI Press: 712-718 [DOI: 10.24963/ijcai.2018/99]
- Guehairia O, Ouamane A, Dornaika F and Taleb-Ahmed A. 2020. Deep random forest for facial age estimation based on face images//Proceedings of the 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP). El Oued, Algeria: IEEE: 305-309 [DOI: 10.1109/CCSSP49278.2020.9151621]
- Gui Q, Wu X T and Niu B N. 2022. Class-aware pseudo labeling for non-random missing labels in semi-supervised learning//Proceedings of 2022 IEEE Eighth International Conference on Multimedia Big Data (BigMM). Naples, Italy: IEEE: 138-143 [DOI: 10.1109/BigMM55396.2022.00031]
- Guo G D, Mu G W, Fu Y and Huang T S. 2009. Human age estimation using bio-inspired features//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 112-119 [DOI: 10.1109/CVPR.2009.5206681]
- Guo Y D, Zhang L, Hu Y X, He X D and Gao J F. 2016. MS-Celeb-1M: a dataset and benchmark for large-scale face recognition//Proceedings of the 14th European Conference on Computer Vision—ECCV 2016. Amsterdam, the Netherlands: Springer: 87-102 [DOI: 10.1007/978-3-319-46487-9_6]
- Gustafsson F K, Danelljan M, Bhat G and Schön T B. 2019. DCTD: deep conditional target densities for accurate regression [EB/OL]. [2025-01-10]. <https://arxiv.org/pdf/1909.12297v1.pdf>
- He R D, Han Z Y, Lu X K and Yin Y L. 2023. Safe-student for Safe deep semi-supervised learning with unseen-class unlabeled data//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 14565-14574 [DOI: 10.1109/CVPR52688.2022.01418]
- Jiang T, Chen L Y, Chen W Q, Meng W J and Qi P H. 2023. ReliaMatch: semi-supervised classification with reliable match. *Applied Sciences*, 13(15): #8856 [DOI: 10.3390/app13158856]
- Kang Z Q, Fini E, Nabi M, Ricci E and Alahari K. 2023. A soft nearest-neighbor framework for continual semi-supervised learning//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 11834-11843 [DOI: 10.1109/ICCV51070.2023.01090]
- Kärkkäinen K and Joo J. 2021. FairFace: face attribute dataset for balanced race, gender, and age for bias measurement and mitigation//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE: 1547-1557 [DOI: 10.1109/WACV48630.2021.00159]
- Kong C, Luo Q M and Chen G L. 2022. Learning deep contrastive network for facial age estimation//Proceedings of 2022 International Joint Conference on Neural Networks (IJCNN). Padua, Italy: IEEE: 1-7 [DOI: 10.1109/IJCNN55064.2022.9892308]
- Kwon Y H and Lobo D V. 1994. Age classification from facial images//Proceedings of 1994 IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 762-767 [DOI: 10.1109/CVPR.1994.323894]
- Li K, Lian Q Z and Gao C. 2023a. Deep semi-supervised learning with fine-grained dynamic weights and pseudo-label constraints//Proceedings of 2023 International Conference on Communications, Computing and Artificial Intelligence (CCCAI). Shanghai, China: IEEE: 1-6 [DOI: 10.1109/CCCAI59026.2023.00009]
- Li P P, Hu Y B, Wu X, He R and Sun Z N. 2020. Deep label refinement for age estimation. *Pattern Recognition*, 100: #107178 [DOI: 10.1016/j.patcog.2019.107178]
- Li Z K, Qi L, Shi Y H and Gao Y. 2023a. IOMatch: simplifying open-set semi-supervised learning with joint inliers and outliers utilization//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 15824-15833 [DOI: 10.1109/ICCV51070.2023.01454]
- Lin C Z, Gou J, Fan Z W and Liao Y X. 2024. A feature fusion-based ResNet using the pooling pyramid for age estimation//Proceedings of 2024 International Joint Conference on Neural Networks (IJCNN). Yokohama, Japan: IEEE: 1-8 [DOI: 10.1109/IJCNN60899.2024.

- 10650545]
- Lin Y M, Shen J, Wang Y J and Pantic M. 2022. FP-Age: leveraging face parsing attention for facial age estimation in the wild. *IEEE Transactions on Image Processing*, 34: 4767-4777 [DOI: 10.1109/TIP.2022.3155944]
- Liu H, Ma M, Gao Z X, Deng Z Y, Li F J and Li Z D. 2023. Siamese graph learning for semi-supervised age estimation. *IEEE Transactions on Multimedia*, 25: 9586-9596 [DOI: 10.1109/TMM.2023.3256065]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021. Swin Transformer: hierarchical vision Transformer using shifted windows//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]
- Liu Z W, Luo P, Wang X G and Tang X O. 2015. Deep learning face attributes in the wild//*Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE: 3730-3738 [DOI: 10.1109/ICCV.2015.425]
- Lowe D G. 1999. Object recognition from local scale-invariant features//*Proceedings of the 7th IEEE International Conference on Computer Vision*. Kerkyra, Greece: IEEE: 1150-1157 [DOI: 10.1109/ICCV.1999.790410]
- Ma Q K, Gao J Y, Zhan B, Guo Y P, Zhou J L and Wang Y. 2023. Rethinking safe semi-supervised learning: transferring the open-set problem to a close-set one//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 16324-16333 [DOI: 10.1109/ICCV51070.2023.01500]
- Ojala T, Pietikainen M and Harwood D. 1994. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions//*Proceedings of the 12th International Conference on Pattern Recognition*. Jerusalem, Israel: IEEE: 582-585 [DOI: 10.1109/ICPR.1994.576366]
- Pan H Y, Han H, Shan S G and Chen X L. 2018. Mean-variance loss for deep age estimation from a face//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 5285-5294 [DOI: 10.1109/CVPR.2018.00554]
- Qin J, Jiao Y, Li Z P and Mao Z Y. 2025. Age estimation of face image based on attention ConvLSTM model. *Computer Applications and Software*, 42(1): 383-390 (秦瑾, 焦勇, 李泽鹏, 毛智勇. 2025. 基于注意力 ConvLSTM 模型的人脸图像年龄估计研究. *计算机应用与软件*, 42(1): 383-390) [DOI: 10.3969/j.issn.1000-386x.2025.01.053]
- Ricanek K and Tesafaye T. 2006. MORPH: a longitudinal image database of normal adult age-progression//*Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*. Southampton, UK: IEEE: 341-345 [DOI: 10.1109/FGR.2006.78]
- 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 (ICCVW)*. Santiago, Chile: IEEE: 252-257 [DOI: 10.1109/ICCVW.2015.41]
- Saito K, Kim D and Saenko K. 2021. OpenMatch: open-set consistency regularization for semi-supervised learning with outliers [EB/OL]. [2025-01-10]. <https://arxiv.org/pdf/2105.14148.pdf>
- Saito K and Saenko K. 2021. OVANet: one-vs-all network for universal domain adaptation//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, Canada: IEEE: 8980-8989 [DOI: 10.1109/ICCV48922.2021.00887]
- Shi C J, Zhao S W, Zhang K and Feng X H. 2023. Multi-task multi-scale attention learning-based facial age estimation. *IET Signal Processing*, 17(2): #12190 [DOI: 10.1049/sil2.12190]
- Shin N H, Lee S H and Kim C S. 2022. Moving window regression: a novel approach to ordinal regression//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, USA: IEEE: 18739-18748 [DOI: 10.1109/CVPR52688.2022.01820]
- Shou Y T, Cao X Y, Liu H and Meng D Y. 2025. Masked contrastive graph representation learning for age estimation. *Pattern Recognition*, 158: #110974 [DOI: 10.1016/j.patcog.2024.110974]
- Singh A and Chakraborty S. 2021. Deep active learning with relative label feedback: an application to facial age estimation//*Proceedings of 2021 International Joint Conference on Neural Networks (IJCNN)*. Shenzhen, China: IEEE: 1-8 [DOI: 10.1109/IJCNN52387.2021.9533657]
- Sohn K, Berthelot D, Li C L, Zhang Z Z, Carlini N, Carlini N, Cubuk E D, Kurakin A, Zhang H and Raffel C. 2020. FixMatch: simplifying semi-supervised learning with consistency and confidence//*Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: 596-608 [DOI: 10.5555/3495724.3495775]
- Tan Z C, Yang Y, Wan J, Guo G D and Li S Z. 2019. Deeply-learned hybrid representations for facial age estimation//*Proceedings of the 28th International Joint Conference on Artificial Intelligence*. Macao, China: AAAI Press: 3548-3554 [DOI: 10.24963/ijcai.2019/492]
- Wang H Y, Sanchez V and Li C T. 2022. Improving face-based age estimation with attention-based dynamic patch fusion. *IEEE Transactions on Image Processing*, 31: 1084-1096 [DOI: 10.1109/TIP.2021.3139226]
- Wen X, Li B Y, Guo H Y, Liu Z W, Hu G S, Tang M and Wang J Q. 2020. Adaptive variance based label distribution learning for facial age estimation//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 379-395 [DOI: 10.1007/978-3-030-58592-1_23]
- Yang L H, Zhao Z, Qi L, Qiao Y, Shi Y H and Zhao H S. 2023. Shrinking class space for enhanced certainty in semi-supervised learning//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France: IEEE: 16141-16150 [DOI:

- 10.1109/ICCV51070.2023.01483]
- Yao Y, Shen J Y, Xu J, Zhong B and Xiao L. 2022. CLS: cross labeling supervision for semi-supervised learning [EB/OL]. [2025-01-10]. <https://arxiv.org/pdf/2202.08502.pdf>
- Yuan K, Guo S P, Liu Z W, Zhou A J, Yu F W and Wu W. 2021. Incorporating convolution designs into visual transformers//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 559-568 [DOI: 10.1109/ICCV48922.2021.00062]
- Zhang K, Liu N, Yuan X F, Guo X Y, Gao C, Zhao Z B and Ma Z Y. 2020. Fine-grained age estimation in the wild with attention LSTM networks. IEEE Transactions on Circuits and Systems for Video Technology, 30 (9) : 3140-3152 [DOI: 10.1109/TCSVT.2019.2936410]
- Zhang K P, Zhang Z P, Li Z F and Qiao Y. 2016. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23 (10) : 1499-1503 [DOI: 10.1109/LSP.2016.2603342]
- Zhang Z F, Song Y and Qi H R. 2017. Age progression/regression by conditional adversarial autoencoder//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 4352-4360 [DOI: 10.1109/CVPR.2017.463]
- Zhang Y P, Shou Y T, Ai W, Meng T and Li K Q. 2025. GroupFace: imbalanced age estimation based on multi-hop attention graph convolutional network and group-aware margin optimization. IEEE Transactions on Information Forensics and Security, 20: 605-619 [DOI: 10.1109/TIFS.2024.3520020]
- Zhao Q L, Dong J Y, Yu H and Chen S. 2021. Distilling ordinal relation and dark knowledge for facial age estimation. IEEE Transactions on Neural Networks and Learning Systems, 32(7): 3108-3121 [DOI: 10.1109/TNNLS.2020.3009523]
- Zheng M K, You S, Huang L, Wang F, Qian C and Xu C. 2022. Sim-Match: semi-supervised learning with similarity matching//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 14451-14461 [DOI: 10.1109/CVPR52688.2022.01407]
- Zhu H, Hu B, Song Y N and Zhao X F. 2024. Curriculum paradigm based on the dynamic weights of samples for semi-supervised learning. Chinese High Technology Letters, 34(4): 342-355 (朱徽, 胡斌, 宋怡宁, 赵晓芳. 2024. 基于样本动态权重的课程式半监督学习方法. 高技术通讯, 34(4): 342-355) [DOI: 10.3772/j.issn.1002-0470.2024.04.002]

作者简介

张珂,男,教授,主要研究方向为计算机视觉、电力计算机视觉和电力人工智能。E-mail: zhangkeit@ncepu.edu.cn

郭玉荣,通信作者,女,讲师,主要研究方向为计算机视觉、电力计算机视觉和电力人工智能。

E-mail: guoyurong@ncepu.edu.cn

梁龙萍,女,硕士研究生,主要研究方向为计算机视觉。

E-mail: lianglongping20@163.com

王子念,男,硕士研究生,主要研究方向为计算机视觉。

E-mail: wznbnxxx@163.com