

文章编号:1673-9469(2025)05-0086-09

DOI:10.3969/j.issn.1673-9469.2025.05.011

基于双分支残差网络的人脸年龄估计

池静¹, 何江¹, 赵伟康¹, 池佳稷², 高松³

(1. 河北工程大学 信息与电气工程学院, 河北 邯郸 056038; 2. 拉彭兰塔理工大学 电气工程学院, 南卡累利亚 拉彭兰塔 53850; 3. 邯郸市邯三建筑工程有限公司, 河北 邯郸 056001)

摘要: 针对现有卷积神经网络难以捕捉人脸长距离依赖关系的问题, 提出一种基于大卷积核的双分支残差网络, 并用于人脸年龄估计。为突破传统小卷积核感受野较小的局限, 首先, 采用大卷积核对深度学习模型(ResNet)的残差模块进行改进, 以此扩大网络的有效感受野, 从而更高效地捕捉人脸图像中的全局信息与长距离依赖关系。其次, 考虑到人脸细微特征对年龄估计的关键作用, 引入细节下采样模块, 该模块能在网络初始阶段最大程度减少细节信息的丢失。在网络结构设计上, 创新性地将 ResNet 的原始残差模块与改进后的大卷积残差模块进行并联, 构建形成双分支残差网络, 并利用注意力模块实现两个分支之间的特征融合。为深度挖掘人脸年龄特征, 在双分支残差网络之后串联两个大卷积残差模块, 通过递进式特征抽象强化模型对复杂年龄模式的建模能力。最后, 针对年龄估计中标签序数特性带来的挑战, 将所构建的双分支残差网络与序数回归方法结合, 通过将年龄值转化为有序标签序列进行建模, 有效提升了模型对年龄变化的分辨能力。实验结果表明, 所提出的方法在 UTK-FACE 数据集上的平均绝对误差最多降低了 0.46, 在 FG-NET 数据集上的平均绝对误差最多降低了 0.09。

关键词: 年龄估计; 序数回归; 卷积神经网络; 有效感受野; 注意力机制

中图分类号: TP391

文献标识码: A

Face Age Estimation Based on Two-branch Residual Network

CHI Jing¹, HE Jiang¹, ZHAO Weikang¹, CHI Jiaji², GAO Song³

(1. School of Information and Electrical Engineering, Hebei University of Engineering, Handan, Hebei 056038, China;
2. School of Electrical Engineering, Lappeenranta University of Technology, Lappeenranta, South Karelia Region
53850, Finland; 3. Handan Hansan Construction Engineering Co., Ltd., Handan, Hebei 056001, China)

Abstract: To address the problem that existing convolutional neural networks have difficulty in capturing long-range dependencies in human faces, a dual-branch residual network based on large convolution kernels is proposed and applied to face age estimation. Firstly, to overcome the limitation of small receptive fields of traditional small convolution kernels, large convolution kernels are adopted to improve the residual module of the deep learning model (ResNet), thereby expanding the effective receptive field of the network and more efficiently capturing the global information and long-range dependencies in face images. Secondly, considering the crucial role of facial fine features in age estimation, a detail downsampling module is introduced, which can minimize the loss of detailed information at the initial stage of the network. In the design of the network structure, the original residual module of ResNet and the improved large convolution residual module are innovatively connected in parallel to form a dual-branch residual network, and an attention module is utilized to achieve feature fusion between the two branches. To further enhance the deep mining of face age features, two large convolution residual modules are concatenated after the dual-branch residual network, and through progressive feature abstrac-

收稿日期: 2024-05-24

基金项目: 河北省自然科学基金资助项目(F2023402011)

第一作者: 池静(1973—), 女, 河北邯郸人, 硕士, 副教授, 主要研究方向为计算机视觉。

tion, the model's ability to model complex age patterns is strengthened. Finally, in response to the challenges brought by the label ordinality characteristics in age estimation, the constructed dual-branch residual network is combined with ordinal regression methods. By converting age values into ordered label sequences for modeling, the model's ability to distinguish age changes is effectively improved. Experimental results show that the proposed method reduces the mean absolute error (MAE) by up to 0.46 on the UTK-FACE dataset and by up to 0.09 on the FG-NET dataset.

Key words: age estimation; ordinal regression; convolutional neural network; effective receptive field; attention mechanism

人脸年龄估计是计算机视觉领域的重要任务之一,被广泛应用于人脸识别、社交媒体分析和年龄相关服务等领域。年龄估计的目标在于找出年龄模式和面部外观的变化,以得到更准确的年龄预测值。Cao 等^[1]将年龄估计问题转化为一系列有序的多个子分类问题,每个子问题仅判断是否比某个年龄大,这推动了与序数回归相关的年龄估计研究工作的开展^[2-3]。He 等^[4]通过融合表观双峰分布和时间顺序单分布,获得了更稳健的年龄标签分布,提高了年龄估计的性能。Kancharlapalli 等^[5]通过使用深度卷积神经网络(Convolutional Neural Network, CNN),改善了年龄和性别估计的性能。Foggia 等^[6]提出了一种多任务 CNN,通过人脸图像识别性别、年龄、种族和情感,与单任务 CNN 相比,在准确率、处理时间和存储空间等方面,都具有明显的优势。杜希婷等^[7]将注意力机制融入残差网络进行年龄估计,提高了表观年龄估计的性能。朱昱等^[8]提出基于对抗正则化的人脸年龄估计模型,将年龄段进行分类,提高了年龄段分类的精度。

综上所述,多数学者基于 CNN 开展人脸年龄估计研究,有效感受野作为 CNN 中的一个重要概念,其特性对模型性能具有关键影响。根据有效感受野理论^[9],感受野中的像素对输出的贡献并不相同,中心像素对输出影响更显著,称为有效

感受野,即感受野的有效区域面积。年龄估计需考虑全局的面部特征和结构,大的有效感受野可更好捕获全局信息和长距离依赖。目前, CNN 主要采用小卷积核^[10-13],其有效感受野较小,无法充分捕捉人脸的全局特征和长距离依赖关系。因此,本文提出利用大卷积核对深度学习模型(ResNet)^[11]的原始残差模块进行改进,以扩展其有效感受野。然而,大卷积残差模块会忽略一些细节信息。为了解决这个问题,本文构建了双分支残差网络,该网络由大卷积残差模块分支和原始残差模块分支组成,分支之间通过注意力模块实现特征融合,最后与序数回归方法结合进行人脸年龄估计。

1 基于双分支残差网络的人脸年龄估计

1.1 基于大卷积核的双分支残差网络

基于大卷积核的双分支残差网络模型的总体架构如图 1 所示,该网络主要由细节下采样模块(Detail Downsampling Module, DDM)、大卷积残差模块、原始残差模块、注意力模块以及序数回归等组件组成。原始残差模块由小卷积核组成,可以更好地提取人脸的细节特征,将原始残差模块作为大卷积残差模块的补充,可以综合利用全局和局部特征,提高人脸年龄预测的性能。大卷积残差模块与原始残差模块之间通过注意力模块进行特征交互。

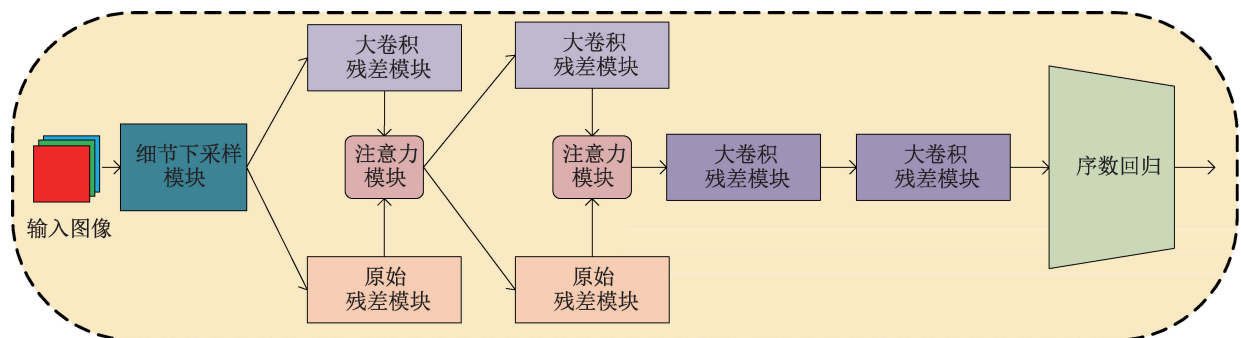


图 1 模型总体架构

Fig. 1 Overall architecture of model

细节下采样模块由 3×3 卷积核、批量归一化层 (Batch Normalization, BN)、线性整流函数 (Rectified Linear Unit, ReLU)、 3×3 逐通道 (Depthwise, DW) 卷积核以及逐点 (Pointwise, PW) 卷积组成, 如图 2 所示。选择较小的卷积核而非原始的 7×7 大卷积核, 是为了在初始阶段保留更多的图像信息和细节特征, 而较为全局的特征则由大卷积残差模块提取。这样的设计旨在提高模型的细节感知能力和特征提取效果。

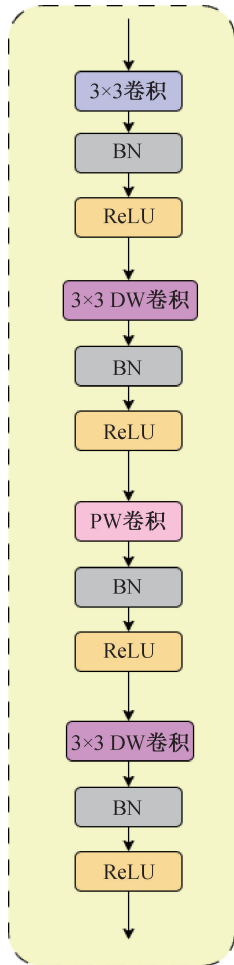


图 2 细节下采样模块

Fig. 2 Detail downsampling module

改进后的大卷积残差模块由 1×1 卷积核、BN 层、ReLU、DW 大卷积核分支、DW 小卷积核分支以及 PW 卷积组成, 如图 3 所示。BN 层和 ReLU 激活函数可确保特征图有相同的分布范围并去除负值, 从而提高模型的稳定性和泛化性能。该模块利用大尺寸的卷积核扩展有效感受野, 捕捉人脸图像的全局信息和长距离依赖关系, 不同的大卷积残差模块使用不同尺寸的卷积核, 以提取不同层次和尺度的年龄特征。大卷积核可以捕捉人脸年龄的全局特征, 但容易忽略局部细节特征, 因

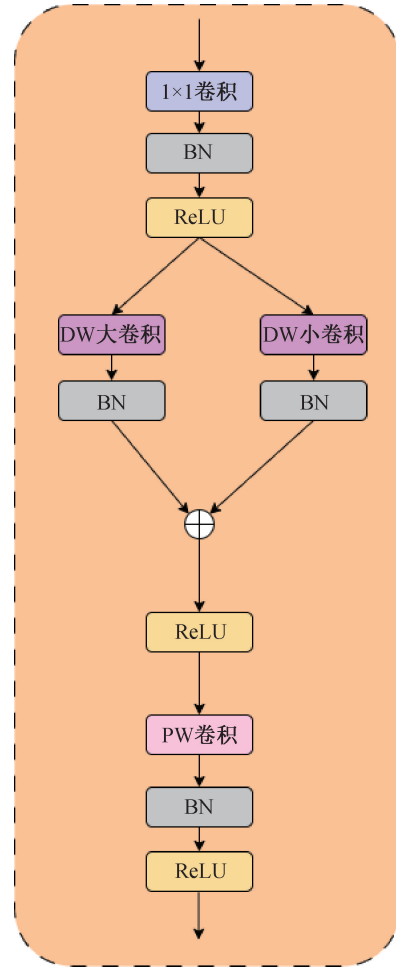


图 3 大卷积残差模块

Fig. 3 Large convolution residual module

此, 在大卷积残差模块内部增加小卷积核分支, 进一步辅助提取细节特征。最后通过残差连接融合两个分支的特征, 实现跨层连接, 解决深度神经网络退化问题。

为了减少参数量和运算成本, 在改进的细节下采样模块和大卷积残差模块中均采用深度可分离卷积^[14]技术, 该技术将卷积操作分解为 DW 卷积和 PW 卷积。为提高模型推理效率, 本文采用结构重参数化方法^[9,15], 将大卷积和小卷积的平行结构转换为另外一个等价的结构。结构重参数化需要使用 BN 层的参数, 其计算公式见式(1)。

$$f(x) = \gamma \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (1)$$

式中: x 为输入的图像信息; μ 为输入数据 x 的平均值; σ^2 为输入数据 x 的方差; γ 为放缩系数; β 为平移系数; ε 为非常小的正数 (通常取 $1e-5$), 避免计算中因 $\sigma^2 = 0$, 出现除以 0 错误。

令 $t = \gamma/\text{std}$, 则重参数化的计算公式为

$$W_{out} = W_{in} \times t \quad (2)$$

$$B_{out} = \beta - \mu \times t \quad (3)$$

式中: std 为 x 的标准差, 即 $\text{std} = \sqrt{\sigma^2 + \varepsilon}$; W_{in} 为输入卷积的权重; W_{out} 、 B_{out} 分别为重参数化后的新卷积、新偏置项权重。

结构重参数化的示例如图 4 所示, 首先, 将 7×7 大卷积核的参数和 3×3 小卷积核的参数分别输入到式(1)中, 得到归一化后的特征图参数(BN 层参数); 然后, 将两个卷积核参数及其 BN 层的参数分别输入到式(2)(3)中, 得到新的大卷积核、小卷积核参数; 最后, 将新的小卷积核参数与大卷积核参数融合, 形成一个等价于原始平行结构的新结构。

1.2 基于序数回归的人脸年龄估计

基于序数回归方法将年龄估计转换为排序问题, 使用年龄轴方法进行人脸年龄估计。具体做法是将年龄估计转换为一系列二元分类问题, 最终得到的聚合二元分类结果可用于年龄分类^[1]。

设 $r_k \in \{r_1, r_2, \dots, r_K\}$ ($1 \leq k \leq K$) 为升序排列的整数年龄序列, y_i ($1 \leq i \leq N$) 为第 i 个样本的真实年龄标签, 其中 K 为年龄类别数量, N 为样本数量。序数回归的本质是将 y_i 转换为 $K-1$ 个二元分类标签 $g_i^k \in \{0, 1\}$, g_i^k 表示年龄 y_i 是否大于第 k 个年龄 r_k , 例如 $g_i^k = 1$ 表示 $y_i > r_k$ 。

通过将真实年龄标签 y_i 转换成 $K-1$ 个二元分类标签 g_i^k , 即将直接预测年龄 \hat{y}_i 的问题转换成了预测 $K-1$ 个二元分类标签 g_i^k 的二元分类问题。最后, 基于模型预测得出的 $K-1$ 个二元分类标签 \hat{g}_i^k , 通过式(4)聚合 $K-1$ 个二元分类结果, 从而计算出最终的年龄预测值 \hat{y}_i 。序数回归模型通过式(5)进行优化。

$$\hat{y}_i = \sum_{k=1}^{K-1} \hat{g}_i^k \quad (4)$$

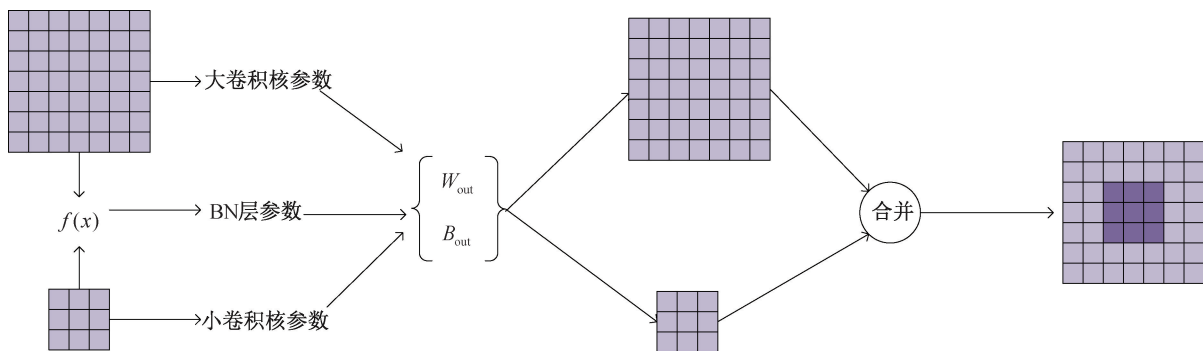


图 4 结构重参数化示例

Fig. 4 Example of structure reparameterization

$$L_{OR} = - \sum_{i=1}^N [\log(S(o_i))g_i + \log(1 - S(o_i))(1 - g_i)] \quad (5)$$

式中: L_{OR} 为序数回归的损失函数, 用来衡量模型预测值与真实值之间的误差; S 代表 Softmax 函数, 其公式为

$$S(o_i) = \frac{e^{o_{i,k}}}{\sum_{j=1}^{K-1} e^{o_{i,j}}} \quad (k = 1, 2, \dots, K-1) \quad (6)$$

式中: o_i 表示第 i 个样本所有二分类器的输出; e 为自然常数。

2 实验

2.1 数据集

UTK-FACE 数据集^[16] 包含超过 20 000 张人脸图像, 标注了年龄、性别和种族信息。该数据集的年龄范围为 0~116 岁, 具有较长的年龄跨度, 其图像多样性高, 涵盖了姿势、表情、光照、遮挡、分辨率等方面的变化。

FG-NET 数据集^[17] 包含 82 个个体的 1 002 张彩色或灰度人脸图像, 年龄范围为 0~69 岁。

2.2 评价指标

人脸年龄估计是一项具有极大挑战性的视觉任务, 为了客观评估年龄估计的效果, 研究者们提出了两个重要的评价指标, 分别是平均绝对误差 (Mean Absolute Error, MAE) 和累积分数 (Cumulative Score, CS)。这两个评价指标已成为人脸年龄估计领域的主流评估标准, 其中, MAE 是使用最广泛的评估指标。

MAE 是指预测年龄和真实年龄之间的绝对误

差的平均值,MAE 越小表示预测结果越准确,性能越好。MAE 的计算公式见式(7)。

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (7)$$

式中: y_i 为第 i 张人脸图像的真实年龄,岁; \hat{y}_i 为第 i 张人脸图像的估计年龄,岁; N 为样本数量,个。

CS 反映了模型在给定误差范围内预测正确的样本数量和总样本数量之间的比值,对于同一个阈值,CS 的值越大,表明模型的性能越好。CS 的计算公式见式(8)。

$$CS = \frac{N_\theta}{N} \quad (8)$$

式中: N_θ 为测试集中预测年龄和真实年龄的绝对误差不超过 θ 年的样本数量,个。通常, θ 取 10 以内的整数,本文将阈值 θ 设置为 5^[2]。

2.3 实验环境及参数设置

实验环境:操作系统使用 Ubuntu 20.04,深度学习框架为 PyTorch 1.11.0,配备 NVIDIA GeForce RTX 3090(24 GB) GPU,CUDA 版本为 11.3。本文实验所有数据集均基于 Retinaface^[18] 的人脸对齐和裁剪等预处理操作,并按照 8:2 的比例划分为训练集和验证集。在训练阶段,使用水平翻转和随机裁剪等数据增强技术提高模型泛化能力。本文选择 AdamW 优化器作为优化算法,设置批处理大小为 64,动量和权重衰减分别设置为 0.9 和 0.0001,初始学习率设置为 0.01。

2.4 实验结果及分析

2.4.1 大卷积残差模块消融实验

为了验证改进模块的有效性,本节在 UTK-FACE 数据集上进行了大卷积残差模块的消融实验。评估指标使用 MAE 和 CS,实验结果如表 1 所示。其中,ResNet50/101 表示原始的 ResNet 模型,ResNet50/101-K1/K2 表示使用大卷积残差模块改进后的 ResNet 模型,且 K1=[13, 13, 9, 7],K2=[21, 19, 17, 13],K1 和 K2 表示两种不同的大卷积核尺寸设置。ResNet50/101-K1/K2 + DDM 表示使用 DDM 替换 ResNet 前四层(7×7 卷积、BN 层、ReLU 激活函数和最大池化层)。

根据实验结果,当使用 K1 的大卷积核设置时,ResNet50-K1 的 MAE 比 ResNet50 降低了 0.33,CS 提高了 2 个百分点。ResNet101-K1 的 MAE 比 Resnet101 降低了 0.4,CS 提高了 2.68 个

百分点;当使用 K2 的大卷积核设置时,ResNet50-K2 取得了比 ResNet50 更高的 MAE 和更低的 CS,虽然 ResNet101-K2 取得了比 ResNet101 更好的效果,但其性能低于 ResNet101-K1;当使用 DDM 时,各个模型的性能都得到了显著的提升。实验结果表明,增加模型的有效感受野能够提升人脸年龄估计的性能,但是通过增大卷积核来扩展模型的有效感受野存在一定限制。需根据具体的任务确定合适的大卷积核尺寸,否则会导致模型的性能变差。此外,新的 DDM 能够保留更多的细节信息,可进一步提高人脸年龄估计的性能。

表 1 在 UTK-FACE 数据集上的消融实验结果

Tab. 1 Results of ablation experiments on the

UTK-FACE dataset

模型	平均绝对误差 MAE	累计分数 CS/%
ResNet50	5.73	62.84
ResNet101	5.66	63.18
ResNet50-K1	5.40	64.84
ResNet50-K2	5.82	61.70
ResNet101-K1	5.26	65.86
ResNet101-K2	5.34	64.52
ResNet50-K1+DDM	5.15	66.36
ResNet50-K2+DDM	5.25	66.07
ResNet101-K1+DDM	5.08	66.75
ResNet101-K2+DDM	5.19	66.17

2.4.2 注意力模块消融实验

根据表 1 的实验结果,本节选择 ResNet101-K1+DDM 构建双分支残差网络,并依托 UTK-FACE 数据集对注意力模块进行了消融实验。评估指标使用 MAE 和 CS,注意力模块选用了流行的卷积注意力模块(Convolutional Block Attention Module, CBAM)^[19]、挤压-激励注意力机制(Squeeze-and-Excitation Networks, SE)^[20]、高效通道注意力机制(Efficient Channel Attention, ECA)^[21]以及坐标注意力机制(Coordinate Attention, CA)^[22],实验结果如表 2 所示。

表 2 注意力模块消融实验

Tab. 2 Ablation experiment of the attention module

模型	平均绝对误差 MAE	累计分数 CS/%
ResNet101-K1+DDM + CBAM	5.02	66.60
ResNet101-K1+DDM + SE	4.97	66.57
ResNet101-K1+DDM + ECA	5.00	66.32
ResNet101-K1+DDM + CA	4.91	66.96

根据表2实验结果,双分支残差网络使用CBAM、SE、ECA、CA等4种不同的注意力模块进行特征融合,都取得了比ResNet101-K1+DDM更低的MAE。ResNet101-K1+DDM+CA的CS为66.96%,高于ResNet101-K1+DDM的66.75%。综上,与使用CBAM、SE、ECA注意力机制进行特征融合相比,CA注意力机制表现出更显著的效果。因此,本文最终选择CA注意力机制来实现分支之间的特征融合。

2.4.3 不同阈值 θ 取值对CS的影响

为分析不同 θ 取值对CS的影响,本节计算了阈值 θ 取1~10的整数时,人脸年龄估计模型在UTK-FACE和FG-NET数据集上的CS,实验结果如表3所示。随着阈值 θ 的增加,模型在UTK-FACE和FG-NET数据集上的CS呈现明显的上升趋势。这是由于CS反映了模型在给定误差范围内预测正确的样本数量和总样本数量之间的比值,更大的阈值意味着更大的允许误差范围。然而,过大或过小的阈值均不能准确地反映出模型的性能,一般将阈值设置为5是较为合适的选择。因此,后续实验中的CS的阈值均设置为5。

表3 不同 θ 取值对CS的影响

Tab. 3 Influence of different θ values on CS

阈值 θ	UTK-FACE	FG-NET
1	29.43	42.46
2	41.47	54.45
3	50.78	71.29
4	59.33	77.08
5	66.96	83.44
6	73.01	85.51
7	76.94	87.72
8	80.43	90.39
9	84.52	91.45
10	87.12	93.06

2.4.4 与其他方法对比

本文模型是以ResNet为基础进行改进的,为了验证改进方法的有效性,对VGG-16、DenseNet121和ConvNext-base等主流CNN模型进行复现。基于本文提出的人脸年龄估计模型以及复现的主流CNN模型,在UTK-FACE和FG-NET数据集上进行实验,并将MAE和CS作为评估指标。为了与其他主流的年龄估计方法进行比较,本次研究选择Coral^[1]、CloFormer-OR^[3](表4)、ALD-Net^[4]、

DOEL3groups^[23]、DEX^[24]以及MV^[25](表5)等方法,并提取其在已发表文章中的实验结果。由于这些研究工作中没有使用CS评估指标,因此,在对比时,未对CS指标作进一步的对比分析。实验结果如表4和表5所示。

表4 与其他方法在UTK-FACE数据集上的对比

Tab. 4 Comparison with other methods on the

UTK-FACE dataset

模型	平均绝对误差 MAE	累计分数 CS/%
VGG-16 ^[10]	13.69	34.89
DenseNet121 ^[12]	5.70	61.48
ConvNext-base ^[13]	6.11	60.69
Coral ^[1]	5.47	—
CloFormer-OR ^[3]	5.37	—
本文模型	4.91	66.96

表5 与其他方法在FG-NET数据集上的对比

Tab. 5 Comparison with other methods on the

FG-NET dataset

模型	平均绝对误差 MAE	累计分数 CS/%
VGG-16 ^[10]	10.29	36.98
DenseNet121 ^[12]	3.86	77.43
ConvNext-base ^[13]	3.91	80.21
ALD-Net ^[4]	3.25	—
DOEL3groups ^[23]	3.44	—
DEX ^[24]	4.30	—
MV ^[25]	4.10	—
本文模型	3.16	83.44

根据表4可知,本文建立的模型在UTK-FACE数据集上的MAE为4.91,CS为66.96%。与CloFormer-OR相比,MAE降低了0.46。与表中同类CNN模型相比,本文模型均取得了更低的MAE和更高的CS。根据表5可知,本文建立的模型在FG-NET数据集上的MAE为3.16,CS为83.44%,与ALD-Net相比,MAE降低了0.09。与表中同类CNN模型相比,同样取得了更低的MAE和更高的CS。

2.4.5 模型的MAE收敛情况对比

MAE被广泛用于年龄估计,通过观察MAE的收敛情况,可以准确地反映出年龄估计模型的收敛程度。本节对所提出的模型以及复现的CNN模型在UTK-FACE、FG-NET数据集上的MAE收敛情况进行了可视化分析,如图5所示。

由图5可知,随着迭代次数的增加,DenseNet121、ConvNext-base和本文模型的MAE呈现相似的下降趋势。值得注意的是,本文模型的MAE始终

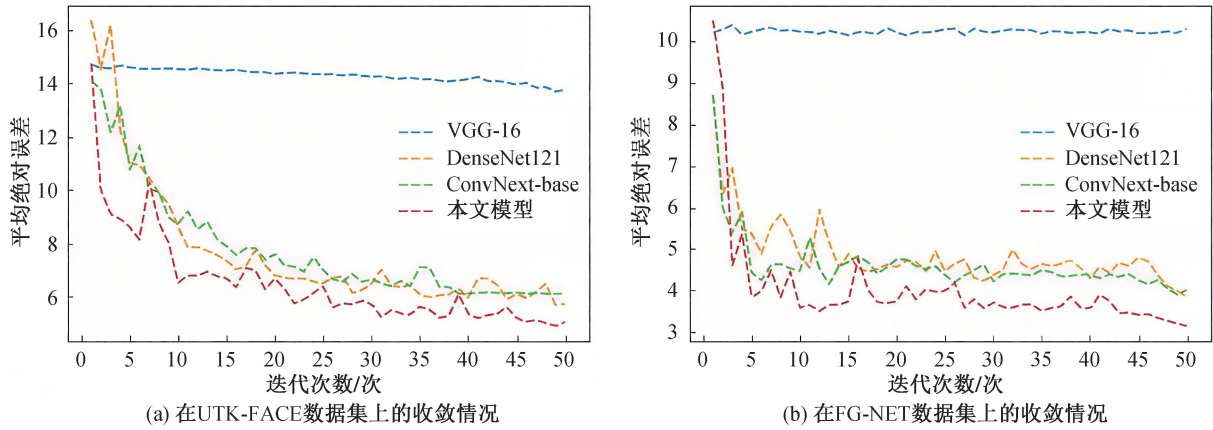


图5 模型 MAE 收敛情况对比

Fig. 5 Comparison of MAE convergence of model

低于其他几个 CNN 模型。与此同时,可以观察到 VGG-16 的 MAE 明显高于其他模型,且随着迭代次数的增加保持稳定,无明显下降趋势。实验结果进一步验证了本文提出的模型在人脸年龄估计任务上具有更出色的性能。

2.4.6 不同年龄段的年龄估计性能评价

本节评估了建立的人脸年龄估计模型在 0~5 岁、6~11 岁、12~17 岁、18~23 岁、24~33 岁、34~45 岁、46~59 岁、60 岁以上等 8 个年龄段内(年龄都为整数)的准确率和 CS,且用于评估的模型未单独针对所划分的 8 个年龄段进行训练。其中,准确率=无误差下预测正确的样本数/总样本数 \times 100%。

表 6 为本文提出的人脸年龄估计模型在 UTK-FACE、FG-NET 等数据集上的实验结果,可以看出,模型在 0~5 岁年龄段的准确率最高;在 0~5 岁和 6~11 岁年龄段,其 CS 均取得了最优结果。值得注意的是,在 18~23 岁年龄段,模型在两个数据集上的准确率仅为 22.25% 和 22.22%,但 CS 却表现较好,分别达到 76.51% 和 74.07%。这是因

为当年龄位于年龄区间边界时,预测年龄和真实年龄可能会被划分到相邻的两个年龄段。例如,当真实年龄为 23 岁,预测年龄为 24 岁时,23 岁属于 18~23 岁年龄段,24 岁则属于 24~33 岁年龄段。与此同时,CS 允许预测年龄有一定范围的误差,当阈值 θ 设置为 5 时,预测年龄在真实年龄上下 5 岁范围内都将被接受。

由表 6 可以观察到,随着年龄段的增加,CS 和准确率的表现并不理想,这主要由以下几个原因造成:首先,数据集样本的分布存在不均衡性,青少年样本数量较多,而其他样本数量随年龄的增加而减少,尤其是老年人样本,这可能导致模型在训练时偏向某些年龄段,而其他年龄段则存在欠拟合问题;其次,在 0~23 岁年龄段内人脸的各项特征具有明显的变化,之后年龄段的变化并不十分显著(在相对较短的年龄范围而言);最后,本文建立的模型并未针对划分的年龄段进行专门训练,缺乏对特定年龄段特征的学习;此外,光照、姿势、种族、图像分辨率和数据集本身的标注错误都会对模型的预测结果产生很大影响。本文提出的

表 6 不同年龄段的年龄估计性能评价

Tab. 6 Evaluation of age estimation performance in different age groups

年龄/岁	UTK-FACE		FG-NET	
	准确率/%	累积分数 CS/%	准确率/%	累积分数 CS/%
0~5	89.74	96.34	86.36	97.73
6~11	50.00	77.42	70.73	100.00
12~17	42.40	58.52	59.46	86.49
18~23	22.25	76.51	22.22	74.07
24~33	77.76	78.92	70.59	70.59
34~45	42.88	44.88	46.15	38.46
46~59	55.67	47.13	20.00	20.00
≥ 60	66.12	47.43	33.33	33.33

人脸年龄估计模型引入了较大的卷积核以增加模型的有效感受野,同时设计了双分支残差网络以补充细节特征,虽然整体上提高了年龄估计的性能,但对于细节特征的捕捉还不够充分,且对于复杂环境下的样本较为敏感。尽管存在以上局限,本文建立的模型仍取得了不错的结果,表明了该模型在年龄估计任务上具有一定的鲁棒性。

3 结论

1)通过引入大卷积核改进 ResNet 残差模块,极大扩展模型的有效感受野,从而有效捕捉人脸全局信息与长距离依赖关系。

2)新的细节下采样模块,可保留更多人脸细节信息,弥补传统方法在细节留存上的不足。

3)构建由改进大卷积残差模块与原始残差模块并联的双分支残差网络,借助注意力模块完成分支间特征融合,为年龄估计提供更丰富的特征支撑。

4)将双分支残差网络与序数回归结合,实现了更准确的年龄估计。与 VGG-16, DenseNet121 等方法相比,本文建立的模型在 UTK-FACE 数据集上的平均绝对误差均有所降低,最多降低了 0.46。与 VGG-16、DenseNet121、ConvNext-base 等的方法相比,本文建立的模型在 FG-NET 数据集上的平均绝对误差同样均有所降低,最多降低了 0.09。

参考文献:

- [1] CAO W, MIRJALILI V, RASCHKA S. Rank consistent ordinal regression for neural networks with application to age estimation [J]. *Pattern Recognition Letters*, 2020, 140: 325-331.
- [2] SHIN N H, LEE S H, KIM C S. Moving window regression: a novel approach to ordinal regression [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Piscataway, NJ: IEEE, 2022: 18760-18769.
- [3] 付帅, 郭小英, 白茹意, 等. 改进的 CloFormer 模型与有序回归相结合的年龄评估方法 [J]. *计算机应用*, 2024, 44(8): 2372-2380.
FU S, GUO X Y, BAI R Y, et al. Age estimation method combining improved CloFormer model and ordinal regression [J]. *Journal of Computer Applications*, 2024, 44(8): 2372-2380.
- [4] HE J, HU C, WANG L. Facial age estimation based on asymmetrical label distribution [J]. *Multimedia Systems*, 2023, 29(2): 753-762.
- [5] KANCHARLAPALLI T P, DWIVEDI P. A Novel Approach for Age and Gender Detection using Deep Convolution Neural Network [C]//*2023 10th International conference on computing for sustainable global development (INDIACom)*. New Delhi, India: IEEE, 2023: 873-878.
- [6] FOGGIA P, GRECO A, SAGGESE A, et al. Multi-task learning on the edge for effective gender, age, ethnicity and emotion recognition [J]. *Engineering Applications of Artificial Intelligence*, 2023, 118: 105651.
- [7] 杜希婷, 张德, 甄庆凯. 基于改进残差网络的人脸表情年龄估计 [J]. *传感器与微系统*, 2023, 42(5): 135-138+142.
DU X T, ZHANG D, ZHEN Q K. Estimation of face apparent age based on improve residual network [J]. *Transducer and Microsystem Technologies*, 2023, 42(5): 135-138+142.
- [8] 朱昱, 樊航, 王鹏, 等. 基于深度对抗丢弃正则化的年龄估计 [J]. *电子测量技术*, 2022, 45(1): 145-152.
ZHU Y, FAN H, WANG P, et al. Age estimation based on deep adversarial dropout regularization [J]. *Electronic Measurement Technology*, 2022, 45(1): 145-152.
- [9] DING X, ZHANG X, HAN J, et al. Scaling up your kernels to 31x31: revisiting large kernel design in cnns [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Piscataway, NJ: IEEE, 2022: 11963-11975.
- [10] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]//*International conference on learning representations*. Ithaca, New York: OpenReview.net, 2015: 1-14.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway, NJ: IEEE, 2016: 770-778.
- [12] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway, NJ: IEEE, 2017: 4700-4708.
- [13] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Piscataway, NJ: IEEE, 2022: 11976-11986.
- [14] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway, NJ: IEEE, 2018: 4510-4520.

- [15] DING X, ZHANG X, MA N, et al. Repvgg: making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 2021: 13733-13742.
- [16] ZHANG Z, SONG Y, QI H. Age progression/regression by conditional adversarial autoencoder[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 2017: 5810-5818.
- [17] LANITIS A, TAYLOR C J, COOTES T F. Toward automatic simulation of aging effects on face images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 442-455.
- [18] DENG J, GUO J, VERVERAS E, et al. Retinaface: single-shot multi-level face localisation in the wild[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 2020: 5203-5212.
- [19] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). Berlin: Springer, 2018: 3-19.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 2018: 7132-7141.
- [21] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 2020: 11534-11542.
- [22] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 2021: 13713-13722.
- [23] XIE J C, PUN C M. Deep and ordinal ensemble learning for human age estimation from facial images[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 2361-2374.
- [24] ROTHE R, TIMOFTE R, VAN GOOL L. Dex: deep expectation of apparent age from a single image[C]//Proceedings of the IEEE international conference on computer vision workshops. Piscataway, NJ, 2015: 10-15.
- [25] PAN H, HAN H, SHAN S, et al. Mean-variance loss for deep age estimation from a face[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 2018: 5285-5294.

(责任编辑 王利君)