

遮挡条件下的人脸年龄估计研究

刘嘉伟, 于双飞

青岛大学计算机科学技术学院, 山东 青岛

收稿日期: 2024年12月24日; 录用日期: 2025年1月23日; 发布日期: 2025年1月29日

摘要

在日常生活中, 口罩遮挡口部和鼻部、眼镜遮挡眼睛等情况对基于人脸的年龄估计模型造成了较大的挑战。为应对这一问题, 本文设计了一种基于知识蒸馏的自监督特征重建模型。由于缺少包含年龄标记的遮挡人脸数据集, 我们选择在没有遮挡的原始人脸图像数据集上合成口罩和眼镜作为遮挡数据集, 并通过自监督学习的方式实现特征重建。在方法中, 我们使用主流的年龄估计模型作为教师模型, 训练一个结构相同但附加了特征重建模块的学生模型。通过知识蒸馏, 将教师模型的知识迁移至学生模型, 并利用特征重建模块对遮挡区域进行特征补全, 以提升模型对遮挡人脸图像的年龄估计效果。实验结果表明, 提出的方法显著提升了模型在遮挡图像上的年龄估计性能。特征重建模块有效缓解了常见遮挡对年龄估计的负面影响, 从而增强了模型的实用性和鲁棒性。

关键词

面部年龄估计, Transformers, 知识蒸馏, 遮挡处理, 特征重建

Research on Facial Age Estimation under Occlusion Conditions

Jiawei Liu, Shuangfei Yu

College of Computer Science and Technology, Qingdao University, Qingdao Shandong

Received: Dec. 24th, 2024; accepted: Jan. 23rd, 2025; published: Jan. 29th, 2025

Abstract

In daily life, situations where masks cover the mouth and nose, and glasses cover the eyes, pose significant challenges to facial based age estimation models. To improve the robustness of the facial age estimation model under occlusion, this paper proposes a self-supervised feature reconstruction model based on knowledge distillation. Due to the lack of age labeled occluded face datasets, we chose to synthesize masks and glasses as occluded datasets on the original unobstructed face image

dataset, and achieved feature reconstruction through self-supervised learning. In the method, we use mainstream age estimation models as teacher models and train a student model with the same structure but additional feature reconstruction modules. By knowledge distillation, the knowledge of the teacher model is transferred to the student model, and the feature reconstruction module is used to complete the features of the occluded area, in order to improve the age estimation effect of the model on occluded face images. The experimental results show that the proposed method significantly improves the age estimation performance of the model on occluded images. The feature reconstruction module effectively alleviates the negative impact of common occlusion on age estimation, thereby enhancing the practicality and robustness of the model.

Keywords

Facial Age Estimation, Transformers, Knowledge Distillation, Occlusion Handling, Feature Reconstruction

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

面部年龄估计(Facial Age Estimation, FAE)是计算机视觉和人工智能领域的重要研究方向,具有广泛的应用前景。通过分析面部图像或视频中提取的特征,FAE技术能够准确预测个体年龄,在人口统计、个性化营销、用户界面优化及生物识别等方面发挥着重要作用。在实际应用中,面部遮挡因素,如口罩、太阳镜和围巾等,显著影响了FAE的准确性。遮挡物不仅掩盖了面部的关键区域,还干扰了模型对面部特征的全面捕捉,导致传统方法难以实现准确的年龄估计。很多现代的面部年龄估计方法依赖于深度神经网络,尤其是卷积神经网络(Convolutional Neural Networks, CNN)和自注意力机制(如Transformer)。这些模型往往需要通过大量的面部细节来进行训练和推理。然而,在有遮挡的情况下,这些模型可能需要额外的学习机制来补偿缺失的特征,或通过其他方式推测遮挡区域的年龄信息。尽管深度学习技术近年来显著推进了面部年龄估计(FAE)的发展,但在面对人脸部分遮挡的情况时,模型的鲁棒性仍然不足,这限制了FAE技术在各种环境下的广泛应用[1]。

当前针对带有遮挡的人脸年龄估计的研究主要集中在提升模型对遮挡区域的鲁棒性,方法包括遮挡检测与定位、基于GAN或Transformer的图像修复,以及利用遮挡特性进行多任务学习或区域特征掩蔽。此外,数据增强和对比学习通过合成遮挡样本或对比无遮挡与遮挡图像来增强模型泛化能力,而知识蒸馏和迁移学习利用无遮挡数据上的教师模型指导学生模型适应遮挡场景。近年来,Transformer架构的引入、自监督学习以及对抗训练等新技术也展现了处理遮挡问题的潜力,为进一步提升模型鲁棒性提供了新方向。

本文提出了一种基于知识蒸馏与Transformer特征重建的自监督学习方法,用于解决遮挡条件下的人脸年龄估计问题。首先,通过无遮挡数据集上生成带有口罩或眼镜的遮挡数据,构建输入数据集。教师模型在正常人脸图像上提取年龄相关的特征,并将这些特征作为学生模型的学习目标。学生模型通过层级知识蒸馏逐层模仿教师模型的特征表达,同时结合基于Transformer的特征重建模块,对遮挡区域的特征进行补全和对齐。该模块采用Encoder-Decoder架构,通过多头自注意力机制建模全局特征关系,并通过重建损失优化遮挡区域的特征恢复。整体方法有效结合了知识蒸馏和Transformer的优点,显著提升

了模型在遮挡条件下的特征表达能力和年龄估计准确性。

研究的主要贡献包括:

1. 提出了一种全新的思路, 通过特征重建的方式处理遮挡的面部图像。这一方法在面部年龄估计领域尚未被广泛采用, 突破了传统方法的局限, 提高了在遮挡条件下的年龄估计准确性。

2. 引入 Transformer 作为特征重建模块, 利用其自注意力机制恢复遮挡区域的面部特征。Transformer 模块能够精确捕捉遮挡区域与未遮挡区域之间的关系, 增强了模型对部分遮挡面部图像的鲁棒性。

3. 通过对比实验, 展示了所提出的特征重建方法在处理遮挡面部图像时, 相较于现有技术具有更好的性能。

本文的结构安排如下: 首先, 概述 FAE 的研究现状, 重点介绍最新进展和遮挡对面部识别的具体挑战。其次, 深入研究提出的方法, 详细介绍所提出的深度学习模型、面部数据的预处理和模型训练过程。接着, 介绍实验设计, 包括所使用的数据集描述、应用的评估指标, 以及对模型在非遮挡和遮挡面部图像上的性能进行全面分析。然后, 讨论结果, 深入探讨模型在处理遮挡方面的有效性及其与现有方法的比较优势。最后, 总结 FAE 的研究结果和未来研究的潜在方向。

2. 相关工作

本章概述了与我们的研究相关的关键领域, 特别关注面部年龄估计和 Transformers 在特征重建中的应用。旨在提升模型在面部遮挡条件下的年龄估计准确性。FAE 技术通过分析面部特征预测个人年龄, 在多个领域具有重要应用。然而, 现有方法在面部被遮挡时表现不佳, 影响了模型性能。为此, 本研究采用知识蒸馏技术, 将教师模型的知识迁移到学生模型, 并增加特征重建模块以处理遮挡条件下的面部图像, 从而提高学生模型的鲁棒性。

2.1. 年龄估计

面部年龄估计是计算机视觉中非常重要且富有挑战性的问题。当前的年龄估计方法大致分为四种: 1) 分类方法, 例如 DEX [2]将年龄估计建模为一个包含 101 个类别的分类任务, 而 AL [3]通过引入 LSTM 单元来提取年龄敏感区域的局部特征, 从而提升了预测准确性。2) 回归方法, 比如 BridgeNet [4]借助具有重叠子空间的局部回归器和桥树结构, 通过门控网络有效地挖掘年龄标签之间的连续关系。3) 排序方法, OR-CNN [5]和 Ranking-CNN [6]将年龄处理为排序问题, 在预测年龄时, 通过逐一判断目标是否大于某个年龄值来获得最终年龄估计。4) 分布学习方法, 为应对年龄标签的不确定性, MV [7]和 AVDL [8]使用分布学习来实现稳健的年龄预测。此外, DRF [9]和 DLDLF [10]将年龄估计视为非线性回归任务, 利用 CNN 中的分割节点连接顶层来学习输入相关的数据分割, 在叶节点进行年龄分布学习, 从而更好地处理数据的非均质性。

2.2. 遮挡图像处理

面部遮挡处理的研究主要分为面部修复和面部去遮挡两大方向。在面部修复方面, 基于自编码器和生成对抗网络(GAN)的方法取得了较好效果, 如 Context Encoder [11]成功修复了缺失区域。为了增强模型对遮挡部分的关注, 提出了全局和局部判别器[12][13], 但这些方法主要适用于规则的矩形遮挡。部分卷积层[14]和门控卷积[15]被提出以处理不规则遮挡, 然而, 由于 CNN 在建模远程纹理关联上的局限, 修复结果常常存在伪影。Contextual Attention 和 Self Attention [16]等模块通过增强长程依赖建模, 改善了修复效果, 部分方法还引入面部真实值[17]来确保面部拓扑结构的准确性。尽管如此, 大多数方法仍依赖手动掩膜, 无法自动检测遮挡区域。

面部去遮挡方面, LSTM 自编码器和半监督方法通过无需地面真实掩膜来进行去遮挡。两阶段 GAN 方法[18]则结合修复和去遮挡来实现去遮挡效果。然而, 由于训练集中的遮挡类型有限, 这些方法的修复效果往往不理想, 尤其在大角度的人脸图像中。针对特定遮挡类型, 如眼镜或口罩, 有些方法取得了较好效果, 但总体上, 面部遮挡处理技术仍面临遮挡类型单一、修复质量不高等问题, 未来研究需提高自动化和修复效果的质量。

3. 提出的方法

3.1. 总体架构

由于缺乏含有年龄标签的遮挡数据集, 本实验通过在原始无遮挡数据集上添加遮挡图像(例如口罩或眼镜), 生成一个遮挡数据集。这个步骤为自监督学习提供了输入数据。

提出的方法中, 学生模型的任务是从部分遮挡的图像中恢复缺失的特征, 但不同于传统的自监督任务, 本文通过将教师模型提取的正常特征作为监督信号, 引导学生模型的学习。具体来说, 教师模型通过在正常、无遮挡的图像上进行训练, 学习到人脸图像中年龄相关的特征, 并将这些特征作为目标输出。而学生模型则处理带有遮挡的图像, 其目标是重建被遮挡区域的特征, 使得学生模型的输出尽可能接近教师模型在正常图像上提取的特征。

本文设计的自监督任务旨在引导学生模型不仅依赖于遮挡区域外的信息推断被遮挡区域的特征, 还利用教师模型提供的正常特征进行精确指导。通过度量学生模型输出特征与教师模型正常特征之间的差异, 逐步优化学生模型的特征表征能力, 从而在处理遮挡图像时实现更加精确的特征恢复。在此框架下, 教师模型的正常特征本质上作为自监督学习中的标签, 引导学生模型在遮挡条件下高效完成特征重建。

由于教师模型和学生模型的网络结构相似, 最终的分层器也相同。所以学生模型的年龄估计的性能取决于教师模型, 对于不同的教师模型, 学生模型展现出的最终性能也会略有不同。

3.2. 知识蒸馏

为了使学生模型在特征表达能力上更接近教师模型, 从而获得更好的泛化能力和准确性, 本文使用了一种基于阶段性知识蒸馏(Stage-wise Knowledge Distillation)的方法。与传统的逐层对齐特征表示的知识蒸馏方法不同, 阶段性知识蒸馏通过将训练过程划分为多个阶段, 在每个阶段中逐步优化学生模型的特定部分, 并在每个阶段后冻结已训练部分的参数, 以确保训练过程的稳定性和高效性。

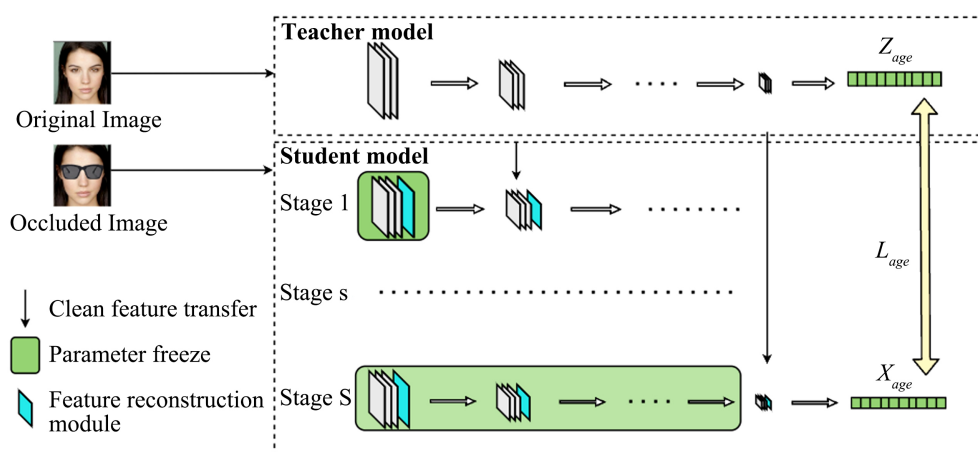


Figure 1. Diagram of stage wise knowledge distillation

图 1. 阶段性知识蒸馏图示

如图 1 所示, 阶段性知识蒸馏的关键在于在每个训练阶段对齐教师模型和学生模型的特定部分的特征表示。在实际操作中, 学生模型和教师模型的网络结构可能并不完全一致, 因此通常选择对学生模型和教师模型的关键层进行对齐。在本文的实验中, 由于教师模型和学生模型的结构相似度较高, 我们选择了对应的卷积层进行特征对齐。这一设计的一个优势在于, 特征重建模块的加入并没有改变特征的维度, 从而保证了蒸馏过程中模型特征的一致性。

由于引入了特征重建模块, 每个阶段训练的损失包括学生模型的任务损失(如交叉熵损失或回归损失)以及特征蒸馏损失。通过最小化蒸馏损失, 学生模型逐步学习教师模型的特征表达能力, 最终实现较好的泛化性能。总蒸馏损失可以表示为:

$$L_{\text{distill}}^{(s)} = \frac{1}{N} \sum_{i=1}^N \left(F_t^{(s)}[i] - F_s^{(s)}[i] \right)^2 \tag{1}$$

其中, $F_t^{(s)}$ 和 $F_s^{(s)}$ 分别是教师模型和学生模型在第 s 阶段对应部分的特征映射, N 是该层特征图的元素数量。

对于每个阶段的蒸馏损失, 将每一阶段的损失加权求和, 得到总蒸馏损失:

$$L_{\text{stage-wise distill}} = \sum_{s=1}^S \lambda_s \cdot L_{\text{distill}}^{(s)} \tag{2}$$

其中, S 是阶段的总数, λ_s 是第 s 阶段的权重。

最终的训练损失由学生模型与真实标签的任务损失和阶段性蒸馏损失组成:

$$L = (1 - \alpha) L_{\text{CE}}(y, z_s) + \alpha L_{\text{stage-wise distill}} \tag{3}$$

其中, $L_{\text{CE}}(y, z_s)$ 是学生模型的任务损失(例如与真实年龄标签 y 的交叉熵损失或回归损失), α 是任务损失和蒸馏损失之间的平衡系数。

通过这种阶段性知识蒸馏方法, 学生模型能够在每个阶段逐步学习教师模型的特征表示, 使得学生模型在遮挡图像的处理能力与计算效率之间达到良好的平衡。阶段性蒸馏不仅提高了学生模型的泛化能力, 还有效地避免了计算开销过大的问题。

3.3. 特征重建模块

为应对遮挡条件对人脸特征完整性的影响, 本研究提出了一种基于 Transformer 的特征重建模块, 如图 2 所示, 通过 Encoder-Decoder 架构实现遮挡区域特征的补全与对齐。

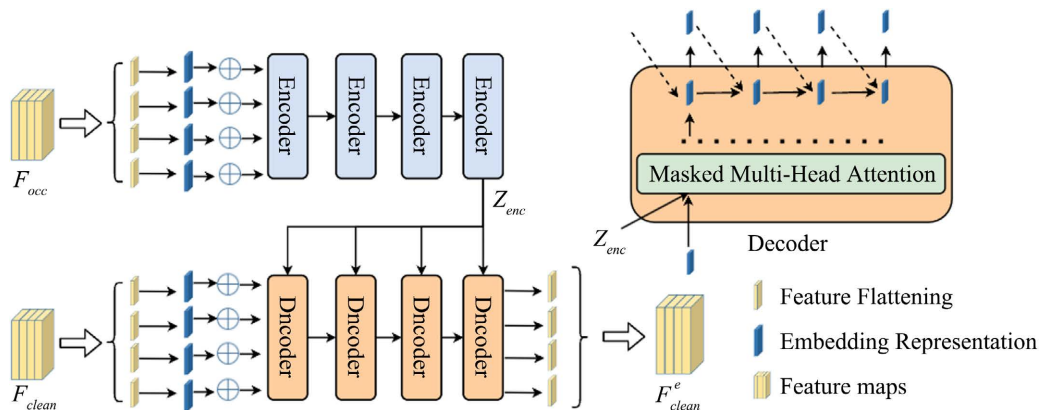


Figure 2. Feature reconstruction module
图 2. 特征重建模块

输入的遮挡图像特征和干净图像特征分别表示为 $F_{occ} \in \mathbb{R}^{H \times W \times C}$ 和 $F_{clean} \in \mathbb{R}^{H \times W \times C}$ 。首先, 这些特征通过展平操作转化为序列化表示:

$$F_{occ}^f = \text{Flatten}(F_{occ}), F_{clean}^f = \text{Flatten}(F_{clean})$$

其中 $F_{occ}^f, F_{clean}^f \in \mathbb{R}^{N \times C}$, $N = H \times W$ 表示序列长度, C 为通道数。随后, 通过线性映射和位置编码, 得到嵌入表示:

$$F_{occ}^e = \text{Linear}(F_{occ}^f) + PE, F_{clean}^e = \text{Linear}(F_{clean}^f) + PE,$$

其中 PE 为位置编码, 用于保留位置信息, 嵌入维度为 d 。

在编码器部分, 遮挡特征 F_{occ}^e 输入多层 Transformer 编码器, 每层通过多头自注意力机制和前馈网络建模全局上下文关系, 其计算过程为:

$$Z_l = \text{MHSA}(F_{occ}^e) + F_{occ}^e, \\ F_{occ}^e = \text{FFN}(Z_l) + Z_l, l = 1, \dots, L,$$

其中 L 为编码器的层数, 最终输出全局特征 $Z_{enc} \in \mathbb{R}^{N \times d}$ 。

解码器结合干净特征 F_{clean}^f 和编码器输出 Z_{enc} , 通过屏蔽注意力(Masked Attention)和编码器-解码器注意力(Encoder-Decoder Attention)逐步恢复遮挡区域的特征。逐步恢复遮挡区域的特征。为了更有效地恢复遮挡区域特征, 我们采用了 mask 机制和循环预测策略。

在每一层解码器的计算过程中, 首先使用 mask 机制对干净特征 F_{clean}^e 中的遮挡区域进行屏蔽。遮挡区域的特征通过逐步推理得到补全, 避免遮挡区域直接影响模型的计算。具体地, 解码器的每一层计算如下:

$$Y_l = \text{Masked Attention}(F_{clean}^e \odot (1 - M_t)) + F_{clean}^e \quad (4)$$

其中 \odot 表示逐元素相乘, $(1 - M_t)$ 表示遮挡区域被 mask 掉, 确保模型只关注非遮挡区域进行计算。接下来, 通过编码器-解码器注意力机制, 解码器将通过 Encoder 获得的上下文特征结合干净图像特征逐步修复遮挡区域的特征:

$$Q_l = \text{Encoder Decoder Attention}(Y_l, Z_{enc}) + Y_l \quad (5)$$

最后, 通过前馈网络(FFN)对解码器输出进行进一步处理, 得到重建特征:

$$F_{recon}^e = \text{FFN}(Q_l) + Q_l, l = 1, \dots, L$$

最终, 重建特征 $F_{recon}^e \in \mathbb{R}^{N \times d}$ 被对齐到干净特征 F_{clean}^e , 完成遮挡区域的特征补全。

此外, 为了进一步优化特征重建效果, 本模块引入了循环预测机制, 即每个时间步的预测不仅依赖于输入的干净特征和遮挡特征, 还依赖于上一时间步的重建结果。这种自回归的方式帮助模型在每一时间步逐渐修正遮挡区域的特征, 最终实现更加精准的补全。具体来说, 解码器的输入在每个时间步更新, 通过上一步生成的重建特征逐步修复遮挡区域:

$$Y_l^t = \text{Masked Attention}(F_{clean}^e \odot (1 - M_t)) + F_{clean}^e \quad (6)$$

$$Q_l^t = \text{Encoder Decoder Attention}(Y_l^{t-1}, Z_{enc}) + Y_l^{t-1} \quad (7)$$

其中 Y_l^{t-1} 是上一时间步的输出。通过这种方式, 解码器在每一时间步迭代修正遮挡区域的特征, 直到达到最终的重建结果。

为了优化模块性能, 模型引入了重建损失, 用于约束重建特征和干净特征之间的差异。重建损失采

用 L_2 损失进行优化:

$$L_{recon} = \frac{1}{N} \sum_{i=1}^N \|F_{recon,i}^e - F_{clean,i}^e\|_2^2$$

通过最小化重建损失, 模型能够有效地学习到遮挡区域和干净区域之间的互补关系。

特征重建模块依托 Transformer 的全局建模能力, 结合 mask 机制和循环预测机制, 有效挖掘遮挡特征与干净特征之间的互补关系, 逐步补齐遮挡区域特征。实验表明, 该模块显著提升了模型在遮挡条件下的鲁棒性, 为年龄估计任务提供了更为全面的特征表达。

4. 实验

4.1. 遮挡数据集

模型的训练主要包括两个阶段, 第一阶段是训练教师模型, 在这阶段我们需要正常人脸数据集。第二阶段是训练学生模型, 这一部分需要受遮挡的数据集。为此, 我们选择了两个数据集: MORPH-2 和 CACD。同时, 我们对这两个数据集进行了数据增强, 通过对图像添加遮挡(墨镜或口罩等)来模拟生活中存在的受遮挡或不完整的人脸图像, 训练学生模型在遮挡条件下的人脸年龄估计的性能。图 3 显示了添加的遮挡的样例, 有两种类型, 一种(第一行和第三行)是通过在面部图像的眼睛或嘴巴上添加太阳镜或口罩来模拟现实生活中的情况。另一种(第二行和第四行)是直接黑色像素块挡住眼睛或嘴巴, 模拟图像不完整的情况。

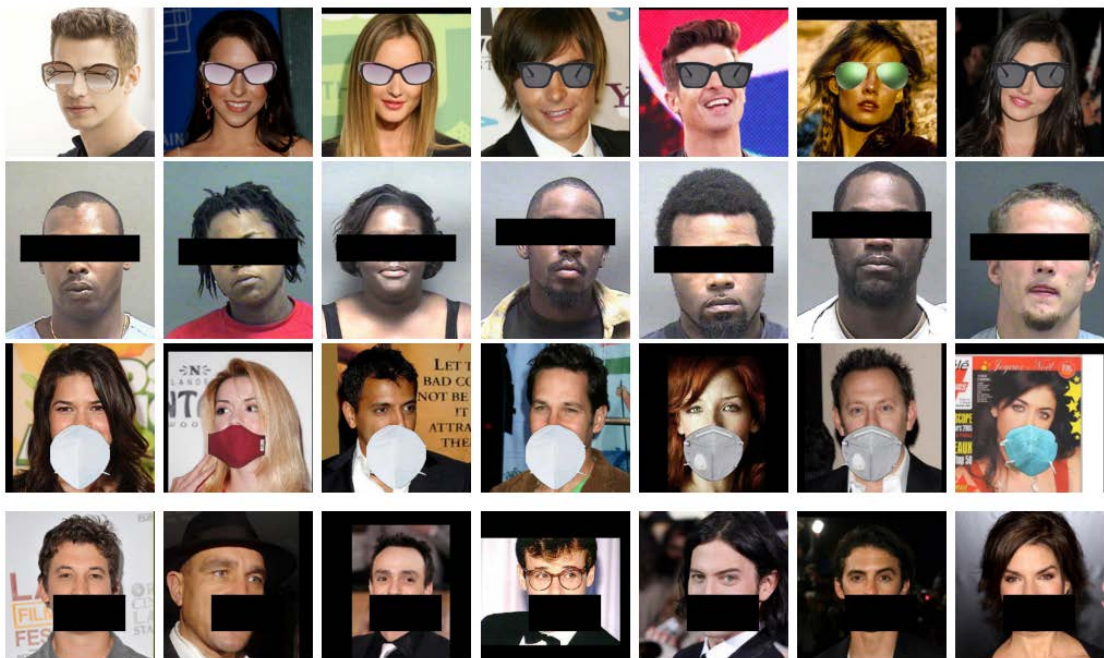


Figure 3. Diagram of occluded dataset

图 3. 遮挡数据集图示

在实验中, 我们选择分别遮挡眼睛和嘴巴, 而不进行同时遮挡。这一设计是考虑到同时遮挡眼睛和嘴巴会导致面部年龄相关信息的大量丢失, 从而使得年龄估计任务变得更加困难, 并可能引发结果的不稳定性。通过至少保留一个区域, 模型能够从剩余的面部特征中提取有效的年龄信息。

4.2. 评估指标

为了对模型进行评估和比较, 我们计算了在最后一轮训练后的测试集上的平均绝对误差(Mean Absolute Error, MAE)和累积分数(CS):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (8)$$

其中 \hat{y}_i 和 y_i 分别是第 i 张测试图像的估计年龄和真实年龄。CS 定义为正确估计的准确率:

$$\text{CS} = \frac{C}{N} \times 100\% \quad (9)$$

其中 C 表示预测正确的图片数量, 即 $|\hat{y}_i - y_i| \leq g$, 在我们的实验设置中 $g \in \{1, 2, \dots, 30\}$, N 代表预测的图片的总数。

MAE (平均绝对误差) 衡量预测年龄与真实年龄之间的平均差异, 对异常值不敏感, 适用于数据中存在异常点的情况。MAE 值越小, 表明模型的准确性越高。CS (准确率) 则衡量模型在特定误差范围内的预测正确率, CS 值越高, 表示模型在预测时越精确。MAE 和 CS 结合使用, 可以从准确性和误差容忍度两个维度全面评估年龄估计模型的性能。

4.3. 对比实验结果

为评估所提出方法在遮挡条件下的人脸年龄估计性能, 我们分别在正常数据集与遮挡数据集上进行了实验。实验中, 我们采用了 MAE (Mean Absolute Error) 作为性能指标, 重点研究了教师模型与学生模型在不同条件下的表现, 并通过消融实验揭示了特征重建模块对性能的提升作用。

4.3.1. 教师模型性能评估

首先, 我们在正常的 MORPH-2 和 CACD 数据集上评估了主流年龄估计模型(教师模型)的性能。表 1 展示了不同模型在正常数据集上的表现, 结果显示各个模型在正常数据集上有较好的准确性, CORAL-CNN 和 DLDL 的 MAE 值较低, 表明其在年龄估计任务中的较好表现。

Table 1. Performance of teacher model on normal dataset

表 1. 教师模型在正常数据集的性能

Methods	MAE		CS	
	MORPH-2	CACD	MORPH-2	CACD
CORAL-CNN	2.31	3.35	91.62	86.81
MWR	2.26	3.38	94.61	91.23
OR-CNN	3.11	4.22	87.57	81.07
DLDL	2.78	3.49	90.36	87.52

为了进一步评估遮挡对教师模型性能的影响, 我们在遮挡数据集上进行了基准测试。表 2 展示了教师模型在两种遮挡设置下(设置 A: 添加墨镜或口罩, 设置 B: 随机添加黑色像素块)在遮挡数据集上的表现。结果表明, 经过微调后的教师模型在遮挡数据集上性能有所提升, 但与正常数据集的表现相比, 仍存在较大差距。

Table 2. Performance benchmark of teacher model on occlusion dataset
表 2. 教师模型在遮挡数据集上的性能基准

Methods	Setting	MAE			
		MORPH-2		CACD	
		Original	Fine tuned	Original	Fine tuned
CORAL-CNN	A	5.13	4.98	6.15	6.02
	B	5.15	4.97	6.18	5.97
MWR	A	5.44	5.19	6.45	5.78
	B	5.42	5.31	6.44	5.95
OR-CNN	A	6.01	5.82	7.78	7.25
	B	5.98	5.83	7.84	7.27
DLDL	A	5.62	5.55	7.52	6.98
	B	5.59	5.38	7.49	6.84

4.3.2. 学生模型性能评估

接下来, 我们评估了学生模型在遮挡数据集上的表现。表 3 展示了学生模型在相同的遮挡设置下(设置 A: 添加遮挡, 设置 B: 添加黑色像素块)的 MAE 和 CS (准确率)值。与表 2 中的教师模型相比, 使用提出的方法改进后的学生模型在遮挡条件下的表现得到了显著提升, 尤其是在 MORPH-2 数据集上, MAE 值较低, 且 CS 值较高, 接近于教师模型在正常数据集上的表现。

Table 3. The performance of student models on occluded datasets
表 3. 学生模型在遮挡数据集上的性能

Methods	Setting	MAE		CS (%)	
		MORPH-2	CACD	MORPH-2	CACD
		CORAL-CNN	A	3.13	4.12
B	3.15		4.35	87.87	86.33
MWR	A	3.44	4.78	86.09	85.82
	B	3.42	4.84	86.21	85.34
OR-CNN	A	4.01	4.88	82.07	82.23
	B	4.58	4.95	82.10	81.39
DLDL	A	3.62	3.98	87.59	85.41
	B	3.59	3.95	87.72	86.15

图 4 展示了学生模型在遮挡数据集上部分正确预测的样本。在这些样本中, 模型能够从未遮挡的人脸部分提取特征, 并对人脸年龄进行正确预测。该结果验证了学生模型在遮挡条件下的鲁棒性和有效性。



Figure 4. Partial prediction of correct samples by student model on occlusion dataset
图 4. 学生模型在遮挡数据集上的部分预测正确的样本

4.3.3. 综合性能评估

最后, 我们在综合数据集上进行性能测试。该数据集包含了正常和遮挡图像, 且遮挡图像按四种不同的遮挡类型(如墨镜、口罩等)平衡分布(每个数据集上都同时用正常数据集和遮挡数据集(测试数据中正常图像和遮挡图像比例为 1 比 1)。其中遮挡图像部分我们同时使用了四种不同的遮挡, 四种遮挡的图像比例为 1:1:1:1。通过 MTCNN 进行人脸检测, 若图像未遮挡, 则由教师模型进行预测; 若图像带有遮挡, 则由学生模型进行处理。表 4 展示了综合数据集上的性能评估结果。

Table 4. Performance on a comprehensive dataset (including normal and occluded datasets)
表 4. 在综合数据集(包含正常数据集和遮挡数据集)上的性能表现

Methods	MAE		CS	
	MORPH-2	CACD	MORPH-2	CACD
CORAL-CNN	2.89	3.44	89.26	86.98
MWR	3.52	3.97	86.87	83.38
OR-CNN	3.85	4.36	84.87	81.16
DLDL	3.78	4.18	84.98	80.95

表 4 的结果表明, 提出的方法在综合数据集上的性能略低于正常数据集的教师模型和遮挡数据集的学生模型的平均值。这可能是由于 MTCNN 未能完全准确地分类所有遮挡数据, 导致教师模型对遮挡数据的预测不准确, 从而影响了综合表现。然而, 整体来看, 所提出的教师-学生模型结构在综合数据集上

的表现仍然优于单独使用教师模型或学生模型, 验证了该方法能够有效提升模型在遮挡条件下的鲁棒性。

4.4. 消融实验

为了评估特征重建模块对学生模型性能的影响, 特别是在遮挡条件下对人脸年龄估计准确性的提升, 本节设计了消融实验, 重点考察特征重建模块的数量对模型性能的影响。

本实验以我们实验中的 CORAL-CNN 为例, CORAL-CNN 网络包含 34 层, 其中有 17 个残差块, 每个残差块包含两层卷积。特征重建模块的插入位置根据网络的残差块进行选择。具体来说, 实验中对 CORAL-CNN 进行了如下设置:

1. 单个特征重建模块: 该模块插入到网络中心位置, 即在第 9 个残差块(第 17 层卷积)后插入。
2. 两个特征重建模块: 第一个模块插入到第 6 个残差块(第 11 层卷积)后, 第二个模块插入到第 12 个残差块(第 23 层卷积)后。
3. 三个特征重建模块: 模块分别插入到第 4、9、14 个残差块(第 7、17、27 层卷积)后。

对于其他两种算法, 我们遵循相同的插入规则, 确保它们的特征重建模块插入位置与 CORAL-CNN 一致, 从而公平地评估不同数量特征重建模块的影响。

表 5 展示了不同设置下, 特征重建模块数量对四种方法在 MORPH-2 和 CACD 数据集上的 MAE (平均绝对误差)的影响。

Table 5. The impact of different quantities of feature reconstruction modules on performance
表 5. 不同数量特征重建模块对性能的影响

Methods	MAE					
	Single reconstruction module		Two reconstruction modules		Three reconstruction modules	
	MORPH-2	CACD	MORPH-2	CACD	MORPH-2	CACD
CORAL-CNN	3.07	4.13	2.75	3.97	2.49	3.74
MWR	3.46	4.77	2.94	4.04	2.82	3.77
OR-CNN	4.05	4.96	3.71	5.27	3.49	4.96
DLDL	3.59	3.88	3.13	4.72	2.96	4.08

实验结果表明, 添加单个特征重建模块后, 学生模型在遮挡条件下的性能显著提升(对比教师模型在遮挡数据集上的性能), 特别是在恢复遮挡区域特征方面。随着两个特征重建模块的加入, 性能进一步改善, 但与单一模块相比, 提升幅度较小。而当加入三个特征重建模块时, 模型的性能接近教师模型在正常数据集上的表现, 尽管性能增幅趋于平缓, 但整体表现有所接近。

综上所述, 在计算资源有限的情况下, 建议使用单个特征重建模块, 其能够在不增加过多计算开销的情况下显著提升模型性能。对于计算资源充足且精度要求较高的场景, 添加多个特征重建模块虽然能够进一步接近教师模型的表现, 但由于训练时长的显著增加, 必须权衡精度提升与计算开销之间的关系。

5. 结论

本研究提出了一种基于教师-学生模型的面部年龄估计方法, 旨在提升在遮挡条件下的年龄估计性能。通过引入特征重建模块, 我们显著增强了学生模型在遮挡图像上的鲁棒性, 使其在遮挡条件下的表

现接近教师模型在正常数据集上的性能。实验结果表明, 改进后的学生模型在遮挡图像上的表现显著提升, 尤其在 MORPH-2 数据集上, 实验验证了方法的有效性。综合测试结果显示, 提出的方法显著提升了模型在复杂条件下的预测能力。消融实验进一步表明, 特征重建模块的加入有效提高了模型的性能, 尤其在多阶段配置下表现更为优越。总体而言, 所提出的方法能够有效应对遮挡带来的挑战, 提升了人脸年龄估计任务在复杂环境中的鲁棒性, 具有较好的应用前景。

基金项目

山东省自然科学基金面上项目 ZR2023MF106。

参考文献

- [1] Zhao, Q., Dong, J., 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**, 3108-3121. <https://doi.org/10.1109/tnnls.2020.3009523>
- [2] Rothe, R., Timofte, R. and Gool, L.V. (2015) DEX: Deep EXpectation of Apparent Age from a Single Image. 2015 *IEEE International Conference on Computer Vision Workshop (ICCVW)*, Santiago, 7-13 December 2015, 252-257. <https://doi.org/10.1109/iccvw.2015.41>
- [3] Zhang, K., Liu, N., Yuan, X., Guo, X., Gao, C., Zhao, Z., et al. (2020) Fine-Grained Age Estimation in the Wild with Attention LSTM Networks. *IEEE Transactions on Circuits and Systems for Video Technology*, **30**, 3140-3152. <https://doi.org/10.1109/tcsvt.2019.2936410>
- [4] Li, W., Lu, J., Feng, J., Xu, C., Zhou, J. and Tian, Q. (2019) Bridgenet: A Continuity-Aware Probabilistic Network for Age Estimation. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 1145-1154. <https://doi.org/10.1109/cvpr.2019.00124>
- [5] Niu, Z., Zhou, M., Wang, L., Gao, X. and Hua, G. (2016) Ordinal Regression with Multiple Output CNN for Age Estimation. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 4920-4928. <https://doi.org/10.1109/cvpr.2016.532>
- [6] Chen, S., Zhang, C., Dong, M., Le, J. and Rao, M. (2017) Using Ranking-CNN for Age Estimation. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 742-751. <https://doi.org/10.1109/cvpr.2017.86>
- [7] Pan, H., Han, H., Shan, S. and Chen, X. (2018) Mean-variance Loss for Deep Age Estimation from a Face. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 5285-5294. <https://doi.org/10.1109/cvpr.2018.00554>
- [8] Wen, X., Li, B., Guo, H., Liu, Z., Hu, G., Tang, M., et al. (2020) Adaptive Variance Based Label Distribution Learning for Facial Age Estimation. *Computer Vision—ECCV 2020*, Glasgow, 23-28 August 2020, 379-395. https://doi.org/10.1007/978-3-030-58592-1_23
- [9] Shen, W., Guo, Y., Wang, Y., Zhao, K., Wang, B. and Yuille, A. (2018) Deep Regression Forests for Age Estimation. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2304-2313. <https://doi.org/10.1109/cvpr.2018.00245>
- [10] Shen, W., Guo, Y., Wang, Y., Zhao, K., Wang, B. and Yuille, A. (2019) Deep Differentiable Random Forests for Age Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 404-419. <https://doi.org/10.1109/TPAMI.2019.2937294>
- [11] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. and Efros, A.A. (2016) Context Encoders: Feature Learning by Inpainting. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 2536-2544. <https://doi.org/10.1109/cvpr.2016.278>
- [12] Li, Y., Liu, S., Yang, J. and Yang, M. (2017) Generative Face Completion. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 5892-5900. <https://doi.org/10.1109/cvpr.2017.624>
- [13] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X. and Huang, T.S. (2018) Generative Image Inpainting with Contextual Attention. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 5505-5514. <https://doi.org/10.1109/cvpr.2018.00577>
- [14] Liu, G., Reda, F.A., Shih, K.J., Wang, T., Tao, A. and Catanzaro, B. (2018) Image Inpainting for Irregular Holes Using Partial Convolutions. *Computer Vision—ECCV 2018*, Munich, 8-14 September 2018, 89-105. https://doi.org/10.1007/978-3-030-01252-6_6

- [15] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X. and Huang, T. (2019) Free-form Image Inpainting with Gated Convolution. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 4470-4479. <https://doi.org/10.1109/iccv.2019.00457>
- [16] Zhang, H., *et al.* (2018) Self-Attention Generative Adversarial Networks. *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, 9-15 June 2019, 7354-7363.
- [17] Yang, Y. and Guo, X. (2020) Generative Landmark Guided Face Inpainting. *Pattern Recognition and Computer Vision*, Nanjing, 16-18 October 2020, 14-26. https://doi.org/10.1007/978-3-030-60633-6_2
- [18] Dong, J., Zhang, L., Zhang, H. and Liu, W. (2020) Occlusion-Aware GAN for Face De-Occlusion in the Wild. 2020 *IEEE International Conference on Multimedia and Expo (ICME)*, London, 6-10 July 2020, 1-6. <https://doi.org/10.1109/icme46284.2020.9102788>