

## 基于双有序性约束的人脸年龄估计研究

王 荀, 黄振生

南京理工大学 数学与统计学院, 南京 210094

**摘要:**目的 人类年龄是人类识别和搜索任务中的重要特征, 现有研究一般将人脸年龄估计视为传统的分类任务, 忽略了年龄之间的有序特征, 导致估计年龄与真实年龄之间的差距较大, 因此, 有必要寻找一种方法以缩小估计年龄与实际年龄的差距。方法 提出一种基于双有序性约束卷积神经网络模型(DO-CNN)的人脸图像年龄估计方法。首先, DO-CNN 使用基于广义 Logistic 分布的有序回归模型作为卷积神经网络的分类器, 并验证比其他有序分类器在人脸估计任务上的优越性; 接着, 进一步提出有序竞争比损失函数, 在传统竞争比损失函数上, 通过引入风险项使损失函数注意到预测年龄与真实年龄的误差, 进而指导模型缩小估计年龄与真实年龄的差距。结果 在开源人脸图像年龄数据集 FGNET 和 AgeDB 上的对比实验显示: 相比现有研究方法, DO-CNN 分别提升约 12% 和 3% 的准确率, 当允许的误差范围扩大后, 该优势依然保持。此外, 基于广义 Logistic 分布的有序回归分类器相比基于其他分布的有序回归分类器具有明显提升。结论 实验结果表明: 基于双有序性约束的卷积神经网络模型可以明显提升人脸年龄估计的准确率, 并减少年龄估计的实际误差。

**关键词:**人脸年龄估计; 有序回归; 卷积神经网络; 竞争比损失函数; 深度学习

**中图分类号:** TP391 **文献标识码:** A **doi:** 10.16055/j.issn.1672-058X.2024.0002.011

### Study on Face Age Estimation Based on Double-ordinality Constraints

WANG Xun, HUANG Zhensheng

School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing 210094, China

**Abstract:** **Objective** Human age is an important feature in human recognition and search tasks. Existing research generally treats age estimation in facial images as a traditional classification task, ignoring the ordered characteristics of age and resulting in a large gap between the estimated age and the actual age. Therefore, it is necessary to find a method to reduce the gap between the estimated age and the actual age. **Methods** This paper proposed a method for age estimation of face images based on a double-ordinality constrained convolutional neural network (DO-CNN) model. Firstly, DO-CNN used an ordered regression model based on the generalized Logistic distribution as a classifier for convolutional neural networks and verified its superiority over other ordered classifiers for face estimation tasks. Then, an ordered competitive ratio loss function was further proposed. By introducing a risk term into the traditional competitive ratio loss function, the loss function considered the error between the predicted age and the actual age, thus guiding the model to reduce the gap between the estimated age and the actual age. **Results** Comparative experiments on the open-source facial image age datasets, FGNET and AgeDB, showed that compared with existing research methods, DO-CNN improved the accuracy by about 12% and 3% respectively, and this advantage remains even when allowing for a larger error range. In addition, the ordered regression classifier based on the generalized Logistic distribution exhibited

**收稿日期:** 2023-01-12 **修回日期:** 2023-03-02 **文章编号:** 1672-058X(2024)02-0086-10

**基金项目:** 全国统计科学研究重大项目(2018LD01)。

**作者简介:** 王荀(1999—), 女, 河南信阳人, 硕士研究生, 从事统计建模与模式识别研究。

**引用格式:** 王荀, 黄振生. 基于双有序性约束的人脸年龄估计研究[J]. 重庆工商大学学报(自然科学版), 2024, 41(2): 86—95.

WANG Xun, HUANG Zhensheng. Study on face age estimation based on double-ordinality constraints[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(2): 86—95.

significant improvements compared with ordered regression classifiers based on other distributions. **Conclusion** The experimental results show that the convolutional neural network model based on double-ordinality constraints can significantly improve the accuracy of age estimation in facial images and reduce the actual errors in age estimation.

**Keywords:** face age estimation; ordered regression; convolutional neural networks; competing ratio loss function; deep learning

## 1 引言

人脸图像中包含着如表情、年龄、性别等丰富的生物学特征,它们可以作为鉴别身份的重要凭据。随着信息技术在社会生活中的广泛应用,人们对从人脸图像中提取生物特征的需求与日俱增。其中,年龄作为面部图像的重要属性之一,在安全控制、人机交互和电子客户关系管理等领域具有十分重要的应用价值<sup>[1]</sup>。例如,基于人脸年龄估计技术,可以鉴别消费者的年龄是否可以参与游乐园的危险项目;通过对不同年龄段的人群制定个性化服务,迎合各年龄层顾客的需要;在刑事侦查中,通过年龄估计技术估计视频中嫌疑人的大致年龄,从而缩小怀疑范围等。人脸年龄估计指利用各种机器学习方法建立一个年龄估计模型,从训练数据中学习出面部特征与年龄之间的数学映射,并将其表达成模型经过训练后的参数。对于未知年龄的人脸图像,可以根据训练好的模型对其进行年龄预测。

目前人脸年龄估计的技术流程主要分为3个阶段:图像预处理、年龄特征提取以及年龄估计。图像预处理即为通过像素亮度变换、几何变换、局部邻域预处理等方式去除图像中的噪声,方便后续年龄特征提取与年龄估计<sup>[2]</sup>。

年龄特征提取的目的是从面部图像中提取出能充分表现人脸年龄变化的特征。1999年,Kwon等<sup>[3]</sup>利用人体测量学模型率先进行人脸图像年龄估计研究,但该方法对图像像素的要求太高,难以在实际中应用。随后,主动外观模型<sup>[4]</sup>、年龄子空间模型<sup>[5]</sup>、年龄流形学习<sup>[6]</sup>以及生物启发特征<sup>[7]</sup>等方法相继被提出,但是这些人工定义的特征提取方法往往存在一定的局限性,无法捕获到更深层的人脸特征信息。不同于传统机器学习中需要人工定义特征,以卷积神经网络为代表的深度学习模型可以通过训练自动从数据中提取深层特征信息。2015年,Wang等<sup>[8]</sup>首次应用卷积神经网络自动提取人脸图像特征,然后采用流形学习方法进行降维,最后分别采用支持向量机和支持向量回归进行年龄段分类以及年龄估计,并获得了不错的效果。

传统人脸年龄估计方法主要将问题建模成回归任务或分类任务并设计相应的算法。针对回归任务,Guo等<sup>[6]</sup>提出一种局部调整的鲁棒回归器,通过对人脸老化趋势的预测,显著提高了年龄估计的准确性;Guo

等<sup>[9]</sup>使用支持向量回归作为模型的回归器,在不同种族的人脸图像数据集上获得了更好的鲁棒性。针对分类任务,Gunay等<sup>[10]</sup>使用 $k$ 最近邻分类算法,在相同年龄的不同图像测试中取得了较高的性能成效;Yi等<sup>[11]</sup>提出基于多尺度的卷积神经网络模型,采用端到端的方式进行人脸图像年龄估计,并取得了显著的效果。然而,回归任务将人脸图像年龄估计约简为线性任务以至忽略了年龄特征通常不具备线性特征,而分类任务完全忽略了人脸年龄标签中的有序信息,因此人脸年龄估计的准确率都不够理想。近几年,研究人员们尝试在卷积神经网络中使用不同的分类器或损失函数,以促使模型在分类任务中尽可能地关注年龄标签中的有序信息。2017年,Chen等<sup>[12]</sup>在卷积神经网络架构的模型中引入排序分类器,证明了年龄标签中的有序信息是提升人脸图像年龄估计任务模型准确率的重要元素;2020年,Vargas等<sup>[13]</sup>采用de la torre等<sup>[14]</sup>在2018年提出的连续二元加权kappa损失函数,并结合卷积神经网络作为特征提取器,对3种不同的累积链接函数应用到有序数据上的效果进行评估(其中包括年龄数据集);2021年,Zhang等<sup>[15]</sup>在负对数似然比损失<sup>[16]</sup>中添加了超参数,并证明了该损失函数在不同深度卷积神经网络结构下,对分类任务有一定的有效性和鲁棒性。然而,单独对分类器或者损失函数的调整,难以在有效利用有序信息的同时,识别模型估计结果与真实年龄标签的实际差异,导致模型的准确率较差,且估计结果与实际年龄存在较大误差。

针对以上问题,本文进一步提出一种基于双有序性约束的卷积神经网络(Double-Ordinality Constrained Convolutional Neural Networks, DO-CNN)人脸图像年龄估计方法,通过有序回归和有序竞争比损失函数将有序性约束融入人脸年龄估计中。首先,使用基于广义Logistic分布的有序回归模型作为卷积神经网络分类器,基于广义Logistic分布的有序回归模型可以使卷积神经网络模型在分类任务上关注到年龄标签中的有序信息,并且通过对偏度参数和尺度参数的调节有效提升模型的泛化能力;此外,提出一种有序竞争比损失函数,在传统竞争比损失函数基础上,通过引入风险项使损失函数注意到预测年龄与真实年龄的误差,进而指导模型缩小估计年龄与真实年龄的差距;最后,在开源

数据集 FGNET 和 AgeDB 上评估 DO-CNN 的有效性和在有序回归中使用不同链接函数的影响。实验表明:相比现有方法,DO-CNN 具有更高的准确率,且当允许的误差范围扩大时,DO-CNN 依然保持优势。基于广义 Logistic 分布的有序回归分类器相比使用其他分布函数具有更好的效果。DO-CNN 是一种通用的有序性增强方法,可以用来对主流 CNN 模型都进行有序性增强,并且理论上可以应用在除年龄估计任务外的其他有序分类任务中。

## 2 相关工作

### 2.1 人脸年龄估计

Lanitis 等<sup>[17]</sup>应用主动外观模型<sup>[4]</sup>从面部图像中提取形状和外观特征;Geng 等<sup>[5]</sup>使用主动外观模型对同一个人使用一系列年龄递增的面部图像模拟其衰老过程,以构造与不同人衰老过程相关的特定模型;此外,为了解释一个人的长期衰老子空间,Geng 等<sup>[18]</sup>提出了衰老模式子空间技术。

研究初期,由于特定的人可用衰老图像通常非常有限,因此许多研究人员专注于开发非特定人的年龄估计技术。Yang 等<sup>[19]</sup>采用具有局部二进制模式的 AdaBoost 算法<sup>[20]</sup>;Lu 等<sup>[21]</sup>提出一种基于序数判别特征学习的方法;Guo 等<sup>[6]</sup>建议使用带有局部调整的鲁棒衰老流形回归器。

最近,随着大型人脸年龄数据集的出现,基于卷积神经网络的方法在年龄估计领域得到广泛应用,并被证明相比传统方法具有显著提升。Yi 等<sup>[11]</sup>提出一种使用相对浅的卷积神经网络多任务学习方法;Wang 等<sup>[22]</sup>训练了一个更深的卷积神经网络来提取人脸图像的不同层次特征;Levi 等<sup>[23]</sup>验证了卷积神经网络在无约束人脸图像上的性能。

### 2.2 卷积神经网络

目前深度学习领域有多种卷积神经网络模型。这

些模型确切的形式可能会有所不同,但主要组件和计算是相似的。源自 LeNet<sup>[24]</sup>的卷积神经网络模型包含交替组合的卷积层和池化层,然后是全连接层将前几层的特征输入映射为输出结果。在训练过程中,根据损失函数计算当前样本的损失,并通过反向传播算法同时更新各层中的权重。

卷积神经网络已广泛用于各个领域。在自然语言处理方面,SENN 系统在各种任务上取得了最先进的效果<sup>[25]</sup>;在文本分类中,卷积神经网络已被广泛应用并取得了卓越的成果;在计算机视觉领域,图像分类<sup>[26]</sup>、目标检测<sup>[27]</sup>、人脸识别<sup>[28]</sup>和图像分割<sup>[29]</sup>采用卷积神经网络取得了巨大成功。最近,随着卷积神经网络在 GPU 上被实现<sup>[26]</sup>,具有深度架构的卷积神经网络模型在大规模图像数据集(例如 ImageNet 数据集)中的目标检测问题上取得了突破。

## 3 DO-CNN 模型

图 1 描述了基于双有序性约束的卷积神经网络 (DO-CNN) 架构。DO-CNN 的架构包括前半部分的传统卷积神经网络及其变体 (CNNs, 用于抽取人脸特征)、全连接层以及后半部分的有序回归模型分类器与有序竞争比损失函数。流行的卷积神经网络变体,如 ResNet<sup>[30]</sup>、VGG 等均可作为 DO-CNN 架构中的特征抽取模型,它们主要包括卷积层和池化层。之后,为了将 CNNs 学习到的特征映射到样本标记空间中,需要连接几个全连接层,得到的输出值与有序回归分类器连接起来,以进一步得到各类别的概率值,最后采用有序竞争比损失函数计算损失值,并通过反向传播算法优化参数。DO-CNN 在通过有序回归分类器学习人脸年龄相关有序信息的同时,使用有序竞争比损失函数指导模型关注估计结果与实际年龄类别之间的差距。

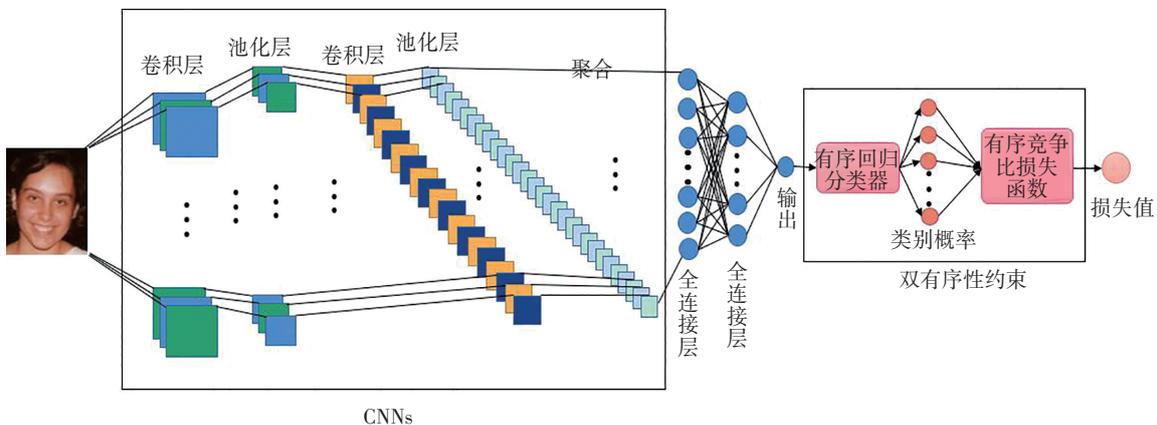


图 1 DO-CNN 模型结构

Fig. 1 Structure of DO-CNN model

本节将依次介绍基于广义 Logistic 分布的有序回归分类器和有序竞争比损失函数的原理以及 DO-CNN 如何应用有序回归分类器、有序竞争比函数指导模型学习年龄标签的有序信息。

### 3.1 基于广义 Logistic 分布的有序回归分类器

有序回归的任务是预测多个存在顺序关系的类别标签<sup>[31]</sup>。现有的有序回归方法分为3类,分别为朴素方法、序数二元分解法和阈值模型。而阈值模型又分为支持向量机模型和有序回归的广义线性模型,目前建模有序问题最流行的方法是后者。因此,本文采用基于广义线性模型的阈值模型方法解决年龄估计的有序问题,其中,选择基于广义 Logistic 分布的链接函数作为累积链接模型中的链接函数。

对于一个有序分类问题,给定  $N$  个样本的训练集  $D = \{(\mathbf{x}_n, y_n)\}_{n=1}^N$ , 其中,  $\mathbf{x}_n = (x_{1n}, \dots, x_{kn})$  为在  $\mathcal{X} \subseteq \mathbf{R}^k$  中取值的随机变量,  $y_n \in Y = \{C_1, C_2, \dots, C_Q\}$  为第  $n$  个样本的标签值。在有序回归中的自然标签排序为  $C_1 < C_2 < \dots < C_Q$ , 其中“ $<$ ”表示顺序关系。有序问题的目标是找到一个分类规则或函数来预测新样本的多个存在顺序关系的类别标签。

累积链接模型<sup>[32]</sup> (Cumulative Link Models, CLMs) 可以预测连续类别组的概率,并考虑到有序尺度,估计累积概率  $P(y \leq C_q | \mathbf{x}_n)$  可以与标准概率直接相关:

$$P(y \leq C_q | \mathbf{x}_n) = P(y = C_1 | \mathbf{x}_n) + \dots + P(y = C_q | \mathbf{x}_n)$$

$$P(y = C_q | \mathbf{x}_n) = P(y \leq C_q | \mathbf{x}_n) - P(y \leq C_{q-1} | \mathbf{x}_n)$$

其中,  $P(y = C_1 | \mathbf{x}_n) = P(y \leq C_1 | \mathbf{x}_n)$ ,  $P(y \leq C_Q | \mathbf{x}_n) = 1$ ,  $q = 2, \dots, Q$ 。

决策规则  $r: \mathcal{X} \rightarrow Y$  不是直接拟合的,而是通过以下一般模型形式满足空间  $\mathcal{X}$  的随机排序:

$$g^{-1}(P(y \leq C_q | \mathbf{x}_n)) = b_q - f(\mathbf{x}_n)$$

其中,  $g^{-1}: [0, 1] \rightarrow (-\infty, +\infty)$ , 是一个单调函数,通常被称为逆链接函数;  $b_q$  是为类  $C_q$  定义的阈值,  $q = 1, 2, \dots, Q-1$ 。该模型基于潜在变量的概念,将输入变量的线性变换作为一维映射。因此,考虑潜在变量  $Z = f(\mathbf{x}_n) + \varepsilon$ , 其中,  $\varepsilon$  是误差为 0 的随机误差分量,记分布为  $F_\varepsilon$ , 当且仅当  $Z \in [b_{q-1}, b_q]$ , 其中潜在变量  $Z$  和  $b = (b_0, b_1, \dots, b_{Q-1}, b_Q)$  从数据中确定时,输入数据被预测为标签  $C_q$ 。假设  $b_0 = -\infty$  和  $b_Q = +\infty$ , 因此,由  $f(\mathbf{x}_n)$  定义的实线被划分为  $Q$  个连续区间,每个间隔对应一个类别。约束条件  $b_1 \leq b_2 \leq \dots \leq b_{Q-1}$  确保了  $P(y \leq C_q | \mathbf{x}_n)$  随着  $q$  的增加而增加。

为了在保证  $b_1 \leq b_2 \leq \dots \leq b_{Q-1}$  的同时应用无约束优化器,可以重新定义阈值。所有阈值都可以从以下形式

的第一个阈值中派生出来<sup>[13]</sup>:

$$b_j = b_1 + \sum_{q=1}^{j-1} \alpha_q^2, j = 2, \dots, Q \quad (1)$$

其中,  $b_1$  是对应于第一个阈值的学习参数,  $\alpha$  是用于获得其余阈值的可学习参数,  $Q$  是类别数。

考虑基于不同分布的链接函数适应不同的数据类型,本文将 Logistic 分布函数的泛化形式引入,可以通过修改两个实际参数来表示广泛使用的分布函数,加入的尺度参数和偏度参数增加了分布函数的灵活性,这种分布被称为广义 Logistic 分布<sup>[33]</sup>, 它的概率分布函数和累积分布函数的形式如下:

$$f_{\lambda, \mu, \gamma}(x) = \frac{\lambda \gamma \exp(-\lambda(x-\mu))}{(1 + \exp(-\lambda(x-\mu)))^{\gamma+1}}$$

$$F_{\lambda, \mu, \gamma}(x) = \frac{1}{(1 + \exp(-\lambda(x-\mu)))^\gamma}$$

其中,  $\lambda$ ,  $\gamma$  和  $\mu$  分别表示尺度参数、偏度参数和位置参数。

根据 Gupta 和 Kundu<sup>[34]</sup> 的研究,经过适当的归一化后,  $n$  个独立同广义 Logistic 分布随机变量的最大值和最小值分布趋向于 Gumbel 的极值分布。因此,它可以用于建模左右倾斜的潜在变量 (Gumbel (max) 分布和 Gumbel (min) 分布)。当  $\gamma = 1$  且  $\lambda = 1$  时,广义 Logistic 分布 ( $\mu = 0$ ) 等同于标准的 Logistic 分布,因此它也可以用于建模标准的 Logistic 潜在变量。同样,通过改变  $\lambda$  和  $\gamma$  的值,该分布可以建模基于不同分布的潜在变量,以提升模型的泛化能力。

由于位置参数是本文中所采用的基于广义 Logistic 分布的潜在变量值的转换,所以可以得到广义 Logistic 分布下,本文的链接函数和概率函数的形式:

$$g \log it(P(y \leq C_q | \mathbf{x}_n)) = b_q - f(\mathbf{x}_n), q = 1, 2, \dots, Q-1$$

$$P(y \leq C_q | \mathbf{x}_n) = \frac{1}{(1 + \exp(-\lambda(b_q - f(\mathbf{x}_n))))^\gamma} \quad (2)$$

其中,  $\gamma$  为偏度参数,  $\lambda$  为尺度参数。本文将选择式 (2) 表示的基于广义 Logistic 分布的链接函数作为累积链接模型的链接函数。

经过有序回归分类器后,样本被分到类  $C_j$ , 即第  $j$  个类别的概率可以表示为

$$p(y = C_j | \mathbf{x}_n) = P(y \leq C_j | \mathbf{x}_n) - P(y \leq C_{j-1} | \mathbf{x}_n) \quad (3)$$

其中,  $j = 2, 3, \dots, Q$ 。

### 3.2 有序竞争比损失函数

对于交叉熵损失函数、负对数似然比损失函数以及竞争比损失函数,都未考虑年龄的有序性特征。因此,在对年龄数据集进行估计时,会产生一定的误差。

为了将有序性关系进一步融入分类任务中,以贝叶斯风险准则为基础,采用动态权重损失函数将类别信息和序数信息进行融合<sup>[35]</sup>,得到有序竞争比损失函数。

首先,从贝叶斯风险准则角度出发,将分类任务的损失表示为条件风险:

$$R(\mathbf{x}_n, C_j) = \sum_{k=1}^Q r(C_j, C_k) p(C_k | \mathbf{x}_n)$$

其中,  $r(C_j, C_k) = \begin{cases} 0, & C_j = C_k \\ 1, & C_j \neq C_k \end{cases}$ ,  $C_k$  表示根据样本  $\mathbf{x}_n$  得到的预测值,  $r(C_j, C_k)$  表示将样本真实类别  $C_j$  判断为类别  $C_k$  的风险,  $p(C_k | \mathbf{x}_n)$  表示样本  $\mathbf{x}_n$  被分到类别  $C_k$  的概率。若将有序性标签引入,则风险项不再是单位 1<sup>[35]</sup>。在有序性约束中,将儿童错分为中年人和老年人的损失是不同的,并且前者的损失要小于后者。因此,本文考虑与真实值和预测值之间距离呈正比的如下风险项:

$$r(C_j, C_k) = 1 + \frac{(C_j - C_k)^2}{(Q-1)^2}$$

根据文献[15],竞争比损失函数表示为

$$L_{\text{CRL}} = \log \frac{(\alpha + \sum_{k=1, k \neq c}^C \hat{p}(y_k | x))^\beta}{\hat{p}(y_c | x)}$$

其中,  $c$  表示真实类别,  $C$  表示总类别的数量,  $\hat{p}(y_c | x)$  表示被分到真实类别的概率,  $\hat{p}(y_k | x)$  表示被分到第  $k$  个类别的概率,  $\sum_{k=1, k \neq c}^C \hat{p}(y_k | x)$  表示根据模型输出值得到的竞争错误类别的总概率,  $\alpha \geq 1$  保证损失值持续为正,即 CRL 的优化是一个最小化问题,超参数  $\beta$  修正了竞争比的权重。

然后,将风险项融入竞争比损失函数中,赋予竞争的错误类别一个风险权重,得到有序竞争比损失函数(Ordered Competitive Ratio Loss Function, OCRL):

$$L_{\text{OCRL}} = \log \frac{\left( \alpha + \left( 1 + \frac{(C_j - C_k)^2}{(Q-1)^2} \right) * \sum_{k=1, k \neq j}^Q \hat{p}(C_k | \mathbf{x}_n) \right)^\beta}{\hat{p}(C_j | \mathbf{x}_n)} = \beta \log \left( \alpha + \left( 1 + \frac{(C_j - C_k)^2}{(Q-1)^2} \right) * (1 - \hat{p}(C_j | \mathbf{x}_n)) \right) - \log \hat{p}(C_j | \mathbf{x}_n)$$

其中,  $\alpha$  的默认值为 1.5,  $\beta$  的默认值为 1.0,  $C_k$  表示预测的类别,  $C_j$  表示真实类别,  $Q$  表示类别的总数量,  $\hat{p}(C_j | \mathbf{x}_n)$  表示被分到真实类别的概率,  $\hat{p}(C_k | \mathbf{x}_n)$  表示被分到第  $k$  个类别的概率。当  $C_j = C_k$  时,与原竞争比损失函数的表达式相同。

### 3.3 模型学习

用  $f_{(j)} = (b_j - f(\mathbf{x}_n))$  表示分类器的输入值,其中  $f(\mathbf{x}_n)$  表示神经网络的输出值,有序回归分类器真实类别的输出概率为  $p_j$ ,根据神经网络的链式法则,有

$$\frac{\partial L_{\text{OCRL}}}{\partial f(\mathbf{x}_n)} = \frac{\partial L_{\text{OCRL}}}{\partial p_j} \times \frac{\partial p_j}{\partial f(\mathbf{x}_n)}$$

对于有序回归的阈值,根据 3.1 节的约束条件式

$$(1), \text{可以得到: 当 } j=1 \text{ 时, } \frac{\partial f_{(j)}}{\partial b_1} = 1; \text{ 当 } j>q \text{ 时, } \frac{\partial f_{(j)}}{\partial \alpha_q} = 2\alpha_q。$$

根据式(2)和式(3)可以得到:

$$\frac{\partial p_j}{\partial f(\mathbf{x}_n)} = -\frac{\lambda \gamma \exp(-\lambda f_{(j)})}{(1 + \exp(-\lambda f_{(j)}))^{\gamma+1}} + \frac{\lambda \gamma \exp(-\lambda f_{(j-1)})}{(1 + \exp(-\lambda f_{(j-1)}))^{\gamma+1}}$$

即可以得到,当  $k=j$  时,

$$\frac{\partial L_{\text{OCRL}}}{\partial f(\mathbf{x}_n)} = \left( -\frac{\beta}{\alpha+1-p_j} - \frac{1}{p_j} \right) \times \left( -\frac{\lambda \gamma \exp(-\lambda f_{(j)})}{(1 + \exp(-\lambda f_{(j)}))^{\gamma+1}} + \frac{\lambda \gamma \exp(-\lambda f_{(j-1)})}{(1 + \exp(-\lambda f_{(j-1)}))^{\gamma+1}} \right)$$

当  $k \neq j$  时,

$$\frac{\partial L_{\text{OCRL}}}{\partial f(\mathbf{x}_n)} = \left( -\frac{\beta(1+\omega_{j,k})}{\alpha+(1+\omega_{j,k}) \times (1-p_j)} - \frac{1}{p_j} \right) \times \left( -\frac{\lambda \gamma \exp(-\lambda f_{(j)})}{(1 + \exp(-\lambda f_{(j)}))^{\gamma+1}} + \frac{\lambda \gamma \exp(-\lambda f_{(j-1)})}{(1 + \exp(-\lambda f_{(j-1)}))^{\gamma+1}} \right)$$

其中,  $\omega_{j,k} = (C_j - C_k)^2 / (Q-1)^2$ 。

可以看到:在梯度更新过程中,有序竞争比损失函数加入的风险项,也起到了一定的作用,在反向传播过程中,影响着神经网络参数的更新。同时,可以通过改变广义链接函数中超参数的值,以适应不同样本的分布。

## 4 实验设置与结果分析

### 4.1 数据集选择

FGNET 数据集由塞浦路斯大学采集,其中包含 82 个个体在不同年龄时期的 1 002 张彩色图像,每个人含有 6~18 张照片,这些照片在不同光照以及不同神态的条件下拍摄,年龄范围为 0—69 岁。根据年龄范围划分为 6 类,年龄分布表如表 1 所示。

表 1 划分后的 FGNET 数据集年龄分布表

Table 1 Age distribution table for the delimited FGNET dataset

年龄段	0-6	7-12	13-17	18-26	27-39	>39
数量	274	207	159	182	111	69

AgeDB 数据集包含来自西方国家的 568 位名人的 16 488 张图片,每张图片都注明了年龄信息,每个人平均包含 29 张图片,年龄范围为 3—101 岁。根据年龄的范围划分为 8 类,年龄分布表如表 2 所示。

表 2 划分后的 AgeDB 数据集年龄分布表

**Table 2 Age distribution table for the delimited AgeDB dataset**

年龄段	0-	13-	23-	32-	42-	51-	64-	>73
数量	12	22	31	41	50	63	73	
	95	959	2 695	3 617	2 775	3 236	1 835	1 276

表 3 训练环境配置

**Table 3 Configurations of the training environment**

名称	CPU	GPU	框架
配置	E5-2673 v3	NVIDIA Corporation Device 2184	Pytorch-GPU

4.2 实验设置

4.2.1 模型训练环境配置

本文模型训练的环境配置如表 3 所示,采用 E5-2673 v3 @ 2.40 GHz 服务器进行实验,内存为 64 G,显卡驱动为 CUDA10.2,深度学习框架为 Pytorch-GPU。

4.2.2 网络模型的选择

为了验证本文提出的有序回归模型分类器与有序竞争比损失函数的有效性,采用 Beckham C 等<sup>[36]</sup>使用的 ResNet 作为图 1 前半部分的卷积神经网络,对人脸图像进行特征提取。其模型结构如表 4 所示。

表 4 本文采用的卷积神经网络模型的结构

Table 4 The structure of the convolutional neural network used in this paper

层名称	层类型	说明	输出	数量
Conv	卷积层	卷积核个数:32;卷积核大小:7 * 7;步长:2	Batchsize * 32 * 112 * 112	1
Pool1	最大池化层	池化层大小:3 * 3;步长:2	Batchsize * 32 * 55 * 55	1
Res-Block	残差单元	卷积核个数:32;卷积核大小:3 * 3;步长:1	Batchsize * 32 * 55 * 55	2
Res-Block	残差单元	卷积核个数:64;卷积核大小:3 * 3;步长:2	Batchsize * 64 * 28 * 28	1
Res-Block	残差单元	卷积核个数:64;卷积核大小:3 * 3;步长:1	Batchsize * 64 * 28 * 28	2
Res-Block	残差单元	卷积核个数:128;卷积核大小:3 * 3;步长:2	Batchsize * 128 * 14 * 14	1
Res-Block	残差单元	卷积核个数:128;卷积核大小:3 * 3;步长:1	Batchsize * 128 * 14 * 14	2
Res-Block	残差单元	卷积核个数:256;卷积核大小:3 * 3;步长:2	Batchsize * 256 * 7 * 7	1
Res-Block	残差单元	卷积核个数:256;卷积核大小:3 * 3;步长:1	Batchsize * 256 * 7 * 7	2
Pool2	平均池化层	池化层大小:7 * 7;步长:2	Batchsize * 256 * 1 * 1	1

考虑 Elu 激活函数相比于 Relu 激活函数具有更好的泛化能力,以及它可以减轻梯度弥散问题的影响,因此,将在卷积神经网络中采用 Elu 激活函数。

4.2.3 评价指标的选择

本文将 FGNET 数据集划分为 70% 的训练集,10% 的验证集,20% 的测试集;将 AgeDB 数据集划分为 80% 的训练集,10% 的验证集,10% 的测试集。除此之外,迭代次数设置为 100 次,选择网络的初始学习率为 0.001,每次迭代的批处理大小为 16,训练模型时采用 Adam 优化算法。由于本文研究的是有序分类问题,故主要采用以下几种有序评估指标。

(1) 准确度 (Accuracy,  $A_{cc}$ ):

$$A_{cc} = \frac{N_c}{N} \times 100\%$$

其中,  $N_c$  表示正确分类的数量,  $N$  是样本的数量。

(2) 平均绝对误差 (Mean Absolute Error,  $R_{MAE}$ ):

$$R_{MAE} = \frac{1}{N} \sum_{i=1}^N |y'_i - y_i|$$

其中,  $N$  是样本的数量,  $y'_i$  为第  $i$  个样本的预测概率值,  $y_i$  为第  $i$  个样本采用 one-hot 编码的真实值。

(3) 1-off 准确率 (1-off accuracy, 1-off  $A_{cc}$ ): 若预测类别处于正确的类别上下一段年龄范围内时,预测结果被定为准确。

4.3 实验结果与分析

4.3.1 不同累积链接函数对模型性能的影响

针对广义 Logistic 链接函数中的偏度参数  $\gamma$  以及尺度参数  $\lambda$ , 进行超参数敏感性实验,探索不同链接函数对 DO-CNN 模型的影响。本文使用 4.2 节中提到的模型评估了它们对 FGNET 和 AgeDB 数据集的分类影响。选取  $\gamma$  的值为 0.25、0.5、1.0、1.5 和 2.0,  $\lambda$  的值为 0.25、0.5、1.0、2.0 和 4.0。

图 2、图 3 和图 4 分别展示了 FGNET 数据集中超参数  $\gamma$  和  $\lambda$  的影响。可以明显地看出:在 FGNET 数据集上,当  $\gamma=2$  时,3 个评价指标上的测试结果明显比其他条件下都要差。除此之外,还可以看出:当  $\gamma=0.25$  以及 0.5 时,在评价指标准确率上的测试结果是

最优的,为 39.31%;当  $\gamma=0.25$ 、 $\gamma=0.5$  以及当  $\gamma=1.0$  且  $\lambda=0.25$ 、 $0.5$ 、 $1.0$  以及  $4.0$  时,在评价指标 1-off 准确率上是最优的,为 73.13%;当  $\gamma=0.25$  时,评价指标  $R_{MAE}$  的值均低于其他条件下的结果,且当  $\lambda=4.0$  时,达到最低值,为 0.218 3。

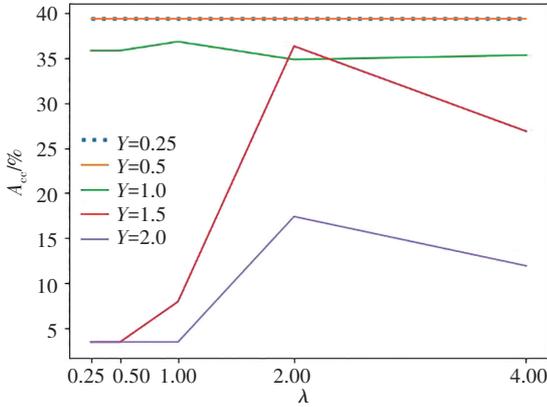


图 2 FGNET 中在不同  $\gamma$  和  $\lambda$  变化下的  $A_{cc}$  性能测试结果  
Fig. 2  $A_{cc}$  performance test results under different variations of  $\gamma$  and  $\lambda$  in FGNET

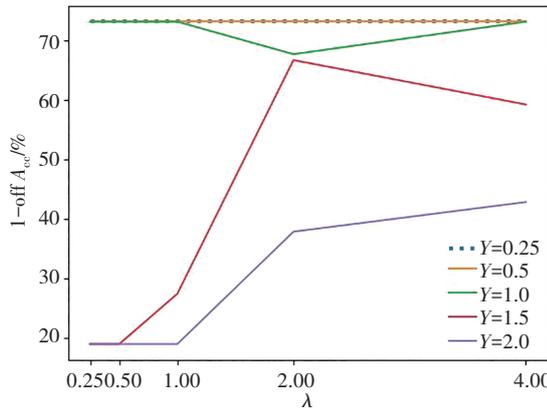


图 3 FGNET 中在不同  $\gamma$  和  $\lambda$  变化下的 1-off  $A_{cc}$  性能测试结果  
Fig. 3 1-off  $A_{cc}$  performance test results under different variations of  $\gamma$  and  $\lambda$  in FGNET

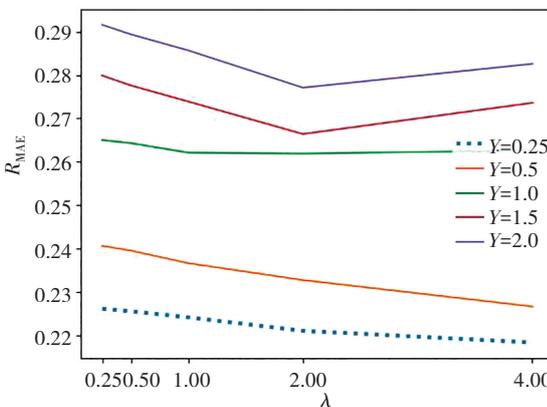


图 4 FGNET 中在不同  $\gamma$  和  $\lambda$  变化下的  $R_{MAE}$  性能测试结果  
Fig. 4  $R_{MAE}$  performance test results under different variations of  $\gamma$  and  $\lambda$  in FGNET

图 5、图 6 和图 7 分别展示了 AgeDB 数据集中超参数  $\gamma$  和  $\lambda$  的影响。同样,可以明显地看出:在 AgeDB 数据集上,当  $\gamma=1.0$  且  $\lambda=0.5$  时,在准确率、1-off 准确率和  $R_{MAE}$  评价指标上的测试结果均达到最优值,分别为 26.49%、61.21%和 0.256 1。

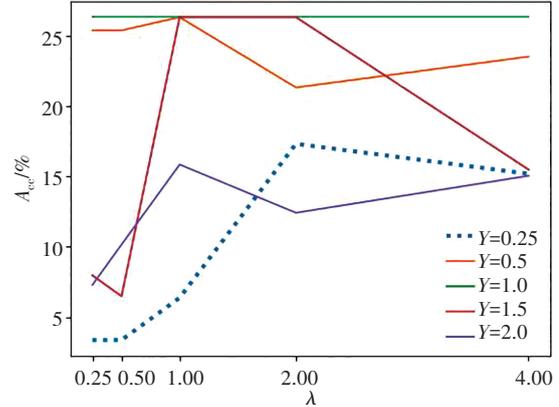


图 5 AgeDB 中在不同  $\gamma$  和  $\lambda$  变化下的  $A_{cc}$  性能测试结果  
Fig. 5  $A_{cc}$  performance test results under different variations of  $\gamma$  and  $\lambda$  in AgeDB

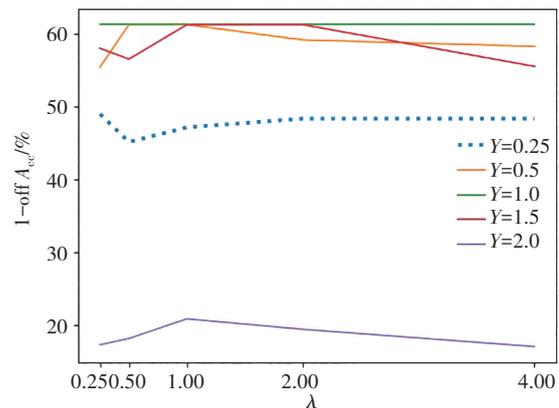


图 6 AgeDB 中在不同  $\gamma$  和  $\lambda$  变化下的 1-off  $A_{cc}$  性能测试结果  
Fig. 6 1-off  $A_{cc}$  performance test results under different variations of  $\gamma$  and  $\lambda$  in AgeDB

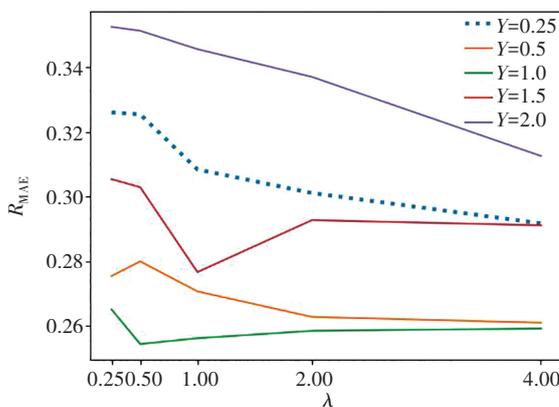


图 7 AgeDB 中在不同  $\gamma$  和  $\lambda$  变化下的  $R_{MAE}$  性能测试结果  
Fig. 7  $R_{MAE}$  performance test results under different variations of  $\gamma$  and  $\lambda$  in AgeDB

综上所述表明:对于 FGNET 数据集,  $\gamma = 0.25$  且  $\lambda = 4.0$  时的广义 Logistic 链接函数是最优的;对于 AgeDB 数据集,  $\gamma = 1.0$  且  $\lambda = 0.5$  时的广义 Logistic 链接函数是最优的。说明选择不同类型的累积链接函数,会对模型的性能产生影响。

### 4.3.2 消融实验

为了验证 DO-CNN 中的有序回归分类器和有序竞争比函数组件对于模型整体的有效性,本文进行了消融实验,结果分别如表 5 和表 6 所示。DO-CNN<sub>reg+soft</sub> 表示在 DO-CNN 中关闭有序回归分类器并使用 Softmax 函数作为模型分类器;DO-CNN<sub>log+probit</sub> 表示在 DO-CNN 中将有序回归中的广义 Logistic 函数替换为 Probit 函数;DO-CNN<sub>cp</sub> 表示将 DO-CNN 中的有序竞争比损失函数退化为普通竞争比损失函数;DO-CNN<sub>ep</sub> 表示 DO-CNN 中使用交叉熵损失函数;DO-CNN<sub>np</sub> 表示 DO-CNN 中使用负对数似然比损失函数。

表 5 FGNET 数据集中的消融实验结果

Table 5 Ablation experiment results in FGNET

模 型	评估指标		
	$A_{cc}/\%$	1-off $A_{cc}/\%$	$R_{MAE}$
DO-CNN <sub>reg+soft</sub>	21.86	56.69	0.2803
DO-CNN <sub>log+probit</sub>	35.82	67.16	0.2420
DO-CNN <sub>cp</sub>	27.52	60.27	0.2613
DO-CNN <sub>ep</sub>	25.96	58.73	0.2911
DO-CNN <sub>np</sub>	27.01	59.48	0.2885
DO-CNN	39.31	73.13	0.2183

表 6 AgeDB 数据集中的消融实验结果

Table 6 Ablation experiment results in AgeDB

模 型	评估指标		
	$A_{cc}/\%$	1-off $A_{cc}/\%$	$R_{MAE}$
DO-CNN <sub>reg+soft</sub>	20.83	54.95	0.3024
DO-CNN <sub>log+probit</sub>	25.32	60.19	0.2817
DO-CNN <sub>cp</sub>	23.30	57.36	0.2988
DO-CNN <sub>ep</sub>	21.09	56.10	0.3079
DO-CNN <sub>np</sub>	22.61	56.95	0.3016
DO-CNN	26.49	61.21	0.2561

带有有序回归和有序竞争比损失函数的方法 DO-CNN 和 DO-CNN<sub>log+probit</sub> 相比其他 4 种消融后的方法在 3 种指标上都具有显著优势,并且基于广义 Logistic 链接函数的有序回归分类器相比基于 Probit 链接函数具有明显提升。总的来说,DO-CNN<sub>reg+soft</sub> 的实验结果在  $A_{cc}$  和 1-off  $A_{cc}$  指标上最弱,这说明年龄标签中有序信息的利用是年龄估计任务中的重要过程。在包含有序回归分类器的方法中,DO-CNN<sub>cp</sub>、DO-CNN<sub>ep</sub> 和

DO-CNN<sub>np</sub> 的实验结果  $R_{MAE}$  均较大,相反包含有序竞争比函数的方法的  $R_{MAE}$  较小,这表明有序竞争比损失函数能够有效指导模型缩小估计年龄与真实年龄之间的距离。

### 4.3.3 不同算法的对比实验分析

对比实验给出了 4 种年龄估计方法在相同实验条件下的 3 种不同评价指标的结果,如表 7 和表 8 所示。用 Ranking-CNN 表示 Chen 等<sup>[12]</sup> 提出的排序卷积神经网络;用 CRL-CNN 表示 Zhang 等<sup>[15]</sup> 提出的引入竞争比损失函数的卷积神经网络模型,根据该文献,超参数选择为  $\alpha = 1.5, \beta = 1.0$ ;用 Kappa-CNN 表示 de La Torre 等<sup>[14]</sup> 提出的引入二次加权 Kappa 损失函数的神经网络模型。

从表 7 和表 8 可以看出:在 3 个评价指标上,DO-CNN 在 FGNET 数据集和 AgeDB 数据集上表现的结果均要优于其他 3 种方法。Ranking-CNN 通过排序卷积神经网络关注年龄标签中的有序信息,其结果仅次于 DO-CNN,也证明了在年龄估计任务中引入有序信息的必要性。Kappa-CNN 和 CRL-CNN 中使用的分类器和损失函数,依然是在传统分类任务上的改进,无法关注到年龄标签中的有序信息,因此结果相对较差。这一结果表明:本文提出的 DO-CNN 模型在解决基于分类模型的人脸年龄估计是有效的,并在序数评价指标上的表现具有一定的优越性。

表 7 4 种人脸估计方法在 FGNET 中的  $A_{cc}$ 、1-off  $A_{cc}$  和  $R_{MAE}$  的实验结果

Table 7 Experimental results of  $A_{cc}$ , 1-off  $A_{cc}$  and  $R_{MAE}$  for four face estimation methods in FGNET

模 型	评估指标		
	$A_{cc}/\%$	1-off $A_{cc}/\%$	$R_{MAE}$
Ranking-CNN	27.86	62.69	0.2513
CRL-CNN	25.36	60.19	0.2490
Kappa-CNN	15.92	55.22	0.2803
DO-CNN	39.31	73.13	0.2183

表 8 4 种人脸估计方法在 AgeDB 中的  $A_{cc}$ 、1-off  $A_{cc}$  和  $R_{MAE}$  的实验结果

Table 8 Experimental results of  $A_{cc}$ , 1-off  $A_{cc}$  and  $R_{MAE}$  for four face estimation methods in AgeDB

模 型	评估指标		
	$A_{cc}/\%$	1-off $A_{cc}/\%$	$R_{MAE}$
Ranking-CNN	23.90	59.87	0.2599
CRL-CNN	21.45	58.69	0.2702
Kappa-CNN	13.98	50.84	0.3065
DO-CNN	26.49	61.21	0.2561

图 8 和图 9 分别显示了 4 种对比方法在真实数据集 FGNET 和 AgeDB 上的年龄估计结果案例。其中第二行表示目标图像属于的真实年龄段,后续 4 行表示针对目标图像各对照方法的估计结果。结果表明:DO-CNN 的估计结果相比其他 3 种方法更接近真实年龄。此外,DO-CNN 在 AgeDB 数据集上的效果明显弱于 FGNET 数据集上的效果,这是由于 AgeDB 中的数据包含了较多不同种族、非正面姿态以及夸张表情特征的图像,而这些因素将干扰年龄估计任务的准确性,例如 DO-CNN 以及之前的模型都在图 9 中的 G3(大笑表情)和 G6(侧脸图像)上效果不好。



Figure	F1	F2	F3	F4	F5	F6
Label	0-6	7-12	13-17	18-26	27-39	>39
Ranking-CNN	7-12	7-12	27-39	18-26	27-39	>39
CRL-CNN	0-6	13-17	27-39	18-26	27-39	>39
Kappa-CNN	0-6	18-26	13-17	>39	>39	27-39
DO-CNN	0-6	7-12	13-17	18-26	27-39	>39

图 8 4 种对照方法在 FGNET 数据集上的预测结果案例

Fig. 8 Examples of prediction results of four comparison methods on FGNET dataset



Figure	G1	G2	G3	G4	G5	G6	G7	G8
Label	0-12	13-22	23-31	32-41	42-50	51-63	64-73	>73
Ranking-CNN	0-12	32-41	42-50	32-41	64-73	64-73	>73	>73
CRL-CNN	0-12	23-31	42-50	32-41	42-50	>73	51-63	64-73
Kappa-CNN	13-22	23-31	51-63	42-50	32-41	>73	64-73	51-63
DO-CNN	0-12	13-22	42-50	32-41	42-50	64-73	64-73	>73

图 9 4 种对照方法在 AgeDB 数据集上的预测结果案例

Fig. 9 Examples of prediction results of four comparison methods on AgeDB dataset

#### 4.4 结论

通过实验,可以看出 DO-CNN 在人脸图像年龄估计任务上具有较好的效果,双有序性约束能够加强模型对有序信息的学习能力。DO-CNN 着重解决如何有效利用人脸年龄中的有序信息问题,无法减轻表情、种族等特征的干扰,需要具体的 CNN 模型在 DO-CNN 中解决干扰特征问题。

#### 5 结束语

本文提出基于双有序性约束卷积神经网络模型的人脸图像年龄估计方法:DO-CNN。通过使用有序回归作为分类器,使模型在学习关注年龄标签中的有序信息;通过引入有序竞争比函数,从风险项的角度将有序性融入损失函数中,指导模型在学习关注估计年龄与真实年龄之间的差距。实验表明:DO-CNN 在人脸图像年龄估计任务上具有较好的效果。

#### 参考文献(References):

- [1] 房国志,孙康瞳. 多尺度 YOLO 人脸年龄估计方法研究[J]. 计算机工程与应用, 2019, 55(21): 135—141.  
FANG Guo-zhi, SUN Kang-tong. Research on face age estimation based on multi-scale YOLO model [J]. Computer Engineering and Applications, 2019, 55(21): 135—141.
- [2] 张晓宁. 基于机器学习的年龄估计算法的研究[D]. 成都: 电子科技大学, 2019.  
ZHANG Xiao-ning. Research on age estimation algorithm based on machine learning[D]. Chengdu: University of Electronic Science and Technology of China, 2019.
- [3] KWON Y H, DA VITORIA LOBO N. Age classification from facial images[J]. Computer Vision and Image Understanding, 1999, 74(1): 1—21.
- [4] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 681—685.
- [5] GENG X, ZHOU Z H, SMITH-MILES K. Automatic age estimation based on facial aging patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(12): 2234—2240.
- [6] GUO G, FU Y, DYER C R, et al. Image-based human age estimation by manifold learning and locally adjusted robust regression [J]. IEEE Transactions on Image Processing, 2008, 17(7): 1178—1188.
- [7] GUO G, MU G, FU Y, et al. Human age estimation using bio-inspired features [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 112—119.
- [8] WANG X, GUO R, KAMBHAMETTU C. Deeply-learned feature for age estimation [C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. IEEE, 2015: 534—541.
- [9] GUO G, MU G, FU Y, et al. Human age estimation using bio-inspired features [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 112—119.
- [10] GUNAY A, NABIYEV V V. Automatic age classification with LBP[C]//Proceedings of the 23rd International Symposium on Computer and Information Sciences. IEEE, 2008: 1—4.
- [11] YI D, LEI Z, LI S Z. Age estimation by multi-scale convolutional network [C]//Asian Conference on Computer Vision. Springer, Cham, 2014: 144—158.
- [12] CHEN S, ZHANG C, DONG M, et al. Using ranking-CNN for age estimation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5183—5192.

- [13] VARGAS V M, GUTIERREZ P A, HERVAS-MARTINEZ C. Cumulative link models for deep ordinal classification [J]. *Neurocomputing*, 2020, 401(7): 48—58.
- [14] DE LA TORRE J, PUIG D, VALLS A. Weighted kappa loss function for multi-class classification of ordinal data in deep learning[J]. *Pattern Recognition Letters*, 2018, 105: 144—154.
- [15] ZHANG K, GUO Y, WANG X, et al. Competing ratio loss for discriminative multi-class image classification [J]. *Neurocomputing*, 2021, 464(2): 473—484.
- [16] YAO H, ZHU D, JIANG B, et al. Negative log likelihood ratio loss for deep neural network classification [C]// *Proceedings of the Future Technologies Conference*. Springer, Cham, 2019: 276—282.
- [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] GENG X, ZHOU Z H, ZHANG Y, et al. Learning from facial aging patterns for automatic age estimation [C]// *Proceedings of the 14th ACM International Conference on Multimedia*. 2006: 307—316.
- [19] YANG Z, AI H. Demographic classification with local binary patterns [C]// *International Conference on Biometrics*. Berlin, Heidelberg Springer, 2007: 464—473.
- [20] AHONEN T, HADID A, PIETIKAINEN M. Face description with local binary patterns: Application to face recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 2037—2041.
- [21] LU H. Learning ordinal discriminative features for age estimation [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012: 2570—2577.
- [22] WANG X, GUO R, KAMBHAMETTU C. Deeply-learned feature for age estimation [C]// *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2015: 534—541.
- [23] LEVI G, HASSNER T. Age and gender classification using convolutional neural networks [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015: 34—42.
- [24] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278—2324.
- [25] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch [J]. *Journal of Machine Learning Research*, 2011, 12(ARTICLE): 2493—2537.
- [26] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84—90.
- [27] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014: 580—587.
- [28] TAIGMAN Y, YANG M, RANZATO M, et al. Deepface: Closing the gap to human-level performance in face verification [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014: 1701—1708.
- [29] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 3431—3440.
- [30] 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*. 2016: 770—778.
- [31] GUTIERREZ P A, PEREZ-ORTIZ M, SANCHEZ-MONEDERO J, et al. Ordinal regression methods: Survey and experimental study [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(1): 127—146.
- [32] MCCULLAGH P. Regression models for ordinal data [J]. *Journal of the Royal Statistical Society Series B (Statistical Methodological)*, 1980, 42(2): 109—127.
- [33] FERNANDEZ-NAVARRO F. A generalized logistic link function for cumulative link models in ordinal regression [J]. *Neural Processing Letters*, 2017, 46(1): 251—269.
- [34] GUPTA R D, KUNDU D. Theory & methods: Generalized exponential distributions [J]. *Australian & New Zealand Journal of Statistics*, 1999, 41(2): 173—188.
- [35] ZHANG C, ZHU C, LIU Y, et al. Image ordinal estimation: classification and regression benefit each other [C]// *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2017: 275—278.
- [36] BECKHAM C, PAL C. Unimodal probability distributions for deep ordinal classification [C]// *Proceedings of the International Conference on Machine Learning*. PMLR, 2017: 411—419.