

基于区域的年龄估计模型研究

孙劲光 荣文钊

(辽宁工程技术大学电子与信息工程学院 辽宁 葫芦岛 125105)

摘 要 随着年龄特征提取和年龄特征分类模式研究的不断深入,为了进一步满足基于年龄信息的人机交互系统在现实生活中的应用需求,构建有效的机器学习算法已成为人脸图像年龄估计技术的研究热点之一。首先,通过分析人脸图像的多个区域特征随年龄变化的规律,将面部分为前额区域、眼部区域、面中部区域及人脸整体区域,并分别构建深度卷积神经网络特征提取模型,实现每个区域年龄的特征提取;其次,以 Morph 人脸库为样本集,将其划分为 10~19 岁、20~29 岁、30~39 岁、40~49 岁、50~59 岁、60 岁以上 6 个年龄段,完成多区域年龄特征提取网络模型的训练及测试;最后,依据多区域网络年龄特征分类的准确率,确定基于区域的动态权重年龄估计模型。实验表明:所提模型在 Morph 人脸库中的年龄估计准确率达到 72.6%,也将该人脸库的年龄分类类别由 4 个提升到 6 个。

关键词 年龄估计,年龄特征,动态权重,Morph 年龄库,深度学习

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.08.008

Research on Regional Age Estimation Model

SUN Jin-guang RONG Wen-zhao

(School of Electronic and Information Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China)

Abstract With the further research on age feature extraction and age feature classification pattern, in order to make further efforts to meet the application demand of human-computer interaction system based on age information in real life, constructing an effective machine learning algorithm has become a research focus in age estimation technology of face image. Firstly, this paper analyzed the rule of multiple regional features changing with age, and divided the face into prefrontal region, eye region, central region and integrated region. Then, it constructed features extraction model of deep convolutional neural network models separately to extract age features of each region. Thirdly, taking Morph face database as the sample set, this paper divided it into 6 stages aged 10~19, 20~29, 30~39, 40~49, 50~59, and 60 years or older to train and test age feature extraction network model in multiple regions. Finally, according to the accuracy of age feature classification model, this paper defined the region-based dynamic weight age estimation model. The experiment shows that the accuracy of age estimation on Morph face database is 72.6%, and the age classification category has been raised from 4 to 6.

Keywords Age estimation, Age characteristics, Dynamic weights, Morph age database, Deep learning

1 引言

在生物信息识别领域中,使用人脸图像进行年龄估计已经成为一个重要的研究课题。深度学习算法在计算机视觉、模式识别等领域被广泛应用,并且取得了前所未有的识别精度,使用深度学习算法进行人脸图像年龄估计是当下年龄估计研究中最受关注的研究内容之一。

1980 年以来,国内外关于人脸识别、表情识别、姿态识别等人脸图像的相关问题已取得了丰硕的研究成果。相对而言,年龄估计的研究成果较少,其原因集中表现为:在人类成长过程中,引起人面部变化因素的复杂度不一,且缺少有效的定义方法;人脸面部年龄特征受性别、种族、生活环境、职业、

遗传等多方因素共同影响;人类年龄具有“伪装性”,肉眼很难对一个人的年龄进行准确判断。随着计算机视觉和深度学习的飞速发展,科研人员也在年龄估计领域取得了巨大突破和进展,常见的年龄估计方法包括基于人体测量学模型、外观模型、特征建模和深度学习等方法。

(1) 基于人体测量学模型的方法

通过测量提取头颅从婴儿到成年人生长过程中的变化,并计算眼睛、鼻子和嘴之间的 6 种几何比例,在利用 Sobel 算子边缘检测提取到眼睛、鼻子和嘴巴位置的基础上计算两种器官间距离的比值作为 BP 神经网络训练的输入,以此区分婴儿和成年人;通过提取面部前额区与下眼睑区域的皱纹特征区分年轻人与老年人^[1-2];通过定位眼睛、鼻子、嘴、下巴、脸

部边缘以及额头 6 个特征点,提取它们之间的 7 种比值作为 BP 神经网络训练参数;并使用 Canny 算子边缘检测方法提取人脸面部中前额、眼尾和脸颊 3 个区域的皱纹特征,以皱纹密度作为 BP 神经网络的训练参数。最后,训练一个 4 层 BP 神经网络作为分类器对 15 岁及以下、16~30 岁、31~50 岁、50 岁及以上年龄阶段进行分类,获得了 86.64% 的平均分类正确率^[3]。

(2) 基于外观模型的方法

基于外观模型的方法主要针对人的肤色、头发颜色和皱纹分布等进行建模。通过 Sobel 算子进行边缘检测^[4-5],提取局部纹理特征并计算其密度,通过主成份分析法(PCA)和多元判别分析法(MDA)对灰度直方图进行降维以获取颜色特征^[5],通过支持向量机(SVM)在 20~30 岁、31~40 岁、41~50 岁、51~60 岁 4 个年龄段获得 72.52% 的平均分类准确率^[6]。使用 Gabor 滤波更加精细地提取了不同年龄人脸的皮肤纹理特征,在 YGA 年龄库进行年龄估计取得了 3.47a 的男性平均绝对误差、3.91a 的女性平均绝对误差,在 FG-NET 年龄库上取得了 4.77a 的平均绝对误差^[7]。

(3) 基于特征建模的方法

基于特征建模方法的主要代表模型有:从回归角度和分类角度出发,基于典型相关分析和代价敏感学习分别处理年龄估计问题,并使用集成学习法将两种方法所得结果结合起来,获得了 3.4863a 的平均绝对误差^[8];通过改进的非负矩阵分解法(NMF),建立具有年龄代表力的子空间对原图像进行表示,并采用径向基神经网络(RBF)进行训练和预测,获得了 6.23a 的平均绝对误差^[9-10];由同一个对象在不同年龄段的人脸图像按时间顺序进行排列构成该对象的年龄模式,使用 PCA 对不同对象的年龄模式进行人脸图像降维,通过提取特征向量得到年龄模式子空间(AGing pattErn Subspace, AGES),并通过定位估计其年龄,在 FG-NET 年龄库上获得了 6.22a 的平均绝对误差,在 MORPH 年龄库上获得了 8.83a 的平均绝对误差^[11]。

(4) 基于深度学习的方法

随着深度学习在图像处理领域取得的巨大成功,2015 年以后年龄估计方法研究在深度学习算法的推动下取得了突破性的进展。Levi 等^[12]使用 19487 张社交照片训练了一个由 3 个卷积层和 2 个全连接层组成的网络结构,其在 0~2 岁、4~6 岁、8~13 岁、15~20 岁、25~32 岁、38~43 岁、48~53 岁、60 岁以上 8 个年龄段获得了 (50.7±5.1)%, 1-off(84.7±2.2)% 的平均准确率。Yi 等^[13]使用不同尺寸的人脸图像训练一个卷积神经网络,获得了 3.63a 的平均绝对误差。在 ChaLearn LAP 2015 表观年龄估计竞赛中,Rothe 等^[14]首先对 20 个 VGG-16 结构使用 ImageNet 数据集进行预训练,然后在 IMDB-WIKI 和 LAP 数据集上进行微调,最后由 20 个 VGG-16 网络的平均值确定最终输出结果,其凭借着 3.221a 的平均绝对误差获得了 ChaLearn LAP 2015 表观年龄估计竞赛的第一名。Liu 等^[15]提出一个名为 AgeNeg 的年龄估计结构,使用 CASIA-WebFace 数据集预训练后,分别在 CACD, Morph-II 以及 WebFaceAge 数据集上进行微调,最终凭借平均绝对误差 3.3345a 获得了 ChaLearn LAP 2015 表观年龄估计竞赛的第二名。Antipov 等^[16]训练 14 个 VGG-16 结构,其

中 3 个 VGG-16 网络只负责识别 0~12 岁儿童,其余 11 个网络用于 0~99 岁的识别,凭借着 ϵ -score 0.2411 的成绩获得了 ChaLearn LAP 2016 表观年龄估计竞赛的第一名。

2 相关工作

2.1 卷积神经网络的基本计算

在构建卷积神经网络时,为了达到良好的实验效果,在卷积神经网络中使用 Relu 激活函数,在卷积层中加入局部响应归一化层(Local Response Normalization, LRN),并在输出层使用 Softmax 分类器。

(1) Relu 激活函数

激活函数的计算公式为:

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (1)$$

因为其近似于生物神经激活函数,Relu 激活函数成为卷积神经网络中最常用的激活函数。其主要优点有:

- 1) 无饱和区域,收敛速度快;
- 2) 计算复杂度小,计算速度快。

(2) 局部响应归一化

局部响应归一化层^[17]是为了提高模型泛化能力,模仿生物神经系统的侧抑制机制,在局部神经元活动中加入竞争机制,防止神经元在学习过程中过饱和。局部响应归一化计算的公式为:

$$b_{x,y}^i = \frac{a_{x,y}^i}{(k + \alpha \sum_{j=\max(0, i-\frac{n}{2})}^{\min(N-1, i+\frac{n}{2})} (a_{x,y}^j)^2)^\beta} \quad (2)$$

其中, N 为卷积核层数, $a_{x,y}^i$ 为第 i 层特征图中 (x, y) 位置的激活值, $b_{x,y}^i$ 为局部响应归一化值,公式中的参数^[17]分别为 $k=2, n=5, \alpha=0.0001, \beta=0.75$ 。

(3) Softmax 分类器

$$s(a_j) = \frac{e^{a_j}}{\sum_{i=0}^{I-1} e^{a_i}} \quad (3)$$

其中, a_j 为最后一层全连接层的输出值, I 为分类的类别个数。Softmax 分类器基于 Softmax 回归模型,是将 logistic 回归模型推广到多分类问题,分类的类别个数不再仅限于 2 个,可以达到多个。

2.2 面部皱纹主观分级评价标准

Lemperle^[18]面部皱纹主观分级评价标准如表 1 所列,其优点是覆盖全面,可使不同区域皮肤皱纹呈现出明显不同的特点。本文利用 Lemperle 面部皱纹主观评价标准对面部皱纹进行准确的刻画。

表 1 Lemperle 面部皱纹的主观评价标准

Table 1 Subjective evaluation criterion of Lemperle facial wrinkles

面部皱纹类型	皱纹级别	特点描述
前额水平纹	0	无皱纹
	1	轻微可见皱纹
眼角皱纹	2	浅纹
	3	中等深度皱纹
面颊褶	4	边界清晰深度皱纹
法令纹	5	多褶皱重度深皱纹

3 动态权值年龄估计模型

本文通过构建年龄估计模型实现年龄估计。首先,为了更加准确地对人脸图像进行区域划分,对人脸图像进行标准化;其次,对人脸面部的不同区域年龄特征进行分析,并在此基础上进行人脸区域划分;最后,使用各区域图像及标准化后的人脸图像训练 4 个卷积神经网络,并在动态权值年龄估计方法的基础上构建动态权值年龄估计模型,从而实现利用人脸图像进行年龄估计的目的。

3.1 人脸图像的标准化

人脸图像在采集过程中会受到多种不可控因素的影响,如不同人脸姿态、不同拍摄距离等,这会在特征提取时产生噪声,进而影响年龄估计精度。人脸图像标准化就是为了减少这些噪声带,以人脸特征点为基础,对人脸图像进行双眼连接线水平矫正、人脸大小统一、人脸图像裁剪。

设采集后的人脸图像像素点集合为:

$$F = \{f_{i,j} | i=0, \dots, M-1; j=0, \dots, N-1\}$$

经标准化后的人脸图像像素点集合为:

$$F_s = \{f_{i,j}^s | i=0, \dots, H-1; j=0, \dots, W-1\}$$

其中, M 和 N 为采集后图像的高和宽, H 和 W 为标准化后图像的高和宽。

人脸图像中每个像素点可被视为二维平面上的一个点,以图像左上角的点为坐标系原点,由坐标 (x_i, y_j) 唯一表示。

利用 ASM 模型确定人脸面部的 78 个特征点 $F_p = \{p_i | i=1, \dots, 78\}$, p_i 的坐标为 (x_{pi}, y_{pi}) 。其中, p_1 到 p_{19} 为面部轮廓点; p_{20} 到 p_{27} 为左眉轮廓点; p_{28} 到 p_{35} 为右眉轮廓点; p_{36} 到 p_{43} 为左眼轮廓点; p_{44} 为左眼瞳孔点; p_{45} 到 p_{52} 为右眼轮廓点; p_{53} 为右眼瞳孔点; p_{54} 到 p_{62} 为鼻轮廓点; p_{63} 为鼻尖点; p_{64} 到 p_{75} 为嘴唇轮廓点; p_{76} 到 p_{78} 为上下唇中间点。

标准化后人脸面部的 78 个特征点为 $F_p^s = \{p_i^s | i=1, \dots, 78\}$ 。

3.1.1 双眼连线水平矫正

脸图像双眼瞳孔连接线与水平线夹角如图 1(a)所示,图像旋转至双眼连接线时处于水平位置。以人脸图像像素点集合 F 与特征点集合 F_p 为基础,矫正双眼连线水平的具体过程如下。

(1)根据左瞳孔点 p_{44} 与右瞳孔点 p_{53} 坐标计算左、右瞳孔连接线的斜率 k :

$$k = \frac{y_{p53} - y_{p44}}{x_{p53} - x_{p44}} \quad (4)$$

(2)当左、右瞳孔连接线斜率 k 非零时,计算左、右瞳孔连接线与水平线之间的夹角 θ :

$$\theta = \arctan(k) * 180/\pi \quad (5)$$

(3)以左眼瞳孔点为旋转中心,通过式(6)给出的变换矩阵得到变换后的人脸图像点集合 F' 及特征点集合 F_p' 。

$$\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \quad (6)$$

$$F' = \{f_{i,j}' | i=0, \dots, M-1; j=0, \dots, N-1\}$$

$$F_p' = \{p_i' | i=1, \dots, 78\}$$

旋转效果如图 1(b)所示。

3.1.2 人脸大小统一化

对人脸图像进行双眼位置水平矫正后,还需要统一人脸图像的大小,为年龄特征区域的划分奠定基础。根据水平矫正后的人脸图像与人脸特征点,以人脸中标准瞳孔点距离与未标准化瞳孔点距离的比值为缩放比例进行缩放变换,以统一人脸大小。以水平矫正后人脸图像像素点集合 F' 与特征点集合 F_p' 为基础,统一人脸大小的具体过程如下。

(1)根据水平矫正后左瞳孔点 p_{44}' 与右瞳孔点 p_{53}' 计算左、右瞳孔间距 D_{new} ,如图 1(c)所示。

$$D_{\text{new}} = x_{p_{53}'} - x_{p_{44}'} \quad (7)$$

(2)根据标准化瞳孔距离 D_{first} 与水平矫正后的瞳孔距离 D_{new} ,计算缩放比例因子 s :

$$s = \frac{D_{\text{first}}}{D_{\text{new}}} \quad (8)$$

(3)根据式(9)给出的变换矩阵得到变换后的人脸图像点集合 F'' 及特征点集合 F_p'' 。

$$\begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} \quad (9)$$

$$F'' = \{f_{i,j}'' | i=0, \dots, M-1; j=0, \dots, N-1\}$$

$$F_p'' = \{p_i'' | i=1, \dots, 78\}$$

3.1.3 人脸图像裁剪

为减小背景对特征提取造成的影响,对人脸部分进行裁剪。根据统一大小后的人脸图像与人脸特征点确定裁剪框左上角点的位置,由高、宽同为 120 像素的矩形框对人脸部分进行裁剪。以统一大小后的人脸图像像素点集合 F'' 与特征点集合 F_p'' 为基础,裁剪人脸图像的具体过程如下。

(1)由人脸轮廓特征点 p_1'' 到 p_{19}'' 中的最左特征点确定矩形框左上角点的横坐标 x 。

$$x = \min(x_{p_i}'') - 5 \quad (10)$$

其中, $i=1, \dots, 19$ 。

(2)由左眼瞳孔特征点 p_{44}'' 确定矩形框左上角点的纵坐标 y 。

$$y = y_{p_{44}''} - 36 \quad (11)$$

(3)使用左上角点坐标 (x, y) , 高、宽同为 120 像素点的矩形框进行人脸裁剪,人脸图像像素点集合 F'' 与特征点集合 F_p'' 经裁剪变换后得到集合 F_s 与 F_p^s , $H=W=120$,如图 1(d)所示。



图 1 人脸图像标准化示意图

Fig. 1 Sketch map of face image standardization

3.2 基于年龄特征的人脸区域划分

3.2.1 年龄特征划分的依据

在日常生活中主要以人头面部为依据来估计人的年龄,这表明人的面部包含了大量的年龄特征。随着年龄的增长,面部脂肪、胶原蛋白和弹力纤维逐渐减少、消失,会在人面部

形成皮肤松弛、皱纹、上/下眼睑下垂、鼻唇沟加深等衰老特征,这些随岁月在人面部留下的印记可作为人脸年龄特征,从而使得年龄估计成为可能。图 2 给出了同一个人在 16 岁、22 岁、29 岁和 43 岁时前额区域皱纹、眼尾皱纹、下眼袋、法令纹、面颊褶的变化情况。



图 2 不同年龄段的面部特征示意图

Fig. 2 Sketch map of facial characteristics with different ages

文献[19-23]以 10 岁为一个年龄段对人头部特征随年龄增长的变化规律进行研究,发现人脸面部变化与年龄增长相关,但每一个面部区域的变化趋势并不统一,因此在估计年

龄时对各面部区域应分开考虑。由于年龄增加导致眼部皮肤松弛,使眼睛外角间的宽度值逐渐减小,上眼皮褶增加,下眼袋变大,鼻宽、面颊皮褶随年龄增长呈现规律变化。眼部、鼻部与面颊都包含大量年龄特征,但由于鼻宽变化会扯动面颊皮肤随之发生变化,因此鼻部、面颊部区域应与眼部区域分开考虑,鼻部、面颊部区域综合称为面中部区域,故眼部区域与面中部区域是年龄估计应考虑的区域;不同年龄段的面部前额纹理可以作为年龄估计的依据,面部前额纹理呈水平状,随着年龄的增加,前额纹理数明显增多,纹理明显加深,因此面部前额区也是年龄估计应考虑的区域之一。为了更准确地刻画各年龄段之间的区别,本文对人脸图像按照 6 个年龄段,根据前额区域、眼部区域、面中部区域的特征进行细化后,归纳出各年龄段的皱纹、下眼袋、上眼皮褶、眼外角间宽、鼻宽等特点,如表 2 所列。

表 2 不同区域年龄的特征

Table 2 Characteristics of different regional ages

区域	前额	眼部				面中部		
年龄段	皱纹	眼尾皱纹	上眼皮褶	眼外角间宽	下眼袋	法令纹	面颊褶	鼻宽
10~19 岁	0	0	无统计	无统计	无	0	0	无统计
20~29 岁	0/1	0/1/2	基础	基础	无	0/1	0/1	基础
30~39 岁	1/2	1/2/3	增加	减小	无/较轻	1/2	0/1/2	增加
40~49 岁	3/4	2/3/4	增加	减小	较轻/下睑松弛 但膨隆感 不明显	2/3	2/3/4	增加
50~59 岁	4/5	3/4/5	增加	减小	轻中度膨隆 伴有明下垂 并伴有皱纹	3/4/5	4/5	增加
60 岁以上	4/5	4/5	增加	减小	中重度膨隆并有 皮肤松弛 眼眶间出现凹陷	4/5	4/5	增加

综上所述,将人脸图像分为前额区域(Forehead Region, FR)、眼部区域(Ocular Region, OR)和面中部区域(Centrale Region of Face, CR)。前额区域具有前额皱纹以及前额皮肤等年龄变化的特征;眼部区域具有眼角皱纹、上眼皮褶、眼外角宽度、下眼袋等年龄变化特征;面中部区域具有法令纹、面颊褶、鼻宽等年龄变化特征。

3.2.2 基于年龄特征的人脸区域划分方法

以集合 F_s 与 F_b 为对象,将人脸图像进行前额区域、眼部区域和面中部区域的划分,结果如图 3 所示。其中,标准化图像的高为 H ,左上角坐标为 $(0,0)$ 。

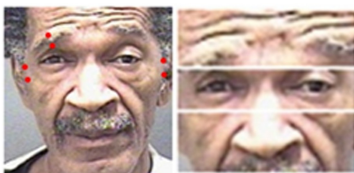


图 3 人脸区域划分示意图

Fig. 3 Sketch map of face region division

(1) 前额区域划分方法

以标准化图像的 $1/6 H$ 作为 FR 的高,右眉上部最右特征点 p_{31}^s 与左眉上部最左特征点 p_{23}^s 作为前额区域的宽,以左眉上部最左特征点 p_{23}^s 为区域的左下角点确定 FR 区域。FR 区域高、宽的计算公式为:

$$\begin{cases} h_{fr} = \frac{1}{6} H \\ w_{fr} = x_{p_{31}}^s - x_{p_{23}}^s \end{cases} \quad (12)$$

其中, h_{fr} 为 FR 的高, w_{fr} 为 FR 的宽,人脸图像高为 H 像素。

FR:以区域左上角点 $(x_{p_{23}}^s, y_{p_{23}}^s - h_{fr})$ 为基准点,以 h_{fr} 为高、 w_{fr} 为宽进行裁剪。

(2) 眼部区域划分方法

以左眉轮廓下部中心点 p_{26}^s 下方 $1/6 H$ 作为 OR 的高,脸部轮廓左第一高点 p_1^s 、右第一高点 p_{19}^s 作为 OR 的宽,以脸部轮廓左第一高点 p_1^s 和左眉下部中心点 p_{26}^s 确定 OR 区域。OR 区域高、宽的计算公式为:

$$\begin{cases} h_{or} = \frac{1}{6} H \\ w_{or} = x_{p_{19}}^s - x_{p_1}^s \end{cases} \quad (13)$$

其中, h_{or} 为 OR 的高, w_{or} 为 OR 的宽,人脸图像高为 H 像素。

OF:以区域左上角点 $(x_{p_1}^s, y_{p_{26}}^s)$ 为基准点,以 h_{or} 为高、 w_{or} 为宽进行裁剪。

(3) 面中部区域划分方法

以脸部轮廓左第二高点 p_2^s 与鼻子下部中心点 p_{58}^s 和嘴唇上部中心点 p_{67}^s 的中心距离作为 CR 的高,脸部轮廓左第二高点 p_2^s 与右第二高点 p_{18}^s 的距离作为 CR 的宽,以脸部轮廓左第二高点 p_2^s 为 CR 的左上角点确定 CR 区域。CR 区域高、

宽的计算公式为:

$$\begin{cases} h_{cr} = y_{p_{28}}^s - y_{p_2}^s + (y_{p_{67}}^s - y_{p_{28}}^s) * 0.5 \\ w_{cr} = x_{p_{19}}^s - x_{p_1}^s \end{cases} \quad (14)$$

其中, h_{cr} 为眼部区域的高, w_{cr} 为前额区域的宽。

CR 区域:以左上角点 $(x_{p_2}^s, y_{p_2}^s)$ 为基准点,以 h_{cr} 为高、 w_{cr} 为宽进行裁剪。

3.3 动态权值年龄估计模型

动态权值年龄估计模型(DW-AEM)由 4 个结构不同的卷积神经网络 ACNNsS, ACNNsF, ACNNsE, ACNNsM 与其对应动态权值 w_S, w_F, w_E, w_M 构成。由于不同子网络结构在不同年龄段的估计能力不一,因此 DW-AEM 模型的输出结果由 4 个网络输出结果与权值综合确定。卷积神经网络 ACNNsS, ACNNsF, ACNNsE, ACNNsM 通过标准化后的人脸图像、前额区域图像、眼部区域图像和面中部区域图像训练获得。动态权值 w_S, w_F, w_E, w_M 由本次估计结果中每个子网络在不同年龄段估计准确率所占 4 个子网络准确率之和确定。DW-AEM 的结构如图 4 所示。

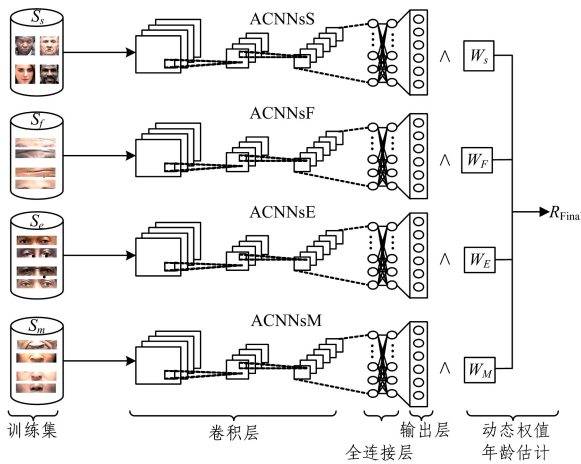


图 4 DW-AEM 结构示意图

Fig. 4 Sketch map of DW-AEM structure

3.3.1 ACNNsS 结构设计

在人脸区域年龄特征学习任务中,使用标准化人脸图像,分别以 0 至 5 作为 10~19 岁、20~29 岁等 6 个年龄段的年龄标签进行训练。ACNNsS 网络结构采用三层卷积层、两层全连接层结构,使用 Relu 激活函数,在全连接层加入 Dropout 层以防止过拟合。网络结构如图 5 所示。

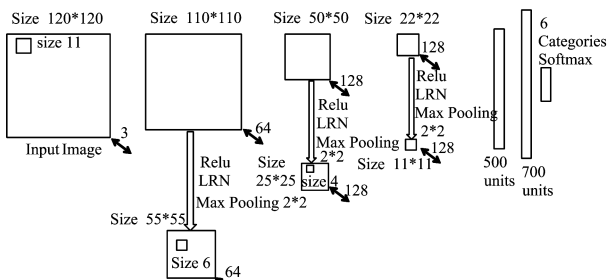


图 5 ACNNsS 结构示意图

Fig. 5 Sketch map of ACNNsS structure

(1) 输入层

将图像统一缩放到 120 * 120 像素大小。

(2) 卷积层

第一层卷积层:使用 64 个大小为 11 * 11 的卷积核,以 1 为步长对输入层图像进行卷积,得到 64 个大小为 110 * 110 的特征图;对第一次卷积后得到的特征图进行激活和局部响应归一化计算,然后计算最大值下采样,得到 128 个大小为 55 * 55 的特征图。

第二层卷积层:使用 128 个大小为 6 * 6 的卷积核,以 1 为步长对第一层卷积层得到的特征图进行卷积,得到 128 个大小为 50 * 50 的特征图;对第二次卷积后得到的特征图进行激活和局部响应归一化计算,然后计算最大值下采样,得到 128 个大小为 25 * 25 的特征图。

第三层卷积层:使用 128 个大小为 4 * 4 的卷积核,以 1 为步长对第二层卷积层得到的特征图进行卷积,得到 128 个大小为 22 * 22 的特征图;对第三次卷积后得到的特征图进行激活和局部响应归一化计算,然后计算最大值下采样,得到 128 个大小为 25 * 25 的特征图。

(3) 全连接层

全连接层与前一层特征图中的每个神经元相连接,两个全连接层分别有 500 个神经元和 700 个神经元。

3.3.2 ACNNsF 结构设计

在前额区域的特征学习过程中,使用前额区域图像,分别以 0 至 5 作为 10~19 岁、20~29 岁等 6 个年龄段的年龄标签进行训练。ACNNsF 网络结构如图 6 所示。输入层图像大小为 72 * 20,由于输入层改变,因此在第一层卷积层中采用 7 * 7 的卷积核,在第二层卷积层中采用 4 * 4 的卷积核。

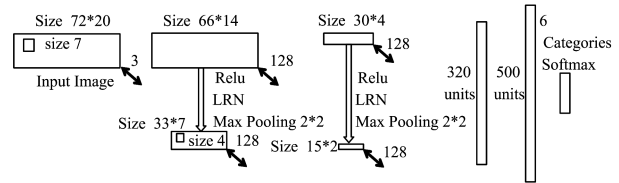


图 6 ACNNsF 结构示意图

Fig. 6 Sketch map of ACNNsF structure

3.3.3 ACNNsE 结构设计

在眼睛区域年龄特征学习任务中,使用眼部区域图像,分别以 0 至 5 作为 10~19 岁、20~29 岁等 6 个年龄段的年龄标签进行训练。ACNNsE 采用与 ACNNsF 相似的网络结构,如图 7 所示,包括两层卷积层和两层全连接层。输入层图像的大小为 100 * 20,前两层卷积层中使用 5 * 5 的卷积核。

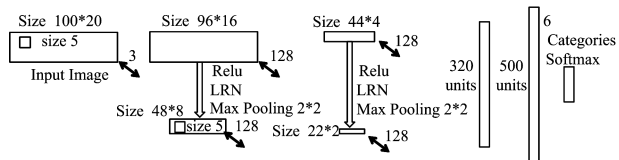


图 7 ACNNsE 结构示意图

Fig. 7 Sketch map of ACNNsE structure

3.3.4 ACNNsM 结构设计

在面中部区域年龄特征学习任务中,使用面中部区域图像,分别以 0 至 5 作为 10~19 岁、20~29 岁等 6 个年龄段的年龄标签进行训练。ACNNsM 网络结构采用与 ACNNsS 相似的结构,如图 8 所示,包含三层卷积层和两层全连接层

的若干张图片,图片大小统一为 200×240 ,共有 52100 张 24 位彩色图片。

4.1.2 年龄估计评价

在基于年龄段分类的性能评价中,常用到年龄估计准确率评价函数^[25],年龄估计准确率为正确估计的图像占总体估计图像数量的百分比,计算公式为:

$$R = \frac{N_r}{N_s} \times 100\% \quad (20)$$

其中, N_r 为正确估计的图像数量, N_s 为测试样本数量,年龄估计准确率值越高,年量估计模型的性能就越好。

4.1.3 硬件环境及软件环境

本实验使用的软件开发环境为:Caffe(Convolutional Architecture for Fast Feature Embedding)深度学习算法框架;Visual Studio2013 开发环境;OpenCV 计算机视觉库;C++ 语言以及 Python 语言。

硬件环境为: Intel(R) Core(TM) i7-4790 的 4 核 8 线程处理器, 3.6 GHz 处理器主频; DDR3 1600 MHz 16 GB 内存; Geforce GTX 1080 主频为 1721-1860 MHz GPU, 10010 MHz 显存频率, 8 GB 显存容量。

4.2 网络结构对分类结果的影响

第一组实验是为了验证网络中不同全连接层数以及网络神经元个数对网络分类结果的影响,使用前额区域年龄特征样本集,在保证网络中每层卷积核个数相同的情况下分别对使用 3 个卷积层和 1 个全连接层的结构网络、3 个卷积层和 2 个全连接层的结构网络进行训练,以对比全连接层数对分类结果的影响;在保证网络结构相同的情况下,使用前额区域年龄特征样本集对 ACNNsF 进行训练,各层分别使用 16-16-64-128 卷积核个数、128-128-320-500 卷积核个数、256-256-500-1000 卷积核个数,根据分类效果对比卷积核个数对分类结果的影响。

表 3 列出了不同网络结构的卷积层的卷积核个数和全连接层神经元个数。通过比较年龄估计准确率发现,在卷积神经网络中训练数据是决定网络结构的主要因素之一,只有网络大小适合数据集大小时才能获得比较好的训练效果。如果网络太小,则会因为没有学习到足够的特征而导致估计效果

不佳;如果网络结构过大,则会因为没有足够的训练集作为支撑而出现训练过拟合,导致年龄估计效果不佳。本文通过多次实验比较,最终确定了适合每个子网络的网络结构与网络大小。

表 3 ACNNsF 子网络的结构与准确率

Table 3 Structure and accuracy of ACNNsF subnetwork

网络名称	第一层卷积核个数	第二层卷积核个数	第一层全连接层神经元个数	第二层全连接层神经元个数	准确率/%
ACNNsF-1	16	16	64	128	54.11
ACNNsF-2	128	128	0	500	69.43
ACNNsF-3	128	128	320	500	71.76
ACNNsF-4	256	256	500	1000	70.75

从表 3 中可以看出, ACNNsF-1 由于第一层和第二层只有 16 个卷积核,因此准确率是 4 个网络中最低的,只有 54.11%;虽然 ACNNsF-3 网络结构不是最大的,但准确率在 4 个网络中最高,为 71.76%; ACNNsF-2 因为少了一层全连接层,可能导致特征映射不足,造成在分类时没有最大程度地利用已经训练得到的特征,从而其准确率比 ACNNsF-3 的准确率低; ACNNsF-4 由于网络结构增大,学习样本没有增加而导致网络过拟合,网络将干扰噪声也作为特征进行学习,在分类时准确率降低到 70.75%。

4.3 子网络年龄估计能力的验证

第二组实验是为了验证子网络 ACNNsS, ACNNsF, ACNNsE, ACNNsM 在每个年龄段的年龄估计准确率。将每个子网络模型测试集样本按年龄段分类,并分别测试每个子模型在各个年龄段的估计准确率。

在人的衰老过程中,不同面部区域在每个年龄段会呈现出不同的特点,导致使用不同面部区域进行年龄估计时在每个年龄段的估计准确率不同。因此,通过计算使用不同面部区域训练的网络在各个年龄段的估计准确率,得到每个子网络在不同年龄段的年龄估计能力,并以此作为在多个子网络年龄估计时计算动态权值的基础。

表 4 列出了 ACNNsS, ACNNsF, ACNNsE, ACNNsM 在每个年龄段的估计准确率。

表 4 每个模型在不同年龄段的估计准确率

Table 4 Estimation accuracy rate of each model in different ages

网络名称		10~19 岁	20~29 岁	30~39 岁	40~49 岁	50~59 岁	60 岁以上
ACNNsS	N_r/N_s	367/573	361/416	341/855	419/653	204/374	66/128
	$R/\%$	86.78	86.77	39.88	64.16	55.35	51.56
ACNNsF	N_r/N_s	100/218	221/289	899/1059	612/907	267/413	104/114
	$R/\%$	45.87	76.47	84.89	67.48	64.65	91.23
CNNsE	N_r/N_s	265/481	268/356	649/773	268/779	149/441	89/170
	$R/\%$	55.09	75.28	83.96	34.40	33.78	52.35
ACNNsM	N_r/N_s	21/383	168/466	299/411	537/677	398/935	62/128
	$R/\%$	5.48	36.05	72.74	79.32	42.57	48.44

通过横向比较,可以得出每个子网络最擅长估计的年龄段。ACNNsS 在 10~19 岁、20~29 岁分别获得了 86.78%, 86.77% 的估计准确率,最擅长通过面部图像进行这两个年龄段的估计。ACNNsF 在 20~29 岁, 30~39 岁, 60 岁以上几个年龄段都获得了较高的准确率,特别是在 60 岁以上年龄段

获得了 91.23% 的准确率。可以看出,利用前额特征可有效区分青年人和老年人,但 10~19 岁与 20~29 岁获得了两个明显不同的准确率,这也说明 29 岁以下青年人之间的前额特征并不存在明显的区分度,前额区域年龄特征很难用于对 29 岁以下的年龄进行估计。ACNNsE 在 30~39 岁年龄段获得

了最高准确率,其最擅长通过眼部区域进行该年龄段的估计,但其在 40~49 岁、50~59 岁的准确率只有 34.4%和33.78%,从而证明单纯利用眼部年龄特征不能很好地对这两个年龄段进行估计。ACNNsM 在 30~39 岁、40~49 岁获得了较高的估计准确率,说明 ACNNsM 最擅长通过面中部区域对这两个年龄段进行估计。

因为表 4 中年龄估计准确率值代表每个子网络在不同年龄段的估计能力,所以在计算动态权值时,每个子网络对应各个年龄段的估计准确率由表 4 中的年龄估计准确率确定。

4.4 不同模型融合方式的对比

分别采用动态权值模型与平均融合模型对 4 个子网络的估计结果进行融合,通过对比实验来验证动态权值模型的年龄估计效果与优势。共进行 5 组实验,每组实验每次在年龄库的各年龄段中随机抽取 100 张图像,分别采用两种不同的融合方式进行年龄估计,并分别计算两种年龄估计模型的估计准确率。

在年龄估计时,由于每个子网络在面部不同区域的估计能力各有所长,因此在多个子网络进行融合时要尽最大可能来扬长避短,故在多模型融合时,利用其在每个年龄段的估计准确性计算动态权值,从而进行多模型融合。

表 5 列出了使用平均融合模型进行年龄估计得到的各年龄段的估计准确率。可以看出,在不同年龄段平均融合模型的年龄估计表现存在很大差异,在 20~29 岁年龄段取得了 87%的最高年龄估计平均准确率,而在 10~19 岁只获得了 23.6%的最低平均准确率,正因为这种极其不稳定的估计表现,导致了平均融合模型只获得了 58.27%的平均估计准确率。

表 5 使用平均融合模型得到的各年龄段的估计准确率

Table 5 Estimation accuracy rate in different ages by using average fusion model

	1st	2st	3st	4st	5st	各年龄段的 平均 准确率/%
10~19 岁	24/100	22/100	25/100	22/100	25/100	23.6
20~29 岁	89/100	88/100	86/100	84/100	88/100	87.0
30~39 岁	86/100	88/100	85/100	82/100	87/100	85.6
40~49 岁	68/100	74/100	67/100	70/100	77/100	71.2
50~59 岁	42/100	46/100	36/100	43/100	42/100	41.8
60 岁以上	41/100	37/100	45/100	42/100	37/100	40.4
平均 准确率/%	58.33	59.17	57.33	57.16	59.33	58.27

表 6 列出了使用 DW-AEM 模型在每个年龄段进行年龄估计得到的各年龄段的估计准确率。从模型估计平均准确率可以看出,与平均融合模型相比,估计准确率得到了很大幅度的提升,5 次实验中 DW-AEM 模型的平均准确率达到 72.6%,提高了 14 个百分点,这说明将多个子网络融合进行年龄估计时,把每个子网络的估计准确率纳入估计结果的判断中可以有效提高估计的准确率。从每个年龄段的平均估计准确率可以看出,与平均融合模型相比,在 10~19 岁,20~29 岁,30~39 岁,60 岁以上 4 个年龄段的估计准确率均有提升,特别是在 10~19 岁、60 岁以上这 2 个年龄段,估计准确率分

别提升了 44 个百分点和 14 个百分点,这也说明了 DW-AEM 模型可以在模型融合时利用每个子网络在各个年龄段的估计准确率来动态计算权值,过滤掉模型中估计能力弱的子网络,充分发挥每个子网络各自的估计能力。但是,DW-AEM 模型在 40~49 岁、50~59 岁两个年龄段的平均估计准确率有所下降,分别下降了 6 个百分点和 1.6 个百分点。通过分析实验结果,估计正确率下降的原因可能是在这两个年龄段中没有估计能力特别强的子网络导致的。从而说明,在年龄估计时,对每个子网络更好的估计准确率是整个模型进行准确估计的基础。

表 6 DW-AEM 模型在各年龄段的估计准确率

Table 6 Estimation accuracy of DW-AEM model in different ages

rate table	1st	2st	3st	4st	5st	各年龄段的 平均 准确率/%
10~19 岁	64/100	68/100	66/100	73/100	69/100	68.0
20~29 岁	97/100	96/100	97/100	96/100	97/100	96.6
30~39 岁	86/100	89/100	88/100	90/100	85/100	87.6
40~49 岁	69/100	70/100	60/100	61/100	66/100	65.2
50~59 岁	39/100	36/100	38/100	47/100	41/100	40.2
60 岁以上	75/100	78/100	83/100	78/100	76/100	78.0
平均 准确率/%	71.67	72.83	72.00	74.17	72.33	72.60

4.5 与其他年龄估计方法的对比

为验证 DW-AEM 模型的有效性,测试数据使用年龄库各年龄段中随机抽取的 100 张图片,将 DW-AEM 模型从 3 个方面与其他经典年龄估计方法进行对比:1)与采用相同年龄库的经典年龄估值算法进行对比;2)与采用同样年龄分组的算法进行对比;3)与采用深度学习的年龄估计方法进行对比。

通过上述实验,DW-AEM 模型获得了 72.6%的平均分类准确率。为了验证本文方法的有效性,将其与其他经典算法进行对比,结果如表 7 所列。

表 7 不同年龄估计方法的比较

Table 7 Comparison of different age estimation methods

方法	准确率/%	分类数
AAM-LFF ^[6]	72.52	4
FIEI-NN ^[5]	61.50	6
CNN-AGE ^[15]	50.70	8
DW-AEM(本文)	72.60	6

实验结果表明,与传统的年龄估计方法相比,本文算法具有一定有效性,在平均估计准确率方面具有明显优势。

结束语 本文通过构建包含 4 个子网络的年龄估计模型,实现了利用人脸图像自动进行年龄估计的目标,并通过实验验证了本文提出的年龄估计模型能达到 72.6%的平均估计准确率。在未来的工作中,可以对非限定条件下的年龄估计与精确年龄估计进行深入研究。

参 考 文 献

- [1] KWON Y H. Age classification from facial images[C]// 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994(CVPR'94). IEEE, 1994:762-767.
- [2] HORNG W B. Classification of age groups based on facial features[J]. 淡江理工学刊, 2001, 4(3): 183-192.

- [3] DEHSHIBI M M, BASTANFARD A. A new algorithm for age recognition from facial images[J]. *Signal Processing*, 2010, 90(8):2431-2444.
- [4] HAYASHI J, YASUMOTO M, ITO H, et al. Age and gender estimation based on wrinkle texture and color of facial images[C]//16th International Conference on Pattern Recognition, 2002. IEEE, 2002:405-408.
- [5] NAKANO M, YASUKATA F, FUKUMI M. Age classification from face images focusing on edge information[C]//Knowledge-based Intelligent Information and Engineering Systems. Springer Berlin/Heidelberg, 2004:898-904.
- [6] TXIA J D, HUANG C L. Age estimation using AAM and local facial features[C]//Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2009 (IIH-MSP'09). IEEE, 2009:885-888.
- [7] GUO G, MU G, FU Y, et al. Human age estimation using bio-inspired features[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2009(CVPR 2009). IEEE, 2009:112-119.
- [8] ZHANG Y, ZHOU Z H. A new age estimation method based on ensemble learning[J]. *Acta Automatica Sinica*, 2008, 34(8):997-1000. (in Chinese)
张宇, 周志华. 基于集成的年龄估计方法[J]. *自动化学报*, 2008, 34(8):997-1000.
- [9] YU Q, DU J X. Age estimation of facial images based on an improved non-negative matrix factorization algorithms[J]. *Journal of Image and Graphics*, 2008, 13(10):1865-1868. (in Chinese)
余庆, 杜吉祥. 基于一种改进 NMF 算法的人脸年龄估计方法[J]. *中国图象图形学报*, 2008, 13(10):1865-1868.
- [10] DU J X, YU Q, ZHAI C M. Age estimation of facial images based on non-negative matrix factorization with sparseness constraints[J]. *Journal of Shandong University (Natural Science)*, 2010, 45(7):65-69. (in Chinese)
杜吉祥, 余庆, 翟传敏. 基于稀疏性约束非负矩阵分解的人脸年龄估计方法[J]. *山东大学学报(理学版)*, 2010, 45(7):65-69.
- [11] 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.
- [12] 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.
- [13] YI D, LEI Z, LI S Z. Age estimation by multi-scale convolutional network[C]//Asian Conference on Computer Vision. Springer International Publishing, 2014:144-158.
- [14] 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, 2015:10-15.
- [15] LIU X, LI S, KAN M, et al. Aenet: Deeply learned regressor and classifier for robust apparent age estimation[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015:16-24.
- [16] ANTIPOV G, BACCOUCHE M, BERRANI S A, et al. Apparent age estimation from face images combining general and children-specialized deep learning models[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016:96-104.
- [17] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems, 2012:1097-1105.
- [18] LEMPERLE G, HOLMES R E, LEMPERLE S S M. A Classification of Facial Wrinkles[J]. *Plastic and Reconstructive Surgery*, 2001, 108(6):1735-1750.
- [19] LI Y L, ZHENG L B, YU K L, et al. Variation of head and facial morphological characteristics with increased age of Han in Southern China[J]. *Chinese Science Bulletin*, 2013, 58(4/5):517-524. (in Chinese)
李咏兰, 郑连斌, 宇克莉, 等. 南方汉族人头部面部形态特征的年龄变化[J]. *科学通报*, 2013, 58(4):336-343.
- [20] SONG X, ZHANG X H, YU K L, et al. Age variations of morphological traits of head and face in Qiang nationality in Sichuan[J]. *Journal of Tianjin Normal University (Natural Science Edition)*, 2016, 36(2):69-74. (in Chinese)
宋雪, 张兴华, 宇克莉, 等. 四川羌族头面部特征的年龄变化[J]. *天津师范大学学报(自然科学版)*, 2016, 36(2):69-74.
- [21] LI Y L, LU S H, ZHENG L B, et al. Age variation of head and facial morphology in Han nationality in Jiangxi province[J]. *Acta Anthropologica Sinica*, 2012, 31(2):193-201. (in Chinese)
李咏兰, 陆舜华, 郑连斌, 等. 江西汉族人头部面部形态特征的年龄变化[J]. *人类学学报*, 2012, 31(2):193-201.
- [22] 李咏兰, 郑连斌, 宇克莉, 等. 中国北方汉族人头部面部形态特征的年龄变化[C]//中国解剖学会 2015 年年会论文文摘汇编, 2015.
- [23] ZHENG L B, ZHANG X H, HU Y, et al. Variation of Morphological Traits in Head-facial Values of Han in Qionglai of Sichuan Province[J]. *Journal of Sun Yat-sen University (Medical Sciences)*, 2011, 32(6):729-734. (in Chinese)
郑连斌, 张兴华, 胡莹, 等. 四川邛崃汉族头面部形态特征的年龄变化[J]. *中山大学学报(医学科学版)*, 2011, 32(6):729-734.
- [24] RICANEK K, TESAFAYE T. Morph: A longitudinal image database of normal adult age-progression[C]//7th International Conference on Automatic Face and Gesture Recognition, 2006 (FGR 2006). IEEE, 2006:341-345.
- [25] WANG X M, LIANG L Y, WANG Z L, et al. Age estimation by facial image: a survey[J]. *Journal of Image and Graphics*, 2012, 17(6):603-618. (in Chinese)
王先梅, 梁玲燕, 王志良, 等. 人脸图像的年龄估计技术研究[J]. *中国图象图形学报*, 2012, 17(6):603-618.