DOI: 10.7641/CTA.2017.70072

融合人脸五官信息的深度年龄估计

李云飞[1,2†], 卢朝阳[1], 李　静[1]

(1. 西安电子科技大学 通信工程学院, 陕西 西安 710071; 2. 渭南师范学院 网络安全与信息化学院, 陕西 渭南 714099)

摘要: 本文提出了一种新型的基于人脸五官辅助的深度年龄估计方法, 将传统的人脸五官区域特征提取加分类器设计方法与基于深层卷积神经网络(convolutional neural network, CNN)的端到端分类方法进行融合来解决年龄估计问题, 增强了系统模型的泛化能力. 该方法将面部关键点生成的局部对齐的人脸图像块作为CNN的输入, 直接从图像的像素点评估年龄, 采用多尺度分析网络结构极大地提高了性能, 同时又利用传统算法增强了五官区域的信息. 最后通过在MORPH AlbumⅡ上的实验表明文中提出方法比其他同类研究方法更加优秀.

关键词: 年龄估计; 五官辅助; 卷积神经网络; 多尺度; 多任务

中图分类号: TP273　　**文献标识码**: A

Deep age estimation by fusing facial information

LI Yun-fei[1,2†], LU Zhao-yang[1], LI Jing[1]

(1. School of Telecommunications Engineering, Xidian University, Xi'an Shaanxi 710071, China;
2. School of Network Security and Information, Weinan Normal University, Weinan Shaanxi 714099, China)

Abstract: The paper presents a new mode of solution for deep age estimation by facial features auxiliary, which fuses the traditional facial information with the convolutional neural network (CNN) to achieve the age estimation, in order to reinforce the generalization ability of system model. The solution estimates age from image pixels directly, which makes the locally aligned face image block generated by the key points of the face as the input of the CNN. The system improves the performance significantly by using the multi-scale CNN network structure. At the same time, it apply the traditional method to strengthen the information of facial areas. The experiments on MORPH AlbumⅡ illustrate the superiorities of the proposed method over other state-of-the-art methods.

Key words: age estimation; facial features auxiliary; convolutional neural network; multi-scale; multi-task

1　引言(Introduction)

通过人脸图像进行年龄估计方面的工作最早在Kwon和Lobo1994年发表的论文[1]中描述, 但是他们只是简单的将年龄分为几个范围来进行相关研究. 与其他的人脸分析技术相似, 年龄估计受许多内在和外在因素的影响, 是一个复杂的问题, 要寻找一个稳定准确的函数来将图像像素和其相应的年龄信息关联起来是很困难的. 由于受到面部分析技术的影响, 早期的方法主要利用几何特征来判断人脸图像的年龄范围, 常用的几何特征包括下巴位置、鼻子位置等[1−2]. 随着分类准确性的提高, 研究者们逐渐开始估测准确的年龄, 而不仅仅是估计年龄的大致范围; 同时, AAM算法[3]的出现使得构建人脸图像的形状纹理模型变得容易, 后续许多新的方法都受到了该算法的启发, 例如AAM + quadratic estimator[4], aging pattern subspace(AGES)[5]等. 2000年后, 在新的模型(与面部分析相关的特征及分类器)的推动下, 年龄估计的性能水平不断提高. 随着性能的改进, 年龄估计衍生出许多新的应用, 比如人口统计分析、商业用户管理、视频安全监控等.

大多数现有的基于人脸图像的年龄估计方法分为两个步骤: 局部特征提取和回归分析. 最近, 许多研究者重点关注特征提取后的回归过程, 如支持向量回归(support vector regression, SVR)[6]、最小二乘法(partial least squares, PLS)[7]、典型相关分析(canonical correlation analysis, CCA)[8]等, 回归分析方法经常被用于估计人脸图像的年龄. 局部特征提取是指获取一个相关因素稳定的表达方式, 包括个体、性别种族、表情、姿态和光照等, 局部特征的维度通常可以通过特征选择或者下采样来降低. 以局部特征为基础,

收稿日期: 2017−02−10; 录用日期: 2017−07−21.

†通信作者. wnlff@126.com; Tel.: +86 13571381191.

国家自然科学基金项目(61502364), 渭南师范学院科研基金项目(16YKS001)资助.

Supported by National Natural Science Foundation of China (61502364) and Scientific Research Foundation of Weinan Normal University (16YKS 001).

研究者主要的关注点在通过分类或者回归分析来进行年龄估计, 从提取到的特征中预测年龄的范围或者准确的年龄. 而在分类的方法中, SVM和SVR是最常用和有效的. 通过使用BIF+SVM, Guo等人[9]在YGA数据库[10]上的平均绝对误差(mean absolute error, MAE)可以达到3.47年(男性)和3.91年(女性), 并且通过使用BIF+SVR, 可以在FG–NET上达到4.77年. Cao等人[11]将年龄估计构造成了一个排序问题, 并且提出了一个基于Rank-SVM[12]的新方法, 在MORHP Album Ⅱ 的子集上, 取得了 MAE = 5.12 年的优秀性能. 最近, 主流的年龄估计方法多采用回归分析, 例如线性回归分析[10], SVR[6], PLS[13]和CCA[14]等. 通过结合AGES和LDA, Geng等人[5]在FG–NET上的MAE达到了6.22年. 然而, AAM是一个基于像素的方法, 会导致基于AAM的方法对环境变化的稳定性不足. 在2007年之后, 局部特征逐渐成为此领域的主流, 比如Gabor[15], LBP[16], SFP(spatially flexible patch)[17]和BIF[10]等.

在现有的年龄估计方法中, 最具有代表性的工作是BIF+CCA[14], 经过细致的参数调整, 该方法在MORPH数据库[18]上获得了3.98年的平均绝对误差(MAE), 达到了很高的性能, 在实际应用中拥有比较好的效果, 但该方法仍然有进一步提升性能的空间. BIF+CCA方法包括3个步骤: Gabor滤波[19]、Max+Std池化和CCA, 可以将其看作一个3层的网络, 由卷积层、池化层和全连接层组成. 深度神经网络也被用来解决年龄估计[20]和性别分类[21]问题, 但是它的作用并未完全发挥.

近年来, 卷积神经网络在计算机视觉多个领域取得巨大成功, 一些研究者也开始使用卷积神经网络(convolutional neural network, CNN)来解决年龄估计问题. Yang等人[20]使用CNN来进行监控场景下的年龄估计, 但是他们的工作重点是人脸跟踪, 而且只使用了最简单的CNN 模型, 并未进行任何的改进, 因此准确性比[9]低. Wang等人[22]使用CNN 进行特征学习, 但是该方法将学习到的特征输入到其他的分类器中来估计最终的结果. Rothe等人[23]提出了一个名为DEX(deep expectation) 的方法, 该方法的模型基于VGG–16, 并且取得很好的性能. 然而该方法需要大量数据来进行预训练.

本文提出了一个新型的基于人脸五官辅助的深度年龄估计方法, 将人脸五官区域提取的传统特征与CNN学习得到的特征进行融合使用. 为了提升年龄估计方法的准确性和普适性, 在充分挖掘CNN的潜力的同时, 把传统的面部分析方法融入到了其中. 与同类研究方法比较, 文中提出的思想具有以下优势: 第一, 使用面部的关键点生成局部对齐的人脸图像块, 把它们作为CNN的输入, 提高所选方法对图像形变和姿态差异的鲁棒性; 第二, 将人脸图像裁剪成多个多尺度的小块用回归分析算法进行联合分析, 由于不同的尺度和位置存在着信息的互补, 所以利用多尺度深度网络结构可以显著地提高CNN的性能; 第三, 利用人脸的对称性增加数据库的数据量, 从而提高CNN的性能; 第四, 采用关键点定位以及人脸先验知识提取五官区域的图像, 包括眼睛、鼻子和嘴巴, 并提取此区域的特征, 此类特征的性能不受训练数据影响, 利用该类特征辅助年龄估计能提升本文方法的泛化能力. 最后, 通过训练得到一个多任务的CNN模型, 并融合五官特征得到最终的模型, 该模型拥有很好的适应能力, 可以高速准确地估计年龄、性别和种族.

本文采用的CNN考虑了传统算法的3个问题: 多尺度分析网络、局部对齐的人脸图像块和面部对称性. 与本文提出的深度网络相比较, 一个相似的多尺度CNN用在了解决场景标定任务中[24], 但其中的多尺度分析仅用在了测试阶段, 而本文既用在了测试阶段, 也用在了训练阶段. 局部对齐图像块在很多解决无限制人脸识别问题方法中[25]都取得了成功, 人脸对称性也被广泛关注, 用来解决人脸姿态问题[26]或者增加训练数据量[27].

由于CNN的复杂性比传统方法要高, 因此为得到一个性能优秀的网络, 减小CNN的过拟合问题, 需要使用大量训练数据. 在现有的年龄数据库中, MORPH Album Ⅱ 数据库的规模最大, 包含55132张人脸. 本文选用MORPH Album Ⅱ 作为实验数据库, 在该数据库上达到了目前比较好的结果, MAE为3.50年. 同时, 由于拥有更大的核矩阵, CNN的速度比BIF+CCA快.

2　多尺度卷积网络(Multi-scale convolutional network)

本文采用的卷积神经网络的结构如图1所示, 它包含了许多子网络.

2.1　局部对齐人脸图像块(Local aligned face patches)

人脸关键点检测对于构建一个好的人脸识别算法非常重要, 尤其在无约束条件下, 以准确的面部关键点为基础, 可以对面部图像进行姿态矫正或者构建一个对姿势鲁棒的面部描述子. 最简单高效的方法是使用局部对齐策略, 通过在每个关键点周围剪裁小块图像, 可以获得一些在局部坐标中对齐的小块. 对于不同的人脸图像, 这些小块拥有相同的高层语义特征, 它比完整的人脸图像更适宜用于训练.

由于人脸关键点检测技术在人脸识别问题中的成功, 可以把面部图像裁剪成许多局部对齐的小块, 并把它们作为本文算法的输入. 对于一张人脸图像, 首先使用SDM[28]定位49个关键点, 检测到的关键点及其编号如图2所示. 为了便于后续的训练步骤, 将这些关键点按照人脸的对称性分组, 即关于人脸中线对称的点分为一组, 而位于中线上的关键点则各自成对.

由于颜色信息不稳定并且对于年龄估计帮助不大,

所以本文和其他工作保持一致, 将彩色图像转化为灰度图. 所有图像按照关键点11和35间的距离被归一化为60, 42, 30, 22像素这样4个尺度, 其他的关键点也随着图像的归一化而进行改变. 在各个尺度的归一化图像中, 以关键点为中心裁剪48 × 48的小图像块. 对于右半边脸的关键点对应的图像块, 通过镜像从而和左半边脸保持一致, 具体的分组情况如图2所示. 这样对于每张人脸图像, 就可以得到27 × 2 = 54个局部对齐的多尺度小块, 这些小块拥有人脸图像的多尺度特征, 并且对于刚性和柔性的扭曲具有鲁棒性.

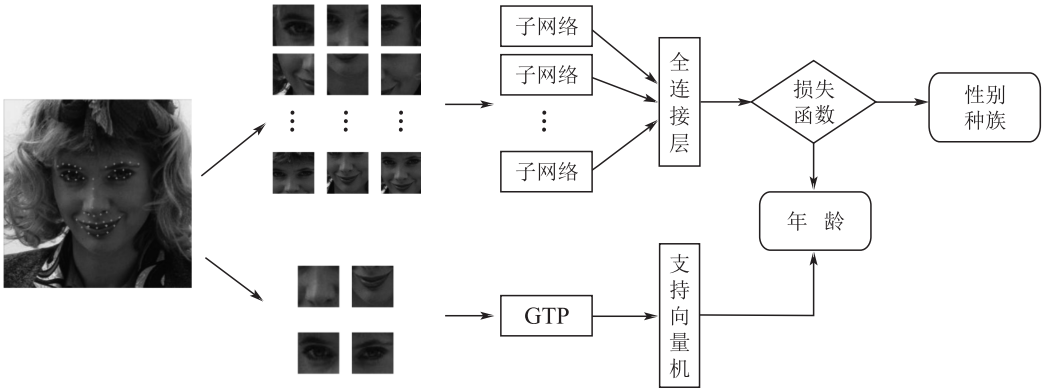

图 1 本文方法的总体架构

Fig. 1 The structure of the proposed network

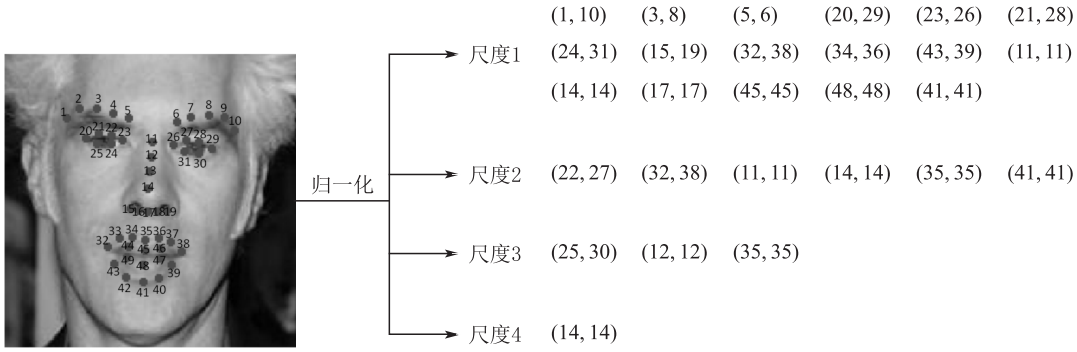

图 2 根据关键点裁剪出的27 × 2 = 54个图像块(分辨率均为48 × 48, 右边面部区域的图像块做镜像操作以增加数据量)

Fig. 2 27 × 2 = 54 patches cropped from a face image based on the landmarks(the resolution is 48 × 48, the patches from the right half of face are mirrored to argument the data)

2.2 用于年龄估计的卷积神经网络(Convolutional network for age estimation)

图1展示了本文提出的网络结构, 每一层的细节在图3中描述.

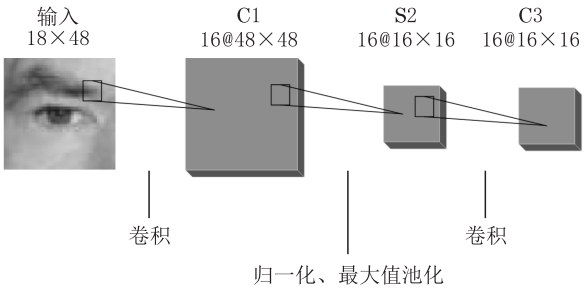

图 3 本方法子网络结构(输入为人脸图像块, 输出作为图1中的全连接层输入)

Fig. 3 The structure of sub-network for each patch (the input of sub-network are face patches and the output are sent to the F4 layer in Fig. 1)

文中按图3创建了27个子网络, 分别用来处理27对图片小块, 并且在最后的连接层融合它们的响应. 这个结构有两个好处: 1) 27组子网络中, 每一个可以只学习特定的特征; 2) 最后的连接层把所有的子网络融合在了一起, 可以使它们互补. 27个子网络的变量和图1 、图3中的最终子层, 在整个处理过程中都是优化过的.

每一个小块的子网络都包含一个基础网络, 该网络由一个卷积层、一个最大值池化层和一个局部连接层. 卷积层、池化层和局部连接层的通道数均为16. 在卷积之前, 用0值填充输入图像的周边, 以确保输出和输入拥有相同的大小. C1层的滤波器的尺寸为7 × 7, C3层的尺寸为3 × 3. 使用ReLU[29]作为C1和C3层的激活函数. S2层的步长为3, 所以经过卷积后图像的尺寸从48 × 48降低到16 × 16, 并且S2 层包含了跨通道归一化单元. 而 C3 层的输出是

$16 \times 16 \times 16 = 4096$维特征向量, 所以图1中全连接层的输入为$4096 \times 27 = 110592$维. 全连接层采用平方差来计算损失函数, 可以视为线性回归层. 在实际应用中, 应该注意全连接层的输出和目标任务的数量级. 一般来说, 需要引入一个尺度因子来确保它们处于同一个数量级. 本文的深度网络采用随机梯度下降(stochastic gradient descent, SGD)进行优化.

除了上述基础网络, 本文在后续实验中还会使用在此基础上附加局部连接层和全连接层的网络结构进行实验. 因为卷积适用于获取整张图片的统计特征, 但是人脸图像的统计特性在图像的不同位置是不同的(对于年龄估计而言, 不同的区域对于年龄的分类作用可能也是不同的), 局部连接层可以更有针对性的对这种差异性进行描述, 之后的实验结果将会验证本文的想法.

2.3　多任务学习(Multi-task learning)

年龄、性别和种族是人的3个重要的特征. 在从人脸图像中估计上述特征的时候, 这3个特征并非是相互独立的. Guo等人[30] 首先从人脸照片中估计性别和种族, 然后把它们送入基于性别和种族的年龄估计器, 获得了很好的效果. 其他一些方法[13–14]使用PLS+CCA来同时检测上述的3个特征, 并获得了比之前的方法更好的效果. 联合地估计性别、年龄和种族有以下优点: 1) 通过共享3个任务的模型来提高学习和计算的速度; 2) 多标签可以为数据库提供更多的信息, 以便充分利用多个任务之间的相关性; 3) 当某个任务训练数据不充足时, 相比单个任务能取得更好的性能.

具体到本文采用的网络, 把损失函数扩展为一个多任务的函数来联合估计性别、年龄和种族信息. 图1全连接层的输出从1维调整成了3维, 分别与年龄、性别和种族相对应. 本文的多任务损失函数由3个部分组成: 年龄均方误差、性别和种族交叉熵损失[31]. 由于MORPH Album Ⅱ数据库中的大部分图片都是黑人和白人的(96%), 所以本文的种族分类只涉及黑人和白人. 损失函数的表达式如下:

$$\begin{aligned} J_1 = {} & (C(X,W)_{\text{age}} - L_{\text{age}})^2 + \\ & \alpha \ln(\mathrm{e}^{-2C(X,W)_{\text{gender}} L_{\text{gender}}} + 1) + \\ & \beta \ln(\mathrm{e}^{-2C(X,W)_{\text{ethnicity}} L_{\text{ethnicity}}} + 1), \end{aligned} \quad (1)$$

其中: $C(X,W)$表示网络的函数; X是输入人脸图像; W是网络的参数; 下标分别表示了该网络的3个输出年龄、性别和种族; L是训练集中的三维标签; $L_{\text{gender}} \in (-1,1)$, –1代表男性, 1代表女性; $L_{\text{ethnicity}} \in (-1,1)$, –1代表黑人, 1代表白人; α和β是用来调节每项重要性的超参数.

因为式(1)是可导的, 所以可以采用用SGD对目标进行优化. 如果需要处理其他数据库的多分类问题, 可以使用一个softmax回归和一个负的对数似然作为损失函数.

2.4　五官辅助(Facial features auxiliary)

由于年龄估计领域的数据库相比于其他计算机视觉任务规模较小, 单纯地通过数据驱动的方法训练得到的模型易受训练数据的影响而无法具有较好的泛化能力. 一旦测试库的样本分布与训练库差别较大, 性能则无法保证. 为了提升本文方法的性能, 在使用卷积神经网络算法的同时, 将用于传统分析的面部五官区域作为辅助特征与其进行融合, 这种传统特征不受训练数据的影响, 可以提升本文方法在不同应用场景下的适应能力. 图4展示了五官辅助的具体流程, 通过关键点定位提取五官区域的图像, 包括眼睛、鼻子和嘴巴, 并在此4个区域提取Gabor三元模式(gabor ternary pattern, GTP)特征.

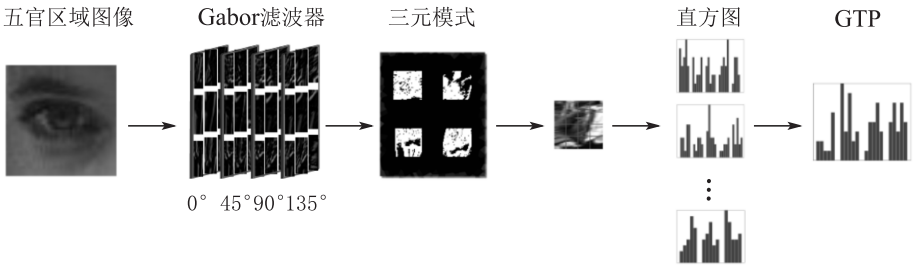

图 4　提取五官的GTP特征流程图(图中仅以提取眼部区域的GTP特征为例)

Fig. 4　The flowchart of exacting the GTP features (only taking the eye area for example)

以右眼区域为例说明: 通过图2中的关键点对(21, 24)得到一个48×48的区域. 由于本文处理的区域相对较小, 所以采用单一尺度和4个方向的Gabor核[32]. 待计算出每个区域的GTP后, 将48×48的区域均分为$4 \times 4 = 16$个子区域, 每个子区域的大小为12×12, 其余的参数保持与文献[33]的相同. 每个

方向都有3个模式, 共有$3^4 = 81$个模式, 这样将每个子区域的直方图连接可形成一个1296维($4 \times 4 \times 81$)的特征向量. 不同于文献[33], 本文在得到1296维的特征向量后, 不再使用PCA进行降维. 特征得到后, 本文保持与文献[9]相同的做法, 采用线性SVM做年龄分类和SVR做年龄回归, 最终在分数层融合两类方法的结果, 五官辅助的比重为30%.

3 实验(Experiment)

本文选择规模较大的MORPH Album Ⅱ作为训练数据库, 然后通过实验与其他先进方法做了对比, 结果表明了多尺度分析和局部对齐的优点, 以及本文所提出模型的优越性和多任务网络的高效性.

3.1 数据库及训练集建立(Database and setup)

MORPH Album Ⅱ包括55132张人脸照片和超过13000个目标, 照片的拍摄时间从2003年到2007年, 年龄从16岁到77岁. 虽然这是一个规模很大的数据库, 但其中性别和人种的分布是不平衡的, 男女比例为5.5:1, 黑白人比例为4:1, 除了黑人和白人, 其他种族所占的比例很低(约4%).

为了有效利用数据库, 本文采用与文献[30]相同的方法来处理数据库, 并把它随机地划分为3个不重叠的子集S1, S2和S3. 首先, 在MORPH中的所有图片都经过预处理, 因为MORPH包含一些无人脸图片(比如纹身), 它们在经过这一步后被移出数据库, 然后使用SDM方法对人脸图像进行面部关键点检测, 局部对齐的图像块也按照第3.1节中所描述的那样被裁剪. 按照以下两个准则构建S1, S2, S33个子集: 1) 使男女比例约为3; 2) 白人黑人比例等于1, 子集的构成情况如表1所示. 在所有的实验中, 重复如下步骤: 1) 由S1训练, S2+S3测试; 2) 由S2训练, S1+S3测试. 对于年龄估计和性别检测, 表1中所有的图像都用到了. 对于种族分类, 忽略“其他”类中的图像.

表 1 处理后的MORPH Album 2子集构成情况

Table 1 The information of the pre-processed MORPH Album 2 and S1, S2, S3 subsets

分类	男 性			女 性		
黑人	S1: 4012	S2: 4012	S3: 28835	S1: 1305	S2: 1305	S3: 3166
白人	S1: 4012	S2: 4012	S3: 0	S1: 1305	S2: 1305	S3: 0
其他	S3: 1845			S3: 130		

3.2 年龄估计(Age estimation)

如第2.1节中所描述, 根据人脸的关键点, 每张人脸图像生成$27 \times 2 = 54$个多尺度图像块. 把非镜像化的小块称为左边小块, 镜像化的小块称为右边小块. 在训练阶段, 左边小块和右边小块可以被视为增加训练集规模的方法. 在测试阶段, 可以分别对左边小块和右边小块进行预测, 根据预测结果取平均值.

3.2.1 结构(Structure)

卷积神经网络的结构决定着模型的性能, 而且要同时考虑训练集规模, 所以如何设计一个好的结构是至关重要的. 本文比较了五种不同结构的模型. 结构的对比如下:

1) C–P–F: 卷积+最大值池化+全链接;

2) C–P–C–F: 卷积+最大值池化+卷积+全链接;

3) C–P–L–F: 卷积+最大值池化+局部层(有着不共享的权重局部连接层)+全链接;

4) C–P–L–F–S(本文方法): 卷积+最大值池化+局部层+全链接+五官辅助;

5) C–P–C–P–L–F: 卷积+最大值池化+卷积+最大值池化+局部层+全链接.

C意为convolution, P意为max pooling, F意为full connection, L 意为 local layer, S 意为structure of facial features auxiliary.

“C–P–L–F–S”是本文最终采用的结构. 在所有的网络中, 所用的滤波器的数量都是16, 训练时候 epoch 的数量为30, 在前 20 个epoch时, 学习率为0.01, 后10个epoch学习率为0.001. 在训练之前, 所有的图像都会减去训练集图片的像素平均值.

在保证相同建库规则的情况下, 随机选择建库样本, 进行10次重复实验并取平均, 上述的5种结构的性能如表2所示.

表 2 5种模型结构的性能对比

Table 2 Comparison of 5 networks with different architectures

结构	训练集	测试集	年龄错误率	平均错误率
C–P–F	S1	S2+S3	3.69	3.67
	S2	S1+S3	3.64	
C–P–C–F	S1	S2+S3	3.95	3.91
	S2	S1+S3	3.87	
C–P–L–F	S1	S2+S3	3.58	3.56
	S2	S1+S3	3.54	
C–P–L–F-S (本文方法)	**S1**	**S2+S3**	**3.46**	**3.50**
	S2	**S1+S3**	**3.53**	
C-P-C–P–L–F	S1	S2+S3	3.76	3.71
	S2	S1+S3	3.66	

从表2中可以看到本文所提出的“C–P–L–F–S”有比其他结构更低的错误率, 因此在接下来的实

验中选择该结构. 从图表中还能看出, 在全链接层之前使用一个局部连接层能有效提升性能. 例如“C–P–L–F”比“C–P–F”性能优异, “C–P–C–P–L–F”比“C–P–C–F”优异, 这说明了局部连接层的重要性, 从而证实本文前面章节的论述. 而“C–P–C–P–L–F”的网络更深, 反而比浅网络结构性能差, 原因可能是更深的结构需要更大规模的训练数据.

3.2.2 多尺度分析(Multi-scale analysis)

如图2所示, 切割后的图像块有4种尺度. 每一种尺度的小块的数量分别是17, 6, 3和1. 为了说明多尺度分析的作用, 对每一种尺度的性能进行了评估. 除了子集的数量, 每个尺度的结构和多尺度的实验配置相同. 各个尺度的实验结果如表3 所示. 从实验结果发现小尺度通常都比大尺度的性能要好. 原因可能有如下3个: 1) 小尺度图像块包含更多的纹理信息, 它们可能和人的年龄紧密相关; 2) 小尺度图像块比大尺度图像块更容易对齐; 3) 小尺度图像块的数量更多. 然而采用多个尺度则能够明显提升性能, 主要由于不同尺度的图像训练出的模型有一定的互补性, 而且更大尺度的图像可以更好的描述图像的结构信息.

表 3 各个尺度的性能对比

Table 3 Comparison of the performance with different scales

尺度	数量	训练集	测试集	年龄错误率	平均错误率
尺度1	17	S1	S2+S3	3.80	3.82
		S2	S1+S3	3.84	
尺度2	6	S1	S2+S3	4.36	4.40
		S2	S1+S3	4.45	
尺度3	3	S1	S2+S3	4.45	4.35
		S2	S1+S3	4.25	
尺度4	1	S1	S2+S3	5.74	5.62
		S2	S1+S3	5.50	
多尺度 (本文方法)	27	**S1**	**S2+S3**	**3.46**	**3.50**
		S2	**S1+S3**	**3.53**	

虽然不同的尺度有不同的性能, 但它们彼此之间是互补的. 当融合 4 种尺度的信息之后, MAE 从3.82年降低到了3.50年.

3.2.3 局部对齐分析(Locally alignment analysis)

除了多尺度分析, 局部对齐是提升性能的另一个重要因素. 为了公平起见, 本文基于平均形状将人脸图像剪裁成小的图像块. 由于平均形状对于每一个人脸图像来说并不准确, 所以剪裁得到的图像块不能很好地对齐, 可以用这些非对齐的图像块训练网络用于比较. 本文所采用的对齐网络和非对齐网络拥有相同的结构, 它们的输入都是$27\times48\times48$维.

表4展示了对齐网络和非对齐网络的性能对比. 显然, 对齐网络的MAE比非对齐网络的低. 由于非对齐的图像块有更多的变化, 非对齐网络在训练过程中会更多地去学习一些不变的特征, 从而降低年龄估计的性能; 相反, 对齐网络的输入图片都是对齐的, 因此网络可以更多地关注年龄估计. 另外, 非对齐网络更加难以收敛, 在本文的实验中, 非对齐网络需要超过30轮训练才能收敛.

表 4 对齐网络和非对齐网络的性能比较

Table 4 The performance of two networks trained on aligned and non-aligned patches

方法	训练集	测试集	年龄错误率	平均错误率
非对齐	S1	S2+S3	3.92	3.89
	S2	S1+S3	3.86	
对齐	S1	S2+S3	3.46	3.50
	S2	S1+S3	3.53	

与一些现有方法的总体性能进行比较, 同样选择MORPH Album Ⅱ数据库, 如表5所示. 从表中的最后一列可以看出本文采用的融合五官信息的深度年龄估计方法将MAE显著地减小到了3.50年. 通过比较表 5 和表 3 可以看到, 即使是只使用单一尺度(表3中的尺度1), 本文的MAE = 3.82, 仍然有着不错的竞争力.

3.3 年龄、性别、种族联合估计(Joint estimation of age, gender and ethnicity)

最终, 本文按照第2节中所描述的方法, 训练得到了一个可以同时从人脸图像中估计年龄、性别和种族的多任务模型. 训练过程与上文实验相似, 结构采用“C–P–L–F–S”, 一共训练30轮, 使用随机梯度下降方法进行优化. 式(1)中的α和β对多任务网络的影响不大, 所以实验中都设置为1. 多任务网络的性能如表5所示. 当联合性别和种族信息进行优化之后, 平均错误率和单任务网络的相同, 都是3.50年. 同时, 性别和种族分类的准确率也优于目前同类方法的研究结果.

与本文提出的方法相比较, BIF+KCCA 和 BIF+KPLS 的测试速度都很慢. 在 Intel 酷睿 2 CPU 2.1 GHz 的处理器上进行特征提取时间的统计, KCCA和KPLS的测试时间是72515 s和72516 s, 本文的测试时间是12916 s, 比KCCA和KPLS快.

表 5 本文方法与先进方法的比较

Table 5 Comparison of the proposed method to state-of-the-art methods

方法	训练集	测试集	性别准确率/%	种族准确率/%	年龄错误率	平均错误率
BIF+CCA[14]	S1	S2+S3	95.2	97.8	5.39	5.37
	S2	S1+S3	95.2	97.8	5.35	
BIF+rCCA[14]	S1	S2+S3	97.6	98.7	4.43	4.42
	S2	S1+S3	97.6	98.6	4.40	
BIF+kCCA[14]	S1	S2+S3	98.5	98.9	4.40	3.98
	S2	S1+S3	98.4	99.0	3.95	
BIF+PLS[13]	S1	S2+S3	97.4	98.7	4.58	4.56
	S2	S1+S3	97.3	98.6	4.54	
BIF+KPLS[13]	S1	S2+S3	98.4	99.0	4.07	4.04
	S2	S1+S3	98.3	99.0	4.01	
BIF+LSVM[14]	S1	S2+S3	—	—	5.06	5.09
	S2	S1+S3	—	—	5.12	
BIF+KSVM[14]	S1	S2+S3	—	—	4.89	4.91
	S2	S1+S3	—	—	4.92	
CNN[20]	S1	S2+S3	—	—	4.64	4.60
	S2	S1+S3	—	—	4.55	
CNN+LSVR[22]	S1	S2+S3	—	—	4.69	4.72
	S2	S1+S3	—	—	4.75	
	S1	**S2+S3**	**98.2**	**99.1**	**3.46**	**3.50**
本文方法	**S2**	**S1+S3**	**97.9**	**98.4**	**3.53**	

4 总结(Conclusions)

本文中提出了一种融合人脸五官信息的深度年龄估计方法. 该方法具有更深的结构和可以学习的参数, 并且特别强调了人脸五官区域特征在年龄估计任务中的重要性, 显著地降低了错误率. 为了能够使CNN在年龄估计中发挥其高效性能, 设计了一个适合年龄估计任务的网络结构, 包括“C–P–L–F–S”结构、多尺度分析和局部对齐图像块, 实验证明, 该结构与同类研究方法比较, 性能和速度都有明显的提升. 同时, 本文为年龄、性别和种族特征共同构建了一个新颖的损失函数, 通过训练一个多任务网络, 在解决年龄估计问题的同时, 在性别和种族分类任务上同时达到高准确率. 未来的工作将会关注如何设计一个更加合理的针对年龄、性别和种族的联合多任务网络.

参考文献(References):

[1] KWON Y, VITORIA LOBO N. Age classification from facial images [C] //*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, 1994: 762 – 767.

[2] KWON Y, VITORIA LOBO N. Age classification from facial images [J]. *Computer Vision & Image Understanding*, 1999, 74(1): 1 – 21.

[3] COOTES T, EDWARDS G, TAYLOR C. Active appearance models [C] //*European Conference on Computer Vision*. Berlin: Springer, 1998: 484 – 498.

[4] LANITIS A, DRAGANOVA C, CHRISTODOULOU C. Comparing different classifiers for automatic age estimation [J]. *IEEE Transactions on Systems Man & Cybernetics, Part B: Cybernetics a Publication of the IEEE Systems Man & Cybernetics Society*, 2004, 34(1): 621 – 628.

[5] GENG X, ZHOU Z H, SMITH-MILES K. Automatic age estimation based on facial aging patterns [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2007, 29(12): 2234 – 2240.

[6] AWAD M, KHANNA R. Support vector regression [J]. *Neural Information Processing Letters & Reviews*, 2007, 11(10): 203 – 224.

[7] GELADI P, KOWALSKI B R. Partial least-squares regression: a tutorial [J]. *Analytica Chimica Acta*, 1986, 185(86): 1 – 17.

[8] HARDOON D R, SZEDMAK S, SHAWE-TAYLOR J. Canonical correlation analysis: an overview with application to learning methods [J]. *Neural Computation*, 2004, 16(12): 2639 – 2664.

[9] GUO G, MU G, FU Y, et al. Human age estimation using bio-inspired features [C] //*IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL, USA: IEEE, 2009: 112 – 119.

[10] FU Y, HUANG T S. Human age estimation with regression on discriminative aging manifold [J]. *IEEE Transactions on Multimedia*, 2008, 10(4): 578 – 584.

[11] CAO D, LEI Z, ZHANG Z, et al. Human age estimation using ranking SVM [C] //*Chinese Conference on Biometric Recognition*. Berlin: Springer, 2012: 324 – 331.

[12] HERBRICH R. Large margin rank boundaries for ordinal regression [J]. *Advances in Large Margin Classifiers*, 2000, 88(1): 115 – 132.

[13] GUO G, MU G. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression [C] //*Computer Vision and Pattern Recognition*. Colorado Springs, CO, USA: IEEE, 2011: 657 – 664.

[14] GUO G, MU G. Joint estimation of age, gender and ethnicity: CCA vs. PLS [C] //*IEEE International Conference and Workshops*

on Automatic Face and Gesture Recognition. Shanghai: IEEE, 2010: 1 – 6.

[15] GAO F, AI H. Face age classification on consumer images with gabor feature and fuzzy LDA method [C] //*International Conference on Advances in Biometrics*. Berlin: Springer, 2009: 132 – 141.

[16] GUNAY A, NABIYEV V V. Automatic age classification with LBP [C] //*International Symposium on Computer and Information Sciences*. Istanbul, Turkey: IEEE, 2008: 1 – 4.

[17] YAN S, LIU M, HUANG T S. Extracting age information from local spatially flexible patches [C] //*IEEE International Conference on Acoustics, Speech and Signal Processing*. Las Vegas, NV, USA: IEEE, 2008: 737 – 740.

[18] RAWLS A. MORPH: development and optimization of a longitudinal age progression database [C] //*Joint Cost 2101 and 2102 International Conference on Biometric Id Management and Multimodal Communication*. Berlin, Heidelberg: Springer, 2009: 17 – 24.

[19] DAUGMAN J G. Complete discrete 2–D Gabor transforms by neural networks for image analysis and compression [J]. *IEEE Transactions on Acoustics Speech & Signal Processing*, 1988, 36(7): 1169 – 1179.

[20] YANG M, ZHU S, LV F, et al. Correspondence driven adaptation for human profile recognition [C] //*Computer Vision and Pattern Recognition*. Colorado Springs, CO, USA: IEEE, 2011: 505 – 512.

[21] DUFFNER S. *Face image analysis with convolutional neural networks* [D]. Albert Ludwigs: University Freiburg, 2008.

[22] WANG X, GUO R, KAMBHAMETTU C. Deeply-learned feature for age estimation [C] //*Applications of Computer Vision*. Waikoloa, HI, USA: IEEE, 2015: 534 – 541.

[23] ROTHE R, TIMOFTE R, GOOL L V. DEX: deep expectation of apparent age from a single image [C] //*IEEE International Conference on Computer Vision Workshop*. Santiago, Chile: IEEE, 2016: 252 – 257.

[24] FARABET C, COUPRIE C, NAJMAN L, et al. Learning hierarchical features for scene labeling [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 35(8): 1915 – 1929.

[25] CHEN D, CAO X, WEN F, et al. Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification [C] //*Computer Vision and Pattern Recognition*. Portland, OR, USA: IEEE, 2013: 3025 – 3032.

[26] HUSSAIN S U, TRIGGS B. Visual recognition using local quantized patterns [C] //*British Machine Vision Conference*. Guildford, UK: Springer, 2012: 716 – 729.

[27] LI S Z, YI D, LEI Z, et al. The CASIA NIR–VIS 2.0 face database [C] //*IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Portland, OR, USA: IEEE Computer Society, 2013: 348 – 353.

[28] XIONG X, TORRE F D L. Supervised descent method and its applications to face alignment [C] //*Computer Vision and Pattern Recognition*. Portland, OR, USA: IEEE, 2013: 532 – 539.

[29] KRIZHEVSKY A, SUTSKEVER I, Hinton G E. ImageNet classification with deep convolutional neural networks [C] //*International Conference on Neural Information Processing Systems*. Lake Tahoe, Nevada, USA: Curran Associates Inc, 2012: 1097 – 1105.

[30] GUO G, MU G. Human age estimation: what is the influence across race and gender? [C] //*Computer Vision and Pattern Recognition Workshops*. San Francisco, CA, USA: IEEE, 2010: 71 – 78.

[31] FRANKLIN J. The elements of statistical learning: data mining, inference and prediction [J]. *Mathematical Intelligencer*, 2010, 173(3): 693 – 694.

[32] LI Yunfei, LU Zhaoyang, LI Jing, et al. Personal recognition with nose area biometrics [J]. *Journal of Xidian University (Natural Science)*, 2014, 165(4): 20 – 25.
(李云飞, 卢朝阳, 李静, 等. 鼻子区域生物特征识别 [J]. 西安电子科技大学学报(自然科学版), 2014, 165(4): 20 – 25.)

[33] LIAO S, JAIN A K, LI S Z. Partial face recognition: alignment-free approach [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 35(5): 1193 – 1205.

作者简介:

李云飞　(1974–), 男, 副教授, 研究方向为图像处理、基于人脸的生物特征识别, E-mail: wnlff@126.com;

卢朝阳　(1963–), 男, 教授, 博士生导师, 研究方向为图像分析与图像理解、图像与视频编码; 基于指纹、虹膜及人脸的生物特征识别; 基于图像分析的智能交通系统应用和自然环境文字分析与识别, E-mail: zhylu@xidian.edu.cn;

李　静　(1979–), 女, 副教授, 硕士生导师, 研究方向为图像处理与模式识别、图像配准、文字识别、增强现实,E-mail: jinglinwpu@163.com.