

文章编号: 2096-1472(2016)-08-14-05

## 基于深度卷积网络的人脸年龄分析算法与实现

曹 楠, 莫雅婷, 黄 晨, 魏子涵

(南京理工大学计算机科学与工程学院, 江苏 南京 210094)

**摘 要:**作为一种新兴的生物特征识别技术,基于人脸图像的年龄估计技术在目前已经成为计算机视觉、人机交互等领域的一个重要研究课题。2006年以来,深度卷积网络在图像识别、语音识别和自然语言处理等领域广泛使用,取得了很好的效果。本文基于深度卷积网络的人脸年龄分析算法,构建一个多层卷积神经网络,通过卷积神经网络获取深度卷积激活特征,作为人脸年龄估计的特征,并利用支持向量机(SVM)的方法训练年龄估计模型,得到年龄估计结果,在人脸识别权威数据集Morph上获得了91.3%的正确率,同时也对比了在不同条件下对实验结果的影响。

**关键词:**深度学习;卷积神经网络;年龄估计

**中图分类号:** TP301

**文献标识码:** A

## Algorithm and Implementation of Age Estimation from Facial Images Based on Deep Convolutional Neural Network

CAO Lei, MO Yating, HUANG Chen, WEI Zihan

(Nanjing University of Science and Technology, Nanjing 210094, China)

**Abstract:** As an emerging biometric technology, the age estimation based on facial images has become an important research topic in the field of computer vision, human computer interaction, etc. Since 2006, Deep Convolutional Neural Network has been widely used and obtained good effects in the field of image recognition, speech recognition, natural language processing, etc. In age estimation algorithm based on Deep Convolutional Neural Network, the paper constructs a multilayered convolutional neural network. Taking the deep convolutional activation features as the facial distinctions of age estimation, the paper applies Support Vector Machine (SVM) to train the age estimation model and acquire estimation results. The algorithm has obtained 91.3% accuracy in Morph, the authoritative facial recognition dataset. Meanwhile, the experiment results in different conditions have been compared.

**Keywords:** deep learning; convolutional neural network; age estimation

### 1 引言(Introduction)

人脸年龄分析算法是利用计算机技术根据人脸图像随年龄变化的规律进行建模,从而使机器能够根据人脸图像计算出人的大致年龄或所属的年龄范围。在很多领域,年龄信息作为人体的一种重要生物特征有着巨大的应用潜力<sup>[1-3]</sup>。基于人脸图像的年龄估计技术作为一种新兴的生物特征识别技术,目前已经成为计算机视觉、人机交互等领域的一个重要研究课题,近几年关注年龄估计的研究越来越多。

深度卷积网络作为一种有效的方法,将局部连接、权值共享和空间下采样结合起来,可以简化人脸特征提取流程,但是由于各种干扰条件,可能使得实验结果不是很理想,本文实验重点是探究侧脸,图像分辨率等因素对于实验结果的干扰情况。

### 2 人脸图像预处理与特征提取(Face image pretreatment and characteristic picking-up)

基于人脸图像的年龄估计系统是一种非常典型的机器学习系统,其流程如图1所示。

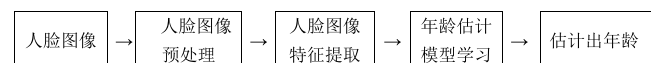


图1 人脸年龄估计流程图

Fig.1 Process of age estimation

#### 2.1 预处理

获取图像的第一个步骤是人脸图像预处理,其主要目的是消除图像中的冗余信息,滤除各种干扰、噪声,提高有关信息的可检测性,从而提高识别的可靠性。这一过程采用的方法有五种:①灰度化;②旋转图片使两眼连线处于水平位

置；③将图片按两种方式进行裁剪(包括嘴和不包括嘴)；④尺度归一化，固定两眼之间的距离及眼睛与鼻尖(包括嘴时)或眼睛与鼻子下端(不包括嘴)时的距离；⑤直方图均衡化。

### 2.1.1 裁剪

裁剪包括嘴时：分别以双眼连线中点、鼻尖的位置坐标作为参照中心，以两眼连线距离及眼睛与鼻尖之间的距离作为参照大小，裁剪出一定大小的人脸图像，裁剪示意图如图2所示。

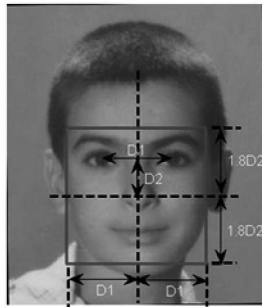


图2 包括嘴的裁剪示意图

Fig.2 Diagram include mouth

裁剪不包括嘴时：分别以两眼连线中点、鼻子下端坐标作为参照中心，以两眼连线距离及眼睛与鼻子下端之间的距离作为参照大小，裁剪出一定大小的人脸图像，裁剪示意图如图3所示。

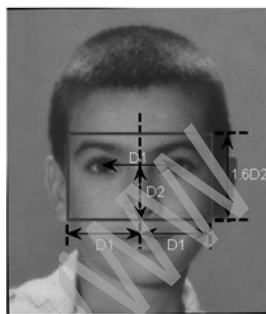


图3 不包括嘴的裁剪示意图

Fig.3 Diagram not include mouth

### 2.1.2 灰度化

彩色图像中的每个像素的值由红、绿、蓝三个通道上的值来决定，而灰度图像的每个像素的值为亮度信息，存储空间仅为彩色图像的三分之一。灰度化可以将彩色图片每个像素的RGB值转化为灰度值，转化公式如下：

$$f = 0.2989 * r + 0.5870 * g + 0.1140 * b$$

其中， $r, g, b$ 分别表示输入的彩色图像的红、绿、蓝三通道上的值， $f$ 为输出的灰度图像对应的灰度值。

## 2.2 特征提取

面部特征提取大致分为三个步骤：

### (1)特征点定位

采用多尺度总和评估ASM方法，在采用一系列的特征点来描述之前，首先针对特定目标建立形状模型之后，这称为点分布模型，然后对形状模型中的每个特征点，建立特征点附近的灰度模型，即建立模型和特征点匹配的过程。

### (2)获取形状信息向量

对由上述过程得到的一组训练样本的特征点位置做主分量分析，得到的由前 $m$ 维特征向量组成的矩阵，通过任意人脸形状在特征空间的投影，即该测试样本通过特征点的坐标向量在这 $m$ 维主元空间上的投影，即可得到一组 $m$ 维的表征形状信息的向量。

### (3)获取纹理信息向量

通过将训练样本的特征点位置拉伸到平均位置，可去除形状信息，此时，图像的灰度信息只包含了训练样本的纹理信息，再对它做主向量分析，可以得到的由前 $n$ 维特征向量组成的矩阵，之后通过该测试样本通过特征点的坐标向量在这 $n$ 维主元空间上的投影，即可得到一组 $n$ 维的表征纹理信息的向量。

## 3 深度卷积网络(Convolutional neural network)

1989年Yan Lecun等人提出可以将BP成功用于训练深度网络的结构——CNN(Convolutional Neural Network)。2012年A.Krizhevsky<sup>[4-6]</sup>等将CNN构造成深度卷积神经网络(Deep Convolutional Neural Network, DCNN)。在这个结构中，并不需要传统的显式特征提取，原始图像的像素可以直接作为输入数据，避免了在浅层学习结构人工提取图像特征时带来的局限性问题，这有利于分类的准确性的提高和模型的泛化能力的增强<sup>[7]</sup>。

### 3.1 卷积

#### 3.1.1 定义

设 $f(x), g(x)$ 是 $R$ 上的两个可积函数，作积分：

$$\int_{-\infty}^{\infty} f(t)g(x-t)dt$$

易证，对于所有的实数 $x$ ，上述积分是存在的。随着 $x$ 的不同取值，我们定义一个新函数 $h(x)$ ，称为函数 $f$ 与 $g$ 的卷积，记为 $h(x) = (f * g)(x)$ 。

因为卷积与傅里叶变换之间有密切的关系，所以利用

一点性质, 即两函数的傅里叶变换的乘积等于它们卷积后的傅里叶变换, 可以使得傅里叶分析中许多问题的处理得到简化。

利用由卷积得到的函数  $f * g$  一般要比  $f$  和  $g$  都光滑这个性质。对于任意的可积函数  $f$ , 都可以简单地构造出一列逼近于  $f$  的光滑函数列  $f_s$ , 这种方法称为函数的光滑化或正则化。

在数列、广义函数以及测度中, 卷积都有很好的应用。

已知两个给定的数列  $\langle f_0, f_1, f_2, \dots \rangle = \langle f_n \rangle$  和  $\langle g_0, g_1, g_2, \dots \rangle = \langle g_n \rangle$ ,

其卷积就是数列  $\langle f_0 g_0, f_0 g_1, f_1 g_0, \dots \rangle = \langle \sum_k f_k g_{n-k} \rangle$ , 易得, 数列的卷积与他们的生成函数的乘积相对应, 这个事实使得许多和式变得易于计算, 如果不是卷积, 这些和式是很难处理的。

### 3.1.2 卷积定理

函数傅里叶变换的乘积是函数卷积的傅里叶变换。即, 一个域中的乘积相当于另一个域中的卷积, 例如时域中的卷积就对应于频域中的卷积。

$$F(g(x) * f(x)) = F(g(x))F(f(x))$$

其中,  $F$  表示的是傅里叶变换, 星号  $*$  表示卷积。

## 3.2 深度卷积网络的基本理论

一个简单的深度卷积网络结构<sup>[7]</sup>如图4所示。

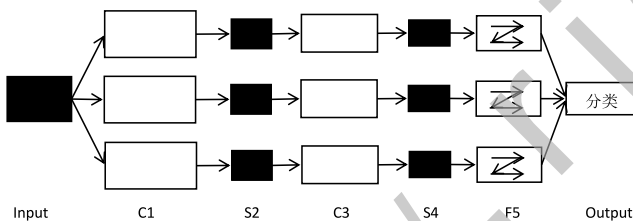


图4 深度卷积网络结构

Fig.4 Structure of convolutional neural network

在这个网络中, 第一层是输入层, 其后是若干个卷积层和下采样层, 最后所有汇到一个分类器。通常一个卷积层后面跟一个下采样层。在深度卷积网络中, 局部连接和权值共享是一个可以降低计算的复杂性、减少网络的训练参数的重要特征。卷积运算得到的结果通过Sigmoid等激活函数的输出达到这一层的特征图, 作为下一层的输入。在下采样层, 将前一层对应的特征图中相邻若干个特征通过池化操作合并成一个特征, 以减小特征的分辨率, 从而实现平移、缩放、扭曲的不变性。

### 3.2.1 卷积层和下采样层

在上面描述的结构细节中, 卷积神经网络涉及的最重要

的前向传播运算是卷积层操作和下采样<sup>[8]</sup>层操作, 其详细过程如下:

#### (1) 卷积层操作

在一个卷积层中, 可学习的卷积核将上一层的特征图进行卷积, 组合这些卷积运算, 然后通过一个激励函数就可以得到一个输出特征图, 公式表达如下:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right)$$

其中,  $x_j^l$  表示第  $l$  层的第  $j$  个特征图,  $f()$  表示激活函数, “ $*$ ” 表示卷积运算,  $M_j$  代表选择的输入图的集合,  $k_{ij}^l$  表示第  $l$  层第  $j$  个特征图与第  $l-1$  层第  $i$  个特征图连接之间的卷积核,  $b_j^l$  表示第  $j$  个特征图对应的偏置项。

#### (2) 下采样层操作

在下采样层, 有  $N$  个输入图, 每个输出图都变小, 也有  $N$  个输出图, 只是每个输出图都变小了, 公式表达为:

$$x_j^l = f(\beta_j^l \text{down}(x_j^{l-1}) + b_j^l)$$

其中, 如  $\text{down}()$  表示下采样函数, 典型的操作就是对输入图像的不同  $n \times n$  大小的区域内的所有像素进行求和。  $\beta_j^l$  表示第  $l$  层第  $j$  个特征图对应的系数,  $b_j^l$  为其对应的偏置项。

### 3.2.2 卷积神经网络的反向传播

卷积神经网络的参数求解形式同经典神经网络一样, 采用梯度下降原则去更新参数, 同样, 最关键的是求出代价函数对每个参数的偏导, 然而经典神经网络不一样的是, 卷积神经网络涉及到的连接个数是多于权值个数的, 所以不能按经典神经网络的反向传播算法计算每层的残差及所需要的偏导, 而这个区别主要是网络的卷积层和下采样层引起的, 所以重点是研究卷积层和下采样层的梯度计算过程。

#### (1) 卷积层的梯度计算

假定每个卷积层  $l$  后面都跟一个下采样层  $l+1$ , 根据经典神经网络的反向传播算法, 要想得到第  $l$  层的每个神经元对应权值的梯度, 就必须首先求第  $l$  层中每一个神经节点的残差  $\delta$ 。

为了求这个残差, 必须先对下一层的节点(连接到当前层  $l$  的感兴趣节点的第  $l+1$  层节点)的残差求和(得到  $\delta^{l+1}$ ), 然后乘以这些连接对应的权值(连接层第  $l$  层节点和第  $l+1$  层节点的权值)  $W$ , 再乘以当前层  $l$  的该神经元节点的输入  $u$  的激励函数  $f$  的导数值(也就是反向传播公式中  $\delta^l$  的求解), 这样也就可以得到

当前层 $l$ 每个神经节点对应的残差 $\delta^l$ 。

由于下采样的存在，采样层的一个神经元节点对应的残差 $\delta$ 对应于上一层的输出特征图的一块区域(下采样窗口大小)。因此，层 $l$ 中每个特征图的每个节点只与 $l+1$ 层中的一个节点连接。上述的残差其实是代价函数对每个神经元的偏导，所以特征图中每个像素都有一个对应的残差，也就组成了特征图对应的残差图。为了有效计算层 $l$ 的残差，需要上采样下采样层对应的残差图，然后将这个上采样得到的残差图与当前层 $l$ 的特征图的激励值的偏导数逐元相乘。下采样层特征图的权值都取一个相同值 $\beta$ ，所以将上一步骤得到的结果乘 $\beta$ 就完成了第 $l$ 层残差的计算。这个过程用公式描述为：

$$\delta_j^l = \beta_j^{l+1} (f(u_j^l) * up(\delta_j^{l+1}))$$

其中， $up()$ 表示点乘运算，即每个像素相乘， $n$ 表示一个上采样操作，过程是若下采样的采样因子是 $n$ 的话，将其每个像素水平和垂直方向上拷贝 $n$ 次，实际上，这个函数可以用Kronecker乘积“ $\otimes$ ”来实现：

$$up(x) = x \otimes I_{n \times n}$$

这样，对于卷积层一个给定的特征图，就可以计算得到其残差图，通过这个残差图，就可以求得该特征图对应偏置项的梯度及对应卷积核的梯度，见下面公式。

$$\frac{\partial J}{\partial b_j} = \sum_{u,v} (\delta_j^l)_{uv}$$

$$\frac{\partial J}{\partial k_{ij}^l} = \sum_{u,v} (\delta_j^l)_{uv} (p_i^{l-1})_{uv}$$

其中， $J$ 为公式的代价函数， $(p_i^{l-1})_{uv}$ 项 $x_i^{l-1}$ 中在卷积时与 $k_{ij}^l$ 逐元素相乘的一块区域， $u, v$ 表示特征图中的像坐标。公式的过程可以通过Matlab中的卷积函数实现：

$$\frac{\partial J}{\partial k_{ij}^l} = rot180(conv2(x_i^{l-1}, rot180(\delta_j^l), 'valid'))$$

#### (2)下采样层的梯度计算

下采样的前向过程涉及的参数是每个特征图对应的一个乘性参数 $\beta$ 和一个偏置项 $b$ ，若是求得此层的残差图，这两个参数的梯度就很容易求得。在计算卷积核的梯度时需要找到输入图中哪个区域对应输出图中哪个像素，这里需要找到当前层的残差图中哪个区域对应下一层的残差图中的给定像素，这样便能将残差反向传播回去，另外，需要乘以输入区域与输出像素之间的权值，公式表达为：

$$\delta_j^l = f'(u_j^l) * conv2(\delta_j^{l+1}, rot180(k_j^{l+1}), 'full')$$

在得到残差图后，乘性因子 $\beta$ 和偏置项 $b$ 的梯度计算过程如下。

$$\frac{\partial J}{\partial b_j} = \sum_{u,v} (\delta_j^l)_{uv}$$

$$\frac{\partial J}{\partial \beta_j} = \sum_{u,v} (\delta_j^l * d_j^l)_{uv}$$

其中， $d_j^l = down(x_j^{l-1})$ 。

### 3.3 基于深度卷积网络的人脸年龄估计算法

现在在人脸年龄估计中，主流的分类方式大部分都是基于统计特征<sup>[9]</sup>的，这就意味着在进行分辨前必须首先提取某些特征。然而，显式的特征提取并不容易在一些应用问题中也并非总是可靠的。

基于深度卷积网络的人脸年龄估计算法创新点主要着眼于特征提取部分，通过构建卷积网络提取出一种新的更有效的隐式年龄估计特征——深度卷积激活特征，并在该特征的基础上采用分类的方法利用支持向量机实现人脸的年龄估计。该算法具体的系统框架流程图<sup>[10]</sup>如图5所示。

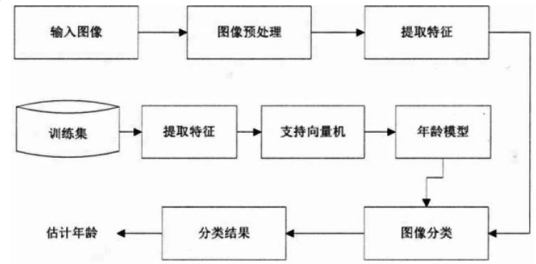


图5 基于深度卷积网络的年龄估计算法框架图

Fig.5 Frame diagram of age estimation based on CNN

## 4 实验结果与分析

### 4.1 数据库

年龄数据库是年龄估计的基础，是年龄估计性能评价的一个基准。作为年龄估计图片库，需要满足三点要求：其一，覆盖尽可能大的年龄区间，通常需要0到80岁；其二，对于各个年龄段的图片数量要基本均等；其三，图片质量尽量高。

#### 4.1.1 FG-NET数据库

FG-NET数据库<sup>[11]</sup>的图像年龄分布如表1所示。从表中可以看出，该数据库覆盖的年龄范围较大，但各年龄段的图片数量相差较大。并且由于图片是通过扫描得到的，图片质量不高。

表1 FG-NET图像库年龄分布表

Tab.1 Age distribution of FG-NET

| 年龄范围  | 人脸图片数 |
|-------|-------|
| 0—9   | 371   |
| 10—19 | 339   |
| 20—29 | 144   |
| 30—39 | 70    |
| 40—49 | 46    |
| 50—59 | 15    |
| 60—69 | 8     |
| 总计    | 1003  |

4.1.2 Morph数据库

Morph数据库<sup>[7]</sup>包含Album1和Album2两部分，其中Album2是目前为止公开的最大的人脸年龄估计数据库，其年龄分布如表2所示。

表2 Morph Album2图像库年龄分布

Tab.2 Age distribution of Morph Album2

| 年龄范围  | 人脸图片数 |
|-------|-------|
| 0—9   | 0     |
| 10—19 | 4971  |
| 20—29 | 14480 |
| 30—39 | 17884 |
| 40—49 | 13668 |
| 50—59 | 4098  |
| 60—69 | 456   |
| 70—79 | 51    |
| 总计    | 55608 |

4.1.3 其他主流数据库

除了FG-NET和Morph两大常用数据库，还有很多其他的用于人脸年龄估计的数据库。

(1)WebFace数据库，是B.Ni等人在2009年再进行有关图像挖掘的方法进行人脸年龄估计的实验是构建的。该数据库的资源来自于网络，利用了一系列有关于年龄的提问在Flicker和Google上采集的。该数据库包含了77021张图像，年龄跨度为1—80岁，是一个比较完备的数据库。

(2)YGA数据库，这个数据库是Y.Gu等人构建的，包含八千张高分辨率的户外人脸图像，共涉及1600个亚洲对象，其中有男性800名，女性800名，年龄跨度为0—93岁，每一个人在同一个年龄段有约五张正脸图像，虽然这个数据库数量并不庞大，但是图像的质量都非常好，有利于人脸年龄估计算法的研究。

(3)WIT-DB数据库，即早稻田人机交互数据库。这个数据库由K.Ueki等人在2006年为研究人机交互技术所构建的人脸数据库，数据库一共包含26222张图片，其中男性为14214张，女性为12008张，涉及5500个不同的日本对象。这些人脸图像基本上都是自然表情，并且是在较为广泛的光照条件下获得，年龄范围包括3—85岁。这个数据库年龄跨度大，图片质量好，是人脸年龄估计的较好数据库之一。

4.2 实验环境配置

实验中硬件配置为Intel i7-6700处理器，主频2.60 GHz，8 GB内存，显卡Nvidia GTX960。软件配置为Linux Mint 17.3 (类似Ubuntu 14.04)，Cuda6.5 GPU并行计算库，网络训练工具采用Tensorflow。

4.3 在数据库下的评测结果

本文主要是针对FG-NET和Morph两大常用数据库进行测试，从表3可知影响识别精度的各项因素。预处理操作十分重要，在正确率上提高巨大。其中侧脸的影响最大，灰度和正确率影响较小。采用Morph数据集训练后，精度上也有20%的提升。

表3 条件对精度的影响

Tab.3 Influences in different condition on accuracy

| 训练集    | 侧脸 | 灰度 | 分辨率 | 正确率   |
|--------|----|----|-----|-------|
| FG-NET | 是  | 高  | 低   | 21.3% |
| FG-NET | 是  | 原色 | 低   | 22.4% |
| FG-NET | 是  | 低  | 低   | 29.4% |
| FG-NET | 是  | 原色 | 适中  | 33.8% |
| FG-NET | 否  | 低  | 低   | 68.8% |
| FG-NET | 否  | 高  | 低   | 67.1% |
| FG-NET | 否  | 高  | 适中  | 71.1% |
| FG-NET | 否  | 原色 | 适中  | 80.5% |
| FG-NET | 否  | 低  | 高   | 70.8% |
| FG-NET | 否  | 低  | 适中  | 77.9% |
| Morph  | 是  | 原色 | 适中  | 33.1% |
| Morph  | 否  | 原色 | 适中  | 91.3% |
| Morph  | 否  | 高  | 适中  | 88.3% |

5 结论(Conclusion)

本文对人脸年龄估计领域做了一个综述，之后实现出一种简单的卷积神经网络来对比在不同的条件下的正确率。实验结果显示，即使在有限的数

据支持下，本文方法依然能够在实际场景中取得不错的性能，同时，侧脸图片灰度不

(下转第8页)