

# 基于多特征深度学习的人脸性别识别

蒋雨欣<sup>1,2</sup>, 李松斌<sup>1,2+</sup>, 刘 鹏<sup>1,2</sup>, 戴琼兴<sup>1</sup>

(1. 中国科学院声学研究所南海研究站, 海南 海口 570105;

2. 中国科学院声学研究所 国家网络新媒体工程技术研究中心, 北京 100190)

**摘 要:** 为解决人脸性别识别这一传统模式识别领域中富有挑战性的难题, 提出一种联合人脸高层特征学习和低层特征学习的深层网络模型。利用卷积和下采样交替提取人脸局部的、抽象的特征, 重建原始的输入人脸图像, 以学习人脸全局的特征作为补充。联合两类特征时, 加入一个可训练的权重进行调节, 利用这种多特征结构进行最终的性别分类。实验结果表明, 该方法在学习能力和泛化能力上具有超越现有人脸性别识别方法的优秀性能。

**关键词:** 人脸性别识别; 深度学习; 多特征学习; 学习能力; 泛化能力

**中图分类号:** TP391.41 **文献标识号:** A **文章编号:** 1000-7024 (2016) 01-0226-06

**doi:** 10.16208/j.issn1000-7024.2016.01.043

## Face gender recognition based on multi-feature deep learning

JIANG Yu-xin<sup>1,2</sup>, LI Song-bin<sup>1,2+</sup>, LIU Peng<sup>1,2</sup>, DAI Qiong-xing<sup>1</sup>

(1. Haikou Lab, Institute of Acoustics, Chinese Academy of Sciences, Haikou 570105, China; 2. National Network New Media Engineering Research Center, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** To address the challenging problems of face gender recognition based on facial images in the traditional pattern recognition field, a deep neural network model that learnt the joint features of human face was proposed. The convolution together with sub-sampling was adopted to obtain the human face's abstract and local features. The human face's global feature was extracted through reconstructing the facial image as further information. When combining these two kinds of features, a trainable weight was added in the network to improve the final gender recognition performance. Experimental results verify that the proposed method has better learning ability and generalization ability than current face gender recognition methods.

**Key words:** gender recognition; deep learning; multi-feature learning; learning ability; generalization ability

## 0 引 言

现有的人脸性别识别方法一般包括人脸检测、特征提取、分类器分类 3 个步骤, 而分类准确率是性别识别中最重要的评价指标。Mäkinen 和 Raisamo<sup>[1]</sup>对 4 种现有主要的性别识别方法进行了综合的评估比较。这 4 种方法分别为局部二值模式 (local binary pattern, LBP) 特征结合支持向量机 (support vector machine, SVM) 分类器的方法 (以下简称 LBP+SVM 方法)、像素特征结合 SVM 分类器的方法 (以下简称 SVM 方法)、基于人工神经网络 (neural networks, NNs) 的方法 (以下简称 NNs 方法)、基于 Haar-like 特征结合 Adaboost 分类器的方法。同时他们还

实验了人脸对齐操作、图像大小、对齐操作时机等因素对结果的影响。根据他们得到的结论, 自动人脸对齐的预处理操作并不能有效提高性别分类的准确率。文中的实验结果表明, SVM 方法在多种实验条件下的平均准确率最高。后来他们通过更多的实验<sup>[2]</sup>进一步指出, 现有的人脸性别识别方法存在泛化能力较弱的缺点, 在高质量的公共人脸数据集上的图片实验时能取得理想的性能, 而在质量较低的网络人脸数据集上的图片测试时, 性能就会明显下降。

另一方面, 近年来基于深度学习思想的方法得到了极大的发展和应用, 在计算机视觉和物体识别等研究领域中取得了突破性的成果<sup>[3-6]</sup>, 为计算机科学向智能化方向发展

**收稿日期:** 2015-03-17; **修订日期:** 2015-05-22

**基金项目:** 国家自然科学基金项目 (61303249); 海南省重大科技基金项目 (JDJS2013006)

**作者简介:** 蒋雨欣 (1989-), 男, 重庆人, 硕士研究生, 研究方向为多媒体信号处理、计算机视觉; +通讯作者: 李松斌 (1981-), 男, 福建漳州人, 副研究员, CCF 会员, 研究方向为多媒体信号处理与取证; 刘鹏 (1989-), 男, 河南新乡人, 博士研究生, 研究方向为多媒体信号处理; 戴琼兴 (1950-), 男, 海南澄迈人, 研究员, 研究方向为信号与信息处理。E-mail: anakin320@126.com

提供了算法上的支持。目前深度学习的方法被越来越多的研究者引入到人脸识别相关研究领域中, 取得了诸多令人满意的成果。Zhang 等<sup>[7]</sup>提出了一种基于深层卷积神经网络(convolutional neural networks, CNNs)并结合多目标学习策略的方法进行人脸检测, 对多姿态、多角度的人脸能够取得非常好的检测效果; Sun 等<sup>[8]</sup>设计了一种3种等级的卷积神经网络, 能够精确定位人脸的关键标注点, 极大地提高了基于标注点进行的人脸检测准确率; Huang 等<sup>[9]</sup>提出一种卷积神经网络结合深度卷积信念网络以及受限制玻尔兹曼机的深层网络模型, 在人脸鉴定和人脸识别问题上取得了非常优异的性能; Sabzevari 等<sup>[10]</sup>提出了一种基于深度信念网络的人脸表情识别系统, 能够快速、准确

地实现人脸表情识别。然而, 目前基于深度学习展开的人脸性别识别研究则非常少, 蔡等<sup>[11]</sup>尝试引入卷积神经网络进行人脸性别分类, 并对网络输出进行置信度评估以提高准确率, 取得了一定的效果。受到深度学习在人脸识别相关领域的应用所启发, 同时为了更全面、深入地研究基于深度学习的人脸性别识别, 本文提出了一种多特征学习深层神经网络来进行人脸性别识别。

## 1 多特征学习深层神经网络

多特征学习深层神经网络主要由3部分组成, 高层特征提取模块, 低层特征提取模块和联合表决输出模块, 其整体结构如图1所示。

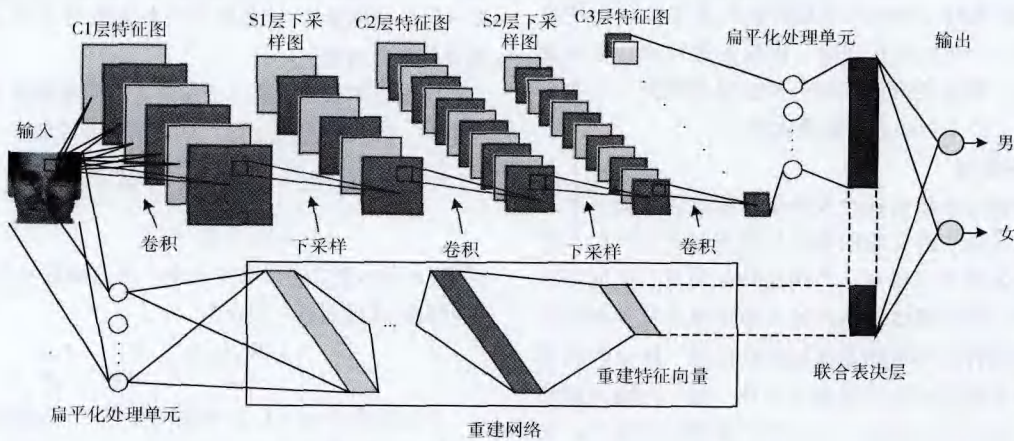


图1 多特征学习深层神经网络结构

### 1.1 高层特征提取

高层特征提取模块主要由卷积层和下采样层组成, 每一层由多个二维平面组成, 而每个平面又由多个独立神经元组成。当卷积层和下采样层组合在一起交替对输入的图像进行卷积、下采样的操作时, 可以提取出输入人脸图像的高层抽象特征。

卷积层中, 每一个神经元只与上一层的一块局部接收域相连。假设接收域(也称作卷积核或者滤波窗口, filter)大小为 $5 \times 5$ , 当输入图像为 $36 \times 36$ 大小时, 连接输入的卷积层将会用这 $5 \times 5$ 的filter卷积图片所有可能的 $32 \times 32$ 位置的像素点, 生成一张 $32 \times 32$ 的局部特征图, 当采用多个不同权值的filter时, 就能够提取出原图的不同局部特征。通过权值共享机制, 可大大减少需要训练的参数量。本文在卷积层中采用sigmoid函数进行神经元的激活, 具体的操作可表示为

$$a_j^l = \text{sigm}(\sum_{i \in M_j} a_i^{l-1} \otimes k_{ij}^l + b_j^l) \quad (1)$$

式中: “ $\text{sigm}()$ ”表示sigmoid函数;  $a_j^l$ 表示第 $l$ 层第 $j$ 张图的输出,  $a_i^{l-1}$ 表示第 $(l-1)$ 层即上一层的输出,  $k_{ij}^l$ 表示第 $l$ 层连接上一层第 $i$ 张图和这一层第 $j$ 张图的卷积核, 而

“ $\otimes$ ”表示卷积,  $b_j^l$ 表示第 $l$ 层第 $j$ 张图对应偏置, 同样在该卷积层中的所有特征图共享一个偏置值。

下采样层中, 每一个神经元仍然只与上一层的一块局部接收域相连, 对上一层的特征图进行下采样, 降低空间分辨率。这样可以充分利用图像的局部相关性, 在减少数据处理量的同时保留有用信息。本文采用平均下采样策略。

高层特征提取模块通过卷积层和下采样层的结合, 具有局部接收域, 权值共享, 空间子采样3个突出的特点, 使得最后提取到的特征获得了一定程度上的位移、尺度以及形状的不变性。具体实现时, 其结构包括5层:  $C_1$ 、 $S_1$ 、 $C_2$ 、 $S_2$ 和 $C_3$ 。其中 $C_1$ 、 $C_2$ 、 $C_3$ 层为卷积层,  $S_1$ 和 $S_2$ 为下采样层。 $C_1$ 层由6张特征图构成,  $C_2$ 层由12张特征图构成,  $C_3$ 层中则含有120张特征图。所有卷积核的大小都设置为 $5 \times 5$ 。下采样层生成的下采样图与上一层的特征图一一对应, 下采样层的尺度均为 $2 \times 2$ 。

### 1.2 低层特征提取

低层特征提取模块主要由扁平化处理单元和重建网络构成。扁平化处理单元首先将输入的人脸二维图像结构转化为一维的人脸原始特征向量, 然后输入到重建网络进行向量的重建。



重建网络采用人工神经网络结构,每一层由多个神经元组成,层与层之间采取全连接的方式,每一个神经元采用 sigmoid 函数进行激活。人脸原始特征向量中含有人脸低层原始特征和全局信息,同时也含有对于性别分类无用的冗余信息以及噪声,故需对其进行重建,提取出对性别分类更为有效的特征信息。通过多层神经元的处理,对人脸原始特征向量进行重新表示和描述,得到最终的重建特征向量。具体实现时,重建网络采用单层 240 个神经元的结构。当采用两层或者更多层的复杂结构组成重建网络时,得到的重建特征包含有复杂的中层人脸特征或更高层的人脸特征,不能起到对高层特征提取模块提取到的高层特征进行补充的作用,反而会导致准确率下降。当单层结构由 160 个神经元组成时,提取到重建特征向量中包含的信息不够;而由 480 个神经元组成时,提取到的重建特征向量又包含了噪声,都会使得最终的分类准确率降低。故本文最终采用单层 240 个神经元构成重建网络。

### 1.3 联合表决输出

联合表决输出模块由联合表决层和输出层组成。高层特征提取模块对输入的人脸图像进行卷积和下采样后,最后通过扁平化处理单元得到一个高层特征向量。而低层特征提取模块经过重建网络对输入的人脸图像进行重新表示后,得到的重建特征向量即是低层特征向量。联合表决层将高层特征向量和低层特征向量合并在一起,与输出层的两个神经元(对应男女两类)全连接。在神经网络中,某一神经元的输入等于由上一层与之连接的所有神经元的加权和,可看作前一层的所有神经元的“表决”结果。那么,本文的多特征学习深度神经网络的输出即是通过联合表决层结合了高层特征和低层特征进行最终的分类判决。

在输出层本文仍采用 sigmoid 函数激活,输出神经元表示某一类的概率  $p_i$  为

$$p_i = \text{sigm}(\sum_k w_k^{l-1} x_k^{l-1} + b_i^l) \quad (2)$$

式中:  $x_k^{l-1}$  ——第  $(l-1)$  层第  $k$  个神经元的输出,  $w_k^{l-1}$  ——第  $(l-1)$  层第  $k$  个神经元与输出层第  $i$  个神经元的连接权值,  $b_i^l$  ——对应偏置。

由于整个网络训练时本文采取随机梯度下降(stochastic gradient descent, SGD)策略,每次输入网络的是一批样本图片,那么网络的输出也是一批样本图片的预测结果,可表示为矩阵形式

$$\mathbf{o} = \text{sigm}(\mathbf{w}_1 \cdot \mathbf{h}_f + \mathbf{w}_2 \cdot \mathbf{l}_f + \mathbf{b}) \quad (3)$$

式中:  $\mathbf{o}$  表示网络输出矩阵(每一列表示一个样本的输出);“ $\text{sigm}()$ ”在这里表示对矩阵中每一元素进行 sigmoid 函数激活;  $\mathbf{w}_1$  表示输出层与高层特征提取模块的输出相连接的权值矩阵;“ $\cdot$ ”表示矩阵乘法;  $\mathbf{h}_f$  表示高层特征提取模块的输出矩阵(每一列表示一个样本的高层特征向量);  $\mathbf{w}_2$  表示输出层与低层特征提取模块的输出相连接的权值矩阵;

$\mathbf{l}_f$  表示低层特征提取模块的输出矩阵(每一列表示一个样本的低层特征向量);  $\mathbf{b}$  表示输出层偏置矩阵。

考虑到低层特征主要是作为高层特征的补充信息,对最后分类的决定作用占比例较小,故本文在低层特征提取模块的输出中加入一个调节权重  $\beta, 0 \leq \beta \leq 1$ , 以抑制低层特征对最终分类的影响。网络输出最终变为

$$\mathbf{o} = \text{sigm}(\mathbf{w}_1 \cdot \mathbf{h}_f + \beta \times \mathbf{w}_2 \cdot \mathbf{l}_f + \mathbf{b}) \quad (4)$$

### 1.4 网络的训练

本文采用误差反向传播算法进行网络的训练。由于深层网络的训练需要大量训练样本的支持,实际运算时采用 SGD 策略实现参数的更新。以样本标签值与网络输出值之间的均方误差作为误差函数  $J(W, b; \beta)$  (由权值  $W$ , 偏置  $b$  以及  $\beta$  等参数确定), 计算网络中各参数对  $J(W, b; \beta)$  的梯度来更新参数的值。

由于调节权重  $\beta$  是一个实数,当使用梯度下降法更新  $\beta$  的值时,涉及到矩阵运算,具体更新公式如下

$$\begin{cases} \beta_{\text{new}} = \beta_{\text{old}} - \alpha \cdot \frac{\partial J(W, b; \beta)}{\partial \beta} \\ 0 \leq \beta_{\text{new}}, \beta_{\text{old}} \leq 1 \end{cases} \quad (5)$$

式中:  $\alpha$  表示整个网络的学习率,误差函数对  $\beta$  求偏导的式子可由链式规则进一步化为

$$\frac{\partial J(W, b; \beta)}{\partial \beta} = \frac{\partial J}{\partial \mathbf{o}} \cdot \frac{\partial \mathbf{o}}{\partial \beta} \quad (6)$$

$\beta$  与矩阵  $(\mathbf{w}_2 \cdot \mathbf{l}_f)$  相乘可看作一个与矩阵  $(\mathbf{w}_2 \cdot \mathbf{l}_f)$  相同行列数的矩阵  $\mathbf{A}$  与矩阵  $(\mathbf{w}_2 \cdot \mathbf{l}_f)$  点乘,  $\mathbf{A}$  中元素值全都为  $\beta$ 。这样,式(6)最终可化为

$$\begin{cases} \frac{\partial J(W, b; \beta)}{\partial \beta} = \text{mean}(\mathbf{B}(:)) \\ \mathbf{B} = (\mathbf{w}_2 \cdot \mathbf{l}_f) \cdot f'(\mathbf{o}) \\ f'(\mathbf{o}) = \mathbf{o} \cdot (1 - \mathbf{o}) \end{cases} \quad (7)$$

式中:“ $\text{mean}(\mathbf{B}(:))$ ”表示对矩阵  $\mathbf{B}$  中所有元素取平均值运算,矩阵  $\mathbf{B}$  行列数与网络输出  $\mathbf{o}$  的行列数相同,  $\mathbf{B}$  中所有元素值为更新后的  $\beta$ 。“ $\cdot$ ”表示矩阵的点乘运算。 $f'(\mathbf{o})$  表示对输出函数求导,当激活函数采用 sigmoid 函数时,其导数形式为  $f'(x) = f(x)(1 - f(x))$ 。

网络中剩余参数的值的训练,都遵循 SGD 策略进行,直到最终输出误差满足精度要求,网络收敛时结束训练过程。为了确保网络最终能够收敛,整个网络的学习率  $\alpha$  被设置为可变变量,它的值随着训练迭代次数的增加而减小。实验结果显示当  $\alpha$  的初始值大于 0.9 时,网络在完成 150 或更多次训练迭代后可以保证收敛,故本文最终设定  $\alpha$  的初始值为 1。

## 2 实验

参考 Mäkinen 和 Raisamo 的结论,本文选取了 LBP+SVM 方法, SVM 方法作为主要比较对象,两者代表了现

有的“浅层学习”类人脸性别识别方法。同时,为更好地体现本文提出的深层网络结合了高层特征学习和低层特征学习的优点,蔡等的CNNs方法和NNs方法这两种分别代表着高层特征提取和低层特征提取的算法,也被选为对比方法。

### 2.1 实验样本介绍

本文首先使用了 FERET、CAS-PEAL、Labeled Face in the Wild (LFW) 以及自行从互联网搜集的人脸图片数据集 1 (以下简称为 IFD-1), 4 种不同的数据源混合构成了一个 13 500 张的混合人脸数据集, 进行不同方法之间的学习能力对比实验。该混合人脸数据集男女比例为 1:1, 包含有各个年龄段, 不同肤色和种族的人脸。

为进一步比较各方法的泛化能力, 本文使用了另外 3 个公共人脸数据集作为新的测试集。3 个公共人脸数据集分别为 AR 数据集, BioID 数据集以及 IMM 数据集。AR 数据集含有 1680 张图片, 其中男性图片 910 张, 女性图片 770 张; BioID 数据集含有 1521 张照片, 人脸检测最终提取出 1275 张人脸图片, 其中男性图片 776 张, 女性图片 499 张; IMM 人脸数据集含有 240 张图片, 人脸检测提取到了全部 240 张人脸, 其中男性图片 198 张, 女性图片 42 张。因为混合人脸数据集以及 3 个测试用的公共人脸数据集中, 每个人都有多张图片, 为了最大程度地贴近实际应用场景, 本文再使用了两个由网络人脸图片构成的数据集 2 和 3 (以下简称为 IFD-2 和 IFD-3) 作为测试集。IFD-2 和 IFD-3 中的人脸都是一人一张图片, 且与 IFD-1 中的人脸独立不重复。IFD-2 总共包含 1801 张较清晰的图片, 其中男性 786 张, 女性 1015 张; IFD-3 总共有 3842 张质量较差的图片, 其中男性 2305 张, 女性 1537 张。这样, 不同方法间泛化能力的对比实验就在 3 个公共人脸数据集和 2 个网络人脸数据集上展开。上述数据集部分样本如图 2 所示。



图2 本文实验数据集部分样本

### 2.2 实验步骤

本文对所有实验图片首先进行人脸检测提取图片中的人脸, 本文直接采用 Zhu 和 Ramanan<sup>[12]</sup> 的方法进行人脸检测。然后对人脸图片进行简单预处理。根据 Mäkinen 和

Raisamo 的结论, 预处理时没有采取任何的自动人脸对齐操作, 仅仅对图片进行了灰度化处理, 调整图片大小至  $36 \times 36$ , 最后对图片做直方图均衡化处理。

本文中提出的多特征学习深层神经网络具体结构如前所述, 训练时采用 SGD 策略, 每次迭代过程中将训练集的图片划分为每 50 张一组, 输入到网络中进行训练。为了更好地实时监控训练过程, 本文记录每一次参数更新后, 网络输出与样本标签的均方误差 (mean squared error, MSE) 的变化, 并画出其曲线图, 如图 3 所示。在画曲线图时对 MSE 做了如下的平滑处理

$$MSE_{new} = MSE_{old} \times 0.99 + MSE_{current} \times 0.01 \quad (8)$$

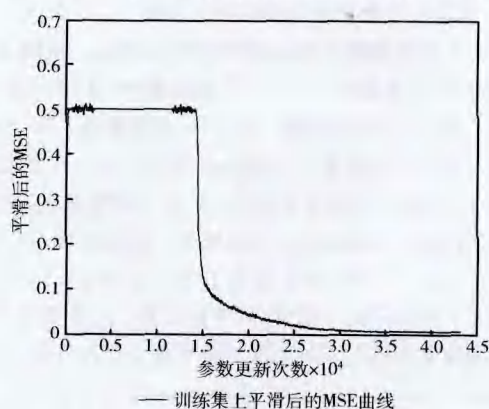


图3 训练过程中平滑后的 MSE 曲线

通过观察 MSE 的变化曲线图, 发现当权值更新次数大约为 3 万次左右, 也即总迭代次数大约为 150 次左右时, 网络在训练集上收敛, 继续增加迭代次数网络在训练集上的性能没有明显提升, 本文最终设置总迭代次数 (epoch) 为 200 次。

实现 LBP+SVM 方法, SVM 方法以及 NNs 方法时, 本文均复现 Mäkinen 和 Raisamo 的做法。在应用 SVM 分类器算法时, 直接使用 LIBSVM 工具进行实验, 每次使用该工具时, 通过调整参数  $C$  和  $\gamma$  的值使得 SVM 最终的分类准确率到达最大值。

实现 CNNs 方法时本文复现了蔡等的卷积神经网络, 网络的结构包括两层卷积层  $C'_1$  及  $C'_2$ ,  $C'_1$  卷积核大小为  $5 \times 5$ ,  $C'_1$  由 10 张特征图构成,  $C'_2$  卷积核大小为  $3 \times 3$ ,  $C'_2$  由 26 张特征图构成。网络结构还包括两层下采样层  $S'_1$  和  $S'_2$ , 都为  $2 \times 2$  的下采样尺度。最后, 网络还包括两个全连接层  $N1$  和  $N2$ , 这两层的激活函数均采用如下的双曲线正切激活函数

$$\phi(X) = 1.7159 \times \tanh(2/3 \times X) \quad (9)$$

训练该卷积神经网络时, 划分训练集为每 50 张图片一组进行参数的更新, 总共进行了 250 次迭代。

当进行学习能力比较测试时, 以混合人脸数据集为基



础,对所有方法进行了5折交叉实验。这样在相同来源的训练集和测试集下进行多组实验,最终得到的平均准确率的高低能够直观地反映出不同方法学习能力的优劣。当进行泛化能力比较测试时,本文重点比较了本文的多特征学习深度学习神经网络方法,LBP+SVM方法以及SVM方法的泛化能力。本文直接使用这3种方法在混合人脸数据集上训练好的模型,然后在3个公共人脸数据集以及两个网络人脸数据集上进行测试。这样训练集和测试集分别选取自不同数据源进行实验的方式能够客观地体现出不同方法间泛化能力的强弱。

## 2.3 实验结果与分析

### 2.3.1 混合人脸数据集实验结果与分析

混合人脸数据集的实验结果如图4所示。从图4中可看出,SVM方法取得了87.76%的准确率,LBP+SVM方法取得了83.49%的准确率,而NNs方法取得了85.78%的准确率,这3者的结果与Mäkinen和Raisamo中的结论基本相符。CNNs方法最终可达到86.87%的准确率,而本文提出的多特征学习深度学习神经网络取得了最高的89.78%的准确率,比LBP+SVM方法提高了6个百分点,比NNs方法提高了4个百分点,比SVM方法提高了2个百分点,也比同属深度学习的CNNs方法提高了近3个百分点。

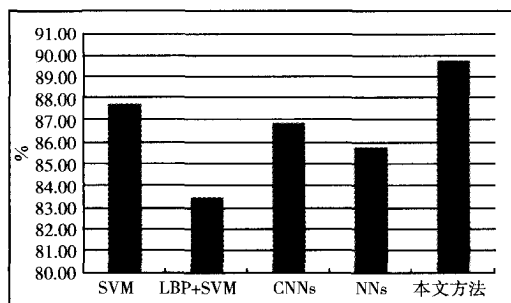


图4 混合人脸数据集上不同方法的平均性别分类准确率

图5则详尽地显示了本文的多特征学习深度学习神经网络在交叉实验的某一折训练过程中,分别在训练集和测试集上的性能变化。

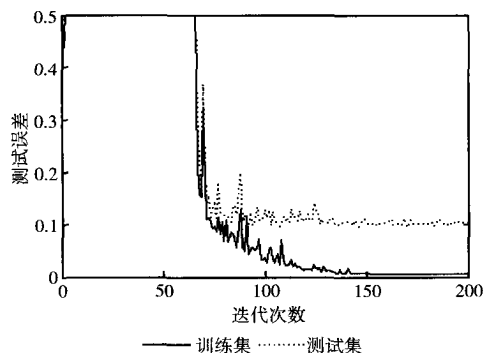


图5 多特征学习深度学习神经网络的性能变化曲线

混合人脸数据集的实验结果显示,应用同样属于深度学习的CNNs方法进行人脸性别识别时,取得了接近现有的“浅层学习”方法SVM方法的性能。然而,CNNs方法学习提取到的仅仅是人脸的高层特征,忽略了人脸中低层特征对分类的有效信息,所以最终的性能仍不理想。同样,如果应用仅仅提取人脸的低层特征的NNs方法,其性能仍未超过SVM方法。只有当使用本文提出的多特征深度学习神经网络,同步学习人脸的高层特征和低层特征,两者联合在一起形成一种多特征的结构,应用这种全方位、多层次的人脸特征进行性别识别时才会取得超过现有的“浅层学习”方法的最高准确率。混合人脸数据集的实验结果证明了本文方法的有效性和优秀的学习能力。

### 2.3.2 公共人脸数据集实验结果与分析

3种方法在3个公共人脸数据集上的测试结果如图6所示。本文的多特征学习深度学习神经网络在AR公共人脸数据集上取得了最高准确率70.6%,超过SVM方法所取得的67.83%的准确率和LBP+SVM方法所取得的66.01%的准确率。同样,本文方法在BioID公共人脸数据集上也取得了最高准确率87.92%,超过SVM方法所取得的86.17%的准确率和LBP+SVM方法所取得的82.93%的准确率。在IMM公共人脸数据集上,本文方法与SVM方法取得了同样高的准确率87.08%,超过LBP+SVM方法所取得的81.25%的准确率。SVM方法在3个公共人脸数据集的平均准确率为80.36%,LBP+SVM方法的平均准确率为76.73%。而本文方法的平均准确率为81.87%,与LBP+SVM方法相比,提高了5个百分点,与SVM方法相比,提高了1.6个百分点。

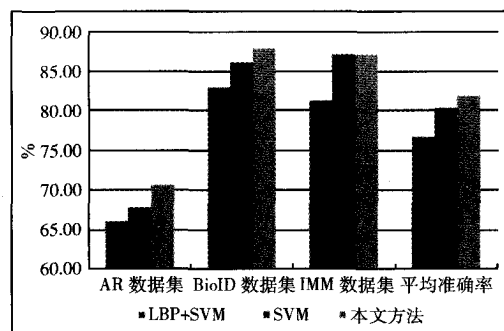


图6 不同方法在3个公共人脸数据集上测试的性别分类准确率

### 2.3.3 网络人脸数据集实验结果与分析

3种方法在两个网络人脸数据集上的测试结果如图7所示。本文的多特征学习深度学习神经网络在IFD-2数据集上取得了最高准确率89.84%,超过LBP+SVM方法所取得的86.45%的准确率和SVM方法所取得的85.17%的准确率。本文方法在IFD-3数据集上也取得了最高的88.8%的准确率,超过LBP+SVM方法所取得的85.27%的准确率和

SVM方法所取得的85.14%的准确率。LBP+SVM方法在两个网络人脸数据集的平均准确率为85.86%, SVM方法的平均准确率为85.16%。而本文方法的平均准确率为89.32%, 与SVM方法相比, 提高了4个百分点, 与LBP+SVM方法相比, 提高了3.5个百分点。

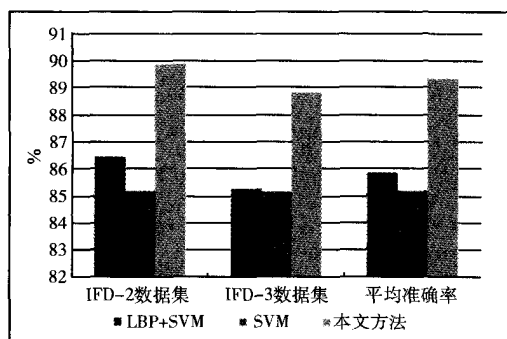


图7 不同方法在两个网络人脸数据集上测试的性别分类准确率

公共人脸数据集的实验结果和网络人脸数据集的实验结果, 也即泛化能力的测试结果验证了本文提出的方法具有超过现有的“浅层学习”类方法的泛化能力, 这表明本文提出的方法不仅有着比现有方法更出色的学习能力, 也有着比现有方法更强的泛化能力。

### 3 结束语

笔者基于深度学习提出了一种多特征学习深层神经网络来解决人脸性别识别这一难题。相比于现有的“浅层学习”类的方法, 该深层网络能够同时学习到人脸的高层特征, 同时结合人脸的低层特征一起构成一种多特征结构, 最后应用这种全方位、多层次的人脸特征进行性别识别能取得优秀的性能。实验结果显示, 该方法的识别准确率比现有的性能最优的“浅层学习”类方法提高了2个百分点, 比现有的“深度学习”类方法提高了近3个百分点, 验证了方法具有超过现有方法的学习能力, 同时也验证了多特征学习的优越性。进一步, 跨数据集进行的泛化能力实验结果表明, 该方法的识别准确率比现有的性能最优的方法提高了3.5个百分点, 其不仅具有很强的学习能力, 还具有超过现有方法的泛化能力。

### 参考文献:

[1] Mäkinen E, Raisamo R. Evaluation of gender classification methods

with automatically detected and aligned faces [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30 (3): 541-547.

- [2] Mäkinen E, Raisamo R. An experimental comparison of gender classification methods [J]. Pattern Recognition Letters, 2008, 29 (10): 1544-1556.
- [3] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks [C] //Advances in Neural Information Processing Systems, 2012: 1106-1114.
- [4] Ciresan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification [C] //IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3642-3649.
- [5] Susskind J, Mnih V, Hinton G. On deep generative models with applications to recognition [C] //IEEE Conference on Computer Vision and Pattern Recognition, 2011: 2857-2864.
- [6] Sermanet P, Kavukcuoglu K, Chintala S. Pedestrian detection with unsupervised multi-stage feature learning [C] //IEEE Conference on Computer Vision and Pattern Recognition, 2013: 3626-3633.
- [7] Zhang C, Zhang Z. Improving multiview face detection with multi-task deep convolutional neural networks [C] //IEEE Winter Conference on Applications of Computer Vision, 2014: 1036-1041.
- [8] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection [C] //IEEE Conference on Computer Vision and Pattern Recognition, 2013: 3476-3483.
- [9] Huang GB, Lee H, Learned-Miller E. Learning hierarchical representations for face verification with convolutional deep belief networks [C] //IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2518-2525.
- [10] Sabzevari M, Toosizadeh S, Quchani SR, et al. A fast and accurate facial expression synthesis system for color face images using face graph and deep belief network [C] //International Conference on Electronics and Information Engineering, 2010: 354-358.
- [11] CAI Shiwei, GUO Tailing, YAO Jianmin. Gender recognition method based on convolutional neural networks [J]. Video Engineering, 2014, 38 (19): 188-191 (in Chinese). [蔡诗威, 郭太良, 姚剑敏. 一种基于卷积神经网络的性别识别方法 [J]. 电视技术, 2014, 38 (19): 188-191.]
- [12] Zhu X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild [C] //IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2879-2886.