

基于改进深层网络的人脸识别算法

李倩玉, 蒋建国, 齐美彬

(合肥工业大学计算机与信息学院, 安徽合肥 230009)

摘要: 目前的人脸识别算法在其特征提取过程中采用手工设计(hand-crafted)特征或利用深度学习自动提取特征. 本文提出一种基于改进深层网络自动提取特征的人脸识别算法, 可以更准确地提取出目标的鉴别性特征. 算法首先对图像进行ZCA(Zero-mean Component Analysis)白化等预处理, 减小特征相关性, 降低网络训练复杂度. 然后, 基于卷积、池化、多层稀疏自动编码器构建深层网络特征提取器. 所使用的卷积核是通过单独的无监督学习获得的. 此改进的深层网络通过预训练和微调, 得到一个自动的深层特征提取器. 最后, 利用Softmax回归模型对提取的特征进行分类. 本文算法在多个常用人脸库上进行了实验, 表明了其在性能上比传统方法和普通深度学习方法都有所提高.

关键词: 人脸识别; 改进的深层网络; 卷积; 池化; 多层稀疏自动编码器

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2017)03-0619-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.03.017

Face Recognition Algorithm Based on Improved Deep Networks

LI Qian-yu, JIANG Jian-guo, QI Mei-bin

(School of Computer & Information, Hefei University of Technology, Hefei, Anhui 230009, China)

Abstract: Current face recognition algorithms use hand-crafted features or extract features by deep learning. This paper presents a face recognition algorithm based on improved deep networks that can automatically extract the discriminative features of the target more accurately. Firstly, this algorithm uses ZCA (Zero-mean Component Analysis) whitening to preprocess the input images in order to reduce the correlation between features and the complexity of the training networks. Then, it organically combines convolution, pooling and stacked sparse autoencoder to get a deep network feature extractor. The convolution kernels are achieved through a separate unsupervised learning model. The improved deep networks get an automatic deep feature extractor through preliminary training and fine-tuning. Finally, the softmax regression model is used to classify the extracted features. This algorithm is tested on several commonly used face databases. It is indicated that the performance is better than the traditional methods and common deep learning methods.

Key words: face recognition; improved deep networks; convolution; pooling; stacked sparse autoencoder

1 引言

人脸识别是近年来计算机视觉领域和机器学习领域中最富挑战性的课题之一, 受到了研究者的广泛关注. 成功有效的人脸识别具有广阔的应用前景, 可在国防安全、视频监控、人机交互和视频索引等场景发挥巨大作用.

一方面, 传统的人脸识别研究, 大都基于 hand-crafted 特征, 例如: Deng 等^[1] 提出对变化具有较强适应性的 PCA (Principal Component Analysis) 特征脸, 用于人脸特征表示, 但其识别性能在光照、姿态变化较大时下降

明显. Zhou 等^[2] 利用 Gabor 小波表示人脸, 计算量较大. Yang 等^[3] 证明了几何特征和纹理特征的融合在人脸识别上的有效性, 构造难度较大. 以上浅层算法在训练样本和计算单元有限时很难有效地表达复杂函数^[4], 揭示了浅层网络的局限性. 同时, 新 hand-crafted 特征的构建消耗大量的时间, 且很难取得明显的效果.

另一方面, 深度学习在动作识别^[5]、目标检测^[6]上的成功应用, 让我们看到了它在模式识别、语音识别和信息检索上面的光明前景. 深度学习网络通过模拟大脑的深度组织结构, 构建多层非线性映射层, 对复杂函数拥有强大的表达能力, 具有从大量样本中学习数据

鉴别性特征的明显优势. 实验研究表明深度网络结构比浅层结构更加有效^[7], 在模式识别等领域表现突出^[8]. 文献[9]使用了 CNN 方法检测人脸关键点, 由于是有监督学习算法, 需要大量有标记样本. 文献[10, 11]以大量无标记样本为支撑, 通过 DBN 方法提取人脸特征, 同时利用部分有标记样本对网络进行微调.

基于此, 本文以深度学习方法为基础, 构建了改进的深层网络, 提出了一种新的有效的人脸识别系统: 首先, 对人脸库进行预处理, 用于减小特征相关性, 降低网络计算复杂度. 然后, 构建一个改进的深层网络特征提取器, 包含卷积层、池化层和双层稀疏自动编码层. 网络训练前, 本文对图片进行采样, 利用无监督学习方法学习网络参数, 获得各类卷积核. 通过此特征提取器获得的人脸特征, 能够全方位、多层次地对人脸进行刻画, 具有较强的鲁棒性. 最后, 利用所得特征, 训练 Softmax 回归模型, 对人脸进行类别判断. 改进的深层网络结构模型可以更加准确地提取出人脸的特征信息, 在识别过

程中获得更好的性能.

2 系统概述

图 1 是本文的算法框架图, 基于改进深层网络的人脸识别算法由三个部分组成: 预处理模块、深层网络特征提取器以及 Softmax 回归模型. 样本经过直方图均衡化、尺度归一化和 ZCA 白化等预处理, 变换到大小为 64×64 、灰度值范围为 $[0, 1]$ 的图像集. 深层网络特征提取器完成样本深层特征的提取, 用来训练 Softmax 回归模型, 对测试集的人脸类别进行判决输出.

预训练阶段利用无监督学习和 L-BFGS 优化算法学习深层网络的参数. 微调阶段利用有监督学习对深层网络参数进行优化, 同时提取深层特征, 输入 Softmax 回归模型进行训练; 测试阶段将测试集数据预处理后经深层网络特征提取器提取特征, 输入 Softmax 回归模型进行类别判决.

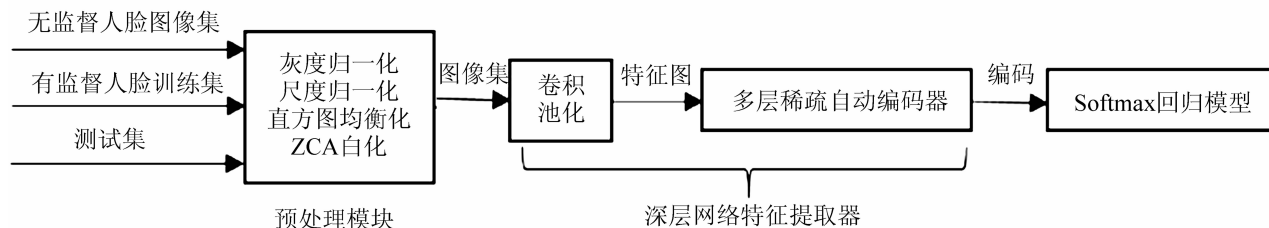


图1 算法框架图

3 预处理模块

预处理模块主要由归一化处理 and ZCA 白化构成. 归一化处理采用尺度归一化、灰度归一化和直方图均衡化, 降低整体网络的计算复杂度, 同时在保留原有信息的基础上减少无用信息. ZCA 白化首先如式(1)所示, 通过 PCA 变换去除了各个特征之间的相关性; 然后利用式(2)使输出特征具有单位方差; 再把数据旋转回去, 得到 ZCA 白化的处理结果, 如式(3)所示, 降低输入的冗余性.

$$\mathbf{x}_{\text{rot}, i} = \mathbf{U}^T \mathbf{x}_i \quad (1)$$

$$\mathbf{x}_{\text{PCAWhite}, i} = \frac{\mathbf{x}_{\text{rot}, i}}{\sqrt{\lambda_i}} \quad (2)$$

$$\mathbf{x}_{\text{ZCAWhite}} = \mathbf{U} \mathbf{x}_{\text{PCAWhite}} \quad (3)$$

输入通过预处理模块得到大小为 64×64 、灰度值范围为 $[0, 1]$ 的图像集. 这样得到的待处理人脸图像数据具有统一的格式, 可以有效降低网络的复杂度和计算难度, 从而节省网络训练用时; 同时减小了特征间的相关性, 为理想特征的提取奠定基础.

4 深层网络特征提取器

4.1 深层网络的介绍

特征提取是各类识别算法的核心步骤, 对最终的识别效果影响巨大. 深度学习通过组合低层特征形成更加抽象的高层特征, 以实现数据的分布式特征表示, 最终提高分类和预测的准确性. 随着计算机内存的扩大和 GPU 计算技术的发展, 各类深度学习算法的研究得以开展, 例如: Deep Belief Networks (DBN)^[10, 11]、Deep Boltzmann Machine (DBM)、Recurrent Neural Networks (RNN)、Sparse Auto Encoder (SAE)^[12, 13] 等无监督深度学习算法和 Convolutional Neural Networks (CNN)^[14] 等有监督深度学习算法.

由于大量的人脸样本是无标记的, 且含有丰富的变化. 故本文采用 SAE (稀疏自动编码器)^[12, 13] 这一无监督学习方法, 对大量无标记样本进行自动学习, 预训练获得深层网络的参数; 采用卷积和池化操作来扩大样本集数量. 卷积和池化使提取的特征具有对位移、尺度以及形变的不变性, 减少网络需要训练的参数个数, 降低模型复杂度. 本文对样本进行子块采样, 并运用单

层降噪稀疏自动编码器方法,学习得到有价值的卷积核.后面实验部分表明深层网络特征提取器中的稀疏自动编码器级数为 2 级时,模型效果最优.基于本文算法结合了卷积、池化和稀疏自动编码等操作,并加以改进,形成了多层的网络结构模型,故称本模型是深层网络结构,提取出的特征为深层特征.

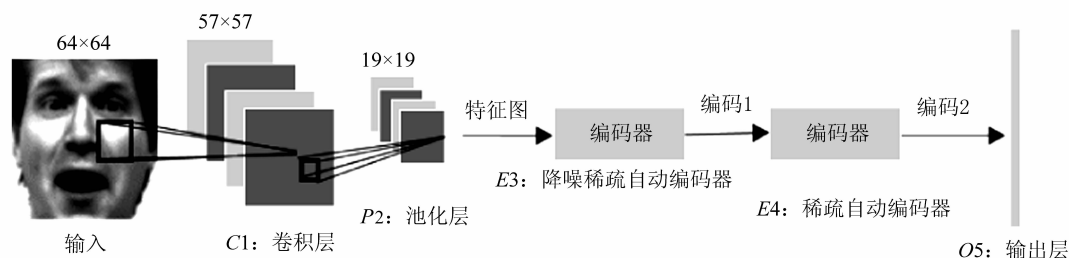


图2 深层网络结构

4.2.1 卷积层、池化层的结构与功能

传统的 CNN 网络中,卷积层的卷积核是随机初始化的,而本文是通过单层降噪稀疏自动编码器进行学习,得到网络的参数,即各类卷积核.同时,本文在卷积和池化过程中使用了修正线性单元^[14](ReLU)对传统的结果进行修正,从而得到对目标具有更强刻画能力的特征,提高识别的准确率.

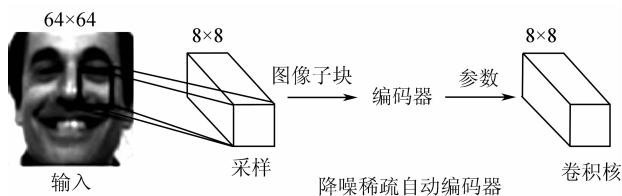


图3 卷积核生成

图 3 显示了各类卷积核生成的过程.针对通过预处理模块的输入图像集,首先进行采样,得到 3 万个 8×8 的图像子块;然后,将这些图像子块输入降噪稀疏自动编码器,结合 L-BFGS 优化算法进行无监督学习,得到网络的参数;最后,将这些参数输出,其结果便是我们学习到的多种卷积核.

在深层特征提取过程中,采用上述学习得到的卷积核实现卷积层的功能.卷积操作时,用一个卷积核 \mathbf{K} 去卷积一个输入图像 \mathbf{x} ,再加上一个偏置 \mathbf{b}_x ,然后通过一个修正线性单元(ReLU)得到卷积层的特征图 \mathbf{C}_x ,如式(4)所示:

$$\mathbf{C}_x = \max(0, \mathbf{K} * \mathbf{x} + \mathbf{b}_x) \quad (4)$$

卷积后得到的特征图的大小为 57×57 ,每个像素与输入图像中 8×8 的邻域(局部感受野)相连.图 2 中 P2 是池化层,通过一个 3×3 的池化窗口对卷积后得到的特征图进行非重叠池化,即每九个像素组成的邻域

4.2 深层网络的实现

深层网络是一个多层的神经网络结构,具有优异的特征学习能力,其学习到的特征对数据有更本质的刻画,从而有助于分类和可视化.如图 2 所示,深层网络包括卷积层 C1、池化层 P2、降噪稀疏自动编码器 E3、稀疏自动编码器 E4 和输出层 O5.

求和得到一个新的像素,然后通过标量 w_{x+1} 加权,再增加偏置 b_{x+1} ,最后通过 ReLU 激活函数,产生一个 19×19 大小的特征映射图 \mathbf{S}_{x+1} .

$$\mathbf{S}_{x+1} = \max(0, \sum \mathbf{C}_x * \mathbf{w}_{x+1} + \mathbf{b}_{x+1}) \quad (5)$$

4.2.2 双层稀疏自动编码器的训练

自动编码器是深度学习的热门算法之一.其主要思想是捕捉可以代表输入信息的主要成分,尽可能复现输入信息.本文一方面在其基础上加入稀疏限制,即满足 L1 的 Regularity 限制条件,使每一层中的节点大部分都是 0,只有少数不为 0.另一方面,我们选用降噪稀疏自动编码器作为第一层自动编码器,在训练数据中人为加入噪声,使网络学习去除这种噪声而获得真正没有被噪声污染过的输入,从而网络具有更强的泛化能力,可以学习输入信号更加鲁棒的表达;第二层自动编码器无需加入噪声,采用一般的稀疏自动编码器.其具体实现过程如下,包括预训练^[15]和微调两个阶段:

预训练阶段:

步骤 1 给定无标签数据,用无监督学习方法学习特征:

如图 4(a) 所示,本文选用搜集的各类人脸样本共 1 万张,经预处理、卷积和池化后输入图中单层降噪稀疏自动编码器.输入 \mathbf{X} 经无监督学习,得到编码 \mathbf{h} ,再经解码器重建输入,计算输入与重建结果间的重构误差,构建代价函数 $L(\mathbf{X}; \mathbf{W})$.

$$\mathbf{h} = \mathbf{W}^T \mathbf{X} \quad (6)$$

$$L(\mathbf{X}; \mathbf{W}) = \|\mathbf{W}\mathbf{h} - \mathbf{X}\|^2 + \lambda \sum_j |h_j| \quad (7)$$

利用 L-BFGS 算法,调整编码器和解码器的参数,使代价函数 $L(\mathbf{X}; \mathbf{W})$ 最小.保存此时的第一层网络

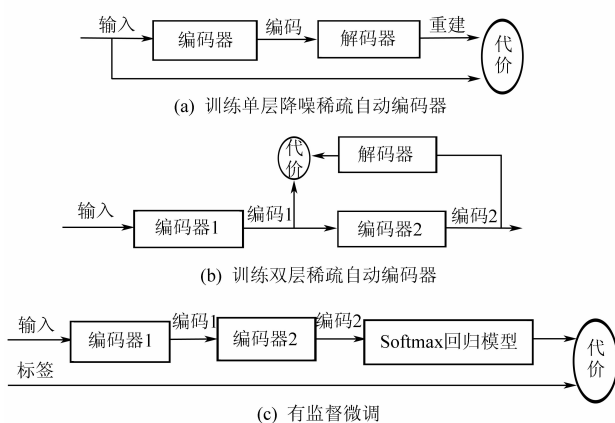


图4 自动编码器的预训练与微调

参数.

步骤 2 训练下一层,实现逐层训练:

第 1 层网络训练完成后,开始第 2 层网络的训练,如图 4(b)所示,将第 1 层网络输出的编码 1 作为第 2 层的输入信号,重复步骤 1,得到原输入信息的第 2 层表示编码 2,并保存此时的第二层网络参数.

微调阶段:

步骤 3 有监督微调:

利用实验所选标准库中的训练集进行微调,如图 4(c)所示,将编码 2 输入 Softmax 回归模型,通过标准的多层神经网络的监督训练方法去训练. 本文采用 end-to-end 学习方法,通过有标签样本,微调整个系统. Softmax 回归模型具体训练步骤见后文.

完成上述步骤后,得到的网络便可用来提取深层特征.

5 基于 Softmax 回归模型的分类判决

5.1 Softmax 回归模型简介

传统的 logistic 回归模型用于处理非线性二分类问题,在给出分类结果的同时给出结果的概率. Softmax 回归模型^[16]是 logistic 回归模型在多分类问题上的推广,在多分类问题中,类标签 y 可以取 k 种不同的值 ($k > 2$). 因此,对于训练集 $\{(\mathbf{x}^{(1)}, y^{(1)}), \dots, (\mathbf{x}^{(m)}, y^{(m)})\}$, 有 $y^{(i)} \in \{1, 2, \dots, k\}$.

Softmax 回归模型是有监督学习得到的,它要求所需的分类类别之间严格相互排斥,也就是两种类别不能同时被一个样本占有. 对于给定的输入 \mathbf{x} , 我们想用假设函数针对每一个类别 j 估算出概率值 $p(y=j|\mathbf{x})$. 也就是说,估计 \mathbf{x} 的每一种分类结果出现的概率. 因此,我们的假设函数将要输出一个 k 维的向量(向量元素的和为 1)来表示这 k 个估计的概率值. 具体地说,我们的假设函数 $h_{\theta}(\mathbf{x})$ 形如式(8),其中 $\theta_1, \theta_2, \dots, \theta_k \in \mathbf{R}^{n+1}$ 是模型的参数,而前面的乘积项用于对概率分布

进行归一化,使得所有概率之和为 1.

$$h_{\theta}(\mathbf{x}^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | \mathbf{x}^{(i)}; \theta) \\ p(y^{(i)} = 2 | \mathbf{x}^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | \mathbf{x}^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T \mathbf{x}^{(i)}}} \begin{bmatrix} e^{\theta_1^T \mathbf{x}^{(i)}} \\ e^{\theta_2^T \mathbf{x}^{(i)}} \\ \vdots \\ e^{\theta_k^T \mathbf{x}^{(i)}} \end{bmatrix} \quad (8)$$

Softmax 回归模型的代价函数见式(9). 不同于传统 Softmax 回归模型中选择的代价函数,本文选用的代价函数添加了权重衰减项,即式(9)中第二个加数,会惩罚过大的参数值,使代价函数变成严格的凸函数,保证得到唯一的解,从而避免陷入局部最优解的局面,以获得全局最优解.

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} l \right] + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2$$

$$l = \log \frac{e^{\theta_j^T \mathbf{x}^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T \mathbf{x}^{(i)}}} \quad (9)$$

对于代价函数 $J(\theta)$ 的最小化问题,目前还没有闭式解法. 因此,本文使用迭代的优化算法 L-BFGS. 经过求导,我们得到梯度式(10),将其值代入 L-BFGS 算法中,来最小化 $J(\theta)$, 获得相应参数,实现一个可用的 Softmax 回归模型.

$$\nabla \theta_j J(\theta) = -\frac{1}{m} \sum_{i=1}^m \mathbf{x}^{(i)} P_{\text{pby}} + \lambda \theta_j \quad (10)$$

$$P_{\text{pby}} = 1\{y^{(i)} = j\} - p(y^{(i)} = j | \mathbf{x}^{(i)}; \theta)$$

5.2 训练 Softmax 回归模型

训练样本分为多类,将其经预处理模块和深层网络模块提取的特征输入 Softmax 回归模型. 利用有监督学习算法学习参数,实现一个可用的 Softmax 回归模型. 其训练过程如下:

- (1) 初始化类别数、样本数和权重衰减项等模型参数;
- (2) 加载训练数据及其对应的标签;
- (3) 计算代价函数、对应的梯度函数和预测标签;
- (4) 进行梯度函数的验证,保证数学上的梯度和实际的梯度差异较小;
- (5) 利用 L-BFGS 算法求得代价函数最小值及模型

参数;

此时,完成了 Softmax 回归模型的训练,测试过程中可以直接使用,同时本训练过程也是深层网络特征提取器的微调部分.

6 实验部分

本文实验采用公用人脸识别数据库 Yale、Yale-B 和 CMU-PIE 作为实验对象,来评估本文算法的性能. 硬件配置为: Intel (R) Core (TM) i5-3230M CPU @ 2.60GHz, 8.00GB 内存.

6.1 实验数据和网络参数设定

本文首先收集了 1 万张无标记的人脸用来完成网络的预训练,即获得图 1 中深层网络特征提取器的参数. 针对公用人脸识别数据库进行实验时,将其分为训练集和测试集两部分. 训练集用来微调深层网络特征提取器参数和训练 Softmax 回归模型,测试集用来测试整个系统的性能. 所有样本通过预处理模块,统一到 64×64 大小、灰度值为 $[0, 1]$ 的图像集.

训练深层网络特征提取器中使用的卷积核时,输入设为 8×8 , lambda 设为 $1e-4$,级数设为 1 级,便能得到各类 8×8 大小的卷积核. 池化层窗口大小设为 3×3 ,池化类型为非重叠池化. 训练稀疏自动编码器时,lambda 设为 $1e-4$,级数设为 2 级. 利用 L-BFGS 算法优化网络参数时,动量设为 0.5,学习率设为 0.2. 利用训练好的改进深层网络提取训练样本的特征,将其输入 Softmax 回归模型,利用 L-BFGS 算法进行代价函数

最小化求解,得到最优的网络参数,形成最终的系统模型. 图 5 为本文实验中采用的人脸库部分人脸图片.

6.2 实验结果

6.2.1 不同算法在各类人脸库上的实验结果

(1) Yale 人脸库上的实验结果

Yale 人脸库由 15 个人,每人 11 幅图像,共 165 幅灰度图像构成. 每个人拥有 6 种不同的表情,3 种不同的光照,戴眼镜和不戴眼镜的人脸. 实验中,从每个人脸子集中随机选取 5 幅图像作为训练样本,剩下的作为测试样本. 针对此数据集,使用不同方法进行对比实验.

图 5(a)是本数据库的部分人脸,结合表 1 分析可知:本文算法效果较传统算法和普通深度学习算法都有所提高,说明其对表情、光照和眼镜遮挡等情况具有一定的适应性.

(2) Yale-B 人脸库上的实验结果

Yale-B 人脸库由 10 个人,每人 64 幅,共 640 幅灰度图像构成. 每个人拥有 5 个子集,包含不同光照条件的图像. 实验中,每个人从 5 个子集中分别选取 5 张、6 张、7 张和 10 张,共 34 张作为训练样本,剩下的作为测试样本.

表 1 Yale 库上不同算法时正确识别率

方法	识别率
PCA ^[1]	80.00%
LGBPHS	81.11%
SVM ^[17]	94.44%
LBP	82.22%
分块 LBP	93.33%
SAE ^[13]	96.67%
本文算法	97.78%

表 2 Yale-B 库上不同算法时正确识别率

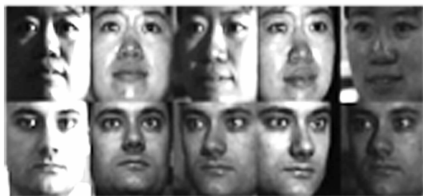
方法	识别率
PCA	66.67%
LGBPHS	73.33%
SVM	80.00%
LBP	82.33%
分块 LBP	84.67%
SAE	93.00%
本文算法	95.00%



(a) Yale 部分人脸



(b) Yale-B 部分人脸



(c) CMU-PIE 部分人脸

图 5 部分实验人脸

图 5(b)是本数据库的部分人脸,可见本数据库中

图像光照变化明显. 结合表 2 分析可知: 针对光照条件改变较大的情况, 本文方法仍然有不错的表现, 鲁棒性良好.

(3) CMU-PIE 人脸库上的实验结果

CMU-PIE 人脸库由 68 个人, 共 41,368 张灰度图像构成, 选取其中 11,554 张完成实验. 该人脸库包含多姿态、光照和表情图像, 其中的姿态和光照变化图像也是在严格控制的条件下采集的. 实验中, 选取其中 7,000 张作为训练样本, 剩下的作为测试样本.

表 3 CMU-PIE 库上不同算法时正确识别率

方法	识别率
PCA	80.14%
LGBPHS	80.06%
SVM	81.51%
LBP	90.54%
分块 LBP	92.93%
SAE	94.20%
本文算法	96.17%

图 5(c) 是本数据库的部分人脸, 可见本数据库含有丰富的姿态和光照变化. 实验结果如表 3, 反应了本文算法对多姿态、光照和表情具有很好的特征表达能力, 效果较其他方法有所提高.

综合分析以上三组实验数据, 表明相对于 hand-crafted 特征和传统深度学习算法, 本文中利用改进的深层网络学习并提取的特征, 对光照、表情、姿态等的改变具有较强的适应性, 平均准确率高, 系统具有较强的泛化能力和鲁棒性.

6.2.2 不同级数在各类人脸库上的实验结果

随着深层网络级数的改变, 所提取的特征具有不同的表达能力. 本文针对稀疏自动编码层的级数进行研究, 在人脸库上进行实验. 如表 4 所示.

表 4 不同级数在各类人脸库上的准确率

稀疏自动编码器的级数	各类人脸库上的识别率		
	Yale 人脸库	Yale-B 人脸库	CMU-PIE 人脸库
1	77.78%	83.33%	85.62%
2	97.80%	95.00%	96.17%
3	95.56%	92.67%	90.82%
4	91.12%	88.33%	87.90%
5	83.33%	81.00%	87.52%
6	75.56%	77.33%	78.17%

分别设定稀疏自动编码层的级数为 1~6 级, 在各类人脸库上完成实验. 由实验结果可见: 起初随着级数

增加, 效果有所提高, 但到达一定级数后, 反而下降. 结合本文所使用的数据集, 我们最终确定使用两级结构的稀疏自动编码器.

7 结束语

本文提出了一个改进的深层网络模型自动学习和提取特征, 并利用 Softmax 回归方法进行准确分类, 完成人脸识别系统. 本文将利用大样本集进行无监督学习, 预训练获得深层网络特征提取器的参数, 与利用实验数据集的训练集进行微调相结合, 提取出具有鉴别性的特征. 本文将改进的深层网络作为特征提取器, 捕捉具有极强表达能力的深层特征, 有助于准确分类. 预处理、卷积、池化和多层稀疏自动编码器的结合及改进, 使系统可以很好地表征人脸. 结合加入权重衰减项的 Softmax 回归模型, 使得本文算法具有较强的泛化能力和优异的性能. 其在公用数据集上的测试结果也证明了本文方法的有效性.

参考文献

- [1] Deng W, Hu J, Lu J, et al. Transform-Invariant PCA: A unified approach to fully automatic face alignment, representation and recognition [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2014, 36(6): 1275-1284.
- [2] 周家锐, 纪震, 等. 基于 Gabor 小波与 Memetic 算法的人脸识别方法 [J]. 电子学报, 2012, 40(4): 642-646. Zhou Jia-rui, Ji Zhen, et al. Face recognition using Gabor wavelets and Memetic algorithm [J]. Acta Electronica Sinica, 2012, 40(4): 642-646. (in Chinese)
- [3] 杨飞, 苏剑波. 人脸显性特征的融合构造方法及识别 [J]. 电子学报, 2012, 40(3): 466-471. Yang Fei, Su Jian-bo. Face recognition based on explicit facial features by fusion construction method [J]. Acta Electronica Sinica, 2012, 40(3): 466-471. (in Chinese)
- [4] Braverman M. Poly-logarithmic independence fools bounded-depth boolean circuits [J]. Communications of the ACM, 2011, 54(4): 108-115.
- [5] Zhang H, Zhou F, Zhang W, et al. Real-time action recognition based on a modified Deep Belief Network model [A]. International Conference on Information and Automation [C]. Hailar, HulunBuir, Inner Mongolia, China: IEEE, 2014. 225-228.
- [6] Ouyang W, Wang X, Zeng X, et al. Deepid-net: Deformable deep convolutional neural networks for object detection [A]. Conference on Computer Vision and Pattern Recognition [C]. Boston, MA, USA: IEEE, 2015. 2403-2412.
- [7] BENGIO Y. Learning deep architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-12.

- [8] Bengio Y, Le Cun Y. Scaling learning algorithms towards AI[J]. Large-Scale Kernel Machines, 2007, 34(5): 321 – 359.
- [9] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection[A]. Conference on Computer Vision and Pattern Recognition[C]. Portland, OR, USA; IEEE, 2013. 3476 – 3483.
- [10] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527 – 1554.
- [11] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[A]. Conference on Computer Vision and Pattern Recognition[C]. Columbus, OH, USA; IEEE, 2014. 1701 – 1708.
- [12] Deng J, Zhang Z, Marchi E, et al. Sparse autoencoder-based feature transfer learning for speech emotion recognition[A]. Humaine Association Conference on Affective Computing and Intelligent Interaction[C]. Geneva, Switzerland; IEEE, 2013. 511 – 516.
- [13] Phan H T, Duong A T, Le N D H, et al. Hierarchical sparse autoencoder using linear regression-based features in clustering for handwritten digit recognition[A]. International Symposium on Image and Signal Processing and Analysis[C]. Melbourne, Australia; IEEE, 2013. 183 – 188.
- [14] Zhang C, Zhang Z. Improving multiview face detection with multi-task deep convolutional neural networks[A]. IEEE Winter Conference on Applications of Computer Vision[C]. Steamboat Springs, CO, USA; IEEE, 2014. 1036 – 1041.
- [15] Dahl G E, Yu D, Deng L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 30 – 42.
- [16] Hinton G E, Salakhutdinov R R. Replicated softmax: an undirected topic model[A]. Advances in Neural Information Processing Systems[C]. Vancouver, BC, Canada; NIPS, 2009. 1607 – 1614.
- [17] Heisele B, Ho P, Poggio T. Face recognition with support vector machines: Global versus component-based approach[A]. International Conference on Computer Vision[C]. Vancouver, Canada; IEEE, 2001. 688 – 694.

作者简介



李倩玉 女, 1991 年生于安徽合肥, 合肥工业大学研究生, 研究方向为数字图像处理与分析、人脸检测与识别。

E-mail: qianyu@mail.hfut.edu.cn



蒋建国 男, 1955 年生于安徽黄山, 教授、博士生导师, 合肥工业大学计算机与信息学院。研究方向为视频信息分析与处理、多媒体智能监控系统、分布式智能系统、DSP 技术应用等。

E-mail: jgjiang@hfut.edu.cn



齐美彬 男, 1969 年生于安徽池州, 教授、硕士生导师, 合肥工业大学计算机与信息学院。研究方向为运动目标检测与跟踪、视频编码等。

E-mail: qimeibin@163.com