

[文章编号]1000-1832(2016)02-0070-07

[DOI]10.16163/j.cnki.22-1123/n.2016.02.016

基于卷积神经网络的人脸识别方法

陈耀丹,王连明

(东北师范大学物理学院,吉林 长春 130024)

[摘 要] 研究并实现了一种基于卷积神经网络的人脸识别方法.该网络由 2 个卷积层、2 个池化层、1 个全连接层和 1 个 Softmax 回归层组成,它能自动提取人脸特征并进行分类.网络通过批量梯度下降法训练特征提取器和分类器,各隐层应用“dropout”方法解决了过拟合问题.应用于 ORL 和 AR 人脸数据库的人脸识别率分别达到 99.50%和 99.62%,识别单张人脸的时间均小于 0.05 s,而且对光照差异、面部表情变化、有无遮挡物等干扰具有鲁棒性.

[关键词] 人脸识别;卷积神经网络;图像处理;人脸数据库

[中图分类号] TP 391.4 [学科代码] 520·2040 [文献标志码] A

0 引言

人脸识别作为生物特征识别(虹膜识别、指纹识别、掌纹识别等)技术之一,由于具有成本低、用户易接受、可靠性高等优点,在身份认证、安防监控、人机交互等领域具有广泛的应用前景.^[1]目前人脸识别算法大致可分为两类:

(1) 基于表象(appearance-based)的方法.基本思想是将二维的人脸输入转换到另一个空间,然后用统计方法分析人脸模式,例如 eigenface^[2]、fisherface^[3]和 SVM^[4]等方法.

(2) 基于特征(feature-based)的方法.一般是提取局部或者全局特征,然后送入一个分类器进行人脸识别,例如基于几何特征的识别^[5]和 HMM^[6]等方法.

卷积神经网络是一个受生物视觉启发、以最简化预处理操作为目的的多层感知器的变形,本质是一个前向反馈神经网络.卷积神经网络与多层感知器的最大区别是网络前几层由卷积层和池化层交替级联组成,模拟视觉皮层中用于高层次特征提取的简单细胞和复杂细胞交替级联结构.卷积层的神经元对前一层输入的一部分区域(称为局部感受野,区域之间有重叠)有响应,提取输入的更高层次特征;池化层的神经元对前一层输入的一部分区域(区域之间无重叠)求平均值或者最大值,抵抗输入的轻微形变或者位移.卷积神经网络的后几层一般是若干个全连接层和一个分类器构成的输出层.近年来,卷积神经网络已经成功应用于字符识别^[7]、人脸识别^[8]、人体姿态估计^[9]和目标检测^[10].

卷积神经网络用于人脸识别是一种基于特征的方法.区别于传统的人工特征提取和针对特征的高性能分类器设计,它的优点是逐层卷积降维进行特征提取,然后经过多层非线性映射,使网络可以从未经特殊预处理的训练样本中,自动学习形成适合该识别任务的特征提取器和分类器.该方法降低了对训练样本的要求,而且网络层数越多,学习到的特征更具有全局性.

LeNet-5^[11]是 LeCun 提出的一个具有高识别率的用于手写数字识别的卷积神经网络.本文借鉴 LeNet-5 的模型结构,提出 6 层卷积神经网络用于 ORL 人脸库和 AR 人脸库的人脸识别.网络前 4 层由卷积层和池化层交替级联用于特征提取,随后接入一个全连接层,最后一层采用非线性分类能力强的 Softmax 分类器.该 6 层卷积神经网络采用 sigmoid 激活函数和固定的学习速率.

[收稿日期] 2015-05-17

[基金项目] 国家自然科学基金资助项目(21227008);吉林省科技发展计划项目(20130102028JC).

[作者简介] 陈耀丹(1993—),女,硕士研究生,主要从事模式识别研究;通讯作者:王连明(1972—),男,博士,教授,主要从事智能信息处理及嵌入式系统领域研究.

1 6层卷积神经网络模型

本文提出的6层卷积神经网络模型由2个卷积层、2个池化层、1个全连接层和1个Softmax回归层组成,如图1所示.卷积层和池化层由多个特征图(即二维平面)组成,每个特征图由多个神经元组成,每一层的特征图作为下一层的输入.卷积层的特征图可能与前一层的若干特征图建立关系.

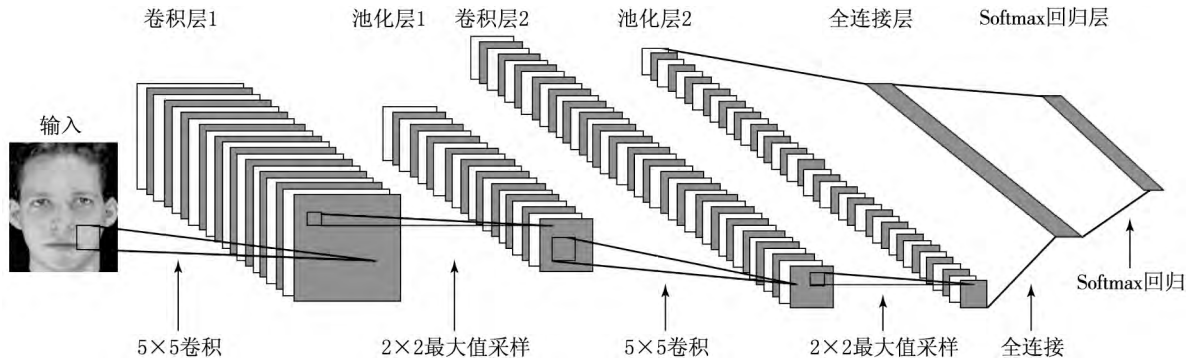


图1 6层卷积神经网络模型

1.1 卷积层

卷积层通过局部连接和权值共享的方法,模拟具有局部感受野的简单细胞,提取一些初级视觉特征的过程.局部连接指卷积层上的每个神经元与前一层特征图中固定区域的神经元建立连接;权值共享指同一特征图中的神经元用一组相同的连接强度与前一层局部连接,可以减少网络训练参数.上述一组相同的连接强度即为一个特征提取器,在运算过程中表现为一个卷积核,卷积核数值先随机初始化,最后由网络训练确定.

与LeNet-5相比,考虑到人脸识别比数字识别更复杂,卷积层为了更充分地提取人脸特征进行了以下改进:(1)增加2个卷积层的特征提取器——卷积核的数量;(2)卷积层2的特征图与池化层1的全部特征图均建立关系.

卷积层每个神经元的输入来自于前一层特征图中固定区域的神经元,区域的大小由卷积核大小 $K_x \times K_y$ 决定.卷积层1的 m 张特征图由1张输入图片跟 m 个可学习的卷积核卷积,加上偏置后,通过激活函数作用得到.卷积层2的 n 张特征图,由池化层1的 m 张特征图分别跟 $n \times m$ 个卷积核卷积,每 m 个卷积结果结合,再加上偏置,通过激活函数作用得到.卷积层操作如图2所示.

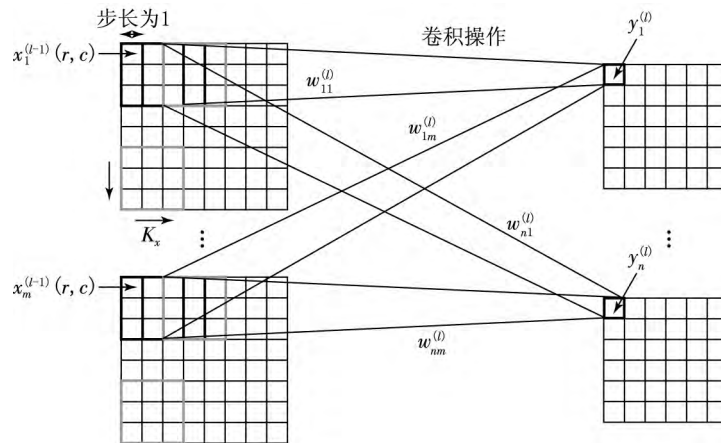


图2 卷积层操作

该层的数学表达式^[8]为

$$y_j^{(l)} = f\left(\sum_{i \in M_{j(l-1)}(u,v)} \sum_{(u,v) \in K^{(l)}} w_{ji}^{(l)} \times x_i^{(l-1)}(c+u, r+v) + b_j^{(l)}\right). \quad (1)$$

其中: $K^{(l)} = \{(u,v) \in \mathbb{N}^2 \mid 0 < u < K_x, 0 < v < K_y\}$; K_x 和 K_y 分别是卷积核的宽度和高度; l 表示当前层数; $b_j^{(l)}$ 是卷积层第 j 张特征图上神经元的偏置; $M_{j(l-1)}(u,v)$ 是与卷积层第 j 张特征图建立关系的前一层特征图的集合.

1.2 池化层

池化层模拟复杂细胞是将初级的视觉特征筛选并结合成更高级、抽象的视觉特征的过程,在网络中通过采样实现. 经过池化层的采样后,输出特征图的数量不变,但是特征图的尺寸会变小,有减少计算复杂度、抵抗微小位移变化的作用. 本文提出的池化层采用最大值采样,采样大小为 2×2 ,即把输入的特征图分割成不重叠的 2×2 大小的矩形,对每个矩形取最大值,所以输出特征图的长和宽均是输入特征图的一半. 本文定义池化层中的神经元不具备学习功能,数学表达式为

$$y_j^{(l)} = \text{down}(y_j^{(l-1)}). \quad (2)$$

其中 down 表示最大值采样的函数,该层运算不包含可学习的权值和阈值.

1.3 全连接层

为增强网络的非线性映射能力,同时限制网络规模的大小,网络在 4 个特征提取层提取特征后,接入一个全连接层. 该层的每一个神经元与前一层的所有神经元互相连接,同层神经元之间不连接. 数学表达式^[8]为

$$o_j^l = f\left(\sum_{i=1}^n x_i^{(l-1)} \cdot w_{ji}^{(l)} + b_j^{(l)}\right). \quad (3)$$

其中: n 是前一层的神经元个数; l 表示当前层数; $w_{ji}^{(l)}$ 是该层神经元 j 与前一层神经元 i 的连接强度; $b_j^{(l)}$ 是该层神经元 j 的偏置; $f(\cdot)$ 表示激活函数.

1.4 Softmax 回归层

LeNet-5 最后一层是 RBF 分类器,采用由 ASCII 码表中的数字作为模板算出的纠错码或者以等概率的方式任选 -1 和 1 作为 RBF 的权值,每个 RBF 输出单元计算输入特征和参数向量之间的欧式距离. 但是人脸特征比数字特征更复杂,而且人脸类别多又没有统一的模板,所以网络最后一层采用非线性分类能力强的 Softmax 回归作为分类器.

假定可分为 k 类的 m 个样本组成训练集 $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$, 样本 $x^{(i)} \in R^{n+1}$, 类标 $y^{(i)} \in \{1, 2, \dots, k\}$, n 是样本维数, Softmax 回归假设函数^[12]为

$$h_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} = \begin{bmatrix} \exp(\theta_1^T x^{(i)}) \\ \vdots \\ \exp(\theta_k^T x^{(i)}) \end{bmatrix} \bigg/ \sum_{j=1}^k \exp(\theta_j^T x^{(i)}). \quad (4)$$

其中: $p(y^{(i)} = j | x^{(i)}; \theta)$ 代表样本 $x^{(i)}$ 属于第 j 类的概率; $\theta_j^T \in R^{n+1}$ 代表模型参量. 将 $\theta_1, \theta_2, \dots, \theta_j$ 按行写成矩阵形式 $\theta \in R^{k(n+1)}$.

$$\text{模型的代价函数为} \quad J(\theta) = \frac{-1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{t=1}^k e^{\theta_t^T x^{(i)}}} \right]. \quad (5)$$

其中: $1\{\cdot\}$ 为示性函数,若 $\{\cdot\}$ 内值为真的表达式,则示性函数值为 1;反之,则为 0. 本文通过批量梯度下降法求解使代价函数 $J(\theta)$ 最小化时的 θ .

2 网络训练

2.1 数据预处理

ORL 人脸库由 40 人的 400 张图片组成,即每人 10 张图片,人脸有表情、微小姿态等变化. 取 AR 人脸库的一个子集,由男、女各 50 人组成的 2 600 张图片,即每人的图片为 26 张,人脸有表情、遮挡、光照等变化. 图 3 是 ORL 人脸库中人脸样例,图 4 是 AR 人脸库中人脸样例.



图 3 ORL 人脸库人脸样例



图 4 AR 人脸库人脸样例

对2个数据库同做以下处理:随机抽取库中90%的人脸作为训练集,剩下的10%的人脸作为测试集,然后标准化2个集合中的人脸,如图5所示。



图5 人脸标准化过程

“min-max”标准化灰度值,即对图片中每个像素点的灰度值标准化为 $[0,1]$ 。用 x 和 x' 分别表示当前和标准化后的灰度值,min和max分别表示图片中的最小和最大的灰度值,标准化公式为

$$x' = \frac{x - \min}{\max - \min}. \quad (6)$$

2.2 网络训练算法

由于2个人脸库训练样本较多,本文采用的是在实践中收敛速度较快的批量随机梯度下降法。对于ORL人脸库的卷积神经网络的训练,批处理块大小为40,动量为0.9,学习速率恒为0.12;对于AR人脸库的卷积神经网络的训练,批处理块大小为65,动量为0.9,学习速率恒为0.15。每次迭代会遍历训练集的所有批处理块,遍历完一个批处理块更新一次网络参数。更新公式为

$$w_{i+1} = \varepsilon \cdot w_i - \eta \cdot \left(\frac{\partial L}{\partial w_i} \right)_{D_i}. \quad (7)$$

其中: w_i 是当前参数, w_{i+1} 是更新后的参数, ε 是动量, η 是学习速率, $\left(\frac{\partial L}{\partial w_i} \right)_{D_i}$ 是第 i 个批处理块 D_i 中误差对 w_i 偏导的平均值。

本文用于ORL人脸库的某个卷积神经网络包含 104.210×10^4 个待训练参数;用于AR人脸库的某个卷积神经网络包含 150.266×10^4 个待训练参数。针对参数远大于训练样本数量可能导致的过拟合问题,Hinton提出了“dropout”^[13],即在网络训练时以一定的概率 $p(0 \leq p \leq 1)$ 将输入层或者隐层的神经元输出置为0,被置为0的神经元不再参与网络的前向传播和误差反向传播,而在测试时,将该输入层或者隐层的神经元输出乘以 p 作为该层的输出。由于每输入一张图片网络中神经元随机被置0,所以就要求隐层神经元必须提取图片中最本质的特征,因此可以用“dropout”来解决过拟合问题。

3 实验结果

本文采用Windows XP下的Matlab 2010a作为实验环境,计算机CPU为3.3 GHz的Intel i3-3220,内存为4 GB,2个卷积层中卷积核大小为 5×5 ,2个采样层采用最大值池化,采样大小为 2×2 ,激活函数采用sigmoid函数,即

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (8)$$

网络权值采用高斯初始化,网络阈值初始化为0,网络的代价函数 $J(\theta)$ 见公式(5),卷积层、全连接层和Softmax回归层均采用“dropout”方法,概率 p 分别设为0.5,0.2和0.5。

6层卷积神经网络模型可简约描述为“ $a-b-c$ ”, a 表示卷积层1有 a 个卷积核,经过该层得到 a 张特征图,再经过池化层1采样得到缩小的 a 张特征图; b 表示卷积层2有 $b \times a$ 个卷积核,经过该层得到 b 张特征图,再经过池化层2采样得到缩小的 b 张特征图; c 表示全连接层的神经元个数。输出层神经元个数是人脸库的模式类个数。本文分别改变全连接层神经元个数 c 和2个卷积层的特征图数量 a 和 b ,找到最适合2人脸库的网络模型。

3.1 改变全连接层神经元个数对网络的影响

ORL人脸库的6层卷积神经网络用20-40-300,20-40-500,20-40-800,20-40-1 100,20-40-1 500模型测试。AR人脸库的6层卷积神经网络用20-40-800,20-40-1 000,20-40-1 500,20-40-2 000模型测试。

图6描述ORL人脸库中每次迭代对应的代价函数 $J(\theta)$ 的值,图7描述AR人脸库中每次迭代对应的代价函数 $J(\theta)$ 的值。图6和7显示各个网络模型从迭代次数为1 900至2 000时基本训练。选取迭代次数

为 2 000 时,测量各个模型对于相应人脸库训练集和测试集的正确识别率,结果如表 1 和 2 所示.

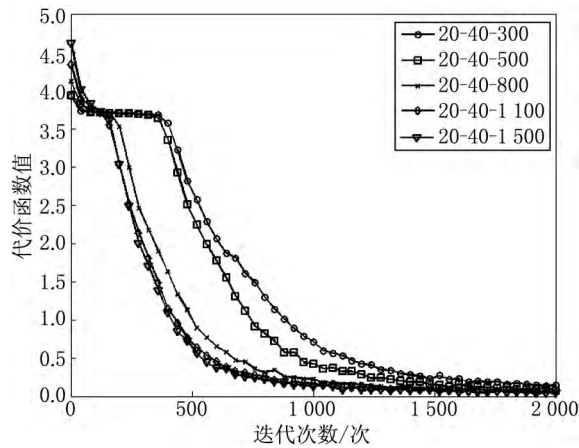


图 6 改变全连接层的 ORL 人脸库的 $J(\theta)$

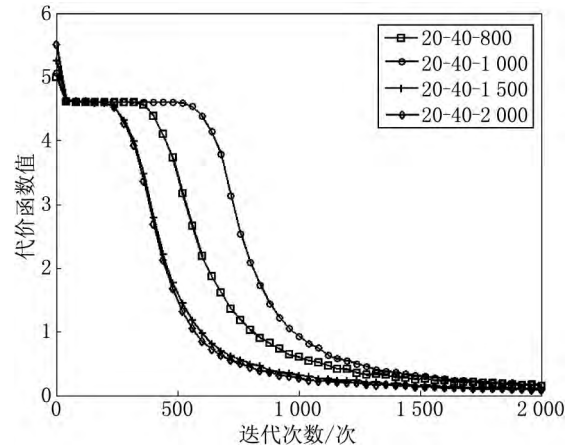


图 7 改变全连接层的 AR 人脸库的 $J(\theta)$

表 1 改变全连接层的 ORL 人脸库的正确识别率 %

模型	20-40-300	20-40-500	20-40-800	20-40-1 100	20-40-1 500
训练集识别率	98.33	99.72	99.44	99.44	99.72
测试集识别率	97.50	97.50	97.50	97.50	97.50

表 2 改变全连接层的 AR 人脸库的正确识别率 %

模型	20-40-800	20-40-1 000	20-40-1 500	20-40-2 000
训练集识别率	99.87	99.74	100.00	100.00
测试集识别率	94.62	94.62	95.77	96.15

对于 ORL 人脸库,最佳的卷积神经网络模型为 20-40-1 500;对于 AR 人脸库,最佳的卷积神经网络模型为 20-40-2 000. 增加全连接层神经元个数可以加快网络的训练速度,在一定程度内能提高正确识别率.

3.2 改变 2 个卷积层特征图数量对网络的影响

ORL 人脸库的 6 层卷积神经网络用 10-15-1 500,15-15-1 500,15-25-1 500, 20-30-1 500, 20-40-1 500 模型测试. AR 人脸库的 6 层卷积神经网络用 15-25-2 000,15-30-2 000,20-30-2 000,20-40-2 000 模型测试.

图 8 描述 ORL 人脸库中每次迭代对应的代价函数 $J(\theta)$ 的值,图 9 描述 AR 人脸库中每次迭代对应的代价函数 $J(\theta)$ 的值. 同理,选取迭代次数为 2 000 时,测量各个模型对于相应人脸库训练集和测试集的正确识别率,结果如表 3 和 4 所示.

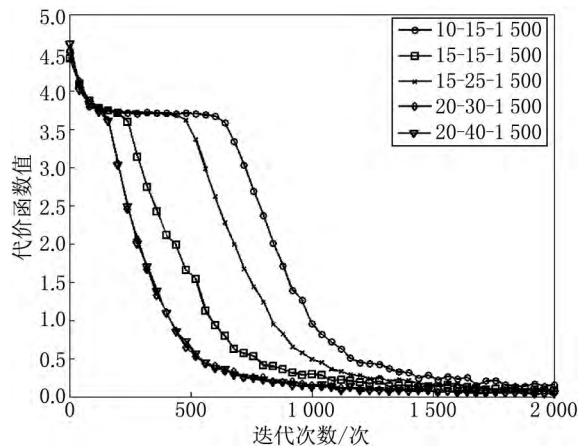


图 8 改变 2 个卷积层的 ORL 人脸库的 $J(\theta)$

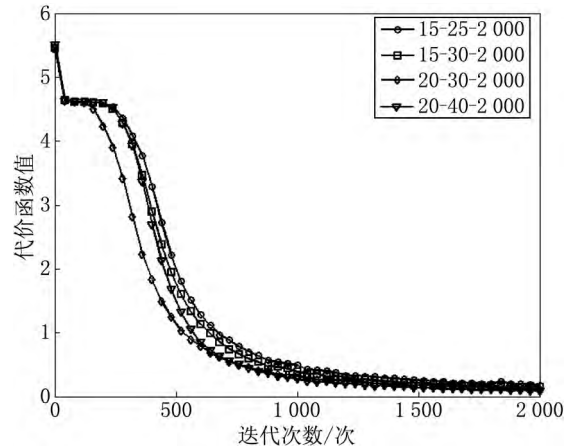


图 9 改变 2 个卷积层的 AR 人脸库的 $J(\theta)$

表 3 改变 2 个卷积层的 ORL 人脸库的正确识别率

%

模型	10-15-1 500	15-15-1 500	15-25-1 500	20-30-1 500	20-40-1 500
训练集识别率	93.06	98.89	96.67	99.17	99.72
测试集识别率	87.50	92.50	90.00	95.00	97.50

表 4 改变 2 个卷积层的 AR 人脸库的正确识别率

%

模型	15-25-2 000	15-30-2 000	20-30-2 000	20-40-2 000
训练集识别率	99.62	99.96	99.96	100.00
测试集识别率	92.69	95.77	96.15	96.15

对于 ORL 人脸库,最佳的卷积神经网络模型为 20-40-1 500;对于 AR 人脸库,最佳的卷积神经网络模型为 20-40-2 000.适当增加卷积层特征图数量可以提高正确识别率.

3.3 20-40-1 500 卷积神经网络模型与其他算法的比较

20-40-1 500 的卷积神经网络模型对 ORL 人脸库全部样本的识别率为 99.50%,20-40-2 000 的卷积神经网络模型对 AR 人脸库全部样本的识别率为 99.62%.表 5 是本文提出的网络与其他算法在 ORL 库识别率的比较,表 6 是本文提出的网络与其他算法在 AR 库识别率的比较.可以看出本文提出的网络在 2 个人脸库上的识别率较好,并且网络抵抗光照差异、面部表情变化、有无遮挡物等干扰的能力较强.

表 5 20-40-1 500 卷积神经网络模型
与用于 ORL 库的其他算法比较

识别方法	正确识别率/%
Eigenface ^[14]	97.50
Fisherface ^[14]	98.50
ICA ^[14]	93.75
2DPCA ^[15]	98.30
20-40-1 500 CNN	99.50

表 6 20-40-1 500 卷积神经网络模型
与用于 AR 库的其他算法比较

识别方法	正确识别率/%
PCA ^[16]	85.53
2DPCA ^[15]	<96.10
PCA+GSRC ^[17]	97.14
LC-KSVD ^[18]	97.80
20-40-2 000 CNN	99.62

4 总结

本文提出了一个 6 层卷积神经网络用于 ORL 和 AR 人脸库的人脸识别,对测试样本的识别率分别为 99.75%和 96.15%,对全部样本的识别率分别为 99.50%和 99.62%.用于 AR 人脸库的卷积神经网络对光照差异、面部表情变化、有无遮挡物等干扰具有鲁棒性.采用 Matlab 2010a 作为实验环境,在 CPU 为 3.3 GHz 的 Intel i3-3220,内存为 4 GB 的计算机上,ORL 和 AR 人脸库的单张人脸识别时间均小于 0.05 s,识别实时性好,为人脸识别提供一种新思路.

[参 考 文 献]

- [1] 陈海霞,崔茜.基于 Gabor 小波和 PCA 的人脸识别[J].东北师大学报(自然科学版),2014,46(4):77-80.
- [2] 孟繁静,王连明.视频人脸识别系统中的最佳人脸捕获[J].东北师大学报(自然科学版),2015,47(3):90-95.
- [3] WANG X, TANG X. Dual-space linear discriminant analysis for face recognition[C]//CVPR 2004 Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, IEEE, 2004, 2: 564-569.
- [4] HEISELE B, HO P, POGGIO T. Face recognition with support vector machines: Global versus component-based approach[C]//ICCV 2001 Proceedings, Eighth IEEE International Conference on Computer Vision, Los Alamitos, IEEE, 2001, 2: 688-694.
- [5] ABATE A F, NAPPI M, RICCIO D, et al. 2D and 3D face recognition a survey[J]. Pattern Recognition Letters, 2007, 28(14): 1885-1906.
- [6] NEFIAN A V, HAYES III M H. Hidden markov models for face recognition[C]//1998 IEEE International Conference on Acoustics, Speed and Signal Processing, Washington, IEEE, 1998: 2721-2724.
- [7] AHRANJANY S S, RAZZAZI F, GHASSEMIAN M H. A very high accuracy handwritten character recognition system for Farsi/Arabic digits using Convolutional Neural Networks[C]//Theories and Applications (BIC-TA), 2010 IEEE Fifth

- International Conference on Bio-Inspired Computing, Beijing:IEEE,2010:1585-1592.
- [8] SYAFEEZA A R, KHALIL-HANI M, LIEW S S, et al. Convolutional neural network for face recognition with pose and Illumination Variation[J]. International Journal of Engineering & Technology, 2014, 6(1): 44-57.
- [9] TOSHEV A, SZEGEDY C. Deeppose: Human pose estimation via deep neural networks[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos:IEEE, 2014: 1653-1660.
- [10] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat integrated recognition, localization and detection using convolutional networks[J]. Neural Networks, 2003, 16(5): 555-559.
- [11] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [12] HUANG V S, SHADMEHR R, DIEDRICHSEN J. Active learning: learning a motor skill without a coach[J]. Journal of neurophysiology, 2008, 100(2): 879-887.
- [13] SRIVASTAVA S, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [14] YANG M H. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods[C]//2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Washington: IEEE, 2002: 0215-0215.
- [15] YANG J, ZHANG D, FRANGI A F, et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2004, 26(1): 131-137.
- [16] ROLI F, MARCIALIS G L. Semi-supervised PCA-based face recognition using self-training[M]. Structural, Syntactic, and Statistical Pattern Recognition. Berlin Heidelberg: Springer, 2006: 560-568.
- [17] YANG M, ZHANG L. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary[M]. Computer Vision – ECCV 2010. Berlin Heidelberg: Springer, 2010: 448-461.
- [18] JIANG Z, LIN Z, DAVIS L S. Learning a discriminative dictionary for sparse coding via label consistent K-SVD[C]//2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington: IEEE, 2011: 1697-1704.

Convolutional neural network for face recognition

CHEN Yao-dan, WANG Lian-ming

(School of Physics, Northeast Normal University, Changchun 130024, China)

Abstract: Feature extraction and classification are two key steps in face recognition. A convolutional neural network composed of two convolutional layers, two pooling layers, one full-connection layer and one Softmax regression layer for face recognition is proposed. The neural network can automatically extract facial features and classify face, with trained feature extractors and the classifier using batch gradient descent. And the network adopts dropout method in hidden layers to avoid the overfitting problem. Experimental results show that proposed network achieves 99.50% recognition accuracy on ORL database and an accuracy of 99.62% on AR database, and it can complete one-time facial recognition in less than 0.05s. More importantly, the network is robust to illumination variances, facial expressions and occlusions.

Keywords: face recognition; convolutional neural network; image processing; face database

(责任编辑:石绍庆)