

基于深度神经网络的特征加权融合人脸识别方法

孙劲光^{1,2}, 孟凡宇^{1*}

(1. 辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125000; 2. 辽宁省数字化矿山装备工程技术研究中心, 辽宁 阜新 123000)

(* 通信作者电子邮箱 mengfanyu1991@163.com)

摘 要: 针对目前难以提取到适合用于分类的人脸特征以及在非限制条件下进行人脸识别准确率低的问题, 提出了一种基于深度神经网络的特征加权融合人脸识别方法(DLWF)。首先, 应用主动形状模型(ASM)提取出人脸面部的主要特征点, 并根据主要特征点对人脸不同器官区域进行采样; 然后, 将所得采样块分别输入到对应的深度信念网络(DBN)中进行训练, 获得网络最优参数; 最后, 利用 Softmax 回归求出各个区域的相似度向量, 将多区域的相似度向量加权融合得到综合相似度评分进行人脸识别。经 ORL 和 WFL 人脸库上进行实验验证, DLWF 算法的识别准确率分别达到 97% 和 88.76%, 与传统算法主成分分析(PCA)、支持向量机(SVM)、DBN 及 FIP + 线性判别式分析(LDA)相比, 无论是限制条件还是非限制条件下, 识别率均有提高。实验结果表明, 该算法具有高效的人脸识别能力。

关键词: 人脸识别; 非限制条件; 深度信念网络; 加权融合; 主动形状模型

中图分类号: TP391.41 **文献标志码:** A

Face recognition based on deep neural network and weighted fusion of face features

SUN Jinguang^{1,2}, MENG Fanyu^{1*}

(1. School of Electronic and Information Engineering, Liaoning Technical University, Huludao Liaoning 125000, China;

2. Liaoning Digital Mining Equipment Engineering Technology Research Center, Fuxin Liaoning 123000, China)

Abstract: It is difficult to extract suitable face feature for classification, and the face recognition accuracy is low under unconstrained condition. To solve the above problems, a new method based on deep neural network and weighted fusion of face features, namely DLWF, was proposed. First, facial feature points were located by using Active Shape Model (ASM), then different organs of face were sampled according to those facial feature points. The corresponding Deep Belief Network (DBN) was trained by the regional samples to get optimal network parameters. Finally, the similarity vector of different organs was obtained by using Softmax regression. The weighted fusion of multiple regions in the similarity vector method was used for face recognition. The recognition accuracy got to 97% and 88.76% respectively on the ORL and LFW face database; compared with the traditional recognition algorithm including Principal Components Analysis (PCA), Support Vector Machine (SVM), DBN, and Face Identity-Preserving (FIP) + Linear Discriminant Analysis (LDA), no matter under the constrained condition or the unconstrained condition, recognition rates were both improved. The experimental results show that the proposed algorithm has high efficiency in face recognition.

Key words: face recognition; unconstrained condition; Deep Belief Network (DBN); weighted fusion; Active Shape Model (ASM)

0 引言

人脸识别一直是计算机视觉与模式识别领域的研究热点。未经降维的人脸图像特征维度较高, 进行正确识别难度大, 所以设计有效的人脸图像特征提取器成为传统人脸识别方法的关键。最典型的方法为主成分分析(Principal Components Analysis, PCA)法^[1], 其基本思想为先把人脸图像按行或列拉直成一个高维列向量, 对整个训练样本的高维列向量构造出协方差矩阵, 通过求解协方差矩阵的特征向量得到变换矩阵, 再把人脸的高维数据特征通过变换矩阵投影到低维空间, 从而得到人脸图像的低维特征。在人脸识别中, 应用 PCA 对特征进行降维处理, 取得了良好的效果, 并被认为是成功的线性鉴别分析方法之一。但是实际应用中, 通

常样本个数远远小于模式特征的维数, 应用 PCA 算法进行线性鉴别分析会产生小样本问题(即样本数远远小于特征维数)。当人脸图像在非限制条件下提取, 或者人脸面部姿态具有较大变化时, 应用 PCA 算法提取的特征并不一定适合作为分类、识别的特征。

随着深度学习模型不断成熟, 人们在利用深度神经网络进行人脸识别时, 主要利用深度神经网络进行特征提取再利用不同分类器进行识别^[2-6], 或者直接利用深度神经网络进行识别^[7-8]。文献[9]中利用卷积神经网络进行人脸验证, 在 LFW 人脸库上已经达到 99.15% 的识别率; 该方法解决了人脸识别中的二分类问题, 仍需解决照片中人物是谁的多分类问题。

为了能够提取到分类特征好的局部人脸抽象特征, 并且

收稿日期: 2015-08-29; 修回日期: 2015-09-19。 基金项目: 国家科技支撑计划项目(2013BAH12F02)。

作者简介: 孙劲光(1962-), 女, 辽宁阜新新人, 教授, 博士生导师, 博士, CCF 会员, 主要研究方向: 计算机图像处理、计算机图形学、知识工程; 孟凡宇(1991-), 男, 辽宁鞍山人, 硕士研究生, 主要研究方向: 计算机图像处理。

能够对面脸的多姿态和复杂背景下识别表现出鲁棒性,本文提出了一种基于深度神经网络的特征加权融合人脸识别方法。主要的解决思路是首先找出人脸面部的主要特征点,进行面部器官区域划分,利用各区域构建的深度神经网络分别求出各区域的隶属度向量,最后对隶属度向量进行加权求和出最终识别结果。本文的主要工作是:分别构建人脸器官区域的深度神经网络模型,求取不同区域的分类隶属度向量;通过对隶属度加权融合进行人脸识别,提出了基于相似度加权评分的深度网络模型(Deep Learning and Weighted Fusion, DLWF)人脸识别方法。

1 基于相似度加权评分深度网络模型

1.1 问题描述

传统深度信念网络(Deep Belief Network, DBN)^[10-11]在进行人脸识别时,输入整张人脸图像的像素级特征,忽略了图像的局部结构,使网络难以学习到人脸图像局部特征^[12];由于人脸的非刚体特性,当人脸面部表情变化时,面部各个器官都会发生不同程度的形变。在自然条件下,通常取得的人脸图像都受光照、背景等其他因素不同程度的影响,很难区分哪些面部器官在人脸识别过程中起的作用比重大,哪些器官起的作用比重小,使得在实际工作中相似度加权的权值确定变得更加复杂。因此本文重点解决适用于提取人脸局部特征的深度学习模型及多特征加权识别方法两个问题。

1.2 基本思想

DLWF 分为以下四个主要的阶段:

1) 提取人脸特征点。利用主动形状模型算法(Active Shape Model, ASM)^[13]提取出人脸面部的主要特征点:眼角、眼睛中心、嘴角、鼻尖、下巴轮廓等。

2) 划分人脸子区域。利用提取的特征点确定面部器官的位置,根据各个器官大小选择合适像素大小的采样窗,提取出各子区域的采样块。

3) 求取面部区域隶属度向量。将各子区域的采样块输入到不同器官所对应的 DBN 中,求出表现每一器官隶属度的相似度概率向量。

4) 加权求和得出判别结果。将相似度概率加权求和得出对应于每个人的综合相似度概率,用于判别测试样本的类别归属。

1.3 基本原理

1.3.1 主动形状模型

主动形状模型方法是由 Cootes 等^[13-15]在 1995 年提出的一种基于统计学、灰度和形状可分离的可变形模型。该模型主要由全局形状模型和局部纹理模型构成。

1) 全局形状模型。

ASM 用点分布模型来表示一个人脸形状 $S_i = (x_1, y_1, x_2, y_2, \dots, x_n, y_n)^T$ (n 为标定特征点的数目),包含 N 个训练样本的训练集记为 $\Omega = \{S_1, S_2, \dots, S_N\}$ 。

训练样本对齐后,用主成分分析(PCA)方法建立形状模型,过程如下:

a) 计算对齐后训练样本的平均形状:

$$\bar{S} = \frac{1}{N} \sum_{i=1}^N S_i$$

其中: S_i 为一个人脸的形状, n 为训练样本个数。

b) 计算对齐后形状向量的协方差矩阵:

$$C = \frac{1}{N-1} \sum_{i=1}^N (S_i - \bar{S})(S_i - \bar{S})^T$$

c) 对协方差矩阵 C 进行特征分解,求得特征值 λ_k ($\lambda_k \geq \lambda_k + 1, \lambda_k \neq 0, k = 1, 2, \dots, 2n$) 和对应的特征向量 P_k 。

d) 取前 t 个特征值 $[\lambda_1, \lambda_2, \dots, \lambda_t]$ 使得:

$$\sum_{i=1}^t \lambda_i / \sum_{i=1}^{2n} \lambda_i \geq \eta$$

其中: η 为所选特征占总特征的比例,一般为 95% ~ 98%。求得相应的特征向量 $P = [p_1, p_2, \dots, p_t]$,最终获得全局形状模型为:

$$S = \bar{S} + Pb$$

其中 b 为主分量参数,用来控制前 t 个模式的系数,且 b_i (b 的分量)的取值为:

$$-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i}$$

统计形状模型的含义是:通过控制相对低维的形状参数,可以在平均形状的基础上叠加主成分特征向量方向上的形变,获得新的形状向量。控制形状参数的范围,可以保证新形状向量描述的形状与训练样本形状类似。

全局形状模型的作用,主要是在搜索更新的过程中改变形状参数找到最佳匹配模型,并将形状的变化约束在模型的允许范围内。

2) 局部纹理模型。

主动形状模型的局部纹理模型构建过程如下:

a) 对第 i 个训练样本的第 j 个标定点,以该点为中心,在与相邻两点连线垂直的方向上一定长度范围(也称 Profile 邻域)内外各取 k 个点的灰度作为该点的灰度信息,记为:

$$g_{ij} = [g_{ij,1}, g_{ij,2}, \dots, g_{ij,2k+1}]$$

b) 求 g_{ij} 的梯度 dg_{ij} :

$$dg_{ij} = [g_{ij,2} - g_{ij,1}, \dots, g_{ij,2k+1} - g_{ij,2k}]$$

对 dg_{ij} 进行归一化处理得:

$$G_{ij} = \frac{1}{\sum_{l=1}^{2k} |dg_{ij,l}|} dg_{ij}$$

其中: $dg_{ij,l} = g_{ij,l(l+1)} - g_{ij,l}$ 。

对训练样本集中所有图像上的第 j 个特征点进行相同的操作,可得到第 j 个特征点的 N 个局部纹理 $G_{1j}, G_{2j}, \dots, G_{Nj}$ 。

c) 最终求得标记点 j 对应的平均纹理 \bar{G}_j 和协方差矩阵 C_{gj} ,即为该点的局部纹理模型。

$$\bar{G}_j = \frac{1}{N} \sum_{i=1}^N G_{ij}$$

$$C_{gj} = \frac{1}{N} \sum_{i=1}^N (G_{ij} - \bar{G}_j)(G_{ij} - \bar{G}_j)^T$$

局部纹理模型的作用,主要是在搜索更新的过程中,确定各个特征点的最佳候选定位点。对特征点的 Profile 邻域内每一个候选点,均计算其局部纹理与训练样本局部纹理之间的马氏距离。马氏距离 $d(G_j')$ 最小的候选点,就是该特征点的最佳候选点。

$$d(G_j') = (G_j' - \bar{G}_j)^T C_{gj}^{-1} (G_j' - \bar{G}_j)$$

其中 G_j' 是由未知图像 j 点附近采样得到的归一化纹理向量。

3) 搜索过程。

主动形状模型的搜索过程如下:

a) 加载平均形状作为初始形状 $S_{i=0}$ 。

b) 对 S_i 的每个特征点, 在其轮廓法线的两边各搜索 m 个像素点, 将马氏距离最小的点作为该特征点的新位置, 得到 S_i' 。

c) 调整伸缩、旋转和平移等变换参数及形状参数 (s, θ, S_c, b) , 使得模型形状 $M(s, \theta)[S] + S_c$ 与形状 S_i' 最佳匹配, 则当前形状 $S_{i+1} = M(s, \theta)[S] + S_c$ 。其中: $S_c = (x_c, y_c, x_c', y_c', \dots, x_c'', y_c'')^T$, (x_c, y_c) 是图像模型的中心位置。

d) 如果 S_i 和 S_{i+1} 之间的差别足够小或迭代次数达到了 N_{\max} 则停止, 否则转到 b)。

在本文当中利用上述模型方法提取的 77 个面部特征点如图 1 所示。



图1 人脸面部特征点

1.3.2 深度信念网络

DBN 是典型的深度学习网络, 它模拟人的大脑组织结构, 能够由低级到高级地提取输入数据的特征^[16]。DBN 由多个受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM) 堆叠而成, RBM 作为一个系统在一组给定状态 (v, h) 下的能量定义为:

$$E(v, h; \theta) = - \sum_{i=1}^{n_v} a_i v_i - \sum_{j=1}^{n_h} b_j h_j - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} w_{ij} v_i h_j \quad (1)$$

其中: n_v, n_h 分别表示可见层和隐含层中包含的神经元数目, v_i 表示可见层第 i 个神经元的状态, h_j 表示隐含层第 j 个神经元的状态, a_i 表示可见层中第 i 个神经元的偏置, b_j 表示隐含层中第 j 个神经元的偏置, w_{ij} 表示可见层中第 i 个神经元与隐含层中第 j 个神经元之间的连接权重, $\theta = \{a, b, w\}$ 为需要训练调整的参数。

用矩阵-向量的形式表示为:

$$E(v, h; \theta) = -a^T v - b^T h - h^T w v$$

当模型参数 $\theta = \{a, b, w\}$ 确定时, 应用式(1) 定义的能量函数, 可以给出可见层 v 和隐含层 h 的联合概率分布:

$$p(v, h; \theta) = \frac{1}{Z_\theta} e^{-E(v, h; \theta)}$$

$$Z_\theta = \sum_{v, h} e^{-E(v, h; \theta)}$$

其中 Z_θ 为归一化因子或称配分函数。

当可视层节点确定时, 隐含层节点 h_i 为 1 或者 0 的条件概率分布公式为:

$$\begin{cases} p(h_i = 1 | v) = \sigma(a_i + \sum_j w_{ij} v_j) \\ p(h_i = 0 | v) = 1 - p(h_i = 1 | v) \end{cases}$$

当隐含层节点确定时, 可视层节点 v_i 为 1 或者 0 的条件概率分布公式为:

$$\begin{cases} p(v_i = 1 | h) = \sigma(b_i + \sum_j w_{ij} h_j) \\ p(v_i = 0 | h) = 1 - p(v_i = 1 | h) \end{cases}$$

第 k 层和第 $k+1$ 层的隐藏单元满足:

$$\begin{cases} p(h^{(k)} | h^{(k+1)}) = \prod_i p(h_i^{(k)} | h^{(k+1)}) \\ p(h_i^{(k)} = 1 | h^{(k+1)}) = \sigma(b_i^{(k)} + \sum_j w_{ij}^{(k)} h_j^{(k+1)}) \end{cases}$$

其中: $\sigma(x) = 1/(1 + e^{-x})$, $b_i^{(k)}$ 表示第 k 层偏置, $w_{ij}^{(k)}$ 为第 k 层和第 $k+1$ 层之间的权值。在 DBN 中将 $p(h^{(l-1)} | h^{(l)})$ 视为一个 RBM 模型, 其结构如图 2 所示。

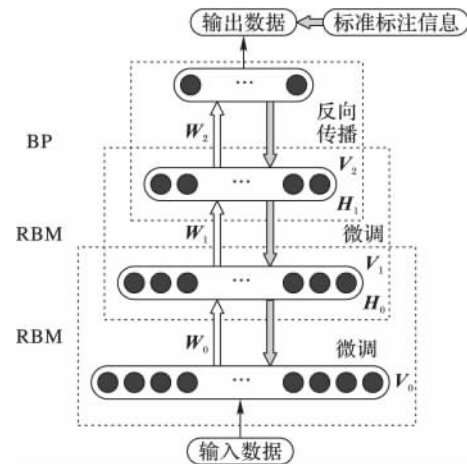


图2 DBN 结构模型

1.3.3 Softmax 回归

Softmax 回归是 Logistic 回归的一般化, 将只能解决二分类问题的 Logistic 回归扩展至能够解决多分类问题。

假定训练 Softmax 回归模型的样本来自 k 个不同的类, 共 m 个, 则这 m 个训练样本组成的训练集为 $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$, 其中第 i 个训练样本为 $x^{(i)} \in \mathbf{R}^{n+1}$, 其标签为 $y^{(i)} \in \{1, 2, \dots, k\}$, 则 Softmax 回归的假设为:

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

其中: 假设向量 $h_\theta(x^{(i)})$ 每一个元素 $p(y^{(i)} = j | x^{(i)}; \theta)$ 代表样本 $x^{(i)}$ 属于第 j 类的概率, $\theta_1, \theta_2, \dots, \theta_k$ 为模型参数向量, 其中 $\theta_i \in \mathbf{R}^{n+1}$, 将这些向量写成矩阵形式可得:

$$\theta = \begin{bmatrix} \theta_1^T \\ \theta_2^T \\ \vdots \\ \theta_k^T \end{bmatrix}$$

则模型的代价函数定义如下:

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] +$$

$$\frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2$$

其中: $1\{\}$ 为指示函数, 如果花括号之中的表达式为真, 那么指示函数值为 1, 否则指示函数值为 0; 加号后面的部分是为了防止模型对训练样本过拟合的权衰减, 其中 λ 是模型的参

数,需要在实验部分确定。分析可知无法直接求得使 $J(\theta)$ 最小的 θ 的解析解,可以通过迭代优化算法来最小化求解。通过计算,代价函数的梯度为:

$$\nabla_{\theta_j} = -\frac{1}{m} \sum_{i=1}^m [x^{(i)} (1\{y^{(i)} = j\} - p(y^{(i)} = j | x^{(i)}; \theta))] + \lambda \theta_j$$

1.3.4 双线性内插法

对于一个给定目标像素 (x, y) ,通过逆投影得到浮点坐标为 $(i + u, j + v)$ ($i \in \mathbf{Z}^+, j \in \mathbf{Z}^+, u \in [0, 1], v \in [0, 1]$),则该像素值 $f(i + u, j + v)$ 可由源图中坐标为 (i, j) 、 $(i + 1, j)$ 、 $(i, j + 1)$ 、 $(i + 1, j + 1)$ 所对应的4个像素的值确定。以 $f(i, j)$ 表示源图坐标 (i, j) 处的像素值:

$$f(x, y) = f(i + u, j + v) = (1 - u)(1 - v)f(i, j) + (1 - u)vf(i, j + 1) + u(1 - v)f(i + 1, j) + uvf(i + 1, j + 1)$$

1.4 本文算法 DLFW

1.4.1 面部区域划分

本文利用主动形状模型确定人脸面部的主要77个特征点,由此对于每一张人脸面部图像可以得到一个特征点位置向量:

$$x_i = (x_{i,0}, y_{i,0}, x_{i,1}, y_{i,1}, \dots, x_{i,(n-1)}, y_{i,(n-1)})$$

其中: $(x_0, y_0) \sim (x_{0+12}, y_{0+12})$ 为脸部轮廓; $(x_{13}, y_{13}) \sim (x_{13+2}, y_{13+2})$ 为额头轮廓; $(x_{16}, y_{16}) \sim (x_{16+5}, y_{16+5})$ 为左边眉毛轮廓; $(x_{22}, y_{22}) \sim (x_{22+6}, y_{22+6})$ 为右边眉毛轮廓; $(x_{29}, y_{29}) \sim (x_{29+9}, y_{29+9})$ 为左眼轮廓,其中 (x_{38}, y_{38}) 为左眼中心点; $(x_{39}, y_{39}) \sim (x_{39+8}, y_{39+8})$ 为右眼轮廓,其中 (x_{39}, y_{39}) 为右眼中心点; $(x_{47}, y_{47}) \sim (x_{47+11}, y_{47+11})$ 为鼻子的轮廓,其中 (x_{52}, y_{52}) 为鼻尖; $(x_{59}, y_{59}) \sim (x_{59+17}, y_{59+17})$ 为嘴部轮廓,其中 (x_{59}, y_{59}) 和 (x_{65}, y_{65}) 分别为左右嘴角。

1) 左、右眼子区域划分方法:根据左右眼角确定采样框的长度,眉毛轮廓最高点和眼睛轮廓最低点确定采样框的宽;为了取得眼睛区域更多的细节特征,长和宽分别向外延伸 L_{leye} 、 L_{reye} 个像素点;此区域归一化到某一固定像素大小。左眼区域长宽求取公式如下:

$$\begin{cases} w_{leye} = x_{30} - x_{34} + L_{leye} \\ h_{leye} = y_{36} - y_{17} + L_{leye} \end{cases}$$

其中: w_{leye} 为左眼子区域的宽, h_{leye} 为左眼子区域的长。

右眼区域长宽求取公式如下:

$$\begin{cases} w_{reye} = x_{44} - x_{40} + L_{reye} \\ h_{reye} = y_{46} - y_{24} + L_{reye} \end{cases}$$

其中: w_{reye} 为右眼子区域的宽, h_{reye} 为右眼子区域的长。

2) 鼻子子区域划分方法:根据鼻子轮廓的最低点和眉毛轮廓最前端的特征点确定了鼻子区域采样框的长度,根据鼻子轮廓最左边的特征点和最右边的特征点确定了采样框的宽度;长和宽分别向外延伸 L_{nose} 个像素点;此区域归一化到某一固定像素大小。此区域长宽求取公式如下:

$$\begin{cases} w_{nose} = x_{54} - x_{58} + L_{nose} \\ h_{nose} = y_{56} - y_{22} + L_{nose} \end{cases}$$

其中: w_{nose} 为鼻子子区域的宽, h_{nose} 为鼻子子区域的长。

3) 嘴子区域划分方法:根据嘴部特征点左右两边的嘴角确定采样框的长度,根据嘴部轮廓最上面和最下面特征点位

置确定采样框的宽;长和宽分别向外延伸 L_{mouth} 个像素点;此区域归一化到某一固定像素大小。此区域长宽求取公式如下:

$$\begin{cases} w_{mouth} = x_{65} - x_{59} + L_{mouth} \\ h_{mouth} = y_{74} - y_{62} + L_{mouth} \end{cases}$$

其中: w_{mouth} 为嘴部子区域的宽, h_{mouth} 为嘴部子区域的长。

4) 下巴子区域划分方法:根据与嘴角平齐的下巴轮廓上的两个特征点确定下巴采样框的长,根据这两个特征点与下巴轮廓的最低点确定下巴采样框的宽;长和宽分别向外延伸 L_{chin} 个像素点;此区域归一化到某一固定像素大小。此区域长宽求取公式如下:

$$\begin{cases} w_{chin} = x_9 - x_3 + L_{chin} \\ h_{chin} = y_6 - y_{74} + L_{chin} \end{cases}$$

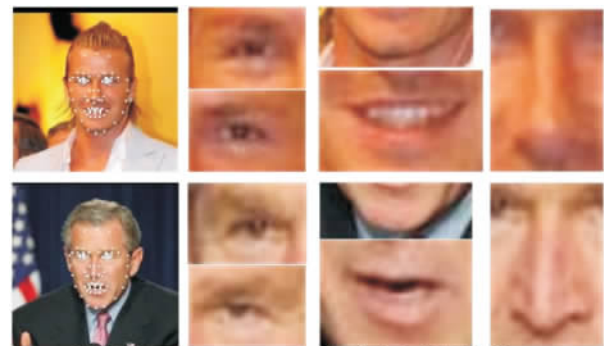
其中: w_{chin} 为下巴子区域的宽, h_{chin} 为下巴子区域的长。

将一个 $m \times n$ 的图像矩阵 I 按上述长宽截取5个子区域,即:

$$\{I_{leye}, I_{reye}, I_{nose}, I_{mouth}, I_{chin}\} \in I$$

其中: I_{leye} 、 I_{reye} 、 I_{nose} 、 I_{mouth} 、 I_{chin} 表示截取后的5个子区域像素矩阵。

以上采样框归一化均采用的是双线性内插法,经过上述步骤得到的5个不同区域采样块,它们构成了训练样本集,如图3所示。



(a) 面部特征点的提取结果

(b) 5个不同子区域采样块

图3 子区域划分示意图

1.4.2 隶属度求取的网络模型

由于人脸面部5不同区域的大小不同,拉直成列向量作为输入层,节点个数也就不同,所以对于5个人脸面部区域需要分别构建出相应的DBN,网络的结构如图4所示。

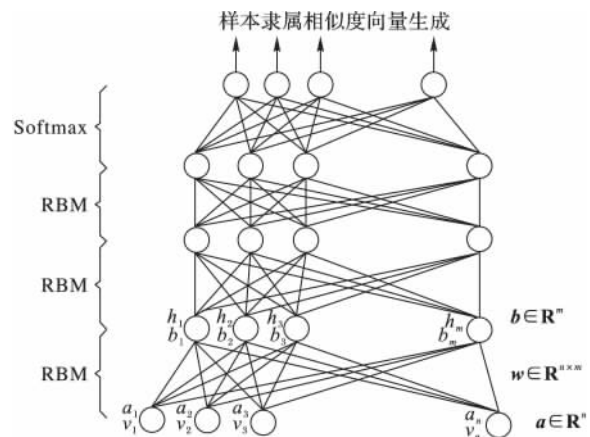


图4 子区域网络结构

隶属度求取的网络是一个5层的DBN,由一个输入层、

一个输出层和三个隐含层组成: 网络输入层神经元个数为各子区域长、宽之积; 隐含层的前两个隐含层的神经元个数相同, 第三个隐含层的神经元个数是前两层的两倍; 输出层的神经元个数与类别数相同。

将提取得到的5个区域采样块及整个人脸分别输入到对应DBN中; 隐含层作用是经过自下而上逐层提取得到更抽象特征; 输出层作用是利用Softmax分类器求出某一区域特征对于各类的隶属度, 形成隶属度向量 $P_i = (p_{1i} \ p_{2i} \ \dots \ p_{ni})^T$ 。对每个区域均求出隶属度向量便组成了隶属度矩阵 P 。

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & p_{n4} & p_{n5} \end{bmatrix}$$

其中: 5列分别代表5个区域, 行数代表有 n 个类别。

1.4.3 隶属度的融合方法

本文提出利用在求隶属度过程中构建的DBN, 求出各个面部器官的识别率, 识别率越高说明在识别时此器官起的作用较大, 越低说明起的作用越小, 并将此识别率作为各个器官采样区域权值系数确定的主要依据。具体方法公式如下:

$$\omega_i = \frac{p_i}{p_1 + p_2 + p_3 + p_4 + p_5}$$

其中 $p_i (i = 1 \ 2 \ 3 \ 4 \ 5)$ 为子区域的隶属度。

这样使得 $\omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5 = 1$ 。在对人脸图像进行识别时, 主要根据上面求得的权值矩阵 P 隶属度向量 ω 相乘, 公式如下:

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & p_{n4} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \end{bmatrix} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ \vdots \\ q_n \end{bmatrix}$$

从而得出此待识别人脸图像隶属于各个类别的综合评分向量 $Q = (q_1 \ q_2 \ \dots \ q_n)$ 。找出其中评分最高的类, 即为识别的最终结果。

1.5 算法描述

基于深度神经网络的特征加权融合人脸识别方法的具体算法实现过程如下:

算法1 训练算法。

1) 利用ASM标注LFW人脸库中全部13233幅图像, 人脸面部特征点向量 $x_i = (x_{i0} \ y_{i0} \ x_{i1} \ y_{i1} \ \dots \ x_{i(n-1)} \ y_{i(n-1)})$ 。

2) 根据面部特征点向量 x_i 确定5个采样区域位置, 并计算5个采样框的长和宽 $w_{leye} \ h_{leye} \ w_{reye} \ h_{reye} \ w_{nose} \ h_{nose} \ w_{mouth} \ h_{mouth} \ w_{chin} \ h_{chin}$ 。

3) 在灰度化后原图像 I 中根据2) 求出的采样框大小, 进行采样得到5个子区域 $\{I_{leye} \ I_{reye} \ I_{nose} \ I_{mouth} \ I_{chin}\} \in I$, 以眼睛区域为例:

$$I_{leye} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1 \mu_{leye}} \\ x_{21} & x_{22} & \dots & x_{2 \mu_{leye}} \\ \vdots & \vdots & & \vdots \\ x_{h_{leye}1} & x_{h_{leye}2} & \dots & x_{h_{leye} \mu_{leye}} \end{bmatrix}$$

其中 x_{ij} 表示采样后子区域内像素值。

4) 将5个子区域利用双线性内插法归一化到规定大小, 并拉直成列向量。

$$I_{leye} = (x_{11} \ x_{12} \ \dots \ x_{h_{leye} \mu_{leye}})^T$$

5) 对5个子区域进行灰度化、灰度拉伸等预处理操作, 形成初调训练样本集, 从将要用于测试的类别中选取一部分作为精调训练样本集。

6) 利用得到的5个初调训练样本集分别对5个DBN网络应用对比散度方法^[17]进行预训练调整各层权值矩阵 W_i 及偏置矩阵 $B_i (i$ 表示第 i 层)。

7) 应用反向传播方法, 在精调训练样本集上对DBN各层权值矩阵 W_i 及偏置矩阵 B_i 精调。

8) 训练完成后, 将各层权值矩阵 W_i 及偏置矩阵 B_i 保存, 供识别时使用。

算法2 分类算法。

1) 对于一张给定的待识别人脸图像, 利用ASM提取出人脸面部特征点向量 $x = (x_0 \ y_0 \ x_1 \ y_1 \ \dots \ x_{(n-1)} \ y_{(n-1)})$ 。

2) 根据面部特征点向量 x 确定5个采样区域位置, 并计算5个采样框的长和宽 $w_{leye} \ h_{leye} \ w_{reye} \ h_{reye} \ w_{nose} \ h_{nose} \ w_{mouth} \ h_{mouth} \ w_{chin} \ h_{chin}$ 。

3) 在原图像 I 中根据上步求出的采样框大小, 进行采样得到5个子区域 $\{I_{leye} \ I_{reye} \ I_{nose} \ I_{mouth} \ I_{chin}\} \in I$ 。

4) 将5个子区域利用双线性内插法归一化到规定大小, 而后进行灰度化、灰度拉伸等预处理操作。

5) 将步骤4) 中得到的子区域拉直成列向量, 输入到由训练完成的权值矩阵 W_i 及偏置矩阵 B_i 构造的DBN中, 求出隶属度矩阵 P 。

6) 将隶属度矩阵 P 与权值向量 $\omega = (\omega_1 \ \omega_2 \ \omega_3 \ \omega_4 \ \omega_5)$ 相乘求出此待识别人脸图像隶属于各个类别的综合评分向量 $Q = (q_1 \ q_2 \ \dots \ q_n)$ 。

7) 找出评分向量 Q 中获得评分最高的类, 作为分类结果。

2 实验结果及分析

为了验证本文提出算法的有效性, 分别在LFW (Labeled Faces in the Wild Database) 人脸库和ORL (Olivetti research laboratory) 人脸库上进行人脸识别仿真实验。选用LFW人脸库进行实验的目的是为了验证本文算法在非限制条件下的有效性; 同时, 由于LFW人脸库中有13233幅图像, 为DBN的预训练提供了充足的训练样本, 能保证DBN参数达到全局最优^[18]。为了进一步评估本文算法在限制条件下的识别能力, 还在ORL人脸库上进行了测试和比较。实施本文实验的硬件环境为: 3.50 GHz的Intel Core i5-4690 CPU 8 GB内存。

2.1 LFW人脸库上的实验

LFW中共包含5749人的13233幅图像, 其中1680人有两幅及以上的图像, 4069人只有一幅图像; 图像分辨率为 250×250 , 以JPEG格式存储, 绝大多数为彩色图, 少数为灰度图像。LFW主要用于非限制条件下人脸识别研究, 已成为学术界(工业界)评价识别性能的基准, 它能充分表现真实条件下的人脸图像变化, 如光照、姿态、表情、种族、背景、性别等。

本文选取库中具有图像数目大于或等于25幅的人作为实验对象, 其中包括29个人共2458幅图像。每人随机选取

5 幅图像作为训练样本,其余为测试样本。这些图像先进行人脸的特征点提取,然后根据特征点的位置进行器官区域划分,然后归一化到固定的大小(下巴区域 88×33 、左眼和右眼区域 44×31 、嘴部区域 51×34 、鼻子区域 32×46),由此作为训练、测试样本。

1) 隐含层单元数不同时对比实验。

DBN 的分类能力主要由自下向上、多层学习而得到的特征有效性决定。为了得到区域更多细节特征,各区域长宽分别扩大了 $L_{\text{eye}} \setminus L_{\text{re}} \setminus L_{\text{nose}} \setminus L_{\text{mouth}} \setminus L_{\text{chin}}$,它们取值是各区域长的 0.2 倍。DBN 的构建、各隐含层神经元个数的确定至关重要,本文网络构建是在原神经网络构建方法基础上的拓展。因为 LFW 人脸库中的图像都是在非限制条件下采集的,所以人脸图像特征复杂,且含有较多的噪声,为了能有效地提取人脸图像特征,首先应讨论网络构建问题中不同隐含层神经元个数对网络性能的影响。实验结果如表 1 所示。

表 1 LFW 库上不同网络结构的正确识别率

3 个隐含层 神经元个数	识别率/%	
	DBN	DLWF
$1000 \times 1000 \times 2000$	55.75	56.25
$1500 \times 1500 \times 3000$	64.50	66.75
$2000 \times 2000 \times 4000$	68.00	70.25
$2500 \times 2500 \times 5000$	71.00	74.50
$3000 \times 3000 \times 6000$	82.50	85.25

由表 1 可知,随着隐含层单元个数增加,深度神经网络能够更好地表达人脸图像特征,但是计算量也逐渐增大。同时,在隐含层单元数相同时,本文算法的识别率均高于输入整张人脸图像像素级特征的传统深度神经网络算法,并且没有因为特征的不完整性导致识别率下降。

2) 训练样本数不同时对比实验。

主要测试本文算法在不同数量的训练样本条件下,与传统算法在识别率方面的优劣程度。由于通常条件下很难采集到大量有标签的训练样本,所以在训练样本个数较少的条件下能取得较好的识别率就变得尤为关键。表 2 为深度神经网络结构为 $3000 \times 3000 \times 6000$ 时,在不同训练样本个数条件下,与传统人脸识别方法的比较结果。

由表 2 可知,本文算法经过非监督预训练,为 DBN 的参数提供了良好的初始化值,再经过少量有标签的训练样本有监督的微调,便能达到较好的识别效果。随着训练样本数增加,开始时识别率上升然后趋于不变,表明本文算法在非限制条件下人脸识别具有较好的鲁棒性。同时,表 2 也说明了本文算法比输入整张人脸图像像素级特征的 DBN 识别效果更好。传统人脸识别算法 PCA、支持向量机 (Support Vector Machine, SVM) 等在样本数较少时,提取的特征分类能力并不强;当训练样数增加时,提取的特征分类能力有所增加,但在非限制条件下算法的鲁棒性并不强。参考文献 [5] 中的方法先利用深度学习提取出人脸图像的 FIP (Face Identity-Preserving) 特征,然后再选择其他的分类方法进行身份的判别。本文对比实验采用的分类方法是线性判别式分析 (Linear Discriminant Analysis, LDA)。与参考文献 [5] 中提出的方法相比,本文方法在识别率方面略有提升,表明本文算法在非限制条件下具有高效的人脸识别能力。

表 2 LFW 库上不同训练样本数正确识别率 %

方法	微调阶段训练样本数		
	5	10	15
PCA	23.00	27.0	29
SVM	35.50	44.0	55
DBN	82.50	83.5	86
FIP + LDA	83.00	85.5	86
DLWF	85.25	87.5	88

3) 类别数不同时对比实验。

本实验的目的主要是为了验证本文算法在非限制条件下进行人脸识别的有效性和普适性。从 LFW 人脸库中选择拥有图像数目大于或等于 25 张的 39 个人作为实验候选对象,从中随机选择 10、20、30 个人作为实验对象,每人随机选取 5 幅图像作为训练样本,其余作为测试样本。多次随机选择不同的人进行实验取平均值作为最终的结果。本次实验采用的 DBN 结构与实验 2 相同,实验结果如表 3 所示。

表 3 LFW 库上类别数不同时不同算法正确识别率

方法	类别数		
	10	20	30
PCA	27.32	25.74	24.63
SVM	41.65	38.23	36.46
DBN	83.71	84.05	85.21
FIP + LDA	84.61	86.72	87.51
DLWF	85.93	87.16	88.76

由表 3 可知,在与传统人脸识别算法相比较时,本文提出的算法识别率最高,在类别数为 30 时识别率可以达到 88.76%,说明本文算法具有较好的识别能力;同时,由于本次实验采用了不同的训练、测试对象,体现了本文算法在人脸识别中具有较好的鲁棒性、稳健性。

经过以上在 LFW 人脸库上的实验可知,本文算法在非限制条件下人脸识别的准确率和鲁棒性均高于传统的人脸识别方法。本文通过将人脸面部进行区域划分,分别提取特征的方法,比传统只输入整张人脸图像像素级特征的方法更容易学习到人脸图像的局部特征。

2.2 ORL 人脸库上的实验

ORL 人脸图像库又称 AT&T 人脸图像库,是剑桥大学历经两年拍摄完成,包含了 40 个人,每个人 10 幅不同姿态、表情、光照的 400 幅图像,其中人脸的姿态和光照变化相对较小;图像分辨率为 112×92 ,灰度级为 256,图像背景均为黑色,存储格式为 PGM 格式。在实验中,神经网络首先是用 LFW 人脸库所有人脸图片进行预训练,然后在 ORL 库中每人随机选取 5 幅图像作为微调阶段训练样本,剩下的作为测试样本。

1) 隐含层单元数不同时对比实验。

本实验是为了验证本文算法的通用性,及在 ORL 人脸库上不同的网络结构所表现的识别性能。表 4 为 DBN 隐含层单元个数不同时识别结果。从表 4 可知,当隐含层节点逐渐增加时,网络逐层学习人脸图像抽象特征的分类能力也随之增强。当网络的结构为 $3000 \times 3000 \times 6000$ 时取得了较高识别率 97%,较 PCA、SVM、传统 DBN 分别提高了 13%、8.5%、3%,说明本文算法具有较好的通用性。

2) 与其他人脸识别方法对比实验。

为了验证本文算法对传统人脸识别算法优越性,分别与

PCA、SVM 以及传统的 DBN 算法作了对比实验,实验结果如表 5 所示,可以看出本文算法具有较高的通用性,与传统的人脸识别方法相比,在识别率方面具有明显优势。

表 4 ORL 库上不同网络结构时正确识别率

3 个隐含层 神经元个数	识别率/%	
	DBN	DLWF
1 000 × 1 000 × 2 000	44	57
1 500 × 1 500 × 3 000	52	64
2 000 × 2 000 × 4 000	71	77
2 500 × 2 500 × 5 000	82	86
3 000 × 3 000 × 6 000	94	97

表 5 ORL 库上不同算法识别率比较

方法	识别率/%	方法	识别率/%
PCA	84.0	FIP + LDA	96.0
SVM	88.5	DLWF	97.0
DBN	94.0		

3 结语

本文提出的基于深度神经网络和特征加权融合人脸识别方法,首先通过提取到的特征点将人脸图像分为多个局部区域采样块,然后将采样块经过归一化处理后分别输入到不同 DBN,求出相似度向量,最后进行加权求和得出最后的识别结果并计算识别率。通过在 LFW 人脸库上的实验结果证明了本文算法在非限制条件下的有效性,能够自下而上提取到人脸图像的抽象适合于分类的特征,识别率可以达到 88.76%。通过对人脸的分块使 DBN 能够提取得到人脸图像的局部特征。同时在 ORL 人脸库上的实验证明了本文算法的有效性和通用性,识别率可以达到 97%,与传统人脸识别算法相比表现出了更高的优越性。

参考文献:

- [1] TURK M, ALEX P. Eigenfaces for recognition [J]. Journal of Cognitive Neuroscience, 1991, 3(2): 71–86.
- [2] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10 000 classes [C]// CVPR 14: Proceedings of the 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 1891–1898.
- [3] HU J, LU J, TAN Y-P. Discriminative deep metric learning for face verification in the wild [C]// CVPR 14: Proceedings of the 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 1875–1882.
- [4] HUANG G, LEE H, LEARNED-MILLER E. Learning hierarchical representations for face verification with convolutional deep belief networks [C]// CVPR 12: Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2012: 2518–2525.
- [5] ZHU Z, LUO P, WANG X, et al. Deep learning identity-preserving face space [C]// ICCV 2013: Proceedings of the 2013 IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2013: 113–120.
- [6] 张雯,王文伟. 基于局部二值模式和深度学习的人脸识别[J]. 计算机应用, 2015, 35(5): 1474–1478. (ZHANG W, WANG W W. Face recognition based on local binary pattern and deep learning [J]. Journal of Computer Applications, 2015, 35(5): 1474–

1478.)

- [7] SUN Y, WANG X, TANG X. Hybrid deep learning for face verification [C]// ICCV 2013: Proceedings of the 2013 IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2013: 1489–1496.
- [8] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: closing the gap to human-level performance in face verification [C]// CVPR 14: Proceedings of the 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 1701–1708.
- [9] SUN Y, CHEN Y, WANG X, et al. Deep learning face representation by joint identification-verification [C]// NIPS 2014: Advances in Neural Information Processing Systems 27. Cambridge, MA: MIT Press, 2014: 1988–1996.
- [10] HINTON G E, OSINDERO S, THE Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527–1554.
- [11] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504–507.
- [12] AREL I, ROSE D C, KARNOWSKI T P. Deep machine learning — a new frontier in artificial intelligence research [research frontier] [J]. IEEE Computational Intelligence Magazine, 2010, 5(4): 13–18.
- [13] COOTES T F, TAYLOR C J, COOPER D H, et al. Active shape models—their training and application [J]. Computer Vision and Image Understanding, 1995, 61(1): 38–59.
- [14] MILBORROW S, NICOLLS F. Active shape models with SIFT descriptors and MARS [C]// VISAPP 2014: Proceedings of the 2014 9th International Conference on Computer Vision Theory and Applications. Singapore: International Association of Computer Science and Information Technology, 2014 [2015-06-16]. <http://milbo.org/stasm-files/active-shape-models-with-sift-and-mars.pdf>.
- [15] MILBORROW S, BISHOP T E, NICOLLS F. Multiview active shape models with SIFT descriptors for the 300-W face landmark challenge [C]// ICCVW 13: Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops. Washington, DC: IEEE Computer Society, 2013: 378–385.
- [16] BENGIO Y, DELALLEAU O. On the expressive power of deep architectures [C]// DS 2011: Proceedings of the 14th International Conference on Discovery Science, LNCS 6926. Berlin: Springer-Verlag, 2011: 18–36.
- [17] HINTON G E. Training products of experts by minimizing contrastive divergence [J]. Neural Computation, 2002, 14(8): 1771–1800.
- [18] BENGIO Y. Learning deep architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1–127.

Background

This work is partially supported by the National Key Technology R&D Program (2013BAH12F02).

SUN Jinguang, born in 1962, Ph. D., professor. Her research interests include computer image processing, computer graphics, knowledge engineering.

MENG Fanyu, born in 1991, M. S. candidate. His research interests include computer image processing.