

文章编号: 1007-5321(2017)01-0084-05

DOI: 10.13190/j.jbupt.2017.01.015

# 基于深度卷积神经网络的跨年龄人脸识别

李 亚<sup>1</sup>, 王广润<sup>2</sup>, 王 青<sup>2</sup>

(1. 广州大学 计算机科学与教育软件学院, 广州 510006; 2. 中山大学 数据科学与计算机学院, 广州 510006)

**摘要:** 提出了一种应用于跨年龄人脸识别的联合学习方法, 该方法由深度卷积神经网络构建而成, 能在特征学习的同时学习到最优的测度函数, 从而避免不合适的固定阈值所带来的匹配错误. 针对有限的内存、过拟合和计算复杂性问题, 在模型训练过程中采用了多种新颖和有效的训练策略. 实验证实了该联合学习方法的有效性, 在公开数据库 MORPH-II 上的识别正确率达到了 93.6%.

**关键词:** 人脸比对; 人脸识别; 跨年龄; 深度卷积神经网络; 联合学习

**中图分类号:** TP391.4

**文献标志码:** A

## A Deep Joint Learning Approach for Age Invariant Face Verification

LI Ya<sup>1</sup>, WANG Guang-run<sup>2</sup>, WANG Qing<sup>2</sup>

(1. School of Computer Science and Educational Software, Guangzhou University, Guangzhou 510006, China;

2. School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China)

**Abstract:** A joint learning approach (JLA) based on deep convolutional neural network (CNN) for age-invariant face verification was proposed. Feature representation, distance metric and decision function can be learned simultaneously thereafter. Comparing with traditional approaches, it uses fix threshold, so the match errors caused by unfit threshold can be avoided. Some strategies to overcome insufficient memory capacity, prevent over-fitting and reduce computational cost were also introduced. Experiment demonstrates the effectiveness of this approach; the rank-1 recognition accuracy is improved to 93.6% on the MORPH-II database.

**Key words:** face verification; face recognition; age invariant; deep convolutional neural network; joint learning

随着年龄的增长, 人的容貌变化较大, 而不同年龄段的变化特点也有所不同: 从出生到成年, 最大的变化是颅面变化, 这种头部轮廓的变化是形状的变化; 而从成年到老年, 可观察到的变化就是皮肤的逐渐老化, 是面部纹理的变化. 同一个体的这些变化称为类内差异性. 不同人在相近的年龄段有可能看起来比较相似, 称为类间相似性. 除了受类间相似性和类内差异性的影响外, 人脸图像还受光照、表

情、姿态、发型等诸多因素的影响, 因此, 跨年龄人脸识别问题非常具有挑战性.

## 1 相关研究

近年来, 跨年龄人脸识别问题引起了越来越多学者的关注. Li 等<sup>[1]</sup>提出多特征区分分析法 (MF-DA, multi-feature discriminant analysis); Gong 等<sup>[2]</sup>提出潜在因素分析法 (HFA, hidden factor analysis);

收稿日期: 2016-03-01

基金项目: 广州市属高校科研项目 (1201620302); 广东省科技计划项目 (2013B010406005, 2015B010128009)

作者简介: 李 亚 (1980—), 女, 讲师, E-mail: liya@gzhu.edu.cn; 王 青 (1973—), 男, 副教授.

Chen 等<sup>[3]</sup>提出了一种称为跨年龄引用编码法 (CARC, cross-age reference coding) 的数据驱动方法; Gong 等<sup>[4]</sup>在 HFA 的基础上, 进一步设计了最大熵特征描述子 (MEFD, maximum entropy feature descriptor), 并采用身份因素分析匹配法 (IFA, identity factor analysis) 进行跨年龄人脸识别. 最近, Li 等<sup>[5]</sup>提出了一种新的层次学习模型法, 并采用局部特征选择描述子 (LPS, local pattern selection) 进行跨年龄人脸识别.

以上方法中特征提取、特征变换和匹配识别各阶段都是独立进行的. 实际上, 特征提取和匹配识别之间是相互影响的, 采用的特征不同, 匹配的效果也会有所不同. 如果构建特征和匹配联合学习的模型, 提取最符合目标要求的特征, 势必会提高匹配的效果. 传统手工设计的特征很难达到这样的要求, 而目前流行的深度卷积神经网络 (CNN, convolutional neural network) 可较为方便地构建高效的“端到端”的联合学习框架. 深度卷积神经网络法已经在诸多计算机视觉任务, 如图像分类、物体检测、语义分割中都取得了良好的效果, 但目前还没有将基于 CNN 的联合学习模型应用于跨年龄识别的任务中. 为了能在特征学习的同时学习到最优的测度函数, 提出一种基于 CNN 的特征学习和距离测度学习的联合学习方法 (JLA, joint learning approach), 并将其应用到跨年龄的人脸识别研究中.

## 2 距离测度学习及优化目标

Xing 等<sup>[6]</sup>首次提出马氏距离形式的距离测度学习算法. 假设  $x$  和  $y$  分别为图像对  $(x, y)$  的特征向量, 马氏距离为  $d(x, y) = (x - y)^T M (x - y)$ , 其中  $M$  为半正定矩阵. 距离测度学习方法一般通过相似样本对和不相似样本对的监督信息来学习马氏距离的参数矩阵  $M$ . 在匹配阶段通常需要一个固定阈值  $d$  来确定  $x$  和  $y$  是否匹配, 如果距离小于或等于阈值  $d$  则匹配; 否则, 不匹配, 可公式化为

$$(x - y)^T M (x - y) \leq d, M \geq 0 \quad (1)$$

然而, 现实中很难确定一个合适的固定阈值. Li 等<sup>[7]</sup>提出学习一个自适应的阈值函数  $d$ ,  $d$  为关于  $(x, y)$  的函数, 而不是固定阈值, 从而避免由阈值不合适引起的判定错误. 由此, 式 (1) 变为  $(x - y)^T \times M (x - y) \leq d(x, y)$ , 而匹配问题的判定函数可写为

$$f(x, y) = d(x, y) -$$

$$(x - y)^T M (x - y) \begin{cases} \geq 0, & \text{若 } l(x) = l(y) \\ < 0, & \text{其他} \end{cases} \quad (2)$$

其中:  $l(x)$  和  $l(y)$  分别为图像  $x$  和  $y$  的类别标签,  $l(x) = l(y)$  表示  $x$  和  $y$  为同一类.  $M$  本身是二次的, 假设  $d(x, y)$  为简单的二次型, 有

$$d(x, y) = \frac{1}{2} x^T \tilde{A} x + \frac{1}{2} y^T \tilde{A} y + x^T \tilde{B} y + c^T (x + y) + b$$

代入式 (2) 可得

$$f(x, y) = \frac{1}{2} x^T A x + \frac{1}{2} y^T A y + x^T B y + c^T (x + y) + b \quad (3)$$

其中:  $A = \tilde{A} - 2M$ ,  $B = \tilde{B} + 2M$ , 且  $A$  为半正定矩阵,  $B$  为半负定矩阵,  $A$  和  $B$  可进一步分解为  $L_A^T L_A$  和  $-L_B^T L_B$ . 因此, 式 (3) 进一步转化为

$$f(x, y) = \frac{1}{2} (L_A x)^T (L_A x) + \frac{1}{2} (L_A y)^T (L_A y) - (L_B x)^T (L_B y) + c^T x + c^T y + b \quad (4)$$

经过以上分析, 通过有监督训练学习到矩阵  $L_A$ ,  $L_B$  和向量  $c$  后, 运用式 (4) 就可确定图像的匹配问题. 跨年龄人脸识别问题的目标为: 对于人物  $P$  的某年龄实例  $z$ , 学习一个重识别的模型从而能正确地识别  $P$  的另一个年龄实例  $z'$ , 也就是说, 希望学习到与年龄无关的有关人物  $P$  的身份信息. 为此, 在构造训练样本时, 每个人应包含尽可能多的年龄实例, 使模型对于年龄的变化更健壮, 并规定正样本为同一人不同年龄的图像组成的图像对, 负样本为不同人的图像组成的图像对. 给定训练集  $Z = \{(z_i, y_i)\}_{i=1}^N$ , 其中  $y_i$  为类标签, 也就是人物身份标识 (ID, identification),  $N$  为总人数, 则图像对集合可表示为  $\Omega = \{\Omega_k = (z_i, z_j)\}$ ,  $k$  为  $\Omega$  中第  $k$  个图像对. 当  $z_i$  和  $z_j$  为同一个人时, 图像对  $\Omega_k$  的标签  $l(\Omega_k)$  等于 1; 否则等于 -1, 即

$$l(\Omega_k) = l(z_i, z_j) = \begin{cases} 1, & \text{若 } l(z_i) = l(z_j) \\ -1, & \text{其他} \end{cases} \quad (5)$$

以  $\Omega$  作为输入, 可通过最大化  $l(\Omega_k) \times f(\Omega_k)$ , 使同一人的  $f(\Omega_k)$  尽可能大, 而不同人的  $f(\Omega_k)$  尽可能小. 更进一步, 把  $(L_A, L_B, c)^T$  记为  $W$ , 忽略  $l(\Omega_k) \times f(\Omega_k) > 1$  的图像对 (这里的约束值为 1, 并不重要, 其可为任意的正值  $t$ , 只不过学到的矩阵乘以  $t$  即可), 则目标函数就可用类似铰链的损失函数表示为

$$H(W) = \sum_{\Omega} \max \{0, 1 - l(\Omega_k) \times f(\Omega_k)\},$$

$$k=1, 2, \dots, N^2 \quad (6)$$

损失函数也可写为  $\sum_{\Omega} [1 - l(\Omega_k) f(\Omega_k)]_+$ , 表明经验损失值取正数, 即当  $1 - l(\Omega_k) f(\Omega_k) \leq 0$  时, 经验损失值为 0.

### 3 联合学习方法介绍

#### 3.1 网络结构

深度卷积神经网络一般由 1 个或多个卷积层和顶端的全连接层组成, 同时也包括关联权重和池化层.  $L_A$ 、 $L_B$  和  $c$  可看成是全连接层的权重, 网络的参数采用梯度下降法和链式求导法进行更新.

包含人物年龄信息的公共数据集规模都不大, 因此设计的网络不很深, 共包括 7 层: 第 1 层为卷积层, 包括 32 个尺寸  $5 \times 5 \times 3$  的卷积核, 卷积的步长为 2 个像素; 第 2 层为最大值池化层; 第 3 层也为卷积层, 包括 32 个尺寸为  $5 \times 5 \times 32$  的卷积核, 卷积的步长为 1 个像素; 第 4 层仍然为最大值池化层; 第 5~7 层包括 2 个 400 维的和 1 个 801 维的全连接层, 如图 1 所示. 在每个卷积层后采用修正线性单元 (ReLU, rectified linear unit) 进行非线性处理. 修正线性函数的处理方法为: 如果卷积计算的值小于 0, 则让其等于 0; 大于 0, 则保持原来的值不变. ReLU 的做法虽然简单, 但却能得到很好的稀疏性. 实际上, 这样 7 层的网络可分成特征学习、共同空间映射和测度学习 3 部分. 特征学习阶段通过堆叠 2 个卷积-池化层和 1 个全连接层学习到 400 维的特征, 第 2 个全连接层起到了把特征映射到共同空间的作用, 第 3 个 801 维的全连接层实现了测度的学习. 通过 3 个阶段同步学习, 同时学到矩阵  $L_A$ 、 $L_B$  和向量  $c$ .

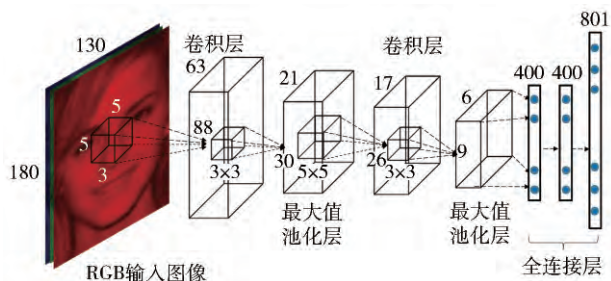


图 1 网络结构

#### 3.2 训练策略

##### 3.2.1 最小批选择策略

存储有限的内存和大规模训练样本间的不匹配

促使采用最小批选择策略. 若有  $N$  个训练样本, 样本对数目为  $O(N^2)$ , 即使对于中等规模的数据集来说也不可能把所有的样本对都调入内存中. 采用随机选择的方法选择样本对方法最简单, 然而随机选择会使负样本数远远大于正样本数. 例如, 对于总人数为  $P$ , 每人平均  $N$  张图像的带标签的数据集来说, 负样本的数目为正样本的  $(p-1)^2$  倍. 随机选择会使正负样本数严重不均衡.

为了解决上述问题, 在每次迭代中随机选择固定数目的人, 然后仅在这些人的图像中产生图像对. 图像对保证穷举所有的正样本, 负样本随机选择, 但负样本数要和正样本数保持一定的比例, 避免正负样本严重不平衡.

##### 3.2.2 数据增强策略

采用 2 种数据增强策略来增加训练样本数, 减少过拟合现象. 第 1 种策略为提取正样本对的图像块, 使用这些图像块进行训练. 图像块的高度和宽度与原图像成一定的比例, 如输入图像大小为  $150 \times 200$ , 图像块占原图的 90%, 图像块的大小为  $135 \times 180$ , 这样正样本对的数目就扩大到原来的 300 倍. 第 2 种策略是对正样本对的水平映射. 理论上来说, 正脸图像具有左右对称性, 若分别对左右脸进行水平映射, 原来的图像数就增加了 2 倍. 相应地组合后的图像对数目为原来的 9 倍. 但现实情况下有很多非正脸图像存在, 随机选择 1/4 的正样本对进行水平映射, 这样正样本对的数目就扩大到原来的 3 倍. 经过这 2 种策略的数据增强, 原来的正样本数扩大到原来的 900 倍, 在一定程度上缓解了数据不足造成的过拟合问题.

##### 3.2.3 部分神经元退出策略

部分神经元退出策略是指在模型训练时随机让网络某些隐含层节点退出不起作用, 退出的那些节点可以暂时认为不是网络结构的一部分, 但它们的权重需要保留下来, 因为下次样本输入时它可能又起作用. 这样的策略可减少神经元间的相互适应, 使模型更鲁棒, 因为隐含节点都是以一定的概率随机出现, 并不能保证相邻的节点每次都同时出现, 因此使权值的更新不依赖于有固定关系隐含节点的共同作用.

对第 1 卷积层的输出按照 0.5 的比例置 0, 退出的神经元对于前馈网络没有贡献, 也不参加反向传播. 在模型测试阶段, 使用均值网络来得到隐含

层的输出,也就是在网络前向传播过程中的隐含层节点的输出值都要减半。

### 3.3 学习算法

算法每次迭代首先选择最小批输入网络,然后构建图像对,再根据损失函数采用统计梯度下降法更新模型的参数。批处理形式下的损失函数可写为

$$H(W) = \sum_{i,j \in m} [1 - l(I_i, I_j) f(F(I_i), F(I_j))] \quad (7)$$

其中  $\langle I_i, I_j \rangle$  为图像对,  $F(I_i)$  为图像  $I_i$  经过处理后的输出,  $\{I_i\}$  为被选的最小批中所有不同的图像集合,  $m$  为图像数。式(7)只与图像对有关,因此可简记为

$$H(W) = \sum_{i,j \in m} J(F(I_i), F(I_j)) \quad (8)$$

更新参数最直观的做法是根据损失函数先计算每个图像对的梯度,然后把所有图像对的梯度相加求和得到整体的梯度。图像对的损失可通过 2 幅图像的输出来定义,由于每个图像对包含的 2 幅图像会引起 2 次前向和反向传播,如果某个图像出现在多个图像对中,按图像对来计算损失函数时,该图像就会被重复计算多次。为了提高计算效率,可采用按最小批中不同的图像来计算损失,这样重复出现在多个图像对中的图像其前向和反向传播得以重复利用。为此式(8)可进一步写为

$$H(W) = H(F(I_1), F(I_2), \dots, F(I_m)) \quad (9)$$

模型参数采用统计梯度下降法进行更新为

$$W = W - \alpha \frac{\partial}{\partial W} H(W) \quad (10)$$

其中:  $\alpha$  为学习率,  $\frac{\partial H}{\partial W}$  的计算采用标准的反向传播算法。为了更详细地说明参数更新的过程,引入符号  $W^{(l)}$  为第  $l$  层的参数,图像  $I_i$  在第  $l$  层的特征映射记为  $X_i^l$ 。注意  $X_i^l$  为中间层的输出,而  $F(I_i)$  为最后 1 层也就是最终的输出。  $\frac{\partial H}{\partial W}$  为最小批中  $m$  个不同图像的梯度之和,可通过反向传播算法和链式法则求出:

$$\frac{\partial H}{\partial W^{(l)}} = \sum_{i=1}^m \frac{\partial H}{\partial X_i^l} \frac{\partial X_i^l}{\partial W^{(l)}} \quad (11)$$

$$\frac{\partial H}{\partial X_i^l} = \frac{\partial H}{\partial X_i^{l+1}} \frac{\partial X_i^{l+1}}{\partial X_i^l} \quad (12)$$

#### 算法 1 深度联合学习算法

输入:

训练集  $X = \{(x_i, I_i)\}$ , 初始化  $W, \alpha, t = 0$

输出:

网络参数  $W$

```

1 while  $t < T$  do
2    $t \leftarrow t + 1$ 
3   从  $X$  中随机选择参加训练的人脸图像
4   从  $\{\langle I_i, I_j \rangle\}$  中选择图像对训练集  $\Omega_k$ 
5   for all  $\{I_i\}$  do
6     利用前向传播算法计算图像  $I_i$  的各层输出
        $X_i^{(1)}, \dots, X_i^{(l)}, \dots, F(I_i)$ 
7   end for
8   for all  $\{I_i\}$  do
9      $\frac{\partial H}{\partial F(I_i)} = 0$ 
10    for 所有图像对  $\Omega_k$ , i. e.  $\langle I_p, I_q \rangle$  do
11      if  $f(\Omega_k) > 1$  then
12        if  $I_i = I_p$  then
13           $\frac{\partial H}{\partial F(I_i)} + = \frac{\partial H}{\partial F(I_p)}$ 
14        else if  $I_i = I_q$  then
15           $\frac{\partial H}{\partial F(I_i)} + = \frac{\partial H}{\partial F(I_q)}$ 
16        end if
17      end if
18    end for
19    根据式(10)和式(11)计算  $\frac{\partial H}{\partial X_i^l} \frac{\partial X_i^l}{\partial W^{(l)}}$ 
20     $\Delta W = \Delta W + \frac{\partial H}{\partial X_i^l} \frac{\partial X_i^l}{\partial W^{(l)}}$ 
21  end for
22   $W^{(t)} = W^{(t-1)} - \alpha_t \Delta W$ 
23 end while
```

## 4 实验

实验采用了 MORPH-II 数据库。MORPH-II 包含超过 1.3 万人的共 5.5 万多个人脸图像,年龄分布范围为 16~77 岁,平均每人 4 张图。选择数据库中 1 万个人,每人挑选年龄差距最大的 2 张图片,共计 2 万张图片做训练数据。从数据库中剩余的人中选择 3 000 个人,每个人年龄最大的图像做查询图像,年龄最小的图像组成检索数据库,也就是在检索数据库中每人仅有 1 张图片。为使实验结果具有可比性,实验设置与文献[3]相同。在该数据库上采用累积匹配特性(CMC, cumulative match characteris-

tic) 曲线评价算法性能. CMC 曲线画出了前  $k$  (top- $k$ ) 个查询结果的累积匹配正确率. 曲线由点  $\{(i, a_i) \mid i=1, 2, \dots, k\}$  组成,  $a_i$  为前  $i$  个查询结果的累积正确率, 即只要前  $i$  个查询结果中包含与查询图像是同一人的图像, 则正确. 实验首先验证不同训练策略对效果的影响, 然后和目前较好的其他几种方法进行比较.

#### 4.1 最小批选择对结果的影响

图2的CMC曲线展示了最小批中不同人数和图像对数组合对结果的影响, 其中括号里的值分别代表人数和图像对数. 为了比较的公平性, 每种策略的正确率都是经过相同迭代次数获得的. 每人在图像对中出现的次数都相同, 例如(60, 4 800)表示批中共有60个人, 4 800个图像对, 每人出现在80个图像对中. 该组实验中的正负样本的比率固定为1:1. 从图2可以看出, (80, 11 200)的选择策略效果较优, 但实际上, 对top-10以上的性能, 选择策略的影响相对已经比较小了.

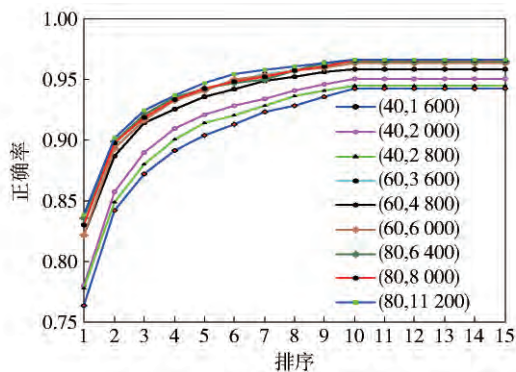


图2 最小批中人数和图像对数对结果的影响

#### 4.2 与其他方法的比较

在MORPH-II数据库上, JLA法也和其他跨年龄人脸识别方法做了比较, 包括CARC法<sup>[3]</sup>、HFA法<sup>[2]</sup>、MFDA法<sup>[1]</sup>、Li等<sup>[5]</sup>提出的LPS+HFA方法和Park等<sup>[8]</sup>提出的方法. 比较结果如表1所示.

表1 MORPH数据库上Rank-1的识别正确率

方法	识别正确率/%
Park等 <sup>[8]</sup>	79.8
MFDA <sup>[1]</sup>	83.9
HFA <sup>[2]</sup>	91.1
CARC <sup>[3]</sup>	92.8
LPS+HFA <sup>[5]</sup>	94.8
<b>JLA</b>	<b>93.6</b>

实验结果显示了JLA法的有效性, 其性能接近目前已发表的最好方法LPS+HFA. 对于top-10和top-20, JLA法的准确率分别为98.8%和99.34%, 相应的CMC曲线图如图3所示. 实际上, top-10和top-20的准确率已经基本达到了实用的要求. 与跨年龄人脸识别相关的诸多应用, 如逃犯追捕系统, 如果能成功推荐前10, 甚至前20个相似的人脸图像, 会大大降低工作的难度, 并提高追捕的成功率.

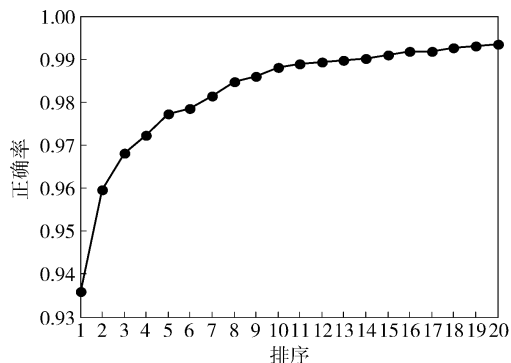


图3 JLA法在MORPH数据库上的CMC曲线

## 5 结束语

笔者提出的基于深度CNN的联合学习方法JLA在特征学习的同时可学习到最优的测度函数, 避免了不合适的固定阈值带来的匹配错误, 提高了匹配识别的正确率. 此外, 介绍了在模型学习过程中采用的多种新颖和有效的训练策略. JLA方法应用于跨年龄人脸识别的研究, 在公开数据库MORPH-II上的识别正确率达到了93.6%.

目前, 实验所用的数据库还有很大的局限性, 能满足人数众多、年龄跨度大2点要求的数据还比较少. 因此, 将致力于数据库的构建工作, 收集、整理多年龄人脸图像, 进一步增加可用数据. 此外, 将继续优化模型, 使模型更加健壮, 进一步提高其识别正确率.

#### 参考文献:

- [1] Li Zhifeng, Park U, Jain A K. A discriminative model for age invariant face recognition [J]. IEEE Transactions on Information Forensics and Security, 2011, 6(3): 1028-1037.
- [2] Gong Dihong, Li Zhifeng, Lin Dahua, et al. Hidden factor analysis for age invariant face recognition [C]// ICCV 2013. Sydney: IEEE, 2013: 2872-2879.

(下转第110页)

- [6] Qiu Tao, Xu Wenjun, He Zhiqiang, et al. Graph-based spectrum sharing for multiuser OFDM cognitive radio networks[C]//Wireless Communication and Signal Processing. [S. l.]: [s. n.], 2011: 1-5.
- [7] Elnainay M Y. Island genetic algorithm-based cognitive networks [D]. Blacksburg: Virginia Polytechnic Institute and State University, 2009.
- [8] 柴争义, 刘芳. 基于免疫克隆选择优化的认知无线网络频谱分配[J]. 通信学报, 2010, 31(11): 92-100.  
Chai Zhengyi, Liu Fang. Spectrum allocation of cognitive wireless network based on immune clone selection optimization[J]. Journal on Communications, 2010, 31(11): 92-100.

---

(上接第88页)

- [3] Chen Borchun, Chen Chusong, Hsu W H. Cross-age reference coding for age-invariant face recognition and retrieval[C]//ECCV 2014. Zurich: Springer, 2014: 768-783.
- [4] Gong Dihong, Li Zhifeng, Tao Dacheng. A maximum entropy feature descriptor for age invariant face recognition [C]//CVPR 2015. Boston: IEEE, 2015: 5289-5297.
- [5] Li Zhifeng, Gong Dihong, Li Xuelong, et al. Aging face recognition: a hierarchical learning model based on local patterns selection [J]. IEEE Transactions on Image Processing, 2016, 25(5): 2146-2154.
- [6] Xing E P, Jordan M I, Russell S, et al. Distance metric learning with application to clustering with side-information [C] // NIPS 2002. Vancouver: NIPS Foundation, 2002: 521-528.
- [7] Li Zhen, Chang Shiyu, Liang Feng, et al. Learning locally adaptive decision functions for person verification [C]//CVPR 2013. Portland: IEEE, 2013: 3610-3617.
- [8] Park U, Tong Yiyang, Jain A K. Age-invariant face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(5): 947-954.