

# 基于深度特征的快速人脸图像检索方法

李振东<sup>1,2\*</sup>, 钟勇<sup>1,2</sup>, 陈蔓<sup>1,2</sup>, 曹冬平<sup>1,2</sup>

<sup>1</sup>中国科学院成都计算机应用研究所, 四川 成都 610041;

<sup>2</sup>中国科学院大学, 北京 100049

**摘要** 针对计算机视觉领域的人脸图像检索问题,提出了一种基于深度特征的快速人脸图像检索方法。该方法首先使用人脸图像训练集对深度卷积神经网络模型进行人脸分类训练;在此基础上采用三元组损失方法对已训练好的人脸分类网络模型进行微调,使得网络能够更加有效地提取人脸特征构建高效的特征向量进行人脸检索初步过滤;最后,为了提高系统检索性能,提出一阶段查询扩展方法对待检索人脸图像特征向量进行融合加强。在两个公用人脸数据集(CASIA-3D FaceV1 和 Labeled Faces in the Wild dataset)上进行详尽的实验验证,结果表明,基于深度特征的人脸图像检索方法不仅能够显著提高检索结果的准确率,而且该方法简单可靠,能够快速实现人脸检索任务。

**关键词** 图像处理; 图像检索; 人脸检索; 卷积神经网络; 特征向量

中图分类号 TP391.2

文献标识码 A

doi: 10.3788/AOS201838.1010004

## Fast Face Image Retrieval Based on Depth Feature

Li Zhendong<sup>1,2\*</sup>, Zhong Yong<sup>1,2</sup>, Chen Man<sup>1,2</sup>, Cao Dongping<sup>1,2</sup>

<sup>1</sup>Chengdu Institute of Computer Application, Chinese Academy of Sciences,  
Chengdu, Sichuan 610041, China;

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract** In order to solve the problem of face image retrieval in the field of computer vision, a face image retrieval method based on the deep features is proposed. Firstly, the convolutional neural network model is trained for face classification by face image training data set. Based on this, the triplet loss method is used to fine-tuning the trained face classification network model so that the network can be more efficient to extract face features of different people and construct efficient feature vectors for preliminary face retrieval filtering stage. In order to further improve the performance of system retrieval, the one-stage query expansion method is proposed to reconstruct the eigenvectors of face images to be retrieved. Through exhaustive experimental verification on two public face datasets (CASIA-3D FaceV1 and Labeled Faces in the Wild dataset), the results show that the face image retrieval method based on deep features improves the accuracy of the retrieval results significantly. Moreover, this method is simple and reliable, and can quickly realize the task of face retrieval.

**Key words** image processing; image retrieval; face retrieval; convolutional neural network; eigenvector

**OCIS codes** 100.2000; 100.3008; 200.4260

## 1 引 言

人脸图像检索是在人脸图像数据库中查找与给定查询内容相同的人脸图像,是计算机视觉领域中的一个基本的研究问题。特别是为了从大型人脸图像数据库中提取有用的信息,这些人脸图像需要进行组织并且是可搜索的。但是,随着待检索对象数量的不

断增加,数据库的规模也不断增大,形成了海量的人脸图像数据库。因此,如何提高海量人脸图像数据库的检索速度和准确率,是大型人脸图像数据库检索向实用化发展过程中必须解决的问题。

针对人脸图像数据库的快速检索问题,刘小华等<sup>[1]</sup>提出了基于  $L-K$  均值层次聚类算法,此算法把大型人脸数据库划分成一些子类数据集,对处于类

收稿日期: 2018-01-22; 修回日期: 2018-03-28; 录用日期: 2018-05-25

基金项目: 四川省科技厅科技成果转化项目(2014CC0043)、四川省科技创新苗子工程项目(SCMZ2006012)

\* E-mail: lizhendong13@mails.ucas.ac.cn

边界的数据,采用冗余技术和预设阈值再重新分配到相应的类中,从而使检索过程只在一个或几个子类中进行,但是该算法随着类内数据量和检索节点的增加导致检索时间显著增加,无法实时检索。杨之光等<sup>[2]</sup>提出了一种基于聚类的家庭数码相机人脸图像检索算法,该算法利用归一化分割在每个时间段内分别对人脸图像进行聚类,使同一个人不同情况下的人脸图像聚为一类。其次采用连续AdaBoost 算法学习得到的人脸识别分类器度量人脸之间的相似度,并进一步提出查询人脸与人脸聚类之间的相似度用于检索。该算法在小的家庭相册人脸检索准确率和时效方面得到了一定的提升,但无法满足海量人脸图像检索的实用性和可靠性要求。

近年来,深度卷积神经网络(CNN)已被证明可以在许多计算机视觉任务中实现最优的性能,如图像分类<sup>[3-5]</sup>、目标检测和跟踪<sup>[6-8]</sup>,以及人脸识别<sup>[9-10]</sup>等任务。深度卷积神经网络经过大量数据训练后的用于学习特征表示的模型可以被泛化使用,甚至用于解决未被训练的任务<sup>[11]</sup>。特别是对于图像检索,Yandex 等<sup>[12]</sup>采用了基于图像分类任务预先训练的CNN 模型,其解决方案为通过从深度卷积神经网络的卷积层提取图像的深度卷积特征进行检索,实现了在流行的检索基准上达到最好的性能,但是该方案在提取深度卷积特征后需要一定的时间对提取到的图像深度卷积特征进行相应的特征池化、降维,以及正则化以实现深度卷积特征的聚合操作。

综上所述,鉴于 CNN 在图像分类、人脸检测和人脸识别等领域的突出贡献,本文提出一种基于深度卷积神经网络的快速可靠人脸图像检索方法。该

方法通过深度卷积神经网络提取人脸图像深度特征,构建人脸图像检索的特征向量,弥补了手工设计特征表征图像效率低、检索准确度差和时效性差等缺陷。该方法利用深度卷积神经网络提取数据库中保存的人脸图像及待检索人脸图像的深度特征,构建人脸图像的深度特征向量描述符,并将其与数据库中保存的人脸图像的深度特征向量描述符进行相似度度量比较得到初始的检索结果列表;在此基础上,采用查询扩展(QE)方法将初始检索结果列表中排序前  $m$  个人脸图像的深度特征向量描述符与待检索人脸图像的深度特征向量描述符进行均值融合后再次执行检索操作,进一步提高系统的检索性能。

## 2 人脸图像检索方法

设计深度卷积神经网络并使用人脸图像数据集对网络模型进行人脸分类训练,使得网络能够有效地提取人脸图像深度特征进行人脸图像检索任务,算法整体结构如图 1 所示。首先通过对人脸图像进行深度卷积运算提取人脸图像高维语义特征,然后将卷积特征激活响应值作为输入送入网络最后两层全连接层进一步提取人脸图像高维稀疏特征,接着对网络最后一层全连接层提取的人脸图像高维稀疏特征进行预处理编码得到最终的人脸图像深度特征向量,将得到的人脸图像深度特征向量与数据库中人脸图像集的深度特征向量进行相似度计算执行人脸检索任务得到初始检索结果,进一步采用查询扩展方法对待检索的人脸图像特征向量进行均值融合后再次执行检索,得到最终的人脸图像检索结果。

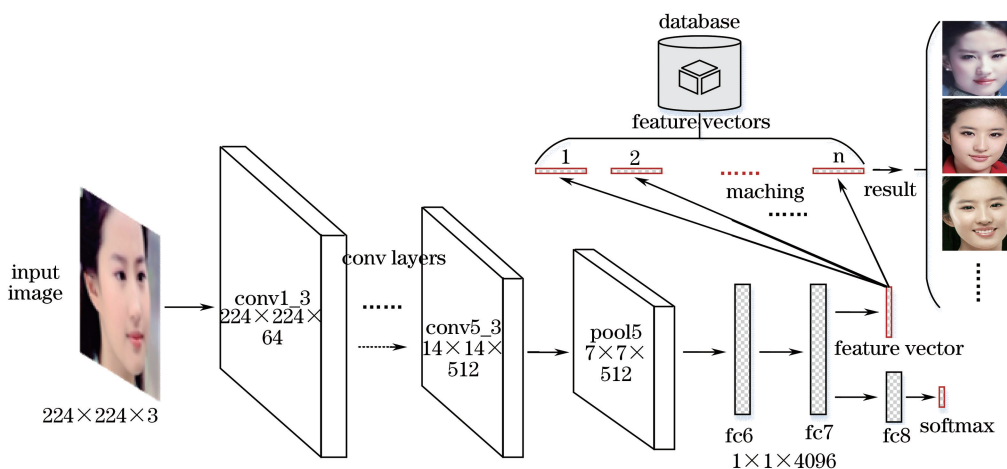


图 1 人脸图像特征提取

Fig. 1 Face image feature extraction

### 2.1 人脸特征提取

对于给定像素大小为  $W_1 \times H_1$  的输入图像  $I$  的卷积层激活响应值是一个三维张量  $W \times H \times K$ ,  $K$  表示输出特征通道数,即多维滤波器。空间像素  $W \times H$  是由网络结构和输入图像像素大小决定的。图 1 所示网络结构中,每个卷积层都利用前一层的输出作为本层的输入,定义为

$$\mathbf{h}_{W,b}^{m(l)}(x) = f\left(\mathbf{b}^{m(l)} + \sum_n \mathbf{W}^{mn(l)} * \mathbf{h}_{W,b}^{n(l-1)}\right), \quad (1)$$

式中  $\mathbf{h}_{W,b}^{m(l)}$  和  $\mathbf{h}_{W,b}^{n(l-1)}$  分别为  $l$  层的第  $m$  个输出通道和  $l-1$  层的第  $n$  个输入通道,  $\mathbf{W}^{mn(l)}$  和  $\mathbf{b}^{m(l)}$  为相应的卷积核滤波器和偏置项,符号  $*$  为卷积运算符。为了保证所有前景元素的非负性,卷积层激活函数选择校正线性激活函数<sup>[3]</sup>  $f(x)$ :

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (2)$$

由于从卷积层直接提取的图像特征数据量大、数据维度高,容易出现过拟合及计算量大等问题,因此在相应的卷积层后采用最大特征值池化方法对卷积层的不同位置进行特征聚合统计:

$$\mathbf{h}_{(i,j)}^{m(l)} = \max_{\forall (p,q) \in \Omega(i,j)} \{\mathbf{h}_{(p,q)}^{m(l)}\}, \quad (3)$$

式中  $\Omega(i,j)$  为索引为  $(i,j)$  的特定区域,  $(p,q)$  为在  $\Omega$  上具体位置的索引。

通过上述方法在网络的最后一层卷积层(Conv5\_3)后同样执行最大特征值池化操作(pool5)得到尺寸大小为  $7 \times 7 \times 512$  维的人脸图像特征激活响应值矩阵,并将该矩阵作为网络全连接层(fc6)的输入,进行人脸图像稀疏特征向量的提取。

### 2.2 人脸特征相似度度量

人脸图像特征向量基于聚合人脸图像  $I$  未经过处理的深度特征。如图 1 所示,从深度卷积神经网络的全连接层 fc7 层提取维度为 4096 维的人脸图像  $I$  的稀疏特征向量  $\mathbf{v}(I) = [x_1, x_2, \dots, x_i, \dots, x_{4096}]$ 。然后,将人脸图像  $I$  的稀疏特征向量进行  $l_2$  正则化得到最终用以人脸图像检索任务的深度特征向量:

$$\mathbf{v}(I) = \frac{\mathbf{v}(I)}{\|\mathbf{v}(I)\|_2}. \quad (4)$$

通过上述方法便可以从深度卷积神经网络的全连接层提取待检索人脸图像  $I'$  的深度稀疏特征向量  $\mathbf{v}(I')$  和数据库中保存的人脸图像数据集  $\mathbf{Z} = \{I_1, \dots, I_k, \dots, I_N\}$  所对应的深度特征向量  $\{\mathbf{v}(I_1), \dots, \mathbf{v}(I_k), \dots, \mathbf{v}(I_N)\}$ ,  $k$  表示数据库中第  $k$  张人脸,  $N$  的大小代表人脸数据库中数据集的大

小。得到相应的人脸图像深度稀疏特征向量后,计算待检索的人脸图像  $I'$  的深度稀疏特征向量  $\mathbf{v}(I')$  和数据库中保存的人脸图像数据集对应的深度特征向量  $\{\mathbf{v}(I_1), \dots, \mathbf{v}(I_k), \dots, \mathbf{v}(I_N)\}$  之间的余弦相似度:

$$d = \text{sim}[\mathbf{v}(I'), \mathbf{v}(I_k)] = \frac{\sum \mathbf{v}(I') \times \mathbf{v}(I_k)}{\sqrt{\sum [\mathbf{v}(I')]^2} \times \sqrt{\sum [\mathbf{v}(I_k)]^2}} = \frac{\sum_{i=1}^{4096} (x'_i \times x_i)}{\sqrt{\sum_{i=1}^{4096} (x'_i)^2} \times \sqrt{\sum_{i=1}^{4096} (x_i)^2}}. \quad (5)$$

根据人脸图像深度特征向量之间的余弦相似度量计算结果的大小进行排序即可得到人脸图像检索初始排序结果列表  $\mathbf{Z}' = \{I'_1, \dots, I'_i, \dots, I'_N\}$  及对应的深度特征向量列表  $\{\mathbf{v}(I'_1), \dots, \mathbf{v}(I'_i), \dots, \mathbf{v}(I'_N)\}$ 。排序结果列表  $\mathbf{Z}'$  中人脸图像  $I'_1$  表示在数据库中保存的人脸图像数据集中与待检索人脸图像  $I'$  最相似、检索结果最匹配的人脸图像。相反,人脸图像  $I'_N$  则表示与待检索人脸图像  $I'$  最不相似、检索结果最不匹配的人脸图像。

### 2.3 特征均值查询扩展

根据以上所述,待检索人脸图像特征向量  $\mathbf{v}(I')$  与检索结果列表中的人脸图像特征向量之间满足距离函数:

$$d[\mathbf{v}(I'), \mathbf{v}(I'_k)] \leq d[\mathbf{v}(I'), \mathbf{v}(I'_{k+1})], \quad 1 \leq k \leq N. \quad (6)$$

在检索结果列表  $\mathbf{Z}'$  中排在最前面的人脸图像是与待检索的人脸图像  $I'$  最相似的人脸图像,由于待检索的人脸图像  $I'$  的深度特征向量  $\mathbf{v}(I')$  不能完全表征同一人在不同自然环境下的人脸图像,因此将初始检索结果排序前  $m$  ( $1 \leq m \leq 10$ ) 个人脸图像的深度特征进行融合,求得均值特征,能够对同一人的脸图像进行一定的表征,利用该均值特征向量作为待检索人脸图像  $I'$  的特征向量  $\mathbf{v}(I')$  进行查询扩展以提高人脸图像检索结果的准确率:

$$\mathbf{v}(I') = \frac{1}{m+1} [\mathbf{v}(I') + \sum_{i=1}^m \mathbf{v}(I'_i)]. \quad (7)$$

通过人脸图像的深度特征均值查询扩展方法将均值特征作为待检索人脸图像  $I'$  的深度特征  $\mathbf{v}(I')$  再次执行检索任务得到最终的检索结果列表  $\mathbf{Z}' = \{I'_1, \dots, I'_i, \dots, I'_N\}$ 。

人脸图像检索一阶段查询扩展方法对待检索人

脸图像的特征向量进行融合加强,能够进一步提高检索结果的准确率。

### 3 人脸检索网络模型训练

为了更好地获得人脸图像的特征表示,提高人脸图像检索结果的性能,受到 Simonyan 等<sup>[4]</sup>提出的用于 ImageNet 大型图像数据集分类的深度卷积神经网络模型的启发,实验中采用人脸图像数据集对图 1 中所示的含有 13 层卷积层和 3 层全连接层的深度卷积神经网络进行训练,使得网络更加有效地表达人脸图像特征,提高人脸图像检索的准确率。

#### 3.1 人脸分类训练

考虑深度神经网络  $f$  识别  $N$  个不同人的脸问题,将网络输出设置为  $N$  路分类问题。深度卷积神经网络通过包含  $N$  个线性预测器(每个标识一个类别)的全连接层(图 1 中 fc8 层)将每个训练图像  $I_t, t=1, \dots, T$  与分数向量  $s_t = \mathbf{W}f(I_t) + \mathbf{b}$  相关联,其中  $\mathbf{W} \in \mathbb{R}^{N \times D}$  为全连接层 fc8 层的权重,  $\mathbf{b} \in \mathbb{R}^N$  为网络偏置项。通过计算 softmax 对数损失  $L(f)$ , 将这些得分与真实标注类别  $C_t \in \{1, \dots, N\}$  进行比较, softmax 对数损失  $L(f)$  计算如下:

$$L(f) = - \sum_t \{ \log [\exp(\langle c_t, \mathbf{x}_t \rangle)] - \log [\sum_{q=1, \dots, N} \exp(\langle c_q, \mathbf{x}_t \rangle)] \}, \quad (8)$$

式中  $\mathbf{x}_t = f(I_t) \in \mathbb{R}^D$  为类别分数向量。

网络模型训练学习之后,可以使用欧氏距离比较分数向量  $\mathbf{x}_t = f(I_t)$  进行人脸验证。虽然,使用三元组损失方案微调网络模型在欧氏空间中进行验证,可以显着提高得分,获得良好的整体性能。但是,将网络首先作为分类器进行预训练会使后续微调训练变得更加容易和快速。

#### 3.2 三元组损失微调网络

三元组损失微调网络模型学习能够在最终应用中表现良好的分数向量,即通过在欧氏空间中比较人脸特征描述符进行身份验证。与众多度量学习方法相似,三元组损失用于学习具有独特性和紧凑性的投影,同时可实现降维。

实验中所使用的三元组损失微调方法与文献<sup>[9]</sup>和<sup>[10]</sup>相似,一个三元组  $(a, p, n)$  包含同一个人的人脸图像的训练样本  $a$ , 以及一个与  $a$  同一人的人脸图像的正样本  $p$ , 和一个与  $a$  不同人的脸图像的负样本  $n$ 。如 3.1 节所述,预先训练的 CNN 的

输出  $f(I_t) \in \mathbb{R}^D$  经过  $l_2$  正则化,并使用仿射投影:

$$\mathbf{x}_t = \mathbf{W}' \frac{f(I_t)}{\|f(I_t)\|_2}, \mathbf{W}' \in \mathbb{R}^{L \times D}, \quad (9)$$

将  $f(I_t)$  投影到更低维  $L (L < D)$  维空间。与 3.1 节的线性预测器不同的是,  $L$  不等于类别的数量,但其为描述符嵌入的大小(实验中设置  $L = 1024$ )。投影  $\mathbf{W}'$  被训练最小化三元组损失:

$$\begin{cases} L(\mathbf{W}') = \sum_{(a,p,n) \in T} \max\{0, \alpha + \|\mathbf{x}_a - \mathbf{x}_p\|_2^2 - \|\mathbf{x}_a - \mathbf{x}_n\|_2^2\}, \\ \mathbf{x}_i = \mathbf{W}' \frac{f(I_i)}{\|f(I_i)\|_2} \end{cases}, \quad (10)$$

式中  $\alpha \geq 0$  为学习边界的固定标量,  $T$  为训练三元组的集合。三元组损失的目的是通过学习,让训练样本  $a$  和  $p$  特征表达之间的距离尽可能小,而  $a$  和  $n$  的特征表达之间的距离尽可能大,并且要让  $a$  和  $n$  之间的距离和  $a$  和  $p$  之间的距离之间有一个最小的间隔,即  $\alpha$ , 公式化的表示为

$$\alpha + \|\mathbf{x}_a - \mathbf{x}_p\|_2^2 < \|\mathbf{x}_a - \mathbf{x}_n\|_2^2. \quad (11)$$

当  $a$  和  $n$  之间的距离小于  $a$  和  $p$  之间的距离加  $\alpha$  时,就会产生损失;当  $a$  和  $n$  之间的距离大于等于  $a$  和  $p$  之间的距离加  $\alpha$  时,损失为零,如(10)式所示。

## 4 实验与结果分析

所有实验均在以下平台实现:CPU 为 Intel(R) Core(TM) i5-6600 CPU @ 3.30 GHz,内存大小为 16 G,12 G 内存的 GPU Nvidia Titan X 平台,操作系统为 Ubuntu 14.04.4 LTS,代码是在 Caffe 环境下采用 python 编程语言实现的。

#### 4.1 实验数据集

采用以下两个人脸图像数据集对本文方法进行性能评估:

1) Labeled Faces in the Wild dataset (LFW)<sup>[14]</sup>: 该数据集包含 5749 个不同人的 13233 张人脸图像,是人脸图像识别的标准验证基准。

2) CASIA-3D FaceV1: CASIA-3D FaceV1 (<http://biometrics.idealtest.org/>)是由中国科学院自动化研究所收集的由 123 人组成的 3D 人脸数据集,该数据集在构建过程中,不仅考虑姿态、表达和照明的单一变化,而且还考虑照明下的表情和表情的组合变化,部分人脸图像如图 2 所示。



图 2 人脸图像示例

Fig. 2 Examples of face images

#### 4.2 实验参数设置

学习  $N$  路人脸分类器(3.1 节)遵循文献[3-4]的步骤和修改建议。目标是找到最小化 softmax 层之后的平均预测对数损失的网络参数。首先是网络的配置过程,通过随机梯度下降和 0.9 的动量系数<sup>[15]</sup>优化。权重衰减系数为  $5 \times 10^{-4}$ ,在 fc6 和 fc7 两个全连接层之后以 0.5 的比率将节点输出置 0。学习率初始值设置为  $10^{-2}$ ,在验证集精度停止增加时减少到  $10^{-3}$ 。网络滤波器的权重通过零均值和标准偏差为  $10^{-2}$  的高斯分布随机采样初始化。网络偏置项初始化为零。

#### 4.3 实验结果验证

使用人脸分类训练和三元组损失微调后的 CNN 在 CASIA-3D FaceV1 和 LFW 两个数据集上进行了人脸检索性能验证评估。将实验分为两种情况进行验证,如表 1 所示。表 1 中 QE 表示查询扩展, $m$  表示采用初始检索结果列表中的前  $m$  个人脸图像的深度特征的均值进行查询扩展;Speed/image 表示检索每张人脸图像所用的总时间(单位为 s);top20-AP 表示检索结果前 20 的平均准确率,mAP

表示所有待检索人脸图像的正样本被正确检索的平均准确率。

首先在 CASIA-3D FaceV1 数据集上进行了人脸检索和采用一阶段查询扩展方法的人脸检索性能评估。从表 1 第二行可以看出,采用查询扩展方法可以进一步提高检索结果的性能,当选取  $m=10$  的情况下取得最优的检索结果,并且没有牺牲太多的检索时间代价去换取检索准确率,mAP 和 top20-AP 分别提高到 98.86%和 99.52%,可见提出的人脸检索方法性能具有一定的可靠性和时效性。

为了进一步验证所提人脸检索方法的可靠性和时效性,将 CASIA-3D FaceV1 和 LFW 两个人脸图像数据集进行合并得到一个更大的人脸数据集,并在合并后的人脸数据集上执行人脸检索性能和时效性验证。从表 1 第三行可以看出,经过在扩展后的人脸数据集上进行人脸检索验证仍然可以得到稳定的检索结果和时效性。因此,本文人脸检索方法具有一定的可靠性和时效性,随着人脸数据集的增加,检索结果性能稳定,时间开销并未明显增加。

表 1 人脸检索结果

Table 1 Face retrieval results

ImageSet	QE	Speed /s <sup>-1</sup>	Top20-AP	mAP
CASIA-3D FaceV1	No	0.081	98.19	95.44
	$m=5$	0.085	99.3	98.1
	$m=10$	0.089	99.52	98.86
CASIA-3D FaceV1+LFW	No	0.1	98.19	95.44
	$m=5$	0.105	99.29	98.06
	$m=10$	0.118	99.52	98.84

在图 3 中列出了三组人脸图像检索结果列表中排序前 15 张的检索结果示例比较图,其中每行的第一列人脸图像包围边框为蓝色的表示待检索的人脸图像,第一列后的包围边框为绿色的人脸图像表示相对应的被正确检索的结果,包围边框为红色的人脸图像表示被错误检索的结果。每组图像上面一排是不

使用查询扩展检索的快速检索结果,下面一排是采用第一阶段快速检索结果列表中前 10( $m=10$ )幅图像进行特征融合后进行查询扩展的检索结果。从图中可以看出,采用查询扩展方法,将待检索图像的深度特征和最相似的 10 张图像的深度特征进行融合再次执行查询,能够进一步提高检索的准确率。



图 3 部分人脸检索结果。(a)待检索的人脸图像(蓝色框);(b)正确检索的结果(绿色框);(c)错误检索的结果(红色框)

Fig. 3 Part of the face retrieval results. (a) Face image to be retrieved (blue box);

(b) correctly retrieved results (green box); (c) error retrieved results (red box)

#### 4.4 实验结果及时效性比较

将本文人脸图像检索方法与其他文献中的检索方法进行比较,表 2 所示为算法在 CASIA-3D FaceV1+LFW 人脸数据集上执行检索结果的平均准确率和检索每张人脸图像所用平均时间的时效性对比。

表 2 结果及时效性比较

Table 2 Results and timeliness comparison

Method	mAP	Speed /s <sup>-1</sup>
L-K one node <sup>[1]</sup>	77.5	0.33
L-K two node <sup>[1]</sup>	86.0	0.67
L-K three node <sup>[1]</sup>	91.0	1.05
L-K four node <sup>[1]</sup>	93.5	1.32
L-K five node <sup>[1]</sup>	95.1	1.65
Ref. [13] method	92.11	0.331
Ref. [13] method+QE	95.7	0.338
DCNNFR	95.44	0.1
DCNNFR+QE	98.84	0.118

如表 2 中实验结果所示,刘小华等<sup>[1]</sup>提出的用于人脸图像检索的 L-K 均值层次聚类方法在随着类内数据量和检索节点的增加时所得到的检索结果明显提高,在没有时间限制的情况下最好的检索准确率可以达到 95.1%,但是该方法随着检索准确率的提升,时间消耗显著增加,其是以时间开销为代价换取检索平均准确率。因此该方法在现实应用场景中,在保证一定检索准确率的同时,不能实现快速实时的检索。Tolias 等<sup>[13]</sup>提出的基于图像分类任务预先训练的 CNN 模型,通过从深度卷积神经网络的卷积层提取图像的深度卷积特征执行检索任务,从表 2 中可以看出,该方法在保证一定检索准确率的同时,提高了检索速度,但是由于该方法在提取

深度卷积特征后,还需要对提取到的图像深度卷积特征进行相应的特征池化、降维以及正则化等特征聚合操作,这将增加一定的时间开销,从而影响检索速度。而本文提出的方法通过训练好的深度神经网络人脸检索(DCNNFR)模型直接提取网络全连接层 fc7 层特征,平均提取和构建每张人脸图像的深度特征向量只需要 0.02 s,然后通过待检索人脸图像的特征向量与数据库中保存的人脸图像集的特征向量之间的余弦相似度度量和查询扩展方法进行快速的人脸检索,从表 2 中可以看出,该方法在保证一定检索准确率的同时能够达到很好的时效性,检索过程简单快速无需人为参与设计特征的提取和检索过滤。

从表 2 中可以看出,本文方法平均准确率可以达到 98.84%,而 Tolias 等提出的方法最高的平均准确率达到 95.7%,这是因为 Tolias 等提出的方法从深度卷积神经网络提取卷积特征经过特征聚合操作后只有 512 维的人脸图像特征向量表征图像,而本文方法是从深度卷积神经网络的全连接层 fc7 层直接提取 4096 维的人脸图像特征向量表征图像,因此本文所提方法提取人脸图像高维度特征向量更具有表征性,更利于检索任务。

## 5 结 论

提出了一种基于深度卷积特征的人脸图像检索方法。通过使用人脸图像训练集对深度卷积神经网络模型进行人脸分类训练,在此基础上采用三元组损失方法对已训练好的人脸分类网络模型进行微调,使网络能够更加有效地提取人脸特征,构建高效的特征向量进行人脸检索过滤,最后通过查询扩展

方法对待检索人脸图像特征向量进行融合,进一步提高检索性能。实验结果证明,该方法不仅能够显著提高检索结果的准确率,而且简单有效,能够快速地实现人脸图像检索任务。

## 参 考 文 献

- [1] Liu X H, Zhou C G, Zhang L B, *et al.* Method of quick searching in a huge scale face database [J]. Journal of Jilin University (Engineering and Technology Edition), 2010, 40(1): 183-188.  
刘小华, 周春光, 张利彪, 等. 海量人脸数据库的快速检索[J]. 吉林大学学报(工学版), 2010, 40(1): 183-188.
- [2] Yang Z G, Ai H Z. Cluster-based face image retrieval and its relevance feedback [J]. Acta Automatica Sinica, 2008, 34(9): 1033-1039.  
杨之光, 艾海舟. 基于聚类的人脸图像检索及相关反馈[J]. 自动化学报, 2008, 34(9): 1033-1039.
- [3] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]//Proceedings of the 25<sup>th</sup> International Conference on Neural Information Processing Systems, 2012: 1097-1105.
- [4] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv: 1409.1556, 2014.
- [5] Liu D W, Han L, Han X Y. High spatial resolution remote sensing image classification based on deep learning [J]. Acta Optica Sinica, 2016, 36 (4): 0428001.  
刘大伟, 韩玲, 韩晓勇. 基于深度学习的高分辨率遥感影像分类研究 [J]. 光学学报, 2016, 36 (4): 0428001.
- [6] Zhang K P, Zhang Z P, Li Z F, *et al.* Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.
- [7] Li H X, Lin Z, Shen X H, *et al.* A convolutional neural network cascade for face detection [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5325-5334.
- [8] Lin S Z, Zheng Y, Lu X F, *et al.* Adaptive tracking algorithm for aerial small targets based on multi-domain convolutional neural networks and autoregression model[J]. Acta Optica Sinica, 2017, 37(12): 1215006.  
蔺素珍, 郑瑶, 禄晓飞, 等. 基于多域卷积神经网络与自回归模型的空中小目标自适应跟踪方法 [J]. 光学学报, 2017, 37(12): 1215006.
- [9] Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition[C]//British Machine Vision Conference, 2015: 1-12.
- [10] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 815-823.
- [11] Razavian A S, Azizpour H, Sullivan J, *et al.* CNN features off-the-shelf: An astounding baseline for recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014: 512- 519.
- [12] Yandex A B, Lempitsky V. Aggregating local deep features for image retrieval[C]//Proceedings of IEEE International Conference on Computer Vision, 2015: 1269-1277.
- [13] Tolias G, Sivic R, Jégou H. Particular object retrieval with integral max-pooling of CNN activations [J]. arXiv preprint arXiv: 1511.05879, 2015.
- [14] Huang G B, Mattar M, Berg T, *et al.* Labeled faces in the wild: a database for studying face recognition in unconstrained environment [C] // Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition. 2008.
- [15] LeCun Y, Boser B, Denker J S, *et al.* Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541-551.