带相关反馈的基于深度神经网络模型的人脸检索方法①

沈旭东[1], 范守科[2], 夏海军[3], 苏金波[3]

[1](中国科学技术大学 自动化系, 合肥 230022)
[2](中国人民解放军 63791 部队, 西昌 615000)
[3](合肥市公安局 网安支队, 合肥 230022)

摘 要: 针对大规模人脸检索问题, 提出了一种带相关反馈的基于深度神经网络模型的人脸检索方法. 首先利用卷积神经网络对人脸进行特征提取, 再利用传统的检索方法进行人脸检索, 在检索环节之后加入相关反馈环节. 根据用户反馈的结果, 将样本分成正例和负例, 作为反馈环节的训练样本, 完成反馈环节的训练. 实验表明, 该方法能够显著提高人脸检索的准确率.

关键词: 人脸检索; 卷积神经网络; 哈希检索; 相关反馈

Face Retrieval Method Base on Deep Neural Networks with Relevance Feedback

SHEN Xu-Dong[1], FAN Shou-Ke[2], XIA Hai-Jun[3], SU Jin-Bo[3]

[1](Department of Automation, USTC, Hefei 230022, China)
[2](63791 Unit of PLA, Xichang 615000, China)
[3](Public Security Bureau of Hefei, Hefei 230022, China)

Abstract: In this paper, a face retrieval method based on deep neural networks with relevance feedback has been presented to solve the problem of large-scale human face retrieval. Firstly, convolutional neural networks has been used for feature extracting, then, traditional search methods will be used in face retrieval. A feedback model is added after the retrieval stage. According to the users' feedback, result samples are divided into positives and negatives, which are be used for training the feedback model. Experiments show that this method can significantly improve the accuracy of face retrieval.

Key words: face retrieval; CNN; hash retrieval; relevance feedback

近些年来, 随着计算机技术的不断发展, 计算机视觉(CV)问题越来越受到人们的关注, 例如物体识别[1,2]、图像检索[3,4]、图像匹配[5,6]等, 而在所有的计算机视觉问题中, 人脸识别与检索方法由于其与人身份的密切联系而受到研究者更加广泛的关注.

目前的人脸检索方法主要包括三个部分, 人脸图像预处理, 人脸特征提取, 特征检索. 而这其中人脸特征提取部分得到的人脸特征的优劣直接决定整个人脸检索系统的性能. 也正是由于这一点, 多年来研究者们纷纷提出了多种多样的特征提取方法.

总结这些特征提取方法, 主要有两个研究方向, 一是人工设计特征, (如 LBP[7], SIFT[8]等), 另一个是学习特征. 人工设计特征是根据图像自然具有的颜色, 纹理, 形状等特征, 通过一定的数学方法, 设计出来的一种特征抽取方法, sift 特征便是这其中较为出色的特征抽取方法. 人工特征虽然具有理论基础清晰的优点, 但是, 人工特征的设计需要大量的理论知识和深厚的数学功底, 这制约了该方法的进一步发展.

2006 年, 以 Geoffrey Hinton 在 Science 发表文献[9], 提出深度置信网络(Deep Belief Networks, DBN)可使用非监督的逐层贪心算法来训练为标志, 研究人员开始将深度学习用于图像特征提取, 并在图像分类问题上

① 基金项目:中国科学院"NGB 有线无线融合应用"重点部署项目子课题(KGZD-EW-103-5(5));国家科技支撑计划子课题(2012BAH73F02);安徽省科技攻关项目(1301b042012);"核高基"重大专项(2012ZX01034-00-001)

收稿时间:2015-04-16;收到修改稿时间:2015-06-03

取得了惊人的效果.

2014 年, Xiaogang Wang, Xiaoou Tang 等人发表文章[10],利用多层卷积神经网络提取人脸图像的特征, 并在 LFW 上验证其分类效果, 实验表明, 文中提出的深度网络进一步提高了人脸分类的准确率.

2014 年, Xiaogang Wang, Xiaoou Tang 等人发表文章[10], 利用多层卷积神经网络提取人脸图像的特征, 并在 LFW 上验证其分类效果, 实验表明, 文中提出的深度网络进一步提高了人脸分类的准确率.

基于上述人脸特征提取方法, 本文提出了一种带相关反馈的深度学习人脸检索方法, 该方法设计了一种多层的 CNN 网络, 利用打好类别标签的人脸图片数据集训练该网络, 此深度网络能提取人脸图像的特征, 基于此特征, 再利用传统的检索方法, 得出待检索人脸的检索结果. 我们发现, 该结果虽然比以往的基于人工特征的检索方法具有更好的检索准确率, 但是仍然具有较大的提升空间, 因此, 在检索之后, 加入反馈环节, 利用相关反馈算法获取带标签数据, 对该反馈网络进行训练, 最终得到一个带反馈的深度学习网络.

1　相关概念

1.1　卷积神经网络

卷积神经网络 Convolutional Neural Networks (CNN)是一种特殊的深层神经网络模型, 它的特殊性体现在两个方面, 一方面它的神经元间的连接是非全连接的, 另一方面, 同一层中某些神经元之间的连接权重是共享的. 卷积网络是为识别二维形状而特殊设计的一个多层感知器, 这种网络结构对平移、比例缩放、倾斜或者共他形式的变形具有高度不变性.

1.2　相关反馈算法

人脸检索领域的反馈即是使用一种判别标准(如人工判断)对检索结果的正确性进行判别, 再将判别结果回送到检索系统, 优化检索系统参数, 从而起到对检索结果不断修正的作用.

相关反馈算法, 一方面, 通过对最佳的查询方向估计来调整查询的方向, 使其不断向用户反馈的正例靠近, 而远离反例; 另一方面, 利用反馈信息修改距离公式中各分量的权值, 突出重要的分量[11,12].

2　带相关反馈的基于深度神经网络模型的人脸检索方法

根据文献[10]的思想, 我们设计了一个 8 层的卷积神经网络, 利用这一网络结构来实现人脸图像特征的提取. 网络结构包含 1 个输入层, 2 个卷积层, 2 个下采样层, 2 个全连接层和 1 个输出层. 网络中的卷积层和下采样层是经过专门设计来提取局部特征和全局特征的. 最终抽取出一个 256 维的特征向量用来表示输入的人脸图片. 网络结构如图 1 所示.

为了训练 CNN 网络, 使参数达到最优, 假设有一个固定样本集 $\{(x^{(1)},y^{(1)}),...,(x^m,y^m)\}$, 它包含 m 个样本, 我们使用批量梯度下降法来训练神经网络. 对于单个样本, 定义代价函数如式(1)所示.

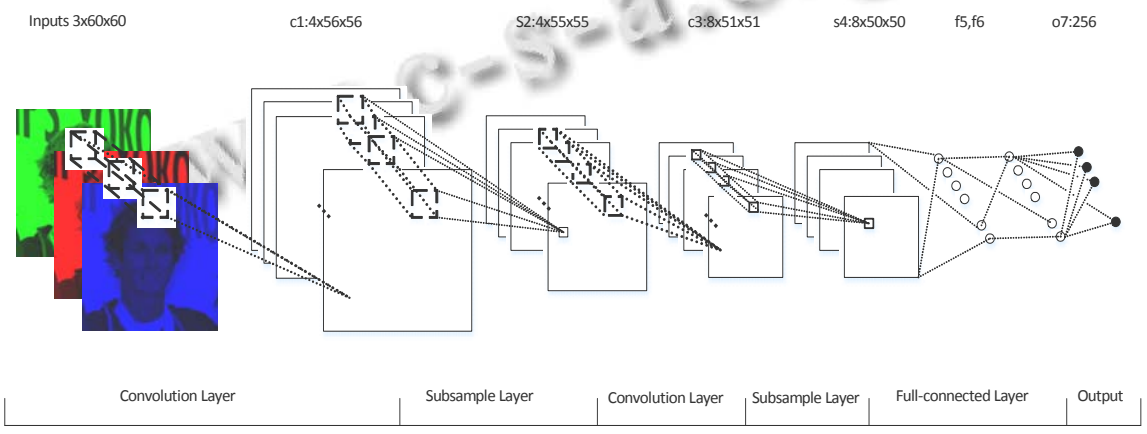

图 1　CNN 网络结构

$$J(W,b;x,y)=\frac{1}{2}\left\|h_{W,b}(x)-y\right\|^2 \tag{1}$$

对于一个 m 个样本的数据集, 定义整体代价函数如式(2)所示.

$$J(W,b)=\left[\frac{1}{m}\sum_{i=1}^{m}J\left(W,b;x^i,y^i\right)\right]+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}\left(W_{ji}^{(l)}\right)^2 \tag{2}$$

上述公式中 $h_{w,b}(x)$ 是网络的实际输出, y 是训练

标签值，λ 为权重衰减参数，用来控制公式中两项的相对重要性，W 为网络的连接权值，b 为偏置值，训练的目的就是为了获取最佳的 W 和 b.

基于上述网络结构得到的人脸特征进行人脸检索，我们利用文献[13]中的有监督哈希检索方法. 该文献的思想是将高维数据投影成二进制码，通过对带有相关性标签的训练样本对的学习，相似样本对之间的汉明距离最小，而不相似的样本对之间的汉明距离最大.

将上述方法应用到人脸检索中，使用哈希方法获得待检索人脸样本的哈希编码，再计算这个哈希编码与检索库中其他检索样本哈希编码之间的汉明距离，通过距离的大小来判断检索库中哪些样本是与待检索样本相似的结果.

选择人脸数据集 $\chi=\{x_1,\ldots,x_n\}\subset \mathrm{R}^d$，其中 x_i 是 d 维人脸特征样本. 利用核函数 $k:R^d\times R^d\mapsto R$. 首先，根据局部敏感哈希(KLSH)算法. 定义预测函数

$$f(x)=\sum_{j=1}^{m}k(x_j,x)a_j-b \tag{3}$$

其中，$x_1,\ldots,x_m$ 是从 χ 中随机选取的 m 个样本，m 是一个比样本个数 n 小很多的参数，这样做的目的是确保哈希的快速性. $a_j\in R$ 是系数，b 是偏置，基于 f，使用

$$h(x)=\mathrm{sgn}(f(x)) \tag{4}$$

构建每一个哈希比特的哈希函数. 最后，利用上述方法生成的哈希编码，得出与检索目标相似的检索结果.

每次检索过程， 一张待检图片都会得出若干个最为相似的检索结果，而这些结果中有部分是正确的检索结果，而另外一部分则是错误的. 根据文献[11]提到的相关反馈算法，检索用户能够很容易判断这些检索结果的正误，且能够通过简单的操作将这些结果进行分类(正确或者错误). 多次检索会积累一定量的此类数据，以往的检索方法没有考虑这些数据，而经验告诉我们，这些数据应该会对往后的检索结果有帮助. 因此，我们设计了一个反馈环节，利用这些数据去训练反馈环节，不断提升整个系统的检索性能.

相关反馈能够运用于人脸检索，正是由于人脸检索库中存在的人物一般是具有身份标签的，每一个人脸都会属于其中的一个身份的人，也就是属于所有类别中的一类，检索库中存在多少个人也就分成多少个类. 在检索过程中，如果用户判断检索出的结果和用户提交的检索图像属于同一个人，则认为是相关图像，否则认为是无关图像. 所有的检索结果，用户认为相关则标记为正例，无关则标记为负例.

本文采用的方法是首先将待检目标人脸，利用前文提到的方法得出一个初步的检索结果，再根据相关反馈算法，由用户对检索结果进行标定，用户认为结果正确，就标为正例，反之则是负例. 再将这些打过标签的检索结果组成的训练集输入到反馈环节中，训练产生一个反馈分类器，之后的检索结果就可以通过这个反馈分类器，判断出更多正确的结果.

反馈环节是一个分类器，提升反馈环节的性能可以使用提升分类器性能的方法. 在一定范围内提升参与分类器训练的样本、调节分类器参数、使用更加优秀的度量函数都可以达到效果. 由于本文的论述重点在于反馈分类器能够使整个系统获得随着检索结果的不断积累而使性能不断优化的功能. 对反馈分类器的分类效果不满意时，每次检索得出的结果都可以在用户反馈后加入训练集对反馈分类器进行重新训练. 系统性能能够随着检索次数的增加而不断提升. 因此，本文主要通过改变样本数量来仿真系统性能的提升. 详细的算法流程如表 1.

表 1 详细算法步骤

算法: 带反馈的深度卷积神经网络检索方法
输入: 60x60 的三通道人脸图片
处理过程:
S1 将检索样本集中的图片输入到已训练完成的深度卷积神经网络，获取一个 256 维的向量，所有测试样本的特征向量组成图片特征集合 $\chi=\left\{x_i\in R^d\right\}_{i=1}^{n}$，其中 d =256，n 为测试样本个数.
S2 将 χ 中的所有样本利用文献[13]中提到的方法进行哈希映射，获取哈希编码.
S3 将测试样本集中的图片获取哈希编码后，进行检索，得出检索结果
S4 用户对检索出的前 N 个人脸图像进行标记得到相关人脸图像集 I+，无关人脸图像集 I-;
S5 准备反馈环节的训练样本集 (x_i,y_i)，$y_i=\begin{cases}+1 & if\ \ x_i\in I^+\\ -1 & if\ \ x_i\in I^-\end{cases}$
S6 利用训练样本训练反馈分类器：$f(x)=sign(\sum_i T_i y_i k(x_i,x)+b)$
S7 根据分类器的输出结果判断每个样本所属类别. 若为+1 则为正确的检索结果，若为-1 则为错误的检索结果.

3　实验

为了测试本文提出的带相关反馈的基于深度神经网络的人脸检索方法的性能，需要首先对深度卷积神经网络进行训练，本次实验使用的训练集由 LFW 上的部分图片和在互联网上下载的图片组成，图片一共有大约 50000 张．部分图片如图 2 所示．测试数据集我们使用的是 YouTube Faces Database[14]随机选取的 20000 张图片，这些图片包含 1595 个不同的人．分别打上 1 到 1595 的标签，数字相同的表示同一个人．实验结果如下．

图 2　用于训练的图片预览

3.1　输出不同的检索结果数，检索准确率对比

人脸检索是通过输入一张待检索图片，输出用户需要的一系列被检索图片．这里输出检索结果图片数量 range 变化，对检索准确率具有直接的影响．一般来说，检索准确率随着输出结果数增加而下．然而，加入反馈环节之后，能够在一定范围内提升整个系统的检索性能．

本次实验为了验证本文所提方法的上述性能，设计使 range 从 10 变化到 20 过程中，记录加入反馈环节前后检索准确率的变化．

实验结果如图 3 中未加反馈曲线所示，其中横坐标表示 range 的变化，纵坐标表示检索准确率．实验表明，随着输出图片数量(range)的不断增加，未加反馈时检索的准确率不断下降．加入反馈，使用前述相同的数据进行实验，其结果如图 4 中加反馈曲线所示，前后两次结果的对比表明，本文提出的带反馈环节的检索方法在输出多个结果时，依然能够显著提升检索的准确率．

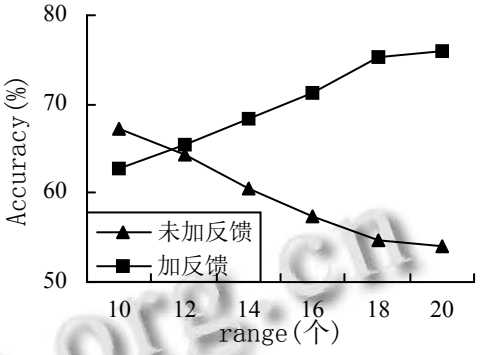

图 3　输出不同结果数对比实验

3.2　不同数量的样本集下反馈环节对检索性能的影响

为了验证样本个数增加对相关反馈算法的性能影响，我们选择测试样本数据集中样本总数分别为 5000,10000,20000,30000 个．再选择样本集中的 80%对哈希检索函数进行训练，20%进行检索测试输出 range=20 的结果，收集这些输出检索结果利用相关反馈算法打上标签，对反馈环节进行训练．记录加相关反馈前后检索准确率．检索准确率如图 4 所示．

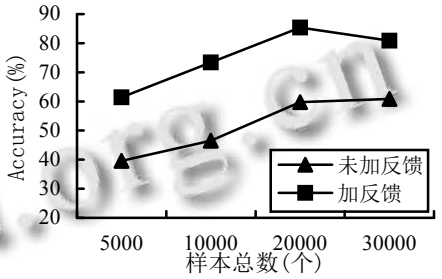

图 4　不同样本下检索准确率对比实验

从图 4 的曲线中，可以发现，未加反馈环节时，随着参与哈希函数训练的样本数据不断增加，检索的准确率也是呈不断上升态势的，因此，提高样本总数，能够提升哈希检索的准确率．但是，无限制地提升样本个数必然会以牺牲检索时间为代价的．另外，获取大量的加标签的人脸图片也是非常困难的工作．而本文方法训练反馈环节的标签样本是多次检索积累下来的，获取比较容易．而加入反馈环节后，实验数据表明，相同的数据量检索准确率有显著提升，且随着数据量的增加，检索准确率也是不断提升的，直到样本

数到达 20000 附近时, 反馈环节参数已达最优, 准确率达到峰值.

4 结语

本文在利用卷积神经网络提取人脸特征并进行人脸检索的基础上, 加入了反馈环节, 利用相关反馈的算法以用户对检索结果是否感兴趣为标准为样本打上标签, 将打过标签的样本输入反馈环节, 训练产生反馈分类器. 经过实验验证, 上述方法能很好的提升系统的检索性能, 从而证明我们提出的方法是有效的.

参考文献

1 Torralba A, Fergus R, Freeman WT. 80 million tiny images: a large data set for nonparametric object and scene recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2008, 30(11): 1958–1970.

2 Torralba A, Fergus R, Weiss Y. Small codes and large image databases for recognition. IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE. 2008. 1–8.

3 Kulis B, Jain P, Grauman K. Fast similarity search for learned metrics. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009, 31(12): 2143–2157.

4 Xu H, Wang J, Li Z, et al. Complementary hashing for approximate nearest neighbor search. 2011 IEEE International Conference on Computer Vision (ICCV). IEEE, 2011. 1631–1638.

5 Korman S, Avidan S. Coherency sensitive hashing. 2011 IEEE International Conference on Computer Vision (ICCV). IEEE. 2011. 1607–1614.

6 Strecha C, Bronstein AM, Bronstein MM, et al. LDAHash: Improved matching with smaller descriptors. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012, 34(1): 66–78.

7 Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971–987.

8 Lowe DG. Object recognition from local scale-invariant features. Proc. of the Seventh IEEE International Conference on Computer vision. IEEE. 1999, 2. 1150–1157.

9 Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science, 2006, 313(5786): 504–507.

10 Taigman Y, Yang M, Ranzato MA, et al. Deepface: closing the gap to human-level performance in face verification. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2014. 1701–1708.

11 许月华,李金龙,陈恩红,等.一种新的基于 SVM 的相关反馈图像检索算法.计算机工程,2005,30(24):116–118.

12 张磊,林福宗.基于支持向量机的相关反馈图像检索算法.清华大学学报(自然科学版),2002,42(1):80–83.

13 Liu W, Wang J, Ji R, et al. Supervised hashing with kernels. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2012. 2074–2081.

14 Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2011. 529–534