

DOI:10.11992/tis.201706025

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20171021.1350.012.html>

人脸识别背后的数据清理问题研究

夏洋洋¹, 龚勋¹, 洪西进^{1,2}

(1. 西南交通大学 信息科学与技术学院, 四川 成都 611756; 2. 台湾科技大学 资讯工程系, 台湾 台北 10607)

摘要:人脸识别技术在深度卷积神经网络(deep convolution neural networks, DCNN)的快速发展下取得了显著的成就。这些成果主要体现在更深层次的 DCNN 架构和更大的训练数据库。然而,由大多数私人公司持有的大型数据库(百万级)并不对外公开,即使当前部分开放的大型数据库,因为标注信息过少,无法保证精度,会影响 DCNN 的训练。本文提出了一种易于使用的多角度清理图像方法来提高数据的准确性:通过人脸检测算法清除掉无法检测到人脸的图像;在清理后的数据集上利用已有模型提取图像特征,并计算相似度,进而统计出一类人脸图像中每一张图像与其他图像不相似的数目,根据改进参数清理数据。实验表明,清理后的数据库训练模型在 LFW 和 Youtube Face 数据集上测试的精度得到了提升,使用较小规模数据集情况下,在 LFW 数据集上取得了 99.17% 的准确率,在 Youtube Face 数据集也达到了 93.53% 的准确率。

关键词:深度卷积神经网络;DCNN;清理图像;人脸识别;大型数据库

中图分类号:TP391.4 **文献标志码:**A **文章编号:**1673-4785(2017)05-0616-08

中文引用格式:夏洋洋,龚勋,洪西进.人脸识别背后的数据清理问题研究[J].智能系统学报,2017,12(5):616-623.

英文引用格式:XIA Yangyang, GONG Xun, HONG Xijin. Research on the data cleansing problem for face recognition technology [J]. CAAI transactions on intelligent systems, 2017, 12(5): 616-623.

Research on the data cleansing problem for face recognition technology

XIA Yangyang¹, GONG Xun¹, HONG Xijin^{1,2}

(1. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China; 2. Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei 10607, China)

Abstract: Face recognition technology has made a significant progress in the rapid development of deep convolution neural networks (DCNN). These developments are mainly focused toward a denser DCNN architecture and larger training database. However, DCNN training is affected because the large-scale database held by most private companies are not publically accessible. Moreover, current large-scale open databases are not accessible because of the slight availability of the labeled information and hard-to-guarantee accuracy. This study presents an easy-to-use image cleansing method to improve the accuracy of data from the following perspectives: First, deleting the face image that cannot be detected by face detection; second, using the existing model to extract the features of an image on the cleaned dataset and calculate the similarity; and finally, counting the number of images that are unlike the other images. The data were cleansed according to the improved parameters extracted from the abovementioned perspectives. The experimental results reveal that the cleansed database training model has improved the accuracy of face recognition in LFW (labeled faces in the wild) and YouTube face database. In the case of using a small-scale dataset, an accuracy of 99.17% and 93.53% was achieved on the LFW and YouTube face datasets, respectively.

Keywords: deep convolution neural network; DCNN; cleansing image; face recognition; large database

构建一个完整的人脸识别系统主要包括图像采集、人脸检测、人脸对齐、特征表示与分类识别等步骤。人脸识别技术的核心在于特征提取和分类

识别,而特征提取则是人脸识别的重点关注点^[1-3]。深度学习与传统人脸识别技术的主要区别在于:深度学习是通过从海量数据中有监督的训练学习来获取能够有效表示人脸信息的特征,不再需要大量的人脸识别先验知识来设计特征,这种人脸面部特征学习更具鲁棒性^[4-6]。深度卷积神经网络(DCNN)作为一种深度学习架构,已成功应用于人脸识别和其他计算机视觉任务,并逐渐成为一种

收稿日期:2017-06-08. 网络出版日期:2017-10-21.

基金项目:国家自然科学基金项目(61202191);计算智能重庆市重点实验室开放基金项目(CQ-LCI-2013-06);国家重点研发计划项目(2016YFC0802209).

通信作者:龚勋.E-mail: xgong@swjtu.edu.cn.

通用算法^[1,4-5,7-10]。虽然 DCNN 是通过从大量数据中自动学习更具区分力的特征,人脸识别的主要目的也是通过从海量数据中学习具有泛化能力的人脸特征,但最大的挑战是如何训练更好的 DCNN 算法。目前可以通过以下两种方式进行提升:设计更强表达能力的网络结构;使用规模更大的数据集,对训练集进行更精确的预处理。

1 人脸识别技术与数据库现状

2009 年发布的 LFW 数据库的样本图片来自于互联网名人图像,其中有 5 700 余人 13 000 多张图像,采用十折平均精度作为人脸识别性能指标,是在真实条件下最具权威的人脸识别数据库之一^[11]。有许多 DCNN 的开源实现已经公布,2014 年 Facebook^[9] 和香港中文大学^[12] 使用 DCNN 技术在人脸识别中取得重大突破,分别在 LFW 上获得了 97.35% 和 97.45% 的平均分类精度。随后,在 LFW 数据集上平均分类精度分别由 VGGface、CaffeFace 和 Google FaceNet 刷新到 97.27%、99.28% 和 99.63%,百度公司甚至得到 99.77% 的准确率^[13]。

DCNN 算法之所以需要大量的数据,主要原因在于越大规模的数据越能够学习到更加抽象、更加鲁棒的特征。表 1 中的数据充分体现了这一趋势:DeepID 系列从 20 万~45 万,CaffeFace 使用 70 万,百度使用 120 万,DeepFace 为 4 亿,VGGFace 有 260 万,谷歌达到 2 亿,Facebook 使用超过 10 万人的 50 亿图像数据库^[14]。

表 1 几种经典的 DCNN 模型在 LFW 数据集上的测试结果
Table 1 The test results of several classic DCNN models on the LFW data set

名称	年份	训练集 数据量/万	网络 数量	LFW 平均 精度/%
DeepFace ^[9]	2014	400	3	97.35
DeepID ^[2]	2014	20	25	97.45
DeepID2 ^[1]	2014	20	25	99.15
DeepID2+ ^[15]	2015	45	25	99.47
VGGFace ^[4]	2015	260	1	98.95
FaceNet ^[16]	2015	20 000	1	99.63
CaffeFace ^[17]	2016	70	1	99.28
Facebook ^[14]	2015	50 000	-	98.00
FaceNet_centerloss ^[18]	2017	505	1	99.30

可以看出,DCNN 需要大量数据进行训练,虽然很多数据可以从网络上抓取,但是没有组织,需要手动标注信息。可用性很强并且标注信息丰富的大规模数据库往往由 Google、Facebook 和百度这样的大型公司持有,但因为版权和隐私的问题,它们并

没有公开发布。由于在获取大规模标注信息丰富、准确度高的数据库方面遇到了较大的障碍,人脸识别技术在学术领域处于被动状态,甚至落后于工业界。为了克服这个问题,越来越多的研究机构也相继发布了一些大型的、多样化的数据集。目前,表 2 是公开的一些大型人脸数据集及其标注信息,香港中文大学团队发布了 CelebFaces+ 数据集,包含了大约 10 000 个身份的 20 万张图像,标注信息丰富,身份信息未划分,主要可以用来研究特征点定位和人脸属性学习。2014 年中科院自动化研究所发布并建立了一个大规模的 CASIA-webface 人脸识别数据集,包含了大约 10 500 个身份的 49 万张图像,并表示这个数据集是作为大规模训练集来使用。在文献[4]中,作者也公布了 2 622 个身份的 260 万张图像。MegaFace^[20] 数量大于 CASIA,但是被设计为测试集,每个人提供的图像很少,因此它从未用于训练 DCNN 系统。目前,这些数据集仍有很多噪声,身份标签错误,单张图像含有多个人脸等(图 1)问题,这导致数据库的准确性很难得到保证。



图 1 数据库存在的问题示例

Fig.1 Examples of problems that exist with the database

表 2 人脸检测和识别数据集

Table 2 Face Detection and Identification Data Set

数据集	图像数 量(万)	人数	注释属性
VGG Face ^[4]	260	2 622	人脸边框和姿态
CASIA WebFace ^[19]	49	10 575	—
MSCeleb ^[21-22]	1 000	100 000	—
UMDfaces ^[23]	37	8 501	人脸边框,21 特征点,性别和 3D 姿态
MegaFace ^[20]	100	690 572	人脸识别测试集合
Youtube face ^[24]	3 425 videos	1 595	非限制场景、视频
Low-shot face recognition ^[25]	100-200	21 000	面部区域被 裁剪和对齐
MF2 Training Dataset ^[26]	470	672 000	人脸边框和 68 个特征点

DCNN 的训练集通常需要进行大量的预处理工作,包括图像反转、随机裁剪、多尺度、彩色渲染、标准化数据等。通过不同方式的预处理产生的结果也是有很大偏差的,主要表现为人脸检测错误、特征点定位错误、身份归类错误等,这些偏差会对训练结果造成较大的影响。为了研究数据库准确性对 DCNN 训练的影响,本文提出了一种多角度评估数据清理方法,主要贡献总结包括:1)从人脸检测、特征点定位和人脸相似性等方面对数据库进行了清理;2)验证清理图像数据库的有效性,以提高训练准确性。

2 多角度评价清理图像

首先,采用当前主流的人脸检测和特征点定位算法对数据集进行清理。其次,利用公开的网络模型提取特征计算相似度,并使用相似度统计算法进行图像清理。如图 2 是数据清理的具体流程。

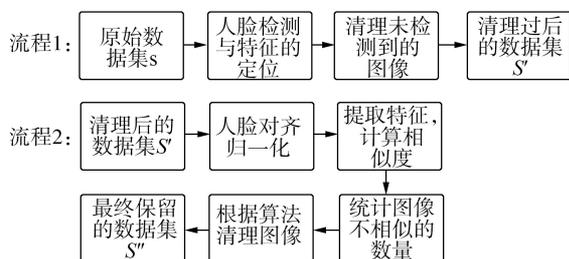


图 2 数据清理流程

Fig.2 Data cleansing process

表 3 5 个关键点的测试平均差、标准差和效率

Table 3 Five key points of the test mean difference, standard deviation and efficiency

算法	DCNN 算法			Seetaface 算法			MTCNN		
	平均差	标准差	效率/ $f \cdot s^{-1}$	平均差	标准差	效率/ $f \cdot s^{-1}$	平均差	标准差	效率/ $f \cdot s^{-1}$
HELEN	0.043 2	0.013 6	3.6	0.036 2	0.017 2	5.1	0.042 1	0.025 0	15.4
AFW	0.068 0	0.020 2	4.1	0.055 5	0.038 3	7.0	0.062 9	0.037 5	18.2
LFPW	0.059 4	0.015 6	3.9	0.049 0	0.022 6	6.5	0.048 2	0.018 2	17.5

在本文中训练深度卷积神经网络需要大规模数据集预处理工作,因此需要考虑面部检测算法的综合性能,主要包括效果和效率。表 4 进一步进行了两次大规模数据集的测试,从表 4 可以看出, MTCNN 算法对大数据的稳定性较好, CelebA 共计 20 多万图像,测试 CelebA 时 MTCNN 算法实际检测的面部图像为 182 387 张, Seetaface 算法实际检测的脸部图像为 180 032 张, MTCNN 算法在大规模数据集中具有较好的效果。

表 4 大规模数据集上 5 个特征点的比较

Table 4 Comparison of five feature points on large-scale data

评价指标	数据集	MTCNN	Seetaface
平均误差	Lfpw_net	0.049 2	0.055 5
	CelebA	0.048 8	0.049 0
标准误差	Lfpw_net	0.025 1	0.038 3
	CelebA	0.024 8	0.032 6

2.1 脸检测与特征点定位

人脸检测和人脸对齐已经成为独立的研究方向,人脸特征提取的前提就需要对这些算法进行研究,其中必不可少的就是人脸检测。人脸对齐(校准)或面部特征点定位主要是通过瞳孔、眉毛、嘴巴等人脸上的面部特征点对人脸进行校准,以促进正面特征提取,特别是对于局部关键特征提取。

本节主要对香港中文大学 DCNN Seetaface、中国科学院的面部检测算法和深圳先进学校 MTCNN 算法^[27]通过比较分析实际面部检测表现,并针对接下来的大规模脸数据集选择适当的人脸检测和关键点检测算法,从不同角度对 DCNN、Seetaface、MTCNN 这 3 种算法进行测试和分析,验证 3 种算法的性能。

表 3 是 3 种算法在 3 个不同测试数据集上 5 个关键点的平均差、标准差和效率。从表 3 中可以看出,DCNN 算法的标准差均小于 Seetaface 算法和 MTCNN 算法, MTCNN 算法略好于 Seetaface 算法,在这 3 个小数据集的测试中 DCNN 算法的稳定性最好,其次是 MTCNN;从平均误差来说,表现最好的是 Seetaface 算法,表明 Seetaface 算法可以适应更广泛的场景。

通过比较人脸检测和特征点定位的时间效率可以看出, MTCNN 检测的时间最短, DCNN 算法检测的时间最长。

根据上述实际测试的数据,本文选择 MTCNN 算法作为所有数据库的面部检测算法。

2.2 基于相似度统计的图像清理

基于相似度统计的图像清理算法的 3 个主要步骤如下。

1) 人脸对齐

所有的人脸区域都需要对齐到 112×96 的 RGB 图像,这里使用来自于文献[17]给出的对齐好的示例图像的 5 个关键点作为标准人脸模板,根据式(1)求出 5 个关键点的平均值 \bar{x}_i, \bar{y}_i , 其中, x_i^j, y_i^j 表示对应图像点的坐标, $i = 1, 2, 3, 4, 5$ 分别表示这 5 个关键点, j 表示第 j 张图像, n 表示所有的示例图像。

$$\begin{cases} \bar{x}_i = \sum_{j=1}^n x_i^j / n \\ \bar{y}_i = \sum_{j=1}^n y_i^j / n \end{cases} \quad (1)$$

在求出标准模板之后,根据式(2)求出原图到目标图像的仿射变换矩阵:

$$T = \text{cp2tform}(\text{Opt}, \text{Cpt}, 'similarity') \quad (2)$$

式中:Opt 是原图像的 5 个特征点,Cpt 是标准模板的 5 个点,T 是求得的变换矩阵,cp2tform 是 MATLAB 的标准函数。

根据式(3)得出对齐之后的人脸图像:

$$\text{crop} = \text{imtransform}(I, T, 'XData', [1, \text{imgsize}(2)], 'YData', [1, \text{imgsize}(1)], 'size', \text{imgsize}) \quad (3)$$

式中:I 是原图像, imgsize 是要对齐的图像尺寸, crop 是对齐之后的人脸图像, imtransform 是 MATLAB 的标准函数。

2) 人脸特征提取与相似度计算

本文使用的网络结构来自于文献[17]在 github 上提供的开源文件,如图 3 所示。

C: 卷积层
P: 池化层
LC: 局部卷积
FC: 全连接层

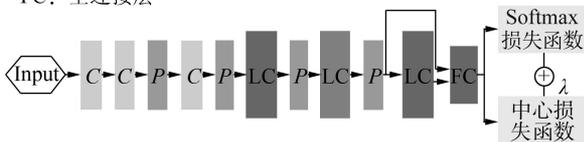


图 3 本文中人脸识别任务中使用的 DCNN 架构

Fig.3 The DCNN architecture used in face recognition tasks in this article

Softmax 损失函数和中心损失函数的联合监视,可以训练足够强大的 DCNN,并且能够得到两个关键学习目标的深层特征以及尽可能高的类间分散性和类内紧密性。这些特性对人脸识别来说至关重要。提取特征的模型也采用了文献[17]公布的网络模型,部分超参数设置由实际数据库来确定。

对于特征提取,按照惯例在进入网络模型之前,RGB 图像像素值将减去 127.5,并将每个像素值除以 128。深度特征来自第一 FC 层的输出,提取每个图像的特征和翻转图像的特征,并将其作为特征向量进行连接。在提取所有图像的特征之后,对从两个不同图像提取的特征向量进行相似性计算。所使用的度量是余弦距离,(4)式是两个特征向量 A 和 B 的角度,直接计算余弦距离作为相似度 S,即

$$S(A, B) = \cos(\theta) = \frac{AB}{\|A\| \|B\|} \quad (4)$$

使用等误差率(EER)作为判断是否为同一个人的阈值。当相似度超过该阈值 T 时,认定为同一个人,小于阈值 T 时不为同一个人。对于每种图像,进行相似度计算,1 表示两个图像是同一个人,0 表示两个图像不是一个人。

式(5)表明如何判断两个图像是否为同一个人。

$$L_{ij} = \begin{cases} 1, & \text{similarity}(P_i, P_j) \geq T \\ 0, & \text{similarity}(P_i, P_j) < T \end{cases}, \quad i, j = 1, 2, \dots, n \quad (5)$$

式中: P_i, P_j 表示一类图像中的第 i 和 j 张图像的特征向量, L_{ij} 表示第 i 和 j 张图像是否为同一个人。

3) 根据不相似的图像的比例确定是否清理

根据式(6)统计同一个人的人脸图像 C 中每一张图像与其他图像不相似的数目。

$$N_i = \text{count}(L_{ij} = 0), \quad j = 1, 2, \dots, n \quad (6)$$

式中: N_i 表示第 i 张图像与其他图像不相似的数目, $i = 1, 2, \dots, n$ 。

最后,根据式(7)判断是否清理该图像。

$$\text{对于 } C \begin{cases} \text{delete}, & N_i \geq \lambda n \\ \text{hold}, & N_i < \lambda n \end{cases}, \quad i = 1, 2, \dots, n \quad (7)$$

式中: n 表示一类人脸所有图像数目;参数 λ 的值需要根据不同的数据库来确定,具体 λ 值将在下一章实验部分具体分析。

3 实验与分析

本节选取了 CASIA-webface 数据集、UMDFace 数据集和 MSceleb 数据集作为实验数据库,并选取公共权威人脸识别测试数据集 LFW 和 Youtube face 作为测试数据集。

LFW 收集的数据集包含 5 749 个不同的人,有姿态、表情和光照的变化,总共 13 233 个图像,旨在研究人脸图像数据库中的困难的人脸识别问题。根据 LFW 提供的测试方案——unrestricted with labeled outside data^[28],我们测试了 6 000 个人脸对,人脸如图 4 所示。



(a) 匹配的对

(b) 不匹配的对

图 4 LFW 测试协议人脸对

Fig.4 The LFW test protocol face pair

Youtube 脸(YTF)数据集包含 3 495 个不同人的视频,平均每人 2.15 个视频。剪辑持续 48~6 070 帧,平均长度为 181.3 帧。它旨在研究视频下人脸识别的问题。根据 YTF 提供的测试协议——

unrestricted with labeled outside data,测试5 000个视频对,如图5所示。



图5 YTF 测试协议人脸视频对

Fig.5 The YTF test protocol face video pair

3.1 训练细节

1) 预处理

CASIA-webface 和 MSceleb 使用 2.1 节提出的人脸检测方法进行初步筛选,删除未检测到的人脸图像,并使用 2.2 节步骤 1 提出的人脸对齐将所有的人脸图像对齐成 112×96 的 RGB 图像。UMDface 数据集是经过包括人工判断在内的深度清理过的人脸图像数据集,清理过程在文献[23]中具体描述,提供了 21 个关键点的标注信息,我们选取了与 MTCNN 算法检测相同的 5 个点作为 UMDface 原始图像的特征点,并使用 2.2 节步骤 1 提出的方法进行人脸对齐。

2) 数据集设置

本文使用经过 2.1 节方式清理过的 CASIA-webface 作为第 1 个数据集 A,共有 10 575 个人,491 582 张图像。使用 UMDface 数据集作为第 2 个数据集 B,共有 8 501 个人,367 919 张图像。第 3 个数据集 C 来自于 2.2 节得出最终清理的数据集。经过 2.1 节方式清理过的 MSceleb 数据集作为第 4 个数据集 D,共有近 90 000 个人,8 500 000 余张图像。所有数据集按照 11:1 的比例分为训练集和验证集,至少使用一个图像作为验证集。

3) 网络设置和参数

本文使用的是 windows 环境下配置的 caffe 平台^[29]来训练模型,本节中的所有 CNN 模型都是相同的架构,详细信息如图 3 所示。学习率从 0.1 开始,学习策略(learning policy)多步衰减,迭代次数到 15 000、24 000、32 000 次时学习率分别除 10,权重腐蚀(weight decay)为 0.000 5,记忆因子(momentum)为 0.9。

3.2 不同准确度数据库训练实验

本实验是研究不同精度数据库对训练的影响,

使用 UMDface 数据库和 Webface 数据库,这两个数据库在保证数据准确性的程度上表现出很大的区别,UMDface 的数据库精度由大量人工信息标注,并且类别精确,Webface 数据库则没有任何人脸标注信息,很难保证数据的准确性。MSceleb 的准确性由于规模庞大,更加难以保证数据准确性,3 个数据库的详细数据对比如表 5 所示。

表 5 3 个数据库的属性比较

Table 5 Comparison of the properties of the three data sets

数据库	人数	图像总数	是否清理	噪声规模
UMDface	8 501	367 920	是	极少
Webface	10 575	494 414	否	少量
MSceleb	89 076	8 549 824	否	大量

在表 6 中,EER 是等错误率的准确率,这种精度是面部识别性能的重要指标。当 FPR 较小时,较大的 TPR 可以进一步解释模型的影响,应用场景不同对 FPR 的要求也不一样。因此,从表 6 可以看出,UMDface 以最小的规模实现了最高的精度,而 MSceleb 数据库规模最大,准确率也是最低的,数据的准确性对训练的影响是非常大的,即使数据规模再大,无法保证较高的数据精度,也无法获得较高的准确率,说明进行数据库清理是非常有必要的。

表 6 不同模型在 LFW 上的准确率

Table 6 The accuracy of different models on LFW

训练数据集	EER/%	FPR=0.01	FPR=0.001
UMDface	99.07	99.02	96.10
Webface	98.53	98.10	92.77
MSceleb	98.43	97.80	90.90

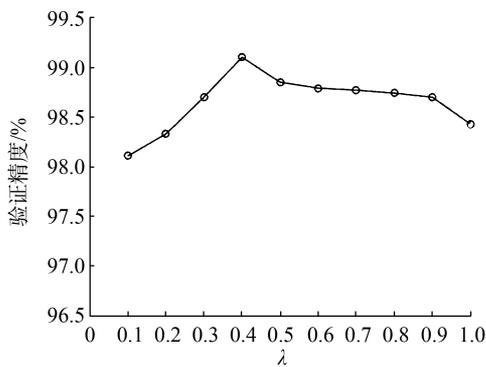
3.3 参数 λ 和一类图像的最小灵敏度

参数 λ 在清理图像数量方面起着重要作用,这可能会影响训练结果,某类图像的最小数量也可能影响训练。

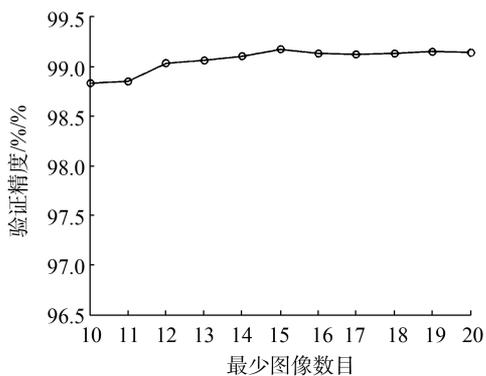
在第 1 个实验中, λ 从 0.1~1 来完成数据集 A 的深度清理工作,并使用清理后的数据集来训练模型,这些模型在 LFW 数据集上的准确率如图 6(a) 所示。显然,不做任何清理($\lambda=1$)准确率最差;正确选择 λ 的值可以提高 DCNN 特征的准确率;进行不同程度的清理(λ 不同)在一定范围内准确率保持稳定,清理过度会导致准确率下降。本文建议将 λ 的

值设定在 0.3~0.6,然后使用不同程度的清理数据集进行训练,并选取最佳 λ 值。

在第 2 个实验中,将 λ 值设为 0.4,再继续清理掉某些类图像数目少于一定值(从 10 到 20)的类,然后使用清理后的数据集来训练模型,这些模型在 LFW 数据集上的准确率如图 6(b)所示。同样,这种模式的准确性在一定范围内可以稳定,最好的结果是每个类别中不少于 15 个人,参数 $\lambda = 0.4$,清理后的 CASIA-webface 数据集有 9 240 个人,一共 400 000 张图像作为数据集 C 。



(a) 不同的 λ 模型



(b) $\lambda = 0.4$ 不同的最少图像数目模型

图 6 LFW 数据集上验证准确率

Fig.6 Verification accuracy on LFW datasets

3.4 在 LFW 和 Youtube face 上的测试结果

表 7 是在 LFW 上的测试结果和其他算法结果的对比,模型 A 是由数据集 A 作为训练集训练出来的模型,模型 B 是由数据集 B 作为训练集训练出来的模型,模型 C 是由数据集 C 作为训练集训练出来的模型。

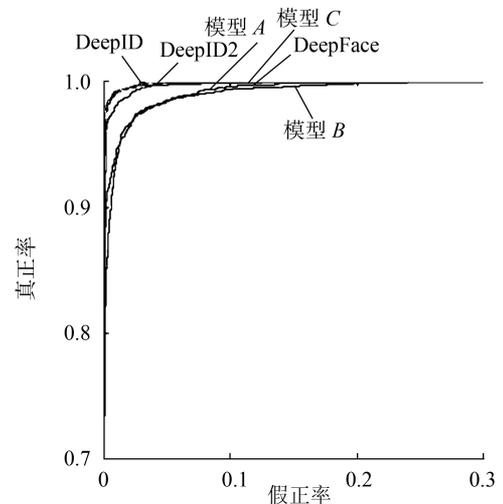
图 7 显示了不同方法的 ROC 曲线,我们可以从表 7 和图 7 的测试结果中得出以下结论:首先,模型 C 在 3 个实验模型中达到了最高的精度,模型 B 精度高于模型 A,表明数据清理明显提高了人脸识别训练模型的准确性;其次,原来的 Caffeface 训练所使用的数据集包含其他数据集,数据量更大,类别

也是本文的 2 倍,而在测试时,Caffeface 的结果是经过特征提取后的 PCA 处理,本文的结果未被 PCA 处理,也与原来的结果非常接近,进一步证明了数据清理的有效性。

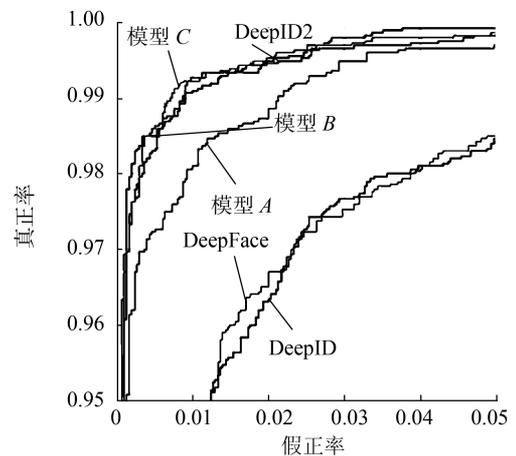
表 7 LFW 上的人脸验证结果

Table 7 Face verification results on LFW

方法名称	数据是否清理	训练人数	训练数量/万	网络个数	LFW 准确率%
DeepID	是(人工)	10 177	20	25	97.45
DeepID2	是(人工)	10 177	20	25	99.15
Deepface	是	4 030	400	3	97.35
Caffeface	是	17 189	70	1	99.28
模型 A	否	10 575	49	1	98.47
模型 B	是	8 501	36	1	99.07
模型 C	是	9 240	40	1	99.17



(a) 假正率 0~0.3 的 ROC 曲线



(b) 假正率 0~0.05 的 ROC 曲线

图 7 LFW 上不同方法的 ROC 曲线

Fig.7 ROC curves of different methods on LFW

表8是在 Youtube face 数据集上本文得出的测试结果和其他算法结果的对比,测试协议按照 3.1 节对 Youtube face 数据库的描述进行。

表8 Youtube face 上的人脸验证结果

Table 8 Face verification results on Youtube face

方法名称	是否清理	训练人数	训练数量/万	网络个数	准确率/%
CNN-3DMM ^[30]	否	10 575	49	1	88.80
DeepID2+	是	10 177	45	25	93.20
Deepface	是	4 030	400	3	91.40
Caffeface	是	17 189	70	1	94.90
模型 A	否	10 575	49	1	90.80
模型 B	是	8 501	36	1	92.50
模型 C	是	9 240	40	1	93.54

图8显示了不同方法的 ROC 曲线,我们还可以从表8和图8的测试结果中得出以下结论:模型C在3个模型中已经达到了最高的准确度,远远超过其他两个模型的 ROC 曲线,在非限制条件下的视频场景中,数据清理仍然会提高人脸识别精度。在目前的主流人脸识别算法中,模型C也显示出显著的成果,比3DMM算法、DeepID2+算法、Deepface算法,以及Caffeface算法训练的效果好。图像清理后,对数据集进行训练,干扰因子较小,对于基于深卷积神经网络的人脸识别非常有用。

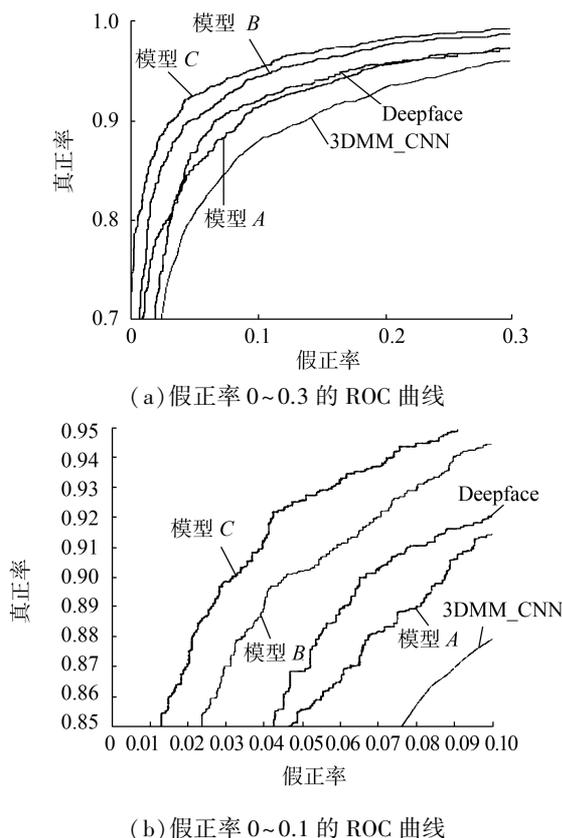


图8 Youtube face 上不同方法的 ROC 曲线

Fig.8 ROC curves of different methods on Youtube face

4 结束语

深度卷积神经网络的发展得益于大数据,因为数据量够大,计算机够强大,机器本身才能学习出各种复杂的特征。而数据的准确性,也就是数据标签、标注信息、类别等的准确性,也会对训练的模型结果造成一定的影响,因此本文对数据库清理方面做了研究。本文提出将数据库进行多角度清理后再训练的方法,通过与未清理和其他方法的比较发现,清理后的数据库在训练上结果更精确。实验证明,清理后的数据集能够提高网络识别率。

目前,大多数公开的数据集仍含有很多噪声,大规模数据去除噪声仍是一个值得重视的问题。本文数据清理方法是否对所有数据库具有普适性,是否已经存在更高效更准确的数据清理方法需要进一步探究。下一步的工作可以考虑将多个清理后的数据集进行合并来扩大数据量。

参考文献:

- [1] SUN Y, WANG X, TANG X. Deep learning face representation by joint identification-verification[J]. Advances in neural information processing systems, 2014, 27: 1988-1996.
- [2] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10,000 classes[C]// IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2014: 1891-1898.
- [3] 王晓刚,孙祎,汤晓鸥. 从统一子空间分析到联合深度学习: 人脸识别的十年历程[J]. 中国计算机学会通讯, 2015, 11(4): 8-15.
WANG Xiaogang, SUN Hui, TANG Xiaoou. From unified subspace analysis to joint depth learning: ten years of face recognition[J]. China computer society newsletter, 2015, 11(4): 8-15.
- [4] PARKHI O M, VEDALDI A, ZISSERMAN A. Deep face recognition[C]//British Machine Vision. London, Britain, 2015: 411-4112.
- [5] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering [C]// IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 815-823.
- [6] DING C, TAO D. A comprehensive survey on pose-invariant face recognition[J]. Acm transactions on intelligent systems and technology, 2015, 7(3): 37.
- [7] SUN Y, LIANG D, WANG X, et al. DeepID3: face recognition with very deep neural networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 963-971.
- [8] SUN Y, WANG X, TANG X. Hybrid deep learning for face

- verification[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(10): 1997–2009.
- [9] TAIGMAN Y, YANG M, RANZATO M, et al. Deepface: closing the gap to human-level performance in face verification[C]//IEEE Conference on Computer Vision and Pattern Recognition. Columbia, American, 2014: 1701–1708.
- [10] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition[C]//ECCV Conference on Computer Vision. Amsterdam, Holand, 2016: 499–515.
- [11] HUANG G B, MATTAR M, BERG T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments[J]. Month, 2007.
- [12] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10,000 classes[C]//IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2014: 1891–1898.
- [13] LIU J, DENG Y, BAI T, et al. Targeting ultimate accuracy: face recognition via deep embedding[C]//European Conference on Computer Vision. Amsterdam, Netherlands, 2016: 499–515.
- [14] TAIGMAN Y, YANG M, RANZATO M, et al. Web-scale training for face identification[C]//Computer Vision and Pattern Recognition. Columbus, USA, 2014: 2746–2754.
- [15] SUN Y, WANG X, TANG X. Deeply learned face representations are sparse, selective, and robust[C]//Computer Vision and Pattern Recognition. Boston, USA, 2015: 2892–2900.
- [16] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: a unified embedding for face recognition and clustering[C]//IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 815–823.
- [17] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition[C]//European Conference on Computer Vision. Amsterdam, Netherlands, 2016: 499–515.
- [18] SEITZ S M, MILLER D, et al. The megaface benchmark: 1 million faces for recognition at scale[C]//Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 4873–4882.
- [19] BORJI A, IZADI S, ITTI L. iLab-20M: a large-scale controlled object dataset to investigate deep learning[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2221–2230.
- [20] KEMELMACHERSHLIZERMAN I, SEITZ S M, MILLER D, et al. The megaface benchmark: 1 million faces for recognition at scale[C]//Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 4873–4882.
- [21] GUO Y, ZHANG L, HU Y, et al. MS-Celeb-1M: challenge of recognizing one million celebrities in the real world[C]//Electronic imaging. San Francisco, USA, 2016: 1–6.
- [22] GUO Y, ZHANG L, HU Y, et al. MS-celeb-1M: a dataset and benchmark for large-scale face recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 113–124.
- [23] BANSAL A, NANDURI A, CASTILLO C, et al. UMDFaces: an annotated face dataset for training deep networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 976–984.
- [24] WOLF L, HASSENER T, MAOZ I. Face recognition in unconstrained videos with matched background similarity[C]//Computer Vision and Pattern Recognition. Colorado Springs, USA, 2011: 529–534.
- [25] GUO Y, ZHANG L. One-shot face recognition by promoting underrepresented classes[J]. Computer vision and pattern recognition, arxiv:1707.05574, 2017.
- [26] NECH A, Kemelmachersh Lizerman I. Level playing field for million scale face recognition[J]. Computer vision and pattern recognition, arxiv:1705.00393, 2017.
- [27] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE signal processing letters, 2016, 23(10): 1499–1503.
- [28] HUANG G. B., LEARNED-MILLER E. Labeled faces in the wild: updates and new reporting procedures[R]. Technical report UM-CS-2014-003.
- [29] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[J]. Eprint arxiv, 2014: 675–678.
- [30] LIU F, ZENG D, ZHAO Q, et al. Joint face alignment and 3D face reconstruction[C]//European Conference on Computer Vision. Amsterdam, Netherlands, 2016: 545–560.

作者简介:



夏洋洋,男,1990年生,硕士研究生,主要研究方向为深度学习、图像处理、人脸识别。



龚勋,男,1980年生,副教授,博士,主要研究方向为图像处理及模式识别、三维人脸建模、人脸图像分析及识别。获国家发明专利2项,发表学术论文30余篇,出版专著1部。



洪西进,男,1957年生,特聘教授,博士,主要研究方向为信息安全、生物辨识、云计算与大数据、智能图像处理。发明专利13项,发表SCI期刊学术论文80余篇,国际学术会议论文110余篇。