

基于深度学习的人脸识别系统

欧阳军林,黄井滔,张啸

(湖南科技大学 计算机科学与工程学院,湖南 湘潭 411201)

摘要:以深度学习为基础,将人脸检测方法(MTCNN)、OpenCV 仿射变换方法、人脸识别方法(FaceNet)、相似性查找方法(HNSW)和人脸属性识别方法(FAN)有机结合,构建了一个人脸识别系统。在深度学习开发框架(Caffe)上进行模型训练与测试,并将预训练模型应用于人脸注册、人脸检索和人脸属性识别任务,通过实例测试表明在现实场景中具有一定的鲁棒性。该系统在课堂考勤、身份认证、门禁系统、登录解锁和社交媒体平台等方面也具有巨大的应用潜力。并将该研究成果向学生进行专题讲解,以期开拓学生的视野,增加学生学习科研兴趣,提高学生学术水平,锻炼动手能力,有利于学生今后的就业。

关键词:人脸识别;软件系统开发;深度学习;深度学习开发框架

中图分类号:G642

文献标志码:A

文章编号:1674-5884(2019)06-0133-06

随着谷歌、苹果、脸书、亚马逊和微软(GAFAM)等大型互联网公司在人脸识别领域不断取得突破性的进展,并且国内已经将人脸识别技术大规模应用于社会生活的方方面面,基于深度学习的人脸识别技术已然成为当今最受追捧的技术之一。人脸识别用途的爆炸式的增长带来了各式各样的现实需求,人脸识别可以应用于身份认证、银行安全、法医调查、刷脸支付、火灾预警、登录解锁、执法、门禁系统、舆情监控、社交媒体平台和疾病诊断等。已有人脸识别方法^[1-4]已经取得了显著的成果,随着深度学习技术的不断发展,许多深度方法都表现出了优于人类识别的准确度。本文在对深度学习方法进行深入研究探讨的基础上,将理论应用于实践,设计开发了一套人脸识别系统。并将该研究成果向本科及研究生进行专题讲解,以期开拓学生的视野,增加学生学习与科研兴趣,提高学生学术水平,锻炼动手能力,有利于学生今后的就业。

1 人脸识别系统总体功能与设计

1.1 人脸识别系统总体功能

在本系统中,依照功能划分为以下五个部分(如图1所示):

1)人脸采集模块,人脸图像的来源可以是来自于本地文件、网络URL或是视频帧。主要通过html标签以表单的形式上传到服务器,或者利用OpenCV函数从视频文件和摄像头中获取。

2)人脸矫正模块,由于截取的人脸图像有可能是歪的,为了提高识别准确度,需要用到OpenCV中提供的一个仿射变换方法,利用检测到的人脸关键点把截取的人脸矫正到一个标准的位置。

3)人脸检测模块,在该模块用来对输入图像进行预处理,判断图片中的人脸数量,以及检测出人脸所在的位置和人脸关键点坐标,并把人脸图像从背景中截取下来。同时在现实场景中应保持较好的鲁棒性和实时检测的性能。

4)人脸识别模块,人脸识别模块一般由两部分组成,即人脸特征提取和建立人脸索引库。在

收稿日期:20190905

基金项目:湖南省自然科学基金项目(2017JJ2099、2017JJ3091);湖南省大学生创新研究项目(105-S143300);湖南科技大学教研教改项目(905-G31709)

作者简介:欧阳军林(1977-)男,湖南湘潭人,讲师,博士,主要从事图形图像处理、深度学习等研究。

得到预处理完成的人脸图像后,将人脸输入到预训练好的模型中去,提取出 512 维的人脸特征,并为其建立索引。

5)人脸属性模块,首先对待检测的人脸图像

进行预处理,得到一个较为标准的人脸图像,然后将该人脸图像输入预训练好的 Caffe 模型中去,最终将得到 40 个预测值,分别表示人脸图像是否拥有该属性。

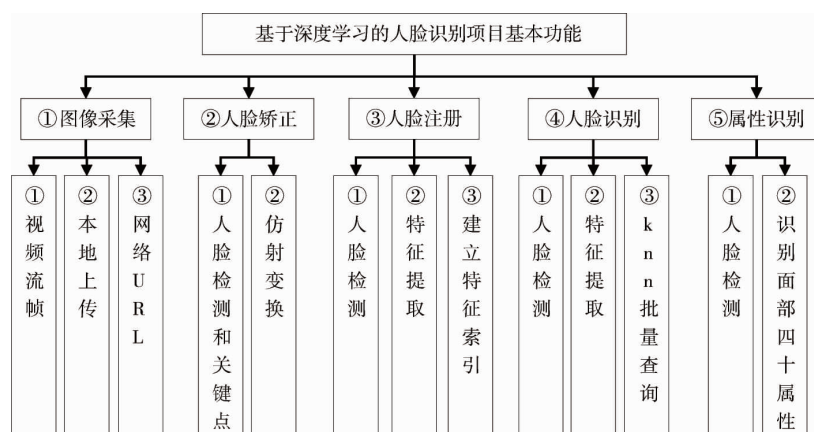


图1 人脸识别系统总体功能

1.2 人脸识别系统设计

本系统研究了人脸识别的相关算法,并在此基础上结合实际情况进行项目开发,本系统采用的开发环境为 Ubuntu14.04/OpenCv/Python2.7/Caffe/Tensorflow。通过 Python 代码将 Caffe 模型应用于人脸识别系统,采用 PythonFlaskWeb 框架将各个功能模块集成起来,为人脸识别系统开发了一个后端服务器。在前端部分,则采用 JQuery 代码为前端编写业务逻辑和界面,实现了人脸注

册、人脸检索和面部属性识别功能。

如图2所示,系统设计了三个页面用于图像采集和结果展示等操作,分别为:人脸注册页面、人脸检索页面和人脸属性识别页面。对于上传的图像可以是来自于照片、视频帧或摄像头等情况,后端接收图像后进行预处理,根据不同场景分别对人脸图像进行特征提取、人脸注册、人脸检索、面部属性提取等操作,最后将得到的处理结果返回给前端进行展示。

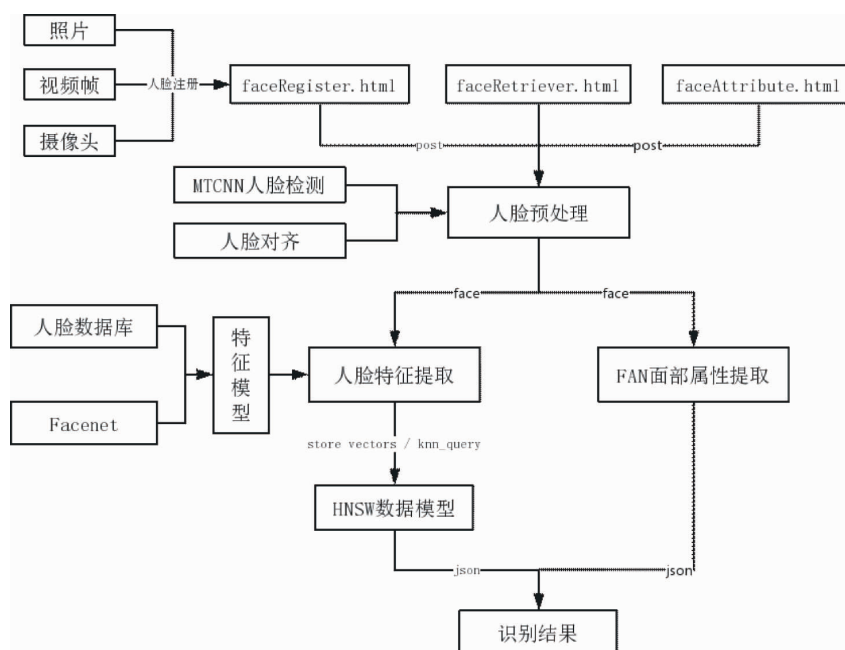


图2 人脸识别系统构架流程

2 模型训练与测试

在本节将完成三个 Caffe 模型的训练或转换,并在相应数据集上进行测试。其中 MTCNN 模型在 WiderFace 数据集上进行训练,在 Fddb 数据集上的识别准确率为 91%;FaceNet 模型通过一个旧预训练模型转换而得到;FAN 模型在 CelebA 数据集上进行训练与测试,在 CelebA 测试数据集上的识别准确率为 91%。

2.1 基于 MTCNN 的人脸检测

MTCNN 是 Kaipeng Zhang 在文献[5]中提出的方法,简而言之,文中采用了一个基于多任务的级联架构,将人脸检测和关键点定位任务分三个步骤来实现,即一步一步从粗到细实现人脸检测和关键点定位的任务,生成最终的人脸预测框和人脸关键点坐标。将 WiderFaceTraining 作为训练数据集,在 Fddb 数据集上测试模型的识别准确率。首先在 Fddb 数据集上生成 MTCNN 模型的检测结果文件,接着将得到的检测结果进行官方测评,并将检测结果绘制成 ROC 图。测试结果如图 3 所示:

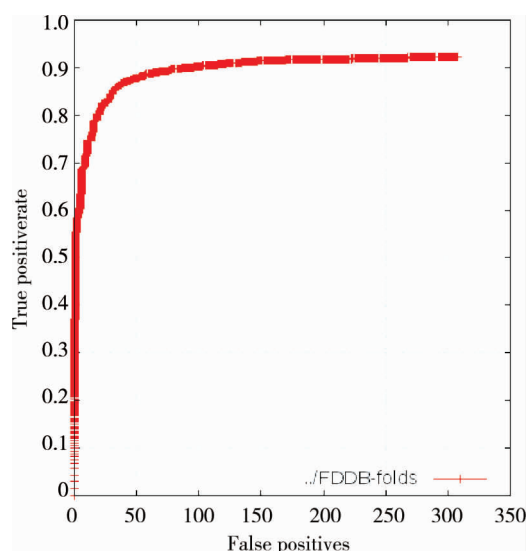


图3 MTCNN 模型测试结果

2.2 基于 FaceNet 的人脸识别

FaceNet 是 Florian Schroff 在文献[6]中提出的,它可以通过卷积神经网络的学习将人脸面部图像映射到一个欧几里得空间中,并在欧几里得空间中具有如下特性,相同的人脸图像在空间中相邻,而不同的人脸则相距较远。在此基础上可以采用机器学习中的方法来为人脸特征建立特征索引,由此完成人脸注册、识别、验

证等任务。将 Inception_resnet_v1 模型转换为 Caffemodel,另外似乎具有 128D 输出的旧预训练模型比具有 512D 输出的新模型有更高的识别效果,所以这里选用的是将 Tensorflow 的预训练模型 20170512-110547 转换为具有 512D 输出的 Caffe 模型。

2.3 基于 FAN 的人脸属性识别

FAN 是 Keke He 在文献[7]中提出的,采用了一个双通道深卷积神经网络神经网络,需要有两部分的输入,即原始图像和经过 GAN 网络生成的抽象图像。为了加快模型参数的收敛,准备了一个旧的 FAN 模型用于微调,并在 CelebA 数据集完成模型训练和测试。在 CelebA 测试数据集上测试模型的识别准确率,将测试数据集输入到模型中进行检测,生成一个包含人脸属性预测结果的 TXT 文件,最后将预测结果和官方的标签文件进行比较,测试结果如图 4 所示:

```

caffe@caffe-Precision-T3610: ~/intel/caffe-master/data/FaceNet-Caffe-master/FaceAttrib
pred_file=./result/test_with_abstraction_full.txt
python prediction_dual_path.py --gpu=4 \
--attr-num=40 \
--prototxt-path=./outputs/deploy_dual.prototxt \
--caffemodel-path=./outputs/dual_path_parse_resnet_celeb
--test-file=${gt_file} \
--pred-file=${pred_file} \
--feature-layer="pred" \
--root-folder=./data/ \
--mean-file=./data/pretrained/ResNet_mean.binaryproto
python cal_accuracy.py --gt-file=${gt_file} --pred-file=${
('mean result', 91.805806031459767)
caffe@caffe-Precision-T3610:~/intel/caffe-master/data/FaceNet-Caffe-master/Face
Attribute/FAN/evaluation5

```

图4 FAN 模型测试结果

3 人脸识别系统设计与实现

本节详述了图像预处理、人脸特征提取、建立人脸特征索引、人脸检索、人脸属性和视频人脸检测等主流业务的详细设计过程,并将各部分功能有机结合起来,实现了从本地视频、摄像头、本地图片或网络 URL 等方式读取数据,实时地人脸检测和识别系统。

3.1 图像预处理

图像预处理部分主要由两部分组成,即人脸检测和人脸对齐。利用 MTCNN 获得截取后的人脸以及关键点坐标,并根据人脸关键点坐标利用 OpenCV 仿射变换将人脸矫正到一个标准的位置。图像预处理部分流程如图 5 所示:

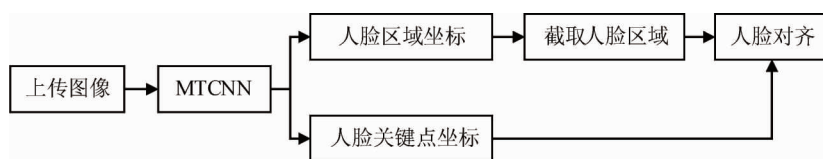


图5 图像预处理流程

3.2 FaceNet 特征提取

利用 3.1 节中提供的方法对人脸图像进行预处理,此时可能得到多张人脸图像,将截取的每一张人脸图像输入到 FaceNet 网络中,并从 FaceNet 的 flatten 层提取人脸特征向量(512 维)。具体流程如图 6 所示:

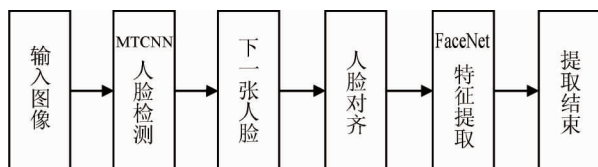


图6 特征提取流程

3.3 HNSW 人脸特征索引

为了完成人脸识别任务,需要为提取到的人脸特征向量建立索引。由于相同人脸建立的索引都是相似的,在此基础上,有很多机器学习的方法来完成人脸分类和识别功能,例如 knn, annoy 算法。本文采用 HNSW[8] 算法为人脸特征向量建立索引,并利用余弦相似度来判别是否为同一张人脸。hnsplib.Index(space, dim) 方法在 space 具有整数维度的空间中创建非初始化的 HNSW 索引 dim。通过计算人脸特征向量的余弦相似度,然后设置相应的阈值,从而判断是否为相同人脸。

3.4 人脸注册

人脸注册主要分两步来实现,即提取人脸特征向量和建立人脸特征索引。需要处理的人脸图像主要有两类,即来自本地文件上传和网络 URL,同时支持批量上传文件。对于上传图像,允许存在多张人脸,并将对图片中的多张人脸进行注册。获得上传图像后,首先需要对人脸图像进行预处理,再将矫正的人脸图像 embedding(映射)为人脸特征向量(512 维)。然后为人脸特征向量建立索引,并将它保存到 HNSW 模型中,调用 add_items() 函数增添人脸特征向量(注册人脸),最后调用 save_index() 函数持久化保存。在前端展示中,将标注出已注册人脸的区域框和关

键点位置,人脸注册页面如图 7 所示:



图7 人脸注册页面

3.5 人脸检索

图像上传和预处理阶段,同人脸注册时基本相似,但对于待检索的图像只允许存在一张人脸。同样将预处理后的人脸图像利用 FaceNetembedding 成 512 维的人脸特征向量。利用 hnsplib 提供的 knn_query() 函数批量查询 k 个距离最近的元素,并返回余弦相似度和该图像的标签(注册时存入的)。例如在下图中选择了《新三国演义》中司马懿的图像作为样例,根据注册情况,后端检索完成后会向前端返回原始注册图像,并在图片正下方标注人脸相似度(80+%)。人脸检索页面如图 8 所示:



图8 人脸检索页面

3.6 人脸属性识别

在人脸属性识别部分允许同时输入抽象人脸和原始人脸进行人脸属性识别,也可以只输入原始图像。在得到预处理完成的人脸图像后,将人脸图像输入到 FAN 网络中提取出 40 种人脸属性。最后将 40 种属性值和人脸图片一起返回给前端展示。在测试中选择了胡歌的图像作为样

例,人脸属性识别页面如图9所示:



图9 人脸属性识别页面

3.7 视频人脸识别

通过 OpenCV 从视频流中读取视频帧,并将读取到的视频帧输入到 MTCNN 神经网络中进行人脸检测和关键点定位,在得到检测结果后,根据检测到的人脸框坐标将人脸截取下来,并保存在相应目录中,同时将人脸图像进行对齐,将对齐后的人脸提取人脸特征,再检索人脸是否已经注册,最后将检测结果在视频中展示。下面详细介绍实现的流程:

第一步是读取数据。在这一部分可以是来自于视频文件,也可以是从摄像头中获取,只需调用 OpenCV 的 CV2 中的 VideoCapture() 函数。当设置参数为 0 时,从摄像头中获取数据,当需要加载本地视频时,参数需设置为视频文件所在的路径。

第二步是人脸检测。在这一部分每隔一帧进行一次人脸,需要注意的是,对于每一帧图像需要先完成通道转换 (RGB→BGR)。然后调用基于 MTCNN 实现的人脸检测方法,得到人脸框坐标和人脸关键点坐标。

第三步是人脸对齐。在得到人脸图像后,并利用人脸关键点通过仿射变换方法进行人脸对齐。

第四步是人脸识别。在这一部分又可以分成两步,首先将预处理后的人脸图像按照 2.2 节中实现的方法提取到 512 维的人脸特征向量,然后根据人脸余弦相似度判断人脸是否已经在 HNSW 模型中注册过。

第五步是视频中展示。利用 cv2 的 rectangle() 和 circle() 函数在视频中标注出人脸区域以及人脸关键点位置。



图10 视频人脸识别页面

4 结语

为了更好地提高人脸识别系统的检测效果和实时性能,方便人们在现实生活中使用。本文从深度学习的应用出发,研究了人脸识别的相关算法,并在此基础上结合实际情况进行项目开发,有机地结合了人脸检测方法 (MTCNN)、OpenCV 仿射变换方法、人脸识别方法 (FaceNet)、相似性查找方法 (HNSW) 和人脸属性识别方法 (FAN),构建了一个实时的人脸识别系统。本系统具有很高的实用价值,很适合推广应用于课堂考勤、身份认证、门禁系统、登录解锁和社交媒体平台等方面。并将该研究成果面向本科生及研究生进行专题讲解,以期开拓学生的视野,增加学生学习科研兴趣,提高学生学术水平,锻炼动手能力,有利于学生今后的就业。

参考文献:

- [1] Zhou E, Cao Z, Yin Q. Naive-deep face recognition: Touching the limit of LFW benchmark or not? [EB/OL]. (2015-01-20) [2018-07-09]. <https://arxiv.org/pdf/1501.04690.pdf>.
- [2] Taigman Y, Yang M, Ranzato M, et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification [C]// Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2014.
- [3] Sun Y, Liang D, Wang X, et al. Deepid3: Face recognition with very deep neural networks [EB/OL]. (2015-02-03) [2018-05-05]. <http://de.arxiv.org/pdf/1502.00873>.
- [4] Masi I, Tran A T, Hassner T, et al. Do we really need to collect millions of faces for effective face recognition? [C]//European Conference on Computer Vision. Springer, Cham, 2016: 579.

- [5] Zhang K , Zhang Z , Li Z , et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks [J]. IEEE Signal Processing Letters, 2016 (10):1499–1503.
- [6] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.
- [7] Keke He , Yanwei Fu , Wuhao Zhang , et al. Harnessing Synthesized Abstraction Images to Improve Facial Attribute Recognition[C]//Proceedings of the Twenty–Seventh International Joint Conference on Artificial Intelligence, {IJCAI–18}, 2018.
- [8] Malkov Y A , Yashunin D A . Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs [EB/OL]. (2016 – 03 – 30) [2018 – 09 – 09]. <https://arxiv.org/ftp/arxiv/papers/1603/1603.09320.pdf>.
- [9] Yang S, Luo P, Loy C C, et al. Wider face: A face detection benchmark[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
- [10] Liu Z, Luo P, Wang X, et al. Deep learning face attributes in the wild[C]//Proceedings of the IEEE international conference on computer vision, 2015.
- [11] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments[C]//Workshop on faces in ‘Real–Life’ Images: detection, alignment, and recognition, 2008.
- [12] Jain V, Learned–Miller E. Fddb: A benchmark for face detection in unconstrained settings [R]. UMass Amherst Technical Report, 2010.
- [13] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]//Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014.

Face Recognition System Based on Deep Learning

OUYANG Junling, HUANG Jingtao, ZHANG Xiao

(School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, Hunan)

Abstract: Based on deep learning, this paper develops a face recognition system that combines face detection method (MTCNN), OpenCV affine transformation method, face recognition method (FaceNet), similarity finding method (HNSW) and face attribute recognition method (FAN). Model training and testing are carried out on the deep learning development framework (Caffe), and the pre-training model is applied to face registration, face retrieval and face attribute recognition tasks. The example test shows that it has certain robustness in real scenes. The system also has great application potential in classroom attendance, identity authentication, access control systems, login and unlocking, and social media platforms. The research results will be given to undergraduate students to broaden students’ horizons, increase students’ interest in studying and scientific research, improve student’ academic performance, and exercise their hands-on ability, which is conducive to students’ future employment.

Key words: face recognition; software system development; deep learning; Caffe

(责任校对 蒋云霞)