

基于深度学习的证件照人脸识别方法^①

张晓林, 范宇, 刘惟锦, 王春华

(北京航天长峰科技工业集团有限公司, 北京 100039)

通讯作者: 范宇, E-mail: alex3936@163.com

摘要: 为解决安防业务中对人脸识别技术的迫切需求, 进行了基于深度学习的证件照人脸识别方法的研究与实现. 对人脸识别各关键技术环节进行了实现并进行了算法对比, 同时提出了使用 Siamese 网络进行特征重映射的方法进一步提升人脸识别精度. 实验结果表明, 使用证件照构建的人脸库, 通过基于深度学习的算法实现了高效精确的人脸识别.

关键词: 人脸识别; 深度学习; 证件照; 人脸检测; 相似度衡量

引用格式: 张晓林, 范宇, 刘惟锦, 王春华. 基于深度学习的证件照人脸识别方法. 计算机系统应用, 2018, 27(5): 203-208. <http://www.c-s-a.org.cn/1003-3254/6357.html>

Face Recognition for Identification Photos Based on Deep Learning

ZHANG Xiao-Lin, FAN Yu, LIU Wei-Jin, WANG Chun-Hua

(Beijing Changfeng Aerospace Science and Industry Co. Ltd., Beijing 100039, China)

Abstract: To solve the urgent technical requirement of face recognition in security industry, this research and implementation of face recognition based on deep learning for identification photos are carried out. The key technologies and algorithms of face recognition are implemented and compared. At the same time, to further improve accuracy, the method of feature remapping using Siamese network is proposed. Finally, the experimental results show that by using deep learning algorithms and face-library constructed by identification photos, efficient and accurate face recognition is achieved.

Key words: face recognition; deep learning; identification photos; face detection; similarity measure

引言

人脸识别^[1,2]技术是计算机视觉领域的研究热点, 同时在安防领域也有大量的应用需求. 最近几年, 随着公安部门在传感技术、云计算技术、视频智能监控技术、图像联网技术等方面的迅猛发展, 针对公安系统领域安防业务需求的人脸识别技术就显得尤为重要, 在嫌疑人追逃、可疑人员确认、重点场所布控等方面, 人脸识别都有着不可替代的作用.

早期使用较多的人脸识别方法主要是局部二值模型 (LBP)^[3]特征结合支持向量机 (SVM) 分类器的方法,

该方法在多种实验条件下都有很好的识别准确率, 但是在真实环境中由于光照、遮挡以及其他不确定因素, 导致识别性能显著下降. 近年来随着深度学习的发展, 基于深度学习的人脸识别算法相比于早期传统算法, 在准确率和鲁棒性方面都有了显著的提升.

在人脸识别的研究中, 最常用的图像库是 LFW^[4], 该数据库中共有 5749 人的 13 233 幅图像, 其中 1680 人拥有两幅及以上的图像. 该数据库所采集的照片是生活中人脸的图片, 目的是能够提高在自然条件下人脸识别的精确度.

① 收稿时间: 2017-09-11; 修改时间: 2017-09-30; 采用时间: 2017-10-12; csa 在线出版时间: 2018-04-23

目前在该图像库下,国内外的很多研究团队的实验室利用深度学习算法的识别准确率已经高达95%乃至99%以上. Z Zhu^[5]等人提出了一种新的深度学习模型,该模型可以在保持个体之间差异的同时,极大地减少单个个体的人脸图像之间的差异,该方法在LFW数据集上取得了96.45%的成绩. Y Taigman^[6]等人通过使用额外的3D模型改进了人脸对齐的方法,然后使用采集到的四百万张人脸图片训练一个9层的网络模型,该方法在LFW数据集上取得了97.35%的准确率. Y Sun^[7]等人同时使用了VGG Net和GoogLeNet相结合来进行人脸识别,利用这两种框架在LFW数据集上取得了97.45%的成绩. E Zhou^[8]等人使用的是常规的卷积神经网络模型,但他们将从互联网上搜集到的五百万张人脸图片作为数据集用来训练该网络,使其在LFW数据集上的准确率达99.50%. F Schroff^[9]等人开发出一种新的人脸识别系统,该系统将人脸图像直接映射到欧式空间,空间中距离的大小代表人脸图像相似性的大小,当该映射空间形成之后,在此基础上进行人脸识别就显得轻松而且自然,该方法在LFW数据集上的准确率为99.63%.

从上述研究成果来看基于深度学习算法的人脸验证准确率已经达到甚至超过了人眼,但并不能武断地说计算机进行人脸识别的准确率已经超越了人类.这主要是基于两点原因,首先是在LFW数据集上是进行人脸验证,是判断一对输入图像中出现的是否是同一个人,而人脸识别是在一个人脸库中找到和输入人脸最相近的图像,相比而言,后者的难度要远大于前者,

尤其是在百万甚至千万量级的人脸库中.其次,在实际的业务应用中,获取个人大量的标注身份信息的图像是不现实的,很多时候只能使用单张的个人身份证照片构建人脸库,而在进行人脸识别时,获取的测试人脸图像由于光线、分辨率、清晰度、倾斜角度和时间跨度等因素,都会大大增加人脸识别的难度.

针对公安部门在日常业务中对基于证件照进行人脸识别的迫切需求,本文进行了基于深度学习的证件照人脸识别的研究与实现.

1 人脸识别系统算法设计

人脸识别系统从输入图像中自动识别图像中人的身份信息,本质上是从构建的人员身份库中找到和输入图像中包含的最相似的人脸,人脸识别要做的是计算两幅人脸图像的相似度是否足够高.

如图1所示,人脸识别系统的实现一般包括人脸检测、特征点定位、人脸对齐、特征向量提取和相似度衡量五个过程.其中人脸检测找到脸在哪里,即找到图像中人脸的精确位置;特征点定位找到五官在哪里,分别找到左眼中心、右眼中心、鼻尖、左嘴角和右嘴角的位置;人脸对齐实现测试图像和参考图像的五官位置;特征向量提取将人脸图像表示为特征向量;相似度衡量比较两幅人脸图像特征向量之间的相似度是否足够高.本文使用MTCNN算法^[10]完成人脸检测和特征点定位,使用普氏分析进行人脸对齐,使用FaceNet进行人脸特征提取,计算特征向量的欧式距离进行人脸图像相似度衡量.

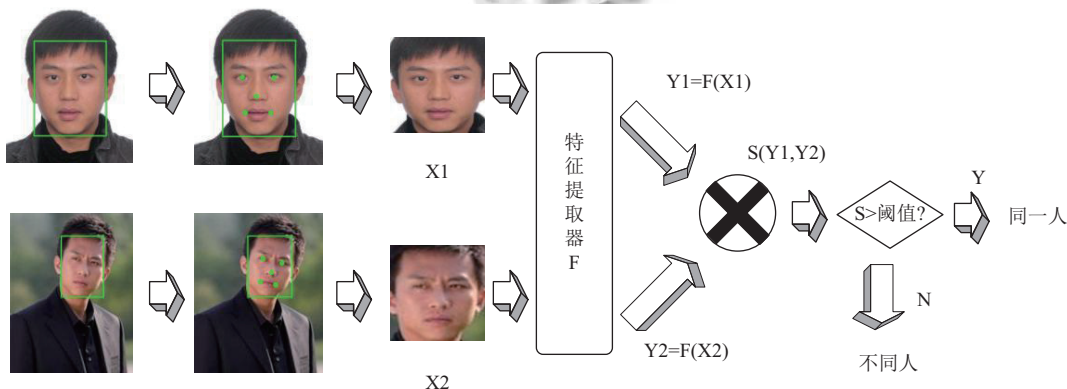


图1 人脸识别流程

1.1 基于MTCNN的人脸检测算法

在人脸检测阶段,分别采用了P-Net, R-Net和O-

Net, 3个不同的网络结构三阶级联,如图2所示.

首先是Proposal Network (P-Net),该网络的主要功

能是初步获取人脸区域范围以及边界框的回归向量,可以同时并快速地获得多个包含人脸的边界框. 然后通过非极大值抑制对这些候选的边界框进行校准,这样可以筛选出重叠度较高的边界框并将它们进行合并. 其次是 Refine Network (R-Net), 该网络的主要功能依然是通过边界框以及非极大值抑制的方法将不满足要求的边界框进行合并,但是由于该网络框架相比

Proposal Network 在最后多了一个全连接层,因此其筛选能力更强,可以将之前无法识别出的干扰边界框进行合并与排除. 最后是 Output Network (O-Net), 该网络也是有着同样的目的,不过其相比于 R-Net 多了一个卷积层,因此在处理时会更加精细. 经过上述 3 个网络的处理,可以得到检测出的人脸边界框以及 5 个关键点(双眼、鼻尖、左右嘴角)的坐标.

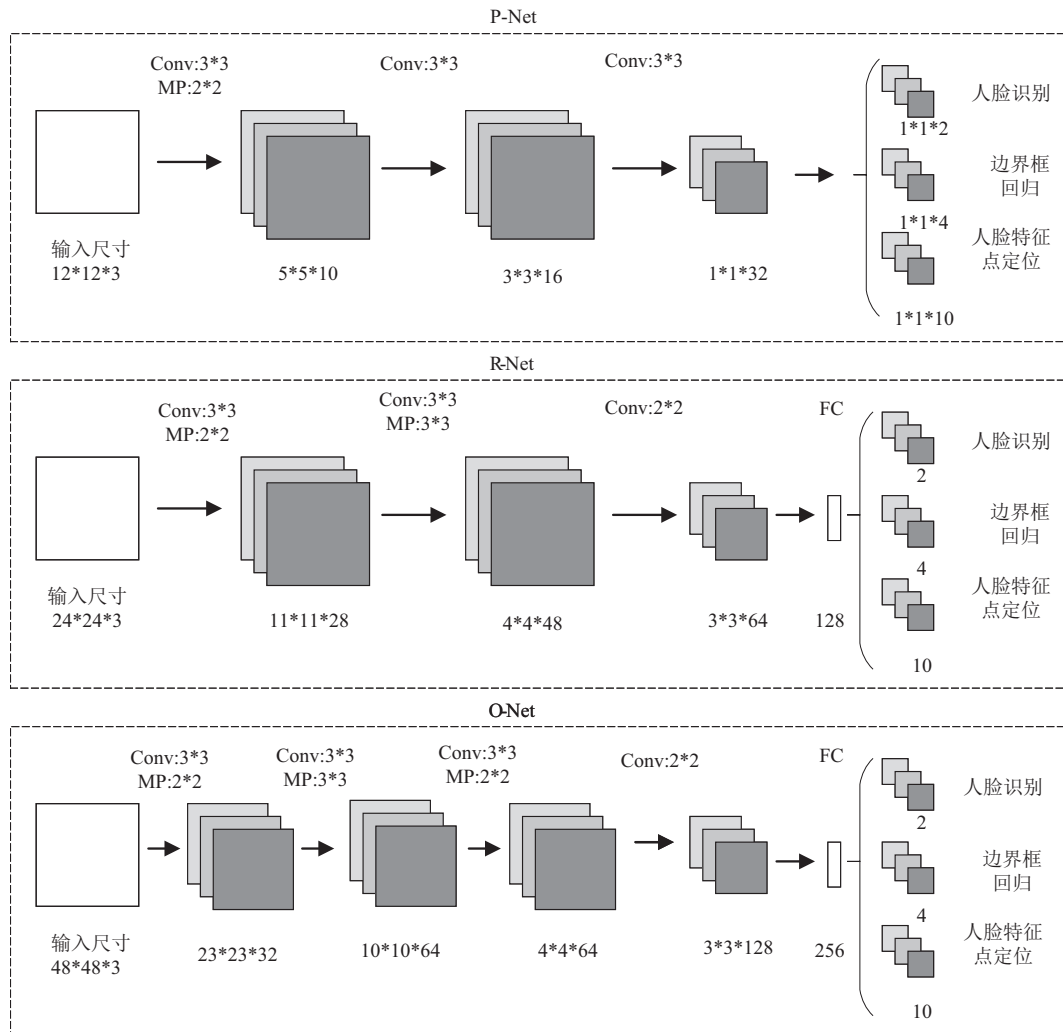


图2 P-Net, R-Net, O-Net 的网络结构

同时,每个网络也都是一个多任务的网络,所处理的任务分别为人脸有无的检测、人脸边界框的确定以及对五个特征点的定位.

首先在判断图像是否包括人脸时采用的交叉熵损失函数为:

$$L_i^{\text{det}} = -(y_i^{\text{det}} \log(p_i) + (1 - y_i^{\text{det}})(1 - \log(p_i))) \quad (1)$$

在人脸边界框回归时,采用的欧氏距离损失函数为:

$$L_i^{\text{box}} = \left\| \widehat{y}_u^{\text{box}} - y_u^{\text{box}} \right\| \quad (2)$$

在确定人脸的特征点时采用的欧式距离损失函数为:

$$\min \sum_{i=1}^N \sum_{j \in \{\text{det}, \text{box}, \text{landmark}\}} \alpha_j \beta_i^j L_i^j \quad (3)$$

其中, α_j 表示三个任务在当前阶段的网络中所占的重要性, $\beta_i^j \in \{0,1\}$ 为当前采样的指示器。

由于本文中采用的数据集为身份证照片,即能够保证数据集中每张图像有且只有一张很清晰的人脸,因此在提取特征向量的人脸检测过程中可以省略三组网络中的前两组,只保留最后的 O-Net,经过实验证实检测的准确率依然可以满足要求。

1.2 普氏分析法进行人脸对齐

普氏分析^[11]是一种用来分析形状分布的统计方法,从数学上来讲,就是利用最小二乘法寻找形状 A 到形状 B 的仿射变换。

仿射变换是空间中直角坐标变换的一种,它是二维坐标之间的一个线性变换。因此,仿射变换也可以通过一系列的原子变换来叠加复合实现。具体的原子变换包括平移、缩放、旋转和倒置等。

仿射变换将原坐标(x,y)变换为新坐标(x',y')的计算方法为:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} a_{00} * x + a_{01} * y + a_{02} \\ a_{10} * x + a_{11} * y + a_{12} \\ 1 \end{bmatrix} \quad (4)$$

其中变换矩阵 $A = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ 0 & 0 & 1 \end{bmatrix}$ 。

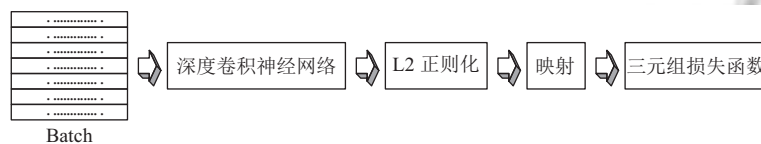


图3 FaceNet 模型结构

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 \quad (5)$$

$f(x_i^a)$ 是表示 Anchor 样本的特征向量, α 表示间隔。最终要最小化的损失函数为:

$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right] \quad (6)$$

1.4 特征向量相似度衡量

人脸图像经过前面的人脸检测、特征点定位、人脸对齐、特征向量提取之后,已经被表示成为 128 维的特征向量,接下来计算测试图像特征向量与人脸库各图像特征向量之间的欧式距离,再选取欧式距离最

使用普氏分析进行人脸对齐时,根据获取的人脸特征点的位置,通过最小二乘法求解参考图像特征点与测试图像特征点之间的变换矩阵,然后对测试图像进行变换以获得对齐之后的测试图像。

1.3 基于 FaceNet 的人脸特征提取

FaceNet 是 Google 提出的人脸识别算法,该算法是基于海量人脸数据训练的深度卷积神经网络,可以将人脸图像映射成 128 维的特征向量。相比于其他人脸识别算法,FaceNet 并没有使用 softmax 进行分类学习,而是采用端对端的方式,直接将图像编码为欧式空间的特征向量,进而基于学习到的特征向量进行人脸验证、识别等应用。其网络结构如图 3 所示,在网络的最前面就是一个传统的卷积神经网络,去掉 softmax 层后,经过 L2 范数的归一化就可以得到特征表示,最后得到一个三元组损失函数 (Triplet Loss)。Triplet 是由三个名为 Anchor, Positive, Negative 三个样本组成的训练三元组,其中 Anchor 是随机选取的一个训练样本,Positive 表示和 Anchor 样本同属一个类别的训练样本, Negative 表示和 Anchor 样本属于不同类别的训练样本。Triplet Loss 基于一个 Anchor 样本的特征向量和其 Positive 样本的特征向量之间的距离一定小于其和 Negative 样本的特征向量之间的距离,并且至少要小于一个阈值以增强模型的判别能力,如公式 (5) 所示。

小的图像最为最佳匹配结果即可完成最终的人脸识别。

2 实验结果与分析

本节对比了人脸识别系统中各环节不同算法对最终识别准确率的影响。人脸库选用 15 000 幅身份证照片,测试图像使用 1416 幅图像。测试图像集由包含多种分辨率、清晰度、角度和年龄跨度的人员生活照片组成。在实验过程中,使用 MTCNN 算法进行人脸检测和特征点定位,对比了两种不同的人脸对齐算法和三种不同的人脸识别模型对识别精度的影响。其中,第一种人脸对齐算法为旋转人脸图像双眼特征点为水平关

系,下文简称为“旋转对齐”,第二种对齐算法为上文介绍的普氏分析法.三种人脸识别模型分别为 Vgg-Face, Caffe-Face 和 Facenet. 人脸识别精度统计了两种识别结果,分别是识别结果中置信度第一的识别结果为正确结果的概率和识别结果中前十置信度的识别结果中包含正确结果的概率,分别简称为 top-1 和 top-10.

本文所有的实验过程都基于 Windows10 操作系统,硬件环境为 Intel Xeon E3-1230(3.40 GHz) CPU, 32 GB RAM, NVIDIA GeForce GTX TITAN X, 采用的深度学习框架为 TensorFlow 1.2.1.

2.1 不同对齐方法识别准确率对比

实验过程中首先对比了两种对齐算法对识别精度的影响,实验过程使用 Vgg-Face 进行人脸特征提取,使用欧式距离进行相似度衡量.实验结果如表 1 所示.

表 1 不同对齐方法人脸识别准确率(单位: %)

	旋转对齐	普氏分析对齐
Top-1	30.9	45.9
Top-10	54.5	66.8

从表 1 中可以看出,使用普氏分析对齐进行人脸识别的精度明显优于旋转对齐方法.这是由于旋转对齐时只是旋转图像使得双眼水平,并没有将各图像的特征点进行严格意义的对齐.而普氏分析法通过最小二乘法迭代减小测试图像与参考图像特征点之间的距离,从而实现测试图像和参考图像之间的严格意义的对齐.基于表 1 的实验结果,下文的实验都使用普氏分析法进行人脸对齐.

2.2 不同人脸特征提取算法识别准确率对比

实验过程中对比了 Vgg-Face, Caffe-Face 和 FaceNet 进行人脸特征提取对识别准确率的影响,实验过程中使用欧式距离衡量相似度,结果如表 2 所示.从表中可以看出,FaceNet 识别准确率效果最好.

表 2 不同人脸特征提取算法识别准确率(单位: %)

	Vgg-Face	Caffe-Face	FaceNet
Top-1	45.9	66.7	81.2
Top-10	66.8	84.1	88.8

2.3 不同特征相似度衡量方法识别准确率对比

本节对比了两种特征相似度衡量的识别准确度.第一种方法是上文提到的欧式距离,第二种方法是基于人脸图像提取的 128 维特征进行了特征重映射,通过训练 Siamese^[12]网络将特征映射为 512 维的特征,网

络包含两个添加了 BN 层的全连接层,损失函数使用对比损失.表 3 为使用 FaceNet 进行人脸特征提取之后原始特征和使用 Siamese 网络重映射之后的特征进行相似度衡量得到的识别精确率对比.

表 3 使用原始特征与 Siamese 重映射特征精确率对比(单位: %)

	原始特征	Siamese重映射特征
Top-1	81.2	82.1
Top-10	88.8	91.3

综合表 3 和表 4 可以看出,在计算时间效率基本接近的情况下,使用 Siamese 进行特征重映射优于原始进行相似度衡量的识别准确率,尤其是 top-10 准确率提升了 2.5%.现阶段受限于训练样本的数量,为避免过拟合无法增大模型的容量,后续可以增大样本量进行更大模型容量的实验.

表 4 总 1416 幅测试图像原始特征与 Siamese 重映射特征相似度衡量时间效率对比(单位: s)

	原始特征	Siamese重映射特征
时间	1.45	1.58

3 结束语

本文针对安防领域对人脸识别的业务需求,进行了基于深度学习算法的证件照人脸识别的研究与实现.对人脸检测、特征点定位、人脸对齐、特征向量提取和相似度衡量五个人脸识别的关键环节都进行了实现并对比了多种算法,在此基础上,本文又进行了人脸特征向量的 Siamese 重映射以进一步提升识别精度.实验结果表明,本文实现的人脸识别系统,可以基于公安证件照进行良好的人脸识别,同时具有良好的时间效率,实现了对人脸图像的高效精确识别,具有良好的应用价值和市场前景.

参考文献

- 1 李月龙, 靳彦, 汪剑鸣, 等. 人脸特征点提取方法综述. 计算机学报, 2016, 39(7): 1356-1374. [doi: 10.11897/SP.J.1016.2016.01356]
- 2 吴翔, 唐焕华, 刘锦高. 基于双目视觉的立体人脸识别系统. 计算机系统应用, 2009, 18(5): 61-64.
- 3 黄金钰, 张会林, 闫日亮. LBP 直方图与 PCA 的欧式距离的人脸识别. 计算机系统应用, 2012, 21(6): 202-204, 198.
- 4 Huang GB, Mattar M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in

- unconstrained environments. Technical Report 07-49. Amherst: University of Massachusetts, 2007.
- 5 Zhu ZY, Luo P, Wang XG, *et al.* Deep learning identity-preserving face space. Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, NSW, Australia. 2013. 113–120.
 - 6 Taigman Y, Yang M, Ranzato MA, *et al.* DeepFace: Closing the gap to human-level performance in face verification. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 1701–1708.
 - 7 Sun Y, Wang XG, Tang XO. Deep learning face representation from predicting 10,000 classes. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 1891–1898.
 - 8 Zhou EJ, Fan HQ, Cao ZM, *et al.* Extensive facial landmark localization with coarse-to-fine convolutional network cascade. Proceedings of 2013 IEEE International Conference on Computer Vision Workshops. Sydney, NSW, Australia. 2013. 386–391.
 - 9 Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 815–823.
 - 10 Zhang KP, Zhang ZP, Li ZF, *et al.* Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 2016, 23(10): 1499–1503. [doi: [10.1109/LSP.2016.2603342](https://doi.org/10.1109/LSP.2016.2603342)]
 - 11 Gower JC. Generalized Procrustes analysis. Psychometrika, 1975, 40(1): 33–51. [doi: [10.1007/BF02291478](https://doi.org/10.1007/BF02291478)]
 - 12 Bertinetto L, Valmadre J, Henriques JF, *et al.* Fully-convolutional Siamese networks for object tracking. In: Hua G, Jégou H, eds. Computer Vision – ECCV 2016 Workshops. Cham: Springer, 2016. 850–865.