Journal of Computer Applications
计算机应用,2020,40(3):704 - 709
ISSN 1001-9081
CODEN JYIIDU
2020-03-10
http://www.joca.cn

文章编号:1001-9081(2020)03-0704-06
DOI:10.11772/j.issn.1001-9081.2019071272

基于深度学习的轻量级和多姿态人脸识别方法

龚　锐[1,2*],丁　胜[1,2,3],章超华[1,2],苏　浩[1,2]

(1.武汉科技大学 计算机科学与技术学院,武汉 430065; 2.智能信息处理与实时工业系统湖北省重点实验室(武汉科技大学),武汉 430065; 3.福建省大数据管理新技术与知识工程重点实验室(泉州师范大学),福建 泉州 362000)

(*通信作者电子邮箱516123131@qq.com)

摘　要:目前基于深度学习的人脸识别方法存在识别模型参数量大、特征提取速度慢的问题,而且现有人脸数据集姿态单一,在实际人脸识别任务中无法取得好的识别效果。针对这一问题建立了一种多姿态人脸数据集,并提出了一种轻量级的多姿态人脸识别方法。首先,使用多任务级联卷积神经网络(MTCNN)算法进行人脸检测,并且使用MTCNN最后包含的高层特征做人脸跟踪;然后,根据检测到的人脸关键点位置来判断人脸姿态,通过损失函数为ArcFace的神经网络提取当前人脸特征,并将当前人脸特征与相应姿态的人脸数据库中的人脸特征比对得到人脸识别结果。实验结果表明,提出方法在多姿态人脸数据集上准确率为96.25%,相较于单一姿态的人脸数据集,准确率提升了2.67%,所提方法能够有效提高识别准确率。

关键词:深度学习;人脸识别;多姿态;轻量级;多任务级联卷积神经网络;ArcFace

中图分类号:TP183　　**文献标志码:**A

Lightweight and multi-pose face recognition method based on deep learning

GONG Rui[1,2*], DING Sheng[1,2,3], ZHANG Chaohua[1,2], SU Hao[1,2]

(1. *College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan Hubei* 430065, *China*;
2. *Hubei Provincial Key Laboratory of Intelligent Information Processing and Real-time Industrial System (Wuhan University of Science and Technology), Wuhan Hubei* 430065, *China*;
3. *Fujian Provincial Key Laboratory of Data Intensive Computing (Quanzhou Normal University), Quanzhou Fujian* 362000, *China*)

Abstract: At present, the face recognition methods based on deep learning have the problems of large model parameter size and slow feature extraction speed, and the existing face datasets have the problem of single pose, which cannot achieve good recognition effect in the actual face recognition task. Aiming at this problem, a multi-pose face dataset was established, and a lightweight multi-pose face recognition method was proposed. Firstly, the MTCNN (Multi-Task cascaded Convolutional Neural Network) algorithm was used by the method for face detection, and the high-level features included in the last network of MTCNN were used for face tracking. Then, the face pose was judged according to the positions of the detected face key points, the current face features were extracted by the neural network with ArcFace as loss function, and the current face features were compared with the face features of the corresponding pose in the face database to obtain the face recognition result. The experimental results show that the accuracy of the proposed method is 96.25% on the multi-pose face dataset, which is 2.67% higher than that on the face dataset with single pose. It shows that the proposed multi-pose face recognition method can effectively improve the recognition accuracy.

Key words: deep learning; face recognition; multi-pose; lightweight; Multi-Task cascaded Convolutional Neural Network (MTCNN); ArcFace

0　引言

人脸识别是基于人的脸部特征信息进行身份验证的一种生物识别技术,通过带有摄像头的终端设备拍摄人的行为图像,使用人脸检测算法,从原始行为图像中得到人脸区域,用特征提取算法提取人脸的特征,并根据这些特征确认身份。传统的人脸识别方法包含主成分分析(Principal Component Analysis, PCA)[1]、拉普拉斯特征映射[2]、局部保持映射(Locality Preserving Projection, LPP)[3]、稀疏表示[4]等。这些传统人脸识别算法在提取特征时,往往是通过人工手段去获取样本特征,而人为设定的特征在特征提取和识别过程存在提取特征过程复杂、识别效率低等不足,而且受人脸的姿态转动、光照变化、遮挡等复杂场景下影响较大,这些方法所使用的特征属于浅层特征,不能从原始图像中获取本征的高层语义特征。

自2012年AlexNet在ImageNet[5]比赛上取得冠军后,卷积

收稿日期:2019-07-22;**修回日期:**2019-09-26;**录用日期:**2019-09-26。

基金项目:福建省大数据管理新技术与知识工程重点实验室开放课题(BD201805)。

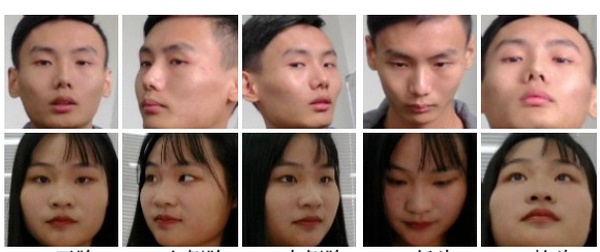
作者简介:龚锐(1995—),男,湖北赤壁人,硕士研究生,主要研究方向:深度学习、计算机视觉; 丁胜(1975—),男,湖北武汉人,副教授,博士,CCF会员,主要研究方向:计算机视觉; 章超华(1996—),男,江西抚州人,硕士研究生,主要研究方向:深度学习、计算机视觉; 苏浩(1996—),男,湖北武汉人,硕士研究生,主要研究方向:迁移学习、计算机视觉。

神经网络愈发火热。随着 VGGNet(Visual Geometry Group Networks[6]、GoogleNe[7]和 ResNet(Residual Neural Networks)[8]的相继提出,深度卷积神经网络将多个计算机视觉任务的性能提升到了新的高度,总体的趋势是为了达到更高的准确性构建了更深更复杂的网络。人脸识别作为一种重要的身份认证技术,被应用于越来越多的移动端和嵌入式设备上,如设备解锁、应用登录、刷脸支付等。由于移动设备计算能力以及存储空间的限制,部署在移动设备上的人脸识别模型要满足准确率高、模型小、特征提取速度快的条件,这些大型网络在尺度和速度上不能满足移动设备的要求。MobileNetV1[9]提出了深度可分离卷积(depthwise separable convolutions),将标准卷积分解成深度卷积和逐点卷积,大幅度地降低了参数量和计算量。ShuffleNet[10]使用分组逐点卷积(group pointwise convolution),通过将卷积运算的输入限制在每个组内,显著地减少了计算量,再使用通道打乱(channel shuffle)操作,解决了分组卷积中不同组间信息不流通的问题。MobileFaceNet[11]从感受野的角度,分析了人脸识别任务中全局平均池化的缺点,使用全局深度卷积代替。

此外,人脸图像较大的类内变化和较小的类间差异是人脸识别任务的一个困难点。由于人脸的角度、光照、表情、年龄、化妆、遮挡、图片质量等变化,同一个人的不同人脸图像具有很大差异,如何让计算机在较大的类内变化的干扰下依然能够辨识到比较微弱的类间变化,是人脸识别的主要挑战。人脸识别要从两个方向优化:一是增加不同人之间的距离(类间距离);二是降低同一个人在不同人脸图像上的距离(类内距离)。基于度量学习的 Triplet Loss[12]的目标是使同类样本之间的距离尽可能缩小,不同类样本之间的距离尽可能放大。Center Loss[13]提出为每一个类提供一个类别中心,并最小化每个样本与该中心的距离。二者都是基于 Softmax Loss 的,存在不足。后来出现的损失函数 SphereFace[14]、CosFace(Large Margin Cosine Loss)[15]和 ArcFace(Additive Angular Margin Loss)[16]从根本上改进,要求扩大分类面之间的 margin,增强 Softmax 损失函数的判别能力,能够较好地满足增大类间距离、减小类内距离的要求。

1　多姿态人脸数据集

现有的深度学习方法是将对齐后的人脸图像通过深层网络提取人脸特征,由于忽略了人脸的姿态特征,因而导致识别精度不够高。针对这个问题,本文建立了一个多姿态的人脸数据集,如图 1 所示。



(a) 正脸　(b) 左侧脸　(c) 右侧脸　(d) 低头　(e) 抬头

图 1　多姿态人脸数据集

Fig. 1　Multi-pose face dataset

该数据集由 800 名志愿者每人 5 张人脸图像(5 个姿态,分别为正脸、左侧脸、右侧脸、抬头、低头)组成,共 4 000 张。采集过程为:在光照条件良好的环境下放置摄像头,通过多任务级联卷积神经网络(Multi-Task cascaded Convolutional Neural Network, MTCNN)[17]算法对视频流进行人脸检测,根据检测到的人脸关键点(左右眼、鼻尖、左右嘴角)判断人脸姿态,判断方法见 1.3 节,采集完 5 个姿态后将对齐后的 5 张图像缩放到 112×112 大小,保存起来,同时记录对应人脸的姓名年龄性别作为人脸识别的标签。

1.1　MTCNN 算法

MTCNN 算法用于人脸检测任务,采用三个级联的网络 P-Net(Proposal Network)、R-Net(Refine Network)和 O-Net(Output Network)由粗到细,通过减少滤波器数量、设置小的卷积核和增加网络结构的深度,在较短的时间内获得很好的性能,在光照变化、部分遮挡和人脸转动等情况下也能得到很好的人脸检测结果。

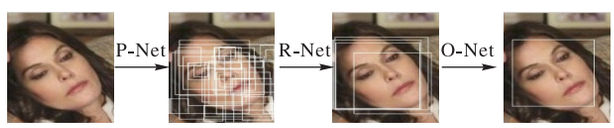


图 2　MTCNN 算法流程

Fig. 2　Flow chart of MTCNN algorithm

1.2　人脸对齐

MTCNN 算法可以检测出人脸在图像中的位置,同时还能检测人脸关键点。人脸关键点可以用来做人脸对齐,将不同角度的人脸对齐后有利于人脸识别,人脸对齐步骤如图 3 所示。根据左右眼睛和鼻子的位置,通过仿射变换,将人脸对齐,并裁剪缩放到 112×112。

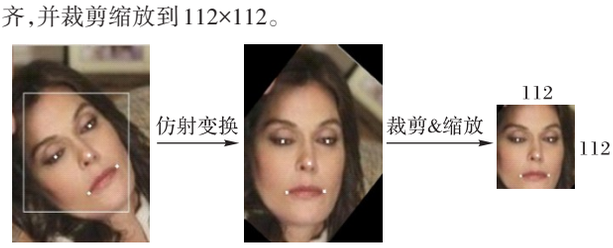


图 3　人脸对齐

Fig. 3　Face alignment

1.3　人脸姿态判断

MTCNN 算法检测的原图和检测后关键点位置如图 4 所示。本文使用左右眼、鼻尖和左嘴角这四个关键点判断人脸姿态方向,记左眼坐标为(x_1, y_1),右眼坐标为(x_2, y_2),鼻尖坐标为(x_3, y_3),左嘴角坐标为(x_4, y_4)。以正脸图为基准,从横轴看,鼻尖近似在左右眼坐标中心;从纵轴看,鼻尖近似在左眼和左嘴角坐标中心,因此满足如下式子:

$$\begin{cases}(x_1 + x_2)/2 \approx x_3\\(y_1 + y_4)/2 \approx y_3\end{cases}$$

同理,当前人脸图像是左侧脸图的条件是:

$$\begin{cases}(x_1 + x_2)/2 = x_3 + D;\ D > 0\\(y_1 + y_4)/2 \approx y_3\end{cases}$$

当前人脸图像是右侧脸图的条件是:

$$\begin{cases}(x_1 + x_2)/2 + D = x_3;\ D > 0\\(y_1 + y_4)/2 \approx y_3\end{cases}$$

当前人脸图像是低头图的条件是:

$$\begin{cases}(x_1 + x_2)/2 \approx x_3 \\ (y_1 + y_4)/2 + D \approx y_3;\ D > 0\end{cases}$$

当前人脸图像是抬头图的条件是：

$$\begin{cases}(x_1 + x_2)/2 \approx x_3 \\ (y_1 + y_4)/2 \approx y_3 + D;\ D > 0\end{cases}$$

在上面4个式子中，D是一个大于0的值，表示中心点与鼻尖的横轴距离或纵轴距离，在本实验中取值为10，即中心pixel点与鼻尖距离相差10 px(pixel)就算满足条件。

(a) 原图

(b) 关键点

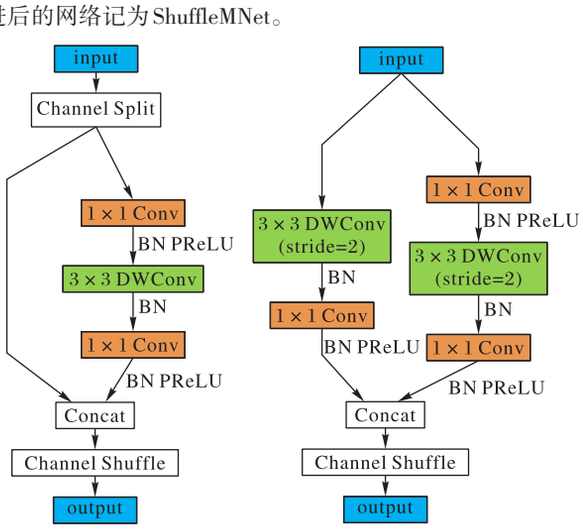
图 4 原图与人脸检测后的关键点

Fig. 4 Original image and key points after face detection

2 主要方法

2.1 特征提取网络

文献[18]提出了一种高效网络模块，如图5所示，在本文中记为mainblock。本文结合ShuffleNetV2的优点对MobileFaceNet进行了改进，改进后的网络结构如表1所示，改进后的网络记为ShuffleMNet。



(a) 基础模块 (b) 下采样模块

图 5 ShuffleMNet的mainblock

Fig. 5 Mainblock of ShuffleMNet

在网络的开始先使用步长为2的卷积核对输入进行下采样，再接上多个mainblock模块得到7×7的特征图，使用全局深度卷积(Global Depthwise Convolution，GDC)取代全局平均池化(Global Average Pooling，GAP)得到1×1的特征图，最后接上一个全连接层得到128维特征向量。

每个mainblock模块先对输入进行1次如图5(b)所示的步长为2的下采样，之后重复n−1次图5(a)所示的操作。图5(a)中首先对输入进行通道分割(Channel Split)，将通道为c的输入分成2组通道为c/2的分支，左边直接映射，右边是一个类Inverted residuals[19]模块，然后左右两边特征图合并，最后进行通道打乱(Channel Shuffle)使得各个通道之间的信息相互交通。图5(b)中左边分支是步长为2的深度可分离卷积，右边分支是步长为2的类Inverted residuals模块，之后通过左右特征图合并、通道打乱得到输出。其中两个类Inverted residuals模块除了空间下采样时第一个1×1卷积进行升维，其他所有卷积操作输入输出通道数都相同。使用参数化修正线性单元(Parametric Rectified Linear Unit，PReLU)作为激活函数。

表 1 ShuffleMNet结构

Tab. 1 ShuffleMNet architecture

Input	Operator	c	n	S
$112^2 \times 3$	Conv3×3	64	1	2
$56^2 \times 64$	depthwise Conv3×3	64	1	1
$56^2 \times 64$	mainblock	64	1	2
			4	1
$28^2 \times 64$	mainblock	128	1	2
			3	1
$14^2 \times 128$	mainblock	128	1	2
			6	1
$7^2 \times 128$	Conv1×1	512	1	1
$7^2 \times 512$	linear GDConv7×7	512	1	1
$1^2 \times 512$	linear Conv1×1	128	1	1

2.2 人脸匹配

将特征提取得到的128维特征与人脸数据库中的所有人脸特征进行比对。相较于欧氏距离，余弦距离使用两个向量夹角的余弦值衡量两个个体间差异的大小，后者更加注重两个向量在方向上的差异，因此，本实验人脸特征比对采用的度量方法是计算余弦距离远近，距离越近(相似度越大)表示两张人脸越相似。本次实验设置阈值设为0.7，若余弦相似度大于0.7，则表示这张图像匹配到了数据库中对应的人。如果存在多张人脸与当前人脸的余弦相似度大于阈值，取相似度最高的为匹配结果。

2.3 人脸跟踪

文献[20]使用核相关滤波(Kernelized Correlation Filters，KCF)方法用于人脸跟踪，文献[21]通过在MTCNN算法后新加一个网络做人脸跟踪。本文使用更简单的方法做人脸跟踪。

MTCNN最后一个网络O-Net的网络结构如图6所示。由于MTCNN算法是通过级联网络由粗到细完成人脸检测任务，O-Net第一个全连接层是256维的高层特征，包含了人脸bounding box回归和人脸关键点的信息，可以利用这256维特征作为人脸跟踪的依据，并判断两个人脸框中心点的欧氏距离进一步确定是否是同一人脸。

人脸跟踪算法流程。

提取第t+1帧的所有人脸框的O-Net的全连接层特征

FOR 第t+1帧的每个特征

  与第t帧的所有特征计算余弦相似度

  IF 最大相似度大于0.8

    IF 两个人脸框的中心点欧氏距离小于25 px

      两帧对应的人脸是同一目标

    ELSE

      将当前人脸标记为新目标

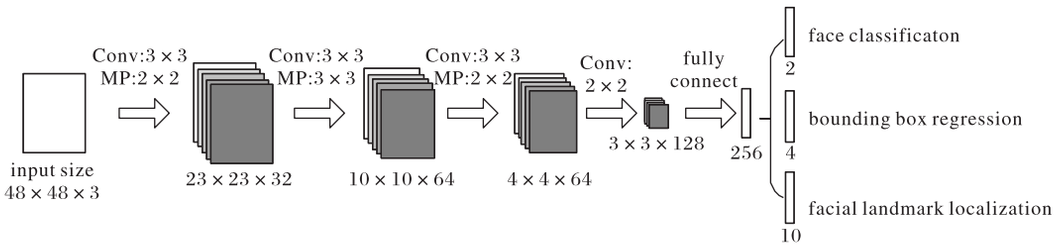


图 6　MTCNN的最后一个网络O-Net

Fig. 6　MTCNN's last network O-Net

3　实验

3.1　实验环境

本实验使用远程服务器训练深度神经网络模型,该服务器配有64位的CentOS系统,配备了Inter Xeon E5 2620 v4处理器,64 GB内存,8张Tesla V100显卡,每张显卡显存为32 GB。

3.2　训练数据集与校验数据集

使用由DeepGlint公开的Asian-Celeb人脸数据集作为训练数据集训练深度卷积神经网络模型。该数据集包含93 979个纯亚洲人,共2 830 146张已经对齐好的人脸图像,规模有93.8 GB。本文提出的多姿态人脸数据集总800人共4 000张,取其中640人共3 200张人脸图像作为训练数据集,其他800张图像作为测试集。

校验数据集包含LFW(Labeled Faces in the Wild)[22]数据集、CFP(Celebrities in Frontal Profile)[23]数据集和Age-DB(Age Database)[24]数据集。其中,LFW数据集包含5 749人共13 233张人脸图像,姿态表情和光照变化较大,使用其中的6 000对人脸图像作为校验数据集;CFP数据集包含500个身份,每个身份有10个正脸,4个侧脸,本实验使用CFP数据集中的frontal-profile(FP)人脸验证,即CFP-FP(CFP with Frontal-Profile)数据集;AgeDB数据集包含440人共12 240张人脸图像,本次实验使用AgeDB-30,包含300正样本对和300负样本对。

3.3　训练前的设置

3.3.1　预处理设置

为了提高人脸识别方法的鲁棒性,使用图像增强的方法对训练集进行数据扩充,使用的方法为左右翻转和模拟光照变化改变图像明暗度,使得训练集数量直接扩大了6倍。之后在输入时对输入图像RGB三通道的像素值进行归一化处理,即对原来RGB三个通道的像素值都减去127.5再除以128,使像素值都在(−1,1)区间内。

3.3.2　训练样本设置

相较于Asian-Celeb数据集,本文提出的多姿态人脸数据集样本数量远少于前者。由于实际应用场景往往与标准数据集的场景不同,需要提高训练时本文数据集所占的比重,所以改进了训练时样本选取的方式。之前样本选取方式完全随机,并且每一轮每个样本仅出现一次,多姿态数据集对整体权重贡献比较小。改进后,通过修改源程序使得每个batch(batch size为256)都会包含2个来自多姿态数据集的随机样本。这样大大提升多姿态数据集对整个网络权重的贡献,提高了其所占的比重。

3.3.3　网络参数设置

网络的下训练分2步。首先使用Softmax loss预训练,可以学习到可分的深度人脸特征,然后使用ArcFace loss训练,学习使得类内紧凑、类间分离的人脸特征。

Softmax训练是使用的学习率为0.1,共训练120 000步,ArcFace训练初始学习率为0.1,迭代到120 000、160 000、180 000步时学习率每次都除以10,总共训练200 000步。全局深度可分离卷积层的权重衰减参数设置为4E−4,其他权重衰减参数设置为4E−5;使用随机梯度下降策略(Stochastic Gradient Descent,SGD)优化模型,动量设置为0.9;batch size设置为256。

3.4　实验结果

第一步采用Softmax loss训练的训练集精度和验证集精度如图7所示,网络训练到42 000步时,训练集精度已经达到81.5%,LFW精度为99.03%,AgeDB-30精度为88.45%,CFP-FP精度为84.84%,此时停止Softmax训练开始使用ArcFace loss进行训练。使用ArcFace loss训练的验证集精度如图8所示,网络训练到160 000步时网络已基本收敛,停止训练,此时LFW精度为99.58%,AgeDB-30精度为96.08%,CFP-FP精度为88.46%。

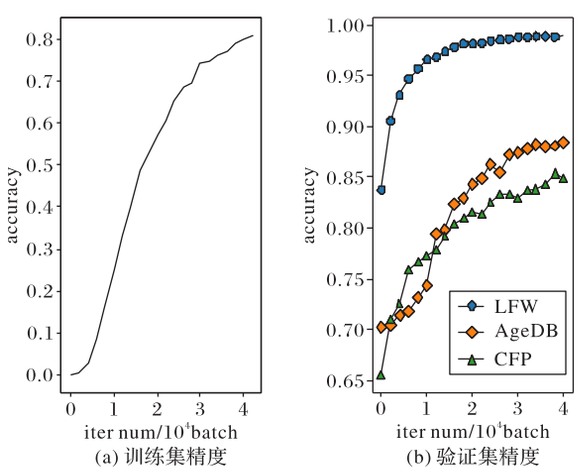


图 7　使用Softmax loss的训练曲线

Fig. 7　Training curve using Softmax loss

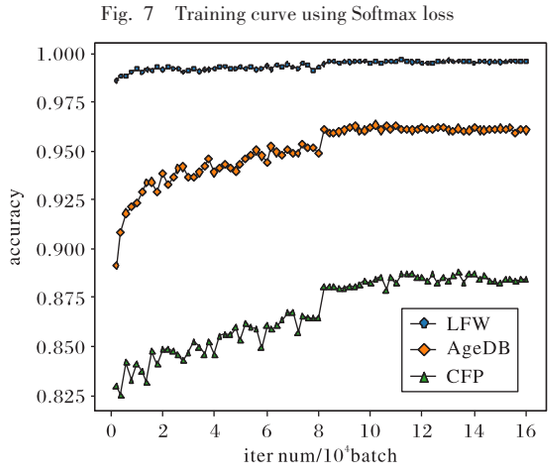


图 8　使用ArcFace loss的训练曲线

Fig. 8　Training curve using ArcFace loss

为了验证本文提出的轻量级网络的有效性，实验并记录了使用其他轻量级网络同样使用上面二步式训练的结果，训练时并未加入本文建立的数据集，仅使用Asian-Celeb数据集训练，实验结果如表2所示。

表 2 ShuffleMNet与其他轻量级网络精度对比 单位：%

Tab. 2 Accuracy comparison of ShuffleMNet and other lightweight networks unit：%

网络	LFW	AgeDB-30	CFP
MobileNetV1	98.69	90.38	87.01
MobileNetV2	98.57	90.23	86.86
MobileFaceNet	99.50	95.91	88.34
ShuffleMNet	**99.58**	**96.08**	**88.57**

3.5 人脸识别

3.5.1 人脸数据库

模型训练好后，使用模型对测试集的800张人脸提取128维人脸特征，将人脸姿态（正脸、左侧脸、右侧脸、低头、抬头）与人脸信息（姓名、性别、年龄）作为标记一起保存到数据库中。

3.5.2 实地测试

重新对测试集中的160名志愿者进行人脸识别测试。在光照条件良好的场景布置摄像头，再次获取这些志愿者的5种人脸姿态图像做测试。使用两种不同的方法进行人脸识别，并统计结果。

第一种直接对每个志愿者的5张人脸图像提取特征后，分别和人脸数据库中的所有特征进行比对，阈值取0.7，余弦相似度大于0.7且人脸信息（姓名、性别、年龄）相同则表示人脸识别正确。

第二种是根据5张人脸图像中每张人脸图像的姿态（由1.3节的方法自动判断人脸姿态，不再是手动获取姿态）选择对应姿态的特征进行比对，即假设当前要对比的正脸图，则只选择人脸数据库中所有正脸图对应的人脸特征进行比对，比对方法同上。

两种方法的实验结果如表3所示，其中使用多姿态数据集意味着在训练时按照3.3.2节中的方法加入多姿态数据集。实验结果表明，第二种比对方法进行人脸识别效果更好。

表 3 不同方法的准确率对比 单位：%

Tab. 3 Accuracy comparison of different methods unit：%

比对方法	使用多姿态数据集	不使用多姿态数据集
根据姿态比对	**96.25**	94.73
直接比对	93.75	92.36

3.6 时间测试

本实验采用的测试平台是虚拟机64位Ubuntu 18.04系统，处理器是Inter Core i5-8400 CPU @2.80 GHz，内存为2 GB。使用NCNN框架作神经网络前向推导。

测试结果如表4所示。从表可以看出，ShuffleMNet比MobileFaceNet速度要快9 ms，此外，由2.3节可以知道，人脸跟踪提取了MTCNN第三个网络全连接层的特征，这里仅仅是前向推导时顺便提取出来的，不需要耗费其他时间做这个特征提取，剩下的就是相似度计算匹配，因此本文的人脸跟踪算法几乎不花费额外的时间，记录检测和识别的总时间仅仅是二者相加而已。

表 4 不同网络及其组合的时间对比 单位：ms

Tab. 4 Time comparison of different networks and their combinations unit：ms

使用方法	时间
MTCNN（*minsize*=40）	40
MobileFaceNet	37
ShuffleMNet	**28**
MTCNN+ MobileFaceNet+人脸跟踪	77
MTCNN+ ShuffleMNet +人脸跟踪	68

4 结语

本文建立了一个多姿态人脸数据集，解决了目前单一姿态数据集存在的人脸识别准确率不高的问题。另外，本文提出的网络ShuffleMNet相较于MobileFaceNet等其他轻量级网络精度更高，速度更快，可以很好地在移动设备上部署。同时，在数据采集时采集了志愿者的年龄性别，有助于后续的人脸年龄性别识别研究。

参考文献（References）

[1] JOLLIFFE I T. Principal component analysis[J]. Journal of Marketing Research, 2002, 25(4):513.

[2] BELKIN M, NIYOGI P. Laplacian eigenmaps and spectral techniques for embedding and clustering[C]// Proceedings of the 14th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2001:585-591.

[3] HE X, NIYOGI P. Locality preserving projections[C]// Proceedings of the 16th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2003:153-160.

[4] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.

[5] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.

[6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2018-04-10]. https://arxiv.org/pdf/1409.1556.pdf.

[7] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015:1-9.

[8] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9908. Cham: Springer, 2016:630-645.

[9] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. [2018-10-15]. https://arxiv.org/pdf/1704.04861.pdf.

[10] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[EB/OL]. [2018-10-07]. https://arxiv.org/pdf/1707.01083.pdf.

[11] CHEN S, LIU Y, GAO X, et al. MobileFaceNets: efficient CNNs for accurate real-time face verification on mobile devices[EB/OL]. [2018-06-15]. https://arxiv.org/ftp/arxiv/papers/1804/1804.07573.pdf.

[12] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern

Recognition. Piscataway: IEEE, 2015: 815-823.

[13] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition[C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9911. Cham: Springer, 2016: 499-515.

[14] LIU W, WEN Y, YU Z, et al. SphereFace: deep hypersphere embedding for face recognition[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017:6738-6746.

[15] WANG H, WANG Y T, ZHOU Z, et al. CosFace: large margin cosine loss for deep face recognition[EB/OL]. [2018-04-03]. https://arxiv.org/pdf/1801.09414.pdf.

[16] DENG J, GUO J, XUE N, et al. ArcFace: additive angular margin loss for deep face recognition[EB/OL]. [2019-02-09]. https://arxiv.org/pdf/1801.07698.pdf.

[17] 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.

[18] MA N, ZHANG X, ZHENG H, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[EB/OL]. [2018-12-10]. https://arxiv.org/pdf/1807.11164.pdf.

[19] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018:4510-4520.

[20] 任梓涵,杨双远. 基于视觉跟踪的实时视频人脸识别[J]. 厦门大学学报(自然科学版), 2018, 57(3):438-444. (REN Z H, YANG S Y. Real-time face recognition in videos based on visual tracking[J]. Journal of Xiamen University (Natural Science), 2018, 57(3):438-444.)

[21] 方国康,李俊,王垚儒. 基于深度学习的ARM平台实时人脸识别[J]. 计算机应用, 2019, 39(8): 2217-2222. (FANG G K, LI J, WANG Y R. Real-time face recognition based on ARM platform based on deep learning[J]. Journal of Computer Applications, 2019, 39(8): 2217-2222.)

[22] HUANG G B, MATTAR M, BERG T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments[EB/OL]. [2019-01-20]. https://people.cs.umass.edu/~elm/papers/lfw.pdf.

[23] SENGUPTA S, CHEN J C, CASTILLO C, et al. Frontal to profile face verification in the wild[C]// Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2016: 1-9.

[24] MOSCHOGLOU S, PAPAIOANNOU A, SAGONAS C, et al. AgeDB: the first manually collected, in-the-wild age database[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2017: 1997-2005.

This work is partially supported by the Open Project of the Fujian Provincial Key Laboratory of Data Intensive Computing (BD201805).

GONG Rui, born in 1995, M. S. candidate. His research interests include deep learning, computer vision.

DING Sheng, born in 1975, Ph. D., associate professor. His research interests include computer vision.

ZHANG Chaohua, born in 1996, M. S. candidate. His research interests include deep learning, computer vision.

SU Hao, born in 1996, M. S. candidate. His research interests include transfer learning, computer vision.