

基于轻量级卷积神经网络的人证比对

高凌飞¹, 王海龙², 王海涛¹, 刘 强³, 张鲁洋¹, 王怀斌¹

(1. 南京航空航天大学自动化学院, 南京 211106; 2. 江苏长城计算机系统有限公司, 南通 226001;
3. 江苏铭远轨道交通设备有限公司, 南京 210044)

摘要: 在证件审核场景中, 常规的深度学习人脸识别方法人证比对精度低且在嵌入式设备运行效率差。为解决上述问题, 本文提出了改进的轻量级卷积神经网络 Lightnet, 并采用了迁移学习方法。Lightnet 是结合深度可分离卷积、线性瓶颈结构和注意力模块构成的轻量级卷积神经网络模块, 引入附加角度裕量的损失函数 AM-Softmax 监督训练后, 网络模型能够保持较高的验证精度, 并有效解决标准卷积神经网络参数冗余、计算量大的问题。迁移学习通过冻结预训练模型的卷积层权重, 并在自制的人证数据集微调, 提高了网络模型的人证场景的识别性能。实验结果表明, 所设计的轻量级人证比对算法在验证精度、参数量以及运行效率等方面取得了很好的效果, 且对生活场景有较好的鲁棒性。

关键词: 人证比对; 轻量级卷积神经网络; 深度学习; 迁移学习

中图分类号: TP391.4

文献标志码: A

文章编号: 1005-2615(2021)05-0751-08

Scene-Identity Face Matching Based on Efficient Convolutional Neural Network

GAO Lingfei¹, WANG Hailong², WANG Haitao¹, LIU Qiang³, ZHANG Luyang¹, WANG Huaibin¹

(1. College of Automation, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China;

2. Jiangsu Great Wall Computer System Co. Ltd., Nantong 226001, China;

3. Jiangsu Mingyuan Rail Transit Co. Ltd., Nanjing 210044, China)

Abstract: In the scene of document verification, the standard deep learning face recognition method has low accuracy and poor real-time performance on embedded devices. To solve these problems, this paper proposes a modified efficient convolutional neural network (CNN) called Lightnet and adopts the transfer learning method. Lightnet is an efficient CNN module composed of depthwise separable convolution, linear bottleneck structure and attention module. After introducing the loss function AM-Softmax with additive angle margin, the network model can effectively solve the problems of redundancy parameter and vast calculation for standard CNN in the foundation of ensuring the high accuracy of face recognition. The transfer learning method can enhance the scene-identity face matching performance by freezing all the convolution layer weights of the pre-trained model and fine-tuning training in the self-made scene-identity face matching dataset. The experimental results show that the designed efficient scene-identity face matching algorithm has achieved good results in terms of verification accuracy, parameters and verification speed, and has good robustness in life scenarios.

Key words: scene-identity face matching; efficient convolutional neural network; deep learning; transfer learning

收稿日期: 2020-07-10; **修订日期:** 2021-03-24

通信作者: 王海涛, 男, 教授, 博士生导师, E-mail: htwang@nuaa.edu.cn。

引用格式: 高凌飞, 王海龙, 王海涛, 等. 基于轻量级卷积神经网络的人证比对[J]. 南京航空航天大学学报, 2021, 53(5): 751-758. GAO Lingfei, WANG Hailong, WANG Haitao, et al. Scene-identity face matching based on efficient convolutional neural network[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53(5): 751-758.

随着计算机和深度学习技术^[1-2]的快速发展,社会对智能化设备的要求不断提高。人脸识别技术作为一种高效的特征识别技术^[3]广泛地应用到各个领域中,如金融、安防^[4]、统计^[5]和反恐防暴等,受到国家高度重视。很多证件审核的场景对智能化的人证比对技术有着迫切的需求。

传统的人证信息核验通常采用人为比对的方式,这种比对方式消耗大量的时间、人力、物力,劳作重复冗余,效率低下^[6]。此外,人为比对的方式强烈依赖人的验证经验,无法对证件自身做出鉴别,错误率较高且存在一定的安全隐患。随着人脸识别技术的发展,人脸识别在人证场景得到了广泛的应用。通过智能算法的自动识别,大量减少了人的冗余劳作。然而,基于深度神经网络的人脸识别系统依赖复杂的网络模型,整个系统不得不基于个人计算机来进行开发,便携性较差。而嵌入式设备具有便携性、易用性且兼具一定的运算能力的优点,这使得发展嵌入式人脸识别技术成为一种时代趋势^[7]。

常规深度学习人脸识别系统的卷积神经网络模型基于高性能服务器在大规模公开人脸数据训练获得。此类算法具有网络层次深、泛化能力好、识别精度高的特点。但是针对嵌入式人证核验的场景,常规深度学习人脸识别方法存在以下问题:

(1)现有的深度卷积神经网络结构大多在网络深度和复杂度上进行研究,模型参数冗余且计算量大,无法适用于嵌入式平台。

(2)常规人脸识别受光照、姿态和表情等干扰,而人证比对是将相对约束的正面拍摄照片与严格约束的身份证照片相匹配。与常规的人脸识别相比,人证场景面临以下3方面的挑战:如老化带来的脸部变化、身份证照片压缩造成的信息丢失以及身份证数据集私有化。

为了解决上述问题,本文做了以下几个方面的工作:改进轻量级卷积神经网络 MobilenetV2 的网络结构,使模型更加轻量、更加精确;引入 AM-Softmax 人脸损失函数进行微调,使人脸特征类内紧凑、类间可分;自制人证数据集并采用小规模人证数据迁移学习方法,使模型更适用于人证比对场景。

1 原理和方法

本文研究的是基于轻量级卷积神经网络的人证比对算法。设计轻量级卷积神经网络使模型在精度和速度上取得平衡,引入 AM-Softmax 损失函数扩大类间边距,减小类内方差。为了使预训练模

型在人证比对领域适用,自建人证比对数据集并采用迁移学习方法。

1.1 深度可分离卷积

深度可分离卷积^[8]如图1所示,其核心是将标准卷积拆解为深度卷积和逐点卷积两个部分,图1中的 D_k 、 N 和 M 分别为卷积核的尺寸、卷积核的个数和卷积层通道数。深度卷积主要完成逐个通道的二维特征图的信息过滤。由于深度卷积只针对逐个通道进行卷积,缺乏对信息空间维度的利用,因此引入逐点卷积对深度卷积操作后的特征图在维度空间完成加权融合以及通道转换。

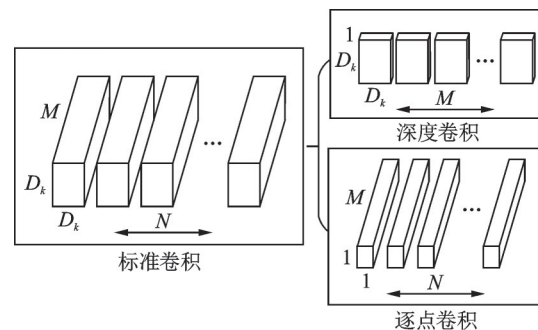


图1 深度可分离卷积

Fig.1 Depthwise separable convolution

图1中标准卷积的卷积核尺寸为 $D_k^2 \times M \times N$, 假设输入一个 $D_F^2 \times M$ 的特征图 F , 并得到一个 $D_G^2 \times N$ 的输出特征图 G 。其中 D_F 表示输入特征图的宽高, M 表示输入的通道数, D_G 表示输出特征图的宽高, N 表示输出的通道数。设步长为1, 标准卷积的参数数量为 $D_k^2 \times M \times N$, 计算量 R_1 为

$$R_1 = D_k^2 \times M \times N \times D_F^2 \quad (1)$$

深度可分离卷积运算由深度卷积和逐点卷积组成。深度卷积的卷积核尺寸为 $D_k^2 \times M$, 所以深度卷积参数数量为 $D_k^2 \times M$, 计算量为 $D_k^2 \times M \times D_F^2$ 。逐点卷积输入特征图尺寸为 $D_F^2 \times M$, 逐点卷积的卷积核尺寸为 $1 \times 1 \times M \times N$, 所以逐点卷积参数数量为 $1 \times 1 \times M \times N$, 计算量为 $1 \times 1 \times M \times N \times D_F^2$ 。因此深度可分离卷积的总参数数量为 $D_k^2 \times M + M \times N$, 计算量 R_2 为

$$R_2 = D_k^2 \times M \times D_F^2 + M \times N \times D_F^2 \quad (2)$$

深度可分离卷积的参数量下降倍数 B_1 、计算量下降倍数 B_2 分别为

$$B_1 = \frac{D_k^2 \times M + M \times N}{D_k^2 \times M \times N} = \frac{1}{N} + \frac{1}{D_k^2} \quad (3)$$

$$B_2 = \frac{D_k^2 \times M \times D_F^2 + M \times N \times D_F^2}{D_k^2 \times M \times N \times D_F^2} = \frac{1}{N} + \frac{1}{D_k^2} \quad (4)$$

由式(3,4)可以看出,相较于标准卷积,深度可

分离卷积大幅降低了参数量和计算量,提升了模型的运算效率。

1.2 人脸特征提取网络

随着人脸研究的不断深入,更多优化技术的出现,本文基于 MobileNetV2^[9]作以下3个方面的优化。改进后的网络命名为 Lightnet,其网络结构如表1所示,表1中 Input 表示输入特征图的尺寸;Operator 表示对特征图展开的操作,其中 Conv2d 表示二维卷积,DW3×3 为深度卷积,Bottleneck 为瓶颈结构,AMBottleneck 为融入注意力模块的瓶颈结构,GDC 为全局深度卷积; t 表示扩张,即瓶颈结构的扩张因子; c 表示输出特征图通道的数目; n 表示重复,即当前层的重复次数; s 表示卷积层的步长。

表1 Lightnet网络结构

Table 1 Lightnet network structure

Input	Operator	t	c	n	s
$112^2 \times 3$	Conv2d		32	1	2
$56^2 \times 32$	DW3×3		32	1	1
$56^2 \times 32$	Bottleneck	2	64	1	2
$28^2 \times 64$	Bottleneck	2	64	4	1
$28^2 \times 64$	Bottleneck	4	128	1	2
$14^2 \times 128$	AMBottleneck	2	128	6	1
$14^2 \times 128$	Bottleneck	4	128	1	2
$7^2 \times 128$	Bottleneck	2	128	2	1
$7^2 \times 128$	Conv2d 1×1		512	1	1
$7^2 \times 512$	GDC7×7		512	1	1
$1^2 \times 512$	FC(Conv2d 1×1)		128	1	1

(1)调整结构尺寸。人证场景中身份证机芯证件照分辨率为102像素×126像素,照片信息不具备 $224 \times 224 \times 3$ 的信息量。且由式(1,2)可知,卷积神经网络的计算量与输入特征图有关,若输入的宽和高减半,计算量将减少4倍,为此本文将输入调整为 $112 \times 112 \times 3$ 。此外,MobileNetV2中Bottleneck的扩张因子 t 为定值6,造成了较多的冗余参数和计算,本文根据层间需求动态调整扩张因子 t 。

(2)避免全局平均池化。MobileNetV2在池化合并特征之前,Bottleneck 输出了一个7像素×7像素特征图。虽然每个特征点具有相同大小的感受野^[10],但中心特征点的感受野包括了完整的图片信息,而边角特征点的感受野只有部分信息。MobileNetV2采用全局平均池化将权重平均,网络的表征能力自然会下降。因此本文引入全局深度卷积代替全局平均池化,让网络自己学习权重,模型泛化能力更强。

(3)融入注意力模块。本文在 MobileNetV2 主特征提取模块 Bottleneck 的基础上,将注意力模块

融入其中并命名为 AMBottleneck,具体结构如图2所示。

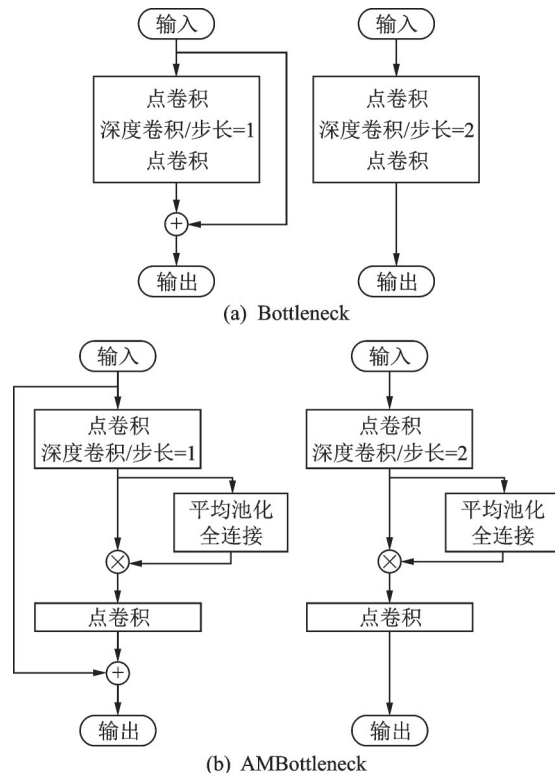


图2 Bottleneck和AMBottleneck模块结构

Fig.2 Module structures of Bottleneck and AMBottleneck

注意力模块就是通过卷积训练获取各个特征通道的重要程度。在模型训练完毕部署时候,模型就可以根据训练数据保存的权重信息自动去强化有用的特征,抑制一些不必要的干扰。注意力模块如图3所示,图中输入的是由一系列卷积后获得的特征图 $F_{\text{map}} = [u_1, u_2, \dots, u_c]$,尺寸为 $D_k \times D_k \times C$, u_c 表示特征图 F_{map} 中的第 C 个二维矩阵, C 表示通道数。

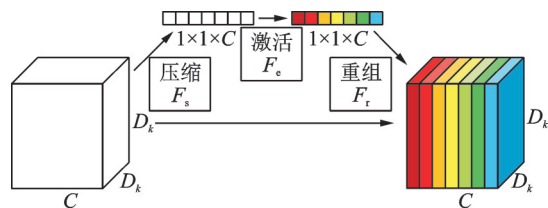


图3 注意力模块

Fig.3 Attention module

注意力模块分为3步:压缩、激活和重组。

第1步是压缩操作,将特征图的全局空间信息沿着特征通道的方向压缩为一个通道实数点,该实数拥有全局特征通道的感受野,并且输出的维度和输入的特征通道数相匹配。形式上可通过使用全局池化层来实现统计数据,由式(5)表示。

$$F_s = \frac{1}{D_k \times D_k} \sum_{i=1}^{D_k} \sum_{j=1}^{D_k} u_c(i, j) \quad (5)$$

第2步是激活操作,为了利用压缩操作捕获的特征通道方面的全局描述特征,需要另外一种运算来抓取通道之间的关系。通过设置全连接层,经过网络训练后学习到权重 W ,再利用带有门控机制的 Sigmoid 激活函数获得激活值(0~1),也就是每个特征通道的权重 s 。激活操作可由式(6)表示

$$s = F_c(F_s, W) = \sigma(g(F_s, W)) \quad (6)$$

第3步是重组操作,根据激励操作的输出结果对输入的特征图按通道进行乘积操作,即完成特征图权重加权处理,重新设定特征图的每个通道的重要程度。重组操作可以由式(7)表示。

$$F_r = u_k \cdot s \quad (7)$$

1.3 AM-Softmax 损失函数

Softmax 损失函数在多分类的场景中使用广泛,在模型完成特征的提取后,Softmax 通过映射这些特征划归各个类的概率实现类别的区分。Softmax 损失函数如式(8)所示。

$$L_s = -\frac{1}{n} \sum_{i=1}^n \log \frac{e^{\mathbf{w}_{y_i}^T f_i}}{\sum_{j=1}^c e^{\mathbf{w}_j^T f_i}} = -\frac{1}{n} \sum_{i=1}^n \log \frac{e^{\|\mathbf{w}_{y_i}\| \|f_i\| \cos(\theta_{y_i})}}{\sum_{j=1}^c e^{\|\mathbf{w}_j\| \|f_i\| \cos(\theta_j)}} \quad (8)$$

式中: n 表示样本数量; i 表示第 i 个样本; j 表示第 j 个类别; $\mathbf{w}_{y_i}^T f_i$ 表示全连接层输出; 其中 $\mathbf{w}_{y_i}^T$ 表示第 i 个样本的权重; f_i 第 i 个样本的特征; y_i 表示第 i 个样本的真实标签。通过式(8)可以发现,提高 $\mathbf{w}_{y_i}^T f_i$ 的具体数值,整体 L_s 就会得以提升,从而获得对第 i 个样本更好的分类性能。由式(8)可知 Softmax 强调了分类的准确率,在人证核验这样的多分类场景中缺乏约束边界的能力,导致了模型验证精度不高。随着度量学习和分类损失的研究^[11-14],研究者由式 $\mathbf{w}^T f_i = \|\mathbf{w}\| \cdot \|f_i\| \cdot \cos \theta$ 发现全连接的输出信息中包含了角度 θ ,由此展开了更多的在角度空间的研究,发现通过在角度空间引入一个最大化分类边距 $\text{Margin}(m)$ 来控制类别间的间隔,可以使模型获得更加具有分辨力的特征值。最终 AM-Softmax 的表达式如式(9)所示。

$$L_{\text{AMS}} = -\frac{1}{n} \sum_{i=1}^n \log \frac{e^{s(e^{\mathbf{w}_{y_i}^T f_i} - m)}}{e^{s(e^{\mathbf{w}_{y_i}^T f_i} - m)} + \sum_{j=1, j \neq y_i}^c e^{s e^{\mathbf{w}_j^T f_i}}} \quad (9)$$

1.4 人证数据集制作

由于身份证信息的隐私性,较难有开源的人证

数据可供训练,为此本文自制一个近似的人证数据集。数据集由 CASIA-FaceV5、Public-IvS^[15] 和个人收集的人证数据构成。CASIA-FaceV5 由中科院提供,数据由中国人构成,拍摄场景为蓝底证件背景,与本文的人证场景相似,数据样张如图4所示。Public-IvS 是由 Zhu 等收集的证件照数据集^[15],每个类别包含一张证件照和多张生活照片,证件照和生活照存在时间差,有衰老带来的人脸变化,数据样张如图5所示。人证数据集制作方法如图6所示(本人的身份证照片仅作为制作方法演示),主要包括以下3步:



图4 CASIA-FaceV5数据集样张

Fig.4 Sample of CASIA-FaceV5 dataset



图5 Public-IvS数据集样张

Fig.5 Sample of Public-IvS dataset

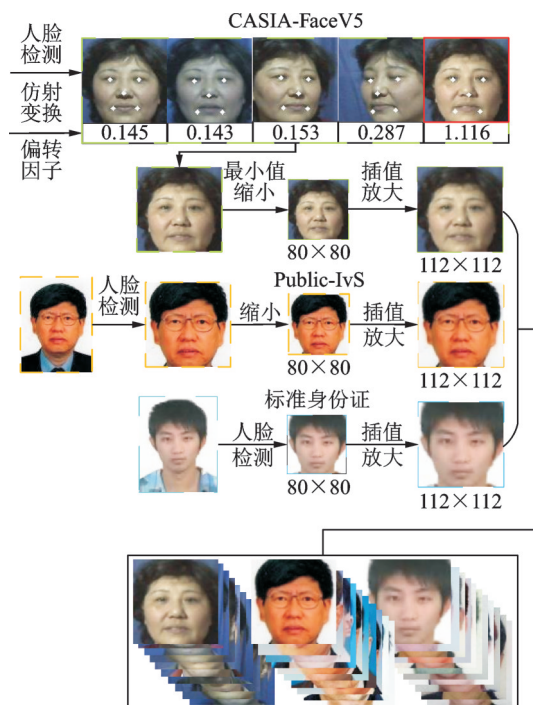


图6 人证数据集制作

Fig.6 Production of the FBVID dataset

(1) 数据清洗和对齐。将样本输入到MTCNN进行人脸检测、对齐并统一裁剪为112像素×112像素,同时人为清洗数据的噪声数据。

(2) 设定证件标签。将CASIA-FaceV5的同组图像送入人脸角度判别算法,设定位姿端正图片作为证件照。人脸角度判别算法如图4所示,根据人脸检测器MTCNN提供的各姿态人脸的关键点:左眼坐标 (x_1, y_1) 、右眼坐标 (x_2, y_2) 、鼻尖坐标 (x_3, y_3) 、左嘴角坐标 (x_4, y_4) 计算如式(10)所示的人脸偏转因子 P_{score} ,人脸偏转越大, P_{score} 值越大。最终选出 P_{score} 值最小的图片作为该组ID的证件照。

$$P_{score} = \sqrt{\frac{(2x_3 - x_1 - x_2)^2}{4(x_2 - x_1)^2} + \left(\frac{y_3 - y_1}{y_4 - y_1} - 0.6\right)^2} \quad (10)$$

(3) 压缩。真实场景的身份证照片人脸区域为80像素×80像素,而CASIA-FaceV5、Public-IvS中的证件照图像分辨率较高,包含的人脸信息丰富。因此本文将CASIA-FaceV5和Public-IvS中的证件照人脸图像缩放至80像素×80像素,再差值放大到112像素×112像素,形成一个和真实身份证人脸相似的模糊人脸数据。

经过数据清洗和增广后,本文自制的人证数据集约1900个ID、1.6万张图片。

1.5 小规模人证数据迁移学习

深度学习实际应用基于一个非常重要的前提,即模型的训练数据和模型部署的应用数据来自相同的特征空间以及具有相同的分布。当特征空间或分布规律无法满足时,大多数的深度学习模型需要大量分布改变后的目标域数据重新训练。然而在很多情况下,重新收集训练数据成本很高或难以实现。

以人证比对任务为例,通用人脸识别模型由无约束的公开人脸数据集训练获得,且数据大多为非东方人数据。因此通用人脸识别模型在人证场景不完全适用,识别精度达不到安全要求。但是自制的人证数据集样本数据有限,在此小规模数据上直接训练易出现过拟合。虽然通用人脸数据集和人证数据集存在差异,但是人脸整体上结构一致,数据属性皆相同。且网络中较多网络层都是提取人脸的通用特征,对于识别起到基础的认知作用。因此可迁移该部分参数作为公共知识帮助求解人证比对的训练问题,网络仅需要少量数据更新顶层的网络层权重即可获得人证场景的分类性能。具体做法如图7所示,主要包括以下3个步骤:

(1) 利用Softmax和AM-Softmax损失函数在

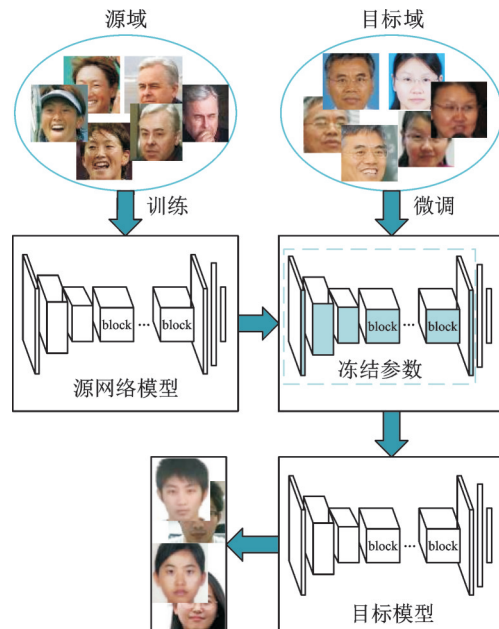


图7 迁移学习流程

Fig.7 Transfer learning process

源域(大规模公开数据集)中训练和微调,得到预训练的人脸分类模型。

(2) 利用AM-Softmax损失函数,在第1步中获得预训练的模型进行目标域(小规模人证数据集)迁移训练,得到迁移学习后的网络模型。迁移学习通过冻结最后一层全连接层前所有网络层的权重,只训练全连接层和分类层。如此可以保留预训练模型的基础权重,减少学习的参数,使得网络训练能够快速收敛并在小样本的人证数据集上拟合,获得人证场景的分类性能。

(3) 通过第2步迁移训练后得到的模型提取人证数据特征,进行人证比对。

2 实验与分析

本文基于深度学习框架MXNet进行模型训练和测试,采用Python编程语言实现。计算机硬件配置CPU为Intel酷睿I7 7800X,主频3.5 GHz,运行内存为32 GB,系统环境为Ubuntu 16.04位操作系统,GPU型号为Nvidia GeForce GTX 1080Ti,显存为11 GB。

2.1 数据集

采用MS-Celeb-1M作为训练数据,数据集包括99 892个ID,8 456 240张图片。由于样本分布不均且包含部分噪声信息,为此本文保留类别数目大于30的数据,剔除不能有效检测人脸或包含2个及多个人脸的图片,对截取的人脸进行仿射变换对齐并统一调整为112像素×112像素。经数据处理后,训练集包括约8.8万个ID,580万张人脸图像。

测试验证采用LFW, AgeDB和NJU-ID对各模型进行评估。LFW(Labeled faces in the wild)包含5 749个ID, 13 233张图片, 数据取自非限制场景, 具有多姿态、光照、表情和遮挡等干扰因素。本文从LFW取3 000对正样本、3 000对负样本形成测试集。AgeDB(Age database)包含568个ID, 16 488张图片, 数据呈现一定的年龄分布, 本文从AgeDB取3 000对正样本、3 000对负样本形成测试集。NJU-ID^[16-18](Nanjing University ID Card Face)是南京大学为研究身份证人脸验证而开发的数据集, 包含256个ID, 每个ID包含一张身份证照片和一张生活照片。身份证照片是IC读卡器读取的二代身份证芯片中存储的低分辨率图像, 生活照是在不同的光照、不同的背景下随意拍摄的高分辨率图像。另外, 两幅匹配图像的拍摄时间间隔也是任意的, 这使得人证比对具有较大的挑战性。本文从NJU-ID取256对正样本、256对负样本形成测试集。

2.2 源域模型训练结果

源域模型训练分2步。第1步基于Softmax损失函数训练3个基线模型, BM1(Base model):使用全局平均池化; BM2:使用全局深度卷积; BM3:使用全局深度卷积和注意力模块。基线模型均从零训练, 权重刷设置为0.000 04; batch设置为256; 使用动量为0.9的随机梯度下降(Stochastic gradient descent, SGD)优化模型。初始学习率设置为0.1, 然后学习率在12万次、14万次迭代后下降10倍, 共训练16万次迭代。第2步基于AM-Softmax损失函数对基线模型BM3进行微调训练, 初始学习率设置为0.05, 然后学习率在12万次、14万次迭代后下降10倍, 共训练20万次迭代。测试阶段比较两张人脸图像时采用特征值的欧式距离作为度量, 各模型验证结果如表2所示。

表2 不同源域模型的识别率

Table 2 Recognition rate of different source domain models

模型	LFW/%	AgeDB/%	NJU-ID/%
BM1	98.08	89.68	90.32
BM2	98.52	90.12	90.78
BM3	99.06	90.70	91.43
AM-Softmax	99.48	95.05	95.60

由表2数值结果可知, 加入了全局深度卷积的BM2相对于基线模型BM1在3个验证集分别提升了0.48%、0.44%和0.36%, 验证了避免全局平均池化的有效性。模型BM3相对于模型BM2在3个

验证集分别提升了0.46%、0.58%和0.65%, 验证了融入注意力模块的有效性。经过AM-Softmax损失函数微调后, 由于最大化分类边距的引入, 微调模型相对基线模型BM3在3个验证集分别提升了0.42%、4.35%和4.17%, 证明了改进损失函数的有效性。此外, 为了可视化AM-Softmax损失函数使类间距离增大、类内方差缩小的效果, 采用微调模型和基线模型BM3分别提取LFW, AgeDB-30和NJU-ID正、负样本对的特征, 计算欧式距离并作直方图, 各数据集样本分布如图8所示。从图中可以看出, 相较于Softmax训练的基线模型, AM-Softmax微调模型有效减少了样本间的重叠部分, 减少了人脸错分的风险, 增加了人脸识别模型的相似度阈值选取的鲁棒性。

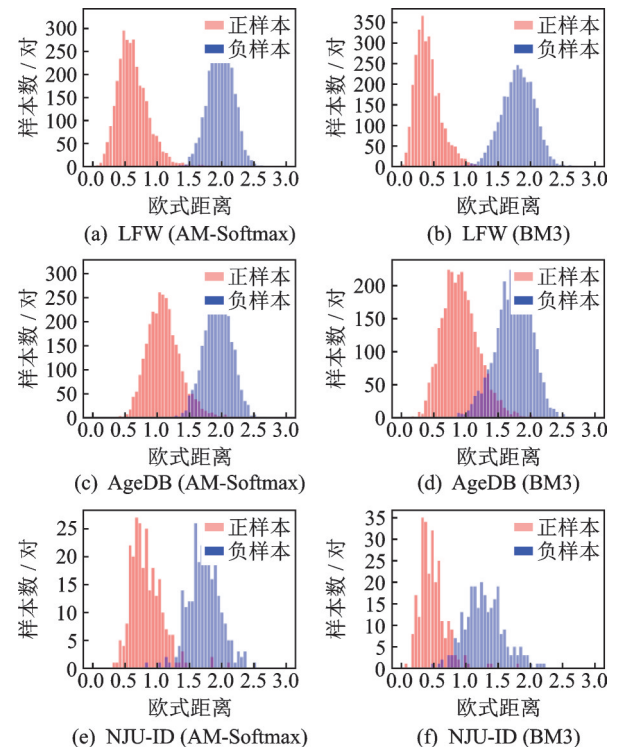


图8 各数据集正负样本分布

Fig.8 Distribution of positive and negative samples in each dataset

为客观了解本文轻量级网络微调训练后模型的识别性能和运行效率, 实验验证了最先进的大型网络ResNet50和其他轻量级网络的识别效果, 计算各网络的参数量, 并基于RK3399(4×A53+2×A72)移动端Android7.1系统在NCNN前向推理加速框架统计各网络的运行效率, 结果如表3所示。

由表3数值结果可知, ResNet50虽然在验证集中取得了最高的验证精度, 但由于其较大的复杂度, 其在移动端推理速度为647.68 ms, 无法满足移动端实时性的需求。由表中本文网络和其他轻量

表 3 轻量级神经网络模型性能对比

Table 3 Performance comparison of diffident efficient CNN models					
网络名称	LFW/ %	Age- DB/%	NJU- ID/%	参数 量/10 ⁶	速度/ ms
ResNet50	99.80	97.70	97.07	25	647.68
MobileNetV1	98.63	88.95	89.92	4.2	113.41
MobileNetV2	99.16	90.67	90.79	3.4	98.29
ShuffleNetV1	98.70	88.81	89.86	1.9	65.22
ShuffleNetV2	99.12	90.54	90.98	2.3	72.98
Lightnet	99.48	95.05	95.60	1.1	42.48

级网络的性能对比可知^[19-20],得益于有效的改进方法,本文的网络模型在相对于表中的其他轻量级网络实现了更好的验证精度和推理速度,证明了本文网络是一种高效的特征提取网络。

2.3 目标域迁移训练结果

源域模型在人证数据集 NJU-ID 上取得了 95.60% 的验证精度,但人证核验往往基于严格的正式场景,需对持证人员精确识别。为此本文利用自建人证数据进行迁移训练。为了验证所采用的迁移学习方法的有效性,本文加入另外两种训练方式:(1)不使用源域模型,直接在人证数据重新训练;(2)不冻结权重,所有层微调训练。各模型验证结果如表 4 所示。

表 4 不同训练方式的模型性能对比

Table 4 Performance comparison of models based on diffident training methods			
训练方式	LFW/%	AgeDB/%	NJU-ID/%
原模型	99.48	95.05	95.60
重新训练	77.13	57.42	65.87
所有层微调	98.06	90.52	94.43
迁移学习	99.04	93.50	97.66

由表 4 数值结果可知,本文采用的迁移学习方法在牺牲了少量通用数据集的识别性能,在目标数据集 NJU-ID 的精度提升了 2.16%,最终在 NJU-ID 的准确率达到 97.66%,对人证场景具有较好的分类性能。验证了常规人脸识别和人证比对是性质不同的两个识别任务,证明了本文提出的迁移学习方法在人证目标域迁移的有效性。而另外两种训练方式表现很差或者出现全面性能下降。这是由于源域模型是在大规模数据集上训练所得,底层网络对人脸特征的抽象和提取已经具备优良的性能,整体模型对人脸识别任务有较强的泛化能力。而重新训练或所有层微调的训练方式改变了这些权重参数,导致模型性能失活或者下降。

2.4 生活场景的人证比对

人证比对是将严格约束的身份证照片和相对约束的正面照片相匹配。尽管证件照是严格约束的条件,但相对约束的人脸正面照还是可能存在轻微的倾斜、光照、发型变化等情况。为测试所提出的人证比对算法的泛用性,本文收集复杂的生活场景数据,照片取自课题组 24 位同学,每个 ID 包含 1 张身份证照片 20 张生活照。数据共组成 480 对正样本和 480 对负样本,识别结果如表 5 所示。

表 5 生活场景的识别效果

Table 5 Performance of real scene					
识别总数	阈值	正确 识别数	错误 拒绝数	错误 接受数	识别率/%
960	1.51	919	10	31	95.73

由表 5 数值结果可知,本文的网络模型在复杂的生活场景的人证识别率达到了 95.73%,具备一定的泛用性,能够满足日常生活的验证需求。

3 结 论

本文从嵌入式的人证比对场景出发,提出一种基于轻量级卷积神经网络的人证比对算法 Lightnet。(1)Lightnet 基于 MobilenetV2 调整网络结构,引入注意力机制和 AM-Softmax 损失函数,在保持较低的复杂度和较快的运行效率情况下取得了较高的识别精度。(2)自制人证数据集并进行迁移学习,通过人证数据的微调进一步增强了模型人证场景下的验证精度,在标准人证数据集 NJU-ID 的准确率达到 97.66%,对人证场景具有了较好的分类性能。(3)复杂生活场景下人证比对验证取得了 95.73% 的验证精度,证明了本文网络模型具备一定的泛用性,能够满足日常生活的验证需求。

参考文献:

[1] 余永红,殷凯宇,王强,等. 基于深度学习的欧几里得嵌入的推荐算法[J]. 南京航空航天大学学报, 2020, 52(5):729-735.
YU Yonghong, YIN Kaiyu, WANG Qang, et al. Recommendation algorithm of euclidean embedding based on deep learning[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(5): 729-735.

[2] 谢润山,王士同. 基于函数链神经网络的深度分类器[J]. 南京航空航天大学学报, 2020, 52(5):736-745.
XIE Runshan, WANG Shitong. Functional-link neural network based deep classifier[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(5):736-745.

- [3] 刘铎. 基于CUDA的实时人脸识别系统[J]. 计算机与现代化, 2015(6):122-126.
LIU Yi. CUDA-based real-time face recognition system [J]. Computer and Modernization, 2015(6): 122-126.
- [4] 李东锋. 人工智能技术在建筑施工现场安全管理中的应用探索[J]. 住宅与房地产, 2020(12):140-141.
LI Dongfeng. Application and exploration of artificial intelligence technology in safety management of construction site[J]. Housing and Real Estate, 2020(12): 140-141.
- [5] 吴凡. 智慧图书馆人脸识别技术应用研究[J]. 图书馆学刊, 2019(7):100-104.
WU Fan. Research on the application of face recognition technology in intelligent library [J]. Journal of Library Science, 2019(7):100-104.
- [6] 戴健. 证件照人脸识别的算法研究[D]. 南京:东南大学, 2016.
DAI Jian. Research on algorithm of face recognition based on ID photo[D]. Nanjing: Southeast University, 2016.
- [7] 刘程, 谭晓阳. 一种基于深度学习的移动端人脸验证系统[J]. 计算机与现代化, 2018(2):107-111.
LIU Cheng, TAN Xiaoyang. A face verification system on mobile terminal based on depth learning [J]. Computer and Modernization, 2018(2):107-111.
- [8] HOWARD A, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17) [2020-12-04]. <https://arxiv.org/abs/1704.04861>.
- [9] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [EB/OL]. (2018-01-13) [2020-12-04]. <https://arxiv.org/abs/1801.04381>.
- [10] LUO W, LI Y, URTASUN R, et al. Understanding the effective receptive field in deep convolutional neural networks [EB/OL]. (2017-01-15) [2020-12-04]. <https://arxiv.org/abs/1701.04128>.
- [11] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10000 classes [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, D. C., USA: IEEE, 2014:1891-1898.
- [12] 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 (CVPR). Washington, D. C., USA: IEEE, 2015: 815-823.
- [13] 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 (CVPR). Honolulu, USA: IEEE, 2017:200-205.
- [14] WANG F, CHENG J, LIU W, et al. Additive Margin Softmax for Face Verification [J]. IEEE Signal Processing Letters, 2018, 25(7):926-930.
- [15] ZHU X, LIU H, LEI Z, et al. Large-scale bisample learning on ID vs. spot face recognition [EB/OL]. (2018-01-13) [2020-12-04]. <https://arxiv.org/abs/1801.04381>.
- [16] HUO J, GAO Y, SHI Y, et al. Ensemble of sparse cross modal metrics for heterogeneous face recognition [C] // Proceedings of the 2016 ACM on Multimedia Conference. USA: ACM, 2016: 1405-1414.
- [17] HUO J, GAO Y, SHI Y, et al. Heterogeneous face recognition by margin-based cross-modality metric learning [J]. IEEE Transactions on Cybernetics, 2018, 48(6):1814-1826.
- [18] HUO J, GAO Y, SHI Y, et al. Cross-modal metric learning for AUC optimization [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(10):1-13.
- [19] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices [EB/OL]. (2017-07-04) [2020-12-04]. <https://arxiv.org/abs/1801.04381>.
- [20] MA N, ZHANG X, ZHENG H, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design [C] // Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: IEEE, 2018: 116-131.

(编辑:张蓓)