

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0049

# 基于 MobileFaceNet 网络改进的人脸识别方法

张子昊<sup>1,2</sup>, 王蓉<sup>1,2,\*</sup>

(1. 中国人民公安大学 警务信息工程与网络安全学院, 北京 100038;

2. 安全防范技术与风险评估公安部重点实验室, 北京 100038)

**摘 要:** 为了解决训练过程中卷积模型参数较多、收敛速度较慢的问题,提出了一种基于 MobileFaceNet 网络改进的人脸识别方法。首先,使用 MobileFaceNet 网络提取人脸特征,在提取特征的过程中,通过引入可分离卷积减少模型中卷积层参数的数量;其次,通过在 MobileFaceNet 网络中引入风格注意力机制来增强特征的表达,同时使用 AdaCos 人脸损失函数来训练模型,利用 AdaCos 损失函数中的自适应缩放系数来动态地调整超参数,避免了人为设置超参数对模型的影响;最后,分别在 LFW、AgeDB 和 CFP-FF 测试数据集上对训练模型进行评估。实验结果显示:改进后的模型在 LFW、AgeDB 和 CFP-FF 测试数据集上的识别精度分别提升了 0.25%、0.16% 和 0.3%,表明改进后的模型相较于改进前的模型在精度和鲁棒性上有所提高。

**关 键 词:** 人脸识别; 深度学习; MobileFaceNet; AdaCos; 卷积神经网络

**中图分类号:** O235; TP183

**文献标志码:** A

**文章编号:** 1001-5965(2020)09-1756-07

人脸特征具有易获取、易捕捉、易处理以及非接触式等特性,已经受到学者们的广泛关注,并在公共安全等领域应用日趋广泛。但是,在实际的应用场景下,人脸识别精度依然受到不同的姿态、光照、遮挡等因素的影响。因此,如何提取更具鲁棒性的特征来有效地辨识人脸便成为解决问题的关键。传统的人脸识别方法主要有主成分分析法(PCA)<sup>[1]</sup>以及局部二值模式(LBP)<sup>[2]</sup>。主成分分析法就是将高维人脸信息通过正交变化投影到低维子空间中,形成特征脸,然后通过分类器对低维的特征脸进行分类;局部二值法是将检测窗口划分为许多小区域,对每一个像素将其与周围相邻像素比较,并重新赋值,通过计算每个区域的直方图,并使用机器学习的方法对其进行分类。近年来,随着深度学习应用的日趋广泛,使用深度学

习的方法来进行人脸识别成为计算机视觉领域的热门方向。

与传统方法相比,深度学习的方法通过海量的数据来训练模型,使模型提取到的人脸特征更具有泛化性。最具代表性的是香港中文大学汤晓鸥团队提出的 DeepID<sup>[3]</sup>系列。该团队中的 Sun 等提出 DeepID1<sup>[4]</sup>通过使用单一的卷积网络来进行特征提取并进行分类识别,之后,又在文献[4]的基础上提出了 DeepID2<sup>[5]</sup>模型,通过将人脸认证信号和人脸验证信号引入网络,以达到增大类间距缩小类内距的要求。随后,为了进一步提高模型提取特征的代表性和鲁棒性,Google 提出了 FaceNet<sup>[6]</sup>模型,将模型映射到欧几里德空间,并使用三元组损失 Triplet loss<sup>[7]</sup>来增大类间距,从而提高模型性能;为了解决不同特征向量大小对

收稿日期: 2020-02-25; 录用日期: 2020-04-18; 网络出版时间: 2020-04-21 11:39

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200421.0911.001.html

基金项目: 公安部技术研究计划项目(2017JSYJB01); 中央高校基本科研业务费专项资金(2019JKF111)

\* 通信作者: E-mail: dbdxwangrong@163.com

引用格式: 张子昊,王蓉. 基于 MobileFaceNet 网络改进的人脸识别方法[J]. 北京航空航天大学学报,2020,46(9): 1756-1762.  
ZHANG Z H, WANG R. Improved face recognition method based on MobileFaceNet network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(9): 1756-1762 (in Chinese).

分类效果的影响,Deng 等提出 Arcface<sup>[8]</sup> 损失,将特征向量归一化到超球面上,使模型在角度空间对分类的边界进行最大化;为了解决训练过程中超参数的选择问题,Zhang 等提出 AdaCos<sup>[9]</sup> 人脸损失函数,引入自适应动态缩放参数来解决超参数的选择问题。但是以上介绍的方法中主干网络大多比较复杂,模型参数较多,增加了计算量,影响计算效率以及模型的收敛速度。

本文在 MobileFaceNet<sup>[10]</sup> 网络的基础上加入风格注意力机制,将改进后的模型作为特征提取网络,使提取到的特征更具代表性和鲁棒性。同时引入 AdaCos 人脸损失函数来对模型进行训练,通过动态自适应缩放系数的调节优化训练,减少了人为调节超参数对训练的影响,提高了训练效率。

## 1 基于 MobileFaceNet 网络的人脸识别方法

### 1.1 MobileNetV1 和 MobileNetV2 网络

MobileNetV1<sup>[11]</sup> 是 Google 发布的网络架构,提出深度可分离卷积的概念。其核心就是把卷积拆分为 Depthwise + Pointwise 两部分。其中 Depthwise 卷积是指不跨通道的卷积,即 Feature Map 的每个通道有一个独立的卷积核,并且这个卷积核作用且仅作用在这个通道之上,输出 Feature Map 的通道数等于输入 Feature Map 的通道数,因此它并没有升维或者降维的功能。Pointwise 卷积用于特征合并以及升维或者降维,一般使用  $1 \times 1$  的卷积来实现功能。由此,可以将一个普通的卷积拆分成如图 1 所示的结构,图中: $N$ 、 $M$  和  $D_k$  分别为卷积的个数、卷积通道数和卷积的宽、高。

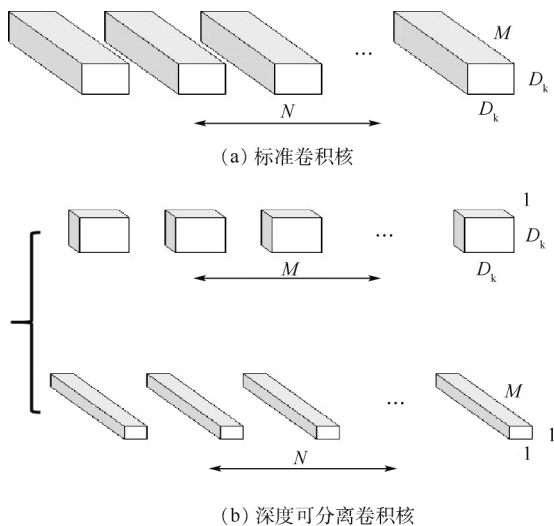


图1 可分离卷积示意图

Fig. 1 Schematic diagram of detachable convolution

标准卷积的参数数量是  $D_k \times D_k \times M \times N$ ; 卷积核的尺寸是  $D_k \times D_k \times M$ , 一共有  $N$  个, 每一个都要进行  $D_w \times D_h$  次运算, 所以标准卷积的计算量  $P_1$  为

$$P_1 = D_k \times D_k \times M \times N \times D_w \times D_h \quad (1)$$

式中:  $D_w$  和  $D_h$  分别为标准卷积的宽和高。

深度可分离卷积的参数数量是:  $D_k \times D_k \times M + 1 \times 1 \times M \times N$ ; 深度可分离卷积的计算量是由深度卷积和逐点卷积两部分组成: 深度卷积的卷积核尺寸  $D_k \times D_k \times M$ , 一共要做  $D_w \times D_h$  次乘加运算; 逐点卷积的卷积核尺寸为  $1 \times 1 \times M$ , 有  $N$  个, 一共要做  $D_w \times D_h$  次乘加运算, 所以深度可分离卷积的计算量  $P_2$  为

$$P_2 = D_k \times D_k \times M \times D_w \times D_h + M \times N \times D_w \times D_h \quad (2)$$

参数数量和乘加操作的运算量下降倍数  $P_3$  为

$$P_3 = \frac{P_2}{P_1} = \frac{D_k \times D_k \times M \times D_w \times D_h + M \times N \times D_w \times D_h}{D_k \times D_k \times M \times N \times D_w \times D_h} = \frac{1}{N} + \frac{1}{D_k^2} \quad (3)$$

从式(3)可以看出,深度可分离卷积在参数数量以及计算量上与标准卷积相比,大幅度的降低,提升了模型的运算效率。

深度卷积本身没有改变通道的能力,即输入通道等于输出通道若是输入通道数较少,则深度卷积只能在低维度工作,最终卷积效并不会很好。所以 MobileNetV2<sup>[12]</sup> 在 MobileNetV1 的深度卷积层之前,使用  $1 \times 1$  的卷积进行升维操作,使网络在一个更高维的空间进行特征提取,这就使得提取到的特征更具全局性。同时,为了解决低维度 Relu 运算产生的信息丢失问题,将最后一个 Relu 替换成 Linear 线性激活函数。MobileNetV2 网络结构如表 1 所示,表中:  $z$  为通道扩张倍数;  $c$  为输出通道数;  $n$  为重复次数;  $v$  为步长;  $f$  为输入通道数。

表1 MobileNetV2 网络结构

输入	卷积操作	$z$	$c$	$n$	$v$
$224 \times 224 \times 3$	Conv2d		32	1	2
$112 \times 112 \times 32$	bottleneck	1	16	1	1
$112 \times 112 \times 16$	bottleneck	6	24	2	2
$56 \times 56 \times 24$	bottleneck	6	32	3	2
$28 \times 28 \times 32$	bottleneck	6	64	4	2
$28 \times 28 \times 64$	bottleneck	6	96	3	1
$14 \times 14 \times 96$	bottleneck	6	160	3	2
$7 \times 7 \times 160$	bottleneck	6	320	1	1
$7 \times 7 \times 320$	Conv2d $1 \times 1$		1280	1	1
$7 \times 7 \times 1280$	Avgpool $7 \times 7$			1	
$1 \times 1 \times f$	Conv2d $1 \times 1$		$f$		

## 1.2 MobileFaceNet 人脸特征提取网络

MobileFaceNet 是 MobileNetV2 的改进版本。由于 MobileNetV2 中使用的是平均池化层,但是针对同一张图片,不同像素点的权重是不同的,全局的平均池化将权重平均,网络的表现能力自然会下降。所以在 MobileFaceNet 中,使用一个  $7 \times 7 \times 512$  的可分离卷积代替原本的全局平均池化层。同时使用 Prelu 代替 Relu 激活函数,并引入归一化层来加快模型收敛,防止模型过拟合。MobileFaceNet 的网络结构如表 2 所示。

表 2 MobileFaceNet 网络结构

Table 2 MobileFaceNet network structure

输入	卷积操作	$z$	$c$	$n$	$v$
$112 \times 112 \times 3$	Conv $3 \times 3$		64	1	2
$56 \times 56 \times 64$	Depthwise Conv $3 \times 3$		64	1	1
$56 \times 56 \times 64$	bottleneck	2	64	5	2
$28 \times 28 \times 64$	bottleneck	4	128	1	2
$14 \times 14 \times 128$	bottleneck	2	128	6	1
$14 \times 14 \times 128$	bottleneck	4	128	1	2
$7 \times 7 \times 128$	bottleneck	2	128	2	1
$7 \times 7 \times 128$	Conv $1 \times 1$		512	1	1
$7 \times 7 \times 512$	Linear GDCov $7 \times 7$		512	1	1
$1 \times 1 \times 512$	Linear Conv $1 \times 1$		128	1	1

网络结构不仅保留了 MobileNetV2 的升降维层以及线性激活函数,同时还在全连接层之前引入  $7 \times 7$  的可分离卷积来替代原本的平均池化层,使网络提取到的特征更具泛化性和全局性。

## 1.3 AdaCos 人脸损失函数

Softmax 目前被广泛应用于图片分类中,主要通过输出样本属于各个类的概率大小来区分不同类别之间的特征。但是 Softmax 并没有对类内和类间距离进行约束,这就大大影响了人脸识别的精度。最新的人脸损失函数 Arcface 以及 AdaCos 就是基于 Softmax 改进的。

### 1.3.1 Arcface 损失函数

Softmax 函数为

$$L_s = -\frac{1}{m} \sum_{i=1}^m \ln \left( \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} \right) \quad (4)$$

式中:  $m$  为每一个训练批次的大小;  $i$  为图像编号;  $j$  为当前样本的真实标签;  $W_{y_i}^T x_i + b_{y_i}$  表示全连接层输出,  $W_{y_i}^T x_i$ 、 $b_{y_i}$  和  $y_i$  分别为第  $i$  个图像类别的权重、图像的特征、图像类别的偏差和图像的真实类别的标签。通过提高  $W_{y_i}^T x_i + b_{y_i}$  所占有的比重,提高对该类别的分类效果。但是 Softmax 只考虑到样本的分类正确性,而对于人脸识别这样的多分类问题,缺乏类内和类间距离约束。又因

为  $W^T x = \|W\| \cdot \|x\| \cdot \cos \theta$  为特征向量  $x$  和权重  $W$  之间的夹角,即特征向量相乘包含有角度信息。为了让卷积网络提取到的特征可以学习到更可分的角度特性,将特征的权重值固定成一个定值  $s$ ,同时为了简化模型参数,将偏置值设为 0。这就将提取到的特征转化到了角度空间,使得决策边界只与角度有关。最后,通过引入一个新参数  $t$  ( $t > 0$ ) 来控制角度余弦量值的大小,使得网络可以学习到更有区分度的特征。最终 Arcface 损失函数为

$$L_{\text{Arcface}} = -\frac{1}{m} \sum_{i=1}^m \ln \left( \frac{e^{s(\cos(\theta_{y_i} + t))}}{e^{s(\cos(\theta_{y_i}))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}} \right) \quad (5)$$

### 1.3.2 AdaCos 损失函数

AdaCos 损失函数在 Arcface 的基础上引入一个动态自适应缩放系数  $s_d$ ,使得在训练过程中可以动态地调整超参数,同时去掉了控制角度余弦量值的参数  $t$ 。最终 AdaCos 损失函数为

$$L_{\text{AdaCos}} = -\frac{1}{m} \sum_{i=1}^m \ln \left( \frac{e^{\tilde{s}_d^{(t)} \cos \theta_{i,j}}}{\sum_{k=1}^{C_1} e^{\tilde{s}_d^{(t)} \cos \theta_{i,k}}} \right) \quad (6)$$

式中:  $C_1$  为类别的数量;  $\theta_{i,j}$  为特征向量  $x_i$  和对应权重之间的夹角;  $\tilde{s}_d^{(t)}$  为一个动态缩放系数,会随着迭代次数  $t_d$  的不同,取值有 2 种情况,如下:

$$\tilde{s}_d^{(t)} = \begin{cases} \sqrt{2} \ln(C_1 - 1) & t_d = 0 \\ \frac{\ln B_{\text{avg}}^{(t_d)}}{\cos\left(\min\left(\frac{\pi}{4}, \theta_{\text{med}}^{(t_d)}\right)\right)} & t_d \geq 1 \end{cases} \quad (7)$$

其中:  $\theta_{\text{med}}^{(t)}$  为引入的一个调试指示变量,是所有对应类别的中心角度。当中心角度大的时候,表示网络参数距离最优还很远,需要一个弱监督信号;当中心角度较小时,表示网络参数距离最优已经很近了,需要一个较强的监督信号。 $B_{\text{avg}}^{(t)}$  为所有非对应 label 的 Logits 之和的平均值,其计算公式为

$$B_{\text{avg}}^{(t)} = \frac{1}{N_1} \sum_{i \in N(t)} \sum_{k_1 \neq y_i} e^{\tilde{s}_d^{(t-1)} \cos \theta_{i,k_1}} \quad (8)$$

式中:  $N(t)$  为一个批次中人脸的类别数量;  $N_1$  为类别总数;  $k_1$  为第  $i$  个图像的真实类别。

由 AdaCos 的公式可以看出,动态自适应参数在每次迭代的时候对分类概率的影响是不同的,由此就可以根据模型的收敛程度动态的产生合理的缩放系数,加快模型的收敛速度。

## 2 改进方法

本节在 MobileFaceNet 人脸特征提取网络的

基础上,将风格注意力机制重新校准模块(Style-based Recalibration Module, SRM)<sup>[13]</sup>加入到 MobileFaceNet 网络中,进一步提高了提取特征的鲁

棒性和全局性。人脸识别方法流程如图2所示,多任务卷积神经网络(MTCNN)是一种人脸检测方法。

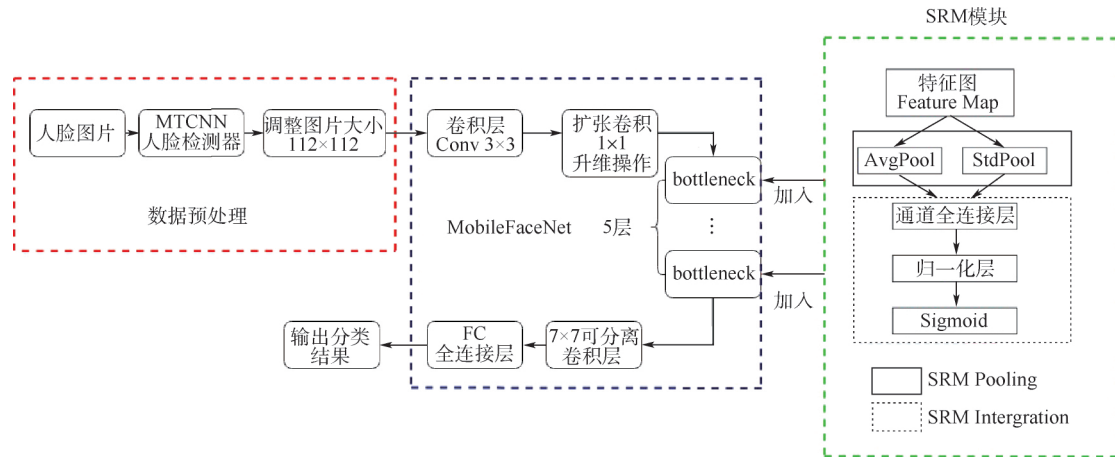


图2 人脸识别方法流程

Fig. 2 Face recognition method flowchart

方法流程主要分为3个部分,即数据预处理、基于SRM模块改进的MobileFaceNet人脸特征提取网络、输出人脸分类结果。之后使用自适应人脸损失函数AdaCos来监督训练过程,减少了设置超参数对训练效果的影响,同时进一步扩大类间间距,缩小类内间距。方法的整体实现是基于pytorch框架。

## 2.1 SRM 模块

通常来说,不同的人脸图像会有不同的属性特征,卷积网络层通过提取图片特征来进行分类和识别。然而,针对不同的人脸图像会有不同的风格,这些风格属性会进一步影响卷积网络对特征的提取,进而影响之后的识别和分类精度。

SRM通过特征重新校准的形式明确地将风格信息合并到卷积神经网络(CNN)表示中。根据不同人脸图像的风格属性,动态地估计各个风格特征的相对重要性,然后根据风格的重要性动态的调整特征权重,这使得网络可以专注于有意义的风格信息,而忽略不必要的风格信息。SRM模块结构如图3所示, $C$ 、 $H$ 、 $W$ 和 $d$ 分别为通道的个数、特征图高度、特征图的宽度和风格特征的数量。

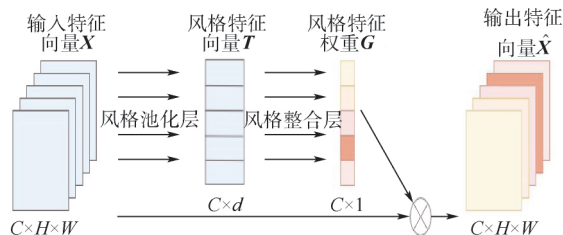


图3 SRM 模块结构

Fig. 3 SRM module structure

SRM 模块由2个主要组件组成:风格池化层(Style Pooling)和风格整合层(Style Integration)。风格池化层通过汇总不通空间维度的特征响应,从每个通道提取风格特征;风格整合层通过逐通道操作,针对图片的不同位置,利用风格特征生成相应的风格权重。同时根据不同的风格权重重新校准特征映射,以强调或隐藏它们的信息

### 2.1.1 风格池化层

风格池化层包括2个部分,即平均池化层和标准差池化层。对于每一个输入为 $X \in \mathbf{R}^{N_d \times C \times H \times W}$ 的特征图, $N_d$ 为小批次中样本的个数,首先通过平均池化层,具体操作如下:

$$\mu_{nc} = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W x_{nchw} \quad (9)$$

式中: $\mu_{nc}$ 为原特征图平均池化后的结果; $x_{nchw}$ 为原特征图 $n$ 、 $c$ 、 $h$ 和 $w$ 分别为 $N_d$ 、 $C$ 、 $H$ 和 $W$ 的分量。其次,将平均池化层的结果与原输入特征图进行标准差的计算,具体过程如下:

$$\sigma_{nc} = \sqrt{\frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W (x_{nchw} - \mu_{nc})^2} \quad (10)$$

式中: $\sigma_{nc}$ 为经过标准池化层后的结果。最后,经过风格池化层的结果如下:

$$t_{nc} = [\mu_{nc} \sigma_{nc}] \quad (11)$$

其中: $t_{nc}$ 为经过风格池化层后的特征向量。

### 2.1.2 风格整合层

风格整合层包括3个部分,即通道全连接层(Channel-wise Fully Connected, CFC)、归一化层(Batch Normalization, BN)<sup>[14]</sup>以及激活层。将从风格池化层输出的结果作为风格整合层的输入,

具体操作如式(12)~式(16)所示。

$$\mathbf{z}_{nc} = \mathbf{w}_c \cdot \mathbf{t}_{nc} \quad (12)$$

$$\boldsymbol{\mu}_c^{(z)} = \frac{1}{N_d} \sum_{n=1}^{N_d} \mathbf{z}_{nc} \quad (13)$$

$$\boldsymbol{\sigma}_c^{(z)} = \sqrt{\frac{1}{N_d} \sum_{n=1}^{N_d} (\mathbf{z}_{nc} - \boldsymbol{\mu}_c^{(z)})^2} \quad (14)$$

$$\hat{\mathbf{z}}_{nc} = \gamma_c \left( \frac{\mathbf{z}_{nc} - \boldsymbol{\mu}_c^{(z)}}{\boldsymbol{\sigma}_c^{(z)}} \right) + \beta_c \quad (15)$$

$$\mathbf{g}_{nc} = \frac{1}{1 + e^{-\hat{\mathbf{z}}_{nc}}} \quad (16)$$

式中:  $\mathbf{z}_{nc}$  为与可学习的参数  $\mathbf{w}_c$  点乘后的特征向量;  $\boldsymbol{\mu}_c^{(z)}$  为取平均值后的特征向量;  $\boldsymbol{\sigma}_c^{(z)}$  为取标准差后的特征向量;  $\hat{\mathbf{z}}_{nc}$  为经过归一化层后的特征向量;  $\mathbf{g}_{nc}$  为经过激活层后的特征向量;  $\mathbf{w}_c$  为可学习的参数,  $\mathbf{w}_c \in \mathbf{R}^{c \times 2}$ ;  $\gamma_c, \beta_c$  为归一化层中的仿射变换参数。

## 2.2 基于风格注意力机制改进的 MobileFaceNet 人脸特征提取网络

MobileFaceNet 人脸特征提取网络中有 5 个 bottleneck 层, 其结构如图 4 所示。

将 SRM 模块加入到 bottleneck 层中后的结构如图 5 所示。

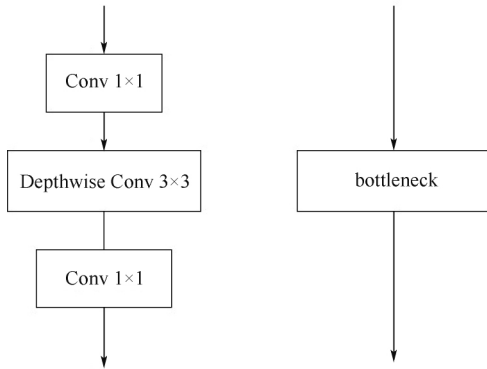


图 4 MobileFaceNet 中 bottleneck 层结构

Fig. 4 Structure of bottleneck layer in MobileFaceNet

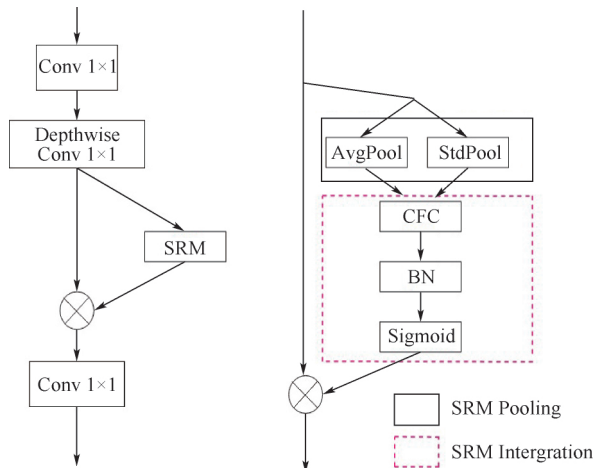


图 5 MobileFaceNet-SRM bottleneck 层结构

Fig. 5 MobileFaceNet-SRM bottleneck layer structure

将 SRM 结构放在了 bottleneck 层中的 Depth-wise 操作之后, 使得特征提取网络可以根据特征图的风格信息来动态地增强有用的特征表达, 抑制可能的噪声, 提高了特征提取的鲁棒性和代表性。这一操作与 MobileNetV3<sup>[15]</sup> 的网络架构相似, 不同的是 MobileNetV3 的网络架构加入 SE 模块<sup>[16]</sup>。

## 3 仿真实验与结果分析

### 3.1 实验设置

#### 3.1.1 数据预处理

本文使用 CASIA-Webface 人脸数据集作为训练数据集, 包含了 10 575 个人的 494 414 张图像。使用 MTCNN<sup>[17]</sup> 人脸检测方法对 CASIA-Webface 中的图片进行再检测, 并将检测到的人脸图片裁剪成 112 × 112 个像素大小。

测试数据集选用的是 LFW、CFP-FF、以及 AgeDB 3 个人脸数据集来进行模型的评估。LFW 数据集包含 5 749 人共 13 233 张人脸图像; 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 FP (Frontal-Profile) 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。

#### 3.1.2 实验环境及参数设置

实验中所有方法都是在 pytorch 框架下通过 python 语言实现的。训练和测试步骤在 NVIDIA GTX2080Ti GPU 上运行。训练总轮数 epoch 设置为 50, 训练初始学习率设置为 0.1, 迭代到 10、25、40 个 epoch 时, 学习率每次除以 10。训练中的 batchsize 设置为 64, 总的权重衰减参数设置为  $5 \times 10^{-4}$ ; 使用随机梯度下降策略 SGD 优化模型, 动量参数设置为 0.9。

### 3.2 在不同 backbone 上的结果

不同的特征提取网络分别为 Resnet-50、MobileFaceNet、MobileFaceNet-SRM, 并使用 AdaCos 人脸识别损失来监督训练过程。ResNet50 是 AdaCos 使用的基础网络; MobileFaceNet 是改进前的特征提取网络; MobileFaceNet-SRM 是加入 SRM 模块后的特征提取网络。评估指标包括在 3 个数据集上的测试准确度以及模型大小, 具体表现如表 3 所示。从表 3 可以看出, 基于 MobileFaceNet 人脸识别方法在 3 个数据集上的测试准确度与 ResNet50 相比, 在 LFW、CFP-FF、AgeDB 3 个数据集上分别提升了 0.49%、0.81%、2.48%, 同时模型参数数量大大降低; 在将 SRM 模块加入到 MobileFaceNet 后, 与改进前的模型相



表3 基于 AdaCos 的损失函数不同卷积框架  
人脸识别模型性能比较

Table 3 Performance comparison of face recognition models with different convolution frames based on AdaCos loss function

网络结构	测试准确度 / %			模型大小 / MB
	LFW	CFP-FF	AgeDB	
ResNet50	98.41	97.88	88.46	161.5
MobileFaceNet	98.90	98.69	90.94	4.9
MobileFaceNet-SRM	99.15	98.85	91.24	5.2

比,模型参数数量略有提升,但是识别精度在 LFW、CFP-FF、AgeDB 3 个数据集上分别提升了 0.25%、0.16%、0.3%,可以看出改进后的模型在精度上有所提升,证明了改进后模型的有效性。

图6是3种主干网络在人脸损失函数 AdaCos 监督训练下,在 AgeDB 评测数据集上的受试者工作特征 (Receiver Operating Characteristic, ROC) 曲线。横轴为假正率 (FPR),指的是分类器识别出的假正实例占有负实例的比例;纵轴为真正率 (TPR),指的是识别出的真正实例占有正实例的比例。从图6可以看出,MobileFaceNet 相较于 ResNet50 网络识别精度有明显提升,同时模型参数数量大大减少,计算效率有所提高。在向 MobileFaceNet 中引入风格注意力机制后,准确度相较于改进前的模型有所提升,这表明改进后的 MobileFaceNet 人脸识别方法的有效性。

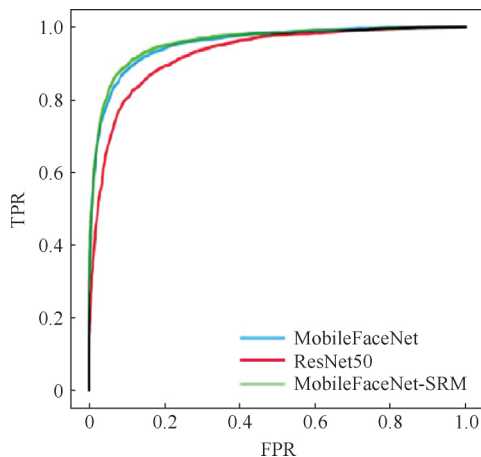


图6 三种主干网络在 AgeDB 数据集上的 ROC 曲线

Fig.6 ROC curves of three backbone networks on AgeDB dataset

## 4 结 论

1) 本文基于 MobileFaceNet 网络提出了一种新的人脸识别方法,通过引入风格注意力机制来增强特征的表达,将改进后的模型作为特征提取网络,大大减少了模型的参数数量,提高了模型的

计算效率,同时提高了提取特征的鲁棒性和代表性。

2) 使用自适应缩放损失函数 AdaCos 作为人脸损失函数来监督训练,通过动态自适应缩放系数,在训练过程中动态调整超参数,加快了模型的收敛速度。在 LFW、CFP-FF 和 AgeDB 3 个人脸数据集上对算法性能进行评测。

3) 实验结果表明,本文方法相较于之前的人脸识别方法,在识别精度有所提升的基础上,大大降低了模型参数的数量,有效地减少了模型的计算量以及模型复杂度。

## 参考文献 (References)

- [1] 徐竟泽,吴作宏,徐岩,等.融合 PCA、LDA 和 SVM 算法的人脸识别[J].计算机工程与应用,2019,55(18):34-37.  
XU J Z, WU Z H, XU Y, et al. Face recognition combined with PCA, LDA and SVM algorithms[J]. Computer Engineering and Applications, 2019, 55(18): 34-37 (in Chinese).
- [2] 任飞凯,邱晓晖.基于 LBP 和数据扩充的 CNN 人脸识别研究[J].计算机技术与发展,2020,30(3):62-66.  
REN F K, QIU X H. Research on CNN face recognition based on LBP and data augmentation[J]. Computer Technology and Development, 2020, 30(3): 62-66 (in Chinese).
- [3] SUN Y, WANG X G, TANG X O. Deep learning face representation from predicting 10 000 classes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 1891-1898.
- [4] SUN Y, WANG X G, TANG X O, et al. Deep learning face representation by joint identification-verification[C]//Advances in Neural Information Processing Systems, 2014: 1988-1996.
- [5] SUN Y, WANG X G, TANG X O. Deeply learned face representations are sparse, selective, and robust[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 2892-2900.
- [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. Piscataway: IEEE Press, 2015: 815-823.
- [7] WANG S S, CHEN Y. A joint loss function for deep face recognition[J]. Multidimensional Systems and Signal Processing, 2019, 30(3): 1517-1530.
- [8] DENG J, GUO J K, XUE N N, et al. Arcface: Additive angular margin loss for deep face recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4690-4699.
- [9] ZHANG X, ZHAO R, QIAO Y, et al. AdaCos: Adaptively scaling cosine logits for effectively learning deep face representations[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 10823-10832.
- [10] CHEN S, LIU Y, GAO X, et al. MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices[C]//

- Chinese Conference on Biometric Recognition, Berlin: Springer, 2018: 428-438.
- [11] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17) [2020-02-20]. <https://arxiv.org/abs/1704.04861>.
- [12] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks [EB/OL]. (2018-01-13) [2019-03-21]. <https://arxiv.org/abs/1801.04381>.
- [13] LEE H J, KIM H E, NAM H. SRM: A style-based recalibration module for convolutional neural networks [EB/OL]. (2019-05-26) [2020-02-20]. <https://arxiv.org/abs/1903.10829?context=cs>.
- [14] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-03-02) [2020-02-20]. <http://arxiv.org/abs/1502.03167>.
- [15] CHU X X, ZHANG B, XU R J. MoGA: Searching beyond MobileNetV3 [EB/OL]. (2019-08-04) [2020-02-20]. <https://arxiv.org/abs/1908.01314v2>.
- [16] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [J]. (2017-09-05) [2020-02-20]. <https://arxiv.org/abs/1709.01507>.
- [17] 常思远, 李有乘, 孙培岩, 等. 一种基于 MTCNN 的视频人脸检测及识别方法 [J]. 许昌学院学报, 2019, 38(2): 154-157.
- CHANG S Y, LI Y S, SUN P Y, et al. A video face detection and recognition method based on MTCNN [J]. Journal of Xuchang University, 2019, 38(2): 154-157 (in Chinese).

作者简介:

张子昊 男 硕士研究生。主要研究方向: 行人再识别。

王蓉 女 博士 教授 博士生导师。主要研究方向: 模式识别、人工智能。

## Improved face recognition method based on MobileFaceNet network

ZHANG Zihao<sup>1,2</sup>, WANG Rong<sup>1,2,\*</sup>

(1. School of Police Information Engineering and Network Security, People's Public Security University of China, Beijing 100038, China;

2. Key Laboratory of Security Technology & Risk Assessment, Beijing 100038, China)

**Abstract:** In order to solve the problem of more convolutional model parameters and slower convergence speed during training, an improved face recognition method based on MobileFaceNet network is proposed. First, we use the MobileFaceNet network to extract facial features. In the process of extracting features, the number of convolutional layer parameters in the model is reduced by introducing separable convolution. Then, the style attention mechanism is introduced in the MobileFaceNet network to enhance the expression of features. At the same time, the AdaCos face loss function is used to train the model, and the adaptive scaling factor in the AdaCos loss function is used to dynamically adjust the hyperparameters to avoid the effect of artificially setting hyperparameters on the model. Finally, we evaluate the training model on the LFW, AgeDB and CFP-FF test dataset, respectively. The experimental results show that the recognition accuracy of the improved model on the LFW, AgeDB and CFP-FF test dataset has increased by 0.25%, 0.16% and 0.3%, respectively, indicating that the improved model has higher accuracy and robustness than the model before improvement.

**Keywords:** face recognition; deep learning; MobileFaceNet; AdaCos; convolutional neural network

**Received:** 2020-02-25; **Accepted:** 2020-04-18; **Published online:** 2020-04-21 11:39

**URL:** [kns.cnki.net/kcms/detail/11.2625.V.20200421.0911.001.html](http://kns.cnki.net/kcms/detail/11.2625.V.20200421.0911.001.html)

**Foundation items:** Ministry of Public Security Technology Research Project (2017JSYJB01); the Fundamental Research Funds for the Central Universities (2019JKF111)

\* **Corresponding author.** E-mail: dbdxwangrong@163.com