光电技术及应用

DOI：10.16818/j.issn1001－5868.2021110204

基于改进 MobileFaceNet 的人脸识别方法

胡佳荣，孟　文*，赵晶晶

（西南交通大学 机械工程学院，成都 610031）

摘　要： 针对目前大多数人脸识别算法参数多、计算量大，难以部署到移动端和嵌入式设备中的问题，提出了一种基于改进 MobileFaceNet 的人脸识别方法。通过对 MobileFaceNet 模型结构的调整，将 bottleneck 模块优化为 sandglass 模块，改良深度卷积和逐点卷积的相对位置，适当增大 sandglass 模块的输出通道数，从而减少特征压缩时的信息丢失，增强人脸空间特征的提取。实验结果表明：改进后的方法在 LFW 测试数据集上准确率达 99.15%，模型大小和计算量分别仅为原算法的 61%和 45%，验证了所提方法的有效性。

关键词： 人脸识别；MobileFaceNet；深度学习；卷积神经网络；深度可分离卷积

中图分类号： TP391　**文章编号：** 1001－5868(2022)01－0164－05

Face Recognition Method Based on Improved MobileFaceNet

HU Jiarong，MENG Wen，ZHAO Jingjing

（School of Mechanical Engineering，Southwest Jiaotong University，Chengdu 610031，CHN）

Abstract： Most current face recognition algorithms are difficult to deploy to mobile terminals and embedded devices because of their many parameters and large amount of calculation，thus a face recognition method based on improved MobileFaceNet is proposed. By adjusting the structure of the MobileFaceNet model，the bottleneck module is optimized to the sandglass module，the relative position of depthwise convolution and pointwise convolution are improved，and the number of output channels of the sandglass module is appropriately increased，thereby reducing information loss during feature compression and enhancing extraction of the facial spatial features. The experimental results show that the improved method achieves an accuracy of 99.15% on the LFW test data set，and the model size and calculation amount are only 61% and 45% of that of the original algorithm，verifying the effectiveness of the proposed method.

Key words： face recognition；MobileFaceNet；deep learning；convolutional neural network；depthwise separable convolution

0　引言

人脸识别是根据人脸部形状特征进行身份识别的生物识别技术，相较于指纹、虹膜等生物识别技术具有无接触性、无强制性等优点，已广泛应用于金融、安防、通行、公安等多个领域。人脸识别的关键是需要在不同的光照、角度、姿态等因素下从人脸图像中提取有效的识别特征，在减小类内差距的同时增大类间差距，从而区分不同个体[1]。

传统的人脸识别大致经历了基于几何结构特征、基于子空间分析、基于特征描述等发展阶段[2]。Belhumeur 提出的 FisherFace 采用主成分分析法（PCA）对人脸图像降维，再采用线性判别分析（LDA）对低维子空间进行特征提取[3]。Ahonen 则是提取人脸图像各个子区域内的局部二值模式

收稿日期：2021－11－02.

* 通信作者：孟　文

E-mail：mengwen@home.swjtu.edu.cn

(LBP)直方图，而后将它们连接成一个单一的直方图，以此作为人脸判别特征[4]。以上这些特征提取方法往往需要根据人的经验来设计提取特征，具有一定的局限性。

随着深度学习的应用，人脸识别研究得到了快速发展。Facebook 团队提出的 DeepFace 算法[5]，利用卷积神经网络对对齐后的人脸分类。Google 提出的 FaceNet[6]则是利用主干网络将人脸图像映射到 128 维的特征向量上，并引入三元组损失(Triplet Loss)学习人脸特征间的可分性，在 LFW 数据集上取得了 99.63%的准确率。Liu[7]提出的 SphereFace 采用了 A-Softmax Loss，将角度特征空间作为人脸判别特征，在 LFW 数据集上取得了 99.42%的准确率。Deng[8]在 SphereFace 的基础上进一步提出了 ArcFace(Additive Angular Margin Loss)，改进了特征向量归一化和加性角度间隔。但这些人脸识别算法采用的主干网络较为复杂、模型参数多、计算量大，难以应用于移动端或嵌入式设备。目前，常用的轻量级网络主要有 MobileNet[9]，ShuffleNet[10] 等。MobileFaceNet[11] 采用了 MobileNetV2[12]的 bottleneck 模块，并对网络进行了改进，使得模型大小仅有 4 M，同时模型准确率达到了 99.55%。

本文对 MobileFaceNet 进行了改进，将 bottleneck 模块优化为 sandglass[13]模块，能更有效地提取人脸特征信息，在不损失精度的情况下，模型大小和计算量大大减少。

1　基于改进 MobileFaceNet 的人脸识别方法

1.1　MobileFaceNet

MobileFaceNet 网络采用 MobileNetV2 的倒残差瓶颈模块(Inverted Residual and Linear Bottleneck，后面简称为 bottleneck 模块)，如图 1 所示。Bottleneck 模块借鉴了深度可分离卷积(Depthwise Separable Convolution)思想，先通过逐点卷积(Pointwise Convolution)增大通道数，再使用深度卷积(Depthwise Convolution)在高维特征图上提取丰富的特征信息，最后采用逐点卷积进行降维。同时为了减少特征信息的丢失，最后一个逐点卷积层采用 linear 激活函数。图中的捷径分支只有在输入特征图和输出特征图尺寸相同时才能相加。

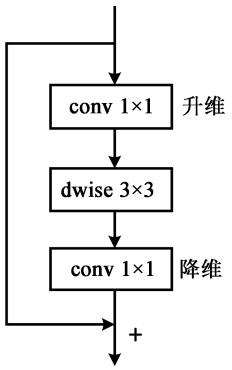

图 1　bottleneck 模块

表 1 表示一个尺寸为 $D_F \times D_F \times M$ 的特征图经过 bottleneck 模块处理后得到 $\frac{D_F}{s} \times \frac{D_F}{s} \times N$ 的特征图，其中，s 为步距，t 为扩张系数。

表 1　bottleneck 模块处理过程

输入	卷积操作	输出
$D_F \times D_F \times M$	1×1 conv，ReLU6	$D_F \times D_F \times tM$
$D_F \times D_F \times tM$	3×3 dwise s，ReLU6	$\frac{D_F}{s} \times \frac{D_F}{s} \times tM$
$\frac{D_F}{s} \times \frac{D_F}{s} \times tM$	1×1 conv，linear	$\frac{D_F}{s} \times \frac{D_F}{s} \times M$

基于 bottleneck 模块，MobileFaceNet 提出的模型架构如表 2 所示，t 为扩张系数，c 为输出通道数，n 为 bottleneck 的序列数，s 为 bootleneck 序列第一层的步距，之后的步距为 1。

表 2　MobileFaceNet 模型架构

输入	卷积操作	t	c	n	s
$112^2 \times 3$	conv 3×3	—	64	1	2
$56^2 \times 64$	dwise conv 3×3	—	64	1	1
$56^2 \times 64$	bottleneck	2	64	5	2
$28^2 \times 64$	bottleneck	4	128	1	2
$14^2 \times 128$	bottleneck	2	128	6	1
$14^2 \times 128$	bottleneck	4	128	1	2
$7^2 \times 128$	bottleneck	2	128	2	1
$7^2 \times 128$	conv 1×1	—	512	1	1
$7^2 \times 512$	linear GD conv 7×7	—	512	1	1
$1^2 \times 512$	linear conv 1×1	—	512	1	1

传统人脸识别算法多采用全局平均池化层处理特征图，但 MobileFaceNet 认为特征图不同位置的重要性也是不同的，平均池化层会弱化关键信息的提取。因此 MobileFaceNet 对最后得到的 7×7×512 的特征图采用了全局深度卷积，能够更有效地学习人脸重要位置的特征。全局深度卷积的公式如下：

$$G_m = \sum_{i,j} K_{i,j,m} \cdot F_{i,j,m} \tag{1}$$

式中，F 为输入特征图，K 为卷积核，其中 K 的尺寸和 F 的尺寸相同，都为 $W\times H\times M$。经过全局深度卷积得到尺寸为 $1\times1\times M$ 的特征 G。

MobileFaceNet 的损失函数采用 ArcFace 损失函数，公式如下：

$$L=-\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))}+\sum_{j=1,j\neq y_i}^{n}e^{s\cos\theta_j}} \tag{2}$$

式中，N 为 batch size 的大小，n 为类别数量，y_i 为真实类别，θ 为特征权重和特征之间的夹角，s 和 m 分别为特征尺度和角度间隔。

1.2 Sandglass 优化模块

本文在 MobileFaceNet 的基础上进行改进，将图 1 所示 bottleneck 模块优化为图 2 所示 sandglass 模块。虽然 bottleneck 模块先对特征图升维，在更深的通道上提取丰富的空间信息，但是在将高维空间压缩到低维空间的过程中，会导致空间信息的丢失，引起梯度混淆。为了解决这个问题，sandglass 模块采用两次深度卷积，并置于两个 1×1 逐点卷积之外，从而减少空间信息编码的丢失，保证更多的特征信息从输入端传递到输出端。

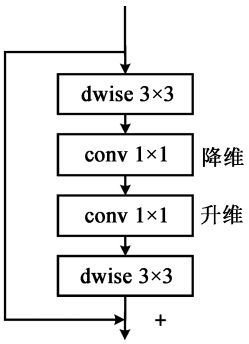

图 2 sandglass 模块

表 3 表示一个尺寸为 $D_F\times D_F\times M$ 的特征图经过 sandglass 模块处理得到 $\frac{D_F}{s}\times\frac{D_F}{s}\times N$ 的特征图，s 为步距，t 为缩小系数。公式如下：

$$\hat{G}=\phi_{1,p}\phi_{1,d}(F) \tag{3}$$

$$G=\phi_{2,d}\phi_{2,p}(\hat{G})+F \tag{4}$$

式中，$\phi_{i,p}$ 和 $\phi_{i,d}$ 分别表示第 i 个逐点卷积和第 i 个深度卷积。

在不考虑捷径分支的情况下，一个尺寸为 $D_F\times D_F\times M$ 的特征图经过卷积操作得到 $D_F\times D_F\times N$ 的特征图，bottleneck 模块和 sandglass 模块的计算量分别为

$$\begin{aligned}C_{btn}=&1\times1\times M\times t_eM\times D_F\times D_F+\\&D_K\times D_K\times t_eM\times t_eM\times D_F\times D_F+\\&1\times1\times t_eM\times N\times D_F\times D_F\end{aligned} \tag{5}$$

$$\begin{aligned}C_{sg}=&D_K\times D_K\times M\times M\times D_F\times D_F+1\times1\times\\&M\times\frac{M}{t_r}\times D_F\times D_F+1\times1\times\frac{M}{t_r}\times N\times D_F\times\\&D_F+D_K\times D_K\times N\times N\times D_F\times D_F\end{aligned} \tag{6}$$

二者计算量之比为

$$\eta=\frac{C_{sg}}{C_{btn}}=\frac{D_K^2(M^2+N^2)+\left(\frac{M^2}{t_r}+\frac{MN}{t_r}\right)}{t_e^2D_K^2M^2+t_e(M^2+MN)} \tag{7}$$

式中，D_K 为卷积核的宽度和高度，通常取 3；t_e，t_r 分别为 bottleneck 的扩大系数和 sandglass 的缩小系数，通常取 2～6，而一般情况下

$$M\leqslant N\leqslant 2M \tag{8}$$

因此，可以得到：

$$\eta\leqslant\frac{2D_K^2M^2+\frac{2M^2}{t_r}}{t_e^2D_K^2M^2+3t_eM^2}\leqslant\frac{2D_K^2+2}{t_e^2D_K^2+3t_e}=\frac{20}{9t_e^2+3t_e} \tag{9}$$

可以看到 t_e 的值越大，采用 sandglass 模块能够减少的计算量就越明显。即使 bottleneck 模块的 t_e 取较小值 2，sandglass 模块的计算量仍能缩小为 bottleneck 模块的一半。

表 3 sandglass 模块处理过程

输入	卷积操作	输出
$D_F\times D_F\times M$	3×3 dwise，ReLu6	$D_F\times D_F\times M$
$D_F\times D_F\times M$	1×1 conv，linear	$D_F\times D_F\times\frac{M}{t}$
$D_F\times D_F\times\frac{M}{t}$	1×1 conv，ReLu6	$D_F\times D_F\times N$
$D_F\times D_F\times N$	3×3 dwise s，linear	$\frac{D_F}{s}\times\frac{D_F}{s}\times N$

大量实验表明，线性瓶颈结构有利于避免出现特征零化现象，减少信息丢失。因此，在 sandglass 模块的第一个逐点卷积层和第二个深度卷积层采取线性输出。而 PReLU 函数已经证明在人脸识别算法中的性能优于 ReLU6 函数。因此 sandglass 模块另外两层改用 PReLU 激活函数。

1.3 改进的 MobileFaceNet

基于 sandglass 模块，对 112×96 的人脸图像，本文提出了改进的 MobileFaceNet 模型架构，如表 4 所示。可以看到，改进的 MobileFaceNet 结构与原来结构几乎相同，只是将其中的 bottleneck 模块优化为 sandglass 模块，并对 sandglass 序列的重复

次数进行了调整。同时，考虑到 sandglass 模块采用的深度卷积通道数过少，会使空间信息提取不充分，导致人脸识别的准确率下降，因此将最后一个 sandglass 模块的特征图通道数扩大为 256。但即使这样，在保证准确率的情况下，模型的参数量和计算量相较于改进前减少很多。

表 4　改进的 MobileFaceNet

输入	卷积操作	t	c	n	s
112×96×3	conv 3×3	—	64	1	2
56×48×64	dwise conv 3×3	—	64	1	1
56×48×64	sandglass	2	64	5	2
28×24×64	sandglass	6	128	2	2
14×12×128	sandglass	2	128	4	1
14×12×128	sandglass	6	256	4	1
14×12×256	sandglass	2	256	2	2
7×6×256	conv 1×1	—	512	1	1
7×6×256	linear GDconv 7×6	—	512	1	1
1×1×512	linear conv 1×1	—	512	1	1

2　实验结果与分析

2.1　数据集

实验使用的训练数据集为 CASIA-Webface 数据集，共有 10 575 个人，494 414 张人脸图像。实验使用的测试数据集为 LFW 数据集，共有 5 749 个人，13 232 张人脸图像，从中随机选择 6 000 对人脸组成人脸识别图像对，其中 3 000 对属于同一个人的 2 张图像，3 000 对属于不同的两个人的图像。本文中的所有人脸图像都按照 SphereFace 的处理方法，采用多任务卷积神经网络（MTCNN）[14] 进行人脸检测，再通过相似变换进行人脸对齐，最后将图像裁剪成 112×96 大小。

2.2　实验过程

本文模型在 pytorch1.4.0 深度学习框架上实现，实验使用的配置为 NVIDIA GTX 1060 GPU，Intel Core(TM) i7-8750H CPU，8 GB 内存。为了保证实验的公平性，模型训练策略参考了 MobileFaceNet 的训练过程，将最后一层权重衰减参数设置为 4×10^{-4}，其余层的权重衰减参数设置为 4×10^{-5}。采用 SGD 优化器，动量系数设置为 0.9，初始学习率设置为 0.1，epoch 设置为 60，当迭代到 32，52 和 58 个 epoch 的时候，学习率除以 10，batch size 设置为 128。

2.3　结果分析

模型的性能对比如表 5 所示。可以看到 MobileFaceNet 作为一款优秀的移动端人脸识别模型，在 LFW 数据集上的准确率为 99.18%，同时模型大小和 Mult-Adds（衡量计算成本的指标）只有 3.94 和 387 M。将 bottleneck 模块优化为 sandglass 模块后，改进模型的准确率仍能达到 99.15%，而模型大小和 Mult-Adds 分别为 2.4 和 176 M，仅为原模型的 61% 和 45%。实验结果表明，基于改进的 MobileFaceNet 切实可行，在保证准确率的情况下，大大减少了模型大小和计算量，提升了模型性能。

表 5　模型性能对比

模型	准确率/%	Mutl-Adds	模型大小
MobileFaceNet	99.18	387 M	3.94 M
改进 MobileFaceNet	99.15	176 M	2.4 M

3　结论

本文对 MobileFaceNet 模型进行了改进，将 bottleneck 模块优化为 sandglass 模块，减少了特征压缩时的信息丢失，增强了人脸空间特征的提取。实验结果表明，改进方法在识别精度不变的情况下，模型大小和计算量大大降低，有效提升了模型性能。

参考文献：

[1] 余璀璨，李慧斌. 基于深度学习的人脸识别方法综述[J]. 工程数学学报，2021，38(4)：451-469.
Yu Cuican，Li Huibin. Deep learning based 2D face recognition：A survey [J]. Chinese J. of Engineering Mathematics，2021，38(4)：451-469.

[2] 景晨凯，宋　涛，庄　雷，等. 基于深度卷积神经网络的人脸识别技术综述[J]. 计算机应用与软件，2018，35(1)：223-231.
Jing Chenkai，Song Tao，Zhuang Lei，et al. A survey of face recognition technology based on deep convolutional neural networks[J]. Computer Applications and Software，2018，35(1)：223-231.

[3] Belhumeur P N，Hespanha J P，Kriegman D J. Eigenfaces vs. fisherfaces：Recognition using class specific linear projection [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence，1997，19(7)：711-720.

[4] Ahonen T，Hadid A，Pietikainen M. Face recognition with localbinary patterns [C]// The 8th European Conf. on Computer Vision，Prague，Czech Republic. Berlin，Heidelberg：Springer-Verlag，2004：469-481.

[5] Taigman Y，Yang M，Ranzato M，et al. DeepFace：closing the gap to human-level performance in face verification[C]// 2014 IEEE Conf. on Computer Vision and Pattern Recognition，2014：1701-1708.

[6] Schroff F，Kalenichenko D，Philbin J. Facenet：A unified

embedding for face recognition and clustering[C]// IEEE Conf. on Computer Vision and Pattern Recognition, 2015: 815-823.

[7] Liu W Y, Wen Y D, Yu Z D, et al. SphereFace: Deep hypersphere embedding for face recognition[EB/OL]. (2018-01-29). https://arxiv.org/abs/1704.08063.

[8] Deng J, Guo J K, Xue N N, et al. ArcFace: Additive angular margin loss for deep face recognition[EB/OL]. (2019-02-09). https://arxiv.org/abs/1801.07698.

[9] Howard A G, Zhu M L, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17). https://arxiv.org/abs/1704.04861v1.

[10] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[EB/OL]. (2017-10-07). http://https://arxiv.org/abs/1707.01083.

[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/abs/1804.07573.

[12] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[EB/OL]. (2019-03-21). https://arxiv.org/abs/1801.04381.

[13] Zhou D Q, Hou Q B, Chen Y P, et al. Rethinking bottleneck structure for efficient mobile network design[EB/OL]. (2020-11-27). https://arxiv.org/abs/2007.02269.

[14] Zhang K P, Zhang Z P, Li Z F, et al. Joint face detection and alignment using multi-task cascaded convolutional networks[EB/OL]. (2016-04-11). https://arxiv.org/abs/1604.02878.

作者简介：

胡佳荣(1996—),男,浙江省绍兴市人,硕士研究生,研究方向为深度学习和图像处理;

孟　文(1966—),男,四川省德阳市人,博士,副教授,硕士研究生导师,研究方向为智能控制技术及应用。