叶继华, 郭凤,黎欣,等. DTFBNet:一种面向智能终端的轻量级人脸识别新方法[J]. 江西师范大学学报(自然科学版),2022,46(2):126-133.
YE Jihua,GUO Feng,LI Xin,et al. DTFBNet:the new lightweight face recognition method for smart terminals[J]. Journal of Jiangxi Normal University(Natural Science),2022,46(2):126-133.

文章编号:1000-5862(2022)02-0126-08

DTFBNet:一种面向智能终端的轻量级人脸识别新方法

叶继华,郭 凤,黎 欣,江 蕗,江爱文
(江西师范大学计算机信息工程学院,江西 南昌 330022)

摘要:对于智能终端资源不足的问题,目前有许多解决方法,但普遍存在依赖样本数据和参数量的问题. 为此,该文先构造了一个深度卷积和传统卷积融合的模块 DTFBlock(depthwise convolution and traditional convolution fusion Block);然后在该基础上提出了一种基于 MobileFaceNet 的改进方法 DTFBNet,DTFBNet 参数量较小,在网络的识别效果较好;最后,在面部识别数据集 CASIA-Webface 和 LFW 上进行实验,结果表明:该算法的最高准确率达到了99.40%,达到在同等参数量上具有竞争力的分类准确率.

关键词:DTFBNet;DTFBlock;融合损失;轻量级;人脸识别
中图分类号:TP 391.4 **文献标志码**:A **DOI**:10.16357/j.cnki.issn1000-5862.2022.02.03

0 引言

人脸识别是一种重要的身份认证技术,已经应用于越来越多的智能终端,如设备解锁、应用登录、移动支付等. 一些配备了人脸验证技术的移动应用程序(如智能手机解锁),需要离线运行. 为了在有限的计算资源下实现用户友好性,智能终端本地部署的人脸识别模型不仅要准确,而且要小而快. 然而,现代高精度人脸识别模型建立在深度和大卷积神经网络(CNN)上,在训练阶段由损失函数监督. 由于需要大量计算资源的大型 CNN 模型并不适用于许多移动和嵌入式应用程序,因此,在嵌入式领域受计算能力和高吞吐量要求限制的环境下,部署基于深度学习的人脸识别模型仍然具有挑战性[1-2].

最近,通过利用计算机视觉任务而设计的轻量级深度学习模型架构,在设计高效的人脸识别解决方案方面取得了较大的进展,如 MobileNetV2[3]、ShuffleNet[4]、VarGNet[5] 等. MobileFaceNet[6] 是最早提出的高效人脸识别模型,具有大约 1 M 参数和 439 M FLOPs. MobileNetV2 的架构是基于反向残差结构和深度可分离卷积[7]. AirFace[8]、ShuffleFaceNet[9] 和 VarGFaceNet[10-11] 模型架构分别由 MobileNetV2、ShuffleNetV2 和 VarGNet 来构建,使用具有大约 1 G FLOPs 计算复杂度的紧凑模型并且达到较高准确度. ShuffleNetV2 利用 ShuffleNetV1 提出的通道 Shuffle 操作,在准确性和计算效率之间进行折中处理. VarGNet 建议固定在每个组卷积中的输入通道数量,而不是固定总组数,以平衡卷积块内部的计算强度. LFR 的 deepglint-light 轨道通过 1 G FLOPs 的计算复杂度和20 M 的内存占用(约5 M 的可训练参数)来实现在环境约束的条件下的人脸识别. CondenseNet[12] 将密集连接与学习组卷积相结合,以促进特征重用,同时消除了冗余连接. Sun Ke 等[13] 提出了在参数和计算方面都较为有效的交错组卷积,而文献[14]引入了移位操作来取代昂贵的空间卷积. VarGFaceNet 考虑了不同组数对提取有效特征的影响,MobileNetV2 考虑了深度可分离卷

收稿日期:2022-01-03
基金项目:国家自然科学基金(62167005,61966018)和江西省教育厅重点科研课题(GJJ200302)资助项目.
作者简介:叶继华(1966—),男,江西上饶人,教授,博士生导师,主要从事普适计算、物联网技术、数据融合、图像处理等研究. E-mail:yjhwcl@163.com

积. 在深度可分离卷积中的深度卷积和点卷积相当于 2 次卷积的过程,并且与输入通道和输出通道、卷积核大小相同的传统卷积相比,深度可分离卷积的参数量会少于传统卷积的参数量. 深度卷积是特殊的 group convolution,深度卷积的输入通道数、输出通道数、组数相同,这表明通过深度卷积计算后的特征会有信息损失,从而导致需要更多的参数才能学习到具有正确分类能力的特征.

基于以上问题,本文在 MobileFaceNet 的基础上将深度可分离模块替换成 DTFBlock 模块. 该模块考虑了深度卷积无法提取含有更多细节的低级特征的问题. DTFBlock 模块对损失的信息进行补偿,可以提取到含有更多细节的低级特征. DTFBlock 模块将深度卷积提取的特征和 2 个系统模块提取的特征进行融合,从而得到含有更多细节的低级特征,这样有利于 DTFBNet 网络提取更具有识别性的特征.

在 LFW[15] 测试数据集上进行实验,实验结果表明本文改进的方法提升了模型的准确性.

1　相关工作

传统卷积同批次输入的特征会一次处理成所需输出特征,但深度可分离卷积并不是对输入的特征一次处理成所需输出特征,而是将传统卷积过程分解成深度卷积和点卷积(见图 1). 深度卷积计算每个输入的特征,输出中间特征,所有的中间特征通过点卷积得到最终的输出特征. 这样不仅计算了图像的空间维度,还计算了图像深度维度. 输入的特征 $\boldsymbol{x} \in \mathbf{R}^{C_{\text{in}} \times h \times w}$,卷积核的大小为 $k \times k$,输出的通道数为 C_{out}. 标准卷积的参数量为

$$N_{\text{sc}} = k \times k \times C_{\text{out}} \times C_{\text{in}}, \tag{1}$$

深度卷积运算的参数量为

$$N_{\text{dc}} = C_{\text{in}} \times k \times k, \tag{2}$$

点卷积运算的参数量为

$$N_{\text{pc}} = C_{\text{in}} \times C_{\text{out}}, \tag{3}$$

通过深度可分离卷积的总参数量为

$$N_{\text{dsc}} = N_{\text{dc}} + N_{\text{pc}}. \tag{4}$$

从式(1) ~ (4) 可以得出,用深度可分离卷积来代替传统卷积,参数减少数量为

$$F = N_{\text{dsc}}/N_{\text{sc}} = 1/C_{\text{out}} + 1/k^2. \tag{5}$$

由式(5)可知:相比于传统卷积,深度可分离卷积减少参数量,具有计算成本优势.

MobileNetV2 和 Xception[16] 等模型均采用了深度可分离卷积,用于智能终端以减少模型的参数量,从而降低模型的复杂度.

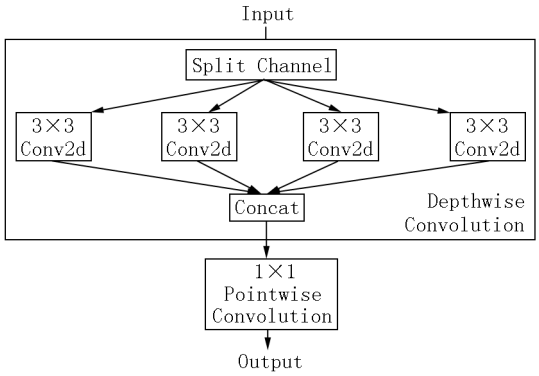

图 1　深度可分离卷积

2　DTFBNet 模型

在 MobileFaceNet 网络中使用了深度卷积,利用深度卷积降低模型参数量. 但深度卷积相对于传统卷积提取的特征信息不够丰富,会阻碍网络的优化. 针对深度卷积提取的特征存在信息损失的问题,借鉴恒等映射的思想和在 MobileFaceNet 中轻量级人脸识别网络设计原则,该文提出了 DTFBNet. DTFBNet 模型主要分为 4 个部分:(i)网络的架构设计;(ii)针对深度卷积提取特征存在信息损失的问题提出了 DTFBlock(depthwise convolution and traditional convolution fusion Block);(iii)特征融合,讨论最佳特征融合方式;(iv)损失函数,在 DTFBNet 模型中使用 ArcFace 损失函数和融合损失函数,其中融合损失函数是针对 DTFBlock 模块在特征融合过程中出现信息损失的问题引入的. 下面将从网络架构、DTFBlock 模块、特征融合和损失函数 4 个方面来介绍 DTFBNet 模型.

2.1　网络架构

在轻量级人脸识别中,MobileFaceNet 是经典的轻量级架构之一,因此,本文选用 MobileFaceNet 架构作为本文模型改进的基础模型来构建 DTFBNet,用于人脸识别的轻量级 CNN,具体的网络结构如图 2 所示,网络中的 DTFBlock 如图 3 所示. DTFBlock 是由 3×3、3×3、3×3、1×1 的卷积核和一个将 2 个输出特征进行融合的结构组成,其中 1×1 卷积核的输出通道数量是输入通道数量的 2 倍,其余 3×3 卷积核的输出通道数量和输入通道数量相同. 网络中的 bottleneck 是倒置的残差结构,它从 MobileNetV2 中引入了线性约束,但是扩展因子比 MobileNetV2 的更小. 倒置的残差结构是由 1×1、3×3、1×1 的卷积核和 1 个将输入特征和输出特征进行融合的结构

组成. 在该结构中,通道数量的变化和结构与残差结构的相反,因此被称为倒残差结构. 此外,使用 PReLU 作为激活函数,并在批量训练时进行批量归一化处理. 在网络中的 Conv 1 ×1 是由 1 ×1 的卷积核构成,该结构将 bottleneck 输出特征进行融合并提取高级语义特征. 在网络中的 Linear GDConv 7 ×6 是由 7 ×6的卷积核构成,该卷积核偏执设置为 false. 该结构的作用是将特征进行下采样,下采样后的特征大小为 1 ×1. 在网络中最后一层 Linear Conv 1 ×1 是由1 ×1的卷积核构成,该卷积核的偏执设置为 false.

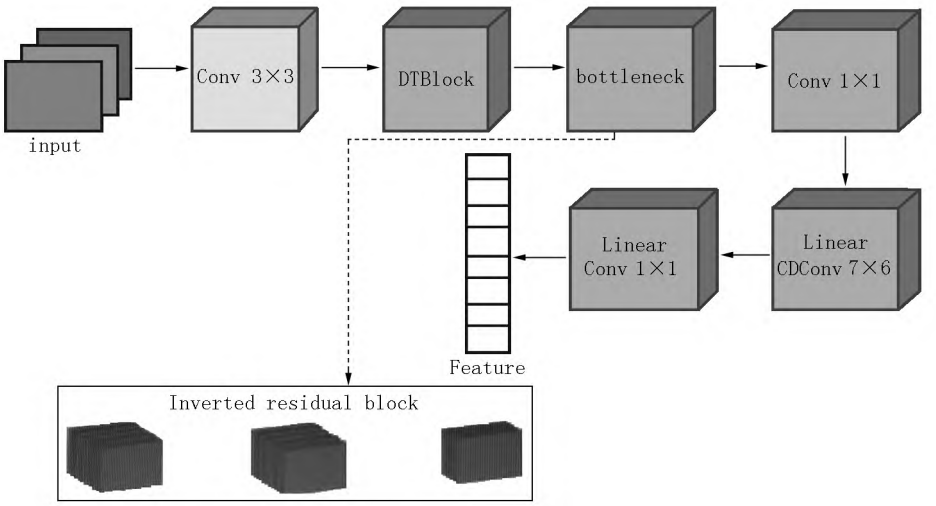

图 2 DTFBNet

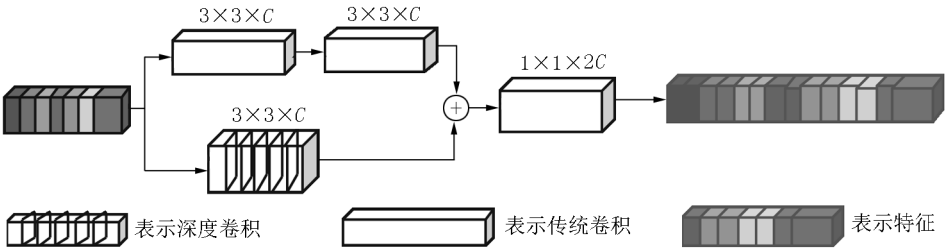

图 3 DTFBlock

在网络中的 bottleneck 可以提取信息丰富的特征,这是由 DTFBlock 输出的特征输入到 bottleneck 中所致. 在 Conv 1 ×1 输出通道数量为输入通道数量的 4 倍情况下,即输出通道数量为 512,会出现特征冗余. 实验结果表明:将 DTFBlock 提取的特征输入 bottleneck 网络结构中,bottleneck 可以提取的特征含有更多细节的信息,这将导致在 Conv 1 ×1 中提取的特征出现特征冗余,模型的精确率没有上升,反而有下降的趋势;当将 Conv 1 ×1 的输出通道数量设为 256 时,效果最佳. 减少 Conv 1 ×1 输出通道数将直接影响 Linear GDConv 7 ×6 的输入通道数量,同时减少 Conv2 输出通道数和 Linear GDConv 7 ×6 的输入通道数是本文减少参数量的关键. 虽然在本文中 DTFBlock 模块引入了深度卷积更多的参数量,但是 DTFBlock 提取的特征相对更加丰富,也更加具有判别力,因此,Conv 1 ×1 的输出通道数和 Linear GDConv 7 ×6 的输入通道数减少的参数量比 DTFBlock 引入的参数量更多,因此,DTFBNet 的总参数数量比 MobileFaceNet 的总参数数量更少.

2.2 DTFBlock

设计一个高效并且参数量较少的模型一般采取的卷积核大小为 1 ×1 和 3 ×3,而 3 ×3 的卷积核相比 1 ×1 的卷积核有更大的感受野. 感受野大的卷积核提取的特征含有全局信息,而感受野较小的卷积核提取的特征属于局部信息,DTFBlock 提取的低级特征需要包含了全局信息. 因此,在 DTFBlock 中采用卷积核大小为 3 ×3 的传统卷积.

本文网络中的深度卷积是一个卷积核大小为 3 ×3、步长为 1,输入通道数和输出通道数均为 32 的卷积. 深度卷积的计算特点是每个卷积核对应一个特征通道,因此,深度卷积没有计算输入特征之间的通道关系,通过深度卷积提取到的特征会有信息损失. 基于深度卷积的特点,本文设计了 DTFBlock. DTFBlock 的设计思想来源于残差网络中的恒等映射的思想,该思想主要是:深层的网络不应比浅层的网络表现得更差,因此设计出了残差块. 本文借助恒等映射的思想设计了 DTFBlock 模块,该模块可以提取含有更多细节的低级特征. DTFBlock 模块主要由

2 种卷积运算构成:一种是传统卷积运算,另一种是深度卷积运算. 如图 4 所示,深度卷积运算是将每个通道对应一个卷积核进行卷积. 因此,每个通道提取的特征只能体现每个通道的重要性或者特殊性,从而忽略了在输入特征中的空间信息. 传统卷积的计算方式如图 5 所示,传统卷积的计算方式是所有特征图和一个卷积核计算,即所有通道与一个卷积核进行计算,每个通道计算后求和得出每个卷积核提取的特征. 传统卷积提取的特征既有通道间的信息又有空间的信息,因此本文借助恒等映射的思想将 2 种卷积运算进行恒等运算,将这 2 种卷积提取的特征进行融合,融合后的特征既含有空间信息又突出了每个通道的特殊性. 在 DTFBlock 中选择 2 个传统卷积,主要原因在于 MobileNetV2 中的 ReLU 会破坏低维空间的数据. 因此,本文借助该思想选择 2 个传统卷积提取特征,这样既不会引入太多的参数量又可以提取包含更多的低维信息的特征.

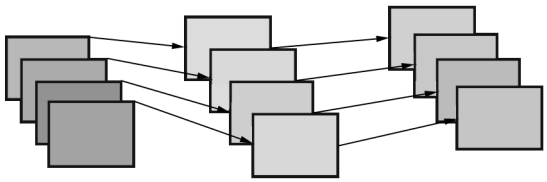

图 4　深度卷积运算过程

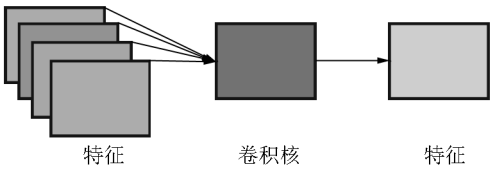

图 5　传统卷积运算过程

2.3　特征融合

随着 CNN 的发展,一系列从低级到高级的特征提取器可以从大规模数据中通过卷积运算的方式自动训练出来. 在高级特征中包含了更多高级信息,如语义等. 虽然在高级特征中也包含有少量的细节,但在低级特征中包含了更多的细节,如纹理、颜色等. DTFBlock 可以提取含有更多细节的低级特征,它通过将深度卷积提取的特征和传统卷积提取的特征进行融合,特征以 add 和 concat 方式进行融合.

add 融合是将深度卷积提取的特征 X_i 和传统卷积提取的特征 Y_i 组合而成,组合方法为

$$Z_{\text{add}} = \sum_{i=1}^{c} (X_i + Y_i) K_i.$$

add 方式融合的特征既包含了先验信息,又在传统卷积提取特征的基础上增强了通道间的信息.

concat融合是将深度卷积提取的特征 X_i 和传统卷积提取的特征 Y_i 进行拼接,拼接方法为

$$Z_{\text{concat}} = \sum_{i=1}^{c} X_i K_i + \sum_{i=1}^{c} Y_i K_{i+c}.$$

concat 特征融合保留了深度卷积提取的特征和传统卷积提取的特征.

2.4　损失函数

本文采用 ArcFace 损失函数和融合损失的方法. ArcFace 在训练样本时可以减小同类别样本之间距离和增大不同类别样本之间距离. 融合损失可以加快网络寻找到 DTFBlock 融合后的最优特征,在特征融合过程中一些重要的信息会丢失,因此可以通过融合损失得到最优的特征.

2.4.1　ArcFace　人脸识别是细粒度识别任务,每个人脸和人脸之间的相似度比不同物种之间的相似度更高,因此本文采用了 ArcFace 损失函数. ArcFace 损失函数的核心思想是扩大类间距离、减小类内距离. ArcFace 损失函数为

$$L_{\text{ArcFace}} = \frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{i=1}^{n} e^{s\cos\theta_{y_i}}},$$

其中 y_i 为第 i 个样本对应的标签;N 为总样本的数量;s 和 m 为超参数,分别表示常数缩放因子和用来控制余弦间隔的常数间隔项;θ_{y_i} 为特征 $\boldsymbol{x}_{j,i} \in \mathbf{R}^{128\times1}$ 经过归一化之后和 $\boldsymbol{W}_{j,i} \in \mathbf{R}^{128\times1}$ 经过归一化之后的点乘积,同时也代表特征 $\boldsymbol{x}_i$ 和对应权重 $\boldsymbol{W}_{j,i}$ 之间的角度.

$$\cos\theta_{yi} = \boldsymbol{x}_i / \|\boldsymbol{x}_i\| \times \boldsymbol{W}_j / \|\boldsymbol{W}_j\|,$$

其中 $\boldsymbol{x}_{j,i}$ 表示在第 j 个类中的第 i 个人脸样本提取到的特征值;$\boldsymbol{W}_{j,i}$ 表示在第 j 个类中的第 i 个人脸的权重值;$\|\boldsymbol{x}_i\|$ 表示归一化之后的特征;$\|\boldsymbol{W}_j\|$ 表示归一化之后的权重.

2.4.2　融合损失　在DTFBlock 中,深度卷积提取的特征不够丰富,因此需添加一个模块进行信息弥补,本文采用 2 个 3 × 3 的传统卷积组成的模块来提取该特征,从而弥补了深度卷积的通道信息. 因此,针对信息弥补定义融合损失,$\boldsymbol{f}_i$ 表示在输入 DTFBlock 中的特征,$\boldsymbol{f}_{fu}$ 表示融合之后的特征,则图像的融合损失函数为

$$L_{fu} = \|\boldsymbol{f}_i - \boldsymbol{f}_{fu}\|.$$

2.4.3　损失函数　综合上述的 ArcFace 损失和融合损失,本文的总损失函数为

$$L_{总} = L_{\text{ArcFace}} + \lambda L_{\text{fu}}, \tag{11}$$

其中 $\lambda(\leqslant 1)$ 表示超参数.

3 实验结果与分析

3.1 实验数据集

本文使用 CASIA-WebFace 数据集进行训练,LFW 数据集作为验证数据集. 在 CASIA-WebFace 数据集中包含了 10 575 个人的 494 414 幅图像.

LFW 数据集是人脸验证的公共基准数据集,其中包含 5 749 个人超过 13 000 幅人脸图像,每 1 幅人脸都有姓名,大约 1 680 个人有超过 2 幅人脸图像. LFW 数据集被广泛地用于人脸识别,本文用 LFW 数据集的标准协议(LFW-pairs)来验证实验结果,LFW 标准协议包含了总共 6 000 对人脸图像,其中 3000 对属于同一个人 2 幅人脸照片(正样本对),3 000 对属于不同的人每人 1 幅人脸照片(负样本对).

3.2 数据预处理

在 CASIA-WebFace 数据集中含有大量的类别,同时每个类别都包含了几十幅图像,但在这些图像中包含了较多的噪声,这些噪声有姿态、光照、多人脸等. 这些噪声若不处理则将导致网络学习的特征无法正确分类,因此需要对特征进行预处理. 在本次预处理过程中本文主要对多人脸和姿态等问题进行预处理,不对光照等噪声进行预处理. 在预处理过程中本文使用 MTCNN[17] 进行预处理,该方法使用了级联的方式对人脸进行检测,在检测过程中使用了人脸边框回归和面部关键点检测[18-20],将人脸姿态进行矫正和在多个人脸的图像中选用其中一个人脸进行预测.

本文预处理过程将所有的图像裁剪成 112 × 96 像素大小的图像.

3.3 实验设置

整个网络的训练和验证都是在 pytorch 框架下进行的. 在训练阶段本文使用 SGD 优化器对模型进行优化,动量设置为 0.9,设置权重衰减参数为 4×10^{-5},最后全局操作(GDConv)权重衰减参数设置为 4×10^{-4}. 本文将初始的学习率设置为0.1,并且在训练轮次为 36、52、58 时学习率除以 10. 本文模型一共训练 70 轮,用 ArcFace 损失函数[21] 作为目标函数优化整个网络,在 ArcFace 中的 s 设置为 32,m 设置为 0.5.

本文实验环境为 CPU Inter Core i7-9500,内存为 16 G,显卡类型为 NVIDIA GeForce GTX 2080SUPER,操作系统为 Ubuntu16.04.

人脸验证常用的模型评价指标是人脸验证的准确率,该准确率的计算方式为准确率 = 预测正确的样本数/总样本数.

使用 10 折交叉验证的方法验证本文的模型:首先得到每 1 折验证部分的准确率,即在该折中预测正确的数量与在该折中所有图像数量的比值,然后将得到的准确率的平均值作为本文的最终评价指标.

3.4 在 LFW 数据集上的结果与分析

表 1 显示了 MobileNetV1、ShuffleNet、MobileNetV2、MobileFaceNet、MobiFace[22]、AirFace[8] 以及本文模型在 LFW 测试数据集上的识别率. 本文将从批量大小、数据集、模型参数量 3 个方面进行比较. 当 MobileFaceNet 在 2 个大小为 5.8 M 的数据集上进行训练(见表 1)时,训练模型使用不同批量大小的实验结果表明:使用批量大的训练得到模型效果比使用批量小的训练得到的模型效果更好,但提升效果不明显(约提升了 0.03%).

由表 1 可知:本文使用的训练批量最小,批量大小为 128. MobileFaceNet、AirFace、WFaceNet[23] 训练的批量大小分别为 512、1 024 和 128,是本文模型训练批量大小的 4 倍、8 倍和 1 倍,但是本文模型在 LFW 数据集上测试效果比 MobileFaceNet、AirFace 和 WFaceNet 的都更好. 因此,在训练批量大小上,本文的模型效果是较优的. 从在不同练数据训集上的比较发现:MobileFaceNet 在不同数据集上的训练效果相差较大. MobileFaceNet 在 5.8 M 的数据集上的效果为 99.55%,而在 0.5 M 的数据集上的效果为99.28%,虽然在这 2 个数据集上的批量大小不同,但是从上述对批量大小分析中可知,批量大小对测试结果的影响程度并不是很大. 因此,这里忽略批量大小对测试结果带来的影响. MobileFaceNe 对这 2 个不同数据集进行训练,发现在 LFW 数据集上的测试结果相差 0.27%. 由此可以看出,相比批次大小对实验结果的影响,训练数据集对模型的测试结果的影响更大;同时这也表明,在使用大的数据集来训练模型的性能方面 DTFBNet 在 0.5 M 的数据集上效果为 99.40%,MobileFaceNet、MobileNetV2、AirFace、ShuffleNet、WFaceNet 在 0.5 M 的数据集上的效果分别提升了0.12%、0.82%、0.14%、0.70%、0.04%. 实验结果表明:DTFBNet 在相同的训练数据集上的效果有一定的提升,然而比使用较大训练数据集训练得到的网络性能更差. 接下来在模型参数量上进行比较,MobileNet 和 MobileNetV2 都在 0.5 M 的数据集上进行训练,批量大小也相同,批量为 512,MobileNet 模型的参数量大小为 3.20 M,MobileNetV2 模型的参数量大小为 2.10 M. MobileNet

模型在 LFW 数据集上的测试效果比 MobileNetV2 模型在 LFW 数据集上的测试效果更好,提升了 0.05%. MobileNet 模型的参数量比 MobileNetV2 模型的参数量增加了1.1 M. WFaceNet[23]在 LFW 数据集上的测试效果为 99.36%,比 MoileNetV2 的提高了 0.78%,并且参数量也比 MobileNetV2 的减少了 0.94 M. 这表明:在模型的参数量较小的情况下,模型的参数量越小,模型提升的性能越有限. DTFBNet 模型在 LFW 数据集上的测试效果为99.40%,参数量大小为 0.92 M. 在比 WFaceNet 参数量减少了 0.24 M的情况下,DTFBNet 模型的网络效果提升了 0.04%;在比 MobileFaceNet 参数量减少了 0.07 M 的情况下,DTFBNet 模型的网络效果提升了0.12%. 这些实验表明:DTFBNet 在参数量降低的情况下,DTFBNet 的性能有一点提升. 因此,本文提出的 DTFBNet 模型无论在训练批量大小、训练数据集上还是在模型参数量大小上都达到了较好的效果.

表 1　LFW 测试结果

人脸识别方法	图像数量/M	参数数量/M	Batch_Size	LFW 准确率/%
ArcFace	5.8	65.00	512	99.83
MobileFaceNet	5.8	0.99	160	99.52
MobileFaceNet	5.8	0.99	256	99.55
LMobileNetE	3.8	26.70	512	99.50
MobiFace	3.8	2.40	1 024	99.70
MobileFaceNet	0.5	0.99	512	99.28
AirFace	0.5	–	1 024	99.26
ShuffleNet(1×1,$g=3$)	0.5	0.83	512	98.70
MobileNet	0.5	3.20	512	98.63
MobileNetV2	0.5	2.10	512	98.58
WFaceNet	0.5	1.16	128	99.36
本文	0.5	0.92	128	99.40

3.5　消融实验

3.5.1　*λ 对实验结果的影响分析*　对融合损失函数与 ArcFace 损失函数的参数 λ 进行了实验,以判断 λ 的选取对实验结果的影响,本次实验模型有 DTFBlock 模块,同时 Conv 1×1 的输出通道数为 256. 从图 6 可知:当 λ 取 0.3 时实验可以取得最好的效果,若逐渐提升或降低 λ 的值则实验效果都会较差. 这表明融合损失对于模型的提升有局限性.

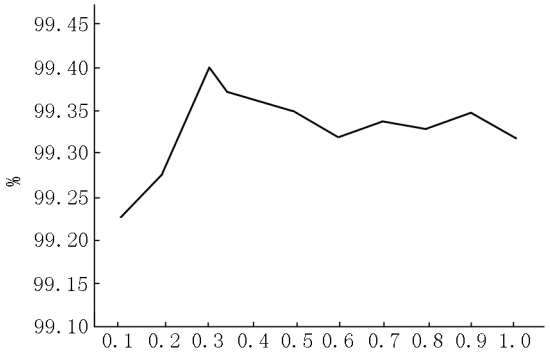

图 6　λ 对实验结果的影响分析

3.5.2　*DTFBlock 模块和损失函数对实验效果的影响分析*　本文提出了 DTFBlock 模块和融合损失函数,为了判断 DTFBlock 模块和融合损失函数在本文提出模型中的意义,设计了消融实验,实验结果如表 2 所示.

表 2　模块比较效果

DTFBlock	融合损失	LFW/%
×	×	99.21
√	×	99.38
√	√	99.40

在表 2 中,没有 DTFBlock 和融合损失的基础模型是 MobileFaceNet. 由表 2 可知:将深度卷积替换成 DTFBlock 会带来实验效果的提升,即在 LFW 数据集上的识别率提升了 0.17%. 这表明 DTFBlock 可以提取到含有更多细节的低级特征,即 DTFBlock 模块更有效. 随后,本文使用了融合损失函数(融合损失超参 λ 设置为 0.3),同时也使用 DTFBlock,在 LFW 数据集上的识别率提升了 0.02%,有较小幅度的提升. 这表明融合损失函数是有效的.

3.5.3　*Conv 1×1 输出通道数对实验结果的影响分析*　本文将 Conv 1×1 模块的输出通道数和 Linear GDConv 7×6 的输入通道数进行改进. 通道数的变化如表 3 所示,其中模型使用了 DTFBlock 和融合损失函数. 由表 3 可知:当 Conv 1×1 中的输出通道数为输入通道数的 2 倍时网络的性能最好,而当输出通道数为输入通道数的 1 倍时网络的性能最差. 这表明在通道倍数为 1 时网络提取的特征比在其他通道倍数时网络提取的特征差. 通道倍数为 2 的网络可以提取更加有效的特征,在通道倍数为 3 和通道

倍数为 4 时的网络性能次于在通道倍数为 2 时的网络. 这表明:在通道倍数为 3 和通道倍数为 4 的网络提取的特征中含有冗余特征,从而导致最终网络的性能不能达到最佳. 在通道倍数为 2 的网络中由于 DTFBlock 提取含有许多细节的特征,所以在 Conv 1×1模块中提取更具有分类能力的高级特征.

表 3 Conv2 输出通道数倍数

倍数	LFW/%
1	99.18
2	99.40
3	99.30
4	99.26

3.6 可视化

为了展示 DTFBlock 可以提取含有更多细节的低级特征,本文在 LFW 数据集上可视化 DTFBlock 提取的特征. 本文从 LFW 数据集中随机选取了 5 幅图像进行可视化(见图 7). 第 1 层是 MobileFaceNet 中深度可分离卷积提取出来的特征可视化图,第 2 层是 DTFBNet 中 DTFBlock 提取出来的特征可视化图. 从图 7 可以看出:DTFBlock 提取出来的特征可以较为清晰地看出每个人的轮廓和一些纹理. 这表明 DTFBlock 可以提取人脸的含有丰富细节的低级特征. 而 MobileFaceNet 中的深度可分离卷积提取的低级特征含有的细节信息不丰富,看不出完整的纹理结构;提取的特征有一定的信息缺失,这会导致后续网络中判别器会出现较多的预测错误. 含有丰富细节的低级特征是提取有判别性高级语义特征的重要前提. 因此,本文提出的 DTFBlock 可以有效地提取含有丰富细节的低级特征.

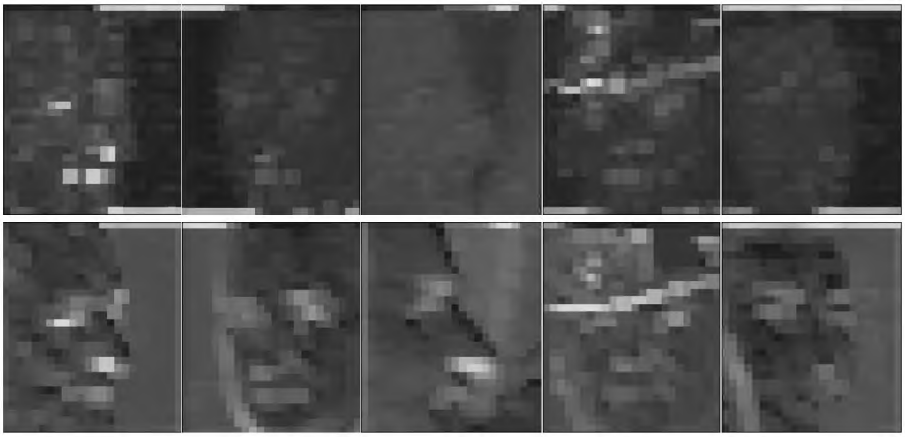

图 7 可视化

4 总结

本文提出了有效的轻量级人脸识别网络模型 DTFBNet. DTFBNet 是基于 MobileFaceNet,将 MobileFaceNet 中的深度卷积替换成本文提出的 DTFBlock. DTFBlock 考虑深度卷积提取的特征没有计算通道信息,因此出现了含有细节信息不丰富的现象. DTFBlock 将深度卷积提取的特征和 2 个传统卷积提取的特征进行融合. DTFBNet 相比 MobileFaceNet 在 Conv 1×1 上减少了输出通道数量,由原来的 512 减少到 128,这导致 Linear GDConv 7×6 的输入通道数也相应地减少. 但是,DTFBlock 引入了深度卷积增多的参数量少于 Conv 1×1 和 Linear GDConv 7×6 减少的参数量,因此,整个模型的参数量减少,但是整体模型在 LFW 数据集上的准确率有所提升. 这表明 DTFBlock 模型提取了更丰富的特征,更加有利于分类器进行分类. 本文提出的 DTFBNet 的参数量只有0.92 M,在 LFW 数据集上的效果也有所提升. 这表明本文的模型可以用于在移动设备和嵌入式设备上的实时人脸验证.

5 参考文献

[1] DENG Jiankang,GUO Jia,ZHANG Debing,et al. Lightweight face recognition challenge [EB/OL]. [2021-11-17]. https://ieeexplore.ieee.org/document/9022288.

[2] MÉNDEZ-VÁZQUEZ H,MARTÍNEZ-DÍAZ Y,NICOLÁS-DÍAZ M,et al. Bench-marking lightweight face architectures on specific face recognition scenarios [J]. Artificial Intelligence Review,2021,54(8):6201-6244.

[3] SANDLER M,HOWARD A,ZHU Menglong,et al. MobileNet V2:inverted residuals and linear bottlenecks [EB/OL]. [2021-11-17]. https://arxiv.org/pdf/1801.04381v3.pdf.

[4] MA Ningning,ZHANG Xiangyu,ZHENG Haitao,et al. ShuffleNet V2:practical guidelines for efficient CNN architecture design [EB/OL]. [2021-11-17]. https://arxiv.org/pdf/1807.11164.pdf.

[5] ZHANG Qian,LI Jianjun,YAO Meng,et al. VarGNet:variable group convolutional neural network for efficient em-

bedded computing [EB/OL]. [2021-11-17]. https://arxiv.org/abs/1907.05653.
[6] CHEN Sheng,LIU Yang,GAO Xiang,et al. MobileFaceNets: efficient CNNs for accurate real-time face verification on mobile devices [EB/OL]. [2021-11-17]. https://arxiv.org/ftp/arxiv/papers/1804/1804.07573.pdf.
[7] HOWARD A G,ZHU Menglong,CHEN Bo,et al. MobileNets:efficient convolutional neural networks for mobile vision applications [EB/OL]. [2021-11-17]. https://arxiv.org/pdf/1704.04861.pdf.
[8] LI Xianyang, WANG Feng, HU Qinghao, et al. AirFace: lightweight and efficient model for face recognition [EB/OL]. [2021-11-17]. https://www.researchgate.net/publication/339768456_AirFaceLightweight_and_Efficient_Model_for_Face_Recognition.
[9] MARTÍNEZ-DÍAZ Y,LUEVANO L S,MÉNDEZ-VÁZQUEZ H,et al. ShuffleFaceNet:a lightweight face architecture for efficient and highly-accurate face recognition [EB/OL]. [2021-11-17]. https://ieeexplore.ieee.org/document/902220.
[10] YAN Mengjia,ZHAO Mengao,XU Zining,et al. VargFaceNet: an efficient variable group convolutional neural network for lightweight face recognition [EB/OL]. [2021-11-17]. https://ieeexplore.ieee.org/document/9022149/.
[11] LI Yuancheng, WANG Yimeng, LI Daoxing. Privacy-preserving lightweight face recognition [J]. Neurocomputing, 2019,363:212-222.
[12] HUANG Gao,LIU Shichen,VAL DER MAATEN L,et al. CondenseNet:an efficient DenseNet using learned group convolutions [EB/OL]. [2021-11-17]. https://arxiv.org/pdf/1711.09224.pdf.
[13] SUN Ke,LI Mingjie,LIU Dong,et al. IGCV3:interleaved low-rank group convolutions for efficient deep neural networks [EB/OL]. [2021-11-17]. https://arxiv.org/pdf/1806.00178.pdf.
[14] WU Bichen,ALVIN W,YUE Xiangyu,et al. Shift:a zero FLOP,zero parameter alternative to spatial convolutions [EB/OL]. [2021-11-17]. https://arxiv.org/pdf/1711.08141.pdf.
[15] HUANG G B,RAMESH M,BERG T,et al. Labeled faces in the wild:a database for studying face recognition in unconstrained environments [EB/OL]. [2021-11-17]. http://cs.brown.edu/courses/cs143/2011/proj4/papers/lfw.pdf.
[16] CHOLLET F. Xception:deep learning with depthwise separable convolutions [EB/OL]. [2021-11-17]. https://arxiv.org/pdf/1610.02357.pdf.
[17] ZHANG Kaipeng,ZHANG Zhanpeng,LI Zhifeng,et al. Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters,2016,23(10):1499-1503.
[18] VO D M,LEE S W. Robust face recognition via hierarchical collaborative representation [J]. Information Sciences, 2018,432:332-346.
[19] 杜星悦,董洪伟,杨振. 基于深度网络的人脸区域分割方法 [J]. 计算机工程与应用,2019,55(8):171-174.
[20] 王小玉,韩昌林,胡鑫豪. 加权特征融合的密集连接网络人脸识别算法 [J]. 计算机科学与探索,2019,13(7):1195-1205.
[21] DENG Jiankang,GUO Jia,XUE Niannan,et al. ArcFace: additive angular margin loss for deep face recognition [EB/OL]. [2021-11-17]. https://arxiv.org/pdf/1801.07698.pdf.
[22] DUONG C N,QUACH K G,JALATA I,et al. MobiFace:a lightweight deep learning face recognition on mobile devices [EB/OL]. [2021-11-17]. https://arxiv.org/pdf/1811.11080.pdf.
[23] WANG Xiaobo,FU Tianyuu,LIAO Shengcai,et al. Exclusivity-consistency regularized knowledge distillation for face recognition [EB/OL]. [2021-11-17]. https://link.springer.com/chapter/10.1007/978-3-030-58586-0_20.

DTFBNet:the New Lightweight Face Recognition Method for Smart Terminals

YE Jihua,GUO Feng,LI Xin,JIANG Lu,JIANG Aiwen
(School of Computer and Information Engineering,Jiangxi Normal University,Nanchang Jiangxi 330022,China)

Abstract:There are many solutions to the problem of insufficient intelligent terminal resources,which are dependent on sample data and the number of parameters. The depthwise convolution and traditional convolution fusion block (DTFBlock) is proposed. Hence,an improved method DTFBNet based on MobileFaceNet is proposed. The DTFBNet proposed in the paper has a smaller number of parameters and better network results. Experiments on face recognition datasets CASIA-Webface and LFW show that the highest accuracy rate of the algorithm proposed in this paper reaches 99.40%,which is already a competitive classification accuracy for the same parameter amount.

Key words:DTFBNet;DTFBlock;fusion loss;lightweight;face recognition

(责任编辑:冉小晓)