

DOI:10.19651/j.cnki.emt.2209356

基于 Ghostnet 轻量级人脸识别算法研究*

赵 锋¹ 张 鹏¹ 张 冉²

(1. 中北大学仪器与电子学院 太原 030051; 2 中北大学大数据学院 太原 030051)

摘要: 为了提高人脸识别在嵌入式设备中的识别精度和速度,提出一种基于 Ghostnet 轻量级人脸识别算法—Ghostfacenet。首先,通过预设卷积生成固定数目的内在特征;针对卷积运算计算消耗大的问题,使用计算成本低廉的线性操作代替卷积运算,产生一系列与内在特征相关联的特征信息;其次,基于 Ghostnet 中的 Ghost 模块以及深度可分离卷积设计出 Ghostfacenet-Bottleneck,并且由其构建出 Ghostfacenet 轻量级卷积神经网络;最后,联合 Softmax 损失函数和 Arcface 损失函数进一步增加人脸类内紧凑性以及类间差异,同时使得轻量级模型有更好的收敛性以及泛化能力。实验结果表明,Ghostfacenet 在嵌入式设备的识别速度分别是 Resnet50、Efficientnet、MobilenetV2 和 Mobilefacenet 的 11.08 倍、8.57 倍、2.75 倍和 2.82 倍。在不显著降低识别性能同时能够显著提高运行效率,非常适用于资源有限的嵌入式设备中。

关键词: 深度学习;损失函数;轻量级卷积神经网络;人脸识别;嵌入式系统

中图分类号: TP391.41 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Research on Ghostnet-based lightweight face recognition algorithm

Zhao Feng¹ Zhang Peng¹ Zhang Ran²

(1. School of Instrument and Electronic, North University of China, Taiyuan 030051, China;

2. School of Date Science and Technology, North University of China, Taiyuan 030051, China)

Abstract: In order to improve the recognition accuracy and recognition speed of face recognition in embedded devices, a Ghostnet-based lightweight face recognition algorithm called Ghostfacenet is proposed. Firstly, a fixed number of intrinsic features are generated by pre-determined convolution. To address the problem of computationally intensive convolutional operations, linear operations with low computational cost are used instead of convolutional operations to generate a series of feature information associated with intrinsic features. Secondly, the Ghostfacenet-Bottleneck is designed based on the Ghost module in Ghostnet and the depthwise separable convolution. And the Ghostfacenet lightweight convolutional neural network is constructed from Ghostfacenet-Bottleneck. Finally, the Arcface loss function and the Airface loss function are combined to further increase the intra-class compactness of faces as well as inter-class differences. It also allows for better convergence and generalization capabilities of lightweight models. The experimental results show that Ghostfacenet is 11.08 times, 8.57 times, 2.75 times and 2.82 times faster than Resnet50, Efficientnet, MobilenetV2 and Mobilefacenet respectively in embedded devices. This is a significant improvement in operational efficiency without a significant reduction in recognition performance and is ideal for embedded devices with limited resources.

Keywords: deep learning; loss function; lightweight convolutional neural networks; face recognition; embedded system

0 引 言

人脸识别是应用最广泛的生物识别技术之一,被广泛应用于军事、金融、安防以及日常生活等领域。同时深度卷积神经网络(deep convolutional neural network, DCNN)的

发展使得人脸识别取得了巨大的成功。传统的人脸识别方法已经不能满足当前复杂环境对高性能和实时性的要求^[1-4]。因此,高精度人脸识别算法都使用了非常复杂的卷积神经网络,例如 Vggface2^[5]、Resnet^[6]、LS-CNN^[7]等。然而这些复杂的卷积神经网络带来了巨大的参数量以及计

收稿日期:2022-03-21

* 基金项目:技术领域基金(2021-JCJQ-JJ-0726)项目资助

算量,因此高性能模型通常需要大内存以及性能强劲的 GPU 进行加速。例如 ResNet-50 需要 0.26 亿参数和 41 亿浮点计算量(FLOPs)去处理大小为 224×224 的图片。因此,这种大型的复杂模型难以部署到资源受限的嵌入式设备或者移动设备中。于是在不明显降低准确度的前提下设计一种适用于人脸识别的高效轻量级卷积神经网络已经成为了一种趋势。

为了解决这个问题,一些专用于人脸识别的高效轻量级卷积神经网络相继被提出,如 Mobilefacenet^[8]、Shufflefacenet^[9]、Vargfacenet^[10]等。虽然这些算法在准确度和速度两个方面取得了较好的平衡,但是都没有考虑卷积过程中存在大量的特征映射冗余问题。因此,针对以上问题,本文提出了一种 Ghostfacenet 高效轻量级卷积神经网络。该网络结合现有轻量级卷积神经网络的优点以及应用 ShufflenetV2^[11]中的指导原则,设计出在准确性和速度均衡的轻量级卷积神经网络。Ghostfacenet 主要参考 Ghostnet^[12]中提出的 Ghost 模块进行网络的构建,并且在设计过程中应用了 ShufflenetV2 的指导原则,同时使用“通道混洗”操作对不同支路信息进行通信,增强特征信息的表达力。在训练过程中,为了缓解轻量级卷积神经网络在使用当前主流损失函数出现损失值不能收敛到很小值这一情况,我们提出了 Arcface^[13]与 Softmax 联合训练策略从而缓解这一问题同时能够提升模型的泛化能力。

本文的其余部分安排如下。第 2 节介绍了轻量级卷积神经网络以及专用人脸识别的轻量级卷积神经网络;第 3 节详细介绍了 Ghostfacenet 的模型结构以及联合训练的策略;第 4 节描述了实验细节以及结果分析;第 5 节对全文进行了总结以及未来展望。

1 相关工作

在设计轻量级卷积神经网络时,通常是在准确性和速度两个方面寻求最佳平衡点。而对于人脸识别的轻量级卷积神经网络通常是由分类任务中迁移过来的。

1.1 轻量级卷积神经网络

设计轻量级卷积神经网络的一个重要方法就是直接构建高效且成本低的模块,通过模块的堆叠实现一个参数和计算量都很小的卷积神经网络。

Mobilenet^[14-16]系列引入深度可分离卷积作为传统卷积的有效代替,在准确度和效率方面都实现了很好的平衡;在后续的改进中引入了倒置残差结构以及线性瓶颈层并且实现了更加高效的性能。Shufflenet^[17]利用逐点组卷积和“通道混洗”操作在保证准确性的同时极大地减少了计算成本;ShufflenetV2^[11]充分考虑了影响模型速度的其他因素,如内存访问成本(MAC)以及平台特性等,提出了设计轻量级卷积神经网络的有效准则并且在准确性和速度中实现了更好的均衡。Vargnet^[18]通过固定组卷积中的通道数目在嵌入式设备中实现了较好的实用价值。Ghostnet^[12]以一系

列廉价的线性操作生成更多的特征,在与其它轻量级模型具有相近计算成本的情况下具有更高的性能。

1.2 专用人脸识别的轻量级卷积神经网络

虽然人脸识别取得了巨大的成功,但是高性能模型庞大的计算量以及参数量使得其难以部署在嵌入式设备或者移动设备中。因此,为人脸识别设计高效轻量级卷积神经网络是十分迫切的。

Light CNN^[19]为了减少权重数量并从输入中选择噪声较小的权重,引入了最大特征映射函数(MFM)。但是其仍旧存在参数量和计算量大等问题,难以部署到嵌入式或者移动设备中。Mobilefacenet^[8]在 MobileNetV2 的基础上进行设计并且将全局平均池化替换为全局深度卷积层,在移动设备中实现了很好的性能表现。Shufflefacenet^[9]在 ShuffleNetV2 基础上进行改进并且使用 PReLU 代替了 ReLU,在保证高性能的同时极大提高了运行速度。Mobiface^[20]为了最大化最后特征向量信息以及减少计算消耗,采用带有扩展层的残差瓶颈层进行网络的构建,同时采用了快速下采样策略去快速降低特征维度。Vargfacenet^[10]使用可变组卷积进行网络设计,该网络在降低计算成本和网络参数同时也可以解决小计算成本和模块内部计算密度不平衡等问题。虽然以上工作都在准确性和速度两个方面取得了较好的平衡,但是,大部分工作都没有考虑到卷积过程会产生很多冗余映射。这些冗余映射仍旧需要大量的卷积运算,这将会带来额外的计算开销。

2 主要工作

2.1 Ghostfacenet 网络设计

由于嵌入式等设备的资源限制使得高性能的模型难以在其上部署,因此研究设计高效的轻量级卷积神经网络十分必要。如图 1 所示,在卷积过程中会出现许多冗余映射(例如第 1 行第 1 列和第 3 行第 1 列、第 1 行第 3 列和第 2 行第 5 列、第 4 行第 5 列和第 5 行第 3 列),而这些冗余映射消耗的计算资源是十分巨大的。因此采用廉价的操作代替卷积操作去产生这些相似的特征是减少计算消耗的有效途径。本文参考 GhostNet 中的 Ghost 模块以廉价的操作生成更多特征同时以 ShuffleNetV2 指导原则进行设计轻量级模型。

1) Ghostfacenet-Bottleneck 设计

在 Ghostnet 中的 Ghost 模块中,输入特征首先通过预设的逐点卷积生成固定数目 c' 的内在特征 $Y \in \mathbb{R}^{w \times h \times c'}$;

$$Y = X * f \quad (1)$$

其中, $X \in \mathbb{R}^{w \times h \times c}$ 为输入原始特征, $f \in \mathbb{R}^{k \times k \times c \times c'}$ 为卷积核。同时为了简便处理,忽略了偏置项。接着由 c' 个内在特征通过一系列线性操作产生与之相关联的 s 个特征信息 y_{ij} :

$$y_{ij} = \phi_{ij}(y'_i), i = 1, 2, \dots, m, j = 1, 2, \dots, s \quad (2)$$

其中, y'_i 是固有特征 Y 中的第 i 个特征, ϕ_{ij} 是生成第 j

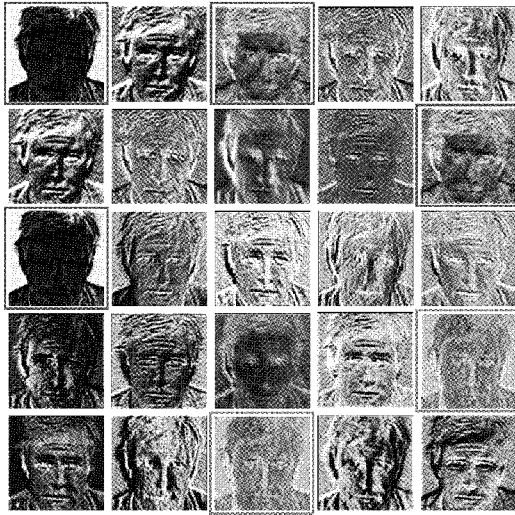


图 1 VGG-16 第五层卷积层产生的部分可视化特征图

个关联特征的线性操作。通过线性操作产生的特征信息 y_{ij} 与内在特征 Y 连接在一起并输出特征信息。Ghost 模块操作流程如图 2 所示。

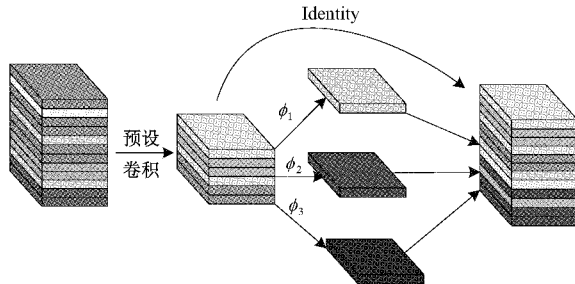


图 2 Ghost 模块

Ghostfacenet-Bottleneck 结构设计如图 3 所示。在 $\text{stride}=1$ 的结构中,输入特征首先经过预设卷积模块生成固定数目的内在特征并且分为了两个分支。其中一个分支包含了线性操作以及深度可分离卷积 DWConv 以及 1×1 的卷积操作。并且这三个部分的输入输出通道数目都是相同的,根据 ShuffleNetV2 指导准则可以得知这样的操作可以减少内存访问次数。之后通过 Concat 操作连接两个支路,最后通过“通道混淆”操作使得两个支路通道信息进行通信,以便提高模型的准确率。

对于空间下采样,设计了 $\text{stride}=2$ 的结构。与 $\text{stride}=1$ 结构不同的是,我们在另外一个分支增加了 DWConv 和 1×1 的卷积操作,并且 DWConv 的步长为 2。两个分支输出的特征通过 Concat 操作连接起来进一步增加输出特征的维度,最后通过“通道混淆”操作。并且在两种结构中使用了批量归一化(BN)操作和 ReLU 非线性操作。同时,根据 ShuffleNetV2 中的指导准则,两种结构中的 1×1 卷积操作都不是在组卷积,在连接两个支路特征信息时为了减少内存访问次数没有使用 Add 操作而是选择

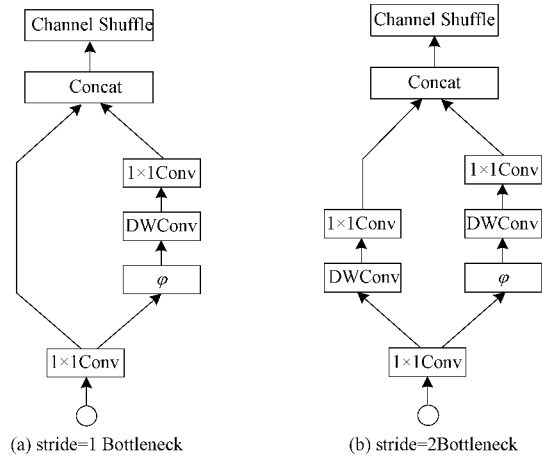


图 3 Ghostfacenet-Bottleneck 结构设计

了 Concat 操作。

2) Ghostfacenet 网络设计

基于 Ghostfacenet-Bottleneck,构建了 Ghostfacenet 轻量级卷积神经网络,具体配置参数如表 1 所示。该网络主要由 Ghostfacenet-Bottleneck 的堆叠组成,第一层是标准的卷积用来进行快速下采样,降低输入特征的维度。卷积神经网络的宽度对于人脸识别是非常重要的,因此后续操作中使用了一系列的 Ghostfacenet-Bottleneck 用来提取丰富的人脸特征以及增加特征的通道数日。最后我们使用线性全局深度卷积操作与线性 1×1 卷积操作输出 128 维度特征向量用来进行人脸识别。

表 1 Ghostfacenet 网络参数配置表

输入	操作	输出通道	重复次数	步长
$112^2 \times 3$	Conv 3×3	64	1	2
$56^2 \times 64$	G-Bottleneck	64	1	2
$28^2 \times 64$	G-Bottleneck	64	4	1
$28^2 \times 64$	G-Bottleneck	128	1	2
$14^2 \times 128$	G-Bottleneck	128	5	1
$14^2 \times 128$	G-Bottleneck	256	1	2
$7^2 \times 256$	G-Bottleneck	256	4	1
$7^2 \times 256$	Conv 1×1	1 024	1	1
$7^2 \times 1 024$	GDConv 7×7	1 024	1	1
$1^2 \times 1 024$	Linear 1×1	128	1	1

2.2 联合损失函数

损失函数的设计在人脸识别中有着十分重要的地位。最广泛应用的 Softmax 损失函数如下:

$$L_{Softmax} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{w_{x_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{w_j^T x_i + b_j}} \quad (3)$$

其中, x_i 是第 y_i 类的第 i 个样本; W_j 是权重的第 j 列, b_j 是偏置项; N 和 n 分别是批次大小和类别总数。由

于 Softmax 没有明确的优化类内方差,因此会导致人脸识别算法性能下降。为了增加了类内紧凑性以及类间差异性,Arctface 在样本 x_i 和样本中心 W_{y_i} 中增加了加性角裕度 m 。Arctface 公式如下:

$$L_{Arctface} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cos(\theta_{y_i} - m)}}{e^{s \cos(\theta_{y_i} - m)} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}} \quad (4)$$

其中, θ_{y_i} 是样本 x_i 与权重 W_{y_i} 之间的夹角,即 $W_{y_i}^T x_i = \|W_{y_i}\| \|x_i\| \cos \theta_{y_i}$; s 是超参数。但是,当直接使用 Arctface 损失函数训练轻量级卷积神经网络时候,会出现损失值发散或者损失值不能收敛到很小的值,影响模型的泛化能力。因此,采取先用 Softmax 损失函数训练模型,在此基础上用 Arctface 去再次训练模型,进一步缩小类内方差,扩大类间方差。

3 实验与结果分析

为了验证本文算法的有效性,选取 MS1MV3^[21]作为训练数据集, LFW^[22]、CPLFW^[23]、CFP^[24]作为测试集。首先分析了超参数对于模型性能的影响,之后与近年提出的轻量级模型进行了性能对比并且在 IJBB/C^[25]数据集上进行性能测试。最后,在嵌入式系统 AX7020 中进行了单张图片识别的速度测试。

3.1 实验参数设置

Ghostfacenet 由 Pytorch 实现。批次大小设置为 256; 初始学习率设置为 0.1 并且在每次完整的迭代结束后都除以 10; momentum 动量设置为 0.9, 权重衰减设置为 5×10^{-4} ; 训练设备为搭载一块 NVIDIA 3090 (24 G) GPU 工作站。

表 3 不同轻量级模型性能比较

方法	LFW/%	CPLFW/%	CFP_FF/%	CFP_FP/%	PARAM/M	FLOPs/G
Resnet50 ^[6]	99.64	90.57	99.63	94.94	40.29	2.19
Efficientnet ^[26]	99.53	90.92	99.50	96.32	6.58	1.14
MobilenetV2 ^[16]	99.55	89.34	99.48	93.17	2.26	0.43
Mobilefacenet ^[8]	99.53	90.34	99.55	95.26	1.0	0.45
MobilefacenetV1 ^[27]	99.40	87.17	99.50	95.80	—	—
Shufflefacenet ^[9]	99.70	88.50	99.60	96.30	2.60	0.58
VarGfacenet ^[10]	99.70	88.55	99.50	96.90	—	—
Ghostfacenet	99.53	89.13	99.09	94.21	0.85	0.15

与复杂网络以及近年流行的轻量级网络相比, Ghostfacenet 的性能有所下降,尤其是在包含姿态变换的测试集中。Ghostfacenet 的性能相比于表现最好的轻量级模型在 LFW、CPLFW、CFP_FF 和 CFP_FP 数据集分别下降了 0.17%、1.21%、0.51% 和 2.69%。在 LFW 和 CFP_FF 测试集中, Ghostfacenet 性能下降并不明显。但是对于跨姿态测试集中, Ghostfacenet 性能下降较为明显的主要原因是姿态变换需要更多的通道去表达丰富的特征信息,

3.2 损失函数超参数设置实验

首先,使用 Softmax 对模型进行了训练。其次,我们在 Softmax 预训练基础上研究了 Arctface 超参数 s 、 m 对于模型的影响。LFW 测试集准确率如表 2 所示。

表 2 Arctface 中不同的 s 、 m 对模型性能的影响

s	m	准确率/%	s	m	准确率/%
16	0.25	99.45	32	0.10	99.52
25	0.25	99.49	32	0.15	99.53
32	0.25	99.50	32	0.20	99.52
64	0.25	99.48	32	0.25	99.50
75	0.25	99.48	32	0.30	99.52
80	0.25	99.45	32	0.35	99.50

如表 2 实验结果所示,发现针对所提出的 Ghostfacenet,当 m 值固定时候,模型性能随着 s 取值变换而变换。当 s 取值为 32 时候模型性能达到最佳性能。因此我们选择 0,当 m 值变换时候,模型性能在 $m = 0.15$ 时稍好于其它情况。这是因为针对 Ghostfacenet,太大的角裕度会使得网络训练变得十分困难,而太小的角裕度会使得类内差异性变大,导致模型性能变差。因此,我们选取 $s = 32$ 、 $m = 0.15$ 作为 Arctface 的超参数。

3.3 轻量级人脸识别算法对比实验

比较了近年来流行的卷积神经网络,例如 Resnet50、Efficientnet、MobilenetV2、Mobilefacenet、MobilefacenetV1、Shufflefacenet 和 VarGfacenet。除了比较近年提出的轻量级网络,我们还比较了应用十分广泛的复杂的网络 Resnet50 和 Efficientnet。模型性能对比的实验结果如表 3 所示。

而 Ghostfacenet 中通道数目相比于其他模型较少不足以表达姿态变换中所需的特征信息。然而, Ghostfacenet 的参数量和计算量少于近年流行的轻量级模型,尤其是 Ghostfacenet 的计算量比 Mobilefacenet 少了 3 倍,比 Shufflefacenet 少了大约 3.87 倍。虽然 Ghostfacenet 稍微损失了部分性能,但是性能下降不显著,可以满足人脸识别技术对性能的要求。同时 Ghostfacenet 极大的减少了模型的参数和计算量,使得其能够方便的部署在资源受限

的嵌入式或者移动设备中。

3.4 IJB/C 数据集性能测试实验

IJB-B 数据集包含 1 845 个对象的 11 754 张人脸图像、55 026 个视频帧、7 011 个视频和 10 044 张非人脸图像。IJB-C 数据集包含 3 531 个目标的 21 294 张人脸图像和 10 040 张非人脸图像。给出了 Ghostfacenet 与其它轻量级模型在 IJB/C 测试集上进行了性能比较。Ghostfacenet 的 TAR (@ FAR = 1×10^{-5} , 1×10^{-4} , 1×10^{-3}) 与其它模型比较结果如图 4 所示。

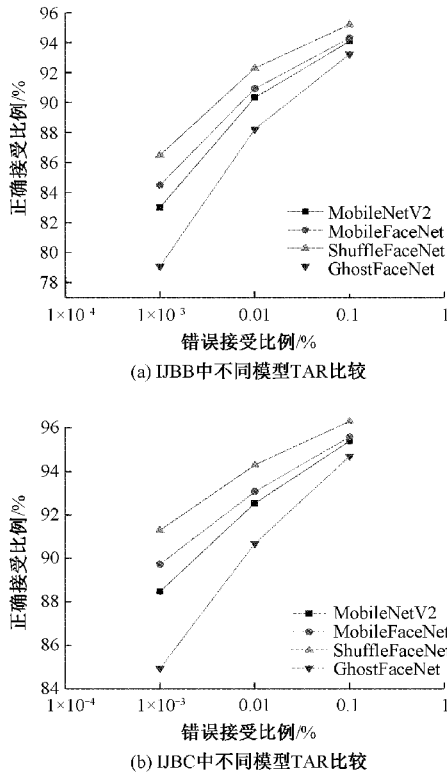


图 4 不同 FAR 下的轻量级模型的 TAR 比较

由图 4 可知,其他轻量级模型性能是优于 Ghostfacenet,这是因为 Ghostfacenet 中通过预设卷积生成内在特征,内在特征通过廉价的线性操作产生与之相关俩的特征。虽然在这一过程中可以减少卷积操作产生的大量运算,但是也不可避免的降低了模型的性能。

除此之外,还给出了 Ghostfacenet 在 IJB/C 测试集中的受试者工作曲线(receiver operating characteristic curve,ROC),如图 5 所示。ROC 曲线代表着对同一信号刺激的不同的感受性。ROC 曲线以负正类率(false positive rate)为横坐标,以真正类率(ture positive rate)为纵坐标。FPR 和 TPR 计算公式如下:

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

其中,FP 代表假正类, TN 代表真负类, TP 代表真

正类, FN 代表假负类。ROC 曲线下方的面积是 area under curve(AUC)。AUC 的取值范围为 0.1~1 并且 AUC 越大代表模型性能越好。如图 5 所示,从 Ghostfacenet 在 IJB/C 中的 ROC 曲线可以看出, $FPR = 1 \times 10^{-5}$ 时, IJB-B 中 TPR 接近 0.8, IJB-C 中 TPR 接近 0.85。并且 Ghostfacenet 在 IJB/C 中的 AUC 分别为 99.5105% 和 99.6082%。ROC 曲线结构说明了 Ghostfacenet 性能可以满足人脸识别技术的性能要求。

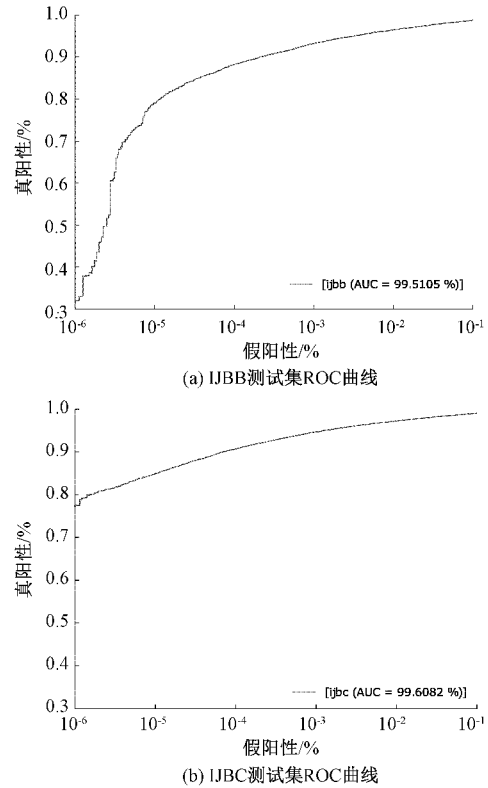


图 5 IJB/C 测试集上的 ROC 曲线

3.5 硬件测试实验

1) 嵌入式软硬件系统介绍

嵌入式硬件系统主要由 XILINX ZYNQ7000 开发平台的 AX7020 开发板、屏幕以及 USB 摄像头构成,如图 6 所示。其中嵌入式平台 AX7020 采用 ARM+FPGA SOC 技术,将 ARM Cortex-A9 与 FPGA 可编程逻辑集成在一颗芯片上,拥有两片大容量得 DDR3(共 8 Gbit)以及 256 Mbit deIQSPI FLASH,同时具有丰富得硬件资源和外围接口。对于嵌入式软件部分,利用 Vivado 构建该硬件系统的专属硬件工程。Petalinux 使用 Vivado 生成得硬件描述文件定制专属的 Linux 内核和 Uboot。其次,为了降低配置软件环境的难度,使用 buildroot 构建根文件系统并且将所需的 Qt5、Opencv3.4 等软件环境一同交叉编译。同时,为了部署模型方便,我们使用 NCNN 进行 Pytorch 模型转换。最终,轻量级的人脸识别验证程序由 Qt5 进行编写。其次,我们还比较了在计算机平台上速度的对比。

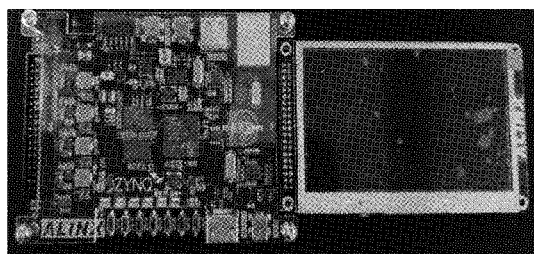


图 6 嵌入式系统硬件

2) 硬件系统测试实验

分别在计算机平台(CPU 为 i9-12900K)和嵌入式平台下进了 1:1 的速度测试,实验结果分别如表 4 和 5 所示。除此之外,还在 PC 端进行精度测试实验,测试数据集为 LFW 数据集。

表 4 PC 端轻量级模型性能测试

方法	准确度/%	速度/ms
Resnet50 ^[6]	99.64	—
Efficientnet ^[26]	99.53	36.50
MobilenetV2 ^[16]	99.55	25.60
Mobilefacenet ^[15]	99.53	23.67
Ghostfacenet	99.53	15.30

在 PC 端,Ghostfacenet 准确度表现稍有下降,但是其单张识别速度为 15.3 ms,分别是比 Efficientnet、MobilenetV2 和 Mobilefacenet 快 2.39 倍、1.67 倍和 1.55 倍。因此在不明显降低准确率的情况下,实现了 PC 端的快速实时识别,保证了人脸识别技术对精度的要求。

将 Ghostfacenet 模型转换为适合于嵌入式平台模型时候,我们并没有采取模型量化操作,模型大小未发生变化。因此,模型准确度性能几乎不变,在嵌入式设备中的测试实验主要进行速度对比实验,实验结果如表 5 所示。

表 5 嵌入式平台下轻量级模型速度测试

方法	嵌入式平台/s
Resnet50 ^[6]	3.09
Efficientnet ^[26]	2.40
MobilenetV2 ^[16]	0.77
Mobilefacenet ^[15]	0.79
Ghostfacenet	0.28

在嵌入式设备中,Ghostfacenet 的识别速度为 0.28 s,分别比 Resnet50、Efficientnet、MobilenetV2 和 Mobilefacenet 快 11.04、8.57、2.75 和 2.82 倍。因此,Ghostfacenet 在嵌入式设备中的速度表现优于其他的轻量级模型,对部署的嵌入式设备资源要求小,适合于部署在嵌入式设备中。同时其性能下降不显著可以满足人脸识别技术对精度的要求。

4 结 论

提出了一种高效的适用于嵌入式或者移动设备的轻量级卷积神经网络,其主要通过预设卷积产生内在特征,内在特征通过廉价线性操作产生更多的与之相关联的特征信息,进而减少卷积带来的大量计算。除此之外,我们利用 ShufflenetV2 指导原则进行网络设计,减少了内存访问次数,直接提高了运行时间。我们采取 Softmax 损失函数和 Arcface 损失函数联合训练模型,进一步提高了模型的泛化能力。与近年提出的轻量级人脸识别算法相比,虽然 Ghostfacenet 的性能稍有下降,但是其参数数量和计算量都大幅度少于其他算法,在 PC 端和嵌入式平台中的单张识别速度都明显优于其他的轻量级模型。但是,本研究仍旧有可以提高的空间,在未来的工作中,可以适当增加网络的层数以及添加注意力机制来提高模型的性能。除此之外,在嵌入式系统中,并未大量使用 FPGA 的资源。在未来的工作中,可以使用 FPGA 中的资源进一步提高模型的运行速度。同时,这也是人脸识别算法部署在嵌入式设备的重要研究方向。

参考文献

- [1] 马博宇,尉寅玮. 基于 AdaBoost 算法的人脸识别系统的研究与实现[J]. 仪器仪表学报, 2016, 37(S1): 162-167.
- [2] 张倩,吕丽平. 基于 ORB 与局部仿射一致性约束的快速人脸特征配准[J]. 电子测量与仪器学报, 2019, 33(6): 38-44.
- [3] 李成勇,王莎,陈成瑞. 基于 OpenCV 的人脸识别系统设计及实现[J]. 国外电子测量技术, 2021, 40(11): 168-172.
- [4] 吕开云,鞠厦轶,龚循强,等. 基于 PCA 和 IGG 权函数的人脸图像鲁棒线性回归分类方法[J]. 国外电子测量技术, 2021, 44(21): 168-172.
- [5] CAO Q, SHEN L, XIE W, et al. Vggface2: A dataset for recognising faces across pose and age[C]. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, 2018: 67-74.
- [6] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016: 770-778.
- [7] WANG Q, GUO G. LS-CNN: Characterizing local patches at multiple scales for face recognition[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 1640-1653.
- [8] 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, Cham, 2018: 428-438.
- [9] MARTINDEZ-DIAZ Y, LUEVANO L S, MENDEZ-VAZQUEZ H, et al. Shufflefacenet: A lightweight face architecture for efficient and highly-accurate face recognition [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, 2019.
- [10] YAN M, ZHAO M, XU Z, et al. Vargfacenet: An efficient variable group convolutional neural network for lightweight face recognition[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, 2019.
- [11] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]. Proceedings of the European Conference on Computer Vision (ECCV), Munich, 2018: 116-131.
- [12] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020: 1580-1589.
- [13] DENG J, GUO J, XUE N, et al. Arcface: Additive angular margin loss for deep face recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019: 4690-4699.
- [14] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, 2019: 1314-1324.
- [15] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[J]. Chinese Conference on Biometric Recognition Springer, Cham, 2018: 428-438.
- [16] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018: 4510-4520.
- [17] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018: 6848-6856.
- [18] ZHANG Q, LI J, YAO M, et al. Vargnet: Variable group convolutional neural network for efficient embedded computing [J]. ArXiv Preprint, 2019, ArXiv:1907.05653.
- [19] WU X, HE R, SUN Z, et al. A light cnn for deep face representation with noisy labels [J]. IEEE Transactions on Information Forensics Security, 2018, 13(11): 2884-2896.
- [20] DUONG C N, QUACH K G, JALATA I, et al. Mobiface: A lightweight deep learning face recognition on mobile devices[C]. 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems(BTAS), 2019: 1-6.
- [21] DENG J, GUO J, ZHANG D, et al. Lightweight face recognition challenge[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, 2019.
- [22] HUANG G B, MATTAR M, BERG T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments [C]. Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, 2008.
- [23] ZHENG T, DENG W. Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments[J]. Beijing University of Posts Telecommunications, DOI: 10.48550/arXiv.1708.08197.
- [24] SENGUPTA S, CHEN J C, CASTILLO C, et al. Frontal to profile face verification in the wild[C]. 2016 IEEE winter Conference on Applications of Computer Vision(WACV), New York, 2016: 1-9.
- [25] WHITELAM C, TABORSKY E, BLANTON A, et al. Iarpa janus benchmark-B face dataset[C]. proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Hawaii, 2017: 90-98.
- [26] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning, 2019: 6105-6114.
- [27] MARTINEZ-DIAZ Y, NICOLAS-DIAZ M, MENDEZ-VAZQUEZ H, et al. Benchmarking lightweight face architectures on specific face recognition scenarios [J]. Artificial Intelligence Review, 2021, 54(8): 6201-6244.

作者简介

赵锋, 硕士研究生, 主要研究方向为目标检测和识别。

E-mail: 18406585517@163.com

张鹏(通信作者), 博士, 副教授, 硕士生导师, 主要研究方向为智能微系统制作。

E-mail: sxyczhangpeng@126.com

张冉, 硕士研究生, 研究方向为信号与图像处理。

E-mail: sxdtzr123@163.com