

基于 MobileNetV2 与树莓派的人脸识别系统^①



赵洋, 许军

(沈阳化工大学 计算机科学与技术学院, 沈阳 110142)

通讯作者: 许军, E-mail: xujun1483@163.com

摘要: 人脸识别技术在安防, 商业, 金融等领域都有广泛的应用. 针对目前人脸识别系统成本高, 易用性低等现象, 提出了基于树莓派 (Raspberry Pi) 实现人脸识别的方案. 首先利用 OpenCV 计算机视觉库中的 Harr 级联方法, 对图像中的人脸进行定位; 然后利用改进的 MobileNetV2 网络模型对人脸进行特征提取和分类, 得到一个优化的人脸识别模型; 最后将模型移植到 Raspberry Pi 进行人脸识别. 该模型对图库中的人识别准确率为 95%, 对陌生人识别准确率为 80%. 实验结果表明该系统进行人脸识别工作稳定, 识别速度快, 应用场景广.

关键词: 图像处理; 树莓派; MobileNetV2; 人脸检测; 人脸识别

引用格式: 赵洋, 许军. 基于 MobileNetV2 与树莓派的人脸识别系统. 计算机系统应用, 2021, 30(8): 67-72. <http://www.c-s-a.org.cn/1003-3254/8156.html>

Face Recognition System Based on MobileNetV2 and Raspberry Pi

ZHAO Yang, XU Jun

(College of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang 110142, China)

Abstract: Face recognition technology is widely used in security, business, finance and other fields. In view of the high cost and low ease of use of existing face recognition systems, a scheme based on Raspberry Pi for face recognition is proposed. First, the Harr cascade method is adopted in the OpenCV computer vision library to locate the faces in images. Then the improved MobileNetV2 model is employed to extract and classify the faces to obtain an optimized face recognition model. Finally, the model is ported to Raspberry Pi for face recognition. The model has an accuracy of 95% for people in the gallery while 80% for strangers. Experimental results show that the system remains stable in face recognition, with fast recognition speed and wide application scenarios.

Key words: image processing; Raspberry Pi; MobileNetV2; face detection; face recognition

自 2006 年, LeCun 等发现通过逐层预训练的方式可以较好的训练多层神经网络, 提出了深度学习 (deep learning) 的概念^[1]. 从此, 人脸识别技术得到了快速的发展, 并且已经普遍应用到人们的日常生活中. 文献 [2] 提出基于 ARM 进行人脸识别的方法. 该方法首先通过 ViBe 算法提取监控区域内的运动目标, 然后再用局部二值模式 (Local Binary Pattern, LBP) 算法对人脸进行检测和识别. 该方法虽然能识别较远距离的人脸, 但是识别准确率不高. 文献 [3] 提出基于达芬奇平台

OMAP3530 进行人脸识别的方法. 该方法首先进行可见光图像和近红外图像的双路图像采集, 然后通过 Harr-like 特征的 AdaBoost 算法进行人脸目标的检测, 最后利用 LBP 算法进行特征提取. 该方法只克服了光照环境对识别精度的影响, 然而在复杂环境下, 识别效果欠佳. 文献 [4] 提出了一种利用云计算平台实现人脸识别的方法. 该方法首先将采集的图像传入嵌入式平台, 然后和 FaceCore 人脸识别云计算平台交互, 利用云计算平台进行人脸的检测和识别. 该方法在一定程

① 收稿时间: 2020-11-30; 修改时间: 2020-12-28; 采用时间: 2021-01-07; csa 在线出版时间: 2021-07-31

度上减少了嵌入式设备的计算量,但是需要通过网络和云计算平台进行数据传输,增加了等待的时间,实效性差.文献[5]首先把采集的图像传入互联网端,利用Dlib官方自带的人脸检测功能定位出人脸图像,然后使用Dlib官方训练好的深度残差网络模型识别人脸.该方法虽然很大程度上简化了开发的复杂度,但无法同时对多张人脸检测.由以上分析可知,在嵌入式设备或者移动端部署人脸识别系统还需要进一步的研究,特别是在保证成本低廉,易用性的前提下,如何提高识别速度仍是亟待解决的问题.

文献[6]提出了LBP, KNN和BP神经网络相结合的人脸识别方法.首先将图像分为局部区域并提取每个局部区域的特征纹理,然后将提取的特征图传入神经网络.LBP和KNN这种相结合的方式减小了原始训练图像之间的相关性.传统的人脸识别算法[7]虽然在人脸特征提取方面取得了很好的效果,但是这些方法在光照,表情,识别速度,准确率等方面还存在着一些不足.文献[8]提出了一种Adacos的人脸损失函数,使用自适应动态缩放参数来解决模型在训练过程中超参数的选择问题,但是模型在训练过程中参数比较多,模型的收敛速度比较慢.文献[9]提出了一种引进感知模型的人脸识别方法,该方法通过在孪生卷积神经网络上引入感知模型,使得网络能够学习到更加详细的特征,解决了在非限定条件下人脸识别困难的问题.文献[10]提出了一种基于空间注意力机制的3D人脸识别方法,该方法采用类间正则化损失函数,降低了不同人脸之间的特征相似度.虽然在一定程度上解决了低质量人脸图像的问题,但是其中的超参数设定只能根据经验设定,没有固定的计算公式.一些大型的深度学习网络模型[11]集成了特征提取和图像分类方法,采用自学习的方式进行训练,在准确率和速度方面得到了很大的提升,但是运行在计算能力有限的嵌入式设备或者移动设备上还有些困难.本文以MobileNetV2网络模型为研究对象,提出了一种基于Raspberry Pi的人脸识别方法.该方法首先对图库中的图像进行预处理,利用OpenCV库中的Haar级联方法对图像中的人脸进行定位,然后把数据送入MobileNetV2网络进行迁移学习,通过训练得到一个健壮的模型,最后将模型移植到Raspberry Pi,利用Raspberry Pi完成人脸识别的任务.

1 系统设计

1.1 系统硬件设计

考虑到成本、算力、功耗等各方面的原因,本系统选用Raspberry Pi作为主控制器[12].整个系统由图像采集模块,供电模块,图像处理模块以及图像显示模块组成.图像采集模块负责人脸图像的读入;供电模块为系统的正常运行提供稳定的电源;图像处理模块对读入的图像进行预处理,然后经过特征提取,对人脸进行识别;显示模块主要对读入的图像进行显示,将识别的结果显示出来.Raspberry Pi作为硬件和软件的核心部分,连接各个功能模块,搭载1.4 GHz的64位四核处理器,10/100以太网接口,双频802.11ac无线局域网,蓝牙4.2, HDMI结构,4个USB2.0接口D-SI显示连接器, RCA av端口音频输出,USB电源接口.

1.2 系统软件设计

系统的数据集是本研究室自己采集的.在不同的环境下,对每名成员采集1000张图片,以姓名命名,以JPG的格式存放在本地.MobileNetV2网络属于轻量级的神经网络,拥有更少的参数,更高的准确率,更快的速度,所以选择该网络结构作为我们的模型.模型采用迁移学习的方式进行训练,加载MobileNetV2网络的初始化参数,这样能够提高模型训练的速度,让模型更快的收敛.然后通过添加辅助层和修改Softmax层的分类类别,调整超参数,得到训练的模型.由于训练模型产生大量的参数,对设备的硬件资源和算力要求比较高,因此选用离线的方式在PC端训练模型,通过训练得到一个性能优越的模型.模型训练过程如图1所示.

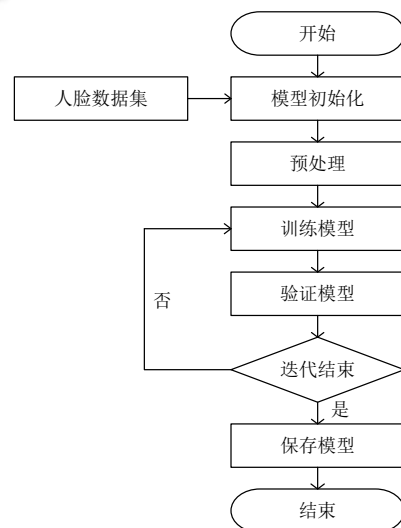


图1 模型训练过程

2 算法流程

2.1 图像预处理

人脸识别技术是基于人的脸部信息进行识别的一种生物技术,而人脸是保证人脸识别算法性能的重要因素之一.在人脸图像的采集过程中,摄像头的像素,拍摄的角度,光照强度等因素会影响整个系统的识别效果.为了增强系统的鲁棒性,避免光照强度对图像的影响,首先对图像进行归一化;然后利用 Haar 级联方法对人脸进行定位^[13],筛选出定位错误或者不满足条件的图像,最后把图库中的图像统一调整为 224×224 的大小.

2.2 网络模型

MobileNet 网络是 Google 团队在 2017 年提出的,专注于移动端或者嵌入式设备的轻量级网络,相比于传统的卷积神经网络,在准确率小幅度降低的前提下大大减少了模型参数和运算量. MobileNetV2 网络结构如表 1 所示.

表 1 MobileNetV2 网络结构

Input	Operator	t	c	n	s
224 ² ×3	conv2d	—	32	1	2
112 ² ×32	bottleneck	1	16	1	1
112 ² ×16	bottleneck	6	24	2	2
56 ² ×24	bottleneck	6	32	3	2
28 ² ×32	bottleneck	6	64	4	2
14 ² ×64	bottleneck	6	96	3	1
14 ² ×96	bottleneck	6	160	3	2
7 ² ×160	bottleneck	6	320	1	1
7 ² ×320	conv2d 1×1	—	1280	1	1
7 ² ×1280	avgpool 7×7	—	—	1	—
1×1×1280	conv2d 1×1	—	k	—	—

表 1 中 Input 为输入特征图的大小, bottleneck 为倒残差网络结构, t 为扩展因子, c 是输出特征矩阵的深度, n 是 bottleneck 重复的次数, s 是步长.

MobileNetV2 网络不但借鉴了 MobileNetV1^[14] 网络的深度可分离卷积^[15,16], 还借鉴了 ResNet 网络结构.深度可分离卷积可分成 DW (DepthWise) 卷积和 PW (PointWise) 卷积.如图 2 所示, DW 卷积是图像一个通道对应一个卷积核,卷积核的深度为 1,输入特征矩阵的通道数等于卷积核个数等于输出特征矩阵的个数.

相对于传统的卷积运算, DW 卷积它只对每个通道进行了独立的卷积运算,没有很好的利用空间上的特征信息,不能得到表达能力较强的 Feature Map,所以还要经过 PW 卷积(如图 3 所示).PW 卷积主要是利

用 $1 \times 1 \times M$ (上一层的 Maps 深度) 的卷积核对上一层的 Maps 逐点进行卷积,这样就能很好的对空间位置上的特征信息进行融合,得到表达能力较强的特征信息.

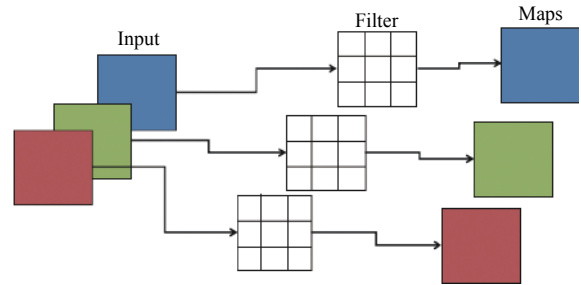


图 2 DW 卷积

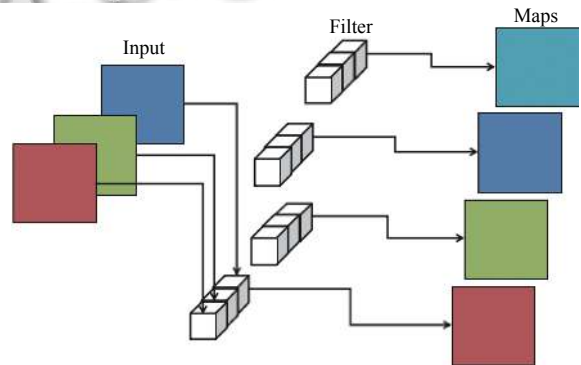


图 3 PW 卷积

深度可分离卷积大大减少了运算量,提高了运算速度.如图 4 所示, D_F 表示输入图像的大小, D_K 表示卷积核的大小, M 表示输入图像的深度, N 表示输出特征矩阵的深度.传统卷积计算量为 $D_F \times D_F \times D_K \times D_K \times M \times N$, 而深度可分离卷积的计算量为 $D_F \times D_F \times D_K \times D_K \times M + D_F \times D_F \times 1 \times 1 \times M \times N$.两者相比可以明显的发现,深度可分离卷积的计算量远远小于传统卷积的计算量.

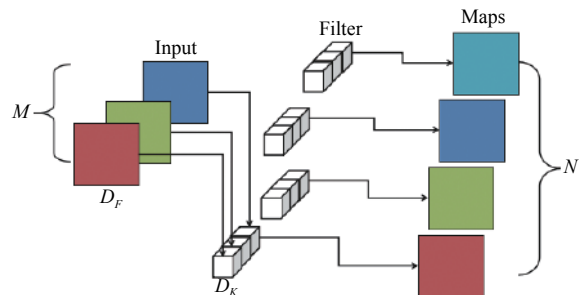


图 4 传统卷积

倒残差网络结构相比于残差网络结构,它先经过 1×1 的卷积进行升维,然后再通过 3×3 的 DW 卷积,最

后再通过 1×1 的卷积进行降维. 结构中的 Shortcut 操作只有当步距为 1 且输入特征矩阵和输出特征矩阵的形状相同时, 才会执行剪枝操作. 如图 5 所示.

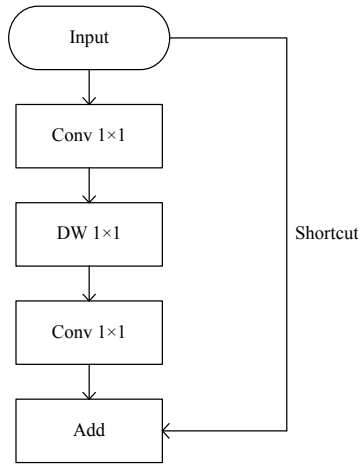


图 5 倒残差网络结构

2.3 训练数据和训练环境

本研究室自己制作的数据集主要通过摄像机拍摄的视频, 将采集的视频分帧, 取出不同时段的图像作为训练样本. 数据集共有 5000 张图片, 5 个类别. 每个类别的图像都各不相同, 包含不同的姿势, 不同的表情, 不同的年龄. 图像的尺寸都是 512×512 , 按 7:3 的比例把数据集分为训练集和验证集. 实验环境为 Windows 10 操作系统, 8 GB 内存, NVIDIA GTX1660 Ti 显卡, 在 Tensorflow^[17] 深度学习框架上实现本文的 Mobile-NetV2 模型并完成对模型的训练.

2.4 模型训练结果分析

在 Tensorflow 深度学习框架上使用 Adam 方法训练模型, 学习率为 0.01, 迭代次数为 50 次, 得到如图 6 所示的模型训练结果. 模型的准确率为 98.3%.

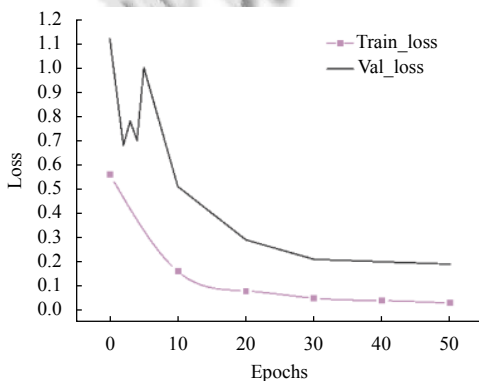


图 6 模型训练结果

由图 6 的训练和验证损失曲线可知, 模型在前 10 次迭代过程中损失不断下降, 说明模型在不断学习. 模型训练次数到达 40 次的时候, 交叉熵损失下降缓慢, 基本保持水平, 说明模型达到局部最优. 但是验证集的损失明显大于训练集的损失, 导致模型在训练的过程中效果比较好, 验证的过程中效果比较差, 这是一种过拟合现象. 综合分析, 模型的过拟合问题主要原因是我们的数据集只有 5 个类, 全连接层的更改导致模型从 $1 \times 1 \times 1280$ logits 回归到 $1 \times 1 \times 5$ 的全连接层的过程中有很大的信息特征突变. 为了解决模型过拟合问题, 采用正则化的方法对模型进行优化, 通过添加 Dropout 层^[18] 的方式, 得到训练后的特征. 不但在一定程度上避免模型过拟合问题, 也能提高模型的泛化能力.

由图 7 训练损失和验证损失曲线可以看出, 通过将学习率修改为 0.001 和添加 Dropout 层得到优化的模型, 在一定程度上降低了模型的过拟合问题. 由图可知训练损失和验证损失慢慢趋向于持平, 并且准确率相比优化前提高了 1.1%. 优化后的模型就可以移植到树莓派设备, 从而在实际的场景下进行测试.

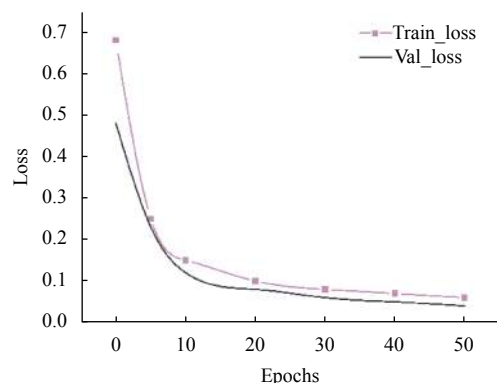


图 7 优化后的模型

3 实验结果与分析

3.1 测试平台

测试平台是 Raspberry Pi-3B+, 运行 Linux 操作系统. 搭载图像处理库 OpenCV3.4.6, Tensorflow2.0.0 深度学习框架. 具体参数如下:

CPU: 1.4 GHz 64 位 4 核 ARM Cortex-A53 处理器.

GPU: Broadcom VideoCoreIV, OPenGL Es2.0.

内存: 1 GB LPDDR2, 扩展储存卡: 16 GB.

3.2 人脸识别的流程

人脸识别系统的具体识别过程如下:

(1) 读入摄像头前的图像.

(2) 图像预处理, 首先对图像进行归一化操作, 去除光照对图像的干扰.

(3) 利用 OpenCV 计算机视觉库中的 Haar 级方法对图像进行人脸的定位, 并将图像大调整为 224×224.

(4) 把图像传入训练好的 MobileNetV2 网络, 对图像进行特征提取并且分类.

(5) 通过和图库中图像进行相似度比较, 相似度达

到设定的阈值就认为识别成功, 否则识别失败.

3.3 系统准确率测试

为了验证系统对人脸识别的准确率, 对图库中图像和非图库中图像进行测试. 图 8(a) 表示对图库中的人进行测试, 图 8(b) 表示对非图库中的人进行测试, 分别在不同的表情, 不同的光照条件下进行人脸辨别. 其中没有任何标识的表示识别成功, 出现标识 Input error 表示识别结果为非图库中的人.



图 8 准确率测试

从图 8(a) 中可以看出, 该系统对图库中的人脸整体测试结果良好, 偶尔出现误判, 但这并不影响系统的性能. 因为摄像头前的人可以在屏幕中看到自己的识别结果, 当出现误判现象, 可以通过矫正身姿, 调整检测距离等方法重新进行识别. 图 8(b) 结果显示, 对于陌生面孔很容易出现误判. 造成这种现象的发生一方面可能是因为训练数据集不足, 实验类别比较少, MobileNetV2 网络会将陌生人归为其中的一类; 另一方面, 网络结构过于复杂, 导致网络退化. 多次测试表明, 通过修改预测图像和实际图库中图像的相似度阈值, 可以在一定程度上提高准确率. 经过多次实验发现, 将相似度阈值设置为 0.9, 不但能够提高对图库中图像识别的准确率, 而且对陌生人识别的准确率也得到了提高.

3.4 系统稳定性测试

为了进一步对系统进行稳定性测试, 在自制数据集和 FERET 数据集上分别进行测试. 分别记录识别类型,

识别成功, 识别失败的次数和识别准确率. 成功次数表示的是识别人物和识别类型匹配成功的次数, 准确率=成功次数/(成功次数+失败次数). 自制数据集包含的对象比较少, 在复杂环境下只对 4 名成员进行 20 次测试, 分别标号为 1~4 号. 1~3 号为图库中人, 4 号为陌生人. 每名成员在不同的环境下进 20 次人脸识别实验, 结果见表 2. FERET 数据集包含 200 个类, 共 1400 张图片. 将数据分为 7 组, 5 组为图库中人, 每组分别为 10 人, 50 人, 100 人, 150 人, 200 人; 2 组为陌生人, 每组分别为 10 人, 50 人. 每人进行 10 次人脸识别实验. 结果见表 3. 不同的网络对数据集上的测试结果见表 4.

从表 4 中可以看出, 虽然 ShuffleNetV2 网络在数据集上的准确率略高于其它网络结构, 但是模型大于其它网络结构, 系统的识别速度会较慢. 结合表 2 和表 3 可以明显看出, MobileNetV2 网络整体识别准确率也比较高, 性能比较优良, 对于图库中的人识别准确率

达到95%以上,对于陌生人,识别准确率达到80%以上.针对于个别识别失败的现象,主要的原因在于模型的过拟合,因此下一步还要提高模型的泛化能力.

表2 自制数据集人脸识别测试

序号	识别类型	成功次数	失败次数	准确率(%)
1号	图库中人	20	0	100
2号	图库中人	19	1	95
3号	图库中人	19	1	95
4号	陌生人	16	4	80

表3 FERET数据集人脸识别测试

人数	识别类型	成功次数	失败次数	准确率(%)
10	图库中人	93	7	93
50	图库中人	471	29	94.2
100	图库中人	953	47	95.3
150	图库中人	1422	78	94.8
200	图库中人	1902	98	95.1
10	陌生人	79	21	79
50	陌生人	401	99	80.2

表4 轻量化神经网络模型在数据集上的对比

模型	自制数据集 准确率(%)	FERET数据集 准确率(%)	参数量 (M)
SqueezeNet	90	86.8	4.8
ShuffleNetV2	95	96.8	7.4
MobileNetV1	92	90.6	4.2
MobileNetV2	95	94	3.4

4 结语

针对目前人脸识别系统成本高,移植性低的情况,提出基于MobileNetV2与Raspberry Pi的人脸识别系统方案.首先对图库中的图像进行归一化,然后通过不断的调整参数,多次训练MobileNetV2网络,使损失降到最低,从而得到一个最优模型,最后将模型移植到Raspberry Pi,利用Raspberry Pi对未知人脸图像进行检测和分类.实验表明该系统性能稳定,能满足用户的基本需求,应用范围广,可应用于门禁,考勤等应用场景.

本文主要是以MobileNetV2网络结构为研究对象,基于Raspberry Pi进行嵌入式开发,实现实时人脸识别.然而,由于MobileNetV2网络结构比较深,对人脸进行特征提取的过程中会发生特征突变,容易对陌生的面孔造成误判.因此,如何提高准确率,避免特征突变对识别结果所带来的影响,是以后重点研究的问题.

参考文献

1 LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015,

- 521(7553): 436–444. [doi: 10.1038/nature14539]
- 2 江烂达, 储珺, 缪君. ARM平台下人脸识别智能监控系统. 计算机工程与设计, 2018, 39(2): 590–595.
- 3 徐志鹏, 黄敏, 朱启兵. 基于嵌入式的近红外人脸识别系统设计. 数据采集与处理, 2015, 30(1): 211–218.
- 4 陆娇蓝, 陈军, 杨著. 基于云计算的嵌入式人脸识别系统构建与研究. 计算机测量与控制, 2016, 24(4): 146–148.
- 5 孙玥, 杨国为. 基于人脸识别的学生考勤系统的研究. 现代电子技术, 2020, 43(10): 116–118, 123.
- 6 Abuzneid MA, Mahmood A. Enhanced human face recognition using LBPH descriptor, multi-KNN, and back-propagation neural network. IEEE Access, 2018, 6: 20641–20651. [doi: 10.1109/ACCESS.2018.2825310]
- 7 姚立平, 潘中良. 基于改进的HOG和LBP算法的人脸识别方法研究. 光电子技术, 2020, 40(2): 114–118, 124.
- 8 Zhang X, Zhao R, Qiao Y, et al. AdaCos: Adaptively scaling cosine logits for effectively learning deep face representations. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. 2019. 10815–10824.
- 9 徐先峰, 张丽, 郎彬, 等. 引入感知模型的改进孪生卷积神经网络实现人脸识别算法研究. 电子学报, 2020, 48(4): 643–647. [doi: 10.3969/j.issn.0372-2112.2020.04.004]
- 10 龚勋, 周炆. 面向低质量数据的3D人脸识别. 电子科技大学学报, 2021, 50(1): 43–51. [doi: 10.12178/1001-0548.2020321]
- 11 李倩玉, 蒋建国, 齐美彬. 基于改进深层网络的人脸识别算法. 电子学报, 2017, 45(3): 619–625. [doi: 10.3969/j.issn.0372-2112.2017.03.017]
- 12 李舜酩, 郭海东, 李殿荣. 振动信号处理方法综述. 仪器仪表学报, 2013, 34(8): 1907–1915. [doi: 10.3969/j.issn.0254-3087.2013.08.031]
- 13 何瑶, 陈湘萍. 基于OpenCV的人脸检测系统设计. 新型工业化, 2018, 8(6): 83–89.
- 14 Howard A G, Zhu M L, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv: 1704.04861.
- 15 Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9.
- 16 Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 1800–1807.
- 17 郭敏钢, 宫鹤. 基于Tensorflow对卷积神经网络的优化研究. 计算机工程与应用, 2020, 56(1): 158–164. [doi: 10.3778/j.issn.1002-8331.1906-0214]
- 18 Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors. arXiv: 1207.0580.