文章编号:1671-4598(2019)03-0220-05 DOI:10.16526/j.cnki.11—4762/tp.2019.03.045 中图分类号:TP393 文献标识码:A

一种组网人脸识别门禁系统的设计

邹优敏[1,2,3]，贵寅超[3]，马啸宸[4]，倪启东[1,2,3]，罗 恒[1,2,3]，刘晨旭[3]
(1. 苏州科技大学 江苏省建筑智慧节能重点实验室，江苏 苏州 215009；2. 苏州市移动网络技术与应用重点实验室，江苏 苏州 215009；3. 苏州科技大学 电子与信息工程学院，江苏 苏州 215009；4. 南京邮电大学 海外教育学院，南京 210000)

摘要：针对传统人脸识别门禁中识别率不高的情况，设计了一种具有环境针对性的可靠人脸识别网络，采用卷积神经网络技术和基于TCP/IP的云应用技术，对人脸识别过程中的人脸提取进行了改进，采用了先进的MTCNN模型，实验结果表明，在2代和10代训练后识别结果收敛效果不明显，存在较大梯度，20代时训练后结果收敛，准确率逼近100%，loss值也逼近0%，表明20代训练模型确实已经完全收敛，且训练的拟合速度适中，既没有出现过快或者梯度消失的现象，也没有出现过慢或者不收敛甚至反升的情况，可以说模型的学习率也已经取到最佳效果；经工程测试，该系统具有更高的识别准确性和注册兼容性，可以满足小区、公司等人员数量不多且较为固定的场所的应用需求。

关键词：人脸识别；卷积神经网络；TCP/IP；移动客户端

Design of Access Control System Based on Face Recognition and Cloud Technology

Zou Youmin[1,2,3], Fei Yinchao[3], Ma Xiaochen[4], Ni Qidong[1,2,3], Luo Heng[1,2,3], Liu Chenxu[3]
(1. Jiangsu Province Key Lab of Intelligent Building Energy Efficiency, Suzhou 215009, China; 2. Suzhou Key Lab of Mobile Networking and Applied Technology, Suzhou 215009, China; 3. Suzhou University of Science and Technology, Suzhou 215009, China; 4. Nanjing University of Posts and Telecommunications, Nanjing 215000, China)

Abstract: An environment adaptive face detection and recognition system is proposed in this paper to improve the recognition precision in traditional door access control systems. The two-stage convolutional neural network technology with precise division of labor and the cloud application technology based on TCP/IP is adopted to improve the performance of traditional systems. Experiment results, with the advanced MTCNN model, show that after 2 generations and 10 generations of training, it can't reach the convergence, while the results are better after 20 generations of training. Practical tests demonstrate the potential applications in communities and companies where the number of occupants maintains is comparatively small due to the high precision and compatibility.

Keywords: face recognition; convolutional neural network; TCP/IP; mobile client

0 引言

随着深度学习相关研究的不断发展，人工智能已经逐步从理论研究开始转入大规模应用，应用领域主要包括自然语言理解、视觉系统、问题求解、博弈、学习和知识展示等[1]，人脸识别门禁系统就是人工智能在视觉系统中的应用。

人脸是具有唯一性的生物标识[2]，近年来，人脸识别作为生物识别的关键技术之一，凭借其独特发展优势，已经被应用于信息安全、出入口控制[3]等领域，并且其仍具有广大的应用前景[4]，其中较为突出的就是门禁行业。

但是，当前市场中的门禁系统大多采取离线的形式，给用户在操作系统的时候带来了很多不便，且当前已有的人脸门禁产品，在使用后存在识别精度不够、容错率较低以及场景针对性不强的等不足，针对这些问题，系统使用MTCNN[5]与识别CNN分工处理组合方法[6]，使得识别更为精准，且容错率高，在样本数量较小的前提下，实现精确识别，可以有效适应小区或者公司这些人数相对较少和固定的应用场合。

收稿日期:2018-09-10； **修回日期**:2018-09-29。

基金项目:国家自然科学基金项目(61602334,61502329,61401297);住房与城乡建设部科学技术项目(2015—K1—047);江苏省自然科学基金项目(BK20140283)。

作者简介:邹优敏(1982-),女,江西省萍乡人,硕士,讲师,主要从事人工智能理论与方法及其在建筑节能中的应用方向的研究。

1 模型设计

1.1 人脸提取

目前人脸提取的主要困难主要有两方面：一是易受外部环境的影响，如人脸获取不全、外界光照的变化等；二是耗时过长，人脸必须用多个角度进行定位才能获得较高的提取成功率，导致提取耗时过长。

MTCNN，即多任务级联卷积网络，它具有三个连续深度CNN的框架：提案网，剩余净额和输出净额[7]，是一种通过调整输入图片的大小来形成不同的尺度的神经网络。具有经济和精度较高等优势。MTCNN使用3*3大小的卷积核，有效减少了计算的负担，论精度，图1给出了MTCNN的功能。

如图所示，整个网络将人脸提取的功能分成三个部分，即判断人脸、框出人脸和定位特征点，这三个功能由三层

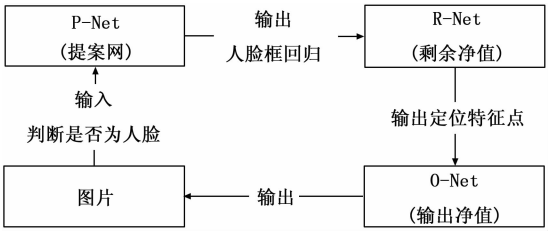

图 1　MTCNN 框架图

连续深度的卷积神经网络实现，每一层网络输入的都是 RGB 图，图片大小由式（1）确定：

$$minL = org_L * (12/minsize) * factor^{(n)} \quad (1)$$

其中：minsize 是可认为设计的最小人脸尺寸，factor 为缩放因子，org _ L 为输入图片的最小边，并且每一层的输入图片大小均不同。相比直接利用 haar 特征进行的人脸检测的方式，MTCNN 通过分工加强了每一层功能的精确度，能够在一定程度上消除外界影响诸如光照变化或者小部分的人脸缺失，但是这种方式拉长了识别网络的深度，使得运算时间大大增加，但这并不影响最终结果，系统使用的 MTCNN 的参数设置如表 1 所示。

表 1　MTCNN 参数表

Net_layer	Input_Size	Conv_Size
P－Net_Conv_1	(12,12,3)	(3,3)
P－Net_Conv_2	(5,5,10)	(3,3)
P－Net_Conv_3	(3,3,16)	(3,3)
R－Net_Conv_1	(24,24,3)	(3,3)
R－Net_Conv_2	(11,11,28)	(3,3)
R－Net_Conv_3	(4,4,48)	(2,2)
R－Net_dense_1	(3,3,64)	(None,128)
O－Net_Conv_1	(48,48,3)	(3,3)
O－Net_Conv_2	(23,23,32)	(3,3)
O－Net_Conv_3	(10,10,64)	(3,3)
O－Net_Conv_4	(4,4,64)	(2,2)
O－Net_dense_1	(3,3,128)	(None,256)

1.2　图像处理

为了提高模型训练的效率，需要将经过 MTCNN 网络输出的人脸图片进行进一步的处理，以减轻内存压力，减少时间成本。原本输出的图片是 48 * 48 * 3 大小的 RGB 图，先将图片转化为灰度图，即 48 * 48 * 1 的图像，之后使用双线性插值的图像缩放算法[8]，引入一个缩放因子 t，这样每个像素新的灰度值就是 $P(x/t, y/t)$，把新的 x、y 值设为 x_1，y_1，由于 x_1，y_1 必须为整数，所以可以淘汰掉一部分的像素点，这样可以找到四个与它相邻的灰度 f_1、f_2、f_3、f_4，然后通过双线性插值算法：

$$f(x,y) \approx (1-x \quad x)\begin{pmatrix} f(0,0) & f(0,1) \\ f(1,0) & f(1,1) \end{pmatrix}\begin{pmatrix} 1-y \\ y \end{pmatrix} \quad (2)$$

便可得到缩放后该点的灰度值。完成缩放后的图片尺寸为 128 * 128 * 1，虽然转为灰度图会减小肤色对训练的影响，但是从总体的精度来看肤色特征并不构成较大影响。

1.3　人脸识别

从图 2 可以看出，模型使用两个卷积卷积层和两个最大池化层来提去人脸中的一般特征，之后使用 Flatten 层将图片一元化，使用两个全连接层 Dense 进一步提取不同人脸的深度特征，最后运用 softmax 函数输出分布概率。完成整个人脸识别的过程。

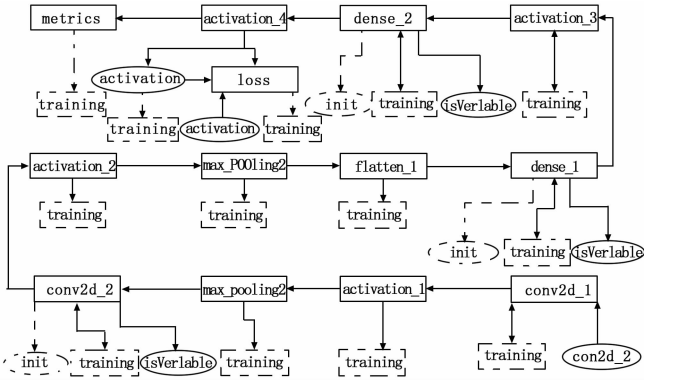

图 2　模型网络图

（1）卷积层：

设计使用两层卷积[9]，卷积核大小为 5 * 5，第一层一共有 32 个卷积核，第二层有 64 个。识别通过这样一个小核多核的设计，来尽可能地保留特征同时减少参数，也起到了平滑噪点的功能，防止过拟合，提高模型的泛化能力。

在经过一个二维卷积核的卷积操作后，又使用一个最大池化函数来进一步减少参数。所谓最大池化即是取出池化区域的元素的最大值，使用了一个 2 * 2 的池化区域，也就是没经过一次池化，将是池化作用的二维数据量缩小为 1/4。这样一个卷积层与池化函数的组合既捕捉到了主要特征，又使得数据规模进一步的减小，使得到全连接层可以更方便快捷得获取特征。

（2）全连接层：

表 2 给出了识别网络数据结构，由表可得，经过 Flatten 层将原本的三维数据压缩为一维之后，进一步使用一个 Dense 层获得一个 512 空间的输出。

表 2　模拟燃料组件测量实验

Layer(type)	Output Shape	Param
conv2D_1(Conv2D)	(None,32,128,1)	102432
activation_1(Activation)	(None,32,128,1)	0
maxpooling_1(Maxpooling)	(None,16,64,1)	0
conv2D_2(Conv2d)	(None,16,64,64)	1664
activation_2(Activation)	(None,16,64,64)	0
maxpooling_2(Maxpooling)	(None,8,32,64)	0
flatten_1(Flatten)	(None,16384)	0
dense_1(Dense)	(None,512)	8389120
activation_3(Activation	(None,512)	0
dense_2(Dense)	(None,num)	1026
activation_4(Activation)	(None,num)	0
Total params:8494242 Trainable params:8494242 Non－trainable params:0		

（3）Relu 函数：

图 3 所示为 Rule 函数曲线，由图可见，在 $x<0$ 时硬饱和，在 $x>0$ 时 $f(x)$ 的导数由为 1，所以梯度不衰减，从而缓解了梯度爆炸，又能更快地收敛，由于硬件能力的制约，采取的样本较少，Relu 函数就能非常好地解决模型训练中容易梯度爆炸的问题。

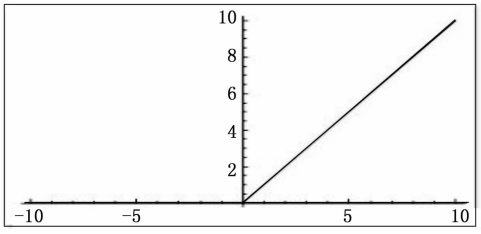

图 3 Relu 函数曲线

由表 2 可见，使用 Relu 函数进行激活，通过 Relu 激活的神经元特征进一步保留并映射到下一层的 Dense。最后一个 Dense 层最终收敛为所有的类别，并通过 softmax 的激活函数计算出各个类别的概率，表 1 中的 num 为训练人脸种类的数量。

（4）loss 函数：

与 softmax 相对应的，设计使用 categorical _ crossentropy[10] 函数作为 loss 函数（或称结果函数），在使用这个函数之前应该将所有的标签先 one－hot 化，也就是说将标签的判断 0、1 化，与该标签吻合为 1，不吻合为 0，并将标签组转化为 one－hot 数据类型。Softmax 和 categorical _ crossentropy 函数的组合被广泛应用于这样的多分类问题之中，使用 loss 进行分层再使用 softmax 作为输出层之后便能将各个标签的概率都表示出来，而不是简单的判断到底是哪一个标签，这就是这个组合的优势。

1.4 模型测试

图 4 所示为成品模型训练之后的 acc－loss 曲线图，图中横轴为训练的代数，也就是说横轴代表着第几次循环训练，纵轴则是 loss/acc 值。从图中可以清晰地看出，在 2 代和 10 代训练明显没有最终收敛，还有很大的梯度，而 20 代时训练已经基本完成，准确率逼近 100%，而 loss 值也已逼近 0，表明 20 代训练模型确实已经完全收敛。且训练的拟合速度适中，既没有出现过快或者梯度消失的现象，也没有出现过慢或者不收敛甚至反升的情况，可以说模型的学习率也已经取到最佳。

图 5 展示的是实际预测结果，模型的预测结果还是比较准确的，并且就算稍稍偏头或者缺失了一下部分的脸的 MTCNN 算法还是可以将脸部抓取出来，然后将矩形框和识别结果绘制在图片上。

从图 5 两次典型场景测试中可以发现，完成所使用的模型预测并不受到用户周围环境、穿戴等环境因素影响，这得益于人脸提取和人脸识别的分工、按序完成，即系统在进行测试识别之前已经将人脸提取出来，环境怎样已经

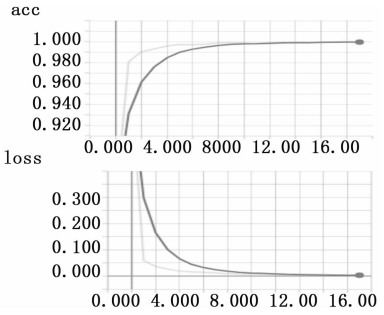

图 4 acc－loss 曲线图

图 5 典型场景实测结果图

无关紧要，并且采用 MTCNN 进行人脸提取容错率也很高，不会因为偏头或者人脸的部分缺失甚至异物遮挡而导致提取失败。

2 硬件系统设计

2.1 硬件总体设计

门禁系统，从功能上讲可以分为单向门禁系统和双向门禁系统，也就是说系统是涉及进出还是只能监控进门；从技术上讲又可分为单机和组网门禁系统。系统采用的是单向组网涉及，具体结构可见图 6。

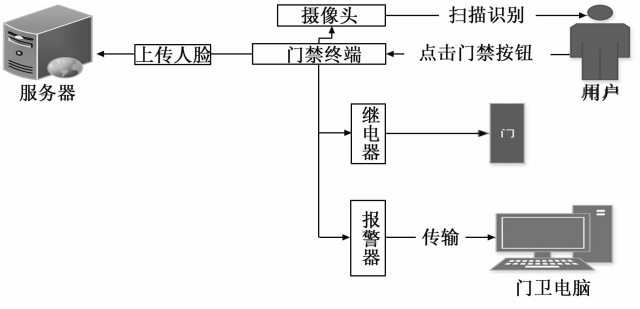

图 6 硬件系统总体设计图

由图 6 可见，门禁终端设备是以 X3288 芯片为核心的开发板。该芯片体积小，且有充足的 GPIO 口，支持前兆有线以太网，也支持电池休眠唤醒，非常适合作为需要长期待机的门禁系统的控制核心，且该开发板也支持多种操作

系统，如 Linux 系统、安卓系统和 ubuntu 系统，十分便于开发。使用安卓系统作为控制终端的操作系统，使用安卓编程可以非常方便地进行前后交互以及用户界面的业务处理，对开发者比较友好。图 6 还给出了硬件系统实现的功能，当用户点击设备显示屏上的门禁按钮时，门禁设备将调用摄像头捕捉用户的人脸，同时上传后端并由后端判断并返回判断结果，设备将按照结果判断是调用继电器开门还是报警并向门卫室传递信息，详细的功能流程将在软件系统设计中进行解释。

图 7 给出了硬件结构原理图。如图所示，门禁设备主体部件由摄像头、报警器、指示灯、继电器和网络设备组成。由于时刻与后端保持通信，需要让终端设备保持网络畅通。在网络连接上使用无线连接方式，主要原因是设计选择了安卓系统作为设备主控系统，使用有线连接并不方便操作；同时，无线连接又确实方便可靠，并且在每个门禁设备相距不远的情况下，又可以共用一个路由器，简化了线路的复杂性，有利于维护和拆装。大多数需要大量运算的识别和训练过程均不在前端设备中进行，也是出于经济考虑，但是设备端仍然存储有人脸识别最新的模型并且将实时更新，此是针对断网的情况，使设备不至于停止工作。

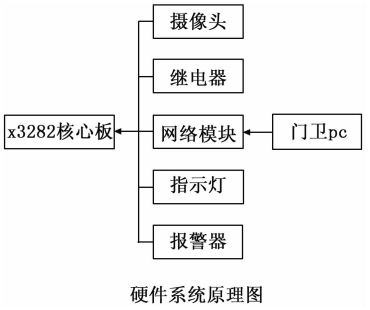

图 7　硬件系统原理图

2.2　软件系统设计

采取传统的网络设计模式，即用户的移动客户端＋后端＋设备端的形式，客户端选择微信小程序＋iOS App＋Android App 的形式，给用户多种选择，也加强的平台兼容性。后端使用 Tomcat 作为 Web 服务器的容器，采取 javaEE＋mysql＋python 的编程形式，javaEE 和 mysql 完成多种业务的处理，并由 java 调用 python 完成模型的更新训练与调度。设备端则使用 Android 编程，完成的功能与用户客户端的功能类似。

（1）移动客户端：

移动客户端的主要功能是将待识别人员的脸部图像上传至服务器，用以更新模型，流程如图 8 所示。由图可见，用户打开 App 后，需进行登录验证。若系统针对一个公司或者小区的内部系统，则不开放注册，用户的账号和密码均由管理员在后台输入，用户只需使用账号密码登录即可。进入用户界面之后，有两种上传的方式，分别可以使用现场拍摄或者已有的正脸照，用户可重复上传自己的照片，识别精度随着用户上传照片数量增加而提高。使用 python 接口对照片进行微调由一张照片生成几十张人脸，与已有的人脸库一起进行训练，更新识别模型和版本号，这个版本号与设备端的模型更新有关。

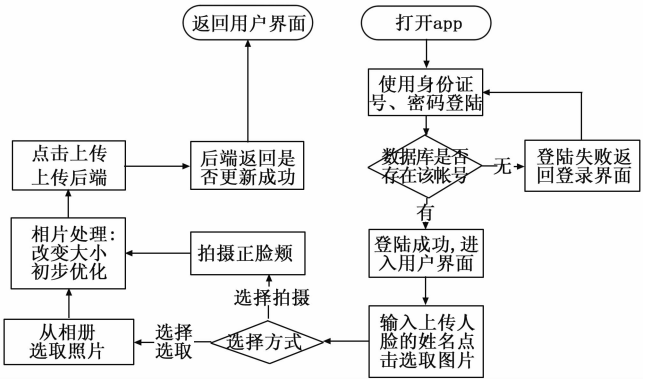

图 8　客户端功能流程图

（2）设备端：

设备端与后端的结合可以简化设备的较多功能，原设备不仅负责实现识别的功能，但需将识别的模型转移到后端去，而本地的模型只是负责在断网之时的维持功能，详细的设备端与后端的前后交互流程可见图 9。图 9 中拍摄照片并不是直接拍摄，而是调用相机进入预览模式，然后从预览的视频中截取一张图片上传后端，后端在接收到图片之后会对图片进行截取人脸、转灰度图、降噪处理，在截取人脸中将返回一个结果信息，即有没有检测到人脸。然后系统调用后端模型对照片进行预测，预测结果将返回结果信息，代表人脸标签库中是否有这个人脸，如果有将返回该人脸的姓名和吻合的可能性（使用小于 1 的小数表示），如果没有则返回不认识。之后将结果信息返回给前端设备并由前端设备进行判断。

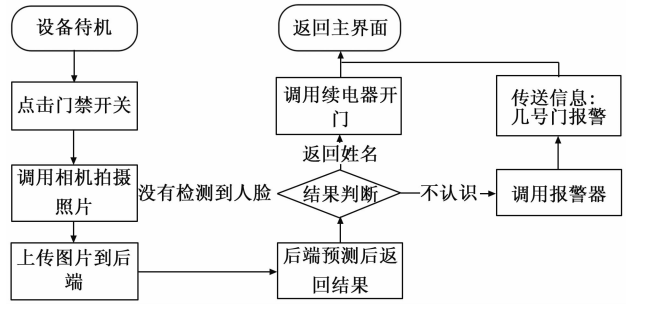

图 9　设备操作流程图

在设备端还有一个备用模型，这个备用模型是在断网之时由设备调用对人脸进行预测的。设备端的人脸识别模型是如何更新的可见图 10，系统使用了心跳连接技术，即在设备的程序中开一个线程，每 30 秒向后端发送一次请求，用存储在设备中的版本号与客户端使用流程中介绍的后端的版本号进行一次减法运算，如果小于 0，代表设备端

的模型版本低于系统模型的版本，于是设备端再次发送请求将系统的新模型和版本号下载到本地进行更新；如果等于 0，代表如今已是最新版本，无需更新，设备端继续等待 30 秒到下一次心跳连接。

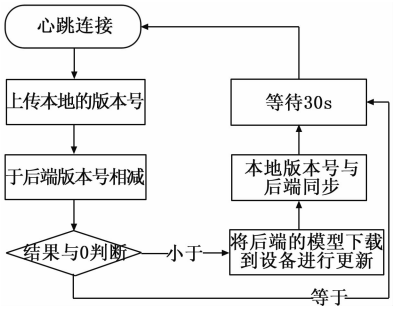

图 10 本地模型更新流程

3 实验结果与分析

在实际应用中，人脸识别的识别精度会受到各种外在环境的影响。实验将采用控制变量法，测试在脸部反射光强差异大、有无遮挡物和脸部表情发生变化这 3 种影响较大的环境变数下，产品是否还能实现精确的识别，并且还将使用极限测试的方法，条件将逐渐苛刻，以此更好地测试设备的性能。

3.1 人脸反射光强对识别的影响

实验对象始终控制为同一个人，在正常表情并佩戴眼镜的情况下进行测试，分别让摄像头对光和背光对人脸进行跟踪采集识别，每隔三秒记录一次人脸识别的置信度即识别率。图 11 位 10 次数据采集获得的识别率的折线图，图像表明在脸部反射光较强的情况下识别率明显优于反射光较弱的情况。对采集数据计算数学期望，反射光较强的情况下数学期望为 0.8504，反射光较弱时数学期望为 0.7673，虽然识别率在两种情况下仍有差距，但就算在反射光较弱的时候，识别率仍然非常高，识别结果可信。

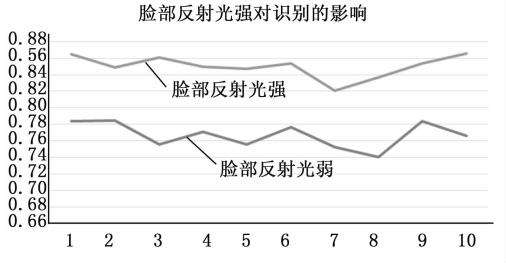

图 11 间隔三秒测得识别率折线图

3.2 表情变化对识别的影响

根据上一次实验的结果，实验控制人脸反射光较弱的时候进行极限测试。为了突出表情的变化，实验采用正常表情和咧嘴笑这两种脸部特征区别较大的情况进行测试。结果如图 12 所示，折线图反映表情变化对识别率的影响并不是很大，计算两种情况数据的数学期望，正常表情下数学期望为 0.7673，咧嘴笑时数学期望为 0.7658，仍然保持较高的识别精度。这是由于在模型训练之前是使用 MTCNN 网络进行脸部的时候，并不是提取完整的人脸，而是抓取特征点密集的部分，造成在提取之时忽略了一部分特征点，导致模糊了表情变化的影响。

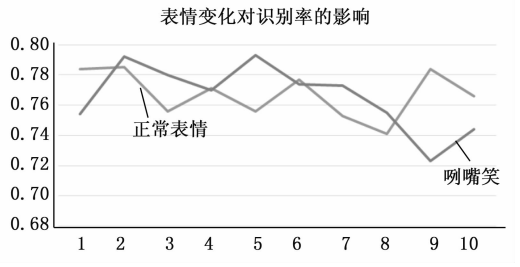

图 12 不同表情下测得识别率折线图

3.3 遮挡物对识别的影响

鉴于表情影响并不大，实验仍然控制正常表情并控制较弱的人脸反射光。图 13 的第一张图是在用户只上传了一次图像数据的情况下遮挡物对识别率的影响，并且上传的数据为佩戴眼镜的图像数据。结果表明识别率的折线在遮挡物的影响下确有震荡，裸眼情况下识别率明显低于佩戴眼镜的情况。测量数学期望值，在裸眼时的数学期望为 0.710 3，而在佩戴眼镜之时的数学期望为 0.767 3，可见遮挡物对识别率的影响较大。但是本系统支持用户多次上传数据更新模型，实验又进行了一次变动，实验对象再次上传了一些在正常光照下裸眼的图像数据进行模型更新，再次进行测试，图 13 的第二张图为实验结果，可见模型更新后遮挡物的影响明显变弱，两种情况下识别率并无大出入，裸眼时候的数学期望也上升至0.769 1，故用户多次上传不同情况下的图像数据可以很好地消除一些环境影响。但即使在最恶劣的识别环境下，模型的预测预测置信度仍然在 0.710 3的较高水平，故产品具有很好的环境适应性。

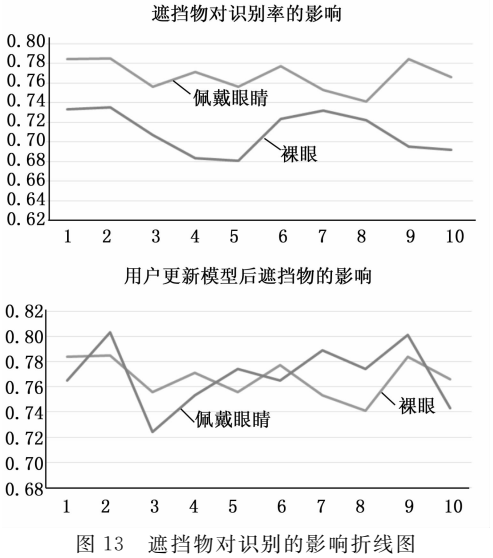

图 13 遮挡物对识别的影响折线图

（下转第 230 页）