计 算 机 研 究 与 发 展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202440415
61(11): 2821−2834, 2024

一种面向耳戴式设备的用户安全连续认证方法

王 勇[1,2] 熊 毅[1] 杨天宇[1] 沈益冉[3]
[1](哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001)
[2](电子政务建模仿真国家工程实验室(哈尔滨工程大学) 哈尔滨 150001)
[3](山东大学软件学院 济南 250100)
(wangyongcs@hrbeu.edu.cn)

A User Security Continuous Authentication Method for Earable Devices

Wang Yong[1,2], Xiong Yi[1], Yang Tianyu[1], and Shen Yiran[3]
[1](*College of Computer Science and Technology*, *Harbin Engineering University*, *Harbin* 150001)
[2](*Modeling and Emulation in E-Government National Engineering Laboratory* (*Harbin Engineering University*), *Harbin* 150001)
[3](*School of Software*, *Shandong University*, *Jinan* 250100)

Abstract Earable devices are used as typical AIoT edge sensing devices. Protecting the privacy of legitimate users and preventing illegal use has become extremely important. In response to the current user authentication methods for earable devices, which are limited by input interfaces, sensor costs, and device power consumption, resulting in insufficient security, low universality, and poor user experience, a user authentication model based on the built-in inertial measurement unit (IMU) of earable devices is proposed. This model extracts user-specific information by collecting vibration signals generated by users performing facial interaction gestures, and achieves diversified implicit continuous user authentication based on intelligent analysis of the above information. To extract accurate and reliable user-specific information, a deep neural network feature encoder based on a Siamese network is proposed, which maps gesture samples of the same user closer in the feature space and enlarges the distance between gesture samples of different users, achieving effective encoding of user-specific information. For continuous user authentication based on user-specific information, a weighted voting strategy based on the distance of the one-class support vector machine hyperplane is proposed, which can adaptively optimize the discrimination boundary to better capture the contained features and structures. The confidence level of the sample is determined based on the distance of the sample points inside and outside the hyperplane, and a weighted voting is designed for authentication. Experimental results show that the method in this paper achieves an authentication accuracy of 97.33% in a single vote, and achieves an authentication accuracy of 99.993% after seven rounds of continuous authentication, which is better than all the methods compared in this paper. It provides a smoother user experience and a higher level of security without the need for passwords, and has a high practical application value.

Key words AIoT; earable devices; user continuous authentication; deep neural network feature encoder; weighted voting strategy

收稿日期: 2024-05-31; **修回日期**: 2024-09-06

基金项目: 国家自然科学基金项目(61672179); 教育部人文社会科学研究青年基金项目(20YJCZH172); 中国博士后科学基金项目(2019M651262); 黑龙江省省属本科高校优秀青年教师基础研究支持计划资助(YQJH2023302)
This work was supported by the National Natural Science Foundation of China (61672179), the Youth Fund Project of Humanities and Social Sciences Research of the Ministry of Education of China (20YJCZH172), the China Postdoctoral Science Foundation (2019M651262), and the Funding from the Basic Research Support Program for Outstanding Young Teachers in Provincial Undergraduate Universities in Heilongjiang Province(YQJH2023302).

通信作者: 杨天宇(yangtianyu@hrbeu.edu.cn)

摘 要 耳戴式设备作为典型智能物联网边端感知设备应用场景众多,保护其合法用户隐私以及防止非法使用至关重要.针对当前耳戴式设备用户身份认证方法受输入界面、传感器成本以及设备功耗等限制导致安全性不足、普适性不高以及用户体验不佳等问题,提出一种基于耳戴式设备内置惯性测量单元的用户身份认证方法,该方法通过采集用户执行面部交互手势所产生的振动信号来提取用户特异性信息,并基于上述信息的智能分析实现多样化的隐式用户连续身份认证.为了提取精准可靠的用户特异性信息,提出了一种基于孪生网络的深度神经网络特征编码器,将同一用户的手势样本映射到特征空间中更近的位置,放大不同用户的手势样本之间的距离,实现用户特异性信息的有效编码.对于基于用户特异性信息的用户身份连续认证,提出了一种基于单类支持向量机超平面距离的加权投票策略,能够自适应地优化判别边界来更好地捕捉蕴含的特征和结构,根据超平面内外样本点与超平面的距离决定该样本的置信程度,以此设计加权投票实现认证.实验结果表明,所提方法在单次投票中实现了97.33%的认证准确率,7轮投票的连续认证后取得99.993%的认证准确率,优于对比的所有方法,无需密码的同时提供更流畅的用户体验和更高级别的安全性,具有较高的实际应用价值.

关键词 智能物联网;耳戴式设备;用户连续认证;深度神经网络特征编码器;加权投票策略

中图法分类号 TP391

近年来,伴随智能物联网(artificial intelligence of things, AIoT)的飞速发展,AI网络代理正在以席卷之势颠覆传统的软件行业,改变着物联网用户与计算机交互的方式[1].耳戴式设备(earable device),尤其是无线网络耳机,正在成为最受欢迎的智能可穿戴设备之一,因其在人机交互(human-computer interaction, HCI)领域的潜力,在未来很可能成为用户与代理之间交流的主要桥梁,实现人-机-物互联互通.

当前市场中已经出现了支持交互式触摸的无线耳机,提供了方便的HCI接口,用以操控智能手机的部分功能.典型的如AirPods[2]是苹果公司的主流蓝牙耳机,该产品支持点击操作,例如双击播放/暂停音乐、接通/挂断电话等功能.Freebuds[3]是华为公司的旗舰产品,在双击和长按的基础上,该产品还支持滑动操作从而调节音量.LinkBuds[4]是索尼旗下的一款蓝牙耳机,该产品在保留了基本的耳机面板触控外,特别提供了广域点击功能,允许用户点击耳朵周围的皮肤,实现各种各样的智能操控.

正因耳戴式设备在交互功能上的快速发展,隐私和安全问题将变得比现在更紧迫.因为其便携性,耳戴式设备容易丢失或被盗窃,这将直接损害用户的隐私和财产安全[5-6].然而,目前的学界和工业界还没有对耳戴式设备的安全给予足够的重视[7].现存的方法或受高昂的成本限制,难以部署到市场上的绝大多数耳戴式产品;或未取得优越的认证准确率,如若部署也将影响用户的使用体验.因此,为耳戴式设备开发真正具有实际应用的身份认证系统,防止攻击者未经授权访问隐私数据或使用隐私服务是至关重要的[8].

本文提出了一种基于耳戴式设备的用户连续认证方法,依赖不同用户执行在面部的交互手势的特异性实现用户身份的认证.本文方法以孪生网络[9]为框架训练了一个深度神经网络作为特征编码器,在训练过程中实现了用户特征与手势特征的解耦合,实现对用户特异信息的显著编码[10].经过广泛的实验表明,特征编码器所编码的特征优越地将不同用户映射到了不同的特征空间分布中,其提取用户特异性信息的能力相比主流的机器学习模型取得更优的效果.在此基础上,本文提出了基于单类支持向量机(one-class support vector machine, OCSVM)[11]的认证方法,该方法所生成的超平面可切分合法用户与攻击者.根据内外样本点与超平面的距离,本文方法设计出基于超平面距离的新颖加权投票方法,以此实现了连续身份认证.广泛的实验表明,本文方法在平滑权衡后,单次投票中实现了0.026 7的等错率,即97.33%认证准确率;同时在连续认证的任务上,7轮加权投票实现7×10^{-5}的等错率,即99.993%的认证准确率,超越了现有最好方法的表现.

当下市场上的智能无线网耳戴式设备基本都集成了惯性测量单元(inertial measurement unit, IMU),本文提出的身份认证模式无需引入额外传感器负担.虽然目前大多数的耳戴式设备无法负担像深度神经网络这样昂贵的计算,但随着移动芯片的快速发展,可以期待耳戴式设备未来将能够运行复杂的算法来认证其合法的用户.

本文的贡献体现在3个方面:1)定义了7种应用于耳戴式设备的面部交互手势,提出一种面向边端耳戴式设备内置的IMU的用户身份认证模式,并提

供了一种新颖的认证方法，实现多样化的隐式身份认证，具备更流畅的用户体验和更高级别的安全性，能够适配于市面上大部分耳戴式设备. 2)提出了一种基于孪生网络的深度神经网络特征编码器，实现了用户特征与手势特征的解耦合，对用户特异信息的显著编码，其编码效果优于本文对比的主流机器学习模型. 3)根据内外样本点与超平面的距离，设计出基于超平面距离的新颖加权投票方法，以此实现了连续身份认证. 本文方法在单次投票中实现 97.33% 的认证准确率，7 轮投票的连续认证后取得 99.993% 的认证准确率，具备较高的实际应用价值.

1 相关工作

近年来，物联网感知层设备的交互模式创新受到了广泛关注，因此出现许多面向感知设备如耳戴式设备的用户认证工作. 其中利用 IMU 采集的振动信号、麦克风采集声学信号或其他轻量化传感信号都成为了用户认证的媒介. 表 1 展示了本文参考的部分相关工作.

Table 1 Methods Used in the Related References

表 1 相关文献所采用的方法

方法	传感器	生物信号	主要算法	连续认证	设备成本
文献 [12]	电容传感	手势	逻辑电路		高
文献 [13]	IMU	手势	DNN	√	低
文献 [14]	IMU	手势	域对抗	√	低
文献 [15]	IMU	动作	机器学习		低
文献 [16]	IMU	动作	RNN		低
文献 [17]	IMU	头部震动	CNN	√	低
文献 [18]	IMU	下颌震动	CNN	√	低
文献 [19]	麦克风	声波	机器学习		低
文献 [20]	麦克风	牙齿声纹	DNN		低
文献 [21]	麦克风	手势震动	孪生网络	√	低
文献 [22]	麦克风	耳内外声	线性回归	√	低
文献 [23]	PPG	心脏活动	SVM	√	高
文献 [24]	PPG	声波	残差门控网		高
本文	IMU	手势	机器学习	√	低

表 1 中部分基于人体手势特异性进行交互和认证的工作取得了不错的进展. 文献 [12] 提出了一种新的耳部硬件设备识别手势，原型包括一个弧形的触摸板，上面有 12 个触摸点，以及一个灵活的设计，允许用户根据自己的耳朵形状弯曲和调整耳塞，能够精确地与耳朵上的特定区域进行交互，但是该方法将设备感知域固定在耳部硬件的识别范围内，在交互方式上缺乏一定的灵活性与易用性，可能会影响用户体验效果. 文献 [13] 提出了一种通过采集耳机麦克风数据识别手势进行交互的系统，它通过无线耳塞的麦克风来识别用户在面部和耳朵上执行的手势，从而实现与设备的交互. 在安静和有背景噪音的环境中收集了 20 名参与者的这 8 种手势的数据，其优化后的分类器达到了 95.3% 的准确率. 然而该方案采用了更多参数的 DenseNet 模型，模型具有一定的复杂度，运算成本和部署成本均较高，在资源能耗有限的环境下使用受限. 文献 [14] 提出了一种耳机上的基于域对抗方法的手势认证系统，创造性地提出了基于不同手势样本距离的权重投票法进行持续认证，实现了 99.9% 的认证正确率，但是和本文相比该工作缺乏更多的认证模型对比.

目前的智能物联网边端设备大多配置了 IMU，基于此设备，学界也做出了许多具有创新性的工作. 文献 [15] 提出了一种基于可穿戴加速度计传感器的位置无关人体活动识别模型，通过应用线性和非线性变换减少传感器位置和方向的敏感性，并在 PAMAP2 数据集上验证了其高效性，实现了平均 94.7% 和 91.7% 的位置依赖和位置无关的活动识别准确率. 与本文相比，该工作的数据采集存在样本分布不均的问题，需要扩充数据集来进一步提升识别效果. 文献 [16] 基于智能手表中的 IMU 采用用户输入密码时的运动数据进行用户识别，使用深度循环神经网络(recurrent neural network, RNN)来学习特征提取器，并通过复合损失函数来训练网络，使其能够在有限的训练数据上工作，在少量数据上也取得了比较好的效果，然而与本文相比，该方案由于缺乏足量数据，导致认证准确率难以达到实际应用水平. 文献 [17] 提出了一种新型的基于振动信号的智能头戴设备用户认证方式，利用头戴设备产生的振动信号来识别用户，该系统使用线性马达生成振动信号，IMU 传感器采集这些信号，并通过卷积神经网络(convolutional neural network, CNN)模型进行处理和分类，实验结果显示，使用短振动信号(≤1 s)时，系统具有很高的认证准确率，误接受率(false accept rate, FAR)和误拒绝率(false refuse rate, FRR)均约为 5%，持续或频繁使用振动信号可能会导致一些用户的不适，可能限制系统在长时间使用中的实用性. 文献 [18] 采用耳机 IMU 捕捉下颌振动信号进行用户认证，提供单次认证和持续认证方法，增强了智能耳机的安全性，但是针对不同年龄段的用户需要进行优化.

但是上述方法在功耗和实际应用场景仍然存在提升空间,需要进一步开发以实现商业化.

面向声学设备和方法的认证和交互也是一个热门的方向,基于声学信号的工作,如文献 [19] 提出了一种耳戴式运动感知系统,该系统利用耳内耳机捕捉人体运动产生的声波,通过 31 名受试者收集的数据,研究显示 OESense 分别在计步召回率达到 99.3%,在 5 种活动识别召回率达到 98.3%,以及在人脸上的 5 种敲击手势识别召回率达到 97.0%. 然而该方案选取的声音信号易受外界噪声的影响,在实际生活环境中使用时抗噪能力将会受到影响. 文献 [20] 采用了一系列代表性的牙齿手势,并通过耳穿戴设备的内向式麦克风来可靠地捕捉这些声波. 系统提取多级声学特征以反映牙齿的内在信息,并用于认证,由于每个人牙齿结构独一无二,因此也具有较容易识别的个体差异,通过 25 名参与者的实验,研究显示 ToothSonic 仅使用 1 种牙齿手势就能达到 95% 的准确率. 但是这些模型尚未在实际中广泛应用,具体使用效果尚不明确. 文献 [21] 设计了一种耳机上的基于声音指纹的认证系统,基于摩擦声学理论,将声学指纹特征分为 3 类:模式类(pattern-class)、脊沟类(ridge-groove-class)和耦合类(coupling-class). 基于此,将认证分成捕获声学指纹、信号预处理、特征提取和用户认证,这 4 个阶段采用深度学习编码器适应不同用户特征,EarSlide 系统的平均平衡准确率达到了 98.37%,证明了其作为一种鲁棒且安全的耳穿戴设备认证系统的潜力,但与本文相比,该方案拥有复杂的数据预处理部分,实时性方面需要进一步提升. 文献 [22] 提出了一种基于耳戴式设备的非侵入式用户认证方法,该方法通过比较耳内和耳外声音的差异来估计个体特有的耳道传输函数. 它使用线性回归捕捉不同频率带上的内外耳声音的缩放因子,并将其作为耳道传输函数用于用户认证,通过 12 名受试者在 6 种不同嘈杂环境下的数据收集,EarSlide 系统实现了 4.84% 的平衡误差率(balanced error rate, BER). 然而该方案在实验中可以发现其易受外界嘈杂环境的干扰,在实际嘈杂环境如海浪声或音乐声的应用中表现退化. 考虑到耳戴式设备的应用场景,能播放音乐基本可认定为耳戴式设备的固有功能,上述基于声学信号进行用户认证的方案难以规避音乐播放带来的模型表现退化的现状. 相较于声学信号,本文基于 IMU 采集的振动信号不需要做出引人注意的行为,具备一定的隐蔽性和用户友好性. 同时,手势引起的振动信号不容易被其他动作干扰,具备较好的抗噪能力.

除了上述传统的传感器和交互方式,近年来一些新型传感器也开始集成到智能物联网设备中,为人机交互提供了新思路. 文献 [23] 中 EarPass 通过耳内光电容积图(photoplethysmography, PPG)传感器捕获信号,开发了一个新颖的认证流程,包括特征提取和分类,以实现连续用户认证. 该方法首先对输入的耳内 PPG 信号进行预处理,以便于特征提取和分类. 此外,提出了一种检测和消除由头部运动引起的运动伪迹(motion artifacts, MAs)的方法. 文献 [24] 中 EarPPG 通过检测和分析耳内 PPG 信号的变化来进行用户认证,这些变化由说话时的面部肌肉活动和血管变形引起,研究团队采用了 1D ReGRU(一维残差门控循环单元)网络来构建用户认证模型,并在 25 名参与者中进行了全面的现实世界评估,取得了 94.84% 的准确率和 95% 的 $F1$ 分数,但是头部运动可能会对 PPG 信号造成影响,且成本较高,和本文采用的 IMU 设备相比,鲁棒性和实用性方面需要进一步提升,尚未集成在当前成熟的头戴设备中.

综上所述,本文提出一种面向耳戴式设备的用户安全连续认证方法. 该方法通过基于孪生网络的特征提取器提取出用户相关的关键信息,相比之前的工作将认证定义为分类问题的框架,本文将认证定义为异常检测. 由于异常检测算法更关注输入样本的整体数值特征而非分类的正确率和支持向量机自身的泛化性,因此面对复杂样本往往具备更高的鲁棒性. 该方法在单次投票中实现 97.33% 的认证准确率,7 轮投票的连续认证后取得 99.993% 的认证准确率. 进行了大量对比试验,验证其有效性. 同时,在实际环境中进行了性能测试,结果表明其能够保持较高的续航能力.

2 用户身份认证方法

为解决耳戴式设备认证用户身份的问题,本文提出了一种面向耳戴式设备内置的 IMU 的用户身份认证方法. 该方法的总体架构如图 1 所示,包括 2 个主要阶段:注册阶段和认证阶段. 注册阶段先进行特征编码器训练,如两条黑实线所示;而后训练用户判别器并生成超平面,如绿色虚线所示. 认证阶段红色圆点线表示正样本输入,蓝色点划线表示负样本输入. 在注册阶段系统使用志愿者搜集的数据进行孪生网络的预训练,使得孪生网络具备根据用户输入提取出用户的特异信息,利用已经训练好的特征编

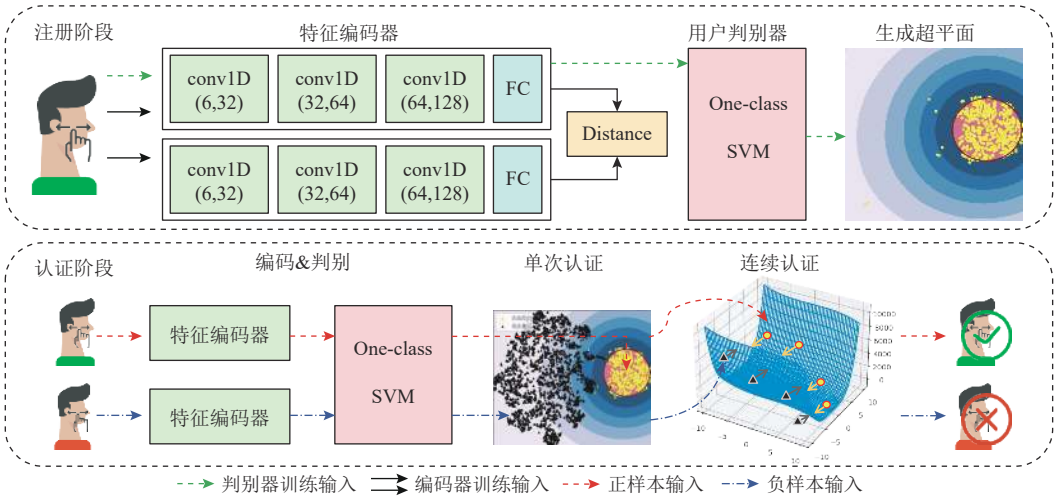


Fig. 1 Illustration of our system

图 1 本文系统示意图

码提取出该合法用户的特征并输入用户判别器中进行模型训练. 在认证阶段, 使用者完成手势活动, 对其采集之后输入特征编码器, 接着输入用户判别器, 根据其结果进行用户身份认证.

2.1 数据定义与收集

在认证之前, 首先需要定义适用于大部分耳戴式设备的面部交互手势. 其灵感来自于索尼公司的蓝牙耳机 LinkBuds 上的广域交互操作. 具体来说, 当用户采取不同的手势触摸耳朵附近的皮肤来控制多媒体内容的播放时, 手指和皮肤之间的摩擦会产生细微的振动, 这些振动可以被嵌入在智能耳机里的 IMU 传感器接收到. 本文考虑了 7 种不同的面部交互手势, 它们是点击、向左划动、向右划动、向上划动、向下划动、顺时针转圈和逆时针转圈, 如图 2 所示.

在实验中, 用户被要求佩戴智能耳机, 在面部执行 7 种交互手势. 手势振动波通过耳机和用户的头部可以被 IMU 传感器内置的三轴加速度计和陀螺仪检测到. 考虑到振动信号的传播延迟, 为确保正确捕捉振动信号, 本文方法采用滑动窗口来实现上述目标. 对于 HI226DK 型号的 IMU, 其频率为 200 Hz, 即每 5 ms 进行一次信号数据采集. 在一轮信号采集后生成的时间序列中, 其中间部分为手势振动脉冲. 本文方法使用 50 ms 的滑动窗口, 每次移动 5 ms, 可以根据频率方差的跳变和下降来识别手势振动脉冲.

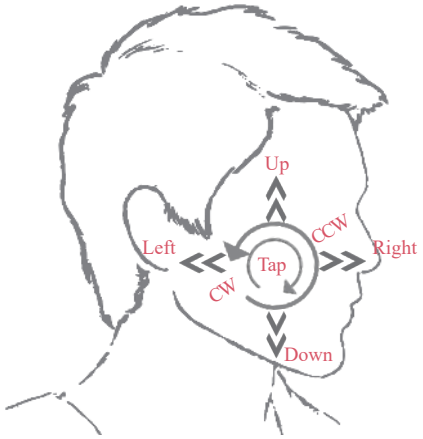


Fig. 2 Seven facial interaction gestures

图 2 7 种面部交互手势

具体而言, 振动信号在传播过程中, 受到用户行为习惯和生理特征的双重影响. 行为习惯的差异性体现在不同用户在执行相同手势时所展现的不同力度和轨迹上. 这种差异性是用户长期行为模式的积累, 具有较高的稳定性和个体性. 生理特征则涉及到用户面部的骨骼结构和脂肪含量等生物学属性, 这些因素共同决定了振动信号在面部传导过程中的特性. 由于面部骨骼结构和脂肪含量在短期内难以发生显著变化, 因此, 这些生理特征同样构成了用户身份认证中的稳定生物标志, 这些信息包含在了用户进行手势执行的过程中. 通过对用户手势信息进行处理和提取, 可以获得该用户的特异性信息并基于这种信息区分已经注册的用户和未注册用户(非法攻击者). 针对已经注册的用户, 模型提取出其手势振动信号中包含的特异信息, 根据先前注册阶段采集的注册样本判定其为合法样本. 而非法用户并未进行注册, 模型抽取出来的生物信息与注册用户差异较大, 因此会被判别器判定为离群值从而攻击失败.

2.2 注册特征编码器

注册阶段的目标是得到个性化的特征编码器以及注册后的用户判别器. 接下来对特征编码器的注册过程进行详细介绍.

合法用户被要求为每个手势提供一定量的注册

样本用于注册,同时训练样本中存在一定量的预训练样本,在注册前允许加载预训练参数以优化特征编码器的表现,而预训练样本不参与后续认证实验的模拟.

所采用的数据集是用陀螺仪和加速度计分别从x轴,y轴,z轴3个方向采集的用户手势信号,因此手势样本存在6个维度,即手势信号在陀螺仪和加速度计各自3个轴向上的变化.这种多维度的数据特性为手势识别提供了丰富的信息,故此,数据定义为

$$\boldsymbol{X}=(\boldsymbol{x}_{\mathrm{Gx}},\boldsymbol{x}_{\mathrm{Gy}},\boldsymbol{x}_{\mathrm{Gz}},\boldsymbol{x}_{\mathrm{Ax}},\boldsymbol{x}_{\mathrm{Ay}},\boldsymbol{x}_{\mathrm{Az}}\}). \tag{1}$$

其中,下标G表示陀螺仪,下标A表示加速器,下标x,y,z表示陀螺仪和加速度的三轴方向.

针对数据的这一特性,采用一维卷积神经网络捕捉空间维度的通道信息和时间维度的时序信息,并以此为基础构建特征提取器.一维卷积神经网络的公式为

$$\boldsymbol{h}[n]=f(\boldsymbol{X},\boldsymbol{W},n)=\sum_{m=-k}^{k}\boldsymbol{X}[n-m]\cdot\boldsymbol{W}[m]\,, \tag{2}$$

其中,$\boldsymbol{h}$表示卷积操作的结果,n是输出信号的索引,m是卷积核中元素的索引,$\boldsymbol{X}[n-m]$是输入信号在位置$n-m$上的值,$\boldsymbol{W}[m]$是卷积核在位置m上的权重,k是卷积核的大小.

单个样本在实际的神经网络实现中,卷积操作通常会伴随一个偏置项$\boldsymbol{b}$,并且可能应用一个激活函数ϕ,如ReLU.因此,卷积层的输出可以表示为

$$\boldsymbol{h}[n]=\phi\left(\sum_{m=-\infty}^{\infty}\boldsymbol{X}[n-m]\cdot\boldsymbol{W}[m]+\boldsymbol{b}\right). \tag{3}$$

本文设计的特征编码器的模块依次包括3层一维卷积网络,输入通道分别为6个、32个、64个,输出通道分别为32个、64个、128个.每层一维卷积后分别连接一个批量归一化层和一个ReLU激活层,最后经过一个最大池化层.结束一维卷积后,对特征向量进行扁平化,通过2048个节点和512个节点的两层全连接层作为输出层,其输出作为对比损失的输入.

因为特征编码器的目标是对用户特异特征和手势特异特征解耦合,将侧重点聚焦于不同用户的生物信息,所以如果仅采用常规深度神经网络和传统分类模型对模型进行预训练的效果并非理想,因此采用孪生网络进行预训练处理.

孪生网络架构如图3所示,每次输入2个或3个样本(取决于损失函数的设置),通过孪生的特征编码网络计算特征之间的距离.如果2个样本的标签相同,希望它们之间的距离越近越好;如果并非相同标签,那么希望它们之间距离越远越好.这需要设置一个阈值($margin$),超过这一阈值则表示足够远,损失为0.

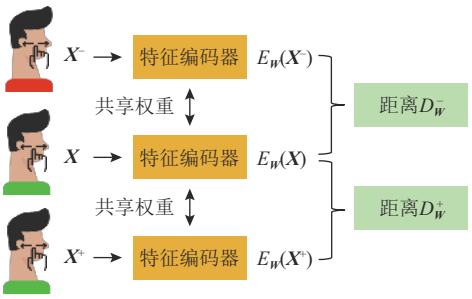


Fig. 3 Illustration of Siamese network

图3 孪生网络示意图

通常采用欧式距离作为损失度量,距离函数定义为

$$D_W(\boldsymbol{X}_1,\boldsymbol{X}_2)=\|E_W(\boldsymbol{X}_1)-E_W(\boldsymbol{X}_2)\|_2\,, \tag{4}$$

其中D_W是距离函数,E_W是特征编码器.

采用对比损失(contrastive loss)[25-27]作为孪生网络的优化损失,函数为

$$L(Y,\boldsymbol{X}_1,\boldsymbol{X}_2)=\frac{1}{2}(1-Y)D_W(\boldsymbol{X}_1,\boldsymbol{X}_2)^2+\frac{Y}{2}\mathrm{MAX}(0,margin-D_W(\boldsymbol{X}_1,\boldsymbol{X}_2)^2), \tag{5}$$

其中,$\boldsymbol{X}_1$,$\boldsymbol{X}_2$为样本向量,其中Y表示$\boldsymbol{X}_1$和$\boldsymbol{X}_2$是否属于同一类,$Y=1$表示属于同一类,$Y=0$表示属于异类,$margin$表示模型如何定义距离足够远的阈值.

$margin$的选择会影响孪生网络抽取特征的能力,针对这一超参数我们从原始的输入数据出发,发现其正则化之后的样本特征经过相加求和之后数值小于8,因此我们利用网格化的方法从[2,10]的范围内以线性步长搜索表现最好的超参数用于后续实验.

2.3 注册用户判别器

在实现特征编码器的训练后,注册阶段的另一个目标是得到注册后的用户判别器.

对于攻击者的手势通过特征编码器产生的特征向量,显然迥异于合法用户的特异性特征,可视作异常样本.因此将用户身份认证问题转移为异常值判别问题,注册后的用户判别器可以要求接受合法用户的样本,拒绝攻击者的样本.

考虑到合法用户接触设备时间远大于攻击者所能接触的时间,期待系统可以充分利用日常大规模合法用户提供的样本,用户判别器原则上应选取无监督模型以适应参数的定期更新.

经过广泛大量的比对,采用无监督的异常值判

别模型 OCSVM 作为用户判别器的基准模型，数据描述器选取稀疏奇异值分解(sparse singular value decomposition, SSVD).

OCSVM 和传统的支持向量机(support vector machine, SVM)[28]类似，其目的是优化出一个超平面，使超平面尽可能地包含所有的训练样本. 给定训练样本 $S=\{\boldsymbol{x}_i, i=1,2,\cdots,m\}$，其中的样本 $\boldsymbol{x}_i$ 为 $\boldsymbol{X}$ 经过特征编码器之后的输出，SSVD 将会找到一个尽可能小的包含绝大多数训练样本的超平面. 超平面之内的样本为正常样本，超平面之外的样本为异常样本. 利用注册样本优化后的 OCSVM 超平面降维示意图如图 4 所示. 使用 $\boldsymbol{a}$ 表示超球体的中心，R 表示超球体的半径. 求解该超平面的公式可以被表示成为以下二次优化问题：

$$\begin{aligned}&\min_{\boldsymbol{a},R}\left\{\frac{R^2}{2}+C\sum_{i=1}^{N}\xi_i\right\},\\&\text{st. } \xi_i\geqslant 0, \text{for all } i=1,2,\cdots,N,\\&\|\boldsymbol{x}_i-\boldsymbol{a}\|^2\leqslant R^2+\xi_i, \text{for all } i=1,2,\cdots,N.\end{aligned}\tag{6}$$

其中 ξ_i 表示松弛变量，允许一部分训练样本点出现在超平面之外；C 表示在最小化超球体的半径和允许一部分样本在超球面之外权衡的超参数；$\boldsymbol{a}$ 在数学上表示超球体的中心，在本系统中表示正常样本点的中心位置；R 表示超球体的半径，在本系统中表示正常样本的范围半径.

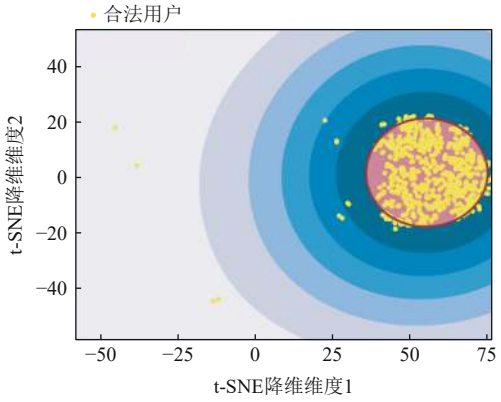


Fig. 4 Illustration of OCSVM training

图 4 OCSVM 训练示意图

引入拉格朗日算子之后得到了公式：

$$\begin{aligned}&L(\alpha,R,\boldsymbol{a},\nu,\xi)=\\&\frac{R^2}{2}+C\sum_{i=1}^{N}\xi_i-\sum_{i=1}^{N}\alpha_i\left(R^2+\xi_i-\|\boldsymbol{x}_i-\boldsymbol{a}\|^2\right)-\sum_{i=1}^{N}\nu_i\xi_i,\end{aligned}\tag{7}$$

其中 $\alpha>0, \nu_i>0$ 表示拉格朗日算子.

接下来对引入拉格朗日算子的目标函数中的 $R,\boldsymbol{a},\lambda_i$ 求偏导并令其为 0，则得到了 KKT 条件：

$$\begin{aligned}&\sum_{i=1}^{N}\alpha_i=1,\\&\boldsymbol{a}=\sum_{i=1}^{N}\alpha_i\boldsymbol{x}_i,\\&C-\alpha_i-\nu_i=0.\end{aligned}\tag{8}$$

式 (8) 中的第 2 式可以视 α_i 为 $\boldsymbol{x}_i$ 的权重，那么球心 $\boldsymbol{a}$ 就是对所有样本的加权求和. 将式 (8) 代入拉格朗日函数中，则优化函数可以写成：

$$L(\alpha)=\sum_{i=1}^{N}\alpha_i\boldsymbol{x}_i\boldsymbol{x}_i-\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\boldsymbol{x}_i\boldsymbol{x}_j\,.\tag{9}$$

原始数据几乎不太可能按照球形分布，需要使用核函数来将原始数据映射到高维空间之后再进行计算，增强模型的表达能力，引入核函数后，对偶问题可表述为：

$$\begin{aligned}&\max_{\theta}L(\alpha)=\sum_{i=1}^{N}\alpha_iK(i,i)-\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jK(i,j)\,,\\&\text{s.t.}\quad\sum_{i=1}^{N}\alpha_i=1\,,\\&0\leqslant\alpha_i\leqslant C,\forall i\,.\end{aligned}\tag{10}$$

其中 $K(i,j)$ 为核函数.

对该对偶问题的求解，可以获得全部 α，那么 R^2 计算为

$$\begin{aligned}&R^2=\|\phi(\boldsymbol{x}_i)-\boldsymbol{a}\|^2=\\&K_{k,k}-2\sum_{i=1}^{N}\alpha_iK_{k,i}+\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jK_{i,j},\end{aligned}\tag{11}$$

其中 ϕ 为映射函数.

最终决策函数为

$$f(\boldsymbol{x})=\|\phi(\boldsymbol{x})-\boldsymbol{a}\|-R^2.\tag{12}$$

如果 $f(\boldsymbol{x})<0$ 则为正常样本，$f(\boldsymbol{x})\geqslant 0$ 则为异常样本. $f(\boldsymbol{x})$ 在优化后得到了能够区分合法用户和攻击者的用户判别器. 具体的算法如算法 1 所示.

算法 1. 合法和非法用户区分算法.

输入：有 IMU 采集的手势特征 $\boldsymbol{x}$，特征编码器 *encoder*，高斯核函数 ϕ，判别器参数 $\boldsymbol{a}$，R；

输出：$y\subset(0,1)$.

① $\boldsymbol{X}_1\leftarrow$*Normalization*($\boldsymbol{x}$); /*归一化操作*/

② ***feature***$\leftarrow$*encoder*($\boldsymbol{X}_1$); /*将归一化的结果输入编码器*/

③ *distance*$\leftarrow\|\phi($***feature***$)-\boldsymbol{a}\|$;

④ if *distance*$<R$

⑤ $y\leftarrow 1$;

⑥ else

⑦ $y\leftarrow 0$;

⑧ end if

2.4 用户认证

在得到注册后的特征编码器和用户判别器之后，对于每个未知样本，均可通过特征编码，根据特征向量是否被用户判别器生成的超平面包含来判定样本来自合法用户还是攻击者.

合法用户和攻击者的样本特征降维示意图如图 5 所示，特征编码器通过注册样本个性化，已实现合法用户与攻击者的分离，用户判别器在此基础上生成了分隔二者的超平面，实现了合法用户身份的认证.

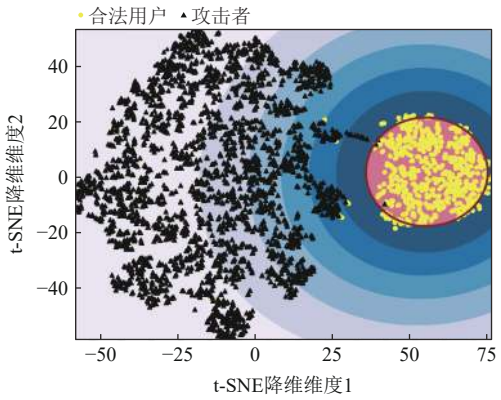


Fig. 5 Illustration of OCSVM authentification

图 5 OCSVM 身份认证示意图

在耳戴式设备身份认证系统中，确保高精度检测攻击行为的同时，又要尽量减少对合法用户体验的干扰，是系统设计中的一个关键挑战.

基于这一考虑，本文认为单一的手势样本进行身份验证难以满足上述要求. 这是因为，单一样本可能受到偶然性因素的影响，导致认证结果的不稳定. 本文从集成学习算法[29-32]中获得启发，集成学习算法是机器学习算法中的重要分支和优秀的优化范式，通过构建一系列独立或者关联的分类器，并在训练过程中通过预定的规则融合输出，这一方法被认为可以广泛增强单一分类器的表现力，其中投票法是集成学习算法中最常见的融合方法，可分为硬投票和软投票 2 种，硬投票根据分类器输出的类别进行多数表决，软投票根据分类器输出的预测向量进行加权判断. 为了提高认证的准确性和鲁棒性，本文设计出基于超平面距离的新颖加权投票方法如图 6 所示，相较于集成学习算法的多模型投票，基于用户可以多次输入的特点，本系统采取多次输入进行投票的方式实现了连续身份认证.

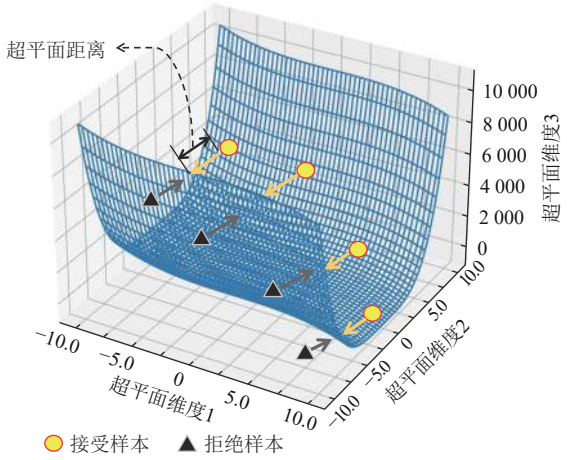


Fig. 6 Illustration of hyperplane-based voting

图 6 基于超平面的投票示意图

该策略的核心思想是，通过对用户连续执行的多个手势样本进行身份验证，并将这些认证结果进行综合分析，从而实现对用户身份的更准确判断. 由于连续样本均来自同一用户，因此，这种方法能够在一定程度上降低偶然误差的影响，并且增强系统对欺骗行为的识别能力.

在图 6 中，样本点距离超平面越远，对应于接受或拒绝数据点的置信度越高.

通过计算与超球体球心的距离，并与半径对比，设计出基于超平面距离的加权投票方法，以提高正确性. 公式化描述为

$$f(\boldsymbol{x})=\frac{\left(\sum_{i=1}^{N}\|\phi(\boldsymbol{x}_i)-\boldsymbol{a}\|\right)}{N}-R^2. \tag{13}$$

如果$f(\boldsymbol{x})>0$则接受为合法用户，反之判定为攻击者. 采取基于超平面距离的加权投票方法相对于硬投票的方式可以减少低置信样本对于投票结果的影响，使得最终的投票结果倾向于接受高置信度样本的输出结果.

3 实 验

本节对实验设置、攻击者模拟设置进行详细介绍，通过大量的实验来评估系统的性能，并演示每个关键组件的功效.

3.1 实验设置

数据采集选用型号为 HI226DK 的 IMU 板，连接在用户的耳廓上，模拟佩戴耳戴式设备，收集不同面部交互手势产生的振动信号.

为了评估本文方法，本实验招募了 20 名志愿者，

志愿者基本信息如表 2 所示，包括 8 名女性和 12 名男性，年龄为 18～28 岁，均为来自学校的在校学生和教职工. 志愿者在不被要求固定姿势的前提下执行预先定义的耳戴式设备交互手势. 这些手势包括点击、上划、下划、左划、右划、顺时针绕圈和逆时针绕圈.

本实验从志愿者身上收集 2 次数据，在每次数据收集过程中，所有志愿者执行 7 种预先定义的交互手势，每个参与者为每个手势提供 100 个样本，总共 28 000 个样本. 数据集以 csv 文件存储，其格式如表 3 所示. 每条手势振动信号数据由 IMU 内置的加速度计和陀螺仪完成，输出单位分别为“m/s^3”和“(°)/s”，采集的时间粒度为 5 ms.

每条手势振动信号的数据形状为 1×3 600，由 6 个长度为 600 的时间序列组成，分别对应了加速度计和陀螺仪的 x, y, z 三个轴. 对于后续的评估，本文使用第 1 次采集的样本作为训练数据，第 2 次采集的样本作为测试数据. 考虑到智能耳机通常在室内使用，数据采集是在室内环境中进行的.

Table 2 Basic Information of the 20 Volunteers

表 2 20 名志愿者的基本信息

身高/cm			体重/kg			年龄			职业			性别	
160～170	170～180	180～190	50～60	60～70	70～80	18～22	22～25	25～28	在校学生	教职工	校外人士	男	女
6 人	12 人	2 人	4 人	14 人	2 人	15 人	4 人	1 人	19 人	1 人	0 人	12 人	8 人

Table 3 Data Set Format

表 3 数据集格式

序号	加速度计 x 轴			加速度计 y 轴			加速度计 z 轴			陀螺仪 x 轴			陀螺仪 y 轴			陀螺仪 z 轴			标签
手势 1	0.964	…	0.972	−0.284	…	−0.265	0.053	…	0.057	−0.8	…	−4.9	0.1	…	0.7	−0.4	…	−0.1	0
手势 2	0.968	…	0.980	−0.261	…	−0.257	0.022	…	0.068	−1.6	…	−2.4	−1	…	0.3	−0.5	…	−0.3	1
⋮	⋮		⋮	⋮		⋮	⋮		⋮	⋮		⋮	⋮		⋮	⋮		⋮	⋮
手势 n	0.976	…	0.984	−0.261	…	−0.265	0.034	…	0.04	−0.8	…	2.2	0.4	…	−1.1	−0.1	…	−0.1	19

模型测试包括 3 个主要步骤：模型分类能力评估、基于模仿攻击的单次认证能力评估和基于模仿攻击的连续认证能力评估. 这些模型使用 PyTorch 框架编码，在一台带有 2 个 12 核 2.3 GHz Intel Xeon Gold CPU 和 4 个 NVIDIA GeForce RTX 3 090 GPU 的服务器上进行训练和运行.

本文选取的定量分析指标包括：

1)分类准确率(classification accuracy)：正确分类用户的概率.

$$Accuracy=\frac{TP+TN}{TP+TN+FN+FP}. \tag{14}$$

2)假阳率(false positive rate, FPR)：攻击者被误认为合法用户的比例.

$$FPR=\frac{FP}{TN+FP}. \tag{15}$$

3)假阴率(false negative rate, FNR)：合法用户被误认为攻击者而被拒绝的比例.

$$FNR=\frac{FN}{TP+FN}. \tag{16}$$

4)等错率(equal error rate, EER)：假阳率与假阴率权衡之后的平均错误率，在上下文中与认证准确率同义，方便理解本系统实验结果.

$$EER=\frac{FPR+\alpha\times FNR}{2}. \tag{17}$$

式中 TP 代表真正例，即一个实例是正类且被判定为正类；TN 代表真负例，即一个实例是负类且被判定为负类；FP 代表假正例，即一个实例是负类但被判定为正类；FN 代表假负例，即一个实例是正类但被判定为负类. α 表示在假阳率和假阴率之间进行权重平衡，一般 α=1.

3.2 攻击者模拟

在本节中，本文主要探讨了物联网边端耳戴式设备的用户认证，因此针对离线模式下已完成训练的边端设备设计攻击场景.

本文系统的核心应用场景是防御模拟攻击，即在耳戴式设备已经完成用户注册的情况下，潜在的攻击者无法渗透至注册阶段，因此只能尝试通过身份欺骗来误导认证系统. 在这种情况下，之前的许多工作考虑到了多种欺骗攻击模式，其中最具代表性的就是模拟攻击和重放攻击 2 种模式[33-36]，以及二者融合进行的混合攻击[33]. 本文重点考虑上述 2 种具有代表性的身份欺骗攻击手段.

首先，针对重放攻击，本文认为由于皮肤上的手势产生的振动信号具有高度的复杂性和随机性，这

使得信号的记录和回放对于攻击者而言存在极大的难度. 具体而言, 振动信号不仅包含了手势的物理运动信息, 还蕴含了用户的生理特征和行为习惯, 这些因素共同构成了一个多维度的生物识别特征. 因此, 本文设计的系统具备了较高的抵抗重放攻击的能力.

其次, 就模拟攻击而言, 潜在的攻击者需要经历一系列复杂的步骤才能尝试对系统进行欺骗. 攻击者首先需要密切观察并学习合法用户的耳戴式设备的交互手势, 包括手势的运动轨迹、速度、力度等细节特征. 接着, 攻击者必须通过一定时间的练习, 以模仿合法用户的手势行为. 最终, 攻击者还需获得合法用户的耳塞设备, 并尝试在设备上复现用户的手势动作, 以欺骗认证系统. 在实验阶段本文主要考虑系统受到模拟攻击.

3.3 模型分类能力评估

参照 3.1 节的量化指标, 选取许多现有的机器学习模型与本文的特征编码器进行了比较, 包括卷积神经网络、极限梯度提升树(XGBoost)、随机森林(random forest, RF)、支持向量机(SVM)和长短期记忆网络(long short-term memory, LSTM), 以展示其在用户分类和认证方面的性能.

本文认为特征编码器对不同用户的分类能力越强, 其提取用户特异信息的能力就越强. 为了测试不同模型分类能力的强弱, 本文在用户分类评估阶段纳入了 20 名志愿者的全部数据, 以确保评估的全面性和可靠性.

在构建训练数据集的过程中, 本文采取了一种增量式的样本积累方法. 具体而言, 本文从每个手势的 1 个样本开始, 逐步增加样本数量, 直至每个手势达到 100 个样本. 这种方法允许更准确地评估模型在不同数据规模下的性能表现. 在每个阶段, 样本的增加相当于每个用户在数据集中的样本数量从 7 个递增至 700 个.

在测试过程中, 为了尽可能考虑到所有的情况, 本文使用了全部的样本进行测试. 这样可以减少由于样本选择偏差带来的影响, 让测试结果更加准确和可靠.

然而, 在进行实验时, 总会存在一些偶然因素的影响. 为了避免这些偶然因素带来的影响, 进行了 10 次独立实验, 以获得更稳定和平衡的指标结果. 这些结果如图 7 所示.

通过对实验结果的分析发现, 首先, RF 和 XGBoost 算法在小样本训练集中有较高的准确率, 这意味着这 2 种算法在处理小样本数据时有着出色的性能. 其次, 在中等规模的数据中, CNN 表现出了不

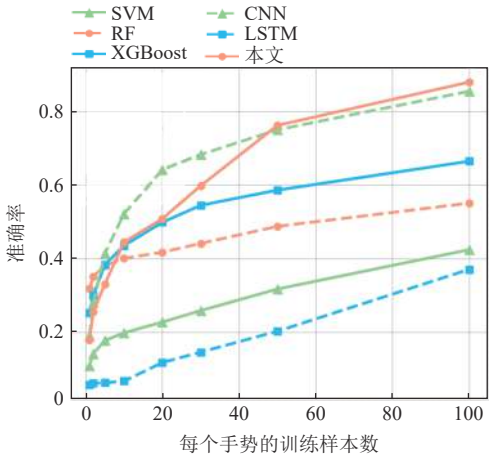


Fig. 7 User classification accuracy of different number of training samples

图 7 不同数量训练样本的用户分类准确率

错的效果, 这可能是因为 CNN 在处理这种规模的数据时, 可以更好地编码出数据的特征. 最后, 本文选用的孪生网络方法在较大规模的数据集上获得了最高的效果, 说明在较大数据规模的情况下, 孪生网络的特征提取能力效果是当前对比模型中最佳的.

但是, 孪生网络在小样本中的优势并不明显. 这可能是因为孪生网络需要采样足够多的样本进行对比学习, 以抽取样本中的生物特征进行分类. 因此, 当样本量不足时, 孪生网络的表现可能就会受到一定的影响.

3.4 基于模拟攻击的认证实验

在认证实验的评估中, 志愿者被分成 2 组. 第 1 组由 15 名志愿者组成, 从第 1 组收集的数据用于预训练特征编码器, 剩下的 5 名志愿者组成第 2 组.

在身份认证评估过程中, 第 2 组的 5 名志愿者依次被选为合法用户来注册身份验证系统, 其余 4 名志愿者作为欺骗者, 执行模拟攻击. 本文的系统必须做的就是认证出合法用户并阻止欺骗者. 具体来说, 在注册阶段, 仅允许被选作合法用户的志愿者提供注册样本进行注册, 即提供手势样本特征微调特征编码器, 以及生成能区分合法用户和欺骗者的用户判别器; 在认证阶段, 纳入 5 名志愿者全部手势样本进行实验, 通过统计模型在合法用户与欺骗者在随机样本上的表现, 评估模型接受合法用户同时拒绝欺骗者的认证能力.

3.4.1 单次认证实验

在单次认证实验中, 旨在评估本文方法在单次手势执行的身份认证上效果如何, 因此模型接收单个手势样本即可进行用户身份判别.

选取孤立森林(isolation forest, IF)、高斯异常检

测(Gaussian anomaly detection, GAD)模型、局部异常因子(local outlier factor, LOF)等主流的异常值检测模型作为用户判别器，将当前成熟的耳戴式设备用户认证模型 BudsAuth[9] 与本文方法作对比.

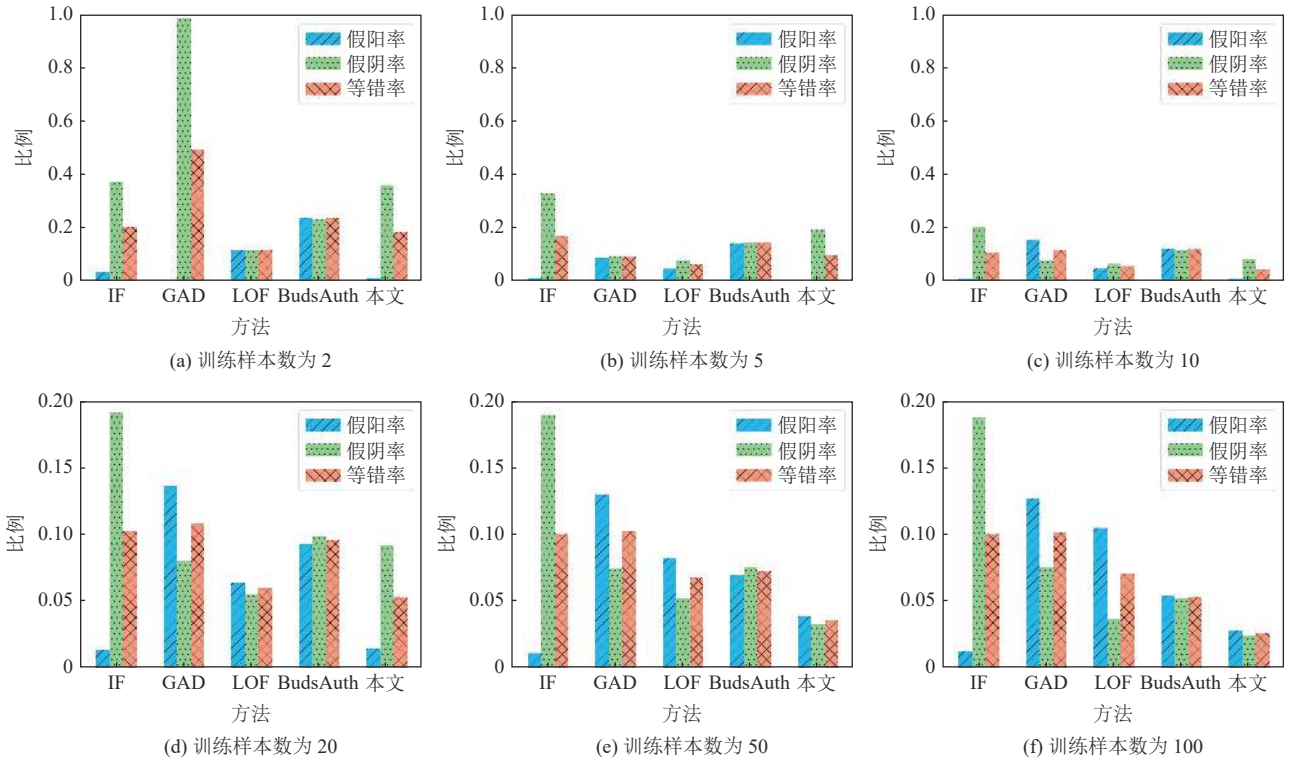
具体结果如图 8 所示，相较于采用传统的分类方法进行验证，异常检测模型会表现出更强的鲁棒性和安全性. 分类模型只要其对应类别数值比其他类别数值稍高即可欺骗认证机制，但是异常检测模型不仅要求其对应用户类别数值比其他类别数值更高，而且要求其他类别数值均处于较低状态，换言之更关注样本的多维度特征，而不仅仅是数值特征.

针对不同的异常检测模型进行了多组试验，分别测试在不同数量的训练样本上进行机器学习异常检测，观察其不同的指标，主要包括假阳率，假阴率和等错率.

如图 8 所示，可以看出 IF 方法和 GAD 方法以及本文选取的用户判别器 OCSVM 的指标，随着训练数据量的上升都有所下降，但是在只有少量样本数据的情况下依旧获得了不错效果，这也说明异常检测模型相较于分类模型具备较好的泛化性和鲁棒性. LOF 方法在小样本数据的情况下效果更好，主要原因在于经过特征提取器之后的样本特征在样本空间中更加集中、密度更高，因此小样本即可以获得比较高的等错率，但是随着所有样本引入之后部分离异常样本距离较近的正常样本更容易干扰检测，本系统不采用 LOF 方法的主要原因在于其表现不太稳定，如果特征提取器抽取样本特征能力较差，LOF 方法的表现不如 OCSVM，换句话说 LOF 方法方差较大，针对不同用户的泛化性和鲁棒性较差，而且该方法时间复杂度高，综合来看，OCSVM 的准确率的综合指标表现更好，在所有样本进行训练的情况下，在平滑权衡后的等错率为 0.026 7，即 97.33% 认证准确率. 因此，选择 OCSVM 作为系统的认证模型. 本文提出的模型相较于 BudsAuth，采用孪生网络架构作为编码器相较于域对抗模型所需要进行的训练时间更少，用户认证效果更优越. 不同于 BudsAuth 直接使用 z-score 算法进行离群值检测，本文考虑更多异常检测算法用作判别器对比，通过采用核函数技巧和 SVM 将输入特征映射到高维空间，使得数据线性可分，增强了模型的鲁棒性，能较好地应对未知数据，从图 8 可以看出，本文的模型相较于 BudsAuth 在同样的数据量上表现更好，等错率更低.

3.4.2 连续认证实验

在连续认证实验中，旨在评估本文方法在连续手势执行的身份认证能力，这主要考验模型输入多个样本手势特征后，能否结合先后不同手势的不同



(a) 训练样本数为 2　(b) 训练样本数为 5　(c) 训练样本数为 10

(d) 训练样本数为 20　(e) 训练样本数为 50　(f) 训练样本数为 100

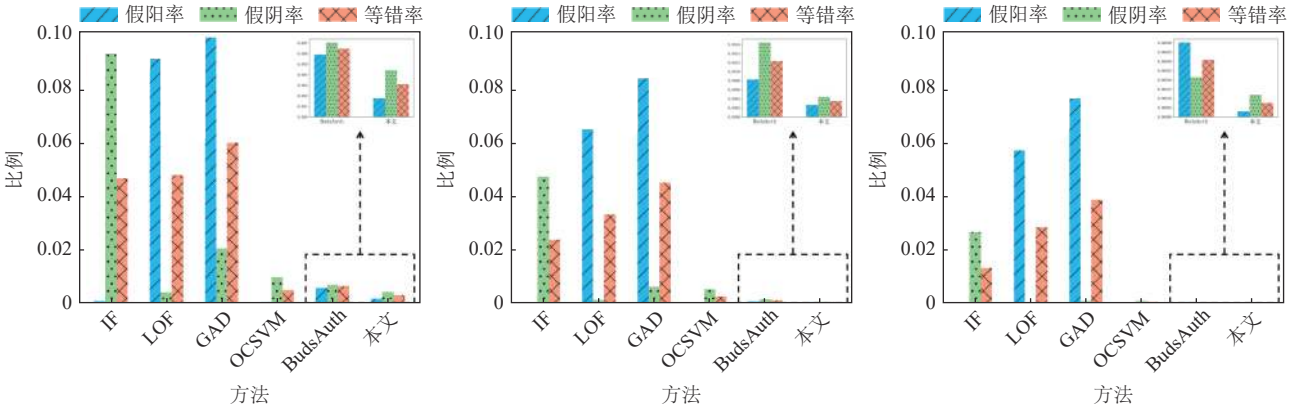
Fig. 8　*FPR*, *FNR*, *ERR* of different number of samples

图 8　不同数量训练样本的假阳率、假阴率、等错率

置信程度,综合决策给出可靠的判别结果.

基于 OCSVM 可以通过样本距离超平面的远近输出模型对于该样本点的类别的置信度,由此本文受集成学习算法多模型单样本进行投票思想的影响,考虑到同一模型多次输入进行加权投票以减少某些样本由于噪声存在导致模型分类错误的现象,减少预测方差. 本文提出了一种基于超平面距离的加权投票策略,与上述模型以及现有效果较好的模型 BudsAuth 作对比.

从图 9 可以看到通过多次投票法在几乎所有异常检测模型上都有较高的提升,由于模型本身效果优越,即使硬投票也有不错的效果提升,而硬投票由于每次投票只能+1 或者−1 操作,对高置信度样本和低置信度样本具有相同的影响,这显然会影响认证的准确率,而采用软投票方法相当于赋予了低置信度样本较低的权重,有效避免了误判,因此可以看到基于 OCSVM 的用户判别器和 BudsAuth 的软投票结果最优,在 7 轮投票的情况下 OCSVM 实现了 7×10^{-5} 的等错率,即 99.993% 的认证准确率,超越了当前最好模型 BudsAuth 的效果.



(a) 投票轮数为3 (b) 投票轮数为5 (c) 投票轮数为7

Fig. 9 *FNR*, *FPR*, *EER* values for continuous authentication in different voting rounds

图 9 不同投票轮次的连续认证的假阳率、假阴率、等错率

3.5 部署实验

为在真实场景下进行本文认证方法的系统开销测试,并且弥补现阶段耳戴式设备算力不足的差距,本节以智能手机作为载体,展示本文方法部署在安卓系统上的实际表现,以验证其实际应用价值. 此类替代仅限于部署测试,上文分类及认证实验所采用的数据集真实来源于耳戴式设备搭载的 IMU 收集,而非智能手机. 系统部署测试可作为本文方法实际应用中功耗、内存占比和推理时间的参考,从而与现有的方法进行比对.

为了测试本文方法在物联网边端设备的实际应用效果,将本文方法部署到选用了 OPPO A93s 智能手机作为离线测试平台. 该手机配备了天玑 700 八核处理器,拥有 256 GB 的存储空间和 5 000 mAh 的电池容量. 在存储空间的占用方面,通过 Android Studio 开发的环境进行编码和部署后,本文的用户身份智能认证系统仅占用了 41.12 MB 的存储空间,而在用户注册过程中,系统额外增加了 3.24 MB 的用户数据存储需求. 即便如此,总的存储空间占用率仅为手机总内存的 0.02%,这一存储开销几乎可以忽略不计,显示出系统在存储效率方面的优化成果.

本文设定了一组极端使用场景的假设条件: 用户每天佩戴耳戴式设备长达 10 h,并在此期间持续不断地与设备进行手势交互,以每秒 1 次的频率累计产生 36 000 次系统调用. 在这一假设下,本文对系统的日常能耗进行了理论估算,得出系统每日的能量消耗量大约为 46.8 J. 为了将这一能耗数据转换为更直观的电池消耗百分比,本文以 HUAWEI FreeBuds Pro 3 耳机为例,该款耳机配置电池盒容量为 510 mAh,本文发现系统的日常能耗仅占其电池盒容量的 1.6%. 上述结果表明,即便在极端使用条件下,系统对电量的消耗也是微乎其微的,在完全可以接受的范围内.

如表 4 所示,编码器采用了深度神经网络模型,占据大部分能耗,要求更大的内存和运行速度,而判

Table 4 Deployment Testing on Smartphones

表 4 在智能手机上的部署测试

系统模块	耗电量/mJ	内存消耗/MB	推理时间/ms
特征编码器	1.06	42.1	32
用户判别器	0.24	12.5	12
总体	1.30	54.6	44

别器的消耗只是使用机器学习算法，运算开销更小，总体而言本文系统开销满足日常使用需求.

4 结　论

本文提出了一种面向耳戴式设备的用户安全连续认证方法. 该方法基于内置 IMU 采集面部手势振动信号，能够更加方便有效地对用户特异性信息进行显著编码，通过设计新颖的加权投票策略实现连续的身份认证. 在真实数据集上进行实验，结果表明本文方法具有更高的准确率，在模型编码和身份认证能力方面均有优势，能够为用户提供更加流畅和更高级别的安全认证，具有较高的实际应用潜力.

但是本文方法还存在需要完善的方面，未来的研究工作中可以进一步深入探讨：1）耳戴式设备在使用过程中存在头部位移和音乐播放干扰，可以进一步优化降噪处理提升模型表现；2）如何利用用户日常使用过程中的数据，综合考虑数据存储、算力限制、更新周期等因素，对模型参数进行流式更新.

作者贡献声明：王勇负责论文构思、论文指导和论文修改；熊毅负责论文构思，完成实验并撰写部分论文；杨天宇提供论文主体思路，撰写和修改论文；沈益冉提供数据及实验平台.王勇和熊毅具有相同贡献.

参　考　文　献

[1] The Gates Notes LLC. AI is about to completely change how you use computers [EB/OL]. [2024-05-27]. https://www.gatesnotes.com/AI-agents

[2] Apple Inc. Apple AirPods [EB/OL]. [2024-05-27]. https://www.apple.com. cn/airpods/

[3] Huawei Terminal Co., Ltd.. Huawei Freebuds official website [EB/OL]. [2024-05-27]. https://consumer.huawei.com/cn/audio/
(华为终端有限公司. 华为 Freebuds 官网 [EB/OL]. [2024-05-27]. https://consumer.huawei.com/cn/audio/)

[4] 2024 Sony China Co., Ltd.. LinkBuds official website [EB/OL]. [2024-05-27]. https://www.sonystyle.com.cn/products/headphone/index. html (in Chinese)
(2024 索尼(中国)有限公司. LinkBuds 官网 [EB/OL]. [2024-05-27]. https://www.sonystyle.com.cn/products/headphone/index.html)

[5] Zhang Yuqing, Zhou Wei, Peng Anni. Survey of Internet of things security[J]. Journal of Computer Research and Development, 2017, 54(10): 2130−2143 (in Chinese)
(张玉清, 周威, 彭安妮. 物联网安全综述[J]. 计算机研究与发展, 2017, 54(10): 2130−2143)

[6] Zhou Jun, Shen Huajie, Lin Zhongyun, et al. Research advances on privacy preserving in edge computing[J]. Journal of Computer Research and Development, 2020, 57(10): 2027−2051 (in Chinese)
(周俊, 沈华杰, 林中允, 等. 边缘计算隐私保护研究进展[J]. 计算机研究与发展, 2020, 57(10): 2027−2051)

[7] Dong Xiaolei. Advances of privacy preservation in Internet of things[J]. Journal of Computer Research and Development, 2015, 52(10): 2341−2352 (in Chinese)
(董晓蕾. 物联网隐私保护研究进展[J]. 计算机研究与发展, 2015, 52(10): 2341−2352)

[8] Liu Qixu, Jin Ze, Chen Canhua, et al. Survey on Internet of things access control security[J]. Journal of Computer Research and Development, 2022, 59(10): 2190−2211 (in Chinese)
(刘奇旭, 靳泽, 陈灿华, 等. 物联网访问控制安全性综述[J]. 计算机研究与发展, 2022, 59(10): 2190−2211)

[9] Bromley J, Guyon I, LeCun Y, et al. Signature verification using a "Siamese" time delay neural network [C] // Proc of the 6th Int Conf on Neural Information Processing Systems (NIPS'93). San Francisco, CA: Morgan Kaufmann Publishers Inc., 1993: 737–744

[10] Wang Huiyong, Tang Shijie, Ding Yong, et al. Survey on biometrics template protection[J]. Journal of Computer Research and Development, 2020, 57(5): 1003−1021 (in Chinese)
(王会勇, 唐士杰, 丁勇, 等. 生物特征识别模板保护综述[J]. 计算机研究与发展, 2020, 57(5): 1003−1021)

[11] Tax D M, Duin R P. Support vector data description[J]. Machine Learning, 2004, 54: 45−66

[12] Lissermann R, Huber J, Hadjakos A, et al. EarPut: Augmenting ear-worn devices for ear-based interaction [C] // Proc of the 26th Australian Computer-Human Interaction Conf on Designing Futures: The Future of Design. New York: Association for Computing Machinery, 2014: 300−307

[13] Xu Xuhai, Shi Haitian, Yi Xin, et al. EarBuddy: Enabling on-face interaction via wireless earbuds [C] // Proc of the 2020 CHI Conf on Human Factors in Computing Systems. New York: Association for Computing Machinery, 2020: 1−14

[14] Wang Yong, Yang Tianyu, Wang Chunxiao, et al. BudsAuth: Toward gesture-wise continuous user authentication through earbuds vibration sensing[J]. IEEE Internet of Things Journal, 2024, 11(12): 22007−22020

[15] Qamar N, Siddiqui N, Ehatisham-ul-Haq M, et al. An approach towards position-independent human activity recognition model based on wearable accelerometer sensor[J]. Procedia Computer Science, 2020, 177: 196−203

[16] Lu Chris Xiaoxuan, Du Bowen, Zhao Peijun, et al. Deepauth: In-situ authentication for smartwatches via deeply learned behavioural biometrics [C] // Proc of the 2018 ACM Int Symp on Wearable Computers. New York: Association for Computing Machinery, 2018: 204−207

[17] Li Feng, Zhao Jiayi, Yang Huan, et al. VibHead: An authentication scheme for smart headsets through vibration[J]. ACM Transactions on Sensor Networks, 2024, 20(4): 1−12

[18] Liu Jianwei, Song Wenfan, Shen Leming, et al. Secure user verification and continuous authentication via earphone IMU[J]. IEEE Transactions on Mobile Computing, 2023, 22(11): 6755−6769

[19] Ma D, Ferlini A, Mascolo C. OESense: Employing occlusion effect for in-ear human sensing [C] //Proc of the 19th Annual Int Conf on Mobile Systems, Applications, and Services. New York: Association for Computing Machinery, 2021: 175−187

[20] Wang Zi, Ren Yili, Chen Yingying, et al. ToothSonic: Earable authentication via acoustic toothprint [C] // Proc of ACM Int Conf on Interactive Mobile, Wearable and Ubiquitous Technologies. New York: Association for Computing Machinery, 2022: 1−24

[21] Wang Zi, Wang Yilin, Yang Jie. EarSlide: A secure ear wearables biometric authentication based on acoustic fingerprint [C] // Proc of ACM Int Conf on Interactive Mobile, Wearable and Ubiquitous Technologies. New York: Association for Computing Machinery, 2024: 1−29

[22] Hu Changshuo, Ma Xiao, Ma Dong, et al. Lightweight and non-invasive user authentication on earables [C] // Proc of the 24th Int Workshop on Mobile Computing Systems and Applications. Newport Beach, CA: Association for Computing Machinery, 2023: 36−41

[23] Li Jiao, Liu Yang, Li Zhenjiang, et al. EarPass: Continuous user authentication with in-ear PPG [C] // Proc of the 2023 ACM Int Joint Conf on Pervasive and Ubiquitous Computing & the 2023 ACM Int Symp on Wearable Computing. Cancun, Quintana Roo, Mexico: Association for Computing Machinery, 2023: 327−332

[24] Choi S, Yim Junghwan, Jin Yincheng, et al. EarPPG: Securing your identity with your ears [C] // Proc of the 28th Int Conf on Intelligent User Interfaces. Sydney, NSW: Association for Computing Machinery, 2023: 835−849

[25] Wang Qihong, Jia Hongjie, Huang Longxia, et al. Semantic contrastive clustering with federated data augmentation[J]. Journal of Computer Research and Development, 2024, 61(6): 1511−1524 (in Chinese)
(王气洪, 贾洪杰, 黄龙霞, 等. 联合数据增强的语义对比聚类[J]. 计算机研究与发展, 2024, 61(6): 1511−1524)

[26] Huang Xuejian, Ma Tinghuai, Wang Gensheng. Multimodal learning method based on intra- and inter-sample cooperative representation and adaptive fusion[J]. Journal of Computer Research and Development, 2024, 61(5): 1310−1324 (in Chinese)
(黄学坚, 马廷淮, 王根生. 基于样本内外协同表示和自适应融合的多模态学习方法[J]. 计算机研究与发展, 2024, 61(5): 1310−1324)

[27] Du Jinming, Sun Yuanyuan, Lin Hongfei, et al. Conversational emotion recognition incorporating knowledge graph and curriculum learning[J]. Journal of Computer Research and Development, 2024, 61(5): 1299−1309 (in Chinese)
(杜金明, 孙媛媛, 林鸿飞, 等. 融入知识图谱和课程学习的对话情绪识别[J]. 计算机研究与发展, 2024, 61(5): 1299−1309)

[28] Bao Han, Wang Yijie. A fast construction method of the erasure code with small cross-cloud data center repair traffic[J]. Journal of Computer Research and Development, 2023, 60(10): 2418−2439 (in Chinese)
(包涵, 王意洁. 低跨云数据中心修复流量的纠删码的快速构造方法[J]. 计算机研究与发展, 2023, 60(10): 2418−2439)

[29] Ganaie M A, Hu M, Malik A K, et al. Ensemble deep learning: A review[J]. Engineering Applications of Artificial Intelligence, 2022, 115: 105−151

[30] Zhao Langcheng, Lyu Rui, Lin Qi, et al. mmArrhythmia: Contactless arrhythmia detection via mmWave sensing [C] // Proc of the 24th ACM Int Conf on Mobile Computing and Networking. New York: Association for Computing Machinery, 2024: 1−25

[31] Wang Yuexin, Zheng Jie, Wang Danni, et al. Multi-objective planning model based on the soft voting ensemble learning algorithm [C] // Proc of the 2023 IEEE Int Conf on Electrical, Automation and Computer Engineering (ICEACE). Piscataway: IEEE, 2023: 1104−1107

[32] Guo Husheng, Zhang Yang, Wang Wenjian. Two-stage adaptive ensemble learning method for different types of concept drift[J]. Journal of Computer Research and Development, 2024, 61(7): 1799−1811 (in Chinese)
(郭虎升, 张洋, 王文剑. 面向不同类型概念漂移的两阶段自适应集成学习方法[J]. 计算机研究与发展, 2024, 61(7): 1799−1811)

[33] Xie Yadong, Li Fan, Wu Yue, et al. User authentication on farable devices via bone-conducted occlusion sounds[J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(4): 3704−3718

[34] Lee S, Choi W, Lee D H. The vibration knows who you are! A further analysis on usable authentication for smartwatch users[J]. Computers & Security, 2023, 125(C): 103−040

[35] Wang Zi, Yang Jie. Ear wearable (earable) user authentication via acoustic toothprint [J/OL]. 2022 [2024-05-27]. https://api.semanticscholar.org/CorpusID:248218666

[36] Xu Xiangyu, Yu Jiadi, Chen Yingying, et al. TouchPass: Towards behavior-irrelevant on-touch user authentication on smartphones leveraging vibrations [C] //Proc of the 26th Annual Int Conf on Mobile Computing and Networking (MobiCom'20). New York: Association for Computing Machinery, 2020: 1−13



Wang Yong, born in 1983. PhD, associate professor. Senior member of CCF. His main research interests include artificial intelligence, privacy computing, and Internet of things.

王 勇, 1983 年生. 博士, 副教授. CCF 高级会员. 主要研究方向为人工智能、隐私计算、物联网.



Xiong Yi, born in 2003. Bachelor. His main research interests include smart wearable devices, intelligent Internet of things, and machine learning.

熊 毅, 2003 年生. 学士. 主要研究方向为智能可穿戴设备、智能物联网、机器学习.



Yang Tianyu, born in 2002. Bachelor. His research interests include smart wearable devices, intelligent Internet of things, and machine learning.

杨天宇, 2002 年生. 学士. 主要研究方向为智能可穿戴设备、智能物联网、机器学习.



Shen Yiran, born in 1987. PhD, professor. Member of CCF, senior member of IEEE. His main research interests include mobile computing and virtual reality.

沈益冉, 1987 年生. 博士, 教授. CCF 会员, IEEE 高级会员. 主要研究方向为移动计算、虚拟现实.