基于行为生物特征的自由手势认证增强方案

姚景予[1, 2, 3]，汪定[1, 2, 3*]，李昕杨[1]
（1. 南开大学密码与网络空间安全学院，天津 300350；
2. 数据与智能系统安全教育部重点实验室，天津 300350；
3. 天津市网络与数据安全技术重点实验室，天津 300350）

摘要：由于拥有更复杂的键入模式和更大的密钥空间，自由手势在智能设备环境中被视为有望取代传统口令认证的新型身份认证方案。然而，最新研究表明，在线字典猜测攻击可以在 20 次尝试内攻破 10.33% 的自由手势，这对自由手势认证系统的安全性和可用性造成了潜在威胁。为了设计更加安全且实用的自由手势认证方案，基于用户的行为生物特征构造了一个增强自由手势认证系统 StrokeFG。除了识别手势轨迹相似度之外，StrokeFG 能够利用用户输入自由手势的行为生物信息进一步验证用户的身份。实验结果表明，在相同类别手势的测试集中，StrokeFG 报告了 5.64% 和 6.81% 的等错误率。相较于基线，StrokeFG 的假阳率降低了 86.56%~91.24%。在攻击者已知手势的前提下，StrokeFG 在 $\alpha \in [0.1, 0.4]$ 区间内提供了 6.50~8.21 比特的安全水平，有效提升了基线模型的安全边界。

关键词：自由手势；行为生物认证；智能移动设备；深度学习
中图分类号：TP391.4，TP393.08 **文献标志码：**A
DOI: 10.20172/j.issn.2097-3136.250403

Enhanced free-form gesture authentication scheme based on behavioral biometric features

YAO Jingyu[1, 2, 3], WANG Ding[1, 2, 3*], LI Xinyang[1]
（1. College of Cryptology and Ciber Science, Nankai University, Tianjin 300350, China；
2. Tianjin Key Laboratory of Network and Data Security, Tianjin 300350, China；
3. Key Laboratory of Data and Intelligent System Security, Tianjin 300350, China）

Abstract: Free-form gesture authentication, offering more complex input patterns and a larger key space, is considered a promising alternative for password authentication in mobile environments. However, recent research reveals that online dictionary attacks can compromise over 10.33% of gestures within 20 attempts, posing a potential threat to the security and usability of gesture authentication systems. To build a more secure yet practical solution, we present StrokeFG, an enhanced free-form gesture authentication scheme that leverages behavioral biometrics. In addition to measuring trajectory similarity, StrokeFG exploits users' unique behavioral traits exhibited during gesture input to verify identity. Experiments on same-category gesture datasets show that StrokeFG achieves an equal-error rate of 5.64%~6.81%, reducing the recognition error of the baseline by 64.25%~70.67%. Against an adversary who knows the gesture, StrokeFG achieves 6.50~8.41 bits of α-guesswork entropy under online dictionary attacks, effectively raising the security boundary of the baseline model.

Keywords: free-form gestures; behavioral biometrics authentication; smart mobile devices; deep learning

* **通信作者**. E-mail: wangding@nankai.edu.cn

基金项目：国家自然科学基金（62222208）；天津市自然科学基金（25JCJQJC00230）

引用格式：姚景予，汪定，李昕杨. 基于行为生物特征的自由手势认证增强方案 [J]. 网络空间安全科学学报，2025，3（4）：29 - 42.

Citation Format：YAO J Y，WANG D，LI X Y. Enhanced free-form gesture authentication scheme based on behavioral biometric features[J]. Journal of Cybersecurity，2025，3（4）：29 - 42.

0 引言

智能手机如今已逐渐成为人们日常生活不可或缺的一部分，其内部存储了大量的个人隐私和敏感数据。目前，6 位个人认证码（Personal Identification Number，PIN）[1]、九宫格口令[2]、指纹识别[3]、面部识别[4]是保护手机数据安全的第一道防线。然而，大量研究表明，这些方法面临着各种潜在威胁，包括侧信道攻击[5]、窥视攻击[6]、弱口令[1,7]、呈现攻击[8]、口令猜测攻击[9-12]等，引发了人们对传统口令保护下智能手机设备安全性的担忧。为了提高智能手机的安全性，近年来，众多替代或补充的认证方案被提出，如基于声感知的移动终端身份认证方案[13]、基于触摸和运动传感器信息的隐式认证方案[14-15]、基于动态手势的认证方案[16]、基于自由手势的认证方案[17]。

在智能设备环境中，自由手势认证（Free-form Gesture Based Authentication）被视为一种有前景、可替代传统口令（如 PIN 和九宫格）的解决方案[17-20]。自由手势认证系统允许用户在空白触摸屏显示器上用一个或多个手指绘制任意形状或图案[18]。研究表明，自由手势在安全性、可用性和易记性方面优于 PIN 和九宫格口令[17-20]。然而，Cheon 等[17]揭示了用户倾向于创建简单、可预测的手势。这些手势聚集在一个“弱子空间”内，在 20 次在线字典猜测攻击中，攻击者通过猜测常用的自由手势发起攻击，攻击成功率高达 23.13%。随后，Cheon 等[21]设计了一个自由手势的强度评估工具 GestureMeter，其可在牺牲部分用户体验的条件下将在线字典猜测攻击的成功率降至 10.33%。尽管自由手势认证范式相较于 PIN 和九宫格口令具有更大的潜力，但仍须进一步研究以提升其安全性。

如图 1 所示，本文首次探索利用自由手势的触摸行为生物特征进一步认证用户身份，从而削弱在线字典猜测攻击带来的负面影响。现有的研究表明，用户的行为模式中包含独特的行为生物信息[22]。智能手机中配备了大量传感器，如压力传感器，电容传感器等。当用户与智能手机交互时，这些传感器捕获的信号包含用户独特的生理和行为信息[22]。在认证期间，传感器可在不影响用户体验的情况下在后端收集认证所需的信息。由于生理和行为信息的隐蔽性和独特性，即使攻击者可能猜测到了手势的类别，也因生理特征与行为模式的差异而难以成功攻破认证系统，从而有效降低了在线字典猜测攻击的成功率，提升了系统的安全水平。

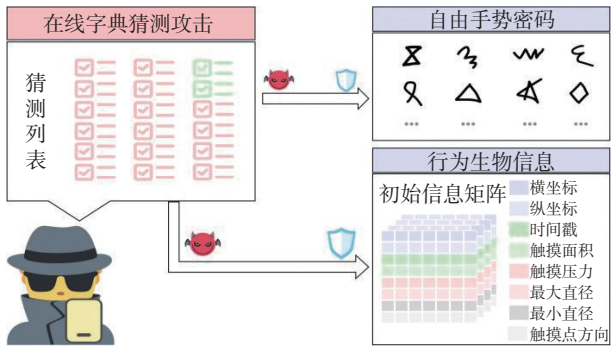

图 1 在线字典猜测攻击

Fig. 1 Online dictionary guessing attack

基于触摸行为生物特征构建自由手势认证增强方案面临以下挑战。

（1）有限样本。注册阶段仅能获得少量目标用户手势，且攻击者数据分布无法提前获得。因此，我们无法获得足够的有效标注数据为每个目标用户训练专用的认证模型。

（2）未知手势。由于用户注册手势的多样性以及同一类手势绘制方式的差异性，认证系统难以提前获取全部用户可能注册的手势的类别和模式。因此，认证系统应具有较强的泛化能力，从而保障当用户注册系统面对未见过的手势时，认证系统仍能够识别并有效认证用户身份的合法性。

面对上述挑战，本文采用少样本学习技术，旨在利用有限的样本训练模型，使模型从不同类别的样本中学习具有区分力的通用特征[23]。少样本学习技术能够在每个类别仅有有限样本的条件下训练模型，使模型不断学习新任务，并提高模型的泛化能力，从而识别未知手势。孪生网络[23]（Siamese network，SN）已被用于处理每类仅有少量样本的少样本分类或表示学习问题，其包含两个拥有完全相同架构和参数的子网络。本文以卷积神经网络（Convolutional Neural Network，CNN）作为子网络，基于孪生网络构造特征提取器，其能够将原始的行为生物信息映射为特征向量。

结合动态时间规整（Dynamic Time Warping，DTW），本文提出一个新的基于行为生物特征的自由手势认证增强方案 StrokeFG。首先，StrokeFG 通过 DTW 将待测手势的嵌入序列与注册模板进行弹性时间对齐，以评估手势轨迹的相似性。随后，利用上述特征提取器从用户行为生物信息中提取特征向

量，并输入由多层感知机构建的认证器，完成身份验证。在识别手势轨迹的基础上，StrokeFG 可以在不影响用户体验的情况下通过用户认证期间的行为生物特征进一步验证用户的身份。

本文的主要贡献如下：

（1）一个新的基于行为生物特征的自由手势认证增强方案。基于 DTW，本文首次利用触摸行为生物特征构造了一个自由手势认证增强方案 StrokeFG。StrokeFG 不要求额外的硬件支持，仅需手机上已有的传感器收集的信息。

（2）安全性和健壮性评估。本文收集了一个包含用户行为生物特征的自由手势数据集，其包含 97 个用户在两种行为状态下绘制的 30 种手势的行为生物数据。基于自由手势数据集，本文在实验阶段系统性地评估了所提的认证方案的健壮性和安全性。结果表明，在相同手势构成的测试样本集中，相较于基线，StrokeFG 的假阳率降低了 86.56%~91.24%。在敌手已知手势类别的前提下，StrokeFG 在在线字典猜测攻击下实现了猜测复杂度为 6.50~8.41 比特（$\alpha \in [0.1, 0.4]$）的安全水平。

（3）数据收集应用程序。为了提升本文工作的可复现性，公开了自研的基于安卓系统的手势数据收集应用程序，其能够捕获用户在智能手机指定区域绘制手势时的行为生物特征，包括用户的触摸坐标、时间戳、面积、压力、最大直径、最小直径、触摸点方向。相关细节见第 4 节。

1　相关工作

自由手势是一种二维空间的手势，自 2014 年由 Sherman 等[20]研究后，引起了一些学者的关注。在本节中，除了介绍与自由手势相关的前沿研究工作外，还阐述了与本文研究相关的行为特征认证工作。

1.1　自由手势认证

自由手势允许用户在触摸屏幕指定空白区域的任意位置绘制任意的图案[20]，并且没有网格或模板提示。研究表明，相较于 PIN 和九宫格口令，自由手势具有更大的口令空间和更强的安全性[17]。2014 年，Sherman 等[20]收集了 63 名参与者的自由手势，并研究了手势的安全性和记忆性。结果表明，用户对于签名和简单形状的记忆性更好。

2016 年，为了研究在自然环境下自由手势认证系统与其他认证系统的区别，Yang 等[18]利用经验取样法研究了 91 名参与者在多账户条件下创建口令和认证的行为。研究表明，参与者在真实环境中生成的口令比在实验室研究中生成的口令更弱，用户更倾向于创建更简单的手势。91 个参与者创建的手势口令中有 49.28% 与形状有关，24.07% 与字母有关，15.76% 与线条有关。2017 年，Liu 等[19]证明了缩放和旋转会对认证系统产生负面影响，降低认证准确率，而对齐预处理不会影响系统的性能。

2020 年，Cheon 等[17]收集了 2 594 名参与者创建的手势数据，并采用了两个广泛应用的识别器：DTW[24]和量角器（Protractor）[25]。为了评估手势认证方案的安全性，Cheon 等利用 n 元马尔可夫模型[7]评估了破解比例为 α 的部分猜测熵。研究结果显示，随着 α 值从 0.1 增长至 1.0，自由手势的部分猜测熵相较于 4 位数字口令和九宫格口令提升了至少 25%。然而，自由手势识别器在抵抗基于聚类的字典攻击方面表现出了脆弱性。攻击者利用字典可以在 20 次尝试内攻破 23.13% 的账户；利用黑名单策略，敌手攻击的成功率下降至 14.93%。

2023 年，为了帮助用户创建更安全的自由手势，Cheon 等[21]提出了一种新的自由手势强度评估工具 GestureMeter。实验表明，在假阴率为 11.54% 的特定决策阈值下，自由手势猜测成功率可降至 10.33%。2025 年，Cheon 等[26]利用单类支持向量机分类器探索了从加速度传感器、重力传感器以及旋转速度获取的行为特征的有效性。结果表明，在用户输入过程中捕获的运动传感器特征可将肩窥攻击者的成功率降低至 2.11%~4.90%。

1.2　行为生物认证

由于用户独特的生理结构和行为习惯，所以在与智能手机交互时会产生特有的身份信息，并可被手机内置的各类传感器捕获。在现有研究中，行为生物信息主要分为三类：一是触摸信息，通过屏幕传感器（如压力、电容传感器等）采集；二是运动信息，借助运动传感器（如陀螺仪、加速度传感器等）获取；三是结合触摸与运动的多模态信息。

触摸认证系统在可用性与安全性方面已得到广泛研究。2013 年，Frank 等[27]和 Li 等[22]对其可用性进行了探讨，表明触摸坐标、时间戳、面积和压力等数据可用于身份验证。后续研究进一步分析了特征选择[27]、实验环境[28]、窗口大小[29]、触摸方向[30]对触摸认证系统的影响，并评估了聚类攻击[31]、模仿攻击[32]、随机输入攻击[33]对系统的威胁。

基于运动信息的认证系统常称为步态认证[34]。研究表明，运动传感器所捕获的用户行走特征可有

效用于身份认证。然而，多数研究数据来自特定实验环境，虽有个别工作采集了自然状态下的数据[35]，但是由于用户所处环境的复杂性，注册阶段常需大量数据[23]，限制了实用价值。此外，用户行为姿势多样（如坐、站、躺、跑），进一步加大了构建稳健步态认证系统的难度[23]。

多模态认证系统结合了触摸信息和步态信息，能够捕获更加丰富的用户行为信息[36-38]。研究表明，相较于触摸认证系统和步态认证系统，多模态认证系统具有更高的认证准确性。

2 系统模型

本节主要介绍了所提的 StrokeFG 的系统架构和模型细节。

2.1 系统架构

如图 2 所示，StrokeFG 系统架构分为三个阶段：训练阶段、注册阶段和登录阶段。在训练阶段，服务提供商负责离线训练和测试模型的性能。在注册阶段，系统同步采集用户注册的自由手势的轨迹以及行为生物信息。行为生物信息经归一化处理后由特征提取器提取特征，提取的特征记录在模板池中。在登录阶段，系统首先验证手势轨迹的合法性，若匹配成功，进一步验证输入特征是否与注册特征一致，据此判断用户身份的合法性。

（1）训练阶段。StrokeFG 的认证部分由轨迹认证模块和行为生物认证模块组成。作为一种入口点认证机制，StrokeFG 的部署方式与基于 PIN、人脸或指纹的认证机制相似，其预先集成于手机系统中，而无须在用户端重训练模型，具有良好的可拓展性。本文聚焦于验证 StrokeFG 的可行性，不涉及具体部署细节。

特征提取器是采用以卷积神经网络为子网络构建的 SN，适用于小样本分类或表示学习任务。SN 结构对称，两个子网络权值共享，旨在衡量两个输入样本之间的相似性，学习更具泛化能力的特征。属于同一类别的样本会提取相似特征，不同类别则提取差异特征，特征相似性以欧氏距离度量。本文使用三元损失函数优化子网络权重。特征提取器训练完成后，服务提供商固定其权重，并利用二元交叉熵损失函数优化认证器（Decision Network，DN）。SN 与 DN 的详细结构见 3.2 节。

（2）注册阶段。与 PIN、人脸或指纹认证类似，用户选择一个自由手势并在不同的行为状态下重复输入若干次以完成注册。用户在输入自由手势时，智能手机内置传感器采集相应的行为生物特征，包括触摸点位置、面积、压力、时间、手指朝向等。触摸点位置序列构成手势轨迹，经尺寸和位置归一化预处理以消除尺度与平移变化。特征提取器同步

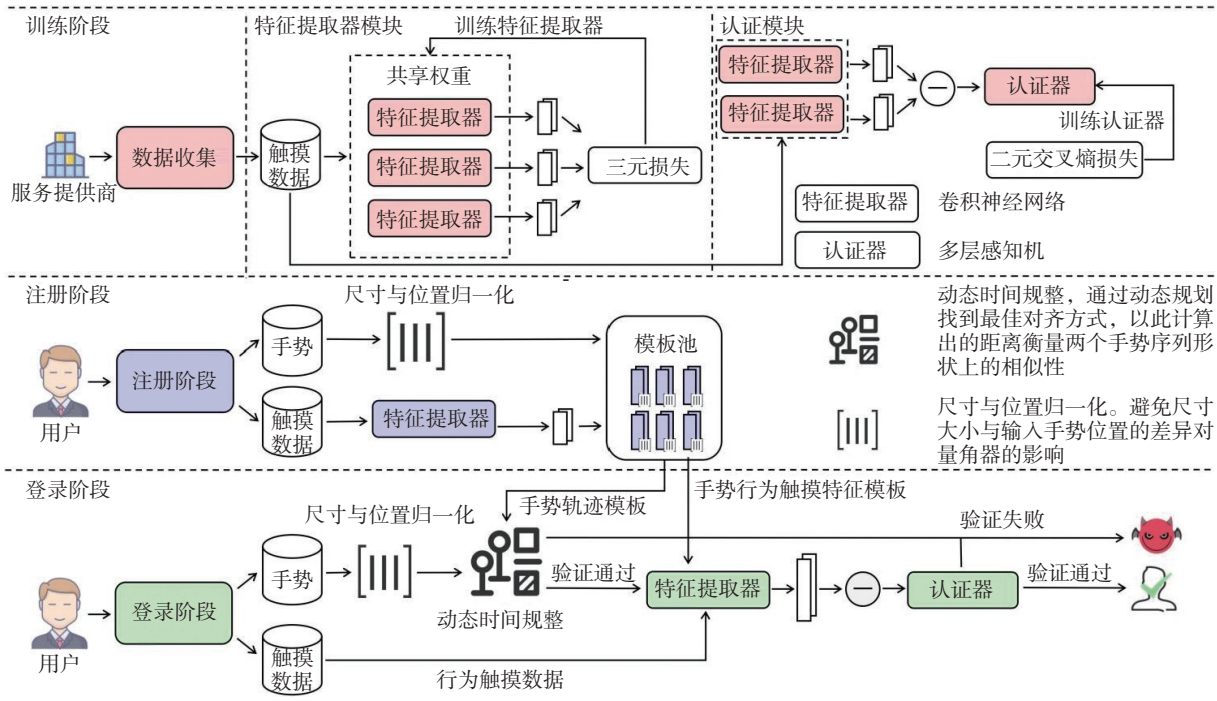

图 2 系统架构

Fig. 2 System structure

将行为生物信息映射为隐式特征向量，捕获原始数据中用户特有的特征。手势轨迹与特征向量均存储于模板池，用于登录验证。

（3）登录阶段。StrokeFG 采用两阶段认证结构。用户输入手势后，系统首先使用 DTW 匹配轨迹与模板池中是否一致。若失败，拒绝访问；若成功，则提取输入手势的行为生物特征向量。认证器以输入特征与模板特征的差值向量作为输入，判断是否来自同一用户，从而决定是否通过认证。

2.2　模型细节

图 3（a）展示了 SN 子网络的设计细节，该子网络以归一化的行为生物数据作为输入，输出一个 256 维的特征向量。由于本文所收集的手势样本数据长度至多为 221，且原始特征共有 8 种，因此子网络的输入维度为 (221, 8)。

输入层	输入维度	(221, 8)
卷积层	输入维度	(221, 8)
	输出维度	(221, 256)
池化层	输入维度	(221, 256)
	输出维度	(110, 256)
卷积层	输入维度	(110, 256)
	输出维度	(110, 512)
池化层	输入维度	(110, 512)
	输出维度	(55, 512)
卷积层	输入维度	(55, 512)
	输出维度	(55, 1 024)
全局平均池化层	输入维度	(55, 1 024)
	输出维度	(1 024)
全连接层	输入维度	(1 024)
	输出维度	(512)
输出层	输入维度	(512)
	输出维度	(256)

(a) 特征提取器
(a) Feature extractor

输入层	输入维度	(256)
全连接层	输入维度	(256)
	输出维度	(512)
全连接层	输入维度	(512)
	输出维度	(512)
全连接层	输入维度	(512)
	输出维度	(256)
Dropout层	输入维度	(256)
	输出维度	(256)
输出层	输入维度	(256)
	输出维度	(1)

(b) 决策网络
(b) Decision network

图 3　行为特征认证模块

Fig. 3　Behavioral feature authentication module

网络包含三层卷积层，卷积核数量分别为 256、512 和 1 024，大小均为 3，采用“same”填充以保持序列长度不变，并使用 ReLU 激活函数。第一层卷积用于初步提取局部特征，后两层进一步提升特征抽象与表示能力。池化层在减少计算量的同时保留重要特征。全局平均池化层将每个通道的特征图压缩为一个标量，实现特征的全局聚合。全连接层与输出层进一步去除冗余，保留鉴别信息。

本文采用多层感知机作为 DN，基于隐特征空间中向量的差值进行分类决策，完成二分类任务，输入为两个隐特征向量的差值，输出为它们属于同一用户的概率。系统可通过设定阈值，根据该概率判定用户身份的合法性。

如图 3（b）所示，DN 结构包括输入层、三个全连接层、Dropout 层和输出层。输入维度为 256。全连接层神经元数分别为 512、512 和 256，激活函数为 ReLU。Dropout 层丢弃率为 0.2，以有效防止过拟合，提升泛化能力。输出层含一个神经元，使用 Sigmoid 函数将输出映射至 [0,1]，表示属于同一用户的概率。

3　威胁模型

本节描述了攻击者的能力以及对自由手势认证系统的威胁模型所做的假设。

（1）现有的手机认证系统将数据存储与处理置于一个安全区域中。例如，苹果的指纹识别数据存储在 Secure Enclave 中，这是一个独立的安全芯片，与主操作系统隔离。安卓设备也会利用安全区域（如 TrustZone）保护敏感数据。与现有的研究保持一致[23]，本文假设触摸传感器数据的存储和处理是在安全区域内进行的。手势认证模块可以在 ARM TrustZone 等安全区域中运行，以限制安全区域之外的任何对传感器和手势认证模块的访问。安全区域的存在限制了攻击者的攻击范围，可以有效抵抗主密钥攻击[39]。然而，我们无法限制攻击者通过传感器接口执行攻击。因此，本文在安全性分析阶段评估了“狼”攻击[39]的有效性。

（2）基于现有的工作，本文将攻击分为三类：零知识攻击、猜测攻击、模仿攻击。零知识攻击是指攻击者没有目标用户自由手势的任何信息，攻击者仅能通过不断尝试各种类别的自由手势与行为生物特征进行攻击。猜测攻击包括离线猜测攻击和在线猜测攻击。由于用户注册的手势样本仅存储在本地，本文假设认证系统受到安全区域的保护，从而避免离线猜测攻击。攻击者可基于收集到的自由手势猜测字典发动在线字典猜测攻击。模仿攻击是指攻击者不仅拥有目标的图案信息，还能够通过反复观察录像等方式尝试模仿用户的行为模式特征。零知识攻击是最弱的，模仿攻击是最强的。本文仅针对在线字典猜测攻击设计自由手势认证增强方案，模仿攻击不在研究范围内，相关工作可参考文献 [19] 和 [20]。

（3）与现有的工作保持一致[23]，通信信道通过使用标准的安全通信协议（如 Secure Sockets Layer/Transport Layer Security，SSL/TLS）得到保护，因此被认为不受攻击者影响。此外，攻击者也无法篡改

手势认证模块已从传感器读取的数据，他们也不能在设备上实施中间人攻击或重放攻击。发动此类攻击需要嗅探安全通信信道，这意味着需要破解 SSL/TLS 等信道加密协议。

4 数据收集实验

为了支持本文的研究，我们设计了数据收集实验，以收集用户绘制自由手势时的行为生物特征。与现有的公开数据集不同，本文不仅收集了用户与智能手机交互过程中的触摸点坐标与时间戳，还额外记录了多种行为特征数据，包括触摸面积、触摸压力、最大直径、最小直径以及触摸点方向。本节的其他部分介绍了数据收集实验关键部分的设计。

4.1 设备

数据采集实验使用 5 台屏幕大小为 6.47 英寸、分辨率为 2 340×1 080、搭载了 Android 系统的智能手机。为了避免手机硬件及软件的差异影响数据采集的质量，所有设备均为华为 P30 pro。

4.2 应用程序

基于 Android 系统，开发了一款自由手势行为生物特征收集应用程序。如图 4（a）所示，应用程序界面分为两部分：上半部分显示提示图案及用户需要提前阅读的注意事项，下半部分灰色区域为用户绘制手势的区域。用户需要在该灰色区域内依照提示，按个人偏好输入手势。该应用程序能够以最高 120 Hz 的频率主动采集行为生物特征信息。

(a) 应用 (a) Application　(b) 手势 (b) Gestures　(c) 实验 (c) Experiment

图 4 数据采集应用及手势类别

Fig. 4 Data-collection application and gesture categories

4.3 任务细节

实验要求用户输入 30 类手势，具体类别如图 4（b）所示。这些手势在 Cheon 等[17] 从 2 594 名用户中收集的手势中最常见。每类手势需要用户连续重复输入 7 次。实验开始前，为了使用户在实验过程中尽可能保持与自然环境下使用手机相似的行为模式，说明如下：

（1）可在灰色区域内任意位置绘制手势，无须局限于中心区域。为了避免用户因实验环境而过度依赖提示图案的位置，鼓励用户以舒适姿势按自身偏好绘制，不预先固定起始点。

（2）提示图案仅供参考，不必严格遵循其形状规则。部分用户可能倾向于按提示样式等比例放大绘制，但我们鼓励用户依个人偏好自由绘制，不限制角度与大小。

（3）选择最常用的手势姿势。常用姿势有助于用户以日常使用手机的自然模式完成实验，从而提高数据的真实性和代表性。

为了研究用户行为状态对自由手势认证系统的影响，要求用户在行走和坐两种状态下完成上述实验。每位参与者在实验结束后可获得 50 元人民币作为报酬。

4.4 伦理考虑

实验开始前，向用户详细说明实验内容与目的，明确告知数据收集事宜。所有实验均在用户知情同意的前提下进行。尽管数据已进行匿名化处理，但因涉及用户生理与行为信息，仍属于敏感数据。公开数据集可能存在隐私泄露风险，并为攻击者提供额外信息。因此，本文不公开数据集具体细节，仅报告相关统计信息，以避免个人隐私泄露。为了保障实验结果的可复现性，在 github（https://github.com/Litraveler/StrokeFG.git）上公开了数据收集应用程序及实验代码。

5 数据集

共有 97 名用户参与实验并提供了数据，其中 90 名用户在行走和坐两种状态下完成了实验，另有 7 名用户因时间限制仅完成了坐状态下的实验。实验共收集到 39 060 条手势数据，为了保障实验的完整性，最终仅采用 90 名用户的数据，共计 37 800 条。为了减少因用户不熟悉手势而导致的输入偏差，对每名用户、每种行为状态下的每类手势数据，均过滤了前 2 条记录。在这 90 名用户中，60 名用户在行走和坐状态下以相同方式持握手机完成实验，另有

30 名用户则改变了持机方式。

6　实验评估

本节通过实验系统评估了 StrokeFG 中行为生物认证模块的性能与安全性。具体地说，首先，评估了不同样本量对行为生物认证模块的影响；其次，通过对比 DTW[24] 与 Protractor[25] 方法，验证了所提出行为生物特征认证模块在识别相同类别手势方面的优势；再次，引入 Van Hamm 等[39] 提出的行为生物认证系统安全性评估框架，对 StrokeFG 的行为生物认证模块抵抗在线字典猜测攻击的安全性进行了量化；最后，评估了未知手势、不同行为状态以及跨行为状态对行为生物认证模块的影响。

6.1　训练策略

如图 2 所示，StrokeFG 的行为生物认证部分包含特征提取模块和认证器模块。本文使用三元组损失函数训练特征提取模块，使用二元交叉熵损失函数训练认证模块。三元组损失函数的目标是使同类样本在嵌入空间中彼此靠近，不同类样本彼此远离，从而有效实现用户身份验证。

使用三元组损失函数训练 SN 需要选取三类样本，分别是锚点样本、正样本、负样本。训练 SN 旨在减少锚点样本和正样本之间的距离，同时增大锚点样本和负样本之间的距离。如式（1）所示，a表示锚点样本，p表示正样本，n表示负样本，α表示边界值，$D(a,p,n)$表示三元组损失函数。α确保锚点样本与正样本之间的距离与锚点样本与负样本之间的距离之差至少为α，防止模型退化为平凡解。

$$D(a,p,n)=\max(\|f(a)-f(p)\|^2-\|f(a)-f(n)\|^2+\alpha,0) \quad (1)$$

本文使用二元交叉熵损失函数训练 DN。二元交叉熵损失函数是一种常见的损失函数，其主要用于二分类问题，能够很好地衡量模型预测的概率分布与真实分布之间的差异，并处理输出为概率值的情况。二元交叉熵损失函数的定义如式（2）所示。N是批量大小，y_i是真实标签，模型的预测概率为$\widehat{y_i}$。$\widehat{y_i}$由模型的输出层产生并经过 Sigmoid 激活函数处理，以确保其范围在 0 ~ 1 之间。DN 输出的值代表两个输入样本属于同一个用户的概率，1 代表输入的两个样本属于同一用户，0 则反之。

$$L=-\frac{1}{N}\sum_{i=1}^{N}[y_i\log(\widehat{y_i})+(1-y_i)\log(1-\widehat{y_i})] \quad (2)$$

6.2　数据处理

训练集与测试集的用户比例设定为 2∶1，并确保训练用户样本不出现在测试集中。在超参数优化阶段，从训练用户数据中进一步划分出验证集，其余作为训练集，训练集与验证集的比例也为 2:1。由于 SN 使用三元组损失函数，而 DN 使用二元交叉熵损失函数，因此，两者的训练机制不同。为了满足不同实验的需求，本文采用了 5 种不同的样本生成策略。

（1）三元组类型一（T1）。分别处理行走和坐两种状态的数据。对每一用户每一状态下每一手势的 5 个样本，两两组合形成 10 个锚点样本-正样本对，并从其他用户同状态同手势的样本中随机选取一个负样本，构成三元组。遍历所有状态、手势和用户后，得到完整的三元组数据集。

（2）三元组类型二（T2）。合并行走和坐两种状态的数据。每一用户每一手势有 10 个样本，组合生成 45 个锚点样本-正样本对。负样本生成方式与 T1 相同。

（3）对比样本类型一（C1）。分别处理行走和坐两种状态。每一用户每一状态下每一手势的 5 个样本组合出 10 个正样本对，并从其他用户同状态同手势样本中为每对随机选取 2 个负样本。遍历所有状态、手势和用户，得到对比样本集。

（4）对比样本类型二（C2）。合并行走和坐两种状态的数据。每一用户每一手势的 10 个样本生成 45 个正样本对，并从其他用户同状态同手势样本中为每对随机选取 5 个负样本。

（5）对比样本类型三（C3）。针对每一用户的每一手势，分别将坐状态样本作为锚点样本，行走状态样本作为正样本，生成所有可能的正样本对。集合仅含正样本对，用于测试状态变化对认证性能的影响。

6.3　超参数优化

本节旨在优化 SN 与 DN 的超参数，其预设值分别如表 1 和表 2 所示，重点考察了学习率、批次大小、边界值及输出维度的影响。本文采用网格搜索方法遍历预定义的超参数空间，最终确定最优超参数组合，如表 3 所示。

SN 与 DN 的超参数优化是独立进行的，具体过程如下：将训练数据划分为训练集和验证集，基于三元组损失函数在各超参数组合下训练并验证 SN 的性能。确定 SN 的最优超参数组合后，固定该类超参数并训练优化 SN 的模型参数。SN 参数优化完成后，固定其模型参数，从而评估 DN 在各超参数组合下的性能，最终确定 DN 的最优超参数组合。

表 1 特征提取器超参数

Table 1 Hyperparameters of the feature extractor

超参数	预设值
学习率	0.1、0.01、0.001、0.000 1
批次大小	32、4、128
边界值	1.0、0.5、0.1、0.05、0.01
输出维度	64、128、256

表 2 决策网络超参数

Table 2 Hyperparameters of the decision network

超参数	预设值
学习率	0.1、0.01、0.001、0.000 1
批次大小	32、64、128

表 3 最优超参数组合

Table 3 Optimal hyperparameter combinations

网络	学习率	批次大小	边界值	输出维度
孪生网络	0.000 1	128	0.01	256
决策网络	0.001	32	—	—

6.4 评估标准

本文使用现有研究中常见的评估指标，每个指标都基于正确分类的良性认证事件的数量（真正例）和错误分类的良性认证事件的数量（假负例），以及正确检测到的攻击尝试的数量（真负例）和未识别出的攻击尝试的数量（假正例）。具体使用了准确率（Accuracy）和等错误率（Equal Error Rate，EER）两种指标。假阳率（False Positive Rate，FPR）和假阴率（False Negative Rate，FNR）两种指标用于辅助计算。EER 表示在特定阈值下 FPR 和 FNR 相等的点，值越低意味着模型性能越好。本文考虑了两种错误的指标，可用于比较不同模型的性能。

准确率的定义如下：

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \tag{3}$$

FPR 的定义如下：

$$\text{FPR}=\frac{\text{FP}}{\text{FP}+\text{TN}} \tag{4}$$

FNR 的定义如下：

$$\text{FNR}=\frac{\text{FN}}{\text{FN}+\text{TP}} \tag{5}$$

6.5 样本量的影响

本节评估了样本量与行为状态对 StrokeFG 的影响。行走和坐两组数据分别独立实验，采用 T1 生成训练 SN 所需三元组，C1 生成训练 DN 所需对比样本，并用 C1 构造测试集，训练轮次为 20。

图 5 和图 6 箱线图显示，样本量由 6 400 增至 12 800 时，30 类手势准确率上升且分布收紧（第三四分位数与第一四分位数的差与须长均缩小）。继续增至 19 200~32 000，准确率反而略降，这可能是因为在训练数据量不断提升但数据多样性有限的情况下，模型过度拟合训练数据。

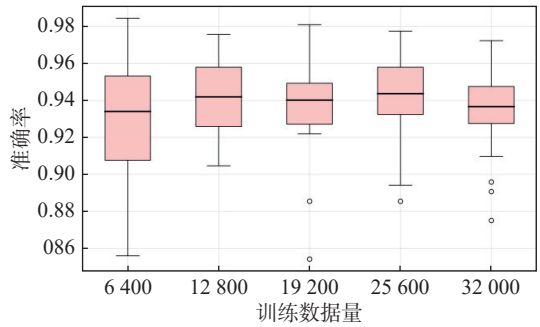

图 5 训练数据量对准确率的影响（坐）

Fig. 5 Impact of training set size on accuracy（Seated）

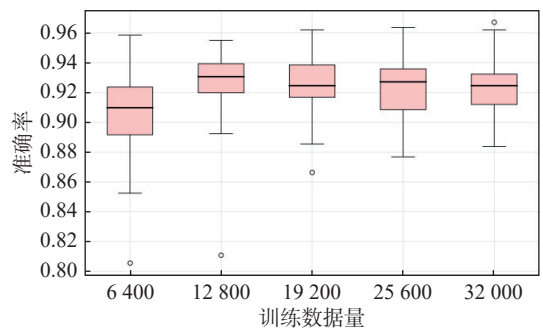

图 6 训练数据量对准确率的影响（行走）

Fig. 6 Impact of training set size on accuracy（Walking）

在相同样本量条件下，坐状态下 StrokeFG 报告的准确率高于行走状态，可能的原因是用户在行走状态下的重心移动导致用户绘制的手势更加多样化且不稳定。当数据量为 12 800 时，认证准确率最高可达 97.56%。在图 7 和图 8 中，坐状态下模型的 EER 中位数最低为 4.03%，行走状态下模型的 EER 均值最低为 6.42%。实验结果表明，StrokeFG 能够在用户无感的状况下进一步验证其身份，具有良好的可部署性。

6.6 对比分析

DTW[24] 和 Protractor[25] 通过判断手势轨迹是否相

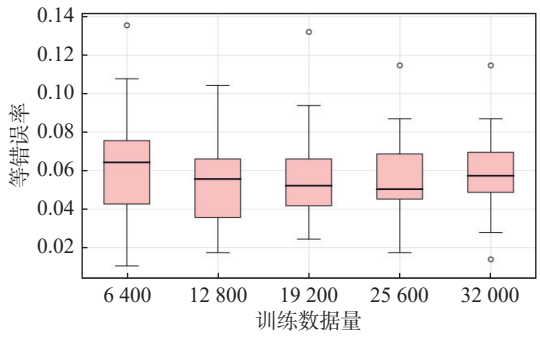

图 7　训练数据量对等错误率的影响（坐）

Fig. 7　Impact of training set size on EER （Seated）

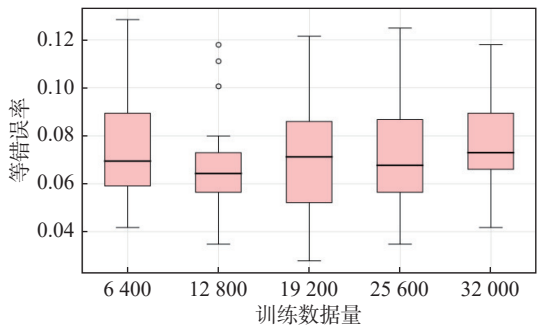

图 8　训练数据量对等错误率的影响（行走）

Fig. 8　Impact of training set size on EER （Walking）

似来识别用户，该方法在自由手势认证领域已被广泛采用[17,19,26]。与上述方法不同，本文方案在识别手势类别的基础上，进一步引入用户特有的触摸行为生物特征以识别潜在冒充者，从而提升系统的安全性。因此，在同一手势测试集中，将本文方案与基线方法 Protractor 和 DTW 进行了性能比较。测试集已进行位置与尺寸归一化，以排除其对基线方法性能的干扰，详见文献 [17]。

图 9 和图 10 展示了各分类器在 C1 测试集上的表现。在文献 [17] 中，Cheon 等在包含 2 594 名用户生成的手势的测试集上实现了 3.59%（DTW）和 4.14%（Protractor）的 EER。本文评估了当 FNR 分别为 3.59% 和 4.14% 时，DTW 和 Protractor 在相同类别手势测试集中的 FPR。如图 9 所示，在坐姿测试集中，StrokeFG 的 EER 为 5.64%。DTW 的 FPR 为 50.68%，Protractor 的 FPR 为 56.58%。与 DTW 和 Protractor 相比，StrokeFG 报告的 FPR 分别降低了 89.86% 和 91.24%。如图 10 所示，在行走测试集中，StrokeFG 的 EER 为 6.81%，DTW 的 FPR 为 50.68%，Protractor 的 FPR 为 56.58%。与 DTW 和 Protractor 相比，FPR 分别降低了 86.56% 和 87.96%。这表明 StrokeFG 中的行为生物认证模块可以弥补 DTW 和 Protractor 在识别相同类别手势方面的不足，进一步提升了系统整体的安全性。

与侧重于区分不同类别手势的 DTW 和 Protractor 不同，StrokeFG 能够进一步捕捉用户特有的行为特征差异。因此，在识别不同用户的相同类别手势时，StrokeFG 错误率降低 86.56%~91.24%。

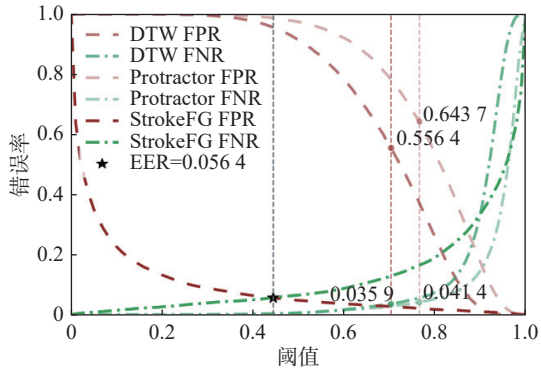

图 9　各分类器性能比较（坐）

Fig. 9　Classifier performance comparison （Seated）

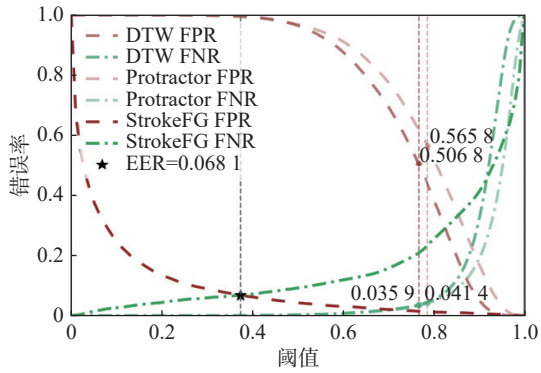

图 10　各分类器性能比较（行走）

Fig. 10　Classifier performance comparison （Walking）

现有研究表明，在线字典攻击可在 20 次尝试内攻破超过 10.33% 的自由手势[21]。在假设攻击者为零知识的前提下，StrokeFG 可在攻击者猜测到手势类别的情况下有效降低其成功率。然而，在触摸认证领域[31,40]中，攻击者可利用已有的行为生物特征信息发起在线字典猜测攻击（类似于“狼”攻击[39]）。由于 StrokeFG 同样依赖用户行为生物特征作为增强认证机制，当攻击者猜测或提前获知用户手势类别时，可凭借自身知识库发起在线字典猜测攻击。因此，在 7.7 节中使用熵评估了 StrokeFG 对抗在线字典猜测攻击的能力，以量化其安全性。

6.7　安全性分析

熵是衡量认证系统安全性的重要指标之一。不同于字符串口令[41]、PIN[1] 等确定性口令认证方案，

生物特征识别具有模糊匹配[39]的特性。因此，本节采用 Van Hamme 等[39]于 2024 年在信息取证与安全学报（Transactions on Information Forensics and Security，TIFS）上提出的评估框架，以计算 StrokeFG 的有效密钥空间。

为了评估 StrokeFG 为行为生物认证模块的安全性，本节参考了文献 [2,17,42] 所采用的评估指标及相关测量结果。为了便于比较分析，选用 α-猜测复杂度作为衡量指标，该指标的详细介绍参见文献 [39]。在评估 StrokeFG 时，假设攻击者已知一半手势类别中一半用户的行为特征样本。该假设合理，因为攻击者穷举所有手势类别并获取所有目标用户的历史手势样本较为困难。攻击者可基于已掌握的行为生物信息库，选取最有可能的攻击序列。在评估中，攻击者拥有的样本量为 675，攻击目标的样本量为 1 350。

表 4 展示了 StrokeFG 的行为生物认证模块的 α-猜测复杂度评估结果。受样本数量限制，未评估 α = 0.7 和 α = 1.0 时 StrokeFG 的安全性。如表 4 所示，在 α-猜测复杂度指标下，StrokeFG 的行为生物认证模块在两种行为状态下分别提供 6.60~8.41 比特与 6.50~8.21 比特的安全性。这意味着即使攻击者已知手势类别，StrokeFG 仍可提供 6.50~8.41 比特的安全强度。尽管受限于文献 [17] 中数据集未公开，本文无法进一步评估在攻击者未知手势类别及行为生物特征时 StrokeFG 对抗在线猜测攻击的能力，但对行为生物认证模块的安全性评估仍为 StrokeFG 的整体安全性提供了重要参考。

表 4 α-猜测复杂度评估

Table 4 Entropy evaluation of α-guesswork

方案	α			
	0.1	0.2	0.3	0.4
Baseline 3×10[17]	7.55	11.06	14.47	16.23
Blacklist 3×10[17]	8.29	12.99	15.62	16.99
4-digit PINs[42]	5.19	7.04	8.37	9.38
Patterns[2]	5.04	5.82	6.54	7.19
行为生物认证模块（坐）	6.60	6.98	7.43	8.14
行为生物认证模块（行走）	6.50	7.01	7.68	8.21

6.8 未知手势的影响

在实际应用场景中，由于无法预先获取用户的选择，用户可能选择训练集中未出现过的手势。为此，本节研究了 StrokeFG 在未知手势样本上的识别性能。使用 T1 和 C1 生成训练与测试样本，训练集规模为 12 800，测试手势类别数为 12 ~ 26。图 11~图 14 展示了坐与行走状态下模型在测试集上的表现，每个箱线图汇总了 10 次重复实验的结果。图 11 和图 12 展示了识别准确率随手势类别数的变化情况，图 13 和图 14 为 EER 的相应变化。

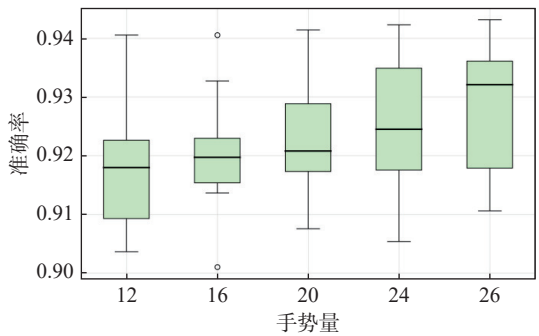

图 11 手势量对准确率的影响（坐）

Fig. 11 Impact of gesture quantity on accuracy（Seated）

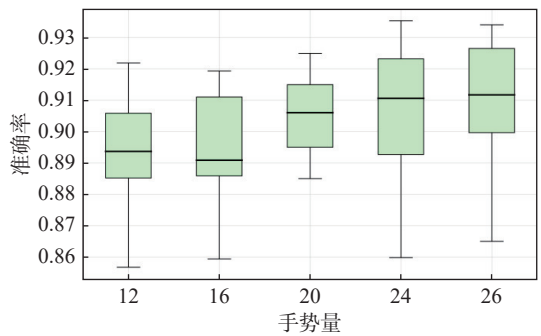

图 12 手势量对准确率的影响（行走）

Fig. 12 Impact of gesture quantity on accuracy（Walking）

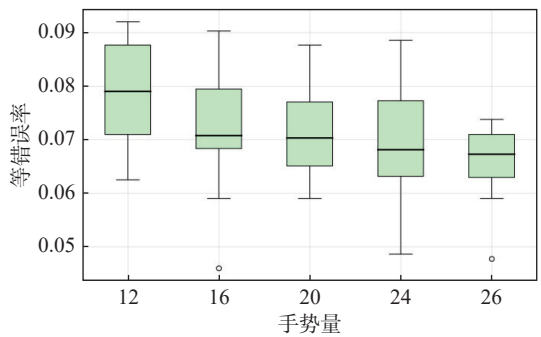

图 13 手势量对等错误率的影响（坐）

Fig. 13 Impact of gesture quantity on EER（Seated）

实验结果表明，随着手势类别数的增加，模型识别准确率逐渐上升，EER 逐渐下降。这一结果符合预期，因为在总训练数据量固定的情况下，更多类别的自由手势提升了数据的多样性，使模型能够

学习更具泛化能力的隐式特征，从而增强了 StrokeFG 从未知手势中提取有价值的通用特征的能力。

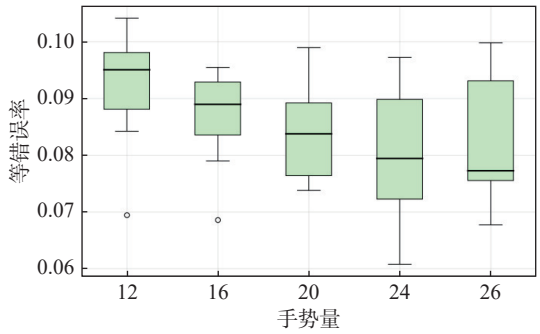

图 14　手势量对等错误率的影响（行走）

Fig. 14　Impact of gesture quantity on EER（Walking）

从图 11~图 14 可见，模型识别准确率的中位数随着手势量的增加而提高，EER 的中位数随着手势量的增加而下降。这是符合预期的，因为随着手势类别的增加，模型更能够从多样化的数据中学习到通用特征。然而，当手势量从 12 增加至 26 时，模型识别准确率的最小值和最大值呈现一定的起伏。这表明部分模型在尾部用户数据上的表现存在明显的波动性。可能原因是当总训练样本量固定为 12 800 时，手势类别数增加意味着每个手势类的平均样本数减少。当手势类别数从 12 增至 26，每类手势样本数降低，类别增加但总样本量不变，优化过程更容易受到噪声和离群样本干扰，尤其是当少数用户手势质量较低或分布偏斜时，模型难以同时保持对所有用户的高精度判别，从而在局部最优解附近震荡。然而，总体而言，模型的性能随手势多样性的增加而提升，表明多样化的手势有利于提升模型的性能。

与坐状态相比，模型在行走状态下的准确率较低，EER 较高。这是因为行走时用户绘制手势的不确定性和变异性更大，而坐时身体更稳定，手势的类内差异较小，从而获得更好的识别效果。

6.9　行为状态的影响

本节探讨用户行为状态对模型认证性能的影响。在实际部署中，认证系统须在注册阶段收集用户手势数据。由于用户认证时可能处于不同的行为状态，注册与认证时的状态不一致情况常有发生，因此需评估模型在此差异下的身份认证能力。

数据集中记录了用户坐和走路两种状态下的手势数据。其中，60 名用户在两种状态下以相同方式使用手机（如均用右手操作），30 名用户则改变了使用方式（如从右手换为左手持机并操作）。

本文设置了三种实验场景（图 15）。场景 1 利用以相同方式使用手机的用户的数据，划分为训练集与测试集，训练样本通过 T1 和 C1 生成，测试样本通过 C3 生成。因 T1/C1 将坐和行走数据分开处理，模型未学习行为变化对手势的影响。

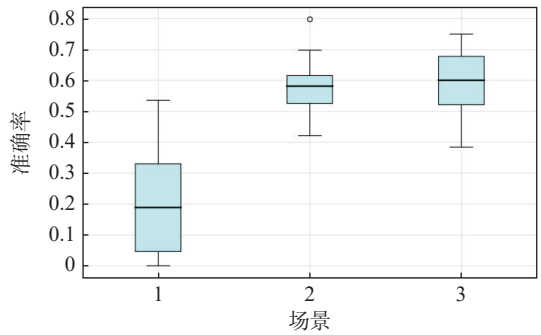

图 15　不同场景下的认证准确率

Fig. 15　Authentication accuracy in different scenarios

场景 2 使用的数据集与场景 1 相同。与场景 1 不同的是，场景 2 使用 T2 和 C2 生成了训练样本集。测试样本集的生成与场景 1 相同。由于 T2 和 C2 融合了用户在坐和行走两种状态下的手势数据，这使得模型有机会从因用户行为变化导致手势变化的信息中学习跨行为状态的特征。

场景 3 利用在坐和行走两种状态下以不同方式使用手机的用户对应的数据集。训练样本通过 T2 和 C2 生成，测试样本集通过 C3 生成。场景 3 与场景 2 的区别在于用户使用手机方式的差异。在场景 3 中，用户在行为变化的同时切换了使用手机的方式，而在场景 2 中，用户并未因行为变化切换使用手机的方式。场景 3 的目的是研究用户使用手机的方式的变化对模型性能的影响。

在训练样本量固定为 12 800 的条件下，图 15 展示了 StrokeFG 在三种场景下识别合法用户的准确率。可以看出，场景 2 中的模型性能显著高于场景 1，场景 1 的准确率集中在 5%~33%，而场景 2 则达到 52%~62%。场景 3 与场景 2 的性能接近。

场景 1 与场景 2 的性能差异表明，即使用户以相同方式绘制手势，不同行为状态下的手势仍存在明显差异。尽管在训练集中加入了行为变化带来的手势变异，StrokeFG 的准确率仍低于第 6.5 节中的结果，说明不同行为状态下用户手势多样性的增加，导致类间差异减小、类内差异增大，从而降低了模型性能。另外，场景 3 与场景 2 表现相似，说明即使用户改变手机使用方式，不同方式下绘制的手势仍具有一定的相似性。总体而言，不同行为状态使得用户手势更趋多样，增加了身份识别的难度。

然而，如第 6.5 节和第 6.9 节所示，同一行为状态下用户手势具有较高的稳定性，跨行为状态带来的差异可通过同时收集用户坐和行走两种状态下的注册样本有效缓解。6.9 节的结果证明了 StrokeFG 的泛化能力，StrokeFG 可以从用户不同行为状态的手势模板中提取具有区分力的特征并验证用户的身份。

7 结束语

本文基于卷积神经网路、多层感知机以及孪生网络技术，设计了一个增强自由手势认证系统。除了识别手势轨迹的相似度，该系统能够进一步从行为生物数据中自动提取具有区分力的隐式特征并有效识别用户的身份。相较于 DTW 和 Protractor，StrokeFG 能够有效处理不同用户的相同手势。在坐状态下，假阳率分别降低了 89.86% 和 91.24%；在行走状态下，假阳率分别降低了 86.56% 和 87.96%。在 α-猜测复杂度指标下，StrokeFG 在坐和行走两种行为状态下分别提供了 6.60~8.41 比特与 6.50~8.21 比特的安全性，有效提升了模型的安全边界。此外，StrokeFG 对未见手势展现出良好的泛化能力。虽然行为状态变化会导致 StrokeFG 的性能下降，但该问题可以通过注册不同行为状态下的模板样本缓解，以增强 StrokeFG 的鲁棒性。

参考文献：

[1] WANG D，GU Q，HUANG X，et al. Understanding human-chosen pins：Characteristics，distribution and security[C]//Proceedings of ACM ASIACC. New York，NY，USA：ACM，2017：372-385.

[2] CHO G，HUH J H，CHO J，et al. Syspal：System guided pattern locks for android[C]//2017 IEEE symposium on security and privacy（S&P）. San Jose，CA，USA：IEEE，2017：338-356.

[3] WU C，HE K，CHEN J，et al. Liveness is not enough：Enhancing fingerprint authentication with behavioral biometrics to defeat puppet attacks[C]//Proceeding of USENIX Security Symposium 2020. Boston，MA，USA：USENIX，2020：2219-2236.

[4] WU Z，CHENG Y，ZHANG S，et al. UniID：Spoofing face authentication system by universal identity[C]//Proceedings of ISOC NDSS 2024. San Diego，CA，USA：ISOC，2024：1-15.

[5] ZHOU M，SU S，WANG Q，et al. PrintListener：Uncovering the vulnerability of fingerprint authentication via the finger friction sound[C]//Proceedings of ISOC NDSS 2024. San Diego，CA，USA：ISOC，2024：1-16.

[6] DE LUCA A，VON ZEZSCHWITZ E，NGUYEN N D H，et al. Back-of-device authentication on smartphones[C]//Proceedings of ACM CHI 2013. New York，NY，USA：ACM，2013：2389-2398.

[7] UELLENBECK S，DÜRMUTH M，WOLF C，et al. Quantifying the security of graphical passwords：The case of android unlock patterns[C]//Proceedings of ACM CCS 2013. New York，NY，USA：ACM，2013：161-172.

[8] RAMACHANDRA R，BUSCH C. Presentation attack detection methods for face recognition systems：A comprehensive survey[J]. ACM Computing Surveys（CSUR），2017，50（1）：1-37.

[9] ROY A，MEMON N，ROSS A. Masterprint：Exploring the vulnerability of partial fingerprint-based authentication systems[J]. IEEE Transactions on Information Forensics and Security（TIFS），2017，12（9）：2013-2025.

[10] WANG D，ZHANG Z，WANG P，et al. Targeted online password guessing：An underestimated threat[C]//Proceedings of ACM CCS 2016. New York，NY，USA：ACM，2016：1242-1254.

[11] XIU K，WANG D. PointerGuess：Targeted password guessing model using pointer mechanism[C]//Proceedings of USENIX Security Symposium 2024. Philadelphia，PA，USA：USENIX，2024：5555-5572.

[12] 龚雪鸾，陈艳姣，王涛，等. SeqGANPass：使用序列生成式对抗网络进行口令猜测 [J]. 电子学报，2023，51（5）：1148-1153.
GONG X L，CHEN Y J，WANG T，et al. SeqGANPass：Password guessing with sequence generative adversarial Nets[J]. Acta Electronica Sinica，2023，51（5）：1148-1153.

[13] 周满，李向前，王骞，等. 基于声感知的移动终端身份认证综述 [J]. 软件学报，2025，36（5）：2229-2253.
ZHOU M，LI X Q，WANG Q，et al. Survey on acoustic-sensing-based authentication on mobile devices[J]. Journal of Software，2025，36（5）：2229-2253.

[14] 姜奇，文悦，张瑞杰，等. 面向智能手机的自适应触屏持续认证方案 [J]. 电子学报，2022，50（5）：1131-1139.
JIANG Q，WEN Y，ZHANG R J，et al. An adaptive touchscreen based continuous authentication scheme for smart phones[J]. Acta Electronica Sinica，2022，50（5）：1131-

1139.

[15] 高焕芝，曹秀莲，王磊，等. 基于动态手势的身份认证方法及其在智能手机上的应用 [J]. 电子学报，2014，42（9）：1857-1862.
GAO H Z，CAO X L，WANG L，et al. An identity authentication method based on dynamic gesture and its application in mobile phone[J]. Acta Electronica Sinica，2014，42（9）：1857-1862.

[16] SONG W，KANG W. Depthwise temporal non-local network for faster and better dynamic hand gesture authentication[J]. IEEE Transactions on Information Forensics and Security（TIFS），2023，18（3）：1870-1883.

[17] CHEON E，SHIN Y，HUH J H，et al. Gesture authentication for smartphones：Evaluation of gesture password selection policies[C]//Proceedings of IEEE S&P 2020. San Francisco，CA，USA：IEEE，2020：249-267.

[18] YANG Y，CLARK G D，LINDQVIST J，et al. Free-form gesture authentication in the wild[C]//Proceedings of ACM CHI 2016. New York，NY，USA：ACM，2016：3722-3735.

[19] LIU C，CLARK G D，LINDDQVIST J. Where usability and security go hand-in-hand：Robust gesture-based authentication for mobile systems[C]//Proceedings ACM CHI 2017. New York，NY，USA：ACM，2017：374-386.

[20] SHERMAN M，CLARK G，YANG Y，et al. User generated free-form gestures for authentication：Security and memorability[C]//Proceedings of ACM MOBISYS 2014. New York，NY，USA：ACM，2014：176-189.

[21] LI Y. Protractor：A fast and accurate gesture recognizer[C]//Proceedings of ACM CHI 2010. New York，NY，USA：ACM，2010：2169-2172.

[22] LI L，ZHAO X，XUE G. Unobservable re-authentication for smartphones[C]//Proceedings of ISOC NDSS 2013. San Diego，California：ISOC，2013：57-59.

[23] FEREIDOONI H，KÖNIG J，RIEGER P，et al. Authentisense：A scalable behavioral biometrics authentication scheme using few-shot learning for mobile platform[C]//Proceedings of ISOC NDSS 2023. San Diego，California：ISOC，2023：1-16.

[24] TARANTA II E M，SAMIEI A，MAGHOUMI M，et al. Jackknife：A reliable recognizer with few samples and many modalities[C]//Proceedings ACM CHI 2017. New York，NY，USA：ACM，2017：5850-5861.

[25] CHEON E，HUH J H，OAKLEY I. GestureMeter：Design and evaluation of a gesture password strength meter[C]//Proceedings ACM CHI 2023. New York，NY，USA：ACM，2023：1-19.

[26] CHEON E，OAKLEY I. Securing gesture passwords against shoulder surfing using behavioral features[C]//Proceedings ACM CHI 2025. New York，NY，USA：ACM，2025：1-8.

[27] FRANK M，BIEDERT R，MA E，et al. Touchalytics：On the applicability of touchscreen input as a behavioral biometric for continuous authentication[J]. IEEE Transactions on Information Forensics and Security（TIFS），2012，8（1）：136-148.

[28] SHEN C，ZHANG Y，GUAN X，et al. Performance analysis of touch-interaction behavior for active smartphone authentication[J]. IEEE Transactions on Information Forensics and Security（TIFS），2015，11（3）：498-513.

[29] SHEN Z，LI S，ZHAO X，et al. IncreAuth：Incremental learning-based behavioral biometric authentication on smartphones[J]. IEEE Internet of Things Journal，2023，11（1）：1589-1603.

[30] FIERREZ J，POZO A，MARTINEZDIAZ M，et al. Benchmarking touchscreen biometrics for mobile authentication[J]. IEEE Transactions on Information Forensics and Security（TIFS），2018，13（11）：2720-2733.

[31] SERWADDA A，PHOHA V V. When kids' toys breach mobile phone security[C]//Proceedings of ACM CCS 2013. New York，NY，USA：ACM，2013：599-610.

[32] KHAN H，HENGARTNER U，VOGEL D. Targeted mimicry attacks on touch input based implicit authentication schemes[C]//Proceedings of ACM MOBISYS 2016. New York，NY，USA：ACM，2016：387-398.

[33] ZHAO B Z H，ASGHAR H J，KAAFAR M A. On the resilience of biometric authentication systems against random inputs[C]//Proceedings of ISOC NDSS 2020. San Diego，CA，USA：ISOC，2020：1-18.

[34] LI Y，HU H，ZHOU G. Using data augmentation in continuous authentication on smartphones[J]. IEEE Internet of Things Journal，2018，6（1）：628-640.

[35] ZOU Q，WANG Y，WANG Q，et al. Deep learning based gait recognition using smartphones in the wild[J]. IEEE Transactions on Information Forensics and Security（TIFS），2020，15（4）：3197-3212.

[36] SHEN Z，LI S，ZHAO X，et al. MMAuth：A continuous authentication framework on smartphones using multiple modalities[J]. IEEE Transactions on Information Forensics and Security（TIFS），2022，17（3）：1450-1465.

[37] SHEN C，LI Y，CHEN Y，et al. Performance analysis of multi-motion sensor behavior for active smartphone authentication[J]. IEEE Transactions on Information Forensics and Security（TIFS），2017，13（1）：48-62.

[38] ZHAO C，GAO F，SHEN Z. AttAuth：An implicit authentication framework for smartphone users using multi-modality data[J]. IEEE Internet of Things Journal，2023，11（4）：6928-6942.

[39] VAN HAMME T，GAROFALO G，RÚA E A，et al. A novel evaluation framework for biometric security：Assessing guessing difficulty as a metric[J]. IEEE Transactions on Information Forensics and Security（TIFS），2024，19（9）：8369-8384.

[40] NEGI，PARIMARJAN，et al. K-means++ vs. behavioral biometrics：One loop to rule them all[C]//Proceedings of ISOC NDSS 2018. San Diego，California：ISOC，2018：1-18.

[41] DUAN Y，WANG D，FU Y. Security analysis of master-password-protected password management protocols[C]//Proceedings of IEEE S&P 2025. San Francisco，CA，USA：IEEE，2025：701-719.

[42] KIM，HYOUNGSHICK，JUN HO HUH. PIN selection policies：Are they really effective?[J]. Computers & Security，2012，31（4）：377-390.