两步认证：采用声音和人脸识别的多生物特征系统

Kuan Wei Chen[1], Ting Yi Lin[1], Wen Ren Yang[1], Aryan Kesarwani[2], Riya Singh[2]

[1]国立彰化师范大学电子工程系，台湾彰化[2]Apex Institute of Technology，昌迪加尔大学，Gharuan，印度旁遮普 elaping5691@gmail.com, brian104120@gmail.com, wry87e@cc.ncue.edu.tw, kesarwaniaryan76@gmail.com, riya0502singh@gmail.com,

关键词：多生物特征、认证、识别、机器学习 G

摘要

本研究提出一种认证系统，整合人脸和语音识别技术，为配备摄像头和麦克风的小型群体提供更安全的认证方式，并着重实现成本效益。通过实际应用和严格测试，本项目展示了多种生物特征模态的有效整合，兼顾安全性与易用性。人脸识别模型和语音识别模型均具有较高的准确率。未来，语音验证过程可纳入实时认证流程，通过使用随机词语作为验证码，增强双重认证安全。代码和模型已公开发布于 https://github.com/NCUE-EE-AIAL/Two-step-Authentication-Multi-biometric-System。

1. 引言

与以往旨在提高多生物特征系统准确率的研究不同， [1], 本研究侧重于通过两步生物特征识别流程降低硬件成本。为彰显多模态系统的优势， [2, 3], 我们的决策级融合系统利用手机和笔记本电脑的内置摄像头与麦克风。这种方法受 [4] 启发，提供了一种经济高效且稳健的解决方案。

图 1 展示了该系统的工作流程，首先进行人脸识别。这简化了后续的语音认证流程，因为语音识别模型只需验证已识别个体的声音。该方法使我们的系统能够高效、准确地识别小群体中的个体

.

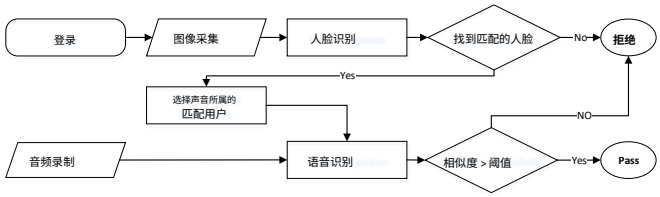


图.1 系统工作流程

2. 方法

2.1 数据集

收集了包含五位人物图像的数据集。为了提高机器学习模型的鲁棒性，对原始照片应用了多种数据增强技术。这些增强技术扩大了初始数据集，使最终照片数量达到 924 张，从而增强了模型对新数据的泛化能力。为了减少背景噪声并使数据集聚焦于相关部分，使用 MTCNN算法从图像中识别并裁剪出面部区域。

使用 LibriSpeech数据集评估语音模型的性能。[5] 训练过程中使用了包含 921 位说话者的 train-other-360 子集。验证时使用了包含 40 位说话者的 test-clean 子集。

2.2 神经网络架构 表 1 下方展示了声音

模型架构以卷积神经网络（CNN）为核心。首先应用一个步幅为2×2的5×5卷积滤波器，随后是由三个卷积层组成的残差块：1×1、3×3和1×1，且步幅均为1×1。在均值层和仿射层之后，应用具有512维的三元组损失。[6]

表 1 语音识别模型的架构

层	结构	步幅	参数
卷积64	$5 \times 5,\ 64$	2×2	1.9K
残差64	$\begin{bmatrix} 1 \times 1,\ 64 \\ 3 \times 3,\ 64 \\ 1 \times 1,\ 64 \end{bmatrix} \times 3$	1×1	46K × 3
Conv128	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3, \end{bmatrix}$ 5 × 5, 128 128 × 3	2×2	205K
Res128	128 1 × 1,	1×1	228K × 3
256 通道卷积层	$5 \times 5, 256$	2×2	821K
256 通道残差块	$\begin{bmatrix} 1 \times 1,\ 256 \\ 1 \times 3,\ 256 \\ 1 \times 1,\ 256 \end{bmatrix} \times 3$	1×1	953K × 3
512 通道卷积层	$5 \times 5, 512$	2×2	3.28M
512 通道残差块	$\begin{bmatrix} 1 \times 1,\ 512 \\ 1 \times 3,\ 512 \\ 1 \times 1,\ 512 \end{bmatrix} \times 3$	1×1	2.89M × 3
mean	-	-	0
仿射	2048×512	-	1M
三元组	-	-	0
总计	-	-	16.8M

图2展示了所提出的人脸识别模型，该模型基于VGG-16架构。顶层和全连接层采用专门设计的自定义层构建，用于通过学习和区分数据集中五个人物独特的面部特征来完成分类。数据层经过全面剪枝，旨在获得最佳准确率。为适应缩小至224×224像素的照片，对输入层进行了调整。由于该架构所需资源较少且能够产生准确结果，因此具有可行性。

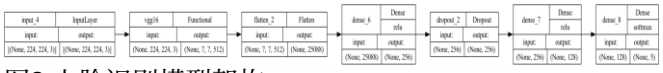


图2 人脸识别模型架构

3. 结果 3.1

人脸识别　　识

表2展示了通过运行人脸识别模型的多个训练轮次获得的VGG-16结果，具体信息见图3中的混淆矩阵。为评估人脸识别模型的性能，下面给出了图4。

表2 人脸识别结果 系统 准确率

F1-分数		e	召回率	精确率
VGG16 95.135%	95.732%		95.153%	96.317%

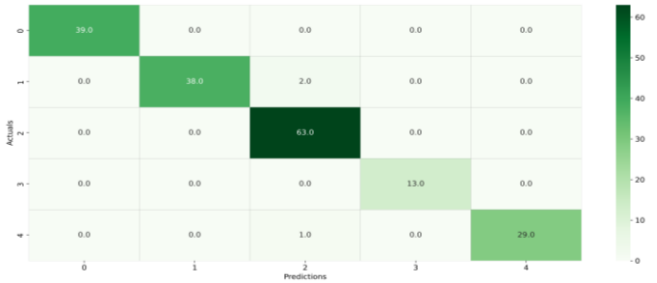


图 3 人脸识别模型的混淆矩阵

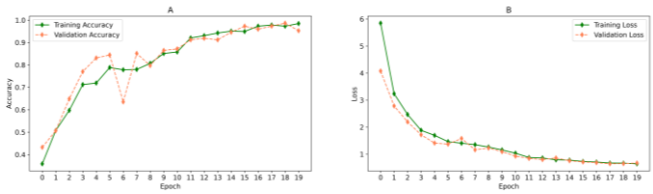


图4 人脸识别模型训练轮次图，显示（a）训练准确率和验证准确率（b）训练损失和验证损失。

3.2 语音识别 音频经过处理

并使用基于能量的语音活动检测来去除静音部分。随后提取特征，生成 64 维滤波器组系数。表 3 展示了语音识别模型经过 60 轮训练后的结果。准确率和等错误率（EER）体现了出色的性能，而召回率和精确率则凸显了模型的稳定性。训练采用默认参数的 Adam优化器，模型使用小批量为 32、间隔为 0.1 的三元组损失 α。图 5 展示了语音识别模型的评估结果。

表3 语音识别结果 系统 准确率 等

错误率			召回率	精确率
模型	98.9%	3.456%	86.48%	88.65%

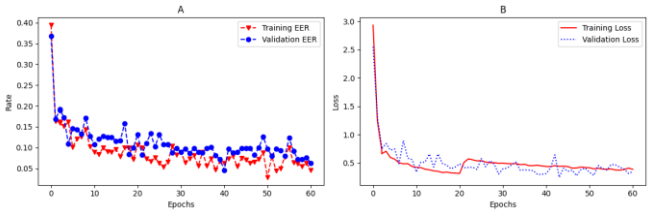


图 5 语音识别模型训练轮次图：(a) 训练和验证等错误率 (b) 训练和验证损失。

4. 结论多模型生

生物特征认证系统融合了人脸识别和语音识别技术。借助设备摄像头和麦克风，该系统提供了一种经济高效的解决方案。人脸识别模型的准确率达到 95.1%，而语音识别模型的等错误率为 3.456%。这些结果凸显了我们的多模态生物特征认证系统在提供稳健、可靠的身份识别能力方面的有效性。

此外，在语音验证过程或连续多模态生物特征认证系统中使用随机句子作为验证码，可实现 [2], 中提及的实时认证流程，防止机器攻击，使其适用于通用设备。

5. 致谢本项目部分由

由台湾教育部通过台湾体验教育计划（TEEP）项目编号 1122505013 资助。

6. 参考文献[1] Rihan FKhalil F

Aoude M.：《运用多模态技术提高生物特征认证准确率》，ACTEA 2023，DOI：10.1109/ACTEA58025.2023.[2] R. Ryu、S. Yeom、S. -H. Kim 和 D. Herbert：《连续多模态生物特征认证方案：系统综述》，发表于 IEEE Access，第 9 卷，第 34541-34557 页，2021 年，doi：10.1109/ACCESS.2021.3061589。[3] Alam RAkilarasu G：《评估人脸与语音识别中生物特征认证的应用》，（2022）[4] 10194130Zhang XCheng DJia P 等：《一种高效的基于安卓的多模态生物特征认证系统，结合人脸与声音》，DOI：10.1109/ACCESS.2020.2999115[5] V. Panayotov、G. Chen、D. Povey 和 S. Khudanpur：《Librispeech：一个基于公版有声读物的自动语音识别语料库》，2015 年 IEEE声学、语音与信号处理国际会议，澳大利亚昆士兰州南布里斯班，2015 年，doi：10.1109/ICASSP.2015.7178964。[6] R. Li、N. Li、D. Tuo、M. Yu、D. Su 和 D. Yu：《用于文本无关说话人验证的边界判别大间隔余弦损失》，ICASSP 2019—2019 年 IEEE 国际声学、语音与信号处理会议（ICASSP），英国布莱顿，2019 年，第 6321-6325 页，doi：10.1109/ICASSP.2019.8682749。