InMyFace：基于惯性传感器与机械肌动图的传感器融合，用于可穿戴面部活动识别

Hymalai Bello[a, *]，Luis Alfredo Sanchez Marin[a]，Sungho Suh[a, b]，Bo Zhou[a, b]，Paul Lukowicz[a, b]

[a]德国人工智能研究中心（*DFKI*），*67663* 凯撒斯劳滕，德国[b]凯撒斯劳滕-兰道莱茵兰-普法尔茨技术大学计算机科学系，*67663* 凯撒斯劳滕，德国

arXiv:2302.04024v1 [cs.LG] 8 Feb 2023

摘要

面部活动识别是一个已得到充分研究（但并非易事）的计算机视觉问题。然而，要实现可靠的解决方案，需要摄像头能够清晰拍摄面部，而这在可穿戴场景中往往难以实现。此外，在可穿戴应用中，系统会伴随用户进行日常活动，摄像头持续运行可能会带来隐私（及法律）方面的问题。本文提出了一种替代方案，通过融合可穿戴惯性传感器、平面压力传感器和声学机械肌电图（肌肉声音）实现面部活动识别。这些传感器被隐蔽地安装在一顶运动帽中，用于监测与面部表情相关的面部肌肉活动。我们介绍了集成式可穿戴传感器系统，阐述了数据融合和分析方法，并通过一项实验评估了该系统。实验参与者为来自不同文化背景（八个国家）的十三名受试者，男女均有（六名女性、七名男性）。采用每位用户单独训练一个模型的方案和后期融合方法，在融合所有传感模态时，系统的平均F1分数达到85.00%。采用跨用户验证和所有用户共用一个模型的方案时，十三名参与者（六名女性、七名男性）的F1分数达到79.00%。此外，采用混合融合（跨用户）方法并识别六个类别时，八名用户的平均F1分数达到82.00%。在跨用户研究中，实验结果可与最先进的非摄像头解决方案相媲美。此外，我们独特的参与者群体体现了该方法的包容性和普适性。

关键词：多模态融合、面部表情、活动识别、机械肌动图

1. 引言

面部表情是人际互动中的一种关键交流方式；有时，它们比明确的语言交流更能传达信息 [3, 4]。面部表情会影响我们的社交智能，以及建立人际情感联系的能力 [5]。因此，在许多普适计算应用中，人们都对面部表情监测进行了研究。例如，新型人机界面 [6, 7], 、学习反馈 [8], ，以及识别汽车驾驶员的疲劳或情绪 [9, 10]。

面部表情识别是一个复杂的问题，原因有很多。首先，也是最重要的，其个体间差异很大 [11]。它受到文化背景 [12, 13], 、年龄、性别 [14], 、种族以及其他个人特征的影响，因此许多相关研究采用了依赖用户的解决方案 [7]（见表 1）。近期， [15],作者正式研究了性别均衡的数据集对面部表情识别（快乐、悲伤、惊讶、恐惧、厌恶和愤怒）结果公平性的影响。他们得出结论，使用混合数据集进行训练在所有情况下都能取得最佳结果。此外，使用高度偏倚的数据集进行训练会损害公平性，尤其是在分类特定表情时。此外，不同的表情类别之间可能只有细微差异（例如，愤怒和悲伤，如图1所示）。

摄像头是最广泛使用的表情识别方案，预期准确率超过 90.00% [16, 17, 18]。然而，在可穿戴应用中，摄像头并非总是合适的方案。首先，要将摄像头安装在用户身上，既不引人注意、适合日常使用，又能同时清晰拍摄整张脸，这几乎难以实现。光照条件和遮挡也可能造成问题。此外，Google Glass 的例子表明，在日常生活中始终佩戴摄像头可能会让人感到尴尬，在某些国家甚至属于违法行为。与非视觉传感器方案（例如 [19]中提出的方案）相比，基于计算机视觉的方法通常也会占用更多内存，功耗更高。

可穿戴设备领域对用于面部监测的非视觉感知方法表现出浓厚兴趣。然而，正如相关研究所述，大多数系统难以兼顾可穿戴性限制与获取干净且信息丰富的信号这一需求。问题的核心在于，面部肌肉运动涉及面部的大面积区域，而传感器无法在不引人注目的情况下覆盖这些区域。此外，传感器放置仅限于以下区域：

*通讯作者电子邮箱：hymalai.bello@dfki.de（Hymalai Bello）



图1：带有传感器信号示例的面部肌肉活动词典；来自华沙照片集的 7 种面部表情 [1] 和来自 [2]的 2 种手势。双通道原始音频数据、十三个梅尔频率倒谱系数（双通道单声道音频）、力敏电阻、压电薄膜、方向和加速度。

围绕眼镜、头部（例如帽子下面）以及耳部周围。挑战在于利用这些有限区域，捕捉与整体面部活动相关的信息。此前研究过的方法包括惯性测量单元（IMU） [20], 超声波 [21], 和光反射式传感器 [22]。与摄像头相比，面部监测的声学感知方案更加紧凑，数据量和计算需求更小，功耗也更低 [21, 23]。然而，基于摄像头/音频的设计带来的隐私顾虑仍限制着用户的接受度。

量化肌肉运动的一种标准方法是测量其电激活模式，即通过肌电图（EMG）进行感知。EMG测量依赖于检测人体微弱的电信号。这项技术需要与皮肤表面保持稳定的电连接，通常通过嵌入肌肉中的微针，或至少使用湿电极来实现。这两种方案在日常可穿戴场景中都可能面临挑战 [24]。另一方面，肌机械图（MMG）测量的是肌纤维横向运动产生的低频机械信号 [25, 26]。MMG无需与皮肤建立电连接，并可通过多种类型的传感器进行检测，例如IMU、压电传感器、麦克风、压力传感器、激光传感器等 [27]。在 [2], 中，有人提出了面部肌肉声学肌机械图（AMMG）。作者在颞肌、额肌和咬肌周围布置听诊器麦克风，以检测十种面部表情。跨用户情形（八名志愿者）的F1分数为54.00%，展现出一种被动且注重隐私的方法在检测面部肌肉运动方面的良好前景。然而，这种方案佩戴起来仍不舒适，也远称不上便携。 [28], 中提出了一种压力肌机械图（PMMG）方案，其对七种华沙照片集面部表情的分类准确率为38.00% [1]。作者展示了一种隐蔽、被动、基于纺织品头带的设计在监测面部肌肉活动方面的潜力。

或者，人们研究了多模态方法，以克服独立传感模态的局限性。其思路是结合多个具有互补信息的数据源，以减少歧义、补充所研究情境的完整性（“表征增益”）、提高信噪比（假设误差来源彼此独立），并增强对模型决策的信心（“鲁棒性增益”）。在 [29], 中，作者提出将惯性测量单元与智能眼镜中的 16 个光学传感器相结合，以检测八种时序面部动作。对于每人使用一个模型来识别面部动作单元（AU：AU12、AU27、LP、AU1+2、AU4、AU43、AU46R、AU61），所得 F1分数为 91.10 % [30]。用于检测面部肌肉运动的光学传感器使系统只能在稳定的光照条件下运行。在 [31], 中，另有一种基于惯性测量单元与眼电图融合的方案，用于识别亲吻手势；在跨用户设置下，其准确率达到 74.33%。惯性测量单元是 the 多模态方案中的共同基础，因此它是需要考虑的信息来源。

本研究提出了一种多模态系统，将互补的传感模态融合在一起，例如惯性传感 IMU（方向和加速度）、AMMG 和 PMMG（力敏电阻和压电薄膜），并将其集成到运动帽配件中。待评估的面部表情来自华沙照片集（七种表情）和 [2]（两种面部运动）；因此，更便于与未来的解决方案进行比较，见图1。

总之，**我们的贡献**如下：

- 我们提出了一种用于监测面部肌肉运动的多模态传感替代方案，该方案基于惯性、平面压力和声学传感器，并将这些传感器分布在不太妨碍活动的可穿戴配件（运动帽）中。此外，该方案还可调整，并有望以不显眼的方式集成到其他头戴式平台（如眼镜和发带）中，同时不影响传感能力。

- 我们采用基于依赖传感器的神经网络的模块化多模态融合方法，使用后期融合方案，内存占用较低（$\leq$ 2 MB），以简化该方案未来在尺寸极小的可穿戴/嵌入式设备上的部署，并减少内存需求（4 MB 至 16 MB 闪存）。

- 我们对来自不同文化背景（八个国家）且性别各异（六名女性和七名男性）的十三位参与者开展了用户研究。我们独特的数据集展现了研究结果的包容性和普适性。在评估中，我们使用华沙照片集 [1] 以及来自 [2] 的两种面部动作。

- 我们使用混合融合方法评估系统：针对每种传感模态进行降维，并采用局部连接的Inception模块，以识别图1所示面部表情词典中八位最佳模仿者的表情，以及六个类别。

本文结构如下。第2节回顾可穿戴设备和机械肌动图（MMG）在面部表情监测方面的相关研究。第3节详细介绍材料与方法，包括装置、实验设计和信息融合方法。第4节展示实验结果并讨论研究局限。第5节总结全文，并对未来工作作出一些展望。

2. 相关工作

本节将探讨与 InMyFace 相关的既有研究项目，范围涵盖可穿戴设备上的面部监测以及基于压力/音频的机械肌动图。

表 1：与面部活动识别领域最先进的可穿戴传感方法的比较

研究	设备	表情	方法	性能
耳道(主动式)[21]	耳塞中的超声波	**21** 定制	SVM **用户-相关，用户 = 11**	**62.5 %** 对应 21 种面部表情-次，并且 **90.00 %** 达 6
ExpressEar（被动式）[32]	耳塞式惯性测量单元	**30**个动作单元	每位用户一个时序卷积网络，**用户 = 12**	准确率**89.90 %**
Face 命令（主动式）[29]	16 光学传感器和惯性测量单元智能眼镜	**8** 时序手势	每位用户一个卷积神经网络，**用户 = 13**	F1值 **91.10 %**
耳塞式耳机（被动式）[20]	耳部的惯性测量单元	**3** 头部运动	层次分类-化 **依赖用户，用户 = 21**	微笑、说话和打哈欠的F1分数 **84.79 %**
FaceListener（主动式）[33]	超声波 头部-手机	**7** 面部表情	一个长短期记忆网络，**每位用户，用户 = 5**	准确率**80.00 %**
Expressure（被动式）[28]	纺织品 压力 in 头带	**7** 面部表情	SVM **用户 = 20** **跨用户** ，	准确率**38.10 %**
麦克风（被动式）[2]	6 听诊器 微型-型面部	**10** 面部活动	SVM **用户 = 8** **跨用户，**	F1分数 **54.00 %**
FaceGlasses（被动式）[34]	2 眼镜中的摄像头形状设备	**7** 伪面部图像	比较 with 存储图像 **用户-依赖型，用户 = 6**	准确率 **87.50 %** **中性，66.70 % 快乐表情和 71.40 % 惊讶**
Neckface（主动式）[35]	红外线 摄像头位于项链和颈带中	**8** 定制 面部运动与 Head 旋转	卷积神经网络实时跟踪 **依赖用户，用户 = 13**	平均绝对误差 **30.29**项链和 **25.61** 颈带
C-Face(被动式)[18]	2 耳部安装式摄像头-eras	**8** 面部 表情 (表情符号)	One 双向长短期记忆网络 **per 用户，用户 = 9**	准确率**88.60 %**
EmotGlass(被动式)[36]	摄像头， EDA and PPG	象限 of the 唤醒度–效价平面	SVM **用户-相关，用户 = 20**	准确率**76.09 %**
PhotGlasses（主动式）[22]	17 照片 反思性《传感器》	**8** 面部表情	一个支持向量机 **每位用户，用户 = 8**	准确率**78.10 %**
CanalSense（被动式）[37]	气压计 in ear-手机	**11 Face 相关动作**	One 随机 森林 **每位用户，用户 = 12**	准确率**87.60 %**
KissGlass（被动式）[31]	惯性测量单元与眼电眼镜中的图	**9** 亲吻手势和行走	**跨用户、用户 = 5**	准确率**74.33 %**
InfraredEar（主动式）[38]	耳部红外传感器阵列配件	**9** 面部动作	一个支持向量机 **每位用户，用户 = 5**	F1分数**97.00 %**
CapGlasses（主动式）[39]	眼镜中的电容通道眼镜	**12** 面部与头部手势-特征	One 随机 森林 **每位用户，用户 = 10**	准确率**89.60 %**
EarIO（主动式）[23]	声学 感知 in 耳机	**9** 定制 面部运动	神经网络 **依赖用户回归，用户 = 12**	平均绝对误差 **25.90**
Our 方法 Sportscap（被动式）	惯性测量单元、力敏电阻、压电传感器电学-tric 和麦克风	**参见图1**	卷积神经网络**针对每位用户及跨用户，用户 = 13**	针对每位用户的 F1 分数 **85.00 %**，跨用户 F1 分数**79.00 %** 针对 13 名用户，F1分数 **82.00%** 适用于 8 名用户和六个类别

2.1. 可穿戴设备上的面部监测

可穿戴设备领域已探索将不同的传感模态嵌入配件中，以进行面部表情监测。各种传感方法包括摄像头[16, 35]，惯性测量单元 [31, 20]，光 [22]，电容式 [39]，压电传感器 [40]，肌电图（EMG） [41]，机械肌动图 [28, 2]，基于音频的解决方案 [2, 33, 42]，以及许多其他方法。表 1 对最先进的方法进行了比较。

眼镜是一种常用于部署传感器的配件。最近， [39]在眼镜中嵌入了电容式传感器，用于识别 12 种面部动作和头部运动。他们探索了两种部署电极的方式：一种是向人体注入 12V 交流信号，以提高人体电势并提升信噪比；另一种则是在靠近面部/眼睛处产生强电场，以消除电容式传感中的接地依赖性 [43, 44] 并获得良好信号。 [22], 介绍了眼镜上的另一种主动传感方法，该方法使用 17 个光反射传感器对八种面部表情进行分类，每位用户的识别结果为 78.00–92.00%。 [18]通过眼镜在面部前方部署了紧凑型可穿戴摄像头。他们专注于追踪面部肌肉运动，并生成虚拟化身。基于摄像头和光传感器的方法都容易受到光照条件影响。

入耳式可穿戴设备正成为一种趋势，相关研究探讨了面部表情识别以及惯性测量单元的嵌入

[20], 红外线、 [38], 电场 [45] 和耳道内的超声波传感器 [21]。文献 [33] 提出了一种耳机超声设备，可利用朝面部发射的主动连续声音信号受到的扰动来追踪面部形变。文献 [23]提出了与[33] 类似的方法。作者使用自行设计的耳机，在两侧分别配备成对的扬声器和麦克风，使声波穿过面部传播。随后，他们根据麦克风输入信号的失真，追踪特定面部运动所引起的皮肤形变。文献 [23] 的结果表明，低功耗主动音频系统在追踪面部运动方面前景广阔。

主动式解决方案可能会带来负面后果或长期健康影响，而医生尚未对此进行直接调查。一些研究表明，有必要对可穿戴设备中使用的标准技术开展广泛研究，例如身体周围的辐射源/波（如 WiFi 和超声波） [46, 47, 48, 49, 50]。尽管在当前可穿戴技术的发展阶段，这种谨慎似乎过度或不切实际，但我们仍希望继续努力，减少志愿者和我们自身接触尚未被证明无风险的信号。

在本研究中，我们重点关注被动式解决方案，例如 [28], 中介绍的方法：将纺织品压力矩阵（机械肌动图）融入头带设计，以对七种面部表情进行分类，其准确率最高达到 38.00%。在 [2] 中，研究人员采用了另一种被动技术——声学机械肌电图（AMMG），对十种面部肌肉活动进行分类，F1分数为 54.00%。压电薄膜（PEF）已被用于实时 [19] ，以检测并分类皮肤形变，从而解码肌萎缩侧索硬化症患者的面部运动。该项工作旨在应用于临床环境中的非言语交流和神经肌肉监测。PEF 传感技术重量轻、可定制，并具备机械能量采集能力 [51]；因此，我们认为 PEF 技术值得在特定应用中开展研究和探索。在此，我们提出融合被动式传感技术，例如使用力敏电阻（ FSR）和压电薄膜（PEF）的压力肌机械图（PMMG）、基于方向和加速度的惯性传感，以及声学机械肌电图（ AMMG）。

2.2. 机械肌动图（*MMG*）

肌肉收缩与放松的周期性模式常见于许多日常人类活动中，例如骑行、跑步和大笑等。通过肌肉活动模式在强度、持续时间和频率上的变化来理解人类行为，是许多研究的主题。在实验室环境中，肌电图（EMG）是测量肌肉活动的主流选择；标准做法是在肌肉中植入表面电极或针电极。肌电图（EMG）在成本方面存在局限，通常仅适用于受控环境，而且信号强度容易受到皮肤阻抗、年龄和个体重量的影响 [52, 53]。在 [52], 中，作者将肌电图（EMG）与基于 MMG 和加速度数据的融合方法进行比较，以实现自动分割，并得出结论：MMG+加速度组合的结果与肌电图（EMG）相似，因此 MMG 适用于家庭等自然环境。

机械肌动图（MMG） 同样适用于评估骨骼肌活动，以及监测肌肉运动中产生的疲劳和力 [54, 55]。在相关文献中，MMG 的测量通常采用 IMU [55], 压电传感器 [56], 麦克风、压力传感器、激光测距传感器等 [27]。多模态 MMG 解决方案已有研究，例如 [56]；作者结合前臂加速度与分别置于肱二头肌、肱桡肌和胸大肌上的压电 MMG 传感器，研究帕金森病患者的肌肉机械振荡。在 [24], 中，提出了一种肌电融合（MMG+EMG）方法，以克服不可靠的传感器

皮肤界面在上肢假肢操控中的问题。综上所述，MMG 是一个广泛且活跃的研究领域。在我们的工作中，我们尤其关注 MMG 的两个类别：声学机械肌电图（AMMG）和压力肌机械图（PMMG）。

声学机械肌电图（AMMG）基于使用低频信号（ 2-200 Hz）和低于 50 Hz 的功率信号监测肌肉力量（收缩/放松） [57]。在 [58] 的举重活动中，使用麦克风记录参与者肱二头肌的活动，结果表明，AMMG 可作为测量肌肉疲劳的肌电图替代方法。在 [59], 中，作者发现，与相应的加速度计频谱相比，AMMG 受运动伪影的影响显著较小（$p \leq 0.05$）。他们得出结论：在减轻运动伪影影响至关重要的情况下，电容麦克风更适合用于 MMG 记录。对于涉及肢体运动的运动机能学研究（例如骑行），这一点尤为适用。在[2]中，研究人员首次将 AMMG 用于面部肌肉运动识别。 [2]中的基于听诊器的设计能够增强肌肉运动声音的强度，但由于直径较大，会牺牲用户舒适度。我们提出在额肌上使用两个 Inter-IC Sound（I2S）数字麦克风来感知 AMMG。与 [2], 中的驻极体麦克风相比，I2S 麦克风尺寸更小；此外，由于它们是数字麦克风，其输出受可穿戴配件细微运动的干扰也更小。

压力机械肌电图（PMMG）的主要用途是通过压力模式转换肌肉振动变化。PMMG 已用于 [28] 中的面部表情和认知负荷研究。作者在二十名志愿者的额头上使用纺织品压力矩阵（TPM），以识别华沙照片集 [1]中的七种面部表情。采用一种不显眼的纺织品被动传感方法识别七种面部肌肉活动，准确率为 38.00%（随机水平为 14.00 %），结果令人鼓舞。

3. InMyFace 方法

我们的系统结合惯性、压力和音频传感器，通过一顶不显眼的运动帽平台识别面部肌肉活动。图1展示了面部肌肉运动词典。惯性测量单元已展现出区分各种面部和头部相关动作的潜力 [29, 31]。PMMG 为面部手势识别提供了一种柔性且舒适的解决方案，并具有中等准确率 [28]。[2] 中采用了 AMMG 进行面部肌肉活动识别。然而，为实现较高的传感准确率，其设计体积庞大且引人注意。将三种传感模态与适当的传感器融合流程相结合，使我们能够在系统不显眼的情况下实现高准确率。

3.1. 系统架构与实现

原型硬件如图 2 所示。使用运动帽作为可穿戴配件，将力敏电阻传感器分布在额肌和颞肌上。两个采样率为 44.1千赫兹的 Inter-IC Sound (I2S) 麦克风按照图 2 所示分别放置在帽子的左侧和右侧。我们使用基于 Teensy 4.1 [1] 和 LoRaESP32 [2] 的定制印刷电路板 (PCB) 对传感器数据进行采样。Teensy 4.1 的集成 SD 卡用于存储声学数据。麦克风开口朝向额肌，以捕获机械声音信息 (AMMG)。分析的声音信息来自肌肉运动，因此这是一个注重隐私的系统（不采集语音或环境声音）。成对的力敏电阻传感器和 PEF 放置在颞肌的左右两侧。选择力敏电阻传感器和 PEF 来测量面部肌肉运动产生的 PMMG。运动帽帽檐处安装了一个惯性测量单元，用于捕获与面部表情相关的头部运动。力敏电阻传感器、PEF 和惯性测量单元的数据通过蓝牙串口 (BT) 传输至手机应用程序 (Flutter 框架[60])，并保存为 JSON文件以供后续分析。力敏电阻传感器、PEF 和惯性测量单元数据的采样率约为 100 Hz。蓝牙通信方案限制了数据采集能力，但速度仍足以捕获持续时间分别为 $\leq 200ms$ 和 ≥ 200 毫秒的微表情和宏表情 [61]。详细的传感器数据采集示意图见图 2。

传感器在头部周围的分布，以及所选传感器的尺寸和重量（见表 2），使我们的设计适合集成到其他头部配件中，例如发带和眼镜。在我们的运动帽中，所选电池（1300 mAh）和原型的主板（6.5 x 4.3 cm）对尺寸和重量有不利影响；未来的设计可以缩小两者，从而提升用户舒适度。

原型在 4.97V 下的功耗约为 0.22A（连续模式下为 1.09 瓦） [3]。基于整体设计理念，有多种进一步降低功耗的可能性。一方面，力敏电阻-PEF 数据可触发惯性测量单元和音频数据采集；当颞肌未检测到压力时，功耗可降至 4.99V 下的 0.06A（0.3 瓦）。另一方面，PEF 是一种机械能来源。机械能被视为无处不在的环境能量，可转化为电能 [51]。利用压电技术进行机械能量采集是一个活跃的研究领域 [51, 62, 63]。此前也已有研究展示了从人体运动中采集能量并同时对这类运动进行分类 [64, 65, 62, 66]，这也可用于进一步降低功耗。

除了缓解功耗问题外，力敏电阻还具有抗运动伪影能力强的优点。因此，我们建议使用力敏电阻信号的斜率/梯度，作为自动分割输入数据并避免运动伪影的一种可能触发条件。例如，文献 [52] 中曾自动使用加速度数据的信号斜率来分割肌机械图数据。最后，尽管我们使用现成的非纺织品力敏电阻和 PEF 传感器进行快速原型设计，但值得注意的是，这两种技术目前都已应用于纺织品中 [28, 65, 67]。

3.2. 多模态传感器融合

根据融合在处理链中的位置，多模态传感器融合大致可分为早期融合和后期融合 [68]。我们主要采用的方法称为后期融合。融合在各个经过独立训练的网络的决策阶段（置信度分数）进行。后期融合方法的优点是能够独立提取每个传感器的特定模式，并可将每个依赖传感器的神经网络并行部署在多个微控制器上，从而降低识别延迟和每个微控制器的内存需求（每个网络 ≤ 2 MB）。后期融合的主要缺点是提取不同传感模态与通道之间互相关的潜力有限。此外，我们还探索了一种混合融合方案。在神经网络隐藏层中、决策层之前进行的融合称为混合融合。在我们的研究中，混合融合结构及其评估是根据后期融合的性能结果进行设计的。图3展示了后期融合和混合融合示意图。图3a **左侧**展示了使用集成神经网络在决策阶段之后拼接依赖传感器的模型。具体的依赖传感器模型将在第3.2.1节中说明。图3b **右侧**展示了混合融合方法的主要模块。这些模块包括传感器输入、Inception模块以及

[1]Teensy 4.1，Paul Stoffregen：https://www.pjrc.com/store/teensy41.html DLA：2023年2月7日[2]LoraESP32, Sparkfun：https://www.sparkfun.com/products/18074 DLA：2023年2月7日

[3]通用串行总线数字功率计：https://www.az-delivery.de/en/products/charger-doktor DLA：2023年2月7日

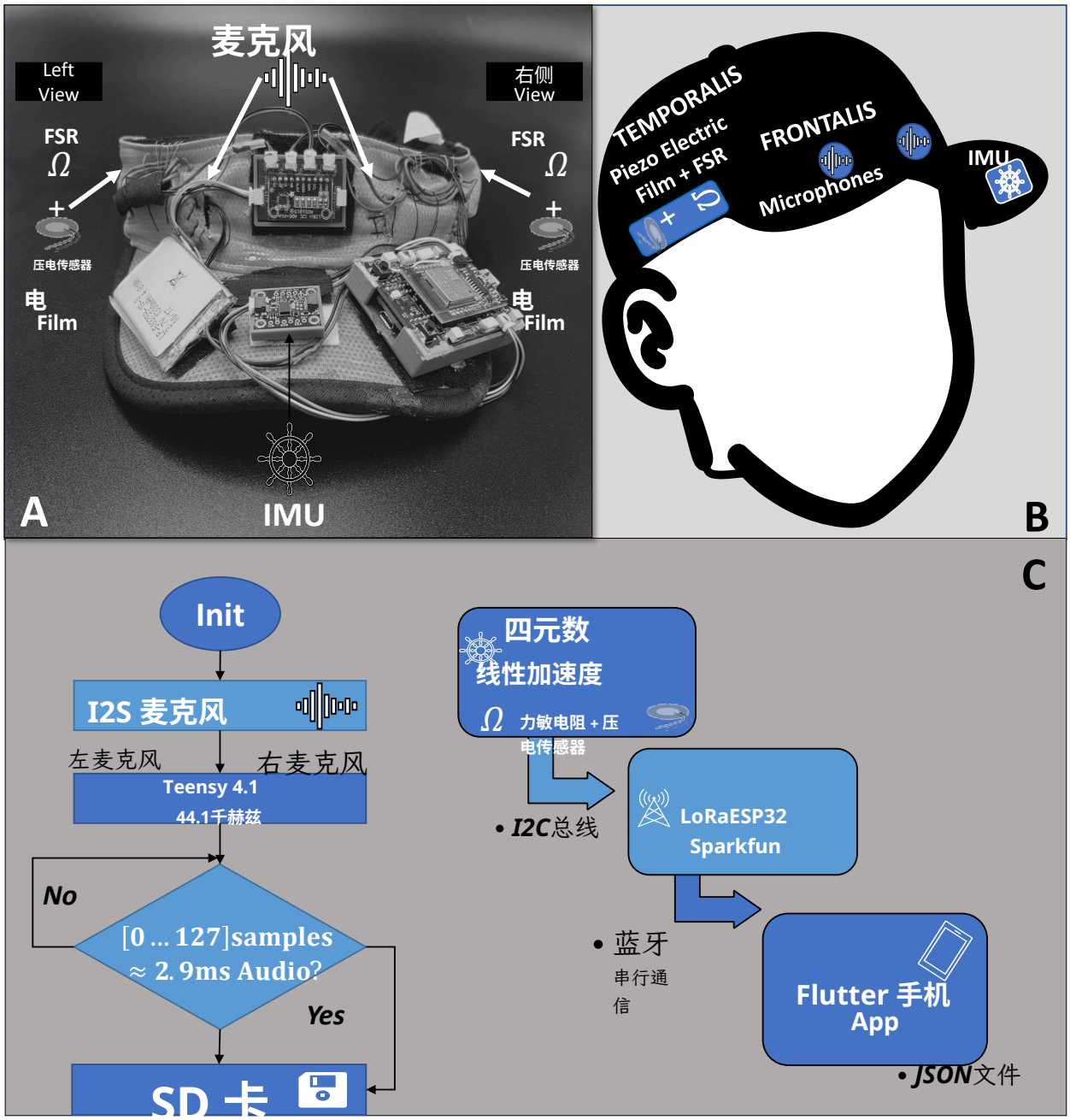
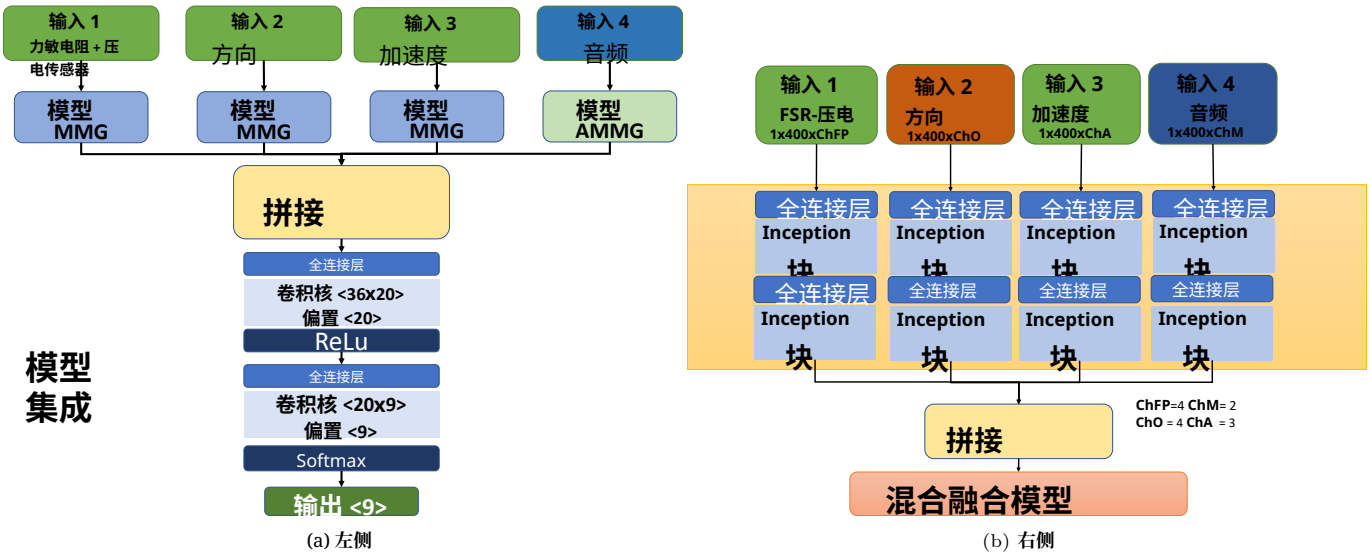


图 2：硬件原型和数据采集示意图。（**A**）传感器分布于运动帽上。**B** 参与者额肌和颞肌上的传感器放置。（**C**）使用定制印刷电路板的数据采集示意图。

表 2：传感器特性

传感器	制造商名称	尺寸（厘米）	重量（克）	优点
力敏电阻 [1]	Alpha MF01A-N-221-A01	直径 1.25	0.26	超薄且柔性
PEF [2]	TE SDT1-028K 屏蔽	4.45 x 1.97 x 0.32	0.30	低噪声，屏蔽且柔性
麦克风 [3]	Knowles SPH0645LM4H	0.35 x 0.26 x 0.09	0.40	高信噪比，达 65dB(A)，平坦的频率响应，全向
惯性测量单元 [4]	Bosch BNO055	0.38 × 0.52 × 0.11	0.15	输出融合后的传感器数据

[1] https://www.mouser.de/datasheet/2 13 01 __ 3_ 2_ 2_96_ 2_ 1_ 01 1915118 07, 2023 / /MF A c a c c a A - .pdf DLA：2 月 [2] https://www.te.com/usa-en/product-CAT-PFS0010 07, 2023 [3] .html DLA：2 月 https://media.digikey.com/pdf/Data%20 07, 2023 [4] 20 20 0645 4 Sheets/Knowles% Acoustics% PDFs/SPH LM H-B.pdf DLA：2 月 https://www.mouser.de/datasheet/2/783/BST_BNO055_DS000-1509603.pdf DLA：February07, 2023



(a) 左侧　　(b) 右侧

图3：(a) **左侧** 集成多模态传感器融合模型概览；在预测阶段进行融合。(b) **右侧** 混合多模态传感器融合模型概览；在隐藏层内以及预测前进行融合。

混合融合神经网络。各块的详细信息见第3.2.2节。

3.2.1. 多模态集成晚期传感器融合

我们采用基于传感器的晚期融合 [69]，如图3a **左侧**和图4所示。数据按传感器类型划分，并在图4中实现了四个依赖传感器类型的神经网络（NN）模型。这种方法使每个神经网络都能学习各个模态的独特属性，并便于采用简单的融合方法。在初始步骤中，依赖传感器类型的神经网络输入以特定于传感器的方式进行预处理，具体如下所述。每个动作的数据均作为一个完整实例进行处理。神经网络采用 TensorFlow 框架（2.9.2版）开发。训练中采用提前停止，耐心值设为30，并启用恢复权重选项，以避免过拟合；训练共进行500轮。在个体模型中，针对每位参与者手动调整学习率；而在一个适用于所有参与者的集成模型（跨用户情形）中，学习率设为0.03。采用分类交叉熵损失函数和 Adagrad优化器来优化依赖传感器类型的神经网络。集成神经网络由一个包含20个单元的全连接层、Adam优化器（0.01）以及具有九个概率输出的 softmax 函数组成，见图3a**左侧模型集成**。依赖传感器类型的神经网络的详细信息将在下文说明。

压力机械肌动图（PMMG）和惯性测量单元：我们结合使用 FSR 和 PEF 来感知 PMMG。对于惯性传感，选用四元数表示方向，以避免万向节锁定问题[70] 并提高稳定性。FSR、PEF 以及惯性传感器（四元数和加速度）的数据均通过减去手势首个值（起始点）和末个值（结束点）的平均值进行归一化。由于每个面部事件都是长度可变的时间序列，因此采用动态重采样过程将其重采样为 400 个采样[43, 44]。随后，将重采样后的信号输入截止频率为 5Hz 的一阶巴特沃斯低通滤波器，以去除重采样过程在信号边缘产生的振铃峰值，并突出低频范围。

FSR 和 PEF 信号被视为一对，因此共用一个神经网络。方向和加速度则分别由独立的神经网络处理。总共训练了三个网络，用于 PMMG 和惯性传感。神经网络结构基于经过修改的 1D-LeNet5 网络 [71]，如图4 **模型肌动描记**所示。该网络由卷积（conv）—最大池化（maxpool）—卷积—最大池化—卷积—全连接（fc）—全连接—Softmax 层组成，并在卷积层中采用批归一化和随机失活。卷积层包含 40 个滤波器，卷积核大小为 10，激活函数为 ReLU。对于最大池化，第一次卷积（400，40）对应的池化大小为（40，40），第二次卷积（40，40）对应的池化大小为（4，40）。第三次卷积的大小为（4，40），不进行池化。经过一个大小为 160 的展平层后，接入一个包含 100 个神经元的全连接层。随后，全连接层和 Softmax 函数将图1中对应不同面部肌肉活动的九个输出转换为概率。图4 **顶部模型肌动描记**展示了 PMMG 和惯性测量单元所用传感器相关神经网络结构的详细视图。

声学机械肌动图（AMMG）： 如图 2 所示，将两只以44.1千赫兹 采样的 I2S 麦克风安装在运动帽上，以覆盖志愿者的额肌。按照文献 [2] 提出的方法，将音频信息用作 AMMG，但未使用听诊器来放大音频。两个音频通道均重新采样至 52000（1.17 秒）。肌肉音频数据被转

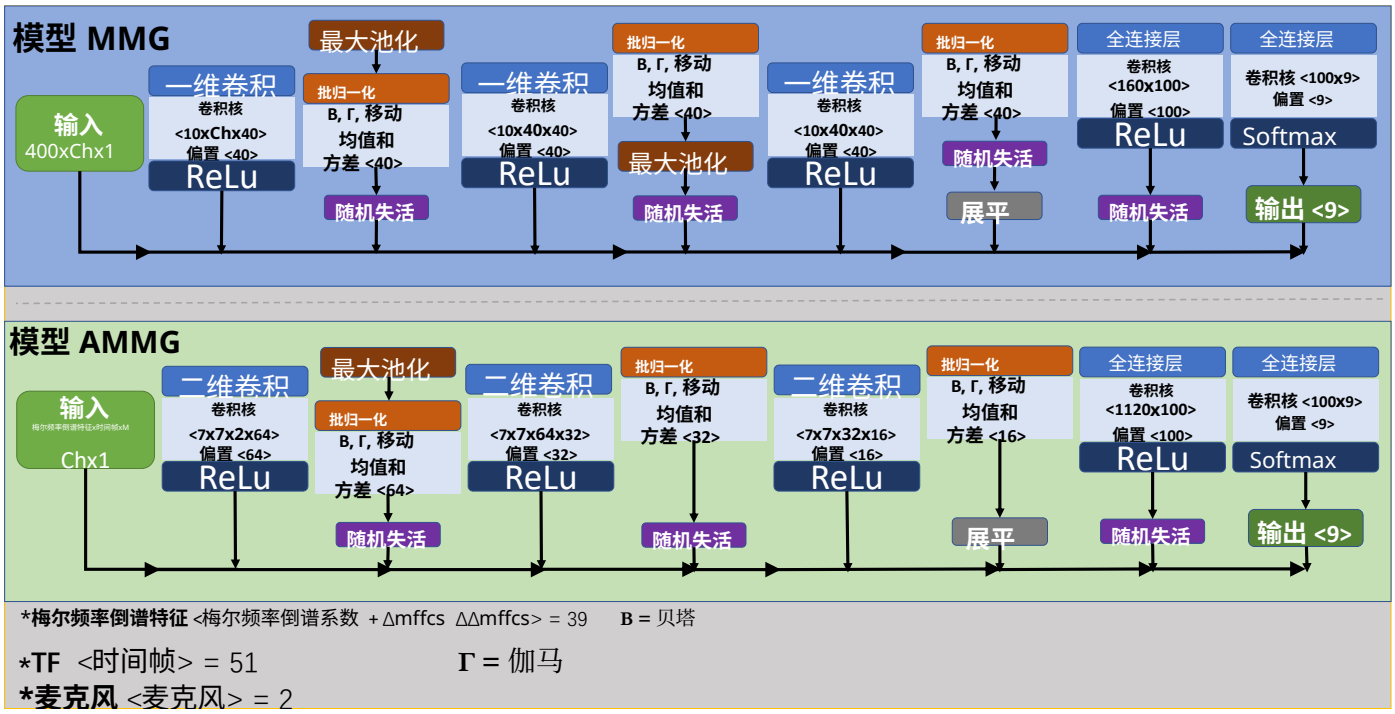


图4：传感器依赖型神经网络模型。**（模型肌动描记顶部）**用于处理 FSR-PEF（PMMG）和惯性测量单元（方向与加速度）信息的神经网络。**（底部模型 AMMG）**用于处理音频梅尔频率倒谱系数的神经网络（AMMG）。

换为梅尔频谱，以降低音频维度，并加快小型数据集上神经网络的收敛速度。梅尔频率倒谱系数（梅尔频率倒谱系数）已用于许多应用 [72, 73]。在音频分析文献中，常用 13 个梅尔滤波器 [74, 73]。短时傅里叶变换（STFT）参数设置如下：采样率为 44.1千赫兹，使用汉明窗，窗长为 4096 （ 93 毫秒），帧移为 1024 （ 23 毫秒）。虽然我们未使用语音信息，为简化起见，仍选择与 [75] 中定义的默认配置（语音分析的典型设置）一致的 STFT 配置。除梅尔频率倒谱系数外，我们还计算了其一阶和二阶导数，以提高识别准确率 [76]。神经网络的输入形状为（ 39,51,2）；包含 13 个梅尔频率倒谱系数、13 个梅尔频率倒谱系数的一阶差分和 13 个二阶差分、51 个时间帧以及两个音频通道。深度学习模型由二维卷积（conv2D）-最大池化-conv2D-conv2D-fc-fc-Softmax 层构成，并在卷积层中采用批归一化和随机失活。如图 4 所示，第一、第二和第三个 conv2D 分别包含 64、32 和 16 个滤波器，卷积核大小为 7，激活层为 ReLU。**底部模型 AMMG**。

在未来实现实时运行的情况下，模块化方法能够灵活地将各神经网络模型的计算任务分配到多个微控制器（MCU）之间。本研究开发的神经网络（NN）小于 2 MB，适合嵌入式设备使用（其闪存通常为 4 至 16 MB），并且借助 TensorFlowLite 框架——work[4]，可以压缩后用于移动设备/嵌入式设备。与大小为 162 MB 的可穿戴摄像头方案相比， [35]，我们的 2 MB 分布式神经网络小了两个数量级。因此，我们可以将 FSR-PEF 和 IMU 模型（图4**模型肌动描记**）部署到微控制器一中，并将音频模型（图4**模型 AMMG**）部署到微控制器二中，最终由支持蓝牙的手机利用集成模型将其合并。由于这些模型轻量且可分别部署在微控制器中，嵌入式设备只会发送预测结果，从而保护用户数据的隐私。

3.2.2. 多模态混合融合

早期/混合融合方法将所有传感模态对识别性能的贡献作为一个整体进行学习，并利用信息源之间的互相关。另一方面，早期/混合多模态传感器融合所采用的神经网络结构可能较为复杂，难以获得具有竞争力的结果，从而增加计算时间 [69, 68]。在某些情况下，融合低级特征可能与任务无关，进而降低融合能力 [77]。因此，进行合并的层级将影响所融合信息的质量。此外，这种融合方法要求所有传感器的数据进入同一处理流水线，从而降低并行性，并增加在嵌入式设备上的部署复杂度。挑战在于保持

[4]移动和边缘设备机器学习 - TensorFlow Lite https://www.tensorflow.org/lite

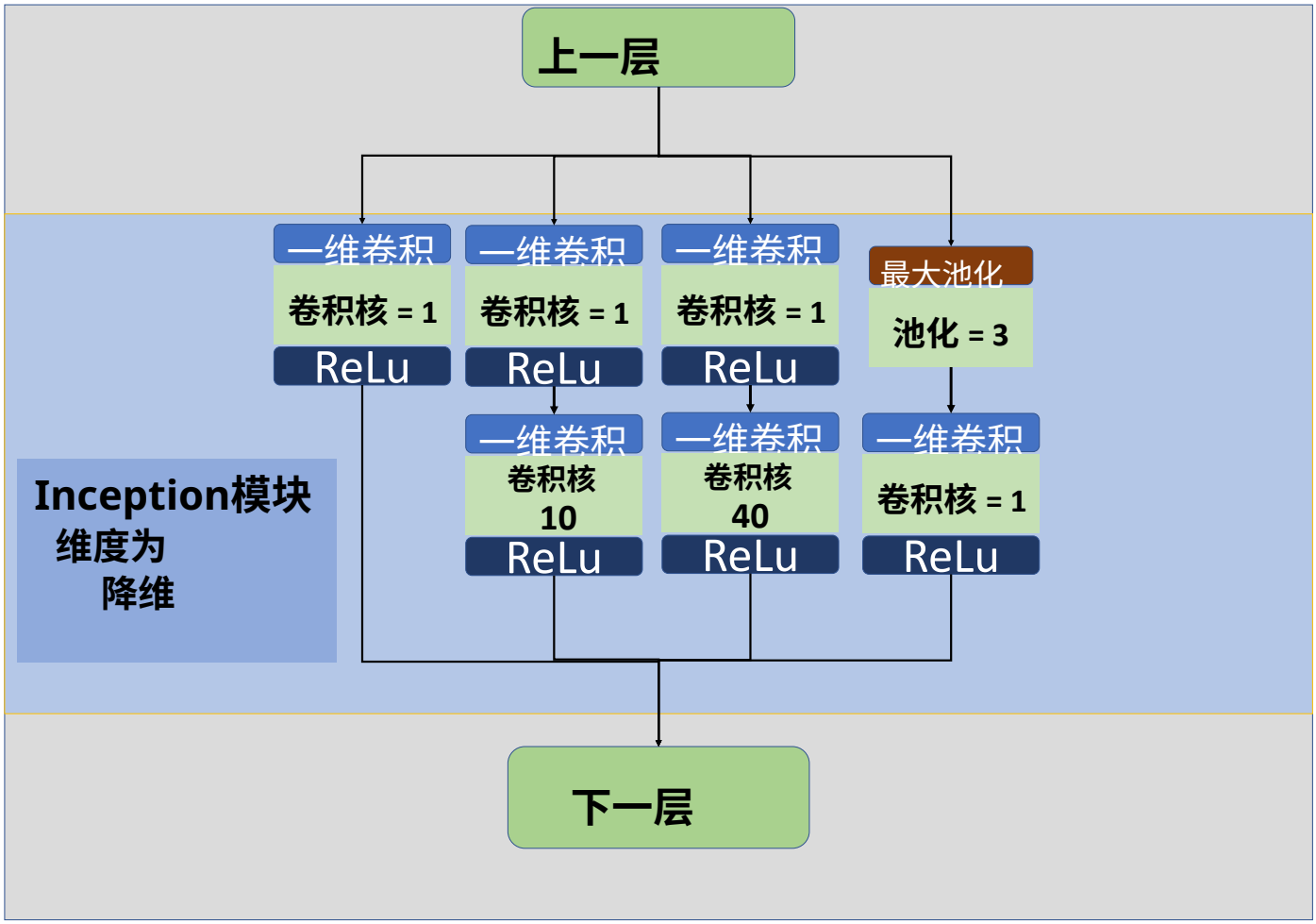


图5：采用降维和按传感模态局部连接的改进型 Inception模块

低复杂度模型的同时，实现与后期融合方法相当的性能。在本节中，我们评估这种传感器融合方法的性能，以测试我们系统识别面部肌肉运动的能力。

我们的实验基于模仿与典型面部表情相关的面部运动。因此，志愿者（没有表演经验）的模仿能力可能会对真实标签产生负面影响。本节中，我们选取了八位模仿能力最强的参与者，以测试系统的可行性，避免参与者模仿不准确所造成的离群影响。筛选依据是对志愿者模仿面部表情词典的表现进行视觉评估，并由大多数参与者判断其相似度。减少志愿者人数的主要缺点是会增加过拟合风险；对于我们的特定任务而言，由于混合融合建模比后期融合复杂得多，这一风险本来就很高。因此，我们采用了旨在降低过拟合风险的技术，包括：耐心值设为 30 并启用恢复权重选项的提前停止、批归一化和最大池化。评估方案采用留一会话法，且神经网络结构基于轻量级神经网络，旨在减少参数数量。我们重点采用一种最受欢迎的轻量级神经网络的主块作为主干网络 {v1}，并将其局部连接到每种传感模态；该模块称为 Inception模块，如图5所示。这些主块专门设计用于减少参数数量，同时兼顾延迟与准确率，从而能够部署在移动/嵌入式系统中。此外，损失函数采用软标签（也称为标签平滑），将 10% 的权重平均分配给相反类别。硬标签是为类别成员分配标签的常见方式，此时类别归属为二元值，即标签为真/假。在软标签中，类别归属基于分配给每个成员的概率分数。因此，10% 的软标签意味着，例如，当类别 1 为真实标签时，类别 1 的概率为 90%，其余 10 % 则平均分配给相反类别。

留出一个会话，神经网络结构基于一种旨在减少参数数量的轻量级神经网络。我们重点采用一种最受欢迎的轻量级神经网络的主块作为主干网络，并将其局部连接到各个传感模态 [78]，该模块称为 Inception 模块，如图5所示。这些主块经过专门设计，可在减少参数数量的同时，平衡延迟与准确度，从而能够部署在移动/嵌入式系统中。此外，损失函数采用软标签（也称为标签平滑），将 10% 的概率均匀分配给相反类别。硬标签是为类别成员分配标签的常见方式，其中类别归属是二元的，即标签为真/假。在软标签中，类别归属基于分配给每个成员的概率分数。因此，10% 的软标签意味着，例如，当类别1为真实值时，类别1的概率为90%，其余10%则均匀分配给相反类别。

为了降低混合模型结构的复杂度，我们根据类别之间的混淆风险减少了类别数量。由于视觉感知相似，喜悦和惊讶这两个类别被合并。

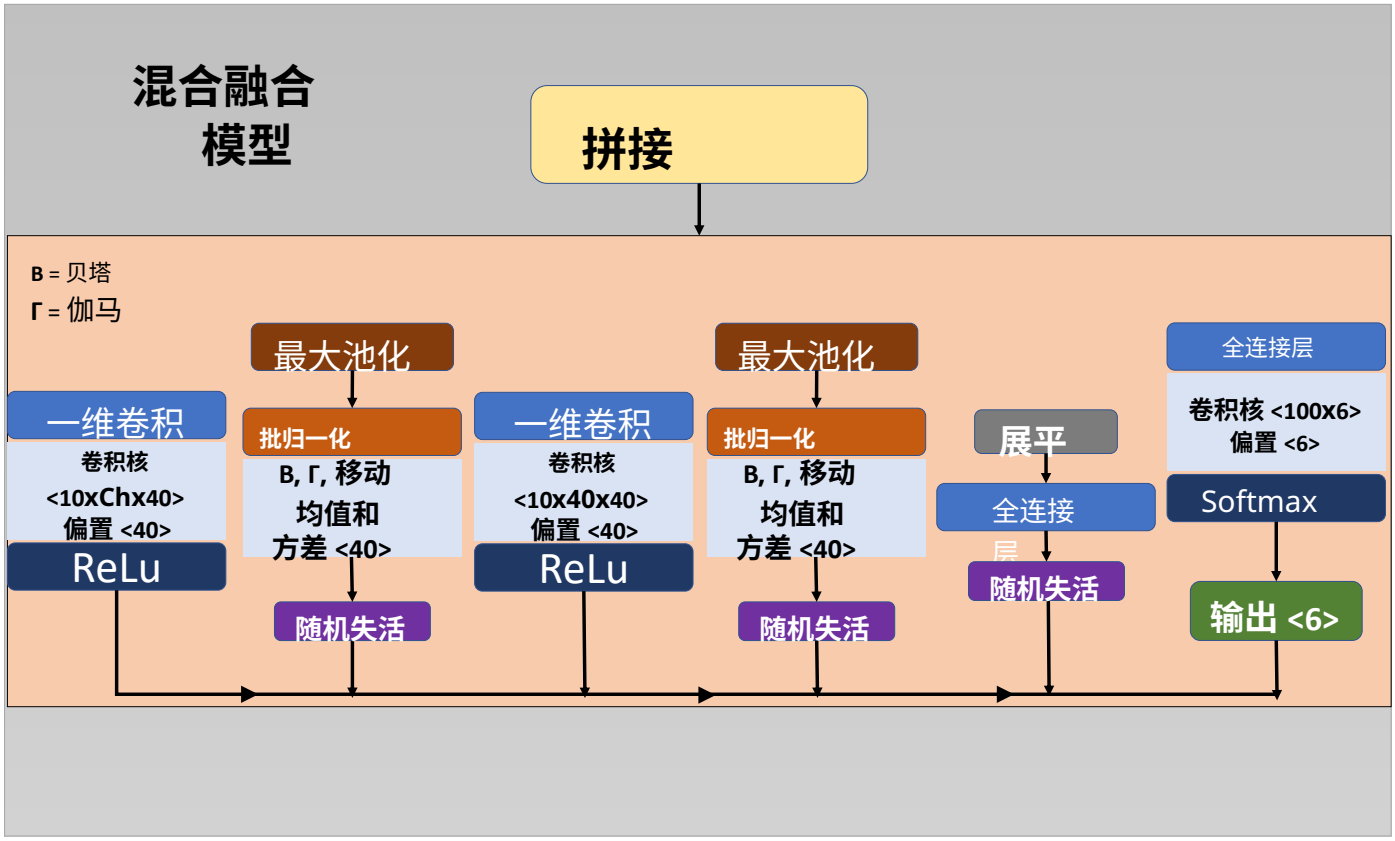


图6：多模态混合融合模型

在面部动作中，这两个类别仅在嘴部运动上有所不同，而传感器安装在前额上，使得对它们的识别只能间接进行，因此两者之间可能会产生相当大的混淆。此外，后续第4节表明，当组合所有传感模态时，后期融合方法中的混淆主要集中在两对类别上，即“厌恶-愤怒”和“悲伤-愤怒”，因此我们将这些类别合并。识别的类别包括中性、喜悦 + 惊讶、厌恶 + 悲伤 + 厌恶、眨眼、恐惧和服药，总共六个类别。尽管采用了所有参数和复杂度降维技术，我们的混合融合模型仍有 939,706 个参数（为后期融合参数的 2.35 倍）。不过，按参数数量来看，与通常认为较小的网络相比，这仍是一个小型网络，例如拥有 350 万个参数的 MobileNetV2 [79]。

如图3b所示，我们采用了与每种传感模态局部连接的 Inception 块，随后进行拼接。然后将输出输入混合融合模型/神经网络。在进入神经网络之前，针对每种传感器类型执行与第3.2.1节相同的信号处理步骤。由于通道数/分辨率以及采样频率不同，我们的数据源在时空上并未对齐。因此，使用卷积神经网络拼接这些数据之前，需要进行一定程度的预处理。因此，所有传感器的数据形式为（时间步 = 400，通道）。对于 PEF 和 FSR、方向、线性加速度和音频，通道数分别为四、四、三和二。局部连接的 Inception 模块结构如图5所示。随后，Inception模块连接混合融合模型，如图6所示。该混合模型由一维卷积-最大池化-一维卷积-最大池化-展平-全连接层（fc）-（fc）-Softmax层组成，并在卷积层中使用批归一化和随机失活。卷积层包含 40 个滤波器，卷积核大小为 10，激活函数为 ReLU。对于最大池化，第一次卷积（400，40）对应的池化大小为（40，40），第二次卷积（40，40）对应的池化大小为（4，40）。160 个单元的展平层后接一个包含 100 个单元的全连接层。随后，通过全连接层和 Softmax函数，将六个类别（中性、喜悦+惊讶、愤怒 +厌恶 +悲伤、眨眼、恐惧和服药）的六个输出转换为概率。训练进行了 200 个训练轮次，并启用了提前停止，耐心值设为 30，同时启用了恢复权重选项。使用分类交叉熵损失函数和学习率为 0.09 的 AdaDelta优化器来优化传感器类型神经网络。AdaDelta优化器是一种针对每个维度采用自适应学习率的方法，其主要优点是不需要设定全局学习率。此外，它还能处理训练过程中学习率持续衰减的固有情况 [80]。

图8 **右侧**展示了用于构建混合神经网络结构和进行训练的各项技术的概览图。

4. 结果与讨论

4.1. 实验流程

来自不同背景（德国、意大利、秘鲁、印度、法国、中国、韩国和委内瑞拉）的13名参与者佩戴我们的运动帽原型，并在每次会话中按照随机顺序模仿图1中定义的面部肌肉运动。图1中的词典包含[1] 提出的七种面部表情，以及来自 [2]的另外两种手势。需要强调的是，与以往硬件相关的研究（见表 1）一样，为了测试融合传感模态的可行性，该系统识别的是有意识模拟的面部表情，而非真实的面部表情。因此，我们强调，系统识别的是与形成类似于真实面部表情特征相关的“面部肌肉运动”。

参与者的肌肉运动声音和压力模式均在未施加任何额外条件的情况下记录，唯一要求是尽可能准确地模仿词典中的动作。受试者无需发出声音，也不受完成面部肌肉运动的特定时限限制。因此，即使是同一参与者，每个手势所需的时间也各不相同。所有参与者均完成了五次会话。每次会话中，词典里的每种面部手势都会以随机打乱的顺序出现四次，以避免肌肉疲劳，并避免同一手势的不同实例之间产生相关性。每位志愿者共采集了180个实例。中性面部表情标记每个手势的起点和终点。总共采集了2160个有效的面部运动。

机械肌动图为实验提供了时间安排要求，以避免参与者因疲劳而导致数据失真 [58]。因此，每次会话持续五至七分钟。平均而言，受试者在会话之间至少休息 10 分钟（期间不佩戴运动帽）。部分志愿者在两天内完成了实验。志愿者包括六名年龄为 21 至 30 岁的女性和七名年龄为 21 至 35 岁的男性（均值 27.00 ± 4.11），头部直径为 52 至 61 厘米（均值 55.54 ± 2.56），发型各异（直发、卷发和无发）。实验在办公室内进行，参与者在实验期间保持坐姿。所有参与者均签署了协议，遵循该大学保护人体受试者委员会的政策以及《赫尔辛基宣言》 [81]。实验过程进行了录像，以供后续保密分析。观察者和参与者遵循符合实验当天强制性公共卫生指南的伦理／卫生规程。有效会话的最低数量为 4 次（志愿者 3、4、8 和 10），典型情况为五次有效会话。训练／测试方案采用五折分层交叉验证，并使用留一会话法 [82], ，与 [23] 中的跨会话验证类似。留一会话交叉验证增强了训练过程对于系统重复佩戴的鲁棒性，这也是可穿戴设备的常见使用情形。

4.2. 结果

图7 **A** 展示了单独模型（按用户划分）与集成模型的性能对比。与最佳依赖传感器的模型相比，集成模型的平均提升幅度为14.30%（F1分数）。图7**B** 中的混淆矩阵显示，单独的拼接模型平均F1分数为85%。针对所有用户的集成模型（跨用户，见图7**C**）的F1分数为79.00%。此外，针对13位参与者的单一集成模型，与最佳依赖传感器类型的神经网络相比，F1分数提升了16.00%（见图7 **C**）。

具体而言，针对13位参与者的依赖传感器的神经网络，其F1值如下：FSR+PEF =51.00%、方向 $= 58.00\%$、加速度 $= 63.00\%$以及麦克风 $= 44.00\%$。对于受试者2、3、6、9和11，集成模型使F1结果最高提升了28.00%。在某些情况下，集成模型的结果仅达到参与者（7和8）中最佳传感器类型神经网络（惯性传感）的性能，这表明集成模型在这些特定情况下并未提升性能。

在图7 **C** 中，最明显的错误分类发生在悲伤和愤怒之间，混淆率高达 17.00%。这可能是由于做出悲伤表情时，眉眼动作被误解所致：许多志愿者并没有抬起眉毛，而是把眉毛向下压，且手势强度也有影响。愤怒和厌恶表情之间的混淆率为 13.00%，如图1所示，这两种表情较为相似。另一个值得注意的错误分类（10.00%）发生在惊讶和快乐表情之间，这两种表情的区别在于嘴部运动。中性面部表情（空类）的识别率为 98.00%，表明该设计具有很高的特异性，因此适合将空类识别用于实时自动数据分段。

AMMG传感模态的性能最弱。造成这一结果的原因可能有多种；在自然环境中，面部表情通常会伴随音频模式，而在模拟实验中可能不存在这种音频模式。本研究中的音频分析仅考虑了使用13个滤波器提取的梅尔频率倒谱系数，以加快神经网络的收敛速度，但忽略了整个频谱中的频谱信息。由于数据集较小，未探索更复杂的图像处理技术。未来强烈建议进行数据增强和迁移学习，以充分利用音频信息。

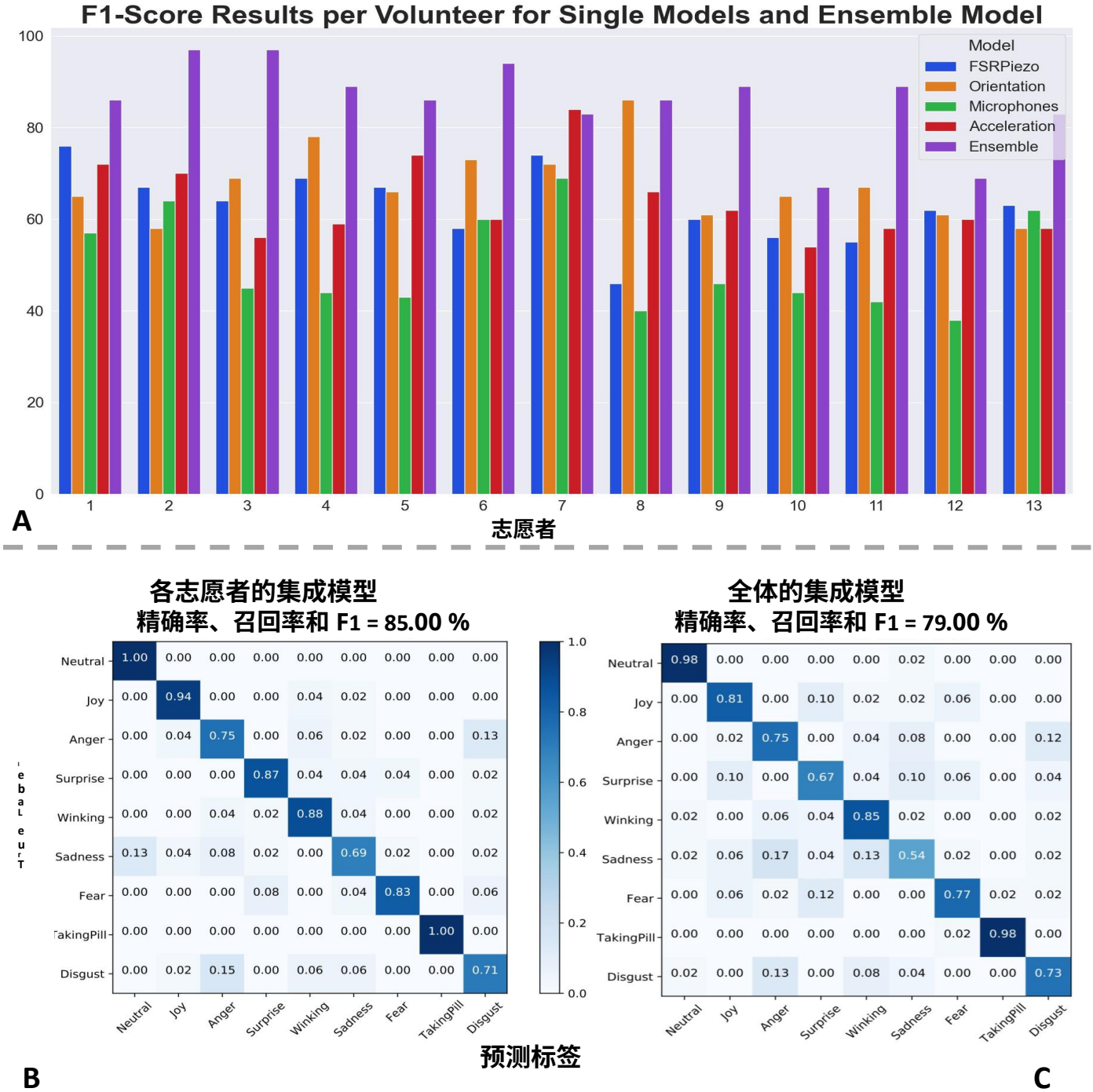


图7：传感器相关神经网络与集成神经网络的结果 (A) 各志愿者传感器相关神经网络的 F1分数对比。(B) 各参与者集成模型的召回率结果，平均值为 F1=85.00%。(C) 所有参与者共用一个集成模型的召回率结果 F1=79.00%

图8 **左侧** 展示了采用留出会话评估方案的混合融合模型性能。结果的F1分数为 = 82 %，这是五次迭代/会话的平均值。主要混淆出现在以下几组类别之间："恐惧-喜悦 + 惊讶"和"眨眼-愤怒 +厌恶 +悲伤"，其混淆率分别为23%和19%。值得注意的是，发生混淆的类别具有相似的眉毛运动。第一组由以上扬眉毛运动为主的表情构成，而下扬眉毛第二类主要由这些动作主导。然而，混淆主要出现在样本较少的类别中。由于合并了类别，模型能够区分互补类别；对于"喜悦 + 惊讶"，召回率超过 80.00 %，对于"愤怒 + 厌恶 + 悲伤"，召回率最高可达 88.00 %。尽管我们的数据集规模较小，F1分数超过 80 % 仍是令人鼓舞的结果。如果向神经网络提供更多数据，结果还可以得到改善。

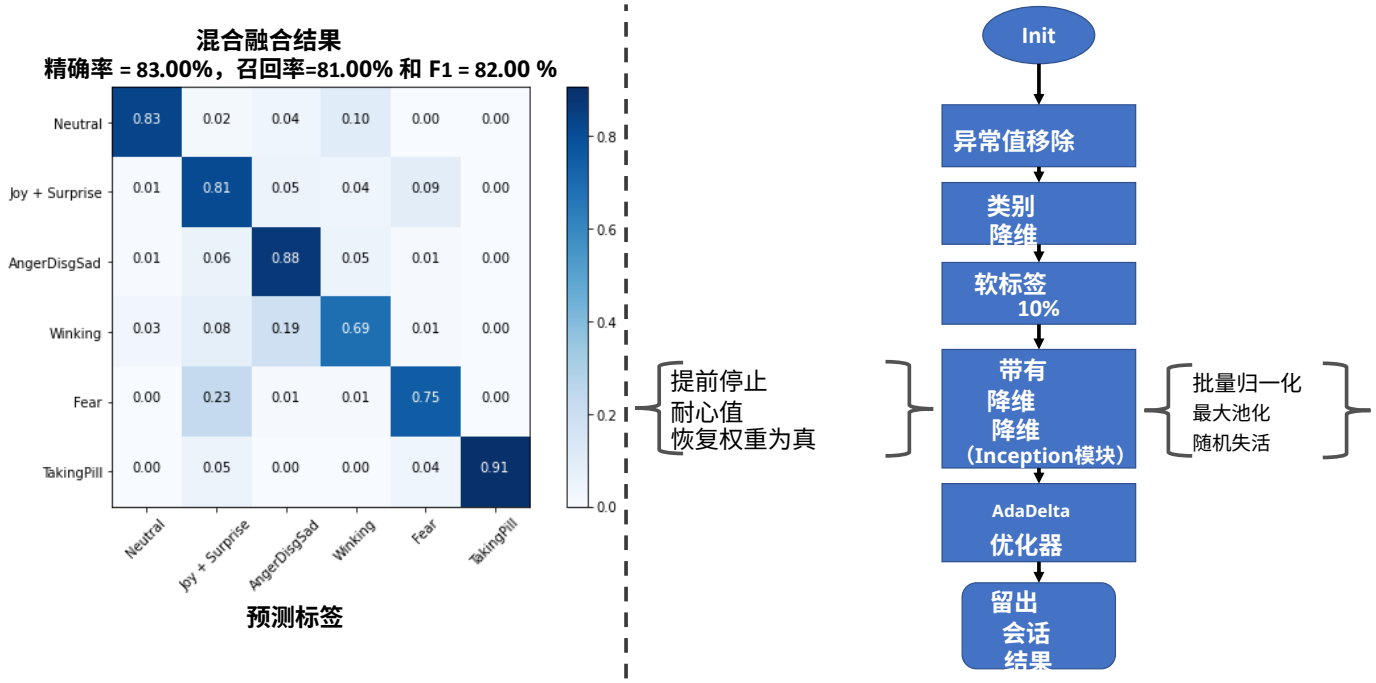


图8：**左侧** 混合融合模型留出会话对八位最佳模仿者的召回率结果 F1 = 82%。**右侧** 混合融合建模流程中所应用的数据分析技术。

混合融合的结果是图8 **右侧**所示数据分析流程的输出。因此，无法直接比较后期融合与混合融合。不过，模块化后期融合复杂度低且具有并行性——每个微控制器上运行一个传感器依赖型神经网络——因此是我们在嵌入式设备中更倾向采用的方案。此外，为了利用传感模态之间的相关性，混合融合所用参数比早期融合多 2.35 倍。未来，如果向神经网络提供更多数据，混合模型的结果或许能够得到改善。值得探索神经网络划分技术，以解决混合融合模型缺乏并行性的问题。在 [83], 中，作者提出利用多个嵌入式／边缘设备的算力来运行大型卷积神经网络。他们提出了一个框架，可自动将卷积神经网络模型划分为多个子模型，并生成代码以便在多个边缘设备（可能是异构设备）上运行这些子模型，同时支持在边缘设备之间及设备内部利用并行性。

4.3. 讨论

需要注意的是，这些结果基于一项面部表情模仿实验，参与者来自不同文化背景且没有表演经验。尽管每位参与者对面部活动的理解各不相同，我们发现，互补的传感模态能够检测到其中的模式。

由于原型的设计，我们的方案存在若干局限。其中之一是，使用运动帽作为可穿戴配件时，脸颊和嘴部的运动只能通过间接方式捕捉，且无法完整描绘面部活动。定制电子电路板的尺寸以及所选电池使系统过于笨重，不适合日常实际使用。要在自然面部表情下的真实场景中检验这一构想，缩小原型电路板的尺寸势在必行。额肌和颞肌上分散布置的传感模态也可能导致 MMG 信号强度出现差异。只有在传感器位置相同的情况下，才能明确判定哪种传感器模态最佳，但至少就我们目前的硬件而言，这会牺牲用户舒适度。一种可行的替代方案是在微型电路原型中将所有传感器叠放（夹层式排列）；这样便可定量比较传感模态与活动类型。

另一方面，基于令人鼓舞的离线结果，下一步可以考虑实现模型的实时运行。不过，在现阶段，系统需要对整个面部手势（起点/终点）进行分段，才能完成识别。未来，测试在线推理这一设想的计划是使用 FSR梯度作为触发信号，进行自动数据分段（不安状态识别），然后再进行推理。

总的来说，这些传感方法和模块化技术可应用于许多不同领域，例如心理学，用于监测学生在教室中的学习过程 [8]。不过，临床医生/认知专家仍需重新设计实验，使其符合心理学要求。总体而言，我们的设计目标是探索该系统检测面部肌肉运动的可行性。

5. 结论与展望

本研究提出了一种保护隐私、内存占用低、功耗低且不显眼的面部表情检测替代系统。尽管该设计使用了麦克风，但麦克风仅用于采集面部运动时肌肉发出的声音，因此语音或环境声音都不会影响隐私要求。我们的系统已证明，使用安装在运动帽上的非摄像头传感器能够检测面部表情。此外，该系统适用于不同性别和不同文化背景的参与者，体现了该方法的包容性和普适性。在个体和跨用户测试中，运动帽的重新佩戴会对结果产生影响；因此，我们的结果能够经受可穿戴设备日常反复佩戴的考验。结果表明，基于所提出传感器的多模态方法非常适合利用不显眼的可穿戴传感器系统识别面部活动。

与基于单一传感模型的方案相比，我们的集成方法将性能提升了 16%。惯性测量单元和 PMMG 是对结果贡献最大的两种传感方法。未来研究中一个值得探讨的问题是，如何在保持低内存占用和保护隐私的设计前提下，从 AMMG 中提取更多信息。具体而言，可以采用数据增强和迁移学习方法，更好地理解 AMMG。此外，还需要考虑头部周围不同的传感器布置方式，以便详细比较各种传感方法。一种很有前景的方法是设计一个微型原型，将压力传感器、压电传感器、机械麦克风和惯性测量单元组成传感器单元堆栈。传感器堆栈将简化所选传感器与待监测肌肉活动类型之间的对应关系。

我们的结果来自离线分析。因此，系统部署可能会面临一些障碍，例如环境声音、功耗、移动场景和用户舒适度。未来，理想的做法是将系统小型化，以便在更多人群和真实（非办公环境）场景中开展实验。采用完整的手势实例数据处理技术，优势在于模型更轻量、速度更快（因为该技术较为简单）。与此同时，这种做法也使设计依赖自动分割技术，以便在实时手势识别时确定面部手势的起点和终点。对于自动手势分割，可以选择以下一种或组合多种方案：利用FSR压力数据梯度作为触发信号，检测用户处于静止状态，然后发出数据采集开始/停止信号，以进行预测。FSR梯度方案也适用于降低功耗。该方案还可以与加权置信系统结合，并将PEF梯度与FSR配合使用；需要注意的是，PEF容易受到所佩戴配件的运动影响，在我们的案例中，该配件是一顶运动帽。FSR-PEF 扰动检测系统可以作为分层流程的第一步：先使用FSR-PEF模型检测空类（13名志愿者的召回率为83.00%），然后启动集成/融合模型。

我们的研究采用了七种经过验证的面部表情（华沙照片集），以便将结果与未来的方案进行比较。尽管许多研究论文采用个性化的面部／头部活动，影响了与相关研究进行公平比较的效果，但我们认为，我们的结果具有与当前先进水平相竞争的实力。要将该系统立即部署到嵌入式设备中，仍面临若干挑战；不过，基于本研究的结果，我们相信这一构想具有潜力。

致谢

本文所述研究得到了德国联邦教育与研究部（BMBF）SocialWear 项目的资助。

参考文献

[1] M. Olszanowski, G. Pochwatko, K. Kuklinski, M. Scibor-Rylski, P. Lewinski, R. K. Ohme，华沙情绪面部表情图片集：面部表情照片的验证研究，《心理学前沿》5 (2015) 1516。[2] H. Bello, B. Zhou, P. Lukowicz，使用可重构差分听诊器麦克风识别面部肌肉活动，《传感器》20 (17) (2020) 4904。[3] R. G. Harper, A. N. Wiens, J. D. Matarazzo，《非言语交流：现状》，John Wiley & Sons，1978。[4] S. St J. Neill，面部表情和姿势对儿童报告的教师非言语交流反应的影响，《英国教育研究杂志》15 (2) (1989) 195–204。[5] K. L. Schmidt, J. F. Cohn，人类面部表情作为适应：面部表情研究中的进化问题，《美国体质人类学杂志：美国体质人类学家协会官方出版物》116 (S33) (2001) 3–24。[6] A. Ghorbandaei Pour, A. Taheri, M. Alemi, A. Meghdari，人机面部表情交互平台：针对自闭症儿童的案例研究，《国际社会机器人学杂志》10 (2) (2018) 179–198。[7] G. Zen, L. Porzi, E. Sangineto, E. Ricci, N. Sebe，学习用于面部表情分析和手势识别的个性化模型，《IEEE多媒体汇刊》18 (4) (2016) 775–788。[8] C. Pabba, P. Kumar，通过面部表情识别监测大型课堂教学中学生参与度的智能系统，《专家系统》39 (1) (2022) e12839。[9] H. K. S. Juboori, L. Kulkarni，利用面部表情视频分析的驾驶员疲劳检测系统，载于：2017年计算、通信、控制与自动化国际会议（ICCUBEA），IEEE，2017年，第1–9页。[10] Z. Huang, R. Li, W. Jin, Z. Song, Y. Zhang, X. Peng, X. Sun，Face2multi-modal：基于面部表情的车载多模态预测器，载于：第十二届汽车用户界面与交互式车载应用国际会议，2020年，第30–33页。[11] H. Zhang, W. Su, Z. Wang，用于面部表情识别的弱监督局部-全局注意力网络，IEEE Access 8 (2020) 37976–37987。[12] T. Shioiri, T. Someya, D. Helmeste, S. W. Tang，面部表情的误解：一项跨文化研究，《精神病学与临床神经科学》53 (1) (1999) 45–50。

[13] K. Ito, T. Masuda, K. Hioki，情境中的情感信息与面部表情判断：北美人与东亚人之间情境效应的文化相似性与差异，《跨文化心理学杂志》43 (3) (2012) 429–445。[14] J. Swenson, F. L. Casmir，文化同质性、性别、出国旅行和学术背景对解读他人情绪面部表情能力的影响，《传播季刊》46 (2) (1998) 214–230。[15] C. Manresa-Yee, S. Ramis Guarinos, J. M. Buades Rubio，面部表情识别：性别对公平性与表情的影响，载于：第二十二届人机交互国际会议，2022年，第1–8页。[16] S. Hickson, N. Dufour, A. Sud, V. Kwatra, I. Essa，Eyemotion：使用眼动追踪摄像头对虚拟现实中的面部表情进行分类，载于：2019年IEEE冬季计算机视觉应用会议（WACV），IEEE，2019年，第1626–1635页。[17] J. H. Cheong, S. Brooks, L. J. Chang，FaceSync：使用头戴式摄像头记录面部表情的开源框架，F1000Research 8 (2019)。[18] T. Chen, B. Steeper, K. Alsheikh, S. Tao, F. Guimbretière, C. Zhang, C-Face：通过深度学习耳戴式微型摄像头拍摄的面部轮廓，持续重建面部表情，载于：第33届ACM用户界面软件与技术研讨会论文集，2020年，第112–125页。[19] T. Sun, F. Tasnim, R. T. McIntosh, N. Amiri, D. Solav, M. T. Anbarani, D. Sadat, L. Zhang, Y. Gu, M. A. Karami等，利用柔性压电界面对面部应变进行解码，《自然·生物医学工程》4 (10) (2020) 954–972。[20] S. Gashi, A. Saeed, A. Vicini, E. Di Lascio, S. Santini，利用耳塞式耳机进行头部手势和面部表情识别的层级分类与迁移学习，载于：2021年国际多模态交互会议论文集，2021年，第168–176页。[21] T. Amesaka, H. Watanabe, M. Sugimoto，利用耳道传递函数进行面部表情识别，载于：第23届可穿戴计算机国际研讨会论文集，2019年，第1–9页。[22] K. Masai, K. Kunze, Y. Sugiura, M. Ogata, M. Inami, M. Sugimoto，智能眼镜面部表情识别在面部方向变化、重复性和位置漂移方面的评估，ACM交互式智能系统汇刊（TiiS）7 (4) (2017) 1–23。[23] K. Li, R. Zhang, B. Liang, F. Guimbretière, C. Zhang，EarIO：一种低功耗声学传感耳戴设备，可持续追踪细致的面部运动，ACM交互式、移动、可穿戴与普适技术论文集6 (2) (2022) 1–24。[24] W. Guo, X. Sheng, H. Liu, X. Zhu，机械肌动图辅助的上肢假肢肌电传感：一种混合方法，《IEEE传感器杂志》17 (10) (2017) 3100–3108。[25] C. Orizio, D. Liberati, C. Locatelli, D. De Grandis, A. Veicsteinas，表面机械肌电图反映肌纤维抽搐的总和，《生物力学杂志》29 (4) (1996) 475–481。[26] R. Uwamahoro, K. Sundaraj, I. D. Subramaniam，利用电刺激和机械肌动图评估肌肉活动：系统性综述，《生物医学工程在线》20 (1) (2021) 1–47。[27] I. Talib, K. Sundaraj, C. Lam，针对不同类型肌肉活动的机械肌动图传感器选择，《电信、电子与计算机工程杂志》（JTEC）10 (1-13) (2018) 79–82。[28] B. Zhou, T. Ghose, P. Lukowicz，Expressure：利用额头织物压力机械肌电图检测与情感和认知活动相关的表情，《传感器》20 (3) (2020) 730。[29] K. Masai, K. Kunze, D. Sakamoto, Y. Sugiura, M. Sugimoto，Face Commands：面向智能眼镜的用户定义面部手势，载于：2020年IEEE混合与增强现实国际研讨会（ISMAR），IEEE，2020年，第374–386页。[30] P. Ekman, W. V. Friesen，面部动作编码系统，Environmental Psychology and Nonverbal Behavior（1978）。[31] R. Li、J. Lee、W. Woo、T. Starner，《Kissglass：使用智能眼镜进行问候手势识别》，载于：增强人类国际会议论文集，2020年，第1–5页。[32] D. Verma、S. Bhalla、D. Sahnan、J. Shukla、A. Parnami，《Expressear：利用耳戴设备感知细粒度面部表情》，ACM交互、移动、可穿戴与普适技术论文集 5（3）（2021年）1–28。[33] X. Song、K. Huang、W. Gao，《FaceListener：通过商用耳机上的声学感知识别人类面部表情》，载于：第21届ACM/IEEE传感器网络信息处理国际会议（IPSN），2022年。[34] H. Aoki、A. Ohnishi、N. Isoyama、T. Terada、M. Tsukamoto，《Facerecglasses：利用紧凑型可穿戴摄像头识别自身面部表情的可穿戴系统》，载于：2021年增强人类会议，2021年，第55–65页。[35] T. Chen、Y. Li、S. Tao、H. Lim、M. Sakashita、R. Zhang、F. Guimbretiere、C. Zhang，《Neckface：通过颈戴式可穿戴设备持续追踪完整面部表情》，ACM交互、移动、可穿戴与普适技术论文集 5（2）（2021年）1–31。[36] J. Kwon、J. Ha、D.-H. Kim、J. W. Choi、L. Kim，《利用多通道面部反应，通过眼镜式可穿戴设备进行情绪识别》，IEEE Access 9（2021年）146392–146403。[37] T. Ando、Y. Kubo、B. Shizuki、S. Takahashi，《Canalsense：基于耳道气压感知的面部相关运动识别系统》，载于：第30届ACM用户界面软件与技术年度研讨会论文集，2017年，第679–689页。[38] K. Futami、K. Oyama、K. Murao，《一种利用耳部配饰上的红外距离传感器阵列识别面部手势的方法》，载于：第23届信息集成与网络智能国际会议，2021年，第650–654页。[39] D. J. Matthies、C. Weerasinghe、B. Urban、S. Nanayakkara，《Capglasses：通过智能眼镜实现无系留电容感知》，载于：2021年增强人类会议，2021年，第121–130页。[40] J. Scheirer、R. Fernandez、R. W. Picard，《Expression glasses：用于面部表情识别的可穿戴设备》，载于：CHI'99计算系统中的人为因素扩展摘要，1999年，第262–263页。[41] M. Perusquía-Hernández、S. Ayabe-Kanamura、K. Suzuki、S. Kumano，《面部肌电图的隐含潜力：区分摆拍微笑与自发微笑时肌电图和计算机视觉的比较》，载于：2019年CHI计算系统中的人为因素会议论文集，2019年，第1–9页。[42] W. Xie、Q. Zhang、J. Zhang，《面向智能眼镜的基于声学的上部面部动作识别》，ACM交互、移动、可穿戴与普适技术论文集 5（2）（2021年）1–28。[43] H. Bello、B. Zhou、S. Suh、P. Lukowicz，《Mocapaci：利用作为电容天线的纺织电缆，在宽松服装中检测姿势与手势》，载于：2021年可穿戴计算国际研讨会，2021年，第78–83页。[44] H. Bello、B. Zhou、S. Suh、L. A. Sanchez Marin、P. Lukowicz，《随特雷门琴而动：利用特雷门琴和嵌入式纺织电缆天线，在宽松服装中识别身体姿势与手势》，计算机科学前沿 4（2022年）。[45] D. J. Matthies、B. A. Strecker、B. Urban，《Earfieldsensing：一种新型入耳式电场感知技术，通过面部表情丰富可穿戴手势输入》，载于：2017年CHI计算系统中的人为因素会议论文集，2017年，第1911–1922页。[46] A. Swerdlow等，《接触超声波和次声波的健康影响》（2016年）。[47] R. J. Lubner、N. S. Kondamuri、R. M. Knoll、B. K. Ward、P. D. Littlefield、D. Rodgers、K. G. Abdullah、A. K. Remenschneider、E. D. Kozin，《常规人类听觉范围之外接触声学能和电磁能后出现的听觉前庭症状综述》，神经学前沿 11（2020年）234。[48] D. B. Moyano、D. A. Paraiso、R. A. González-Lezcano，《可能的

超声波接触对健康的影响、工作环境中的风险因素及职业安全综述》，载于：Healthcare，第10卷，MDPI，2022年，第423页。[49] K. K. Kesari、A. Agarwal、R. Henkel，《辐射与男性生育能力》，生殖生物学与内分泌学 16（1）（2018年）1–16。[50] D.-K. Li、H. Chen、J. R. Ferber、R. Odouli、C. Quesenberry，《接触磁场非电离辐射与流产风险：一项前瞻性队列研究》，《科学报告》7（1）（2017年）1–7。[51] N. Sezer、M. Koç，《压电能量采集技术现状综合综述》，《纳米能源》80（2021年）105567。[52]R. B. Woodward、M. J. Stokes、S. J. Shefelbine、R. Vaidyanathan，《利用惯性数据分割机械肌动图测量中的肌肉活动阶段》，《科学报告》9（1）（2019年）1–10。[53] K. A. Lamkin-Kennard、M. B. Popovic，《传感器：天然与合成传感器》，《生物机电一体化》（2019年）81–107。[54] I. Talib、K. Sundaraj、C. K. Lam、S. Sundaraj，《利用机械肌动图评估肱二头肌活动的系统综述》，《肌肉骨骼与神经元相互作用杂志》18（4）（2018年）446。[55] I. Shin、S. Ahn、E. Choi、J. Ryu、S. Park、J. Son、Y. Kim，《基于机械肌动图的股四头肌疲劳分析》，《国际精密工程与制造期刊》17（4）（2016年）473–478。[56] L. V. Schaefer、N. Löffler、J. Klein、F. N. Bittmann，《与对照组相比，机械肌动图和加速度显示帕金森病患者在无震颤情况下执行单侧运动任务时存在肢间不对称》，《科学报告》11（1）（2021年）1–15。[57]R. B. Woodward、S. J. Shefelbine、R. Vaidyanathan，《运动与肌肉激活的普遍监测：惯性数据与肌音图融合》，《IEEE/ASME机电一体化汇刊》22（5）（2017年）2022–2033。[58] Z. F. Yang、D. K. Kumar、S. P. Arjunan，《用于识别肌肉活动与疲劳的机械肌动图》，载于：2009年IEEE生物医学工程学会年度国际会议，IEEE，2009年，第408–411页。[59] A. Posatskiy、T. Chau，《运动伪影对机械肌动图的影响：麦克风与加速度计的比较研究》，《肌电图与运动机能学杂志》22（2）（2012年）320–324。doi：https://doi.org/10.1016/j.jelekin.2011.09.004。URL：
h
t
t
ps://
www.sciencedirect.com/science/article/pii/S1050641111001441[60] M. Napoli，《Flutter简介与入门》，2019年，第1–23页。doi：10.1002/9781119550860.ch1。[61] X. Shen、Q. Wu、K. Zhao、X. Fu，《电生理学证据揭示微表情与宏表情识别之间的差异》，《心理学前沿》7（2016年）1346。[62] J. Zhao、Z. You，《一种嵌入鞋内的压电能量采集器，用于可穿戴传感器》，《传感器》14（7）（2014年）12497–12510。[63] B. Wang、Z. Long、Y. Hong、Q. Pan、W. Lin、Z. Yang，《仿啄木鸟双层带式压电阵列能量采集器，为腕戴式可穿戴设备供电》，《纳米能源》89（2021年）106385。[64] S. V. Fernandez、F. Cai、S. Chen、E. Suh、J. Tiepelt、R. McIntosh、C. Marcus、D. Acosta、D. Mejorado、C. Dagdeviren，《通过创新设计与贴合式结构实现体表压电能量采集器》，《ACS生物材料科学与工程》（2021年）。[65] F. Mokhtari、G. M. Spinks、C. Fay、Z. Cheng、R. Raad、J. Xi、J. Foroughi，《由纳米结构压电纤维制成的可穿戴电子纺织品》，《先进材料技术》5（4）（2020年）1900900。[66] Y. Liu、H. Khanbareh、M. A. Halim、A. Feeney、X. Zhang、H. Heidari、R. Ghannam，《面向自供能上肢可穿戴应用的压电能量采集》，《纳米精选》2（8）（2021年）1459–1479。[67] Y. Wu, Y. Ma, H. Zheng, S. Ramakrishna，柔性与可穿戴电子设备用压电材料：综述，《材料与设计》211（2021年）110164。[68] K. Gadzicki, R. Khamsehashari, C. Zetzsche，多模态卷积神经网络中的早期融合与后期融合，载于：2020年 IEEE 第 23 届国际信息融合会议（FUSION），IEEE，2020年，第 1–6 页。[69] S. Münzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, R. Dürichen，基于卷积神经网络的多模态人类活动识别传感器融合技术，载于：2017 年 ACM 可穿戴计算机国际研讨会论文集，2017年，第 158–165 页。[70] D. S. Brezov, C. D. Mladenova, I. M. Mladenov，关于万向节锁定问题的新视角，载于：AIP 会议论文集，第 1570 卷，美国物理联合会，2013 年，第 367–374 页。[71] Y. LeCun 等，LeNet-5，卷积神经网络，网址：http://yann. lecun. com/exdb/lenet 20 (5) (2015) 14。[72] A. I. Siam, A. A. Elazm, N. A. El-Bahnasawy, G. M. El Banby, A. El-Samie, E. Fathi，利用梅尔频率倒谱系数和神经网络进行基于光电容积描记法的人体识别，《多媒体工具与应用》80 (17) (2021年) 26001–26019。[73] P. M. Chauhan, N. P. Desai，在噪声环境下使用维纳滤波器进行基于梅尔频率倒谱系数（MFCC）的说话人识别，载于：2014 年绿色计算、通信与电气工程国际会议（ICGCCEE），IEEE，2014 年，第 1–5 页。[74] M. R. Hasan, M. Jamil, M. Rahman 等，使用梅尔频率倒谱系数进行说话人识别，《Variations》1 (4) (2004) 565–568。[75] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, O. Nieto，librosa：使用 Python 进行音频与音乐信号分析，载于：第 14 届 Python 科学会议论文集，第 8 卷，2015 年，第 18–25 页。[76] M. A. Hossan, S. Memon, M. A. Gregory，一种新颖的 MFCC 特征提取方法，载于：2010 年第 4 届信号处理与通信系统国际会议，IEEE，2010 年，第 1–5 页。[77] J. Liu, S. Zhang, S. Wang, D. N. Metaxas，用于行人检测的多光谱深度神经网络，arXiv 预印本 arXiv:1611.02644 (2016)。[78] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich，通过卷积走向更深层，载于：IEEE 计算机视觉与模式识别会议论文集，2015 年，第 1–9 页。[79] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen，MobileNetV2：倒置残差与线性瓶颈，载于：IEEE 计算机视觉与模式识别会议论文集，2018年，第 4510–4520 页。[80] M. D. Zeiler，AdaDelta：一种自适应学习率方法，arXiv 预印本 arXiv:1212.5701 (2012年)。[81] D. A. Shephard，1975 年《赫尔辛基宣言》与知情同意。《加拿大医学会杂志》115 (12) (1976) 1191。[82] R. Bey, R. Goussault, F. Grolleau, M. Benchoufi, R. Porcher，用于实现无偏且保护隐私的联邦学习的分折分层交叉验证，《美国医学信息学会杂志》27 (8) (2020年) 1244–1251。[83] X. Guo, A. D. Pimentel, T. Stefanov，在边缘端自动探索并实现分布式卷积神经网络推理，IEEE《物联网期刊》（2023年）。