

基于跨模态生理信号融合的飞行员认知状态检测*

张文泽¹ 王在俊² 蒋宇恒² 杨睿哲²

(1. 中国民用航空飞行学院空中交通管理学院 广汉 618307;

2. 中国民用航空飞行学院民航飞行技术与飞行安全工程技术重点实验室 广汉 618307)

摘要: 飞行员认知状态的精准评估对保障飞行安全至关重要,然而现有方法在多模态生理信号融合上存在局限性。为此,本文提出一种基于双向跨模态注意力的双流深度学习网络。该模型采用并行双分支架构:脑电(EEG)分支通过相位锁定值(PLV)量化大脑功能连接,并采用SE模块增强的密集连接网络进行深度特征提取;心电(ECG)分支则提取心率变异性(HRV)及波形特征,经残差连接多层感知机处理以表征自主神经系统活动。在此基础上,通过创新设计的双向跨模态注意力模块,动态加权融合双路深度特征,实现对注意力集中、分散和惊吓/惊奇3种状态的精准分类。在NASA公开数据集上的实验表明,模型总体识别准确率达97.44%。消融与对比分析证实,该融合策略显著优于单模态分析和简单的特征拼接方法。研究表明,通过注意力机制深度融合EEG功能连接与ECG生理信息,可有效提升认知状态识别性能,该方法为开发客观、高效的飞行员状态监测系统提供了可靠的技术支持,对提升飞行安全具有重要的应用价值。

关键词: 认知状态识别;飞行安全;跨模态注意力;脑电信号;深度学习

中图分类号: TN911.7 **文献标识码:** A **国家标准学科分类代码:** 520.2040

Detection of pilots' cognitive states based on cross-modal physiological signal fusion

Zhang Wenze¹ Wang Zaijun² Jiang Yuheng² Yang Ruizhe²

(1. College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan 618307, China;

2. Key Laboratory of Flight Techniques and Flight Safety, Civil Aviation Flight University of China, Guanghan 618307, China)

Abstract: Accurate assessment of pilot cognitive states is critical for ensuring flight safety, yet existing methods exhibit limitations in fusing multimodal physiological signals. To address this, this paper proposes a dual-stream deep learning network based on bidirectional cross-modal attention. The model adopts a parallel dual-branch architecture: The electroencephalography (EEG) branch quantifies brain functional connectivity through phase locking value (PLV) features and employs a densely connected network enhanced with squeeze-and-excitation (SE) modules for deep feature extraction; the electrocardiogram (ECG) branch extracts heart rate variability (HRV) and waveform features, processed by a residual-connected multilayer perceptron to characterize autonomic nervous system activity. Building upon this, an innovatively designed bidirectional cross-modal attention module dynamically weights and fuses the dual-path deep features to achieve precise classification of three states—concentrated attention, distracted attention, and startle/surprise. Experimental results on the NASA public dataset demonstrate an overall recognition accuracy of 97.44%. Ablation and comparative analyses confirm that the fusion strategy significantly outperforms single-modality analysis and simple feature concatenation methods. The study reveals that deep integration of EEG functional connectivity and ECG physiological information via attention mechanisms effectively enhances cognitive state recognition performance. This approach provides reliable technical support for developing objective and efficient pilot state monitoring systems, holding significant application value for improving flight safety.

Keywords: cognitive state recognition; flight safety; cross-modal attention; electroencephalography; deep learning

0 引言

近年来,人为因素在航空事故中的占比约为60%到

80%^[1]。这些事故的发生往往与飞行员的认知状态密切相关,飞行员认知状态主要涉及以下3个方面:飞行员自身的经验能力、生理状态以及心理状态^[2],而这些因素共同决定

收稿日期:2025-07-16

*基金项目:民航飞行技术与飞行安全重点实验室自主研究项目(FZ2022ZZ02)、四川省民航飞行技术与飞行安全工程技术研究中心项目(GY2024-64E)资助

了人机协同飞行控制系统的效能^[3]。因此,准确评估飞行员的认知状态对于提高飞行安全性和人机协同效率具有重要意义。

为了准确评估飞行员认知状态,传统方法主要依赖主观量表,通过自我报告的方式收集飞行员对任务负荷、注意力分配以及压力水平的主观感受。然而,这种方法存在显著的局限性。一方面,主观量表容易受到个体心理状态、记忆偏差以及社会期望效应的影响,从而导致数据的稳定性和可靠性受到质疑^[4];另一方面,此类评估通常在任务结束后进行,无法实时反映飞行过程中认知状态的动态变化^[5]。

为了克服上述局限性,近年来研究人员开始探索利用生理信号作为客观量化指标来评估飞行员的认知状态。生理信号,如心电(electrocardiogram, ECG)、脑电(electroencephalography, EEG)、肌电(electromyography, EMG)和皮肤电活动(electrodermal activity, EDA)等,已在情绪识别^[6]、疲劳检测^[7]、神经性疾病诊断^[8]、压力检测^[9]等领域展现出广泛的应用潜力。在众多生理信号中,脑电信号和心电信号由于能够分别反映大脑活动和自主神经系统状态,在国内外研究中应用广泛,成为评估飞行员认知状态的核心工具。

EEG作为一种直接测量大脑活动的工具,在飞行员认知状态评估中具有独特优势。Zhao等^[10]在一项关于飞行员学员认知控制的研究中,使用EEG微状态分析技术,发现飞行员学员在面对更高任务需求和不确定性时,其认知控制水平会有所下降,还发现随着训练的进行,飞行员学员在基础任务中的认知控制水平有所提高。Ji等^[11]研究了飞行员在不同飞行阶段的心理工作负荷,他们使用专业飞行模拟器模拟飞行的起飞、巡航、左转、右转和着陆阶段,并记录飞行员在左转和右转任务中的脑电活动。研究发现,与巡航阶段相比,飞行员在左转和右转阶段的心理工作负荷显著增加,表现为beta波能量比和Shannon熵的显著升高。他们还利用支持向量机模型对飞行员的心理工作负荷进行分类,模型在训练集和测试集上的分类准确率分别达到98.92%和93.67%。

与EEG聚焦大脑活动不同,ECG通过心率及心率变异性(heart rate variability, HRV)指标,可以反映飞行员自主神经系统在应激状态下的变化。Shao等^[12]针对飞行员在特殊飞行动作中的应激反应,提出了基于心电信号的评估模型。招募了11名有中级教练机飞行经验的男性飞行员,使用穿戴式心电监测设备采集他们在特殊飞行动作中的心电信号。通过分析心率和心率变异性(HRV)相关指标,发现飞行员在执行特殊飞行动作时,心率会瞬时大幅上升随后下降,心率曲线面积的变化能反映应激强度。利用多尺度算法,构建了应激强度随训练次数增加呈指数衰减并趋于稳定的函数关系。

飞行员的认知状态是一个复杂的多维概念,单一模态

的生理信号,可以在某些特定维度上提供有价值的信息,但难以全面的分析认知状态。因此,近年来双模态及多模态研究在飞行员认知状态评估领域逐渐兴起。Han等^[13]提出了一种基于多模态生理信号(包括EEG、ECG、呼吸和EDA)的飞行员认知状态分类方法。他们在飞行模拟器中诱导飞行员的4种认知状态(分心、工作负荷、疲劳和正常),并设计了一种结合卷积神经网络(convolutional neural network, CNN)和长短期记忆(long short-term memory)模型的MDL架构。实验结果表明,该方法通过融合多模态信号显著提高了分类准确率,平均准确率达到85.2%。Harrivel等^[14]提出了一种基于多模态心理生理信号的实时认知状态预测方法。该研究利用EEG、ECG、GSR和呼吸等信号,在高保真飞行模拟器中采集数据,并通过数据融合和机器学习算法,对飞行员在不同任务状态下的认知控制、注意力分配及工作负荷进行有效分类。结果显示,多模态信号的联合应用显著提高了认知状态的预测准确率。然而,现有方法仍存在两方面的局限性:一是特征融合方式较为简单,多采用直接拼接,未能深入挖掘EEG和ECG信号间的内在关联;二是对EEG信号的处理过于局限,往往仅聚焦于单一维度(如时域、频域或空域)的分析,忽视了跨频段特征间的相互作用。这些不足在一定程度上限制了多模态方法在检测特定认知状态时的表现。

鉴于此,本文在从特征表征与模态融合两个层面进行创新。在特征层面,本文不仅关注传统的时域或频域指标,更深入地探索了基于相位同步的脑电功能连接特性,以捕捉大脑各区域间的动态协作模式。在融合层面,本文构建了一个双流深度网络结构,该结构能够并行处理脑电与心电信号,并通过一个特定的交互模块实现两种信息的深度整合。通过这一系列改进,能更精准地捕捉到飞行员在注意力集中、分散及惊吓等状态下的细微生理变化,从而显著提升识别模型的性能与可靠性。

1 数据集与信号处理

1.1 数据集描述

本文采用了美国NASA公开的飞行员生理数据集^[14],该数据集记录了17名飞行员在注意力集中、注意力分散、惊吓/惊讶3种认知状态下的生理数据。实验中通过B-AlertX24脑电采集设备采集了20导脑电(EEG)信号和单导联心电(ECG)信号,采样频率为256 Hz。采集设备及电极位置如图1所示。

1.2 信号预处理

为了确保脑电和心电信号的质量,为后续特征提取提供可靠基础,本文对原始信号进行了系统化的预处理。

对于EEG信号,首先利用0.5~40 Hz带通滤波器进行滤波,有效去除低频漂移和高频噪声,保留与认知活动相关的关键频段。随后,采用独立成分分析方法分离并剔除眼动、眨眼及肌肉活动等伪迹,确保信号能够真实反映大脑

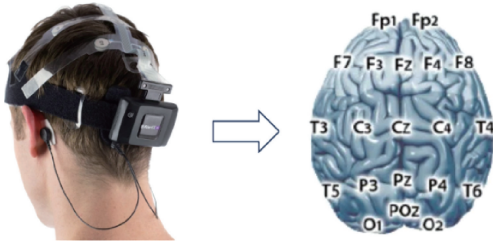


图 1 B-Alert X24 脑电采集设备及 20 通道电极分布

Fig. 1 B-Alert X24 EEG acquisition device and 20-channel electrode distribution

活动。接着,将连续 EEG 信号分割为 1 s 时段,便于后续特征提取和状态分类。

对于 ECG 信号,首先应用 5~15 Hz 带通滤波器滤除基线漂移及 50/60 Hz 工频干扰。随后,采用 Pan-Tompkins 算法精确检测 R 峰并生成 R-R 间隔序列,进而识别并剔除由运动伪迹或设备故障引起的异常 R-R 间隔。在此基础上,为满足短时 HRV 分析需求,将连续 ECG 信号划分为 10 s

非重叠时间窗(epoch),确保每段至少包含 9 个完整心搏周期(以心率 ≥ 54 bpm 为下限)。最终输出标准化 R-R 间期序列集,供多维度特征提取使用。划分后的完整的 epoch 个数如表 1 所示。

表 1 预处理后的生理信号片段

Table 1 Preprocessed physiological signal segments

状态	EEG 片段	ECG 片段
注意力集中	11 810	1 181
注意力分散	1 700	170
惊吓/惊奇	960	96
总片段	14 470	1 447

为协调脑电(EEG)的瞬时性与心电图(ECG)的长时性,本文采用主从式时间窗关联;每个 10 s ECG 窗关联其时间范围内的 10 个连续 1 s EEG 窗,并在训练时,每个 EEG 切片(1 s)共享所属 ECG 窗(10 s)的特征。脑电与心电图信号的完整处理流程如图 2 所示。

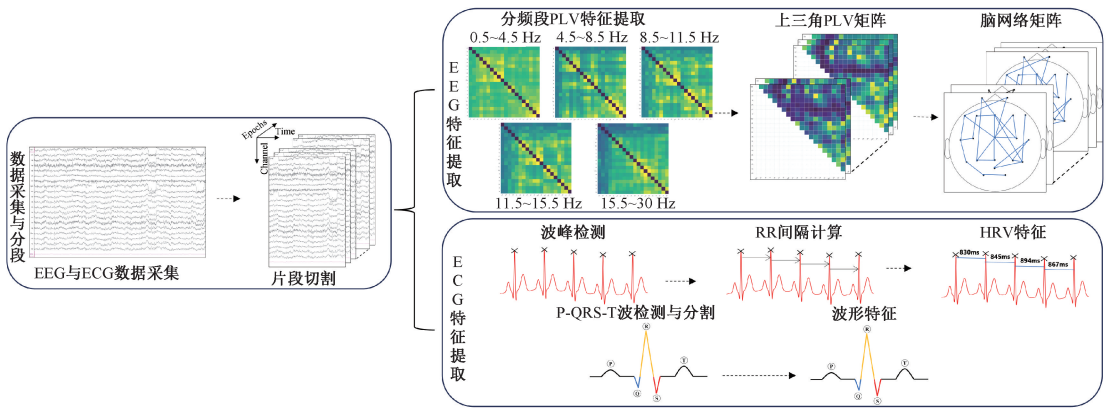


图 2 脑电与心电图信号处理流程图

Fig. 2 Flowchart of EEG and ECG signal processing

1.3 EEG 特征提取

相位锁定值(phase locking value, PLV)是量化两通道脑电信号相位同步性的核心指标,其通过统计特定时间窗内相位差的稳定性,反映神经振荡的功能耦合强度。近年来,PLV 因具有明确的数学物理意义及对非线性信号的强表征能力,被广泛应用于脑功能网络构建^[15]、认知状态解码^[16]等领域。本文 PLV 特征提取流程包括:

首先,基于神经振荡节律的生理意义,采用零相位失真滤波器组将原始 EEG 信号分解为 δ 波(0.5~4.5 Hz)、 θ 波(4.5~8.5 Hz)、 α 波(8.5~11.5 Hz)、低频 β 波(11.5~15.5 Hz)及高频 β/γ 波(15.5~30 Hz) 5 个特征频段,以分离不同认知功能相关的神经振荡成分。随后对各频段信号进行 Hilbert 变换,其数学定义为:

$$H(x(t)) = \frac{1}{\pi} P.V. \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \quad (1)$$

其中, P.V. 表示柯西主值积分。由此构造解析信号:

$$z(t) = x(t) + j \cdot H(x(t)) = A(t)e^{j\phi(t)} \quad (2)$$

其中, $A(t)$ 是瞬时幅度, $\phi(t)$ 是瞬时相位。PLV 值是通过不同通道的相位差来计算的。给定两路信 $\phi_i(t)$ 和 $\phi_j(t)$, 在滑动时间窗 T (本文取 1 s) 内计算相位同步指数:

$$PLV_{ij} = \left| \frac{1}{T} \sum_{t=1}^T e^{j(\phi_i(t) - \phi_j(t))} \right| \in [0, 1] \quad (3)$$

PLV 值的范围在 0~1, 1 表示完全同步, 0 表示完全不同步。最终提取 PLV 矩阵上三角元素, 经 Z-score 标准化消除通道间基线差异, 构建维度统一的相位耦合特征集, 为后续机器学习模型提供高判别性输入。

经特征提取后的每个片段包含的特征数量为 $190 \times 5 = 950$ 个。

1.4 ECG 特征提取

本文针对预处理后的 10 s 心电图片段, 从心跳间隔变化和 ECG 信号形态两个角度出发, 提取了以下 10 个关键的

生理特征,构成用于描述每个 ECG 片段的特征向量。这些特征主要分为两类,即心率变异性特征(HRV features)和波形特征(waveform features)。

1)心率变异性特征

心率变异性特征通过分析 R-R 间期的时间变化,反映自主神经系统(包括交感神经和副交感神经)对心率的调节作用。表 2 为提取的 5 个 HRV 特征:

表 2 心率变异性特征

Table 2 Heart rate variability (HRV) features

特征名称	描述
平均 RR 间期	连续 R 波峰之间的平均时间间隔,反映整体心率。
RR 间期的标准差	测量 RR 间期在整个记录期间的总变异性,反映心率变异性的总体水平。
相邻 RR 间期差值的均方根	反映相邻心跳间隔的短时变化,主要受副交感神经高频调节的影响。
pNN20	相邻 RR 间期差值超过 20 ms 的百分比,评估副交感神经活动。
心搏总数	10 s ECG 片段内检测到的心搏总数,与心率直接相关。

2)波形特征

波形特征通过分析 ECG 信号的形态以及各波段的持续时间,能够准确描述心脏电活动的不同阶段,表 3 为提取的 5 个波形特征。

表 3 波形特征

Table 3 Waveform features

特征名称	描述
QRS 波群平均持续时间	QRS 波群的平均时长,反映心室除极过程所需的时间。
平均校正 QT 间期	使用 Bazett's 公式对 QT 间期进行心率校正,反映心室除极到复极的平均时间。
平均 R 波幅值	R 波的平均高度,反映心室除极的强度。
平均 T 波幅值	T 波的平均幅度,与心室复极过程相关。
平均 P 波持续时间	P 波的平均时长,反映心房除极的时间。

2 跨模态网络设计

为了高效融合飞行员的脑电(EEG)与心电(ECG)生理信号,实现对注意力集中(channelized attention, CA)、注意力分散(diverted attention, DA)及惊吓/惊奇(startle/surprise, SS)3 种核心认知状态的精准检测,本章构建了一个基于双向跨模态注意力的双流深度学习模型。本章将首先介绍模型的整体架构设计,随后将分别详述用于处理脑电与心电信号的两个并行分支:针对脑电 PLV 特征的 SE-DenseNet-EEG 网络和针对心电特征的 RC-MLP-ECG 网络。最后,将重点介绍实现两种模态信息融合与最终分类的跨模态注意力模块。

2.1 模型总体框架

本文提出的飞行员认知状态检测模型总体框架如图 3 所示。

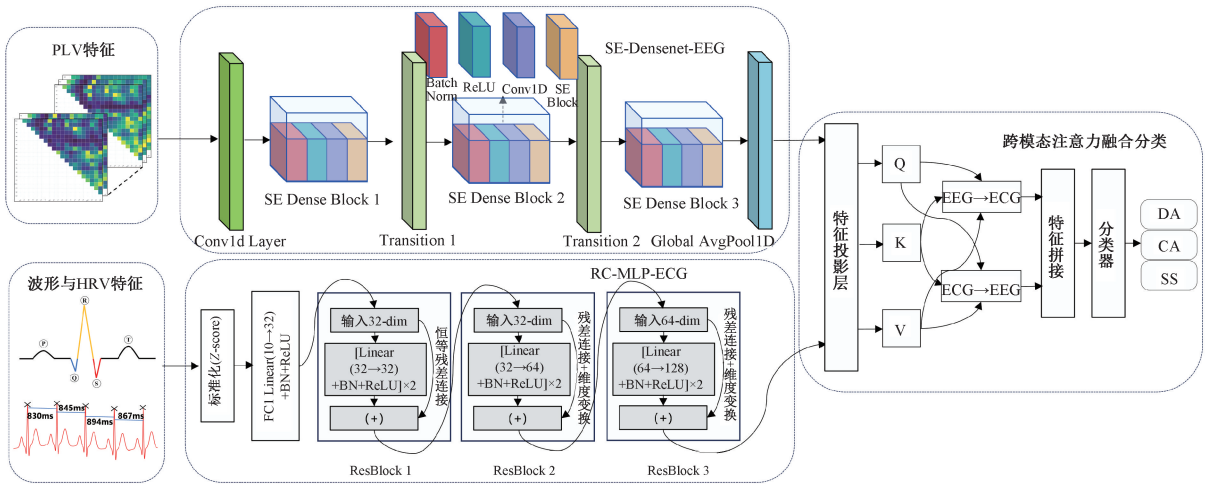


图 3 双流跨模态注意力网络整体架构图

Fig. 3 Overall architecture of the dual-stream cross-modal attention network

该框架遵循“并行提取,融合分类”的设计思想,主要由 3 个核心部分构成:

脑电特征深度提取分支(SE-DenseNet-EEG):该分支

输入从 EEG 信号中提取的相位锁定值(PLV)特征矩阵。通过一个带有挤压激励(squeeze-and-excitation, SE)模块的密集连接网络(densely connected network, DenseNet)^[17]

进行处理,以深度编码大脑功能连接特性,最终输出深度抽象表征 F_{EEG} 。

心电特征深度提取分支(RC-MLP-ECG):该分支输入从 ECG 信号中提取的心率变异性(HRV)与波形特征向量。这些特征通过一个基于残差连接的多层感知机(RC-MLP)进行处理,以反映自主神经系统活动,并输出深度表征 F_{ECG} 。

双向跨模态注意力融合与分类模块:该模块接收来自两个分支的深度特征 F_{EEG} 和 F_{ECG} 。利用一个创新的双向跨模态注意力机制进行信息交互与融合,对两路深度特征进行动态加权,最终通过分类器输出 3 种认知状态的预测结果。

2.2 脑电特征深度提取分支:SE-DenseNet-EEG

脑电信号(EEG)的相位锁定值特征能有效反映脑区之间的功能连接,是认知状态识别的关键依据。然而,原始 PLV 特征存在高维稀疏、跨频段耦合关系未充分挖掘等问题。为此,SE-DenseNet-EEG 采用密集连接(特征复用,缓解高维稀疏)、1D 卷积(沿频段维度滑动,捕捉跨频段耦合)、SE 模块(通道注意力,增强关键脑功能连接)3 大策略,实现 PLV 特征的深度提取。

原始输入为从 20 导联 EEG 信号中提取的 950 维 PLV 特征向量(190 个脑区连接对,即 C_{20}^2)。为适配 1D 卷积操作,将向量重构为三维张量:

$$X_{PLV} \in \mathbf{R}^{L \times F \times C} \quad (4)$$

其中, $L = 190$ (脑区连接对数量,沿“空间维度”排列), $F = 5$ (频段数量,沿“特征维度”排列), $C = 1$ (单通道特征图,用于扩展维度)。

SE-DenseNet-EEG 的核心结构包括初始特征变换层、SE-Dense Block 组、过渡层和输出层,具体设计如下:

1) 初始特征变换层

采用 1D 卷积将输入特征映射到高维空间,捕获脑区连接与频段的初步关联:

$$F_0 = \text{ReLU}(\text{BN}(\text{Conv1D}_{k=7,s=2}(X_{PLV}))) \quad (5)$$

其中,卷积核尺寸 $k = 7$ (覆盖 3 个以上频段,捕获宽频关联),步长 $s = 2$ (压缩空间维度),零填充 $\text{pad} = 3$ (保持空

间尺寸)。输出特征图为 $F_0 \in \mathbf{R}^{95 \times 3 \times 32}$ 。

随后通过最大池化层($\text{MaxPool1D}_{k=3,s=2}$)进一步压缩空间维度,输出 $F_{\text{pool}} \in \mathbf{R}^{48 \times 2 \times 32}$,减少后续计算量。

2) SE-Dense Block 组

采用 3 个串联的 SE-Dense Block (每个 Block 包含 4 个密集连接的卷积单元),通过密集连接(Dense Connection)实现特征复用,通过 SE 模块实现通道级特征校准。每个卷积单元的结构为:

$$\begin{cases} y^{(i)} = \text{ReLU}(\text{BN}(\text{Conv1D}_{k=3,s=1}(x^{(i)}))) \\ z^{(i)} = \text{SE}(y^{(i)}) \\ x^{(i+1)} = [x^{(i)}, x^{(i)}, \dots, z^{(i)}] \end{cases} \quad (6)$$

其中, $x^{(i)}$ 为第 i 层的输入特征(融合了前 i 层的输出), $\text{Conv1D}_{k=3}$ (聚焦相邻频段交互,避免过平滑), $\text{SE}(\cdot)$ 为通道注意力模块,其运算过程为:

$$\begin{cases} s = \sigma(W_2 \delta(W_1 \text{GAP}(z))) \\ z^{\text{weighted}} = s \otimes z \end{cases} \quad (7)$$

式中: GAP (全局平均池化)将每个通道的空间信息聚合为 scalar; $W_1 \in \mathbf{R}^{C \times C/r}$, $W_2 \in \mathbf{R}^{C/r \times C}$ 为全连接层 ($r = 16$ 为压缩比,平衡参数数量与表达能力); δ 为 ReLU 激活函数, σ 为 Sigmoid 函数(生成 $0 \sim 1$ 的通道权重); \otimes 为元素级乘法(对特征图进行通道加权)。

3) 过渡层

位于 SE-Dense Block 之间,用于压缩特征图尺寸与通道数,避免模型过拟合:

$$F_{\text{out}} = \text{AvgPool1D}_{s=2}(\text{Conv1D}_{k=1,s=1}(F_{\text{in}})) \quad (8)$$

其中, $\text{Conv1D}_{k=1}$ (调整通道数,压缩率 $\theta = 0.5$), $\text{AvgPool1D}_{s=2}$ (将空间尺寸减半)。

4) 输出层

通过全局平均池化(GAP)将最后一个 SE-Dense Block 的输出特征图($F^{(L)} \in \mathbf{R}^{12 \times 1 \times 232}$)压缩为固定长度的向量:

$$F_{EEG} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F^{(L)}(i, j) \quad (9)$$

其中, $H = 12$, $W = 1$ 为特征图的空间尺寸,输出 $F_{EEG} \in \mathbf{R}^{232}$ 。

SE-DenseNet-EEG 的参数配置如表 4 所示。

表 4 SE-DenseNet-EEG 的参数配置

Table 4 Parameter configuration of SE-DenseNet-EEG

层级	操作序列	输出尺寸
输入	—	190 × 5 × 1
Conv0	BN-ReLU-Conv1D($k=7, s=2$)	95 × 3 × 32
Pool0	MaxPool1D($k=3, s=2$)	48 × 2 × 32
SE-Dense Block1	[BN-ReLU-Conv1D($k=3$)] × 4 + SE	48 × 2 × 160
Transition1	Conv1D($k=1$) - AvgPool1D($s=2$)	24 × 1 × 80
SE-Dense Block2	[BN-ReLU-Conv1D($k=3$)] × 4 + SE	24 × 1 × 208
Transition2	Conv1D($k=1$) - AvgPool1D($s=2$)	12 × 1 × 104
SE-Dense Block3	[BN-ReLU-Conv1D($k=3$)] × 4 + SE	12 × 1 × 232
输出层	Global AvgPool	232

2.3 心电特征深度提取分支:RC-MLP-ECG模型

心电信号的心率变异性特征和波形特征能反映自主神经系统的状态,但低维特征之间存在复杂的非线性耦合,传统多层感知机难以捕捉这些关系且易出现梯度消失。为此,本文提出残差连接多层感知机(residual-connected MLP, RC-MLP-ECG),其架构如图3所示。

原始输入为从10 s ECG片段中提取的10维心电特征向量,涵盖HRV时域特征与波形形态特征。为消除特征尺度差异,输入特征需进行标准化处理(均值为0,标准差为1),即:

$$\hat{f}_i = \frac{f_i - \mu_i}{\sigma_i} \quad (10)$$

其中, f_i 为原始特征值, μ_i 为特征均值, σ_i 为特征标准差, \hat{f}_i 为标准化后的特征值。标准化后的特征向量为 $\mathbf{x}_{ECG} = [\hat{f}_1, \hat{f}_2, \dots, \hat{f}_{10}]^T \in \mathbf{R}^{10}$ 。

RC-MLP-ECG的核心结构包括输入层与特征映射、残差块组(ResBlock1-3)、输出层,通过残差连接实现跨层级特征融合,具体设计如下:

1) 输入层与特征映射

输入层将10维标准化特征向量映射到高维空间,为后续残差学习提供足够的特征空间:

$$\mathbf{x}^{(0)} = \text{ReLU}(\text{BN}(W_0 \mathbf{x}_{ECG} + b_0)) \quad (11)$$

其中, $W_0 \in \mathbf{R}^{32 \times 10}$ (全连接层权重), $b_0 \in \mathbf{R}^{32}$ (偏置),BN(批归一化)消除特征尺度差异,ReLU激活函数引入非线性。输出特征向量为 $\mathbf{x}^{(0)} \in \mathbf{R}^{32}$ 。

2) 残差块组:跨层级特征融合

采用3个串联的残差块(ResBlock1、ResBlock2、ResBlock3),每个残差块包含两个全连接层与一个残差连接,实现底层特征(即时生理状态)与高层特征(关联信息)的融合。每个残差块的数学定义为:

$$\begin{cases} \mathbf{y}^{(l)} = \text{ReLU}(\text{BN}(W_1^{(l)} \mathbf{x}^{(l)} + b_1^{(l)})) & (\text{第1层:特征变换}) \\ \mathbf{z}^{(l)} = \text{BN}(W_2^{(l)} \mathbf{y}^{(l)} + b_2^{(l)}) & (\text{第2层:特征抽象}) \\ \mathbf{x}^{(l+1)} = \text{ReLU}(\mathbf{z}^{(l)} + F_{\text{shortcut}}(\mathbf{x}^{(l)})) & (\text{残差连接:跨层级融合}) \end{cases} \quad (12)$$

其中, $F_{\text{shortcut}}(x)$ 为shortcut连接(维度适配),定义为:

$$F_{\text{shortcut}}(x) = \begin{cases} x, & \dim(x) = \dim(\mathbf{z}^{(l)}) \\ W_{s,x}, & \text{调整维度} \end{cases} \quad (13)$$

3) 输出层:深度特征生成

通过全局平均池化(global average pooling, GAP)将最后一个残差块的输出($\mathbf{x}^{(3)} \in \mathbf{R}^{128}$)压缩为固定长度的向量,消除空间维度差异:

$$F_{ECG} = \text{GAP}(\mathbf{x}^{(3)}) = \frac{1}{128} \sum_{i=1}^{128} x_i^{(3)} \in \mathbf{R}^{128} \quad (14)$$

RC-MLP-ECG的参数配置如表5所示。

2.4 跨模态注意力融合与分类模块

本模块处理的是每一个1 s EEG片段深度特征 F_{EEG}

表5 RC-MLP-ECG的参数配置

Table 5 Parameter configuration of RC-MLP-ECG		
层级	操作序列	输出尺寸
输入	—	\mathbf{R}^{10}
FC1	Linear(10→32)+BN+ReLU	\mathbf{R}^{32}
ResBlock1	[Linear(32→32)+BN+ReLU]× 2+残差连接	\mathbf{R}^{32}
ResBlock2	[Linear(32→64)+BN+ReLU]× 2+残差连接	\mathbf{R}^{64}
ResBlock3	[Linear(64→128)+BN+ReLU]× 2+残差连接	\mathbf{R}^{128}
输出层	Global average pooling	\mathbf{R}^{128}

与其对应的10 s ECG片段深度特征 F_{ECG} 的融合。脑电深度特征 $F_{EEG} \in \mathbf{R}^{232}$ (来自SE-DenseNet-EEG)编码了瞬时认知状态,心电深度特征 $F_{ECG} \in \mathbf{R}^{128}$ (来自RC-MLP-ECG)表征了持续生理背景。两者信息高度互补,但传统融合方法如直接拼接、加权求和等未考虑模态间动态依赖,导致融合特征判别性不足。

为此,本文提出了一种双向跨模态注意力融合模块,通过两个方向的注意力交互实现多模态信息深度融合。

双向跨模态注意力的核心流程包括特征投影、双向注意力计算、特征融合和分类,具体建模包括:

1) 特征投影

为解决脑电(232维)与心电(128维)的维度异构问题,首先将两者映射到同一高维语义空间,得到查询(Query)、键(Key)、值(Value)矩阵:

$$\begin{cases} \mathbf{Q}_{EEG} = F_{EEG} \cdot \mathbf{W}_Q, \mathbf{W}_Q \in \mathbf{R}^{232 \times 64} \\ \mathbf{K}_{ECG} = F_{ECG} \cdot \mathbf{W}_K, \mathbf{W}_K \in \mathbf{R}^{128 \times 64} \\ \mathbf{V}_{ECG} = F_{ECG} \cdot \mathbf{W}_V, \mathbf{W}_V \in \mathbf{R}^{128 \times 64} \\ \mathbf{Q}_{ECG} = F_{ECG} \cdot \mathbf{W}'_Q, \mathbf{W}'_Q \in \mathbf{R}^{128 \times 64} \\ \mathbf{K}_{EEG} = F_{EEG} \cdot \mathbf{W}'_K, \mathbf{W}'_K \in \mathbf{R}^{232 \times 64} \\ \mathbf{V}_{EEG} = F_{EEG} \cdot \mathbf{W}'_V, \mathbf{W}'_V \in \mathbf{R}^{232 \times 64} \end{cases} \quad (15)$$

其中, $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V, \mathbf{W}'_Q, \mathbf{W}'_K, \mathbf{W}'_V$ 为可学习的投影矩阵,用于将原始特征映射到语义空间,实现模态间的空间对齐。

该维度经过实验调优,为实现模型表达能力与计算效率之间的最佳平衡值。

2) 双向注意力计算

EEG引导的ECG特征增强:计算脑电查询 \mathbf{Q}_{EEG} 与心电键 \mathbf{K}_{ECG} 的相似性,得到脑电对心电的注意力权重:

$$A_{e2c} = \text{softmax} \left(\frac{\mathbf{Q}_{EEG} \cdot \mathbf{K}_{ECG}^T}{\sqrt{d_k}} \right) \quad (16)$$

其中, $\sqrt{d_k}$ 为缩放因子,用于缓解高维度下注意力分数过大导致的softmax饱和问题; $A_{e2c} \in \mathbf{R}^{64 \times 64}$,元素 a_{ij} 表示第 i 个脑电语义对第 j 个心电语义的关注程度(值越大,

该心电图语义与认知状态的相关性越强)。

基于注意力权重 A_{e2c} , 对心电值 V_{ECG} 进行加权求和, 得到增强后的 ECG 特征:

$$F_{ECG}^{enh} = A_{e2c} \cdot V_{ECG} \quad (17)$$

ECG 引导的 EEG 特征增强: 类似地, 计算心电查询 Q_{ECG} 与脑电键 K_{EEG} 的相似性, 得到心电对脑电的注意力权重(反映生理背景对认知特征的筛选强度):

$$A_{c2e} = \text{softmax}\left(\frac{Q_{ECG} \cdot K_{EEG}^T}{\sqrt{d_k}}\right) \quad (18)$$

其中, $A_{c2e} \in \mathbf{R}^{64 \times 64}$, 元素 a_{ij} 表示第 i 个心电图语义对第 j 个脑电语义的关注程度(值越小, 该脑电语义受生理背景的抑制越强)。

基于注意力权重 A_{c2e} , 对脑电值 V_{EEG} 进行加权求和, 得到增强后的 EEG 特征:

$$F_{EEG}^{enh} = A_{c2e} \cdot V_{EEG} \quad (19)$$

3) 特征融合

将增强后的脑电特征 $F_{EEG}^{enh} \in \mathbf{R}^{64}$ 与心电特征 $F_{ECG}^{enh} \in \mathbf{R}^{64}$ 进行拼接(保留两者的互补信息), 得到融合特 $F_{concat} \in \mathbf{R}^{128}$:

$$F_{concat} = F_{EEG}^{enh} \oplus F_{ECG}^{enh} \quad (20)$$

其中, \oplus 表示沿特征维度的拼接操作。

为提取融合特征的高阶语义信息并减少计算量, 采用全连接层对 F_{concat} 进行降维, 并引入 ReLU 激活函数增加非线性:

$$F_{fusion} = \text{ReLU}(W_f \cdot F_{concat} + b_f) \quad (21)$$

其中, $W_f \in \mathbf{R}^{256 \times 128}$ 为降维矩阵, $b_f \in \mathbf{R}^{256}$ 为偏置。

4) 分类器

采用两层全连接网络(MLP)对融合特征 F_{fusion} 进行处理, 最后通过 Softmax 分类器输出认知状态的概率分布:

$$h_1 = \text{ReLU}(W_{c1} \cdot F_{fusion} + b_{c1}), W_{c1} \in \mathbf{R}^{128 \times 256}, b_{c1} \in \mathbf{R}^{128}$$

$$h_2 = W_{c2} \cdot h_1 + b_{c2}, W_{c2} \in \mathbf{R}^{8 \times 128}, b_{c2} \in \mathbf{R}^8$$

$$P = \text{softmax}(h_2) \quad (22)$$

其中, $P = [p_{CA}, p_{DA}, p_{SS}]^T$ 表示认知状态的概率分布。

3 实验结果分析

3.1 实验设置

1) 数据集划分: 采用美国 NASA 公开的 AHPLS 生理数据集^[14]。为进行模型训练与评估, 采用分层抽样策略按 8:1:1 的比例将数据集划分为训练集、验证集和测试集, 确保不同认知状态和不同飞行员样本在 3 个子集中的比例保持一致。

2) 评估指标: 采用以下核心指标评估模型性能:

整体准确率 (overall accuracy, OA): 正确分类的样本数占总样本数的比例。

精确率 (precision, P): 针对类别 c , 预测为正例且正确的比例:

$$P_c = \frac{TP_c}{TP_c + FP_c} \quad (23)$$

其中, TP_c 为类别 c 的真阳性数, FP_c 为假阳性数召回率 (recall, R): 针对类别 c , 正确识别出的正例比例:

$$R_c = \frac{TP_c}{TP_c + FN_c} \quad (24)$$

其中, FN_c 为类别 c 的假阴性数 F1 分数 (F1-Score, F1): 精确率与召回率的调和平均数:

$$F1_c = 2 \times \frac{P_c \times R_c}{P_c + R_c} \quad (25)$$

3) 实验环境

硬件: NVIDIA GeForce RTX 4070 Laptop GPU, Intel[®] Core[™] i7-13700HX Processor CPU, 32 GB RAM。

软件: Python 3.8, PyTorch 1.12.1, scikit-learn 1.0.2。

4) 模型训练

选用 AdamW 优化器解决特征分布差异问题, 初始学习率设为 1×10^{-4} , 辅以余弦退火衰减策略平衡收敛效率与局部最优规避; 依据 EEG 与 ECG 模态的片段比例, 差异化设置批处理量, 确保跨模态梯度均衡更新; 通过 Dropout (概率 0.2) 抑制神经元共适应, 叠加权重衰减 (1×10^{-4}) 压缩参数空间; 考虑到数据集中 3 类认知状态的样本数量存在显著不平衡, 为防止模型在训练中过度关注易于分类的多数类样本而忽略关键的少数类样本, 本研究采用 Focal Loss 作为模型的损失函数。Focal Loss^[18] 是对标准交叉熵损失的一种改进, 它通过引入一个调制因子来降低易分类样本的权重, 从而使模型能够更专注于学习难分类的样本; 设定最大训练轮次 100, 并启用早停策略 (验证集监控, patience=10), 在泛化性能饱和时提前终止, 确保模型最优性。

3.2 实验结果

本文提出的基于跨模态注意力融合的飞行员认知状态识别模型在测试集上取得了优异的性能, 总体分类准确率 (overall accuracy, OA) 达到了 97.44%。模型的训练与验证准确率变化曲线如图 4 所示, 可以看出, 随着训练轮次的增加, 训练准确率和验证准确率均稳步上升并最终收敛, 验证曲线的平稳趋势表明模型具有良好的泛化能力, 未出现过拟合现象。

为深入评估模型对不同认知状态的判别能力, 表 6 展示了模型在各类别的详细分类性能。模型对“注意力集中 (CA)”状态的识别能力最为突出, 其精确率高达 98.48%, 召回率为 98.73%, F1 分数达到 98.60%。这表明模型能够精准捕捉并学习到飞行员在执行常规、专注任务时稳定且独特的生理模式。对于样本量相对较少的“注意力分散 (DA)”和“惊吓/惊奇 (SS)”状态, 模型同样展现了强大的判别力。DA 状态的 F1 分数为 92.26% (精确率 93.37%,

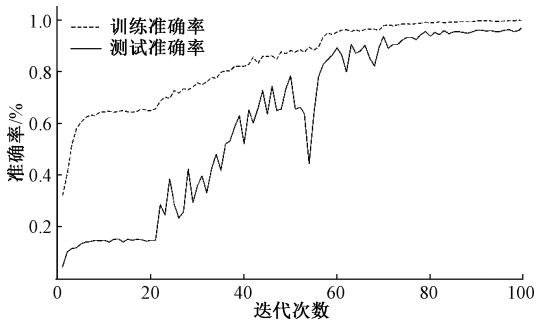


图4 模型的训练、测试曲线图

Fig. 4 Training and validation accuracy curves of the model

召回率 91.18%),而 SS 状态的 F1 分数为 92.23%(精确率 91.75%,召回率 92.71%)。尽管 CA 状态的样本量远超 DA 和 SS,模型对后两类少数样本的召回率依然能保持在 91%以上,这充分证明本研究采用的 Focal Loss 取得了显著效果,有效缓解了数据长尾分布带来的学习偏差,使得模型对突发、关键的认知状态变化同样具备高敏感度。模型在测试集上的分类混淆矩阵如图 5 所示,可直观反映三类认知状态的分类结果与整体判别效果。

表 6 本文模型在各类别上的分类性能

Table 6 Classification performance of the proposed model on each category

类别	精确率 (Precision)	召回率 (Recall)	F1 分数 (F1-Score)
注意力集中	98.48	98.73	98.60
注意力分散	93.37	91.18	92.26
惊吓/惊奇	91.75	92.71	92.23

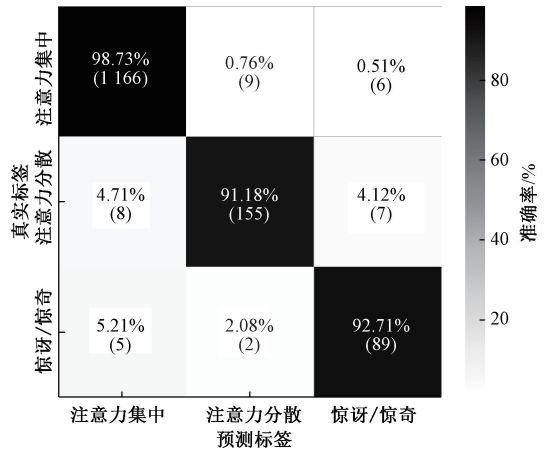


图 5 模型在测试集上的分类混淆矩阵

Fig. 5 Confusion matrix of the model on the test set

3.3 消融实验

为系统地验证模型各核心组件的有效性,本文设计了消融实验。其结果如表 7 所示。

实验结果明确了模式融合的必要性。单独使用脑电信号的 EEG-only 模型准确率为 93.30%,而引入心电信号的完整模型则将准确率提升至 97.44%,增幅超过 4%。这表明,尽管 EEG 是认知分析的核心,但融合能够反映自主神经系统状态的 ECG 信号,提供了更丰富、互补的生理信息维度,从而构建出更具判别力的特征空间。

然而,仅仅融合多模态信息并不足以保证性能提升,融合策略的选择至关重要。实验中,采用简单拼接融合的模式准确率仅为 92.88%,效果甚至劣于单模态模型。这揭示了简单的特征向量拼接可能引入冗余特征并干扰核心信息,导致“1+1<1”的负面效果。与之形成鲜明对比

表 7 消融实验结果

Table 7 Ablation experiment results

实验配置	实验描述	总体准确率/%
EEG-only	仅使用脑电特征分支进行提取分类	93.30
拼接融合	使用 EEG 和 ECG 双流分支、通过直接拼接特征进行融合	92.88
Remove-SE	在本文完整模型中去除了 Densenet 中的 SE 模块	95.78
标准交叉熵损失	在本文完整模型中使用标准的交叉熵损失函数替换 Focal Loss	94.55
本文模型	完整的双流跨模态注意力网络	97.44

的是,本文模型凭借双向跨模态注意力机制,能够智能地识别并增强有效信息、抑制噪声,实现了真正意义上的高效融合。

在此基础上,模型内部各组件的优化作用也得到了验证。在 EEG 分支中,移除 SE 模块会导致模型准确率下降至 95.78%,凸显了其通过通道注意力增强关键特征表达的积极作用。同样地,损失函数的选择也对最终结果产生了决定性影响。为量化 Focal Loss 的贡献,将其与标准交叉熵损失函数对比,发现后者会导致准确率下降近 3%~94.55%。这清晰地表明,Focal Loss 通过其独特的聚焦机

制,成功引导模型克服了数据不平衡的挑战。

综上所述,消融实验系统地证明了:EEG-ECG 双模态融合的必要性、双向跨模态注意力融合的先进性,以及 SE 模块和 Focal Loss 等关键组件的有效性是本文模型取得优异性能的共同保障。

3.4 与其他方法对比

为进一步将本文模型置于现有研究的背景下进行评估,选取了近年来飞行员认知状态识别领域具有代表性的 4 种方法进行性能对比,具体结果如表 8 所示。

表 8 飞行员认知状态识别各方法对比

Table 8 Comparison of different methods for pilot cognitive state recognition

方法	网络架构	生理信号	检测的飞行员认知状态	总体准确率/%
文献[13]	CNN+LSTM	EEG、ECG、呼吸和 EDA	分心、负荷、疲劳与正常状态	85.2
文献[19]	DCAEN	EEG	疲劳状态、正常状态	91.67
文献[20]	机器学习集成模型	EEG、GSR	注意力状态、低认知负荷	90.1
文献[21]	机器学习集成模型	EEG	无任务、注意力状态、惊吓\惊奇状态	86.0
本文	双流跨模态融合模型	EEG、ECG	注意力状态、惊吓\惊奇状态	97.44

对比结果显示,本文提出的双流跨模态融合模型在总体准确率上表现出显著的优越性,达到 97.44%。例如,与同样数据集的文献[20](准确率 86%)和文献[21](准确率 90.1%)相比,本文模型在准确率上取得了超过 7%~11% 的提升。

这种性能优势主要归因于以下两方面:

1) 更高效的模态选择与融合策略。文献[13]虽然采用了包括 EEG、ECG 在内的 4 种模态,但其准确率仅为 85.2%。这侧面印证了并非模态数量越多性能越好,关键在于融合的质量。本文模型仅采用 EEG 和 ECG 两种核心生理信号,但凭借精心设计的双向跨模态注意力机制,实现了对两种模态信息的深度、智能交互与增强,其融合效率远高于文献[13]中较为传统的 CNN 与 LSTM 结合的串联式处理方式。

2) 更具针对性的特征提取网络。相较于采用 DCAEN 或传统机器学习集成模型的文献[19]、[20]和 [21],本文为 EEG 和 ECG 两种异构数据分别设计了专属的深度学习分支。SE-DenseNet-EEG 网络能够充分挖掘高维 PLV 特征中的跨频段耦合关系,而 RC-MLP-ECG 网络则有效捕捉了低维 ECG 生理指标间的非线性关联。这种精细化的特征工程与深度表征学习,为后续的高效融合与精准分类奠定了坚实基础,从而在整体上超越了其他方法。

4 结 论

为解决现有方法在多模态生理信号融合上的局限性,实现对飞行员认知状态的精准评估,本文提出了一种基于双向跨模态注意力的双流深度学习模型。该模型通过深度融合脑电信号的功能连接特性与心电信号的生理状态信息,实现了对注意力集中、分散及惊吓/惊奇等关键状态的高精度识别。在公开的 NASA 飞行员生理数据集上的实验结果证明,该模型总体识别准确率高达 97.44%。本研究为开发新一代客观、高效的飞行员状态监测系统提供了可靠的技术支持,对提升飞行安全、优化人机交互、预防人为因素导致的航空事故具有一定的应用价值。

参考文献

[1] WU Y B, ZHANG SH G, ZHANG X, et al. Analysis on coupling dynamic effect of human errors in aviation

safety[J]. Safety Science, 2023, 192: 107277.

- [2] 李文杰. 模拟飞行任务下人员认知状态分析及多模态生理指标优化[D]. 长沙:国防科技大学, 2022.
- LI W J. Personnel cognitive state analysis and multimodal physiological index optimization under simulated flight task [D]. Changsha: National University of Defense Technology, 2022.
- [3] 蔡正祥, 吴奇, 黄丹, 等. 基于 FPA 优化的 GP 算法的飞行员认知状态识别[J]. 电光与控制, 2016, 23(11): 78-84.
- CAI ZH X, WU Q, HUANG D, et al. Recognition of pilot's cognitive state based on FPA optimized GP[J]. Electronics Optics & Control, 2016, 23(11): 78-84.
- [4] KRIEGLSTEIN F, BEEGE M, REY G D, et al. A systematic meta-analysis of the reliability and validity of subjective cognitive load questionnaires in experimental multimedia learning research[J]. Educational Psychology Review, 2022, 34: 2485-2541.
- [5] SCHUESSLER K, FISCHER V, WALPUSKI M, et al. Investigating construct validity of cognitive load measurement using single-item subjective rating scales[J]. Instructional Science, 2025, 53: 71-97.
- [6] WANG X M, ZHANG J W, HE CH H, et al. A novel emotion recognition method based on the feature fusion of single-lead EEG and ECG signals[J]. IEEE Internet of Things Journal, 2024(5): 8746-8756.
- [7] 冯笑, 代少升, 黄炼. 基于可解释深度学习的单通道脑电跨被试疲劳驾驶检测[J]. 仪器仪表学报, 2023, 44(5): 140-149.
- FENG X, DAI SH SH, HUANG L. Cross-subject driver fatigue detection from single-channel EEG with an interpretable deep learning model [J]. Chinese Journal of Scientific Instrument, 2023, 44(5): 140-149.
- [8] 俞小彤, 赵若辰, 宁晓琳. 多通道权重融合和小波分解的癫痫棘波检测方法[J]. 电子测量与仪器学报, 2024, 38(10): 24-34.
- YU X T, ZHAO R CH, NING X L. Multichannel weight fusion and wavelet decomposition method for detecting

- epileptic spikes[J]. *Journal of Electronic Measurement and Instrumentation*, 2024, 38(10): 24-34.
- [9] 杜扶遥,姜囡,陆思宇. 基于语音脑电的双模态心理压力分级评估研究[J]. *电子测量技术*, 2024, 47(19): 114-122.
- DU F Y, JIANG N, LU S Y. Evaluation of psychological stress level based on speech and EEG signal [J]. *Electronic Measurement Technology*, 2024, 47(19): 114-122.
- [10] ZHAO M T, JIA W J, JENNINGS S, et al. Monitoring pilot trainees' cognitive control under a simulator-based training process with EEG microstate analysis[J]. *Scientific Reports*, 2024, 14: 24632.
- [11] JI L, YI L Y, LI H W, et al. Detection of pilots' psychological workload during turning phases using EEG characteristics[J]. *Sensors*, 2024, 24(16): 5176.
- [12] SHAO SH Y, ZHOU Q X, LIU ZH Q. A new assessment method of the pilot stress using ECG signals during complex special flight operation[J]. *IEEE Access*, 2019, 7: 185360-185368.
- [13] HAN S Y, KWAK N S, OH T, et al. Classification of pilots' mental states using a multimodal deep learning network[J]. *Biocybernetics and Biomedical Engineering*, 2020, 40(1): 324-336.
- [14] HARRIVEL A R, STEPHENS C L, MILLETICH R J, et al. Prediction of cognitive states during flight simulation using multimodal psychophysiological sensing[C]. *AIAA SciTech 2017*, 2017.
- [15] LI C B, TANG T, PAN Y, et al. An efficient graph learning system for emotion recognition inspired by the cognitive prior graph of EEG brain network[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 36(4): 1-15.
- [16] GUPTA A, SIDDHAD G, PANDEY V, et al. Subject-specific cognitive workload classification using EEG-based functional connectivity and deep learning[J]. *Sensors*, 2021, 21(20): 6710.
- [17] HUANG G, LIU ZH, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]. *2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2017: 2261-2269.
- [18] LIN T Y, GPYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. *2017 IEEE International Conference on Computer Vision(ICCV)*, 2017: 2999-3007.
- [19] WU E Q, PENG X Y, ZHANG C ZH, et al. Pilots' fatigue status recognition using deep contractive autoencoder network [J]. *IEEE Transactions on Instrumentation and Measurement*, 2019, 68(10): 3907-3919.
- [20] HARRIVEL A R, LILES C, STEPHENS C L, et al. Psychophysiological sensing and state classification for attention management in commercial aviation [C]. *AIAA SciTech 2016*, 2016.
- [21] ALRESHIDI I, MOULITSAS I, JENKINS K W. Multimodal approach for pilot mental state detection based on EEG[J]. *Sensors*, 2023, 23(17): 7350.

作者简介

张文泽, 硕士, 主要研究方向为生理信号处理及其在飞行安全领域的识别应用。

E-mail: skzwz@qq.com

王在俊(通信作者), 研究员, 硕士生导师, 主要研究方向为飞行技术与飞行安全、智能计算。

E-mail: zaijunwang@cafuc.edu.cn

蒋宇恒, 硕士, 主要研究方向为生理信号处理及其在飞行安全领域的识别应用。

E-mail: 2951168474@qq.com

杨睿哲, 硕士, 主要研究方向为生理信号处理及其在飞行安全领域的识别应用。

E-mail: 764861685@qq.com