

DOI:10.19651/j.cnki.emt.2518752

基于多任务轮换学习的非接触式心理感知方法^{*}

王宇 牛知艺

(中国民用航空飞行学院航空电子电气学院 成都 641450)

摘要: 计算机视觉在智能感知领域发挥着重要作用。现有的心理状态感知方法仅局限于面部表情识别或远程光电容积描记术等单一任务,难以实现多维特征的协同感知;而融合多模态生理信号的方法则面临较高的计算成本。针对这些问题,本文提出一种基于多任务轮换学习的非接触式心理状态感知方法。该方法通过多任务模型处理人脸视频,同时完成 rPPG 心率信号提取、情感指标预测和心理状态分类 3 个任务。实验结果表明,该模型在 rPPG 心率信号提取上的平均绝对误差为 3.78,情感效价和唤醒度预测的准确率分别为 97.47% 和 96.75%,心理状态分类的准确率为 97.42%。该方法为非接触式心理状态感知提供了一种高效的多任务处理方案,具有重要的理论和实践价值。

关键词: 计算机视觉;远程光电容积脉搏波描记法;情绪识别;非接触心理状态感知;多任务学习;轮换训练

中图分类号: TN911.7 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Non contact psychological perception method based on multi-task rotation learning

Wang Yu Niu Zhiyi

(College of Aviation Electronic and Electrical Engineering, Civil Aviation Flight University of China, Chengdu 641450, China)

Abstract: Computer vision plays a crucial role in the field of intelligent perception. Existing methods for psychological state perception are typically limited to single tasks such as facial expression recognition or remote photoplethysmography, making it difficult to achieve collaborative perception of multidimensional features. Additionally, approaches that integrate multimodal physiological signals face high computational costs. To address these challenges, this paper proposes a non-contact psychological state perception method based on multi-task rotation learning. The proposed approach utilizes a multi-task model to process facial video, simultaneously performing three tasks: rPPG heart rate signal extraction, emotional valence and arousal prediction, and psychological state classification. Experimental results show that the model achieves an average absolute error of 3.78 for rPPG heart rate signal extraction, prediction accuracies of 97.47% and 96.75% for emotional valence and arousal, respectively, and a classification accuracy of 97.42% for psychological state. This method provides an efficient multi-task processing solution for non-contact psychological state perception, offering significant theoretical and practical value.

Keywords: computer vision; remote photoplethysmogram; emotion recognition; non-contact psychological state perception; multi-task learning; rotation training

0 引言

在当今的人机交互、医疗健康、交通安全等多个领域,人体心理状态感知已成为一个关键的研究方向。心理状态不仅是影响个体行为的重要因素^[1],在飞行员操控、汽车驾驶及空中交通管制等高负荷任务场景中,更是确保安全与工作效率的核心要素。因此,精准识别心理状态可为个性

化服务、安全预警和健康评估提供依据,提升生活质量与工作效率,并在关键场景保障生命财产安全。

许多研究者通过单模态来感知人物的心理状态,常见的模态包括面部表情和远程光电容积脉搏波描记术(photoplethysmography, PPG)。面部表情识别作为最直观的情绪感知手段,依赖面部动作编码系统通过表情动作单元推断情绪状态^[2]。然而,这类方法面临三重挑战:其

收稿日期:2025-05-07

*基金项目:国家自然科学基金青年基金(62406207)、四川省自然科学基金青年基金(2025ZNSFSC1502)、中央高校基本科研业务费(25CAFUC03023)资助

一,表情易受主观控制或社交情境干扰,如人物在高压环境下的情绪抑制会导致识别偏差;其二,表情中一些细微的变化难以被察觉,并且情绪是一种内在的反应,其变化很多时候不会通过表情完全被反映出来^[3];其三,离散情绪模型难以刻画动态情绪演变,单一表情无法关联深层心理机制。相反,利用生理信号可以较为准确地反映人物的情绪状态,目前远程光电容积脉搏波描记术(remote photoplethysmography, rPPG)是研究者们关注的焦点之一。rPPG 技术通过视频提取血液脉冲信号,进而测量心率、呼吸率以及心率变异性等生理指标^[4],为心理状态感知提供了客观依据。rPPG 信号的收集容易受头部运动、光照变化等噪声的干扰^[5],因此单一使用 rPPG 信号进行情感分析容易受到外界环境的干扰,导致识别结果出现偏差。由此可见,单任务模型在心理感知任务中存在协同性缺失的问题^[6],忽略了其他相关因素对心理状态的影响。

为克服单模态信息易产生识别偏差的问题,当前,多模态信息融合已成为心理状态感知研究的重要趋势。张志雯等^[7]融合脑电、肌电与皮电信号特征进行情绪预测。李奇等^[8]使用脑电信号对情绪进行分类。王宇等^[9]将 rPPG 信号与人脸三维法向量相结合,以提升情绪识别的精度。Mai 等^[10]开发了多融合剩余记忆网络,实现话语级情绪分类。然而,多模态融合模型需要训练并使用不同的模型对不同模态的信息进行收集与测量,这会导致网络臃肿且计算量庞大。对于不同模态信息的融合,往往也需要依赖复杂的融合层,并需经过长时间训练,导致计算成本高且难以进行细粒度任务优化。

因此,研究者们又通过多任务学习的方法来共享特征,同时对多种心理状态评估指标进行测量。然而,经典的联合训练范式在实际应用中面临显著挑战;当多个任务在优化目标、数据分布或学习难度上存在较大差异时,强制共享模型参数易引发负迁移现象,即任务间的干扰导致部分甚至所有任务的性能低于单独训练的水平。尽管现有方法如谢平等^[11]使用的为损失函数分配不同权重来找到最优解,或是 Kendall 等^[12]使用的动态权重分配亦或是更精细的网络结构设计^[13],在一定程度上缓解了冲突,但它们往往计算复杂、调参困难,且在处理特定组合的异构任务时效果不稳定,未能从根本上解决多任务联合优化过程中固有的参数干扰难题。

针对现有研究中单任务协同性缺失、多模态计算效率瓶颈以及多任务联合训练的参数干扰的核心问题,本文提出了多任务轮换学习方法通过创新机制实现突破。传统单任务模型难以整合生理与情绪特征,多模态融合又面临计算复杂的困境,而本方法通过构建跨任务特征协同机制,实现面部血流信号与表情动态的互补表征,既保留生理信号的客观性,又捕捉情绪体验的主观性;同时采用分阶段轮换训练策略,避免多任务优化时的参数冲突。实验表明,本文的方法在 rPPG 心率提取、情绪预测和心理状态分类任务

中均实现较好的表现,较单任务模型性能有所提升,为非接触式心理状态感知提供了全新有效的解决方案,在航空驾驶、医疗监护等场景中具有重要的应用价值。

1 多任务心理状态感知模型的构建

多任务心理状态感知模型的整体架构如图 1 所示,主要由共享特征提取器与 3 个子任务预测头组成。首先,共享特征提取器从视频数据中提取高级时空特征,该共享特征将作为 3 个子任务的输入:1)rPPG 信号提取头:从共享特征中提取与人体面部颜色变化相关的特征,并通过信号处理与回归分析获得人体的 rPPG 信号;2)情绪预测头:对共享特征中的人体表情信息进行时序建模,预测人体的唤醒度和效价值;3)心理状态预测头:综合多模态特征,对人体的心理状态是否适合当前的工作任务进行二分类判断。

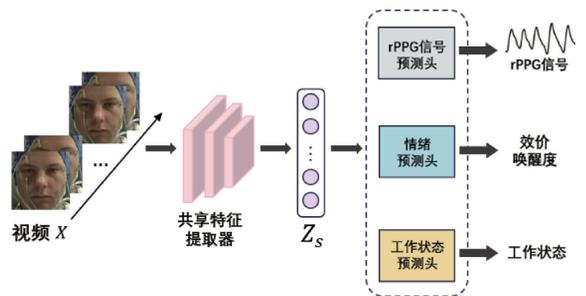


图 1 多任务心理状态感知模型总体框架

Fig. 1 Overall framework of multitask psychological state perception model

1.1 共享特征提取器的构建

共享特征提取器的核心目标是提取公共特征,并将其用于多个相关任务,以实现信息共享、提升模型在各任务上的泛化能力,同时减少参数规模与计算开销。如图 2 所示,本文设计的共享特征提取器由主干特征提取、时空分块标记化、与时序差分 Transformer 3 个模块组成。该提取器能够有效建模连续帧之间的时序变化,融合时空信息,进而形成高质量的共享特征,为后续的子任务提供更加精准和稳定的输入。

1) 主干特征提取模块

视频的每一帧图像包含了丰富的视觉信息,为了减少后续时序建模的计算量,本文首先采用主干特征提取模块提取浅层视觉特征。输入的人脸视频 $X \in \mathbb{R}^{3 \times T \times H \times W}$ 依次通过三维卷积层(3D convolution, 3D Conv)、批量归一化层(batch normalization, BN)、线性整流激活函数(rectified linear unit, ReLU)和三维最大池化层(3D max pooling, 3D MaxPool),最终获得浅层特征表示 X_s ,如式(1)所示。

$$X_s = 3DMaxPool(ReLU(BN(3DConv(X)))) \quad (1)$$

该模块在提取局部纹理信息的同时,能够降低计算成本,为后续时序建模提供高效输入。

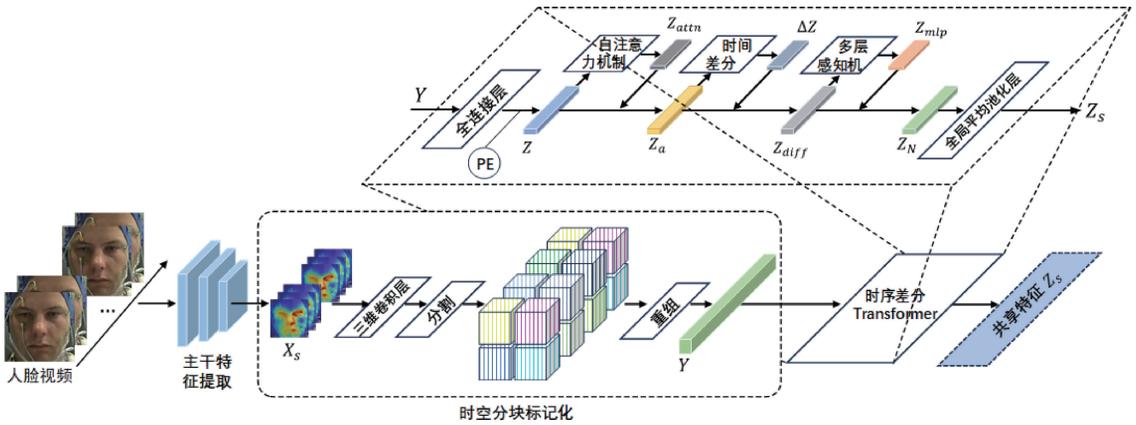


图 2 共享特征提取器结构图

Fig. 2 Structure diagram of shared feature extractor

2) 时空分块标记化模块

时空分块标记化的核心目标是将输入的浅层特征 X_s 压缩并划分为多个时空块,再将其转换为序列化的令牌 Y ,以适配 Transformer 结构。 X_s 首先通过一个三维卷积层,将特征通道数从 C 扩展至 C' ,以增强时空局部特征表达能力。再调换特征维度顺序,将空间维度分割为 H_{seg} 和 W_{seg} ,而后对时间维度进行重组得到序列化的令牌 $Y \in \mathbb{R}^{T \cdot H_{seg} \cdot W_{seg} \times (n_{segment}^2 \cdot C')}$,其中,定义序列长度 $L = T \cdot H_{seg} \cdot W_{seg}$,特征维度 $D = n_{segment}^2 \cdot C'$ 。序列当中的每个令牌可以捕捉局部运动与外观变化,这对 rPPG 信号(依赖细微的面部血流变化)和情绪(依赖面部表情)的联合建模至关重要。

3) 时序差分 Transformer 模块

在获得序列化令牌之后,本文采用 Transformer 模型进行时空特征提取。首先将 Y 经过全连接层投影到标准输入维度 D' ,增强全局语义信息,为后续注意力机制提供更紧凑的表示。同时,为了区分不同时间步的特征,需要为分块特征加入位置编码(positional encoding, PE),得到 Transformer 的输入 Z 。而后计算其自注意力机制得到 Z_{attn} ,并与 Z 进行残差连接得到 Z_a 。如此便可以捕捉序列中不同位置的依赖关系,增强全局建模能力。

由于 Transformer 结构本身不具备时序感知能力,本文提出时间差分机制,通过计算相邻时间步的特征差分 $\Delta Z = Z_a[t] - Z_a[t-1]$, ($t \geq 1$),突出时间维度上的动态变化,并抑制光照变化等背景噪声。而后得到的每一个特征差分 ΔZ 与 Z_a 进行残差连接得到特征。通过这种方式,模型能够更有效地聚焦于面部血流引起的微小颜色变化,同时降低光照干扰的影响。

Z_{diff} 经过一个多层感知机网络来学习更高阶的特征得到 Z_{mlp} , 并与其本身进行残差连接最终得到 Z_l , 在经过 N 层 Z_l 堆叠后,最终编码特征表示为: $Z_N \in \mathbb{R}^{L \times D'}$ 。为了获得全局信息,采用全局平均池化层将序列特征聚合为单一

向量,从而提取任务相关的全局统计量,如平均心率和主导情绪状态,最终得到共享特征 $Z_s \in \mathbb{R}^{D'}$ 。

1.2 多任务预测器的构建

本文设计了 3 个独立的任务预测头,将共享特征转换为特定任务的表征。其结构如图 3 所示。

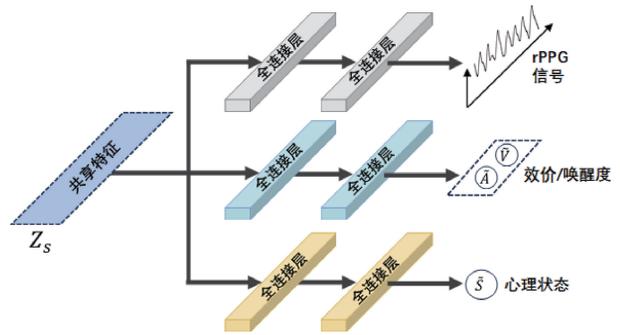


图 3 多任务预测器

Fig. 3 Multi-task predictor

1) rppg 信号提取头

rPPG 信号提取头的核心目标是从共享的时空特征中解码与面部血流动力学相关的周期性生理信号。为了有效提取与面部微血管容积变化相关的全局统计信息,该预测头采用双全连接层对高维特征进行非线性映射,以增强信号的可分辨性。具体而言,共享特征 Z_s 首先通过第 1 层全连接层,让特征维度从 D' 变为 \mathbb{R}^{256} 进行特征增强,并借助 ReLU 激活函数引入非线性表达能力;随后,第 2 层全连接层将特征映射至目标维度 $\mathbb{R}^{128 \times t}$,输出以 128 Hz 采样率记录的 t 秒长 rPPG 波形。

2) 情绪预测头

对于情绪预测任务,本文对效价、唤醒度指标进行预测,其中,效价表示情绪的正负向属性,即个体对刺激的愉悦或厌恶程度,唤醒度表示情绪的激活水平。该预测头由两个全连接层组成,最终输出 2 维向量,分别对应情绪的效价 \tilde{V} 和唤醒度 \tilde{A} 。考虑到 DEAP 数据集的情绪标注范

围为 $[1, 9]$, 本模型采用 S 形曲线函数将预测值归一化至 $[0, 1]$, 随后采用线性缩放函数将其映射至目标区间。

3) 心理状态预测头

心理状态预测模块旨在整合生理和情绪特征, 以评估受试者是否处于适宜执行当前工作任务的心理状态。该预测头通过共享 rPPG 信号预测和情绪预测任务的参数, 建模生理与情绪之间的相互作用, 从而提高工作状态评估的准确性。该预测头由两个全连接层组成, 输出维度为 1, 对应受试者的工作状态预测值 \tilde{S} , 其值为 0 或者 1, 0 代表不适合工作的状态, 1 代表适合工作的状态。

2 多任务轮换训练

传统的多任务学习通常采用同步训练策略, 即所有任务的损失函数被加权求和, 并在训练过程中同时优化。然而, 这种训练方式存在诸多缺陷: 如果任务间的特征需求不一致可能会导致任务之间的冲突; 同时训练多个复杂任务可能会导致训练困难或收敛。

为解决上述问题, 本文提出了一种多任务轮换训练策略如图 4 所示。

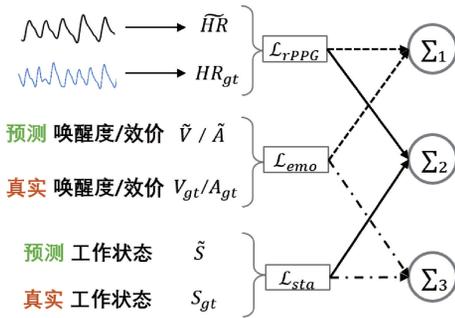


图 4 轮换训练策略

Fig. 4 Rotation training strategy

该策略通过分阶段任务组合的方式逐步优化模型。首先分别计算 3 个任务的损失函数, 如式(2)所示。

$$\begin{cases} L_{rPPG} = RMSE(\hat{HR}, HR_{gt}) \\ L_{emo} = MSE(\tilde{V}, V_{gt}) + MSE(\tilde{A}, A_{gt}) \\ L_{sta} = BCE(\tilde{S}, S_{gt}) \end{cases} \quad (2)$$

为了减少噪声干扰并更精准地捕获心率信息, 本文采用快速傅里叶变换^[14]对预测的 rPPG 信号与真实 PPG 信号进行频域变换, 并分别提取对应的心率信号 \hat{HR} 与 HR_{gt} 。随后, 计算预测心率与真实心率之间的均方根误差(root mean square error, RMSE)作为损失函数。对于情绪预测任务, 本文采用均方误差(mean squared error, MSE)计算预测效价 \tilde{V} 和唤醒度 \tilde{A} 与其真实值 V_{gt} 与 A_{gt} 之间的误差。针对心理状态分类任务, 本文使用二元交叉熵损失(binary cross-entropy, BCE)来衡量预测的心理状态

\tilde{S} 与真实标签 S_{gt} 之间的偏差。

本文采用任务排列组合的方式逐步优化模型。轮换训练策略包括 3 个阶段, 首先训练 rPPG 提取任务和情绪预测任务 $\Sigma_1 = L_{rPPG} + L_{emo}$ 。再训练 rPPG 提取和心理状态预测任务 $\Sigma_2 = L_{rPPG} + L_{sta}$ 。最后训练情绪预测和心理状态预测任务 $\Sigma_3 = L_{emo} + L_{sta}$ 。通过这样的训练策略可以有效减少任务间的冲突, 提高模型的训练稳定性和效率。每个阶段的损失函数只涉及当前阶段的任务组合, 有助于任务间知识的共享和任务的相互促进。

3 实验与结果分析

3.1 数据预处理

本文采用公开的基于生理信号的情绪分析数据库(a database for emotion analysis using physiological signals, DEAP)来训练模型, 本文对该数据集的使用已得到官方的授权。该数据集记录了 32 名受试者在观看 40 段 1 min 音乐视频时的 40 通道生理信号, 以及受试者对效价、唤醒度、控制感和喜好度的自评评分。前 22 名受试者还同步录制了面部表情视频。本研究选择第 39 通道(PPG 信号)作为训练数据。每段视频随机截取连续的 150 帧作为训练样本。由于模型输入为面部视频, 仅选用有面部视频记录的前 22 名受试者的数据, 其中前 15 名用于模型训练, 其余 7 名用于模型测试。

在心理状态分类任务中, 由于 DEAP 数据集未提供真实标签, 本文邀请了 13 位飞行员心理评估专家(包含 3 名主任体检医师), 依据飞行员工作状态标准对每段视频进行飞行工作状态标注。所有专家均持有《民用航空人员体检合格证》(依据 CCAR-67FS 规章颁发)。标注过程参照《民用航空医学手册》(ICAO Doc 8984/AN895)中的第 9 章(心理健康)与第 17 章(疲劳与飞行运行)的标准。每位专家(不包含 3 位主任体检医师)对同一视频单元独立完成状态判别, 当共识度 $\geq 70\%$ 时, 采用多数意见作为结果, 当共识度 $< 70\%$ 时, 则由 3 位主任体检医师重新评估并经多数票决定最终标签。

3.2 实验环境

模型在配备 NVIDIA 4090 GPU 24GB 显存、Intel i7-13700K CPU 和 32 GB 内存的 Windows 11 系统上进行训练。本文使用 Adam 作为优化器, 学习率设置为 0.000 1, 模型一共训练了 500 轮。

3.3 评价指标

对于 rPPG 心率信号提取任务, 本文采用最常用的心率评估指标, 包括心率平均绝对误差(mean absolute error of heart rate, MAE-HR)、心率均方根误差(root mean square error of heart rate, RMSE-HR)和心率皮尔逊相关系数(pearson correlation coefficient of heart rate, r-HR)。其中 MAE-HR 定义如式(3)所示。

$$MAE-HR = \frac{1}{N} \sum_{i=1}^N |HR_{gt_i} - \widehat{HR}_i| \quad (3)$$

式中: N 表示心率信号的长度。

RMSE-HR 定义如式(4)所示。

$$RMSE-HR = \sqrt{\frac{1}{N} \sum_{i=1}^N (HR_{gt_i} - \widehat{HR}_i)^2} \quad (4)$$

r-HR 定义如式(5)所示。

r-HR =

$$\frac{\sum_{i=1}^N (HR_{gt_i} - \overline{HR_{gt}}) (\widehat{HR}_i - \overline{HR})}{\sqrt{\sum_{i=1}^N (HR_{gt_i} - \overline{HR_{gt}})^2} \sqrt{\sum_{i=1}^N (\widehat{HR}_i - \overline{HR})^2}} \quad (5)$$

式中: $\overline{HR_{gt}}$ 、 \overline{HR} 分别为 HR_{gt} 、 \widehat{HR} 的平均值。

在情绪预测任务中,DEAP 数据集的效价与唤醒度值范围在 $[1, 9]$, 将 $[1, 6)$ 范围归类为“低”, $[6, 9]$ 范围归类为“高”。因此,效价被分为“高效价”和“低效价”,唤醒度被分为“高唤醒度”和“低唤醒度”。本文计算效价准确率(valence accuracy, ACC-V)、以及唤醒度准确率(arousal accuracy, ACC-A)。

其中 ACC-V 定义如式(6)所示。

$$ACC-V = \frac{TP_V + TN_V}{N} \quad (6)$$

式中: TP_V 表示预测效价为“高效价”且真实预测效价为“高效价”的视频数, TN_V 表示预测效价为“低效价”且真实预测效价为“低效价”的视频数, N 为视频总数。

ACC-A 定义如式(7)所示。

$$ACC-A = \frac{TP_A + TN_A}{N} \quad (7)$$

式中: TP_A 表示预测效价为“高唤醒度”且真实预测唤醒度为“高唤醒度”的视频数, TN_A 表示预测效价为“低唤醒度”且真实预测唤醒度为“低唤醒度”的视频数。

对于二分类的心理状态分类任务,计算其状态预测值的平均准确率(status accuracy, ACC-Sta)。

ACC-Sta 定义如式(8)所示。

$$ACC-Sta = \frac{TP_s + TN_s}{N} \quad (8)$$

式中: TP_s 表示预测心理状态为 1 且真实预测心理状态为 1 的视频数, TN_s 表示预测心理状态为 0 且真实预测心理状态为 0 的视频数。

3.4 轮换训练对比试验

为验证轮换训练策略在多任务模型中的有效性,本文对比了采用同步训练与轮换训练两种不同方式的模型。两种训练策略的具体实验数据如表 1 所示。

从表格中可以看出,采用轮换训练策略的模型相较于同步训练,r-HR 提高了 0.05,MAE-HR 与 RMSE-HR 分别下降了 0.25 和 0.62。在情绪预测任务中,ACC-V 提高了 0.95%,ACC-A 提高了 0.9%,在状态分类任务中,

ACC-Sta 提升了 0.86%。这些结果表明,轮换训练有效减少了任务间的干扰,使得每个任务能够在更加专注的优化方向上进行训练,从而提高了整体模型的性能。轮换训练在多任务学习场景中展现出了更优的学习效果和更强的任务适应性。

表 1 轮换训练方式与同时训练方式性能对比

Table 1 Performance comparison between alternating training method and simultaneous training method

评价指标	同时训练	轮换训练
MAE-HR	4.03	3.78
RMSE-HR	8.15	7.53
r-HR	0.67	0.72
ACC-V	96.52%	97.47%
ACC-A	95.85%	96.75%
ACC-Sta	96.56%	97.42%

3.5 多任务预测头消融实验

为了验证 3 个多任务预测头对于模型心理状态感知能力提升的有效性,本节设置了 4 个变体进行实验对比:只包含 rPPG 信号预测头、只包含情绪预测头、只包含工作状态预测头、多任务模型(包含全部 3 个预测头)。消融实验的结果如表 2 所示。

表 2 多任务预测头消融实验结果

Table 2 Experimental results of multi task prediction head ablation

评价指标	只含 rPPG 信号预测头	只含情绪预测头	只含工作状态预测头	多任务模型
MAE-HR	4.67	—	—	3.78
RMSE-HR	8.91	—	—	7.53
r-HR	0.59	—	—	0.72
ACC-V	—	95.41%	—	97.47%
ACC-A	—	95.04%	—	96.75%
ACC-Sta	—	—	94.45%	97.42%

在 rPPG 信号提取任务中,多任务模型相较于只含 rPPG 信号预测头模型的 MAE-HR 与 RMSE-HR 分别降低了 0.89 与 1.38,r-HR 提高了 0.13。对于情绪预测任务,多任务模型的 ACC-V 与 ACC-A 相较于只含情绪预测头的模型分别提高了 2.06% 和 1.71%。在心理状态分类任务中,多任务模型的 ACC-Sta 相较于只含工作状态预测头的模型提升了 2.97%。实验结果充分展示了基于多任务学习的模型在所有任务中均优于仅含单任务预测头的模型,验证了多任务学习通过共享特征和协同优化有效提高模型在相关任务上的性能,特别是在情绪预测和心理状态分类任务中,任务间的信息传递进一步增强了模型的整体表现。

3.6 多任务模型与现有单任务方法对比

本节将基于多任务轮换学习方法与当前主流的单任

务方法进行对比。由于实验所涉及的状态分类任务尚无无可比的对同类模型或方法,因此仅对比 rPPG 信号提取和情绪识别两个子任务。

1) rPPG 信号提取性能对比

针对 rPPG 信号提取任务,本文与现有的多种 rPPG 提取方法进行对比。实验结果如表 3 所示,比较对象包括传统算法^[15]、监督学习算法^[16-18]以及无监督学习算法^[19]。

表 3 rPPG 信号提取性能对比

Table 3 Comparison of rPPG signal extraction performance

rPPG 提取方法	MAE-HR	RMSE-HR	r-HR
POS ^[15]	7.39	10.25	0.82
PhysFormer ^[16]	3.03	3.96	0.92
TransPhys ^[17]	6.35	10.51	0.67
FVSC-HR ^[18]	4.05	11.12	0.75
Yue ^[19]	4.79	7.64	0.82
本文	3.78	7.53	0.72

实验结果表明,本文提出的多任务模型在 MAE-HR 和 RMSE-HR 指标上优于传统方法、无监督学习算法和部分监督学习算法,说明该方法在 rPPG 信号提取准确性方面具有一定优势。然而,与当前先进的监督学习算法相比,本文模型在 RMSE-HR 和 r-HR 上的表现仍存在差距,这是因为本文通过快速傅里叶变换从预测波形中间接计算心率,损失函数仅约束时域波形误差,而采用端到端频谱监督的模型,其直接优化频域特征相关性。本文在信号稳定性和时序一致性方面仍有进一步优化空间。

本文将多任务模型提取的 rPPG 信号与真实 PPG 信号波形进行对比,效果图如图 5 所示。从图 5 中可以观察到,模型预测的 rPPG 信号在整体波形趋势、峰谷位置以及信号周期性方面,与真实的 PPG 信号较好吻合。这表明模型能够有效地捕捉 PPG 信号的生理节律。此外,预测信号的振幅波动与真实信号较为接近,表明模型不仅能准确预测心率变化趋势,还能在一定程度上保留 PPG 信号的振幅信息。这进一步说明模型具备较强的血流动力学表征能力。

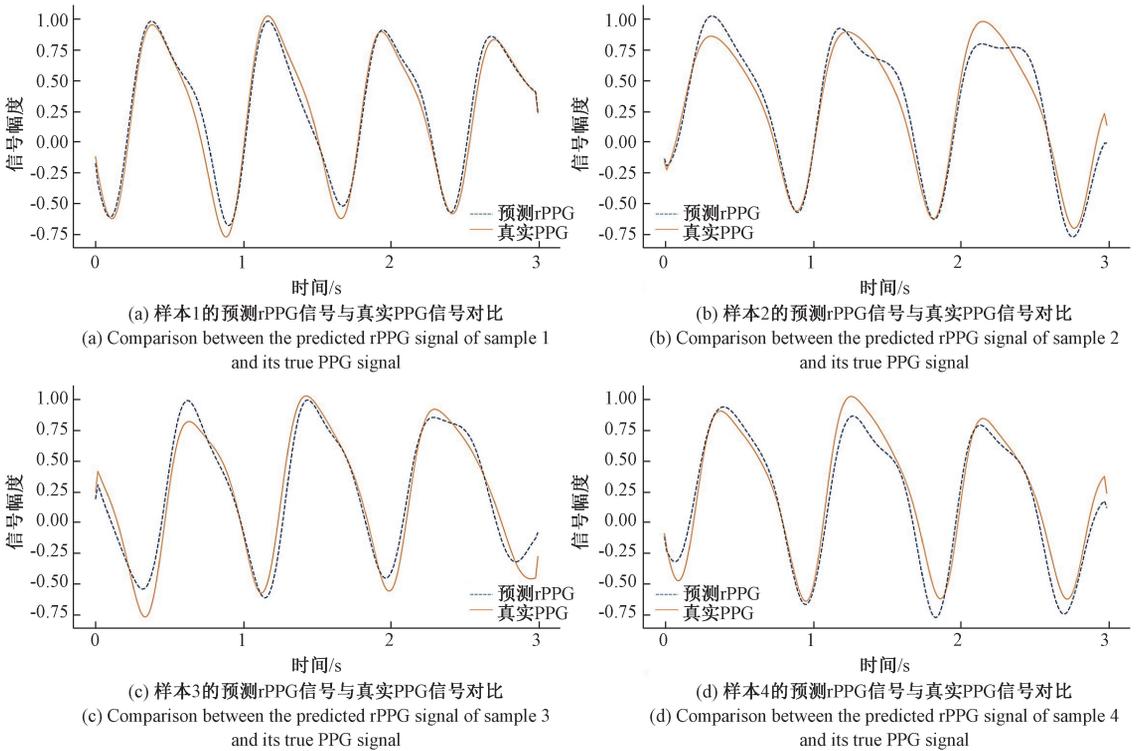


图 5 对预测的 rPPG 信号与其相应的真实 PPG 信号进行视觉比较

Fig. 5 Visual comparison between the predicted rPPG signal and its corresponding real PPG signal

2) 情绪识别性能对比

为进一步验证所提出多任务模型在情绪识别任务中的有效性,本文选取多个代表性方法进行了对比,涵盖基于脑电(electroencephalogram, EEG)^[20-21]、PPG 信号^[22-23]及人脸图像(face)^[24]3 种不同模态的情绪识别方法。各方

法在 DEAP 数据集上的效价与唤醒度预测准确率如表 4 所示。

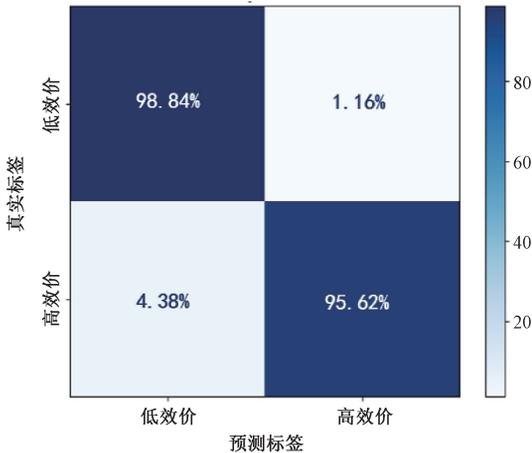
结果显示,PPG 模态的预测性能相对较低,而基于 EEG 和 face 模态则具有更高准确率。值得注意的是,文献[20]提出的多任务学习方法结合了人脸图像与 rPPG

信号,其识别率明显高于其他方法,在此基础上本文在其提出的多任务基础上增加了心理状态预测任务分支,进一步提升了模型的整体学习效果,在效价与唤醒度预测准确率上相比文献[20]提出的方法分别提升了 1.21%与 0.53%。效价、唤醒度-二分类任务的混淆矩阵如图 6(a)与(b)所示。

表 4 情绪识别性能对比

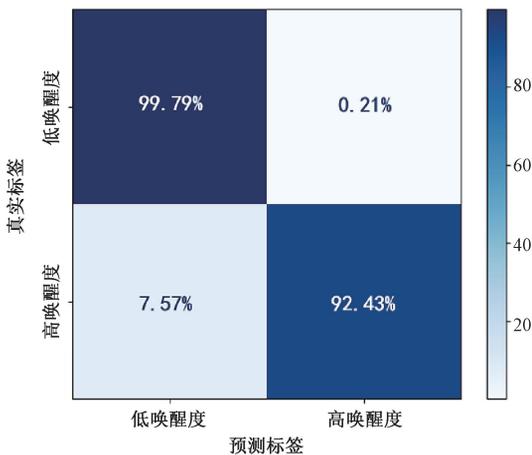
Table 4 Comparison of emotion recognition performance

情绪识别方法	模态	ACC-V/%	ACC-A/%
文献[20]	EEG	96.11	96.29
文献[21]	EEG	94.72	95.68
文献[22]	PPG	75.90	79.30
文献[23]	PPG	82.10	80.90
文献[24]	face	96.26	96.22
本文	face	97.47	96.75



(a) 效价二分类混淆矩阵

(a) Confusion matrix of binary-class valence classification



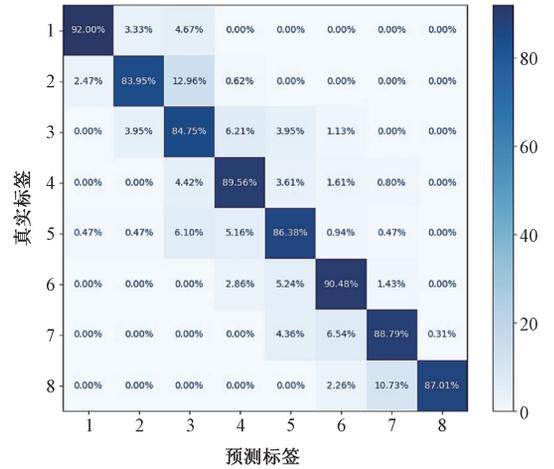
(b) 唤醒度二分类混淆矩阵

(b) Confusion matrix of binary-class arousal classification

图 6 效价、唤醒度-二分类混淆矩阵

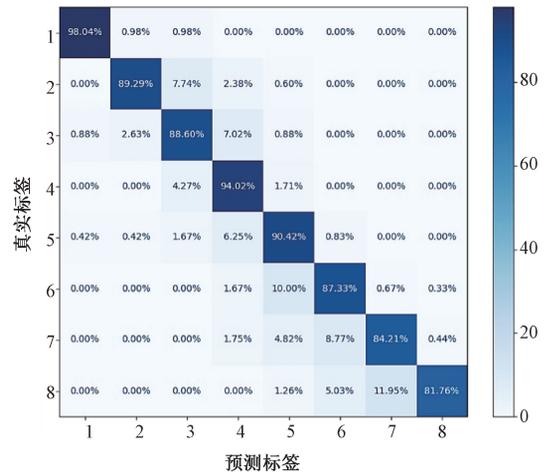
Fig. 6 Confusion matrix of binary-class valence, arousal classification

此外,根据效价与唤醒度的取值范围 [1, 9],本文额外展示了将效价、唤醒度分为八类的混淆矩阵图,如图 7(a)与(b)所示。效价与唤醒度的八分类准确率达到 88.00%、88.78%。该结果表明,本研究提出的模型不仅在二分类任务中表现优异,而且在更精细化的预测任务中也能实现较高的准确率。



(a) 效价八分类混淆矩阵

(a) Confusion matrix of eight-class valence classification



(b) 唤醒度八分类混淆矩阵

(b) Confusion matrix of eight-class arousal classification

图 7 效价、唤醒度-八分类混淆矩阵

Fig. 7 Confusion matrix of eight-class valence, arousal classification

综合以上对比结果可见,引入任务相关性不仅提升了模型的整体性能,同时也有助于捕捉情绪识别中潜在的深层特征结构,验证了本文所提出多任务模型在心理状态感知任务中的有效性和实用价值。

4 结 论

本文提出了一种基于多任务轮换学习的非接触式人体心理状态感知方法,并在 DEAP 数据集上进行了全面的实验验证。通过设计一个共享特征提取器来提取人脸视频中生理及情绪信号的深层时空特征,并通过轮换训练策

略,依次优化 rPPG 心率信号提取、情绪预测和状态分类子任务,有效减少任务间的干扰,提升了各个子任务的训练稳定性和准确性,从而提升了多任务学习模型在心理状态感知中的性能。实验结果表明:rPPG 心率信号提取达到了 3.78 的平均绝对误差与 7.53 的均方根误差;情绪代价与唤醒度预测达到了 97.47%与 96.75%的准确率;心理状态预测达到了 97.42%的准确率。

然而,本文在心理状态分类任务中仅使用了飞行员驾驶任务的标准对数据集进行标签标定,导致模型的适用范围较为局限。未来的研究可以考虑拓展心理状态的标签标定范围,探索其他职业或任务场景中的心理状态,以进一步增强模型的泛化能力和应用场景。

参考文献

- [1] 高创,张萌,韦剑道.轨道驾驶员安全心理评价方法研究[J].心理学进展,2022,12(2):352-359.
GAO CH, ZHANG M, WEI J D. Research on the evaluation method of the railroad drivers' safety psychology[J]. Advances in Psychology, 2022, 12(2): 352-359.
- [2] 薛志超,伊力哈木·亚尔买买提,闫天星.基于 MobileNetV3 多尺度特征融合的人脸表情识别[J].电子测量技术,2023,46(8):38-44.
XUE ZH CH, YILIHAMU Y, YAN T X. Facial expression recognition based on MobileNetV3 multi-scale feature fusion [J]. Electronic Measurement Technology, 2023, 46(8): 38-44.
- [3] 潘礼正,王顺超,丁忆,等.基于全视域特征表征与 ELM-Adaboost 方法的情绪辨识研究[J].仪器仪表学报,2022,43(11):162-171.
PAN L ZH, WANG SH CH, DING Y, et al. Emotion recognition research based on the full-view feature representation and ELM-Adaboost[J]. Chinese Journal of Scientific Instrument, 2022, 43(11): 162-171.
- [4] 耿梦晴,王永雄.基于 MAE 和 Mamba 的自监督预训练的心率信号测量[J/OL].信息与控制,2025,1-10 [2025-06-25], DOI: 10.13976/j.cnki.xk.2024.4971.
GENG M Q, WANG Y X. A self-supervised pre-training method for heart rate signal measurement based on MAE and Mamba[J/OL]. Information and Control, 2025, 1-10[2025-06-25], DOI: 10.13976/j.cnki.xk.2024.4971.
- [5] 成娟,武梦影,宋仁成,等.模型和盲源分离融合的运动鲁棒视频心率检测[J].仪器仪表学报,2023,44(8):289-298.
CHENG J, WU M Y, SONG R CH, et al. Motion-robust heart rate measurement from videos via combining model with blind source separation [J]. Chinese Journal of Scientific Instrument, 2023, 44(8): 289-298.
- [6] 彭军强,张立坤,杨亚楠.基于多模态轻量化混合模型的情绪识别[J].电子测量技术,2024,47(3):9-18.
PENG J Q, ZHANG L K, YANG Y N. Emotion recognition based on multi-modal lightweight hybrid model [J]. Electronic Measurement Technology, 2024, 47(3): 9-18.
- [7] 张志雯,于乃功,边琰,等.基于多模态生理信号特征融合的情绪识别方法研究[J].生物医学工程学报,2025,42(1):17-23.
ZHANG ZH W, YU N G, BIAN Y, et al. Research on emotion recognition methods based on multi-modal physiological signal feature fusion [J]. Journal of Biomedical Engineering, 2025, 42(1): 17-23.
- [8] 李奇,常立娜,武岩,等.基于深层图卷积的 EEG 情绪识别方法研究[J].电子测量技术,2024,47(4):18-22.
LI Q, CHANG L N, WU Y, et al. Research on EEG emotion recognition method based on deep graph convolutional [J]. Electronic Measurement Technology, 2024, 47(4): 18-22.
- [9] 王宇,刘宇昂,赵梦洁,等.融合 rPPG 和人脸三维法向量的非接触情绪感知技术[J].电讯技术,2024,64(10):1667-1676.
WANG Y, LIU Y ANG, ZHAO M J, et al. Non-contact emotion recognition technology integrating rPPG and 3D facial normal vectors [J]. Telecommunication Engineering, 2024, 64(10): 1667-1676.
- [10] MAI S J, HU H F, XU J, et al. Multi-fusion residual memory network for multimodal human sentiment comprehension[J]. IEEE Transactions on Affective Computing, 2020, 13(1): 320-334.
- [11] 谢平,胡锦涛,江国乾,等.基于多任务学习的快速序列视觉呈现脑电图分类[J].仪器仪表学报,2023,44(11):215-223.
XIE P, HU J CH, JIANG G Q, et al. Classification of rapid serial visual presentation based EEG with multi-task learning [J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 215-223.
- [12] KENDALL A, GAL Y, CIPOLLA R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7482-7491.
- [13] 宋建辉,刘鑫,庄爽,等.面向无人驾驶的多任务环境感知算法研究[J].电子测量与仪器学报,2025,39

- (1): 122-132.
- SONG J H, LIU X, ZHUANG SH, et al. Research on multi-task environment perception algorithm for unmanned driving [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(1): 122-132.
- [14] LIU X, NARAYANSWAMY G, PARUCHURI A, et al. Rppg-toolbox: Deep remote ppg toolbox[J]. Advances in Neural Information Processing Systems, 2023, 36: 68485-68510.
- [15] WANG W J, DEN BRINKER A C, STUIJK S, et al. Algorithmic principles of remote PPG [J]. IEEE Transactions on Biomedical Engineering, 2016, 64 (7): 1479-1491.
- [16] YU Z T, SHEN Y M, SHI J G, et al. Physformer: Facial video-based physiological measurement with temporal difference transformer [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 4186-4196.
- [17] WANG R X, SUN H M, HAO R R, et al. TransPhys: Transformer-based unsupervised contrastive learning for remote heart rate measurement [J]. Biomedical Signal Processing and Control, 2023, 86: 105058.
- [18] 王猛, 杨观赐. 基于可分离卷积的人脸视频心率检测方法[J]. 软件导刊, 2025, 24(8): 38-42.
- WANG M, YANG G C. Separable convolutional-based heart rate detection method for face video[J]. Software Guide, 2025, 24(8): 38-42.
- [19] YUE Z J, SHI M J, DING S. Facial video-based remote physiological measurement via self-supervised learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45 (11): 13844-13859.
- [20] LI ZH P, MA W ZH, ZHANG J S, et al. DSRP: A database for stress reduction using physiological signals[J]. IEEE Access, 2024, 12: 135089-135102.
- [21] KAN H N, YU J L, HUANG J J, et al. Self-supervised group meiosis contrastive learning for eeg-based emotion recognition[J]. Applied Intelligence, 2023, 53(22): 27207-27225.
- [22] ZHU ZH B, WANG X Y, XU Y F, et al. An emotion recognition method based on frequency-domain features of PPG[J]. Frontiers in Physiology, 2025, 16: 1486763.
- [23] LEE M S, LEE Y K, LIM M T, et al. Emotion recognition using convolutional neural network with selected statistical photoplethysmogram features [J]. Applied Sciences, 2020, 10(10): 3501.
- [24] JI Y, DONG S Y. Multi-task learning by leveraging non-contact heart rate for robust facial emotion recognition[J]. IEEE Access, 2024, 12: 92175-92180.

作者简介

王宇, 博士, 副教授, 主要研究方向为图像处理、深度学习、人工智能等。

E-mail: scuwangyu@163.com

牛知艺(通信作者), 硕士研究生, 主要研究方向为图像处理、非接触式信号测量、情绪识别。

E-mail: 861719782@qq.com