

# 题目

## 摘要

客观准确的情绪诊断参考对心理学家来说至关重要，尤其是在与因病理原因难以沟通的患者打交道时。然而，当前基于脑电图（Electroencephalogram, EEG）数据用于情感辨别的系统存在一些问题，包括模型复杂性过高、准确性一般和可解释性有限。因此，论文提出了一种新颖而有效的特征融合机制，称为相互交叉注意（Mutual-Cross-Attention, MCA）。结合专门定制的 3D 卷积神经网络（3D-CNN），这种纯数学机制巧妙地发现了 EEG 数据中时域和频域特征之间的互补关系。此外，新设计的 Channel-PSD-DE 3D 功能也有助于实现高性能。所提出的方法最终在 DEAP 数据集上实现了 99.49%（效价）和 99.30%（唤醒）的准确性。本文提出对经过 MCA 融合的特征应用 PCA（Principal Component Analysis, PCA）降维并结合机器学习分类器的改进方法，该方法在 DEAP 数据集上表现更优泛化性能。

**关键词：**情绪识别；注意特征融合；3D-卷积神经网络；脑电图特征

## 1 引言

抑郁症作为一种深刻影响个体身心健康的精神疾病，其对生活质量与社会功能的侵蚀往往具有持续且不可逆转的效应。倘若未能及时干预，个体可能遭遇情感、认知以及社交功能等多维度的严重紊乱，最终导致无法恢复的精神及生活失调。传统的心理评估手段在面对这一复杂现象时，常常面临重重障碍——语言表达不清、情绪波动显著、反刍思维困扰、字面理解的困难及沟通障碍等因素使得精神障碍的诊断充满挑战。因此，借助更为精确且客观的评估工具显得尤为迫切。脑电图（EEG）作为一种反映脑电活动动态变化的生理信号，逐渐成为心理学家与精神健康专家在情感及认知状态评估中的重要工具。EEG 能够揭示大脑在不同情感与认知状态下的电活动模式，这为情感状态的准确识别提供了坚实依据。通过深入分析这些信号所蕴含的情感信息，心理治疗师能够迅速识别患者的情绪状态，并在此基础上精确构建个性化治疗方案，显著提升治疗效果及患者生活质量。

近些年来，情感识别领域的学术研究已取得显著进展，探索并应用了多种先进技术。初期的研究主要依赖于传统的 EEG 特征，如差分熵（DE）和功率谱密度（PSD）[6]，然而，单一特征往往难以全面揭示 EEG 信号的复杂性。因此，近年的研究已转向特征融合方法，通过将多种 EEG 特征有机结合，力求更加全面地捕捉大脑情感活动的全貌，从而提升情感识别的精度。随着深度学习技术的快速发展，特征融合与深度神经网络的结合愈加频繁，尤其是 3D 卷积神经网络（3D-CNN）[6] 在此领域的应用，使情感识别的准确性和效率得到了显著

提高。EEG 信号具有独特的时间-频率-空间三维结构，而 3D-CNN 能够充分挖掘其时空关联性，从而有效提升分类精度与模型的稳定性。

为实现基于 EEG 分析的即时高效情感判断系统，论文提出了一种创新的解决方案，并通过实验验证，证明其为目前领域中的先进方法（SOTA）。其主要贡献包括：

1) 互交注意力机制：提出了一种名为“互交注意力”（MCA）的纯数学特征融合方法。该方法通过在特征的每个维度上应用注意力机制，提升了情感分析的精度。通过互交注意力机制，模型能更加高效地集中于 EEG 信号中的关键特征，从而提升情感识别的精确度。

2) 创新的 3D 特征结构：通过对现有研究成果的深入分析，光谱信息相较于空间信息（通过通道拓扑呈现）在情感识别中具有更为显著的作用。因此，论文提出了一种新的通道-频率-时间（CFT）3D 特征结构。该结构能够有效融合光谱与时间信息，进一步提高情感识别的准确性。结合 3D-CNN 网络，这一创新特征结构充分利用了其强大的时空特征学习能力，能够更准确地捕捉 EEG 信号中的复杂情感模式，从而显著提升情感分类的效果。

3) 高准确率的情绪识别模型：提出的方法最终在情绪识别开源数据集 DEAP 数据集上实现了 99.49%（效价）和 99.30%（唤醒）的准确性。

## 2 相关工作

### 2.1 特征融合

近年来，学术界在情绪识别方面取得了一些进展。最初，研究重点是单一的传统脑电图特征，如差分熵（DE）和功率谱密度（PSD）[6]。随后，人们采用了特征融合和深度学习的方法来提高识别准确率。Amin 等人 [1] 提出了一种多层 CNNs 方法，用于融合具有不同特征和结构的 CNN，以提高 EEG 运动想象分类的准确性。随着这些新技术的应用，分类器的性能得到了提高，但仍存在一些问题。目前，主流的融合方法是采用不同的可学习模型来提取特征映射，然后直接进行串联 [4]。这些策略训练网络自主集中处理信号的重要方面，加重了模型训练的负担，降低了模型输出的效率。考虑到情感分类系统在实际应用中需要即时输出，目前日益复杂的神经网络并无益处。

### 2.2 3D-CNN

利用 3D-CNN 进行序列建模已在视频分析（如动作识别）中得到广泛探索 [3]。受这些研究的启发，3D-CNN 也被引入了这一领域。Salama 等人 [8] 提出了一种多通道 EEG 随机排列成帧的 3D-CNN 模型。连续的帧被进一步串联成一个 3D 立方体。Wang 等人 [11] 也提出了一种名为 Emo-tioNet 的三维 CNN 模型，该模型集成了批量归一化和密集预测机制，以解决协方差偏移和地面实况标签不可靠的问题。将二维帧（通道  $\times$  时间）转换为三维立方体（通道拓扑结构  $\times$  时间），作为模型的输入。前两层采用三维卷积来学习。先学习空间和时间特征，然后进行融合操作，将这些学习到的高层表征融合在一起。因此，这一层的输出只有时间特征，这些特征被输入到下面两层进行高级时间表征学习。在模型的最后，利用密集预测来进行时变情绪状态预测。

## 3 本文方法

### 3.1 本文方法概述

论文提出了两种主要贡献：首先，设计了一种名为相互交叉注意力机制（Mutual-Cross-Attention, MCA）的特征融合方法，结合定制的 3D 卷积神经网络（3D-CNN），有效发现 EEG 数据中时域和频域特征的互补关系。其次，开发了一种新的 3D 特征表示 Channel-PSD-DE，该特征同时呈现谱信息和时序信息，以提高特征表示的丰富性。在方法上，论文设计了一个完整的实验流程，包括数据采集、预处理、特征提取、特征融合和分类。使用 DEAP 数据集进行实验，通过预处理步骤提高数据质量，然后从五个不同的频带中提取差分熵（DE）和功率谱密度（PSD）特征。在特征融合阶段，应用 MCA 机制在每个频带中融合 DE 和 PSD 特征，最后通过定制的 3D-CNN 进行情感预测。

### 3.2 数据采集

本文使用了 DEAP 数据集 [5]，该数据集包含 32 名被试的 EEG 信号及其他生理信号（如眼动、电皮肤反应等）。每位被试在观看 40 个 1 分钟的音乐视频时，根据情绪价值（Valence）和唤醒度（Arousal）对每个视频进行评分。EEG 信号通过 32 个通道采集，采样频率为 512 Hz，记录了被试在观看视频时的脑电活动。视频内容包括不同类型的音乐视频，旨在激发各种情感反应。每个视频播放后，受试者需要根据其观看体验对情绪的情感价值和唤醒度进行评分，评分范围从 1 到 9，涵盖了从最消极到最积极的情绪状态。

### 3.3 数据预处理模块

数据集的预处理采用 Steve Luck 的经典 EEG 信号预处理步骤 [7]，通过过滤信号和去除噪声成分进行清理。首先，应用陷波滤波器（Notch Filter）去除 50 Hz 频率的信号，这通常与交流电源干扰相关。随后，为了减少测量工具的不准确性和环境干扰，设定了 4–45 Hz 的带通滤波器。之后，使用独立成分分析（ICA）对过滤后的 EEG 数据进行处理，以去除眼电图、心电图和肌电图等噪声成分。最后，将原始的 512 Hz 数据下采样至 128 Hz，以减少数据量并加快计算速度。

### 3.4 特征提取模块

功率谱密度、差分熵、功率谱密度不对称、差分熵不对称作为 EEG 信号中常见的特征。本文从 EEG 信号中提取了差分熵（DE）和功率谱密度（PSD）两类特征，并分别在五个频段（ $\theta$ 、 $\alpha$ 、低  $\alpha$ 、 $\beta$ 、 $\gamma$ ）上进行提取，以确保特征的多样性。其中， $\theta$  频段（4 - 7 Hz）与情绪调节有关，适当的  $\theta$  波活动有助于缓解紧张情绪，能使人进入放松和平静的状态。 $\alpha$  频段（8 - 13 Hz）代表大脑处于相对放松但仍保持警觉的状态，当人处于安静、清醒且闭眼放松状态时容易出现。它的变化可以反映注意力集中程度，而注意力的变化也会对情绪产生间接影响，比如注意力集中可能抑制焦虑等情绪。低  $\alpha$  频段（8 - 10 Hz）可能在大脑的状态转换中起作用，从活跃到安静或相反过程中会变化，与大脑内部信息整合有关，这种整合可能会影响情绪反应。 $\beta$  频段（14 - 30 Hz）与人处于清醒、警觉、思考或者紧张状态有关。在紧张思考或高度集中注意力时出现，这种状态下的情绪可能是紧张、焦虑或者兴奋的。 $\gamma$  频段（30 Hz 以

上) 出现在大脑高度活跃状态, 如复杂认知任务、学习新知识或多感官信息整合时。可能在构建复杂认知结构和意识体验方面起关键作用, 这种复杂的认知活动可能引发好奇、兴奋等积极情绪, 也可能因任务困难产生焦虑等消极情绪。

功率谱密度 (PSD) 提取: 采用 Welch 方法 [9] 计算 PSD, 通过快速傅里叶变换 (FFT) 将信号转换为频域信息, 帮助捕捉 EEG 信号中的频率特征。

差分熵 (DE) 提取: 采用高斯分布假设 [10], 计算每个频段内的 DE 值。DE 可以反映信号的复杂性, 从而为情感识别提供有价值的特征。

### 3.5 特征融合模块

在特征融合阶段, 应用一种基于互交注意力机制 (MCA) 的融合方法融合差分熵 (DE) 与功率谱密度 (PSD) 特征。MCA 结构如图 1 所示。MCA 机制应用于每个选定的频带以融合 DE 和 PSD。MCA 机制通过将 DE 和 PSD 视为键 (key) 和查询 (query) 向量, PSD 指定为值 (value) 并执行基本的缩放点积注意力 (Scaled Dot - Product Attention) [2] 操作, 公式如下:

$$Atten(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (1)$$

在 MCA 另一个方向, 反过来将 PSD 用作 Q, DE 用作 K 和 V, 再次执行缩放点积注意力操作。再将两个方向的结果相加得到新特征, 得到一个新的融合特征表示。整个过程可以用数学公式表示为:

$$MCA(f_1, f_2) = Atten(f_1, f_2, f_2) + Atten(f_2, f_1, f_1) \quad (2)$$

执行特征融合操作后, 单一特征表示包含了 60 秒 EEG 信号的特征, 这不利于后续任务的执行, 执行分类任务所需的时间太长。因此, 单一 EEG 特征表示被分割成 20 个 3 秒的数据这一操作使后续任务模型能够每 3 秒输出一次情感预测。

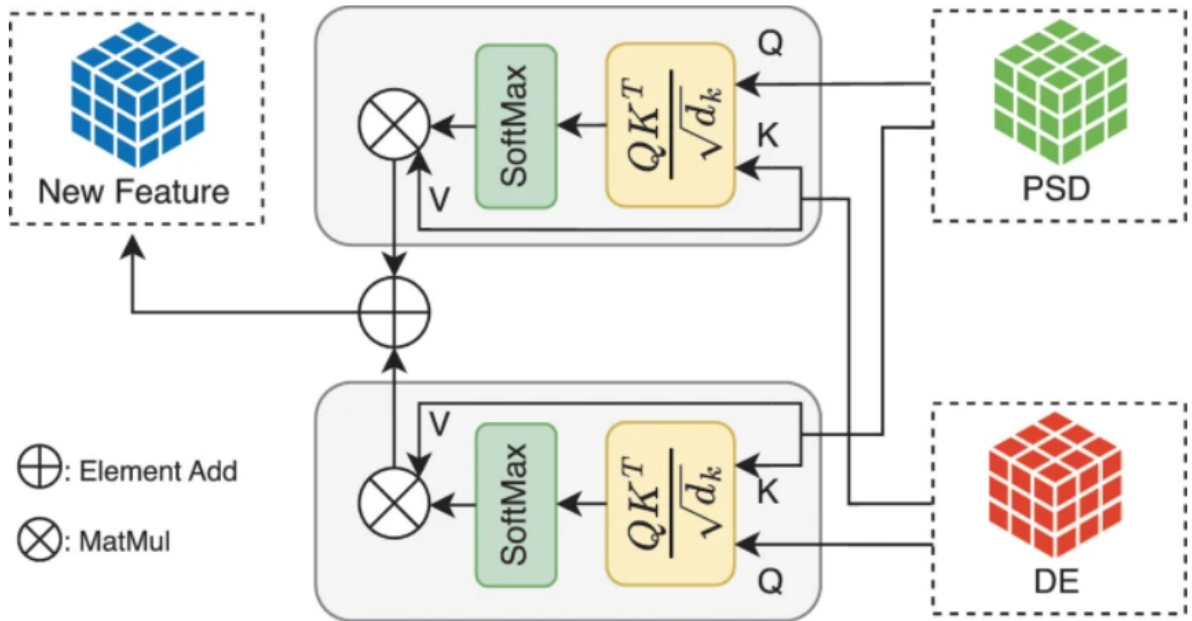


图 1. MCA 结构 [12]



### 3.6 分类器定义

原文分类器模块基于 3D 卷积神经网络 (3D-CNN) 架构,旨在通过多阶段的特征提取和层级性特征压缩,在情感分类任务中实现深度抽象和时效性的双重优化。网络首先通过两层堆叠的 3D 卷积操作,逐步从输入数据中提取局部和全局特征,避免过早的维度压缩,从而保持更丰富的空间信息。卷积层通过增加深度和输出通道数,有效增强了网络的特征学习能力。在特征提取后,Max Pooling 层通过对特征图的下采样,减少了冗余信息,同时保留了关键信息。这一过程通过不对称填充和池化策略,确保了特征的高效压缩与表示的稳定性。网络随后将输出通道数扩展至 64,进一步加深了对复杂模式的理解,为分类任务提供了更加丰富的特征表示。为了提高实时性,特征被切分成 20 个子特征,允许模型每 3 秒输出一次情感预测。最终,通过全连接层对提取的特征进行分类,输出情感预测结果。这一架构的设计不仅提升了分类精度,也优化了计算效率,适应了动态情感识别等需要快速响应的应用场景。

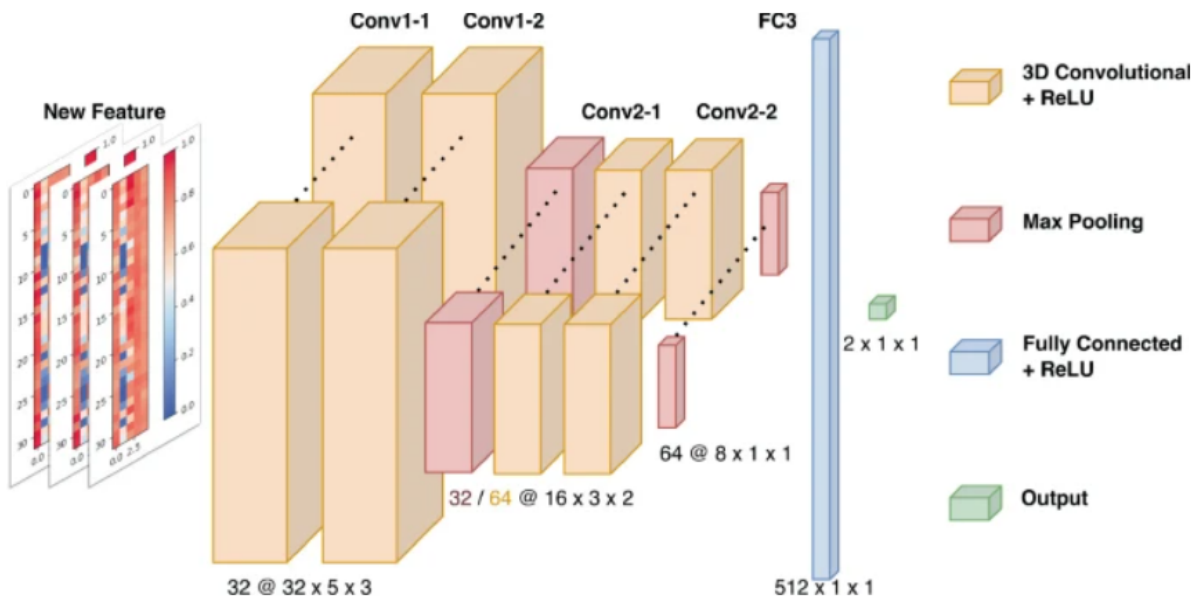


图 2. 3D-CNN 网络结构 [12]

## 4 复现细节

### 4.1 与已有开源代码对比

原论文已有开源代码,对源码进行了复现,并对开源代码进行了改进:

改进一: 本文在原有方法的基础上进行了改进,通过简化原文中复杂的 3D-CNN 网络结构,改用一种更为高效的方法(本文称之为 MCA-PCA-ML)。具体而言,我们将融合特征首先通过主成分分析(PCA)进行降维,之后将降维后的特征输入到机器学习分类器进行处理。相比于原文中复杂的 3D-CNN 网络,这一改动有效简化了模型结构,减少了计算负担,同时通过 PCA 保留了数据中的主要成分,剔除了冗余信息,从而在提高训练效率的同时,优化了分类器的性能。该方法不仅加快了处理速度,还增强了分类器对核心模式的提取能力,提升了模型的精度和泛化能力。

改进二: 在实验设计方面,本文深入扩展了实验的广度和深度。原论文仅将数据集按照

8:2 比例划分训练集、测试集。本文对原论文模型开展 10 折交叉验证实验，对模型进行了更加全面和系统的评估。与传统的简单训练-测试分割不同，10 折交叉验证通过将数据集划分为 10 个子集，分别进行训练和验证，能够有效评估模型在不同数据切分下的稳定性和泛化能力。这种方法不仅增加了实验的可靠性，还能够降低因数据分割不均导致的偶然误差，确保模型性能评估的全面性。通过这种细致的实验验证，论文展示了模型在不同数据集和环境下的鲁棒性，进一步提升了研究的可信度与实用价值。

## 4.2 实验环境搭建

实验环境搭建基于 NVIDIA A6000 显卡，以确保高效的计算能力和大规模数据处理。硬件配置包括强大的 GPU 支持，以加速深度学习模型的训练和推理过程，尤其适用于高计算需求的任务。软件环境方面，采用了 PyTorch 深度学习框架，因其灵活性和高效性，特别适合于构建和训练复杂的神经网络模型。PyTorch 支持 CUDA 加速，能够充分发挥 A6000 显卡的计算潜力，从而提高模型训练的速度和精度。此外，配套的 Python 开发环境以及相关依赖库（如 NumPy、Pandas、Matplotlib 等）也已经配置完毕，确保实验过程中的数据处理、可视化及其他辅助功能的顺利进行。

## 4.3 创新点

本文提出了将 MCA 融合特征经过主成分分析（PCA）降维后，再输入机器学习分类器（随机森林分类器）进行分类的创新方法。特征融合通常会产生高维数据，可能导致计算效率低下以及模型的过拟合风险。为了解决这一问题，我们引入了 PCA 作为一种降维技术，利用其通过线性变换保留数据中的主要成分，去除冗余和不相关的特征，从而减少特征空间的维度。在降维过程中，PCA 能够最大限度地保留数据的原始信息和结构，确保关键信息的传递，同时降低了计算复杂度，提升了模型训练的效率。与传统的直接输入融合特征到分类器的做法不同，PCA 降维后的特征更加精炼，能够使机器学习分类器更专注于数据中的核心模式，有助于提高分类精度。更重要的是，PCA 降维还有效缓解了高维数据可能带来的“维度灾难”，并增强了模型的泛化能力。此外，PCA 与分类器的组合方式，突破了单纯依赖分类器学习复杂特征的局限性，为多维特征的处理提供了更高效的方案，特别是在处理大规模数据集和高维特征时，这一策略展现了明显的优势。

## 5 实验结果分析

为了验证所提方法的有效性，我们使用 DEAP 数据集进行实验。DEAP 数据集中情绪效价和唤醒标签为 1-9 数值，本文将效价值 1-5 划分为低效价，效价值 6-9 划分为高效价，将唤醒值 1-5 划分为低唤醒，唤醒值 6-9 划分为高唤醒。PCA 降维设置选择保留 95% 方差的主成分。数据集分别按照 8:2 比例划分训练集、测试集的范式及 10 折交叉验证的范式下构建效价分类模型和唤醒分类模型。在二分类问题中，可以将样例根据其真实类别和预测类别的组合划分为：真正例（true positive）、假正例（false positive）、真反例（true negative）、假反例（false negative），令 TP、FP、TN、FN 分别表示其对应的样例数。模型采用常见的模型评价指标准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F1 分数（F1-score），指标

公式如下：

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{T}{TP + FN} \quad (5)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (6)$$

PCA 降维使得特征矩阵降维到 162 个特征。数据集分别按照 8:2 比例划分训练集、测试集的范式及 10 折交叉验证的范式下构建模型，结果如表 1、2 所示。

表 1. 8:2 划分训练集、测试集范式下实验结果

	MCA-3dCNN(原文)		MCA-3dCNN(复现)		MCA-PCA-ML(改进)	
	Valence	Arousal	Valence	Arousal	Valence	Arousal
Accuracy (%)	99.49	99.30	98.79	98.98	98.52	99.36
Precision (%)	99.60	99.45	98.87	99.20	98.60	99.01
Recall (%)	99.54	99.31	99.07	99.00	99.93	99.86
F1-score (%)	99.57	99.38	98.97	99.10	98.75	99.43

表 2. 10 折交叉验证范式下实验结果

	MCA-3dCNN(原文)		MCA-3dCNN(复现)		MCA-PCA-ML(改进)	
	Valence	Arousal	Valence	Arousal	Valence	Arousal
Accuracy (%)	-	-	66.36	65.73	67.88	66.62
Precision (%)	-	-	67.55	67.59	68.47	68.02
Recall (%)	-	-	82.57	75.77	84.29	77.37
F1-score (%)	-	-	74.30	71.43	75.55	72.39

将 8:2 比例划分训练集、测试集的范式下实验结果与先进模型对比，MCA-3dCNN 复现方法及 MCA-PCA-ML 改进方法在四个指标中均表现较为出色，准确率均达到了 98%。MCA-3dCNN 与 MCA-PCA-ML 均采用 MCA 的特征融合方法，不同之处在于 MCA-3dCNN 的分类器是采用了一个 3dCNN 去捕捉频谱和时间信息。MCA-PCA-ML 则弃用复杂的神经网络结构，选用更为轻便的机器学习随机森林模型作为分类器。MCA-PCA-ML 和 MCA-3dCNN 的性能指标均达到 98% 以上，根据实验结果可知，MCA-PCA 的性能逼近 MCA-3dCNN 模型。同时，也可知 MCA 的特征融合方法能够较好地捕捉并融合 EEG 特征。

## 6 总结与展望

本文主要研究了基于脑电图 (EEG) 的情感识别方法。论文提出了一种名为相互交叉注意力机制 (MCA) 的特征融合方法，并结合定制的 3D 卷积神经网络 (3D-CNN) 来提高情

感识别的准确性和效率。具体来说，论文的主要贡献包括：1) 特征融合机制：通过 MCA 机制，有效地融合了 EEG 数据中的时域和频域特征，提升了特征的互补性和信息的丰富性。2) 创新的 3D 特征结构：开发了 Channel-PSD-DE 3D 特征表示，同时呈现谱信息和时序信息，进一步提高了特征表示的准确性。3) 高准确率的情感识别模型：在 DEAP 数据集上实现了 99.49%（效价）和 99.30%（唤醒）的准确性。

尽管论文取得了显著的成果，但在实现过程中仍存在一些不足之处：1) 模型复杂度。虽然 3D-CNN 在特征提取方面表现出色，但其复杂的网络结构可能导致计算资源消耗较大，尤其是在大规模数据集上进行训练时，可能会面临计算效率和资源限制的问题。2) 数据依赖性。模型的性能在很大程度上依赖于 DEAP 数据集的特征分布和标注质量。对于其他类型的数据集或不同的情感标注方式，模型的泛化能力可能需要进一步验证和调整。3) 可解释性。尽管 MCA 机制在特征融合方面表现出色，但其内部工作机制的可解释性仍有待提高。对于心理学家和临床医生来说，理解模型如何从 EEG 数据中提取和融合特征以做出情感判断是非常重要的。

为了进一步提升情感识别系统的性能和实用性，未来的研究可以考虑以下几个方向：1) 探索更轻量级的网络结构或算法，以降低计算复杂度和资源消耗，同时保持或提高模型的准确性和稳定性。2) 在更多的数据集上进行实验验证，以评估模型的泛化能力和鲁棒性。3) 加强对 MCA 机制和 3D-CNN 模型的可解释性研究，开发可视化工具或解释方法，帮助用户理解模型的决策过程和特征贡献。

## 参考文献

- [1] Syed Umar Amin, Mansour Alsulaiman, Ghulam Muhammad, Mohamed Amine Mekhtiche, and M Shamim Hossain. Deep learning for eeg motor imagery classification based on multi-layer cnns feature fusion. *Future Generation computer systems*, 101:542–554, 2019.
- [2] Vaswani Ashish. Attention is all you need. *Advances in neural information processing systems*, 30:I, 2017.
- [3] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):221–231, 2012.
- [4] Ziyu Jia, Youfang Lin, Xiyang Cai, Haobin Chen, Haijun Gou, and Jing Wang. Sst-emotionnet: Spatial-spectral-temporal based attention 3d dense network for eeg emotion recognition. In *Proceedings of the 28th ACM international conference on multimedia*, pages 2909–2917, 2020.
- [5] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.



- [6] Xiang Li, Yazhou Zhang, Prayag Tiwari, Dawei Song, Bin Hu, Meihong Yang, Zhigang Zhao, Neeraj Kumar, and Pekka Marttinen. Eeg based emotion recognition: A tutorial and review. *ACM Computing Surveys*, 55(4):1–57, 2022.
- [7] Steven J Luck. *An introduction to the event-related potential technique*. MIT press, 2014.
- [8] Elham S Salama, Reda A El-Khoribi, Mahmoud E Shoman, and Mohamed A Wahby Shalaby. Eeg-based emotion recognition using 3d convolutional neural networks. *International Journal of Advanced Computer Science and Applications*, 9(8), 2018.
- [9] Li-Chen Shi, Ying-Ying Jiao, and Bao-Liang Lu. Differential entropy feature for eeg-based vigilance estimation. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6627–6630. IEEE, 2013.
- [10] Otis M Solomon Jr. Psd computations using welch ’ s method. *NASA STI/Recon Technical Report N*, 92:23584, 1991.
- [11] Yi Wang, Zhiyi Huang, Brendan McCane, and Phoebe Neo. Emotionet: A 3-d convolutional neural network for eeg-based emotion recognition. In *2018 international joint conference on neural networks (IJCNN)*, pages 1–7. IEEE, 2018.
- [12] Yimin Zhao and Jin Gu. Feature fusion based on mutual-cross-attention mechanism for eeg emotion recognition. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 276–285. Springer, 2024.