

# 基于多域信息融合的卷积 Transformer 脑电情感识别模型

张学军<sup>1,2</sup>, 王天晨<sup>1</sup>, 王泽田<sup>1</sup>

(1. 南京邮电大学电子与光学工程学院、柔性电子(未来技术)学院, 南京 210023; 2. 南京邮电大学射频集成与微组装技术国家地方联合工程实验室, 南京 210023)

**摘要:** 当前脑电信号的情感识别方法很少融合空间、时间和频率信息, 并且大多数识别方法只能提取局部的脑电特征, 在全局信息关联方面存在着局限性。本文提出了一种基于多域信息融合的三维特征卷积神经网络 Transformer 机制(3D-CNN-Transformer mechanism, 3D-CTM)模型的脑电情感识别方法。该方法首先根据脑电信号的特性设计了一种三维特征结构, 同时融合脑电信号的空间、时间以及频率信息; 然后采用卷积神经网络模块学习多域信息融合的深层特征, 再连接 Transformer 自注意力模块, 提取特征信息内的全局关联性; 最后利用全局平均池化整合特征信息进行分类。实验结果表明, 3D-CTM 模型在 SEED 数据集上的三分类平均准确率达到 96.36%, 在 SEED-IV 数据集上的四分类平均准确率达到 87.44%, 有效地提高了情感识别精度。

**关键词:** 脑电信号; 情感识别; 卷积神经网络; Transformer; 自注意力

**中图分类号:** TP391 **文献标志码:** A

## Convolutional Transformer EEG Emotion Recognition Model Based on Multi-domain Information Fusion

ZHANG Xuejun<sup>1,2</sup>, WANG Tianchen<sup>1</sup>, WANG Zetian<sup>1</sup>

(1. College of Electronic and Optical Engineering & College of Flexible Electronics (Future Technology), Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 2. Nation-Local Joint Project Engineering Lab of RF Integration & Micropackage, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

**Abstract:** Current emotion recognition methods for electroencephalogram (EEG) signals seldom fuse spatial, temporal and frequency information, and most methods can only extract local EEG features, resulting in limitations in global information correlation. The article proposes an EEG emotion recognition method based on 3D-CNN-Transformer mechanism (3D-CTM) model with multi-domain information fusion. The method first designs a 3D feature structure based on the characteristics of EEG signals, simultaneously fusing the spatial, temporal, and frequency information of EEG signals. Then a convolutional neural network module is used to learn the deep features for multi-domain information fusion, and then the Transformer self-attention module is connected to extract the global correlations within the feature information. Finally, the global average pooling is used to integrate the feature information for classification. Experimental results show that the 3D-CTM model achieves an average accuracy of 96.36% in the SEED dataset for triple classification and 87.44% in the SEED-IV dataset for quadruple

classification, which effectively improves the emotion recognition accuracy.

**Key words:** electroencephalogram (EEG); emotion recognition; convolutional neural network; Transformer; self-attention

## 引言

情感是一种复杂的生理和心理状态,尤其在人与人之间的交往中扮演着重要的角色。情感计算是一个快速兴起的交叉前沿领域,主要目标是让机器拥有模拟人类同理心的能力<sup>[1]</sup>。情感识别是情感计算的一个至关重要的部分,通过对采集到的情感信息进行分析处理和建模,最终输出情感的状态,从而使脑机交互系统更加灵敏地捕捉用户的内在需求和意图,为开发基于脑机接口的智能辅助技术提供重要支持<sup>[2]</sup>。此外,情感识别技术还具有广泛的应用前景,可以拓展至医疗健康、远程教育等领域,帮助诊断心理健康问题、提高教育效果,推动相关领域的智能化发展。近年来,情感识别因其重要意义和广泛应用而成为人工智能、神经科学和模式识别等领域的热门研究课题。以往研究大多针对面部表情<sup>[3]</sup>、语音信号<sup>[4]</sup>和身体姿势<sup>[5]</sup>等非生理因素进行研究,但其可靠性难以保证。而基于脑电信号(Electroencephalogram, EEG)的非侵入式生理电信号是自发的、主观不可控制的,能够更准确地反映真实的情感状态。

基于EEG情感识别方法通常分为特征提取和分类两个步骤,如何获取有鉴别性的特征是EEG情感识别领域中的关键部分<sup>[6]</sup>。典型的特征提取方法主要从时域、频域和空间域进行分析。Feradov等<sup>[7]</sup>研究了10个时域脑电特征对消极情感的适用性,表明归一化信号的一阶差分特征更有利于识别消极情感;Duan等<sup>[8]</sup>提取了多个频带的微分熵(Differential entropy, DE)和功率谱密度(Power spectral density, PSD)等特征分别用于情感识别,发现微分熵特征的效果最好;Li等<sup>[9]</sup>根据脑电电极通道的空间位置关系,将不同通道的脑电特征映射成二维空间矩阵进行训练分类。然而,以上方法都是从单一域特征进行分析,不足以包含EEG中蕴含的丰富信息。

随着深度学习技术的发展,尤其是卷积神经网络(Convolutional neural network, CNN)和循环神经网络(Recurrent neural network, RNN)已普遍应用于EEG分类的各个领域,因此如何将多域特征信息融合逐渐成为EEG特征组织方面的研究热点。Wang等<sup>[10]</sup>提出了一种三维卷积神经网络(EmotionNet),采用3D卷积核从原始EEG中同时提取空间和时间信息;Yang等<sup>[11]</sup>提出了一种结合不同频带信号特征,同时保留通道间空间信息的三维图像表示方法,采用连续卷积神经网络(Continuous CNN, CCNN)来挖掘EEG的空间和频率信息;Wang等<sup>[12]</sup>通过短时傅里叶变换提取EEG的时频特征并构建电极频率分布图,采用基于残差块的深度卷积神经网络进行特征学习;Wu等<sup>[13]</sup>提出了双分支卷积神经网络的多尺度双半球非对称模型(Multi-scale bi-hemispheric asymmetric model, MSBAM),在两个独立的分支中提取不同的时间表征和非对称空间信息;Cui等<sup>[14]</sup>提出了一种区域-非对称卷积神经网络(Regional-asymmetric CNN, RACNN),由时间、区域和非对称特征提取器组成,以此来学习EEG的时间以及空间的局部和非对称信息;Yang等<sup>[15]</sup>提出了一种并行卷积循环神经网络(Parallel convolutional RNN, PCRNN),利用CNN模块将多通道EEG转换为二维矩阵来挖掘相邻通道之间的空间位置信息,利用长短时记忆网络(Long short-term memory, LSTM)模块来挖掘时间信息。虽然以上方法取得了不错的效果,但仍面临着一些挑战:一是现有的多域信息融合方法大多数只考虑了两种特征信息的组合(如空间-时间组合),而忽略了空间-时间-频率信息之间的互补性,在一定程度上限制了模型的分类能力;二是大多方法都是基于CNNs和RNNs的,存在着一些固有的局限性。CNNs由于卷积核大小的限制通常只能学习局部区域的信息,忽略了全局信息,且无法获取特征信息的长期依赖关系。虽然

RNNs可以捕获长期依赖关系,但它的并行计算效率较低,且只作用于先前的记忆和当前的状态,而EEG是连续的,每一个部分都是相关的,包括过去、现在和未来<sup>[16]</sup>。在这种情况下,CNNs和RNNs都不足以有效地挖掘EEG的全局关联性。

近年来,基于自注意力的Transformer模型<sup>[17]</sup>能够很好地学习长期依赖关系,挖掘特征的全局关联信息,在自然语言处理和图像视觉领域均取得了先进的性能。由于Transformer的特性,最近也被提出用于各种生理信号分类任务中。Cheng等<sup>[18]</sup>采用1D-CNN和Transformer编码器对时域心音信号进行特征提取分类,实验结果验证了该方法的优越性;Song等<sup>[16]</sup>从通道-时间维出发,采用注意力机制的Transformer进行特征学习,提高了EEG运动想象分类的精度;Wang等<sup>[19]</sup>利用Transformer编码器对不同电极脑区信息整合并捕获脑区之间的空间依赖性,在EEG情感分类领域取得了不错的效果。这些方法表明了Transformer模型在生理信号分类领域的有效性,但是也忽略了空间-时间-频率信息之间的互补性。

针对以上问题,本文提出了一种多域信息融合的三维特征CNN Transformer机制(3D-CNN-Transformer mechanism, 3D-CTM)脑电情感识别模型,主要贡献如下:

(1)提出了一种新的3D特征结构,通过提取多个频带的DE特征,并结合电极通道空间位置信息和前后时间序列信息组织成3D特征矩阵,融合了EEG的空间、时间和频率信息。

(2)提出了由CNN和Transformer组成的CTM网络,CNN主要用于挖掘深层的多域特征信息,Transformer进一步学习特征信息的全局关联性。

(3)在SEED和SEED-IV数据集进行实验来表明3D-CTM模型的优越性,验证了学习率衰减策略和不同频带组合对模型的显著影响,以及进行消融研究来证明Transformer模块在挖掘特征信息全局关联性的作用。

## 1 方 法

图1所示为3D-CTM模型的总体框架,主要包括3D特征组织、CTM网络和分类(三分类或四分类)。

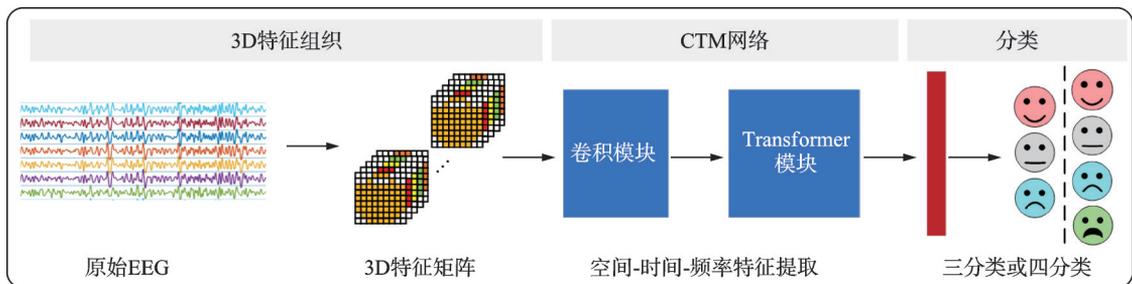


图1 3D-CTM模型总体框架

Fig.1 Overall framework of 3D-CTM model

### 1.1 3D特征组织

图2是对单个原始EEG样本处理生成3D特征结构的具体过程。首先,对于时长为 $T$ 的样本采用 $W$ 的时间窗口进行无重叠分割成 $S$ 段,将每个 $W$ 窗口的EEG视为样本中的一帧,由此一个EEG样本中包含了 $S$ 个连续的帧,从而保持了EEG的时间信息。研究表明,EEG的高频部分对情感识别的影响更大,且多个频带融合的效果好于单个频带。因此提取了每一帧EEG的 $\theta(4\sim 7\text{ Hz})$ 、 $\alpha(8\sim 13\text{ Hz})$ 、

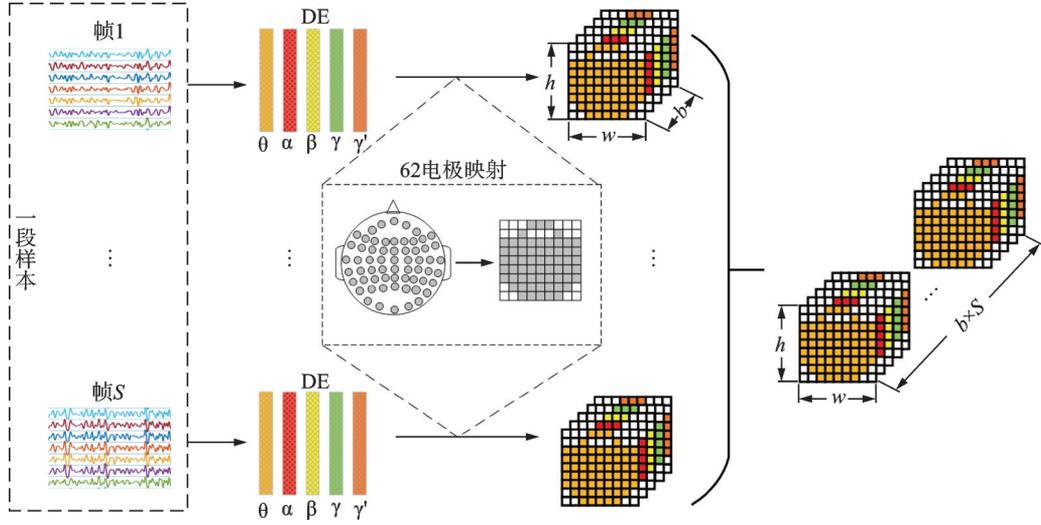


图2 3D特征组织流程图

Fig.2 Flowchart of 3D feature organization

$\beta$  (14~30 Hz)、 $\gamma$  (31~50 Hz)和 $\gamma'$  (51~75 Hz)这5个频带,值得注意的是,以往研究中大多只提取50 Hz以内的频带,而本文将51~75 Hz的频带也进行了应用。另外,文献[20]等研究表明,DE特征是脑电情感识别领域中最适合和最稳定的特征之一,相较于其他常用特征(如PSD)更为优秀。因此计算了每帧EEG在5个频带上的DE特征,计算公式为

$$DE = - \int_x f(x) \log(f(x)) dx \quad (1)$$

式中: $f(x)$ 表示连续信息的概率密度函数。对于一段固定长度的近似服从 $N(\mu, \sigma^2)$ 高斯分布的EEG,其DE可以表示为

$$DE = \frac{1}{2} \log(2\pi e \sigma^2) \quad (2)$$

式中: $e$ 为欧拉常数, $\mu$ 和 $\sigma$ 分别为 $x$ 的均值和标准差。

为了保持脑电电极位置的空间信息,采用了构造二维矩阵的方法,如图2中间虚线框内所示,根据脑电电极在大脑中的位置,将62通道EEG的DE特征转换为二维DE特征矩阵,对于没有值的位置采用0进行填充,以避免其他冗余信息。之后将每一帧EEG的所有频带的二维DE特征矩阵叠加得到 $h \times w \times b$ 大小的三维矩阵,其中 $h$ 和 $w$ 分别表示电极在垂直和水平方向上的最大值<sup>[21]</sup>, $b$ 表示频带的数量。最后,把每帧信号构成的三维特征矩阵按照 $S$ 帧的时间顺序在频带维进行拼接,最终得到了一个样本构成的三维特征矩阵 $X_n \in \mathbb{R}^{h \times w \times (b \times S)}$ ,从而融合了EEG的空间、时间和频率信息。本文中,样本长度 $T$ 和时间窗口 $W$ 分别为6和1,即每个样本有6个连续的帧, $h$ 、 $w$ 和 $b$ 的值分别为9、9和5。

## 1.2 CTM网络

本节提出的CTM网络以融合了空间、时间和频率信息的3D特征矩阵作为输入,包括CNN模块、Transformer模块和分类器模块。CTM网络的具体架构如图3所示。

### 1.2.1 CNN模块

CNN模块主要包含局部空间-时频卷积层、全局空间卷积层和平均池化3个部分。受伪3D卷积<sup>[22]</sup>思想的启发,将传统的大小为 $k \times k \times l$ 的3D卷积核等效分解为2D-CNN的 $k \times k \times 1$ 卷积核和1D-CNN的 $1 \times 1 \times l$ 卷积核,可以有效减少计算成本,加快训练速度。因此,本文将局部空间-时频卷积层的3D卷

积核分解成了局部空间卷积层和时频卷积层。

CNN模块的相关层参数细节如表1所示。其中,局部空间卷积层包含64个大小为 $3 \times 3 \times 1$ 的卷积核,步长为1,该层用于挖掘相邻电极通道的空间信息,并采用零填充模式防止边缘空间信息的丢失。时频卷积层包含64个大小为 $1 \times 1 \times 5$ 的卷积核,步长为1,该层用于提取时频维度上的信息。全局空间卷积层的内核大小为 $9 \times 9 \times 1$ ,即与二维空间矩阵大小相同,该层用于整合全局电极通道的空间信息,且内核数量设置为32,进一步压缩融合特征信息,减少参数量。在所有的卷积层后添加批归一化(Batch normalization, BN)操作来加快训练速度,抑制过拟合。接下来采用内核大小为 $1 \times 1 \times 6$ ,步长为4的平均池化层沿着时频维度进一步压缩特征信息。最后,将得到的特征图进行维度变换以适应Transformer模块的输入。CNN模块的操作公式为

$$Y_n = \text{CNN}_{\text{module}}(X_n) \quad n = 1, 2, \dots, N \quad (3)$$

式中: $\text{CNN}_{\text{module}}(\cdot)$ 表示CNN模块的全部操作, $Y_n \in \mathbb{R}^{6 \times 32}$ 表示提取的深层特征信息,其中6表示特征数量,32表示每个特征的维度信息,也是Transformer网络中的标记(token), $N$ 表示样本的数量。

### 1.2.2 Transformer 模块

Transformer模块采用自注意力机制使得模型可以同时考虑输入序列中不同位置的相关性,从而有效挖掘深层特征信息的全局关联性,以弥补CNN模块只能提取局部信息的局限性,此模块包含位置编码(Positional encoding, PE)和Transformer编码器两部分。其中,位置编码向模型提供有关输入序列中的位置信息。Transformer编码器负责处理输入序列信息编码成高维空间表示,主要包含多头自注意力(Muti-head self attention, MHA)机制和多层感知(Multiple layer perception, MLP)两个模块,其中最重要的是MHA机制,它通过拼接多个自注意力层形成,每个自注意力层通过计算输入序列中各个位置之间的依赖关系,得到每个位置的上下文表示,使得模型在不同位置之间进行交互和信息传递。MLP由两个全连接层组成,第1个全连接层将输入维度转换为更高维度的中间表示,第2个全连接层再将中间表示映射回原始的维度,对增强模型的表示能力和处理复杂关系具有关键作用。此外,在两模块的周围都采用了残差连接和层归一化(Layer normalization, LN)的处理,以保护信息的完整性,加速网络训练。Transformer模块以 $Y_n$ 作为输入,具体操作为

$$Z_n = Y_n + PE_{\text{inf}} \quad n = 1, 2, \dots, N \quad (4)$$

$$W_n = \text{LN}(\text{MHA}(Z_n) + Z_n) \quad n = 1, 2, \dots, N \quad (5)$$

$$M_n = \text{LN}(\text{MLP}(W_n) + W_n) \quad n = 1, 2, \dots, N \quad (6)$$

式中: $PE_{\text{inf}} \in \mathbb{R}^{6 \times 32}$ 添加的位置信息,MHA( $\cdot$ )、MLP( $\cdot$ )和LN( $\cdot$ )分别表示MHA、MLP和LN操作。经

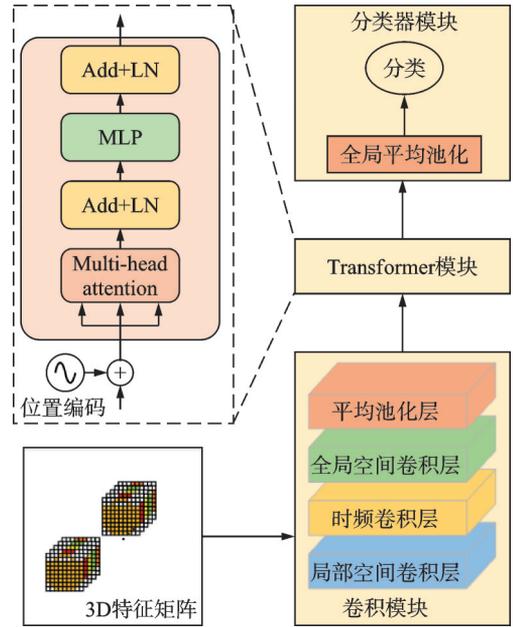


图3 CTM网络架构

Fig.3 Architecture of the CTM network

表1 CNN模块相关层参数

Table 1 Relevant layer parameters of CNN module

相关层参数	输出尺寸
局部空间卷积层, $3 \times 3 \times 1@64$	$9 \times 9 \times 30 \times 64$
时频卷积层, $1 \times 1 \times 5@64$	$9 \times 9 \times 26 \times 64$
全局空间卷积层, $9 \times 9 \times 1@32$	$1 \times 1 \times 26 \times 32$
平均池化层, $1 \times 1 \times 6, 1 \times 1 \times 4$	$1 \times 1 \times 6 \times 32$
Reshape 维度转换	$6 \times 32$

实验验证, MHA 机制的头数设置为 8, MLP 中的第 1 个全连接层的表示维度设置为 128, 最终得到添加了自注意力机制的输出  $M_n$ 。

### 1.2.3 分类器模块

经上述处理,  $M_n$  整合了多域特征信息, 同时也蕴含了特征信息之间的关联性。分类器模块采用全局平均池化的操作, 将  $M_n$  沿着特征维度进行平均输出一维向量, 即

$$V_n = \text{GAP}(M_n) \quad n = 1, 2, \dots, N \quad (7)$$

最后将  $V_n \in \mathbf{R}^{32}$  连接到全连接层, 神经元的数量等于类别的数量, 利用 Softmax 函数进行分类。

## 2 实验验证

### 2.1 数据集与数据预处理

SEED<sup>[20]</sup> 和 SEED-IV<sup>[23]</sup> 数据集均为上海交通大学 BCMI 实验室所采集的公共情感数据集。SEED 数据集共有积极、中性和消极 3 种情感, 包含平均年龄在 23 岁左右的 15 名受试者 (7 男 8 女) 的脑电数据, 采用国际 10-20 体系, 从 62 个脑电电极记录 15 个受试者观看相关视频的 EEG。每个受试者在 3 个不同的时间段分别采集 EEG, 每个时间段受试者观看相同的 15 个视频, 每个视频约 4 min, 其中积极、中性和消极的电影片段各 5 个。视频之间有 45 s 的自我评估时间和 15 s 的休息时间。最后对采集后的数据下采样至 200 Hz, 并应用 0~75 Hz 的带通滤波器进行滤波。本文截取了 SEED 数据集中每个 EEG 的其中 120 s 作为实验数据, 并对每个试验按照时长  $T$  不重叠划分样本。

SEED-IV 数据集共包括快乐、悲伤、恐惧和中性 4 种情感, 包含 15 名受试者 (6 男 9 女) 的脑电数据, 以同样的方式进行采集, 不同的是每个受试者在 3 个不同的时间段分别观看不同的 24 个视频, 每个视频约 2 min。最后将处理后的数据下采样至 200 Hz, 应用了 1~75 Hz 的带通滤波器进行滤波。本文对 SEED-IV 数据集中每个时间段的 EEG 分别进行整体混合, 对每个试验按照  $T$  不重叠划分样本, 并去掉每个试验剩余不足  $T$  的数据。预处理后的 SEED 和 SEED-IV 数据集格式如表 2 所示。

表 2 预处理后的 SEED 和 SEED-IV 数据集格式

Table 2 Format of preprocessed SEED and SEED-IV datasets

数据	数据格式	标签信息
SEED	13 500×62×1 200	13 500×3
SEED-IV-1	8 445×62×1 200	8 445×4
SEED-IV-2	8 265×62×1 200	8 265×4
SEED-IV-3	8 154×62×1 200	8 154×4
组成	样本数量×通道×采样点	样本数量×情感数量

### 2.2 实验设置

实验使用的硬件环境为 NVIDIA Geforce GTX 3090 GPU, 软件环境采用了 Python3.8 的 Keras 框架搭建网络模型。采用 Adam 优化器, 每次迭代的批大小为 128, 学习率初始值设置为 0.000 4, 并采用自定义的学习率衰减策略。对于 SEED 数据集, 采用所有受试者参与的整体分类实验, 此方式更适用于情感脑机接口的实际应用。对于 SEED-IV 数据集, 对每个时间段采用整体分类实验, 最终结果取 3 个时间段的平均值。为了更好地评估模型的稳定性, 所有实验均采用 5 折交叉验证的方式。

### 2.3 学习率衰减策略的影响

本文采用了自定义学习率衰减策略优化算法, 即每当连续 15 个 epoch 验证损失没有下降时, 学习率就变为原来的一半, 能够显著影响实验模型的效果。为了验证学习率衰减策略的对本文模型的影响, 以 SEED 数据集为例进行实验。

图4(a,b)分别为在验证过程当中,3D-CTM模型使用学习率衰减策略和不使用学习率衰减策略的准确率和损失随epoch迭代的趋势图。在训练300个epoch过程中,采用学习率衰减策略时在100个epoch之前有着比较剧烈的波动,之后便基本趋于平稳,在200个epoch之后准确率误差不超过0.2%,损失误差不超过0.002。未采用学习率衰减策略时模型一直存在着剧烈的波动,收敛状况不佳。由此可见,学习率衰减策略可以通过在训练过程中逐渐降低学习率,模型可以更快地找到最优解,有效地加快模型的收敛速度,缩短训练时间。此外,学习率衰减策略还可以提高模型的稳定性,使得模型更加可靠。

2.4 不同频带组合的影响

在以往的频带组合相关研究中,大多只提取50 Hz以内的 $\delta$ (1~3 Hz)、 $\theta$ (4~7 Hz)、 $\alpha$ (8~13 Hz)、 $\beta$ (14~30 Hz)和 $\gamma$ (31~50 Hz)五个频带组合进行研究,往往忽略了51~75 Hz之间的频带,本文为了验证包含 $\gamma'$ (51~75 Hz)频带的组合对情感分类的作用,进行了3组频带组合研究,实验结果如表3所示。从表3可以看出,本文所采用的包含 $\gamma'$ (51~75 Hz)的5种频带组合在SEED以及SEED-IV数据集的3个时间段均取得了最好的效果,在两个数据集的最终准确率分别为96.36%、87.44%。与以往研究常常采用的 $\delta$ 、 $\theta$ 、 $\alpha$ 、 $\beta$ 、 $\gamma$ 五个频带组合的实验结果相比有着明显的提升,在SEED数据集上提升了2.77%,在SEED-IV数据集的3个时间段分别提升了6.49%、6.72%、6.61%,平均提升了6.60%,证明了包含 $\gamma'$ (51~75 Hz)的5种频带组合对情感分类的重要性,也说明了 $\gamma'$ (51~75 Hz)相较于 $\delta$ (1~3 Hz)所包含着对情感分类更具判别性的信息。另外,本文对全部6个频带的组合也进行了实验,所得实验结果并没有达到最高,这可能是由于组合所有频带所包含的与情感分类无关的冗余信息过多导致的,尤其是 $\delta$ 频带可能对判别情感状态并不是那么关键。

2.5 消融研究

为了探讨Transformer模块在分类中的作用,进行了消融研究,即分别使用SEED和SEED-IV数据集去除Transformer模块进行实验,结果如图5所示。在SEED数据集上,去除Transformer模块后,准确率下降了3.07%。同样,对于SEED-IV数据集的3个时间段,去

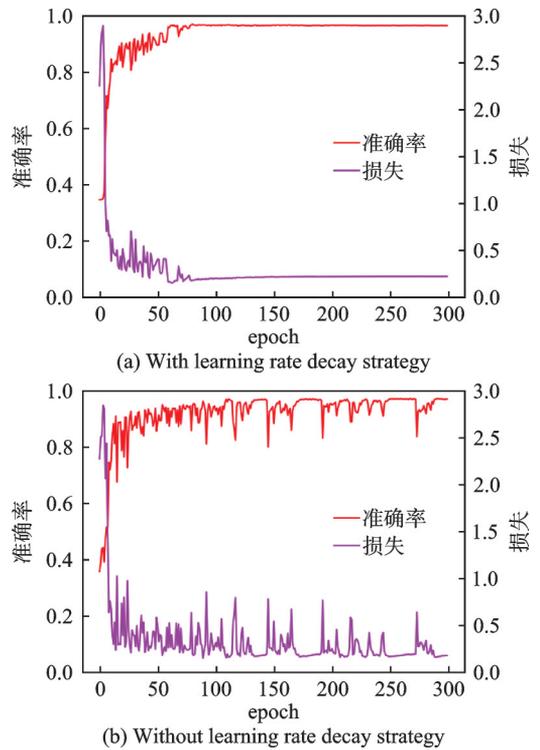


图4 准确率和损失趋势图

Fig.4 Accuracy and loss trend chart

表3 不同频带组合的分类准确率

Table 3 Classification accuracy for different frequency band combinations

数据	$\delta+\theta+$	$\theta+\alpha+$	$\delta+\theta+\alpha+$
	$\alpha+\beta+\gamma$	$\beta+\gamma+\gamma'$	$\beta+\gamma+\gamma'$
SEED	93.59	96.36	95.83
SEED-IV-1	80.76	87.25	85.16
SEED-IV-2	79.09	85.81	84.16
SEED-IV-3	82.66	89.27	88.04
SEED-IV-Ave	80.84	87.44	85.79



图5 Transformer模块消融研究

Fig.5 Ablation study of Transformer module

除 Transformer 模块后,准确率分别下降了 2.93%、4.30% 和 4.00%,最终的平均结果下降了 3.74%。由此说明,Transformer 模块对本模型具有显著作用,能够很好地挖掘 CNN 模块提取的特征信息的全局关联性,有效提高情感分类的准确率。另外也说明了 Transformer 模块能够适用于不同的数据集和时间段,具有良好的鲁棒性和可靠性。

## 2.6 方法对比分析

为了进一步验证本文方法的先进性,以及空间-时间-频率信息融合和 Transformer 对 EEG 情感识别的作用,将分类结果与近 5 年来最先进的情感识别算法进行了比较,结果如表 4 所示。其中,DGCNN 模型和 RGNN 模型分别采用动态图卷积神经网络和正则化图神经网络来学习不同 EEG 通道之间的内在关系,主要是从 EEG 的空间信息方面进行分析。R2G-STNN 模型关注了 EEG 的时间和空间信息,采用双向长短时记忆网络(Bidirectional LSTM, BiLSTM)来学习大脑区域和全局的时空特征进行分类。以上方法仅融合了空间、时间和频率信息的一种或两种,最终所得的结果相对也较低。

表 4 相关方法对比

Table 4 Comparison of relevant methods

名称	方法	SEED 准确率/%	SEED-IV 准确率/%
DGCNN <sup>[24]</sup>	空	90.40	—
RGNN <sup>[25]</sup>	空	94.24	79.37
R2G-STNN <sup>[6]</sup>	空时	93.38	—
4D-CRNN <sup>[26]</sup>	空时频	94.74	—
4D-aNN <sup>[27]</sup>	空时频	96.25	86.77
SST-EmotionNet <sup>[22]</sup>	空时频	96.02	84.92
EEG Conformer <sup>[28]</sup>	时+Transformer	95.30	—
EeT <sup>[29]</sup>	空时+Transformer	96.28	83.27
3D-CTM(本文)	空时频+Transformer	96.36	87.44

4D-CRNN 模型构建了一种 4D 的特征结构,整合了 EEG 的空间、时间和频率信息,使用 CNN 和 LSTM 分别来挖掘空频信息和时间信息;4D-aNN 模型在上述 4D 特征结构的基础上,增添了空间-频谱注意力机制和时间注意力机制,采用 CNN+BiLSTM 的网络架构进行分类;SST-EmotionNet 模型在一个统一的网络框架内采用双结构形式集成了 EEG 的空间、时间和频率信息,具体来说,其中包括空频结构和时空结构,并在两个结构上都增添了注意力机制。以上方法均融合了 EEG 的空间、时间和频率信息,分类结果均有提升。值得注意的是,在添加了注意力机制后,能够自适应地探索有判别性的局部特征,对分类结果有了较好的提升,但这种注意力机制只能挖掘局部的关键信息,而忽略了全局信息的关联性。

EEG Conformer 模型采用 CNN 对通道-时间的 EEG 进行局部特征提取,之后采用 Transformer 对提取的信息进行全局表示,但是忽略了 EEG 的频率信息和电极间的空间位置信息。EeT 模型采用 Transformer 自注意机制直接应用于 EEG 序列的时间和空间信息中,忽略了 EEG 的频率信息。

本文提出的 3D-CTM 模型融合了 EEG 的空间、时间和频率信息,并利用 Transformer 对全局信息进行关注,在 SEED 和 SEED-IV 数据集的平均准确率分别达到了 96.36% 和 87.44%,与近年来的先进算法相比都达到了最好的结果。另外经过分析,EEG 的空间、时间和频率信息以及全局信息的关联性两个方面对情感识别都有着重要的作用。

### 3 结束语

本文提出了一种用于EEG情感识别的3D-CTM模型。首先,提出的3D特征结构充分整合了空间、时间和频率信息,有效利用了不同信息之间的互补性。之后采用不同的CNN结构对3种特征信息进行学习,并进一步利用Transformer挖掘全局特征信息的关联性。在不同的脑电情感数据集上进行实验都取得了显著的效果。此外,学习率衰减策略的应用有效地减少了模型的训练时间、提高了模型的稳定性。通过组合不同的频带证明了先前研究常常忽略的51~75 Hz频带对情感分类的作用。消融研究验证了Transformer模块的有效性。最后通过对相关方法进行了详细的对比分析,说明了3D-CTM模型在融合多域信息以及挖掘信息关联性两个方面对情感识别的结果都有着重要的作用。

除了脑电信号,面部表情、语音信号和各种生理信号都是情感识别领域的重要方向。当前已开展了许多关于多模态情感识别的研究,涵盖了图像、声音和生理信号等多种信息源。未来的研究可以进一步探索不同模态信息之间的相互作用,以实现更全面和精准的情感识别,为情感脑机接口技术的发展提供重要支持。

#### 参考文献:

- [1] 陈景霞,胡修文,唐喆喆,等.基于卷积联合适应网络的脑电信号情感识别[J].数据采集与处理,2022,37(4): 814-824.  
CHEN Jingxia, HU Xiuwen, TANG Zhezhe, et al. EEG emotion recognition based on convolutional joint adaptation network [J]. Journal of Data Acquisition and Processing, 2022, 37(4): 814-824.
- [2] 秦天鹏,生慧,岳路,等.脑电信号情绪识别研究综述[J].计算机工程与应用,2023,59(15): 38-54.  
QIN Tianpeng, SHENG Hui, YUE Lu, et al. Review of research on emotion recognition based on EEG signals[J]. Computer Engineering and Applications, 2023, 59(15): 38-54.
- [3] JAIN D K, SHAMSOLMOALI P, SEHDEV P. Extended deep neural network for facial emotion recognition[J]. Pattern Recognition Letters, 2019, 120: 69-74.
- [4] LIU Zhentao, WU Min, CAO Weihua, et al. Speech emotion recognition based on feature selection and extreme learning machine decision tree[J]. Neurocomputing, 2018, 273: 271-280.
- [5] NOROOZI F, CORNEANU C A, KAMIŃSKA D, et al. Survey on emotional body gesture recognition[J]. IEEE Transactions on Affective Computing, 2018, 12(2): 505-523.
- [6] LI Yang, ZHENG Weiming, WANG Lei, et al. From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition[J]. IEEE Transactions on Affective Computing, 2019, 13(2): 568-578.
- [7] FERADOV F, GANCHEV T. Ranking of EEG time-domain features on the negative emotions recognition task[J]. Annual Journal of Electronics, 2015, 9: 26-29.
- [8] DUAN Ruonan, ZHU Jiayi, LU Baoliang. Differential entropy feature for EEG-based emotion classification[C]//Proceedings of 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). [S.l.]: IEEE, 2013: 81-84.
- [9] LI Jinpeng, ZHANG Zhaoxiang, HE Huiguang. Hierarchical convolutional neural networks for EEG-based emotion recognition [J]. Cognitive Computation, 2018, 10: 368-380.
- [10] WANG Yi, HUANG Zhiyi, MCCANE B, et al. EmotioNet: A 3-D convolutional neural network for EEG-based emotion recognition[C]//Proceedings of 2018 International Joint Conference on Neural Networks (IJCNN). [S.l.]: IEEE, 2018: 1-7.
- [11] YANG Yilong, WU Qingfeng, FU Yazhen, et al. Continuous convolutional neural network with 3D input for EEG-based emotion recognition[C]//Proceedings of 25th International Conference on Neural Information Processing. Siem Reap, Cambodia: Springer, 2018: 433-443.
- [12] WANG Fei, WU Shichao, ZHANG Weiwei, et al. Emotion recognition with convolutional neural network and EEG-based EFDMs[J]. Neuropsychologia, 2020, 146: 107506.
- [13] WU Yihan, XIA Min, NIE Li, et al. Simultaneously exploring multi-scale and asymmetric EEG features for emotion recognition[J]. Computers in Biology and Medicine, 2022, 149: 106002.

- [14] CUI Heng, LIU Aiping, ZHANG Xu, et al. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network[J]. *Knowledge-Based Systems*, 2020, 205: 106243.
- [15] YANG Yilong, WU Qingfeng, QIU Ming, et al. Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network[C]//*Proceedings of 2018 International Joint Conference on Neural Networks (IJCNN)*. [S.l.]: IEEE, 2018: 1-7.
- [16] SONG Yonghao, JIA Xueyu, YANG Lie, et al. Transformer-based spatial-temporal feature learning for EEG decoding[EB/OL]. (2021-06-11). <https://doi.org/10.48550/arXiv.2106.11170>.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998-6008.
- [18] CHENG Jiawen, SUN Kexue. Heart sound classification network based on convolution and transformer[J]. *Sensors*, 2023, 23(19): 8168.
- [19] WANG Zhe, WANG Yongxiong, HU Chuanfei, et al. Transformers for EEG-based emotion recognition: A hierarchical spatial information learning model[J]. *IEEE Sensors Journal*, 2022, 22(5): 4359-4368.
- [20] ZHENG Weilong, LU Baoliang. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks[J]. *IEEE Transactions on Autonomous Mental Development*, 2015, 7(3): 162-175.
- [21] 蔡冬丽, 钟清华, 朱永升, 等. 基于混合神经网络的脑电情感识别[J]. *华南师范大学学报(自然科学版)*, 2021, 53(1): 109-118.
- CAI Dongli, ZHONG Qinghua, ZHU Yongsheng, et al. EEG emotion recognition based on hybrid neural networks[J]. *Journal of South China Normal University (Natural Science Edition)*, 2021, 53(1): 109-118.
- [22] JIA Ziyu, LIN Youfang, CAI Xiyang, et al. SST-EmotionNet: Spatial-spectral-temporal based attention 3D dense network for EEG emotion recognition[C]//*Proceedings of the 28th ACM International Conference on Multimedia*. [S.l.]: ACM, 2020: 2909-2917.
- [23] ZHENG Weilong, LIU Wei, LU Yifei, et al. EmotionMeter: A multimodal framework for recognizing human emotions[J]. *IEEE Transactions on Cybernetics*, 2019, 49(3): 1110-1122.
- [24] SONG Tengfei, ZHENG Wenming, SONG Peng, et al. EEG emotion recognition using dynamical graph convolutional neural networks[J]. *IEEE Transactions on Affective Computing*, 2018, 11(3): 532-541.
- [25] ZHONG Peixiang, WANG Di, MIAO Chunyan. EEG-based emotion recognition using regularized graph neural networks[J]. *IEEE Transactions on Affective Computing*, 2020, 13(3): 1290-1301.
- [26] SHEN Fangyao, DAI Guojun, LIN Guang, et al. EEG-based emotion recognition using 4D convolutional recurrent neural network[J]. *Cognitive Neurodynamics*, 2020, 14: 815-828.
- [27] XIAO Guowen, YE Mengwen, XU Bowen, et al. 4D attention-based neural network for EEG emotion recognition[EB/OL]. (2021-01-14). <https://doi.org/10.48550/arXiv.2101.05484>.
- [28] SONG Yonghao, ZHENG Qingqing, LIU Bingchuan, et al. EEG conformer: Convolutional transformer for EEG decoding and visualization[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022, 31: 710-719.
- [29] LIU Jiyao, WU Hao, ZHANG Li, et al. Spatial-temporal transformers for EEG emotion recognition[C]//*Proceedings of the 6th International Conference on Advances in Artificial Intelligence*. New York, USA: ACM, 2022: 116-120.

#### 作者简介:



张学军(1969-),男,博士,教授,硕士生导师,研究方向:智能信息处理、脑机接口技术、无线射频识别技术等, E-mail: xjzhang@njupt.edu.cn。



王天晨(2000-),男,硕士研究生,研究方向:智能信息处理、脑电信号处理。



王泽田(1999-),通信作者,男,硕士研究生,研究方向:智能信息处理、脑电信号处理, E-mail: wangzt1019@163.com。

(编辑:王静)