

文章编号:1671-6833(2026)01-0073-08

# 基于时间序列成像的多任务学习驱动情感识别

许胜新<sup>1</sup>, 梁弼政<sup>2</sup>, 胡飞<sup>3</sup>, 徐华兴<sup>3</sup>

(1. 中国电子科学研究院 社会安全风险感知与防控大数据应用国家工程研究中心, 北京 100041; 2. 中国空间技术研究院 通信与导航卫星总体部, 北京 100094; 3. 郑州大学 电气与信息工程学院, 河南 郑州 450001)

**摘要:** 针对脑电情感识别依赖特征抽取或时频谱图导致的计算复杂度高的问题, 提出一种基于时间序列成像(TSI)的多任务学习驱动情感识别方法。通过格拉姆角场、马尔可夫转移场以及模式差分场, 实现一维脑电信号到二维图像的直接编码; 基于 ResNet18 分类骨干网络, 设计多任务特征融合框架, 融合 3 种 TSI 编码提取的特征表示。实验结果表明: 在 DEAP 数据集上, 所提方法在 Valence 和 Arousal 的二分类中的平均分类准确率分别为 96.51% 和 97.22%, 在 AMIGOS 数据集上分别为 98.59% 和 99.64%, 扩展到四分类和八分类时, DEAP 上的平均分类准确率分别为 91.06% 和 87.43%, AMIGOS 上的平均准确率分别为 97.41% 和 89.84, 在脑电情感识别中具备良好的鲁棒性。

**关键词:** 脑电; 情感识别; 时间序列成像; 多任务; 特征融合

**中图分类号:** TP181; G202 **文献标志码:** A **doi:** 10.13705/j.issn.1671-6833.2026.01.005

生理信号能够真实刻画个体的生理变化, 在情感识别中展现出独特优势。其中, 脑电 (electroencephalogram, EEG) 信号因直接关联大脑活动而更具情感相关性, 采自左、右颞区和额叶区的脑电信号已被广泛用于表征情感状态<sup>[1]</sup>, 体现出较高的客观性和可靠性, 成为情感识别研究的重要手段。

传统的脑电情感识别研究通常包括特征提取和分类模型构建两个阶段。常用特征主要包括时域、频域以及时频域等。经典机器学习分类器包括决策树 (decision tree, DT)<sup>[2]</sup>、多层感知机 (multi-layer perceptron, MLP)<sup>[3]</sup>、 $k$  邻近 ( $k$ -nearest neighbor, KNN)<sup>[4]</sup>、随机森林 (random forest, RF)<sup>[5]</sup> 以及支持向量机 (support vector machine, SVM)<sup>[6]</sup>。传统的情感识别方法特征提取过程复杂, 分类器依赖反复调优, 对复杂情境下的情感分类能力有限。

深度学习的发展使其在脑电情感识别中的应用已成为研究的重点方向。总体而言, 基于深度学习的情感识别主要分为两大类。一类是先提取 EEG 特征, 再利用深度学习模型完成分类。魏琛等<sup>[7]</sup>基

于 6 级小波包变换提取特征, 集成 6 个不同深度的卷积神经网络 (convolutional neural networks, CNN) 进行分类; Song 等<sup>[8]</sup>提出动态图卷积神经网络 (dynamical graph convolutional neural networks, DGCNN) 模型对微分熵、功率谱、不对称差、不对称商以及微分因果这 5 类特征进行分类; Yang 等<sup>[9]</sup>构建时域卷积神经网络 (temporal convolutional network, TCN) 对固定长度、多个频带的微分熵 (differential entropy, DE) 特征进行分类。另一类是通过时频谱图将 EEG 转化为高维特征矩阵, 采用二维 CNN、三维 CNN 或 GCNN 等不同结构的网络完成分类。Salama 等<sup>[10]</sup>先将多通道脑电信号处理为二维矩阵, 再将多个二维矩阵组成三维矩阵, 构建 3D-CNN 对其进行分类。Ramzan 等<sup>[11]</sup>通过集成 CNN 和长短期记忆递归神经网络实现情感分类。

尽管基于深度学习的脑电情感识别在分类精度方面取得了显著提升, 但仍面临诸多问题与挑战。第一类方法仅将深度学习模型作为分类器, 难以充分发挥其端到端学习的优势, 且与传统方法类似, 在

收稿日期: 2025-04-02; 修订日期: 2025-05-05

基金项目: 国家重点研发计划 (2022YFC3502400); 中央本级重大增减支项目 (2060302-1802-03)

作者简介: 许胜新 (1991—) 男, 河南灵宝人, 中国电子科学研究院工程师, 博士, 主要从事人工智能、大数据等研究, E-mail: xushengxin@cetc.com.cn。

通信作者: 徐华兴 (1988—) 男, 河南确山人, 郑州大学副教授, 博士, 主要从事人工智能与信号处理应用研究, E-mail: xuhuaxing@zzu.edu.cn。

引用本文: 许胜新, 梁弼政, 胡飞, 等. 基于时间序列成像的多任务学习驱动情感识别[J]. 郑州大学学报(工学版), 2026, 47(1): 73-80. (XU S X, LIANG B Z, HU F, et al. Multi-task learning-driven emotion recognition based on time series imaging[J]. Journal of Zhengzhou University (Engineering Science), 2026, 47(1): 73-80.)

最优特征选择上存在较大的不确定性。而第二类方法则依赖多种时频分析方法,缺乏统一的标准,不同实验对象往往需要针对性地调整分析方法和参数设置。此外,部分基于图卷积的网络在构建中高度依赖专业知识,尤其是在图结构设计和节点特征定义方面,对领域经验要求较高。

为了充分利用情感脑电信号中的丰富信息,降低特征提取和模型构建的复杂度,本文提出基于时间序列成像(time series imaging, TSI)的多任务学习驱动方法。首先,一维脑电信号通过3种不同的时间序列成像算法转换为二维图像;其次,为实现不同成像算法之间的优势互补,基于多任务学习,设计多任务特征融合框架,构建多个并行 ResNet18 骨干网络和融合模块,提高情感识别的准确性;最后,本文在 DEAP 和 AMIGOS 数据集上对基于时间序列成像的多任务学习驱动情感识别方法开展情感识别二分类、四分类和八分类任务的性能评估,并系统分析了不同频带、时间尺度、特征融合策略以及数据增强方式的影响。

## 1 时间序列成像

时间序列成像将时间序列数据划分为多个窗口,并将每个窗口数据转换为图像,最终将这些图像合并成一个整体图像,从而直观展示时间序列数据的演变趋势、周期性和异常情况。本文以时间序列成像在标准时间序列数据集的分类性能作为选取依据,采用格拉姆角场、马尔可夫转移场、模式差分场3种算法将脑电信号转化为图像。

### 1.1 格拉姆角场

格拉姆角场(Gramian angular field, GAF)变换是一种将时间序列通过极坐标映射为角度,利用角度和余弦与角度差正弦分别生成对称与反对称矩阵的图像编码方法。该变换在保留数据时序相关性的同时,能够捕捉序列的空间连续性。GAF 编码的步骤如下。

(1)最小-最大归一化。将时间序列  $\mathbf{X} = \{x_i\}_{i=1}^n$  缩放到  $[-1, 1]$ , 归一化后的时间序列  $\bar{\mathbf{X}} = \{\bar{x}_i\}_{i=1}^n$ , 其中,

$$\bar{x}_i = \frac{x_i - \max(\mathbf{X}) + x_i - \min(\mathbf{X})}{\max(\mathbf{X}) - \min(\mathbf{X})}。 \quad (1)$$

(2)极坐标转换。

$$\begin{cases} \phi = \cos^{-1}(\bar{x}_i), -1 \leq \bar{x}_i \leq 1, \bar{x}_i \in \bar{\mathbf{X}}; \\ r = \alpha t_i, t_i \in \mathbf{N}. \end{cases} \quad (2)$$

式中:  $\phi$  为转换后的角度;  $\alpha$  为正则化极坐标跨度的

常数因子;  $t_i$  为  $x_i$  的时间戳。

(3)三角和差。通过角度和余弦与角度差正弦分别得到格拉姆角和场(Gramian angular summation field, GASF)矩阵  $\mathbf{O} \in \mathbf{R}^{n \times n}$  与格拉姆角差场(Gramian angular difference field, GADF)矩阵  $\mathbf{H} \in \mathbf{R}^{n \times n}$ , 其元素分别为

$$\begin{cases} O_{ij} = \cos(\phi_i + \phi_j), 1 \leq i, j \leq n; \\ H_{ij} = \sin(\phi_i - \phi_j), 1 \leq i, j \leq n. \end{cases} \quad (3)$$

式中:  $\phi_i$  和  $\phi_j$  分别为时刻  $i$  和时刻  $j$  处的归一化信号值通过式(2)转换的角度。

GASF 为对称图像,反映了相邻时刻信号的相似性;GADF 为反对称图像,强调相邻时刻信号的趋势差异。GAF 变换的运算过程如图 1 所示。

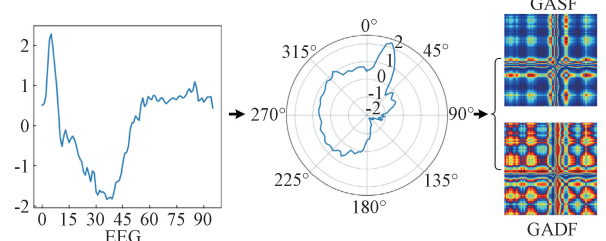


图 1 基于 GAF 的脑电信号图像编码

Figure 1 EEG signal image encoding with GAF

### 1.2 马尔可夫转移场

马尔可夫转移场(Markov transition field, MTF)借助马尔可夫转移概率来捕捉时间序列的动态变化,将一维时间序列转换为二维的转移概率图像。MTF 的图像编码步骤如下。

(1)分位离散化。将时间序列  $\mathbf{X} = \{x_i\}_{i=1}^n$  的值域范围通过  $Q$  分位离散化为状态空间  $\{q_k\}_{k=1}^Q$ , 分配  $x_i$  到其对应的  $q_k$  构造新的状态序列  $\mathbf{S} = \{s_i\}_{i=1}^n, s_i \in \{q_k\}_{k=1}^Q$ 。

(2)计算转移概率。根据状态序列  $\mathbf{S}$ , 基于一阶马尔可夫链,计算从状态  $q_k$  转移到状态  $q_l$  的转移概率  $p(q_k \rightarrow q_l)$ :

$$p(q_k \rightarrow q_l) = \frac{\#(q_k \rightarrow q_l)}{\sum_{m=1}^Q \#(q_k \rightarrow q_m)}; \quad (4)$$

$$\#(q_k \rightarrow q_l) = \sum_{i=1}^n \mathbf{1}\{s_i = q_k \text{ 且 } s_{i+1} = q_l\}。 \quad (5)$$

式中:  $\#(q_k \rightarrow q_l)$  为状态从  $q_k$  转移到  $q_l$  的转移次数;  $1 \leq k, l \leq Q$ 。

(3)构造 MTF 矩阵  $\Phi \in \mathbf{R}^{Q \times Q}$ , 其元素为

$$\Phi_{kl} = p(q_k \rightarrow q_l)。 \quad (6)$$

4 分位 MTF 变换的运算过程如图 2 所示。

### 1.3 模式差分场

模式差分场(motif difference field, MDF)划分时

间序列为相同长度的重叠子序列,通过计算子序列中相邻点的欧氏距离生成差分矩阵,并按相邻点的时间间隔维度堆叠,构造多幅图像。该方法的主要优点在于能够捕捉时间序列中的局部模式和重复性,并且不需要先验知识或模型。MDF 的图像编码步骤如下。

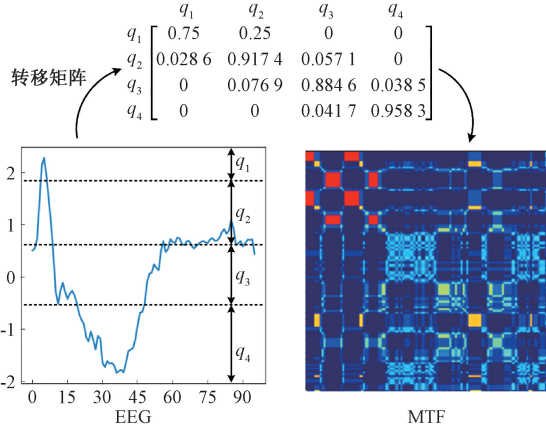


图 2 基于 MTF 的脑电信号图像编码

Figure 2 EEG signal image encoding with MTF

(1) 差分子序列。基于时间序列  $X = \{x_i\}_{i=1}^n$ , 构造起始点为  $s$ , 步长为  $d$ , 长度为  $m$  的子序列  $M_{d,s}^m \in \mathbf{R}^m$ :

$$M_{d,s}^m = \{x_j\}, j = s, s + d, \dots, s + (m - 1)d. \quad (7)$$

式中:  $1 < m \leq n; s = 1, 2, \dots, n - (m - 1)d, 1 \leq d \leq d_{\max}; d_{\max} = \lfloor (n - 1) / (m - 1) \rfloor$ 。

通过将  $M_{d,s}^m$  按行堆叠, 可得步长为  $d$  的子序列矩阵  $M_d^m \in \mathbf{R}^{(n-(m-1)d) \times m}$ :

$$M_d^m = [[M_{d,1}^m]^T, [M_{d,2}^m]^T, \dots, [M_{d,n-(m-1)d}^m]^T]^T. \quad (8)$$

对  $M_{d,s}^m$  进行相邻点差分, 可得差分子序列矩阵  $dM_d^m \in \mathbf{R}^{(n-(m-1)d) \times (m-1)}$ :

$$dM_d^m = [[dM_{d,1}^m]^T, [dM_{d,2}^m]^T, \dots, [dM_{d,n-(m-1)d}^m]^T]^T. \quad (9)$$

式中:  $dM_{d,s}^m$  的第  $i (i = 1, 2, \dots, m - 1)$  个元素为  $x_{s+id} - x_{s+(i-1)d}$ 。

(2) 零填充。由于  $dM_d^m$  中子序列的个数为  $n - (m - 1)d$ , 最多为  $n - (m - 1)$ , 为了使得  $dM_d^m$  中子序列的维度相同, 均为  $(n - (m - 1))(m - 1)$ , 构造新序列  $P_{d,s}^m$  满足以下条件:

$$P_{d,s}^m = \begin{cases} dM_{d,s}^m, & 1 \leq s \leq n - (m - 1)d; \\ \mathbf{0}, & n - (m - 1)d < s \leq n - (m - 1). \end{cases} \quad (10)$$

相对应的, 新的差分矩阵为

$$P_d^m = [[P_{d,1}^m]^T, [P_{d,2}^m]^T, \dots, [P_{d,n-(m-1)}^m]^T]^T. \quad (11)$$

将不同步长的新差分矩阵堆叠, 可得模式差分场:

$$J^m(:, :, d) = P_d^m, d = 1, 2, \dots, d_{\max}. \quad (12)$$

(3) 图像修正。由以上分析可知, 长度为  $m$  的差分子序列  $P_{d,s}^m$  有  $m - 1$  个元素, 所以  $J^m$  可以生成  $m - 1$  个通道图像。第  $k$  个通道图像矩阵可以表示为

$$G_k^m = [P_1^m(:, k), P_2^m(:, k), \dots, P_{d_{\max}}^m(:, k)]^T. \quad (13)$$

式中:  $G_k^m \in \mathbf{R}^{d_{\max}(n-(m-1))}; 1 \leq k \leq m - 1$ 。为了修正填充的 0 元素,  $k$  通道的图像矩阵可更改为

$$L_k^m = G_k^m + K^m \odot G_k^m. \quad (14)$$

式中:  $\odot$  为 Hadamard 积;  $G_k^m$  由  $G_k^m$  旋转  $180^\circ$  后得到;  $K^m$  为掩码矩阵, 其元素表示为

$$K_{ds}^m = \begin{cases} 0, & 1 \leq s \leq n - (m - 1)d; \\ 1, & n - (m - 1)d < s \leq n - (m - 1). \end{cases} \quad (15)$$

子序列长度为 3 的 MDF 变换的运算过程如图 3 所示。

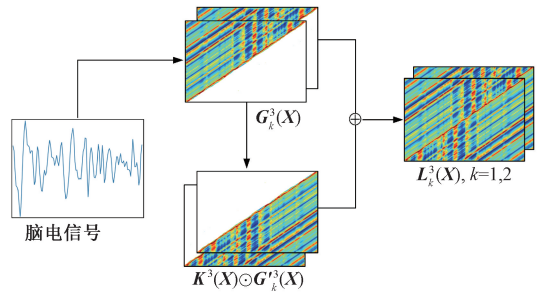


图 3 基于 MDF 的脑电信号图像编码

Figure 3 EEG signal image encoding with MDF

## 2 多任务特征融合

GAF 通过极坐标映射, 衡量数据点之间的相关性和连续性; MTF 利用各离散状态间的转移概率描述脑电信号整体变化趋势; MDF 通过求相邻脑电片段的差值来判断相邻脑电序列的相似性。不同的 TSI 算法侧重于刻画脑电信号的不同特征, 生成的图像具有互补性。

为充分融合不同 TSI 算法编码图像的特征, 进一步提高模型的分类能力, 设计多任务特征融合框架, 如图 4 所示。

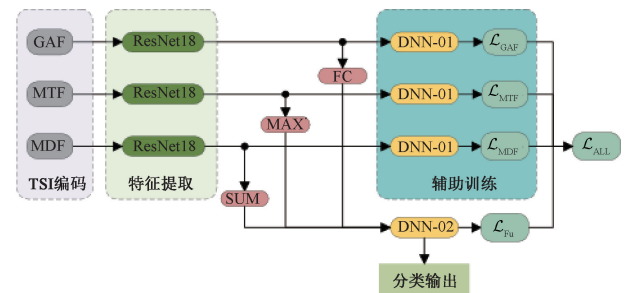
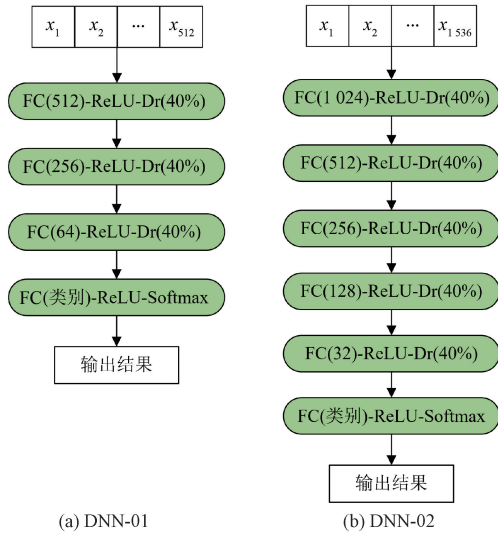


图 4 基于 TSI 的多任务特征融合网络

Figure 4 Multi-task feature fusion network based on TSI  
借助多任务学习的思想, 3 类 TSI 编码的脑电

图像分别送入 ResNet18 特征提取网络抽取图像特征,提取到的图像特征在经过特征融合模块后送入下层 DNN-O2 网络的同时,分别通过 DNN-O1 网络进行辅助训练,最终由 DNN-O2 输出分类结果。DNN-O1 和 DNN-O2 的网络结构如图 5 所示。



(a) DNN-O1

(b) DNN-O2

图 5 DNN-O1 和 DNN-O2 的网络结构

Figure 5 Network structure of DNN-O1 and DNN-O2

特征融合模块采用 3 种方式:特征叠加(SUM)、特征最大化(MAX)以及特征全连接(FC)。其中,SUM 融合假设不同 TSI 编码的重要性一致,在不改变特征向量大小的同时,能够减轻网络特征选择的压力;MAX 融合从高级特征向量中选择最重要的特征,以减少特征维度,突出关键特征,避免噪声和非关键特征对模型的不良影响;FC 融合通过全连接层学习特征之间的非线性关系,动态分配不同特征的权重,增强重要特征的表达能力。3 种融合方式如下所示:

$$\begin{cases} F_{\text{SUM}} = f_{\text{GAF}} + f_{\text{MTF}} + f_{\text{MDF}}; \\ F_{\text{MAX}} = \text{MAX}(f_{\text{GAF}}, f_{\text{MTF}}, f_{\text{MDF}}); \\ F_{\text{FC}} = \text{FC}(f_{\text{GAF}}w_{\text{GAF}} + f_{\text{MTF}}w_{\text{MTF}} + f_{\text{MDF}}w_{\text{MDF}}). \end{cases} \quad (16)$$

式中: $f_{\text{GAF}}$ 、 $f_{\text{MTF}}$ 、 $f_{\text{MDF}}$  为分别从使用 GAF、MTF、MDF 编码后的图像中提取到的特征; $w_{\text{GAF}}$ 、 $w_{\text{MTF}}$ 、 $w_{\text{MDF}}$  分别为对应全连接层中的权重; $F_{\text{SUM}}$ 、 $F_{\text{MAX}}$ 、 $F_{\text{FC}}$  均为融合后的特征。

为了让单个特征和融合后的特征之间保持平衡,设计了一个总损失函数 $\mathcal{L}_{\text{ALL}}$ :

$$\mathcal{L}_{\text{ALL}} = a(\mathcal{L}_{\text{GAF}} + \mathcal{L}_{\text{MTF}} + \mathcal{L}_{\text{MDF}}) + b\mathcal{L}_{\text{Fu}}. \quad (17)$$

式中: $a$  和  $b$  为平衡权重,此处分别设为  $1/3$  和  $1.0$ ,使它们在训练时对网络参数的优化贡献相等; $\mathcal{L}_{\text{GAF}}$ 、 $\mathcal{L}_{\text{MTF}}$ 、 $\mathcal{L}_{\text{MDF}}$  分别为 GAF、MTF、MDF 各自训练时的损失函数,用来辅助优化各自对应的训练模型; $\mathcal{L}_{\text{Fu}}$  为特征融合后的损失函数;整个特征融合网络由 $\mathcal{L}_{\text{ALL}}$  优化。

### 3 实验设置

#### 3.1 情感识别数据集

DEAP 数据集共有 32 名被试者参与,男女各半。实验中采集 40 通道生理信号,其中前 32 个通道为脑电信号,采集位置按照 10-20 系统选择。原始采样率 512 Hz 的信号经降采样预处理至 128 Hz。观看完每段视频后,被试者按 9 点评分量表,从 Valence、Arousal、Dominance、Liking 这 4 个情感维度进行自我评分。

AMIGOS 数据集包含短视频与长视频两类实验,分别有 40 名和 37 名被试者参与。每位被试者在 10-20 系统电极位置布置 14 个电极。观看每段视频后,被试者按 9 点评分量表从 Valence、Arousal、Dominance、Familiarity、Liking 这 5 个维度评分。

DEAP 和 AMIGOS 数据集的数据格式及内容如表 1 所示。

表 1 数据集格式

Table 1 Dataset format

数据集	名称	形状	描述
DEAP	生理信号	40×40×8 064	视频/通道/长度
	标签	40×4	视频/标签
AMIGOS	生理信号	16×16×(2 060~9 000)	视频/通道/长度
	标签	16×5	视频/标签

#### 3.2 数据预处理

为降低低频噪声和个体差异对情感识别的干扰,需要对脑电数据进行基线预处理<sup>[12]</sup>。DEAP 经基线去除和分割处理后,每个样本包含 32 个通道,128 个采样点,每名被试者的采样数据经分割后生成 2 400 个样本。AMIGOS 为 14 个通道,128 个采样点,每名被试者的采样数据生成 1 378 个样本。实验均基于 Valence 和 Arousal 两个情感维度进行,各种情感标签的阈值均为 5( $\leq 5$  表示低, $> 5$  表示高)。

#### 3.3 脑电图像拼接

由于不同通道捕捉的神经活动信息存在差异,为了提高情感识别精度,需要将各通道的 TSI 图像按照对应通道的物理位置进行图像拼接,具体拼接方式如图 6 所示。

研究表明,与情感相关的 EEG 信号主要分布在额叶、颞叶和顶叶<sup>[6]</sup>,因此,DEAP 选取 Fp1、Fp2、F3、F4、P3、P4、F7、FC5、FC1、T7、P7、PO3、PZ、FZ、F8、FC6、FC2、T8、P8、PO4 共 20 个通道;AMIGOS 选取 F7、F3、FC5、T7、P7、P8、T8、FC6、F4、F8 共 10 个通道。

#### 3.4 参数设置

模型训练基于开源深度学习框架 Pytorch,使用

Adam 优化器,学习率设置为  $10^{-4}$ , dropout 比率设置为 0.4。模型评估采用准确率指标,通过十折交叉完成验证。所有实验都在一台运行 Ubuntu Linux 21.04 的服务器上进行,配置为 Intel Xeon (R) Gold 5218 CPU、128 GB RAM 和 NVIDIA GTX 2080Ti 显卡。

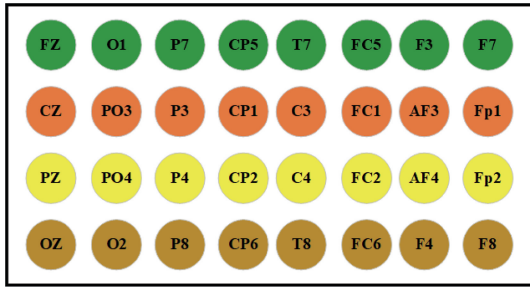


图 6 多通道图像拼接方式

Figure 6 Multi-channel image stitching method

## 4 实验结果与分析

### 4.1 脑电频带

脑电信号在频域上可划分为多个频带,不同频带反映不同的神经活动。情感状态会引起特定频带的变化,合理选择频带有助于提升情感识别的准确性。全频带下不同时间序列成像脑电情感识别的性能如图 7 所示,可以看出,GADF 整体上优于 GASF,MTF 优于 GAF, MDF 在 3 种时间序列算法中表现最佳,且在子序列长度为 4 时效果最优。结果表明,时间序列成像算法可提取稳定且显著的情感特征,即使是浅层的识别模型也能有效学习,体现出时间序列成像算法将脑电信号转化为图像进行分类的优势。

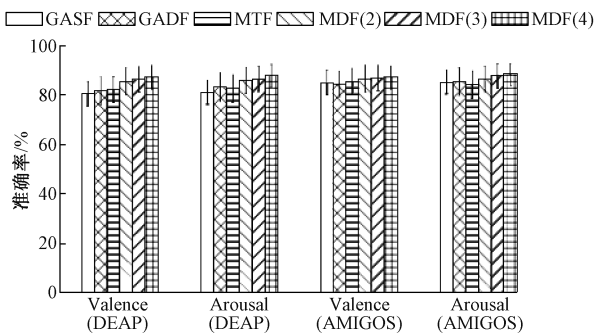


图 7 DEAP 和 AMIGOS 数据集上不同时间序列成像的分类准确率

Figure 7 Classification accuracies of different TSI methods with DEAP and AMIGOS datasets

相关研究表明, Theta (4 ~ 8 Hz)、Alpha (8 ~ 13 Hz)、Beta (13 ~ 30 Hz)、Gamma (30 ~ 50 Hz) 这 4 个频带更能反映情感特性,但其脑电特征信息存在一定的差异<sup>[13]</sup>,需要进一步确定最适合时间序列成像的脑电频带。实验使用零相位巴特沃斯四阶带通滤波器对 DEAP 和 AMIGOS 数据集进行分频,在

基线预处理和通道选择的基础上,选用 MDF(4) 算法将全部通道的脑电数据转化为图像,并用 Res-Net18 进行分类。

图 8 的频带选择实验结果显示,分类准确率随着频率的增加而提升,使用全频段时的分类效果最佳,在 DEAP 和 AMIGOS 数据集上, Valence 和 Arousal 两个维度的平均分类准确率分别为 87.52% 和 87.73%。这表明基于 TSI 的脑电情感识别需要综合不同频率范围的信息。

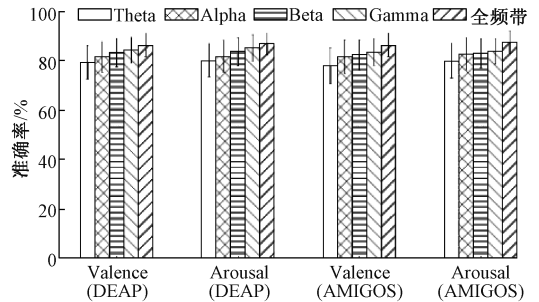


图 8 DEAP 和 AMIGOS 数据集上不同频带的分类准确率  
Figure 8 Classification accuracies across different frequency bands with DEAP and AMIGOS datasets

### 4.2 多尺度处理

传统单一尺度的时域和频域分析无法捕获脑电信号的局部特征。对脑电信号进行多尺度处理<sup>[14]</sup>之后的情感识别结果如图 9 所示。当时间尺度为 2 时,分类效果最佳, DEAP 数据集上 Valence 和 Arousal 两个维度的平均分类准确率分别为 89.33% 和 89.50%; AMIGOS 数据集上分别为 92.37% 和 92.87%。随着时间尺度的增大,分类准确率逐渐降低,这是因为时间尺度过大会导致不同尺度上提取的特征出现重叠,造成信息冗余,信号中的噪声也会随之增加,从而影响分类效果。当时间尺度超过 4 时,分类准确率甚至低于直接使用原始脑电信号,此时局部动态信息被较长的时间尺度所平滑,导致特征信息丢失。

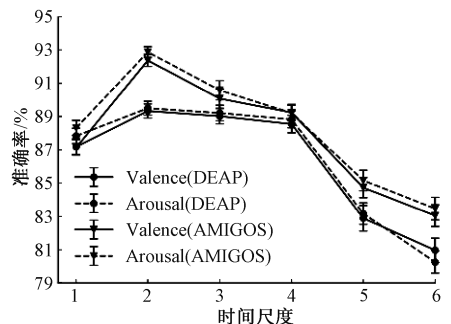


图 9 DEAP 和 AMIGOS 数据集上不同时间尺度的分类准确率

Figure 9 Classification accuracies across different time scales with DEAP and AMIGOS datasets

### 4.3 不同特征融合脑电情感识别结果

为了评估3种特征融合方式对多任务特征融合分类性能的影响,基于4.1节的识别结果,选用GADF、MTF、MDF(4)(记为GMM)和MDF(2)、MDF(3)、MDF(4)(记为MMM)这两种组合进行多任务特征融合实验。通过表2中的分类结果可知,MMM组合优于GMM组合,且无论采用哪种多任务组合方式,使用FC融合进行特征融合的分类效果始终

最好。具体来说,在FC融合方式下,GMM组合在DEAP和AMIGOS上的平均分类准确率分别为91.11%和93.03%,而MMM组合的平均分类准确率分别为94.87%和96.66%,分别提升了3.76个百分点和3.63个百分点。与单模型的最佳分类结果(图8)相比,基于MMM组合的多任务特征融合模型的情感分类在DEAP和AMIGOS上的平均分类准确率分别提高了7.35个百分点和8.93个百分点。

表2 DEAP和AMIGOS数据集上不同特征融合方式的分类准确率

Table 2 Classification accuracies of different feature fusion methods with DEAP and AMIGOS datasets

多任务	特征融合	DEAP 分类准确率/%		AMIGOS 分类准确率/%	
		Valence	Arousal	Valence	Arousal
GMM	SUM	90.15±4.27	90.51±4.11	92.13±3.21	92.26±3.11
	MAX	90.77±4.12	91.02±4.04	92.57±3.14	92.69±3.03
	FC	<b>90.97±3.93</b>	<b>91.24±3.86</b>	<b>92.94±2.86</b>	<b>93.12±2.91</b>
MMM	SUM	92.91±3.87	93.18±3.86	94.47±2.95	94.52±3.12
	MAX	93.35±2.95	94.78±2.63	95.26±2.63	94.28±2.69
	FC	<b>94.51±2.61</b>	<b>95.23±2.43</b>	<b>96.94±2.22</b>	<b>96.37±2.08</b>

### 4.4 数据增强

Mixup数据增强<sup>[15]</sup>能够将不同的样本以一定的比例进行混合,从而产生新的训练样本。图10所示为经过Mixup数据增强后的MMM+FC融合的分类结果,可以看出,随着混合系数的增加,分类准确率也随之提升,在0.4时最高。

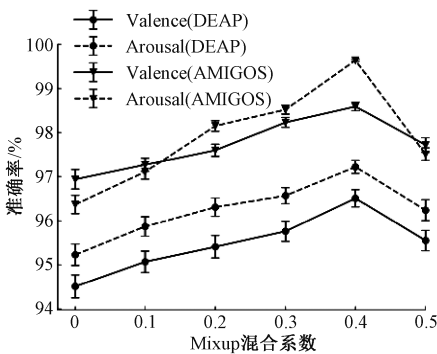


图10 DEAP和AMIGOS数据集上不同Mixup混合系数的分类准确率

Figure 10 Classification accuracies with different Mixup mixing coefficients with DEAP and AMIGOS datasets

### 4.5 与现有模型方法的二分类结果对比

表3对比了多任务特征融合模型与当前常用的、准确率较高的机器学习和深度学习方法的的情感识别分类结果。由表3可知,传统机器学习在DEAP和AMIGOS数据集上的分类准确率普遍低于90%,而结合LSTM、CNN等深度学习模型整体上可有效提升识别性能。

本文提出的基于时间序列成像的脑电情感识别方法取得了较高的分类准确率,TSI基线模型在DEAP和AMIGOS数据集上的平均分类准确率分别超过了89%和92%,与一些基于时频谱图进行分类的方法相比,性能相当。而利用多任务特征融合的平均分类准确率可分别提升至96%以上和98%以上,充分体现了时间序列成像在脑电情感识别任务中的有效性。

脑电情感识别时频特征提取通常依赖较多参数,如脑电信号分帧长度、帧移等,同时存在较多的二维时频特征可供选取。相比之下,时间序列成像算法不仅能有效捕捉脑电信号复杂的时序特征,并且计算简单、参数设置较少,比如GAF不需要额外参数设置;MTF只需要设置转移概率值域划分范围;MDF仅需要设置子片段长度和时间间隔。

### 4.6 四分类和八分类实验结果对比

当前基于脑电信号的情感识别研究通常被视为二分类问题,只能判断具有或不具有特定情感状态的类别。为了进一步探究多任务学习特征融合模型的性能,将识别任务扩展到四分类和八分类。由表4可知,不管是四分类任务还是八分类任务,特征融合方式使用FC时效果均最好。四分类的平均分类准确率均在90%以上,但是八分类的平均分类准确率均有一定的降低。这是因为当类别数越多时,单个类别的数据量就越少,此时数据量限制了模型分类能力。

表 3 不同方法在 DEAP 和 AMIGOS 数据集的平均分类准确率对比

Table 3 Comparison of average classification accuracies of different methods with DEAP and AMIGOS datasets

数据集	类别	方法	输入	平均分类准确率/%	
				Valence	Arousal
DEAP	特征提取+机器学习	DT <sup>[2]</sup>	分频带的 DE 特征	71.63	73.70
		SVM <sup>[6]</sup>	分频带的 DE 特征	89.33	89.99
		MLP <sup>[3]</sup>	DE 特征	87.73	88.88
	高维矩阵+深度学习	CNN <sup>[16]</sup>	二维频域特征矩阵	88.03	88.68
		CNN+LSTM <sup>[12]</sup>	二维电极脑电矩阵	89.15	89.96
		DGCNN <sup>[7]</sup>	DE 特征邻接矩阵	92.55	93.50
		CNN+双向 LSTM <sup>[17]</sup>	四维 DE 特征矩阵	94.02	94.86
	时间序列成像	TSI 基线模型 (本文)	二维 MTF 图像	89.33	89.50
多任务特征融合 (本文)		二维 MMM 图像	96.51	97.22	
AMIGOS	特征提取+机器学习	KNN <sup>[4]</sup>	滑动模态奇异谱特征	85.26	92.01
		RF <sup>[5]</sup>	DE 特征	85.98	83.00
	特征提取+深度学习	FCN <sup>[18]</sup>	预训练 Transformer 特征	85.00	89.00
	高维矩阵+深度学习	CNN <sup>[19]</sup>	全息特征图	92.92	90.76
		3D CNN <sup>[20]</sup>	二维电极拓扑图	96.96	97.52
	时间序列成像	TSI 基线模型 (本文)	二维 MTF 图像	92.37	92.87
		多任务特征融合 (本文)	二维 MMM 图像	98.59	99.64

表 4 四分类和八分类的平均分类准确率

Table 4 Average classification accuracies of four-class and eight-class classification

特征融合	DEAP 平均分		AMIGOS 平均分	
	类准确率/%		类准确率/%	
	四分类	八分类	四分类	八分类
SUM	87.44	83.91	92.75	82.96
MAX	88.85	84.47	95.16	84.61
FC	91.06	87.43	97.41	89.84

### 5 结论

本文提出基于时间序列成像的脑电情感识别方法,利用相位映射、状态转移以及局部差分 3 类策略实现原始脑电信号图像编码,具有较高的计算效率和良好的情感特征表达能力,并结合多任务特征融合框架,实现多策略联合建模,进一步增强时间序列成像的脑电情感识别能力。在二分类、四分类和八分类识别任务中,平均分类准确率分别超过 96%, 91% 和 87%,验证了时间序列成像应用于脑电情感识别的有效性。下一步研究将考虑扩大样本数量,建立多样化的脑电数据库,引入图像分类更优的 Transformer 架构,提高多类别情感识别的分类性能。

### 参考文献:

[1] WANG Y M, ZHANG B, DI L M. Research progress of EEG-based emotion recognition: a survey [J]. ACM Computing Surveys, 2024, 56(11): 1-49.  
 [2] LIU Y, WEI Y, LI C, et al. Bi-CapsNet: a binary cap-

sule network for EEG-based emotion recognition [J]. IEEE Journal of Biomedical and Health Informatics, 2023, 27(3): 1319-1330.  
 [3] YANG Y L, WU Q F, FU Y Z, et al. Continuous convolutional neural network with 3D input for EEG-based emotion recognition[J]. Lecture Notes in Computer Science, 2018, 11307(1):433-443.  
 [4] SHARMA L D, BHATTACHARYYA A. A computerized approach for automatic human emotion recognition using sliding mode singular spectrum analysis [J]. IEEE Sensors Journal, 2021, 21(23): 26931-26940.  
 [5] GALVÃO F, ALARCÃO S M, FONSECA M J. Predicting exact valence and arousal values from EEG [J]. Sensors, 2021, 21(10): 3414.  
 [6] TAO W, LI C, SONG R C, et al. EEG-based emotion recognition via channel-wise attention and self attention [J]. IEEE Transactions on Affective Computing, 2023, 14(1): 382-393.  
 [7] 魏琛, 陈兰岚, 张傲. 基于集成卷积神经网络的脑电情感识别 [J]. 华东理工大学学报 (自然科学版), 2019, 45(4): 614-622.  
 WEI C, CHEN L L, ZHANG A. Emotion recognition of EEG based on ensemble convolutional neural networks [J]. Journal of East China University of Science and Technology, 2019, 45(4): 614-622.  
 [8] SONG T F, ZHENG W M, SONG P, et al. EEG emotion recognition using dynamical graph convolutional neural networks [J]. IEEE Transactions on Affective Computing, 2020, 11(3): 532-541.  
 [9] YANG L Q, LIU J W. EEG-based emotion recognition

- using temporal convolutional network[C]//2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS). Piscataway: IEEE, 2019: 437-442.
- [10] SALAMA E S, EL-KHORIBI R A, SHOMAN M E, et al. EEG-based emotion recognition using 3D convolutional neural networks[J]. *International Journal of Advanced Computer Science and Applications*, 2018, 9(8): 329-337.
- [11] RAMZAN M, DAWN S. Fused CNN-LSTM deep learning emotion recognition model using electroencephalography signals[J]. *International Journal of Neuroscience*, 2023, 133(6): 587-597.
- [12] YANG Y L, WU Q F, QIU M, et al. Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network[C]//2018 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE, 2018: 1-7.
- [13] LU Y F, ZHENG W L, LI B B, et al. Combining eye movements and EEG to enhance emotion recognition[C]//International Joint Conference on Artificial Intelligence. Buenos Aires: IJCAI, 2015: 1170-1176.
- [14] BUSA M A, VAN EMMERIK R E A. Multiscale entropy: a tool for understanding the complexity of postural control[J]. *Journal of Sport and Health Science*, 2016, 5(1): 44-51.
- [15] ZHANG H Y, CISSE M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization[EB/OL]. (2018-04-27)[2025-03-30]. <https://arxiv.org/abs/1710.09412>.
- [16] GAGLIARDI G, ALFEO A L, CATRAMBONE V, et al. Improving emotion recognition systems by exploiting the spatial information of EEG sensors[J]. *IEEE Access*, 2023, 11: 39544-39554.
- [17] CUI F C, WANG R Q, DING W W, et al. A novel DE-CNN-BiLSTM multi-fusion model for EEG emotion recognition[J]. *Mathematics*, 2022, 10(4): 582-593.
- [18] VAZQUEZ-RODRIGUEZ J, LEFEBVRE G, CUMIN J, et al. Emotion recognition with pre-trained transformers using multimodal signals[C]//2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII). Piscataway: IEEE, 2022: 1-8.
- [19] TOPIC A, RUSSO M, STELLA M, et al. Emotion recognition using a reduced set of EEG channels based on holographic feature maps[J]. *Sensors*, 2022, 22(9): 3248.
- [20] ZHAO Y X, YANG J, LIN J L, et al. A 3D convolutional neural network for emotion recognition based on EEG signals[C]//2020 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE, 2020: 1-6.

## Multi-task Learning-driven Emotion Recognition Based on Time Series Imaging

XU Shengxin<sup>1</sup>, LIANG Bizheng<sup>2</sup>, HU Fei<sup>3</sup>, XU Huaxing<sup>3</sup>

(1. National Engineering Center for Risk Perception and Prevention (RPP), China Academy of Electronics and Information Technology, Beijing 100041, China; 2. Institute of Telecommunication and Navigation Satellites, China Academy of Space Technology, Beijing 100094, China; 3. School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China)

**Abstract:** To overcome the high computational complexity of EEG-based emotion recognition methods based on feature extraction or time-frequency representations, a multi-task learning-driven method for emotion recognition based on time series imaging (TSI) was proposed. EEG signals were directly transformed into two-dimensional images using Gramian angular field, Markov transition field, and motif difference field. Built upon the ResNet18 architecture, a multi-task feature fusion framework was designed to integrate features from the three imaging methods to enhance emotional feature representation. Experimental results showed that with the DEAP dataset, the proposed method achieved average classification accuracies of 96.51% and 97.22% for binary classification of Valence and Arousal, respectively, and with the AMIGOS dataset, the accuracies reached 98.59% and 99.64%. When extended to four-class and eight-class classification tasks, the proposed method achieved average accuracies of 91.06% and 87.43% with DEAP, and 97.41% and 89.84% with AMIGOS, respectively. These results demonstrated the robustness of the proposed method in EEG-based emotion recognition.

**Keywords:** EEG; emotion recognition; time series imaging; multi-task; feature fusion