

基于条件双重对齐注意力U-Net的跨被试脑电情绪识别方法

许子豪, 廖志强, 赵孟君

五邑大学电子与信息工程学院, 广东 江门

收稿日期: 2026年3月26日; 录用日期: 2026年4月23日; 发布日期: 2026年4月30日

摘要

针对跨被试脑电情绪识别中的个体差异与特征分布偏移问题, 本文提出一种条件双重对齐注意力U-Net (CDA-AUNet)模型。该模型由脑电特征编码器与双重域对齐模块构成。特征编码阶段构建了融合压缩-激励(SE)通道注意力机制的U-Net架构, 以实现多频带和空间通道特征的自适应加权; 并在深层引入信息瓶颈机制以抑制生理噪声、缓解过拟合。在跨域对齐阶段, 模型结合条件域对抗网络(CDAN)与相关对齐(CORAL)算法, 通过特征与预测概率的张量外积对齐联合条件分布, 并在二阶协方差维度约束特征偏移, 实现源域与目标域特征的深度协同对齐。在SEED数据集上的留一被试交叉验证结果表明, CDA-AUNet的平均分类准确率达到85.18%, 显著优于现有主流基线模型。消融实验进一步验证了各核心模块在提升跨被试泛化能力中的有效性。

关键词

脑电信号, 情绪识别, 跨被试, 条件双重对齐, 注意力U-Net

Cross-Subject EEG Emotion Recognition Method Based on Conditional Dual-Aligned Attention U-Net

Zihao Xu, Zhiqiang Liao, Mengjun Zhao

School of Electronics and Information Engineering, Wuyi University, Jiangmen Guangdong

Received: March 26, 2026; accepted: April 23, 2026; published: April 30, 2026

Abstract

To address the challenges of individual differences and feature distribution shifts in cross-subject

文章引用: 许子豪, 廖志强, 赵孟君. 基于条件双重对齐注意力 U-Net 的跨被试脑电情绪识别方法[J]. 计算机科学与应用, 2026, 16(4): 563-573. DOI: 10.12677/csa.2026.164153

EEG-based emotion recognition, this paper proposes a Conditional Dual-Aligned Attention U-Net (CDA-AUNet) model. The proposed model consists of an EEG feature encoder and a dual-domain alignment module. During the feature encoding phase, a U-Net architecture integrated with a Squeeze-and-Excitation (SE) channel attention mechanism is constructed to perform adaptive weighting across multi-band and spatial channel features. Furthermore, an information bottleneck mechanism is introduced in the deep layers to suppress physiological noise and mitigate overfitting. During the cross-domain alignment phase, the model integrates a Conditional Domain Adversarial Network (CDAN) with the Correlation Alignment (CORAL) algorithm. By utilizing the tensor outer product of deep features and prediction probabilities to align the joint conditional distribution, and constraining feature shifts at the second-order covariance dimension, the model achieves deep collaborative alignment of feature distributions between the source and target domains. Leave-one-subject-out (LOSO) cross-validation results on the SEED dataset demonstrate that the CDA-AUNet achieves an average classification accuracy of 85.18%, significantly outperforming current mainstream baseline models. Ablation studies further validate the effectiveness of the core modules in enhancing cross-subject generalization capabilities.

Keywords

EEG, Emotion Recognition, Cross-Subject, Conditional Dual Alignment, Attention U-Net

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

情绪是个体对客观环境刺激产生的主观体验与复杂的生理反应，在人类的认知处理与行为决策中发挥着重要作用[1]。在人机交互(Human-Computer Interaction, HCI)领域，赋予设备感知人类情感的能力是当前的研究热点。相比于易受主观控制的面部表情与语音信号，脑电图(Electroencephalogram, EEG)信号能够直接反映中枢神经系统活动，具备较高的客观性与可靠性，已成为构建情感脑机接口(Affective Brain-Computer Interface, aBCI)系统的重要生理数据源[2] [3]。

然而，脑电情绪识别领域面临多重技术挑战。首先，脑电信号本身具有微弱、低信噪比及非平稳等特性，在采集过程中易受到眼电(EOG)、肌电(EMG)以及外界工频干扰的影响，导致有效情感信息被掩盖。其次，跨被试变异性(Cross-subject Variability)是当前制约模型泛化的核心瓶颈。由于不同被试在大脑解剖结构与认知模式上存在显著差异，导致跨被试场景下脑电特征存在严重的分布偏移(Domain Shift) [4] [5]。因此，如何克服跨被试脑电特征的分布偏移，构建具有较强泛化能力的跨被试情感识别模型，是当前脑机接口领域亟待突破的技术难点[6]。

在脑电特征提取方面，早期研究主要依赖信号处理与传统机器学习算法。研究人员多从时域、频域或时频域中提取手工特征[7] [8]。例如，利用功率谱密度(PSD)量化各频段能量，或通过提取微分熵(Differential Entropy, DE)度量脑电信号在特定频段内的生理复杂度。然而，传统机器学习方法依赖人工特征设计，且浅层分类模型(如支持向量机、K近邻、随机森林等)难以充分挖掘高维脑电信号中复杂的空间关联信息，致使其在跨被试任务中的泛化能力受限。

近年来，深度学习技术通过端到端框架，在脑电信号解码中得到广泛应用。其中，卷积神经网络(CNN)常被用于局部特征提取[9]。例如，EEGNet结合深度可分离卷积和逐点卷积，在较小参数量下提取时空特征[10]；ShallowConvNet则针对振荡信号设计大尺度卷积层以提取频带能量信息[11]。但在处理复杂脑电

信号时,传统 CNN 仍存在一定局限性。一方面,脑电信号具有多尺度特性,单一固定感受野的卷积算子较难全面提取复杂的脑电模式信息;另一方面,由于脑电信号信噪比低,常规深度卷积网络易提取到大量与情感任务无关的冗余信息。这不仅增加了计算开销,也容易对有效的情感表征造成干扰。因此,在特征提取阶段引入注意力机制与特征重构模块,以抑制生理噪声并增强有效特征的表达,是现阶段优化的方向之一[12]。

为解决跨被试特征的分布偏移问题,迁移学习与无监督域适应(Unsupervised Domain Adaptation, UDA)技术常被用于提升模型的跨域泛化能力。UDA 旨在无目标域标签的情况下,利用源域数据对齐两者的特征分布。目前,基于对抗学习的域对抗神经网络(Domain-Adversarial Neural Network, DANN)应用较为广泛。DANN 通过引入梯度反转层(GRL),使特征提取器与域判别器进行对抗训练,促使网络学习域不变特征,从而在一定程度上实现源域与目标域特征在一阶边缘分布(Marginal Distribution)上的对齐[13]。

然而,基于对抗的域适应算法通常假设源域与目标域具有相似的特征流形,仅依赖一阶边缘分布对齐存在一定局限。在跨被试脑电情绪识别任务中,个体差异不仅体现在整体分布的偏移,其类内的二阶统计特性(如条件协方差结构)也存在差异。DANN 等模型在进行全局特征对齐时,易忽略数据的局部结构信息,在优化对抗损失的过程中可能会改变原有的特征空间分布,导致目标域样本发生错分。这种单一对齐边缘分布的方法容易破坏局部条件分布的一致性,进而引发“负迁移”现象并导致识别性能下降[14]。因此,如何在减小域间整体分布差异的同时,保持特征空间的拓扑结构与二阶统计特征一致性以缓解负迁移问题,是当前跨被试域适应研究中需要解决的关键难点。

针对传统卷积网络易引入信息冗余,以及单一对抗对齐在跨被试迁移中易模糊类别边界而引发负迁移的问题,本文提出了一种条件双重对齐注意力 U-Net (CDA-AUNet)模型。该模型分别从特征重构与高阶域适应层面进行改进,旨在跨被试场景下实现更稳健的情绪解码。

在脑电特征提取阶段,本研究构建了融合通道注意力机制(Squeeze-and-Excitation, SE)的 U-Net 特征编码器。该模块通过计算通道间的相关性,对多频带、多通道脑电特征进行自适应加权,以突出情绪相关特征并抑制生理噪声。此外,网络深层引入了信息瓶颈(Information Bottleneck)结构,利用降维机制过滤被试的特异性噪声,提升了深层特征的判别能力。

在跨域特征迁移阶段,本研究设计了结合条件域对抗网络(Conditional Domain Adversarial Network, CDAN)与相关对齐(Correlation Alignment, CORAL)的双重对齐机制。其中,CDAN 分支将深度特征与分类器预测概率进行多线性映射(外积运算),实现源域与目标域联合条件分布的对齐;CORAL 分支则通过最小化两域特征矩阵的协方差距离,在二阶统计维度上约束特征的相关性结构。该联合策略弥补了单一边缘分布对齐在保持类别边界上的局限性,在减小域间差异的同时保持了特征空间的几何拓扑结构,进而缓解了跨域迁移中的负迁移现象。

本研究的主要研究贡献如下:

(1) 设计了一种基于注意力 U-Net 的脑电特征提取模型。结合 SE 通道注意力与深层信息瓶颈结构,实现了对高维脑电信号的自适应特征筛选与空间通道重构,降低了传统卷积提取过程中的信息冗余,提高了特征的抗噪性能。

(2) 提出了一种基于 CDAN 与 CORAL 的条件双重对齐域适应机制。通过多线性映射对齐联合条件分布,并辅以二阶协方差结构约束,有效改善了单一分布对齐在跨被试迁移中易产生的特征混淆现象,为解决负迁移问题提供了一种联合优化方案。

(3) 在公开脑电情感数据集 SEED 上开展了留一被试交叉验证(LOSO)实验。结果表明,CDA-AUNet 模型有效提升了跨被试情绪分类准确率,优于多种现有基线模型。

2. 实验设置

2.1. 数据集描述

本研究采用的数据集为上海交通大学 BCMI 实验室公开发布的 SEED 情感数据集[15]。该数据集包含了 15 名健康被试在观看视频刺激时的脑电(EEG)数据。每位被试参与了三次不同时间段的实验会话(Session), 每次观看 15 段诱发情绪的电影片段。该数据集将情绪划分为三种基本状态: 积极、中性和消极。

由于原始脑电信号具有非平稳性, 且易受肌电、眼电及工频的干扰, 本研究对多源域脑电数据进行了统一的去噪预处理与特征提取。在特征选择上, 本研究采用微分熵(Differential Entropy, DE)作为核心特征。原始信号经过去噪与下采样后, 通过短时傅里叶变换提取五个标准频带(Delta、Theta、Alpha、Beta 和 Gamma)的 DE 特征。最终, 每个样本提取的 5 个频带特征按 62 个电极通道进行组织, 构建为维度为 62×5 (通道数 \times 频带数)的二维特征矩阵, 作为 CDA-AUNet 模型的输入。

2.2. 实验环境设置

为评估模型在跨被试场景下的泛化性能, 本研究采用留一被试(Leave-One-Subject-Out, LOSO)交叉验证协议。实验基于 PyTorch 深度学习框架实现, 并在单张 NVIDIA GeForce RTX 3070 GPU 上完成训练。模型训练的批次大小(Batch Size)设置为 128, 训练总轮次(Epochs)为 50。网络参数的更新统一采用 AdamW 优化器, 初始学习率设为 0.0002, 并引入系数为 0.001 的 L2 正则化以缓解模型在复杂脑电特征上的过拟合现象。此外, 学习率采用余弦退火衰减策略(Cosine Annealing)进行动态调整, 以保障网络在训练后期的平稳收敛。在联合损失优化中, 条件域对抗损失与相关对齐损失的权重分别固定为 1.0 与 0.5。为保证对比的公平性, 所有基线模型均在同一硬件环境与评估协议下运行, 其核心超参数遵循原始文献设定。

2.3. 对比模型

SVM [16]: SVM 是一种经典的传统机器学习基线模型。它通过在特征空间中寻找最大间隔超平面来实现情绪类别的划分, 代表了基于手工特征工程的传统分类范式。

CNN [11]: CNN 是一种基础的深度学习解码模型。本研究参考经典卷积网络架构, 将一维卷积核作用于频带特征维度以提取局部频谱规律, 克服了浅层分类器的局限性。

DANN [13]: DANN 是一种无监督域适应网络模型。它通过引入梯度反转层(GRL), 使特征提取器与域判别器进行对抗博弈, 强制网络学习跨被试的“域不变特征”, 旨在缓解脑电信号的个体差异与分布偏移问题。

DGCNN [17]: DGCNN 是一种基于图神经网络的脑电特征建模方法。该模型将脑电通道视为图节点, 通过动态学习不同脑区(电极通道)间的内在拓扑连接矩阵, 从而有效挖掘脑电信号中复杂的空间关联特征。

DGGN [18]: DGGN 是一种融合图卷积神经网络(GCNN)、长短期记忆网络(LSTM)与对抗学习的跨被试特征解码模型, 通过联合挖掘复杂的空间拓扑特征并对齐域间分布, 以减小跨被试数据间的分布偏移。

3. 模型架构

针对跨被试脑电(EEG)情感识别中存在的分布偏移与个体差异问题, 本研究提出了一种条件双重对齐注意力 U-Net (Conditional Dual-Aligned Attention U-Net, CDA-AUNet)模型。如图 1 所示, 该模型主要由三个核心模块构成: 基于注意力机制的特征提取编码器、条件域对抗网络(CDAN)分支以及二阶协方差对齐(CORAL)分支。模型旨在通过联合一阶条件分布与二阶统计特征的协同对齐, 在提取高判别性情感特征的同时有效缓解跨域负迁移现象。

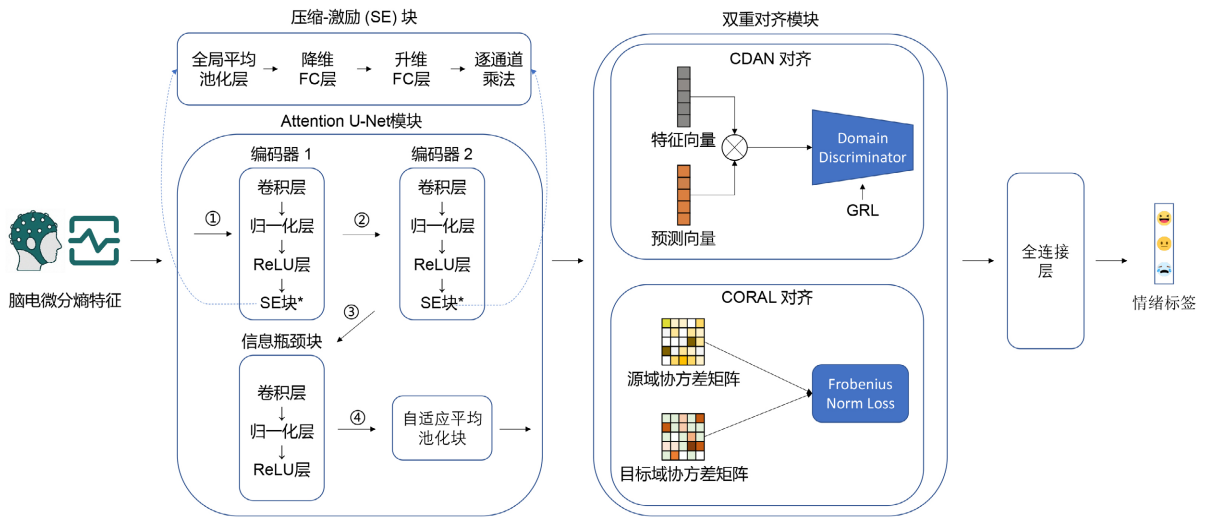


Figure 1. The overall architecture of CDA-AUNet
图 1. CDA-AUNet 整体模型框架图

3.1. 基于注意力 U-Net 的特征提取编码器

脑电信号在频域和空域上具有多尺度动态特征。其中，高频带(如 Beta、Gamma)多与高唤醒度的情绪状态相关，低频带(如 Delta、Theta)则主要反映基准认知活动。传统的浅层或单一尺度卷积难以充分提取此类跨频带的非线性关联信息。为此，本研究设计了融合通道注意力机制的一维特征编码器(Attention U-Net Encoder)。该模块沿用 U-Net 的收缩路径(Contracting Path)，利用逐层下采样与多尺度卷积，将局部频带能量映射为全局情绪表征。此外，其层次化结构为通道注意力机制提供了多尺度感受野，使模型能够自适应量化多频带及空间通道的特征权重，从而实现针对脑电生理特性的有效特征编码。

设输入的脑电信号为 $X \in \mathbb{R}^{b \times c \times f}$ ，其中 b 为批次大小， c 为电极通道数， f 为频带特征维度。在本模型中，一维卷积核沿特征的频带维度进行局部感知，以提取不同频带间的非线性关联。在收缩路径中，模型通过两层通道数递增(通道数分别设为 c_1 和 c_2)的卷积模块对特征进行映射，并通过逐层下采样扩大感受野，以提取多尺度空频特征。

为增强特征的判别性，本研究在每个卷积与批归一化(Batch Normalization)层后引入 SE 通道注意力模块。对于输入 SE 模块的特征图 $U \in \mathbb{R}^{b \times c \times l}$ ，首先通过全局平均池化(Global Average Pooling)将其在空间维度上压缩，获得通道描述符 $v \in \mathbb{R}^{c'}$ ：

$$v_c = \frac{1}{l} \sum_{i=1}^l U_c(i) \tag{1}$$

随后，通过包含两个全连接层的门控机制学习通道间的非线性依赖关系，并利用 Sigmoid 函数生成各通道的激活权重 s ：

$$s = \sigma(W_2 \delta(W_1 v)) \tag{2}$$

其中， δ 为 ReLU 激活函数， $W_1 \in \mathbb{R}^{\frac{c'}{r} \times c'}$ 和 $W_2 \in \mathbb{R}^{c' \times \frac{c'}{r}}$ 为可学习的权重矩阵， r 为降维比例。最终，将权重 s 与原始特征图 U 按通道相乘，得到重标定后的特征 \tilde{U} 。该机制通过特征重标定，实现了多频带及空间通道特征的自适应加权。

在网络深层的信息瓶颈(Information Bottleneck)处，本研究采用自适应平均池化层(AdaptiveAvgPool1d)

将多维特征降维为紧凑的特征向量 $\mathbf{z} \in \mathbb{R}^d$ 。该降维结构旨在过滤个体特异性高频噪声，提取全局情感特征表征。

3.2. 条件双重特征对齐机制

获取深度特征 \mathbf{z} 后，引入基于 CDAN 与 CORAL 的条件双重特征对齐模块，以减小源域与目标域之间的数据分布差异。

(1) 基于 CDAN 的联合条件分布对齐

常规的域对抗神经网络(DANN)主要对齐源域与目标域的特征边缘分布 $P(\mathbf{z})$ ，未考虑类别条件分布 $P(\mathbf{z}|y)$ 的差异，可能导致跨域迁移时类别边界重叠。本研究采用条件域对抗网络(CDAN)进行改进。

特征 \mathbf{z} 经分类器输出的情感类别预测概率 $p \in \mathbb{R}^K$ (K 为类别数)计算如下：

$$p = \text{Softmax}(W_{cls}\mathbf{z} + b_{cls}) \quad (3)$$

对深度特征 \mathbf{z} 与预测概率 p 进行外积运算，生成条件联合特征 $\mathbf{h} \in \mathbb{R}^{d \times K}$ ：

$$\mathbf{h} = \mathbf{z} \otimes p \quad (4)$$

外积特征 \mathbf{h} 包含了特征表示与语义预测的交互信息。将其输入域判别器 \mathcal{D} ，并结合梯度反转层(Gradient Reversal Layer, GRL)进行对抗训练。CDAN 的对抗损失函数 \mathcal{L}_{cdan} 定义为：

$$\mathcal{L}_{cdan} = -\frac{1}{n_s} \sum_{i=1}^{n_s} \log(D(\mathbf{h}_i^s)) - \frac{1}{n_t} \sum_{j=1}^{n_t} \log(1 - D(\mathbf{h}_j^t)) \quad (5)$$

其中， n_s 与 n_t 分别为源域和目标域的样本数， \mathbf{h}_i^s 与 \mathbf{h}_j^t 分别为对应源域与目标域的条件联合特征。

(2) 基于 CORAL 的二阶协方差分布约束

对抗学习在全局对齐时，往往会改变特征的局部空间结构。为此，本研究采用相关对齐(CORAL)算法对特征的二阶统计量进行约束，以维持原有的数据分布形态。

设 $\mathbf{M}_s \in \mathbb{R}^{n_s \times d}$ 与 $\mathbf{M}_t \in \mathbb{R}^{n_t \times d}$ 分别为源域与目标域经过编码器提取的特征矩阵。源域与目标域的协方差矩阵 Σ_s 与 Σ_t 计算如下：

$$\Sigma_s = \frac{1}{n_s - 1} \left(\mathbf{M}_s - \frac{1}{n_s} \mathbf{1}\mathbf{1}^T \mathbf{M}_s \right)^T \left(\mathbf{M}_s - \frac{1}{n_s} \mathbf{1}\mathbf{1}^T \mathbf{M}_s \right) \quad (6)$$

$$\Sigma_t = \frac{1}{n_t - 1} \left(\mathbf{M}_t - \frac{1}{n_t} \mathbf{1}\mathbf{1}^T \mathbf{M}_t \right)^T \left(\mathbf{M}_t - \frac{1}{n_t} \mathbf{1}\mathbf{1}^T \mathbf{M}_t \right) \quad (7)$$

其中， $\mathbf{1}$ 为与样本数对应维度的全 1 列向量。CORAL 损失 \mathcal{L}_{coral} 通过计算源域与目标域协方差矩阵的 Frobenius 范数平方，以衡量域间二阶分布差异：

$$\mathcal{L}_{coral} = \frac{1}{4d^2} \|\Sigma_s - \Sigma_t\|_F^2 \quad (8)$$

通过最小化 \mathcal{L}_{coral} ，模型能够有效对齐特征的协方差结构，保持特征流形的拓扑一致性。

3.3. 联合损失函数与优化策略

模型的总体损失函数 \mathcal{L}_{total} 由分类损失 \mathcal{L}_{cls} 、条件域对抗损失 \mathcal{L}_{cdan} 与二阶协方差损失 \mathcal{L}_{coral} 三部分联合构成。其中分类损失 \mathcal{L}_{cls} 采用标准的交叉熵损失函数计算源域样本的分类误差。总体损失函数 \mathcal{L}_{total} 定义如下：

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \lambda_{adv} \mathcal{L}_{cdan} + \lambda_{coral} \mathcal{L}_{coral} \quad (9)$$

其中, λ_{adv} 与 λ_{coral} 为平衡各损失项的超参数。

此外, 为防止模型在训练初期因分类器预测概率 p 不准确而导致对抗域适应产生负迁移, 本研究在 GRL 中引入了动态退火衰减因子 α :

$$\alpha = \frac{2}{1 + \exp(-\gamma \cdot \tau)} - 1 \quad (10)$$

其中, $\tau \in [0, 1]$ 为当前训练轮次占总轮次的比例, γ 为控制退火速率的常数。在训练初期, α 接近于 0, 模型主要侧重于分类任务的学习; 随着训练深入, α 逐渐趋近于 1, 域对抗机制逐步主导特征对齐过程, 以促进模型平稳收敛。

4. 实验结果

各模型在 SEED 数据集上进行留一被试(LOSO)交叉验证的平均准确率如表 1 所示。结果表明, CDA-AUNet 模型取得了 85.18% 的平均分类准确率, 显著优于其他对比基线模型。相较于表现次优的图网络基线模型 DGCNN (77.12%) 以及经典深度学习模型 CNN (73.78%), CDA-AUNet 分别实现了 8.06% 和 11.40% 的性能提升。此外, 与传统机器学习方法 SVM (74.42%) 相比, CDA-AUNet 也表现出更好的分类性能。

在域适应策略的有效性验证方面, CDA-AUNet 的准确率相较于基础对抗模型 DANN (75.01%) 提升了 10.17%。结果表明, 在跨被试脑电解码任务中, 单一的边缘分布对齐容易导致特征边界模糊; 本研究引入的条件双重对齐机制(CDAN + CORAL)通过联合条件分布与二阶协方差结构的协同对齐, 更有效地缓解了跨域负迁移现象, 从而大幅提升了模型在面对未知被试时的泛化能力。

Table 1. Classification accuracy of different models on the SEED dataset
表 1. 各模型在 SEED 数据集上的分类准确率

模型	准确率(%)
SVM	74.42/6.77
CNN	73.78/7.54
DANN	75.01/9.12
DGCNN	77.12/6.62
DGGN	72.41/6.63
CDA-AUNet	85.18/7.67

为验证本模型(CDA-AUNet)中各个关键模块的作用, 本研究进行了消融实验如表 2 所示。实验结果表明, 当完全移除域适应机制时, 模型性能下降最为明显(降低了 6.02%)。在双重对齐组件中, 仅移除条件域对抗网络(w/o CDAN)导致分类准确率下降了 3.62%, 而仅移除相关对齐模块(w/o CORAL)使性能下降了 1.67%。这表明这两种对齐策略均对提升跨被试泛化能力做出了积极贡献, 且 CDAN 模块在其中发挥着主导作用。

此外, 为了验证条件联合分布对齐的优势, 本研究移除条件联合策略(w/o Cond), 即将 CDAN 模块退化为传统的边缘分布对齐模块(DANN + CORAL)。结果显示, 在缺失条件语义信息的情况下, 模型准确率下降了 3.06%, 这直接证明了条件多线性映射在防止跨域特征混淆方面的必要性。最后, 在特征提取阶段, 移除通道注意力模块(w/o SE-Att)也导致了 1.81% 的性能损失。这些结果共同表明, 双重对齐机制与注意力特征提取的协同运行大幅提升了跨被试脑电情绪识别的性能。

Table 2. Ablation study results of CDA-AUNet on the SEED dataset
表 2. CDA-AUNet 在 SEED 数据集上进行的消融研究结果

模型	准确率(%)
w/o 域适应机制	79.16/7.79
w/o 条件域对抗网络	81.56/7.82
w/o 条件联合策略	82.12/8.01
w/o 通道注意力模块	83.37/6.53
w/o 相关对齐模块	83.51/8.21
CDA-AUNet	85.18/7.67

图 2 展示了 CDA-AUNet 模型的混淆矩阵。结果表明，CDA-AUNet 模型对积极(Positive)和消极(Negative)情绪的识别具有较高的敏感性，准确率分别达到了 89.71% 和 83.75%。相比之下，中性(Neutral)情绪的分类准确率略低(81.82%)，其主要误差源于与消极状态的特征重叠。具体而言，有 14.59% 的中性样本被误判为消极情绪，同时有 12.04% 的消极样本被错误归类为中性。这一现象表明，在跨被试场景中，消极与中性情绪的神经生理特征存在一定的重叠与相似性。综上所述，该模型在维持高准确率的同时，在各个情绪类别上均展现出了较高的判别能力。

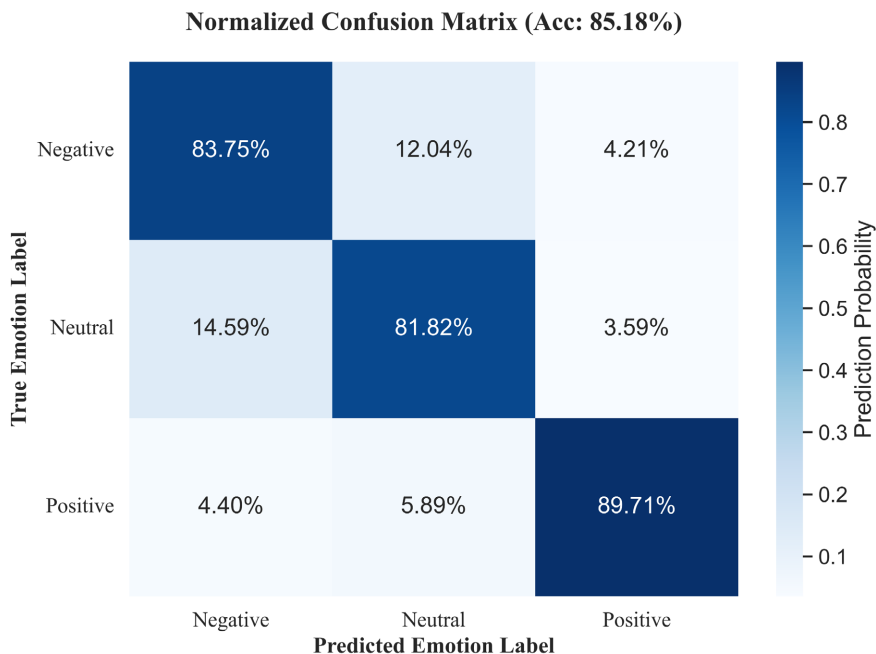


Figure 2. Confusion matrix of the CDA-AUNet
图 2. CDA-AUNet 的混淆矩阵图

为进一步展示模型在特征提取与跨域对齐上的效果，本研究将原始脑电特征与 CDA-AUNet 模型全连接层前提取的高维特征通过 t-SNE 投影至二维空间[19]，如图 3 所示。从图 3 左侧的原始脑电特征分布可以看出，未经模型处理的特征在不同情绪类别间相互交叠，缺乏较好的可分性；此外，相同类别下源域与目标域的样本存在明显的空间偏移，反映出跨被试脑电数据之间存在分布差异。图 3 右侧展示了经

过 CDA-AUNet 模型处理后提取的深层语义特征分布。由图可知, 特征在不同类别间表现出更好的可分性, 形成了三个相对清晰的聚类簇; 且在每个情绪类别簇内, 源域样本与目标域样本的重叠度较高。可视化结果表明, 本研究提出的条件双重对齐机制(CDAN + CORAL)能够有效缓解被试间的特征分布差异, 并在跨域场景下学习到更具区分性的情感特征表示。

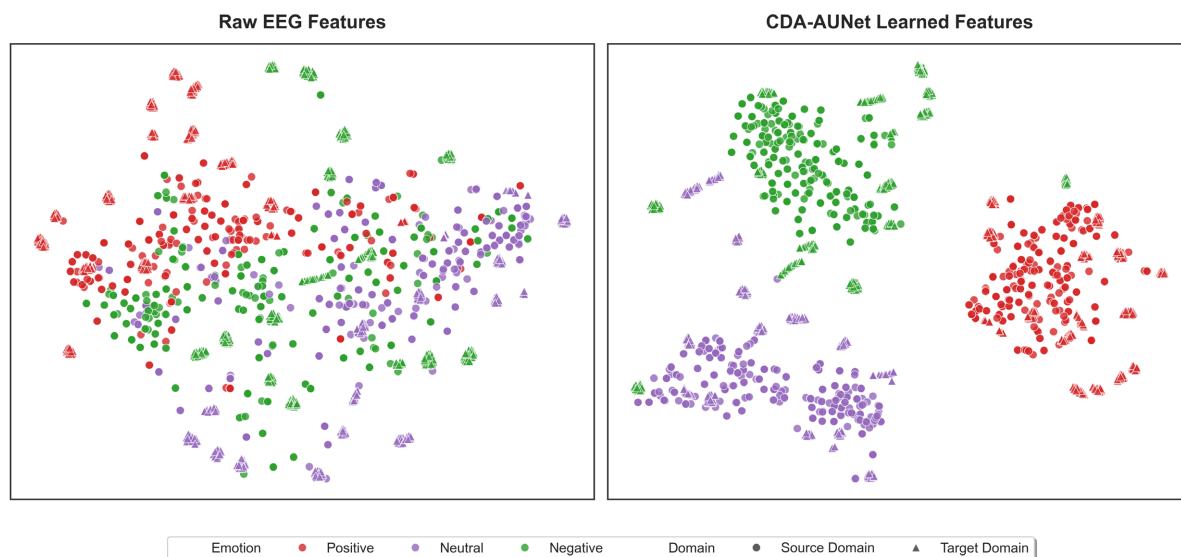


Figure 3. Model feature extraction visualization (Subject 6)

图 3. 模型提取特征可视化(被试 6)

5. 讨论

5.1. 核心模块的协同作用分析

CDA-AUNet 模型通过整合特征提取与域适应模块, 构建了网络结构的联合优化框架。在跨被试情绪识别任务中, 针对脑电信号个体差异显著及背景生理噪声高的问题, 直接全局域对齐易破坏原始特征的空间结构。为此, 模型前端引入 SE 通道注意力机制对空频特征进行自适应加权, 抑制冗余频带与通道信息, 从而提升输入域适应阶段的特征信噪比。

在跨域对齐阶段, 为解决单一边缘分布对齐(如 DANN)易导致目标域分类边界模糊的问题, 模型联合引入 CDAN 与 CORAL 约束。具体而言, CDAN 利用多线性映射捕获深度特征与预测概率的条件依赖关系, 实现联合条件分布对齐; 同时, 引入 CORAL 作为二阶协方差正则化项, 限制对齐过程中的特征偏移, 以防止对抗学习破坏特征局部结构。通过上述特征重标定与双重分布对齐策略, CDA-AUNet 有效缓解了单一边缘对齐引发的负迁移现象。

5.2. 模型可解释性与神经生理学分析

为探究模型决策的空间特征依据, 本研究基于输入梯度的显著性映射方法提取脑电通道判别权重, 并绘制三种情绪状态下的空间显著性地形图。如图 4 所示, CDA-AUNet 在解码不同效价情绪时, 呈现出差异化的空间拓扑模式。

消极情绪下, 显著性权重集中于右侧前额叶及大范围顶枕叶区域。顶枕叶的强激活与 SEED 数据集的视频诱发范式直接相关: 被试处理负性刺激时伴随较高认知负荷, 需调用更多视觉注意力与感觉统合资源, 进而引发脑后部视觉皮层及相关神经网络的强烈响应。同时, 右侧前额叶的权重分布特征符合情

绪神经科学的“右半球侧化假说”，即负面情感加工主要由右半球主导。作为对比，积极情绪的空间激活模式更为局限，证明模型有效提取了效价特异性的神经表征。

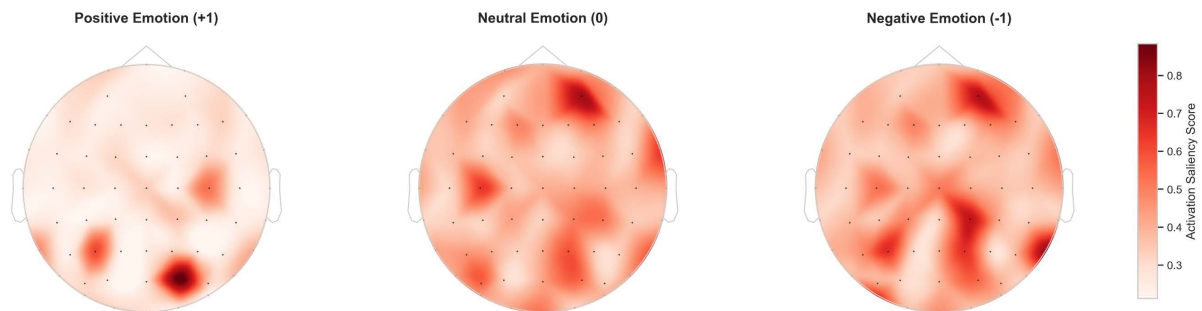


Figure 4. Spatial saliency topomaps of features

图 4. 特征空间显著性地形图

对于实验中易混淆的中性与消极情绪，地形图显示两者全局能量分布相似，但在局部极值点(如右侧顶枕叶及部分额叶节点激活强度)差异显著。实验结果表明，在跨被试场景下，模型主要依赖关键脑区的拓扑边界差异进行决策，而非受限于全局能量均值。

上述空间显著性分析结果与认知神经科学理论相符，进一步证实 CDA-AUNet 提取的特征在提升分类性能的同时，具备神经生理学合理性。

6. 结论

针对跨被试脑电情绪识别中的信息冗余与数据分布偏移问题，本文提出了一种条件双重对齐注意力 U-Net (CDA-AUNet)模型。在特征提取阶段，该模型引入通道注意力机制与信息瓶颈结构以抑制背景噪声，提升输入特征的信噪比；在域适应阶段，模型联合使用 CDAN 与 CORAL 进行分布对齐。其中，CDAN 通过多线性映射实现联合条件分布对齐，CORAL 作为二阶协方差正则化项限制特征偏移。该双重对齐策略有效缓解了单一边缘分布对齐易引发的负迁移现象。在 SEED 数据集上的跨被试交叉验证结果表明，CDA-AUNet 的平均准确率达到 85.18%，优于现有主流基线模型。消融实验证实了各模块的性能增益，特征可视化分析进一步验证了模型决策特征与神经生理学先验的一致性。本研究表明，CDA-AUNet 具备良好的跨被试泛化能力，可为情感脑机接口的跨域解码算法设计提供有效参考。

参考文献

- [1] Adolphs, R., Mlodinow, L. and Barrett, L.F. (2019) What Is an Emotion? *Current Biology*, **29**, R1060-R1064. <https://doi.org/10.1016/j.cub.2019.09.008>
- [2] Ahern, G.L. and Schwartz, G.E. (1985) Differential Lateralization for Positive and Negative Emotion in the Human Brain: EEG Spectral Analysis. *Neuropsychologia*, **23**, 745-755. [https://doi.org/10.1016/0028-3932\(85\)90081-8](https://doi.org/10.1016/0028-3932(85)90081-8)
- [3] Alarcao, S.M. and Fonseca, M.J. (2019) Emotions Recognition Using EEG Signals: A Survey. *IEEE Transactions on Affective Computing*, **10**, 374-393. <https://doi.org/10.1109/taffc.2017.2714671>
- [4] Kamrud, A., Borghetti, B. and Schubert Kabban, C. (2021) The Effects of Individual Differences, Non-Stationarity, and the Importance of Data Partitioning Decisions for Training and Testing of EEG Cross-Participant Models. *Sensors*, **21**, Article 3225. <https://doi.org/10.3390/s21093225>
- [5] Yin, Y., Zheng, X., Hu, B., Zhang, Y. and Cui, X. (2021) EEG Emotion Recognition Using Fusion Model of Graph Convolutional Neural Networks and LSTM. *Applied Soft Computing*, **100**, Article 106954. <https://doi.org/10.1016/j.asoc.2020.106954>
- [6] Chen, H., He, H., Zhu, J., et al. (2022) Weight-Based Channel-Model Matrix Framework Provides a Reasonable Solution for EEG-Based Cross-Dataset Emotion Recognition. arXiv:2209.05849.

-
- [7] Yu, M., Zhang, D., Zhang, G., Zhao, G., Liu, Y., Han, Y., *et al.* (2019) A Review of EEG Features for Emotion Recognition. *Scientia Sinica Informationis*, **49**, 1097-1118. <https://doi.org/10.1360/n112018-00337>
- [8] Jenke, R., Peer, A. and Buss, M. (2014) Feature Extraction and Selection for Emotion Recognition from EEG. *IEEE Transactions on Affective Computing*, **5**, 327-339. <https://doi.org/10.1109/taffc.2014.2339834>
- [9] Craik, A., He, Y. and Contreras-Vidal, J.L. (2019) Deep Learning for Electroencephalogram (EEG) Classification Tasks: A Review. *Journal of Neural Engineering*, **16**, Article 031001. <https://doi.org/10.1088/1741-2552/ab0ab5>
- [10] Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P. and Lance, B.J. (2018) EEGNet: A Compact Convolutional Neural Network for EEG-Based Brain-Computer Interfaces. *Journal of Neural Engineering*, **15**, Article 056013. <https://doi.org/10.1088/1741-2552/aace8c>
- [11] Schirrneister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggenesperger, K., Tangermann, M., *et al.* (2017) Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. *Human Brain Mapping*, **38**, 5391-5420. <https://doi.org/10.1002/hbm.23730>
- [12] Tao, W., Li, C., Song, R., Cheng, J., Liu, Y., Wan, F., *et al.* (2023) EEG-Based Emotion Recognition via Channel-Wise Attention and Self Attention. *IEEE Transactions on Affective Computing*, **14**, 382-393. <https://doi.org/10.1109/taffc.2020.3025777>
- [13] Ganin, Y. and Lempitsky, V. (2015) Unsupervised Domain Adaptation by Backpropagation. *International Conference on Machine Learning*, Lille, 6-11 July 2015, 1180-1189.
- [14] Long, M., Cao, Z., Wang, J., *et al.* (2018) Conditional Adversarial Domain Adaptation. *Advances in Neural Information Processing Systems*, **31**, 1640-1650.
- [15] Zheng, W.L. and Lu, B.L. (2015) Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development*, **7**, 162-175. <https://doi.org/10.1109/tamd.2015.2431497>
- [16] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1023/a:1022627411411>
- [17] Song, T., Zheng, W., Song, P. and Cui, Z. (2018) EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. *IEEE Transactions on Affective Computing*, **11**, 532-541. <https://doi.org/10.1109/taffc.2018.2817622>
- [18] Gu, Y., Zhong, X., Qu, C., Liu, C. and Chen, B. (2023) A Domain Generative Graph Network for EEG-Based Emotion Recognition. *IEEE Journal of Biomedical and Health Informatics*, **27**, 2377-2386. <https://doi.org/10.1109/jbhi.2023.3242090>
- [19] Van der Maaten, L. and Hinton, G. (2008) Visualizing Data Using t-SNE. *Journal of Machine Learning Research*, **9**, 2579-2605.