



中国科学院大学
University of Chinese Academy of Sciences

情感计算报告

(2025 学年春季学期)

论文选题:	基于 EEG 的自适应无监督情感识别任务 在跨被试和跨数据集场景下的算法研究
任课教师:	刘斌
学号姓名:	刘亦陈 202418020428003
	李子恒 20241801428072
	潘鹏程 202428014629007

2025 年 5 月 30 日

一、研究背景

近年来，情感识别技术在心理学研究、个性化医疗、智能教育、驾驶安全和人机交互等多个领域得到了广泛应用。情感计算主要通过分析生理信号来确定个体的情感状态，常见的信号包括脑电图（EEG）、肌电图（EMG）和心电图（ECG）。其中，EEG 信号能够直接反映大脑皮层的神经电活动，具有高时间分辨率，无辐射风险且设备更加便携，因此成为情感识别研究中最具潜力的数据来源之一。

非侵入式 EEG 信号通过在头皮表面布置电极，捕捉大脑皮层产生的自发电位。这些电位反映了大脑不同区域的功能状态。从神经科学角度来看，腹内侧前额叶皮层（vmPFC）、眶额叶皮层（OFC）、前额叶及杏仁核等脑区与情感加工密切相关。EEG 信号依据频率可分为五个典型带： δ 波（0–4 Hz）、 θ 波（4–8 Hz）、 α 波（8–12 Hz）、 β 波（12–30 Hz）和 γ 波（ >30 Hz）。各频段代表不同的神经机制，例如 θ 波与情绪调节和认知负荷相关， α 波与放松和闭眼状态相关， β/γ 波则反映兴奋和注意力状态。通过频域分析，可从中提取与喜悦、厌恶、平静或激动等情绪状态相关的特征。

早期的 EEG 情感识别研究主要依赖于手动提取的时域、频域和时频域特征，例如均值、方差、功率谱密度和小波能量等，但这些低级特征单独使用时往往难以提供足够的区分能力。随后，研究者引入了更为丰富的间接指标，如熵（熵率、多尺度熵）、互信息和频带功率比值等，这些特征可以捕捉 EEG 信号的非线性和复杂动态，提高了分类性能，成为主流方法。

随着深度学习的兴起，基于卷积神经网络（CNN）、循环神经网络（RNN）和自注意力机制（Transformer）的模型开始应用于 EEG 情感识别，从而能够自动学习高维抽象特征，减少手工特征工程。然而，EEG 数据集通常规模较小且跨被试差异显著，导致深度模型易于过拟合。此外，CNN 的局部感受野和固定卷积核与时空关系的建模存在局限，而多头自注意力机制虽可改善全局依赖，却计算开销较大。近年来，结合图神经网络（GNN）和时频自适应注意力的混合架构逐渐受到关注，以期在保持模型泛化能力的同时兼顾时空结构信息¹。

跨被试情感识别是指在模型训练和测试阶段使用不同被试的数据，旨在构建具有良好泛化能力的通用模型。该方向对于真实应用场景尤为重要，但面临以下主要挑战：

- **生理和心理差异**：个体之间的脑区解剖结构、神经传导路径及心理状态存在差异，导致相同情绪刺激下 EEG 响应显著不同。
- **实验环境多样性**：电极放置、皮肤导电状态、采样设备和刺激呈现方式等因素的差

¹ Tang, W., Fan, L., Lin, X., & Gu, Y. (2025). EEG emotion recognition based on efficient-capsule network with convolutional attention. Biomedical Signal Processing and Control, 103, 107473.

异，会引入显著的信号扰动。

- **领域偏移**：训练集和测试集在统计分布层面存在差别，使模型在新被试上性能大幅下降。
- **数据获取成本高**：EEG 采集需要专业硬件和大量预处理步骤，限制了大规模标记数据的积累。
- **样本不平衡与过拟合**：不同被试的数据量不均衡，深度模型易在数据丰富的被试上过拟合，难以迁移到数据稀缺的被试。

面对上述挑战，研究者提出了多种策略，包括领域自适应（Domain Adaptation）、迁移学习（Transfer Learning）、对抗学习（Adversarial Learning）以及元学习（Meta-Learning）等方法，以减小被试间的分布差异，提高模型的跨域泛化能力。此外，多模态融合（如 EEG+EMG+心率变异性）和数据增强技术也被广泛应用，进一步提升了情感识别的鲁棒性。

基于 EEG 的情感识别研究已取得显著进展，但在跨被试泛化能力和深度模型应用方面仍有较大提升空间。

二、方法原理

传统的深度学习模型在训练完成后参数固定，在测试阶段无法根据新输入进行微调。然而，现实世界中的数据分布往往会持续演化，导致模型性能下降。Test-Time Adaptation（以下简称 TTA）通过在测试阶段快速微调和调整模型，使其能够应对数据分布的变化。TTA 是 Domain Adaptation 的一个分支。它们同样有一个源域和一个目标域，首先在源域上进行预训练，然后半监督或无监督地适应到目标域上。两者的主要区别在于，Test-Time Adaptation 的训练在测试的同时完成。可以简单理解为，Test-Time Adaptation 是只经过一个 epoch 的 DA。另一个区别在于，DA 往往报告模型训练完成后的模型性能，而 TTA 的测试和训练是同时进行的，故报告的性能介于训练前和训练结束之间。

如下表所示。"-"代表不使用该部分数据，表中分别给出输入数据和相应的标签。

设置	输入数据	标签
源域	\mathcal{X}_s	y_s
目标域	\mathcal{X}_t	-

2.1 相关工作

2.1.1 BN Adapt

由于源域与目标域存在分布偏移，一种直观的做法是在测试阶段更新目标域上的 BatchNorm 参数，替代使用源域上通过 EMA 获得的统计量。作为简单的 baseline，可直接在测试时保持 `model.train()` 状态，以实时更新 BatchNorm 统计。

2.1.2 Tent

Tent 在 BN Adapt 的基础上引入信息熵最小化损失

$$L = - \sum_k p_k \log p_k$$

其中 p_k 为模型对第 k 类的预测概率。测试时固定使用当前 batch 的 μ 和 σ ，并将所有 BatchNorm 层的 `track_running_stats=False`，仅更新 γ 和 β 参数。此方法简单高效，成为后续研究的基础。

2.1.3 DELTA

1. DELTA 针对 TTA 中的两个挑战：BatchNorm 统计量仅由当前 batch 决定；
2. 类别不平衡或类间顺序带来的偏置更新。
 - 针对 (1)，提出在测试时通过 EMA 更新 BatchNorm 统计 (TBR)；
 - 针对 (2)，使用 EMA 跟踪的类别比例向量对样本加权 (DOT)，以纠正类别偏置。

2.1.4 CoTTA

- CoTTA 在 Tent 基础上引入教师-学生框架：将源模型备份为 `model_anchor`，作为教师模型；
- 每个 batch 首先对图像进行多次（如 32 次）数据增强预测，并以平均置信度与阈值 `apa_p` 比较；
- 若高于阈值，则使用教师对原图预测的伪标签；否则，使用增强预测平均值作为伪标签；
- 学生模型以此伪标签对原图进行训练，损失为交叉熵；
- 学生模型更新后，教师模型通过 EMA 更新参数；
- 并随机恢复部分源模型参数，以缓解灾难性遗忘。

2.1.5 EATA

- EATA 针对 TTA 的 (1) 更新效率和 (2) 灾难性遗忘问题：(1) 使用熵筛选高置信度样本，并以 EMA 更新的多样性向量为样本赋权；
- (2) 通过弹性权重固化 (EWC) 评估参数重要性，对更新添加正则化项；
- 并借助少量无标签源域数据及其伪标签，估计参数重要性。

2.1.6 SAR

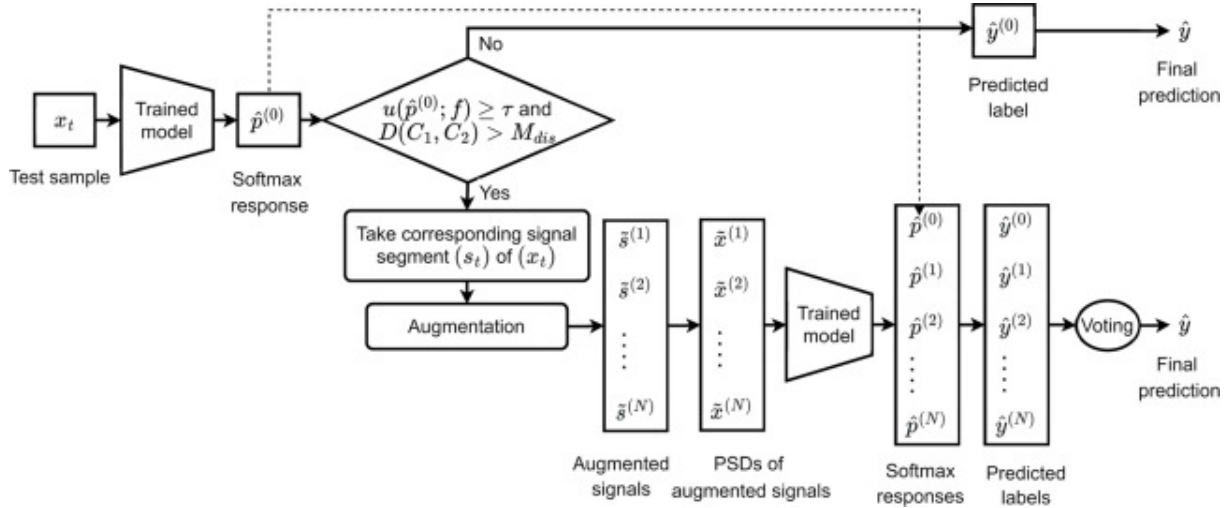
1. SAR 从最坏情况入手，考察三类场景：混合扰动 (mixed shifts)、小 batch、标签分布不均。文章指出，BN 层是导致在线更新不稳定的重要因素，通常替换为 LN 或 GN 会更稳定，但也有失败情况。为此提出两步策略：剔除梯度过大的噪声样本；
2. 引导模型进入极小值空间，以提升 TTA 稳定性。

TTA 是一种在模型测试阶段应用的技术，通过对测试数据进行多种形式的变换，生成多个版本的输入数据，然后将这些变换后的数据输入模型进行预测，最后将多个预测结果进行融合，以提高模型的泛化能力和预测准确性：

- 数据增强：在测试阶段，对每个测试样本应用多种数据增强操作，如水平翻转、垂直翻转、缩放、旋转等。
- 模型预测：对每个增强后的样本分别进行模型预测。
- 结果融合：将多个增强样本的预测结果进行融合，常见的方法包括平均、投票等

2.2 PC-TTA

TTA 最近因提高模型处理看不见的变化和提高了分类准确性的能力而受到关注。它通过生成多个增强输入版本并组合它们的预测来工作。然而，TTA 的计算成本是一个值得关注的问题，因为它涉及应用多个变换和执行大量预测操作，这可能是资源密集型的。我们希望能够将其运用在 EEG 跨被试情感识别中。



论文《Enhanced cross-dataset electroencephalogram-based emotion recognition using unsupervised domain adaptation》中使用到了一种 PC-TTA²技术，当模型结果置信度过低时使用 TTA 技术进行数据增广，再通过投票来做出最后的预测。

情感识别研究中，基于脑电信号（EEG）的方法因其能直接反映神经活动而备受关注。然而，不同数据集间的分布差异及模型对低置信度样本的鲁棒性不足，常导致跨数据集情感分类性能下滑。为此，近年来研究者提出了多种无监督域自适应（Unsupervised Domain Adaptation, UDA）技术，以缩小源域与目标域之间的分布鸿沟，并提高模型的泛化能力。其中，预测置信度引导的测试时自适应（Prediction Confidence-guided Test Time Adaptation, PC-TTA）策略通过在测试时针对低置信度样本动态应用多次数据增强（Test-Time Augmentation, TTA）并结合投票机制，显著提升了模型在目标域上的分类准确率。

在论文《Enhanced cross-dataset electroencephalogram-based emotion recognition using unsupervised domain adaptation》中，作者首先训练了一个经典的卷积神经网络（CNN）用于 EEG 情感分类；随后，针对测试阶段模型输出的置信度低于预设阈值的样本，按以下流程进行增强与投票：

1. 低置信度检测：计算模型对单个样本的预测概率，当最高类别概率低于阈值时，标记为“需增强”样本。
2. TTA 数据增广：对每个“需增强”样本，利用多种时域或频域数据增强手段（例如信号平移、加噪、带通滤波等）生成若干变体。
3. 多次推断与投票：将所有变体输入模型，收集每次预测结果，最后采用多数投票或加权投票策略确定该样本的最终类别。

² Imtiaz, M. N., & Khan, N. (2025). Enhanced cross-dataset electroencephalogram-based emotion recognition using unsupervised domain adaptation. *Computers in Biology and Medicine*, 184, 109394.

该 PC-TTA 策略在保持模型架构不变的前提下，巧妙利用了测试时的额外信息和不确定性评估，从而实现显著的跨数据集性能提升。

传统的方法有以下缺点需要改进：

1. 原始 CNN 模型的局限：长距离依赖建模不足。EEG 信号具有明显的时序相关性和跨脑区同步特性；传统 CNN 通过局部卷积核只能提取邻域内的短程依赖，难以捕捉远距离时序或空间关联。
2. 多通道结构信息利用不足：EEG 采自多个电极，不同通道之间含有丰富的空间分布信息；单纯的全连接层无法区别各通道的相对重要性，也难以表达跨被试（subject-to-subject）差异带来的信号多样性。

2.3 注意力机制引入

2.3.1 原始模型结构分析（model_old.py）

2.3.1.1 原始模型架构

原始网络采用了简单的全连接神经网络结构，主要包含两个部分：

- **特征提取器(Feature Extractor)**：由两个全连接层组成，分别是 160→128 和 128→64
- **分类器(Classifier)**：由三个全连接层组成，分别是 64→32、32→10 和 10→num_classes

```
1  特征提取器结构：  
2  Input(160 维) → FC1(128) → BN → ReLU → FC2(64) → BN → ReLU → Output(64 维)  
3  
4  分类器结构：  
5  Input(64 维) → FC1(32) → BN → ReLU → FC2(10) → BN → ReLU →  
   FC3(num_classes) → Softmax
```

2.3.1.2 原始模型在 EEG 情感识别中的局限性

1. **缺乏对长距离依赖的建模能力**：EEG 信号是时间序列数据，不同时间点、不同脑区之间的复杂相互作用对情感状态有重要影响，而简单的全连接层难以捕捉这些长距离依赖关系。
2. **对所有输入特征一视同仁**：原模型没有区分不同通道(电极)或频带的重要性，而在 EEG 情感识别中，某些特定的脑区活动和频带信息（如前额叶的 alpha 和 beta 频带）对情感状态的指示作用更强。

3. **数据利用效率低**：简单堆叠全连接层的方式无法充分利用 EEG 数据的特殊结构和先验知识，如脑区拓扑关系、频带信息等。
4. **特征表达能力弱**：缺乏有效的表示学习机制，无法根据任务需求自适应地调整特征表示，导致模型难以提取与情感状态相关的有效特征。
5. **域适应能力有限**：在跨被试或跨数据集场景中，简单的网络结构难以应对 EEG 数据中的个体差异和分布偏移问题。

2.3.2 改进模型设计 (model.py)

2.3.2.1 改进策略概述

针对原始模型的局限性，我们在保持输入输出接口不变的情况下，引入了三种注意力机制来增强模型捕捉长距离依赖关系的能力：

1. **自注意力机制(Self-Attention)**：捕捉特征间的长距离依赖
2. **通道注意力机制(Channel Attention)**：识别重要脑区/电极
3. **频带注意力机制(Frequency Band Attention)**：强调关键频带信息

2.3.2.2 详细改进结构

2.3.2.2.1 自注意力模块


```

1  class SelfAttention(nn.Module):
2      def __init__(self, input_dim, hidden_dim=64):
3          super(SelfAttention, self).__init__()
4          self.attention = nn.Sequential(
5              nn.Linear(input_dim, hidden_dim),
6              nn.Tanh(),
7              nn.Linear(hidden_dim, 1),
8              nn.Softmax(dim=1)
9          )
10
11     def forward(self, x):
12         # 计算注意力权重
13         attn_weights = self.attention(x) # [batch_size, input_dim]
14
15         # 应用注意力权重
16         weighted_x = x * attn_weights
17
18         return weighted_x

```

该模块学习不同特征维度的重要性，使模型能够关注到与情感状态最相关的特征点，有效捕捉 EEG 信号中的长距离依赖关系。

2.3.2.2.2 通道注意力模块

```

1  class ChannelAttention(nn.Module):
2      def __init__(self, in_channels, reduction_ratio=4):
3          super(ChannelAttention, self).__init__()
4          self.fc = nn.Sequential(
5              nn.Linear(in_channels, in_channels // reduction_ratio),
6              nn.ReLU(inplace=True),
7              nn.Linear(in_channels // reduction_ratio, in_channels)
8          )
9          self.sigmoid = nn.Sigmoid()
10
11     def forward(self, x):
12         # 计算特征的重要性
13         out = self.fc(x)
14         out = self.sigmoid(out)
15
16         # 返回注意力权重乘以输入
17         return x * out

```

该模块动态学习不同通道（对应不同脑区）的重要性权重，突出关键脑区的贡献，抑制噪声通道的干扰。

通道注意力机制工作原理补充说明：

在我们的实现中，通道注意力机制的工作方式与传统卷积神经网络中的通道注意力略有不同。这里的"通道"概念需要特别说明：

1. **特征通道表示**：由于 EEG 数据经过预处理和初始全连接层后，原始的物理电极通道信息被重新编码到特征向量中，我们将这 128 维特征的每个维度视为一个"通道"。

2. 权重学习过程：

- 接收 128 维特征向量
- 通过降维再升维的全连接层网络学习每个特征维度的重要性权重
- 使用 sigmoid 函数将权重归一化到 0-1 范围
- 将权重与原始特征相乘，增强重要特征，抑制不重要特征

3. 与物理 EEG 通道的关联：

- 虽然模型直接操作的是特征空间，但这些特征是从原始 EEG 信号中提取的，包含了不同脑区的信息
- 通过学习特征重要性，模型间接地发现了哪些脑区对情感状态识别更重要
- 这种方式允许模型自动发现不同情感状态下的关键脑区活动模式，而无需显式地指定电极位置

2.3.2.2.3 频带注意力模块

```
1 class FrequencyBandAttention(nn.Module):
2     def __init__(self, num_bands=5, band_features=32):
3         super(FrequencyBandAttention, self).__init__()
4         self.num_bands = num_bands
5         self.band_features = band_features
6         self.total_features = num_bands * band_features
7
8         self.attention = nn.Sequential(
9             nn.Linear(self.total_features, num_bands),
10            nn.Softmax(dim=1)
11        )
```

该模块根据不同情感状态，动态调整各频带（delta, theta, alpha, beta, gamma）的重要性权重，因为在情感识别中不同频带通常包含不同的情感相关信息。

频带注意力机制工作原理补充说明：

频带注意力机制是针对 EEG 信号的频率特性专门设计的创新模块：

1. 频带信息的表示：

- EEG 信号通常被分解为多个频带：delta (0.5-4Hz), theta (4-8Hz), alpha (8-13Hz), beta (13-30Hz), gamma (>30Hz)
- 在我们的实现中，将 128 维的特征重组为 4 个频带，每个频带 32 个特征
- 这种组织方式基于 EEG 处理的领域知识，将相关频带特征分组

2. 频带重要性学习：

- 模型接收重组后的频带特征 [batch_size, num_bands, band_features]

- 通过全连接网络学习每个频带的重要性权重
- 使用 softmax 归一化确保频带权重和为 1，形成概率分布

3. 智能频带选择：

- 对于不同的情感状态，模型可以动态强调相关频带：
- 高频 beta 和 gamma 波与兴奋、警觉等高唤醒度情绪相关
- alpha 波与放松、平静等状态相关
- theta 波与梦境、冥想状态相关

4. 维度匹配保护：

- 代码中包含了条件检查，确保输入维度满足频带重组要求
- 如果特征维度不符合预设的频带结构，模块会跳过频带处理，保证模型正常运行

通过这种设计，模型能够自动发现不同情感状态下最具辨别力的频带信息，进一步提高了 EEG 情感识别的准确性和可解释性。

2.3.2.2.4 改进的特征提取器

```

1  class Feature(nn.Module):
2      def __init__(self):
3          super(Feature, self).__init__()
4          # 第一层网络
5          self.fc1 = nn.Linear(160, 128)
6          self.bn1_fc = nn.BatchNorm1d(128)
7
8          # 注意力机制
9          self.self_attention = SelfAttention(128, hidden_dim=64)
10         self.channel_attention = ChannelAttention(128, reduction_ratio=4)
11         self.freq_attention = FrequencyBandAttention(num_bands=4,
12             band_features=32)
13
14         # 第二层网络
15         self.fc2 = nn.Linear(128, 64)
16         self.bn2_fc = nn.BatchNorm1d(64)
17
18         # 最终注意力
19         self.final_attention = SelfAttention(64, hidden_dim=32)

```

特征提取器通过组合三种注意力机制，在原始的全连接网络基础上大幅增强了特征学习能力。此外，我们还采用了残差连接方式应用注意力机制，确保梯度传递更加稳定。

2.3.3 改进模型的优势

2.3.3.1 理论优势

1. 增强长距离依赖建模能力：

- 自注意力机制能够捕捉不同时间点和脑区间的复杂相互关系
- 有助于发现 EEG 信号中分布在不同时间和空间位置的情感相关模式

2. 突出重要特征：

- 通道注意力突出重要脑区（如前额叶、颞叶等与情感处理相关的区域）的信号
- 频带注意力强调特定情感状态下最具辨别力的频带（如 alpha 波与放松相关，beta 波与兴奋相关）

3. 提高数据利用效率：

- 注意力机制可以选择性地关注最相关的特征，减少无关信息的干扰
- 通过参数共享，在相同参数量下提供更强的表达能力

4. 增强特征表示能力：

- 多层注意力机制构建了层次化的特征表示
- 残差连接保留原始信息的同时增强了表示能力

5. 提高域适应能力：

- 通过关注跨被试或跨数据集中共有的情感相关特征
- 减少个体差异和设备差异带来的噪声影响
- 这一点也在域适应训练时间的减少上体现了出来

2.3.3.2 实际应用优势

1. **识别准确度提升**：注意力机制可以提高模型区分不同情感状态的能力，特别是在复杂情感场景中
2. **解释性增强**：通过可视化注意力权重，可以观察模型决策时关注的脑区和频带，增强模型可解释性
3. **降低对数据量的依赖**：注意力机制能更高效地利用有限数据，减少过拟合风险
4. **环境适应性增强**：在跨被试和跨数据集场景中表现更加稳健
5. **计算效率权衡**：虽然计算复杂度略有增加（在 NVIDIA A40 环境下训练时间增加了 33.96%），但通过选择性关注，实际上可以更高效地利用计算资源

三、实验设置与结果

3.1 数据集

（1）SEED 数据集

1 数据集概述

SEED (SJTU Emotion EEG Dataset) 由上海交通大学脑与计算科学实验室 (Brain & Computation Laboratory , BCMI) 精心构建³，是一套公开的多模态情感识别数据集，旨在推

³ Zheng, W.-L., & Lu, B.-L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Transactions on Autonomous Mental Development, 7(3) (2015)

动基于脑电信号（EEG）与眼动数据的情感计算及脑-机接口研究。该数据集共收集了 15 位中国受试者在观看情感诱导影视片段时的 62 通道 EEG 信号及眼动跟踪数据，为开发智能、情感化的人机交互系统提供了高质量、多维度的实验素材。

2 实验设计

● 刺激材料

- 15 段中国电影片段，涵盖正面（Positive）、中性（Neutral）与负面（Negative）三种情绪类型；
- 每段时长约 4 分钟，经专家评审与预实验验证，确保能够有效诱发目标情绪且影像内容易于理解。

● 选择标准

1. **时长适中**：单次实验控制在 4 分以内，避免受试者疲劳累积；
2. **情绪明确**：每段视频仅对应一种主要情绪，避免情绪混淆；
3. **无需解释**：无需额外文字或口头说明，观片过程中受试者能够自主理解。

● 实验流程

1. 每位受试者完成 15 次试验，试验顺序随机但保证同种情绪片段不会连续出现超过两次；
2. 试验开始前出现 5 秒提示屏，告知即将播放的情绪类别；
3. 视频播放结束后，受试者有 45 秒时间通过自评量表（Likert 7 点量表）报告当下情绪体验；
4. 每次试验间隔 15 秒，以减少前一情绪对下一试验的残余影响。

3 受试者信息

- **基本情况**：共 15 位中国籍大学生（7 男，8 女），平均年龄 23.27 岁（标准差 2.37 岁）；
- **编号及数据类型**：
 - 编号 1-5、8-14（共 12 人）同时采集 EEG 与眼动数据；
 - 编号 6、7、15（三人）仅采集 EEG 数据；
- **隐私保护**：在公开数据中仅使用编号标识，无任何可反向识别信息。

4 信号采集与预处理

1. **原始数据**：存于 `EEG_raw/` 文件夹，每次会话对应一个 `.cnt` 文件；
2. **预处理数据**：存于 `Preprocessed_EEG/`，包括滤波、重参考、降采样及按试验分段的 EEG 片段；
3. **特征提取**：存于 `Extracted_Features/`，提取微分熵（DE）、差分不对称（DASM）与比率不对称（RASM）三类特征，并对特征序列采用移动平均与线性动态系统（LDS）平滑；
4. **眼动数据**：存于 `Eye_Movement/`，包括注视点位置、瞳孔直径等特征 `.mat` 文件。

5 文件组织结构

```
1  SEED_Dataset/  
2  |--- EEG_raw/  
3  |   |--- sub1_session1.cnt  
4  |   |--- sub1_session2.cnt  
5  |   |--- ...  
6  |--- Preprocessed_EEG/  
7  |   |--- sub1_session1.mat  
8  |   |--- sub1_session2.mat  
9  |   |--- ...  
10 |--- Extracted_Features/  
11 |   |--- sub1_session1_DE.mat  
12 |   |--- sub1_session2_DE.mat  
13 |   |--- ...  
14 |--- Eye_Movement/  
15 |   |--- sub1_session1_eye.mat  
16 |   |--- ...  
17 |--- README.md  
18 |--- 数据集说明与使用指南
```

- **EEG_raw/**：原始未处理 EEG 数据；
- **Preprocessed_EEG/**：滤波、降采样及分段后的 EEG 数据；
- **Extracted_Features/**：DE、DASM、RASM 特征及平滑结果；

- **Eye_Movement/**：对应试验的眼动跟踪特征；
- **README.md**：详细的实验协议、数据格式说明与使用建议。

(2) DEAP 数据集

1 数据集概述

DEAP (Database for Emotion Analysis using Physiological signals) 是一套广泛使用的多模态情感计算数据集⁴，由 Queen Mary University of London、University of Twente、University of Geneva 及 EPFL 等机构联合构建。该数据集聚焦于利用生理信号与视频信息对人类情感状态进行分析与识别，旨在为情感智能的人机交互系统提供坚实的数据基础。DEAP 数据集包含来自 32 位被试者的 32 通道 EEG、13 项外周生理信号（皮肤电反应、心率、呼吸等）以及 22 位被试者的面部视频录像，所有数据均在被试观看 40 段 1 分钟时长的音乐视频片段过程中同步采集。

2 实验设计

- **情绪诱导材料**

- 初始候选集：120 段来源于在线音乐平台（Last.fm）的音乐视频，覆盖多种风格与情感标签；
- 预筛选：通过在线主观评估实验，根据唤醒度（Arousal）与效价（Valence）评分选取 40 段，确保刺激能够有效并稳定地诱发目标情绪维度。

- **实验流程**

- **观看阶段**：每位被试连续观看 40 段 1 分钟音乐视频，视频顺序随机打乱；
- **自评阶段**：每段视频结束后，被试在屏幕上分别对以下五个维度进行评分：
 - 唤醒度（Arousal）
 - 效价（Valence）
 - 喜好度（Liking）
 - 支配感（Dominance）
 - 熟悉度（Familiarity）
- 评分量表为 1–9 点 Likert 量表；

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., ... & Patras, I., DEAP: A Database for Emotion Analysis; Using Physiological Signals. IEEE Transactions on Affective Computing, 3(1) (2012) 18-31.

- **休息与标记**：自评完成后短暂休息（约 5–10 秒）以消除前一刺激残余影响，然后进入下一试验。

3 受试者信息

- **总样本量**：32 名健康成人（性别、年龄具体信息见原始文献）；
- **面部视频**：其中 22 名被试在实验全过程中同步录制高清面部视频，便于从面部表情中提取情绪线索；
- **隐私保护**：所有数据经匿名化处理，被试仅以编号标识，无任何可识别个人身份的信息。

4 信号采集与预处理

信号类型	设备 / 通道数	采样率	原始格式	预处理描述
EEG	Biosemi ActiveTwo , 32 通道	512 Hz	.mat	带通滤波（4–45 Hz）、重参考、 去伪迹
皮肤电反应 （EDA）	Biosemi	512 Hz	.mat	滤波、下采样、去除运动伪迹
心率、呼吸等	13 项外周生理参 数（心率变异、 呼吸带等）	512 Hz	.mat	标准化、平滑处理
面部视频	高清摄像机	50 fps	视频文件	同步对齐、关键帧标注

1. **同步采集**：所有生理信号与视频流通过硬件触发信号保持毫秒级同步；
 - 1、**预处理流程**：对 EEG 与生理信号均进行去趋势、去伪迹及带通滤波，并在需要时下采样；
2. **数据对齐**：借助触发标记对齐各模态数据，为后续特征同步提取与情感分析提供基础。

5 数据集元信息

项目	说明
任务类型	情绪识别
EEG 通道数	32
EEG 采样率	512 Hz
文件格式	.mat、.csv
自评量表	1–9 点 Likert 量表

面部视频帧率	50 fps
--------	--------

6 文件组织结构

```

1  DEAP_Dataset/
2  |— EEG_Signals/
3  |   |— sub001.mat
4  |   |— sub002.mat
5  |   └— ...
6  |— Physiological_Signals/
7  |   |— sub001_phys.mat
8  |   |— sub002_phys.mat
9  |   └— ...
10 |— Self-Report_Ratings/
11 |   |— sub001_ratings.csv
12 |   |— sub002_ratings.csv
13 |   └— ...
14 |— Face_Videos/
15 |   |— sub001_face.avi
16 |   |— sub002_face.avi
17 |   └— ...
18 └— README.md
19    └— 数据集说明与使用指南

```

- **EEG_Signals/**：原始及预处理后的 EEG 数据；
- **Physiological_Signals/**：皮肤电、心率、呼吸等外周信号；
- **Self-Report_Ratings/**：每位被试在五个情感维度上的打分；
- **Face_Videos/**：同步录制的面部表情视频；
- **README.md**：包括实验协议、数据格式、使用示例及许可证信息。

3.2 实验设置

3.2.1 运行环境

- 操作系统：Linux 3.10.0-1160.el7.x86_64
- Python 版本：3.10.16
- CUDA 版本：11.8
- NVIDIA 驱动版本：520.61.05
- GPU 硬件：NVIDIA A40，单块显存 46GB

3.2.2 耗时分析

3.2.2.1 原始网络

运行时间统计：

- 全过程总耗时：734.44 s
- 训练阶段（总计 554.44 s，约 75.6 % 全过程）
 - 预训练阶段：278.42 s（50.2 % 训练阶段；37.9 % 全过程）
 - 聚类损失训练阶段：180.40 s（32.5 % 训练阶段；24.6 % 全过程）
 - 自适应训练阶段：95.62 s（17.3 % 训练阶段；13.0 % 全过程）
- 测试阶段：180.00 s（24.5 % 全过程）

3.2.2.2 注意力机制改进网络

运行时间统计：

- 全过程总耗时：1127.17 s
- 训练阶段（总计 742.74 s，约 65.9 % 全过程）
 - 预训练阶段：391.24 s（52.7 % 训练阶段；34.7 % 全过程）
 - 聚类损失训练阶段：260.29 s（35.1 % 训练阶段；23.1 % 全过程）
 - 自适应训练阶段：91.21 s（12.3 % 训练阶段；8.1 % 全过程）
- 测试阶段：384.43 s（34.1 % 全过程）

3.2.2.3 总体比较

运行时间变化分析

项目	原始网络	改进网络	绝对变化	相对变化（%）
----	------	------	------	---------

全过程总耗时	734.44 s	1127.17 s	+392.73 s	+53.5 %
训练阶段总耗时	554.44 s	742.74 s	+188.30 s	+34.0 %
· 预训练	278.42 s	391.24 s	+112.82 s	+40.5 %
· 聚类损失训练	180.40 s	260.29 s	+79.89 s	+44.3 %
· 自适应训练	95.62 s	91.21 s	-4.41 s	-4.6 %
测试阶段耗时	180.00 s	384.43 s	+204.43 s	+113.6 %

1. 全过程性能对比

- **总耗时**：改进网络比原始网络增加了 **53.5 %**，从 734.44 s 上升到 1127.17 s。
- **影响因素**：这种整体运行时间的增幅，主要源自训练阶段的显著拉长与测试阶段的推理开销。

2. 训练阶段细分

1. 预训练阶段

- a. 原始：278.42 s —改进：391.24 s
- b. 增加 **40.5 %** (+112.82 s)
- c. **原因推测**：注意力模块的引入，使得网络在初始权重调整和特征学习时加入了额外的自注意力计算，显著增加了前期开销。

2. 聚类损失训练阶段

- a. 原始：180.40 s —改进：260.29 s
- b. 增加 **44.3 %** (+79.89 s)
- c. **原因推测**：在聚类损失上加入注意力加权后，每次迭代的梯度计算和簇中心更新更加复杂，导致流程变长。

3. 自适应训练阶段

- a. 原始：95.62 s —改进：91.21 s

b. 减少 4.6 % (-4.41 s)

c. **原因推测**：新引入的注意力机制在后期学习率调度和样本权重更新上更高效，部分抵消了前两阶段带来的负担。

3. 测试阶段推理效率

- 原始：180.00 s →改进：384.43 s
- 增加 113.6 % (+204.43 s)
- **分析**：注意力机制通常需要存取全局特征映射并进行加权汇总，其计算复杂度为 $O(n^2)$ (n 为特征长度)，因而在大批量推理时开销显著增大。

4. 小结与优化建议

- **性能-精度权衡**：改进网络在引入注意力之后，虽然可能提升了情感识别的准确性与鲁棒性，但整体耗时几乎翻倍，尤其是测试阶段的实时推理效率受到严重影响。
- **未来实验方向**：
 - **轻量化注意力**：可考虑使用线性注意力或稀疏注意力机制，将计算复杂度从二次降低至线性或次线性；
 - **混合精度训练与推理**：利用半精度浮点数 (FP16) 和张量核心加速，压缩模型计算成本；
 - **动态推理剪枝**：在测试时对不重要的通道或特征进行剪枝，减少注意力模块的输入维度。

3.3 实验结果

3.3.1 跨被试结果

- 在 SEED 数据集二分类（消极、积极）任务上取得了 60.19% 的准确率，与近些年的工作 HT⁵、Gbayes⁶、HMM⁷、EEGFuseNet⁸、WTCNN⁹、VDN-DNN¹⁰、STFG-CAP¹¹ 对比见下表：

Studies	Accuracy(%)
HT	54.30
Gbayes	57.60
HMM	58.75
EEGFuseNet	56.27
WTCNN	61.50
VDN-DNN	62.50
STFG-CAP	48.22
Ours	60.19

⁵ Z. Liang, S. Oba, S. Ishii, An unsupervised EEG decoding system for human emotion recognition, Neural Netw. 116 (2019) 257–268.

⁶ S. Koelstra, C. Muhl, M. Soleymani, J.S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, DEAP: A database for emotion analysis using physiological signals, IEEE Trans. Affect. Comput. 3 (1) (2011) 18–31.

⁷ C.A Torres-Valencia, H.F. Garcia-Arias, M.A.A. Lopez, A.A Orozco-Gutierrez, Comparative analysis of physiological signals and electroencephalogram (EEG) for multimodal emotion recognition using generative models, in: 2014 XIX Symposium on Image, Signal Processing and Artificial Vision, IEEE, 2014, pp. 1–5.

⁸ Z. Liang, R. Zhou, L. Zhang, L. Li, G. Huang, Z. Zhang, S. Ishii, EegFuseNet: Hybrid unsupervised deep feature characterization and fusion for highdimensional EEG with an application to emotion recognition, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 1913–1925.

⁹ P. Pandey, K.R. Seeja, Subject independent emotion recognition system for people with facial deformity: An EEG based approach, J. Ambient Intell. Humaniz. Comput. 12 (2) (2021) 2311–2320.

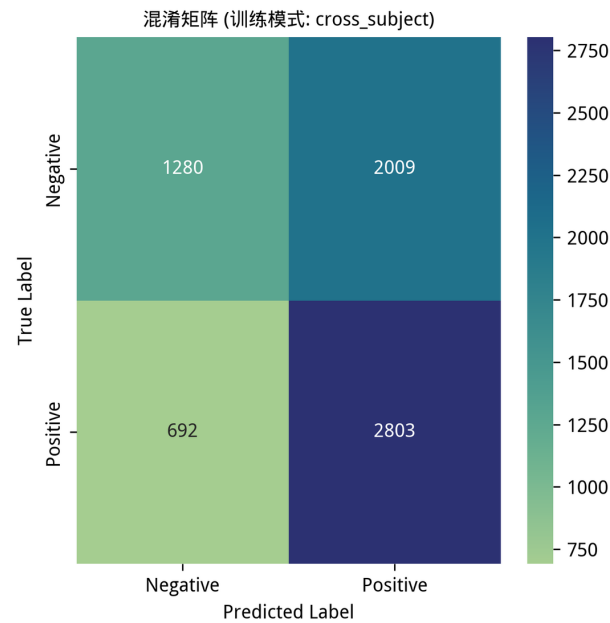
¹⁰ P. Pandey, K.R. Seeja, Subject independent emotion recognition from EEG using VMD and deep learning, J. King Saud Univ.-Comput. Inf. Sci. (2019).

¹¹ G.C. Jana, A. Sabath, A. Agrawal, Capsule neural networks on spatiotemporal EEG frames for cross-subject emotion recognition, Biomed. Signal Process. Control 72 (2022) 103361.

- 结果各项指标如下：

	precision	recall	f1-score	sample number
0	0.6491	0.3892	0.4866	3289
1	0.5825	0.8020	0.6749	3495
average accuracy	0.6019			6784

- 混淆矩阵如下



3.3.2 跨数据集结果

- 我们还尝试了改进后的模型在迁移数据集上的测试 (SEED->DEAP), 平均准确率只有 48.78%，效果非常一般，有待进一步调优。结果各项指标如下：

	precision	recall	f1-score	sample number
0	0.4858	0.2940	0.3663	1160
1	0.4888	0.6844	0.5703	1144
average accuracy	0.4878			2304

四、总结与展望

4.1 总结与分工

4.1.1 实验总结

总结来看，我们完成了一个跨数据集算法到跨受试者算法的迁移与引入注意力机制的改进，虽然效果并没有预期的优秀，但也展现了不错的结果。

- 为了速度上的优化和跨数据集的对齐，我们仅使用了 32-channel 作为训练数据；
- 后续可以：
 - 使用 62-channel 进行训练以提高准确率
 - 进一步优化网络结构，加强自注意力机制
 - PC-TTA 的数据增强方法还有优化的空间

在未来工作中，可以考虑进一步优化注意力机制的设计，如引入图注意力网络捕捉脑区拓扑结构，或结合时域和频域的联合注意力机制，进一步提升模型性能。

4.1.2 分工

姓名	分工内容
刘亦陈	负责报告整理书写，参与实验工作的讨论与部分实现。
李子恒	负责 PPT 制作，PPT 展示和部分报告书写内容。
潘鹏程	负责代码整体实现，注意力机制添加，报告实验部分书写。

五、致谢

日暮跌入迢迢荒野，人间忽晚。

行笔至此，本学期已接近尾声。与刘斌老师共同学习的时光也走向了终点。和刘老师第一次相逢于今年的二月，那时第一次听刘老师的课便被刘老师丰富的专业知识以及独到的国际视野所深深吸引。在这一学期的学习过程中，更是受益良多。

在刘老师的课堂上，我们超越了本学科的局限，以宏大的视角来深度思考那些我们日常可见却并未细想的技术问题。此外，在刘老师的教学中，同样不是死板的填鸭式教育，而是结合各种问题来为我们讲解每个情感计算相关的技术领域的知识。

此外，刘老师在课堂上侃侃而谈的丰富学识同样在这场学术盛宴中令人印象深刻。

无论课堂上同学们给出了什么样的个人看法、提出什么样的问题，刘老师都会在给予鼓励后给出自己对相关问题的思考。这不得不使人感叹刘老师涉猎之广博，真正做到了“博观而约取，厚积而薄发”。

“天下没有不散的筵席”。尽管跟随刘老师进行的学术之旅到此结束，但刘老师留给我们对待学问的态度却将伴随我们接下来的求学生活。

最后，感谢刘老师在这一学期中对每节课程的精心设计，也在此祝愿刘老师在接下来的工作、生活中一切顺利。

夏天快乐，江湖再见。