

**硕士学位论文**

**(专业学位论文)**

**基于自监督学习的脑电图鲁棒分类研究**

**ROBUST CLASSIFICATION OF  
ELECTROENCEPHALOGRAM BASED  
ON SELF-SUPERVISED LEARNING**

**李楚阳**

**哈尔滨工业大学  
2025 年 5 月**

国内图书分类号：TM301.2  
国际图书分类号：62-5

学校代码：10213  
密级：公开

## 硕士学位论文

# 基于自监督学习的脑电图鲁棒分类研究

硕士研究生：李楚阳  
导 师：李修成  
申 请 学 位：工学硕士  
类 别：计算机技术  
所 在 单 位：哈尔滨工业大学（深圳）  
答 辩 日 期：2025 年 05 月 12 日  
授予权利单位：哈尔滨工业大学

Classified Index: TM301.2

U.D.C: 62-5

## Dissertation for the Master Degree

# ROBUST CLASSIFICATION OF ELECTROENCEPHALOGRAM BASED ON SELF-SUPERVISED LEARNING

<b>Candidate:</b>	Chuyang Li
<b>Supervisor:</b>	Dr. Xiucheng Li
<b>Academic Degree Applied for:</b>	Master of Engineering
<b>Specialty:</b>	Computer Technology
<b>Affiliation:</b>	Harbin Institute of Technology, Shenzhen
<b>Date of Defense:</b>	June, 2025
<b>Degree-Conferring-Institution:</b>	Harbin Institute of Technology

## 摘要

脑电图（EEG）作为一种非侵入性的神经生理检测技术，以其毫秒级时间分辨率在如癫痫检测中具有不可替代的价值。然而，在实际临床环境中，EEG 分类模型普遍会受到由噪声和预处理不一致导致的数据分布偏移的影响，部署时常出现性能退化问题。同时，专业标注的稀缺性进一步限制了传统监督学习方法在 EEG 分类任务中的应用，模型难以充分挖掘数据潜在模式。亟需开发鲁棒性强、适应性高的分类模型以应对真实医疗场景挑战。

基于上述背景，本研究提出了一种面向脑电时序建模的鲁棒预训练框架，旨在解决 EEG 信号处理中标签稀缺性和数据分布偏移的双重困境。研究动机源于现有自监督学习方法对缺少对噪声模式的判别、忽略时频通道关系鲁棒性差异以及时频特征利用不充分的问题。为此，本研究创新性地融合了额外噪声负样本对比学习、跨域重构与通道对齐三大技术，构建多任务联合优化的预训练架构。额外噪声负样本对比学习通过引入额外噪声负样本，显式建模噪声分布特征，增强模型对干扰模式的辨识能力；跨域重构任务设计时域到频域及频域到时域的双向映射机制，实现时频特征深度融合，避免单一域内信息损失；通道对齐策略利用频域通道关系的稳定性校正时域表征偏移，进一步提升对数据偏移的鲁棒性。实验在天普大学医院脑电图数据集的癫痫发作检测与异常分类任务上进行，涵盖多种噪声和偏移条件下的性能评估。实验表明，本框架在原始数据及数据偏移环境下均显著优于现有自监督和监督学习基准方法，尤其对带通滤波导致的分布偏移展现出卓越鲁棒性，性能下降幅度最小。本研究首次将噪声显式建模引入时间序列对比学习，建立了时频特征交互与通道校准机制，为 EEG 信号鲁棒分类提供了新视角与实践方案，对推动脑机接口、神经疾病诊断等领域的智能化应用具有重要理论价值与临床意义。

此外，本研究进一步探索了基于残差混合的对比学习增强方法，旨在通过在嵌入空间内实施残差混合操作，缓解传统数据增强策略多样性不足的局限性，增强模型鲁棒性。研究通过分析监督模型中间层特征偏移分布，发现噪声强度与特征偏移间的潜在关联，并据此设计了嵌入空间残差混合机制以引入语义噪声。尽管实验结果显示该方法在不同模型架构中表现不一，性能提升有限，但明确了残差混合策略的适用边界，揭示了高维嵌入空间特征混合可能引入的语义歧义风险，为后续对比学习优化提供了重要经验参考。

**关键词：**时间序列分类；脑电图；鲁棒性；自监督学习

## Abstract

Electroencephalogram (EEG), as a non-invasive neurophysiological detection technology, holds irreplaceable value in applications such as epilepsy detection due to its millisecond-level temporal resolution. However, in real-world clinical environments, EEG classification models are often affected by data distribution shifts caused by noise and inconsistent preprocessing, leading to performance degradation during deployment. Additionally, the scarcity of professional annotations further limits the application of traditional supervised learning methods in EEG classification tasks, making it difficult for models to fully explore the latent patterns in the data. There is an urgent need to develop robust and adaptive classification models to address the challenges of real medical scenarios.

Against this backdrop, this study proposes a robust pre-training framework for EEG time-series modeling, aiming to tackle the dual challenges of label scarcity and data distribution shifts in EEG signal processing. The motivation for this research stems from the limitations of existing self-supervised learning methods, which lack discrimination of noise patterns, overlook robustness differences in time-frequency channel relationships, and underutilize time-frequency features. To address these issues, this study innovatively integrates three key techniques—additional noise negative sample contrastive learning, cross-domain reconstruction, and channel alignment—into a multi-task joint optimization pre-training architecture. The additional noise negative sample contrastive learning explicitly models noise distribution characteristics by introducing extra noise negative samples, enhancing the model’s ability to identify interference patterns. The cross-domain reconstruction task designs bidirectional mapping mechanisms between the time domain and frequency domain, achieving deep fusion of time-frequency features and avoiding information loss within a single domain. The channel alignment strategy leverages the stability of frequency-domain channel relationships to correct time-domain representation shifts, further improving robustness against data shifts. Experiments were conducted on the Temple University Hospital EEG dataset for seizure detection and anomaly classification tasks, covering performance evaluations under various noise and shift conditions. The results demonstrate that the proposed framework significantly outperforms existing self-supervised and supervised learning benchmarks under both raw data and data shift conditions, particularly exhibiting superior robustness against distribution shifts caused

---

---

Abstract

by bandpass filtering, with the smallest performance degradation. This study pioneers the integration of explicit noise modeling into time-series contrastive learning, establishes mechanisms for time-frequency feature interaction and channel calibration, and provides a novel perspective and practical solution for robust EEG signal classification. It holds significant theoretical value and clinical relevance for advancing intelligent applications in brain-computer interfaces and neurological disease diagnosis.

Furthermore, this study explores a contrastive learning augmentation method based on residual mixing, aiming to mitigate the limitations of traditional data augmentation strategies in terms of diversity by implementing residual mixing operations in the embedding space to enhance model robustness. By analyzing the intermediate-layer feature shift distributions of supervised models, a potential correlation between noise intensity and feature shifts was identified, based on which an embedding space residual mixing mechanism was designed to introduce semantic noise. Although experimental results show inconsistent performance across different model architectures with limited improvement, the study clarifies the applicability boundaries of residual mixing strategies and reveals the risk of semantic ambiguity introduced by feature mixing in high-dimensional embedding spaces. These findings provide valuable empirical insights for future optimization of contrastive learning.

**Keywords:** Time-series classification, Electroencephalography, Robustness, Self-supervised learning

# 目 录

第1章 绪论.....	1
1.1 课题背景及研究的目的和意义 .....	1
1.2 国内外研究现状.....	2
1.2.1 脑电图分类研究现状.....	2
1.2.2 自监督学习研究现状.....	4
1.2.3 时间序列自监督学习研究现状.....	6
1.2.4 现有研究存在的主要问题.....	7
1.3 本文的主要研究内容.....	8
1.4 本文的章节安排.....	9
第2章 相关技术基础 .....	12
2.1 自监督学习.....	12
2.1.1 对比学习.....	12
2.1.2 生成式自监督学习.....	13
2.2 Transformer.....	15
2.3 脑电图与癫痫检测.....	18
2.4 本章小结.....	19
第3章 面向脑电时序建模的鲁棒预训练框架 .....	20
3.1 引言.....	20
3.2 问题定义.....	20
3.3 研究动机.....	21
3.3.1 额外噪声负样本.....	21
3.3.2 跨域重构.....	23
3.3.3 时频通道的鲁棒性差异.....	23
3.4 模型设计.....	24
3.4.1 数据预处理.....	25
3.4.2 跨域重构学习.....	26
3.4.3 时频通道对齐.....	27
3.4.4 附带额外噪声负样本的对比学习.....	28
3.4.5 模型训练设计与优化目标.....	30
3.5 实验与分析.....	31
3.5.1 数据集设置.....	31
3.5.2 基准模型与评价标准.....	32
3.5.3 训练设置.....	33
3.5.4 实验结果与分析.....	33

3.5.5 消融实验.....	34
3.5.6 预训练有效性验证.....	37
3.6 本章小结.....	37
第 4 章 基于残差混合的对比学习鲁棒增强方法 .....	38
4.1 引言.....	38
4.2 研究动机.....	38
4.3 附带残差混合的对比学习架构 .....	39
4.4 实验与分析.....	44
4.4.1 数据集设置.....	44
4.4.2 基准模型与评价标准.....	44
4.4.3 超参数分析.....	44
4.4.4 实验结果与分析.....	46
4.4.5 可视化视角下的残差混合鲁棒性退化分析 .....	49
4.5 本章小结.....	51
结论.....	52
参考文献.....	54
附录 A 第四章实验中各混合模型的超参数分析图.....	61

## 第1章 绪论

### 1.1 课题背景及研究的目的和意义

脑电图作为一种非侵入性检测手段，能够准确记录大脑电活动的时空模式，因此在神经功能障碍诊断中具有不可替代的作用。该技术通过头皮电极阵列检测神经元群体活动产生的微弱电位变化，这些信号经过放大和数字化处理后形成可分析的时域波形。其突出优势在于毫秒级时间分辨率，可实时监测大脑功能状态，尤其适用于癫痫等以异常放电为特征的神经系统疾病诊断。此外，脑电图技术还广泛应用于睡眠障碍评估、麻醉深度监测、意识状态判定及脑功能损伤诊断等领域。脑电图数据规模随着医疗信息化建设的推进呈现爆发式增长。如何从海量数据中高效提取有价值信息，构建精准预测模型以实现疾病早期诊断和风险预警，已成为医疗健康研究领域的重要挑战。

数字化转型浪潮下，人工智能技术作为第四次工业革命的核心驱动力，正与脑电图分析实现深度融合。深度时间序列分析是人工智能的重要分支，专注于研究基于深度学习的时间序列建模、预测与解释，而其中时间序列分类问题因其在临床诊断中的关键作用而备受关注。大量临床实践表明，深度时间序列分类技术在提升诊断效能方面具有显著优势，如心电图对心律失常的识别、脑电图对癫痫发作的检测等典型应用案例，充分证明了该技术在提高诊断准确率和实现早期筛查方面的重要价值<sup>[1-4]</sup>。

然而，脑电图数据的特殊性给分类任务带来了挑战。首先是专业标注的稀缺性。脑电图数据集通常需要临床神经科学专家进行标注，这一过程不仅耗时费力，而且成本高昂<sup>[5]</sup>。其次，脑电图采集和处理过程中普遍存在的数据分布偏移问题也不容忽视，即训练数据与实际应用场景中的数据分布存在显著差异。在脑电图领域，数据分布偏移来源具有多源性特征：既包括受试者个体生物节律的固有差异，也涉及采集设备的参数设定、采集环境、数据预处理方法的多样性以及电极部署的异质性等多方面因素。首先，噪声源会导致数据偏移问题。例如，受试者的个体差异可能导致信号频率特征的变化；采集设备的不同参数设定或环境干扰（如电源线噪声、肌肉活动干扰）可能引入伪迹；而电极部署的异质性则可能导致信号的空间分布不一致。此外，训练和部署的数据预处理方法的选择和实现方式不一致也会导致数据偏移问题。例如，滤波技术（如带通滤波、低通滤波）常用于去除特定频段的噪声，但如果模型部署和训练采用不同的滤波参数，数据的频谱分布也会不一样，模型的性能也因此会受损；重参考方法通过改变参考电极的位

置来减少共模噪声，但不同的参考方案可能引入新的偏差。上述导致的数据分布偏移问题可能在实际医疗环境中引发模型性能的显著退化<sup>[6]</sup>。

在当前研究背景下，传统时间序列分类方法在处理标签稀缺的脑电图数据集时存在显著的局限性<sup>[7]</sup>。深度学习中基于监督学习的模型依赖大量标注数据进行训练，标签稀缺导致监督学习方法的性能受到严重制约，难以充分挖掘数据中的潜在模式。自监督学习范式为解决这一困境提供了新的技术路径，该范式通过构建预测性预训练任务，能够从海量未标注数据中学习具有丰富语义的特征表示<sup>[8-10]</sup>。近年来，自监督学习在时间序列分析领域展现出巨大潜力，针对脑电图数据，自监督学习有望构建更具鲁棒性和可扩展性的分析模型，有效应对脑电信号固有的复杂性，从而规避传统方法中繁重的人工标注过程<sup>[11,12]</sup>。

然而，现有的自监督学习方法不能很好地应对实际医疗场景下脑电图的数据分布偏移问题。模型尽管在离线测试时表现良好，但在实际部署时由于数据分布偏移而表现不佳。若模型缺乏足够的鲁棒性，则难以适应真实临床场景中的复杂条件，进而可能导致诊断结果的准确性下降，甚至对患者造成潜在风险。因此，研究如何利用自监督学习提升脑电图分类模型的鲁棒性，对于推动人工智能技术在脑电图的实际应用具有重要理论价值和实践意义。

## 1.2 国内外研究现状

### 1.2.1 脑电图分类研究现状

脑电图（EEG）信号作为一种典型的生理时间序列，其分析与分类是生物医学工程与计算机科学交叉领域的重要研究课题。脑电图分类本质上属于时间序列分类的范畴，具体而言，是医疗时间序列分类的一个子问题。当前，脑电图分类的研究方法主要沿袭时间序列分类领域的技术路线：基于传统机器学习的经典方法和基于深度学习的现代方法。

在传统机器学习范式下，研究者们发展了多种具有代表性的技术路径。早期时间序列分类研究中，基于距离度量的方法占据重要地位，主要包括基于欧氏距离的策略以及基于动态时间规整（DTW）<sup>[13]</sup>的策略。动态时间规整算法凭借其弹性匹配机制，能够有效处理时间序列在时间轴上的非线性形变问题。此类方法在小规模数据集上表现出较好的分类性能，然而，当应用于大规模数据集时，其计算效率与分类精度均面临显著挑战，这一点在处理大规模脑电数据集时尤为突出。另一重要的研究方向是基于特征的方法。其中，基于 Shapelet<sup>[14]</sup> 特征的方法通过挖掘具有判别性的局部子序列模式进行分类，其核心在于识别能够最有效地区分

不同类别时间序列的代表性片段。基于字典的方法则采用符号化表征策略，将原始时间序列转化为符号序列，并利用符号特征进行建模，BOSS<sup>[15]</sup> 算法即为该类方法的典型代表，它通过构建符号化字典来捕捉时间序列的全局与局部特征模式。此外，基于区间统计特征的方法着眼于从时间序列的特定区间提取统计量作为分类依据，例如 TSF<sup>[16]</sup> 算法通过计算各区间段的均值、方差和斜率等统计特征，并结合随机森林分类器实现有效分类。尽管这些传统方法在特定应用场景下取得了一定的成功，但它们普遍面临若干根本性局限。这些方法高度依赖人工设计的特征提取过程，需要领域专家投入大量精力进行特征工程。更为重要的是，在面对脑电图分析中常见的大规模、标签稀缺且存在数据分布偏移的数据集时，这些传统方法的性能往往难以满足实际应用的需求。

随着深度学习技术的迅猛发展，基于神经网络的时间序列分类方法逐渐成为研究主流，为克服传统方法的局限性提供了新的途径。卷积神经网络（CNN）因其卓越的局部特征提取能力在该领域备受关注。例如，OS-CNN<sup>[17]</sup> 采用一维卷积神经网络，通过滑动卷积核有效捕获时间序列的局部时序模式；InceptionTime<sup>[18]</sup> 模型借鉴了计算机视觉领域的 Inception 结构，构建多尺度特征提取网络，以适应时间序列的多尺度特性；RTFN<sup>[19]</sup> 模型则专注于挖掘时间序列局部特征之间的关联性，从不同数据位置发现潜在的时序规律。然而，传统卷积神经网络在处理时间序列的长程依赖性方面存在固有局限。为弥补这一不足，研究者们开始探索基于 Transformer<sup>[20]</sup> 架构的新型方法。此类方法利用自注意力机制，能够有效建模时间序列中的长距离依赖关系<sup>[21]</sup>，例如将标准 Transformer 框架应用于多元时间序列分类问题，以及通过结合一维卷积与注意力机制的混合架构来增强全局特征表示能力<sup>[22]</sup>。

在脑电图领域，Varl<sup>[23]</sup> 等人提出了一种结合时频图像变换与原始时域信号联合建模的深度学习混合模型。该方法将连续小波变换和短时傅里叶变换生成的时频图像特征与长短期记忆网络提取的时序特征相融合，构建了 CNN-RNN 联合架构，在小型的癫痫分类数据集上优于传统机器学习方法。Siddhad<sup>[24]</sup> 针对脑电信号分类中传统方法严重依赖手工特征提取的问题，创新性地将 Transformer 网络架构迁移至 EEG 数据分析领域，通过构建端到端的分类模型，在未经特征提取的原始脑电数据上实现了年龄性别分类与心理负荷分类，显著降低了传统脑电分析对复杂预处理流程的依赖性。

尽管深度学习方法在脑电图分类任务中取得了显著进展，但仍面临两个关键挑战：其一，监督学习范式对标注数据的依赖性限制了其在标签稀缺场景下的应用潜力；其二，现有方法对脑电图中普遍存在的数据分布偏移问题缺乏有效的鲁

棒性保障机制。这些问题的存在促使研究者们开始探索更具适应性和鲁棒性的新型时间序列分类范式。

### 1.2.2 自监督学习研究现状

自监督学习分类学尚未确定，本章选取与本研究相关的自监督学习方法加以介绍：对比学习作为自监督学习领域影响力较大的技术范式之一，其理论体系和研究方法近年来取得了突破性进展。该范式的核心思想源自计算机视觉领域提出的实例判别任务，通过构建样本间的对比关系实现无监督特征学习<sup>[25]</sup>。从理论发展脉络来看，对比学习技术的发展呈现出明显的阶段性特征。早期研究主要采用基于负样本的对比学习架构，其核心机制是通过在特征空间构建样本间的对比关系来实现特征学习。具体而言，这类方法首先对原始样本进行数据增强生成不同视图，然后将同一样本的不同增强视图作为正样本对，不同样本的视图作为负样本对。通过优化特征表示使得正样本对在隐空间中相互靠近而负样本对相互远离，从而学习到具有判别性的特征表示。

SimCLR<sup>[26]</sup> 系列研究做出了开创性贡献，系统阐述了对比学习的四个关键组件设计原则：一是强数据增强策略，包括随机裁剪、色彩抖动、高斯模糊等组合；二是灵活的编码器网络，通常采用 ResNet 等主干网络；三是非线性投影头，将高维特征映射到低维对比空间；最后是归一化温度对比损失，平衡难易负样本的梯度贡献。SimCLR 证明了大批量训练的重要性，但同时也面临计算资源需求高的挑战。随后 MoCo<sup>[27]</sup> 系列工作通过创新的内存队列机制和动量编码器设计，巧妙解决了批量大小限制问题。MoCo 将对比学习重新表述为动态字典查询过程，其中查询编码器负责当前批次表示，而动量更新的键编码器维护一个大型队列，存储历史批次的表示作为负样本。这种设计使模型能够使用较小批量实现大量负样本对比，显著提升了训练效率和表示质量。紧接着，SwAV<sup>[28]</sup> 开创了基于聚类的对比学习新范式。与传统对比学习直接比较样本对不同，SwAV 通过学习原型（prototypes）向量集合，将表示空间隐式划分为多个语义聚类。其核心创新在于“多视角代码交换”机制：对于同一图像的不同增强视图，SwAV 不是直接拉近它们的特征表示，而是要求这些视图分配到相同的原型聚类，实现一种软聚类一致性约束。随着研究的深入，学者们逐渐认识到过度依赖强数据增强可能导致原始数据结构失真，产生增强样本与原始样本之间的分布差异，反而会损害模型性能<sup>[29]</sup>。针对这一问题，最新研究<sup>[30]</sup> 提出了基于弱增强监督信号的新型对比学习框架。该框架通过将弱增强样本作为监督信号来指导强增强样本的特征学习，有效缓解了强增强可能导致的过拟合问题。

生成式人工智能作为当前人工智能领域最具突破性的研究方向之一，其核心在于通过数据驱动的方式学习复杂概率分布，进而创造具有高度真实性的新内容。生成式架构通过遮挡输入数据的部分区域，训练神经网络重构被遮蔽部分，从而实现对数据分布的隐式建模<sup>[31]</sup>。已有研究表明，理想的嵌入表示应当对输入数据的各类变换（例如视觉领域的旋转、裁剪或其他预定义任务扰动）保持稳定性，即学习到的特征表示应具有变换不变性<sup>[32]</sup>。当前主流的生成式技术范式主要包括生成对抗网络（Generative Adversarial Networks）、基于注意力机制的Transformer架构以及变分自编码器（Variational Autoencoder）三大类。

生成对抗网络作为机器学习框架的重要分支，已成为实现生成式人工智能的主流方法之一。该框架的创新性主要体现在其不依赖于大量标注训练数据的特性上，与传统基于监督学习的深度神经网络相比具有显著差异<sup>[33]</sup>。该网络体系主要由生成器与判别器两大核心组件构成，其中生成器通过潜在空间中的随机噪声持续合成新的数据样本，而判别器则负责鉴别生成样本与真实样本之间的差异。在训练过程中，判别器以目标领域的真实数据分布作为参照基准，通过二元分类任务来评估生成器输出的真实性。整个系统的优化过程通过对抗性训练机制使得生成器与判别器在动态博弈中不断进化，直至判别器无法有效区分生成样本与真实样本为止<sup>[34]</sup>。这种独特的对抗机制从根本上改变了生成模型的训练范式，为构建高保真度的数据生成系统提供了新的技术路径。

在技术发展初期，基于生成对抗网络的模型面临着训练不稳定性和模式崩溃等核心挑战<sup>[35]</sup>。训练不稳定性主要表现为生成器与判别器难以达到动态平衡状态，导致模型输出结果出现显著振荡；模式崩溃则体现为生成器陷入局部最优解，仅能生成有限多样性的重复样本。针对这些技术瓶颈，研究者提出了基于Wasserstein距离的改进架构 WGAN<sup>[36]</sup>。该架构采用地距离作为模型优化的目标函数，通过引入 Lipschitz 连续性约束使得梯度更新更加平滑稳定。其创新性地采用评价网络替代传统判别器结构，输出连续评分而非离散概率值，这一改进使得模型在视频序列生成、三维几何建模以及高分辨率纹理合成等复杂任务中展现出更强的鲁棒性和泛化能力。

由 Vaswani 等人提出的 Transformer<sup>[20]</sup> 架构标志着生成式人工智能技术的重大突破。在自然语言处理应用中，研究者通过将位置编码信息与输入词向量相结合，使模型能够有效捕捉序列数据中的长程依赖关系。这种基于自注意力机制的架构已成为构建生成式预训练 Transformer 的基础，使得模型能够生成语义连贯且上下文相关的文本内容<sup>[37]</sup>。值得注意的是，包括 BERT 和 GPT 系列在内的多个里程碑式模型均建立在 Transformer 架构之上，这些模型通过大规模预训练和微调策略，在各类自然语言生成任务中取得了突破性进展。

Kingma 等人提出的变分自编码器是生成式人工智能领域的另一重要技术路线<sup>[38]</sup>。该模型采用编码器-解码器结构作为核心框架，其中编码器负责将高维输入数据映射到低维潜在空间，解码器则负责从潜在表示重构原始数据分布。在模型优化过程中，通过引入标准正态分布作为潜在空间的先验约束，采用变分推断方法最大化证据下界。这种概率化建模方法不仅提供了学习数据本质特征的理论保证，还能通过潜在空间的随机采样实现新样本的生成。

自监督学习虽然可以不依赖于人工标注进行模型训练，能够有效缓解标注数据稀缺的问题，但在实际医疗的复杂场景中，面对成像设备噪声、运动伪影等干扰因素时，现有方法的鲁棒性仍然面临严峻挑战。这主要体现在模型对噪声分布的敏感性以及特征表示的稳定性等方面，需要通过改进网络架构和训练策略来进一步提升模型的抗干扰能力。

### 1.2.3 时间序列自监督学习研究现状

时间序列自监督学习作为新兴的研究方向，近年来取得了显著的理论突破和技术创新。

在时间序列表示学习框架方面，研究者们提出了多种创新性方法。其中一种代表性工作是将时间序列解耦为频域周期分量和时域趋势分量的双流架构<sup>[39]</sup>，通过精心设计的数据增强策略降低模型对虚假相关性的依赖。另一项重要研究则针对时间序列特性设计了专门的对比学习框架<sup>[40]</sup>，该框架同时采用强增强策略（如分段重排加噪）和弱增强策略（如幅值扰动），分别用于学习上下文相关性和时序依赖性，从而获得更具判别力的特征表示。其中研究<sup>[21]</sup>首次将 Transformer 架构引入时间序列自监督学习的研究。该研究对标准 Transformer 进行了三项关键改进：首先，仅保留 decoder 结构以适应时间序列特性；其次，将层归一化替换为批归一化以提升训练稳定性；最后，采用基于伯努利分布的序列掩码策略，使模型能够同时学习数据上下文关系和多元时间序列间的交互模式。这一创新为时间序列表示学习开辟了新的技术路径。

在损失函数设计方面，基于对比预测编码<sup>[41]</sup>的双向对比损失函数<sup>[42]</sup>展现出独特优势。该损失函数通过在隐状态空间构建正负样本对，有效提升了特征表示的紧凑性和判别性。此外，基于变分自编码器的解耦表示学习方法<sup>[43]</sup>也取得了重要进展，该方法通过约束特征空间的独立性，不仅提高了模型性能，还增强了特征的可解释性。特别值得关注的是 Data2vec<sup>[44]</sup>框架，该框架摒弃了传统方法中对特定目标的预测方式，转而学习预测输入数据的整体潜在表示。

尽管这些方法在噪声鲁棒性方面有所提升，但在面对实际医疗场景特有的数据分布偏移问题时仍存在局限。脑电图数据具有显著的主体间差异性和环境敏感

性，这促使研究者开发专门的主体感知学习技术。其中一种创新方法通过主体特定的对比损失函数和对抗训练策略<sup>[45]</sup>，有效提升了模型的主体不变性。另一种名为 CLOCS<sup>[46]</sup> 的对比学习方法则通过同时优化空间、时间和患者三个维度的表示相似性，显著改善了跨主体泛化能力。

在脑电信号处理领域，研究者们提出了多种针对性的自监督学习策略。包括通过通道级信号变换（如掩码、幅值调整、加噪和相位移动）构建正样本对的技术<sup>[47]</sup>，以及基于脑半球信号相似性<sup>[48]</sup>、大脑状态识别和脑年龄预测等辅助任务的预训练方法。针对癫痫检测这一特定应用，最新研究提出了双向自回归对比学习框架<sup>[42]</sup>，该框架通过多层次特征学习和基于脑电电极图的消息传递机制，有效缓解了标签不平衡和噪声干扰问题，同时考虑了癫痫发作时的神经信号传播特性。为了应对脑电图数据的复杂性，徐晴等人<sup>[49]</sup> 将一维脑电数据转化为二维图像，利用视觉网络对脑电图像进行分类，达到了良好的分类效果。

#### 1.2.4 现有研究存在的主要问题

脑电图信号在实际采集环境中容易受到噪声的影响，这些噪声不仅干扰了信号的质量，更严重制约了后续分析的准确性，噪声主要源于被试者在数据采集过程中的生理状态变化以及设备配置差异等多重因素。已有研究表明，脑电噪声会引发数据分布偏移现象，其具体表现形式包括但不限于、窄带阻抗噪声干扰、宽带随机噪声污染等，同时，数据预处理的多样性也会引发数据分布偏移，例如带通滤波的参数设置不同会严重影响数据的频谱特性<sup>[6]</sup>。这种数据偏移对维持机器学习模型的稳定性能构成严峻挑战，特别是在带通滤波导致的分布偏移普遍存在的条件下，其负面影响更为突出。

传统去伪迹方法，如滤波、重参考和独立成分分析（ICA），在脑电信号处理中已被广泛应用，其主要优势在于基于明确的信号模型和假设，能够针对特定伪迹（如眼动或工频干扰）实现精准去除。然而，这些方法通常对数据分布的稳定性有较高要求，且在滤波参数选择或参考电极设置上存在主观性，容易因参数不一致导致数据分布偏移，进而影响模型性能。相比之下，自监督学习通过从无标注数据中学习特征表示，不仅减少了对专业标注的依赖，还能增强模型对复杂噪声环境的适应性。尽管自监督方法在计算成本和训练复杂度上可能高于传统方法，但在实际临床场景中，其对数据分布偏移的鲁棒性和泛化能力提供了显著优势。

尽管近年来针对脑电信号的自监督学习已开展大量研究工作<sup>[48,50,51]</sup>，但现有方法普遍存在一个关键缺陷：未能解决医疗场景下数据偏移导致的模型在部署时性能下降的问题。当前的自监督学习范式虽然能够在一定程度上缓解标注数据不

足的困境，但在处理 EEG 信号固有的数据偏移时仍表现出明显的局限性。这种局限性主要体现在模型对噪声分布的敏感性以及特征表示的稳定性等方面，导致其在真实场景中的泛化能力受到显著制约。

这一研究空白凸显了两个问题：首先，如何在自监督学习框架中有效纳入对数据偏移具有鲁棒性的机制；其次，如何设计具有鲁棒的特征表示方法以应对 EEG 信号的数据分布偏移。解决这些问题对于推动脑机接口、神经疾病诊断等实际应用具有至关重要的意义。特别是在医疗健康领域，由于脑电信号采集环境的复杂性（如医院设备电磁干扰、患者运动伪影等），开发具有强鲁棒性的自监督学习方法显得尤为迫切。针对上述问题，现有方法的不足主要表现在三个层面：在信号预处理阶段，传统方法往往依赖于固定的噪声假设，难以适应实际场景中的动态变化；在特征学习阶段，多数自监督任务未能充分考虑 EEG 信号的时频特性与噪声分布的关联性；在模型优化层面，缺乏针对噪声鲁棒性的专项损失函数设计。这些局限性共同导致了当前方法在复杂噪声环境下的性能退化问题。虽然已经有一些工作<sup>[52-54]</sup> 对提到模型的鲁棒性做出了贡献，但是这些工作并不能很好地提高模型对实际医疗场景的数据分布偏移的鲁棒性。

开发新型自监督学习框架，使其能够应对数据分布偏移带来的挑战，成为当前研究的重要方向。这需要从特征表示和模型优化等多个维度进行系统性创新，包括但不限于：建立更符合实际噪声分布的数据生成模型，设计具有偏移不变性的自监督预训练任务，以及开发能够区分有效神经活动与噪声干扰的特征学习机制。这些技术创新将为脑电信号分析在实际医疗环境中的可靠应用奠定重要基础。

### 1.3 本文的主要研究内容

本文围绕脑电图在数据分布偏移下的鲁棒分析展开，旨在解决脑电图信号处理中面临的标签稀缺性与数据分布偏移问题。针对现有自监督学习方法在处理脑电信号数据分布偏移方面的不足，本研究重点探索了基于自监督学习的鲁棒增强的时间序列分类算法，主要研究内容包括以下几个方面：

首先，本研究提出了一种面向脑电时序分类的鲁棒预训练框架。该框架创新性地整合了时频域双向重构学习机制、通道注意力对齐策略和含噪样本的对比学习方法，形成一个多任务联合优化的预训练体系。具体而言，该框架在预训练阶段创新性地采用时频域双向重构学习机制，使特征表示能够同时捕获时域和频域的关键信息；通过设计通道注意力对齐策略，有效提升模型对时域噪声的鲁棒性；此外，通过引入含噪样本的对比学习框架，进一步增强模型对噪声模式的辨识能

力。相较于现有方法，本研究所提出的解决方案不仅在噪声鲁棒性方面表现优异，而且在分类性能上也有显著提升。

其次，本研究探索了基于残差混合的对比学习增强方法。该方法通过分析监督模型中间层特征偏移的分布特性，探索噪声强度与特征偏移之间的潜在关联，并设计在嵌入空间实施残差混合的操作机制，尝试建立更灵活的噪声建模方式。与传统数据增强策略不同，该方法在嵌入空间内进行操作，有潜力捕捉更复杂的语义噪声模式。虽然实验表明该方法在不同模型架构中的表现存在差异，但仍为对比学习范式的改进提供了新的研究思路。

最后，本研究在多个脑电图数据集上进行了系统的实验验证，实验结果表明，本研究所提出的鲁棒预训练框架在癫痫发作检测和脑电异常分类任务上均优于现有基准方法，尤其在带通滤波导致的分布偏移等环境下表现出卓越的鲁棒性。通过消融实验进一步验证了各组件对模型整体性能的贡献，并探讨了残差混合策略在不同对比学习框架中的适用边界。

## 1.4 本文的章节安排

本研究聚焦于构建基于自监督学习的脑电图鲁棒分类，文章的章节如下所示：

第一章绪论系统阐述了基于自监督学习的脑电图鲁棒分类研究的背景意义、理论基础和问题。在课题背景部分，重点分析了脑电图技术在神经系统疾病诊断中的核心价值及其面临的挑战。作为非侵入性检测手段，脑电图凭借毫秒级时间分辨率在癫痫诊断、睡眠障碍评估等领域具有不可替代的优势。然而脑电图存在专业标注稀缺和数据偏移的双重困境，传统分类方法在标签稀缺场景下面临根本性局限。自监督学习范式通过预测性预训练任务从海量未标注数据中学习特征表示，为解决这一困境提供了新的技术路径。相关理论发展概况部分深入剖析了时间序列分类与自监督学习两大领域的研究进展。现有研究存在的主要问题部分明确指出，当前方法在应对实际医疗场景特有的复杂噪声组合时表现欠佳。脑电图信号受设备参数、电极部署等多因素产生的数据分布偏移，进而使模型在实际部署中出现性能退化，现有研究在处理带通滤波导致的分布偏移、阻抗噪声等实际干扰时仍缺乏系统性解决方案。这一研究空白凸显了开发具有噪声鲁棒性的自监督学习框架的迫切需求，也为本研究的创新方向提供了明确指引。

第二章阐述了本研究涉及的两大核心技术基础：自监督学习与 Transformer 架构。在自监督学习部分，重点分析了对比学习和掩码重构学习两大技术路线的原理与特点。对比学习通过构建正负样本对优化特征空间分布，其 InfoNCE 损失函数源自互信息最大化原理，能够学习具有良好迁移性和鲁棒性的特征表示。掩码

重构学习则属于生成式自监督学习范畴，通过预测输入数据中被掩盖部分来挖掘数据内在结构，在时间序列处理中展现出独特优势。研究表明，两种方法具有互补性，混合策略往往能获得更优性能。Transformer 架构部分详细解析了其核心组件和工作原理。自注意力机制通过动态计算序列元素间相关性权重，有效解决了长距离依赖问题；多头设计进一步增强了模型对复杂序列模式的理解能力。位置编码采用三角函数为序列注入顺序信息，前馈网络则提供非线性变换能力。结合残差连接和层归一化技术，该架构在并行计算和全局依赖建模方面展现出显著优势，为后续研究奠定了坚实的技术基础。这些核心技术共同构成了本研究的理论支撑和方法框架。

第三章系统性地提出了基于自监督学习的脑电图鲁棒分类框架，针对脑电信号分析中存在的数据分布偏移、标注依赖性强及时频特征利用不充分等核心问题，创新性地融合了噪声增强对比学习、跨域重构与通道对齐三大关键技术。本章首先明确了研究问题的形式化定义，将 EEG 信号表征为多通道时间序列矩阵，并建立了带通滤波导致的分布偏移、阻抗噪声和高斯噪声三类典型干扰的数学模型。通过对现有方法局限性的深入分析，揭示了传统对比学习在处理带通滤波导致的分布偏移时存在的固有缺陷，这为后续方法创新提供了理论依据。在方法设计层面，本研究提出了三大核心创新机制：首先，引入额外噪声负样本的对比学习策略通过显式建模噪声分布特征，显著提升了模型对复杂干扰的判别能力。其次，设计的双向跨域重构任务构建了时频特征的系统性映射机制，通过时域到频域和频域到时域的双向预测损失，实现了时频特征的深度融合。最后，创新的通道对齐方法利用频域通道关系的稳定性来校正时域表征，解决了噪声环境下通道关系偏移问题。模型实现采用分层处理架构，包含数据预处理、跨域重构学习、时频通道对齐和噪声增强对比学习四个关键模块。预处理阶段创新性地在频域执行掩码操作，通过带阻滤波提取背景噪声成分。跨域重构模块采用时频双流 Transformer 编码器，分别处理时域和频域信号，并通过预测器实现域间映射。通道对齐模块基于注意力机制建立时频通道关系矩阵，采用梯度截断技术确保优化稳定性。对比学习模块则通过温度缩放交叉熵损失，在时频双域同步实施噪声增强的对比优化。这些创新为生物医学信号处理领域提供了重要的方法学参考，也为解决实际医疗环境中的信号变异问题开辟了新途径。

第四章进一步聚焦于自监督学习中的对比学习范式，针对其在脑电信号分析中的噪声鲁棒性问题探索性地提出了基于残差混合的增强策略，并通过系统性实验揭示了该方法在不同模型架构中的差异化表现。与第三章提出的整体框架不同，本章的研究是对对比学习范式本身的局部优化，旨在通过在嵌入空间实施残差混

合操作，缓解传统数据增强方法在多样性不足的局限性，两章之间形成“整体框架设计”与“局部技术改进”的递进关系，共同推动了脑电图鲁棒分类领域的研究进展。研究动机部分通过监督模型中间层特征偏移分析，发现不同噪声强度导致的特征偏移具有可替代性特征，这为残差混合策略提供了理论依据。在方法设计层面，本章提出在嵌入器内部选取特定网络层进行残差混合操作的技术路线。尽管最终性能提升有限，本章研究仍具有重要的方法论价值：其一，通过对照实验明确了残差混合策略的适用边界，为后续研究提供了关键参考；其二，揭示了嵌入空间数据增强的潜在风险，警示了简单迁移计算机视觉技术的局限性。这些发现不仅丰富了对比学习理论体系，也为脑电图分析的鲁棒性研究开辟了新视角。

## 第2章 相关技术基础

### 2.1 自监督学习

自监督学习作为无监督学习的重要分支，近年来在深度学习领域展现出强大的特征学习能力。其核心思想是通过设计合理的预训练任务，从无标注数据中自动生成监督信号，从而学习具有丰富语义的特征表示。与传统监督学习相比，自监督学习最大的优势在于能够充分利用海量无标注数据进行预训练，有效缓解医疗等领域标注数据稀缺的问题。

从方法体系来看，自监督学习有两条重要的技术路线：生成式和对比式两大技术路线。生成式方法通过重构输入数据或其变换形式来学习特征表示，典型代表包括自编码器和掩码模型。这类方法通过最小化重构误差迫使模型捕捉数据的关键特征，在图像和文本领域取得了显著成功。对比式方法则通过构建正负样本对，学习区分不同样本的特征表示。其核心思想是使相似样本在特征空间中彼此靠近，而不相似样本相互远离。

#### 2.1.1 对比学习

对比学习的基本思路是构建语义相关的正样本对和语义无关的负样本对，通过最大化正样本对在特征空间中的相似度，同时最小化与负样本对的相似度，从而学习具有区分性的特征表示。对比损失函数源自互信息最大化原理，旨在最大化不同视角下同一数据点的互信息。SimCLR<sup>[26]</sup> 框架代表了对比学习方法的重大突破。如图2-1所示，该框架采用了一种高效的小批量处理策略：对每个原始样本进行两次独立增强，生成  $2N$  个增强样本 ( $N$  为训练批次的大小)。该方法的损失函数基于归一化的余弦相似度度量构建，通过引入温度系数调节的 softmax 运算实现对比优化。

其对应的数学表达式为公式 (2-1)。

$$\mathcal{L}_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)} \quad (2-1)$$

式中  $N$ ——训练批次的大小；

$\text{sim}$ ——相似度计算函数；

$\mathcal{L}_{i,j}$ ——锚点样本与正样本的距离损失函数；

$\tau$ ——温度系数。

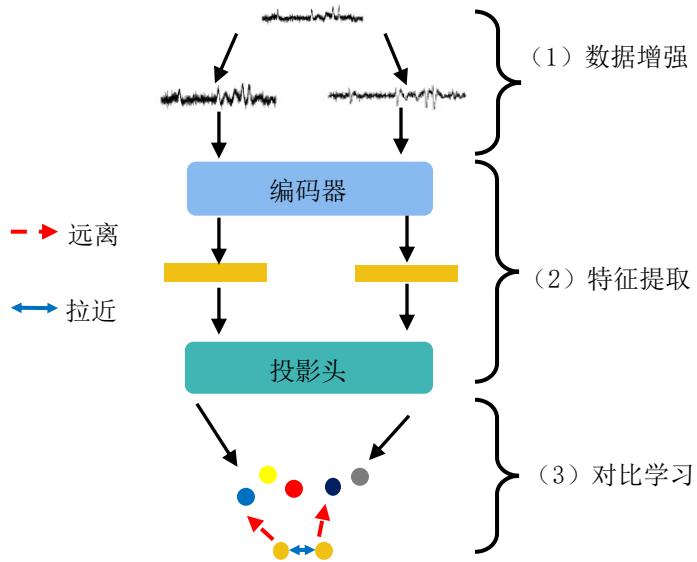


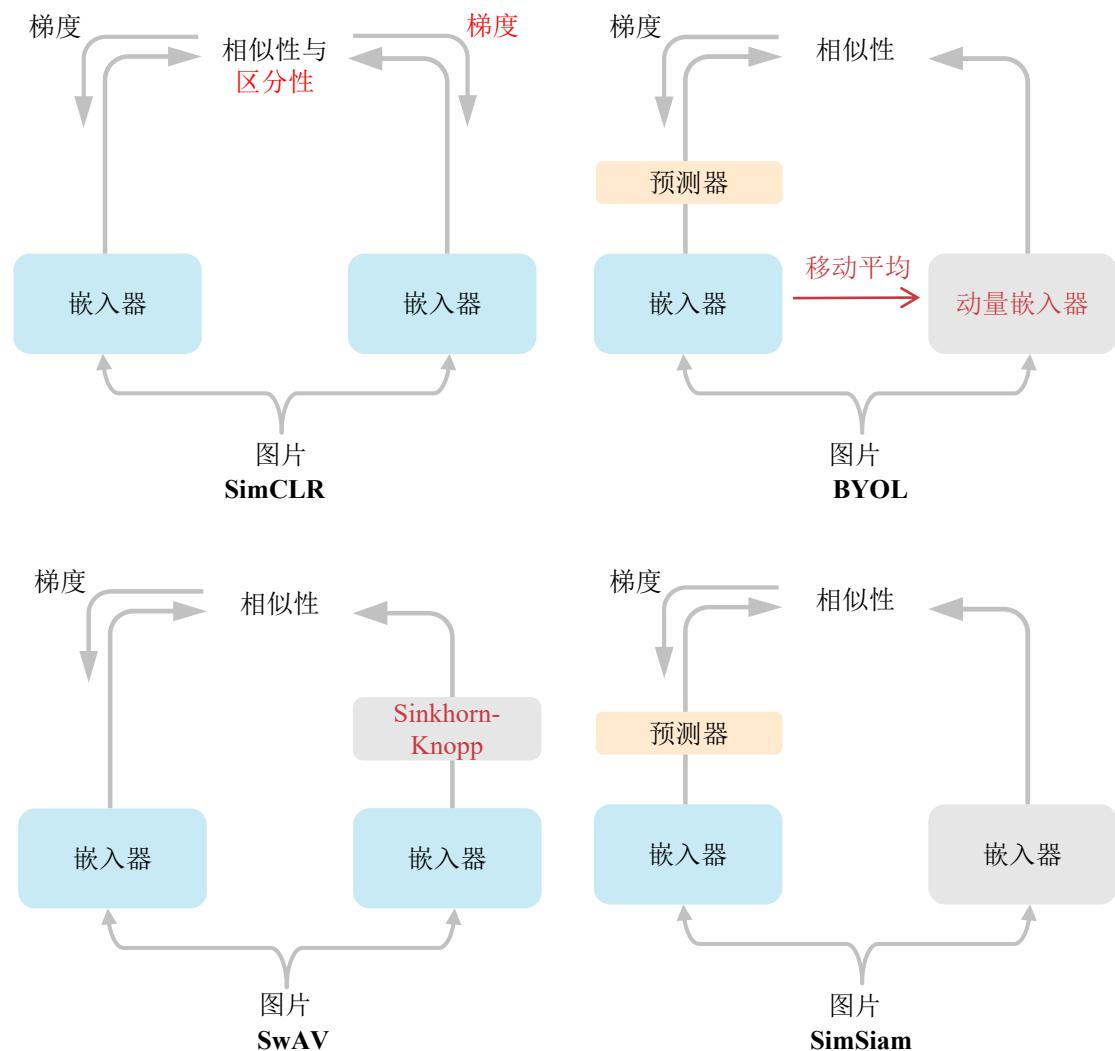
图 2-1 SimCLR 架构图

在 SimCLR 的基础上, BYOL<sup>[55]</sup>、SimSiam<sup>[56]</sup> 和 SwAV<sup>[28]</sup> 作为两种代表性框架, 进一步推动了对比学习的发展。BYOL 无需使用负样本对即可学习有效的视觉表征, 而是通过构建在线网络与目标网络的双分支结构, 使前者预测后者对相同图像不同视图的表示, 并通过指数移动平均更新机制维持目标网络参数, 从而实现表征学习的自举过程。SimSiam 是一种无需负样本对比、不依赖大批量训练的自监督表征学习方法, 其核心在于采用孪生网络结构, 通过对同一图像的两个增强视图进行表征学习, 并巧妙运用停止梯度操作防止模型坍塌, 实现了高效且有效的特征提取能力。SwAV 则创新性地引入聚类机制, 通过“交换赋值”策略对同一数据的不同增强视图进行在线聚类, 利用伪标签指导特征学习, 避免显式构建负样本对。这两种方法分别通过动态字典和聚类任务优化了对比学习的目标, 在无监督特征学习任务中均展现出优异性能, 为后续研究提供了重要启发。上述对比学习方法如图2-2所示。

对比学习在实践中展现出多方面的优势。首先, 对比学习能够学习到具有良好迁移性的特征表示, 这些特征在下游任务中表现出色, 尤其是在标注数据有限的情况下。其次, 对比学习习得的特征往往具有更强的鲁棒性, 对噪声、域漂移等干扰因素表现出更好的抵抗力。最后, 对比学习的训练过程更加稳定, 相比生成式方法, 不易陷入嵌入空间坍塌等训练困境。

### 2.1.2 生成式自监督学习

生成式自监督学习是自监督学习中另一个重要的技术路线, 其核心思想是通过预测或重构输入数据中被掩盖的部分来学习有效的特征表示。与对比学习不同,

图 2-2 几种对比学习方法架构的比较<sup>[56]</sup>

生成式自监督学习属于生成式自监督学习范畴，它不依赖样本间的比较，而是通过自我监督的方式，从数据本身挖掘监督信号。

生成式自监督学习的基本范式可以追溯到经典的自编码器模型，该模型通过重构完整的输入数据来学习特征表示。然而，简单的重构任务往往导致模型学习到的是数据的浅层统计特性，而非深层语义信息。为了克服这一局限，生成式自监督学习引入了掩码操作，通过随机遮盖输入数据的部分内容，强制模型从可见部分推断被掩盖的信息，从而学习到更有意义的特征表示。生成式自监督学习在实现上涉及多个关键组件，其中最重要的是编码器-解码器架构，其中编码器负责提取输入数据的特征表示，解码器则基于这些特征重构被掩盖的内容。根据具体应用场景，编码器和解码器可以采用不同的网络结构。

在时间序列领域，时间序列数据的连续性和时序依赖性为掩码重构提供了天

然的基础。研究者通常采用多种掩码策略来适应时间序列的特性。最常见的方法是随机掩盖时间序列中的部分时间点或时间段，如图2-3所示，然后训练模型从上下文信息中重构这些缺失值。这种方法不仅能够学习时间序列的时序模式，还能增强模型对缺失值和异常值的鲁棒性。

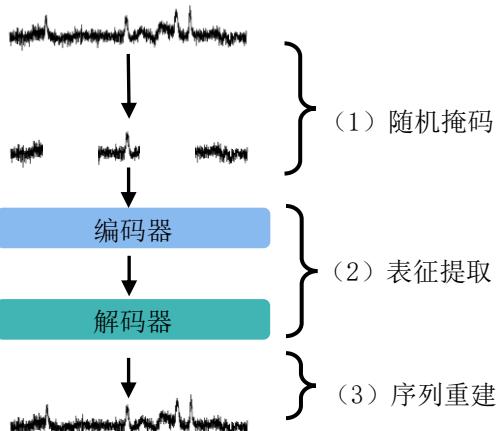


图 2-3 时间序列掩码重构的抽象模型

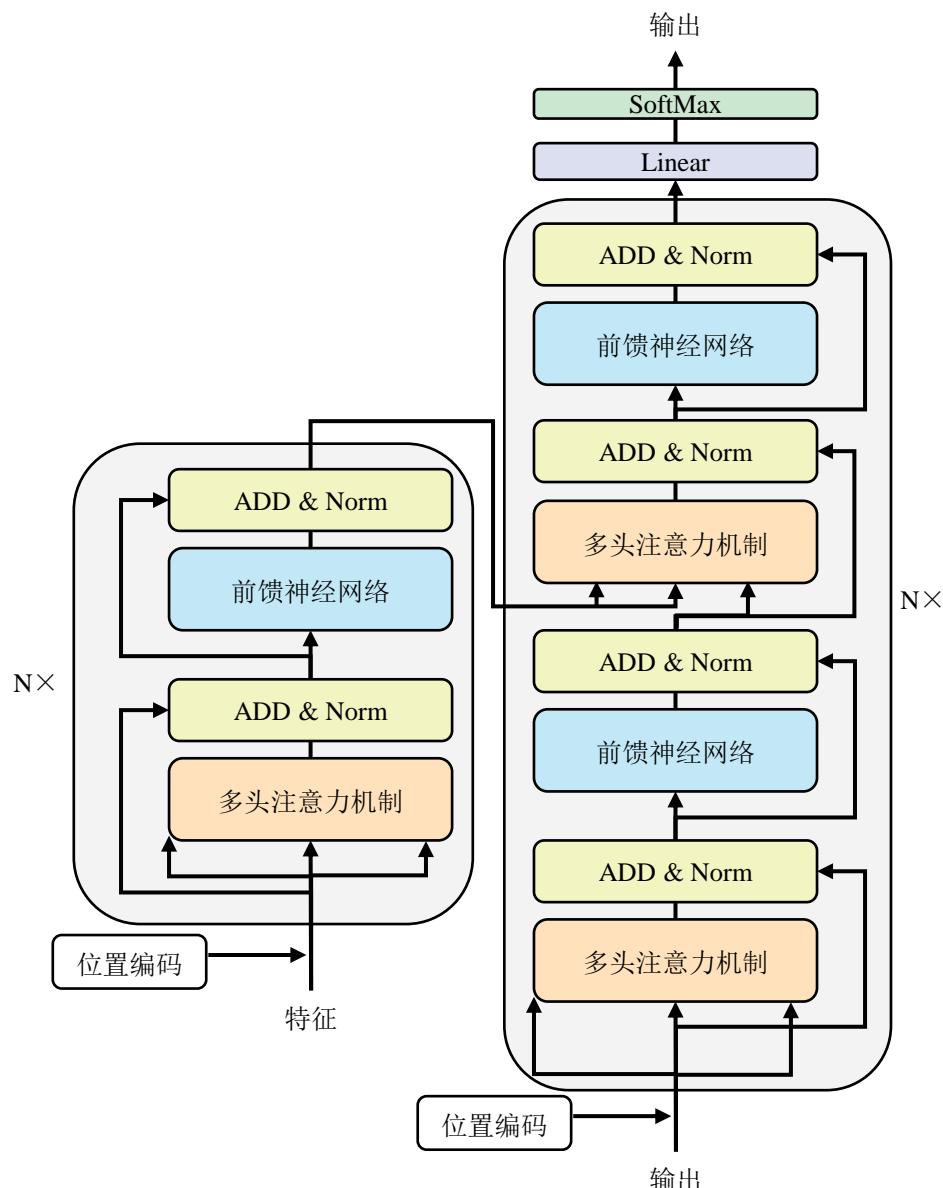
生成式自监督学习与对比学习并非相互排斥，而是可以互补结合。近年来，研究者提出了多种混合方法，既利用生成式自监督学习捕捉数据的内在结构，又通过对比学习增强特征的判别性。这种混合策略在多个领域展现出优于单一方法的性能，这也反映了不同自监督学习范式各自的优势和局限。

## 2.2 Transformer

Transformer<sup>[20]</sup>作为一种基于自注意力机制的神经网络架构，其核心创新在于完全摒弃了传统的循环和卷积结构，仅依赖注意力机制进行序列建模，在并行计算能力和长程依赖建模方面展现出显著优势。该架构通过特殊的设计对序列数据进行高效处理，使其在自然语言处理、计算机视觉等多个领域取得了突破性进展。

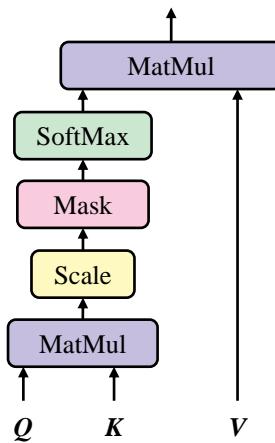
如图2-4所示，Transformer 的核心组件包括自注意力层、位置编码、前馈网络以及残差连接与层归一化机制等。其中，自注意力层通过计算序列元素之间的注意力权重，实现对全局依赖关系的动态捕捉。在传统循环神经网络中，信息需要沿着序列逐步传递，而自注意力机制则允许模型直接建立任意两个位置之间的关联，从根本上解决了长距离依赖问题。

如图2-5所示，自注意力层作为Transformer 架构的核心组件，通过动态计算序列元素间的相关性权重，实现了对长距离依赖的高效建模。该机制摒弃了传统循环神经网络的串行处理方式，允许序列中所有位置直接交互，在并行计算能力和上下文感知方面展现出独特优势。注意力计算的基本形式可以表示为矩阵乘法与

图 2-4 Transformer 架构图<sup>[20]</sup>

缩放 softmax 操作的组合，由公式 (2-2) 所示，其中查询矩阵、键矩阵与值矩阵构成了注意力机制的三个核心要素。通过缩放因子的引入，该机制有效稳定了梯度传播过程，提升了模型训练的稳定性。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2-2)$$

图 2-5 自注意力机制<sup>[20]</sup>

式中  $Q$ ——查询矩阵；

$K$ ——键矩阵；

$V$ ——值矩阵；

$d_k$ ——键向量的维度，缩放因子用于稳定梯度。

多头注意力机制的引入进一步增强了模型的表征能力。通过并行运行多组独立的注意力计算单元，每个注意力头可以学习不同的关注模式和特征关联，从而捕获序列中的多种语义关系。最终，所有注意力头的输出被拼接并通过线性变换层融合，形成更为丰富和综合的特征表示。这种多头设计显著提升了模型对复杂序列模式的理解能力，特别是在处理具有多级语义结构的数据时表现尤为突出。

位置编码是 Transformer 架构中另一个关键创新。由于模型完全基于自注意力机制而不包含任何循环或卷积结构，序列中的顺序信息容易丢失。为解决这一问题，Transformer 通过注入精心设计的位置编码向量来保持序列的顺序信息。常用的正弦位置编码利用三角函数的特性，如正弦编码(2-3)，为不同位置生成唯一的编码向量，使模型能够感知元素在序列中的相对与绝对位置。这种编码方式不仅提供了位置信息，还保持了相对位置之间的平滑关系，便于模型学习位置相关的模式。

$$\text{PE}_{(\text{pos},2i)} = \sin(\text{pos}/10000^{2i/d_{\text{model}}}) \quad (2-3a)$$

$$\text{PE}_{(\text{pos},2i+1)} = \cos(\text{pos}/10000^{2i/d_{\text{model}}}) \quad (2-3b)$$

式中  $\text{pos}$ ——位置索引；

$\text{PE}$ ——位置编码。

前馈网络作为 Transformer 中的另一核心组件，对每个位置独立应用参数共享的两层全连接网络，在模型中引入非线性变换能力。这种点式前馈网络增强了模

型的表达能力，使其能够学习更复杂的特征变换。与此同时，Transformer 还采用了残差连接和层归一化技术，每个子层输出都会与其输入进行残差相加，并应用层归一化操作。这种设计有效缓解了深层网络训练中的梯度消失问题，提升了训练稳定性和收敛速度。

综合而言，Transformer 通过自注意力机制、位置编码、多头设计以及归一化等创新组件，构建了一种高效且灵活的序列建模框架。其并行计算特性和全局依赖捕获能力使其成为现代深度学习中最具影响力的架构之一，为后续的大型语言模型和多模态模型奠定了坚实基础。

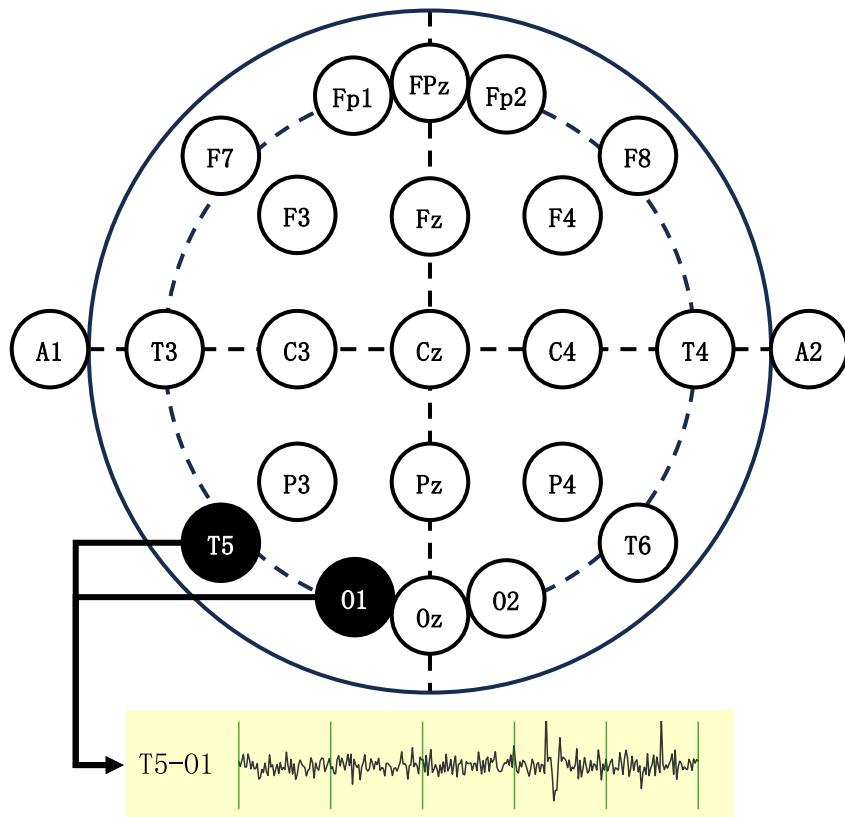
### 2.3 脑电图与癫痫检测

脑电图 (EEG) 是一种非侵入性的神经生理学检测技术，通过在头皮上放置电极记录大脑皮层的电活动，为研究和诊断脑部相关疾病提供了重要手段。特别是在癫痫研究领域，脑电图因其能够捕捉大脑异常放电活动而成为不可或缺的工具。相较于其他脑成像技术，脑电图具有成本低、操作简便以及高时间分辨率的优点，使其在临床和科研中广泛应用。脑电图的核心原理是通过头皮电极测量神经元群体活动产生的电位变化。这些电位变化通常以微伏级别存在，因此需要精密的放大和记录设备来获取准确数据。在记录过程中，脑电图能够反映大脑在不同状态下的活动模式，例如正常状态、睡眠状态以及癫痫发作时的异常放电。

在实际应用中，脑电图数据的采集通常遵循国际 10-20 系统进行电极放置。该系统的命名源于电极间距为鼻根到枕骨隆起总距离的 10% 或 20%，通过头皮上的解剖标志点（如鼻根和枕骨隆起）确定电极位置，确保数据在空间上的一致性和可比性。本研究采用的数据集——天普大学医院脑电图数据集 (TUH EEG)，即遵循这一标准进行数据采集。图2-6展示了 TUH EEG 数据集的电极配置，上半部分呈现了基于 10-20 系统放置的电极位置，下半部分则分别展示了睁眼和闭眼状态下的双极电位差图像，直观反映了不同生理状态下脑电信号的差异。

TUH EEG 数据集由天普大学医院提供，包含了大量的临床脑电图记录，是癫痫检测研究的重要资源。该数据集的信号通常通过多个通道采集，每个通道对应头皮上的特定位置，通道数量根据配置不同可在 20 至 30 个之间，常见采样率为  $250Hz$  或  $256Hz$ ，足以捕捉脑电图信号的细微变化。此外，TUH EEG 数据集的一个显著特点是其标注成本较高。癫痫相关事件的标注需由专业神经科医生手动完成，涉及识别发作期 (ictal)、发作前 (preictal) 和发作间期 (interictal) 等状态，这一过程不仅耗时长，且对专业知识要求极高。

综上所述，脑电图作为一种高效的脑活动监测手段，在癫痫研究中具有不可

图 2-6 天普大学医院脑电图数据集电极配置设置图<sup>[7]</sup>

替代的价值。随着时间序列自监督学习等技术的不断进步，脑电图将在神经科学和临床医学领域发挥更为重要的作用，为癫痫等脑部疾病的精准诊断和个性化治疗提供有力支持。

## 2.4 本章小结

本章系统回顾了自监督学习、Transformer 架构以及脑电图与癫痫检测的相关技术基础。自监督学习通过对比式和生成式方法有效挖掘无标注数据潜力；Transformer 以自注意力机制为核心，显著提升序列建模能力；脑电图则是本研究的研究载体。这些技术为后续研究奠定了理论与实践基础。

## 第3章 面向脑电时序建模的鲁棒预训练框架

### 3.1 引言

实际临床环境中采集的脑电信号普遍存在复杂的噪声干扰，严重影响了信号分析的准确性和可靠性。当前基于深度学习的 EEG 分析方法主要面临三个关键挑战：首先，传统监督学习方法高度依赖大规模标注数据，而医疗领域高质量标注数据的获取成本高昂且存在专业壁垒；其次，现有方法对噪声干扰的鲁棒性不足，特别是对频域选择性噪声（如带通滤波偏差）的抵抗能力较差<sup>[6]</sup>；最后，大多数模型未能充分利用 EEG 信号在时域和频域的互补特征，导致信息利用不充分。这些问题严重制约了脑电分析模型在真实临床场景中的应用效果。

本研究中提到的“带通滤波偏差”并非传统意义上的噪声，而是指由于脑电设备或数据预处理中带通滤波参数设置的差异（如低截断频率和高截断频率的不一致），导致信号频谱内容的部分丢失或限制，进而引发数据分布偏移的现象。这种偏差在实际临床环境中普遍存在，可能显著影响模型的泛化性能。为区分其与高斯噪声等传统噪声的差异，本文将其定义为一种特殊的预处理偏差，并在实验中系统评估其对模型性能的影响。

针对上述挑战，本研究提出了面向脑电时序建模的鲁棒预训练框架。该框架引入额外噪声负样本的对比学习策略，通过建模噪声模式与有效信号模式的关系，增强模型对噪声干扰的鲁棒性。同时，该框架使用跨域重构任务进行预训练，强制模型在时域和频域之间建立双向去噪，实现时频特征的深度融合与提升模型鲁棒性；最后，该框架在训练中使用时频通道对齐任务进行预训练，提高模型的通道关系的鲁棒性。实验结果说明了本框架的有效性与鲁棒性。

### 3.2 问题定义

具体而言，给定一个包含  $N$  个数据实例的 EEG 数据集  $X = \{x_1, \dots, x_N\}$ ，其中每个样本  $x$  可以表示为  $\mathbb{R}^{C \times T}$  维的矩阵形式。该矩阵包含  $C$  个通道的 EEG 信号记录，每个通道的特征则由  $T$  个连续时间步的观测值构成。本研究的主要目标在于构建一个参数化的映射函数  $f_\theta$ ，该函数能够将原始的 EEG 样本  $x$  有效地转换到  $M$  维的潜在表示空间，即实现  $f_\theta : x \rightarrow z$  的映射过程，其中  $z \in \mathbb{R}^M$ 。随后， $f_\theta$  就可以与分类头连接，在下游分类任务数据集上进行微调进而实现时间序列分类。在信号处理过程中，本研究特别关注两种重要的数据变换形式：首先， $\hat{x}$  表示经过

数据增强处理后的样本；其次， $x^F$  代表原始信号经过傅里叶变换后得到的频域表示。

通过对现有文献的系统梳理<sup>[6]</sup>，本研究发现主要的噪声干扰可以归纳为三类典型模式：其一是符合高斯分布特性的随机噪声；其二是阻抗噪声，其三是带通滤波偏差。首先，符合高斯分布特性的随机噪声通常表现为宽带噪声，这种噪声往往源于多种环境因素或仪器设备的随机干扰，其频率范围覆盖较广，对脑电信号的整体质量具有显著影响。其二，阻抗噪声主要表现为窄带低频随机噪声，通常与传感器与头皮接触不良或干/湿电极的选择密切相关。这种噪声的频率范围较为集中，常在低频段内对脑电信号产生干扰，其影响程度与接触阻抗的大小以及采集环境的稳定性密切相关。最后，带通滤波偏差则直接关联于脑电设备中硬件层面的带通滤波设置，表现为由于不同设备或制造商在滤波频率范围设置上的差异所导致的信号频谱内容限制。这种噪声的影响主要体现在对信号特定频率成分的削弱或丢失，从而可能对后续的信号分析和特征提取产生一定程度的偏差。这些噪声模式的模型与参数及其影响程度已在表格 3-1 中进行了详细总结和分析。

表 3-1 典型 EEG 噪声及定义表

EEG 噪声	定义	参数	描述
带通滤波偏差	$n_{bp} := \Psi(x, f_L, f_H)$	$f_L \in \{0.5, 1\}$ $f_H \in \{25, 30\}$	$\Psi$ : 带通滤波函数 $f_L, f_H$ : 低, 高截断频率
阻抗噪声	$n_{in} := x + \epsilon$	$\sigma \in \{0.1, 0.2\}$ $z_\sigma \sim \mathcal{N}(0, \sigma^2)$ $\epsilon = \Psi(z_\sigma, 0, 1)$	$\sigma$ : 噪声强度 $z_\sigma$ : 高斯噪声 $\epsilon$ : 高斯噪声中的 0-1Hz 成分
高斯噪声	$n_{bn} := x + \epsilon$	$\sigma \in \{0.1, 0.2\}$ $\epsilon \sim \mathcal{N}(0, \sigma^2)$	$\sigma$ : 噪声强度 $\epsilon$ : 高斯噪声

### 3.3 研究动机

#### 3.3.1 额外噪声负样本

当前时间序列对比学习技术在应对各类噪声干扰时表现出明显的局限性，尤其对带通滤波偏差的处理能力存在显著不足。为验证这一现象，本研究设计了对比实验，实验以采用了典型对比学习方法的 SimMTM<sup>[57]</sup> 为基准模型，首先在天普医院癫痫脑电数据集上完成预训练流程，随后分别输入原始训练样本和经过带通滤波偏差干扰的污染样本，分别测量并比较它们与正样本的相似度。实验结果的

详细可视化展示见图3-1。通过在对比学习空间中正样本与负样本、噪声样本的相似性分布分析发现：以红色区域表示的常规负样本与正样本的相似性分布呈现出相似度小的特点，这表明模型在处理标准负样本时具备良好的区分能力；然而，以蓝色区域表示的噪声样本与正样本的相似性分布呈现出相似度较大的特点，这显示模型难以有效识别和区分噪声模式与有效信号模式。带通滤波偏差的分布特征与标准负样本存在显著差异，这一差异凸显了现有时序对比学习模型在有效区分噪声模式与典型信号模式方面的能力局限。

针对这一关键问题，本研究提出在频域和时域中采用引入额外背景噪声负样本的对比学习的策略。该方法的核心理念是在传统对比学习框架的基础上，同时在时域和频域两个维度上引入经过额外背景噪声样本作为补充负样本。具体而言，在模型训练过程中，本研究不仅保留原有的标准负样本对，还系统地构建了包含各类噪声的负样本对，通过这种显式的噪声模式建模方式，迫使学习模型必须同时关注信号特征和噪声特征的鉴别问题。从特征学习的角度来看，它显著强化了模型对噪声模式与有效信号模式的辨识能力；从数据多样性的角度来看，通过引入接近真实应用场景的噪声干扰，扩展了训练数据的覆盖范围。尽管增广负样本机制在计算机视觉领域已有相关探索<sup>[9]</sup>，但在时间序列分析领域专门针对噪声鲁棒性的应用尚属首次。实验表明，本方法有望显著提升时间序列分析模型在实际噪声环境中的稳定性和预测准确性，为解决这一长期存在的技术难题提供了新的思路。

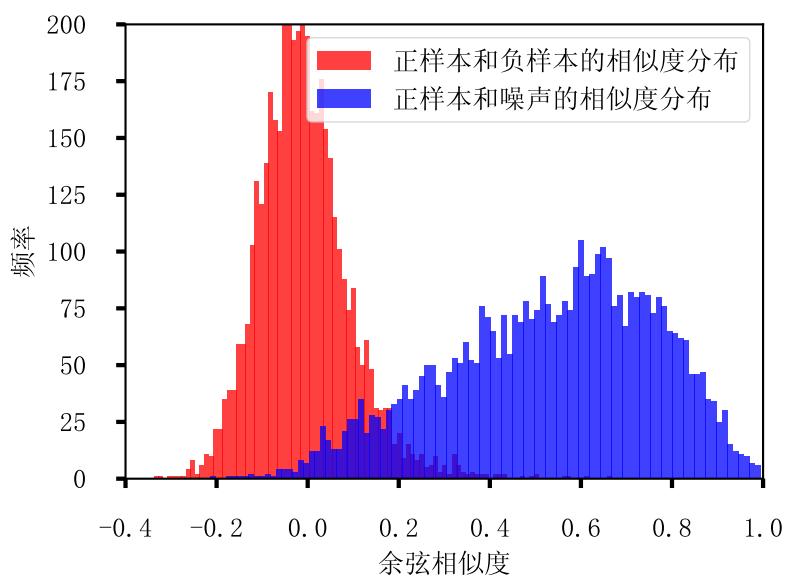


图 3-1 负样本和噪声分别与正样本相似度分布图

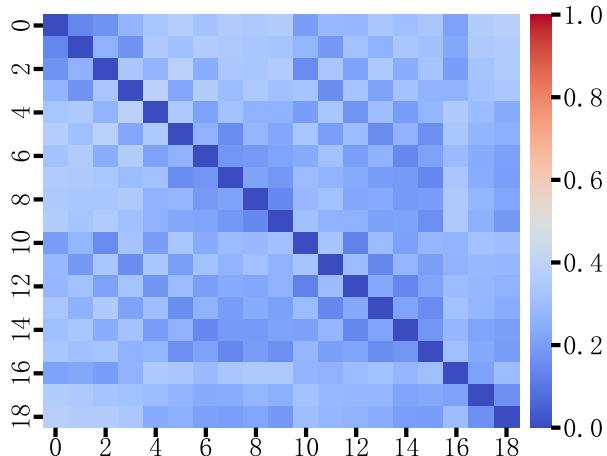
### 3.3.2 跨域重构

跨域重构方法其核心价值在于充分挖掘时间序列数据在时域与频域的双重表征优势。虽然傅里叶变换及其逆变换在数学层面已经建立了时频两域间的转换桥梁，但神经网络对两域数据的特征提取机制存在本质差异：在时域分析维度，神经网络能够有效捕捉信号幅值随时间演变的动态特征，这种时序建模能力对于癫痫诊断中关键神经电活动（如棘波、尖波等瞬态异常波形）的识别至关重要；而在频域分析维度，神经网络则擅长解析信号中各频率成分的能量分布特征，这种谱分析能力特别适用于检测与特定神经状态密切相关的节律模式（如 $\alpha$ 波、 $\beta$ 波等特征性节律）。当前主流的深度学习方法往往局限于单一域内的信号重构任务<sup>[57]</sup>，这种处理范式难以充分发掘跨域信息协同的潜在价值。为突破这一技术瓶颈，本研究创新性地提出了一种基于双向跨域映射的自监督学习任务框架。该框架的核心设计思想是通过构建时频双域间的受损-完整信号重建机制，即利用时域受损信号重构完整频域信号，同时利用频域受损信号重构完整时域信号，迫使神经网络必须建立跨域的特征关联模型。

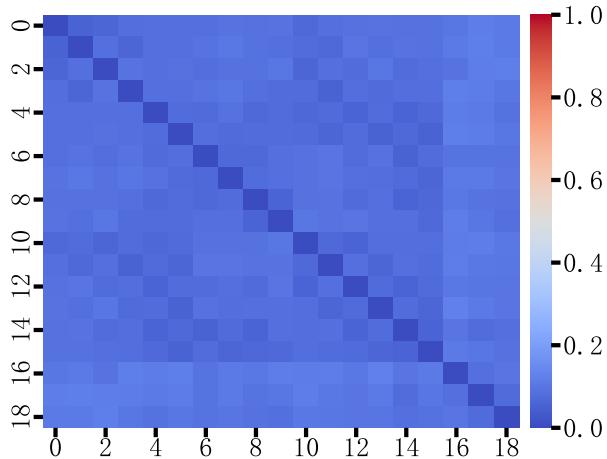
与直接在原始信号空间进行重构的传统方法相比，本方案创新性地在压缩的嵌入空间执行跨时频域双向映射。在特征表示层面，嵌入空间的跨域映射能够有效避免误差累积问题。具体而言，当在原始信号空间进行重构时，模型需要对每个时间步或频点进行精确重建，这种逐点重建要求会导致无关紧要的细节特征对整体损失函数产生过度影响。而在压缩的嵌入空间中进行特征映射，则允许模型专注于具有判别性的高层特征，忽略无关紧要的细节波动。

### 3.3.3 时频通道的鲁棒性差异

带通滤波偏差对脑电信号时域与频域通道间关系的影响存在显著差异。为验证这一现象，本研究设计了如下实验方案：首先基于余弦相似度度量方法构建原始信号各通道间的关系矩阵，随后引入特定频带的带通滤波偏差干扰，定量分析通道关系矩阵的偏移程度。实验结果的详细可视化呈现参见图3-2。实验结果表明带通滤波偏差干扰下频域通道关系展现出更强的稳定性。图中左半部分展示时域通道相似性偏移，右半部分则呈现频域通道相似性偏移。研究发现，频域通道关联性在噪声干扰下保持相对稳定，而时域通道关系则表现出明显的扰动。这一现象揭示了频域表征对带通滤波偏差具有天然的鲁棒性特征。基于此发现，本研究提出创新的跨域校准方法，通过建立频域稳定通道关系与时域动态通道关系之间的映射机制，实现对时域通道关系的有效校正。



(a) 时域通道关系偏移



(b) 频域通道关系偏移

图 3-2 噪声引起的时频关系偏移对比图

### 3.4 模型设计

如图3-3所示，本研究提出的框架采用时域与频域数据并行处理的预训练策略。框架中的时域处理模块和频域处理模块在结构上保持一致，但使用独立的参数进行优化。整个处理流程可分为四个主要阶段：首先，对输入的原始样本执行特定频带的掩码处理，并通过带阻滤波器提取样本中的背景噪声成分，随后利用逆傅里叶变换将经过频域掩码处理的三种信号转换回时域表示。在第二阶段，模型对多通道样本中的每个通道进行独立处理。每个通道的信号分别输入到时域编码器和频域编码器中，生成对应的通道级嵌入表示。这些嵌入随后用于跨域重构任务，通过域间映射促使模型学习更具判别性的特征表示。第三阶段将样本各通道嵌入

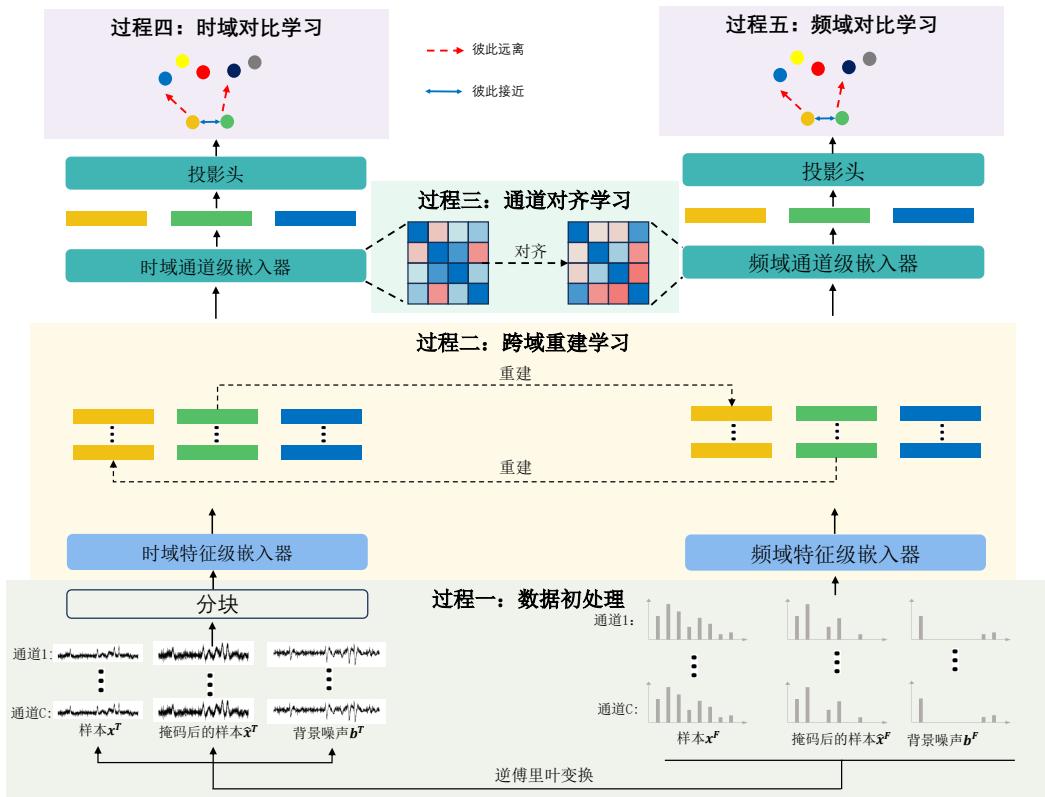


图 3-3 鲁棒预训练框架

合并成样本嵌入，然后将样本级别的嵌入映射到对比学习空间。在此空间中，模型同时在时域和频域两个维度上分别执行带有额外噪声负样本的对比学习。整个框架的设计充分考虑了时域和频域特征的互补性，通过多阶段、多层次的信号处理与特征学习，最终实现对 EEG 信号的有效表征。各模块的实现细节将在后续章节中详细说明。

### 3.4.1 数据预处理

首先，对于给定样本的频域数据  $x_r^F$ ，模型首先采用零值对部分频带进行掩码以生成增强频域数据  $\hat{x}_r^F$ 。随后利用带阻滤波器生成背景数据  $b_r^F$ ，本研究设定 1Hz 以下或 30Hz 以上的频率成分为数据的背景噪声。最终通过逆傅里叶变换将所有频域数据转换至时域，获得时域样本数据  $x_r^T$ ，时域增强数据  $\hat{x}_r^T$ ，时域噪声数据  $b_r^T$ 。与<sup>[58]</sup> 中直接在时域进行数据增强不同，本方法创新性地在频域执行增强操作，并将增强后的频域数据转换至时域。

其次，模型在时域和频域进行分块操作：对于原始时域数据  $x_r^T \in R^{C \times T}$ ，其中  $C$  代表通道数目， $T$  表示时间点数量，本模型采用分块处理方式，通过设定固定尺寸  $P$  将数据重塑为  $x^T \in R^{C \times L \times P}$  的三维结构，此处  $L = T/P$  表示分块后的块数。在

频域数据表示方面， $x^F \in \mathbb{R}^{C \times F \times 2}$  的维度结构中， $F$  对应于频率成分的数量，而最后一个维度 2 则用于存储振幅与相位信息。到此模型得到了输入时域输入数据  $x^T$ ,  $\hat{x}^T$ ,  $b^T$  和频域输入数据  $x^T$ ,  $\hat{x}^T$ ,  $b^T$ 。

最后，模型会将多通道数据拆分成单通道数据进行处理。因为模型需要融合所有通道的信息，但由于不同通道具有异质性生理特征量，在跨域重建任务中，混合其他通道信息时难以学习单一通道的语义特征，且在各通道信息未独立编码前难以建模通道间关系。因此，在模型初始阶段实施通道分离操作：对每个通道  $x_m \in \mathbb{R}^{L \times P}$  (频域表示为  $x_m^F \in \mathbb{R}^{F \times 2}$ ，其中  $m = 1, \dots, M$ ) 进行独立编码。

### 3.4.2 跨域重构学习

跨域重构学习模块致力于建立时域与频域特征之间的双向映射关系，通过强制模型执行跨域信号重建任务来获取更具判别性的特征表示。具体实现过程可分为三个关键步骤：首先，将经过预处理的单通道数据分别输入时域与频域编码器。具体而言，原始时域信号  $x_m$  及其掩码版本  $\hat{x}_m$  输入时域通道编码器  $\text{Encoder}_T$ ，生成对应的时域嵌入特征  $z_m = \text{Encoder}_T(x_m)$  和  $\hat{z}_m = \text{Encoder}_T(\hat{x}_m)$ ；同时，原始频域信号  $x_m^F$  及其掩码版本  $\hat{x}_m^F$  输入频域通道编码器  $\text{Encoder}_F$ ，产生频域嵌入特征  $z_m^F = \text{Encoder}_F(x_m^F)$  与  $\hat{z}_m^F = \text{Encoder}_F(\hat{x}_m^F)$ 。本研究中设计的通道编码器均采用标准的 Transformer 架构，其核心创新在于将每个数据块视为独立的 Token 进行序列建模，这种处理方式既保留了局部时序特征，又能够捕捉长程依赖关系。

跨域重构任务通过精心设计的双向预测机制实现，包含两个互补的映射过程：其一是从频域到时域的特征转换，由预测器  $P_{ft}$  完成；其二是从时域到频域的特征转换，由预测器  $P_{tf}$  实现。这两个过程分别通过以下损失函数进行优化：

$$\mathcal{L}^r = \sum_{i=1}^{NM} \|P_{ft}(\hat{z}_i^F) - z_i^T\| \quad (3-1)$$

式中  $P_{ft}$ ——将频域特征映射到时域特征的预测器；  
 $\mathcal{L}^r$ ——时域特征重建损失。

$$\mathcal{L}_F^r = \sum_{i=1}^{NM} \|P_{tf}(z_i) - \hat{z}_i^F\| \quad (3-2)$$

式中  $P_{tf}$ ——将时域特征映射到频域特征的预测器；  
 $\mathcal{L}_F^r$ ——频域特征域重建损失。

从功能角度看， $P_{ft}$  预测器负责将受干扰的频域特征映射回纯净的时域表示，而  $P_{tf}$  预测器则执行逆向的映射过程。这种双向重构机制迫使模型必须深入理解时频域

特征的内在关联，从而学习到更具鲁棒性的联合表征。在实现细节上，两个预测器均采用双层多层感知机结构，这种设计在保证模型表达能力的同时，有效控制了参数规模。特别需要指出的是，通过这种跨域重构约束，模型不仅能够保留各域特有的特征提取优势，更重要的是建立了时频特征之间的系统性关联，为后续的通道对齐和对比学习奠定了坚实基础。

### 3.4.3 时频通道对齐

时频通道对齐旨在解决频带掩蔽条件下时域与频域通道关系不一致的关键问题。在获得各通道的嵌入表征后，本研究旨在对齐频带掩蔽条件下时域和频域之间的通道关系偏移。在完成各通道的嵌入表征提取后，时域中的所有通道嵌入  $z_m$ （其中  $m = 1, \dots, M$ ）通过专用的通道编码器进行特征融合，最终生成样本级别的时域嵌入表示  $z \in \mathbb{R}^T$ ；相应地，频域处理流程也采用相同的架构生成对应的频域嵌入表示  $z^F$ 。本方案采用的通道编码器基于标准的 Transformer 架构实现，具体包括时域序列编码器  $\text{SeqEncoder}_T$  和频域序列编码器  $\text{SeqEncoder}_F$  两个独立模块。在这两个编码器中，各通道的嵌入表示被视为 Token，而整个通道序列则作为输入序列进行处理，这种设计既保留了通道间的关联性，又充分利用了 Transformer 架构在序列建模方面的优势。

在特征提取过程中，时域通道编码器内部会计算并输出通道间的注意力权重矩阵  $A_T$ ，该矩阵定量刻画了时域中各通道信号之间的动态关联强度；与之对应，频域通道编码器也采用相同的处理流程生成频域通道注意力矩阵  $A_F$ 。这两个关键矩阵的计算过程可通过公式组 (3-3) 进行定义。

$$A_T = \text{TempAttn}([z_1, \dots, z_M], [z_1, \dots, z_M], [z_1, \dots, z_M]) \quad (3-3a)$$

$$A_F = \text{FreqAttn}([z_1^F, \dots, z_M^F], [z_1^F, \dots, z_M^F], [z_1^F, \dots, z_M^F]) \quad (3-3b)$$

式中  $\text{TempAttn}$ —— $\text{SeqEncoder}_T$  中的注意力函数；

$\text{FreqAttn}$ —— $\text{SeqEncoder}_F$  中的注意力函数。

基于上述获得的时频双域注意力矩阵，本研究提出创新的通道关系对齐策略。该策略的核心思想是利用频域通道关系的稳定性来校正时域通道表征，具体通过最小化时域注意力矩阵  $A_T$  与频域注意力矩阵  $A_F$  之间的差异来实现。这一过程通过专门设计的通道对齐损失函数  $\mathcal{L}^a$  进行优化，由公式 (3-4) 所示。

$$\mathcal{L}^a = \sum_{i=1}^N |A_T - \text{Detach}(A_F)| \quad (3-4)$$

式中  $\mathcal{L}^a$ ——通道对齐损失函数；

Detach——梯度截断操作。

需要特别说明的是，在优化过程中本研究采用了梯度截断技术对  $A_F$  进行梯度冻结，这一设计具有双重优势：首先确保了频域注意力矩阵在训练过程中保持相对稳定，避免双向优化导致的振荡问题；其次使得时域通道关系能够单向地向更为稳定的频域通道关系对齐，从而提升模型在噪声干扰下的鲁棒性。

### 3.4.4 附带额外噪声负样本的对比学习

为了增强模型的鲁棒性与避免特征在训练时坍塌，本模型最后分别在时域和频域下进行附带额外噪声负样本的对比学习。本模型利用模型的背景噪声作为额外的负样本，原始 EEG 信号经过预处理后生成的背景噪声信号  $b$  与  $b^F$ ，分别经由时域编码器和频域编码器进行特征提取，再通过通道编码器的进一步处理，最终形成具有判别性的噪声样本嵌入表示  $z_b$  和  $z_b^F$ 。在时域处理分支中，原始样本嵌入  $z^T$ 、掩码样本嵌入  $\hat{z}^T$  以及噪声样本嵌入  $z_b^T$  共同输入时域投影器  $P_T$ ，完成从特征空间到对比学习空间的非线性映射，相应生成时域对比特征  $h^T$ 、 $\hat{h}^T$  与  $h_b^T$ 。频域处理分支采用对称架构，通过频域投影器  $P_F$  的转换，获得频域对比特征  $h^F$ 、 $\hat{h}^F$  及  $h_b^F$ ，其数学表达如公式 (3-5) 所示。

$$h^T = P_T(z^T) \quad (3-5a)$$

$$h^F = P_F(z^F) \quad (3-5b)$$

式中  $z^T$ ——锚点样本的时域特征；

$z^F$ ——增强样本的频域特征；

$h^T$ ——对比学习空间中的其他样本的时域特征；

$h^F$ ——对比学习空间中的其他样本的频域特征；

$P_T$ ——时域投影器；

$P_F$ ——频域投影器。

为确保模型在噪声干扰下仍能保持稳定的特征判别能力，本研究提出附带额外噪声样本的对比优化准则：对于任意样本  $x_i$ ，其嵌入特征应当与经过掩码处理的增强版本  $\hat{x}_i$  在表示空间中保持邻近关系，而与其余样本  $x_j$ （其中  $j \neq i$ ）及其对应的背景噪声  $x_b$  的嵌入特征则需维持足够的距离。基于这一准则，本研究形式化定义了时域背景噪声嵌入集合  $B = \{h_{b,i}^T \mid i = 1, \dots, N\}$  以及频域背景噪声嵌入集合  $B^F = \{h_{b,i}^F \mid i = 1, \dots, N\}$ 。在时域对比学习的实现过程中，正样本集合  $S^{T+}$  由原始样本与其掩码版本的特征对构成，而负样本集合  $S^{T-}$  则包含跨样本特征对及所有噪声样本特征，其数学定义如公式 (3-6) 所示。

$$S^{T+} = \{(h_i^T, \hat{h}_i^T) \mid i = 1, \dots, N\} \quad (3-6a)$$

$$S^{T-} = \{(h_i^T, h_j^T) \mid i = 1, \dots, N \wedge i \neq j\} \cup B \quad (3-6b)$$

式中  $h_i^T$ ——对比学习空间中的锚点样本的时域特征；

$\hat{h}_i^T$ ——对比学习空间中的增强样本的时域特征；

$h_j^T$ ——对比学习空间中的其他样本的时域特征；

$B$ ——时域噪声样本集合；

$S^{T+}$ ——时域正样本集合；

$S^{T-}$ ——时域负样本集合。

与在时域的对比学习类似，频域中的对比学习的正负样本集的数学表达如下所示：

$$\text{Positive pairs set: } S^{F+} = \{(h_i^F, \hat{h}_i^F) \mid i = 1, \dots, N\} \quad (3-7a)$$

$$\text{Negative pairs set: } S^{F-} = \{(h_i^F, h_j^F) \mid i = 1, \dots, N \wedge i \neq j\} \cup B^F \quad (3-7b)$$

式中  $h_i^F$ ——对比学习空间中的锚点样本的频域特征；

$\hat{h}_i^F$ ——对比学习空间中的增强样本的频域特征；

$h_j^F$ ——对比学习空间中的其他样本的频域特征；

$B^F$ ——频域噪声样本集合；

$S^{F+}$ ——频域正样本集合；

$S^{F-}$ ——频域负样本集合。

频域对比学习采用与时域对称的设计理念，其正负样本集合的构建遵循相同的原则。正样本集合  $S^{F+}$  包含原始频域特征与其掩码版本的配对，负样本集合  $S^{F-}$  则由不同样本间的特征组合及全部频域噪声样本组成，具体形式化表达见公式(3-7)。在损失函数设计方面，本研究采用经过标准化处理的温度缩放交叉熵损失，该损失函数通过引入可调节的温度参数  $\tau$ ，能够精确控制正负样本在嵌入空间中的分布密度，这一技术手段在自监督学习领域已被广泛验证其有效性。时域对比损失函数  $\mathcal{L}_T^C$  的具体数学形式如公式(3-8)所示，其核心优化目标在于最大化正样本对  $(h_i^T, \hat{h}_i^T)$  的相似度测度，同时系统性地抑制与负样本集  $\{h_j^T\}_{j \neq i} \cup \{h_{b,k}^T\}$  的相似性关联。

频域对比损失函数  $\mathcal{L}_F^C$  的设计与时域保持高度对称，其具体表达式见公式(3-9)。这种双域对称的对比学习机制不仅确保了时频特征在表示空间中的一致性，更重要的是通过显式地引入噪声负样本，强制模型建立对噪声模式的判别能力，从而显著提升在真实噪声环境下的鲁棒性。从理论分析的角度来看，这种设计使得模型必须同时学习两类关键特征：其一是原始信号的本质特征表示，其二是噪声

干扰的模式特征，二者的协同优化最终促成具有强判别性和高鲁棒性的特征表示空间。

$$\mathcal{L}_T^C = - \sum_{i=1}^N \log \frac{\exp(\text{sim}(h_i^T, \hat{h}_i^T)/\tau)}{\sum_{h_b \in B} \exp(\alpha \cdot \text{sim}(h_i^T, h_b^T)/\tau) + \sum_{i \neq j} \exp \text{sim}(h_i^T, h_j^T)/\tau} \quad (3-8)$$

式中  $h_i^T$ ——对比学习空间中的锚点样本的时域特征；  
 $\hat{h}_i^T$ ——对比学习空间中的增强样本的时域特征；  
 $h_j^T$ ——对比学习空间中的其他样本的时域特征；  
 $h_b^T$ ——对比学习空间中的噪声样本的时域特征；  
 $\tau$ ——为温度调节参数；  
 $\text{sim}(\cdot, \cdot)$ ——向量间的余弦相似度函数。

$$\mathcal{L}_F^C = - \sum_{i=1}^N \log \frac{\exp(\text{sim}(h_i^F, \hat{h}_i^F)/\tau)}{\sum_{h_b \in B^F} \exp(\alpha \cdot \text{sim}(h_i^F, h_b^F)/\tau) + \sum_{i \neq j} \exp \text{sim}(h_i^F, h_j^F)/\tau} \quad (3-9)$$

式中  $h_i^F$ ——对比学习空间中的锚点样本的频域特征；  
 $\hat{h}_i^F$ ——对比学习空间中的增强样本的频域特征；  
 $h_j^F$ ——对比学习空间中的其他样本的频域特征；  
 $h_b^F$ ——对比学习空间中的噪声样本的频域特征；  
 $\tau$ ——为温度调节参数；  
 $\text{sim}(\cdot, \cdot)$ ——向量间的余弦相似度函数。

### 3.4.5 模型训练设计与优化目标

预训练阶段的总体损失函数包含两个组成部分。其一，跨域重构损失  $\mathcal{L}^r$  与  $\mathcal{L}_F^r$  构成模型训练的基础目标。该损失函数促使模型在时域和频域之间建立有效的特征映射关系，通过强制进行域间信号重构，不仅增强了模型抵抗频域滤波攻击的能力，还实现了时频双域特征的有机融合。这种设计使得模型能够充分利用时域信号的局部时序特征和频域信号的全局谱特征。其二，对比损失  $\mathcal{L}_T^C$  和  $\mathcal{L}_F^C$  引入了噪声判别机制，通过将背景噪声作为额外负样本纳入对比学习过程。这种设计具有双重优势：一是有效避免了模型坍塌到平凡解，另一方面提升了模型对噪声模式的判别能力。两个对比损失项分别在时域和频域空间实施，确保模型在两个特征空间都能保持判别性。综合而言，预训练损失函数定义为：

$$\mathcal{L} = (\mathcal{L}^r + \mathcal{L}_F^r) + \mathcal{L}^a + (\mathcal{L}_T^C + \mathcal{L}_F^C) \quad (3-10)$$

式中  $\mathcal{L}^r$ ——时域重构损失函数；

$\mathcal{L}_F^r$ ——频域重构损失函数；

$\mathcal{L}^a$ ——通道对齐损失函数；

$\mathcal{L}$ ——预训练损失函数。

在预训练过程中，Encoder<sub>T</sub>（时域编码器）、Encoder<sub>F</sub>（频域编码器）、SeqEncoder<sub>T</sub>（时域序列编码器）和 SeqEncoder<sub>F</sub>（频域序列编码器）均参与参数学习。当完成预训练并对新样本  $x$  进行处理时，模型将 Encoder<sub>F</sub> 的输出与 SeqEncoder<sub>T</sub> 的输出进行拼接，形成最终的联合嵌入表示  $z = [z^T, z^F]$ 。

## 3.5 实验与分析

本节将在两类真实 EEG 数据集（癫痫发作检测与脑电异常分类任务）上，针对三种实际噪声条件系统评估模型的性能表现。

### 3.5.1 数据集设置

本实验采用天普大学医院脑电图数据集 (TUH EEG)<sup>[7]</sup> 中的两个重要子集进行模型验证与性能评估。作为目前公开规模最大的癫痫脑电数据集，天普大学医院癫痫脑电数据集 (TUSZ)<sup>[7]</sup> 是天普大学医院脑电图数据集的重要子集，该数据集收录的脑电记录总时长达到 435.55 小时，其中所有癫痫发作事件均经过专业标注，精确标记了每次发作的起始与终止时间，一共包含 1832 条临床脑电图记录，其中标注有 324 次癫痫发作事件，涵盖 8 种不同的癫痫发作类型。为简化分类任务复杂度并提高临床实用性，本研究将所有癫痫发作类型合并处理，构建二分类任务框架。其二，天普大学医院异常脑电语料库 (TUAB) v3.0.0<sup>[7]</sup> 作为天普大学医院脑电图数据集的另一个重要子集，专门用于临床脑电信号的正常与异常模式分类研究。该数据集规模庞大，共收录了来自 2383 名受试者的 2993 次检测会话，包含 2993 份脑电记录文件，数据总量达到 58.6GB。TUAB 提供了临床诊断标注为正常或异常的脑电图记录，据此构建第二个二分类任务。为评估模型鲁棒性，本实验将表 3-1 所述三种噪声（含不同参数配置）分别注入两个数据集的形成噪声测试集。

表 3-2 TUSZ 数据集的类别分布表

类别	训练集	验证集	测试集
正常	220762	68826	25914
癫痫发作	13882	4561	2183

参照<sup>[59]</sup>的研究方法，本研究对 TUSZ 数据集进行预处理：首先将所有脑电信

号采样频率统一降采样至 200Hz。随后通过以下步骤获取时频域特征片段及对应标签：采用滑动窗口法提取脑电片段，设置窗口长度为 12 秒，步长与窗口长度一致。当窗口超出原始信号长度时，舍弃末端不完整片段。针对癫痫发作标注，若某片段时间范围内癫痫发作持续时间超过该片段总时长的 1/5（即 2.4 秒），则标记为阳性样本（标签 1），否则标记为阴性样本（标签 0）。TUSZ 与 TUAB 已经划分好了训练集、验证集和测试集，经过裁剪与筛选后，各数据集的样本大小与类别分布由表 3-3 与表 3-2 所示。

表 3-3 TUAB 数据集的类别分布表

类别	训练集	验证集	测试集
正常	122525	31535	14190
异常	125703	30477	16572

在特征标准化阶段，本研究采用基于训练集的 z-score 归一化方法，即利用训练集数据计算得到的均值与标准差对所有脑电片段进行标准化处理。考虑到 TUAB 与 TUSZ 数据集均源自 TUEG 数据库，本研究对 TUAB 数据集采用与 TUSZ 相似的预处理流程，确保数据处理方法的一致性。

### 3.5.2 基准模型与评价标准

为全面评估所提模型的性能优势，本研究设计了系统的对比实验方案，重点考察模型在噪声鲁棒性和跨任务泛化能力两个关键维度上的表现。在基线方法选择方面，本实验选取了当前最具代表性的两类方法进行对比分析。自监督学习领域选择了三种先进方法：TS-TCC<sup>[40]</sup>、TFC<sup>[58]</sup> 和 SimMTM<sup>[57]</sup>，其中 TS-TCC 作为一种自监督模型，其核心目标是通过时序对比与上下文对比学习对抗强扰动的鲁棒表征。该模型通过强增强与弱增强两种数据增强方式，为同一锚样本生成两个不同的视图进行对比学习；TFC 模型创新性地将时序对比学习与频域对比学习相结合，并在统一的空间中实现两者的协同优化。这种双域对比机制能够更全面地捕捉时间序列数据的本质特征；SimMTM 采用自监督掩码建模策略，通过利用流形上多个随机掩码的相邻样本来避免破坏关键的时间动态特征。该方法有效解决了传统掩码建模可能导致的时序信息损失问题。监督学习领域则选取了 TimesNet<sup>[60]</sup>、PathTST<sup>[61]</sup> 和 Informer<sup>[62]</sup> 作为对比基准，其中 Informer 是基于 Transformer 架构的长序列时间预测模型，其创新性地采用 ProbSparse 自注意力机制与自注意力蒸馏技术，显著提升了长序列建模的效率与性能；PatchTST 同样是基于 Transformer 的时间序列预测模型，通过子序列级别的分块处理和通道独立建模策略，实现了对时序特征的高效捕捉。这种方法在保持模型轻量化的同时提升了预测精度；TimesNet

作为 Transformer 架构的改进模型，其核心创新在于通过参数高效的 Inception 模块，从具有多周期特性的时间序列数据中提取二维时空变化特征。这种设计使模型能够更全面地捕捉复杂的时序模式。

针对临床脑电数据普遍存在的类别不平衡问题，本研究采用 AUROC 作为评估指标。该指标能够有效避免分类器在数据分布偏斜情况下的评估偏差，尤其适合 TUSZ 数据集呈现的长尾分布特性。在具体实验设置中，除了记录模型在原始测试集上的性能表现外，还重点测量了各方法在不同噪声强度干扰下的表现，以此量化模型的噪声鲁棒性。

### 3.5.3 训练设置

本实验基于 PyTorch 实现，并且使用 NVIDIA RTX A6000 进行训练。实验对模型采用 Adam 优化器进行优化，学习率设定为  $3 \times 10^{-4}$ 。频段掩蔽比例设置为 0.15，通过带阻滤波技术滤除 1Hz 至 30Hz 范围外的频率成分以生成噪声信号，数据分块尺寸为 200。通道编码器采用两层标准 Transformer 编码器结构，序列编码器配置单层 Transformer 编码器。跨域映射模块使用包含 256 个隐藏单元的双层 MLP 实现，对比学习投影器采用单层 MLP 结构，同样配置 256 维隐藏单元，通过降维操作获得紧凑的特征表示。

### 3.5.4 实验结果与分析

表 3-4 TUSZ 数据集的实验结果

EEG Shifts	TS-TCC	TFC	SIM-MTM	Informer	Patch-TST	Times-Net	本模型
No shift (baseline)	0.6021	0.8398	0.8299	0.6438	0.7379	0.7073	<b>0.8707</b>
$n_{bn}(\sigma = 0.1)$	0.6021	0.8390	0.8299	0.6438	0.7380	0.7073	<b>0.8708</b>
$n_{bn}(\sigma = 0.2)$	0.6021	0.8399	0.8299	0.6440	0.7379	0.7074	<b>0.8705</b>
$n_{in}(\sigma = 0.1)$	0.6021	0.8398	0.8299	0.6415	0.7380	0.7076	<b>0.8707</b>
$n_{in}(\sigma = 0.2)$	0.6021	0.8398	0.8299	0.6424	0.7380	0.7078	<b>0.8707</b>
$n_{bp}(f_L = 0.5, f_H = 30)$	0.6361	0.8067	0.8199	0.6181	0.6906	0.7800	<b>0.8604</b>
$n_{bp}(f_L = 0.5, f_H = 25)$	0.6337	0.8060	0.8165	0.6241	0.6889	0.7795	<b>0.8617</b>
$n_{bp}(f_L = 1, f_H = 30)$	0.6621	0.8088	0.7990	0.6432	0.6556	0.8295	<b>0.8600</b>
$n_{bp}(f_L = 1, f_H = 25)$	0.6591	0.8097	0.7967	0.6438	0.6539	0.8291	<b>0.8606</b>

通过系统性的实验评估，本研究提出的模型展现出卓越的性能优势，具体表现可从三个维度进行阐述。在分类性能方面，如表 3-4 和表 3-5 所示，ROTM 模型在 TUAB 和 TUSZ 两个数据集上均实现了最优的分类准确率。值得注意的是，TUAB 数据集虽然具有相对平衡的标签分布，但本模型仍超越表现最佳的监督学习方法

表 3-5 TUAB 数据集的实验结果

EEG Shifts	TS-TCC	TFC	SIM-MTM	Informer	Patch-TST	Times-Net	本模型
No shift(baseline)	0.7349	0.8278	0.7837	0.8184	0.8288	0.8164	<b>0.8392</b>
$n_{bn}(\sigma = 0.1)$	0.7350	0.8278	0.7837	0.8183	0.8288	0.8164	<b>0.8392</b>
$n_{bn}(\sigma = 0.2)$	0.7350	0.8278	0.7837	0.8184	0.8288	0.8164	<b>0.8392</b>
$n_{in}(\sigma = 0.1)$	0.7350	0.8278	0.7837	0.8184	0.8288	0.8164	<b>0.8392</b>
$n_{in}(\sigma = 0.2)$	0.7350	0.8278	0.7837	0.8184	0.8288	0.8165	<b>0.8392</b>
$n_{bp}(f_L = 0.5, f_H = 30)$	0.7482	0.770	0.7928	0.8193	0.7923	0.8075	<b>0.8391</b>
$n_{bp}(f_L = 1, f_H = 30)$	0.7427	0.7739	0.7801	0.8162	0.7934	0.8073	<b>0.8388</b>
$n_{bp}(f_L = 0.5, f_H = 25)$	0.7427	0.7829	0.7801	0.7937	0.7877	0.7883	<b>0.8384</b>
$n_{bp}(f_L = 1, f_H = 25)$	0.7553	0.7795	0.7745	0.7897	0.7895	0.7882	<b>0.8382</b>

PatchTST。在标签分布严重不平衡的 TUSZ 数据集上，本模型的优势更为显著。这一结果表明本模型在分类性能上超越了现有的方法。就噪声鲁棒性而言，实验结果揭示了本模型对 3 种噪声的卓越的鲁棒性。综合来看高斯噪声与阻抗噪声对所有模型影响微弱，而带通滤波偏差会显著降低所有模型性能。TUSZ 数据集上的实验结果表明，本模型的鲁棒性超越所有自监督学习方法，其性能下降幅度最小。在 TUAB 分类任务中，本模型在带通滤波偏差干扰下性能保持稳定，而其他模型则出现显著性能衰退。

综合实验结果表明本模型在平衡与不平衡 EEG 分类任务中均优于监督式与自监督学习方法。该模型展现出对多种噪声类型的强鲁棒性，特别是对其他模型影响严重的带通滤波偏差。本模型能有效处理真实场景中 EEG 信号的变异特性，为实际应用中的 EEG 分析提供了可靠且稳健的解决方案。

### 3.5.5 消融实验

表 3-6 TUAB 数据集上的消融实验结果

EEG Shifts	w/o C	w/o A	w/o N	本模型
No shift(baseline)	0.8324	0.8227	0.8368	<b>0.8392</b>
$n_{bn}(\sigma = 0.1)$	0.8324	0.8227	0.8368	<b>0.8392</b>
$n_{bn}(\sigma = 0.2)$	0.8324	0.8227	0.8368	<b>0.8392</b>
$n_{in}(\sigma = 0.1)$	0.8324	0.8227	0.8368	<b>0.8392</b>
$n_{in}(\sigma = 0.2)$	0.8324	0.8227	0.8368	<b>0.8392</b>
$n_{bp}(f_L = 0.5, f_H = 30)$	0.8309	0.8225	0.8338	<b>0.8391</b>
$n_{bp}(f_L = 1, f_H = 30)$	0.8302	0.8223	0.8331	<b>0.8388</b>
$n_{bp}(f_L = 0.5, f_H = 25)$	0.8280	0.8214	0.8304	<b>0.8384</b>
$n_{bp}(f_L = 1, f_H = 25)$	0.8272	0.8212	0.8296	<b>0.8382</b>

为深入理解模型各组件的作用机制，本研究设计了系统的消融实验，重点考

察损失函数不同组成部分对模型性能的影响。通过控制变量法，在保持其他条件不变的情况下，依次移除损失函数的特定组件进行对比分析。实验设置包含三种变体模型：移除跨域预测损失的模型记为 w/o C，移除附加噪声样本的模型记为 w/o N，移除通道对齐模块的模型记为 w/o A。完整模型与各变体模型的对比实验结果见表 3-6，结果表明，所有损失组件都对最终性能具有不可替代的贡献。

为进一步验证引入附加噪声样本的有效性，本研究额外设置了针对附加噪声样本的消融实验。本实验从数据和模型两个维度进行了控制：数据维度上，本实验准备了 TUSZ 的测试集和由带通滤波偏差构成的噪声数据集；模型维度上，本实验准备了使用额外噪声样本的原始完整模型与剔除该功能的对照模型 w/o N。实验过程中，本实验将原始数据和噪声数据分别输入上述两种模型架构，通过分别提取对比学习空间中原始数据嵌入和噪声数据嵌入，系统对比原始数据嵌入和噪声数据嵌入之间的相似度关系。如图3-4所示，红色分布表示原始数据嵌入与常规负样本之间的相似度分布，蓝色分布则反映原始数据嵌入与噪声数据嵌入之间的相似度分布。通过对比可以发现，使用了附加噪声样本的原模型能对噪声模式和原始信号模式进行区分，而不适用附加噪声样本的对照模型虽然能正确区分正样本和负样本，但是对噪声模式不能有效区分。这个实验表明了附加噪声样本的对比学习能让模型对噪声模型进行更好的区分，以此增强模型的鲁棒性。

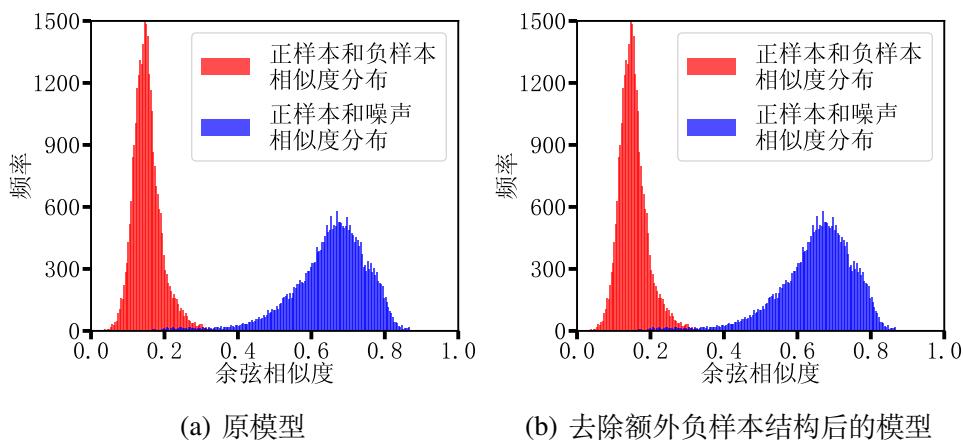


图 3-4 消融实验中负样本和噪声分别与正样本相似度分布对比图

为进一步验证通道对齐模块是否加强了模型的通道关系鲁棒性，本研究设针对通道对齐设计了额外的消融实验，从数据和模型两个维度进行了控制：数据维度上，本实验准备了 TUSZ 的测试集和添加了带通滤波偏差的偏移数据集；模型维度上，本实验准备了集成通道对齐功能的原始完整模型与剔除该功能的对照模型 w/o A。实验过程中，本实验将原始数据和偏移数据分别输入上述两种模型架

构，通过分别提取时域特征级嵌入器和频域特征级嵌入器处理后的各个通道特征表征，系统考察偏移数据对通道间关系结构的平均干扰影响程度。结果如图3-5所示。图3-5 (a)展示了在缺乏通道对齐机制情况下，模型的时域通道关系对数据偏移表现出的显著敏感性；与之形成鲜明对比的是图3-5 (c)所示的集成通道对齐后的模型表现，数据显示通道对齐模块能够有效抑制时域通道关系受偏移干扰的波动幅度，从而提升模型在时域层面的通道关系稳定性。在频域分析方面，图3-5 (b)呈现了不具备通道对齐功能的模型在频域通道关系上表现出的偏移敏感性；而对比图3-5 (d)的结果可以发现，尽管在模型训练过程中我们对通道对齐模块施加了频域梯度锁定措施，该模块仍然能够间接增强频域通道关系的抗干扰能力。综合实验证据表明，通道对齐模块作为本研究框架的核心组成部分，能够增强模型对数据偏移的鲁棒性。该模块不仅在时域特征空间中建立了更为稳健的通道间关联结构，而且能够协同增强频域特征空间的通道关系稳定性，从而在时频双域层面共同提升模型的整体抗干扰能力。

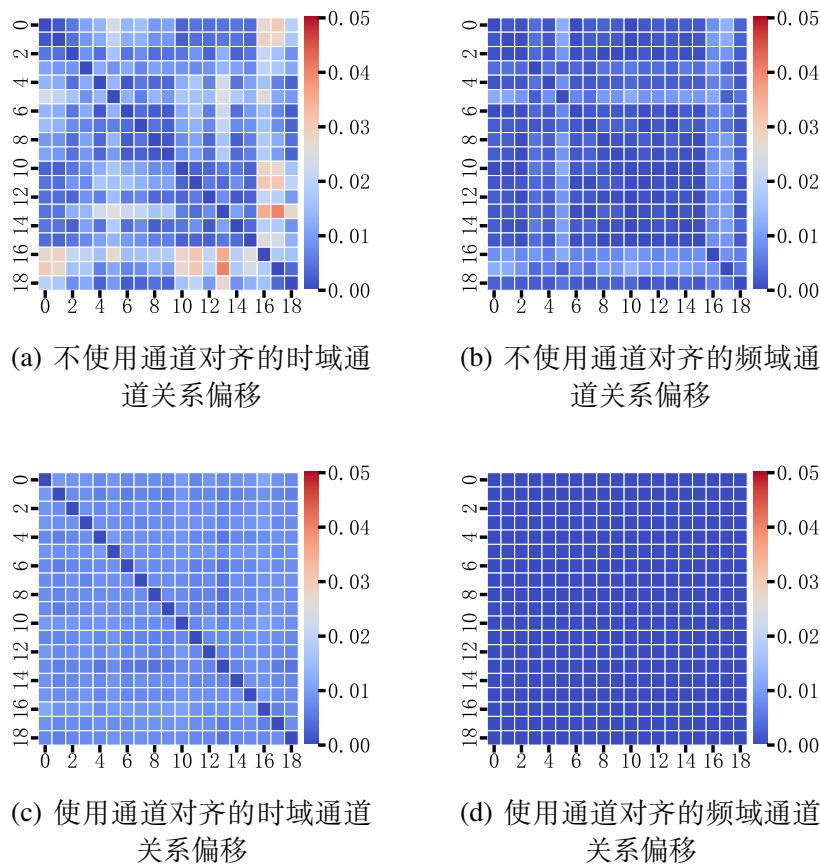


图 3-5 通道对齐前后的时域和频域通道关系偏移对比图

### 3.5.6 预训练有效性验证

为探究预训练在下游脑电图分类任务的有效性，本研究设计了如下实验：模型在 TUAB 上进行预训练后，分别将 5000 条正常数据和 5000 条异常数据输入进模型，提取最终的嵌入向量用 UMAP 进行降维可视化。实验结果如图3-6所示，红色圆圈代表异常特征，蓝色三角代表正常特征。图中显示经过预训练后，模型的嵌入器已经能对下游任务中不同类别的表征进行差别嵌入，预训练能够增强模型对下游脑电图分类任务的表现。

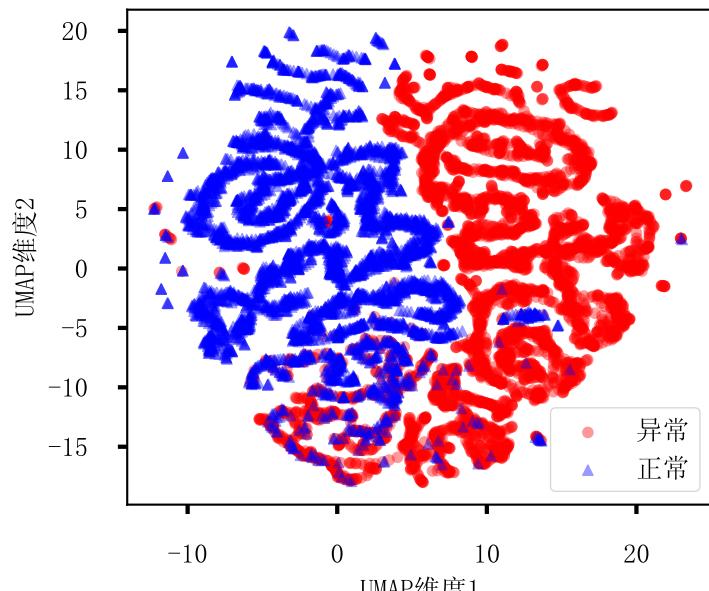


图 3-6 预训练后数据嵌入表征图

## 3.6 本章小结

本章系统阐述了面向脑电时序建模的鲁棒预训练框架。针对医疗场景下 EEG 分类中噪声敏感、标注依赖性强的问题，提出了的鲁棒预训练框架创新方案。理论上，明确了研究问题的形式化定义，分析了 EEG 信号中典型噪声特性及其影响。方法上，设计了联合跨域重构学习、时频通道对齐和附带额外噪声负样本的对比学习策略的预训练框架。实验在 TUH EEG 数据集上验证了方法在癫痫检测与异常分类任务中的优越性，尤其在强噪声环境下展现出显著鲁棒性，消融研究进一步证实各组件贡献。本章首次将噪声显式建模引入时间序列对比学习，建立了时频特征交互方法与通道校准技术，后续将探索多元信息融合应用。

## 第4章 基于残差混合的对比学习鲁棒增强方法

### 4.1 引言

当前基于对比学习的 EEG 分析方法面临多样性不足的数据增强策略和语义噪声建模能力有限两大挑战，导致模型在实际应用中性能退化。为此，本研究提出基于残差混合的对比学习增强方法，从分析特征偏移分布、设计嵌入空间残差混合机制及验证方法适应性三个方向进行研究。然而，实验结果显示该方法在 TFC<sup>[58]</sup> 和 COMET<sup>[63]</sup> 框架下有改进潜力，但在 TSCC<sup>[40]</sup> 和 SimMTM<sup>[57]</sup> 框架中未达预期，揭示了嵌入空间残差度量、线性混合操作及模型架构敏感性等问题。尽管如此，本研究仍具探索价值，通过实验明确了残差混合策略的适用边界，揭示了高维嵌入空间特征混合的风险，并为后续研究提供了经验参考，对理解对比学习在脑电图分析中的局限性具有启示意义。

### 4.2 研究动机

当前对比学习方法在脑电图信号分析中存在的关键局限主要表现为数据增强策略的多样性不足及其对噪声分布先验假设的过度依赖。传统的增强方法如随机掩码、添加噪声等通常依赖于预设参数，难以有效适应医疗场景中复杂的非平稳噪声分布特性。

为验证上述问题，本研究设计了一系列初步实验，系统分析了噪声对特征表示空间产生偏移的影响模式。实验以监督模型 OS-CNN<sup>[64]</sup> 为基准，在 TUSZ 数据集上完成训练后，随机选取测试集样本，引入不同强度的高斯噪声并观察模型中间层特征的偏移分布特性。实验结果如图4-1所示，数据 1 在标准差为 0.175 时的特征偏移可被数据 2 在标准差为 0.1 时的特征偏移所近似模拟。这一现象表明，不同样本在噪声影响下的特征偏移具有一定的可替代性，由此启发本研究可以通过混合不同样本的特征残差来模拟多样化的语义噪声，而无需显式设计复杂的噪声分布。这一策略不仅能够提高正样本对的丰富性，还能有效缓解对噪声参数先验假设的依赖问题。

从理论视角分析，残差混合方法与对比学习的互信息最大化目标高度契合。对比学习通过对比损失优化正样本对的相似性，同时拉远负样本对的距离，而残差混合通过在嵌入空间内引入语义噪声，有效扩展了正样本对的分布范围，从而潜在地增加了正样本与负样本之间的区分度。这种操作相当于在特征空间中构建了

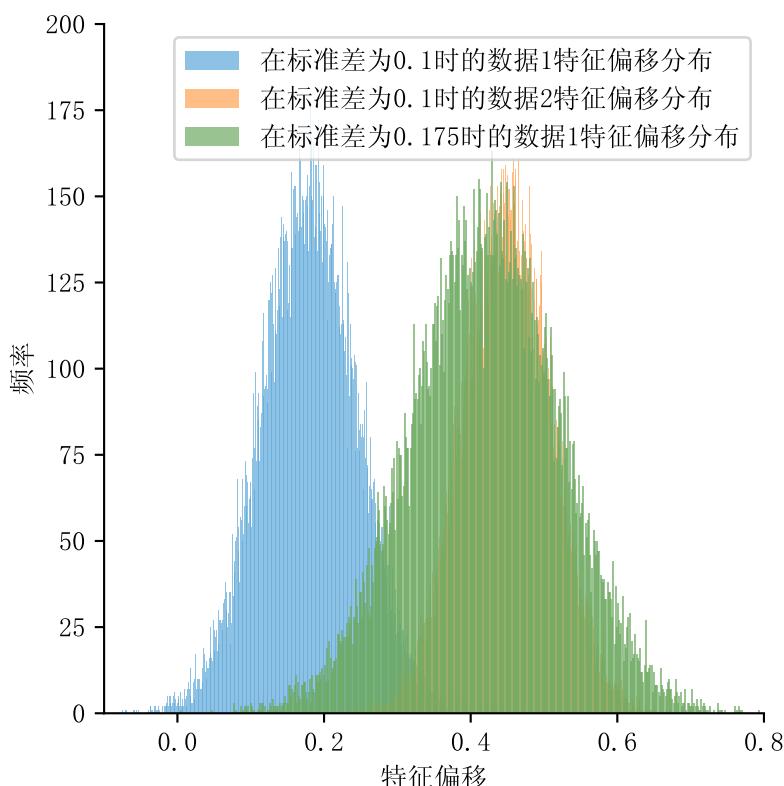


图 4-1 嵌入向量受高斯噪声影响的分布图

“软增强”视图，为模型提供了更为丰富的学习信号。然而，需要注意的是，嵌入空间的高维性和非欧几里得特性可能导致线性混合操作无法准确捕捉复杂的语义关系，后续实验将对此进一步探究。

### 4.3 附带残差混合的对比学习架构

在对比学习框架中，数据增强策略对模型鲁棒性和特征表示的多样性至关重要。然而，现有对比学习方法在应对实际医疗场景中的复杂噪声时，往往受限于数据增强策略的多样性不足，模型对特征偏移的模式识别不充分，导致模型在数据分布偏移环境下性能退化。为此，本研究提出了一种附带残差混合的对比学习架构，旨在通过在嵌入空间内引入语义噪声，增强正样本对的多样性，从而提升模型对噪声干扰的鲁棒性。

如图4-2所示，典型的对比学习框架包括嵌入器(Encoder)和投影器(Projector)两大核心组件。嵌入器负责将输入样本映射到高维特征空间，提取具有语义信息的中间表示；投影器则进一步将特征映射到低维对比空间，用于计算样本间的相似性并进行损失优化。在训练过程中，原始样本通过数据增强生成不同视图，同一样本的不同视图构成正样本对，而不同样本的视图及增强视图构成负样本对。通

过最大化正样本对的相似性并最小化负样本对的相似性，模型学习到具有区分能力的特征表示。在下游任务微调阶段，投影器将被丢弃，仅保留嵌入器结合分类头进行优化。

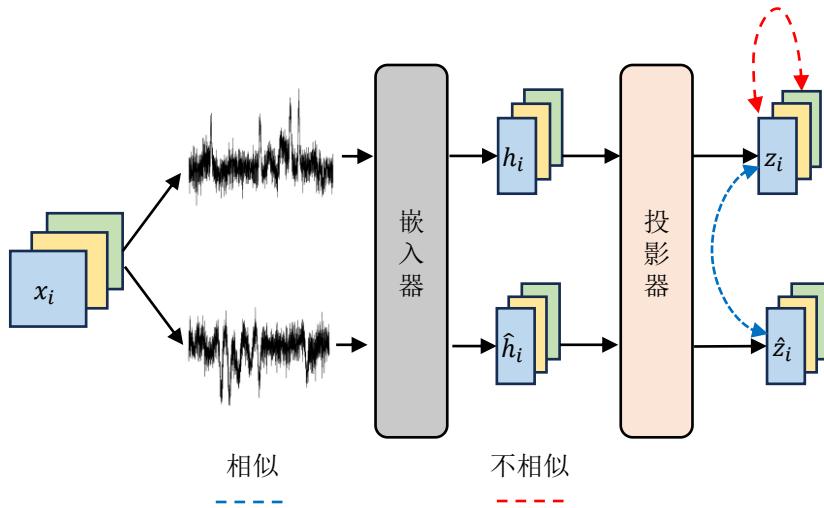


图 4-2 典型的对比学习架构图

本研究提出的附带残差混合的对比学习架构在上述基础框架上进行了创新性改进，核心在于在嵌入器内部引入残差混合操作模块，通过混合原始样本和增强样本之间的特征残差，生成新的增强视图，从而在嵌入空间内引入语义噪声增强样本的多样性。以下将分阶段详细描述整个附带残差混合的对比学习架构的实现细节。

(1) 输入空间的数据增强：在本架构中会有两次数据增强，首先是输入空间的数据增强。给定训练批次中的原始样本  $x_i$ ，通过数据增强（例如随机时间段掩码、幅值扰动）得到其增强样本  $\hat{x}_i$ 。

(2) 中间特征提取：在完成数据预处理和增强后，原始样本  $x$  及其增强版本  $\hat{x}$  被送入特征提取模块，在本架构中，特征提取模块被抽象为一个分层结构，包含多个处理阶段，每一阶段提取不同抽象程度的特征表示。假设特征提取模块包含  $L$  个处理层，则第  $k$  层 ( $1 \leq k \leq L$ ) 的特征由公式 (4-1) 所示。

$$z^{(k)} = f_{\text{enc}}^{(k)}(z^{(k-1)}) \quad (4-1a)$$

$$\hat{z}^{(k)} = f_{\text{enc}}^{(k)}(\hat{z}^{(k-1)}) \quad (4-1b)$$

式中  $f_{\text{enc}}^{(k)}$ ——第  $k$  层特征提取函数；  
 $z^{(k-1)}$ ——原始样本的第  $k-1$  层特征；  
 $\hat{z}^{(k-1)}$ ——增强样本的第  $k-1$  层特征；  
 $z^{(k)}$ ——原始样本的第  $k$  层特征；  
 $\hat{z}^{(k)}$ ——增强样本的第  $k$  层特征。

(3) 残差混合操作：残差混合操作是本架构的核心创新，旨在通过在中间特征空间内引入语义噪声，进一步增强正样本对的多样性，从而提升模型对复杂噪声环境的鲁棒性。与传统在输入空间实施的数据增强不同，残差混合操作在特征提取模块的中间阶段进行，利用特征表示中已提取的语义信息，通过跨样本的残差注入模拟多样化的语义扰动。图4-3展示了残差混合模块在嵌入器内部的具体位置和操作流程。具体而言，残差混合操作在特征提取模块的特定层（记为第  $K$  层）实施。

首先，计算原始样本和增强样本在第  $K$  层的特征残差  $r_i^{(K)}$ ，其具体数学表达如公式 (4-2) 所示。

$$r_i^{(K)} = \hat{z}_i^{(K)} - z_i^{(K)} \quad (4-2)$$

式中  $\hat{z}_i^{(K)}$ ——增强样本的中间层特征；  
 $z_i^{(K)}$ ——原始样本的中间层表示；  
 $r_i^{(K)}$ ——中间层残差。

其中  $r_i^{(K)}$  反映了数据增强操作对特征表示的影响，可视为一种“语义噪声”向量，刻画了增强操作引入的特征偏移特性。残差向量的计算在训练批次内对所有样本对执行，形成一个残差集合  $R = \{r_1^{(K)}, r_2^{(K)}, \dots, r_N^{(K)}\}$ ，其中  $N$  为批次大小。

随后，进行残差混合操作，从残差集合  $R$  中随机选取一个残差向量  $r_j^{(K)}$  ( $j \neq i$ )，并将其与当前增强样本的中间特征  $\hat{z}_i^{(K)}$  进行线性混合，生成新的增强特征表示  $\tilde{z}_i^{(K)}$ ，其具体数学表达如公式 (4-3) 所示。

$$\tilde{z}_i^{(K)} = \lambda \cdot \hat{z}_i^{(K)} + (1 - \lambda) \cdot r_j^{(K)} \quad (4-3)$$

式中  $\hat{z}_i^{(K)}$ ——增强样本的中间层特征；  
 $r_j^{(K)}$ ——其他样本的中间层残差；  
 $\tilde{z}_i^{(K)}$ ——新的增强样本的中间层特征。

其中  $\lambda \in [0, 1]$  为混合系数，控制原始增强特征和外部残差的贡献比例。为避免过度混合导致特征表示失真，本研究通过均匀分布  $\text{Uniform}(\beta, 1)$  采样  $\lambda$  ( $\beta$  通常大于等于 0.75)，确保原始特征占主导地位，仅引入少量外部残差作为语义噪声补充。

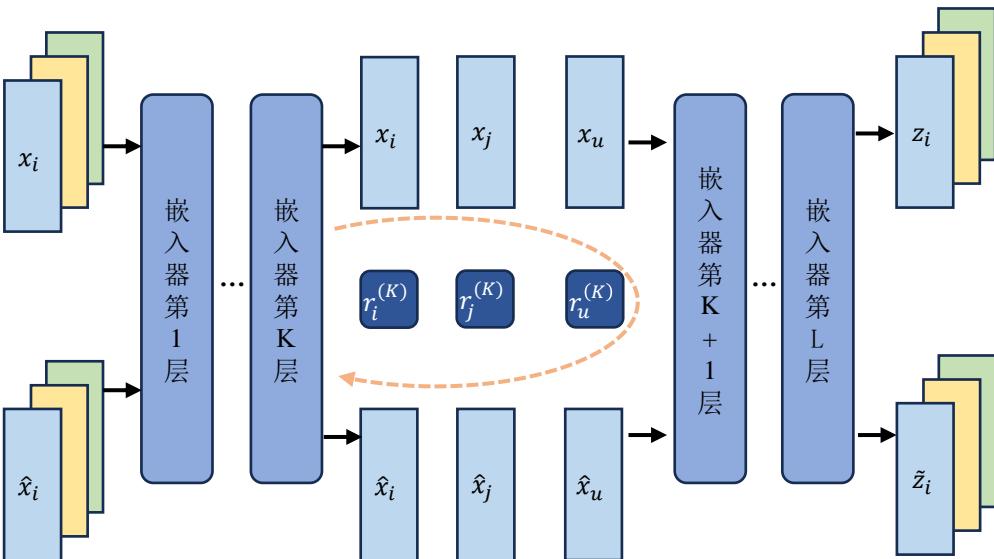


图 4-3 在嵌入器内部进行残差混合的示意图

这种高  $\lambda$  值的设计旨在增强多样性的同时，尽量保留增强样本  $\tilde{z}_i^{(K)}$  的核心语义信息，避免混合操作对特征表示的破坏。

(4) 后续特征处理：完成残差混合后，新的增强特征  $\tilde{z}_i^{(K)}$  替代原增强特征  $\hat{z}_i^{(K)}$ ，被送入特征提取模块的后续层，继续完成特征抽象过程，最终生成完整的特征表示  $\tilde{z}_i$ ，形式化为公式 (4-4)。

$$\tilde{z}_i = f_{\text{enc}}^{\text{post}}(\tilde{z}_i^{(K)}) \quad (4-4)$$

式中  $\tilde{z}_i^{(K)}$ ——经过残差混合后的增强样本的特征；

$f_{\text{enc}}^{\text{post}}$ ——第  $K$  层之后的嵌入器；

$\tilde{z}_i$ ——增强样本的嵌入向量。

通过这种方式，残差混合的影响被融入到最终特征表示中，为后续对比优化提供了更丰富的学习信号。

(5) 对比损失优化：如公式 (4-5) 在特征提取和残差混合完成后，原始样本的特征表示  $z_i$  和经过残差混合的增强特征表示  $\tilde{z}_i$  被进一步映射到对比空间，以便进行相似性度量和对比损失优化。

$$h_i = P(z_i) \quad (4-5a)$$

$$\tilde{h}_i = P(\tilde{z}_i) \quad (4-5b)$$

式中  $P$ ——将嵌入向量映射到对比空间的投影头；

$z_i$ ——原样本的嵌入向量；

$\tilde{z}_i$ ——增强样本的嵌入向量；

$h_i$ ——对比空间中原样本的嵌入向量；

$\tilde{h}_i$ ——对比空间中增强样本的嵌入向量；

对比学习的核心目标是通过损失函数的优化，使得正样本对在对比空间中尽可能靠近，而负样本对尽可能远离，从而学习具有高判别性和鲁棒性的特征表示。在本架构中，正样本对由原始样本的投影特征  $h_i$  和经过残差混合的增强特征  $\tilde{h}_i$  构成，形式化为  $(h_i, \tilde{h}_i)$ ；负样本对则由训练批次中不同样本的特征对构成，形式化为  $(h_i, h_j)$  或  $(h_i, \tilde{h}_j)$ ，其中  $i \neq j$ 。这种正负样本对的构建方式具有高度灵活性，可根据具体任务需求或架构特性进行调整。

损失函数的设计采用基于温度缩放的交叉熵损失（即 InfoNCE 损失），旨在通过优化正样本对的相似性，同时抑制与负样本对的相似性，来提升特征表示的判别能力。形式化地，对比损失函数定义为：

$$\mathcal{L}_{\text{contra}} = - \sum_{i=1}^N \log \frac{\exp(\text{sim}(h_i, \tilde{h}_i)/\tau)}{\sum_{i \neq j} \exp(\alpha \cdot \text{sim}(h_i, h_j)/\tau) + \sum_{i \neq j} \exp \text{sim}(h_i, \tilde{h}_j)/\tau} \quad (4-6)$$

式中  $h_i$ ——对比学习空间中的锚点样本的特征；

$\tilde{h}_i$ ——对比学习空间中的增强样本的特征；

$h_j$ ——对比学习空间中的其他样本的特征；

$\tilde{h}_j$ ——对比学习空间中的增强样本的特征；

$\tau$ ——为温度调节参数；

$\text{sim}(\cdot, \cdot)$ ——向量间的相似度函数。

需要说明的是，本架构中残差混合操作作为一个插件式模块，其设计目标是与现有的多种自监督时间序列学习方法兼容。目前，时间序列自监督学习领域已涌现出众多优秀的架构设计，这些方法在特征提取、投影映射和对比空间构建等方面各具特色。无论底层架构如何设计，残差混合操作均可灵活嵌入到特征提取的中间阶段，通过对中间特征的增强操作，为后续的投影映射提供更丰富的语义信息。这种插件式特性确保了残差混合方法能够适配包括第三章所提出的架构在内的自监督时间序列学习框架，无需对原始架构的核心组件进行重大调整，仅通过在特定阶段引入残差混合模块即可实现鲁棒性增强。

## 4.4 实验与分析

### 4.4.1 数据集设置

本研究采用阿尔茨海默病 (AD)<sup>[65]</sup> 脑电图数据集进行模型评估，该数据集包含 23 名受试者的临床脑电记录，其中 12 名为确诊 AD 患者，11 名为健康对照者。每位受试者平均完成 30.0 次独立试验，每次试验记录包含 1280 个时间点（采样频率 256Hz）的 16 通道脑电信号。

在数据预处理阶段，首先将原始试验数据分割为 9 个重叠样本段，重叠率为 50% 以确保时序连续性；其次对每个样本段应用 z-score 进行通道标准化；最后采用基于受试者的分层划分策略，随机选取 1 名患者和 1 名健康对照者的样本作为验证集，另外 1 名患者和 1 名健康对照者的样本构成测试集，其余样本用于训练集构建。最终训练集有 4096 个样本，验证集和测试集有 512 个样本。

为系统评估模型鲁棒性，实验设置两个测试场景：首先使用原始预处理数据构建标准测试数据集，其次在测试集上添加  $n_{bp}(f_L = 1, f_H = 25)$  的带通滤波偏差构建噪声测试数据集。

### 4.4.2 基准模型与评价标准

本实验择采用了对比学习的预训练模型 TFC<sup>[58]</sup>、COMET<sup>[63]</sup>、TS-TCC<sup>[40]</sup> 和 SimMTM<sup>[57]</sup> 作为基准模型。其中 TFC、TS-TCC、SimMTM 已经在第三章的实验部分进行了详细介绍，这里不再赘述。COMET 采用了病人、采样和样本的多级层次对比学习机制来保证脑电数据的数据集结构一致性。本实验所用模型的嵌入器架构如表 4-1 所示。

本研究将残差混合操作设计为可插拔模块，同时对所有基准模型的嵌入器内部进行修改以支持残差混合操作。为了检验残差混合的有效性，本模块选取 mix-up<sup>[66]</sup> 方法作为参照方法，其核心是在输入空间将不同样本进行混杂，而非本方法在嵌入空间进行残差混杂，因此对每个模型设置以下实验组：(1) 采用原始架构的模型 (2) 在输入空间添加 mix-up 方法的模型 (3) 在嵌入空间进行 mix-up 方法的模型 (4) 在输入空间进行残差混合的模型 (5) 在嵌入空间进行残差混合的模型。采用混合策略的模型的混合系数被设置为 0.3。每个实验配置采用 5 次不同随机初始化的训练结果取平均值，以确保结论的统计显著性。

### 4.4.3 超参数分析

使用了残差混合的 COMET 一共有 4 个参数，分别是预训练批次大小  $B$ ，微

表 4-1 实验中各基线模型的架构表

模型	嵌入器架构	嵌入器层数	采用残差混合操作的层数
TFC	1D ResNets <sup>[67]</sup>	3	0, 1, 2
COMET	TCN <sup>[68]</sup>	10	0, 3, 6
SimMTM	Transformer	3	0, 1, 2
TS-TCC	CNN	3	0, 1, 2

调批次大小  $B$ ，学习率  $lr$  和混合系数  $\lambda$ 。本小结通过在 AD 数据集上调整这些参数的取值来确定使模型性能最优的参数取值，在 AD 数据集上进行实验，并使用 AUROC 评价分类效果。图 4-4 使用箱线图展示了在 AD 数据集上这些参数的不同取值对分析效果的影响。

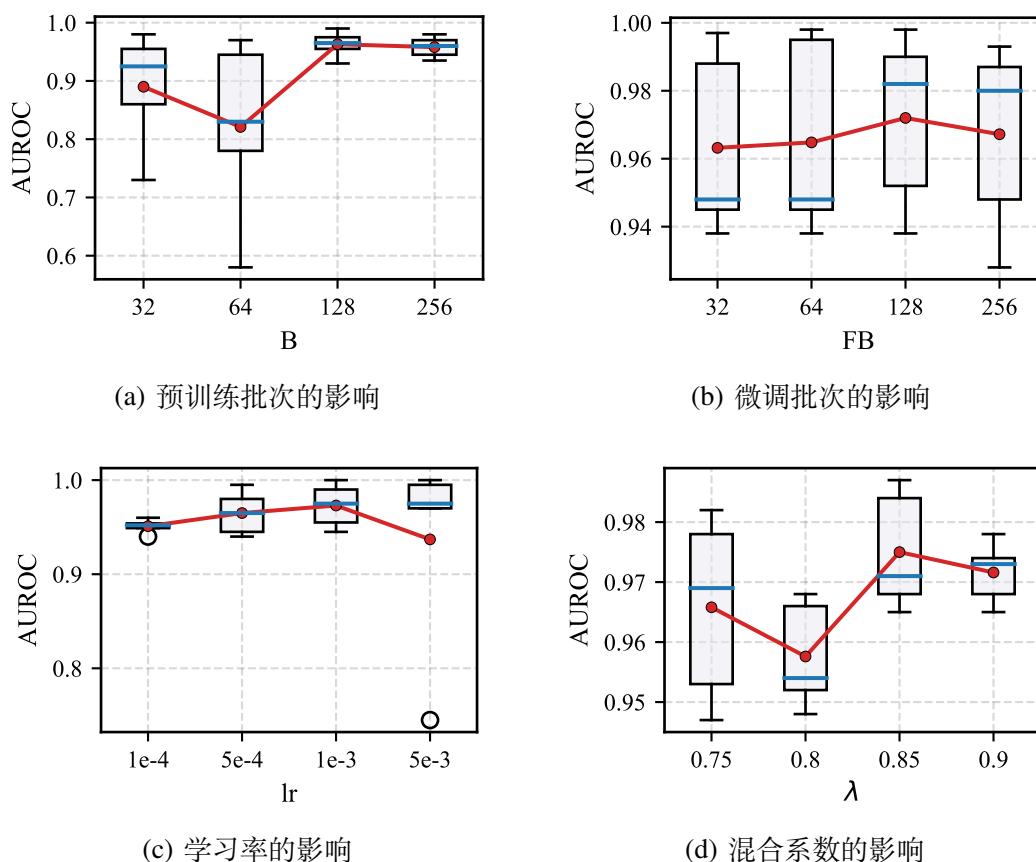


图 4-4 使用了残差混合的 COMET 的超参数分析箱线图

在采用了对比学习的模型的预训练中，预训练批次大小  $B$  起到了重要作用，对比学习核心原理是通过拉近正样本对、推远负样本对来学习有效表示，大多数对比学习方法依赖同批次其他样本作为负样本，预训练批次过小导致可用负样本数量极其有限，降低学习表示的有效性。本节将  $B$  的范围设置为  $\{32, 64, 128, 256\}$  进

行测试。从图4-4 (a)可以看出，随着预训练批次的增大，模型的分类效果变得越来越好且稳定。因此本实验将使用了残差混合的 COMET 的预训练批次设置为 256。

在模型微调的过程中，若微调批次 FB 的数量较少，则会引发较大的随机性，从而使得模型收敛变得困难。随着批次数量的逐渐增多模型的精度会逐步提升，然而，这同时也可能导致模型陷入局部最优的困境。本节将 FB 的范围设置为 {32, 64, 128, 256} 进行测试。从图4-4 (b)可以看出，随着微调批次的变化，虽然模型的分类效果没有出现很大的变化，但是 FB 为 128 时平均分类效果最好，所以本实验将使用了残差混合的 COMET 的微调批次设置为 128。

学习率在模型参数更新中具有至关重要的作用。若学习率设定过低，将导致模型收敛速度变得缓慢；反之，若学习率设定过高，则可能导致参数在最优解附近反复振荡。本节将学习率 lr 的大小设置为 1e-4, 5e-4, 1e-3, 5e-3。从图4-4 (c)可以看出，当学习率逐渐增大，模型学习变得越来越不稳定。因此本实验选择将使用了残差混合的 COMET 的学习率设置为 1e-4。

在残差混合操作中，混合系数  $\lambda$  非常重要，如果  $\lambda$  设置过高，则对样本的增强效果不足，而设置过低则会破坏原样本的语义，造成失真。本节将混合系数的分布设置为 {Uniform(0.75, 1), Uniform(0.8, 1), Uniform(0.85, 1), Uniform(0.9, 1)}。从图4-4 (d)可以看出， $\lambda$  的取值会对模型的分类性能起到产生重大影响。本实验选择 Uniform(0.85, 1) 对  $\lambda$  进行采样。

图4-5展示了使用了残差混合的 TFC 模型的参数分析。与使用了残差混合的 COMET 的参数选取思路相同。本实验将使用了残差混合的 TFC 模型的预训练批次大小设置为 256，将微调批次大小设置为 64，将学习率设置为 3e-4，将混合系数  $\lambda$  按 Uniform(0.9, 1) 进行采样。

TS-TCC 和 SimMTM 超参数分析箱线图详见附录，后文仅提供超参数分析的实验结果分析。图A-1展示了使用了残差混合的 TS-TCC 模型的参数分析。与使用了残差混合的 COMET 的参数选取思路相同。本实验将使用了残差混合的 TS-TCC 模型的预训练批次大小设置为 256，将微调批次大小设置为 512，将学习率设置为 3e-4，将混合系数  $\lambda$  按 Uniform(0.9, 1) 进行采样。

图A-2展示了使用了残差混合的 SimMTM 模型的参数分析。与使用了残差混合的 COMET 的参数选取思路相同。本实验将使用了残差混合的 TS-TCC 模型的预训练批次大小设置为 256，将微调批次大小设置为 128，将学习率设置为 3e-4，将混合系数  $\lambda$  按 Uniform(0.85, 1) 进行采样。

#### 4.4.4 实验结果与分析

实验结果表中的 \_Mixup\_ 和 \_Remix\_ 代表不同的混合策略实现方式。例如，

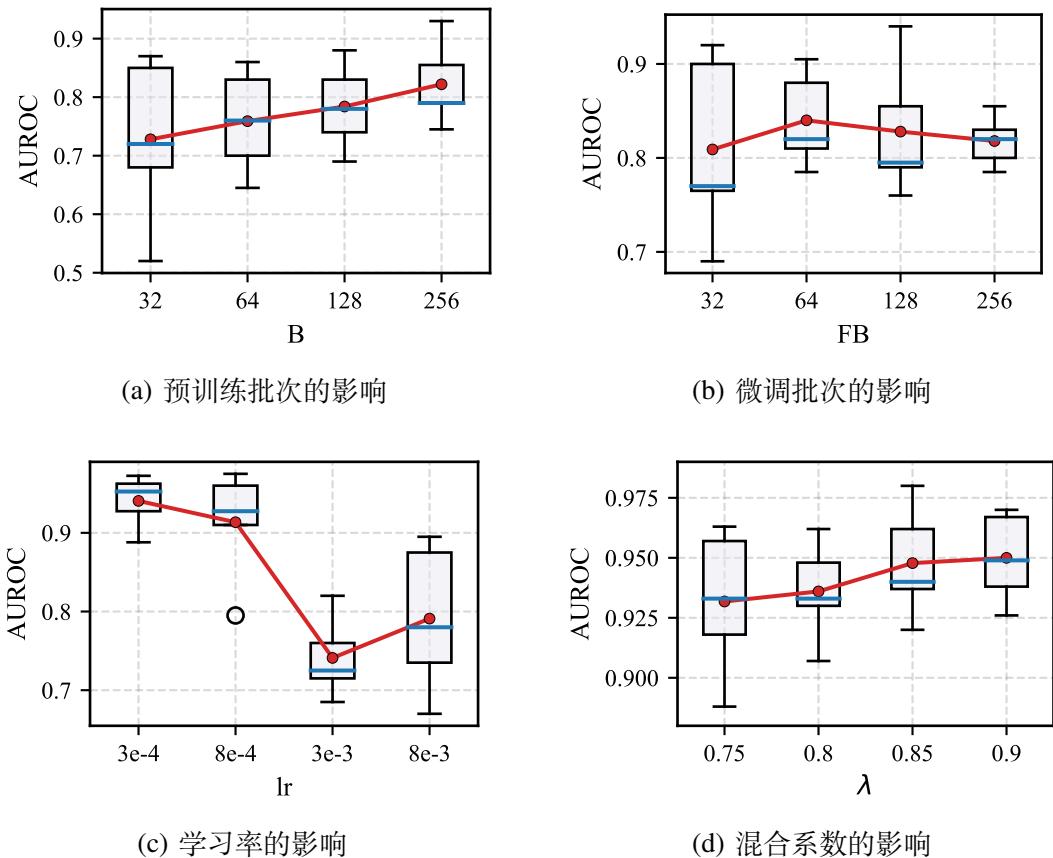


图 4-5 使用了残差混合的 TFC 的超参数分析箱线图

在 TFC 模型中，TFC\_Remix\_1 表示在嵌入器第一层实施残差混合；而在 COMET 模型中，COMET\_Mixup\_1/3 则表示在嵌入器深度  $1/3$  处进行混合操作。这种细粒度的混合位置设计有助于精确分析残差混合在不同网络深度的影响效果。

表 4-2 展示了残差混合插件在 TFC 模型上的分类表现与鲁棒性。基于 TFC 模型残差混合实验的结果表明，残差混合策略在标准数据集上显著提升了模型性能，但在噪声数据集上表现出较大的性能退化。从分类性能上看：虽然混合策略显著提升基础性能，但是残差混合对分类性能提升的效果不如 mix-up，同时在嵌入空间内进行残差混合也不如在输入空间进行残差混合。无论是 Mixup 还是残差混合，TFC 变体在标准数据集上的 AUROC 均大幅超越了原始 TFC 模型。其中 TFC\_Mixup\_0（输入层混合）达到了 0.9961 的最高 AUROC 值。数据显示混合操作在输入层实施（\_0 后缀）比在中间层实施（\_1 后缀）效果更佳，这一现象在 Mixup 和残差混合两种策略中均有体现。从鲁棒性的角度看：所有采用混合策略的模型变体在噪声环境下的 AUROC 值均显著下降，从标准数据集的 0.95 以上降至噪声数据集的约 0.55 左右，这一下降幅度远超原始 TFC 模型。

表 4-2 在 TFC 上使用残差混合的实验结果

模型	AUROC	AUROC_noise
TFC	0.8236	<b>0.7271</b>
TFC_Mixup_0	<b>0.9961</b>	0.5561
TFC_Mixup_1	0.9864	0.5227
TFC_Remix_0	0.9904	0.5820
TFC_Remix_1	0.9557	0.5829

表 4-3 呈现了在 COMET 架构上应用残差混合策略的实验结果。从分类性能上看：在嵌入空间内进行残差混合能显著提高 COMET 模型的分类性能与鲁棒性。在嵌入器深度 2/3 处实施残差混合的 COMET\_Remix\_2/3 配置表现最佳，AUROC 达到 0.9807。从鲁棒性的角度看：残差混合在噪声条件下的表现优异，达到 0.9134。传统的 Mixup 策略反而导致性能下降，这一对比清晰地表明，在 COMET 架构中，残差混合策略不仅能有效提升模型在标准测试集上的性能，更显著增强了模型面对噪声干扰的鲁棒性。

表 4-3 在 COMET 上使用残差混合的实验结果

模型	AUROC	AUROC_noise
COMET	0.9761	0.8909
COMET_Mixup_1/3	0.9473	0.8551
COMET_Mixup_2/3	0.9488	0.8533
COMET_Remix_0	0.9786	0.9103
COMET_Remix_1/3	0.9782	0.9071
COMET_Remix_2/3	<b>0.9807</b>	<b>0.9134</b>

表 4-4 展示了残差混合策略在 TS-TCC 架构上的应用效果。与 COMET 架构不同，TS-TCC 模型在应用残差混合后表现出了复杂的结果模式。从分类性能上看，虽然残差混合方法能略微提高 TS-TCC 的分类效果，但是提升幅度原不如基线 mix-up 方法。残差在不同层的混合的分类结果表明，在内层混合可以提高分类性能。从鲁棒性角度看，所有残差混合均无法避免模型失效，不具有任何鲁棒性上的提升。

表 4-5 呈现了残差混合在 SimMTM 模型上的实验结果。无论从分类性能角度看还是鲁棒性上看，与其他模型不同，残差混合造成了 SimMTM 模型的性能下降和鲁棒性的下降，而且在输入空间进行残差混合的分类结果要比在嵌入空间内进行残差混合的实验结果好。与基线方法 mix-up 相比，mix-up 方法反而提升了模型的分类性能。

以上所有实验结果表明，残差混合增强策略在不同模型架构中呈现出显著的性能差异，整体而言未能展现出普遍性的性能优势与鲁棒增强的特点。残差混合

表 4-4 在 TS-TCC 上使用残差混合的实验结果

模型	AUROC	AUROC_noise
TS-TSCC	0.8601	0.6788
TS-TSCC_Mixup_2	0.8500	<b>0.6812</b>
TS-TSCC_Mixup_1	<b>0.9007</b>	0.6495
TS-TSCC_Remix_0	0.8627	0.6592
TS-TSCC_Remix_1	0.8741	0.6659
TS-TSCC_Remix_2	0.8683	0.6727

表 4-5 在 SimMTM 上使用残差混合的实验结果

模型	AUROC	AUROC_noise
SimMTM	0.8257	0.7417
SimMTM_Mixup_1	<b>0.8208</b>	<b>0.7366</b>
SimMTM_Mixup_2	0.8104	0.7222
SimMTM_Remix_0	0.8120	0.7301
SimMTM_Remix_1	0.8093	0.6933
SimMTM_Remix_2	0.8008	0.7195

插件仅在 COMET 模型上达到了最佳分类效果与最强的鲁棒性。对于在嵌入器内部混合是否有优势的问题，答案也是否定的：就算进行残差混合，在嵌入器内部进行混合效果与在输入空间混合相比也没有优势。对于此否定的研究结果，本研究分析可能的原因包括：首先，在深度嵌入空间中，基于欧式距离的残差度量可能无法准确反映样本间的语义关系。高维嵌入空间中的距离度量存在“维度灾难”问题，可能导致混合操作引入不稳定性。其次，噪声样本的残差混合可能产生语义歧义。特别是在医疗数据等专业领域，简单的线性混合可能破坏原始样本的关键特征模式。最后，不同模型架构的特征表示能力存在差异，对于某些框架（如 COMET），深层特征的混合可能保留更多语义信息，而对于其他框架，相同操作可能导致特征退化。

#### 4.4.5 可视化视角下的残差混合鲁棒性退化分析

本节深入探究残差混合操作与现有对比学习模型结合时导致鲁棒性退化的内在机理。表 4-2、表 4-4 与 表4-5 显示在嵌入空间采用残差混合时，模型在噪声数据集上的性能显著低于原始模型。为阐明这一现象的原因，本研究设计了针对性实验，分析残差混合操作对模型鲁棒性的影响机制。

本实验选取了鲁棒性退化最为显著的 TFC 模型作为实验对象。首先在 AD 数据集上完成不含残差混合的 TFC 模型预训练与微调过程。随后选择两对样本：一对包含标签为正常的原始样本及其对应的数据增强样本，另一对包含标签为异常

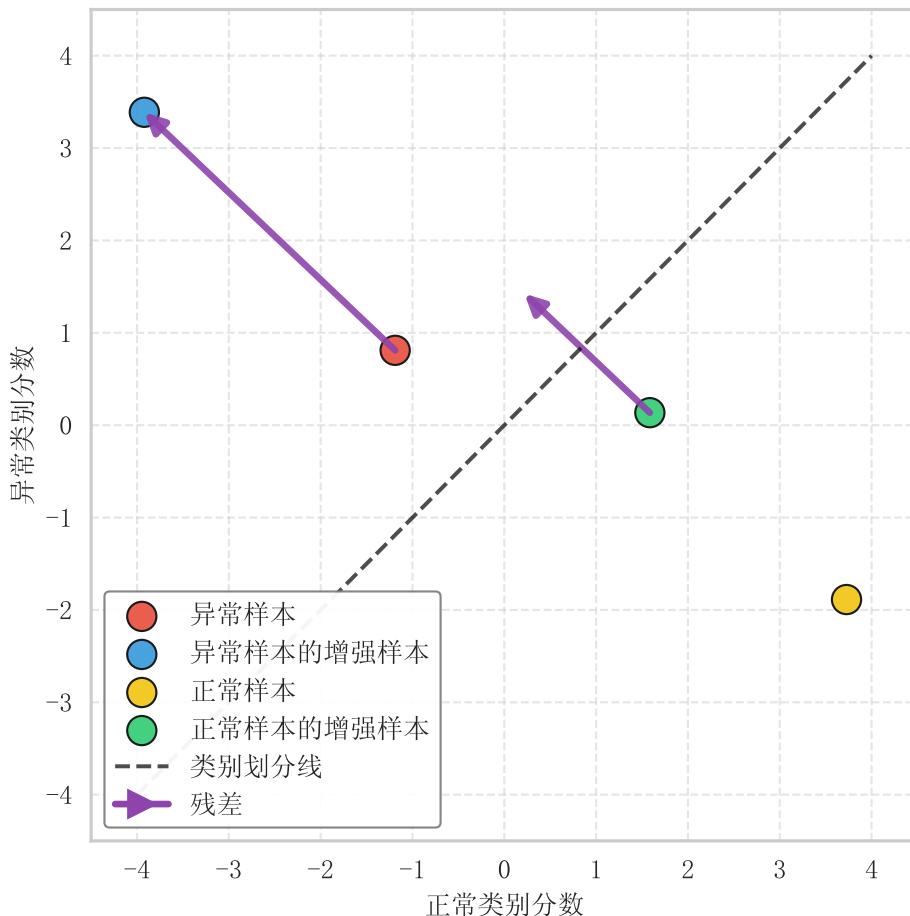


图 4-6 残差混合对 TFC 嵌入空间分布的影响

的原始样本及其对应的数据增强样本。将这四组数据输入经训练的模型，并从分类头末端提取其嵌入表示进行可视化分析。

可视化结果如图4-6所示，横纵坐标分别表示二分类任务中各类别的判别分数。异常样本及其增强版本均被准确分类至左上半区域，而正常样本及其增强版本则被正确归类至右下半区域，表明模型对四个样本均实现了准确分类。紫色的箭头表示异常样本对之间产生的残差，如果其被混合至绿色的点，其类型信息会发生改变。

然而，当 TFC 模型引入残差混合操作后，异常样本对产生的残差可能被混入正常样本的增强样本中，导致后者的类别信息发生偏移。这种混合效应使得在对比学习过程中，正常样本与其经残差混合后的增强样本之间的距离被不当地拉近。这种距离度量的扭曲破坏了原有类型信息的边界，最终导致模型对噪声样本的鲁棒性显著下降。

此分析结果揭示了残差混合操作在特定场景下可能产生的负面效应，为进一步优化对比学习框架中的数据增强策略提供了依据。

## 4.5 本章小结

本章针对脑电图分析中对比学习方法的噪声鲁棒性问题，提出了基于残差混合的数据增强策略。动机部分验证了噪声影响下特征偏移可替代性特征。但实验揭示了在嵌入器内部进行残差混合对分类性能和模型鲁棒性的改变是不稳定的，残差混合策略效果依赖模型架构，普适性有限。本研究虽然得到了否定性的研究结论，但是探索了在嵌入空间进行数据增强的效果，揭示了嵌入空间进行数据增强的风险，为后续研究提供了重要参考。未来可探索更合适的语义距离度量、非线性混合策略及自适应混合机制，以突破现有局限，推动脑电图分析技术发展。

## 结 论

在脑电图（EEG）信号分析领域，特别是在神经系统疾病诊断和临床应用中，高效处理标签稀缺性和复杂噪声干扰问题一直是研究的重点和难点。本研究针对这些挑战，提出了一种基于自监督学习的鲁棒预训练框架，并探索了基于残差混合的对比学习增强方法，旨在提升脑电图时间序列分类模型在实际医疗场景中的性能和鲁棒性。通过系统性的理论分析、方法设计和实验验证，本研究取得了显著的成果，为脑电图分析的智能化和临床应用提供了新的技术路径和理论支持。

首先，本研究提出了一种面向脑电时序建模的鲁棒预训练框架，该框架通过整合跨域重构学习、时频通道对齐以及附带额外噪声负样本的对比学习策略，成功实现了时频特征的深度融合和噪声模式的有效判别。跨域重构任务通过构建时域到频域和频域到时域的双向映射机制，使模型能够充分利用 EEG 信号在不同域中的互补特性；时频通道对齐策略则利用频域通道关系的稳定性校正时域表征偏移，进一步提升了模型对噪声的抵抗能力；引入额外噪声负样本的对比学习方法则通过显式建模噪声分布，显著增强了模型对复杂干扰的辨识能力。这些创新机制的协同作用，使得模型在特征表示的鲁棒性和判别性上均取得了突破。实验表明本模型在原始数据和多种噪声干扰条件下均优于现有的自监督和监督学习基准方法。特别是在带通噪声等复杂噪声环境下，本模型表现出显著的鲁棒性，性能下降幅度远低于其他方法。

其次，本研究探索了基于残差混合的对比学习增强方法，旨在通过在嵌入空间内实施残差混合操作，缓解传统对比学习中数据增强策略多样性不足的局限。实验结果是否定的，在嵌入空间进行残差混合不能有效地提供分类性能和鲁棒性上的改善。尽管如此，该探索仍明确了探索了在嵌入空间内进行混合的有效性，并指出了高维嵌入空间中特征混合可能引入的语义歧义风险。

综上所述，本研究通过创新的自监督学习框架和系统性的实验分析，为 EEG 信号鲁棒分类领域提供了新的理论视角和实践方案。未来研究可进一步探索多元信息融合技术、设计更具适应性的噪声建模方法，以及优化嵌入空间增强策略，以突破现有局限，推动脑电图分析技术在神经疾病诊断和脑机接口等领域的广泛应用。

本研究仍存在一些值得进一步探索和改进的方向：

首先，可以进一步优化自监督学习框架的设计，特别是在噪声建模和特征表示的适应性方面。管本研究通过引入额外噪声负样本的对比学习策略有效提升了

模型对复杂噪声的辨识能力，但当前方法对噪声分布的建模仍较为静态，难以完全适应真实临床环境中噪声的动态变化和多源性特征。因此，可以探索基于生成对抗网络或变分自编码器的动态噪声生成模型，通过数据驱动的方式模拟更贴近实际场景的噪声分布，从而进一步增强模型的鲁棒性。

其次，针对跨域重构学习机制，可以引入更复杂的时频特征交互技术，例如基于图神经网络的时频通道关系建模，以更精细地捕捉 EEG 信号在时域和频域中的多层次关联特性。

## 参考文献

- [1] HENRY J C. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, Fifth Edition [J]. Neurology, 2006, 67 (11) : 2092.
- [2] HAMALAINEN M, HARI R, ILMONIEMI R J, et al. Magnetoencephalography: Theory, Instrumentation, and Applications to Noninvasive Studies of the Working Human Brain [J]. Reviews of Modern Physics, 1993, 65 (2) : 413-497.
- [3] SATAPATHY S K, DEHURI S, JAGADEV A K. An Empirical Analysis of Different Machine Learning Techniques for Classification of EEG Signal to Detect Epileptic Seizure [J]. International Journal of Applied Engineering Research, 2016, 11 (1) : 120-129.
- [4] 邱锐琴. 视频脑电图监测在癫痫患儿诊断及定位中的作用初探 [J]. 中国实验诊断学, 2020, 24 (02) : 205-207.
- [5] TAVARES T P. Electroencephalography (EEG) [G] //Encyclopedia of Personality and Individual Differences. 2020 : 1266-1269.
- [6] WAGH N, WEI J, RAWAL S, et al. Evaluating Latent Space Robustness and Uncertainty of EEG-ML Models under Realistic Distribution Shifts [C/OL] // Advances in Neural Information Processing Systems, New Orleans, USA, 2022 [2023-09-04] . openreview.net/forum?id=KRk0lBRPpOC.
- [7] SHAH V, von Weltin E, LOPEZ S, et al. The Temple University Hospital Seizure Detection Corpus [J]. Frontiers in Neuroinformatics, 2018, 12 : 83.
- [8] ASSRAN M, DUVAL Q, MISRA I, et al. Self-supervised learning from images with a joint-embedding predictive architecture [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, SF, USA, 2023 : 15619-15629.
- [9] GE S, MISHRA S, LI C-L, et al. Robust Contrastive Learning Using Negative Samples with Diminished Semantics [C] //Advances in Neural Information Processing Systems : Vol 34, CA, USA, 2021 : 27356-27368.
- [10] UELWER T, ROBINE J, WAGNER S S, et al. A Survey on Self-Supervised Representation Learning [J]. arXiv preprint arXiv:2106.00750, 2023 (arXiv:2308.11455).

- [11] TANG S, DUNNMON J, SAAB K K, et al. Self-Supervised Graph Neural Networks for Improved Electroencephalographic Seizure Analysis [C/OL] // International Conference on Learning Representations, Held online, 2022 [2023-02-15] . openreview.net/forum?id=k9bx1EfHI\_-.
- [12] MOHSENVAND M N, IZADI M R, MAES P. Contrastive Representation Learning for Electroencephalogram Classification [C] // Proceedings of the Machine Learning for Health NeurIPS Workshop : Vol 136, Held online, 2020 : 238-253.
- [13] CHEN Y, HU B, KEOGH E, et al. DTW-D: Time Series Semi-Supervised Learning from a Single Example [C] // Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2013 : 383-391.
- [14] YE L, KEOGH E. Time series shapelets: a new primitive for data mining [C] // Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, 2009 : 947-956.
- [15] SCHAFER P. The BOSS Is Concerned with Time Series Classification in the Presence of Noise [J] . Data Mining and Knowledge Discovery, 2015, 29 (6) : 1505-1530.
- [16] DENG H, RUNGER G, TUV E, et al. A Time Series Forest for Classification and Feature Extraction [J] . Information Sciences, 2013, 239 (12) : 142-153.
- [17] YANG Y, XU Z. Rethinking the value of labels for improving class-imbalanced learning [C] // Proceedings of the 34th International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 2020.
- [18] ISMAIL FAWAZ H, LUCAS B, FORESTIER G, et al. InceptionTime: Finding AlexNet for Time Series Classification [J] . Data Mining and Knowledge Discovery, 2020, 34 (6) : 1936-1962.
- [19] XIAO Z, XU X, XING H, et al. RTFN: A Robust Temporal Feature Network for Time Series Classification [J] . Information Sciences, 2021, 571 (C) : 65-86.
- [20] VASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 2017 : 6000-6010.
- [21] ZERVEAS G, JAYARAMAN S, PATEL D, et al. A Transformer-Based Framework for Multivariate Time Series Representation Learning [C] // Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 2021 : 2114-2124.

- [22] SHANKARANARAYANA S M , RUNJE D. Attention Augmented Convolutional Transformer for Tabular Time-Series [C] //2021 International Conference on Data Mining Workshops (ICDMW), Auckland, New Zealand, 2021 : 537-541.
- [23] VARL M , YLMAZ H. Multiple classification of EEG signals and epileptic seizure diagnosis with combined deep learning [J] . Journal of Computational Science, 2023, 67 : 101943.
- [24] SIDDHAD G , GUPTA A , DOGRA D P , et al. Efficacy of transformer networks for classification of EEG data [J] . Biomedical Signal Processing and Control, 2024, 87 : 105488.
- [25] 张重生, 陈杰, 李岐龙. 深度对比学习综述 [J] . 自动化学报, 2023, 49 (01) : 15-39.
- [26] CHEN T , KORNBLITH S , NOROUZI M , et al. A Simple Framework for Contrastive Learning of Visual Representations [C] //International Conference on Machine Learning, Held online, 2020 : 1597-1607.
- [27] HE K , FAN H , WU Y , et al. Momentum Contrast for Unsupervised Visual Representation Learning [C] //2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Held Online, 2020 : 9726-9735.
- [28] CARON M , MISRA I , MAIRAL J , et al. Unsupervised learning of visual features by contrasting cluster assignments [C] //Proceedings of the 34th International Conference on Neural Information Processing Systems, NY, USA, 2020 : 13.
- [29] TIAN Y , SUN C , POOLE B , et al. What makes for good views for contrastive learning? [C] //Proceedings of the 34th International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 2020 : 13.
- [30] WANG X , QI G-J. Contrastive Learning With Stronger Augmentations [J] . IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45 (5) : 5549-5560.
- [31] ASSRAN M , CARON M , MISRA I , et al. Masked Siamese Networks for Label-Efficient Learning [C] //European Conference on Computer Vision, Tel Aviv, Israel, 2022 : 456-473.
- [32] MISRA I , van der MAATEN L. Self-Supervised Learning of Pretext-Invariant Representations [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Held online, 2020 : 6707-6717.

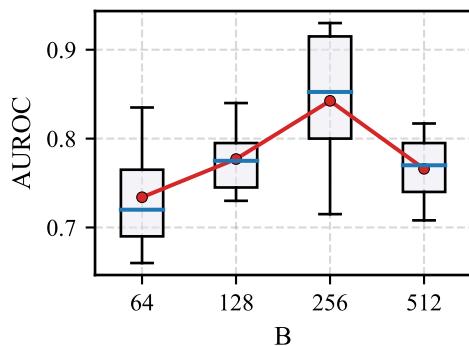
- [33] CRESWELL A, WHITE T, DUMOULIN V, et al. Generative Adversarial Networks: An Overview [J]. IEEE Signal Processing Magazine, 2018, 35 (1) : 53-65.
- [34] GOODFELLOW I J, Pouget-Abadie J, MIRZA M, et al. Generative Adversarial Networks [J]. Communications of the ACM, 2020, 63 (11) : 139-144.
- [35] MESCHEDER L M, GEIGER A, NOWOZIN S. Which Training Methods for GANs do actually Converge? [C] //International Conference on Machine Learning, Vienna, Austria, 2018 : 3481-3490.
- [36] ARJOVSKY M, CHINTALA S, BOTTOUL. Wasserstein Generative Adversarial Networks [C] //Proceedings of the 34th International Conference on Machine Learning, Held online, 2017 : 214-223.
- [37] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving Language Understanding by Generative Pre-Training [J]. arXiv preprint arXiv:1418.22162, 2018.
- [38] KINGMA D P, WELLING M. Auto-Encoding Variational Bayes [J]. arXiv preprint arXiv:1507.11271, 2022 (arXiv:1312.6114).
- [39] WOO G, LIU C, SAHOO D, et al. CoST: Contrastive Learning of Disentangled Seasonal-Trend Representations for Time Series Forecasting [C] //ICLR 2022, Held online, 2022.
- [40] ELDELE E, RAGAB M, CHEN Z, et al. Time-Series Representation Learning via Temporal and Contextual Contrasting [C] //Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, Held online, 2021 : 2352-2359.
- [41] FRAIKIN A F, BENNETOT A, ALLASSONNIERE S. T-Rep: Representation Learning for Time Series using Time-Embeddings [C] //The Twelfth International Conference on Learning Representations, Vienna, Austria, 2024.
- [42] CHEN J, YANG Y, YU T, et al. Brainnet: Epileptic wave detection from seeg with hierarchical graph diffusion learning [C] //Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, NY, USA, 2022 : 2741-2751.
- [43] LI Y, CHEN Z, ZHA D, et al. Towards Learning Disentangled Representations for Time Series [C] //KDD, New York, NY, USA, 2022 : 3270-3278.

- [44] BAEVSKI A, HSU W-N, XU Q, et al. Data2vec: A General Framework for Self-Supervised Learning in Speech, Vision and Language [C] //International Conference on Machine Learning, Baltimore MD, USA, 2022 : 1298-1312.
- [45] CHENG J Y, GOH H, DOGRUSOZ K, et al. Subject-Aware Contrastive Learning for Biosignals [J] . arXiv preprint arXiv:2007.04871, 2020.
- [46] KIYASSEH D, ZHU T, CLIFTON D A. Clocs: Contrastive Learning of Cardiac Signals across Space, Time, and Patients [C] // International Conference on Machine Learning, Held online, 2021 : 5606-5615.
- [47] MOHSENVAND M N, IZADI M R, MAES P. Contrastive Representation Learning for Electroencephalogram Classification [C] // Proceedings of the Machine Learning for Health NeurIPS Workshop : Vol 136, Held online, 2020 : 238-253 [2023-08-31] .
- [48] WAGH N, WEI J, RAWAL S, et al. Domain-Guided Self-Supervision of Eeg Data Improves Downstream Classification Performance and Generalizability [C] //Machine Learning for Health, Held online, 2021 : 130-142.
- [49] 徐晴, 葛成, 蔡标. 基于深度学习的癫痫脑电信号分类 [J] . 数据采集与处理, 2022, 37 (04) : 787-797.
- [50] BANVILLE H, CHEHAB O, HYVARINEN A, et al. Uncovering the Structure of Clinical EEG Signals with Self-Supervised Learning [J] . Journal of Neural Engineering, 2021, 18 (4) : 46020.
- [51] XU J, ZHENG Y, MAO Y, et al. Anomaly Detection on Electroencephalography with Self-Supervised Learning [C] // 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Online, 2020 : 363-368.
- [52] SETO S, THEOBALD B-J, DANIELI F, et al. REALM: Robust entropy adaptive loss minimization for improved single-sample test-time adaptation [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Hawaii, USA, 2024 : 2062-2071.
- [53] ANG Y, HUANG Q, BAO Y, et al. TSGBench: Time Series Generation Benchmark [J] . Proc. VLDB Endow., 2023, 17 (3) : 305-318.
- [54] REN W, ZHAO T, QIN W, et al. T-sas: Toward shift-aware dynamic adaptation for streaming data [C] // Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, Birmingham, UK, 2023 : 4244-4248.

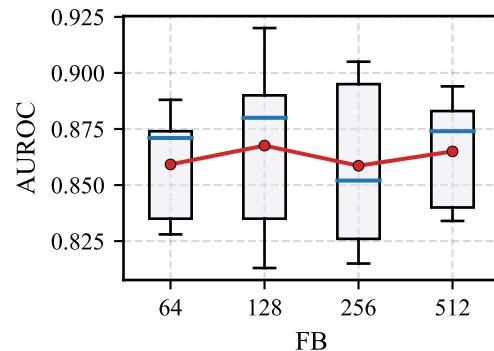
- [55] ZBONTAR J, JING L, MISRA I, et al. Barlow twins: Self-supervised learning via redundancy reduction [C] //International conference on machine learning, Held Online, 2021 : 12310-12320.
- [56] CHEN X, HE K. Exploring simple siamese representation learning [C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Held Online, 2021 : 15750-15758.
- [57] DONG J, WU H, ZHANG H, et al. SimMTM: A Simple Pre-Training Framework for Masked Time-Series Modeling [C] //Thirty-Seventh Conference on Neural Information Processing Systems, LA, USA, 2023 : 30.
- [58] ZHANG X, ZHAO Z, TSILIGKARIDIS T, et al. Self-Supervised Contrastive Pre-Training For Time Series via Time-Frequency Consistency [C/OL] // Advances in Neural Information Processing Systems, New Orleans, USA, 2022 [2024-03-15] . openreview.net/forum?id=OJ4mMfGKLN.
- [59] TANG S, DUNNMON J, SAAB K K, et al. Self-Supervised Graph Neural Networks for Improved Electroencephalographic Seizure Analysis [C] // International Conference on Learning Representations, Held online, 2022.
- [60] WU H, HU T, LIU Y, et al. Timesnet: Temporal 2d-Variation Modeling for General Time Series Analysis [C/OL] //ICLR, Kigali, Rwanda, 2023 [2024-01-25] . openreview.net/forum?id=ju\_Uqw384Oq.
- [61] NIE Y, NGUYEN N H, SINTHONG P, et al. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers [C/OL] //The Eleventh International Conference on Learning Representations, Kigali, Rwanda, 2023 [2024-04-14] . openreview.net/forum?id=Jbdc0vTOcol.
- [62] ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting [C] //Proceedings of the AAAI Conference on Artificial Intelligence : Vol 35, Held online, 2021 : 11106-11115.
- [63] WANG Y, HAN Y, WANG H, et al. Contrast Everything: A Hierarchical Contrastive Framework for Medical Time-Series [C] //Thirty-Seventh Conference on Neural Information Processing Systems, NY, USA, 2023 : 24.
- [64] TANG W, LONG G, LIU L, et al. Omni-Scale CNNs: a simple and effective kernel size configuration for time series classification [C] //International Conference on Learning Representations, Held online, 2022.

- [65] ESCUDERO J, ABASOLO D, HORNERO R, et al. Analysis of Electroencephalograms in Alzheimer's Disease Patients with Multiscale Entropy [J]. *Physiological measurement*, 2006, 27 (11) : 1091.
- [66] WICKSTROM K, KAMPFFMEYER M, MIKALSEN K O, et al. Mixing Up Contrastive Learning: Self-Supervised Representation Learning for Time Series [J]. *Pattern Recognition Letters*, 2022, 155 : 54-61.
- [67] RAMANATHAN A, MCDERMOTT J. Fall detection with accelerometer data using Residual Networks adapted to multi-variate time series classification [C] // 2021 International Joint Conference on Neural Networks (IJCNN), Held Online, 2021 : 1-8.
- [68] BAI S, KOLTER J Z, KOLTUN V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling [J]. *CoRR*, 2018, 1803.01271.

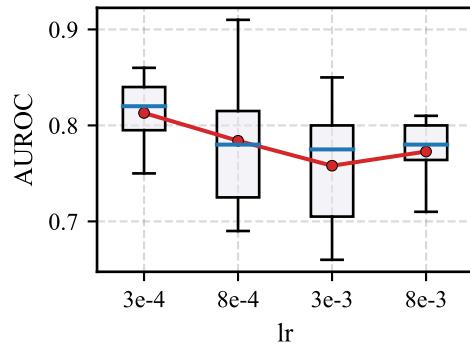
## 附录 A 第四章实验中各混合模型的超参数分析图



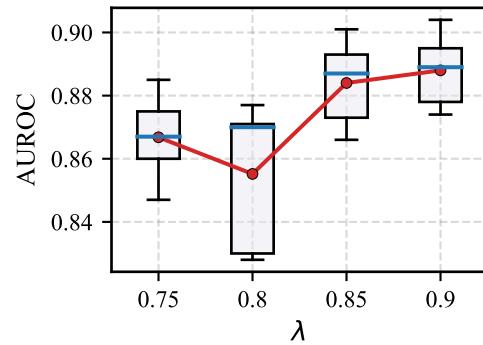
(a) 预训练批次的影响



(b) 微调批次的影响



(c) 学习率的影响



(d) 混合系数的影响

图 A-1 使用了残差混合的 TS-TCC 的超参数分析箱线图

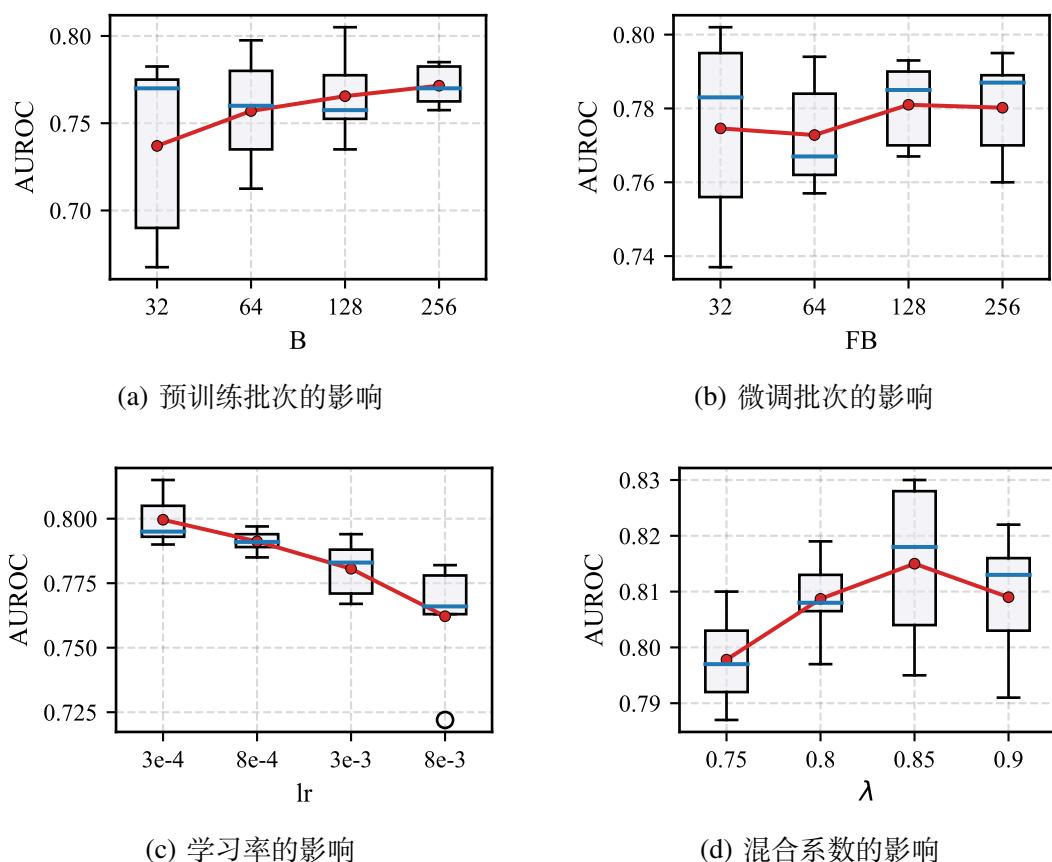


图 A-2 使用了残差混合的 SimMTM 的超参数分析箱线图