

学号： 222320045

学校代码： 10336

密级：

杭州电子科技大学

硕士学位论文

(专业学位)



论文题目： 基于 3D 卷积神经网络和生成对抗
网络的运动想象脑电信号解码研究

作者姓名： 黄飞

指导教师： 孙曜 高级实验师

专业学位类别： 电子信息

专业领域： 控制工程

所在学院： 圣光机学院

完成时间： 2025 年 4 月

杭州电子科技大学硕士学位论文

**基于 3D 卷积神经网络和生成对抗网络
的运动想象脑电信号解码研究**

研 究 生：黄飞

指导教师：孙曜 高级实验师

2025 年 04 月

**Dissertation Submitted to Hangzhou Dianzi
University for the Degree of Master**

**Research on Decoding Motor Imagery
EEG Signals Based on 3D
Convolutional Neural Network and
Generative Adversarial Network**

Candidate: Huang Fei

Supervisor: Sun Yao

April, 2025

摘要

运动想象（Motor Imagery, MI）脑机接口（Brain-Computer Interface, BCI）系统通过采集受试者在进行运动想象任务时产生的脑电信号（MI-EEG），并将其解码为可用于控制外部设备的指令信号，从而实现人与环境的交互。该技术在运动康复、远程控制和人体增强方面展现出重要应用潜力和独特价值。实现 MI-BCI 系统的高效运行，依赖于对足量的 MI-EEG 信号进行特征提取及准确分类。但是，MI-EEG 由于不同受试者的大脑活动的差异难以泛化，现有的分类方法缺乏受试者自适应的特征提取机制，同时，还忽略了 MI-EEG 信号的重要空间拓扑信息和全局时间变化趋势，这些问题限制了分类的准确性。另一个难点是，MI-EEG 具有非平稳性和信噪比低的特点，这导致难以获取到训练高性能的深度学习模型所需的大量数据，最终影响到分类精度。本文针对 MI-EEG 解码研究中存在的两类问题进行的主要研究工作如下：

（1）针对 MI-EEG 解码研究中出现的缺乏时空特征提取和泛化能力差的问题，本文设计了基于多尺度特征的 3D 卷积神经网络（Multi scale feature 3D convolutional neural network, MSF3D）。该网络可以在大脑区域中自适应地为与运动想象任务相关的空间通道和时间采样分配更高的权重，从而获取到 MI-EEG 的多尺度时空特征。实验结果表明该模型性能优于其他传统机器学习和多尺度卷积神经网络解码模型。在 MI-EEG 四分类数据集 BCICIV_2A（2A）和 High Gamma Dataset（HGD）上平均分类准确率分别达到 92.89% 和 95.39%，而在二分类数据集 BCICIV_2B(2B) 上平均分类准确率也达到 86.55%。

（2）针对 MI-EEG 解码研究中出现的时序数据不充足的问题，本文设计了基于生成对抗网络（Generative Adversarial Network, GAN）的数据增强网络模型 WTGAN。该模型一方面引入 Wasserstein 距离和梯度惩罚机制缓和传统 GAN 训练时出现的梯度消失现象，另一方面引入的注意力模块通过动态权重分配机制，可以同步捕获电极间空间关联性与 MI-EEG 信号的多尺度时序依赖。实验结果表明该模型性能优于其他几种具有代表性的数据增强模型，在数据集 BCICIV_2A（2A）取得了最高的平均分类准确率。

关键词：脑机接口，运动想象，脑电信号解码，卷积神经网络，生成对抗网络

ABSTRACT

Motor Imagery (MI)-based Brain-Computer Interface (BCI) systems acquire electroencephalography (EEG) signals generated during motor imagery tasks (MI-EEG) and decode them into control commands for external devices, thereby enabling human-environment interaction. This technology demonstrates significant potential and unique value in applications such as motor rehabilitation, remote control, and human enhancement. The efficient functioning of MI-BCI systems relies on effective feature extraction and accurate classification of a sufficient amount of MI-EEG signals. However, due to individual differences in brain activity, MI-EEG signals exhibit poor generalizability. Existing classification methods often lack subject-adaptive feature extraction mechanisms and fail to capture the critical spatial topological information and global temporal dynamics inherent in MI-EEG signals, which limits classification accuracy. Another major challenge lies in the non-stationary nature and low signal-to-noise ratio of MI-EEG, which makes it difficult to obtain large-scale datasets required to train high-performing deep learning models, ultimately affecting classification performance.

To address these challenges in MI-EEG decoding, this study focuses on two primary research directions:

(1) To tackle the issues of insufficient spatiotemporal feature extraction and poor generalization in MI-EEG decoding, this paper proposes a Multi-Scale Feature 3D Convolutional Neural Network (MSF3D). This network adaptively assigns higher weights to spatial channels and temporal samples associated with motor imagery tasks within brain regions, thereby capturing multi-scale spatiotemporal features of MI-EEG signals. Experimental results demonstrate that the proposed model outperforms traditional machine learning methods and other multi-scale CNN decoding models. It achieves average classification accuracies of 92.89% on the four-class MI dataset BCICIV_2A (2A) and 95.39% on the High Gamma Dataset (HGD), and 86.55% on the binary-class dataset BCICIV_2B (2B).

(2) To address the issue of insufficient temporal data in MI-EEG decoding, this paper introduces a data augmentation model based on Generative Adversarial Networks (GAN), named WTGAN. The model incorporates Wasserstein distance and gradient penalty to mitigate the gradient vanishing problem commonly observed in traditional

GAN training. Additionally, an attention mechanism is introduced to dynamically allocate weights, enabling the simultaneous capture of inter-electrode spatial correlations and multi-scale temporal dependencies of MI-EEG signals. Experimental results show that the proposed model outperforms several representative data augmentation methods and achieves the highest average classification accuracy on the BCICIV_2A (2A) dataset.

Keywords: Brain-Computer Interface; Motor Imagery; EEG Signal Decoding; Convolutional Neural Network; Generative Adversarial Network

目 录

第 1 章 绪论.....	1
1.1 研究背景与意义.....	1
1.2 MI-EEG 解码研究现状.....	2
1.2.1 MI-EEG 预处理方法.....	3
1.2.2 基于传统机器学习的 MI-EEG 特征提取方法	4
1.2.3 基于传统机器学习的 MI-EEG 特征分类方法	6
1.2.4 基于深度学习的 MI-EEG 解码方法	7
1.2.5 基于生成对抗网络的 MI-EEG 数据增强方法	8
1.3 研究内容和文章架构.....	9
第 2 章 相关理论与技术	12
2.1 MI-EEG 原理及特点.....	12
2.2 事件相关同步/去同步现象	13
2.3 CNN 基本原理	14
2.4 GAN 基本原理.....	16
2.5 本章小节.....	18
第 3 章 基于 MSF3D 的 MI-EEG 解码研究	19
3.1 引言.....	19
3.2 MSF3D 网络模型结构.....	19
3.2.1 MI-EEG 的 3D 表示.....	20
3.2.2 3D 空间注意力模块.....	21
3.2.3 多尺度时间注意力模块.....	23
3.2.4 分类模块.....	25
3.3 数据集及实验设置介绍.....	25
3.3.1 BCICIV ₂ A 数据集	26
3.3.2 BCICIV ₂ B 数据集	27
3.3.3 High Gamma Dataset 数据集.....	28
3.3.4 实验设置.....	29
3.4 实验结果与分析.....	30
3.4.1 对比实验.....	31
3.4.2 消融实验.....	36
3.4.3 留一交叉验证实验.....	40
3.5 本章小节.....	42
第 4 章 基于 WGAN 的 MI-EEG 数据增强研究	43
4.1 引言.....	43

4.2 WTGAN 网络模型结构.....	43
4.2.1 WGAN-GP 优化思路.....	44
4.2.2 编码器和解码器.....	47
4.2.3 生成器和判别器.....	49
4.3 数据集及实验设置介绍.....	50
4.3.1 BCICIV ₂ A 数据集.....	50
4.3.2 实验设置.....	50
4.4 实验结果与分析.....	51
4.4.1 数据增强效果分析.....	52
4.4.2 分类准确性分析.....	53
4.4.3 不同混合比影响.....	55
4.5 本章小节.....	55
第 5 章 总结与展望	57
5.1 总结.....	57
5.2 展望.....	57
参考文献.....	58

第 1 章 绪论

1.1 研究背景与意义

脑机接口（BCI）是一种直接连接人脑与外部设备的交互系统^[1]，能够将大脑的神经活动转换为计算机可识别的指令，从而实现对外部设备的控制。BCI 作为一种颠覆性的人机交互技术，已成为科学研究的热点前沿。其应用领域正逐步突破传统的临床康复训练，不再仅限于帮助神经或肌肉功能障碍患者，而是逐步扩展到工业控制、教育教学、智能生活、娱乐休闲等多个领域，为广大群众提供智能化和便捷化的生活生产方式^{错误!未找到引用源。-4]}。一个典型的 BCI 系统如图 1.1 所示，主要包含信号采集、预处理、特征提取、特征分类以及信号表达与信息反馈这五个模块。

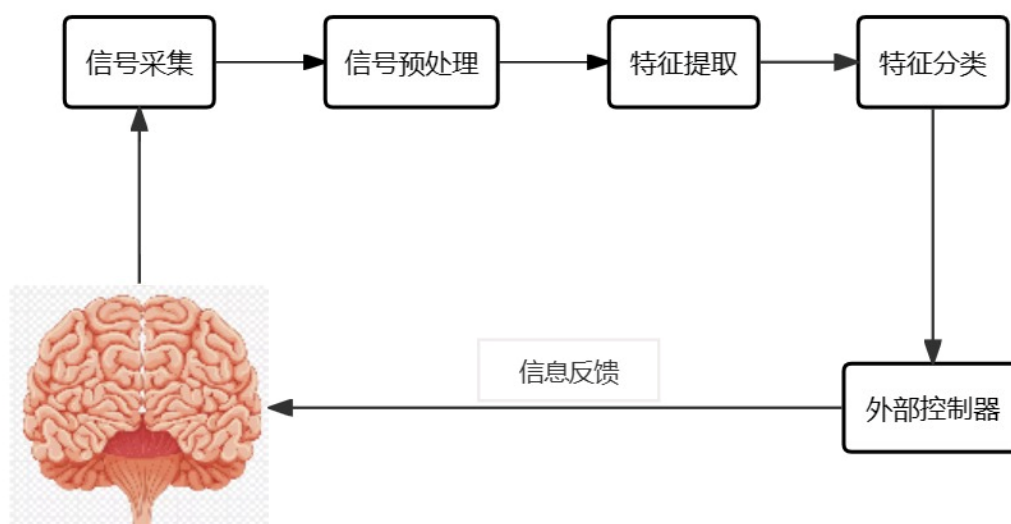


图 1.1 典型的 BCI 系统示意图

在信号采集环节，根据提取大脑神经生理学信息的方法和种类，可以分为侵入式、半侵入式和非侵入式。侵入式记录脑电信号如植入微电极等虽然快速高效，但是会使受试者机体产生不可逆的损伤，目前只用于严重残疾人员，所以数据量较少^[5-7]。非侵入式则使用外部设备记录大脑神经活动，如通过布置在大脑不同分区的电极采集电信号的脑电图（Electroencephalography, EEG）^[8]和通过脑磁仪器采集不同神经活动磁场变化的脑磁图（Magnetoencephalography, MEG）^[9]。还有介于二者之间的半侵入式如皮层脑电（Melectrocorticography, ECoG）^[10-12]。在这些方法中，脑电图由于其非侵入式及时间和空间分辨率较为出色的特点，因而成为最常用的脑电信号范式^[13]。在脑电图中有一类特殊的种

类也即运动想象（Motor Imagery, MI）脑电图（MI-EEG）。运动想象是一种在没有实际运动的情况下，仅通过大脑意念来模拟运动的心理过程，因此 MI-EEG 属于自发式 BCI 系统。在此过程中，大脑皮层相关区域（如初级运动皮层、辅助运动区等）会产生特定的神经活动，可通过脑电信号进行记录和分析。基于 MI-EEG 的脑机接口技术可以解码这些脑电信号，并将其转换为控制外部设备的指令，在医疗康复和人体增强等领域具有广泛应用价值^[14]。

近年来，MI-EEG 技术取得了显著进展。例如，研究表明 MI-EEG 信号可用于控制假肢，使截肢者恢复基本的手部功能。此外，康复训练设备结合 MI-BCI 技术后，患者的运动恢复效率提高了 30%-50%，显著加速了康复进程。除了用于中风、瘫痪或假肢患者的医疗康复之外，MI-BCI 还有广泛的应用，例如 MI-EEG 结合 BCI 技术可用于远程控制机器人，实现高效的人机交互。例如，在工业生产和危险环境，如核电站维护中，操作人员可以通过 MI-BCI 控制机器人执行复杂任务。此外，MI-EEG 可用于增强现实（AR）和虚拟现实（VR）交互。例如，用户可通过 MI-EEG 控制虚拟角色的运动，提高游戏沉浸感^[15]。在 VR 训练系统中，MI-BCI 可帮助飞行员和宇航员进行模拟训练，并提高注意力和反应速度。未来，MI-EEG 可集成至智能穿戴设备，如脑控耳机、脑控眼镜等，使用户能够通过意念操作电子设备。例如，某些研究正在探索 MI-BCI 在智能家居系统中的应用，使用户可以通过运动想象控制灯光、空调等家电。

总而言之，MI-EEG 在医疗康复和人体增强领域展现出巨大潜力。从中风康复到智能假肢，再到远程机器人控制和 VR 交互，MI-BCI 正在逐步改变人类与技术交互的方式。随着技术的不断发展，未来 MI-BCI 预计将在医疗、工业、军事等多个领域发挥更大的作用，进一步提升人类的生活质量和工作效率。在这个过程中，MI-EEG 的解码任务即通过各种方法实现脑电信号的分类识别工作，准确地分辨出运动意图就显得尤为重要。

1.2 MI-EEG 解码研究现状

MI-EEG 解码是 MI-BCI 技术应用落地的重要基础，这一环节由两部分组成，即特征提取和特征分类。特征提取旨在将原始 MI-EEG 数据转换为更具代表性且易于分析的形式，通过降维减少冗余信息，并提取其中关键特征，用以提升后续分类的效果。而特征分类主要目的则是识别不同类别的 MI-EEG，实现与外部设备的交互，两者相辅相成，为 MI-BCI 的应用奠定了基础。

MI-EEG

前者的主要特征是各个环节有明显的区分，一般是先通过算法对时域、频域或时频域进行特征提取，接着用机器学习领域应用广泛的分类器，例如 SVM、KNN 等完成分类任务。而后者依托深度神经网络强大的非线性拟合能力既可以

结合传统的特征提取方法完成分类任务，也可以设计成一种端到端的神经网络结构，集特征提取和特征识别为一体，避免手动提取所依赖的先验知识和实践经验，减少了人为因素在过程中的干扰。但两者都需要在解码任务开始之前对原始数据进行预处理，去除噪声、眼电伪迹或数据降维，将原始脑电信号处理为解码可接受的标准格式。

1.2.1 MI-EEG 预处理方法

脑电信号在采集过程中会受到诸多不可控因素的影响，难免会掺杂噪声。因此，在进行下一步信号解码前，需要进行预处理以去除干扰。在实践中常见的预处理方法包括以下几种：

（1）阈值法

阈值法是一种简单且常用的伪迹去除方法，主要用于消除由于眼动、肌肉活动或其他外部干扰导致的异常信号。在实验前，通常会让受试者多次执行特定的运动，如眨眼、头部移动或肢体轻微活动，并记录相应的脑电信号，以此建立参考基线并确定合理的阈值范围。在实际数据处理中，如果某些信号分量的振幅超过设定阈值，则认为该信号受到伪迹影响，需要进行舍弃或进一步处理。这种方法具有计算简单、执行效率高的优点，适用于快速去除强烈干扰信号。然而，由于阈值的设定通常基于经验，难以精准适应个体差异，且对于幅度较低但仍存在干扰的伪迹可能无法有效识别，因此在一些高精度要求的应用场景中，通常会结合其他信号处理方法。

（2）带通滤波

在 EEG 信号预处理中，带通滤波是一种常见且有效的信号处理方法，主要用于去除特定频段以外的噪声和伪迹，以增强目标信号的质量。在应用带通滤波时，首先需要将原始脑电信号从时域转换到频域，通常采用傅里叶变换方法，以便分析不同频率成分的分布情况。

在频域中，根据研究需求设定适当的频率范围，例如 0.5Hz-50Hz 作为脑电信号的有效频段，然后对该范围之外的信号进行滤除，以抑制低频漂移、工频干扰（如 50Hz 电源噪声）以及高频肌电伪迹。经过频域处理后，再使用逆变换如逆傅里叶变换将信号转换回时域，此时的信号已经完成了伪迹去除，保留了目标频段内的有效信息，使后续特征提取和分类更加准确。

（3）主成分分析

主成分分析（Principal Component Analysis, PCA）^[16]是一种常用的数据降维方法，主要用于提取信号的主要特征，同时减少冗余信息和噪声干扰。其基本原理是通过线性变换寻找原始信号数据中方差最大的方向，并将数据投影到这些方向上，从而实现信息的有效压缩，同时尽可能保留原始数据的主要特征。

在具体操作过程中，PCA 首先计算原始脑电数据的协方差矩阵，以分析不同信号通道或数据维度之间的相关性。接着，通过特征值分解或奇异值分解（SVD），提取出数据分布中方差最大的特征方向，即主成分，并按照贡献度排序。这些主成分代表了数据中的主要变化模式，通常前几个主成分就能涵盖大部分原始信息。因此，可以选取前几个主成分，构成一个低维且互不相关的新特征向量空间，以此完成降维。

PCA 具有计算高效、操作简单的优势，广泛应用于脑电信号分析中的噪声去除、特征提取以及数据可视化等任务。在 MI-EEG 应用中，PCA 可用于提取关键的 EEG 频率成分，以提高信号分类的准确性。

（4）独立成分分析

独立分量分析（Independent Component Analysis, ICA）^[17]是一种基于空间滤波的信号处理方法，在 MI-EEG 预处理中被广泛应用，特别是在去除伪迹和提高信号质量方面具有显著优势。ICA 主要用于将多通道 EEG 数据分解为一组空间固定且时间上相互独立的成分，从而更好地识别和去除非神经源性的干扰信号，如眼动伪迹、肌电噪声和电源干扰。

其基本原理是假设观测到的 EEG 信号是多个独立信号源的线性混合，而 ICA 通过寻找这些独立信号源，使得分解后的成分在统计意义上彼此独立。具体而言，ICA 采用矩阵分解的方法，将原始 EEG 数据投影到一个新的坐标空间，使得每个独立成分（Independent Component, IC）在时间序列上具有最大的独立性。经过分解后，可以通过分析每个独立分量的时空特性来识别伪迹分量，并通过设定阈值或人工筛选的方式去除不需要的成分，从而获得更加纯净的 EEG 信号。

1.2.2 基于传统机器学习的 MI-EEG 特征提取方法

预处理通过特征选择实现了对原始数据的降维操作，以减少数据冗余并提高后续特征提取的效率和准确性。这一过程能够有效去除无关或冗余的信息，使得模型能够更专注于关键特征，提高信号分析的稳定性和泛化能力。在 MI-EEG 信号处理中，特征提取是信号分析的核心环节，常见的特征提取方法主要包括时域、频域、时频域和空域方法。不同的 MI-EEG 特征具有不同的表现形式，因此需要根据特征的具体性质选择合适的特征提取方法。

（1）时域特征提取方法

由于 MI-EEG 是直观的时序数据，因此时域方法是最早用于信号分析的特征提取方式，直接提取信号的波形特征，如振幅、均值、方差、波峰、极大极小值等，计算简单，能够直观反映 MI-EEG 信号的时域信息。然而，该方法无法表征脑电信号的频率成分及其变化趋势，难以全面刻画 MI-EEG 信号特征，

因而存在一定局限性。

常用的时域特征提取方法有：过零点法、直方图法、峰值检测法和波形识别法等^[18]，在这些方法中应用较广泛的是 Chai 等人^[19]提出的自回归模型。该模型假设当前信号值是由其历史值的加权和加上一个随机误差项所决定的，即利用过去若干时刻的 EEG 信号来预测当前信号，从而捕捉数据的时间相关性和动态特征。进一步地，Atyabi 等人^[20]将自回归模型中高维度和低维度的特征进行混合，得到更能表征原始信号的混合特征以解决低维度特征的信号不充足和高维度特征噪声多的难题。

（2）频域特征提取方法

由于 EEG 信号在不同频段具有不同的生理意义，因此通过频域特征提取可以有效获取不同频段脑电信号的频率、相位等特征信息。同理，片面的频域特征提取忽略了时域信息，造成 EEG 特征不充分。

经典的频域特征提取方法以快速傅里叶变换^[21]为代表。除此之外还有功率谱密度^[22]、谱矩心^[23]和波段功率分析^[24]。其中应用较多的功率谱密度可以量化 EEG 信号在不同频率上的功率分布。在 MI-BCI 中，不同运动想象任务会引起 α 频段（8~13Hz）和 β 频段（14~30Hz）的事件相关同步（Event Related Synchronization, ERS）和事件相关去同步（Event Related Desynchronization, ERD）现象，功率谱密度可用于检测这些变化。

（3）时频域特征提取方法

MI-BCI 脑电信号是一种非平稳信号，具有时变性，即信号的频率成分会随时间动态变化。因此，仅使用时域或频域方法提取 EEG 特征可能无法全面捕捉其信息。时频域特征提取方法结合了时域和频域分析的优势，能够同时提供 EEG 信号在时间和频率上的分布信息。

经典时频域特征提取方法有：短时傅里叶变换^[25]、连续小波变换^[26]和小波包分解^[27]。此外，时频域提取到的特征时往往作为下一步神经网络分类器的输入。例如，Pattnaik 等人^[28]在二分类 MI-EEG 特征提取任务中使用离散小波变换（Discrete Wavelet Transform, DWT）结合深度学习算法 ANN 进行特征分类，取得了 80% 以上的分类准确率。相比于连续小波变换，离散小波变换通过二进制尺度的分解方式大大减少计算量。

（4）空域特征提取方法

除了时频域信息，在采集脑电信号时不同空间位置的电极通道也表征了空间维度的特征。为了有效提取此类特征，学者进行大量的尝试。

共空间模式（Common Spatial Pattern, CSP）^[29]在 MI-EEG 空域特征提取方面是不可绕开的经典算法。CSP 的核心思想是通过一组线性变换即空间滤波器

对多通道 EEG 信号进行处理，以最大化不同类别之间的方差差异，同时最小化同类别内部的方差变化，从而提取最具判别力的特征。进一步地，为了提升 CSP 的适应性，研究者提出了多种改进算法。例如，Razi 等人^[30]采用了一种多类别分类策略，结合 OVO（One-Versus-One）和 OVR（One-Versus-Rest）方法，并利用 Dempster-Shafer 理论进行二分类。实验结果显著优于传统机器学习算法。Ang 等人^[31]提出了滤波器组共空间模式（Filter Bank Common Spatial Pattern, FBCSP）算法。FBCSP 先利用一组带通滤波器将 EEG 信号分解为多个频带，每个频带通常覆盖 8~30Hz 的脑电活动范围，随后对各频带数据分别应用 CSP 进行特征提取。最终，将所有 CSP 提取的特征进行拼接，提高了 MI-EEG 的解码性能。此外，Novi 等人^[32]在 FBCSP 基础上进一步优化，提出了子带共空间模式（Sub-Band Common Spatial Pattern, SBCSP）。该方法结合频域分解和 CSP，在不同子频带上执行 CSP 分析，提取更丰富的空频特征。

1.2.3 基于传统机器学习的 MI-EEG 特征分类方法

在特征提取完成之后紧接着就要进行特征分类，当前应用较多的基于传统机器学习的 MI-EEG 分类算法有线性判别分析（Linear Discriminant Analysis, LDA）、支持向量机（Support Vector Machine, SVM）以及 K 近邻算法（K-Nearest Neighbors, KNN）等。三类方法的特点和适用范围各不相同，下面详细论述。

LDA 的核心原理是寻找一个最优投影方向，使得不同类别的 EEG 数据在该方向上尽可能分开，同时同类别数据尽可能聚集，从而提高分类的可分性。LDA 通过计算不同类别的均值向量和协方差矩阵，最大化类间方差，最小化类内方差，以构建一个最优的线性判别边界。LDA 假设 EEG 数据在各类别内服从高斯分布，且不同类别的数据具有相同的协方差矩阵，因此在数据线性可分的情况下，能够实现较好的分类效果。如 Long 等人^[33]利用 LDA 在 MI-BCI 系统中对 MI-EEG 进行分类从而实现轮椅操控

SVM 的核心原理是通过寻找一个最优超平面，将不同类别的 EEG 数据进行最大间隔分隔，以提高分类的泛化能力。SVM 通过构造一个决策边界，使得边界两侧的 EEG 数据尽可能远离分隔超平面，从而增强分类的鲁棒性。当 EEG 数据是线性可分时，SVM 直接利用线性核函数进行分类；当数据存在非线性特征时，可以通过核技巧如径向基函数核、Sigmoid 核、多项式核等将 EEG 数据映射到高维特征空间，使其在高维空间中变得线性可分。Ghaemi 等人^[34]通过改进的二元引力搜索算法来寻找左右手分类中的最佳通道并结合小波分析和 SVM 等算法完成分类，在 BCICIV_2B 数据集中平均准确率达 76.24%。

KNN 用于 EEG 数据分类的基本原理是基于测量样本之间的相似性进行决

策。首先，将 EEG 数据通过特征提取方法如 PCA 等转换为特征向量。然后，在分类阶段，对于一个待分类的样本，KNN 算法计算其与训练集中所有样本的距离，该距离通常用欧氏距离或余弦相似度，并选取最近的 K 个邻居。最后，依据这 K 个邻居的多数类别决定该样本的最终分类标签。KNN 是一种非参数方法，适用于 EEG 数据这种高维、非线性和小样本的情况。Dash 等人^[35]将 KNN 算法用于癫痫发作的表层脑电信号分类。但由于 KNN 计算过程复杂度较高，需结合降维或高效索引结构优化性能。

总之，LDA 适用于线性可分 EEG 分类任务，SVM 适用于高维复杂 EEG 数据，而 KNN 更适用于非线性数据但计算量较大。在具体应用中，需要根据 EEG 数据特性选择合适的分类方法。

1.2.4 基于深度学习的 MI-EEG 解码方法

针对传统机器学习分类精度较低且依赖人工特征提取的问题，深度学习算法在 MI-EEG 解码任务中得到了广泛应用。其中，卷积神经网络（Convolutional Neural Network, CNN）作为深度学习的代表模型，能够自主学习脑信号的多维特征，无需手工设计特征提取器。在 MI 任务中，CNN 可以学习大脑信号中相关模式间的特殊依赖关系，提取更具判别力的特征表示^[36]。

Schirrmeister 等人^[37]验证了使用 CNN 解码 MI-EEG 数据的可行性。研究中使用了三种 CNN 架构，并针对这些架构评估了几种特定的设计选择。Lawhern 等人^[38]使用深度和可分离卷积构建了一个紧凑的轻型 CNN 网络 EEGNet，用于对脑电图信号进行精确分类。在 EEGNet 的基础上，Ingolfsson 等人^[39]融合时间卷积网络（Temporal Convolutional Network, TCN）创新地提出 EEG-TCNet 模型，该模型能够对 MI-EEG 时域特征进行提取，平均分类准确率有所提升。Sun 等人^[40]提出自适应时空图卷积网络 AST-GCN，同时利用时域特征和通道间的关联信息，以提升分类精度。Dai 等人^[41]为了进一步增强 MI-EEG 的时域、频域和空域的信息，设计了一种混合尺度的 CNN 分类架构 HSCNN。

然而，以上的这些方法使用的输入都是原始 MI-EEG 表示的 2D 数组。该数组以时间采样点的数量作为阵列宽度，以电极的数量作为阵列高度，通常会省略 MI-EEG 数据的空间信息。于是一些研究者试图将拓扑结构整合到深度学习架构中。Zhang 等人^[42]提出将带有电极位置信息的 1D 矢量 EEG 格式映射到 2D 网状 EEG 信号，该信号进一步用于形成 3D 格式数据，并将该数据输入到设计的 3D CNN 网络。Zhao 等人^[43]通过将原始脑电图信号转换为包含所有电极空间结构的 2D 数组序列，提出了一种 3D 表示。设计了具有不同感受野大小的多分支 3D CNN 来提取高级 MI EEG 相关特征。

此外，Li 等人^[44]证明了用注意力机制提高分类性能的可行性。他们将注意

力机制引入多尺度融合卷积神经网络。该网络从多脑区域表示信号中提取多尺度特征，并辅以密集融合策略以保留最大的信息。值得注意的是，Transformer 架构^[45]虽源于自然语言处理中的机器翻译任务，但其卓越的序列建模能力使其在脑电信号处理领域展现出独特优势。该架构通过自注意力机制实现全局依赖关系建模，这一特性已被 CTNet^[46]模型成功验证。其采用 CNN 进行局部特征提取，结合 Transformer 编码器构建全局表征，显著提升了模型性能。

1.2.5 基于生成对抗网络的 MI-EEG 数据增强方法

目前，大多数公开 MI-EEG 数据集的受试者数量通常不超过百人。尽管各种识别算法在这些数据集上取得了较高的识别精度，但这并不意味着它们能在更大规模的数据集中保持同样的性能。因此，一个关键挑战是如何获取足够的 MI-EEG 数据，以优化现有的识别算法，同时满足更大规模训练的需求。

受试者数量受限的主要原因在于 MI-EEG 信号的采集需要专门的脑机接口设备，这些设备不仅价格昂贵而且使用不便。此外，数据采集通常要求特定的实验范式，并依赖实验人员协助佩戴 EEG 设备，用户也需要严格配合。由于 EEG 信号信噪比较低，往往需要重复执行相同任务，以确保信号质量，这使得数据采集过程费时费力，影响了 EEG 解码技术的实际应用和用户接受度。因此，如何扩充 EEG 数据成为关键。

在 MI-EEG 数据增强方面，传统方法与图像数据增强方式相似，通常对时间序列进行几何变换，确保不改变信号的频率、空间分布及功率成分。例如，通过在 EEG 数据中添加噪声或旋转空间电极分布来模拟仿射变换，从而生成新的 EEG 数据样本。此外，实验任务的组合和调整也可用于合成额外的数据，以扩充训练集。

此外，窗口滑动也是一种常见的数据增强技术，该方法通过沿时间轴滑动固定大小的窗口截取数据，而不改变原始信号本身。例如，Fan 等人^[47]采用 0.5 秒滑动窗口对 EEG 进行数据扩展，并在 PhysioNet 数据集上验证了其有效性。该方法使系统仅需 0.5 秒的静息态 EEG 数据即可完成快速识别，并在 14 通道 EEG 设置下实现 99.32% 的平均准确率和 0.18% 的误差率，相比传统方法大幅提升了计算效率和实用性。

生成对抗网络（Generative Adversarial Network, GAN）在图像生成领域发展迅速，它可以有效地学习真实图像的分布特征，并生成具有相同分布的图像样本。同时，它可以生成人类视觉无法区分的图像，并使用增强的数据集来提高分类器的性能，使分类器拟合得更快。正因如此，近年来，GAN 在 EEG 数据增强中的应用逐渐增多。例如，Piplani 等人^[48]通过优化损失函数，使 GAN 生成具有真实功率谱特性的 EEG 数据，进而提升分类模型的准确率，从 90.8% 提

高至 95.0%。Hartmann 等人^[49]则提出 EEG-GAN 框架，基于 Wasserstein GAN 进行稳定训练，尽管该方法可生成接近真实的 EEG 信号，但仍局限于单通道数据。为改进 GAN 在 EEG 领域的应用，Abdelfattah 等人^[50]引入递归 GAN（RGAN），以 RNN 代替传统全连接层，从而更有效地捕获 EEG 的时间特征。Mirza 等人^[51]则在 WGAN-GP 结构中加入分类器设计出 CC-WGAN-GP，提高 GAN 生成的 EEG 数据在分类任务中的适用性。此外，Harada 等人^[52]采用基于长短时记忆（LSTM）的 GAN 结构，使其生成的 EEG 序列数据更能保留时序信息，在癫痫检测数据集上的实验结果显示，该方法可有效提升分类精度。

在 EEG 频谱图生成方面，Zhang 等人^[53]使用 DCGAN 生成 EEG 信号的频谱图图像，并应用于 MI-EEG 解码任务的数据扩展。与其他方法相比，该模型生成的频谱图与真实 EEG 频谱图更为相似，同时提高了分类精度。此外，Zhang 等人^[54]提出深度对抗数据增强框架 DADA，以优化小样本 EEG 任务的训练。其方法在 BCI 竞赛数据集上，将 CNN 分类器的准确率从 74.8% 提高到 79.3%，主要得益于改进的损失函数设计，使生成数据更加多样化。

Luo 等人^[55]提出了一种基于条件 Wasserstein GAN（CWGAN）的 EEG 数据增强框架，该方法通过梯度惩罚 Wasserstein GAN 生成微分熵特征，并在情绪识别任务中取得了显著效果。相比传统几何变换，CWGAN 能够更深入地学习 EEG 数据分布，提高分类模型的准确性。此外，Zhang 等人^[56]进一步提出多生成器 MG-CWGAN，结合 LDA 特征选择，在 SEED 数据集上进行实验，相较于 CWGAN，该方法收敛更快，并在 SVM 分类任务中表现更优。

在 EEG 时频特征增强方面，Zhang 等人^[57]基于 cDCGAN 生成 EEG 频谱图，并用于 MI 任务分类，实验表明该方法有效提高分类精度。Luo 等人^[58]则提出基于遗传算法的条件边界均衡 GAN 模型 CBEGAN，专注于 EEG DE 特征生成。其主要优势在于克服了传统 GAN 的训练不稳定问题，并显著加快了收敛速度。在多模态情感数据集上的实验结果显示，该方法在三分类任务和五分类任务中的准确率分别提升 4.6% 和 8.9%。

综合来看，GAN 在 EEG 数据增强中的应用日益广泛，结合不同任务需求和优化策略，这些方法为 EEG 识别系统提供了更丰富的数据支持，并推动了实际应用的发展。

1.3 研究内容和文章架构

通过前文对 MI-EEG 解码研究的背景及研究现状的详细描述，目前该领域研究中存在以下问题：

问题一：现有的基于深度学习的 MI-EEG 解码方法，其网络模型输入多是将 MI-EEG 表示为二维数组，这种 2D 表示忽略了电极通道之间的空间拓扑依赖

性。并且，使用的 2D 卷积核无法完全反映附近电极通道之间的相关性。因此，MI-EEG 分类准确性受到影响。同时，每个电极通道 EEG 信号都是一个时间序列，其特征也随时间变化，并在不同的个体和不同的任务之间表现出显著差异。受试者通常在某个时间集中注意力，但在其他时间分心，并且不同受试者在试验中的不同时间集中注意力。因此，当受试者专注于试验而忽略其他切片时，强调 EEG 时间切片对于提取时域特征是必要的。但现有解码方法忽视了受试者专注于试验时的 EEG 时间切片的特殊性，而从全流程时间切片中充分提取时间不变的高级时间特征来编码时间信息，导致针对性不足，分类性能受到影响。

问题二：由于 MI-EEG 自身非线性，非稳定，信号微弱，信噪比低，抗干扰能力差等多方面的原因导致数据量不充足，不能满足越来越复杂的深度学习网络模型的需要。并且 MI-EEG 是一种时间分辨率远高于空间分辨率的信号，但是当前的基于 GAN 的时序脑电数据扩增方法中，脑电时序数据生成领域的 GAN 模型研究很少，生成时序脑电数据能力还比较有限。并且传统 GAN 在训练过程中由于使用 JS 散度作为优化目标会导致梯度消失，梯度消失问题会严重阻碍 GAN 的训练，导致模式坍塌或生成样本质量低下，最终影响数据增强的效果。

针对上述问题，本文主要贡献如下：

针对问题一：本文提出了提取多尺度特征的 3D 卷积神经网络 MSF3D 模型结构，该模型可以通过 3D 表示和注意力模块提取电极通道间的依赖关系并在大脑区域自适应地为运动想象相关的空间通道和时间采样切片分配比运动无关的更高的权重。

针对问题二：本文提出了基于生成对抗网络的数据增强网络模型 WTGAN。该模型一方面引入 Wasserstein 距离和梯度惩罚机制缓和传统 GAN 训练时出现的梯度消失现象，另一方面引入的注意力模块通过动态权重分配机制，可以同步捕获电极间空间关联性与 MI-EEG 信号的多尺度时序依赖。

本文分为 5 个章节，论文组织结构见图 1.2，各章内容简介如下：

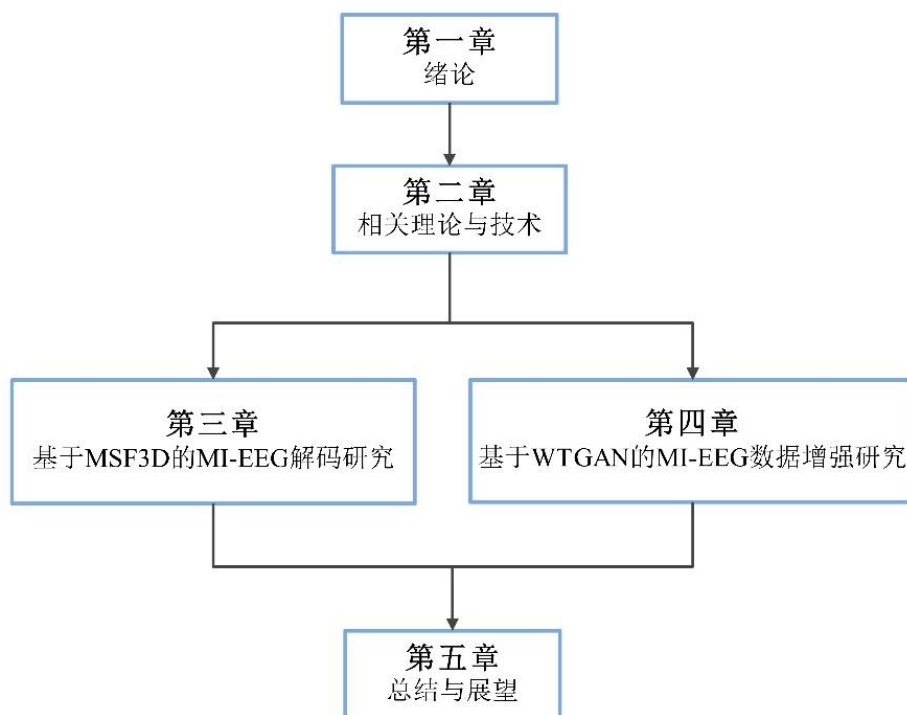


图 1.2 论文结构图

第 1 章：绪论。本章首先论述了 MI-BCI 技术的研究背景与意义，其次详细介绍了 MI-EEG 解码和数据增强算法的研究现状，接着通过分析指明该领域目前存在的问题，然后由问题导向本文的主要研究内容，最后给出文章行文顺序和组织结构。

第 2 章：相关理论与技术。本章首先论述了 MI-EEG 的产生机理和特点，接着介绍了 MI-EEG 涉及的特殊现象及解码任务和数据增强应用最广泛的两种模型原理。

第 3 章：基于 MSF3D 的 MI-EEG 解码研究。本章首先论述了提出 MSF3D 的原因和构建思路，接着分模块介绍各自的具体作用和训练流程，最后为了验证 MSF3D 的分类性能，在公开实验数据集上进行对比，消融和留一交叉验证等一系列实验并作结果分析和原因解释。

第 4 章：基于 WTGAN 的 MI-EEG 数据增强研究。本章首先论述了提出 WTGAN 的原因和构建思路，接着分模块介绍各自的具体作用和训练流程，最后为了验证 MSF3D 的数据增强效果，用多种维度进行分析，论证数据增强的有效性和在分类任务中的贡献。

第 5 章：总结与展望。本章主要对全文内容进行总结，并对后续可能的研究方向进行合理展望。

第 2 章 相关理论与技术

2.1 MI-EEG 原理及特点

大脑皮层所产生的脑电信号通常可分为诱发性脑电信号和自发性脑电信号两类。诱发性脑电信号是指在外部刺激（如图像、声音、闪光、按键等）作用下，大脑对特定事件做出反应所产生的电活动。它通常是时间锁定的，即信号在某个特定事件发生后的数十至数百毫秒内出现特定波形。诱发性脑电信号广泛用于被动型 BCI 系统，如 P300 拼写器、视觉诱发电位系统等。这类系统依赖于外界刺激的设计，用户只需进行被动感知，不需要主观意图调动，因此通常训练门槛较低。而自发性脑电信号是指在无外部特定刺激的情况下，大脑自然产生的神经活动电信号。这种信号通常反映的是个体的内在状态，如注意力水平、情绪状态、意识状态等。由于自发性脑电信号具有的这些特点，其广泛用于运动想象、情绪识别、注意力监测等任务中。这类信号的获取不依赖任何视觉、听觉或其他感官刺激，仅通过用户自身的想象或认知活动即可激发，非常适合构建主动型 BCI 系统。

在 MI-EEG 构建的 BCI 系统中，受试者需有意识地想象自己执行某项肢体动作，如手部或脚部的移动，而实际并不会发生肌肉收缩或运动输出。这一过程完全是在认知层面进行的，是一种纯粹由神经系统驱动的内部运动表征机制。神经生理学研究发现，真实运动与运动想象之间在皮层电位表现上具有显著相似性，主要源于两者在神经调控通路中存在共享机制，例如都涉及感觉运动皮层（SMC）和辅助运动区（SMA）等区域的活跃。换句话说，尽管运动想象过程中并未发生实际动作，但由于其激活的大脑区域与实际运动相重叠，因此仍会引起大脑皮层电信号的显著变化，表现为特定频段的节律活动增强或抑制。正因如此，MI-EEG 成为 BCI 系统中广泛应用的一种控制信号来源，尤其在无需外部感官刺激的 BCI 应用中具有独特优势。

MI-EEG 具有如下特点：

（1）具有多通道协同特性，各个电极通道之间并非相互独立，而是存在显著的空间协同关系。深入挖掘通道间的互信息与时空依赖性，对于提升脑电信号的特征提取与分类精度具有重要意义；

（2）具有明显的非线性、非平稳特点：大脑是非线性系统，由于人类的认知、情绪和注意状态在不断变化，导致脑电活动呈现出时间上的动态波动。此外，个体对外部环境刺激的适应性也使得脑电节律呈现复杂多变的特性；

(3) 信号微弱，极易受干扰，一方面易受到如心电、肌电、眼电伪迹等生理电信号干扰，另一方面也会受到采集过程中的电磁信号等物理性干扰。多种干扰叠加造成 MI-EEG 信噪比低，获取难度大；

(4) 频域信息丰富且结构清晰，频率范围通常在 1~100 Hz 之间。根据不同的频段，脑电活动被划分为五类典型波形： δ (Delta) 波、 θ (Theta) 波、 α (Alpha) 波、 β (Beta) 波和 γ (Gamma) 波，每种波形对应不同的认知与生理状态。如表 2.1 所示。在 MI-EEG 的频域特征中， α 和 β 频段被认为与运动想象活动密切相关。

表 2.1 MI-EEG 的频段划分

模式	频率	大脑状态
Delta(δ)	1~3Hz	醉酒、深度睡眠状态
Thete(θ)	4~7Hz	困倦、轻度睡眠状态
Alpha(α)	8~13Hz	闭上眼睛，放松状态
Beta(β)	14~30Hz	积极思考、专注状态
Gamma(γ)	> 30Hz	高度警觉、焦虑状态

2.2 事件相关同步/去同步现象

肢体动作的执行，甚至是肌肉的微弱收缩，都会激活大脑中对应的感觉运动皮层区域。事实上，即使没有实际动作，仅在大脑中进行运动的准备或想象，也足以引发这些区域的神经振荡活动，这一现象被称为感觉运动节律 (Sensory Motor Rhythm, SMR)。SMR 主要集中在 α 和 β 频段，其中振荡强度的上升被称为事件相关同步 (Event-Related Synchronization, ERS)，而强度下降则被称为事件相关去同步 (Event-Related Desynchronization, ERD)。研究表明，在进行运动想象任务时，个体大脑特定区域内的 α 和 β 波段通常会出现 ERD 现象，而处于放松状态时则更易诱发 ERS 响应^[59]。

此外，大脑在处理不同肢体运动想象时，其激活的皮层区域是有明确分工的。例如，左手与大脑右侧 C4 区域相关，而右手则对应左侧 C3 区域。如图 2.1 所示，当个体想象左手运动时，C4 区域会出现显著的 ERD 变化；相应地，想象右手运动时，则会在 C3 区域观察到类似的去同步现象。相比之下，双脚的运动想象主要由 Cz 区域控制，但由于左右脚在皮层上的功能区域非常接近，因此在脑电图中往往难以区分。值得注意的是，皮层区域的面积在一定程度上决定了该区域所控制肢体的灵活性和信号可识别度。通常而言，像双手、舌头或双脚这类灵活部位，其在皮层中所占区域较大，尤其是左右手，因此更易在 EEG 信号中展现出明显的区分特征。

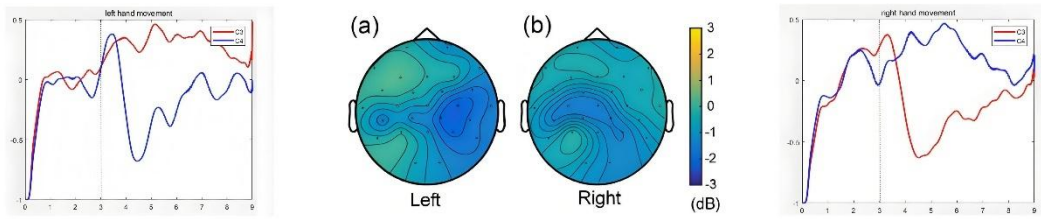


图 2.1 左/右手运动想象时 ERD/ERS 现象

2.3 CNN 基本原理

卷积神经网络（Convolutional Neural Network, CNN）是一种专门用于处理具有类似网格结构数据的深度学习模型，比如图像（2D 像素网格）或时间序列信号（1D 结构），也常被用于 EEG 的处理与分类。CNN 采用了一种高效的特征提取机制，其核心设计包含两个关键特性：局部感知和参数共享。不同于传统神经网络的全连接方式，CNN 中的每个神经元仅与输入数据的局部区域相连，这种结构显著降低了模型的复杂度。同时，同一卷积核在整个输入数据上共享参数，进一步减少了需要训练的权重数量。

在特征提取过程中，CNN 通过卷积运算逐层捕获数据的空间特征，配合非线性激活函数实现复杂的特征变换。这种架构能够自动学习从低级到高级的层次化特征表示。如图 2.2 所示，典型的 CNN 由多个功能互补的组件构成：接受数据的输入层，负责特征提取的卷积层，进行特征压缩的池化层，实现分类决策的全连接层，以及引入非线性因素的输出层。这些组件的协同工作使得 CNN 在保持计算效率的同时，具备强大的特征学习能力。

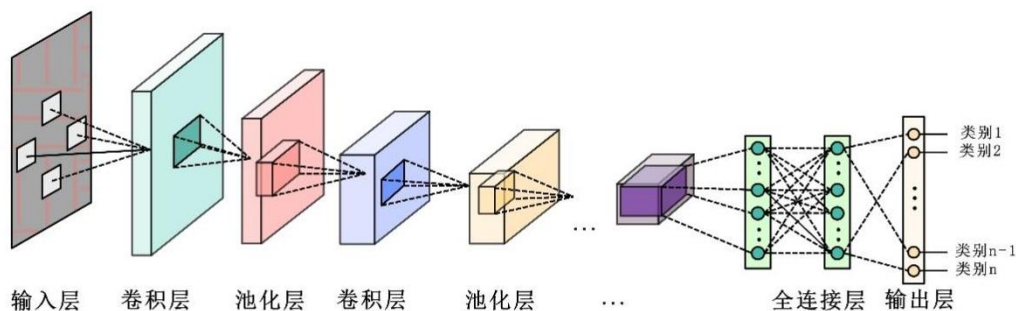


图 2.2 CNN 基本结构图

输入层是 CNN 的数据入口，输入层不包含可训练参数，但直接影响网络对原始数据的理解方式，是特征提取流程的起点。现代网络框架通常将预处理步骤集成在输入层实现。在 MI-EEG 解码任务中，输入层则主要是将原始脑电信

号处理为 CNN 能接受的形式，大多是传统的 2D 表示，即二维数组。两个维度分别代表时间序列的样本点和不同的信号通道。

卷积层是 CNN 的核心组成部分，其核心思想是利用局部感受野和权重共享来提取输入数据的局部特征。每个卷积核在输入图像上滑动，进行局部区域的点积运算，生成对应的特征图，图 2.3 以二维卷积为例。通常卷积层有多种作用不同的卷积核，例如，一个检测水平边缘的卷积核会在图像中高亮显示水平方向的亮度变化区域。多个不同的卷积核可以并行工作，提取边缘、纹理或是颜色渐变等不同的特征。由于卷积核的权重在整个图像上共享，CNN 相比全连接网络大幅减少了参数量，同时保留了空间信息，使其特别适合处理图像、语音或 EEG 等具有局部相关性的数据。

池化层紧接着卷积层，主要用于降维和增强特征鲁棒性，通过局部区域的下采样，如图 2.4 所示，用一个 2×2 窗口减少数据量，同时保留关键特征。池化方式一般有两种：最大池化和平均池化。最大池化选取区域内的最大值，突出显著特征并赋予模型平移不变性即目标微小移动不影响识别；平均池化则计算区域均值，平滑特征响应。池化层能有效减少计算量、防止过拟合，并通过逐步降低特征图分辨率，帮助网络聚焦于更高级别的语义信息，而非细节位置。

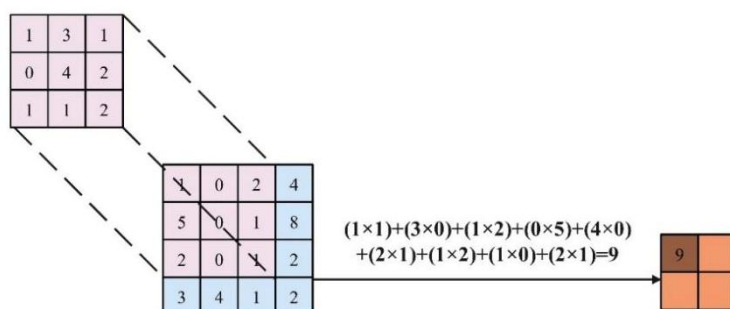


图 2.3 二维卷积示意图

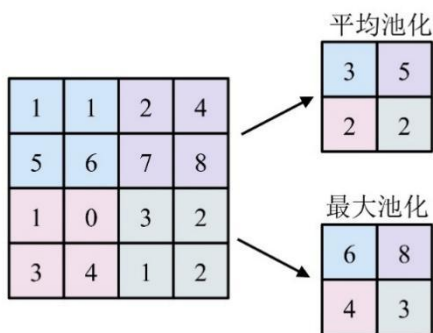


图 2.4 平均池化与最大池化示意图

全连接层通常位于 CNN 的末端，承担着从特征提取到最终决策的关键转换。该层首先将前层输出的多维特征矩阵展开为连续的一维特征向量，随后通过密集连接实现全局特征的综合分析。其核心机制在于建立当前层所有神经元与前一层的全连接拓扑，每个连接都配有可训练权重参数，这些参数在训练过程中不断优化以捕捉特征间的深层关联。这种设计使网络能够整合空间和通道维度的全部信息，最终输出适用于分类或回归任务的高层语义表示。全连接层的参数规模通常占网络总参数的主要部分，体现了从局部特征到全局决策的转换过程。

输出层是 CNN 的最终决策层，其作用是将前面各层提取的抽象特征映射为具体任务结果。对于分类任务，通常采用 Softmax 函数，输出每个类别的概率分布。同时通过损失函数如交叉熵、均方误差等与真实标签对比，也可以为整个网络提供反向传播的梯度起点。

CNN 中的激活函数是为了引入非线性变换，使网络能够学习复杂的特征模式。以 ReLU 为代表的激活函数通过其非线性特性实现特征选择和过滤，同时保持梯度稳定流动，有效解决了深层网络的梯度消失问题。这种非线性转换机制与卷积层的线性滤波形成功能互补，共同构建了 CNN 强大的特征提取能力，使网络能够逐层抽象从边缘到语义的高级特征。常用的激活函数有 Softmax、logSoftmax 和 SNL，如图 2.9 所示：

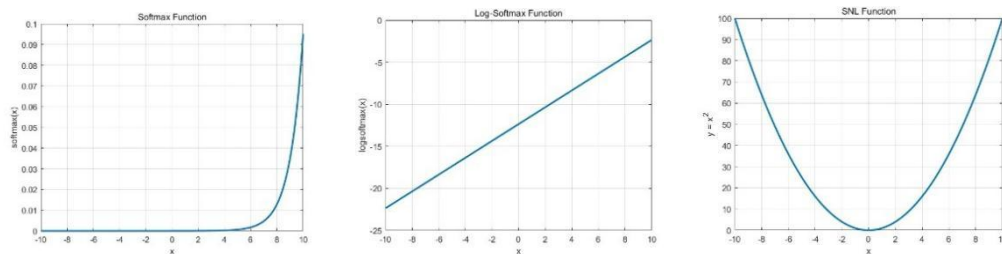


图 2.5 Softmax 函数、logSoftmax 函数和 SNL 函数图像

2.4 GAN 基本原理

生成对抗网络（Generative Adversarial Network, GAN）由 Goodfellow 等人^[60]于 2014 年提出，是一种基于博弈论思想的无监督学习框架，旨在通过对抗性训练方式学习数据分布并生成具有高质量的拟合样本。如图 2.6 所示，图中是一个简单的 GAN 网络结构，它由两个主要组成部分构成：生成器 G （Generator）与判别器 D （Discriminator）。其中，生成器负责从噪声空间中生成与真实数据分布相似的样本，而判别器的任务是区分输入数据是真实样本还是由生成器伪造的样本。

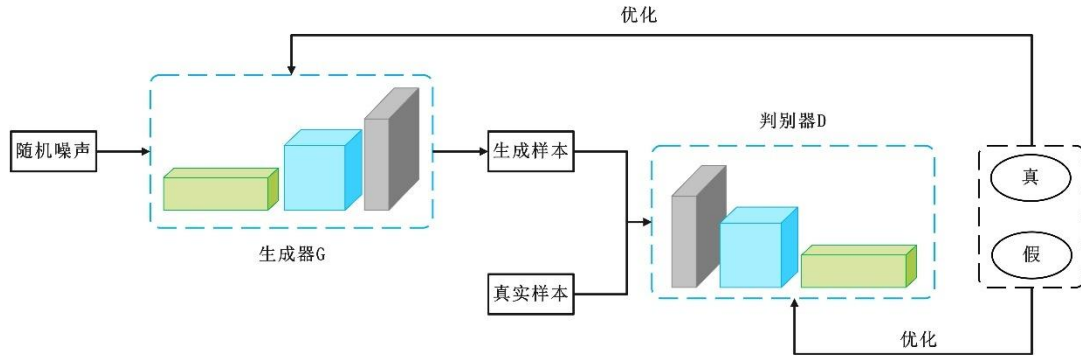


图 2.6 生成对抗网络结构图

在训练过程中，生成器与判别器通过零和博弈相互竞争：生成器试图欺骗判别器，而判别器则不断提高其识别伪造样本的能力，在训练过程中将判别器的输出作为反馈优化生成器和判别器的这种能力。该机制能促使生成器逐步生成更具真实性的样本，从而逼近真实数据的分布。

GAN 在图像生成、语音合成、数据增强等领域表现出色。其最大的优势在于能够捕捉复杂的高维数据分布并生成多样化、高质量的样本，在样本不足的任务中尤为重要。在 MI-EEG 解码任务中，GAN 被广泛用于增强数据集、解决数据不平衡问题、提高分类器的泛化能力。

GAN 的目标函数表达式如下：

$$\min_G \max_D V(D, G) = E_{x \sim p_r(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (2.1)$$

其中 x 代表真实数据， $p_r(x)$ 代表真实数据的分布， z 代表输入的随机噪声， $p_z(z)$ 代表生成器输入噪声的分布，一般为标准正态分布或 $[0,1]$ 均匀分布。

在训练过程中，生成器 G 首先接受输入噪声 z ，输出虚假样本 $G(z)$ ，虚假样本和真实样本的维数一致。接着，真实样本 x 和虚假样本 $G(z)$ 同时被馈送到判别器 D 中，经过判别后会针对每一个样本输出一个 $0 \sim 1$ 之间的实数。该实数越接近于 1，则表示输入样本越接近于真实样本，反之越趋近于 0 则说明输入样本与真实样本差距越大。最终目标是让生成器学习真实数据的分布 $p_r(x)$ ，使得判别器无法区分真实数据和生成数据 $p_G(x)$ 。

目标函数的前一项 $E_{x \sim p_r(x)} [\log D(x)]$ 表示判别器对真实样本判别结果的期望值，后一项 $E_{z \sim p_z(z)} [\log (1 - D(G(z)))]$ 则代表判别器对生成样本判别结果的期望值。于是，当 $x \sim p_r(x)$ ，也即输入样本为真实样本时， $D(x)$ 取值为 1， $E_{x \sim p_r(x)} [\log D(x)]$ 取到最大值，代表判别器识别真实样本的能力最强；而当 $z \sim p_z(z)$ ，也即输入的是随机噪声时， $D(G(z))$ 取值为 0， $E_{z \sim p_z(z)} [\log (1 - D(G(z)))]$ 取到最大值，代表判别器能够把生成样本识别为假样本。

为了保证 GAN 的训练效果，我们定义优化函数： $V(G, D) = \log(D(x)) +$

$\log(l - D(G(z)))$ 。其中 $D_G^* = \arg\max_D V(D, G)$, D_G^* 指最优判别器。当 D_G^* 固定不变时, 就能得到最佳的生成器 $G_D^* = \arg\min_G \left(\arg\max_D V(D, G) \right)$ 。于是 GAN 训练过程就转化为如公式 2.1 的最值求解问题。

为了度量 GAN 生成的数据分布与真实数据分布之间的距离, GAN 的提出者 Goodfellow 等人引入了 KL 散度以及 JS, 其表达式如下:

$$KL(p_r \parallel p_G) = E_{x \sim p_r} \log \frac{p_r}{p_G} \quad (2.2)$$

$$JS(p_r \parallel p_G) = \frac{1}{2} KL\left(p_r \parallel \frac{p_r + p_G}{2}\right) + \frac{1}{2} KL\left(p_G \parallel \frac{p_r + p_G}{2}\right) \quad (2.3)$$

Goodfellow 等人证明了 KL 散度能衡量两个概率分布的相似性, 但其问题在于当 $p_r(x)$ 存在但 $p_z(z)$ 极小或趋近 0 时, $\log \frac{p_r}{p_z}$ 会趋于无穷大, 导致 KL 极高。反之, 当 $p_r(x)$ 很小或趋近 0 而 $p_z(z)$ 存在时, KL 值接近 0, 这导致梯度更新很慢。

JS 散度在 GAN 训练中的问题则是: 当两个分布几乎不重叠, 即 p_r 和 p_G 相距较远时, JS 散度几乎恒定为 $\log 2$, 导致梯度为 0, 生成器无法得到有效的学习信号, 训练停滞。只有当两个分布有较大的重叠区域时, JS 散度才会产生足够的梯度, 推动 GAN 更新。这种情况导致 GAN 训练早期生成器难以获得有效梯度, 学习过程停滞, 难以收敛, 即梯度消失问题。

2.5 本章小节

本章首先指出研究对象 MI-EEG 的产生机理和数据特点, 接着介绍 MI-EEG 分类任务的理论依据, 即运动想象过程中出现的事件相关同步/去同步现象。随后则分别对第 3 章和第 4 章采用方法的基本原理进行具体阐述。

第3章 基于 MSF3D 的 MI-EEG 解码研究

3.1 引言

通过 1.2.4 节基于深度学习的 MI-EEG 解码方法的论述我们意识到当前的 MI-EEG 解码任务仍然存在以下问题：

1. MI-EEG 的传统 2D 表示忽略了电极通道之间的空间拓扑依赖性，此外，使用传统的 2D 卷积核无法完全反映附近电极通道之间的内部联系。因此，MI-EEG 分类系统的性能受到影响。

2. EEG 电极通道选择的目的是确定最具鉴别力的 EEG 节点，而运动区域 C3、C4 和 Cz 是位于大脑运动区域的三个常见的手动选择通道。但是这些通道可能会因为个体差异或 MI-EEG 任务的不同而表现不佳。

3. 受试者通常在某个时间集中注意力，但在其他时间分心，并且不同的受试者在试验中的不同时间集中注意力。因此，当受试者专注于试验而忽略其他切片时，强调 EEG 时间切片对于成功的 EEG 分析是必要的。因此，以往那些试图从全部时间切片中充分提取时间不变的高级时间特征来编码时间信息，以便对不同主题进行更高、更稳健的分类是有问题的。

针对上述的问题，本章提出一种基于多尺度特征 3D 卷积神经网络模型。

3.2 MSF3D 网络模型结构

本章节提出的多尺度特征的 3D 卷积神经网络（Multi scale feature 3D convolutional neural network, MSF3D）模型结构如图 3.1 所示，该网络模型具体由 3D 空间注意力模块、多尺度时间注意力模块和分类模块组成。3D 空间注意力模块和多尺度时间注意力模块的目标是在所有大脑区域自适应地为运动想象相关的空间通道和时间采样切片分配比运动无关的更高的权重。它们可以在空间和时间域中定义 MI-EEG 的新的紧凑特征表示，并防止生物和环境伪影的影响，以提高分类性能。最后将得到的多尺度空间和时间特征进行特征组合后输入分类模块，得到提取多尺度特征的分类结果。MSF3D 各模块结构及作用将在下面各个小节具体描述。

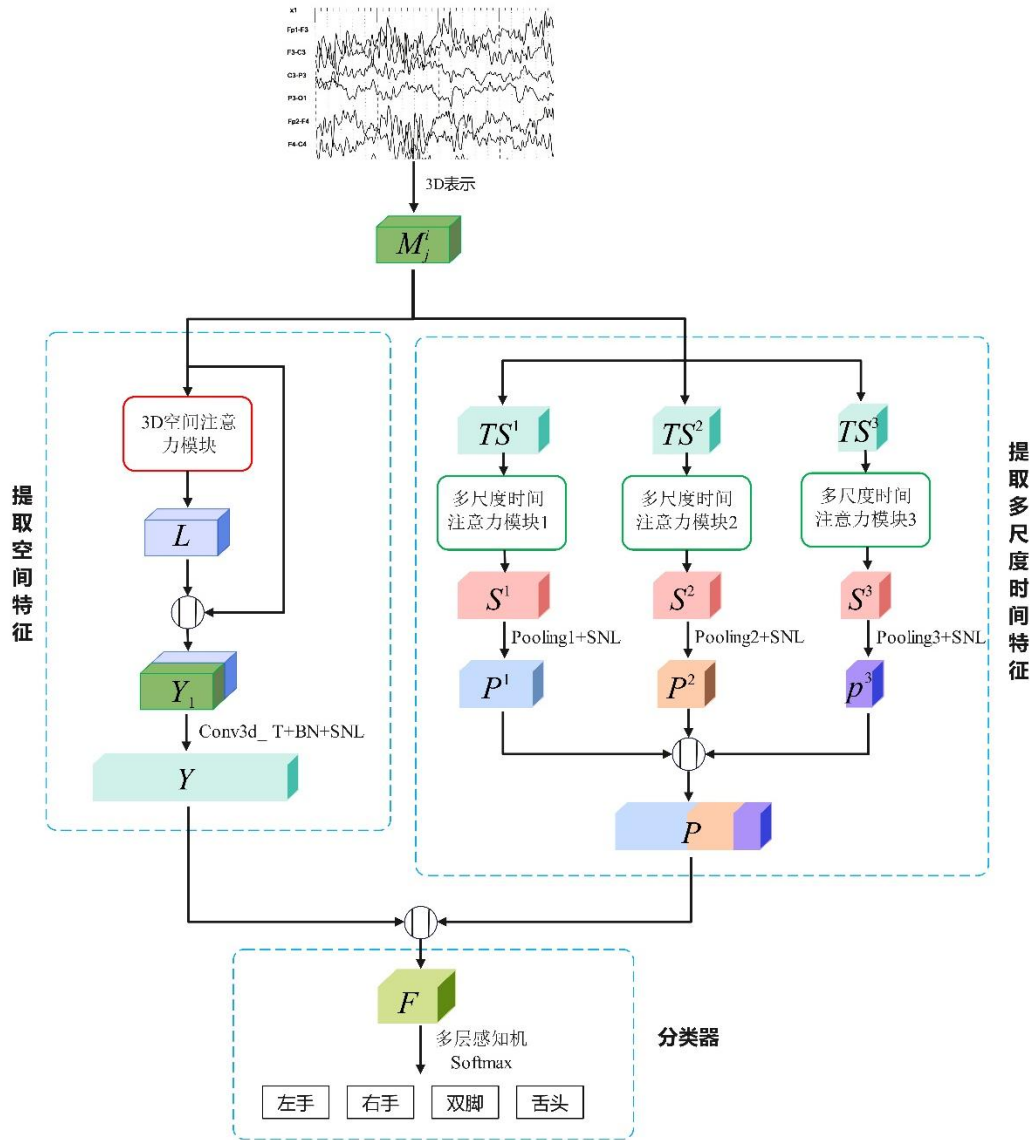


图 3.1 MSF3D 网络模型结构

3.2.1 MI-EEG 的 3D 表示

3D 表示的定义如下：设 X_j^i 为原始 MI-EEG 信号，其中 $X_j^i \in \mathbb{R}^{C \times T}$ 将第 j 个受试者的第 i 次试验表示为传统的 2D 矩阵，其中矩阵的宽度为离散时间采样点的数量（ T ），矩阵的高度为电极通道的数量（ C ）。然而，如上所述，传统的 2D 表示 X_j^i 就无法完全反映附近电极通道之间的相关性。因此，我们使用脑电图国际标准 10-20 中 EEG 电极通道的位置，并将其映射到 3D 表示的对应编号位置实现把传统的 2D 矩阵 X_j^i 扩展为 3D 张量 M_j^i 的转换，如图 3.2 所示。例如， M_j^i 中的色彩数字 k 表示它与电极通道 k 具有相同的相对位置，其时间采样值在 M_j^i 中形成为连续数据。 M_j^i 的蓝色数字 0 表示没有电极通道，其时间取样值为零。将零添加到 M_j^i 的目的是将 M_j^i 保留为 3D 立方体张量，以支持在不引入任何噪声的

情况下使用 3D 卷积。

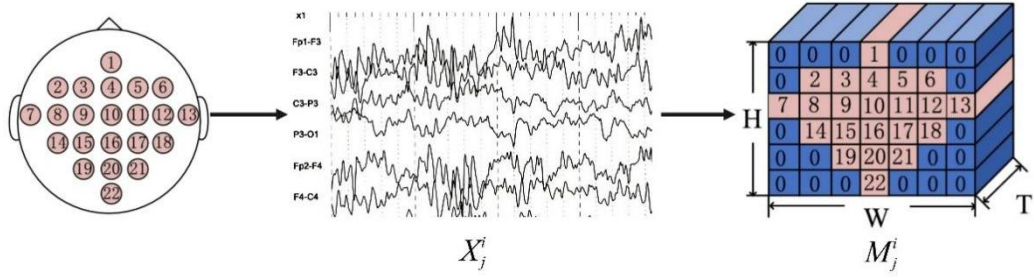


图 3.2 3D 表示示意图

这种 3D 表示不仅使用电极分布来显式保留电极通道之间的相对空间拓扑信息，而且还使用时间采样值的顺序形式来保留时间信息。因此，它很容易用于使用 3D 卷积提取时空特征。

3.2.2 3D 空间注意力模块

3D 空间注意力（3D spatial attention, 3Dsa）模块学习了 M_j^i 的新 3D 空间表示，它会自动为大多数与运动相关的通道分配较高的权重，为与运动无关的通道分配较低的权重，如图 3.3 所示，详细步骤如下：

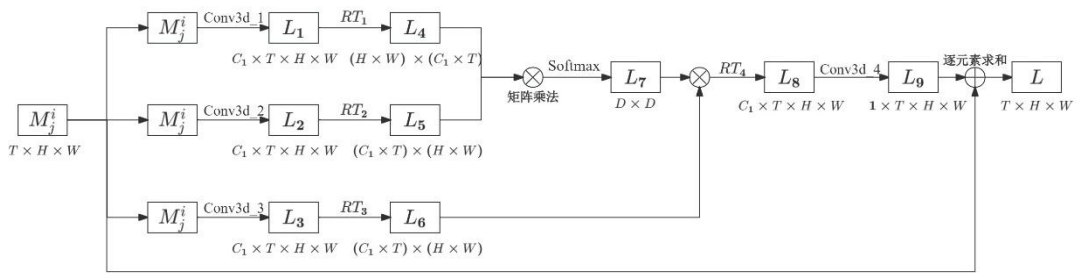


图 3.3 3D 空间注意力模块结构示意图

3D 张量 M_j^i 首先被输入到三个可分离的 3D 卷积块 $Conv3d_1$, $Conv3d_2$, $Conv3d_3$ 生成不同的 3D 空间特征块 L_1 , L_2 和 L_3 。 L_1 , L_2 和 L_3 属于 $\mathbb{R}^{C_1 \times T \times H \times W}$ 且 C_1 为 4，表示特征块的数量。接着， L_1 , L_2 和 L_3 将重构并转置（ RT_1 , RT_2 和 RT_3 ）转换为不同的大小，形如 $\mathbb{R}^{(H \times W) \times (C_1 \times T)}$, $\mathbb{R}^{(C_1 \times T) \times (H \times W)}$ 和 $\mathbb{R}^{(C_1 \times T) \times (H \times W)}$ ，用于在它们之间执行矩阵乘法（ Mat ）。然后，将 softmax 函数应用于 L_4 和 L_5 以获取空间注意力权重矩阵（ $L_7 \in \mathbb{R}^{D \times D}$ ）。

$$L_7^{ij} = \frac{Mat(L_4^i, L_5^j)}{\sum_{j=1}^D Mat(L_4^i, L_5^j)} \quad (3.1)$$

其中 D ($H \times W$) 表示 M_j^i 中的电极通道数量。 $L_7^{ij} \in L_7$ 的范围为 0 到 1，表示 M_j^i 的第 i 个和第 j 个电极通道之间的相似性权重。电极相关特性越高，权重 L_7^{ij} 就越大。

执行 L_6 和 L_7 之间的另一个矩阵乘法以获得 L_8 。 L_8 是 M_j^i 的新的基于注意力的空间表示，它通过根据空间注意力权重矩阵（ L_7 ）自适应地聚合其他电极通道的 3D 空间特征来更新每个电极通道。

一个 1×1 卷积核（*Conv3d_4*）是因为它具有强大的降低特征块维数的能力，而 1×1 卷积操作只关注特征块维数，输入数据的数量是恒定的。因此，它适用于高效地处理大量输入通道。

最后，通过将可学习参数 γ_1 乘以 L_9 ，并对原始 EEG 3D 表示 M_j^i 进行逐元素求和运算，获得了基于注意力的自适应空间特征，如下所示：

$$L = M_j^i + L_9 \times \gamma_1, L \in \mathbb{R}^{(T \times H \times W)} \quad (3.2)$$

γ_1 是一个实数，在训练所提出的架构中会自动学习乐观值。与其他空间特征提取（如 2D 卷积）相比，3D 空间注意力模块侧重于现实世界中电极通道的位置。它基于 3D 空间信息增强了有价值的运动相关特征，抑制了无用的运动无关特征，这与现实情况是更相符合的，即不同的大脑功能区域可能对不同受试者的不同运动想象任务有一定的影响。因此，通过创新地将注意力机制引入 3D 卷积，该模块实现了对现有 3D 神经网络的补充。

基于注意力的空间自适应特征和原始 EEG 3D 表示 M_j^i 通过级联组合为 $Y_1 = \{M_j^i, L\}$ 。接着对得到的 Y_1 通过 *Conv3d_T* + *BN* + *SNL* 进行处理。其中 *Conv3d_T* 是一个使用 3D 卷积提取空间特征并降低维度，用于将大小重塑为 $32 \times T \times 1 \times 1$ ，获得更紧凑的特征。但输入特征的分布可能会发生变化，数据分布的偏移会影响网络的训练^[61]。于是紧接着使用批量归一化 *BN*（Batch Normalization）^[62]和平方非线性激活函数 *SNL*（Square Non-Linearity）进一步处理，最后得到输出空间特征 Y 。

*BN*是一种归一化技术，在每一层网络的中间层进行批量统计计算，主要作用是加速训练、稳定梯度、减少内协变量偏移，从而增强特征提取能力。

（1）稳定梯度，提高训练速度

*BN*对每个小批量进行均值为 0、方差为 1 的归一化，减少梯度消失或梯度爆炸问题，使模型更容易训练。其数学表达为（假设给定一批输入 x ）：

$$\hat{x} = \frac{x - \mathbb{E}[x]}{\sqrt{\text{Var}[x] + \epsilon}} \quad (3.3)$$

其中 $\mathbb{E}[x]$ 和 $\text{Var}[x]$ 是小批量的统计均值和方差， ϵ 是防止除零的小值。

（2）减少内协变量偏移，增强特征稳定性

由于深度神经网络的中间层分布不断变化（即“内协变量偏移”），*BN*通过强制归一化，使得每层的输入分布更稳定，提高模型对不同数据的适应能力。

（3）增强特征的分布均匀性

BN 使得特征在所有神经元上具有相似的方差，防止某些特征通道对学习过程的影响过大，提高网络提取多样性特征的能力。

(4) 具有一定的正则化效果

BN 类似于 $Dropout$ 的正则化效果，能减少过拟合，提高泛化能力。

平方非线性激活函数 SNL ($f(x) = x^2$) 是一种非线性变换方法，在信号处理、神经网络、EEG 信号分析等任务中常用。相比 $ReLU$ 、 $Sigmoid$ 等激活函数，平方非线性激活函数具有不同的特征提取机制。主要作用是增强信号幅度，放大特征差异。例如，在 EEG 处理或图像处理任务中，平方非线性会增强输入特征的对比度，使得高幅值信号更加明显，而小幅值信号被进一步减小从而放大重要特征，过滤掉较小的无关信息。

3D 空间注意力的详细网络参数如表 3.1 所示。

表 3.1 3D 空间注意力网络参数

Input	Layer	Output	Feature Blocks	Kernel	Stride	Padding
$M_j^i(1,1125,6,7)$	Conv3d_1	$L_1(4,1125,6,7)$	4	(3,3,3)	(1,1,1)	(1,1,1)
$L_1(4,1125,6,7)$	RT_1^1	$L_4(42,4500)$	/	/	/	/
$M_j^i(1,1125,6,7)$	Conv3d_2	$L_2(4,1125,6,7)$	4	(3,3,3)	(1,1,1)	(1,1,1)
$L_2(4,1125,6,7)$	RT_2^1	$L_5(4500,42)$	/	/	/	/
$L_4(42,4500) L_5(4500,42)$	MNS^2	$L_7(42,42)$	/	/	/	/
$M_j^i(1,1125,6,7)$	Conv3d_3	$L_3(4,1125,6,7)$	4	(3,3,3)	(1,1,1)	(1,1,1)
$L_3(4,1125,6,7)$	RT_3^1	$L_6(4500,42)$	/	/	/	/
$L_6(4500,42) L_7(42,42)$	MRT^3	$L_8(4,1125,6,7)$	4	/	/	/
$L_8(4,1125,6,7)$	Conv3d_4	$L_9(1,1125,6,7)$	1	(1,1,1)	(1,1,1)	(0,0,0)
$L_9(1,1125,6,7)$	EWS^4	$L(1,1125,6,7)$	1	/	/	/

3.2.3 多尺度时间注意力模块

从脑电图 3D 表示中得到的 3D 张量 M_j^i 被切割成三个时间片 ($TS^i \in \mathbb{R}^{32 \times T_i \times N_i}, i = \{1,2,3\}$, T 是每个时间片的长度, N_i 是沿时间维度具有尺度 i 的时间片的数量), 并将它们分别送到设计的多尺度时间注意力 (Multi scale temporal attention, Msta) 模块进行时间特征提取。

为了更好地提取每个时间片内的时不变高级特征, 我们利用注意力机制为不同的时间片分配自适应权重, 以满足 EEG 分析的要求, 其中不同的受试者专注于不同的时间段。与之前的方法相比, 多尺度时间注意力模块不依赖于受试者或任务, 对新的受试者和任务也更稳健。

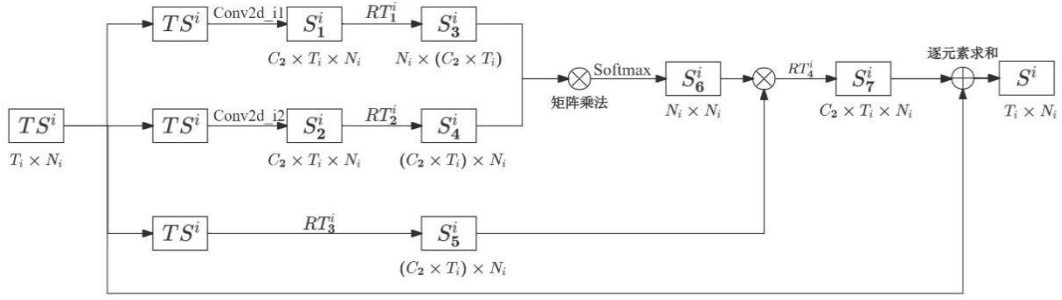


图 3.4 多尺度时间注意力模块结构示意图

以第 i 个的多尺度时间注意力模块处理为例，如图 3.4 所示， TS^i 被送到两个单独的 2D 卷积（ $Conv2d_i1$ 和 $Conv2d_i2$ ）中，以生成初始时间特征 S_1^i 和 S_2^i ，它们的大小都是 $\mathbb{R}^{C_2 \times T_i \times N_i}$ 。然后，对 S_1^i, S_2^i 和 TS^i 进行重构和转置（ RT_1^i, RT_2^i 和 RT_3^i ）生成 S_3^i, S_4^i 和 S_5^i ，其大小为 $\mathbb{R}^{N_i \times (C_2 \times T_i)}$ ， $\mathbb{R}^{(C_2 \times T_i) \times N_i}$ 和 $\mathbb{R}^{(C_2 \times T_i) \times N_i}$ ，用于在它们之间执行矩阵乘法（ Mat ）。然后，将 softmax 函数应用于 S_3^i 和 S_4^i 以获取时间注意力权重矩阵（ $S_6^i \in \mathbb{R}^{N_i \times N_i}$ ）。

$$S_6^i(k, l) = \frac{Mat(S_3^i(k), S_4^i(l))}{\sum_{j=1}^D Mat(S_3^i(k), S_4^i(l))} \quad (3.4)$$

其中 N_i 是尺度 i 下的时间片数。 $S_6^i(k, l)$ 是试验中第 k 个和第 l 个时间片之间的相似性权重。 S_6^i 专注于特定的与运动相关的时间切片，这些切片比其他与运动无关的切片更容易区分。 $S_6^i(k, l)$ 越大，第 k 个和第 l 个时间片彼此越相似。然后，计算所有 EEG 时间切片的加权和（ S_5^i ）来学习基于注意力的时间表示（ S_7^i ）。最后，另一个可学习参数 γ_2^i 与 S_7^i 相乘，并与 TS^i 进行逐元素求和运算，以获得尺度 i 下的基于注意力的自适应时间特征，如下所示：

$$S^i = TS^i + S_7^i \times \gamma_2^i, S^i \in \mathbb{R}^{(32 \times T_i \times N_i)} \quad (3.5)$$

接下来，将第 i 个基于注意力的时间自适应表示（ S^i ）馈送到具有平方非线性激活函数 SNL 的平均池化层中，以并行聚合时间维度的特征。它进一步减少了时间维度，将低级特征转换为高级抽象特征（ P^1, P^2 和 P^3 ），这些特征被连接为多尺度时间特征 $P = \{P^1, P^2, P^3\}, P \in \mathbb{R}^{(32 \times 1 \times N)}$ 。

多尺度时间注意力模块的详细网络参数如表 3.2 所示。

表 3.2 多尺度时间注意力模块网络参数

Input	Layer	Output	Feature Blocks	Kernel	Stride
$TS^1(32, 25, 45)$	Conv2d_11	$S_1^1(32, 25, 45)$	32	(1, 1)	(1, 1)
$S_1^1(25, 25, 45)$	RT_1^1	$S_3^1(45, 800)$	/	/	/
$TS^1(32, 25, 45)$	Conv2d_12	$S_2^1(32, 25, 45)$	32	(1, 1)	(1, 1)
$S_2^1(32, 25, 45)$	RT_2^1	$S_4^1(800, 45)$	/	/	/

续表 3.2

Input	Layer	Output	Feature Blocks	Kernel	Stride
$S_3^1(45,800)S_4^1(800,45)$	MNS^2	$S_6^1(45,45)$	/	/	/
$TS^1(32,25,45)$	RT_3^1	$S_5^1(800,45)$	/	/	/
$S_5^1(800,45)S_6^1(45,45)$	MRT^3	$S_7^1(32,25,45)$	32	/	/
$TS^1(32,25,45)$ $S_7^1(32,25,45)$	EWS^4	$S^1(32,25,45)$	32	/	/
$TS^2(32,25,45)$	Conv2d_21	$S_1^2(32,25,45)$	32	(1,1)	(1,1)
$S_1^2(32,45,25)$	RT_1^2	$S_3^2(25,1440)$	/	/	/
$TS^2(32,45,25)$	Conv2d_22	$S_2^2(32,45,25)$	32	(1,1)	(1,1)
$S_2^2(32,45,25)$	RT_2^2	$S_4^2(1440,25)$	/	/	/
$S_3^2(45,1440)S_4^2(1440,45)$	MNS^2	$S_6^2(25,25)$	/	/	/
$TS^2(32,45,25)$	RT_3^2	$S_5^2(1440,25)$	/	/	/
$S_5^2(1440,45)S_6^2(25,25)$	MRT^3	$S_7^2(32,45,25)$	32	/	/
$TS^2(32,45,25)$ $S_7^2(32,45,25)$	EWS^4	$S^2(32,45,25)$	32	/	/
$TS^3(32,75,15)$	Conv2d_31	$S_1^3(32,75,15)$	32	(1,1)	(1,1)
$S_1^3(32,75,15)$	RT_1^3	$S_3^3(15,2400)$	/	/	/
$TS^3(32,75,15)$	Conv2d_32	$S_2^3(32,75,15)$	32	(1,1)	(1,1)
$S_2^3(32,75,15)$	RT_2^3	$S_4^3(15,2400)$	/	/	/
$S_3^3(15,2400)S_4^3(2400,15)$	MNS^2	$S_6^3(15,15)$	/	/	/
$TS^3(32,75,15)$	RT_3^3	$S_5^3(2400,15)$	/	/	/
$S_5^3(2400,15)S_6^3(15,15)$	MRT^3	$S_7^3(32,75,15)$	32	/	/
$TS^3(32,75,15)$ $S_7^3(32,75,15)$	EWS^4	$S^3(32,75,25)$	32	/	/

3.2.4 分类模块

将 3.2.2 节获得的 MI-EEG 空间特征 Y 和 3.2.3 节获得的多尺度时间特征 P 进行特征融合，这样获得的全局时空特征 F 就既具有时频特征，同时也具有空间特征。本小节采用的分类模型是先通过多层感知机获取分类信息，再经由典型的 Softmax 函数将特征转换为多个标签的条件概率来用于二分类或多分类任务。

3.3 数据集及实验设置介绍

为了验证提出模型的性能表现，本章节选用经过历代学者验证的三个公开 MI-EEG 数据集：BCICIV_2A（2A）^[63]、BCICIV_2B（2B）^[64]和 High Gamma Dataset（HGD），下面分别对各数据集及实验设置做详细介绍。

3.3.1 BCICIV_2A 数据集

BCI Competition IV 2A (BCICIV_2A) 数据集是脑机接口领域的一个经典公开数据集，由柏林工业大学提供，广泛用于评估和比较脑电信号分类算法的性能。该数据集特别适用于运动想象任务的研究，具有良好的实验设计和数据质量，已被大量研究引用和验证。

该数据集包含来自 9 名编号为 A01-A09 健康被试者的 EEG 数据，每位被试参与了两个不同时段实验。任务类型包括左手、右手、双脚和舌头的运动想象，共四个类别。这两个时段的数据分别用于模型训练与测试。被试在每一轮次实验执行 48 次运动想象任务，每个类别 12 次，即每个被试两个时段合计执行 576 次运动想象任务，其中训练集包含 288 次的数据，测试集包含 288 次的数据。数据采用 22 个通道的 EEG 电极采集，采样率为 250 Hz，同时提供了 3 个 EOG (Electro-Oculogram, EOG) 通道用于眼动伪迹检测。电极分布如图 3.5 所示，其中图 3.5 (a) 显示了 MI-EEG 信号的电极布置，图 3.5 (b) 则为眼动电信号 EOG 的电极分布图。实验中所使用的采集设备采样频率为 250 Hz，所采集的信号经过 0.5~100 Hz 的带通滤波处理，同时应用 50 Hz 陷波滤波以抑制工频干扰。

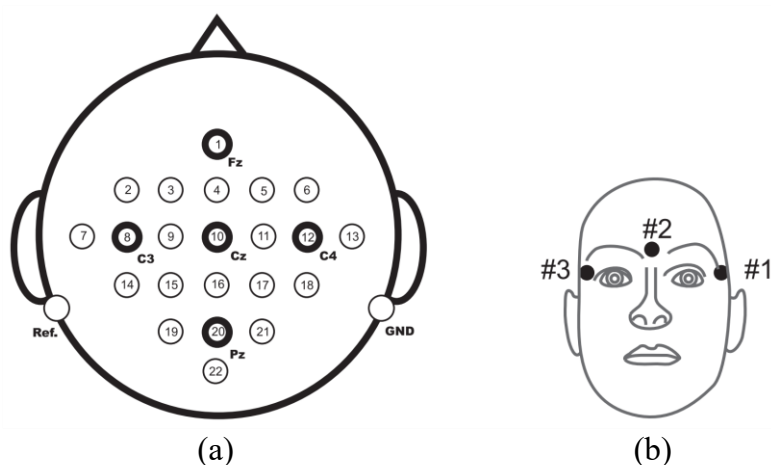


图 3.5 数据集 BCICIV_2A 电极通道空间分布

实验流程如图 3.6 所示。被试端坐于扶椅上，正对电脑屏幕。实验开始时 ($t=0\text{ s}$)，屏幕中央出现一个固定的十字，伴随一声提示音；两秒后 ($t=2\text{ s}$)，屏幕上呈现一个箭头提示，持续约 1.25 s，用于指示本轮的运动想象类别：箭头朝左、右、下、上分别对应左手、右手、双脚及舌头的运动想象。接下来，被试根据指令执行相应想象任务，直到提示十字消失 ($t=6\text{ s}$)。每轮任务结束后进入短暂休息阶段，整个实验周期约为 8 秒。

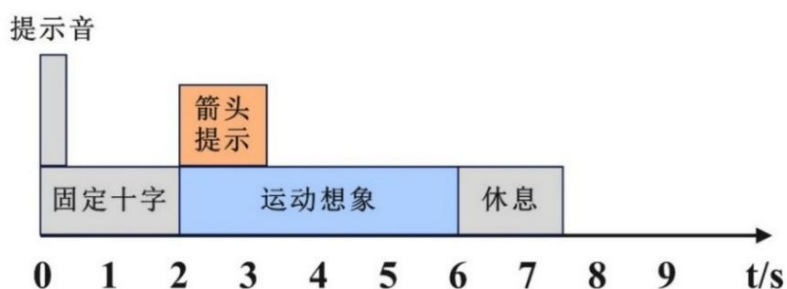


图 3.6 BCICIV_2A 实验流程图

本文实验中使用的是每个实验流程中 2~6.5 秒的 MI-EEG 数据，并对数据执行 4~38Hz 带通滤波以去除 EOG、肌电及工频干扰，得到实验数据样本。

3.3.2 BCICIV_2B 数据集

BCI Competition IV 2B (BCICIV_2A)，该数据集同样是由柏林工业大学提供，是 MI-EEG 分析领域的标准公开数据集之一。数据集共包含 9 名编号为 B01-B09 健康受试者的运动想象任务脑电记录，任务类型为左手与右手的二分类运动想象。

脑电信号通过 3 个通道（C3、C4 和 Cz）进行采集，电极位置如图 3.7 所示，同理也记录了 3EOG 通道以用于伪迹检测。采样频率为 250 Hz，数据预处理步骤同 BCICIV_2A。每位被试共完成 5 个实验时段，其中包含 3 个训练时段和 2 个评估时段，前两个时段为无视觉反馈的 MI-EEG 数据，后三个时段是包含视觉反馈的 MI-EEG 数据。无视觉反馈的每个时段实验分为 4 个轮次，被试在每一轮次实验执行 40 次运动想象任务，每个类别 20 次，即每个时段共包含 160 次运动想象任务的数据。包含视觉反馈的每个时段实验分为 6 个轮次，被试在每一轮次实验执行 20 次运动想象任务，每个类别 10 次，即每个时段共包含 120 次运动想象任务的数据

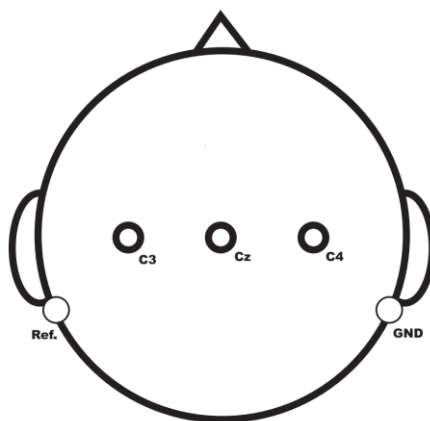


图 3.7 BCICIV_2B EEG 电极位置分布图

图 3.8 (a) 展示了无视觉反馈条件下的 MI-EEG 数据采集流程。被试端坐于距屏幕约一米处。实验起始时刻 ($t = 0\text{ s}$)，屏幕中央呈现一个固定的十字，2 秒后伴随短暂提示音，提示被试即将开始运动想象任务。在 $t = 3\text{ s}$ 时，十字会变为指向左或右的箭头，分别对应左手或右手的运动想象指令，提示持续约 1.25 秒。随后，被试根据提示开始运动想象，直至 $t = 7\text{ s}$ 十字消失，随后进入约 1.5 秒的休息阶段。每轮试验的总时长共计 9 秒。

图 3.8 (b) 为带视觉反馈条件下的实验流程示意图。起始阶段 ($t = 0\text{ s}$) 屏幕显示一个灰色笑脸图案，2 秒后伴随提示音进入任务准备阶段。接着在 $t = 3\text{ s}$ 时，屏幕呈现左右方向提示，引导被试通过想象使屏幕中的灰色笑脸向对应方向移动。若方向正确，笑脸颜色变为绿色；若错误，则变为红色哭脸，形成即时反馈。到 $t = 7.5\text{ s}$ 时，屏幕恢复为空白，并随机进入 1 至 2 秒的休息阶段。

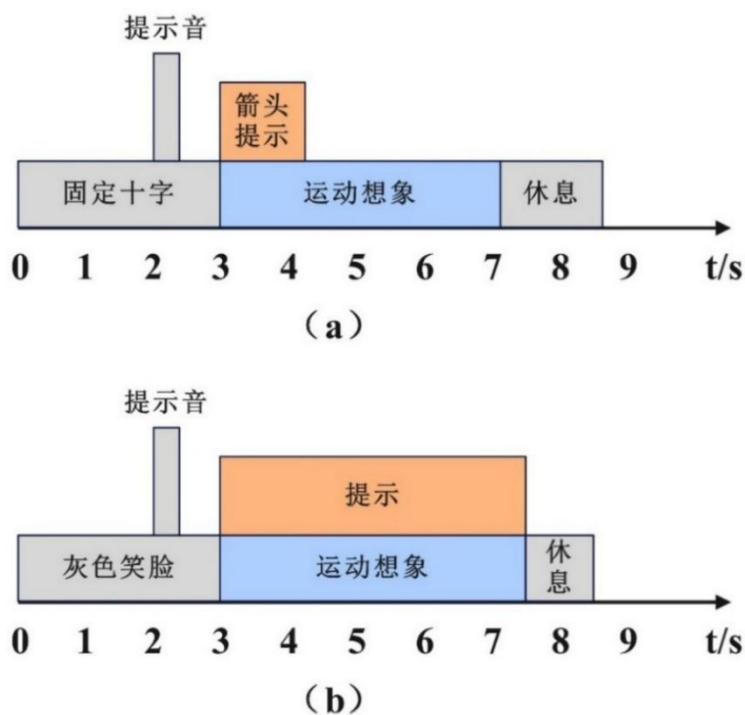


图 3.8 BCICIV_2B 实验流程图

本文实验中使用的是每个实验流程中 3~7 秒的 MI-EEG 数据，对数据执行 4~38Hz 带通滤波以去除 EOG、肌电及工频干扰，得到实验数据样本。

3.3.3 High Gamma Dataset 数据集

HGD 数据集包含来自 14 名编号为 H01-H14 受试者的 MI-EEG 数据。实验任务共设四类：左手、右手、双脚运动想象，以及静息状态。每位被试共参与 13 轮实验，总计完成约 1040 次试次，每个类别对应约 260 次。

在每轮试验中，屏幕上会呈现一个方向箭头，提示被试执行为期 4 秒的运

动想象任务。箭头方向左、右、下、上分别对应左手、右手、双脚和休息状态。被试在完成想象任务的同时，还需根据提示方向配合相应动作，如敲击手指、反复屈伸脚趾或保持放松，以激活认知过程。这些动作虽然涉及的近端肌肉活动极少，但足以帮助被试维持注意力。每次想象任务后为3至4秒的休息时间，两者交替进行，直至本轮实验结束。

脑电信号采集使用了128个通道，采样频率为500 Hz。该数据集的一个显著优势在于其较低的噪声水平，得益于高分辨率放大器、电磁干扰屏蔽以及全光隔离技术的应用。为与原数据采集者的处理方法保持一致，本文将EEG信号重新采样至250 Hz，并选取44个关键通道用于后续分析，以减少冗余。最终提取每次运动想象持续的4秒MI-EEG数据，并通过4~38 Hz带通滤波获得实验所用的数据样本。

3.3.4 实验设置

（1）实验数据设置

本章所有实验均使用单一被试数据对各模型进行训练和测试，评估指标为分类准确率，以衡量模型的特征提取效果。在2A数据集中，每位被试的第一时段MI-EEG数据作为训练集，第二时段数据用于测试。对于2B数据集，训练集由每位被试的前三个时段数据组成，其中包括两个无视觉反馈的时段和一个有视觉反馈的时段，测试集由最后两个有视觉反馈的时段数据构成。对于HGD数据集，每位被试包含1040次运动想象任务的数据，其中880次（每个类别220次）用作训练集，剩余160次（每个类别40次）作为测试集。

（2）对比方法选择

为了验证本章提出方法的分类性能和效果，选取的对比方法应该既包含传统机器学习的经典算法，如滤波器组共空间模式（FBCSP）算法和其改进算法子带共空间模式（SBCSP），也应有一些经典的CNN模型，如EEG-TCNet和其改进模型EEG-TCNet以及动态联合域适应CNN模型DJ DAN^[65]，同时也要有与本方法采取类似思路的算法，如首次纳入3D表示的M3DCNN和提取时域、空间和频域信息的HSCNN。

（3）总损失函数

对于分类任务，我们使用负对数似然损失函数（NLLoss）来评估所提出的网络结构，总损失函数定义如下。

$$Loss(\theta) = \frac{1}{\Omega} \sum_{i=1}^{\Omega} NLLoss(O_j^i - \overline{O_j^i}) + \gamma \cdot \|\theta\|^2 \quad (3.6)$$

其中， Ω 表示试验次数， O_j^i 表示第 j 个受试者的第 i 次试验的真实标签。 $\overline{O_j^i}$

是分类模块 M-ConvNet 的预测结果。 γ 是一个超参数，其值介于 0 和 1 之间（ $\gamma = 0.01$ ）。 $\|\theta\|^2$ 表示为正则化项，以减轻过拟合。我们的目标是找到一个最优参数 θ ，从而获得最小的损失 $Loss(\theta)$ 。

（4）评价指标

本章采用的性能评价指标包括分类准确率 acc、标准差 Std 以及 Kappa 系数，它们的定义式和作用如下：

分类准确率 acc 可以衡量模型对所有样本的整体识别正确率，是最常用的性能指标之一，能够直观反映分类器的总体预测能力。其定义式如下：

$$acc = \frac{TP + TN}{TP + TN + FP + FN} = \frac{\sum_{i=1}^c \frac{TP_i}{I_i}}{c} \quad (3.7)$$

式中， TP 、 TN 、 FP 、 FN 分别表示分类结果中的真正、真负、假正、假负样本数量。 c 是分类的类别数， TP_i 是类别 i 中正确预测的样本数， I_i 是类别 i 中样本数。

标准差 Std 用于衡量模型在多次实验或交叉验证中的稳定性，Std 越小，说明模型性能越稳定。其定义式如下：

$$Std = \sqrt{\frac{1}{N} \sum_{i=1}^N (acc - Avg)^2} \quad (3.8)$$

式中，Avg 表示平均分类准确率， N 代表受试者数。

Kappa 系数用于衡量分类结果与随机猜测之间的改进程度，剔除偶然一致带来的影响，更真实地反映模型的有效性，特别适用于类别分布不均的场景。

$$Kappa = \frac{acc - p_e}{1 - p_e} \quad (3.9)$$

式中， p_e 是偶然一致性，即各类别分别对应的实际样本数量与预测样本数量的乘积除以样本总数平方之和。

（5）网络配置和实验环境

该网络模型使用 Adam 优化器进行模型优化，卷积层的参数由 Xavier 算法初始化，初始学习率设置为 0.0001，衰减权重为 0.01。

实验采用的 CPU 为 Intel Core i9-9900X 3.50Ghz，显卡为 NVIDIA 2080Ti 12G，深度学习框架为 Pytorch。

3.4 实验结果与分析

本节将围绕提出的 MSF3D 网络模型在上述三个公开 MI-EEG 数据集上分类性能及相关实验结果展开分析。首先，第 3.4.1 小节将 MSF3D 模型与其他相关

模型和优秀算法进行对比实验，通过比较这些方法在不同数据集上的分类准确率、标准差及 Kappa 系数，探讨其性能差异。接着，第 3.4.2 小节通过模块消融实验，进一步验证 MSF3D 各关键组件对整体模型性能的贡献。最后，第 3.4.3 节利用留一交叉验证来测试 MSF3D 在面对陌生对象时的跨被试泛化能力。

3.4.1 对比实验

上述 3.3.4 节选用的各方法在数据集 2A、2B 以及 HGD 上的 MI-EEG 分类实验结果，包含各受试者的准确率 acc，各方法的平均准确率 Avg，标准差 Std 以及 Kappa 系数如表 3.3、表 3.4 和表 3.5 所示，将其转化为更直观的折线图则如图 3.9、图 3.10 和图 3.11 所示。

表 3.3 各方法在数据集 2A 上的分类性能（最佳表现用粗体显示）

被试	FBCSP	SBCSP	EEGNet	EEG-TCNet	DJDAN	M3DCNN	HSCNN	MSF3D
A01	77.44	75.10	83.24	88.62	85.48	90.14	91.67	95.52
A02	49.54	66.32	59.58	79.86	69.43	81.24	79.65	83.56
A03	82.2	79.45	91.56	95.43	92.64	95.28	96.58	97.37
A04	55.76	59.05	78.94	80.24	84.42	88.76	88.76	91.87
A05	49.25	65.21	57.71	75.76	74.65	96.24	96.95	93.24
A06	52.34	43.55	48.35	74.27	65.74	85.54	86.14	82.08
A07	81.62	76.25	88.86	93.45	94.38	91.44	93.25	97.54
A08	82.55	82.15	84.75	90.27	84.27	94.21	91.54	91.34
A09	78.15	62.48	81.34	84.5	82.74	91.61	93.61	96.45
Avg	67.65	67.73	75.15	84.99	81.53	90.72	90.35	92.89
Std	13.18	11.98	14.74	6.78	9.11	4.85	5.34	5.20
Kappa	0.562	0.563	0.6687	0.7998	0.754	0.8729	0.8673	0.906

表 3.4 各方法在数据集 2B 上的分类性能（最佳表现用粗体显示）

被试	FBCSP	SBCSP	EEGNet	EEG-TCNet	DJDAN	M3DCNN	HSCNN	MSF3D
B01	71.20	74.86	76.43	77.52	78.30	80.47	81.30	83.71
B02	59.46	56.47	60.19	63.07	64.46	63.84	65.26	66.88
B03	61.74	57.44	58.24	62.58	66.59	64.28	67.52	65.64
B04	96.17	97.65	95.79	97.86	96.88	96.04	97.68	98.64
B05	93.23	80.24	91.87	95.26	95.76	93.66	92.46	94.61

续表 3.4

被试	FBCSP	SBCSP	EEGNet	EEG-TCNet	DJDAN	M3DCNN	HSCNN	MSF3D
B06	81.52	89.23	86.76	84.63	85.21	87.89	86.12	91.02
B07	77.26	84.34	84.94	83.79	84.42	83.71	82.74	89.47
B08	91.43	92.51	93.79	93.42	90.64	94.87	92.84	94.75
B09	88.18	88.97	89.84	87.69	89.68	91.29	89.22	93.24
Avg	80.47	80.19	82.2	83.09	83.1	84.23	84.35	86.55
Std	13.45	14.33	13.13	12.41	11.29	11.6	11.11	11.22
Kappa	0.6094	0.6038	0.644	0.6618	0.662	0.6846	0.687	0.731

表 3.5 各方法在数据集 HGD 上的分类性能（最佳表现用粗体显示）

被试	FBCSP	SBCSP	EEGNet	EEG-TCNet	DJDAN	M3DCNN	HSCNN	MSF3D
H01	80.74	93.46	94.28	84.47	94.55	88.97	94.87	96.45
H02	92.03	92.28	96.41	92.89	93.14	92.35	95.05	94.14
H03	92.65	96.47	95.39	96.88	94.27	92.47	88.74	98.01
H04	93.44	95.71	97.74	98.24	96.01	93.86	98.13	97.76
H05	90.64	96.47	94.64	91.26	92.54	92.27	94.12	95.59
H06	92.16	95.31	92.68	92.55	94.05	94.06	95.76	94.42
H07	85.62	93.41	93.41	92.18	93.17	91.18	92.45	95.87
H08	91.31	96.17	93.94	92.06	96.07	92.47	95.41	96.89
H09	95.29	96.84	96.83	93.12	97.18	94.69	95.75	98.71
H10	86.22	91.17	90.34	87.59	87.68	87.25	93.14	92.81
H11	87.42	79.92	79.77	92.89	84.71	89.14	88.86	93.14
H12	92.61	95.10	94.22	94.45	93.24	90.16	92.71	95.86
H13	90.07	92.68	92.75	92.01	94.14	89.87	96.22	96.05
H14	84.27	80.17	91.26	86.12	93.04	86.09	92.86	94.34
Avg	90.02	92.52	93.47	92.53	93.12	90.91	93.9	95.39
Std	4.04	5.4	4.93	3.59	3.47	2.61	2.89	1.86
Kappa	0.8669	0.9003	0.913	0.9005	0.9082	0.8782	0.9187	0.9385

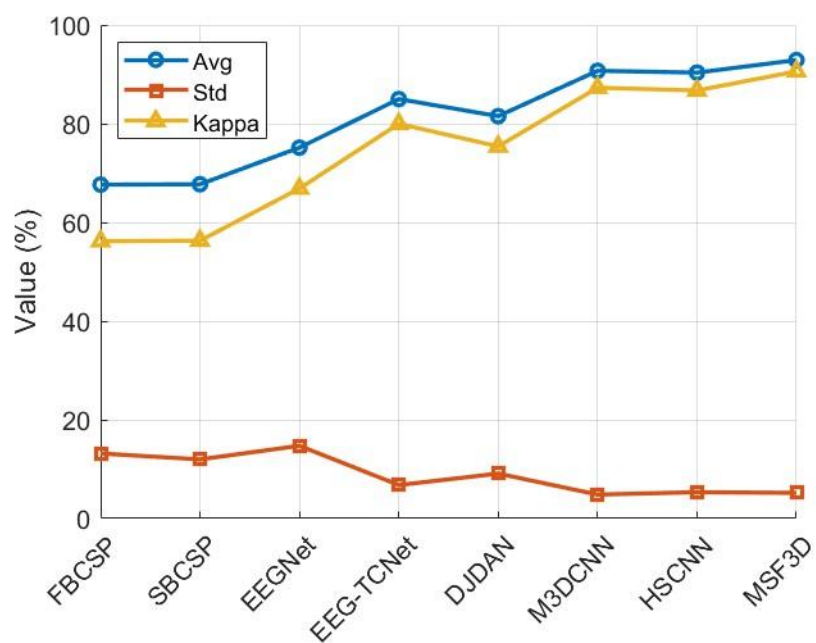


图 3.9 各方法在数据集 2A 上的分类性能折线图

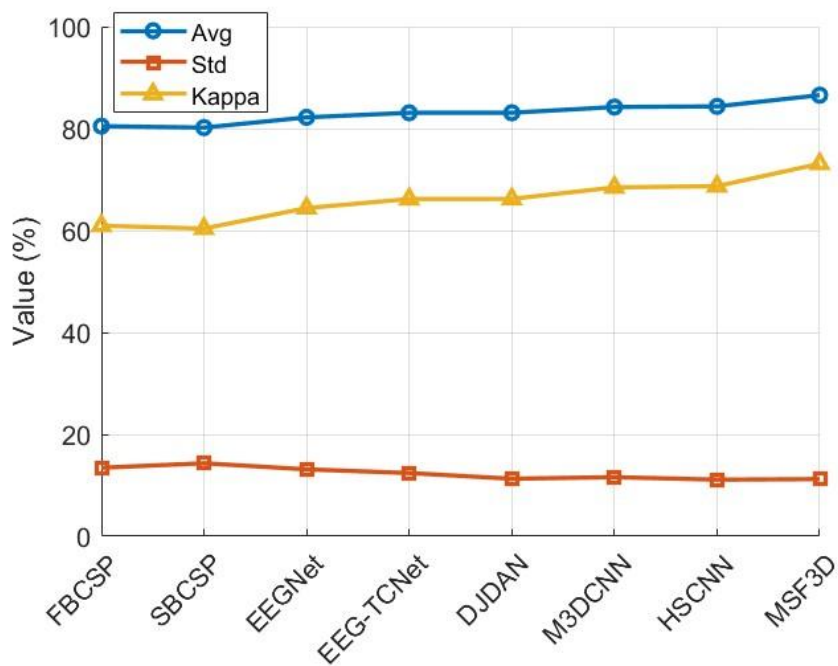


图 3.10 各方法在数据集 2B 上的分类性能折线图

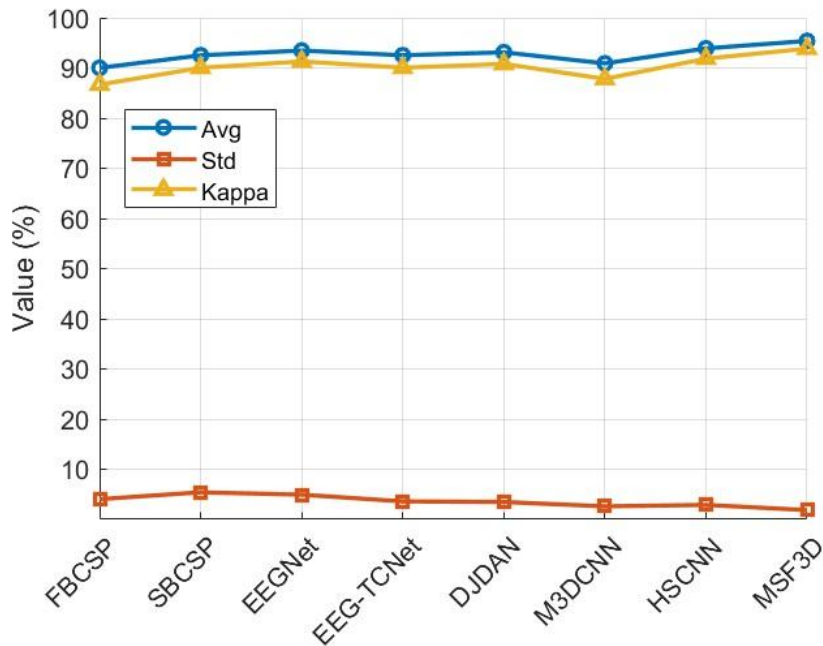


图 3.11 各方法在数据集 HGD 上的分类性能折线图

在 2A 数据集上，所对比的八种方法在平均准确率 Avg、标准差 Std 和 Kappa 系数三项指标上表现差异显著。传统方法如 FBCSP 与 SBCSP 虽然有计算成本低的优势，但碍于 CSP 的局限性，分类准确率较低，仅 67% 左右，且稳定性较差，说明其对复杂 MI-EEG 信号特征的建模能力有限。而深度学习方法中，EEGNet 的优点是能够自动从脑电图数据中提取时空特征，但由于结构简单限制了特征提取能力。EEG-TCNet 结合了时序卷积和图卷积网络，能够高效地提取脑电图的时空特征并捕捉不同通道间的空间依赖性，但模型复杂度较高对数据预处理要求过高，故表现不佳。DJDAN 则是因为在有限的卷积层中，单个接受核的大小限制了提取高级特征来提高分类性能，准确率仅有 81.53%。本章提出的 MSF3D 取得最高的平均准确率 92.89% 与 Kappa 系数 0.906，并保持较低的标准差 5.2%，显示出优秀的分类性能与真实有效性。同时在 9 名受试者中的 6 名（A01、A02、A03、A04、A07 和 A09）中取得了最高的分类准确率。如图 3.12 所示，以受试者 A01 为例分别进行混淆矩阵和 t-SNE 可视化分析。M3DCNN 与 HSCNN 模型也表现出色，得益于 M3DCNN 首次将 3D 数据结构引入脑电信号过程，对原始 MI-EEG 进行更深入、更复杂的表示。而 HSCNN 则采取使用随距离分布的 3 个核大小在时域、空间和频域中提取 EEG 多尺度信息方式有效提高了准确率。但两者均和 MSF3D 有 2.17%、2.54% 的 Avg 差距和 0.0331、0.0387 的 Kappa 系数差距。

为了测试模型的泛用性，在三通道的二分类数据集 2B 以及分辨率更高、噪声更低的数据集 HGD 上也进行对比实验。结果表明本章提出的 MSF3D 模型也

表现不俗。在 2B 数据集上达到最佳的 86.55%和 0.731 的平均分类准确率和 Kappa 系数，比起第二名的 HSCNN 分别高出 2.2%和 0.044，并且在 9 名受试者中的 6 名（B01、B02、B04、A06、A07 和 A09）中取得了最高的分类准确率。如图 3.13 所示，以受试者 B07 为例分别进行混淆矩阵和 t-SNE 可视化分析。而在数据集 HGD 也取得了最佳的 95.39%和 0.9385 的平均分类准确率和 Kappa 系数，相比其他方法至少提高了 1.49%和 0.0198，并且在 14 名受试者中的 8 名（H01、H03、H07、H08、H09、H11、H12 和 H14）中取得了最高的分类准确率。如图 3.14 所示，以受试者 H09 为例分别进行混淆矩阵和 t-SNE 可视化分析。除此之外，MSF3D 模型在数据集 2B 和 HGD 分别得到了略高于最佳表现 0.11 的 11.22 和最佳的 1.86 的标准差，这表明 MSF3D 分类性能非常稳定，泛化能力强。

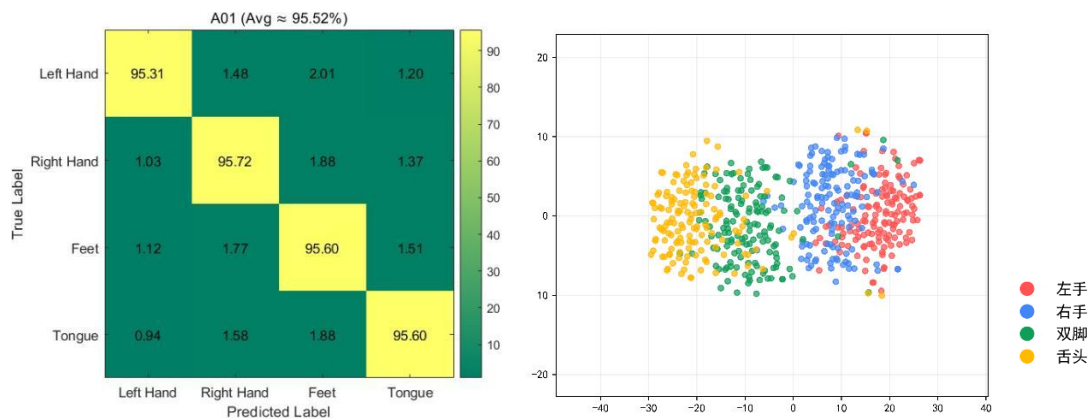


图 3.12 受试者 A01 在数据集 2A 的混淆矩阵和 t-SNE 可视化分析

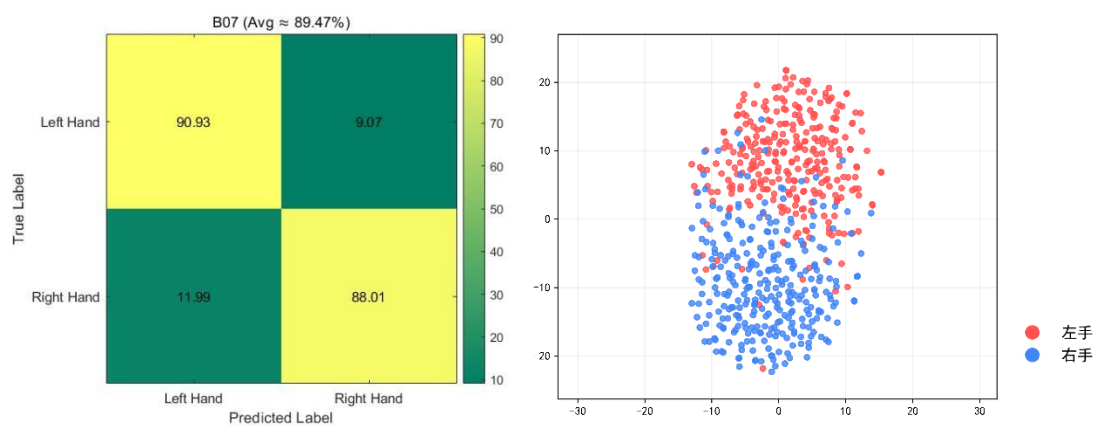


图 3.13 受试者 B07 在数据集 2B 的混淆矩阵和 t-SNE 可视化分析

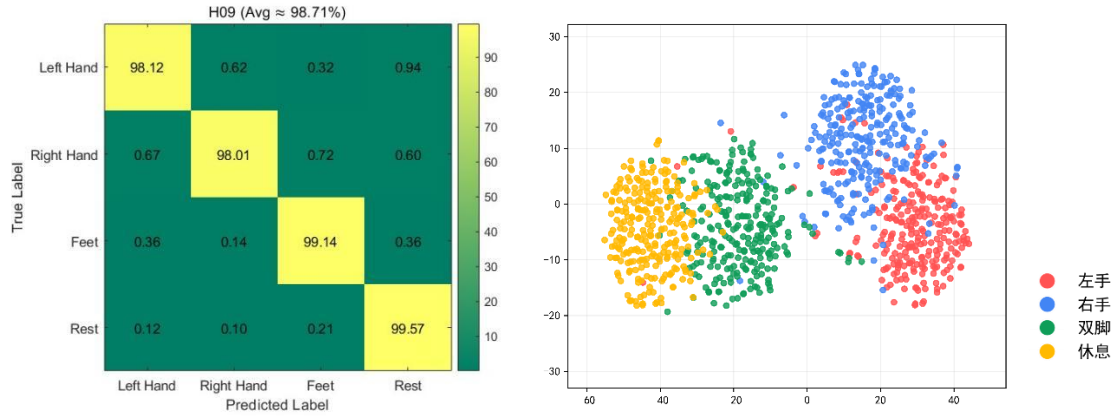


图 3.14 受试者 H09 在数据集 HGD 的混淆矩阵和 t-SNE 可视化分析

3.4.2 消融实验

消融实验可以评估模型各个组成模块对整体性能的贡献。通过逐步移除或替换模型中的某一部分，并观察性能变化，可以明确该组件在模型中的作用与重要性，从而验证设计的合理性并为后续优化提供依据。对本章提出的 MSF3D 模型，本节将分别移除 3D 空间注意力模块，记作 w/o. 3Dsa；移除多尺度时间注意力模块，记作 w/o. Msta 然后和完整的 MSF3D 模型进行对比。通过列表的定量分析所有对象的平均分类准确率，结果如表 3.6 所示。然后选取 2A、2B、和 HGD 数据集上的受试者 A07、B07 和 H03 三个对象，通过 t-SNE 图的定性分析来确认各模块在特征分布和分类的作用，结果图 3.15、图 3.16 和图 3.17 所示。

其中，t-SNE (t-distributed Stochastic Neighbor Embedding) 特征图的作用是将高维数据映射到二维或三维空间，以可视化其潜在结构和分布特征。通过保持原始数据在高维空间中的局部邻近关系，t-SNE 能够直观展示不同类别样本之间的聚类情况、类别分界以及数据的分布模式，常用于评估特征提取效果或模型学习到的表示能力。在 MI-EEG 解码任务中，t-SNE 图可用于验证模型是否成功地将不同状态的特征进行有效区分。

表 3.6 MSF3D 模型消融实验平均分类准确率(%)

数据集	MSF3D w/o.3Dsa	MSF3D w/o.Msta	MSF3D
2A	86.74	89.68	92.89
2B	85.36	84.21	86.55
HGD	87.18	90.66	95.39

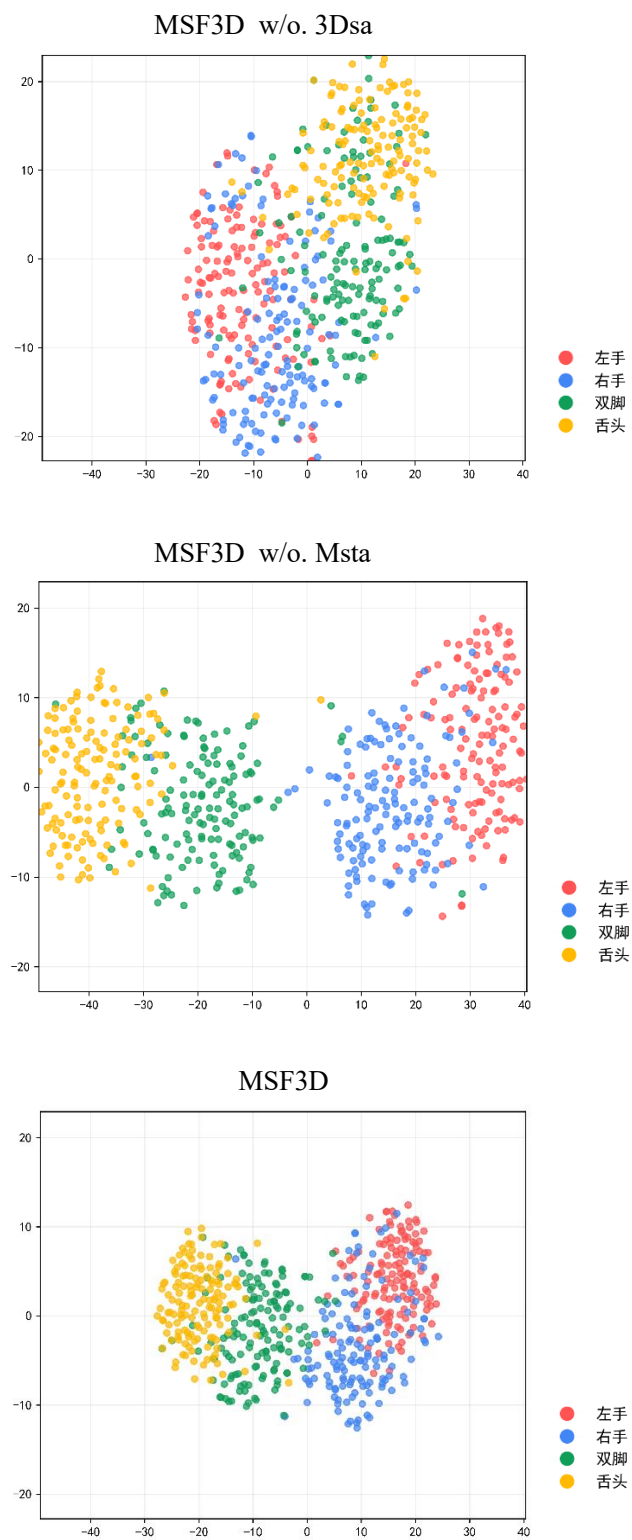


图 3.15 受试者 A07 消融实验 t-SNE 可视化图

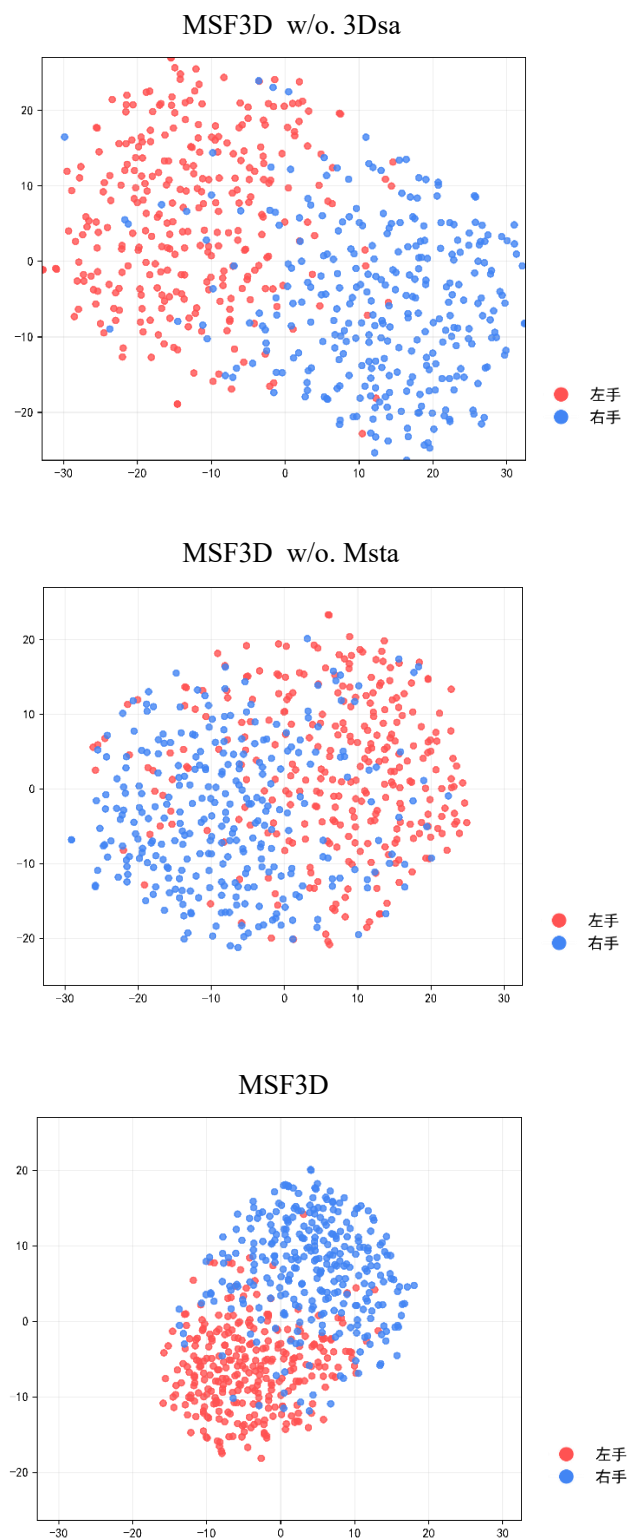


图 3.16 受试者 B07 消融实验 t-SNE 可视化图

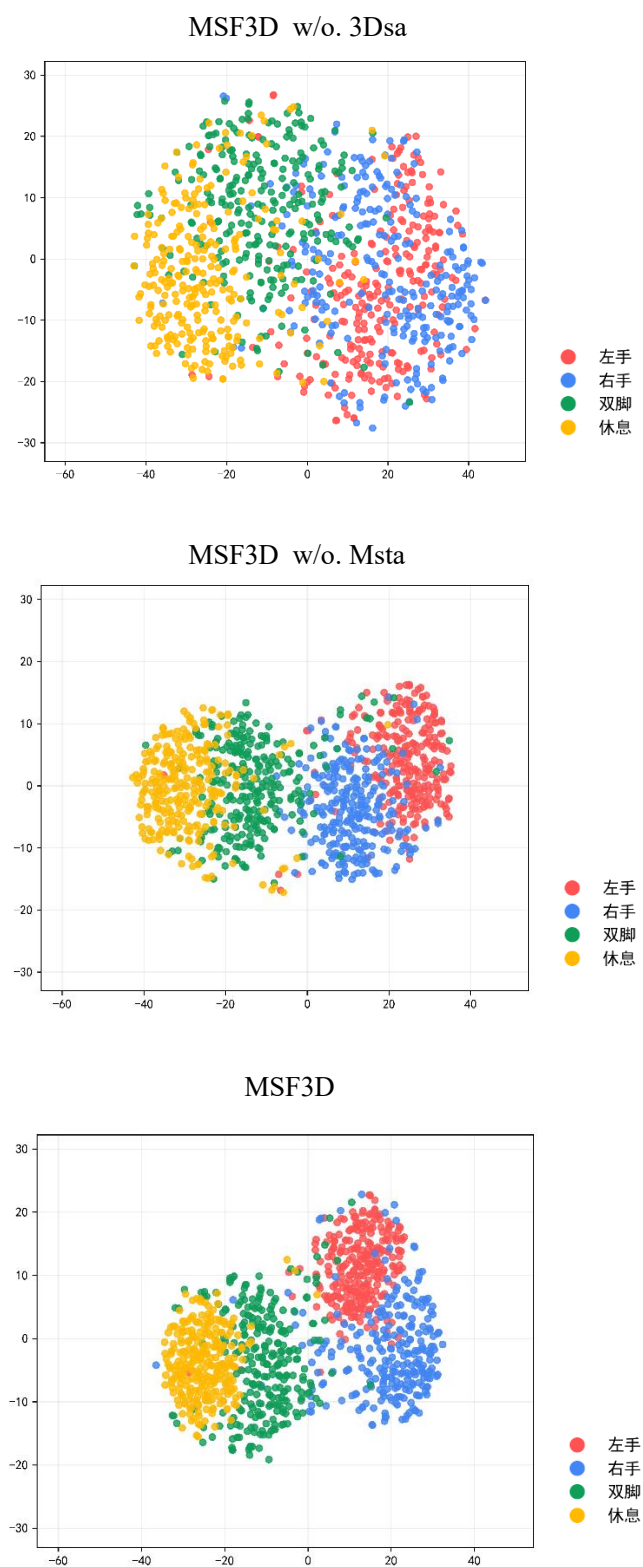


图 3.17 受试者 H03 消融实验 t-SNE 可视化图

首先，我们通过表 3.6 的数据做定量分析：总体来看，完整的 MSF3D 模型在三个数据集上均表现出最高的平均分类准确率，比移除部分模块表现更佳。

接着展开局部分析，在移除 3D 空间注意力模块后，2A 和 HGD 数据集的平均分类准确率骤降了 6.15% 和 8.21%，而 2B 数据集受影响的程度很小，这主要是由于 2B 数据集仅含三个电极通道的 MI-EEG 数据，通道间的空间依赖关系不强，因此即便移除 3D 空间注意力模块准确率也仅仅下降了 1.19%。而在移除了多尺度时间注意力模块后，三个数据集上的平均分类准确率也均有不同程度的下降。这是由于移除后模型的泛化能力下降，在应对新的分类对象时表现不稳定。

其次，我们可以通过图做定性分析：总体来看，完整的 MSF3D 模型在三个数据集的特征区分性上表现最优，特征分布更加清晰、聚类效果更显著，也侧面说明两个模块在整体架构中起到了不可替代的作用。其中，3D 空间注意力模块能通过给不同电极通道动态赋值提取深层空间特征，多尺度时间注意力模块则是通过给不同时间切片动态赋值将那些和 MI-EEG 相关的时间切片区分出来，并对其特征进行提取。两者都作为整体模型的重要组分提高了分类性能。

综合消融实验的定量和定性分析，可以得出结论：MSF3D 模型在 3.4.1 对比试验中的出色表现离不开 3D 空间注意力模块和多尺度时间注意力模块的支撑，该模型的所有组分共同构成了一个高效、鲁棒的整体，显著提升了 MI-EEG 的解码性能。

3.4.3 留一交叉验证实验

为全面评估模型在跨被试条件下的泛化能力，本文采用留一交叉验证（Leave One Subject Out, LOSO）策略。具体地，在每轮验证中，将一名被试的全部数据作为测试集，其余被试的数据用于模型训练。该过程对所有被试轮流进行一次，最终对各轮实验结果进行平均，得到模型的整体性能指标。如图 3.18 所示，在计算受试者 A01 的不依赖主题分类准确率时，A01 的测试样本即图中的蓝色矩形用作性能评价的测试样本，同时，其他受试者的所有数据集也即图中的黄色矩形都用作训练样本。

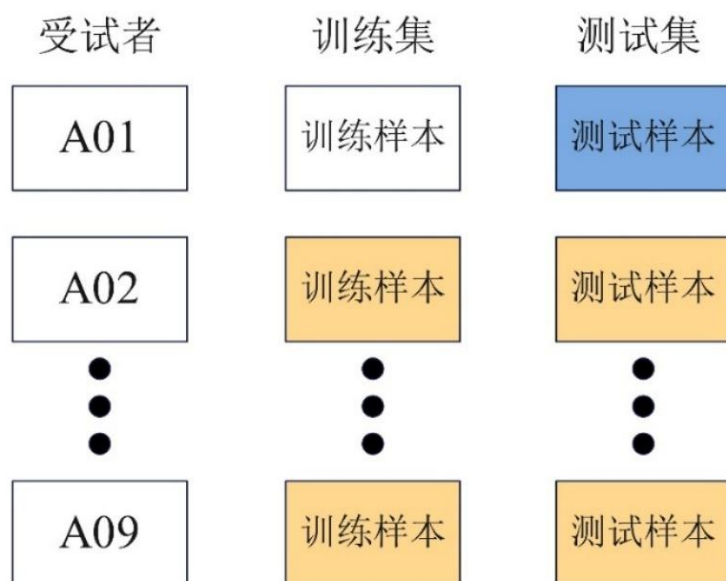


图 3.18 留一交叉验证流程示意图

LOSO 能有效避免训练和测试数据间的个体重叠，从而防止信息泄露，更加真实地反映模型在面对陌生被试时的适应能力。在 MI-EEG 解码任务中用该方法可以研究模型的跨被试泛化性能。

本节选取 Schirrmeister 等人的 DeepCNN，Lawhern 等人的 EEGNet，Amin 等人设计的 MCNN 作为比较方法，分析各个方法留一交叉验证后得到的准确率 acc 如表 3.7 所示

表 3.7 各方法在数据集 2A 上的留一交叉验证结果（最佳表现用粗体显示）

被试	DeepCNN	EEGNet	MCNN	MSF3D
A01	46.24	/	61.07	72.65
A02	32.41	/	42.41	48.47
A03	40.32	/	62.87	80.69
A04	33.47	/	51.26	55.74
A05	42.74	/	48.96	53.81
A06	35.24	/	38.41	54.58
A07	42.75	/	61.85	64.27
A08	49.25	/	58.49	71.74
A09	50.75	/	68.98	67.92
Avg	41.46	40.00	54.92	63.32

受试者不依赖主题分类准确率低的主要原因是 MI-EEG 信号随时间变化，且因受试者即便是同一受试者也不时变化，无法确定哪些电极通道或时间切片与 MI 最相关。因此，传统方法如 DeepCNN、EEGNet 和 MCNN 在独立于受试者的分类中的性能有限。相比之下，本章提出的方法 MSF3D 主要特点便是通过 3D 空间注意力模块 3Dsa 和多尺度时间注意力模块 Msta 可以自动为大多数与运动相关的电极通道和时间段分配权重。提高了受试者不依赖主题分类的平均准确率，达到了 63.32%，远高于其他方法，这也表明 MSF3D 在面对陌生的受试者时表现出良好的性能，模型的跨被试泛化性能较强。

3.5 本章小节

本章首先对当前特征提取算法出现的一些问题进行总结，并针对性地设计出一种具有多尺度特征的端到端的 3D 卷积神经网络模型 MSF3D。其次分模块介绍各模块的结构、参数和训练流程。接着通过和经典的、相关的、先进的算法在 2A、2B 和 HGD 三个公共数据集进行一系列对比，验证了 MSF3D 模型具备相当的准确性、稳定性。下一步则是通过消融实验证明 MSF3D 的每一个模块都在解码任务中发挥了重要作用，共同支撑起整体的优良表现。最后通过留一交叉验证实验证明了模型具有较强的跨被试泛化性能。

第4章 基于 WTGAN 的 MI-EEG 数据增强研究

4.1 引言

通过 1.2.5 节基于生成对抗网络的 MI-EEG 数据增强方法和 2.1 节 MI-EEG 的原理及特点的论述我们意识到由于 MI-EEG 具有非线性、非稳定，信号微弱，信噪比低等特点造成采集难度大，数据不充分，不能满足越来越复杂的深度神经网络的训练需要。同时，现有的数据增强研究又存在如下一些问题：

1. 传统的 MI-EEG 数据增强方法如噪声添加、时间平移、通道丢弃等存在生理合理性不足、忽视时空特性、生成多样性有限等问题。这些方法往往无法真实模拟 MI-EEG 信号的节律性、相位同步性和非平稳性，可能破坏多通道拓扑关系或长程时间依赖，且依赖人工参数调整，导致增强数据失真或过拟合风险。

2. 尽管 GAN 在图像生成领域已展现出强大能力，能够学习真实数据的分布特征并生成高质量样本，有效提升分类器性能，但是，目前的大多数算法仅在频域上进行数据增强，在时序信号生成领域，尤其是针对脑电信号这种具有高度非平稳性和随机性的时间序列数据，GAN 的应用研究仍较为有限。

3. 传统 GAN 在训练过程中使用 JS 散度作为优化目标，当真实数据分布和生成数据分布几乎没有重叠时，JS 散度会饱和，导致梯度消失。梯度消失问题会严重阻碍 GAN 的训练，导致模式坍塌或生成样本质量低下，最终影响数据增强的效果。

针对上述的问题，本章提出一种基于 WTGAN 的 MI-EEG 数据增强方法。

4.2 WTGAN 网络模型结构

本章节提出的基于 WTGAN 的数据增强方法的流程图如图 4.1 和图 4.2 所示，其中绿色箭头表示正向传播，黄色箭头表示损失优化。其最大的特征是在判别器的优化函数 \mathcal{L}_u 中引入梯度惩罚，此外，相较于传统的生成对抗网络，WTGAN 额外增设了编码器和解码器，同时在原本 TGAN 的框架基础上增加了基于注意力机制的特征提取模块。

引入梯度惩罚可以防止判别器过强，有效缓解 GAN 训练中出现的梯度消失问题。为了更有效地捕捉数据的潜在结构，模型中引入了编码器-解码器模块。该模块首先将真实样本通过编码器压缩至低维潜在空间表示，再通过解码器对其进行重构，保留原始数据的关键特征。在此基础上，生成对抗网络所生成的数据被映射到相同的隐空间，并与真实数据的编码特征进行对齐，通过监督机

制引导生成器学习潜在空间的真实分布特性，该训练过程如图 4.1 所示。经过生成器与判别器的对抗优化后，生成器输出的样本再经由解码器还原为完整的数据样本，这个数据生成的过程如图 4.2 所示。而增加的基于注意力机制的特征提取模块通过动态权重分配机制，同步捕获电极间空间关联性与 MI-EEG 信号的多尺度时序依赖。

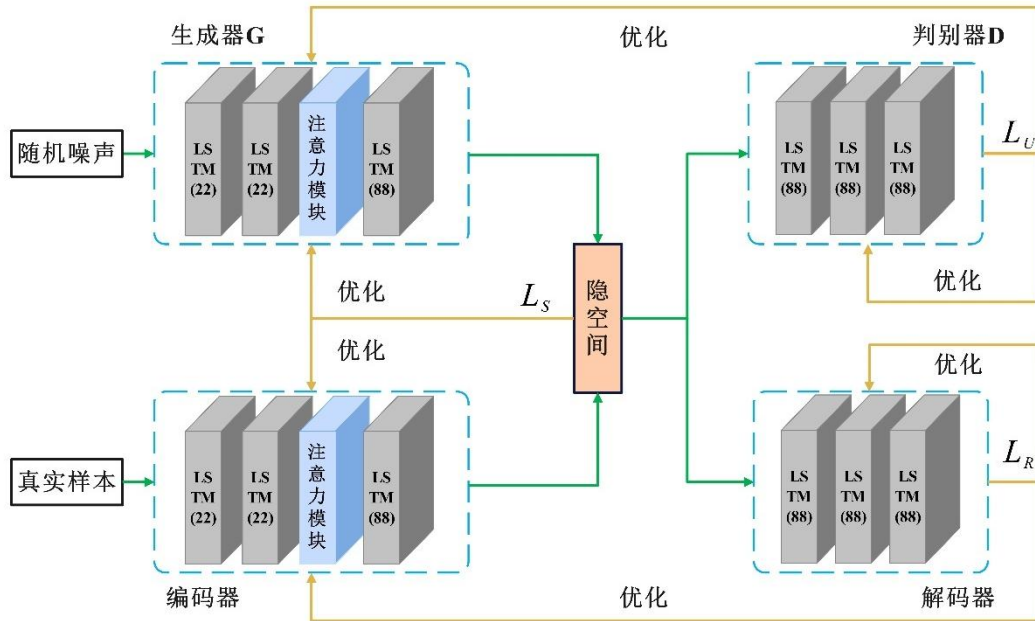


图 4.1 WTGAN 网络训练流程图

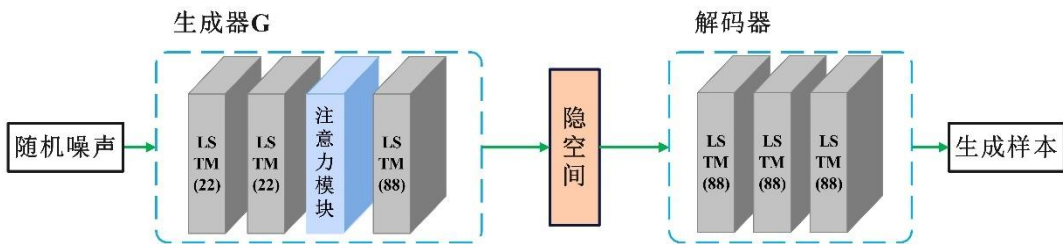


图 4.2 WTGAN 数据生成流程图

4.2.1 WGAN-GP 优化思路

Arjovsky 等人^[66]提出的 Wasserstein GAN (WGAN) 通过引入 Wasserstein 距离作为损失函数，有效解决了传统 GAN 训练中的关键问题。相较于使用 JS 散度的原始 GAN，Wasserstein 距离有效避免了 JS 散度导致的梯度消失问题，使生成器和判别器的训练更加平衡稳定。其次，Wasserstein 距离随训练过程单调递减，可作为训练效果的量化指标使训练进程可供监测。最后，大量实验证明：WGAN 能产生比原始 GAN 更高质量的生成样本。

Wasserstein 距离通过衡量概率分布之间的最优传输代价，提供了更平滑的梯度信号，这是其性能优势的理论基础。该改进使 GAN 模型在脑电信号生成等复杂任务中展现出更强的鲁棒性和可靠性，其定义式如下：

$$W(p_r, p_g) = \inf_{\gamma \sim \Pi(p_r, p_g)} E_{(x,y) \sim \gamma} [\|x - y\|] \quad (4.1)$$

其中， p_r, p_g 同 2.4 节定义，分别是真实数据分布和生成数据分布， $\Pi(p_r, p_g)$ 代表所有可能的联合分布 $\gamma(x, y)$ 的集合，其中 x 服从真实数据分布 p_r ， y 服从生成分布 p_g 。 $\gamma(x, y)$ 描述了从 p_r 到 p_g 的移动距离。Wasserstein 距离是所有可能传输方案中最优解。接下来用一个例子来说明，假设存在这样两个分布 P 和 Q ，它们分别满足 $\forall (x, y) \in P, x = 0.1, y \sim U(0, 1)$ 和 $\forall (x, y) \in Q, x = \theta, 0 \leq \theta \leq 1, y \sim U(0, 1)$ 。于是在 $\theta \neq 0$ 时， P 和 Q 完全独立，此时 KL 散度，JS 散度和 Wasserstein 距离有：

$$\begin{aligned} KL(P||Q) &= \sum_{x=0, y \sim U(0,1)} 1 \cdot \log \frac{1}{0} = +\infty \\ KL(Q||P) &= \sum_{x=0, y \sim U(0,1)} 1 \cdot \log \frac{1}{0} = +\infty \\ JS(P, Q) &= \frac{1}{2} \left(\sum_{x=0, y \sim U(0,1)} 1 \cdot \log \frac{1}{\frac{1}{2}} + \sum_{x=0, y \sim U(0,1)} 1 \cdot \log \frac{1}{\frac{1}{2}} \right) = \log 2 \\ W(P, Q) &= |\theta - 0.1| \end{aligned} \quad (4.2)$$

Wasserstein 距离相比 JS 散度具有显著优势，主要体现在以下方面：首先，即使真实分布 p_r 和生成分布 p_g 的支撑集完全不重叠即测度为零，Wasserstein 距离仍能提供连续且有意义的空间距离度量。其次，Wasserstein 距离的变化具有连续性，它随着分布间差异增大，距离值单调递增至无穷，而不会像 JS 散度那样收敛到固定值。当两个分布完全重合 $\theta = 0$ 时，Wasserstein 距离与 KL、JS 散度均为零。特别值得注意的是，在分布部分重叠或接近分离的情况下，W 距离能够提供比 JS 散度更精确的分布相似性评估，这对 GAN 训练的稳定性至关重要。这种特性使得 W 距离成为衡量生成模型性能的更优指标，特别是在分布匹配的早期阶段。

Wasserstein 距离定义式在实际应用中存在计算困难。为解决这一问题，可借助 Kantorovich-Rubinstein 对偶原理[44]将其转化为等价形式：

$$W(p_r, p_g) = \frac{1}{K} \sup_{\|f\|_L \leq K} (E_{x \sim p_r} [f(x)] - E_{x \sim p_g} [f(x)]) \quad (4.3)$$

接着引入一个依托 w 的函数 f_w ， f_w 满足 K-Lipschitz 连续性条件，记作 $\|f\|_L \leq K$ 也即 $|f(a) - f(b)| \leq K|a - b|$ 。其中 K 为正实数。值得注意的是，函数 f_w 可以采

用任意形式，包括深度神经网络等复杂非线性映射。于是，公式 4.3 重新表述为以下问题：

$$W(p_r, p_g) = \max_{w: \|f_w\|_L \leq K} E_{x \sim p_r}[f_w(x)] - E_{x \sim p_g}[f_w(x)] \quad (4.4)$$

在这个时候，判别器的角色发生了根本性转变：从传统的真假样本分类器演变为学习满足 K-Lipschitz 条件用以计算 Wasserstein 距离的函数。随着优化过程的推进，Wasserstein 距离的逐步缩小促使生成器产生更符合真实数据分布的样本。

以上的转化均基于一个前提，即此时的判别器 f_w 必须严格满足 K-Lipschitz 连续性条件，这可以通过两种技术实现：权重裁剪或是梯度惩罚。

权重裁剪核心思想是对判别器网络参数施加硬性边界约束。例如某次训练更新需要将网络权值参数控制在 $[-0.1, 0.1]$ 中：

$$w_{update}^{clip} = \begin{cases} 0.1, & w_{update} > 0.1 \\ w_{update}, & -0.01 \leq w_{update} \leq 0.1 \\ -0.1, & w_{update} < -0.1 \end{cases} \quad (4.5)$$

大量研究表明，权重裁剪技术存在明显的局限性：首先，该方法会导致判别器参数集中分布在裁剪边界附近，例中即 ± 0.01 处，严重制约了深度神经网络的表征能力；其次，不恰当的阈值设置会引发优化问题：过小的边界值导致梯度消失，而过大的边界值则可能引发梯度爆炸，这些都会显著影响生成样本的质量。

为克服这些缺陷，梯度惩罚 WGAN 网络（WGAN-GP）被提出，其核心思想是在判别器目标函数中引入正则化项，判别器的损失函数更新为：

$$L = E_{x \sim p_g}[D(x)] - E_{x \sim p_r}[D(x)] + \lambda E_{x \sim p_{\hat{x}}}[(\|\nabla_x D(x)\|_2 - 1)^2] \quad (4.6)$$

梯度惩罚方法在理论层面需要对整个输入空间施加约束，但实现中采用了一种高效的近似策略：仅针对真实样本与生成样本之间的插值样本 \hat{x} 施加梯度约束。通过在每次迭代中用真实样本和随机样本的如下计算式来拟合 \hat{x} 的值：

$$\hat{x} = \alpha x + (1 - \alpha)G(z) \quad (4.7)$$

式中， α 是 $[0, 1]$ 范围内的随机数。

WGAN-GP 框架的核心改进在于将判别器重构为 Wasserstein 距离估计器，同时将生成器的优化目标转化为最小化该距离度量。这种设计不仅有效避免了传统 GAN 训练中的梯度消失问题，还通过 Wasserstein 距离的单调递减特性为训练过程提供了可靠的量化指标。在实际应用中，该方法显著提升了模型训练的稳定性和生成样本的质量，同时保持了较低的计算复杂度。

4.2.2 编码器和解码器

TGAN 首次在传统 GAN 结构中引入编码器和解码器，两者可以用于捕捉 MI-EEG 信号时间序列的高维潜在表示，使生成的时序特征更加多样，提高数据增强的效果，进一步地，增强基于时序特征的 MI-EEG 识别模型的训练效果和分类准确率。编码器、解码器均是采用 LSTM 网络结构搭建的。LSTM (Long Short-Term Memory) 网络是一种特殊的循环神经网络 (Recurrent Neural Network, RNN)，专为处理和记忆长序列数据中的长期依赖关系而设计。如图 4.3 所示，它通过引入门控机制包括输入门、遗忘门和输出门来控制信息的流动，有效缓解了传统 RNN 在处理长序列时容易出现的梯度消失或梯度爆炸问题，使其在序列建模任务中，如时间序列预测和 EEG 分析等领域表现出色。

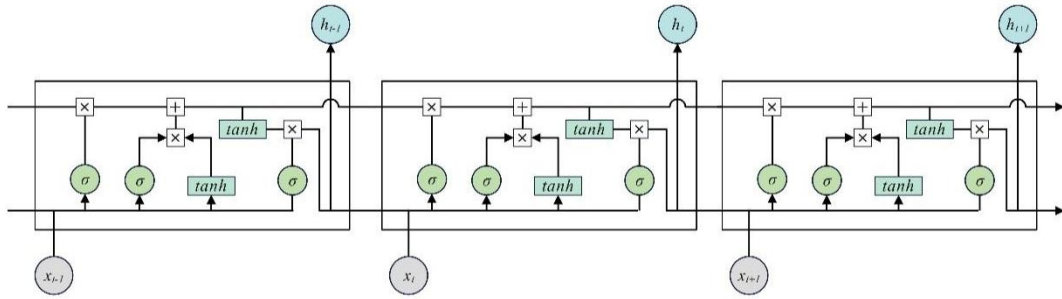


图 4.3 LSTM 结构图

首先，遗忘门用于判断前一时刻的记忆状态中哪些信息应被保留或丢弃。具体地，当前输入向量 x_t 与前一时刻的隐藏层输入 h_{t-1} 状态共同输入至 Sigmoid 激活函数，输出值在 $[0,1]$ 之间，用于表示信息保留的程度。输出值越小，表明该部分信息的重要性越低，当值为 0 时表示完全舍弃。该机制本质上类似一种注意力加权策略，用于学习历史状态对当前任务的影响权重，并据此调整前一单元状态的保留程度，更新后的遗忘门的输出 C_{t_1} 如下：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4.8)$$

$$C_{t_1} = f_t^* C_{t-1} \quad (4.9)$$

其次，输入门控制当前输入对细胞状态的更新程度。Sigmoid 函数决定当前时刻输入中哪些信息需要写入记忆单元，Tanh 函数则生成新的候选记忆 \tilde{C}_t 。二者的乘积形成当前时刻应被纳入的记忆内容 C_{t_2} 。随后，遗忘门输出的 C_{t_1} 与输入门更新项 C_{t_2} 相加，构成当前时刻的细胞状态 C_t ，用于存储长期依赖信息，详细计算过程如下：

$$\begin{aligned}
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
\tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\
C_{t_2} &= i_t^* \tilde{C}_t \\
C_t &= C_{t_1} + C_{t_2}
\end{aligned} \tag{4.10}$$

最后，输出门决定当前时刻的隐藏状态输出 h_t ，即短期记忆的内容。该门控同样通过 Sigmoid 函数确定输出比例，同时使用 Tanh 函数将细胞状态 c_t 映射至合适的输出空间，最终得到当前隐藏层状态 h_t ，计算如下：

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{4.11}$$

$$h_t = o_t^* \tanh(c_t) \tag{4.12}$$

EEG 作为大脑神经元群体放电活动的宏观反映，具有极高的时间分辨率，信号波动快速且与个体生理和心理状态紧密相关，呈现出明显的短期与长期依赖特性。因此，LSTM 网络因其对时序关系建模能力强，特别适用于脑电信号中时间动态特征的提取。在基于 LSTM 的 MI-EEG 解码过程中，首先将脑电信号预处理为离散的时序样本作为输入，利用 LSTM 学习其时序模式，再通过全连接层与 Softmax 分类器进行身份判别。为了进一步提升识别性能，可通过结构改进手段，如引入注意力机制、加深网络层数等方式增强模型对关键时序特征的表达能力，从而提升分类的准确性与鲁棒性。

本章使用的编码器功能区别于 TGAN 中利用低维时序特征作为驱动，转而将隐空间映射到相应的高维隐空间以获取更深层、更详实的时序特征信息，其具体实现如下：设原始数据空间为 X ，其中 $x \in X$ 表示一个真实样本，编码器将该样本映射至隐空间 H_x 。具体而言，编码器 e 实现从原始数据空间到隐空间的映射，即 $e: \Pi_t X \rightarrow \Pi_t H_x$ ，用于提取样本的高维时间序列特征。该映射关系如公式(4.13)所示：

$$h_{1:T} = e(X_{1:T}) \quad h_t = e(h_{t-1}, x_t) \tag{4.13}$$

本章节中的编码器采用三层结构的 LSTM 网络实现，并在第二层输出与第三层输入之间嵌入一个注意力机制模块，以增强模型对关键信息的关注能力。该注意力机制包含两个维度：时间维度和电极通道维度。

首先，为强化对局部时间依赖与静态信息的建模，引入时间注意力机制，其通过对第二层 LSTM 输出 M_o 进行加权，得到时间维度的权重矩阵 W_1 。计算过程如下所示：

$$W_1 = R_s \left(\text{mean} \left(\text{softmax} \left(F_s(M_o) \right) \right) \right) \tag{4.14}$$

其中， F_s 表示在电极通道维度上进行全连接操作，softmax 为归一化激活函数，mean 表示对通道维度求均值， R_s 表示将权重矩阵按电极通道维度进行复制扩展，最终输出为 24×22 的权重矩阵 W_1 。

其次，为突出不同通道的重要性，进一步设计了电极通道注意力机制，其权重矩阵 W_2 通过时序维度的全连接计算得到，如公式(4.15)所示：

$$W_2 = R_T \left(\text{mean} \left(\text{softmax} (F_t(M_o)) \right) \right) \quad (4.15)$$

同理， F_t 表示沿时序维度进行全连接运算， R_T 表示按时序维度复制扩展。最终，时间和通道维度的注意力权重被共同作用于第二层输出，得到注意力模块最终输出 M_i ，作为第三层 LSTM 的输入，其计算公式如下：

$$M_i = M_o^* (W_1^* W_2) \quad (4.16)$$

第三层 LSTM 基于上述融合后的时空特征继续建模，以获取更丰富、更具判别性的表示。

对于解码器部分，模型采用三层 LSTM 结构对编码器生成的高维特征进行解码，输出与原始数据特征分布尽可能接近的样本。为简化模型并提高训练效率，解码器并未引入注意力机制，但其性能仍可保持稳定。解码器 r 实现将隐空间特征映射回原始数据空间，即 $r: \Pi_t X \rightarrow \Pi_t H_x$ ，其数学表达式如下：

$$\tilde{X}_{1:T} = r(h_{1:T}) \quad \tilde{x}_t = r(h_t) \quad (4.17)$$

为了确保编码器映射得到的隐空间特征在经过解码器后能够尽可能还原原始数据的分布，需要设计一个损失函数 \mathcal{L}_R ，用于最小化原始样本 x_t 与重构样本 \tilde{x}_t 之间的差异。该损失函数的计算方式如公式(4.18)所示：

$$\mathcal{L}_R = \mathbb{E}_{x_{1:T} \sim p} \left[\sum_t \|x_t - \tilde{x}_t\|_2 \right] \quad (4.18)$$

4.2.3 生成器和判别器

鉴于传统 GAN 在建模时间序列特征方面存在一定局限性，本研究中所设计的生成器旨在生成与编码器输出相对应的隐空间特征，以更有效地捕捉数据的时序结构与通道间依赖信息。设 Z 为生成器输入的随机向量空间， $z \in Z$ 为其中一个样本，生成器 g 实现从随机向量空间到高维隐空间 H_x 的映射，即 $g: \Pi_t Z \rightarrow \Pi_t H_x$ ，目标是使生成的特征分布尽可能接近真实样本在隐空间中的分布。该映射过程如公式(4.19)所示：

$$\tilde{h}_{1:T} = g(Z_{1:T}) \quad \tilde{h}_t = g(h_{t-1}, z_t) \quad (4.19)$$

为了确保生成器输出的特征与编码器映射得到的隐空间特征具有相似的内部结构，生成器在架构设计上与编码器保持一致，均采用三层注意力机制 LSTM 网络构建。注意力模块插入于第二层 LSTM 的输出与第三层输入之间，其结构与编码器中的注意力模块完全相同，用于增强模型对关键时间点和重要通道的关注。

判别器部分用于判断输入的隐空间特征是否来源于真实样本或生成器生成

的伪样本。其功能由判别函数 d 表示，完成从隐空间 H_x 到判别输出空间 $[0,1]$ 的映射，即 $d: \prod_t H_x \rightarrow \prod_t [0,1]$ 。判别器由三层 LSTM 网络与一个全连接分类器构成，其判别输出记为 \tilde{y}_t ，计算方式如下：

$$\tilde{y}_t = d(u_t^1, u_t^2) \quad (4.20)$$

其中， $u_t^1 = c^1(\tilde{h}_t, u_{t+1}^1)$ ， $u_t^2 = c^2(\tilde{h}_t, u_{t-1}^2)$ 分别表示最后一层 LSTM 的前向与后向隐藏层的状态，函数 c^1 和 c^2 表示判别器中的多层 LSTM 模块。按照传统的 GAN 思路，生成器与判别器将通过无监督的对抗损失函数进行联合训练，使生成器能够持续优化其隐空间输出特征，使之更贴近真实数据的特征分布，其损失函数定义式如公式(4.21)所示：

$$\mathcal{L}_u = \mathbb{E}_{x_{1:T} \sim p} \left[\sum_t \log y_t \right] + \mathbb{E}_{x_{1:T} \sim \hat{p}} \left[\sum_t \log(1 - \hat{y}_t) \right] \quad (4.21)$$

但是，根据我们在第 2.4 节和第 4.2.1 节中的论述，传统 GAN 使用 JS 散度量真实分布和生成分布间距离，在训练过程中难免会出现梯度消失的问题，使扩充数据集的质量达不到预期。于是，我们在这里引入 Wasserstein 距离取代 JS 散度，同时加入梯度惩罚策略将上述损失函数更新为：

$$\begin{aligned} \mathcal{L}_u = \mathbb{E}_{x_{1:T} \sim p} \left[\sum_t \log y_t \right] + \mathbb{E}_{x_{1:T} \sim \hat{p}} \left[\sum_t \log(1 - \hat{y}_t) \right] \\ + \lambda \mathbb{E}_{x_{1:T} \sim \hat{p}} \left[\left(\left\| \nabla_t \sum_t \log(1 - \hat{y}_t) \right\|_2 - 1 \right)^2 \right] \end{aligned} \quad (4.22)$$

4.3 数据集及实验设置介绍

为了验证提出模型的性能表现，本章节选用的是 BCICIV_2A (2A)。下面分别所用数据集及实验设置做详细介绍。

4.3.1 BCICIV_2A 数据集

相关介绍见第 3.3.1 节。

4.3.2 实验设置

(1) 实验数据设置

本文从数据集 2A 中选取 A01-A09 每位受试者前 10000 个采样点作为基础数据，在经过去噪声去伪迹和数据标准化后，从原始信号中截取训练样本，用于构建 WTGAN 的训练集以及识别系统的原始样本集。此处选用的数据标准化方式是最大最小归一化 (Min-Max scaling)。最大最小归一化主要用于将原始数据按比例缩放到指定的范围内，通常是 $[0,1]$ 。这种归一化方法可以消除不同特征之间量纲不一致带来的影响，使模型在训练时更容易收敛，提高学习效率，其

计算公式如下：

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (4.23)$$

其中， x 是原始数据， x_{\min} 和 x_{\max} 分别表示数据集中出现的最小值和最大值，而 x' 是归一化后的数值。

（2）对比方法选择

为了验证本章提出方法的数据增强效果，选取的对比方法应该既包含传统的脑电裁剪方法（Crop），也应有一些经典的 GAN 模型衍生模型，如仅依靠卷积结构来捕捉时序特征的 WaveGAN，同时也要有与本方法相关的算法，如专门针对时序数据生成的生成对抗网络 TGAN。

（3）总损失函数

鉴于编码器与生成器均以同一高维隐空间作为映射目标，为确保生成器所生成的样本在隐空间中的分布能够有效对齐于真实数据经编码器变换后的特征分布，本文引入监督损失项以最小化二者在隐空间特征表示上的差异。通过该监督机制，在联合训练过程中可促使生成器学习到与真实样本相一致的潜在特征分布，即趋近于 $p(H_X)$ 的理想分布。具体而言，采用最大似然估计对监督损失函数进行优化，其计算过程如下：

$$\mathcal{L}_S = \mathbb{E}_{x_{1:T} \sim p} \left[\sum_t \|h_t - g(h_{t-1}, z_t)\|_2 \right] \quad (4.24)$$

值得强调的是，监督损失项 \mathcal{L}_S 在本质上建立了编码器与生成器之间的直接联系，使二者能够共同映射至统一的潜在空间，并持续通过损失函数进行对齐优化。

（4）评价指标

本章采用的性能评价指标包括分类准确率 acc、预测分数以及判别分数。判别分数与预测分数这两种指标可以精确度量生成样本与真实样本之间的距离，以实现更为客观的定量分析。

（5）网络配置和实验环境

在 3.4.2 节选用分类器是 Deep4 和 Shallow 网络模型，两者都使用 RMSProp 优化器进行模型优化，卷积层的参数由 Xavier 算法初始化，初始学习率设置为 0.0005，衰减权重为 0.01。

实验采用的 CPU 为 Intel Core i7-8700K 3.70Ghz，显卡为 RTX2080Ti 16G，深度学习框架为 Pytorch。

4.4 实验结果与分析

本节将围绕提出的 WTGAN 网络模型在 2A 数据集上的数据增强效果，数据

有效性分析及用于分类实验的表现展开分析，同时也会探索分类实验中最佳的生成数据和真实数据混合比。首先，第 3.4.1 小节会通过判别分数和预测分数来描述真实数据和生成数据的差距进而得到数据增强的效果。其次，第 3.4.2 小节将混合数据集作为神经网络分类器的输入，然后通过分类准确率的比较分析 WTGAN 能否增强 MI-EEG 解码任务的分类效果。最后第 3.4.3 节则探索不同生成数据和真实数据混合比对分类结果的影响。

4.4.1 数据增强效果分析

本节用判别分数和预测分数作为依据来对数据增强效果进行分析。

判别分数 (Discriminative Score): 为了精确度量生成样本与真实样本之间的距离，本文构建了一个双层 LSTM 时序分类器，并将其作为判别模型。在训练阶段，将原始样本标记为实数，生成样本则标记为非实数，将该问题建模为标准的二分类任务。分类器在训练完成后用于评估模型对两类数据的区分能力，从而反映生成数据与真实数据之间的差异程度。判别分数越小，表示差距越小，数据增强效果越好。判别分数如表 4.1 所示。

表 4.1 各方法的判别分数（最佳表现用粗体显示）

被试	Crop	WaveGAN	TGAN	WTGAN
A01	0.3445	0.2214	0.1855	0.0987
A02	0.2587	0.2631	0.3002	0.1725
A03	0.5821	0.3562	0.1834	0.1856
A04	0.4568	0.2457	0.2641	0.1245
A05	0.3985	0.3002	0.2597	0.2314
A06	0.4785	0.3947	0.4124	0.2414
A07	0.3658	0.2766	0.1798	0.1249
A08	0.5173	0.4439	0.3687	0.2471
A09	0.4712	0.3312	0.2714	0.2214
Avg	0.4304	0.3148	0.2695	0.1831

预测分数 (Predictive Score): 除衡量生成样本与真实样本的相似性外，还需评估生成数据在继承原始数据动态特性方面的表现。考虑到数据为时间序列形式，本文采用生成样本训练一个双层 LSTM 预测模型，目标为预测输入序列的下一时间步特征向量。随后，将该模型在真实数据集上进行测试，并使用平均绝对误差 (MAE) 作为性能指标，以衡量生成数据在时序模式学习方面的泛化能力。预测分数越小，表示生成数据保留原始数据特征的能力越强，数据增

强效果越好。预测分数如表 4.2 所示。

表 4.2 各方法的预测分数（最佳表现用粗体显示）

被 试	Crop	WaveGAN	TGAN	WTGAN
A01	0.0245	0.0212	0.0132	0.0133
A02	0.0164	0.0158	0.0012	0.0095
A03	0.0311	0.0264	0.0146	0.0142
A04	0.0201	0.0169	0.0094	0.0076
A05	0.0335	0.0301	0.0167	0.0152
A06	0.0314	0.0224	0.0178	0.0151
A07	0.0241	0.0261	0.0168	0.0124
A08	0.0301	0.0198	0.0251	0.0154
A09	0.0274	0.0284	0.0214	0.0194
Avg	0.0265	0.0230	0.0151	0.0136

从判别分数的结果可以看出，在四种方法中，WTGAN 在 2A 数据集 9 个被试的 8 个中具有最低的判别分数，说明其生成的数据样本与真实样本之间的分布差异更小，具备更高的相似性。在预测分数方面，WTGAN 同样在 9 个被试的 8 个表现最佳，反映出其在捕捉真实数据波动趋势方面具有更强的能力。综上，基于判别分数与预测分数这两个定量指标，ATGAN 在脑电数据生成质量上整体优于其他方法，展示出更优的建模性能和更强的时序特征学习能力。

4.4.2 分类准确性分析

为验证所生成数据的有效性，本小节在原始数据集与扩充后的数据集上采用相同的分类网络进行对比实验。为更直观地观察数据扩充对分类器性能的提升作用，本研究选用在 MI-EEG 时域信号数据增强应用广泛的两种分类器 Deep4 和 Shallow 网络模型。

在实验过程中，采用训练好的生成模型分别生成与原始训练集规模相当的样本数据用于扩充训练集，即生成数据和真实数据混合比为 1，随后在相同的测试集上评估分类器的识别性能。分类器在不同数据集上的训练条件保持一致，除训练数据来源不同外，其余超参数均一致。各方法在使用 Deep4 和 Shallow 作为分类器的准确率如表 4.3 和表 4.4 所示：

表 4.3 Deep4 作为分类器时各方法分类准确率（最佳表现用粗体显示）

被试	原始数据	Crop	WaveGAN	TGAN	WTGAN
A01	67.45	66.14	68.74	68.77	70.43
A02	51.32	52.33	52.17	54.01	56.95
A03	54.67	54.74	56.21	55.12	57.05
A04	95.12	94.69	97.14	96.25	96.66
A05	81.72	82.38	80.69	81.25	83.16
A06	74.21	74.49	75.79	76.68	78.21
A07	71.64	70.98	72.66	74.81	75.86
A08	80.62	81.06	80.13	83.02	84.11
A09	72.08	73.51	73.69	75.98	73.47
Avg	72.09	72.26	73.02	73.99	75.10

表 4.4 Shallow 作为分类器时各方法分类准确率（最佳表现用粗体显示）

被试	原始数据	Crop	WaveGAN	TGAN	WTGAN
A01	71.88	71.54	72.24	73.47	74.86
A02	53.02	52.97	52.82	53.84	52.69
A03	57.71	56.94	56.46	58.05	59.43
A04	93.85	94.21	94.71	94.45	95.11
A05	82.43	82.22	81.79	84.12	83.92
A06	74.33	73.86	75.84	78.66	78.84
A07	71.64	71.20	72.33	73.09	74.55
A08	81.89	81.63	82.34	83.44	84.82
A09	71.11	72.01	72.74	73.88	75.22
Avg	73.10	72.95	73.47	74.78	75.49

由表格可知，无论是 Deep4 还是 Shallow 作为分类器，几乎所有的数据增强方法都对分类准确率提高有贡献。其中脑电裁剪方法 Crop 对准确率的影响比较小，这可能是由于 Crop 方法是通过对原始信号进行裁剪来扩充训练样本数量，本质上并没有提升数据的多样性，并且由于样本片段的长度及样本总量直接受

到裁剪窗口的大小与步长参数的影响，当裁剪片段过短时，可能无法充分覆盖关键的时域脑电特征，导致分类性能下降，例如 Shallow 作为分类器时使用 Crop 方法反而使平均准确率下降了 0.15%。在这些方法中表现最出色的是本章提出的 WTGAN，不仅平均准确率在两种分类器下均位列第一，且在 9 个被试者中的多个个人准确率也保持领先。同时纵向对比，各个方法在 Shallow 分类器的表现均高于 Deep4。总之，实验结果表明 TWGAN 在提升 MI-EEG 分类精度方面表现显著，也再次验证了该模型生成的脑电数据具有较高的质量。

4.4.3 不同混合比影响

为进一步评估 WGAN 在提升模型分类性能方面的效果，本文将探索不同的生成数据和真实数据混合比，也即不同数据增强程度情况下，平均分类准确率将会呈现怎样的变化。

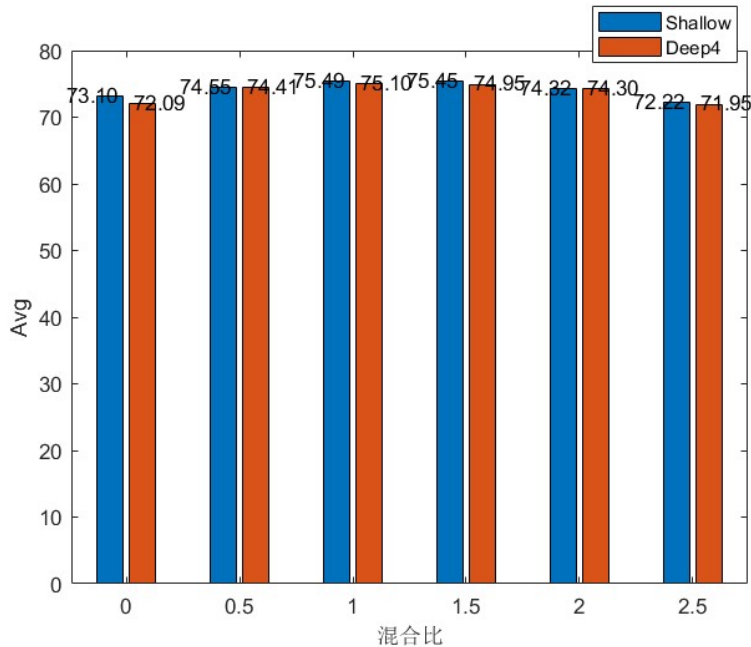


图 4.4 不同混合比下的平均分类准确率

如图 4.4 所示，实验结果表明，当生成数据和真实数据的混合比在 1 以下时，数据增强带来的分类性能增益较为显著；而当生成样本量超过原始数据的 1.5 倍后，分类性能提升趋于饱和甚至略有减弱，甚至在混合比为 2.5 时出现了反向效果的情况，这显示出增强规模与性能提升之间存在一定的边际效应。

4.5 本章小节

本章首先对当前基于 GAN 的 MI-EEG 数据增强算法中出现的一些问题进行总结，并针对性地基于 TGAN 设计出性能更佳的模型 WTGAN。其次分模块介绍各模块的结构、作用和训练流程。接着通过和经典的、相关的、先进的算法在 2A 公共数据集进行一系列对比，验证了 WTGAN 模型生成的数据质量较高。

下一步则是通过加入分类器来探究数据增强对 MI-EEG 解码准确率的影响。最后通过进一步地改变混合比发现了增强规模和性能提升之间的边际效应。

第 5 章 总结与展望

5.1 总结

本文通过说明原理，进行实验和结果阐释完成了以下的工作：

（1）提出基于 MSF3D 的 MI-EEG 解码模型，有效解决了传统 2D 表示作为输入时忽视电极通道相关性及不重视含运动想象时间切片造成的特征提取不完善的问题。通过和经典的、相关的、先进的算法在 2A、2B 和 HGD 三个公共数据集进行一系列对比，验证了 MSF3D 模型具备相当的准确性、稳定性。接着通过消融实验证明 MSF3D 的每一个模块都在解码任务中发挥了重要作用。最后通过留一交叉验证实验证明了模型具有较强的跨被试泛化性能。

（2）提出基于 WTGAN 的 MI-EEG 数据增强模型，有效解决了传统 GAN 梯度消失的问题，同时引入注意力模块提取空间和时序特征。接着通过和经典的、相关的、先进的算法在 2A 公共数据集进行一系列对比，验证了 WTGAN 模型生成的数据质量较高。下一步则是通过加入分类器来探究数据增强对 MI-EEG 解码准确率的影响。最后通过进一步地改变混合比发现了增强规模和性能提升之间的边际效应。

5.2 展望

首先，在基于多尺度 3D 卷积神经网络的 MI-EEG 的解码研究中，虽然通过描述各模块结构及计算过程，展示了空间注意力和时间注意力模块对实验分类精度的贡献，但是缺乏对注意力机制的可视化验证，后续可以通过拓扑图等形式呈现时空维度的注意力权重分布。这样做可以明确特征选择偏好，提升模型可解释性。

其次，在基于 WTGAN 模型的数据增强研究中，如何有效评估其生成数据的质量是下一步需要解决的重要问题。在图像生成任务中，例如人脸或自然场景图像，研究人员可以依靠视觉直观判断生成样本的清晰度与真实感，从而评价模型的训练效果。然而，在某些专业领域，如本文涉及的 MI-EEG，由于其复杂性和高度专业性，单纯依赖肉眼观察难以准确评估生成样本的质量。因此，除了本文设计的分类精度实验外，需要提出更多的验证方法来有效评估生成数据的质量。

参考文献

- [1] Saha S, Mamun K A, Ahmed K, et al. Progress in brain computer interface: Challenges and opportunities[J]. Frontiers in Systems Neuroscience, 2021, 15: 578875.
- [2] Korczowski L, Barachant A, Andreev A, et al. Brain Invaders 2: an open source Plug & Play multi-user BCI videogame[J]. 2016.
- [3] Van d L, B, Gurkok H, Plass-Oude Bos D, et al. Experiencing BCI Control in a Popular Computer Game[J]. IEEE Transactions on Computational Intelligence and AI in Games, 2013, 5(2): 176-184.
- [4] Tsui C S L, Gan J Q, Hu H. A Self-Paced Motor Imagery Based Brain-Computer Interface for Robotic Wheelchair Control[J]. Clinical Eeg & Neuroscience, 2011, 42(4): 225-229.
- [5] 梅意城. 基于脑机接口技术的手臂康复研究[D]. 北京工业大学, 2014.
- [6] 王磊. 基于运动想象的脑电信号分类与脑机接口技术研究[D]. 河北工业大学, 2009.
- [7] 冯春辉. 基于脑电节律的模式识别方法研究[D]. 燕山大学, 2011.
- [8] Saba, Moghimi, Azadeh, et al. A Review of EEG-Based Brain-Computer Interfaces as Access Pathways for Individuals with Severe Disabilities[J]. Assistive Technology, 2013.
- [9] Mellinger J, Schalk G, Braun C, et al. An MEG-based brain-computer interface (BCI)[J]. Neuroimage, 2007, 36(3): 581-593.
- [10] Coyle S, Ward T, Markham C, et al. On the suitability of near-infrared (NIR) systems for next-generation brain-computer interfaces[J]. Physiological Measurement, 2004, 25(4): 815-822.
- [11] Wilson J A, Felton E A, Garell P C, et al. ECoG factors underlying multimodal control of a brain-computer interface[J]. IEEE Transactions on Neural Systems & Rehabilitation Engineering A Publication of the IEEE Engineering in Medicine & Biology Society, 2006, 14(2): 246.
- [12] Zhang D, Song H, Xu R, et al. Toward a minimally invasive brain-computer interface using a single subdural channel: A visual speller study[J]. Neuroimage, 2013, 71(Complete): 30-41.
- [13] Clerc M. Brain computer interfaces principles and practise[J]. Biomedical

- Engineering Online, 2013, 12(1): 1-4.
- [14] Lee M, Jeong J H, Kim Y H, et al. Decoding Finger Tapping With the Affected Hand in Chronic Stroke Patients During Motor Imagery and Execution[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29: 1099-1109.
- [15] Li M, Li F, Pan J, et al. The mindgomoku: An online P300 BCI game based on Bayesian deep learning[J]. Sensors, 2021, 21(5): 1613.
- [16] Wold S, Esbensen K, Geladi P. Principal component analysis[J]. Chemometrics and intelligent laboratory systems, 1987, 2(1-3): 37-52.
- [17] Jung T P, Humphries C, Lee T W, et al. Extended ICA removes artifacts from electroencephalographic recordings[J]. Advances in neural information processing systems, 1998: 894-900.
- [18] 楼恩平. 抑郁症脑电信号特征提取及分类研究[D]. 金华: 浙江师范大学, 2009.
- [19] Chai R, Naik G R, Nguyen T N, et al. Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system[J]. IEEE Journal of Biomedical and Health Informatics, 2016, 21(3): 715-724.
- [20] Atyabi A, Shic F, Naples A. Mixture of autoregressive modeling orders and its implication on single trial EEG classification[J]. Expert Systems with Applications, 2016, 65: 164-180.
- [21] Yudhana A, Muslim A, Wati D E, et al. Human emotion recognition based on EEG signal using fast fourier transform and K-Nearest neighbor[J]. Advances in Science, Technology and Engineering Systems Journal, 2020, 5(6): 1082-1088.
- [22] Demuru M, La Cava S M, Pani S M, et al. A comparison between power spectral density and network metrics: an EEG study[J]. Biomedical Signal Processing and Control, 2020, 57: 101760.
- [23] Riaz F, Hassan A, Rehman S, et al. EMD-based temporal and spectral features for the classification of EEG signals using supervised learning[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2015, 24(1): 28-35.
- [24] Newson J J, Thiagarajan T C. EEG frequency bands in psychiatric disorders: a review of resting state studies[J]. Frontiers in Human Neuroscience, 2019, 12: 521.
- [25] Furman Ł, Duch W, Minati L, et al. Short-time fourier transform and embedding method for recurrence quantification analysis of EEG time series[J]. The European

- Physical Journal Special Topics, 2023, 232(1): 135-149.
- [26] Almanza-Conejo O, Almanza-Ojeda D L, Contreras-Hernandez J L, et al. Emotion recognition in EEG signals using the continuous wavelet transform and CNNs[J]. Neural Computing and Applications, 2023, 35(2): 1409-1422.
- [27] Subasi A, Jukic S, Kevric J. Comparison of EMD, DWT and WPD for the localization of epileptogenic foci using random forest classifier[J]. Measurement, 2019, 146: 846-855.
- [28] Pattnaik S, Dash M, Sabut S. DWT-based feature extraction and classification for motor imaginary EEG signals[C]// Proceedings of the 2016 International Conference on Systems in Medicine and Biology (ICSMB). Kharagpur, India: IEEE, 2016: 186-201.
- [29] Guger C, Ramoser H, Pfurtscheller G. Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface[J]. IEEE Transactions on Rehabilitation Engineering, 2000, 8(4): 447-456.
- [30] Razi S, Mollaei M R K, Ghasemi J. A novel method for classification of BCI multi-class motor imagery task based on Dempster-Shafer theory[J]. Information Sciences, 2019, 484: 14-26.
- [31] Ang K, Chin Z Y, Wang C, et al. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b[J]. Frontiers in neuroscience, 2012, 6: 39.
- [32] Novi Q, Guan C, Dat T H, et al. Sub-band common spatial pattern (SBCSP) for brain-computer interface[C]. //2007 3rd International IEEE/EMBS Conference on Neural Engineering, Kohala Coast, HI, USA. Piscataway, NJ: IEEE, 2007. 204-207.
- [33] Long J, Li Y, Wang H, et al. A hybrid brain computer interface to control the direction and speed of a simulated or real wheelchair[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2012, 20(5): 720-729.
- [34] Ghaemi A, Rashedi E, Pourrahimi A M, et al. Automatic channel selection in EEG signals for classification of left or right hand movement in Brain Computer Interfaces using improved binary gravitation search algorithm[J]. Biomedical signal processing and control, 2017, 33:109-118.
- [35] Dash D P, Kolekar M H, Jha K. Surface EEG based epileptic seizure detection using wavelet based features and dynamic mode decomposition power along with KNN classifier[J]. Multimedia Tools and Applications, 2022, 81(29): 42057-

42077.

- [36] Lotte F, Cichocki A, Bougrain L, et al. A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update[J]. Journal of Neural Engineering, 2018, 15(3): 031005.
- [37] Schirrmester RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggersperger K, Tangermann M, Hutter F, Burgard W, Ball T, et al. Deep learning with convolutional neural networks for eeg decoding and visualization[J]. Hum Brain Mapp, 2017, 38(11):5391-5420.
- [38] Lawhern V J, Solon A J, Waytowich N R, et al. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces[J]. Journal of Neural Engineering, 2018, 15(5): 056013.
- [39] Ingolfsson T M, Hersche M, Wang X, et al. EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces[C]. Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2020: 2958-2965.
- [40] Sun B, Zhao X, Zhang H, Bai R, Li T, et al. Eeg motor imagery classification with sparse spectrotemporal decomposition and deep learning[J]. IEEE Trans Autom Sci Eng, 2021, 18(2):541-551.
- [41] Dai G, Zhou J, Huang J, Wang N Hs-cnn: a cnn with hybrid convolution scale for eeg motor imagery classification[J]. Neural Eng, 2020, 17(1):016025.
- [42] Zhang D, Yao L, Chen K, Wang S, Chang X, Liu Y, et al. Making sense of spatio-temporal preserving representations for eeg-based human intention recognition[J]. IEEE Trans Cybern, 2021, 50(7):3033-3044.
- [43] Zhao X, Zhang H, Zhu G, You F, Kuang S, Sun L, et al. A multi-branch 3d convolutional neural network for eeg-based motor imagery classification[J]. IEEE Trans Neural Syst Rehabil Eng, 2019, 27(10):2164-2177.
- [44] Li D, Xu J, Wang J, Fang X, Ying J, et al. A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of eeg signals decoding[J]. IEEE Trans Neural Syst Rehabil Eng, 2020, 28(12):2615-2626.
- [45] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [46] Zhao W, Jiang X, Zhang B, et al. CTNet: a convolutional transformer network for EEG-based motor imagery classification[J]. Scientific Reports, 2024, 14(1):

- 20-23.
- [47] Fan Y, Shi X, Li Q. CNN-Based Personal Identification System Using Resting State Electroencephalography[J]. Computational Intelligence and Neuroscience, 2021:20-21.
- [48] Piplani T, Merrill N, Chuang J. Faking it, making it: fooling and improving brain-based authentication with generative adversarial networks[C]//2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE, 2018: 1-7.
- [49] Hartmann K G, Schirrmeister R T, Ball T. EEG-GAN: Generative adversarial networks for EEG brain signals[J]. ArXiv, 2018.
- [50] Abdelfattah S M, Abdelrahman G M, Wang M. Augmenting the size of EEG datasets using generative adversarial networks[C]//2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018: 1-6.
- [51] Mirza M, Osindero S. Conditional generative adversarial nets[J]. ArXiv, 2014.
- [52] Haradal S, Hayashi H, Uchida S. Biosignal data augmentation based on generative adversarial networks[C]//2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, 2018: 368-371.
- [53] Zhang K, Xu G, Han Z, et al. Data augmentation for motor imagery signal classification based on a hybrid neural network[J]. Sensors, 2020, 20(16): 4485.
- [54] Zhang X, Wang Z, Liu D, et al. Dada: Deep adversarial data augmentation for extremely low data regime classification[C]// IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2019: 2807-2811.
- [55] Luo Y, Lu B L. EEG data augmentation for emotion recognition using a conditional Wasserstein GAN[C]//2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, 2018: 2535-2538.
- [56] Zhang A, Su L, Zhang Y, et al. EEG data augmentation for emotion recognition with a multiple generator conditional Wasserstein GAN[J]. Complex & Intelligent Systems, 2021: 1-13.
- [57] Zhang Q, Liu Y. Improving brain computer interface performance by data augmentation with conditional deep convolutional generative adversarial networks[J]. ArXiv, 2018.
- [58] Luo Y, Zhu L Z, Lu B L. A GAN-based data augmentation method for

- multimodal emotion recognition[C]//Advances in Neural Networks–ISNN 2019: 16th International Symposium on Neural Networks, ISNN 2019, Moscow, Russia, July 10-12, 2019, Proceedings, Part I 16. Springer International Publishing, 2019: 141-150.
- [59] Pfurtscheller G, Da Silva F H L. Event-related EEG/MEG synchronization and desynchronization: basic principles[J]. Clinical Neurophysiology, 1999, 110(11): 1842-1857.
- [60] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]. Advances in neural information processing systems. 2672-2680.
- [61] Santurkar S, Tsipras D, Ilyas A, Madry A, et al. In Advances in Neural Information Processing Systems[M]. 2018: 2483-2493
- [62] Bjorck N, Gomes CP, Selman B, Weinberger KQ, et al. In Advances in Neural Information Processing Systems[M]. 2018: 7694-7705
- [63] Brunner C, Leeb R, Müller-Putz G, et al. BCI competition 2008–graz data set a[J]. Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology, 2008, 16: 1-6.
- [64] Leeb R, Brunner C, Müller-Putz G, et al. BCI competition 2008–graz data set b[J]. Graz University of Technology, Austria, 2008, 16: 1-6.
- [65] Hong X, Zheng Q, Liu L, Chen P, Ma K, Gao Z, Zheng Y, et al. Dynamic joint domain adaptation network for motor imagery classification[J]. IEEE Trans Neural Syst Rehabil Eng, 2021, 29:556-565
- [66] Arjovsky M, Chintala S, Bottou L, et al. Wasserstein gan[J]. ArXiv: 2017.