

分 类 号 \_\_\_\_\_

密 级 \_\_\_\_\_

U D C \_\_\_\_\_

编 号 10486

# 武汉大学

## 硕士 学位 论文

### 基于双分支特征融合和注意力机制的运动想象脑电图分类网络研究

研 究 生 姓 名: 刘梓轩

学 号: 2021202110015

指导教师姓名、职称: 李石君 教授

专 业 名 称: 计算机科学与技术

研 究 方 向: 机器学习与应用

二〇二四年四月







## 论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

学位论文作者（签名）：

年      月      日



## 摘要

论文的摘要是对论文研究内容和成果的高度概括。摘要应对论文所研究的问题及其研究目的进行描述，对研究方法和过程进行简单介绍，对研究成果和所得结论进行概括。摘要应具有独立性和自明性，其内容应包含与论文全文同等量的主要信息。使读者即使不阅读全文，通过摘要就能了解论文的总体内容和主要成果。

论文摘要的书写应力求精确、简明。切忌写成对论文书写内容进行提要的形式，尤其要避免“第1章……；第2章……；……”这种或类似的陈述方式。

关键词是为了文献标引工作、用以表示全文主要内容信息的单词或术语。关键词不超过5个，每个关键词中间用分号分隔。

**关键词：**脑机接口；运动想象；脑电图；注意力；深度学习



## ABSTRACT

An abstract of a dissertation is a summary and extraction of research work and contributions. Included in an abstract should be description of research topic and research objective, brief introduction to methodology and research process, and summary of conclusion and contributions of the research. An abstract should be characterized by independence and clarity and carry identical information with the dissertation. It should be such that the general idea and major contributions of the dissertation are conveyed without reading the dissertation.

An abstract should be concise and to the point. It is a misunderstanding to make an abstract an outline of the dissertation and words “the first chapter”, “the second chapter” and the like should be avoided in the abstract.

Keywords are terms used in a dissertation for indexing, reflecting core information of the dissertation. An abstract may contain a maximum of 5 keywords, with semi-colons used in between to separate one another.

**Key words:** BCI; Motor Imagery; EEG; attention; Deep Learning



## 目 录

摘要	I
ABSTRACT	III
1 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.2.1 基于先验知识的运动想象脑电图特征研究现状	3
1.2.2 基于深度学习的运动想象脑电图分类研究现状	4
1.3 研究内容	6
1.4 论文组织结构	6
2 运动想象脑电图分类和深度神经网络基础	7
2.1 脑电生理基础	7
2.1.1 人脑结构与运动想象	7
2.1.2 脑电图信号及其特性	7
2.1.3 脑电图信号与运动想象	12
2.2 深度神经网络基础	13
2.2.1 卷积神经网络	13
2.2.2 循环神经网络	17
2.2.3 注意力机制	19
2.3 本章小结	20
3 基于特征融合和注意力机制的运动想象脑电图分类网络构建	21
3.1 基于特征融合和注意力机制的端到端 MI-EEG 分类网络 HA-FuseNet	21
3.2 基于多尺度密集连接和混合注意力的分类网络 DIS-Net	22
3.2.1 基于 Inception 的基础网络	22
3.2.2 多尺度密集连接	25
3.2.3 混合注意力 svSE	28
3.3 基于 LSTM 和全局自注意力的分类网络 LS-Net	32
3.3.1 全局自注意力 SCoT	32
3.3.2 基于 LSTM 和 SCoT 的分类网络	36
3.4 基于 ShuffleNet 和 GhostNet 的网络轻量化	38
3.4.1 经典轻量级网络结构	38

3.4.2 基于 ShuffleNet 和 GhostNet 改进的轻量化卷积模块 .....	39
3.5 本章小结 .....	41
<b>4 实验结果分析与模型评估 .....</b>	<b>43</b>
4.1 实验环境 .....	43
4.2 数据与实验准备 .....	43
4.2.1 运动想象脑电图数据集 .....	43
4.2.2 数据预处理 .....	46
4.2.3 实验准备 .....	48
4.3 实验设计 .....	48
4.3.1 DIS-Net 实验设计 .....	48
4.3.2 LS-Net 实验设计 .....	51
4.3.3 各模块消融实验 .....	51
4.3.4 超参数优化实验 .....	51
4.3.5 BCI Competition IV Dataset 2A 数据集上的对比试验 .....	51
4.3.6 BCI Competition IV Dataset 2B 数据集上的对比试验 .....	51
4.4 本章小结 .....	51
<b>5 总结与展望 .....</b>	<b>53</b>
<b>参考文献 .....</b>	<b>55</b>
<b>致谢 .....</b>	<b>63</b>

# 1 绪论

## 1.1 研究背景及意义

大脑是一个结构和功能都很复杂的器官，是人体内外环境信息获得、存储、处理、加工及整合的中枢<sup>[1]</sup>，其神经活动蕴含着丰富的信息，直接反映了人类的思维、情绪和行为意图。对大脑的研究，不仅可以防止大脑的衰退以及脑疾病的产生，而且可以通过模拟大脑促进人工智能的发展<sup>[2]</sup>。

现今世界，各国政府及科研机构都高度重视脑科学研究的发展，美国于 2013 年启动“创新性神经技术大脑研究”计划，旨在开发和应用新的工具和技术，彻底改变人类对大脑的理解<sup>[3]</sup>。日本于 2014 年启动“脑/思维”计划，研究集中在三个领域：狨猴大脑的结构和功能图谱，用于脑图绘制的创新神经技术，以及人类脑图谱<sup>[4]</sup>。欧盟于 2013 年启动“人类脑计划”，最初目标为通过超级计算机对人脑进行模拟，该项目已于 2023 年 9 月结束，其在神经科学领域取得了多项重大进展<sup>[5]</sup>。中国于 2021 年将“脑科学与类脑研究”列为科技创新 2030 重大项目<sup>[6]</sup>，正式启动了“中国脑计划”，其内容包括认知神经机制的基础研究，脑疾病诊断和干预的转化研究以及脑启发智能（类脑）技术<sup>[7]</sup>。

脑机接口（Brain Computer Interface, BCI）是脑科学与信息科学交叉产生的新兴学科领域，研究如何在大脑与外部设备之间建立直接的通信和控制通道，实现脑与设备之间的双向信息传输。脑机接口系统的目标是将人类的思维、意愿或行动指令实时转化为可识别的控制信号，从而使人类可以直接由大脑而非神经肌肉通路来控制计算机或设备，双向脑机接口系统不仅可以实现大脑控制，也为通过神经接口调节中枢神经系统提供了一种可能的方案<sup>[8]</sup>。脑机接口的基础在于采集大脑活动产生的生物信号，例如脑电图（Electroencephalography, EEG）、功能性磁共振成像（Functional Magnetic Resonance Imaging, fMRI）、近红外光谱（Near-Infrared Spectroscopy, NIRS）等，这些生物信号从不同角度反映了大脑的认知过程、情绪状态和意图表达，其中，脑电图以其实时性、灵活性、便携性、低成本、非侵入性、高时间分辨率等优点，在脑机接口研究领域中受到了广泛的应用。

运动想象（Motor Imagery, MI）是脑机接口研究的主要方向之一，其表征的是一种运动意图，即个体在不实际执行物理动作的情况下，在大脑中想象执行特定动作的一种心理过程。运动想象的研究源于对大脑功能区域的认知神经科学探索，研究发现，当人们想象执行某个动作时，即使身体并未实际做出该动作，大脑中的特定脑区仍会有所激活。这种现象为基于运动想象的脑机接口系统提供了理论基础，通过识别和解码运动想象相关的脑电信号，可以基于使用者的自主意识实现对辅助设备的非侵入式控制，其在运动功能代偿、运动功能修复<sup>[9]</sup>、智能人机协作、认知神经科学研究、游戏和虚拟现实

等领域具有广阔的应用前景。

近年来，随着认知神经科学的持续发展以及各国政府对脑科学的日益重视，运动想象研究领域逐渐吸引了广泛的关注。在基于脑电图信号进行运动想象分类的研究中，鉴于脑电信号固有的复杂性与异质性，传统机器学习方法通常需要依据神经科学领域的先验知识进行人工特征设计，随着深度学习的迅速发展，研究者们开始将深度神经网络应用于运动想象分类任务中，期望能够自动化提取和学习脑电图信号中蕴含的潜在特征，然而，受限于数据集规模、脑电数据质量、实时响应性能等因素，深度学习在该领域的应用效果尚未达到理想的水平。

综上所述，尽管运动想象研究已积累了一系列有价值的科研成果，但是依然存在一定的进步空间。目前基于脑电图的运动想象分类任务仍然存在以下问题：

(1) 当前方法的实施常受限于多种条件，如特征提取依赖于专业神经科学知识与实践经验，对滤波的特定频率范围处理要求严苛，以及需要较高的电极采样密度。然而，在实际应用如家庭级和个性化的运动想象脑机接口系统中，往往不具备专家校验、标准滤波、高时空分辨率、高计算性能等理想条件，导致在实际场景的普适性上存在局限，应用场景较窄。

(2) EEG 信号固有的非平稳性和被试特异性导致现有分类方法在不同个体间的性能表现差异较大，即使在部分被试上达到较高识别精度，但在其他被试上可能显著降低，这阻碍了运动想象分类模型在广泛人群中的稳定应用。因此，仍然需要进一步提升分类方法的精确度、稳健性和一致性，从而在各类被试中获得更为满意的表现，取得较为稳定均衡的性能。

(3) 基于深度学习的运动想象脑电图分类模型通常拥有庞大的参数规模，这在计算资源有限的边缘设备上运行效率低下，影响实时响应性能，从而制约了运动想象脑机接口系统的普及推广。因此，有必要研发对模型的参数量进行精简，以实现性能与效率的优化。

(4) EEG 数据集因隐私保护、被试生理心理状态变化等因素，普遍具有规模偏小且样本多样性不足的特点。尽管数据增强是应对小规模数据集的有效策略，但针对具有被试特异性的 EEG 信号，某些增强方法可能破坏原始信号中蕴含的有价值信息。此外，数据增强过程会加重训练阶段的数据处理压力，延长训练周期，对运动想象脑机接口系统的实时响应性能有所制约。因此，如何在有限的小规模数据集上实现高效且稳定的性能提升，是需要考虑的问题。

## 1.2 国内外研究现状

不同于常规的时间序列数据，运动想象脑电图数据具有独特的生理学特征，因此，对基于神经科学先验知识进行的特征提取研究进行回顾，有利于运动想象领域的深度神经网络的设计。因此，本节将从脑电图特征提取和深度神经网络应用两方面来系统性地

介绍运动想象领域的国内外研究现状。

### 1.2.1 基于先验知识的运动想象脑电图特征研究现状

传统机器学习中，对运动想象脑电图信号（MI-EEG）的处理通常分为预处理、特征提取和分类三个主要步骤<sup>[10]</sup>，其中，分类算法与其它领域内广泛采用的传统机器学习技术类似，因此，论文主要介绍与 MI-EEG 信号生理特性相关的预处理技术和特征提取方法的相关研究。

MI-EEG 信号的信息密度在不同通道和频段间呈现出显著差异，并且数据的采集极易受到设备、环境、个体生理状态等因素的干扰，从而产生大量噪声，因此，对 MI-EEG 信号进行预处理是有必要的。预处理步骤包括一系列操作，例如通道选择（为 MI 任务选择最有价值的 EEG 通道）、信号滤波（为 MI 工作选择最有意义的频率范围）、信号归一化（围绕时间轴对每个 EEG 通道进行归一化）、伪影去除（从 MI-EEG 信号中去除噪声）<sup>[10]</sup>、基线矫正（消除 EEG 数据漂移带来的影响）等。其中，伪影去除的经典方法是独立成分分析（Independent Component Analysis, ICA）和离散小波变换<sup>[11]</sup>。

MI-EEG 信号的特征主要分为三类：时域特征、频域特征和空间域特征<sup>[10]</sup>，此外，通过对原始数据的进一步加工和转换，可以构建一系列复合特征，如时频特征、时空特征、时频空特征等。

时域特征反映 EEG 信号随时间变化的特性，其包括均值、方差、标准差、峰度、偏度等统计量，此外，Bo Hjorth 提出了一种快速计算时变信号的三个重要特征的方法，即活动性、移动性和复杂性，统称为 Hjorth 参数<sup>[12]</sup>，Luke 等人使用卡尔曼滤波来处理 EEG 信号的不确定性<sup>[13]</sup>，考虑到 EEG 信号具有分形性质，Hsu 提出了一种将分形维数和离散小波变换相结合的方法<sup>[14]</sup>。空域特征反映 EEG 信号在不同脑电极上的分布情况，共空间模式（Common Spatial Pattern, CSP）是空域特征提取的经典方法，其核心思想是寻找一组最优的空间滤波器，通过这组滤波器对原始多通道 EEG 数据进行空间投影，从而最大化两类信号在投影后空间的方差差<sup>[15]</sup>。研究者对 CSP 算法进行了进一步优化，提出了滤波器组共空间模式（FBCSP）<sup>[16]</sup>、共稀疏空间谱模式（CSSSP）<sup>[17]</sup>、共空间谱模式（CSSP）<sup>[18]</sup>、小波共空间模式（WCSP）<sup>[19]</sup> 等方法，在多分类、异步脑机接口等领域取得了更好的效果。频域特征反映 EEG 信号在不同频带的功率，通常包括功率谱（Power Spectral）、功率谱密度（Power Spectral Density, PSD）、高阶谱（Higher-order Spectral）、微分熵（Differential Entropy）、傅里叶变换（Fourier Transform）等方法。Wang 等人提出了一种基于快速傅里叶变换（Fast Fourier Transform, FFT）的功率谱特征提取方法<sup>[20]</sup>，Kroupi 等人使用基于 Welch 方法的功率谱密度（Power Spectral Density）估计来进行脑电情感分析<sup>[21]</sup>，Herman 等人比较分析了不同的频率特征在运动想象分类中的效果，并证实了功率谱密度具有最好的鲁棒性<sup>[22]</sup>。时频特征由于结合了时域特征和频域特征，在基于非平稳 EEG 信号的脑机接口研究中获得了广泛的使用<sup>[23]</sup>，其主要包括匹配滤波（Matched Filtering, MF）、

自回归模型（Autoregressive Model, AR）、短时傅里叶变换（Short-time Fourier Transform, STFT）、小波变换（Wavelet Transform, WT）等方法，时频特征在提升了分类精度的同时，也提高了计算的复杂度。

尽管关于 EEG 信号特征提取的研究已积累了大量成果，然而传统预处理及特征提取方法仍存在着计算复杂度较高以及信息丢失的可能性，同时，EEG 信号的特征提取高度依赖神经科学领域的先验知识，从而限制了特征提取的自动化程度<sup>[10]</sup>。在这种背景下，深度学习被引入运动想象分类领域，以克服传统方法存在的问题。

### 1.2.2 基于深度学习的运动想象脑电图分类研究现状

随着深度学习的迅速发展，基于深度学习的方法开始被应用于运动想象脑电图分类任务中，并取得了良好的效果。基于深度学习的运动想象脑电图分类方法主要包括三个部分：数据预处理，网络输入模式，以及深度神经网络架构。

在数据预处理阶段，由于与运动想象相关的脑电图事件相关去同步化（Event-Related Desynchronization, ERD）和事件相关同步化（Event-Related Synchronization, ERS）主要发生在 Mu 节律（8-12Hz）和 Beta 节律（18-26Hz）中<sup>[10]</sup>，多数深度学习方法采用带通滤波技术来提取相关频率成分。然而，这种方法可能会干扰或削减脑电图信号中的有效信息。此外，部分研究通过人工或自动手段去除脑电信号中的伪迹，例如 Ma 等人<sup>[24]</sup>采用了自动伪迹去除工具箱<sup>[25]</sup>。

在网络输入模式方面，研究<sup>[26-31]</sup> 使用 CSP、FBCSP、PSD、FFT、离散小波变换（Discrete Wavelet Transformation, DWT）等算法进行特征的提取，并将提取的特征作为深度神经网络的输入。然而，这些特征提取算法在一定程度上依赖于神经科学的先验知识和专家经验，并且人工设计的算法可能导致对某些重要信息的遗漏。Xu 等人<sup>[32]</sup> 使用 WT 算法将脑电图信号转化为频谱图，以更全面地表征时域和频域特征。Miao 等人<sup>[33]</sup> 则使用 FFT 算法获取空频图，表征脑电图信号中的空域特征和频域特征。Li 等人<sup>[34]</sup> 更进一步地将不同通道的数据进行组合，形成了时域、空域、频域的三维表示。尽管将脑电信号转换为图像形式有助于更好地表征 EEG 信号中不同域的信息，但这也会加剧计算资源的消耗，延长训练的时间。Amin<sup>[35]</sup> 等人基于深度学习自动提取特征的概念，将 EEG 原始数据直接传入深度神经网络中进行训练。近年来，端到端的深度学习网络在运动想象脑电图分类任务中日益受到重视。

对于深度神经网络架构，卷积神经网络（Convolutional Neural Network, CNN）、循环神经网络（Recurrent Neural Network, RNN）等深度网络架构在运动想象脑电信号分类任务中被广泛应用。

卷积神经网络是深度学习中最常见的模型架构之一，擅长捕捉信号的局部特征，并随着网络层次加深逐步提取更高层次的抽象特征。在运动想象分类任务中，CNN 不仅可以处理时频图、空频图和时空频三维图等转换后的数据，还可以直接应用于原始 EEG

数据。Schirrmeister 等人<sup>[36]</sup> 提出了一系列基于 CNN 的网络架构，包括 ShallowConvNet 和 DeepConvNet，它们通过堆叠轴向卷积层替代传统的卷积层，实现了无需预先进行人工特征提取的情况下对脑电信号进行分类。Lawhern 等人<sup>[37]</sup> 提出了 EEGNet，将深度卷积和可分离卷积引入网络架构中，构建了一种紧凑的通用型模型。Riyad 等人<sup>[38]</sup> 在 EEGNet 的基础上提出了 MI-EEGNet，针对运动想象分类任务进行了进一步的改进。

卷积神经网络是深度学习中最常见的模型架构之一，通过卷积和池化操作来提取图像的局部特征，并通过全连接层将这些特征进行组合和分类。随着网络加深，CNN 获取的特征也越来越高级。CNN 在计算机视觉领域应用广泛，在运动想象分类任务中，CNN 被用来处理时频图、空频图、时频空三维图，也被用来处理 EEG 原始数据。Schirrmeister 等人<sup>[36]</sup> 提出了一系列基于 CNN 的网络架构，包括 ShallowConvNet 和 DeepConvNet，其结构较为简单，为卷积层的堆叠，对卷积核进行了修改，将原本的卷积核修改为轴向卷积，在不进行人工特征提取的情况下进行分类。Lawhern 等人<sup>[37]</sup> 提出了 EEGNet，将深度卷积和可分离卷积引入网络架构中，构建了一种紧凑的通用型模型。Riyad 等人<sup>[38]</sup> 在 EEGNet 的基础上提出了 MI-EEGNet，针对运动想象分类任务进行了进一步的改进。Mane 等人<sup>[39]</sup> 提出了 FBCNet，将 FBCSP 算法的思想融入网络设计中，通过多个窄带滤波器对原始脑电图进行滤波，得到 EEG 信号的多视图表示。Zhang 等人<sup>[40]</sup> 使用了多个分支进行多尺度的特征提取。Song 等人<sup>[41]</sup> 为了捕获脑电图信号中的长期依赖关系，将 Transformer<sup>[42]</sup> 模型引入 CNN 模型之后，直接处理经过一维时间和空间卷积层后的序列。Musallam 等人<sup>[43]</sup> 基于残差神经网络（Residual Network, ResNet）<sup>[44]</sup> 构建了一种专用于运动想象脑电图分类的模型。

循环神经网络是一类专门设计用于处理序列数据的神经网络架构，它的特点是网络内部包含循环结构或反馈连接，使得信息可以在时间步之间传递和累积，特别适合处理具有时间依赖性或顺序结构的数据。RNN 的变体有长短期记忆网络（Long Short-Term Memory, LSTM），门控循环单元（Gated Recurrent Unit, GRU）等。LSTM 模型是一种能够学习长期关系的 RNN 网络，克服了传统 RNN 的梯度消失问题，其在自然语言处理领域应用广泛，在运动想象任务中，LSTM 也被用来处理 EEG 序列数据，获取其中蕴含的时域信息。Ma 等人<sup>[45]</sup> 使用滑动窗口对 EEG 原始数据进行截取从而进行数据增强，随后采用 LSTM 进行运动想象的分类。Wang 等人<sup>[46]</sup> 提出了一种基于一维聚合近似（One Dimension-Aggregate Approximation, 1d-AX）的 LSTM 模型进行运动想象脑电图分类任务。LKumar 等人<sup>[47]</sup> 提出了一种使用 CSP 算法进行特征提取，线性判别分析（Linear Discriminant Analysis, LDA）进行特征约简，支持向量机（Support Vector Machine, SVM）作为分类器的 LSTM 模型，在 GigaDB<sup>[48]</sup> 和 BCI Competition IV-1<sup>[49]</sup> 数据集上分别取得了 68.19% 和 82.52% 的准确率。

### 1.3 研究内容

针对现阶段运动想象脑电图分类任务中仍然存在的问题，论文的研究目标为基于现有方法提出一种改进的新模型，其应当具有以下特点：

(1) 端到端：只进行很少数据预处理的端到端模型，旨在减轻对神经科学先验知识和专家经验的依赖，简化数据处理步骤，降低计算成本，并拓宽模型在不同应用场景下的适用范围。

(2) 提升分类精度与泛化能力：在现有方法的基础上进一步提高分类精度，确保模型能在多样化的被试群体中都展现出优秀且稳定的性能，无论被试个体的特异性如何，都能获得较为理想的分类效果。

(3) 轻量化：具有较少的参数规模，使得模型能够在计算资源有限的边缘设备上高效运行，从而提升实时响应性能，促进运动想象脑电图分类技术在实际应用中的部署与普及。

为了实现研究目标，论文基于现有方法进行改进，

### 1.4 论文组织结构

论文主要分为以下五个章节：

第一章，绪论。首先介绍了运动想象脑电图分类研究的背景与意义，其次梳理了该领域的国内外研究进展，分别围绕基于神经科学先验知识的运动想象脑电图特征提取研究和基于深度学习方法的运动想象脑电图分类技术研发进行了介绍。最后介绍了论文的研究目标、研究内容以及论文的整体组织结构。

第二章，运动想象脑电图分类和深度神经网络基础。首先介绍了与运动想象脑电图分类任务相关的神经科学基础知识，其次介绍了卷积神经网络、循环神经网络、注意力机制等深度神经网络的相关基础知识，最后介绍了运动想象脑电图分类任务中的一些特殊网络结构。

第三章，基于双分支特征融合和注意力机制的运动想象脑电图分类网络构建。分析了现阶段运动想象脑电图分类任务中仍然存在的问题，并针对这些问题，提出了进一步的改进方法。

第四章，实验结果分析与模型评估。介绍了实验所需的软硬件环境、所使用的运动想象脑电图数据集，以及运动想象脑电图分类任务的评价指标体系。其次介绍了运动想象脑电图数据处理方法，然后依次展开实验过程和结果讨论，包括对比实验、消融实验和泛化性实验，验证了论文提出的方法在运动想象脑电图分类任务上的有效性。

第五章，总结与展望。对论文研究工作进行了全面总结，分析了所提方法的优点与局限性，并在此基础上提出了未来可能的研究方向和改进策略。

## 2 运动想象脑电图分类和深度神经网络基础

### 2.1 脑电生理基础

#### 2.1.1 人脑结构与运动想象

大脑作为神经活动的中枢处理器，其活动产生的多元生物电信号揭示了人体的认知行为、感官体验、情绪调控等多种思维活动和行动指令。通过现代神经科学技术，如脑电图（EEG）、功能性磁共振成像（fMRI）、近红外光谱成像（NIRS）等手段，人类能够探测并解读这些生物信号，从而了解大脑内部的运作机制。

大脑皮层是大脑最外部的灰质结构，承载了复杂的高级认知功能。大脑皮层中与运动相关的结构主要包括初级运动皮层（Primary Motor Cortex, M1）、辅助运动区（Supplementary Motor Area, SMA）、运动前区（Premotor Cortex, PMC）、顶叶皮层（Parietal Lobe）、前额叶皮层（Prefrontal Lobe, PFC），此外，位于基底神经节的纹状体（Striatum）也参与其中。运动想象（Motor Imagery, MI）是一种心理活动，指的是在没有实际执行肢体运动的情况下，个体在脑海中模拟或重现某项运动的过程，研究发现，当个体执行运动想象任务时，这些脑区仍然会有激活，其中，主要激活的是辅助运动区，辅助运动区在运动执行任务以及运动想象任务中与运动前区、感觉运动皮层区、默认网络结构都存在双向连接<sup>[50]</sup>，并且存在指向纹状体的单向连接<sup>[51]</sup>。相对于运动执行任务，运动想象任务中主运动皮层也存在激活，但激活程度不及运动执行任务<sup>[50][52]</sup>。大脑皮层中各运动区的位置如图 2.1<sup>[53]</sup> 所示。

#### 2.1.2 脑电图信号及其特性

脑电图信号（Electroencephalogram, EEG）由皮质内大量神经元突触后电位同步总和形成，是很多神经元共同活动的结果<sup>[54]</sup>。EEG 信号可以分为深部 EEG，皮层 EEG 和

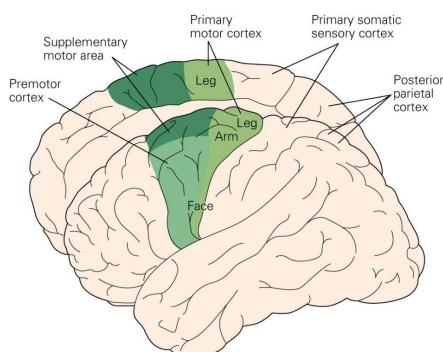


图 2.1 大脑皮层运动相关区域

头皮 EEG<sup>[55]</sup>。相比深部 EEG 和皮层 EEG，从头皮采集的 EEG 信号需穿透颅骨和头皮组织，因此具有相对较低的信号质量，但头皮 EEG 无需进行任何开创性手术，极大地降低了医疗风险，保证了高度的安全性，并且实施便捷，获取数据更为简易。因此，头皮 EEG 成为了神经系统疾病诊断、脑功能研究等领域广泛应用的首选工具。后文中所提到的 EEG 信号，如无特殊说明，都是指头皮 EEG。

EEG 信号的采集设备为脑电采集系统，其样式如图 2.2a 所示，通过在受试者头上放置单个或多个脑电极（或称通道），以连续、实时的方式记录一段时间内的大脑皮层产生的电信号。图 2.2b 展示了脑电极在头皮上的一种排列方式。EEG 信号的采集方式决定了其数据的形状，通常为脑电极（通道）与时间构成的二维数据，其可视化如图所示。对于单通道采集的数据，EEG 信号可视化以电位为纵轴，时间为横轴，表示该通道上电位随时间的变化，图 2.3 为包含 16 个通道的 EEG 信号可视化，可视化中相邻的通道在空间上未必具有相邻关系。

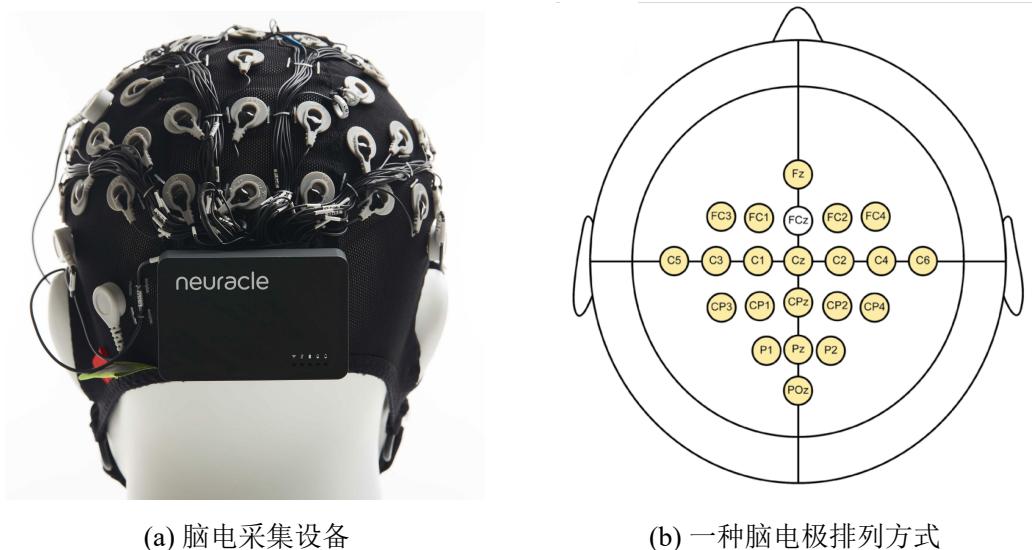


图 2.2 脑电信号采集

正常脑电信号是多个具有生物学意义的神经振荡节律的组合<sup>[55]</sup>。神经科学的研究发现，根据脑电信号的频率范围可以划分出五种主要节律：

(1) Delta：频率范围为 0.5-4Hz。研究发现，Delta 波的起源主要有两个，分别是皮层和丘脑，后者主要与睡眠相关<sup>[56]</sup>。Delta 节律主要出现在深度睡眠阶段，尤其是在儿童和婴儿的脑电图中尤为明显。成年人在某些病理状态下也可观察到增强的 Delta 节律，如严重的器质性脑疾病。

(2) Theta：频率范围为 4-8Hz。Theta 节律不仅出现在轻度睡眠和快速眼动睡眠阶段，还可见于青少年和儿童清醒时的脑电图中，以及成年人在专注、疲劳或焦虑状态下的头顶前部中线位置<sup>[57][58]</sup>。

(3) Alpha：频率范围为 8-13Hz。Alpha 节律依据起源位置的不同可分为两大类：一

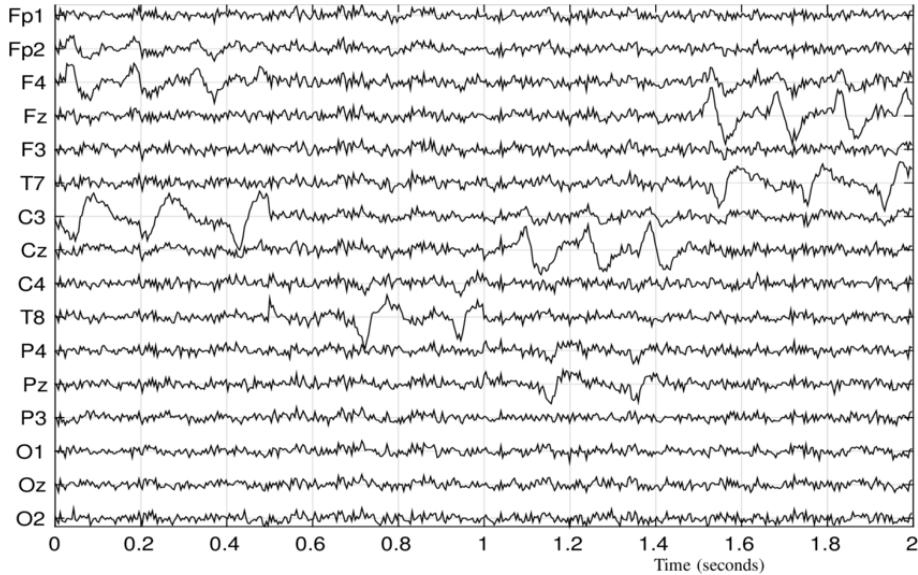


图 2.3 EEG 信号可视化

类是源自初级躯体感觉皮层（Primary Somatosensory Cortex, S1）的 Mu 节律（或称运动 Alpha 节律）<sup>[59]</sup>，手部运动、运动想象等活动可以阻断 Mu 节律，而在肌肉放松状态下则会增强 Mu 节律；另一类则是来源于视觉皮层或枕叶的 Alpha 节律，通常在个体放松、闭眼精细状态出现，而在睁眼或注意力集中时受到抑制，表现为振幅减小。此外，有研究指出，脑力劳动的增加也将导致 Alpha 节律活动的减弱<sup>[55]</sup>。

(4) Beta: 频率范围为 13-30Hz。Beta 节律根据分布区域的不同可分为两大类：一类是在感觉运动皮层区域表现最强的运动 Beta 节律（Rolandic Beta），其活动强度受运动任务影响，在运动想象、运动准备和运动执行过程中，Rolandic Beta 节律的幅值功率都会有相应变化；另一类是在额叶区域表现最强的额叶 Beta 节律（Frontal Beta），此类节律主要出现在清醒、警觉、专注以及执行认知任务时的大脑活动中，其强度的增加通常与认知活动和心理负荷的上升相关联。

(5) Gamma: 频率范围通常定义为超过 30Hz，有时也指 25Hz 至 100Hz 甚至更高的频率范围。Gamma 节律广泛参与到认知处理、意识维持、注意力集中、感知整合和记忆编码等复杂脑功能过程中，被认为与高级认知功能和信息整合紧密相关。Gamma 节律可在多个脑区观察到，包括但不限于前额叶、顶叶和颞叶区域。

需要说明的是，各类脑电波的频率范围并不是绝对分割的，而是存在着一定的重叠与交织。在实际的大脑生理活动中，某一时刻或时间段内的脑电信号往往包含了多种不同的节律形态，不仅限于 Delta、Theta、Alpha、Beta、Gamma 五类，例如，Rolandic Beta 节律和 Frontal Beta 节律可以同时出现。实际上，脑电极采集到的脑电信号通常是不同频率节律相互交织、叠加而成的复杂组合。

除却具有脑生理活动信息的 EEG 节律，脑电极会同时采集到各类噪声信息，称之为 EEG 伪迹。常见的 EEG 伪迹有以下几种：

(1) 眼动伪迹：当眼睛移动（如眨眼、扫视等）时，眼周的肌肉和眼睑的电位变化会在电极上产生强烈的电位波动。这种伪迹在额部电极附近特别明显，特征是突然出现的尖峰或方波形状的电位变化。眼动信号可以通过眼电图（Electrooculogram, EOG）记录。EOG 的幅度一般是 EEG 信号的几倍，频率通常分布在 3-16Hz，幅值通常分布在 0.5-4mV。

(2) 肌电伪迹：肌电伪迹主要来自于头颈部及面部肌肉的伸展和收缩，如吞咽、呼吸、讲话等。肌电伪迹通常表现为高频、大振幅的波动，与实际脑电活动相比更加剧烈和不稳定。肌电信号可以通过肌电图（Electromyogram, EMG）记录。EMG 的频率在 0-200Hz 广泛分布。

(3) 心电伪迹：脑电极分布在脑部血管附近，心脏搏动会在头皮电极上产生明显的电位波动，尤其是在靠近耳朵后部的乳突区域。心电伪迹通常表现为周期性出现的正弦波形，与心脏搏动频率（大约每分钟 60-100 次）相符。心电信号可以通过心电图（Electrocardiogram, ECG）记录。ECG 的波形类似 EEG 波形，频率在 0.05-100Hz 广泛分布。

(4) 工频伪迹：工频干扰来源于电源线和其他电气设备释放的 50Hz（欧洲和亚洲大部分地区）或 60Hz（北美和其他部分地区）的交流电噪声。这种干扰在 EEG 信号中表现为稳定的、与电源频率相同的周期性波动。

为了减少 EEG 信号中的伪迹干扰，通常会在数据预处理阶段运用一系列针对性算法和技术。例如，利用独立成分分析识别并排除源于眼动和肌电活动的伪迹成分，通过低通滤波抑制心电伪迹在特定频率范围内的影响等。然而，需要说明的是，即便是基于信号处理技术和神经科学先验知识，伪迹的去除过程仍有可能导致某种程度的真实脑电信息丢失。尤其是在处理复杂、微妙的脑电活动时，很难做到既能彻底清除伪迹又不损害有价值信号的完整性。

综上所述，EEG 信号是一种蕴含丰富信息却又存在诸多问题的复杂生物电信号，它不仅涵盖着从 Delta 至 Gamma 等多种频率范围的脑电节律，且这些节律在实际记录中存在交错重叠的现象。此外，EEG 信号容易受到多种噪声源的显著影响，包括眼动伪迹、肌电伪迹、心电伪迹等生理伪迹以及工频干扰等非生理伪迹。鉴于 EEG 信号的独特生理特征和采集特点，它展现出一些鲜明的特性：

(1) 低信号强度。EEG 信号采集的是神经元活动产生的微弱电压波动，其强度一般在  $\mu\text{V}$  级别 ( $5-100\mu\text{V}$ )<sup>[60]</sup>。这使 EEG 信号易受噪声影响，提高了信号处理和解析的难度。

(2) 宽带特性，低信噪比。EEG 信号所蕴含的有价值信息具有显著的时间动态性，即这些信息会随时间持续演变并在整个时间域内传播，这使得 EEG 信号在频谱分析中呈现出宽广的频率范围，其中包含的主要成分通常分布在大约 0.5 至 45Hz 的频率范围内<sup>[55]</sup>，涵盖了 Delta、Theta、Alpha、Beta 及部分 Gamma 节律。然而，由于伪迹噪声普

遍存在且分布广泛，EEG 信号中蕴含的有用信息时常与伪迹及其他信息在不同频带上相互交叉混叠。与此同时，由于人体脑电信号本身强度微弱，加之复杂的生理和心理状态变化，以及检测技术在实际应用中的不稳定性，进一步导致了在实际采集的 EEG 信号数据中，不同脑电成分之间易于互相干扰和混淆。这些现象显著降低了 EEG 信号的信噪比，即有用信号与噪声的比例较低。这为 EEG 信号的分析带来了困难，难以完整地将有用信号从背景噪声中剥离，并且需要依赖经验和神经科学先验知识，耗费相当的开销。

(3) 非平稳性，被试特异性。不同于平稳随机信号，EEG 信号的统计特性随时间发生变化，其频率、幅度等特征在不同任务条件下呈现出显著的动态变化，表明 EEG 信号具有明显的非平稳性。与此同时，不同个体（被试）所产生的 EEG 信号在频率成分、幅度强度、响应时间等方面通常表现出个体间的差异性；不仅如此，即使是同一被试在不同时间段内执行同样的任务，其所产生的 EEG 信号也可能存在变化。因此，EEG 信号具有显著的被试特异性 (Subject-Specificity)，意味着每位被试的脑电活动模式在很大程度上是个体独有的。这使得很难使用一种固定的方法对 EEG 信号进行分析。

(4) 非线性。EEG 信号不是单一过程的结果，而是多种脑内过程或同源过程非线性耦合与叠加的产物，呈现出显著的非线性特征。这种非线性特性使得各种信号成分之间的相互作用和影响更为复杂化，极大地提升了 EEG 信号分析的复杂度和解析难度。

(5) 多通道，时空分布不均。在实际采集过程中，EEG 信号通常利用多个脑电极（通道）进行同步记录，形成了多通道数据属性。同时，EEG 信号的时空分布不均匀，大脑在执行特定功能时，会在特定时间和脑区产生显著的电压波动，例如，进行眼球运动时，在额部电极附近会产生即时且明显的眼电活动。脑电极的分布通常使用国际 10-20 标准导联确定，其中，“10”和“20”指相邻电极之间的实际距离设定为头骨前后或左右总距离的 10% 或 20%，标准 10-20 导联系统包含了 21 个电极，这些电极捕获的信号在不同大脑活动中蕴含的信息量不尽相同。此外，EEG 信号在时间域的数据量往往要高于空间域的数据量，且时间域蕴含的有效信息更为丰富。这种时空分布的不均衡性提升了筛选和提取有效信息的困难性。

EEG 信号具有时间域、空间域、频率域三个维度的特征。在时间域层面，EEG 信号揭示了神经活动随时间推移的动态变化过程；在空间域层面，EEG 信号揭示了神经活动在不同头皮电极位置上的强度分布差异，反映了大脑不同区域的功能活动；而在频率域层面，EEG 信号反映了信号中蕴含的不同脑电节律。脑电采集设备所采集到的原始 EEG 信号直观地反映了时间域和空间域的特征，如图 2.3 所示。频率域的特征则需要通过快速傅里叶变换等频谱分析技术对原始数据进行转换。相较于直接处理原始的 EEG 信号，提取频率域特征通常需要更多的计算资源和处理时间，耗费更高的开销。

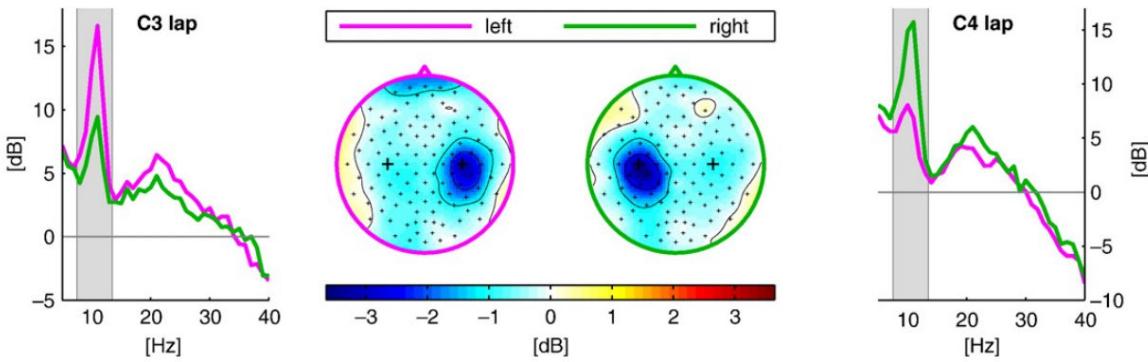


图 2.4 左右手运动想象的 ERD 现象

### 2.1.3 脑电图信号与运动想象

事件相关去同步化(Event-Related Desynchronization, ERD)和事件相关同步化(Event-Related Synchronization, ERS)是在研究EEG信号过程中发现的两种脑活动现象，它们反映了大脑在特定任务或事件触发后，特定频段脑电活动的动态变化。

(1) 事件相关去同步：ERD是指在执行一项特定的任务，如运动想象或实际运动时，原本在一定频段上占主导地位的同步脑电活动（即电位振荡趋于一致）出现了减少或消失，表现为脑电功率的下降。这一现象通常与大脑皮层的兴奋性增加和特定区域活动增强有关。在运动想象或实际运动时，初级运动皮层和辅助运动区的Alpha节律(Mu节律)、Beta节律通常会出现ERD现象。

(2) 事件相关同步：ERS与ERD相反，是指在特定任务或事件发生后，某一频段的脑电活动同步性增强，表现为脑电功率的升高。ERS可能表示大脑某些区域的活动暂时减少或进入了一种休眠状态，为其他区域的活动让渡资源，或者代表了大脑在执行某些认知或精神任务时的资源重新配置。在运动想象或实际运动时，ERS现象同样出现在Alpha节律(Mu节律)和Beta节律中。

在进行单侧肢体的运动想象任务时，大脑呈现出不对称的激活模式：对侧大脑半球（即与执行想象运动的肢体相对应的脑区）通常会出现更为显著的ERD现象，意味着该区域的特定频段脑电活动强度减弱<sup>[61]</sup>。与此同时，执行运动想象任务的同侧大脑半球则会出现一定程度的ERS现象。实际上，在运动想象任务结束后的放松阶段，ERS现象更为明显，此时大脑可能在先前活跃区域呈现出电活动同步性的增强，这一现象反映了大脑从运动想象状态向静息状态转变时的活动调整和资源分配。进行左右手运动想象时，脑功率谱密度变化如图2.4所示，当想象左手运动时，大脑皮层右侧(C4电极附近)出现ERD现象，相关区域能量减小；当想象右手运动时，大脑皮层左侧(C3电极附近)出现ERD现象，相关区域能量减小。

综上所述，与运动想象相关的脑电图信号主要是出现在感觉运动皮层的Mu节律和Beta节律，进行运动想象时，主要会出现ERD、ERS这两种脑电活动现象。

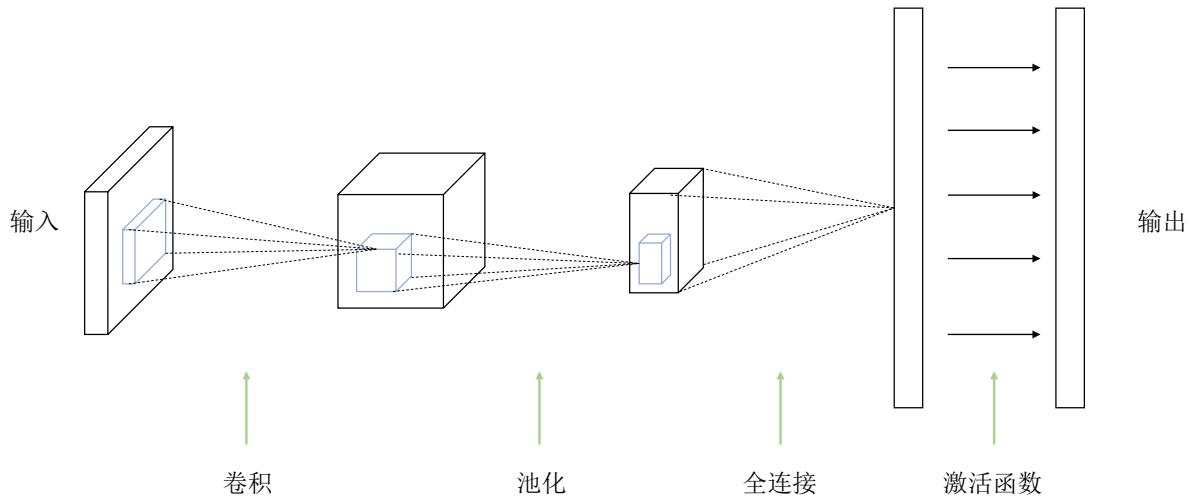


图 2.5 多分类卷积神经网络基本结构

## 2.2 深度神经网络基础

深度神经网络（Deep Neural Network, DNN）是一种模拟人脑神经网络设计的多层非线性模型，在计算机视觉、自然语言处理、语音交互等领域有着广泛的应用。相较于传统的神经网络，深度神经网络通过多个隐藏层逐层对输入数据进行特征提取和抽象，模型复杂度和表达能力显著提高，能够拟合更复杂的非线性关系，取得更为优异的性能和效果。

### 2.2.1 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）是深度学习领域经典且重要的模型，其设计灵感来自于生物学中对动物视觉系统的理解，设计目的是模拟人类的视觉处理过程。CNN 在计算机视觉领域取得了巨大成功，广泛应用于图像识别、图像分类、物体检测、语义分割、视频分析等任务中，其优势在于能够自动学习图像特征，无需进行人工特征提取，从而简化了流程，并提高了性能。

CNN 的结构主要包括输入层（Input Layer）、卷积层（Convolutional Layer）、池化层（Pooling Layer）、激活函数（activation Function）、全连接层（Fully Connected Layer）、输出层（Output Layer），此外还可包括 Dropout 层、归一化层（Normalization Layer）等。一个简单的用于分类任务的 CNN 网络结构如图 2.5 所示。论文将对其中的重要结构进行说明。

#### (1) 卷积层

卷积层是 CNN 的核心组成部分之一，它对人类视觉系统中的感受野（Receptive Field）进行模拟，对图像的局部特征进行提取。卷积层使用一组可学习的滤波器（Filter），或称卷积核（Kernel）在输入数据上进行滑动，并与覆盖窗口内的数据做元素间乘法和

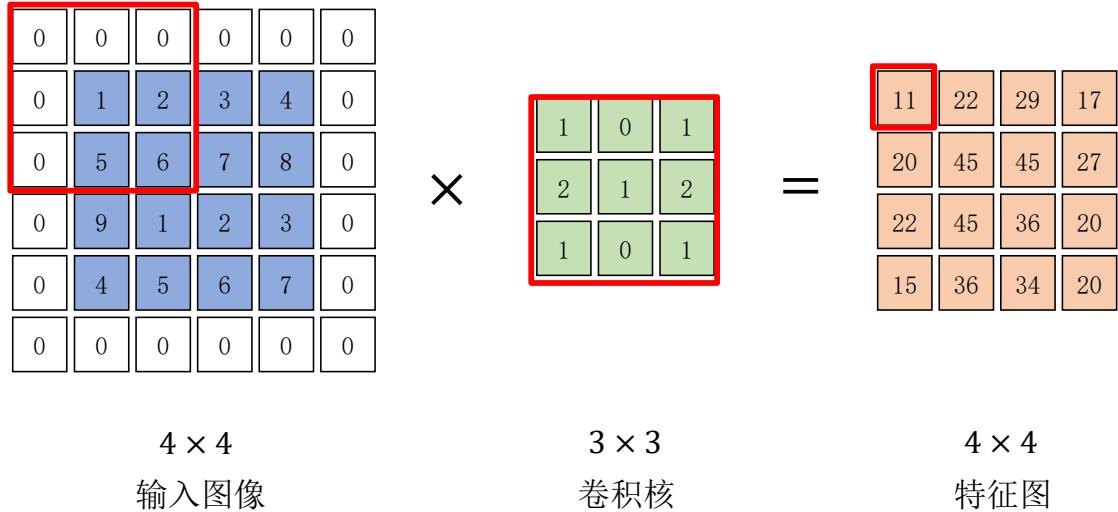


图 2.6 二维卷积运算

加法运算，得到一个特征图（Feature Map）或激活图（Activation Map），这种计算过程被称为卷积。卷积操作中有许多可设置的参数，如卷积核的大小（Kernel Size），卷积核滑动的步长（Stride），数据填充（Padding）等。卷积核的大小决定了局部感受野的大小，较大的卷积核可以捕获更大范围的上下文信息，而较小的卷积核在捕获局部细节上更具优势；步长决定了卷积核在数据上移动的间距，当步长为 1 时，进行逐元素的卷积；填充是在数据外缘填充额外的数据，一般用于对数据大小进行扩充，使得卷积后能够获得与原数据大小相同的特征图。图 2.6 展示了在一个  $4 \times 4$  的输入数据上进行卷积的过程，卷积核大小为  $3 \times 3$ ，步长为 1，填充为 1，红框部分的计算过程如公式 2.1 所示。

$$11 = 1 \times 0 + 0 \times 0 + 1 \times 0 + 2 \times 0 + 1 \times 1 + 2 \times 2 + 1 \times 0 + 0 \times 5 + 1 \times 6 \quad (2.1)$$

在处理如彩色图像这类多通道数据时（如 RGB 三通道），卷积核的通道数与输入图像的通道数相匹配，卷积核的每个通道分别与输入图像的相应通道进行卷积操作，然后将各个通道的结果进行逐元素相加，形成最终的输出特征图，其过程如图 2.7 所示，其中使用了一个具有三通道的卷积核。在经过卷积操作生成的特征图中，任一元素实质上是由上一层特征图中的  $N \times N$  个相邻元素加权组合而成，使得该元素具有  $N \times N$  大小的感受野。随着卷积层的叠加，感受野逐渐扩大，意味着在深层的特征图中，一个元素能够捕获到输入中更广阔区域的信息，提取到的特征更为高级和抽象。图 2.8 展示了感受野随层次加深的变化。

## (2) 池化层

池化层是 CNN 的重要组成部分之一，用来对特征图进行降维，从而在保留重要信息的同时减少参数的数量。池化层的主要类型为最大池化（Max Pooling）和平均池化（Average Pooling），此外还有最小池化（Minimum Pooling）、L2 范数池化（L2-Normalized Pooling）等，但在实际应用中并不常见。

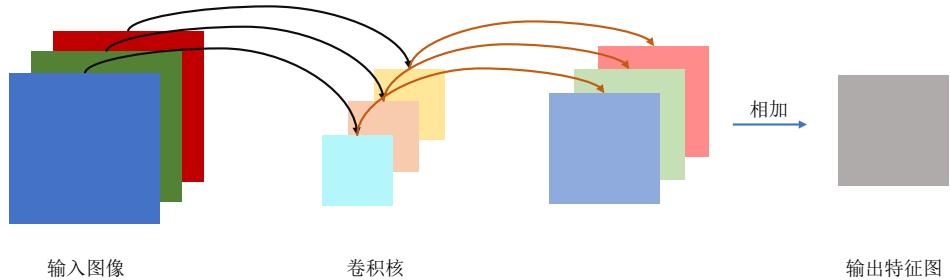


图 2.7 多通道卷积运算

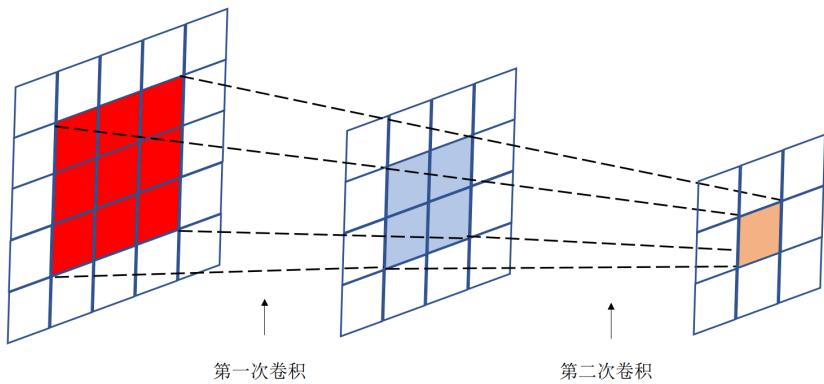


图 2.8 感受野随层次加深示例

在最大池化中，池化层将输入特征图中每个区域的最大值作为输出，能够保留特征图中最为明显的特征；在平均池化中，池化层将输入特征图中每个区域的平均值作为输出，相较于最大池化，平均池化更为平滑，能够降低特征图中的噪声，但可能会丢失一些细节信息。池化层中的可调参数通常是池化核的大小和步长，池化核的大小决定了池化操作所覆盖的局部区域的大小，步长决定了池化核每次移动的距离，二者共同决定了输出特征图的空间维度。图展示了最大池化和最小池化的计算过程，其中，池化核的大小都为 $2 \times 2$ ，步长都为1。在CNN中，池化层通常与卷积层交替堆叠，构成特征提取的基本结构。

### (3) 激活函数

激活函数是神经网络中的一种非线性函数，它对神经元的输出进行非线性变换，使得神经网络能够学习和表示更加复杂的函数关系。常用的激活函数有以下几种：

a. Sigmoid 函数：Sigmoid 函数将输入信号映射  $(0, 1)$  范围，具有平滑、连续、连续可微等特点，常用于二分类任务中。其公式为：

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2.2)$$

Sigmoid 函数的缺点在于：当输入较大或较小时，Sigmoid 的导数趋近于 0，导致反向传播时神经元的权重难以更新，容易出现梯度消失现象；Sigmoid 的输出不以 0 为中心，导

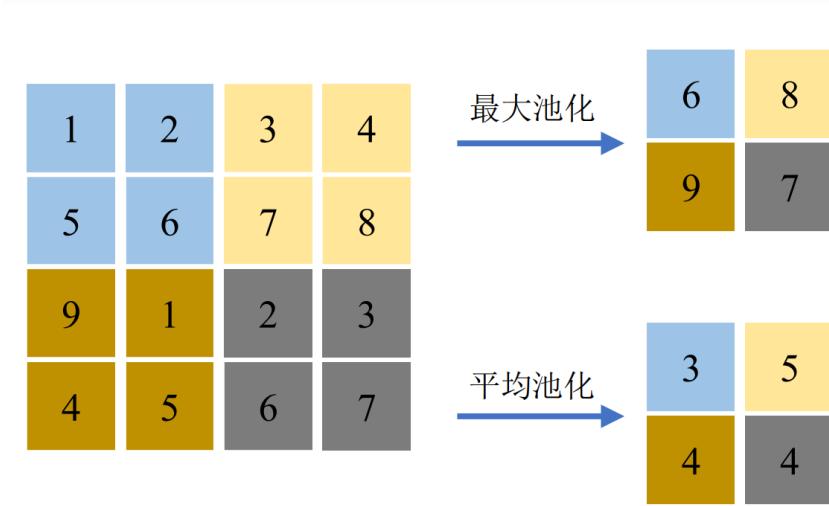


图 2.9 最大池化与平均池化

致收敛速度变慢；函数中的指数级运算使得计算成本较高。

b. Softmax 函数：Softmax 函数将输入信号映射  $(0, 1)$  范围，并且使得它们的总和为 1，具有平移不变性、对数凸函数等特点，常用于多分类任务中。其公式为：

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}} \quad (2.3)$$

其中， $x_i$  是输入中的一个元素， $N$  是输入的维度。Softmax 函数的缺点在于，当某个类别的预测概率接近 1 时，对应的梯度几乎为 0，使得梯度下降缓慢，出现梯度饱和现象等。

c. ReLU 函数<sup>[62]</sup>：ReLU 函数将负数映射为 0，正数保持不变，具有运算简单、收敛迅速的特点，在主流模型中应用广泛。其公式为：

$$\text{ReLU}(x) = \max(0, x) \quad (2.4)$$

ReLU 函数的缺点在于，在输入为负值时，ReLU 的导数恒为 0，导致部分神经元无法更新参数，出现神经元坏死现象。

d. GELU 函数<sup>[63]</sup>：GELU 函数是一种基于高斯分布的激活函数，主要目标在于结合 ReLU 函数及其变种的优点，同时避免其中存在的一些问题，例如 ReLU 函数的神经元坏死，Sigmoid 函数的梯度消失现象等。GELU 的公式定义为：

$$\text{GELU}(x) = x\Phi(x) \quad (2.5)$$

其中， $P(X \leq x)$  表示  $x$  的高斯正态分布的累计分布函数，由于计算困难，一般使用公式进行近似计算。

$$\text{GELU}(x) = \frac{1}{2}x \left( 1 + \tanh \left( \sqrt{\frac{2}{\pi}} \cdot (x + 0.044715x^3) \right) \right) \quad (2.6)$$

GELU 的优点在于：具有更为光滑的导数，使得梯度可以更容易地进行传播，避免了零点导数不连续的问题，有助于解决梯度消失和梯度爆炸的现象；保持了 ReLU 在正数的线性增长特性，同时在负数也有非零梯度，解决了 ReLU 的神经元坏死现象；基于自然分布提出，能够更好地对人脑神经元进行模拟。鉴于这些优点，GELU 函数当前在越来越多的模型中得到应用，如无特殊说明，论文使用的激活函数都是 GELU 函数。

#### (4) 归一化层

归一化层是 CNN 的常见组成部分，能够减少因数据分布的变化而带来的内部协变量偏移、梯度消失、梯度爆炸等问题，用于提升训练的稳定性、加速收敛以及增强模型的泛化能力。常见的归一化层有以下三种：

- a. 批量归一化（Batch Normalization, BN）：BN 对每一层神经网络的输入按批次（Batch）进行归一化，即对每个批次内的每一个特征图（通道）分别计算其均值和方差，然后对数据进行归一化，最后可以通过两个可学习的参数  $\gamma$  和  $\beta$  进行缩放和平移，以恢复数据的表达能力。BN 将输入数据的分布调整为均值为 0，方差为 1 的标准正态分布。BN 具有一定的正则化效果，可以减少对 Dropout 层的依赖。
- b. 层归一化（Layer Normalization, LN）：不同于 BN 对整个批次进行归一化，LN 是对单个样本的同一层的所有神经元的输出进行归一化，特别适用于循环神经网络之类用以处理序列数据的模型，在使用小批量数据进行训练时也可以有效应用。
- c. 组归一化（Group Normalization, GN）：GN 通常与分组卷积<sup>[64]</sup>一起使用，将特征图（通道）分为若干组，对每一组内的神经元输出进行归一化。GN 兼顾了 BN 和 LN 的优点，在小批量数据训练等场景具有良好的性能。

#### (5) Dropout 层

Dropout 是深度神经网络中常用的正则化技术，主要用于缓解神经网络在训练过程中可能产生的过拟合现象。Dropout 的思想在于，通过在训练过程中随机地将一些神经元的输出“丢弃”，使每个神经元能够独立地学习到更多的特征，避免神经元之间的相互适应。具体实现中，Dropout 为每个神经元分配一个保留概率  $p$ ，表示神经元的输出被保留下参与后续计算的概率。在每个训练批次中，Dropout 随机选择一些神经元，将它们的输出以  $p$  的概率丢弃（即置为 0），并将其余神经元的输出值除以  $1 - p$  进行缩放，从而保持整体的期望输出值不变。而在测试阶段，Dropout 不再进行随机丢弃，而是保持所有神经元的输出，并将其乘以  $p$ ，用以保持与训练时相同的期望输出。

### 2.2.2 循环神经网络

循环神经网络（Recurrent Neural Network, RNN）是一种用于处理序列数据的深度神经网络结构，其内部的循环连接使得网络在处理每个时间步的输入时都能够保留之前的信息状态，即具有“记忆”能力。这种特性使得 RNN 特别适合用来处理具有先后顺序的数据，如时间序列数据、自然语言文本等。

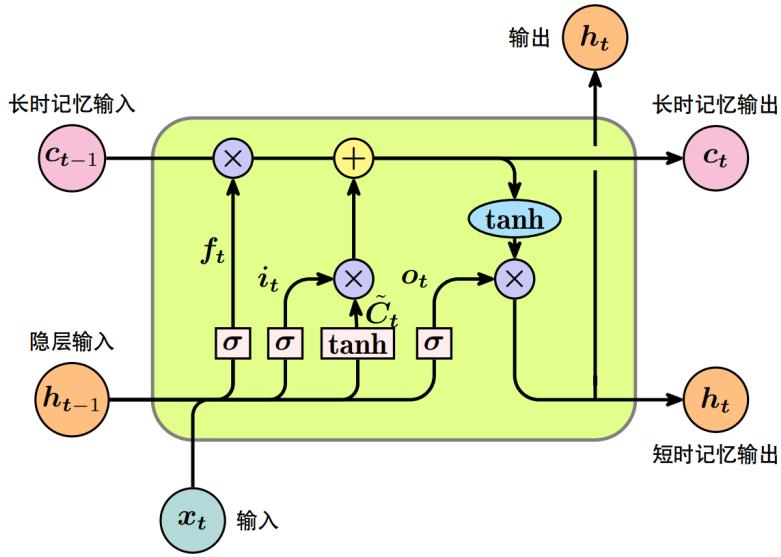


图 2.10 LSTM 元胞结构

原始的 RNN 模型中的循环连接使得梯度在进行传播时容易呈现指数级衰减或增长的问题，即容易出现梯度消失或梯度爆炸现象，从而导致模型难以训练，难以捕捉到序列数据中的长期依赖关系。长短期记忆网络（Long Short-Term Memory, LSTM）针对这些问题进行了改进，成为现阶段被广泛使用的基于 RNN 架构的模型之一。LSTM 引入了门控机制，通过一系列门控单元控制信息的存储、读取和更新，从而实现对长期和短期信息的有效管理。LSTM 的主要组成部分包括单元状态（Cell State）、遗忘门（Forget Gate）、输入门（Input Gate）和输出门（Output Gate），LSTM 一个元胞的结构如图 2.10 所示，其中， $h_{t-1}$  表示前一时间步的隐藏状态， $c_{t-1}$  表示前一步的单元状态， $x_t$  表示当前时间步的输出， $f_t$  表示忘记门， $i_t$  表示输入门， $o_t$  表示输出门， $\sigma$  为 Sigmoid 激活函数，代表门控的程度， $\tanh$  为双曲正切激活函数，用于更新细胞状态。多个元胞前后相连，就组成了 LSTM。

### (1) 单元状态

单元状态用于存储和传递序列数据中的长期依赖信息，其通过遗忘门、输入门和输出门的控制，来决定如何更新和传递记忆信息。更新单元状态的公式如公式 2.7 所示。

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{C}_t \quad (2.7)$$

其中， $\odot$  表示逐元素相乘的 Hadamard 乘法， $\tilde{C}_t$  表示候选单元状态，其计算方式如公式 2.8 所示，其中， $W_c$  是用于单元状态更新的权重参数， $b_c$  是偏置参数。

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2.8)$$

### (2) 遗忘门

遗忘门用于控制哪些信息应当从单元状态中“遗忘”。遗忘门的输出值在 0 和 1 之间，代表每个单元状态中的信息应该被保留或遗忘的程度，其计算过程如公式 2.9 所示，

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.9)$$

其中， $W_f$  是遗忘门的权重参数， $b_f$  是偏置量。

### (3) 输入门

输入门用于控制哪些新信息应当被“记住”。输入门决定了候选单元状态被记忆的程度，即有多少信息被加入单元状态中。输入门的计算如公式 2.10 所示，

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.10)$$

其中， $W_i$  是输入门的权重参数， $b_i$  是偏置量。

### (4) 输出门

输出门用于控制从单元状态传递到当前时间步的隐藏状态（Hidden State）的信息量，输出门的计算如公式所示，

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2.11)$$

其中， $W_o$  是输出门的权重参数， $b_o$  是偏置量。

### (5) 隐藏状态

隐藏状态（Hidden State）用于捕捉序列数据中的信息和特征，并传递给下一个时间步或输出层。隐藏状态通过循环连接获取上一个时间步的隐藏状态，能够存储序列数据中的历史信息，其计算过程如公式 2.12 所示。

$$h_t = o_t \cdot \tanh(c_t) \quad (2.12)$$

## 2.2.3 注意力机制

注意力机制的提出受到人类认知科学的启发，其核心理念在于模拟人类大脑在处理信息过程中的选择性关注机制，即并非均匀地分配注意力处理对待所有输入，而是将注意力主动、动态且有选择性地聚焦于最重要或最相关的信息上。自二十世纪被提出以来<sup>[65]</sup>，注意力机制在计算机视觉、自然语言处理等领域得到了广泛的应用，并表现出了优秀的效果。

注意力机制的基本思想为计算每个位置的权重，代表该位置对输出的重要性，权重通常可以由可学习的函数或神经网络生成，是一个与输入相同大小的矩阵，使用权重对输入数据进行加权操作，从而为重要的数据分配更多的关注。注意力机制通常包括查询矩阵（Query）、键矩阵（Key）和值矩阵（Value），其计算如公式 2.13 所示，

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}V\right) \quad (2.13)$$

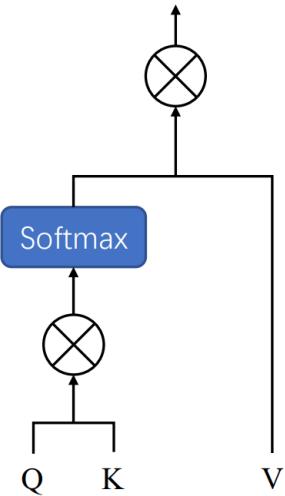


图 2.11 注意力机制

其中， $Q$  为查询矩阵， $K$  为键矩阵， $V$  为值矩阵， $d_k$  为键的维度。输入数据首先通过查询、键和值三个矩阵进行变换，然后计算出每个位置的注意力得分（Attention Score），这个得分衡量了数据中每个位置与当前处理位置之间的关联强度，所有位置的得分构成注意力矩阵。接着，使用 Softmax 函数获取注意力矩阵的归一化表示，最后，使用注意力矩阵对输入数据进行加权求和，从而得到最终的结果。注意力机制的结构如图 2.11 所示。

注意力机制有多种变体，其中最常见的是点积注意力（Dot-Product Attention）和加性注意力（Additive Attention）。点积注意力使用输入序列和当前输出的点积来计算注意力权重，而加性注意力则使用一个全连接神经网络来学习权重。此外，还有自注意力（Self Attention）<sup>[42]</sup> 机制，其计算每个位置上的元素与序列中所有其他元素的相关性来动态获取上下文信息，以捕获序列内部的依赖关系。

### 2.3 本章小结

本章主要介绍了与论文研究工作相关的理论和方法。首先，介绍了运动想象脑电图分类相关的生理基础知识，包括运动想象与人脑神经系统的关系、脑电图信号及其特性，为论文后续的研究进行铺垫。其次，论文介绍了深度神经网络的一些理论和模型，包括卷积神经网络、循环神经网络和注意力机制，也是论文后续研究中会用到的方法。通过本章内容，能够了解运动想象脑电图分类和深度神经网络的基础知识和常用方法，为后续的研究和应用奠定基础。

### 3 基于特征融合和注意力机制的运动想象脑电图分类网 络构建

论文前两个章节主要对运动想象脑电图 (Motor Imagery Electroencephalography, MI-EEG) 分类领域的基础知识以及相关研究做了一定的介绍，并且对该领域仍然存在的问题进行了分析。本章对这些问题进行进一步的探讨，针对这些问题提出了一种端到端的新模型 HA-FuseNet：首先，在 Inception 模型的基础上引入密集连接，在进行多尺度特征提取的同时，对 EEG 信号的浅层特征和高级语义信息加以利用，以更完整地获取 EEG 信号的特征；其次，在以上结构的基础上，针对 EEG 信号的特性构建了 svSE 混合注意力模块，将方差信息和时空轴向分离引入注意力机制中，提高模型对 EEG 信号中的重要特征的关注度；随后，针对 EEG 信号的时序特性，使用 LSTM 获取长短期时序依赖，同时，构建了全局自注意力模块 SCoT 以更好地利用 EEG 信号的全局时空信息；最后，为了取得更好的实时性能，对模型进行了轻量化设计。经过以上构建，HA-FuseNet 能够同时利用 EEG 信号的局部特征细节和全局依赖信息，实现更为丰富的特征融合效果，获得更好的 MI-EEG 分类效果。

#### 3.1 基于特征融合和注意力机制的端到端 MI-EEG 分类网络 HA-FuseNet

原始的 EEG 信号通常为二维数据，包括通道（电极）和时间两个维度，具体而言，在 EEG 信号矩阵中，行代表分布在头皮不同位置的采样通道，列为时间序列数据，每个采样点对应一个时间戳下的生物电信号（通常为电压值），因此，一列数据就是一个特定时间点下所有通道同步采集到的电压读数。原始的 EEG 信号经过预处理之后，可以转换为时频图、头皮点位拓扑图等输入模式，尽管经过转换的输入相较于原始输入能够更全面地体现 EEG 信号的时频空信息，但这一过程往往需要具有神经科学背景的人工参与，在增加了人工成本同时，限制了模型自适应学习 EEG 信号中蕴含的复杂时空特征的能力，此外，复杂的预处理环节也增加了计算开销和应用成本，难以满足 BCI 系统即时响应的需求。因此，端到端网络在 MI-EEG 分类领域受到越来越多的重视，这类网络不经过或者仅仅经过很少的预处理步骤，而由深度学习算法自适应地提取关键特征并作出预测。

为此，论文构建了基于特征融合和注意力机制的端到端 MI-EEG 分类网络 HA-FuseNet，其结构如图所示。HA-FuseNet 有两条并行分支，分别是基于卷积神经网络搭建的分支与基于长短记忆网络的分支。卷积神经网络分支基于 Inception 结构搭建，通过在 Inception 分支引入密集连接结构，进行多尺度特征提取，同时利用高级语义信息与浅层特征；通过引入反转瓶颈层，在深度维度上促进时空特征的融合；通过引入

svSE 混合注意力模块促进网络对重要特征的关注度，svSE 模块采取时空特征分离的策略，同时利用了 EEG 信号的方差信息，以获得针对性的结果。长短期记忆网络分支用以获取时间域的长期依赖信息，并且加入了 SCoT 全局注意力模块获取时空域的全局上下文信息。双分支的特征通过 C2R 模块和 R2C 模块进行交互，并随后在深度维度进行融合，以在利用卷积神经网络提取到的局部信息的同时有效地建立全局依赖关系，从而获取更为精准的分类效果。

通过 HA-FuseNet，可以更完整地获取 MI-EEG 信号的有效特征，提升分类效果。在下文中，将依次阐述论文的改进思路与构建方法。

## 3.2 基于多尺度密集连接和混合注意力的分类网络 DIS-Net

卷积神经网络具有强大的特征提取能力，在计算机视觉等领域应用广泛。EEG 信号在时空域上具有局部相关性，即相邻的电极点和相近的时间窗内的信号往往携带相似的信息（但数据中相邻的行未必在空间域中相邻），因此，可以将 EEG 信号视为二维图像，使用卷积神经网络提取其局部特征。

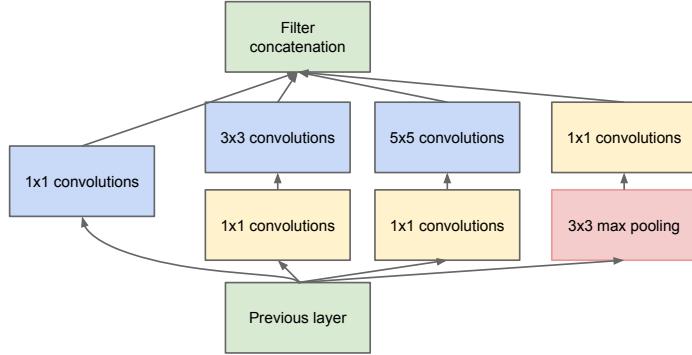
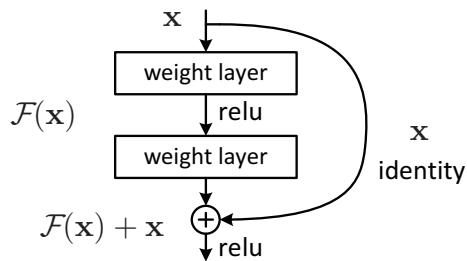
### 3.2.1 基于 Inception 的基础网络

ShallowConvNet<sup>[36]</sup> 是一个专为端到端解码脑电图 (EEG) 信号而设计的深度学习架构，其构思源自 EEG 信号解码研究领域中广泛使用的经典特征提取方法——滤波器组共空间模式 (Filter Bank Common Spatial Pattern, FBCSP)<sup>[16]</sup>。ShallowConvNet 具有 FBCSP 算法对频带功率特征高效提取的特性，在实验中证明了能够学习频带功率变化的时间结构特性<sup>[36]</sup>，研究发现，该特性有助于提高分类性能<sup>[66]</sup>。实验证明，ShallowConvNet 在 MI-EEG 分类领域具有优良的性能<sup>[37]</sup>，其采用四步流程对原始二维输入数据进行处理。具体而言，ShallowConvNet 首先通过时间卷积层捕获信号的时间域特征，再通过空间卷积层捕获这些时间特征在不同通道间的空间关联性，随后通过平均池化层进行下采样，最后通过全连接层将多维特征映射至分类输出空间。ShallowConvNet 提出的时间卷积与空间卷积相分离的策略是由于 EEG 原始输入的时域与空域之间的相关性较低，后续的研究大多沿袭了这种方法，论文同样遵循这一思路对 EEG 信号进行处理。

将 EEG 原始输入视为具有空间信息的图像数据，论文参考以下几种计算机视觉领域的经典模型，用于网络的特征提取基础结构：

#### (1) Inception 网络

Inception 模块起源于经典的 GoogLeNet 模型<sup>[67]</sup>，并在计算机视觉图像分类任务中取得了优异的效果。传统卷积神经网络倾向于通过加深和拓宽网络结构以增进性能，然而这种做法伴随着参数数量的激增，不仅加大了计算负担，还可能导致过拟合问题。在这种背景下，Inception 模块提出了多尺度特征并行抽取的策略，旨在保持网络稀疏性的同时，充分利用密集矩阵运算的高性能。典型的 Inception-V1 模块的结构如图 3.1 所示，

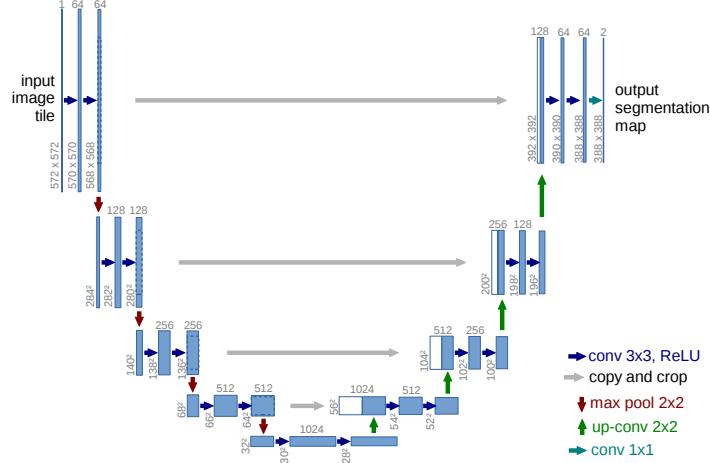
图 3.1 Inception 结构<sup>[67]</sup>图 3.2 残差块结构<sup>[44]</sup>

其将不同大小的卷积层和最大池化层并行排列，并行地对输入数据执行多种卷积和池化运算，继而将提取到的不同尺度特征在深度维度上进行拼接。这种设计能够在单层网络内并行地提取输入数据在不同层次和粒度的特征信息，从而在高效扩展网络的深度和宽度的同时，有效削减参数规模，提升计算速度。此外，Inception 模块中引入了  $1 \times 1$  卷积核，用以实现深度上的特征转化和降维，这种方式能够让模型学习到更为丰富的特征，同时降低计算成本。后续的论文中，Inception 模块不断迭代优化，陆续引入了批归一化、深度可分离卷积、矩阵因子分解等技术，进一步提升了模型的性能<sup>[68][69]</sup>。

### (2) 残差神经网络

残差神经网络（Residual Network, ResNet）<sup>[44]</sup> 是计算机视觉图像识别领域的一个经典模型。ResNet 研究发现了深度神经网络的退化现象（Degradation），即随着网络深度不断增加，模型准确率起初随深度上升，却在达到峰值后急剧下滑。针对这种现象，ResNet 提出了残差学习框架，其核心思想是引入残差块（Residual Block），每个残差块通过快捷连接（Shortcut Connection）将输入信息直接输送至输出层，使得网络只需要专注学习输入与输出之间的残差信息，而非完整的映射关系。基础的 ResNet 由一系列残差块堆叠而成，残差块的结构如图 3.2 所示。通过快捷连接，ResNet 在训练过程中，梯度能够从深层网络直接回传至浅层，避免网络深度增加带来的训练困难和性能下降问题，从而提升深度神经网络的性能表现和训练效率。

### (3) U-Net

图 3.3 U-Net 结构<sup>[70]</sup>

U-Net 模型<sup>[70]</sup>最初是为生物医学图像分割任务而设计，其具有优秀的性能，尤其在细胞、器官和病变区域的精确标注上表现出色，是医学图像分割领域的主流模型之一。U-Net 的独特之处在于其采用了对称的编码-解码结构（Encoder-Decoder）和跳跃连接（skip connection），其结构如图 3.3 所示。编码器通过连续的卷积和下采样层对输入图像进行深度特征提取和空间压缩，提炼出高级抽象特征；解码器部分则通过上采样和卷积恢复到与输入图像相同的空间分辨率，同时保留详细的定位信息。跳跃连接将编码器各阶段的特征图直接传递给相应的解码器阶段，有效地结合了包含更多细节信息的浅层特征和包含更多高级语义信息的深层特征，从而在图像分割任务中能够取得更为精细的分割效果。同时，U-Net 模型结构简单，易于训练，能够缓解小样本数据集上的过拟合问题。

在这三种模型中，Inception 和 ResNet 均在图像分类任务中展现出了优秀的性能。Inception 通过同一层网络内的多尺度特征并行抽取，在不显著增加网络深度的前提下，实现了特征提取的广度与效率的提升。ResNet 通过引入快捷连接，解决了深度神经网络训练过程中的梯度消失和退化问题，增强了深层次网络的训练效率和性能表现。U-Net 则在生物医学图像分割领域取得了优秀的表现，医学图像的语义信息较为简单，且结构较为固定，因此高级语义信息和低级特征都相对重要，U-Net 通过跳跃连接保留并融合了这两类信息，同时，U-Net 参数量较小，不容易在小样本数据集上发生过拟合现象。论文选择将 U-Net 迁移至 MI-EEG 分类任务中，是因为 EEG 信号具有与生物医学图像类似的生理特性，如特征相对简单、数据集规模偏小等。

为了验证 Inception、ResNet 与 U-Net 在 EEG 信号分类任务中的性能，论文在 BCI Competition IV Dataset 2A 数据集上进行实验对比。在实验设置中，统一将三种模型的网络深度调整为三层，并对其他关键参数如卷积核大小、学习率等进行了固定，此外，对这三种模型的原始代码进行了调整，使得其适应 MI-EEG 分类任务。实验结果如表 3.1 所示，主要展示准确率（Accuracy, ACC）和 Kappa 一致性系数（Kappa）指标，这两项

指标是数据集中九位被试的平均表现。

**表 3.1 Inception、ResNet、U-Net 实验结果对比**

Models	ACC(%)	Kappa
Inception	<b>67.40</b>	<b>0.56</b>
ResNet	56.94	0.43
U-Net	62.27	0.50

实验数据显示，Inception 模型在这三种模型中具有最优的性能表现，U-Net 次之，ResNet 的表现则相对较差。这可能是因为同样的网络深度下，Inception 模型得益于多尺度并行特征提取机制，能更全面地捕获 EEG 信号的多种特征。相比之下，U-Net 虽然通过跳跃连接有效地结合了 EEG 信号的低层特征和高层语义信息，但在解码器阶段，U-Net 将特征图重建至原始空间尺寸的过程可能为分类任务引入了不必要的复杂性。ResNet 的快捷连接在较浅层网络结构中可能未能完全发挥其优势，更适用于深层次网络。实验结果与过往研究中关于浅层网络更适合 MI-EEG 分类任务的研究结论相互印证。综上所述，论文选用 Inception 模块作为 MI-EEG 信号特征提取的基础结构，旨在保持模型简洁高效的同时，在 MI-EEG 分类任务中取得更好的性能。

EEG 信号的空间分辨率较为不稳定，例如，在 BCI Competition IV Dataset 2B<sup>[71]</sup> 数据集中，仅仅使用了三个电极采集 MI-EEG 信号，使得空间信息相对时间信息更为稀疏。为了减少对高空间分辨率的依赖，论文采取更关注时间特征的策略，即将 Inception 模块应用于时间卷积层中，使得时间卷积层的复杂度高于空间卷积层的复杂度，从而保持相对均衡的特征提取。

文献<sup>[36, 37]</sup>指出，在 EEG 信号解码任务中，增加神经网络的深度有利于提升解码精度。瓶颈层（Bottleneck Layer）是深度神经网络中的常见结构<sup>[44, 72]</sup>，通常用于对数据的降维和升维，由于采用了  $1 \times 1$  卷积进行操作，瓶颈层能够有效地减少神经网络的参数。不同于原始 Inception 模块中通过瓶颈层进行数据降维的操作，论文使用瓶颈层对数据进行升维操作，并将瓶颈层提取至卷积和池化操作之前，其目标为在深度维度上促进时空信息的融合。此外，论文在模型中引入了批量归一化层和 Dropout 层，用以加快网络训练速度，并避免小数据集下过早的过拟合。

论文将改进后得到的基础模型称为 BaseNet，其结构如图 3.4 所示。需要注意的是，Inception 模块的层次数量和分支数量是影响其性能表现的两项可调的超参数。

### 3.2.2 多尺度密集连接

在构建 BaseNet 时，论文发现 U-Net 在处理 MI-EEG 分类任务时同样展现了一定的优势。由于 EEG 信号的特征相对简单，因此低级特征与高级语义信息都相对重要，U-Net 因其特殊的跳跃连接结构有效地融合了这两种信息，然而，U-Net 中通过解码器阶段将

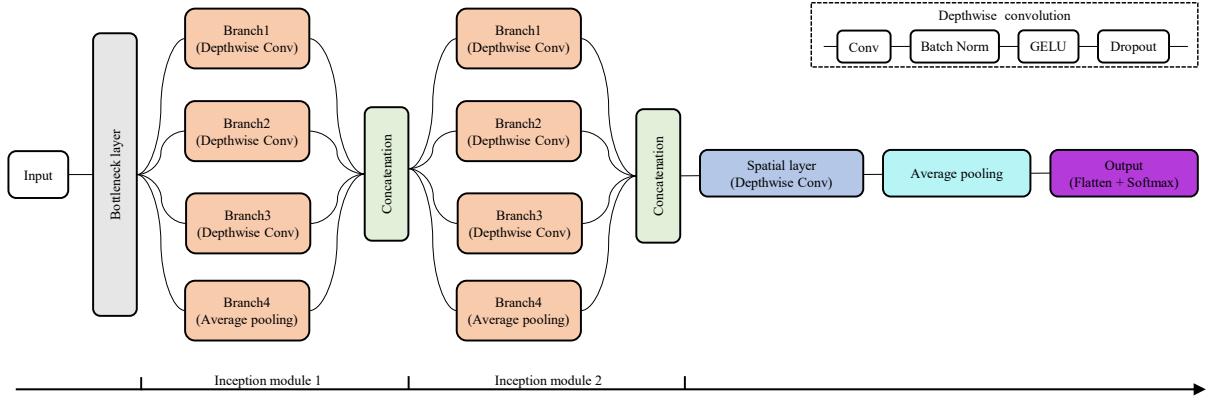
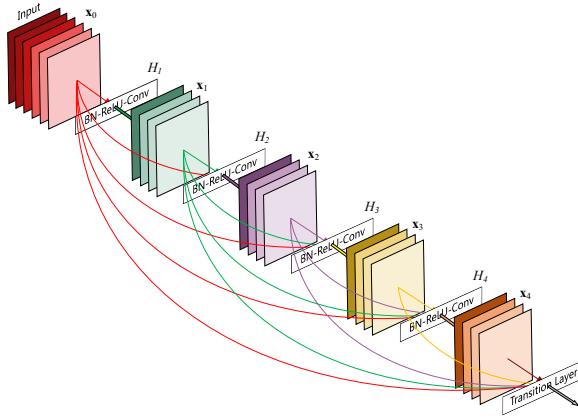


图 3.4 BaseNet 结构

图 3.5 密集连接结构<sup>[72]</sup>

特征图恢复至原始空间尺寸的操作并非必要，因为在分类任务中，这种重建过程可能导致额外的计算负担且对分类性能的提升效果不明显。因此，论文从 U-Net 兼顾低级特征与高级语义信息的策略中得到启发，同时对不必要的特征图空间尺寸还原过程进行规避，以构建一个既能充分利用 EEG 信号中各级别特征信息，又具备高效计算能力的改进模型。

Gao Huang 等人于 2016 年提出了密集连接网络<sup>[72]</sup>（Dense Convolutional Network, DenseNet）。在 ResNet 的基础上，DenseNet 提出了一种更为激进的连接模式：引入从任意层到后续层的直接连接，即密集连接（Dense Connection）。DenseNet 的第  $l$  层接收所有前序特征图为输入，其输出为  $x_l$ ：

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (3.1)$$

其中， $[x_0, x_1, \dots, x_{l-1}]$  代表第  $0, \dots, l-1$  层的输出特征图， $H_l(\cdot)$  代表非线性转换复合函数，可能包括一系列的批量归一化、ReLU、池化及卷积操作。密集连接的结构如图 3.5 所示。

在密集连接模块中，所有的前序特征图通过 Concentrate 在深度维度上连接在一起，

因此，在一个密集连接模块中所有特征图的大小是相同的。DenseNet 提出了 Transition 模块用于连接两个相邻的密集连接模块，并通过池化操作对特征图进行下采样，从而减小特征图的大小。

通过密集连接的方式，网络中的每一层都能够访问并整合所有前序层提取出的特征信息，进而充分利用 EEG 信号中的低层特征细节和高层语义信息。相较于 U-Net 中的编码-解码结构，密集连接无需经历数据空间的重建过程，即可实现特征的有效复用。此外，DenseNet 中的 Transition 模块也为实现特征图的下采样提供了思路。

为了同时利用 Inception 模块多尺度并行特征提取和密集连接兼顾深层与浅层特征的优点，论文选择将 DenseNet 嵌入 Inception 模块中，替代原本的时间卷积核。为了简洁起见，下文称之为密集连接模块。

原始的密集连接模块在计算机视觉图像分类任务中都取得了优秀的表现，在设计上，它考虑到了图像数据在空间维度上的局部相关性，然而，EEG 信号的时空域局部相关性较低，需要对始密集连接模块的卷积核进行改造。研究指出<sup>[37]</sup>，当设计用于提取 EEG 特征的卷积核时，将时间卷积核长度设置为 EEG 信号采样率的一半，可以有效地捕获 2Hz 及以上频段的信号信息。因此，在处理 EEG 信号时，时间卷积核的长度应当依据 EEG 信号的实际采样率来灵活设定，以便提取不同频率成分的信号特征。

将密集连接模块中的卷积核概念转化为针对时间序列的一维卷积（尽管在实现时仍采用二维卷积），为了更好地匹配 EEG 信号的采样特性及其内在频率成分，依据 EEG 信号的采样频率  $sfreq$  来动态调整位于 Inception 第  $i$  个分支上的密集连接模块的时间卷积核大小  $kernel_i$ ，具体公式如下：

$$kernel_i = \left\lfloor \frac{s freq}{2 \times i} \right\rfloor, \quad i \leq 5 \quad (3.2)$$

其中， $s freq$  是 EEG 信号的采样频率。这样设置卷积核大小是为了更全面地捕捉 EEG 信号中不同频率成分的特征信息，同时避免因卷积核大小过于接近而导致提取的特征之间重叠度过高。例如，当  $i = 1$  时，卷积核大小将是采样率的一半，进而能够有效地捕获到 2Hz 及更高频段的 EEG 信号特征。

由此，密集连接模块的非线性转换复合函数  $H(\cdot)$  定义如公式 3.3 所示，

$$H(\cdot) = BN + GELU + 1 \times kernel_i conv + Dropout \quad (3.3)$$

经过以上改进，新构建模型的结构如图 3.6 所示，将其命名为 DI-Net。Dense-Inception module 构成了时间卷积层，其内部的各个分支均由一系列改进版的 Bottleneck Dense Block 紧密堆叠而成，形成密集连接结构，用以同时获取浅层和深层的时间特征。每个分支提取的特征图在深度维度上进行聚合，并通过 Transition 模块进行深度压缩和特征整合，以进一步提升模型的表达能力和计算效率。

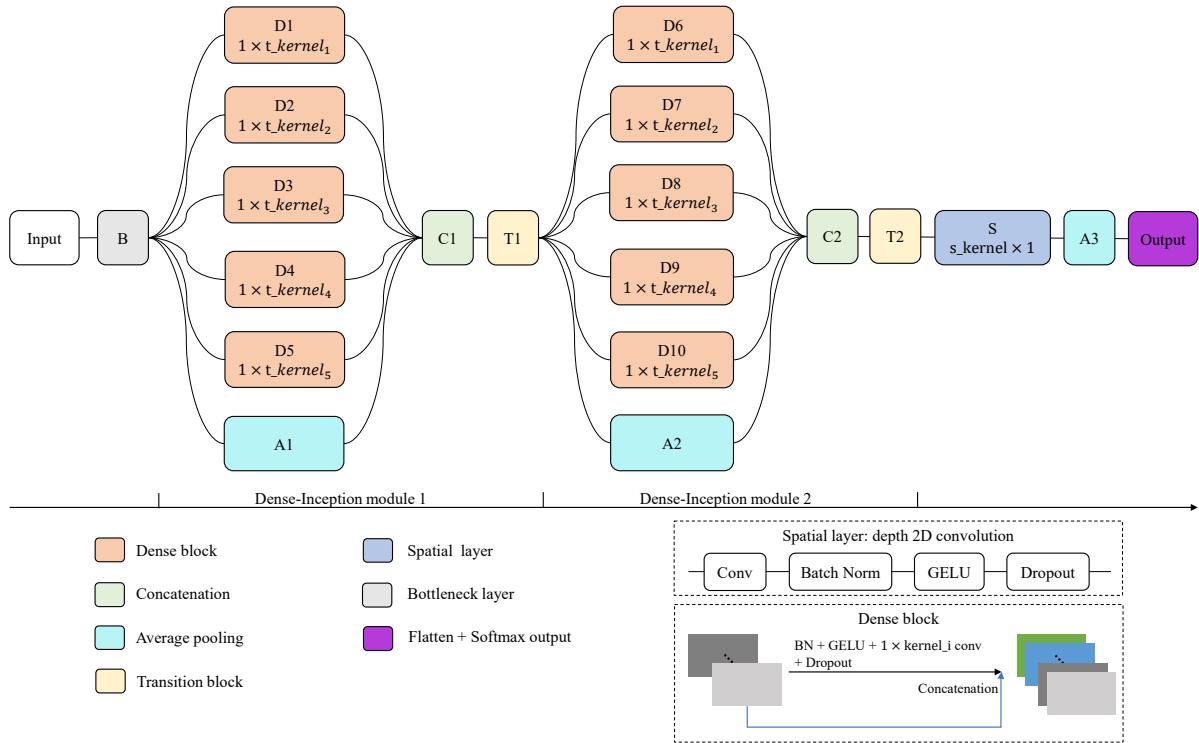


图 3.6 DI-Net 结构

### 3.2.3 混合注意力 svSE

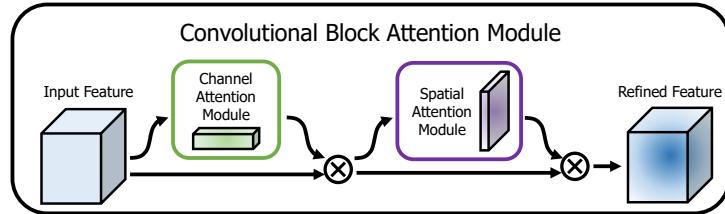
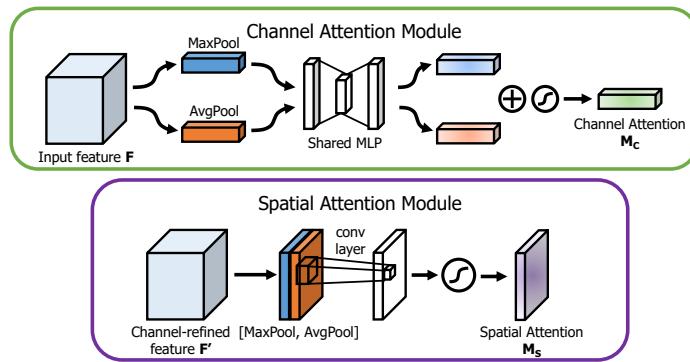
根据神经科学先验知识, EEG 信号中不同的通道和采样点具有不同的重要性, 这为在 MI-EEG 分类领域应用注意力机制提供了理论依据, 此外, 将二维 EEG 信号视为一种由通道和时间两个维度构成的特殊图像, 使得在 MI-EEG 分类领域能够迁移应用计算机视觉领域中的注意力机制。

计算机视觉领域中经常使用的注意力机制有:

#### (1) 通道注意力机制

不同于 EEG 信号中代表电极的通道, 计算机视觉领域的通道代表图像的不同特征映射。通道注意力机制用于调整不同特征通道的重要性, 通常会对每一个特征通道计算一些全局统计量, 如均值、方差等, 再将这些统计量经过非线性变换层进行编码, 最后将编码向量进行转换并用于各个特征通道的加权。通道注意力机制的经典模型是压缩和激励网络 (Squeeze-and-Excitation Networks, SENet)<sup>[73]</sup>, 其主要思想即是压缩 (Squeeze) 和激励 (Excitation), SENet 首先通过压缩操作获取全局上下文信息, 然后通过激励操作对每个通道独立生成权重系数。具体而言, 在压缩操作中, SENet 在空间维度执行全局池化操作, 将每个通道的特征图汇总成一个标量值; 然后, 在激励操作中, SENet 通过一个全连接网络生成每个通道的权重系数, 这些权重系数用于重新加权每个通道的特征图, 以增强有用的特征并抑制无用的特征。

在后文中, 为避免与计算机视觉领域中的概念相混淆, 在 MI-EEG 分类任务中, 用

图 3.7 CBAM 结构<sup>[75]</sup>图 3.8 CBAM 模块结构<sup>[75]</sup>

深度来代表 EEG 信号的不同特征映射，而通道仍然代表电极。

### (2) 空间注意力机制

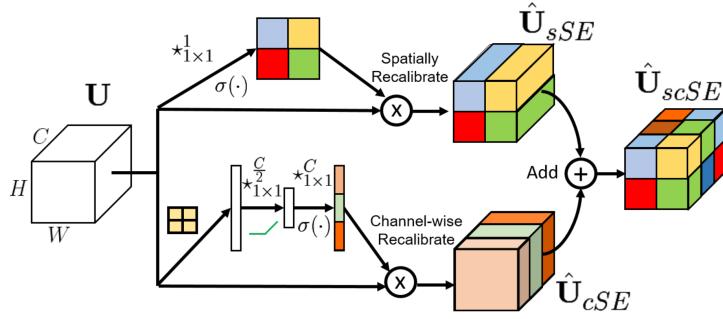
在计算机视觉领域中，空间注意力机制用于调整图片、视频等输入数据在空间维度中不同区域的重要性，通常会在深度维度上通过全局池化、卷积、特征融合等操作生成一个与特征图尺寸相同的注意力图，其值反映了空间维度中不同区域的注意力强度，最后，将注意力图进行转换，并用于原始特征图的加权。空间注意力机制的经典模型是空间变换网络（Spatial Transformer Network, STN）<sup>[74]</sup>，其具有对输入数据进行空间变换的能力，能够自动捕获重要区域的特征。

### (3) 混合注意力机制

混合注意力机制是一种集成多种注意力机制（如空间注意力、通道注意力及自注意力等）的方法，旨在更全面地捕获和整合输入数据在不同维度的有效信息。混合注意力机制通常会使用不同的注意力机制分别计算原始特征图的注意力权重，再将这些注意力权重进行融合，最后将融合后的注意力权重用于原始特征图的加权，或者将不同的注意力权重用于原始特征图加权，再将加权特征图进行融合。混合注意力机制的经典模型有卷积注意力机制模块（Convolutional Block Attention Module, CBAM）<sup>[75]</sup>、空间与通道压缩与激励模块（Spatial and Channel Squeeze-and-Excitation, scSE）<sup>[76]</sup> 等。

CBAM 结合了通道注意力机制与空间注意力机制，其结构如图 3.7 所示，输入特征图首先经过通道注意力模块进行加权，再通过空间注意力模块进行加权，从而得到最终结果。

具体而言，在通道注意力模块中，输入特征图分别进行空间维度上的全局最大池化

图 3.9 scSE 结构<sup>[76]</sup>

和全局平均池化，再将得到的统计值分别通过一个共享权重的全连接层，最后经过逐点加和与非线性变换得到通道注意力权重，用于输入特征图的加权。空间注意力模块的输入是经过通道注意力加权的特征图，首先在通道维度上进行全局最大池化和平均池化，再将得到的统计值在通道维度进行拼接，最后经过卷积降维与非线性变换得到空间注意力权重，与特征图加权后得到最终结果。CBAM 的模块结构如图 3.8 所示。

scSE 同样结合了通道注意力机制与空间注意力机制，基于 SENet 提出了一种通道注意力模块（Channel Squeeze-and-Excitation, cSE）和一种空间注意力模块（Spatial Squeeze-and-Excitation, sSE），其结构如图 3.9 所示，不同于 CBAM，scSE 的两个子模块并行处理原始输入，分别在空间维度和通道维度对原始输入进行加权，最后再进行特征图的融合。具体而言，cSE 模块中，原始输入依次经过了空间维度的全局平均池化，通道维度的卷积降维与升维，以及非线性变换，以得到通道注意力权重。sSE 模块中，直接通过深度卷积在通道维度进行降维，再经过非线性变换以得到空间注意力权重。

注意力机制通过动态分配权重，使得模型能够聚焦于输入数据中的关键信息，削弱噪声的影响，混合注意力机制则结合了多种注意力机制的优点，从而能够更全面地捕获和整合不同维度的数据特征，并在许多情况下展现出优于单一注意力机制的性能。因此，论文将升维处理后的 EEG 信号视作具有深度信息的图像数据，采用结合了深度注意力和空间注意力的混合注意力机制对 BaseNet 进行改进。

CBAM 模块和 scSE 模块均为轻量级注意力模块，且均兼顾深度注意力和空间注意力，但 scSE 模块在参数数量上更具优势。与此同时，文献<sup>[76]</sup>研究发现 scSE 模块在语义分割任务上表现出色，特别是在与 EEG 信号拥有相似生理特性的医学图像的分割任务，其性能优于 CBAM 模块。基于以上理由，论文选择基于 scSE 模块进行改进，提出了一种新的注意力机制 svSE（Separate Variance-Informed Spatial and Channel Squeeze-and-Excitation）模块，其结构如图 3.10 所示。

针对 cSE 模块，采用全局最大池化取代全局平均池化操作，用以突出显著特征，得到权重图  $Att_c \in \mathbb{R}^{D \times 1 \times 1}$ 。针对 sSE 模块，论文提出两种方式进行改进，并将两种方式

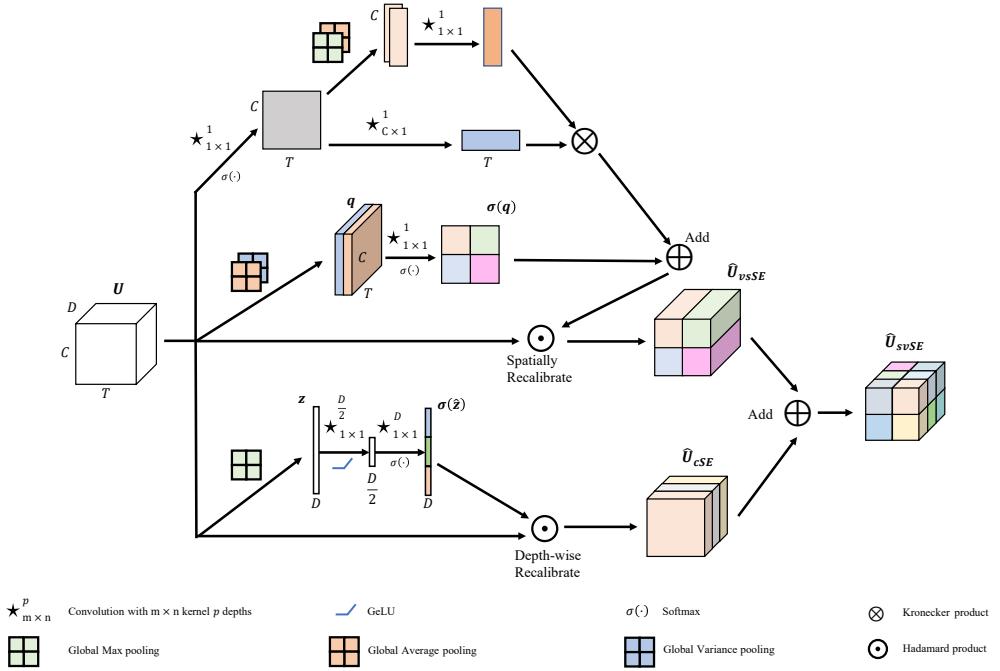


图 3.10 svSE 结构

所得的权重相结合以获取最终的输出：

(1) 由 CBAM 模块的多维全局池化思想以及 FBCNet 模型的方差层设计<sup>[39]</sup> 得到启发，对于输入  $X \in \mathbb{R}^{D \times C \times T}$ ，采用深度维度上的全局平均池化和全局方差计算操作代替原模块中的压缩操作，得到  $X_{pool} \in \mathbb{R}^{2 \times C \times T}$ ，随后通过  $1 \times 1$  卷积对特征图在深度维度进行聚合，得到的权重图  $Att_v \in \mathbb{R}^{1 \times C \times T}$ ，更好地表征 EEG 信号特性；

(2) 考虑 EEG 信号中的时空权重低相关性，即空间特征权重代表电极重要程度，时间特征权重代表采样点重要程度，分两个维度提取特征，获取轴向注意力。对于输入  $X \in \mathbb{R}^{D \times C \times T}$ ，在空间维度，首先使用  $1 \times 1$  卷积进行深度压缩操作获得  $X_{sf} \in \mathbb{R}^{1 \times C \times T}$ ，随后通过时间维度上的平均池化和最大池化得到两个特征图，获得  $X_{spool} \in \mathbb{R}^{2 \times C \times 1}$ ，通过  $1 \times 1$  卷积对这两个特征图进行融合，得到  $X_s \in \mathbb{R}^{1 \times C \times 1}$ 。对于时间维度，进行空间维度上的卷积操作，以得到时序权重  $X_t \in \mathbb{R}^{1 \times 1 \times T}$ 。最后，将空间权重与时序权重以克罗内克积（Kronecker）的方式相乘，恢复维度，得到最终的权重图  $Att_s \in \mathbb{R}^{1 \times C \times T}$ 。

其中， $D$  为输入的深度， $C$  为通道， $T$  为时间。此外，使用 Softmax 激活函数替换 Sigmoid 激活函数，旨在更好地利用全局信息。由此，整个 svSE 模块的公式如公式 3.4 所示，其中， $\oplus$  为逐元素相加， $\odot$  为逐元素相乘， $X_{svSE}$  为加权后的输出。

$$\begin{aligned}
 Att_{vs} &= Att_v \oplus Att_s \\
 X_1 &= \text{Expand}(Att_{vs}) \odot X \\
 X_2 &= \text{Expand}(Att_c) \odot X \\
 X_{svSE} &= X_1 \oplus X_2
 \end{aligned} \tag{3.4}$$

### 3.3 基于 LSTM 和全局自注意力的分类网络 LS-Net

卷积神经网络倾向于模拟人类视觉系统，卷积层具有局部感受野，能够出色地捕获信号中的局部时空特征，但对于那些跨越较长时间跨度和空间分布的复杂交互信息，其建模能力受限。EEG 信号作为随时间变化的信号，具有时间序列属性，长短期记忆网络能够捕获时间域的长期依赖，适合处理序列数据，但缺乏对空间信息进行提取的能力。为此，论文选择 LSTM 结合全局自注意力的方式，为模型加入时空全局信息。

#### 3.3.1 全局自注意力 SCoT

LSTM 能够一定程度上捕获序列数据的长期依赖关系，但 EEG 信号在具有长距离时间依赖的同时具有全局空间特性，使得 LSTM 无法充分挖掘并整合全局时空域的依赖信息。因此，论文基于 Non-local<sup>[77]</sup> 和 Contextual Transformer (CoT)<sup>[78]</sup> 这两种全局自注意力机制进行改进，针对 EEG 信号的特点提出了 SCoT (Separate Contextual Transformer) 模块，旨在更完整地捕捉 EEG 信号在时空域内的全局依赖关系，从而增强模型的全局建模能力。

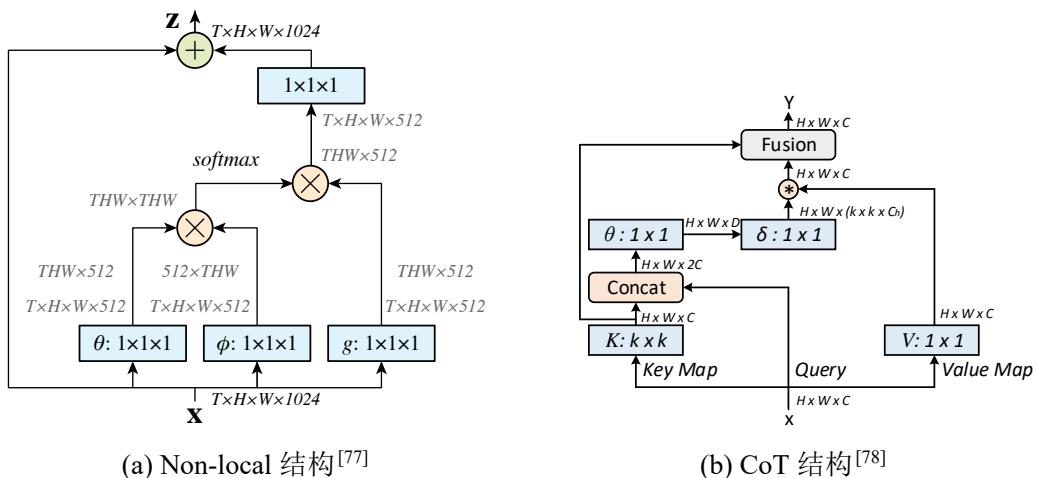


图 3.11 两种全局自注意力

#### (1) Non-local

Non-local 是将自注意力<sup>[42]</sup> 应用于计算机视觉领域的一项经典模型，能够捕捉特征图中任意两个位置之间的依赖关系，其结构如图 3.11a 所示。对于输入  $X \in \mathbb{R}^{C \times H \times W}$ ，首先通过三个  $1 \times 1$  卷积将通道数量由  $C$  压缩为  $\frac{C}{2}$ ，分别得到代表查询 (Query)、键 (Key) 和值 (Value) 矩阵的三个特征图  $X_\theta$ 、 $X_\phi$  和  $X_g$ 。随后对  $X_\theta$ 、 $X_\phi$ 、 $X_g$  进行展平 (Flatten)，并通过点积操作计算  $X_\theta$  与  $X_\phi$  的相似度矩阵  $S$ ，该矩阵代表输入特征图中各个位置之间的联系。接着，使用 Softmax 函数对相似度矩阵  $S$  进行归一化，使得其转化为概率分布，并与值矩阵  $X_g$  相乘，得到结果矩阵  $Y \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ ，最后，通过  $1 \times 1$  卷积恢复  $Y$  的维度至与  $X$  相同，以逐元素相加的方式得到加权后的结果。Non-local 通过

直接计算两个位置之间的远程依赖关系，能够获取全局自注意力，但在数据量较大的情况下，存在计算量偏大的问题。

### (2) CoT

CoT 考虑到 Non-local 忽略了相邻键上下文信息的问题，提出了一种将上下文信息挖掘能力集成入自注意力机制中的方式，其结构如图 3.11b 所示。对于输入  $X \in \mathbb{R}^{C \times H \times W}$ ，CoT 的改动主要在于使用  $3 \times 3$  卷积获取特征图的静态上下文信息，将其作为键矩阵  $K$ ，同时，通过  $1 \times 1$  卷积获取到值矩阵  $V$ ，直接获取了查询矩阵  $Q$ ，然后对  $K$  和  $Q$  进行深度维度的聚合，并通过两个连续的  $1 \times 1$  卷积获取注意力权重图  $A$ ，通过注意力权重图  $A$  和值矩阵  $V$  相乘获取特征图的动态上下文信息，最后将静态上下文信息和动态上下文信息进行了融合。

对于运动想象 MI-EEG 分类任务来说，问题在于 EEG 信号的特性：EEG 信号的时域和空域上具有较低的局部相关性；在空域上，尽管相邻的电极具有空间相关性，但在 EEG 原始输入中，相邻的电极未必排列为相邻的通道，因此，在端到端网络的情况下，空域内具有较低的局部相关性；在时域上，EEG 信号具有局部相关性，即相邻的采样点往往蕴含相似的信息。因此，Non-local 模块和 CoT 模块无法直接迁移至运动想象 MI-EEG 分类任务中：Non-local 模块能够计算任意两个位置之间的远程依赖关系，但在较大的数据量下，其计算代价高昂；而 CoT 模块则是基于图像数据的局部相关性，将局部上下文信息引入了自注意力机制中，并通过卷积操作减少了计算量。论文借鉴 Non-local 和 CoT 的思想，提出了 SCoT 模块。

综合考量 EEG 信号的时空特性与 Non-local、CoT 各自的特点，论文提出的 SCoT 注意力模块采取了时空域分步计算全局自注意力的策略，具体来说，针对空域中局部相关性较弱且数据规模有限的情况，对 Non-local 模块进行改良以适应空间自注意力计算；而对于具有强局部相关性且数据规模较大的时域信息，基于 CoT 进行改进以计算时间自注意力。SCoT 的结构如图 3.12 所示。

SCoT 分两个过程计算 EEG 信号的全局时空注意力，首先计算空域上的空间自注意力，使用空间自注意力对输入进行加权后，再基于加权数据计算时空域自注意力。空间自注意力能够为重要的电极（通道）分配更多的注意力，在局部相关性较弱的情况下突出不同通道的相对重要性。时空域自注意力能够充分利用时间序列数据中的局部相关性，以及已经由空间自注意力加权强化过空间特征信息，从而更加精细而全面地捕获 EEG 信号中的全局时空依赖关系。此外，两个过程的计算也有利于降低计算的复杂度。

SCoT 模块计算注意力的整体流程如公式 3.5 所示，其中， $X$  为 EEG 信号输入， $C$  为通道， $T$  为时间， $Att_s$  为空间自注意力模块， $Att_{st}$  为时空自注意力模块。

$$SCoT(X) = Att_{st}(Att_s(X)), X \in \mathbb{R}^{1 \times C \times T} \quad (3.5)$$

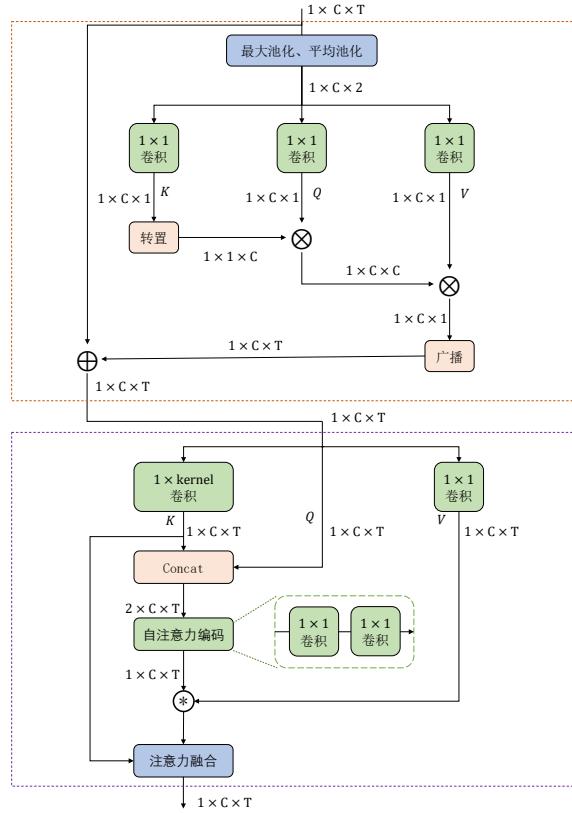


图 3.12 SCoT 结构

在空间自注意力模块中，首先进行混合池化操作，对输入进行转换：

$$X' = \text{Concat}[\text{MaxPool}(X), \text{AvgPool}(X)], X' \in \mathbb{R}^{2 \times C \times T} \quad (3.6)$$

EEG 信号具有低信噪比的特性，混杂了大量的噪声和干扰成分，通过结合最大池化操作和平均池化这两种操作，能够能够在最大程度保留 EEG 信号中关键特征的同时，对噪声起到一定的抑制作用，从而在不同的特征信息之间取得平衡。在取得池化表示  $X'$  后，通过公式获得键 (Key)、查询 (Query)、值 (Value) 矩阵  $K$ 、 $Q$  和  $V$ ，其中， $W_K$ 、 $W_Q$ 、 $W_V$  分别是对应的权重参数，通过  $1 \times 1$  卷积获得。

$$K = X'W_K, Q = X'W_Q, V = X'W_V \quad (3.7)$$

随后，将  $K$  经过转置后与  $Q$  相乘，提取其相似性特征，并使用 Softmax 进行归一化，得到权重矩阵  $A$ ，得到的计算过程如公式 3.8 所示，其中， $\otimes$  代表克罗内克积：

$$A = \text{Softmax}(Q \otimes K^T), A \in \mathbb{R}^{1 \times C \times C} \quad (3.8)$$

$A$  随后与值矩阵  $V$  相乘，实现对特征图的加权，并通过广播方式扩展到与输入特征图相同的尺寸，与输入特征图  $X$  加和后得到空间自注意力的输出，计算过程如公式 3.9 所示。

$$Att_s(X) = X + \text{Expand}(A \otimes V) \quad (3.9)$$

空间自注意力的输出被作为时空域自注意力模块的输入，以在时空域中利用经过强化的空间特征。首先通过  $1 \times kernel$  卷积提取时间域上的局部上下文特征，使得生成的键矩阵  $K$  具有输入特征图的静态上下文信息，对查询矩阵  $Q$  不做变换，使用  $1 \times 1$  矩阵获取值矩阵  $V$ 。对于经过空间自注意力加权后的输出  $Y \in \mathbb{R}^{1 \times C \times T}$ ，三个矩阵的计算过程如公式 3.10 所示， $W_K, W_V$  分别是键矩阵和值矩阵对应的权重参数，在计算上，使用卷积进行获取，其过程如公式 3.11 所示。

$$K = YW_K, Q = Y, V = YW_V \quad (3.10)$$

$$W_K = Conv_{1 \times kernel}(Y), W_V = Conv_{1 \times 1}(Y), kernel = \left\lfloor \frac{sfreq}{4} \right\rfloor \quad (3.11)$$

其中， $kernel$  的值取决于采样率  $sfreq$ ，在此设定为  $\left\lfloor \frac{sfreq}{4} \right\rfloor$ ，是为了捕获 4Hz 以上的频率信号。

随后， $K$  和  $Q$  在深度维度进行聚合，提取其相似度特征，并于  $V$  矩阵进行融合，获取特征图的动态上下文信息  $L$ ，其计算过程如公式 3.12 所示：

$$L = Fusion(V, (Conv_{1 \times 1}(Conv_{1 \times 1}(Concat[K, Q])))), L \in \mathbb{R}^{1 \times C \times T} \quad (3.12)$$

其中， $\odot$  表示逐元素相乘的 Hadamard 乘法。通过两个  $1 \times 1$  卷积进行自注意力编码运算，学习动态多头注意力矩阵。 $Fusion$  为局部矩阵乘法运算，用于度量局部网格中每个 query 与相应 key 之间的成对关系，在计算上，公式表示了  $Fusion$  运算的集合，在时间域和空间域分别使用轴向卷积进行了局部矩阵乘法运算，随后进行了  $BN$  批量归一化和  $Softmax$  归一化。

$$Fusion = Conv_{channels \times 1} + Conv_{1 \times kernel} + BN + Softmax \quad (3.13)$$

最后，将输入特征  $Y$  与  $L$  进行融合，得到时空域自注意力模块的输出  $Att_{st}$ ，如公式 3.13 所示。其中， $\odot$  表示逐元素相乘的 Hadamard 乘法。融合方式参考了 Transformer 的残差连接思想，将  $Y$  与  $L$  经过 Concat 聚合之后由卷积进行融合，从而实现对特征的校准，最后，通过逐元素相乘得到校准后的加权输出。计算期间，对数据进行了多项变形（Reshape）操作。

$$Att_{st} = Softmax(Conv_{1 \times 1}(AvgPool(Concat[Y, L]))) \odot Y \quad (3.14)$$

对运动想象脑电图分类任务来说，传统的全局自注意力机制忽视了 EEG 信号时域与空域的低局部相关性，导致网络无法很好对重要数据进行关注。SCoT 通过两个过程进行全局时空注意力的提取，对 EEG 信号具有更好的适应性，同时降低了计算的复杂度。空间全局自注意力能够突出不同通道数据的重要性，为时空域的全局自注意力提供先验信息，时空域的全局自注意力利用时间域的局部上下文信息和轴向卷积，将输入特征图的静态上下文和动态上下文进行结合，能够增强输出特征图的表征能力。

### 3.3.2 基于 LSTM 和 SCoT 的分类网络

长短期记忆网络 (Long Short-Term Memory, LSTM) 属于循环神经网络, 能够利用当前时间步及其之前的信息, 并且解决了 RNN 在处理长期依赖问题时存在的梯度消失和梯度爆炸问题。LSTM 通过单元状态存储和更新信息, 使得模型能够记住历史信息, 并且有选择地进行遗忘或者更新; 通过门控机制 (输入门、遗忘门、输出门), LSTM 能够对信息的流动进行有效的控制, 从而保留和传递长距离信息, 提升对长期依赖关系的建模能力。考虑到 LSTM 具有出色的对时序数据进行建模的能力, 论文选择使用 LSTM 来获取 EEG 信号中的时序长期依赖信息。

LSTM 用于处理形状为 (*sequence\_length*, *batch\_size*, *input\_size*) 的三维序列数据, 其中 *sequence\_length* 为序列的长度, *batch\_size* 为批量处理的样本数量, *input\_size* 为每个时间步的特征维度。其输出同样是三维序列数据, 形状与输入序列类似, 具体为 (*sequence\_length*, *batch\_size*, *hidden\_size*), 其中, *hidden\_size* 为 LSTM 内部的隐藏维度。LSTM 由一系列元胞连接而成, 这些元胞按时间顺序逐个处理输入序列, 在每一个时间步内, LSTM 元胞进行以下操作: 计算当前时间步的输入允许计入单元状态的程度; 判断单元状态中应当被丢弃的信息; 计算候选单元状态; 更新单元状态; 生成当前时间步的隐藏状态, 以及将隐藏状态传递给下一个时间步或输出。

EEG 信号是使用脑电采集设备采集的大脑在一段时间内的电压变化, 天然具有时序属性, 原始 EEG 信号为时间和通道的二维数据, 其中, 时间维度可以对应 LSTM 输入的序列长度, 通道维度可以对应 LSTM 输入的特征维度。由此, 可以将 EEG 信号以 (*time\_samples*, *batch\_size*, *channels*) 的形式与 LSTM 的输入对齐, 即将通道维度视为对应时间步 (采样点) 上数据所具有的特征。

虽然 LSTM 通过其内部的单元状态和门控结构缓解了 RNN 中存在的长期依赖的问题, 但在处理较长的时间序列数据时, 其性能仍存在局限性。LSTM 在遍历时间序列的过程中, 对每个时间步的输入信息均采用统一的方式通过隐藏状态进行累积和更新, 并不具备针对当前任务为各个时间步的输入赋予不同重要性的能力。因此, 有必要在 LSTM 模型中引入全局时空注意力机制 SCoT 模块, 使得模型能够动态地对 EEG 信号中的重要数据进行关注, 理解不同位置之间的深层次依赖关系, 捕获更为完整的长距离信息, 从而获得更为精确和全面的特征表示, 提升模型处理 EEG 信号分类任务时的性能和泛化能力。

LSTM 结合 SCoT 得到的网络结构如图 3.13 所示, 将其称之为 LS-Net (LSTM-SCoT Net)。LS-Net 的输入为  $X \in \mathbb{R}^{C \times T}$ , 其中,  $C$  为通道数量, 表示每个时间采样点所具有的特征数,  $T$  为时间, 表示表示时间轴上的采样点序列长度。在处理过程中, LSTM 按时间序列递进地对每个采样点进行加工, LSTM 的每个层 (Layer) 的输出结果为  $Y \in \mathbb{R}^{H \times T}$ , 其中,  $H$  为隐藏层数量。在经过每个 Layer 处理后,  $Y$  由 SCoT 模块计算全局时空自注意力。需要说明的是, 尽管  $H$  来自于 LSTM 的隐藏层维度, 但在 SCoT 模块中, 仍然将

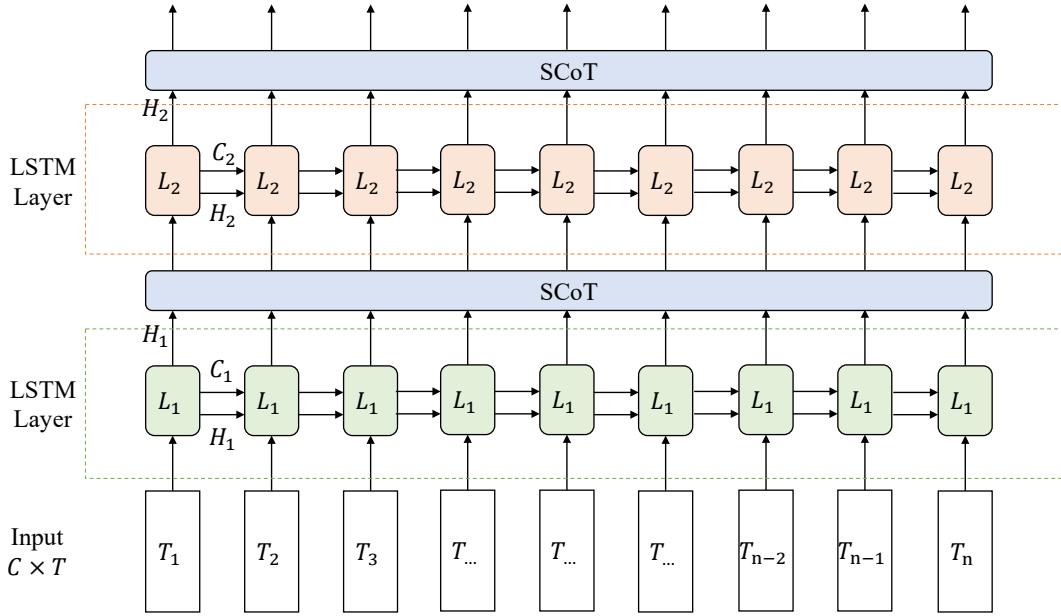


图 3.13 LS-Net 结构

其视为通道维度特征进行处理，以对来自 LSTM 输出的特征进行整合，从而提升网络捕捉输入 EEG 信号全局时空依赖关系的能力。

为了综合 LS-Net 和 DIS-Net 的优点，论文选择以并行分支的形式将 LS-Net 引入 DIS-Net 网络结构中。这是因为直接以串联形式加入 LS-Net 可能导致模型仅能捕捉到经过深层卷积获得的特征图的依赖关系，而非充分利用全局信息。并行分支的设计旨在全面利用 LS-Net 对全局依赖的建模能力，同时避免干扰卷积神经网络本身的局部特征提取功能。

将 LS-Net 模块引入 DIS-Net 后获得的新模型命名为 HA-FuseNet，其结构如图 ?? 所示。

为了整合 DIS-Net 与 LS-Net 所提取的特征信息，将二者获取的特征图在深度维度进行特征聚合。为了确保两个分支的数据能够聚合，需要对数据形状进行适配处理。此外，由于需要使用历史信息，LSTM 无法进行并行计算，因此同样需要对于输入 LSTM 的数据进行一定的处理，以对计算效率进行平衡。

论文设计了 C2R-Block 和 R2C-Block 两种模块，用于并行分支间的交互融合。在 C2R-Block 中，对数据执行时间维度的池化操作以实现下采样，将原始样本  $X \in \mathbb{R}^{C \times T}$  在时间轴上进行压缩，转化为  $X' \in \mathbb{R}^{C \times T'}$ ，其中， $C$  代表通道数（电极数）， $T$  代表原始时间序列长度， $T'$  则代表下采样后的时间序列长度。另一方面，在 R2C-Block 中，采用反卷积技术对 LS-Net 模块产生的输出进行时间维度上的上采样，使其与 DIS-Net 分支的数据维度相符。最后，在深度维度对经过上述处理的两类特征进行聚合，从而充分融合卷积神经网络捕获的局部特征信息、长短期记忆网络揭示的时序依赖特性，以及全

局自注意力模块发掘的全局时空依赖关系，以期全面提升模型对 EEG 信号的综合理解与分析能力。

### 3.4 基于 ShuffleNet 和 GhostNet 的网络轻量化

考虑到 BCI 系统对即时响应具有较高的要求，虽然 HA-FuseNet 的轴向卷积削减了一定的模型参数量，但为了追求更加卓越的实时性能表现，仍有进行进一步轻量化的必要。HA-FuseNet 模型的参数主要来源于密集连接模块、长短期记忆网络和全局自注意力。针对长短期记忆网络和全局自注意力，论文通过下采样削减了输入数据的规模。与此同时，对于密集连接模块以及其他卷积层，论文进行针对性的轻量化卷积设计，并将在下文中对轻量化卷积模块的设计思路与方案加以阐述。

#### 3.4.1 经典轻量级网络结构

MobileNet<sup>[79]</sup>由 Google 团队提出，其核心思想在于引入了深度可分离卷积（Depth-wise Separable Convolution），通过深度卷积处理单个输入通道，然后使用点积卷积（也称为逐点卷积或  $1 \times 1$  卷积）跨通道整合信息。这种分解大幅减少了计算成本和模型大小，同时保持了较高的精度。在 V2 版本中，引入了反转残差（Inverted Residual）结构<sup>[80]</sup>，不同于常规残差块的设计，它在网络较窄的地方采用线性瓶颈层增加深度，而非在宽度较大的地方。此外，MobileNetV2 取消了线性瓶颈层后的小维度特征上的 ReLU 激活函数，以此来保持更多的非线性表达能力。V3 版本在前两代的基础上，通过结合人类专家指导的网络设计原则与自动化的网络结构搜索（Network Architecture Search, NAS）技术<sup>[81]</sup>，进一步定制了模型结构，即每个卷积层的具体通道数和操作类型都根据 NAS 的结果确定。此外，MobileNetV3 还引入了 Hard Swish 这一新的激活函数以及其他优化策略。

ShuffleNet<sup>[82]</sup>由旷视科技的研究团队所开发，其设计目标在于实现模型计算效率与预测准确性的均衡优化。为了解决分组卷积可能导致的组间特征孤立问题，ShuffleNet 引入了通道混洗（Channel Shuffle）技术，即在执行完分组卷积之后，对各组内的特征图进行有序重组，以促进不同组间的特征信息相互交织和传播，从而增强模型的特征表达能力。此外，ShuffleNet 针对分组卷积的配置进行了精细化调整，科学地配置组的数量和各组内的通道数，以便在有限的计算资源条件下最大限度地提升模型性能。在 V2<sup>[83]</sup>版本中，研究团队提出了一套高效卷积神经网络设计原则，并据此对 ShuffleNetV1 架构做出了改进。他们摒弃了  $1 \times 1$  的分组卷积操作，转而采用通用卷积来替代，并提出了通道分割（Channel Split）策略。具体而言，模型将输入特征图均匀划分为两个部分，一部分不经额外计算直接向下传输，另一部分则经历一系列计算后再与前者合并。在 concatenate 操作之后，继续实施通道混洗操作，从而有效提升模型在轻量化条件下的性能表现。ShuffleNet 的结构如图 3.14 所示，图 3.14 (a) 是 ShuffleNet 的基本单元，图 3.14 (b)

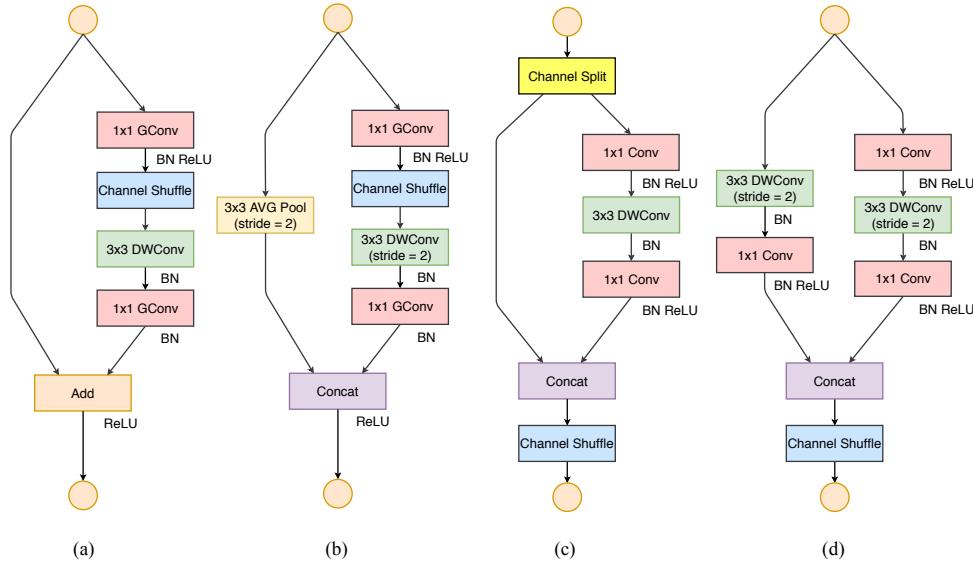


图 3.14 ShuffleNet 结构<sup>[83]</sup>

在 ShuffleNet 基本单元的基础上增加了下采样，图 3.14 (c) 是 ShuffleNetV2 的基本单元，图 3.14 (d) 在 ShuffleNetV2 基本单元的基础上增加了下采样。

**GhostNet**<sup>[84]</sup> 由华为诺亚方舟实验室提出，其核心思想在于提出了幻影模块（Ghost Module），该模块针对深度神经网络中存在的潜在冗余和相关性较高的特征图问题，通过高效的计算策略提炼出“幻影特征图”。研究团队发现，在许多情况下，多个特征图可能蕴含着相似的模式信息，这意味着部分特征图的信息实质上可以从其他特征图中推衍得出，犹如“幻影”。鉴于此，GhostNet 摒弃了对每组特征图均采用标准卷积的传统做法，转而采用一种更为经济高效的计算方式去合成这类“幻影特征图”。在具体实施中，首先运用有限数量的标准卷积层提取基础特征图，以此严格控制参数规模，接下来通过对基础特征图施加一组精心设计的线性变换（通常是深度可分离卷积中的逐点卷积操作），高效地生成大量辅助特征图，这些辅助特征图被视为原始特征图的“幻影”。最后，Ghost 模块将“幻影特征图”与标准卷积产生的少量核心特征图整合在一起，以有效模拟传统卷积所带来的丰富表征能力，在保持高精度的同时，降低了模型的计算复杂性和参数规模。Ghost 模块的结构如图 3.15 所示，其中，Identity 操作由标准卷积生成的特征图， $\Phi$  为廉价线性变换。

### 3.4.2 基于 ShuffleNet 和 GhostNet 改进的轻量化卷积模块

在 HA-FuseNet 架构中，已经通过轴向卷积削减了一定的模型参数量。因此，本节在现有成果的基础上，主要结合 ShuffleNetV2 的基础单元和 GhostNet 所提出的 Ghost 模块进行了针对性改良。

ShuffleNetV2 的优势体现在其利用通道混洗操作促进不同特征图组之间的信息交互，特别是在 Channel Split 机制中，它将特征图分割成两个部分，其中一个部分不经额外

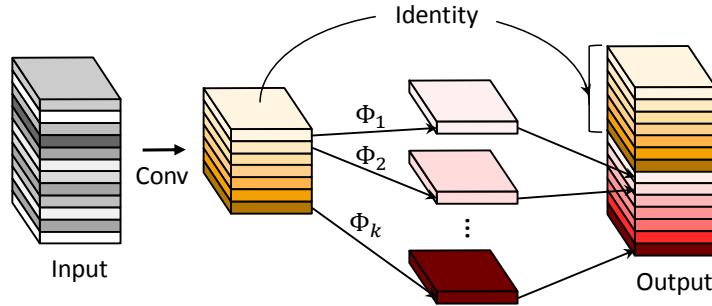
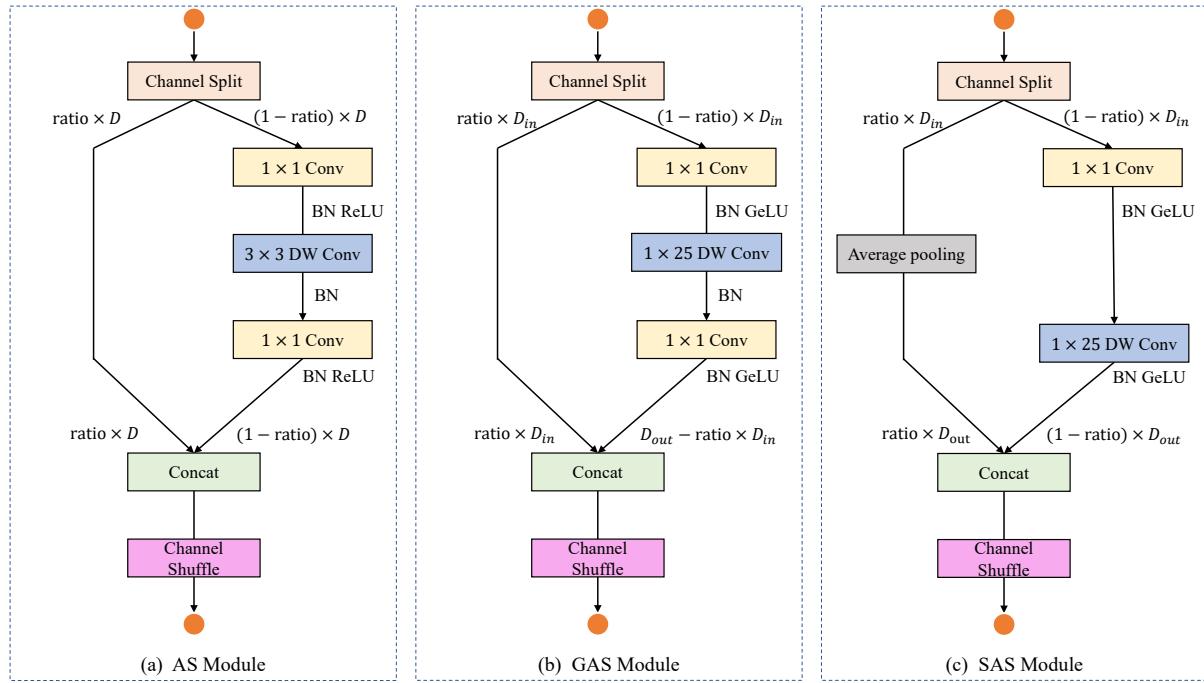
图 3.15 Ghost 模块结构<sup>[84]</sup>

图 3.16 AS、GAS、SAS 模块结构

计算直接向下进行传递，这种设计有助于减少模型分支内的参数量。然而，ShuffleNetV2 在进行 Channel Split 时，通常是对特征图进行均匀划分，这种方式难以灵活调整直传特征图的比例，可能导致对重要特征信息的有效利用率不足，并且在极端情况下可能影响模型精度。此外，ShuffleNetV2 的基础结构在深度维度上的特征变换主要依赖于  $1 \times 1$  卷积，这在一定程度上限制了其内在的深度特征变换能力。

为了提升 ShuffleNetV2 基本模块在深度特征变换方面的能力，论文提出了一种可调节分支比例的 Shuffle 模块（Adaptive Shuffle Module, AS 模块），其架构如图 3.16 (a) 所示。假设卷积层的输入深度为  $D_{in}$ ，输出深度为  $D_{out}$ ，针对不同的输入输出深度关系（即  $D_{in} \leq D_{out}$  和  $D_{in} > D_{out}$ ），论文对 AS 模块进行了优化，分别形成了图 3.16 (b) 所示的 GAS 模块（Growing Adjustable Shuffle Module）和图 3.16 (c) 所示的 SAS 模块（Straight Adjustable Shuffle Module）。图中的 DW Conv 指深度卷积（Depthwise Convolution）。

对于 GAS 模块，在保留原有直接传递分支的同时，对进行深度变换的分支进行了如

下改进：根据采样频率特性调整卷积核大小，为了描述的简洁性，以时间域特征提取卷积模块为例，并将其卷积核大小设定为  $1 \times 25$ ，对于采样频率为 250Hz 的 BCI Competition IV Dataset 2A 数据集而言，意味着该卷积核以 0.1 秒的时间窗口进行特征捕获。进入变换分支的特征图首先通过  $1 \times 1$  卷积实现深度变换，紧接着通过  $1 \times 25$  的深度卷积进行特征提取，最后再次通过  $1 \times 1$  卷积将深度变换至输出需要的深度。在此过程中，加入了批归一化和 GELU 激活函数。这样设计的目的是优化 GAS 模块对 MI-EEG 分类任务的性能。

对于权重参数为  $ratio$  的 GAS 模块，其直接传递分支的特征图输入数量  $BS_{in}$  设定为  $BS_{in} = D_{in} \times ratio$ ，对应的输出数量  $BS_{out}$  同样为  $BS_{out} = D_{in} \times ratio$ 。而变换分支的特征图输入数量  $BT_{in}$  则设定为  $BT_{in} = D_{in} - BS_{in}$ ，其输出特征图数量  $BT_{out}$  计算为  $BT_{out} = D_{out} - BS_{out}$ 。这样的设计使得 GAS 模块能在维持模型轻量化的同时，更精准地针对 MI-EEG 分类任务进行深度特征提取与整合。

对于 SAS 模块，其在变换分支的设计上采用了与 GAS 模块一致的改进措施，但去掉了最后的  $1 \times 1$  卷积层，而由第一个  $1 \times 1$  卷积层将深度降维至输出维度，再由  $1 \times 25$  卷积层进行特征的提取。然而，在直接传递分支上，进一步引入了深度维度上的平均池化操作，目的在于在不增添额外参数的前提下，有效减少特征图的数量。SAS 模块的池化分支的特征图输入数量  $BS_{in}$  设定为  $BS_{in} = D_{in} \times ratio$ ，对应的输出数量  $BS_{out}$  设定为  $BS_{out} = D_{out} \times ratio$ 。而变换分支的特征图输入数量  $BT_{in}$  则设定为  $BT_{in} = D_{in} - BS_{in}$ ，其输出特征图数量  $BT_{out}$  计算为  $BT_{out} = D_{out} - BS_{out}$ 。相较于原始的 Shuffle unit，GAS 模块和 SAS 模块在功能上进行了扩展，不仅包含了特征维度的升维和降维处理，而且还引入了一个可调节的权重参数  $ratio$ ，以动态控制两个分支的特征图数量。这样一来，这两种模块在处理 MI-EEG 分类任务时，能够更加灵活和高效地调整特征空间的分布和深度特征的提取，从而更好地适应任务需求和优化模型性能。

论文针对通道混洗算法进行了优化升级，使其具备了对不同数量分组及不同数量组内特征图进行灵活、均衡混洗的能力。这使得无论组及组内特征图的数量如何变化，该算法都能够实现有效的特征交互与信息整合，从而增强了网络对多种复杂场景的适应性和泛化能力。

### 3.5 本章小结



## 4 实验结果分析与模型评估

### 4.1 实验环境

论文构建的 HA-FuseNet 是在实验室搭建的服务器上进行搭建和训练的，所用的软/硬件环境如表 4.1 所示：

表 4.1 实验环境

软/硬件名称	型号/版本
操作系统	Ubuntu 20.04.6 LTS
CPU	Intel(R) Xeon(R) Gold 5218R CPU @ 2.10GHz
GPU	NVIDIA GeForce RTX 3090
内存	128GB
显存	24GB
CUDA	11.8
Python	3.11.5
Pytorch	2.0.1
MNE	1.6.0
Numpy	1.26.3

### 4.2 数据与实验准备

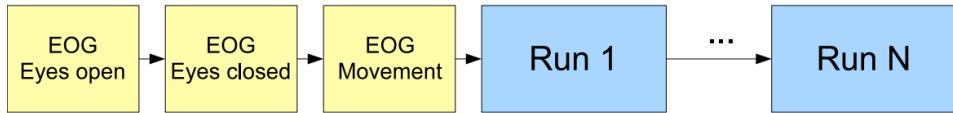
#### 4.2.1 运动想象脑电图数据集

运动想象脑电图是使用脑电采集设备从头皮上获取的人类大脑的神经元活动时产生的生物电位信号，能够反映大脑皮层和深层结构的功能状态及其异常变化。脑电图信号的采集过程主要包括以下几个步骤：

(1) 采集准备：根据国际 10-20 标准导联系统或其他标准定位方案，将电极安放在被试头皮的不同位置，以捕获不同脑区的电位信号。电极通常通过电极帽或电极盘固定，以确保位置的稳定和正确。由于人体脑电信号强度微弱，通常会通过与脑电采集设备相连的放大器对脑电信号进行增强和记录；

(2) 信号记录：当大脑神经元兴奋或抑制时，会产生微弱的电位变化，这些电位变化传导到头皮表面，形成可测量的电压差，由脑电采集系统进行捕获和放大。脑电采集系统通常以每秒进行  $N$  次连续采集的方式工作，即采集频率为  $N\text{Hz}$ ；

(3) 数字化：根据脑电采集系统的设置，对被捕获的脑电信号进行一定的处理，包括通过模数转换器（Analog to Digital Converter, ADC）将模拟信号转换为数字信号。

图 4.1 2A session 采集模式<sup>[85]</sup>

在临床和科研应用中，脑电图信号因其非侵入性、实时监测性、对大脑功能活动的敏感性等特点，已经在大脑解码领域获得了广泛的应用。运动想象领域有多个公开的脑电图信号数据集，论文主要选取 BCI Competition IV Dataset 2A<sup>[85]</sup> 数据集和 BCI Competition IV Dataset 2B<sup>[86]</sup> 数据集作为模型训练和测试的数据集。BCI Competition（脑机接口竞赛）是一项由德国柏林洪堡大学和柏林工业大学发起的国际性脑机接口技术竞赛，旨在推动脑机接口技术的创新和发展。在第四届比赛（BCI Competition IV）中，主办方提供了多项数据集，其中，运动想象领域的 2A 和 2B 数据集在相关研究中被广泛使用。

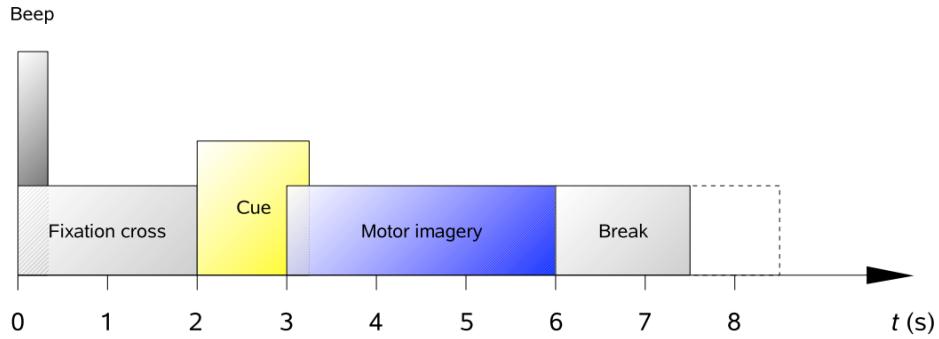
### (1) BCI Competition IV Dataset 2A

BCI Competition IV Dataset 2A 数据集采集了 9 名被试的 MI-EEG 信号，包括 4 类运动想象任务，即左手、右手、双脚和舌头。每名被试在不同的日期进行了两次采集/会话 (session)，两个 session 分别被作为训练集 (T) 和测试集 (E)，数据以.gdf 的格式存储，因此，每名被试具有两个文件，如对于一号被试而言，存在 A01T.gdf 和 A01E.gdf 两个文件，其中，训练集 A01T.gdf 包含标签信息，A01E.gdf 不包含标签信息，使用额外的 A01E.mat 文件提供标签。

session 采集的过程如图 4.1 所示，在采集开始时，进行约五分钟的 EOG 记录，包括两分钟的睁眼模式、一分钟的闭眼模式和一分钟的眼球运动模式。其中，四号被试的测试集只具有眼球运动的 EOG 记录。

在 EOG 记录之后，进行六次采集 (run)，在每个 run 中，各进行 12 次每类运动想象任务，这些任务的顺序是随机的，由此，一个 run 包含 48 次试验 (trial)，一个 session 包含 288 次试验 (trial)。一个 trial 的流程如图 4.2 所示，测试开始时 ( $t=0s$ )，一个十字图形出现在屏幕上，伴有简短的提示音；两秒后 ( $t=2s$ )，运动想象任务的指示箭头（分别指向左、右、下、上，对应左手、右手、双脚及舌头运动）出现在屏幕上，持续约 1.25 秒，每名被试进行运动想象任务直到十字图形消失 ( $t=6s$ )。在这个过程中没有任何反馈。

采集过程中，以 250Hz 进行 EEG 信号采样，并在 0.5Hz 至 100Hz 之间进行了带通滤波，放大器的灵敏度被设置为  $100\mu V$ ，并使用了 50Hz 的陷波滤波器用以抑制线路噪声。头皮电极的位置按照国际 10-20 标准导联确定，共使用 22 个电极（通道），此外，还使用了 3 个不参与分类的 EOG 电极用以记录眼电信号。

图 4.2 2A trial 与无视觉反馈的 2B trial 的采集模式<sup>[85, 86]</sup>

在 BCI 领域，事件描述某一种波形或任务的起始点，数据上，事件表现为一个三元组：第一个元素以整数来描述的事件起始采样点；第二个元素对应当前事件来源的刺激通道（Stimulus Channel）的先前值（Previous Value），大多数情况下为 0；第三个元素表示该事件的类型（Identify, id）。BCI Competition IV Dataset 2A 数据集中共包含 11 类事件（Event），其中，与数据处理相关的事件如表 4.2 所示，其中，拒绝试验是指由于质量欠佳或受试者未能有效完成而被专家标注出的试验数据。

表 4.2 2A 事件类型列表

事件类型	描述
768	trial 开始
769	左手运动想象任务 (class 1)
770	右手运动想象任务 (class 2)
771	双脚运动想象任务 (class 3)
772	舌头运动想象任务 (class 4)
783	未知运动想象任务 (测试集)
1023	拒绝试验

论文遵循 BCI Competition 竞赛设置，使用 T 文件为训练集，E 文件为测试集，针对每名被试进行被试内实验和被试间实验。则在不剔除拒绝试验数据的情况下，每名被试的训练集和测试集切片数量分别为：288, 288。

## (2) BCI Competition IV Dataset 2B

BCI Competition IV Dataset 2B 采取了类似 2A 数据集的采集方式，采集了 9 名被试的 2 类 MI-EEG 信号（左手、右手），每名被试在不同时间进行了五次 session，数据以.gdf 格式存储，每名被试有五个文件，例如，对于一号被试而言，B0101T.gdf、B0102T.gdf 和 B0103T.gdf 为训练集，B0104E.gdf 和 B0105E.gdf 为测试集。其中，前两个文件不包含视觉反馈，后三个文件包含视觉反馈，即在试验过程中，通过屏幕上的笑脸图案对运

动想象任务是否被正确执行予以反馈。无视觉反馈的 session 包含 6 次 run，每个 run 包含 20 次 trial（左手、右手各 10 次，随机排布），有视觉反馈的 session 包含 4 次 run，每个 run 包含 40 次 trial（左手、右手各 20 次，随机排布）。无视觉反馈的 trial 的采集模式与 2A 相同，如图 4.2 所示。

BCI Competition IV Dataset 2B 的事件类型与 2A 的事件类型相似，但不包含双脚和舌头运动想象任务。此外，不同于 2A 数据集的 22 个电极，2B 数据集仅使用 3 个电极记录数据。

为了保证实验数据只涉及运动想象，而不涉及视觉反馈，论文选择无视觉反馈的两个 session 分别作为训练集和测试集，则在不剔除拒绝试验数据的情况下，每名被试的训练集和测试集切片数量分别为：120，120。

综上所述，论文所使用的数据的信息如表 4.3 所示。

表 4.3 数据集信息

数据集	BCI IV 2A	BCI IV 2B
被试数量	9	9
类别数量	4	2
通道数量	22	3
频率范围	0.5-100Hz	0.5-100Hz
采样频率	250Hz	250Hz
训练集数据量	288（每被试）	120（每被试）
测试集数据量	288（每被试）	120（每被试）

#### 4.2.2 数据预处理

EEG 信号具有低信噪比、非平稳、空间变异性等特性，并且通常具有较小规模的数据集，一般来说，对数据进行一定的预处理有助于后续分类任务。然而，论文基于端到端网络的思想构建模型，因此尽可能地对预处理操作进行削减。论文进行的数据预处理方法如下：

##### (1) 数据提取与切片

MI-EEG 原始信号储存在.gdf 格式的文件中，除目标脑电信号之外，原始文件还包括 EOG 信号、间隙缺失值、与运动想象任务不直接相关的事件等。因此，在提取数据时，有必要进行一定的处理：

首先，剔除 EOG 通道数据，仅保留 EEG 通道数据，对于用于分隔 run 的缺失值（编码为非数字，NaN），使用对应通道的均值替代，以确保数据的连续性和完整性；

其次，对与运动想象任务直接相关的事件进行筛选和提取，并对相关时段进行切

分。对于 BCI Competition IV 的 2A/2B 数据集，直接相关事件为四类/二类运动想象任务事件，并由提示信号（Cue）标识任务的开始，论文对连续数据进行切片，提取从 Cue 出现后的第 1 秒至第 4 秒（trial 周期的第 3 秒至第 6 秒）的数据段，作为事件对应的运动想象任务持续期间产生的 EEG 信号，在后续加以分析。对于采样率为 250Hz 的数据集而言，3 秒的数据区间将包含 750 个采样点。

需要说明的是，论文在数据提取过程中并未对拒绝试验进行剔除，同时未使用滤波器对 EEG 信号的频率进行过滤，以对真实应用情境中可能出现的多样化数据表现进行模拟，对模型自主识别各种频率成分并提取有效特征的能力进行评估。

## （2）标准化

归一化操作的目的在于对数据的特征尺度进行统一，从而消除奇异数据导致的不良影响，提高模型的训练效率和稳定性。EEG 信号常用的归一化方法有 Z-score 标准化、最大最小值归一化等。

Z-score 标准化的操作过程如公式 4.1 所示，其中， $x$  为原始数据， $\mu$  表示数据的均值， $\sigma$  表示数据的标准差，Z-score 通过均值和标准差对数据进行操作，使得处理后的数据符合均值为 0，方差为 1 的标准正态分布。

$$x = \frac{x - \mu}{\sigma} \quad (4.1)$$

最大最小值归一化的操作过程如公式 4.2 所示，其是一种线性变换操作，将数据映射至  $[0, 1]$  区间，其中， $X$  为一组通道数据。最大最小值归一化计算简单，但对具有波动性的 EEG 信号而言，将数据缩放至  $[0, 1]$  区间容易导致数据特征的损失。

$$x = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (4.2)$$

因此，为了尽可能保留 EEG 信号的特征，论文采用 Z-Score 标准化方法对 EEG 信号进行处理，以提高模型训练的速度和稳定性。

## （3）数据增强

MI-EEG 信号通常具有较小规模的数据集，因此，通过一定的数据增强操作扩大数据规模，有助于提升网络训练效果，防止过拟合现象的发生。然而，在 BCI 系统的实际应用中，数据增强操作可能会导致训练阶段数据处理压力的增长，因此，在论文中，数据增强是一项可选操作，其目的在于提高模型分类精度，论文将分别使用进行数据增强和不进行数据增强的数据集进行训练。

EEG 信号的数据增强方法主要有以下几种：

（1）添加随机噪声：在 EEG 信号上叠加高斯白噪声、有色噪声等不同类型的噪声，模拟真实环境中的噪声干扰，能够提升模型的抗噪性和鲁棒性；

(2) 滑动窗口：设定时间滑动窗口的长度小于运动想象任务持续的时间长度，将滑动窗口内的数据视作一次事件，通过滑动窗口对数据进行切片，有助于模型学习 EEG 信号随时间变化的特征；

(3) 频率混叠：在保持 EEG 信号特性的前提下，将多种不同频率成分进行混叠，使得模型能够学习更广泛的频率特征；

(4) 生成对抗网络（Generative Adversarial Network, GAN）：基于 GAN 生成拟真的 EEG 数据，对原始数据进行扩充。

由于在预处理阶段没有采取频率滤波的操作，论文采取频率混叠的方式对 EEG 信号进行增强，从而加强模型识别不同频率成分的能力。具体来说，

**Data:** A set  $E_o = \{e_1, e_2, \dots, e_n\}$  and a similarity threshold  $\theta$ .

**Result:** The Uniqueness-Index  $u(E_o)$  of the set  $E_o$ .

*counter*  $\leftarrow 0$ ;

**for** each item  $e_i \in E_o$  **do**

**for** each item  $e_j \in E_o$  where  $j \neq i$  **do**

*similarity*  $\leftarrow \text{CosSim}(v_i, v_j)$ ;

**if** *similarity*  $< \theta$  **then**

*counter*  $\leftarrow i + 1$ ;

**end**

**end**

**end**

$u(E_o) \leftarrow \frac{2 \times C}{n \times (n-1)}$ ;

**return**  $u(E_o)$ ;

算法 1 数据增强算法

#### 4.2.3 实验准备

(1) 评价指标

(2) 损失函数

(3) 参数配置

### 4.3 实验设计

#### 4.3.1 DIS-Net 实验设计

为了确定模型架构的最佳设置，本节对 DIS-Net 构建过程中不同的模型架构进行实验，从而逐步构建起性能最优的模型。

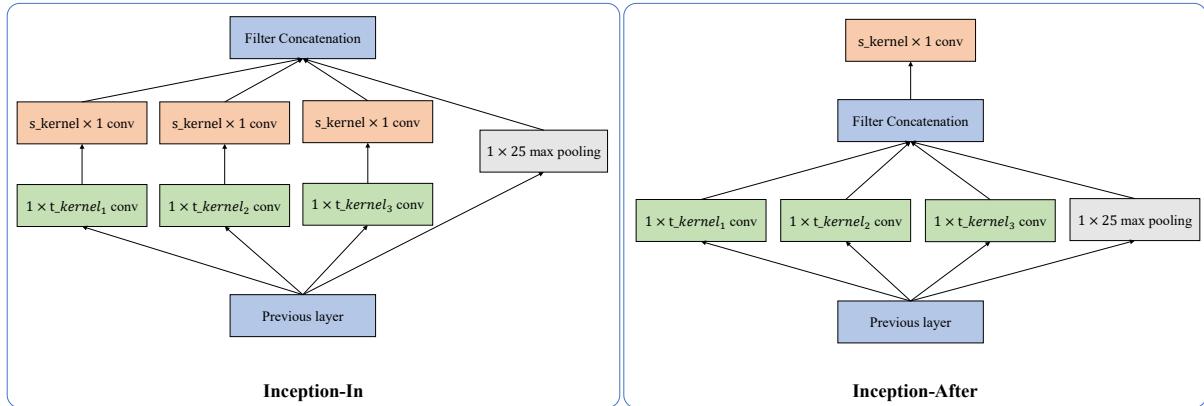


图 4.3 Inception 模块引入空间卷积层的方式

### (1) Inception 模块引入空间卷积层的方式

在 DI-Net 中，空间卷积层有两种不同的方式融入基于 Inception 改进的时间卷积层之后，一种是在每个 Inception 模块内部的分支结构上增加空间卷积层，另一种则是在整个 Inception 模块之后附加空间卷积层。图 4.3 展示了这两种引入方式的区别，将这两种方式分别称为分支内融合（Inception-In）和模块后融合（Inception-After），需要说明的是，图中省略了网络的其他结构，如瓶颈层等，以尽可能简洁地展现不同引入方式的差异。其中， $s\_kernel$  表示空间卷积核， $t\_kernel_i$  表示第  $i$  个分支的时间卷积核。

为了比较 Inception-In 与 Inception-After 的性能差异，论文在 BCI Competition IV Dataset 2A 数据集上设计实验进行对比。在实验设置阶段，固定了 Inception 模块的层次数量、分支数量等参数，实验结果如表 4.4 所示。在此，重点关注两项评价指标——准确率（Accuracy, ACC）和 Kappa 一致性系数（Kappa），这两项指标均基于数据集中九位被试的平均表现。实验结果显示，Inception-After 方式在准确率和一致性系数上均表现更优。这一优势可能源自两方面的原因：一方面，虽然 Inception-In 模式借鉴了 FBCSP 算法的分频段处理思路，但在 Inception 分支内部直接进行空间特征提取的过程中，损失了部分空间全局信息；另一方面，Inception-In 结构具有相对更大的参数规模，这可能导致模型在有限样本条件下更容易出现过拟合现象。基于以上分析和实验验证，论文选择以 Inception-After 的方式布局时间卷积层与空间卷积层。

表 4.4 Inception-In、Inception-After 实验结果对比

Models	ACC(%)	Kappa
Inception-In	63.31	0.51
Inception-After	<b>75.35</b>	<b>0.70</b>

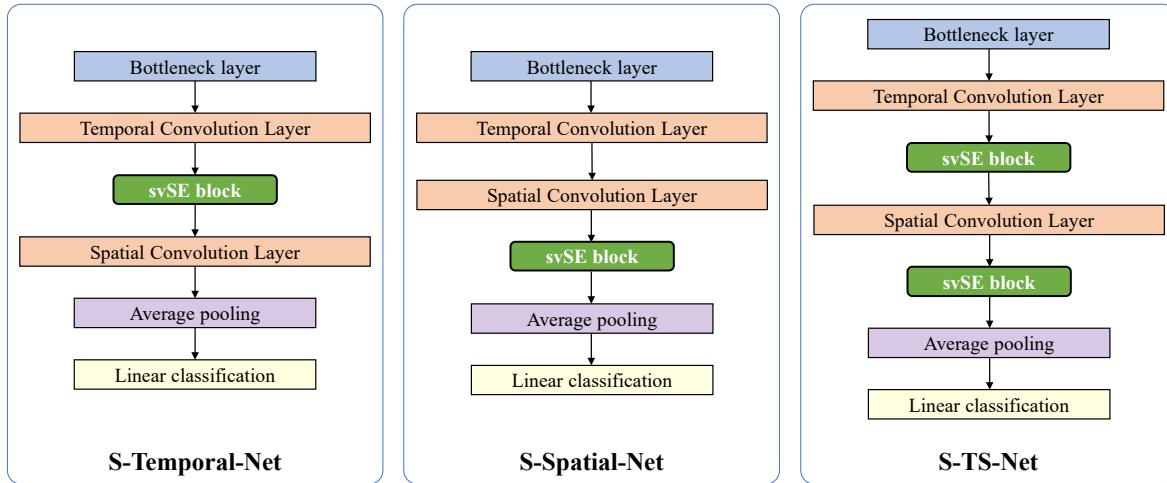


图 4.4 DI-Net 引入注意力模块的方式

## (2) svSE 模块的引入方式

DIS-Net 中，由于 DI-Net 的特征提取过程分为时间卷积和空间卷积两个阶段，svSE 模块可采取以下三种引入方式：其一是在时间卷积层后引入；其二是在空间卷积层后引入；其三是同时在时间卷积层和空间卷积层之后引入。图 4.4 展示了这三种引入 svSE 模块的方式，从左至右分别是时间卷积层后引入 svSE 模块、空间卷积层后引入 svSE 模块，以及在时间卷积和空间卷积层后均引入 svSE 模块。将这三种引入方式对应的模型分别简称为 S-Temporal-Net、S-Spatial-Net、S-TS-Net。

表 4.5 展示了 S-Temporal-Net、S-Spatial-Net、S-TS-Net 三种模型在 BCI Competition IV Dataset 2A<sup>[71]</sup> 数据集上的对比实验结果。实验采用固定的参数，表格中展示的准确率（Accuracy, ACC）和 Kappa 一致性系数（Kappa）指标为数据集中九位被试的平均表现，标准差（Standard Deviation, SD）则为准确率的标准差。从准确率和一致性分析，S-ST-Net 模型的效果优于其他两种模型，与经验相符。此外，S-Temporal-Net 模型的效果优于 S-Spatial-Net 模型，其原因可能在于，空间卷积层沿通道维度的卷积使得数据损失了部分特征，进而减弱了 svSE 模块提取关键特征权重的能力，而时间卷积层保留了大部分深度信息和通道信息，因此，在时间卷积层之后加入 svSE 模块能够帮助模型更好地捕捉深度和空间的特征。从标准差分析，S-TS-Net 模型的准确率波动幅度较小，对不同被试的 MI-EEG 分类效果相对均衡，另外两种模型在不同被试间的分类精度则存在较为明显的差异。实验数据显示，S-TS-Net 模型取得了更好的效果，因此，论文采用同时在时间卷积层和空间卷积层之后引入 svSE 模块的方式，将这种结构的模型称为 DIS-Net。

表 4.5 svSE 模块引入位置对比

Models	ACC(%)	Kappa	SD
S-Temporal-Net	78.09	0.71	10.38
S-Spatial-Net	77.16	0.69	10.24
S-TS-Net	<b>78.55</b>	<b>0.71</b>	<b>9.46</b>

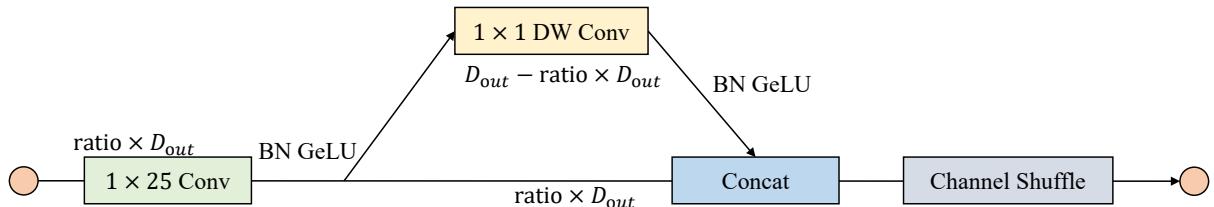


图 4.5 SG 模块结构

#### 4.3.2 LS-Net 实验设计

#### 4.3.3 各模块消融实验

#### 4.3.4 超参数优化实验

#### 4.3.5 BCI Competition IV Dataset 2A 数据集上的对比试验

#### 4.3.6 BCI Competition IV Dataset 2B 数据集上的对比试验

### 4.4 本章小结

GhostNet 的优势在于其 Ghost 模块设计，该模块凭借对特征图间内在相似性的有效利用，实现了低成本的特征转换操作，同时，GhostNet 中同样使用了权重参数  $ratio$ ，用以控制经由廉价线性转换进行操作的特征图的比例，其中，使用普通卷积的分支其特征图数量为  $D_{out} \times ratio$ ，其中， $D_{out}$  为输出特征图数量，使用深度卷积的分支其特征图数量为  $D_{out} - D_{out} \times ratio$ 。

然而，Ghost 模块在运用线性变换与传统卷积相结合的方式来产生新特征图的过程中，尚存在一定的局限性，即生成的特征图之间交互作用不充分。鉴于此，论文进一步改良了 Ghost 模块，引入了通道混洗机制，构建出 SG 模块（Shuffle Ghost Module）。SG 模块的结构如图 4.5 所示，通过通道混洗机制增强不同特征图之间的相互作用，进而提升网络的整体表现力和特征学习能力，图中的 DW Conv 指深度卷积。此外，SG 模块采用与 GAS 模块相同的方式对 Ghost 模块内部的卷积核大小进行了修改，并加入了批归一化和 GELU 激活函数。

为了验证更适合的轻量化卷积模块，论文在 BCI Competition IV Dataset 2A 数据集

上进行了实验验证。由于轻量化卷积主要针对密集连接进行改进，实验使用 DI-Net，并固定了 Inception 层数、密集连接层数、学习率、*ratio* 等超参数。表展示了 GAS/SAS、SG 和原始（Origin）模块的对比实验结果，其指标为参数量（Parameters）、浮点运算数（Floating Point Operations, FLOPs）和准确率（Accuracy, Acc），其中，准确率为九位被试的平均值。

**表 4.6 轻量化卷积模块实验结果对比**

Models	Paramters	FLOPs	ACC(%)
GAS/SAS(0.5)	7.31K	130.37M	74.61
GAS/SAS(0.4)	7.83K	143.44M	73.80
GAS/SAS(0.6)	<b>6.80K</b>	<b>117.75M</b>	<b>74.61</b>
SG(0.5)	7.27K	129.93M	73.26
SG(0.4)	7.67K	139.73M	72.22
SG(0.6)	7.27K	129.93M	73.00
Origin	29.99K	690.37M	<b>75.62</b>

## 5 总结与展望



## 参考文献

- [1] 蒲慕明. 脑科学的未来[J]. 心理学通讯, 2019, 2: 80-83.
- [2] 蒲慕明. 脑科学研究的三大发展方向[J]. 中国科学院院刊, 2019, 34: 807-813.
- [3] JORGENSON L A, NEWSOME W T, ANDERSON D J, et al. The brain initiative: developing technology to catalyse neuroscience discovery[J]. Philosophical Transactions of the Royal Society B: Biological Sciences, 2015, 370(1668): 20140164.
- [4] OKANO H, MIYAWAKI A, KASAI K. Brain/minds: brain-mapping project in japan[J]. Philosophical Transactions of the Royal Society B: Biological Sciences, 2015, 370(1668): 20140310.
- [5] NADDAF M. Europe spent€ 600 million to recreate the human brain in a computer. how did it go?[J]. Nature, 2023, 620(7975): 718-720.
- [6] 中华人民共和国科学技术部. 科技部关于发布科技创新 2030 — “脑科学与类脑研究”重大项目 2021 年度项目申报指南的通知[EB/OL]. 2021[2022-06-01]. [https://service.most.gov.cn/kjjh\\_tztg\\_all/20210916/4583.html](https://service.most.gov.cn/kjjh_tztg_all/20210916/4583.html).
- [7] MING POO M, LIN DU J, IP N, et al. China brain project: Basic neuroscience, brain diseases, and brain-inspired computing[J/OL]. Neuron, 2016, 92(3): 591-596. <https://www.sciencedirect.com/science/article/pii/S0896627316308005>.
- [8] HE B, YUAN H, MENG J, et al. Brain-computer interfaces[J]. Neural engineering, 2020: 131-183.
- [9] PICHIORRI F, MORONE G, PETTI M, et al. Brain-computer interface boosts motor imagery practice during stroke recovery[J]. Annals of neurology, 2015, 77(5): 851-865.
- [10] ALTAHERI H, MUHAMMAD G, ALSULAIMAN M, et al. Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals: A review[J]. Neural Computing and Applications, 2023, 35(20): 14681-14722.
- [11] SAI C Y, MOKHTAR N, AROF H, et al. Automated classification and removal of eeg artifacts with svm and wavelet-ica[J]. IEEE journal of biomedical and health informatics, 2017, 22(3): 664-670.
- [12] HJORTH B. Eeg analysis based on time domain properties[J/OL]. Electroencephalography and Clinical Neurophysiology, 1970, 29(3): 306-310. <https://www.sciencedirect.com/science/article/pii/0013469470901434>.
- [13] LUKE R, WOUTERS J. Kalman filter based estimation of auditory steady state response parameters[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering,

2017, 25(3): 196-204.

- [14] HSU W Y. Eeg-based motor imagery classification using neuro-fuzzy prediction and wavelet fractal features[J/OL]. Journal of Neuroscience Methods, 2010, 189(2): 295-302. <https://www.sciencedirect.com/science/article/pii/S0165027010001858>.
- [15] WANG Y, GAO S, GAO X. Common spatial pattern method for channel selelction in motor imagery based brain-computer interface[C]//2005 IEEE engineering in medicine and biology 27th annual conference. IEEE, 2006: 5392-5395.
- [16] ANG K K, CHIN Z Y, ZHANG H, et al. Filter bank common spatial pattern (fbcsp) in brain-computer interface[C]//2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). IEEE, 2008: 2390-2397.
- [17] DORNHEGE G, BLANKERTZ B, KRAULEDAT M, et al. Combined optimization of spatial and temporal filters for improving brain-computer interfacing[J]. IEEE transactions on biomedical engineering, 2006, 53(11): 2274-2281.
- [18] LEMM S, BLANKERTZ B, CURIO G, et al. Spatio-spectral filters for improving the classification of single trial eeg[J]. IEEE transactions on biomedical engineering, 2005, 52(9): 1541-1548.
- [19] MOUSAVI E A, MALLER J J, FITZGERALD P B, et al. Wavelet common spatial pattern in asynchronous offline brain computer interfaces[J]. Biomedical Signal Processing and Control, 2011, 6(2): 121-128.
- [20] WANG S J, CAI Y X, SUN Z R, et al. Tinnitus eeg classification based on multi-frequency bands[C]//Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Proceedings, Part IV 24. Springer, 2017: 788-797.
- [21] KROUPI E, YAZDANI A, EBRAHIMI T. Eeg correlates of different emotional states elicited during watching music videos[C]//Affective Computing and Intelligent Interaction: Fourth International Conference, ACII 2011, Memphis, TN, USA, October 9–12, 2011, Proceedings, Part II. Springer, 2011: 457-466.
- [22] HERMAN P, PRASAD G, MCGINNITY T M, et al. Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2008, 16(4): 317-326.
- [23] PAWAR D, DHAGE S. Feature extraction methods for electroencephalography based brain-computer interface: A review.[J]. IAENG International Journal of Computer Science, 2020, 47(3).
- [24] MA X, QIU S, WEI W, et al. Deep channel-correlation network for motor imagery decoding from the same limb[J]. IEEE Transactions on Neural Systems and Rehabilitation

- Engineering, 2019, 28(1): 297-306.
- [25] GÓMEZ-HERRERO G, DE CLERCQ W, ANWAR H, et al. Automatic removal of ocular artifacts in the eeg without an eog reference channel[C]/Proceedings of the 7th Nordic signal processing symposium-NORSIG 2006. IEEE, 2006: 130-133.
- [26] LUO T J, ZHOU C L, CHAO F. Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network[J]. BMC bioinformatics, 2018, 19: 1-18.
- [27] OLIVAS-PADILLA B E, CHACON-MURGUIA M I. Classification of multiple motor imagery using deep convolutional neural networks and spatial filters[J]. Applied Soft Computing, 2019, 75: 461-472.
- [28] SHE Q, HU B, LUO Z, et al. A hierarchical semi-supervised extreme learning machine method for eeg recognition[J]. Medical & biological engineering & computing, 2019, 57: 147-157.
- [29] MA X, WANG D, LIU D, et al. Dwt and cnn based multi-class motor imagery electroencephalographic signal recognition[J]. Journal of neural engineering, 2020, 17(1): 016073.
- [30] CHU Y, ZHAO X, ZOU Y, et al. A decoding scheme for incomplete motor imagery eeg with deep belief network[J]. Frontiers in neuroscience, 2018, 12: 680.
- [31] HASSANPOUR A, MORADIKA M, ADELI H, et al. A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals[J]. Expert Systems, 2019, 36(6): e12494.
- [32] XU B, ZHANG L, SONG A, et al. Wavelet transform time-frequency image and convolutional network-based motor imagery eeg classification[J]. Ieee Access, 2018, 7: 6084-6093.
- [33] MIAO M, HU W, YIN H, et al. Spatial-frequency feature learning and classification of motor imagery eeg based on deep convolution neural network[J]. Computational and mathematical methods in medicine, 2020, 2020.
- [34] LI F, HE F, WANG F, et al. A novel simplified convolutional neural network classification algorithm of motor imagery eeg signals based on deep learning[J]. Applied Sciences, 2020, 10(5): 1605.
- [35] AMIN S U, ALSULAIMAN M, MUHAMMAD G, et al. Deep learning for eeg motor imagery classification based on multi-layer cnns feature fusion[J]. Future Generation computer systems, 2019, 101: 542-554.
- [36] SCHIRRMEISTER R T, SPRINGENBERG J T, FIEDERER L D J, et al. Deep learning with convolutional neural networks for eeg decoding and visualization[J]. Human brain mapping, 2017, 38(11): 5391-5420.

- [37] 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.
- [38] RIYAD M, KHALIL M, ADIB A. Mi-eegnet: A novel convolutional neural network for motor imagery classification[J]. Journal of Neuroscience Methods, 2021, 353: 109037.
- [39] MANE R, CHEW E, CHUA K, et al. Fbcnet: A multi-view convolutional neural network for brain-computer interface[J]. arXiv preprint arXiv:2104.01233, 2021.
- [40] ZHANG C, KIM Y K, ESKANDARIAN A. Eeg-inception: an accurate and robust end-to-end neural network for eeg-based motor imagery classification[J]. Journal of Neural Engineering, 2021, 18(4): 046014.
- [41] SONG Y, ZHENG Q, LIU B, et al. Eeg conformer: Convolutional transformer for eeg decoding and visualization[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022, 31: 710-719.
- [42] VASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [43] MUSALLAM Y K, ALFASSAM N I, MUHAMMAD G, et al. Electroencephalography-based motor imagery classification using temporal convolutional network fusion[J]. Biomedical Signal Processing and Control, 2021, 69: 102826.
- [44] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [45] MA X, QIU S, DU C, et al. Improving eeg-based motor imagery classification via spatial and temporal recurrent neural networks[C]//2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, 2018: 1903-1906.
- [46] WANG P, JIANG A, LIU X, et al. Lstm-based eeg classification in motor imagery tasks [J]. IEEE transactions on neural systems and rehabilitation engineering, 2018, 26(11): 2086-2095.
- [47] KUMAR S, SHARMA A, TSUNODA T. Brain wave classification using long short-term memory network based optical predictor[J]. Scientific reports, 2019, 9(1): 9153.
- [48] CHO H, AHN M, AHN S, et al. Eeg datasets for motor imagery brain–computer interface [J]. GigaScience, 2017, 6(7): gix034.
- [49] BLANKERTZ B, DORNHEGE G, KRAULEDAT M, et al. The non-invasive berlin brain–computer interface: fast acquisition of effective performance in untrained subjects[J]. NeuroImage, 2007, 37(2): 539-550.
- [50] SOLODKIN A, HLUSTIK P, CHEN E E, et al. Fine modulation in network activation

- during motor execution and motor imagery[J]. Cerebral cortex, 2004, 14(11): 1246-1255.
- [51] WATANABE T, HANAJIMA R, SHIROTA Y, et al. Effects of rtms of pre-supplementary motor area on fronto basal ganglia network activity during stop-signal task[J]. Journal of Neuroscience, 2015, 35(12): 4813-4823.
- [52] KASESS C H, WINDISCHBERGER C, CUNNINGTON R, et al. The suppressive influence of sma on m1 in motor imagery revealed by fmri and dynamic causal modeling[J]. Neuroimage, 2008, 40(2): 828-837.
- [53] PENFIELD W, RASMUSSEN T. The cerebral cortex of man; a clinical study of localization of function.[J]. 1950.
- [54] 周思捷,白红民. 事件相关去同步化和同步化方法在脑电信号分析中的研究进展[J]. 中国微侵袭神经外科杂志, 2018, 23: 141-143.
- [55] 郑潜. 自适应时频分析及脑电信号处理应用研究[D]. 浙江大学, 2022.
- [56] KROPOTOV J D. Quantitative eeg, event-related potentials and neurotherapy[M]. Academic Press, 2010.
- [57] ISHIHARA T. Activation of abnormal eeg by mental work[J]. Rinsho Nohha (Clin. Electroencephalogr.), 1966, 8: 26-34.
- [58] ISHIHARA T. The interaction between paroxysmal eeg activities and continuous addition work of uchida-kraepelin psychodiagnostic test[J]. Med. J. Osaka Univ., 1967, 18: 75-85.
- [59] HARI R, SALMELIN R. Human cortical oscillations: a neuromagnetic view through the skull[J]. Trends in neurosciences, 1997, 20(1): 44-49.
- [60] 胡广书, 汪梦蝶. 生物医学信号处理研究综述[J]. 数据采集与处理, 2015, 30: 915-932.
- [61] PFURTSCHELLER G, ARANIBAR A. Event-related cortical desynchronization detected by power measurements of scalp eeg[J]. Electroencephalography and clinical neurophysiology, 1977, 42(6): 817-826.
- [62] GLOROT X, BORDES A, BENGIO Y. Proceedings of the fourteenth international conference on artificial intelligence and statistics[C]//PMLR: volume 15. 2011: 315-323.
- [63] HENDRYCKS D, GIMPEL K. Gaussian error linear units (gelus)[J]. arXiv preprint arXiv:1606.08415, 2016.
- [64] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [65] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20 (11): 1254-1259.
- [66] SAKHAVI S, GUAN C, YAN S. Parallel convolutional-linear neural network for motor imagery classification[C]//2015 23rd European signal processing conference (EUSIPCO).

- IEEE, 2015: 2736-2740.
- [67] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [68] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
- [69] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI conference on artificial intelligence: volume 31. 2017.
- [70] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer, 2015: 234-241.
- [71] TANGERMAN M, MÜLLER K R, AERTSEN A, et al. Review of the bci competition iv[J]. Frontiers in neuroscience, 2012, 6: 21084.
- [72] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
- [73] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.
- [74] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks [J]. Advances in neural information processing systems, 2015, 28.
- [75] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]// Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [76] ROY A G, NAVAB N, WACHINGER C. Concurrent spatial and channel squeeze & excitation in fully convolutional networks[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I. Springer, 2018: 421-429.
- [77] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794-7803.
- [78] LI Y, YAO T, PAN Y, et al. Contextual transformer networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 1489-1500.
- [79] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
- [80] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and lin-

- ear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.
- [81] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-1324.
- [82] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6848-6856.
- [83] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.
- [84] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.
- [85] 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.
- [86] 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.



## 致谢

以简短的文字表达作者对完成论文和学业提供帮助的老师、同学、领导、同事及亲属的感激之情。