

分 类 号_____

密 级_____

U D C_____

编 号 10486

武汉大学

硕 士 学 位 论 文

基于 Inception 和混合注意力的运动想象
脑电图分类研究

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	2
1.2.2 基于深度学习的运动想象脑电图分类研究现状	3
1.3 研究内容	3
1.4 论文组织结构	3
2 运动想象脑电图分类和深度神经网络基础	5
2.1 脑电生理基础	5
2.2 深度神经网络基础	5
2.3 基于深度学习的运动想象脑电图分类基础	5
2.4 本章小结	5
3 基于脑电图信号的运动想象脑电图分类网络构建	7
3.1 基于 Inception 和混合注意力融合改进的端到端 MI-EEG 分类网络 HIT-Net	7
3.1.1 基础网络 BaseNet	9
3.1.2 局部混合注意力 svSE	14
3.1.3 基于密集连接优化 Inception 模块	19
3.1.4 基于 Transformer 构建并行分支	22
3.2 基于 HIT-Net 轻量化的端到端 MI-EEG 分类网络 HIT-lightNet	25
3.2.1 经典轻量级网络结构	25
3.2.2 ShuffleNet 结合 GhostNet 的轻量化卷积模块	27
3.3 基于 MEKT 算法改进的 MI-EEG 迁移学习算法 SV-MEKT	27
4 实验结果分析与模型评估	29
5 总结与展望	31
参考文献	33

致谢	39
----------	----

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^[1]，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^[2]。

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^[3]。日本于 2014 年启动“脑/思维”计划，研究集中在三个领域：猿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^[4]。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^[5]。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^[6]，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^[7]。

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^[8]。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^[9]、智能人机协作、认知神经科学研究等领域具有广阔的

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. 脑电信号,运动想象的数据集
2. 端到端,实时响应性能
3. 分类准确率
4. 可解释性

针对这些问题,论文进行了

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^[10],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)^[10]、基线校正(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^[11]。

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^[10],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^[12]，Luke 等人使用卡尔曼滤波来处理 EEG 信号的不确定性^[13]，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^[14]。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^[15]。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）^[16]、共稀疏空间谱模式（CSSSP）^[17]、共空间谱模式（CSSP）^[18]、小波共空间模式（WCSP）^[19] 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、功率谱密度、高阶谱、微分熵、傅里叶变换等方法。Wang 等人提出了一种基于快速傅里叶变换的功率谱特征提取方法^[20]，Kroupi 等人使用基于 Welch 方法的功率谱密度（PSD）估计来进行脑电情感分析^[21]，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^[22]。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^[23]，其主要包括匹配滤波、自回归模型、短时傅里叶变换、小波变换等方法，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^[10]。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

1.3 研究内容

1.4 论文组织结构

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

2.2 深度神经网络基础

2.3 基于深度学习的运动想象脑电图分类基础

2.4 本章小结

3 基于脑电图信号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, 针对这些问题进行了以下研究:

(1) 针对运动想象脑电图分类任务精度有待提升的问题, 构建了一种端到端的新模型 HIT-Net: 首先, 以 ShallowConvNet 模型为基础, 引入多尺度并行特征提取模块 (Inception 模块)^[24]、瓶颈层、深度可分离卷积和混合注意力机制, 旨在并行提取多尺度时间特征, 促进时空特征的深度融合, 同时提高模型对重要特征的关注度; 其次, 在以上模型的基础上, 将密集连接引入 Inception 模块的分支中, 以更好地利用 EEG 信号的局部细节和浅层特征; 最后, 引入并行多分支的 Transformer 模块以更好地利用 EEG 信号的全局信息, 并且针对 EEG 信号信噪比低的特点, 引入软阈值模块对 Transformer 进行改进, 从而减少噪声的干扰。

(2) 针对运动想象脑电图分类任务即时响应的需求, 对 HIT-Net 进行了轻量化设计, 构建了一种紧凑型端到端模型 VSNet: 首先, 删去参数量较大的 Inception 模块和 Transformer 分支, 保留网络的其他结构; 其次, 使用单层的深度可分离卷积替代 Inception 模块作为时间卷积层, 并对瓶颈层进行调整。

(3) 针对运动想象脑电图缺乏标注数据的问题, 设计了一种半监督的迁移学习算法: 首先, 通过 ICA 算法和 Wasserstein 距离^[25] 选择与目标域分布相似的 N 个源域; 其次, 将源域与目标域的特征在黎曼流形上对齐并映射到切线空间, 并通过源域与目标域特征的相似性计算特征的权重; 最后, 基于 MEKT 算法获取特征映射矩阵, 并通过加权投票获取最终的结果。

本章分三个部分对这些研究进行介绍。

3.1 基于 Inception 和混合注意力融合改进的端到端 MI-EEG 分类网络 HIT-Net

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, 一行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

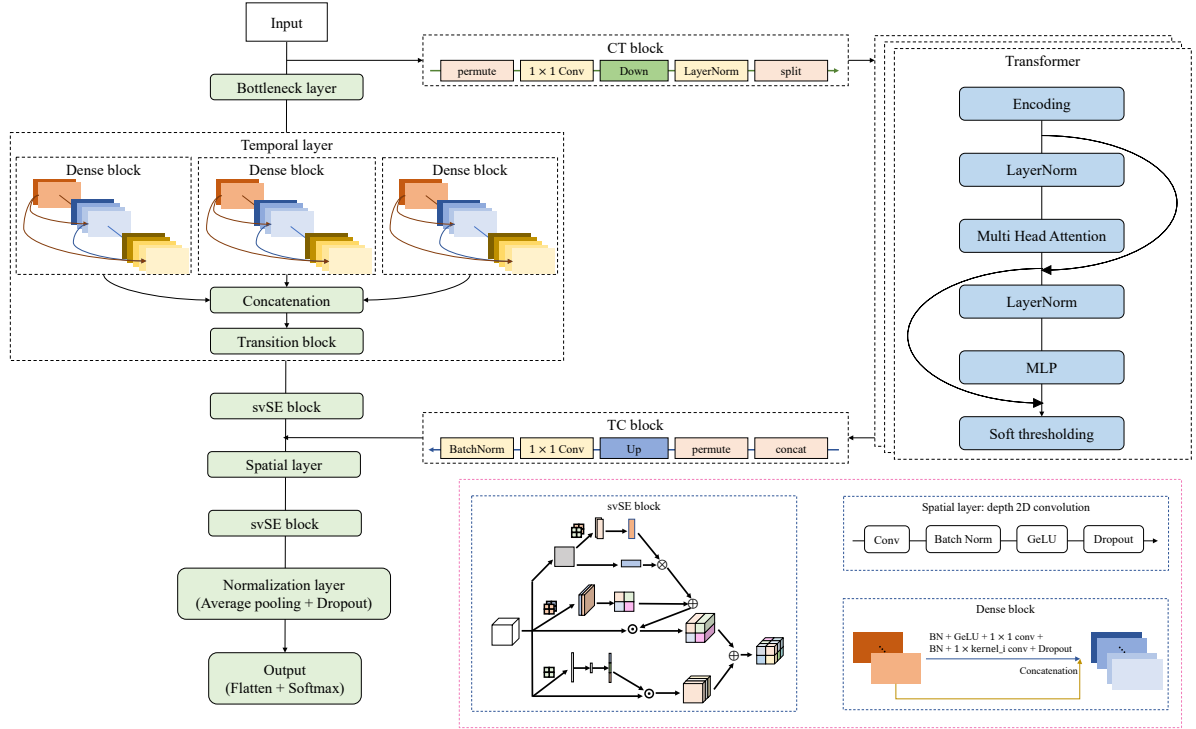


图 3.1 HIT-Net 结构

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 Inception 和混合注意力融合改进的端到端 MI-EEG 分类网络 HIT-Net (HybridAttention-Inception-Transformer-MI-EEGNet)，其结构如图 3.1 所示。HIT-Net 有两条并行分支，分别是卷积神经网络分支与 Transformer 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，同时通过深度可分离卷积等方法减小网络的参数规模；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。Transformer 分支用以获取全局上下文信息，同时采用了时空特征分离的策略，以并行分支方式进行通道上的时间长距依赖信息；通过引入软阈值函数以抑制噪声。卷积神经网络分支与 Transformer 分支通过 CT-Block 与 TC-Block 进行特征的融合，以同时利用 Transformer 提取的全局信息与卷积神经网络提取的局部细节，构建长短依赖的混合注意力机制。

通过 HIT-Net，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。

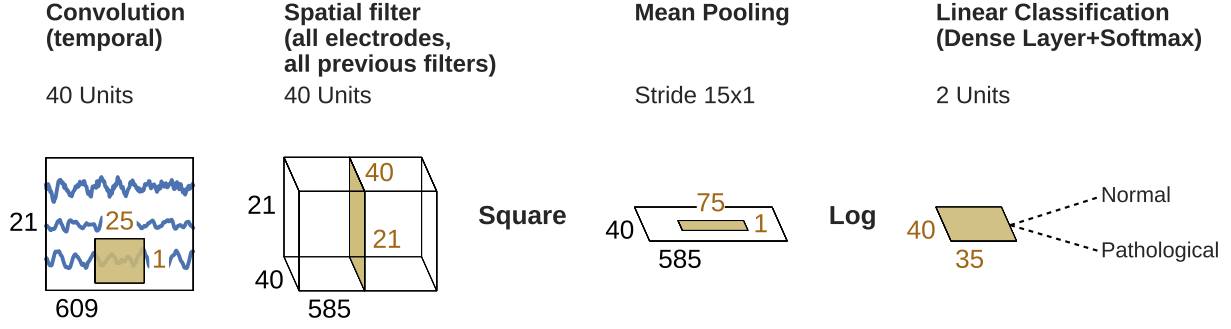


图 3.2 ShallowConvNet 结构

3.1.1 基础网络 BaseNet

ShallowConvNet^[26] 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）^[16]。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^[26]，研究发现，该特性有助于提高分类性能^[27]。实验证明，ShallowConvNet 在 MI-EEG 分类领域具有优良的性能^[28]，同时具有较少的参数量，因此论文参考 ShallowConvNet 设计 MI-EEG 分类网络。

ShallowConvNet 的结构如图 3.2 所示。ShallowConvNet 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 采取的时间卷积和空间卷积相分离的策略有效地减少了模型参数量，同时，空间卷积核能够学习其对应的时间卷积核提取的特征，这种设计隐含了对 FBCSP 算法核心思想的借鉴与实现。

由于 EEG 原始输入具有时间和通道两个维度的信息，可以被视为具有空间信息的图像数据，因此，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^[24]，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 3.3 所示，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

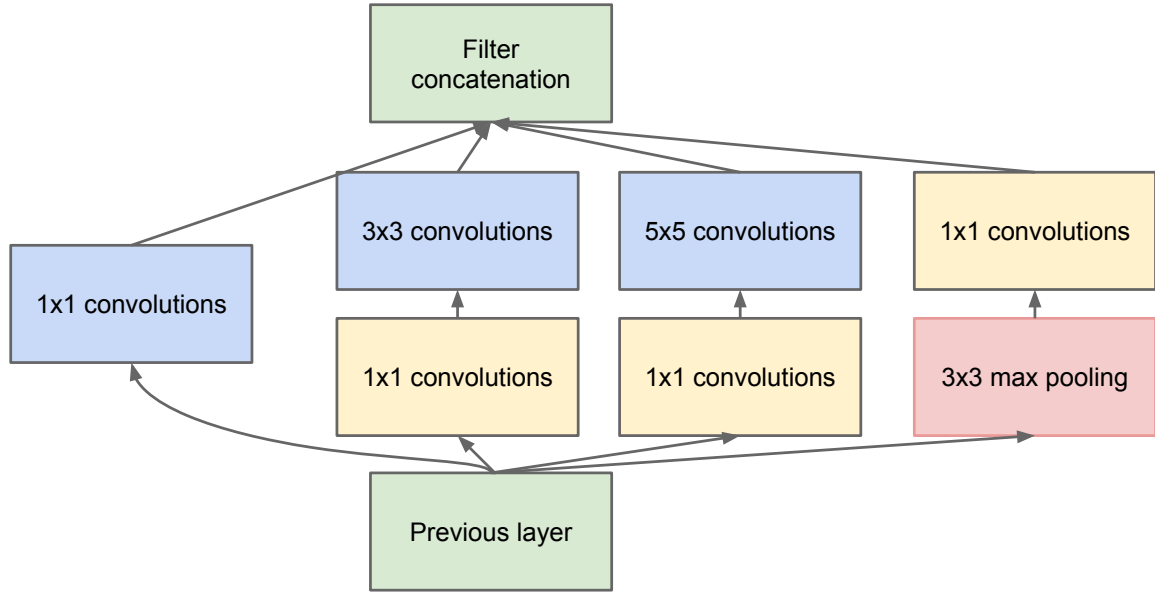


图 3.3 Inception 结构

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 1×1 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^{[29][30]}。

(2) 残差神经网络

残差神经网络 (Residual Network, ResNet)^[31] 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 3.4 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

U-Net 模型^[32] 最初是为生物学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 3.5 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

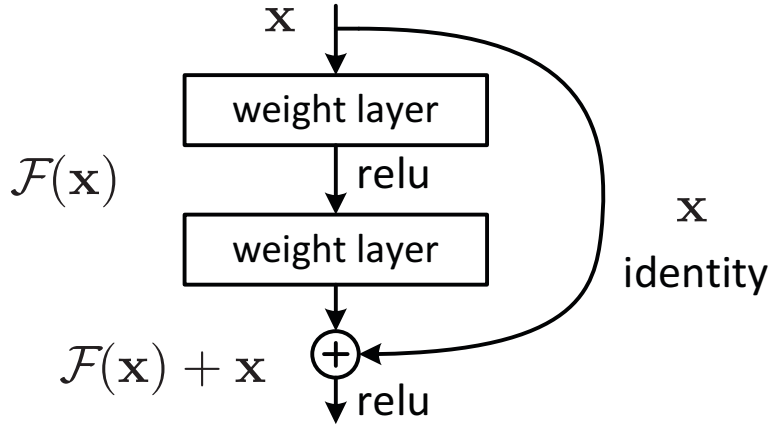


图 3.4 残差块结构

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受试者的平均表现。

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

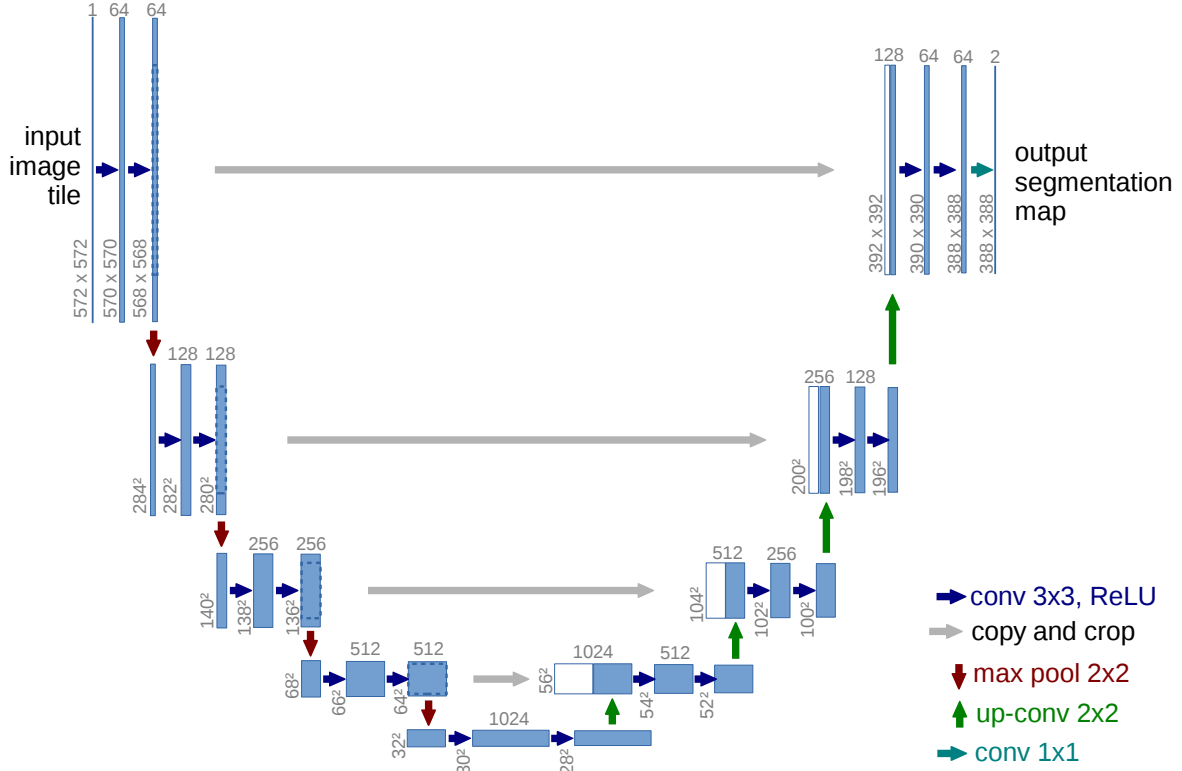


图 3.5 U-Net 结构

表 3.1 Inception、ResNet、U-Net 实验结果对比

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

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 信号的空间特征复杂度通常低于时间特征复杂度，时空信息具有不平衡性。例如，在 BCI Competition IV Dataset 2B^[33] 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。因此，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要高于空间卷积层的复杂度，这一策略的目的在于使得模型具有捕捉高维时空特征能力的同时，防止模型因复杂度过高而在小样本数据集上过早地发生过拟合现象。

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

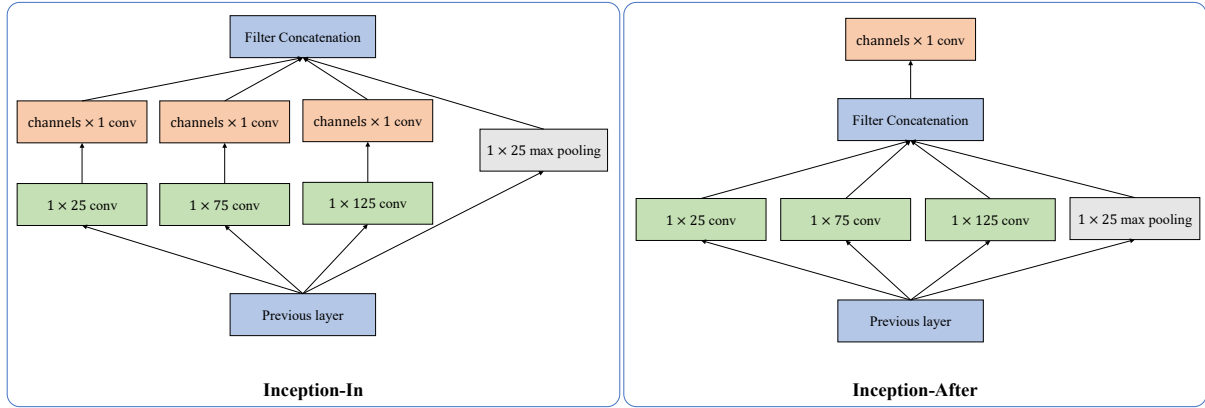


图 3.6 Inception 模块引入空间卷积层的方式

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 3.6 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。

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 3.2 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受试者的平均表现。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 3.2 Inception-In、Inception-After 实验结果对比

Models	ACC(%)	Kappa
Inception-In	63.31	0.51
Inception-After	75.35	0.70

文献^{[28][26]}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^{[31][34]}，通常用于对数据的降维和升维，由于采用了 1×1 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

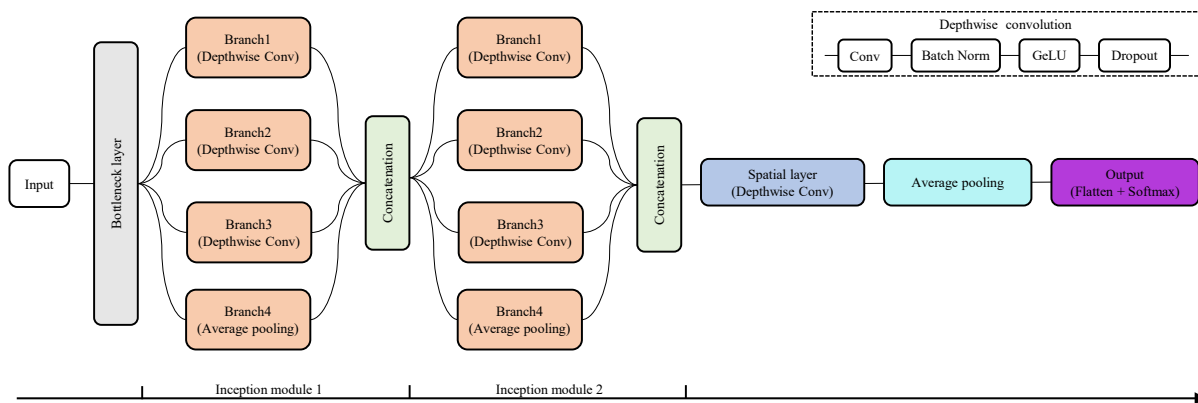


图 3.7 BaseNet 结构

进时空信息的融合。

为了加快网络训练速度，并避免小数据集下过早的过拟合，论文在引入瓶颈层的基础上，对网络的结构进行了进一步的调整：

(1) 借鉴 EEGNet^[28] 的结构，将时间卷积和空间卷积替换为参数量更少的深度可分离卷积^[28]；

(2) 在模型中引入了一系列正则化技术，包括批量归一化（Batch Normalization, BN）和随机 Dropout 等。

此外，论文将 ShallowConvNet 中的对数激活函数替换为 GELU 激活函数，前者基于 FBCSP 算法中的对数方差计算提出，后者基于高斯分布和自然梯度流提出，能够更自然地模拟神经元的概率行为，具有更好的连续性和光滑性^[35]，实验证明了 GELU 激活函数较之对数激活函数具有更好的效果。

论文将改进后得到的基础模型称为 BaseNet，其结构如图 3.7 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。

3.1.2 局部混合注意力 svSE

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^[36]，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。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）^[37]，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）^[38]，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）^[39]、空间与通道压缩与激励模块（Spatial and Channel Squeeze-and-Excitation, scSE）^[40] 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 3.8 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

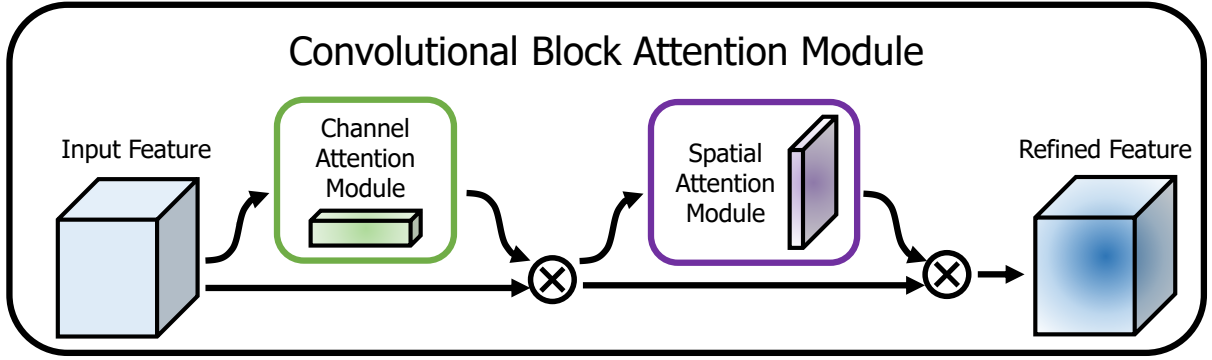


图 3.8 CBAM 结构

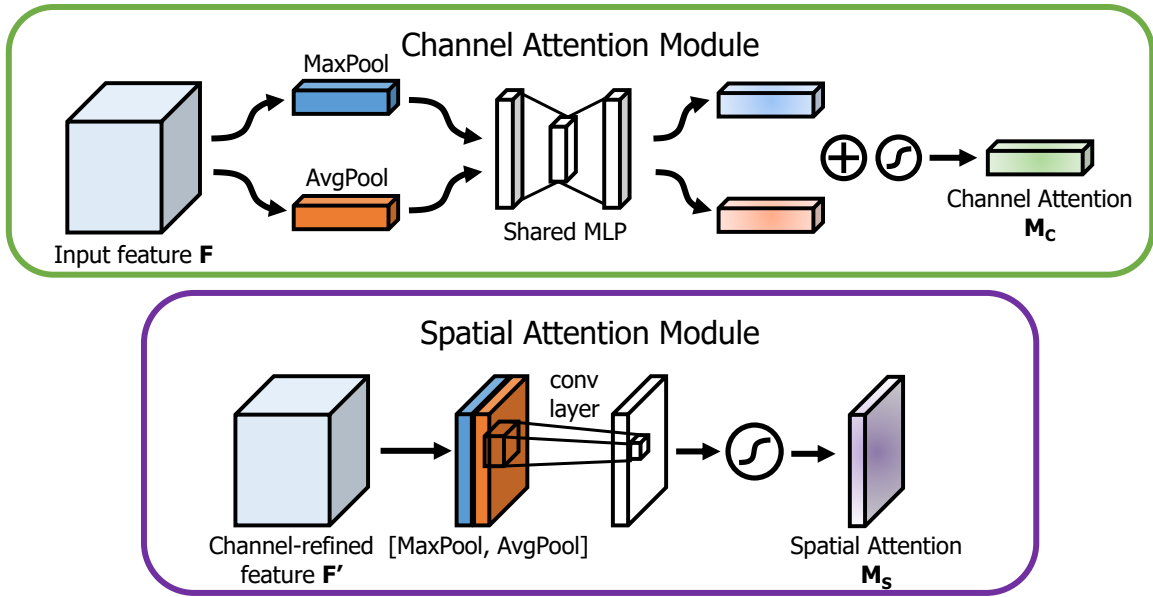


图 3.9 CBAM 模块结构

力权重，与特征图加权后得到最终结果。CBAM 的模块结构如图 3.9 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 3.10 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

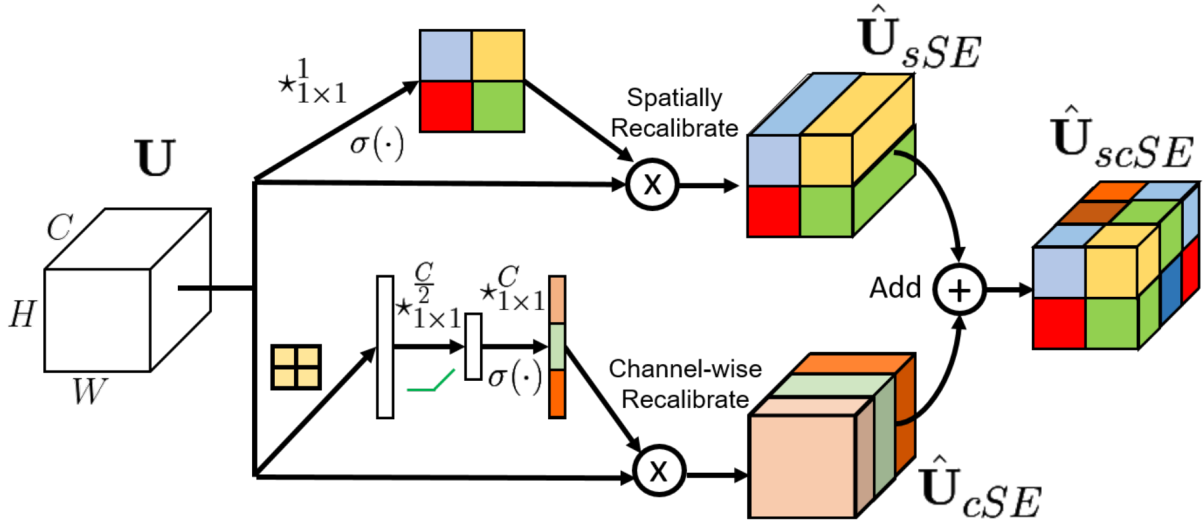


图 3.10 scSE 结构

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^[40]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，论文选择将 scSE 模块引入 BaseNet 模型中。由于 BaseNet 的特征提取过程分为时间卷积和空间卷积两个阶段，scSE 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 3.11 展示了这三种引入 scSE 模块的方式，从左至右分别是时间卷积层后引入 scSE 模块、空间卷积层后引入 scSE 模块，以及在时间卷积和空间卷积层后均引入 scSE 模块。将这三种引入方式对应的模型分别简称为 S-Temporal-BaseNet、S-Spatial-BaseNet、S-TS-BaseNet。

表 3.3 展示了 S-Temporal-BaseNet、S-Spatial-BaseNet、S-TS-BaseNet 三种模型在 BCI Competition IV Dataset 2A^[33] 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受试者的平均表现，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-BaseNet 模型的效果优于其他两种模型，与经验相符。此外，S-Temporal-BaseNet 模型的效果优于 S-Spatial-BaseNet 模型，其原因可能在于，空间卷积层沿通道维度的卷积和沿深度维度的降维使得数据损失了部分特征，进而减弱了 scSE 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 scSE 模块能够帮助模型更好地捕捉深度和空间的特征。从标准

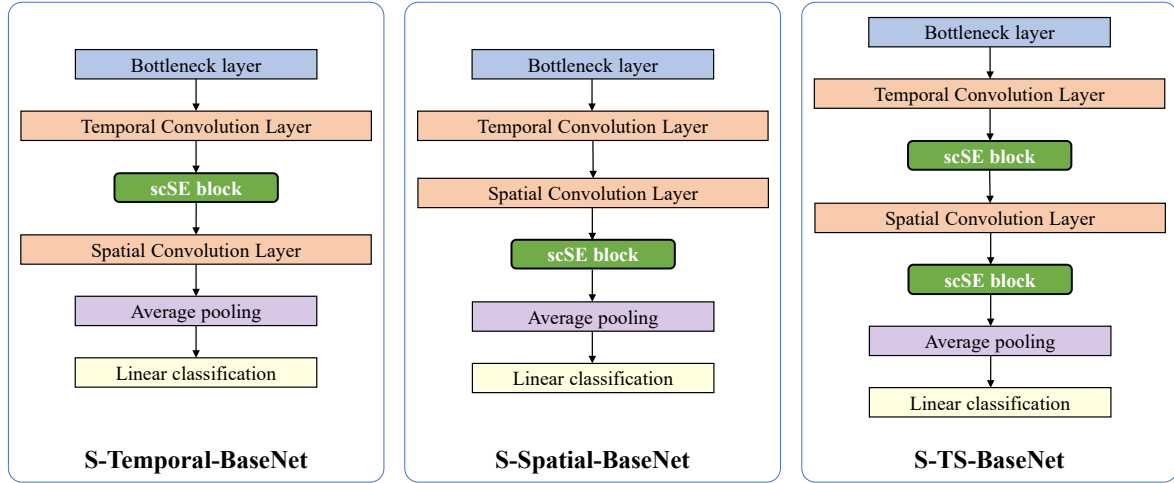


图 3.11 BaseNet 引入注意力模块的方式

差分析，S-TS-BaseNet 模型的准确率波动幅度较小，对不同受试者的 MI-EEG 分类效果相对均衡，另外两种模型在不同受试者间的分类精度则存在较为明显的差异。实验数据显示，S-TS-BaseNet 模型在增加了少量参数的情况下，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 scSE 模块的方式，将这种结构的模型称为 S-BaseNet。

表 3.3 scSE 模块引入位置对比

Models	ACC(%)	Kappa	SD	Parameters
S-Temporal-BaseNet	78.09	0.71	10.38	4702
S-Spatial-BaseNet	77.16	0.69	10.24	4357
S-TS-BaseNet	78.55	0.71	9.46	4765

为取得更好的效果，对 scSE 模块进行改进。针对 cSE 模块，采用全局最大池化取代全局平均池化操作，实验证明了这一改动使模型具有更好的效果。此外，将 scSE 模块中的 Sigmoid 激活函数替换为 Softmax 激活函数，旨在更好地利用全局信息。针对 sSE 模块，论文提出两种方式进行改进，并将两种方式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^[41]得到启发，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，随后通过深度卷积对深度维度的特征图进行聚合，在更好地表征 EEG 信号特性的同时，进一步降低了参数规模；

(2) 考虑 EEG 信号中的时空权重无关性，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。对于空间维度，首先进行深度压缩操作，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，通过 1×1 卷积对这两个特征图进行融合。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。

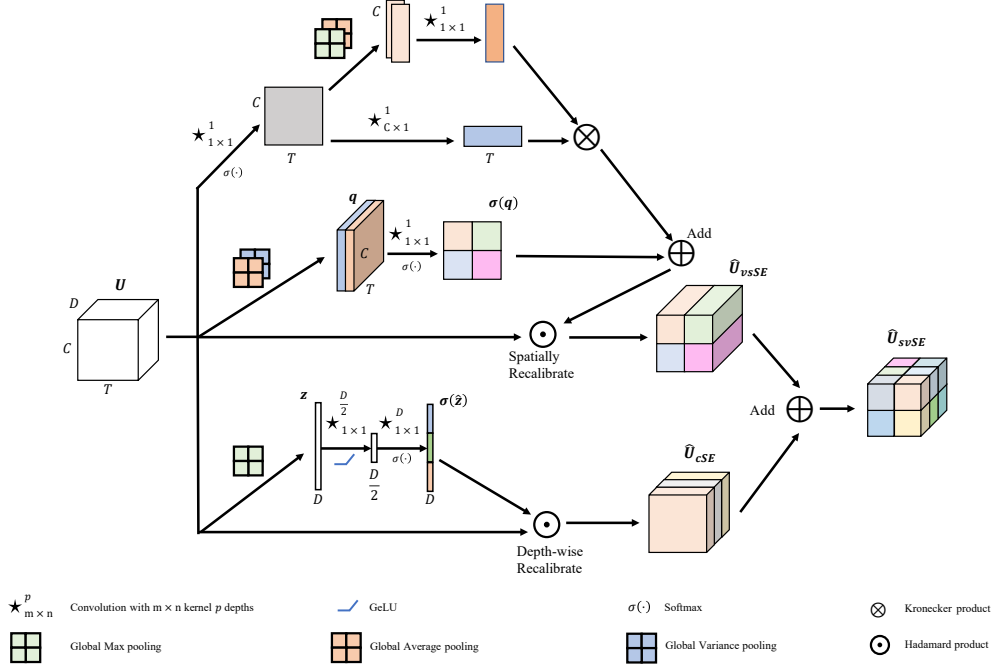


图 3.12 svSE 结构

最后，将空间权重与时序权重相乘，恢复维度。

得到论文将改良后的 scSE 模块称为 svSE (Separate Variance-Informed Spatial and Channel Squeeze-and-Excitation) 模块，其结构如图 3.12 所示。使用 svSE 模块替换 S-BaseNet 模型中的 scSE 模块。

3.1.3 基于密集连接优化 Inception 模块

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，在对 VSNet 进行改进时，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^[34] (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 ：

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

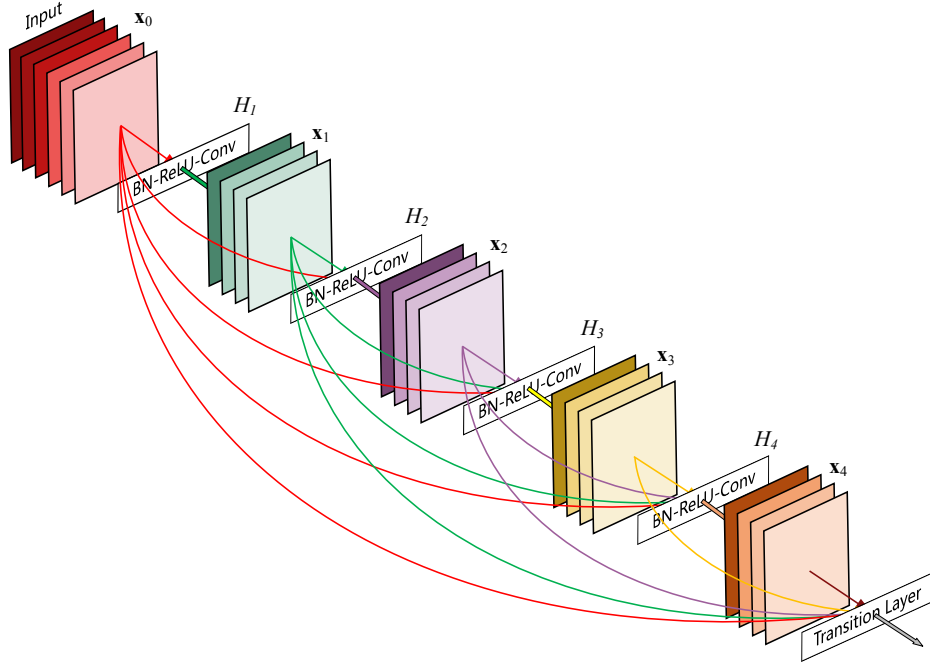


图 3.13 密集连接结构

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 3.13 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。

在 DenseNet 中, 有两种不同构造的密集连接模块, 分别是基础密集连接模块 (**Basic Dense Block**) 和带有瓶颈层的密集连接模块 (**Bottleneck Dense Block**), 它们的区别主要体现在非线性转换复合函数 $H(\cdot)$ 的设计上:

(1) 基础密集连接模块 (**Basic Dense Block**): 其非线性转换复合函数定义为: $H(\cdot) = BN + ReLU + 3 \times 3 \text{ conv}$ 。

(2) 瓶颈密集连接模块 (**Bottleneck Dense Block**): 其非线性转换复合函数更为复杂, 包括两个卷积层、两次 BN 和 ReLU 激活: $H(\cdot) = BN + ReLU + 1 \times 1 \text{ conv} + BN + ReLU + 3 \times 3 \text{ conv}$ 。加入瓶颈层的主要原因在于降低特征维度的数量, 从而提升模型计算效率。

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 Bottleneck Dense Block 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具有等价的重要性。然而, EEG 信号的时空信息具有不平衡的重要性, 通常情况下, 时间维度往往比空间维度蕴含更为丰富的信息。与此同时, 研究指出^[28], 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。

鉴于上述 EEG 信号与图像数据的特性差异, 对原始密集连接模块的卷积核进行改造。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:

$$kernel_i = sfreq \times time_unit_i, \quad time_unit_i \in \{0.1, 0.3, \dots, 0.9\} \quad (3.2)$$

其中, $sfreq$ 是 EEG 信号的采样频率, $time_unit_i$ 则为卷积核对应的时间单位, 其取值范围分布在 0.1 秒至 0.9 秒的奇数间隔内。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, 当采样周期设定为 0.5 秒时, 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 $H(\cdot)$ 定义如下:

$$\begin{aligned} H(\cdot) &= BN + GeLU + 1 \times 1 \text{ conv} + BN + 1 \times kernel_i \text{ conv} + Dropout \\ kernel_i &= sfreq \times time_unit_i, \quad time_unit_i \in \{0.1, 0.3, \dots, 0.9\} \end{aligned} \quad (3.3)$$

相较于原始模块, 论文在前文研究的基础上做出了以下改进, 分别为: 根据 EEG 信号的特性调整了卷积核大小; 选用性能更优的激活函数 GeLU 替代 ReLU; 引入了 Dropout 层, 以进一步缓解过拟合现象。

经过以上改进, 新构建模型的结构如图 3.14 所示, 将其命名为 DIS-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。需要说明的是, Dense-Inception module 的数量, 以及 Dense-Inception module 的分支数量、Dense block 的数量, 都是可调的超参数。

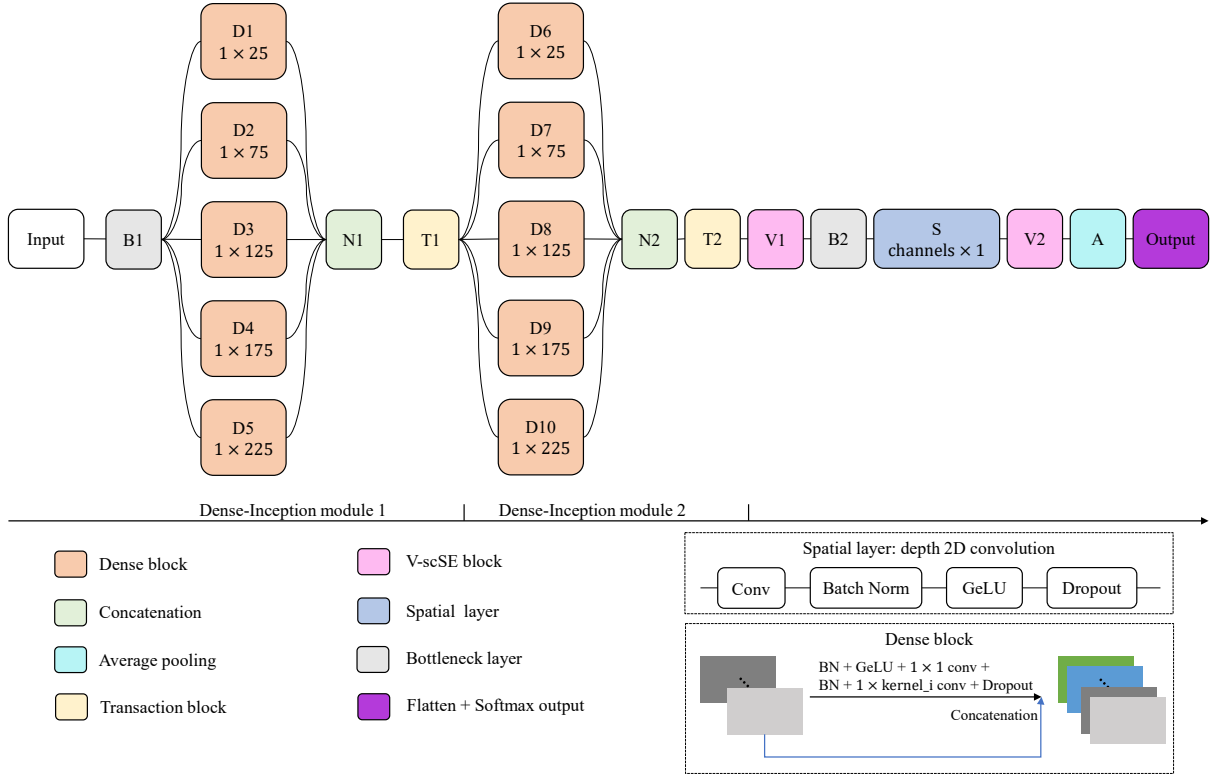


图 3.14 DI-VSNet 结构

3.1.4 基于 Transformer 构建并行分支

卷积神经网络中，尽管卷积层因其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。而 Transformer 模型^[42]因其自注意力机制（Self-Attention）可以在不考虑输入位置顺序的情况下，灵活地捕捉并编码任意两点之间的相互依赖关系，从而能够更有效地理解和处理长距离依赖，但与此同时，其可能无法充分提取局部时空特征。因此，论文选择将两者进行结合，用于 MI-EEG 分类任务。

原始的 Transformer 模型是编码器-解码器（Encoder-Decoder）结构，编码器模块负责接收输入序列，并通过自注意力机制捕捉并整合整个序列中的语义和上下文信息；解码器则依据这些编码后的信息逐步生成对应的输出序列。在 MI-EEG 分类任务中，不涉及目标序列的生成，因此只需要使用 Transformer 的编码器，其结构如图 3.15 所示。编码器由以下几部分组成：

- (1) 输入嵌入与位置编码层（Input Embedding & Positional Encoding）：将输入序列转化为稠密向量表示，通过位置编码为输入添加位置信息。
- (2) 多头注意力（Multi-Head Attention）：使用多个并行运行的自注意力（Self-Attention）子层（头），提取不同尺度的长距离依赖，对输入进行注意力加权。
- (3) 残差连接与层归一化（Add & Norm）：在每个子层之后应用残差连接与层归一

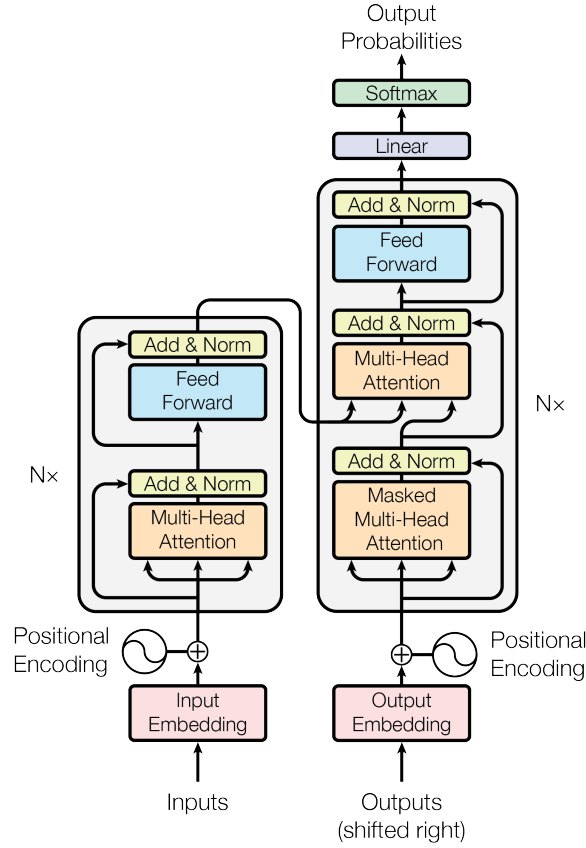


图 3.15 Transformer Encoder 结构

化，提高模型训练效率和稳定性，避免网络退化。

(4) 前馈神经网络 (Feed Forward): 全连接神经网络，对多头注意力的输出进行非线性转换。

对于 MI-EEG 信号而言，空间维度的特征反映了电极位置特异性，时间维度的特征则体现了与事件相关去同步 (ERD) 和事件相关同步 (ERS) 等生理现象相关的时序变化。因此，重要的长依赖关系主要存在于在时间维度，即时序特征。因此，论文使用的 Transformer 模块沿袭经典 Transformer 结构。

为了更好地获取时序长依赖，将一个输入样本 $X \in \mathbb{R}^{C \times T}$ 中的第 i 个时间序列视为一个 Token $x_i \in \mathbb{R}^T$ ，则一个样本需要 C 个 Transformer 处理，表达为：

$$f(X) = \sum_i^C f_i(x_i) \quad (3.4)$$

。其中， C 代表数据的通道数， T 代表时间序列长度。考虑单个 Transformer，获得其对应的 x 的词嵌入表示 $E \in \mathbb{R}^{T \times d_{model}}$ ，其中， d_{model} 为词嵌入维度，随后对 Token E 的处理过程如下：

首先，为了编码序列的位置信息，论文对 E 进行可学习位置编码，即通过一个可学习的参数矩阵表征 Token 的位置信息，将其记为 PE ，通过相加的方式与 Token 相融合，获得最终的 Token Y : $Y = E + PE$ 。

随后，对于 Y ，通过含多个自注意力机制的多头注意力层（MHA）进行处理。对于第 i 个头（head），首先通过三组线性变换，获得 Query Q_i ，Key K_i ，Value V_i ，如公式 3.5 所示：

$$Q = W_i^Q Y, K = W_i^K Y, V = W_i^V Y \quad (3.5)$$

其中， $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d_{model} \times d_k}$ ， $d_k = d_{model} \div h$ ， h 是 head 的数量。通过公式 3.6 得到当前 head 的输出，记为 A_i ：

$$A_i = softmax(\frac{Q_i K_i^T}{\sqrt{d_k}} V_i) \quad (3.6)$$

随后，根据公式 3.7 对所有 head 的结果进行合并：

$$MHA(Y) = Concat(A_1, ..., A_h) W^O \quad (3.7)$$

其中， $W^O \in \mathbb{R}^{T \times d_{model}}$ ，输出经过残差连接和层归一化：

$$Z = LayerNorm(Y + MHA(Y)) \quad (3.8)$$

MHA 的输出会在全连接前馈神经网络（FFN）中进行转换，并同样经过残差连接和层归一化，由公式 3.9 得到最终的输出。

$$\hat{Z} = LayerNorm(Z + FFN(Z)) \quad (3.9)$$

论文从深度残差收缩网络（Deep Residual Shrinkage Network, DRSN）^[43] 中得到启发，对于具有低信噪比的 EEG 数据，将软阈值函数（Soft Thresholding）引入 Transformer 中。软阈值函数是信号处理中常用的去噪方式^[44]，其公式如下：

$$y = \begin{cases} x - \tau, & x > \tau \\ 0, & -\tau \leq x \leq \tau \\ x + \tau, & x < -\tau \end{cases} \quad (3.10)$$

其中， x 为输入， y 为输出， τ 为阈值。软阈值函数将绝对值小于阈值的输入转为零值，同时对绝对值接近阈值的输入有抑制作用，从而能够减少噪声对特征的影响。论文在 Transformer 模块之后添加软阈值函数，通过公式 3.11 计算阈值 τ ：

$$\tau = \alpha \times Average(\hat{Z}), \quad \alpha \in (0, 1) \quad (3.11)$$

其中， α 是一个可供学习的参数。通过这种方式， τ 是一个不太大的正数，从而避免输出全部为零的情况。

不同于 Song 等人将 Transformer 嵌入卷积神经网络之后的方式^[45]，论文选择以并行分支的形式引入 Transformer 模块。这是因为直接在卷积层之后加入 Transformer 可能

导致模型仅能捕捉到深层卷积特征图的依赖关系，而非充分利用全局信息。并行分支的设计旨在全面利用 Transformer 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 Transformer 模块引入 DIS-Net 后获得的新模型命名为 HIT-Net，其结构如图 3.1 所示。为了综合 CNN 和 Transformer 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，因此，有必要对数据进行一定的处理，使得两个分支的数据形状相匹配。论文设计了 CT-Block 和 TC-Block 用于并行分支数据的交互融合。在 CT-Block 中，数据通过时间维度的池化和通道维度的卷积进行下采样，将一个样本 $X \in \mathbb{R}^{C \times T}$ 沿通道维度压缩至 $X' \in \mathbb{R}^{C' \times T}$ ，其中， C 代表通道数（电极数）， T 代表时间序列长度， C' 是一个大于等于 1 且小于等于 C 的整数，通常设定为 $\left\lfloor \frac{C}{N} \right\rfloor$ ， N 为输出类别数量，这样设计的目的是减少计算开销和保留一定的空间特征信息。在 TC-Block 中，则通过反卷积对 Transformer 模块的输出进行重建，将时间维度和通道维度上采样至与 CNN 分支相同的大小。

3.2 基于 HIT-Net 轻量化的端到端 MI-EEG 分类网络 HIT-lightNet

考虑到 BCI 系统对即时响应具有较高的要求，虽然 HIT-Net 已运用了瓶颈层、深度可分离卷积等手段有效地缩减了模型参数量，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。HIT-Net 模型参数的主要来源在于 Transformer 模块和密集连接模块。针对 Transformer 模块的轻量化策略包括但不限于减少原始层级数目、量化参数类型及运用知识蒸馏技术。然而，面对小样本规模的 EEG 信号数据集，由于缺乏足够的大数据训练支撑，论文主要采取了减少原始层数这一轻量化途径。具体来说，对于 CT-Block 设计了更为高效的下采样方案，将原本由多个 Transformer 单元组成的 Transformer 分支精简为仅包含一个 Transformer 单元。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。

3.2.1 经典轻量级网络结构

MobileNet 由 Google 团队提出，其核心思想在于引入了深度可分离卷积（Depthwise Separable Convolution）^[46]，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 1×1 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^[47]，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

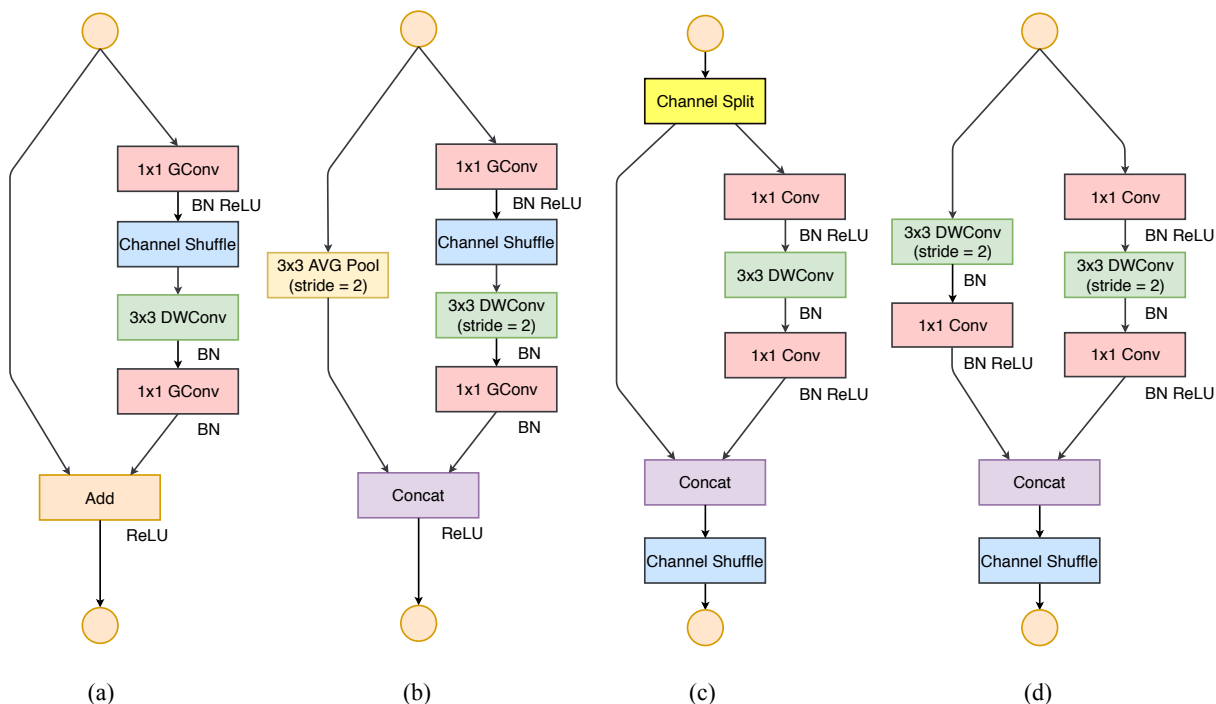


图 3.16 ShuffleNet 结构

术^[48]，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的优化策略。

ShuffleNet^[49] 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每组内的通道数，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^[50] 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 1×1 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 3.16 所示，图 3.16 (a) 是 ShuffleNet 的基本单元，图 3.16 (b) 在 ShuffleNet 基本单元的基础上增加了下采样，图 3.16 (c) 是 ShuffleNetV2 的基本单元，图 3.16 (d) 在 ShuffleNetV2 基本单元的基础上增加了下采样。

GhostNet 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）^[51]，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，

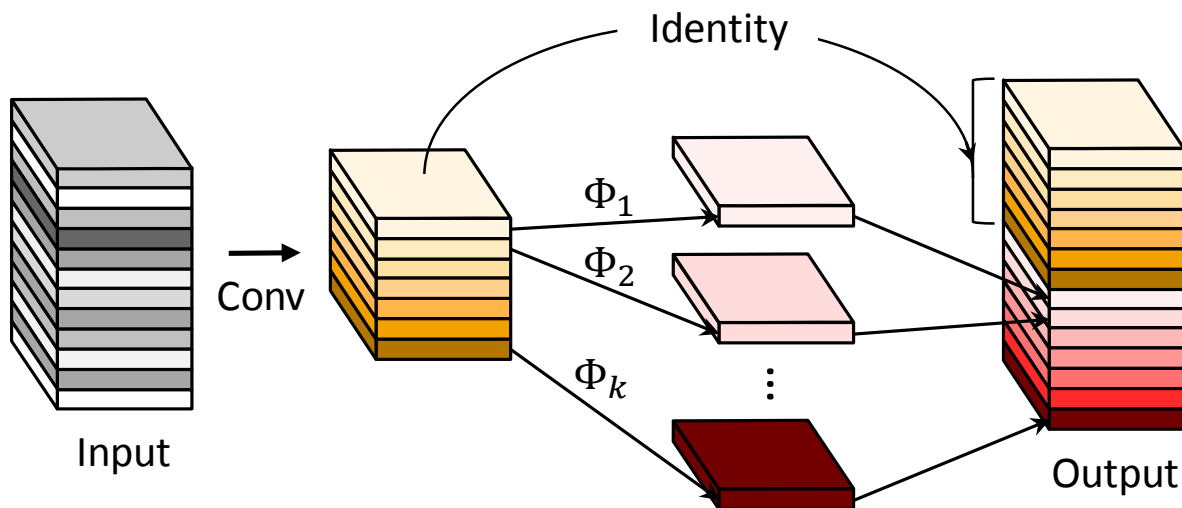


图 3.17 Ghost 模块结构

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 3.17 所示，其中，Identity 操作由标准卷积生成的特征图， Φ 为廉价线性变换。

3.2.2 ShuffleNet 结合 GhostNet 的轻量化卷积模块

3.3 基于 MEKT 算法改进的 MI-EEG 迁移学习算法 SV-MEKT

4 实验结果分析与模型评估

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] JORGENSEN 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.
- [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 selection 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] 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.

- [25] RUBNER Y, TOMASI C, GUIBAS L J. The earth mover's distance as a metric for image retrieval[J]. International journal of computer vision, 2000, 40: 99-121.
- [26] SCHIRRMESTER 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] TANGERMANN M, MÜLLER K R, AERTSEN A, et al. Review of the bci competition iv[J]. Frontiers in neuroscience, 2012, 6: 21084.
- [34] 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.
- [35] HENDRYCKS D, GIMPEL K. Gaussian error linear units (gelus)[J]. arXiv preprint arXiv:1606.08415, 2016.
- [36] 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.
- [37] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Confer-

- ence on Computer Vision and Pattern Recognition. 2018: 7132-7141.
- [38] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks [J]. Advances in neural information processing systems, 2015, 28.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [43] ZHAO M, ZHONG S, FU X, et al. Deep residual shrinkage networks for fault diagnosis [J]. IEEE Transactions on Industrial Informatics, 2019, 16(7): 4681-4690.
- [44] WRIGHT S J, NOWAK R D, FIGUEIREDO M A. Sparse reconstruction by separable approximation[J]. IEEE Transactions on signal processing, 2009, 57(7): 2479-2493.
- [45] 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.
- [46] 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.
- [47] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.
- [48] 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.
- [49] 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.
- [50] 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.
- [51] 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.

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。