

DOI:10.19651/j.cnki.emt.2416701

基于多模态脑影像的阿尔茨海默病分类方法^{*}

钟海杰 吴晓红 卿磊波 陈洪刚 何小海

(四川大学电子信息学院 成都 610065)

摘要: 阿尔茨海默病(AD)是一种神经退行性疾病,是痴呆的主要原因,准确的诊断出AD具有重要意义。利用来自大脑的氟脱氧葡萄糖正电子发射断层成像(FDG-PET)和结构磁共振成像(sMRI)的多模态数据可以从不同角度提供关于病变的全面信息,提高诊断的准确性。然而影像数据具有大量冗余信息,并且不同模态的特征也有显著差异。使用传统的卷积神经网络和简单的特征拼接方法无法有效利用多模态数据的互补信息,从而限制AD的诊断性能。针对此问题,提出了一种结合sMRI与FDG-PET的多模态影像AD分类方法。该方法通过嵌入坐标注意力机制和空间-通道重构卷积来捕捉影像中的特异性区域并限制冗余信息;同时设计了一种并行交互网络,不仅可以增强每个模态自身的特征,还可以根据其他模态的特征进行自适应调整,从而实现模态之间的有效交互。在ADNI数据集上评估了提出网络的分类性能,准确率、敏感性和特异性分别达到了93.66%、91.67%和95.41%,实验结果表明,与现有的AD分类网络相比,本文提出的网络具有更优异的性能。

关键词: 阿尔茨海默病;多模态影像;深度学习;磁共振成像;正电子发射断层成像

中图分类号: TN911;TP391.4 **文献标识码:** A **国家标准学科分类代码:** 510.4030

Classification of Alzheimer's disease based on multimodal brain images

Zhong Haijie Wu Xiaohong Qing Linbo Cheng Honggang He Xiaohai

(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: Alzheimer's disease (AD) is a neurodegenerative disease that is a significant contributor to dementia. Accurate diagnosis of Alzheimer's disease (AD) is of great significance. The integration of multimodal data from Fluorodeoxyglucose positron emission tomography (FDG-PET) and structural magnetic resonance imaging (sMRI) of the brain provides a comprehensive information of lesions from multiple perspectives and enhancing diagnostic accuracy. However, the image data is highly redundant, and the features of the various modes are also significantly disparate. Traditional convolutional neural networks and simple feature concatenation methods are unable to effectively utilize the complementary information of multi-modal data, consequently, this limits the diagnostic performance of AD. To solve this problem, we propose a multimodal image AD classification network combining sMRI and FDG-PET. The network incorporates coordinate attention mechanisms and spatial-channel reconstruction convolution to capture specific regions in images and limit redundant information. A parallel interaction network is also designed, which not only enhances each modality's own features, but also adaptively adjusts itself according to the features of other modalities, thus realizing effective interaction between modalities. The classification performance of the proposed network is evaluated on the ADNI dataset, and the accuracy, sensitivity, and specificity reach 93.66%, 91.67% and 95.41%, respectively, the experimental results show that the proposed network in this paper has a superior performance compared to the existing AD classification networks.

Keywords: Alzheimer's disease; multimodal imaging; deep learning; magnetic resonance imaging; positron emission tomography

0 引言

阿尔茨海默病(Alzheimer's disease, AD)是一种进行

性的神经退行性疾病,是老年人群罹患的主要疾病之一。随着疾病的发展,患者通常会会出现记忆衰退和认知功能的恶化,还可能

收稿日期:2024-08-20

* 基金项目:四川省重点研发计划(2023YFS0195)项目资助

失^[1]。研究表明,全球约有5 000万人患有痴呆,预计到2050年患者人数将增加两倍,这将极大地增加疾病负担和医疗保健成本^[2-3]。而在AD的早期阶段进行一定的药物、锻炼和记忆训练可以缓解疾病进展并改善患者的生活质量^[4]。临床上一般依靠医生结合诊断数据进行综合分析来诊断AD,这要求医生具有丰富的临床经验,同时医生的主观判断也有可能存在误诊。在医疗资源紧张的今天,迫切需要一种客观的诊断方法辅助医生对AD患者的诊疗提供支持。

目前,神经影像检查是AD诊断的常用手段之一。其中,最常用的成像技术是结构磁共振成像(structural magnetic resonance imaging, sMRI)与正电子发射断层成像(positron emission computed tomography, PET)。sMRI可以揭示AD病情发展中的脑部结构性变化,清晰直观的展示脑组织的萎缩和病变^[5-6]。而PET可以显示大脑代谢和功能状态的情况,如18F-氟脱氧葡萄糖正电子发射断层成像(fluorodeoxyglucose positron emission computed tomography, FDG-PET)可以在脑萎缩之前检测到与AD相关的糖代谢水平变化^[7-8]。两种成像技术的结合,可以提供一种有效的AD诊断方法。

近年来,卷积神经网络(convolutional neural network, CNN)在计算机辅助诊断^[9]、医学图像分割^[10-12]等领域的应用已经取得了成熟的进展。CNN网络是指通过堆叠具有卷积计算的卷积层进行特征提取的一类网络的总称,如ResNet、VGG、DenseNet等经典网络都属于CNN网络。与传统的机器学习方法相比,基于CNN的方法可以实现更好的特征提取,也消除了复杂且耗时的人工特征提取过程。AD的病变主要发生在大脑区域,而3D脑影像包含的信息极其复杂,并且有价值的病变信息往往非常细微,传统的CNN网络难以有效地提取脑影像特征。因此需要通过改进网络或增强数据的方式来加强特征提取能力。例如,邓爽等^[13]利用改进的VGG网络,通过弱监督注意力学习、数据增强和双线性注意力池化,实现细粒度分类。然而AD是一种复杂的疾病,涉及到大脑的结构和功能变化,仅依靠单模态数据无法提供足够的信息。

现今的AD分类研究仍然以单模态MRI影像为主,但越来越多的研究表明,多模态数据可以从多个角度提供互补信息,有效提高分类性能^[14-15]。张响泉等^[16]分别提取sMRI影像特征和临床数据非影像特征,再将两种特征进行级联后进行分类。Narazani等^[17]使用多模态3D ResNet网络,基于sMRI和FDG-PET影像数据实现AD评估,并通过实验表明FDG-PET数据比sMRI数据表现更好。Lin等^[18]设计了一种基于极限学习机的分级方法,该方法将患者的MRI、PET、脑脊液(cerebrospinal fluid, CSF)和遗传数据作为输入,每种数据分别获得分类分数,然后将这些分数连接起来用于AD分类。这些研究在融合多模态数据特征时采用简单的级联和加权求和,这可能导致模型被来自

某些模态的信息主导,而忽略来自其他模态的信息。针对这个问题,Zhang等^[19]提出了一种多模态AD诊断框架,整合PET、MRI、CSF的信息,提高了诊断准确率。Aviles-Rivero等^[20]引入双重嵌入策略来构建一种新型的半监督学习框架,该框架可以整合多模态影像数据和非影像数据的高阶关系。Cheng等^[21]提出了一种基于去相关约束的AD诊断模型,从FDG-PET和sMRI数据中获取不同的病变特征,使用去相关约束函数降低不同模态之间的特征相似性。虽然这些方法使模型可以综合多种模态数据的信息来提高分类性能,但多模态数据之间的异质性和冗余噪声也会导致模型难以区分和有效利用多模态数据之间的互补信息,从而影响融合特征的表达能力^[22-23]。

现阶段,关注多模态数据互补融合的AD分类研究较少。其中,Zhang等^[24]提出了一种基于金字塔注意机制的生成对抗网络模型PA-GAN,通过在生成对抗网络中嵌入金字塔注意力来生成PET影像,然后将生成的PET影像和真实的MRI影像直接进行映射得到融合影像,再使用CNN网络对融合影像进行特征分类。该方法实现了多模态影像融合,但生成的PET影像实质上是对MRI影像信息的深度挖掘,信息的真实性和丰富度无法与真实的PET影像相比。Liu等^[25]提出了一种基于分层注意力的多模态融合模型HAMMF,使用ResNet-18来提取各模态影像的浅层特征,然后引入通道和空间注意力提取更细粒度的病理特征,并在通道和空间注意力中嵌入Transformer的多头注意力机制来捕捉不同模态之间的相关性。这种方法可以有有效的捕捉脑影像中的细微病变,但因为特征提取的同时也进行特征融合,会导致部分病变特征丢失,且影响融合效果。

针对上述问题,本文提出了一种新的端到端的多模态影像AD分类方法。针对脑影像的特点,本文方法设计了一个基于3D CNN架构的特征提取子网络,通过堆叠卷积计算层、激活函数和归一化层进行特征提取,并在此基础上嵌入坐标注意机制^[26](coordinate attention, CA),加强网络对病变区域的捕捉能力,同时引入空间-通道重构卷积^[27](spatial and channel reconstruction convolution, SCConv),有效减少特征提取过程中冗余信息的干扰。为应对多模态数据的异质性,本文采用一个跨模态并行交互子网络以实现两种影像特征之间的有效融合,从而充分利用它们的互补信息以实现AD分类。

1 结合sMRI和FDG-PET的多模态影像分类网络

1.1 网络总体结构

本文提出了一种新的端到端的多模态影像AD分类网络。网络总体框架结构如图1所示。

首先,使用特征提取子网络分别从sMRI影像和FDG-PET影像中提取出病理特征 F_m 和 F_p 之后,将两种特征输

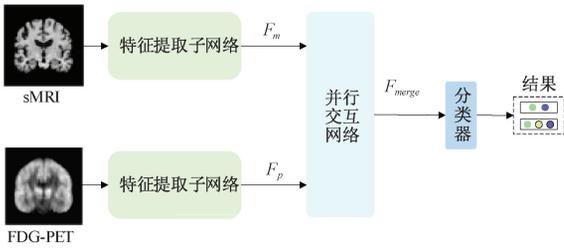


图 1 网络总体框架图

Fig. 1 Overall network framework diagram

入并行交互网络进行交互融合得到融合特征 F_{merge} , 最终将融合特征输入分类器实现 AD 分类。提出的方法可以由式(1)表示。

$$P(Y | MRI, PET) = Classifier(F_{pi}(F_{ex}(MRI, PET))) \quad (1)$$

其中, P 表示分类结果, $Classifier$ 表示分类器, F_{pi} 表示并行交互子网络, F_{ex} 表示特征提取子网络。

1.2 影像特征提取子网络

本文的特征提取子网络基于 3D CNN 架构, 通过堆叠卷积层进行特征提取, 具体结构与 ResNet-18 相似, 都是由输入卷积层、四层残差结构、平均池化层组成。本文每个残差结构包含两个残差块, 其中每个残差块又由两个卷积层和跳跃连接构成。残差结构可以有效的解决网络训练过程中的梯度消失和梯度爆炸等问题。子网络框架如图 2 所示。

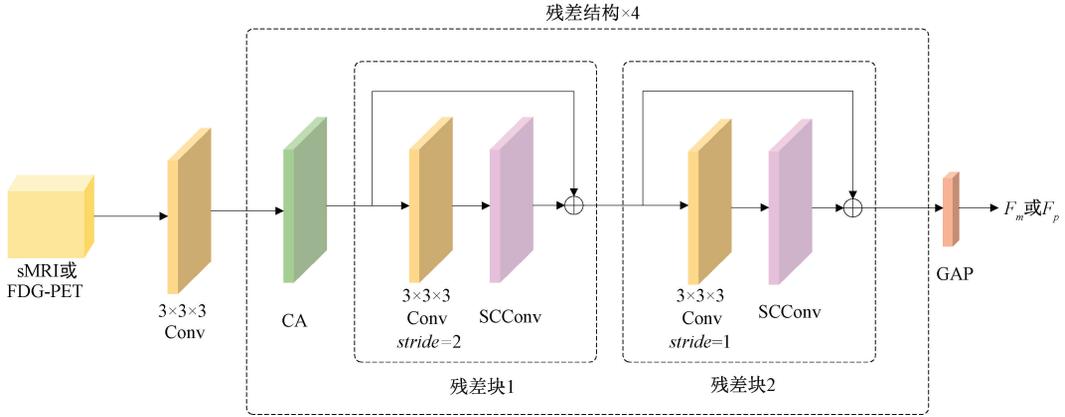


图 2 基于 3D CNN 架构的特征提取子网络

Fig. 2 Feature extraction subnetwork based on 3D CNN architecture

特别的, 本文的特征提取子网络针对 AD 分类任务进行了改进。首先, 与常规图像不同, 脑影像的特征提取更关注于大脑的细微变化, 为了避免丢失有效信息, 本文的输入卷积层应用 $3 \times 3 \times 3$ 卷积核替换 $7 \times 7 \times 7$ 卷积核进行浅层特征提取。卷积层包括卷积计算层、BatchNorm 层和 ReLU 激活层。然后, 考虑到脑影像的复杂性, 本文在残差结构中嵌入了坐标注意机制与空间-通道重构卷积。改进后的残差结构通过嵌入坐标注意机制加强网络对于病变区域的捕捉能力, 同时使用空间-通道重构卷积降低特征提取过程中的冗余信息。网络通过堆叠四层残差结构以实现深层特征提取, 四层残差结构的通道数分别为 32、64、128、256。

为了更好的表示每层残差结构的计算过程, 这里假设上一个卷积层的输出特征图为 $f_c \in \mathbb{R}^{B \times C \times D \times H \times W}$ (其中 B 、 C 、 D 、 H 、 W 分别表示训练批次大小、通道数、特征图深度、高度和宽度)。首先使用坐标注意机制对特征图进行特征加权, 为了适应本文的 3D 影像数据, 本文将坐标注意机制从 2D 结构改进为 3D 结构。改进后的具体结构如图 3 所示。利用 3 个一维的池化操作将输入特征分解为沿 3 个方向的位置感知特征 $f_d \in \mathbb{R}^{B \times C \times D \times 1 \times 1}$ 、 $f_h \in \mathbb{R}^{B \times C \times 1 \times H \times 1}$ 和 $f_w \in \mathbb{R}^{B \times C \times 1 \times 1 \times W}$, 这 3 个位置感知特征都嵌入了特定的方

向信息, 随后分别编码为 3 个注意力图, 每个注意力图都沿一个空间方向捕获输入特征的长距离依赖关系。最后通过乘法将 3 个注意力图应用到输入特征上得到加权后的特征 $f_c' \in \mathbb{R}^{B \times C \times D \times H \times W}$ 。

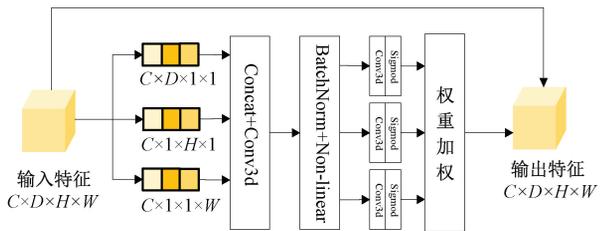


图 3 坐标注意机制

Fig. 3 Coordinate attention

如图 2 所示, 经注意力加权后, 进一步将特征输入级联的两个残差块中。两个残差块具有相同的结构, 不同的是只有第一个残差块会进行下采样操作。对于第一个残差块, 输入特征先由一个卷积核为 $3 \times 3 \times 3$, stride 为 2 的卷积层下采样为 $f_c' \in \mathbb{R}^{B \times C \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}}$, 然后进行空间-通道重构卷积, 如图 4 所示, 展示了空间-通道重构卷积的架构图。空间-通道重构卷积 (SCConv) 是一种轻量级卷积模

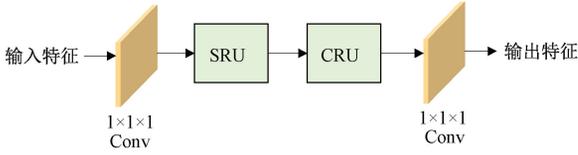


图4 空间-通道重构卷积

Fig. 4 Spatial and channel reconstruction convolution

块,它由两个核心单元组成:空间重建单元(spatial reconstruction unit, SRU)和通道重建单元(channel reconstruction unit, CRU)。SRU采用分离重构的策略来抑制空间冗余。它将整个特征图分解为包含不同有效信息量的两个特征,之后再两个特征进行重构,剔除冗余信息。CRU采用分割、变换、融合的策略来减少通道冗余。它允许网络更有效的组合来自独一无二的通道的信息,同时通过轻量的卷积运算和特征重用方案处理冗余信息。在大量的实验中,将SCConv嵌入到各种最先进的图像分类和目标检测方法中,都展现了卓越的性能。

为了减少计算量同时实现跨通道信息交互,本文先使用 $1 \times 1 \times 1$ 卷积压缩通道数,得到特征图 $f_r \in \mathbb{R}^{B \times C' \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}}$ (其中, C' 为压缩后的通道数),随后再应用空

间-通道重构卷积进行冗余特征压缩。之后,再使用 $1 \times 1 \times 1$ 卷积将重构后的特征升维为 $f_r \in \mathbb{R}^{B \times C \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}}$ 。第2个残差块具有相似的流程。

输入的原始影像数据在经过浅层特征提取和四层残差结构的深层特征提取后,再经过全局平均池化(global average pooling, GAP)得到最终的病理特征 F_m 和 F_p 。

1.3 并行交互网络

对于AD分类任务,不同模式的数据可以提供互补信息,但如何有效地整合这些互补信息是如今研究的重点。代理注意力^[28]是一种新的自注意力机制,通过引入代理向量的方式聚合深层特征,比广泛采用的Transformer自注意力机制更有效,同时保留了全局上下文建模的能力。近年来,交叉注意力的思路在图像和文本之间的跨模态任务中表现出了优异的性能。受此启发,本文将交叉注意力的思路应用到代理注意力上,设计了一个并行交互网络,使用一种模式的特征来引导另一种模式特征进行代理注意力计算,从而实现了两种成像特征的跨模态交互以及它们各自特征的自适应调整。并行交互网络的具体框架如图5所示。

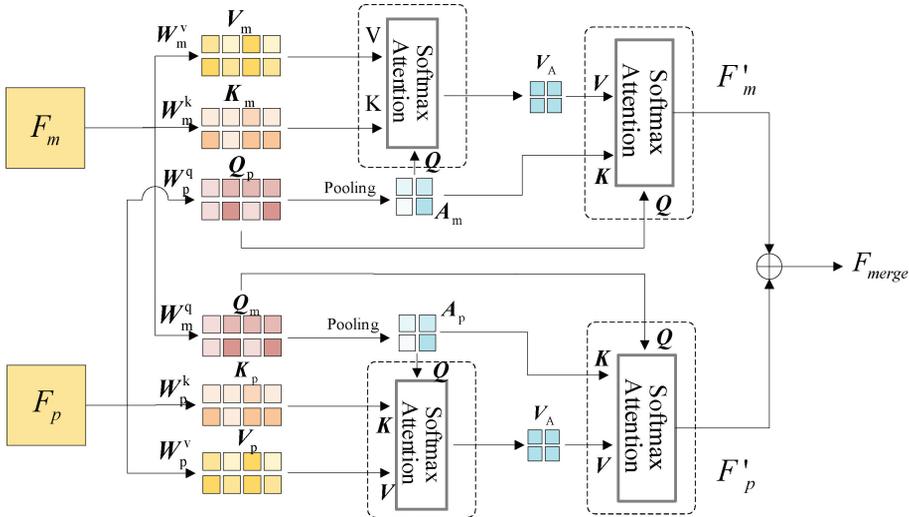


图5 并行交互网络

Fig. 5 Parallel interactive network

该部分包含两个并行分支: MRI分支和PET分支,每一个分支内采用多头交叉注意力的形式来进行跨模态信息交互。以PET分支为例,多头注意力包含多个子网络,对于单个子网络,首先使用3个不同的线性变换分别将输入的MRI病理特征 F_m 映射为查询矩阵 Q_m ,将PET病理特征 F_p 映射为键矩阵 K_p 和值矩阵 V_p ,计算过程如式(2)所示。

$$Q_m = W^q F_m, K_p = W^k F_p, V_p = W^v F_p \quad (2)$$

其中, W^q, W^k, W^v 表示网络待学习的权重变换矩阵。

与传统注意力机制不同的是,代理注意力引入了额外的代理向量 A ,代理向量 A 可以充当 Q 的代理以聚合来自 K 和 V 的信息,然后将信息广播回 Q ,这种方式集成了Softmax注意力和线性注意力。如图5所示,本文采用简单的池化策略从查询 Q 中提取出代理 A ,将代理 A 视为查询,并在 A, K, V 之间执行注意力操作,以从 K 和 V 中聚合出代理特征 V_A ,计算公式如式(3)所示。

$$V_A = \text{Softmax} \left(\frac{A(K)^T}{\sqrt{d_k}} \right) V \quad (3)$$

其中, d_k 表示 \mathbf{A}, \mathbf{K} 的维度大小。随后, 利用 \mathbf{A} 作为键, \mathbf{V}_A 作为查询 \mathbf{Q} 第二次注意力计算中的值, 这样可以将来自代理特征 \mathbf{V}_A 的全局信息广播回查询 \mathbf{Q} , 并得到该子网络的最终输出 h_n , 计算过程如式(4)所示。

$$h_n = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{A}^\top}{\sqrt{d_k}}\right)\mathbf{V}_A \quad (4)$$

最后, 将多个子网络的输出特征 h_n 进行拼接并经过线性变换将它们映射到与输入相同的特征空间以得到最终特征 F'_p , 如式(5)所示。

$$F'_p = \text{Concat}(h_1, h_2 \cdots h_n)\mathbf{W}^o \quad (5)$$

其中, h_i 表示第 i 个子网络的输出特征, \mathbf{W}^o 表示线性映射的权重参数。

MRI 分支与 PET 分支具有对称的操作流程, 最终两个分支分别生成对方特征交互引导的特征 F'_m 和 F'_p , 再将两个特征进行相加, 得到最终的融合特征 F_{merge} 。该网络的并行对称结构不仅可以促进两种模态特征之间的交互, 还可以保留自身模态有价值的信息, 从而在 AD 分类中实现两种成像信息的相互补充和增强。最后, 本文采用全连接层作为分类器, 融合特征 F_{merge} 作为输入, 经过全连接层和 Softmax 归一化后得到分类概率。

2 数据集构建

2.1 ADNI 数据集

本文使用的 sMRI 和 FDG-PET 数据均为 3D 影像, 来自于阿尔茨海默病神经影像学计划 (Alzheimer's disease neuroimaging initiative, ADNI)。该数据集是目前最大规模的阿尔茨海默病研究数据集之一, 于 2004 年由美国生物医学成像与生物工程研究所、美国国立卫生衰老研究所等组织联合创立。本文从 ADNI 中选择了具有配对 FDG-PET 和 T1 加权的 sMRI 影像数据的受试者。

2.2 数据集划分

本文以受试者为单位进行训练集、验证集和测试集划分, 避免出现数据泄露问题^[29-30]。同时为了确保训练集、验证集和测试集中数据的均衡分布, 在样本的多个统计参数上进行均衡划分, 统计参数包括年龄、性别、临床量表评分。其中临床量表采用简易精神状态检查量表 (mini-mental state examination, MMSE) 和临床痴呆评定量表 (clinical dementia rating, CDR), 这两种量表都是临床上医生进行 AD 诊断的重要参考依据。划分后的数据集统计情况如表 1 所示。构建的数据集共有 1 052 例样本, 其中

表 1 数据集统计情况

Table 1 Dataset statistics

数据集	类别	样本量/例	年龄/岁	MMSE	CDR
训练集	CN	347	74.49(7.46)	28.96(2.09)	0.02(0.17)
	AD	280	75.90(7.27)	21.80(2.44)	0.89(0.18)
验证集	CN	115	75.44(6.89)	28.98(1.17)	0.04(0.17)
	AD	105	74.81(7.61)	21.67(3.62)	0.88(0.23)
测试集	CN	109	75.04(5.96)	28.76(1.18)	0.03(0.21)
	AD	96	74.75(6.86)	21.11(3.16)	0.91(0.27)

包括认知正常 (cognitively normal, CN) 样本 571 例, 阿尔茨海默病 (AD) 样本 481 例。表 1 中, 年龄、MMSE、CDR 括号外为平均值, 括号内为标准差。

2.3 数据预处理

由于成像设备与受试者个体差异等因素的影响, 不同受试者的神经影像之间存在着多方面的差异, 因此需要进行一系列的预处理操作。本文的预处理操作基于 clinica 软件平台^[31]进行, 该平台支持多种神经影像处理工具。预处理过程包括: 1) 格式转换。将数据集中的原始 DICOM 文件转换为 NIFTI 格式的扫描。2) 偏差校正。3) 配准与空间标准化。通过刚性和非刚性配准方法将不同影像数据对齐, 使不同受试者的影像在相同空间中对应相同位置, 再映射到标准空间模板。4) 颅骨剥离。5) 灰度值归一化等。经过这些预处理步骤, 可以尽可能地消除影像由外部因素引起的差异。如图 6 所示, 展示了预处理前后的

sMRI 和 FDG-PET 影像。

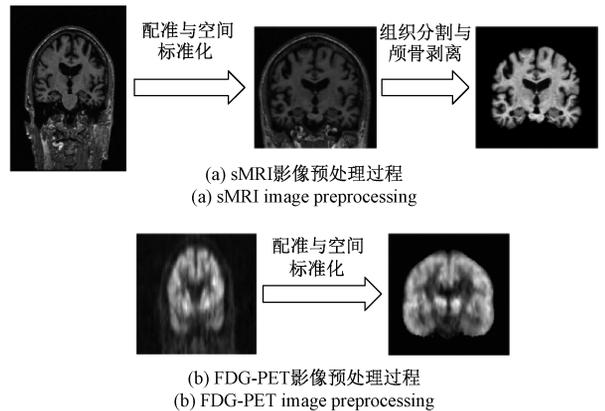


图 6 预处理前后的 sMRI 和 FDG-PET 影像

Fig. 6 sMRI and FDG-PET images before and after preprocessing

3 实验与结果分析

3.1 实验设置

本文的模型基于 24 GB GeForce RTX 3090 GPU, 采用 python 3.8.1、pytorch1.13.1 框架实现。操作系统为 Ubuntu20.04, CUDA 版本为 11.4。训练过程采用 Adam 优化器进行优化, 初始学习率为 0.001, 通过余弦退火策略调节学习率, 震荡周期为 50, 总训练次数为 300 轮。

本文采用交叉熵损失函数计算预测概率和真实标签的损失值, 交叉熵损失函数如式(6)所示。

$$H(p, q) = - \sum_x (p(x) \log q(x) + (1-p(x)) \log(1-q(x))) \quad (6)$$

其中, p 表示模型期望输出概率, q 表示模型实际输出概率。

3.2 评价指标

本文以二分类(AD/CN)任务进行实验, 为了准确衡量模型的性能, 使用以下指标进行评估: 准确率(accuracy, ACC)、敏感性(sensitivity, SEN)、特异性(specificity, SPE)。

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (7)$$

$$SEN = \frac{TP}{TP + FN} \times 100\% \quad (8)$$

$$SPE = \frac{TN}{TN + FP} \times 100\% \quad (9)$$

其中, TP 、 FP 、 TN 、 FN 分别代表真阳性、假阳性、真阴性、假阴性。准确率即为模型正确分类出患者类别的准确率。敏感性表示在患有疾病的所有人中, 模型分类正确的人所占有的比例, 敏感性越高漏诊的概率越小。特异性表示在未患有疾病的所有人中, 模型诊断正确的人所占的比例, 特异性越高误诊的概率越小。本文主要以准确率作为参考指标。

3.3 实验结果及分析

为了验证本文提出方法的性能和有效性, 本文与相关研究进行了对比实验, 实验结果如表 2 所示, 评价指标为准确率(ACC)、敏感性(SEN)与特异性(SPE)。表中的对比研究均与本文一样采用的 ADNI 数据库, 由于各研究都是从 ADNI 中选取的数据子集构成自己的数据集, 各数据集的患者存在部分不同, 且作者未公开源码, 所以只能进行粗略的对比。其中 MRI 表示 sMRI 影像数据, PET 表示 FDG-PET 影像数据, D 表示人口统计信息(demographics), AGE 表示年龄, APOE 表示风险基因(apolipoprotein E), 它是罹患阿尔茨海默病的重要遗传风险因素。

表 2 不同分类方法的结果对比

Table 2 Comparison of results of different categorization methods

对比方法	数据	AD/CN		
		ACC	SEN	SPE
RevGAN ^[32]	MRI+PET	92.28	90.38	94.37
TPA-GAN ^[33]	MRI+PET	92.00	89.10	94.00
M-ResNet ^[17]	MRI+PET	89.60	—	—
MHDN-DP ^[20]	MRI+PET+D+APOE	92.11	92.80	91.33
H-MACAT ^[34]	MRI+AGE	90.50	84.60	95.0
PA-GAN ^[24]	MRI+PET	93.40	97.50	92.80
HAMMF ^[25]	MRI+PET	93.15	93.15	93.57
本文	MRI+PET	93.66	91.67	95.41

从表 2 中可以看出, 本文提出的方法准确率高于其他对比方法。其中, PA-GAN^[24]、RevGAN^[32] 与 TPA-GAN^[33] 均使用了生成对抗网络来生成 PET 影像。PA-GAN 方法生成影像后, 将生成的 PET 影像与真实的 MRI 影像进行简单的映射得到融合影像, 然后使用 3D CNN 网络提取融合影像特征进行分类; RevGAN 方法将 PET 影像和 MRI 影像看做两个不同的通道, 通过通道融合将它们堆叠为一个 4D 图像输入到 CNN 网络中进行分类; TPA-GAN 方法则将生成的 PET 影像与真实的 MRI 图像通过一个路径传递的密集卷积网络实现多模态特征的融合; 与本文的方法相比, 这 3 种方法可以容许 PET 影像数据的缺失, 但在特征提取时都使用了较为简单的 CNN 网络, 忽略

了脑影像的特征复杂性, 且生成影像的信息真实性和丰富度都低于本文使用的真实影像。

M-ResNet^[17] 方法采用 ResNet 网络作为特征提取网络, 对每个模态数据独立的应用一个 ResNet 网络来进行特征提取, 之后使用简单的特征拼接方式来实现特征融合。该方法的特征融合策略过于简单, 导致分类性能较低。MHDN-DP^[20] 方法通过双重嵌入策略来生成影像数据和非影像数据的嵌入向量, 这些不同模态的嵌入向量通过高阶图扩散的形式进行融合以提升分类性能。该方法可以整合影像数据和非影像数据的高阶关系, 但对多模态数据中可能存在的冗余噪声干扰较为敏感。

H-MACAT^[34] 方法则提出了一种具有多尺度注意力

卷积和 Transformer 子网络的混合网络用于 AD 分类,该网络在 3D CNN 架构中采用多尺度注意力卷积来捕捉脑影像的局部病变,使用 Transformer 子网络将年龄信息嵌入到影像特征中,最后基于混合特征实现分类。与本文的方法相比,该方法可以提取多尺度的病变特征,但可能会忽略病变区域的相关性,同时年龄信息与影像特征的异质性过大,即使嵌入了年龄信息,对分类性能的提升也较小。

HAMMF^[25]方法使用 ResNet-18 来提取各模态影像的浅层特征,然后引入通道和空间注意力提取更细粒度的病理特征,并在通道和空间注意力中嵌入 Transformer 的多头注意力机制来捕捉不同模态之间的相关性。与本文的方法相似,该方法也在 3D CNN 架构中引入注意力机制来加强网络对脑影像细微病变特征的捕捉能力,不同的是,本文还引入了一个重构卷积来抑制冗余信息干扰。此外,该方法在提取深层特征的同时进行特征融合,这可能会导致丢失部分病变特征,且影响融合效果。而本文的方法在特征提取之后再独立的进行特征融合,两个阶段互不干扰,可以实现更好的特征提取及融合。

综上所述,本文提出的方法可以有效的提取脑影像病变特征,充分整合来自 MRI 和 PET 数据的互补信息以实现 AD 分类。

为了探究 sMRI 和 FDG-PET 影像数据各自对 AD 分类的影响,在构建的多模态影像数据集基础上,对比了使用多模态影像和单独使用 sMRI 影像或 FDG-PET 影像的分类结果,具体实验结果如图 7 所示。其中,MRI 和 PET 表示使用本文提出的特征提取子网络提取 sMRI 或 FDG-PET 影像特征后直接输入分类器进行分类。

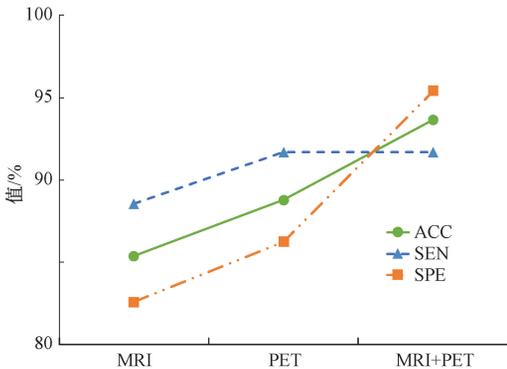


图 7 多模态数据与单模态数据分类对比结果
Fig. 7 Comparison results of multi-modal data and single-modal data classification

使用多模态数据的准确率为 93.66%,而单独使用 sMRI 数据和 FDG-PET 数据的准确率分别为 85.36% 和 88.78%。上述结果表明,使用多模态数据可以比单模态数据提供更多的病理信息,从而得到更准确的分类结果。同时,还可以发现,FDG-PET 可以产生比 sMRI 相对更好的整体性能,这与 Narazani 等^[17]的发现相符,可能的原因

是 PET 影像可以在脑部发生结构性病变前检测到代谢功能异常。

在本文提出的特征提取子网络中,通过嵌入坐标注意机制来增强模型的空间敏感性,同时使用空间-通道重构卷积以消除特征的冗余信息。为了验证本文提出的特征提取子网络的有效性,进行了消融实验,实验结果如表 3 所示。其中,CA 表示坐标注意机制;SC 表示空间-通道重构卷积;AgentAtt 表示本文的并行交互网络。表 3 中的数据表示分别使用消融的特征提取网络提取 sMRI 和 FDG-PET 的影像特征 F_m 和 F_p 后,再将两种影像特征输入本文提出的并行交互网络进行融合分类后得到的结果。

表 3 特征提取子网络的消融实验结果
Table 3 Results of ablation experiments for the feature extraction sub-network

CA	SC	AgentAtt	AD/CN		
			ACC	SEN	SPE
—	—	✓	90.24	86.46	93.58
✓	—	✓	91.71	90.62	92.66
✓	✓	✓	93.66	91.67	95.41

从表 3 可知,本文的特征提取子网络相较于基础的 CNN 网络,准确率提高了 3.42%,敏感性和特异性分别提高了 5.21% 和 1.83%。实验结果可以有效地证明本文的特征提取网络总体性能优于原基础网络。AD 的病变通常发生在额叶、顶叶皮质和海马体等区域,因此如何让网络关注这些区域的有效信息并抑制其他区域的冗余信息至关重要。嵌入坐标注意机制(CA)可以帮助网络关注到受疾病影响的各病变区域之间的空间位置关系,并学习到它们的潜在联系。空间-通道重构卷积则在网络提取深层特征的过程中,对特征在空间和通道维度上进行卷积压缩,消除特征的冗余信息并增强特征的表达能力。

为了验证本文提出的并行交互网络的有效性,本文分别使用 3 种不同的特征融合方式进行了对比实验,实验结果如表 4 所示。其中,Concat 表示简单的特征拼接方式,CrossAtt 表示传统的交叉注意力融合方法,AgentAtt 表示本文提出的并行交互网络。表 4 中的数据表示使用本文的特征提取子网络提取两种影像特征后分别使用这 3 种特征融合方式的结果。

表 4 不同特征融合方法的对比实验结果

Table 4 Comparative experimental results of different feature fusion methods

特征融合方法	AD/CN		
	ACC	SEN	SPE
Concat	90.73	89.58	91.74
CrossAtt	92.68	91.67	93.58
AgentAtt	93.66	91.67	95.41

对比表4中的 AgentAtt 与 Concat 实验结果,可以观察到本文提出的并行交互网络有效地提高了 AD 分类性能,准确率、敏感性和特异性分别提高了 2.93%、2.09% 和 3.67%。Concat 的实验结果最差,因为简单的特征拼接方式无法有效地聚合两种成像特征,甚至可能会出现其中一种特征完全占据主导地位的情况。对比 AgentAtt 和 CrossAtt 的实验结果,本文的并行交互网络也展现了更优异的性能。与传统交叉注意力相比,本文的并行交互网络通过代理的方式,集成了强大的 *Softmax* 注意力和高效的线性注意力,同时保留了全局上下文建模的能力,从而让网络实现了更好的性能。

在深度学习方法的视觉任务中,通过可视化注意力热图来验证模型的学习效果是一种常见的分析手段。如图8所示,展示了将3D梯度加权类激活映射(3D gradient-weighted class activation mapping, 3D Grad CAM)^[35]技术应用到本文提出模型的效果。

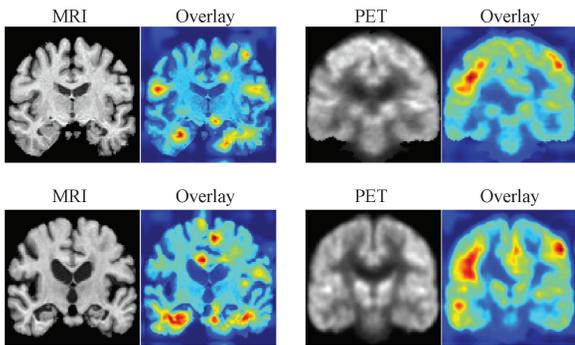


图8 可视化注意力热图

Fig.8 Visual heat map of attention

从图8中可以看出,模型注意力区域更集中在全脑组织上,同时在 sMRI 影像中更关注于包括颞叶和海马体等区域,在 FDG-PET 影像主要关注在顶叶、额叶和颞叶等区域,这些区域与 AD 影像学标志物特征的临床表现相符合,也对本文提出方法在分类性能上出色的表现提供了一定的可解释性。

4 结 论

阿尔茨海默病(AD)是一种复杂的疾病,患病期间,脑部会发生显著的结构萎缩和代谢异常。本文基于 sMRI 影像和 FDG-PET 影像提出了一种多模态影像 AD 分类方法。首先引入坐标注意机制和空间-通道重构卷积来捕捉影像中的特异性区域并限制冗余信息。之后,构建了跨模态并行交互网络来融合多模态互补信息进行 AD 分类。实验结果表明,本文提出的方法具有一定的先进性和有效性。此外,本文的方法也具有局限性。除了 sMRI 影像和 FDG-PET 影像外,还可以结合其他非影像数据来进行 AD 分类。同时,AD 也是一个长期渐进性的疾病,会随着时间发生变化,因此可以收集患者多个时期的数据进行纵向分

析。针对于此,应在未来开展更多的研究以进一步提高 AD 分类的准确性。

参考文献

- [1] LIM B, PRASSAS I, DIAMANDIS E P. Alzheimer disease pathogenesis: The role of autoimmunity[J]. The Journal of Applied Laboratory Medicine, 2021, 6(3): 756-764.
- [2] PATTERSON C. World Alzheimer report 2018[R/OL]. London: Alzheimer's Disease International, 2018. <https://www.alzint.org/resource/world-alzheimer-report-2018>.
- [3] ZHANG X X, TIAN Y, WANG Z T, et al. The epidemiology of Alzheimer's disease modifiable risk factors and prevention[J]. The Journal of Prevention of Alzheimer's Disease, 2021, 8: 313-321.
- [4] MARTÍ-JUAN G, SANROMA-GUELL G, PIELLA G. A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in Alzheimer's disease [J]. Computer methods and programs in biomedicine, 2020, 189: 105348.
- [5] SALVATORE C, CERASA A, CASTIGLIONI I. MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia 24 months before probable diagnosis [J]. Frontiers in aging neuroscience, 2018, 10: 135.
- [6] DUBOIS J, ALISON M, COUNSELL S J, et al. MRI of the neonatal brain: A review of methodological challenges and neuroscientific advances[J]. Journal of Magnetic Resonance Imaging, 2021, 53 (5): 1318-1343.
- [7] OU Y N, XU W, LI J Q, et al. FDG-PET as an independent biomarker for Alzheimer's biological diagnosis: A longitudinal study [J]. Alzheimer's Research & Therapy, 2019, 11(1): 1-11.
- [8] VEITCH D P, WEINER M W, AISEN P S, et al. Using the Alzheimer's disease neuroimaging initiative to improve early detection, diagnosis, and treatment of Alzheimer's disease[J]. Alzheimers Dement, 2022, 18: 824-857.
- [9] 杜淑慧,何小海,赵晓玲,等.基于时空和频域特征的 EEG 帕金森疾病识别 [J]. 电子测量技术, 2023, 46(3):121-127.
- [10] DU SH H, HE X H, ZHAO X L, et al. EEG Parkinson's disease identification based on spatiotemporal and frequency domain feature networks[J]. Electronic Measurement Technology, 2023, 46(3):121-127.

[10] 王肖,张俊华,王泽彤.并行多尺度特征融合的肺炎 CT 分割方法[J]. 国外电子测量技术,2023,42(11):

- 15-23.
WANG X, ZHANG J H, WANG Z T. Parallel multiscale feature fusion for pneumonia CT segmentation [J]. Foreign Electronic Measurement Technology, 2023, 42(11):15-23.
- [11] 黄昆, 张俊华, 普钟. 基于深度学习的脊椎 CT 图像分割[J]. 电子测量技术, 2022, 45(20):151-159.
HUANG K, ZHANG J H, PU ZH. Vertebra CT image segmentation based on deep learning [J]. Electronic Measurement Technology, 2022, 45(20):151-159.
- [12] 纪秋浪, 王继红, 杨晨, 等. 多尺度双重注意力网络医学图像分割模型[J]. 国外电子测量技术, 2022, 41(6):65-71.
JI Q L, WANG J H, YANG CH, et al. Multi-scale dual attention network medical image segmentation model [J]. Foreign Electronic Measurement Technology, 2022, 41(6):65-71.
- [13] 邓爽, 何小海, 卿粼波, 等. 基于改进 VGG 网络的弱监督细粒度阿尔兹海默症分类方法[J]. 计算机应用, 2022, 42(1):302-309.
DENG SH, HE X H, QING L B, et al. Weakly supervised fine-grained classification method of Alzheimer's disease based on improved visual geometry group network [J]. Journal of Computer Applications, 2022, 42(1):302-309.
- [14] JAHAN S, ABU T K, KAISER M S, et al. Explainable AI-based Alzheimer's prediction and management using multimodal data[J]. PLOS ONE, 2023, 18(11): e0294253.
- [15] GE CH, XU J Y, HU J Y, et al. MDMA: Multimodal data and multi-attention based deep learning model for Alzheimer's disease diagnosis[C]. 2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), IEEE, 2023: 120-127.
- [16] 张响泉, 吴晓红, 唐荔莉, 等. 基于多模态数据的阿尔茨海默病分类方法[J]. 计算机应用, 2023, 43(S2): 298.
ZHANG Y X, WU X H, TANG L L, et al. Alzheimer's disease classification method based on multimodal data [J]. Journal of Computer Applications, 2023, 43(S2): 298.
- [17] NARAZANI M, SARASUA I, PÖLSTERL S, et al. Is a PET all you need? A multi-modal study for Alzheimer's disease using 3D CNNs[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022: 66-76.
- [18] LIN W M, GAO Q Q, YUAN J N, et al. Predicting Alzheimer's disease conversion from mild cognitive impairment using an extreme learning machine-based grading method with multimodal data[J]. Frontiers in Aging Neuroscience, 2020, 12: 77.
- [19] ZHANG J, HE X H, LIU Y, et al. Multi-modal cross-attention network for Alzheimer's disease diagnosis with multi-modality data[J]. Computers in Biology and Medicine, 2023, 162: 107050.
- [20] AVILES-RIVERO A I, RUNKEL C, PAPADAKIS N, et al. Multi-modal hypergraph diffusion network with dual prior for Alzheimer classification [C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022: 717-727.
- [21] CHENG J Y, WANG H B, WEI SH CH, et al. Alzheimer's disease prediction algorithm based on decorrelation constraint and multi-modal feature interaction[J]. Computers in Biology and Medicine, 2024, 170: 108000.
- [22] GOLOVANEVSKY M, EICKHOFF C, SINGH R. Multimodal attention-based deep learning for Alzheimer's disease diagnosis [J]. Journal of the American Medical Informatics Association, 2022, 29(12): 2014-2022.
- [23] 薛景瑜. 基于深度学习的多模态阿尔兹海默症预测方法研究[D]. 济南: 齐鲁工业大学, 2021.
XUE J Y. Research on multi-modal Alzheimer's disease prediction method based on deep learning[D]. Jinan: Qilu University of Technology, 2021.
- [24] ZHANG M Y, SUN L J, KONG ZH K, et al. Pyramid-attentive GAN for multimodal brain image complementation in Alzheimer's disease classification[J]. Biomedical Signal Processing and Control, 2024, 89: 105652.
- [25] LIU X, LI W M, MIAO SH, et al. HAMMF: Hierarchical attention-based multi-task and multi-modal fusion model for computer-aided diagnosis of Alzheimer's disease [J]. Computers in Biology and Medicine, 2024, 176: 108564.
- [26] HOU Q B, ZHOU D Q, FENG J SH. Coordinate attention for efficient mobile network design [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [27] LI J F, WEN Y, HE L H. SCConv: Spatial and channel reconstruction convolution for feature redundancy[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6153-6162.
- [28] HAN D CH, YE T ZH, HAN Y Z, et al. Agent attention: On the integration of softmax and linear

- attention [J]. ArXiv preprint arXiv: 2312.08874, 2023.
- [29] BÄCKSTRÖM K, NAZARI M, GUI Y H, et al. An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images[C]. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018: 149-153.
- [30] 王斌,吴晓红,辜蕊,等. 基于改进 ResNet 的阿尔兹海默症分类网络[J]. 智能计算机与应用, 2023, 13(3): 69-76, 82.
WANG B, WU X H, GU R, et al. Alzheimer's disease classification network based on improved Residual Network [J]. Intelligent Computer and Applications, 2023, 13(3): 69-76, 82.
- [31] ROUTIER A, BURGOS N, DÍAZ M, et al. Clinica: An open-source software platform for reproducible clinical neuroscience studies [J]. Frontiers in Neuroinformatics, 2021, 15: 689675.
- [32] LIN W Y, LIN W M, CHEN G, et al. Bidirectional mapping of brain MRI and PET with 3D reversible GAN for the diagnosis of Alzheimer's disease [J]. Frontiers in Neuroscience, 2021, 15: 646013.
- [33] GAO X Y, SHI F, SHEN D G, et al. Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in Alzheimer's disease[J]. IEEE Journal of Biomedical and Health Informatics, 2021, 26(1): 36-43.
- [34] GAO X Y, CAI H J, LIU M H. A hybrid multi-scale attention convolution and aging transformer network for Alzheimer's disease diagnosis[J]. IEEE Journal of Biomedical and Health Informatics, 2023, 27(7): 3292-3301.
- [35] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]. IEEE International Conference on Computer Vision, 2017: 618-626.

作者简介

钟海杰, 硕士研究生, 主要研究方向为计算机视觉及医学图像处理。

E-mail: 2758442065@qq.com

吴晓红, 副教授, 硕士生导师, 主要研究方向为图像处理与识别、计算机视觉。

E-mail: wxh@scu.edu.cn

卿粼波, 教授, 博士生导师, 主要研究方向为多媒体通信与信息系统、人工智能与计算机视觉。

E-mail: qing_lb@scu.edu.cn

陈洪刚, 副研究员, 博士, 主要研究方向为计算机视觉、人工智能、多媒体通信。

E-mail: honggang_chen@scu.edu.cn

何小海(通信作者), 教授, 博士生导师, 主要研究方向为图像处理与网络通信、人工智能与大数据分析。

E-mail: hxh@scu.edu.cn