DOI:10.19651/j.cnki.emt.2518731

基于多模态影像的阿尔兹海默症分类研究*

陈 洛 王正勇 卿粼波 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)是一种神经系统疾病,主要影响人的脑细胞,是痴呆症的主要形式,由于其不可逆的特性,早期诊断对于减缓疾病进展至关重要。结构磁共振成像(sMRI)与氟脱氧葡萄糖正电子发射断层扫描(FDG-PET)是目前在神经退行性疾病研究中被广泛应用的两种成像技术,结合这两种影像来评估大脑状态能提高结果的准确性。本文提出了一种基于 Vision Transformer 的多模态融合框架,通过自注意力视觉变换器从单模态影像中提取特征,同时利用交互注意力融合网络专注于两种影像特征的相似性,既能强化各模态的独立表征能力,还能提高两种模态的交互性。同时使用深度置信网络降低提取特征的冗余性,提高不同模态的信息互补,最后采用集成分类器做出 AD 分类结果。选取 ADNI 数据集,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 94.65%、93.24%和 95.62%,与目前的融合方法相比,所提出的方法在 AD 分类任务中取得了更优异的结果。

关键词:阿尔茨海默病;多模态影像;深度学习;多模态分类
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

Multimodal imaging-based Alzheimer′s disease classification research

Chen Luo Wang Zhengyong Qing Linbo Chen Honggang He Xiaohai
(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: Alzheimer′s disease (AD) is a neurological disorder that primarily affects a person′s brain cells and is the main form of dementia; due to its irreversible nature, early diagnosis is critical to slowing the progression of the disease. Structural magnetic resonance imaging (sMRI) and fluorodeoxyglucose positron emission tomography (FDG-PET) are two imaging techniques that are widely used in neurodegenerative disease research, and combining these two images to assess the brain state can improve the accuracy of the results. In this paper, we propose a multimodal fusion framework based on Vision Transformer, which extracts features from unimodal images through a self-attentive vision transformer, and at the same time focuses on the similarity of the features of the two images by using an interactive attentional fusion network, which strengthens the independent characterization ability of each modality, and also improves the interactivity of the two modalities. At the same time, a deep confidence network is used to reduce the redundancy of the extracted features and improve the complementary information of different modalities, and finally an integrated classifier is used to make AD classification results. The ADNI dataset is selected and the classification performance of the proposed network is evaluated, and the accuracy, sensitivity and specificity reach 94.65%, 93.24% and 95.62%, respectively, and the proposed method achieves superior results in the AD classification task compared to current fusion methods.

Keywords: Alzheimer′s disease; multimodal imaging; deep learning; multimodal classification

0 引 言

阿尔茨海默病(Alzheimer′s disease, AD)是老年人中最常见的痴呆症之一,它会逐渐导致不可逆的脑损伤并影响正常的脑功能,大部分患者为 65 岁以上的人群。AD 通常表现为记忆、思维和情绪的异常,影响人们的正常活动。AD 不仅严重降低患者的生活质量,还会给护理人员带来困扰。全球至少有 5 000 万人可能患有 AD 或其他痴呆症。为痴呆症患者提供的医疗服务、长期照护以及临终关怀所产生的总体费用支出持续上升,给医疗系统和社会资源带

收稿日期:2025-04-30
*基金项目:四川省重点研发计划(2023YFS0195)、成都市重大科技应用示范项目(2019-YF09-00120-SN)资助

来显著压力。到 2050 年，AD 患者的数量可能将达到 1.15 亿[1]。因此，早期诊断和治疗阿尔兹海默症至关重要，利用计算机辅助进行 AD 诊断，有利于医生做出正确结果。

医学影像技术是识别脑部疾病进展的有力工具。具体地说，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）可以辅助诊断疾病并监测其进展。sMRI 可以很好地量化 AD 患者的脑组织萎缩[2]。Klöppel 等[3]利用受试者的 sMRI 图像生成大脑灰质密度图，并利用支持向量机（support vector machine，SVM）实现了对 AD 的识别。PET 可以监测人体葡萄糖代谢的变化。Ou 等[4]提取 PET 图像特征，通过逻辑回归区别健康对照者和 AD。对于单一模态的特征，观察到的特征信息通常仅从某个角度提供，而多模态的特征信息可以实现对人脑更全面的研究。因此，开发基于多模态医学图像的 AD 诊断模型已成为一种新趋势。最近的一些研究表明，利用多模态脑成像数据做出的结果比单模态数据具有更好的效果[5-7]。

现今许多研究采用深度学习的方法，使用多模态神经影像数据进行脑疾病诊断[8-11]。卷积神经网络（convolutional neural network，CNN）在 AD 诊断和预测方面取得了令人瞩目的表现，在图像分割领域也取得诸多成效[12-13]。在处理多模态信息时，大多数方法在图像级别执行早期融合或在特征级别执行后期融合。需要注意的是，由于 CNN 的局部性，早期融合会丢失不同模态之间的全局交互，而后期融合缺乏中间特征之间的交互，因此不能充分利用多模态信息。与 CNN 相比，Transformer 通过注意力机制可以捕捉到隐藏在多模态特征中的长程依赖关系。通常用于特征融合阶段，然而，不同模态的数据分布可能有很大差异，并且提取的特征通常位于它们自己的空间中，这使得很难通过注意力机制有效地学习多模态数据的互补信息[14]。

近年来，研究通过结合 sMRI 和 PET 影像的深度学习方法推动了 AD 诊断的发展。针对多模态融合，提出了不同的策略，Lu 等[15]使用多尺度深度神经网络融合 MRI 和 PET 提取一维特征。Lin 等[16]提出 3D 可逆生成对抗网络（generating adversarial networks，GAN）来弥补缺失数据，并使用 sMRI 和 PET 数据的通道级早期融合进行 AD 诊断。Song 等[17]通过将 MRI 中灰质与 FDG-PET 数据叠加，并将其输入 3DCNN 进行分类。Liu 等[18]提出一个级联框架，包括多个深度 3DCNN 学习局部图像特征，并结合 2DCNN 对高层特征进行融合。Feng 等[19]结合 3D CNN 和 LSTM，对 MRI 和 FDG-PET 数据进行晚期融合来进行 AD 诊断。Huang 等[20]提出基于 3DVGG 的早期和晚期融合方法。Narazani 等[10]在研究基于 3DCNN 的 MRI 和 PET 融合方法后发现，这些现有的多模态融合技术的诊断性能尚未优于单独使用 PET 的效果。

Transformer 在医疗影像任务中应用广泛，在使用多模态影像进行 AD 诊断中取得成果，Li 等[21]结合 CNN 和 Transformer 模块进行多模态医学图像融合。Zhang 等[22]提出一个端到端 3D ResNet 框架，利用注意力机制在 MRI 和 PET 数据间融合多级特征。Gao 等[23]提出多模态 Mul-T，使用 DenseNet 和空间注意力提取全局和局部特征，再通过跨模态 Transformer 融合 T1、T2-MRI 和 PET 数据。Miao 等[24]提出多模态多尺度 Transformer 融合网 MMTFN，结合 CNN 残差模块和 Transformer，联合学习多模态数据进行 AD 诊断。Tang 等[25]通过 3DCNN 提取结构性 MRI 和 PET 图像的深度特征表示，并利用改进的 Transformer 渐进学习特征间的全局相关信息。这些方法通常采用 CNN 进行初步特征提取，再利用 Transformer 实现特征融合。然而这些组合方式无法完全发挥 Transformer 在多模态学习中的潜力，导致部分病理特征丢失，产生了一些冗余信息，降低分类的准确性，最终导致融合效果不够好。

为了应对所提及的问题，本文构建了一种基于多模态影像 AD 分类方法。针对脑影像的特点，设计了一个基于 Vision Transformer 架构的特征融合网络[26]，自注意力变换器由一系列变换器块组成，每个块都具有多头自注意力机制，独立提取 sMRI 和 FDG-PET 影像的特征，更加充分地提取病理信息。同时引入交互注意力融合网络，专门捕捉两种模态的重叠信息中的相关性，实现更有效的融合。由于交互注意力模块尤其关注模态间的潜在相似性，得到的高维特征可能存在冗余信息、噪声或弱相关特征，本文使用深度置信网络（deep belief network，DBN）通过无监督学习方式对特征进行进一步优化，增强关键特征的表征能力，同时减少不必要的特征干扰，最终输入集成分类器得到分类结果，本文提出的新型融合网络有效获取两种影像的互补信息，提高了 AD 分类的准确率。

1　多模态影像分类网络

1.1　网络整体框架

本研究构建了一种新颖的端到端多模态影像分类网络，用于 AD 诊断，网络的整体架构如图 1 所示。该方法首先通过自注意力视觉变换器从 sMRI 和 FDG-PET 影像中分别提取病理特征 $\boldsymbol{F}_{MRI}$ 和 $\boldsymbol{F}_{PET}$。同时这两种影像被输入到交互注意力融合网络中进行信息融合，得到融合特征 $\boldsymbol{F}_{MP}$。最后，经过 DBN 处理的三种特征被输入至集成分类器，完成阿尔茨海默症的分类任务。该计算过程可通过式(1)表示。

$$P(Y)=Classifier(F_{DBN}(F_{Self}(m,p)+F_{Interactive}(m,p))) \tag{1}$$

式中：P 代表分类结果，$Classifier$ 表示分类器，F_{DBN} 表示深度贝叶斯网络，F_{Self} 表示特征提取的自注意力视觉变换器，$F_{Interactive}$ 表示交互注意力融合网络，用 m、p 表示 sMRI 和 FDG-PET。

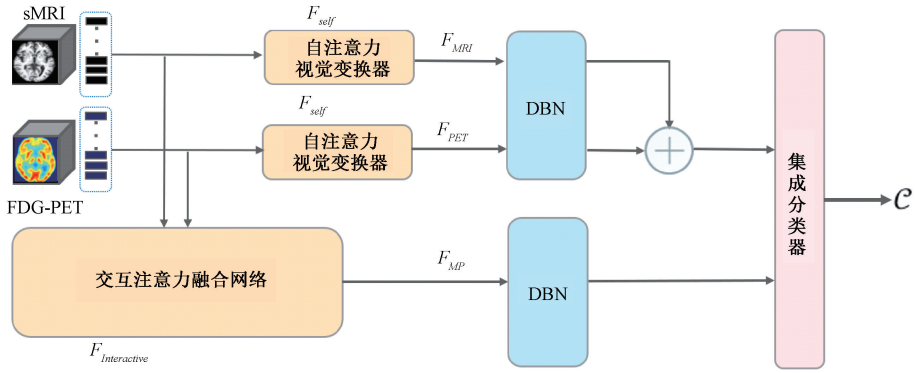

图 1　网络总体框架图

Fig. 1　General framework diagram of the network

1.2　自注意力视觉变换器

本文的自注意力视觉变换器基于 Vison Transformer，该架构由一系列变换器块组成，每个块都具有多头自注意力机制[27]，如图 2 所示。输入图像首先被分成不重叠的块，然后线性嵌入并输入到变换器块中。此外，该模型使用可学习的位置嵌入对每个块的空间信息进行编码。有效捕捉 sMRI 和 FDG-PET 影像中的复杂空间依赖关系，能够提取丰富的特征信息。与传统的卷积网络不同，Vison Transformer 可以捕捉长距离依赖关系，特别适用于高维医学影像数据的处理。

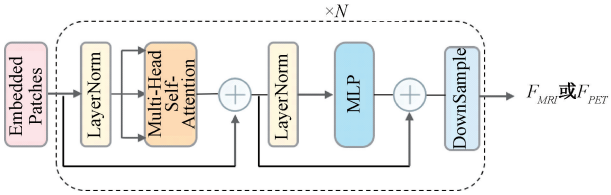

图 2　自注意力视觉变换器

Fig. 2　Self-attention vision transformer

对于 sMRI 和 FDG-PET 影像 $\boldsymbol{m}, \boldsymbol{p} \in R^{H\times W\times D}$，本文将其划分为大小为 $h\times w\times d$ 的非重叠体素补丁。然后，每个补丁被展平并投影到一个维度为 f_e 的特征空间，形成输入张量 $\boldsymbol{X}_m \in \boldsymbol{R}^{N\times f_e}$，其中，$N$ 表示补丁的总数，特征嵌入通过一个可学习的线性投影获得，确保每个补丁都被映射到一致的特征维度。CNN 和它有所不同，这种基于补丁的嵌入方式能够保留更多的全局上下文信息。为了提取 sMRI 和 FDG-PET 的特征，在 Vsion Transformer 编码器内利用自注意力机制。以 sMRI 支路为例，首先输入特征通过层归一化得到式(2)：

$$\hat{\boldsymbol{X}}_m = LayerNorm(\boldsymbol{X}_m) \tag{2}$$

每个补丁的归一化特征经过线性投影，分别映射为查询、键和值：$\boldsymbol{Q}_m = \hat{\boldsymbol{X}}_m \boldsymbol{W}_Q$，$\boldsymbol{K}_m = \hat{\boldsymbol{X}}_m \boldsymbol{W}_K$，$\boldsymbol{V}_m = \hat{\boldsymbol{X}}_m \boldsymbol{W}_V$。接着计算自注意力权重 A_m 为 $Softmax\left(\frac{\boldsymbol{Q}_m \boldsymbol{K}_m^{\mathrm{T}}}{\sqrt{f_e}}\right)$，其中，$softmax$ 归一化确保特征交互具有概率权重。最终的注意力输出计算如式(3)所示。

$$\boldsymbol{S}_m = \boldsymbol{A}_m \boldsymbol{V}_m \tag{3}$$

这一机制允许每一个补丁关注影像中的相关区域，与传统 CNN 只能关注局部区域不同，Vison Transformer 通过自注意力机制让每个补丁与整个影像交互，提取全局信息，提高了特征的判别能力，使模型更容易捕捉远程依赖关系，从而提升医学影像分类的效果。为了增强特征的表达能力，本研究采用多层编码器，每一层由自注意力、多层感知机(multi-layer perceptron，MLP)和残差连接组成，通过残差连接将 $\boldsymbol{S}_m$ 和 $\hat{\boldsymbol{X}}_m$ 结合起来，再经过层归一化和 MLP 得到输出如式(4)所示。

$$\boldsymbol{X}_m^{out} = MLP(LN(\boldsymbol{S}_m + \hat{\boldsymbol{X}}_m)) \tag{4}$$

其中，MLP 模块包含一个两层的前馈网络，并采用(gaussian error linear unit，GELU)激活函数，以进一步优化特征表达。为了同时捕捉精细特征和高级上下文信息，本研究引入多尺度特征提取策略。在自注意力视觉变换器之间使用下采样算子，这样可以在不同尺度上逐步聚合信息，提高分类任务的鲁棒性。经过多层变换器最终得到特征 $\boldsymbol{F}_{MRI}$ 和 $\boldsymbol{F}_{PET}$，经过 DBN 再相加得到特征 X_{self}^{out}。自注意力机制允许模型捕捉远距离区域之间的依赖关系，对识别医学影像中的细微病变尤为重要。自注意力视觉变换器分支分别处理 sMRI 和 FDG-PET 数据，确保提取互补特征，同时减少跨模态冗余。层级化、多尺度特征学习策略提高了模型对图像分辨率和对比度变化的适应性。

1.3　交互注意力融合网络

在 AD 分类任务中，不同模态数据具有互补性，能够提供多层次的病理信息。如何高效融合这些异质信息，是目前的研究中的关键问题。本文设计了交互注意力融合网络以充分利用 sMRI 和 FDG-PET 影像的互补信息，克服了单模态特征提取的局限性。融合网络的主要优势是捕捉跨模态相关性，仅靠自注意力视觉变换器提取 sMRI 和

FDG-PET 会导致信息不完整，降低特征融合效果。本研究提出的交互注意力融合网络包含两个独立的分支，分别处理来自不同模态的输入数据。每个分支内部包含多个堆叠的 Vision Transformer 编码器层，如图 3 所示。

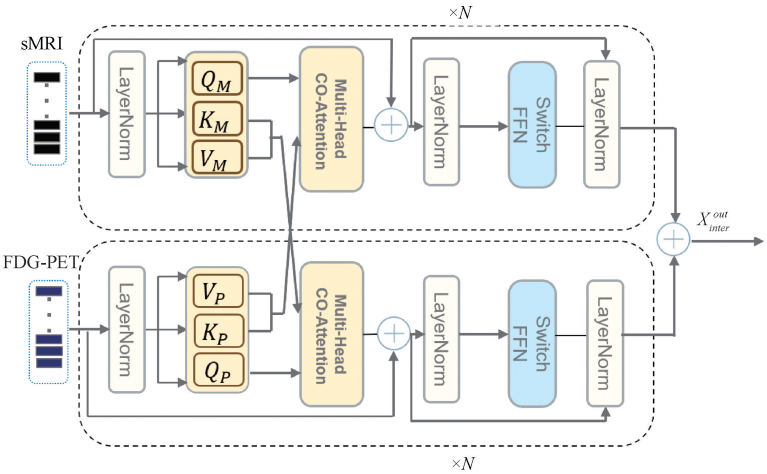

图 3　交互注意力融合网络

Fig. 3　Interactive attention fusion network

首先将 sMRI 和 FDG-PET 影像分别输入到各自的嵌入层进行处理。每个影像会被切割成若干个块，通过线性映射到嵌入空间。接着，为每个模态添加位置编码如式(5)、(6)，这样可以补充空间信息，然后为了确保输入数据的稳定性，对每个模态的嵌入进行归一化处理。

$$\boldsymbol{y}_{MRI} = \boldsymbol{x}_{MRI} + \boldsymbol{PE}_{MRI} \tag{5}$$

$$\boldsymbol{y}_{PET} = \boldsymbol{x}_{PET} + \boldsymbol{PE}_{PET} \tag{6}$$

在多头联合注意力机制(collaborative attention，CA)中，sMRI 模态的查询 $\boldsymbol{Q}$ 将与 $\boldsymbol{y}_{PET}$ 的键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$ 进行加权点积计算，以捕捉两种模态之间的交互信息。查询 $\boldsymbol{Q}_{MRI}$ 是通过对 $\boldsymbol{y}_{MRI}$ 输入进行线性变换得到式(7)，FDG-PET 的键和值是通过线性变换从 $\boldsymbol{y}_{PET}$ 输入获得式(8)：

$$\boldsymbol{Q}_{MRI} = Linear_Q(\boldsymbol{y}_{MRI}) \tag{7}$$

$$\boldsymbol{K}_{PET}, \boldsymbol{V}_{PET} = Linear_{KV}(\boldsymbol{y}_{PET}) \tag{8}$$

然后对 sMRI 模态的查询向量 $\boldsymbol{Q}_{MRI}$ 和键向量 $\boldsymbol{K}_{PET}$ 进行加权点积计算，以衡量 sMRI 模态对 FDG-PET 模态关键特征的关注程度，即计算注意力分数，计算方式如式(9)，其中，d 表示查询和键维度。

$$\boldsymbol{A}_{MRI-PET} = \frac{\boldsymbol{Q}_{MRI} \cdot \boldsymbol{K}_{PET}^{\mathrm{T}}}{\sqrt{d}} \tag{9}$$

然后对注意力分数进行 Softmax 归一化，以将其转换为注意力权重矩阵。该步骤的目的是对 sMRI 查询向量与 FDG-PET 不同区域的匹配得分进行归一化，使 sMRI 关注的 FDG-PET 特征更加突出，同时抑制无关信息，Softmax 操作是针对点积注意力得分矩最后一个维度($dim = -1$)进行的，如式(10)所示。

$$\boldsymbol{A}_{MRI-PET-Softmax} = Softmax(\boldsymbol{A}_{MRI-PET}) \tag{10}$$

利用这些注意力权重对 PET 模态的值进行加权平均，得到融合输出式(11)：

$$\boldsymbol{O}_{MRI-PET} = \boldsymbol{A}_{MRI-PET-Softmax} \cdot \boldsymbol{V}_{PET} \tag{11}$$

另一条支路同样地计算得出 $\boldsymbol{O}_{PET-MRI}$，然后进入切换前馈神经网络(switch feedforward neural network，Switch FFN)。Switch FFN 是一个可选择性激活不同前馈网络的机制，能够在降低计算开销的同时保持表示能力。为了保持训练的稳定性，再次对 Switch FFN 的输出进行 LayerNorm 操作，两条支路如式(12)、(13)所示，相加得到最终输出 $\boldsymbol{X}_{inter}^{out}$。

$$\boldsymbol{X}_{MP}^{out} = LN(SwitchFFN(\boldsymbol{O}_{MRI-PET})) \tag{12}$$

$$\boldsymbol{X}_{PM}^{out} = LN(SwitchFFN(\boldsymbol{O}_{PET-MRI})) \tag{13}$$

此过程将在多个 Vision Transformer 编码器层中进行堆叠，并为最终任务(如分类、回归等)提供有效的特征表示。经过多层编码后，输出能够捕捉 sMRI 和 FDG-PET 影像之间的复杂关系，并为后续任务提供有意义的信息。联合注意力机制利用每个模态的查询与另一个模态的键和值计算加权点积注意力，实现模态间信息的交互和融合。Switch FFN 通过选择性激活前馈网络层，提高计算效率的同时保持表达能力。LayerNorm 在每个子模块前后应用确保网络的训练稳定性。此架构实现了 sMRI 和 FDG-PET 影像的有效融合，适用于多模态影像处理任务，能够捕捉和利用不同模态之间的细粒度关系。

1.4　集成分类器

前文中提到的，在多模态特征处理框架中，经过自注意力视觉变换器处理得到的 $\boldsymbol{X}_{self}^{out}$，以及经过交互注意力融合模块处理后的 $\boldsymbol{X}_{inter}^{out}$，再送入深度置信网络进行进一步的特征提取与表示学习。DBN 是一种无监督深度学习模型，它由多个受限玻尔兹曼机(restricted Boltzmann machine，RBM)堆叠构成。能够在高维特征空间中捕获复杂的非线性关系，并学习到更具判别力的特征表示。经过 DBN 提取的深度特征输入到集成分类器如图 4 所示，进行最终的分类。

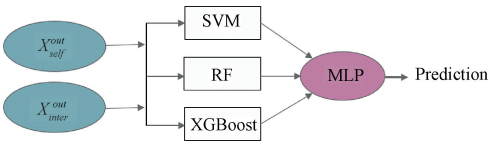

图 4　集成分类器

Fig. 4　Integrated classifier

集成分类器应用广泛[28],本文设计了一个典型的 Stacking 集成学习框架作为最终的分类器,用于对深度置信网络输出的特征进行分类预测。该框架采用两层模型结构,包括基分类器层和元分类器层,通过多模型集成的方式提高模型的泛化能力和预测准确性。其中,基分类器层包括 SVM、随机森林(random forest,RF)和梯度提升决策树(eXtreme gradient boosting,XGBoost)3 种不同类型的分类器,这些分类器以并行方式对 DBN 提取的输入特征进行处理,分别生成初步的预测结果。不同的基分类器具有各自的优势,例如 SVM 在高维数据中表现优异,随机森林在处理非线性关系和防止过拟合方面具有优势,而 XGBoost 则因其强大的特征学习能力和对不平衡数据的良好适应性而被广泛应用。这种多样化的基分类器设计能够有效捕捉数据中的不同模式,通过模型的多样性来增强整体的泛化能力。

最后元分类器层以第一层基分类器的预测结果作为输入,在这一层中,本研究选用了 MLP 作为元分类器,它具备更强的非线性学习能力,适合处理复杂的特征交互信息。进一步学习预测特征间的关系,从而得到最终的分类概率。

2　数据集构建

2.1　ADNI 数据集

阿尔茨海默病神经影像学计划(Alzheimer′s disease neuroimaging initiative, ADNI)是一个由多个研究机构共同参与的大型多中心合作项目,旨在加速阿尔茨海默病(AD)相关研究的进展。该项目于 2004 年正式启动,得到了美国国立衰老研究所(NIA)和美国国立生物医学成像与生物工程研究所(NIBIB)等机构的资助和支持。通过收集和共享多模态神经影像、生物标志物、认知评估等数据,以促进 AD 早期诊断和疾病进展研究。本研究从 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

为防止数据泄露,本研究在构建训练集、验证集和测试集时,以受试者为单位进行划分。为了确保各子集在统计特征上的分布一致,样本在年龄、性别及临床量表评分等多个参数维度上实现了均衡划分。所采用的临床量表包括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),这两项评估工具是临床医生进行 AD 诊断时的重要依据。划分后数据集的统计信息如表 1 所示。该数据集共包含 449 名受试者样本,其中认知正常(cognitively normal,CN)样本为 233 例,AD 样本为 216 例。表 1 中列出了年龄、MMSE 和 CDR 的分布情况,其中括号外为均值,括号内为标准差。

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	164	73.36(5.36)	27.97(2.07)	0.02(0.16)
	AD	150	74.80(6.67)	21.50(2.34)	0.79(0.19)
验证集	CN	34	76.44(6.78)	27.96(1.17)	0.04(0.18)
	AD	33	75.83(7.51)	22.67(3.52)	0.98(0.24)
测试集	CN	35	74.02(5.86)	28.66(1.38)	0.03(0.11)
	AD	33	74.65(6.75)	21.31(3.30)	0.82(0.17)

2.3　数据预处理

由于成像设备的不同以及个体间的生理差异,不同受试者的神经影像可能存在多种不同,因此需要进行一系列预处理步骤以提高数据的一致性和可比性。本研究采用了 clinica 软件平台进行预处理[29],该平台整合了多种神经影像处理工具。具体预处理流程包括:1)格式转换:将原始 DICOM 文件转换为 NIFTI 格式,以便后续分析;2)偏差校正:纠正影像中的非均匀信号强度;3)配准与空间标准化:采用刚性和非刚性配准方法,使不同受试者的影像在同一空间中对应相同的解剖结构,并映射到标准空间模板;4)颅骨剥离:去除颅骨等非脑组织,提高影像质量;5)灰度值归一化:调整影像灰度值,使其在不同受试者间保持一致。以上步骤共同构成了影像数据的预处理流程。可最大程度减少外部因素对影像的影响,提高数据的可靠性和分析的准确性。如图 5、6 所示。

3　实验与结果分析

3.1　实验设置

本研究模型的训练在配备 24 GB 显存的 GeForce RTX 3090 GPU 上进行,实验平台基于 Python 3.10.6 编

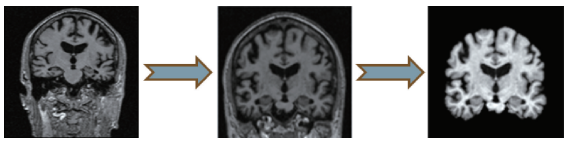

图 5　预处理前后的 sMRI 影像

Fig. 5　sMRI images before and after preprocessing

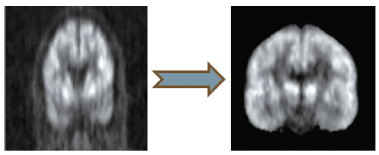

图 6　预处理前后的 FDG-PET 影像

Fig. 6　FDG-PET images before and after preprocessing

程语言，采用 PyTorch 1.12.1 深度学习框架，操作系统为 Ubuntu 20.04，CUDA 版本为 11.3。训练过程中，优化器选择 Adam，初始学习率设置为 5×10^{-4}，并通过余弦退火策略进行动态调整，震荡周期设定为 50 轮，总训练轮数为 150 轮。

为计算模型预测结果与真实标签之间的误差，本文选用交叉熵损失函数，其数学表达形式如式(14)所示。

$$H(p,q)=-\sum_{x}(p(x)\log q(x)+(1-p(x))\log(1-q(x))) \tag{14}$$

式中：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

本文聚焦于 AD/CN 二分类任务，为了评估模型的有效性，使用准确率(accuracy，ACC)、敏感性(sensibility，SEN)、特异性(specificity，SPE) 作为评估指标。

$$ACC=\frac{TP+TN}{TP+TN+FP+FN}\times100\% \tag{15}$$

$$SEN=\frac{TP}{TP+FN}\times100\% \tag{16}$$

$$SPE=\frac{TN}{TN+FP}\times100\% \tag{17}$$

式中：TP 表示正确识别阳性病例的数量，FP 映误判阴性样本为阳性的情况，TN 指准确判定阴性样本的数量，FN 指的是模型未能正确识别阳性病例，即漏诊的误判。准确率用于评估模型在正确识别各类患者方面的整体分类能力。敏感性反映模型在所有实际患病个体中成功识别出的比例，较高的敏感性意味着漏诊风险较低。特异性衡量模型在所有实际未患病个体中准确识别的比例，特异性越高，误诊的可能性越小。为全面评估模型的分类性能，本文将分类准确率作为主要评价指标。

3.3　实验结果及分析

为了验证本文提出方法的有效性和准确度，本研究开展了有关多模态的对比实验，评估指标涵盖准确率、灵敏性及特异性，具体数据结果详如表 2 所示。需特别说明的是，所有参照研究均基于 ADNI 数据库选取不同子集构建独立数据集，由于受试者筛选标准差异及原始算法代码未对外公开，且受试者群体存在部分差异，故本研究仅开展大致对比分析。本文 MRI 代表 sMRI 影像，PET 代表 FDG-PET 影像。

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
BOKM[30]	MRI+PET	90.60	90.50	90.70
MiSePyNet[31]	MRI+PET	93.13	90.32	95.49
Mul-T[23]	MRI+PET	94.40	93.00	95.50
Multi-Modality[32]	MRI+PET	90.10	90.85	89.21
FSBi-LSTM[19]	MRI+PET	86.36	83.33	88.78
3D-Class[16]	MRI+PET	92.28	90.38	94.37
MMTFN[24]	MRI+PET	91.67	**93.33**	86.66
Ours	MRI+PET	**94.65**	93.24	**95.62**

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。BOKM[30] 提出了一种基于核组合的新型 AD 和 MCI 多模态数据融合和分类方法。与传统的直接特征连接方法相比，该方法提供了一种统一的方式来组合异构数据，特别是对于不同类型的数据无法直接连接的情况。该方法可以灵活的对不同的数据模态使用不同的权重。这类方法依赖于手工制作的特征，通常会导致次优结果。基于深度学习的模型，尤其是卷积神经网络在自动 AD 诊断和预测方面取得了令人瞩目的表现。在处理多模态信息时，大多数方法在图像级别执行早期融合或在特征级别执行后期融合。需要注意的是，由于 CNN 的局部性，早期融合可能会丢失不同模态之间的全局交互，而后期融合缺乏中间特征之间的交互，因此不能充分利用多模态信息。

MiSePyNet[31] 方法遵循分解卷积的思想，为每个视图部署可分离的 CNN、切片和空间 CNN。这种设计的好处是，能够联合考虑轴向、冠状和矢状视图而不会丢失空间信息。此外，每个视图都以多尺度网络为特征，以捕捉不同的变化并扩大感受野的范围，从而增强判别特征图。模型采用了多个卷积层和多尺度网络设计，提高了模型的表达能力，但增加了模型的复杂性。设计中考虑了轴向、冠状和矢状视图，但并没有提到如何处理和选择这些视图中的关键信息。不同视图可能包含不同类型的空间信息，模型需要有效地从这些视图中提取相关特征，避免无效信息干扰。而且，尽管多视图联合学习提高了模型的鲁棒性，但不同视图之间的相关性可能会导致冗余特征，影响计算效率。

Mul-T[23] 方法用多级引导生成对抗网络(multi-level

guidance generative adversarial network,MLG-GAN)和多模态变换器分别用于不完整图像生成和疾病分类。首先,文章提出用 MLG-GAN 来生成缺失数据,并以来自体素、特征和任务的多级信息为指导。除了体素级监督和任务级约束之外,还提出了一个特征级自回归分支来嵌入目标图像的特征以实现精确生成。利用完整的多模态图像,文章提出了一种用于疾病诊断的 Mul-T 网络,它不仅可以结合全局和局部特征,还可以通过跨模态注意机制对从一种模态到另一种模态的潜在相互作用和相关性进行建模。尽管 Mul-T 网络通过跨模态注意机制来建模模态之间的相互作用和相关性,但这种跨模态建模方法是否能在所有任务和数据集上都有效,尤其是在模态间差异较大的情况下(例如,MRI 与 PET 图像的成像原理和特征差异较大)。不同模态数据之间的非线性关系可能比当前方法所能捕捉的更加复杂,因此可能导致某些信息的丢失或误判。

Multi-Modality[32]方法提出了一种利用卷积神经网络来整合海马区 MRI 和 FDG-PET 图像中包含的所有多模态信息,用于 AD 的诊断。与传统的机器学习算法不同,该方法不需要手动提取特征,而是利用 3D 图像处理 CNN 来学习用于 AD 诊断或预处理后的特征。然而在特征提取过程中采用了相对简单的 CNN 架构,未能充分表征脑影像中固有的特征复杂性。此外,其生成影像在信息的真实性与丰富度方面也不及本研究所使用的真实影像。

FSBi-LSTM[19]方法设计了一个新颖的深度学习框架。具体来说,利用了 3DCNN 和全堆叠双向长短期记忆网络的优点。首先,设计了一个 3DCNN 架构来从 MRI 和 PET 中获取深度特征表示。然后将 FSBi-LSTM 应用于深度特征图中的隐藏空间信息,以进一步提高其性能。最后,在 ADNI 数据集上验证了方法。但是将 3DCNN 与 FSBi-LSTM 结合的设计相对复杂。3DCNN 负责从 MRI 和 PET 图像中提取空间特征,而 FSBi-LSTM 用于处理序列数据的时间依赖性,模型架构的多层次和多模块组合可能会增加训练过程中的调参难度,并容易导致过拟合或梯度消失等问题。

3D-Class[16]方法利用多模态互补信息,首先采用可逆生成对抗网络模型来重建缺失数据。然后使用一种具有多模态输入的 3D 卷积神经网络分类模型来执行 AD 诊断。该方法使用合成数据来弥补缺失的模态信息,虽然可以提高诊断准确性,但过度依赖合成数据可能会使模型失去对真实临床数据的敏感性。在多模态图像融合时,3D CNN 可能会将两种模态的特征过度压缩,尤其是在深层网络中。虽然 3D CNN 能够提取空间上的高维特征,但如果没有合适的融合机制,它可能会丢失一些关键信息。例如,在 PET 图像和 MRI 图像的融合中,可能会出现两种模态的信息无法完全有效地互补的情况,影响最终的分类效果。

MMTFN[24]提出了一种基于 Transformer 的多模态多尺度自注意力融合方法 MMTFN,用于阿尔茨海默病(AD)诊断,该方法利用多个大脑图像扫描和多层 Transformer 进行 AD 分析。MMTFN 整合了来自多个模态在不同阶段和层次的细粒度特征,以解决当前多模态融合方法的局限性。MMTFN 充分利用了每个模态在不同尺度上的特征图信息,在图像数据融合中构建了不同尺度之间的联合表示。其中,3D 多尺度残差块通过多种尺寸的膨胀卷积提取多尺度的细粒度表示。多尺度融合网络使用在不同尺度和不同模态下提取的细粒度特征表示,并构建了不同特征之间的依赖关系。

上述多种融合方法存在一些不足,而本文的框架基于视觉 Transformer,有效整合 sMRI 和 PET 数据。该架构具有自注意力视觉变换器和一种新颖的交互注意力融合模块,协同融合 sMRI 和 PET 数据,同时引入多模态归一化方法以减少冗余依赖,从而提升诊断性能。

综上所述,本文所提出的方法能够有效提取脑影像中的病理特征,并充分融合 sMRI 与 FDG-PET 两种模态间的互补信息,从而 AD 分类的性能。为进一步分析不同模态对 AD 分类任务的贡献,本文基于构建的多模态影像数据集,分别评估了多模态输入与单独使用 sMRI 或 FDG-PET 数据时的分类效果。实验结果如表 3 所示,展示了各方案在分类性能上的差异与优势。

表 3　单模态和多模态的对比实验

Table 3　Comparative experiments between unimodal and multimodal　%

AD/CN	ACC	SEN	SPE
sMRI	91.56	88.37	90.70
FDG-PET	92.28	91.38	89.52
Multi-modality	**94.65**	**93.24**	**95.62**

使用多模态数据的准确率为 94.65%,仅输入 sMRI 数据或者 FDG-PET 数据,它们的准确率分别为 91.56% 和 92.28%。实验结果说明多模态数据的融合不仅提升了模型的分类能力,提出的方法还增强了对不同影像模态互补信息的利用,有助于更全面地了解疾病特征,实验结果验证了本文提出的多模态方法在 AD 诊断中的优势。

在本文提出的 AD 分类框架中有自注意力视觉变换器,用来专注每一个模态的特征,还使用交互注意力融合模块来提取 sMRI 影像和 FDG-PET 影像的结合特征,后阶段使用 DBN,它通过无监督学习逐层提取更加抽象的特征,从而增强 MRI 和 FDG-PET 影像特征的表征能力。为了验证本文提出的特征提取子网络的有效性,表 4 展现了网络架构的消融实验结果。

其中,Y_{Self} 表示前文提到的自注意力视觉变换器,$Y_{Interactive}$ 表示交互注意力融合网络,DBN 是深度信念网络,Stacking 表示集成分类器,数据结果可以展现出每一部分

表 4　网络架构的消融实验结果

Table 4　Results of ablation experiments on network architectures　%

AD/CN	ACC	SEN	SPE
Y_{Self} ＋DBN＋Stacking	90.56	90.31	90.20
$Y_{Interactive}$ ＋DBN＋Stacking	91.67	92.55	86.64
Y_{Self} ＋$Y_{Interactive}$ ＋Stacking	94.21	92.47	94.82
Y_{Self} ＋$Y_{Interactive}$ ＋DBN＋Stacking	**94.65**	**93.24**	**95.62**

对 AD 分类结果的影响。

从表 4 可知，实验结果充分证明了交互注意力融合网络 $Y_{Interactive}$ 在多模态信息交互中的关键作用，相比于单独使用独立自注意力视觉变换器，$Y_{Interactive}$ 能够更有效地利用 sMRI 和 FDG-PET 之间的互补信息，提高分类性能。同时，DBN 在深度特征融合方面的重要性也得到了验证，其能够增强不同模态特征的联合表示能力，进一步提升分类效果。此外，完整模型在 AD/CN 识别任务上表现最佳，表明本研究提出的总体架构具有优势。

为了验证本文提出的集成分类器的有效性，比较不同分类器在 AD/CN 识别任务中的性能，本研究设计了不同分类器的对比实验，结果如表 5 所示。包括传统机器学习分类器 SVM、RF、XGBoost、神经网络分类器 MLP，以及基于集成学习的 Stacking 方法。

表 5　不同分类器对比实验结果

Table 5　Comparative experimental results of different classifiers　%

分类器种类	AD/CN		
	ACC	SEN	SPE
SVM	88.28	87.38	85.37
RF	87.04	85.10	88.53
XGBoost	90.61	89.45	90.46
MLP	84.11	84.80	90.38
Stacking1(SVM＋RF＋XGBoost→逻辑回归)	93.51	92.69	95.02
Stacking2(SVM＋RF＋XGBoost→MLP)	**94.65**	**93.24**	**95.62**

在单一分类器的实验中，SVM 和 RF 在分类任务中性能较为接近，ACC 分别达到了 88.28% 和 87.04%，而 MLP 的表现相对较低，它在实际任务中波动大，ACC 为 84.11%。相比之下，单个分类器 XGBoost，在捕捉数据方面具有一定优势，ACC 提升至 90.61%。结合三种单一分类器的优势组成集成分类器的第一层，鉴于 MLP 在建模非线性关系方面具有显著优势，尝试将它放在最后一层，在集成学习方法中，本研究测试了两种 Stacking 方案，Stacking1 将多个基础分类器的输出输入逻辑回归进行最终决策，Stacking2 即使用多层感知机作为最后一层。实验结果表明，Stacking 方法能够有效整合不同分类器的优势，其中 Stacking2 达到了最佳性能，优于所有单一分类器和其他集成方法。这一实验结果表明，基于 Stacking 的集成学习方法能够充分利用多个分类器的互补特性，提高模型的泛化能力，尤其是使用 MLP 作为最终分类器时，能够进一步提升分类性能。因此，在本研究中最终选择 Stacking(SVM＋RF＋XGBoost→MLP)作为最终分类模型。

4　结　　论

磁共振成像和正电子发射断层扫描都是广泛用于早期诊断阿尔茨海默病的成像方式。本文提出了一种基于 sMRI 和 FDG-PET 的新型多模态融合模型，用于 AD 的早期诊断。通过自注意力视觉变换器和交互注意力融合网络，有效地学习多模态数据。自注意力视觉变换器从单一模态中提取不同的特征，本研究提出的交互注意力融合网络专注于多模态之间的相似性，旨在捕捉它们在疾病特定依赖性上的关联。实验结果表明，本文提出的方法具有一定的研究价值和创新性。在一定程度上取得了成效，但也存在不足之处。本文只利用了 sMRI 和 FDG-PET 影像，未能融合更多模态或其他关键信息，未来的研究会进一步结合其他类型的非影像数据来提升 AD 的分类精度。考虑到 AD 具有潜伏性，而且是长期渐进发展的疾病，病理特征会伴随时间产生变化，因此获取患者在不同疾病阶段的数据再进行全面分析，可能会为模型的进一步优化提供新的视角，帮助研究人员更全面地了解其发展过程。鉴于上述情况将在未来开展更多的研究对 AD 疾病分类做出贡献。

参考文献

[1] ZHANG Y, SUN K, LIU Y, et al. Transformer-based multimodal fusion for early diagnosis of Alzheimer′s disease using structural MRI and PET[C]. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). IEEE, 2023: 1-5.

[2] 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.

[3] KLÖPPEL S, STONNINGTON C M, CHU C, et al. automatic classification of mr scans in alzheimer′s disease[J]. Brain, 2008, 131(3): 681-689.

[4] 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: 1-11.

[5] FAN Y, RAO H, HURT H, et al. Multivariate

examination of brain abnormality using both structural and functional MRI[J]. NeuroImage, 2007, 36(4): 1189-1199.

[6] ZHOU T, THUNG K H, ZHU X, et al. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis[J]. Human Brain Mapping, 2019, 40(3): 1001-1016.

[7] ZHANG D, WANG Y, ZHOU L, et al. Multimodal classification of Alzheimer's disease and mild cognitive impairment[J]. NeuroImage, 2011, 55(3): 856-867.

[8] 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.

[9] 张昀枭,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.

[10] 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. Springer, Cham, 2022: 66-76.

[11] 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.

[12] 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.

[13] 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.

[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] LU D, POPURI K, DING G W, et al. Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images[J]. Scientific Reports, 2018, 8(1): 5697.

[16] LIN W, LIN W, 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.

[17] SONG J, ZHENG J, LI P, et al. An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis[J]. Frontiers in Digital Health, 2021, 3: 637386.

[18] LIU M, CHENG D, WANG K, et al. Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis[J]. Neuroinformatics, 2018, 16: 295-308.

[19] FENG C, ELAZAB A, YANG P, et al. Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM[J]. IEEE Access, 2019, 7: 63605-63618.

[20] HUANG Y, XU J, ZHOU Y, et al. Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network[J]. Frontiers in Neuroscience, 2019, 13: 509.

[21] LI W, ZHANG Y, WANG G, et al. DFENet: A dual-branch feature enhanced network integrating transformers and convolutional feature learning for multimodal medical image fusion[J]. Biomedical Signal Processing and Control, 2023, 80: 104402.

[22] ZHANG Y, HE X, LIU Y, et al. An end-to-end multimodal 3D CNN framework with multi-level features for the prediction of mild cognitive impairment[J]. Knowledge-Based Systems, 2023, 281: 111064.

[23] GAO X, SHI F, SHEN D, et al. Multimodal transformer network for incomplete image generation and diagnosis of Alzheimer's disease[J]. Computerized Medical Imaging and Graphics, 2023, 110: 102303.

[24] MIAO S, XU Q, LI W, et al. MMTFN: Multi-modal multi-scale transformer fusion network for Alzheimer's disease diagnosis[J]. International Journal of Imaging Systems and Technology, 2024, 34(1): e22970.

[25] TANG Y, XIONG X, TONG G, et al. Multimodal

diagnosis model of Alzheimer′s disease based on improved Transformer[J]. BioMedical Engineering OnLine, 2024, 23(1): 8.

[26] ZHANG Y, SUN K, LIU Y, et al. Transformer-based multimodal fusion for early diagnosis of Alzheimer′s disease using structural MRI and PET[C]. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). IEEE, 2023: 1-5.

[27] JUN W, TIANLIANG Z, JIAHUI Z, et al. Hierarchical multiples self-attention mechanism for multi-modal analysis[J]. Multimedia Systems, 2023, 29(6): 3599-3608.

[28] AZAM M, ALI R, MUQADDAS R, et al. Predictive modelling of diabetes using ensemble classifiers: Findings from a large and integrated data set[C]. 2024 International Conference on Engineering and Emerging Technologies(ICEET). IEEE, 2024: 1-6.

[29] 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.

[30] ZHANG D, WANG Y, ZHOU L, et al. Multimodal classification of Alzheimer′s disease and mild cognitive impairment[J]. Neuroimage, 2011, 55(3): 856-867.

[31] PAN X, PHAN T L, ADEL M, et al. Multi-view separable pyramid network for AD prediction at MCI stage by 18 F-FDG brain PET imaging[J]. IEEE Transactions on Medical Imaging, 2020, 40(1): 81-92.

[32] HUANG Y, XU J, ZHOU Y, et al. Diagnosis of Alzheimer′s disease via multi-modality 3D convolutional neural network[J]. Frontiers in Neuroscience, 2019, 13: 509.

作者简介

陈洛，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：1870276027@qq.com

王正勇，副教授，硕士生导师，主要研究方向为图像处理与模式识别、计算机视觉。

E-mail：690728634@sina.com

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，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