

面向肝细胞癌微血管侵犯评估的高效多模态 贡献度感知网络研究

贾熹滨^{1,2,3}, 于高远^{1,2,3}, 王璐¹, 邓玉辉⁴, 杨大为⁴, 杨正汉⁴

(1. 北京工业大学信息学部, 北京 100124; 2. 多媒体与智能软件技术北京市重点实验室(北京工业大学), 北京 100124;
3. 北京人工智能研究院(北京工业大学), 北京 100124; 4. 首都医科大学附属北京友谊医院放射科, 北京 100050)

摘要: 微血管侵犯(MicroVascular Invasion, MVI)是肝细胞癌(HepatoCellular Carcinoma, HCC)切除或移植患者出现早期复发和长期预后不良的重要影响因素,因此在HCC患者术前评估是否存在MVI具有非常重要的临床价值.近年来,深度学习为MVI影像诊断评估提供了有价值的解决方法,但受数据标注收集困难等因素的影响,目前研究多独立利用电子计算机断层扫描(Computed Tomography, CT)或核磁共振成像(Magnetic Resonance Imaging, MRI)手段采集影像中的单模态序列,缺乏对各成像手段中多模态序列的综合应用.在小样本场景下,为有效利用多模态序列数据,提高诊断效能,本文提出一种高效多模态贡献度感知网络.该网络可以利用模态分组卷积和高效多模态自适应加权模块,在极少计算开销的引入下,学习CT或MRI的各模态影像信息在复杂多样的MVI表象下的诊断贡献.本文在三甲医院收集的临床数据集上进行实验,结果表明该网络模型可以在少量有标注数据的支持下,取得优于多种基于注意力机制的神经网络模型的MVI诊断性能,为专业医师的诊断分析提供了有效参考.

关键词: 微血管侵犯评估;多模态融合;高效多模态贡献度感知;模态分组卷积;高效多模态自适应加权

基金项目: 国家自然科学基金(No.82071876, No.62171298, No.61871276),北京市医管中心青苗人才项目(No.QML20200108)

中图分类号: TP18

文献标识码: A

文章编号: 0372-2112(2024)06-2053-14

电子学报URL: <http://www.ejournal.org.cn>

DOI:10.12263/DZXB.20220919

Efficient Multimodal Contribution Aware Network for Assessment of Microvascular Invasion in Hepatocellular Carcinoma

JIA Xi-bin^{1,2,3}, YU Gao-yuan^{1,2,3}, WANG Luo¹, DENG Yu-hui⁴, YANG Da-wei⁴, YANG Zheng-han⁴

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;

2. Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing University of Technology, Beijing 100124, China;

3. Beijing Artificial Intelligence Research Institute, Beijing University of Technology, Beijing 100124, China;

4. Department of Radiology, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China)

Abstract: Microvascular invasion (MVI) is an important factor for early recurrence and poor long-term prognosis in patients with hepatocellular carcinoma (HCC) after resection or transplantation. Therefore, it is of great clinical value to evaluate whether MVI exists in patients with HCC before operation. In recent years, deep learning has provided a valuable solution for MVI image diagnosis and evaluation. Nevertheless, due to the difficulties of data annotation and collection, the current researches mostly use computed tomography (CT) or magnetic resonance imaging (MRI) methods to collect single modal sequences in images independently, which lacks the comprehensive application of multimodal sequences in various imaging methods. In order to make more effective use of multimodal data of CT and MRI images and improve diagnosis efficiency under few-shot scenarios, an efficient multimodal contribution aware network is proposed in this paper. The modality grouping convolution and efficient multimodal adaptive weighting module in this network are used to learn the diagnostic contribution of each modal information of CT or MRI under complex and diverse MVI representation with little computational cost introduced. The experiment is carried out on the clinical dataset collected by the third-class hospital. Re-

sult show that with the support of a small amount of labeled data,our method can achieve better MVI diagnostic performance than many deep neural networks based on attention mechanism,which provides an effective reference for professional doctors' diagnostic analysis.

Key words: microvascular invasion evaluation; multimodal fusion; efficient multimodal contribution aware; modality grouping convolution; efficient multimodal adaptive weighting

Foundation Item(s): The National Natural Science Foundation of China (No.82071876, No.62171298, No.61871276); Beijing Medical Management Center Young Talents Project (No.QML20200108)

1 引言

肝细胞癌(HepatoCellular Carcinoma, HCC)是最常见的原发性肝脏恶性肿瘤之一,也是全球第三大肿瘤相关死亡原因^[1,2]. 手术切除和移植目前被认为是肝癌的最佳治疗选择. 然而,由于具有高复发率,肝细胞癌的术后早期和长期预后效果仍然欠佳^[3]. 有研究^[4,5]证实,微血管侵犯(MicroVascular Invasion, MVI)是肝癌切除或移植患者早期复发和长期预后不良的重要因素^[6,7]. 然而,临床MVI主要通过HCC外科切除术后或肝移植术后的病理进行判断,且不同研究者的分析存在一致性较低和主观性较强的问题. 研究人员^[8]发现,即使是经验丰富的放射科医师,在使用核磁共振影像(Magnetic Resonance Imaging, MRI)评估肝细胞癌MVI时,不同医师的评估结果之间仍存在较大差异. 目前仍需一种无创、高度准确的工具来评估HCC患者在进行治疗手术前是否存在MVI.

随着医学成像技术的发展,出现了计算机断层扫描(Computed Tomography, CT)和核磁共振成像(MRI)等多种成像方式,每种成像方式又包含多种模态序列,如动脉期、门静脉期等. 受成像机理、扫描方式和对比剂注入时间等因素的影响,不同模态序列所反映的病理特征各不相同. 例如,T1加权成像(T1-Weighted Image, T1WI)有助于捕获组织的解剖结构,而T2加权成像(T2-Weighted Image, T2WI)则对出血区域较为敏感. 进一步分析临床阅片过程,有经验的临床医师通常会根据待诊断疾病,有偏向性地重点关注几个影像模态序列,并在对比不同模态序列的表象后综合作出诊断决策.

随着人工智能、深度学习等前沿技术的快速发展,一些研究者利用CT或MRI影像构建了基于深度学习的肝细胞癌MVI无创诊断评估方法,并取得初步成果^[9-13]. 然而,现有肝细胞癌MVI智能诊断评估方法多独立利用CT或MRI影像中的各模态序列,缺乏将差异化对待多模态影像并综合判断的临床诊断经验纳入模型的设计. 近年来,随着通道注意力机制^[14]在计算机视觉领域的广泛应用,已有研究工作^[15,16]将通道注意力机制应用于医学影像智能诊断任务中,学习不同特征通道上的重要性权重,提升特征表征能力和网络性能. 但目前压缩激励等主流通道注意力计算机制存在较大

的模型计算开销^[17]. 与此同时,因医学影像数据隐私性和标注要求专业性等原因,收集大规模的标注数据用于训练深度神经网络较为困难. 根据Vapnik-Chervonenkis(VC)维度理论,当网络参数量增加但标注数据不足时,网络易出现过拟合、泛化性能不佳等现象. 因此,如何在不引入过多冗余参数的前提下,关注不同模态影像在不同病变下的重要程度差异,并高效计算各模态影像对最终诊断评估任务的贡献度是一个亟待解决的问题.

本文以临床MVI诊断主要采用的增强CT和多参数MRI影像为处理对象,建模基于这两种扫描手段的诊断模型. 重点针对各扫描手段,研究利用各扫描手段中的多模态影像序列、将医师的模态差异化关注的临床诊断经验引入模型结构的设计,同时降低特征提取网络的计算开销以解决训练样本不足问题. 本文探讨基于组卷积特征提取方法和高效的模态贡献权重计算机制,从模态视角提取支持MVI判别的特征信息并增加重要特征权重、降低冗余信息干扰,并进一步建立性能优异的肝细胞癌MVI评估模型. 本文主要贡献包括:(1)在采用模态分组卷积(Modality Grouping Convolution, MGC)提取CT或MRI多模态序列的本征特征基础上,提出了一种模态内信息交互与跨模态信息交互结合的高效模态注意力计算机制,并基于该机制设计、构建了高效多模态自适应加权(Efficient Multimodal Adaptive Weighting, EMAW)模块,以轻量的计算方式提升多模态特征表征能力.(2)将EMAW模块与主流MVI诊断框架结合,提出了一种高效多模态贡献度感知网络(Efficient Multimodal Contribution Aware Network, EMCANet),在降低模型复杂度的同时提升MVI诊断性能.(3)在CT多模态数据、MRI多模态数据、CT与MRI混合数据3种场景下验证了EMCANet的诊断效能. 通过与现有的主流深度学习网络框架相比,本文的EMCANet在参数体量和模型性能的平衡下达到了最优效果,结果如图1所示,为处理医学影像多模态融合问题提供了一种轻量级的诊断框架.

后续内容将按以下结构安排进行详细阐述. 第2节介绍相关工作的研究现状,包括基于深度学习的肝细胞癌微血管侵犯评估方法、通道注意力机制以及医学影像多模态融合方法. 第3节介绍本文所提出的方法,具

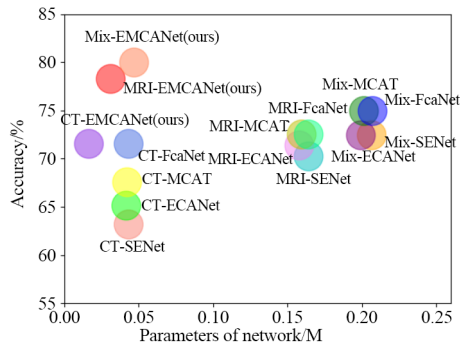


图1 各模型参数体量与MVI诊断性能之间比较

体介绍基于模态分组卷积的特征提取方法,高效多模态自适应加权(EMAW)模块和两种高效模态注意力计算机制以及高效多模态贡献度感知(EMCA)网络的构建.第4节介绍实验设置及结果分析.最后给出研究结论.

2 相关工作

2.1 基于深度学习的肝细胞癌微血管侵犯评估方法

近年来,深度学习方法在人工智能的不同领域取得了令人瞩目的进展^[18-20],已有研究者将深度学习技术应用在肝细胞癌微血管侵犯的评估任务之中. Wei等人^[9]分别使用对比增强计算机断层扫描(Contrast-Enhanced Computed Tomography, CE-CT)和钆塞酸二钠增强核磁共振成像(gadoxetic acid-enhanced Magnetic Resonance Imaging, EOB-MRI)对Resnet18进行了端到端的训练,通过实验比较了两种成像形式对MVI进行术前评估的能力,同时验证了深度学习模型在MVI评估方面的有效性. Zhou等人^[10]利用3个独立的3D-CNN分别提取核磁共振影像平扫期(pre-contrast phase)、动脉期(Arterial Phase, AP)和门静脉期(Portal Vein Phase, PVP)3个期相的模态特征,并将提取到的各模态特征拼接,用于MVI的分类诊断. Zhang等人^[11]利用一个3D-CNN对T2加权成像、T2压脂成像(T2-Spectral Presaturation with Inversion Recovery, T2-SPIR)和动态对比增强成像(Dynamic Contrast-Enhanced, DCE)3种核磁共振影像序列进行了特征提取,但未考虑不同序列模态的融合方法,而是将同一病例的3种序列数据视为3个训练样本,共同训练了一个3D-CNN诊断评估模型. Liu等人^[13]则是利用CT动脉期(Arterial Phase, AP)影像和患者临床因素(Clinical Factors, CF)等参数构建了ResNet18和支持向量机(Support Vector Machine, SVM)模型,并系统比较了两种模型对术前MVI的预测性能. 本文对以上利用深度神经网络进行MVI诊断的研究工作进行了总结,各方法的骨干网络和内部验证

集上的诊断准确率如表1所示.由此可见,结合深度学习的医学影像智能分析技术可以为准确评估HCC的MVI情况提供一个有效的解决方案.

然而,目前对肝细胞癌MVI进行诊断评估的主流深度模型,多利用直接特征拼接^[10]或共享网络参数训练^[11]等形式处理多模态影像信息,对临床多种影像模态序列在诊断中作用的差异性考虑不足.因此,如何更有效地发挥不同模态信息的特点,并提升肝细胞癌MVI诊断评估模型的性能仍有待进一步研究.

表1 基于深度学习的MVI诊断方法诊断准确率

方法	数据量	骨干网络	诊断准确率
CE-CT/EOB-MRI Resnet18 ^[9]	421/444	Resnet18	71.3%/75.7%
多模态3DCNN ^[10]	117	3D CNN	87.50%
多模态3DCNN ^[11]	237	3D CNN	71%
AP+CF Resnet18/SVM ^[13]	309	Resnet18/SVM	72%/60%

2.2 通道注意力机制

注意力机制有助于提升神经网络对感兴趣特征的代表能力,因其方法的通用性和有效性,已被广泛应用于多个计算机视觉领域.通道注意力作为一种实现形式,通过从通道维度对特征信息进行筛选——关注重要通道的特征信息、抑制非重要通道的特征信息,达到提升特征表征质量的目的. Hu等人^[14]提出了一种压缩激励模块(Squeeze-and-Excitation Block, SE Block),其结构先利用全局平均池化对特征进行压缩,再利用一个衰减率为 r 的降维再恢复操作来计算各通道的权重. Woo等人^[21]在压缩激励的计算方式基础上,提出从全局平均池化和最大平均池化两个视角来增加特征聚合信息,从而综合计算通道权重. Qin等人^[22]提出利用频域压缩机制来代替传统的平均池化操作,使得压缩操作考虑更多频域的信息. Wang等人^[17]提出了一种轻量级的高效通道注意力模块(Efficient Channel Attention Block, ECA Block),并指出降维再恢复操作割裂了通道与其对应权值之间的直接关系,不利于通道注意力的计算.此外,已有研究工作^[23]将压缩激励的通道注意力思想引入了医学影像多模态融合领域,提出了一种多模态感知加权模块(Multimodal Adaptive Weighting Module, MAWM),通过计算不同模态特征贡献度,提升了肝细胞癌病理分级无创诊断性能.

综上所述,通道注意力机制作为一种简单有效的方式被广泛集成于医学影像智能诊断的骨干模型之中,而模态注意力机制作为其基础上融合临床经验的衍生产物,已在一些医学任务上取得了初步成果.通道注意力机制和模态注意力机制可以从不同视角关注重要特征信息、抑制冗余特征信息,为高效利用与诊断MVI有关的各模态信息、提升复杂表象MVI的评估性

能提供了基础解决思路。

2.3 医学影像多模态融合方法

医学影像融合一直是影像处理的重要问题,旨在研究通过数据融合、特征融合或决策融合等方式对多模态影像数据进行融合计算,并在病变定性诊断、病灶区域分割等方面取得了重要成果.例如 Guo 等人^[24]将 MRI 影像、CT 影像和 PET 影像(正电子发射断层成像)分别利用3种方式进行融合,并对软组织肉瘤病变进行分割. Zhou 等人^[25]将压缩激励模块(SE Block)和密集跨接模块(Dense Block)结合,对由动态增强 MRI 影像中动脉期、静脉期和延迟期3个模态组成的3D数据建立HCC分级模型.

综上所述,多模态医学影像融合研究作为综合利用多种模态数据、提升模型诊断性能提供了重要的方法支持.但因成像机理不同,影像模态间存在较大差异性,如何在小样本下克服模态间差异,利用模态间关联关系建立有效的融合模型,学习出各模态中支持目标任务的判别特征,仍有待进一步研究.此外,如何有效融合多模态多维度信息,同时避免过多额外计算开销的引入,提升模型面向小样本多模态数据的诊断能力也是一个亟待解决的问题.

3 高效多模态贡献度感知机制的神经网络

针对小样本场景下,多模态影像肝细胞癌微血管侵犯评估深度学习方法,本文提出利用高效的模态注意力机制计算医学影像多模态加权融合特征.首先,为独立提取各模态影像的本征特征并减低网络复杂度,提出利用模态分组卷积代替普通卷积进行特征提取工作.随后,为充分利用模态内的信息约束和模态间的信息约束,提出模态内信息交互和跨模态信息交互两种模态权重计算方式,以此构建了高效多模态自适应加权(EMAW)模块.最后,将EMAW模块与现有的骨干网络结合,建立一种高效多模态贡献度感知网络(EMCANet),并将其用于医学多模态影像分类任务之中.该网络的具体设计细节如下.

3.1 问题描述与网络整体框图

面向肝细胞癌微血管侵犯评估问题,旨建立一种适用于多模态影像序列的网络模型,用于支持临床综合利用医学影像多模态序列进行肝细胞区域是否发生MVI的自动诊断计算.这里,多模态影像序列采用临床MVI评估使用的增强CT序列或多参数MRI平扫和增强的多种序列.考虑不同医院实际诊断手段差异,本文重点探讨采用单一扫描手段情形下多模态扫描序列作为输入的设计.

采用多模态融合数据作为输入,有助于综合不同模态的多视角信息以提升MVI评估性能,然而受限于MVI影像临床采样和数据共享的难度,易出现训练样本有限而导致的特征表征能力差的问题.本文效仿专业医师重点观察并综合应用多种模态影像的临床经验,研究探讨一种高效注意力计算机制,在避免引入额外计算开销的同时,提升特征可判别性.在整体结构设计方面,考虑到 Zhou 等人^[10]提出的多分支特征提取网络结构相对 Zhang 等人^[11]利用同一特征提取网络提取不同模态特征的方法而言,更有利于提取各模态专有特征,从而利用模态间特征的互补性;与之相对,利用同一特征提取网络有利于同时关注多模态重要特征,且避免大规模的特征提取网络参数.因此,结合两者优势,提出在同一特征提取网络内利用模态分组卷积进行多模态融合特征的提取,并利用模态注意力加权计算融合特征,从而保证在提取各模态专有特征并考虑模态差异性的同时避免过多的网络参数体量.整体框架如图2所示.首先利用一个 3×3 的模态分组卷积(MGC)提取浅层的各模态专有特征,随后将浅层特征送入多个高效多模态贡献度感知计算块(Efficient Multimodal Contribution Aware Block, EMCA Block)中,其结构中的两个 3×3 MGC和EMAW模块会计算深层多模态融合特征并为各模态特征加权,随后利用残差结构将EMAW模块输出的深层多模态加权融合特征与原始浅层特征相加,作为下一个EMCA Block的输入,最终将经由多个EMCA Block的输出特征用于MVI的阳性/阴性分

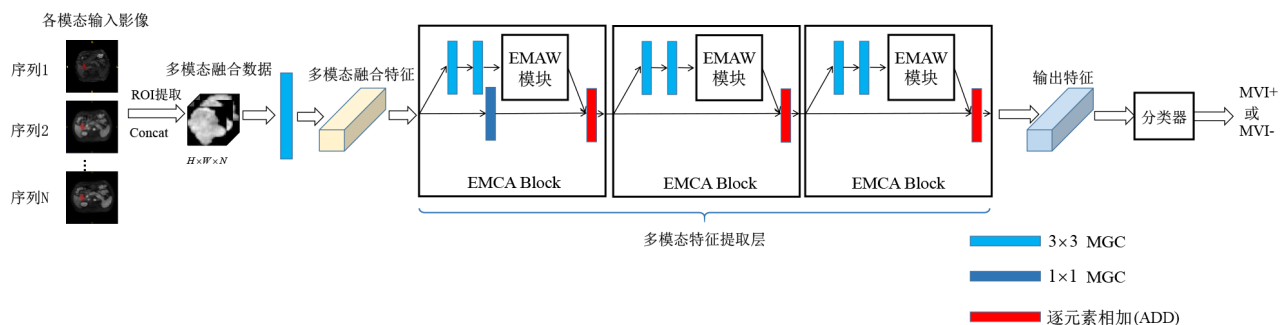


图2 EMCANet结构整体框架(仅展示一层多模态特征提取层)

类评估工作.

3.2 基于模态分组卷积(MGC)的多模态融合特征提取

根据影像医生标注的HCC区域,对多模态影像序列数据横切面扫描层进行裁剪,建立大小为 $H \times W$ 的HCC图像样本,并将不同序列的切面图像 $x_i \in \mathbf{R}^{H \times W}$ 进行拼接,如图3所示,从而得到多模态融合数据 $[x_1, x_2, \dots, x_N]$, $[\]$ 表示通道维度拼接操作, N 表示序列的模态数.

将多模态融合数据送入神经网络,如利用普通卷积操作直接进行特征提取,则各输出通道为加权各模态影像信息后的综合特征.即:给定多模态融合数据记

为 $\mathbf{X} = [x_1, x_2, \dots, x_N]$,利用普通卷积核直接进行特征提取如式(1)所示,最终得到输出特征图 $\mathbf{Y} = [y_1, y_2, \dots, y_C]$.其中 $\text{conv}_{j,i}()$ 表示参与第 j 个输出通道运算的第 i 个卷积核, C 表示输出特征的通道数.

$$y_j = \sum_{i=1}^N \text{conv}_{j,i}(x_i), j \in [1, C] \quad (1)$$

可以发现,输出特征各通道上的信息 y_j 均由输入数据的各模态信息 x_j 与对应卷积核相乘再相加获得.这种全连接卷积计算网络的参数规模较大,在样本有限的情况下不易优化,同时所提取特征缺乏对模态专

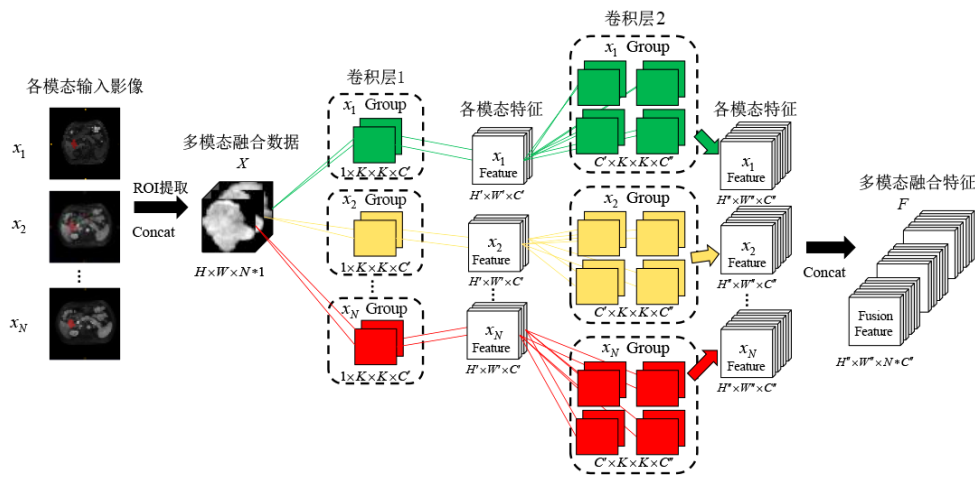


图3 模态分组卷积提取多模态融合特征

为更好提取各模态专有特征,同时降低特征提取模块的参数规模,避免为不同模态单独维护一个特征提取网络带来的成倍参数增长,提出利用以模态数量 N 为分组数的组卷积进行多模态融合特征的提取,具体过程如图3所示,即输入特征沿通道维按顺序分为 N 组,其中 N 为特征包含的模态数,之后每组特征内部进行卷积运算,得到相应的输出特征 $f_{z,j}$.其计算过程如式(2)所示.其中 $f_{z,j}$ 表示第 z 个分组中第 j 个输出通道的特征. $\text{conv}_{z,j,i}()$ 表示第 z 个模态分组中参与第 j 个输出通道运算的第 i 个卷积核. C' 和 C'' 分别表示输入特征的通道数与输出特征的通道数.

$$f_{z,j} = \sum_{i=1+(z-1)C'/N}^{z * C'/N} \text{conv}_{z,j,i}(x_i), z \in [1, N], j \in [1, \frac{C''}{N}] \quad (2)$$

最终各模态特征沿通道维拼接得到多模态融合特征 $\mathbf{F} = [f_{1,1}, \dots, f_{1, \frac{C''}{N}}, f_{2,1}, \dots, f_{N, \frac{C''}{N}}]$.这种模态分组卷积可以在单独提取各模态影像特征的同时降低所需优化的卷积参数数量.例如,利用卷积大小为 K 的传统卷积提取特征所需的参数量为 $C' \times K \times K \times C''$,而利用相同卷

积大小且分组数为 N 的组卷积提取特征所需的参数量为 $(\frac{C'}{N} \times K \times K \times \frac{C''}{N}) \cdot N = (C' \times K \times K \times C'')/N$,即模态分组卷积的方法可避免特征提取阶段各模态的相互干扰.在提取各模态本征特征的同时减少网络的训练参数,在数据较少的情况下降低模型过拟合的风险,改善医学小样本场景下的诊断模型优化性能.

3.3 高效模态自适应加权(EMAW)模块的设计

为建模医师模态差异化观察诊断的经验,采用注意力机制计算支持MVI诊断的可判别特征.受高效通道注意力^[17]的启发,提出利用自适应一维卷积捕获模态交互信息,建立高效的模态注意力计算方法,以提升其计算的效率和有效性.考虑到同模态特征交互信息与不同模态特征交互信息对权重计算的指导作用不同,提出了两种模态注意力计算策略:(1)利用模态内互信息的自注意力机制计算各模态权重.(2)利用跨模态交互信息计算各模态权重,并最终融合两种计算策略所得的各模态权重.

根据所提出的模态内信息交互的高效模态注意力(Efficient Multimodal Adaptive Weighting by Intra-Modal

information interaction, EMAW-IM)与跨模态信息交互的高效模态注意力(Efficient Multimodal Adaptive Weighting by Cross-Modal information interaction, EMAW-CM)两种计算方式,设计了一种高效多模态自适应加权模块(EMAW module),其整体结构如图4所示.首先将由多层模态分组卷积提取到的初步多模态融合特征输入,EMAW模块会利用模态内信息交互机制(EMAW-IM)和跨模态信息交互机制(EMAW-CM)分别计算模态注意力权重.EMAW-IM会利用密集一维卷积构建同一模态的不同特征与当前模态权重的映射关系,利用每个模态内的交互信息指导当前模态权重的计算.而EMAW-CM则会利用跨模态交互的一维空洞卷积提取当前模态特征和其他模态特征的局部长距离依赖关系,从而在考虑不同模态特征信息的前提下,综合计算当前模态的贡献权重.两种模态注意力计算过程如下.

(1)模态内信息交互的高效模态注意力各影像模态刻画不同诊断视角下的病理特征,而同一模态影像在同一参数设置下成像,产生的特征图具有模态表征的一致性,故提出将同一模态不同通道的特征信息利用自适应卷积核大小 K 和滑动步长 S 的一维

卷积进行线性组合,如图4所示.随后将所有模态的权重参数进行归一化,得到各模态对最终诊断任务的贡献程度,即模态内交互权重.具体计算过程如式(3)所示.

$$\omega = \sigma(\text{C1D}_{K,S}(\text{GAP}(F))) \quad (3)$$

其中, $\text{GAP}()$ 表示通道维度的全局平均池化操作, $\text{C1D}_{K,S}()$ 表示自适应卷积核大小 K 和滑动步长 S 的一维卷积,自适应卷积核的大小 K 和滑动步长 S 均为 $(N \cdot C')/N = C'$, N 表示当前特征所含的模态数, C' 表示每种模态所含的特征通道数,其值会随着网络深度的加深而逐渐扩大, σ 表示Sigmoid函数.

(2)跨模态信息交互的高效模态注意力因不同成像模式下各模态产生的特征图具有各异性和互补性,临床诊断过程中专业医师会综合观察各模态表象后,再重点关注重要程度更高的模态序列,故提出利用当前模态特征信息及其邻接模态特征信息,共同计算当前模态的权重,如图4所示.本文采用给定卷积核大小 K 和自适应空洞率 D 的一维卷积来实现模态注意力的计算,以高效的方式完成跨模态信息的交互过程,最终得到跨模态交互权重.具体计算过程如式(4)所示.

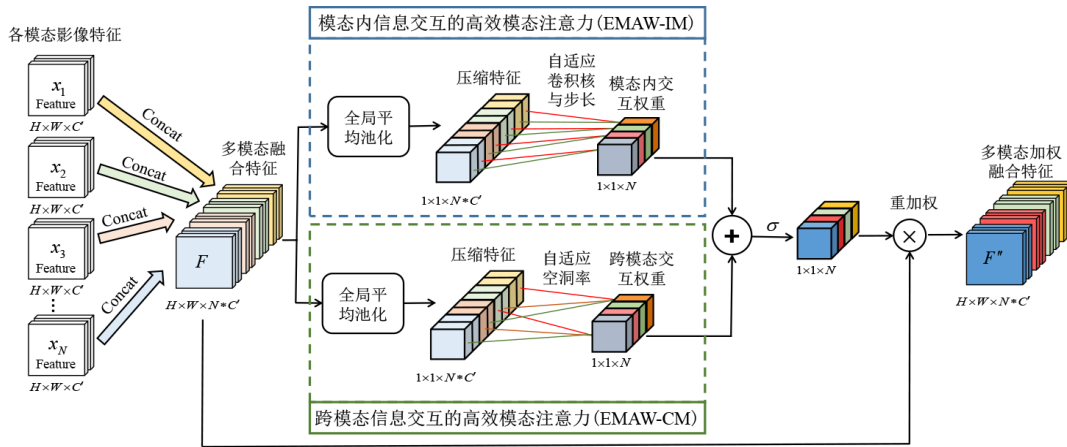


图4 高效多模态自适应加权模块(EMAW module)结构示意图

$$\omega = \sigma(\text{C1D}_{K,D}(\text{GAP}(F))) \quad (4)$$

其中, $\text{GAP}()$ 表示通道维度的全局平均池化操作, $\text{C1D}_{K,D}()$ 表示卷积核大小为 K 且自适应空洞率为 D 的一维卷积,自适应空洞率 D 的计算过程如式(5)所示,推导原理如附录A所示, σ 表示Sigmoid函数.

$$D = \left\lceil \frac{N \cdot C' - N}{K - 1} \right\rceil + 1 \quad (5)$$

对以上两种高效模态注意力机制的计算效能进行评估,结果如表2所示.通过对比可以发现,本文提出的两种高效模态注意力计算方法所需维护的计算参数均要小于MAWM模态注意力^[23],且随着输入数据模态数 N 增多,这种计算轻量化的优势会越来越明显.同时随着

网络层数加深和特征编码通道数 C ($C = N \cdot C'$)的增长,本文的高效模态注意力始终只需维护 N 个注意力权重,相较于SE block^[14]和ECA block^[17]在每层均计算 C 个通道注意力权重的方法更加高效,从计算复杂度和维护权重两个角度避免过多冗余参数导致的网络性能不佳的问题.

将两种模态注意力机制得到的权重相加并进行归一化操作,从而得到最终各模态权重.最终,对原始的多模态融合特征 F 沿通道维度按所计算的模态权重进行重加权,得到最终的多模态加权融合特征 F'' ,为后续的诊断评估工作提供基础,EMAW模块的计算过程如式(6)所示.

表 2 各注意力模块的计算开销

注意力类型	方法名称	计算方式	计算复杂度($N \ll C$)	需计算权值的数量	模块轻量化
通道注意力	SE block ^[14]	$\sigma(W_2 \text{Relu}(W_1 y))$	$2 \times C^2/r$	C	×
	ECA block ^[17]	$\sigma(\text{CID}_K(y))$	K	C	√
模态注意力	MAWM ^[23]	$\sigma(\text{Relu}(W_1 y + b))$	$(C+1)N$	N	×
	EMAW-IM	$\sigma(\text{CID}_{K,S}(y))$	C/N	N	√
	EMAW-CM	$\sigma(\text{CID}_{K,D}(y))$	K	N	√

$$F'' = F \otimes_m \text{Sigmoid}(\text{CID}_{K,S}(\text{GAP}(F)) + \text{CID}_{K,D}(\text{GAP}(F))) \quad (6)$$

其中, $F \in \mathbf{R}^{H \times W \times N \times C'}$ 表示输入特征, N 表示模态数, C' 表示特征提取阶段后每个模态产生的特征通道数, \otimes_m 表示为每个模态产生的特征赋予对应的模态权重, $F'' \in \mathbf{R}^{H \times W \times N \times C'}$ 表示对原始特征重加权后得到的多模态加权融合特征.

3.4 高效多模态贡献度感知网络

本文将所提出的EMAW模块与主流的医学影像智能诊断骨干网络进行结合,提出了一种高效多模态贡献度感知计算块(Efficient Multimodal Contribution

Aware Block, EMCA Block). 如图 5(a)所示,通过在传统残差基础块(Resnet BasicBlock)中第二个卷积操作之后添加EMAW模块,建立高效多模态贡献度感知计算块(EMCA Block),如图 5(b)所示. 该计算块可以利用计算的多模态加权融合特征增强原始特征信息,综合不同模态信息对最终诊断任务的贡献程度,从而突出更有助于当前诊断任务的模态特征信息、抑制无关模态特征信息,提升特性的可判别性. 同时,考虑特征提取过程中的降维操作,在EMCA Block的残差分支上加入了一个 1×1 卷积,并使第一个 3×3 卷积的stride为2,得到一种含降采样的EMCA残差计算块如图 5(c)所示,缓解网络层数加深带来的计算问题.

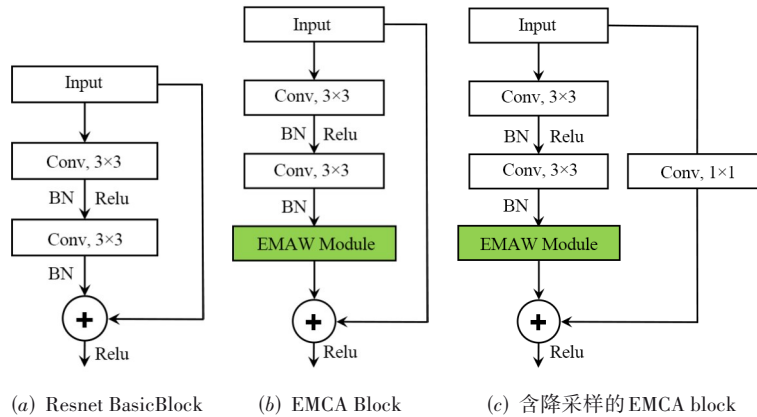


图 5 不同残差计算块的比较

与 Resnet 基础框架结构相似,将所提出的EMCA Block进行堆叠,得到一种高效多模态贡献度感知网络(EMCANet),网络结构如表 3 所示. 该网络由特征编码模块和分类评估模块构成,其中特征编码模块由一个卷积操作和 3 层多模态特征提取层组成,每层的多模态特征提取层包含 3 个 EMCA Block,其中每层中第一个 Block 含有降采样操作;分类评估模块由一个多层感知器组成,即输入提取到的多模态加权融合特征,通过两层的全连接层和 Softmax 函数输出 MVI 阳性/阴性的分类结果,其如式(7)所示.

$$R_j = e^{g(E(x))} / \sum_{i=1}^N e^{g(E(x))} \quad (7)$$

其中, R_j 表示样本 x 属于类别 j 的概率, $E()$ 表示特征提

取网络, $g(\cdot)$ 表示多层感知器, N 表示多层感知器输出向量的维度,即待分类的类别个数,本文设置为 2. 因网络参数设置的低维性和多模态融合的高效性,本文提出的EMCANet适用于少样本多模态医学影像智能诊断任务,该模型可以利用少量参数实现多模态特征的计算,采用轻量化计算的方式缓解了对大规模标注数据的需求和过拟合风险.

4 实验与结果分析

4.1 数据集

由于个人隐私和伦理道德等原因,目前尚无可利用的公开肝细胞癌微血管侵犯数据集,发表成果均在自建数据集上进行. 本研究采用某三甲医院放射科提

表3 EMCANet结构与参数设置

模块名称	网络结构	输出大小	参数设置
特征编码模块		64 × 64	3 × 3 MGC, 16, stride 1, padding 1
	多模态特征提取层1	32 × 32	$\begin{cases} 3 \times 3 \text{ MGC, modality} * 2 \\ 3 \times 3 \text{ MGC, modality} * 2 \\ \text{EMAW, modality} * 2 \end{cases} \times 3$
	多模态特征提取层2	16 × 16	$\begin{cases} 3 \times 3 \text{ MGC, modality} * 4 \\ 3 \times 3 \text{ MGC, modality} * 4 \\ \text{EMAW, modality} * 4 \end{cases} \times 3$
	多模态特征提取层3	8 × 8	$\begin{cases} 3 \times 3 \text{ MGC, modality} * 8 \\ 3 \times 3 \text{ MGC, modality} * 8 \\ \text{EMAW, modality} * 8 \end{cases} \times 3$
		1 × 1	Average pooling, 64-d fc
分类评估模块	分类器	1 × 1	16-d fc
			2-d fc, softmax

供的临床患者腹部CT及MRI影像资料,收集并整理了103位肝细胞癌患者的影像数据,其中包括只做CT检测的患者38例、只做MRI检测的患者27例以及同时做CT和MRI检测的患者38例。

每种检测方式均采用多模态序列/期相影像扫描,CT检测包含在注射造影剂后的平扫期(Plain Scan, PS)、动脉期(Arterial Phase, AP)、门静脉期(Portal Venous Phase, PVP)成像,共计3种动态期相的影像模态;MRI检测数据均包含扩散加权成像(Diffusion Weighted Imaging, DWI)、T1加权成像(T1-Weighted Imaging, T1WI)、T2加权成像(T2-Weighted Imaging, T2WI)以及在静脉注入二乙烯三胺五乙酸后的动脉期(Arterial phase, AP)、门静脉期(Portal Venous Phase, PVP)、平衡期(Equilibrium Phase, EP)成像,共计6种模态的序列/期相影像,本文序列与期相采用相同处理手段,视为同类对象,故统称为序列。其中CT扫描成像中AP模态影像的HCC层厚范围为4~6张,而PS和PVP模态的HCC层厚范围均大于7张;MRI扫描成像中DWI与T2WI模态的HCC层厚范围为3~6张,其余模态的HCC层厚均

大于10张。本文根据以上影像学资料构建了一个多成像多模态的肝细胞癌微血管侵犯数据集。

(1)数据集构建方法:根据专业放射科医师标定的ROI区域,对各病例的最大HCC横切面进行提取,共计得到仅CT影像HCC样本49例、仅MRI影像HCC样本32例、包含两种影像的HCC样本40例。为扩充数据集样本数量且避免无效肝背景区域样本的引入,以CT成像中AP模态与MRI成像中DWI/T2WI模态的层厚为基准,将每个病例中以包含最大HCC横切面为中心的相邻4个扫描切片的二维HCC横切面引入,使样本量扩充为原来5倍。最终得到CT影像HCC样本245例、MRI影像HCC样本160例,同时包含两种影像的混合HCC样本200例。样本提取过程采用统一尺寸64×64对各扫描切片的HCC区域进行裁剪,即输入网络的肝细胞癌样本图像的长(H)和宽(W)均为64。按照4:1比例将三部分数据拆分成各自的训练数据和测试数据,拆分过程以肝细胞癌个体为单位,即防止同一HCC的相邻二维切面同时出现在训练数据和测试数据之中。结果如表4所示。

表4 肝细胞癌微血管侵犯数据集构建情况

检测方式	病例数量	HCC横切面(MVI+ MVI-)	训练数据(MVI+ MVI-)	测试数据(MVI+ MVI-)
CT	38	90 155	70 125	20 30
MRI	27	60 100	45 80	15 20
CT+MRI	38	70 130	55 105	15 25

注:MVI+表示HCC切面为MVI阳性,MVI-表示HCC切面为MVI阴性。

4.2 评价指标与对比方法

为评估模型的诊断性能,本文采用常用的分类评价指标:准确度(Accuracy)、敏感度(Sensitivity)、精准度(Precision)和F1分数(F1-Score),如式(8)~式(11)所示。

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

$$F1 - Score = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \quad (11)$$

TP、TN、FP、FN 表示预测结果中真阳样本、真阴样本、假阳样本、假阴样本的数量。准确率反应的是预测正确的样本占有所有样本的比重；敏感度反应的是真实值为正的样本中，预测正确的比重；精确度反应的预测结果为正的样本中预测正确样本的比重；F1 分数是综合考虑敏感度和精确度的结果，其取值范围为[0, 1]，F1 分数的值越接近于1则表示模型预测性能越好。

将所提出的方法与包含注意力机制的神经网络架构^[14, 17, 22, 23]进行了对比实验，从模型参数量、模型对肝细胞癌 MVI 的诊断评估性能和临床适用性等角度对现有方法和本方法进行总结分析。

4.3 实验设置

4.3.1 实验方法及参数设置

本文实验均基于 Python 以及 Pytorch 0.4.0 框架，在 Resnet 基础模块上进行实验。将实验数据分为三部分，即仅做 CT 成像的 HCC 样本 245 例、仅做 MRI 成像的 HCC 样本 160 例和同时包含 CT 与 MRI 成像的 HCC 样本 200 例，并在三部分数据上分别进行实验。在具体实验中，先对各部分训练数据进行 5 折交叉验证，以验证集上的平均 F1 分数作为衡量指标，从而选取最优的模型超参数。随后本文利用最优超参数和各部分整体训练数据构建 MVI 诊断模型，并在各部分独立测试集上进行测试。以上模型的训练和测试过程重复 5 次，利用 5 次实验各评价指标的平均数值来评判模型的优劣。考虑到训练好的 CT 诊断模型和 MRI 诊断模型可以辅助混合诊断模型的学习，故在混合数据的实验中将各方法已训练过的 CT 诊断模型和 MRI 诊断模型作为混合模型两个分支的预训练模型进行加载，之后再在混合数据上进行训练和测试。

值得注意的是，因本文的 CT 多模态影像的模态数为 3，MRI 多模态影像的模态数为 6。为使 EMAW-CM 实现跨模态信息交互，故将 EMAW-CM 中的给定卷积核的大小设置为 3，填充率设置为 $(K-1)/2$ 。为了公平起见，本文所有对比方法的基础框架均利用参数量较少的 Resnet18 网络进行实现，同时各方法模型每层的特征编码尺寸均设置成一致大小。

4.3.2 损失函数

在仅 CT 数据或仅 MRI 数据的实验中，采用二分类交叉熵损失作为约束，对网络参数进行优化学习，如式(12)所示。

$$L = - \sum_{i=1}^N y_i \log(R_i) \quad (12)$$

其中， L 表示分类评估模块的输出预测结果与样本真实标签之间的交叉熵损失， y 表示样本类别 x 的 one-hot 标

签编码， N 表示类别个数， R_i 表示样本属于类别 i 的概率。

在混合数据的实验中，CT 和 MRI 两个分支的诊断结果用一个可学习比例超参数 α 进行加权融合，得到最终的综合诊断结果 $R_{Final} = \alpha \times R_{CT} + (1 - \alpha) \times R_{MRI}$ ，其中 R_{CT} 和 R_{MRI} 表示 CT 分支和 MRI 分支的输出，超参数 α 初始化为 0.5。本文采用融合决策结果和各自分支的诊断结果与标签真实值 one-hot 编码计算的交叉熵损失进行约束，如式(13)所示。

$$L_{Total} = L_{Fusion} + L_{CT} + L_{MRI} \quad (13)$$

其中， L_{Total} 表示混合网络整体的约束损失， L_{Fusion} 表示混合网络输出的加权结果 R_{Final} 与样本真实标签之间的交叉熵损失， L_{CT} 与 L_{MRI} 分别表示 CT 分支输出结果和 MRI 分支输出结果与真实标签之间的交叉熵损失。

4.4 对比实验

4.4.1 与其他采用注意力机制的网络的对比实验

为验证本文提出的高效多模态贡献度感知网络 (EMCANet) 是否优于传统网络框架，将 EMCANet 与其他利用注意力机制得到的 CT 诊断模型、MRI 诊断模型和混合诊断模型进行了对比，各模型在其对应测试数据上 5 次实验后的平均诊断性能如表 5 所示。

通过实现结果发现，无论是 CT 影像诊断模型、MRI 影像诊断模型还是混合影像诊断模型，本文提出的 EMCANet 均在模型参数量最少的前提下，实现了最优的诊断性能，其诊断准确率分别达到了 71.6%、78.29% 和 80%。其中利用 MRI 影像得到的高效模态贡献度感知网络 (MRI-EMCANet) 性能提升最为明显，其平均诊断准确率相比于 SENet、ECANet、MACT、FcaNet 的 MRI 诊断模型分别提升了 8%、6.86%、5.72% 和 5.72%。这表明了本文所提出的 EMCANet 的有效性，证明了 EMCANet 的设计结构与目前自然图像领域主流注意力网络 SENet、ECANet 和 FcaNet 相比更符合医学多模态小样本场景，在有标注医学影像数据有限的情况下，利用轻量级的计算框架和高效关注不同模态特征对诊断任务贡献的策略提升了模型的诊断性能。而应用于医学多模态数据场景的 MCAT 模型，其卷积和注意力计算方式的参数复杂度较高，在小样本场景下难以获得理想的优化学习性能。此外，由于目前尚没有公开的 MVI 数据集，将得到的模型诊断性能与已公开的 MVI 诊断性能(表 1)直接进行对比。结果发现，由 EMCANet 构建的 3 种诊断模型均取得了有竞争力的应用诊断性能，从可用性角度为所提方法的有效性提供了一定的证明。

4.4.2 消融实验

为评估本文提出的模态分组卷积 (MGC) 和高效多模态自适应加权 (EMAW) 模块中两种计算机制的性

表 5 各模型对肝细胞癌 MVI 的诊断评估性能(测试数据)

模型	网络名称	网络参数量	Accuracy /%	Sensitivity /%	Precision /%	F1 score
CT 诊断模型	SENet ^[14]	42 862	63.20±6.88	63.33±8.15	63.14±8.17	0.63±0.07
	ECANet ^[17]	41 425	65.20±6.27	63.83±5.41	65.25±4.44	0.64±0.07
	MACT ^[23]	41 755	67.60±5.57	66.33±5.91	66.40±5.79	0.68±0.06
	FcaNet ^[22]	42 910	71.60±5.85	69.67±6.00	70.56±6.29	0.71±0.06
	EMCANet	16 030	71.60±6.74	72.50±7.78	71.78±7.60	0.72±0.07
MRI 诊断模型	SENet ^[14]	163 840	70.29±8.59	70.17±9.09	70.60±9.37	0.70±0.09
	ECANet ^[17]	157 987	71.43±6.26	72.17±5.49	72.18±5.55	0.71±0.06
	MACT ^[23]	159 358	72.57±7.14	74.00±6.27	75.29±5.97	0.72±0.08
	FcaNet ^[22]	164 008	72.57±3.43	73.17±2.66	75.38±2.40	0.72±0.04
	EMCANet	30 880	78.29±7.14	77.83±7.39	79.32±6.95	0.78±0.07
混合诊断模型	SENet ^[14]	206 702	72.50±5.00	64.66±7.77	72.35±21.10	0.67±0.10
	ECANet ^[17]	199 412	72.50±5.00	70.00±6.67	70.33±6.00	0.72±0.06
	MACT ^[23]	201 113	75.00±0.00	72.00±2.67	75.81±4.95	0.74±0.02
	FcaNet ^[22]	206 918	75.00±0.00	73.33±0.00	73.33±0.00	0.75±0.00
	EMCANet	46 910	80.00±6.12	77.33±4.90	80.67±8.98	0.80±0.06

能,本实验采用 Resnet18 作为 Baseline 网络,分别在 CT 和 MRI 数据上进行消融实验,对比方法包括 Baseline 模型、仅在 Baseline 上加入模态分组卷积的网络模型、在模态分组卷积基础上分别加入模态内信息交互的高效模态注意力(EMAW-IM)或跨模态信息交互的高效模态注意力(EMAW-CM)的模型,以及本文提出的同时应用模态分组卷积和两种高效模态注意力计算方式的 EMCANet 模型,实验结果如表 6 所示. 其中,在对特征提取网络中卷积层的计算方法进行消融的实验中,Baseline+MGC 表示基准网络中的传统卷积操作全部用 MGC 代替. 实验表明,以模态分组卷积作为医学影像多模态特征提取方法的轻量化网络架构相比于传统卷积具有更优的性能. 同时,加入两种注意力计算机制后的网络均

展示出了优于仅加入 MGC 的网络性能,这是因为 Baseline+MGC 的方法可以视为轻量级的直接特征拼接^[10],而两种注意力机制在其基础上均关注到了不同模态特征的差异性,并利用加权计算的方法关注重要模态的作用,提升多模态特征表征能力. 在两种计算机制中,EMAW-CM 的计算方式展示出了更优的性能指标,从多模态特征互补性的角度出发,该方式相较于 EMAW-IM 更有助于捕获不同模态特征的局部长距离依赖关系,使得邻接辅助模态特征可以为当前模态特征权重的计算提供有效参考. 此外,同时应用两种计算机制的 EMCANet 展示出了最优的性能指标,这表明在模态权重的计算过程中,同时采用模态内部特征和跨模态特征交互指导可以提升多模态 MVI 影像的诊断评估性能.

表 6 高效模态贡献度计算机制消融实验

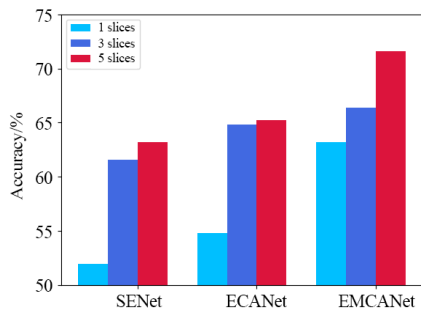
	方法名称	Accuracy /%	Sensitivity /%	Precision /%	F1-score
CT 影像	Baseline	61.60±4.08	58.50±3.63	59.68±4.90	0.61±0.04
	Baseline+MGC	62.80±8.63	60.17±9.30	60.03±10.87	0.61±0.10
	Baseline+MGC+EMCA-IM	65.60±7.74	63.33±8.25	64.05±8.17	0.65±0.08
	Baseline+MGC+EMCA-CM	67.60±7.74	65.50±8.31	68.61±8.06	0.66±0.08
	EMCANet	71.60±6.74	72.50±7.78	71.78±7.60	0.72±0.07
MRI 影像	Baseline	66.86±1.40	68.83±0.67	69.71±0.46	0.66±0.02
	Baseline+MGC	68.00±10.13	67.17±10.44	69.55±11.59	0.67±0.10
	Baseline+MGC+EMCA-IM	73.71±7.96	72.17±8.14	74.78±9.77	0.73±0.08
	Baseline+MGC+EMCA-CM	74.29±8.08	73.33±8.36	74.96±8.52	0.74±0.08
	EMCANet	78.29±7.14	77.83±7.39	79.32±6.95	0.78±0.07

4.4.3 样本量与模型性能对比实验

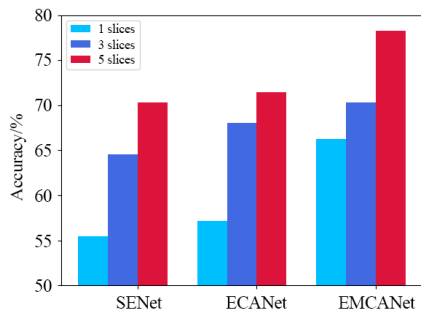
为了评估训练集样本数量对最终模型性能的影响程度,本文在测试集数据和训练策略不变的情况下,分别在不扩充额外样本(1 slices)、额外扩充两个样本

(3 slices)和额外扩充 4 个样本(5 slices)的 3 种场景下,对比了 3 种网络模型训练收敛后在测试数据上的诊断性能(5 次重复实验平均准确率),其中各模型在 3 种不同扩充样本训练数据集上的 5 次重复实验平均准确率

均大于 96%, 在独立测试数据上准确率对比实验结果如图 6 所示. 实验结果显示, 不论对于 CT 诊断模型还是 MRI 诊断模型, 扩充额外的 HCC 区域数据均能一定程度提高模型的诊断性能, 特别对于 SENet 和 ECANet 网络模型, 当额外样本量从零样本扩充至两个样本时, 两种网络结构在对应测试数据上的诊断性能均明显提高, 这说明了引入病灶中心扫描层邻接切片有助于扩充有效样本, 补充了与中心扫描层切片样本具有一定差异性的病灶样本, 在一定程度提升了小样本数据下模型的泛化性. 值得注意的是, 在不扩充额外样本的情况下, 由 SENet 和 ECANet 训练得到的 CT 诊断模型的测试集诊断准确率仅分别为 52.0% 和 54.8%, MRI 诊断模型测试集诊断准确率仅分别为 55.4% 和 57.1%, 均存在过拟合问题. 而本文提出的 EMCANet 在同样不进行额外样本扩充的情况下, 其 CT 诊断模型和 MRI 诊断模型的诊断准确率分别达到了 63.2% 和 66.3%, 相比 SENet 与 ECANet 均有显著的提升. 这也在一定程度上说明, 本文根据医生诊断经验设计的轻量化 EMCANet 框架可以利用更少的网络参数适应医学小样本场景, 在样本量有限的情况下缓解网络过拟合风险, 提高诊断网络在测试数据上泛化性.



(a) CT 数据模型



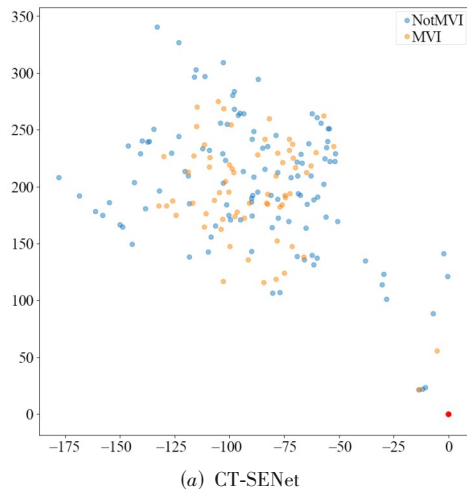
(b) MRI 数据模型

图 6 训练数据量对 CT 诊断模型或 MRI 诊断模型性能(测试数据)影响情况

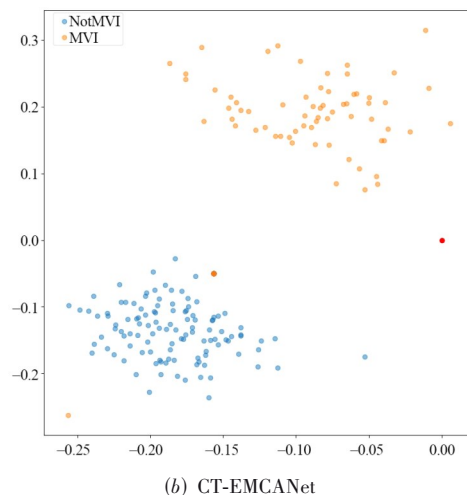
4.5 可视化分析

为进一步探讨高效多模态自适应加权模块的特征表征性能, 在相同实验设置的情况下, 对比了由 SENet、EMCANet 构建的 CT 诊断模型和 MRI 诊断模型特征空间的 T-SNE 图, 如图 7 所示. 图中分别展示了 CT 或 MRI 数据中 MVI 阳性样本和 MVI 阴性样本在训练后的 SENet、EMCANet 特征嵌入空间中的二维映射情况.

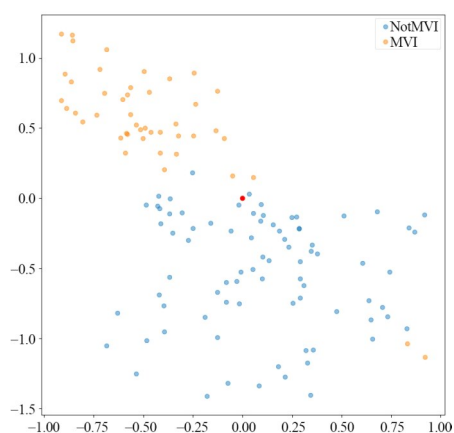
观察发现, 无论 CT 还是 MRI 数据, EMCANet 的特征空间(图 7(b)与图 7(d))相对 SENet 的特征空间(图 7(a)与图 7(c))更具有可判别性, 不同类别样本在训练后得到的特征空间分布上更具可分离性, 相同类别样本在其特征空间分别内聚性更好. 这使得分类器易于找到分类的决策边界, 从而提升分类的准确性. 与之相对的是, SENet 编码得到的特征空间没有显示出明显的规则性. 这说明在高效多模态自适应加权模块的作用下, 网络模型提取到了更具判别力的多模态序列加权融合特征, 为后续的分类评估模块提供了良好的判别基础.



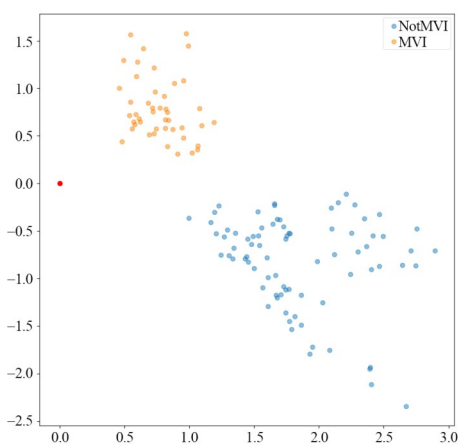
(a) CT-SENet



(b) CT-EMCANet



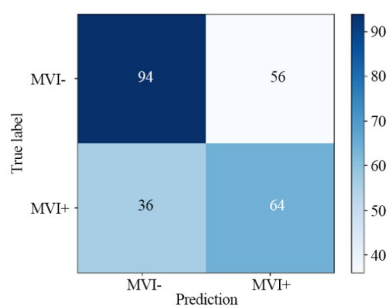
(c) MRI-SENet



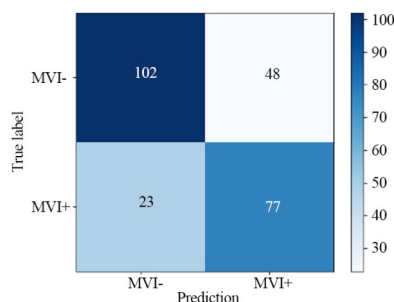
(d) MRI-EMCANet

图7 SENet与EMCANet的CT诊断模型或MRI诊断模型特征空间对比情况

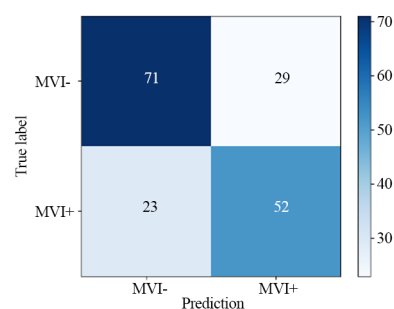
此外,本文对由SENet、EMCANet构建的MVI诊断模型的预测混淆矩阵进行了可视化,如图8所示.观察发现,由EMCANet构建的CT诊断模型相较于SENet构建的CT诊断模型,其误诊率由37.33%下降为32%,漏诊率由36%下降为23%.由EMCANet构建的MRI诊断模型相较于SENet构建的MRI诊断模型,其误诊率由29%下降为19%,漏诊率由30.67%下降为25.33%.这说明了高效多模态贡献度感知网络在多模态MVI诊断



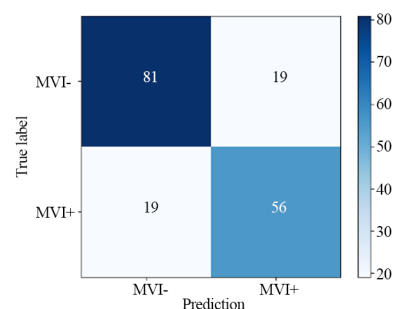
(a) CT-SENet



(b) CT-EMCANet



(c) MRI-SENet



(d) MRI-EMCANet

图8 SENet与EMCANet的CT诊断模型或MRI诊断模型混淆矩阵(5次实验)

问题上具有一定的优越性.

5 结论

本文提出了一种高效多模态贡献度感知网络(EMCANet),该网络参考医生诊断经验,利用模态分组卷积与考虑模态内信息交互和跨模态信息交互的高效多模态自适应加权模块完成多模态加权融合特征的计算.所提出的模态分组卷积提升了对各模态专有特征的代表性能,降低了多模态分支结构参数规模增加带来的模型过拟合风险;在此基础上,利用注意力机制关注各模态显著特征的作用,通过模态内及模态间交互融合提升小样本情景下特征的可判别性能.与其他采用注意力机制模型的对比实验与模块消融实验,验证了EMCANet可以利用较少的模型参数实现性能优异的MVI诊断模型,特别是在小样本情景下利用多模态影像数

据取得了有竞争力的肝细胞癌 MVI 诊断实验结果. 同时本文所提出的模型既适用于 CT 影像又适用于 MRI 影像进行 MVI 的诊断判别工作, 为微血管侵犯评估提供了有效的影像诊断技术支持.

针对医学影像领域样本数量有限且类别不平衡的客观问题, 在下一步研究工作中, 将探究小样本下基于数据知识迁移的元学习方法, 改进现有的肝细胞癌微血管侵犯评估模型. 通过利用其他肝病病变数据构建与目标任务相似的基类任务, 研究包含常见肝脏局灶性病变等多模态数据到 MVI 数据的知识迁移方法, 以基类任务辅助目标 MVI 评估任务的学习, 并重点探究各单模态诊断知识到多模态融合知识的迁移方法, 进一步提升微血管侵犯评估模型的泛化性.

参考文献

- [1] GOLABI P, FAZEL S, OTGONSUREN M, et al. Mortality assessment of patients with hepatocellular carcinoma according to underlying disease and treatment modalities[J]. *Medicine*, 2017, 96(9): e5904.
- [2] EL-SERAG H B. Epidemiology of viral hepatitis and hepatocellular carcinoma[J]. *Gastroenterology*, 2012, 142(6): 1264-1273.
- [3] IGUCHI T, SHIRABE K, AISHIMA S, et al. New pathologic stratification of microvascular invasion in hepatocellular carcinoma: Predicting prognosis after living-donor liver transplantation[J]. *Transplantation*, 2015, 99(6): 1236-1242.
- [4] IMAMURA H, MATSUYAMA Y, TANAKA E, et al. Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy[J]. *Journal of Hepatology*, 2003, 38(2): 200-207.
- [5] CUCCHETTI A, PISCAGLIA F, CATURELLI E, et al. Comparison of recurrence of hepatocellular carcinoma after resection in patients with cirrhosis to its occurrence in a surveilled cirrhotic population[J]. *Annals of Surgical Oncology*, 2009, 16(2): 413-422.
- [6] SUMIE S, KUROMATSU R, OKUDA K, et al. Microvascular invasion in patients with hepatocellular carcinoma and its predictable clinicopathological factors[J]. *Annals of Surgical Oncology*, 2008, 15(5): 1375-1382.
- [7] OISHI K, ITAMOTO T, AMANO H, et al. Clinicopathologic features of poorly differentiated hepatocellular carcinoma[J]. *Journal of Surgical Oncology*, 2007, 95(4): 311-316.
- [8] MIN J H, LEE M W, PARK H S, et al. Interobserver variability and diagnostic performance of gadoteric acid-enhanced MRI for predicting microvascular invasion in hepatocellular carcinoma[J]. *Radiology*, 2020, 297(3): 573-581.
- [9] WEI J W, JIANG H Y, ZENG M S, et al. Prediction of microvascular invasion in hepatocellular carcinoma via deep learning: A multi-center and prospective validation study[J]. *Cancers*, 2021, 13(10): 2368.
- [10] ZHOU W, JIAN W W, CEN X P, et al. Prediction of microvascular invasion of hepatocellular carcinoma based on contrast-enhanced MR and 3D convolutional neural networks[J]. *Frontiers in Oncology*, 2021, 11: 588010.
- [11] ZHANG Y X, LV X F, QIU J L, et al. Deep learning with 3D convolutional neural network for noninvasive prediction of microvascular invasion in hepatocellular carcinoma[J]. *Journal of Magnetic Resonance Imaging*, 2021, 54(1): 134-143.
- [12] HU H T, ZHENG Q, HUANG Y, et al. A non-smooth tumor margin on preoperative imaging assesses microvascular invasion of hepatocellular carcinoma: A systematic review and meta-analysis[J]. *Scientific Reports*, 2017, 7(1): 1-8.
- [13] LIU S C, LAI J, HUANG J Y, et al. Predicting microvascular invasion in hepatocellular carcinoma: A deep learning model validated across hospitals[J]. *Cancer Imaging*, 2021, 21(1): 1-16.
- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [15] ZHANG Y L, LI K, LI K P, et al. MR image super-resolution with squeeze and excitation reasoning attention network[C]//Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 13425-13434.
- [16] CHEN Y, DU W W, DUAN X J, et al. Squeeze-and-excitation convolutional neural network for classification of malignant and benign lung nodules[J]. *Journal of Advances in Information Technology*, 2021, 12(2): 153-158.
- [17] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 11531-11539.
- [18] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2012, 60(6): 84-90.
- [19] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Conference on computer vision and pattern recognition. Piscataway: IEEE,

2016: 770-778.

- [20] YANG D W, JIA X B, XIAO Y J, et al. Noninvasive evaluation of the pathologic grade of hepatocellular carcinoma using MCF-3DCNN: A pilot study[J]. *BioMed Research International*, 2019, 2019(1):1-12.
- [21] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.
- [22] QIN Z, ZHANG P Y, WU F, et al. FcaNet: Frequency channel attention networks[C]//International Conference on Computer Vision. Piscataway: IEEE, 2021:763-772.
- [23] JIA X B, SUN Z, MI Q, et al. A multimodality-contribution-aware triplet for histologic grading of hepatocellular carcinoma[J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021, 19(4): 2003-2016.
- [24] GUO Z, LI X, HUANG H, et al. Deep learning-based image segmentation on multimodal medical imaging[J]. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2019, 3(2): 162-169.
- [25] ZHOU Q, ZHOU Z Y, CHEN C M, et al. Grading of hepatocellular carcinoma using 3D SE-DenseNet in dynamic enhanced MR images[J]. *Computers in Biology and Medicine*, 2019, 107: 47-57.

附录 A

式(5)的推导过程

$$C_{out} = \frac{C_{in} + 2 \times \text{Padding} - K'}{\text{Stride}} + 1$$

$$\text{AND } K' = K + (K - 1) \times (D - 1)$$

Where $K = \text{odd}$, $\text{Padding} = (K - 1) // 2$, $\text{Stride} = 1$ in our

paper

$$\text{So } C_{out} = \frac{C_{in} + (K - 1) - (K + (K - 1) \times (D - 1))}{1} + 1$$

$$\text{So } C_{out} = C_{in} - (K - 1) \times (D - 1)$$

Where $C_{in} = N \times C'$, $C_{out} = N$ in our paper

$$\text{So } D = \left\lceil \frac{N \times C' - N}{K - 1} \right\rceil + 1$$

作者简介



贾熹滨 女, 1969年出生, 山西太原人. 现为北京工业大学信息学部计算机学院正教授. 主要研究方向为视觉信息认知与计算、智能医学图像分析诊断、情感计算.
E-mail: jiaxiabin@bjut.edu.cn



于高远 男, 1997年出生, 河北石家庄人. 现为北京工业大学信息学部计算机学院在读研究生. 主要研究方向为机器学习、医学图像智能诊断、小样本学习.
E-mail: yugaoyuan@emails.bjut.edu.cn



王璐 男, 1990年出生, 北京人. 现为北京工业大学信息学部计算机学院讲师. 主要研究方向为计算机视觉、深度学习、跨模态语义提取.
E-mail: wangluo@bjut.edu.cn



邓玉辉 女, 1985年出生, 湖南郴州人. 现为黑龙江省医院医学影像部副主任医师. 主要研究方向为肝细胞癌微血管侵犯的深度学习.
E-mail: yuhuida@163.com



杨大为 男, 1983年出生, 湖南岳阳人. 现为首都医科大学附属北京友谊医院放射科副主任医师、副教授. 主要研究方向为腹部疾病影像诊断与智能医学影像.
E-mail: yangdawei@cemu.edu.cn



杨正汉 男, 1968年出生, 浙江江山人. 现为首都医科大学附属北京友谊医院放射科主任、正教授. 主要研究方向为临床MRI检查技术及图像判读规范、消化系统疾病影像诊断与智能医学影像.
E-mail: yangzhenghan@vip.163.com