

因果干预与分层特征聚合的弱监督脑肿瘤语义分割

李媛, 郝凡昌

(山东建筑大学 计算机科学与技术学院 山东 济南 250101)

摘要:近年来,全监督学习方法在脑肿瘤语义分割中得到了广泛应用,但逐像素标注十分耗时。为减轻标注负担,基于类激活映射(CAM)的弱监督语义分割(WSSS)方法逐渐受到关注。然而,由于脑肿瘤边界模糊、形态复杂,现有WSSS方法定位精度较低。针对上述问题,提出一种基于因果干预与分层特征聚合的网络(CHFA-Net)。首先,模型提取骨干网络各层特征图,融合脑肿瘤图像的浅层灰度信息与深层语义信息;其次,建立因果图,利用因果干预,分解因果路径以间接估计因果效应,获得因果特征;最后,设计特征聚合模块,以捕捉边界的细微变化,从而提升分割精度。在MSD Brain和BraTS2021数据集上的实验结果表明,CHFA-Net的分割精度明显优于现有WSSS算法。

关键词:弱监督;因果干预;特征聚合;语义分割;脑肿瘤

DOI:10.11907/rjtk.251134

扫描二维码阅读全文:



中图分类号:TP391.41

文献标识码:A

文章编号:1672-7800(2026)006-0202-08

Weakly Supervised Semantic Segmentation of Brain Tumor Combining Causal Interventions and Hierarchical Feature Aggregation

LI Yuan, HAO Fanchang

(School of Computer Science and Technology, Shandong Jianzhu University, Ji'nan 250101, China)

Abstract: In recent years, fully supervised learning methods have been widely adopted for brain tumor semantic segmentation, yet pixel-level annotation remains time-consuming. To alleviate labeling burdens, weakly supervised semantic segmentation (WSSS) approaches based on Class Activation Maps (CAM) have gained increasing attention. However, due to the blurred boundaries and complex morphology of brain tumors, existing WSSS methods still suffer from limited localization accuracy. To address these challenges, this study proposes a Causal Intervention and Hierarchical Feature Aggregation Network (CHFA-Net). First, the model extracts multi-level feature maps from the backbone network, integrating shallow grayscale information and deep semantic features of brain tumor images. Subsequently, a causal graph is constructed to eliminate interference from confounding factors on prediction outcomes through causal intervention. Finally, a feature aggregation module is designed to capture subtle boundary variations, thereby enhancing segmentation precision. Experimental results on the MSD Brain and BraTS2021 datasets demonstrate that CHFA-Net achieves significantly superior segmentation accuracy compared to existing WSSS algorithms.

Key Words: weak supervision; causal intervention; feature aggregation; semantic segmentation; brain tumor

0 引言

脑肿瘤是中枢神经系统的一种严重疾病,随着现代医学影像技术的发展,脑肿瘤的早期发现和治疗变得更加可

行^[1]。磁共振成像(Magnetic Resonance Imaging, MRI)技术是脑肿瘤诊断中最常用的成像方法,能够提供高分辨率的脑部图像。深度学习在医学图像处理中具有重要作用,利用深度学习能够实现自动识别和分割病变或肿瘤区域,从而辅助医生进行诊断和治疗^[2]。

收稿日期:2025-04-22

基金项目:山东省自然科学基金面上项目(ZR2022MF272);山东省重点研发计划公益类项目(2019GCX101068)

作者简介:李媛(1999-),女,CCF学生会会员,山东建筑大学计算机科学与技术学院硕士研究生,研究方向为机器学习与数据挖掘;郝凡昌(1981-),男,博士,山东建筑大学计算机科学与技术学院副教授、硕士生导师,研究方向为机器学习与数据挖掘、算法分析与设计。本文通讯作者:郝凡昌。

传统的脑肿瘤分割方法主要基于图像的浅层特征,如纹理和结构信息。常用的技术包括基于阈值^[3]、区域、边缘^[4]的分割算法。但这些方法采用人工设计的特征,难以挖掘图像的深层信息,易受伪影、噪声等干扰因素影响,分割效果不够理想。

近年来,随着深度学习技术的发展,卷积神经网络(Convolutional Neural Networks, CNN)已广泛应用于医学图像检测、分割和分类等任务,自动化的脑肿瘤分割方法逐渐成为研究重点^[5]。大多数用于分割的深度学习方法依赖于完全标注的数据集,而该标注过程往往非常耗时且成本高昂。为此,弱监督学习逐渐成为医学图像语义分割研究中的一个重要方向。在各种弱监督标签中,图像级标签是最容易获得的,这也是本文唯一使用的标签^[6-7]。

目前,许多使用图像级标签的弱监督语义分割(Weakly Supervised Semantic Segmentation, WSSS)方法都是基于类激活图(Class Activation Map, CAM)的^[8]。针对自然图像, Selvaraju等^[9]利用类别特定的梯度信息为卷积特征图分配权重,通过加权求和生成激活图。但激活图在梯度饱和或消失时,生成的激活图无法准确反映重要区域。Wei等^[10]通过不同CAM的集合来扩展种子区域,生成更完整的激活图。但对于目标尺度差异较大的图像,激活效果不佳。Zhang等^[11]通过对图像进行局部擦除,迫使模型重新学习和发现目标物体的特征,以形成新的激活区域。但对抗擦除生成的互补区域可能包含噪声,尤其是在初始CAM质量较低的情况下,噪声会被放大并传递到后续训练中。Wang等^[12]利用类别预测分数直接对特征图进行加权,从而生成更精确的激活区域,但其需要进行多次前向传播,计算开销较大。Jiang等^[13]结合了多个卷积层的特征来生成更精确的激活图,通过从不同的卷积层提取特征,提供了更细粒度和多尺度的语义信息,以更好地捕捉复杂对象的结构和边界。然而,错误的卷积层选择会导致不准确的激活图。针对医学图像, Pate等^[14]提出利用跨模态等变约束,在多模态图像场景中利用自监督来显著增强原始CAM。然而,其直接对特征图进行平均,使得网络在不同位置上关注相同的特征图值,导致CAM的性能较差。Li等^[15]提出了OEEM(Online Easy Example Mining),通过为损失较低的样本分配更高权重,迫使分割模型从可信的监督信号中进行学习。Prabakaran等^[16]利用无监督语义分割模型进一步增强图像的显著性表达,从而优化CAM预测结果,生成更高质量的训练标签。这两种方法虽然都在一定程度上改善了模糊边界问题,但是无法完全解决分类与分割之间差距的问题。

上述方法都是基于分类网络产生CAM,难以在具有混杂因素(各种复杂组织)的上下文中学习到像素与标签间真实的因果关系。这导致模型容易错误地将无因果关系但正相关的像素与标签关联起来,从而产生错误的激活区域。

由于因果干预考虑因果关系而不是模型的统计学关联关系,因此为深度学习提供了一种新方法^[17]。近年来,结构因果模型在计算机视觉领域得到了广泛关注,已应用于图像分类^[18-19]、视觉语言任务^[20-21]、视频分析^[22-23]、图像识别^[24]以及语义分割^[25]等多个方向。Zhang等^[26]在WSSS中引入后门调整,但由于上下文中的混杂因子无法直接观测到,无法确定其类别,因此只能通过假设对这些混杂因子进行近似估计。Sun等^[27]使用后门调整来学习图像的常识表示,但需要大量图像和丰富的上下文信息作为训练数据,才能有效地结合视觉特征和常识知识,因此不适用于数据匮乏的医学图像。Wang等^[28]使用前门调整以避免假设混杂因子,但是模型过度依赖局部特征和不相关的上下文信息,导致目标过度分割,将背景中的一些噪声或细微变化误认为目标的一部分。

针对上述问题,本文提出一种新的融合因果干预和分层特征聚合的弱监督语义分割网络,旨在提取并融合骨干网络不同层级的特征,并对融合特征进行因果干预,从而捕捉病灶在不同层级网络上边界的细微变化,提高模型的边缘感知能力和细节识别能力。

本文的主要贡献如下:

(1)结合脑肿瘤图像的特点,融合骨架网络的多层级特征,提出对融合后的特征进行因果干预,通过前门调整间接估计图像与类激活图CAM的因果效应,获取鲁棒的因果特征,提高了模型可解释性。

(2)提出一种特征聚合模块(Feature Aggregation, FA),该模块通过注意力模块CA获取通道注意力权重,利用中心损失学习脑肿瘤的形状和大小,从而提高模型对肿瘤边界的判别能力。

(3)在MSD Brain^[30]和BraTS 2021^[31]数据集上进行对比实验,结果验证了本文方法的有效性,取得了较好的分割结果。

1 CHFA-Net

1.1 网络结构

本文提出一种融合因果干预和分层特征聚合的网络(Causal Intervention and Hierarchical Feature Aggregation Network, CHFA-Net)用于2D脑肿瘤图像分割,模型结构如图1所示。模型的输入是经过切片处理的脑肿瘤灰度图像,记为 $I \in \mathbb{R}^{B \times H \times W}$ 。CHFA-Net包括分层特征提取(Hierarchical Feature Extraction, HFE)、因果干预(Causal Intervention, CI)、特征聚合(Feature Aggregation, FA)3个模块。

1.2 分层特征提取

产生高质量的CAM是弱监督医学图像分割的基础。如图2所示,本文构建了一个分类网络,用于融合从骨架网络提取的不同层级的CAM。在分层特征提取(HFE)阶段,本文使用二值分类网络ResNet-18,得到输入图像X的

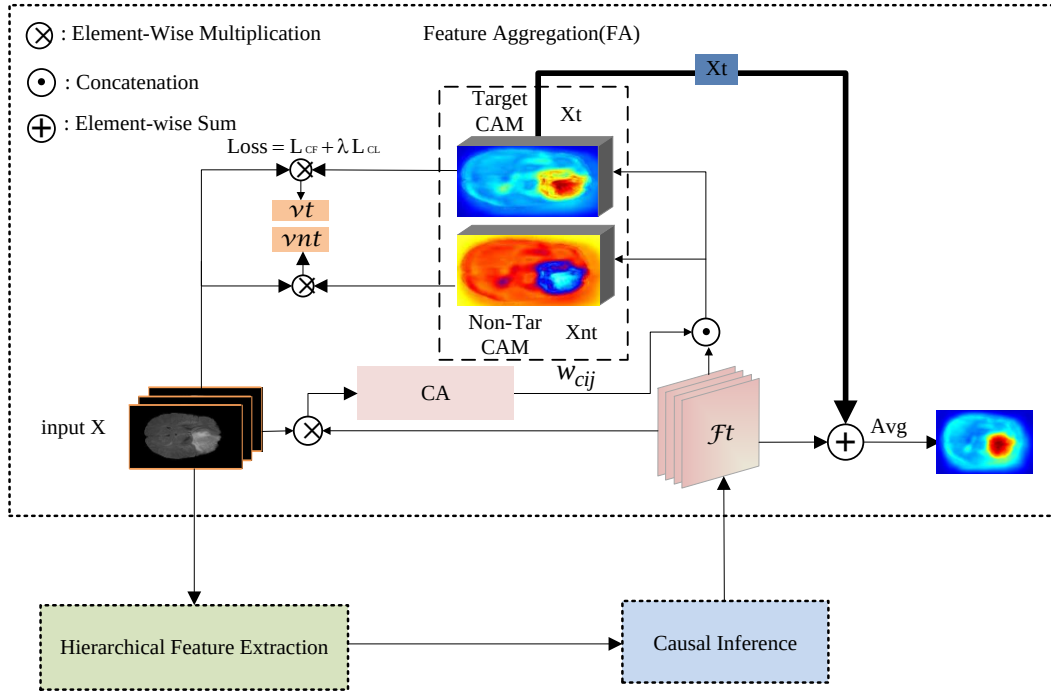


Fig. 1 Structure of CHFA-Net model

图1 CHFA-Net网络模型结构

类概率。

为了提取骨架网络每个层级的特征,每个残差块都被分配一个单独的分类器(Classifier, CLS)用于二值分类,并生成CAM。通过 1×1 的卷积层得到4个层级的特征图 f_i , 分别表示为 f_1, f_2, f_3, f_4 。然后,通过全局平均池化(Global Average Pooling, GAP)函数得到相应的预测分数 S_i 。最后,沿着通道维度拼接4种不同层级的特征图,生成融合的特征图 f 。分类器CLS损失表示为 L_{CF} , 损失定义为:

$$L_{CF} = \sum_i L_{CE}(S_i, G) \quad (1)$$

其中, L_{CE} 为交叉熵损失, S_i 为类的预测值, G 为真实值。

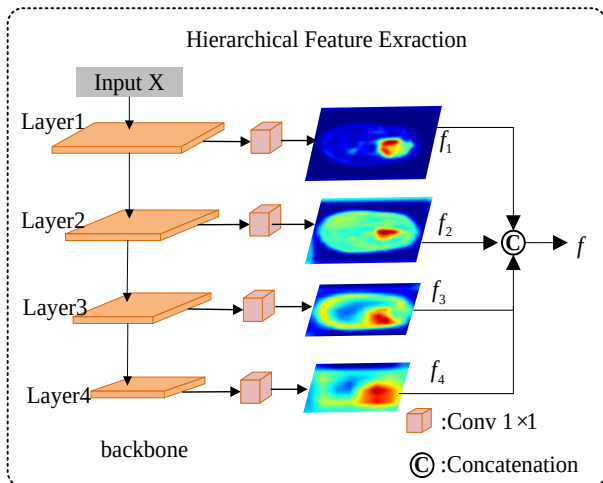


Fig. 2 Hierarchical feature extraction module

图2 分层特征提取模块

1.3 因果干预

由于图像中的混杂因子无法观测,本文采用前门调整,通过控制中介变量T间接估计图像X与类激活图Y间的因果效应,从而避免假设混杂因子。如图3所示,首先绘制了一个因果图,以简化对给定变量间因果关系的理解^[29]。根据前门调整,C是无法观测的混杂因子,影响X和Y; $X \rightarrow T$ 的路径无混杂因子; $T \rightarrow Y$ 的路径存在混杂因子; X只能通过T间接影响Y。弱监督语义分割($X \rightarrow T \rightarrow Y$)通过预测像素X对应的标签T,从而产生像素级定位Y。然而,图像中无法观测的混杂因子C会导致标签预测过程中产生混淆效应。这种效应可被视为是一个后门路径($X \leftarrow C \rightarrow Y$),降低了目标的定位精度。为了阻断路径($X \leftarrow C$),消除C对X的影响,需要控制中介变量T,从而间接估计X对Y的因果效应。即在干预X的条件下求Y的概率分布,表示为 $P(Y|do(X))$, 此因果效应不受C的影响。

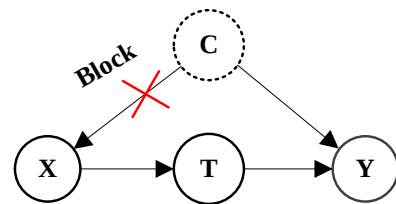


Fig. 3 Causal graph

图3 因果图

由上文内容可知,本文方法是针对二值分类任务定义的,分类器预测标签T可能为肿瘤(tumor)或正常(normal)。 $P(Y|do(X))$ 推导过程如下:

(1) 拆分因果路径。根据全概率公式,将干预 X 后 Y 的分布分解为所有可能的中介变量 $T=t$ 的贡献之和:

$$P(Y|\text{do}(X)) = \sum_t P(Y|\text{do}(X), T=t)P(T=t|\text{do}(X)) \quad (2)$$

(2) 消除干预 $\text{do}(X)$ 。根据前门调整, $X \rightarrow T$ 的路径无混杂, X 直接决定 T 的分布,有:

$$P(T=t|\text{do}(X)) = P(T=t|X) \quad (3)$$

干预 X 后, Y 仅受 T 影响,故:

$$P(Y|\text{do}(X), T=t) = P(Y|\text{do}(T=t)) \quad (4)$$

将式(3)、式(4)的结果代入式(2),则:

$$P(Y|\text{do}(X)) = \sum_t P(T=t|X)P(Y|\text{do}(T=t)) \quad (5)$$

(3) 消除 $T \rightarrow Y$ 路径混杂。 $T \rightarrow Y$ 路径存在后门路径 ($T \leftarrow X \leftarrow C \rightarrow Y$)。为阻断后门路径,需控制变量 X 。根据后门调整公式:

$$P(Y|\text{do}(T=t)) = \sum_{X'} P(Y|T=t, X=X')P(X') \quad (6)$$

将式(6)的结果代入式(5),得:

$$P(Y|\text{do}(X)) = \sum_t P(t|X) \sum_{X'} P(Y|t, X')P(X') \quad (7)$$

其中, $P(t|X)$ 表示图像 X 属于类别 t 的概率,可由分类器 Classifier 计算得到。假设每个训练样本是等概率的, $P(X')$ 服从均匀分布,即每张类别 t 的图像出现的概率为 $\frac{1}{N_t}$ 。 $P(Y|t, X')$ 为类别 t 图像 X 的定位 Y 的概率分布。 $\sum_{X'} P(Y|t, X')P(X')$ 为类别 t 所有图像的定位 Y 的期望值,表示为 $E(Y_t)$,也可由分类器计算得到。 $P(Y|\text{do}(X))$ 的值可以通过式(8)转化为预测分数 S 来训练分类器,使交叉熵损失最小化。

$$S = \text{GAP}(\sum_t P(t|X) \sum_{X'} P(Y|t, X')P(X')) \quad (8)$$

对 HFE 模块得到的特征 f 应用因果干预,分别计算类别 t 为肿瘤和正常的期望,即 $E(Y_{\text{tumor}})$ 、 $E(Y_{\text{normal}})$,再计算期望值与类别概率 $P(t|X)$ 的乘积和,即 $P(\text{tumor}|X) \times E(Y_{\text{tumor}}) + P(\text{normal}|X) \times E(Y_{\text{normal}})$,从而得到因果特征图 F_t 。

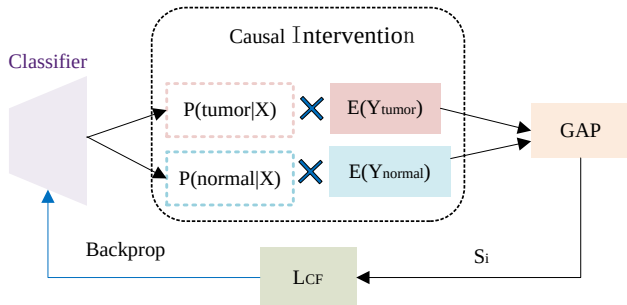


Fig. 4 Causal intervention module

图4 因果干预模块

1.4 特征聚合

在特征聚合模块中,对输入的灰度图像 $\text{input } X$ 与因果特征图 F_t 进行逐元素相乘,如图1所示。这把模型的注意力集中在具有较高响应的肿瘤区域,从而抑制无关信息。此过程类似于空间注意力机制,将生成的结果输入到 (Channel Attention, CA) 模块。

CA 模块是由多个 (Feature Extraction, FE) 块组成的序列模块,每个 FE 块由一个 3×3 卷积层、批归一化层和 ReLU 激活函数组成,如图5所示。CA 模块利用多次卷积进一步获取重要特征,最后通过 Softmax 函数在通道维度上生成注意力权重 w_{cij} 。

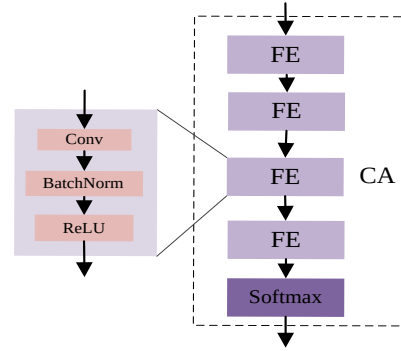


Fig. 5 CA module

图5 CA模块

之后,将由 CA 模块生成的通道注意力权重 w_{cij} 和特征图 F_t 在通道维度上进行逐元素相乘,并对乘积进行求和,得到最终的注意力图 Target CAM X_t 和 Non-Tar CAM X_{nt} 。其中, X_t 为肿瘤激活图, X_{nt} 为正常区域激活图。 X_t 、 X_{nt} 是对特征图 F_t 加权求和的结果,代表了模型对输入图像各个像素的重要性判断。接下来将 X_t 、 X_{nt} 与输入的灰度图像 $\text{Input } X$ 进行加权,得到目标区域特征 v_t (肿瘤区域) 与非目标区域特征 v_{nt} (正常区域)。为了增强模型对目标区域与非目标区域的识别能力,本文使用中心损失 L_{CL} 结合交叉熵损失 L_{CLS} 将同一类别的特征向量聚集到类别中心,使得类内特征更加紧凑,降低了类内的方差,从而间接提升了类间可分性。损失函数定义如下:

$$\text{Loss} = L_{CLS} + \lambda L_{CL} \quad (9)$$

$$L_{CL} = \frac{1}{2} \|v_i - c_i\|^2 \quad (10)$$

其中, L_{CLS} 是 HFE 模块的交叉熵分类损失。 λ 是一个超参数,用于权衡交叉熵分类损失和中心损失的影响。对于 L_{CL} , v_i 是样本 i 的特征, c_i 是样本 i 所属类别的中心向量。

2 实验结果与分析

2.1 数据集

本文使用 MSD Brain 和 BraTS 2021 数据集进行实验。MSD Brain 数据集包含 484 例训练样本和 266 例测试样本,包括胶质母细胞瘤 (GBM) 及低级别胶质瘤 (LGG) 患者的

磁共振图像。每个病例包含 4 种不同 MRI 模态的序列: T1 加权序列 (T1w)、对比增强后的 T1 加权序列 (T1gd)、原生 T2 加权序列 (T2w) 和 T2 液体衰减反转恢复序列 (flair)。本文实验将官方训练集按照 8:1:1 的比例划分为训练集、验证集和测试集,然后沿着 z 轴对每个序列进行切片,总共形成 75 020 张二维图像。

BraTS 2021 数据集是一个大规模的脑部多模态 MR 脑胶质瘤分割数据集,包含 1 251 例训练样本、219 例验证样本以及 530 例测试样本,每个病例分别包括: T1 加权序列 (T1)、T2 加权序列 (T2)、T1 加权增强序列 (T1ce) 和 FLAIR 序列 (flair)。本文实验使用官方验证集作为测试集,并沿着 z 轴对每个序列进行切片,共形成 193 905 张二维图像。

2.2 评价指标

实验采用 3 个广泛使用的指标评估模型性能,分别为 Dice、IoU 和 Hausdorff95。其中, Dice 和 IoU 的目的是评估预测分割结果与真实分割结果之间的相似程度,该数值越大说明模型分割效果越好。Hausdorff95 用于衡量两个集合(分割结果与真实标签)之间最远的边界距离,该数值越小越好。Dice、IoU 和 Hausdorff95 分别被定义为:

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \quad (11)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (12)$$

$$\text{Hausdorff95} = \max_{p \in P} (95\%) \max_{p \in P} \|p - t\|, \max_{t \in T} (95\%) \max_{p \in P} \|t - p\| \quad (13)$$

其中, TP、FP、FN 分别为真阳性、假阳性、假阴性; t 表示预测分割的边界体素, p 表示真实分割的边界体素。实验结果采用平均值来概括评价指标的分布。

2.3 实验参数

CHFA-Net 模型采用 Pytorch 框架实现,使用 ResNet-

18 作为主干分类器,数据集中所有图像的分辨率为 240×240。基于经验标准,排除了数据集中实际区域像素少于 50 的切片,同时确保来自同一患者的切片只出现在训练、验证或测试集中。CHFA-Net 仅使用图像类别标签作为监督信号进行训练,所有指标都是在测试集上获得的。CHFA-Net 在 Nvidia RTX 2080ti 11gb 上使用 Adam 优化器进行训练,其中权重衰减设置为 0.000 1,初始学习率设置为 0.000 5,使用余弦退火算法逐步调整学习率, batchsize 设置为 256。将训练数据归一化后,对模型进行 100 次训练。

2.4 对比实验

将本文提出的 CHFA-Net 与 4 种先进的弱监督分割方法 ScoreCAM^[12]、LayerCAM^[13]、C²AM^[28] 和 CMER^[14] 的分割性能进行比较,同时与全监督 U-net^[32] 进行比较,该方法可为各类 WSSS 方法提供性能上限参考。在 MSD Brain 和 BraTS2021 数据集上进行对比实验,结果如表 1、表 2 所示。

在 MSD Brain 数据集上, CHFA-Net 在 3 个指标上均取得了最佳结果。其中, Dice 分别达到了 60.90%、62.72%、74.89%、78.80%。表 1 中的数据表明, CHFA-Net 在 T2w 和 flair 模态上表现出较大优势。具体来看,性能最好的基线方法是 CMER, 本文方法的平均 Dice 与其相比提高了 10% 以上。ScoreCAM 的 Dice 系数最低,表明其难以获取准确的类激活图。C²AM 和 CMER 性能接近,精度略低于 CMER。在 BraTS2021 数据集上, CHFA-Net 在 T1 模态上取得了次优结果, Dice 比 CMER 低 2.75%,但在剩余 3 个模态上取得了最优结果。

本文方法优异的性能归功于 CHFA-Net 中的分层特征提取、因果干预和特征聚合模块,这些模块在关注脑部整体形态特征的同时,也能够兼顾局部的病变肿瘤,显示出 CHFA-Net 在学习高级语义特征和底层细节方面的强大表

Table 1 Result comparison of different methods on MSD Brain dataset

表 1 MSD Brain 数据集上不同方法结果比较

方法	Dice/%				IoU/%				HD95/mm			
	T1w	T1gd	T2w	flair	T1w	T1gd	T2w	flair	T1w	T1gd	T2w	flair
ScoreCAM	25.43	30.28	38.85	45.47	14.57	17.84	24.11	29.42	71.35	59.89	44.76	38.85
LayerCAM	50.35	45.23	55.05	56.45	33.65	29.22	37.98	39.32	48.49	39.01	29.96	27.16
C ² AM	45.67	57.74	59.42	58.76	29.59	33.10	42.27	41.60	38.72	49.24	23.86	25.36
CMER	50.89	54.38	63.83	60.82	34.13	37.34	46.88	43.70	44.57	30.07	21.43	23.48
Ours	60.90	62.72	74.89	78.80	43.78	45.69	59.86	65.02	23.66	25.25	14.74	11.96
U-net	68.76	73.95	86.57	88.58	55.96	58.67	76.32	79.50	19.05	15.32	8.56	8.69

Table 2 Result comparison of different methods on BraTS 2021 dataset

表 2 BraTS 2021 数据集上不同方法结果比较

方法	Dice/%				IoU/%				HD95/mm			
	T1	T1ce	T2	flair	T1	T1ce	T2	flair	T1	T1ce	T2	flair
ScoreCAM	27.15	42.08	46.73	53.15	15.71	26.65	30.49	36.19	64.78	51.08	42.32	33.25
LayerCAM	56.43	54.53	64.52	65.08	39.30	37.49	47.62	48.24	25.35	30.03	25.08	24.78
C ² AM	51.27	59.85	58.76	68.53	34.47	42.70	41.60	52.13	29.66	35.67	20.52	21.23
CMER	61.33	65.79	60.82	67.65	44.23	49.02	43.70	51.11	22.23	28.07	18.55	17.48
Ours	58.58	68.77	84.58	75.63	41.42	52.40	73.28	60.81	23.91	20.12	11.53	6.96
U-net	81.54	82.93	93.14	88.29	68.83	70.84	87.16	79.04	13.16	10.30	6.56	4.69

征能力,这在医学图像分割中是至关重要的。最后,在所有实验中完全监督的U-net获得了最高的平均Dice和IoU值以及最低的Hausdorff95值,表明WSSS方法与完全监督的方法之间仍然存在性能差距,WSSS方法在未来仍有改进的空间。性能差距的核心在于WSSS方法缺乏精确的空间监督信号,导致模型难以准确学习肿瘤的细节与边界。此外,WSSS通常需进行多阶段训练,流程复杂,伪标签的误差会在训练中积累,影响后续分割模型的优化。

图6、图7分别显示了不同分割方法在MSD Brain数据集和BraTS2021数据集上分割结果的视觉对比。Score-

CAM会产生过大的激活区域,掩码质量最低。LayerCAM定位能力良好,但激活区域较小,难以完整覆盖肿瘤区域。CMER和C²AM的定位能力接近,CMER的覆盖准确率略高于C²AM。图6第3行的分割结果显示,当图像前景与背景对比度低且边缘模糊时,所有方法的分割形状均不准确,但本文方法定位最准确,且覆盖了最大的肿瘤区域。图中第3列展示了本文方法的分割结果。与其他弱监督方法相比,CHFA-Net分割掩码与真值掩码最为接近,取得了最准确的分割结果,这归因于模型强大的定位能力和边缘感知能力。

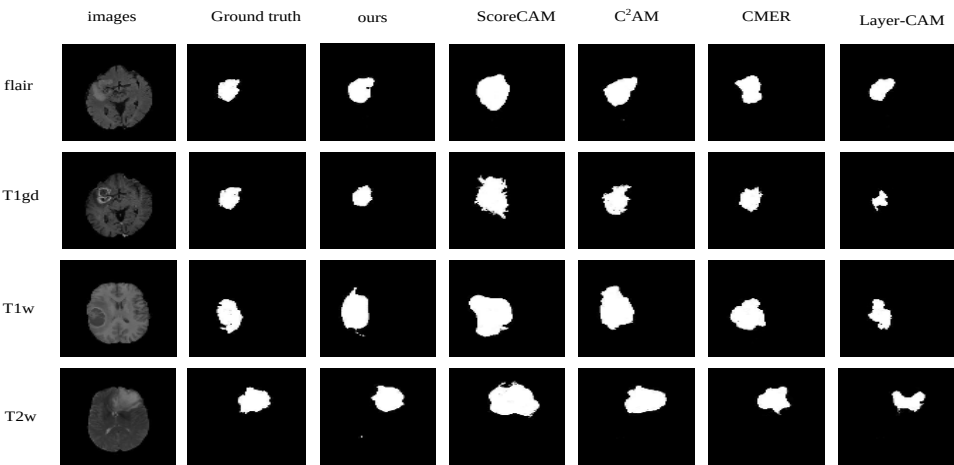


Fig. 6 Visual comparison of segmentation results on the MSD Brain dataset
图6 MSD Brain数据集分割结果视觉对比

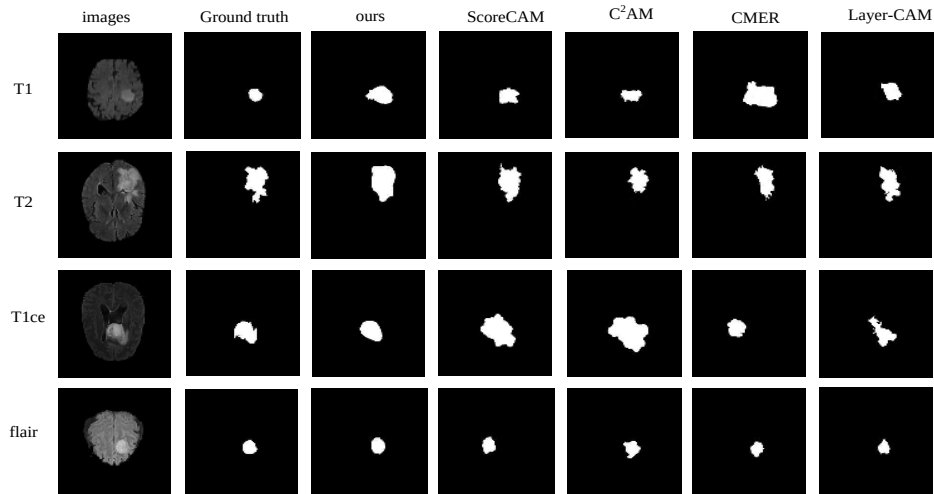


Fig. 7 Visual comparison of segmentation results on the BraTS 2021 dataset
图7 BraTS 2021数据集分割结果视觉对比

2.5 消融实验

为了验证CHFA-Net中不同模块的有效性,在MSD Brain数据集上进行了消融实验,比较了使用不同模块的分割结果。CHFA-Net中每个模块的消融实验结果如表3所示。结果表明,在4个序列数据集上,与纯CAM方法相比(无任何改进),在CAM中引入分层特征提取、因果干预和特征聚合3个模块都不同程度上提高了分割的准确率,

Table 3 The ablation experiment results of CHFA-Net							
表3 CHFA-Net消融实验结果							
CAM	HFE	CI	FA	Dice/%			
				flair	T1gd	T1w	T2w
√	×	×	×	46.65	38.32	35.67	52.34
√	×	√	×	60.37	57.90	48.68	61.04
√	√	×	√	72.48	60.79	56.23	70.27
√	√	√	√	78.80	62.72	60.90	74.89

从而证明了3个模块的有效性。具有因果干预模块的性能相比纯CAM方法提高了近10%,说明鲁棒的因果特征是提升分割精度的基础。同时具有分层特征提取和特征聚合模块的性能提高了20%以上,这是由于浅层与深层信息具有互补性,浅层灰度信息可通过高分辨率细节精准定位肿瘤边界,有效抑制了深层语义边界模糊的问题。深层语义信息则通过大感受野解析肿瘤与周围脑组织之间的整体结构关系,能矫正浅层噪声导致的误分割。因此,浅层特征和深层特征的有效融合对于医学图像分割具有重要意义。

图8的a、b、c、d组分别展示了纯CAM、CAM+CI、CAM+HFE+FA、CAM+HFE+FA+CI(本文方法)的可视化CAM对

比。a组显示纯CAM方法激活图只定位模型认为最具辨识性的区域,因而只覆盖了部分肿瘤区域。此外,由于图像比较模糊且具有各种噪声,产生的CAM还一定程度上错误激活了非肿瘤区域。b组相较于a组覆盖的肿瘤区域更完整,而且错误激活的区域较少,说明因果干预能提高像素预测的准确性,证明了因果干预模块的有效性。c组能基本正确定位肿瘤区域,覆盖的肿瘤区域较为完整,精度高于b组的CAM,缺陷是对肿瘤边界和形状的激活结果不够精细,不能很好地分割肿瘤边界。d组相较于c组对边界的分割更细致,没有错误激活非肿瘤区域,对肿瘤形状特征的分割更精准,证明了本文提出的分层特征提取和聚合模块的有效性,本文方法取得了最精准的分割结果。

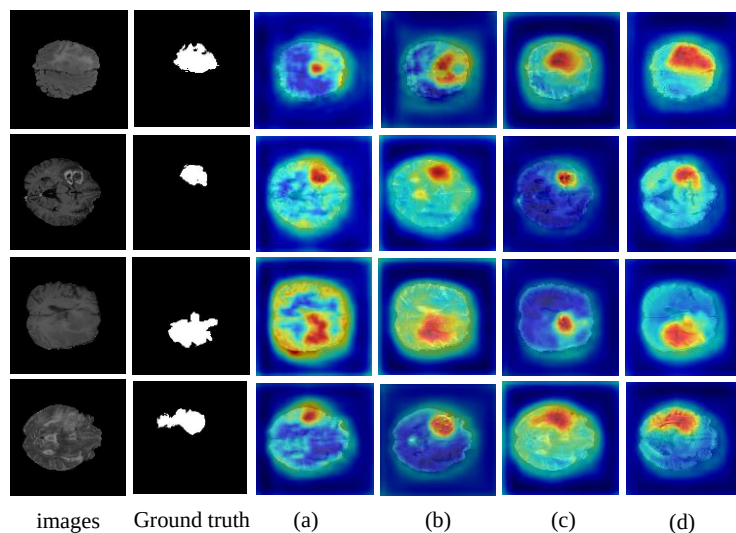


Fig. 8 Visualization comparison of different modules in ablation experiment

图8 消融实验不同模块可视化对比

3 结语

本文提出一个融合因果干预和分层特征聚合的弱监督语义分割网络CHFA-Net,其只需要图像级标签进行脑部肿瘤分割。首先,通过前门调整间接估计图像与类激活图CAM的因果效应,获取了鲁棒的因果特征。其次,提出一种特征聚合模块,设计了注意力模块CA以获取通道注意力权重,促使模型专注于响应强度较高的肿瘤区域,抑制无关信息。最后,利用中心损失学习病灶的形状和大小,提高模型对目标边界和形状的判别能力,以达到最佳的分割精度。实验结果表明,与其他先进的WSSS方法相比,本文方法具有最高的分割精度。但与全监督方法相比,本文方法仍具有一定差距。为弥补弱监督信号信息量少的缺陷,未来将考虑融合多种模态的医学图像,通过不同模态的互补信息提升模型性能。

参考文献:

[1] LAPOINTE S, PERRY A, BUTOWSKI N A. Primary brain tumours in adults[J]. The Lancet, 2018, 392: 432-446.

[2] ISENSEE F, JAEGER P F, KOHL S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature Methods, 2021, 18(2): 203-211.

[3] OTSU N. A threshold selection method from gray-level histograms[J]. Automatica, 1975, 11: 23-27.

[4] MICHAEL K, WITKIN A, TERZOPOULOS D. Snakes: active contour models [J]. International Journal of Computer Vision, 1988 (4): 321-331.

[5] WU X H, SHU X, FAN Y, et al. A dual-channel feature perception network for medical image segmentation[J]. Control and Decision, 2025, 40 (3): 871-879.

武相虎, 束鑫, 范燕, 等. 面向医学图像分割的双通道特征感知网络[J]. 控制与决策, 2025, 40(3): 871-879.

[6] LIU X, LIU Q, ZHANG Y, et al. TSSK-Net: weakly supervised biomarker localization and segmentation with image-level annotation in retinal OCT images [J]. Computers in Biology and Medicine, 2023, 153: 106467.

[7] KWEON H, YOON S H, YOON K J. Weakly supervised semantic segmentation via adversarial learning of classifier and reconstructor[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 11329-11339.

- [8] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2921–2929.
- [9] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 618–626.
- [10] WEI Y, XIAO H, SHI H, et al. Revisiting dilated convolution: a simple approach for weakly- and semi-supervised semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7268–7277.
- [11] ZHANG X, WEI Y, FENG J, et al. Adversarial complementary learning for weakly supervised object localization [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1325–1334.
- [12] WANG H, WANG Z, DU M, et al. Score-CAM: score-weighted visual explanations for convolutional neural networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 24–25.
- [13] JIANG P T, ZHANG C B, HOU Q, et al. LayerCAM: exploring hierarchical class activation maps for localization [J]. IEEE Transactions on Image Processing, 2021, 30: 5875–5888.
- [14] PATEL G, DOLZ J. Weakly supervised segmentation with cross-modality equivariant constraints [J]. Medical Image Analysis, 2022, 77: 102374.
- [15] LI Y, YU Y, ZOU Y, et al. Online easy example mining for weakly-supervised gland segmentation from histology images [C]//Switzerland: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022: 578–587.
- [16] PRABAKARAN B S, OSTROWSKI E, SHAFIQUE M. BoundaryCAM: a boundary-based refinement framework for weakly supervised semantic segmentation of medical images [DB/OL]. <https://arxiv.org/pdf/2303.07853v1>.
- [17] PEARL J, GLYMOUR M, JEWELL N P. Causal inference in statistics: a primer [M]. Hoboken: John Wiley & Sons, 2016.
- [18] TANG K, HUANG J, ZHANG H. Long-tailed classification by keeping the good and removing the bad momentum causal effect [J]. Advances in Neural Information Processing Systems, 2020, 33: 1513–1524.
- [19] CHEN K, SUN S, ZHAO J. Camil: causal multiple instance learning for whole slide image classification [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2024: 1120–1128.
- [20] YANG X, ZHANG H, QI G, et al. Causal attention for vision-language tasks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 9847–9857.
- [21] SUN K, XIE Z, GUO C, et al. SDGIN: structure-aware dual-level graph interactive network with semantic roles for visual dialog [J]. Knowledge-Based Systems, 2024, 286: 111251.
- [22] YANG X, FENG F, JI W, et al. Deconfounded video moment retrieval with causal intervention [C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021: 1–10.
- [23] ZANG C, WANG H, PEI M, et al. Discovering the real association: multimodal causal reasoning in video question answering [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 19027–19036.
- [24] LIANG T, LIU T, WANG J, et al. Causal deep learning for explainable vision-based quality inspection under visual interference [J]. Journal of Intelligent Manufacturing, 2025, 36: 1363–1384.
- [25] KIM J, LEE B K, RO Y M. Causal unsupervised semantic segmentation [DB/OL]. <https://arxiv.org/html/2310.07379>.
- [26] ZHANG D, ZHANG H, TANG J, et al. Causal intervention for weakly-supervised semantic segmentation [J]. Advances in Neural Information Processing Systems, 2020, 33: 655–666.
- [27] SUN K, XIE Z, GUO C, et al. SDGIN: structure-aware dual-level graph interactive network with semantic roles for visual dialog [J]. Knowledge-Based Systems, 2024, 286: 111251.
- [28] WANG Y. Causal class activation maps for weakly-supervised semantic segmentation [DB/OL]. <https://arxiv.org/pdf/2009.12547>.
- [29] PEARL J. Models, reasoning and inference [M]. Cambridge: Cambridge University Press, 2000.
- [30] ANTONELLI M, REINKE A, BAKAS S, et al. The medical segmentation decathlon [J]. Nature Communications, 2022, 13(1): 4128.
- [31] BAID U, GHODASARA S, MOHAN S, et al. The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification [DB/OL]. <https://arxiv.org/pdf/2107.02314>.
- [32] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234–241.

(责任编辑:黄健)