

基于图卷积与 Transformer 的缺失模态脑肿瘤分割

王彩莲, 郑文斌

(成都信息工程大学软件工程学院, 四川 成都 610225)

摘要:多模态磁共振成像(MRI)在脑肿瘤分割中具有重要应用,但模态缺失问题在临床中普遍存在,导致肿瘤区域细节丢失及模态间关联性遭到破坏等问题,显著降低分割性能。为此,该文提出一种基于图卷积与 Transformer 融合的缺失模态脑肿瘤分割模型。模型首先采用三维卷积与 MBCConv 结合的编码器模块提取局部特征,并利用三维适配器捕获脑肿瘤的空间信息,增强模态内特征表达能力。随后,设计了动态图卷积-Transformer 融合模块;其中,动态图卷积用于生成模态相关性矩阵,捕获模态间局部依赖关系,并结合掩码机制有效处理模态缺失;同时,Transformer 模块则通过自注意力机制建模模态与肿瘤区域的全局相关性上下文信息,从而显著提升特征融合鲁棒性。在 BraTS2018 和 BraTS2020 数据集上的实验结果表明,该模型在完整肿瘤、肿瘤核心及增强肿瘤区域的平均 Dice 系数较 M2FTrans 分别提升 1.36 百分点、1.28 百分点和 1.39 百分点,验证了其在模态缺失场景下的优越性。

关键词:脑肿瘤分割;缺失模态;三维适配器;动态图卷积;Transformer

中图分类号:TP391.41;R739.41 文献标识码:A 文章编号:1673-629X(2025)07-0173-09

doi:10.20165/j.cnki.ISSN1673-629X.2025.0064

Missing Modality Brain Tumor Segmentation Based on Graph Convolution and Transformer

WANG Cai-lian, ZHENG Wen-bin

(School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Multimodal Magnetic Resonance Imaging (MRI) has significant applications in brain tumor segmentation. However, the issue of missing modalities is common in clinical practice, leading to the loss of tumor region details and the disruption of inter-modality correlations, which significantly reduces segmentation performance. To address this, we propose a missing modality brain tumor segmentation model based on the fusion of Graph Convolutional Networks (GCN) and Transformer. The model first employs an encoder module combining 3D convolution and MBCConv to extract local features, and utilizes a 3D adapter to capture spatial information of brain tumors, enhancing the intra-modality feature representation capability. Subsequently, a dynamic graph convolution-Transformer fusion module is designed, where the dynamic graph convolution generates a modality correlation matrix to capture inter-modality local dependencies and effectively handles missing modalities using a masking mechanism. Meanwhile, the Transformer module models the global correlations and contextual information between modalities and tumor regions through a self-attention mechanism, thereby significantly enhancing the robustness of feature fusion. Experimental results on the BraTS2018 and BraTS2020 datasets demonstrate that the proposed model achieves an average Dice coefficient improvement of 1.36 percentage points, 1.28 percentage points, and 1.39 percentage points for the whole tumor, tumor core, and enhancing tumor regions, respectively, compared to M2FTrans, validating its superiority in scenarios with missing modalities.

Key words: brain tumor segmentation; missing modality; three-dimensional adapter; dynamic graph convolution; Transformer

0 引言

脑肿瘤是一种高危疾病,其位置、大小和生长速度各异,病程迅速,最终可危及生命。磁共振成像能够以最小的侵袭性清晰获得肿瘤的软组织病变的不同区

域,在脑肿瘤诊断中有明显优势。常用的 MRI 模态包括 T1 加权 (t1)、对比增强 T1 加权 (t1ce)、T2 加权 (t2) 和液体衰减反转恢复 (flair),这些模态提供互补信息,显著提高了分割的准确性和临床评估的可靠性,

收稿日期:2024-11-28

修回日期:2025-03-28

基金项目:四川省自然科学基金(2022NSFSC0571);四川省科技厅项目(2018JY0273)

作者简介:王彩莲(2000-),女,研究生,CCF 会员(Y0983G),研究方向为计算机视觉;通讯作者:郑文斌(1973-),男,博士,副教授,CCF 会员(25051S),研究方向为深度学习、图像分析。

支持诊断和手术规划。近年来,随着科学技术发展,多模态脑肿瘤分割取得了很大的进展。然而,受限于不同的扫描协议、数据缺损和患者配合度等因素,临床中难以获得完整的 MRI 图像^[1]导致脑肿瘤分割性能下降。在模态缺失的情况下,需要开发适用于临床实际的缺失模态脑肿瘤分割方法。

现有的缺失模态脑肿瘤分割方法在融合策略上存在差异。例如,HeMIS 模型^[2]通过计算每个可用脑肿瘤模态的均值和方差实现特征融合,但未能有效地聚合缺失模态的脑肿瘤特征。Ding 等人^[3]提出了区域融合网络,利用卷积融合可用模态,但难以捕获全局信息。为此,Zhang 等人^[1]引入 Transformer 模块,捕获脑肿瘤的全局信息和长期依赖性,但自注意力机制难以有效减轻不完整模态的影响。Shi 等人^[4]通过模态掩码和 Transformer 机制模拟多种缺失模态场景,并利用可学习融合令牌实现长程依赖,灵活整合跨模态信息。尽管直接融合可用的模态特定特征以学习共享的潜在表示可以对所有模态组合子集都具有鲁棒性的共同特征,现有方法在跨模态相关性建模上仍存在不足。这些方法难以捕获模态间局部的依赖关系,进一步限制了特征融合的有效性。此外,模态缺失会导致脑肿瘤细节信息^[5]丢失。这种对细节信息的忽视,无论是在模态融合还是单独分析^[6]中,都可能导致关键信息的遗漏,从而影响分割的准确性。

针对上述问题,该文提出了一种融合图卷积和 Transformer 的分割网络(Graph Convolutional Transformer Fusion, GCTF),旨在捕获脑肿瘤模态的局部和全局依赖信息。首先,在编码阶段,网络通过普通卷积层和 MBConv 层提取局部特征,并引入三维适配器(Three-Dimensional Adapter, T-Adapter)帮助模型更好地捕获脑肿瘤空间特征,为特征融合提供高质量表示。其次,为了捕获模态间的局部细节特征,提出了一种动态图卷积 Transformer 融合模块(Dynamic Graph Transformer Fusion Module, DGTFM),通过动态图卷积捕获模态间的局部依赖关系,并结合 Transformer 捕获全局上下文信息。最后,使用空间权重注意和通道融合变换器重新加权脑肿瘤特征,减轻特征冗余并输入到解码器进行脑肿瘤分割。

1 相关工作

在临床环境中,脑肿瘤图像的模态缺失问题较为常见。为解决这一问题,近年来提出了多种缺失模态脑肿瘤分割方法,主要分为三类:知识蒸馏方法、生成对抗网络方法和基于共享表示的方法。

1.1 知识蒸馏方法

知识蒸馏方法通过将完整模态训练的神经网络知

识迁移到处理缺失模态的学生网络中。ACN^[7]采用对抗性知识蒸馏,对齐完整模态与缺失模态的潜在表示;SMU-Net^[8]则将全模态与缺失模态编码到同一潜在空间,并通过风格匹配机制将全模态信息转移至缺失模态,校准表示空间。尽管知识蒸馏能够增强学生网络的特征学习能力,但学生模型在提取缺失模态关键特征时仍存在较大挑战。

1.2 生成对抗网络方法

生成对抗网络(Generative Adversarial Networks, GAN)通过可用模态重建缺失的脑肿瘤模态。HyperGAN^[9]、CollaGAN^[10]和 MM-GAN^[11]分别采用一对一、多对一和多对多的 GAN 架构,从现有模态中合成缺失模态。然而,这些方法通常需要针对不同缺失模态场景设计专用网络,计算资源消耗较大。此外,GAN 方法高度依赖可用模态,难以准确合成缺失模态中的关键特征。

1.3 基于共享表示的方法

基于共享表示的方法通过将多个可用模态投影到共享子空间,学习模态的不变表示。U-HMIS^[2]通过均值和方差融合可用模态的特征。Yao 等人^[12]引入离散共享空间,捕获细粒度表示并提高跨模态检索准确性。Chen 等人^[13]将特征分解为模态不变和模态特定特征,并通过门控策略优化表示。Zhang 等人^[1]基于 Transformer 设计了模态内与模态间编码器,对齐不同模态的全局相关性以提取模态不变表示。尽管这些方法在缺失模态的脑肿瘤分割中取得显著进展,但忽略了模态内外的跨模态局部细节信息,仍然影响分割精度。

2 方法

该文提出了一个融合图卷积与 Transformer 的缺失模态脑肿瘤分割框架,以解决 15 种不同模态组合下的分割问题。图 1 展示了框架整体架构,主要包括编码器、跳跃连接和解码器。首先,在编码器中引入三维适配器,以进一步处理和增强模态内的局部特征。其次,将动态图卷积和 Transformer 结合,实现模态间的局部和全局信息的协同捕获。图卷积通过动态建模可用模态间的邻接关系,从其他模态中动态补充所需信息,弥补信息不足;而 Transformer 则利用自注意力层捕获模态的全局依赖关系^[14],进一步增强跨模态特征融合能力。最后,融合后的特征被送入解码器,通过逐步上采样进行脑肿瘤分割。

2.1 编码器模块

编码器采用三维卷积层和 MBConv 层提取局部特征,并引入三维适配器,帮助模型更好地捕获脑肿瘤空间信息。具体而言,编码器第一级由三层步长为 1

的卷积层构成;第二至第五级则由一层步长为 2 的 MBConv 层和两层步长为 1 的 MBConv 层。在每个编码器的第五层卷积块后,通过三维适配器^[15]帮助模型更好地提取脑肿瘤空间信息。编码器从 m 模态中提取的第 k 级特征表示为特征向量 $E_m^k \in \mathbb{R}^{c \times h \times w \times d}$, 其中 $m \in \{t1, t2, t1ce, flair\}$ 。每个模态通过独立编码器提取特征。

三维适配器的结构如下:首先,将编码器第五级卷积层生成的特征 $E_m^5 \in \mathbb{R}^{c \times h \times w \times d}$ 输入三维适配器,其中, h 、 w 、 d 分别表示特征的高度、宽度和深度, c 为通道数。三维适配器通过增强模态内局部细节信息,提升了缺失模态脑肿瘤的分割精度。具体而言,特征 E_m^5 首先经过归一化处理 and 降维投影,得到特征 \hat{E}_m^5 。

$$\hat{E}_m^5 = W_1 \cdot \text{LN}(E_m^5) + b_1 \quad (1)$$

其中, LN 表示归一化操作, W_1 和 b_1 分别为权重矩阵和偏置向量。接下来,对特征 \hat{E}_m^5 进行三维卷积处理,并通过非线性激活函数捕获局部空间信息。经过非线性变换后的结果进行上投影,得到特征 $\widetilde{E}_m^5 \in \mathbb{R}^{c \times h \times w \times d}$ 。

$$\widetilde{E}_m^5 = W_2 \cdot \delta[\text{Conv3}(\hat{E}_m^5)] + b_2 \quad (2)$$

其中, W_2 为权重, b_2 为偏置, δ 表示 GELU 激活函数, Conv3 表示三维卷积操作。

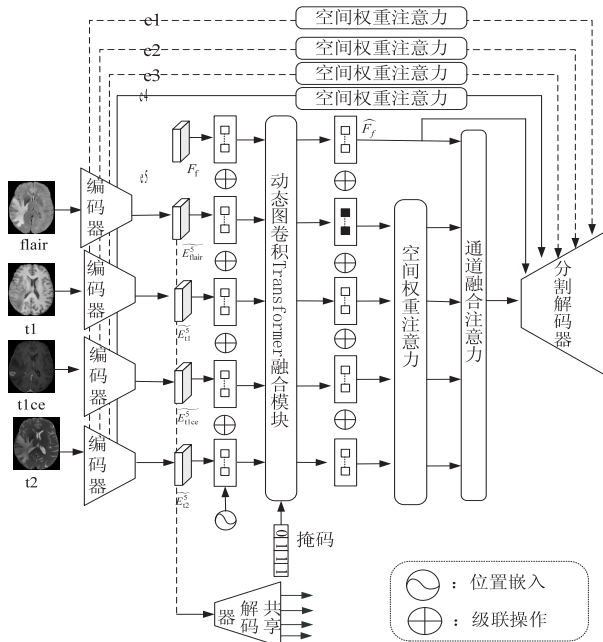


图 1 GCTF 模型的总结构

2.2 动态图卷积 Transformer 融合模块

尽管掩码 Transformer 模块在捕捉缺失模态脑肿瘤的全局信息方面表现良好,但在提取模态间的局部依赖关系时仍然存在不足。已有研究表明^[16], MRI 模态之间存在强相关性。受 M2GCNet 模型的启发^[6], 该文设计了动态图卷积 Transformer 融合模块。动态

图卷积通过基于掩码的邻接矩阵动态建模模态间的局部依赖,弥补模态缺失时的信息不足;Transformer 则建模跨模态的全局依赖,增强信息融合的完整性。具体而言,动态图卷积网络专注于不同模态节点特征的高效融合,精准捕获模态间相关性,提升脑肿瘤边界的分割精度;Transformer 利用自注意力机制提取全局信息,进一步提升分割性能,尤其在模态缺失场景下效果显著。

2.2.1 动态图卷积模块

该文提出的动态图卷积网络旨在提取模态间的局部依赖特征,其核心机制是通过邻居节点的交互和相关性动态建模模态关系,从而实现了对多模态数据局部特征的高效整合。图 2 展示了该网络的结构框架以及提取跨模态局部特征的具体步骤。其中,虚线表示权重较低的边,反映了动态图卷积在权重动态调整中的灵活性, $\{V_m^1, V_m^2\}$ 表示模态节点。

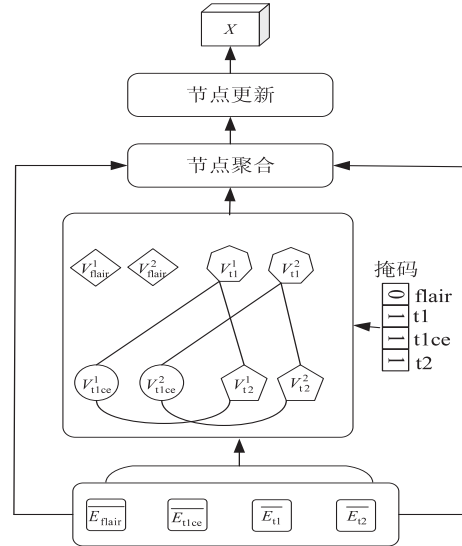


图 2 动态图卷积模块架构

首先,将独立编码器生成的模态特征 \widetilde{E}_m^5 进行展平和转置操作,得到特征 $\overline{E}_m \in \mathbb{R}^{N \times c}$, $N = h \times w \times d$ 。随后,在模态维度上堆叠四个模态特征,生成张量 $\overline{E}_4 \in \mathbb{R}^{4 \times N \times c}$ 。根据 \overline{E}_4 构建邻接矩阵 L , 邻接矩阵的生成过程如下:首先将四个模态进行模态加权,对输入特征 \overline{E}_4 进行线性变换得到特征 \overline{E}_4' , 并将 \overline{E}_4' 与可学习的权重相乘,以重新加权模态特征,突出重要的模态。

$$\overline{E}_4' = \overline{E}_4 \odot w \quad (3)$$

其中, \odot 表示逐元素相乘, w 是通过反向传播优化的可学习权重。然后应用掩码过滤缺失模态,利用掩码矩阵 M 对加权后的模态特征进行处理得到特征 \overline{E}_4'' , 将缺失模态置为 0, 使生成邻接矩阵不会影响其他模态。

$$\overline{E}_4 = \overline{E}_4 \odot (M_i + \mu) \quad (4)$$

其中, μ 是一个极小的正值, 避免梯度问题或计算错误。掩码矩阵 M 生成过程如下: $M \in \mathbb{R}^4$, 根据输入数据确定模态是否可用, 若模态 i 可用, 则 $M_i = 1$, 若模态 i 缺失, 则 $M_i = 0$ 。随后通过矩阵乘法计算模态间的邻接矩阵, 对处理后的模态特征 \overline{E}_4 计算内积, 生成模态间的邻接矩阵 L 。

$$L = \overline{E}_4 \cdot \overline{E}_4 \quad (5)$$

将得到的邻接矩阵 L 进行 softmax 归一化后再次利用掩码机制对缺失模态进行稀疏化。

$$L_{i,j} = \left(\frac{\exp(L_{i,j})}{\sum_k \exp(L_{i,k})} \right) \cdot M_i \cdot M_j \quad (6)$$

其中, k 是矩阵 $L_{i,j}$ 的列索引, i 和 j 表示模态。最后, 对邻接矩阵 $L_{i,j}$ 进行对角线正则化, 避免梯度消失, 生成最终的邻接矩阵 L 。

将 \overline{E}_4 和邻接矩阵 L 输入图卷积模块, 使用邻接矩阵对各模态的特征进行加权求和, 以聚合模态邻居节点的信息。这一过程可以表示为:

$$X = L \cdot (\overline{E}_4 \cdot W) + b \quad (7)$$

其中, \overline{E}_4 是输入特征矩阵, W 是权重矩阵, b 表示偏置, X 是聚合了邻居节点信息的特征矩阵。在图卷积模块中, 使用邻接矩阵有效编码模态间的关系, 图卷积网络通过特征聚合, 既捕获模态节点自身的特征, 又利用跨模态邻居节点的交互获取关键信息。这种机制使得图卷积网络能够充分捕获模态间局部依赖信息, 从而全面地捕捉脑肿瘤的多模态特征。

2.2.2 掩码 Transformer 模块

将经过图卷积模块处理后的特征 X 分解出每个模态特征 $\widetilde{E}_m \in \mathbb{R}^{N \times c}$, 并引入可学习的融合令牌 $F_f \in \mathbb{R}^{N \times c}$ 。随后, 将四个模态的特定特征 \widetilde{E}_m 与可学习的融合令牌 F_f 级联, 并结合位置嵌入 $P_E \in \mathbb{R}^{5N \times c}$, 得到联合特征 $Z \in \mathbb{R}^{5N \times c}$ 。具体的级联操作如下所示:

$$Z = (F_f \oplus T) + P_E \quad (8)$$

其中, $T = \widetilde{E}_{flair} \oplus \widetilde{E}_{t1} \oplus \widetilde{E}_{t1cc} \oplus \widetilde{E}_{t2}$ 表示将四个模态的特征级联, \oplus 表示级联操作。随后, 将特征 Z 输入掩码 Transformer 模块。

在掩码 Transformer 模块中, 首先采用掩码注意力模块对缺失模态的脑肿瘤进行全局特征提取。掩码注意力的操作过程如图 3 所示。具体而言, 将联合特征 Z 通过自注意力机制投影为查询向量 Q 、键向量 K 和价值向量 V 。为应对模态缺失, 采用对应于 Z 标记间成对依赖关系的二进制注意力掩码 $A \in \{0, 1\}^{5N \times 5N}$, 其

中掩码用于确定标记 i 和 j 是否属于缺失模态。掩码机制的具体规则如下: 若查询向量 q_i 和键向量 k_j 属于同一可用模态, 则 $A_{i,j} = 1$, 表示同一模态内令牌通过自注意力交互, 强化模态特异性特征; 若 q_i 属于可学习的融合令牌而 k_j 属于可用模态, 则 $A_{i,j} = 1$, 可学习的融合令牌主动聚合所有可用模态特征, 促进跨模态信息融合; 若 q_i 或 k_j 中某一个属于缺失模态, 则 $A_{i,j} = 0$, 屏蔽缺失模态的注意力权重, 避免模型建立无效依赖。根据掩码机制, 掩码注意力模块计算如下:

$$\text{att}(Q, K, i, j) = \frac{A_{i,j} \exp\left(\frac{q_i \cdot k_j}{\sqrt{H}}\right)}{\sum_{j:A_{i,j}=1} \exp\left(\frac{q_i \cdot k_j}{\sqrt{H}}\right)} \quad (9)$$

$$\text{att}[i, j] = \begin{cases} -\infty, & \text{if mask}[i, j] = 0 \\ \text{att}(Q, K, i, j), & \text{otherwise} \end{cases} \quad (10)$$

$$x = \text{softmax}(\text{att}[i, j]) \cdot V \quad (11)$$

$$x = \text{Linear}(x) \quad (12)$$

其中, j 表示该位置对应的模态可用, H 表示多头掩码注意力模块的头数, Linear 函数表示线性变换。 x 表示通过掩码注意力模块融合的特征。

然后, 将生成的特征 x 经过掩码注意力机制 (Multi-Head Masked Attention, MHMA) 和前馈神经网络 (Feed-Forward Network, FFN) 后, 得到增强的特征 X' 。

$$x_1 = \text{MHMA}(x) + Z \quad (13)$$

$$X' = \text{FFN}(\text{LN}(x_1)) + x_1 \quad (14)$$

其中, LN 是层归一化操作。通过 3 层掩码 Transformer 模块后, 可学习的融合令牌 F_f 被更新和整形为具有跨模态信息的 $\hat{F}_f \in \mathbb{R}^{c \times h \times w \times d}$, 模态特定特征 \widetilde{E}_m 被更新和整形为包含更多模态内全局特征的 $\hat{E}_m \in \mathbb{R}^{c \times h \times w \times d}$ 。

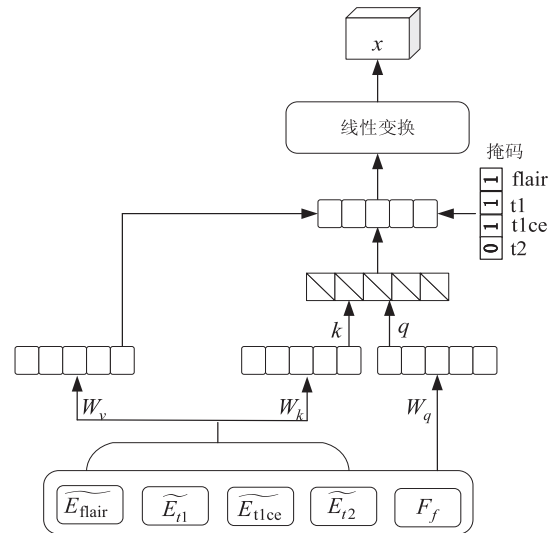


图 3 掩码注意力模块

2.3 空间权重注意力

在模态交互过程中,尽管不同模态的脑肿瘤特征具有相同的注意力权重,但各模态特征在分割任务中的重要性不同。因此,在每层特征融合时,采用空间权重注意力机制对不同模态的脑肿瘤特征进行加权。根据文献[17]的研究,模态交互的掩码自注意力矩阵列向量可表示对应模块在整个脑组织图像中的重要性。因此,该文利用掩码自注意力矩阵的列向量来表示相应脑肿瘤特征在整体图像中的重要性,其过程如下:

$$I = \sum_{j=1}^{4N} I(j) = \sum_{j=1}^{4N} \sum_H \sum_{i=4N+1}^{5N} \text{att}_1(i,j), I \in \mathbb{R}^{1 \times 4N} \quad (15)$$

$$I_m = \text{split}(I), I_m \in \mathbb{R}^{1 \times N} \quad (16)$$

其中: $\text{att}_1(i,j)$ 表示第一层掩码注意力的权重, $I(j)$ 是掩码注意力权重第 j 列求和得到的令牌 j 的权重; I 表示所有模态中标号为 j 的权重, I_m 是每个模态的令牌权重,并将其整形为与 \hat{E}_m 尺寸相同的矩阵 \hat{I}_m 。将整形后的每个模态的令牌权重 \hat{I}_m 与脑肿瘤每个模态的特定特征 \hat{E}_m 的对应位置元素相乘,以重新加权脑肿瘤模态特征得到增强的模态特定特征 \widehat{E}_m , 可以表示为:

$$\widehat{E}_m = \hat{E}_m \odot \hat{I}_m \quad (17)$$

其中, \odot 表示逐元素相乘。

2.4 通道融合注意力

上述空间权重注意力减轻了空间冗余,但通道方向仍存在冗余。因此,该文引入通道融合注意力,它关注的是不同模态或特征通道之间的冗余信息。首先,将模态特定特征 \widehat{E}_m 沿着空间维度整形为 $\bar{E}_m \in \mathbb{R}^{c \times N}$, 并将 4 个模态的特征级联为 $\bar{Z} \in \mathbb{R}^{4c \times N}$ 。接着,将 \bar{Z} 投影为键向量 \mathbf{K} 和值向量 \mathbf{V} 。同时,将可学习的融合令牌 \hat{F}_f 整形为 $\hat{F}_f^0 \in \mathbb{R}^{c \times N}$, 并通过自注意力机制投影为查询向量 \mathbf{Q} 。投影过程如下:

$$\lambda(x) = \omega(\eta(\varphi(\cdot))) \quad (18)$$

其中, ω 和 η 是卷积核大小为 1 的逐点卷积, φ 是卷积核大小为 3 的深度卷积。类似于上述掩码注意力机制,该方法沿通道维度引入了一个二进制注意力掩码 $\Omega \in \{0,1\}^{c \times 4c}$, 其中 i 和 j 分别对应于 \mathbf{Q} 和 \mathbf{K} 中 q_i 和 k_j 之间的关系。具体规则如下:若 i 和 j 属于相同的可用模态,则 $\Omega_{i,j} = 1$;若 i 属于可学习的融合令牌而 j 属于可用模态,则 $\Omega_{i,j} = 1$;若 i 和 j 任何一方属于缺失模态,则 $\Omega_{i,j} = 0$ 。基于二进制掩码,掩码注意力模块的计算过程如下:

$$\tilde{Z} = \lambda(\text{CFT}(\mathbf{Q}, \mathbf{K}) \cdot \mathbf{V}) + \tilde{Z} \quad (19)$$

$$\tilde{Z}_0 = \text{FFN}(\tilde{Z}) + \tilde{Z} \quad (20)$$

$$\text{CFT}(\mathbf{Q}, \mathbf{K}, i, j) = \frac{\Omega_{i,j} \exp(\frac{q_i \cdot k_j}{\sqrt{N}})}{\sum_{j, \Omega_{i,j}=1} \exp(\frac{q_i \cdot k_j}{\sqrt{N}})} \quad (21)$$

其中, j 表示该位置为可用的脑肿瘤模态, $N = h \times w \times d$ 。FFN 表示前馈神经网络。 \tilde{Z}_0 是经过 1 层通道融合注意力得到的结果,经过 2 层通道融合注意力后,可学习的融合令牌 \hat{F}_f 进一步增强为 $\bar{F}_f \in \mathbb{R}^{c \times h \times w \times d}$ 。

2.5 解码器

解码器包括分割解码器和共享解码器。为减少模态缺失带来的偏差,该文引入共享解码器 D_s 进行正则化处理。具体来说,各模态特定特征被投影到一个共享潜在空间,同时每个模态特征仍能独立分割,从而减轻模态缺失对分割效果的影响。设计了四个模态特定编码器和一个共享解码器,并通过以下方式联合训练。

$$L_r = \sum_{m \in M} L_{\text{Dice}}(D_s(E_m^5), Y) + L_{\text{WCE}}(D_s(E_m^5), Y) \quad (22)$$

其中, Y 表示真实的分割结果, L_{Dice} 表示 Dice 损失, L_{WCE} 表示加权交叉熵损失。分割解码器将编码器的模态特征依次应用动态图卷积 Transformer 融合模块、空间权重注意力和通道融合注意力学习到高质量融合特征 \hat{F}_f 和 \bar{F}_f , 随后将融合特征进行上采样并与重新加权的跳跃连接特征重新加权。该方法有效保留了脑肿瘤的底层空间信息,从而提升了缺失模态下的分割精度。分割解码器的训练损失为:

$$L_{\text{seg}} = L_{\text{Dice}}(\text{Concat}(Y_m), Y) + L_{\text{WCE}}(\text{Concat}(Y_m), Y) \quad (23)$$

其中, Y_m 四个模态编码器的特征级联后依次通过 e5 层的图卷积掩码模块、空间权重注意力和通道融合注意力进行处理;在 e1、e2、e3 层经过空间权重注意力进行处理, e4 层经过空间权重注意力和通道融合注意力进行处理。五层的结果全部输入分割解码器,通过上采样得到预测的分割结果。 Y 表示真实的分割结果。Concat 是级联操作。

为提高分割解码器的稳定性和鲁棒性,该文你引入了深度监督方法,在神经网络的不同层分别加入损失函数,防止梯度消失并使浅层特征也能学习到有用信息。深度监督损失 L_{DS} 由 Dice 损失 L_{Dice} 和加权交叉熵损失 L_{WCE} 组成:

$$L_{\text{DS}} = \sum_{i=1}^5 L_{\text{Dice}}(\text{Up}_{2^{i-1}}(\hat{F}_f^i), Y) + L_{\text{WCE}}(\text{Up}_{2^{i-1}}(\hat{F}_f^i), Y) \quad (24)$$

其中, $\text{Up}_{2^{i-1}}$ 表示 2^{i-1} 倍的上采样操作,用于对不同层

的特征进行上采样,使低级特征映射到与最终输出一致的分辨率。 \hat{F}_f^t 表示分割解码器第 t 层的融合特征。

2.6 总损失

该模型的总训练损失由共享解码损失 L_r 、分割解码损失 L_{seg} 和深度监督损失 L_{DS} 组成,可以表示为:

$$L = L_r + L_{DS} + L_{seg} \quad (25)$$

3 实验

3.1 数据集

本实验选取了多模态脑肿瘤分割挑战的两个公开数据集: BraTS2018^[18] 和 BraTS2020^[19]。BraTS2018 包含 285 例患者, BraTS2020 包含 369 例患者。每例数据包括四种 MRI 模态 (flair、t1、t1ce、t2) 及对应的脑肿瘤标签,标注为坏死及非增强肿瘤组织、水肿和增强肿瘤区域。分割任务将脑肿瘤区域划分为完整肿瘤 (Whole Tumor, WT)、肿瘤核心 (Tumor Core, TC) 和增强肿瘤 (Enhancing Tumor, ET)。数据按 7 : 2 : 1 划分为训练集、验证集和测试集。

3.2 实验细节

该文采用 Dice 相似性系数 (Dice Similarity Coefficient, DSC) 和 95% 豪斯多夫距离 (95% Hausdorff Distance, 95% HD) 进行性能评价。所有实验硬件平台为 NVIDIA GeForce RTX 3090 GPU (24 GB 显存), 软件环境基于 Python 3.10 及 PyTorch 1.12.0 框架实现。在数据预处理阶段,原始图像经归一化处理采用随机裁剪策略构建样本。具体来说,提出的 GCTF 方法与对比方法 M2FTrans^[4] 采用 $80 \times 80 \times 80$ 的随机裁剪尺寸; 而 mmFormer^[1] 方法和 MMMViT^[20] 方法通过实验证明当裁剪尺寸为 $128 \times 128 \times 128$ 时获得最优性能。为提升模型的泛化性能,所有方法均采用随机翻转、强度偏移和镜像翻转等数据增强策略。网络训练采用 AdamW 优化器,设置权重衰减系数为 $1e-4$ 。不同方法的初始学习率不同,文中方法、mmFormer 方法和 M2FTrans 方法均采用 $2e-4$ 的初始学习率,而 MMMViT 方法的最佳初始学习率为 $1e-3$ 。所有实验均采用统一训练配置: 批量大小设置为 2, 最大训练轮次为 1 000 epochs。

3.3 对比实验

为验证文中方法的有效性,表 1 ~ 3 分别展示了在 BraTS2018 和 BraTS2020 数据集上提出的融合图卷积 Transformer 模型与 mmFormer^[1]、M2Ftrans^[4] 和 MMMViT^[20] 方法在 95% HD 和 DSC 的比较结果。mmFormer 方法首次将 Transformer 架构引入缺失模态的脑肿瘤分割任务。该方法通过构建混合模态特异性编码器和模态相关编码器,有效捕获模态内及跨模态

间的长程依赖关系,提取具有鲁棒性的模态不变表示,最终建立对任意输入模态子集均具有稳定分割能力的模型框架; M2FTrans 方法融合了 CNN 与 Transformer 优势,采用掩码机制有效缓解缺失模态带来的负面影响,并引入可学习融合令牌实现跨模态特征的深度交互与动态融合; MMMViT 在模型架构方面进行优化,通过由 4 个 CNN-Transformer 混合模块构成的模态内编码器,实现局部特征提取与全局多尺度特征学习的有机结合,并设计了包含多模态相关建模模块和多模态表示融合模块的模态间编码器,有效学习跨模态共享的潜在空间表示。与 mmFormer、M2FTrans 和 MMMViT 方法相比,文中方法在继承传统 Transformer 长程依赖建模优势的基础上,并行设计了动态图卷积模块,利用模态间的相关性,强化模态间的局部依赖关系,此外也使用 CNN 和三维适配器组合进一步捕获局部空间细节信息。根据表 1 所示,在 BraTS2018 数据集上,文中方法相比次优的 M2FTrans 方法,在 WT、TC 和 ET 区域的平均 95% HD 分别降低了 0.72 mm、1.13 mm 和 0.72 mm; 表 2 中的结果进一步表明,文中方法在 WT、TC 和 ET 区域的平均 DSC 分别提升了 1.36 百分点、1.28 百分点和 1.39 百分点。对于 BraTS2020 数据集,表 1 和表 3 中的结果也呈现出类似的趋势,文中方法始终优于其他方法,与次优的 M2FTrans 方法相比,WT、TC 和 ET 区域的 DSC 平均分别增加了 0.93 百分点、1.05 百分点和 1.13 百分点; 同时,WT、TC 和 ET 区域的平均 95% HD 分别降低了 0.64 mm、0.88 mm 和 0.42 mm。从表 2 和表 3 中可以清晰看到,ET 区域的分割效果有明显提升,这进一步验证了动态图卷积模块捕获不同模态间局部依赖信息的有效性。总体而言,文中方法在大多数情况下都表现出优异的性能。

表 1 BraTS2018 与 BraTS2020 数据集的 95% 豪斯多夫距离对比

区域	方法	BRATS2018	BRATS2020
WT	mmFormer	7.30	7.71
	M2Ftrans	6.40	5.68
	MMMViT	7.42	8.05
	GCTF	5.68	5.04
TC	mmFormer	8.42	8.09
	M2Ftrans	7.30	6.49
	MMMViT	8.66	8.21
ET	GCTF	6.17	5.61
	mmFormer	7.31	6.11
	M2Ftrans	5.95	5.02
	MMMViT	7.13	5.72
	GCTF	4.93	4.60

表 2 BraTS2018 数据集 Dice 相似系数分割结果对比

可用 模态	mmFormer			M2Ftrans			MMViT			GCTF		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
flair	85.01	64.80	31.59	87.15	69.89	37.98	88.10	63.10	41.10	88.10	71.85	40.99
t1ce	75.24	79.01	75.37	77.78	82.6	78.92	75.30	77.60	73.40	83.98	83.12	79.58
t1	73.36	62.06	32.45	77.21	66.24	37.24	66.90	57.20	34.50	83.79	70.74	39.52
t2	84.28	67.97	37.19	86.92	72.37	46.41	82.00	66.90	43.20	87.26	74.43	49.89
flair, t1ce	88.11	82.04	76.83	88.85	84.78	82.05	89.40	77.80	75.00	89.49	85.70	82.52
flair, t1	87.39	72.72	38.76	88.37	74.08	43.48	88.20	68.60	46.90	88.44	76.99	46.67
flair, t2	88	71.93	41.09	89.24	75.4	49.12	89.10	70.30	48.00	89.95	76.01	51.06
t1ce, t1	78.6	81.37	76.3	81.06	83.45	80.77	74.80	78.10	77.10	82.68	83.55	81.01
t1ce, t2	86.1	82.26	74.47	88.07	85.23	80.93	81.50	80.10	76.20	88.74	85.90	81.65
t1, t2	86	71.38	40.26	87.45	74.45	47.23	84.60	72.30	46.10	87.58	76.89	48.59
flair, t1ce, t1	88.51	83.42	79.53	88.95	85.26	82.19	88.50	81.10	75.90	89.60	85.84	82.83
flair, t1ce, t2	89.01	83.23	77.17	89.78	85.29	80.56	89.30	80.90	77.40	90.13	84.97	80.93
flair, t1, t2	88.49	74.09	43.01	89.39	76.48	49.79	88.90	74.40	54.90	90.04	77.36	50.80
t1ce, t1, t2	86.61	83.43	74.88	88	85.46	80.82	85.60	83.80	76.30	88.40	86.27	81.43
whole	89.19	84	77.69	89.73	85.67	80.61	91.20	85.80	80.20	90.15	86.17	81.41
平均	84.93	76.25	58.44	86.53	79.11	63.87	84.20	74.50	61.70	87.89	80.39	65.26

表 3 BraTS2020 数据集 Dice 相似系数分割结果对比

可用 模态	mmFormer			M2Ftrans			MMViT			GCTF		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
flair	86.27	67.80	38.39	88.70	72.20	43.39	89.30	65.70	43.80	89.56	73.62	45.48
t1ce	74.86	80.74	76.45	78.80	81.85	82.57	76.60	78.80	75.20	80.51	82.82	83.25
t1	74.91	64.24	34.78	79.15	66.75	40.87	70.40	57.10	36.50	81.84	69.27	41.40
t2	85.37	70.21	46.12	87.20	72.31	51.50	83.60	67.30	44.70	87.72	73.90	51.57
flair, t1ce	89.03	84.44	77.22	90.38	85.54	83.07	90.90	78.20	78.30	91.42	86.95	84.59
flair, t1	88.91	71.83	41.17	90.30	74.44	47.04	90.60	67.90	48.40	91.75	75.84	48.20
flair, t2	89.04	72.82	47.98	90.56	75.42	53.87	90.20	71.50	50.00	91.07	76.55	55.73
t1ce, t1	79.61	82.00	77.12	82.40	83.70	83.81	76.30	78.90	79.70	84.72	84.75	84.59
t1ce, t2	87.35	84.30	75.29	88.67	84.62	82.35	83.10	79.80	77.90	89.49	84.80	83.58
t1, t2	87.19	72.61	48.07	88.34	73.56	49.90	85.40	72.30	46.10	88.90	74.91	51.51
flair, t1ce, t1	89.61	84.59	77.13	91.00	85.82	84.12	88.50	81.10	75.90	91.21	86.16	85.93
flair, t1ce, t2	89.82	84.63	76.36	91.16	85.27	81.23	90.60	85.90	78.20	91.33	86.05	81.67
flair, t1, t2	89.96	73.90	48.77	90.91	76.14	53.33	89.70	75.80	54.80	91.00	77.01	54.32
t1ce, t1, t2	87.85	84.28	76.69	89.01	84.90	82.36	89.60	84.10	80.30	89.70	85.01	83.97
whole	90.31	84.49	78.11	91.36	85.43	82.17	91.60	85.90	81.50	91.56	86.17	82.77
平均	86.01	77.52	61.31	87.86	79.20	66.77	85.80	75.40	63.40	88.79	80.25	67.90

3.4 消融实验

本节通过消融实验验证了各组件的有效性。实验首先以基线模型为对照,逐步引入 DGTFM 和 T-Adapter,并分析其对脑肿瘤分割性能的影响。结果表明,DGTFM 和 T-Adapter 均能显著提升分割精度,而

两者结合后的 GCTF 模型进一步提高了缺失模态下的分割效果。具体而言,如表 4 所示,在 BraTS2018 数据集上,加入 T-Adapter 后,WT、TC 和 ET 区域的平均 DSC 分别提高了 0.33 百分点、0.78 百分点和 0.65 百分点;加入 DGTFM 后分别提高了 1.09 百分点、1.07

百分点和 1.10 百分点;当两者结合形成 GCTF 模型时,WT、TC 和 ET 区域的平均 DSC 分别提高了 1.36 百分点、1.28 百分点和 1.39 百分点。类似地,在 BraTS2020 数据集上(见表 5),引入 T-Adapter 后,WT、TC 和 ET 区域的结果分别提高了 0.46 百分点、0.55 百分点和 0.27 百分点;引入 DGTFM 后分别提高了 0.70 百分点、0.73 百分点和 0.77 百分点;两者结

合的 GCTF 模型,WT、TC 和 ET 区域的平均 DSC 分别提高了 0.93 百分点、1.05 百分点和 1.13 百分点。这些实验结果表明,DGTFM 能够有效捕获模态间的局部依赖信息,而 T-Adapter 在增强模态内局部细节提取方面发挥了重要作用。两者的结合在缺失模态条件下对脑肿瘤分割的性能提升尤为显著。

表 4 BraTS2018 数据集上消融研究结果

方法	DSC/%				95% HD/mm			
	WT	TC	ET	均值	WT	TC	ET	均值
Baseline	86.53	79.11	63.87	76.50	6.40	7.30	5.95	6.55
Baseline+T-Adapter	86.86	79.89	64.52	77.09	6.01	6.82	5.53	6.12
Baseline+DGTFM	87.62	80.18	64.97	77.59	5.87	6.45	5.19	5.84
GCTF	87.89	80.39	65.26	77.85	5.68	6.17	4.93	5.59

表 5 BraTS2020 数据集上消融研究结果

方法	DSC/%				95% HD/mm			
	WT	TC	ET	均值	WT	TC	ET	均值
Baseline	87.86	79.20	66.77	77.94	5.68	6.49	5.02	5.73
Baseline+T-Adapter	88.32	79.75	67.04	78.37	5.49	6.11	4.89	5.5
Baseline+ DGTFM	88.56	79.93	67.54	78.68	5.26	5.84	4.74	5.28
GCTF	88.79	80.25	67.90	78.98	5.04	5.61	4.60	5.08

3.5 可视化展示

如图 4 所示,在 BraTS2018 数据集上的实验结果通过分割结果可视化进行了展示。可视化分析表明,当仅有一个模态或者两个模态时,脑肿瘤的分割效果较差,而随着可用模态数量的增多,分割结果逐渐接近真实标签。这证明了提出的 GCTF 网络的有效性,特别是在处理模态缺失的情况下。具体而言,GCTF 网络通过动态图卷积模块构建了跨模态局部依赖,同时,掩码 Transformer 模块能够有效捕获全局信息。两个模块并行工作,即使在模态缺失的情况下,模型仍然实现高精度的脑肿瘤分割。

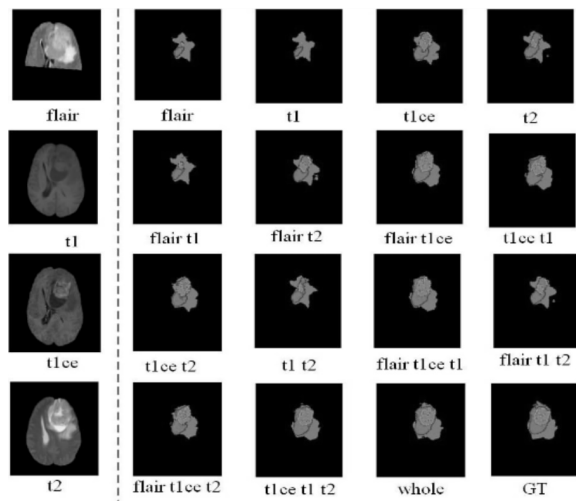


图 4 GCTF 方法 15 种模态组合分割结果的可视化

4 结束语

该文提出了一种融合图卷积与 Transformer 的缺失模态脑肿瘤分割方法,通过结合三维适配器和动态图卷积 Transformer 融合模块,在模态缺失的情况下根据模态间相关性捕获局部依赖信息和细节特征。具体而言,三维适配器模块位于模态特定编码器和动态图卷积 Transformer 融合模块之间,能够有效提取脑肿瘤的空间细节特征,提升特征融合效果;动态图卷积 Transformer 融合模块则利用脑肿瘤模态间的相关性,捕获不同模态间的局部依赖信息和全局信息,提高缺失模态分割的精度。此外,空间加权注意力机制对可用模态特征重新加权,进一步优化特征融合效果。基于公开的 BraTS2018 和 BraTS2020 数据集的实验表明,该模型在整体性能上优于现有方法。未来,将进一步完善该方法,适应不同场景下的临床需求,并期待与医院合作,在实际临床环境中优化分割模型。

参考文献:

[1] ZHANG Y, HE N, YANG J, et al. Mmformer: multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation[C]//International conference on medical image computing and computer-assisted intervention. Cham: Springer, 2022: 107-117.

[2] HAVAEI M, GUIZARD N, CHAPADOS N, et al. Hemis;

- hetero-modal image segmentation [C]//Medical image computing and computer-assisted intervention - MICCAI 2016; 19th international conference. Athens; Springer, 2016; 469-477.
- [3] DING Y, YU X, YANG Y. RFNet; region-aware fusion network for incomplete multi-modal brain tumor segmentation [C]//Proceedings of the IEEE/CVF international conference on computer vision. Montreal; IEEE, 2021; 3975-3984.
- [4] SHI J, YU L, CHENG Q, et al. MFTrans; modality-masked fusion transformer for incomplete multi-modality brain tumor segmentation [J]. IEEE Journal of Biomedical and Health Informatics, 2023, 28(1); 379-390.
- [5] 黄莉. 基于 U-Net 改进模型的多模态脑肿瘤分割方法 [J]. 计算机技术与发展, 2022, 32(11); 58-63.
- [6] ZHOU T. M2GCNet; multi-modal graph convolution network for precise brain tumor segmentation across multiple MRI sequences [J]. IEEE Transactions on Image Processing, 2024, 7(1); 278-287.
- [7] WANG Y, ZHANG Y, LIU Y, et al. ACN; adversarial co-training network for brain tumor segmentation with missing modalities [C]//Medical image computing and computer assisted intervention - MICCAI 2021; 24th international conference. Strasbourg; Springer, 2021; 410-420.
- [8] AZAD R, KHOSRAVI N, MERHOF D. SMU-Net; style matching U-Net for brain tumor segmentation with missing modalities [C]//International conference on medical imaging with deep learning. Zurich; Springer, 2022; 48-62.
- [9] YANG H, SUN J, YANG L, et al. A unified hyper-GAN model for unpaired multi-contrast MR image translation [C]//Medical image computing and computer assisted intervention - MICCAI 2021; 24th international conference. Strasbourg; Springer, 2021; 127-137.
- [10] LEE D, MOON W J, YE J C. Assessing the importance of magnetic resonance contrasts using collaborative generative adversarial networks [J]. Nature Machine Intelligence, 2020, 2(1); 34-42.
- [11] SHARMA A, HAMARNEH G. Missing MRI pulse sequence synthesis using multi-modal generative adversarial network [J]. IEEE Transactions on Medical Imaging, 2019, 39(4); 1170-1183.
- [12] YAO T, HAN Y, WANG R, et al. Efficient discrete supervised hashing for large-scale cross-modal retrieval [J]. Neurocomputing, 2020, 385; 358-367.
- [13] CHEN C, DOU Q, JIN Y, et al. Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion [C]//Medical image computing and computer assisted intervention - MICCAI 2019; 22nd international conference. Shenzhen; Springer, 2019; 447-456.
- [14] 刘倩. 基于 Transformer 模型的多模态脑肿瘤 MRI 影像处理及应用 [D]. 武汉; 华中农业大学, 2023.
- [15] GONG S, ZHONG Y, MA W, et al. 3DSAM-adapter; holistic adaptation of SAM from 2D to 3D for promptable tumor segmentation [J]. Medical Image Analysis, 2024, 98; 103324.
- [16] ZHOU T, CANU S, VERA P, et al. Latent correlation representation learning for brain tumor segmentation with missing MRI modalities [J]. IEEE Transactions on Image Processing, 2021, 30; 4263-4274.
- [17] BAKAS S, REYES M, JAKAB A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge [J]. arXiv:1811.02629, 2018.
- [18] LI J, ZHANG L, ZHONG K, et al. A discrepancy-aware self-distillation method for multi-modal glioma grading [J]. Knowledge-Based Systems, 2024, 295; 111858.
- [19] ZHANG D, HUANG G, ZHANG Q, et al. Exploring task structure for brain tumor segmentation from multi-modality MR images [J]. IEEE Transactions on Image Processing, 2020, 29; 9032-9043.
- [20] QIU C, SONG Y, LIU Y, et al. MMMViT; multiscale multi-modal vision transformer for brain tumor segmentation with missing modalities [J]. Biomedical Signal Processing and Control, 2024, 90; 105827.