

基于 U-Net 和 Transformer 结合的不完整多模态脑肿瘤分割方法

汤占军¹, 蹇洪¹, 王健²

(1. 昆明理工大学信息工程与自动化学院, 昆明 650500; 2. 昆明理工大学电力工程学院, 昆明 650500)

摘要: 由于患者个体差异、采集协议多样性和数据损坏等因素, 现有基于磁共振成像 (Magnetic resonance imaging, MRI) 的脑肿瘤分割方法存在模态数据丢失问题, 导致分割精度不高。为此, 本文提出了一种基于 U-Net 和 Transformer 结合的不完整多模态脑肿瘤分割 (Incomplete multimodal brain tumor segmentation based on the combination of U-Net and Transformer, IM TransNet) 方法。首先, 针对脑肿瘤 MRI 的 4 个不同模态设计了单模态特定编码器, 提升模型对各模态数据的表征能力。其次, 在 U-Net 中嵌入双重注意力的 Transformer 模块, 克服模态缺失引起的信息不完整问题, 减少 U-Net 的长距离上下文交互和空间依赖性局限。在 U-Net 的跳跃连接中加入跳跃交叉注意力机制, 动态关注不同层级和模态的特征, 即使在模态缺失时, 也能有效融合特征并进行重建。此外, 针对模态缺失引起的训练不平衡问题, 设计了辅助解码模块, 确保模型在各种不完整模态子集上均能稳定高效地分割脑肿瘤。最后, 基于公开数据集 BRATS 验证模型的性能。实验结果表明, 本文提出的模型在增强型肿瘤、肿瘤核心和全肿瘤上的平均 Dice 评分分别为 63.19%、76.42% 和 86.16%, 证明了其在处理不完整多模态数据时的优越性和稳定性, 为临床实践中脑肿瘤的准确、高效和可靠分割提供了一种可行的技术手段。

关键词: 注意力机制; 脑肿瘤分割; 多模态; U-Net; Transformer

中图分类号: R739.41

文献标志码: A

Incomplete Multimodal Brain Tumor Segmentation Method Based on the Combination of U-Net and Transformer

TANG Zhanjun¹, JIAN Hong¹, WANG Jian²

(1. School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China;
2. School of Electric Power Engineering, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Given inherent variations among patients, discrepancies in imaging protocols, and potential data corruption, existing brain tumor segmentation methods based on magnetic resonance imaging (MRI) are often challenged by the issue of missing modality data, resulting in low segmentation accuracy. To address this, an innovative incomplete multimodal brain tumor segmentation method based on the combination of U-Net and Transformer (IM TransNet) is proposed. Firstly, a modality-specific encoder is developed for four distinct MRI modalities to enhance the model's ability to capture unique characteristics of each modality. Secondly, a dual-attention Transformer module is embedded within the U-Net to mitigate the

issue of incomplete information arising from missing modalities, thus alleviating the limitations imposed by long-range context interactions and spatial dependencies within the U-Net framework. Additionally, a skip-cross attention mechanism is incorporated into the U-Net's skip connections to dynamically focus on features from various hierarchical levels and modalities, effectively facilitating feature fusion and reconstruction even in the presence of missing modalities. Furthermore, an auxiliary decoding module is devised to counteract the training imbalance induced by missing modalities, ensuring that the model can consistently and effectively segment brain tumors across diverse subsets of incomplete modalities. Finally, the model's performance is validated on the publicly accessible BRATS dataset. Experimental results indicate that the proposed model attains average Dice scores of 63.19%, 76.42%, and 86.16% for enhancing tumor, tumor core, and whole tumor, respectively, highlighting its superiority and robustness in handling incomplete multimodal data. This approach offers a viable technical solution for accurate, efficient, and reliable brain tumor segmentation in clinical practice.

Key words: attention mechanism; brain tumor segmentation; multimodal; U-Net; Transformer

引言

脑肿瘤是一种在大脑或其他颅内组织中形成的异常生长物,其中大部分具有恶性特征,对患者的健康构成严重威胁^[1]。在医学图像分割领域^[2-3]中,对脑肿瘤的精准分割和识别在临床诊断、治疗规划以及预后评估阶段起着至关重要的作用。磁共振成像(Magnetic resonance imaging, MRI)作为常用的神经影像技术,广泛应用于脑部肿瘤的检查。在脑部MRI中,存在4种常见的模态数据:T1加权(T1)、对比增强T1加权(T1c)、T2加权(T2)和液体衰减反转恢复(Fluid attenuated inversion recovery, FLAIR)。它们各自在揭示脑肿瘤特征方面具有独特的优势,并且能够互补性地提供更全面的肿瘤信息^[4]。因此,使用互补信息的多模态数据可以显著提高临床评估的准确性和可靠性。

然而,在传统的肿瘤区域分割过程中,临床专家的手动注释是必不可少的环节。这一过程不仅耗时费力,而且在效率和成本方面存在明显的局限性。为了克服这些挑战,研究人员开发了基于深度学习的多模态脑肿瘤分割方法^[5-10]。尽管深度学习方法在完整模态数据下取得了显著的进展,但在实际情况中,由于成像数据的损坏、采集协议的差异以及患者自身条件的影响,往往无法获取到一组完整的MRI图像数据,这种模态的缺失问题使得现有的全模态分割方法应用于不完整数据集时会导致脑肿瘤分割性能严重下降。

针对该问题,研究人员提出了一些基于缺失模态的脑肿瘤分割方法,Havaei等^[11]通过计算可访问多模态特征的均值和方差来构建融合特征,进而用于解码端的分割任务。在此基础上,Dorent等^[12]进一步提出U-HVED模型,此模型利用多模态变分自编码器处理缺失模态,通过高斯求积的方式整合模态的均值和方差,并引入跳过连接来捕捉更为丰富的特征信息。Chen等^[13]则将多模态特征分解为模态特有的外观代码和模态无关的内容代码,通过融合内容代码形成共享特征表示,从而实现缺失模态下的任意模态重建以及肿瘤分割。Zhou等^[14]则引入注意力机制,以突出不同可访问模态中重要肿瘤特征。此外,Ding等^[15]还提出了一种区域感知的融合网络(Region-aware fusion network, RFNet),它能够自适应且有效地利用不同组合的多模态数据进行肿瘤分割。同时,还有一些研究探索了从完整模态到不完整模态的知识蒸馏方法^[16-18]。尽管这些方法取得了一定成果,但它们并未深入挖掘肿瘤区域与图像模态之间的内在联系,不同的图像模态呈现出各异的外观特征,对肿瘤区域的敏感度也有较大差异,从而导致肿瘤分割和识别精度不够。

鉴于CNN在视觉识别任务中的局限性,ViT (Vision Transformer)作为一种新架构,能够弥补这些不足。Chen等^[19]提出的TransUNet通过引入Transformer改进了U-Net,旨在解决CNN在建模长期依赖性方面的问题。然而,直接在三维图像上应用Transformer会增加计算复杂度和过拟合风险,特别是在小规模数据集中过拟合风险急剧增大。为了消除这一限制,Wang等^[20]提出了TransBTS,利用3D CNN作为编码器处理MRI图像切片,提取局部空间特征,同时减轻Transformer的计算负担。针对不完整多模态脑肿瘤分割,Zhang等^[21]提出了一种mmFormer模型,该模型利用模态间和模态内的关系来进行建模,以实现分割方法。然而,尽管Transformer在视觉处理领域取得了显著进展,但其在全模态脑肿瘤分割领域的应用仍有限,尤其是处理缺失模态的情况。

受上述工作的启发,本文提出U-Net和Transformer结合的不完全多模态脑肿瘤分割(Incomplete multimodal brain tumor segmentation based on the combination of U-Net and Transformer, IM TransNet)方法。首先,在编码阶段,使用单模态特定编码器模块,分别对脑肿瘤MRI的4个模态进行特定的编码,旨在提取出更为丰富的多模态特征。其次,在U-Net的跳跃连接中引入跳跃交叉注意力(Skip cross-attention, SCA)机制,使信息传递过程中能够动态关注不同层级和模态的特征。通过这种机制,即使在模态缺失时,也能融合不同分辨率和模态的特征,充分利用现有信息进行特征重建。此外,传统的U-Net结构需要完整的输入模态进行特征提取,模态缺失会导致特征不完整,从而影响模型性能,并难以有效捕捉全局长距离的上下文信息,尤其在处理复杂的三维图像时。因此,本文提出了双重注意力的Transformer模块来解决这些问题。最后,为解决模态缺失导致的训练不平衡问题,引入了一个辅助解码模块。在公开数据集BRATS2020和BRATS2021上进行了广泛的实验,结果证明了IM TransNet方法的有效性。

1 网络模型

1.1 整体架构

本文提出的IM TransNet结构如图1所示,其核心组件包括单模态特定编码器模块、交叉跳跃注意力模块、双重注意力的Transformer模块以及辅助解码模块。首先,单模态特定编码器模块通过交替堆叠普通三维卷积与空洞三维卷积,能够充分提取各个模态中丰富的肿瘤特征。随后,这些特征被送入双重注意力的Transformer模块中,该模块通过空间和通道上的双重注意力机制实现了对长距离依赖关系的有效建模。紧接着,交叉跳跃注意力模块在信息传递方面发挥了关键作用,它能够将底层空间信息高效地传递给各个解码器,从而在恢复细粒度细节的同时保持信息的完整性。最后,辅助解码模块为整个模型提供了强大的鲁棒性。在一个或多个模态缺失的情况下,该模块能确保模型的分割能力不会受到严重影响,从而使得IM TransNet在各种场景下都能表现出稳定的分割性能。

1.2 单模态特定编码器

在处理多模态数据的传统架构中,全模态数据通常被一并送入编码器进行统一编码。针对脑部MRI的4个模态,为了让每个模态的数据在编码过程中保持其独特性,设计了一个单模态特定编码器对其进行特征提取,从而更有效地捕捉各模态特有的信息。

具体来说,本文所构建的五级单模态特定编码器通过堆叠卷积块来实现,如图2所示。每一级的卷积块都遵循特定的结构:第一个卷积块融合了实例归一化层、Leaky ReLU层以及普通三维卷积层,其中卷积层的内核大小设为3,填充大小设为1,步长设为1,以确保特征的有效提取;第二个卷积块则包含实例归一化层、Leaky ReLU层和空洞三维卷积层,为保证空洞卷积的输出与普通三维卷积的输出在尺寸上能完美匹配,将空洞卷积层的内核大小设为5,填充大小设为2,步长设为1。这种设置使得模型能

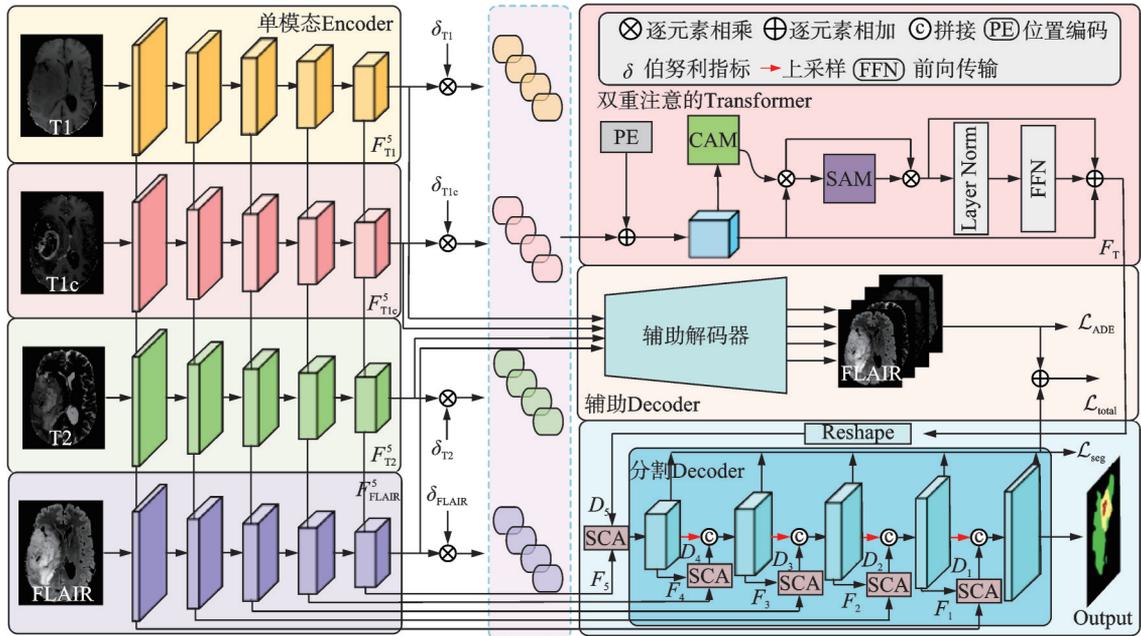


图1 IM TransNet整体结构

Fig.1 Overall structure of IM TransNet

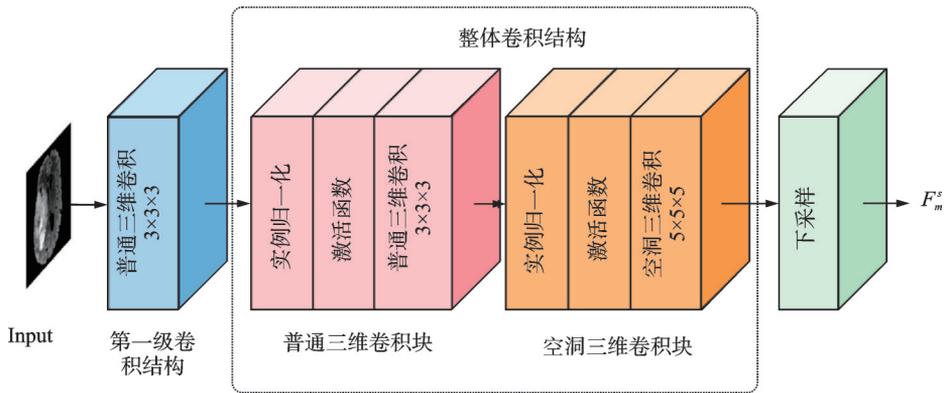


图2 单模态特定编码器的卷积块结构

Fig.2 Convolutional block structure of single modality-specific encoder

能够在卷积过程中拥有更大的感受野,从而捕获到图像中更为丰富的上下文信息。值得注意的是,第一级中的第一个卷积结构包含实例归一化层、Leaky ReLU层以及普通三维卷积层,作为特征提取的起点。在两个连续的卷积模块之间,采用步长为2的普通三维卷积层来对上一级得到的特征图进行下采样,这一步骤有助于逐步缩小特征图的尺寸,同时保留关键信息。在编码器的每一级,将滤波器的数量设计为16、32、64、128和256。

给定输入图像为 $X_m \in \mathbb{R}^{1 \times D \times H \times W}$, $m \in M$, 用 $M = \{T1, T1c, T2, FLAIR\}$ 表示模态的完整集合。利用单模态特定编码器对其进行处理,以生成具有局部上下文的紧凑特征映射。这一步骤旨在从图像中提取出各模态的关键信息,并保持其局部空间关系,为后续处理提供丰富的特征表达。单模态特定编码器 E_m^{Conv} 产生每个模态的局部上下文的特征图,可定义为

$$F_m^s = E_m^{Conv}(X_m; \theta_m^{Conv}) \tag{1}$$

式中: $F_m^s \in \mathbf{R}^{C \times \frac{D}{2^{s-1}} \times \frac{H}{2^{s-1}} \times \frac{W}{2^{s-1}}}$; $s \in \{1, 2, 3, 4, 5\}$ 为单模态特定编码器中的级数; C 为通道维度; D 、 H 和 W 分别为体积的深度、高度和宽度; θ_m^{Conv} 为 E_m^{Conv} 的计算参数。

1.3 跳跃交叉注意力模块

在U-Net结构中,编码器和解码器之间巧妙地运用了跳跃连接,这一设计不仅可实现网络低级特征与高级特征的融合,而且能有效地防止信息在网络传递过程中的丢失。然而,当处理三维图像时,尤其是在模态缺失的情况下,直接将编码器和解码器相连,不仅会增加网络的计算复杂度,还会使跳跃连接中的信息不完整,导致解码器难以重建高质量的特征。为了避免这一问题,本文采用高效注意力构建了一个SCA模块,其结构如图3所示。

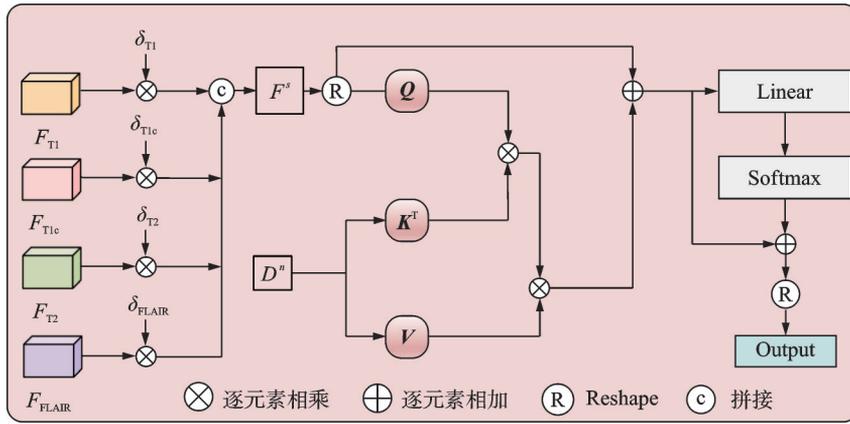


图3 跳跃交叉注意力结构

Fig.3 Skip cross-attention structure

传统的自注意力机制在计算复杂度上达到了 $O(N^2)$,这一特性导致其在处理高分辨率图像时效率低下,无法充分发挥其优势。然而,高效注意力机制则能有效避免这一问题,它首先对键和查询进行规范化处理,以消除尺度差异,提升计算的稳定性;随后,将处理后的键与值进行相乘操作,从而得到一个全局上下文向量;最后,将这个全局上下文向量与查询相乘,生成新的表示,这一过程进一步丰富了特征的语义信息。与传统的点积注意力机制不同,高效注意力机制并不需要预先计算输入特征中每对位置之间的相似性,相反,它将键表示为 d_k 个注意力图 K_j^T ,其中 j 指输入特征中的位置。这些注意力图能够反映整个输入特征的语义分布,而不仅仅局限于特定位置的相似性。这种转换方式不仅降低了注意力机制的计算复杂性,使得处理高分辨率图像时更加高效,同时也保持了较高的表征能力,使得模型能够更好地理解和处理图像信息。高效注意力的编码过程定义为

$$E(Q, K, V) = \rho_q(Q)(\rho_k(K^T)V) \tag{2}$$

式中 $\rho_q(\cdot)$ 和 $\rho_k(\cdot)$ 分别为查询和键的归一化函数。

本文所提出的SCA模块不仅能够有效地提供空间信息,使模型更加专注于肿瘤等关键信息,而且在解码器恢复细粒度细节的过程中,能够生成精确的输出掩码。更重要的是,SCA模块在保持网络性能的同时,降低了计算复杂度,提高了处理效率。在SCA模块中,通过对键、查询和值使用不同的输入,使SCA模块实现了对编码器层输出 F^s 和解码器层输出 D^n 的有效融合。为确保 F^s 和 D^n 在嵌入维度上的一致性,采用线性层对 F^s 进行缩放,而将 F^s 作为查询的输入, D^n 作为键和值的输入,则是在高效注意力块内对多级特征进行建模的关键步骤。其建模过程定义为

$$F^s = [\delta_{T1} F_{T1}^s, \delta_{T1c} F_{T1c}^s, \delta_{T2} F_{T2}^s, \delta_{FLAIR} F_{FLAIR}^s] W \tag{3}$$

$$X = FC(D^n), K, V = \text{proj}(X), Q = \text{proj}(F^s), E = \rho_v(V) \rho_k(K^T) Q \quad (4)$$

式中: W 为权重矩阵; $s \in \{1, 2, 3, 4, 5\}$ 为单模态编码器的级数; $n \in \{1, 2, 3, 4, 5\}$ 为分割解码器的级数; F^s 为编码层的输出; D^n 为解码层的输出; $\rho_v(\cdot)$ 为归一化函数; $\text{proj}(\cdot)$ 为投影函数; $FC(\cdot)$ 为全连接函数; $\delta_m \in \{0, 1\}$ 为伯努利指标, 在模型训练过程中, 通过将 δ_m 随机置 0 来模拟模态的缺失。

1.4 双重注意力的Transformer结构

基于U-Net设计的网络架构在处理大尺寸图像和缺失模态时存在一定的局限性。首先, 它主要依赖完整的输入模态进行特征提取和分割, 因此缺失模态会导致特征提取不完整, 从而影响分割的准确性; 其次, U-Net主要依赖卷积操作, 虽然能捕捉局部特征, 但难以捕捉长距离依赖关系, 尤其是在处理复杂的三维脑肿瘤图像时。这些局限性对图像分割性能产生了巨大影响, 使得分割结果可能不够精确或全面。为了解决上述问题, 本文提出了一种双重注意力的Transformer结构, 如图4所示。它包括1个双重注意力机制和1个前馈网络

(Feed forward network, FFN), 其中双重注意力机制由通道和空间注意力以级联的方式构成。通道注意力机制的核心在于自适应地调整不同通道的权重, 以凸显输入数据中关键通道的重要性, 同时削弱冗余信息的影响, 从而优化模型性能。而空间注意力机制则侧重于捕捉输入数据中不同位置的重要性差异, 通过增强对关键空间区域的关注度, 使得模型能够更有效地处理空间维度的信息。这两种注意力机制的结合, 不仅提升了模型对关键信息的处理能力, 还使得模型在复杂的数据环境中表现出更高的鲁棒性和准确性。

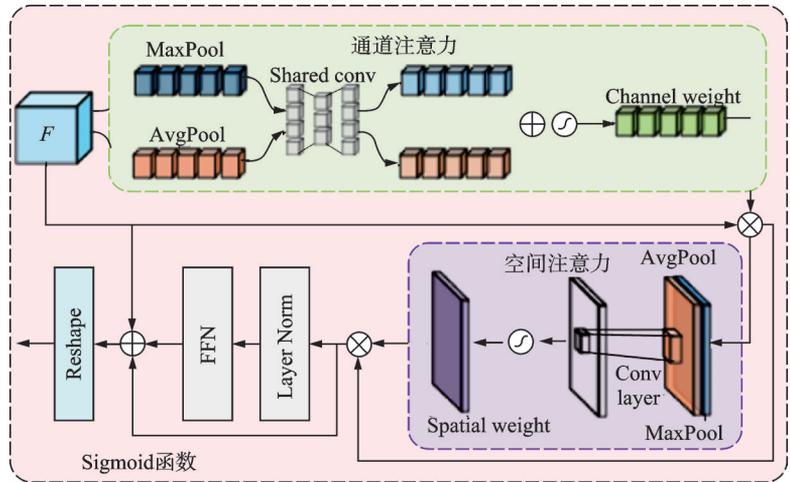


图4 双重注意力Transformer结构

Fig.4 Transformer structure with dual attention

Transformer以序列到序列的方式进行数据处理, 因此需要将卷积编码器生成的局部特征图 F^{5*} 展平成一维序列, 并通过线性投影映射到输出空间。然而, 这一展平过程会导致空间信息的损失。为了弥补这一缺陷, 将局部特征图 F^{5*} 输入双重注意机制的Transformer之前, 借助标记器实施可学习的位置嵌入 P_m 操作。这种位置嵌入能够捕捉并保留原始特征的空间信息, 通过与展平后的特征进行逐元素求和的融合, 使得空间信息的完整性在一定程度上得以恢复。此过程被定义为

$$F^{5*} = [\delta_{T1} F_{T1}^5, \delta_{T1c} F_{T1c}^5, \delta_{T2} F_{T2}^5, \delta_{FLAIR} F_{FLAIR}^5] W \quad (5)$$

$$F = F^{5*} W_m + P_m \quad (6)$$

式中 W_m 为线性投影的权重。

双重注意力Transformer结构的精髓在于对通道和空间信息的精准提取。首先, 该结构将融合后的特征图 F 输入至通道注意力机制进行全局平均池化和最大池化操作, 以提取特征图的通道维度信息。接着, 这些信息被送入卷积层以学习并捕捉各通道的重要性, 经过激活函数的映射后, 生成通道注意力权重矩阵 W_c , 将 W_c 与特征图 F 相乘产生通道注意力机制的输出 F_c 。然后, F_c 作为输入进入空间注意力机制。在这一阶段, 再次执行通道维度上的全局平均池化和最大池化操作, 以捕获空间维度的特征。

这些特征经过融合后,通过卷积操作进一步提取和整合特征图中的信息。再经过激活函数的映射处理,得到空间注意力权重矩阵 W_s 。将 W_s 与 F_c 相乘产生空间注意力机制的输出 F_s 。最后, F_s 经过线性层和前向传输网络的处理后,与原始特征 F 和空间注意力机制的输出 F_s 进行残差连接。这一步骤有效地结合了原始特征、通道注意力以及空间注意力的优势,从而生成了双重注意力 Transformer 的最终输出结果 F_T 。整个流程充分利用了双重注意力机制在通道和空间两个维度上的特长,显著提升了模型在图像分割任务上的性能,其编码过程定义为

$$W_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (7)$$

$$F_c = W_c(F) \otimes F \quad (8)$$

$$W_s(F_c) = \sigma(f^{7 \times 7}([\text{AvgPool}(F_c); \text{MaxPool}(F_c)])) \quad (9)$$

$$F_s = W_s(F_c) \otimes F_c \quad (10)$$

$$F_T = \text{FFN}(\text{LN}(F_s)) + F_s + F \quad (11)$$

式中: $W_c(\cdot) \in \mathbb{R}^{C \times 1 \times 1}$; $W_s(\cdot) \in \mathbb{R}^{1 \times H \times W}$; σ 为 Sigmoid(\cdot) 激活函数; $\text{MLP}(\cdot)$ 为卷积操作; $\text{AvgPool}(\cdot)$ 为平均池化操作; $\text{MaxPool}(\cdot)$ 为最大的池化操作; “ \otimes ” 表示逐元素相乘; $f^{7 \times 7}$ 为 7×7 的卷积算子; $\text{FFN}(\cdot)$ 为一个具有 GeLU 激活的双层感知器; $\text{LN}(\cdot)$ 表示线性计算。

1.5 辅助解码器

分割网络通常依赖于特定的模态来识别脑肿瘤区域,然而,一旦这种模态缺失,模型的分割性能会大幅下降。为解决这一问题,本文引入了辅助解码器模块,其结构如图 5 所示,该模块将卷积编码器的输出共享并融入其中。在辅助解码器的设计中,采用采样因子为 2 的上采样操作,以增强特征图的分辨率,随后通过卷积层进一步细化特征。此外,还引入了跳跃连接,将其与一个配备 $3 \times 3 \times 3$ 大小过滤器的卷积层相结合,以充分利用多尺度的特征信息。每级解码层都配备了实例归一化层和 Leaky ReLU 激活函数层,以加速训练过程并提高模型的鲁棒性。

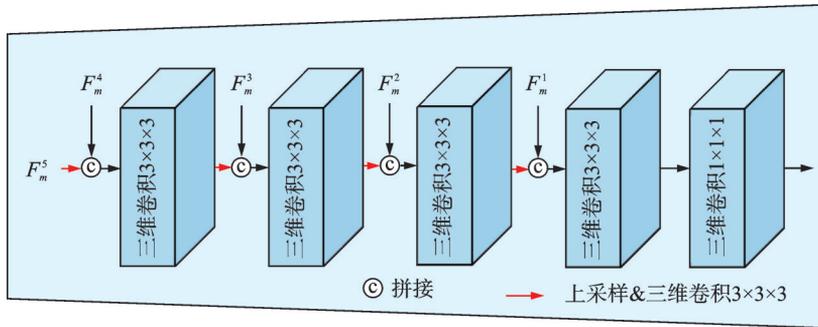


图 5 辅助解码器结构

Fig.5 Auxiliary decoder structure

为了提升模型的性能表现,引入 L_1 范数作为辅助解码损失的度量标准。这一范数有效地量化了辅助解码器输出数据与原始数据之间的细微差异。通过不断缩小辅助解码损失,可使网络模型得到强大的激励,以便更精准地捕捉并恢复原始数据的核心特征。本文将辅助解码的损失定义为

$$\mathcal{L}_{\text{ADE}} = \sum_{m \in M} \|\hat{x}_m - x_m\|_1 \quad (12)$$

式中: \hat{x}_m 为辅助解码器输出的预测图像; x_m 为输入的真实图像; $M = \{T1, T1c, T2, \text{FLAIR}\}$ 。

1.6 分割解码结构

在像素级分割任务中,分割解码结构扮演着至关重要的角色,它承担着将高级语义信息精准地映射回原始输入尺寸的关键任务。首先,对双重注意力Transformer的输出 F_T 进行线性变换得到 D^5 ,这一步骤为后续的解码过程奠定了坚实的基础。接着,将 D^5 与第五级编码器的输出 F^5 共同送入SCA模块中,通过该模块的处理能够进一步提取和整合特征信息。随后,处理后的结果被送入分割解码结构,在此本文利用上采样操作来逐步提升特征图的空间分辨率,确保在恢复过程中不丢失关键信息。为进一步提升分割的准确性,在各级编码器和分割解码结构之间都采用了SCA模块进行连接。这种连接方式巧妙地利用注意力权重,使模型能够更好地保留低层细节信息,从而进一步增强分割任务的性能。

1.7 损失函数

本文采用加权交叉熵(Weighted cross-entropy, WCE)和Dice相似系数(Dice similarity coefficient, DSC)相结合的策略,作为分割网络的优化目标。这一策略旨在通过精细调整网络参数,提升像素分类的准确性。损失函数被应用于每个解码器层,以逐层引导网络在解码过程中逐渐细化分割结果。具体来说,分割损失可定义为

$$\mathcal{L}_{\text{seg}} = \sum_{i=1}^N \mathcal{L}_{\text{WCE}}^i + \mathcal{L}_{\text{Dice}}^i \quad (13)$$

式中 $N=5$ 为分割解码器的级数。

L_{WCE} 和 L_{Dice} 可分别定义为

$$\mathcal{L}_{\text{WCE}}(X, Y) = - \sum_{k=1}^K w_k \cdot Y_k \cdot \log X_k \quad (14)$$

$$\mathcal{L}_{\text{Dice}}(X, Y) = 1 - \frac{2 \cdot |X \cap Y| + \epsilon}{|X| + |Y| + \epsilon} \quad (15)$$

式中: K 为类别的数量; w_k 为第 k 个类别的权重; X_k 为第 k 个类别的模型的预测值; Y_k 表示第 k 个类别的真实值; $|X|$ 为预测的肿瘤分割掩码; $|Y|$ 为标签的肿瘤分割掩码; $|X \cap Y|$ 为预测的肿瘤分割掩码与标签的肿瘤分割掩码的交集; ϵ 为防止零除错误而引入的平滑项。总体损失为

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{ADE}} + \mathcal{L}_{\text{seg}} \quad (16)$$

2 实验部分

2.1 实验细节

在实验中,本文使用Windows 10操作系统,并在配备12 GB显存的NVIDIA GeForce RTX 3060显卡上基于PyTorch 1.8.1进行模型的训练、验证和测试。为增强模型的泛化能力,数据预处理阶段采用多种数据增强方法,包括中心裁剪,以0.5的概率对冠状面、矢状面和轴向面进行随机镜像翻转,以及在 $[-0.1, 0.1]$ 范围内应用随机强度偏移。

在训练过程中,将图像裁剪为 $128 \times 128 \times 128$ 体素大小的标准化输入,优化了数据处理效率和模型的适应性。同时,使用Adam优化器对网络进行优化,初始学习率设为0.000 2,训练共进行400轮。然而,IM TranNet模型的参数量约为 $42.42\text{M} \times 10^6$ 个,在单个12 GB显存的GPU上训练时,显存消耗较大,因此必须采用合适的优化策略来保证训练的稳定性 and 效率。为解决这一问题,将Batch size设置为1,通过累积多个小批量的梯度来模拟大Batch size的训练效果。这种方法降低了每次前向传播的显存占用,使模型在有限的显存资源下能够更频繁地更新,从而确保IM TranNet能够在单个12 GB显存的GPU上高效、稳定地进行训练。

2.2 数据集

本文采用国际医学图像计算和计算机辅助干预协会提供的BRATS2020和BRATS2021数据集,以验证所提出的方法的有效性。BRATS2020数据集包含369个训练案例,它们被划分为训练集、验证集和测试集,具体比例为258:37:74;而BRATS2021数据集则包含更为丰富的1251个训练案例,本文从中随机选取了450个案例,并按照315:45:90的比例分为训练集、验证集和测试集。为确保数据的统一性和可比性,对这些数据集进行预处理,包括模版配准、头骨剥离以及重新采样到统一的分辨率(1 mm^3)。

在BRATS数据集中,每个训练案例均包含4种模态的图像,分别是T1、T1c、T2和FLAIR,以及对应的真实标签(Label),如图6所示。在标签图像中,红色区域代表非增强肿瘤,对应标签1;绿色区域代表水肿,对应标签2;黄色区域代表增强肿瘤,对应标签4;而背景区域则对应标签0。在评价过程中,将这4种结构(水肿、增强肿瘤、坏死和非增强肿瘤核心)合并为3个主要的肿瘤区域进行评价,分别为增强肿瘤区域(Enhancing tumor, ET)、肿瘤核心区域(Tumor core, TC)和肿瘤整体区域(Whole tumor, WT)。具体来说,ET对应标签图像中的黄色区域;TC对应标签图像中红色和黄色区域的合并部分;而WT则对应红色、绿色和黄色这3个区域的合并部分。

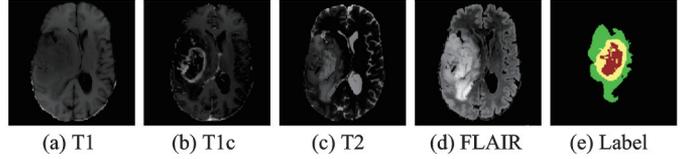


图6 不同模态的脑肿瘤图像和标签

Fig.6 Brain tumor images and labels from different modalities

2.3 评价指标

Dice相似系数是语义分割任务中的重要评价指标之一。它能够有效衡量两个样本之间的相似程度,为模型性能的评估提供有力的依据。具体而言,DSC指标的数值越接近1,就意味着模型在分割任务中的精度越高,即预测结果与真实标签之间的重合度越高,其定义为

$$\text{DSC}(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (17)$$

式中: $|X|$ 为预测肿瘤的分割掩码; $|Y|$ 为标签肿瘤的分割掩码; $|X \cap Y|$ 为预测肿瘤的分割掩码和标签肿瘤的分割掩码的交集。

95%豪斯多夫距离(95% Hausdorff distance, HD95)是衡量两个点集之间相似度的一种指标。它特别常用于图像分割领域,通过度量预测结果和真实标签的边界之间的相似度来衡量分割性能,其数值越小,表示边界分割越精确,可定义为

$$\text{HD95}(X, Y) = \max \left\{ Q_{95\%} \left(\inf_{y \in Y} d(x, y) \mid x \in X \right), Q_{95\%} \left(\inf_{x \in X} d(y, x) \mid y \in Y \right) \right\} \quad (18)$$

式中: X 为预测肿瘤的分割掩码; Y 为标签肿瘤的分割掩码; $Q_{95\%}$ 表示计算距离集合中的第95个百分点数; $d(x, y)$ 表示点 x 和 y 之间的欧氏距离;inf表示取下确界(即最小的最大距离)。

本文采用DSC和HD95评估模型的分割精度。

2.4 对比实验

在数据集BRATS2021上,本文将IM TransNet与U-HVDE^[12]、ACN^[18]、RFNet^[15]、mmFormer^[21]和M³AE^[22]进行了比较,以DSC和HD95为该实验的评价指标,比较结果如表1所示。在表1中,U-H表示U-HVED;RFN表示RFNet;mmF表示mmFormer;M³A表示M³AE;M代表模态;Avg表示14种不同缺失模态和1种完整模态的平均值;“√”和“×”分别表示可用和缺失的模态;最佳结果以粗体显示。

表1 本文方法与最先进方法在数据集BRATS2021上的性能对比

Table 1 Performance comparison of the proposed method with state-of-the-art methods on BRATS2021 dataset

M				ET											
T1	T1c	T2	F	U-H		ACN		RFN		mmF		M ³ A		Ours	
				DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm
√	×	×	×	14.98	20.85	39.97	9.23	37.40	11.24	38.56	10.54	37.56	11.56	38.75	9.68
×	√	×	×	60.39	12.43	73.95	2.20	74.18	3.87	67.04	3.86	72.77	3.21	79.27	2.86
×	×	√	×	26.09	11.3	44.06	8.21	46.63	8.06	42.34	7.78	46.61	7.51	42.04	8.02
×	×	×	√	22.64	16.43	41.82	8.66	37.21	10.34	40.64	10.73	35.63	11.02	40.12	10.46
√	√	×	×	64.32	8.01	71.70	3.56	74.70	2.76	68.80	2.36	74.75	2.65	80.06	2.12
√	×	√	×	27.10	10.11	46.19	6.45	45.31	6.56	43.72	6.79	48.73	6.01	46.89	6.21
√	×	×	√	24.14	13.35	41.17	9.92	39.82	7.73	44.21	9.34	41.26	8.12	44.94	7.60
×	√	√	×	67.36	5.87	72.05	3.43	73.27	2.70	71.04	2.46	74.38	2.67	80.34	1.95
×	√	×	√	67.55	10.45	72.10	4.68	74.19	4.09	71.17	3.44	75.60	3.47	79.92	3.14
×	×	√	√	30.48	9.25	44.72	8.89	47.17	6.80	45.58	8.21	45.48	9.92	45.60	8.20
√	√	√	×	68.74	5.01	72.20	2.24	74.26	3.15	70.96	2.32	75.45	2.21	80.53	2.01
√	√	×	√	68.57	6.89	71.25	5.35	74.23	3.34	70.54	3.35	74.56	3.95	80.35	3.27
√	×	√	√	30.81	8.84	41.66	11.32	48.10	5.76	46.39	7.44	48.85	5.78	48.92	5.42
×	√	√	√	68.91	7.23	71.09	5.03	75.24	4.12	70.96	3.35	73.81	4.01	80.14	3.05
√	√	√	√	69.91	5.98	72.48	4.98	76.80	4.04	70.23	3.31	75.50	4.55	80.11	1.96
Avg				47.46	10.13	58.42	6.28	59.90	5.64	57.47	5.69	60.06	5.78	63.19	5.06
M				TC											
T1	T1c	T2	F	U-H		ACN		RFN		mmF		M ³ A		Ours	
				DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm
√	×	×	×	40.79	24.54	67.79	10.78	65.50	12.14	57.20	13.03	63.16	11.56	62.29	12.14
×	√	×	×	65.38	18.86	79.52	7.42	81.03	6.91	80.66	7.32	81.97	6.56	86.38	6.21
×	×	√	×	53.16	13.15	70.34	9.03	69.57	8.18	59.47	9.97	66.98	9.12	62.35	9.67
×	×	×	√	51.37	20.25	67.70	10.41	66.13	13.01	57.82	12.10	65.46	12.97	59.67	12.68
√	√	×	×	69.84	13.91	81.29	8.84	82.56	4.65	83.53	5.18	83.46	8.54	87.52	4.50
√	×	√	×	56.28	11.74	72.02	8.33	72.31	7.30	62.80	8.66	71.83	8.57	67.64	8.93
√	×	×	√	55.08	15.84	67.66	12.64	72.29	11.03	62.96	9.92	70.81	11.14	66.64	11.46
×	√	√	×	74.58	9.37	82.83	7.13	82.95	4.30	84.76	5.09	84.26	4.30	80.15	5.79
×	√	×	√	74.31	14.56	81.38	7.63	83.41	3.43	83.24	6.67	85.43	8.14	86.91	5.35
×	×	√	√	59.17	11.46	70.23	10.43	72.94	8.81	63.11	8.80	70.91	10.23	65.64	10.89
√	√	√	×	76.57	7.96	84.12	8.38	83.60	4.54	84.68	4.81	84.45	6.72	88.24	4.12
√	√	×	√	75.83	9.30	81.65	10.37	84.12	5.56	83.92	6.73	84.14	8.92	87.73	5.10
√	×	√	√	60.79	10.35	68.08	11.62	74.49	8.91	65.05	8.43	72.72	9.56	69.11	10.03
×	√	√	√	76.73	9.23	81.75	11.01	83.85	5.64	84.16	5.87	84.62	9.03	88.10	5.02
√	√	√	√	77.77	8.31	83.42	8.75	83.85	5.52	84.14	5.79	84.52	9.06	88.04	5.12
Avg				64.51	13.26	75.98	9.52	77.24	7.33	73.16	7.89	76.98	8.96	76.42	7.80

续表

M				WT											
				U-H		ACN		RFN		mmF		M ³ A		Ours	
T1	T1c	T2	F	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm	DSC/ %	HD95/ mm
√	×	×	×	57.01	24.22	77.41	11.58	76.35	11.20	70.81	11.79	74.45	11.24	74.77	11.20
×	√	×	×	59.94	20.92	78.74	8.72	73.82	9.73	70.81	11.06	75.45	8.81	75.66	8.95
×	×	√	×	79.51	12.45	84.21	11.43	83.45	6.75	83.17	6.68	75.80	11.89	84.52	6.61
×	×	×	√	81.36	18.23	86.31	7.78	86.70	8.54	85.58	7.79	86.79	7.78	86.81	7.54
√	√	×	×	65.33	18.15	79.61	10.06	79.87	8.13	75.53	9.11	77.28	8.59	78.64	8.19
√	×	√	×	81.65	9.42	85.30	7.89	86.87	5.90	84.18	6.48	86.70	7.42	85.81	6.32
√	×	×	√	83.38	13.44	86.29	11.83	89.37	5.26	87.06	6.25	89.10	4.98	89.32	4.90
×	√	√	×	82.06	8.54	85.93	6.96	86.64	5.75	84.97	6.08	86.35	6.82	86.93	5.51
×	√	×	√	85.54	13.15	87.26	7.53	89.56	5.30	87.81	5.69	89.71	5.91	90.19	4.83
×	×	√	√	86.82	9.43	87.11	7.12	89.62	5.21	87.81	5.58	88.91	7.74	89.77	5.56
√	√	√	×	82.93	7.25	85.83	11.30	87.33	6.04	85.28	5.24	85.73	7.01	87.75	4.96
√	√	×	√	86.49	11.40	87.12	10.03	90.26	4.83	88.23	5.32	88.94	8.21	90.29	4.64
√	×	√	√	87.68	7.49	86.51	10.08	90.30	4.40	88.26	5.47	89.97	8.45	90.15	5.17
×	√	√	√	88.15	7.79	87.44	7.56	90.46	5.40	89.05	5.31	90.28	5.78	90.92	5.10
√	√	√	√	88.40	6.23	87.77	8.11	90.40	5.07	89.16	5.10	90.11	5.82	90.88	4.49
Avg				79.57	12.54	84.85	9.20	86.06	6.50	83.84	6.86	85.03	7.76	86.16	6.26

U-HVED引入多模态变分自动编码器,旨在将不完整的多模态图像投影到共享的潜在空间中。ACN则利用对抗学习来对齐潜在空间表示,从而增强从完整到缺失模态的信息提取能力。RFNet进一步对区域和模态之间的关系进行建模,以实现自适应的特征聚合。M³AE提出了多模态掩蔽自动编码器,专注于丢失模态下的鲁棒多模态表示自监督学习。然而,这些方法都没有考虑到模型的长距离依赖的问题。而mmFormer利用模态间和模态内的关系进行分割建模,以实现远程的依赖。尽管如此,这些现有方法在处理低级特征以及长距离依赖时仍显不足。本文通过结合SCA模块和双重注意力Transformer结构,使该方法不仅解决了远程依赖关系的建模问题,还能够在解码时更准确地还原低级特征中的掩码结构。如表1所示,所提出的模型在ET、TC和WT上的平均Dice评分分别为63.19%、76.42%和86.16%,而平均HD95距离分别达到5.06、7.80和6.26 mm。与mmFormer相比,本文模型在ET、TC和WT上的平均Dice评分分别提高了5.72%、3.26%和2.32%,平均HD95距离分别下降了0.63、0.09和0.24 mm。本文通过对15种不同模态缺失情况的对比实验表明,该方法在性能上均优于其他现有方法。

为了评估本文所提出方法的鲁棒性,在BRATS2020数据集上也进行了实验。本文比较了U-HVDE^[12]、RobustSeg^[13]、RFNet^[15]和mmFormer^[21]等4种先进模型。如表2所示,本文使用DSC和HD95作为评价指标,所提出的模型在ET、TC和WT上的平均DSC评分分别为64.23%、80.11%和87.56%,而平均HD95距离分别达到5.42、6.31和5.91 mm。与RFNet相比,本文模型在ET、TC和WT上的平均DSC评分分别提高了2.80%、2.02%和0.98%,平均HD95距离分别下降了0.21、1.01和0.58 mm。实验结果表明,该方法在增强肿瘤和肿瘤核心区域方面取得了较好的改善。

本文对于RFNet^[15]、mmFormer^[21]以及本文所提出的模型IM TransNet,在仅使用1种模态(FLAIR)、2种模态组合(T1c+FLAIR和T2+FLAIR)、3种模态组合(T1c+T2+FLAIR)以及全模态

表 2 本文方法与最先进方法在数据集 BRATS2020 上的性能对比

Table 2 Performance comparison of the proposed method with state-of-the-art methods on BRATS2020 dataset

模型	DSC/%				HD95/mm			
	ET	TC	WT	AVG	ET	TC	WT	AVG
U-HVED	47.85	67.16	80.51	65.17	11.85	14.32	13.21	13.13
RobustSeg	54.56	73.89	84.10	70.85	11.04	14.62	14.18	13.28
RFNet	61.43	78.09	86.58	75.37	5.63	7.32	6.59	6.51
mmFormer	47.71	73.18	83.39	68.09	11.55	9.65	7.89	9.70
Ours	64.23	80.11	87.56	77.30	5.42	6.31	6.01	5.91

数据(Complete)的情况下,进行分割结果的可视化展示,如图7所示。可视化结果表明,在大多数情况下,本文的方法可以获得准确的分割结果。与此同时,进一步分析了不同模态缺失的影响。从图7可以看出,FLAIR模态在整个肿瘤分割中起着重要作用,即使在缺少其他3种模态的情况下,FLAIR模态也可以获得更准确的分割结果。在面临两种模态数据缺失的情况下,黄色区域的分割结果与真实标签之间仍呈现出明显的差异。然而,相比之下,本文所提出的方法所生成的预测掩码表现出了更高的精确度。不仅如此,与红色区域相比,本文方法在性能上通常优于其他两种模型。随着可利用模态数量的逐步增加,本文方法对于所有肿瘤区域的分割精度也愈发接近真实标签,展示了其在处理复杂情境下的卓越能力和高适应性。

为了更加直观地验证所提出模型的有效性,本文展示了15个单一和组合模态的分割场景,如图8所示。在各种模态数据可能存在缺陷的情况下,本文的模型依然展现出了出色的分割性能。从图8可

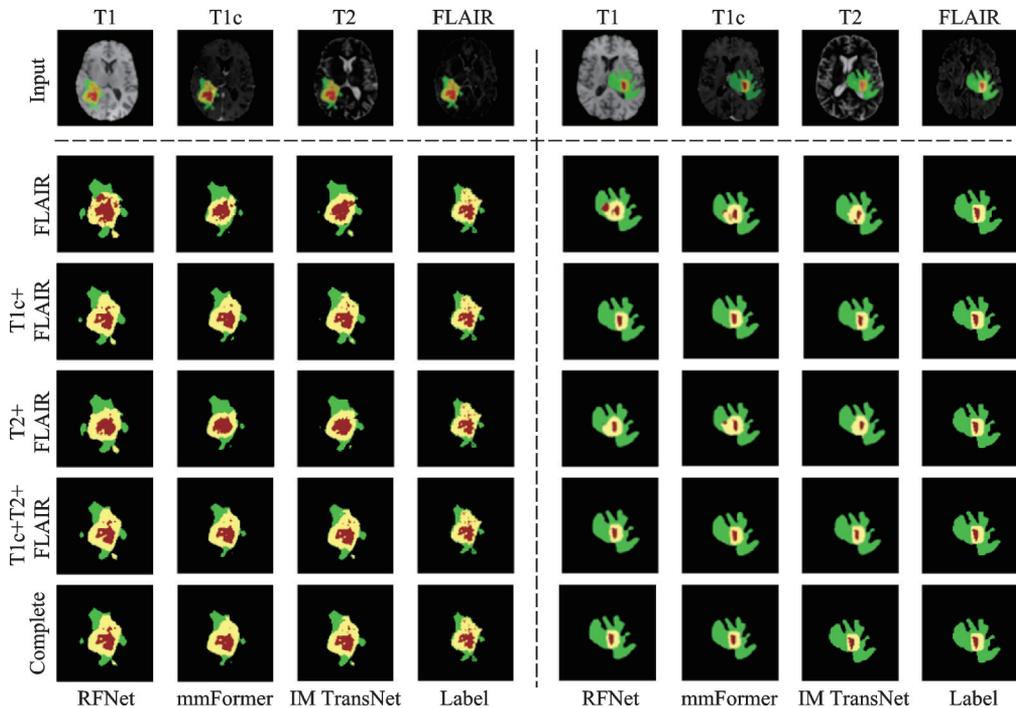


图7 本文方法与RFNet和mmFormer的分割对比图

Fig.7 Comparison of segmentation results of the proposed method with RFNet and mmFormer

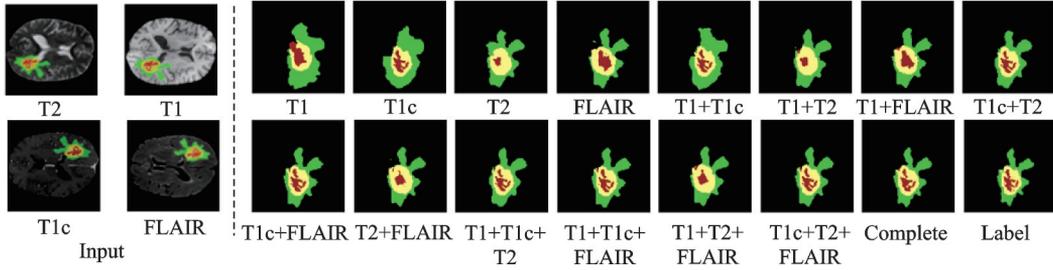


图8 15个单一和组合模态分割结果图

Fig.8 Segmentation results of 15 individual and combined modalities

以清晰看到,即便是在仅有 T2 或 FLAIR 单一模态的情况下,模型依然能够精准地完成分割任务。而随着可利用模态数量的逐步增加,模型的分割精度也随之显著提升,这进一步验证了其在处理多模态数据时的强大能力。

2.5 消融实验

本文采用 3D U-Net^[23] 网络架构作为基准模型构建了一个新型网络模型 IM TransNet,以充分利用三维图像中的空间信息实现数据的高效传播。为进一步验证本文提出的 IM TransNet 网络模型中各模块的有效性,在 BRATS2021 数据集上进行了一系列消融实验,实验结果如表 3 所示。表 3 中,用单模态特定编码器来替换 3D U-Net 的通用编码器得到模型 1;在模型 1 的基础上引入跳跃交叉注意力机制得到模型 2;在模型的基础上加入双重注意力的 Transformer 得到模型 3;在模型 3 中引入辅助解码器得到模型 4。图 9 展示了消融实验中不同组合分割结果的可视化结果,其中包括使用 4 种不同输入模式的分割结果。左栏显示了 4 个输入通道(T1、T1c、T2 和 FLAIR);右栏显示了不同模型在各种可用通道组合下的分割结果。

表 3 本文模型的消融实验结果

Table 3 Ablation experimental results of the proposed model

模型 编号	基准 模型	单模态 特定编 码器	交叉 注意 力	双重注意力 的 Trans- former	辅助 解码 器	DSC/%				HD95/mm			
						ET	TC	WT	Avg	ET	TC	WT	Avg
0	✓	×	×	×	×	57.38	70.78	83.21	70.45	7.64	10.64	9.93	9.40
1	✓	✓	×	×	×	59.21	72.34	83.56	71.70	7.44	9.41	8.96	8.60
2	✓	✓	✓	×	×	60.23	73.33	84.25	72.60	6.32	9.35	8.16	7.94
3	✓	✓	✓	✓	×	61.89	74.97	85.28	74.04	5.49	8.64	7.79	7.31
4	✓	✓	✓	✓	✓	63.19	76.42	86.16	75.25	5.06	7.80	6.26	6.37

模型 1 采用单模态特定编码器来替代 3D U-Net 网络架构中的通用编码器,使其能够从多模态医学图像中提取出特定模态特征。实验结果显示,模型 1 相较于基准模型在 BRATS2021 数据集上的平均 DSC 评分提高了 1.25%,而平均 HD95 距离下降了 0.80 mm。模型 2 是在模型 1 的基础上增加了一个跳跃交叉注意力机制以连接编码器和解码器,使得模型能够在医学图像分析中对关键信息的关注度和空间信息的传递效率。实验结果显示,模型 2 相较于模型 1 在 BRATS2021 数据集上的平均 DSC 评分提升了 0.90%,而平均 HD95 距离下降了 0.66 mm。模型 3 在模型 2 的基础上加入了双重注意力的 Transformer 模块,该模块建立远程依赖关系捕获通道和空间特征,从而提升了模型在医学图像分割任务中捕捉和整合远距离信息的能力。实验结果表明,模型 3 相较于模型 2 在 BRATS2021 数据集上的平均 DSC 提升了 1.44%,而平均 HD95 距离下降了 0.63 mm。在模型 3 的基础上增加了一个辅助解码器结构得到

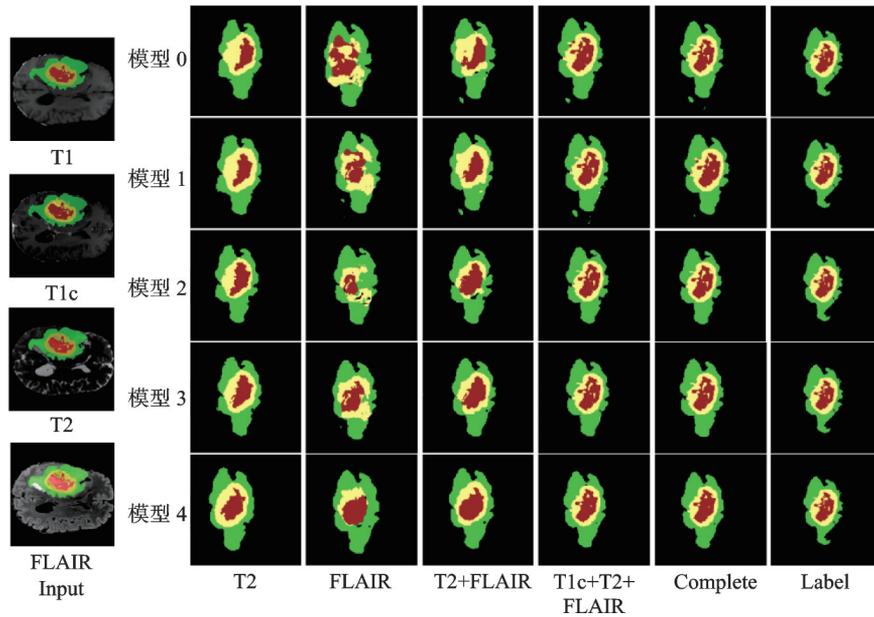


图9 消融实验的可视化结果对比

Fig.9 Comparison of visualization results of ablation experiments

模型4,该结构解决了在实际应用中一个或多个模态缺失导致的模型训练不平衡问题。从而保持其在多模态数据上的稳定性能。实验结果显示,模型4相较于模型3在BRATS2021数据集上的平均DSC评分增加了1.21%,而平均HD95距离下降了0.94 mm。最终得到的IM TransNet模型相较于基准模型在平均DSC评分上提升了4.8%,而平均HD95距离下降了3.03 mm。这不仅验证了本文所提模块的有效性,同时也彰显了它们在模型整体性能提升中的重要作用。

3 结束语

本文提出了一种U-Net和Transformer结合的不完整多模态脑肿瘤分割方法,通过多种实验的分析,主要结论如下:

(1)本文提出的IM TransNet模型实现了缺失模态脑肿瘤的准确分割,在BRATS2021数据集上的验证结果表明,IM TransNet相较于其他模型展现出显著的优势,在增强型肿瘤(ET)、肿瘤核心(TC)以及全肿瘤(WT)的分割任务上,分别取得了63.19%、76.42%和86.16%的平均DSC评分,并且平均HD95距离分别达到5.06、7.80和6.26 mm。

(2)本文提出的IM TransNet模型在BRATS2020数据集上进行实验,完成了对ET、TC以及WT的分割任务,分别取得了64.23%、80.11%和87.56%的平均DSC评分,且平均HD95距离分别达到5.42、6.31和5.91 mm。通过在不同数据集上进行实验,验证了该模型具有良好的泛化能力。

(3)通过一系列消融实验,证明了本文所提模块的有效性和实用性,且该方法具备高度的适应性,能够灵活应对各种模态缺失的场景。

参考文献:

[1] MENZE B H, JAKAB A, BAUER S, et al. The multimodal brain tumor image segmentation benchmark (BRATS)[J]. IEEE Transactions on Medical Imaging, 2014, 34(10): 1993-2024.
 [2] 赵凡, 张学典. 集成自注意力机制的医学图像分割方法[J]. 数据采集与处理, 2024, 39(5): 1240-1250.

- ZHAO Fan, ZHANG Xuedian. Medical image segmentation method with integrated self attention[J]. *Journal of Data Acquisition and Processing*, 2024, 39(5): 1240-1250.
- [3] 叶怀泽, 周子奇, 祁磊, 等. 基于特定领域解码的域泛化医学图像分割方法[J]. *数据采集与处理*, 2023, 38(2): 324-335.
- YE Huaize, ZHOU Ziqi, QI Lei, et al. Domain generalization via domain specific decoding for medical image segmentation[J]. *Journal of Data Acquisition and Processing*, 2023, 38(2): 324-335.
- [4] MOESKOPS P, VIERGEVER M A, MENDRIK A M, et al. Automatic segmentation of MR brain images with a convolutional neural network[J]. *IEEE Transactions on Medical Imaging*, 2016, 35(5): 1252-1261.
- [5] LIU Z, TONG L, CHEN L, et al. CANet: Context aware network for brain glioma segmentation[J]. *IEEE Transactions on Medical Imaging*, 2021, 40(7): 1763-1777.
- [6] ZHUANG Y, LIU H, SONG E, et al. A 3D cross-modality feature interaction network with volumetric feature alignment for brain tumor and tissue segmentation[J]. *IEEE Journal of Biomedical and Health Informatics*, 2022, 27(1): 75-86.
- [7] HATAMIZADEH A, TANG Y, NATH V, et al. UNETR: Transformers for 3D medical image segmentation[C]// *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Piscataway, NJ: IEEE Press, 2022: 574-584.
- [8] ZHOU T, RUAN S, VERA P, et al. A tri-attention fusion guided multi-modal segmentation network[J]. *Pattern Recognition*, 2022, 124: 108417.
- [9] HU J, GU X, WANG Z, et al. Active consistency network for multi-source domain generalization in brain tumor segmentation [J]. *Biomedical Signal Processing and Control*, 2023, 86: 105132.
- [10] WANG Y, CHEN J, BAI X. Gradient-assisted deep model for brain tumor segmentation by multi-modality MRI volumes[J]. *Biomedical Signal Processing and Control*, 2023, 85: 105066.
- [11] HAVAEI M, GUIZARD N, CHAPADOS N, et al. HeMIS: Hetero-modal image segmentation[C]// *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens: Springer, 2016: 469-477.
- [12] DORENT R, JOUTARD S, MODAT M, et al. Hetero-modal variational encoder-decoder for joint modality completion and segmentation[C]// *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*. Shenzhen, China: Springer, 2019: 74-82.
- [13] CHEN C, DOU Q, JIN Y, et al. Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion [C]// *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI*. Shenzhen, China: Springer, 2019: 447-456.
- [14] ZHOU T, CANU S, VERA P, et al. Brain tumor segmentation with missing modalities via latent multi-source correlation representation[C]// *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*. Lima: Springer, 2020: 533-541.
- [15] 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*. Piscataway, NJ: IEEE Press, 2021: 3975-3984.
- [16] CHEN C, DOU Q, JIN Y, et al. Learning with privileged multimodal knowledge for unimodal segmentation[J]. *IEEE Transactions on Medical Imaging*, 2021, 41(3): 621-632.
- [17] HU M, MAILLARD M, ZHANG Y, et al. Knowledge distillation from multi-modal to mono-modal segmentation networks [C]// *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*. Lima: Springer, 2020: 772-781.
- [18] WANG Y, ZHANG Y, LIU Y, et al. ACN: Adversarial co-training network for brain tumor segmentation with missing modalities[C]// *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*. Strasbourg: Springer, 2021: 410-420.
- [19] CHEN J, LU Y, YU Q, et al. TransUNet: Transformers make strong encoders for medical image segmentation[EB/OL]. (2021-02-08). <https://doi.org/10.48550/arXiv.2102.04306>.
- [20] WANG W, CHEN C, DING M, et al. TransBTS: Multimodal brain tumor segmentation using transformer[C]// *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*. Strasbourg: Springer, 2021: 109-119.

- [21] ZHANG Y, HE N, YANG J, et al. mmFormer: Multimodal medical Transformer for incomplete multimodal learning of brain tumor segmentation[C]//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2022: 107-117.
- [22] LIU H, WEI D, LU D, et al. M3AE: Multimodal representation learning for brain tumor segmentation with missing modalities [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2023, 37(2): 1657-1665.
- [23] ÇIÇEK Ö, ABDULKADIR A, LIENKAMP S S, et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens: Springer, 2016: 424-432.

作者简介:



汤占军(1969-),男,高级工程师,硕士生导师,研究方向:新能源发电技术研究以及图像处理, E-mail: tzj504@163.com。



蹇洪(1999-),男,硕士研究生,研究方向:医学图像分割的应用。



王健(1994-),通信作者,男,讲师,研究方向:可靠性评估及深度学习算法研究, E-mail: Jianwangzx@163.com。

(编辑:王静)