

基于注意力机制的缺失模态下脑肿瘤分割网络

张宇恒, 刘 恒

安徽工业大学 计算机科学与技术学院, 安徽 马鞍山 243000

摘要:目的 多模态核磁共振成像之间由于包含差异性和互补性信息,在临床实践上往往被综合运用来进行诊断和脑肿瘤分割。然而,在真实世界中,可能无法获取到完整模态的脑部核磁成像,而一种或者多种关键模态的缺失会导致分割性能的下降甚至崩溃。为了解决这一问题,提出了一种引入基于注意力的解耦再重构模块的分割网络,希望能够在缺失模态下仍能保持对脑肿瘤的分割精度。方法 网络使用 U 型架构,并引入基于注意力机制的解耦再重构模块。在解耦再重构模块中,首先通过各模态的高维特征融合获得共有基本特征,然后在此基础上借助差异注意力机制获得各模态的个性特征,最后将基本特征和个性特征经过交叉注意力整合重构后送入解码器中获取最终分割结果。同时,还设计了可学习的跳连接门,以使编码器中多种模态的浅层特征经过充分交互后传向解码器。结果 在一系列实验中,分割网络在 BraTS2020 脑胶质瘤数据集上表现出了出色的性能,缺失模态下仍能完成对脑肿瘤的分割,评价指标相较其他先进的方法提高约 2% 以上,高达 85.3%。结论 因此,基于注意力机制的脑肿瘤分割网络能够有效地在缺失模态下完成对脑肿瘤的分割,具有重要的实际意义。

关键词:注意力机制;脑肿瘤分割;缺失模态;核磁成像

中图分类号:TP391.4 文献标识码:A doi:10.16055/j.issn.1672-058X.2026.0001.006

Brain Tumor Segmentation Network Based on Attention Mechanism under Missing Modalities

ZHANG Yuheng, LIU Heng

School of Computer Science and Technology, Anhui University of Technology, Maanshan 243000, Anhui, China

Abstract: Objective Multi-modal magnetic resonance imaging (MRI) is often used in clinical practice for diagnosis and brain tumor segmentation due to its containing differential and complementary information. However, in real-world scenarios, complete modalities of brain MRI may not always be available, and the absence of one or more key modalities can lead to a decrease or even collapse in segmentation performance. To address this issue, a segmentation network incorporating an attention-based decoupling and reconstruction module is proposed to maintain the segmentation accuracy for brain tumors even in the presence of missing modalities. **Methods** The network adopted a U-shaped architecture and integrated a decoupled and reconstitution module based on attention mechanisms. In this decoupling and reconstitution module, high-dimensional features from each modality were first fused to obtain common basic features. Subsequently, by employing a differential attention mechanism, individual features unique to each modality were extracted from the shared common basic feature. Finally, basic features and individual features were integrated through cross-attention and fed into the decoder to obtain the final segmentation results. Additionally, a learnable skip connection gate was designed to allow the shallow features of multiple modalities in the encoder to interact sufficiently before being passed to the decoder.

收稿日期:2024-02-27 修回日期:2024-06-22 文章编号:1672-058X(2026)01-0047-08

基金项目:安徽省自然科学基金(2008085MF190)项目资助。

作者简介:张宇恒(1998—),男,硕士研究生,从事医学图像分割研究。

通信作者:刘恒(1976—),男,博士,教授,从事计算机视觉与模式识别研究。Email:Hengliusk@aliyun.com。

引用格式:张宇恒,刘恒.基于注意力机制的缺失模态下脑肿瘤分割网络[J].重庆工商大学学报(自然科学版),2026,43(1):47-54.

ZHANG Yuheng, LIU Heng. Brain tumor segmentation network based on attention mechanism under missing modalities [J].

Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(1): 47-54.

Results In a series of experiments, the segmentation network demonstrated excellent performance on the BraTS2020 brain glioma dataset, successfully segmenting brain tumors even under missing modalities. The evaluation metrics showed an improvement of over 2% compared with other advanced methods, reaching up to 85.3%. **Conclusion** Therefore, the attention-based brain tumor segmentation network can effectively segment brain tumors in incomplete modalities, which is of great practical significance.

Keywords: attention mechanism; brain tumor segmentation; missing modality; magnetic resonance imaging

分割任务是医学影像领域重要的任务之一,深度学习的出现极大地促进了这一领域的发展。脑肿瘤因其发病位置的特殊性和病理机制,一直被医生密切关注着。脑肿瘤最为常见和有效的检测手段就是核磁共振(Magnetic Resonance Imaging, MRI),因此如何在MRI成像中准确高效地分割出脑肿瘤,成为当前热门的研究方向。

由于MRI本身成像过程中的物理参数变化会对最终成像结果有较大的影响,因此产生了对应不同物理参数的MRI成像模态,如T1、T1ce、T2、FLAIR(Fluid Attenuated Inversion Recovery)。由于成像参数不同,不同的模态包含的信息特点各不相同。不同种模态之间可以相互补充,为医学诊断提供了更全面的信息。因此,在脑肿瘤分割中,许多工作都旨在利用多种模态的联合学习来提升分割的效果。

随着深度学习技术的进步,脑肿瘤分割也得到了长足的发展。而卷积神经网络在其中发挥了十分重要的作用。U-Net的提出奠定了一种主流且高效的分割网络,即采用多尺度的编码器-解码器框架,并引入跳连接补充细节信息^[1]。随后在此基础上改进并且实现更优越性能的网络被提出,包括能够对数据进行自适应处理的mn U-Net^[2]和能够更好地针对以三维(3-Dimension, 3D)体积输入的医学影像处理的3D U-Net^[3]。尽管这些基于U-Net的方法能够快速精准地进行分割,但是由于U-Net中卷积操作的特性,导致了特征提取时感受野受限,这在某些情况下影响了分割的性能。而基于注意力机制的Transformer^[4]可以更好地捕捉全局信息,文献[5-7]引入了Transformer以探索图像中的远程依赖关系。例如,Hatamizadeh等^[5]用Transformer替代原本的卷积编码器,能够更好地捕获上下文信息。然而,上述方法都是在单模态核磁输入下进行分割的,没有充分利用多模态核磁成像之间的差异和互补信息,其分割性能还有进一步提升空间。

在多模态核磁信息利用上,当前的一些主流方法探索了在特征维度上进行融合的策略,这样可以更好的发挥深度学习的能力^[8-10]。例如,Dolz等^[8]采取了层级层次的融合策略,并且不局限于在单个层之间进行融合,而是交叉在不同模态的各个层级之间,这允许网络可以学习模态之间更复杂的组合特征。但是,实际应用场景中,针对每个患者往往拿不到完整并且高质量的多种模态。在某一种或者几种关键模态的缺失下,这些方法对肿瘤的分割性能很难得到保证。现有的方法选择将模态特征抽象到潜在的特征空间中,再从这个特征空间中生成出包括肿瘤分割在内的数据^[11-13]。例如,Havaei等^[11]使用与模态数量无关的统计特征对每个模态提取出来的特征进行编码,再基于编码之后的特征进行分割。还有一些方法尝试将模态特征进行解耦来解决模态缺失问题^[14-16]。

尽管上述方法在面对不完整模态时仍能保证一定的分割性能,但大多数方法只能针对一种模态缺失的情况进行处理;有些方法则需要针对每种模态缺失的情况专门设计和训练网络,计算资源消耗巨大。为此,本文提出了一种缺失模态下的脑肿瘤分割网络,克服了上文中提到的问题,不需要针对每种模态缺失情况进行专门训练并且仍然能够保持高精度的分割性能。网络基于U型架构,包括解耦再重构模块和可学习的多模态跳连接门。基于注意力机制的解耦再重构模块从各模态特征中提取出共有的基本特征以保持分割的基本性能,并在此基础上引入有效模态的个性特征。跳连接门则保证了在流入解码器前编码器中各种模态的充分交互。实验结果证明了本文所提出方法的有效性。

1 网络模型架构设计

如图1所示,网络由三个主要组成部分构成,分别是:编码器、特征解耦再重构模块和解码器。此外,可学习跳连接门作为连接编码器和解码器的数据通路。

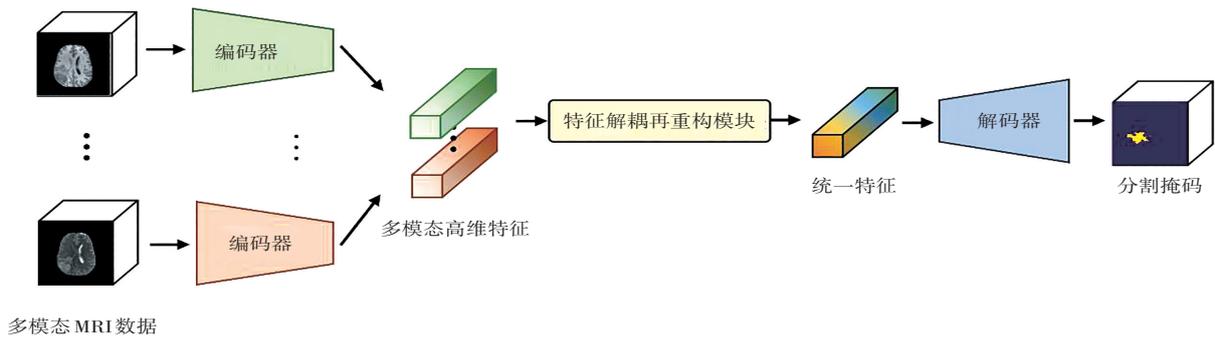


图 1 网络整体结构图

Fig. 1 Overall structure diagram of the network

1.1 特征解码器

编码器旨在通过搭建连续的卷积操作来提取特定模态内的高级特征。为了更好地关联同一个病例连续切片之间上下文的信息,分割编码器由三维卷积块堆叠构成。同时,为了能对每种模态都有更适应模态特点的特征提取策略,针对每个模态均有独立的参数不共享的编码器。

具体来说,编码器由 4 层基本编码层依次组成,每个编码层结构相同,如图 2 所示。

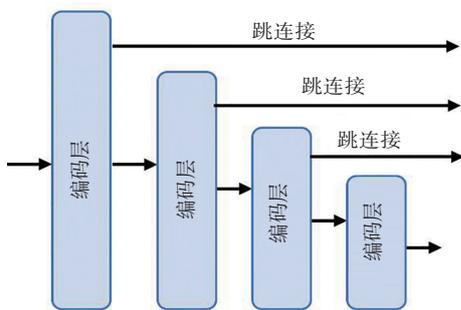


图 2 编码器示意图

Fig. 2 Structure diagram of the encoder

编码层主要由卷积组成,卷积核尺寸均为 3,其中一个步长为 2 的卷积同时负责下采样和提取特征。除此之外,为了改善模型的性能和泛化能力,编码器的每一层中还包含了 LeakyReLU 层和 Dropout 层,归一化层出于同样的目的也被加入其中,具体结构如图 3 所示。

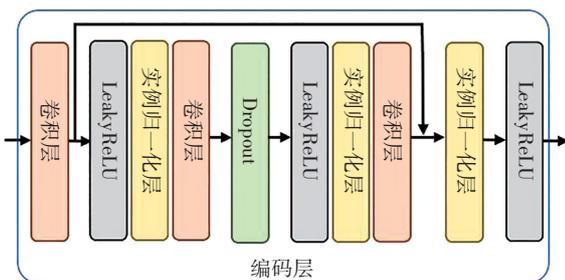


图 3 编码层详细结构

Fig. 3 Specific details of the encoding layer

特征每经过一个编码层后,尺寸会减半,而通道数则会翻倍,这一操作完成了对高维语义特征的聚集,提升了非线性特性,同时也实现了对多尺度信息的提取。

每一个编码层的输出除了作为下一层的输入,还会通过跳连接传给解码器,这一设计有助于将编码器中低级别的细节信息作为编码器中高级特征的补充,提高网络对细微结构的感知能力。

1.2 特征解耦再重构模块

现有的多模态分割方法可能太过依赖某个决定性模态,当该模态缺失之后发生的性能崩塌会使分割精度骤降。为了解决这一问题,首先保证模态共有基本特征的有效性,并在此基础上增强对模态个性特征的利用。

如图 4 所示,特征解耦再重构过程中,首先将各模态的高维特征一同输入基础特征聚合模块得到包含模态之间相同信息的特征作为基本特征,然后利用基本特征和各模态特征经过差异注意力得到模态个性特征。再重构过程即是将这两种特征建立起非线性的关联从而获得统一特征,交叉注意力帮助实现了这一过程。

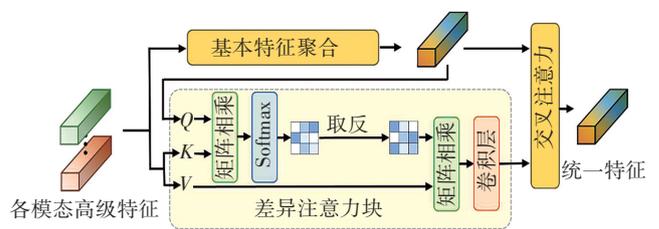


图 4 特征解耦再重构模块示意图

Fig. 4 Schematic diagram of feature decoupling and reconstitution module

1.2.1 基本特征聚合

为了获取基本特征,首先将各模态的高维特征沿通道拼接在一起,然后通过 1×1 可逆卷积^[18]压缩通道数来完成基本特征聚合。由于 1×1 卷积的可逆性,可

以保持梯度的流动,因此不会引入梯度消失或梯度爆炸问题,可以更好地支持深层网络的训练。同时,使用 1×1 卷积网络来完成特征融合可以在减少网络计算量的前提下,允许来自不同模态的特征进行跨通道的交互,使网络能够学习到更一般化的特征表示。

对于基本特征的提取与使用保证了模型在不管缺失何种模态的情况下都能获得基本的分割性能,摆脱了对某种关键模态的依赖,增强了网络的鲁棒性。基本特征聚合具体如式(1)所示:

$$\mathbf{F}_{\text{base}} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_m] \mathbf{W} \quad (1)$$

其中, \mathbf{F}_{base} 表示基本特征; \mathbf{F}_i 表示第 i 种模态对应的高维特征; \mathbf{W} 表示卷积操作的权重, $[\dots]$ 表示拼接操作。

1.2.2 差异注意力

各模态共有基本特征保证了网络的基本分割性能,个性特征则是在此基础上寻求分割精度的进一步提升。包含每个模态特有信息的个性特征帮助网络更加全面地对目标肿瘤进行定位和分割。每个模态可能会更好地突出某些特定的肿瘤特征,充分利用这些特异信息可以增强网络对这些特定特征的感知能力。此外,在某种模态缺失的情况下,其他模态的特异信息可能会起到部分弥补的作用。

为了获得个性特征,本文设计了一种差异注意力来寻找出各模态源特征与基本特征之间的差异信息。

常见的注意力机制通过计算权重图的方式来获取特征之间具有更高相关性的信息,这与目的正好相反。因此,在传统注意力机制的基础上,选择将权重图取反来获取差异信息。

传统注意力机制的注意力权重图是 Query 和 Key 矩阵做积而来,代表着两个矩阵的相似度。在经过 Softmax 层之后,权重图的取值范围在 $[0, 1]$ 之间,越接近 1 代表两个矩阵对应位置相似度越高,在后续的计算中权重也越高。而对于两个矩阵的差异信息,则是更应该关注接近 0 的位置,因此通过用 1 减去权重图来完成取反操作,使得相似度越低的差异区域权重越高。差异注意力机制可以被定义为式(2):

$$\mathbf{F}_{\text{indiv}} = \left(1 - \text{Softmax} \left(\frac{\mathbf{Q}\mathbf{K}_i^T}{\sqrt{d_K}} \right) \right) \mathbf{V}_i \quad (2)$$

其中, $\mathbf{F}_{\text{indiv}}$ 表示个性特征 (Individual Feature); \mathbf{Q} 表示 Query 矩阵,由基本特征经过前向传播层变换而来, \mathbf{K}_i 和 \mathbf{V}_i 分别表示第 i 个模态对应的 Key 和 Value 矩阵,由模态对应的源特征经过前向传播层变换而来; d_K 则表示矩阵 \mathbf{K} 的维度。

1.2.3 交叉注意力

对于基本特征和个性特征的融合,选择在高维空间进行,而不是在各自得到分割结果之后的后端融合。这是因为特征级别的融合可以更加充分地利用两者之间的相关性和互补性。高维空间也可以提供更多维度的信息表示,提供模型的表达能力和泛化能力,更好地适应复杂多变的数据分布。

借鉴于 Chen 等^[19]提出的 CrossViT 方法,本文选择使用交叉注意力机制作为融合层。交叉注意力是注意力机制的一种拓展,被提出后在后续的研究中广泛应用于视觉任务中。交叉注意力允许模型在特征融合过程中动态地调整不同来源特征的权重,将基本特征和个性特征融合再重构为统一特征。

基本特征和个性特征作为交叉注意力层的两个输入,并且生成作为计算因子的 \mathbf{K} , \mathbf{Q} 和 \mathbf{V} 矩阵。其中, \mathbf{Q} 矩阵由统一特征经过前向传播层变换得到, \mathbf{K} 和 \mathbf{V} 矩阵由个性特征经过前向传播层变换得到。计算如式(3)所示:

$$\mathbf{F}_{\text{unified}} = \text{Softmax} \left(\frac{\mathbf{Q}_{\text{base}} \mathbf{K}_{\text{indiv}}^T}{\sqrt{d_K}} \right) \mathbf{V}_{\text{indiv}} \quad (3)$$

其中, $\mathbf{F}_{\text{unified}}$ 表示统一特征 (Unified Feature); \mathbf{Q}_{base} 表示由基本特征 (Base Feature) 变换得来的 \mathbf{Q} 矩阵; $\mathbf{K}_{\text{indiv}}$, $\mathbf{V}_{\text{indiv}}$ 表示由个性特征 (Individual Feature) 得到的 \mathbf{K} 和 \mathbf{V} 矩阵; d_K 则表示矩阵 \mathbf{K} 的维度。

1.3 可学习跳连接与解码器

本方法和 U-Net 一样,编码器接受来自解码器对应层级的跳连接以获得更多的上下文信息和细节特征。但是,在多模态输入下,每个模态的独立编码器都会产生上下文特征,如何将多个模态的特征和解码器中对应层级的特征进行维度匹配就成了问题。一些方法通过简单的求最大值的方式完成多种模态跳连接的融合^[20]。但是,这种方法没有考虑各模态的贡献度。为此,本文设计了可学习的卷积权重图将这些特征结合在一起。

解码器旨在从抽象的高维特征逐层恢复出分割掩码。如图 5 所示,解码器采取和编码器基本对称的结构,即 4 层基本解码层,每层解码层主要由 3D 卷积块堆叠构成。对于编码器中的跳连接,解码器使用沿通道维度进行拼接的方式进行结合。与编码器中通过卷积完成下采样不同,解码器中专门设置了上采样层并通过双线性插值的方式完成高维特征分辨率的逐层重建。其他的卷积层、LeakyReLU 层和归一化层则给解码

器提供了更强的学习能力和泛化能力。具体结构如图 6 所示。

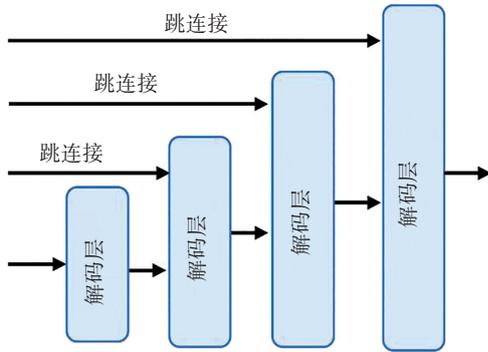


图 5 解码器示意图

Fig. 5 Schematic diagram of the decoder

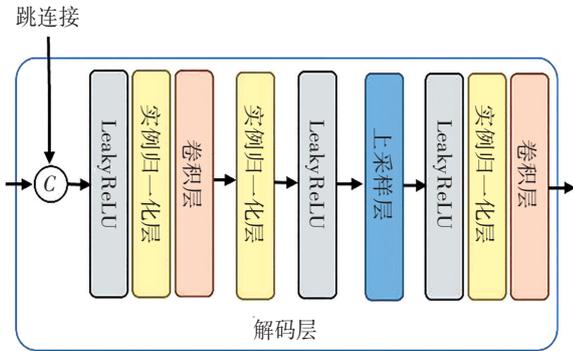


图 6 解码层具体结构图

Fig. 6 Specific details of the decoding layer

1.4 损失函数

分割训练损失包括二值交叉熵损失 (Binary Cross Entropy Loss, BCELoss) 和 Dice 分数损失。

二值交叉熵损失是分割的常见损失,适用于分割这种像素级别的二分类任务。但是,二值交叉熵损失会使模型在肿瘤分割这种背景像素数量更多的分割情况下,更偏向于背景,而且对物体边界和细节的保留能力较弱。因此,另外引入了 Dice 分数损失,Dice 分数损失通过计算预测结果和真实结果之间的交叉区域和总区域的比值来计算损失,只用于前景背景不平衡的情况。整体的损失函数 ℓ_{total} 可以概括如下:

$$\ell_{BCE}(y, \hat{y}) = -\frac{1}{n} \sum [y \ln \hat{y} + (1-y) \ln (1-\hat{y})]$$

$$\ell_{Dice}(y, \hat{y}) = 1 - \text{Dice}(y, \hat{y}) = 1 - \frac{2y\hat{y} + \varepsilon}{y + \hat{y} + \varepsilon}$$

$$\ell_{total}(y, \hat{y}) = \ell_{BCE}(y, \hat{y}) + \ell_{Dice}(y, \hat{y}) \quad (4)$$

其中,其中 y 表示真实的肿瘤分割掩码, \hat{y} 表示预测的肿瘤分割掩码, ε 表示防止零除错误引入的平滑项。

2 仿真实验与结果分析

2.1 实验数据集

本文使用 BraTS2020 数据集进行性能评估, BraTS2020 数据集是一个广泛使用的医学影像数据集,用于脑胶质瘤的分割任务^[21]。本数据集是对胶质母细胞瘤或低级别胶质瘤患者进行核磁扫描而构成。本数据集中所有病例的脑部核磁共振成像扫描均由完整的 4 个模态组成,其中包括 T1、T1c、T2 和液体衰减反转恢复序列 (Fluid Attenuated Inversion Recovery, FLAIR)。

如图 7 所示是同一病例在 4 个模态的展示。肿瘤标签则由专家标注而形成,包括 4 个分类,分别是背景、非增强肿瘤区域、水肿区域和增强肿瘤区域。实验选择由 3 个肿瘤区域共同组成的完整肿瘤区域 (Whole Tumor, WT) 作为分割目标。在训练过程中,按照 8 : 1 : 1 的比例配置训练集、验证集和测试集。

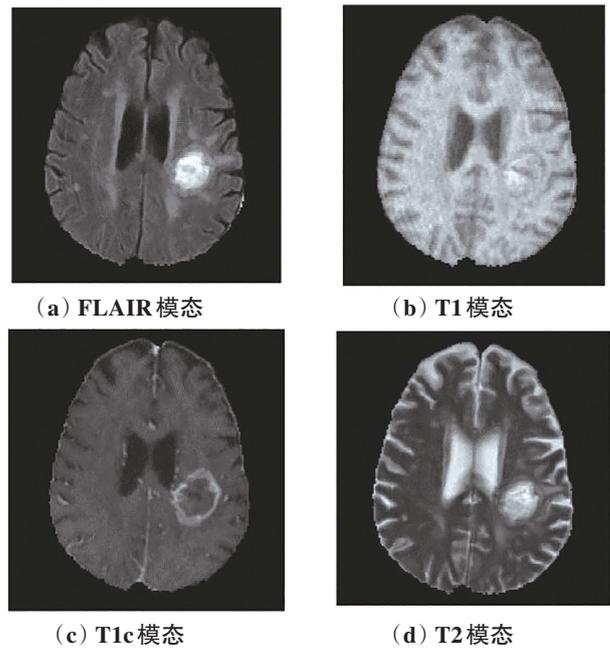


图 7 同一病例不同模态展示图

Fig. 7 Display of the same case in different modalities

2.2 数据集预处理和数据增强

对于数据集中的 MRI 图像,首先进行空间标准化、归一化等预处理,得到具有更清晰的脑部 MRI 影像。然后,在训练前将数据集进行随机变换以完成数据增强,具体的数据增强内容为:随机旋转、随机翻转、随机缩放、随机裁剪和随机弹性形变。同时,为了模拟在实际应用中出现的模态缺失情况,在训练前将数据集中每例 MRI 的 M 种模态分别乘上服从伯努利分布的系数 δ_i ,具体的流程可以用式(5)表示:

$$X_i = \delta_i X_i \quad (5)$$

其中, X_i 表示输入病例 X 的第 i 种模态, $i \in \{1, 2, \dots, M\}$ 。

2.3 实验环境与配置

所有实验均基于运行 Ubuntu 操作系统的 Python 3.8 环境, 并使用 PyTorch 1.6。硬件方面采用了单张 RTX 3090 显卡, 具有 24GB 的显存。初始学习率设置为 0.000 1, 采用随着 epoch 增加而衰减的“poly”衰减策略。默认的优化器为 Adam。训练的最大 epoch 数设置为 1 000。损失函数已在前文中进行了详细描述。

2.4 结果与讨论

2.4.1 定量实验结果

实验采用测试集的平均 Dice 分数和平均交并比 (IoU) 作为评价指标, Dice 和 IoU 是分割任务中常见的评价指标。两者取值范围均在 $[0, 1]$ 之间, 分数越接近 1, 说明预测结果和目标的真实分割掩码越接近。

实验中, 模态缺失通过置零来实现, 并选取了 3 个使用共享潜在空间的代表性模型来进行对比^[11-12, 16]。在 BraTS2020 数据集上的表现如表 1 所示, 包含了全部

的 15 种不完整/完整模态下的分割性能, 与其中的 14 种方法相比, 本文所提出方法的分割精度均高于对比方法, 平均 Dice 分数领先性能最好的对比方法 2%, 平均 IoU 分数领先 4%。可以发现: 随着可用模态的数量增加, 模型也表现出更好的分割效果, 这表明本方法对各模态个性特征的设计起到了正面效果。同时, 在极端情况下, 即只有一种可用模态时, 模型依然保持基本的分割性能, 证明了其鲁棒性。以上结果表明: 本文中针对各模态高维特征进行解耦再重构的设计在不完整模态下可以发挥较好的作用。

此外, 从表 1 的结果能看出, 在只有一种可用模态时, 各模态的 Dice 分数从高到低依次是 FLAIR、T2、T1c、T1。后续依次增加可用模态时, 不同情况的对比也呈现了这一趋势。这在一定程度上反映了各个模态对肿瘤分割过程中的贡献度和重要性, 同时也符合从病理学的角度对各个模态的评价。验证了所提出分割方法和医生临床诊断的一致性, 说明所提出方法具有一定的临床指导意义。

表 1 定量结果

Table 1 Quantitative results

可用模态	完整肿瘤区域 Dice 分数 [%]/IoU 分数 [%]			
	HeMIS ^[11]	U-HVED ^[12]	mmFormer ^[16]	Ours
Flair	52.48/35.57	84.39/72.99	86.10/75.59	88.36/79.14
T1c	61.53/44.43	53.62/36.63	72.22/56.51	76.80/62.33
T1	57.62/40.47	49.51/32.89	67.52/50.96	68.16/51.69
T2	80.96/68.01	79.83/66.43	81.15/68.27	80.87/67.88
Flair, T1c	68.99/52.66	85.93/75.33	87.30/77.46	90.16/82.08
Flair, T1	64.62/47.73	85.71/74.99	87.06/77.08	88.86/79.95
Flair, T2	82.95/70.86	87.58/77.90	87.59/77.92	88.71/79.71
T1c, T1	68.47/52.05	64.22/47.29	74.42/59.26	79.87/66.48
T1c, T2	82.48/70.18	81.32/68.52	82.99/70.92	86.43/76.10
T1, T2	82.41/70.08	81.56/68.86	82.20/69.77	84.10/72.56
Flair, T1c, T1	72.31/56.60	86.72/76.55	87.33/77.50	90.20/82.14
Flair, T1c, T2	83.85/72.19	88.09/78.71	88.14/78.79	90.26/82.28
Flair, T1, T2	83.43/71.57	88.07/78.68	87.75/78.17	88.89/80.00
T1c, T1, T2	83.94/72.32	82.32/69.95	82.71/70.51	87.14/72.21
Flair, T1c, T1, T2	84.74/73.52	88.46/79.30	89.64/81.22	90.32/82.34
平均	74.05/58.79	79.16/65.50	82.94/70.85	85.28/74.33

为了验证网络的泛化能力, 对于训练完成的模型, 还选取了 BraTS2021 数据集中新增的部分病例进行分割性能测试, 在多种情况下平均 Dice 能达到 84.71%, 说明本方法也具有较强的泛化能力。

如表 2 所示, 还对模型的参数量和浮点计算量

(FLOPS) 进行了分析, 并和通用的分割模型 nnU-Net^[2] 进行比较。结果显示: 本文的模型在参数量和计算量上均更占优。但是在训练时, 由于注意力机制中矩阵相乘的计算操作对显存占用很大, 所以导致网络对硬件设备要求较高。

表 2 模型参数量和浮点计算量
Table 2 Model parameters and FLOPS

模 型	参数量/M	计算量/G
nnU-Net	31.2	539.7
ours	5.19	486.6

2.4.2 定性实验结果分析

图 8 展示了网络的分割效果图。因为篇幅有限,只分别在 1、2、3、4 种可用模态的情况下对一例测试病例进行分割可视化,而没有展示全部的 15 种情况。从完整模态下的分割可视化结果可以看出,模型在保证分割肿瘤主体的前提下,对复杂边缘和邻近小肿瘤区域的高分割精度,这得益于可学习跳连接门对于编码器浅层特征中细节信息的充分运用。同时,在仅有一种模态的情况下,网络依然保持了较高的分割精度,验证本文提出方法的可行性。

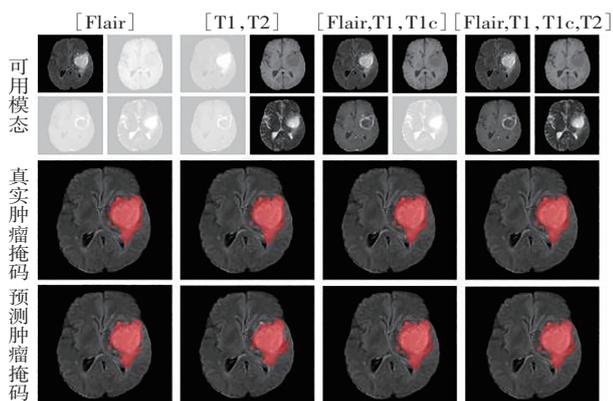


图 8 不同可用模态下分割结果可视化

Fig. 8 Visualization of segmentation results under different available modalities

如图 9 所示,本文还挑选了分割效果不佳的病例进行可视化展示,所选病例在全模态可用的情况下 Dice 分数为 61.30%。从真实标签和预测标签的对比图可以看出,在面对多个分散的小区域肿瘤时,网络的分割精度有限。虽然对肿瘤的主体和较大的区域仍然能分割,但是精度有待提升。这表明了网络在面对这种复杂情况病例时,仍需要进一步改进。

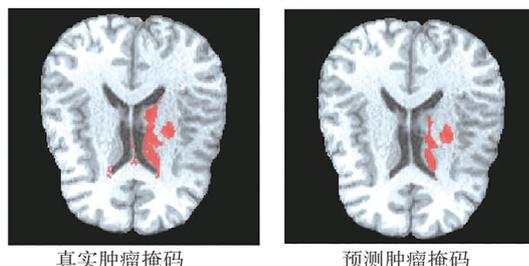


图 9 分割效果不佳病例可视化

Fig. 9 Visualization of a poorly segmented case

2.4.3 消融实验

为了进一步验证网络中各模块的有效性,设计了一系列消融实验,实验结果如表 3 所示。从表 3 可以看出:在多种模块组合情况下,解耦再重构模块使分割精度提升了大约 2%,可学习跳连接贡献了大约 1%的提升,完整的网络则提升了 2%~3%。此外,还将特征解耦在重构模块中的差异注意力更换为了普通的注意力机制,实验得到多种模态下平均 Dice 为 83.62%。上述进行的一系列实验表明了基于交叉注意力和差异注意力的解耦再重构模块和可学习跳连接的有效性。

表 3 消融实验结果
Table 3 Results of ablation study

网络组成	平均 Dice/%
基线网络	82.95
基线网络+可学习跳连接	83.79
基线网络+解耦再重构(差异注意力)	84.52
基线网络+解耦再重构(普通注意力)	83.62
基线网络+解耦再重构(差异注意力)+可学习跳连接	85.28

3 结 论

本文提出了一种基于注意力机制的缺失模态下脑肿瘤分割方法。本文方法在 U 型架构网络的基础上引入解耦再重构模块和可学习跳连接。通过各模态的基本特征保持在关键模态缺失下的基本性能,并进一步借助模态个性特征提升分割性能,展现了多种不完整模态情况下稳定的分割精度。在 BraTS2020 数据集上的实验证明了本文提出方法相较于其他算法的优越性,平均 Dice 分数达到了 85%。同时,在其他数据集上进行测试的实验效果验证了模型具备较好的泛化能力。一系列消融实验也证明了提出模块的有效性。这些都表明本文方法在医学图像分割领域有着广阔的应用前景。

本文所提出方法具有一定的局限性。虽然在分割性能上表现优异,但是注意力机制存在着训练困难和显存占用较大的问题,这意味着本网络需要更长的训练时间和面对潜在的梯度消失/爆炸问题。在后续的工作中,将会进一步寻求对网络的轻量化和训练的稳定性。此外,如定性实验所示,网络在遇到多个小面积的分散肿瘤区域时,可能会出现漏检或对边缘分割不准确的情况,后续的研究中会增强网络对于这种细小目标的关注。

参考文献(References):

- [1] RONNEBERGER O, FISCHER P, BROX T. Unet: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI. Springer International Publishing, 2015: 234–241.
- [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] ÇIÇEK Ö, ABDULKADIR A, LIENKAMP S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI. Springer International Publishing, 2016: 424–432.
- [4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017(12): 5998–6008.
- [5] 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. 2022: 574–584.
- [6] WANG W, CHEN C, DING M, et al. Transbts: Multimodal brain tumor segmentation using transformer[C]//Medical Image Computing and Computer Assisted Intervention – MICCAI. Springer International Publishing, 2021: 109–119.
- [7] ZHANG Y, PEI Y, ZHA H. Learning dual transformer network for diffeomorphic registration[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI. Springer International Publishing, 2021: 129–138.
- [8] DOLZ J, GOPINATH K, YUAN J, et al. HyperDense-Net: a hyper-densely connected CNN for multi-modal image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2018, 38(5): 1116–1126.
- [9] ZHANG Y, YANG J, TIAN J, et al. Modality-aware mutual learning for multi-modal medical image segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI. Springer International Publishing, 2021: 589–599.
- [10] XING Z, YU L, WAN L, et al. NestedFormer: Nested modality-aware transformer for brain tumor segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. 2022: 140–150.
- [11] HAVAEI M, GUIZARD N, CHAPADOS N, et al. Hemis: Hetero-modal image segmentation[C]//Medical Image Computing and Computer – Assisted Intervention-MICCAI. Springer International Publishing, 2016: 469–477.
- [12] DORENT R, JOUTARD S, MODAT M, et al. Hetero-modal variational encoder-decoder for joint modality completion and segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI. Springer International Publishing, 2019: 74–82.
- [13] CHARTSIAS A, JOYCE T, GIUFFRIDA M V, et al. Multimodal MR synthesis via modality-invariant latent representation[J]. *IEEE Transactions on Medical Imaging*, 2017, 37(3): 803–814.
- [14] NING Z, ZHONG S, FENG Q, et al. SMU-Net: Saliency-guided morphology-aware U-Net for breast lesion segmentation in ultrasound image[J]. *IEEE Transactions on Medical Imaging*, 2021, 41(2): 476–490.
- [15] 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(2): 4263–4274.
- [16] ZHANG Y, HE N, YANG J, et al. Former: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. 2022: 107–117.
- [17] YANG H, SUN AND J, XU Z. Learning unified hyper-network for multi-modal MR image synthesis and tumor segmentation with missing modalities[J]. *IEEE Transactions on Medical Imaging*, 2023, 42(12): 3678–3689.
- [18] KINGMA D P, DHARIWAL P. Glow: Generative flow with invertible 1×1 convolutions[J]. *Advances in Neural Information Processing Systems*, 2018(7): 10236–10245.
- [19] CHEN C F R, FAN Q, PANDA R. Crossvit: Cross-attention multi-scale vision transformer for image classification [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 357–366.
- [20] LACHINOV D, VASILIEV E, TURLAPOV V. Glioma segmentation with cascaded UNet[C]//International MICCAI Brainlesion Workshop. 2018: 189–198.
- [21] 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.

责任编辑:代小红