

基于深度学习通道交换的多模态脑肿瘤图像融合的分割模型

张顾瀛, 贺光华

绍兴文理学院数理信息学院, 浙江 绍兴

收稿日期: 2023年3月19日; 录用日期: 2023年4月20日; 发布日期: 2023年4月27日

摘要

脑部肿瘤是患者面临极高致死率的疾病, 手术切除是最佳治疗手段。术前精确分割整个脑部肿瘤区域对于手术切除至关重要。本文提出了一种基于深度学习通道交换的多模态脑肿瘤图像融合分割模型, 该模型采用通道交换和Swin-Transformer技术, 对不同模态进行分类, 提取不同类别的特征模态并进行特征通道交换, 以获得图像局部互补性特征。同时, 该模型利用Swin-Transformer进一步融合全局和局部特征。实验结果表明, 在BraTS2021数据集下, 该模型具有竞争力的分割精度, 整个肿瘤的Dice值为91.07。

关键词

深度学习, 多模态融合, Swin-Transformer

Multimodal Brain Tumor Image Fusion Segmentation Model Based on Deep Learning Channel Swapping

Guying Zhang, Guanghua He

College of Mathematics & Information, Shaoxing University, Shaoxing Zhejiang

Received: Mar. 19th, 2023; accepted: Apr. 20th, 2023; published: Apr. 27th, 2023

Abstract

Brain tumor is a disease with extremely high mortality rate for patients, and surgical resection is the best treatment option. Accurate preoperative segmentation of the entire brain tumor region is crucial for surgical resection. In this paper, a multimodal brain tumor image fusion segmentation

model based on deep learning channel swapping is proposed. The model utilizes channel swapping and Swin-Transformer technology to classify different modalities, extract feature modalities of different categories, and perform feature channel swapping to obtain locally complementary features. Meanwhile, the model utilizes Swin-Transformer to further fuse global and local features. Experimental results show that the model has competitive segmentation accuracy on the BraTS2021 dataset, with a Dice value of 91.07 for the entire tumor.

Keywords

Deep Learning, Multimodal Fusion, Swin-Transformer

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 研究背景以及意义

大脑是人类神经系统中最复杂、最先进的器官,在中枢神经系统中发挥着至关重要的作用和功能。脑瘤,也称为颅内肿瘤或脑癌,是新形成的生长物,可以起源于颅内腔内的各种组织,例如脑组织、脑膜、神经和血管。脑瘤是常见的神经系统疾病,对患者的生命构成重大威胁。磁共振成像(MRI)目前被广泛认为是检测和诊断颅内病变最有效的成像诊断方法。

在脑瘤的诊断和治疗中,医学图像分割技术发挥着至关重要的作用。因此,开发自动化、高效、准确的脑瘤分割算法对于改善脑瘤的诊断和治疗效果至关重要。手术切除是治疗脑瘤的最佳方法之一,因此,在术前准确分割出需要切除的整个脑瘤区域至关重要。传统的分割方法是人工手动分割,这一方法需要大量的人力、物力和时间资源,并且分割结果受到医生个人经验和主观因素的影响,从而导致一些误差和不确定性。

基于深度学习的图像分割算法可以自动学习和提取图像特征,从而减少人为干预,提高脑瘤分割的准确性和精度。基于深度学习的算法具有良好的学习能力,可以从大量的医学图像数据中学习更准确、可靠的特征,从而提高脑瘤的分割准确性和自动化水平。这种方法可以有效完成脑瘤分割任务,缩短医生的诊断时间,减轻医生的工作负担,同时也降低医疗成本。因此,基于深度学习的脑瘤分割研究已成为医学图像处理领域的热点和挑战之一。因此,基于深度学习的脑瘤分割研究具有重要的理论和实践意义,并将在未来医学诊断和治疗中发挥越来越重要的作用。

1.2. 国内外研究现状

1.2.1. 深度学习在医学图像分割中的应用

近年来,基于深度学习的计算机视觉图像处理已成为研究的热点之一,特别是主要基于卷积神经网络(CNN)以自动从大量的医学图像中学习特征,并且可以实现端到端的训练和预测。其中,由多层卷积叠加构成的全卷积神经网络在医学图像处理的领域更是备受关注。全卷积神经网络在提取图像特征信息上取得了不小的进展。卷积运算中卷积核参数的共享方式,不仅能大幅度降低参数量,还能保证学习到的参数具有稳定性,而卷积本身能学习到图像数据的结构不变性。因此,全卷积网络结构是当前应用最广泛的结构之一,例如,由 Olaf 等人首次提出的 U-Net [1]结构解决了深层特征抽象和浅层特征提取不足

的问题, 实现了更明显和具体的特征提取, 从而提高了分割精度。此后, 许多研究人员在 U-Net 结构的基础上增加不同类型的卷积模块, 从而提出新的基于卷积计算的神经网络结构[2] [3], 例如, 提出图像块分割后, 对每个块进行卷积运算后重新拼接的重组块卷积算法[3]。虽然全卷积结构在医学图像分割方面取得了巨大的成功, 但由于卷积计算的特殊性, 它是通过滑动图像块的方式对像素块进行计算, 导致特征学习受限于局部信息, 难以学习到全局特征。

1.2.2. Swin-Transformer 结构

当输入图像分辨率较高, 像素点较多时, 使用全局自注意力的 Transformer 将导致计算量急剧增加, 因此 Transformer 常用于提取较低维度的底层特征图, 例如 TransBTS 模型[4]。然而, 高维特征图中包含的全局信息同样重要, 因此研究学者从多个角度提出各种方法来减少 Transformer 计算时的数据量。例如 Swin-Transformer 结构, Swin-Transformer [5]是由香港中文大学、微软亚洲研究院等机构的研究人员共同提出的一种基于 Transformer 算法的一个新的架构。用于解决当 Transformer 计算量大的问题。Swin-Transformer 模型通过将特征图用不同的方法分块, 再将这些块进行 Transformer 的计算, 从而达到降低参数量的效果。解决了在原始特征图上进行 Transformer 计算参数量大的问题; 在此之后, 这些 swin_Transformer 模块被直接添加到网络中, 直接替换网络中的卷积模块数据进行特征提取[6]。这些新的神经网络已经被广泛的应用于医学图像分割[7] [8] [9] [10] [11]。

1.2.3. 多模态图像融合

多模态图像融合是指将不同类型的图像信息结合在一起, 以获取更准确、可靠的结果。这种方法特别适用于医学图像处理, 其中许多研究人员致力于使用深度学习技术来提取和融合多模态医学图像的特征。多模态图像融合方法可以分为前期像素级别融合和中期特征级别融合。前期像素级别融合方法将图像像素点与像素点结合起来, 但效率低且容易产生模糊效果。而中期特征级别融合方法更加灵活, 适用于处理不同的融合任务, 可以帮助分割模型学习代表性的特征, 并消除图像中的不确定性和歧义信息, 以获取更准确的判别结果。基于深度学习技术图像融合可以融合由不同成像方式而来的不同模态图像信息。例如, 利用通道交换的方法对 RGB、Depth、Normal 等不同的图像模态进行融合[12]。对于多模态的脑肿瘤图像来说, 不同的成像模态可以为病理区域提供独特的特征, 并且能够揭示隐藏在各种形式中的互补性信息[13]。例如, 一些研究采用 Transformer 和 U-net 的下采样分支分别提取大脑图像的四个模态的特征信息, 然后将这些特征信息拼接计算以获得最终的分割图像[12]。

2. 模型以及数据处理

2.1. 数据集和数据预处理

本文使用宾夕法尼亚大学生物医学图像计算与分析中心(CBICA)提供的公开脑肿瘤数据集 BraTs2021 [14] [15]来验证我们的方法。在 BraTs2021 数据集中, 每个病人数据包含四个模态图像(T1、T1ce、FLAIR、T2)和一个肿瘤标签图像。肿瘤标签包括背景(标签 0), 坏死和非增强性肿瘤(标签 1), 肿瘤周围水肿(标签 2), GD 增强性肿瘤(标签 4), 分别用红色、绿色和黄色标记。标签由四名经验丰富的放射科专家医生按照统一的标准进行标注。一般选取增强区域(ET, 标签 4), 肿瘤核心区域(TC, 标签 1 和 4), 以及整个肿瘤区域(WT, 标签 1、2 和 4)作为研究对象, 或者分别单独作为研究对象。在 BraTs2021 的训练集中, 即有分割标签的训练集(BraTs2021 共 1251 个数据)中, 随机选取 70%作为训练数据对网络进行训练; 10%作为验证集, 将训练好的模型用于数据集其余 20%的数据进行预测。

文中使用的实验数据为三维图像, 原始尺寸 $240 \times 240 \times 155$, 经过随机裁剪后变为 $128 \times 128 \times 128$ 作为输入图像, 这种裁剪的方式在一定程度上可以减少计算成本, 增加训练数据量, 并且易于输入模型。

训练数据进行数据增强, 进行随机翻转, 增加随机噪声, 数据增强可以提高模型的泛化能力。采用 Adam 优化器, 初始学习率的设置为 0.0002, 其中初始速率通过每次迭代以 0.9 的幂衰减。模型正则化则采用 L2 范数, 权重衰减率为 10^{-5} 。Batch size 为 1, 并且使用两个 3090 GPU 进行并行训练, 即每个 GPU 每次只放入一张图像进行技术。训练将最后 4 个模型的训练结果进行平均得到最后的分割结果。

如图 1 所示, 研究选取了 BraTS2021 数据集中随机的一个病人样本, 该样本包含了四种模态图像, T1 模态, T1ce 模态, T2 模态和 Flair 模态。图像从左到右分别是 T1 模态, T1ce 模态, T2 模态, Flair 模态, 从上到下分别是大脑不同模态的灰度图像和带分割标签的脑肿瘤图像。肿瘤标签图像标记了坏死和非增强性肿瘤标签(标签 1, 红色)、肿瘤周围水肿标签(标签 2, 绿色)以及 GD 增强性肿瘤标签(标签 4, 黄色)。T1 和 T1ce 模态在检测脑肿瘤水肿方面有其优势, 医生通常会结合 T1 和 T1ce 模态进行观察。T1ce 可以突显血量较高的区域, 并强化显示肿瘤部位, 进一步展现肿瘤内部情况。T2 模态信号强度与水含量相关, 易于确定病灶位置和大小, 水肿区域边界清晰可见, 但容易与脑脊液混淆。为了解决这一问题, 可以结合观察 T2 和 Flair 模态, 以便更容易区分脑脊液和肿瘤, Flair 模态有助于观察肿瘤边缘。同理, 医生通常会结合 T2 模态和 Flair 模态进行观察, 以便更准确地判断胶质瘤的实际范围。这种方法在 2022 年提出的 CKD-TransBTS 模型中被证实其有效性和必要性[16]。

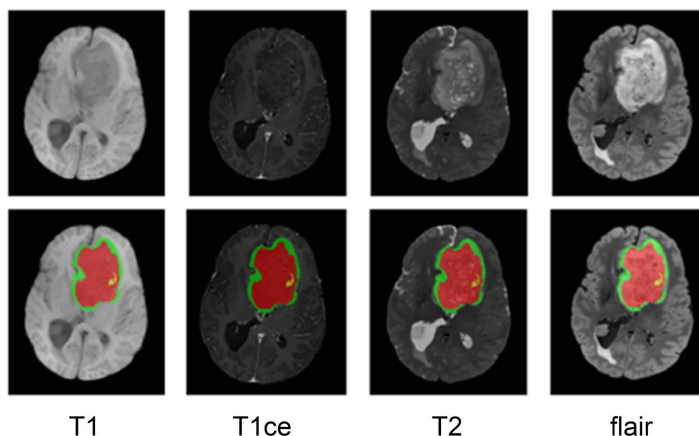


Figure 1. Four modal images of a patient in the BraTS2021 dataset

图 1. BraTS2021 数据集中一个病人的四个模态图像

2.2. 网络结构的整体框架

图 2 是本文提出的网络结构的整体框架, 这个框架具有一个 U 形结构, 由一个 encoder 分支, 一个跨越桥梁分支, 一个 decoder 分支组成。

在 encoder 部分, 将一个病人的脑肿瘤 MRI 图像作为输入, 将 4 个模态的 MRI 图像作为 4 个分支输入, 首先对其分类, T1, T1ce 拼接作为第一个组合类别, T2, Flair 拼接作为第二个组合类别。这两个类别作为输入, 两个分类分别经过 Baseconv block 后, 再经过三次 modal_fusion block down。modal_fusion block down 是利用卷积算法提取不同模态中的细节特征, 再利用通道交换的方法进一步融合模态。经过三次这样的模块输出三个不同尺度, 具有较好的不同模态融合特征的特征图。

在跨越桥梁部分。将 encoder 输出的三个特征图分别经过 Swin_Transformer 模块, 其中与传统的 Swin_Transformer 模块不同, shifted window 的取法做出改变, 利用扩大视野的方法, 使得这个模块提取到最佳的图像全局特征。跨越桥梁的输出与输入做残差连接, 降低学习到的特征过于抽象, 对模型分割精度的影响

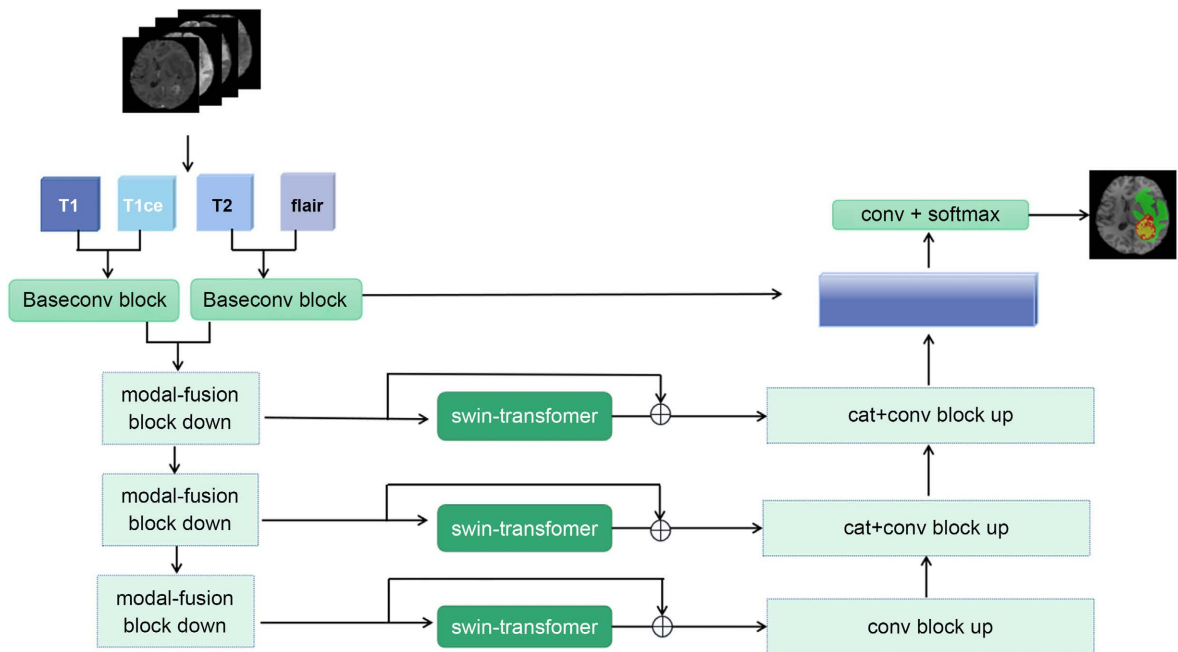


Figure 2. The framework of the proposed network structure
图 2. 提出模型的整体框架

在 decoder 部分。利用残差连接的方式，融合全局特征和局部特征，作为 decoder 的拼接特征图。由最底层的低维度特征图，通过反卷积的方法，恢复图像大小。升维的特征图与跨越桥梁的输出进行拼接。再经过反卷积，向上恢复维度。这样的拼接做两次，特征图向上恢复维度做三次，即 conv block up 模块。最后利用一层卷积核 SoftMax 函数输出分割概率矩阵，在网络中采用的模块包括 Baseconv block, modal_fusion block down, Swin_Transformer, cat + conv block up 模块。其中 modal_fusion block down, 和的结构 Swin_Transformer 如图 3 所示。

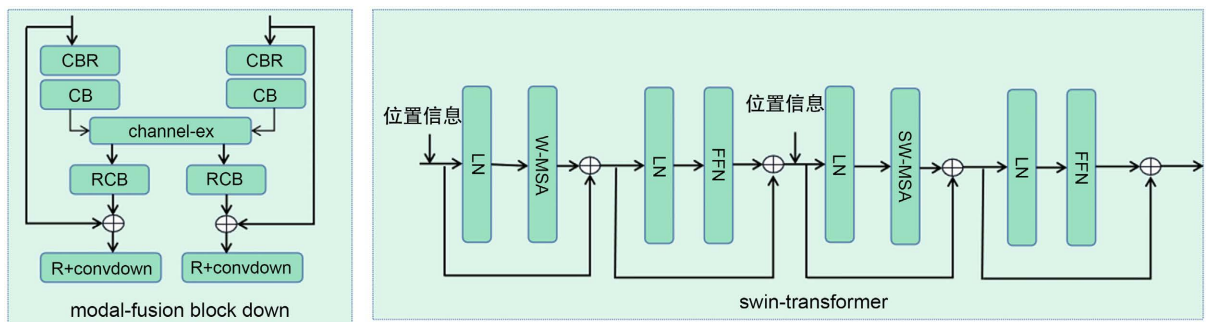


Figure 3. The internal structure of each module in the proposed network
图 3. 各个模块的内部结构细节

2.3. 网络结构细节

2.3.1. Encoder

在 Encoder 部分，对模态进行分组，将分组后的数据，设计通道交换的特征提取路径，来提取同一分组数据下相似的特征和不同分组数据的融合特征。模型的输入为 $X \in R^{4 \times W \times H \times D}$ ，即 4 种模态的 MRI 图像。将 4 个模态分为两组，T1 模态，T1ce 模态为一组，T2 模态，Flair 模态为一组，进行特征提取，即

输入数据分别为 T1 模态, T1ce 模态分组 $X_1 \in R^{2 \times W \times H \times D}$, T2 模态, Flair 模态分组 $X_2 \in R^{2 \times W \times H \times D}$ 。 X_1, X_2 两组数据分别经过 Baseconvblock 进行特征提取。 Baseconvblock 是由组合 $3 \times 3 \times 3$ 卷积, Batch Norm, ReLu 激活函数, 两个相同组合构成。得到同一分组下, 分组一提取的特征 $x_{1,0}$ 和分组二提取的特征 $x_{2,0}$, 其中 $x_{i,0} \in R^{4 \times W \times H \times D}$, $x_{i,j}$ 表示第 i 个分组的第 j 层特征, $i \in \{1,2\}$, $j \in \{0,1,2,3\}$, 其中输入 modal_fusion block down 的初始特征 $j=0$ 。随后再经过 3 次 modal_fusion block down 模块, 即模态融合模块, 其具体结构如图 3 左图框架所示, 图中 C 表示 $3 \times 3 \times 3$ 卷积, B 表示 Batch Norm, R 表示 ReLu 激活函数。在第一层 modal_fusion block down 模块中, 特征 $x_{1,0}$ 和 $x_{2,0}$ 分别经过 CBR 进行通道扩张, 由 4 扩展为 8, 得到的特征经过 CB 后进行通道交换。

每个模态组合的每个通道有一个可学习的参数, 这个参数的大小说明了这个通道学习到的特征的重要性。即这个通道特征越重要性越大参数越大。当这个学习到的参数小于设定的阈值 ε , 训练时设置 $\varepsilon = 1e-2$ 。参数小于设定的阈值说明这个通道的特征的重要性不大。则将这个通道用其他模态的对应通道取代。如公式 1 所示, $x_{i,j,k}$ 表示第 i 组模态 $i \in \{1,2\}$, 第 j 层特征, $j \in \{0,1,2,3\}$, 第 k 个通道。通道交换细如图 4 所示。通道交换消除了特征学习期间的冗余数据, 同时进行不同组合模态之间的通道交换, 完成不同模态组合的特征融合。

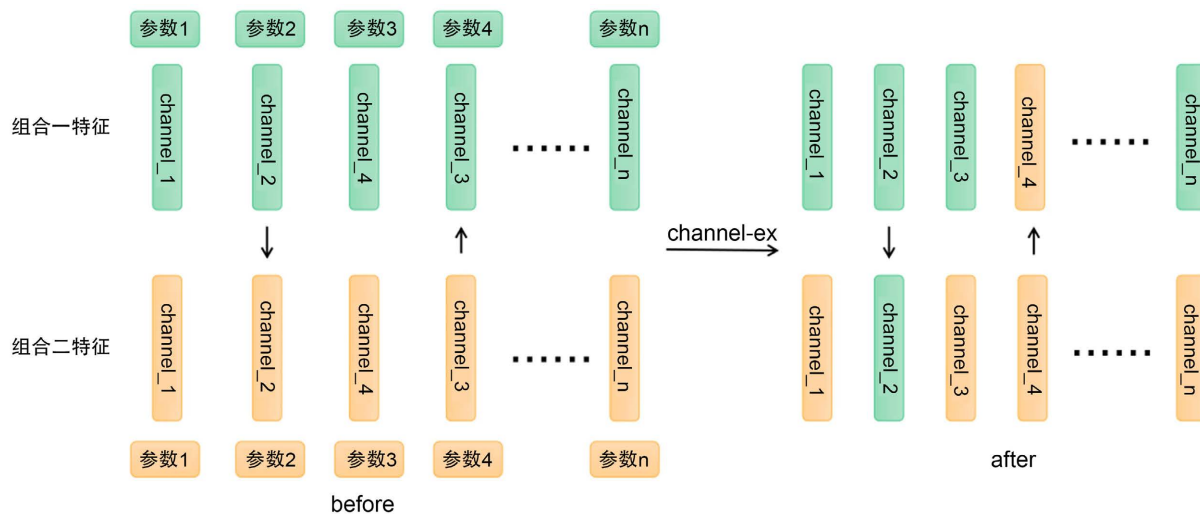


Figure 4. Diagram of the practical teaching system of automation major
图 4. 通道交换细节

$$x_{1,j,k} = \begin{cases} x_{1,j,k} \alpha_{1,j,k} \geq \varepsilon \\ x_{2,j,k} \alpha_{1,j,k} < \varepsilon \end{cases}, x_{2,j,k} = \begin{cases} x_{2,j,k} \beta_{1,j,k} \geq \varepsilon \\ x_{1,j,k} \beta_{1,j,k} < \varepsilon \end{cases} \quad (1)$$

将通道交换后的两组融合特征分别经过 RCB。得到的特征与输入的特征 $x_{1,0}$ 和 $x_{2,0}$ 进行残差相加后, 经过一个 ReLu 激活函数, 再通过步长为 2 的 $3 \times 3 \times 3$ 卷积, 缩小图像。输出下一层输入特征 $x_{1,1}$ 和 $x_{2,1}$, $x_{i,1} \in R^{8 \times W \times H \times D}$ 。重复以上计算三次, 获得三种不同分辨率大小的特征, 使得每一层的特征维度为 $x_{i,j} \in R^{C \times \frac{W}{2^j} \times \frac{H}{2^j} \times \frac{D}{2^j}}$, 通道数为 $C = 4 \times 2^j$, $j \in \{0,1,2,3\}$ 。第一层输出特征 $x_{1,1}$ 和 $x_{2,1}$, $x_{i,1} \in R^{8 \times 64 \times 64 \times 64}$, 第二层输出特征 $x_{1,2}$ 和 $x_{2,2}$, $x_{i,2} \in R^{16 \times 32 \times 32 \times 32}$, 第三层输出特征 $x_{1,3}$ 和 $x_{2,3}$, $x_{i,3} \in R^{32 \times 16 \times 16 \times 16}$ 。

2.3.2. 跨越桥梁

在 Encoder 和 Decoder 桥梁部分采用 swin-Transformer 来提取数据全局特征。swin-Transformer 结构

如图 5 右图框架所示。首先对数据进行 window patch 取块, 对每个块加入位置信息, 经过一层 Layer Norm, 再对每个 patch 进行 Multi_Attention 计算, 输出与输入进行残差相加后, 经过一层 Layer Norm 和一层 FFD (FeedForward), 残差相加后输出。输出数据进行 s_window patch 取块, 进行相同的计算。

对于本文来说, 即对 Encoder 部分输出的第一层特征 $x_{1,1}$ 和 $x_{2,1}$, 将特征, 按照通道方向进行拼接作为 swin-Transformer 的输入数据。第二层, 第三层相同。swin-Transformer 计算细节如图 5 左图所示。图 5 左图为 Transformer 中 Multi_Attention 计算图解。

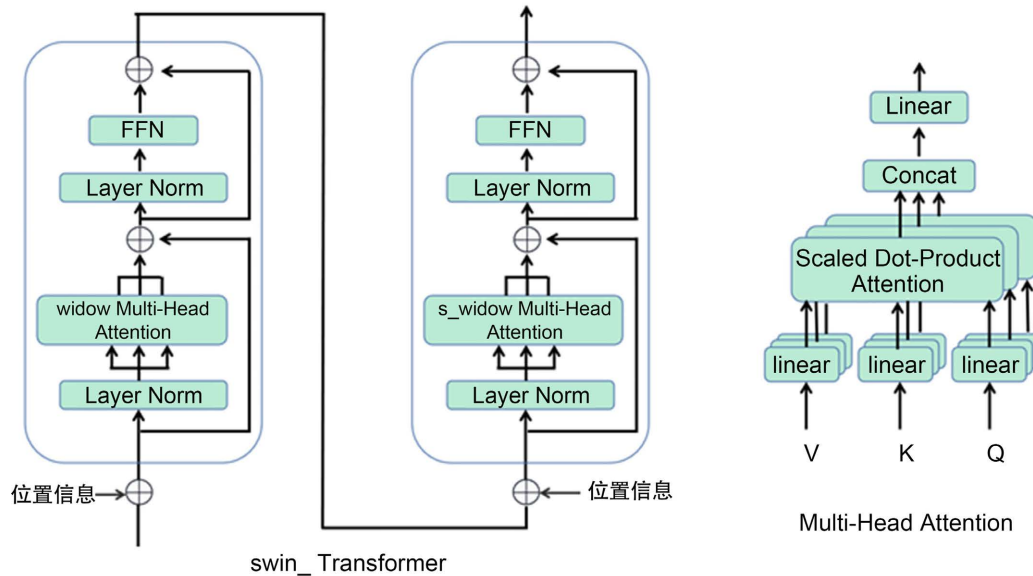


Figure 5. swin_Transforme details
图 5. swin_Transformer 细节

实质上 swin-Transformer 的计算是用输入数据以不同的形式取 patch, 对每一个 patch 进行 Transformer 计算。其中两次取 patch, 分别是 Window patch 和 s_window patch, 其取法如图 6。如图 6 左图, Window patch 是用不重叠的方式将特征块分成多个 patch。传统 swin-Transformer 中取 s_window patch (shifted window)的方法, 是以重叠的方式滑动 window, 从而得到多个 patch, 得到的每一个 patch 都是由连续的像素点构成。与传统方法不同, 本文提出的 s_window patch (space lattice window)是以不重叠的方式将图像分成多个块, 取出每个块中相同位置的像素点拼成一个 patch。即是以点阵的形式取 patch 后进行 Transformer 计算。这种方法与传统方法相比, 保留了图像的全局特征, 增大了特征学习的视野, 即每一个 patch 都是一个低分辨率的原始图像。

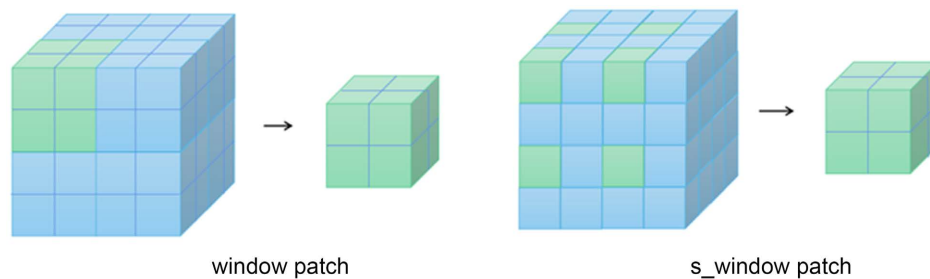


Figure 6. patch details
图 6. patch 取法细节

三层不同 swin-Transformer 分别输出三个不同大小的模态融合的全局特征 $x_j \in R^{C \times \frac{W}{2^j} \times \frac{H}{2^j} \times \frac{D}{2^j}}$, 通道数为 $C = 4 \times 2^j \times 2$, $j \in \{1, 2, 3\}$ 。将输入 swin-Transformer 的特征和输出的特征进行残差相加。得到三层既包含局部的模态融合后的特征, 又包含全局的特征。这三层特征用于 Decoder 反卷积时的长连接。

2.3.3. Decoder

在 Decoder 部分, cat + conv block up 模块细节如图 7 所示。Encoder 和 Decoder 桥梁输出的最底层即第三层特征经过 conv block up 模块后, 与桥梁输出第二层特征按通道维度进行拼接, 再经过 conv block up 模块。同理, 输出特征与桥梁输出第一层特征按通道维度进行拼接, 再经过 conv block up 模块。得到的特征与整个模型输入的原始特征进行相加后, 经过 SoftMax 激活函数得到分割概率矩阵。根据概率得到最后分割结果。

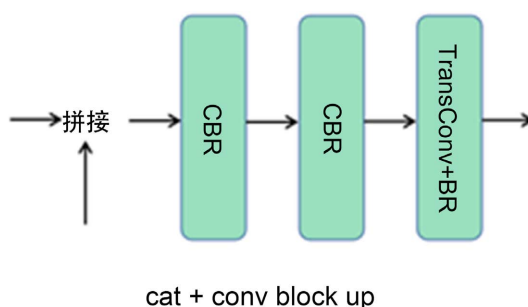


Figure 7. Details of cat + conv block up module in Decoder

图 7. Decoder 中 cat + conv block up 模块细节

3. 实验与结果分析

3.1. 与基线模型比较

除模型中采用各个标签 0 和 1, 2, 4 融合标签 WT 的 softmax Diceloss 之和作为损失函数对网络进行训练, 如公式(2)所示, 其中 Dice0, DiceWT, 表示输出标签 0, 和融合标签 WT 与精标准之间的 Dice 值。每个模型我们都进行了 400 次 epoch 的迭代。

$$\text{Loss} = 1 - \text{Dice0} + 1 - \text{DiceWT} \quad (2)$$

我们将提出的模型与 TransBTS [5], 2016 年提出的 U-net 结构[1]进行比较。

表 1 展示了三个模型在 BraTs2021 数据集下的分割结果, 我们提出的 channel_ex 模型在整个肿瘤区域(WT)获得 Dice score 为 91.05。可以看出我们提出的模型与 TransBTS、2016 年提出的 U-net [12]结构对比获得了有竞争性的结果。即三个模型中我们的模型具有最佳性能。并且我们做了模块的消融研究, 当模型中不含有 swin_Transformer 模块, 只含有通道交换模块即表 1 中 Our model_chex 的模型, 其 WT 的分割精度小于带有 swin_Transformer 模块的模型, 证明了 swin_Transformer 模块在模型中对 WT 区域分割精度提高的有效性。

Table 1. Comparison of segmentation results of each model on BraTS 2021 data

表 1. 各模型在 BraTS 2021 数据上分割结果比较表

Method	TransBTS	U_net	Ourmodel_chex	Our model
Dice score WT (%)	90.75	90.16	91.00	91.05

图 8 显示了在 BraTs2021 数据集上与其他基线模型相比较的定性结果, 红色区域表示标签 1, 即 WT 标签部分。上到下分别是三个病人同一层的脑部图像, 从左到右分别是 TransBTS, U-net 结构, 本文提出的模型 2 的二分类分割结果, GT 为精标准。绿色方框区域表示我们的结果比基线模型的细节刻画优越的区域。

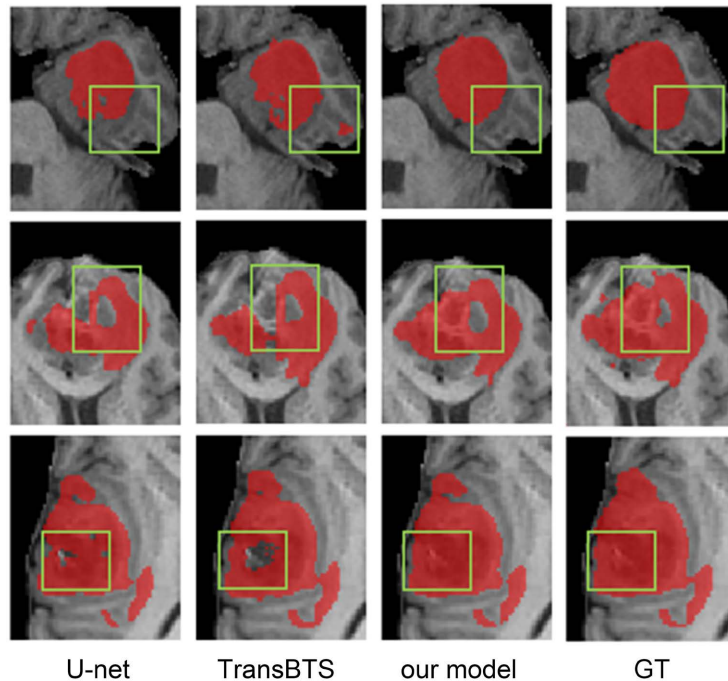


Figure 8. Details of cat + conv block up module in Decoder

图 8. 在 BraTS2021 数据集上不同模型的分割结果的可视化比较

从可视化结果的绿色方框区域结果可以看出, 我们的模型学习到比其他模型更为完整的分割信息。

我们将四个模态分为两类, 对每一类提取特征信息的同时不同类之间进行通道交换, 以融合模态中的特征信息。模型中的 Swin_Transformer 结构能够学习到更为全局, 且刻画更加细节的特征, 从而获得精度更高的分割结果。

3.2. 模型消融研究

公对于模型中通道交换方法和 swin_Transformer 方法分别进行了消融研究, 研究结果如表 2 所示。每个消融实验的研究我们都进行了 400 次 epoch 的迭代。我们对模型中的三层 swin_Transformer 中每一层的窗口(window)大小进行了研究。三层的基础窗口边长数为 8, 研究不同层的窗口大小对实验结果的影响。

Table 2. swin_Transformer window ablation study

表 2. swin_Transformer window 消融研究

swin_Transformer	第一层	第一层	第一层	WT (%)
Window_size	8	8	8	90.75
	8	8	16	89.76
	8	16	8	89.58
	16	8	8	91.05
	8	16	16	90.20

4. 结论与展望

根据医生手动标注的经验, 脑肿瘤四个模态被划分为两类, 同类模态之间具有更准确的分割特征。为了更好地提取同一分类图像的特征信息, 需要通过去除学习到的冗余信息, 利用通道交换的方式去除不必要的通道, 并将另一类重要的通道交换进去, 使得每个通道学习到的特征都非常重要, 并且可以同时融合不同的模态信息。在模型的长连接部分, 使用点阵 window 的 swin-Transformer 模型对图像提取全局信息。该模型在第一章介绍的公开的数据集 BraTs2021 上进行了验证。通过将该方法与其他两个基线网络 TransBTS 和 U-net 进行比较。结果表明, 该方法比其他方法更能获得更精确的分割结果。同时对这个模型进行了消融研究, 实验表明, 通道交换和点阵 window 的 swin-Transformer, 在模型中对提高分割精度具有重要作用。并且对通道交换的以及 swin-Transformer window 大小进行了消融研究, 获得最佳模型方案。

由于训练数据集是模态完整的脑肿瘤图像, 在未来期望随机丢掉部分模态, 测试模型对缺失模态是否具有泛化能力。

参考文献

- [1] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *MICCAI 2015: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Lecture Notes in Computer Science*, Vol. 9351, Springer, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [2] Jwaid, W.M., Al-Husseini, Z.S.M. and Sabry, A.H. (2021) Development of Brain Tumor Segmentation of Magnetic Resonance Imaging (MRI) Using U-Net Deep Learning. *Eastern-European Journal of Enterprise Technologies*, **4**, 23-31. <https://doi.org/10.15587/1729-4061.2021.238957>
- [3] Pereira, S., Pinto, A., Amorim, J., Ribeiro, A., Alves, V. and Silva, C.A. (2019) Adaptive Feature Recombination and Recalibration for Semantic Segmentation with Fully Convolutional Networks. *IEEE Transactions on Medical Imaging*, **38**, 2914-2925. <https://doi.org/10.1109/TMI.2019.2918096>
- [4] Wang, W., Chen, C., Ding, M., Yu, H., Zha, S. and Li, J. (2021) TransBTS: Multi-Model Brain Tumor Segmentation Using Transformer. In: Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y. and Essert, C., Eds., *MICCAI 2021: Medical Image Computing and Computer Assisted Intervention—MICCAI 2021, Lecture Notes in Computer Science*, Vol. 12901, Springer, Cham, 109-119. https://doi.org/10.1007/978-3-030-87193-2_11
- [5] Liu, Z., et al. (2021) Swin Transformer: Hierarchical Vision Transformer Using shifted Windows. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 9992-10002. <https://doi.org/10.1109/ICCV48922.2021.00986>
- [6] Cao, H., et al. (2021) Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. (Preprint)
- [7] Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. and Xu, D. (2022) Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: Crimi, A. and Bakas, S., Eds., *BrainLes 2021: BrainLesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Lecture Notes in Computer Science*, Vol. 12962, Springer, Cham, 272-284. (Preprint) https://doi.org/10.1007/978-3-031-08999-2_22
- [8] Peiris, H., Hayat, M., Chen, Z., Egan, G. and Harandi M. (2021) A Robust Volumetric Transformer for Accurate 3D Tumor Segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S. and Li, S., Eds., *MICCAI 2022: Medical Image Computing and Computer Assisted Intervention—MICCAI 2022, Lecture Notes in Computer Science*, Vol. 13435, Springer, Cham, 162-172. (Preprint) https://doi.org/10.1007/978-3-031-16443-9_16
- [9] Wu, Y., et al. (2022) D-Former: A U-Shaped Dilated Transformer for 3D Medical Image Segmentation. *Neural Computing and Applications*, **35**, 1931-1944. (Preprint) <https://doi.org/10.1007/s00521-022-07859-1>
- [10] Lin, A., Chen, B., Xu, J., Zhang, Z., Lu, G. and Zhang, D. (2022) DS-TransUNet: Dual Swin Transformer U-Net for Medical Image Segmentation. *IEEE Transactions on Instrumentation and Measurement*, **71**, 1-15. <https://doi.org/10.1109/TIM.2022.3178991>
- [11] Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I. and Patel, V.M. (2021) Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. In: Bruijne, M., et al., Eds., *MICCAI 2021: Medical Image Computing and Computer Assisted Intervention—MICCAI 2021, Lecture Notes in Computer Science*, Vol. 12901, Springer, Cham, 36-46. https://doi.org/10.1007/978-3-030-87193-2_4

- [12] Zhang, Y., *et al.* (2022) mmFormer: Multimodal Medical Transformer for Incomplete Multimodal Learning of Brain Tumor Segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S. and Li, S., Eds., *MICCAI 2022: Medical Image Computing and Computer Assisted Intervention—MICCAI 2022, Lecture Notes in Computer Science*, Vol. 13435, Springer, Cham, 101-117. (Preprint) https://doi.org/10.1007/978-3-031-16443-9_11
- [13] Calhoun, V.D. and Sui, J. (2016) Multimodal Fusion of Brain Imaging Data: A Key to Finding the Missing Link(s) in Complex Mental Illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, **1**, 230-244, <https://doi.org/10.1016/j.bpsc.2015.12.005>
- [14] Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J. and Wiest, R. (2015) The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, **34**, 1993-2024, <https://doi.org/10.1109/TMI.2014.2377694>
- [15] Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M. and Rozycki, M. (2018) Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. (Preprint)
- [16] Lin, J., Lin, J., Lu, C., Chen, H., Lin, H., Zhao, B., Shi, Z., Qiu, B., Pan, X. and Xu, Z. (2023) CKD-TransBTS: Clinical Knowledge-Driven Hybrid Transformer with Modality-Correlated Cross-Attention for Brain Tumor Segmentation. *IEEE Transactions on Medical Imaging*, 1-1. (Preprint) <https://doi.org/10.1109/TMI.2023.3250474>