

# 基于融合去噪的多模态脑肿瘤 MR 图像分割

申俊丽, 蔺崇玉, 海玉曼

(河南师范大学 计算机与信息工程学院(人工智能学院), 河南 新乡 453007)

**摘要:** 针对多模态脑肿瘤 MR 图像分割面临模态融合困难和融合过程中易产生噪声的问题及模态处理后全局和局部特征提取不充分造成分割精度降低的问题, 首先设计了结合 ViT(vision Transformer)和反卷积(transposed convolution)的模块, 用于实现多模态 MR 图像的高效融合与深度去噪, 然后创建了融合 Transformer skip connection 模块(TSC), 利用 Transformer 的多头交叉注意力机制, 对传统跳跃连接进行了革新, 用于捕获融合去噪后模态特征图的全局和局部特征, 进一步提升分割精度. 最后在公开数据集 BraTS2021 上的实验结果表明, 该方法在分割肿瘤时, 精度达到了 88.39%, Dice 指数为 83.44%, Hausdorff 距离仅为 2.356 6.

**关键词:** 脑肿瘤 MR; 融合去噪; U-Net; Transformer; 图像分割

**中图分类号:** R318; TP391.41

**文献标志码:** A

**文章编号:** 1000-2367(2026)02-0135-08

在脑肿瘤的医学图像分析中, 磁共振成像通常用于获取不同模态的图像(T1、T2、T1CE、FLAIR), 每种模态提供有关肿瘤的不同信息. 具体来说, T1 模态描绘肿瘤的解剖结构, T2 模态显示病灶的位置和大小, T1CE 模态揭示肿瘤的内部状况, FLAIR 模态则突出显示肿瘤周围及水肿区域. 然而, 多模态图像在融合过程中往往会产生噪声和伪影, 使得有效的特征融合和噪声处理成为当前的研究热点.

鉴于脑肿瘤大小、位置分布不均, 边界复杂多变, 因此精准分割极具挑战性. 深度学习技术, 尤其是 U-Net<sup>[1]</sup> 模型及其变体, 在脑肿瘤分割领域展现出巨大潜力. 其中, UNet++<sup>[2]</sup> 模型通过引入密集连接路径, 扩展原始的跳跃连接, 缩小语义编码器与解码器子网络特征图之间的差距, 构建更为复杂的嵌套 U 型结构, 提升分割的精度<sup>[3]</sup>. Attention Gate ResU-net<sup>[4]</sup> 首次在跳跃连接中引入注意力机制, 考虑不同模态特征之间的关联, 提高了多模态脑肿瘤图像的分割精度.

本文设计了 2 个模块: 1) VTC(vision Transformer transposed convolution) 模块, 在编码器阶段融合了 ViT(vision Transformer)<sup>[5]</sup>, 利用全局自注意力机制<sup>[6]</sup> 对特征融合和重建过程进行优化. 解码器结合反卷积层实现上采样, 并有效减少噪声干扰. 同时, 利用自编码器和跳跃连接提升图像质量. 2) TSC(Transformer skip connection) 模块, 融合多头交叉注意力机制的跳跃连接块, 捕获全局上下文信息, 解决了图像分割过程中造成的语义信息缺失问题, 并在瓶颈层引入 ResNet<sup>[7]</sup> 块, 有效缓解了深层网络训练中的梯度问题. 在 BraTS2021<sup>[8]</sup> 数据集上的实验结果显示不仅显著提升了图像处理和分割的精度与效率, 还为临床诊断提供了技术支撑.

**收稿日期:** 2024-07-17; **修回日期:** 2024-08-15.

**基金项目:** 国家自然科学基金(11601130); 河南省自然科学基金(242300420249; 252300420329); 河南师范大学国家级科研项目培育基金(20210372).

**作者简介(通信作者):** 申俊丽(1979-), 女, 河南淮阳人, 河南师范大学副教授, 博士, 研究方向为医学图像处理, E-mail: zuoyawen1215@126.com.

**引用本文:** 申俊丽, 蔺崇玉, 海玉曼. 基于融合去噪的多模态脑肿瘤 MR 图像分割[J]. 河南师范大学学报(自然科学版), 2026, 54(2): 135-142. (Shen Junli, Lin Chongyu, Hai Yuman. Multimodal brain tumor MR image segmentation based on fusion denoising[J]. Journal of Henan Normal University(Natural Science Edition), 2026, 54(2): 135-142. DOI: 10.16366/j.cnki.1000-2367.2024.07.17.0001.)

# 1 模型

## 1.1 整体网络架构

图 1 给出了整体网络架构.在图像处理阶段,选择了去噪自编码器(denoising autoencoder, DAE),这是一种用于去除数据噪声并提取有用特征的深度学习模型.它是一种自编码器的变体,专门用于处理噪声数据,并提高模型的鲁棒性和性能.与生成对抗网络等模型相比,这种变体在图像重建方面能够表现出更好的性能.在此基础上,以多模态脑肿瘤 MR 图像(FLAIR、T1、T1c 和 T2)作为输入,每种模态的图像通过独立的卷积层进行初步特征提取.

为了进一步提升图像质量并实现多模态图像融合去噪,设计了 VTC 模块.该模块在编码器部分嵌入了 ViT,并在解码器部分添加了反卷积层,以获得融合去噪的高质量单模态图像.在获得这些高质量单模态图像后,进一步构建了一个专门用于图像分割的网络架构.该网络在 U-Net 架构的基础上进行改进,它将处理后的单模态图像输入编码器进行深度特征提取,随后解码器对这些特征进行还原和优化.该模块在编码器和解码器之间引入了一个包含 Transformer 的跳跃连接.由于传统 U-Net 在各个阶段的编码器和解码器之间存在语义差距,从而限制了分割性能,用加入了 Transformer 的跳跃连接取代了普通的跳跃连接,在医学图像分割任务中取得了更好的性能.

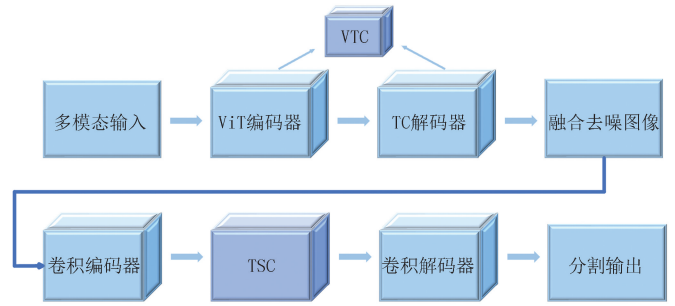


图1 整体网络架构图

Fig.1 Overall network architecture diagram

该模块在编码器和解码器之间引入了一个包含 Transformer 的跳跃连接.由于传统 U-Net 在各个阶段的编码器和解码器之间存在语义差距,从而限制了分割性能,用加入了 Transformer 的跳跃连接取代了普通的跳跃连接,在医学图像分割任务中取得了更好的性能.

## 1.2 用于融合去噪的 VTC 模块设计

在多模态医学图像处理过程中,由于多模态图像的特征分散,有必要整合来自不同模态的信息,以获得包含所有区域特征的单一模式图像.通过从该单一模式图像中提取特征,可以提高分割性能.通过多模态融合技术,能够学习每个通道的特征,从而增强对特定区域的检测能力.在图像融合过程中,噪声干扰是常见问题.通过使用去噪自编码器和噪声生成方法,可以提高图像重建质量,增强图像保真度,从而改进病灶区域的检测和分割.

针对噪声干扰的难题,ATLASON 等<sup>[9]</sup>提出了一种基于卷积自编码器的无监督脑病变分割方法.该方法通过自编码器在磁共振图像中学习深度特征,并利用解码器重建病变区域,从而显著减少对大量标注数据的依赖,为处理稀缺标注的医学图像数据集开辟了路径.此外,KASCENAS 等<sup>[10]</sup>在大量正常脑部 MR 图像上训练去噪自编码器后,发现经典的去噪自编码器模型具有应用潜力,尤其强调了跳跃连接在提高图像保真度方面的重要性.他们设计了一种噪声生成方法,可以将低分辨率噪声升级,从而实现高质量的重建.通过适当的噪声生成,去噪自编码器的重建误差可以推广到高强度病灶分割,并在脑部 MR 图像的无监督肿瘤检测中达到最先进的性能,超过了最新的分割方法.

在多模态融合领域,DOSOVITSKIY 等<sup>[5]</sup>提出的 ViT 首次将全局自注意力机制应用于完整图像,在 ImageNet 分类任务中取得了卓越表现.ZHOU 等<sup>[11]</sup>通过在通道层次上融合多模态图像,学习各通道的特征,从而提升分割效果.HUANG 等<sup>[12]</sup>提出了差分特征图,通过放大多模态图像中的肿瘤区域进行特征提取,增强对肿瘤区域的检测能力.毋小省等<sup>[13]</sup>提出融合注意力机制的多模态脑肿瘤 MR 图像分割方法,通过三重注意力模块(TAM)捕捉不同模态间的特征联系,增强特征间的语义相关性,有效减少了不相关信息的干扰.ZHANG 等<sup>[14]</sup>研发的 mmFormer(多模态医学 Transformer)模型,进一步扩展了 Transformer 的应用范围,实现了跨模态的局部与全局上下文建模.全局 Transformer 通过多头交叉注意力机制,全面捕捉图像的整体结构与细节信息,充分融合不同模态的特征,便于进行下游任务.

虽然上述基于多模态融合和去噪的方法在脑肿瘤分割中已经取得了较好的效果,但仍存在未充分考虑

各模态间深层联系、特征捕获不全面、对噪声等干扰处理不完善的问题.为了进一步解决这些问题,设计了 VTC 模块,该模块框架如图 2 所示.

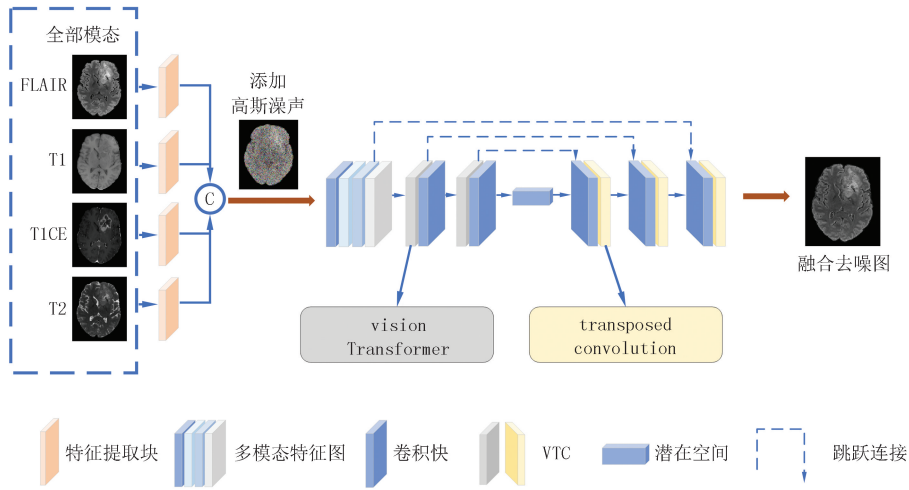


图2 图像预处理模块VTC

Fig.2 Image preprocessing module VTC

多模态脑肿瘤 MR 图像提供了完整的结构信息,对脑肿瘤的精确诊断和分割至关重要.然而,多模态数据的异质性和噪声问题增加了处理的难度.提出了一种新的网络架构,用于多模态图像处理.去噪自编码器包括编码器和解码器 2 个部分,在此基础上,将 Transformer 层集成到编码器中,实现多模态特征的高效融合,并在解码器中使用反卷积层进行去噪和上采样.

为了提升网络对噪声的鲁棒性,在输入层引入了高斯噪声,将随机生成的噪声添加到每个输入图像中.去噪自编码器的任务是去除这些噪声并重建原始输入图像.通过学习区分健康大脑图像模式和随机噪声模式来执行去噪.因此,在测试阶段,噪声生成对成功检测异常至关重要.噪声被添加到输入图像中.对于每个输入图  $I_m$ ,添加高斯噪声生成噪声图像  $\tilde{I}_m$ ;  $\tilde{I}_m = I_m + N(0, \sigma^2)$ ,其中,  $m \in \{1, 2, \dots, N\}$ .  $N(0, \sigma^2)$  是从均值为 0, 方差为  $\sigma^2$  的正态分布中采样的噪声图像.

在编码器部分加入 ViT 以捕捉全局上下文信息,每个头独立计算注意力权重,然后将结果拼接在一起:

$$\mathbf{F}_{\text{attention}}(\mathbf{Q}_m, \mathbf{K}_m, \mathbf{V}_m) = \mathbf{F}_{\text{softmax}}\left(\frac{\mathbf{Q}_m \mathbf{K}_m^T}{\sqrt{d_{k_m}}} + \mathbf{N}_M(\tilde{\mathbf{I}}_m)\right) \mathbf{V}_m, \mathbf{X}_{\text{head}, i} = \mathbf{F}_{\text{attention}}(\mathbf{Q}_m^i, \mathbf{K}_m^i, \mathbf{V}_m^i),$$

其中,  $\mathbf{Q}_m, \mathbf{K}_m, \mathbf{V}_m$  分别为不同模态的查询、键和值矩阵,  $d_{k_m}$  表示键的维度,  $\mathbf{N}_M(\tilde{\mathbf{I}}_m)$  是输入图像噪声的掩码,表示噪声对注意力权重的修正,并可以根据噪声水平进行调整.

解码器包括多个反卷积层,用于上采样和去噪.滤波器对输入特征进行反卷积操作,并通过噪声掩码抑制噪声:  $\mathbf{Y}_{\text{deconv}} = \mathbf{F}_{\text{deconv}}(\mathbf{Z}_{\text{latent}} + \mathbf{W}_{\text{deconv}}) + \mathbf{b} - \mathbf{N}_M(\mathbf{Z}_{\text{latent}})$ ,其中,  $\mathbf{Z}_{\text{latent}}$  是编码器输出的潜在表示,  $\mathbf{W}_{\text{deconv}}$  是反卷积核,  $\mathbf{N}_M(\mathbf{Z}_{\text{latent}})$  表示潜在表示中噪声的掩码,用于减少噪声对反卷积输出的影响.最终,在输出层得到多模态融合后去噪的高分辨率图像.

### 1.3 用于脑肿瘤 MR 图像分割的 TSC 模块设计

基于 VTC 模块,进一步设计 TSC 模块,该模块充分考虑模态融合去噪过程中可能忽略的全局和局部特征,针对融合去噪后的单模态特征图进行分割处理,提高分割精度.

以往对于跳跃连接的改进,ZHOU 等<sup>[2]</sup>提出的 U-Net++,基于 U-Net 分割架构,以其深度监督的特性和嵌套的密集跳跃连接,显著缩小了编码与解码特征图之间的语义差异,展现了卓越性能.MultiResUNet<sup>[15]</sup>聚焦于解决跳跃连接中的语义不匹配问题,通过引入残差路径(ResPath)优化特征融合,进一步提升分割效果.Attention U-Net<sup>[16]</sup>在传统 U-Net 基础上融入注意力机制,通过注意力门控制跳跃连接中的特征传递,强化有用特征,抑制无关特征,从而精准定位病变区域,提高分割精度.

基于注意力机制的改进,GCAUNet<sup>[17]</sup>提出的细节恢复(DR)路径和群体跨通道注意力(GCA)模块,结

合多尺度输入特征,与注意力机制共同作用,从多个维度增强了对图像细节与重要特征的捕捉能力.TransUNet<sup>[6]</sup>与 TransBTS<sup>[18]</sup>等模型将 Transformer 的强大全局建模能力引入医学图像分割领域,通过编码器-解码器架构与 Transformer 的深度融合,实现了对图像局部与全局信息的捕获。

通过设计 TSC 模块,解决在融合去噪过程中模态的全局和局部特征丢失问题.通过利用跳跃连接优化特征融合,解决了编码和解码特征图之间的语义差异和特征传递问题,从而提高了分割精度.引入 Transformer<sup>[19]</sup>的全局建模能力,实现对图像局部与全局信息的捕获,提升脑肿瘤分割的效果.利用 ResNet 块,缓解深度网络训练中的梯度问题,确保训练稳定性和分割精度,并提升网络架构性能<sup>[20]</sup>.通过综合利用多头注意力机制和深层网络架构的优势,解决了全局上下文信息缺失的问题,增强对图像细节与重要特征的捕捉,为脑肿瘤核磁共振图像的精准分割提供有力支持,该模块架构如图 3 所示。

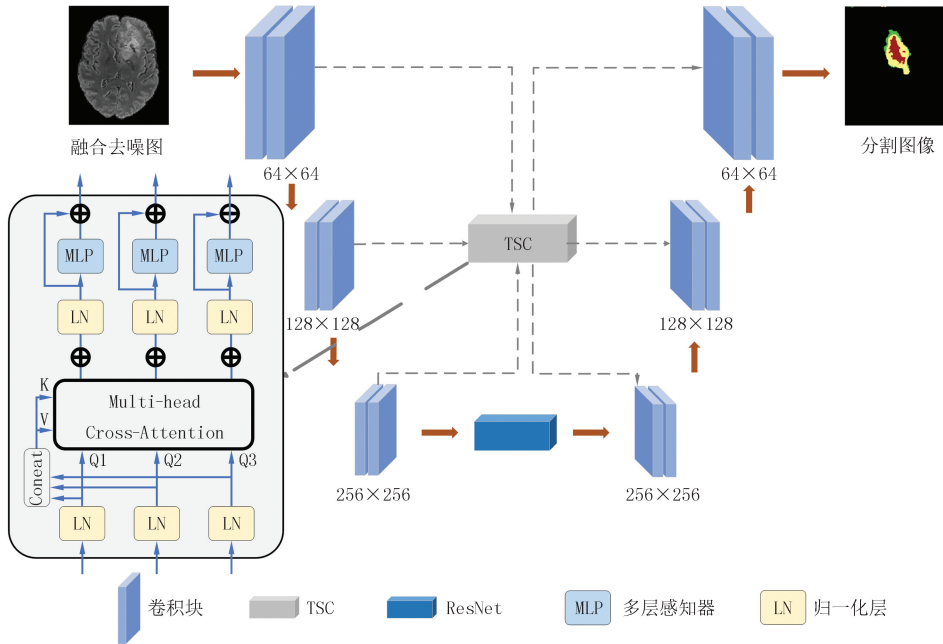


图3 图像分割模块TSC

Fig.3 Image segmentation module TSC

在图像分割过程中,使用改进的 U-Net 结构对去噪后的图像进行分割.为了解决跳跃连接问题,设计了新的模块 TSC,它融合了多尺度编码器的特性和 Transformer<sup>[19]</sup>的长距离依赖优点.模块 TSC 包含 3 个关键点:多尺度特征嵌入、多头交叉注意机制和多层感知器<sup>[21]</sup>,其中,多尺度特征嵌入能够同时捕捉不同层次的信息,提高模型对多样化输入的适应能力。

TSC 模块的主要组成部分是多头交叉注意力机制.由结构图可见,来自不同尺度的特征被传送到模块 TSC,经过多头交叉注意力机制处理后,再传送至多头交叉注意力模块,最终进入具有残差结构的多层感知器,进行通道和依赖关系的编码,从而精炼每个 U-Net 编码器级别的多尺度特征.TSC 模块的具体计算过程:

$$M_{TSC,i} = \sigma \left[ \psi \left( \frac{Q_i^T K}{\sqrt{C_s}} \right) \right] V^T$$

其中,  $\psi(\cdot)$  和  $\sigma(\cdot)$  分别表示实例归一化和 softmax 函数,  $C_s$  为缩放因子.与原始自注意力机制不同,该机制在通道轴而非补丁轴上进行注意力运算,并采用实例归一化.这种机制对相似性映射中的每个实例矩阵进行归一化,从而有效传播梯度.多头交叉注意后的输出计算:

$$M_{MCA,i} = (M_{TSC,i}^1 + M_{TSC,i}^2 + \dots + M_{TSC,i}^N) / N_{TSC}$$

其中,  $N_{TSC}$  为 TSC 模块的数量。

ResNet 残差块中的跳跃连接允许梯度直接通过网络层反向传播,有助于保持梯度大小,使网络训练更加充分.残差块能够帮助网络更快地收敛,减少了训练时间,帮助优化器更容易找到局部最优解,最大限度加快模型的训练速度,提高图像分割性能。

## 2 实验

为了评估多模态融合去噪效果及基于 U-Net 的脑肿瘤图像分割方法的性能,进行了以下实验。

### 2.1 实验环境与数据准备

实验中,使用 BraTS2021<sup>[8]</sup> 数据集和单个 Nvidia GPU RTX 4080(32G),通过 PyTorch 深度学习环境进行数据处理和训练,ITK-SNAP 用于图像可视化。BraTS2021 数据集中的每个图像经过专业医学专家手动标注,提供了精确的肿瘤分割标签,包括增强肿瘤核心区、非增强肿瘤核心区、坏死区和肿瘤水肿区。加载单个患者的 MR 图像(T1、T2、T1CE、FLAIR)及其标签。每个模态的数据大小均为  $240 \times 240 \times 155$ ,并共享分割标签。分割标签为:[0,1,2,4];label0:背景;label1:坏疽(necrotic tumor core, NT);label2:浮肿区域(peritumoral edema, PE);label4:增强肿瘤区域(enhancing tumor, ET),4种模态数据的可视化结果及地面实况(见图4)。

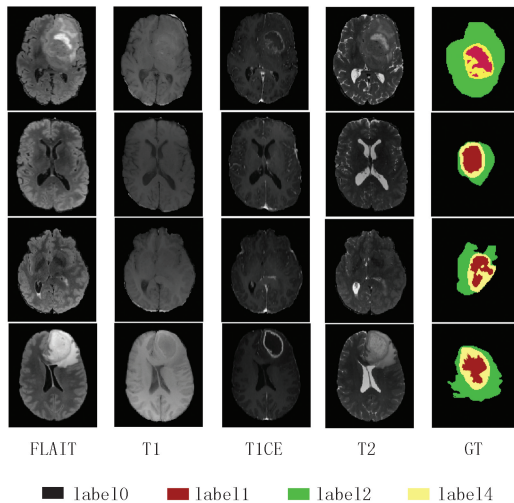


图4 数据集可视化

Fig.4 Dataset visualization

### 2.2 数据预处理

实验设置:采用去噪自编码器(DAE)作为最初的研究框架。DAE是一种常见的无监督学习方法,通过学习数据的隐层表示来实现自动编码和解码,从而实现图像去噪和异常检测。对每个模态的图像进行归一化处理,并添加噪声以模拟实际情况。模型架构输入层将每个模态的图像大小设置为  $64 \times 64$ 。编码器中的 Transformer 层主要使用多头自注意力机制进行特征融合,头数设置为 8,隐藏层大小为 256。前馈神经网络隐藏层大小设置为 256,进行层归一化。潜在空间经过编码器处理后,输出的潜在特征维度为  $256 \times 256$ 。解码器使用的反卷积层包括 3 个反卷积操作,内核大小为  $3 \times 3$ ,步长为 2,每层后接 ReLU 激活函数和批归一化。噪声掩码对潜在特征进行去噪处理。最后输出去噪后的高分辨率图像,大小为  $64 \times 64$ 。

评价指标:精度( $A_p$ ) $0 \sim 1$ ,数值越高表示模型性能越好。

$$A_p = \sum_n (R_n - R_{n-1}) P_n,$$

其中, $n$  根据相关文档的排名进行索引, $R_n$  表示检索结果中相关文档的数量, $P_n$  是检索结果的准确率。Dice 指数( $I_{Dice}$ )是一种用于衡量 2 个样本相似度的度量方法:

$$I_{Dice} = (2 \times |A \cap B|) / (|A| + |B|),$$

其中, $A$  表示预测区域,即模型预测为异常的区域, $B$  是真实区域。 $I_{Dice}$  取值范围为  $[0, 1]$ ,数值越靠近 1 表明 2 个样本相似性越高,模型性能越好。

损失函数:VTC 模块使用均方误差(MSE)损失函数,计算预测值与真实值之间差值的平方的平均值,损失函数  $L_{VTC}$ <sup>[10]</sup> 为

$$L_{VTC} = \frac{1}{N_{VTC}} \sum_{i=1}^N ((y_i - \hat{y}_i)^2 \cdot m_i),$$

其中, $N_{VTC}$  表示批次中样本的数量, $y_i$  表示真实标签, $\hat{y}_i$  表示模型的预测值, $m_i$  是掩码,用于在计算损失时忽略某些不重要或不相关的像素区域,从而专注于图像中有意义的部分, $L_{VTC}$  表示最终所有样本损失的平均值。

实验结果:在实验中,将对比 LRc-GAN<sup>[22]</sup>、AE-GAN<sup>[23]</sup>、DAE<sup>[10]</sup> 模型,以评估本文方法的性能(见图 5 和表 1)。CC filtering 指连通分量滤波,±表示 3 次运行的标准差,↑表示越大越好,↓表示越小越好。从上面的可视化实验结果可以看出,VTC 模块相比于原始 DAE,A 提高了约 11 个百分点,Dice 指数提高了约 12 个百分点,证明了 VTC 模块在图像融合去噪处理中是有效的。

### 2.3 图像分割

实验设置:使用 U-Net 作为基准模型,以对比本文提出的融合 Transformer 和跳跃连接的改进方法。

分割网络使用改进的 U-Net 架构,训练过程采用的损失函数为 Dice 损失函数( $L_{Dice}$ )和交叉熵损失函数( $L_{CE}$ )的加权和( $L_{TSC}$ ),学习率设置为 0.001,动量参数 $\beta_1=0.900$ , $\beta_2=0.999$ ,批量大小为 2.为了充分学习注意力机制,训练 2 000 轮。

评价指标: $A_P$ 、 $I_{Dice}$  和 Hausdorff 距离( $D_H$ ). $D_H$  用于衡量 2 个点集之间相似性,常用于评估图像分割结果的精度.对于 2 个点集  $A$  和  $B$ ,定义  $D_H(A, B) = \max(\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|b - a\|)$ ,为这 2 个点集中最近点之间的最大距离,其中,  $\| \cdot \|$  表示欧氏距离。

表 1 图像处理定量对比(BraTS2021)

Tab. 1 Quantitative Comparison of Image Processing(BraTS2021)

网络模型	$A_P$	$I_{Dice}$	$I_{Dice} (+CC \text{ filtering})$
LRc-GAN <sup>[22]</sup>	0.722 1±0.008	0.719 5±0.010	0.700 8±0.007
AE-GAN <sup>[23]</sup>	0.738 2±0.006	0.739 8±0.004	0.744 5±0.005
DAE <sup>[10]</sup>	0.775 3±0.004	0.756 0±0.008	0.785 7±0.006
VTC(ours)	<b>0.881 4±0.005 ↑</b>	<b>0.877 1±0.006 ↑</b>	<b>0.888 4±0.005 ↑</b>

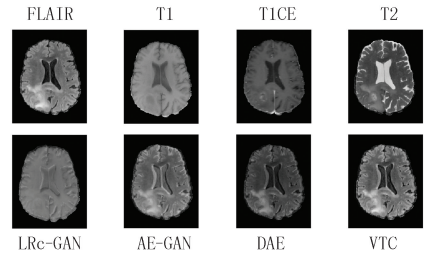


图5 图像处理定性对比

Fig.5 Qualitative comparison of image processing

损失函数:交叉熵损失主要应用于分类问题,它通过衡量预测概率分布与实际标签分布之间的差异来评估模型性能.在图像分割中,每个像素点视为一个分类问题,定义交叉熵损失  $L_{CE} = - \sum_{i=1}^N (y_i \ln p_i + (1 - y_i) \ln(1 - p_i))$ ,其中, $y_i$  表示第  $i$  个像素的真实标签, $p_i$  表示第  $i$  个像素属于目标类的预测概率。

骰子损失是基于指数定义的,用于衡量预测结果与真实标签之间的重叠程度,因此适用于不平衡数据集.定义骰子损失  $L_{Dice} = (2 \times |P \cap G|) / (|P| + |G|)$ ,其中, $P$  和  $G$  分别表示预测结果和真实标签。

为了结合交叉熵损失和骰子损失的优势,定义总体损失函数  $L_{TSC} = \alpha L_{CE} + \beta L_{Dice}$ ,其中, $\alpha$  和  $\beta$  是损失函数的权重系数,设置  $\alpha = \beta = 0.5$ 。

实验结果:对比了 4 种方法 U-Net<sup>[1]</sup>、U-Net++<sup>[2]</sup>、TransBTS<sup>[18]</sup>、UC-TransNet<sup>[21]</sup>(见附录图 S1 和表 2)。

以上方法均使用融合去噪后的图像.实验结果表明 TSC 模块优于其他方法的分割效果。

表 2 定量结果对比(BraTS2021)

Tab. 2 Comparison of quantitative results(BraTS2021)

网络模型	$A_P$	$I_{Dice}$	$D_H$
U-Net <sup>[1]</sup>	0.830 6±0.005	0.769 5±0.004	3.354 1±0.006
U-Net++ <sup>[2]</sup>	0.847 0±0.006	0.779 8±0.004	2.925 6±0.005
TransBTS <sup>[18]</sup>	0.849 3±0.003	0.786 0±0.006	2.880 3±0.003
UC-TransNet <sup>[21]</sup>	0.850 3±0.005	0.786 0±0.004	2.656 4±0.004
TSC(ours)	<b>0.883 9±0.006 ↑</b>	<b>0.834 4±0.005 ↑</b>	<b>2.356 6±0.006 ↓</b>

### 2.4 消融实验

为了验证各模块对最终图像分割的影响,进行了消融实验.(a) baseline(U-Net);(b) baseline+TSC:在 baseline 中集成了 TSC 模块,但未使用多模态融合去噪图像处理方法,而是在改进后的 TSC 模块基础上进行实验,以验证 TSC 模块的分割性能;(c) baseline+TSC+VTC:在 baseline 中同时集成了 TSC 和 VTC 模

块,验证 VTC 的关键作用。

从表 3 可以看出,在加入 TSC 模块后提高了分割的性能,结合对比实验可以证明该模块的有效性。在使用 VTC 模块后得到的实验结果比单独使用 TSC 时的结果要好,表明了该模块在分割任务中的关键作用。

表 3 消融实验定量结果(BraTS2021)

Tab. 3 Ablation experiment quantitative results(BraTS2021)

模块名称	$A_p$	$I_{Dice}$	$D_H$
(a)baseline	0.830 6±0.005	0.769 5±0.005	3.154 1±0.003
(b)baseline+TSC	0.860 1±0.003	0.800 9±0.002	2.664 1±0.002
(c)baseline+VTC+TSC	<b>0.883 9±0.004 ↑</b>	<b>0.834 4±0.003 ↑</b>	<b>2.356 6±0.003 ↓</b>

### 3 结 论

为了实现多模态 MR 图像去噪与精确分割,本文提出了 VTC 和 TSC 模块。相比以往的多模态融合方法,既充分考虑噪声干扰问题,VTC 模块实现多模态 MR 图像的融合去噪,辅助脑肿瘤分割,又解决了分割精度的问题。TSC 模块基于 U-Net 模型,结合 Transformer 和跳跃连接对模态融合和去噪后的图像进行分割,深度提取全局和局部特征,提高了分割效果。实验结果表明,该方法显著提高了含噪声的多模态 MR 图像的分割精度。提出的 2 种方法比同类方法更好地融合了多模态特征,去除噪声,提高脑肿瘤的分割精度。未来,将尝试更多数据集,以提升模型的泛化能力。

附录见电子版(DOI:10.16366/j.cnki.1000-2367.2024.07.17.0001)。

### 参 考 文 献

- [1] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[M]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [2] ZHOU Z W, RAHMAN SIDDIQUEE M M, TAJBAKHSI N, et al. UNet++: a nested U-Net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer, 2018: 3-11.
- [3] 殷晓航,王永才,李德英.基于 U-Net 结构改进的医学影像分割技术综述[J].软件学报,2021,32(2):519-550.
- [4] YIN X H, WANG Y C, LI D Y. Suvery of medical image segmentation technology based on U-Net structure improvement[J]. Journal of Software, 2021, 32(2): 519-550.
- [5] ZHANG J X, JIANG Z K, DONG J, et al. Attention gate ResU-net for automatic MRI brain tumor segmentation[J]. IEEE Access, 2020, 8: 58533-58545.
- [6] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [EB/OL]. [2020-10-22]. <https://arxiv.org/abs/2010.11929>.
- [7] CHEN J, LU Y, YU Q, et al. Transunet: Transformers make strong encoders for medical image segmentation[EB/OL]. [2024-07-08]. <https://arxiv.org/abs/2102.04306>.
- [8] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.
- [9] BAID U, GHODASARA S, MOHAN S, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification[EB/OL]. [2021-07-05]. <https://arxiv.org/abs/2107.02314>.
- [10] ATLASON H E, LOVE A, SIGURDSSON S, et al. Unsupervised brain lesion segmentation from MRI using a convolutional autoencoder [C]//Medical Imaging 2019: Image Processing. San Diego: SPIE, 2019: 52.
- [11] KASCENAS A, PUGEAULT N, ONEIL A Q. Denoising autoencoders for unsupervised anomaly detection in brain MRI[C]// International Conference on Medical Imaging with Deep Learning. [S.l.]: Springer, 2022: 653-664.
- [12] ZHOU C H, DING C X, LU Z T, et al. One-pass multi-task convolutional neural networks for efficient brain tumor segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2018. Cham: Springer, 2018: 637-645.
- [13] HUANG Z, XU H, SU S, et al. A computer-aided diagnosis system for brain magnetic resonance imaging images using a novel differential feature neural network[J]. Computers in Biology and Medicine, 2020, 121: 103818.
- [14] 毋小省,杨奇鸿,唐朝生,等.融合注意力机制的多模态脑肿瘤 MR 图像分割[J].计算机辅助设计与图形学学报,2023,35(9):1429-1438.

- WU X X, YANG Q H, TANG C S, et al. Multimodal brain tumor MR image segmentation network fused with attention mechanism[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2023, 35(9): 1429-1438.
- [14] ZHANG Y, HE N J, YANG J W, et al. mmFormer: Multimodal Medical Transformer for Incomplete Multimodal Learning of Brain Tumor Segmentation[C]// *Medical Image Computing and Computer Assisted Intervention-MICCAI 2022*. Cham: Springer, 2022: 107-117.
- [15] IBTEHAZ N, RAHMAN M S J N N. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation [EB/OL]. [2019-02-11]. <https://arxiv.org/abs/1902.04049>.
- [16] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: Learning where to look for the pancreas[EB/OL]. [2024-07-11]. <https://arxiv.org/abs/1804.03999>.
- [17] HUANG Z, ZHAO Y W, LIU Y H, et al. GCAUNet: a group cross-channel attention residual UNet for slice based brain tumor segmentation[J]. *Biomedical Signal Processing and Control*, 2021, 70: 102958.
- [18] WANG W X, CHEN C, DING M, et al. TransBTS: multimodal brain tumor segmentation using transformer[M]// *Medical Image Computing and Computer Assisted Intervention-MICCAI 2021*. Cham: Springer International Publishing, 2021: 109-119.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2024-06-12]. <https://arxiv.org/abs/1706.03762>.
- [20] 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 2016*. Cham: Springer, 2016: 424-432.
- [21] WANG H, CAO P, WANG J. Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(3): 2441-2449.
- [22] ZHAN B, LI D, WANG Y, et al. LR-cGAN: Latent representation based conditional generative adversarial network for multi-modality MRI synthesis[J]. *Biomedical Signal Processing and Control*, 2021, 66: 102457.
- [23] CAO B, BI Z, HU Q, et al. Autoencoder-driven multimodal collaborative learning for medical image synthesis [J]. *International Journal of Computer Vision*, 2023, 131(8): 1995-2014.

## Multimodal brain tumor MR image segmentation based on fusion denoising

Shen Junli, Lin Chongyu, Hai Yuman

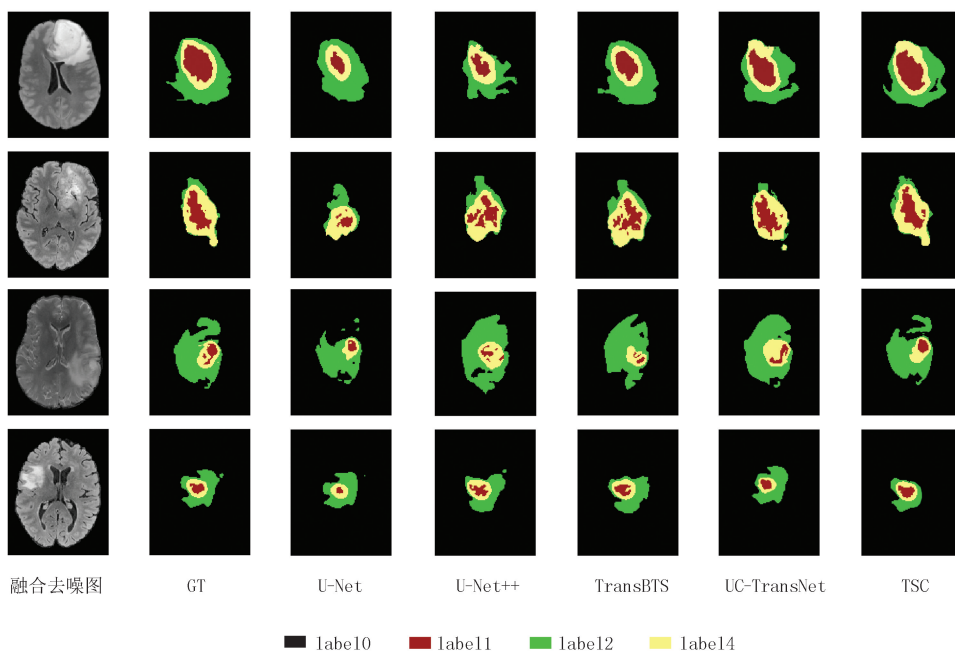
(School of Computer and Information Engineering(School of Artificial Intelligence), Henan Normal University, Xinxiang 453007, China)

**Abstract:** The segmentation of multimodal brain tumor MR images faces the problems of difficulty in modal fusion, easy generation of noise during the fusion process, and insufficient extraction of global and local features after modal processing, resulting in reduced segmentation accuracy. This paper first designs a module(VTC) that combines vision Transformer and deconvolution to achieve efficient fusion and deep denoising of multimodal MR images. Then, a skip connection module(TSC) fused with Transformer is created. The traditional skip connection is innovated by using the multi-head cross-attention mechanism of Transformer to capture the global and local features of the modal feature map after fusion and denoising, further improving the segmentation accuracy. Finally, the experimental results on the public dataset BraTS2021 show that this method achieves an accuracy of 88.39% when segmenting tumors, a Dice coefficient of 83.44%, and a Hausdorff distance of only 2.356 6.

**Keywords:** brain tumor MR; fusion denoising; U-Net; Transformer; image segmentation

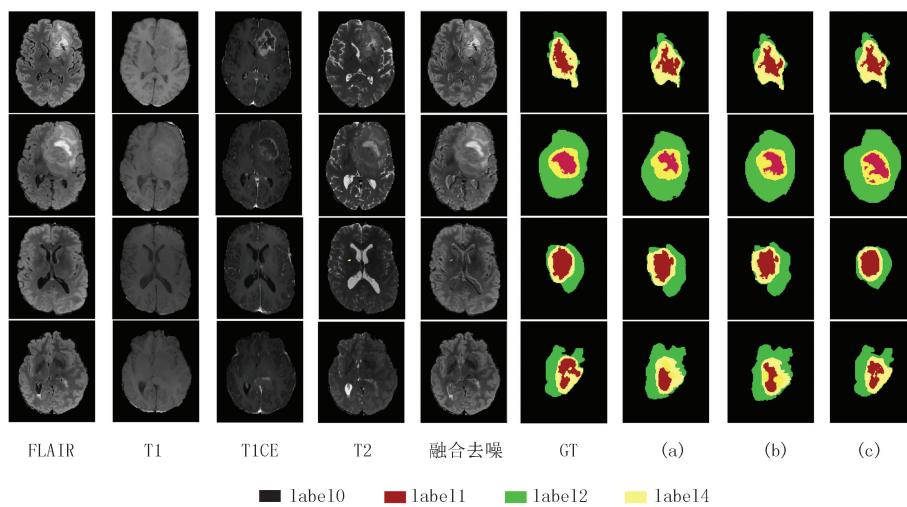
[责任编辑 杨浦 刘洋]

附 录



图S1 5种分割方法的定性结果对比

Fig.S1 Comparison of qualitative results of five segmentation methods



图S2 消融实验定性结果

Fig.S2 Qualitative results of ablation experiments