

DOI:10.3969/j.issn.1005-202X.2024.08.005

医学影像物理

基于交叉注意力的CT与MRI直肠癌病灶分割方法

邓杰夫¹, 奚峥皓¹, 黄陈², 刘翔¹

1. 上海工程技术大学电子电气工程学院, 上海 201620; 2. 上海交通大学附属第一人民医院普外科, 上海 200080

【摘要】针对直肠癌辅助诊断过程中部分医学图像分割模型只适用于单一模态图像的局限性,提出一种基于交叉注意力机制并同时适用于CT与MRI两种模态的医学图像分割方法。首先,针对CT与MRI两种图像特征表示不同,提出一种交叉注意力机制统一两类图像的特征表示;其次,针对直肠癌图像病灶区域较小的特点,提出三支路的改进Swin Transformer分割网络,并将交叉注意力机制加入其中,使模型对两种图像都能进行病灶区域的分割;最后,使用该方法在直肠癌患者的CT与MRI影像数据进行验证,与ADDA、CycleGAN和SIFA方法相比,在CT图像上的准确率分别提高2.94%、3.05%、0.71%,在MRI图像上的准确率分别提高了3.31%、4.55%、1.76%。实验结果表明该方法对于两类图像具有更好的分割性能。

【关键词】直肠癌;语义分割;交叉注意力;Transformer

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2024)08-0953-07

Segmentation of rectal cancer lesions on CT and MRI based on cross attention

DENG Jiefu¹, XI Zhenghao¹, HUANG Chen², LIU Xiang¹

1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China; 2. Department of General Surgery, Shanghai General Hospital, Shanghai Jiao Tong University, Shanghai 200080, China

Abstract: In response to the limitation of some medical image segmentation models for rectal cancer auxiliary diagnosis that are only applicable to single-modality images, a medical image segmentation method based on a cross attention mechanism that is applicable to both CT and MRI modalities is presented. Considering the different feature representations of CT and MRI images, a cross attention mechanism is proposed to unify the feature representations of the two types of images. In view of the small lesions on rectal cancer images, an improved Swin Transformer segmentation network with 3 branches is established, and the cross attention mechanism is incorporated into it, enabling the model to segment lesion areas in both types of images. The proposed method is validated using CT and MRI image data from patients with rectal cancer. Compared with ADDA, CycleGAN, and SIFA methods, the proposed method improves the accuracy by 2.94%, 3.05%, 0.71% on CT images, and 3.31%, 4.55%, 1.76% on MRI images, respectively, demonstrating its superior segmentation performance for both types of images.

Keywords: rectal cancer; semantic segmentation; cross attention; Transformer

前言

直肠癌已成为全球第3大常见癌症,也是第2大

癌症死亡原因^[1]。在临床实践中,放射科医生需要结合病人CT以及MRI等图像手动定位直肠癌病变的区域,为后续的治疗方案做准备,然而人工的定位方式耗时,因此一个能自动分割多类医学图像的系统能够对后续的诊断治疗提供较大的帮助。

近年来,深度学习方法已广泛应用于各个领域且取得显著的成效,其中在医学图像分割方面也取得巨大成功^[2-3]。目前大多数较为成熟的医学图像分割模型,如U-Net模型^[4],都是以全监督学习技术为基础,对训练图像的标签进行学习以后,再对新的图像进行分割,这使得一个模型的训练常常需要大量的有标签的图像,标签常常是费时且昂贵的,而这一

【收稿日期】2024-04-15

【基金项目】国家自然科学基金(12104289);上海申康医院发展中心临床三年行动计划(SHDC2020CR4022);上海市教育委员会高峰高原学科建设计划(20191425);CSCO-青年创新肿瘤研究基金(Y-Young2020-0458);上海市松江区科技攻关项目(2020SJ255)

【作者简介】邓杰夫,硕士研究生,研究方向:计算机视觉、医学图像处理,E-mail: m320121349@sues.edu.cn

【通信作者】奚峥皓,博士,副教授,研究方向:计算机视觉、智能认知学习与控制,E-mail: zhenghaoxi@hotmail.com

问题在医学图像领域尤其突显,受各种原因影响,研究人员常常无法获得大量的、对应的且带有标签的医学图像数据对一个全监督模型进行训练。因此,如何充分利用这些已有的、少量的有标签图像以及对应的无标签图像,成为目前医学图像分析领域的一个重要研究方向。

目前,已有许多工作致力于开发除完全监督学习技术之外的方法,如自监督^[5-9]、半监督^[10-12]和弱监督学习方法^[13-15]。然而,这些方法大多是为单一领域的场景设计的,忽略了实际临床应用场景,图像数据的来源常常是多种多样的,包括不同的医院和不同的成像扫描仪器。这导致了各种图像数据分布在不同的域中^[16-17]。因此,在一个域(即源域)上训练的模型通常不能很好地泛化到另一个域(即目标域)上收集的数据,模型在目标域上的性能严重下降。为了提高模型跨域的泛化性能,一种简单有效的方法是使用目标域新标注的数据对源域数据训练好的模型进行微调^[18]。然而,标记来自新领域的数据会花费额外的时间和费用成本。

无监督域自适应(Unsupervised Domain Adaptation, UDA)方法可以解决在没有目标域数据注释情况下的域转移,通过利用源域有标记的数据和目标域未标记的数据来最小化域之间的差异,从而提高模型跨域的自适应性能。大多数UDA都是基于卷积神经网络(Convolutional Neural Network, CNN)方法^[19],然而基于CNN的方法对目标域生成伪标签的噪音鲁棒性不强,无法实现精准的域对齐^[20]。随着Transformer在自然语言处理(Natural Language Processing, NLP)与计算机视觉(Computer Vision, CV)领域的成功, Li等^[21]发现Transformer中的交叉注意力机制可以对齐不同分布的数据,甚至是不同的模态,例如图像对图像、图像对文本、文本对语音等。结果表明,该方法对伪标签中的噪音具有一定的鲁棒性。因此,本文将Transformer应用于UDA任务,利用其对噪音的鲁棒性以及能够将特征对齐的优点来改良CNN方法中的问题。

综上所述,本文的主要贡献如下:(1)改进了基于Swin Transformer骨干网络的图像分割网络;(2)搭建了一个共享权重的三支路框架,将交叉注意力机制模块加入模型中,提高模型对图像噪音的鲁棒性以提升模型分割性能;(3)广泛挖掘两个领域的知识,将两个特征域的知识自适应融合,实现跨领域的图像分割。将本文模型应用于CT和MRI两种类型的医学图像中,对这两种直肠癌图像中的肿瘤进行精确分割,综合结果和分析表明,本文提出的框架比现有的UDA方法具有更好的性能。

1 直肠癌CT与MRI图像分割方法

1.1 基于自注意力机制的改进交叉注意力模块

为解决直肠癌CT和MRI图像中均存在的病灶区域较小、正负样本不平衡问题,本文在图像特征编码阶段引入Vision Transformer(ViT)网络的自注意力机制^[22],如图1所示,强化图像中正样本之间的特征关系。

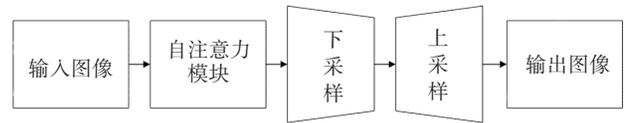


图1 图像特征编码流程

Figure 1 Image feature coding process

令长为 H 像素,宽为 W 像素,通道数为 C 的CT图像组 I_s 和MRI图像组 I_t 构成数据集 $R=\{I_s, I_t\}$ 。利用ViT的编码结构,将 R 中的任意图像 I 分成 N 个图像块(patch),对其中任意第 i 个patch平展为二维向量 $\mathbf{x}_i \in R^{N \times (D^2 \cdot C)}$ 。其中, D 为patch的边长像素, $N = H \cdot W / D^2$ 为 I 的patch数。则 I 的特征可用矩阵 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N]^T$ 表示, \mathbf{X}_s 和 \mathbf{X}_t 分别表示 I_s 和 I_t 的特征可用矩阵。

以ViT网络模型中的自注意力机制为依据^[23],将 \mathbf{X}_s 线性映射为 \mathbf{Q}_s 、 \mathbf{K}_s 和 \mathbf{V}_s 3个特征矩阵:

$$\mathbf{Q}_s = \mathbf{X}_s \mathbf{W}_Q \quad (1)$$

$$\mathbf{K}_s = \mathbf{X}_s \mathbf{W}_K \quad (2)$$

$$\mathbf{V}_s = \mathbf{X}_s \mathbf{W}_V \quad (3)$$

其中, \mathbf{Q}_s 为CT图像 I_s 中由每个patch位置信息构成的矩阵, \mathbf{K}_s 为 I_s 中由每个patch与其他patch间位置关系构成的矩阵, \mathbf{V}_s 为 I_s 中每个patch特征信息构成的矩阵。 \mathbf{W}_Q 、 \mathbf{W}_K 、 \mathbf{W}_V 分别为可通过网络学习实时更新参数矩阵。

令 \mathbf{K}_s 的维度为 d_k ,结合式(1)和式(2)可获得 I_s 的各个patch间相似度关系为:

$$\text{DD} = \frac{\mathbf{Q}_s \mathbf{K}_s^T}{\sqrt{d_k}} \quad (4)$$

对式(4)进行 soft_{\max} 归一化,得到相似度的权重矩阵 \mathbf{W}_s :

$$\mathbf{W}_s = \text{soft}_{\max}(\text{DD}) \quad (5)$$

结合式(3)和式(5)可得式(6),从而实现图像 I_s 的各个patch间关系的关注度,即强化对 I_s 的特征关注。

$$\text{Att}_{\text{self}}(\mathbf{Q}_s, \mathbf{K}_s, \mathbf{V}_s) = \mathbf{W}_s \mathbf{V}_s \quad (6)$$

同理,针对MRI图像组 I_t ,可获得关于 I_t 的3个特

征矩阵 Q_i, K_i 和 V_i , 相似度的权重矩阵 W_i , 则图像 I_i 的自注意力特征表征如式(7)所示:

$$\text{Att}_{\text{self}}(Q_i, K_i, V_i) = W_i V_i \quad (7)$$

式(6)和(7)分别关注 R 中的CT图像组 I_s 和MRI图像组 I_t , 不能互通使用。这也增加了训练过程的工作量和样本规模。

本文据此构建一种基于直肠癌CT和MRI图像的交叉注意力模块, 考虑到在不同类型直肠癌影像数据中, 病灶区具有位置信息大致相同, 但呈现出特征分布不同的特点, 构建式(8)来描述 R 中 I_s 与 I_t 图像组任意两张图像间 patch 关系:

$$\text{Att}_{\text{cross}}(Q_s, K_t, V_t) = \text{soft}_{\text{max}}\left(\frac{Q_s K_t^T}{\sqrt{d_k}}\right) V_t \quad (8)$$

其中, Q_s 由式(1)获得, K_t 和 V_t 由图像组 I_t 经式(2)和式(3)计算获得。交叉注意力模块计算出的每个输出, 包含了 I_s 与 I_t 图像中不同 patch 之间的相似度关

系。由此, 通过交叉注意力模块, 可将分别来自CT和MRI的两张直肠癌图像中的相似特征进行聚合。

1.2 基于 Swin Transformer 的改进分割网络

一般来说, 数据集 R 中的图像分辨率较大, 若沿用 ViT 模型, 一方面不能提取多尺度特征信息; 另一方面在保持 patch 尺寸大小不变的同时, 进行每一层网络实现输入和输出, 会增加 ViT 模型的计算复杂度。相比之下, Swin Transformer 在每一层网络的输出 patch 维度更小, 提取多尺度特征信息的能力更优^[24]。但 Swin Transformer 在解决 R 中直肠癌图像病灶区域较小的问题时, 因 R 的数据集规模较小, 网络训练至深层时会使病灶特征丢失, 导致过拟合。因此, 本文对 Swin Transformer 模型进行 U 型网络结构改进, 并添加跳跃连接, 以达到将网络底层信息直传上层, 优化网络学习的特征质量, 减少训练数据过拟合的目的(图2)。

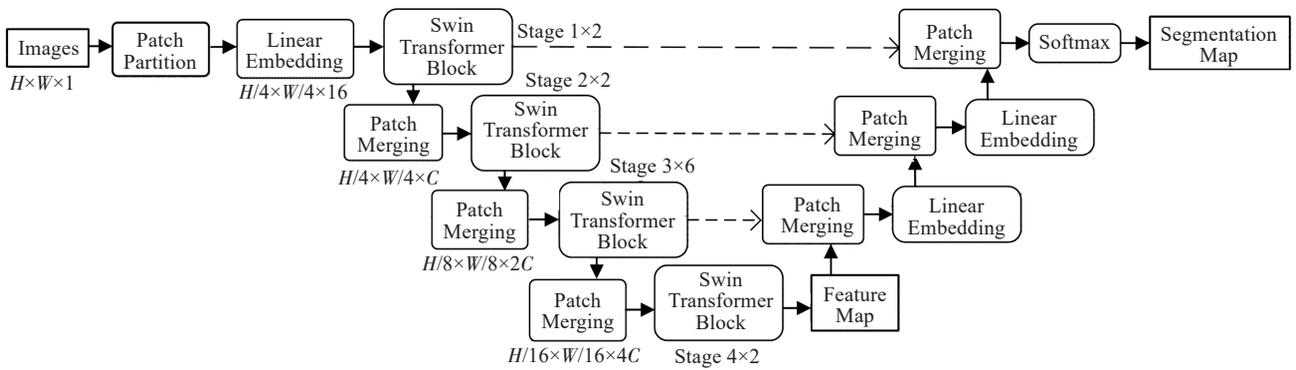


图2 改进的 Swin Transformer 网络结构

Figure 2 Improved Swin Transformer network structure

输入大小为 $H \times W \times 1$ 的 I_s 或 I_t 图像。经过 patch Partition 层后, 可得到 16 个 $H/4 \times W/4$ 大小的 patch。再经 Linear Embedding 层可得 C 个大小为 $H/4 \times W/4$ 的 patch, 即将原有的多个 patch 由全连接层编码成维度更高的特征, 使得图像可用更多数量的 patch 表示。其中, 设 $C=32$, 即为原有 patch 数的两倍。再在 Swin Transformer Block 对每个 patch 做自注意力计算, 即得维度为 $H/4 \times W/4 \times C$ 的 patch。最后, 在 Patch Merging 层, 将每个 patch 拆分成两个大小为 $H/8 \times W/8$ 的 patch, 即获得维度为 $H/8 \times W/8 \times 2C$ 的 patch。将每个 patch 对应的特征点位置进行如图3所示的标序, 并按照序号位置拼接, 可得维度更小, 数量更多的 patch, 重复上述操作, 在4次下采样后可获得维度为 $H/16 \times W/16 \times 4C$, 用于表征图像特征的 patch 向量。

依据 U 型对称原则, 使用插值法在解码器中进行

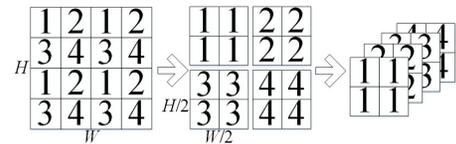


图3 Patch Merging 过程示意图

Figure 3 Patch Merging operation process

4次上采样操作。此过程将两个相邻 patch 拼接为一个, 即对 Patch Merging 反操作, 将特征向量逐步还原为与原图大小相同的分割结果图。

1.3 基于交叉注意力的三支路并行分割网络

若将数据集 R 中的任意图像输入至章节 1.2 所述的单个网络模型进行训练, 会因 I_s 和 I_t 图像组的特征不同, 导致模型混淆训练特征, 使得对病灶区域的分割性能下降。为此, 本文提出一种三支路并行分割网络结构(图4)。

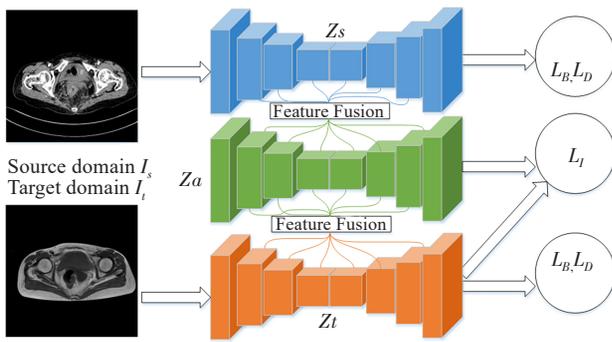


图4 模型整体框架

Figure 4 Overall framework of the model

定义3条支路分别为源分支 Z_s 、目标分支 Z_t 、自适应分支 Z_a 。 Z_s 对应学习 CT 图像 I_s (源域) 数据, Z_t 对应学习 MRI 图像 I_t (目标域) 数据, Z_a 用于两个域公共的数据特征表示, 如图 5 所示。

在 Z_s 与 Z_t 中, 使用自注意力模块来学习一类图像的特征表示。 Z_a 中 Q_s, K_t 和 V_t 依据式 (8) 分别来自于 Z_s 和 Z_t , 即利用交叉注意力模块聚合 Z_s 和 Z_t 支路上的特征信息, 输出将两个域数据对齐以后的特征。

在数据集 R 进行学习时, 当网络传播到深层, 因数据集规模较小, 使得模型容易丢失浅层的特征信息。为此, 本文在三支路并行分割网络中添加残差连接 (Residual Connection), 将每一层的特征输出与

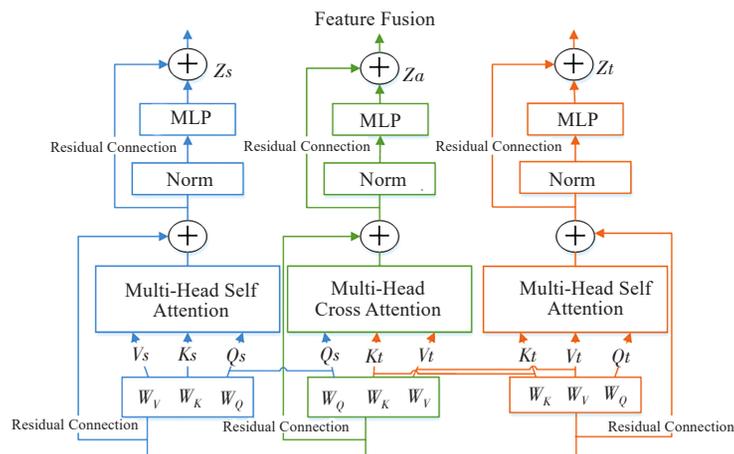


图5 三支路特征编码融合过程

Figure 5 Three-branch feature coding fusion process

上一层的输出叠加, 使网络能够始终在上一层学习结果的基础上进行下一层学习, 从而避免因梯度消失/爆炸、参数无法更新而导致的网络训练失败。

1.4 损失函数

损失函数由 3 部分组成: 二元交叉熵损失函数、Dice 损失函数和结构一致性损失函数。其中, 交叉熵损失函数定义如式 (9) 所示:

$$L_B = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (9)$$

其中, N 为像素点的数量, y_i 为第 i 个像素点真实的类别, p_i 表示模型分割结果中的第 i 个像素点属于真实类别的概率。利用式 (9) 对图像分割结果中的每个像素点进行计算。

另一方面, 在利用深度学习网络解决医学图像分割任务时, 常面临正负样本不平衡问题, 即背景像素点远大于前景像素点。此时, 仅用二元交叉熵损失函数会使模型越来越关注背景像素类的分割准确

率, 而忽略前景像素点。为此, 本文引入如式 (10) 所示的 Dice 损失函数:

$$L_D = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (10)$$

其中, X 表示网络模型分割结果值, Y 表示图像病灶区域经标注后的真实值。 $|X \cap Y|$ 为 X 集合与 Y 集合中具有相同元素的个数, $|X|$ 与 $|Y|$ 分别表示 X 集合与 Y 集合中的元素个数。Dice 损失函数反映了模型的分割结果与真实标注图像之间的整体相似性。

对于目标分支与自适应分支, 希望两个网络 Z_t 、 Z_a 在对同一张图像的分割结果上尽量保持一致, 因此使用结构一致性损失函数:

$$L_I = \frac{1}{N} \sum_{i=1}^N \|(-T_i^s \log(T_i^s)) - (-T_i^t \log(T_i^t))\|^2 \quad (11)$$

其中, T_i^s 表示 Z_s 支路中图像第 i 个像素点的值。综合式 (9)~式 (11), 构建最终损失函数如式 (12) 所示:

$$L = L_B + L_D + L_I \quad (12)$$

2 实验结果与分析

2.1 数据集与实验环境

本实验所用数据集 R 由课题合作单位上海交通大学医学院采集标注,共采集 50 例直肠癌患者的 CT 及 MRI 影像。将所得三维影像数据沿 Z 轴方向进行切片,其中选取包含直肠部位的二维图像共 1 564 张。

本文算法实验运行环境为 Ubuntu 18.04 的操作系统,硬件采用 Intel i7-10700K 处理器和 NVIDIA GeForce GTX3080 显卡,软件架构平台为 PyTorch 1.90。模型训练时,所有图像的分辨率统一为 256×256 ,利用随机梯度下降算法优化训练过程,设置学习率为 0.000 1,权重衰减系数为 0.001,批量大小为 4,共训练 200 个 epoch。

2.2 评估指标

本文采用医学影像分割问题常用的准确率 (ACC)、交并比 (IoU)、Dice 相似性系数 (Dice Similarity Coefficient, DSC) 作为模型的性能评估指标,各指标定义如式 (13)~(15) 所示:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (14)$$

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (15)$$

其中,TP 为真阳性,即模型分割结果与标签同为正样本的数量;TN 为真阴性,即分割结果与标签同为负样本的数量;FP 为假阳性,即分割结果为正样本,标签为负样本的数量;FN 为假阴性,即分割结果为负样本,标签为正样本的数量。

此外,本文还采用参数量 (Parameter, Params) 和浮点运算符 (Giga Floating-Point Operations Per Second, GFLOPs) 指标分别表征模型的大小和计算复杂度。

2.3 对比实验

将本文模型与 ADDA^[25]、CycleGAN^[26] 和 SIFA^[27] 等先进 UDA 方法在数据集 R 上进行比较,结果如表 1 所示。从表中可看出,本文模型在 ACC 上比 ADDA、CycleGAN 和 SIFA 的方法在 CT 图像上分别提高了 2.94%、3.05%、0.71%,在 MRI 图像上的 ACC 分别提高了 3.31%、4.55%、1.76%。特别是在从源域下的 CT 图像自适应到目标域的 MRI 图像时,因交叉注意力模块的作用,将两个域中图像相似部分距离拉近,使模型学到了两个域间的公共特征,从而显著提升分割准确率。但是 3 个网络并行学习方法的使用,也使得本文模型具有较大的参数量与计算复杂度。

表 1 模型评价指标对比结果

Table 1 Model evaluation index comparison results

模型	Params/M	GFLOPs/G	ACC/%		IoU/%		DSC/%	
			CT	MRI	CT	MRI	CT	MRI
ADDA	120.42	94.26	92.46	90.80	75.85	70.89	84.54	79.60
CycleGAN	140.50	57.78	92.35	89.56	72.48	68.16	80.56	75.26
SIFA	163.14	50.34	94.69	92.35	77.54	74.36	87.34	83.50
本文模型	190.36	64.46	95.40	94.11	79.84	77.26	88.36	86.62

图 6 为本文模型以及对比模型在对不同患者的 CT 图像或 MRI 图像分割的可视化结果。从图中可以看出,针对于两种不同模式直肠癌图像的肿瘤分割,本文模型在对 CT 图像或 MRI 图像分割时,都能取得更好的分割效果,最接近人工标注的结果。

2.4 消融实验

对本文模型进行消融实验,如表 2 所示。由表 2 可见,加入残差连接之后,与基线模型相比,模型的 ACC 在 CT 图像上提升 4.97%,但在 MRI 图像分割的 ACC 仅提升 1.87%。而交叉注意力模块能够在模型跨域分割时发挥巨大的作用,使网络能够对 MRI 图

像进行分割时不至于损失太多的性能,在加入了交叉注意力模块以后,与基线模型相比,对 MRI 图像分割的准确率提升 11.61%。

3 结语

本文针对在利用机器视觉作为直肠癌辅助诊断过程中,因 CT 与 MRI 两种模态病灶图像特征表示不同而难以通过深度学习网络统一参数进行分割的问题,提出一种可聚合两种模态特征的交叉注意力机制,以更好地提取 CT 与 MRI 图像的共性特征。同时,本文将该模块引入三支路的改进 Swin

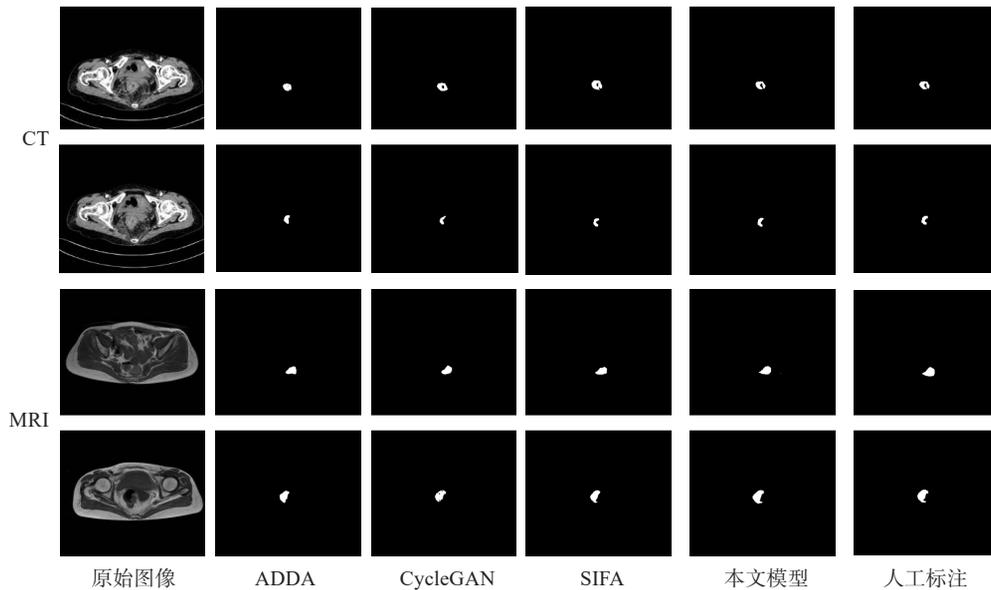


图6 模型分割结果

Figure 6 Segmentation results

表2 不同模块的消融实验结果

Table 2 Ablation study results of different modules

模型	Params/M	GFLOPs/G	ACC/%		IoU/%		DSC/%	
			CT	MRI	CT	MRI	CT	MRI
基线	142.60	52.64	90.26	78.59	76.59	62.16	83.49	71.89
+残差连接	165.35	55.75	95.23	80.46	79.65	64.23	86.90	72.16
+交叉注意力模块	167.43	60.36	92.46	90.20	76.23	73.49	84.05	82.41
+残差连接+交叉注意力模块	190.36	64.46	95.40	94.11	79.84	77.26	88.36	86.62

Transformer 分割网络中,实现可对两种模态图像同时进行分割的网络模型。经实验验证,本文模型在 ACC、IoU、DSC 方面显著优于 ADDA、CycleGAN 和 SIFA 等现有的模型方法。由于使用 3 个网络并行学习,使得模型整体参数量较大、推理速度较慢。未来笔者将在此模型的轻量化方面继续开展后续研究。

【参考文献】

[1] Bhandari A, Woodhouse M, Gupta S. Colorectal cancer is a leading cause of cancer incidence and mortality among adults younger than 50 years in the USA: a SEER-based analysis with comparison to other young-onset cancers[J]. J Investig Med, 2017, 65(2): 311-315.
 [2] Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-net: learning dense volumetric segmentation from sparse annotation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016. Cham: Springer International Publishing, 2016: 424-432.
 [3] Milletari F, Navab N, Ahmadi SA. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV). Piscataway, NJ, USA: IEEE, 2016: 565-571.
 [4] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer

International Publishing, 2015: 234-241.
 [5] Bai WJ, Chen C, Tarroni G, et al. Self-supervised learning for cardiac MR image segmentation by anatomical position prediction [C]// Medical Image Computing and Computer Assisted Intervention-MICCAI 2019. Cham: Springer International Publishing, 2019: 541-549.
 [6] Li XM, Jia MY, Islam MT, et al. Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis[J]. IEEE Trans Med Imaging, 2020, 39(12): 4023-4033.
 [7] Zhou ZW, Sodha V, Pang JX, et al. Models genesis[J]. Med Image Anal, 2021, 67: 101840.
 [8] Zeng Z, Yang XL, Yu QY, et al. SeSe-net: self-supervised deep learning for segmentation[J]. Pattern Recognit Lett, 2019, 128: 23-29.
 [9] Zhu JW, Li YX, Hu YF, et al. Rubik's cube+: a self-supervised feature learning framework for 3D medical image analysis[J]. Med Image Anal, 2020, 64: 101746.
 [10] Bai WJ, Oktay O, Sinclair M, et al. Semi-supervised learning for network-based cardiac MR image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2017. Cham: Springer International Publishing, 2017: 253-260.
 [11] Zhao ZY, Zhang XM, Chen C, et al. Semi-supervised self-taught deep learning for finger bones segmentation [C]//2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). Piscataway, NJ, USA: IEEE, 2019: 1-4.
 [12] Zhao ZY, Zeng Z, Xu KX, et al. DSAL: deeply supervised active learning from strong and weak labelers for biomedical image segmentation[J]. IEEE J Biomed Health Inform, 2021, 25(10): 3744-3751.

- [13] Rajchl M, Lee MC, Oktay O, et al. DeepCut: object segmentation from bounding box annotations using convolutional neural networks[J]. IEEE Trans Med Imaging, 2017, 36(2): 674-683.
- [14] Playout C, Duval R, Cheriet F. A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images[J]. IEEE Trans Med Imaging, 2019, 38(10): 2434-2444.
- [15] Belharbi S, Rony J, Dolz J, et al. Deep interpretable classification and weakly-supervised segmentation of histology images *via* max-min uncertainty[J]. IEEE Trans Med Imaging, 2022, 41(3): 702-714.
- [16] Gibson E, Hu YP, Ghavami N, et al. Inter-site variability in prostate segmentation accuracy using deep learning [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2018. Cham: Springer International Publishing, 2018: 506-514.
- [17] Guan H, Liu MX. Domain adaptation for medical image analysis: a survey[J]. IEEE Trans Biomed Eng, 2022, 69(3): 1173-1185.
- [18] Ghafoorian M, Mehrtash A, Kapur T, et al. Transfer learning for domain adaptation in MRI: application in brain lesion segmentation [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2017. Cham: Springer International Publishing, 2017: 516-524.
- [19] Jiang X, Lao QC, Matwin S, et al. Implicit class-conditioned domain alignment for unsupervised domain adaptation[C]//Proceedings of the 37th International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2020: 447.
- [20] Morerio P, Volpi R, Ragonesi R, et al. Generative pseudo-label refinement for unsupervised domain adaptation[C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2020: 3119-3128.
- [21] Li XY, Hou YH, Wang PC, et al. Trear: transformer-based RGB-D egocentric action recognition[J]. IEEE Trans Cogn Dev Syst, 2022, 14(1): 246-252.
- [22] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2021-06-03). <https://arxiv.org/abs/2010.11929>.
- [23] Brændengen M, Hansson K, Radu C, et al. Delineation of gross tumor volume (GTV) for radiation treatment planning of locally advanced rectal cancer using information from MRI or FDG-PET/CT: a prospective study[J]. Int J Radiat Oncol Biol Phys, 2011, 81(4): e439-e445.
- [24] Liu Z, Lin YT, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 9992-10002.
- [25] Tzeng E, Hoffman J, Saenko K, et al. Adversarial discriminative domain adaptation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 2962-2971.
- [26] Zhu JY, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 2242-2251.
- [27] Chen C, Dou Q, Chen H, et al. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation [J]. IEEE Trans Med Imaging, 2020, 39(7): 2494-2505.

(编辑:薛泽玲)