

基于DPU-Net的直肠癌肿瘤分割与T分期研究

康帅¹, 奚峥皓¹, 黄陈², 傅中懋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】基于MRI图像对直肠癌肿瘤进行分割与T分期识别在直肠癌术前诊断与治疗方案制定中起着重要作用。为了准确分割肿瘤并进行T分期,提出一种多任务学习模型DPU-Net,联合分割与分期任务。在分割分支中,针对直肠癌MRI图像结构复杂的特点,结合注意力机制与多尺度特征加强模型对肿瘤的关注度以及边缘特征提取能力,提高分割效果。在分类分支中,为解决直肠癌肿瘤T分期难的问题,引入诊疗文本,充分利用医疗数据。提出基于动态加权的多模态融合模型,结合图像特征与文本特征对肿瘤T分期识别。将本文模型与主流模型进行对比,实验结果表明,本文模型分割精度DSC为82.88%,相比于U-Net提升了17.96%。分类准确率为76.24%,相比于Dense-Net提高了9.43%。本文模型具备辅助医生诊断的可行性。

【关键词】直肠癌; 分割; T分期; 注意力机制; 多模态融合

【中图分类号】R318; TP391.41

【文献标志码】A

【文章编号】1005-202X(2023)10-1189-09

Rectal tumor segmentation and T staging based on DPU-Net

KANG Shuai¹, XI Zhenghao¹, HUANG Chen², FU Zhongmao², 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 School of Medicine, Shanghai 200080, China

Abstract: The segmentation and T staging of rectal tumor in MRI images are critical in preoperative diagnosis and treatment planning. A multi-task learning model, DPU-Net, is proposed to achieve both accurate tumor segmentation and T staging. In the segmentation path, considering the complex structure of rectal cancer MRI images, attention mechanism and multi-scale features are used to enhance the model's focus on tumor and the edge feature extraction ability, thereby improving segmentation performance. In the classification path, medical treatment texts are introduced to make full use of medical data; and a multi-modality fusion model based on dynamic weight which combines image features and text features is established for T staging. The experimental results show that the proposed model achieves a Dice similarity coefficient of 82.88%, which is 17.96% higher than U-Net, and that the staging accuracy is 76.24%, which is 9.43% higher as compared with Dense-Net. The proposed method is feasible for auxiliary diagnosis.

Keywords: rectal cancer; segmentation; T staging; attention mechanism; multi-modality fusion

前言

根据最新的全球癌症统计报告显示,直肠癌已成为全球范围内第3大常见癌症^[1]。我国是直肠癌

多发国家之一^[2],根据我国癌症中心统计数据显示,2020年我国新增直肠癌患者为55万,确诊人数居恶性肿瘤第2位,仅次于肺癌。直肠癌总死亡人数高达28万人,居恶性肿瘤死亡人数的第5位。直肠癌在临床上定义为从齿状线至直肠乙状结肠交界处之间的恶性肿瘤,具有进展期长的特点,很多患者确诊时即是晚期,致死率高达86%^[3]。直肠癌发病原因除遗传因素外,还与现代人饮食不规律、缺乏锻炼等可控因素有关。然而,由于现代人缺乏对癌症的防范意识和检测意识,加上医疗资源的分布不平衡,导致多数患者确诊即是中晚期,错过了早期诊疗的最佳时间。因此,尽早对直肠癌进行诊断筛查是阻断其发展的重中之重,具有实际意义。

【收稿日期】2023-05-08

【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】康帅,硕士,研究方向:计算机视觉、医学图像处理,E-mail: kstxsr1105@163.com

【通信作者】黄陈,博士,主任医师,研究方向:胃肠外科、胃肠肿瘤影像学组学,E-mail: richard-hc@hotmail.com

近年来,随着计算机技术的发展,通过结合计算机技术和医疗图像进行计算机辅助诊断是当前智慧医疗在中国的核心应用场景之一。通过计算机辅助诊断方法对直肠癌MRI图像进行肿瘤分割与T分期研究能够给医生提供参考,具有重大的临床价值。然而,直肠癌的MRI图像呈现出复杂多变的结构特征,肿瘤呈现相对较小的体积。此外,肿瘤与周围的肠壁、结节等组织结构紧密相连,增添了直肠癌诊断的难度。因此,直肠癌肿瘤分割与T分期是一项具有挑战性的任务。本文通过计算机技术对直肠癌MRI图像进行分析,提出一种多任务深度学习模型DPU-Net (Dual-Path U-Net),对直肠癌肿瘤进行分割与分期识别,辅助医生诊断,提高诊疗效率,缓解医疗资源不平衡的现状,使早筛查早诊断成为可能。本文的主要贡献如下:(1)DPU-Net基于U形结构扩展,联合分割与分类任务,共享特征权重,是首个能够同时处理直肠癌肿瘤分割与T分期的模型。(2)结合注意力机制与多尺度特征,增强肿瘤的感知能力,提高分割效果。(3)充分利用医疗资源,引入诊疗文本数据与图像相结合,提出基于动态加权的多模态融合模型实现肿瘤T分期识别。

1 相关工作

直肠癌肿瘤分割与T分期是医学图像分析领域的重要组成部分。临床上通过MRI图像分析肿瘤对直肠壁的浸润程度来评估癌症T分期,根据肿瘤的发展状况可分为以下4个分期:T1分期,肿瘤侵犯黏膜下层;T2分期,肿瘤侵犯固有肌层;T3分期,肿瘤穿透固有肌层到达浆膜下层;T4分期,肿瘤穿透腹膜脏层或侵犯其他器官。然而,人工阅片方式对影像科医生造成极大负担,容易出现误诊漏诊、主观性过强等问题。并且,医学图像的标注是一项费时费力的任务。因此,目前急需一种能够自动分割直肠癌肿瘤并进行分期的方法。

近年来,深度学习通过学习海量数据能够实现端到端的预测,成为了研究主流。在医学图像分割领域,全卷积神经网络(Fully Convolutional Networks, FCN)是开山之作,其通过反卷积层对特征图进行上采样,可以得到像素级的分割结果^[4]。U-Net模型在其基础上进行改进,极大地推动了该领域的发展。现有工作大多基于U-Net模型进行改进^[5]。例如,Otkay等^[6]结合注意力机制与U-Net,提出了Attn U-Net,增强了肿瘤分割效果。Alom等^[7]在编码器结构中加入循环,提出了R2U-Net,用于眼底血管分割,证明了模型的可行性。Xiao等^[8]结合残差结构对U-Net模型进行改进,增加了网络深度,在视网

膜分割实验中取得了不错效果。对于直肠癌肿瘤分割的研究主要基于U-Net模型进行研究。Zheng等^[9]在U-Net模型前构建了ROI提取模块,然后对肿瘤进行分割。这种做法可以排除无关器官的干扰,但是需要先验知识对图像预处理,无法实现快速的端到端预测。Wei等^[10]基于梯度下降变分法与U-Net提升直肠癌肿瘤分割。该方法需要对变分法调整大量超参数,算法鲁棒性较差,难于实现。Gao等^[11]构建了更为密集连接的U-Net++模型增强对肿瘤的感知能力,但是模型在肿瘤边缘等细节特征处分割效果较差。而在医学图像分类领域,现有工作大部分基于图像分类模型。Zheng等^[12]基于Fast-RCNN对胃癌肿瘤检测与分期。Esteva等^[13]基于Inception-v3对皮肤病良恶性分类进行识别。而随着研究的深入,医疗数据的多样化,更多的学者开始使用不同模态图像进行研究。韩坤等^[14]基于3D-CNN融合了MRI图像和PET图像特征,实现了对阿兹海默病的分类与诊断。Xiao等^[15]基于神经结构搜索的方法将神经胶质瘤的术中近红外线图像(Near-Infrared, NIR)图像与病理图像相融合,实现了术中胶质瘤分级。Wang等^[16]基于双流卷积神经网络(Convolutional Neural Networks, CNN),融合了彩色眼底照相术(Color Fundus Photography, CFP)和光学相干断层扫描(Optical Coherence Tomography, OCT)的图像信息,探讨了年龄相关性黄斑变性的自动分类。对于直肠癌肿瘤分期的研究则少之又少,现有研究无法做到准确的直肠癌肿瘤T分期识别。Liang等^[17]基于影像学特征构建最小绝对收缩和选择算子回归模型(Least Absolute Shrinkage and Selection Operator Regression, LASSO),用于肿瘤早期/晚期的识别。Kim等^[18]构建级联CNN模型,先对肿瘤进行检测,之后进行良恶性分类,但无法做到具体的T分期。本文认为,在实际诊疗中,分割与分类具有同等重要的临床意义。现有研究将两者独立研究,无法有效利用两种任务所提取的特征,导致分割效果不佳、无法实现T分期。因此,本文提出多任务学习模型DPU-Net,联合分割与分类任务,实现肿瘤分割与T分期。

2 本文方法

本文针对直肠癌MRI图像特点,基于U形网络进行扩展,提出了可同时对直肠癌肿瘤进行分割与分期的DPU-Net模型。模型结构如图1所示。整体模型包含两个分支:分割分支与分类分支。分割分支为编-解码器对称的U形网络,用于输出分割掩码。编码器中采用ImageNet预训练的DenseNet-121模型的卷积层来提取特征。并在编码过程中结合注意力

机制增强模型对肿瘤的关注程度。通过使用多核池化模块对编码特征进行处理,融合不同尺度的肿瘤特征增强模型对细节特征的提取能力。解码器由反卷积层构成上采样块,通过带注意力跳跃连接的方式,结合同一阶段编码器的信息对特征进行重建,得到分割结果。而分类分支通过引入诊疗文本数据并

进行特征提取,共享分割过程中的图像编码信息。通过基于 Dynamic Weight 的多模态融合模型将图像与文本特征进行融合增强,预测肿瘤 T 分期。最后通过两层多层感知机输出长度为 1×4 的向量,转化为预测概率值,将其中最大结果作为最后的 T 分期结果。

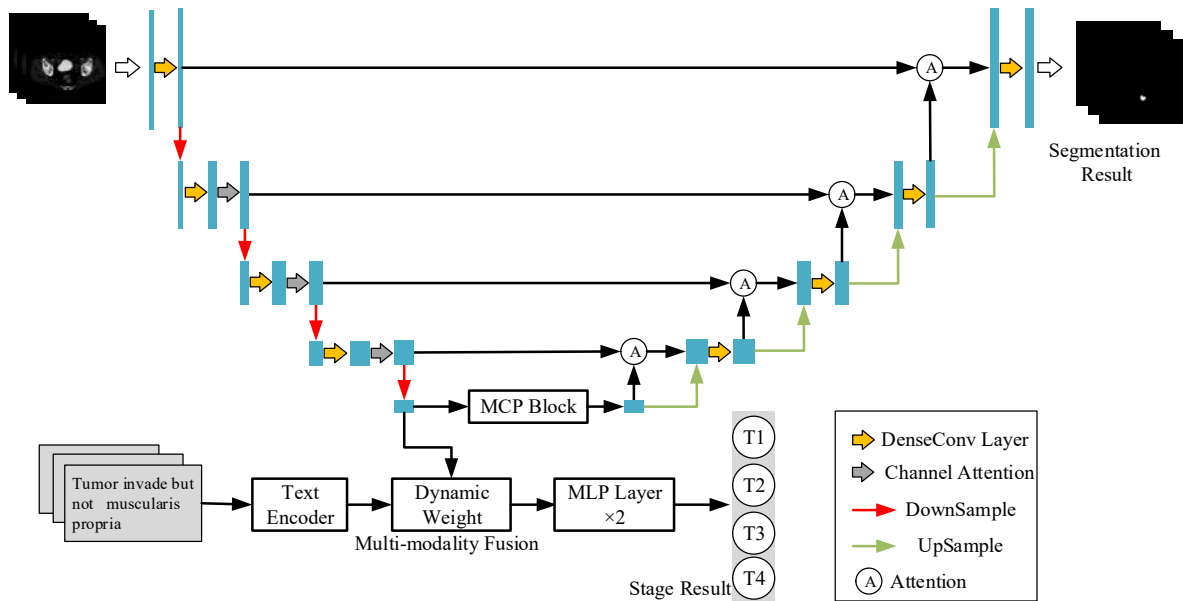


图1 DPU-Net模型结构

Figure 1 Structure of DPU-Net model

2.1 基于通道注意力的图像编码器

U-Net 模型采用传统卷积池化的方式对细胞图像进行特征提取。然而,不同医学图像之间差异较大,且医学图像数据有限,简单的特征提取方式难以在复杂的直肠癌 MRI 图像中学习到有效特征,模型会出现过拟合、欠拟合等现象。因此,本文将 ImageNet 上预训练的 DenseNet-121 迁移至本文模型中,通过采用前 4 层卷积层的初始化参数,能够有效增强网络的鲁棒性。并且,通过构建更为密集连接的卷积方式,能够有效增强网络的训练能力,提高特征的利用率,如图 2 所示。

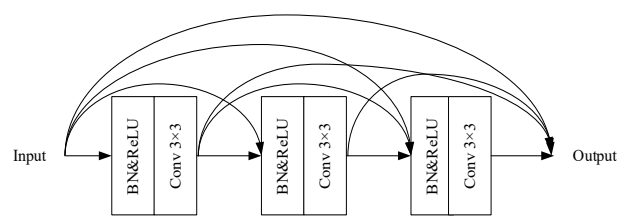


图2 密集连接卷积层

Figure 2 Dense connected convolutional layer

直肠癌 MRI 图像包括了盆腔区域的完整成像,涵盖了直肠及其周围组织结构,如直肠周围脂肪组织、肠壁、血管系统以及邻近的盆骨、髌骨等骨骼结构。相比于肝癌、胃癌等肿瘤,直肠癌肿瘤在 MRI 图像中占比较小,直肠内亦有其他干扰物质。因此,本文针对直肠癌肿瘤形态小而难以定位的问题,在编码过程中结合注意力机制^[19],在卷积层后加入通道注意力模块,加强模型对肿瘤的感知能力,能够准确分割肿瘤区域,如图 3 所示。该模块对卷积输出的特征建立相互依赖关系模型,分为 Squeeze 和 Excitation 两步。对于大小为 $H \times W \times C$ 的输入特征图,首先通过 Squeeze 压缩特征 U ,以全局平均池化 (Global Average Pooling, GAP) 的方式对特征进行压缩和信息汇总,得到通道描述关系,如公式(1)所示:

$$z_c = F_{\text{seq}}(v_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W v_c(i,j) \tag{1}$$

其中,每个通道的二维特征 v_c 大小为 $H \times W$,通过 GAP 操作将特征 v_c 压缩成大小为 $1 \times 1 \times C$ 的数值 z_c 。再对通道描述关系进行 Excitation 操作,如公式(2)所示:

$$s_c = F_{\text{weight}}(z_c, W) = \text{sigmoid}(W_2 \cdot \text{ReLU}(W_1 z_c)) \tag{2}$$

通过两个全连接层 W_1 、 W_2 得到通道注意力权重矩阵,自适应调整每个特征通道权值,得到特征通道间的相互关系 s_c 。并通过 F_{scale} 重新缩放特征图的大

小得到最终的输出特征 X_{new} 。通过该方式可强化模型对目标特征的关注度,抑制了无关器官的干扰,提高了模型的鲁棒性以及对肿瘤的敏感性。

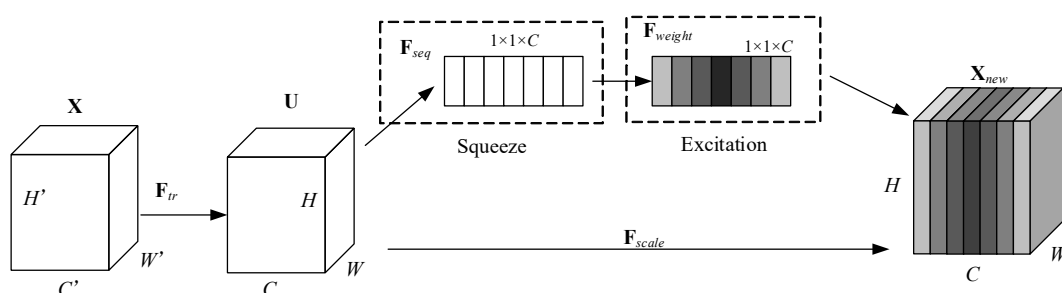


图3 通道注意力模块

Figure 3 Channel attention block

2.2 基于空间注意力的跳跃连接

模型在解码重建特征时,通过跳跃连接的方式结合同一深度的编码特征。在这个过程中,部分图像在成像过程中受到的伪影等干扰噪声也会随目标肿瘤特征一同融合重建,对分割效果造成影响。为避免该过程中掺杂了冗余噪声,本文通过构建基于空间注意力的跳跃连接,提高特征重建质量,如图4所示。

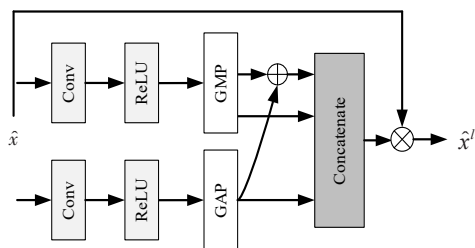


图4 空间注意力模块

Figure 4 Spatial attention block

模块采用并行分支,对第1层编码器特征 x' 进行 GAP 和全局最大池化(Global Max Pooling, GMP)操作,如公式(3)所示:

$$\hat{x}' = \text{MLP}(\text{Concat}(\text{GAP}(\text{Conv}(x')), \text{GMP}(\text{Conv}(x')), x')) \otimes x' \quad (3)$$

原特征图通过池化得到两个大小一致的特征图,并且结合拼接、叠加操作,合并语义信息相似的特征,然后通过多层感知机层(Multi-Layer Perceptron, MLP)对特征通道进行对齐,得到加强后的编码器特征 \hat{x}' 。将同一深度的编码器特征图与解码器特征图进行拼接,融合低维空间特征与高维语义特征,去除噪声干扰,优化解码过程。

2.3 基于多核池化的瓶颈设计

直肠癌肿瘤形态具有多变性,T1分期与T2分期肿瘤在形态上较小,特征不明显。T3分期与T4分期肿瘤在形态结构上表现并不规律,这两个分期的肿瘤边缘与直肠壁的相对位置接近,表现十分模糊,易与直肠壁混淆,难以准确分割并判断对直肠的浸润程度。因此,针对该问题,本文在瓶颈阶段(Bottleneck)构建基于多核池化(Multi-Core Pooling, MCP)模块,增强模型对肿瘤形态、边缘等高维语义特征的敏感性,如图5所示。MCP模块通过不同尺度的池化层对特征图进行处理。本模块中,选取池化核大小为2、3、5的平均池化层对编码结束后的特征图进行处理,经过 1×1 的卷积层进行通道校准并拼接。以不同的感受野提取不同大小的特征。大感受野用于提取肿瘤的全局形态特征,小感受野用于提取肿瘤的细小边缘特征,将大感受野下得到的宏观信息与小感受野下提取的局部细节信息进行融合,能够得到更为丰富的语义信息。通过这种多尺度特征融合方式能够有效加强模型对肿瘤边缘等细节特征的学习能力。

2.4 基于动态加权的多模态融合模型

相比于自然图像,医学图像获取渠道有限,公共数据集匮乏。医学工作者通过手工标注病灶区域构建数据集,不仅费时费力且数据规模有限,无法在深度学习模型上取得更好的效果。本文针对医学图像数据规模有限、获取不易的问题,引入其他研究者很少关注的诊疗文本数据。这部分文本数据通常与影像可同时获取。现有研究证明了基于医学图像特征难以做到直肠癌肿瘤T分期,但诊疗文本中包含丰富的、有意义的信息。通过将图像数据与文本数据相

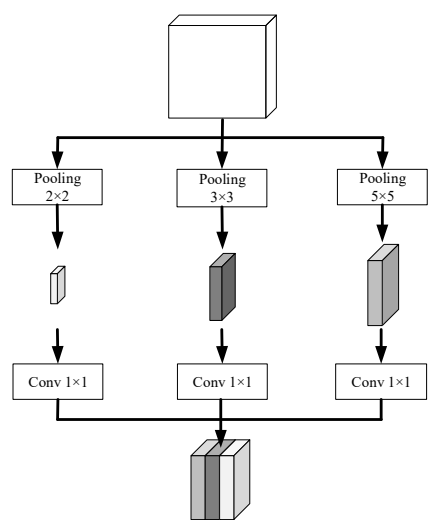


图5 多核池化模块
Figure 5 Multi-core pooling block

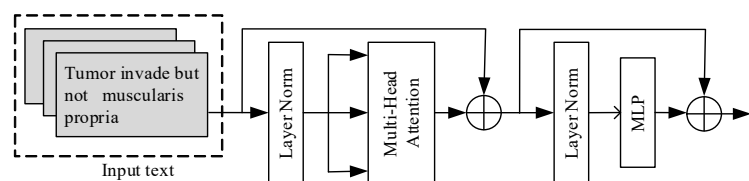


图6 文本特征提取器
Figure 6 Text feature extractor

成联合预测;不同分支的预测概率相乘得到最终概率。这类方法有一个共性:在进行特征交互时,元素之间都是一对一进行交互。这种单一的特征交互方式使得模型并没有完全学习得到全局的特征表示,限制了特征元素融合后的表达能力。因此本文采用一种交叉交互的方式,使得图像和文本中的每一个特征元素都能得到充分学习,增强特征的表达能力,如图7所示。

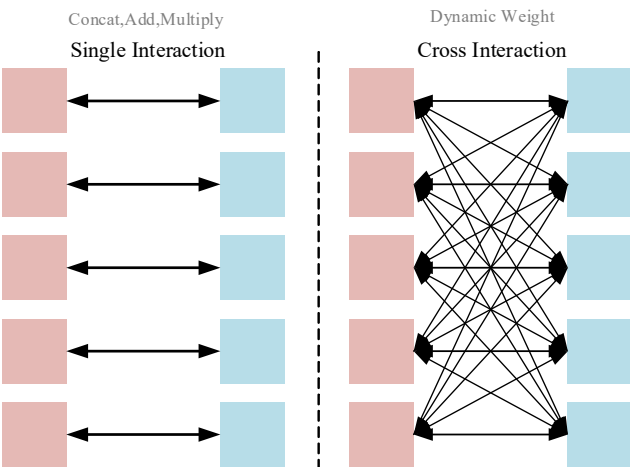


图7 特征交互方式对比
Figure 7 Comparison between feature interaction methods

结合,构建多模态融合模型对肿瘤T分期识别具备可行性,能够充分挖掘医疗资源的多样性。

多模态融合模型中的图像特征来自2.1节中的图像编码器。针对直肠癌肿瘤T分期任务,将肿瘤分期T1、T2、T3、T4作为分类目标,对诊疗文本数据进行预处理。筛选其中与肿瘤分期相关的信息,将描述肿瘤信息的文本作为文本输入。基于Transformer编码器构建文本特征提取器,如图6所示。

首先,对其添加嵌入编码,将其转换为序列向量。然后经过Transformer编码器进行自注意力计算,对特征编码,得到文本特征表示,与图像特征共同作为多模态融合模型的输入。

现有的多模态特征融合方式主要分为3种:不同模态特征的通道拼接;不同分支的预测结果相加形

并基于该种特征元素交互方式,设计了一种动态加权(Dynamic Weight)的方法,用于图像与文本特征融合,如图8所示。

z_t 是文本特征, z_i 是图像特征,如公式(4)、(5)所示:

$$z_i = (x_1, x_2, \cdots, x_m) \tag{4}$$

$$z_t = (y_1, y_2, \cdots, y_n) \tag{5}$$

首先,通过将两种模态的特征进行通道拼接,保留特征中关联性强的特征。然后分别经过MLP层对特征通道进行校正。将文本特征作为辅助特征,用于对图像特征进行投影增强。通过对文本特征 z_t 进行Reshape操作,将其转换为二维的权重矩阵 W ,如公式(6)所示:

$$W = \text{Reshape}(z_t) = \begin{pmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mn} \end{pmatrix} \tag{6}$$

将 W 作为条件权重矩阵,对图像特征进行加权投影。在文本特征权重矩阵中,与目标肿瘤相关的权值大于背景的权重。基于元素特征交叉交互的方式,通过该权重矩阵与图像特征矩阵进行相乘,就能够增强图像特征中与目标肿瘤相关的特征元素,抑制背景特征,得到融合增强特征 \hat{z}_i ,如公式(7)所示:

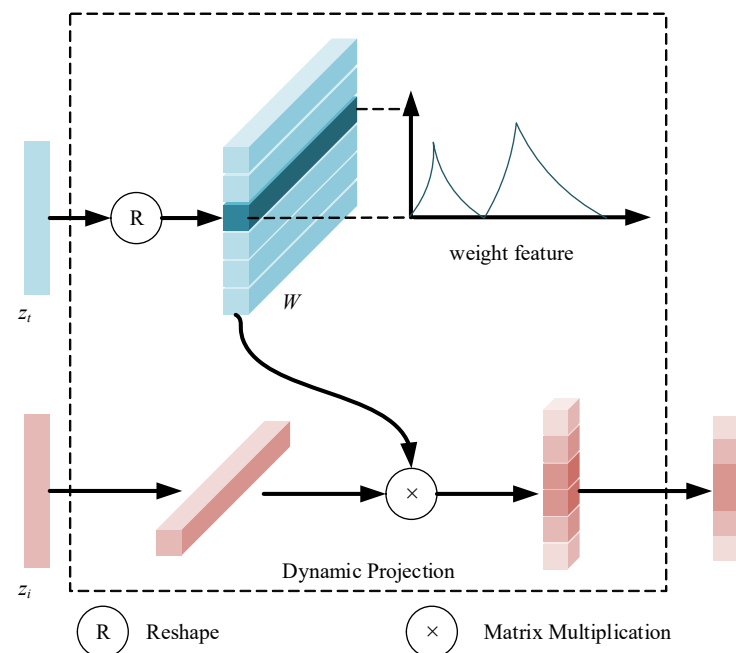


图8 基于动态加权的特征融合
Figure 8 Feature fusion based on dynamic weight

$$\hat{z}_i = D(z_i, W) = (\sum_{i=1}^m x_i \cdot y_{i1}, \dots, \sum_{i=1}^m x_i \cdot y_{in}) \tag{7}$$

在模型训练过程中,通过使用 Dynamic Weight 模块来调节目标特征与背景特征的权重大小,从而调整网络模型参数的更新,就可以对目标特征进行一次增强。在本模型中,重复2次该过程,可以达到动态加强肿瘤特征的效果。

3 实验结果与分析

3.1 实验环境与参数设置

本文实验硬件配置为 Intel i7-9700 CPU, NVIDIA GeForce GTX 3070Ti GPU。软件环境为 Ubuntu18.04 操作系统, Cuda11.3, PyTorch1.9.0, Python3.7。模型训练采用 Adam 优化器进行优化,批次大小设置为 16。初始学习率为 1e-4,学习过程中学习率衰减为 2e-5,一共训练 200 个 epoch,损失函数选用骰子损失和交叉熵损失。

3.2 实验数据与评价标准

本实验所用的数据集收集了上海市第一人民医院 50 名直肠癌患者的 MRI T2 加权图像,数据格式为 NRRD,所有图像均经过脱敏处理,并由专业的影像科医生完成数据标注。每例患者的成像数据包含 40~50 张横截面切片,选取其中包含肿瘤病灶的切片构建数据集,实验数据共包含 500 张直肠癌 MRI 图像,其中 450 张用于模型训练验证,50 张作为测试评估。

本文使用 4 个指标, Dice 相似度系数 (Dice

Similarity Coefficient, DSC) 评价分割精度;准确率 (Accuracy) 衡量分类准确率;精确率 (Precision) 和召回率 (Recall) 评估模型性能,指标定义如下所示:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{11}$$

其中, DSC 指标用于评价模型分割结果和实际肿瘤区域之间的相似程度,其值为 0~1。TP 表示将病灶区域正确识别为肿瘤的像素, TN 表示将背景区域正确识别为背景的像素, FP 表示将正常图像误识别为肿瘤的像素, FN 表示将实际为肿瘤的像素误识别为正常图像的像素。

3.3 实验结果分析

3.3.1 分割结果分析 将本文模型与主流分割模型 U-Net、ResU-Net、Attn U-Net、R2U-Net 进行对比,结果如表 1 所示。

由分割结果可以看出, U-Net 模型由于模型结构简单,在处理复杂的直肠癌 MRI 图像时并不具备良好的泛化性, DSC 为 64.92%, 分割效果最差。模型召回率与精确率均明显低于其他模型,表明 U-Net 无法准确识别肿瘤区域并对其分割。R2U-Net 与 ResU-Net 构建了更为深层的编码器,在分割精度上略有提升。Attn U-Net 加入了注意力机制, DSC 为 73.29%,

表1 不同模型分割结果对比(%)

Table 1 Segmentation performance comparison among various models (%)

| 模型 | DSC | 精确率 | 召回率 |
|------------|-------|-------|-------|
| U-Net | 64.92 | 78.82 | 67.99 |
| ResU-Net | 71.93 | 80.32 | 77.56 |
| R2U-Net | 69.47 | 80.07 | 78.14 |
| Attn U-Net | 73.29 | 82.26 | 81.47 |
| DPU-Net | 82.88 | 87.35 | 85.94 |

这表明注意力机制在学习过程中能够有效增强对目标肿瘤的关注度,提升分割效果。本文模型DPU-Net在所有指标上均取得最优,与U-Net模型相比,分割精度DSC提高了17.96%,取得了较优的分割效果。模型分割可视化结果如图9所示。从图中也可以看出,本文模型分割结果在肿瘤形态、大小等方面均最接近于专家标注金标准。

另外,对本文模型的改进部分进行消融实验,包括仅使用通道注意力模块、仅使用空间注意力模块、

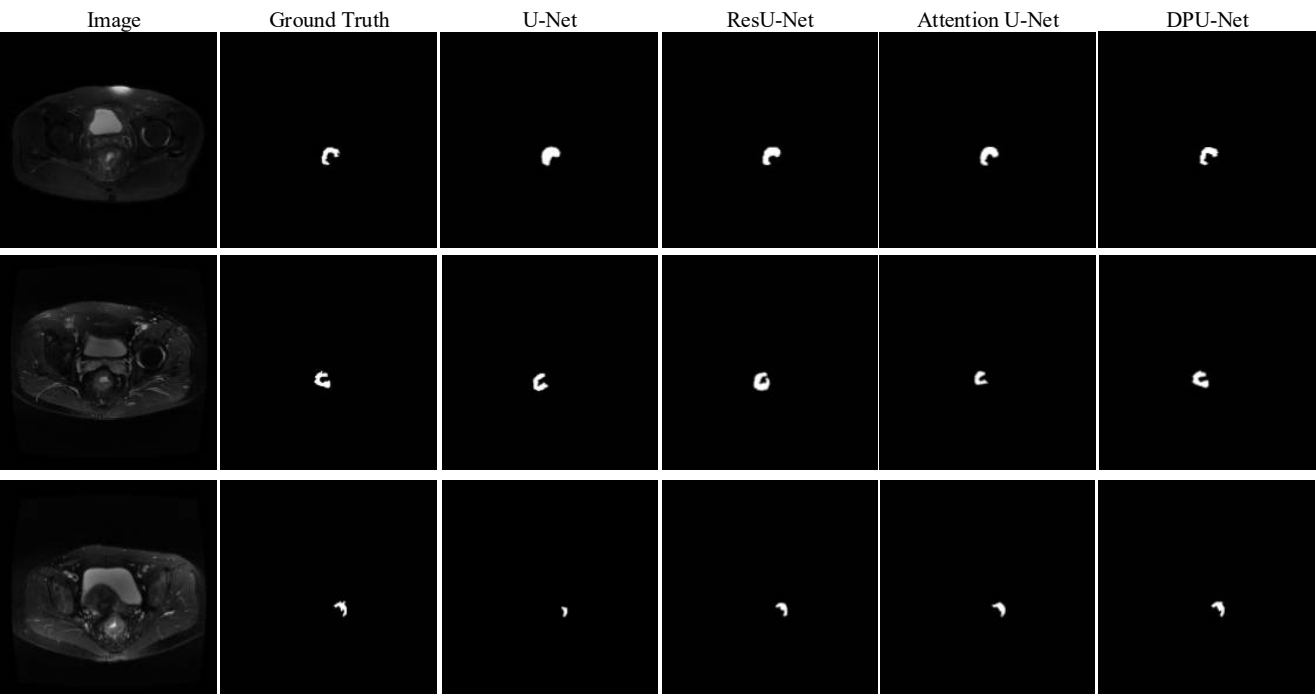


图9 分割结果可视化

Figure 9 Visualization of segmentation results

仅使用MCP模块以及本文所使用的结合空间、通道注意力模块以及MCP模块的方法,从而进一步分析改进模块的有效性以及作用。结果如表2所示。

其中,CA表示通道注意力模块,SA表示空间注意力模块。由结果分析可以得出,3种模块均能提升

分割效果。其中,空间与通道注意力模块在召回率上表现更佳,表明注意力机制能够有效感知肿瘤区域并对其分割。而MCP模块在精确率上表现更加,表明结合多尺度特征能够更精确地分割肿瘤。本文模型结合了空间与通道注意力模块,并结合MCP模块所提取的多尺度特征,在分割精度上取得最优。

3.3.2 T分期结果分析 分期实验中,选取主流的分类模型与本文所提模型进行对比,包括ResNet-18^[20]、VGG-19^[21]、Inception-v3^[22]、DenseNet-121^[23]。结果如表3所示。由结果可以看出,主流分类模型并没有结合分割过程中所得到的丰富的语义特征,分类效果较差。ResNet-18分类准确率仅为61.73%,性能最优的DenseNet-121取得了66.81%的准确率。而本文模型相比于目前最先进的DenseNet-121提升了9.43%,分类准确率达到76.24%。本模型通过5折交叉验证法得到

表2 不同模块的消融实验(%)

Table 2 Ablation experiment of various modules (%)

| 模型 | DSC | 精确率 | 召回率 |
|----------|-------|-------|-------|
| U-Net | 64.92 | 78.82 | 67.99 |
| 骨干网络+CA | 79.81 | 80.83 | 84.71 |
| 骨干网络+SA | 77.57 | 81.56 | 83.46 |
| 骨干网络+MCP | 77.32 | 86.14 | 80.69 |
| 本文模型 | 82.88 | 87.35 | 85.94 |

混淆矩阵如图 10 所示,其中数据被分成 5 个相等的部分。模型被训练 5 次,每次使用其中的 4 个部分作为训练集,剩下的 1 个部分作为验证集。这个过程重复 5 次,确保每个部分都被用作验证集一次。最终,模型的性能由这 5 次验证的性能指标平均值来衡量。在进行 5 折交叉验证时得到 5 个混淆矩阵。这 5 个矩阵代表了在不同验证集上模型的性能表现,用来分析模型在不同类别上的预测准确性,进而评估模型的整体性能。模型验证时,T1 分期与 T2 分期多出现误分类情况,这是因为 T1 与 T2 分期形态上较小,特征不明显,导致分类错误。而 T3 与 T4 分期则因为这两个分期的肿瘤在影像学上均表现为肿瘤侵犯邻近组织。部分 T3 分期肿瘤与 T4 分期肿瘤对膀胱、子宫等器官的侵犯程度十分接

近,导致模型难以对这两个分期准确识别。虽然对不同分期的肿瘤仍旧存在误分类的情况,但相比于其他模型已有明显提升,具备为医生提供参考价值的可行性。

表 3 不同模型分期结果对比(%)
Table 3 Comparison of staging accuracy among various models (%)

| 模型 | 准确率 |
|--------------|-------|
| ResNet-18 | 61.73 |
| VGG-19 | 64.96 |
| Inception-v3 | 63.45 |
| DenseNet-121 | 66.81 |
| 本文模型 | 76.24 |

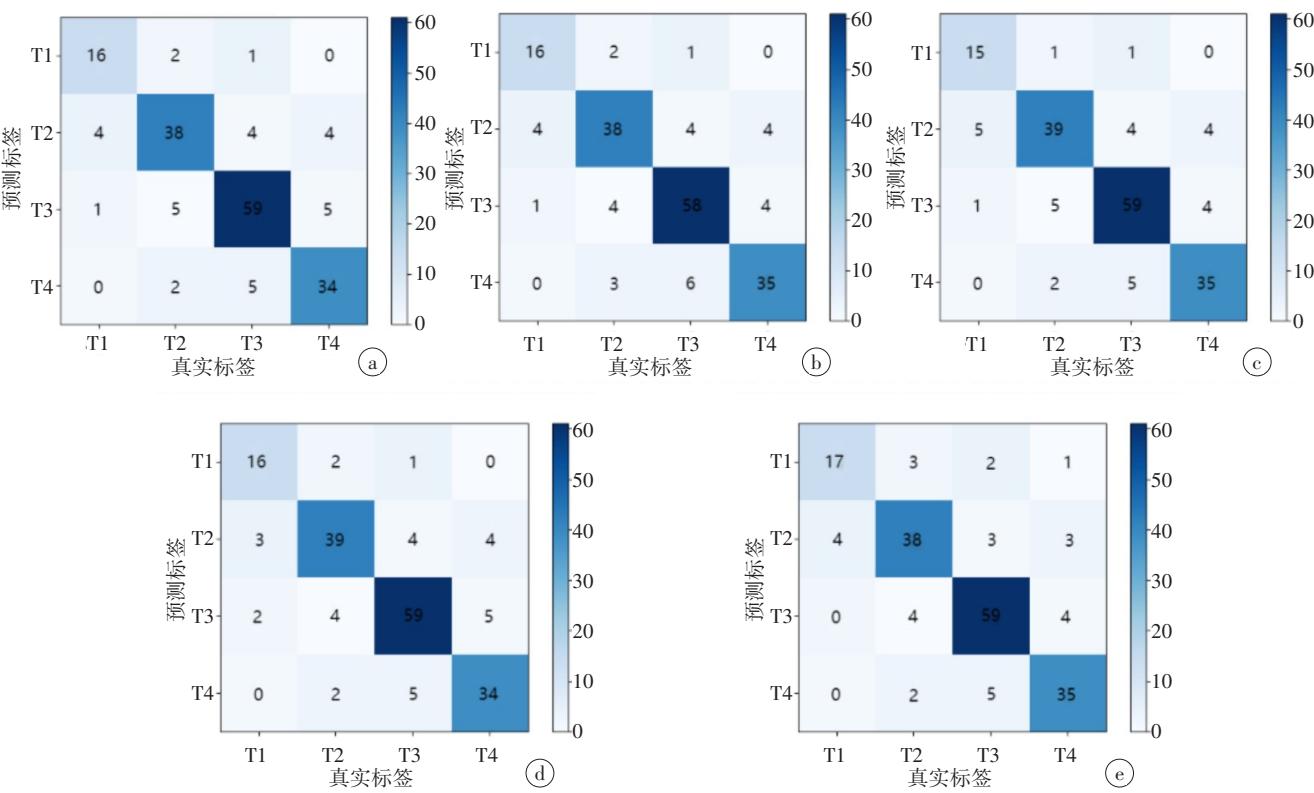


图 10 5 折交叉验证混淆矩阵
Figure 10 Confusion matrix of 5-fold cross-validation
图 a~e 分别代表进行 5 折交叉验证时得到的 5 个混淆矩阵

另外,为验证本文所提出的基于元素交叉交互的动态加权模型对于肿瘤 T 分期识别的有效性,将本文模型与主流多模态融合方式进行结果对比,如表 4 所示。由表 4 结果可以看出,通过结合图像和文本两种模态特征能够提升肿瘤分期效果。通过将特征元素相加(Add)和相乘(Multiply)的方法在结果上十分接近,而通过将特征通道拼接的方式(Concat)相比于前两者具有更好的效果,但提升并不明显。本文所

提的基于特征元素交叉交互的动态加权方式相比于 Concat 方式,在准确率上提升了 3.56%,证明了本文方法的有效性。

4 结 语

本文提出了一种多任务学习模型 DPU-Net,并首次将其应用于直肠癌肿瘤分割与 T 分期,辅助医生进行诊断,为指定医疗方案提供参考。模型针对直肠

表4 不同特征融合方法对比(%)
Table 4 Comparison among different
feature fusion methods (%)

| 方法 | 准确率 |
|----------|-------|
| Add | 71.65 |
| Multiply | 71.94 |
| Concat | 72.68 |
| 本文方法 | 76.24 |

癌MRI图像结构复杂、形态各异的特点,在分割分支中通过结合注意力机制与多尺度特征,提高模型对肿瘤的关注度,加强细节边缘的感知能力,提高肿瘤分割效果。在分类分支中,共享分割信息。引入诊疗文本数据,充分利用医疗资源,构建基于元素交叉交互的动态加权模型,用于图像特征与文本特征的融合,实现肿瘤T分期。实验结果表明,本文模型的分割精度与分类准确率均优于主流模型,有效改善直肠癌肿瘤分割效果不佳、T分期难的问题。未来将根据分割结果进一步拓展至三维重建领域,具备更高的临床价值。

【参考文献】

[1] Siegel RL, Miller KD, Goding SA, et al. Colorectal cancer statistics, 2020[J]. CA Cancer J Clin, 2020, 70(3): 145-164.
[2] Zheng R, Zhang S, Zeng H, et al. Cancer incidence and mortality in China, 2016[J]. J Natl Cancer Cent, 2022, 2(1): 1-9.
[3] Kolligs FT. Diagnostics and epidemiology of colorectal cancer[J]. Visc Med, 2016, 32(3): 158-164.
[4] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2015: 3431-3440.
[5] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-assisted Intervention. Springer, Cham, 2015: 234-241.
[6] Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: learning where to look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.
[7] Alom MZ, Hasan M, Yakopcic C, et al. Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation[J]. arXiv preprint arXiv:1802.06955, 2018.
[8] Xiao X, Lian S, Luo Z, et al. Weighted Res-UNet for high-quality

retina vessel segmentation[C]//2018 9th International Conference On Information Technology in Medicine and Education (ITME). IEEE, 2018: 327-331.
[9] Zheng B, Cai C, Ma L. CT images segmentation method of rectal tumor based on modified U-Net [C]//2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV). IEEE, 2020: 672-677.
[10] Wei B, Nai W, Xing Y, et al. Application of U-Net with variable fractional order gradient descent method in rectal tumor segmentation [C]//2019 6th International Conference on Systems and Informatics (ICSAI). IEEE, 2019: 522-526.
[11] Gao H, Zeng X, Pan D, et al. Covariance self-attention dual path UNet for rectal tumor segmentation [C]//2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021: 11162-11168.
[12] Zheng LB, Zhang XY, Hu JL, et al. Establishment and applicability of a diagnostic system for advanced gastric cancer t staging based on a faster region-based convolutional neural network[J]. Front Oncol, 2020, 10: 1238.
[13] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2017, 542(7369): 115-118.
[14] 韩坤, 潘海为, 张伟, 等. 基于多模态医学图像的 Alzheimer病分类方法[J]. 清华大学学报(自然科学版), 2020, 60(8): 664-671.
Han K, Pan HW, Zhang W, et al. Alzheimer's disease classification method based on multi-modal medical images[J]. Journal of Tsinghua University (Science and Technology), 2020, 60(8): 664-671.
[15] Xiao A, Shen B, Shi X, et al. Intraoperative glioma grading using neural architecture search and multi-modal imaging[J]. IEEE Trans Med Imaging, 2022, 41(10): 2570-2581.
[16] Wang WS, Li XR, Xu ZY, et al. Learning two-stream CNN for multi-modal age-related macular degeneration categorization[J]. IEEE J Biomed Health Inform, 2022, 26(8): 4111-4122.
[17] Liang CS, Huang YQ, He L, et al. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer[J]. Oncotarget, 2016, 7 (21): 31401-31412.
[18] Kim J, Oh JE, Lee J, et al. Rectal cancer: toward fully automatic discrimination of T2 and T3 rectal cancers using deep convolutional neural network[J]. Int J Imaging Syst Technol, 2019, 29(3): 247-259.
[19] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems, 2017: 5998-6008.
[20] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 770-778.
[21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv: 1409.1556, 2014.
[22] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
[23] Zhu Y, Newsam S. Densenet for dense flow [C]//2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 790-794.

(编辑:薛泽玲)