

DOI:10.3969/j.issn.1005-202X.2025.01.005

医学影像物理

基于多尺度卷积与并行反向注意力的医学图像分割

陈梦飞¹, 王尧芬¹, 王海玲¹, 李朋², 宫晓梅²

1. 上海工程技术大学电子电气工程学院, 上海 201620; 2. 同济大学附属上海市肺科医院放疗科, 上海 200433

【摘要】提出一种基于多尺度卷积与并行反向注意力的医学图像分割网络(RPR-MLP)。在编码器阶段,采用Res2net模块和标记化多层感知机模块作为骨干结构,以提取多尺度信息并增强语义特征的多样性。与此同时,通过并行的部分解码器提高解码器中提取语义信息的准确性。此外,反向注意力模块再次强调对重要区域的关注,进一步提高分割结果的精确性。本文提出的网络在Kvasir和ISIC 2018两个公共数据集上的Dice相似系数(DSC)分别为0.8967、0.8762,证明本文网络对医学图像分割的有效性,同时具有较强的泛化能力。将该方法应用于肺肿瘤CT图像LungCancer数据集,评价指标DSC、IoU和F1分别为0.7278、58.83%和67.85%,其结果与基准网络UNeXt和普通CNN网络U-Net、AttU-Net、U-Net++、PraNet相比,DSC、IoU和F1提升幅度分别为0.0301~0.0578、3.16%~4.70%和6.72%~18.53%,结果表明本文提出的网络性能明显优于对比方法。本研究证明RPR-MLP在不同数据集上的有效性和泛化能力,为肺肿瘤图像分割提供重要的技术支持。

【关键词】医学图像分割;多尺度卷积;多层感知机;部分解码器;反向注意力模块

【中图分类号】R318;TP391

【文献标志码】A

【文章编号】1005-202X(2025)01-0027-10

Medical image segmentation based on multi-scale convolution and parallel reverse attention

CHEN Mengfei¹, WANG Raofen¹, WANG Hailing¹, LI Peng², GONG Xiaomei²

1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China; 2. Department of Radiotherapy, Shanghai Pulmonary Hospital Affiliated to Tongji University, Shanghai 200433, China

Abstract: A medical image segmentation network (RPR-MLP) based on multi-scale convolution and parallel reverse attention is presented. In the encoder, Res2net modules and tokenized multi-layer perceptron modules are used as the backbone structure to extract multi-scale information and enhance the diversity of semantic features. Meanwhile, the accuracy of semantic information extraction in the decoder is improved through parallel partial decoder. Additionally, reverse attention module re-emphasizes the focus on important regions for further improving the accuracy of segmentation results. The proposed method achieves Dice scores of 0.8967 and 0.8762 on the Kvasir and ISIC 2018 public datasets, respectively, demonstrating its effectiveness and generalization ability in medical image segmentation. Furthermore, when applied to the lung tumor CT image dataset (LungCancer dataset) collected in the study, the proposed network has Dice score, IoU and F1 score of 0.7278, 58.83% and 67.85%, respectively, outperforming baseline network (UNeXt) and common CNN (U-Net, AttU-Net, U-Net++ and PraNet) by 0.0301-0.0578, 3.16%-4.70%, and 6.72%-18.53%, respectively. The study confirms the effectiveness and generalization ability of RPR-MLP network on different datasets, providing important technical support for lung tumor image segmentation.

Keywords: medical image segmentation; multi-scale convolution; multi-layer perceptron; partial decoder; reverse attention module

【收稿日期】2024-08-21

【基金项目】国家自然科学基金(61803255, 62001284);上海市科委科技创新行动计划(20Y11913600);申康三年行动计划肺科培育项目(SKPY2021006)

【作者简介】陈梦飞, 硕士研究生, 研究方向: 医学图像处理, E-mail: 841239459@qq.com

【通信作者】王尧芬, 博士, 副教授, 研究方向: 医学图像处理、脑机接口, E-mail: rfwangsues@163.com

前言

医学图像分割是医学影像分析中的关键任务之一,其在癌症诊断、治疗规划和疾病监测等方面具有重要应用。肺癌作为全球范围内最常见的癌症之一,早期诊断和准确的肿瘤分割对于治疗和提高生存率至关重要^[1]。根据世界卫生组织(WHO)数据表明,每年有数百万人被诊断患有肺癌,并因此疾病不治而亡。肺癌通常需要根据其扩散程度进行分期,

以确定最佳治疗方案,分期包括4个阶段,从I期(局部肿瘤)到IV期(扩散到其他部位)。通常使用计算机断层扫描(CT)^[2]、X射线^[3]和正电子发射断层扫描^[4]等成像技术对肺癌进行诊断和分期。CT扫描是肺癌诊断的关键工具之一,可以提供有关肿瘤位置、大小和形状的详细信息,有助于医生进行初步评估和规划治疗^[5]。然而,手动分割肺肿瘤CT图像是一项耗时且主观的任务,图像分割方法对肺癌进行准确识别、详细分析、合理诊断和预测具有重要意义。由于成像机器对医学图像呈现的形状、大小和位置的影响,使得图像背景和目标灰度差异不显著,病变区域边界不明显,更重要的是纹理变化复杂,分割精度难以控制。传统分割技术如阈值法、区域生长法、边缘检测法等精度、效率和鲁棒性都较差。随着深度学习的快速发展,卷积神经网络(Convolutional Neural Networks, CNN)的引入为医学图像的分割领域带来革命性的进步^[6]。CNN在众多分割任务中表现出色,其中U-Net作为经典的架构之一,通过编码器和解码器间的跳跃连接,有效地捕获图像的空间特征,展现出卓越的性能。作为基准网络,U-Net^[7]已被广泛应用于医学图像分割任务中,如U-Net++^[8]、U-Net3+^[9]、3D U-Net^[10]、V-Net^[11]和KiU-Net^[12]等网络都以U-Net为基础进行改进。然而,对于肺肿瘤CT图像这样复杂的数据以及噪声的影响,U-Net类网络在边界和细节部分的分割结果常存在一定的不准确,尤其是对于小型肿瘤的检测和分割效果不显著。此外,在处理复杂的图像分割任务时,高级网络产生的特征图分辨率较低,包含的细节信息较少。与此相反,低级网络生成的特征图具有较多噪声,并且其感知范围有限,无法充分捕获语义信息。目前有效的解决方法是将这些不同层次信息进行融合,这种多尺度特征融合的方法有助于更精确地识别分割目标,从而提高分割模型性能。为了实现对小病变和多病变区域的准确分割,文献[13-14]在CNN模型中引入低级特征和高级语义信息相融合的策略。但是随着网络加深,深层网络的特征与浅层网络的特征发生较大的偏移,直接的特征融合会导致语义冲突,这些冲突抑制细节信息的学习,不利于多尺度特征上下文信息的建立,对分割结果有着负面影响。同时,CNN网络需要大量计算资源来进行训练和预测,深层网络 and 大型数据集使得模型计算需求巨大^[15]。近年来,基于多层感知机(Multi-Layer Perceptron, MLP)的MLP-Mixer^[16]、gMLP^[17]和AS-MLP^[18]等新型网络架构被相继提出。MLP-Mixer使用MLP代替CNN的卷积运算,分别通过交叉位置和每个位置操作在标记之间构建上下文

和通道间的相关性。gMLP具有比MLP-Mixer更好的性能,参数更少,门控机制更简单。AS-MLP是第一个迁移到下游任务的MLP架构,通过平行轴向移位操作将不同特征对准同一通道,沿通道位置的全连接实现十字形视场。上述网络虽在参数效率和计算复杂度方面做出一定程度的优化,但仍然面临着平衡模型性能和计算冗余的问题。

基于上述分析,本文提出一种基于多尺度卷积与并行反向注意力的医学图像分割网络(RPR-MLP),采用Res2net模块^[19]和标记化MLP模块^[20]作为骨干结构,引入编-解码器和跳跃连接的多尺度特征融合U型架构。该网络包括两个重要的模块:并行部分解码器(Parallel Partial Decoder, PPD)^[21-23]和反向注意力模块^[22-23],将不同层特征更好地融合,并提高计算效率。具体方法如下:首先,针对深层和浅层特征偏移的问题,本文方法在编码器阶段利用Res2net模块和标记化MLP模块对图像特征进行提取,这些模块有助于保留来自不同空间尺度的信息,并提供多层次的语义特征;同时,在PPD中,舍弃较低级别的图像特征以提高计算速度,然后将较高级别的图像特征进行融合,以增强语义信息的精确性,这种策略可以有效地调整多尺度特征融合中的语义冲突,并提取出完整的全局上下文信息;为了进一步加强分割网络对重要区域的关注,使用反向注意力模块,在跳跃连接中挖掘较高级别图像特征的边界信息,提高分割网络的精确性和泛化能力;最后,该网络在Kvasir、ISIC 2018两个公共数据集和同济大学附属上海市肺科医院提供的肺肿瘤CT图像数据集上进行实验,证明RPR-MLP在医学图像分割上的有效性和较强的泛化能力,在肺肿瘤图像分割任务中有着优秀的性能表现。

1 方法

图1为RPR-MLP网络结构图,该网络由4个部分组成:编码器部分、颈部(PPD模块)、解码器部分、跳跃连接。在编码器部分,通过Res2net模块和标记化MLP模块逐步提取低级特征和具有出色局部的高级特征。5个层次的特征从上到下分为 $\{f_i, i = 1, 2, \dots, 5\}$,其中 $\{f_i, i = 1, 2\}$ 特征为低级特征, $\{f_i, i = 3, 4, 5\}$ 特征为高级特征。网络颈部采用一个全新的先进译码器组件—PPD,通过并行连接来聚合高级特征。在这一部分中,可以生成包含丰富细节信息和高级语义信息特征,并作为上采样的输入。在高级特征的跳跃连接上引入反向注意力模块,通过从输出的特征中消除已经预测的肺肿瘤区域,依

次挖掘互补的区域和细节。这些现有区域是从更深层次的特征图中进行上采样得到的。在解码器的阶段中,对生成的多尺度特征进行双线性插值上采样,

并通过标记化MLP模块和普通卷积得到最终的预测图。网络中 C_1 、 C_2 、 C_3 、 C_4 和 C_5 分别为 32、64、128、160 和 256。

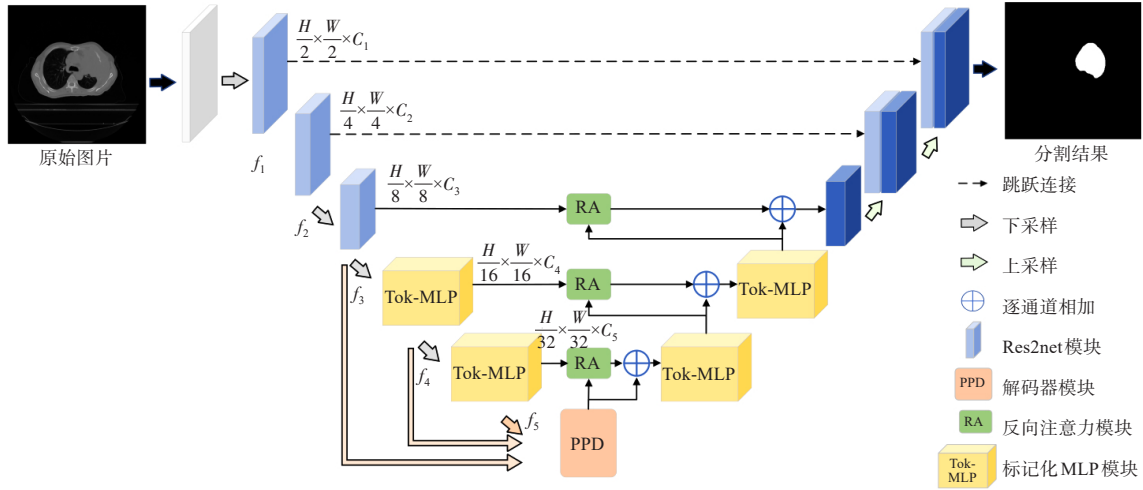


图1 RPR-MLP网络结构图

Figure 1 RPR-MLP network structure diagram

1.1 Res2net模块

编码器部分采用多尺度卷积模块 Res2net 作为网络的特征提取模块,以增强图像的多尺度特征。Res2net 模块结构如图 2 所示,在 Res2net 模块中,通过构建分层的残差连接来处理单个残差块内的特征,从而实现多尺度特征的处理,有助于提取全局和局部信息。为了更好地融合不同尺度的信息,将所有特征信息连接起来,并通过 1×1 卷积传递。Res2net 模块处理特征图过程中,首先经过 1×1 卷积后,将得到的特征图平均分成 s 个特征子集,记为 x_i ,其中 $i \in \{1, 2, \dots, s\}$,每个特征子集 x_i 与输入特征映射相比具有相同空间大小,且通道数为 $1/s$ 。除 x_1 外,每个 x_i 都有一个对应的 3×3 卷积,记为 $K_i(\cdot)$,卷积操作后得到的特征图用 y_i 表示,并连接到下一个子集中进行以上操作。将得到的所有特征图按通道方向进行相加,再使用一个 1×1 卷积融合全部信息,并与原始输入特征连接输出最终的特征图。因此, y_i 如下所示:

$$y_i = \begin{cases} x_i, & i = 1 \\ K_i(x_i), & i = 2 \\ K_i(x_i + y_{i-1}), & 2 < i \leq s \end{cases} \quad (1)$$

1.2 标记化MLP模块

在标记化MLP模块中,将卷积特征映射为标记,然后利用MLP学习特定的特征信息。图3是标记化MLP模块结构示意图,操作过程中首先使用卷积核

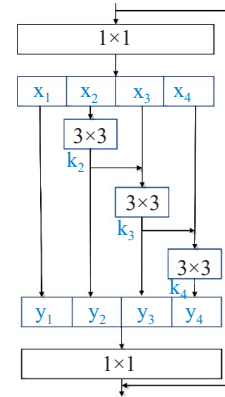


图2 Res2net 模块示意图

Figure 2 Schematic diagram of Res2net module

大小为 3 的卷积将通道维数投影映射为 E 。然后,将特征传递给跨宽度移位 MLP 模块,其中 MLP 的隐藏维度是超参数 W ,其操作与 Swin-Unet^[24]中使用的移动窗口注意力相似。在进行标记化之前,先调整卷积特征图的通道顺序,这有助于 MLP 更专注于特定的位置。首先把特征图划分成 h 个区域,并利用跨宽度移位的 MLP 在特定轴上移动至指定位置。这种移位操作能够更深入地关注卷积特征的局部位置,从而学习更为精细的局部特征信息。经过跨宽度移位 MLP 后,将学习到的特征信息通过深度卷积 (Depthwise Convolution, DWConv) 传递,进一步提取

和加工特征信息,以增强网络的表示能力和学习能力。接下来,特征会经过高斯误差线性单元(Gaussian Error Linear Unit, GELU)函数,随后通过跨高度移位的MLP对特征图进行处理。该MLP首先在特征图的 h 个区域水平方向上移动,然后沿垂直方向移动到特定位置,以实现跨高度处理特征图。

通过这个过程,特征的维度从 W 转换为 H 。然后,利用残差连接将原始标记的特征图与处理后的特征图相结合,以增强特征的表示能力。最后,应用层归一化对输出特征进行标准化,并将其传递给下一个模块。这一系列操作有助于模型更好地捕捉图像的局部和全局特征,提升分割性能^[25-30]。

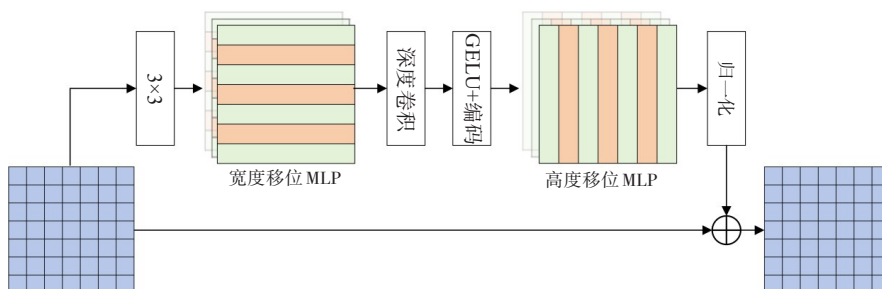


图3 标记化MLP模块示意图

Figure 3 Schematic diagram of tokenized MLP module

在标记化MLP模块中,对于输入特征 T ,将通道划分为 h 个部分,然后沿 W 维轴向移动并进行标记化,得到 T_w 如下所示:

$$T_w = \text{Tokenized}(\text{shift}_w(\rho(T))) \quad (2)$$

其中, ρ 和 shift_w 分别表示沿通道维度的分割和沿 W 维度的移位。 T_w 沿着通道维度执行MLP,将通道数量映射到 h 个维度,然后通过 3×3 深度卷积层和GLEU层^[29]得到 T_1 ,如下所示:

$$T_1 = \text{GLEU}(\text{DWConv}(\text{MLP}(T_w))) \quad (3)$$

其中,DWConv表示 3×3 深度卷积。同样地, T_1 沿着 H 维移动得到 T_H 。 T_H 映射到原始输入特征维度后,将残差连接得到模块输出。该模块通过生成随机窗口,提取特定的局部特征,如下所示:

$$T_H = \text{MLP}(\text{Tokenized}(\text{shift}_H(\rho(T)))) \quad (4)$$

$$\text{output} = T \oplus \text{FC}(T_H) \quad (5)$$

其中,FC和 shift_H 分别表示全连接层和沿 H 维方向移位。 \oplus 表示各元素相加。

1.3 PPD

图4为采用的PPD,将3层高级特征 $\{f_i, i = 3, 4, 5\}$ 分别经过3个并行的感受野模块(Receptive Field Block, RFB)^[22, 26]、上采样和 3×3 卷积等操作,将高级特征 $\{f_i, i = 3, 4, 5\}$ 进一步融合,生成丰富细节信息和高级语义信息的特征。图5所示为RFB模块结构示意图,该模块有4个分支 $\{b_m, m = 1, \dots, 4\}$,为了加快每个分支计算的速度,使用 1×1 卷积层先将通道数减少到16,对于

$\{b_m, m > 1\}$ 条分支,分别增加 $1 \times (2m-1)$ 、 $(2m-1) \times 1$ 卷积层和扩张率为 $(2m-1)$ 的 3×3 扩张卷积层^[28]。将这些分支的输出连接起来,并通过 3×3 和 1×1 卷积层将通道减少到16。然后添加一个短链接将原始的特征图作为残差。

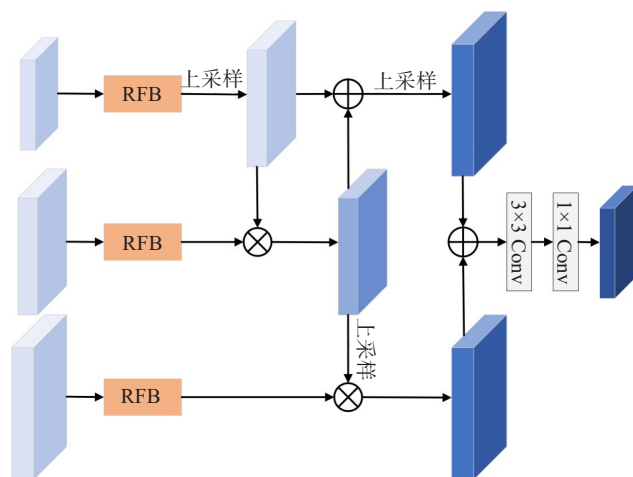


图4 并行部分解码器示意图

Figure 4 Schematic diagram of parallel partial decoder

PPD中,RFB模块输出的特征为 $\{f_i^c, i \in [l, \dots, L], c \in [a, d]\}$,最上面的分支输出的特征为 $\{f_i^{c_1}\}$ 。对于 $(i = L)$ 的特征,设 $f_L^{c_2} = f_L^{c_1}$,对于特征 $\{f_i^{c_1}, i < L\}$,其通过元素的方式将自身与所有更深层的特征相乘输出为 $f_i^{c_2}$,定义如下所示:

$$f_i^{c_2} = f_i^{c_1} \odot \prod_{k=i+1}^L \text{Conv}(\text{Up}(f_k^{c_1})), i \in [L, \dots, L-1] \quad (6)$$

其中, $\text{Up}(\cdot)$ 为上采样特征, 上采样因子为 2^{k-j} , Conv 为 3×3 卷积层。最后利用上采样串联策略来整合多层次特征。

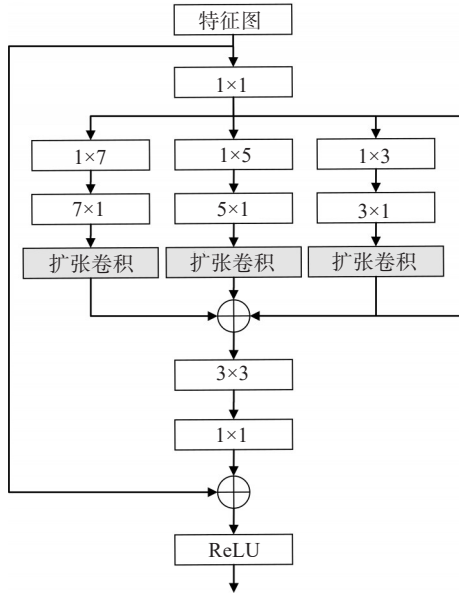


图5 RFB模块结构示意图

Figure 5 RFB module structure diagram

1.4 反向注意力模块

解码器部分的输入来自 PPD 模块输出的特征图, 它只能捕捉肺肿瘤区域的相对粗略的位置, 而没有边界细节。本文引入的反向注意力模块在挖掘互补区域和细节方面发挥关键作用, 它通过擦除高级侧输出特征中当前预测的肺肿瘤显著区域, 从而指导网络更加深入地探索图像中的其他信息。这里的当前预测是通过从更深层次向上采样获得的。这种从高级特征到低级特征的擦除方式最终可以将低分辨率的特征图细化为具有探索区域和细节的完整高分辨率特征图。

将高级特征侧输出特征 $\{f_i, i = 3, 4, 5\}$ 乘以反向注意力权重 A_i 来获得输出的反向注意力特征 R_i :

$$R_i = f_i \odot A_i \quad (7)$$

其中, A_i 为反向注意力权重, R_i 为反向注意力特征, \odot 为特征图元素相乘。反向注意力权重 A_i 计算公式如下所示:

$$A_i = 1 - \text{Sigmoid}(\text{Up}(S_{i+1})) \quad (8)$$

其中, $\text{Up}(\cdot)$ 为上采样操作, $\text{Sigmoid}(\cdot)$ 是 Sigmoid 函数。图 6 展示了反向注意力操作过程细节, 反向注意力驱动的擦除机制能有效地将初始的粗略估计细化为精确且完整的预测图。

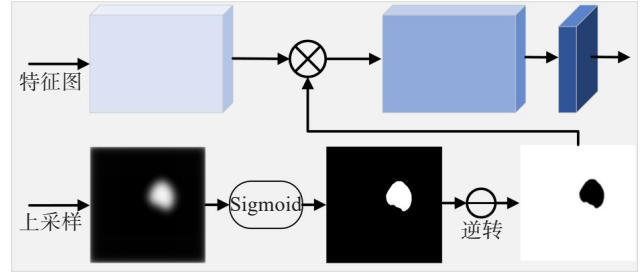


图6 反向注意力模块示意图

Figure 6 Schematic diagram of reverse attention module

1.5 损失函数

损失函数可以衡量标签和真实值的差距, 本文使用二元交叉熵 (Binary Cross Entropy, BCE) 损失函数和 Dice 损失函数相结合的混合损失函数来训练模型。预测值 \hat{Y} 与标签值 Y 之间的损失 L 表示如下所示:

$$L = 0.5\text{BCE}(\hat{Y}, Y) + \text{Dice}(\hat{Y}, Y) \quad (9)$$

BCE 损失函数和 Dice 损失函数是医学图像分割最常用的损失函数。BCE 损失函数特别适用于处理二分类问题, 它在图像分割任务中能够有效地计算预测值与标签值之间的差异, 促使模型更好地学习到图像的整体特征和结构; Dice 损失函数能够衡量预测分割结果与真实分割结果之间的相似程度, 其目标是 minimized 预测分割结果和真实分割结果之间的评价指标 Dice 相似系数的补集 (即 $1 - \text{Dice}$), 从而使 Dice 相似系数最大化, 尤其适用于处理不平衡类别和边缘细节。通过将 BCE 损失函数和 Dice 损失函数结合起来, 可以综合考虑预测结果的整体特征和局部细节, 提高模型在图像分割任务中的性能和泛化能力。因此, 将它们的权重分别设置为 0.5 和 1.0, 可以有效地平衡两种损失函数的影响, 提高模型的训练效果。Dice 损失函数的定义为:

$$\text{Dice}(\hat{Y}, Y) = 1 - \frac{2 \sum_{i=1}^n Y \hat{Y}}{\sum_{i=1}^n Y + \sum_{i=1}^n \hat{Y}} \quad (10)$$

BCE 损失函数定义为:

$$\text{BCE}(\hat{Y}, Y) = - \sum_{i=1}^n [Y \log(\hat{Y}) + (1 - Y) \log(1 - \hat{Y})] \quad (11)$$

Loss 的最小化过程实际上是在推动网络学习提高评价指标 Dice 系数、IoU 和 F1 系数, 即提高分割结果的准确度和重叠程度, 从而达到更好的图像分割效果。Loss 值的结果越小, 说明真实值和预测值重合度越高, 表明分割结果越好。

2 结果与讨论

2.1 数据集

为了验证本文网络的有效性、泛化能力和对肺肿瘤图像分割的性能表现,本实验使用两个公共数据集 ISIC 2018、Kvasir 和 1 个私有数据集非小细胞肺癌 CT 图像(LungCancer)。ISIC 2018 数据集^[27]:该数据集包含 2 594 张用于训练的彩色皮肤镜图像、100 张验证图像和 1 000 张测试图像以及相应的真实标签,将所有图像大小调整为 512×512。Kvasir 数据集^[28]:该数据集包含胃肠道息肉图像和真实标签,像素大小统一调整为 512×512,共 1 000 张训练图像和 196 张测试图像,将图像按 8:2 的比例随机划分为训练集和验证集。LungCancer 数据集:该数据集由同济大学附属上海市肺科医院提供,医院伦理委员会审批号为 L20-316Z,共包括 87 例患者的 1 106 张非小细胞肺癌 CT 图像,训练集、验证集和测试集样本数量分别划分为 711、196 和 199 张,图像分辨率均为 512×512。为保证肿瘤区域的标注质量,该数据集的标签均由经验丰富的医生手动标注。由于训练集样本数量较少,为防止过拟合,分别对训练集和验证集的每张图像及对应的标签进行两次随机旋转,来扩大学习样本数量。然而在测试阶段,希望评估模型在真实环境下的性能,因此不对测试数据进行任何改变,以确保评估结果的准确性和可靠性。最终得到的训练集、验证集和测试集样本数量分别为 2 133、588 和 199 张。

2.2 实验环境

本实验使用深度学习框架 PyTorch 2.0.1、Python 3.11、CUDA 11.8 和 NVIDIA RTX 4070Ti GPU 来训练模型。在训练过程中,参数设定 Batch size 为 20 以及设置 150 个 Epoch。选择 Adam 作为网络优化器,并设置初始学习率为 0.001。此外,训练过程中采用余弦退火学习率调度,确保学习率能够平稳地降至最小值 0.00001。

2.3 评价指标

本文选用 Dice 相似系数(DSC)、IoU 和 F1-Score 3 个常用评价指标来评估模型性能,分别如式(12)~(16)所示:

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (12)$$

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

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

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (16)$$

其中,TP 表示正确预测肿瘤的像素个数,FP 表示预测背景为肿瘤的像素个数,FN 表示预测肿瘤为背景的像素个数。DSC 是衡量预测结果和标签的相似度。IoU 是两组真实值和预测值的交集和并集之比。查准率(Precision)是正确预测肿瘤像素与总预测肿瘤像素之比,反映肿瘤分割结果的可靠性。查全率(Recall)是正确预测的肿瘤像素与真实肿瘤像素比值,反映肿瘤分割的完整程度。F1 定义为查全率和查准率调和平均值,取值为 0~1。参数量(Params)和计算量(FLOPs)用于评估网络复杂度和计算量。

2.4 结果

为了验证本文提出的 RPR-MLP 的有效性和泛化能力,首先在 Kvasir 和 ISIC 2018 两个公共数据集上进行测试,然后将网络应用于非小细胞肺癌 CT 图像的分割,实现对肺肿瘤区域的精确分割。使用最经典的医学分割网络与本文提出的 RPR-MLP 进行比较,对比网络包括广泛使用的 CNN 网络 U-Net^[8]、AttU-Net^[25]、U-Net++^[9]、基于并行反向注意力的 PraNet 网络^[24]和基于 MLP 的 UNeXt 网络^[21]。不仅对各个模型输出的 DSC、IoU、F1 进行比较,还对每个模型的参数量(Params)和计算量(FLOPs)进行比较。

2.4.1 模型性能测试 表 1 为 RPR-MLP 和对比网络在两个公共数据集上 DSC、IoU 和 F1 评价指标的实验结果,依据表 1 分析不同网络在两个公共数据集上的定量结果对比。(1)Kvasir 数据集:CNN 网络 U-Net 以及改进网络 AttU-Net 和 U-Net++ 对 Kvasir 数据集的分割结果 DSC 分别为 0.822 9、0.822 3、0.821 0,总体结果较差。基于 MLP 的 UNeXt 网络,由于 MLP 能够更好地专注于学习局部依赖关系,因此该网络对 Kvasir 数据集的分割结果比 U-Net、AttU-Net 和 U-Net++ 高。然而,PraNet 是针对息肉数据集设计的分割网络,其网络结构更适合于息肉图像的特点,在 Kvasir 数据集上的分割结果较好。而本文提出的 RPR-MLP 在 Kvasir 数据集上 DSC 和 F1 结果与 PraNet 相差不到 0.01,同时,RPR-MLP 网络在 DSC、IoU 和 F1 指标上分别比 U-Net、AttU-Net 和 U-Net++ 提升幅度为 0.073 8~0.075 7、10.90%~12.25% 和 7.24%~11.64%,而且比基准模型 UNeXt 网络的 DSC、IoU 和 F1 分别提高 0.016 3、2.62% 和 4.73%。结果表明,RPR-MLP 对医学图像分割有效。(2)ISIC 2018 数据集:RPR-MLP 在 DSC、IoU 和 F1 指标上均取得最好的结果,分别为 0.876 2、79.53% 和 88.26%,且远高于 U-Net、AttU-Net、U-Net++、UNeXt 和 PraNet 网络。RPR-MLP 通过反向注意力模块和多尺度特征融合的

特点,能够取得较好的综合性能表现。该实验不仅验证 RPR-MLP 对医学图像分割有效,同时证明 RPR-MLP 分割性能的优越性以及较强的泛化能力。

表1 RPR-MLP 和对比网络在两个公共数据集上的实验结果
Table 1 Experimental results of RPR-MLP network and other networks on two public datasets

数据集	网络	DSC	IoU/%	F1/%
Kvasir	U-Net	0.822 9	70.61	83.13
	AttU-Net	0.822 3	69.73	79.29
	U-Net++	0.821 0	69.26	78.73
	PraNet	0.898 0	84.00	91.09
	UNeXt	0.880 4	78.89	85.64
	RPR-MLP	0.896 7	81.51	90.37
ISIC 2018	U-Net	0.832 6	72.41	83.92
	AttU-Net	0.842 0	73.84	85.19
	U-Net++	0.848 0	73.97	83.04
	PraNet	0.846 0	74.98	83.01
	UNeXt	0.864 6	76.64	87.45
	RPR-MLP	0.876 2	79.53	88.26

为了更直观地研究 RPR-MLP 网络的性能优势,表2为各模型参数量大小和计算量结果对比。由表2可以看出,RPR-MLP 的参数量只有 3.08 M,远小于 U-Net++网络(47.18 M),虽然参数量和计算量比基准网络 UNeXt 略大,但 RPR-MLP 综合性能比基准网络更优异。

表2 RPR-MLP 和对比网络的参数量和计算量结果对比
Table 2 Comparison of quantity of parameters and calculation results between RPR-MLP network and other networks

网络	Params/M	FLOPs/G
U-Net	31.04	218.95
AttU-Net	34.88	226.52
U-Net++	47.18	798.57
PraNet	30.50	27.82
UNeXt	1.47	2.29
RPR-MLP	3.08	3.24

图7展示所有网络在 Kvasir 和 ISIC 2018 两个公共数据集上具有代表性的示例。第一行是输入图像,后几行分别是 U-Net、AttU-Net、U-Net++、PraNet、

UNeXt 和 RPR-MLP 网络的分割结果,红线是真实标签,蓝线是网络分割结果。结果表明,U-Net 和 AttU-Net 以及 U-Net++在两个公共数据集上都不能准确分割图像的位置和边界,而 PraNet 在使用注意力模块的情况下,虽比 U-Net、AttU-Net 和 U-Net++的分割相对精准,但分割结果的边界比较模糊,基准网络 UNeXt 在两个公共数据集上对分割结果的位置和边界都有缺陷。而本文提出的 RPR-MLP 在位置和边界方面都最接近真实标签,与其他网络相比,RPR-MLP 网络在两个数据集上分割的边界都更清晰、更准确。

根据以上实验结果分析,本文提出的 RPR-MLP 在 Kvasir 数据集和 ISIC 2018 数据集上均表现出显著的优势。在 Kvasir 数据集上,相比传统卷积网络和改进的网络,RPR-MLP 能够更好地捕获模糊边界特征,并取得与专为息肉数据集设计的 PraNet 相媲美的分割结果,而且网络结构比 PraNet 更轻量化。在 ISIC 2018 数据集上,RPR-MLP 在 DSC、IoU 和 F1 指标上都取得最佳结果,均高于其他网络。以上结果表明 RPR-MLP 在医学图像分割任务中具有显著的性能优势。

2.4.2 肺肿瘤 CT 图像分割结果及分析 在 LungCancer 数据集上,RPR-MLP 的分割效果明显优于对比网络,如表3所示。RPR-MLP 在 DSC、IoU 和 F1 指标上分别达到 0.727 8、58.83% 和 67.85%,相较于其他模型,其分割结果均有所提升,DSC、IoU 和 F1 提升幅度分别为 0.030 1~0.057 8、3.16%~4.70% 和 6.72%~18.53%。图8呈现了在 LungCancer 数据集上的典型示例,展示各个网络对肺肿瘤区域的分割效果,图片右下角为绿框区域局部放大图。通过观察示例图像,可清晰地发现 RPR-MLP 在分割肺肿瘤区域上具有出色的准确性和精确度。RPR-MLP 能够有效地定位和分割肿瘤区域,准确地勾勒出肺部组织和肿瘤区域之间的边界,呈现出更为清晰和精细的分割结果。结果表明,RPR-MLP 在处理肺肿瘤 CT 图像分割任务中具有显著优势。通过利用反向注意力模块和多尺度特征融合策略,RPR-MLP 能更好地捕获肺肿瘤图像的重要特征,并在分割过程中更准确地识别肿瘤区域。这一优势体现在定量指标的提升,并在分割图像的定性分析中得到有效的验证,为进一步提高肺肿瘤图像分割的准确性和可靠性提供有力的技术支持。

2.4.3 消融实验 为了验证 RPR-MLP 中各模块对整体网络的贡献,本节将进行消融实验。实验是在 LungCancer 数据集上进行的,并将它们与基准模型 UNeXt 进行比较,包括:(1) UNeXt: 基准网络;(2)

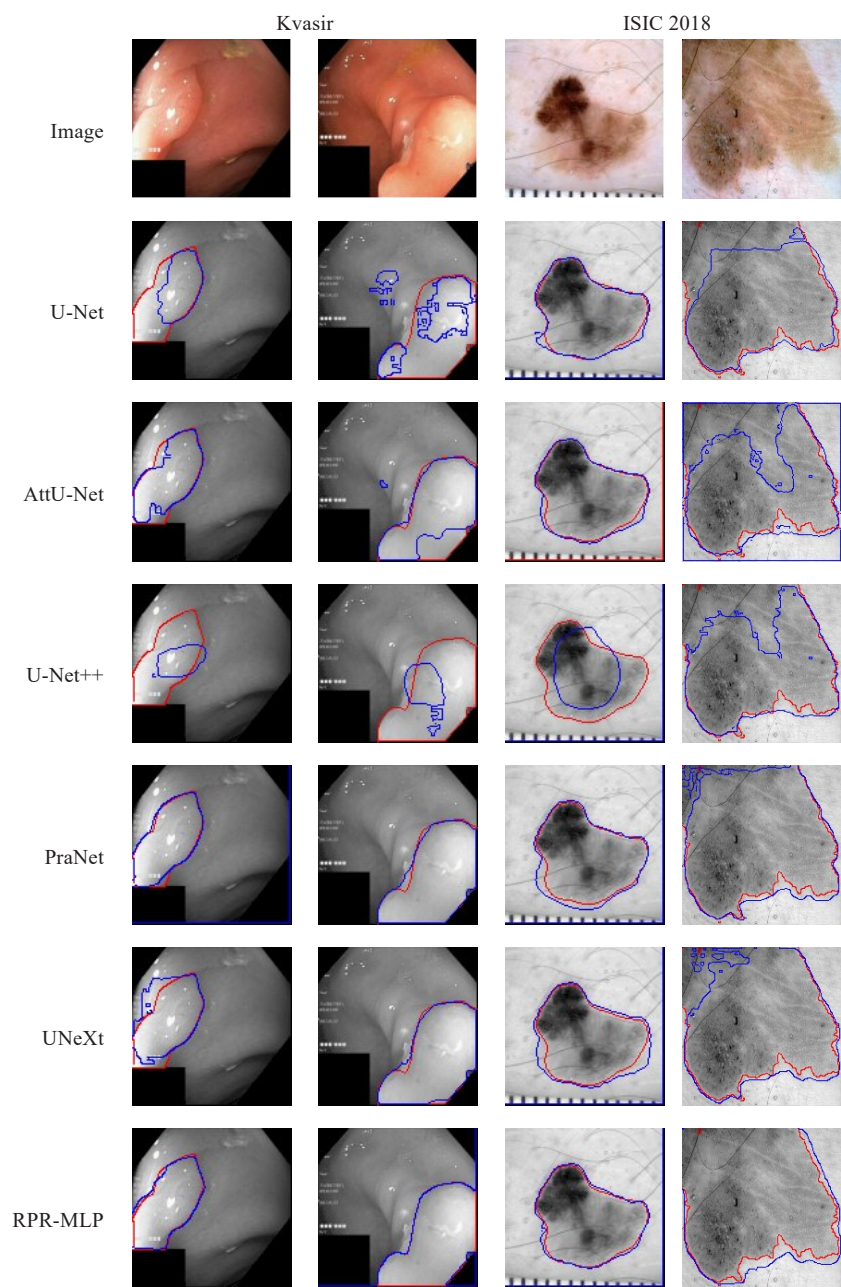


图7 RPR-MLP 和对比网络在两个公共数据集上定性结果比较

Figure 7 Comparison of qualitative results between RPR-MLP network and other networks on two public datasets

表3 RPR-MLP和对比网络在 LungCancer 数据集上的实验结果

Table 3 Experimental results of RPR-MLP network and other networks on the LungCancer dataset

网络	DSC	IoU/%	F1/%
U-Net	0.695 6	55.21	52.90
AttU-Net	0.696 4	55.50	50.68
U-Net++	0.684 3	54.13	50.17
PraNet	0.670 0	55.67	49.32
UNeXt	0.697 7	54.77	61.13
RPR-MLP	0.727 8	58.83	67.85

PPD+UNeXt: 在基准模型 UNeXt 颈部加上 PPD;(3) PPD+RA+UNeXt: 在 PPD+UNeXt 的基础上,在网络的跳跃连接中加上反向注意力模块;(4)RPR-MLP: 本文提出的网络框架。表 4 为 LungCancer 数据集上进行的消融实验结果,引入 PPD 后,PPD+UNeXt 网络的整体性能相比基准网络有所提升,验证 PPD 聚合高级层特征的能力有助于提升网络的分割性能。再进一步加入反向注意力模块后,分割结果在 PPD+UNeXt 网络的基础上有所提升,进一步验证反向注意力模块在挖掘互补区域和细节方面的有效性。最

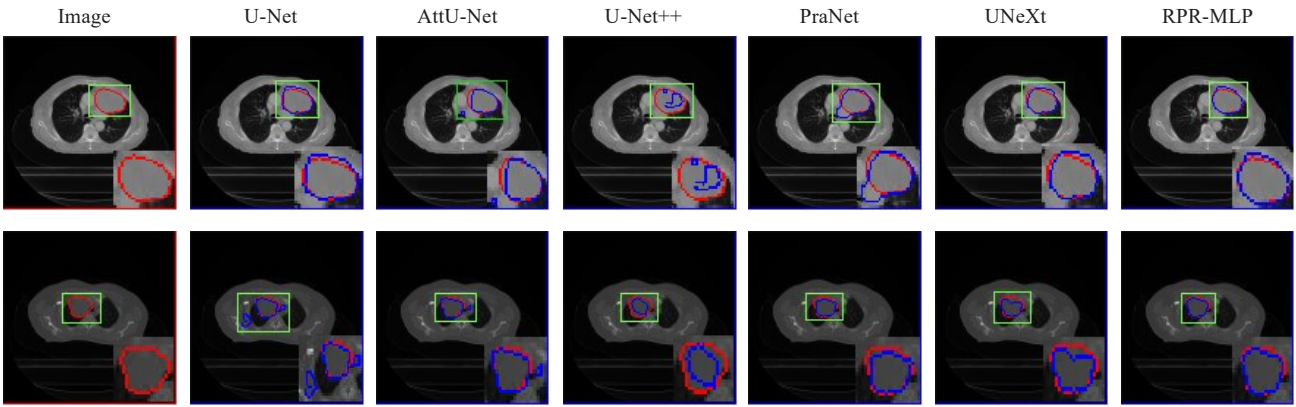


图8 RPR-MLP和对比网络在 LungCancer数据集上定性结果比较

Figure 8 Comparison of qualitative results between RPR-MLP network and other networks on the LungCancer dataset

后,将编码器部分的普通卷积替换为Res2net模块,形成本文提出的网络RPR-MLP。根据表4中的定量分析结果对比,可得出RPR-MLP在不增加太多参数量的情况下,整体性能明显提升。基于以上分析,验证RPR-MLP的有效性,该网络能够有效地对高级层特征进行融合,并挖掘互补区域和细节,从而提高对非小细胞肺癌CT图像中小病灶区域的分割精度。

表4 消融实验结果比较

Table 4 Ablation study results

网络	Params/M	DSC	IoU/%	F1/%
UNeXt	1.47	0.6977	54.77	61.13
PPD+UNeXt	2.31	0.6993	54.65	62.47
PPD+RA+UNeXt	2.93	0.7054	57.11	65.86
RPR-MLP	3.08	0.7278	58.83	67.85

3 总结

本文提出一种新的医学图像分割网络RPR-MLP,首先利用Res2net模块和MLP模块提取更细粒的图像特征,并使用PDD和反向注意力模块来挖掘高级层中的特征信息,显著提升网络特征的提取能力。这些改进方法不仅增强网络对小病灶目标的分割能力,同时解决提取特征信息不足的问题。实验首先利用两个公共医学数据集Kvasir和ISIC 2018对本文提出的网络性能进行测试,结果表明本文网络在两个公共数据集上均取得优异的结果,证明本文网络的有效性和较强的泛化能力。然后将该方法应用于非小细胞肺癌CT图像的分割,结果表明RPR-MLP能够较好地分割出肿瘤区域,与其它方法相比性能

有显著提升,评价指标F1值提升6.72%~18.53%,证明RPR-MLP方法在肺肿瘤CT图像分割任务中有着出色的性能。在接下来的工作中将改进方法,设计更轻量且性能更佳的模块来增强网络的性能,以增加肺肿瘤分割算法的实用性。

【参考文献】

[1] 张玉博, 王志强. 影像组学在非小细胞肺癌中的研究进展[J]. 吉林医药学院学报, 2024, 45(1): 65-68.
Zhang YB, Wang ZQ. Research progress in imaging omics in non-small cell lung cancer[J]. Journal of Jilin Medical College, 2024, 45 (1): 65-68.

[2] 张涛, 袁钰晓, 王国伟, 等. CT在非小细胞肺癌TNM分期的临床应用价值[J]. 中国现代医生, 2023, 61(24): 46-49.
Zhang T, Yuan YX, Wang GW, et al. Clinical value of CT in diagnosing TNM staging of non-small cell lung cancer[J]. China Modern Doctor, 2023, 61(24): 46-49.

[3] 吴晓博, 赵哲, 王瑾瑜, 等. X射线对人非小细胞肺癌细胞的辐射损伤效应研究[J]. 成都医学院学报, 2022, 17(2): 162-165.
Wu XB, Zhao Z, Wang JY, et al. Study on radiation damage effect of X-ray on human non-small cell lung cancer cells[J]. Journal of Chengdu Medical College, 2022, 17(2): 162-165.

[4] 林杰民, 陈俏均, 张雅芝, 等. PET-CT定位在非小细胞肺癌放射治疗中的应用与挑战分析[J]. 中国设备工程, 2023(14): 17-19.
Lin JM, Chen QJ, Zhang YZ, et al. Application and challenge analysis of PET-CT localization in radiotherapy for non small cell lung cancer[J]. China Plant Engineering, 2023(14): 17-19.

[5] 周洁, 郑燕婷, 江舒琪, 等. CT影像组学联合形态学特征模型评估非小细胞肺癌患者预后生存期的价值[J]. 中国医学物理学杂志, 2024, 41(1): 18-26.
Zhou J, Zheng YT, Jiang SQ, et al. Value of CT radiomics combined with morphological features in predicting the prognosis of patients with non-small cell lung cancer[J]. Chinese Journal of Medical Physics, 2024, 41(1): 18-26.

[6] Hesamian MH, Jia WJ, He XJ, et al. Deep learning techniques for medical image segmentation: achievements and challenges[J]. J Digit Imaging, 2019, 32(4): 582-596.

[7] 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.

[8] Zhou ZW, Rahman Siddiquee MM, Tajbakhsh 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 International Publishing, 2018: 3-11.

- [9] Huang HM, Lin LF, Tong RF, et al. UNet 3+: a full-scale connected UNet for medical image segmentation[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2020: 1055-1059.
- [10] 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.
- [11] 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.
- [12] Valanarasu JM, Sindagi VA, Hacihaliloglu I, et al. KiU-net: overcomplete convolutional architectures for biomedical image and volumetric segmentation[J]. IEEE Trans Med Imaging, 2022, 41(4): 965-976.
- [13] Cheng ZK, Wang LJ. Dynamic hierarchical multi-scale fusion network with axial MLP for medical image segmentation[J]. Sci Rep, 2023, 13(1): 6342.
- [14] Zhao HS, Shi JP, Qi XJ, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6230-6239.
- [15] 林嘉豪, 王瑜, 肖洪兵, 等. LRAE-Unet: 轻量级MRI脑肿瘤全自动分割网络[J]. 中国医学物理学杂志, 2024, 41(1): 43-49.
Lin JH, Wang Y, Xiao HB, et al. LRAE-Unet: a lightweight network for fully automatic segmentation of brain tumor from MRI[J]. Chinese Journal of Medical Physics, 2024, 41(1): 43-49.
- [16] Yuan YH, Chen XL, Wang JD. Object-contextual representations for semantic segmentation[C]//Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 173-190.
- [17] Liu HX, Dai ZH, So DR, et al. Pay attention to MLPs[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2024: 9204-9215.
- [18] Lian DZ, Yu ZH, Sun X, et al. AS-MLP: an axial shifted MLP architecture for vision[EB/OL]. (2022-03-17). <https://arxiv.org/abs/2107.08391>.
- [19] Gao SH, Cheng MM, Zhao K, et al. Res2net: a new multi-scale backbone architecture[J]. IEEE Trans Pattern Anal Mach Intell, 2021, 43(2): 652-662.
- [20] Valanarasu JM, Patel VM. UNeXt: MLP-based rapid medical image segmentation network[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2022. Cham: Springer Nature Switzerland, 2022: 23-33.
- [21] Wu Z, Su L, Huang QM. Cascaded partial decoder for fast and accurate salient object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 3902-3911.
- [22] Chen SH, Tan XL, Wang B, et al. Reverse attention for salient object detection [C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 236-252.
- [23] Fan DP, Ji GP, Zhou T, et al. PraNet: parallel reverse attention network for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2020. Cham: Springer International Publishing, 2020: 263-273.
- [24] Cao H, Wang YY, Chen J, et al. Swin-unet: unet-like pure transformer for medical image segmentation[C]//Computer Vision-ECCV 2022 Workshops. Cham: Springer Nature Switzerland, 2023: 205-218.
- [25] Oktay O, Schlemper J, Le Folgoc L, et al. Attention U-net: learning where to look for the pancreas[EB/OL]. (2018-05-20). <https://arxiv.org/abs/1804.03999>.
- [26] Liu ST, Huang D, Wang YH. Receptive field block net for accurate and fast object detection [C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 404-419.
- [27] Codella NC, Gutman D, Celebi ME, et al. Skin lesion analysis toward melanoma detection: a challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC) [C]//2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). Piscataway, NJ, USA: IEEE, 2018: 168-172.
- [28] Jha D, Smedsrud PH, Riegler MA, et al. Kvasir-SEG: a segmented polyp dataset [C]//MultiMedia Modeling. Cham: Springer International Publishing, 2020: 451-462.
- [29] Hendrycks D, Gimpel K. Gaussian error linear units (GELUs)[EB/OL]. (2023-06-06). <https://arxiv.org/abs/1606.08415>.
- [30] Chen LC, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional Nets, atrous convolution, and fully connected CRFs[J]. IEEE Trans Pattern Anal Mach Intell, 2018, 40(4): 834-848.

(编辑:陈丽霞)