

基于边缘约束和改进Swin Unetr的复杂器官分割方法

彭琳娜, 张红云, 苗夺谦

引用本文

彭琳娜, 张红云, 苗夺谦. 基于边缘约束和改进Swin Unetr的复杂器官分割方法[J]. 计算机科学, 2025, 52(4): 177-184.

PENG Linna, ZHANG Hongyun, MIAO Duoqian. Complex Organ Segmentation Based on Edge Constraints and Enhanced Swin Unetr [J]. Computer Science, 2025, 52(4): 177-184.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多尺度融合注意力的多视角文档图像篡改检测与定位](#)

Multi-view and Multi-scale Fusion Attention Network for Document Image Forgery Localization

计算机科学, 2025, 52(4): 327-335. <https://doi.org/10.11896/jsjcx.240100142>

[大选择性核双边网络的长尾分布医学图像分类方法](#)

Long-tail Distributed Medical Image Classification Based on Large Selective Nuclear Bilateral-branch Networks

计算机科学, 2025, 52(4): 231-239. <https://doi.org/10.11896/jsjcx.240700039>

[面向无人机航拍图像的YOLOv8目标检测改进算法](#)

An Improved YOLOv8 Object Detection Algorithm for UAV Aerial Images

计算机科学, 2025, 52(4): 202-211. <https://doi.org/10.11896/jsjcx.240500042>

[基于YOLO-Unet组合网络的牛只个体识别方法研究](#)

Research on Individual Identification of Cattle Based on YOLO-Unet Combined Network

计算机科学, 2025, 52(4): 194-201. <https://doi.org/10.11896/jsjcx.240100144>

[基于虹膜纹理感知的自适应关联学习方法](#)

Adaptive Contextual Learning Method Based on Iris Texture Perception

计算机科学, 2025, 52(4): 185-193. <https://doi.org/10.11896/jsjcx.250100022>

基于边缘约束和改进 Swin Unetr 的复杂器官分割方法

彭琳娜 张红云 苗夺谦

同济大学电子与信息工程学院 上海 201804

(lnpeng@tongji.edu.cn)

摘要 针对腹部 CT 多器官分割任务中器官边缘模糊、器官比例差异过大的问题,提出了基于边缘约束和改进 Swin Unetr 的复杂器官分割方法。为了在不同体素比例的器官上提取精细程度不同的特征,设计了掩码注意力模块,通过计算各个器官的掩码信息,提取对应特征。随后,以数据集先验和掩码信息为基础,在相应的窗口和块大小上进行特征提取,以获得小比例器官分割所需的精细化特征,并与编码器的输出特征融合;同时,输出初步预测的语义分割结果后,为了充分利用边界信息,增强模型对于边界信息的处理能力,输出的语义特征通过卷积层进一步提取出边界信息,通过边缘损失最小化使模型的语义分割结果受到边缘预测任务的约束。在 BTCV 和 TCIA pancreas-CT 数据集上对所提方法进行训练和测试,在基于卷积网络的 UNet++ 和基于 Transformer 的 Swin Unetr 上加入了提出的改进模块并进行训练,与 Unetr 等经典网络进行了对比实验。在 BTCV 数据集上,所提模型 Dice 系数分别达到了 0.8479 和 0.8406,HD 距离分别为 11.76 和 8.35,整体上优于其他对比方法,从而验证了所提方法的有效性和可行性。

关键词: 图像分割;注意力机制;计算机断层扫描分割;深度学习;多器官分割;多任务学习;计算机辅助诊断

中图分类号 TP391

Complex Organ Segmentation Based on Edge Constraints and Enhanced Swin Unetr

PENG Linna, ZHANG Hongyun and MIAO Duoqian

College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China

Abstract To address the challenges of organ edge blurring and significant differences in organ proportions in abdominal CT multi-organ segmentation tasks, this paper proposes a complex organ segmentation approach based on edge constraints and enhanced Swin Unetr. To extract features of varying degrees of granularity from organs with different voxel proportions, this study introduces the Masked Attention Block. By computing the mask information of each organ, the corresponding features are extracted. Subsequently, based on dataset priors and mask information, refined feature extraction is conducted within appropriate window and block sizes to obtain the fine-grained features necessary for segmenting organs with smaller voxel proportions. Upon generating preliminary semantic segmentation predictions, to fully leverage boundary information and enhance the model's ability to handle such information, the semantic features are further extracted through convolutional layers to capture boundary details. The model's semantic segmentation results are constrained by the edge prediction task through edge loss minimization. The proposed method is trained and tested on the BTCV and TCIA pancreas-CT datasets. The enhancement modules are incorporated into the UNet++ based on convolutional networks and the Swin Unetr based on Transformers for training. Comparative experiments are conducted with classic networks such as Unetr. On the BTCV dataset, the Dice coefficients reach 0.8479 and 0.8406, with corresponding Hausdorff distances of 11.76 and 8.35, respectively. Overall, the proposed method outperforms other comparative methods, confirming its effectiveness and feasibility.

Keywords Image segmentation, Attention mechanism, Computed tomography segmentation, Deep learning, Multi-organ segmentation, Multi-task learning, Computer-aided diagnosis

到稿日期:2024-05-31 返修日期:2024-08-26

基金项目:国家自然科学基金(62076182,62376198);国家重点研发计划(2022YFB3104700);上海市自然科学基金(22ZR1466700)

This work was supported by the National Natural Science Foundation of China(62076182,62376198), National Key Research and Development Program of China(2022YFB3104700) and Natural Science Foundation of Shanghai(22ZR1466700).

通信作者:张红云(zhanghongyun@tongji.edu.cn)

1 引言

在计算机辅助诊断、外科手术规划和放射性治疗等领域，都需要对计算机断层扫描的腹部器官进行精确分割^[1]。器官分割的精度直接影响了对病变的定位和分期以及诊断的定量分析。此外，基于器官分割的三维重建可以在外科手术规划和放射治疗中提供可视化的解剖环境，有助于最大限度地减少对周围健康组织的损伤。

由于 CT 图像的手动分割非常耗时且不稳定，因此自动分割方法受到了广泛关注。传统的 CT 图像分割方法通常涉及到手工的图像特征检测^[2-5]，如基于阈值的方法、基于区域的分割方法等，一般是根据强度或梯度的差异来实现的。使用这些方法一般需要大量的专业知识和大量手动设计的特征。例如，基于阈值^[6]的分割结果受阈值选择的影响，在灰度值不均匀、边界模糊的多器官分割中鲁棒性较差，而基于区域^[7]的分割方法容易受噪声和灰度值不均的影响产生空洞或过分割。因此，这些方法仅适用于粗略的器官分割或特定器官的分割。

深度学习模型由于其深层结构具备强大的非线性建模能力，因此在复杂任务中表现出色，使用深度结构使得模型能够更好地捕捉和理解特征的复杂关系^[8]，从而提高了模型的分割性能。随着深度学习模型的发展，以 U-Net^[9] 为代表的模型能够自动提取图像特征，无需复杂的非刚性配准。因此，在多器官分割任务中，这些模型在效率和有效性方面具有巨大优势。

由于人体腹部结构的复杂性，基于 U-Net 的分割方法仍然面临两个主要困难。首先，器官在大小和形状上存在着很大差异，较小的结构通常具有更大的形变差异，这会降低较小器官的分割性能。如图 1 所示，在 BTCV 数据集中，肝脏、胃、脾、左肾和右肾占据了近 90% 的体素，剩余 8 个器官只占据剩

下的 10% 体素，而食管、右肾上腺和左肾上腺的比例均低于 0.5%。这些占比小的器官特征难以捕捉，给分割任务带来了挑战。其次，腹部 CT 的背景复杂，相邻器官之间的对比度低，使得器官边界区域模糊难分，导致属于这些区域的体素更容易受到错误标签预测的影响。图 2(a) 是 TCIA pancreas-CT 数据集中 0002 样本的横断面切片，十二指肠在切片上不连续，在三维空间中通常形变复杂，导致十二指肠分割的精确度低。图 2(b) 和图 2(c) 分别是 BTCV 数据集中 0035 的矢状面和冠状面切片。矢状面切片展现出主动脉边缘的复杂结构，冠状面切片可以看出胃、静脉、胰腺、肾上腺与食管边缘紧密，结构复杂。在医学环境下，这些区域的分割精度往往更有价值，而上述问题导致在实验中复杂小型器官以及器官相邻区域局部细节的分割精度未达到实际医学应用的要求。

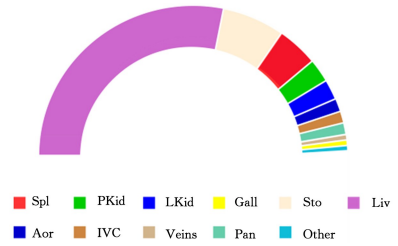


图 1 BTCV 数据集中各器官的比例

Fig. 1 Proportion of each organ in BTCV dataset

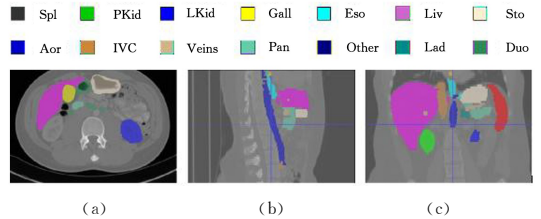


图 2 多器官 CT 切片可视化

Fig. 2 Visualization of multiple-organ CT sections

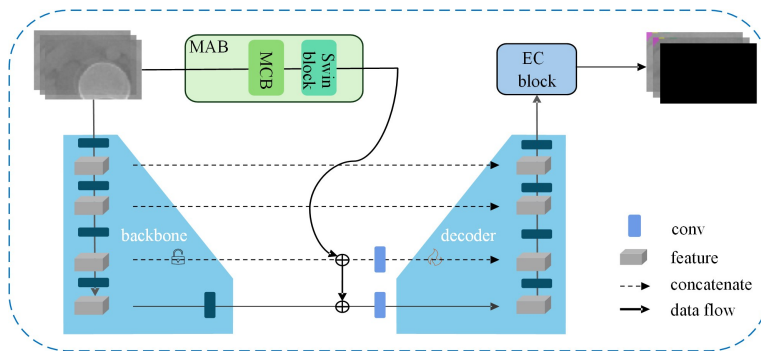


图 3 整体网络结构

Fig. 3 Overall network structure

另一方面，在目前的成熟模型中，边缘区域是在任务中出错率最高的空间位置，因此器官 CT 数据的边缘信息对分割任务具有重要价值。为了充分利用模型参数和提高模块的灵活型，本文进行共享参数的多任务训练，然后利用学习到的边缘信息约束分割结果的预测。

本文提出的基于边缘约束和改进 Swin Unetr 的复杂器官分割方法的整体结构如图 3 所示。输入的 CT 序列通过 MAB 自适应地提取不同体素比例器官的对应特征，将获得的

特征与编码器输出的最后两层特征进行融合，将融合后的特征跳接到解码器。解码器输出的特征通过边界约束(Edge-Constrained, EC)模块进行基于边缘约束的语义分割，然后将学习到的边界特征与语义特征融合后输出最终的分割结果。

与传统的基于 Transformer 的 Unetr 相比，本文提出的基于边缘约束和改进 Swin Unetr 的复杂器官分割方法具有以下两个方面的优势。一方面，针对体素比例小的器官，利用掩码信息获得较准确的分割区域，并使用自适应的窗口注意

力进一步细化器官特征。另一方面,引入边界约束,通过最小化边缘损失约束分割体的生成。在引入少量参数的前提下缓解了模型生成分割结果时边缘模糊的影响。

2 相关工作

目前基于全卷积网络的腹部多器官分割方法一般采用 2D 或 3D 卷积架构^[10-14]。2D 方法以逐片的方式处理 CT 扫描,并预测各个切片上的器官标签^[15]。二维切片方法的优势在于参数量低、训练效率高,但其难以充分利用三维图像的上下文信息。3D 方法通过处理整个三维 CT 图像,并在跳接过程中拼接不同层次的特征来利用丰富的体素上下文,从而获得比 2D 方法更好的分割性能。

现有的 3D 方法主要集中于设计更好的架构来改进表征学习。针对 UNet 网络结构中跳接带来的语义鸿沟问题,Zhou 等^[16]认为高分辨率的特征图和对对应语义的特征图相融合的模式可以捕捉更多细节,提出了 UNet++ 网络,重新设计了一系列嵌套的密集跳接路径,使得模型在连通性上得到提升。Jin 等^[17]提出了 RA-UNet 网络,使用 3D UNet 作为基本框架,堆叠注意力模块以精确提取感兴趣的肝脏体素,但它们从全局范围内探索特征联系的能力仍有欠缺。

卷积网络中卷积层的核大小有限,其对远程信息相关性的建模能力并不理想,这限制了模型处理形变差异大的器官特征的能力。Transformer^[18]模型在多个领域(包括自然语言处理和计算机视觉等)表现出了捕获长距离信息的出色能力,受其影响,Hatamizadeh 等^[19]尝试利用 Transformers 作为编码器来学习输入 CT 体素的序列表示并提出了 Unetr 模型,以有效地捕获全局多尺度特征。Unetr 遵循了编码器和解码器结构的 UNet 网络架构,Transformers 编码器通过不同分辨率的跳接直接连接到解码器,计算获得最终的语义分割结果。受到 Swin Transformer 的启发,Nath 等^[20]提出了 Swin Unetr。Swin Unetr 模型利用 Swin Transformer 作为编码器,利用移动窗口计算自注意力,提取 5 种不同分辨率的特征,并通过跳接连接到基于全卷积网络的解码器。

针对 Swin Unetr 训练过程的有效性受到了训练时间的限制的问题,MetaSwin^[21]用空间池化操作替换 Swin Unetr 中基于注意力的组件并将压缩和激励(Squeeze-and-Excitation, SE)模块合并到解码器中,有效降低了参数量并缩短了训练时长。为了提高模型的可解释性,SDAH-UNet^[22]将移位窗口可变形注意合并到一个混合架构中,以提高分割性能,同时确保可解释性。为了同时利用 Transformer 的长程依赖捕捉能力和卷积的局部特征提取能力,Hiformer^[23]使用 Swin Transformer 模块和基于 CNN 的编码器设计了两种多尺度特征表示,在编码器-解码器轮廓的跳跃连接中提出了双层融合模块。

这些改进网络进一步提高了模型提取器官断层扫描特征的能力,但目前仍存在着器官边缘模糊和器官比例差异过大造成的小器官分割精度低的问题。针对这个问题,本文提出改进的 Swin Unetr 和边缘约束模块。

3 改进 Swin Unetr 模块

本章简要介绍 Swin Unetr 并具体介绍 MAB 模块的细节。

3.1 Swin Unetr

Swin Unetr 采用 U 型网络设计,使用 Swin Transformer

作为编码器和基于 CNN 的解码器,编码器的各级输出跳接到解码器的各级输入。

如图 4 所示,模型的输入是四维的 mri 数据,通过图块划分层(patch partition layer)将 mri 数据分块,然后通过 4 个 Swin Transformer 模块提取不同维度的特征,将多分辨率的特征跳接给 CNN 解码器。如图 4 中的右下部分所示,Swin Transformer 模块由一个基于 W-MSA 和 SW-MSA 的注意力模块和一个带有高斯误差线性单元(Gaussian Error Linear Unit, GELU)非线性激活函数的双层 MLP 组成。在每个 MSA 模块和每个 MLP 模块之前都有一个层归一化(Layer Normalization, LN)层。

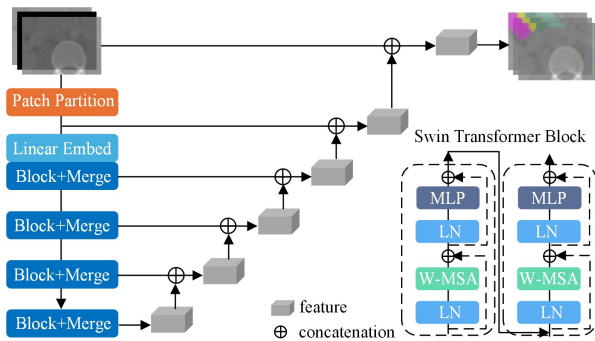


图 4 Swin Transformer 网络结构

Fig. 4 Network structure of Swin Transformer

其中,W-MSA 和 SW-MSA 都是 Swin Transformer 的核心组件。W-MSA 将输入特征分割成固定大小的窗口,然后在每个窗口上应用多头自注意力机制,从而实现全局上下文的捕获。而 SW-MSA 则引入了位移操作,使得相邻窗口之间可以共享信息,提高了模型的参数效率和性能。为了有效计算移位窗口机制,使用了三维循环移动^[24],其中使用到的自注意力的计算式如下:

$$Attention(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

3.2 Masked Attention Block

考虑到某些器官的分割效果较差是因为它们占比小且形态差异大,但一般的网络结构中并没有针对它们进行额外的处理,本文设计了 MAB 来对其所在的局部区域特征进行细化提取。

如图 5 所示,输入数据通过推理层和 softmax 层计算得到对应掩码,通过滚动操作扩大掩码范围,然后提取对应特征,在适合的窗口和块大小上进行精细化特征提取。

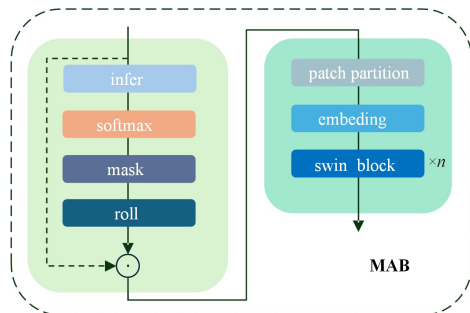


图 5 MAB 模块结构

Fig. 5 MAB module structure

如图 3 所示,提取出的特征与编码器的深层输出融合,以缓解在网络深层细节特征丢失的问题。形式化表示如下:

$$\begin{cases} \hat{\mathbf{y}} = \text{softmax}(l(\mathbf{y})) \\ \mathbf{m}_{h,w,d} = [\hat{\mathbf{y}}_{h,w,d} > \epsilon] \end{cases} \quad (1)$$

其中, \mathbf{y} 是体输入, l 是固定参数的推理层, $\hat{\mathbf{y}} \in \mathbb{R}^{b,k,h,w,d}$ 是经过初步预测得到的分割结果, k 是类别数, $[\]$ 是 Iverson 表示, ϵ 是选定的阈值, $\mathbf{m}_{h,w,d}$ 是 (h, w, d) 处对应的掩码值。

掩码的计算使用了张量滑动以扩大范围,避免因推理误分类而造成的边缘体素丢失。

$$\begin{cases} \mathbf{m}'_{h,w,d} = \mathbf{S}(\mathbf{m}_{h,w,d}, \rho) \\ \mathbf{m}_{h,w,d} = \mathbf{m}_{h,w,d} + \sum(\mathbf{m}'_{h,w,d}) \end{cases} \quad (2)$$

其中, \mathbf{S} 表示张量在 h, w, d 维度滚动 ρ 个体素,如图 6 所示,向 8 个方向滚动后求和,最后利用式(3)计算得到掩码 \mathbf{m} 在 h, w, d 处对应的值。通过式(4)获得最终的输出,其中 \odot 表示哈达玛积。

$$\mathbf{m}_{h,w,d} = [\mathbf{m}_{h,w,d} > 0.5] \quad (3)$$

$$\mathbf{x} = \mathbf{y} \odot \mathbf{m} \quad (4)$$

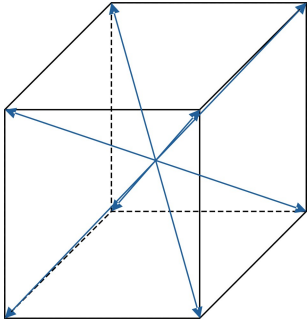


图 6 滚动操作的方向示意图

Fig. 6 Direction chart of rolling operation

以二维图像为例,图 7(a)是推理预测的器官位置,图 7(b)是经过张量滑动后的器官可能区域,图 7(c)是扩大域后将要进行后续特征提取的器官位置。

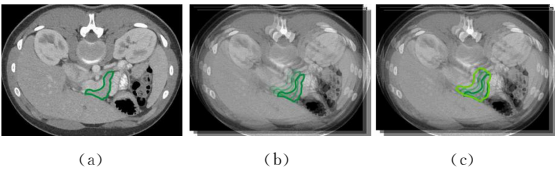


图 7 掩码计算

Fig. 7 Computation of mask

为了更精细地对断层扫描进行分割,不采用传统 Unetrns 对所有特征都使用完全相同的窗口和块来进行注意力计算和特征提取的方法。对于器官 k , $l_k = \max(m_{w,k}, m_{h,k}, m_{d,k})$ 表示掩码中 w, h, d 中长度最大的体素距离,窗口大小根据 l_k 进行自适应调整,patch 大小根据窗口大小进行调整。

$$ws = \lceil \log_2 l_k \rceil \quad (5)$$

$$ps = \lfloor \log_2 ws \rfloor \quad (6)$$

在不同的数据集上, l_k 可由上述计算式计算得到,也可以参考数据集的先验信息,即对应器官的平均范围。改变底数 a 可以进一步进行调整。

$$ws = \lceil \log_a l_k \rceil \quad (7)$$

如图 3 所示,将获得的特征和编码器的最后两层输出连接后与解码器跳接,以缓解编码器深层细节特征丢失的问题,减少细节特征丢失带来的复杂小型器官边缘分割不精确,细节部分被错误分割或者遗漏的问题。

4 共享参数的边缘约束

本章首先介绍器官分割任务中的 EC block,然后介绍本文使用的损失函数,即为 Soft Dice 损失、交叉熵损失和边缘损失的加权和。

4.1 边缘约束网络

多任务学习通常通过硬参数共享和软参数共享来实现。在硬参数共享范式中,多个任务共享一个共同优化的参数集,而针对特定任务的参数则是单独优化的。在软参数共享中,每个任务都使用自己的参数集进行参数化,这些参数集使用约束进行联合正则化。针对边缘约束的分割预测任务,本文使用图 8 所示的网络拓扑结构来控制训练参数大小。在不同基准模型上的实验结果显示,共享解码器参数以及针对两个任务分别使用不同的解码器这两种方法的分割效果在不同的基线网络和不同的器官上都有细微的差异,但整体差别不大,这与文献[25]的实验结果一致,因此本文使用结构更简单灵活的共享编码器解码器结构。

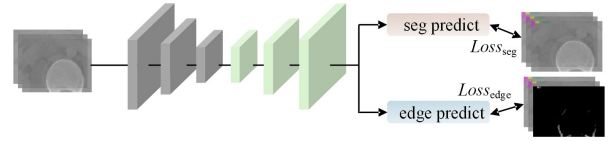


图 8 边缘和分割预测的网络拓扑结构

Fig. 8 Network topologies for edge prediction and segmentation prediction

利用侵蚀操作从真实分割中提取出边缘信息。

$$\hat{\mathbf{x}} = \mathbf{x} \ominus \kappa$$

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$$

其中, \ominus 表示侵蚀操作, κ 是选择的卷积核, \mathbf{e} 是获得的 \mathbf{x} 图像对应的边缘信息。

边缘约束模块的网络结构如图 9 所示。

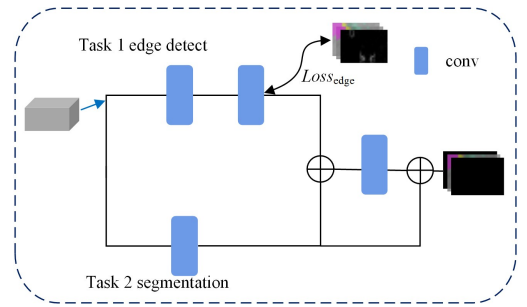


图 9 边缘约束模块

Fig. 9 Edge-constrained block

模型的编码器和解码器在特征提取和特征解码过程中共享参数,输出的是初步的预测结果。为了充分利用边界信息,增强模型对边界信息的处理能力,输出的语义特征通过卷积层进一步提取出对应的各器官的边界信息,通过最小化边界损失使模型的语义分割结果受到边缘预测任务的约束。将学

习到的边界特征与分割特征连接后进行融合,通过卷积层整合特征并使用残差连接将初步分割预测与边缘约束的分割预测相连接作为输出,采用输出结果与真实分割标签来计算交叉熵损失和 Dice 损失。边缘约束模块因为共用分割预测任务的编码器解码器参数,所以有很好的拓展性,在仅添加少量参数的同时,对于各种网络都有很好的性能提升效果。

4.2 损失函数

本文使用的损失函数是 Soft Dice 损失、交叉熵损失和边缘损失的组合。为了提升小型器官的分割效果,对于类别 j , 计算交叉熵时,使用体素比例的倒数作为系数,以提升小型器官分割精确度对模型训练的影响。损失函数可以以体素的方式按照式(8)进行计算。

$$\begin{cases} d_j = \sum_{i=1}^I \mathbf{G}_{i,j} \mathbf{Y}_{i,j} \\ d_j' = \sum_{i=1}^I \mathbf{G}_{i,j}^2 + \sum_{i=1}^I \mathbf{Y}_{i,j}^2 \\ CE_g = I \sum_{j=1}^J \left(\frac{1}{I_j} \sum_{i=1}^I \mathbf{G}_{i,j} \log(\mathbf{Y}_{i,j}) \right) \\ CE_e = \sum_{j=1}^J \sum_{i=1}^I \mathbf{E}_{i,j} \log(\ell(\mathbf{Y}_{i,j})) \\ L(\mathbf{G}, \mathbf{Y}, \mathbf{E}) = 1 - \frac{2}{J} \sum_{j=1}^J (d_j / d_j') - CE_g - \beta CE_e \end{cases} \quad (8)$$

其中, I 是体素数量; J 是类别数目; $\mathbf{Y}_{i,j}$, $\mathbf{G}_{i,j}$ 和 $\mathbf{E}_{i,j}$ 分别表示类别 j 在体素 i 处的预测标签、真实标签和边界标签; ℓ 表示隐藏层; α 和 β 是超参数。

5 实验结果与分析

本章将介绍本文使用的数据集、评价指标以及实现细节和实验结果分析。

5.1 数据集

为了验证所提方法的有效性,本文使用 BTCV 和 TCIA Pancreas-CT 数据集进行腹部 CT 多器官分割任务的训练和测试。

BTCV 数据集由 30 名受试者的腹部 CT 扫描结果组成,这些扫描源自转移性肝癌患者或术后腹壁疝患者,在范德比尔特大学医学中心临床放射科医生的监督下对 13 个器官进行了标注。每次的 CT 扫描都是在门静脉造影剂阶段拍摄的。数据集包含了 85~198 个切片,像素为 512×512 , 平面内分辨率为 $0.54 \times 0.54 \text{ mm}^2 \sim 0.98 \times 0.98 \text{ mm}^2$, 切片厚度为 $2.5 \sim 5.0 \text{ mm}$ 不等。每个切片进行了独立预处理,包含脾、右肾、左肾、胆囊、胃、肝脏、主动脉、下腔静脉、胰腺、门静脉和脾静脉、食管、左肾上腺和右肾上腺 13 个器官。

TCIA Pancreas-CT 数据集是由癌症图像档案(TCIA)提供的,主要包含胰腺 CT 图像。这些图像来自美国国立卫生研究院临床中心,包括来自肾切除手术前的健康肾供者和没有主要腹部病变或胰腺癌患者的腹部 CT 扫描。每张图像由医学生手动标记,并由经验丰富的放射科医生验证和修改。对于原始数据集中不存在的数据,在具有 8 年胃肠道 CT 和 MRI 图像解释经验的认证放射科医生的监督下,由图像研究员使用 Matlab 2015b 和 ITK-SNAP 3.2 进行了交互式的分割。对原始数据集中的分割进行编辑,以确保跨数据集的

一致性^[26]。该数据集包括脾脏、左肾、胆囊、食管、肝脏、胃、胰腺、十二指肠。

5.2 评价指标

本文在实验中使用 Dice 分数和 95% 豪斯多夫距离 (Hausdorff Distance, HD) 来评估分割的准确性。对于一个给定的语义类别, G_i 和 P_i 分别表示体素 i 的真实类别和预测值, G' 和 P' 分别表示真实点集和预测点集。Dice 分数和 95% 豪斯多夫距离的定义分别为:

$$\begin{aligned} Dice(\mathbf{G}, \mathbf{P}) &= \frac{2 \sum_{i=1}^I \mathbf{G}_i \mathbf{P}_i}{\sum_{i=1}^I \mathbf{G}_i + \sum_{i=1}^I \mathbf{P}_i} \\ HD(\mathbf{G}', \mathbf{P}') &= \max \left\{ \max_{g' \in \mathbf{G}'} \min_{p' \in \mathbf{P}'} \|g' - p'\|, \max_{p' \in \mathbf{P}'} \min_{g' \in \mathbf{G}'} \|p' - g'\| \right\} \end{aligned}$$

95% 豪斯多夫距离指使用真实点集和预测点集的距离从大到小排列后,排名为 5% 的距离。因此,在计算 HD 距离时,非常小的异常值子集的影响会被最小化。

5.3 实现细节

本文的实验使用了 PyTorch 和 monai 框架。模型使用 NVIDIA GeForce RTX 3090 进行训练。本文随机地将 BTCV 数据集分为两部分:24 次扫描用于训练,6 次扫描用于测试。然后随机地选择 6 个训练扫描形成一个验证集,用于选择模型的超参数。对于 TCIA 数据集,使用 31 个扫描数据用于训练,4 个扫描用于验证,6 个扫描用于测试。最后的测试集上的实验结果由训练集全体训练获得。

所有模型的训练批大小为 1,使用 AdamW 优化器。对于特定的批大小和迭代次数,平均训练时间是 40h。本文的基于 Swin Unetr 的编码器参考了原始的配置,特征维度设置为 24。对于图 3 中未参加训练的编码器子模块和推理模块,使用了 Swin Unetr 的预训练模型。对于 BTCV 和 TCIA pancreas-CT 数据集,将 CT 值范围归一化到 (0,1),将随机采样大小为 $96 \times 96 \times 96$ 尺寸的图像作为输入,使用了随机翻转、随机旋转 90 度、随机强度偏移等进行数据增强。在推理过程中,使用了滑动窗口方法进行推理。实验中的其他模型均来自官方代码,UNet++ 的特征维度为 (16, 32, 64, 128, 256, 32),其他参数使用默认值。

5.4 实验结果与分析

对于卷积网络和基于 Transformer 的分割网络,EC block 和 MAB 模块对不同的模型在不同程度上提升了分割效果。表 1 列出了不同方法在 6 个测试样本上各器官的平均 Dice 系数。表 1 中的 EM+UNet++ 是在 UNet++ 的基础上增加了 EC block 和 MAB 模块,EM+UNet++ 在脾脏、右肾、左肾、食管、肝脏、主动脉、下腔静脉、胰腺、右肾上腺和左肾上腺上的 Dice 系数均高于 UNet++。其中在中型器官食管(体素占比 0.50%)和胰腺(体素占比 2.66%)上提升超过了 1%,分别为 2.43% 和 3.66%;在小型器官右肾上腺(体素占比 0.13%)和左肾上腺(体素占比 0.16%)上的提升超过了 0.3%,分别为 0.39% 和 0.79%;平均 Dice 系数提升了 0.59%。表 2 列出了各方法的平均 95% 豪斯多夫距离,

EM+UNet++比 UNet++的 HD 距离降低了 0.89,表明新模型的预测分割和真实分割的轮廓差异度更低。

表 1 BTCV 数据集上不同方法的各器官分割 Dice 系数

Table 1 Organ segmentation Dice coefficients of different methods on BTCV dataset

方法	Spl	RKid	LKid	Gall	Eso	Liv	Sto	Aor	IVC	Veins	Pan	Rad	Lad	AVG
UNet++ ^[16]	94.93	94.01	94.01	81.62	72.46	96.03	83.62	89.82	85.55	75.96	77.57	66.86	66.62	84.20
EM+UNet++	95.36	94.48	94.56	81.23	74.89	96.57	82.90	90.10	86.50	74.84	81.23	67.25	67.41	84.79
Unetr ^[19]	91.96	93.16	93.37	69.56	72.49	95.55	80.11	88.09	81.56	68.64	76.89	62.20	59.05	80.88
Swin Unetr ^[20]	93.92	93.67	93.33	78.16	69.48	96.29	78.67	89.18	84.64	73.97	77.86	69.46	59.17	82.67
MetaSwin ^[21]	94.24	93.81	92.56	80.49	70.75	96.09	79.15	89.68	83.58	71.90	78.51	67.99	60.73	81.50
E+Swin Unetr	94.27	93.36	93.71	82.22	76.85	96.39	80.80	90.11	84.86	72.27	80.56	66.94	62.99	83.93
EM+Swin Unetr	94.67	93.62	92.41	84.81	75.20	96.20	77.99	89.16	84.82	73.08	79.84	69.62	65.69	84.06

表 2 BTCV 数据集上不同方法的平均豪斯多夫距离

Table 2 Average 95% Hausdorff distance of different methods on BTCV dataset

方法	HD_AVG
UNet++ ^[16]	12.65
EM+UNet++	11.76
Unetr ^[19]	20.46
Swin Unetr ^[20]	13.30
MetaSwin ^[21]	15.05
E+Swin Unetr	9.41
EM+Swin Unetr	8.35

表 1 中, E+Swin Unetr 是在 Swin Unetr 的基础上添加了 2.2 节中提到的边缘约束模块。表 3 列出了实验中 MAB 模块和边缘约束模块的参数大小。表 3 中, 边缘约束模块的参数大小为 0.04 MB。在 Swin Unetr 上增加了参数量为 0.04 MB 的模块, 使得模型在 14 个类别上的平均 Dice 系数提高了 1.26%, 其中胆囊、食道、胃、胰腺和左肾上腺的 Dice 系数提升超过 2%, 分别为 4.06%, 7.37%, 2.13%, 2.70%, 3.82%。在这之中食管的效果比预期的更好, 这是因为食管是分割的 BTCV13 种器官中形变最大的, 它在不同病人的体内形态差异很大, 且在同一个患者体内常常和其他器官之间的边界紧密, 边缘复杂。通过分析实验结果可知, 基于改进 Swin Unetr 模块和边缘约束模块的方法对食道分割这种区域跨度大、边界预测难度高的任务有很好的提升效果, 因为其不仅能很好地进行长距离特征建模, 还能利用边缘约束模块解决分割难点中的局部模糊边缘问题。

EM+Swin Unetr 是在 E+Swin Unetr 的基础上增加

表 4 TCIA pancreas-CT 数据集上不同方法的各器官分割 Dice 系数

Table 4 Organ segmentation Dice coefficient of different methods on TCIA pancreas-CT dataset

方法	Spl	LKid	Gall	Eso	Liv	Sto	Pan	Duo	AVG
Unetr ^[19]	95.74	94.41	74.23	63.14	95.54	83.80	71.78	55.42	79.26
Swin Unetr ^[20]	96.68	95.08	73.98	71.52	95.69	89.07	77.67	67.43	83.39
MetaSwin ^[21]	96.44	95.00	75.97	71.36	96.03	88.64	77.05	63.89	83.05
E+Swin Unetr	96.48	94.84	77.72	72.46	96.01	89.86	79.51	67.34	84.28
EM+Swin Unetr	96.68	94.92	77.92	71.88	96.06	89.05	79.72	68.79	84.38

图 10 给出了样本中随机选择两个样本在不同模型下分割结果的随机切片可视化。第一行和第二行分别是 CT 样本 0035 和 CT 样本 0040 的随机切片。为了观察实验效果, 需要选择同时有多个器官的切片进行可视化, 因此随机切片位置在腹部的中部。为了更清晰地比较分割结果, 图 10 的下面两

行了 2.1 节介绍的 MAB 模块。如表 3 所列, 实验中增加的模块的参数大小为 36.05 MB。由于推理层和编码器在训练过程中不参与训练, 因此增加的训练参数大小在可接受的范围内。模型各类别的平均 Dice 系数提高了 0.13%, 其中脾脏、右肾、胆囊、门静脉和脾静脉、右肾上腺和左肾上腺的表现超过了 E+Swin Unetr, 分别提高了 0.4%, 0.26%, 2.59%, 0.81%, 2.68%, 2.70%, 且在中型器官胆囊(体素占比 0.84%)和小型器官右肾上腺、左肾上腺上的提升超过了 2%。MAB 模块聚焦于体素比例较小的小型器官, 分析实验结果可以发现, 这些小型器官的分割精度得到了不同程度的提高。如表 2 所列, E+Swin Unetr 的平均 HD 距离为 9.41, EM+Swin Unetr 的平均 HD 距离在此基础上降低了 1.06。

表 3 各模块参数大小

Table 3 Params size of blocks

模块	Params size/MB
MAB	36.05
EC Block	0.04

如表 4 所列, 在 TCIA pancreas-CT 数据集上, E+Swin Unetr 在胆囊、食道、肝脏、胃和胰腺上的分割精度对比均有不同程度的提升, 分别提升了 3.74%, 0.94%, 0.32%, 0.79% 和 1.84%, 平均 Dice 系数提高了 0.89%。EM+Swin Unetr 在中小型器官胆囊、胰腺和十二指肠上有明显的提升, 相比于 E+Swin Unetr, EM+Swin Unetr 在胆囊上有 0.2% 的提升, 在胰腺上有 0.21% 的提升, 在十二指肠上有 1.45% 的提升。

行是上面两行对应区域的放大, 可以看到与真实结果 (Ground Truth, GT) 相比, EM+UNet++ 和 EM+Swin Unetr 的分割精细度高于前面的 4 个模型。分割结果显示, 边缘约束模块使得分割边缘更加平滑, MAB 模块使得分割与 GT 的相似度更高, 不符合抖动减少的真实情况。

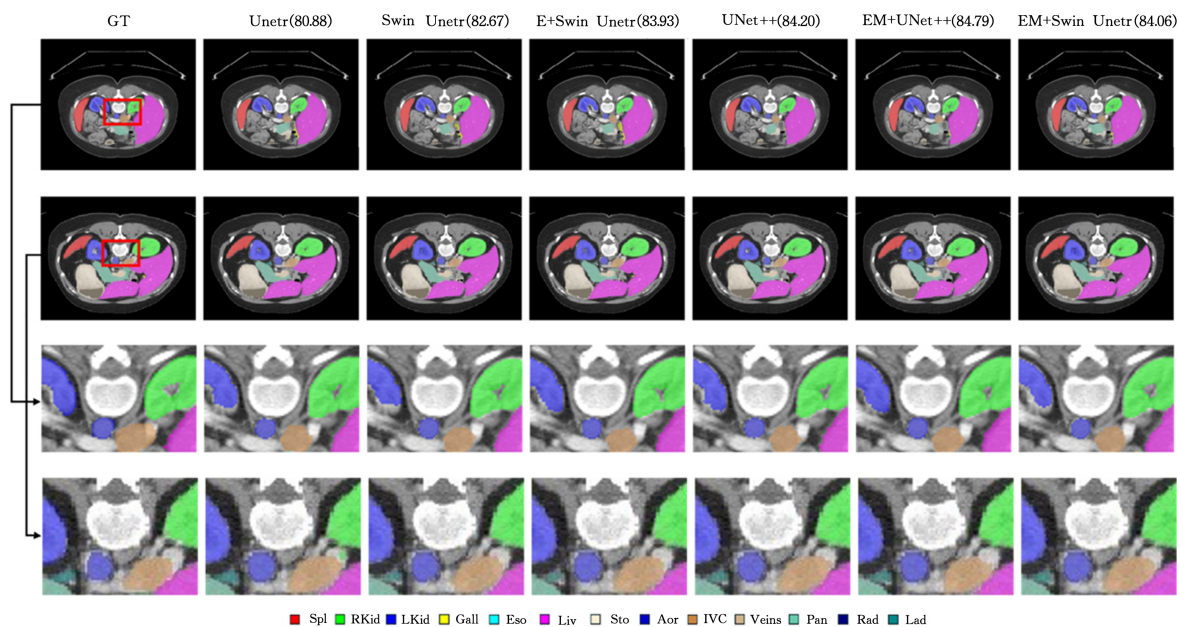


图 10 BTCV 测试数据上不同模型的效果比较

Fig. 10 Comparison of the effects of different models on BTCV test data

β 的影响如表 5 所列,在 BTCV 数据集上使用 Swin Unetr 作为基准模型,当 β 逐渐变大时,在平均和小型器官(包括类别 4,5,10,12,13)上 Dice 逐渐提高,当 β 超过 0.7 时,Dice 值逐渐下降。控制 β 不变,使用 Focal Loss 替代 4.2 节提到的改进的 CE Loss,Dice 在小比例器官上有一定的提高,但是整体没有达到实验预期。

表 5 不同 β 下所有器官的平均 Dice 以及高、中和小比例器官的平均 Dice

Table 5 Mean dice for all organs and mean Dice for organs of high, median and small proportions in different β

β	average	high	median	small
0	81.99	90.10	87.73	71.38
0.4	81.78	90.61	88.10	70.17
0.5	82.41	90.32	87.98	72.10
0.55	82.19	90.72	88.03	71.24
0.6	83.08	90.35	88.48	73.33
0.65	82.97	90.42	88.31	72.39
0.7	82.20	90.92	88.18	70.99
focal loss	82.12	88.95	87.28	72.87

结束语 腹部 CT 多器官分割任务在现代化医疗诊断、外科手术等领域发挥着重要作用。当前基于 U-Net 的分割模型仍未很好地解决小型器官分割精度低的问题,在这方面并未达到理想的医学应用的标准。因此,本文引入了 MAB 模块,利用先验信息和掩码信息自适应地调整窗口和块大小,以增强对体素比例低的小型器官的细节特征提取能力。同时提出了边缘约束模块,共享参数进行边缘预测和分割预测,然后利用边缘信息约束语义分割的预测。

本文在 BTCV 和 TCIA Pancreas-CT 腹部器官的 CT 数据集上验证了所提方法的有效性。实验结果显示,在不同的网络上使用 MAB 模块和边缘约束模块,在参数量较少的同时能取得更好的分割效果,尤其是针对小型器官的分割。总的来说,本文方法展现出了在医学图像分割领域有效分割器

官体素比例失衡数据集的潜力。

基于 Transformer 架构的模型由于存在大量的参数和计算负载而不适合移动医疗应用。为了解决实际医疗环境中计算资源限制带来的挑战,需要实现模型的轻量化。未来的工作包括设计一个更轻量级的网络,并在更多的数据集上进行验证,促进它们在移动医疗及医学其他领域的应用。

参考文献

- [1] VAN G B,SCHAEFER-PROKOP C M,PROKOP M. Computer-aided diagnosis;how to move from the laboratory to the clinic [J]. Radiology,2011,261(3):719-732.
- [2] TONG T,WOLZ R,WANG Z H,et al. Discriminative dictionary learning for abdominal multi-organ segmentation [J]. Medical Image Analysis,2015,23(1):92-104.
- [3] OLIVEIRAB,QUEIRÓS S,MORAIS P,et al. A novel multi-atlas strategy with dense deformation field reconstruction for abdominal and thoracic multi-organ segmentation from computed tomography[J]. Medical Image Analysis,2018,45:108-120.
- [4] CERROLAZAJ J,VILLANUEVA A,CABEZA R. Hierarchical multi-resolution decomposition of statistical shape models[J]. Signal, Image and Video Processing,2015,9:1473-1490.
- [5] OKADA T,LINGURARU M G,HORI M,et al. Abdominal multi-organ segmentation from CT images using conditional shape-location and unsupervised intensity priors[J]. Medical Image Analysis,2015,26(1):1-18.
- [6] SINGH P,BOSE S S. A quantum-clustering optimization method for COVID-19 CT scan image segmentation[J]. Expert Systems with Applications,2021,185:115637.
- [7] BOZKURT F,KÖSE C,SARI A. An inverse approach for automatic segmentation of carotid and vertebral arteries in CTA[J]. Expert Systems with Applications,2018,93:358-375.

- [8] GIBSON E, GIGANTI F, HU Y P, et al. Automatic multi-organ segmentation on abdominal CT with dense V-networks [J]. *IEEE Transactions on Medical Imaging*, 2018, 37 (8): 1822-1834.
- [9] RONNEBERGERO, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C] // *Medical Image Computing and Computer-Assisted Intervention, Medical Image Computing and Computer Assisted Intervention Society*, 2015: 234-241.
- [10] HEINRICH M P, OKTAY O, BOUTELDJA N. OBELISK-Net: Fewer layers to solve 3D multi-organ segmentation with sparse deformable convolutions [J]. *Medical Image Analysis*, 2019, 54: 1-9.
- [11] ZHANG L, ZHANG J, SHEN P, et al. Block level skip connections across cascaded V-Net for multi-organ segmentation [J]. *IEEE Transactions on Medical Imaging*, 2020, 39 (9): 2782-2793.
- [12] FANG X, YAN P K. Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction [J]. *IEEE Transactions on Medical Imaging*, 2020, 39 (11): 3619-3629.
- [13] ISENSEE F, JAEGER P F, KOHL S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation [J]. *Nature Methods*, 2021, 18 (2): 203-211.
- [14] ZHANG J P, XIE Y T, WANG Y, et al. Inter-slice context residual learning for 3D medical image segmentation [J]. *IEEE Transactions on Medical Imaging*, 2020, 40 (2): 661-672.
- [15] SINHA A, DOLZ J. Multi-scale self-guided attention for medical image segmentation [J]. *IEEE Journal of Biomedical and Health Informatics*, 2020, 25 (1): 121-130.
- [16] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHS N, et al. U-net++: Redesigning skip connections to exploit multiscale features in image segmentation [J]. *IEEE Transactions on Medical Imaging*, 2019, 39 (6): 1856-1867.
- [17] JIN Q G, MENG Z P, SUN C M, et al. RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans [J]. *Frontiers in Bioengineering and Biotechnology*, 2020, 8: 605132.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS' 17)*. 2017: 6000-6010.
- [19] HATAMIZADEHA, TANG Y C, NATH V, et al. Unetr: Transformers for 3d medical image segmentation [C] // *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2022: 574-584.
- [20] HATAMIZADEH A, NATH V, TANG Y C, et al. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images [C] // *International MICCAI Brainlesion Workshop*. 2021: 272-284.
- [21] LEE S, LEE M. MetaSwin: a unified meta vision transformer model for medical image segmentation [J]. *PeerJ Computer Science*, 2024, 10: e1762.
- [22] WANG L C, HUANG J H, XING X D, et al. Hybrid Swin Deformable Attention U-Net for Medical Image Segmentation [C] // *2023 19th International Symp on Medical Information Processing and Analysis*. 2023: 1-5.
- [23] HEIDARI M, KAZEROONI A, SOLTANY M, et al. Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation [C] // *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2023: 6202-6212.
- [24] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 10012-10022.
- [25] IRSHAD S, GOMES D P S, KIM S T. Improved abdominal multi-organ segmentation via 3D boundary-constrained deep neural networks [J]. *IEEE Access*, 2023, 11: 35097-35110.
- [26] GIBSON E, GIGANTI F, HU Y P, et al. Automatic multi-organ segmentation on abdominal CT with dense V-networks [J]. *IEEE Transactions on Medical Imaging*, 2018, 37 (8): 1822-1834.



PENG Linna, born in 2001, master candidate. Her main research interests include image segmentation and machine learning.



ZHANG Hongyun, born in 1972, Ph.D., associate professor. Her main research interests include principal curve algorithm, granular computing and fuzzy sets.

(责任编辑:李亚辉)