

通过边缘引导的肾上腺三维 CT 影像分割

王文静¹, 牛四杰¹, 李帆², 曹彩霞³, 丛文斌³, 杨自成³

(1. 济南大学信息科学与工程学院, 山东 济南 250000)

(2. 广州柏视医疗科技有限公司, 广东 广州 510530)

(3. 青岛大学附属医院, 山东 青岛 266000)

[摘要] 计算机断层扫描图像是判断肾脏情况的主要成像方式. 医生可以通过分割出腹部 CT 图像中感兴趣的肾上腺区域, 从而计算出肾上腺的体积、灰度值和表面积来判断肾病的病因. 然而手工标记图像中的病变区域是耗时、繁琐且具有挑战性的, 且病变区域与周围组织极为相似, 勾画出的边界极为模糊. 因此本文采用一种全卷积神经网络模型 MedNeXt——一个受 Transformer 启发的大核分割网络来对肾上腺 3D 数据进行体积分割. 为应对样本类别不平衡问题, 本文还使用对称统一焦点损失替换 Dice 损失, 以提高分割精度. 同时考虑到肾上腺组织与周围组织边界难以区分的问题, 本文提出结合边界损失函数与主体损失函数同时监督分割过程, 使得模型更关注边界的细节信息, 从而提升模型性能, 实现更精确的分割结果. 实验结果表明, 所用方法与近几年最新的模型相比在本文所用肾上腺 3D 数据集上实现了最先进的性能.

[关键词] 全卷积, Transformer, MedNeXt, 类别不平衡, 体积分割

[中图分类号] TP391 **[文献标志码]** A **[文章编号]** 1001-4616(2025)01-0093-07

Edge Guided 3D CT Image Segmentation of Adrenal Gland

Wang Wenjing¹, Niu Sijie¹, Li Fan², Cao Caixia³, Cong Wenbin³, Yang Zicheng³

(1. College of Information Science and Engineering, University of Jinan, Jinan 250000, China)

(2. Perception Vision Medical Technologies Co., Ltd., Guangzhou 510530, China)

(3. The Affiliated Hospital of Qingdao University, Qingdao 266000, China)

Abstract: Computed tomography image is the main imaging method to judge the condition of kidney. Doctors can determine the cause of kidney disease by segmenting the adrenal region of interest in the abdominal CT image and calculating the volume, gray value and surface area of the adrenal gland. However, it is time-consuming, tedious and challenging to manually mark the lesion area in the image, and the lesion area is very similar to the surrounding tissue, and the boundary outlined is extremely fuzzy. Therefore, the method adopted in this paper uses a full convolution neural network model MedNeXt — a transformer inspired large core segmentation network to perform volume segmentation on 3D adrenal data. In order to deal with the problem of unbalanced sample categories, this paper also uses symmetrical unified focus loss to replace Dice loss to improve segmentation accuracy. At the same time, considering the problem that it is difficult to distinguish between adrenal tissue and surrounding tissue boundaries, this paper proposes to combine the boundary loss function and the main body loss function to simultaneously monitor the segmentation process, so that the model pays more attention to the details of the boundary, thus improving the model performance and achieving more accurate segmentation results. Finally, experiments show that the method used in this paper achieves the most advanced performance on the adrenal 3D dataset compared with the latest models in recent years.

Key words: full convolution, Transformer, MedNeXt, unbalanced sample categories, volume segmentation

在临床实践中, 医学图像的精确语义分割起着重要作用. 自动化的器官分割, 如心脏、肺部、肾脏等的分割, 使得医生能够摆脱枯燥的标注工作, 从而提高诊断和治疗的效率. 肾脏是人体重要器官, 其功能包括过滤血液中的杂质、维持体液的平衡, 以及产生尿液并调节血压. 正常成人具有两枚肾脏, 位于腰部两侧后方. 肾脏疾病的出现可能导致肾功能丧失, 其中包括原发性肾病和继发性肾病. 原发性肾病由肾上

收稿日期: 2024-06-07.

基金项目: 国家自然科学基金项目(62101213、62103165、62302191)、山东省高等学校人才引育创新团队项目.

通讯作者: 牛四杰, 博士, 教授, 研究方向: 模式识别, 医学影像分析. E-mail: sjniu@hotmail.com

腺组织的病变引起,如肾小球肾炎;而继发性肾病是由其他疾病影响肾上腺功能,例如糖尿病和高血压,导致肾上腺病变,如增生和腺瘤。所以肾上腺区域的自动分割对于临床诊断来说是极其重要的。

随着深度学习的发展^[1],基于神经网络的方法逐渐成为医学图像分割的主流技术,这些工作主要基于卷积神经网络(CNN)。在这些工作中,UNet^[2]已经成为医学图像分割的主流框架,其编码器部分将图像压缩到潜在特征空间中,该空间携带更多图像的编码信息。解码器部分通过扩展潜在特征并通过跳跃连接利用来自编码器的多级特征图来学习并找到感兴趣区域(ROI)^[3]。此后出现了许多 UNet 变体,如 UNet++^[4],VNet3D^[5],nnUNet^[6]。然而,由于 CNN 受卷积核的限制,无法提取远程依赖。因此,许多研究尝试在基于 CNN 的模型中引入注意力机制,尽管这改善了模型的性能,但仅仅通过这些生成的全局上下文是不够的。随后,Transformer^[7-9]的出现使得提取全局和远程依赖成为可能,ViT^[8]在 NLP 领域脱颖而出,并在医学图像分割领域取得了巨大的成功,进一步提高了模型的性能,但是 Transformer 在捕获局部细节方面存在局限,很难关注到边缘部分。因此本文使用一种在计算机断层扫描(CT)图像中界定肾脏的自动方法,利用图像处理技术和深度卷积神经网络来最大限度地减少误报,即全 ConvNeXt^[10] 3D 分割网络 MedNeXt^[11],一个最大限度地发挥 ConvNeXt 设计的潜力,同时独特地解决医学图像分割中有限数据集的挑战的模型,也是一个可以缩放的编码器-解码器网络。为了进一步增强 MedNeXt 对边界部分的分割能力,引入了一个新的边界损失函数来辅助监督网络进行分割。通过这种方法,本文可以确保特征图包含更丰富的信息,从而改善分割结果的准确性和精细化程度。

1 相关工作

近年来,基于卷积神经网络的方法在医学图像分割领域备受关注。UNet^[2]模型在多个医学图像分割任务上展现出令人期待的性能,引领了其他 UNet 变体的涌现,如 UNet++^[4]、Res-UNet^[12]、Attention UNet^[13]等。UNet++继承了 UNet 的结构,同时借鉴了 DenseNet^[14]的稠密连接方式,通过各层之间的紧密连接,每一层都尽量多的保存细节信息和全局信息并共享给最后一层,实现全局信息和局部信息的保留与重构。Attention Unet 的中心思想就是提出 Attention gate 模块,使用 soft-attention 替代 hard-attention,将 attention 集成到 Unet 的跳跃连接和上采样模块中,实现空间上的注意力机制。通过注意力机制来抑制图像的无关信息,突出局部的重要特征。

Transformer 最近在自然语言处理(NLP)领域取得的成功促进了视觉 Transformer 的发展,以克服 CNN 在图像分割任务中的限制,众多结合 Transformer 的医学分割网络也相继提出,包括 ViT^[8]、UNETR^[15]、TransUNet^[16]等。UNETR 利用纯 Transformers 作为编码器来学习输入量的序列表示并有效地捕获全局多尺度信息。同时也遵循了编码器和解码器成功的“U 型”网络设计,Transformers 编码器通过不同分辨率的跳跃连接直接连接到解码器,以计算最终的语义分段输出。TransUNet 采用 Transformer 以逐切片的方式处理每个 3D 医学图像,借助自注意力操作捕捉全局信息,将这部分信息上采样之后与高分辨率特征图融合,有效的提高了分割任务的效果,实现精准定位。然而,这些模型虽然具有优秀的表现,但它们在精细医学图像分割方面面临巨大挑战,对于边界模糊和不连续区域的分割性能较差。

为了加强边缘引导,Zhang 等^[17]提出了 Edge-Attention Network 网络模型,利用边缘引导模块来学习早期编码层中的边缘注意力表示,然后将其转移到多尺度解码层,使用加权聚合模块进行融合,提高了分割的精确性。Yang 等^[18]提出 EPT-Net,其使用边缘权重引导模块,通过最小化边缘信息函数来提取边缘信息,而不添加网络参数。针对分割不连续区域,Valanaras 等^[19]提出一个双分支网络 KiU-Net,能够更精细的学习边缘细节。对于边界模糊问题,Lee 等^[20]提出了一个最适合目标区域的边界关键点选择算法,使得边界模糊的区域得到了更好的分割。Manzar 等^[21]提出一个新的 U 型网络 BEFU-net,其包含双分支编码器:边缘编码器和主体编码器,增强了整体和边缘信息的融合。他们还使用边界损失和整体损失函数共同监督网络,以实现精确的医学图像分割。上述模型虽然在医学影像分割领域取得了不错的成绩,但也存在一定的不足。许多方法是为了特定任务所建立或在单个数据集上训练,缺少在各种数据集上表现良好的通用的模型。当遇到其他任务时,这些模型需要重新训练且不确定他们的表现是否稳定。

受到 ChatGPT 等大型语言模型的启发,研究人员开发了类似的大规模视觉模型(LVM)。这些模型展现出优秀的零样本和少样本泛化能力,可以通过预训练和微调来快速适应和扩展到目标任务或领域。其

中 CLIP^[22] 提供了一个统一的视觉和语言模型,适用于多种任务,如分类、检测和视觉问答. 该模型通过广泛的文本和图像进行预训练,在多个基准测试中展现了令人印象深刻的表现. 最近,经过预训练的 SAM^[23] 已经成为用于图像分割的 LVM. SAM 展示了强大的零样本功能,可以对任何对象进行交互式或自动分割. 然而,自然图像和医学图像之间存在显着的领域差异,且医学图像数据集较为稀少,将 SAM 应用于医学图像分割是极为困难的.

所以本文选用了全卷积神经网络并加以边缘引导来进行数据的分割,能够有效地结合卷积神经网络强大的特征提取能力和边缘引导的信息,进一步提高分割的精度和准确性.

2 方法

本节主要介绍本文所用的模型以及所使用的损失函数,整体方法如图 1 所示.

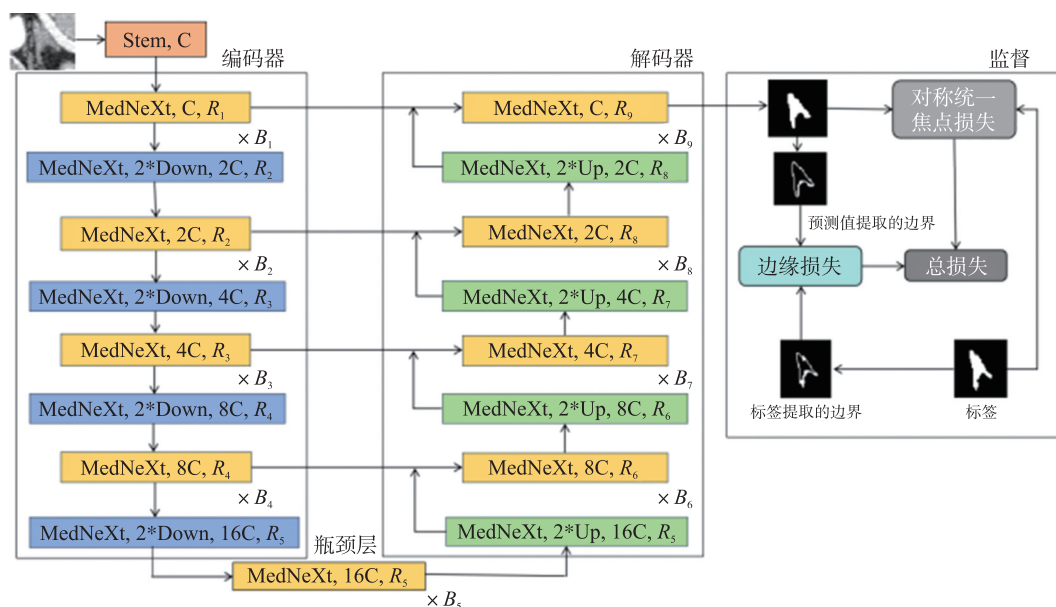


图 1 本文方法整体流程图

Fig. 1 Overall flow chart of our method

2.1 模型架构

本文的任务是对肾上腺 3D CT 图像进行分割,分割出感兴趣的肾上腺区域,从而能够根据分割出来的数据计算出病人肾部的灰度值、体积和表面积等,判断出病人的肾病类型. 为了在节省计算成本的同时得到一个较为精确的分割结果,本文使用的模型是 Roy 等^[11] 所提出的 MedNeXt 网络,一个受 Transformer 启发的大核全卷积分割网络,主要包括四部分:(1)用于医学图像分割的完全 ConvNeXt 3D 编码器-解码器网络,(2)残差 ConvNeXt 上采样和下采样块,以保持跨尺度的语义丰富性,(3)一种通过对小核网络进行上采样来迭代增加内核大小的新技术,防止在有限医学数据集上性能饱和,(4)MedNeXt 多个级别(深度、宽度、内核大小)的复合缩放^[12]. 该模型能够以全卷积的架构取得超出 CNN 和 Transformer 混合架构的性能结果,且没有利用迁移学习或集成学习等技术,在医学分割领域占据了重要地位.

2.2 损失函数

由于 Dice 损失梯度本质上是不稳定的,这在梯度计算涉及小分母的高度类别不平衡数据中最为明显. 所以本文不直接使用原文中标准的 Dice 损失,改用了 Yeung 等^[24] 所提出的基于 Dice 和交叉熵的损失的对称统一焦点损失,该损失函数可以使得模型关注于更难学习的类别,缓解了类别不平衡问题,加快了模型的收敛速度,提高了模型性能. 为了提高模型的边界分割能力,本文进一步提出了结合边界损失函数,引导模型更加关注边缘信息. 下面来详细介绍这两个损失函数.

2.2.1 对称统一焦点损失

混合焦点损失采用基于 Dice 和交叉熵的损失来处理类别不平衡,但因其超参数过多和焦点参数引入的增强或抑制效果适用于所有类别等问题, Yang 等^[18] 提出了对称统一焦点损失函数,解决了混合焦点损

失的问题. 公式如下:

$$L_{Bce} = - \sum_{i=0}^N [(1-y'_i) \log(1-y_i) + y'_i \log y_i], \quad (1)$$

$$L_{mF(p_t)} = \delta(1-p_t)^{1-\gamma} \cdot L_{BCE(p,y)}, \quad (2)$$

$$L_{mFT} = \sum_{c=1}^C (1-mTI)^\gamma, \quad (3)$$

其中 y 和 $y' \in \{0,1\}^N$, 且 y' 为预测值, y 为真值标签.

p_t 为预测为真值的概率, γ 可以同时实现焦点损失抑制和焦点 Tversky 损失增强, δ 参数控制输出不平衡.

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1-p, & \text{if } y = 0 \end{cases}, \quad (4)$$

$$mTI = \frac{\sum_{i=1}^N p_{0i} g_{0i}}{\sum_{i=1}^N p_{0i} g_{0i} + \delta \sum_{i=1}^N p_{0i} g_{1i} + (1-\delta) \sum_{i=1}^N p_{1i} g_{0i}}, \quad (5)$$

其中 p_{0i} 是像素 i 属于前景类的概率, p_{1i} 是像素属于背景类的概率. g_{0i} 为 1 时, 表示前景, 为 0 时表示背景; 相反, g_{1i} 的值为 1 表示背景, 0 表示前景.

$$L_{sUF} = \lambda L_{mF} + (1-\lambda) L_{mFT}, \quad (6)$$

其中 $\lambda \in [0,1]$, 用来确定两个损失的相对权重大小. 本文中使用的 λ 为 0.5, δ 为 0.6, γ 为 0.5.

2.2.2 边界损失

在医学分割领域中, 准确捕捉目标物体的边界信息对于精细分割至关重要. 为了进一步提升分割结果的精细程度, 本文提出添加边界损失以辅助监督模型的分割过程. 所提出的边界损失包括交叉熵损失和 Dice 损失. 交叉熵损失用于度量像素级预测误差, 通过对每个像素的分类结果进行评估, 鼓励网络在像素级别进行准确的分类预测, 有助于提高分割的精度. 在医学图像分割任务中, 不同类别之间的像素数量可能存在明显的不平衡. Dice 损失函数能够更好地处理类别不平衡情况, 因为它是基于像素级别的相似度指标, 而非简单的分类损失函数. 公式描述如下:

$$Dice = \frac{2|X \cdot Y|}{|X| + |Y|}, \quad (7)$$

其中, X 为真值标签, Y 是预测值.

$$L_{Dice} = 1 - Dice. \quad (8)$$

边界损失公式描述如下:

$$L_{edge} = (1 - L_{Dice}) + L_{Bce}, \quad (9)$$

其中, L_{Bce} 为交叉熵损失, 如式(1)中定义.

2.2.3 总体损失

最终的损失函数为:

$$L = \alpha L_{sUF} + (1-\alpha) L_{edge}, \quad (10)$$

其中, α 表示权重系数.

3 实验

3.1 数据集与数据预处理

本文所采用的数据集由青岛大学提供, 共包含 60 例临床患者的 CT 影像和医生勾画结果. 图像空间大小为 512×512 , 层数在 47~836 间. 在本文的医学影像数据集中, 大多数病例的 z 轴间距为 1.25 mm. 然而, 也存在少数例外情况, 间距介于 0.6~5.0 mm. x 轴和 y 轴间距介于 0.6~1.0 mm.

本文对数据和标签进行了填充、重采样间距、尺寸和窗宽窗位处理等操作. 具体来说, 首先读取标签, 然后对每个切片进行填充空洞处理. 之后获取了原始体数据的间距和大小信息, 然后根据指定的输出间

距 $[1.0, 1.0, 1.0]$, 计算出新的输出尺寸. 最后计算窗口的最小值和最大值. 如果选择了归一化, 则将体数据数组归一化到 $[0, 1]$ 范围, 并将超出窗口范围的像素值截断到窗口范围内. 如果未选择归一化, 则将低于窗口最小值的像素值设为 0. 保存处理后的数据和标签. 由于重采样后的三维医学影像数据维度过大, 受限于显卡显存大小, 无法直接送入模型进行训练. 因此, 本文在训练阶段采用图像随机裁剪, 裁剪大小为 $(96, 96, 96)$. 为了避免随机裁剪引入过多的与感兴趣区无关的背景信息, 从而影响模型的性能, 本文找到体数据中每个坐标轴第一个和最后一个不包含背景的索引对数据和标签进一步裁剪. 在测试阶段, 本文采用滑动窗口策略, 其中窗口大小等于训练补丁的大小.

3.2 实验设置

MedNeXt 模型按缩放块数和扩展比分为四种配置, 如表 1 所示, 本文所采用的 MedNeXt 是最简化配置 (MedNeXt-S), 模型基于 Pytorch 深度学习框架, 采用端到端的训练方式. 本文对所有实验的训练集采用 5 折交叉验证策略, 从每个折中选择最终模型, 并简单地平均它们的输出以得出最终分割预测. 所有实验均在 NVIDIA GeForce RTX 3090 GPU 上进行. 为了提高计算速度, 所有训练和测试的数据分辨率均裁剪为 $96 \times 96 \times 96$. 在实验过程中调整了许多超参数, 例如学习率、裁剪的数据大小、数据中所包含的病变区域像素数以及滑动窗口大小等. 传统 Adam 优化器的权重衰减可能会导致过度的参数更新, 而 AdamW 通过独立处理权重衰减, 有助于减少过拟合和提高泛化能力. 因此本文优化器使用 AdamW 优化器, 学习率设置为 0.000 1, 权重衰减为 $1e-4$. AdamW 优化器通过对权重衰减的处理, 可以提供更稳定的训练过程. 为了节省计算资源, 本文将 batch 大小设置为 1, 并将对称统一焦点损失的权重 α 设置为 0.5.

表 1 通过缩放块计数(B)和扩展比(R)得到的 MedNeXt 配置

结构	块数(B)	扩展比(R)
S	$B_{all} = 2$	$R_{all} = 2$
B		$R_1 = R_9 = 2$
M	$B_1 = B_9 = 3$	$R_2 = R_8 = 3$
	$B_{2-8} = 4$	$R_{3-7} = 4$
L	$B_1 = B_9 = 3$	$R_1 = R_9 = 3$
	$B_2 = B_8 = 4$	$R_2 = R_8 = 4$
	$B_{3-7} = 8$	$R_{3-7} = 8$

域像素数以及滑动窗口大小等. 传统 Adam 优化器的权重衰减可能会导致过度的参数更新, 而 AdamW 通过独立处理权重衰减, 有助于减少过拟合和提高泛化能力. 因此本文优化器使用 AdamW 优化器, 学习率设置为 0.000 1, 权重衰减为 $1e-4$. AdamW 优化器通过对权重衰减的处理, 可以提供更稳定的训练过程. 为了节省计算资源, 本文将 batch 大小设置为 1, 并将对称统一焦点损失的权重 α 设置为 0.5.

3.3 评价指标

Dice 相似系数 (Dice) 是通过衡量分割预测和真值的重叠区域来评估分割性能. 所以本文使用 Dice score 来评估分割肾上腺的性能.

$$Dice = \frac{2 \times TP}{|TP + FN| + |TP + FP|}, \quad (11)$$

其中 TP 是正确预测的正样本的数量. TN 、 FP 、 FN 分别表示准确预测的负样本数、预测为正数的负样本数、预测为负数的正样本数.

Hausdorff_95 (豪斯多夫距离) 用于计算两个集合之间的距离, 值越小, 代表两个集合的相似度越高. 主要用来度量边界的分割准确度.

$$HD95 = \max \left\{ \max_{y \in Y} \min_{y' \in Y'} \|y - y'\|, \max_{y' \in Y'} \min_{y \in Y} \|y - y'\| \right\}, \quad (12)$$

其中, $\|$ 表示欧几里德距离函数, y 和 y' 分别表示来自真实值 Y 和预测分割 Y' 的各个体素.

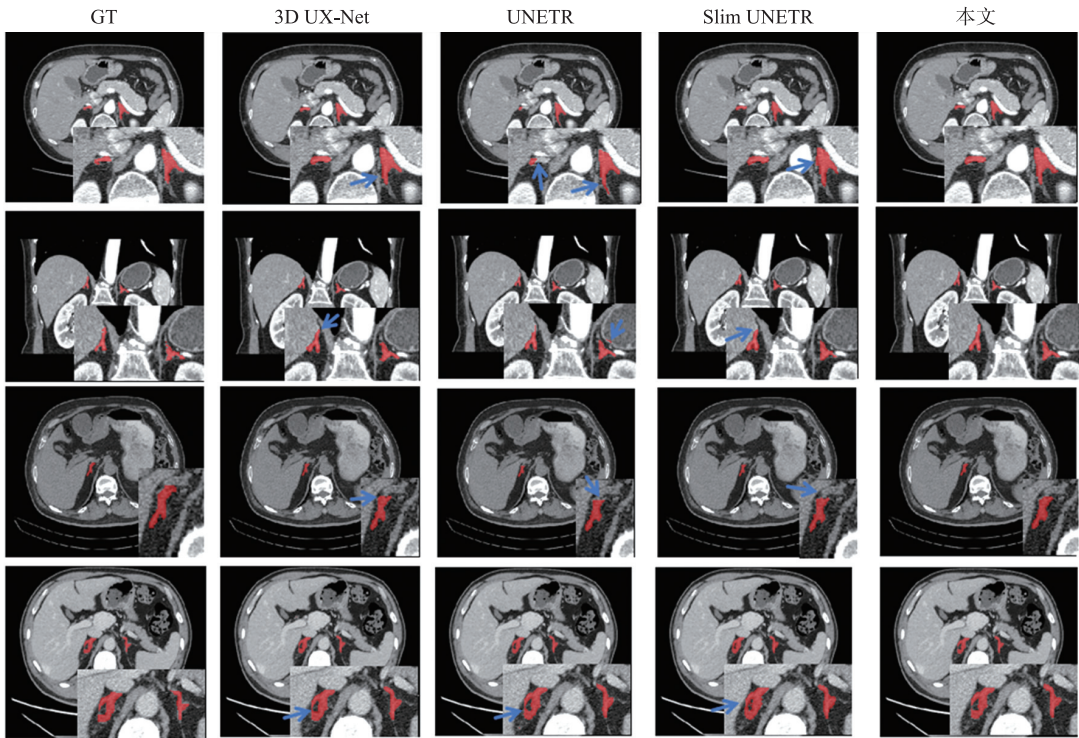
3.4 实验结果

3.4.1 对比实验

将本文所用方法与 3D UX-Net^[14]、UNETR^[16]、TransBTS^[17] 以及 Slim UNETR^[25] 等先进医学图像分割模型在本文的数据集上进行实验比较. 此外, 本文还对 TransBTS 和 Slim UNETR 的原损失函数进行替换, 改为对称统一焦点损失进行实验. 实验结果如表 2 所示. 从结果可以看出, 本文的方法在所有评价指标上均优于现有模型. 相较于替换了损失函数的 Slim UNETR, 本文的方法在 Dice 得分上提高约 2.6%, HD 95 降低了 30.86. 所有方法在数据集上的分割对比结果如图 2 所示. 由于边界的引导作用, 本文所采用的方法在处理细节分割方面表现出色, 模型能够有效地捕捉分割目标的整体特征.

表 2 模型五重交叉实验结果

方法	Avg Dice \uparrow	Avg HD 95 \downarrow
3D UX-Net	0.594	194.426
UNETR	0.673	156.344
Slim UNETR	0.721	105.694
TransBTS+对称统一焦点损失	0.716	87.694
Slim UNETR+对称统一焦点损失	0.727	75.100
本文	0.753	44.236



红色部分标识了分割出的肾上腺区域;蓝色箭头为分割误差区域.

图 2 所有方法在数据集上的分割结果

Fig. 2 Segmentation results of all methods on the data set

3.4.2 消融实验

在这一部分,本文研究了对称统一焦点损失和边界损失函数对模型性能的影响,具体结果如表 3 所示. 首先使用 MedNeXt 模型与原损失函数(Dice Loss)进行实验. 随后,本文将原损失函数替换为对称统一焦点损失. 实验结果显示,对称统一焦点损失显著提升了模型的性能,使得 Dice 得分提高了 2.4%. 最后本文引入边界损失函数,值得注意的是,边界损失的结合带来了实质性的性能改进,与仅利用对称统一焦点损失的方法相比,Dice 得分显著提高了约 2.7%,这表明边界损失有助于模型有效地学习边界信息.

表 3 消融实验结果

Table 3 Ablation results

方法	Avg Dice \uparrow	Avg HD 95 \downarrow
MedNeXt	0.702	89.802
MedNeXt+对称统一焦点损失	0.726	58.566
本文	0.753	44.236

4 结论

本文提出了一种新的 3D 肾上腺体积分割方法,使用全卷积神经网络 MedNeXt 作为主干网络,并将原损失函数替换为对称统一焦点损失和边界损失的组合. 卷积操作有助于编码精确的空间信息,从而提高了模型的分割性能. 边缘辅助监督进一步增强了模型对边界的分割能力,有助于为医生或专家在判断肾病类型时提供辅助信息,使患者能够更快获得诊断和治疗. 实验结果表明,本文所提出的方法在 HD 95 和 DSC 方面相较先进模型具有更大优势. 未来将进一步优化所使用的方法.

[参考文献]

[1] GOODFELLOW I,BENGIO Y,COURVILLE A,et al. Deep Learning[M]. Cambridge,UK:MIT Press,2016.

[2] RONNEBERGER O,FISCHER P,BROX T. U-net:convolutional networks for biomedical image segmentation[C]//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany: Springer, 2015:234-241.

[3] LI B,LIU S K,WU F,et al. RT-Unet: an advanced network based on residual network and transformer for medical image

- segmentation[J]. International journal of intelligent systems,2022,37(11):8565–8582.
- [4] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHS 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. Granada, Spain: Springer International Publishing, 2018:3–11.
 - [5] MILLETARI F, NAVAB N, AHMADI S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C]//Proceedings of the 2016 Fourth International Conference on 3D Vision(3DV), Stanford, CA: IEEE, 2016: 565–571.
 - [6] ISENSEE F, PETERSEN J, KLEIN A, et al. NNU-Net: self-adapting framework for U-Net-based medical image segmentation [J/OL]. arXiv Preprint arXiv:1809.10486, 2018.
 - [7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017:5998–6008.
 - [8] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[J/OL]. arXiv Preprint arXiv:2010.11929, 2020.
 - [9] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE, 2021:10012–10022.
 - [10] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA: IEEE, 2022:11976–11986.
 - [11] ROY S, KOEHLER G, ULRICH C, et al. Mednext: transformer-driven scaling of convnets for medical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Vancouver, Canada: Springer, 2023: 405–415.
 - [12] XIAO X, LIAN S, LUO Z, et al. Weighted res-unet for high-quality retina vessel segmentation[C]//International Conference on Information Technology in Medicine and Education(ITME). Hangzhou, China: IEEE, 2018:327–331.
 - [13] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: learning where to look for the pancreas[C]//Proceedings of the 21th International Conference Medical Image Computing and Computer Assisted Intervention (MICCAI). Granada, Spain: Springer International Publishing, 2018:564–572.
 - [14] 蒋婷, 李晓宁. 采用多尺度视觉注意力分割腹部 CT 和心脏 MR 图像[J]. 中国图像图形学报, 2024, 29(1):268–279.
 - [15] HATAMIZADEH A, TANG Y, NATH V, et al. Unetr: transformers for 3d medical image segmentation[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, HI: IEEE, 2022:574–584.
 - [16] CHEN J, LU Y, YU Q, et al. Transunet: transformers make strong encoders for medical image segmentation[J/OL]. arXiv Preprint arXiv:2102.04306, 2021.
 - [17] ZHANG Z, FU H, DAI H, et al. Et-net: a generic edge-attention guidance network for medical image segmentation[C]//Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, Shenzhen, China: Springer International Publishing, 2019:442–450.
 - [18] YANG J, JIAO L, SHANG R, et al. Ept-net: edge perception transformer for 3d medical image segmentation[J]. IEEE transactions on mmedical imaging, 2023, 42(11):3229–3243.
 - [19] VALANARAS J M J, SINDAGI V A, HACIHALILOGL I, et al. Kiu-net: overcomplete convolutional architectures for biomedical image and volumetric segmentation[J]. IEEE transactions on medical imaging, 2021, 41(4):965–976.
 - [20] LEE H J, KIM J U, LEE S, et al. Structure boundary preserving segmentation for medical image with ambiguous boundary [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020:4817–4826.
 - [21] MANZAR O N, KALEYBAR J M, SAADAT H, et al. BEFUnet: a hybrid CNN–Transformer architecture for precise medical image segmentation[J/OL]. arXiv Preprint arXiv:2402.08793, 2024.
 - [22] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//International Conference on Machine Learning. San Francisco, CA, USA: PMLR, 2021:8748–8763.
 - [23] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE, 2023:4015–4026.
 - [24] YEUNG M, SALA E, SCHÖNLIEB C B, et al. Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation[J]. Computerized medical imaging and graphics, 2022, 95:102026.
 - [25] PANG Y, LIANG J, HUANG T, et al. Slim UNETR: scale hybrid transformers to efficient 3D medical image segmentation under limited computational resources[J]. IEEE transactions on medical imaging, 2024, 43(3):994–1005.