

DOI:10.3969/j.issn.1005-202X.2025.06.008

医学影像物理

肾小球超微结构的半监督语义分割方法

陈翔^{1,2,3}, 张桢泰^{1,2,3}, 龙楷兴^{1,2,3}, 路艳蒙⁴, 耿舰^{5,6}, 周志涛⁴, 曹蕾^{1,2,3}

1. 南方医科大学生物医学工程学院, 广东 广州 510515; 2. 广东省医学图像处理重点实验室, 广东 广州 510515; 3. 广东省医学成像与诊断技术工程实验室, 广东 广州 510515; 4. 南方医科大学中心实验室, 广东 广州 510515; 5. 南方医科大学基础医学院, 广东 广州 510515; 6. 广州华银医学检验中心, 广东 广州 510515

【摘要】肾小球超微结构的精准识别对慢性肾脏病诊断至关重要,但高质量标注数据的获取成本限制了全监督学习的应用。为此,提出一种基于SAM(Segment Anything Model)的多类别半监督语义分割框架MC4S-SAM。首先,对SAM的掩码解码器进行改进,使其在无需提示信息的情况下具备多类别语义分割能力;然后,利用改进后的模型通过自训练(Self-training)策略生成并优化伪标签;最后,构建多级一致性正则化约束提升模型性能。实验结果表明,在肾小球系膜区超微结构分割任务中,使用DeepLabV3+作为分割网络,MC4S-SAM在标注数据量占比为1/16时,平均交并比(mIoU)和平均Dice系数(mDSC)分别比全监督模型提高11.72%和11.45%;在标注数据量占比为1/4时,其mIoU和mDSC分别达到68.91%和78.73%,为慢性肾脏病的辅助诊断提供重要的应用价值。

【关键词】医学图像分割;半监督学习;一致性正则化;肾小球超微结构

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2025)06-0757-09

Semi-supervised semantic segmentation method for glomerular ultrastructure

CHEN Xiang^{1,2,3}, ZHANG Zhentai^{1,2,3}, LONG Kaixing^{1,2,3}, LU Yanmeng⁴, GENG Jian^{5,6}, ZHOU Zhitao⁴, CAO Lei^{1,2,3}

1. School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China; 2. Guangdong Provincial Key Laboratory of Medical Image Processing, Guangzhou 510515, China; 3. Guangdong Provincial Engineering Laboratory for Medical Imaging and Diagnostic Technology, Guangzhou 510515, China; 4. Central Laboratory, Southern Medical University, Guangzhou 510515, China; 5. School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China; 6. Guangzhou Huayin Medical Laboratory Center, Guangzhou 510515, China

Abstract: Accurate identification of the glomerular ultrastructure is critical for the diagnosis of chronic kidney diseases, but the high cost of acquiring high-quality annotated data limits the application of fully-supervised learning. Therefore, a multi-class semi-supervised semantic segmentation framework based on segment anything model (MC4S-SAM) is proposed. After improving the mask decoder of segment anything model to enable multi-class semantic segmentation without requiring prompt information, the improved model is used to generate and refine pseudo-labels through a self-training strategy, and multi-level consistency regularization constraints are incorporated to enhance the model's performance. Experimental results show that, in the task of segmenting the glomerular mesangial ultrastructure, MC4S-SAM outperforms the fully-supervised model by 11.72% in mean intersection over union (mIoU) and 11.45% in mean Dice similarity coefficient (mDSC) when the labeled data accounts for 1/16 of the total. When the labeled data proportion is 1/4, the mIoU and mDSC reach 68.91% and 78.73%, respectively, demonstrating its significant potential for aiding the diagnosis of chronic kidney diseases.

Keywords: medical image segmentation; semi-supervised learning; consistency regularization; glomerular ultrastructure

前言

慢性肾脏病(Chronic Kidney Disease, CKD)正逐

渐成为全球范围内最常见的非传染性慢性疾病之一。据2017年的研究数据显示^[1],全球约有6.975亿人患有CKD,占全球总人口的9.1%。其中,中国的CKD患者数量位居世界首位。根据世界卫生组织的预测,到2040年,CKD将跃升为全球第五大常见慢性疾病^[2]。肾小球超微结构的识别在肾脏疾病的精准诊断中占据着重要地位^[3]。借助透射电子显微镜,病理医生能够在纳米尺度下观察到肾小球超微结构相

【收稿日期】2025-01-18

【基金项目】国家自然科学基金(32071368)

【作者简介】陈翔,硕士研究生,研究方向:医学图像分割, E-mail: qluchenxiang@163.com

【通信作者】曹蕾,博士,副教授,研究方向:医学图像分析, E-mail: caolei@smu.edu.cn

关病理改变,从而为诊断提供关键依据^[4]。目前,已有研究将深度学习技术成功应用于肾小球超微结构的精确分割领域,例如,对基底膜、滤过屏障和电子致密物的分割研究^[5-6]。相比之下,针对系膜区超微结构的分割研究仍显不足,而许多肾脏疾病的诊断,如IgA肾病、狼疮性肾炎和系膜增生性肾小球肾炎,仍依赖于对肾小球系膜区特征性病理变化的识别^[3-4]。因此,深入研究肾小球系膜区的超微结构分割对推动该领域发展和提升诊断精度具有重要的意义。

医学图像的特殊性和复杂性使得获取肾小球系膜区超微结构的高质量分割标注数据成本高昂,进而导致了标注数据的稀缺。这种状况严重限制了全监督深度学习模型在提升性能方面的进展^[7-9]。近年来,针对标注数据稀缺的挑战,半监督学习策略引起了广泛关注和研究^[10-11]。目前半监督学习方法有两个重要分支^[12]:自训练方法和一致性正则化方法。自训练方法^[13]利用在少量标记数据上训练的模型,对大量未标记数据进行预测生成伪标签,并将这些伪标签作为补充的监督信号加入训练过程,以此来逐步提升模型性能。其中,代表性方法有ST++^[14]等。一致性正则化方法^[15]通过增强模型对扰动的鲁棒性,提高模型的学习能力和泛化性能。此类方法的代表包括:提升模型对数据扰动鲁棒性的图像级正则化方法^[16-19]、在图像级正则化基础上加入特征级正则化的混合方法^[20]、融合多模型输出伪标签的模型级正则化方法^[21-22]。

近期,研究者将自训练方法与一致性正则化方法相结合,充分发挥了两者的优势。自训练方法能够提高用于一致性正则化的伪标签质量,而一致性正则化则有助于减少自训练过程中错误的累积。通过将这两种方法结合,能够更有效地利用未标注数据,显著提升模型的性能。Wang等^[23]提出的STCRNet网络通过在自训练过程中引入不同强弱的扰动操作,进行一致性正则化,有效提升变化检测任务的表现;Li等^[24]通过在自训练框架中引入Mean Teacher方法,实现医学图像分割任务的显著性能提升。然而,在这些方法中,由于模型性能不足,导致在训练中产生的伪标签质量较低,由此引入的大量噪声限制了模型性能的提升^[25]。

基础模型SAM(Segment Anything Model)^[26]的出现为提高伪标签的质量提供新的可能。Zhang等^[27]和Li等^[28]在半监督模型基础上直接利用SAM生成的伪标签作为补充监督信息进行左心房分割任务;Chen等^[29]使用SAM微调和自训练方法用于肝脏肿瘤分割。然而,由于自然图像与医学图像之间存

在固有的领域差异,直接应用SAM有时难以保证所生成伪标签的质量。此外,传统的SAM分割需要提示信息才能生成特定类别的分割掩码,在缺乏明确先验信息的情况下会受到限制;而在无提示信息时,SAM生成的伪标签又缺乏语义信息,仅能提供无类别信息的分割标签,不适用于多类别语义分割任务。因此,为了实现肾小球系膜区多类超微结构的自动分割,本文将SAM改进为多类别语义分割SAM(Multi-Class Semantic Segmentation SAM, MCSAM),并将自训练和一致性正则化方法相结合,提出了一种基于SAM的多类别半监督语义分割框架(Multi-class Semi-supervised Semantic Segmentation framework based on SAM, MC4S-SAM)。

本文主要贡献包括:(1)在有限标注数据的条件下,利用所提出的MC4S-SAM,实现了对肾小球系膜区三类超微结构的分割,有效减少对大规模标注数据的依赖;(2)改进了SAM模型的掩码解码器,使其在无需先验提示信息的情况下即可具备多类别语义分割能力。结合自训练策略,利用改进后的模型(MCSAM)自动生成并优化伪标签;(3)将MCSAM生成的伪标签纳入半监督学习框架,基于现有的图像级与特征级一致性正则化方法,进一步加入模型级一致性正则化,构建了多级一致性正则化,有效降低伪标签中噪声的干扰,提升了模型的鲁棒性和泛化能力。实验结果表明,在私有数据集上,MC4S-SAM方法在针对肾小球系膜区三类超微结构的分割任务中,取得了优于现有主流方法的性能表现,展现了较强的分割能力,进一步推动了肾小球超微结构分割领域的研究进展。

1 肾小球超微结构分割方法

本文提出的MC4S-SAM旨在充分提高模型在有限的图像标注数据集 $D_l = \left\{ \left(x_i^l, y_i \right) \right\}_{i=1}^M$,以及未标注数据集 $D_u = \left\{ x_i^u \right\}_{i=1}^N$ 下的分割性能,其中, $x_i^l \in \mathbb{R}^{H \times W \times 3}$ 表示标注图像, $y_i \in \mathbb{R}^{H \times W}$ 表示其对应的标注, $x_i^u \in \mathbb{R}^{H \times W \times 3}$ 表示未标注图像, $M \in \mathbb{N}$ 和 $N \in \mathbb{N}$ 表示标注图像和未标注图像的数量。该框架包含两个部分:MCSAM和MCSAM伪标签增强半监督学习。

1.1 MCSAM

SAM通过在规模庞大且多样化的SA-1B数据集上进行训练,旨在实现分割任务的零样本泛化能力。其模型架构由三大关键部分组成,分别是图像编码器(Image Encoder)、提示编码器(Prompt Encoder)以及轻量级的掩码解码器(Mask Decoder)。如图1所示,这3部分协同工作,使得SAM能够在接收图像及

相应提示信息的情况下,为指定对象生成精确的分割掩码^[26]。此外,SAM分割掩码的生成高度依赖于提示信息,当缺少提示时,难以对特定对象进行有效分割。并且由于SAM缺乏语义理解模块,无法直接输出类别标签。因此,SAM难以胜任肾小球超微结构的多类别语义分割任务。值得注意的是,SAM在提示信息模糊的情况下,能够生成整体、部分和子部分3种不同层级的掩码^[26]。基于此,本文对原始SAM进行针对性的改进和优化,使其能够在无需明确提示信息的情况下,生成每个类别的分割掩码。

MLP和类别标记引导掩码生成。

(1)构建语义MLP:为确保解码器能够输出符合数据集中不同类别目标的语义掩码,本文对原始解码器中的MLP层数进行修改。具体而言,在原有用于生成整体、部分和子部分3类目标的MLP结构基础上,增加1层,使其由3层扩展为4层,从而适应语义分割任务中4类目标的输出需求。经过这一调整后的模块被称为语义MLP(Semantic MLP),如图2a中虚线框②所示。

(2)类别标记引导掩码生成:在没有提示信息输入的情况下,SAM会生成一个随机初始化且可学习的特征用于多个掩码预测,该特征的维度与图像编码器输出的特征一致,其表示形式为四维张量,形状为(批量大小、通道数、特征高度、特征宽度),本文将使用该特征代替提示标记(prompt tokens),同SAM类似,将其称为类别标记(class tokens),如图2a中虚线框①所示。首先,类别标记与图像编码器输出的图像特征一同输入到Transformer解码器^[30]中,在两层Transformer解码器^[30]中进行多次的特征注意力计算更新类别标记和图像特征。然后,再次计算类别标记和图像特征的注意力,并将更新后的输出类别标记传递给语义MLP。最后,该语义MLP输出与经过两次上采样放大后的图像特征进行点乘操作,从而预测生成每个类别的语义掩码,如图2中虚线框③所示。

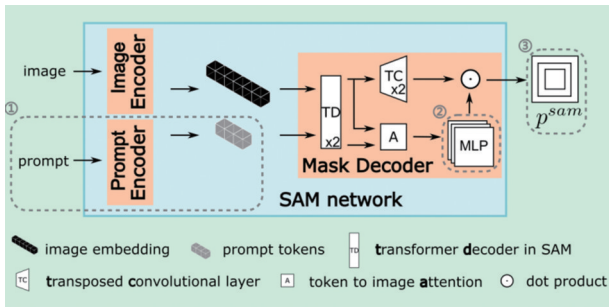


图1 SAM网络结构

Figure 1 SAM network structure

改进后的模型MCSAM如图2a部分所示。具体来说,MCSAM具有两个核心改进,分别是构建语义

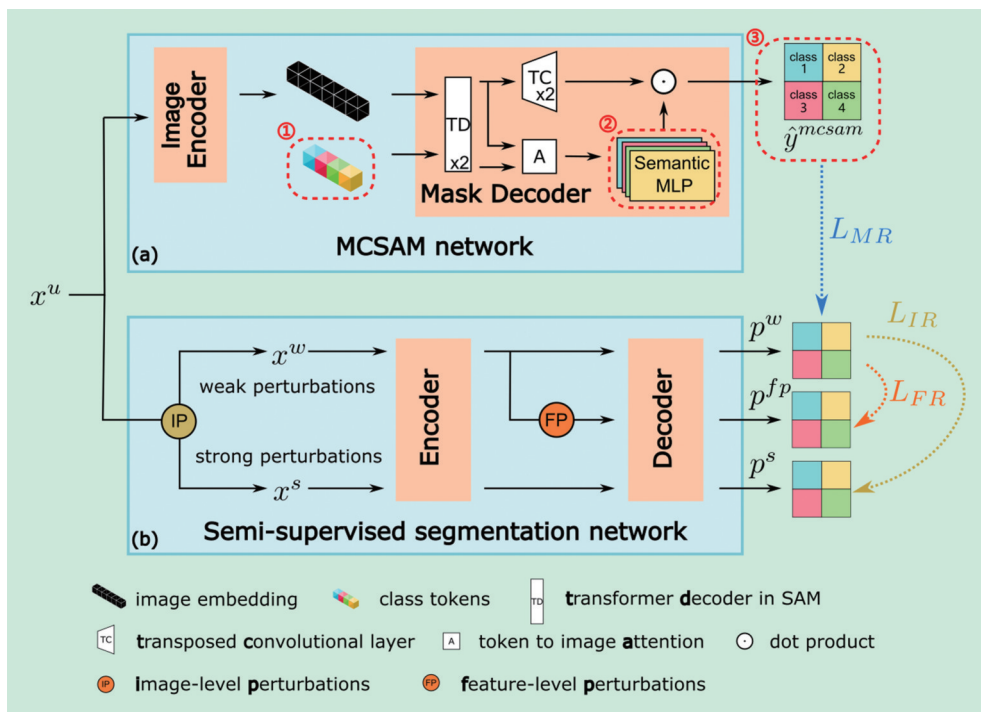


图2 MC4S-SAM方法整体框架

Figure 2 MC4S-SAM framework

1.2 MCSAM 伪标签增强半监督学习

1.2.1 自训练生成并优化伪标签 为了确保 MCSAM 能高效应用于肾小球系膜区超微结构的多类别语义分割任务,本文结合自训练策略生成并优化多类别语义伪标签,为后续一致性正则化奠定坚实基础。自训练策略包括两个阶段。

(1) 第一阶段:伪标签生成。使用标注图像数据集 $D_l = \{(x_i^l, y_i)\}_{i=1}^M$ 对 MCSAM 模型进行微调训练。通过引入真实标注数据,这一过程将原本针对通用场景设计的 SAM 模型调整为专门适用于肾小球系膜区超微结构分割的模型。随后,利用微调后的 MCSAM 对未标注图像数据集 $D_u = \{x_i^u\}_{i=1}^N$ 进行推理,生成伪标签 $\{\hat{y}_i\}_{i=1}^N$,其中 $\hat{y}_i \in \mathbb{R}^{H \times W}$,并将这些图像与伪标签一起添加到训练集中。

(2) 第二阶段:伪标签自训练优化。在第一阶段生成的伪标签可能仍然存在一定的噪声和不确定性,因此在本阶段,本文采用自训练策略对伪标签进行优化。具体而言,第一阶段生成的伪标签被重新加入训练集,与真实标注数据共同用于 MCSAM 的微调,使模型能够更好地适应目标任务。更新后的训练集为 $D_{\text{train}} = D_l \cup \{(x_i^u, \hat{y}_i)\}_{i=1}^N$,其中, \hat{y}_i 为第一阶段推理生成的伪标签。在模型进一步训练后, MCSAM 再次对未标注数据集进行推理,生成最终的优化伪标签集 $D_{\text{mcsam}} = \{(x_i^u, \hat{y}_i^{\text{mcsam}})\}_{i=1}^N$,这些优化后的伪标签用于后续训练,能够更准确地反映超微结构图像的多类别分割对象,为半监督学习提供可靠的监督信息,提升模型的整体分割性能。

1.2.2 多级一致性正则化 现有方法在图像级和特征级一致性正则化过程中,仍然面临伪标签质量较低的问题。为此,本文引入基于 MCSAM 的模型级一致性正则化,通过生成高质量的伪标签来有效引导模型对未标注数据的训练,从而提升正则化效果。本文提出的多级一致性正则化框架如图 2 所示, MC4S-SAM 半监督语义分割框架在现有的图像级与特征级一致性正则化方法基础上,进一步加入模型级一致性正则化,构建多级一致性正则化,以提升模型的鲁棒性与泛化能力,其中,使用 x^u 表示未标注图像集, \hat{y}^{mcsam} 表示对应的 MCSAM 伪标签集。

本文首先基于 UniMatch^[20] 构建图像级扰动 (Image-level Perturbations, IP) 和特征级扰动 (Feature-level Perturbations, FP),如图 2b 所示。首先,对 x^u 施加强扰动操作 $w(\cdot)$ 得到视图 $x^s \in \mathbb{R}^{H \times W \times 3}$,编码后对应的特征记为 $F^w \in \mathbb{R}^{H \times W \times C}$, C 表示特征的维度,解码器对该特征的预测输出为 $p^w \in \mathbb{R}^{H \times W}$,并将

p^w 作为一致性正则化约束过程中基础的伪监督信号。

图像级正则化:对未标注图像 x^u 施加强扰动操作 $s(\cdot)$ 得到视图 $x^s \in \mathbb{R}^{H \times W \times 3}$ 。对于该视图,解码器的预测输出为 $p^s \in \mathbb{R}^{H \times W}$,并通过 p^w 对 p^s 进行监督,构成图像级一致性正则化损失 L_{IR} ,如式(1)所示:

$$L_{IR} = CE(p^s, p^w) \quad (1)$$

其中, $CE(\cdot)$ 表示交叉熵损失。

特征级正则化:使用 Dropout 方法对特征 F^w 进行随机扰动,解码器对该特征的预测输出为 $p^{fp} \in \mathbb{R}^{H \times W}$,通过 p^w 对 p^{fp} 进行监督,构成特征级一致性正则化损失 L_{FR} ,如式(2)所示:

$$L_{FR} = CE(p^{fp}, p^w) \quad (2)$$

在图像级和特征级正则化的过程中,均将 p^w 作为伪监督信号。然而,当模型性能较弱时, p^w 可能包含大量噪声标签,这在训练过程中会干扰一致性正则化,从而影响模型整体性能。因此,为提高伪监督信号质量和正则化约束效果,本文在框架中引入模型级一致性正则化约束。

模型级正则化:对于如图 2b 所示的半监督分割网络 (Semi-supervised segmentation network),将 \hat{y}^{mcsam} 加入到其一致性正则化的训练过程中,构成模型级一致性正则化损失 L_{MR} ,如式(3)所示:

$$L_{MR} = CE(p^{\text{ssn}}, \hat{y}^{\text{mcsam}}) \quad (3)$$

其中, p^{ssn} 表示半监督分割网络的预测。

多级一致性正则化:由于图像级和特征级正则化的存在,半监督分割网络的预测 p^{ssn} 有 p^w 、 p^s 以及 p^{fp} 3 种类型,所以多级一致性正则化的无监督损失函数 L_U 的构建是多样的。在此,本文分为 L_{u1} 和 L_{u2} 两种约束损失进行探究,如式(4)、式(5)所示:

$$L_{u1} = \lambda_1 \cdot L_{IR} + \lambda_2 \cdot L_{FR} + \lambda_3 \cdot L_{MR} \quad (4)$$

$$L_{u2} = \lambda_1 \cdot CE(p^w, \hat{y}^{\text{mcsam}}) + \lambda_2 \cdot CE(p^s, \hat{y}^{\text{mcsam}}) + \lambda_3 \cdot CE(p^{fp}, \hat{y}^{\text{mcsam}}) \quad (5)$$

其中, λ_1 、 λ_2 、 λ_3 为超参数,使用 p^w 作为 p^{ssn} 计算 L_{MR} 损失。

在 L_{u1} 损失中, \hat{y}^{mcsam} 仅对 p^w 进行直接约束,而不对图像扰动预测 p^s 和特征扰动预测 p^{fp} 进行直接约束。在 L_{u2} 中, \hat{y}^{mcsam} 对半监督分割网络中不同扰动项的预测输出均进行直接正则化约束。在本文中,选择 L_{u1} 约束损失作为 MC4S-SAM 语义分割框架的多级一致性正则化损失函数 L_U ,对于 L_{u2} 将在消融实验部分进行讨论。

最终,本文的 MC4S-SAM 语义分割框架损失函数 $L_{\text{MC4S-SAM}}$ 包括监督损失 L_S 和无监督损失 L_U 两部分,如式(6)所示:

$$L_{MCAS-SAM} = \frac{1}{2}(L_S + L_U) \quad (6)$$

监督损失 L_S 应用于标注图像的训练, 计算半监督分割模型预测与真实标签之间的交叉熵损失, 无监督损失 L_U 则应用于未标注图像的训练。

2 实验设置与结果

2.1 数据集介绍及处理

本文为一项回顾性研究, 数据来源于广州华银医学检验中心, 使用 JEOL(日本电子)JEM-1400Flash 电子显微镜在加速电压 100 kV 下观察, 使用 Gatan SC600 CCD camera 进行成像, 图像分辨率为 $2\,048 \times 2\,048$ 。研究共收集了 166 例肾小球疾病患者的 382 张系膜区超微结构图像, 该数据集在患者水平下, 包含狼疮性肾炎(Lupus Nephritis, LN)患者 41 例、

IgA 肾病(IgA Nephropathy, IgAN)患者 41 例、膜性肾病(Membranous Nephropathy, MN)患者 51 例以及其他类型(Others)肾病患者 33 例; 在图像水平下, 包含 LN 图像 83 张、IgAN 图像 118 张、MN 图像 90 张以及其他类型肾病图像 91 张。所有数据均经过脱敏处理, 已完全移除患者的个人隐私信息。

本文按患者例数进行五折交叉验证的数据随机划分, 每一折数据按 3:1:1 的比例划分为训练集、验证集和测试集, 具体数据如表 1 所示。每折训练完成后, 使用该折的最优模型参数对相应的测试集进行评估, 并计算均值和方差作为最终实验结果。此外, 受实验硬件设备限制, 原始 $2\,048 \times 2\,048$ 像素的肾小球系膜区超微结构图像被调整为 $1\,024 \times 1\,024$ 像素, 以满足实验需求。

表 1 五折交叉验证数据介绍[病人数(图像数)]

Table 1 Introduction to five-fold cross validation data [number of patients (number of images)]

| 交叉验证折数 | 训练集 | 验证集 | 测试集 |
|--------|---|---|---|
| 第 1 折 | LN:23(46), IgAN:27(78), MN:32(48), Others:18(57) | LN:8(19), IgAN:3(7), MN:12(23), Others:10(24) | LN:10(18), IgAN:11(33), MN:7(19), Others:5(10) |
| 第 2 折 | LN:20(40), IgAN:21(51), MN:37(62), Others:22(58) | LN:10(18), IgAN:11(33), MN:7(10), Others:5(19) | LN:11(25), IgAN:9(34), MN:7(9), Others:6(23) |
| 第 3 折 | LN:25(48), IgAN:22(57), MN:33(59), Others:20(57) | LN:11(25), IgAN:9(34), MN:7(9), Others:6(23) | LN:5(10), IgAN:10(27), MN:11(13), Others:7(20) |
| 第 4 折 | LN:29(62), IgAN:23(74), MN:26(42), Others:21(66) | LN:5(10), IgAN:10(27), MN:11(13), Others:7(20) | LN:7(11), IgAN:8(17), MN:14(26), Others:5(14) |
| 第 5 折 | LN:26(53), IgAN:33(94), MN:25(32), Others:18(62) | LN:7(11), IgAN:8(17), MN:14(26), Others:5(14) | LN:8(19), IgAN:3(7), MN:12(23), Others:10(24) |

本文目标是对肾小球系膜区中系膜细胞的细胞核、胞质和基质以及周边基底膜 3 类超微结构进行分割。所有数据均由 3 名医学实习生进行初步标注, 并由两名病理专家进行审核, 以确保标注的准确性和一致性。图像标注过程中未区分病人个体差异或图像来源差异。分割标注类别划分标准为: 以细胞核边缘区分细胞核与胞质基质, 以胞质边缘区分胞质基质与基底膜、内皮细胞, 以基底膜的内皮侧边缘和系膜区边缘区分基底膜与其他结构。

2.2 实验设置及评估指标

本文的模型训练均在相同的软硬件环境下进行, 其中 GPU 为 NVIDIA GeForce RTX 3090, 编程语言使用 Python 3.11.8, 所依赖的关键 Python 库包括: torch 2.2.2、torchaudio 2.2.2、torchvision 0.17.2 以及 albumentations 1.4.10。

在伪标签生成阶段和一致性正则化阶段的训练过程中, 由于模型不同, 实验任务不同, 因此, 在每个阶段采用不同的超参数设置和数据增强策略。

(1) 在伪标签生成阶段, 模型训练的超参数设置如下: 训练周期(epoch)为 30, 批次大小(batch size)为 2, 优化器选择 Adam, 学习率为 0.000 1, 动量因子(betas)为 (0.9, 0.999), 权重衰减系数为 0.000 1。数据增强基于 albumentations 库实现, 随机概率 p 为 0.3, 具体如下: 在几何变换方面, 进行随机水平翻转和垂直翻转, 以及进行随机平移、缩放和旋转变换; 在颜色增强方面, 进行随机亮度和对比度调整以及随机 Gamma 调整; 在噪声和模糊方面, 随机引入高斯模糊和高斯噪声。(2) 在一致性正则化阶段, 超参数的设置有所不同: 训练周期为 150, 批次大小同样为 2, 优化器选择 SGD, 学习率为 0.001, 动量为 0.9, 权重衰减系数为 0.000 1, 损失函数系数 λ_1 、 λ_2 、 λ_3 均为 1/3。数据增强相关参数设置与 UniMatch^[20] 方法保持一致。

本文选择平均交并比(mean Intersection over Union, mIoU) 和平均 Dice 相似系数(mean Dice Similarity Coefficient, mDSC) 作为评估模型在肾小球系膜区 3 个关键结构分割性能的分类平均评价指标。

2.3 实验结果与分析

为验证所提出的半监督分割框架MC4S-SAM的优越性,分别在标注数据占比为1/16、1/8和1/4的条件下,将MC4S-SAM与当前主流的半监督分割方法进行对比。同时,为验证模块的有效性进行消融实验。最后,通过可视化分割结果来进一步说明所提分割框架的效果。

2.3.1 对比实验 为了充分展示MC4S-SAM的先进性,本文将与其与多种基于不同分割网络的语义分割

方法进行对比。这些方法包括:以DeepLabV3+^[31]为分割网络的全监督方法,以及ST++^[14]、U2PL^[18]、UniMatch^[20]等半监督方法;以U-Net^[32]为分割网络的全监督方法,以及Mean Teacher^[17]、Yu^[19]、Luo^[22]、CPS^[21]等半监督方法。其中,DeepLabV3+采用ResNet50^[33]作为编码器,U-Net则使用简单的卷积层作为编码器。在肾小球系膜区语义分割数据集上,本文提供不同分割方法的对比实验结果,具体如表2所示。

表2 不同分割方法的对比实验结果 ($\bar{x}\pm s, \%$)

Table 2 Comparative experimental results of different segmentation methods ($Mean\pm SD, \%$)

| 分割网络 | 模型 | mIoU | | | | mDSC | | | |
|------------|------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | | 1/16 | 1/8 | 1/4 | 1/1 | 1/16 | 1/8 | 1/4 | 1/1 |
| DeepLabV3+ | 全监督 | 51.33±1.32 | 58.92±2.23 | 65.16±2.87 | 70.01±1.30 | 62.22±2.70 | 69.29±1.81 | 74.99±2.44 | 67.65±1.38 |
| | ST++ ^[14] | 47.35±3.35 | 55.04±2.59 | 60.41±1.10 | - | 60.36±4.02 | 67.04±1.75 | 71.77±0.60 | - |
| | U2PL ^[18] | 62.13±0.61 | 66.10±2.86 | 70.07±0.50 | - | 73.06±0.72 | 76.36±2.02 | 79.63±0.52 | - |
| | UniMatch ^[20] | 62.70±1.17 | 65.13±1.27 | 67.33±0.60 | - | 73.46±0.90 | 75.69±0.57 | 77.30±0.36 | - |
| | MC4S-SAM | 63.05±1.02 | 67.07±0.39 | 68.91±0.22 | - | 73.67±0.75 | 76.87±0.79 | 78.73±0.82 | - |
| U-Net | 全监督 | 37.34±1.92 | 43.56±2.44 | 45.96±2.50 | 55.96±2.27 | 48.96±2.62 | 55.62±1.73 | 57.77±2.28 | 67.65±1.38 |
| | Mean Teacher ^[17] | 29.67±2.83 | 38.57±2.53 | 40.27±3.84 | - | 41.57±2.92 | 50.59±2.83 | 52.59±3.39 | - |
| | Yu ^[19] | 35.01±2.38 | 40.03±3.47 | 45.32±3.73 | - | 47.50±3.03 | 52.53±4.18 | 58.10±3.34 | - |
| | Luo ^[22] | 41.31±2.83 | 47.60±2.53 | 55.88±3.84 | - | 54.10±2.92 | 60.54±2.83 | 68.02±3.39 | - |
| | CPS ^[21] | 33.93±2.48 | 39.81±2.87 | 45.61±1.51 | - | 46.45±2.58 | 52.02±3.72 | 58.39±1.71 | - |
| | MC4S-SAM | 52.61±1.74 | 53.72±1.54 | 54.10±2.35 | - | 64.90±1.20 | 65.97±1.03 | 66.35±2.34 | - |

1/16、1/8、1/4和1/1表示训练数据集中标注数据所占比例,各组实验中使用的训练数据总量保持一致

与全监督方法相比,在标注数据占比为1/16、1/8、1/4的条件下,使用DeepLabV3+作为分割网络时,MC4S-SAM的mIoU分别提高11.72%、8.15%、3.75%,mDSC分别提高11.45%、7.58%、3.74%;使用U-Net作为分割网络时,MC4S-SAM的mIoU相比于全监督方法分别提高15.27%、10.16%、8.14%,mDSC分别提高15.94%、10.35%、8.58%。结果表明,MC4S-SAM语义分割框架在较低的标注水平下,能够有效利用未标注数据,使得分割模型性能有大幅提升,因此该框架能够有效解决图像标注稀缺问题。

与其他半监督方法相比,在标注数据占比为1/16、1/8的情况下,无论是以DeepLabV3+还是U-Net作为分割网络,MC4S-SAM均取得最优的性能。在DeepLabV3+分割网络下,基于一致性正则化方法均优于基于自训练的ST++方法,这表明通过一致性正则化能够有效降低噪声标签的干扰。MC4S-SAM对比其他一致性正则化方法也有一定的优势,体现出引入MCSAM伪标签的有效性。在与Mean Teacher^[16]、Yu^[19]、Luo^[22]、CPS^[21]等以U-Net作为分割网络的一致性正则化方法对比中,MC4S-SAM的优势

更为明显,在标注数据占比为1/16的情况下,mIoU分别提升22.94%、17.60%、11.30%、18.68%;mDSC分别提升23.33%、17.40%、10.80%、18.45%。该结果表明MC4S-SAM分割框架能够通过MC-SAM模型对U-Net进行有效伪监督,提高U-Net提取特征的能力,使得轻量化编码器也能有效地对不同类别特征进行学习与识别。更为值得一提的是,无论使用DeepLabV3+还是U-Net作为半监督分割网络,MC4S-SAM方法均保持良好的性能表现,展现出较强的模型鲁棒性和灵活性,能够轻松适配多种分割网络架构,为系膜区超微结构语义分割任务提供有力的技术支持。

2.3.2 消融实验 为了验证MC4S-SAM各个模块的有效性,本文使用DeepLabV3+作为分割网络,进行消融实验,如表3所示。具体来说,本文从以下3个方面进行消融实验。首先,对比不同伪标签生成网络(DeepLabV3+、MCSAM),如表3中第2、3行所示;其次,有、无自训练策略生成并优化伪标签,如表3中第3、4行所示;最后,探究多级一致性正则化的直接约束和间接约束损失函数的影响,如表3中第4、5行

所示。由表3可以看出, MCSAM生成伪标签的质量优于DeepLabV3+, 在经过MCSAM自训练的情况下(+MCSAM*+ L_{u_2}), 伪标签的质量进一步提升, 更有利于增强半监督网络模型的性能, 特别是在标注数据较少(标注数据占比为1/16)的情况下, 相较于无自训练的MCSAM(+MCSAM+ L_{u_2}), mIoU和mDSC分别

提升2.06%和2.19%。对于两种不同一致性损失函数的约束, L_{u_1} 比 L_{u_2} 有更好的效果。其原因在于, L_{u_1} 并未直接强制所有输出与MCSAM伪标签对齐, 从而避免了对伪标签的过度拟合, 允许模型从更多样的伪标签中学习, 进而提取出更丰富的特征表示, 有助于在未标注数据上取得更优的表现。

表3 MC4S-SAM的消融实验结果($\bar{x}\pm s, \%$)Table 3 Ablation experiment results of MC4S-SAM (Mean \pm SD, %)

| 模块消融 | mIoU | | | mDSC | | |
|-------------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | 1/16 | 1/8 | 1/4 | 1/16 | 1/8 | 1/4 |
| 基线(UniMatch) | 62.70 \pm 1.17 | 65.13 \pm 1.27 | 67.33 \pm 0.60 | 73.46 \pm 0.90 | 75.69 \pm 0.57 | 77.30 \pm 0.36 |
| +DeepLabV3+ + L_{u_2} | 59.57 \pm 2.95 | 64.28 \pm 0.98 | 66.77 \pm 0.93 | 70.03 \pm 2.61 | 74.40 \pm 1.18 | 76.55 \pm 0.60 |
| +MCSAM+ L_{u_2} | 60.01 \pm 1.89 | 65.44 \pm 0.36 | 68.00 \pm 0.70 | 70.85 \pm 1.50 | 75.65 \pm 0.49 | 77.87 \pm 0.70 |
| +MCSAM*+ L_{u_2} | 62.07 \pm 1.86 | 66.90 \pm 0.63 | 68.38 \pm 0.36 | 73.04 \pm 1.24 | 76.74 \pm 0.69 | 78.14 \pm 0.65 |
| +MCSAM*+ L_{u_1} | 63.05 \pm 1.02 | 67.07 \pm 0.39 | 68.91 \pm 0.22 | 73.67 \pm 0.75 | 76.87 \pm 0.79 | 78.73 \pm 0.82 |

*表示模型进行自训练

虽然本文基于UniMatch^[20]开展, 但也探究了不同特征扰动类型的影响, 并进行消融实验, 如表4所示。根据表4的结果, 添加高斯噪声的实验表现最差。这主要是因为高斯噪声对每个特征值进行扰动, 导致模型在训练过程中难以捕捉到稳定的特征

表示, 从而影响整体性能。相比之下, 使用Dropout进行扰动时, 未被丢弃的部分依然保持了原有的空间一致性和局部结构。这种方式帮助模型在面对部分特征缺失的情况下, 依然能够依赖其他完整的特征进行准确预测, 从而提升模型的性能。

表4 不同特征扰动类型的消融实验结果($\bar{x}\pm s, \%$)Table 4 Ablation experiment results of different feature perturbation types (Mean \pm SD, %)

| 特征扰动类型 | mIoU | | | mDSC | | |
|--------------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | 1/16 | 1/8 | 1/4 | 1/16 | 1/8 | 1/4 |
| 无扰动 | 61.83 \pm 1.14 | 66.18 \pm 0.18 | 68.07 \pm 0.98 | 72.90 \pm 0.88 | 76.40 \pm 0.46 | 77.88 \pm 1.25 |
| 添加高斯噪声($\sigma = 0.1$) | 60.98 \pm 1.64 | 65.61 \pm 0.43 | 68.01 \pm 0.76 | 72.20 \pm 0.98 | 75.85 \pm 0.43 | 77.63 \pm 1.34 |
| Dropout($p = 0.3$) | 61.89 \pm 2.00 | 66.89 \pm 0.44 | 68.64 \pm 0.52 | 73.12 \pm 1.54 | 76.75 \pm 0.61 | 78.49 \pm 0.72 |
| Dropout($p = 0.5$) | 63.05 \pm 1.02 | 67.07 \pm 0.39 | 68.91 \pm 0.22 | 73.67 \pm 0.75 | 76.87 \pm 0.79 | 78.73 \pm 0.82 |

2.4 可视化

为了直观地展示MC4S-SAM与Ground Truth以及其他对比方法之间的差距, 本文对系膜区超微结构分割结果进行可视化分析, 如图3所示。图3中第1、3、5行表示原图、Ground Truth以及各个方法的预测结果, 其中, 粉红色表示系膜区周边的基底膜, 浅绿色表示胞质和基质, 淡蓝色表示细胞核, 第2、4、6行表示各个方法的预测结果与Ground Truth之间的误差, 红色表示分割基底膜的误差, 绿色表示分割胞质和基质的误差, 蓝色表示分割细胞核的误差。

在所有的分割误差图像中, MC4S-SAM的细胞核分割误差最小, 如图3的第2、4、6行中蓝色部分所示, 展现了较高的分割准确性。然而, 当系膜区的边缘与内皮细胞相邻时, 由于二者边界较为模糊, 区分

难度较大, 导致分割错误率较高。但相比其他方法, MC4S-SAM仍然具备一定的优势, 如图3的第4、6行中绿色部分所示, 这种错误在标注数据量较小时更为明显, 如图3的第2行中绿色部分所示。此外, 在分割系膜区周边的基底膜时, 由于其灰度特征与基质及肾小球滤过屏障区域的基底膜具有较高的一致性, 可能会导致部分错误分割, 如图3的第6行中红色部分所示。综合来看, 虽然MC4S-SAM在某些复杂区域仍存在一定的误差, 但相较于其他方法, 依然表现出明显的优势, 尤其在边界清晰的结构方面, 分割精度更高, 误差更小。原因在于, 基于自训练的ST++方法在训练过程中容易导致错误累积, 这进而使得模型拟合了错误的标签; 基于一致性正则化的UniMatch方法, 由于低质量的伪标签不能够进行较

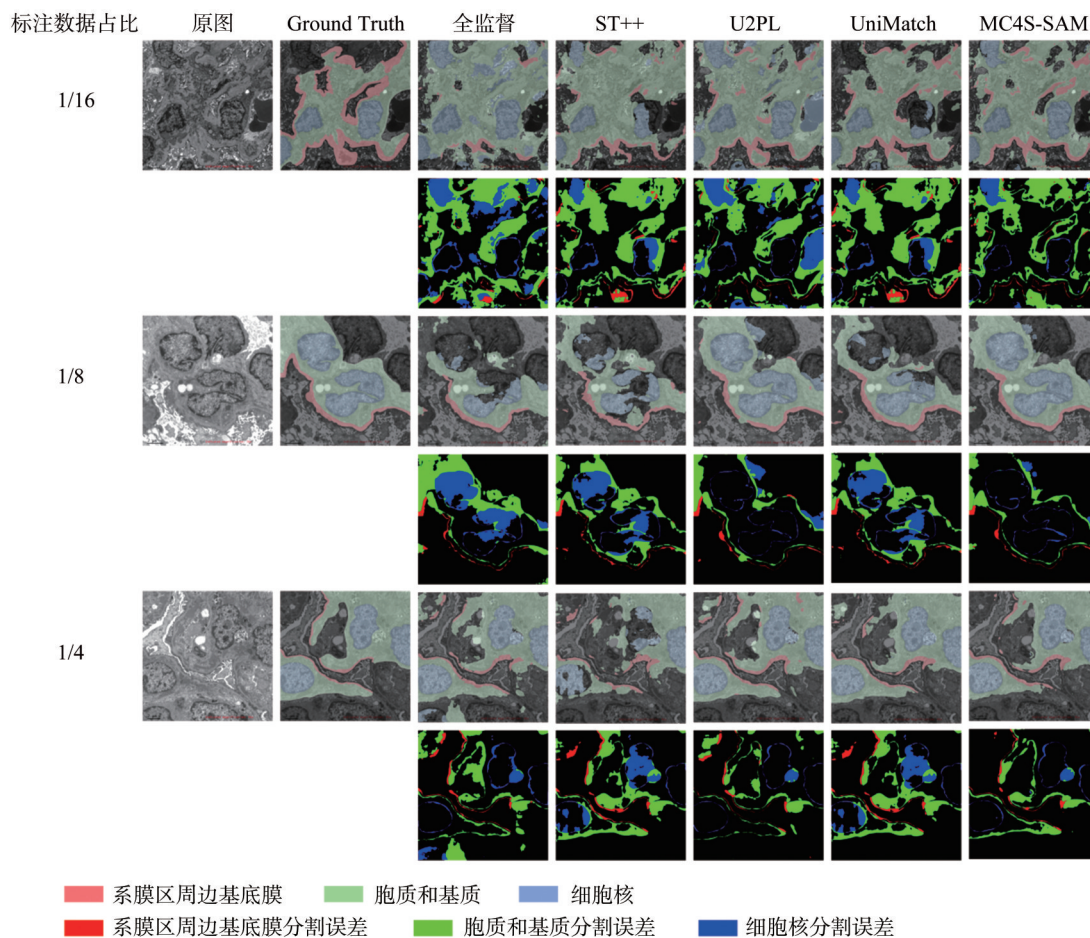


图3 不同方法分割结果对比

Figure 3 Comparison of segmentation results obtained by different methods

好的一致性正则化约束,导致模型性能较弱;U2PL使用对比学习的方式来处理低质量的伪标签,但在标注数据量较小(1/16)时对其性能有一定的限制;而MC4S-SAM方法通过MCSAM模型以自训练方式生成的伪标签,同时实施多级一致性正则化策略避免了伪标签中噪声的影响。

3 结语

肾小球系膜区的精准识别对于慢性肾脏病的辅助诊断至关重要。然而,由于标注数据的匮乏,肾小球系膜区超微结构的语义分割仍面临较大挑战。为了解决这一问题,本文提出一种基于SAM的多类别半监督语义分割框架MC4S-SAM。此框架对SAM的掩码解码器进行优化升级,使其能够在无提示信息的情况下生成适用于多类别语义分割任务的伪标签,并结合自训练策略提升伪标签的质量与可靠性。随后,将这些的伪标签融入半监督学习框架中,通过实施图像、特征和模型的多级一致性正则化约束,最大限度地减少了伪标签中的噪声干扰,从而更好地满足系膜区超微结构分割的需求。这一方法不仅提

升了多类别语义分割模型的性能,也为解决标注数据稀缺问题提供新的思路与解决方案。与当前的其他利用SAM提供伪标签的方法相比,MCSAM模型能够无需借助任何提示即可生成多类别伪标签供半监督模型学习;而且采用多级一致性正则化策略能够在一定程度上降低伪标签中噪声信息的干扰。在系膜区分割数据集上,MC4S-SAM框架在伪标签标注数量较少的情况下优于当前主流半监督分割方法,特别是在轻量级分割模型(U-Net)中表现突出。然而,MC4S-SAM也存在一些局限性。随着标注数据量占比的增大,该框架的优势逐渐减弱;以及受限于数据量及分割对象的复杂性,对肾小球系膜区周边基底膜的分割仍存在一定不足。未来将致力于通过探究更为有效的一致性正则化策略来提升模型性能。

【参考文献】

[1] Kalantar-Zadeh K, Jafar TH, Nitsch D, et al. Chronic kidney disease [J]. Lancet, 2021, 398(10302): 786-802.
 [2] Borg R, Carlson N, Søndergaard J, et al. The growing challenge of chronic kidney disease: an overview of current knowledge[J]. Int J Nephrol, 2023, 2023(1): 9609266.

- [3] 刘志红, 曾彩虹, 梁丹丹. 肾活检病理诊断报告模式专家共识[J]. 中国实用内科杂志, 2017, 37(9): 810-816.
Liu ZH, Zeng CH, Liang DD. Pathological diagnosis reporting consensus on kidney biopsy in China[J]. Chinese Journal of Practical Internal Medicine, 2017, 37(9): 810-816.
- [4] 邹万忠. 肾活检病理学[M]. 第5版. 北京: 北京大学医学出版社, 2021.
Zou WZ. Pathology of renal biopsy[M]. 5th ed. Beijing: Peking University Medical Press, 2021.
- [5] Lin GY, Zhang ZT, Long KX, et al. GCLR: a self-supervised representation learning pretext task for glomerular filtration barrier segmentation in TEM images[J]. Artif Intell Med, 2023, 146: 102720.
- [6] Wang YZ, Liu YQ, Fu YQ, et al. Segmentation and thickness calculation of glomerular basement membrane using RADS-Net in glomerular microscopic images[J]. Biomed Signal Process Control, 2024, 88, Part A: 105557.
- [7] Bortsova G, Dubost F, Hogeweg L, et al. Semi-supervised medical image segmentation via learning consistency under transformations [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2019. Cham: Springer International Publishing, 2019: 810-818.
- [8] Luo XD, Chen JN, Song T, et al. Semi-supervised medical image segmentation through dual-task consistency[C]//Proceedings of the AAAI conference on artificial intelligence. Palo Alto, CA, USA: AAAI Press, 2021: 8801-8809.
- [9] Wu CX, Zhang WX, Han JZ, et al. Multi-consistency training for Semi-Supervised medical image segmentation[J/OL]. J Shanghai Jiaotong Univ Sci. (2024-05-10). <https://doi.org/10.1007/s12204-024-2733-0>.
- [10] Jiao RS, Zhang YC, Ding L, et al. Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation[J]. Comput Biol Med, 2024, 169: 107840.
- [11] Peláez-Vegas A, Mesejo P, Luengo J. A survey on semi-supervised semantic segmentation[EB/OL]. (2023-02-20). <https://arxiv.org/abs/2302.09899>.
- [12] 许华杰, 肖毅烽. 基于多教师网络模型的半监督语义分割方法[J]. 计算机科学, 2023, 50(12): 279-284.
Xu HJ, Xiao YF. Semi-supervised semantic segmentation method based on multiple teacher network model[J]. Computer Science, 2023, 50(12): 279-284.
- [13] Lee D. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks[C]//Workshop on Challenges in Representation Learning. Atlanta: ICML, 2013.
- [14] Yang LH, Zhuo W, Qi L, et al. ST++: make self-training work better for semi-supervised semantic segmentation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 4258-4267.
- [15] Sajjadi M, Javanmardi M, Tasdizen T. Regularization with stochastic transformations and perturbations for deep semi-supervised learning[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2016: 1171-1179.
- [16] Sohn K, Berthelot D, Li CL, et al. FixMatch: simplifying semi-supervised learning with consistency and confidence[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2020: 596-608.
- [17] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017: 1195-1204.
- [18] Wang YC, Wang HC, Shen YJ, et al. Semi-supervised semantic segmentation using unreliable pseudo-labels[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 4238-4247.
- [19] Yu LQ, Wang SJ, Li XM, et al. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation[C]//Medical Image Computing and Computer Assisted Intervention - MICCAI 2019. Cham: Springer International Publishing, 2019: 605-613.
- [20] Yang L, Qi L, Feng L, et al. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
- [21] Chen XK, Yuan YH, Zeng G, et al. Semi-supervised semantic segmentation with cross pseudo supervision[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 2613-2622.
- [22] Luo XD, Hu MH, Song T, et al. Semi-supervised medical image segmentation via cross teaching between CNN and transformer[C]//Proceedings of The 5th International Conference on Medical Imaging with Deep Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2022: 820-833.
- [23] Wang LK, Zhang M, Shi WZ. STCRNet: a semi-supervised network based on self-training and consistency regularization for change detection in VHR remote sensing images[J]. IEEE J Sel Top Appl Earth Obs Remote Sens, 2024, 17: 2272-2282.
- [24] Li BL, Xu YM, Wang Y, et al. The student-teacher framework guided by self-training and consistency regularization for semi-supervised medical image segmentation[J]. PLoS One, 2024, 19(4): e0300039.
- [25] Arazo E, Ortego D, Albert P, et al. Pseudo-labeling and confirmation bias in deep semi-supervised learning[C]//2020 International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ, USA: IEEE, 2020: 1-8.
- [26] Kirillov A, Mintun E, Ravi N, et al. Segment anything[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 3992-4003.
- [27] Zhang YC, Cheng Y, Qi Y. SemiSAM: exploring SAM for enhancing semi-supervised medical image segmentation with extremely limited annotations[EB/OL]. (2023-12-11). <https://arxiv.org/abs/2312.06316v1>.
- [28] Li SM, Qi L, Yu Q, et al. Concatenate, fine-tuning, re-training: a SAM-enabled framework for semi-supervised 3D medical image segmentation[EB/OL]. (2024-03-17). <https://arxiv.org/abs/2403.11229v1>.
- [29] Chen SY, Lin L, Cheng PJ, et al. ASLSEG: adapting Sam in the loop for semi-supervised liver tumor segmentation[C]//2024 IEEE International Symposium on Biomedical Imaging (ISBI). Piscataway, NJ, USA: IEEE, 2024: 1-5.
- [30] Yun S, Han D, Chun S, et al. CutMix: regularization strategy to train strong classifiers with localizable features[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 6022-6031.
- [31] Chen LC, Zhu YK, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 833-851.
- [32] 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.
- [33] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.

(编辑:薛泽玲)