

基于可疑像素相互修正的半监督医学图像分割

杨子瑶^{1,2}, 雷涛^{1,2*}, 杜晓刚^{1,2}, 王梦溪^{1,2}, 孙菲蔓³, 公茂果⁴

(1. 陕西科技大学电子信息与人工智能学院, 陕西西安 710021; 2. 陕西省人工智能联合实验室(陕西科技大学), 陕西西安 710021;
3. 西南交通大学数学学院, 四川成都 611756; 4. 西安电子科技大学协同智能系统教育部重点实验室, 陕西西安 710071)

摘要: 现有的半监督学习方法通常对图像中的所有像素进行同等处理, 忽视了图像内不同区域的复杂度差异. 这导致模型对于预测难度较高的挑战性区域学习不足, 从而降低了其对这些区域的处理能力. 此外, 由于伪标签是基于当前模型对未标注数据的预测结果生成的, 而模型在挑战性区域表现较差, 因此不准确的预测增加了伪标签中引入噪声的风险, 进而降低了伪标签的可信度. 针对上述问题, 提出了一种基于可疑像素相互修正的半监督学习框架 (Suspicious Pixel Mutual Correction, SPMC). 该框架由两个编码器结构相同, 但解码器上采样方式不同的网络构成. 首先, 设计了一个用于处理标注数据的共困难像素筛查 (Common Difficulty Pixel Screening, CDPS) 模块. 通过利用每个网络预测结果中的可疑像素, 精准筛查出两个网络都预测困难的共困难像素, 并使用准确的监督信号对其进行修正, 从而提高模型对挑战性区域预测的准确性. 其次, 设计了一个用于处理未标注数据的软伪标签辅助教学 (Soft Pseudo-Label Assisted Teaching, SPLAT) 模块. 通过利用一个网络生成的软伪标签中的可信像素, 选择性地对另一个网络预测结果中的可疑像素进行伪监督. 两个网络通过高质量交互来更新参数, 从而减少模型的认知偏差并提升伪标签质量. 实验结果表明, 提出的方法在三种公开医学数据集左心房 (Left Atrium, LA)、脑部肿瘤分割 (Brain Tumor Segmentation, BraTS) 和自动心脏挑战 (Automatic Cardiac Diagnosis Challenge, ACDC) 上的性能均优于当前主流的半监督学习方法.

关键词: 半监督学习; 医学图像分割; 一致性正则化; 伪标签; 不确定性; 像素级修正

基金项目: 国家自然科学基金 (No.62271296, No.62201334); 陕西省教育厅科学研究计划项目 (No.23JP022, No.23JP014)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2025)05-1607-15

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20240616

Semi-Supervised Medical Image Segmentation Based on Suspicious Pixel Mutual Correction

YANG Zi-yao^{1,2}, LEI Tao^{1,2*}, DU Xiao-gang^{1,2}, WANG Meng-xi^{1,2}, SUN Fei-man³, GONG Mao-guo⁴

(1. School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, Shaanxi 710021, China;

2. Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, Shaanxi 710021, China;

3. School of Mathematics, Southwest Jiaotong University, Chengdu, Sichuan 611756, China;

4. Key Laboratory of Collaborative Intelligent Systems, Ministry of Education, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract: Existing semi-supervised learning methods typically handle all pixels in the image equally, ignoring the differences in complexity of different regions within the image. This results in the model's insufficient learning of challenging regions with higher difficulty to predict, reducing its ability to process challenging areas. Furthermore, since pseudo-labels are generated based on the model's predictions on unlabeled data, and the model performs poorly in challenging regions, inaccurate predictions increase the risk of introducing noise into the pseudo-labels, thereby reducing their reliability. To address these issues, a semi-supervised learning framework based on suspicious pixels mutual correction (SPMC) is proposed, the framework consists of two networks with identical encoder structures but different upsampling methods in the decoder. Firstly, a common difficulty pixel screening (CDPS) module was designed to handle labeled data. By utilizing suspicious pixels from the prediction results of each network to accurately screen out the hard to predict pixels in both networks. These pixels are then corrected using precise supervision signals, thereby improving the model's prediction accuracy in chal-

lenging regions. Secondly, a soft pseudo-label assisted teaching (SPLAT) module was developed to handle unlabeled data. By utilizing trusted pixels from the soft pseudo-labels generated by one network to selectively pseudo-supervise suspicious pixels in the predicted results of the other network. Two networks update parameters through high-quality interaction, thereby reducing the model's cognitive bias and enhancing the quality of the pseudo-labels. Experimental results on three publicly available medical datasets, left atrium (LA), brain tumor segmentation (BraTS), and automatic cardiac diagnosis challenge (ACDC), show that the proposed method is superior to current mainstream semi-supervised learning methods.

Key words: semi-supervised learning; medical image segmentation; consistency regularization; pseudo-labels; uncertainty; pixel-level correction

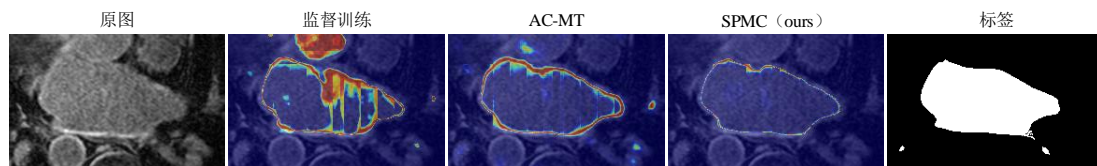
Foundation Item(s): National Natural Science Foundation of China (No.62271296, No.62201334); Scientific Research Program Funded by Shaanxi Provincial Education Commission (No.23JP022, No.23JP014)

1 引言

医学图像分割旨在准确地识别并分离原始图像中的器官、肿瘤等解剖结构,为临床诊疗提供可靠的辅助信息,能够帮助医生更准确地诊断疾病、制定个性化治疗方案,并有效监测治疗效果^[1]。随着深度学习技术的不断发展,基于卷积神经网络的多种新型网络结构相继被提出^[2-4],这些新型网络结构在自动图像分割任务中实现了显著的性能提升。然而,这种基于全监督学习的图像分割方法通常依赖大量的标注数据。由于医学图像的特殊性,公开的数据集相对较少,且医学图像的标注需要具备专业知识的医生来完成,导致高质量的标注数据较为稀缺。相比之下,半监督学习则能够利用少量标注数据和大量未标注数据实现模型训练,从而获得与全监督学习相当甚至更好的分割结果。

目前,基于一致性正则化和伪标签的半监督学习方法已成为该领域的主流方法^[5-9]。尽管这些方法在提高模型性能和降低标注成本方面取得了重要进展,但仍存在以下挑战。首先,医学图像具有复杂的解剖结构,组织和器官的边界较为模糊,并且在成像过程中容

易受到伪影的影响。这些特性导致图像中不同区域的预测难度存在显著差异,其中预测难度较高的区域又称为挑战性区域^[10]。挑战性区域的预测结果通常具有较高的不确定性,如图1所示,在某训练样本的不确定性热图中可以观察到,不确定性较高的像素通常集中在分割目标的边缘区域。因此,如果不对挑战性区域进行单独处理,可能会导致模型对这些区域的处理能力较差。然而,现有的半监督学习方法在进行图像分割时,通常将输入图像中的所有像素视为一个整体,未能充分考虑不同区域像素之间的分割特性和难度,缺乏对挑战性区域的有效处理^[11,12]。其次,伪标签的质量对于未标注数据预测的准确性至关重要。然而,生成的伪标签中往往存在大量噪声。如果不对这些噪声进行适当处理,将影响伪监督的有效性,进而导致模型性能下降。一些方法通过忽略不确定性较高的像素,确保模型在训练过程中只使用不确定性较低的像素作为伪标签,以此来提高伪标签质量并增强模型的稳定性和准确性^[13,14]。然而,这些策略虽然可以减少噪声对伪标签的影响,但不确定性较高的像素仍包含有价值的学习信息,因此,忽略这些像素会导致重要知识的丢失。



注:在LA数据集上使用10%的标注数据分别训练全监督和半监督模型生成的某训练样本的不确定性热图。

图1 某训练样本的不确定性热图

为了解决上述问题,提出了一种基于可疑像素相互修正的半监督学习框架(Suspicious Pixel Mutual Correction, SPMC),其中可疑像素是指图像预测结果中不确定性较高的像素。该框架由两个结构不同且独立更新参数的网络组成,它利用两个网络预测结果中的不确定性信息,来筛选图像中的挑战性区域,两个网络通过相互反馈来修正彼此预测结果中的可疑像素,从而进行高质量的交互学习。具体来说,提出了共困像素筛

查(Common Difficulty Pixel Screening, CDPS)模块和伪标签辅助教学(Soft Pseudo-Label Assisted Teaching, SPLAT)模块,分别用于标注和未标注数据的训练。本文的主要贡献如下:

(1)提出CDPS模块,旨在解决因模型对挑战性区域学习不足而导致的模型预测不准确的问题。与主流方法对图像中所有像素进行同等处理的方式不同,CDPS将两个网络的预测结果中不确定性较高的可疑

像素单独提取出来. 通过筛查出两个预测结果中共同存在的可疑像素(即共因像素), CDPS 引导网络重点关注这些共因像素并对其进行修正, 从而提高模型对复杂和模糊区域预测的准确性.

(2) 提出 SPLAT 模块, 旨在解决伪标签中存在噪声带来的潜在影响. 不同于主流方法仅使用不确定性较低的可信像素作为伪标签进行伪监督, SPLAT 同时对不确定性较高的可疑像素进行信息挖掘. 为了充分利用这些可疑像素, SPLAT 使用一个网络预测中的可信像素, 对另一个网络预测中的可疑像素进行选择性的伪监督, 同时引入软伪标签的思想来增强可信像素的稳定性. 这一过程为网络提供辅助教学信号, 有效提升了模型对未标注数据的学习效率, 进一步提高了伪标签的质量.

(3) 提出的框架集成了以上两个模块, 在 3D 数据集左心房 (Left Atrium, LA)、脑部肿瘤分割 (Brain Tumor Segmentation, BraTS) 和 2D 数据集自动心脏挑战 (Automatic Cardiac Diagnosis Challenge, ACDC) 三个公开数据集上进行了性能评估, 实验证明所提框架的分割性能优于当前主流的半监督学习方法.

2 相关工作

2.1 半监督学习

半监督学习旨在有效利用未标注数据来提升模型的整体性能, 从而实现与完全依赖标注数据训练的监督学习相近的效果. 常见的半监督学习方法可以分为生成对抗方法^[15-17]、对比学习^[18-20]、一致性正则化^[11, 12, 21]、伪标签方法^[22-24]和混合方法^[25-27]五类. 其中一致性正则化和伪标签方法的应用最为广泛.

一致性正则化方法基于平滑假设, 其核心思想是对于同一张图像的不同视角, 模型应该给出相似的预测结果. 因此, 一致性正则化方法根据引入扰动的方式可以分为数据扰动、特征扰动和模型扰动. 首先, 数据扰动是对原始输入图像进行变换, 常见的数据增强技术包括 CutMix^[28]和 ClassMix^[29]等. 这些方法通过混合不同图像来增加训练样本之间的多样性. 然而, 现有的数据增强方法仅在标注或未标注数据内部进行操作, 忽略了标注和未标注数据之间经验分布不匹配的问题. 因此双向复制粘贴 (Bidirectional Copy-Paste, BCP)^[30]将标注数据和未标注数据分别视为前景或背景, 通过双向混合标注和未标注数据来生成新样本, 以减少两种数据的经验分布差异. 其次, 特征扰动是对分割网络的内部特征进行变换. 交叉一致性训练 (Cross-Consistency Training, CCT)^[31]在特征空间进行噪声添加、随机遮挡等操作, 促使主解码器和辅助解码器在不同扰动下的特征空间中产生一致的预测结果. 为了生

成更具挑战性的样本, 扰动和严格均值教师 (Perturbed and Strict Mean Teachers, PS-MT)^[12]引入了一种新的对抗性特征扰动, 用于增加模型在正确分类和对抗性分类结果之间的差异. 最后, 模型扰动是在训练过程中对网络结构或参数进行变换. 交叉伪监督 (Cross Pseudo Supervision, CPS)^[32]通过训练具有相同结构但不同初始化参数的两个网络, 强制网络对同一输入图像产生相似的预测. 而互一致性网络 (Mutual Consistency Network, MC-Net)^[10]是对网络结构进行了改变, 通过使用两个上采样方式不同的解码器对相同的输入图像产生预测偏差, 以此鼓励他们之间的相互一致性. 为了改善现有半监督方法对未标注数据的信息提取能力, 互补一致性网络 (Complementary Consistency Network, CC-Net)^[33]通过在两个辅助解码器中交替使用跳跃连接层来形成互补的模型扰动, 从而帮助模型关注模糊区域. 上述基于一致性正则化的半监督方法仅依赖单一扰动, 无法同时处理多种扰动, 因此自适应双向位移 (Adaptive Bidirectional Displacement, ABD)^[34]被提出. 该方法通过使用可信区域替换模型的不适应区域, 从而生成语义互补的数据, 以此来提升一致性学习的质量.

伪标签方法基于聚类假设, 其核心思想是首先使用标注数据训练模型, 然后利用训练好的模型生成未标注数据的伪标签进行数据集扩展, 并使用新扩展的数据集重新训练模型. 此外, 一些方法还通过当前模型直接将未标注数据的预测结果转化为伪标签. 由于伪标签的质量对于半监督学习的训练至关重要, 因此自训练 (Self-Training, ST)^[22]只选取置信度较高的预测作为未标注数据的伪标签, 以此扩充数据集. 然而只保留置信度较高的像素作为伪标签进行训练, 往往会浪费许多正确的低置信度像素. 为此, 动态互训练 (Dynamic Mutual Training, DMT)^[35]则通过比较不同模型的预测差异识别伪标签中的错误, 从而进行相互训练. 交替多样均值教师 (Alternate Diverse Mean-Teacher, AD-MT)^[36]和可靠半监督互学习 (Reliable Semi-Supervised Mutual Learning, RSSML)^[37]通过评估不同子网预测结果的可靠性, 动态生成高质量的伪标签, 从而充分利用图像中的所有像素, 以此提升半监督学习的效果. 为了进一步简化模型并提高其泛化能力, 先进自训练 (advanced Self-Training, ST++)^[23]选择性使用整体预测稳定性更高的可靠图像进行再训练. 另外, 还有一些方法将伪标签和一致性正则化方法结合起来使用. 如 FixMatch^[13]和修正对比伪监督 (Rectified Contrastive Pseudo Supervision, RCPS)^[38]都鼓励模型在增强数据上的预测结果与弱增强数据生成的伪标签保持一致. 尽管上述方法通过不同的方式提升了伪标签的质量, 但忽视了图像内不同区域的预测复杂度差异, 缺少对

图像中挑战性区域的单独处理.

2.2 不确定性引导的半监督语义分割

在半监督学习中,通过对模型进行不确定性分析,可以充分利用伪标签中包含的信息.常用的不确定性估计方法是蒙特卡洛 Dropout^[39],该方法通过在推理阶段对网络多次应用 Dropout,然后利用生成的多个预测结果的方差来估计不确定性.除此之外,还有一些常见的不确定性估计方法.例如,可以通过计算 KL (Kullback-Leibler) 散度来度量模型预测分布与某个目标分布之间的差异.另外,可以通过计算熵或最大类概率等方式来评估预测结果的置信度水平.并且,还可以通过集成多个不同的模型,使用方差或熵等方式量化其预测结果之间的差异程度.综合运用这些不确定性估计方法,可以更有效地选择和利用未标注数据,最大限度地提升模型的性能和泛化能力.

具体来说,为了让模型学习更可靠的目标,不确定性感知自集成模型 (Uncertainty-Aware self-ensembling Model, UA-MT)^[40]利用蒙特卡洛 Dropout 进行不确定性估计,仅使用不确定性低的像素进行模型训练.由于图像中像素分辨率存在差异,直接进行像素级一致性可能会降低模型的鲁棒性.为了解决这一问题,不确定性修正金字塔一致性 (Uncertainty-Rectified Pyramid Consistency, URPC)^[41]通过计算预测结果的 KL 散度进行不确定性校正,鼓励模型逐步从不同尺度的可靠预测中学习.此外,为了提升伪标签质量,UniMatch^[14]使用最大类概率来评估模型预测结果的置信度,并且只保留弱增强数据预测结果中置信度较高的像素作为伪标签.对于图像中的挑战性区域,网络容易产生错误的预测,因此 MC-Net+^[42]利用不同解码器输出的预测差异来表示模型的不确定性,迫使模型在这些挑战性区域产生一致的预测结果. UA-MT、URPC 和 UniMatch 通过过滤不确定性较高的像素来降低噪声对模型的影响.然而,不确定性较高的像素仍然包含许多有价值的信息,直接过滤会导致信息浪费.与上述方法不同,本文方法将不确定性较高的可疑像素单独筛选出来,鼓励模型对其进行修正,并从中学习额外的知识.

3 方法

3.1 可疑像素相互修正框架

在半监督医学图像分割背景下,定义 3D 医学图像为 $X \in \mathbb{R}^{H \times W \times D}$,其中 H 、 W 和 D 分别指图像的高度、宽度和深度.给定数据集 $D = \{D^l, D^u\}$,其中 $D^l = \{(X_i^l, Y_i^l)\}_{i=1}^N$ 表示标注数据集, $D^u = \{X_i^u\}_{i=1}^M$ 表示未标注数据集, N 和 M 分别表示标注和未标注数据的数量,且 $N \ll M$. $Y_i^l \in \{0, 1, \dots, C-1\}^{H \times W \times D}$ 表示标签, C 表示类数.

当前, CPS 方法^[32]已经成为半监督图像分割的主

流方法之一.该方法通过利用一个网络的预测结果生成伪标签,以此来监督另一个网络的预测结果,从而实现两个网络之间的相互学习,进而提高模型的整体性能.然而,这种方法缺少对图像中挑战性区域的单独处理,使用相同的方式训练同一图像中不同预测难度的像素,可能会给模型带来噪声信息,导致网络在挑战性区域表现不佳.据此, SPMC 使用最大类概率作为置信度对图像中的各个像素进行不确定性估计,并根据置信度大小将不同预测难度的像素分为可疑 (low confidence) 像素和可信 (high confidence) 像素,通过筛选两个网络的预测结果,对标注数据和未标数据中置信度较低的可疑像素进行独立教学,从而修正图像中的不可靠预测,图 2 是 SPMC 的整体框架图.

受前人工作启发^[10,31], SPMC 框架由两个结构不同的独立网络组成,网络 A 和网络 B 使用相同的编码器 E 和上采样方式不同的解码器 D_a 和 D_b .由于不同视角的特征有助于模型降低过拟合风险,并能提高模型的整体性能^[43].因此,为了提升分割网络的泛化能力,在网络 B 的特征图上添加噪声扰动.将原始特征图 F 和扰动版本的特征图 \tilde{F} 分别输入到解码器 D_a 和 D_b 中,从而得到网络的预测结果.该过程表示为

$$F = E(X), \tilde{F} = \text{Noise}(F) \quad (1)$$

$$P_a = D_a(F), P_b = D_b(\tilde{F}) \quad (2)$$

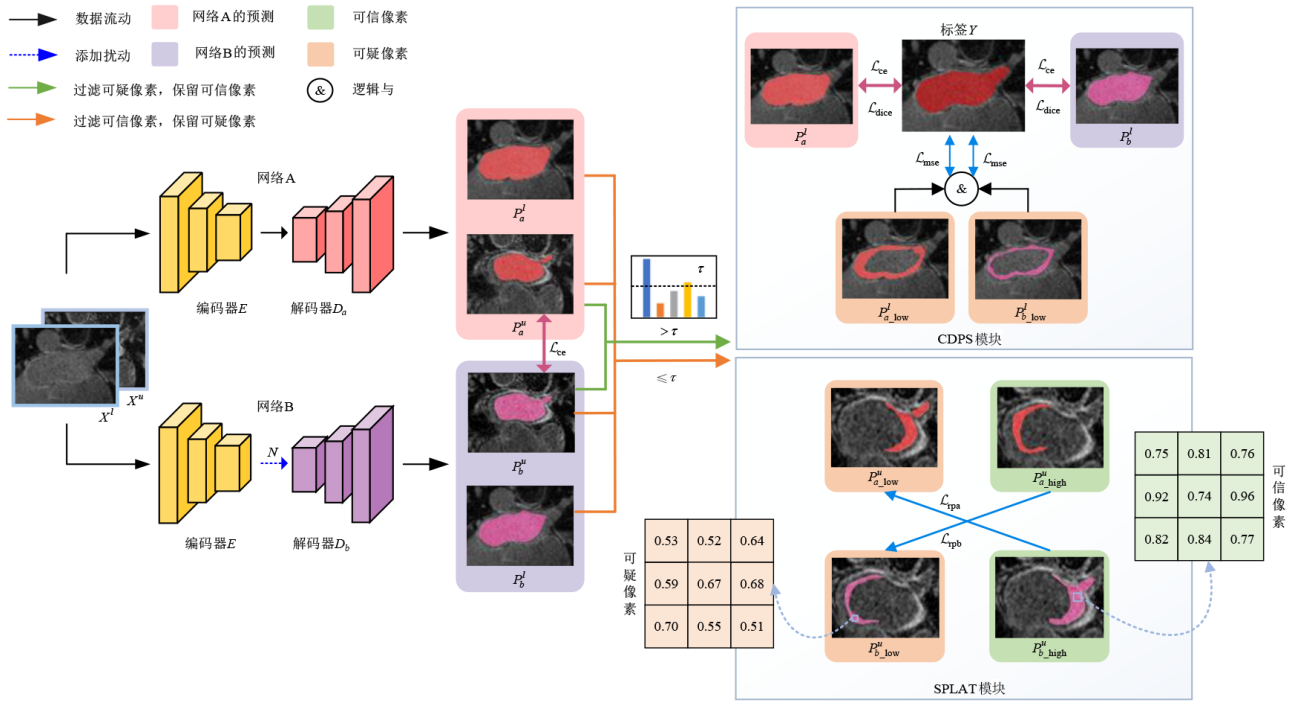
其中,解码器输出包括标注数据和未标注数据的预测结果, $P = P^l \cup P^u$.

通常,半监督分割网络的损失函数包括监督损失 \mathcal{L}_{sup} 和无监督损失 $\mathcal{L}_{\text{unsup}}$.具体来说,对于标注数据的预测结果 P^l ,除了计算交叉熵 (Cross Entropy, CE) 损失和戴斯 (Dice) 损失外,还通过 CDPS 引入均方误差 (Mean Squared Error, MSE) 损失 \mathcal{L}_{mse} ,用于修正网络预测中的共因像素.对于未标注数据的预测结果 P^u ,在计算 CPS 损失 \mathcal{L}_{cc} 的基础上,还使用 SPLAT 引入修正伪监督损失 \mathcal{L}_{rp} ,让网络预测中的可疑像素只向对方的可信像素学习,从而修正两个网络自身的预测偏差.单个网络训练的损失函数为

$$\mathcal{L}_a = \underbrace{\mathcal{L}_{\text{cc}}(P_a^l, Y) + \mathcal{L}_{\text{dice}}(P_a^l, Y) + \beta \mathcal{L}_{\text{mse}}(P_a^l, Y)}_{\mathcal{L}_{\text{sup}}} + \lambda \left(\underbrace{\mathcal{L}_{\text{cc}}(P_a^u, \text{PL}_b) + \beta \mathcal{L}_{\text{rpa}}(P_a^u, \text{SPL}_b)}_{\mathcal{L}_{\text{unsup}}} \right) \quad (3)$$

$$\mathcal{L}_b = \underbrace{\mathcal{L}_{\text{cc}}(P_b^l, Y) + \mathcal{L}_{\text{dice}}(P_b^l, Y) + \beta \mathcal{L}_{\text{mse}}(P_b^l, Y)}_{\mathcal{L}_{\text{sup}}} + \lambda \left(\underbrace{\mathcal{L}_{\text{cc}}(P_b^u, \text{PL}_a) + \beta \mathcal{L}_{\text{rpb}}(P_b^u, \text{SPL}_a)}_{\mathcal{L}_{\text{unsup}}} \right) \quad (4)$$

其中, $\text{PL} = \text{argmax}(P^u)$,表示将预测概率转化成伪标签; SPL 表示使用概率锐化函数^[44]生成的软伪标签;超参数 β 用于平衡 MSE 损失、修正伪监督损失和其他损



注:该框架鼓励模型从可疑像素中学习额外的信息,通过计算预测结果的置信度来获取网络预测中可疑像素的分布情况.此处省略伪标签和软伪标签的转换过程.

图 2 SPMC 整体框架图

失; λ 表示随时间变化的高斯预热函数^[45], 用于平衡监督损失和无监督损失.

SPMC 的总损失函数定义为

$$\mathcal{L} = \mathcal{L}_a + \mathcal{L}_b \quad (5)$$

3.2 CDPS

医学图像具有复杂的解剖结构和特性, 不同区域的预测难度存在显著差异. 对于图像中的挑战性区域, 如果模型不及时修正预测不准确的像素, 这些问题将随着训练的深入不断累积, 从而影响最终的分割性能. 为了解决网络预测偏差的问题, 文献[46]提出将两个网络预测不一致的区域视为可能错误的预测, 并使用标签对这些区域进行自动纠正. 虽然这种方式可以缓解模型的认知偏差, 但它缺少对不同区域预测难度的度量, 当两个网络对同一区域都预测错误时, 这种问题将不能被发现和纠正. 因此, 我们对图像中不同预测难度的像素加以区分, 通过筛查并修正两个网络都无法准确预测的共困像素, 从而解决模型对图像挑战性区域预测困难的问题.

如图 3 所示, 提出了 CDPS 模块, 其中共困像素是指两个网络预测结果中同一位置同时被标记为可疑的像素. 对于两个不同版本的网络预测结果, 可疑像素的分布通常存在差异. 经过特征扰动后的预测结果通常包含更多的可疑像素, 因此, 我们认为只有当两个网络对

同一位置的像素都预测困难时, 才有必要对其进行额外的学习. 由于共困像素是两个网络都难以自我修正的区域, 因此需要使用更为准确的监督信号对其进行修正.

具体来说, 首先对网络 A 和网络 B 的 softmax 输出 P^l 计算其置信度分布. 然后使用阈值 τ 过滤置信度较高的可信像素, 保留置信度较低的可疑像素. 其形式为

$$\mathcal{M}_a^l = \Pi(\max(P_a^l) \leq \tau) \quad (6)$$

其中, $\Pi(\cdot)$ 表示指示函数. 以同样的方式计算 \mathcal{M}_b^l . 然后对得到的两个可疑像素掩膜取交集, 得到一个新的包含两个网络交互信息的掩膜. 其形式为

$$\mathcal{M}^l = \mathcal{M}_a^l \cap \mathcal{M}_b^l \quad (7)$$

所得的掩膜被用于筛查两个网络预测结果中的共困像素

$$P_{a_cdp}^l = P_a^l \odot \mathcal{M}^l, P_{b_cdp}^l = P_b^l \odot \mathcal{M}^l \quad (8)$$

其中, \odot 表示对应位置元素逐个进行乘积运算. 为了获得正确的监督信号, 对标签进行同样的操作.

$$Y_{cdp} = Y \odot \mathcal{M}^l \quad (9)$$

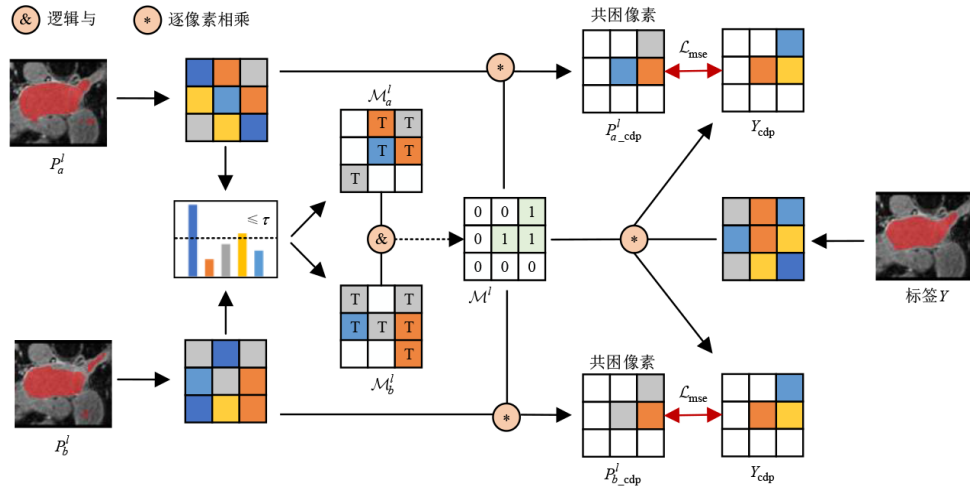
这里使用 MSE 损失对共困像素进行修正, 让网络重点关注这些难以预测的区域并进行正确的引导. 其计算式表示为

$$\mathcal{L}_{mse} = \text{MSE}(P_{a_cdp}^l, Y_{cdp}) \quad (10)$$

$$\mathcal{L}_{\text{mse}} = \text{MSE}(P_{b_cdp}^l, Y_{\text{cdp}}) \quad (11)$$

其中, $\text{MSE}(\cdot)$ 表示 MSE 损失函数.

为了更清楚地表示共困像素的筛查过程, 选取训练集中一样本的预测结果进行展示, 如图 4 所示.



注: 该模块用于筛查两个网络对同一标注数据中都难以预测的区域, 并指导网络对其进行检查与修正. τ 代表置信度过滤阈值. M_a^l 和 M_b^l 分别表示网络 A 和网络 B 的预测结果中可疑像素所在位置. M^l 表示共困像素所在位置.

图 3 CDPS 的结构图

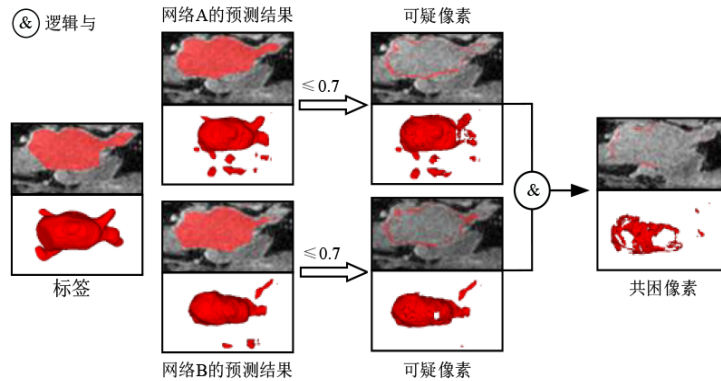


图 4 CDPS 过程图

3.3 SPLAT

文献[32]提出了一种使用两个网络为彼此生成伪标签的 CPS 学习策略. 虽然这种方法可以互补两个网络的预测偏差, 但对于图像中的挑战性区域, 如果两个网络存在类似的错误, 将会影响伪标签的生成质量. 为了提升伪标签的质量, 文献[13]提出只使用不确定性较低的像素作为伪标签, 以减少不准确信息引入的噪声. 虽然这种方式可以提高模型性能, 但不确定性较高的区域可能包含重要信息, 直接丢弃会造成知识浪费. 为了充分利用不确定性较高的像素并提高伪标签质量, 我们对未标注数据中的可疑像素进行额外学习, 为网络提供辅助教学信号, 从而缓解伪标签中噪声带来的影响.

如图 5 所示, 提出了 SPLAT 模块. 由于 SPMC 使用两个结构不同的分割网络, 为了充分挖掘其对相同图

像预测结果的差异性, SPLAT 鼓励每个网络预测结果中的可疑像素只向另一网络对应位置的可信像素学习, 并引入软伪标签思想为网络预测结果中的可疑像素提供更可靠的监督信号. 通过这种方式, 实现了两个网络之间的高质量交互学习, 进而提升了伪标签的生成质量.

具体来说, 首先利用概率锐化函数将网络 A 和网络 B 的 softmax 输出 P_a^u 和 P_b^u 转换为软伪标签 SPL_a 和 SPL_b . 概率锐化函数的计算式为

$$\text{SPL} = \frac{P^{1/T}}{P^{1/T} + (1-P)^{1/T}} \quad (12)$$

其中, T 表示控制模型输出分布的超参数. 通过概率锐化函数, 可以让预测结果的概率分布差异变得更加明显, 使得置信度高的像素变得更加“自信”.

得到软伪标签后, 则可以使用阈值 τ 进行像素筛选.

具体的做法是,首先筛选出两个网络预测结果 P_a^u 和 P_b^u 中小于等于阈值 τ 的像素,即过滤可信像素,保留可疑像素. 其次筛选出软伪标签 SPL_a 和 SPL_b 中大于阈值 τ 的像素,即过滤可疑像素,保留可信像素. 其形式如下:

$$\mathcal{M}_{a_1}^u = \Pi(\max(P_a^u) \leq \tau) \quad (13)$$

$$\mathcal{M}_{b_1}^u = \Pi(\max(P_b^u) \leq \tau) \quad (14)$$

$$\mathcal{M}_{a_2}^u = \Pi(\max(SPL_a) > \tau) \quad (15)$$

$$\mathcal{M}_{b_2}^u = \Pi(\max(SPL_b) > \tau) \quad (16)$$

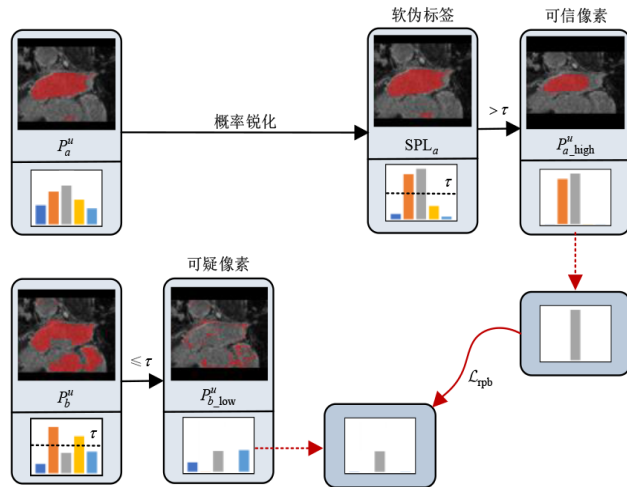
若使用 SPL_b (或 SPL_a) 中的可信像素对 P_a^u (或 P_b^u) 中的可疑像素进行伪监督,必须确保筛选出的可信像素与可疑像素在空间位置上对齐. 因此要保证筛选出的所有像素同时存在于 $\mathcal{M}_{a_1}^u$ 与 $\mathcal{M}_{b_2}^u$ (或在 $\mathcal{M}_{b_1}^u$ 与 $\mathcal{M}_{a_2}^u$) 中. 上述过程通过对不同的掩膜取交集来实现. 具体形式为

$$\mathcal{M}_a^u = \mathcal{M}_{a_1}^u \cap \mathcal{M}_{b_2}^u, \quad \mathcal{M}_b^u = \mathcal{M}_{b_1}^u \cap \mathcal{M}_{a_2}^u \quad (17)$$

这里使用修正伪监督损失对未标注数据中的可疑像素进行指导. 修正伪监督损失表示为

$$\mathcal{L}_{tpa} = \text{MSE}((P_a^u \odot \mathcal{M}_a^u), (SPL_b \odot \mathcal{M}_a^u)) \quad (18)$$

$$\mathcal{L}_{tpb} = \text{MSE}((P_b^u \odot \mathcal{M}_b^u), (SPL_a \odot \mathcal{M}_b^u)) \quad (19)$$



注:红色虚线箭头表示只保留两个过滤后的置信度分布中相同位置的像素.以 SPMC 总框架上方支路为例,该模块利用软伪标签中置信度较高的区域,对另一网络在预测结果置信度较低的区域进行选择性教学.

图5 SPLAT的结构图

4 实验结果与分析

4.1 数据集与预处理

(1) LA 数据集. LA 数据集^[47]由 100 个三维钆增强核磁共振成像扫描及其标签组成,各向同性分辨率为 $0.65 \text{ mm} \times 0.65 \text{ mm} \times 0.65 \text{ mm}$. 遵循 MC-Net^[10]中的实验设置,我们将 80 个样本用于训练,20 个样本用于验

证. 在训练过程中,将数据随机裁剪为 $112 \times 112 \times 80$ 的体素作为输入;在推理阶段使用步长为 $18 \times 18 \times 4$ 的滑动窗口. 本实验使用 10% 和 20% 的训练样本作为标注数据,其余训练样本作为未标注数据.

(2) BraTS 数据集. BraTS 数据集^[48]由 335 例胶质瘤患者的术前核磁共振成像扫描及其标签组成,其中包括 259 例高级别胶质瘤 (High-Grade Glioma, HGG) 和 76 例低级别胶质瘤 (Low-Grade Glioma, LGG). 每位患者的核磁共振成像扫描包括四种模式: T1、T1Gd、T2 和 T2-FLAIR,并由专家逐像素检查标签. 遵循循环原型一致性学习 (Cyclic Prototype Consistency Learning, CPCL)^[49]中的实验设置,我们将 250 个样本用于训练,25 个样本用于验证,60 个样本用于测试. 在训练过程中,将数据随机裁剪为 $96 \times 96 \times 96$ 的体素作为输入;在推理阶段使用步长为 $64 \times 64 \times 64$ 的滑动窗口. 本实验使用 10% 的训练样本作为标注数据,其余训练样本作为未标注数据.

(3) ACDC 数据集. ACDC 数据集^[50]由 100 名患者的心脏动态核磁共振成像扫描及其标签组成,这些标签分为四类,包括背景、右心室、左心室和心肌. 遵循 CTCT (Cross Teaching between CNN and Transformer)^[51]中的实验设置,我们将 70 个样本用于训练,10 个样本用于验证,20 个样本用于测试. 在训练过程中,数据集的 3D 扫描被转化成 2D 切片,所有切片的大小调整为 256×256 像素作为输入. 本实验使用 10% 和 20% 的训练样本作为标注数据,其余训练样本作为未标注数据.

4.2 实验细节和评价指标

所有实验使用 Pytorch 1.7 框架和 NVIDIA GeForce RTX 3090 GPU 实现. 模型使用随机梯度下降 (Stochastic Gradient Descent, SGD) 优化器训练,权重衰减因子设置为 10^{-4} ,动量设置为 0.9,初始学习率设置为 10^{-2} ,迭代次数设置为 30 k.

在 LA 数据集和 BraTS 数据集的分割任务中,主干网络均使用 V-net^[52]. 批大小分别设置为 8 和 4. 当批大小为 8 时,每个批次包含 4 个标注数据和 4 个未标注数据;当批大小为 4 时,每个批次包含 2 个标注数据和 2 个未标注数据. 在 ACDC 数据集的分割任务中,主干网络使用 U-net^[2]. 批大小设置为 16,每个批包含 8 个标注数据和 8 个未标注数据. 高斯预热函数计算式为 $\lambda(t) = 0.1 \times e^{-5(1-t/t_{\max})^2}$,其中 t 表示当前迭代次数, t_{\max} 表示最大迭代次数. 在概率锐化函数中,超参数 T 通常根据经验设置为 0.1. 三个数据集的置信度过滤阈值 τ 依次设置为 0.7、0.65、0.7,损失函数中的超参数 β 均设置为 0.5. 由于网络 A 采用了原始 V-Net (或 U-Net) 结构中的转置卷积 (或双线性) 进行上采样,而网络 B 则使用三线性插值 (或转置卷积) 作为上采样方式,并在特征处理上加

入了噪声扰动. 为了与其他方法公平比较, 我们仅使用网络 A 进行推理.

我们使用四个指标来评估所提方法的性能, 可以将这些指标分为两类: 区域敏感指标和边缘敏感指标. 区域敏感指标包括戴斯相似系数(Dice)和杰卡德相似系数(Jaccard); 边缘敏感指标包括 95% 的豪斯多夫距离(95 HD)和平均表面距离(Average Surface Distance, ASD). Dice 和 Jaccard 指标用于衡量两个样本区域的重叠程度, 因此其值越高则代表分割效果越好. 95 HD 代表两样本之间最近点的最大距离, ASD 代表两个样本表面点之间的平均距离, 因此 95 HD 和 ASD 的值越低则代表分割效果越好.

4.3 实验结果对比

首先在 LA 数据集上对所提方法进行性能评估, 对比方法包括均值教师(Mean Teacher, MT)^[11], 双重任务一致性(Dual-Task Consistency, DTC)^[53], CPS^[32], MC-Net^[10], URPC^[41], 平滑性与类分离网络(Smoothness and class-Separation Network, SS-Net)^[54], 相互校正框架(Mutual Correction Framework, MCF)^[46], 模糊选择一致性均值教师

(Ambiguity-Consistency Mean-Teacher, AC-MT)^[55], 不确定性引导的相互一致性学习(Uncertainty-Guided Mutual Consistency Learning, UG-MCL)^[56], 解剖感知不确定性(Anatomically-aware uncertainty, AAU)^[57]和基于证据的推理学习(Evidential Inference Learning, EVIL)^[58]. 此外, 还使用 V-Net 在 10%、20% 和 100% 的标注数据下训练获得的指标作为参考性能的基线和上限. 从表 1 可得, 与基线模型相比, 我们的方法在四个评估指标上都得到了显著提升. 具体来说, 在使用 10% 的标注数据和 90% 的未标注数据进行训练时, Dice 提高到 90.33%, 提高 13.82 个百分点; 在使用 20% 的标注数据和 80% 的未标注数据进行训练时, Dice 提高到 91.07%, 提高 6.29 个百分点, 接近全监督学习上限 91.55% (使用 100% 的标注数据训练 V-Net). 与最先进的方法相比, SPMC 在两种设定的训练中仍保持最佳性能. 尤其在使用 10% 的标注数据和 90% 的未标注数据训练时, 相较于次优方法(EVIL), Dice 提高 1.58 个百分点. 此外, 由图 6 可知, 我们的方法可以更精细地分割出左心房难以分割的分支或边缘区域, 如图中的画圈部分.

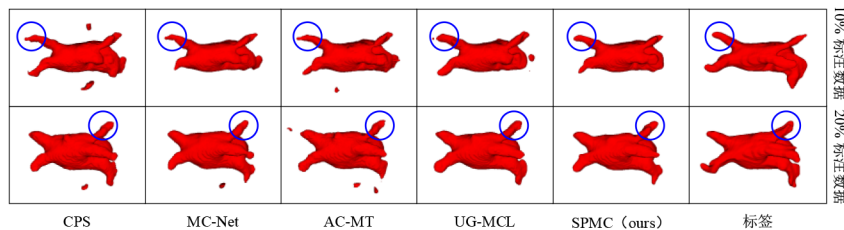


图 6 在 LA 数据集上使用不同方法(部分)训练的可视化结果

为了进一步验证所提方法在不同数据集上的泛化能力, 我们在 BraTS 数据集上进行性能评估, 使用的对比方法与 LA 数据集相同. 相较于 LA 数据集, 脑肿瘤分割更具挑战性, 但我们的方法仍具有较好的表现. 如表 2 所示, 只使用 10% 的标注数据和 90% 的未标注数据进行训练时, SPMC 相较于基线模型, Dice 提高 4.65 个百分点. 此外, 与最先进的方法相比, SPMC 仍占据优势, 相较于次优方法(UG-MCL), Dice 提高 1.03 个百分点. 由图 7 可知, 我们方法的分割效果与真实标签更接近, 尤其对于肿瘤边缘分割的更完整.

除了在 LA 和 BraTS 这样的二分类数据集上进行分割实验外, 我们还在四分类数据集 ACDC 上对所提方法进行评价, 使用的对比方法和 LA 数据集相同. 如表 3 所示, 当使用 10% 的标注数据和 90% 的未标注数据训练时, 与基线模型相比, SPMC 的 Dice 提高 7.04 个百分点; 将标注数据的数量提升至 20% 时, Dice 提高 4.59 个百分点. 与最先进的方法相比, 我们的方法在多个器官的分割任务上仍保持最优. SPMC 在 10%、20% 的标注数据和 90%、80% 的未标注数据下训练时, 与其对应的次优方法

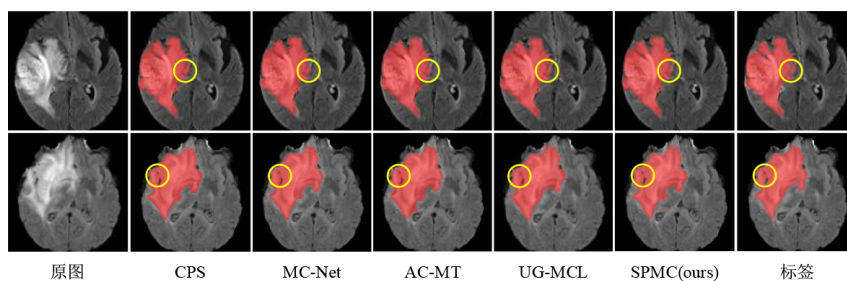


图 7 在 BraTS 数据集上使用不同方法(部分)训练 10% 的标注数据的可视化结果

表 1 在 LA 数据集上与最先进的分割方法比较

方法	标注数据	未标注数据	Dice/% ↑	Jaccard/% ↑	95 HD/体素 ↓	ASD/体素 ↓		
V-Net	8(10%)	0	76.51	62.93	32.17	9.81		
V-Net	16(20%)	0	84.78	74.33	17.37	4.46		
V-Net	80(All)	0	91.55	84.45	5.60	1.86		
MT _[NIPS'2017]	8(10%)	72(90%)	85.27	74.62	14.33	3.39		
DTC _[AAAI'2021]			85.13	74.71	18.95	5.29		
CPS _[CVPR'2021]			86.06	75.99	19.47	5.54		
MC-Net _[MICCAI'2021]			86.49	76.66	11.91	1.99		
URPC _[MIA'2022]			85.02	75.98	15.21	2.95		
SS-Net _[MICCAI'2022]			86.82	76.98	11.62	3.17		
MCF _[CVPR'2023]			86.63	77.00	8.37	2.95		
AC-MT _[MIA'2023]			87.64	78.10	16.59	4.06		
UG-MCL _[AHIM'2023]			87.29	77.65	11.46	3.10		
AAU _[MIA'2024]			86.34	76.32	12.62	3.41		
EVIL _[CBM'2024]			88.75	79.97	11.82	2.80		
SPMC(ours)			90.33 ↑ 1.58	82.46	9.06	1.95		
MT _[NIPS'2017]			16(20%)	64(80%)	88.68	79.83	9.14	2.50
DTC _[AAAI'2021]					88.30	79.34	12.11	3.45
CPS _[CVPR'2021]	89.92	81.93			9.46	2.81		
MC-Net _[MICCAI'2021]	90.35	82.53			6.99	1.86		
URPC _[MIA'2022]	88.43	81.15			8.21	2.35		
SS-Net _[MICCAI'2022]	88.63	79.82			8.89	2.56		
MCF _[CVPR'2023]	89.07	80.76			8.19	2.49		
AC-MT _[MIA'2023]	89.15	80.54			13.86	3.61		
UG-MCL _[AHIM'2023]	89.85	81.73			8.87	2.16		
AAU _[MIA'2024]	88.66	79.92			7.47	2.06		
EVIL _[CBM'2024]	90.28	82.39			6.86	1.82		
SPMC(ours)	91.07 ↑ 0.72	83.74			5.65	1.67		

注:粗体为最优结果,红色为最优结果和次优结果的差值.

表 2 在 BraTS 数据集上与最先进的分割方法比较

方法	标注数据	未标注数据	Dice/% ↑	Jaccard/% ↑	95 HD/体素 ↓	ASD/体素 ↓
V-Net	25(10%)	0	79.23	68.40	15.83	4.34
V-Net	250(All)	0	86.49	77.23	8.68	1.98
MT _[NIPS'2017]	25(10%)	225(90%)	80.49	69.99	11.71	3.46
DTC _[AAAI'2021]			82.53	72.73	12.66	3.12
CPS _[CVPR'2021]			82.70	72.30	12.99	3.46
MC-Net _[MICCAI'2021]			82.51	72.98	14.57	3.78
URPC _[MIA'2022]			82.07	72.11	9.92	2.24
SS-Net _[MICCAI'2022]			81.64	70.95	15.31	4.83
MCF _[CVPR'2023]			81.96	71.76	13.96	3.99
AC-MT _[MIA'2023]			82.09	71.65	15.91	4.81
UG-MCL _[AHIM'2023]			82.85	72.69	11.12	2.28
AAU _[MIA'2024]			82.43	72.33	13.83	4.18
EVIL _[CBM'2024]			82.60	72.46	11.44	2.76
SPMC(ours)			83.88 ↑ 1.03	74.06	10.13	2.72

注:粗体为最优结果,红色为最优结果和次优结果的差值.

表3 在 ACDC 数据集上与最先进的分割方法比较

方法	标注数据	未标注数据	Dice/% ↑	Jaccard/% ↑	95 HD/体素 ↓	ASD/体素 ↓
U-Net	7(10%)	0	81.59	70.74	8.07	2.35
U-Net	14(20%)	0	84.83	75.33	7.61	2.09
U-Net	70(All)	0	91.10	84.04	4.90	1.14
MT _[NIPS'2017]	7(10%)	63(90%)	83.65	73.15	13.45	3.61
DTC _[AAAI'2021]			85.40	75.49	9.91	2.47
CPS _[CVPR'2021]			85.34	75.50	8.78	2.37
MC-Net _[MICCAI'2021]			86.07	76.58	11.48	3.37
URPC _[MIA'2022]			84.25	74.09	4.28	1.31
SS-Net _[MICCAI'2022]			86.95	77.88	4.93	1.56
MCF _[CVPR'2023]			85.18	75.22	10.73	2.78
AC-MT _[MIA'2023]			86.36	76.86	9.64	2.59
UG-MCL _[AHM'2023]			86.84	77.60	7.77	1.94
AAU _[MIA'2024]			86.46	76.99	8.00	2.06
EVIL _[CBM'2024]			87.25	78.18	5.31	1.35
SPMC(ours)			88.63 ↑ 1.38	80.20	4.45	1.19
MT _[NIPS'2017]			14(20%)	56(80%)	86.55	77.49
DTC _[AAAI'2021]	87.67	78.83			6.80	1.93
CPS _[CVPR'2021]	87.03	78.13			6.66	2.12
MC-Net _[MICCAI'2021]	86.58	77.68			15.99	4.93
URPC _[MIA'2022]	86.07	76.77			6.04	1.66
SS-Net _[MICCAI'2022]	87.33	78.45			7.38	2.34
MCF _[CVPR'2023]	87.32	78.43			6.53	2.00
AC-MT _[MIA'2023]	87.99	79.34			7.85	2.14
UG-MCL _[AHM'2023]	88.40	79.86			9.12	2.44
AAU _[MIA'2024]	87.71	78.92			8.51	2.14
EVIL _[CBM'2024]	88.34	79.90			6.46	1.68
SPMC(ours)	89.42 ↑ 1.02	81.42			5.99	1.46

注:粗体为最优结果,红色为最优结果和次优结果的差值.

(EVIL, UG-MCL)相比, Dice 分别提高 1.38 个百分点和 1.02 个百分点.

4.4 消融实验

4.4.1 不同模块的影响

如表 4 所示,我们在 LA 数据集和 ACDC 数据集上对 SPMC 中的不同模块进行了消融实验以显示核心模块的有效性. CPS 通常是利用两个网络预测的差异性进行相互学习,然而如果只使用上采样不同的双网络进行训练,会减小网络的学习空间. 因此我们在第二支路的网络中增加特征扰动,为网络提供新的特征表示以提高其泛化能力. 添加特征扰动后的网络在训练 10% 的标注数据和 90% 的未标注数据时,在 LA 数据集和 ACDC 数据集上, Dice 指标分别提高 0.59 个百分点和 0.52 个百分点.

此外,两个核心模块单独使用时都有助于提高模型的分割性能. 具体来说, CDPS 模块在两个数据集上的 Dice 指标相较于基线模型分别提高 0.56 个百分点和

0.53 个百分点. 为了学习两个网络都难以预测的区域,我们对两个网络预测结果中的共困像素进行额外监督,而不是分别监督每个网络预测结果中的所有可疑像素. 由于这些共困像素的预测难度较大,使用标签对其进行修正是最为准确的. 实验结果表明,这样的方式可以提高网络的分割性能.

仅使用 SPLAT 模块时,模型的分割能力显著增强,相较于基线模型, Dice 指标分别提高 3.14 个百分点和 2.28 个百分点. 未标注数据之间的伪监督是不够准确的,因此我们选择预测结果中的可信像素为另一预测结果中相同位置的可疑像素提供监督信号,确保一个预测中表现较差的像素只向另一预测中表现较好的像素学习,这样有助于降低噪声对模型预测的影响.

当 CDPS 和 SPLAT 模块联合使用时,模型达到最佳的分割效果. 具体来说, CDPS 通过改善模型对标注数据中挑战性区域的识别,提升了模型在分割难度较高区域的预测精度. 这一过程间接为未标注数据提供了更精确

的上下文信息. 由于未标注数据中可能包含与标注数据挑战性区域相似的部分, CDPS 的优化有助于模型在未标注数据中更有效地处理这些区域, 从而提升模型的学习能力, 并进一步帮助模型生成更高质量的伪标签, 增强未标注数据的监督效果. SPLAT 通过提升模型对未标注数据中挑战性区域的学习能力, 显著增强了模型的泛化能力. 在无监督学习场景下, 未标注数据的监督信号相对有限, 导致其挑战性区域的学习难度通常高于标注数据中的类似区域. SPLAT 的优化, 强化了模型在全局范围内对挑战性区域的学习能力, 进一步促进了模型对标注数据的预测性能. 实验结果表明, 当两个模块联合使用时, 在 LA 数据集上的 Dice 指标比基线模型提高 3.6 个百分点, Jaccard 提高 5.46 个百分点, 95 HD 降低 7.48 个体素, ASD 降低 2.76 个体素. 在 ACDC 数据集上, Dice

指标比基线模型提高 3.09 个百分点, Jaccard 提高 4.46 个百分点, 95 HD 降低 9.48 个体素, ASD 降低 2.72 个体素. 值得注意的是, 在 ACDC 数据集上, 相较于单独使用, 两个模块联合使用时的 95 HD 分别降低 6.35 个体素和 5.9 个体素. 由于图像中的可疑像素大多分布在器官的边缘区域, 而 ACDC 是多分类数据集, 它包含的可疑像素比 LA 数据集相对更多, 我们的方法对可疑像素的修正是非常有效的, 因此, ACDC 数据集的边缘敏感指标 95 HD 的改善也相对更好. 如图 8 所示, 我们分别使用 CDPS 和 SPLAT 单独或联合训练时保存的不同迭代次数下的权重进行测试. 结果表明, 随着训练的进行, 单独使用 CDPS 和 SPLAT 进行训练能够修正分割目标的挑战性区域, 并使分割目标的边缘更加清晰 (如图中箭头所指区域), 当两个模块联合训练时能够进一步提升这一效果.

表 4 SPMC 中不同模块的消融实验

数据集	标注数据	未标注数据	V1-V2	Feature-Noise	CDPS	SPLAT	Dice/% ↑	Jaccard/% ↑	95 HD/个体素 ↓	ASD/个体素 ↓
LA	8(10%)	72(90%)	√	×	×	×	86.14	75.97	19.60	5.69
			√	√	×	×	86.73	77.00	16.54	4.71
			√	√	√	×	87.29	77.62	16.49	4.33
			√	√	×	√	89.87	81.72	9.18	2.51
			√	√	√	√	90.33	82.46	9.06	1.95
ACDC	7(10%)	63(90%)	√	×	×	×	85.02	74.92	13.39	3.99
			√	√	×	×	85.54	75.74	13.93	3.91
			√	√	√	×	86.07	76.61	10.80	3.67
			√	√	×	√	87.82	78.97	10.35	2.78
			√	√	√	√	88.63	80.20	4.45	1.19

注: 粗体为最优结果.

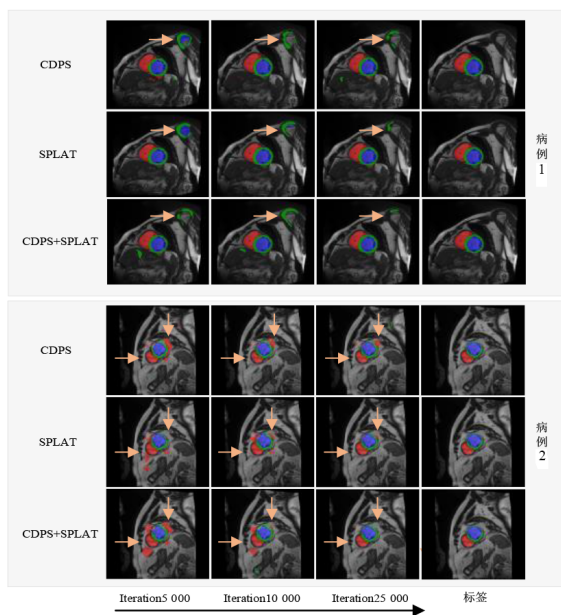


图 8 不同迭代次数下使用不同模块训练 ACDC 数据集的可视化结果

4.4.2 不同辅助教学方式的比较

为了从另一角度展示 SPLAT 模块的有效性, 我们使用不同的 SPLAT 方式进行验证. 如表 5 所示, 分别在 LA 数据集和 ACDC 数据集上使用 10% 的标注数据和 90% 的未标注数据进行训练. 在 SPLAT 中, 通过筛选一个预测结果中置信度较高 (high) 且另一预测中置信度较低 (low) 的像素来获取 \mathcal{M}^l , 即用置信度较高的可信像素监督置信度较低的可疑像素 (high \rightarrow low). 这里我们使用两种不同的方式获取 \mathcal{M}^l 进行对比实验. 第一, 我们仅筛选出一个预测结果中置信度较低的可疑像素获取 \mathcal{M}^l , 对另一预测不做任何的筛选, 不做筛选的像素这里称为普通 (ordinary) 像素, 即用普通像素监督可疑像素 (ordinary \rightarrow low). 第二, 我们仅筛选出一个预测中置信度较高的可信像素获取 \mathcal{M}^l , 对另一预测不做任何的筛选, 即用可信像素监督普通像素 (high \rightarrow ordinary). 实验结果表明, 在 LA 数据集上, 我们的方法比其他两种方法的 Dice 指标分别提高 0.46 个百分点和 3.26 个百分点; 在 ACDC 数据

集上,我们的方法比其他两种方法的Dice指标分别提高0.35个百分点和2.19个百分点,95 HD分别降低0.98个体

素和3.58个体素. 这证明我们的方法使用置信度较高的可信像素去监督置信度较低的可疑像素的有效性.

表5 在LA数据集和ACDC数据集上比较不同SPLAT的方式

数据集	标注数据	未标注数据	方法	Dice/% ↑	Jaccard/% ↑	95 HD/体素 ↓	ASD/体素 ↓
LA	8(10%)	72(90%)	ordinary→low	89.87	81.69	7.03	2.23
			high→ordinary	87.07	77.42	14.28	4.53
			high→low(ours)	90.33	82.46	9.06	1.95
ACDC	7(10%)	63(90%)	ordinary→low	88.28	79.60	5.43	1.39
			high→ordinary	86.44	77.13	8.03	2.90
			high→low(ours)	88.63	80.20	4.45	1.19

注:粗体为最优结果.

4.4.3 超参数的取值分析

本文研究了置信度阈值 τ 在不同的取值下对三种数据集训练指标的影响. 如表6所示,在10%的标注数据和90%的未标注数据上进行实验. 对于LA数据集和ACDC数据集,当置信度阈值 τ 取值为0.7时,Dice和Jaccard指标达到最佳效果. 对于BraTS数据集,当置信度阈值 τ 取值为0.65时,Dice、Jaccard和95 HD指标达到最佳效果. 随着置信度阈值 τ 的逐渐减小,筛选出的可疑像素则越少,此时可能会过滤掉一些有价值的信息,导致模型表现不佳;随着置信度阈值 τ 逐渐增大,筛选出的可疑像素则越来越多,模型可能会更加关注置信度相对较高的像素,导致模型忽略图像中的细微特征,造成图像细节预测不够准确,进而影响模型的整体性能. 因此,在选择置信度阈值 τ 的大小时,需要综合考虑信息保留和模型性能之间的平衡,根据实验结果选择训练效果最佳的

表6 超参数 τ 在三个数据集上的消融实验

数据集	标注数据	未标注数据	τ	Dice/% ↑	Jaccard/% ↑	95 HD/体素 ↓	ASD/体素 ↓
LA	8(10%)	72(90%)	0.60	89.34	80.86	7.98	2.01
			0.65	90.27	82.38	7.16	1.70
			0.70	90.33	82.47	9.06	1.95
			0.75	87.13	77.47	9.54	2.73
			0.80	87.42	77.95	13.19	2.97
BraTS	25(10%)	225(90%)	0.60	82.54	72.64	11.57	2.85
			0.65	83.88	74.06	10.13	2.71
			0.70	83.83	74.15	11.74	3.32
			0.75	83.23	73.12	13.85	3.79
			0.80	82.79	73.02	9.98	2.44
ACDC	7(10%)	63(90%)	0.60	88.37	79.85	4.29	1.18
			0.65	88.44	79.93	3.52	1.07
			0.70	88.63	80.20	4.45	1.19
			0.75	88.24	79.65	3.53	1.17
			0.80	88.55	80.02	6.38	1.55

注:粗体为最优结果.

阈值,从而在不同数据集上取得更优的表现.

此外,本文还研究了损失函数中超参数 β 的影响. 如表7所示,在LA数据集上使用10%的标注数据和90%的未标注数据进行实验,随着 β 的取值在0.3~0.7变化,当 β 取值为0.5时,模型在Dice、Jaccard和ASD指标上表现最佳. 然而,从 β 不同取值下模型性能的整体变化趋势来看,模型对 β 的变化并不敏感. 这可能是因为加权的MSE损失和修正伪监督损失仅作用于图像中少量置信度较低的像素,而常规的监督损失和伪监督损失涵盖图像中的所有像素. 因此,常规的损失项在总损失中占主导地位,尽管超参数 β 发生变化,也不会引起模型整体性能的显著波动. 但是, β 的取值也不能过大或过小,过大会导致模型过度关注挑战性区域,从而忽视整体结构;过小则可能导致模型对挑战性区域学习不足,无法捕捉细节信息.

表7 超参数 β 在LA数据集上的消融实验

β	Dice/% ↑	Jaccard/% ↑	95HD/体素 ↓	ASD/体素 ↓
0.3	90.04	82.05	6.30	1.92
0.4	90.10	82.09	9.40	2.20
0.5	90.33	82.46	9.06	1.95
0.6	90.30	82.42	8.83	2.21
0.7	90.27	82.36	6.46	1.98

注:粗体为最优结果.

5 结束语

在本项研究中,我们提出了SPMC,它充分利用了网络预测中的可疑像素,有效解决了同一图像中不同像素预测难度不一致导致的挑战性区域像素预测不准确的问题,通过对可疑像素进行额外学习,降低了噪声对伪标签质量的影响,进而提高了模型的分割性能. CDPS将标注数据中两个网络预测结果中共同存在的可疑像素视为需要额外学习的区域,引导网络对其进行修正. SPLAT使用未标注数据预测中的可信像素监督另一预测中的可疑像素,以此来提高伪标签质量. 在LA、BraTS和ACDC数据集上的实验结果显示,提出的

SPMC 方法在分割性能上有显著提升. 在未来的研究中,我们计划对可疑像素的筛选方式进一步优化,以提高模型在更具挑战性的分割任务中的性能,并细化对器官或肿瘤边缘的分割.

参考文献

- [1] WANG R S, LEI T, CUI R X, et al. Medical image segmentation using deep learning: A survey[J]. *IET Image Processing*, 2022, 16(5): 1243-1267.
- [2] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[M]//*Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015*. Cham: Springer International Publishing, 2015: 234-241.
- [3] ÇIÇEK Ö, ABDULKADIR A, LIENKAMP S S, et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation[M]//*Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016*. Cham: Springer International Publishing, 2016: 424-432.
- [4] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHSH N, et al. UNet++: Redesigning skip connections to exploit multi-scale features in image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2020, 39(6): 1856-1867.
- [5] LI X Q, HE Q, DAI S M, et al. Semi-supervised semantic segmentation constrained by consistency regularization[C]//*2020 IEEE International Conference on Multimedia and Expo (ICME)*. Piscataway: IEEE, 2020: 1-6.
- [6] LI X M, YU L Q, CHEN H, et al. Transformation-consistent self-ensembling model for semisupervised medical image segmentation[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(2): 523-534.
- [7] SHEN Z Q, CAO P, YANG H, et al. Co-training with high-confidence pseudo labels for semi-supervised medical image segmentation[C]//*Proceedings of the 32nd International Joint Conference on Artificial Intelligence*. Barcelona: International Joint Conferences on Artificial Intelligence Organization, 2023: 4199-4207.
- [8] LU S F, YAN Z Y, CHEN W, et al. Dual consistency regularization with subjective logic for semi-supervised medical image segmentation[J]. *Computers in Biology and Medicine*, 2024, 170: 107991.
- [9] CHEN Y T, MA Y, MEI X G, et al. Triple-task mutual consistency for semi-supervised 3D medical image segmentation[J]. *Computers in Biology and Medicine*, 2024, 175: 108506.
- [10] WU Y C, XU M F, GE Z Y, et al. Semi-supervised left atrium segmentation with mutual consistency training[M]//*Medical Image Computing and Computer Assisted Intervention-MICCAI 2021*. Cham: Springer International Publishing, 2021: 297-306.
- [11] TARVAINEN A, VALPOLA H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 1-10.
- [12] LIU Y Y, TIAN Y, CHEN Y H, et al. Perturbed and strict mean teachers for semi-supervised semantic segmentation[C]//*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2022: 4248-4257.
- [13] SOHN K, BERTHELOT D, LI C L, et al. FixMatch: Simplifying semi-supervised learning with consistency and confidence[EB/OL]. (2020-11-25) [2024-07-02]. <https://arxiv.org/abs/2001.07685v2>.
- [14] YANG L H, QI L, FENG L T, et al. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation[C]//*2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2023: 7236-7246.
- [15] JIN G, LIU C C, CHEN X. Adversarial network integrating dual attention and sparse representation for semi-supervised semantic segmentation[J]. *Information Processing & Management*, 2021, 58(5): 102680.
- [16] XU D, WANG Z L. Semi-supervised semantic segmentation using an improved generative adversarial network[J]. *Journal of Intelligent & Fuzzy Systems*, 2021, 40(5): 9709-9719.
- [17] 黄子臻, 王雷, 张玉坤, 等. 用于医学图像分割的半监督对抗自集成网络[J/OL]. *计算机辅助设计与图形学学报*. (2024-10-29)[2025-04-01]. <https://link.cnki.net/urlid/11.2925.TP.20241029.1352.024>.
- [18] HUANG Z Z, WANG L, ZHANG Y K, et al. Semi-supervised adversarial self-integration network for medical image segmentation[J/OL]. *Journal of Computer-Aided Design & Computer Graphics*. (2024-10-29) [2025-04-01]. <https://link.cnki.net/urlid/11.2925.TP.20241029.1352.024>. (in Chinese)
- [19] WANG W G, ZHOU T F, YU F, et al. Exploring cross-image pixel contrast for semantic segmentation[C]//*2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Piscataway: IEEE, 2021: 7283-7293.
- [20] ZHONG Y Y, YUAN B D, WU H, et al. Pixel contrastive-consistent semi-supervised semantic segmentation[C]//*2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Piscataway: IEEE, 2021: 7253-7262.
- [21] TANG C, ZENG X Y, ZHOU L P, et al. Semi-supervised medical image segmentation via hard positives oriented contrastive learning[J]. *Pattern Recognition*, 2024, 146: 110020.

- [21] LEI T, LIU H L, CHEN Q, et al. Semi-supervised 3D medical image segmentation using shape-guided dual consistency learning[C]//2022 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE, 2022: 1-6.
- [22] LEE D H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks[C]//Workshop on Challenges in Representation Learning. Atlanta: ICML, 2013: 896.
- [23] YANG L H, 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: IEEE, 2022: 4258-4267.
- [24] 魏翔, 王靖杰, 张顺利, 等. ReLSL: 基于可靠标签选择与学习的半监督学习算法[J]. 计算机学报, 2022, 45(6): 1147-1160. WEI X, WANG J J, ZHANG S L, et al. ReLSL: Reliable label selection and learning based algorithm for semi-supervised learning[J]. Chinese Journal of Computers, 2022, 45(6): 1147-1160. (in Chinese)
- [25] LEI T, ZHANG D, DU X G, et al. Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network[J]. IEEE Transactions on Medical Imaging, 2023, 42(5): 1265-1277.
- [26] XIAO H, LI D, XU H, et al. Semi-supervised semantic segmentation with cross teacher training[J]. Neurocomputing, 2022, 508: 36-46.
- [27] 方超伟, 李雪, 李钟毓, 等. 基于双模型交互学习的半监督医学图像分割[J]. 自动化学报, 2023, 49(4): 805-819. FANG C W, LI X, LI Z Y, et al. Interactive dual-model learning for semi-supervised medical image segmentation[J]. Acta Automatica Sinica, 2023, 49(4): 805-819. (in Chinese)
- [28] 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: IEEE, 2019: 6022-6031.
- [29] OLSSON V, TRANHEDEN W, PINTO J, et al. ClassMix: Segmentation-based data augmentation for semi-supervised learning[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2021: 1368-1377.
- [30] BAI Y H, CHEN D W, LI Q L, et al. Bidirectional copy-paste for semi-supervised medical image segmentation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 11514-11524.
- [31] OUALI Y, HUDELLOT C, TAMI M. Semi-supervised semantic segmentation with cross-consistency training[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 12674-12684.
- [32] CHEN X K, YUAN Y H, 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: IEEE, 2021: 2613-2622.
- [33] HUANG H J, CHEN Z G, CHEN C Y, et al. Complementary consistency semi-supervised learning for 3D left atrial image segmentation[J]. Computers in Biology and Medicine, 2023, 165: 107368.
- [34] CHI H Y, PANG J, ZHANG B F, et al. Adaptive bidirectional displacement for semi-supervised medical image segmentation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 4070-4080.
- [35] FENG Z Y, ZHOU Q Y, GU Q Q, et al. DMT: Dynamic mutual training for semi-supervised learning[J]. Pattern Recognition, 2022, 130: 108777.
- [36] ZHAO Z, WANG Z C, WANG L Y, et al. Alternate Diverse Teaching for Semi-Supervised Medical Image Segmentation[M]//Computer Vision-ECCV 2024. Cham: Springer Nature Switzerland, 2024: 227-243.
- [37] HANG W L, BAI K, LIANG S, et al. Reliable semi-supervised mutual learning framework for medical image segmentation[J]. Biomedical Signal Processing and Control, 2025, 99: 106798.
- [38] ZHAO X Y, QI Z X, WANG S, et al. RCPS: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation[J]. IEEE Journal of Biomedical and Health Informatics, 2024, 28(1): 251-261.
- [39] KENDALL A, GAL Y, SHRIDHAR K, et al. What uncertainties do we need in Bayesian deep learning for computer vision? [EB/OL]. (2017-10-05) [2024-07-02]. <https://arxiv.org/abs/1703.04977v2>.
- [40] YU L Q, WANG S J, LI X M, et al. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation[M]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2019. Cham: Springer International Publishing, 2019: 605-613.
- [41] LUO X D, WANG G T, LIAO W J, et al. Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency[J]. Medical Image Analysis, 2022, 80: 102517.
- [42] WU Y C, GE Z Y, ZHANG D H, et al. Mutual consistency learning for semi-supervised medical image segmentation[J]. Medical Image Analysis, 2022, 81: 102530.
- [43] XIA Y D, LIU F Z, YANG D, et al. 3D semi-supervised learning with uncertainty-aware multi-view co-training[C]//

- 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2020: 3646-3655.
- [44] XIE Q Z, DAI Z H, HOVY E, et al. Unsupervised data augmentation for consistency training[EB/OL]. (2020-11-05)[2024-07-02]. <https://arxiv.org/abs/1904.12848v6>.
- [45] LAINE S, AILA T. Temporal ensembling for semi-supervised learning[EB/OL]. (2017-03-15) [2024-07-02]. <https://doi.org/10.48550/arXiv.1610.02242>.
- [46] WANG Y C, XIAO B, BI X L, et al. MCF: Mutual correction framework for semi-supervised medical image segmentation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 15651-15660.
- [47] XIONG Z H, XIA Q, HU Z Q, et al. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging[J]. Medical Image Analysis, 2021, 67: 101832.
- [48] BAKAS S. Brats MICCAI Brain tumor dataset[DB/OL]. IEEE Dataport, (2020-02-28) [2024-07-02]. <https://iee-dataport.org/competitions/brats-miccai-brain-tumor-dataset>.
- [49] XU Z, WANG Y X, LU D H, et al. All-around real label supervision: Cyclic prototype consistency learning for semi-supervised medical image segmentation[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(7): 3174-3184.
- [50] BERNARD O, LALANDE A, ZOTTI C, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?[J]. IEEE Transactions on Medical Imaging, 2018, 37(11): 2514-2525.
- [51] LUO X D, HU M H, SONG T, et al. Semi-supervised medical image segmentation via cross teaching between CNN and transformer[C]//International Conference on Medical Imaging with Deep Learning. New York: PMLR, 2022: 820-833.
- [52] MILLETARI F, NAVAB N, AHMADI S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV). Piscataway: IEEE, 2016: 565-571.
- [53] LUO X D, CHEN J N, SONG T, et al. Semi-supervised medical image segmentation through dual-task consistency[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(10): 8801-8809.
- [54] WU Y C, WU Z H, WU Q Y, et al. Exploring smoothness and class-separation for semi-supervised medical image segmentation[M]//Medical Image Computing and Computer Assisted Intervention - MICCAI 2022. Cham: Springer Nature Switzerland, 2022: 34-43.
- [55] XU Z, WANG Y X, LU D H, et al. Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation[J]. Medical Image Analysis, 2023, 88: 102880.
- [56] ZHANG Y C, JIAO R S, LIAO Q C, et al. Uncertainty-guided mutual consistency learning for semi-supervised medical image segmentation[J]. Artificial Intelligence in Medicine, 2023, 138: 102476.
- [57] ADIGA V S, DOLZ J, LOMBAERT H. Anatomically-aware uncertainty for semi-supervised image segmentation[J]. Medical Image Analysis, 2024, 91: 103011.
- [58] CHEN Y Y, YANG Z Y, SHEN C Y, et al. Evidence-based uncertainty-aware semi-supervised medical image segmentation[J]. Computers in Biology and Medicine, 2024, 170: 108004.

作者简介



杨子瑶 女,2002年9月出生于陕西省西安市。现为陕西科技大学电子信息与人工智能学院硕士研究生。主要研究方向为计算机视觉。

E-mail: 231611038@sust.edu.cn



雷涛 男,1981年11月出生于陕西省渭南市。2011年在西北工业大学获得博士学位,现为陕西科技大学教授,博士生导师。主要研究方向为图像处理、模式识别和计算机视觉。中国电子学会会员编号:E190184479M。

E-mail: leitao@sust.edu.cn



杜晓刚 男,1985年9月出生于陕西省宝鸡市。现为陕西科技大学电子信息与人工智能学院副教授。主要研究方向为机器学习、计算机视觉、医学图像处理。

E-mail: duxiaogang@sust.edu.cn