

doi:10.12068/j.issn.1005-3026.2026.20259022

# 基于MedSAM的高效半监督医学图像 病灶分割方法

贾熹滨<sup>1</sup>, 尹训洁<sup>1</sup>, 范超<sup>1</sup>, 杨正汉<sup>2</sup>

(1. 北京工业大学 计算机学院, 北京 100124; 2. 首都医科大学 附属北京友谊医院, 北京 100050)

**摘要:** 针对半监督病灶分割中教师网络性能较差,难以指导学生网络进行有效分割的问题,本文提出一种高效的半监督医学图像病灶分割方法. 该方法选用特征提取能力更强的MedSAM(medical segment anything model)作为教师网络,构建基于Mamba的轻量级学生网络,通过知识蒸馏提升学生网络分割性能. 针对异构网络特征对齐带来的语义失配问题,提出基于扰动一致的跨架构知识蒸馏策略,将教师特征映射到学生特征空间并对齐扰动响应,提升学生网络特征表达能力以优化分割性能. 此外,针对病灶形态多样及前景背景对比度低导致的分割一致性差问题,提出基于分布的自监督损失进行优化. 在多类医学图像病灶分割数据集上的实验表明,本文方法的分割性能优于现有方法,同时学生网络参数量仅为1.34 M,显著提升了模型效率.

**关键词:** 病灶分割; MedSAM; Mamba; 知识蒸馏; 自监督损失

中图分类号: TP 391.4 文献标志码: A 文章编号: 1005-3026(2026)01-0001-10

## Efficient Semi-supervised Medical Image Lesion Segmentation Method Based on MedSAM

JIA Xi-bin<sup>1</sup>, YIN Xun-jie<sup>1</sup>, FAN Chao<sup>1</sup>, YANG Zheng-han<sup>2</sup>

(1. College of Computer Science, Beijing University of Technology, Beijing 100124, China; 2. Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China. Corresponding author: JIA Xi-bin, E-mail: jiaxibin@bjut.edu.cn)

**Abstract:** In semi-supervised lesion segmentation, the performance of the teacher network is poor, making it difficult for it to guide the student network to perform effective segmentation. To address this issue, an efficient semi-supervised medical image lesion segmentation method was proposed, employing the medical segment anything model (MedSAM), which exhibited superior feature extraction capabilities, as the teacher network. A lightweight student network based on Mamba was constructed, and its segmentation performance was enhanced through knowledge distillation. To address the semantic mismatch caused by feature alignment across heterogeneous networks, a perturbation-consistent cross-architecture knowledge distillation method was introduced. This approach mapped teacher features to the student feature space and aligned perturbation responses, thereby improving the student network's feature representation ability and improving segmentation performance. Additionally, to tackle the challenges of diverse lesion morphologies and low foreground-background contrast, leading to poor segmentation consistency, a distribution-based self-supervised loss was proposed for optimization. Experiments on multiple types of medical image lesion segmentation datasets demonstrate that the proposed method in this paper outperforms existing methods in segmentation performance. Meanwhile, the student network has only 1.34 M parameters, which significantly improves the model efficiency.

**Key words:** lesion segmentation; MedSAM; Mamba; knowledge distillation; self-supervised loss

收稿日期: 2025-06-06

基金项目: 国家自然科学基金资助项目(62476015, 62171298, 82372043, 82371904).

作者简介: 贾熹滨(1969—),女,山西太原人,北京工业大学教授,博士生导师.

通信作者: 贾熹滨, E-mail: jiaxibin@bjut.edu.cn.

医学图像病灶分割因标注数据稀缺而面临严峻挑战.传统全监督深度学习方法在数据不足时难以捕捉病灶的多样性与复杂边界,导致分割精度下降.半监督学习通过结合少量有标数据和大量无标数据成为解决这一问题的有效途径,但其核心难点在于如何生成高质量伪标签.提升无标数据利用效率以及处理病灶区域的前景背景分离.

早期方法主要围绕教师-学生框架优化伪标签质量.MT(mean teacher)<sup>[1]</sup>利用一致性正则化约束学生网络与教师网络的预测一致性,降低标注依赖,但其性能高度依赖教师模型的初始质量,在病灶边界模糊时伪标签噪声显著.UA-MT(uncertainty aware mean teacher)<sup>[2]</sup>在 MT 基础上加入不确定性加权损失,缓解噪声敏感性,但未能有效解决病灶多样性带来的偏差.DAMTN(distribution-aware mean teacher network)<sup>[3]</sup>利用有标数据和无标数据的分布信息来指导模型的学习,以便在训练阶段使模型对有标和无标数据的分割结果的分布尽可能相似,但存在伪标签错误累积循环和双模型计算量大的问题.

后续方法围绕一致性约束与伪标签优化等方面开展,URPC(uncertainty rectified pyramid consistency)<sup>[4]</sup>引入不确定性估计和金字塔一致性约束,增强对无标数据的鲁棒性,但对复杂形态病灶的适应性有限.MSMVT(multi-scale and multi-view Transformer)<sup>[5]</sup>通过生成多尺度原型表示,强化跨尺度一致性以缓解 Transformer 训练不足问题,但对低质量超声影像的边缘检测稳定性不足,且多尺度原型对比学习模块因需多次前向传播导致计算开销较大.SASSNet(shape-aware semi-supervised network)<sup>[6]</sup>通过设计多任务网络,同时预测分割掩码和目标表面的符号距离图,并对输出施加几何形状约束来提升性能,但存在形状先验依赖性强、计算复杂度高、对抗训练容易导致不稳定等问题.MC-Net(mutual consistency network)<sup>[7]</sup>采用多分支解码器模型,通过计算多个解码器输出的统计差异表示模型的不确定性,并在 1 个解码器的概率输出和其他解码器的软伪标签之间应用互一致性进行约束,但在低对比度场景下任务间干扰限制了效果.BCP(bidirectional copy-paste)<sup>[8]</sup>通过有标数据和无标数据间的病灶复制增强样本多样性,但在复杂形态下易引入噪声.CPS(cross pseudo supervision)<sup>[9]</sup>通过交叉网络伪监督实现一致性正则化,在常规

器官分割中显著提升伪标签质量,但对非规则形态病灶的边界适应性不足.CCT(cross-consistency training)<sup>[10]</sup>通过编码器特征扰动一致性训练增强特征判别性,在半监督语义分割中有效提升模型泛化能力,但对纹理稀疏数据的特征分离假设存在局限性.RDMT(residual-driven momentum teacher)<sup>[11]</sup>通过熵-方差双视角不确定性度量联合评估预测可靠性,保留边缘关键结构,并引入 Frobenius 范数优化 EMA(exponential moving average)参数更新,但在训练初期因教师模型未充分收敛可能导致残差补偿误差积累.此外,高计算复杂度限制了其在实时临床应用中的部署效率.CPC-SAM(cross prompting consistency with segment anything model)<sup>[12]</sup>通过交叉提示策略和提示一致性正则化,利用 SAM 的少样本学习能力,显著提升了极少标注场景下的医学图像分割性能,但对提示位置的敏感性可能导致伪标签噪声,尤其在复杂结构区域,提示优化策略的鲁棒性有待提升.SemiSAM(semi-supervised segment anything model)<sup>[13]</sup>将 SAM 作为辅助监督分支,结合均值教师框架,通过自动生成提示点和一致性损失优化分割精度,该方法依赖 SAM-Med3D 的适应性,通用性可能受限,且在噪声提示处理上仍需进一步优化以提高稳定性.

近年来一些方法进一步探索因果推理与边界优化.CauSSL(causality-inspired semi-supervised learning)<sup>[14]</sup>通过建模图像特征因果关系提升模型鲁棒性,在低标注比例下显著降低伪标签噪声,但对非规则边界病灶适应性不足.BaPC(boundary-aware prototype consistency)<sup>[15]</sup>利用边界感知原型学习优化边界分割,在脑肿瘤等模糊边界场景表现优异,但对低对比度皮肤病变适应性有限.

高效架构方面,MedSAM<sup>[16]</sup>基于 Transformer 扩展 SAM 实现通用分割,依赖提示输入且参数量大.Swin Transformer<sup>[17]</sup>通过移窗自注意力建模全局特征,在结直肠息肉分割中表现突出,但实时性不足.Swin UNETR(swin unet Transformer)<sup>[18]</sup>通过分层 Swin Transformer 编码器实现长程依赖建模,在 BraTS(brain tumor segmentation)多模态脑肿瘤分割任务中达到 SOTA(state-of-the-art)性能,但计算复杂度约是传统 3D U-Net 的 3.8 倍,且参数量高达 62 M.Mamba<sup>[19]</sup>基于状态空间模型提供线性复杂度,其在视觉领域的相关方法如 RGU-Mamba(region-guided unet Mamba)<sup>[20]</sup>,

VMamba (visual Mamba)<sup>[21]</sup> 和 H-vmunet (high-order vision Mamba UNet)<sup>[22]</sup>取得了一定成果,但是在医学图像病灶分割上仍然存在缺乏针对性预训练策略、特征分离能力差、小目标漏检率高等问题.SliceMamba<sup>[23]</sup>引入双向切片扫描(BSS)与神经架构搜索(NAS),但动态切片策略导致训练时间增加.此外,这些高效架构在实际应用中往往面临计算资源与性能的权衡,特别是在资源受限的场景下,难以同时满足高精度和低延迟的需求.为了在性能与效率之间取得平衡,知识蒸馏成为一种有效的解决方案,通过将预训练大模型的知识迁移至轻量模型,提升小模型的性能.传统的知识蒸馏方法(如KD(knowledge distillation)<sup>[24]</sup>, FitNet<sup>[25]</sup>)主要聚焦于 logits 层的对齐,无法高效利用教师网络的中间特征和空间信息.针对这一问题,跨架构特征层知识蒸馏逐渐受到关注,相关方法如 PKD-BERT (patient knowledge distillation for BERT)<sup>[26]</sup>, Manifold Distillation<sup>[27]</sup>, Attention-Transfer<sup>[28]</sup>通过中间特征或注意力图的直接对齐,试图保留教师网络的语义和空间细节以提升学生网络的表达能力,但架构异构容易导致语义失配,影响知识迁移效果.同时,医学图像病灶区域的复杂性和前景背景的低对比度进一步加剧了问题,使得直接特征对齐难以捕捉病灶的空间细节和深层语义信息,从而限制了分割性能的提升.

针对上述问题,本文提出了一种半监督医学

图像病灶分割方法,结合 MedSAM 和基于 Mamba 架构的轻量级分割网络 Tiny-MUNet,通过跨架构知识蒸馏和自监督约束实现高效分割.设计了一种闭环半监督训练框架,利用 MedSAM 生成病灶伪标签,驱动 Tiny-MUNet 在有标数据 and 无标数据上的联合优化,有效降低了对有标数据的依赖性,缓解了数据稀缺问题.针对 Transformer 与 Mamba 特征空间不匹配的挑战,提出跨架构知识蒸馏策略,通过特征投影模块将 MedSAM 的 Transformer 特征映射至 Mamba 空间,并以扰动响应对齐替代直接特征对齐,提升知识迁移效率,增强学生网络对复杂病灶的建模能力.针对病灶分割中伪标签一致性不足及前景背景分离困难的难点,引入自监督前背景方差均值损失,约束前景与背景区域的方差最小化并拉大均值差异,利用无标数据的统计特性增强分割一致性.

## 1 方 法

### 1.1 研究基础

MedSAM<sup>[16]</sup>是基于 SAM 改进的通用医学图像分割模型,专门针对医学图像的特点进行了优化,能够通过提示(如边界框)生成高质量病灶分割掩码.其架构由图像编码器、提示编码器和掩码解码器三部分组成,在病灶分割中展现了强大的泛化能力.MedSAM 网络结构如图 1 所示.

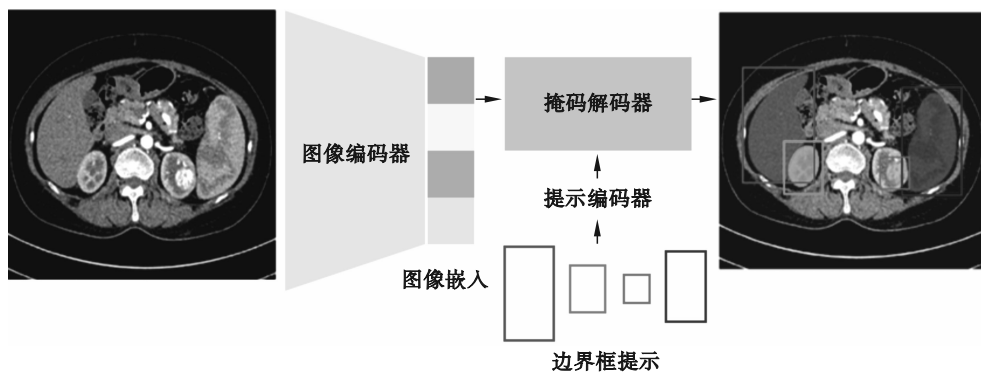


图 1 MedSAM 网络结构

Fig. 1 MedSAM network structure

图像编码器采用 Vision Transformer 结构,将输入图像平均分成  $16 \times 16$  的补丁,通过嵌入层映射为特征向量,随后送入多层 Transformer 模块. Transformer 模块的核心是多头自注意力 (multi-head self-attention, MHSA) 机制,通过计算补丁间的全局依赖关系,捕捉病灶区域的长距离上下文信息.提示编码器将边界框坐标编码为条件嵌

入,与图像特征结合,为掩码解码器提供定位信息.掩码解码器由轻量级 Transformer 模块组成,融合图像编码器的特征和提示编码器的嵌入,通过上采样生成像素级分割掩码.MedSAM 在大量医学图像上预训练,泛化能力强,可为无标数据生成高质量伪标签.

Mamba<sup>[20]</sup>是一种基于选择性状态空间模型

(selective state space model) 的高效序列建模方法,旨在替代 Transformer 的高计算成本,同时保持长程依赖建模能力,其结构如图 2 所示.

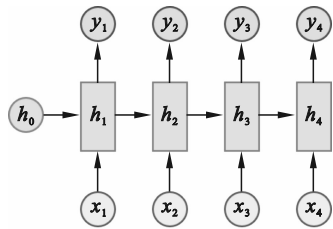


图 2 Mamba 状态空间模型

Fig. 2 Mamba state space model

Mamba 的核心是状态空间模型,通过隐状态逐层传递输入特征并生成输出.其创新点在于选择性机制,通过 MLP (multi-layer perceptron) 动态生成参数矩阵,使模型能自适应输入特征中的关键信息,尤其适合医学图像中病灶区域的针对性建模.此外, Mamba 通过离散化处理和硬件优化的并行扫描算法,将计算复杂度降至线性,显著提升了效率.这种设计在保持长程依赖建模能力的同时,实现了高效计算,使其成为轻量级医学图

像分割网络的理想选择.

## 1.2 网络架构

Tiny-MUNet 是一个基于 Mamba 的轻量级 U 形网络,在降低计算复杂度和参数数量的同时,保持与 MedSAM 特征层数一致以便进行跨架构知识蒸馏.网络由 4 层 Mamba 模块组成编码器,每层提取不同尺度的特征.解码器通过上采样和跳跃连接融合编码器特征,逐步恢复分辨率,最终输出分割掩码.

图 3 展示了半监督医学图像病灶分割网络框架,整体框架以 MedSAM 作为教师网络,利用其预训练能力生成伪标签, Tiny-MUNet 作为学生网络.输入包括有标数据(标注形式为边界框)和无标数据. MedSAM 初始通过边界框提示生成伪标签, Tiny-MUNet 通过跨架构知识蒸馏模块 (FPKDM) 接收知识,并结合自监督前背景方差均值损失 (FBVM) 进行优化.训练若干轮后,学生网络 Tiny-MUNet 的分割结果用于提取无标数据的边界框提示,反馈至 MedSAM 生成更高质量的伪标签,形成闭环训练流程,进一步提升训练效果.

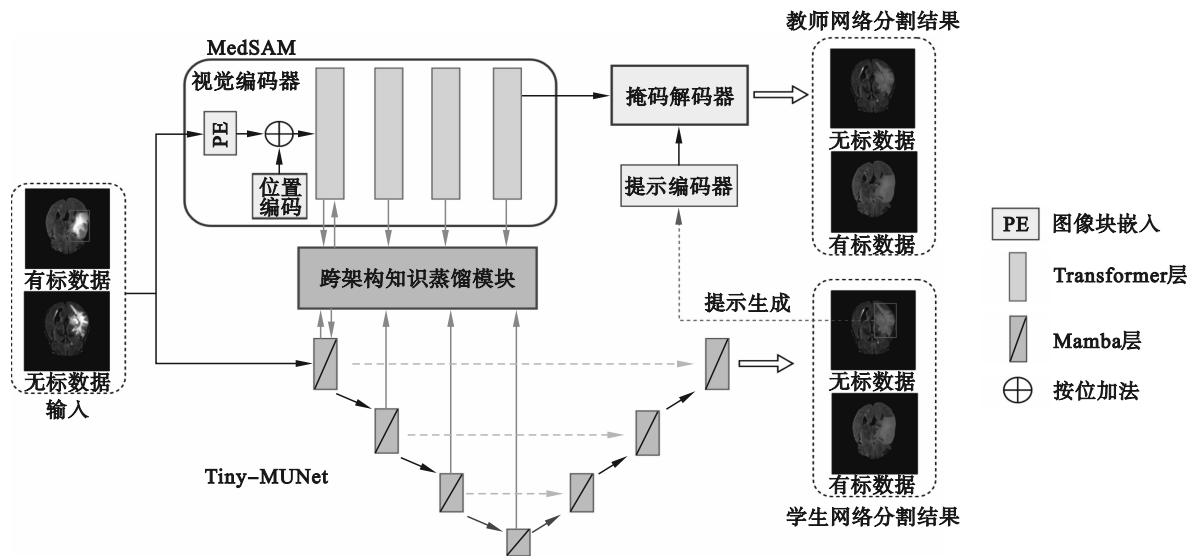


图 3 基于 MedSAM 的半监督医学图像病灶分割网络框架

Fig. 3 Semi-supervised medical image lesion segmentation network framework based on MedSAM

### 1.2.1 跨架构知识蒸馏模块

由于 MedSAM (基于 Transformer) 和 Tiny-MUNet (基于 Mamba) 特征空间异构,受文献[29-30]启发,本文设计了一种跨架构知识蒸馏策略,通过特征扰动知识蒸馏模块 (feature perturbation knowledge distillation module, FPKDM) 实现高效知识迁移,如图 4 所示.

FPKDM 首先对 MedSAM 的特征进行预处理,展平调整格式;然后使用分组线性投影,每小

块的空间位置共享一个全连接层来转换通道数,组的大小设定为 16 个位置共享一层;接着加入一层 Dropout 来提高模型的稳定性;最后输出调整后的特征. MedSAM 的 Transformer 特征经过上述特征空间投影操作映射到 Tiny-MUNet 的 Mamba 特征空间,减少语义失配.随后在投影后的最前层特征上施加随机扰动(如高斯噪声),生成扰动特征,并通过特征重注入前向传播,分别在 Transformer 和 Mamba 特征空间中捕捉后续层的扰动响应差异.通

过元素级减法获取 2 个网络的扰动响应差异来计算损失,优化知识迁移,确保 Tiny-MUNet 有效吸收 MedSAM 的全局建模能力,计算公式如下:

$$L_{RD} = \frac{1}{L} \sum_{l=1}^L (1 - \text{CosSim}(\Delta F_l^t, \Delta F_l^s)), \quad (1)$$

$$\text{CosSim}(x, y) = \frac{x \cdot y}{\|x\|_2 \|y\|_2}. \quad (2)$$

其中:  $L_{RD}$  为扰动响应差异损失;  $\Delta F_l^t$  和  $\Delta F_l^s$  分别为教师和学生网络第  $l$  层特征在扰动前后的变化;  $L$  为特征层数;  $\text{CosSim}(x, y)$  为余弦相似度. MedSAM 的预训练特征为 Tiny-MUNet 提供了丰富的语义信息,提升了知识蒸馏效率和分割性能.

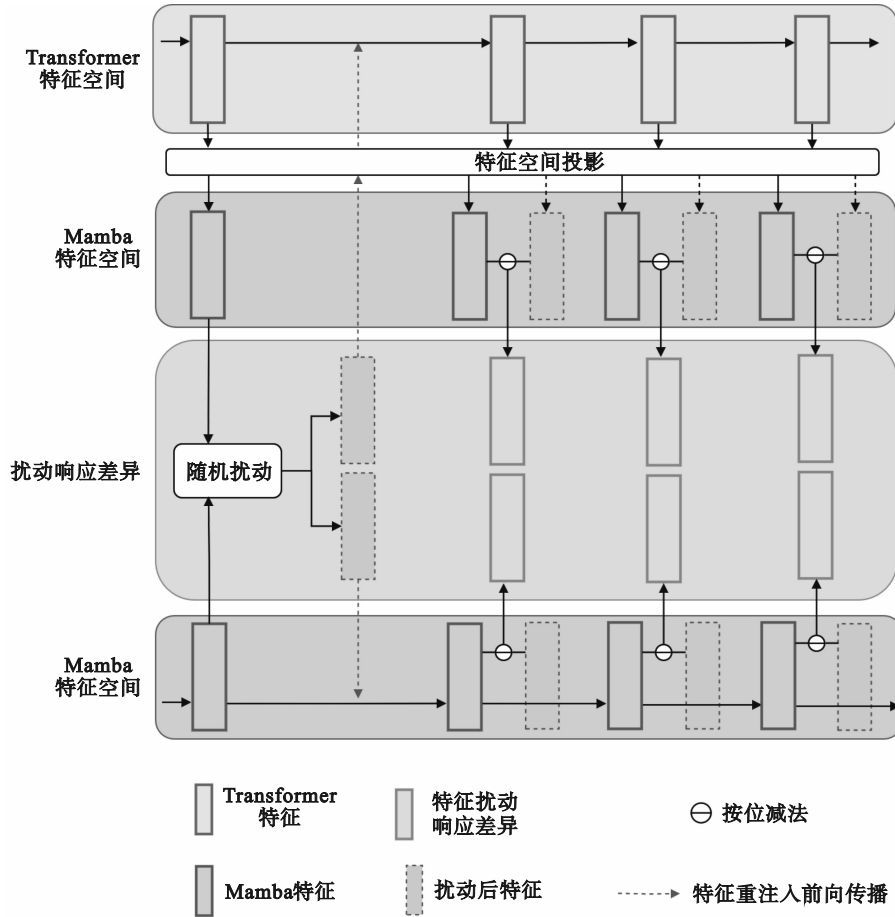


图 4 跨架构知识蒸馏模块

Fig. 4 Cross-architecture knowledge distillation module

### 1.2.2 前背景方差均值损失

为解决伪标签一致性和前景背景分离问题,受文献[31]启发,本文提出了一种自监督前背景方差均值 (foreground-background variance-mean, FBVM) 损失. 该损失以输入原图和分割掩码 (作为前景区域) 为基础, 结合网络的预测输出, 利用分割掩码将原图像素分为前景 (病灶) 和背景两类. 随后, 通过统计约束提升分割性能: 约束前景和背景区域的方差最小化以增强一致性, 同时拉大均值差异以提高分离度. 此外, 通过 Sigmoid 函数对均值差异施加非线性约束, 增强前景背景分离的稳定性, 并通过掩码 (基于输入图像的有效像素) 避免噪声干扰. 对于输入图像的每个通道,

损失计算公式如下:

$$S_{\text{var}} = \frac{1}{N} \sum_{b=1}^N \left( \frac{1}{C} \sum_{c=1}^C (\text{Var}_c(F_b) + \text{Var}_c(B_b)) \right), \quad (3)$$

$$S_{\text{mean}} = \frac{1}{N} \sum_{b=1}^N \left( \frac{1}{C} \sum_{c=1}^C \frac{1}{1 + e^{-k(\text{Mean}_c(F_b) - \text{Mean}_c(B_b))}} \right), \quad (4)$$

$$L_{\text{FBVM}} = S_{\text{var}} - S_{\text{mean}}. \quad (5)$$

其中:  $S_{\text{var}}$  是方差约束项;  $S_{\text{mean}}$  是均值约束项;  $F_b$  为第  $b$  个样本的前景区域像素集合;  $B_b$  为第  $b$  个样本的背景区域像素集合;  $N$  为批量大小;  $C$  为通道数;  $\text{Var}_c$  和  $\text{Mean}_c$  分别为第  $c$  通道的前景背景方差和均值;  $k$  为超参数;  $L_{\text{FBVM}}$  (foreground-background variance-mean loss) 为自监督前背景方差均值损失.

### 1.3 总损失

训练过程采用闭环策略,分为初始训练和闭环优化两个阶段.在初始训练阶段,基于有标数据,MedSAM利用边界框提示生成伪标签,Tiny-MUNet通过相关损失进行有限监督学习,持续10个轮次以提升初始性能.在闭环优化阶段,首先由学生网络推理无标数据,生成预测分割掩码.然后从预测分割掩码中提取边界框,作为MedSAM的新提示输入,生成更高质量的伪标签.使用新伪标签和自监督损失继续优化Tiny-MUNet,形成前后推动的闭环.闭环策略有效解决了MedSAM对无标数据提示生成的问题,逐步提升伪标签质量.总损失分为有标数据损失和无标数据损失两部分,具体损失和总损失如下:

$$L_{lb/ub} = L_{KD} + L_{RD} + L_{FBVM}, \quad (6)$$

$$L_{total} = \lambda_1 L_{lb} + \lambda_2 L_{ub}, \quad (7)$$

其中: $L_{KD}$ 为输出层知识蒸馏损失(基于KL(Kullback-Leibler)散度); $L_{total}$ 为总损失; $L_{lb}$ 和 $L_{ub}$ 分别为有标数据损失和无标数据损失;超参数 $\lambda_1, \lambda_2$ 分别为1和0.5,通过实验调优确定.

## 2 实验与分析

### 2.1 数据集和实验设置

本文使用MSD(medical segmentation decathlon)脑肿瘤、HAM10000皮肤病和Kvasir-SEG内窥镜3个公共数据集验证提出方法的分割性能.MSD脑肿瘤分割子任务数据集包含484例多模态MRI(magnetic resonance imaging)数据(T1, T2, FLAIR(fluid-attenuated inversion recovery)),提供像素级分割掩码,划分为80%训练集(387例)和20%测试集(97例).HAM10000是一个皮肤病分割数据集,包含10 015张图像,涵盖基底细胞癌、黑色素瘤等多种病灶,选取6 000张,划分为80%训练集(4 800张)和20%测试集(1 200张).Kvasir-SEG是一个内窥镜数据集,包含1 000张胃肠道息肉图像及其像素级分割掩码,分辨率从332像素×487像素到1 920像素×1 072像素不等,划分为80%训练集(800张)和20%测试集(200张).这3个数据集覆盖脑肿瘤、多种皮肤病变和胃肠道息肉,模态分别为多序列MRI图像和RGB(red green blue)图像,病灶类型多样,能够全面验证方法的适应性和鲁棒性.数据预处理和实验设置方面,所有输入图像分辨率统一调整为256像素×256像素,并进行标

准化处理(均值为0,标准差为1).采用SGD(stochastic gradient descent)优化器,学习率为0.001,动量为0.99,权重衰减为0.000 03.使用CosineAnnealingLR调度器,最大迭代周期100,最小学习率0.000 01.

### 2.2 评估指标

为定量评估模型的分割性能,本文采用医学图像分割领域常用的2个指标,Dice相似系数(Dice similarity coefficient, Dice)和Hausdorff距离(Hausdorff distance, HD).Dice相似系数衡量预测分割掩码与真实掩码的重叠程度,计算公式如下:

$$\text{Dice} = \frac{2|P \cap G|}{|P| + |G|}, \quad (8)$$

其中: $P$ 为预测分割掩码; $G$ 为真实掩码; $|\cdot|$ 为像素数量.Dice值范围为 $[0, 1]$ ,值越高表示分割精度越高.Hausdorff距离衡量预测分割边界与真实边界之间的最大距离,反映分割结果的边界精确性.计算公式如下:

$$\text{HD}(P, G) = \max(\sup_{p \in P} \inf_{g \in G} d(p, g), \sup_{g \in G} \inf_{p \in P} d(g, p)). \quad (9)$$

其中: $d(p, g)$ 为点 $p$ 和 $g$ 之间的欧几里得距离; $\sup$ 和 $\inf$ 分别表示上确界和下确界.HD值越小,表示边界误差越小,分割结果越精确.

### 2.3 对比实验

为了验证本文方法在半监督医学图像分割领域的优势,将本文方法与多种对比方法在3个数据集上的性能进行了比较(见表1~表3),对比方法包括全监督基线(Baseline U-Net)和半监督方法(UA-MT, SemiSAM-MT, SASSNet, URPC, MC-Net+, BCP, CauSSL, BaPC, CPC-SAM).对比实验在不同标注比例(10%和20%有标数据)下进行,所有结果均为测试集上的平均值,Dice以百分比表示,HD以像素为单位.

表1~3给出了本文方法和其他对比方法在MSD脑肿瘤、HAM10000皮肤病和Kvasir-SEG数据集上的量化指标结果.由实验结果可知,本文方法在3个数据集上均表现出色.在MSD数据集上,10%和20%有标数据下的Dice分别为83.46%和85.65%,HD分别为13.60像素和12.34像素;在HAM10000数据集上,10%和20%有标数据下的Dice分别为87.58%和90.32%,HD分别为10.06像素和9.84像素;在Kvasir-SEG数据集上,10%和20%有标数据下的Dice分别为87.63%和88.45%,HD分别为7.59像素和6.36像素,均显著优于所有对比方法.特别是在低标注

比例下展现出良好的鲁棒性,验证了跨架构知识蒸馏和前背景方差均值损失在提升伪标签质量和分割一致性方面的作用.值得关注的是,在 3 个数据集上,本文方法在半监督设置下的性能均优于全监督的 Baseline U-Net,这一超越全监督性能的关键在于 MedSAM 的预训练能力. MedSAM 在大量医学图像数据上进行预训练,具备强大的泛化能力,能够为无标数据生成高质量伪标签,为 Tiny-MUNet 提供可靠的初始监督信号.结合闭环训练策略, Tiny-MUNet 逐步获取 MedSAM 的通用分割知识,弥补了标注数据不足的限制.

表 1 在 MSD 脑肿瘤分割数据集上的对比实验结果  
Table 1 Comparative experimental results on MSD brain tumor segmentation dataset

方法	标注比例/%	Dice/% ↑	HD/像素 ↓
Baseline U-Net	100	80.28	15.24
	10	64.48	20.53
	20	66.01	19.31
UA-MT	10	70.12	17.65
	20	72.29	16.18
SemiSAM-MT	10	72.47	15.34
	20	74.88	13.97
SASSNet	10	70.04	19.58
	20	71.61	17.66
URPC	10	67.43	18.01
	20	70.07	15.26
MC-Net+	10	72.89	15.52
	20	73.27	13.86
BCP	10	81.48	14.01
	20	82.21	13.57
CauSSL	10	80.18	14.25
	20	81.93	13.87
BaPC	10	80.29	13.96
	20	81.47	13.08
CPC-SAM	10	81.73	12.16
	20	83.68	11.59
Ours	10	<b>83.46</b>	<b>13.60</b>
	20	<b>85.65</b>	<b>12.34</b>

注:黑体表示结果最好;↑表示越大越好;↓表示越小越好.下同.

为进一步评估本文方法的计算效率,本文对比了不同模型在参数量(Para/M)和计算复杂度(MACs/G)方面的性能差异.由表 4 可知,本文方法(Ours)的参数量仅为 1.34 M,计算复杂度为 10.01 G,其参数量为所有模型中最低,MACs 值亦优于所有对比方法.尤其值得注意的是,本文

方法在参数量和计算复杂度上较 MedSAM 实现了显著压缩.这种轻量化突破主要得益于 Tiny-MUNet 的 Mamba 模块设计,其线性复杂度特性相比传统 Transformer 的二次复杂度,在保证分割精度的前提下,将 MACs 从典型 Transformer 模型的百 G 量级压缩至 10.01 G,更适合医学影像的实时处理需求.

表 2 在 HAM10000 皮肤病数据集上的对比实验结果  
Table 2 Comparative experimental results on HAM10000 skin disease dataset

方法	标注比例/%	Dice/% ↑	HD/像素 ↓
Baseline U-Net	100	78.57	10.24
	10	72.47	16.16
	20	74.54	14.02
UA-MT	10	79.15	15.45
	20	81.84	12.30
SemiSAM-MT	10	79.81	14.84
	20	82.05	12.43
SASSNet	10	82.93	15.75
	20	83.36	11.61
URPC	10	77.48	14.26
	20	79.69	11.84
MC-Net+	10	83.71	12.62
	20	84.24	11.04
BCP	10	85.65	12.20
	20	87.03	11.15
CauSSL	10	86.25	11.52
	20	87.56	10.36
BaPC	10	85.87	13.84
	20	87.40	11.28
CPC-SAM	10	86.15	11.79
	20	87.91	10.35
Ours	10	<b>87.58</b>	<b>10.06</b>
	20	<b>90.32</b>	<b>9.84</b>

为直观展示本文方法的分割效果,在 MSD 脑肿瘤数据集上进行了可视化分析,比较了本文方法与对比方法在 10% 和 20% 有标数据比例下的分割结果.图 5 展示了该数据集上的输入图像、真实掩码(ground truth, GT)、对比方法的预测结果和本文方法的预测结果. MSD 脑肿瘤 FLAIR 模态图像边界模糊、对比度低.结果显示,对比方法存在多种问题:在 10% 和 20% 有标数据下, U-Net 和 URPC 漏分割现象严重,尤其在 10% 数据比例下病灶区域丢失较多; UA-MT 和 SASSNet 在模糊边界处误分割明显,特别是在低标注比例下表现更差; MC-Net+, BCP, CauSSL 和 BaPC 存

在边界误差和过分割问题,在 10% 数据比例下过分割现象更为突出.相比之下,本文方法在 2 种数据比例下均表现优异,掩码与真实掩码高度一致,边界平滑,误分割显著减少.CPC-SAM 在 10% 和 20% 有标数据比例下利用 SAM 的少样本学习和交叉提示策略,其分割能力优于 U-Net 和 URPC,但因对提示位置敏感,伪标签噪声明显,导致边界不精确,存在部分病灶漏分问题.SemiSAM 边界优于 UA-MT 和 SASSNet,但噪声提示处理不足,低标注比例下模糊边界易误分,通用性受限.在所有数据集上,本文方法均优于对比方法,边界更精确、病灶更完整、噪声更少,而对比方法普遍存在漏分割、过分割、边界不平滑和背景噪声问题,难以适应复杂场景,尤其在低标注比例下表现不佳.

表 3 在 Kvasir-SEG 数据集上的对比实验结果  
Table 3 Comparative experimental results on Kvasir-SEG dataset

方法	标准比例/%	Dice/% ↑	HD/像素 ↓
Baseline U-Net	100	83.79	6.84
	10	69.52	11.64
	20	73.27	7.01
UA-MT	10	81.39	11.55
	20	83.84	8.95
SemiSAM-MT	10	81.64	10.48
	20	83.46	9.77
SASSNet	10	79.24	10.76
	20	81.44	9.94
URPC	10	78.04	16.97
	20	79.58	13.58
MC-Net+	10	83.49	9.55
	20	84.75	9.07
BCP	10	81.32	10.12
	20	83.12	9.53
CauSSL	10	84.19	12.68
	20	85.36	10.36
BaPC	10	84.95	8.94
	20	85.73	8.01
CPC-SAM	10	83.68	10.24
	20	85.07	9.64
Ours	10	<b>87.63</b>	<b>7.59</b>
	20	<b>88.45</b>	<b>6.36</b>

## 2.4 消融实验

为验证本文方法中各组件的有效性,在 MSD

脑肿瘤、HAM10000 皮肤病和 Kvasir-SEG 数据集(20% 有标数据)上进行了消融实验,分析 MedSAM 预训练能力、跨架构知识蒸馏模块(FPKDM)和前背景方差均值(FBVM)损失的贡献.表 5 展示了不同设置下的 Dice 和 HD 结果,其中“w/o”表示移除对应组件,“w/o MedSAM”表示用未预训练的 SAM 替代 MedSAM.

表 4 不同方法在性能上的对比实验结果  
Table 4 Comparative experimental results on performance of different methods

方法	Para/M ↓	MACs/G ↓
MedSAM	37.50	116.38
U-Net	3.27	12.54
URPC	1.82	11.02
UA-MT	3.27	13.76
SASSNet	3.27	11.71
MC-Net+	3.21	15.26
BCP	5.69	16.61
CauSSL	4.36	15.92
BaPC	7.63	11.47
Ours	<b>1.34</b>	<b>10.01</b>

由表 5 数据可知,MedSAM 预训练是模型性能的核心支撑.移除后,分割精度在模糊边界与复杂形态场景下显著下降,验证了其医学先验知识对伪标签质量的关键作用.进一步去除 FPKDM 时,模型对全局上下文建模能力减弱,尤其在形态多样病灶中出现明显过分割,凸显了扰动对齐策略对知识迁移的有效性.此外,移除前背景统计约束(FBVM)后,低对比度区域的误分割率上升,表明该损失函数通过增强前景背景分离能力,显著提升了边界鲁棒性.完整方法在 3 个数据集上均取得最佳性能,表明 MedSAM 的预训练能力、FPKDM 和 FBVM 的协同作用共同提升了分割性能.MedSAM 预训练提供了高质量伪标签,FPKDM 实现了知识迁移,FBVM 增强了前景背景分离,三者缺一不可.消融实验结果证明本文所提方法的合理性和模块的有效性.

为了进一步评估 Tiny-MUNet 在有限监督学习中通过相关损失训练的初始性能对模型的影响,本文在 3 个数据集上(20% 有标数据)进行消融实验,比较不同训练轮次(10, 20, 30, 40)下的性能,并与不进行初始训练(Ours)进行对比.实验结果如表 6 所示.

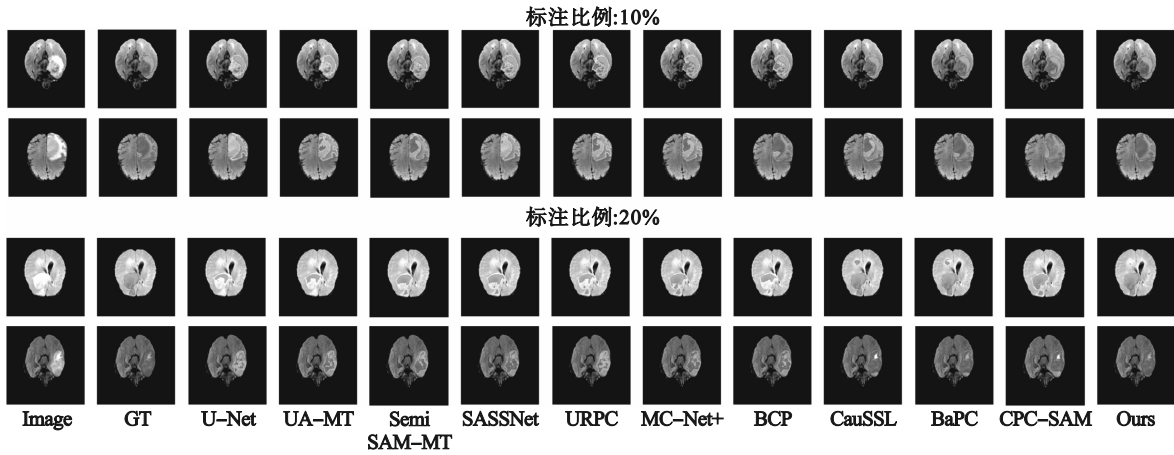


图 5 不同方法在 MSD 数据集上的分割结果对比

Fig. 5 Comparison of segmentation results of different methods on MSD datasets

表 5 MSD 脑肿瘤、HAM10000 皮肤病和 Kvasir-SEG 数据集消融实验结果

Table 5 Ablation experiments results on MSD brain tumor, HAM10000 skin disease, and Kvasir-SEG datasets

方法	(Dice/%)/(HD/像素)		
	MSD	HAM10000	Kvasir-SEG
w/o MedSAM	57.12/30.82	70.34/27.96	60.21/20.48
w/o FPKDM	82.47/13.95	85.25/13.47	84.51/15.63
w/o FBVM	83.50/13.12	87.33/11.04	87.97/13.27
Full	<b>85.65/12.34</b>	<b>90.32/9.84</b>	<b>88.45/6.36</b>

表 6 Tiny-MUNet 初始性能对模型影响的消融实验  
Table 6 Ablation experiments on influence of Tiny-MUNet initial performance on model

轮次	(Dice/%)/(HD/像素)		
	MSD	HAM10000	Kvasir-SEG
10	85.54/13.02	88.74/10.18	88.31/6.77
20	85.58/12.79	88.76/9.97	88.40/6.54
30	85.51/12.71	88.84/10.01	<b>88.47/6.39</b>
40	85.59/ <b>12.28</b>	89.97/9.93	88.43/6.40
0(Ours)	<b>85.65/12.34</b>	<b>90.32/9.84</b>	<b>88.45/6.36</b>

实验结果显示, Tiny-MUNet 的初始性能对模型性能的提升效应较为有限. Dice 和 HD 的改善幅度总体上并不显著, 且在大多数情况下, 某些初始训练轮次下的表现甚至不如没有进行初始训练的结果. 这表明调整 Tiny-MUNet 的初始性能并未带来显著的性能增益, 在某些情况下甚至可能逊于直接联合训练的效果, 反映出模型对初始性能的依赖性较弱, 能够在无需复杂预训练的情况下, 通过有限监督学习直接适应多种数据集, 降低了训练过程中的前期准备成本, 为实际应用提供了更大的便利性.

### 3 结 语

针对医学图像病灶分割面临标注数据稀缺、

病灶形态复杂和前景背景对比度低等难题, 本文提出了一种半监督学习框架, 结合 MedSAM 的预训练能力和基于 Mamba 架构的轻量级分割网络 Tiny-MUNet, 通过 FPKDM 和 FBVM, 有效解决了上述问题. MedSAM 的预训练能力提供了高质量伪标签, FPKDM 实现了知识迁移, FBVM 增强了前景背景分离能力, Tiny-MUNet 的轻量化设计则降低了参数量并确保了计算效率. 实验结果表明, 本文方法在 MSD 脑肿瘤、HAM10000 皮肤病和 Kvasir-SEG 数据集上均取得了优异性能. 以 20% 有标数据为例, 本文方法在 MSD 数据集上的 Dice 达 85.65%, HD 为 12.34 像素; 在 HAM10000 数据集上的 Dice 达 90.32%, HD 为 9.84 像素; 在 Kvasir-SEG 数据集上的 Dice 达 88.45%, HD 为 6.36 像素, 显著优于对比方法. 消融实验进一步验证了各组件的有效性, 参数量和计算复杂度分析则突出了方法的轻量化优势. 本文方法在提升分割精度和效率的同时, 降低了标注需求, 具有重要的学术价值和临床意义, 为医学图像病灶分割提供了高效解决方案, 未来可进一步探索其在更多模态和病种上的应用.

### 参考文献:

[1] Tarvainen A, Valpola H. Mean teachers are better role models: weight-averaged consistency targets improve semi-

- supervised deep learning results [C]// Proceedings of the International Conference on Learning Representations. Vancouver, 2018:6–11.
- [ 2 ] Yu L Q, Wang S J, Li X M, 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, 2019: 605–613.
- [ 3 ] 赵小明, 石培炼, 王丹丹, 等. 分布感知均值教师网络的半监督医学影像分割[J]. 中国图象图形学报, 2025, 30(2): 575–588.  
(Zhao Xiao-ming, Shi Pei-lian, Wang Dan-dan, et al. Distribution-aware mean teacher networks for semi-supervised medical image segmentation [J]. *Journal of Image and Graphics*, 2025, 30(2): 575–588.)
- [ 4 ] Luo X D, Liao W J, Chen J N, et al. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency [C]//Medical Image Computing and Computer Assisted Intervention—MICCAI 2021. Cham: Springer, 2021: 318–329.
- [ 5 ] 李飞翔, 降爱莲. MSMVT: 多尺度和多视图 Transformer 半监督医学图像分割框架[J]. 计算机工程与应用, 2025, 61(2): 273–282.  
(Li Fei-xiang, Jiang Ai-lian. MSMVT: semi-supervised framework with multi-scale and multi-view Transformer for medical image segmentation[J]. *Computer Engineering and Applications*, 2025, 61(2): 273–282.)
- [ 6 ] Li S L, Zhang C Y, He X M. Shape-aware semi-supervised 3D semantic segmentation for medical images [C]//Medical Image Computing and Computer Assisted Intervention—MICCAI 2020. Cham: Springer, 2020: 552–561.
- [ 7 ] 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.
- [ 8 ] 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). Vancouver, 2023: 11514–11524.
- [ 9 ] 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). Nashville, 2021: 2613–2622.
- [ 10 ] Ouali Y, Hudelot C, Tami M. Semi-supervised semantic segmentation with cross-consistency training [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 12671–12681.
- [ 11 ] 徐新辉, 曾智勇, 林征宇. 基于一致性正则化的半监督医学图像分割方法[J]. 中国医学物理学杂志, 2025, 42(6): 784–790.  
(Xu Xin-hui, Zeng Zhi-yong, Lin Zheng-yu. Semi-supervised medical image segmentation method based on consistency regularization [J]. *Chinese Journal of Medical Physics*, 2025, 42(6): 784–790.)
- [ 12 ] Miao J Z, Chen C, Zhang K L, et al. Cross prompting consistency with segment anything model for semi-supervised medical image segmentation [C]//Medical Image Computing and Computer Assisted Intervention—MICCAI 2024. Cham: Springer, 2024: 167–177.
- [ 13 ] Zhang Y C, Yang J, Liu Y C, et al. SemiSAM: enhancing semi-supervised medical image segmentation via SAM-assisted consistency regularization [C]//2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Lisbon, 2024: 3982–3986.
- [ 14 ] Miao J Z, Chen C, Liu F R, et al. CauSSL: causality-inspired semi-supervised learning for medical image segmentation [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, 2023: 21369–21380.
- [ 15 ] Wang Y C, Xiao B, Bi X L, et al. Boundary-aware prototype in semi-supervised medical image segmentation [J]. *IEEE Transactions on Image Processing*, 2024, 33: 5456–5467.
- [ 16 ] Ma J, He Y T, Li F F, et al. Segment anything in medical images[J]. *Nature Communications*, 2024, 15(1): 654–659.
- [ 17 ] Liu Z, Lin Y T, Cao Y, et al. Swin Transformer: hierarchical vision transformer using shifted windows [C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, 2022: 9992–10002.
- [ 18 ] Hatamizadeh A, Nath V, Tang Y C, et al. Swin UNETR: swin transformers for semantic segmentation of brain tumors in MRI images [C]//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Cham: Springer, 2022: 272–284.
- [ 19 ] Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. (2023–12–01) [2024–12–24]. <https://arxiv.org/abs/2312.00752>.
- [ 20 ] Gao J H, Cai Y H, Zhao Z K, et al. RGU–Mamba: an U–Mamba network with region-based training optimized for domain generalization applied to myocardial scar and edema segmentation [C]//Comprehensive Analysis and Computing of Real-World Medical Images. Cham: Springer, 2025: 77–86.
- [ 21 ] Jiao J B, Liu Y, Liu Y F, et al. VMamba: visual state space model [C]//Advances in Neural Information Processing Systems. Vancouver, 2024: 103031–103063.
- [ 22 ] Wu R K, Liu Y H, Liang P C, et al. H–vmunet: high-order vision Mamba UNet for medical image segmentation [J]. *Neurocomputing*, 2025, 624: 129447.
- [ 23 ] Fan C, Yu H Y, Huang Y, et al. SliceMamba with neural architecture search for medical image segmentation [J]. *IEEE Journal of Biomedical and Health Informatics*, 2025, 29(10): 7446–7458.
- [ 24 ] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network [EB/OL]. (2015–03–09) [2024–12–24]. <https://arxiv.org/abs/1503.02531>.
- [ 25 ] Romero A, Ballas N, Kahou S E, et al. FitNets: hints for thin deep nets [EB/OL]. (2014–12–19) [2024–12–14]. <https://arxiv.org/abs/1412.6550>.
- [ 26 ] Sun S Q, Cheng Y, Gan Z, et al. Patient knowledge distillation for BERT model compression [EB/OL]. (2019–08–25) [2024–12–24]. <https://arxiv.org/abs/1908.09355>.
- [ 27 ] Hao Z, Guo J, Jia D, et al. Learning efficient vision transformers via fine-grained manifold distillation [C]// Advances in Neural Information Processing Systems. New Orleans, 2022: 9164–9175.
- [ 28 ] Zagoruyko S, Komodakis N. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer [EB/OL]. (2016–12–12) [2024–12–20]. <https://arxiv.org/abs/1612.03928>.
- [ 29 ] Hao Z, Guo J, Han K, et al. One-for-all: bridge the gap between heterogeneous architectures in knowledge distillation [C]//Advances in Neural Information Processing Systems. New Orleans, 2023: 79570–79582.
- [ 30 ] Liu Y F, Cao J J, Li B, et al. Cross-architecture knowledge distillation [J]. *International Journal of Computer Vision*, 2024, 132(8): 2798–2824.
- [ 31 ] Zhai S W, Wang G T, Luo X D, et al. PA–seg: learning from point annotations for 3D medical image segmentation using contextual regularization and cross knowledge distillation [J]. *IEEE Transactions on Medical Imaging*, 2023, 42(8): 2235–2246.