



多模态论文阅读

研究内容：混合域半监督医学图像分割

——医学影像领域

汇报人：谢宇涵

时间：2025年05月29日



一、文献介绍



《Steady Progress Beats Stagnation: Mutual Aid of Foundation and Conventional Models in Mixed Domain Semi-Supervised Medical Image Segmentation》

-----CVPR, 2025

关键概念

01

半监督医学图像分割 (SSMIS)

SSMIS针对医学图像标注成本高，在某一域上仅有少量的有标注数据和大量的未标记数据，通过一致性正则化和伪标签等方法提升分割精度。

Setting	Limited Annotations	Domain Shift	Unknown Domain Labels
SSMIS	✓	✗	-
UDA	✗	✓	✗
LE-UDA	✓	✓	✗
MiDSS	✓	✓	✓

02

混合域半监督医学图像分割 (MiDSS)

MiDSS是一个更通用的框架，面向由于设备参数、患者群体和疾病严重程度的差异，标记数据和未标记数据来自不同的分布的情况。

有限数量的标记样本来自单个域，而大量未标记样本来自多个混合域。



一、文献介绍

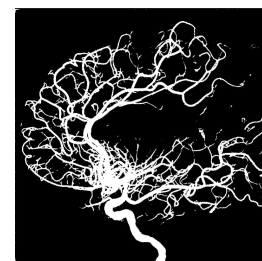
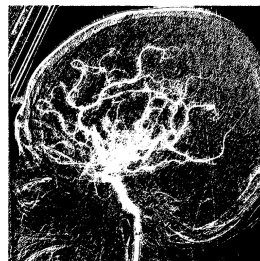


研究背景

01

SSMIS现有方法

从头开始训练传统模型（例如U-Net），使用有标签数据进行预训练，并为无标签数据生成伪标签。



02

现有缺陷

域偏移使得传统模型在早期训练时容易导致模型生成的伪标签质量较差，在缺少正确的伪标签引导学习时，会进一步导致后续模型的训练效果变差。

03

大型预训练视觉基础模型

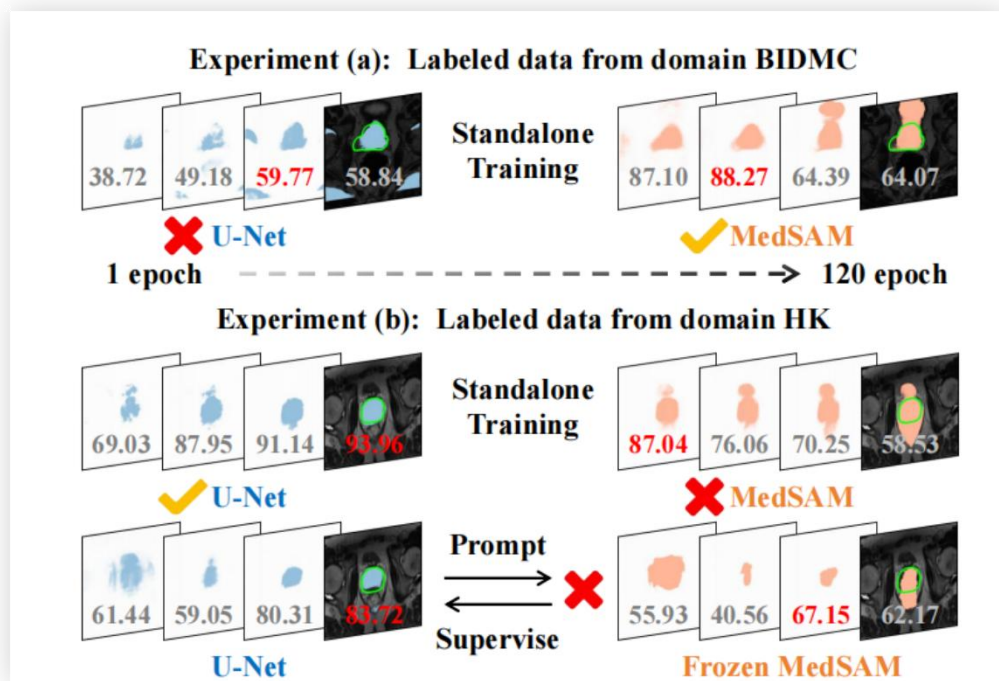
其通过在大规模视觉图像上进行预训练，具备强大视觉特征提取和理解能力，可以有效适应医学分割等下游任务（如MedSAM）。



一、文献介绍



基础模型和传统模型在 MiDSS 中的表现



标记数据来自 BIDMC 域:

- U-Net 会过拟合标记数据, 导致性能不佳。
- MedSAM 由于其固有的大量先验知识, 在训练早期阶段表现出卓越的分割能力。

标记数据来自 HK 域:

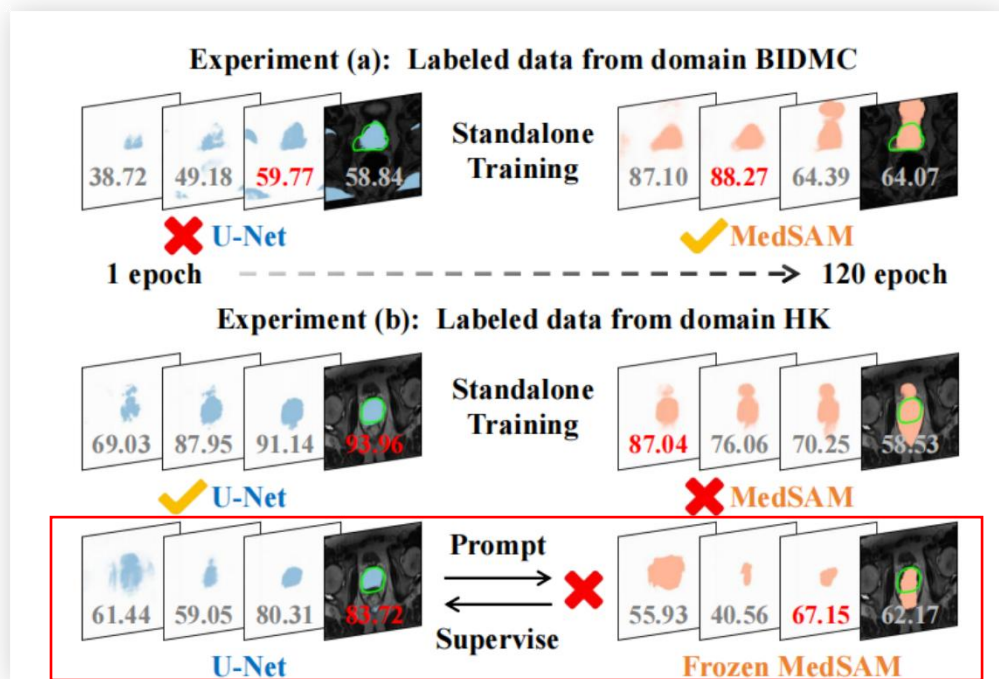
- MedSAM 无法纠正高置信度的错误预测, 阻碍了性能提升。(“自信但错误”)
- U-Net 能够积极纠正高不确定性的错误预测, 达到更高的性能上限。



一、文献介绍



基础模型和传统模型在 MiDSS 中的表现



标记数据来自 HK 域:

- 冻结基础模型的伪标签来指导传统模型的训练。
- U-Net 生成的伪标签为 MedSAM 提供边界框提示，而 MedSAM 反过来为 U-Net 提供额外的监督信号。
- 在这种训练方案下，MedSAM 生成的伪标签会导致高置信度错误的“冻结”传递，进而导致 U-Net “被动适应”这种缺陷。



一、文献介绍



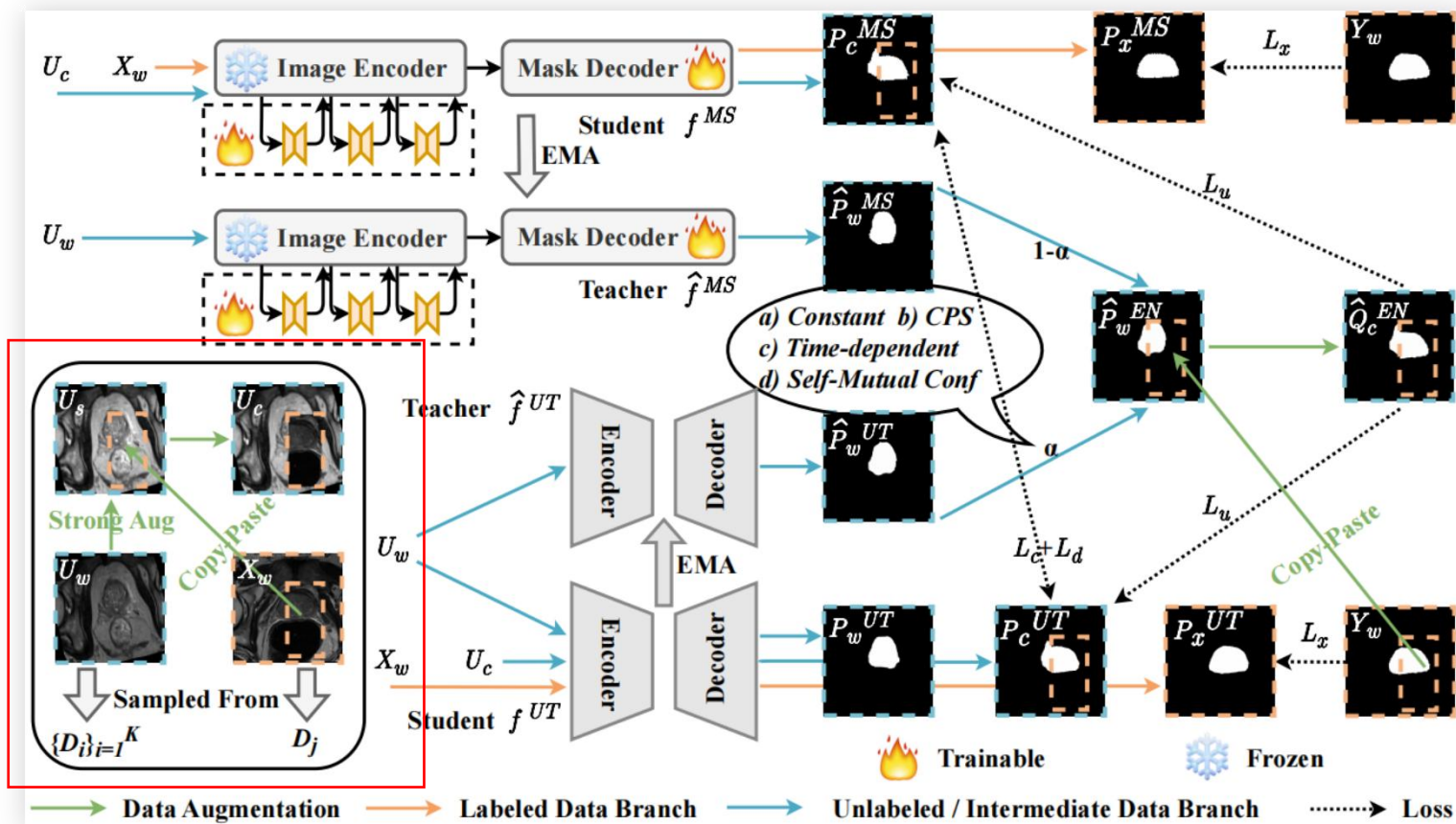
主要贡献

- 本文提出了一种**混合域半监督医学图像分割的协同训练框架**，其中基础模型（例如 MedSAM）和传统模型（例如 U-Net）进行协同训练（SynFoC）。
- 本文在训练过程中动态调整每个模型的主导地位：在训练**早期**，MedSAM 起主导作用，以确保 U-Net 的训练质量；在**后期**，U-Net 起主导作用，纠正高置信度错误，释放 MedSAM 的性能潜力。
- 为了共同提升模型的表征能力，本文采用**一致性正则化**来加强信息共享。对于预测一致的区域，鼓励更高的置信度可以提供更可靠的优化方向。



二、研究方法

混合域半监督医学图像分割的协同训练框架SynFoC



数据增强模块

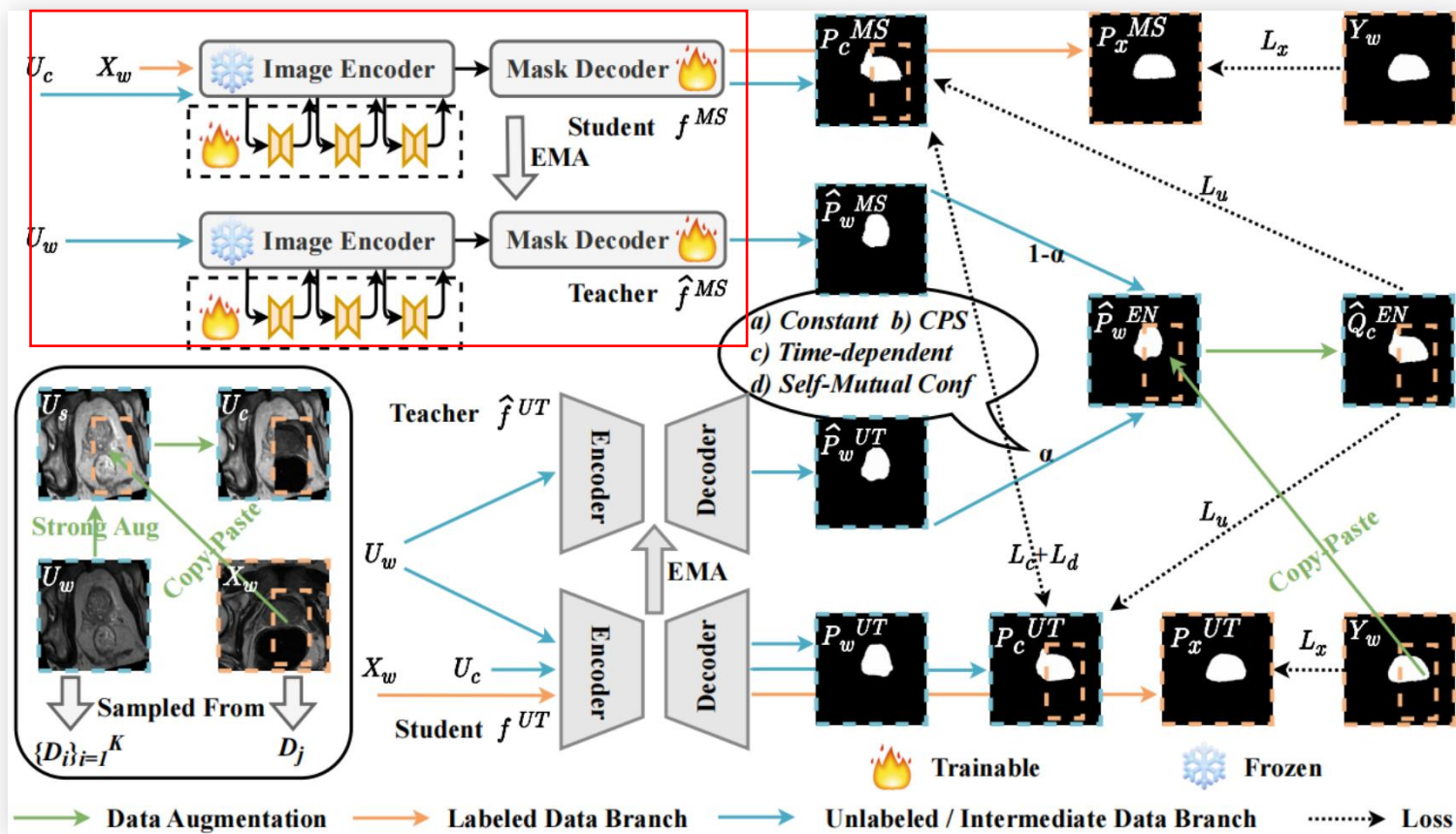
原始数据 $(X, Y), U$ 经过弱增强得到 $(X_w, Y_w), U_w$, U_w 经过强增强得到 U_s , U_s 与 X_w 进行复制粘贴得到中间图像 U_c 。

$$U_c = X_w \odot M + U_s \odot (1 - M)$$



二、研究方法

混合域半监督医学图像分割的协同训练框架SynFoC



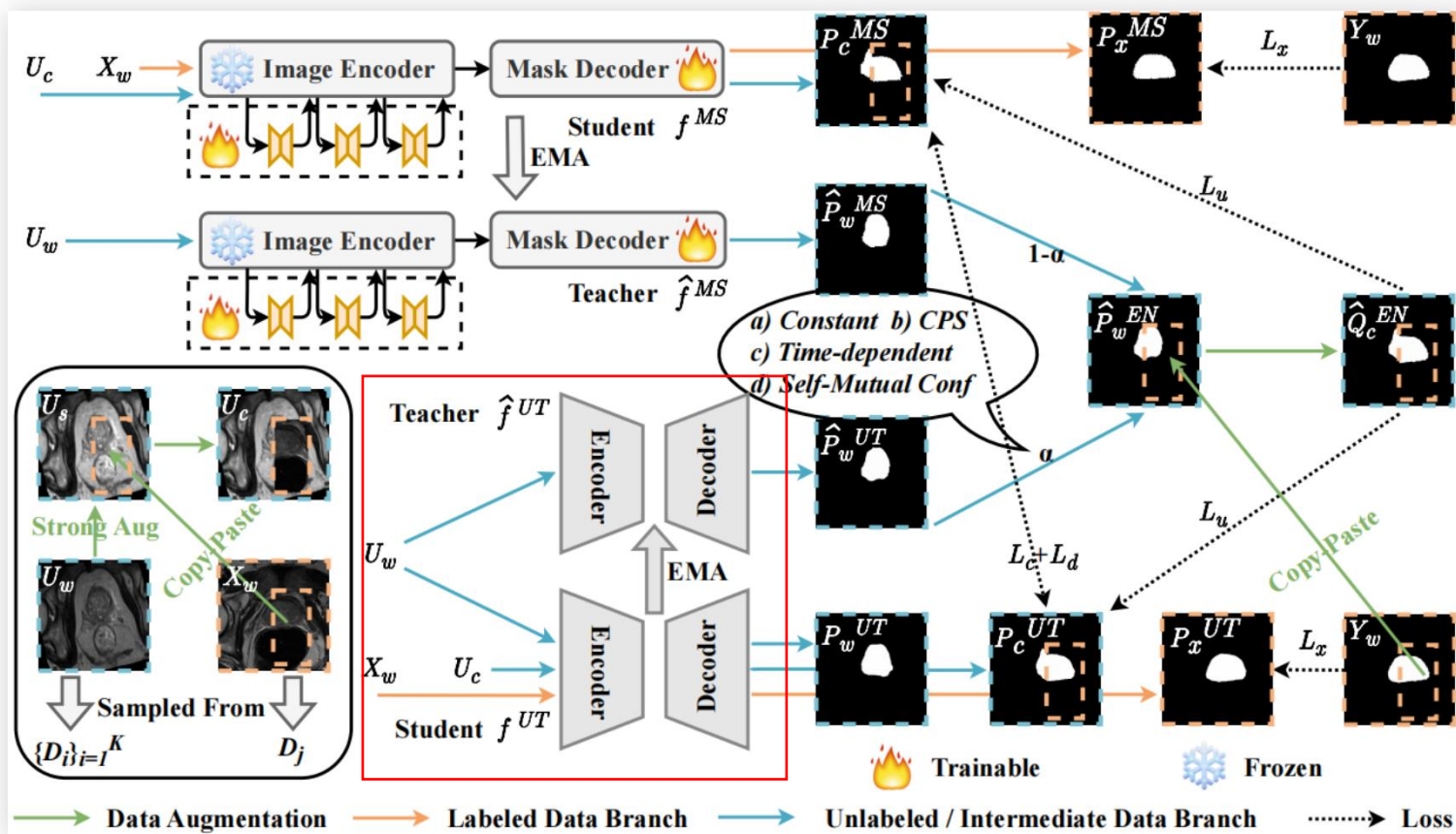
基础模型MedSAM

对冻结的图像编码器采用 LoRA 技术，与掩码解码器一起进行训练，并使用 EMA 更新教师模型和学生模型。



二、研究方法

混合域半监督医学图像分割的协同训练框架SynFoC



传统模型U-Net

采用从头开始训练的方式，使用EMA同步更新教师模型和学生模型。

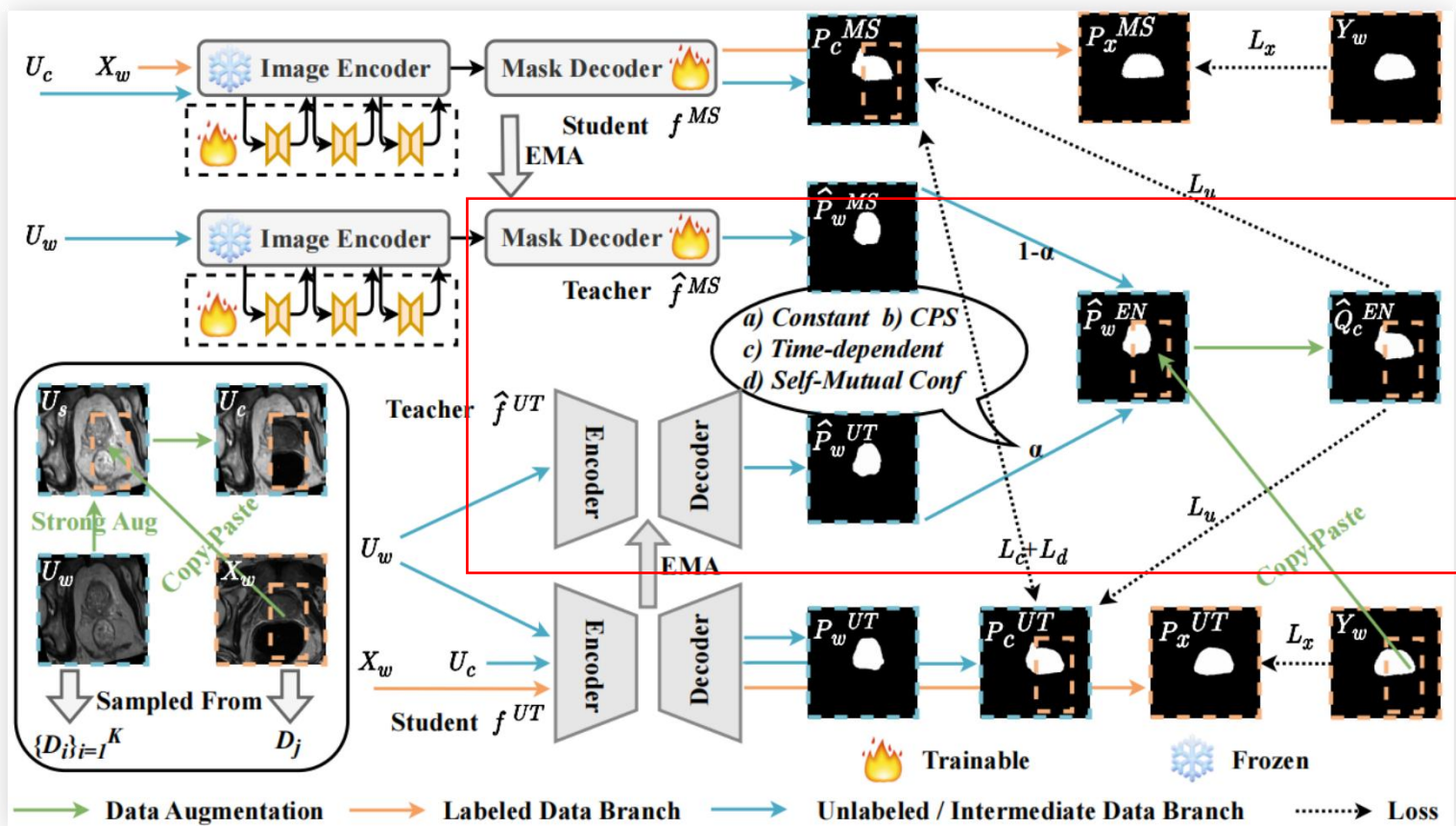


X_w 分别放入MedSAM和UNet的学生模型得到预测 P_x^{MS} 和 P_x^{UT} ，两者独立训练，并与真标签 Y_w 计算监督损失 L_x 。



二、研究方法

混合域半监督医学图像分割的协同训练框架SynFoC



无监督训练-T

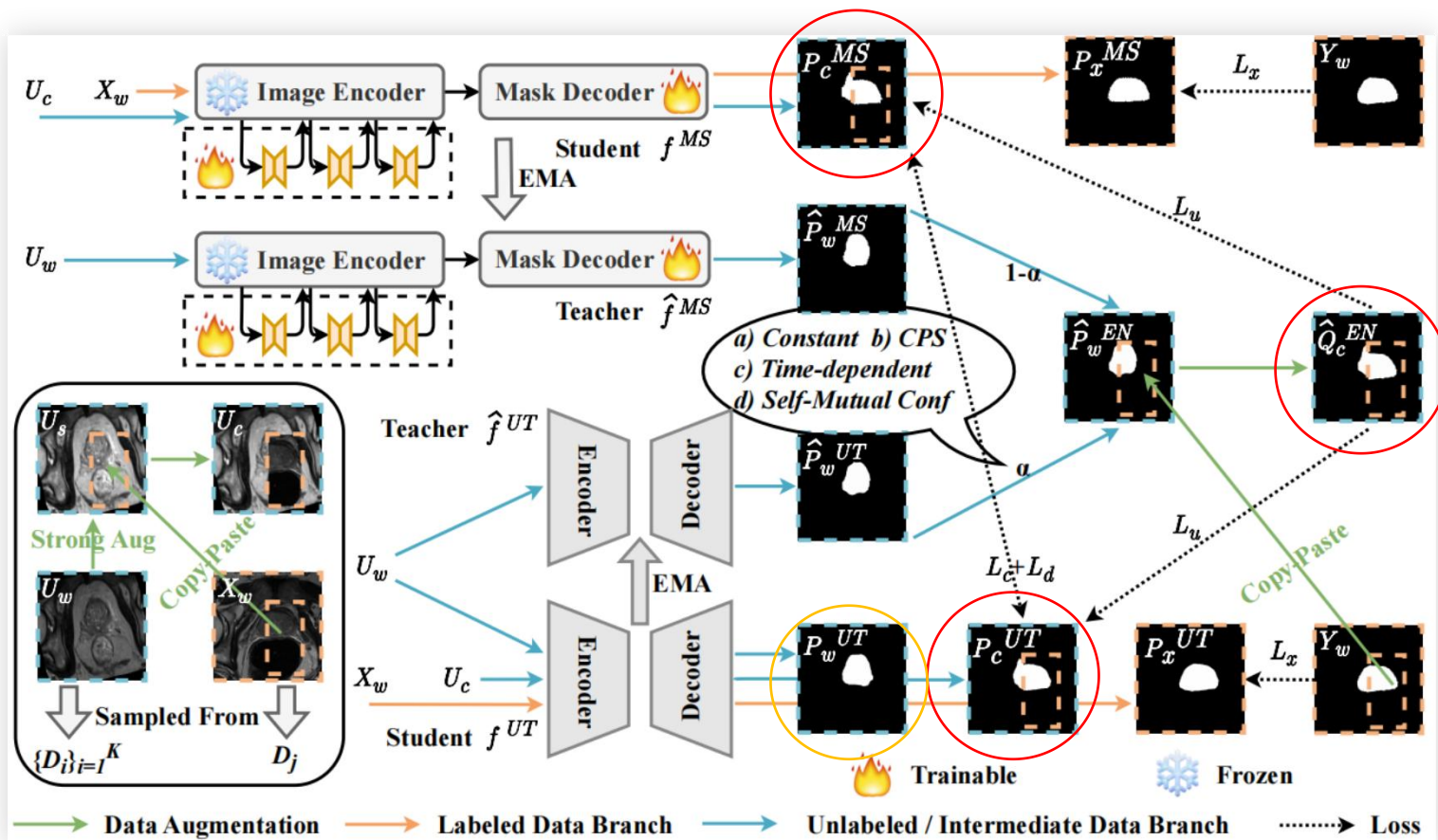
U_w 分别放入MedSAM和UNet教师模型得到 P_w^{MS} 和 P_w^{UT} ,两者通过权重 α 合成伪软标签 P_w^{EN} ,通过二值化得到伪标签 Q_w^{EN} 。进行复制粘贴得到 U_c 伪标签 Q_c^{EN} 。

$$\hat{Q}_c = Y_w \odot M + \hat{Q}_w \odot (1 - M)$$



二、研究方法

混合域半监督医学图像分割的协同训练框架SynFoC



无监督训练-S

U_c 分别放入MedSAM和UNet的学生模型得到 P_c^{MS} 和 P_c^{UT} ,与伪标签 Q_c^{EN} 计算无监督损失 L_u ,并计算一致性损失 L_c 和 L_d 。
 U_w 需经过UNet学生模型预测,计算自置信度。



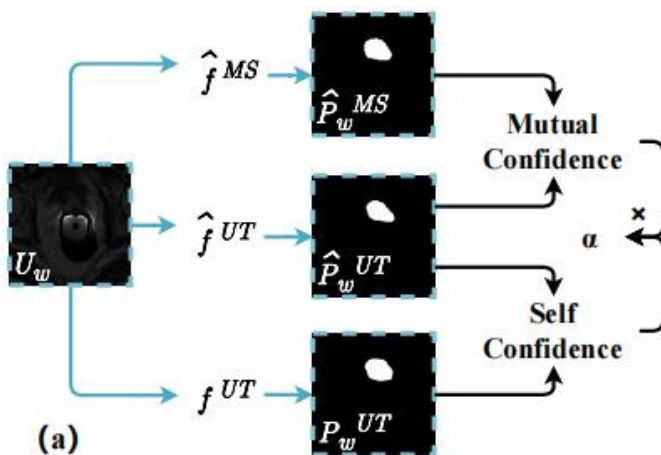
二、研究方法



置信度模块SMC确定集成比例 α

自置信度

自置信度 φ^{self} 衡量UNet教师模型和学生模型对无标签数据 U_w 的预测结果一致性，即计算预测前景类一致的重合度。



置信度趋近于1表示UNet预测可靠

$$\alpha = \varphi^{self} * \varphi^{mut}$$

U_c 的伪标签更倾向于UNet的预测

互置信度

互置信度 φ^{mut} 衡量UNet教师模型和MedSAM教师模型对无标签数据 U_w 的预测结果一致性，同样计算预测前景一致类的重合度。



二、研究方法



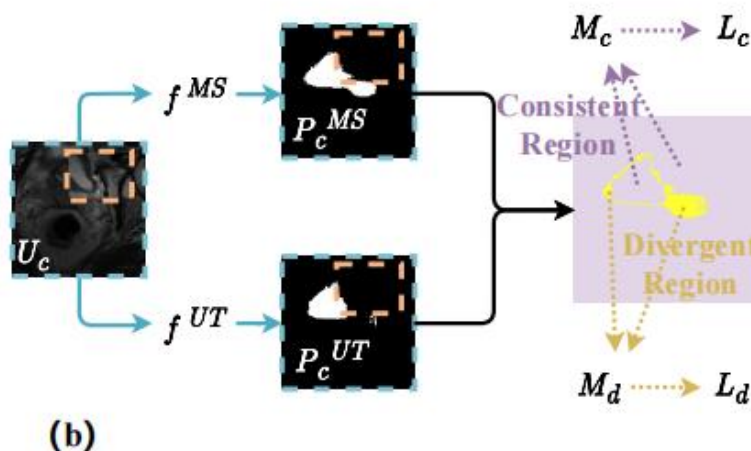
区域特定一致性正则化CDCR

一致性区域 M_c

目标：鼓励模型在预测可靠的区域生成高置信度的预测。

由于低熵表示预测确定性强，最小化 M_c 区域内预测的香农熵，计算损失 L_c 。

$$L_c = -\frac{1}{S} \sum (P_c^{UT} \log P_c^{UT} + P_c^{MS} \log P_c^{MS}) \odot M_c$$



一致区域 M_c 指 P_c^{MS} 和 P_c^{UT} 二值化后的重合区域，分歧区域 M_d 与 M_c 互补。

$$M_c = 1 (Q_s^{UT} = Q_s^{MS}), M_d = 1 - M_c$$

分歧区域 M_d

目标：对于分割困难区域，减少预测差异，以促进表征能力的持续提升。

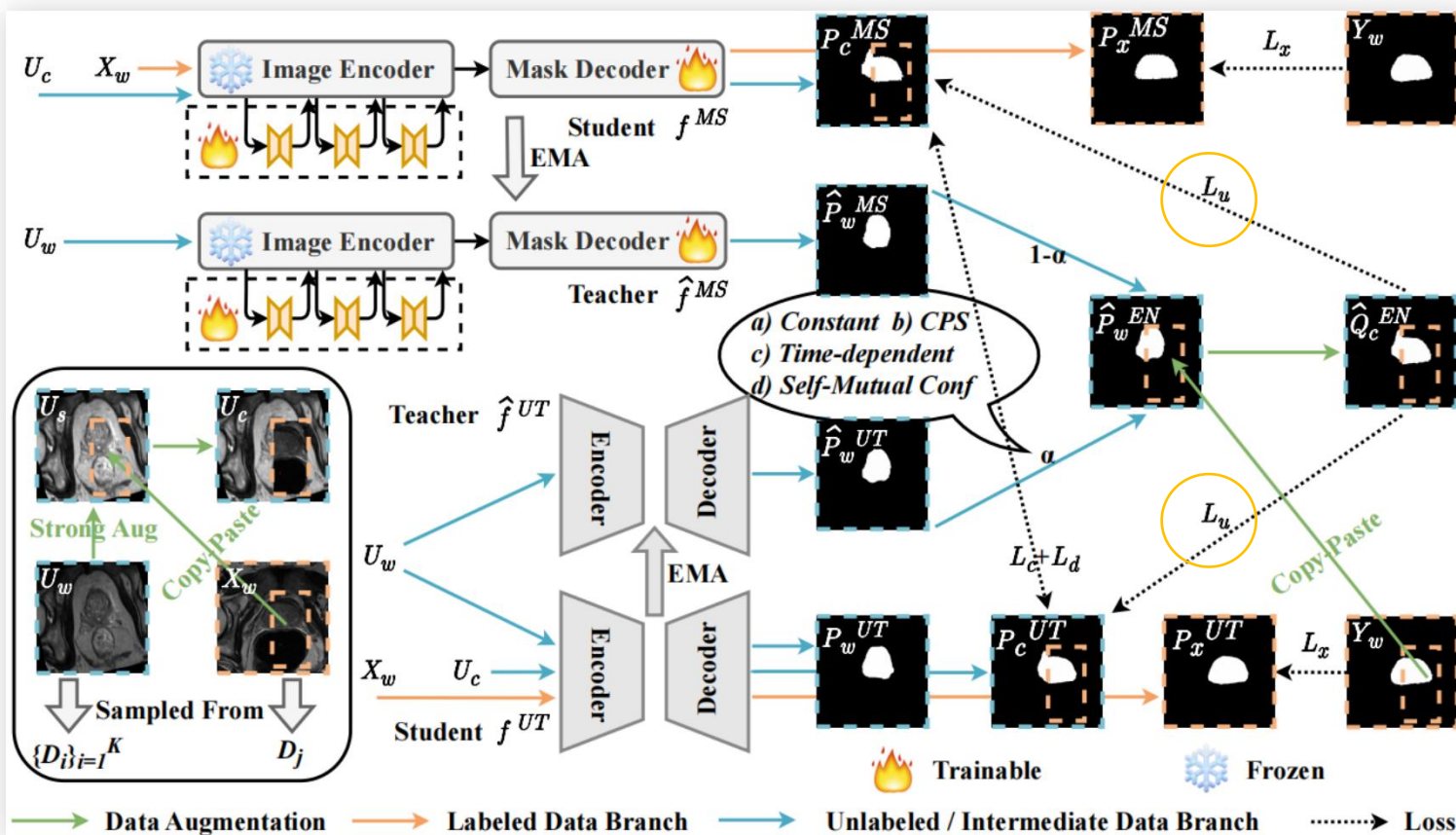
最小化 M_d 区域内的均方误差，计算损失 L_d 。

$$L_d = -\frac{1}{S} \sum \text{MSE} (P_c^{UT}, P_c^{MS}) \odot M_d$$



二、研究方法

混合域半监督医学图像分割的协同训练框架SynFoC



损失函数

总损失函数 L_{total} :

$$L_{total} = L_x + \lambda (L_u + L_c + L_d)$$

其中 λ 与步数 t 相关:

$$\lambda(t) = e^{-5(1-t/t_{max})}$$

无监督损失 L_u 仅计算高置信度

$$\text{区域 } W_c^{EN} = 1 (\max(P_w^{EN}) \geq \tau)$$

τ 设置为 0.95



三、实验配置与结果



实验配置

数据集设置

- ✓ 前列腺数据集：多站点 MRI 数据，6 个域，20 个标注样本，2D 切片分辨率 384×384 。
- ✓ 眼底数据集：4 个医疗中心的眼底图像，视杯 / 视盘分割，20 个标注样本，分辨率 256×256 。
- ✓ M&Ms 数据集：4 个扫描仪供应商的心脏 MRI，5 个标注样本，用于 LV/MYO/RV 分割，分辨率 288×288 。
- ✓ BUSI 数据集：乳腺超声图像，按肿瘤类型分 2 个域，标注样本数为 64 (1/8) 或 128 (1/4)，分辨率 256×256 。

对于各数据集，在训练集中选择其中**一个域的部分作为有标签样本**，取其他域作为无标签样本，测试集包含数据集内全部域。

模型训练参数

- 基础模型：MedSAM (ViT-B 版本)，应用 LoRA 模块降低计算成本，输入分辨率 512×512 。
- 传统模型：U-Net，SGD 优化器 (lr=0.03，动量 0.9)。
- L_x 和 L_u 使用交叉熵损失 L_{ce} 和 L_{dice} 计算。

评价指标

- 骰子相似系数 (DSC)、雅卡尔指数 (Jaccard)、95% 豪斯多夫距离 (95HD)、平均表面距离 (ASD)。



三、实验配置与结果



部分实验结果

前列腺数据集

Methods	Venue	#L	(Prostate Segmentation) DSC ↑						DSC ↑	Jaccard ↑	95HD ↓	ASD ↓
			RUNMC	BMC	HCRUDB	UCL	BIDMC	HK				
SupOnly	-	20	22.11	21.81	19.60	13.87	18.16	26.98	20.42	15.63	118.15	79.70
UA-MT [58]	MICCAI'19	20	19.09	13.66	16.07	37.30	15.23	11.22	18.76	13.44	127.59	85.76
FixMatch [40]	NeurIPS'20	20	81.69	65.27	53.70	70.40	10.20	81.22	60.41	52.53	49.47	28.83
SS-Net [52]	MICCAI'22	20	14.92	11.64	14.49	34.31	15.45	12.52	17.22	12.65	119.73	81.38
BCP [3]	CVPR'23	20	64.79	62.46	50.49	55.08	63.31	57.64	58.96	48.74	56.81	27.77
CauSSL [34]	ICCV23	20	20.36	31.11	15.68	27.27	26.17	26.66	24.54	18.03	116.15	70.57
ABD [14]	CVPR'24	20	53.10	62.28	9.17	59.22	51.92	22.19	42.98	32.35	75.73	47.17
SymGD [33]	CVPR'24	20	<u>88.34</u>	<u>83.26</u>	<u>83.99</u>	<u>85.45</u>	42.03	<u>78.02</u>	<u>76.85</u>	<u>67.88</u>	39.02	21.08
SAMed [60]	arXiv'23	20	63.67	65.62	65.25	68.65	48.83	75.31	64.56	53.49	35.85	15.83
SemiSAM [†] [61]	arXiv'23	20	78.07	67.16	76.49	69.25	65.97	69.76	71.12	59.69	<u>28.19</u>	<u>12.61</u>
H-SAM [13]	CVPR'24	20	56.29	51.23	41.59	60.50	51.94	51.86	52.24	42.37	69.31	36.15
CPC-SAM [35]	MICCAI'24	20	77.23	66.46	66.98	76.79	<u>79.71</u>	75.42	73.77	62.85	31.08	13.16
SynFoC ^{U-Net}	This paper	20	88.09	85.22	84.97	87.74	87.63	87.74	86.90	79.11	11.25	4.82
SynFoC ^{MedSAM}	This paper	20	88.54	85.74	84.89	87.51	87.92	88.34	87.16	79.30	10.26	4.41

- 对比方案先进，DSC提升显著。
- 训练完成的基础模型 MedSAM 效果略高于传统模型 UNet。



三、实验配置与结果



部分实验结果

消融实验

Base	SMC	CDCR	CR	DSC ↑						DSC ↑	Jaccard ↑	95HD ↓	ASD ↓
				RUNMC	BMC	HCRUDB	UCL	BIDMC	HK	Avg.	Avg.	Avg.	Avg.
✓ U-Net				87.74	67.96	82.45	86.31	41.91	84.35	75.12	65.76	54.67	29.08
✓ MedSAM				87.24	79.70	80.20	85.34	71.88	82.46	81.14	71.91	22.78	9.13
✓	✓			88.37	85.09	84.64	87.31	87.45	87.94	86.80	78.85	10.78	4.64
✓			✓	87.99	81.84	82.99	86.18	71.92	86.34	82.88	73.81	17.09	7.21
✓		✓		88.20	84.86	83.36	86.75	87.15	86.82	86.19	77.94	11.91	5.03
✓	✓	✓		88.54	85.74	84.89	87.51	87.92	88.34	87.16	79.30	10.26	4.41

- SMC 表示 U-Net 和 MedSAM 的协同训练，其中伪标签集成权重 α 基于自互置信度确定。
- CDCR 指的是在两个模型之间引入一致性——分歧一致性正则化。
- CR 表示直接对 P_c^{MS} 和 P_c^{UT} 使用均方误差计算损失。



四、总结



总结与思考

01

基础模型和传统模型的动态互补

实验过程发现教师模型生成的伪标签质量很差，导致很多情况下学生模型的测试效果反而弱于教师模型。

可以利用 MedSAM 的泛化能力指导 U-Net 早期训练，同时通过 U-Net 纠正 MedSAM 的高置信度错误，打破“基础模型主导”的传统模式。

02

置信度区域

无监督损失 L_u 提及选择高置信度区域参与训练，这是由于高置信度代表模型对此区域的预测是确信的；若 $\max(P_w^{EN}) < \tau$ ，则说明模型对该区域的分类不确定性高，这类区域直接参与训练会导致模型学习到不准确的信息。



四、总结



总结与思考

03

一致性正则化

根据模型预测的重合度，分区域进行约束：

对模型预测一致区域 M_c 最小化香农熵 L_c ；对分歧区域 M_d 最小化MSE损失 L_d 。

这种方式强化了模型对可靠区域的预测信心，推动了困难区域的表征融合，有助于提升分割精度和模型训练的加速收敛。

04

数据增强

文中不仅使用了弱增强和强增强的方式，还使用了复制粘贴的操作弥补两种样本的域差距，增加数据的多样性，但文中并没有提及复制粘贴使用的掩码 M 是如何确定的。

文章的强增强方式只生成了 U_s ，在后续的计算中只与 U_c 相关，然而 U_c 的伪标签却由弱增强的样本 U_s 预测并和 Y_w 复制粘贴得到，这里的原因不是很确定。