

使用 SAMIHS 进行视频息肉分割

彭雲鵬

摘要

”Segment Anything Model (SAM), 一个在大规模注释数据上训练的视觉领域基础模型, 最近在医学图像分割领域引起了广泛关注。尽管 SAM 在自然场景上具有出色的性能, 但在处理医学图像时, 特别是那些具有模糊边界和低对比度高度不规则区域的图像时, 它的性能会下降。本文复现的论文提出了一种基于 SAM 的参数高效微调方法, 称为 SAMIHS, 用于颅内出血分割。与以前的 SAM 和基于 SAM 的方法不同, SAMIHS 将参数重构适配器引入 SAM 的图像编码器, 并考虑了适配器参数的高效灵活利用。此外, 我们采用了组合损失, 将二元交叉熵损失和边界敏感损失结合起来, 以增强 SAMIHS 识别边界区域的能力。本文基于其开源代码进行了复现, 同时将原有模型应用到视频息肉分隔, 各项指标取得了较为明显的提升。

关键词: SAM; 参数微调; 息肉分割

1 引言

结直肠癌 (CRC) 是全球癌症死亡的主要原因 [1]。大多数结直肠癌由腺瘤性息肉演变而来, 因此早期发现和切除息肉对于预防和治疗结直肠癌至关重要。结肠镜检查是在息肉发展为结直肠癌之前发现和切除息肉的黄金标准 [2]。然而, 由于息肉的形状、大小和质地的多样性, 在结肠镜检查过程中准确识别和分割息肉是一项复杂的任务。这可能导致漏诊或误诊息肉, 这可能严重损害患者的健康。

息肉分割在医学影像处理领域, 尤其是在结肠癌的早期诊断和预防中, 扮演了至关重要的角色。这一过程涉及使用高级影像技术, 如内窥镜图像或计算机断层扫描 (CT), 从中准确地识别和分割出息肉。准确的息肉分割使医生能够早期检测到这些结构, 从而及时干预, 如通过手术切除。这一点对于降低结肠癌的发病率和死亡率至关重要, 因为一旦发展为癌症, 治疗难度和复杂性会大大增加。通过及时移除息肉, 可以在病变恶化之前进行有效的干预。除了早期诊断和预防, 息肉分割在提高诊断过程的效率和精度方面也发挥着关键作用。在内镜检查过程中, 自动或半自动的息肉分割技术可以帮助医生快速识别息肉, 减少由于疲劳或其他因素引起的遗漏和误诊。这种技术的应用提高了检查的准确性, 确保了患者接受适当的治疗。

息肉分割不仅对于结肠癌的早期诊断和预防至关重要, 也对提高医疗诊断的效率、精度以及推动相关医学研究的进展具有重大意义。随着医学影像技术和人工智能的进一步发展, 我们可以期待息肉分割技术在未来对结肠健康管理 and 癌症预防方面发挥更大的作用。

2 相关工作

2.1 任务概述

我们主要关注视频息肉分割任务，它可以定义为二元类视频对象分割任务，即识别息肉和非息肉区域。具体来说，我们的目标是训练一个模型，为每帧的每个像素分配概率预测（即范围从 0 到 1 的非二进制掩码）。

2.2 SAM 模型

虽然 SAM 在广泛的自然场景的分割中表现得非常出色，但它被证明在直接应用于医学领域时不如自然图像表现优秀 [3]。针对这个问题，MedSAM [4] 冻结了 SAM 的图像编码器和提示编码器，并在大量的医学图像上微调了轻量级的掩模解码器。在以前的研究中，参数微调方法被认为是使基础模型适应特定下游任务的有效方法。特别是，将不同形式的适配器整合到基础模型的多个位置，从而以更少的额外参数成本显著提高性能 [5–8]。一些工作已经将这种方法引入到 SAM 的微调中。例如，SAMed [9] 在固定图像编码器中应用了基于低秩次的策略 (LoRA)，在 Synapse 多器官数据集上一起微调了 LoRA 层、提示编码器和掩模解码器。MSA [10] 在图像编码器的每个 Transformer 块中采用了适配器。适配器分别以串联和并行的方式集成到注意力层和 MLP 层。SAMUS [11] 设计了一个与 SAM 的图像编码器平行的可学习的 CNN 分支，同时增加了用于 Transformer 块的特征适配器和用于位置嵌入的位置适配器。我们所复现的这篇论文所提出的 SAMIHS [12] 优化了适配器设计和参数微调损失函数，同时还对颅内出血分割进行了专门的评估。

2.3 视频息肉分割

视频息肉分割在结肠癌早期诊断和预防方面的重要性主要体现在它极大地提高了医学诊断的准确性和效率。这项技术利用高级影像分析和人工智能算法，在内窥镜视频中自动识别和分割息肉，即结肠癌的潜在前体。这种自动化的过程减少了医生在长时间和复杂的内窥镜检查中由于疲劳或其他因素引起的遗漏，从而提高了检测的全面性和准确性。更重要的是，它使得医生能够及时发现并移除这些息肉，从而预防癌症的发展。

视频息肉分割技术还对个性化医疗计划的制定至关重要。通过详细分析息肉的特征，如大小、形状和生长速度，医生可以更好地评估患者的病情并制定更精确的治疗方案。此外，这项技术在医学教育和远程医疗服务中也显示出巨大潜力。它不仅可以用作教学工具，帮助医学生和年轻医生学习如何识别和处理息肉，还可以通过远程诊断支持，使得位于偏远地区的患者也能够获得专业的诊断和治疗建议。

随着人工智能和机器学习技术的不断进步，视频息肉分割的准确性和效率有望进一步提升。这不仅对结肠癌的早期诊断和预防具有重要意义，也为医学影像分析领域带来了革新，标志着现代医疗技术向更高水平的发展。

3 本文方法

在本节中，我们首先介绍了我们复现的论文的概述，然后详细介绍了我们在此基础之上的改进。

3.1 论文方法概述

本文提出了一种基于 SAM 模型参数有效微调方法，称为 SAMIHS，用于颅内出血分割。SAMIHS 的整体结构如图 1 所示。SAM 模型虽然在大多数任务上都表现良好，但对于医学图像而言实际表现效果并不好。为此本文提出了一个 Adapter 称为 Parameter-Refactoring Adapter 用于微调 SAM 模型的参数，使其适应医学图像任务。通过对 Transformer Block 中插入两个所有层共用参数的 Adapter 使得 SAM 模型可以适应医学图像处理的问题。同时使用了交叉熵损失函数和边界损失函数，使得对边界模糊，对比度低的情况下依然能得到不错的分割效果。

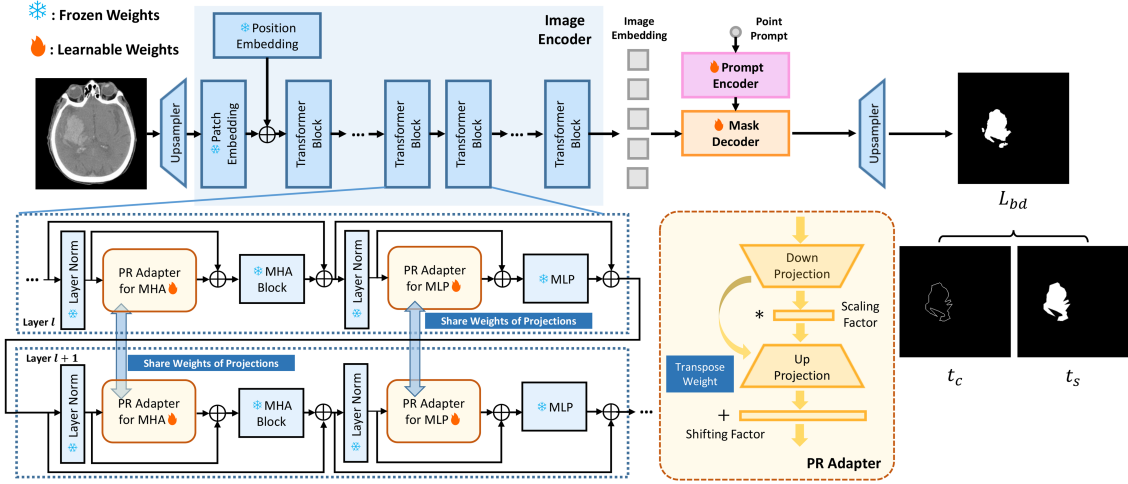


图 1. 方法示意图

3.2 PR Adapter

先前的研究已经以各种形式将单独的可学习参数引入 SAM 的不同层。受这些工作的启发，这篇论文的作者探索了一种新的方法，为不同变压器块中相同位置的适配器配置统一的线性映射设置，以提高适配器的效率。具体地，如图 1 所示，1，给定输入特征 $m_{in} \in \mathcal{R}^{h \times w \times c}$ ，它依次通过对称下投影 $W_{down} \in \mathcal{R}^{c \times c'}$ 和上投影 $W_{up} \in \mathcal{R}^{c' \times c}$ 。P 适配器中的正向过程可以公式化为：

$$m_{out} = m_{int} W_{down} R_l W_{up} + B_l + m_{in} \quad (1)$$

其中 $R_l \in \mathcal{R}^{c'}$ 和 $B_l \in \mathcal{R}^c$ 分别表示缩放因子和移位因子。缩放因子 R_l 和移位因子 B_l 对于每一层中的每个适配器都是独立的，旨在动态重构多层特征以提高自适应性能。由于 SAM 采用了在多头注意力块之前对特征采用窗口注意力进行分块，因此我们在 MHA 和 MLP 之前配置了两个 PR 适配器以便于在每个变压器块内灵活调整。因此，修改后的变压器组可以形式化地描述为

$$m^{(l)'} = MHA(PR_{MHA}(LN(m^{(l-1)}))) + m^{(l-1)} \quad (2)$$

$$m^{(l)} = MLP(PR_{MLP}(LN(m^{(l)'}))) + m^{(l)'} \quad (3)$$

请注意， PR_{MHA} 和 PR_{MLP} 是独立的，而它们在不同 Transformer 块之间的权重是共享的，以提高 PR Adapter 的效率。

3.3 损失函数

与自然场景下 SAM 的学习对象不同，颅内出血与周围正常组织的对比度较低，导致出血边界模糊，容易发生分割错误。受相关研究的启发，作者在 Samish 训练中引入了边界敏感损失函数，并将其与交叉熵损失相结合，以提高出血预测的性能。给定一对包含出血目标的预测 P 和真值 G ，我们的目标是减小差值面积 $P \cup G - P \cap G$ 和增加相交 $P \cap G$ ，特别是增加边界区域的相交。为此，边界敏感损失可以写成

$$L_{bd} = 1 - \frac{2\gamma TP}{2\gamma TP + FP + FN} \quad (4)$$

其中 γ 是调节边界感知灵敏度的动态因子，

$$\gamma = 1 - \frac{t_c}{t_s} \quad (5)$$

其中 t_c 和 t_s 分别表示目标的边界长度和目标的大小。SAMIHS 的总损失如下：

$$L = \lambda_1 L_{bd} + \lambda_2 L_{ce} \quad (6)$$

其中权衡参数 $\lambda_1 = \lambda_2 = 0.5$ 。

4 复现细节

4.1 与已有开源代码对比

由于本次浮现的论文已开源，我们结合论文的基本框架，参考开源代码进行复现。复现的代码相较于原本的代码相比整体增加了两个模块，使用 PNS+ 作为提示输入辅助分割，使分割结果更加精细，误差更小。整体的改进如下：

(1) 改进数据集加载代码，论文原有代码所处理的数据集是颅内出血的超声数据集，所以在其原本的代码里对于数据集的加载主要是对超声数据的预处理然后加载，由于本文复现这篇论文所处理的任务不是颅内出血，而是息肉视频分割，所以对其数据集加载模块进行较大的改动。

(2) 使用 PNS+ 网络分割结果作为 SAM 的提示输入。SAMIHS 原文的主要工作在于对 SAM 的 Image Encoder 进行改进使其适应其他任务。但是论文忽视了 Prompt Encoder 的重要性，在其开源的源代码里，Prompt Encoder 所接受的提示输入只是随机的点提示或框提示。对于特定任务来说，由足够信息量的提示可以使分割结果进一步提升。所以针对这一点，我们使用了此次实验任务此前的一个 baseline 作为提示，使其能够对该任务的分割效果进一步提升。使用 PNS+ 作为提示不仅可以提升分割的效果，同时由于 PNS+ 是视频分割，其也能作为图像分割的 SAM 模型注入时空特征使其能够具有时空一致性。

4.2 实验环境搭建

本次复现工作使用的操作系统是 Ubuntu20.04 系统，在 Anaconda3 虚拟环境下使用 Pytorch1.1 框架复现。使用的显卡是 Nvidia 2080GPU，训练了 15 个 epoch 达到最佳训练效果。实验选用的数据集是 SUN-SEG 数据集，训练时使用其 TrainDataset 里的随机约 40% 数据进行训练。该数据集于 2022 提出，包含 1013 个视频攻击 158690 张图片，并含有各类的标注数据。是目前最大，最具代表性的息肉视频分割数据集。

4.3 创新点

相较于原本的 SAMIHS 的结构，通过增加一个领域相关的 baseline 网络作为提示，使 SAMIHS 能够在对应任务上表现出非常不错的性能。具体的改进结构如图 2 所示。创新部分由于涉及论文 idea 实现，暂不开源，请见谅。

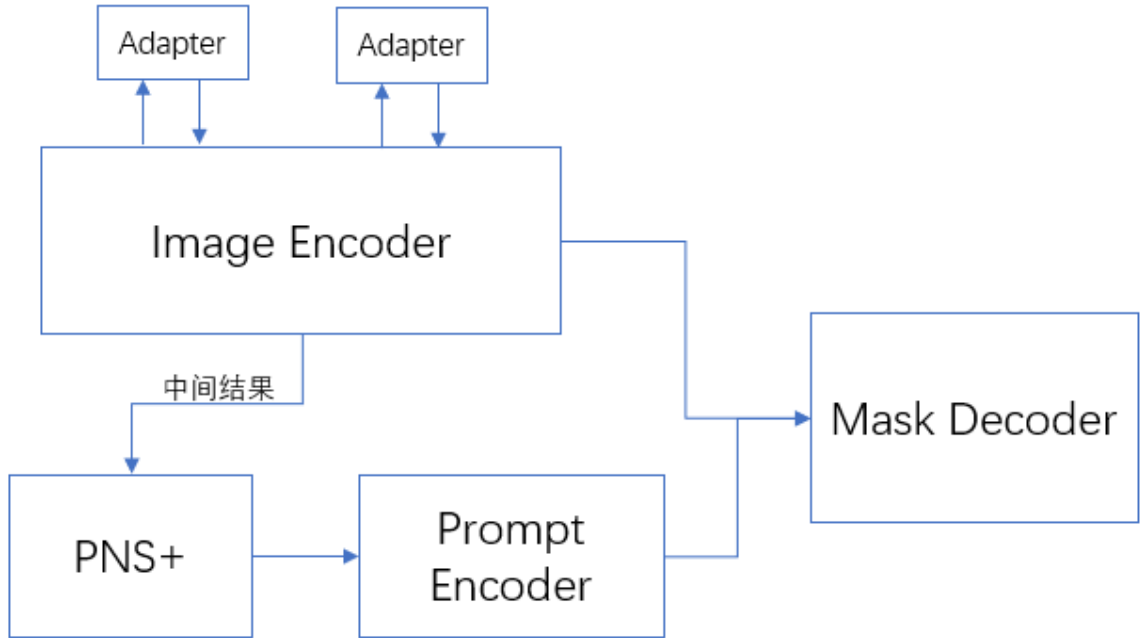


图 2. 整体结构

5 实验结果分析

由于本次复现所改进的部分主要使针对与与原文不同的视频息肉分割任务，所以在此比较的是原本该任务的 sota 框架 AutoSAM [13]。我们分别在目前最大的 SUN-SEG 数据集上测试了 Dice, S_α , Sensitivity 和 F_β^w 这些常用测试指标来评估模型的改进效果。具体数据如表 1。部分分割结果可视化如图 3

表 1. 息肉分割指标比较

method	SUN-SEG-Hard(Unseen)			
	S_{α}	$Dice$	Sen	F_{β}^{ω}
AutoSAM	0.815	0.753	0.672	0.716
ours	0.836	0.799	0.702	0.770

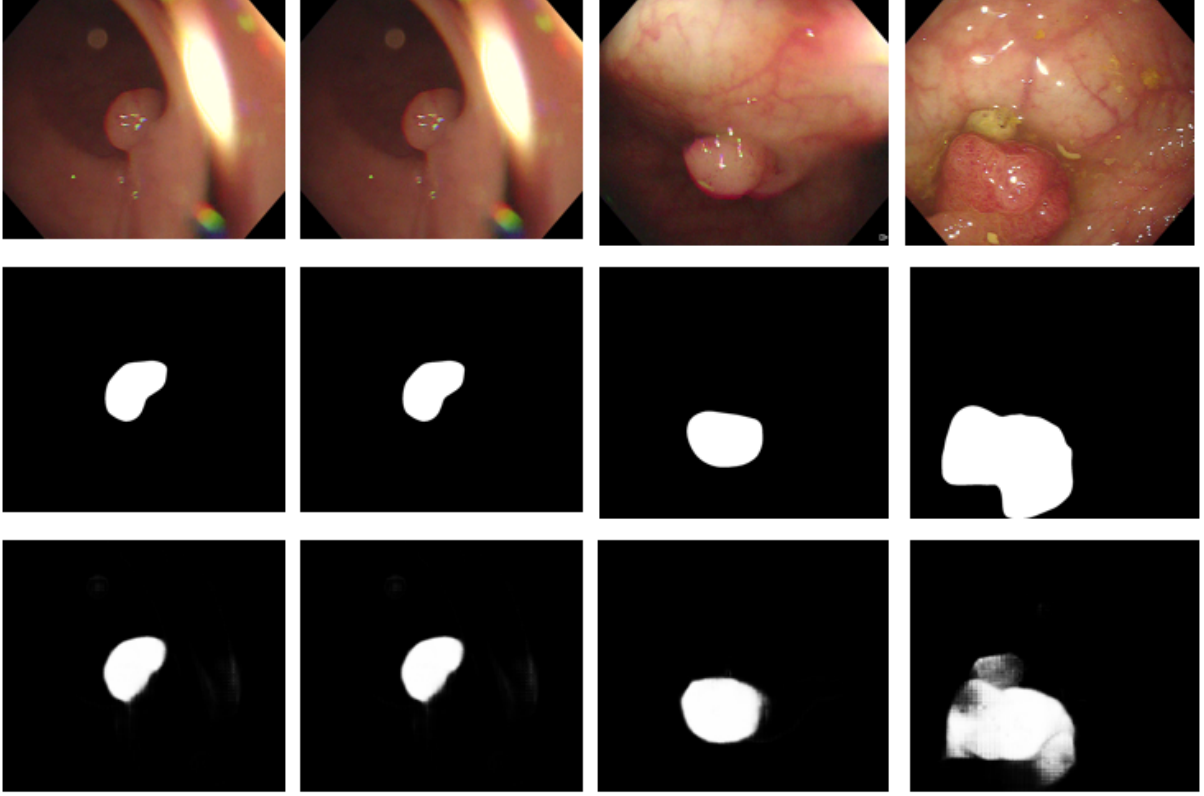


图 3. 实验结果图

6 总结与展望

本文对论文 SAMIHS: ADAPTATION OF SEGMENT ANYTHING MODEL FOR INTRACRANIAL HEMORRHAGE SEGMENTATION 进行了复现，并在开源代码的基础上进行了对特定任务的改进。改进之后的结果相较于该领域的 sota 框架有很显著的提升。但有所不足的在于由于本人的研究方向并不是颅内出血的图像处理，所以并没有针对原论文的任务进行改进和测试。同时虽然针对息肉分割的研究方向而言效果已经由很显著的提升，但是在实际的分割结果中仍有不足，还需要继续改进。未来可能再引入一些后处理的模块加强其分割结果。

参考文献

- [1] Rebecca L. Siegel, Kimberly D. Miller, Hannah E. Fuchs, and Ahmedin Jemal. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, 72(1):7–33, 2022.

- [2] Aasma Shaukat, Charles J. Kahi, Carol A. Burke, Linda Rabeneck, Bryan G. Sauer, and Douglas K. Rex. Acg clinical guidelines: Colorectal cancer screening 2021. *Official journal of the American College of Gastroenterology / ACG*, 116(3), 2021.
- [3] Maciej A. Mazurowski, Haoyu Dong, Hanxue Gu, Jichen Yang, Nicholas Konz, and Yixin Zhang. Segment anything model for medical image analysis: An experimental study. *Medical Image Analysis*, 89:102918, 2023.
- [4] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images, 2023.
- [5] Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Larousilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp, 2019.
- [6] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021.
- [7] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning, 2022.
- [8] Dongze Lian, Daquan Zhou, Jiashi Feng, and Xinchao Wang. Scaling shifting your features: A new baseline for efficient model tuning, 2023.
- [9] Kaidong Zhang and Dong Liu. Customized segment anything model for medical image segmentation, 2023.
- [10] Junde Wu, Wei Ji, Yuanpei Liu, Huazhu Fu, Min Xu, Yanwu Xu, and Yueming Jin. Medical sam adapter: Adapting segment anything model for medical image segmentation, 2023.
- [11] Xian Lin, Yangyang Xiang, Li Zhang, Xin Yang, Zengqiang Yan, and Li Yu. Samus: Adapting segment anything model for clinically-friendly and generalizable ultrasound image segmentation, 2023.
- [12] Yinuo Wang, Kai Chen, Weimin Yuan, Cai Meng, and XiangZhi Bai. Samihs: Adaptation of segment anything model for intracranial hemorrhage segmentation. *arXiv preprint arXiv:2311.08190*, 2023.
- [13] Tal Shaharabany, Aviad Dahan, Raja Giryes, and Lior Wolf. Autosam: Adapting sam to medical images by overloading the prompt encoder, 2023.