

SAM2-FNet: 一种基于频域多专家融合的医学图像分割网络

张子华¹, 李绍丽¹, 葛兵², 陆鹏²

¹沈阳工业大学信息学院, 辽宁 沈阳

²沈阳翰熙机械设备有限公司, 辽宁 沈阳

发布日期: 2025年01月30日

摘要

近年来,深度学习在医学图像分割方面取得了重大进展,但现有方法受限于局部细节与全局语义的耦合难题。为此,本文提出一种基于SAM2-UNet改进的频域专家融合网络SAM2-FNet,通过频域解耦与多专家协同机制实现细粒度多尺度特征增强,显著提升分割精度。具体而言, SAM2-FNet主要具备以下三个功能: (1)频域解耦增强,利用二维傅里叶变换将输入特征分解为高频与低频成分,并结合双分支异构处理结合通道注意力,实现自适应融合; (2)动态专家决策,设计改进型专家融合模块(FEM), FEM模块中构建了包含五个子网络(SubNet)的多专家结构,增强模型对病灶异质性的泛化能力; (3)解码优化架构,引入局部细化模块(LRM),通过空洞卷积金字塔与空间注意力机制弥合编解码器语义鸿沟,提升边缘重建精度。在Kavsir-SEG数据集上的实验表明, SAM2-FNet在Kavsir-SEG数据集上相较原SAM2-UNet模型Dice系数提升3.0%, Iou系数提升4.1%;与umambabot、swinunetr等基准模型相比, SAM2-FNet且在高噪声与低对比度场景下展现出更强鲁棒性。该研究为基于预训练编码器的下游任务改进提供了一种新颖而有效的解决方案,具有较高的理论价值与实际应用前景。

关键词

医学图像分割; SAM2; 频域专家融合模块

SAM2-FNet: A Frequency-Domain Multi-Expert Fusion Network for Medical Image Segmentation

Zihua Zhang¹, Shaoli Li¹, Bing Ge² and Peng Lu²

¹Shenyang University of Technology, Shenyang 110870, China

²Shenyang Hanxi Mechanical Equipment LLC, Shenyang Liaoning, China

Abstract

文章引用: 张子华, 李绍丽, 葛兵, 陆鹏. (2026). SAM2-FNet: 一种基于频域多专家融合的医学图像分割网络. 计算机科学辑要, 1(3), 14-22. ISSN Print: 3079-5362; ISSN Online: 3079-5370
DOI: 10.63313/CS.8023

In recent years, deep learning has achieved substantial progress in medical image segmentation; however, existing methods remain constrained by the challenge of effectively coupling local detail with global semantic context. To address this issue, we propose SAM2-FNet, a frequency-domain expert-fusion network built upon an improved SAM2-UNet. SAM2-FNet attains fine-grained, multi-scale feature enhancement through frequency-domain decoupling and a multi-expert collaborative mechanism, leading to marked improvements in segmentation accuracy. Specifically, SAM2-FNet provides three primary capabilities: (1) Frequency-domain decoupling and enhancement—input features are decomposed into high-frequency and low-frequency components via a two-dimensional Fourier transform; the resulting components are processed by a dual-branch heterogeneous pipeline and adaptively fused using channel-wise attention to preserve both texture details and global structure. (2) Dynamic expert decision-making—an improved Expert Fusion Module (FEM) implements a multi-expert architecture composed of five sub-networks (SubNets), enabling dynamic, data-driven expert weighting and enhancing the model's generalization to lesion heterogeneity. (3) Decoder optimization—a Local Refinement Module (LRM) is introduced in the decoder: an atrous (dilated) convolutional pyramid combined with a spatial-attention mechanism narrows the encoder-decoder semantic gap and improves boundary reconstruction fidelity. Experiments on the Kvasir-SEG dataset demonstrate that SAM2-FNet yields a 3.0% absolute increase in Dice coefficient and a 4.1% absolute increase in IoU relative to the baseline SAM2-UNet. Moreover, compared with benchmark models such as umambabot and SwinUNETR, SAM2-FNet exhibits superior robustness under high-noise and low-contrast imaging conditions

Keywords

Medical image lesion segmentation; SAM2; frequency expert ensemble module

Copyright © 2026 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



1. 引言

医学图像分割作为辅助临床诊断、治疗方案设计及预后评估的重要技术手段[1], 能够帮助医生准确地观察患者的病变区域和器官结构, 从而提高诊断效率与准确性[2]。传统基于 CNN 的分割方法虽然在局部细节捕捉上表现突出, 但其局部感受野的固有限制导致难以充分建模病灶边缘的全局上下文信息[3]。为了克服这一局限性, 近年来部分研究尝试引入 Transformer 模块以增强全局特征建模, 但往往以较大的模型复杂度和计算成本为代价, 同时在多尺度信息融合方面仍存在不足[4]。

另一方面, 基础模型如 Segment Anything Model (SAM) [5]的出现为分割任务带来了新的机遇。尽管 SAM 在自然图像分割中取得了令人瞩目的成果, 但其直接应用于医学图像分割仍面临泛化能力不足、提示依赖性强等问题[6]。

因此, 已经进行了许多尝试来使 SAM 适应医学图像, 产生了有希望的结果, 并证明了 SAM 在医学领域的巨大潜力[7]。例如, 作为将 SAM 的成功扩展到医学图像的初步尝试, Ma 等人[7]提出了通过单独微调 SAM 的掩模解码器来实现 MedSAM, 并证明了与原始 SAM 相比有一定程度的改进。此外, 受自然语言处理 (NLP) 中常用的“适配器”技术的启发, Wu 等人[8]和 Cheng 等人[12]通过使用“适配器”模块引入医学特定知识, 促进了 SAM 的微调, 并证明了显著的增强。然而, 这些初步的努力仍然表现出一些局限性。首先, 他们主要关注器官分割, 而较少关注病灶分割, 这是一个更具挑战性和关键性的任务。其次, 为了促进培训过程, 他们总是依赖大量的注释医学数据 ([7]中的 20 多万个掩码和[9]中的 1970 万个掩码)。最后, 这些方法建立在原始 SAM 的特征表示上, 并且仅使用其瓶颈信息来进行微调, 为了充分发挥 SAM 预训练编码器的强大特征提取能力, 并针对医学图像中病灶异质性、局部细节与全局语义信息

难以兼顾的问题, Xiong 等人[10]提出了 SAM2-UNet, SAM2-UNet 采用 SAM2 的 Hiera 主干作为编码器, 而解码器则采用经典的 U 型设计。此外, 在编码器中还插入了适配器, 以实现参数高效微调。本文提出了一种基于 SAM2-UNet 改进的频域专家融合网络。

2. 方法

2.1. 整体架构

本文提出的 SAM2-FNet 结构 (如图 1 所示) 是一种基于 Transformer 编码器和多尺度特征增强机制的分割网络, 其整体设计充分融合了全局上下文信息和局部细节特征, 以实现高精度的分割效果。该网络首先输入尺寸为 $3 \times 256 \times 256$ 的原始图像, 经过 Transformer 编码器(如图 3 所示)处理后, 在每个 Transformer Block 中嵌入了轻量化的 Adapter 模块, 从而在保持原有多头自注意力机制的同时, 对分割任务相关的细微特征进行高效调优, 进一步提取并强化全局语义信息。编码器在多层次特征提取过程中, 通过下采样操作逐步降低特征图的空间分辨率, 同时增大通道数, 使得低分辨率、高语义的特征图能够涵盖更丰富的信息, 其输出包括四个不同尺度的特征图, 从 144 通道到 1152 通道不等。接下来, 解码器部分采用 FEEM 模块对各尺度特征进行通道压缩与加权融合, 以突出关键特征并抑制冗余信息, 然后通过级联上采样操作逐步恢复特征图的空间分辨率, 在上采样过程中, 深层特征与浅层特征通过拼接与卷积融合, 实现了语义信息与细节信息的有效结合。此外, 在部分上采样阶段, 网络还引入了局部感受野细化模块 LRM, 该模块通过多分支卷积结构 (包含不同尺寸卷积核及膨胀卷积) 捕捉局部多尺度上下文信息, 并对局部特征进行精细化建模和注意力加权, 从而有效提升边缘和细节区域的辨识能力。最终, 网络输出与输入尺寸一致的单通道分割概率图, 通过 1×1 卷积实现目标与背景的精确区分。综合来看, SAM2-FNet 利用 Transformer 的全局建模优势、Adapter 的轻量化增强、以及 FEEM 和 LRM 模块的多尺度特征融合机制, 实现了在复杂场景下高效且精细的目标分割, 为医学图像分割等应用提供了有力的技术支持。

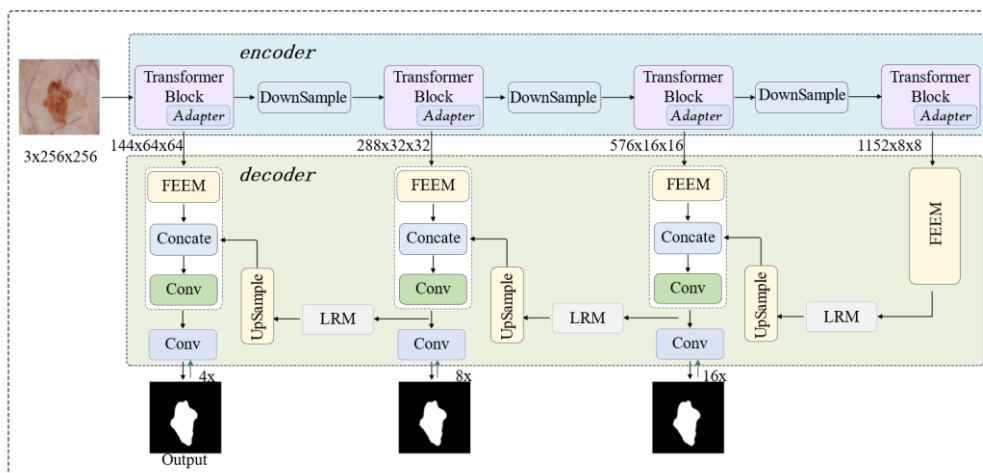


Figure 1. Overall architecture of SAM2-FNet

图 1. SAM2-FNet 的整体架构

2.2. 频域专家融合模块

为有效提升医学图像病灶分割任务中多尺度细节与全局语义信息的协同建模能力, 本文提出了一种频域增强融合模块 (Frequency-enhanced Ensemble Module, FEEM), 与将频域处理作为附加后处理不同, FEEM 将频域分解直接建模为特征提取的核心操作, 从根本上重新界定了特征精炼机制, 其核心结构如图 2 所示。该设计基于一个理论前提: 病理结构在频谱域中呈现出独特的特征分布, 其中病变边界主要对应高频成分, 而解剖学上下文则主要分布在低频波段。因此, 通过可学习的频谱分解与动态专家融合机制, 可以实现对互补频域信息的最优整合。具体而言, FEEM 首先对输入特征图 $X \in \mathbb{R}^{C \times H \times W}$ 经过零相位平移预处理, 以实现频谱中心化并获得正确的频率分量分布。随后, 利用二维快速傅里叶变换 (FFT) 映射至频域, 得到复数域表示:

$$F(X) = \mathcal{F}\{\text{Shift}(X)\} \quad (1)$$

基于频域能量分布特性将频谱分解为高频分量 F_{high} 与低频分量 F_{low} , 表示为:

$$F_{high} = F(X) \odot M, F_{low} = F(X) \odot (1 - M) \quad (2)$$

其中 F_{high} 主要捕捉图像中的边缘、纹理等高频细节信息, 而 F_{low} 则用于编码图像的全局语义信息, 二者构成互补的表征。

接下来, 高频与低频分量分别通过逆傅里叶变换与逆平移 (Unshift) 重构得到频域解耦后的空间域特征, 得到空间域特征表示:

$$X_{high} = \text{Unshift}(\mathcal{F}^{-1}(F_{high})), X_{low} = \text{Unshift}(\mathcal{F}^{-1}(F_{low})) \quad (3)$$

对于高频分支, 首先对还原后的特征 X_{high} 进行层归一化处理 (LayerNorm), 以稳定训练过程并消除不同样本间的尺度差异, 得到归一化特征 $\hat{X}_{high}$; 随后通过一个 3×3 卷积层对归一化后的特征进行通道扩展, 生成扩展特征图:

$$X_{ext} = \text{Conv}_{3 \times 3}(\hat{X}_{high}) \quad (4)$$

接下来, 将扩展后的特征 X_{ext} 在通道维度上均匀切分为两部分, 分别记为主支 X_{main} 和辅支 X_{aux} , 以支持后续的多专家动态融合机制。在多专家动态融合阶段, 构建了一个包含五个异构子网络 (专家) 的并行系统, 每个子网络由 1×1 降维卷积, 本文设计的 LFSI 模块, 1×1 升维卷积三级组成。为了实现动态融合, 辅支特征 X_{aux} 经全局自适应平均池化提取通道统计量, 再输入全连接层生成各专家得分 S_i ; 通过对得分进行 Softmax 和归一化处理, 计算出每个子网络的权重 α_i 。为了提升效率, 采用 Top-K 稀疏选择, 仅保留得分最高的 3 个专家 $w1, w2, w3$, 其余子网络的权重置 0。最终, 选定专家在主支路上的输出通过加权求和得到高频增强特征 X'_{high} 。

$$X'_{high} = \sum_{i \in \text{Top-k}} \alpha_i \text{SubNet}_i(X_{main}) \quad (5)$$

低频分支经相似的流程处理, 最终与高频增强特征在通道维度拼接, 随后通过 LFSI 模块实现跨频段交互与融合, 输出经频域增强的精炼特征。

整体而言, FEEM 通过频域解耦策略, 将特征提取转化为频谱域的分治优化问题, 有效分离高频判别性细节与低频全局语义。同时, 多专家动态融合机制实现了输入自适应的子网络选择, 在放大结构显著信号的同时抑制噪声伪影, 从而为解码器提供更完整、更判别性的特征表达。该设计显著提升了医学图像病灶分割的精度、鲁棒性与跨数据集的泛化能力。

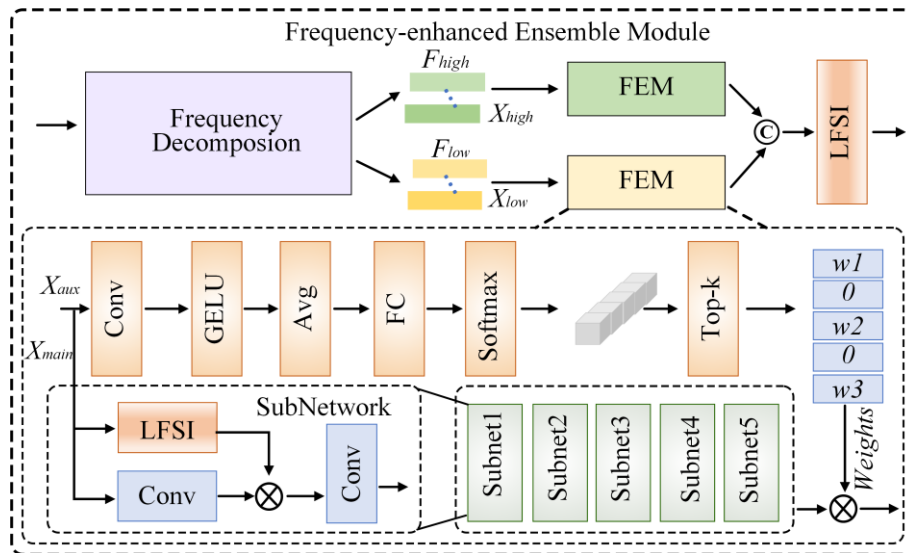


Figure 2. Network architecture diagram of FEEM

图 2. FEEM 的网络结构图

2.3. 局部细化模块

此外, 在解码阶段, SAM2-FNet 还引入了局部细化模块 (Local Refine Module, LRM), 利用空洞卷积金字塔与空间注意力机制, 有效缓解编码器与解码器之间的语义鸿沟, 进一步提升边缘重建的精度。

图 3 中所示的 Local Refine Module (LRM) 旨在从不同尺度的局部区域中提取更丰富的上下文信息, 以增强特征表达能力并精细化目标边界。具体而言, 输入特征首先通过三个平行分支分别进行空洞卷积 ($dilation = 1, 3$ 和 5), 以捕捉不同感受野下的局部细节; 在每个分支中, 特征图在卷积操作后与其原始输入通过元素级运算进行融合, 以保留必要的原始信息并对冗余特征进行初步筛选。随后, 各分支还引入最大池化 (Max Pool) 与平均池化 (Avg Pool) 两种操作, 用于在局部范围内进一步聚合显著特征并抑制噪声; 池化后的特征与先前的卷积结果相结合, 产生对局部区域更具辨别力的响应。最后, 三个分支的输出在通道维度或特征维度上进行整合 (Refine Module), 形成具有多尺度上下文的细化结果, 并通过与原始输入特征的残差连接或加权融合, 得到在边缘和小目标区域表现更加精细的特征表示。通过在网络解码或上采样阶段反复堆叠该模块, 可以逐步强化不同尺度下的局部细节, 有效提升整体分割精度与对复杂结构的识别能力。

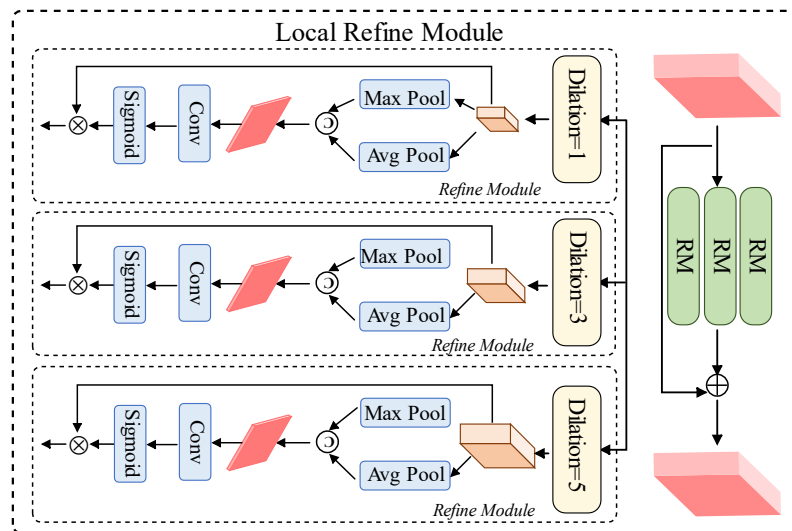


Figure 3. Network architecture diagram of LRM

图 3. LRM 的网络结构图

3. 数据集与实验环境

本文使用的是公共医学图像分割数据集 Kvasir-SEG, Kvasir-SEG[11]的息肉图像由内窥镜设备收集, 并由专业医生和经验丰富的胃肠病学家仔细标记。总体而言, 数据集由 1000 张图像组成, 分辨率从 332×487 像素到 1920×1072 像素不等。在训练设置方面, 我们按照 PraNet[12]将输入图像的大小调整为 352×352 , 并按照 CGMA-Net[13]并将整个数据集分为 80% 的训练集、10% 的验证集和 10% 的测试集。

本文的实验环境基于 ubuntu22.04 操作系统, 实验设备为 NVIDIA GeForce RTX 4090 GPU, 深度学习框架为 pytorch 2.1.0, cuda 版本为 12.1, 使用 python3.10 语言进行编程。具体的训练策略如表 1 所示。

Table 1. Training Parameters

表 1. 训练参数

实验参数	Kvasir-SEG
迭代次数	200

批量大小	8
输入尺寸	352*352
初始学习率	0.0001
优化器	AdamW

对于性能评估, 我们选择了四个指标来衡量预测地图与地面实况之间的相似性, 包括 DSC, IoU, ACC 和 SE (recall)。每个度量的定义如下。

1) 骰子相似系数: 骰子相似系数 (DSC) 通常用于测量地面实况和预测分割之间的相似性。它在下面定义:

$$DSC = \frac{2TP}{2TP + FP + FN} \times 100\%, \quad (6)$$

2) Intersection over Union: IoU, 用来量化预测分割和地面实况之间的重叠。它在下面定义。

$$IoU = \frac{TP}{TP + FP + FN} \times 100\%, \quad (7)$$

3) ACC

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%, \quad (8)$$

4) SE

$$SE = \frac{TP}{TP + FN} \times 100\%, \quad (9)$$

其中 TP、TN、FP 和 FN 分别表示真阳性、真阴性、假阳性和假阴性的数量。具体来说, DSC 和 IoU 说明了预测掩膜和真实掩膜之间的重叠程度, ACC 表示分割的准确性, 而 SE 则衡量识别病变区域的能力。

4. 实验结果与分析

为验证所提 SAM2-FNet 在医学图像分割任务中的泛化性与鲁棒性, 本研究在 Kvasir-SEG[25]公开数据集上与十四种最新方法进行了对比实验。这些方法涵盖了通用分割模型及医学专用模型。定量结果如表 2 所示, 在 Kvasir-SEG 息肉分割任务中, SAM2-FNet 以 91.6% 的 DSC 与 86.2% 的 IoU 刷新性能记录。相较于轻量级模型 HarDNet[42]其 DSC 由 88.5% 提升至 91.6%, 得益于 FEEM 模块动态频域分解对边缘细节的增强; 与参数量接近的 FCBFormer[43]相比, 局部细化模块 (LRM) 有效降低了 MAE, 由 2.5% 减少至 2.1%, 同时准确率提升至 97.3%。特别地, SAM2-FNet 在灵敏度 (SE) 上达到 93.1%, 较基准 SAM2UNet[14]提升 1.2%, 证明其对微小息肉的检出能力。尽管 TransUNet[24]虽然通过跨模态预训练获得了较高性能取得 89.2% 的 DSC, 但其参数量为 SAM2-FNet 的 1.4 倍, 并且在 MAE 和 IoU 上略低于本方法, 进一步凸显了 SAM2-FNet 在效率与精度之间的平衡优势。可视化对比结果如图 4 所示。

性能归因分析表明, SAM2-FNet 的卓越表现主要得益于其频域-空间域协同架构设计。具体来看, FEEM 模块借助二维傅里叶变换实现对输入特征的动态分解, 分别提取高频细节与低频语义信息; 多专家融合机制通过子网络自适应加权, 进一步强化了关键频段特征; 而局部细化模块则补偿了高频处理可能带来的全局信息丢失。在 Kvasir-SEG 数据集中, 对细粒度特征的精准刻画也显著提升了分割性能。

Table 2. Comparison of experiment results for various algorithms

表 2. 各类算法实验结果对比

Method	DSC	IoU	SE	ACC	MAE	Params/M
UNet[14]	79.5	69.4	80.7	91.9	7.8	7.1
UNet++[15]	75.2	64.5	76.0	93.0	6.7	9.0
Att-UNet[16]	87.1	82.0	-	95.5	4.4	23.7

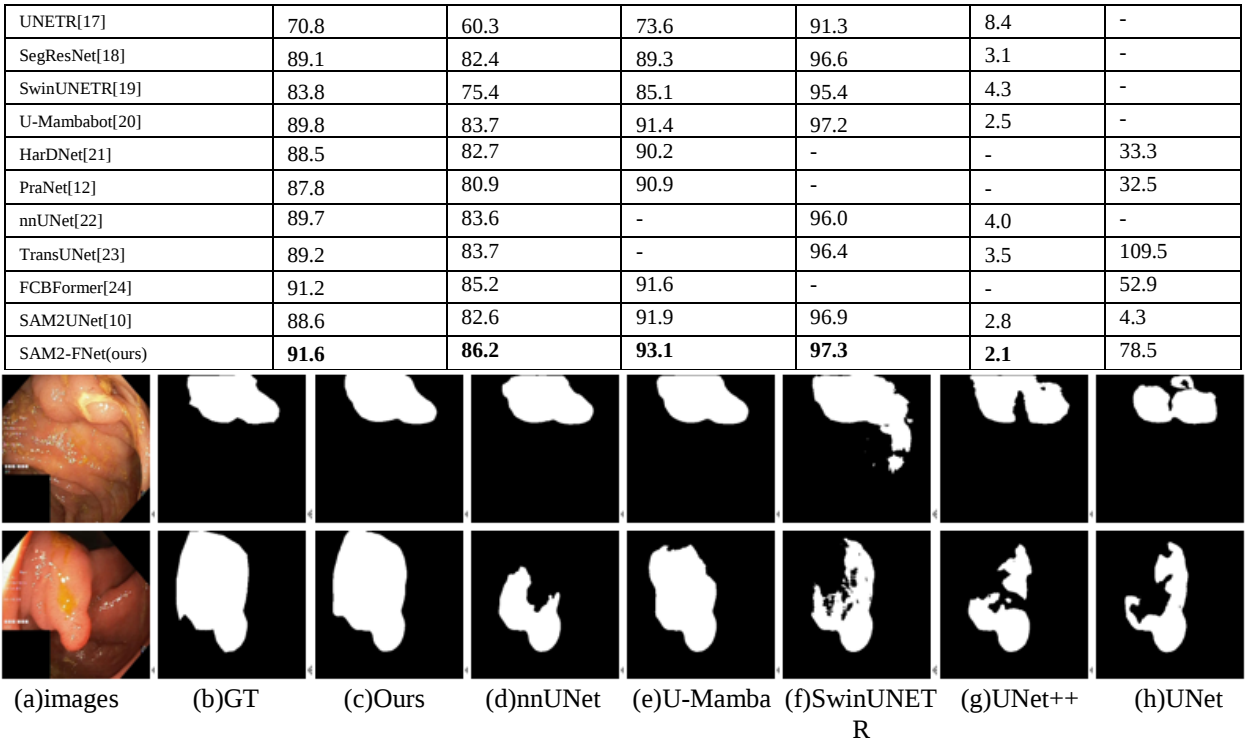


Figure 4. Visualization results
图 4. 可视化结果

为验证频域专家融合模块（FEEM）与局部细化模块（LRM）的有效性，本文在 Kvasir-SEG 数据集上开展了消融实验。实验结果如表 3 显示，在 Kvasir-SEG 数据集中，单独采用 FEEM 模块使模型的 DSC 和 IoU 分别提升了 2.3%和 2.5%；在联合 LRM 模块后，DSC 进一步提升了 0.7%，IoU 提升 0.9%，同时灵敏度提升至 93.1%，提升了 1.2%，验证了两模块协同作用在边界清晰病变分割中的显著优势。

在参数效率方面，仅引入 FEEM 模块使模型的参数量从 4.3M 增加至 77.3M，GFLOPS 从 6.6G 激增至 31.2G，这主要归因于频域变换和多专家子网络并行计算所带来的开销；而 LRM 模块仅使整体参数量增加 1.2M（达 78.5M）且 GFLOPS 仅略增 0.6G（达 31.8G），充分体现了其轻量化设计特性。尽管整体计算成本有所上升，但在 Kvasir-SEG 数据集中 DSC 取得了 3.0%的绝对增益，充分证明了 FEEM 与 LRM 模块的协同机制在频域与空间域跨模态优化下，显著增强了模型对复杂医学图像结构的解析能力。可视化结果如图 5 所示。

Table 3. The ablation experiments on Kavsir-SEG
表 3. 在 Kavsir-SEG 的消融实验

	baseline	FEEM	LRM	DSC	IoU	SE	ACC	Params/M	GFLOPS
①	✓			88.6	82.6	92.1	96.9	4.3	6.6
②	✓	✓		90.9	85.1	91.9	97.6	77.3	31.2
③	✓	✓	✓	91.6	86.2	93.1	97.3	78.5	31.8

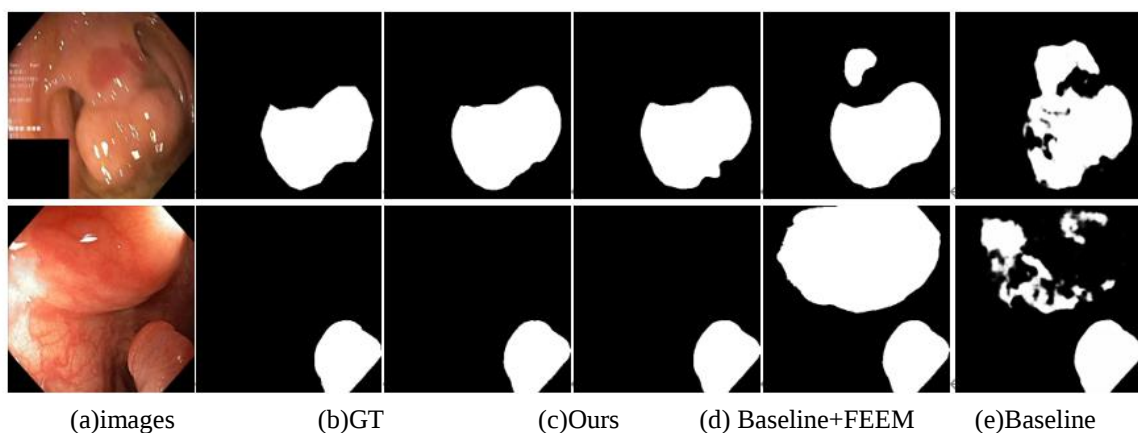


Figure 5. Visualization results of ablation experiments

图 5. 消融实验可视化结果

5. 总结

本文提出的 SAM2-FNet 利用 SAM2 编码器提取高质量多尺度特征, 通过引入频域专家融合模块 (FEEM) 与局部细化模块 (LRM), 实现了全局语义与局部细节信息的协同优化, 在 Kvasir-SEG 数据集上取得了显著的性能提升。实验结果表明, SAM2-FNet 在 DSC、IoU 及准确率等关键指标上均优于传统的 CNN 和 Transformer 模型, 尤其在处理低对比度与边界模糊病灶图像时展现出卓越的鲁棒性, 为临床精准诊断提供了有力的技术支撑。尽管如此, 所提出模型的 FEEM 模块通过多专家子网络并行计算虽显著增强了特征表达, 却也带来了较高的参数量和 GFLOPS, 可能在资源受限的实际环境中导致计算开销较大; 同时, 模型在部分低对比度病灶的检测上仍存在敏感性略有下降的问题, 这提示频域分解策略与专家融合机制之间仍需进一步精细调控。未来工作将致力于探索更高效的模块设计、参数压缩与优化技术, 以降低计算成本并提升实时性能, 同时结合半监督和领域自适应等新兴方法, 扩展模型在各类临床成像场景下的适用性和稳健性。总体而言, 本文为医学图像分割任务提供了一种新颖且高效的解决方案, 并为实现复杂临床场景下的高精度自动诊断指明了未来的发展方向。

参考文献

- [1] Li X, Meng X, Chen H, Fu X, Wang P, Chen X, Gu C, Zhou J. Integration of single sample and population analysis for understanding immune evasion mechanisms of lung cancer[J]. npj Systems Biology and Applications, 2023, 9(1): 4.
- [2] Li X, Feng X, Zhou J, Luo Y, Chen X, Zhao J, Chen H, Xiong G, Luo G. A multi-modal feature fusion method based on deep learning for predicting immunotherapy response[J]. Journal of Theoretical Biology, 2024, 586: 111816.
- [3] Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation[M]. arXiv, 2021.
- [4] Si, C., Yu, W., Zhou, P., Zhou, Y., Wang, X., & Yan, S. (2022). Inception transformer. Advances in Neural Information Processing Systems, 35, 23495-23509.
- [5] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollar P, Girshick R. Segment Anything[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 4015-4026.
- [6] Chen T, Zhu L, Ding C, Cao R, Wang Y, Li Z, Sun L, Mao P, Zang Y. SAM Fails to Segment Anything? -- SAM-Adapter: Adapting SAM in Underperformed Scenes: Camouflage, Shadow, Medical Image Segmentation, and More[M]. arXiv, 2023.
- [7] Ma J, He Y, Li F, Han L, You C, Wang B. Segment Anything in Medical Images[J]. Nature Communications, 2024, 15(1): 654.

- [8] Wu J, Ji W, Liu Y, Fu H, Xu M, Xu Y, Jin Y. Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation[M]. arXiv, 2023.
- [9] Cheng J, Ye J, Deng Z, Chen J, Li T, Wang H, Su Y, Huang Z, Chen J, Jiang L, Sun H, He J, Zhang S, Zhu M, Qiao Y. (2023). Sam-med2d. arXiv preprint arXiv:2308.16184.
- [10] Xiong X, Wu Z, Tan S, Li W, Tang F, Chen Y, Li S, Ma J, Li G. (2024). Sam2-unet: Segment anything 2 makes strong encoder for natural and medical image segmentation. arXiv preprint arXiv:2408.08870.
- [11] Jha D, Smedsrud PH, Riegler MA, Halvorsen P, De Lange T, Johansen D, Johansen HD. Kvasir-SEG: A Segmented Polyp Dataset[M]//Ro Y M, Cheng W H, Kim J, et al. MultiMedia Modeling: Vol. 11962. Cham: Springer International Publishing, 2020: 451-462.
- [12] Fan DP, Ji GP, Zhou T, Chen G, Fu H, Shen J, Shao L. PraNet: Parallel Reverse Attention Network for Polyp Segmentation[M]. arXiv, 2020.
- [13] Zheng J, Yan Y, Zhao L, Pan X. CGMA-Net: Cross-Level Guidance and Multi-Scale Aggregation Network for Polyp Segmentation[J]. IEEE Journal of Biomedical and Health Informatics, 2024, 28(3): 1424-1435.
- [14] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D. Attention U-net: Learning where to look for the pancreas[M]. arXiv, 2018.
- [15] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation[M]//Stoyanov D, Taylor Z, Carneiro G, et al. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Vol. 11045. Cham: Springer International Publishing, 2018: 3-11.
- [16] Attention u-net: Learning where to look for the pancreas[EB].
- [17] Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, Roth HR, Xu D. UNETR: Transformers for 3D Medical Image Segmentation[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2022: 1748-1758.
- [18] Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization[M]. arXiv, 2018.
- [19] Hatamizadeh A, Nath V, Tang Y, Yang D, Roth H, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images[M]. arXiv, 2022.
- [20] Ma J, Li F, Wang B. U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation[M]. arXiv, 2024.
- [21] Huang CH, Wu HY, Lin YL. HarDNet-MSEG: A Simple Encoder-Decoder Polyp Segmentation Neural Network that Achieves over 0.9 Mean Dice and 86 FPS[M]. arXiv, 2021.
- [22] Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature methods, 2021, 18(2): 203-211.
- [23] Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation[M]. arXiv, 2021.
- [24] Sanderson E, Matuszewski BJ. FCN-Transformer Feature Fusion for Polyp Segmentation[M]//Yang G, Aviles-Rivero A, Roberts M, et al. Medical Image Understanding and Analysis: Vol. 13413. Cham: Springer International Publishing, 2022: 892-907.