

SAM-Med3D 模型图示

算法整体框图

图1: SAM-Med3D 算法整体架构示意图。该模型采用完整的3D架构,包括**3D图像编码器**(基于3D ViT)、**提示编码器**和**掩码解码器**三个主要模块¹。输入为体积医学影像(例如CT扫描,尺寸约 $1 \times 128 \times 128 \times 128$),首先经过 Patch Embedding 将体数据划分为如 16^3 体素块并映射为序列嵌入。随后,**3D图像编码器**(Vision Transformer)提取图像的高维特征表示¹(例如产生512个token,每个维度384)。与此同时,**提示编码器**将用户提供的**稀疏提示**(点、框等位置提示)和**稠密提示**(先验掩码)编码为提示向量¹。最后,**3D掩码解码器**(Transformer解码器)接收图像特征和提示嵌入,通过交叉注意力融合提示信息,从而输出目标区域的分割掩码¹。图中标注了关键模块名称及典型的数据维度以供参考。

训练流程图

图2: SAM-Med3D 的模型训练流程图。该流程从带标注的3D医学影像(NIfTI文件,包括图像及其对应的真值掩码)开始。首先进行**数据预处理**,例如将体数据重采样或裁剪到统一大小(如 $128 \times 128 \times 128$),并进行归一化处理。然后基于真值掩码生成**模拟提示**:常用做法是在目标区域内随机采样正例点、在背景区域采样负例点,或生成覆盖目标的边界框作为提示²。接着,将预处理后的图像及生成的提示一同送入模型,执行**前向传播**得到预测掩码。模型预测结果与真值掩码进行比对,计算**损失函数**,这里采用Dice损失和交叉熵损失的加权组合³以同时兼顾区域重叠度和逐像素准确度。随后进行**反向传播**计算梯度,并通过优化器(如Adam)更新模型权重,迭代训练直至模型收敛。训练中使用Dice+CE复合损失已在医学分割任务中被证明是有效的策略³。

推理流程图

图3: SAM-Med3D 的推理流程图。在推理阶段,首先对输入的**测试3D图像**进行与训练时类似的预处理(如重采样到模型输入尺寸)。用户根据需要提供单点或多点的**交互式提示点**(可包括前景点或背景点)。预处理后的图像与用户提示一同输入**SAM-Med3D模型**,经过编码器和解码器模块,输出对应的**预测掩码**¹。对于体积较大的输入(超出 128^3 的范围),模型会采用分块推理策略:例如先截取围绕提示点的 128^3 补丁进行推理,若在补丁边缘检测到分割区域,则以50%重叠滑动窗口扩展推理范围,以平衡精度和效率⁴。模型生成的输出掩码通常为概率形式,后续经过**阈值化**(例如0.5二值阈值)转换为二值分割结果,并应用**连通域分析**移除小的孤立区域噪声,仅保留主要的连通目标⁵。经过上述后处理得到的最终掩码即为目标区域的三维分割结果。

¹ ² ³ ⁵ [openreview.net](https://openreview.net/pdf/8f15c2f7289fa0545a55b1df0304393501097ca8.pdf)

<https://openreview.net/pdf/8f15c2f7289fa0545a55b1df0304393501097ca8.pdf>

⁴ SAM-Med3D: Towards General-purpose Segmentation Models for Volumetric Medical Images

<https://arxiv.org/html/2310.15161v3>