

基于 *Semi-SAM+* 架构的医学图像分割系统开发手册

2025 年 6 月 27 日

摘要

本文档在 *Semi-SAM+* 半监督框架下，给出三维超声图像分割的完整工程实施细节。内容涵盖数据预处理、模型选型、Student / Teacher 数据流、伪标签生成与筛选、不确定性掩膜、一致性损失、EMA 更新、训练调度、风险缓解以及部署推理方案。硬件假设为 NVIDIA TITAN V (12 GB)。

目录

1	数据准备与预处理	3
1.1	数据集概况	3
1.2	典型预处理流程	3
2	模型对比与选择	3
3	整体流程图	4
4	Student 分支: Slicing SAM3D	4
4.1	切分-还原伪代码	4
4.2	粗掩膜公式	4
5	Teacher 分支: 伪标签生成	5
5.1	MedSAM2 专用预处理	5
5.2	多增广 + Prompt 队列伪代码	5
6	损失函数与训练调度	5
6.1	区域一致性损失	5
6.2	总损失组合	5
6.3	Warm-up & 全流程训练计划	6
7	EMA 教师更新	6

8	风险与缓解	6
9	部署与推理	6
9.1	显存预算 (FP16)	6
9.2	推理流程	6

1 数据准备与预处理

1.1 数据集概况

- 私有带标注：2000 份 3D 超声
- 公开带标注：400 – 500 份
- 私有无标注：5000 份
- 验证 / 测试：各 250 份

1.2 典型预处理流程

1. 灰度归一化：线性映射至 $[0, 1]$ ；RF 数据先做包络检测。
2. 重采样：统一体素间距 0.5^3mm ，裁剪/填充至 128^3 。
3. 标签格式：二值或 one-hot。
4. 批次采样： $|\mathcal{D}_L| : |\mathcal{D}_U| \approx 1 : 1$ 。

2 模型对比与选择

模型	编码框架	参数量	推理 (TITAN V)	备注
Slicing SAM3D	MedSAM-ViT-B(frozen)+3D CNN	88 M	0.30 s / 128^3	Student；仅训练 decoder
Tiny SAM-Med3D	TinyViT+MoE	31 M	0.25 s / 128^3	Teacher-1 (轻量)
SAM-Med3D	3D ViT-B+Prompt+3D CNN	93 M	patch sliding	Teacher-2 (零-shot)
MedSAM2	SAM2-ViT+Memory-Attn	600 M	100 slice/s	Teacher-3 (大模型)

表 1: 四种候选 Student / Teacher 模型对比 (FP16, 单 TITAN V)

3 整体流程图

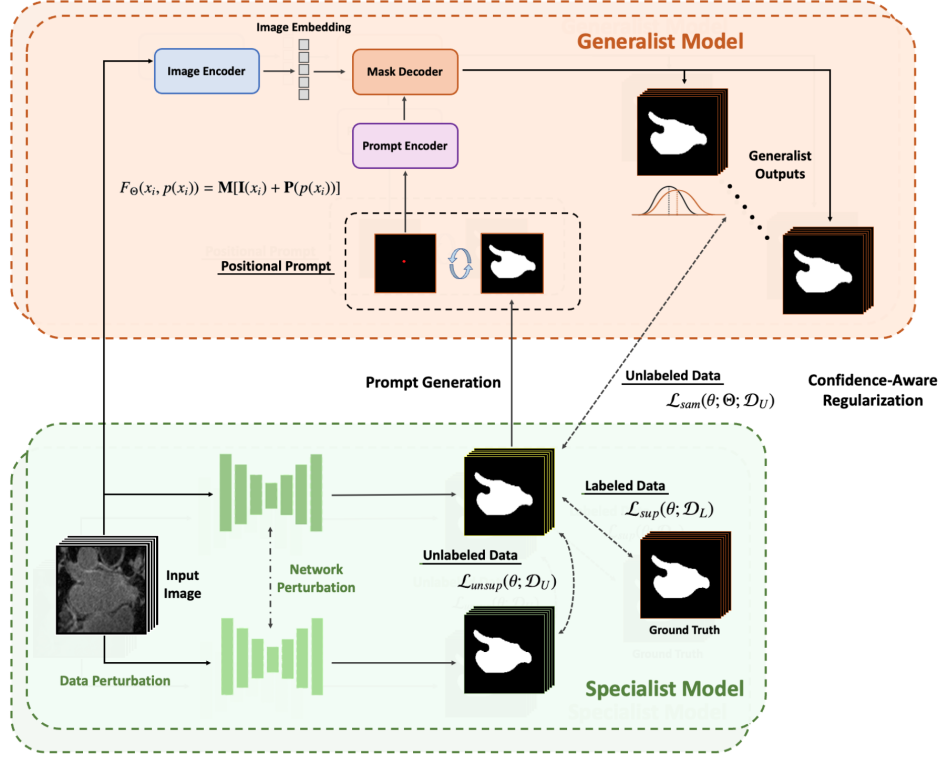


图 1: Semi-SAM+ 训练与推理流程

4 Student 分支: Slicing SAM3D

4.1 切分-还原伪代码

Algorithm 1 3D 切分 → 增强 → 推理 → 掩膜还原

```

1:  $Vol \leftarrow \text{load\_volume}()$  #  $D \times H \times W$ 
2: for  $z = 0$  to  $D-1$  step  $D_s$  do
3:    $blk \leftarrow Vol[z:z+D_s, :, :]$ ;  $pos \leftarrow (z, 0, 0)$ 
4:    $blk^{aug} \leftarrow \text{strong\_aug}(blk)$ 
5:    $\ell_s \leftarrow S_\theta(blk^{aug})$ ;  $p_s = \sigma(\ell_s)$ 
6:    $mask_{blk} = (p_s > 0.5)$ 
7:    $\text{write\_back}(mask_{blk}, pos)$ 
8: end for
9: return 重组后掩膜  $mask_{128^3}$ 

```

4.2 粗掩膜公式

$$M_{\text{coarse}} = (\sigma(S_\theta(x_u^{\text{aug}})) > 0.5).$$

5 Teacher 分支：伪标签生成

5.1 MedSAM2 专用预处理

1. 每帧双线性缩放到 512×512 并复制成 RGB；记录仿射矩阵 A 。
2. 前 K 帧形成记忆向量；其余帧用 *Memory Attention* 加速推理。
3. 推理后按 A^{-1} 将掩膜映射回原分辨率。

5.2 多增广 + Prompt 队列伪代码

Algorithm 2 多增广 Prompt 队列伪标签生成 & 方差筛选

Require: 无标注 3D 体 x_u ，其对应 Student 粗掩膜 M_c ，增广次数 N

```

1: gray# 初始化
2:  $\mathcal{P} \leftarrow []$  {收集所有  $\hat{M}$  的列表}
3: for  $j = 1$  to  $N$  do
4:   gray# 1. 随机一致几何增广
5:    $T_j \leftarrow \text{random\_geom}()$ ;  $img_j = \text{weak\_aug}(x_u, T_j)$ 
6:    $M_c^{(j)} = T_j(M_c)$  {将粗掩膜映射到增广空间}
7:    $p^{(j)} \leftarrow \text{sample\_point}(M_c^{(j)})$ 
8:    $E \leftarrow \text{encode\_image}_{\text{TinySAM}}(img_j)$ 
9:    $\hat{M}_m^{(j)} = \text{decode\_mask}(E, \text{mask} = M_c^{(j)})$ 
10:   $\hat{M}_p^{(j)} = \text{decode\_mask}(E, \text{points} = [p^{(j)}])$ 
11:   $\mathcal{P} += T_j^{-1}(\hat{M}_m^{(j)}), T_j^{-1}(\hat{M}_p^{(j)})$ 
12:  if  $\text{use\_MedSAM2}$  then
13:     $(img_{512}, M_{512}, p_{512}) \leftarrow \text{medsam2\_preproc}(img_j, M_c^{(j)}, p^{(j)})$ 
14:     $\hat{M}_{m2}, \hat{M}_{p2} \leftarrow \text{medsam2.predict}(img_{512}, M_{512}, p_{512})$ 
15:     $\mathcal{P} += T_j^{-1}(\hat{M}_{m2}), T_j^{-1}(\hat{M}_{p2})$ 
16:  end if
17: end for
18:  $\triangleright$  全部伪标签  $\mathcal{P} = \{\hat{M}_i\}_{i=1}^{2N(\text{或}4N)}$ 
19:  $\text{Var}(v) = \text{Var}(\{\hat{M}_i(v)\})$ 
20:  $M_{\text{conf}}(v) = \mathbb{I}[\text{Var}(v) < u_{\text{th}}]$ 
21:  $\text{Pseudo}(v) = \text{mean}\{\hat{M}_i(v)\} \cdot M_{\text{conf}}(v)$ 
22: return  $\text{Pseudo}, M_{\text{conf}}$ 

```

6 损失函数与训练调度

6.1 区域一致性损失

$$L_{rs-con} = \frac{\sum_v M_{\text{conf}}(v) (p_s(v) - p_f(v))^2}{\sum_v M_{\text{conf}}(v)}.$$

6.2 总损失组合

$$L_{\text{total}} = L_{\text{sup}} + \lambda(t)L_{\text{con}} + \beta(t)L_{rs-con}, \quad \begin{cases} \lambda(t) = \lambda_{\text{max}} \cdot \text{sigmoid}(\frac{t}{T_r}) \\ \beta(t) = \beta_{\text{max}} \cdot \exp(-t/T_d) \end{cases}$$

典型 $\lambda_{\max}=1$, $\beta_{\max}=0.3$, $T_r=10$ epochs (ramp-up), $T_d=40$ (ramp-down)。

6.3 Warm-up & 全流程训练计划

- **Phase-0 (Warm-up)**: 1800 带标注, 只训练 decoder 50 epochs; 学习率 2×10^{-3} 。
- **Phase-1 (Semi)**: 4000 无标注 + 1800 标注, Consistency+Pseudo, 100 - 150 epochs, Poly LR。
- **Phase-2 (Finetune)**: 1800+4000 全量, 再 50 epochs, 冻结 Teacher, 只调 Student。

7 EMA 教师更新

$$\theta'_t = 0.99 \theta'_{t-1} + 0.01 \theta_t.$$

每 iteration 后同步; 保存验证集 Dice 最优的 θ' 作为最终权重。

8 风险与缓解

- **显存峰值**: patch sliding、梯度累积、半精度。
- **I/O 瓶颈**: 多线程 dataloader + 缓存。
- **Prompt 生成**: 点+掩膜混合; 若 Teacher 输出抖动大, 升高 u_{th} 。

9 部署与推理

9.1 显存预算 (FP16)

- Student + Tiny SAM \Rightarrow <7 GB
- Student + MedSAM2 \Rightarrow 11 GB (`no_grad`)

9.2 推理流程

1. 预处理 \rightarrow Student 切片推理 \rightarrow 重组掩膜;
2. (可选) 医生点按/框选 \rightarrow MedSAM2 交互精修;
3. 导出 nii.gz / dcmseg。