



河北大学  
HEBEI UNIVERSITY

# 基于 SAM 大模型的肝脏肿瘤分割软件开发

答辩人：王恺

指导老师：刘琨

# 目录

绪论

理论技术基础

实验研究

结果分析与性能评估

总结

# 绪论

## 研究的背景及意义

近年来，深度学习在图像处理和计算机视觉领域取得了显著进展，其在医学图像分割任务中展现出了巨大潜力。众多深度学习模型，在该领域取得了优异的结果。这些研究不仅提高了分割的准确性，而且大大缩短了处理时间。

Meta 提出的 Segment Anything Model (SAM)，作为一种基于深度学习的全新模型架构，已在多个自然图像处理基准上显示出优异的性能。SAM 的引入为医学图像，特别是复杂的肝脏肿瘤分割任务，提供了新的可能性。

SAM 模型是 Meta AI 提出的一个基于 Transformer 的深度学习模型，它是在超过一亿张图像和相应标记的基础之上训练得来，用于执行零样本学习 (Zero-shot Learning) 中的图像分割任务，它能够通过接收简单的标注提示（如点、框等），自动产生精确的图像分割，这在处理医学图像中的不同器官和病变（如肝脏肿瘤）时，表现出独特的优势。使用 SAM 模型应用与医学图像处理具有以下优势：

- 泛化能力强：SAM 通过在大量图像上进行预训练，学习了丰富的视觉特征表示，具备强大的泛化能力，能够应对多样的图像分割任务。
- 少样本学习：在具有少量标注数据的情况下，SAM 仍然能表现出良好的分割效果，这对于医学影像分析尤其重要，因为高质量的医学标注数据往往难以获得。
- 自适应性强：作为一种自适应的模型，SAM 能够根据不同的输入图像和分割任务调整其行为，这使得它在面对肝脏肿瘤这种形态多样、边界模糊的目标时，表现出比传统方法更好的分割精度。



## 国内外研究现状

- 基于传统方法的分割

- 阈值法
- 主动轮廓法
- 区域生长法。

传统的分割方法噪声敏感，特征选择和提取过程复杂且依赖经验，难以确保正确分割。

- 基于深度学习的分割

- U-Net 及其变种
- 基于卷积对抗神经网络（CNN）的分割方法
- 基于 Transformer 的方法（如 SAM）

基于传统深度学习模型的分割方法对数据量需求大，所需数据的标注成本高，尤其是肿瘤分割等任务。泛化能力弱，模型在一个数据集上表现良好，但在其他数据集上性能可能显著下降。

## 研究内创新点

本文的创新之处在于使用了大型自监督学习模型 SAM 开发肝脏肿瘤分割软件，该研究方法的优势体现在以下几个方面：

1. 提高肝脏肿瘤分割的准确性和灵活性：利用 SAM 模型的强大特征提取能力，可以更准确地识别和分割肝脏中的肿瘤组织，特别是在肿瘤边界模糊或与周围组织对比度低的情况下。
2. 减少对标注数据的依赖：由于 SAM 模型基于自监督学习，在训练过程中不完全依赖标注数据，通过利用自监督学习，该软件能够利用未标注的医学图像进一步提升模型的性能和泛化能力。
3. 提升分割速度，支持实时应用：针对 SAM 模型的实现进行了微调优化，使得肝脏肿瘤分割软件能够快速处理图像，满足临床环境中对实时或近实时分析的需求。

# 理论技术基础



## 肝脏肿瘤 CT 图像

计算机断层扫描（Computed Tomography, CT）技术是一种先进的医学成像技术，通过利用 X 射线和计算机技术生成身体内部详细的横断面图像。

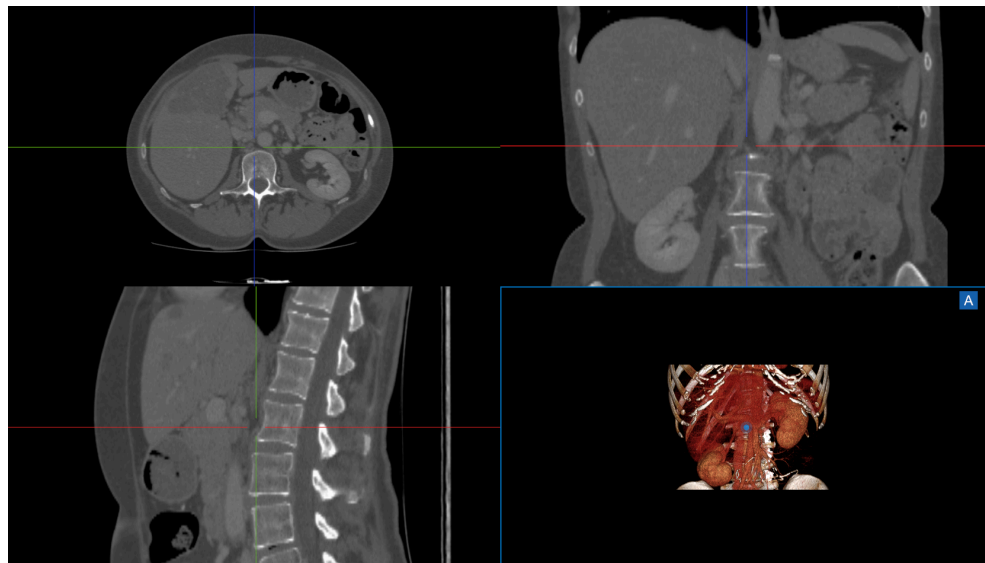


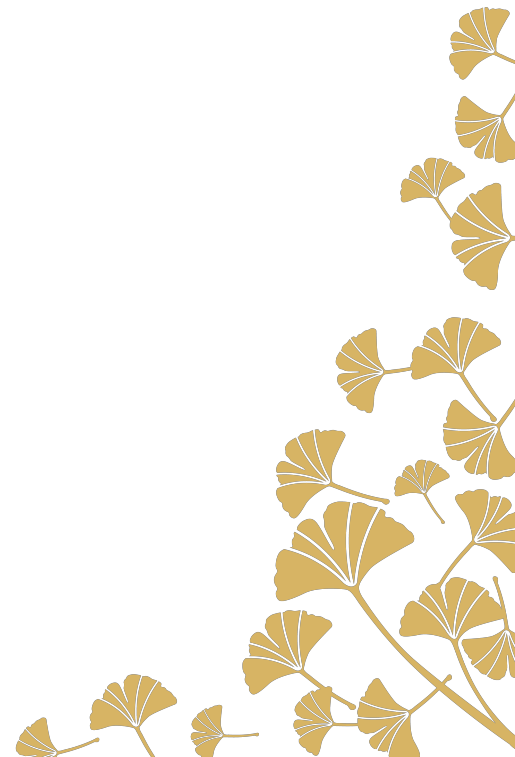
Figure 1: 腹部 CT 图像，左上三图为腹部 CT 图像 MPR 预览，右下为 3D 合成图像预览

CT 图像的优点有以下几点：

1. 高分辨率：能够清晰显示细微结构和病变。
2. 快速成像：现代 CT 扫描仪能够在几秒钟内完成扫描，适合紧急情况。
3. 多平面成像：能够生成横断面、冠状面和矢状面图像，提供多角度视图。
4. 三维重建：通过多个切片图像的组合，生成精确的三维图像。

但肝脏肿瘤的分割面临以下难点：

1. 肿瘤形态多样性
2. 图像噪声和伪影 CT
3. 肝脏结构复杂
4. 肿瘤与正常组织对比度低
5. 患者间差异
6. 数据标注困难



## 深度学习基本原理

深度学习是机器学习的一个分支，基于多层神经网络模拟人脑处理数据的方式。它涵盖从输入层接收数据，通过多个隐藏层处理，到输出层产生结果的过程。

## Pytorch 框架

PyTorch 是一个开源的机器学习库，广泛用于计算机视觉和自然语言处理等领域。

本文研究将使用 PyTorch 框架微调 SMA 模型。



## Transformer

Transformer 是一种基于自注意力机制的神经网络架构，最初由 Vaswani 等人在 2017 年的论文《Attention Is All You Need》中提出。Transformer 模型在自然语言处理（NLP）任务中取得了巨大成功，并被广泛应用于其他领域如计算机视觉（例如 VIT, vision Transformer）。

### Transformer 模型的组成

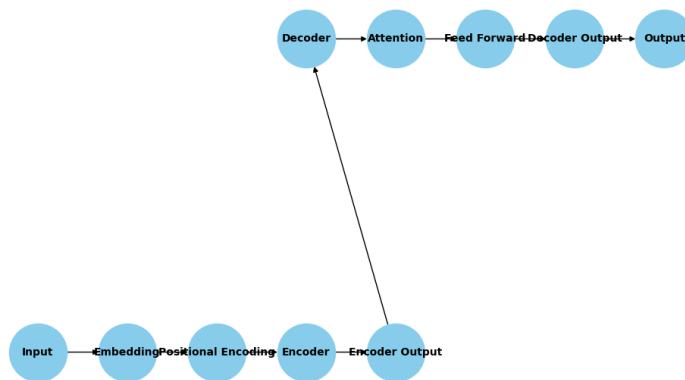


Figure 2: 图 2.3 Transformer 结构图

## (Vision Transformer, VIT)

VIT 模型是 Transformer 模型在 CV 领域的应用，用于把图像映射到特征空间，具体结构如图所示。

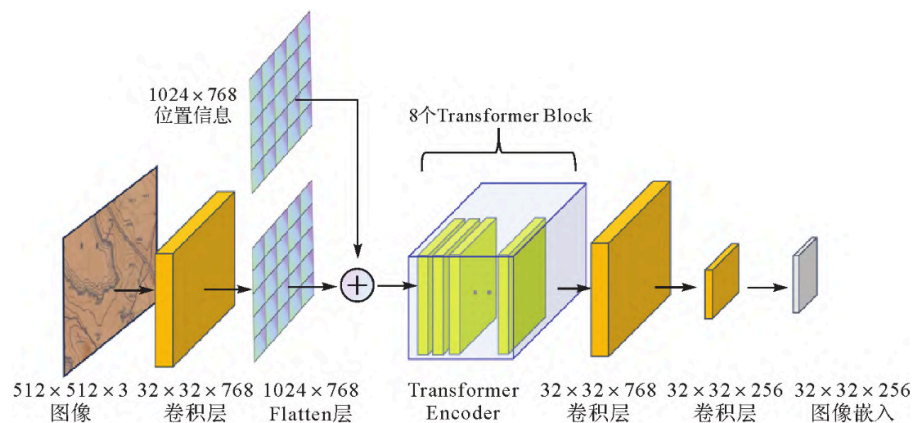


Figure 3: VIT 结构图

## 分割大模型（SAM 大模型）

SAM（Segment Anything Model）是近年来出现的一种新型深度学习模型，由 Meta 的 FAIR 实验室提出，旨在通过单一的模型实现对任何对象的高精度分割

SAM 通常在自然图像上表现优异，但是在特定领域如医疗影像，遥感图像等，由于训练数据集缺乏这些数据，SAM 的效果并不是理想。因此，在特定数据集上微调 SAM 是十分有必要的。

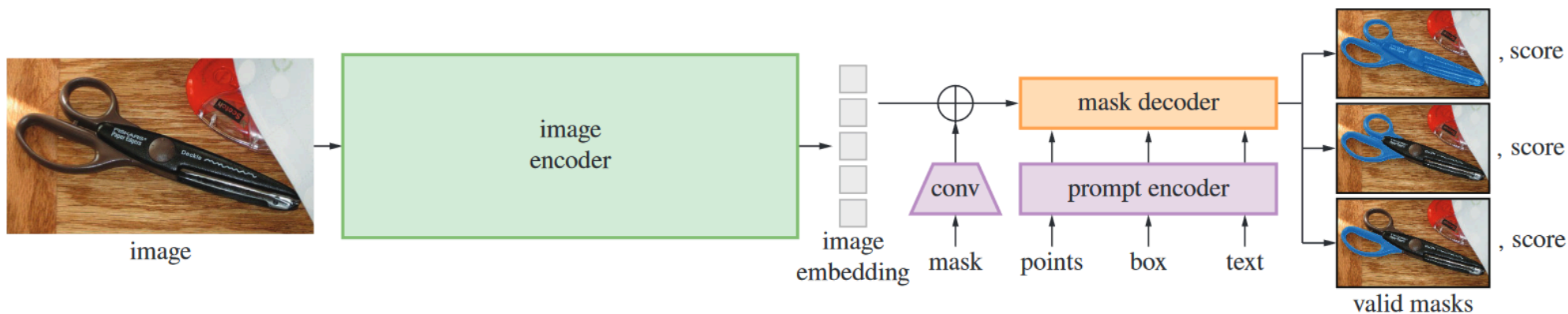


Figure 4: 图 2.5 SAM 模型的结构

## SAM 模型运行步骤

- 输入图像 (Image): 模型接收一幅输入图像。
- 图像编码器 (Image Encoder): 输入图像经过图像编码器 (通常是 VIT), 提取出高维度的图像特征。
- 提示编码器 (Prompt Encoder): 用户提供的提示信息 (如标记点、框等) 经过提示编码器, 提取出提示特征。
- 特征融合与加权 (Feature Fusion and Weighting): 图像特征和提示特征进行融合。具体操作是通过逐元素相乘 (element-wise multiplication), 使提示特征对图像特征进行加权。
- 掩码解码器 (Mask Decoder): 融合后的特征通过掩码解码器, 生成图像的分割掩码。
- 最终输出 (Final Output): 输出最终的分割掩码, 表示图像中的分割区域。

encoding), 加 2 个代表该点是前景/背景的可学习的嵌入 (embedding)。

# 实验研究



## 数据集处理

数据集共包括 20 个病例的 CT 图像，每个图像都附有肝脏肿瘤的标注（如表 1），便于对数据集的训练与测试。

文件名	文件内容	
PATIENT_DICOM	DICOM 格式的匿名患者图像	
LABELLED_DICOM	DICOM 格式分割的各个感兴趣区域对应的标签图像	
MASKS_DICOM	包含每个 mask 的 DICOM 图像的各个感兴趣区域的名称对应的一组新子文件夹	

Table 1: 数据集内容

读取数据集并进行数据分割和图像预处理，CT 值转换、窗口化、直方图均衡化、归一化、定义增强参数。模型训练和工作过程中的数据存储。

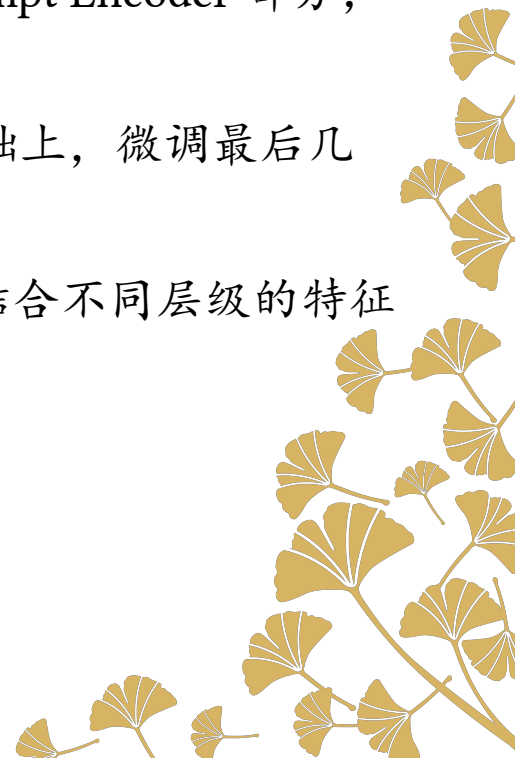
## 模型架构调整

加载 meta 预训练的 SAM 模型作为基础模型。

自定义 Prompt Encoder: 为了更好地适应肝脏 CT 图像特征, 设计或调整 Prompt Encoder 部分, 增加对解剖结构敏感的特征提取层。

微调 Mask Decoder: 在 Mask Decoder 部分, 保持原有 Transformer 架构的基础上, 微调最后几层的权重, 使其对肝脏和肿瘤边缘更加敏感。

多尺度特征融合: 引入多尺度特征融合策略, 如 U-Net 结构中的跳跃连接, 结合不同层级的特征来提高分割精度。



## 训练策略

迁移学习: 从预训练的 SAM 模型开始, 冻结部分或全部编码器层, 仅微调解码器层和可能新增的自定义层。

学习率调度: 使用余弦退火学习率调度 (Cosine Annealing), 初始学习率设为 $1^{-4}$ , 随着训练进行逐渐降低。

• 公式:

$$\text{lr}_t = \text{lr}_{\min} + \frac{1}{2}(\text{lr}_0 - \text{lr}_{\min})(1 + \cos(\frac{T_{\text{cur}}}{T_{\text{max}}}\pi))$$

使用训练数据集进行迭代训练。在训练过程中, 需要监控模型的性能指标如训练损失和验证损失。为了确保模型的泛化能力, 可以在每个训练周期结束后使用验证集进行评估。通过细致地微调, 提升模型对肝脏肿瘤图像特点的学习能力和分割准确性。

## 损失函数设计

**Dice Loss + BCE Loss:** 使用结合了 Dice Loss 和二元交叉熵损失（BCE）的复合损失函数，Dice Loss 关注区域的重叠程度，BCE 则强调分类的准确性。

**加权损失:** 对肿瘤区域施加更高的权重，因为肿瘤区域相对于正常肝脏组织更为关键，且数量较少，这样可以平衡类别不均衡问题。

## Dice 系数

**定义:** Dice 系数（也称为 Dice 相似系数）是一种用于衡量两个样本集合相似度的统计指标，尤其适用于评估图像分割的精度。

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} = \frac{2|A \cap B|}{|A| + |B|}$$

## 二值交叉熵损失 (BCE Loss)

对于每个像素的二值交叉熵损失定义如下：

$$\text{BCE} (p, g) = -(g \log(p) + (1 - g) \log(1 - p))$$

整个图像的 BCE Loss 同样需要通过对所有像素的 BCE 损失求平均得到：

$$\text{BCE Loss} = \frac{1}{N} \sum_{i=1}^N \text{BCE} (p_i, g_i)$$

BCE Loss 适用于二分类问题，常用于分割任务中的每个像素分类（如前景和背景的分类）。

后处理算法: 应用形态学操作（如开运算、闭运算）平滑分割边界，去除小孤立区域，或使用条件随机场（CRF）后处理提高边界界定质量。

# 结果分析与性能评估

为了全面评估基于 SAM 大模型的肝脏肿瘤分割软件的性能，选取主流的分割模型作为比较对象，U-Net 等在医学图像分割领域内已有大量成功应用，与之比较，可以更好的评估模型能力。

## 交并比 (IOU)

交并比 (Intersection Over Union, IOU) 也称为 Jaccard Index，是一种常用的评估指标，用于衡量预测分割与真实分割之间的重叠程度。用于衡量预测分割结果与真实标注之间的一致性。IOU 的计算公式表示如下：

$$\text{IoU} (A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{\sum_i p_i g_i}{\sum_i p_i + \sum_i g_i - \sum_i p_i g_i}$$

## 评估模型性能



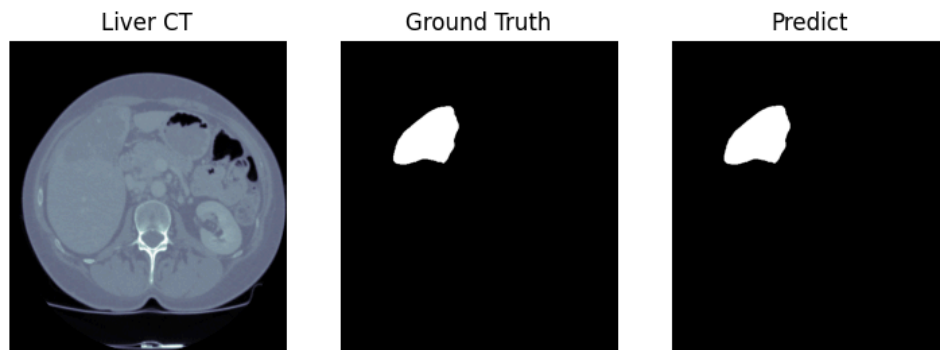


Figure 5: 分割结果展示

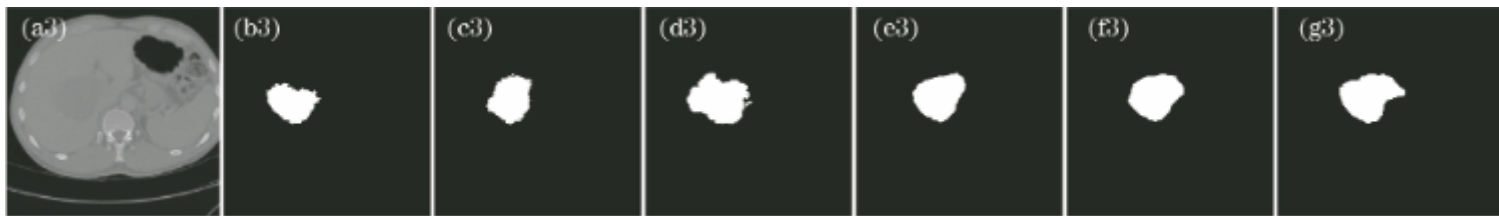


Figure 6: 其他模型的分割结果，从左至右依次为 原 CT 图像、UNet、SegNet、DeepLabv3、FC-DenseNet、堆叠树形结构空洞卷积模型、分割真值标签



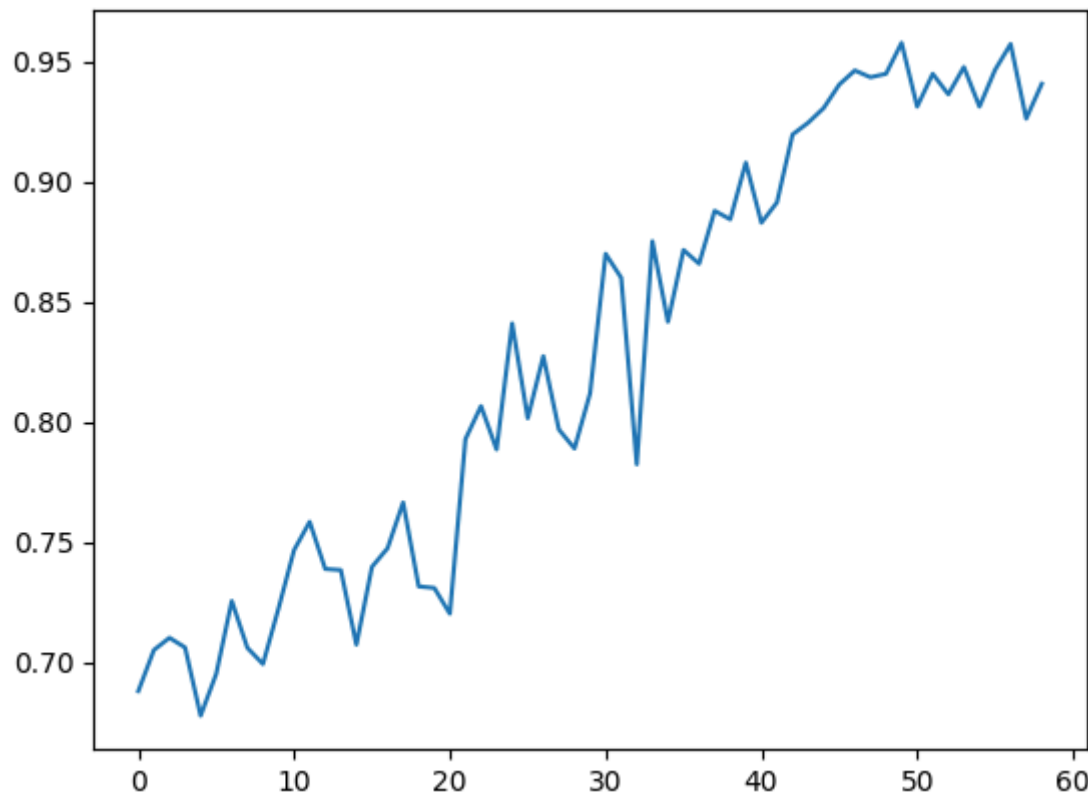


Figure 7: 图 4.3 模型训练过程 IOU 结果，横坐标为训练轮次，纵坐标为验证集 IOU 指标

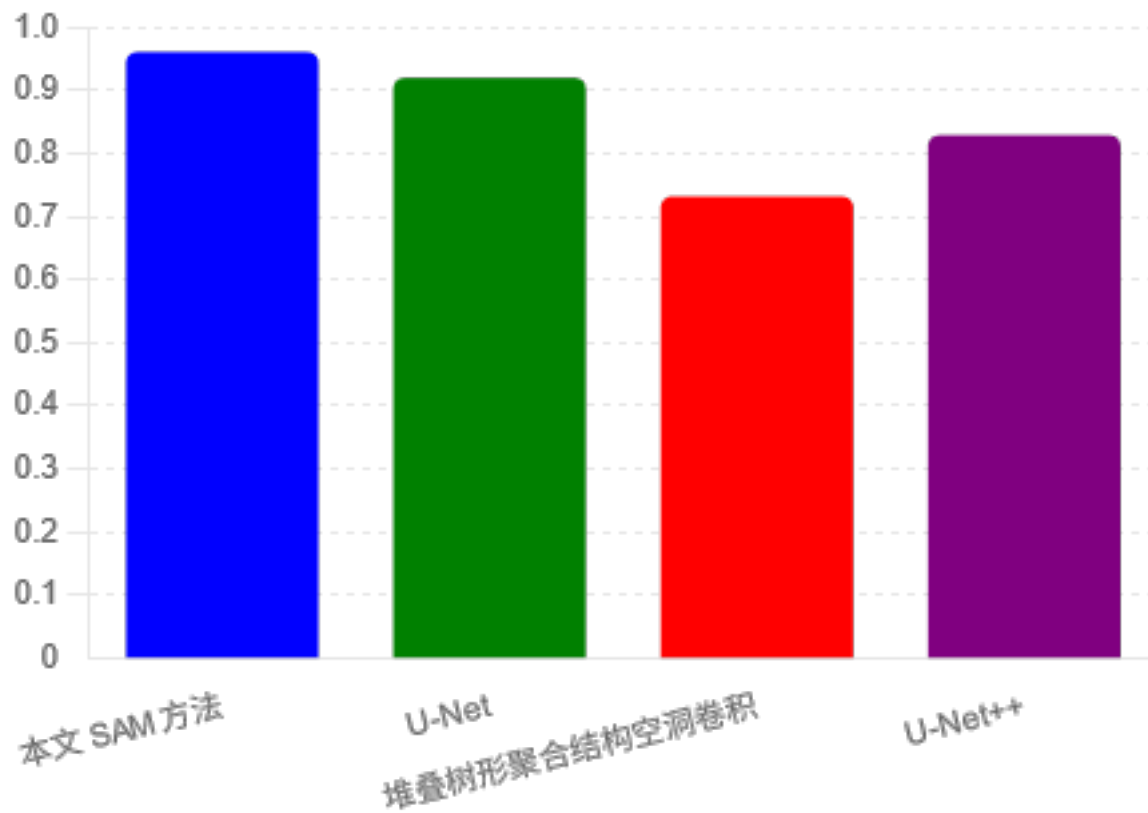
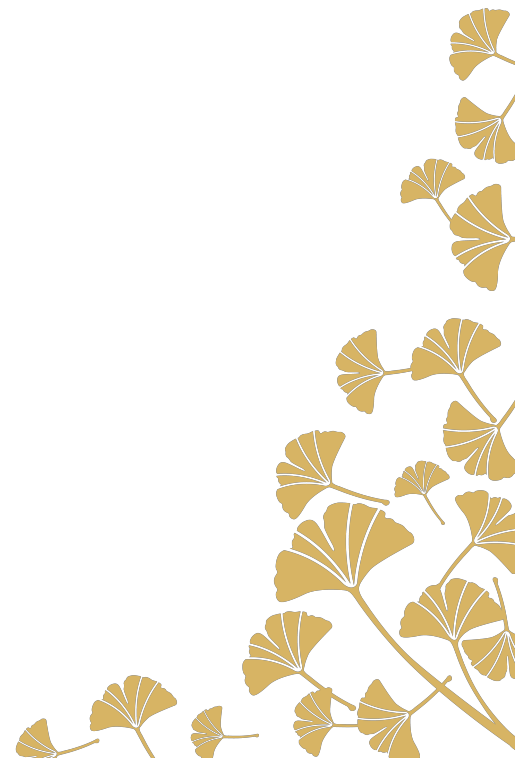


Figure 8: 不同模型的肝脏肿瘤分割 IOU

模型及其微调过程中涉及复杂的神经网络结构和大量的计算资源，训练和推理的时间和成本较高。在训练过程中冻结了模型的大部分内容，这在实际应用中可能限制其能力发挥。未来可以尝试通过模型压缩、剪枝和量化等技术来减少计算资源需求。



# 总结

尽管本研究在利用 SAM 模型进行肝脏肿瘤分割方面取得了显著成果，但仍存在一些不足和挑战。未来的研究可以通过多模态数据融合、迁移学习、实时分割、集成学习和临床合作等多方面的努力，进一步提升模型性能，扩大其应用范围，最终实现更高效、更准确的医学图像分割系统。

在基于 SAM 大模型的肝脏肿瘤分割软件开发上取得了很好的成效，将来可通过对模型的进一步优化微调与前端开发提升其性能与易用性。我希望在以后的学习中能在该研究方面继续不懈学习和研究，也希望能通过自己的努力能够为医疗领域图像处理奉献力量。



感谢老师指导!