



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

研究的背景及意义

使用 SAM 模型应用与医学图像处理具有以下优势：

国内外研究现状

现有智能分割方法及其分类

结论

研究内容

创新点

技术原理学习

深度学习基本原理

Pytorch 框架

Transformer

自注意力机制（Self-Attention Mechanism）

多头注意力机制（Multi-Head Attention）

分割大模型（SAM 大模型）

背景

SAM 模型结构

图像编码器 (Image Encoder)
提示编码器 (Prompt Encoder)

研究方法

准备数据集

实验过程

实验结果

结果分析与性能比较

准确度 (Accuracy)

Dice 系数

结论与展望

不足

展望

总结

研究的背景及意义

医学影像技术，特别是计算机断层扫描（CT）技术，在现代医疗体系中扮演着至关重要的角色。通过分析患者的影像数据，医生可以更好地理解患者的独特生理和病理状态，从而提供更加定制化的治疗方案。

肝脏肿瘤是世界范围内常见的肝脏疾病之一。准确快速地对肝脏肿瘤进行分割，在临床诊断、治疗规划以及手术导航中起着重要作用。肝脏肿瘤的早期诊断和治疗是提高患者生存率的关键。

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。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。



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

研究进展与实例分析

在最近的研究中，有相关论文表明，SAM 在处理带有阴影伪影和不均匀强度分布的超声图像时，能够有效识别并分割出肝脏肿瘤区域。此外，Mazurowski 等的 SAM 医学图像分割分析实验研究进一步验证了 SAM 在不同医学成像数据集上的广泛适用性，尤其是在处理肝 CT 图像时，SAM 显示出较高的 IoU 性能。

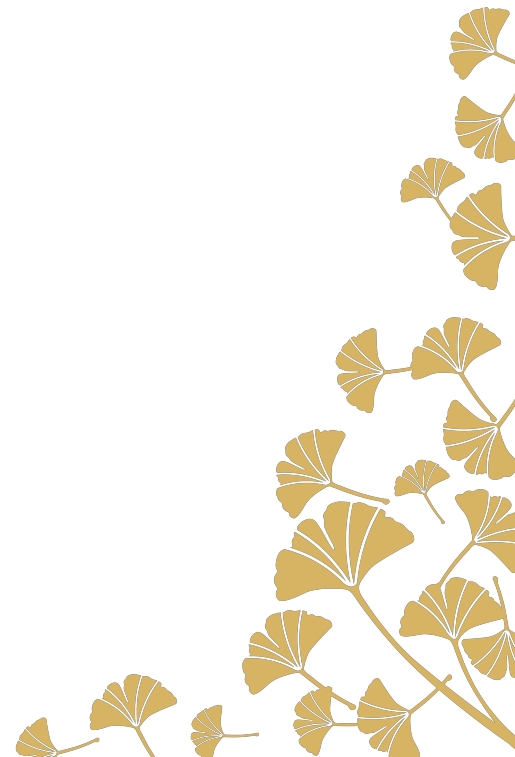




现有智能分割方法及其分类

作为新生的图像处理领域大模型，SAM 在医学图像处理领域还未得到广泛应用，目前主流医学图像分割方案多为使用深度学习技术的神经网络模型，以下是目前常用的分割方法。

- 传统的卷积神经网络 (CNN)
- U-Net 及其变种
- 基于 Transformer 的方法 (如 SAM)
- 集成方法





结论

总体来说，虽然多种方法各有优势，SAM 方法由于其强大的零样本学习能力和优异的上下文捕获能力，展现了在医学图像，特别是在肝脏肿瘤分割方面的独特优势。未来研究可以进一步探索如何优化 SAM 模型以适应具体的医学图像特征，以及如何将这些高级模型与医学专家的临床经验相结合，从而在提高肝脏肿瘤诊断和治疗的准确性方面发挥重要作用。



研究内容

- 第一章主要对介绍自动化肝脏肿瘤分割现有的可行方案和及未来发展方向进行了描述。讲述了深度学习图像处理的工作原理以及分类。讲解了评估模型的性能指标以及计算方法。
- 第二章简单介绍研究所用的各项技术原理，包括深度学习，pytorch 框架，SAM 模型。
- 第三章描述试验方法，首先安装模型与搭建微调必需的包，创建配置文件，然后使用数据集进行训练，最后遍历数据集训练，输出损失函数。
- 第四章选取其他几种领先的分割模型作为比较对象，与本研究模型比较分析，全面评估基于SAM 大模型的肝脏肿瘤开发软件的性能。
- 第五章对全文进行总结和展望。

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。

结论

《基于 SAM 大模型的肝脏肿瘤分割软件开发》通过引入前沿的自监督学习技术，不仅提高了分割的精度和效率，还降低了对标注数据的依赖，提供了一种新的解决方案来解决医学图像分割中的一些长期挑战。这些创新之处使得该软件在医学图像处理领域具有较高的应用价值和推广前景。



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。训练神经网络涉及使用损失函数评估预测误差，并通过反向传播和梯度下降等优化算法调整网络参数，以减少误差。此外，为避免过拟合，常用正则化技术如舍弃法（Dropout）。

深度学习在图像识别、语音处理等领域展示出显著的应用潜力，。深度学习模型能够从大量数据中自动学习到复杂的特征表示，这一点在图像识别、语音识别等任务中尤为显著。



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。它由 Facebook 的人工智能研究团队开发，并得到了包括微软、Salesforce 等多家大公司的支持和贡献。PyTorch 是一种强大的深度学习框架，适合从学术研究到商业应用的广泛用途。它的灵活性、用户友好的设计以及强大的社区支持使其成为当前最受欢迎的深度学习框架之一。

- PyTorch 动态计算图 (Dynamic Computation Graphs) 使得模型设计和调试更为简单直接。
- PyTorch 提供了丰富的 API，这使得构建复杂的神经网络变得更加容易。
- PyTorch 拥有一个活跃的社区可以帮助用户快速开始项目并解决遇到的问题。
- 由于其灵活性和简便性，PyTorch 广受欢迎。
- PyTorch 不仅易于使用，而且在性能方面也非常优秀。
- PyTorch 提供了与其他重要科学计算库的接口，这使得在不同框架之间转换模型变得更加容易。

Transformer

1. 编码器 (Encoder)

- 自注意力层 (**Self-Attention Layer**) : 计算输入序列中每个位置的注意力分数, 捕捉输入序列中的全局依赖关系。
- 前馈神经网络 (**Feed-Forward Neural Network, FFN**) : 对每个位置的输入进行独立的非线性变换。

2. 解码器 (Decoder)

- 自注意力层 (**Self-Attention Layer**) : 与编码器相同, 但只对解码器输入的前缀部分计算注意力分数。 -
- 编码器-解码器注意力层 (**Encoder-Decoder Attention Layer**) : 将解码器的输入与编码器的输出结合, 计算注意力分数, 捕捉输入与输出之间的依赖关系。
- 前馈神经网络 (**Feed-Forward Neural Network, FFN**) : 与编码器相同, 对每个位置的输入进行独立的非线性变换。

自注意力机制 (Self-Attention Mechanism)

自注意力机制是 Transformer 的核心组件，它允许模型在计算每个位置的表示时，考虑整个序列中的其他位置。其工作原理如下：

1. 输入变换：

- 将输入序列（例如词嵌入）变换为查询（query）、键（key）和值（value）向量：

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

其中， W_Q, W_K, W_V 是学习到的权重矩阵。

2. 计算注意力分数：

- 计算查询与键的点积，并进行缩放和软最大化（softmax），得到注意力分数：

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

其中， d_k 是键向量的维度。

3. 加权求和：

- 使用注意力分数对值向量进行加权求和，得到每个位置的新表示。



多头注意力机制 (Multi-Head Attention)

为了捕捉不同子空间的信息，Transformer 使用了多头注意力机制。具体做法是将查询、键、值向量分成多个头，每个头独立计算注意力，然后将结果拼接在一起，并通过线性变换得到最终的输出：

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O$$

其中， $\text{head}_i = \text{Attention}(QW_{Q_i}, KW_{K_i}, VW_{V_i})$ 。



分割大模型（SAM 大模型）

背景

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 大模型通过大规模数据的预训练和针对特定任务的微调，实现了对不同对象类型的准确识别和分割。

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。



SAM 模型结构

SAM 模型结构主要包括以下几个部分：

1. 图像编码器（Image Encoder）：

- 接收输入图像并提取图像特征。
- 通常使用预训练的视觉模型，如 VIT（vision Transformer）。

2. 提示编码器（Prompt Encoder）：

- 处理用户提供的提示信息（如点、框、文本等）。
- 提取提示信息的特征表示。

3. 掩码解码器（Mask Decoder）：

- 将图像特征和提示特征进行融合。
- 生成分割掩码。



4. 特征融合与加权（Feature Fusion and Weighting）:

- 通过逐元素相乘（element-wise multiplication）对图像特征进行加权。
- 利用提示特征对图像特征进行调整，提升分割精度。

5. 最终输出:

- 生成图像的分割掩码。



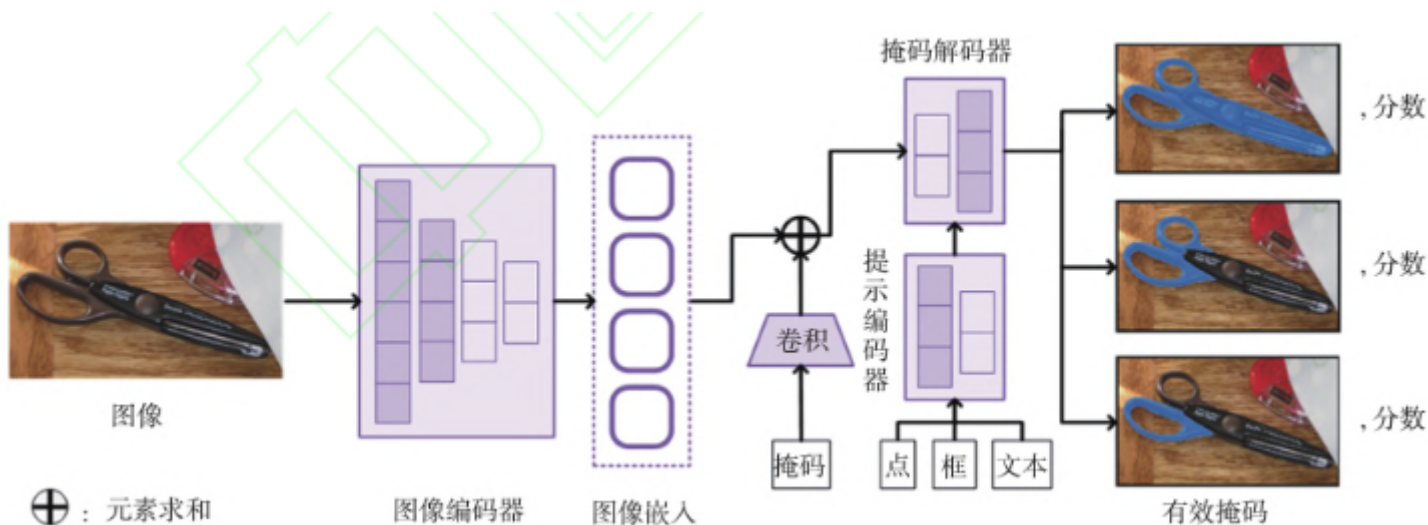
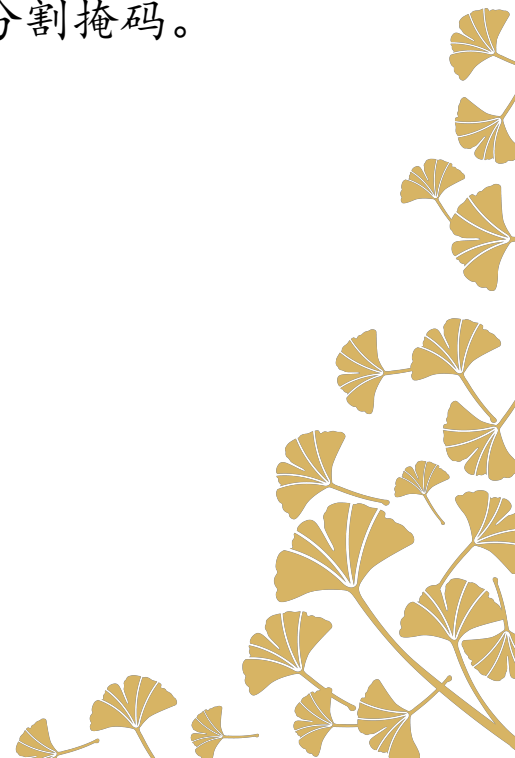


Figure 1: SAM 模型的结构

详细步骤

- 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。

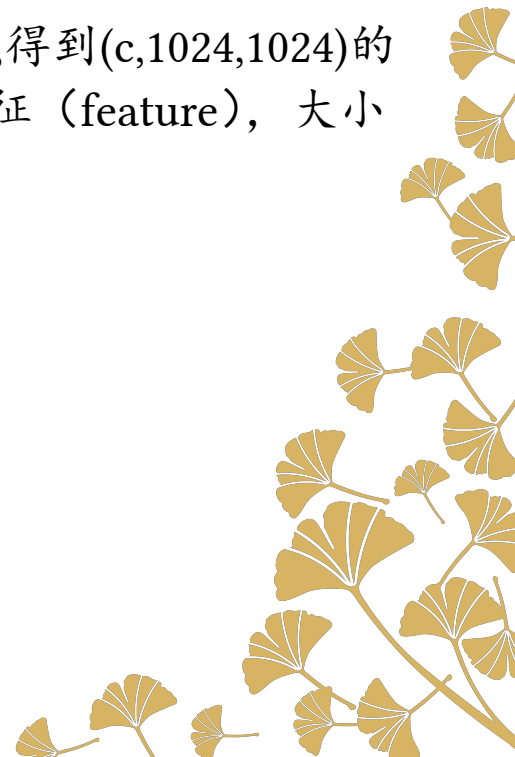
- 提示编码器（Prompt Encoder）：用户提供的提示信息（如标记点、框等）经过提示编码器，提取出提示特征。
- 特征融合与加权（Feature Fusion and Weighting）：图像特征和提示特征进行融合。具体操作是通过逐元素相乘（element-wise multiplication），使提示特征对图像特征进行加权。
- 掩码解码器（Mask Decoder）：融合后的特征通过掩码解码器，生成图像的分割掩码。
- 最终输出（Final Output）：输出最终的分割掩码，表示图像中的分割区域。



图像编码器（Image Encoder）

利用 mae 预训练的初始点（ViT），最低限度适应高分辨率的输入，该编码器在提示编码器（prompt encoder）之前，对每张图像只运行一次。

输入(c,h,w)的图像，对图像进行缩放，按照长边缩放成 1024，短边不够就 pad,得到(c,1024,1024)的图像，经过图像编码器（image encoder）处理，得到对图像 16 倍下采样的特征（feature），大小为(256,64,64)。



提示编码器（Prompt Encoder）

提示编码器分成 2 类：稀疏的(点，box，文本)，稠密的（mask）

- 点（point）：映射到 256 维的向量，包含代表点位置的位置编码（positional encoding），加 2 个代表该点是前景/背景的可学习的嵌入（embedding）。
- 框（box）：用一个嵌入对表示（1）可学习的嵌入代表左上角（2）可学习的嵌入代表右下角
- 文本：通过 CLIP 模型进行文本编码。
- 掩码（mask）：用输入图像 1/4 分辨率的掩码，然后用 (2,2) 卷积核步长为 2，输出通道数为 4 和 16，再用 (1,1) 卷积核将通道数量升到 256。掩码和图像嵌入（image embedding）通过逐元素相乘（element-wise），也就是说用掩码的特征对图像的特征进行加权使特定区域被放大或抑制。

准备数据集

数据集内容包括以下内容

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



实验过程

加载预训练模型

加载 meta 预训练的 SAM 模型作为基础模型。

冻层输入层，只微调顶部层

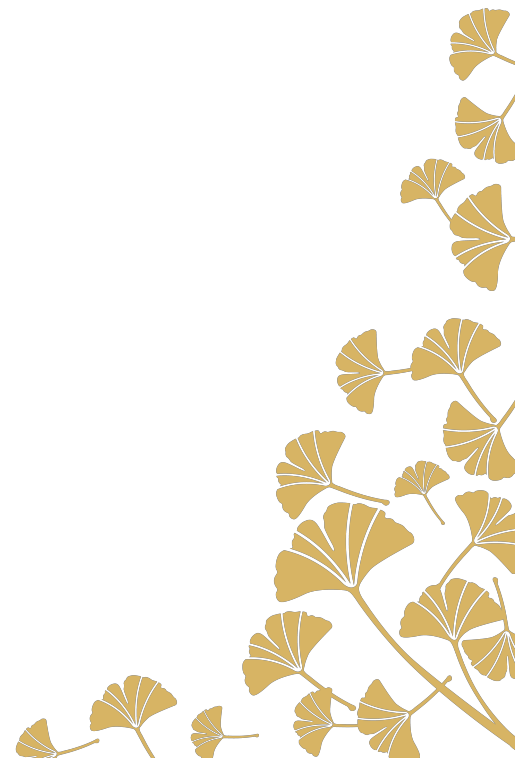
在加载预训练模型后，冻结模型的输入层，只对顶部层进行微调。冻结输入层可以防止在训练过程中更新这些层的权重，从而保留原始模型的特征表示。只微调顶部层有助于模型适应新任务的特定需求，同时减少训练时间和计算资源。

训练模型并监控性能，如损失值

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。如果验证损失在一段时间内不再下降，使用早停策略（Early Stopping）来终止训练。

实验结果

在搭建好的 SAM 大模型上实施肝脏肿瘤的分割任务，使用常见性能指标，准确度（Accuracy）、Dice 系数评价模型性能。通过这些指标，量化分析模型的分割效果，评估模型的训练效果并与其他分割方案比较。



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 U-Net 作为比较对象，U-Net 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。

准确度 (Accuracy)

计算方法：

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

优点：准确度简单直观，适用于类分布均匀的数据集。

缺点：当数据集类别不平衡时（即某些类别的像素数远多于其他类别），准确度可能会产生误导性结果。

Dice 系数

计算方法：

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

- A 为预测的二值掩码。
- B 为真实的二值掩码。
- $|A \cap B|$ 为预测与真实掩码的交集像素数。

优点：Dice 系数在处理类别不平衡的数据集时表现优异。它在较小的目标（如肿瘤）分割中，能够更好地反映模型的实际性能。

缺点：Dice 系数相对复杂，计算需要同时考虑预测结果与真实标签，尤其在处理多类别分割任务时，需要分别计算每个类别的 Dice 系数，再求平均值。

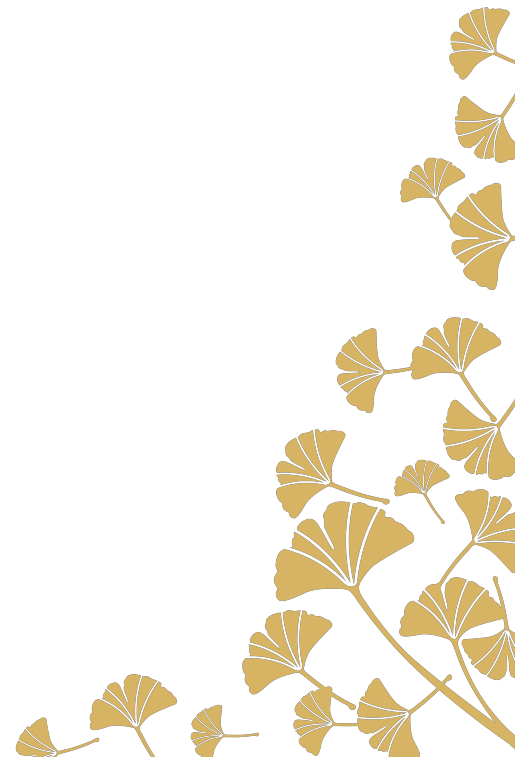
不足

1. 数据集局限性
2. 模型复杂性和计算资源需求
3. 参数调优难度
4. 实际应用验证不足



展望

1. 多模态数据融合:
2. 模型的迁移学习与领域自适应:
3. 实时分割与嵌入式应用:
4. 集成学习与模型融合:
5. 临床合作与反馈:



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!