

基于Adapter微调的轻量型SAM在超声图像分割中的研究

张霖娜

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2025年3月29日; 录用日期: 2025年4月22日; 发布日期: 2025年4月30日

摘要

随着医学影像技术的快速发展, 超声图像在疾病诊断和治疗中发挥着重要作用。然而, 超声图像具有噪声大、对比度低和边界模糊等特点, 给超声图像分割任务带来了巨大挑战。本文提出了一种基于Adapter微调的轻量超声图像分割模型LG-MedSeg, 旨在解决计算资源受限和超声图像分割精度不足的问题。通过知识蒸馏技术将SAM-Med2d模型轻量化, 并引入并行低秩分解Adapte (LRAdapter)模块, 增强模型对超声图像特征的适应能力。实验结果表明, LG-MedSeg在胸腔积液和腹腔积液分割任务中表现出色, Dice系数达到90.85%, IoU达到84.5%, 同时模型参数量仅为13.15 M, 显著降低了计算复杂度, 并在多个公开数据集上的实验验证了LRAdapter的有效性和通用性。本文的研究为医学图像分割提供了一种高效、轻量化的解决方案, 特别适用于资源受限的边缘设备部署。

关键词

超声图像分割, Adapter微调, 轻量化模型

Research on Lightweight SAM Based on Adapter Fine-Tuning for Ultrasound Image Segmentation

Linna Zhang

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 29th, 2025; accepted: Apr. 22nd, 2025; published: Apr. 30th, 2025

Abstract

With the rapid development of medical imaging technology, ultrasound images play a crucial role

in disease diagnosis and treatment. However, ultrasound images are characterized by high noise, low contrast, and blurred boundaries, posing significant challenges to ultrasound image segmentation tasks. This paper proposes LG-MedSeg, a lightweight ultrasound image segmentation model based on Adapter fine-tuning, aiming to address the issues of insufficient computational resources and inadequate segmentation accuracy for ultrasound images. By applying knowledge distillation technology to streamline the SAM-Med2D model and introducing a parallel low-rank decomposition Adapter (LRAdapter) module, the model's adaptability to ultrasound image features is enhanced. Experimental results demonstrate that LG-MedSeg achieves outstanding performance in thoracic effusion and abdominal effusion segmentation tasks, with a Dice coefficient of 90.85% and IoU of 84.5%, while maintaining only 13.15 million parameters. This significantly reduces computational complexity, and experiments on multiple public datasets validate the effectiveness and generalizability of the LRAdapter. Our research provides an efficient and lightweight solution for medical image segmentation, particularly suitable for deployment on resource-constrained edge devices.

Keywords

Ultrasound Image Segmentation, Adapter Fine-Tuning, Lightweight Model

Copyright © 2025 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

随着医学影像学技术的迅速发展,医学影像在疾病诊断、治疗决策以及手术规划等方面发挥着越来越重要的作用[1]。传统的医学影像分析方法依赖于专家的经验 and 人工标注,效率低且容易受到主观因素的影响。近年来,人工智能特别是深度学习技术已被广泛应用于医学图像分析领域,尤其是在医学影像分割任务中取得了显著的成果[2]。深度学习模型能够自动从复杂的医学影像中提取高层次的特征,从而实现高精度的组织或病变区域分割,为临床医生提供决策支持,极大提高了诊断效率和准确性。特别是在重大自然灾害、突发医疗事件等许多应急医疗场景下,经验丰富的医学影像分析专家是紧缺资源,人工智能辅助的医学影像分析可以为现场非专业人员提供专业指导,为患者得到及时的救治提供了强有力的支持。在众多医学影像检查技术中,超声检查技术以其低成本、无辐射、实时成像和便携的特点被广泛应用于各种临床医疗场景中。超声图像分析在监测病情、辅助治疗、患病风险预测等方面具有重要意义,推动了医学影像技术的发展和临床应用的提升。

基础模型是近年来在人工智能领域中取得革命性进展的一种模型,其通过对网络规模数据集的充分预训练和强大的零样本泛化能力,在广泛的下游任务中展现出了惊人的性能[3]。随着 OpenAI 推出 chatGPT,大语言模型迅速发展并取得了巨大成功,成为人工智能领域的里程碑。研究人员受到启发,开始探索计算机视觉领域中的大型视觉模型,Meta 随之也推出了分割一切基础大模型 SAM (Segment Anything) [4],在 1100 万个图像上训练了一个可提示模型,使用了能够实现强大零样本泛化的可提示分割任务。这些基础模型通常在大量数据上进行训练,具有强大的零样本迁移和少量样本迁移的能力,在解决许多下游任务时显示出强大的潜力,有助于加速准确且鲁棒的模型发展[5]。预训练大模型具有极大的价值,许多研究者将目标聚焦在如何合理利用预训练大模型,以提升一些下游任务的性能。

为促进 SAM 在医疗领域的应用,许多研究将目标聚集到对 SAM 进行微调,以适应医学图像的特点。Cheng J [6]等人通过在其搜集和整理的 460 万张图像和 19.7 万个掩码上进行微调得到 SAM-Med2 (SAM

to Medical 2D Images), 相对于 SAM 来说, SAM-Med2d 中蕴含更多的医学图像知识, 在多种成像方式的医学图像上都有较好的表现, 模型参数量约为 271 M。经过微调后的模型已经在多种模态的医学图像上都展现出不错的效果, 但由于许多医疗实际应用中的资源设备大多是边缘设备或便携式设备, 其计算资源是十分有限的, 并不能满足这些模型的资源需求。此外, 尽管 SAM 在多种医学图像模态上展现出良好的性能, 但在超声图像上的表现相对较弱。超声图像具有一些独特的挑战性特点, 例如图像噪声较大、对比度较低、边界模糊以及伪影干扰等, 这些因素使得模型在超声图像上的分割和识别任务更加困难。特别是对于胸腔积液和腹腔积液的研究, 由于这些区域的解剖结构复杂、图像特征不明显, 且数据标注难度较大, 导致相关研究相对较少。因此, 针对模型轻量化和超声图像的特定优化和微调策略仍需进一步探索, 以提升模型在这一领域的性能和应用效果。

针对计算资源限制的问题, 通过知识蒸馏的方式对模型进行轻量化, 可以减少计算资源的需求并加速训练和推理过程。知识蒸馏方法自从被 Hinton 等人[7]提出后, 广泛应用于解决模型轻量化的问题。Lucas 等人[8]通过实验证明, 确保为学生和老师模型提供一致的输入、合适的数据增强和非常长的训练计划是通过知识蒸馏使小模型表现良好的关键。MobileSAM [9]和 EfficientSAM [10]证明了对于 SAM 大模型进行知识蒸馏的有效性。但对于医疗领域, 相关研究仍非常有限。

微调策略的选择和实现方式也面临诸多挑战。在深度学习中, 为了将大规模数据集上预训练模型应用到其他领域, 一般通过微调方法来调整参数以适应目标任务。在医学图像分割领域, 全量微调通过调整全部参数实现对目标数据的充分拟合, 通常能取得较优性能, 但全量微调在实际应用中面临多方面的挑战。首先, 全量微调需要更新全部参数, 这个过程需要消耗大量计算资源; 同时, 全量微调的有效实施依赖于大量标注数据以防止过拟合, 而医学图像标注的专业性要求高、成本昂贵, 获取足量标注数据存在实际困难; 此外, 当目标任务的数据分布与源域存在较大的差异时, 全量微调可能导致模型出现灾难性遗忘现象, 即模型丢失在源域学习到的知识。这些限制因素制约了全量微调在医学图像分割领域的广泛应用。

为了克服全量微调的局限性并提升模型在医学图像分割任务中的适应效率, 研究者们提出了多种高效微调方法, 旨在以更低的计算成本和数据需求实现模型的快速迁移与优化[11]。其中 Adapter [12]微调和 Lora [13]微调被广泛应用。Adapter 模块作为一种独立的小型神经网络结构, 具有高度的灵活性和可扩展性。其模块化设计允许将其无缝集成到预训练模型的任意层级中, 这种特性使得 Adapter 在不同任务间的迁移和切换变得更为便捷, 仅需替换或添加相应的 Adapter 模块即可实现任务适配。相比之下, Lor 方法通过对模型权重进行低秩分解来实现参数更新, 缺乏模块化的特性, 在任务切换时需要重新调整低秩矩阵, 从而限制了其灵活性和实用性。相比之下 Adapter 的模块化架构不仅简化了实现过程, 还提升了系统的可扩展性, 使其能够快速适应不同的实验需求和部署场景。这种特性尤其适用于实时医学图像分割任务, 满足其快速迭代和高效部署的关键需求。

本文旨在针对超声图像分割任务, 特别是胸腔积液和腹腔积液的分割, 提出一种高效的微调策略。为了克服全量微调在计算资源和数据标注上的局限性, 本文将医疗图像领域的分割大模型 SAM-Med2d 通过知识整理的方式进行轻量化, 之后基于 Adapter 微调方法, 对轻量化的模型在超声数据集上进行微调。本文的目标是提升模型在胸腔和腹腔积液分割任务中的性能, 同时满足边缘设备对计算效率的需求, 为超声图像的实时分割应用提供一种可行的解决方案。

2. 模型设计与分析方法

本文提出的基于 Adapter 微调的轻量超声图像分割模型 LG-MedSeg, 其核心思想是对预训练医疗图像分割模型 SAM-Med2d 通过知识蒸馏的方式对其图像编码器进行轻量化得到 LG-MedSeg 模型, 将轻量

级的 Adapter 模块插入到轻量的图像编码器中，通过在超声数据集上进行微调，使模型能够快速适应超声图像分割任务，同时保留其在源域学习到的知识。

2.1. LG-MedSeg 的整体框架

LG-MedSeg 模型由三个主要部分组成：图像编码器、提示编码器和掩码解码器。图像编码器负责提取输入超声图像的层次化特征，提示编码器用于编码用户提供的交互信息(例如点、框等)，掩码解码器则根据图像特征和提示信息生成最终的分割结果(图 1)。

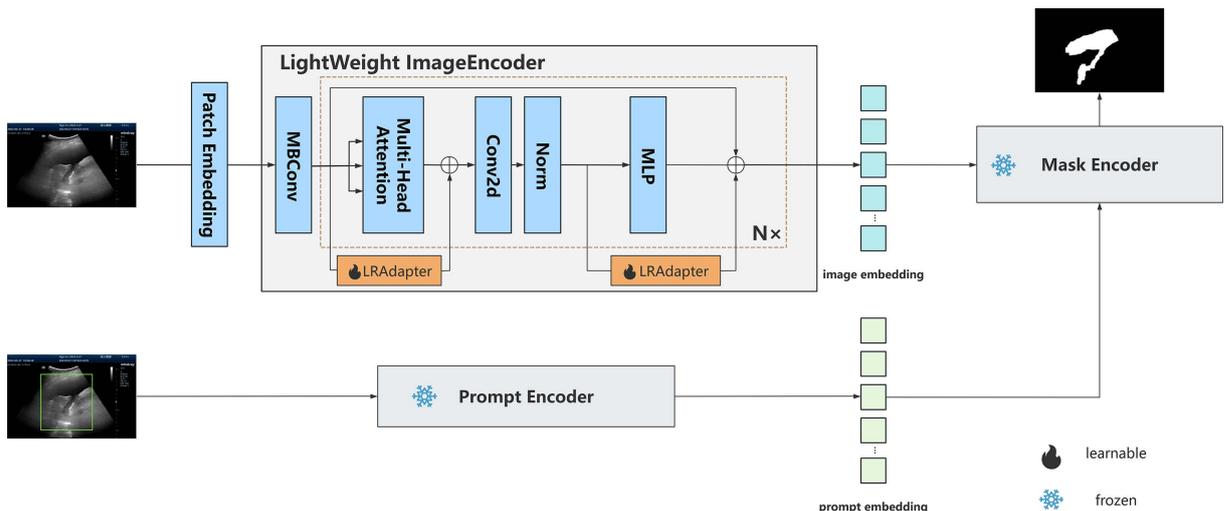


Figure 1. Overall architecture of LG-MedSeg based on Adapter fine-tuning

图 1. 基于 Adapter 微调的 LG-MedSeg 整体架构

为了将 LG-MedSeg 更好地应用到超声图像上，我们对图像编码器部分进行调整，将 LRAdapter 模块并行插入到图像编码器的多头注意力层和 MLP 层，用于对超声图像特征进行自适应调整。LRAdapter 模块是一种轻量级的神经网络模块，将会在 2.3 节中详细介绍其结构。在训练过程中，只需要对 LRAdapter 模块的参数在微调过程中进行更新，而其他部分的参数则保持不变。

2.2. Adapter 结构设计

LG-MedSeg 的图像编码器是一个轻量级的视觉 Transformer 模型，由多个 TinyViT Block 堆叠而成。每个 TinyViT Block 包含多头自注意力机制(MSA)、局部卷积层(Local_Conv)和多层感知机(MLP)。其中 MSA 层负责捕捉图像的长距离依赖关系，而 MLP 层则负责对特征进行非线性变换和特征提取。这两个层对模型的特征提取能力起着至关重要的作用。为了将图像编码器适配到超声图像分割任务，将 Adapter 模块插入到 MSA 层和 MLP 层，可以有效地增强这两个层的特征提取能力，使模型能够更好地适应超声图像分割任务。

全量微调(Full Fine-Tuning)是迁移学习中的一种常见策略，它涉及在预训练模型的基础上，对所有层或大部分层的权重进行进一步训练，以适应特定的下游任务，如图 2(a)所示。这种方法允许模型在保留预训练知识的同时，通过调整其参数来更好地捕捉新任务的特征和细节。然而，由于需要更新大量的参数，其需要较大的计算资源和时间成本，而且在小数据集上全量微调可能会导致过拟合，降低模型的泛化能力。在计算资源相对有限的情况下，Adapter 微调技术是一种更为高效和灵活的替代方案。Adapter 微调通过在预训练模型中插入小型网络模块，在保持原有模型参数不变的情况下，仅对这些新增的小型模

块进行训练，极大降低了计算成本和过拟合风险，同时保持了模型的强大表示能力。Adapter 微调特别适合于资源受限的环境和需要快速适应新任务的场景。

Adapter 模块的插入位置对模型性能至关重要。对 SAM 的图像编码器中，通常采用了在 MSA 后串行插入一个 Adapter 层的方式进行微调，如图 2(b)所示。这种方式在 SAM-Med2d 的任务场景中表现出了良好的效果，能够有效利用 Adapter 模块对特征进行自适应调整，同时保持较高的计算效率。然而，超声图像分割任务具有其独特的挑战，例如图像噪声大、对比度低以及目标边界模糊等，这些特点要求模型能够更灵活地捕捉全局上下文信息和局部细节特征。串行结构虽然能够在一定程度上适应目标任务，但由于 Adapter 模块与主网络串联，可能会对原始特征流造成一定的限制，导致信息融合的灵活性不足，难以充分适应超声图像的特征分布。

相比之下，并行结构能够更灵活地补充特征，增强模型的表达能力，同时保证训练的稳定性。因此，为了进一步提升模型在超声图像分割任务中的性能，本文提出在 MSA 和 MLP 层处分别并行插入 LRAdapter 模块，其结构如图 2(c)所示。这种设计不仅能够独立地对 MSA 层和 MLP 层的输出进行自适应调整，还能避免对原始特征流的干扰，从而更有效地适应超声图像的特征分布，提升分割精度。通过这种方式，本文的方法能够在保留预训练模型强大特征提取能力的同时，快速适应超声图像分割任务的独特需求，实现更高效、更准确的分割效果。

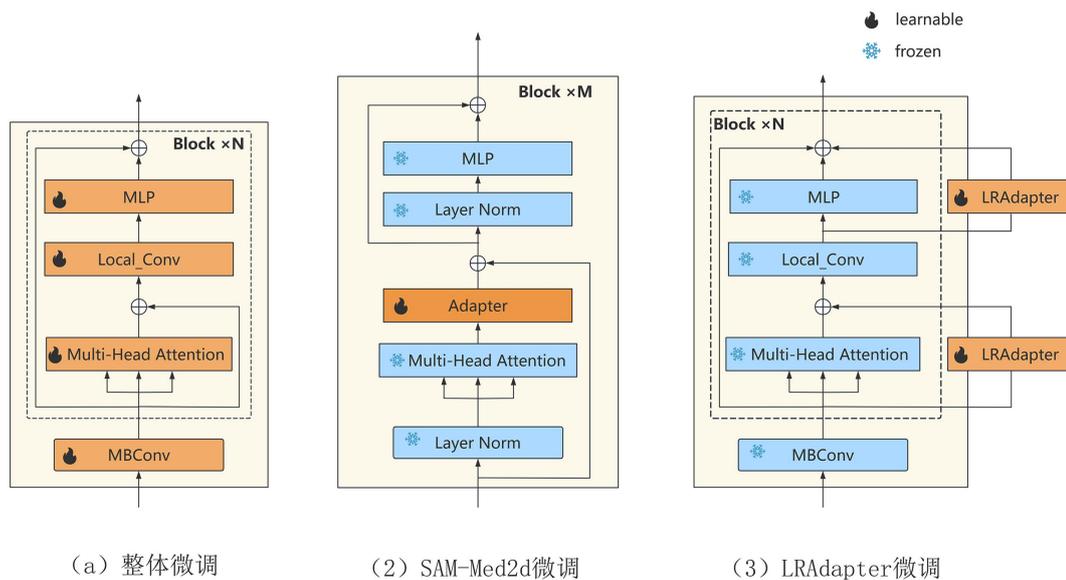


Figure 2. Comparison of different fine-tuning methods
图 2. 不同微调方式对比

2.3. 融入低秩分解的轻量化 Adapter

经典的 Adapter 结构由两个全连接层和非线性层组成，如图 3(a)所示。对于输入的特征向量 $x \in R^{d_m}$ ，原始的 Adapter 结构可以表示为：

$$\text{Adapter}(x) = W_{\text{up}} \cdot \text{NonLinearLayer}(W_{\text{down}} \cdot x) \quad (1)$$

其中，可以是任何非线性函数，例如 ReLU、GeLU、Sigmoid 等。 $W_{\text{down}} \in R^{r \times d_m}$ 是降维矩阵(forward_down_project)，将输入特征从 d_m 维降到 r 维， $W_{\text{up}} \in R^{d_m \times r}$ 是升维矩阵，将特征从 r 维升到 d_m 维。一般来说， r 的取值为 d_m 的 1/8 到 1/4 左右。需要注意的是，单纯为了减少模型参数量而过度降低 r 的值可能会导致特征空间

的表达能力不足。这种降维操作可能会限制模型捕捉输入特征多样性的能力，并导致高维信息的丢失，特别是对细节特征和全局上下文信息的建模能力产生负面影响。

为了保证模型的轻量，本文使用低秩分解 LRAdapter 结构，进一步减少微调过程中的参数量。具体而言，将 W_{down} 分解为两个低秩矩阵 $U_{down} \in R^{r \times k}$ 和 $V_{down} \in R^{k \times d_{in}}$ ：

$$W_{down} \approx U_{down} \cdot V_{down} \tag{2}$$

同样将 W_{up} 进行分解得到 $U_{up} \in R^{d_{in} \times r}$ 和 $V_{up} \in R^{k \times r}$ ：

$$W_{up} \approx U_{up} \cdot V_{up} \tag{3}$$

将低秩分解后的矩阵代入原始 Adapter 结构，得到低秩分解 LRAdapter 的公式：

$$\text{LRAdapter}(x) = (U_{down} \cdot V_{down}) \cdot \text{NonLinearLayer}((U_{up} \cdot V_{up}) \cdot x) \tag{4}$$

通过低秩分解后的 Adapter 参数量极大的减少，矩阵乘法的计算时间复杂度从 $O(d_{in} \cdot k)$ 降低到 $O(d_{in} \cdot k + k \cdot r)$ ，其中 $k \ll r$ 。

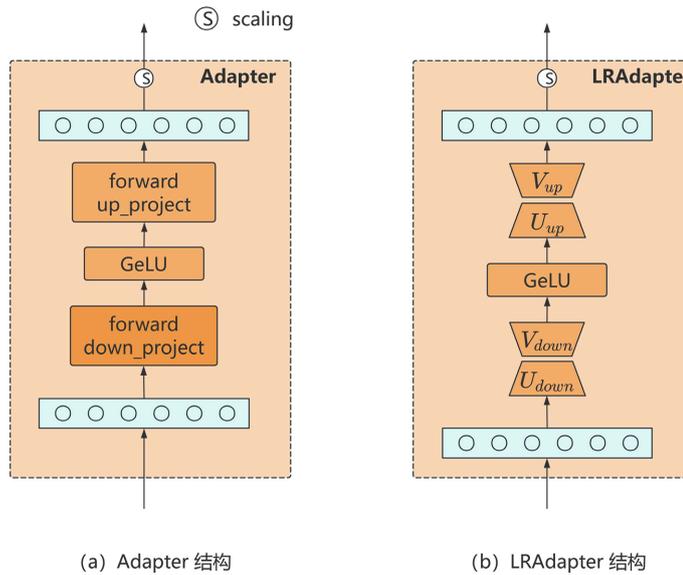


Figure 3. Comparison diagram of Adapter structures
图 3. Adapter 解构对比图

2.4. 损失函数设计

在基于 SAM 模型的训练中，通常使用的损失函数是一种综合的损失函数，该损失函数结合了 Focal Loss、Dice Loss 和 Mask IoU Loss 的优点。在医疗图像分割任务中，目标的边界通常模糊不清，且边界区域的像素数量远少于非边界区域，这导致传统的像素级损失函数难以有效地优化边界区域的分割结果。为了解决这一问题，在 SAM 原有的损失函数上进行优化，融入 Boundary Loss，专门用于增强模型对目标边界的关注能力。

Boundary Loss 的核心思想是通过计算预测边界与真实边界之间的距离，直接优化边界区域的分割结果。与传统的像素级损失函数不同，Boundary Loss 是一种基于边界的损失函数，能够更好地捕捉目标的几何形状和边界信息，通过计算预测边界与真实边界之间的距离来定义，使用 Hausdorff 距离进行距离度量。其定义公式为：

$$L_{\text{Boundary}} = \frac{1}{N} \sum_{i=1}^N H_{\text{dist}}(B_{\text{pred}}^i, B_{\text{gt}}^i) \quad (5)$$

其中 B_{pred}^i 是第 i 个样本的预测边界, B_{gt}^i 是第 i 个样本的真实边界, H_{dist} 是边界之间的 Hausdorff 距离。给定两个点集 A 和 B , Hausdorff 距离定义为:

$$H_{\text{dist}}(A, B) = \max \left(\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|a - b\| \right) \quad (6)$$

其中, \sup 表示上确界(最小上界), \inf 表示下确界(最大下界), $\|a - b\|$ 是点 a 和点 b 之间的欧几里得距离。

Focal Loss 是一种改进的交叉熵损失函数, 通过引入一个调节因子, 减少易分样本的损失贡献, 增加难分样本的损失贡献, 从而提升模型对难分样本的学习能力。其公式如下:

$$L_{\text{Focal}} = -\alpha \cdot (1-p)^\gamma \cdot \log(p) - (1-\alpha) \cdot p^\gamma \cdot \log(1-p) \quad (7)$$

其中 p 是模型预测的概率值, α 是类别权重, 用于平衡正负样本, γ 是 Focal 因子, 用于调节难易样本的权重, 在本文中设置 $\gamma = 2.0$, $\alpha = 0.25$ 。

Dice Loss 是基于 Dice 系数的损失函数, 广泛用于医学图像分割任务。Dice 系数衡量预测结果和真实标签之间的重叠程度, Dice Loss 则是 Dice 系数的补数。其公式如下:

$$L_{\text{Dice}} = 1 - \frac{2 \sum_i p_i \cdot g_i + \varepsilon}{\sum_i p_i + \sum_i g_i + \varepsilon} \quad (8)$$

其中 p_i 是模型预测的第 i 个像素的概率值; g_i 是真实标签的第 i 个像素的值(0 或 1); ε 是一个平滑项, 用于避免分母为零。

Mask IoU Loss 是一种基于交并比(IoU)的损失函数, 用于衡量预测掩码与真实掩码之间的重叠程度。其公式如下:

$$L_{\text{MaskIoU}} = \left(\frac{2 \sum_i p_i \cdot g_i + \varepsilon}{\sum_i p_i + \sum_i g_i - \sum_i p_i \cdot g_i + \varepsilon} - \text{pred_iou} \right)^2 \quad (9)$$

其中, p_i 、 g_i 和 ε 的定义与 Dice Loss 中一致, pred_iou 是模型预测的 IoU 值。

最终的损失函数为:

$$\text{Loss} = \lambda_1 \cdot L_{\text{Focal}} + \lambda_2 \cdot L_{\text{Dice}} + \lambda_3 \cdot L_{\text{MaskIoU}} + \lambda_4 \cdot L_{\text{Boundary}} \quad (10)$$

其中 λ_1 、 λ_2 、 λ_3 、 λ_4 是各损失项的权重。

3. 实验及结果分析

3.1. 数据集

实验部分共使用了四个超声图像数据集, 包括三个公开数据集 BUSI (Breast Ultrasound Images Dataset) [14]、DDTI (Digital Database for Thyroid Ultrasound Images) [15] 和 TN3K (Thyroid Nodule Region Segmentation Dataset) [16], 和一个私有数据集 USEFD (Ultrasound E-Fast Effusion Dataset), 包含了乳腺超声图像、甲状腺超声图像、胸腔积液和腹腔积液超声图像。

BUSI 数据集是一个经典的乳腺超声图像数据集, 包含了 600 名女性患者的 780 张图像, 患者的年龄范围在 25 至 75 岁之间。每张图像都附带有掩码真值图像, 这些掩码图像与原始图像一起呈现, 为乳腺

病变的分割提供了精确的标注。DDTI 数据集专注于甲状腺超声图像的分割任务，包含了 637 张超声图像，每张图像都带有像素级的标注，涵盖了多种甲状腺疾病，包括甲状腺炎、甲状腺肿、结节和癌症等。TN3K 数据集是一个专门针对甲状腺结节超声图像的数据集，包含了来自 2421 名患者的 3493 张图像，该数据集的每张图像至少包含一个甲状腺结节区域，为了避免数据冗余，TN3K 在同一区域或从同一患者的角度拍摄多张图像的情况下，仅保留了一张代表性图像。这种数据筛选策略确保了数据集的多样性和代表性，使得 TN3K 成为甲状腺结节检测和分割任务中的重要基准数据集。

USEFD 是一个专注于胸腔积液和腹腔积液检测的超声影像数据集，由来自上海某医院的 84 名受试者的数据构成。该数据集的目标是通过超声影像快速、准确地检测和定位积液区域，为 E-FAST (Extended Focused Assessment with Sonography in Trauma, 扩展创伤超声重点评估) 超声检查提供支持。数据集由 23,457 张图像和对应的 23,457 张掩码组成。每张图像都附带有精确的像素级标注，标注内容包括胸腔积液和腹腔积液。

3.2. 实验设置

所有模型的训练和推理均基于 Pytorch 深度学习框架，在一台配备 RTX8000 GPU 的服务器上进行。输入图像被统一调整为 256×256 的分辨率，并进行归一化处理，数据增强技术包括随机旋转、水平翻转和随机裁剪等，以提升模型的泛化能力。训练过程中使用 Adam 优化器，初始学习率设置为 5×10^{-4} ，batch size 设置为 16。微调过程中使用的数据集按照 8:1:1 的比例划分为训练集、验证机和测试集。损失函数各个部分的参数设置为 $\lambda_1 = 20.0$ 、 $\lambda_2 = 1.0$ 、 $\lambda_3 = 1.0$ 、 $\lambda_4 = 2.0$ 。实验中所使用的模型由训练 15 个 epoch 得到。

3.3. 实验设置

在 USEFD 数据集上，我们对四种模型(SAM-Med2d, LG-MedSeg, LG-MedSeg + Adapter 和 LG-MedSeg + LRAdapter)的性能进行了全面评估，主要指标包括 Dice 系数(Dice%)、交并比(IoU%)。其中，添加 Adapter 使用串行的 Adapter 结构，添加 LRAdapter 使用并行的 LRAdapter 结构。实验结果如表 1 所示。

Table 1. Performance comparison of various models on the USEFD dataset

表 1. USEFD 数据集上各模型性能对比

模型	Dice (%)	IoU (%)	Params (M)	FLOPs (G)	推理速度(ms)
SAM-Med2d	77.00	65.39	271.24	325.7	89.2
LG-MedSeg	79.34	67.53	13.08	12.8	15.3
W/Adapter	90.25	82.10	13.28	14.6	18.7
2W/LRAdapter (ours)	90.85	84.50	13.15	13.2	16.1

SAM-Med2d 和 LG-MedSeg 作为基准模型，其 Dice 系数分别为 77% 和 79.34%，IoU 分别为 65.39% 和 67.53%。这表明 LG-MedSeg 在分割精度上略优于 SAM-Med2d，但两者的性能仍有一定局限性。通过引入适配器模块提升了性能。串行引入 Adapter 的 Dice 系数达到 90.25%，IoU 达到 82.1% 并行引入 LRAdapter 进一步优化，Dice 系数提升至 90.85%，IoU 提升至 84.5%。添加 LRAdapter 之后，模型在 Dice 系数、IoU 上均有所提升。这表明 LRAdapter 通过更高效的参数利用和特征提取能够更适应超声图像分割。

在参数量方面，SAM-Med2d 的参数量高达 271.24 M，而 LG-MedSeg 的参数量仅为 13.08 M，极大地降低了模型复杂度。LG-MedSeg 在引入 Adapter 后，参数量增加 0.2 M，引入全连接降维 - 升维操作，

FLOPs 增加 14.1%；在引入 LRAdapter 后，参数量仅仅增加 0.07 M，通过低秩分解($k=16$)，FLOPs 仅增加 3.1%，优于普通 Adapter。并且值得一提的是，相对于经典的 Adapter，通过低秩分解的 LRAdapter 在参数量少于 Adapter 的情况下，其性能也略优于普通的 Adapter。并且低秩矩阵乘法非常高效，单张图片的推理速度仅增加 5.2%，接近原始模型。

添加 LRAdapter 之后，模型在 Dice 系数、IoU 上均有所提升。这表明 LRAdapter 通过更高效的参数利用和特征提取能够更适应超声图像分割，进一步优化了分割结果的精度和边界匹配效果。

接下来的实验将对 LRAdapter 中秩 k 的取值进行分析，当 k 太大时，参数量会相对较大，当 k 太小时，模型可能无法捕捉足够的特征，导致性能下降，因此对秩的选择是十分重要的。

Table 2. The impact of different ranks on the performance of LRAdapter
表 2. 不同秩对 LRAdapter 的性能影响

秩	Dice (%)	IoU (%)	Params (K)
$k=4$	68.22	57.68	17.8
$k=8$	84.54	76.22	35.7
$k=16$	90.85	84.50	71.4
$k=32$	90.88	84.55	142.8

由表 2 结果可知，当 k 取值较小时(如 $k=4$)，模型参数量虽少(仅 17.8 K)，但特征提取能力不足(Dice 仅 68.22%)；随着 k 增大($k=8$ 至 $k=16$)，模型性能显著提升，此时参数量仍保持合理范围；但当 k 继续增大($k=32$)时，性能提升趋于饱和，Dice 仅增加 0.03%，参数量却翻倍至 142.8 K。因此，综合性能与效率考虑，在 LRAdapter 中建议将秩设置为 16，该设置能在保持较高分割精度的同时控制模型复杂度。

为了验证 LRAdapter 的有效性，接下来在三个医学图像数据集(BUSI、DDTI 和 TN3K)上对比了有无 LRAdapter 的模型性能。实验结果如表 3 所示。

Table 3. Performance comparison of LRAdapter on the BUSI, DDTI, and TN3K datasets
表 3. LRAdapter 在 BUSI、DDTI 和 TN3K 数据集上的性能对比

数据集	w/o LRAdapter		w/LRAdapter	
	Dice (%)	IoU (%)	Dice (%)	IoU (%)
BUSI	89.12	81.03	93.08	87.35
DDTI	81.97	70.42	92.24	85.76
TN3K	81.62	70.44	91.26	84.62

实验结果表明，引入 LRAdapter 后，模型在多个数据集上的性能均得到了提升。具体而言，在 BUSI 数据集上，Dice 系数从 89.12% 提高至 93.08%，IoU 从 81.03% 提高至 87.35%。在 DDTI 数据集上，Dice 系数从 81.97% 提高至 92.24%，IoU 从 70.42% 提高至 85.76%。在 TN3K 数据集上，Dice 系数从 81.62% 提高至 91.26%，IoU 从 70.44 提高至 84.62%。这些数据充分证明了 LRAdapter 在提升模型性能方面的有效性。

为了验证损失函数在超声图像分割任务中的有效性，接下来的实验中将系统地对比了不同损失函数组合在超声图像分割任务中的性能表现，包括 Focal Loss、Dice Loss、Mask IoU Loss 和 Boundary Loss 的组合。实验在 USEFD 上进行，实验结果如表 4 所示。

Table 4. Ablation experiments on different loss functions
表 4. 不同损失函数消融实验

Focal Loss	Dice Loss	Mask IoU Loss	Boundary Loss	Dice (%)	IoU (%)	Hausdorff Dist
√	×	×	×	85.20	73.80	10.5
√	√	×	×	86.80	77.55	9.8
√	√	√	×	87.02	78.85	9.2
√	√	√	√	90.85	84.50	7.2

当仅使用 Focal Loss 时, 模型在测试集上的 Dice 系数为 85.2%, IoU 为 73.8%。这一结果表明, Focal Loss 能够有效解决类别不平衡问题, 增强模型对难分样本的关注能力。然而, 由于缺乏对分割重叠区域和边界的直接优化, 模型的分割性能相对有限, 尤其是在边界区域的精确度上表现较差(Hausdorff 距离较大)。

在 Focal Loss 的基础上加入 Dice Loss 后, 模型的性能进一步提升。Dice 系数从 85.2% 提升至 86.8%, IoU 从 73.8% 提升至 77.55%, Hausdorff 距离从 10.5 降低至 9.8。这一结果表明, Dice Loss 能够有效优化分割结果的重叠区域, 提升模型对目标区域的整体分割能力。然而, 由于缺乏对边界的直接优化, Hausdorff 距离仍然较高, 表明边界区域的分割精度仍有提升空间。

进一步加入 Mask IoU Loss 后, 模型的性能继续提升。Dice 系数达到 87.02%, IoU 达到 78.85%, Hausdorff 距离降低至 9.2。这一结果表明, Mask IoU Loss 能够有效提升模型对分割边界的精确度, 进一步优化分割结果的整体质量。

在 Focal Loss、Dice Loss 和 Mask IoU Loss 的基础上加入 Boundary Loss 后, 模型的性能达到最优。Dice 系数提升至 90.85%, IoU 提升至 84.5%, Hausdorff 距离进一步降低至 7.2。这一结果表明, Boundary Loss 能够直接优化边界区域的分割结果, 提升模型对边界区域的关注能力。与未使用 Boundary Loss 的模型相比, Hausdorff 距离的有所降低(从 8.5 降低至 7.2)表明, Boundary Loss 在提升边界分割精度方面具有重要作用。特别是在处理超声图像中噪声大、对比度低和边界模糊等问题时, 该损失函数组合表现出了较大的优势。

4. 总结与展望

本文提出了一种基于 Adapter 微调的轻量超声图像分割模型 LG-MedSeg, 通过知识蒸馏和低秩分解技术, 成功在降低模型复杂度的同时保持了较高的分割精度。并行 Adapter 结构的设计使模型能够更好地适应超声图像噪声大、对比度低的特点, 而多任务损失函数的引入则进一步提升了边界分割的准确性。实验验证了该模型在胸腔积液和腹腔积液分割任务中的优越性能, 同时展示了其在多个医学图像数据集上的广泛适用性。

尽管 LG-MedSeg 在实验中表现优异, 但其仍受限于对高质量标注数据的依赖以及在复杂解剖结构或罕见病变场景下的泛化能力有待提升。针对标注数据依赖性强的问题, 未来将探索无监督或弱监督学习方法, 利用未标注或部分标注数据提升模型的训练效率和泛化能力。同时, 结合模型压缩和硬件加速技术, 进一步优化推理效率, 推动 LG-MedSeg 在资源受限设备上的实时应用。

参考文献

- [1] Wei, W., Zhao, S. and Zou, D. (2023) Recommendation System: A Survey and New Perspectives. *World Scientific*

-
- Annual Review of Artificial Intelligence*, **1**, Article ID: 2330001. <https://doi.org/10.1142/s2811032323300013>
- [2] Xia, Q., Zheng, H., Zou, H., Luo, D., Tang, H., Li, L., *et al.* (2025) A Comprehensive Review of Deep Learning for Medical Image Segmentation. *Neurocomputing*, **613**, Article ID: 128740. <https://doi.org/10.1016/j.neucom.2024.128740>
 - [3] Zhang, C., Liu, L., Cui, Y., *et al.* (2023) A Comprehensive Survey on Segment Anything Model for Vision and Beyond. arXiv: 2305.08196.
 - [4] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., *et al.* (2023) Segment Anything. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 3992-4003. <https://doi.org/10.1109/iccv51070.2023.00371>
 - [5] Zhang, Y., Shen, Z. and Jiao, R. (2024) Segment Anything Model for Medical Image Segmentation: Current Applications and Future Directions. *Computers in Biology and Medicine*, **171**, Article ID: 108238. <https://doi.org/10.1016/j.compbiomed.2024.108238>
 - [6] Cheng, J., Ye, J., Deng, Z., *et al.* (2023) SAM-Med2D. arXiv: 2308.16184.
 - [7] Hinton, G., Vinyals, O. and Dean, J. (2015) Distilling the Knowledge in a Neural Network. arXiv: 1503.02531.
 - [8] Beyler, L., Zhai, X., Royer, A., Markeeva, L., Anil, R. and Kolesnikov, A. (2022) Knowledge Distillation: A Good Teacher Is Patient and Consistent. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 10915-10924. <https://doi.org/10.1109/cvpr52688.2022.01065>
 - [9] Zhang, C., Han, D., Qiao, Y., *et al.* (2023) Faster Segment Anything: Towards Lightweight SAM for Mobile Applications. arXiv: 2306.14289.
 - [10] Xiong, Y., Varadarajan, B., Wu, L., Xiang, X., Xiao, F., Zhu, C., *et al.* (2024) EfficientSAM: Leveraged Masked Image Pretraining for Efficient Segment Anything. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 16111-16121. <https://doi.org/10.1109/cvpr52733.2024.01525>
 - [11] Xin, Y., Luo, S., Zhou, H., *et al.* (2024) Parameter-Efficient Fine-Tuning for Pre-Trained Vision Models: A Survey. arXiv: 2402.02242.
 - [12] Houshy, N., Giurghi, A., Jastrzebski, S., *et al.* (2019) Parameter-Efficient Transfer Learning for NLP. *International Conference on Machine Learning*, Long Beach, 10-15 June 2019, 2790-2799.
 - [13] Hu, E.J., Shen, Y., Wallis, P., *et al.* (2022) LoRA: Low-Rank Adaptation of Large Language Models. arXiv: 2106.09685.
 - [14] Al-Dhabyani, W., Gomaa, M., Khaled, H. and Fahmy, A. (2020) Dataset of Breast Ultrasound Images. *Data in Brief*, **28**, Article ID: 104863. <https://doi.org/10.1016/j.dib.2019.104863>
 - [15] Pedraza, L., Vargas, C., Narváez, F., Durán, O., Muñoz, E. and Romero, E. (2015) An Open Access Thyroid Ultrasound Image Database. 10th *International Symposium on Medical Information Processing and Analysis*, Cartagena de Indias, 14-16 October 2014, 188-193. <https://doi.org/10.1117/12.2073532>
 - [16] Gong, H., Chen, G., Wang, R., Xie, X., Mao, M., Yu, Y., *et al.* (2021) Multi-Task Learning for Thyroid Nodule Segmentation with Thyroid Region Prior. 2021 *IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, Nice, 13-16 April 2021, 257-261. <https://doi.org/10.1109/isbi48211.2021.9434087>