

# 《Segment Anything in Medical Images and Videos: Benchmark and Deployment》

## 摘要

医学图像分割是计算机视觉领域的重要研究课题，广泛应用于医学诊断和治疗辅助决策中。随着深度学习和自监督学习技术的快速发展，基于生成式模型的图像分割方法取得了显著进展。本文介绍了最新的医学图像分割模型——SAM2，该模型基于 ViT 架构，并通过创新的内存注意力模块增强了医学视频分割的能力。SAM2 能够通过自监督学习自动适应不同类型的医学图像和视频数据，显著提高了分割精度和鲁棒性。实验结果表明，SAM2 在多个标准医学图像数据集上都表现出色，尤其是在视频分割任务中，相较于传统分割方法和之前的版本（如 MedSAM），SAM2 展现了更优的分割效果。本文还详细分析了 SAM2 的架构、训练过程以及其在实际医疗应用中的潜力，尤其是在自动化影像分析和智能诊断领域的应用前景。最后，本文讨论了该模型的挑战和未来的研究方向。

**关键词：**医学图像分割；ViT；内存注意力；视频分割；SAM2

## 1 引言

医学图像分割是计算机视觉和医学影像分析中的一项关键任务，它通过对医学图像中的感兴趣区域进行精确分割，帮助医生进行早期疾病诊断、治疗规划和疾病监测。医学图像包括但不限于 CT、MRI、PET、超声等影像数据，这些图像提供了人体内部结构和功能的详细信息 [1]。在临床实践中，分割技术广泛应用于癌症检测、器官和肿瘤的精确定位、脑部疾病的分析等多个领域。例如，早期癌症的检测往往依赖于对医学影像中肿瘤或病变区域的准确分割，从而为后续的治疗提供依据。

传统上，医学图像分割依赖于人工标注和规则基础的算法。然而，随着医学影像数据的日益复杂化，传统方法面临着精度和效率的挑战。近年来，深度学习方法，尤其是卷积神经网络（CNN）在医学图像分割任务中取得了显著进展 [2]，通过学习大规模的医学数据，深度学习模型能够自动进行高精度分割，大大提高了分割效率和准确性。

尽管深度学习技术在医学图像分割领域取得了显著进展，现有的医学图像分割模型仍然存在一些局限性。首先，许多现有的分割模型通常是针对特定任务、特定模态或特定解剖结构进行优化的，例如针对肝脏、肺部、脑部等器官的分割模型。这些专用模型在特定任务上能够取得较好的性能，但在其他任务上则表现较差，缺乏跨领域的泛化能力。

其次，传统的医学图像分割模型往往依赖于大量的标注数据进行训练，数据的稀缺性和标注的高成本使得这些模型的推广应用面临困境。在许多情况下，标注数据不仅需要专业知

识，而且每个数据集都需要单独训练模型，这大大增加了模型开发的时间和成本。此外，医学影像的多样性和复杂性（如不同成像模态之间的差异）也使得模型难以处理跨模态的分割任务。

基础模型（如 SAM）的出现提供了解决这些问题的一种可能途径。SAM（Segment Anything Model）[3] 作为一种基础分割模型，旨在通过大规模的训练数据和强大的泛化能力，解决传统专用分割模型的局限性。然而，尽管 SAM 模型在自然图像分割中表现出色，它在医学图像的分割任务中仍然面临一些挑战，例如医学图像的低分辨率、噪声干扰、复杂的解剖结构等问题。

SAM 模型（尤其是 SAM1）是第一种实现跨域分割的基础模型，能够在多个领域内执行分割任务，包括自然图像、视频等。然而，由于其训练数据主要来自自然图像，SAM 在处理医学图像时面临许多挑战。医学图像通常具有低分辨率、图像质量较差、解剖结构复杂等特点，这些特征使得 SAM 在医学领域的表现相对较差。此外，医学图像和自然图像之间存在明显的领域差异，SAM1 训练时未能充分考虑这些差异，因此直接应用于医学图像时常常无法得到理想的分割效果。

为了克服这些挑战，研究者们提出了 MedSAM [4]，它通过迁移学习对 SAM 进行了微调，使其能够适应医学图像分割任务。尽管 MedSAM 在 2D 医学图像分割上取得了显著进展，但它在 3D 医学图像和视频分割上的表现仍然不尽如人意。因此，基于 SAM 模型的新一代版本——SAM2 [5]，应运而生。SAM2 不仅增强了图像分割能力，还加入了视频分割功能，且通过改进的架构和训练方法，旨在克服 SAM1 在医学图像分割中的不足。

本研究的主要目标是评估 SAM2 在医学图像和视频分割中的表现，并与 SAM1 和 MedSAM 进行全面对比。具体来说，我们将重点探讨 SAM2 在处理不同医学图像模态（如 CT、MRI、PET、超声等）中的表现，并通过与 SAM1 和 MedSAM 的比较，分析 SAM2 在 2D 和 3D 医学图像分割任务中的优势和不足；此外，本研究还将探索 SAM2 在医学视频分割中的应用。传统的医学图像分割主要集中在静态图像上，而随着医学影像采集技术的发展，视频分割逐渐成为一个重要的研究方向。SAM2 的引入视频分割能力，能够处理动态医学数据（如超声和内窥镜视频），为医学视频分析提供新的解决方案。因此，本研究还将评估 SAM2 在医学视频分割中的表现，并探讨其在临床应用中的潜力；最后，还将通过迁移学习的实验，探讨如何通过微调 SAM2 来提升其在特定医学图像任务中的表现，特别是在 3D 医学图像分割中的适应性。通过本研究，希望能够深入了解 SAM2 在医学领域中的应用前景，并为未来的研究提供有价值的见解。

## 2 相关工作

### 2.1 以往的医学图像分割方法

医学图像分割是医学图像分析中至关重要的任务，旨在将感兴趣区域（ROI）从复杂的医学图像中分离出来，帮助医生进行疾病诊断、治疗规划和监测。传统的医学图像分割方法大体可以分为两类：基于阈值的方法和基于人工设计特征的机器学习方法。

#### 2.1.1 基于阈值的方法

阈值分割方法通常依据像素强度或颜色信息来进行分割。这些方法的优点在于简单、直

观，计算速度较快，但它们对于噪声和复杂背景的适应能力较差，且无法处理图像中多个结构或解剖部位的分割任务。例如，在 CT 图像中，肿瘤和正常组织的密度差异可能不显著，基于阈值的方法很难准确分割目标。

### 2.1.2 基于人工设计特征的机器学习方法

随着机器学习的发展，许多基于特征提取的传统方法也被广泛应用于医学图像分割。这些方法通常依赖于人工设计的特征（如边缘、形状、纹理等）来进行图像分割。支持向量机 (SVM) 和随机森林 (RF) 是常用的分类算法 [6]。尽管这些方法在处理简单任务时表现良好，但由于它们依赖于特征选择和设计，面对复杂的医学图像时，分割性能和泛化能力较差。

### 2.1.3 基于深度学习的方法

随着深度学习的兴起，特别是卷积神经网络 (CNN) 在图像领域的成功，基于深度学习的医学图像分割方法逐渐成为主流。FCN [7] 是首个应用于像素级图像分割的网络架构，之后 U-Net 架构被提出，并在医学图像分割中取得了显著的成功。U-Net 通过结合跳跃连接和下采样、上采样层，有效地提取和传递图像的局部和全局信息，在医学图像中的器官分割、肿瘤检测等任务中广泛应用 [8]。然而，这些方法通常依赖于大量标注数据，且对数据质量要求较高。

### 2.1.4 小结

尽管深度学习方法取得了巨大成功，传统模型和方法在某些低分辨率图像或小数据集上的表现仍有优势。传统方法通常计算效率较高，适用于资源受限的环境，但深度学习方法显然更适用于大规模医学数据集，尤其是当数据量足够时。

## 2.2 基础分割模型 (SAM)

基础分割模型 (Foundation Models) 是近年来在图像和语言处理领域兴起的一类模型，其通过大规模数据集的训练，能够进行 zero-shot 学习，具有强大的跨任务、跨领域的泛化能力。SAM) 是一种基础图像分割模型，它采用了自监督学习的方法，通过使用大规模的自然图像数据集（包含 1.1 亿个标注掩膜）进行训练，获得了强大的分割能力。SAM 采用了一种新的方法来解决图像分割中的复杂性，允许用户通过简单的交互方式（如点击、框选或遮掩）完成分割任务，从而实现高效且准确的分割。

SAM 的核心优势在于其对自然图像的强大处理能力。无论是物体分割还是场景分割，SAM 都能够通过少量的输入提示（如点击、框选等）生成准确的分割掩膜。此外，SAM 还支持跨多个任务和场景的 zero-shot 学习能力，这使得它能够应用于各种类型的自然图像分割任务，而无需专门的模型训练

尽管 SAM 在自然图像分割中表现优异，但其直接应用于医学图像分割时表现不尽如人意。医学图像与自然图像之间存在显著差异，例如，医学图像往往有较低的分辨率、复杂的结构、噪声干扰等特点。由于 SAM 的训练数据主要来自自然图像，这使得 SAM 模型在处理医学图像时，尤其是在低分辨率、细节复杂的 3D 图像和视频分割中，表现较差。此外，医学

图像的解剖结构和生物标志物的多样性，要求分割模型具有更强的领域适应性，这也是 SAM 需要改进的地方。

### 2.3 MedSAM 与 SAM2 的进展

为了弥补 SAM 在医学图像分割中的局限性，MedSAM 应运而生。MedSAM 通过迁移学习对 SAM 进行微调，使用医学图像数据集（如 CT、MRI 等）进一步优化其在医学领域的表现。MedSAM 在 2D 医学图像分割中表现出色，尤其在处理细粒度的解剖结构时，能够实现较高的分割精度。与传统的专用医学图像分割模型相比，MedSAM 的优势在于其较强的跨模态适应性和更少的对标注数据的依赖。然而，MedSAM 仍然存在一些限制，尤其是在 3D 医学图像和视频分割中的表现。许多医学图像（如 CT 扫描、MRI 等）具有复杂的三维结构，而现有的 MedSAM 只能处理 2D 图像，导致 3D 图像分割性能不足。此外，尽管 MedSAM 在处理静态医学图像时表现良好，但面对动态医学视频（如超声、内窥镜等）的挑战时，MedSAM 的分割精度和效率较低。

SAM2 是 SAM 模型的升级版，它在继承 SAM 和 MedSAM 的优势的基础上，引入了视频分割功能，特别是通过改进的架构和训练方法，提升了模型在医学图像和视频中的表现。SAM2 的引入使得视频分割成为可能，解决了医学视频分割中帧间目标追踪和分割的难题。SAM2 通过引入时间上下文的内存注意力模块，使得每一帧的分割结果可以利用前一帧的信息进行优化，从而提升了视频分割的精度。

相比于 SAM 和 MedSAM，SAM2 在处理 3D 医学图像和视频时表现出了更强的能力。尤其是在通过视频分割的方式处理 3D 医学图像时，SAM2 能够有效地利用视频序列中的时序信息进行掩膜传播，显著提高了分割效率和精度；与 MedSAM 类似，SAM2 也通过迁移学习进一步优化了其在医学领域的适应性。SAM2 能够通过微调适应不同的医学图像任务，从而提高分割精度。通过这种方法，SAM2 不仅能处理 2D 图像分割任务，还能在 3D 医学图像和视频分割中发挥作用，推动医学图像分割技术的跨模态发展。

## 3 本文方法

### 3.1 本文方法概述

如图 1 是 SAM2 的整体输入输出架构图。其中 a 部分是输入 + 提示信息，b 部分是整体架构，c 部分是相关的数据集。SAM2 的架构基于 Transformer 模型的强大能力，特别是引入了 ViT 和 Hiera 架构来增强图像分割能力。以下是 SAM2 架构的详细说明：



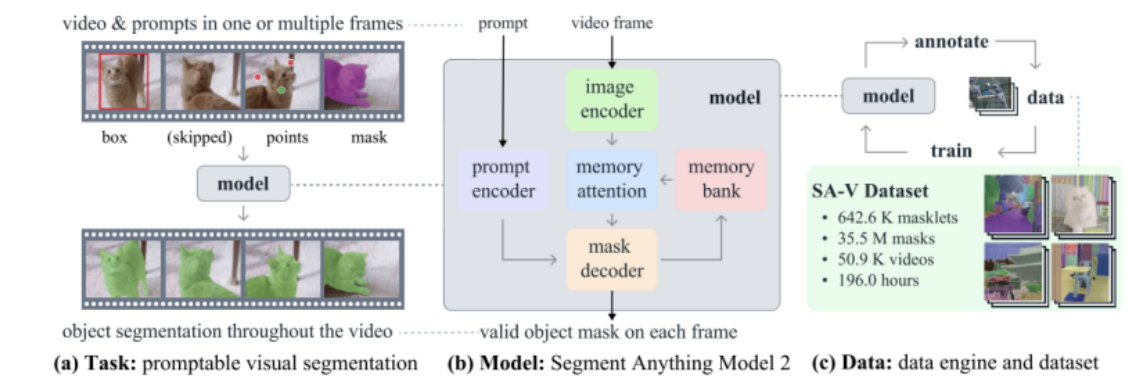


图 1. SAM2 整体输入输出架构图

### 3.1.1 ViT

ViT 最初被提出用于图像分类任务，通过将图像分割成多个小块，然后将每个图像块线性嵌入为特征向量，通过自注意力机制建模图像块之间的关系。ViT 的引入使得 SAM2 能够在多个图像模态中处理复杂的图像特征，从而提高分割精度。相较于传统的 CNN，ViT 具有更强的建模全局信息的能力，能够捕捉更长范围的上下文信息，对于复杂图像的分割尤为有效。

### 3.1.2 Hiera

为了进一步提升分割的细粒度精度，SAM2 采用了 Hiera 架构。如图 2 所示是其架构图。

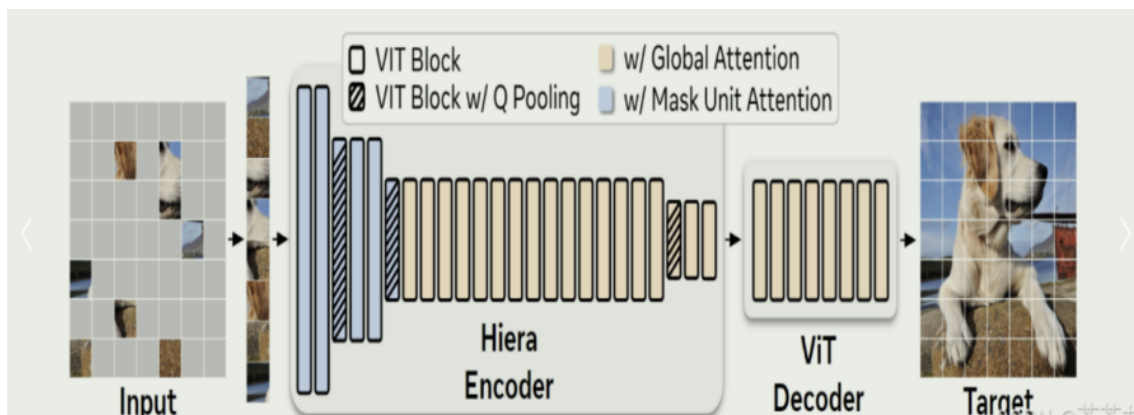


图 2. Hiera 架构图

Hiera 是基于多尺度特征提取的 Transformer 架构，通过多层级的特征聚合，增强了模型对细节信息的捕捉能力。具体来说，Hiera 在图像分割任务中起到了增强分辨率和细节分割的作用，特别适用于医学图像中具有复杂细节的区域（例如肿瘤或器官边界）。

### 3.1.3 内存注意力模块

SAM2 在处理视频分割时，采用了时间上下文的内存注意力模块。该模块通过将每一帧的分割信息与之前帧的特征和预测进行结合，从而增强视频序列中的时序信息理解。这个模

块通过“记忆”之前的分割信息，使得视频中的目标分割可以跨帧保持一致，减少帧间分割的不稳定性。

#### 3.1.4 训练细节

SAM2 采用了 AdamW 优化器，使用学习率调度和早停策略来控制训练过程。在训练过程中，使用了数据增强技术（如随机旋转、缩放和翻转）来增加模型的泛化能力。训练时，将每个医学图像和视频帧调整到标准化尺寸，并进行批量处理，以提高训练效率。

### 3.2 视频与 3D 图像分割

SAM2 在视频和 3D 医学图像分割中表现出色，尤其在处理具有时序信息的医学视频和复杂三维结构的图像时，展现了其强大的分割能力。

#### 3.2.1 3D 医学图像分割

SAM2 利用视频分割技术将 3D 图像分割转化为帧间传播的任务。在处理 3D 医学图像（如 CT、MRI 等）时，SAM2 首先在中间切片上进行分割，然后将该切片的分割掩膜传播到其余的切片。这种方法允许 SAM2 有效地处理 3D 图像的复杂结构，并通过从中间切片获取最大的物体区域，避免了传统逐片分割带来的不一致性和性能损失。

#### 3.2.2 视频分割

视频分割的主要挑战在于如何处理图像中的动态目标。SAM2 的内存注意力模块通过将每一帧的分割结果与上一帧的特征信息结合，确保了视频序列中目标分割的一致性。视频分割不仅能够利用每一帧的图像信息，还能通过时序信息提升分割结果的稳定性和准确性。

### 3.3 迁移学习与微调协议

为了提升 SAM2 在医学图像分割中的适应性，通过迁移学习对其进行微调，尤其是在 3D 医学图像任务中的微调过程。

#### 3.3.1 迁移学习的目标

SAM2 在初始阶段使用了大规模的自然图像和视频数据进行训练，然而，医学图像和视频具有与自然图像不同的特点。因此，通过迁移学习，可以将 SAM2 从自然图像领域迁移到医学图像领域，使其更好地适应医学图像的复杂结构和特征。

#### 3.3.2 微调过程

在迁移学习过程中，冻结了 SAM2 的部分网络层，特别是那些与图像特征提取和图像理解相关的部分（如 Vision Transformer 的编码层），只对图像解码器进行了训练，以适应医学图像的特定任务和数据集。

训练数据集：使用了 MICCAI FLARE22 腹部 CT 数据集进行训练，该数据集包含了大量的 3D CT 图像，覆盖了不同的解剖结构（如肝脏、脾脏、肾脏等）。

微调设置：使用学习率为  $6e-5$  的 AdamW 优化器，批量大小为 16，训练 1000 个 epoch，采用早停机制来避免过拟合。在训练过程中，使用了 Dice 损失和交叉熵损失的加权和，这种复合损失函数已经在许多医学图像分割任务中证明了其稳定性。

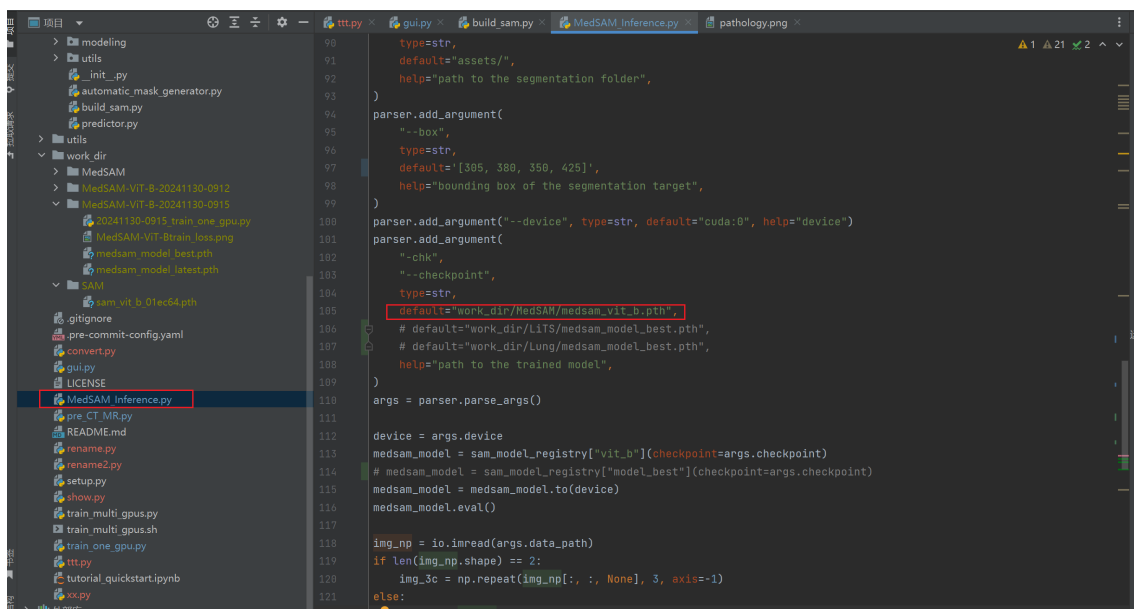
## 4 复现细节

### 4.1 与已有开源代码对比

参考代码链接 [MedSAM](#)，主要应用及修改如下：

#### 4.1.1 利用官方权重推理图像

如图 3 所示，利用官方给的权重 `medsam_vit_b.pth` 进行图像推理，推理结果如图 4 所示。



```
90     type=str,
91     default="assets/",
92     help="path to the segmentation folder",
93 )
94 parser.add_argument(
95     "--box",
96     type=str,
97     default="[305, 380, 350, 425]",
98     help="bounding box of the segmentation target",
99 )
100 parser.add_argument("--device", type=str, default="cuda:0", help="device")
101 parser.add_argument(
102     "--chk",
103     type=str,
104     default="work_dir/MedSAM/medsam_vit_b.pth",
105     # default="work_dir/LITS/medsam_model_best.pth",
106     # default="work_dir/Lung/medsam_model_best.pth",
107     help="path to the trained model",
108 )
109 args = parser.parse_args()
110
111 device = args.device
112 medsam_model = sam_model_registry["vit_b"](checkpoint=args.checkpoint)
113 # medsam_model = sam_model_registry["model_best"](checkpoint=args.checkpoint)
114 medsam_model = medsam_model.to(device)
115 medsam_model.eval()
116
117 img_np = io.imread(args.data_path)
118 if len(img_np.shape) == 2:
119     img_3c = np.repeat(img_np[:, :, None], 3, axis=-1)
120 else:
121     pass
```

图 3. 推理代码截图

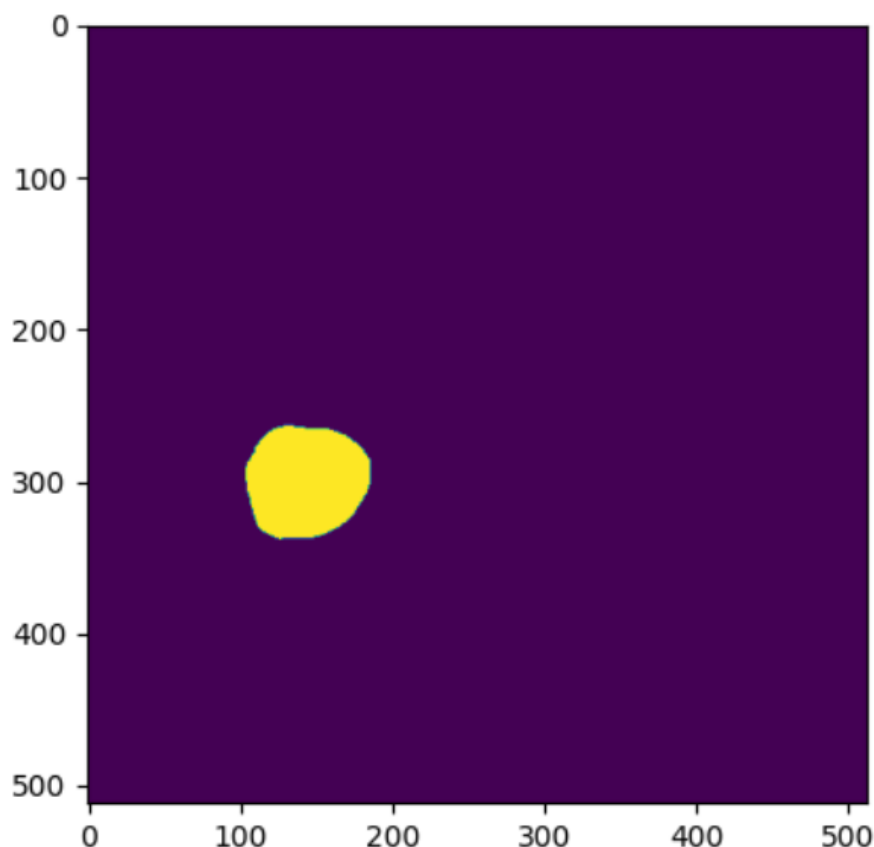


图 4. 推理结果

#### 4.1.2 数据集选择

主要选择了三个数据集：

第一个是官方提供的 FLARE22 数据集，一共包含 2300 张来自超过 20 个中心的三维腹部 CT 影像数据，其中提供 50 例有标注数据和 2000 例无标注数据用于训练模型，50 例用于验证，以及 200 例用于最终的测试和排名。后续的实验选择了其中的 50 例有标注数据进行训练。

第二个是 LiTS (The Liver Tumor Segmentation Benchmark)，它是专注于肝脏及其肿瘤分割的 CT 数据集。该数据集收集了 7 个不同医学中心的数据，包含 131 例训练集和 70 例测试集，其中测试数据标签不公开。

第三个是 MSD 肺癌数据集，它是 MSD (Medical Segmentation Decathlon, 医学图像分割十项全能挑战赛) 中的 Task06 即第 6 个子任务，目标是从 CT 图像中分割出肺部肿瘤，MSD 选择该数据集的原因是“在大的背景中分割出小目标”。该数据集包含 96 例非小细胞肺癌患者的薄层 CT 扫描，官方划分为 64 例训练集和 32 例测试集，其中测试集可以通过官网提交分割结果进行测试。



### 4.1.3 数据预处理

主要完成下面几个作用，划分数据集、窗口宽度和高度的灰度增强、归一化、resize 以及保存到 npy 的 2d 格式，参考代码文件是 pre\_CT\_MR.py。如图 5 所示。

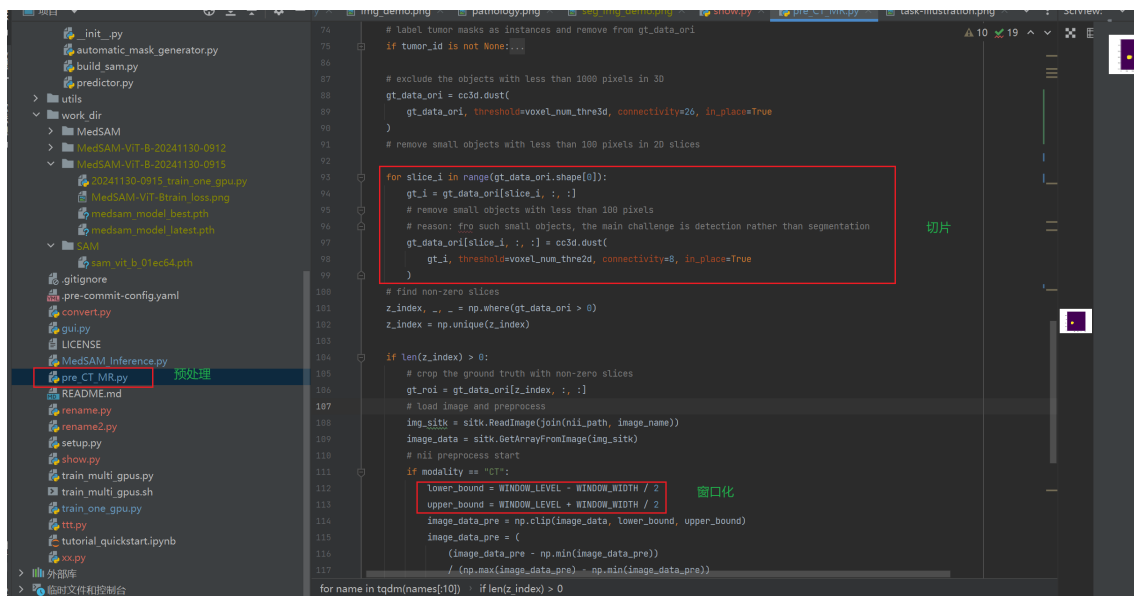


图 5. 预处理截图

### 4.1.4 单 GPU 训练

主要完成在单个 GPU 下进行模型的训练，参考代码文件是 train\_one\_gpu.py。如图 6 所示。

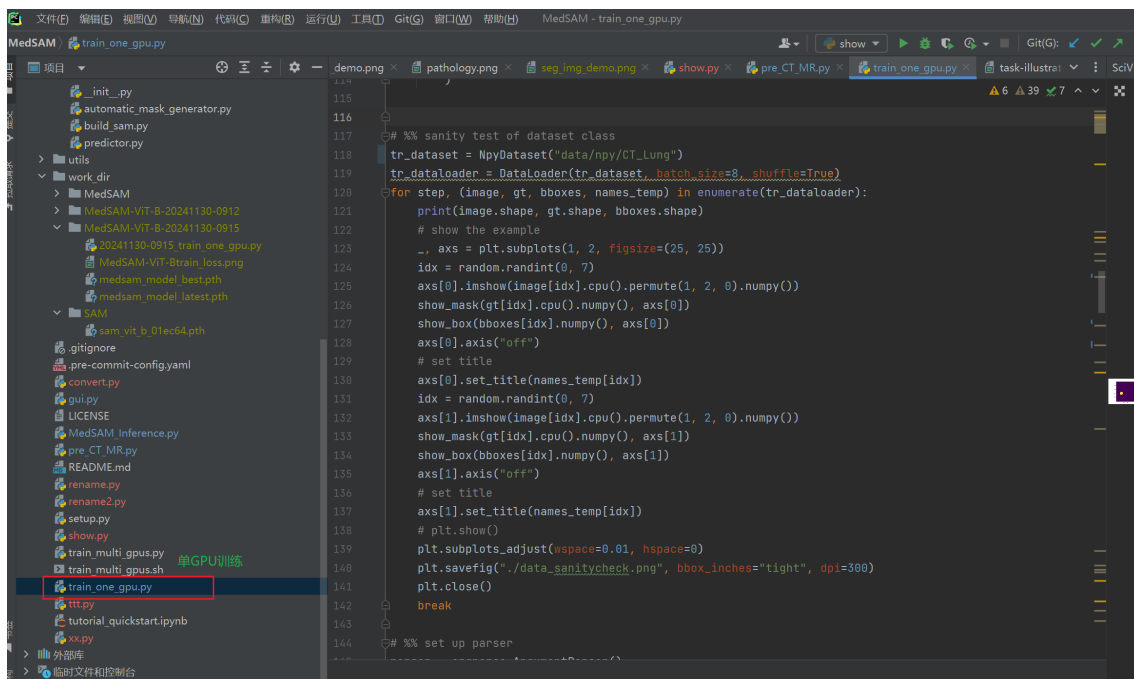


图 6. 训练代码截图

## 4.2 实验环境搭建

1. 包管理器选择 anaconda，虚拟环境配置 python 的版本官方是 3.10，命令如下：

```
1 conda create -n medsam python=3.10 -y
```

2. 激活虚拟环境后 (conda activate medsam)，安装 gpu 版本的 torch，官方的建议 torch 版本需要高于 2.0，应该会用到 2.x 版本的库文件

3. cd 到 MedSAM 目录后，运行下面代码即可配置好基本的环境

```
1 pip install -e .
```

## 4.3 界面分析与使用说明

如图 7 所示，主要是利用训练好的权重去对给定的医学图像进行推理，选择好对应的训练好的模型输入对应的图像即可进行分割，图中的例子是 MSD 肺癌数据集。

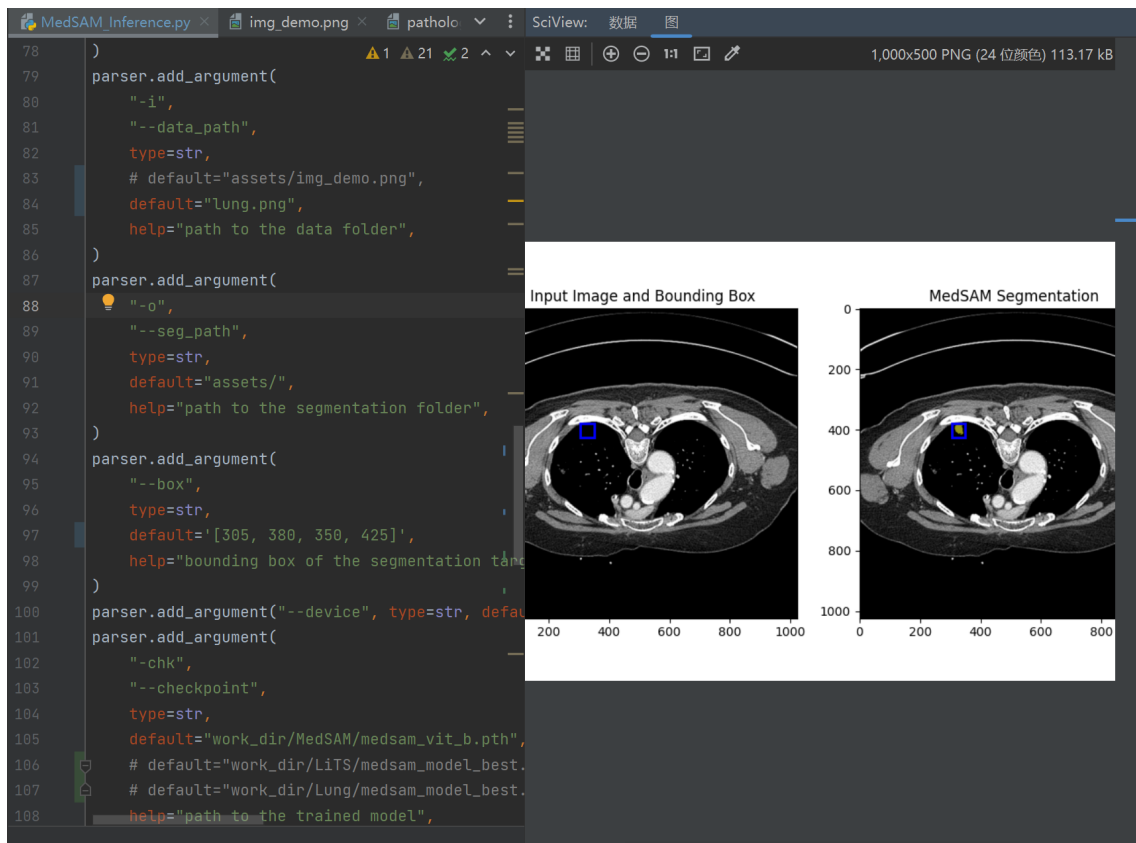


图 7. 界面使用图

## 4.4 创新点

### 4.4.1 引入视频分割能力

SAM2 将视频分割能力引入到医学图像分割中，能够处理动态医学数据（如超声和内窥镜视频）。通过利用视频中的时序信息，SAM2 能够在不同帧之间保持一致性，避免了传统静态图像分割方法在处理动态数据时的局限性。特别是，SAM2 采用了内存注意力模块，使得模型可以跨帧传播分割掩膜，确保视频序列中的目标分割保持一致性和精度。

#### 4.4.2 创新的 3D 图像分割方法

SAM2 通过将 3D 医学图像（如 CT、MRI）处理为视频序列的方式进行分割，显著提高了分割效率和精度。这种方法利用了视频分割的时序信息来处理每个切片，并将中间切片的分割掩膜传播到其他切片。相较于传统的逐片分割，SAM2 能够更好地处理 3D 医学图像的复杂性，减少分割不一致性，提高精度。

#### 4.4.3 迁移学习与微调能力

SAM2 采用了迁移学习技术，通过微调模型来增强其在医学图像上的适应性。特别是在 3D 医学图像任务中，SAM2 通过迁移学习能够快速适应不同数据集，优化模型在医学图像分割任务中的表现。通过这种方法，SAM2 在处理特定医学任务时，能够减少训练时间，同时提高在特定领域（如肝脏、肾脏等器官分割）的分割精度。

#### 4.4.4 高效的临床部署工具

SAM2 被开发为 3D Slicer 插件和 Gradio API，这使得其在临床环境中的应用更加高效和便捷。通过这两种部署工具，医学专业人员可以在不需要编程知识的情况下，使用 SAM2 进行医学图像和视频的分割。这种部署方式不仅简化了操作流程，还提高了模型在临床应用中的可访问性和实用性。

## 5 实验结果分析

Dice 相似系数 (DSC) 用于衡量预测分割与真实分割的重叠程度，DSC 值越高，表示分割越准确；归一化表面距离 (NSD) 用于评估边界匹配的精度，衡量分割掩膜边界与真实边界之间的距离。

FLARE22：如图 8 所示，是官方的预训练权重与从头训练 100 轮 FLARE22 的推理结果对比。图 9 是训练的 loss 变化，在训练轮数不足的情况下，loss 值也还没有达到饱和。可以看到，官方的结果显然效果更好，这主要是因为我的实验条件有限，训练轮数不太够，只有 100 轮，官方的推荐轮数是 1000 轮起。

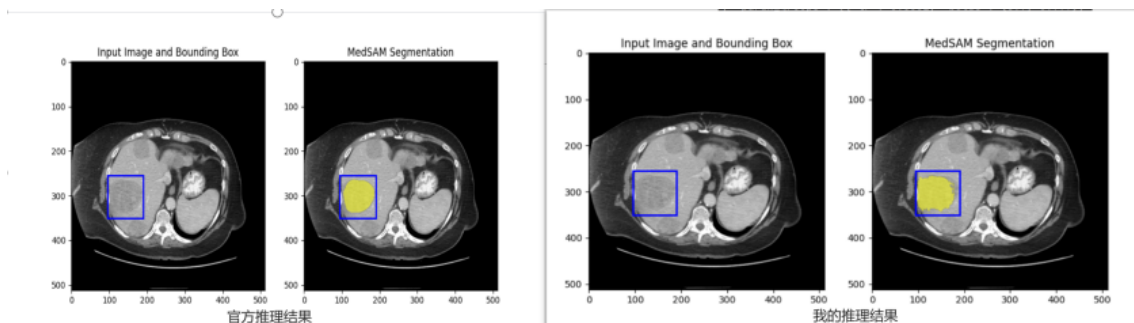


图 8. 官方与我的 FLARE22 推理结果对比

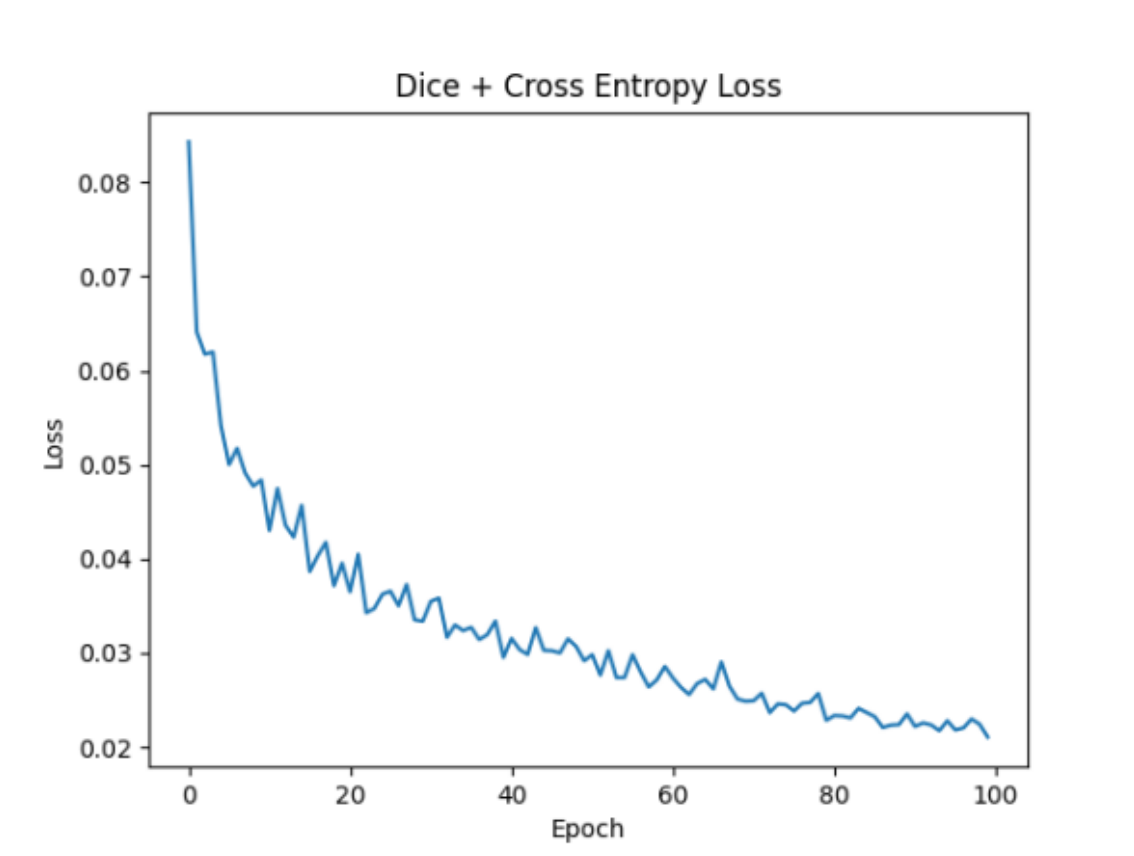


图 9. FLARE22 的 loss 变化

LiTS: 图 10和图 11分别是 LiTS 数据集的推理结果与 loss 变化。我这里 LiTS 数据集和 FLARE22 数据集均没有计算 DSC 和 NSD 指标，主要由于他们数据量庞大，实验条件所限，训练轮数不够，指标的计算意义不大。

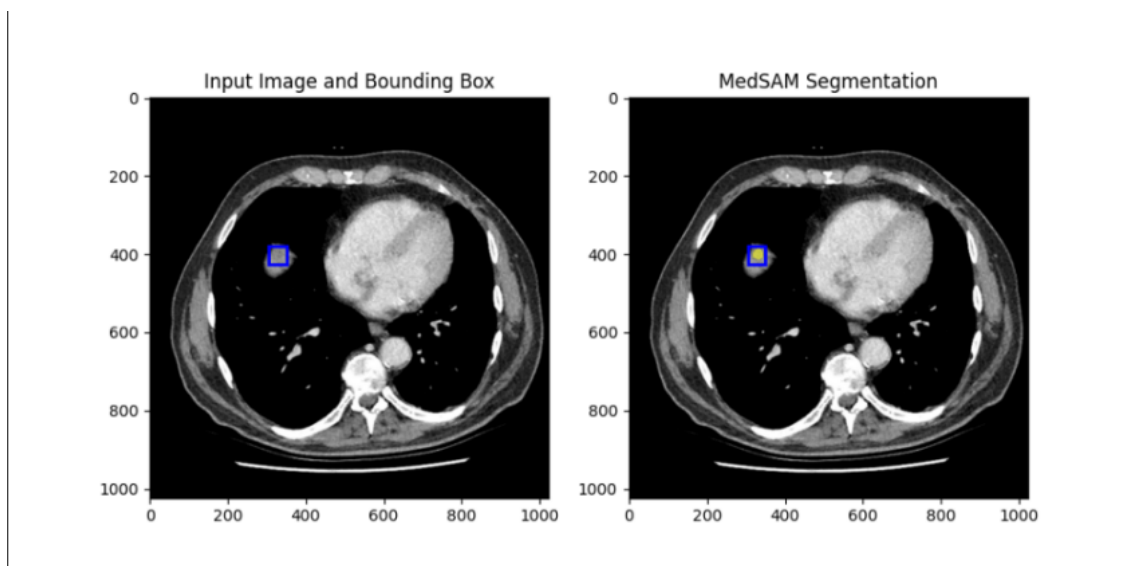


图 10. LiTS 推理结果

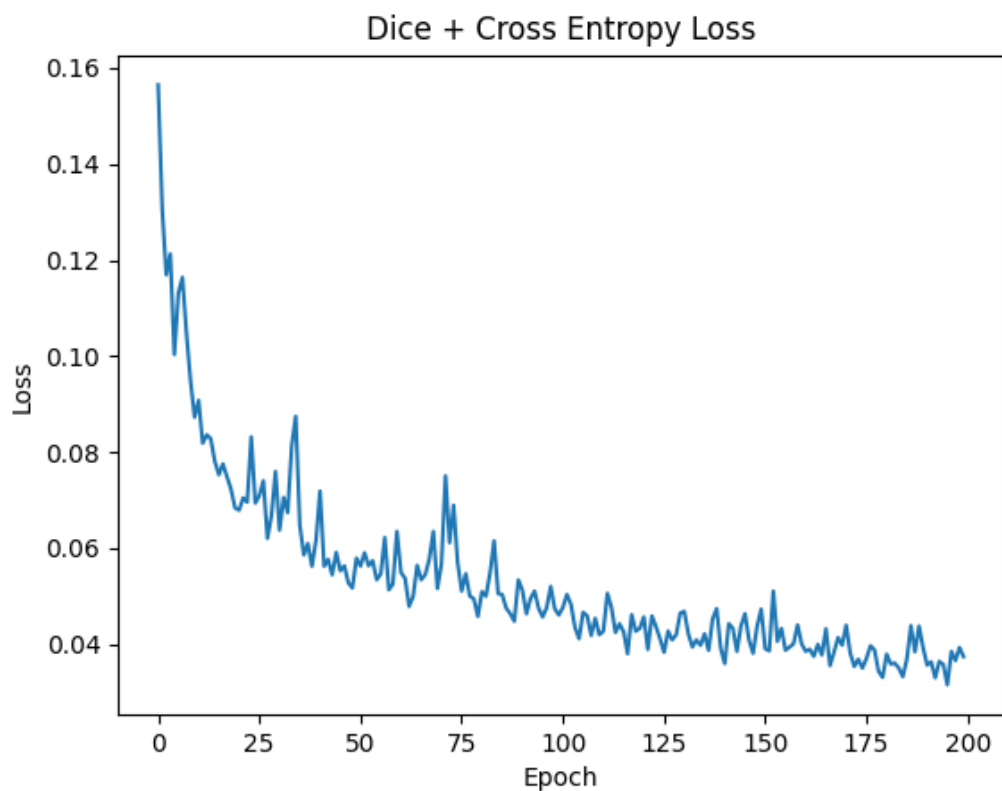


图 11. LiTS 的 loss 变化

MSD 肺癌数据集：图 12和图 13分别是 MSD 肺癌数据集的推理结果与 loss 变化。可以看到，当训练轮数达到 1000 轮时，loss 趋向于饱和，分割效果也很不错，能够有效找到病变部位。表 1是其 DSC 和 NSD 指标，分别达到了 90 多和 60 多，进一步证明了 SAM2 模型分割医学图像的有效性。

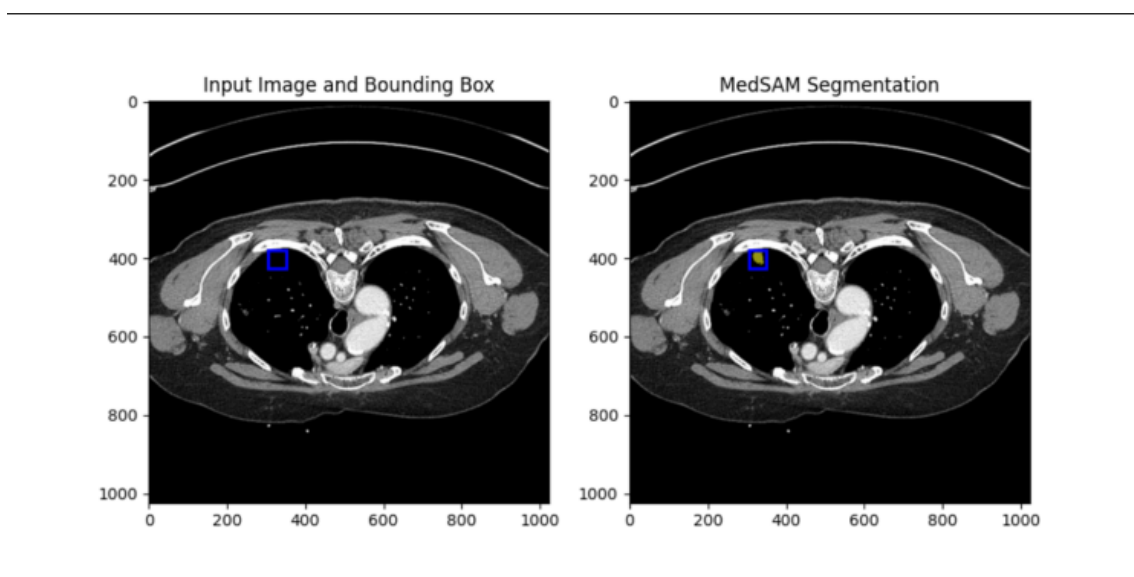


图 12. MSD 肺癌数据集推理结果



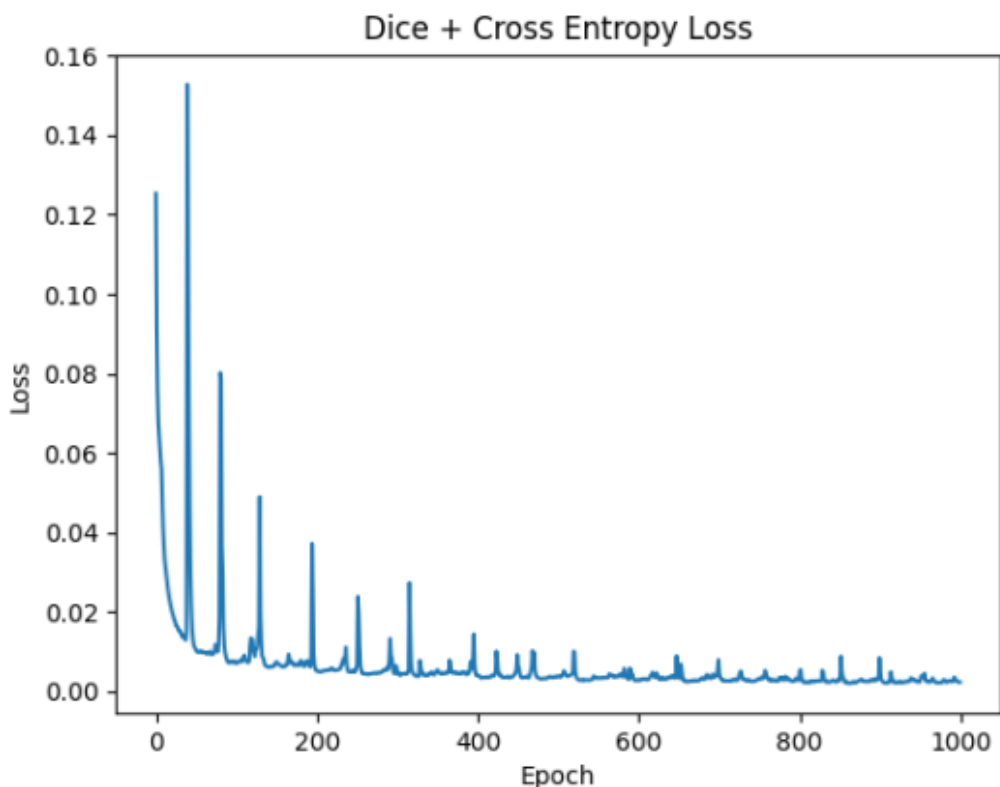


图 13. MSD 肺癌数据集的 loss 变化

表 1. MSD 数据集分割性能评估

MSD 肺癌数据集	DSC	NSD
	90.0568	63.2715

## 6 总结与展望

### 6.1 研究总结

本研究对 SAM2 在医学图像分割任务中的应用进行了深入评估，特别是在 2D、3D 图像和医学视频分割中的表现。主要研究发现如下：

#### 6.1.1 SAM2 在 2D 医学图像分割中的表现

SAM2 在多种医学图像模态（包括 CT、MRI、超声、内窥镜等）上显示出了强大的分割能力。在与 SAM1 和 MedSAM 的对比中，SAM2 在大多数模态中表现优越，尤其是在处理高分辨率图像（如 MRI 和皮肤镜图像）时，SAM2 的分割精度较高。

### 6.1.2 SAM2 在 3D 医学图像分割中的优势

通过采用视频分割技术，SAM2 能够处理 3D 医学图像的复杂结构。在 CT 和 MRI 图像的 3D 分割任务中，SAM2 取得了显著的性能提升，相比 SAM1 和 MedSAM，在 DSC 得分上表现更为出色。特别是在 CT 图像中的分割精度提升显著，展示了其在 3D 医学图像中的应用潜力。

### 6.1.3 SAM2 在医学视频分割中的突破

SAM2 在超声和内窥镜视频的分割任务中展现了优异的性能。通过内存注意力模块的引入，SAM2 能够有效处理视频帧间的时序信息，实现了跨帧的一致性分割。尽管在一些复杂场景（如物体边界不明确或图像对比度低时）中仍存在失败案例，SAM2 在医学视频分割任务中的表现已达到一个新的高度。

### 6.1.4 迁移学习的有效性

通过迁移学习，SAM2 在 3D 医学图像分割任务中的表现得到了显著提升。在特定数据集（如腹部 CT 数据集）的微调过程中，SAM2 在各类器官（如肝脏、脾脏、肾脏等）的分割任务中取得了显著的 DSC 和 NSD 得分，提高了模型在医学图像任务中的适应性。

### 6.1.5 小结

总体而言，SAM2 展示了作为一个基础模型在医学图像分割领域的巨大潜力，尤其在处理复杂的 3D 医学图像和动态视频时，SAM2 的创新性方法和优越性能提供了更高效、精确的解决方案。

## 6.2 贡献与展望

本研究的贡献主要体现在以下几个方面：

### 6.2.1 模型微调管道的构建

提出了一种针对 SAM2 的迁移学习微调管道，显著提升了其在医学图像中的适应性，尤其是针对 3D 医学图像的分割任务。通过对 SAM2 进行精细微调，结合医学图像特定领域的特征（如噪声和低分辨率等问题），我们增强了其在医学图像领域的表现。

### 6.2.2 SAM2 的部署与实际应用

为了促进 SAM2 在临床环境中的应用，开发了两种部署方式：3D Slicer 插件和 Gradio API。3D Slicer 插件为医学专业人员提供了一个无编码的分割工具，可以方便地在医学图像中进行 3D 分割和标注；而 Gradio API 为医学视频分割提供了一个灵活的界面，用户可以上传视频并进行标注，系统会自动处理整个视频序列。这些部署接口使得 SAM2 能够更容易地被医学专业人员应用于实际临床工作中，缩短了从研究到应用的转化时间。

### 6.2.3 视频与 3D 图像分割的新思路

通过引入内存注意力模块，SAM2 能够利用视频中的时序信息进行帧间一致性分割，这对动态医学数据（如超声和内窥镜视频）的处理具有革命性意义。此外，SAM2 采用了基于视频分割的 3D 图像处理方法，突破了传统 3D 医学图像分割的瓶颈，为处理更复杂的三维医学数据提供了新的解决方案。

### 6.2.4 临床应用潜力

本研究展示了 SAM2 在医学图像分割中的广泛适用性，尤其是在多模态数据的处理、复杂三维结构的分割以及动态视频序列的分割等方面，为未来的医学图像分析系统提供了有力的支持。通过将 SAM2 集成到临床软件（如 3D Slicer 和 Gradio API），不仅简化了使用流程，还为临床医生提供了一个更高效、准确的工具，推动了医学图像分析技术的实际应用。

## 6.3 未来的研究方向

尽管本研究展示了 SAM2 在医学图像分割中的巨大潜力，但仍有几个方面值得进一步探索：

### 6.3.1 增强对复杂医疗图像的鲁棒性

尽管 SAM2 在大多数医学图像任务中表现优越，但在处理一些低分辨率或高噪声的图像时，其性能仍然受到一定限制。未来研究可以通过改进图像预处理方法、增强模型的去噪能力以及引入更强大的数据增强技术来提升模型的鲁棒性。

### 6.3.2 多模态学习与跨模态适应性

医学图像通常来自多种不同的成像模态（如 CT、MRI、PET 等），这些图像在信息表示上存在显著差异。未来的研究可以进一步探索如何通过多模态学习技术使 SAM2 能够更好地跨模态适应，进而提高在不同模态下的分割精度。

### 6.3.3 文本提示与自然语言处理的结合

目前 SAM2 支持点、框和掩膜提示，而文本提示（如使用自然语言进行交互）则为更复杂结构的分割提供了极大的灵活性。结合自然语言处理技术，未来的 SAM2 版本可以进一步扩展其在医学领域的应用，提升分割过程的交互性和智能化。

### 6.3.4 实时视频分割与临床应用

随着医学视频数据的日益增加，实时视频分割将成为一个重要的研究方向。未来的工作可以通过优化 SAM2 的推理效率，减少其计算资源需求，使其能够在更广泛的临床环境中实现实时应用。

### 6.3.5 小样本学习与自监督学习

由于医学图像数据的标注成本高且稀缺，小样本学习和自监督学习将成为未来的重要研究方向。通过这些技术，可以在标注数据较少的情况下有效地训练分割模型，从而提高 SAM2 在少标注数据环境下的表现。

## 参考文献

- [1] Sanuwani Dayarathna, Kh Tohidul Islam, Sergio Uribe, Guang Yang, Munawar Hayat, and Zhaolin Chen. Deep learning based synthesis of mri, ct and pet: Review and analysis. *Medical image analysis*, 92:103046, 2024.
- [2] DR Sarvamangala and Raghavendra V Kulkarni. Convolutional neural networks in medical image understanding: a survey. *Evolutionary intelligence*, 15(1):1–22, 2022.
- [3] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.
- [4] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.
- [5] Jun Ma, Sumin Kim, Feifei Li, Mohammed Baharoon, Reza Asakereh, Hongwei Lyu, and Bo Wang. Segment anything in medical images and videos: Benchmark and deployment. *arXiv preprint arXiv:2408.03322*, 2024.
- [6] Haifeng Wang and Sang Won Yoon. A machine learning model for medical image recognition using texture-based features. In *IIE Annual Conference. Proceedings*, pages 1655–1660. Institute of Industrial and Systems Engineers (IISE), 2017.
- [7] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.