

# Segment Anything

## 摘要

**SAM (Segment Anything Model)** 构建了迄今为止最大规模的自然图像数据集SA-1B, 并在计算机视觉的不同子领域有着极强的零样本泛化能力。作为主要用于自然图像分割的视觉大模型, **SAM**在医学图像上表现并不如意。基于医学图像存在低对比度、大量噪声、不规则的掩码形状等问题, 将**SAM**直接应用于医学图像领域将得到灾难性的分割结果, 甚至不如一些基础的医学SOTA框架。而**SAM**强大的预训练权重又可以为医学图像分割带来更强的泛化能力, 因此, 我在已有的对**SAM**模型的改进的基础上添加新一代适配器, 使**SAM**模型更加关注局部信息, 从而大大提高**SAM**模型在医学图像分割领域的性能。

**关键词:** SAM大模型; 医学图像语义分割; 超声心动图

## 1 引言

超声心动图是视频语义分割的难点之一, 它具有超声图像特有的低对比度、噪声大等特点, 同时又有视频分割的时间维度的信息。近年来, 基于卷积神经网络的超声心动图分割有了极大的进步, 许多半监督模型被应用在超声心动图上并已经取得了与医生标注接近的性能。但目前主流的三个超声心动图数据集: CAMUS、EchoNet-Dynamic以及HMC-QU数据集中, 除了CAMUS数据集在今年获得了医生的全标注数据, 其余两个数据集依旧只有舒张末期以及收缩末期的两帧数据有标注数据, 因此, 主流方法都是依照半监督思路来进行分割。诚然, 数据集的天然性质适配于半监督模型, 但在其中一个数据集公开全部标注后, 使用一种通用模型进行分割, 以扩展到其他两个数据集上成为新一种研究思路。

在大模型愈发火热的时代, SAM的出现为计算机视觉领域, 特别是医学图像分割领域提供了新的模式。SAM所拥有的大量预训练权重在经过适当调整后可以在医学数据集上获得超越SOTA框架的性能。但是依旧存在一个问题, 基于SAM改进后的模型SAMUS, 并未考虑到超声心动图中是时间信息。而许多主流方法如Painchaud et al (Painchaud et al 2022)<sup>[1]</sup>提出的一种增强时间一致性的后处理方法, 该方法基于现有模型的分割结果, 通过正则化形状编码来纠正时间不一致性。或者是Wei等人引入的CLAS<sup>[2]</sup>, 这是一种新颖的3D分割网络, 它在仅依赖ED和ES注释的情况下努力保持时间一致性。为了实现这一点, CLAS预测变形场, 并在训练期间使用它来传播注释。由此我们可以看出, 时间信息在超声心动图中占有重要的地位。受到Uniformer<sup>[3]</sup>的启发, 我们提出了一种基于SAMUS的时间模块, 使其能够捕获数据中的时间信息并维持其时间一致性。我们的主要贡献可以概括如下:

(1) 我们基于SAMUS框架提出了TRA模块, 主要用于对时间维度上数据的处理, 以捕获数据帧中的时间信息并有效地在学习好的空间表示上建模时间依赖性。

(2) 我们根据数据集的特点，增加了局部信息注意模块LIA，用于增强模型学习特征信息的表达能力。

(3) 我们在CAMUS公开数据集上进行了实验，结果表明我们的方法略优于现有的SOTA方法。

## 2 相关工作

### 2.1 超声心动图

超声心动图是医生诊断心脏疾病的重要工具，同时计算机辅助成像技术已经可以帮助医生进行疾病诊断。因此，超声心动图分割成为医学图像与计算机视觉交叉领域中的重要问题之一。超声图像与自然图像相比，普遍存在低对比度、大量噪声以及非均匀性等问题。而心脏是一个动态的器官，因此超声心动图作为视频序列，可能包含运动模糊，尤其是在心脏心室收缩与扩张时，这也给模型分割带来了难度。现有基于超声心动图的研究方法中，大部分都是基于半监督的框架，这也是由于目前公开的数据集仅标注了收缩末期及扩张末期的两帧所导致的。但今年CAMUS数据集公开了超声心动图序列的全标注数据，这使得利用大模型进行全监督训练成为可能。

### 2.2 大模型在医学图像分割中的应用

随着大模型不断在自然语言处理领域取得突破，计算机视觉也基于transformer结构开发出了视觉大模型。最新的发展之一是SAM (Kirillov et al., 2023)<sup>[4]</sup>，一种用于通用图像分割的视觉基础模型。通过对10亿个掩码进行预训练，SAM在众多图像分割任务中展示了令人印象深刻的零标签泛化能力。同时，SegGPT (Wang et al., 2023c)和SEEM (Zou et al., 2023)也用于一般图像分割，但与SAM相比，它们是在相对较小的数据集上进行预训练的。相较于传统医学图像分割任务所使用的nnU-Net等模型，大模型所蕴含的丰富的预训练权重可以为不同任务的医学分割带来更大的提升。而SAM模型是在自然图像数据机上加以训练，所以将其直接应用在医学图像处理领域势必会存在性能下降等情况。因此，如何改进大模型，使其适用于医学数据集成为大模型在医学图像分割领域的应用的重点。目前主流方法分为三种：基于适配器、基于卷积与ViT的统一以及与SOTA框架相结合。MA-SAM<sup>[5]</sup>、MSA等模型通过为SAM模型添加适配器、调整掩码解码器等方式，为模型引入关于特定数据集的信息，使其适应医学图像分割任务。SAMUS<sup>[6]</sup>通过为SAM模型添加卷积层，使其关注于局部信息，丰富模型对不同病理数据集的感知力。而nnSAM则是通过将SAM模型的图像编码器与nn-Unet模型相结合，利用SAM的预训练权重辅助nn-Unet进行预测。

## 3 本文方法

### 3.1 本文方法概述

本文基于SAMUS框架，为CAMUS数据集设计了两个辅助模型预测的模块。鉴于SAMUS模型并未将CAMUS数据集视为视频进行分割，因而忽略了数据中丰富的时间信息，因此在卷积模块中加入时间模块，更好的对时间进行建模。同时基于2D图像层面加入增强模块，使模型能够更好地分割出结果。模型整体概述如图1所示。

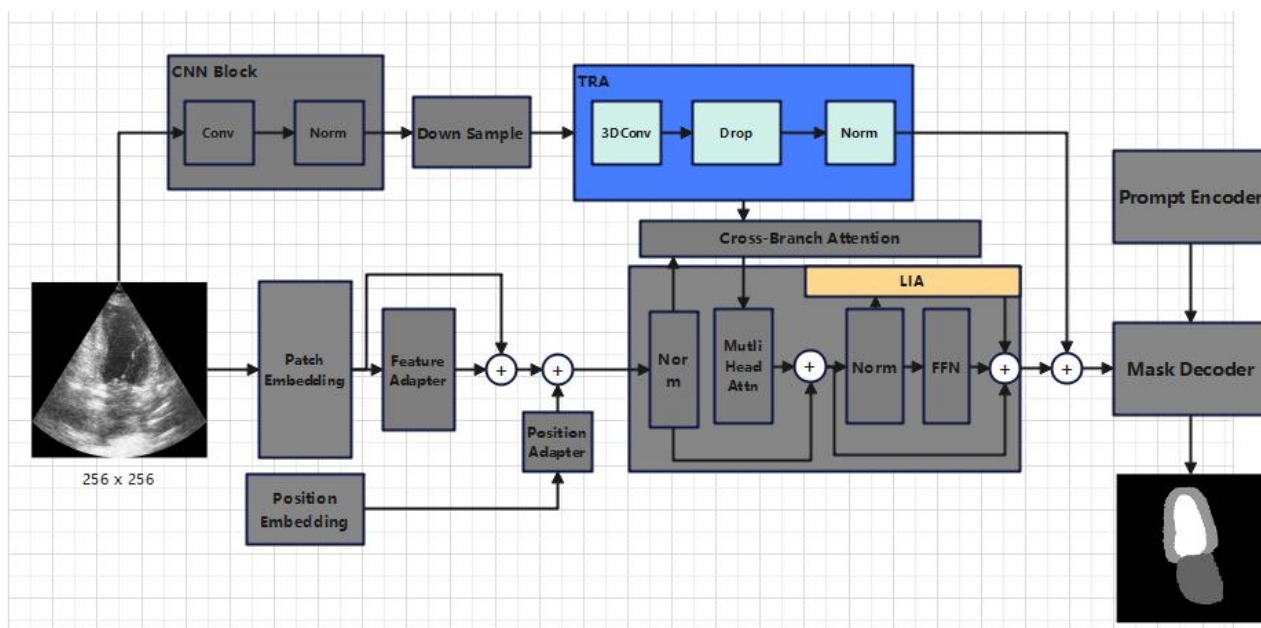


图1 模型整体概述

### 3.2 TRA模块

超声心动图作为视频序列，其中蕴含了丰富的时间信息，因此对时间建模显得尤为重要。为了有效地在超声心动图上建模时间依赖性，本文基于Uniformer中将卷积与注意力机制统一的思想，提出了TRA模块。通过卷积操作获取时间信息并送入注意力机制中进行进一步训练。在设计TRA模块时，本文对比了时间适配器，时间注意力以及TRA模块的性能，发现TRA模块在性能上优于前两种方法。

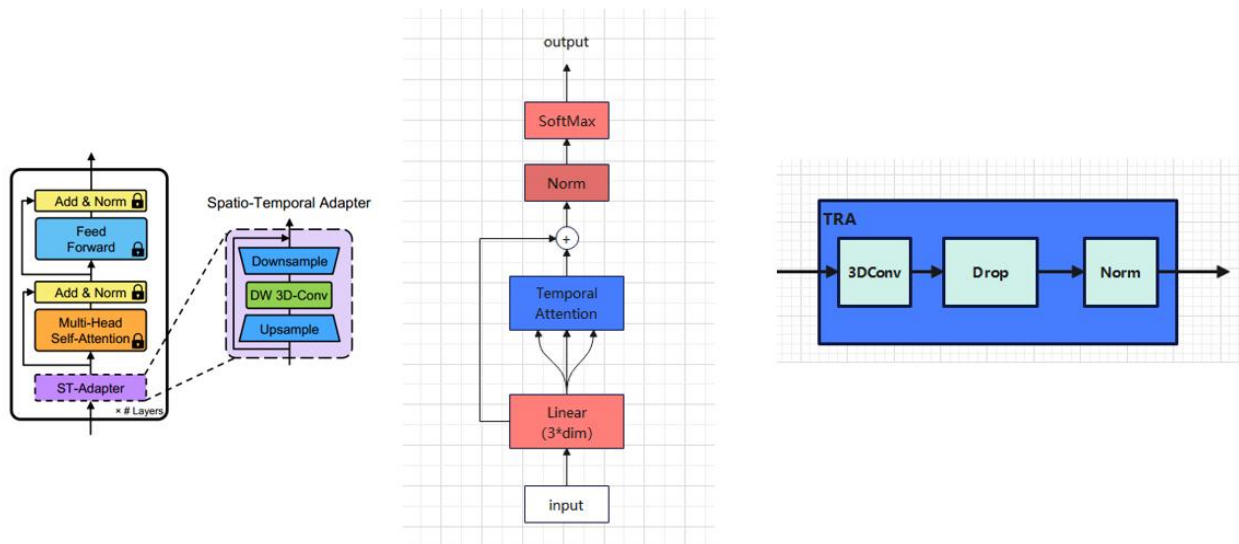


图2 三种时间模块，从左至右依次为时间适配器、时间注意力以及TRA模块

### 3.3 LIA模块

受到坐标注意机制的启发，我们为模型设计了一维特征编码器，通过为横、纵坐标中的每个通道进行编码，产生方向感知特征。这使我们的模块沿着一个空间方向捕获远程依赖关系，并沿着另一个空间方向保留精确的位置信息，这有助于网络更准确地定位感兴趣的对象。

## 4 复现细节

### 4.1 数据集介绍

CAMUS数据集由500名患者的超声心动图组成。为了保证临床数据的真实性，既没有进行预处理也没有进行数据选择。CAMUS数据集的提出主要用于左心室射血分数的计算。数据集由训练集 450 位病人、测试集50 位病人组成，每位病人包括10-20张不等的二腔数据及10-20张不等的四腔数据（如图所示）。在2023年之前,CAMUS数据集仅为舒张末期（ED）和收缩末期（ES）两个时相的数据提供标注，今年更新后的CAMUS数据集拥有了全部数据的真实标签，因此，对CAMUS进行分割的任务性质由半监督转为全监督。在本文中，将数据集划分为7:1:2(训练集:验证集:测试集)

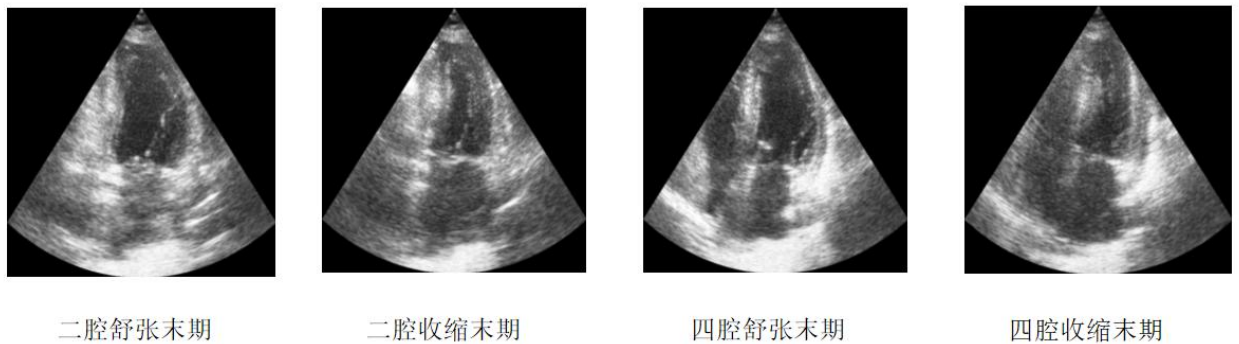


图3 CAMUS数据集心脏数据

### 4.2 训练细节

本文模型使用vit-b预训练权重初始化参数，在8张NVIDIA GeForce RTX 2080 Ti 显卡上训练。在训练过程中，使用组合损失函数进行监督，该函数由Dice损失和交叉熵损失组成，两者比例为8:2。为了便于比较以及实际应用，仅使用一个前景点提示符。模型由Adam优化器训练，学习率为0.0001。

### 4.3 TRA

涉及论文 idea 实现，暂不开源，请谅解。

### 4.4 LIA

涉及论文 idea 实现，暂不开源，请谅解。

## 5 实验结果分析

在本节中，我们将分析我们提出的模型与基准模型以及其他SOTA模型之间的性能差异。

如表1所示，在CAMUS数据集中，我们提出的模型有效地增强了分割结果的时间一致性，同时在dice分数等指标上相较于SAMUS基准框架及其他SOTA模型都有不同幅度的提升。

Method	CAMUS			
	Dice	IoU	HD	TCI
U-Net	91.68	84.96	3.2451	458
CE-Net	91.26	84.07	3.5381	561
SAM	85.1	74.98	8.43	1696
SAMUS	92.76	86.7	3.1184	295
Ours	93.15( $\pm 0.35$ )	87.19( $\pm 0.24$ )	3.0694( $\pm 0.0168$ )	274( $\pm 26$ )

表1 与SOTA模型及基准模型的比较

在消融实验中，我们首先对比了三种不同的对时间建模的方式，分别是时间适配器、时间注意力以及我们提出的TRA模块。由于硬件条件的限制，我们并未在CAMUS全部的数据上进行研究。为了减短时间，我们在CAMUS数据集中随机采样200份病例。在其基础上以与全部数据集相同的比例，即7:1:2，划分出训练集、验证集和测试集。在Mini-CAMUS上进行时间模块的消融研究。值得注意的是，我们为时间模块选择的位置是在ViT与CNN的交叉注意力之前，期望它能在复杂地全局建模之前完善时间方面的建模。

Method	Mini-CAMUS			
	Dice	IoU	HD	Bad case
ST-Adapter	91.57	84.74	3.4791	456
Temp-Att	91.43	84.55	3.3937	458
L-MHRA	91.8	84.89	3.1854	421

表2 时间模块的消融研究

如表2所示，三种时间建模模块中我们提出的TRA模块效果最好，因此我们选择TRA模块作为最终模型的时间建模模块。模型整体的消融研究如表3所示，可以看出TRA模块与HW模块在时间方面以及2D图像分

割方面有不同的提升。我们将成功归因于一下两个原因：1) 时间模块有效地增强了模型的时间建模能力，同时对帧与帧之间的时间依赖性有良好的表示。2) HW模块通过将2D图像分为高与宽两部分1D序列进行建模，增强了模型对变化幅度较大的心脏形状的鲁棒性，使得分割结果更加精确。

Temporal Block	H-W Block	Dice	IoU	HD
X	X	92.76	86.7	3.1184
X	✓	93.24 (+0.48)	87.49 (+0.79)	3.0735 (-0.0449)
✓	X	93.11 (+0.35)	87.32 (+0.62)	3.0589 (-0.595)
✓	✓	93.49 (+0.73)	88.25 (+1.55)	3.0105 (-0.1079)

表3 消融实验

## 6 总结与展望

针对医学图像的分割任务，在SAM大模型的基础上有许多工作，我们在SAMUS模型的基础之上，提出了一种针对超声心动图数据集的分割模型，旨在加强模型对时间一致性方面的建模以及使模型更准确地定位到感兴趣的对象。

首先，针对时间方面的建模方法，我们对比了关于时间适配器、时间注意力以及我们提出的方法，称为时间关系聚合模块，通过卷积操作使模型获得时间尺度上的特征信息，并且提高模型的时间一致性建模能力。其次，针对2D图像的分割领域，我们提出的模块，称之为HW模块，旨在将图像分为不同维度的1D序列，从而使模型能够捕获跨通道信息。最后在CAMUS数据集上与其他SOTA框架做了大量对比实验以及消融实验，证明了我们提出的模块的有效性。

在未来的计算机视觉领域中，大模型注定为医学图像分割领域带来更多的提升。在SAM出现后，使用大量医学数据集训练出的SAM-Med2D已经出现，成为医学图像分割领域的大模型。在此浪潮下，针对具体子领域、不同数据集的大模型调整方案将会是未来研究的主流。在未来的超声心动图研究中，我们的方法可以应用到其他并未全标注的数据集上，为相关的半监督分割方案带来新的思路。同时针对超声心动图的多区域统一分割，我们的模型也有很好的效果。

## 参考文献

- [1] Painchaud, N.; Duchateau, N.; Bernard, O.; and Jodoin, P.-M. 2022. Echocardiography segmentation with enforced temporal consistency. *IEEE Transactions on Medical Imaging*.
- [2] Wei, H.; Cao, H.; Cao, Y.; Zhou, Y.; Xue, W.; Ni, D.; and Li, S. 2020. Temporal-consistent segmentation of echocardiography with co-learning from appearance and shape. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 623–632. Springer.
- [3] Li, KunChang and Wang, Yali and He, Yinan and Li, Yizhuo and Wang, Yi and Wang, Limin and Qiao, Yu. 2023. UniFormerV2: Unlocking the Potential of Image ViTs for Video Understanding. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 1632-1643.
- [4] Kirillov, Alexander and Mintun, Eric and Ravi, Nikhila and Mao, Hanzi and Rolland, Chloe and Gustafson,

Laura and Xiao, Tete and Whitehead, Spencer and Berg, Alexander C. and Lo, Wan-Yen and Dollar, Piotr and Girshick, Ross. 2023. Segment Anything. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 4015-4026.

[5] Cheng Chen, Juzheng Miao, Dufan Wu, Zhiling Yan, Sekeun Kim, Jiang Hu, Aoxiao Zhong, Zhengliang Liu, Lichao Sun, Xiang Li, Tianming Liu, Pheng-Ann Heng, Quanzheng Li. MA-SAM: Modality-agnostic SAM Adaptation for 3D Medical Image Segmentation. arXiv:2309.08842, 2023.

[6] Xian Lin, Yangyang Xiang, Li Zhang, Xin Yang, Zengqiang Yan, Li Yu. SAMUS: Adapting Segment Anything Model for Clinically-Friendly and Generalizable Ultrasound Image Segmentation. arXiv:2309.06824, 2023.