

DOI: 10.13382/j.jemi.B2407869

融合注意力和状态空间的儿科超声心动图分割*

龚瑾儒¹ 翟锦涛¹ 田峰¹ 雷卫瑞¹ 常帅¹ 王润民² 邹孝¹ 钱盛友¹

(1. 湖南师范大学物理与电子科学学院 长沙 410081; 2. 湖南师范大学信息科学与工程学院 长沙 410081)

摘要:不同年龄段的儿童心脏尺寸差异大且儿童心率过快会导致心脏边界相较成人更加模糊,影响超声心动图的分割效果。针对上述问题,对 H2Former 分割模型进行改进,提出了一种时间和位置注意力的分层混合视觉状态空间模型(TPA-H2VSS)对儿科超声心动图左心室进行分割。首先替换 Transformer 模块为视觉状态空间模块,改进医学图像分割的长期依赖性建模关系;其次在模型的编码器与解码器之间搭建时间注意力模块,把超声心动视频左心室的语义信息在时间维度上进行补充和交互;最后,在输出部分加入位置注意力模块,进一步提高网络的分割性能。在儿科超声心动视频数据集 EchoNet-Pediatric 的 PSAX 数据集和 A4C 数据集上分别进行训练、验证和测试。与基线模型 H2Former 相比,在 PSAX 数据集上的 Dice、交并比(IoU)、准确率(accuracy)分别提升了 0.86%、1.41%、0.15%,豪斯多夫距离(HD)降低了 0.219 5;在 A4C 数据集上的 Dice、IoU、Accuracy 分别提升了 0.93%、1.53%、0.2%,HD 降低了 0.167。与其他模型进行比较,该模型能有效分割儿科超声心动图左心室,可以为先天性心脏病辅助诊断提供新的解决方案。

关键词:深度学习;图像分割;儿科超声心动图;注意力机制;状态空间模型

中图分类号: TP391;TN911.7 文献标识码: A 国家标准学科分类代码: 510.40

Pediatric echocardiography segmentation combining attention and state space

Gong Jinru¹ Zhai Jintao¹ Tian Feng¹ Lei Weirui¹ Chang Shuai¹Wang Runmin² Zou Xiao¹ Qian Shengyou¹

(1. School of Physics and Electronics, Hunan Normal University, Changsha 410081, China;

2. School of Information Science and Engineering, Hunan Normal University, Changsha 410081, China)

Abstract: The significant variation in cardiac dimensions across different age groups and the faster heart rate in children result in more blurred cardiac borders compared to adults, impacting the segmentation of echocardiography. To address the above problems, the segmentation model called H2Former is improved, and the model called TPA-H2VSS combining the attention and state space is proposed to segment the left ventricle of pediatric echocardiography. Firstly, this paper replaces the Transformer block with the visual state space (VSS) block to enhance the model's advantage in long-range modeling. Secondly, the temporal attention (TA) module is built between the encoder and decoder in the model to complements and interacts with the semantic information of the left ventricle in the echocardiography video in the temporal dimension. Finally, the positional attention (PA) module is added in the output head to make pediatric echocardiographic left ventricle segmentation more accurate. The experiments were trained, validated, and tested on the pediatric echocardiographic video dataset EchoNet-Pediatrics on the PSAX dataset and the A4C dataset, respectively. Compared with the base model H2Former, Dice, IoU, and accuracy on the PSAX dataset were improved by 0.86%, 1.41%, and 0.15%, respectively, and HD was reduced by 0.219 5. Dice, IoU, and accuracy on the A4C dataset were improved by 0.93%, 1.53%, and 0.2%, respectively, and HD was reduced by 0.167. By comparing with other models, it was demonstrated that the model could effectively segment the left ventricle in pediatric echocardiography and could provide a new solution for the auxiliary diagnosis of congenital heart disease.

收稿日期: 2024-09-30 Received Date: 2024-09-30

* 基金项目: 国家自然科学基金(12274200, 61502164)、湖南省教育厅科研基金(21A0052, 22B0036)、湖南省自然科学基金(2020JJ4057)、长沙市自然科学基金(kq2202239)项目资助

Keywords: deep learning; image segmentation; pediatric echocardiography; attention mechanism; state space model

0 引言

先天性心脏病 (congenital heart disease, CHD) 被认为是最常见的先天性缺陷之一^[1], 超声心动图对先天性心脏病的诊断和评估起着至关重要的作用。准确计算心脏功能检测指标依赖于对超声心动图中左心室轮廓的精确分割^[2]。

超声心动图左心室分割存在难点, 例如超声心动图中存在斑点噪声^[3]、血池和心肌之间对比度较低^[4]、不同患者之间心脏的形状存在差异性^[5]。另外, 对于婴幼儿等儿科患者来说, 患者间心脏大小样本方差较大、心率较快, 超声心动图中心脏边界相较成人更加模糊, 增加了分割难度。

近年来, 以卷积神经网络 (convolutional neural network, CNN) 和 Transformer 为代表的深度学习模型在医学图像分割方面取得了瞩目的成就^[6-7]。Ronneberger 等^[8]提出用于医学图像分割的 UNet, 该网络利用编解码结构和特征融合方法来实现目标分割。Zeng 等^[9]提出 MAEF-Net 对左心室进行分割, 该网络将多注意力机制与高效的特征融合相结合。Guo 等^[10]在编码器部分使用具有通道注意力的双路径特征提取模块来增强特征提取能力, 同时, 在解码器部分采用了一个具有空间注意力的高低级特征融合模块, 以更好地融合高低级特征。Chen 等^[11]首次提出了混合 CNN-Transformer 架构的 TransUNet 网络对医学图像进行分割。但是对于一些混合 CNN-Transformer 结构, 如 UTNet^[12]、BoTNet^[13]、TransUNet, 只考虑令牌自注意力来构建令牌之间的关系, 无法捕捉不同尺度的特征。CrossFormer^[14]虽然采用多尺度特征融合, 但并没有进行特征选择, 导致不同尺度特征之间的冲突。针对上述问题, He 等^[15]提出了一种医学图像分割模型 H2Former, 将 Transformer 的令牌智能注意力、多尺度通道注意力 (multi-scale channel attention, MSCA) 和 CNN 的强归纳偏置相结合形成混合 Transformer 模块 (hybrid Transformer block, HTB), 从不同方面聚合特征提取能力。

H2Former 中的 Transformer 模块在捕获远程依赖关系方面表现出优异的性能, 然而, Transformer 的自注意力机制在图像尺寸上要求二次复杂度, 导致较高的计算成本, 特别是对于医学图像分割等需要密集预测的任务。近年来, 状态空间模型 (state space model, SSM) 引起了研究人员的广泛关注。以 Mamba^[16]为代表的现代状态空间模型不仅建立了长距离依赖关系, 还能将计算复杂度降低到线性。Liu 等^[17]提出了一种通用的视觉主干模型 VMamba, 它使用了基于交叉扫描模块 (cross-scan

module, CSM) 的视觉状态空间 (visual state space, VSS) 模块, 在具有全局接受野的二维图像空间中实现一维选择性扫描。Ye 等^[18]将 Mamba 合并到编码器和解码器分支中进一步捕获全局依赖关系, 提高了左心室分割性能。但是基于 Mamba 的模型在医学视频分割方面还不成熟, 主要集中在单帧分割, 只能提取空间特征, 忽略了儿科超声心动视频连续帧之间的时间一致性。研究者将目光转向利用时序信息的视频分割^[19-20]。基于光流的方法在自然视频分割中的成功使其应用于超声心动视频^[21-22]。然而, 由于心脏运动的不规则性和各向异性以及斑点噪声的存在, 基于光流的方法往往无法有效捕捉超声心动视频中的运动信息, 导致性能不理想。

针对上述问题, 提出了一种新的融合注意力和状态空间的时空深度神经网络, 基于时间和位置注意力的分层混合视觉状态空间模型 (hierarchical hybrid visual state space based on temporal and position attention, TPA-H2VSS) 对儿科超声心动图左心室进行分割以解决上述挑战。该模型是一个以 H2Former 作为基线网络的编解码结构, 这样有利于不同形状和尺度的儿科超声心动图左心室分割。首先, 使用 VSS 模块替代混合 Transformer 模块中的 Transformer 模块以形成新的混合视觉状态空间模块 (hybrid VSS block, HVB), 改进左心室分割的远程依赖关系建模, 在提升分割精度的同时保持了线性复杂度。此外, 为了充分利用和传输标记帧和未标记帧之间的时间上下文, 设计了时间注意力 (temporal attention, TA) 模块对编码器和解码器之间的时间性进行建模, 促进了视频片段语义信息的交互融合, 同时为了更好地捕捉复杂图像空间的多尺度特征, 在时间注意力模块内将得到的注意图通过选择性内核注意力 (selective kernel attention, SKA)^[23]与输入特征融合。最后, 在模型的输出部分引入位置注意力 (positional attention, PA) 模块结合时间序列信息来增强特征表达, 进一步提升网络提取左心室区域的能力。

1 整体思路

1.1 整体网络架构

整体网络框架如图 1(a) 所示, 编码器结构由 4 个阶段组成。提出的基于时间和位置注意力的分层混合视觉状态空间模型的“分层”是指在每个阶段都采用混合视觉状态空间模块, 这有利于医学图像的分割任务。“混合”是指将卷积、多尺度通道注意力、VSS 3 种特征编码策略集成到一个统一的块中。

本文使用 H2Former 作为 TPA-H2VSS 的基线网络。

模型将连续 3 帧超声心动视频片段作为输入, 帧指数分别为 $i-2, i-1, i$ 。模型的输出为一个二值分割, 帧索引为 i 。输入由维度为 $F \in \mathbf{R}^{C \times T \times H \times W}$ 的帧序列组成, 其中 C 表示每一帧的通道数, T 表示输入序列的时间范围, $H \times W$ 表示输入的空间分辨率。在编码器部分, 将混合 Transformer 模块中的 Transformer 模块替换为 VSS 模块形成一个新的混合视觉状态空间模块。在编码器与解码器之间引入 TA 模块从不同角度提取多分辨率的上下文信息, 然后将集成了高级时空表示的输出特征作为解码器的输入。解码器部分包含多个补丁扩展层 (patch expending) 和卷积层 (convolution layers), 通过多个卷积

操作逐步实现特征融合和细化。最后在网络的末端加入 PA 模块结合时间序列信息来增强特征表达。

1.2 混合视觉状态空间模块 (hybrid VSS block)

编码器每个阶段都含有一个混合视觉状态空间模块, 结构如图 1(b) 所示。该结构由 VSS 模块、卷积模块 (convolutional block) 和 MSCA 组成。该模块首先通过卷积模块和 MSCA 提取局部特征和多尺度通道特征, 并将它们聚合在一起。然后, 通过 VSS 模块提取丰富的特征, 以达到同时结合 CNN、多尺度通道智能注意力和 VSS 的全局特征的优势的目的。

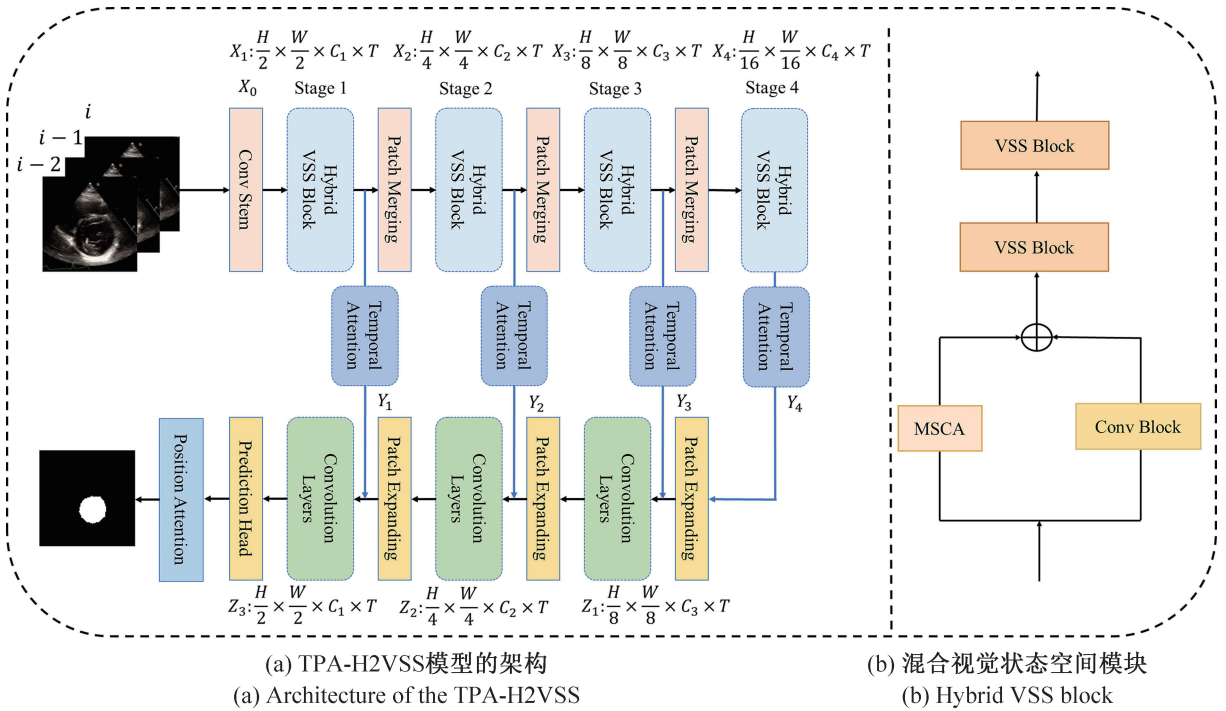


图 1 TPA-H2VSS 模型的架构和混合视觉状态空间模块

Fig. 1 Overall architecture of the proposed TPA-H2VSS and hybrid VSS block

VSS 模块的结构如图 2(a) 所示。输入特征经过层归一化 (layer normalization) 后被分为两个分支。在第 1 个分支中, 输入特征先经过线性层 (linear layer), 然后经过深度可分离卷积 (depth-wise separable convolution, DW-Conv)、激活函数和 2D 选择性扫描 (2D-selective-scan, SS2D) 模块。随后, 使用 layer normalization 对特征进行归一化。在第 2 个分支中, 输入特征先经过线性层和激活函数处理, 然后与第一条路径的输出特征合并。最后, 使用线性层混合特征, 并将此结果与残差连接相结合, 形成 VSS 模块的输出。

采用 SiLU (sigmoid-weighted linear unit)^[24] 作为激活函数。每个阶段 VSS 模块的个数为 $[2, 2, 2, 2]$ 。与视觉

变压器不同, VSS 模块避免了位置嵌入, 而是选择了没有多层感知器 (multi-layer perceptron, MLP) 的流线性结构, 从而在相同的深度预算内实现更密集的模块堆栈。

SS2D 模块包括 3 个步骤, 分别为交叉扫描、S6 块选择性扫描和交叉合并, 如图 3 所示。给定输入图像, 交叉扫描操作首先沿着四条不同的扫描路径将图像展开成序列。然后使用 S6 块处理每个序列, 确保各个方向的信息被彻底扫描, 从而捕获不同的特征。随后, 使用交叉合并操作对输出进行合并, 构建 2D 特征图作为最终的输出。SS2D 通过采用互补的扫描路径, 使图像中的每个像素能够有效地整合来自不同方向的所有其他像素的信息, 从而促进全局接受域的建立。

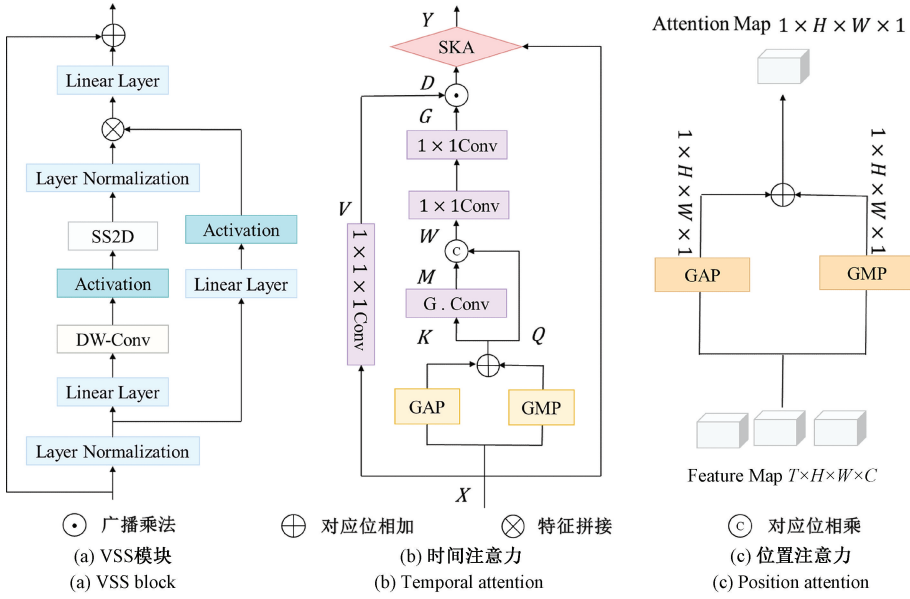


图 2 VSS 模块、时间注意力和位置注意力的具体架构

Fig. 2 Specific architecture for VSS block, temporal attention, and position attention

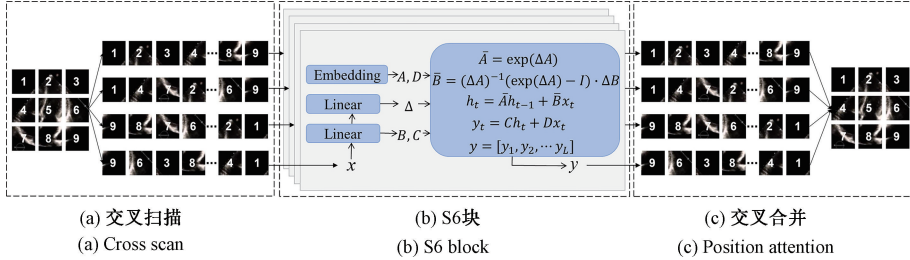


图 3 2D 选择性扫描模块

Fig. 3 Illustration of the SS2D module

1.3 TA

TA 模块的结构如图 2(b) 所示。首先将原始三维特征映射投影到二维平面,目的是过滤掉不相关的背景并显示左心室的关键信息。以经过每个混合视觉状态空间模块后的输出特征 X 为例。使用全局最大池化(global maximum pooling, GMP)和全局平均池化(global average pooling, GAP)沿着时间方向进行时间投影操作得到查询值 Q 和键值 K ,同时通过单次卷积得到 V , V 仍然保持 3D 形状。具体操作如下:

$$K = Q = GMP_T(X_i) + GAP_T(X_i) \quad (1)$$

之后,通过对 K 进行群大小为 4 的群卷积,提取局部时间注意力 M 。将 Q 与 M 拼接起来得到 W ,然后通过 W 进行两个连续 1×1 二维卷积和维数扩展操作,进一步得到全局关注 G 。注意力矩阵 G 不仅编码孤立查询键对内的上下文信息,还编码键内的注意力。接下来,基于二维全局注意力 G ,计算全局时间融合注意力图 D 来增强特征表示。具体操作如下:

$$M = \sigma(GConv(K)) \quad (2)$$

$$W = Concat(M, Q) \quad (3)$$

$$G = Expend(Conv(\sigma(Conv(W)))) \quad (4)$$

$$D = V \odot G \quad (5)$$

式中: $\sigma(\cdot)$ 代表归一化操作; $GConv(\cdot)$ 代表群卷积; $Expend(\cdot)$ 表示维数扩展操作; $Concat(\cdot)$ 表示级联操作; $Conv(\cdot)$ 代表卷积操作; \odot 代表广播乘法。

为了更好地捕捉复杂图像空间的多尺度特征,将得到的全局时间融合注意力图 D 通过 SKA 与输入特征融合进一步得到输出特征 Y 。SKA 允许每个神经元对于不同尺寸的输入信息可以自适应地调整其感受野的大小。SKA 的结构如图 4 所示。

1.4 PA

PA 模块的结构如图 2(c) 所示。使用特征图 Z 作为输入并生成位置注意力图 P 。PA 模块关注输入连续帧的每一帧中关键信息的位置。首先,在时间维度和通道维度对特征图采用 GAP 和 GMP 操作,然后将获得的两个特征图相加。具体操作如下:

$$P = GMP(Z) \oplus GAP(Z) \quad (6)$$

式中: \oplus 表示相加操作。

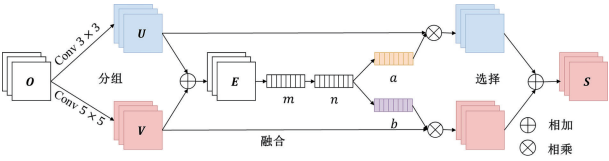


图 4 选择性内核注意力
Fig. 4 Selective kernel attention

2 实验方法和结果分析

2.1 实验数据

使用的数据集为儿科超声心动视频数据集 EchoNet-Pediatric^[25]。该数据集由斯坦福大学于 2023 年提供,包含 4 424 个胸骨旁短轴 (parasternal short axis, PSAX) 超声心动视频和 3 176 个顶端四腔 (pical-4-chamber, A4C) 超声心动视频。该数据集涵盖了典型的超声心动图像相应的标记测量值,包括收缩末期和舒张末期的左心室容积、专家标注的左心室轮廓。专家只在每个视频的收缩末期和舒张末期进行注释。视频每帧的尺寸为 112×112。将训练集、验证集和测试集按 8 : 1 : 1 的比例进行划分。数据集样本及对应左心室轮廓如图 5 所示。

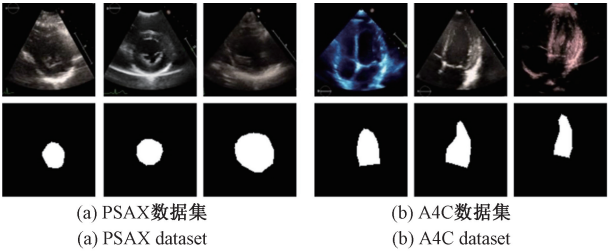


图 5 样本及对应左心室轮廓

Fig. 5 Sample and corresponding left ventricular contour

2.2 实验条件

所有实验都是在 NVIDIA GeForce RTX 3090 GPU 上用 Pytorch1. 13. 1 实现的。整个训练过程使用 AdamW 优化器,初始学习率为 0. 000 1,权重衰减为 0. 000 01。每经过 10 轮,学习率更新 1 次,调整倍数为 0. 9。batch size 设置为 16,训练轮数为 90。将图片输入大小由 112×112 调整为 128×128,输入视频片段是包含心脏收缩末期或舒张末期在内的连续 3 帧。损失函数采用二元交叉熵损失 (binary cross-entropy loss, BCE Loss)。

2.3 评价指标

为了评估所提出的模型的分割性能,选取 Dice 系数 (Dice coefficient)、交并比 (intersection over union, IoU)、准确率 (accuracy) 和豪斯多夫距离 (Hausdorff distance, HD) 作为其性能评价指标。

HD 量化了预测分割与参考分割之间的最大距离。该评价指标用于测量预测分割的边界与参考分割的边界之间的差异,较小的 HD 表示更好的分割结果,公式如下:

$$HD = \text{Max}_{distance} \{ predicted, GT \} \quad (7)$$

式中: $predicted$ 为网络输出的预测结果; GT 为真实标签。

2.4 消融实验

在 TA 模块的全局时间融合注意力图 D 与输入特征 X 的融合部分,分别使用 SimA (simple attention)^[26]、SEA (squeeze-and-excitation attention)^[27]、EMA (efficient multi-scale attention)^[28]、ECA (efficient channel attention)^[29]、CA (comprehensive attention)^[30] 和 SKA 模块进行比较,实验结果如表 1 所示。

表 1 TA 模块在 PSAX 数据集上对实验结果的影响

| Table 1 Impact of TA module on experimental resultson PSAX dataset | | | | |
|--|---------------|---------------|---------------|-----------------|
| 方法 | Dice/% | IoU/% | Accuracy/% | HD/mm |
| SimA | 92. 18 | 85. 84 | 98. 62 | 1. 613 3 |
| SEA | 92. 24 | 85. 94 | 98. 64 | 1. 574 9 |
| EMA | 92. 19 | 85. 87 | 98. 62 | 1. 600 2 |
| ECA | 92. 14 | 85. 80 | 98. 62 | 1. 653 0 |
| CA | 92. 18 | 85. 84 | 98. 63 | 1. 597 8 |
| SKA | 92. 30 | 86. 06 | 98. 64 | 1. 559 4 |

从表 1 可以看出,当融合部分使用 SKA 模块时,模型的 Dice、IoU、Accuracy 和 HD 分别达到了 92. 30%、86. 06%、98. 64% 和 1. 559 4,实现了更精确的分割。这是因为 SKA 注意力能够根据输入动态选择不同大小的卷积核。这种设计使得网络可以根据输入自适应地调整其感受野,从而更有效地捕获不同尺度的特征。因此,在 TA 模块的融合部分,将使用 SKA 模块进一步对左心室进行分割。

为了证明各个模块的有效性,在 PSAX 数据集上进行消融实验,将 VSS 模块、TA 模块和 PA 模块去掉以形成一个以 H2Former 为基线的网络,并将它们逐一加入到基线网络中,实验结果如表 2 所示。

表 2 各种改进方案在 PSAX 数据集上对实验结果的影响

| Table 2 Impact of each improvement scheme on the experimental results on the PSAX dataset | | | | | | |
|---|----|----|---------------|---------------|---------------|-----------------|
| VSS | TA | PA | Dice/% | IoU/% | Accuracy/% | HD/mm |
| — | — | — | 91. 44 | 84. 65 | 98. 49 | 1. 778 9 |
| ✓ | — | — | 91. 76 | 85. 19 | 98. 56 | 1. 681 1 |
| — | ✓ | — | 91. 84 | 85. 30 | 98. 57 | 1. 676 5 |
| — | — | ✓ | 91. 63 | 85. 01 | 98. 54 | 1. 688 7 |
| ✓ | ✓ | — | 92. 10 | 85. 68 | 98. 61 | 1. 628 9 |
| ✓ | — | ✓ | 91. 85 | 85. 30 | 98. 58 | 1. 650 0 |
| — | ✓ | ✓ | 92. 11 | 85. 75 | 98. 61 | 1. 607 6 |
| ✓ | ✓ | ✓ | 92. 30 | 86. 06 | 98. 64 | 1. 559 4 |

从表 2 可以观察到,当仅使用 H2Former 作为基线网络时,模型的 Dice、IoU、Accuracy 和 HD 分别为 91.44%、84.65%、98.49%、1.778 9。将 Transformer 模块替换为 VSS 模块后,与原来的基线网络相比,Dice、IoU、Accuracy 分别提升了 0.32%、0.54%、0.07%,HD 降低了 0.097 8。当单独加入 TA 模块和 PA 模块后,分割精度均有所提升,这表明 TA 模块和 PA 模块可以通过利用时间和空间信息引导网络更加关注儿科超声心动图的左心室区域从而提高网络的分割性能。此外,还研究了模块之间的相互作用对分割的影响,移除单个模块并将剩余模块成对组合。由表 2 的对比结果可知,模块的两两组合并不会降低网络的性能,反而会超过单个模块的性能。最后,当把 3 个模块融合后,网络性能同样得到改善,与基线网络对比,Dice、IoU、Accuracy 分别提升了 0.86%、1.41%、0.15%,HD 降低了 0.219 5。消融实验证明了 VSS 模块、TA 模块和 PA 模块对于提升网络性能具有重要作用,三者相互配合能提高左心室区域的分割准确率。

2.5 不同方法结果比较

为了评估模型的性能,将本文设计的模型与基线模型以及其他 9 种分割模型进行定量比较。

PSAX 数据集实验结果如表 3 所示。本文设计的方法在各项评估指标上均表现优异,Dice、IoU、Accuracy 和 HD 分别达到 92.30%、86.06%、98.64%、1.559 4,与 UNet 相比,Dice、IoU、Accuracy 分别提升了 1.67%、2.78%、0.28%,HD 降低了 0.422 1。这是因为 UNet 采用小卷积核的卷积操作,忽略了全局上下文信息。同时,与性能较优的 DATransUNet 相比,Dice、IoU、Accuracy 分别提高了 1.11%、1.82%、0.19%,HD 降低了 0.307 7。

表 3 不同算法在 PSAX 数据集上的性能指标

| Table 3 The values of performance indexes by different algorithms on the PSAX dataset | | | | |
|---|--------|-------|------------|---------|
| 方法 | Dice/% | IoU/% | Accuracy/% | HD/mm |
| FCN ^[31] | 89.98 | 82.32 | 98.26 | 2.144 7 |
| UNet ^[8] | 90.63 | 83.28 | 98.36 | 1.981 5 |
| DeepLabV3 ^[25] | 89.59 | 81.69 | 98.20 | 2.246 3 |
| TransUNet ^[11] | 90.76 | 83.59 | 98.38 | 2.042 5 |
| SSFormer ^[32] | 90.90 | 83.76 | 98.40 | 1.966 0 |
| DCSAUNet ^[33] | 90.28 | 82.80 | 98.28 | 2.126 7 |
| DATransUNet ^[34] | 91.19 | 84.24 | 98.45 | 1.867 1 |
| AMSUNet ^[35] | 89.07 | 80.88 | 98.07 | 2.473 3 |
| Mamba_UNet ^[36] | 90.89 | 83.78 | 98.40 | 1.949 1 |
| H2Former ^[15] | 91.44 | 84.65 | 98.49 | 1.778 9 |
| TPA-H2VSS | 92.30 | 86.06 | 98.64 | 1.559 4 |

表 4 为 A4C 数据集上的对比实验结果,TPA-H2VSS 模型在各项指标上均优于其他几种分割方法,在 Dice、IoU、Accuracy 上分别达到 90.75%、83.47%、98.10%,同时,TPA-H2VSS 模型获得了最小的 HD,这意味着它有能

力处理模糊的边界。与基线网络 H2Former 相比,Dice、IoU、Accuracy 分别提高了 0.93%、1.53%、0.2%,HD 降低了 0.167。与性能较好的 SSFormer 相比,Dice、IoU、Accuracy 分别提高了 1.14%、1.84%、0.24%,HD 降低了 0.351 4。

通过对比发现,尽管 DATransUNet 和 SSFormer 能够捕获全局上下文,但仍然获得比 TPA-H2VSS 更低的精度。这种观察结果归因于充分利用连续帧之间的时间信息缓解了超声心动图中模糊的边界的影响。此外,VSS 模块可以更好地捕获远程依赖关系,从而更准确地识别左心室轮廓。虽然 DATransUNet 和 SSFormer 两种算法在特定的数据集上能取得较好的效果,但在不同数据集上表现不同,而 TPA-H2VSS 模型在两个数据集上均实现了最优的分割结果,进一步证明了 TPA-H2VSS 的泛化性。

表 4 不同算法 A4C 数据集上的性能指标

| Table 4 The values of performance indexes by different algorithms on the A4C dataset | | | | |
|--|--------|-------|------------|---------|
| 方法 | Dice/% | IoU/% | Accuracy/% | HD/mm |
| FCN ^[31] | 88.44 | 79.84 | 97.61 | 2.784 1 |
| UNet ^[8] | 88.61 | 79.98 | 97.64 | 2.587 6 |
| DeepLabV3 ^[25] | 88.52 | 79.94 | 97.65 | 2.776 3 |
| TransUNet ^[11] | 89.07 | 80.77 | 97.76 | 2.648 7 |
| SSFormer ^[32] | 89.61 | 81.63 | 97.86 | 2.508 8 |
| DCSAUNet ^[33] | 88.44 | 79.86 | 97.60 | 2.949 2 |
| DATransUNet ^[34] | 89.21 | 81.01 | 97.78 | 2.636 8 |
| AMSUNet ^[35] | 87.25 | 78.06 | 97.40 | 3.357 0 |
| Mamba_UNet ^[36] | 89.21 | 81.09 | 97.82 | 2.639 9 |
| H2Former ^[15] | 89.82 | 81.94 | 97.90 | 2.324 4 |
| TPA-H2VSS | 90.75 | 83.47 | 98.10 | 2.157 4 |

为了对实验结果进行更深入的分析,TPA-H2VSS 模型和其他模型在测试集上的可视化图如图 6 所示。其中,白色区域表示专家在儿科超声心动图左心室上的手工分割结果,红色轮廓代表算法自动分割结果。前两行代表 PSAX 数据集上分割的结果,后两行代表 A4C 数据集上分割的结果。结果表明,几种模型在两个数据集上均受伪影的干扰,在左心室舒张末期和收缩末期出现欠分割和过分割现象,相比之下,TPA-H2VSS 在一定程度上降低了伪影的干扰,更接近手动分割结果,同时可以发现该算法受分割区域面积变化的影响较小,在心脏收缩末期左心室区域极小时,表现仍然优于其他几种算法。

除了数值和视觉结果外,还提供了评估指标的统计结果。使用箱线图对所有测试图像的评价指标分布进行分析和对比。图 7 和 8 所示为所有算法在 PSAX 数据集和 A4C 数据集上的 Dice 分布,有效地显示了数据的分散度。箱型图描绘了下四分位数、上四分位数、中位数、最小值和最大值。从图 7、8 可以看出,TPA-H2VSS 明显优于其他模型,并在两个数据集上表现出一致的性能。

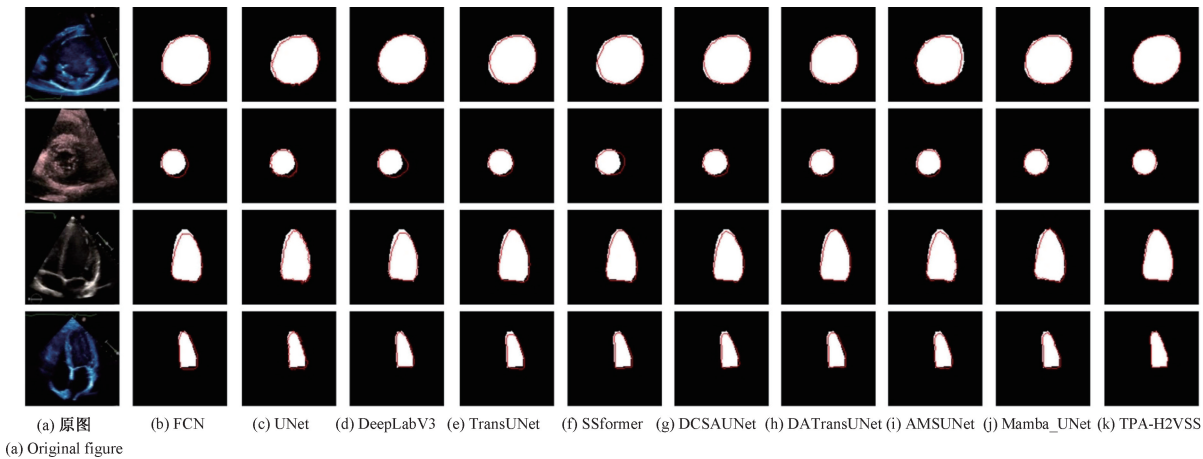


图 6 PSAX 数据集和 A4C 数据集的左心室分割结果
Fig. 6 The segmentation of left ventricle on the PSAX dataset and A4C dataset

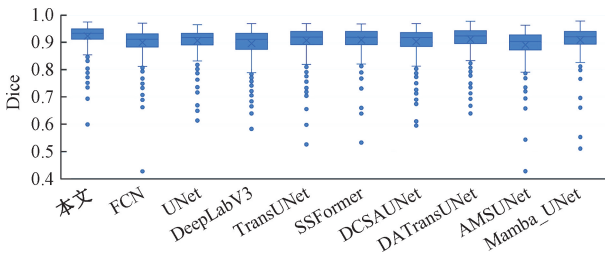


图 7 在 PSAX 数据集上 Dice 的箱型图
Fig. 7 The Boxplots of Dice on PSAX dataset

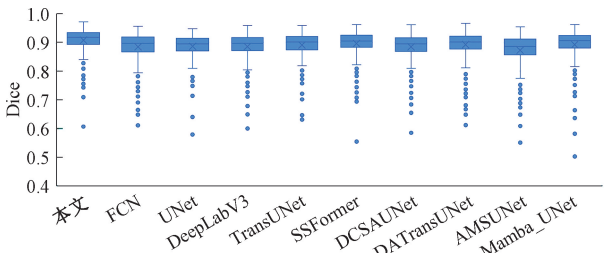


图 8 在 A4C 数据集上 Dice 的箱型图
Fig. 8 The Boxplots of Dice on A4C dataset

3 结 论

以往关于儿科超声心动图的研究主要集中在单帧分割,没有充分探索视频帧之间的时间一致性。此外,Transformer 的自注意力机制的复杂性随着输入图像尺寸的增加呈二次增长,导致较高的计算成本。针对上述问题,提出了一种融合注意力和状态空间的模型 TPA-H2VSS 对儿科超声心动图像左心室进行分割。该算法利用 VSS 对远程依赖的建模能力和 CNN 对局部信息的归纳能力提升分割精度。在此基础上,在编码器和解码器之间加入时间注意力融合当前帧和历史帧的特征,从

不同角度提取上下文信息。最后在网络的输出部分加入位置注意力,进一步结合时间序列信息来增强特征表达,提高网络分割左心室区域的准确性。实验结果表明,本文设计的算法能够取得更准确的分割结果。相对基线模型 H2Former,在 PSAX 和 A4C 数据集上的 Dice 分别提升 0.86% 和 0.93%, IoU 分别提升 1.41% 和 1.53%, Accuracy 分别提升 0.15% 和 0.2%, HD 降低了 0.219 5 和 0.167。此外,本文只针对儿科超声心动图左心室进行研究,未来的研究将聚焦于更为复杂和多样的医学图像,以提升模型在广泛的医学分割任务中的应用能力。

参考文献

[1] 杨涛,王芳,朱玉蓉,等. 中国儿童先天性心脏病流行病学特征研究进展 [J]. 心脏杂志, 2024, 36 (1): 106-111.
YANG T, WANG F, ZHU Y R, et al. Research progress on epidemiological characteristics of congenital heart disease in Chinese children [J]. Chinese Heart Journal, 2024, 36 (1): 106-111.
[2] 庞俊,王永雄,陈丽君,等. 面向儿科超声心动图双侧心室分割的注意力引导网络 [J]. 生物医学工程学报, 2023, 40 (5): 928-937.
PANG J, WANG Y X, CHEN L J, et al. An attention-guided network for bilateral ventricular segmentation in pediatric echocardiography [J]. Journal of Biomedical Engineering, 2023, 40 (5): 928-937.
[3] PAINCHAUD N, SKANDARANI Y, JUDGE T, et al. Cardiac segmentation with strong anatomical guarantees [J]. IEEE Transactions on Medical Imaging, 2020, 39 (11): 3703-3713.
[4] ZHANG J, WANG Y, CHEN L, et al. Dual-branch transV-Net for 3-D echocardiography segmentation [J].

- IEEE Transactions on Industrial Informatics, 2023, 19(12): 11675-11686.
- [5] LIU F, WANG K, LIU D, et al. Deep pyramid local attention neural network for cardiac structure segmentation in two-dimensional echocardiography [J]. Medical Image Analysis, 2021, 67: 101873.
 - [6] 邱拴虎, 杨文瀚, 廖苗, 等. 基于 RA-Unet 的 CT 图像肝脏肿瘤分割[J]. 仪器仪表学报, 2022, 43(8): 65-72.
DI SH H, YANG W H, LIAO M, et al. Liver tumor segmentation in CT images based on RA-Unet [J]. Chinese Journal of Scientific Instrument, 2022, 43(8): 65-72.
 - [7] 何晓云, 许江淳, 陈文绪. 基于改进 U-Net 网络的眼底血管图像分割研究[J]. 电子测量与仪器学报, 2021, 35(10): 202-208.
HE X Y, XU J CH, CHEN W X. Research on fundus blood vessel image segmentation based on improved UNet network [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(10): 202-208.
 - [8] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]. Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. IEEE, 2015: 234-241.
 - [9] ZENG Y, TSUI P H, PANG K, et al. MAEF-Net: Multi-attention efficient feature fusion network for left ventricular segmentation and quantitative analysis in two-dimensional echocardiography [J]. Ultrasonics, 2023, 127: 106855.
 - [10] GUO L, LEI B, CHEN W, et al. Dual attention enhancement feature fusion network for segmentation and quantitative analysis of paediatric echocardiography [J]. Medical Image Analysis, 2021, 71: 102042.
 - [11] CHEN J, LU Y, YU Q, et al. Transunet: Transformers make strong encoders for medical image segmentation[J]. ArXiv preprint arXiv: 2102. 04306, 2021.
 - [12] GAO Y, ZHOU M, METAXAS D N. Utnet: A hybrid transformer architecture for medical image segmentation[C]. Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. IEEE, 2021: 61-71.
 - [13] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 16519-16529.
 - [14] WANG W X, YAO L, CHEN L, et al. CrossFormer: A versatile vision transformer based on cross-scale attention[J]. ArXiv preprint arXiv: 2108. 00154, 2021.
 - [15] HE A, WANG K, LI T, et al. H2Former: An efficient hierarchical hybrid transformer for medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2023, 42(9): 2763-2775.
 - [16] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces [J]. ArXiv preprint arXiv: 2312. 00752, 2023.
 - [17] LIU Y, TIAN Y J, ZHAO Y Z, et al. Vmamba: Visual state space model [J]. ArXiv preprint arXiv: 2401. 10166, 2024.
 - [18] YE Z, CHEN T, WANG D, et al. HFE-Mamba: High-frequency enhanced mamba network for efficient segmentation of left ventricle in pediatric echocardiograms [J]. IEEE Access, 2024, 12: 123038-123048.
 - [19] CHEN F, HAN H, NING G, et al. Immunohistochemical index prediction of breast cancer by using hybrid ultrasound data [J]. IEEE Transactions on Biomedical Engineering, 2022, 70(4): 1401-1412.
 - [20] HAN H, LIAO H, ZHANG D, et al. Thyroid nodule diagnosis in dynamic contrast-enhanced ultrasound via microvessel infiltration awareness [C]. Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. IEEE, 2023: 169-179.
 - [21] LI M, ZHANG W, YANG G, et al. Recurrent aggregation learning for multi-view echocardiographic sequences segmentation [C]. Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. IEEE, 2019: 678-686.
 - [22] PAINCHAUD N, DUCHATEAU N, BERNARD O, et al. Echocardiography segmentation with enforced temporal consistency [J]. IEEE Transactions on Medical Imaging, 2022, 41(10): 2867-2878.
 - [23] LI X, WANG W, HU X, et al. Selective kernel networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510-519.
 - [24] ELFWING S, UCHIBE E, DOYA K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning [J]. Neural Networks, 2018, 107: 3-11.
 - [25] REDDY C D, LOPEZ L, OUYANG D, et al. Video-based deep learning for automated assessment of left ventricular ejection fraction in pediatric patients [J]. Journal of the American Society of Echocardiography, 2023, 36(5): 482-489.
 - [26] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural

- networks [C]. International Conference on Machine Learning, 2021: 11863-11874.
- [27] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [28] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]. International Conference on Acoustics, Speech and Signal Processing. IEEE, 2023: 1-5.
- [29] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [30] GU R, WANG G, SONG T, et al. CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2020, 40 (2): 699-711.
- [31] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [32] WANG J, HUANG Q, TANG F, et al. Stepwise feature fusion: Local guides global [C]. Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. IEEE, 2022: 110-120.
- [33] XU Q, MA Z, NA H E, et al. DCSAU-Net: A deeper and more compact split-attention U-Net for medical image segmentation [J]. Computers in Biology and Medicine, 2023, 154: 106626.
- [34] SUN G, PAN Y, KONG W, et al. DA-TransUNet: Integrating spatial and channel dual attention with transformer U-net for medical image segmentation [J]. Frontiers in Bioengineering and Biotechnology, 2024, 12: 1398237.

- [35] YIN Y, HAN Z, JIAN M, et al. AMSU-Net: A neural network using atrous multi-scale convolution for medical image segmentation [J]. Computers in Biology and Medicine, 2023, 162: 107120.
- [36] WANG Z, ZHENG J Q, ZHANG Y, et al. Mamba-UNet: Unet-like pure visual mamba for medical image segmentation [J]. ArXiv preprint arXiv: 2402.05079, 2024.

作者简介



龚瑾儒, 2022 年于湖南理工学院获得学士学位, 现为湖南师范大学硕士研究生, 主要研究方向为医学图像处理。

E-mail: gongjinru@hunnu.edu.cn

Gong Jinru received her B. Sc. degree from Hunan Institute of Science and Technology in 2022. Now she is a M. Sc. candidate in Hunan Normal University. Her main research interests include medical image processing.



邹孝(通信作者), 2021 年于湖南师范大学获得博士学位, 现为湖南师范大学副教授、硕士生导师, 主要研究方向为生物医学电子学和医学图像处理。

E-mail: shawner@hunnu.edu.cn

Zou Xiao (Corresponding author) received Ph. D. degree from Hunan Normal University in 2021. Now he is an associate professor and M. Sc. supervisor in Hunan Normal University. His main research interests include biomedical electronics and medical image processing.



钱盛友, 1997 年于上海交通大学获得博士学位, 现为湖南师范大学教授、博士生导师, 主要研究方向为生物医学电子学和智能仪器。

E-mail: shyqian@hunnu.edu.cn

Qian Shengyou received Ph. D. degree from Shanghai Jiaotong University in 1997. Now he is a professor and Ph. D. supervisor in Hunan Normal University. His main research interests include biomedical electronics and intelligent instruments.