DDPTransSeg：基于 DAFM 的双路径 Transformer 用于 3D 多模态心脏分割框架

霍连昊 李碧原（通讯作者）

天津职业技术师范大学电子工程学院，天津，300222；

摘要：心血管疾病为全球首要致死原因，准确心脏分割对诊疗至关重要。但心脏形态高度可变、结构边界细微，单模态成像难以应对这一挑战。为此，我们提出 DDPTransSeg 多模态心脏分割方法，基于 CT/MRI 数据构建双路径 Transformer 框架：以 Swin Transformer 为编码器捕获模态特定特征，通过双注意力融合模块（DAFM）动态校准通道贡献，保留互补信息并抑制冗余；SEFA 块进一步强化特征选择，解码器则恢复空间分辨率以实现精确边界定位。在 MM-WHS 2017 数据集上评估，DDPTransSeg 表现卓越：Dice 分数达 82.96%，MIoU 为 72.33%，HD95 降至 8.39 毫米，性能优于现有 CNN-Transformer 模型，证实其在多模态心脏分割中的有效性与临床潜力。

关键词：多模态心脏分割；双路径编码器-解码器；基于 Transformer 的架构

DOI：10.69979/3029-2808.26.01.079

1 介绍

心血管疾病（CVDs）是非传染性疾病首要致死原因[1]，世界卫生组织数据显示，缺血性心脏病与中风长期占据全球死因前两位[2]，这凸显了可靠诊断工具的迫切需求。在此背景下，3D 医学图像分割成为计算机辅助诊断的核心[3]，其对心脏等复杂器官的精准描绘，是疾病检测、个性化治疗规划等下游应用的关键[4]。

心脏分割仍极具挑战：患者心脏形态差异显著，且成像噪声与组织低对比度常导致边界模糊[5]，传统单模态方法难以兼顾稳健性与精度。多模态成像虽能提供互补信息 ——CT 擅长高分辨率解剖可视化，MRI 具备优异软组织对比度[7]，但因强度分布差异、采集协议不同及模态错位，融合难度极大。

深度学习为解决该问题提供新思路。原用于自然语言处理的 Transformer 架构[6]，凭借长距离依赖建模能力进军视觉领域，ViT [7] 与 Swin-Unet[19] 的成功推动其向医学分割延伸。但纯 Transformer 弱于局部细节捕捉，而 CNN 虽擅长提取精细局部特征，却难处理全局空间一致性[9]。二者互补性催生混合架构，这类设计对心肌边界等关键解剖线索的捕捉至关重要[8]，成为医学分割研究热点。

2 相关工作

Transformer 最初为自然语言处理设计[10]，其全局注意力机制可建模长距离依赖，近年在计算机视觉领域展现潜力。视觉变压器（ViT）[7] 开创图像块序列标记范式，Swin-Unet[19] 引入移位窗口分层架构，显著提升高分辨率图像分析效率，二者衍生的 Swin - UNETR[11] 已成功应用于医学图像体积分割。

但 Transformer 缺乏强归纳偏差，依赖大规模标注数据，且难以捕捉医疗所需的精细局部结构。多模态融合虽能整合 CT 与 MRI 互补信息，却面临诸多挑战：早期 HyperDense-Net[12] 密集连接路径参数冗余、训练低效；引入注意力的融合方法[13] 虽优化特征选择，却未能优先考量模态特异性贡献；跨模态转换器[14] 在 3D 场景中计算开销巨大。

CNN 擅长捕捉局部上下文但对高阶特征依赖敏感，ViT 建模上下文能力强却成本高。混合架构如 LMRT-NET[15] 虽减少参数与内存占用，却依赖直接拼接等简单融合策略，忽略模态特征重要性差异（如 MRI 界定心肌边界的关键作用），引入冗余信号，导致复杂解剖结构分割效果不佳。

为此，我们提出 DAFM 模块，在 CT/MRI 特征融合中动态调整通道重要性。通过自适应加权模态线索，DAFM 有效保留临床相关信息、抑制冗余，提升多模态心脏分割的准确性与鲁棒性。

3 方法

3.1 结构概述

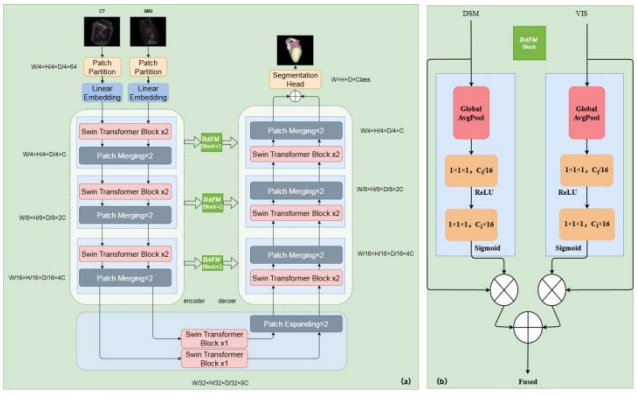

图. 1 DDPTransSeg：基于 DAFM 的双路径变换器用于三维多模态心脏分割框架

如图 1 (a)所示，DDPTransSeg 采用双路径编码器-解码器框架进行多模态心脏分割。首先将 CT 和 MRI 输入数据嵌入为标记，并通过基于 Swin Transformer 的编码器进行独立处理，这些编码器能在不同尺度上捕捉特定模态特征。在编码器与解码器的连接处设置双注意力融合模块（DAFM），通过通道级注意力机制动态平衡 CT 与 MRI 的贡献。随着网络深度增加，SEFA 模块进一步优化融合后的表征。解码器通过块扩展层逐步恢复空间分辨率，利用跳跃连接传递编码器特征实现精准定位。最终，融合的多尺度特征经分割头投影后，生成心脏亚结构的体素级预测结果。

3.2 双注意力融合（DAFM）模块

图 1(b)则呈现了 DAFM 模块的具体结构。在 DDPTransSeg 体系中，作为关键跨尺度特征融合单元的 DAFM 模块，在编码和解码过程中多次部署，旨在提升模型整体性能和三维分割能力。该模块采用基于通道注意力机制的双向特征融合架构，通过独立的 SE 子模块（Spatial Enapsulation 子模块）分别对输入的两组三维特征图（编码器特征与解码器特征）生成通道注意力权重。每个 SE 子模块首先通过全局平均池化（GAP）压缩空间维度，接着利用两个 1×1×1 三维卷积学习通道间关联性，并生成归一化通道权重（范围 0-1）。两组特征图与对应权重相乘后，既能突出关键通道的响应特征，又能抑制冗余信息。最终通过逐元素求和的方式融合加权特征图，从而保留跨层特征的互补信息。针对三维医学图像中不同解剖结构的空间分布特征，通过通道注意力机制对各通道的贡献权重进行动态调整，增强关键区域的响应。

4 实验

4.1 数据集

我们采用广泛认可的多模态心脏分割基准数据集 MM-WHS 2017 [13] 评估 DTransSeg 模型。该数据集含 CT 与 MRI 影像及专家标注的 7 种心脏亚结构（左心室肌 Myo、左心房腔 LA、左心室腔 LV、右心房腔 RA、右心室腔 RV、升主动脉 AA、肺动脉 PA），可覆盖解剖复杂性与模态异质性，共 60 组 3D 心脏图像（20 组标注用于训练，40 组未标注用于测试），各结构以特定体素强度编码（205、420、500、550、600、820、850），背景为 0。我们将数据集随机划分为 16 组训练样本与 4 组测试案例，采用 5 项指标评估：1）Dice 相似系数（衡量区域重叠，对微小结构关键）；2）平均交并比（MIoU，评估整体一致性）；3）95% 豪斯多夫距离（HD95，毫米级，量化边界精度）；4）可训练参数（Params/M，百万级，不含不可训练统计量）；5）浮点运算次数（FLOPs/G，十亿级）。FLOPs 基于输入张量（1,2,128,128,128），通过 thop 库（v0.0.31）计算单次前向传播值。

4.2. 实验结果

如表 1 所示，本文提出的 DDPTransSeg 在所有对比方法中表现最佳。该模型以 82.96%的 Dice 评分和 72.33%的 MIoU 值拔得头筹，超越了 nnFormer（Dice79.5%，MIoU70.87%）等强劲基线模型。此外，其 HD95 值(8.39 毫米）最低，显示出更优的边界精度，这对肺动脉等复杂解剖结构尤为重要。在计算效率方面，DDPTransSeg 需要 61.4M 参数和 376.2 GFLOPs。虽然比 MedNeXt 等轻量级网络（5.5M 参数，138.4 GFLOPs）更重，但其效率优于 Swin-Unet（453.1 GFLOPs），同时保持显著更高的精度。这种精度与效率的平衡突显了双路径设计和基于动态自适应特征融合模块（DAFM）的特征融合策略的有效性。总体而言，这些结果证实 DDPTransSeg 通过有效整合 CT 与 MRI 信息，不仅超越了纯 transformer 架构和混合 CNN-Transformer 模型，更实现了临床可靠边界勾画的顶尖分割精度，展现出在心脏病诊断和手术规划中的广阔应用前景。

表 1 不同网络在数据集上的实验结果（%）

方法	年份	Dice↑	MIoU↑	HD95↓	Params/M	FLOPs/G
nnFormer [17]	2021	79.5	70.87	9.11	37.351	170.575
VT_Unet [18]	2022	76.6	63.02	14.81	20.508	127.078
Swin-Unet [19]	2022	71.28	56.16	17.47	33.579	453.068
MedNeXt [20]	2023	79.22	66.68	12.02	5.542	138.427
DDPTransSeg (ours)	2025	82.96	72.33	8.39	61.427	376.233

5 结论和未来工作

针对心脏分割中形态个体差异大、细微边界模糊及多模态数据融合难的问题，本研究提出 DDPTransSeg——基于双路径 Transformer 与 DAFM 模块的 3D 多模态心脏分割框架。该框架以双路径 Swin Transformer 编码器分别提取 CT（高空间分辨率）与 MRI（高软组织对比度）特征，通过 DAFM 模块动态校准模态贡献权重，抑制冗余信息并保留互补价值，配合 SEFA 模块与解码器实现空间分辨率恢复与精准分割。

在 MM-WHS 2017 数据集（60 组 CT/MRI 数据，含 7 种心脏亚结构）上，DDPTransSeg 表现优异：Dice 系数达 82.96%、MIoU 为 72.33%、HD95 降至 8.39mm，显著优于 nnFormer、Swin-Unet 等主流模型；同时以 61.4M 参数、376.2 GFLOPs 实现精度与效率平衡，为心血管疾病诊断、手术规划提供了可靠的解剖结构分割结果。

后续将从五方面优化：一是扩充多中心、多疾病类型的 CT/MRI 数据，提升模型对罕见病例的泛化性；二是通过稀疏注意力、知识蒸馏实现模型轻量化，适配边缘设备实时分割需求；三是引入半监督学习，利用少量标注数据结合无标注数据训练，降低标注成本；四是整合 PET-CT、超声等模态，构建 “结构 + 功能” 的端到端分割 - 评估任务链；五是开展临床实验，结合医生主观评价与 PACS 系统测试，推进技术落地。

参考文献

[1]Kumar A S, Rekha R. An improved hawks optimizer based learning algorithms for cardiovascular disease prediction[J]. Biomedical Signal Processing and Control, 2023, 81: 104442.

[2]World health statistics 2024: monitoring health for the SDGs, Sustainable Development Goals. Geneva: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO.

[3]Gao Y, Zhang J, Wei S, et al. PFormer: An efficient CNN-Transformer hybrid network with content-driven P-attention for 3D medical image segmentation[J]. Biomedical Signal Processing and Control, 2025, 101: 107154.

[4]Zhang X, Liu J, **an X, et al. PSVT: Pyramid Shifted Window based Vision Transformer for cardiac image segmentation[J]. Biomedical Signal Processing and Control, 2025, 102: 107339.

[5]Ma X, Shan S, Sui D. SAMP-Net: a medical image segmentation network with split attention and multi-layer perceptron[J]. Medical & Biological Engineering & Computing, 2025: 1-14.

[6]Liu Y, Wu Y H, Sun G, et al. Vision transformers with hierarchical attention[J]. Machine Intelligence Research, 2024, 21(4): 670-683.

[7]Li X, Jiang A, Qiu Y, et al. TPFR-Net: U-shaped model for lung nodule segmentation based on transformer pooling and dual-attention feature reorganization[J]. Medical & Biological Engineering & Computing, 2023, 61(8): 1929-1946.

[8]Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.

[9]Fan X, Liu L, Zhang H. multi-modal information interaction for medical image segmentation [J]. arxiv preprint arxiv:2404.16371, 2024.

[10]A. Vaswani, “Attention is all you need,” in Proc. 31st Conf. Neural Inf. Process. Syst., 2017, pp. 1 - 11.

[11]Ali Hatamizadeh, V. Nath, Yucheng Tang, Dong Yang, Holger R. Roth, and Daguang Xu, “Swin

unetr: Swin transformers for semantic segmentation of brain tumors in mri images,”in BrainLes@MICCAI, 2022.
[12]Dolz J, Gopinath K, Yuan J, et al. HyperDense-Net: a hyper-densely connected CNN for multi-modal image segmentation[J]. IEEE transactions on medical imaging, 2018, 38(5): 1116-1126.
[13]Liang M, Yang B, Chen Y, et al. Multi-task multi-sensor fusion for 3d object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 7345-7353.
[14]Zhang J, Liu H, Yang K, et al. CMX: Cross-modal fusion for RGB-X semantic segmentation with transformers[J]. IEEE Transactions on intelligent transportation systems, 2023, 24(12): 14679-14694.
[15]Zhu X, Li Y. A Latent Multi-Scale Residual Transformer Approach for Cross-Modal Medical Image Synthesis[J]. IEEE Access, 2025.
[16]X. Zhuang et al.,“Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge,”Med. Image Anal., vol. 58, 2019, Art. no. 101537.
[17]Zhou H Y, Guo J, Zhang Y, et al. nnformer: Interleaved transformer for volumetric segmentation[J]. arxiv preprint arxiv:2109.03201, 2021.
[18]Peiris H, Hayat M, Chen Z, et al. A robust volumetric transformer for accurate 3D tumor segmentation[C]//International conference on medical image computing and computer-assisted intervention. Cham: Springer Nature Switzerland, 2022: 162-172.
[19]Cao H, Wang Y, Chen J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2022: 205-218.
[20]Saikat Roy, Gregor Koehler, Constantin Ulrich, Michael Baumgartner, Jens Petersen, Fabian Isensee, Paul F. Jaeger, and Klaus H. Maier-Hein,“Mednext: Transformer-driven scaling of convnets for medical image segmentation,”ArXiv, vol. abs/2303.09975, 2023.

作者简介：霍连昊，男，生于 2001 年 3 月，汉族，山东费县人，就读于天津职业技术师范大学，学历：研究生，研究方向：医学图像处理。
李碧原，男，生于 1992 年 1 月，蒙古族，内蒙古自治区赤峰市人，天津职业技术师范大学专职教师，职务：讲师，学历：博士，研究方向：医学图像处理。