



FCT-Net: 基于 CNN 与 Transformer 双分支并行融合的斑马鱼心脏图像分割网络模型

张轩宇¹, 徐凌波^{1,2}

(1. 浙江理工大学理学院, 杭州 310018; 2. 浙江大学流体动力基础件与机电系统全国重点实验室, 杭州 310027)

摘要: 为了解决斑马鱼心脏图像轮廓模糊导致的医学图像分割精度不足问题, 提出了一种融合 CNN 与 Transformer 的新型网络模型——FCT-Net (Fusion convolution-transformer network)。该模型基于经典的编码器-解码器架构, 构建了双分支并行特征融合模块, 其中: CNN 分支用于提取局部组织特征, 并针对单一卷积核难以覆盖多尺度特征的局限, 在卷积模块中引入多尺度特征融合机制, 构建多感受野特征金字塔, 以增强对边缘细节的表征能力; Transformer 分支用于捕捉长距离的全局上下文依赖关系, 实现局部特征与全局语义的有效融合。实验结果表明, FCT-Net 在斑马鱼心脏图像分割任务中的准确率较基准 U-Net 模型提升了 5.8%, 有效提高了心脏轮廓分割精度。该模型具备高精度的斑马鱼心脏分割能力, 可以为后续基于斑马鱼心脏形态学特征的药物筛选研究提供较为可靠的算法支撑。

关键词: 斑马鱼; 心脏图像; CNN; Transformer; 多尺度特征融合

中图分类号: TP399

文献标志码: A

文章编号: 1673-3851 (2025) 07-0571-09

引用格式: 张轩宇, 徐凌波. FCT-Net: 基于 CNN 与 Transformer 双分支并行融合的斑马鱼心脏图像分割网络模型[J]. 浙江理工大学学报(自然科学), 2025, 53(4): 571-579.

Reference Format: ZHANG Xuanyu, XU Lingbo. FCT-Net: A zebrafish heart image segmentation network model based on the parallel fusion of dual branches of CNN and Transformer[J]. Journal of Zhejiang Sci-Tech University, 2025, 53(4): 571-579.

FCT-Net: A zebrafish heart image segmentation network model based on the parallel fusion of dual branches of CNN and Transformer

ZHANG Xuanyu¹, XU Lingbo^{1,2}

(1. School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2. State Key Laboratory of Fluid Power & Mechatronic Systems, Zhejiang University, Hangzhou 310027, China)

Abstract: To address the issue of insufficient accuracy in medical image segmentation caused by the blurred contours of zebrafish heart images, a novel network model called FCT-Net (fusion convolution-transformer network) is proposed, which integrates CNN and Transformer. This model is based on the classic encoder-decoder architecture and a dual-branch parallel feature fusion module is constructed. Specifically, the CNN branch is utilized to extract local tissue features. To overcome the limitation of a single convolutional kernel in covering multi-scale features, a multi-scale feature fusion mechanism is introduced within the convolutional module, and a multi-receptive field feature pyramid is built to enhance the representation capability of edge details. The Transformer branch is employed to capture long-range global contextual dependencies, achieving effective fusion of local features and global semantics. Experimental results demonstrate that FCT-Net improves the accuracy by 5.8% compared to the baseline U-Net model in the task of zebrafish heart image segmentation, effectively enhancing the precision of heart contour segmentation. With its high-precision zebrafish heart segmentation capability, this model can

收稿日期: 2025-03-18 网络出版日期: 2025-04-29

基金项目: 流体动力基础件与机电系统全国重点实验室开放基金项目(GZKF-202313)

作者简介: 张轩宇(2000—), 男, 江苏淮安人, 硕士研究生, 主要从事人工智能在生物医学中的应用方面的研究。

通信作者: 徐凌波, E-mail: xlb@zstu.edu.cn

provide relatively reliable algorithmic support for the subsequent drug screening studies based on the morphological characteristics of the zebrafish heart.

Key words: zebrafish; heart image; CNN; Transformer; multi-scale feature fusion

0 引言

近年来,斑马鱼模型在心脏功能研究领域取得了较为明显的进展,通过分析斑马鱼心脏形态可以为药物筛选和疾病机理探索等生物医学研究提供数据支持^[1-2],但现有的斑马鱼心脏形态分析方法仍面临不少挑战。传统的分析方式依赖于人工操作,将斑马鱼置于显微镜下进行观察,这种方式不仅费时费力,还受操作者主观判断的影响,容易引入主观误差,导致数据准确性和一致性下降,同时也增加了实验成本^[3-4]。因此,现有方法通常采用计算机图像处理来分析斑马鱼的心脏形态,其中关键的步骤是斑马鱼心脏图像分割。然而,斑马鱼体内心脏区域的图像高度透明且结构复杂,图像中各组织结构的边界较为模糊,这给心脏图像分割和特征提取带来了较大难度,限制了对心脏微小结构变化的精准捕捉,不利于后续生物医学研究^[5-7]。另外,由于斑马鱼心脏是动态变化的,在连续图像采集和处理过程中,对数据处理速度、图像质量以及存储等方面提出了更高要求,进一步加大了实验实施的复杂性。因此,亟需构建更有针对性的图像分割模型,以降低人工干预的影响,提高图像处理的准确性和效率,从而实现更精准、标准化的心脏功能评估,进而为药物筛选和心脏疾病治疗研究提供技术支撑^[8-11]。

医学图像分割的主要方法可分为传统图像处理方法与深度学习方法,近年来,深度学习方法凭借强大的特征提取能力,迅速成为该领域的研究热点与应用主流。在众多深度学习模型中,U-Net 模型凭借其对称的 U 型架构和独特的跳跃连接机制,在医学图像分割任务中表现出色,成为目前应用最为广泛的模型。然而,U-Net 在复杂边缘建模及小目标分割方面仍有一定局限性,为此研究者们不断对其架构进行迭代优化,研究重点主要聚焦于多尺度特征融合机制、注意力机制以及动态结构探索等^[12]。U-Net++通过密集跳跃连接重构特征传递路径,并利用嵌套子网络增强跨层级特征的复用性,但该模型也引入了较高的计算复杂度^[13];DP-UNet++进一步整合了可变形卷积与动态池化模块,并通过自适应感受野调整提升对不规则目标的形变建模能力,但动态计算显著影响了推理效率^[14];Attention

U-Net 引入了门控空间注意力机制,通过动态特征加权抑制了背景干扰并聚焦关键区域,但该模型在全局上下文建模上存在不足^[15];TransUNet 融合了 Transformer 与 U-Net 架构,并借助自注意力机制捕捉全局依赖关系,但存在显存与计算资源的巨大消耗问题^[16]。针对斑马鱼心脏图像分割的问题,上述模型表现并不理想,主要归因于两大挑战:一是模型效率与分割精度的平衡难题,这直接关系模型在临床应用中的实用性和可靠性;二是小样本情境下模型的泛化能力不足,特别是在精确描绘斑马鱼心脏边缘结构时表现较差。

针对上述问题,本文提出了一种融合 CNN 与 Transformer 的新型网络模型—FCT-Net (Fusion convolution-transformer network),通过融合不同尺度和层次的图像特征,在提升模型分割精度的同时兼顾运行效率,增强模型在小样本情境下的鲁棒性。FCT-Net 在经典编码器-解码器架构基础上,构建了双分支并行特征融合模块。一方面,通过 CNN 分支捕获局部组织特征,同时针对单一卷积核特征提取的局限性,在卷积模块中引入多尺度特征融合机制,通过构建多感受野特征金字塔来提升心脏边缘特征的表达能力;另一方面,通过 Transformer 分支捕捉全局上下文依赖关系,以实现局部特征与全局语义的有效融合。本文构建的 FCT-Net 模型通过双分支并行特征融合的方法,提升了斑马鱼心脏图像细节结构的分割精度,减少图像模糊导致的误判误差,为后续基于斑马鱼模型的药物筛选研究提供可靠的算法支撑。

1 网络模型构建

本文构建了 CNN 与 Transformer 双分支并行特征融合模块,该模块通过构建卷积局部感知与自注意力全局建模的深度融合机制,在特征编码阶段实现多尺度全局信息的动态融合。其中空间细节增强单元可以有效提取心脏轮廓边缘的纹理特征,而跨层特征交互组件通过建立长程依赖关系精准刻画心脏形态学特征,最终提升对斑马鱼心脏细微结构的分割精度。

1.1 FCT-Net 网络模型

FCT-Net 采用基于 U-Net 的模型框架,针对斑

马鱼心脏图像边缘模糊、局部特征捕捉不足的问题重新进行了设计,通过特征融合编码器模块来处理下采样过程产生的多尺度特征。相比于 U-Net 标准架构,FCT-Net 增强了对斑马鱼心脏局部纹理和全局结构信息的表达能力。

FCT-Net 网络模型如图 1 所示。FCT-Net 网络整体可分为下采样(Conv-Downsampling)、跳层连接(Skip Connection)和上采样(Conv-

Upsampling)等 3 个部分。该网络模型在下采样过程中通过 FCT-Module 进行特征提取,该模块包含两部分:一是利用 CNN 提取斑马鱼心脏图像中的局部特征;二是借助 Transformer 模块捕捉全局上下文依赖关系。上采样部分则将经过融合处理的特征图逐步生成最终的心脏掩码图。而跳层连接部分负责在下采样与上采样之间实现特征的直接传递与融合。

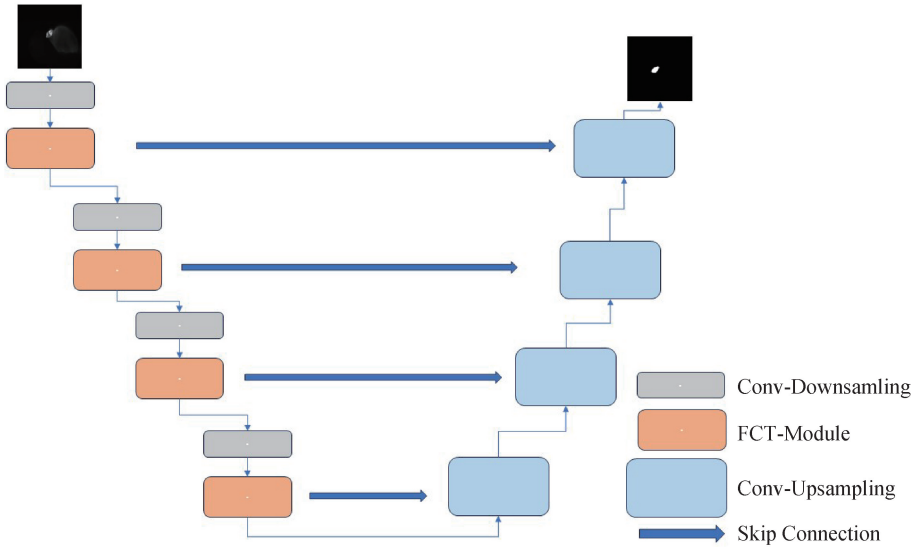


图 1 FCT-Net 网络模型示意图

1.2 CNN 和 Transformer 融合架构的特征融合模块

在斑马鱼心脏图像分割任务中,下采样过程对结构细节的捕获具有决定性作用。针对传统 U-Net

因单一卷积核感受野受限导致的纹理信息衰减问题,以及由于深层特征语义鸿沟引发的边缘模糊现象,本文提出了特征融合模块,其核心设计如图 2 所示。

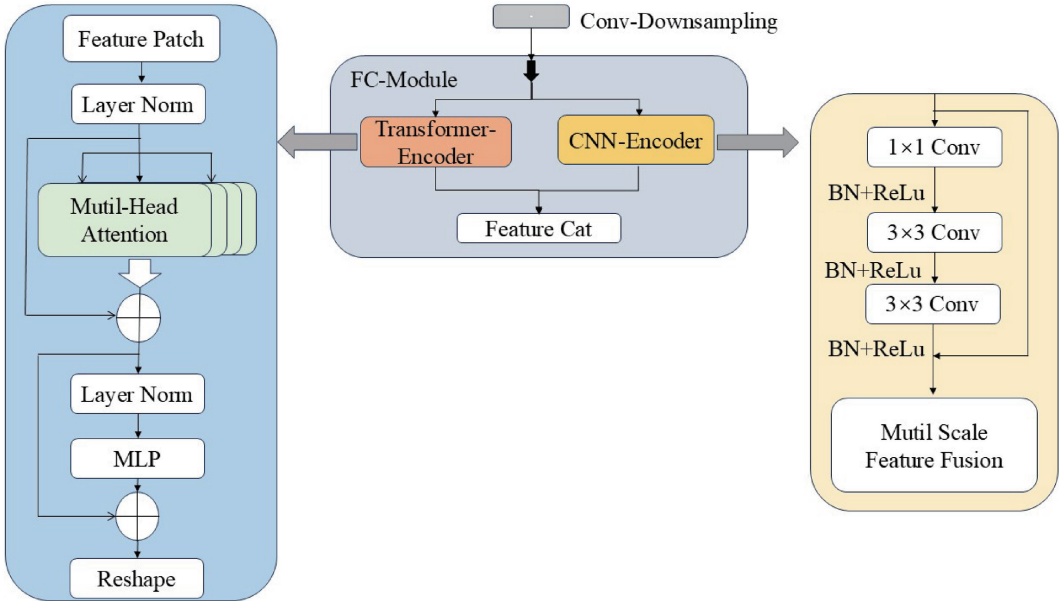


图 2 CNN 和 Transformer 融合架构的特征融合模块结构示意图

1.2.1 CNN 分支模块

CNN 分支模块(CNN-Encoder)如图 2 所示,

CNN 分支采用渐进式特征抽象策略,通过渐进式下采样与通道扩展实现从像素级细节到器官级语义的

编码;每个阶段包含 2 层相同的卷积残差单元,其核心结构由 1×1 通道压缩卷积、 3×3 深度可分离卷积及 3×3 通道扩展卷积串联构成;通道压缩层通过降维减少计算冗余,深度可分离卷积在提取局部空间特征的同时保留心肌细胞边界的亚像素细节,扩展卷积则恢复通道维度以匹配输入尺寸,最终通过残差连接将扩展特征与原始输入相加,有效缓解梯度消失问题。

针对单尺度卷积感受野受限的缺陷,设计多

尺度特征融合模块(Mutil Scale Feature Fusion, MSFF)以增强特征表达能力。其核心架构如图 3 所示,该模块同步采用 3×3 、 5×5 、 7×7 等 3 种卷积核并行提取多粒度特征:小核聚焦心脏边界的高频纹理,中核捕获心室壁的中尺度形态,大核刻画心腔整体轮廓特征,将各自捕获的信息进行串行拼接,在特征通道压缩的同时融合多尺度特征信息,并加入残差连接以防止网络层数过深引起的退化现象。

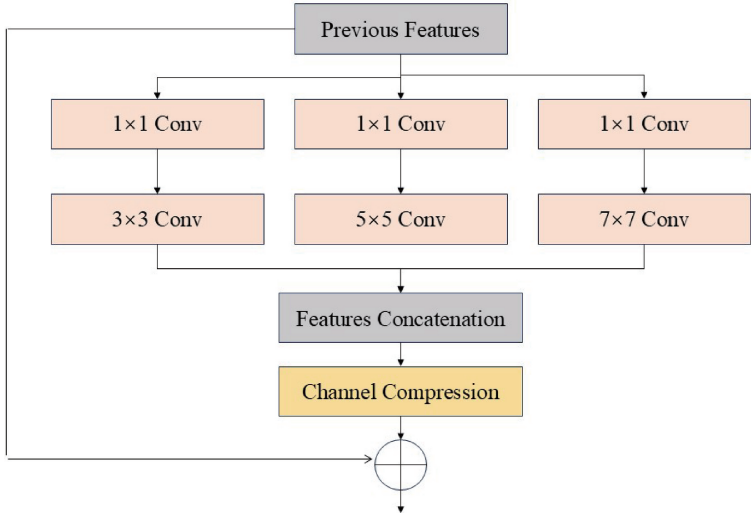


图 3 多尺度特征融合结构示意图

1. 2. 2 Transformer 分支模块

Transformer 分支(Transformer-Encoder)基于 Vision Transformer 框架改进,其核心架构如图 2 所示,通过堆叠 N 个定制化 Transformer 块捕捉全局上下文依赖关系。每个 Transformer 块由多头自注意力(Multi-Head Attention, MSA)模块与局部增强型 Multi-Layer Perceptron(MLP)模块级联构成,其中:在 MSA 模块前通过 LayerNorm 对特征进行实例归一化,采用特征通道压缩策略使得计算复杂度降低,同时保留心室-心房的长程空间关联性;MLP 模块嵌入扩展-压缩全连接层,其中扩展层将特征维度进行提升以增强非线性表达能力,压缩层恢复原始维度以避免信息冗余。为了进一步强化局部特征感知,在 MLP 中插入 3×3 深度卷积层,对特征 Patch 进行隐式编码,替代显式位置,消除因人工设计与心脏结构偏差导致的定位误差。

CNN 与 Transformer 分支输出的多模态特征分别承载了局部细节与全局语义信息,通过通道拼接与动态融合实现互补增强。将同层级的 CNN 特征图与 Transformer 特征图沿通道维度拼接,利用 1×1 卷积执行跨模态特征融合。该卷积层通过可

学习权重矩阵筛选与心脏解剖相关的显著性通道,抑制血流伪影等噪声干扰,最终生成兼具局部精细结构与全局信息的融合特征。该设计使模型在无需显式位置编码的情况下,仍能保持对斑马鱼心脏动态形变的高精度分割能力。

2 实验设计与结果分析

2. 1 实验设计

2. 1. 1 实验环境

实验使用的操作系统为 Windows11 专业版,深度学习框架为 PyTorch1. 12. 1。硬件配置中 CPU 为 AMD Ryzen 7 7840H,自带 Radeon 780M Graphics, 3. 80 GHz,内存 16. 0 GB, GPU 为 NVIDIAGeForce RTX4060。使用 Anaconda 来进行编程环境搭建,在 Anaconda 中安装 Python3. 9 作为编程语言使用 Pytorch 框架,并用 PyCharm 作为集成开发环境,进行基本的程序运行和模型搭建。

本文基于斑马鱼心脏图像构建了一个经过数据增强处理的数据集,该数据集包含 5400 张高质量图像样本。采用严格的数据分割策略,将其按照 8:1:1

的比例划分为训练集、验证集和测试集 3 个独立子集。训练集包含 4320 张图像,用于模型参数学习;验证集和测试集各包含 540 张图像,分别用于超参数调优和最终性能评估。在模型训练阶段,采用具有自适应学习率特性的 Adam 优化器进行参数更新,受限于 GPU 显存容量,批处理量设置为 3,以保证训练稳定性。

2.1.2 实验参数设置

本文建立了系统的验证机制,以监控模型训练动态。在每个训练周期结束后,模型在验证集上执行前向推理并计算平均 Dice 系数,特别是设计模型保存策略时,考虑到训练初期参数空间尚未稳定的特点,设置 80 个 epoch 的缓冲期后才启动最优模型保存机制,即仅当当前 epoch 的验证集 Dice 系数超越历史最佳表现时,才将当前模型参数存档。整个训练过程持续 200 个完整 epoch,最终选取全训练周期中验证集性能最优的模型参数作为最终产出模型。该实验参数设计方法通过缓冲期设置有效规避了早期过拟合风险,同时 200 个 epoch 的充分训练周期保障了模型收敛的完备性;验证集 Dice 系数的持续监测机制为模型选择提供了可靠的量化依据,确保获得泛化性能最优的解决方案。

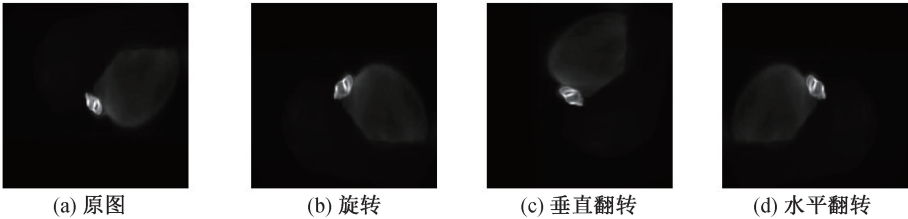


图 4 数据增强处理示例图像

2.2 实验结果

2.2.1 消融实验

为验证特征融合模块(FCT-Module)对纹理信息衰减与边缘模糊问题的改进效果,本文基于 U-Net 架构设计消融实验,设置以下对照组:

- Base:基准模型,使用标准 U-Net 中的上、下采样模块和简单的跳跃连接;
- Inception:Inception 模块;
- ASPP:空洞空间金字塔池化模块;
- MSFF:多尺度特征融合模块;
- TE:Transformer 分支模块;
- CE:CNN 分支模块;
- FCT:CNN 与 Transformer 协同架构的特征融合模块。

本文采用下列指标评价不同模型的分割结果:

2.1.3 数据采集与处理

本文使用实验室定制的自动化采集系统获取斑马鱼心脏图像,该系统配备 NikonAZ100 显微镜,采集高清 PNG 图像。成像过程中,斑马鱼置于毛细玻璃管内,通过步进电机旋转并结合位姿识别算法调整拍摄角度,确保多角度精准采集。同时,系统利用虹吸作用控制水流,结合夹管阀实现斑马鱼的自动化装载与定位,并通过液面差控制与位姿校准机制提升图像质量一致性。

此外,采用多阶段数据处理方案,以提高模型泛化能力。首先,应用旋转、垂直翻转和水平翻转等 3 种数据增强方法,有效提升模型对斑马鱼心脏形态变化的适应能力。数据增强处理示例图像如图 4 所示,这些增强操作在不破坏心脏生物结构真实性的前提下,显著增加了训练数据的多样性,既能防止过拟合,又增强了模型对不同空间方位样本的识别鲁棒性。其次,使用 Labelme 软件抠图生成二值化标签,并通过 JSON 文件高效管理图像-标签关系。为了进一步提升模型性能,利用深度学习特征迁移算法合成虚拟数据,以优化大模型预训练并缓解过拟合问题。最后,完成训练数据集的构建,进行 FCT-Net 网络模型的迭代优化与参数调优。

Pixel Accuracy:像素准确率,衡量像素分类准确程度,正确预测像素数与总像素数比值的百分比。

Dice 系数:衡量预测结果与真实标签重叠程度,取值范围为 0 到 100%,值越高表示分割效果越好。

IoU:交并比,预测结果与真实标签交集和并集比值的百分比,综合反映定位和分割准确性。

Params:模型参数数量,指模型中可学习参数数量,反映模型复杂度,单位为 M(百万)。

表 1 给出了不同模型的分割结果。由表 1 可以看出,特征融合模块(FCT-Module)在斑马鱼心脏图像分割任务中实现了一定的精度提升。基准模型(Base)采用传统 CNN 结构,由于单一尺度卷积核的局限性及浅层特征表达能力的受限,其 Dice 系数仅为 85.31%,IoU 为 83.64%,表明模型在分割精度方面的表现较差,特别是在边缘模糊区域的细节

识别上容易丢失信息。

在此基础之上,加入多尺度特征提取模块(Inception、ASPP、MSFF),模型的局部特征提取能力得到增强,并进行模型对比。其中,Inception 模块通过多尺度并行卷积,增强了对不同粒度特征的建模能力,使 Dice 系数达到 86.58%,IoU 为 85.69%;ASPP 模块利用不同扩张率的空洞卷积,扩展了感受野,在增强全局信息建模的同时保留了更多细节信息,使 Dice 系数达到 86.30%,IoU 为 86.47%;相比 Inception 和 ASPP 结构,多尺度特征融合模块(MSFF)更加强调跨尺度信息的整合,最终 Dice 系数提升至 87.31%,IoU 达到 86.92%,表明该模块在不同尺度特征的融合上更具优势。

为了增强全局信息建模能力,加入 Transformer 分支(TE),Dice 系数提升至 87.62%,IoU 提高至 86.18%。这表明模型在整体形态一致性方面得到了优化,能够更好地捕捉长距离依赖关系。然而,Transformer 主要建模全局信息,在局部边缘的精细处理上仍存在不足,导致心脏边界易受误分割影响。

相比单独使用 Transformer 分支(TE),采用 CNN 分支(CE)后,模型的局部特征提取能力进一步增强,Dice 系数提升至 88.06%,IoU 增加至

87.34%。这种提升主要归因于 CNN 的多尺度卷积结构可以提取不同尺度的信息,提高了模型对边缘和细节特征的刻画能力,使得分割结果更加精细。然而,由于 CNN 主要依赖局部特征提取,在边缘区域仍然存在误分割问题,尤其在目标与背景对比度较低的情况下,模型可能难以区分相似结构,影响整体分割精度。

完整的特征融合模块(FCT-Module)整合了 Transformer 分支和 CNN 分支,通过并行融合方式,使 Dice 系数达到 90.71%,IoU 提升至 89.77%。这一模块的主要优势体现在:结合了 Transformer 的全局建模能力,提升整体形态一致性,使目标区域的边界更加平滑;融合了 CNN 的多尺度特征提取方式,增强局部细节的捕捉能力,使得目标区域的分割更加精准,降低误分割率,在背景复杂、目标易混淆的情况下,提高模型的鲁棒性。

尽管完整的特征融合完整的模块的参数量增加至 78.50 M(较 Base 增加 131%),计算成本有所上升,但其在斑马鱼心脏图像分割任务中的性能提升较为明显,为心脏图像分析提供了更可靠的解决方案。综合来看,完整的特征融合模块在像素级精度、Dice 系数和 IoU 指标上均优于其他方法,展现出更好的全局-局部特征融合能力。

表 1 消融实验分割精度指标

网络模型	Pixel Accuracy/%	Dice 系数/%	IoU/%	Params/M
Base	88.27	85.31	83.64	33.95
Base+TE	92.45	87.62	86.18	56.20
Base+Inception	91.84	86.58	85.69	44.80
Base+ASPP	91.96	86.30	86.47	46.42
Base+MSFF	92.20	87.31	86.92	45.70
Base+CE	92.83	88.06	87.34	47.80
Base+FCT	94.07	90.71	89.77	78.50

消融实验各模块可视化如图 5 所示。图 5 显示:基准模型(Base)在斑马鱼心脏图像分割任务中存在较为明显的边缘模糊和误分割问题,特别是在低对比度区域,心脏与背景的分界不够清晰。引入 Transformer 分支(Base+TE)后,模型的全局形态一致性得到了提升,使得心脏区域的完整性有所改善。然而,由于 Transformer 主要关注长距离依赖关系,在局部细节的处理上仍然存在一定不足,导致边缘模糊的现象仍未得到根本性解决。

采用 Inception 结构(Base+Inception)后,多尺度特征提取能力得到了增强,使得模型可以捕捉不同尺度的心脏边缘信息,分割结果在细节表达上有

所提升。然而,由于其仍然依赖 CNN 进行特征提取,缺乏全局建模能力,在复杂背景下仍可能出现边界不清或误分割的情况。类似地,ASPP 结构(Base+ASPP)通过多尺度空洞卷积增强了对不同尺度目标的感知能力,使得边缘细节略有改善,但仍然存在局部区域误融合的问题。

引入多尺度特征融合(Base+MSFF)后,模型进一步增强了局部细节的分割能力,使得心脏的分割精度得到了提升,并减少了小目标区域的误分割现象。然而,由于缺乏全局特征的有效约束,在形态一致性方面仍存在一定的优化空间。相比之下,CNN 分支(Base+CE)进一步强化了模型对局部特

征的表达能力,使得 Dice 系数达到了 88.06%,边缘细节刻画更加精准,但在复杂区域的分割仍然可能受到背景干扰的影响。

完整特征融合模块(Base + FCT)结合了 Transformer 的全局建模能力和 CNN 的局部特征增强能力,使得心脏的分割精度得到进一步提升,并

在形态一致性上达到了最优。该模块不仅提高了模型的抗干扰能力,还能更有效地区分低对比度区域中的细节,使得 Dice 系数提升至 90.71%,IoU 达到 89.77%,在多个模型中表现最优。这表明,通过融合全局与局部特征,FCT-Net 在斑马鱼心脏图像分割任务中展现出了更强的鲁棒性和精确性。

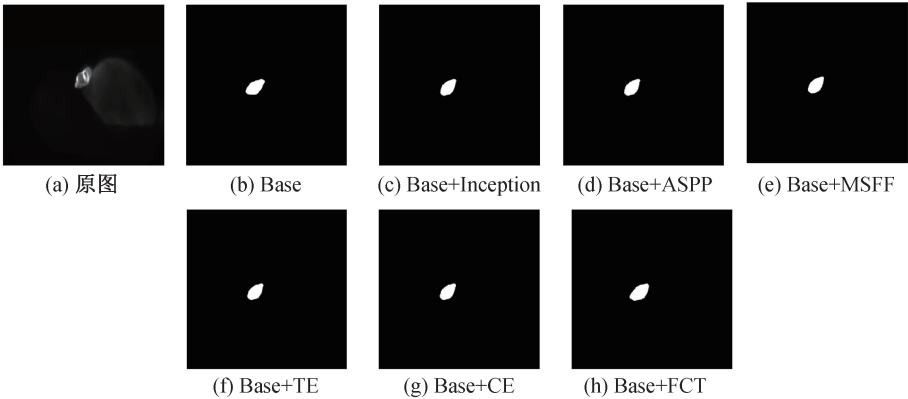


图 5 消融实验各模块可视化

2.2.2 对比实验

本文对比了 3 个经典双分支架构模型 Swin-Unet、CoTr、TransUNet。Swin-Unet 采用纯 Transformer 结构,结合 Swin Transformer 与 U-Net 架构,通过窗口自注意力和移位窗口机制,实现局部与全局特征的交互。其优势在于能够适应高分辨率图像(如病理切片、遥感图像)处理,同时保持 Transformer 强大的全局建模能力。然而,由于窗口化注意力的局限性,对小目标的分割可能不够精准。

TransUNet 采用 CNN 与 Transformer 串联的设计,利用 Vision Transformer 作为编码器提取全局特征,并在解码阶段通过 CNN 进行特征重建。其优势体现在对医学图像的多器官联合分割等需要广泛上下文信息的任务上,能够捕获远程依赖关系。然而,由于 Vision Transformer 需要将输入图像分块,会导致细节信息丢失,影响小目标的分割精度,且计算开销较大。

CoTr 采用 CNN 与 Transformer 交替堆叠的方式,每个阶段先用 CNN 提取局部特征,再用 Transformer 进行全局建模,实现渐进式特征融合。该方法兼顾边缘细节和长距离依赖,适合处理小目标分割任务。虽然该结构在模糊任务上表现优异,但由于其固定的交替堆叠方式,可能在时序建模和动态形变捕捉方面有所限制。

本文搭建的 FCT-Net 网络模型采用 CNN 与 Transformer 并联的设计,在编码器中针对不同尺

度的特征同时使用 CNN 和 Transformer,分别提取局部特征和全局信息,从而实现更加充分的信息融合。相比于 Swin-Unet 的纯 Transformer 结构,FCT-Net 通过并联 CNN 保留了更多的细节信息,增强了对小目标的识别能力;相比 TransUNet 采用 CNN 仅在解码阶段进行重建,FCT-Net 在编码阶段就同时融合了局部和全局特征,提高了语义理解能力;而相较于 CoTr 的交替堆叠方式,FCT-Net 采用并联结构,使得局部和全局特征能够在每个阶段同步提取并融合,提升了分割精度,更加适用于具有复杂纹理结构的图像分割任务。

为系统验证所提出的 FCT-Net 的有效性 with 泛化能力,本文选取医学图像分割领域 5 个代表性模型——U-Net、UNet++、DP-UNet++、TransUNet 与 FCT-Net 进行对比实验;同时,采用了与消融实验相同的训练参数,以便对比。基准模型 U-Net 采用经典的对称编解码结构,通过跳跃连接实现浅层细节与深层语义的初步融合,为后续改进提供基础框架;UNet++通过引入密集嵌套跳跃连接,构建多层次特征聚合路径,在提升微小目标分割精度的同时,因密集连接导致参数量增加约 45.3%;DP-UNet++进一步整合可变形卷积与动态池化技术,其可调节感受野特性增强了对不规则生物组织的形变适应能力;TransUNet 采用 CNN-Transformer 串行架构,利用自注意力机制建立长程空间依赖关系,但其分阶段特征处理模式易导致全局与局部信息交互不足,各网络模型的分割结果如表 2 所示。

表 2 显示:FCT-Net 在斑马鱼心脏分割任务中表现较好,像素精度 (94.07%)、Dice 系数 (90.71%)及 IoU(89.77%)均显著优于对比模型,其中 Dice 系数较侧重形变建模的 DP-UNet++提升 2.66%,较依赖密集连接的 UNet++提升 3.82%,验证了 FCT-Net 采用的多模态特征协同机

制的有效性。同时,FCT-Net 通过 CNN-Transformer 并行架构,在 78.50 M 参数量下实现性能突破——其参数量仅为 TransUNet 的 84.2%,而 Dice 系数反超 3.06%,且较 UNet++系列模型虽参数量增加 47.8%,却通过特征融合实现精度明显提升。

表 2 医学图像分割代表性模型与 FCT-Net 的图像分割结果评价指标

网络模型	Pixel Accuracy/%	Dice 系数/%	IoU/%	Params/M
U-Net	88.27	85.31	83.64	33.95
UNet++	92.18	86.89	85.94	49.35
DP-UNet++	93.45	88.05	87.92	53.12
TransUNet	92.31	87.65	87.16	93.23
FCT-Net	94.07	90.71	89.77	78.50

代表性模型与 FCT-Net 的图像分割结果示例图像如图 6 所示。图 6 表明:FCT-Module 通过局部-全局特征融合的方法,在提升分割精度的同时,控制了模型复杂度,实现了参数量与性能的高效平衡。相较于传统 CNN 或 Transformer 单一架构,FCT-Net 采用 CNN-Transformer 并行架构,能够

充分捕捉局部信息与全局上下文依赖关系,从而在保持较低计算成本的前提下,实现对斑马鱼心脏复杂形态的精确分割。此外,其较高的 Dice 系数和 IoU 表现进一步验证了多模态特征协同机制在医学图像分割任务中的有效性,为未来轻量化、高精度的医学图像分析模型设计提供了新的思路 and 方向。

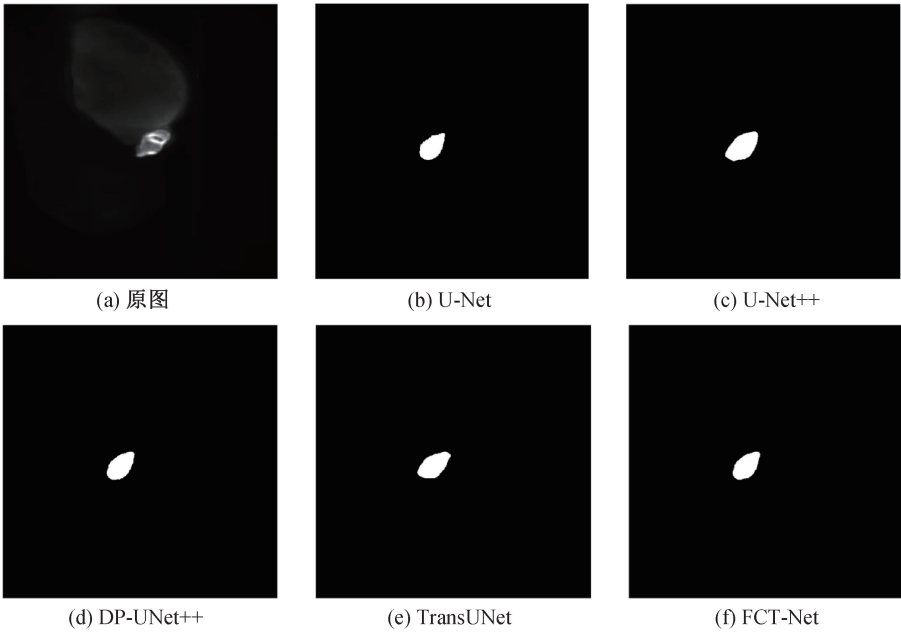


图 6 代表性模型与 FCT-Net 的图像分割结果示例图像

2.2.3 统计显著性检验

为了探究本文所提出的 FCT-Net 网络模型是否具有真实有效性,本文在数据集上进行了五折交叉验证,并以 Dice 系数为基础,开展统计显著性检验,计算对应的 P 值。 P 值是在假定原假设成立的情况下,观察到的样本数据或者更极端情况出现的概率。在本文中,原假设通常是 FCT-Net 网络模型的效果没有真实有效性,即模型的表现是由随机噪声等因素造成的。若 P 值小于 0.05,则说明在统计

学意义上,FCT-Net 的性能表现与噪声影响下的表现存在显著差异,表明模型性能的提升具有可靠性和稳定性,能够有效反映其在目标任务中的处理能力与优势;反之,若 P 值大于等于 0.05,则无法拒绝零假设,即不能排除模型性能提升是由随机因素引起的可能性,此时需进一步审查模型设计、数据特征或实验设置,以寻找更加有效的改进路径与验证方法,确保模型的实际有效性和可应用性。在本次检验中,选取经典基准模型 U-Net 与提出的 FCT-Net 进行对比分

析。其他对比模型已具备较强的应用验证基础,本文重点关注 U-Net 与 FCT-Net 之间的差异性。

U-Net 与 FCT-Net 的五折检验 Dice 系数如表 3 所示。由表 3 可见:FCT-Net 在各折验证中的 Dice 系数均显著高于 U-Net,进一步采用配对样本 t 检验对两个模型在五折交叉验证下的表现进行统计分析,计算得出的 P 值为 1.73×10^{-7} ,远小于显著性水平 0.05,表明 FCT-Net 相较于 U-Net 的性能提升具有较高的统计显著性。这进一步说明了 FCT-Net 的性能优势并非由偶然因素导致,而是体现了其在该任务中的有效性和鲁棒性。

表 3 五折检验的 Dice 系数 %

网络 模型	Dice 系数				
	1 折	2 折	3 折	4 折	5 折
U-Net	84.92	85.15	85.67	86.02	84.79
FCT-Net	90.35	90.52	90.89	91.14	89.87

3 结 语

本文提出了一种融合 CNN 与 Transformer 的斑马鱼心脏图像分割网络模型(FCT-Net),采用经典的编码器-解码器架构,构建了兼顾局部细节与全局特征的双分支并行架构,提升了心脏建模能力,增强对复杂背景和低对比度区域的适应性,且在动态成像伪影干扰下依然实现稳定且高精度的分割效果,较好地解决了斑马鱼心脏图像中因边缘轮廓模糊导致的分割精度不足问题。实验结果表明,与传统的 U-Net 模型相比,该模型在斑马鱼心脏图像分割任务中的 Dice 系数有较为明显的提升,局部边界分割准度得到改善,有效提高了心脏轮廓的分割精度。

尽管本文提出的模型分割精度有所提升,但模型参数量也相应增加。在对比度较低或特征模糊的区域,模型对心脏动态形变的适应性仍显不足,导致分割效果略有下降;同时模型在应对高动态变化场景下的鲁棒性仍有待加强。此外,当前图像标注过程仍依赖人工操作,训练集存在一定的主观性,影响了分割结果的客观性。未来研究可在优化数据标注流程的同时,降低模型计算复杂度,并探索更高效的特征融合方法,以进一步提升系统的实用性与可靠性。

参考文献:

[1] 马海钢,吴家辉,朱亚辉,等. 面向先进生物医学应用的光声显微成像术(特邀)[J]. 激光与光电子学进展,2024,61(6):115-144.
[2] 蔡海丽,张晓朦,刘亚迪,等. 药源性心脏毒性模型的构建与评价

进展[J]. 中国药物警戒,2024, 21(7):765-770.
[3] Castranova D, Samasa B, Venero Galanternik M, et al. Long-term imaging of living adult zebrafish[J]. Development, 2022, 149(4): dev199667.
[4] Hasani H, Sun J, Zhu S I, et al. Whole-brain imaging of freely-moving zebrafish[J]. Frontiers in Neuroscience, 2023, 17: 1127574.
[5] 李浩宇,曲丽颖,华子杰,等. 基于深度学习的荧光显微成像技术及应用[J]. 激光与光电子学进展,2021,58(18):185-204.
[6] Zhang B, Pas K E, Ijaseun T, et al. Automatic segmentation and cardiac mechanics analysis of evolving zebrafish using deep learning[J]. Frontiers in Cardiovascular Medicine, 2021, 8: 675291.
[7] Naderi A M, Bu H, Su J, et al. Deep learning-based framework for cardiac function assessment in embryonic zebrafish from heart beating videos[J]. Computers in Biology and Medicine, 2021, 135: 104565.
[8] Wang N, Dong G, Qiao R, et al. Bringing artificial Intelligence (AI) into environmental toxicology studies: A perspective of AI-enabled zebrafish high-throughput screening[J]. Environmental Science & Technology, 2024, 58(22): 9487-9499.
[9] Kim J, Yuk H, Choi B, et al. New machine learning-based automatic high-throughput video tracking system for assessing water toxicity using Daphnia Magna locomotory responses[J]. Scientific Reports, 2023, 13: 3530.
[10] Wang B, Sun Q, Liu Y, et al. Intelligent larval zebrafish phenotype recognition via attention mechanism for high-throughput screening[J]. Computers in Biology and Medicine, 2025, 188: 109892.
[11] Huang M H, Naderi A M, Zhu P, et al. Assessing cardiac functions of zebrafish from echocardiography using deep learning[J]. Information, 2023, 14(6): 341.
[12] Ni J, Mu W, Pan A, et al. Rethinking the encoder-decoder structure in medical image segmentation from releasing decoder structure[J]. Journal of Bionic Engineering, 2024, 21(3): 1511-1521.
[13] Micallef N, Seychell D, Bajada C J. Exploring the U-net++ model for automatic brain tumor segmentation[J]. IEEE Access, 2021, 9: 125523-125539.
[14] Zhou Z W, Rahman Siddiquee M M, Tajbakhsh N, et al. UNet++: A nested U-Net architecture for medical image segmentation[C]// Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer International Publishing, 2018: 3-11.
[15] Das N, Das S. Attention-UNet architectures with pretrained backbones for multi-class cardiac MR image segmentation[J]. Current Problems in Cardiology, 2024, 49(1): 102129.
[16] Yu J, Qin J, Xiang J, et al. Trans-UNeter: a new decoder of TransUNet for medical image segmentation[C]//2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2023: 2338-2341.