

一种混合CNN-Transformer多尺度特征的肝脏肿瘤分割网络

汤文超*, 丁德锐#, 袁浩成, 刘华卿

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2024年12月14日; 录用日期: 2025年1月5日; 发布日期: 2025年1月15日

摘要

针对肝脏肿瘤分割中肿瘤区域与肝脏正常组织界限模糊、大小和形状的可变性所导致的挑战, 对基于融合多尺度特征的双编码器分割算法进行了研究。提出一种基于双编码器融合多尺度特征的医学图像分割算法MCT-Net (Mix CNN-Transformer Multi-scale Feature Network)。首先, 在编码器阶段, 一方面使用细节特征提取模块DFEM (Detail Feature Extraction Module), 提取详细的肿瘤边缘局部特征; 另一方面引入Transformer编码器, 在保持肿瘤边缘细节分割良好的前提下, 获得了更大的感受野, 进一步提高肿瘤整体感知。其次, 在不同分支分别设计局部空间注意力模块LAM (Local Spatial attention Module)和全局空间注意力模块GAM (Global Spatial attention Module), 在降低计算复杂度的同时, 提高肿瘤特征的表达能力以充分学习可变肿瘤的局部边缘信息和全局语义信息。进而, 在Transformer编码器中添加多轴自注意力Max-SA (Multi-axis self-attention), 将完全密集的注意力机制分解为两个更轻量级的变体。在解码器阶段有效利用并行编码器的不同信息, 设计了多尺度信息交互模块MSIF (Multi-scale Information Fusion), 实现了信息在不同尺度之间的传递和补充, 从而提高边界的分割精度。最后, 所提的方法分别在公开数据集LiTS2017和3D-IRCADb进行试验评估测试和泛化实验, 所提方法在LiTS2017上评价指标Dice和ASD分别为72.16%和3.380 mm。

关键词

肝脏肿瘤分割, 多尺度信息交互, Transformer, 细节特征提取

A Mix CNN-Transformer Multi-Scale Feature Network for Liver Tumor Segmentation

Wenchao Tang*, Derui Ding#, Haocheng Yuan, Huaqing Liu

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

*第一作者。

#通讯作者。

Received: Dec. 14th, 2024; accepted: Jan. 5th, 2025; published: Jan. 15th, 2025

Abstract

Liver tumor segmentation is greatly challenging because of blurred boundaries between tumor areas and normal liver tissues, and variability in its sizes and shapes. To deal with such challenges, research on a multi-scale feature fusion-based medical image segmentation algorithm via double encoders and is proposed MCT-Net (Mix CNN-Transformer Multi-scale Feature Network). Firstly, in the encoder stage, on the one hand, the DFEM (Detail Feature Extraction Module) was used to extract the detailed local features of the tumor edge. Transformer encoder, introduced, on the other hand, in keeping the tumor edge detail segmentation under the premise of good, even more receptive field, further raise the overall awareness. Second, the Local space attention in different branches respectively designs LAM (Local Spatial attention Module) and GAM (Global Spatial attention Module), in reducing computational complexity at the same time, improve the expression ability of tumor features to fully learn the local edge information and global semantic information of variable tumors. Furthermore, Max-SA (Multi-axis self-attention) is added to the Transformer encoder to decompose the fully dense attention mechanism into two more lightweight variants. In effectively using the different Information of parallel encoder, decoder design the MSIF (Multi-scale Information Fusion), realizes the Information transfer and complementary between different scales, so as to improve the segmentation accuracy of the boundary. Finally, the proposed method, respectively, in the open LiTS2017 and 3D-IRCADb experiment assessment and generalization experiment, the proposed method on LiTS2017 Dice and ASD evaluation indexes are 72.16% and 3.380 mm respectively.

Keywords

Liver Tumor Segmentation, Multi-Scale Information Interaction, Transformer, Detail Feature Extraction

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

肝脏肿瘤是一种严重危害人类健康的疾病。随着社会经济的发展和生活方式的改变，肝脏肿瘤的发病率呈上升趋势。据世界卫生组织(World Health Organization, WHO)的统计数据显示，2020 年全球新增肝癌例达 90 余万例，并导致 83 余万人死亡[1]。特别是，在一些发展中国家和地区，由于病毒性肝炎、肝硬化以及不健康的生活方式等因素的影响，肝脏肿瘤的发病率甚至更为严重。早期的肝癌患者五年生存率可达 40%~70%，晚期肝癌患者生存时间只有 6 个月左右[2]。精准放疗需要医生把握放射治疗的靶区，即肿瘤区域，患者如果能在肝癌早期能发现肿瘤区域，为其后续的治疗提供引导，这可以有效地减少肝癌的重症率和死亡率。

目前，针对肝脏病变普遍采用的诊断方式是计算机层析成像(Computed Tomography, CT)。肝脏在 CT 图像上的表现具有高变异性，且肝脏组织与邻近器官和组织的对比度低以及肝脏肿瘤大小、形状、位置高度可变性等问题，使得在医学影像中准确识别和分割肝脏肿瘤成为一项具有挑战性的任务。自动分割方法可以自动定位肿瘤位置并分割出肿瘤边缘，从而辅助医生诊断，制定最为合适的治疗方案。研究高效的肝脏肿瘤自动分割方法对肝癌的临床诊断与治疗具有重要意义。

肝脏 CT 图像中的肿瘤通常具有边界模糊、大小各异、数目不固定等特点。肝脏肿瘤图像传统的分割方法主要包括区域生长法、阈值分割法、贝叶斯估计算法、支持向量机等[3]、Qi 等[4]提出了一种基于半自动贝叶斯的肝脏肿瘤分割方法。Liu 等[5]利用支持向量机对肝脏肿瘤进行粗分割，然后使用基于马尔可夫随机场的全方位可边模型细化分割结果。这类方法难以有效适应肝脏肿瘤的复杂性和多样性，且通常需要人工干预，无法实现肝脏肿瘤区域的自动分割。

近年来，卷积神经网络因其强大的学习能力、对复杂数据的高效处理以及优秀的泛化能力，被广泛运用于医学图像分割领域。2015 年 Ronneberger 等人[6]修改了全卷积网络 FCN，并提出了一种具有对称结构的 U-Net 模型，因为其结构简洁同时具有较好的分割性能。很多基于 U-Net 及其变体的模型与算法也逐渐用于肝脏肿瘤的自动分割任务。Schlemper 等[7]在 U-Net 框架的跳跃链接部分融入注意力机制，提出 Attention U-Net 模型，能够自动学习分割任务相关的区域信息，抑制无关特征，但同时易导致边界分割不准确，小肿瘤难以检出等问题。U-Net++ [8]是在 U-Net 跳跃连接处引入嵌套和密集的跳跃链接，通过叠加提取、融合不同层次的特征，减少编解码器之间的语义鸿沟。虽然这种网络在一定程度上可以增加小肿瘤的检测率，但它仍然面临着无法从多个尺度中提取足够信息、肝脏肿瘤分割精度低以及肿瘤分割困难等问题。Arulapan 等人[9]在不同卷积块之间使用具有不同膨胀率的膨胀卷积来获得局部尺度信息。然而由于缺乏上下文融合信息，在分割小体积肿瘤时受到限制。类似的网络还有 TransUNet [10]、H-DenseUNet [11]、RIU-Net [12]、MS-FANet [13]等。

尽管基于深度学习的肝脏肿瘤自动分割算法取得了一定成功但仍然存在局限性，其面临的挑战主要体现在如下两个方面。一方面，不同患者的肝脏肿瘤在大小、位置和形态等方面存在显著差异，这可能导致误检或漏检现象。另一方面，由于肝脏和邻近组织低对比度，肿瘤的边缘常常模糊不清。这对网络特征提取表征能力提出了更高的要求。Deng 等人[14]提出了一种名为 GAN-UNet 的网络。该网络结合了 U-Net 和生成对抗网络 GAN 的思想，在训练过程中 U-Net 生成的肝脏肿瘤区域图像被送入判别网络，然后借助于判别网络区分生成网络生成的伪标签和真实标签，这种动态的“博弈过程”可以有效提高分割精度，并且能够较为精确地分割出肝脏肿瘤边界。Wang 等人[15]提出的 CPAD-Net 和 Peng 等人[16]提出的 MDA-Net 通过对关联特征加以整理并选择性的聚合上下文切片来辅助分割过程。Zhang 等人[17]提出了一种基于空间聚合模块和不确定区域修复模块的多阶段肝脏肿瘤分割方法。该方法在编码器中利用不同阶段之间的像素级信息进行交互，并利用相邻的具有区别性特征来细化不确定像素，改善肝脏肿瘤过分割和欠分割的情况。然而，利用肝脏 CT 图像数据，深度神经网络算法如何灵活利用肝脏多层级特征，以及全局局部信息交互实现肿瘤分割精度的提升仍是一项重大挑战。

为此，本文提出 MCT-Net (Mix CNN-Transformer Multi-scale Feature Network)，其主要工作如下所述。

- 1) 本文提出一种基于双编码器融合多尺度特征的分割网络 MCT-Net，实现对肝脏肿瘤的自动分割。通过融合卷积和 Transformer 网络的优势，设计了细节特征提取模块 DFEM (Detail Feature Extraction Module)，捕获了丰富的全局上下文信息，充分学习了肿瘤的局部边缘信息和全局语义信息，在兼顾边缘细节分割的同时提高对肿瘤的整体感知；
- 2) 设计了局部空间注意力模块 LAM (Local Spatial Attention Module) 和全局空间注意力模块 GAM (Global Spatial Attention Module)。LAM 模块通过增强卷积编码器对细节和局部特征的捕捉能力，提升了对微小且关键肿瘤特征的解析度；GAM 模块不仅扩展了感受野，还提升了对复杂形态和大范围变异肿瘤的整体感知和精确度，从而改善分割精度；
- 3) 设计了多尺度信息交互模块 MSIF (Multi-Scale Information interaction module)。该模块实现了对高分辨率和低分辨率特征之间的特征差异性的交流和互补，加强了全局信息与局部信息之间的联系；
- 4) 在 Transformer 编码器中添加多轴自注意力 Max-SA (Multi-axis Self-Attention)，从而将完全密集的

注意力机制分解为两个更轻量级的变体 Block attention 和 Grid attention，不仅提升了模型在处理高维数据时的计算效率，还使得模型能够更灵活地聚焦于不同特征轴上的关键信息。

2. 本文方法

2.1. MCT-Net 整体结构

本文提出的肝脏肿瘤分割网络 MCT-Net 是由编码器、解码器两大核心组件构成，如图 1 所示。编码器采用了创新的双分支特征提取结构，结合了卷积神经网络 CNN 和 Transformer 架构的优势，能够更深入的挖掘图像特征。其中，卷积编码器的设计采用了多个深度可分离卷积的 DFEM 模块，这些模块嵌入了局部空间注意力机制 LAM。通过使用 3×3 和 5×5 的卷积核，编码器在保持高性能的同时降低了计算复杂度和参数量。这种设计不仅优化了模型效率，而且为后续的特征融合提供了精细的肿瘤轮廓信息。Transformer 编码器则由多个嵌入全局空间注意力机制 GAM 的 DFEM 模块组成，随后通过 Block attention 和 Grid attention 两个轻量级的注意力模块，有效地扩展了模型的感受野，提升了对复杂形态和大范围变异肿瘤的整体感知和精确度。

解码器部分由多尺度信息融合模块 MSIF 和基础卷积模块构成。MSIF 模块由三个并行分支构成，每个分支通过不同内核的 Maxpooling 和卷积操作调整特征图大小，实现高分辨率和低分辨率特征之间的交流和互补。与传统的跳跃链接结构相比，这种设计能够更有效地捕捉肿瘤边缘的细微变化和局部结构。融合后的特征图通过 3×3 和 5×5 的自适应核注意力 SK-attention 模块进行处理，动态调整不同感受野下的特征响应，确保在多尺度特征融合过程中精确地描述肿瘤与周围组织的边界。最终，通过上采样处理双编码器融合得到的特征图，并与 MSIF 模块输出的特征图进行拼接和逐层传递，形成网络的最终输出，为肝脏肿瘤的精确分割提供了强有力的支持。

具体地，卷积编码器分支(下标记为 C)输入头的输出特征 $E_{C,1}$ 由输入图像 X 首先经过步长为 2 的 Conv 3×3 , BatchNorm 和 Relu 操作，为了表达简洁，这一连续操作记为 CBR(\cdot)，随后经过三个连续嵌入 LAM 局部注意力的 DFEM 模块获得。接下来，编码器的其余阶段的输出记为：

$$E_{C,i+1} = \text{DFEM}_{\times n}(E_{C,i}) \quad (1)$$

Transformer 编码器分支(下标记为 T)输入头的输出特征 $E_{T,1}$ 由输入图像 I 首先经过连续两次的 CBR(\cdot) 操作得到。随后，编码器其余阶段是经过六个连续 DGB (Detail Global Attention Module) 模块获得，其输出记为：

$$E_{T,i+1} = \text{DGB}_{\times n}(E_{T,i}) \quad (2)$$

接下来，一方面将卷积编码器提取的相邻三个阶段特征输入到 MSIF 模块，得到融合之后的特征图。另一方面将来自双编码器获得的特征图进行上采样，进而与来自 MSIF 模块的输出拼接。具体地，首先将获得的相邻特征通过 MSIF 模块进行高低分辨率的交流，得到边缘细节更丰富的特征，记为 E_{fuse} ：

$$E_{\text{fuse}} = \text{MSIF}(E_{T,i-1}, E_{T,i}, E_{T,i+1}) \quad (3)$$

同时将双编码器得到的 $E_{T,4}$ 、 $E_{C,4}$ 的特征拼接在一起：

$$E'_4 = \text{CBR}(E_{T,4}) \text{ ① } E_{C,4} \quad (4)$$

其中 $E_{C,4}$ 、 $E_{T,4}$ 分别代表卷积编码器和 Transformer 编码器第四阶段特征，①表示将输入特征沿着通道维度拼接。进而，将所得特征 E'_4 和 E_{fuse} 拼接在一起，经过 Self-Attention 上采样等操作得到特征 E''_4 。随后将得到的特征 E''_4 经过 Transpose 转置操作与来自 MSIF 模块的特征 E_{fuse} 在通道上进行拼接，得到解码器

D_4 阶段的输出:

$$D_4 = \text{CBR}(\text{Transpose}(E_4'') \odot E_{\text{fuse}}) \quad (5)$$

随后, 利用得到的解码器 D_4 阶段的特征和 MSIF 模块所获得的特征, 重复操作逐阶段进行拼接并逐层传递, 最终输出肿瘤分割图。

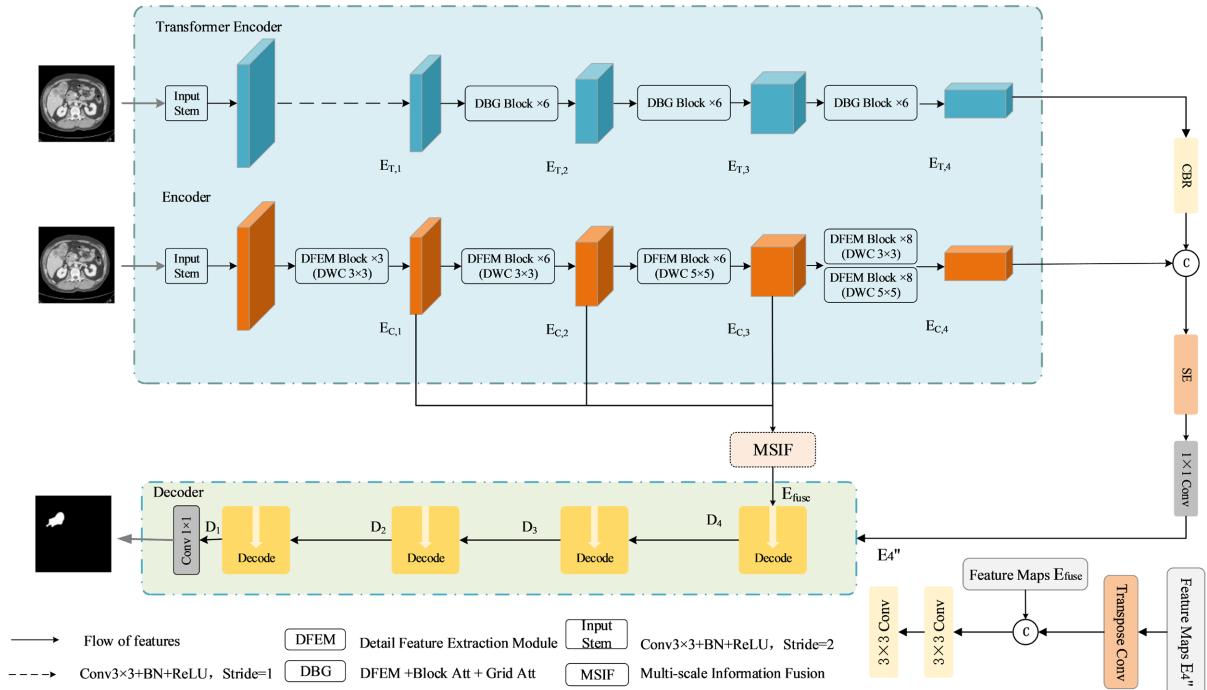


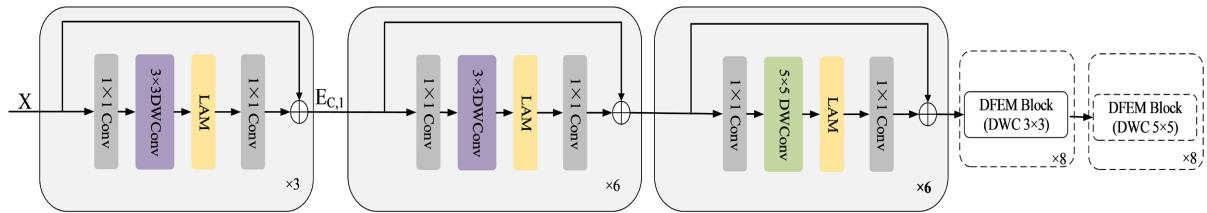
Figure 1. Overall network structure of MCT-Net

图 1. MCT-Net 整体网络结构

2.2. 双编码模块

从腹部 CT 中分割出肝脏肿瘤是一个复杂的医学图像分割任务, 其模糊的边界、不规则的形状、大小, 都要求网络有着优秀的特征提取能力。有效地结合局部特征信息和全局信息的长距离依赖建模, 这是区分病变位置和背景像素的关键。卷积神经网络(CNN)的卷积操作受限于其固有的局部感受野, 限制了在有限像素区域内建立特征关联的能力, 难以捕捉图像的全局信息。这种局限性在处理组织边缘模糊的情况下容易导致误差, 影响分割精度。为了克服这一局限, 解决 CNN 在建立长距离依赖方面的不足, 本文在编码器阶段引入了 Transformer 编码器, 以利用其对长距离依赖的建模能力。进而, 结合卷积编码器的优势, 实现局部和全局特征的全面捕捉。最近的研究表明, 使用相对自注意力[18]-[21]可以有效地代替标准的自注意力[22] [23], 它通过引入注意力权重的相对学习偏差, 在保持较快的推理速度的同时, 在多个视觉任务上展现出优越的性能[18]-[20]。受此启发, 本文在 Transformer 编码分支中引入了多轴自注意力机制, 以进一步提升特征提取的灵活性和效率。

MCT-Net 总体编码器是由卷积编码器和 Transformer 编码器两大分支构成, 其中卷积编码器是由若干个嵌入 LAM 注意力机制的 DFEM 模块构成。DFEM 作为编码器的基本模块, 主要利用 3×3 和 5×5 的深度可分离卷积提取特征。这种设计保持模型性能的同时, 降低计算复杂度并减少参数量。此外, 它还为后面阶段各个特征融合提供更加精细的肿瘤轮廓信息, 如图 2 所示。

**Figure 2.** Feature extraction module**图 2.** 特征提取模块

在卷积特征提取模块，现以 $E_{C,1}$ 阶段为例，具体来说，对于输入 X ，首先使用 1×1 的卷积将输入拓展到更高维的空间：

$$X' = \text{Conv}_{1 \times 1}(X) \quad (6)$$

为了进一步缩减参数量，利用 3×3 的 Depth wise Convolution，单独对每个输入通道进行操作：

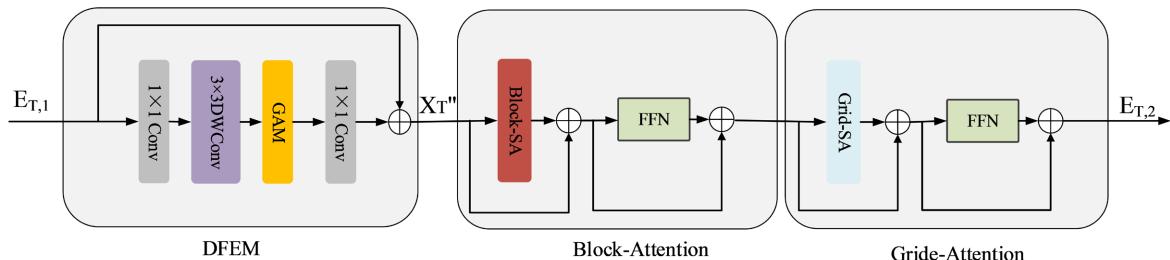
$$X'_C = \text{DWConv}_{3 \times 3}(X') \quad (7)$$

其后添加 Local Spatial Attention，能够对通道特征响应进行自适应重新加权，用 1×1 卷积将特征“压缩”到更低维的空间中，从而仅保留基本特征。为了帮助快速收敛并防止梯度消失问题，使用了类似于 ResNet 的残差连接，基本模块表示如下：

$$E_{C,1} = \text{Conv}_{1 \times 1} \left(\text{LAM} \left(\text{DWConv}_{3 \times 3} \left(\text{Conv}_{1 \times 1}(X) \right) \right) \right) + X \quad (8)$$

在 $E_{C,1}$ 、 $E_{C,2}$ 、 $E_{C,3}$ 阶段，特征提取模块分别经过 3 次、6 次、6 次的堆叠，其中深度可分离卷积分别利用的是 3×3 、 5×5 、 3×3 。 $E_{C,4}$ 阶段较为特殊， $\text{DWConv}_{3 \times 3}$ 特征提取模块和 $\text{DWConv}_{5 \times 5}$ 特征提取模块依次重复 8 次。虽然该编码模块已经具备很强的特征提取能力，但其内在的感受野限制使其对肿瘤灵活边缘特征利用略显不足。

近年来，Transformers 在计算机视觉任务中大放异彩。值得一提的是，Transformer 分支凭借较大的感受野，通过能提取更加全局的特征表达，本文在其中添加了多轴自注意力 Max-SA，总体模块架构如图 3 所示。本质上，它将完全密集的注意力机制分解为两个更轻量级的变体，分别是 Block attention 和 Grid attention。这种重组有效地减轻了与普通注意力相关的二次复杂性，同时保留了关键的非局部性方面，通过空间轴分解将全尺寸注意力分解为局部和全局形式。减少参数量的同时，充分利用局部和全局信息，有效提高模型的表达能力。

**Figure 3.** Overall architecture of Transformer encoder**图 3.** Transformer 编码器模块架构

Transformer 编码器是由若干个嵌入 GAM 注意力机制的 DFEM 模块、Block-SA 和 Grid-SA 构成。现以 $E_{T,2}$ 阶段为例。具体来说，对于输入 $E_{T,1}$ 首先经过一个 $\text{Conv}_{1 \times 1}$ 和 $\text{DWConv}_{3 \times 3}$ ：

$$X'_T = \text{Conv}_{1 \times 1} \text{DWConv}_{3 \times 3}(E_{T,1}) \quad (9)$$

随后特征 X'_T 被传入 GAM 模块，随后经过 1×1 卷积，得到特征 X''_T

$$X''_T = \text{Conv}_{1 \times 1}(\text{GAM}(X'_T)) \quad (10)$$

对于输入特征图 $X''_T \in R^{H \times W \times C}$ ，本文将其划分形状为 $\left(\frac{H}{P} \times \frac{W}{P}, P \times P, C\right)$ 大小的张量。多轴自注意力结构如图 4 所示。划分操作产生不重叠的块，每个块大小是 $P \times P$ 。在这个局部空间维度上使用自我注意力，对于块就相当于在一个有限的窗口内聚焦。这个过程通俗地说就是“块注意力”，有助于促进本地互动。这种局部注意力模型避免了完全自我注意力的大量计算，受到块注意力的启发，本文在其之后引入网格注意力，其避免了固定窗口大小的划分，而是将张量网格化为具有固定 $G \times G$ 均匀网格尺寸的形状 $\left(\frac{H}{G} \times \frac{W}{G}, G \times G, C\right)$ 并且在细分的网格轴上运用自注意力，相当于空间域中的标记的扩张的全局融合。保持一致的窗口和网格尺寸(其中 $P = G = 7$)协调了局部和全局过程之间的计算负荷。在之后通过残差连接与 X''_T 融合得到特征 X''_T ，并送入 Block attention 和 Grid attention 中得到 X_{At} ，最后在堆叠上述操作一次得到阶段输出结果 $E_{T,2}$ ：

$$\begin{cases} X''_T = X''_T + X'_T \\ X_{At} = G(B(X''_T)) \\ E_{T,2} = G(B(DF(X_{At}))) \end{cases} \quad (11)$$

其中， $B(\cdot)$ 表示块注意力操作， $G(\cdot)$ 表示网格注意力操作， $DF(\cdot)$ 表示特征提取模块。

本文所提出 Transformer 分支结合了 DEFM、块注意力和网格注意力模块，使网络能够从浅层到深层捕获局部和全局特征。特征提取模块和 Attention 的组合使用提高了网络的泛化能力和可训练性。另一个好处是将 DEFM 模块层放在注意力之前，其中深度卷积的功能是条件位置编码(CPE)，消除了模型中对单独位置编码层的要求。

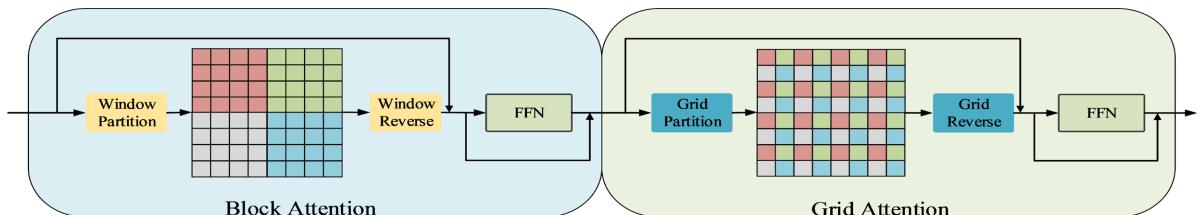


Figure 4. The structure of Multi-axis self-attention
图 4. Max-SA 总体结构

2.3. 多尺度特征信息交互模块

针对肝脏肿瘤分割任务中存在的尺度变化和形态多样性等挑战，传统的单一尺度特征提取方法往往难以充分捕捉到图像的全局信息，并且容易导致信息丢失或不足的问题。因此，为了克服这些挑战并提高分割模型的性能，本文提出了一种多尺度特征融合模块 MSIF。该模块旨在通过有效整合不同尺度的特征信息，提高模型对不同尺度和形态变化的适应能力，加强在特征提取过程中对于目标区域的特征表达，从而改善肝脏肿瘤分割的精度和鲁棒性。以三个不同阶段的特征 $E_{C,1}$ 、 $E_{C,2}$ 、 $E_{C,3}$ 为例，MSIF 模块通过并行处理和特征融合，实现了对这些特征的有效整合，如图 5 所示。

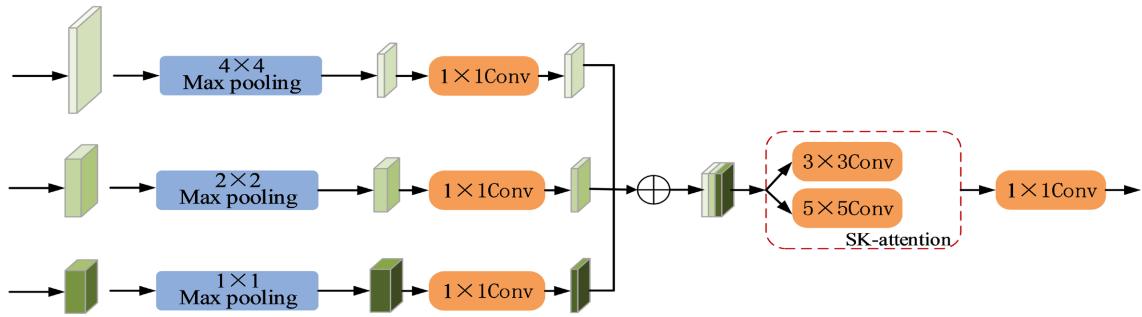


Figure 5. Multi-scale information interaction module
图 5. 多尺度特征信息交互模块

利用多阶段特征和注意力，有利于多层特征的交换和互补，在保持分辨率的同时改进信息流，进而充分融合编解码特征。具体来说，不同阶段的特征具有不同的分辨率和通道，因此在特征增强之前，本文使用多分支特征融合层通过适当地设置通道、池化内核大小和多分支特征的分支数量来实现的将不同的特征映射统一到相同的通道和分辨率。具体来说，利用不同阶段 $E_{C,i}$ ($i \in \{1, 2, 3\}$) 输出的特征。首先，对 $E_{C,1}$ 经过下采样 Maxpooling 将其统一成 $E_{C,3}$ 一样大小，在经过卷积操作统一不同分支特征通道，对特征 $E_{C,2}$ 、 $E_{C,3}$ 操作类似：

$$\begin{cases} M_1 = W_1(K_4(E_{C,1})) \\ M_2 = W_2(K_2(E_{C,2})) \\ M_3 = W_3(K_1(E_{C,3})) \end{cases} \quad (12)$$

进而有

$$U = M_1 \oplus M_2 \oplus M_3 \quad (13)$$

其中， K_1 、 K_2 、 K_4 代表的是 Kernel size 为 1×1 、 2×2 、 4×4 的 Max pooling 操作， W_1 、 W_2 、 W_3 代表的是 3×3 的卷积， \oplus 代表的是拼接操作。

得到的特征 U ，通过两个不同尺度的卷积核进行卷积操作，得到形状同为 $[C, H, W]$ 两个特征图 U_1 、 U_2 ；然后使用拼接操作将不同尺度卷积后的特征图进行融合，对融合后的特征图进行全局平均池化，得到每个通道的全局信息；进而通过两个全连接层和一个 Soft max 层生成两个不同尺度特征图的权重，将生成的权重应用到对应的特征图上，重新加权后的特征图进行加法融合，得到最终的输出特征图 E_{fuse} 。

2.4. 局部全局注意力模块

注意力机制在深度学习中扮演着关键角色，通过动态调整模型对输入数据的关注程度，提高了模型对重要信息的感知能力。面对肝脏 CT 图像分割任务的复杂性、肿瘤的不均匀性和不规则性，以及局部细节与全局一致性的挑战，本文设计了用于卷积编码器的局部空间注意力 LAM 模块和用于辅助分支编码器的全局空间注意力 GAM 模块，旨在强化目标相关信息的同时，捕获更为丰富的全局和局部空间特征。

在肝脏肿瘤分割任务中，由于肝脏肿瘤边缘常常模糊不清，局部注意力机制能够帮助模型更加关注这些模糊的区域。通过引入局部注意力机制，模型可以更有效地处理这些模糊边缘同时，肿瘤的形态和位置往往受到周围组织的影响，因此全局上下文信息对于准确分割肿瘤非常重要。同时结合全局上下文信息，提高对肿瘤的准确性和稳定性。LAM 模块和 GMA 模块如图 6 和图 7 所示。

卷积编码器上引入局部空间注意力模块 LAM 能够在不同的区域之间实现动态的加权，从而更好地

捕捉到具有重要意义的特征。降低对无关区域的关注的同时增强对肿瘤区域的表征能力，提高了分割的准确性和稳定性。在辅助编码器分支利用全局空间注意力 GAM 模块，全局注意力能够帮助模型有效地捕捉到这些全局上下文信息，这种全局性的关注能够帮助模型更好地理解肝脏肿瘤的整体形态和分布规律，从而提高了对肿瘤区域的感知和理解能力。

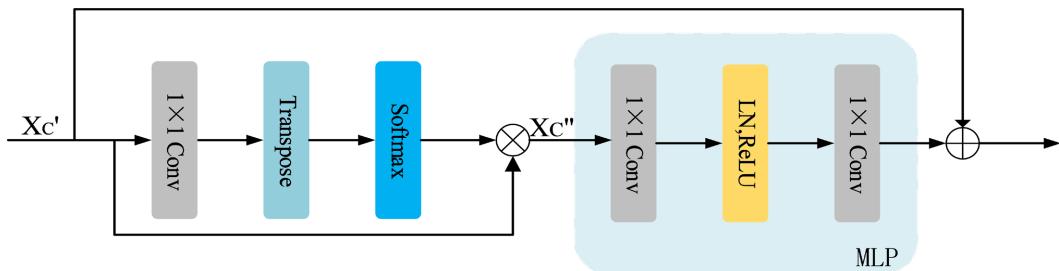


Figure 6. Local spatial attention module
图 6. 局部空间注意力模块

现以 $E_{C,1}$ 阶段为例，在特征提取模块 DFEM 中，经过 3×3 的深度可分离卷积得到的特征 X'_C ，之后经过局部注意力模块 LAM，可以更好地聚焦小肿瘤。具体来说，首先经过级联的 1×1 的卷积和 3×3 的深度卷积，在降维的同时减少计算量以提取更高级别的特征表示：

$$X'_C = DW_{3 \times 3} C_{1 \times 1}(X) \quad (14)$$

接下来，引入残差连接，将经过 1×1 和 3×3 卷积之后的特征与原始输入相加，经过 1×1 卷积后，再经过 Sigmoid 激活函数得到注意力权重图，逐点相乘对输入特征进行加权，加权后的特征与原始输入相加：

$$\begin{cases} X'_C = X'_C + X \\ X''_C = \text{Sigmoid } C_{1 \times 1}(X'_C) \\ Y_{\text{AtL}} = X''_C \odot X + X \end{cases} \quad (15)$$

总体局部注意力 LAM 模块以及输出 Y_{AtL} 可以被概括为：

$$\text{Att}_L(X) = \sigma(C_{1 \times 1}(X)) + X \quad (16)$$

$$Y_{\text{AtL}} = \text{Att}_L(X) \odot X + X \quad (17)$$

其中， $\text{Att}_L(\cdot)$ 表示局部注意力模块 LAM 操作， Y_{AtL} 表示经过局部注意力的输出， $C_{1 \times 1}$ 表示 1×1 的卷积， $\sigma(\cdot)$ 表示 Sigmoid 激活函数， \odot 表示逐点相乘操作。

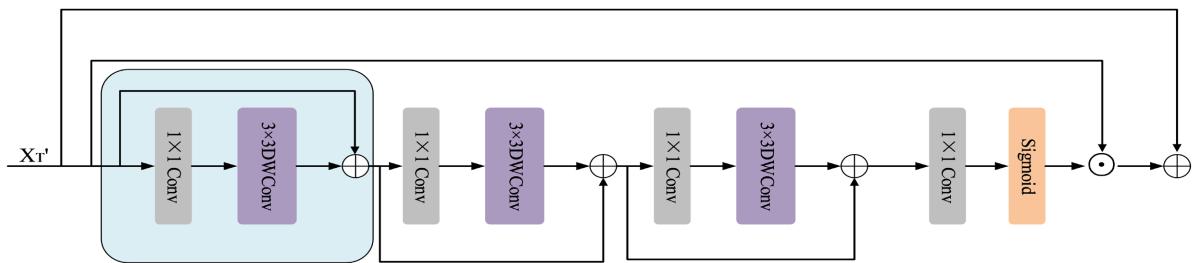


Figure 7. Global spatial attention module
图 7. 全局空间注意力模块

在 Transformer 分支编码器中, 以 $E_{T,2}$ 阶段为例, 输入特征首先经过 1×1 的卷积以及 3×3 的深度可分离卷积得到特征 X'_T , 随后经过全局注意力 GAM 模块, 能更有效捕捉全局上下文信息, 提高模型对肿瘤区域的感知和理解能力。具体来说, 首先经过一个 1×1 的卷积降维, 随后特征经过 Transpose 转置操作, 将特征矩阵的维度进行变换, 以便于后续的注意力计算。

$$X''_T = \text{Transpose } C_{1 \times 1}(X'_T) \quad (18)$$

经过 Softmax 激活函数为每个特征分配权重之后, 接下来经过由两个 1×1 的卷积、BatchNorm、Relu 激活函数构成的 MLP, 其中第一层卷积是以 $r = 2$ 将其投射到更高维空间, 它将特征图从原来的通道数 C 投射到一个更高维的 $2C$, 增强特征表达能力。第二层利用卷积将维度恢复为与输入相同。

$$\begin{cases} X''_T = \text{Softmax}(\text{Transpose}(C_{1 \times 1}(X'_T))) \\ X'''_T = X''_T \otimes X'_T \\ Y_{\text{AtG}} = \text{MLP}(X'''_T) + X'_T \end{cases} \quad (19)$$

整个全局注意力 GAM 模块以及输出 Y_{AtG} 可以被概括为:

$$\text{Att}_G(X'_T) = \text{Softmax}(\text{Transpose}(C_{1 \times 1}(X'_T))) \quad (20)$$

$$Y_{\text{AtG}} = \text{MLP}(\text{Att}_G(X'_T) \otimes X'_T) + X'_T \quad (21)$$

其中, $\text{Att}_G(\cdot)$ 表示全局注意力模块 GAM 操作, \otimes 表示矩阵乘法, MLP 表示 1×1 Conv + BN Relu + 1×1 Conv, Y_{AtG} 表示经过全局注意力的输出。

3. 实验及其结果分析

3.1. 数据集及预处理

本文公开在数据集 LiTS2017 上验证所提出算法的灵活性和有效性, 并且在 3D-IRCADb 进行泛化性能的测试。在训练模型之前, 需要对数据进行预处理。以下是关于数据集和预处理的简要介绍。

采用数据集是肝脏肿瘤分割挑战赛提供的公开数据集 LiTS [24]。其中肝脏和肝脏肿瘤区域是由 4 名经验丰富的放射科医生使用 ITK-ANAP 软件手动标注, 作为分割的金标准。按照 8:1:1 将 131 病例数据随机划分为训练集、验证集和测试集。

训练前将图像及标签统一重采样为 512×512 大小, 然后通过调整窗宽窗位将所有 CT 图像的强度值限制在 $[-200, 200]$ Hounsfield Unit (HU) 范围内, 以增强 CT 图像的对比度, 减少无关组织或器官对肝脏肿瘤分割结果的影响。随后, 对图像进行自适应直方图均衡化, 并将灰度值映射到 $[0, 1]$ 范围内, 以增强肝脏和肿瘤之间的对比度, 从而提高 CT 图像的可视化效果。考虑到实际 CT 采样时人体差异或者站位不同, 对图像随机进行 0° 至 10° 旋转, 预处理前后对比如图 8 所示。

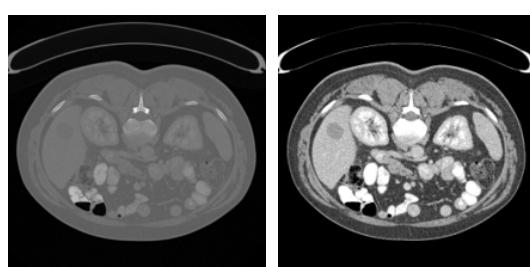


Figure 8. Comparison before and after pre-processing of liver CT images
图 8. 肝脏 CT 图像预处理前后对比

通过一系列数据预处理,如裁剪、调整窗宽和窗位、自适应直方图均衡化以及归一化等操作,可以观察到经过预处理的肝脏肿瘤区域与周围器官的对比更加明显,其轮廓也更加清晰。

3.2. 实验环境及参数设置

本次实验在 Ubuntu 操作系统上进行,使用 PyTorch 3.8.0 深度学习框架构建模型。实验中利用两张 NVIDIA GeForce RTX 2080 显卡进行并行计算。本文算法具体的实验环境与所用设备如表 1 所示。

Table 1. Experimental environment and required apparatus

表 1. 实验环境与所用设备

名称	参数
操作系统	Ubuntu18.04
CPU	Intel(R) Xeon(R) CPU E5-2640
GPU	NVIDIA GeForce RTX 2080
内存	32GB
Cuda 版本	12.2
深度学习框架	PyTorch
Python 版本	3.8.0

在网络训练过程中,本文将输入图像的尺寸设置为 512×512 ,并对网络参数进行了随机初始化。同时采用了 Adam 优化器来更新参数,并将批量大小设置为 8。在训练过程中,将 epoch 设置为 150。初始学习率为 0.001,当连续 5 个 epoch 内验证集的平均损失不再降低时,学习率衰减为原来的 0.1 倍。

3.3. 评价指标

为了定量评价网络分割性能,本文采用的模型性能评价指标有 Dice 系数(Dice coefficient, Dice) [25]、体积重叠误差(Volumetric Overlap Error, VOE) [26]、相对体积差(Relative Volume Difference, RVD) [27]、平均表面距离(Average surface Distance, ASD) [27]。若网络实际的分割结果为 A,真实的标签为 B,则各评价指标的计算公式如下。

1) Dice 系数是医学图像分割领域使用频率最高的评价指标,用于计算网络分割结果和真实标签之间的相似性,Dice 的取值范围为[0, 1]。其值越大,分割效果越好。其数学表达式为:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (22)$$

2) VOE 用于计算网络的分割结果和真实标签体素之间的体积重叠误差值越接近于零,分割效果越好,其公式为:

$$\text{VOE} = 1 - \frac{|A \cap B|}{|A \cup B|} \quad (23)$$

3) RVD 计算网络的分割结果和真实标签的体素之间的相对体积重叠误差,值越接近 0,表示分割结果与真值的体积差异越小分割效果越好,定义为:

$$\text{RVD} = \frac{|B| - |A|}{|A|} \quad (24)$$

4) 平均表面距离 ASD 也是用来衡量两个集合的距离,比较预测标签和真实标的表面相似度。公式

表达如下：

$$ASD(A, B) = \frac{1}{|S(A)| + |S(B)|} \times \left(\sum_{p \in S(A)} d(P, S(B)) + \sum_{q \in S(B)} d(q, S(A)) \right) \quad (25)$$

其中， $S(A)$ 是网络分割结果的边界点集， $S(B)$ 是真实标签的边界点集， $d(P, S(\cdot))$ 表示边界点 P 到 $S(\cdot)$ 的欧氏距离。

综合上述四个指标，可以较为完善的评估所提方法在整体区域以及边界的分割精确度，其中 Dice 和 VOE 数值在 0~1 之间，值越大表示整体区域分割准确度越高，而 RVD 和 ASD，值越小表明边界分割效果越好。

3.4. 实验结果及分析

为了验证本文提出的网络对肝脏肿瘤分割的性能，将本文所提出的方法和 U-Net [6]、Attention U-Net [7]、U-Net++ [8]、TransUNet [10]、CPAD-Net [15]、TD-Net [28]、MS-FANet [13]的分割性能进行比较。对比试验均在本文相同的实验环境和训练参数设置下进行，损失函数采用交叉熵 BCE 损失和 Dice 损失之和，以上不同的算法在 LiTS2017 训练集上训练，并在 LiTS2017 测试集上进行性能测试。对比结果如表 2 所示。

由表 2 结果可知，本文所提出的算法在 Dice、VOE、RVD 和 ASD 四种评价指标上分别达到了 72.16%、40.56、0.078 和 3.380 其中 Dice、RVD 和 ASD 均优于其他算法。相较于 MS-FANet，本文在 Dice 上提高了 1.13%，平均表面距离 ASD 达到 3.380 mm。

Table 2. Comparison results of different models on LiTS2017

表 2. 不同算法在 LiTS2017 数据集分割结果

方法	Dice	VOE	RVD	ASD
U-Net	64.83	47.45	0.204	12.386
Attention U-Net	68.95	42.33	0.165	5.912
U-Net++	70.34	40.67	-0.137	4.425
TransUNet	70.76	40.24	0.146	4.141
CPAD-Net	69.85	41.55	0.155	5.347
TD-Net	70.48	40.92	0.121	4.259
MS-FANet	71.23	39.73	-0.107	3.862
MCT-Net(ours)	72.16	40.56	0.078	3.380

本文针对性提出的细节特征提取模块 DFEM、多尺度信息交互 MSIF 模块以及 GAM 和 LAM 注意力模块，在一定程度上提升了对肿瘤的分割性能，综合上述的评价指标来看，说明本文所提的分割模型有利于肝脏肿瘤 CT 图像分割。

为了更加直观的说明本文提出的分割模型的有效性，选取 6 个不同且具有代表性的 CT 切片使用不同模型进行预测，对得到的预测结果进行可视化分析。如图 9 所示，不难看出切片总体的肝脏肿瘤区域边界较为清晰，由于这些器官与肝脏在图像中的密度相似，总的来说，切片 A 中，肝脏与胃、胰腺、肠道这些器官密度相似，并且它们的边界可能与肝脏肿瘤边界重叠，因此分割算法可能会将它们错误地识别为肿瘤的一部分。在切片 B 中，同样受到胃边界的影响，一些算法的分割结果出现了断层现象同时边缘分割细节不够。而切片 C 的肝脏肿瘤数量众多且小，且因其和周围组织纹理相似对比度低，大部分

算法难以捕捉微小肿瘤具体位置造成漏检误检，在切片 D 和 F 中，因肝脏和相邻器官存在粘连覆盖导致算法忽略了胃部附近重叠部分的肝脏肿瘤分割。在切片 E 中，肝脏区域的左下角存在下腔静脉，且正位于肝脏边缘上，因此干扰了模型的识别判断，因此预测结果对该部分的缺失分割。相较之下，本文提出的算法在这些分割难度较大的 CT 图像中仍表现出较强的分割性能。

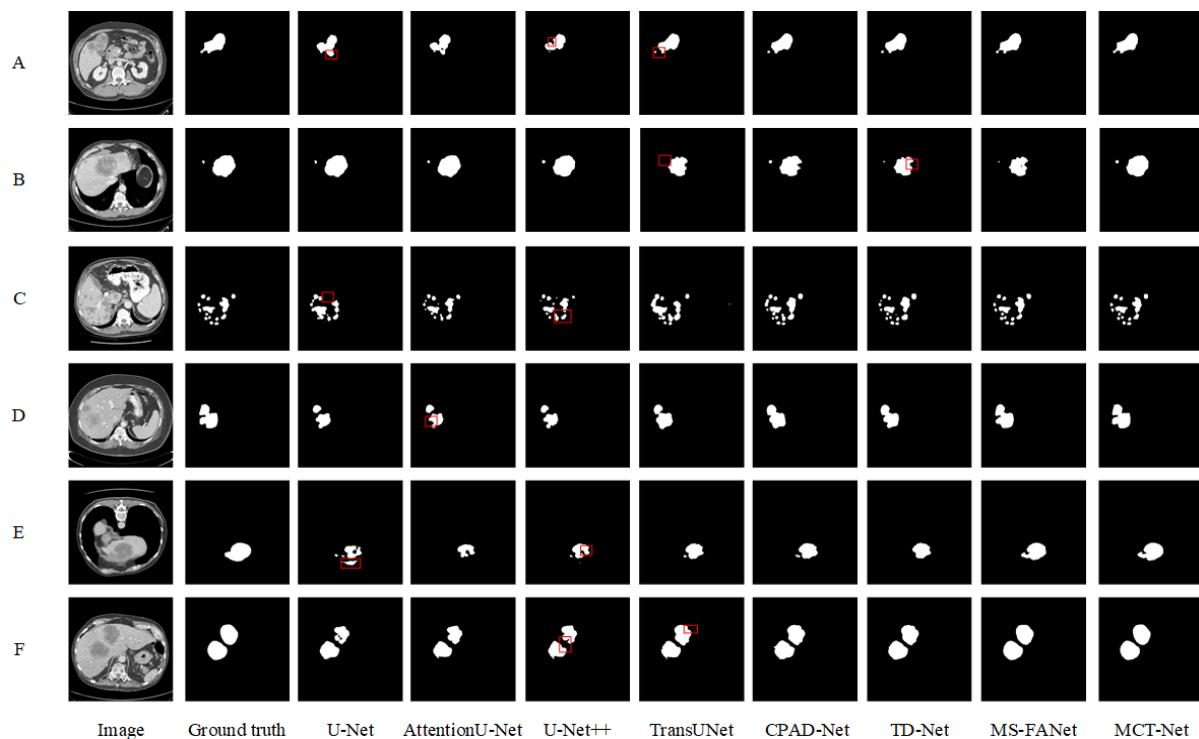


Figure 9. Visualization of segmentation results of different models on LiTS2017

图 9. 不同模型在 LiTS2017 上分割结果可视化

此外，为了进一步验证本文提出模型的泛化能力，本文使用了 3D-IRCADb 数据集进行了不同模型的进一步测试。值得注意的是，3D-IRCADb 数据集包含 20 个患者的腹部 CT 扫描图像，具有高分辨率和详细的器官分割标签。其数据的图像质量、分辨率和患者特征的多样性，为模型的泛化能力提供了严峻的考验，泛化能力测试结果如表 3 所示。

Table 3. Comparison results of different models on 3D-IRCADb

表 3. 不同方法在 3D-IRCADb 数据集上的对比结果

方法	Dice	VOE	RVD	ASD
U-Net	61.22	50.25	0.233	14.249
Attention U-Net	63.95	48.63	0.195	7.812
U-Net++	65.2	47.58	0.172	6.946
TransUNet	66.16	46.74	0.161	6.413
CPAD-Net	65.73	47.06	0.167	7.204
TD-Net	67.51	46.12	0.152	6.126
MS-FANet	68.16	45.35	0.136	5.591
MCT-Net (本文)	70.36	43.87	0.118	4.856

本文所提出的分割方法四种指标均获得了最优的评估结果。在评价指标 Dice 和平均表面距离 ASD 中, 本方法分别达到了 70.36% 和 4.856 mm, 较其他模型有显著提升。可见本文提出的算法具有良好的泛化性能, 尤其是在处理肿瘤模糊边界时, 能够产生更加精确的分割结果。

3.5. 消融实验

为验证本文提出的各个模块对整体分割性能的影响, 在 LiTS2017 数据集上进行了几种不同的消融实验: 分别是去除细节特征提取模块(DFEM)、去除 GAM 和 LAM 两类注意力模块、去除多尺度信息交互模块(MSIF)以及去除边 Transformer 编码分支。

实验结果如表 4 所示。可以看到当模型中引入 DEEM 模块 Dice 上升 1.14%, 引入两类注意力模块 Dice 上升 1.19%, 引入 MSIF 模块 Dice 上升 1.98%, 引入 Transformer 编码分支 Dice 上升 1.51%, 需要注意的是, 尽管引入两类注意力模块在 Dice 系数上的提升相对较小, 但从平均表面距离 ASD 这一边界分割准确度指标的结果可以明显看出, 该模块在提升肿瘤边界分割精度方面起到了显著作用。

Table 4. Ablation research of each module on LiTS2017

表 4. 各个模块在 LiTS2017 数据集上的消融研究

方法	Dice	VOE	RVD	ASD
w/o DFEM	71.02	41.78	0.084	3.674
w/o LMA and GMA	70.97	42.12	0.091	3.745
w/o MSIF	70.18	41.94	0.087	3.682
w/o Transformer encode	70.65	42.04	0.089	3.369
本文	72.16	40.56	0.078	3.380

4. 结论

本文提出的肝脏肿瘤分割网络 MCT-Net 较为有效解决了现有的肝脏 CT 图像中与四周正常组织之间的边界模糊、肿瘤数目大小高度可变性而导致的肿瘤分割方法中存在欠分割、过分割、分割精度低的问题。编码器结合了卷积神经网络(CNN)和 Transformer 架构的优势, 有效地结合局部特征信息和全局信息的长距离依赖建模, 能够更深入地挖掘图像特征, 解决肿瘤形状大小高度可变性问题; 细节特征提取模块(DFEM)保持模型性能的同时, 降低计算复杂度并减少参数量并后续的特征融合提供了精细的肿瘤轮廓信息, 促进整体精确分割; 局部空间注意力模块(LAM)和全局空间注意力模块(GAM)提高肿瘤特征的表达能力以充分聚焦并学习可变肿瘤的局部边缘信息和全局语义信息, 同时降低计算复杂度; 多尺度信息交互模块(MSIF)实现高分辨率和低分辨率特征之间的交流和互补, 能够更有效地捕捉肿瘤边缘的细微变化和局部结构改善模糊边界的分割; 分割网络 MCT-Net 分别在公开数据集 LiTS2017 和 3D-IRCADb 进行试验评估测试和泛化实验, 评价指标 Dice 和 ASD 分别为 72.16% 和 3.380 mm。在多次实验中证明其有效性, 对肝脏 CT 图像肝脏肿瘤的自动检测和辅助诊断具有一定的应用价值。

基金项目

国家自然科学基金资助项目(62373251)。

参考文献

- [1] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., *et al.* (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer*

- Journal for Clinicians*, **71**, 209-249. <https://doi.org/10.3322/caac.21660>
- [2] 乐美琰, 魏千越, 邓炜, 等. 基于电子计算机断层扫描图像的肝癌病灶自动分割方法研究进展[J]. 生物医学工程学杂志, 2018, 35(3): 481-487.
- [3] Aqil Burney, S.M. and Tariq, H. (2014) K-Means Cluster Analysis for Image Segmentation. *International Journal of Computer Applications*, **96**, 1-8. <https://doi.org/10.5120/16779-6360>
- [4] Cremers, D. (2003) A Multiphase Level Set Framework for Motion Segmentation. In: Griffin, L.D. and Lillholm, M., Eds., *Scale Space Methods in Computer Vision*, Springer, 599-614. https://doi.org/10.1007/3-540-44935-3_42
- [5] Liu, Y.W., Mao, J. and Chen, X.L. (2014) Interactive Liver Tumor Segmentation Method Based on Support Vector Machine Classification. *Automation and Instrumentation*, **6**, 166-169.
- [6] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [7] Oktay, O., Schlemper, J., Folgoc, L.L., et al. (2018) Attention U-Net: Learning Where to Look for the Pancreas. arXiv: 1804.03999.
- [8] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N. and Liang, J. (2020) UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, **39**, 1856-1867. <https://doi.org/10.1109/tmi.2019.2959609>
- [9] Arulappan, A. and Thankaraj, A.B.R. (2021) Liver Tumor Segmentation Using a New Asymmetrical Dilated Convolutional Semantic Segmentation Network in CT Images. *International Journal of Imaging Systems and Technology*, **32**, 815-830. <https://doi.org/10.1002/ima.22663>
- [10] Chen, J., Lu, Y., Yu, Q., et al. (2021) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv: 2102.04306.
- [11] Li, X., Chen, H., Qi, X., Dou, Q., Fu, C. and Heng, P. (2018) H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes. *IEEE Transactions on Medical Imaging*, **37**, 2663-2674. <https://doi.org/10.1109/tmi.2018.2845918>
- [12] Lv, P., Wang, J. and Wang, H. (2022) 2.5D Lightweight RIU-Net for Automatic Liver and Tumor Segmentation from CT. *Biomedical Signal Processing and Control*, **75**, Article ID: 103567. <https://doi.org/10.1016/j.bspc.2022.103567>
- [13] Chen, Y., Zheng, C., Zhang, W., Lin, H., Chen, W., Zhang, G., et al. (2023) MS-FANet: Multi-Scale Feature Attention Network for Liver Tumor Segmentation. *Computers in Biology and Medicine*, **163**, Article ID: 107208. <https://doi.org/10.1016/j.combiomed.2023.107208>
- [14] Deng, H., Deng, Y.X., Ding, T.B., et al. (2021) Liver CT Image Segmentation Based on Generative Adversarial Network. *Beijing Biomedical Engineering*, **40**, 367-376.
- [15] Wang, X., Wang, S., Zhang, Z., Yin, X., Wang, T. and Li, N. (2023) CPAD-Net: Contextual Parallel Attention and Dilated Network for Liver Tumor Segmentation. *Biomedical Signal Processing and Control*, **79**, Article ID: 104258. <https://doi.org/10.1016/j.bspc.2022.104258>
- [16] Peng, X.G. and Peng, D.L. (2023) MDA-Net: A Medical Image Segmentation Network That Combines Dual-Path Attention Mechanisms. *Journal of Chinese Computer Systems*, **44**, 2308-2313.
- [17] Kushnure, D.T. and Talbar, S.N. (2021) MS-UNet: A Multi-Scale UNet with Feature Recalibration Approach for Automatic Liver and Tumor Segmentation in CT Images. *Computerized Medical Imaging and Graphics*, **89**, Article ID: 101885. <https://doi.org/10.1016/j.compmedimag.2021.101885>
- [18] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 9992-10002. <https://doi.org/10.1109/iccv48922.2021.00986>
- [19] Dai, Z., Liu, H., Le, Q.V., et al. (2021) CoatNet: Marrying Convolution and Attention for All Data Sizes. *Advances in Neural Information Processing Systems*, **34**, 3965-3977.
- [20] Jiang, Y., Chang, S. and Wang, Z. (2021) Transgan: Two Pure Transformers Can Make One Strong Gan, and That Can Scale up. *Advances in Neural Information Processing Systems*, **34**, 14745-14758.
- [21] Shaw, P., Uszkoreit, J. and Vaswani, A. (2018) Self-attention with Relative Position Representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, New Orleans, June 2018, 464-468. <https://doi.org/10.18653/v1/n18-2074>
- [22] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. arXiv: 1706.03762.
- [23] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2021) An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.

-
- [24] Bilic, P., Christ, P., Li, H.B., *et al.* (2023) The Liver Tumor Segmentation Benchmark (Lits). *Medical Image Analysis*, **84**, Article ID: 102680.
 - [25] Milletari, F., Navab, N. and Ahmadi, S. (2016) V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 *Fourth International Conference on 3D Vision (3DV)*, Stanford, 25-28 October 2016, 565-571. <https://doi.org/10.1109/3dv.2016.79>
 - [26] Jaccard, P. (1901) Eude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société vaudoise des sciences naturelles*, **37**, 547-579.
 - [27] Heimann, T., van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., *et al.* (2009) Comparison and Evaluation of Methods for Liver Segmentation from CT Datasets. *IEEE Transactions on Medical Imaging*, **28**, 1251-1265. <https://doi.org/10.1109/tmi.2009.2013851>
 - [28] Song, L., Liu, G. and Ma, M. (2022) TD-Net: Unsupervised Medical Image Registration Network Based on Transformer and CNN. *Applied Intelligence*, **52**, 18201-18209. <https://doi.org/10.1007/s10489-022-03472-w>