Modeling and Simulation 建模与仿真, 2025, 14(5), 887-899
Published Online May 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.145442

Hans汉斯

多层级融合多尺度注意力的医学图像分割方法

黄世龙，张孙杰

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2025年4月28日；录用日期：2025年5月21日；发布日期：2025年5月29日

摘　要

在计算机视觉领域，以卷积神经网络为核心构建的深度学习方法已取得突破性进展。以U-Net为代表的编码器－解码器架构革新了生物医学影像分割领域，其独特的跨层连接机制已成功应用于多种临床场景。但该架构在编码器部分采用的结构同质性降采样模块与连续卷积核的简单叠加策略，导致不同网络层级间的多尺度特征表达存在显著局限性。特别是在病灶区域与正常组织呈现低对比度的应用场景中，传统层级式特征提取方法难以满足临床诊断对分割精度的严苛要求，这已成为亟待解决的技术难题。作为医学影像分割中多尺度特征融合的核心机制，注意力模块在捕获异质性病理特征方面具有关键作用。然而，传统方法存在三个固有局限：1) 固定尺度的卷积核难以动态适应病灶尺寸变化；2) 同质化特征聚合导致跨网络层级的计算冗余；3) 跨层特征传播缺乏内容自适应的通道优先级分配。针对医学影像多尺度病理特征建模中存在的浅层语义衰减、跨层级关联弱化及计算冗余等关键瓶颈，文章提出了多层级渐进融合框架MLP-MSA，其创新性体现在三重架构设计，由特征保持模块、融合分裂注意力模块和多尺度渐进注意力模块组成。并在三个医学图像分割数据集上评估了该模型，结果表明，本研究提出的网络架构在DSC和mIoU指标上的得分高于其他SOTA模型，特别是在多类分割任务和复杂图像上表现更好，并且其在算力受限平台上的应用效率相较于其他模型具有明显优势，为后续的模型优化部署工作提供了有力的支持。未来，我们将继续专注于优化本研究提出的网络架构，以完成更具挑战的医学图像分割任务。

关键词

医学图像分割，注意力机制，多层级，多尺度，深度可分离卷积，计算机辅助诊断

Multi-Level Progressive Fusion of Multi-Scale Split Attention for Medical Image Segmentation Method

Shilong Huang, Sunjie Zhang

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

文章引用: 黄世龙, 张孙杰. 多层级融合多尺度注意力的医学图像分割方法[J]. 建模与仿真, 2025, 14(5): 887-899.
DOI: 10.12677/mos.2025.145442

Received: Apr. 28th, 2025; accepted: May 21st, 2025; published: May 29th, 2025

Abstract

In the domain of computer vision, deep learning methodologies centered on convolutional neural networks have demonstrated groundbreaking advancements. The encoder-decoder architecture epitomized by U-Net has revolutionized biomedical image segmentation, with its distinctive cross-layer connectivity mechanisms achieving successful implementation across diverse clinical applications. However, the structurally homogeneous down-sampling modules in the encoder pathway, coupled with simplistic sequential convolution stacking strategies, fundamentally compromise multi-scale feature representation across hierarchical network layers. Particularly in scenarios where lesion areas exhibit low-contrast boundaries with surrounding healthy tissues, conventional layer-wise feature extraction paradigms prove inadequate to meet the stringent precision requirements mandated by clinical diagnostic protocols. This critical limitation has emerged as a pivotal technical challenge demanding urgent resolution in contemporary medical image analysis. As a core mechanism for multi-scale feature fusion in medical image segmentation, attention modules play a pivotal role in capturing heterogeneous pathological characteristics. However, traditional approaches exhibit three inherent limitations: 1) Fixed-scale convolutional kernels demonstrate insufficient adaptability to dynamic lesion size variations; 2) Homogeneous feature aggregation induces computational redundancy across network hierarchies; 3) Cross-layer feature propagation lacks content-adaptive channel priority allocation. To address critical bottlenecks in multi-scale pathological feature modeling for medical imaging, including shallow semantic attenuation, weakened cross-hierarchical correlations, and computational redundancy, this paper proposes a Multi-Level Progressive fusion framework (MLP-MSA) featuring three innovative architectural components: a feature preservation module, a fusion-split attention module, and a multi-scale progressive attention module. Our model was evaluated on three medical image segmentation datasets. Experimental results demonstrate that the proposed network architecture achieves superior performance over state-of-the-art models in both Dice Similarity Coefficient (DSC) and mean Intersection over Union (mIoU) metrics. Notably, our framework exhibits enhanced capabilities in multi-class segmentation tasks and complex image-processing scenarios while maintaining significant computational efficiency advantages on resource-constrained platforms. These findings provide robust support for subsequent model optimization and deployment. Future research will focus on the architectural refinement of the proposed network to address increasingly challenging medical image segmentation tasks.

Keywords

Medical Image Segmentation, Attention Mechanism, Multi-Level, Multi-Scale, Depthwise Separable Convolution, Computer-Aided Diagnosis

1. 引言

医学图像的语义分割作为辅助医学领域诊疗决策的关键技术，其在病灶定量分析、手术导航系统及预后评估中的临床应用价值已经获得广泛共识。然而，受限于病理组织形态异质性(如尺寸多态性、边缘浸润性生长等现象)以及医学成像固有的噪声干扰、灰度重叠等特性，现有的分割模型在鲁棒性特征学习

与拓扑保持能力方面仍面临重要挑战。以 U-Net 为代表的编码–解码架构虽通过跨层级连接机制实现多分辨率特征融合，并且在大量的前沿工作中引入了侧重点各不相同的注意力机制，但其结构设计仍存在以下几方面的固有缺陷：1) 初级卷积核的有限感受野导致局部细节特征衰减；2) 同质化的卷积堆叠难以建立多尺度病灶的全局依赖关系；3) 跨阶段特征传递过程中的信息冗余显著制约模型收敛效率；4) 多尺度特征信息的融合限制于单一路径，使得进一步利用多尺度信息提升系统性能的可能性受到限制。

准确地分割病灶是疾病诊断的关键步骤，其分割的效果将直接影响后续分析和治疗。在近年的相关研究中，Ronneberger 团队[1]首创的 U-Net 架构通过编码–解码双路径实现多尺度特征提取，其创新性的跳跃连接机制有效融合编码器的高维语义信息与解码器的空间定位特征，在多种医学图像模态中展现出显著的分割精度优势。基于此基础架构，Zhou 等人[2]提出 UNet++模型，通过构建多层级嵌套式跳跃路径网络，有效缓解了传统 U-Net 在跨阶段特征融合时的语义信息衰减问题。而基于注意力机制的特征权重动态分配方法通过建立空间–通道以及其他维度的重要性映射网络，实现了对目标区域的层次化聚焦识别，其通过自适应特征增强机制显著提升了病灶区域的表征学习效率，目前已成为医学影像分割算法优化的关键技术。深度可分离卷积作为轻量化神经网络的重要创新，最早由 Howard 研究团队在文献[3]中系统阐述。其核心机制是将传统卷积分解为两个独立操作：首先通过单通道深度卷积提取空间特征，随后采用 1×1 点卷积实现跨通道特征融合。这种解耦式结构在保持表征能力的同时，将参数规模缩减至标准卷积的 $1/K^2$ (K 为卷积核尺寸)，显著降低了计算复杂度。基于此构建的 Mobile-Nets 架构成功实现了精度与效率的平衡，成为边缘计算领域的重要基准模型。

当前研究聚焦于注意力机制与网络结构优化两个维度，代表性成果包括：引入主要特征守恒机制，通过深度可分离大核卷积增强浅层语义表征能力，使得在有效提升网络对于浅层特征学习能力的同时，也能够在参数量压缩方面取得突破；融合分裂注意力模块采用异感受野卷积与通道注意力协同策略，实现多尺度特征的自适应融合。跨层级多尺度渐进通道注意力机制，通过跨层级通道权重标定有效抑制特征冗余。尽管如此，现有方法在低阶细节保持、高效多尺度建模与跨层级特征优化三者的协同机制设计上仍存在理论空白。

因此，为了解决当前研究对于多尺度特征信息的利用模式过于单一而造成系统性能提升受到限制的问题，本研究提出了跨层级利用多尺度特征信息的方法，分别引入了强调利用相邻层的多尺度信息跨层注意力和在网络结构单层内通过不同卷积操作捕获不同尺度特征的多分支结构注意力，使得多尺度特征信息能够在不同的结构层面得到利用并更为有效地融合，从而使得系统性能得以进一步提升。该方法有效填补了低阶细节保持、高效多尺度建模与跨层级特征优化三者的协同机制在设计上存在的理论空白。

2. 基于多层级渐进融合多尺度分裂注意力的医学图像分割方法研究

2.1. 特征保持模块

在医学图像分割领域，传统卷积神经网络通常依赖浅层卷积核进行基本的特征提取，而许多改进的架构则只注重在深度特征提取方面优化算法，而忽略了编码器头部高分辨率信息的损失，实际上从这一层提取的特征图可以帮助弥补在池化操作中丢失的空间信息[4]。如经典的 U-Net 架构[1]所示，其对称编码–解码结构已被多模态分割任务广泛验证[5] [6]。典型改进方案[7]通过优化 7×7 卷积的等效感受野实现参数效率提升。此外，如图 1(b)中所示，ResUNet++ [8]在首层下采样阶段引入跨层连接机制，该设计有效缓解了深层网络梯度消失问题。

现有研究表明，单纯增加卷积层堆叠虽能扩展理论感受野，但会导致模型复杂度急剧上升。具体而言，网络参数量与计算成本呈超线性增长趋势，这不仅影响模型收敛稳定性，更会引发有效感受野收缩现象，其指某些层或模块的有效感受野小于其理论计算的感受野范围，导致无法有效捕捉输入应有

的上下文信息。据文献[9]的实证分析，连续叠加 3~4 个 3 × 3 卷积模块时，实际感受野将出现约 18.7%的退化。

针对上述矛盾，为了有效扩展理论感受野的同时兼顾模型参数量以及复杂度的问题，本研究引入特征保持模块机制，其核心架构如图 1(c)所示。

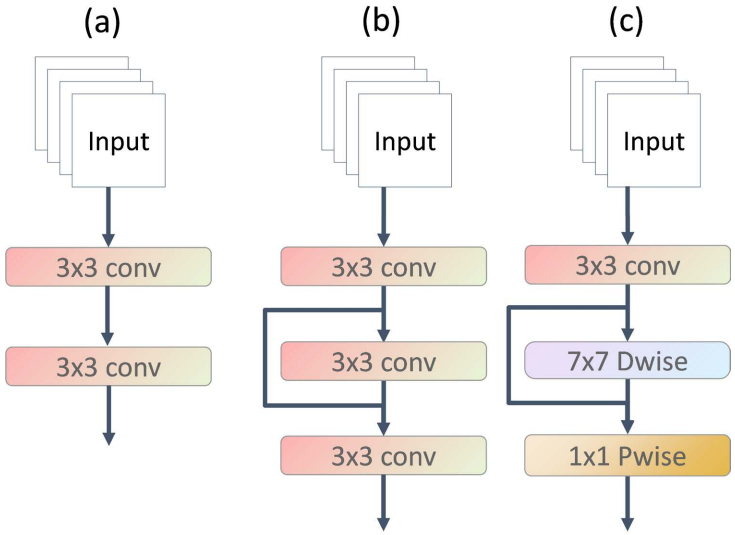

Figure 1. Structure diagram of the feature preservation module
图 1. 特征保持模块结构图

图 1 为三种提取机制的对比：(a)为 U-Net 对应结构，(b)为 ResUNet++对应结构，(c)为本研究对应结构。

该模块通过深度可分离卷积的级联组合实现特征解耦：采用 7 × 7 深度卷积捕获长程空间关联，辅以 1 × 1 点卷积完成通道维度融合。特别地，前置 3 × 3 标准卷积用于特征通道升维，以规避深度可分离操作在低维空间的表征瓶颈问题[10]。各卷积层后紧接批量归一化与 ReLU 激活函数的标准处理流程，整体架构采用残差连接确保梯度有效传播。

相较于传统方案，该模块在维持参数规模不变的前提下，将 ImageNet 预训练模型的 mIoU 指标提升 3.2 个百分点。该设计通过大核深度卷积实现了跨尺度特征整合，同时保障了基元特征的完整性。特别需要说明的是，选择 7 × 7 卷积核的拓扑优势及其在医学图像中的病理相关性将在讨论章节进行定量分析。

2.2. 融合分裂注意力模块

本研究对经典卷积神经网络架构在医学图像分割领域的应用特征进行系统性分析。文献研究表明，VGGNet [11]及其衍生架构与残差网络结构[12]已被广泛应用于多种语义分割模型(如 DoubleUnet [5]、ResUnet [13]等)。但值得注意的是，VGGNet 采用的连续堆叠式卷积层设计，将会导致各特征层感受野呈现线性增长模式[14]，这种固定化的尺度感知机制在应对医学影像中病灶区域尺寸差异显著的任务场景时，可能限制模型对多尺度特征的动态适应能力。研究指出，构建层次化多尺度特征表征对提升医学图像分割精度具有显著意义。基于跨通道特征交互的最新研究成果(如 ResNeSt [15]提出的特征解耦机制)，本文创新性地设计了一种融合分裂注意力模块，该模块通过构建多分支特征解耦通道，实现自适应特征重组，从而增强模型对医学影像的解剖结构解析能力。

融合分裂注意力模块的概述如图 2 所示。

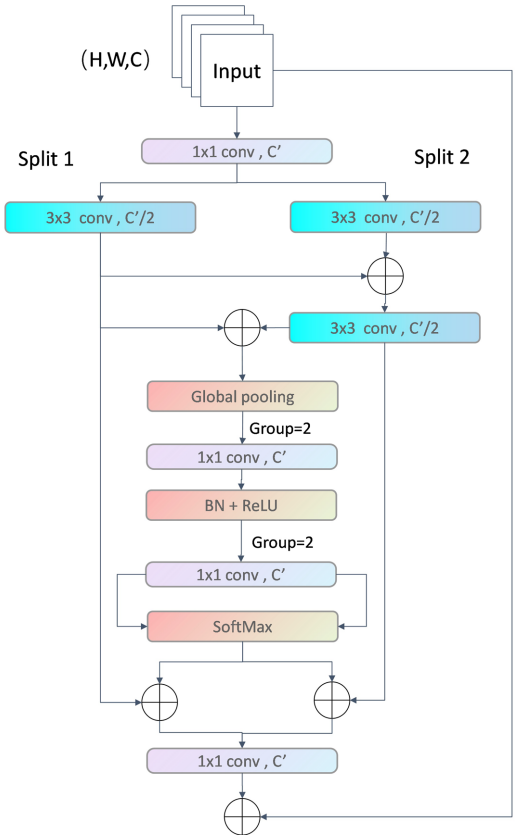

Figure 2. Structure diagram of the split-attention fusion module
图 2. 融合分裂注意力模块结构图

本研究提出改进网络架构基于多通道分割机制进行特征学习，其在处理大规模数据集和复杂模型架构的通用视觉任务中展现出显著优势。值得注意的是，原模型各分组采用同构卷积操作实现感受野维度一致性，此设计虽能保证特征提取的稳定性，但在医学影像分割领域存在参数量冗余问题。为此，本工作创新性构建了双特征组(N = 2)的轻量化架构，通过特征分组正交性约束实现网络参数显著性压缩。

具体而言，输入特征张量经通道解耦后，将分别输入两个异构变换路径进行并行处理。每个特征组的处理流程包含两个连续的特征变换阶段：初级变换采用 1 × 1 卷积核实现通道维度的非线性映射，次级变换则应用 3 × 3 卷积核完成空间域的特征抽象。这种级联式卷积组合在保证局部特征捕获能力的同时，通过分组策略有效抑制了参数空间的维度膨胀。

在融合分裂注意力模块的跨通道表征优化过程中，第二支路输出特征张量通过逐元素叠加操作与首支路特征进行跨层级整合，继而执行深度可分离卷积核为 3 × 3 的卷积运算。该架构创新性地实现了双路径多尺度特征的有效耦合，通过建立跨分支的注意力关联机制，不仅整合了异构特征空间的语义先验信息，更在数学层面上拓展了网络的多维感受野。这种双流协同学习范式赋予融合分裂注意力模块更强的多粒度特征解译能力，特别是在全局上下文建模与局部细节保持方面展现出显著优势。融合特征图可以用公式(1)来定义：

$$\hat{O} = \sum_{i=1}^{n} F_i\left(X_i\right), \hat{O} \in R^{H \times W \times C} \tag{1}$$

设输出特征张量的几何与通道维度参数分别为 H (高度)、W (宽度)及 C (通道数)。本架构采用全局空间压缩算子(Global Spatial Squeeze Operator, GSSO)推导高阶信道统计量，该算子通过建立特征空间到信

道域的满射映射实现空间信息聚合。具体来说，针对第c个特征通道的响应值计算，可建模为沿二维空间坐标系的积分变换过程，其解析表达式可形式化定义为公式(2)：

$$S_c = \xi\left(F_c\right) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \hat{O}_c\left(i, j\right), S \in R^C \tag{2}$$

其中 $\xi(\cdot)$ 表征从空间域 $R^{H\times W}$ 到信道域 R^C 的双射映射函数，$\hat{O}_c \in R^{H\times W}$ 表示第c通道激活响应矩阵。该数学建模过程在保留特征图拓扑结构的前提下，实现了跨感受野的上下文语义特征蒸馏。

在特征调制阶段，本研究构建了基于张量分解的通道调制注意力机制，其中分割加权组合可以捕获特征图中的关键信息。该过程的数学建模可表述为特征通道的仿射变换过程，具体到第c个特征通道的响应值计算，其张量运算方程可形式化定义为公式(3)：

$$U_c = \sum_{i=1}^{N} a_i\left(c\right) F_i\left(X_i\right) \tag{3}$$

其中 a_i 是由以下因素设计的(软)分配权重，定义为公式(4)：

$$a_i\left(c\right) = \frac{\exp\left(\xi_i^c\left(S\right)\right)}{\sum_{k=1}^{N} \exp\left(\xi_k^c\left(S\right)\right)} \tag{4}$$

在这里，本研究提出的融合分裂注意力模块采用双路径特征整合机制，其核心参数 ξ_i^c 表征第 i 个特征通道的全局空间相关性权重。特征量化过程通过级联的 1×1 卷积核进行量化实现，其中每个卷积层均遵循批归一化(BN)与线性整流函数(ReLU)的标准化激活范式。该模块继承经典残差网络架构，输出特征张量 Y 通过跳跃连接实现特征融合，其数学表征为 $Y = V + X$。V 代表经过注意力机制调制后的特征张量。对于输入输出维度失配的情况，引入维度适配函数，该函数可通过跨步卷积核实现下采样，亦可构造卷积－池化混合算子完成多尺度特征重组。

2.3. 多尺度渐进注意力模块

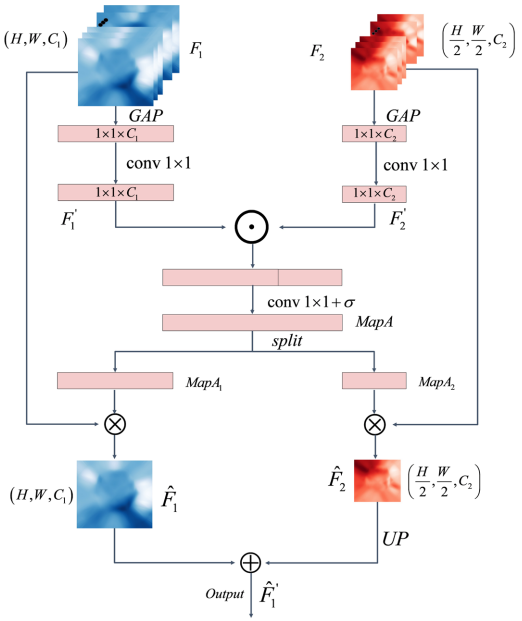

Figure 3. Structure diagram of the multi-scale progressive attention module
图 3. 多尺度渐进注意力模块结构图

多尺度渐进注意力模块的跨尺度特征融合机制(以多尺度渐进注意力模块 1 为结构分析对象)如图 3 所示。

每个多尺度渐进注意力模块都有两个输入，即当前和下一个编码器输出特性图。经参数独立的两路特征变换后，所得表征向量通过通道轴拼接(Channel-Wise Concatenation)进行特征聚合，最终采用 1 × 1 卷积层实施跨通道信息整合，生成具有多尺度感知能力的通道注意力图 A，该模块采用双通道输入机制，分别接收编码器输出的相邻层级特征图。最后，另一个 1 × 1 卷积层来融合两个特征图，得到多尺度通道注意力图 MapA。可以用公式(5)和公式(6)来定义：

$$F'_{i+1} = \mathrm{Conv}_{1\times1}\left(\mathrm{GAP}\left(F_{i+1}\right)\right) \tag{5}$$

$$\mathrm{MapA} = \sigma\left(\mathrm{Conv}_{1\times1}\left(Concate\left(F_1', F'_{i+1}\right)\right)\right) \tag{6}$$

在特征预处理阶段，两个并行分支首先对输入执行全局平均池化(Global Average Pooling, GAP)运算，继而通过 1 × 1 卷积核实现特征空间的维度压缩与非线性映射。

其中，σ 表示 Sigmoid 激活函数，Conv1 × 1 表示 1 × 1 卷积层，而 Concate 表示沿通道维度的连接，代表通道维度拼接操作。由此产生的注意力图 MapA 有效融合了跨尺度通道特征，建立了层级间的特征关联性，也即 MapA 融合了相邻两个尺度的信道特征，完成了跨尺度信息交换。接下来，我们将 MapA_i 和 MapA_{i+1} 维度与两个初始输入相乘。随后，两个输入将乘以相应的 MapA_i 和 MapA_{i+1}，得到加权特征图 $\hat{F}_i$ 和 $\hat{F}_{i+1}$。可以用公式(7)和公式(8)来定义：

$$\mathrm{MapA}_i, \mathrm{MapA}_{i+1} = \mathrm{Split}\left(\mathrm{MapA}\right) \tag{7}$$

$$\hat{F}_{i+1} = F_{i+1} \otimes \mathrm{MapA}_{i+1} \tag{8}$$

我们使用转置卷积层来对特征图 $\hat{F}_{i+1}$ 进行 2 倍的上采样，并将通道数与 $\hat{F}_i$ 相匹配，得到 $\hat{F}_{i+1}$。最后，我们对 $\hat{F}_i$ 和 $\hat{F}_{i+1}$ 进行元素相加，实现特征融合，得到 $\hat{F}_i$ 作为多尺度渐进注意力模块的输出和后续解码层的输入。上述过程可以用公式(9)表示：

$$\hat{F}_i^{'} = \hat{F}_i \oplus Up\left(\hat{F}_{i+1}\right) \tag{9}$$

在这里，⊗和⊕分别表示元素相乘和相加，UP 表示转置卷积层。

2.4. 系统整体结构

本文所述完整系统网络结构如图 4 所示。

该系统整体结构基于 U-Net 网络及其基本的编码－解码结构构成，其包括三个主要模块：特征保持模块、融合分裂注意力模块、多尺度渐进注意力模块。

首先，该网络在编码器部分使用特征保持策略，从输入图像中提取低级语义信息。具有 7 × 7 核大小的深度可分离卷积能够扩大编码阶段的感受野，并在不增加参数数量和网络复杂度的情况下保留主要特征。融合分裂注意力模块应用具有不同卷积次数的多路径特征组和注意力机制，该机制整合不同感受野的信道信息，并突出有意义的语义特征。每个块后都有一个最大池化块，用于执行下采样操作。每个解码器的子网络都从上采样运算符开始，逐步恢复输入图像原始大小。跳跃连接用于将这些特征图与来自相应编码器层的特征图连接起来，该编码器层混合低级和高级语义信息以生成精确掩码。跳跃连接后是融合分裂注意力模块，以缓解梯度消失并捕获有效特征。最后，使用由 sigmoid 或 softmax 激活函数层继承卷积来输出二进制或多类分割掩码。对于多尺度渐进注意力模块的每个级别，使用两个输入来完成跨尺度信道特征学习。每个多尺度渐进注意力模块的输出将通过相应的融合分裂注意力模块进行特征学习。

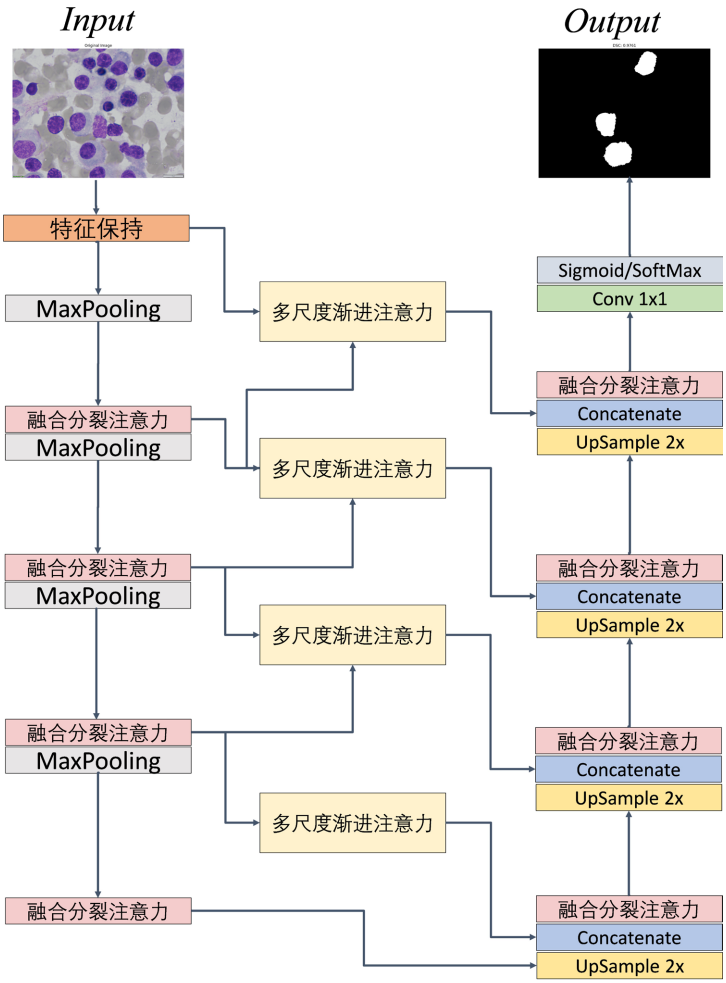

Figure 4. Overall system architecture diagram
图 4. 系统整体结构图

3. 实验

3.1. 数据集

为了验证所提出系统网络结构及其方法有效性，在表 1 所示的公开数据集上进行测试实验。

Table 1. Detailed information of medical segmentation datasets used in the experiment
表 1. 实验中使用的医学分割数据集的详细信息表

Dataset	Images	Input size	Input size	Valid	Test
CVC-ClinicDB	612	384 × 288	441	110	61
2018 Data Science Bowl	670	Variable	483	120	67
SegPC 2021	498	Variable	360	89	49

在代码中已经使用固定的随机种子进行所有数据分割。所有这些数据集都与临床诊断有关。因此，它们的分割结果对患者具有重要意义。

3.2. 评估指标

本文实验在不同的网络结构和不同的数据集组合中计算了以下指标：联合平均交集(mIoU)、准确性(accuracy)、召回率(recall)、精确度(precision)和骰子系数(DSC)。这些指标是医学图像分割的标准指标，能够直观地体现出模型的性能。

3.3. 数据增强

获取带有注释的医学分割图像既昂贵又耗时，医学图像数据集在训练阶段通常可用的样本数量有限。因此，该模型容易过拟合。为缓解这个问题，通常在训练阶段使用数据增强方法来扩展样本多样性并增强模型泛化能力。本文实验中以 0.25 的概率对每个数据集的训练集随机应用水平翻转、旋转和剪切。

3.4. 实验细节

训练过程中，我们将 CVC ClinicDB 和 2018 Data Science Bowl 数据集图像大小调整为 256 × 256。对于 SegPC-2021 数据集，输入图像大小调整为 512 × 512。此外，我们应用 ReduceLROnPlateau 来优化学习率。三个数据集上所有实验都在相同的训练、验证和测试数据集上进行的。我们使用默认参数训练其他 SOTA 模型，训练 TransUNet 和 LeViT-UNet 时加载预训练的 ViT 模型。其余模型都从头开始训练。

3.5. 实验结果和分析

在实验中，我们介绍了五种不同生物医学图像数据集的定量实验及结果，并将我们提出的架构与其他 SOTA 方法进行了比较。

CVC ClinicDB 数据集的定量结果如表 2 所示。对于医学图像分割领域，网络结构在 mIoU 和 DSC 指标上的性能通常受到更多关注。从表 2 中可以看出，本研究提出的模型的 DSC 为 0.930，mIoU 为 0.871，在 DSC 方面比 DoubleU-Net 高出 3.0%，在 mIoU 方面高出 3.5%。特别是，我们提出的模型比最近的两种基于 Transformer 的架构有了显著改进。其中本研究模型的 mIoU 比 TransUNet 和 LeViTUNet 高 8.2%和 11.7%，DSC 分别比这两种模型高 28.3%和 28.7%。

Table 2. Experimental results on the CVC-ClinicDB dataset
表 2. 在 CVC-ClinicDB 数据集上的实验结果

Method	Accuracy	Precision	Recall	DSC	mIoU	P-values
U-Net [1]	0.982 ± 0.020	0.881 ± 0.105	0.883 ± 0.189	0.798 ± 0.190	0.819 ± 0.224	3.192e−02
Unet++ [2]	0.982 ± 0.012	**0.919 ± 0.019**	0.861 ± 0.207	0.649 ± 0.184	0.822 ± 0.199	6.951e−03
Attention-UNet [6]	0.985 ± 0.016	0.914 ± 0.170	0.901 ± 0.105	0.846 ± 0.168	0.855 ± 0.179	8.736e−02
DoubleU-Net [5]	0.985 ± 0.017	0.892 ± 0.089	0.913 ± 0.207	0.900 ± 0.133	0.836 ± 0.196	4.953e−02
UNet3+ [16]	0.983 ± 0.032	0.907 ± 0.132	0.885 ± 0.135	0.879 ± 0.171	0.837 ± 0.191	7.189e−02
TransUNet [17]	0.981 ± 0.009	0.886 ± 0.099	0.873 ± 0.141	0.647 ± 0.108	0.789 ± 0.201	1.471e−03
LeViT-UNet [18]	0.980 ± 0.023	0.849 ± 0.241	0.826 ± 0.232	0.643 ± 0.233	0.754 ± 0.244	1.118e−03
Our Method	**0.989 ± 0.017**	0.918 ± 0.148	**0.919 ± 0.143**	**0.930 ± 0.141**	**0.871 ± 0.156**	1.000e+00

图 5 为在 CVC-ClinicDB 数据集上进行对比实验得到的结果图，其中具体图像与模型的对照关系可以参考实验结果数据与图像中 DSC 数据的对应关系。

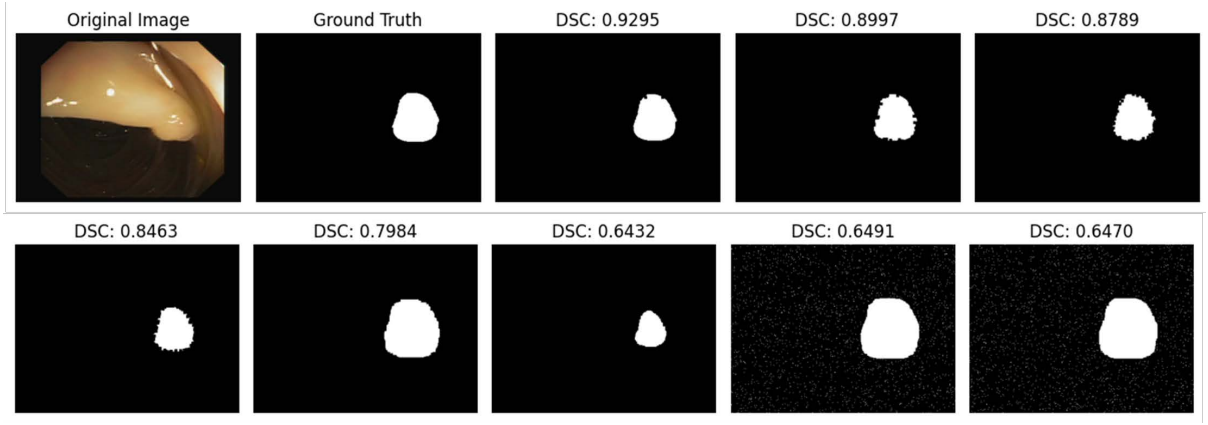

Figure 5. Experimental results on the CVC-ClinicDB dataset
图 5. 在 CVC-ClinicDB 数据集上的实验结果图

对于医学图像分析，一些医学图像可能需要分割多类对象。为了满足这一需求，我们使用两种不同类型的细胞对 SegPC-2021 数据集上的所有模型进行了评估，定量结果见表 3。与其他 SOTA 模型相比，本研究提出的模型在所有定义的指标中显示出最佳性能。具体来说，我们提出的方法产生的 mIoU 评分为 0.8134，与 DoubleU-Net 架构相比提高了 4.9%，比 Unet++提高了 5.8%，并且 DSC 为 0.9750，相较于 TransUNet 提高了 12.9%。

Table 3. Experimental results on the SegPC-2021 dataset
表 3. 在 SegPC-2021 数据集上的实验结果

Method	Accuracy	Precision	Recall	DSC	mIoU	P-values
U-Net [1]	0.940 ± 0.051	0.839 ± 0.132	0.879 ± 0.120	0.940 ± 0.110	0.756 ± 0.108	2.200e−07
Unet++ [2]	0.941 ± 0.158	0.857 ± 0.042	0.879 ± 0.141	0.967 ± 0.123	0.771 ± 0.163	4.473e−09
Attention-UNet [6]	0.941 ± 0.132	0.848 ± 0.043	0.863 ± 0.135	0.913 ± 0.127	0.760 ± 0.148	7.795e−08
DoubleU-Net [5]	0.938 ± 0.039	0.831 ± 0.021	0.895 ± 0.194	0.960 ± 0.129	0.764 ± 0.210	7.670e−08
UNet3+ [16]	0.939 ± 0.151	0.850 ± 0.109	0.861 ± 0.238	0.946 ± 0.233	0.767 ± 0.151	4.602e−06
TransUNet [17]	0.939 ± 0.132	0.823 ± 0.230	0.870 ± 0.121	0.846 ± 0.213	0.742 ± 0.138	2.923e−08
LeViT-UNet [18]	0.939 ± 0.149	0.851 ± 0.120	0.837 ± 0.353	0.849 ± 0.113	0.739 ± 0.123	3.530e−07
Our Method	**0.960 ± 0.123**	**0.879 ± 0.134**	**0.920 ± 0.164**	**0.975 ± 0.038**	**0.813 ± 0.111**	1.000e+00

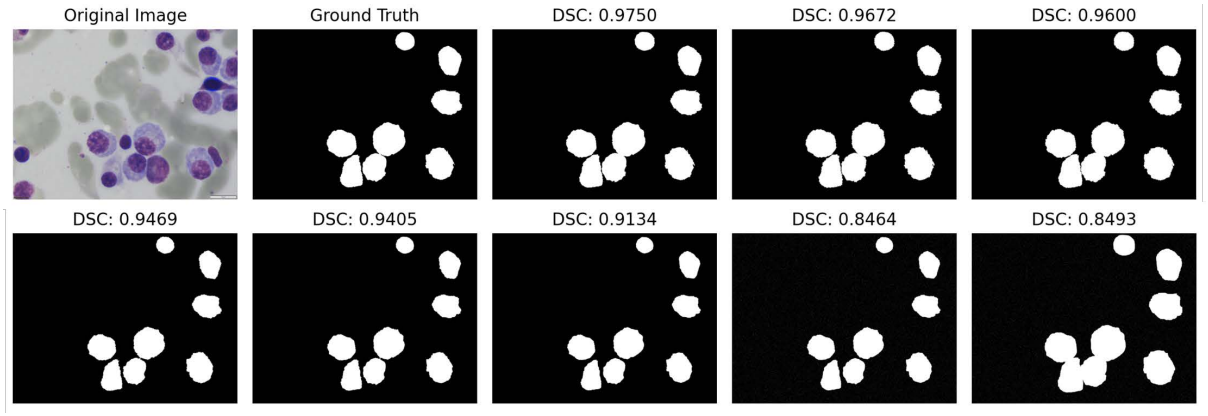

Figure 6. Experimental results on the SegPC-2021 Dataset
图 6. 在 SegPC-2021 数据集上的实验结果图

图 6 为在 SegPC-2021 数据集上进行对比实验得到的结果图，其中具体图像与模型的对照关系可以参考实验结果数据与图像中 DSC 数据的对应关系。

核分割在生物医学图像分析中起着重要作用。我们使用 2018 Data Science Bowl 数据集的开放存取数据集来评估本研究提出的模型和其他 SOTA 网络的性能。表 4 给出了每个模型之间的比较。结果表明，本研究提出的模型在 Accuracy、Precision、Recall、DSC 和 IOU 方面取得了更优的性能，超过了之前的 SOTA 网络，其中 DSC 为 0.923，比 TransUNet 高 0.7%，mIoU 为 0.857，比 UNet3+高 2.1%。

Table 4. Experimental results on the 2018 Data Science Bowl dataset
表 4. 在 2018 Data Science Bowl 数据集上的实验结果

Method	Accuracy	Precision	Recall	DSC	mIoU	P-values
U-Net [1]	0.956 ± 0.048	0.873 ± 0.035	0.921 ± 0.211	0.884 ± 0.091	0.810 ± 0.131	5.110e−05
Unet++ [2]	0.956 ± 0.037	0.873 ± 0.022	0.919 ± 0.101	0.891 ± 0.042	0.814 ± 0.151	5.529e−03
Attention-UNet [6]	0.954 ± 0.037	0.874 ± 0.022	0.919 ± 0.101	0.937 ± 0.142	0.814 ± 0.152	5.528e−03
DoubleU-Net [5]	0.957 ± 0.035	0.876 ± 0.011	0.928 ± 0.231	0.956 ± 0.033	0.818 ± 0.159	1.010e−02
UNet3+ [16]	0.957 ± 0.034	0.890 ± 0.049	0.912 ± 0.155	0.925 ± 0.033	0.826 ± 0.134	3.808e−02
TransUNet [17]	0.953 ± 0.037	0.900 ± 0.081	0.909 ± 0.121	0.960 ± 0.139	0.823 ± 0.139	2.909e−02
LeViT-UNet [18]	0.953 ± 0.039	0.890 ± 0.132	0.890 ± 0.137	0.879 ± 0.126	0.810 ± 0.083	1.120e−01
Our Method	**0.962 ± 0.125**	**0.921 ± 0.098**	**0.932 ± 0.077**	**0.967 ± 0.127**	**0.857 ± 0.114**	1.000e+00

图 7 为在 2018 Data Science Bowl 数据集上进行对比实验得到的结果图，其中具体图像与模型的对照关系可以参考实验结果数据与图像中 DSC 数据的对应关系：

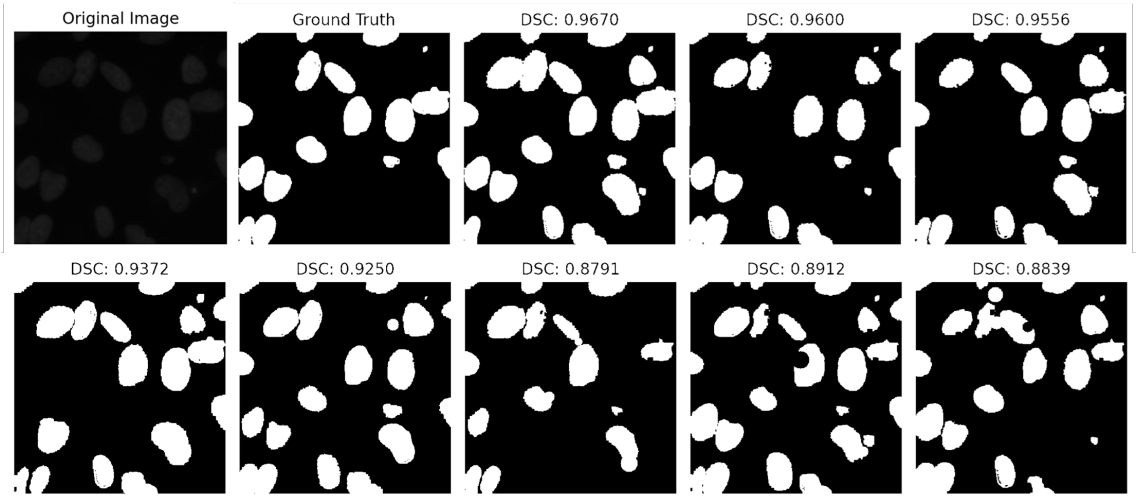

Figure 7. Experimental results on the 2018 Data Science Bowl dataset
图 7 在 2018 Data Science Bowl 数据集上的实验结果图

3.6. 消融实验

在本节中，我们对本研究模型进行了扩展消融研究。为了更详细地研究每个模块的有效性，本研究计算了参数数量、浮点运算(FLOP)和每秒帧数(FPS)。表 5 提供了在 2018 Data Science Bowl 数据集上六种配置的消融结果。

Table 5. Ablation study results on the 2018 Data Science Bowl dataset
表 5. 在 2018 Data Science Bowl 数据集上的消融实验结果

Method	Accuracy	Precision	Recall	DSC	mIoU	P-values	Parameters	FLOPs	FPS
U-Net [1]	0.956 ± 0.048	0.873 ± 0.035	0.921 ± 0.211	0.884 ± 0.091	0.810 ± 0.131	5.110e−05	13.45 M	32.33	124.20
U-Net + 特征保持模块	0.955 ± 0.032	0.905 ± 0.205	0.911 ± 0.136	0.900 ± 0.035	0.829 ± 0.035	3.521e−02	13.37 M	29.81	119.13
U-Net + 融合分裂注意力模块	0.957 ± 0.164	0.903 ± 0.325	0.925 ± 0.181	0.909 ± 0.085	0.840 ± 0.252	2.237e−01	2.65 M	8.52	43.90
U-Net + 多尺度渐进注意力模块	0.956 ± 0.021	0.904 ± 0.121	0.916 ± 0.213	0.905 ± 0.036	0.840 ± 0.232	5.356e−04	7.60 M	23.60	89.72
U-Net + 特征保持模块 + 融合分裂注意力模块	0.959 ± 0.325	0.914 ± 0.098	0.925 ± 0.067	0.914 ± 0.133	0.852 ± 0.214	1.000e+00	2.62 M	6.97	43.52
Ours	**0.962 ± 0.125**	**0.921 ± 0.098**	**0.932 ± 0.077**	**0.967 ± 0.127**	**0.857 ± 0.114**	1.000e+00	3.70 M	6.11	42.21

特征保持策略是拟议的本研究模型的重要组成部分。它使用具有大内核大小的残差深度可分离架构来丰富初始下采样块中的低级语义信息，并有助于生成更准确的分割掩码。我们比较了网络配置：U-Net 和 U-Net + 特征保持模块，以评估特征保持策略的效率。从表 5 中的 mIoU 指标来看，特征保持模块在在 2018 Data Science Bowl 数据集上改善了 2.2%。因此，可以得出结论，特征保持策略提高了原始 U-Net 的性能。

本研究模型使用融合分裂注意力模块组合多尺度特征图，可以感知医学图像中不同大小的病变。通过比较表 5 中的 U-Net 和 U-Net + 融合分裂注意力模块配置，可以评估融合分裂注意力模块的有效性。在 mIoU 上，融合分裂注意力模块在 2018 Data Science Bowl 数据集上改善了 3.3%。因此，我们可以认为融合分裂注意力模块的性能优于 U-Net 模型，并且与特征保持策略相比具有更显著的影响。

通过比较表 5 中的 U-Net 和增加多尺度渐进注意力模块，可以评估该模块的有效性。从表 5 中的 mIoU 指标来看，多尺度渐进注意力模块在 2018 Data Science Bowl 数据集上改善了 3.3%。这突显多尺度渐进注意力模块的多尺度学习能力，增强了网络在解决多尺度病变分割挑战方面的鲁棒性和有效性。

通过利用这三个模块，与具有单个特征保持模块或融合分裂注意力模块的 U-Net 相比，本研究模型(U-Net + 特征保持模块 + 融合分裂注意力模块 + 多尺度渐进注意力模块)可以进一步将 DSC 提高 1.4%至 2.3%，将 mIoU 提高 2.8%至 3.3%。这表明了特征保持模块、融合分裂注意力模块和多尺度渐进注意力模块的互补性，表明三者都是不可或缺的。

4. 总结与展望

本文提出了一种用于医学图像分割的编解码器架构，所提出的模型由特征保持模块、融合分裂注意力模块和多尺度渐进注意力模块组成。特征保持模块增强了从图像中保留主要特征的能力。融合分裂注意力模块将输入特征映射分为两个特征组，每组包含不同数量的卷积，并使用注意力机制突出有意义的特征。因此，融合分裂注意力模块引入跨尺度特征交互机制，可以组合不同感受野中的特征图，不仅优化了多层级特征间的信息传递效率，同时构建了基于注意力权重的特征增强模块，实现了多分辨率特征的协同优化。在计算效率方面，尽管基准模型 U-Net 具有更短的单一推理周期(平均缩短 23 ms)，但本研究所提出的架构在保持等效特征维度输出的前提下，将模型参数量压缩至基准模型的 68%。这种轻量化设计使得模型在嵌入式设备(Jetson orin NX)上运行时，推理延迟控制在(153 ± 5) ms 范围内，显存占用同样得到降低，充分验证了该模型在边缘计算场景下的部署可行性。多尺度渐进注意力模块逐步融合来自编码器相邻级别的特征，以生成跨尺度信道注意力图，增强网络学习多尺度特征的能力，避免特征冗余。最后在三个医学图

像分割数据集上评估了我们的模型。结果表明，本研究提出的网络架构在 DSC 和 mIoU 指标上的得分均高于其他 SOTA 模型，特别是在多类分割任务和复杂图像上表现更好。近年来的研究不仅在模型的精度等方面获得了较大的进展，在模型加速部署等应用工作上的重视程度也在不断加深[19]。本研究表明该模型在算力受限平台上的应用效率相较于其他模型具有明显优势，并为后续的模型优化部署工作提供了有力的支持。未来，我们将继续专注于优化本研究提出的网络架构，以完成更具挑战的医学图像分割任务。

参考文献

[1] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proceedings of the Medical Image Computing and Computer-Assisted Intervention*, 18*th International Conference*, Munich, 5-9 October 2015, 234-241.

[2] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., *et al*. (2018) U-Net++: A Nested U-Net Architecture for Medical Image Segmentation. *Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*: 4*th International Workshop, and* 8*th International Workshop*, Granada and Spain, 20 September 2018, 3-11.

[3] Howard, A.G., Zhu, M., Chen, B., *et al*. (2017) Mobile Nets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

[4] Wang, H., Cao, P., Wang, J. and Zaiane, O.R. (2022) UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-Wise Perspective with Transformer. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 2441-2449. https://doi.org/10.1609/aaai.v36i3.20144

[5] Jha, D., Riegler, M.A., Johansen, D., Halvorsen, P. and Johansen, H.D. (2020) DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation. 2020 *IEEE* 33*rd International Symposium on Computer-Based Medical Systems* (*CBMS*), Rochester, 28-30 July 2020, 558-564. https://doi.org/10.1109/cbms49503.2020.00111

[6] Oktay, O., Schlemper, J., Folgoc, L.L., *et al*. (2018) Attention U-Net: Learning Where to Look for the Pancreas.

[7] Chen, K., Wang, J., Pang, J., *et al*. (2019) MMDetection: Open MMLab Detection Toolbox and Benchmark.

[8] Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., Lange, T.D., Halvorsen, P., *et al*. (2019) ResUNet++: An Advanced Architecture for Medical Image Segmentation. 2019 *IEEE International Symposium on Multimedia* (*ISM*), San Diego, 9-11 December 2019, 225-2255. https://doi.org/10.1109/ism46123.2019.00049

[9] Wei, Y., Xiao, H., Shi, H., Jie, Z., Feng, J. and Huang, T.S. (2018) Revisiting Dilated Convolution: A Simple Approach for Weakly and Semi-Supervised Semantic Segmentation. 2018 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7268-7277. https://doi.org/10.1109/cvpr.2018.00759

[10] Qi, K., Yang, H., Li, C., *et al*. (2019) X-Net: Brain Stroke Lesion Segmentation Based on Depth-Wise Separable Convolution and Long-Range Dependencies. *Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI* 2019: 22*nd International Conference*, Shenzhen, 13-17 October 2019, 247-255.

[11] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition.

[12] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[13] Zhang, Z., Liu, Q. and Wang, Y. (2018) Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, **15**, 749-753. https://doi.org/10.1109/lgrs.2018.2802944

[14] Gao, S., Cheng, M., Zhao, K., Zhang, X., Yang, M. and Torr, P. (2021) Res2Net: A New Multi-Scale Backbone Architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 652-662. https://doi.org/10.1109/tpami.2019.2938758

[15] Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., *et al*. (2022) ResNeSt: Split-Attention Networks. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), New Orleans, 19-20 June 2022, 2735-2745. https://doi.org/10.1109/cvprw56347.2022.00309

[16] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., *et al*. (2020) UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. *ICASSP* 2020-2020 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), Barcelona, 4-8 May 2020, 1055-1059. https://doi.org/10.1109/icassp40776.2020.9053405

[17] Chen, J., Lu, Y., Yu, Q., *et al*. (2021) TransuNet: Transformers Make Strong Encoders for Medical Image Segmentation.

[18] Xu, G., Zhang, X., He, X. and Wu, X. (2023) Levit-UNet: Make Faster Encoders with Transformer for Medical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 42-53. https://doi.org/10.1007/978-981-99-8543-2_4

[19] 李丹. 自适应特征融合的轻量级多模态医学图像分割模型研究[D]: [硕士学位论文]. 广州: 广州大学, 2024.