中北大学学报(自然科学版)
JOURNAL OF NORTH UNIVERSITY OF CHINA(NATURAL SCIENCE EDITION)
2026年 第47卷 第3期(总第227期)
Vol. 47 No. 3 2026(Sum No. 227)

文章编号：1673-3193(2026)03-0385-10 J. North Univ. China, Nat. Sci. Ed., 2026, 47(3): 385-394.

# 轻量级动态多尺度结肠息肉分割模型

刘泽生[1]，余本国[2*]，刘泽晋[1]

(1. 中北大学 软件学院，山西 太原 030051；
2. 海南医科大学 智能医学与技术学院（大数据研究中心），海南 海口 571199)

**摘　要：**结肠息肉的精准分割对结直肠癌的早期筛查和临床诊断具有重要意义。然而，由于息肉在尺寸、形态及边界特征上的显著差异，且常伴随低对比度与边界模糊等问题，现有方法在分割精度与计算效率之间难以取得平衡。针对上述问题，本文提出了一种轻量级动态多尺度分割网络LDMS-Net。该模型基于U形编码器-解码器结构，在编码阶段引入动态多尺度卷积模块(DMM)，通过全局平均池化与两层全连接网络生成自适应权重，对不同膨胀率(1,2,3)及标准卷积分支进行动态加权融合，实现多尺度特征的自适应建模；同时采用结构重参数化策略，在推理阶段将多分支结构融合为单一3×3卷积，以降低计算复杂度。为进一步提升分割性能，设计区域细节聚合器(RDA)和全局上下文整合器(CCI)，分别用于强化局部纹理信息与建模全局语义依赖，并在解码阶段通过层次后期融合策略逐层整合多尺度特征。实验在Kvasir-SEG和CVC-ClinicDB等数据集上进行，在仅使用Kvasir-SEG 800张训练图像的条件下，所提方法在测试集上取得93.65%的平均Dice系数和89.39%的平均交并比，参数量为$6.54\times10^6$，FLOPs为$7.23\times10^9$。结果表明，该方法在保证轻量化的同时具有较高的分割精度和良好的鲁棒性。

**关键词：**结肠息肉分割；轻量化；动态多尺度；特征提取
**中图分类号：**TP391 **文献标识码：**A **doi:** 10.62756/jnuc.issn.1673-3193.2025.09.0016
**引用格式：**刘泽生，余本国，刘泽晋. 轻量级动态多尺度结肠息肉分割模型[J]. 中北大学学报(自然科学版)，2026，47(3)：385-394.
LIU Zesheng, YU Benguo, LIU Zejin. Segmentation model of colon polyp with lightweight dynamic multi-scale[J]. Journal of North University of China(Natural Science Edition), 2026, 47(3): 385-394.

# Segmentation Model of Colon Polyp with Lightweight Dynamic Multi-Scale

LIU Zesheng[1], YU Benguo[2*], LIU Zejin[1]

(1. School of Software, North University of China, Taiyuan 030051, China;
2. School of Intelligent Medicine and Technology (Big Data Research Center),
Hainan Medical University, Haikou 571199, China)

**Abstract:** Precise segmentation of colon polyps is crucial for early screening and clinical diagnosis of colorectal cancer. However, large variations in polyp size, shape, and boundary characteristics, along with low contrast and blurred edges, make it challenging to balance segmentation accuracy and computational efficiency. To address these issues, a lightweight dynamic multi-scale segmentation network (LDMS-Net) was proposed. The model adopted a U-shaped encoder-decoder architecture and introduced a dynamic multi-scale module (DMM) in the encoder. Specifically, global average pooling and a two-layer fully connected network were

**收稿日期：**2025-09-25
**作者简介：**刘泽生(1998—)，男，硕士生，主要从事图像分割方面的研究。
*通信作者：余本国(1976—)，男，副教授，博士，主要从事图像处理、人工智能与大数据方面的研究。E-mail：120487362@qq.com。

used to generate adaptive weights, enabling dynamic fusion of multiple convolution branches with different dilation rates (1, 2, 3) and a standard convolution branch for adaptive multi-scale feature representation. A structural re-parameterization strategy was further applied to merge the multi-branch structure into a single 3×3 convolution during inference to reduce computational cost. In addition, a regional detail aggregator (RDA) and a comprehensive context integrator (CCI) were designed to capture local texture details and global semantic dependencies, respectively, and were integrated in the decoder via a hierarchical late fusion strategy. Experiments on Kvasir-SEG and CVC-ClinicDB datasets show that, using only 800 training images from Kvasir-SEG, the proposed method achieves a mean Dice coefficient of 93.65% and a mean IoU of 89.39%, with $6.54\times10^6$ parameters and $7.23\times10^9$ floating-point operations. The results demonstrate that LDMS-Net achieves high accuracy and robustness while maintaining a lightweight design.

**Key words**: colon polyp segmentation; lightweight; dynamic multi-scale; feature extraction

## 0 引 言

医学影像分割在疾病诊断与治疗规划中扮演着关键角色，通过精准勾画病灶轮廓和定位病变区域，为临床医生提供辅助的诊断依据[1-2]。结肠息肉作为结直肠癌的主要前兆病变，其准确分割对于早期诊断和干预至关重要，有助于降低结直肠癌的发病率和死亡率[3]。然而，结肠息肉在尺寸、形态和边界特征上的复杂性为分割任务带来了显著挑战[4]。息肉尺寸从数毫米到数厘米不等，且常伴随模糊边界或与周围组织对比度低的问题。此外，临床计算机辅助诊断(Computer-Aided Diagnosis, CAD)系统对实时性和低计算资源的需求进一步增加了模型设计的难度。早期结肠息肉分割主要依赖传统图像处理与机器学习方法，如基于主动轮廓模型(ACM)[5]、图割(Graph Cuts)[6]、GrabCut[7]以及随机森林结合手工特征(RF+HF)等。这些方法在简单背景下可取得一定效果，但面对息肉形态多样、边界模糊及低对比度等复杂情况时，分割精度往往受到明显限制，且高度依赖人工设计的特征与后处理，难以满足临床实时性与鲁棒性需求。现有分割方法在处理复杂形态息肉时，常因多尺度特征提取不足或局部细节与全局语义整合不佳，导致过分割、欠分割或边界模糊，限制了其在资源受限环境中的应用。

深度学习技术的快速发展为医学影像分割提供了新的解决方案。基于卷积神经网络(Convolutional Neural Network, CNN)的模型[8]，例如：UNet[9]因其对称的编码器-解码器结构和跳跃连接设计，在捕捉多尺度特征方面表现出色，成为医学影像分割的基准方法。UNet的成功催生了多种变体，如UNet++[10]、ResUNet[11]和Attention UNet[12]，通过引入残差连接或注意力机制提升了模型的鲁棒性。然而，这些模型受限于卷积操作的局部感受野，难以有效建模长程依赖关系[13]，导致在处理复杂形态的息肉时性能受限。为克服这一局限，研究人员尝试将视觉变换器(Vision Transformer, ViT)引入医学影像分割[14]。ViT通过全局自注意力机制建模长程空间关系，在图像分类任务中取得了显著成果。基于Transformer的模型，例如：TransUNet[15]和Swin-UNet[16]通过结合CNN和ViT的优势，改善了全局上下文信息的捕获能力。然而，这些模型通常需要高计算资源和大规模预训练数据支持，参数量和计算复杂度较高，例如：TransUNet的参数量常超过$100\times10^6$，计算量达$10\times10^9$，难以满足临床环境中对轻量化模型的需求[17]。

多尺度特征提取作为医学影像分割的核心技术，其核心挑战在于如何有效处理目标区域显著的尺寸变异问题。传统解决方案主要依靠特征金字塔网络(FPN)[18]或扩张卷积操作来扩展感受野，此类方法虽能捕获多尺度特征，但在局部纹理细节与全局语义信息的协同整合方面仍显不足。当前主流框架广泛采用多尺度策略模块，例如：ASPP[19]和FPN通过并行卷积或特征融合机制提升了模型对异尺度目标的识别能力。值得注意的是，局部纹理信息直接影响分割边界的清晰度，而全局语义线索则决定着病灶定位的准确性，二者的融合成为提升分割性能的关键。近年来，动态特征提取方法[20]通过引入可自适应调整的卷积核参数或注意力权重机制，展现出更强的多尺度上下文建模能力。然而这类方法普遍依赖复杂的模块设计，导致计算成本显著增加，难以适配低功耗设备的实时处理需求。在此背景下，结构重参数化技术(如RepVGG[21])为实现模型效

率优化提供了新思路，通过训练阶段的多分支设计与推理阶段的单路径转换，有效平衡了性能与计算成本的矛盾。当前研究的热点正聚焦于如何在维持模型轻量化的前提下，构建高效的局部-全局特征协同机制，这对推动精准医疗影像分析系统的实际部署具有重要价值。

针对上述结肠息肉分割面临的尺度剧烈变化、边界模糊以及临床部署对轻量化的严苛要求，本文提出轻量级动态多尺度分割网络（Lightweight Dynamic Multi-scale Segmentation Netowrk，LDMS-Net），主要贡献如下：

1）提出动态多尺度卷积模块（DMM），将输入自适应的动态核选择机制与多膨胀率深度卷积相结合，并在推理阶段通过结构重参数化将多分支动态结构完全融合为单个普通的3×3卷积，在实现多尺度高表达能力的同时将推理FLOPs降至$7.23\times10^9$，彻底解决传统静态多尺度模块与动态方法无法兼顾精度与速度的矛盾。

2）设计层次后期融合（Hierarchical Late Fusion）策略，在解码器每一层将局部细节聚合器（RDA）提供的精细纹理与全局上下文整合器（CCI）提供的长程语义进行逐层融合，显著优于传统的早期或中期融合方式，有效缓解了高级语义与低级细节间的语义鸿沟，确保边界清晰度和整体一致性。

3）在仅使用Kvasir-SEG 800张训练图像的常用跨数据集协议下，LDMS-Net以$6.54\times10^6$的参数量获得mDice 93.65%、mIoU 89.39%，在同等轻量级方法中达到当前最佳水平，表明其具备较高的临床应用潜力。

## 1　方　法

### 1.1　整体架构

本文提出的轻量级动态多尺度分割网络LDMS-Net整体架构如图1所示。输入图像首先经U形编码器逐级提取多尺度特征，随后通过区域细节聚合器（Regional Detail Aggregator，RDA）和全局上下文整合器（Comprehensive Context Integrator，CCI）分别捕获局部纹理细节与全局语义信息，最后由解码器采用层次后期融合策略将两者逐层整合并上采样恢复至原分辨率，输出最终分割结果。

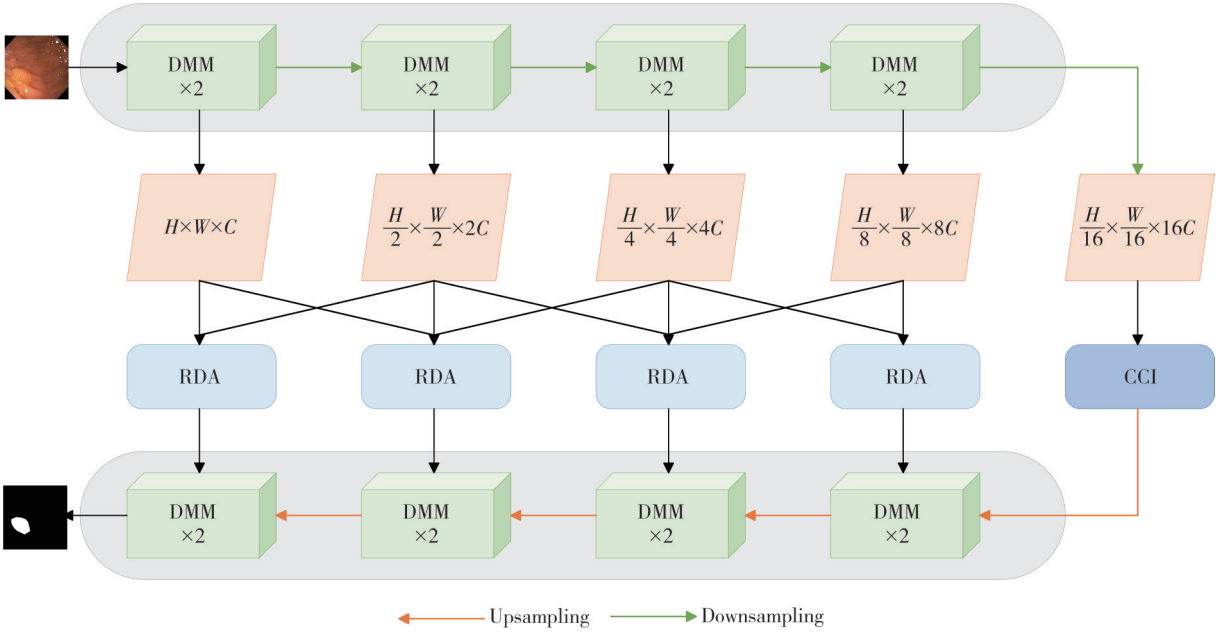



图1　LDMS-Net整体架构

Fig. 1　Overall architecture of the LDMS-Net

编码器包含4个阶段，每阶段由两个动态多尺度卷积模块（DMM）构成，相邻阶段间通过步幅为2的下采样将空间分辨率减半、通道数加倍。对于输入尺寸为$H\times W\times3$的图像，第$k$阶段输出特征图尺寸为$\frac{H}{2^{i-1}}\times\frac{W}{2^{i-1}}\times2^{i-1}C$，从而形成丰富的多尺度特征金字塔。

为进一步缓解边界模糊与过/欠分割问题，本文在编码器与解码器之间引入两个专用模块：

1）区域细节聚合器（RDA，共4个，分别位于4个上采样阶段之前）：每个RDA接收所在阶段及其上一阶段的特征图，构建局部特征金字塔，并

通过局部窗口自注意力捕获精细纹理、弥合高低层语义差距；

2）全局上下文整合器(CCI，仅1个，位于编码器最深层之后)：接收编码器全部5个层级(包括输入图像)的特征图，构建全局特征金字塔，并通过Transformer捕获长程依赖，实现精准病灶定位。

解码器采用层次后期融合(Hierarchical Late Fusion)策略：在每个上采样阶段，先将对应RDA输出的局部细节特征与CCI提供的全局语义特征逐通道拼接或加权融合，再送入上采样模块，从而在每一层级都同时保留清晰边界与整体一致性，最终输出与输入同分辨率的分割概率图。

LDMS-Net选用U形结构而非纯Transformer或非对称架构，是因为实验表明U形编码器-解码器在保持轻量化的同时，能最有效地实现多尺度特征、局部细节与全局语义的协同融合。若改为纯Transformer架构，虽全局建模能力更强，但参数量与计算量将显著增加，难以满足临床实时需求。

## 1.2 DMM

本文提出的动态多尺度卷积模块(Dynamic Multi-scale Convolution Module，DMM)完整前向流程如图2所示。输入特征图首先通过全局平均池化(GAP)生成通道描述符向量$z$，并由此计算动态权重$\alpha_1 \sim \alpha_4$；同时，输入特征被送入4个膨胀率分别为1,2,3以及一个5×5的并行深度卷积核进行多尺度特征提取；随后4个分支输出按对应$\alpha_i$加权融合，再经批归一化BN、SE通道注意力和1×1卷积增强后，与输入残差连接得到最终输出。在推理阶段，所有分支与动态权重通过结构重参数化融合为单个普通的3×3卷积，实现训练时高性能与推理时高效率的统一。为清晰起见，图中仅展示一个通道进行说明。实际中，深度卷积独立应用于每个输入通道，动态权重在所有通道间共享。

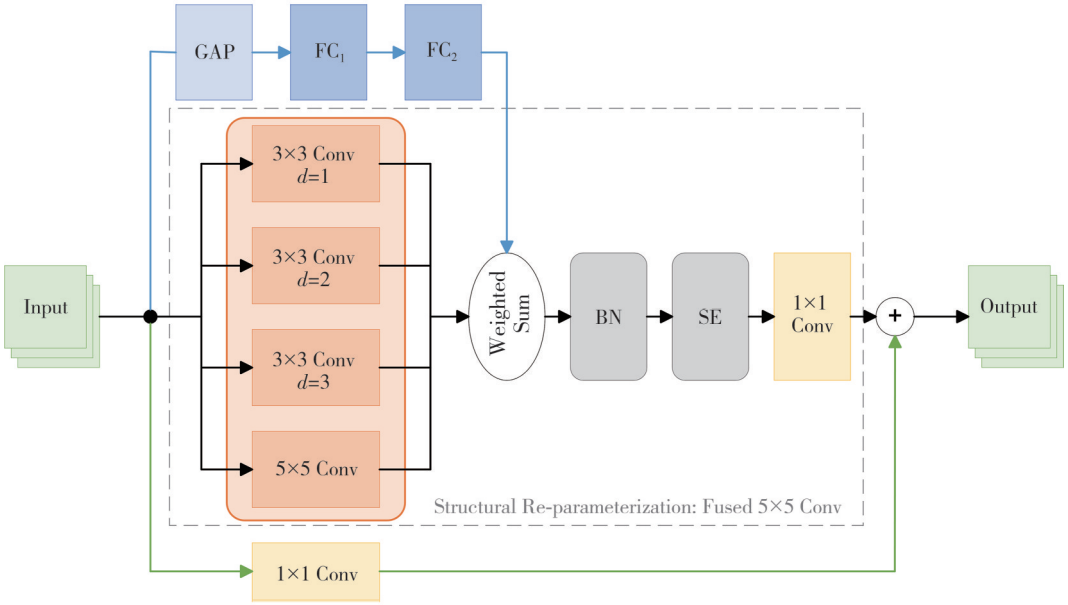



图2 动态多尺度卷积模块

Fig. 2 Dynamic multi-scale convolution module

下面将按信号流顺序依次详细阐述动态核选择机制、深度卷积核集合、特征增强与融合、残差连接与结构重参数化4个组成部分的设计与公式。

### 1.2.1 动态核选择机制

动态核选择机制通过根据输入特征图的全局信息生成卷积核权重，实现自适应多尺度特征提取。该机制显著提升了模块对不同医疗图像模态(如CT、MRI、超声)和目标尺度(如小息肉或大肿瘤)的适应性，是DMM区别于其余多分支模块的关键组件。该机制包括全局特征提取和权重生成两个步骤。

对输入特征图$F \in \mathbf{R}^{C_{in} \times H \times W}$进行全局平均池化(Global Average Pooling，GAP，用于将每个通道的空间信息压缩为一个标量)，生成通道描述符为

$$z = \frac{1}{H \times W} \sum_{h=1}^{H} \sum_{\omega=1}^{W} F_c(h, \omega), c = 1, 2, \ldots, C_{in}, \quad (1)$$

其中，输入特征图$F \in \mathbf{R}^{C_{in} \times H \times W}$，表示具有$C_{in}$个通道和空间分辨率为$H \times W$的特征张量。全局平均池化通过对每个通道的空间维度求均值，生成通道描述符$z \in \mathbf{R}^{C_{in}}$，其第$c$个分量$z_c$为输入特征图第$c$通道的空间均值，在图2中体现在GAP模块的

输出处，用于捕捉整个特征图的全局上下文信息。

通过两层全连接网络处理$z$（其分量$z_c$对应各通道描述符），生成4个卷积核的权重为

$$\alpha = Softmax\Big(FC_2(ReLU(FC_1(z)))\Big), \quad (2)$$

其中，通道描述符$z$首先通过第一层全连接网络$FC_1: \mathbf{R}^{C_{in}} \to \mathbf{R}^{C_{in}/\gamma}$降维，其中缩放因子$\gamma = 8$。随后，ReLU激活函数引入非线性。第二层全连接网络$FC_2: \mathbf{R}^{C_{in}/\gamma} \to \mathbf{R}^4$生成4个权重。Softmax操作归一化权重，确保$\sum_{i=1}^{4}\alpha_i = 1$，输出$\alpha = [\alpha_1, \alpha_2, \alpha_3, \alpha_4] \in \mathbf{R}^4$，每个$\alpha_i$表示对应卷积核的贡献度。

1.2.2　深度卷积核集合

深度卷积核集合是DMM多尺度特征提取的基础，通过4个不同感受野的卷积核捕获从局部细节到全局上下文的特征。DMM引入膨胀卷积以扩大感受野并减少冗余，是模块性能提升的关键组件。

DMM设计了4个并行深度卷积核，以捕获不同尺度的空间特征。这些核包括：第1个$3\times3$卷积核，扩张率为1，生成$3\times3$的感受野，专注于局部细节；第2个$3\times3$卷积核，扩张率为2，形成$7\times7$的感受野，适用于中尺度上下文；第3个$3\times3$卷积核，扩张率为3，扩展至$11\times11$的感受野，针对大尺度目标；一个$5\times5$卷积核，产生$5\times5$的感受野，平衡局部和中尺度特征。这一组合通过膨胀卷积显著增强了多尺度特征提取能力，适应医疗图像分割中多样化的目标尺度。每个核对输入特征图执行深度卷积，表示为

$$F_i = F \otimes K_i, i = 1, 2, 3, 4, \quad (3)$$

式中：深度卷积核$K_i \in \mathbf{R}^{C_{in}\times k_h \times k_w \times C_{out}}$，独立作用于每个输入通道，其中$k_h$和$k_w$为核尺寸，输出通道数$C_{out} = C_{in}$，保持通道数不变。

1.2.3　特征增强与融合

特征增强与融合阶段将并行卷积的输出通过动态加权融合为单一特征图，并通过批归一化（BN）、Squeeze-and-Excitation（SE）层和$1\times1$逐点卷积进一步优化特征表达。该部分可在动态核选择和深度卷积的基础上提升特征质量，是DMM性能优化的重要环节。

4个卷积输出通过动态权重$\alpha_i$进行加权求和（在图2中体现为4个深度卷积分支输出后、加权求和前的乘法操作，$i$=1,2,3,4），即

$$F' = \sum_{i=1}^{4}\alpha_i(F \otimes K_i), \quad (4)$$

其中，$\alpha_i \in \mathbf{R}$来自式（2），控制第$i$个卷积核的贡献度。每个卷积输出$F \otimes K_i$包含特定尺度的特征，融合特征图$F'$整合多尺度信息。

对融合特征图应用BN进行归一化操作，如式（5）所示。其中，运行均值$\mu \in \mathbf{R}^{C_{out}}$和方差$\delta^2$用于统计每个通道的分布。

$$F'' = BN(F') = \frac{\eta}{\sqrt{\delta^2 + \varepsilon}}(F' - \mu) + \beta。 \quad (5)$$

SE层用于增强通道依赖性，分别为

$$s = Conv_2\Big(ReLU(Conv_1(GAP(F'')))\Big), \quad (6)$$

$$F''' = s \cdot F''。 \quad (7)$$

$1\times1$逐点卷积用于调整通道维度，表示为

$$O = Conv_{1\times1}(F''')。 \quad (8)$$

1.2.4　残差连接与结构重参数化

残差连接通过残差结构优化梯度传播，结构重参数化在推理时融合动态核为单一卷积核，确保轻量级部署。

残差连接通过$1\times1$卷积处理输入特征图，与主路径输出相加。推理时融合动态核和BN参数为单一卷积核，具体为

$$K_{fused} = \sum_{i}^{4}\bar{\alpha}_i K_i \cdot \frac{\eta}{\sqrt{\delta^2 + \varepsilon}}, \quad (9)$$

$$b_{fused} = \beta - \mu \cdot \frac{\eta}{\sqrt{\delta^2 + \varepsilon}}, \quad (10)$$

其中，动态权重值$\bar{\alpha}_i$用于加权深度卷积核$K_i$，$3\times3$核通过零填充扩展至$5\times5$以对齐$5\times5$核$K_4$。

### 1.3　区域细节聚合器和全局上下文整合器

多尺度特征金字塔在密集预测任务中的有效性已被多项研究验证。本文设计了一种基于Transformer的全局上下文整合器（CCI）模块，作为编码器与解码器之间的桥梁，以融合编码器的多尺度特征金字塔。通过充分利用金字塔层次结构中的多尺度特征，CCI能够在不同尺度高效捕获全局上下文语义。为优化特征金字塔处理，提出了一种卷积嵌入方法替代标准补丁嵌入，包含重塑、卷积、拼接及平铺操作。如图3下方所示，CCI以编码器5个阶段的特征图构成的全局特征金字塔作为输入。这些特征图被重塑至目标尺寸，并通过$3\times3$卷积聚合。卷积输出被平铺为向量后输入Transformer模块。此设计显著降低了计算复杂性，并弥补了Transformer缺乏局部归纳偏见的不足，从而提升了细粒度细节捕获能力。

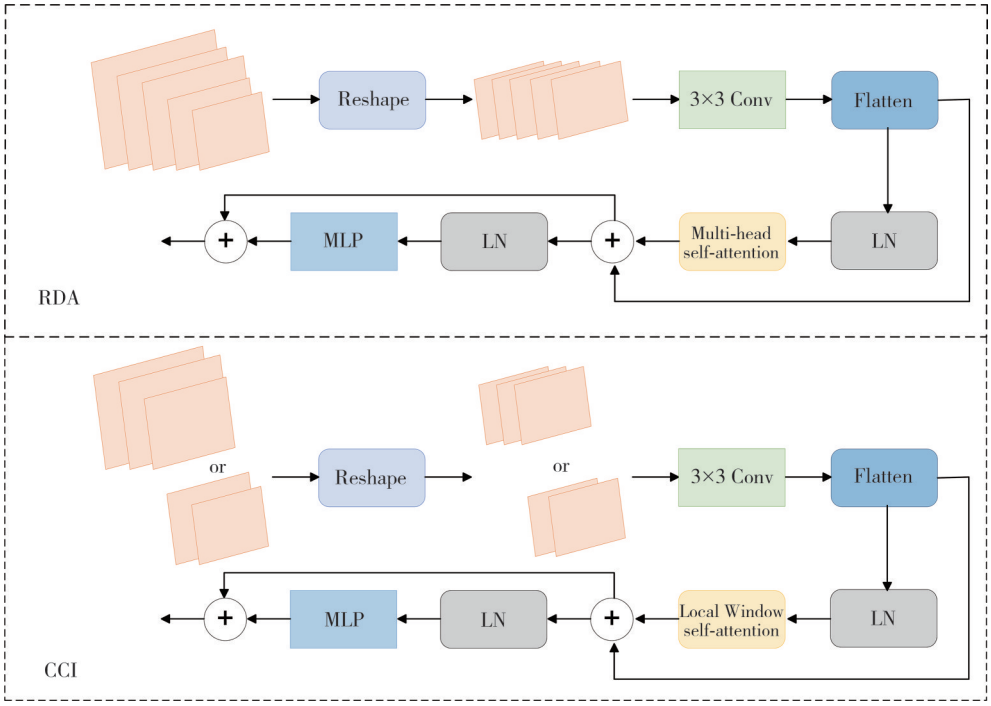



图3 区域细节聚合器和全局上下文整合器

Fig. 3 Regional detail aggregator and comprehensive context integrator module

具体而言，编码器特征图$x_1,x_2,x_3,x_4,x_5$的分辨率相对于输入图像分别为$1,\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{16}$。特征金字塔$X=\{x_1,x_2,x_3,x_4,x_5\}$被下采样至目标尺寸$\frac{H}{S}\times\frac{W}{S}$(其中$S$为相对于输入图像的步幅，如$\frac{H}{16}\times\frac{W}{16}$)。重塑后的金字塔沿通道维度拼接，并通过3×3卷积聚合多尺度特征，最终平铺为向量(序列)。

$$\boldsymbol{X}'=Flatten\Big(Conv(Concat(\mathrm{Reshape}(X)))\Big)。\tag{11}$$

为生成具有全局上下文感知能力的语义信息，平铺的特征图$\boldsymbol{X}'$被输入至Transformer模块。该模块包含多头自注意力(Multi-Head Self-Attention，MHSA)、两层多层感知器(MLP)、两个层归一化(LayerNorm，LN)层及两个残差连接。Transformer模块的处理分别表示为

$$\boldsymbol{X}_a=\boldsymbol{X}'+MHSA(LN(\boldsymbol{X}')),\tag{12}$$

$$\boldsymbol{X}_o=\boldsymbol{X}_a+MLP(LN(\boldsymbol{X}_a)),\tag{13}$$

其中，多头自注意力(MHSA)由查询(Q)、键(K)和值(V)三部分构成，各头的维度由输入序列长度和通道数决定。自注意力机制估计表示为

$$Att=Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V,\tag{14}$$

式中：$d=\frac{C}{M}$为头维度，$M$为头数。输出$X$被重塑为$C$，并传输至解码器。

与CCI类似，设计了一种轻量高效的区域细节聚合器(RDA)模块。如图3上方所示，RDA基于局部自注意力机制，具备局部归纳偏见和线性计算复杂性。为弥合高层次与低层次特征间的语义差距，编码器相邻阶段的输出被组合，构建局部特征金字塔作为RDA输入。局部特征金字塔的层数依据RDA所在阶段而定，第一和第四阶段包含两层，第二和第三阶段包含三层。金字塔特征根据RDA阶段重塑至目标尺寸，沿通道维度拼接，并通过3×3卷积聚合，随后，所得特征输入Transformer模块进一步处理。

此设计具有以下优势：1)局部特征金字塔整合了多尺度空间和语义信息，对医学影像分割至关重要。2)RDA的局部归纳偏见使其能够提供细致边缘和纹理信息，辅助解码器恢复完整空间分辨率。3)解码器通过层次后期融合策略，聚合RDA的局部纹理信息与CCI的全局语义信息，生成精准的分割结果。此策略通过逐层整合多尺度表征，确保了分割边界清晰性和整体一致性。凭借CCI与RDA的协同作用，LDMS-Net在医学影像分割任务中表现出优异性能，在多个基准数据集上展现出较强竞争力。

## 2　实　验

### 2.1　实验数据集

为验证LDMS-Net在结肠息肉分割任务中的有效性，本文选用了5个公开基准数据集：Kvasir-SEG[22]（100张）、CVC-ColonDB[23]（380张）、CVC-ClinicDB[24]（612张）、ETIS-LaribPolypDB[25]（196张）和CVC-300（60张）。这些数据集的真值分割掩码均由专业内镜医师或放射科医生手工精细标注，公开发布的数据集已被结肠息肉分割领域广泛认可并作为标准参考。

为保证公平比较与可复现性，本文严格遵循该领域公认的标准训练/测试划分协议：训练集仅使用Kvasir-SEG的800张训练图像（官方划分），其余4个数据集（CVC-ClinicDB 612张、CVC-300 60张、CVC-ColonDB 380张、ETIS-LaribPolypDB 196张）全部作为独立测试集，用于评估模型跨数据集的泛化性能；Kvasir-SEG剩余的200张图像作为同数据集测试集。

### 2.2　评估指标

为了评估LDMS-Net模型的性能，采用平均Dice相似系数（mDice）、平均交并比（mIoU）、精确率（Precision）、召回率（Recall）、准确率（Accuracy）、参数量以及FLOPs作为评估指标，部分计算公式分别为

$$mDice=\frac{1}{N}\sum_{i=1}^{N}\frac{2TP}{2TP+FP+FN},\tag{15}$$

$$mIoU=\frac{1}{N}\sum_{i=1}^{N}\frac{TP}{TP+FN+FP},\tag{16}$$

$$Precision=\frac{TP}{TP+FP},\tag{17}$$

$$Recall=\frac{TP}{TP+FN},\tag{18}$$

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN}。\tag{19}$$

式中：$TP$为正确预测的息肉区域的像素数；$FP$为错误预测为息肉区域的背景区域的像素数；$TN$为正确预测的背景区域的像素数；$FN$为错误预测为背景区域的息肉区域的像素数。

### 2.3　实验配置

实验在配备NVIDIA RTX 4080 GPU（16 GB显存）的计算平台上进行，基于PyTorch框架实现。

训练过程中，输入图像统一调整为256×256像素的分辨率，批次大小设为16。优化器采用AdamW，初始学习率设为$1\times10^{-4}$。训练轮数设为200轮，每10轮保存一次模型检查点。为增强模型泛化能力，数据增强策略包括随机翻转、旋转、缩放和亮度调整。损失函数采用Dice损失与交叉熵损失的组合，以优化分割精度与模型收敛性。动态核选择器的初始温度参数$\tau$设为1.0，并以每50轮衰减0.1的速率逐渐锐化权重分布。CCI模块的多头自注意力机制设置8个头，RDA模块的局部窗口大小设为7×7，以平衡计算效率与特征表达能力。

### 2.4　实验结果

#### 2.4.1　对比实验

将LDMS-Net与其他传统方法以及先进方法在Kvasir-SEG和CVC-ClinicDB数据集进行比较，包括Unet、Unet++、ResUnet、TransUnet、ssformer、ResUnet++、Swin-Unet、Deeplabv3+、HarDNet-MSEG、U²-Net。实验结果如表1和表2所示，其中最优指标值用加粗字体标示。

表1　不同模型在Kvasir-SEG数据集上的对比

Tab. 1　Comparative of different models on the Kvasir-SEG dataset

| 模型 | *mDice*/% | *mIoU*/% | 精确率/% | 召回率/% | 准确率/% | FLOPs/$10^9$ | Params/$10^6$ |
|---|---|---|---|---|---|---|---|
| Unet[9] | 88.71 | 83.68 | 86.69 | 85.53 | 94.97 | 55.36 | 32.13 |
| Unet++[10] | 90.47 | 86.81 | 87.98 | 87.54 | 95.08 | 35.31 | 9.23 |
| ResUnet[11] | 89.28 | 81.13 | 79.23 | 82.46 | 93.85 | 81.25 | 13.83 |
| TransUnet[12] | 83.77 | 75.65 | 73.07 | 75.98 | 96.13 | 37.84 | 87.16 |
| Ssformer[26] | 90.31 | 87.03 | **90.88** | 89.51 | 96.13 | 42.33 | 29.57 |
| ResUnet++[27] | 89.97 | 81.38 | 82.01 | 81.26 | 94.27 | 70.89 | 14.37 |
| Swin-Unet[16] | 80.12 | 70.89 | 55.38 | 68.07 | 87.01 | 11.49 | 41.67 |
| Deeplabv3+[28] | 93.31 | 88.97 | 85.79 | 86.79 | 95.48 | 14.57 | 7.24 |
| HarDNet-MSEG[29] | 90.43 | 86.50 | 90.85 | **89.91** | 95.23 | **6.09** | **3.51** |
| U²-Net[30] | 91.07 | 87.53 | 90.03 | 89.22 | 95.48 | 34.25 | 43.75 |
| LDMS-Net | **93.65** | **89.39** | 89.64 | 89.01 | **96.30** | 7.23 | 6.54 |

对于Kvasir-SEG数据集，LDMS-Net在平均Dice相似系数、平均交并比、准确率等指标上均居首位，分别达到了93.65%，89.39%，96.30%，优于其他模型，而其参数量和计算量仅为$6.54\times10^6$、$7.23\times10^9$，远小于其余模型。图4展示了LDMS-Net在5个数据集上的可视化典型分割结果，每个数据集中选取两例最具代表性的病例。从图中可以看出，LDMS-Net能够更准确地识别出息肉的轮廓形

状，其划分的边界更为清晰，且结果也更接近真实值，优于其他方法。

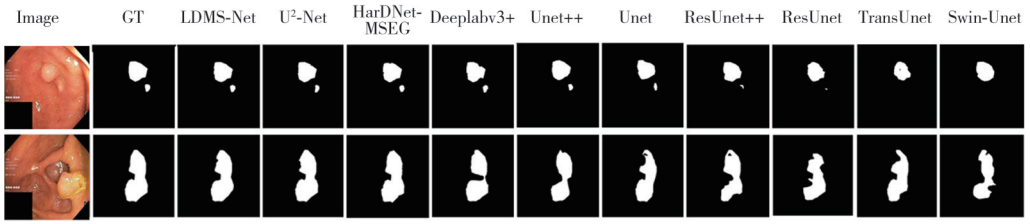
表2 不同模型在CVC-ClinicDB数据集上的对比

Tab. 2 Comparative of different models on the CVC-ClinicDB dataset

| 模型 | *mDice*/% | *mIoU*/% | 精确率/% | 召回率/% | 准确率/% | FLOPs/$10^9$ | Params/$10^6$ |
|---|---|---|---|---|---|---|---|
| Unet[9] | 89.92 | 85.02 | 87.72 | 86.95 | 96.43 | 55.36 | 32.13 |
| Unet++[10] | 91.50 | 87.93 | 89.03 | 88.71 | 96.46 | 35.31 | 9.23 |
| ResUnet[11] | 90.64 | 82.26 | 80.50 | 83.51 | 95.10 | 81.25 | 13.83 |
| TransUnet[15] | 84.87 | 76.81 | 75.43 | 77.01 | 97.45 | 37.84 | 87.16 |
| Ssformer[26] | 91.41 | 88.47 | 91.23 | 90.91 | 98.32 | 42.33 | 29.57 |
| ResUnet++[27] | 91.02 | 82.81 | 83.03 | 82.62 | 95.37 | 70.89 | 14.37 |
| Swin-Unet[16] | 81.30 | 71.94 | 56.74 | 69.41 | 88.26 | 11.49 | 41.67 |
| Deeplabv3+[28] | 94.57 | 90.08 | 86.87 | 88.11 | 96.64 | 14.57 | 7.24 |
| HarDNet-MSEG[29] | 91.22 | 88.09 | 91.62 | 90.79 | 95.78 | **6.09** | **3.51** |
| $U^2$-Net[30] | 92.31 | 89.25 | **92.00** | **91.57** | 96.08 | 34.25 | 43.75 |
| LDMS-Net | **94.75** | **90.17** | 90.76 | 90.50 | **98.41** | 7.23 | 6.54 |

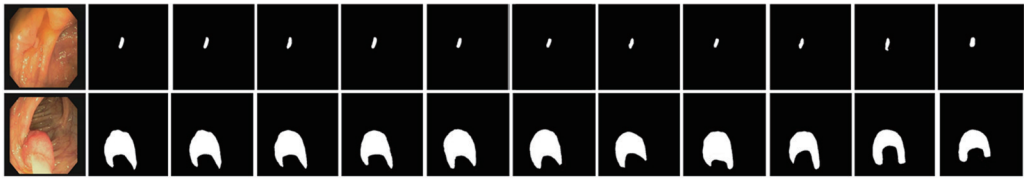
图4不仅展示出本方法对规则形状息肉的准确分割，也能够对不规则形状的息肉进行准确的识别和边界定位，这说明LDMS-Net在处理息肉分割任务时具有更强的能力。

综上，证明了本模型相较于其他方法的显著优势，无论在准确率还是边界区分或其他评价指标的表现上，都显示出其在息肉分割任务中的卓越性能。

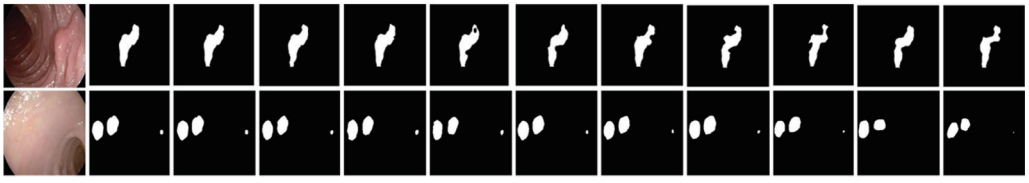



(a) Kvasir-SEG数据集典型分割结果

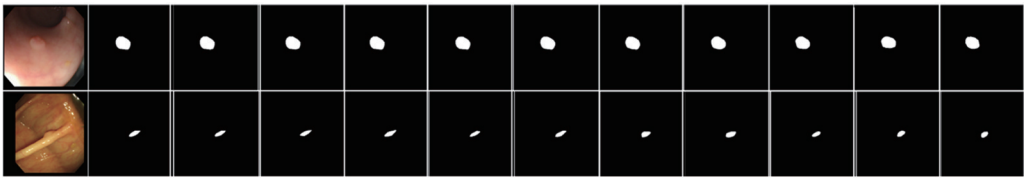
(b) CVC-ColonDB数据集典型分割结果



(c) CVC-ClinicDB数据集典型分割结果



(d) ETIS-LaribPolypDB数据集典型分割结果



(e) CVC-300数据集典型分割结果

图4 不同数据集上的可视化分割结果

Fig. 4 Visualization of segmentation results on different datasets

2.4.2　消融试验

表 3 使用 Unet 作为基线，显示了每个模块对模型性能的影响。添加多分支模块后，mIoU 提高了 2.09 百分点，展示了其在提取多尺度特征以提高泛化能力方面的有效性。

此外，将 RDA 和 CCI 结合起来可以增强表征学习。RDA 和 CCI 的结合实现了最佳的鲁棒性，突出了集成局部纹理和密集预测任务的全局语义。

表 3　每个模块的影响

Tab. 3　Impact of each module

| Unet | 多分支 | RDA | CCI | *mDice*/% | *mIoU*/% |
|---|---|---|---|---|---|
| √ | | | | 89.76 | 82.68 |
| √ | √ | | | 91.34 | 84.77 |
| √ | √ | √ | | 92.01 | 86.40 |
| √ | √ | | √ | 91.93 | 86.18 |
| √ | √ | √ | √ | 93.57 | 89.15 |

表 4 呈现了 LDMS-Net 模块设计选择的细粒度消融实验结果。当所有膨胀率设置为 1 时，mDice 和 mIoU 均显著下降（分别降低 1.55 百分点和 2.37 百分点），这证实了多尺度膨胀卷积对捕获尺度多样性的重要性。将膨胀集合扩展至[1,2,4]虽能略微提升大病灶的分割性能，但未超越基线配置[1,2,3]，表明过大的感受野会产生收益递减效应。类似地，将 5×5 分支替换为纯 3×3 卷积会降低分割精度，而 7×7 分支仅带来微弱的性能提升却需额外计算量和参数开销，因此验证了 5×5 卷积核是实现精度与效率平衡的最佳选择。

表 4　模块级细粒度消融试验

Tab. 4　Fine-grained ablation study on module-level

| 配置 | *mDice*/% | *mIoU*/% | FLOPs/$10^9$ | Params/$10^6$ |
|---|---|---|---|---|
| 基线 | 93.65 | 89.39 | 7.23 | 6.54 |
| 膨胀率[1,1,1] | 92.10 | 87.02 | 7.23 | 6.54 |
| 膨胀率[1,2,4] | 93.21 | 88.60 | 7.23 | 6.54 |
| 3×3 分支 | 93.03 | 88.39 | 7.05 | 6.38 |
| 7×7 分支 | 93.78 | 89.52 | 8.00 | 7.10 |

2.4.3　鲁棒性与部署效率验证

为有力验证 LDMS-Net 在目标尺度剧烈变化以及计算资源受限场景下的独特优势，本文分别针对尺度鲁棒性和嵌入式部署效率进行了专项实验。

首先，按息肉真实面积将 Kvasir-SEG 测试集（200 张）划分为小（<200 $px^2$）、中（200～1 000 $px^2$）、大（>1 000 $px^2$）3 类，与 4 种代表性方法对比，结果如表 5 所示。LDMS-Net 在临床最难的小息肉上 mDice 提升尤为显著（提升 3.1 百分点），充分体现了 DMM 动态多尺度机制的优越性。

其次，在两款临床常见的嵌入式平台（NVIDIA Jetson NX 15W、Jetson Nano 10W）上测试 384×384 输入的推理速度与显存占用（FP32），结果如表 6 所示。得益于推理阶段结构重参数化，LDMS-Net 在 Jetson NX 上达到 85 帧·$s^{-1}$、在更受限的 Nano 上仍保持 30 帧·$s^{-1}$（实时），显存占用最低，验证了其在资源受限设备上的显著部署优势。

表 5　不同尺度息肉分割性能对比

Tab. 5　Segmentation performance on polyps of different scales

| 方法 | mDice/% | | | |
|---|---|---|---|---|
| | Small | Medium | Large | Overall |
| Unet | 81.2 | 91.5 | 94.3 | 89.1 |
| Unet++ | 83.7 | 92.1 | 94.8 | 90.2 |
| HarDNet-MSEG | 85.4 | 93.2 | 95.1 | 91.3 |
| CaraNet | 87.1 | 94.0 | 95.6 | 92.4 |
| LDMS-Net | 90.2 | 95.3 | 96.2 | 93.65 |

表 6　嵌入式平台推理速度与资源占用

Tab. 6　Inference speed and resource consumption on embedded platforms

| 方法 | Params/$10^6$ | FLOPs/$10^9$ | Jetson NX/（帧·$s^{-1}$） | Jetson Nano/（帧·$s^{-1}$） | Memory/MB |
|---|---|---|---|---|---|
| Unet | 31.0 | 28.4 | 31 | 12 | 1 823 |
| HarDnet-MSEG | 8.7 | 9.8 | 68 | 24 | 913 |
| CaraNet | 12.3 | 11.2 | 59 | 21 | 1 107 |
| LDMS-Net | 6.54 | 7.23 | 85 | 30 | 784 |

## 3　结　论

针对息肉尺寸多变、边界模糊及计算资源有限等挑战，本文提出了一种轻量级动态多尺度分割网络 LDMS-Net。通过融合动态多尺度特征提取与局部-全局特征聚合机制，LDMS-Net 实现了精准高效的结肠息肉分割。在多组数据集上的实验表明，该网络在保持低计算复杂度的同时，各项性能指标均优于代表性基线方法。从理论角度分析，解码器对不同编码器层的差异化利用具有显著效果：浅层主要编码高频边界细节，深层则捕获低频语义语境。后期融合策略使这些互补信息在预测阶段得以保留与整合，从而同时确保边界锐利度与病灶定位鲁棒性。这些成果彰显了其

在临床计算机辅助诊断系统中的部署潜力。未来工作将探索该网络在其他医学图像分割任务中的扩展应用,并进一步提升其实时处理能力。

## 利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors declare no relevant conflict of interests.

## 参考文献:


[1] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. Medical Image Analysis, 2017, 42: 60-88.

[2] DI J, ZHU Y, LIANG C. A medical image segmentation model based on SAM with an integrated local multiscale feature encoder[J]. Journal of Measurement Science and Instrumentation, 2025, 16(3): 359-370.

[3] NIIKURA R, HIRATA Y, SUZUKI N, et al. Colonoscopy reduces colorectal cancer mortality: A multicenter, long-term, colonoscopy-based cohort study [J]. PLoS One, 2017, 12(9): e0185294.

[4] BERNAL J, SÁNCHEZ F J, FERNÁNDEZ-ESPARRACH G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation *vs.* saliency maps from physicians [J]. Computerized Medical Imaging and Graphics, 2015, 43: 99-111.

[5] AMELING S, WIRTH S, PAULUS D, et al. Texture-based polyp detection in colonoscopy[C]// Bildverarbeitung für die Medizin 2009. Berlin, Heidelberg: Springer, 2009: 346-350.

[6] BERNAL J, SÁNCHEZ J, VILARIÑO F. Towards automatic polyp detection with a polyp appearance model[J]. Pattern Recognition, 2012, 45(9): 3166-3182.

[7] HWANG S, OH J, TAVANAPONG W, et al. Polyp detection in colonoscopy video using elliptical shape feature [C]//2007 IEEE International Conference on Image Processing, 2007: II-465-II-468.

[8] TAJBAKHSH N, SHIN J Y, GURUDU S R, et al. Convolutional neural networks for medical image analysis: Full training or fine tuning?[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1299-1312.

[9] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Cham: Springer, 2015: 234-241.

[10] ZHOU Z, 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, 2018: 3-11.

[11] ZHANG Z, LIU Q, WANG Y. Road extraction by deep residual U-net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753.

[12] OKTAY O, SCHLEMPER J, LE FOLGOC L, et al. Attention U-Net: Learning where to look for the pancreas [DB/OL]. (2018-05-20) [2025-09-25]. https://arxiv.org/abs/1804.03999.

[13] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7794-7803.

[14] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [DB/OL]. (2021-06-03) [2025-09-25]. https://arxiv.org/abs/2010.11929.

[15] CHEN J, LU Y, YU Q, et al. TransUNet: Transformers make strong encoders for medical image segmentation [DB/OL]. (2021-06-03) [2025-09-25]. https://arxiv.org/abs/2102.04306.

[16] CAO H, WANG Y, CHEN J, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation [C]//Computer Vision-ECCV 2022 Workshops. Cham: Springer, 2023: 205-218.

[17] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical transformer: Gated axial-attention for medical image segmentation [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer, 2021: 36-46.

[18] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 936-944.

[19] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[DB/OL]. (2017-12-05)[2025-09-25]. https://arxiv.org/abs/1706.05587.

[20] YANG B, BENDER G, LE Q V, et al. CondConv: Conditionally parameterized convolutions for efficient inference [DB/OL]. (2017-12-05) [2025-09-25]. https://arxiv.org/abs/1904.04971.


(下转第405页)