

# L-SwinUNet: 融合注意力增强与深度可分离卷积的轻量化分割模型

檀文文<sup>1,2</sup>, 卢 棚<sup>1</sup>, 姜 韦<sup>2</sup>

<sup>1</sup>合肥综合性国家科学中心能源研究院(安徽省能源实验室), 安徽 合肥

<sup>2</sup>安徽理工大学计算机科学与工程学院, 安徽 淮南

收稿日期: 2025年12月17日; 录用日期: 2026年1月26日; 发布日期: 2026年2月6日

## 摘要

针对现有基于Transformer的分割网络普遍存在参数冗余、计算复杂度高、推理效率低等问题, 本文提出一种轻量化医学图像分割网络L-SwinUNet。该网络从三个维度进行协同优化: 首先, 在编码器-解码器跳跃连接中嵌入无参数SimAM注意力模块, 自适应增强浅层空间语义与边界敏感特征; 其次, 在解码器中采用深度可分离卷积替代标准卷积, 以分离式特征提取策略大幅削减参数规模与浮点运算量; 最后, 在上采样阶段引入CARAFE内容感知重组算子, 通过自适应核预测机制精细重建高分辨率边缘细节。实验在Synapse数据集上验证了方法的有效性。结果表明, 相较原始Swin-UNet, L-SwinUNet在减少约48%参数与40%计算量的同时还提升了Dice与HD95指标, 证明其在医学图像分割中的轻量化优势与精度潜力。

## 关键词

医学图像分割, 注意力机制, 深度可分离卷积, 上采样优化

# L-SwinUNet: A Lightweight Segmentation Model that Combines Attention Enhancement with Deep Separable Convolutions

Wenwen Tan<sup>1,2</sup>, Peng Lu<sup>1</sup>, Wei Jiang<sup>2</sup>

<sup>1</sup>Institute of Energy, Hefei Comprehensive National Science Center (Anhui Energy Laboratory), Hefei Anhui

<sup>2</sup>School of Computer Science and Engineering, Anhui University of Science & Technology, Huainan Anhui

\*通讯作者。

Received: December 17, 2025; accepted: January 26, 2026; published: February 6, 2026

## Abstract

To address the prevalent issues of parameter redundancy, high computational complexity, and low inference efficiency in existing Transformer-based segmentation networks, this paper proposes a lightweight medical image segmentation network, L-SwinUNet. The network undergoes synergistic optimisation across three dimensions: firstly, embedding parameter-free SimAM attention modules within the encoder-decoder skip connections to adaptively enhance shallow spatial semantic and boundary-sensitive features; secondly, it employs depthwise separable convolutions within the decoder instead of standard convolutions, significantly reducing parameter size and floating-point operations through a separable feature extraction strategy; finally, it introduces the CARAFE content-aware reorganisation operator during the upsampling stage, which employs an adaptive kernel prediction mechanism to finely reconstruct high-resolution edge details. Experiments on the Synapse dataset validate the method's efficacy. Results demonstrate that compared to the original Swin-UNet, L-SwinUNet achieves approximately 48% fewer parameters and 40% reduced computational load while improving Dice and HD95 metrics, proving its lightweight advantages and accuracy potential in medical image segmentation.

## Keywords

Medical Image Segmentation, Attention Mechanism, Depthwise Separable Convolutions, Upsampling Optimisation

Copyright © 2026 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

医学图像分割是现代智能医疗系统中的关键环节，在靶区勾画、病灶识别、器官建模及术前规划等场景发挥着核心作用。随着深度学习的发展，卷积神经网络(Convolutional Neural Networks, CNN)在医学图像分割任务中取得了显著进展[1]。但传统 CNN 方法难以同时捕获局部细节与全局语义关系，尤其是在复杂解剖结构与多尺度器官的场景中，其性能仍受到一定限制。近年来，Transformer 结构凭借强大的建模能力被引入医学图像分割领域，其中 Swin Transformer 使用分层结构与窗口注意力机制[2]，既保持了高分辨率特征表达，又具备良好的全局交互能力。基于该结构的 Swin-UNet 已成为当前性能表现较为突出的 Transformer 架构医学图像分割模型之一[3]。

尽管 Transformer-UNet 架构提升了医学图像分割性能，但其在实际部署中面临两大瓶颈：结构复杂导致模型体积庞大，以及推理阶段计算资源消耗高。医学图像通常具有更高分辨率与更多结构细节，使得 Transformer-based 网络在临床系统(如放疗 TPS、移动边端设备)上的应用受到限制[4]。因此，如何在保持模型精度的同时有效降低参数规模与计算复杂度，是当前医学图像分割研究的重要方向。

针对上述问题，研究者们提出了多种轻量化策略，包括深度可分离卷积、跨层注意力、轻量化 Transformer 模块以及知识蒸馏等[5]。然而，部分方法依赖复杂的模块设计，增加了工程实现难度；另一些方法在降低参数的同时导致模型表达能力下降，特别是在器官边界恢复、小器官结构描述与多尺度特

征融合方面出现性能退化。因此, 构建一种兼具高效性、轻量化与边界恢复能力的医学图像分割网络仍具有研究价值。

为解决上述问题, 本文在 Swin-UNet 框架基础上提出轻量化分割模型 L-SwinUNet。本研究的改进策略源自对 Swin-UNet 结构的深入分析: 编码器(Swin Transformer Block)在全局建模方面表现优异, 无需进行轻量化修改; 模型的主要计算和参数负担集中在跳跃连接的浅层特征融合、解码器的多层卷积操作以及上采样阶段的特征重建。基于此, 本文从三个关键环节对 Swin-UNet 进行轻量优化, 设计了一套结构清晰、推理轻量、边界恢复能力强的改进方案。

首先, 为提升 skip connection 的特征表达能力, 本文在跳跃连接中引入 SimAM (Simple Attention Module)零参数注意力机制[6]。不同于 SE、CBAM 等需要额外参数的注意力模块, SimAM 基于神经科学启发的能量函数构建像素级空间注意力, 能够在不增加参数与几乎不增加计算量的前提下, 突出跳跃连接中的关键结构特征。浅层特征通常包含丰富的边界信息与纹理细节, 但同时也包含噪声与局部不稳定性。SimAM 能够基于输入特征自身统计特性分配自适应权重, 从而在 skip fusion 阶段为解码器提供更干净、更明确的空间特征输入。

其次, 为降低解码器的计算复杂度, 本文将传统的  $3 \times 3$  卷积替换为深度可分离卷积(Depthwise Separable Convolution, DSC), 构建轻量化解码器(DSC-Decoder) [7]。标准卷积在解码器中占据大量参数与 FLOPs, 而 DSC 将卷积分解为通道内卷积与通道间 pointwise 卷积, 能在保持特征融合能力的同时显著降低计算成本。在医学图像重建场景中, 深度可分离卷积的结构正则化作用还能一定程度提升模型的泛化能力, 使 decoder 更善于恢复器官局部结构。

最后, 为提高模型在上采样阶段的边界细节恢复能力, 本文采用 CARAFE (Content-Aware Reassembly of Features)重组上采样模块代替双线性插值[8]。传统上采样方法(bilinear 或 ConvTranspose2d)通常会导致小器官模糊、长边界断裂等问题, 而 CARAFE 能够通过内容感知的自适应卷积核对特征进行局部重组, 实现更细致、更柔和的空间重建, 尤其适用于腹部器官这类边界复杂且尺度变化大的医学图像。

基于上述三个模块的整合, 本文提出的 L-SwinUNet 在保持 Swin-UNet 全局编码能力的同时, 显著降低了整体复杂度, 并提升了边界恢复能力与小器官分割精度。在 Synapse 和 ACDC 数据集上的实验表明, L-SwinUNet 相较于原始模型在参数减少约 48%与计算量下降约 40%的条件下, 仍提高 Dice 分数与更低的 HD95 指标, 证明其轻量化设计的有效性与实用价值。

综上, 本文的贡献可以总结为以下三点:

- (1) 提出在跳跃连接中引入 SimAM 零参数注意力, 用于无额外开销地强化浅层空间特征;
- (2) 构建基于深度可分离卷积的轻量化解码器, 有效减少计算复杂度;
- (3) 在上采样阶段采用 CARAFE 重组机制, 提高边界与小尺度结构的重建质量;

从而提出一种兼具轻量化、高精度、低推理成本的医学图像分割网络, 为 Transformer-based 医学图像分割的实际部署提供新的结构优化思路。

## 2. 本文方法

### 2.1. 整体网络架构

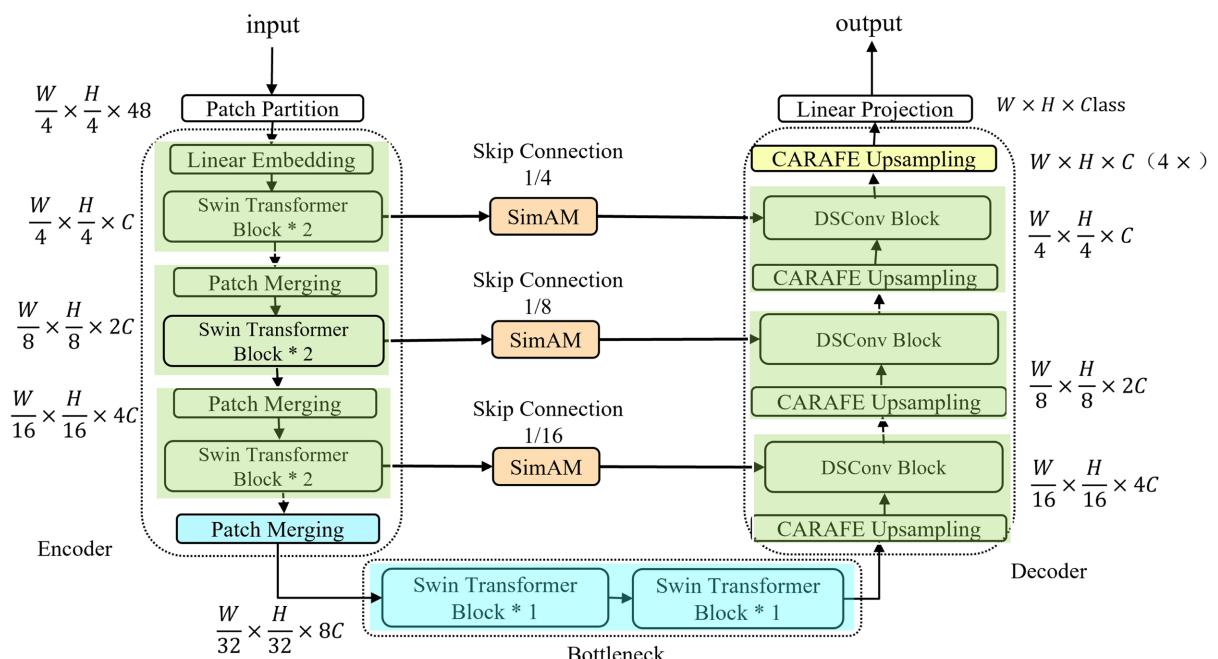
针对为提升医学图像分割任务中对细粒度解剖结构的建模能力并降低网络整体参数量, 本文构建了一种轻量化的改进型 SwinUNet 网络, 其整体架构如图 1 所示。模型仍采用典型的编码器—解码器—跳跃连接的对称式结构, 但在特征压缩、跨层信息交互与上采样重建等关键环节进行了针对性优化。

在编码器部分, 前几级保持原始 Swin Transformer Block, 以维持模型对局部结构特征的高敏感性, 并确保浅层阶段能够充分提取边缘与纹理细节。自中间层开始, 随着特征维度增加及建模复杂度提升, 在各

个尺度的特征输出端引入了 SimAM 注意力机制。SimAM 以无参数的能量函数形式对空间通道进行加权，使模型能够在不显著增加计算开销的前提下提升关键区域的响应强度，从而增强编码特征的判别能力。

瓶颈层仍由精简后的 Swin Transformer Block 构成，用以维持 Transformer 在长程依赖建模方面的优势，但减少了块数以控制整体复杂度。在此基础上，保持 U 型结构跨尺度优势，通过跳跃连接将浅层空间细节引导至对称解码器。

在解码器部分，传统的卷积上采样被轻量化模块替代。具体而言，在每一级特征恢复阶段采用了 CARAFE 自适应重建上采样算子，并在上采样后叠加深度可分离卷积(DSConv)以完成局部结构的细致重建。CARAFE 能够利用动态聚合核实现高质量特征重构，而 DSConv 进一步降低参数量与计算成本，使得解码器在保持结构表达能力的同时显著减轻模型负担。多级上采样过程与编码器特征逐级融合，实现对目标区域结构细节的恢复。



**Figure 1.** Improved SwinUNet model network architecture  
**图 1.** 改进型 SwinUNet 模型网络结构

## 2.2. SimAM 零参数注意力机制

针对医学图像分割任务中编码器 - 解码器架构存在的浅层空间细节特征与深层语义特征融合效率不足的问题，本文在跳跃连接路径中引入 SimAM (Simple Parameter-Free Attention Module)零参数注意力机制。该机制源于神经科学领域的空间抑制理论(Spatial Suppression Theory)，SimAM 通过定量评估每个神经元与其邻域神经元之间的线性可分性来确定其重要程度[10]。

具体而言，对于输入特征图  $X \in R^{(C \times H \times W)}$  中的目标神经元  $t$ ，SimAM 构建如下能量函数来度量其显著性：

其中  $\hat{t} = w_t t + b_t$  和  $\hat{x}_i = w_i x_i + b_i$  分别为目标神经元  $t$  和其他神经元  $x_{-i}$  的线性变换， $M = H \times W$  为该通道的神经元总数， $w_t$  和  $b_t$  为线性变换的权重与偏置。通过最小化该能量函数，可以找到使目标神经元与周围神经元最大程度分离的线性变换参数。为简化计算，采用二值标签  $y_t = 1$  和  $y_o = -1$ ，并添加正则化项，最终能量函数表示为：

该能量函数存在关于  $w_t$  和  $b_t$  的闭式解析解，避免了迭代优化的计算负担：

其中  $\mu_t = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i$  和  $\sigma_t^2 = \frac{1}{M-1} \sum_{i=1}^{M-1} (x_i - \mu_t)^2$  分别为除目标神经元外所有神经元的均值与方差。考

虑到同一通道内所有像素服从相同分布的假设，可以在整个通道上计算统一的统计量并复用，从而显著降低计算成本。因此，目标神经元  $t$  的最小能量可简化为：

其中  $\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i$  和  $\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2$  为通道内所有神经元的全局均值与方差。能量值  $e_t^*$  越低，表明

该神经元越能与周围环境区分，因此其重要性由  $1/e_t^*$  度量。最终的特征细化过程采用增益调制机制：

其中  $E$  汇集了所有空间位置与通道的能量值， $\sigma$  为 sigmoid 激活函数用于限制权重范围[11]， $\odot$  表示逐元素乘法。该注意力机制的核心优势在于：(1) 无需引入任何可学习参数，避免了参数冗余；(2) 计算过程仅涉及通道统计量与逐元素运算，可通过现代深度学习框架高效实现，代码实现少于 10 行；(3) 同时考虑空间与通道维度的三维注意力权重，相比传统的通道注意力(如 SE 模块)或空间注意力具有更强的表达能力。如图 2 所示，SimAM 模块为输入特征生成同时覆盖空间与通道维度的三维注意力权重，从而在不引入额外参数的情况下实现对关键区域的自适应增强。在医学图像分割的跳跃连接中应用 SimAM，可在不增加模型参数的前提下，自适应地强化浅层特征图中对分割任务更关键的解剖结构边界和纹理细节，抑制无关背景噪声，从而提升特征融合质量。

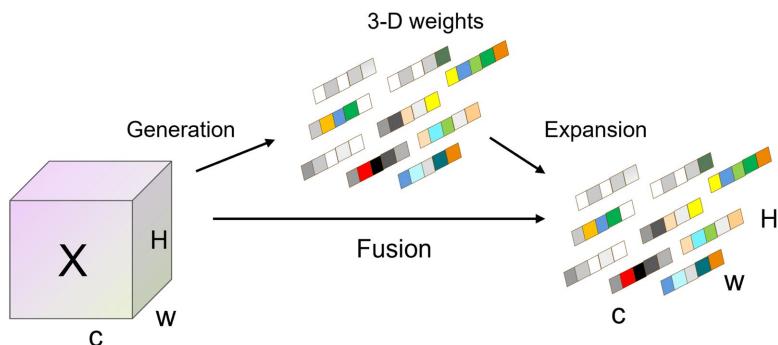


Figure 2. Full 3-D weights for attention

图 2. 注意力全三维权重

### 2.3. 深度可分离卷积的轻量化解码器

为降低解码器的计算复杂度并减少参数冗余，本文采用深度可分离卷积(Depthwise Separable Convolution)替代原始 SwinUNet 解码器中的 Swin Transformer 块。深度可分离卷积将标准卷积分解为深度卷积与逐点卷积两个独立步骤。对于输入特征图，深度卷积首先对每个通道独立进行空间滤波：

$$\hat{G}_{k,l,m} = \sum \hat{K}_{i,j,m} \cdot F_{k+i-1, l+j-1, m} \quad (1)$$

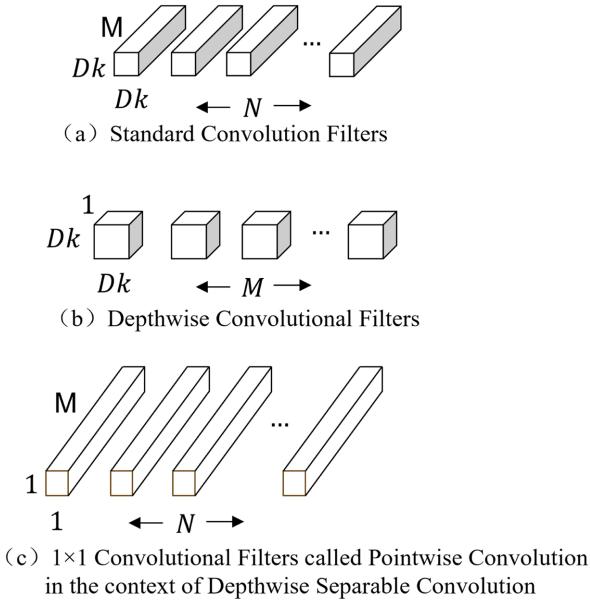
随后通过  $1 \times 1$  逐点卷积进行跨通道信息融合：

$$G_{k,l,n} = \sum_m W_{m,n} \cdot \hat{G}_{k,l,m} \quad (2)$$

该分解策略使得计算复杂度从标准卷积的  $D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F$  降低至  $D_K \cdot D_K \cdot M \cdot D_F + M \cdot N \cdot D_F \cdot D_F$ ，压缩比例可达：

$$\frac{1}{N} + \frac{1}{D_K^2} \quad (3)$$

采用  $3 \times 3$  深度可分离卷积时, 计算量可减少约 8~9 倍。其结构对比如图 3 所示, 深度可分离卷积通过将标准卷积分解为深度卷积与逐点卷积两个阶段, 在显著降低计算复杂度的同时保持有效的特征表达能力。这种架构设计在显著降低模型参数量与浮点运算次数的同时, 仍能维持充分的特征表达能力, 适用于资源受限的临床部署场景。



**Figure 3.** The standard convolutional filters in (a) are replaced by two layers: depthwise convolution in (b) and pointwise convolution in (c) to build a depthwise separable filter

**图 3.** (a)中的标准卷积滤波器被替换为两层: 深度卷积(b)和逐点卷积(c); 从而构建深度可分离滤波器

## 2.4. CARAFE 内容感知特征重组上采样

传统的双线性插值与反卷积上采样算子在特征重建过程中通常依赖固定或位置无关的采样核, 难以根据局部语义内容自适应调整重建方式, 容易造成边界模糊或细节信息丢失。为提升解码阶段高分辨率特征的重建质量, 本文在 SwinUNet 解码器中引入内容感知特征重组机制 CARAFE (Content-Aware Reassembly of Features) [9], 用于替代原有的 Patch Expanding 上采样模块。

CARAFE 首先通过轻量级内容编码器对低分辨率特征的局部上下文进行建模, 并为每个目标位置  $l'$  自适应预测一组大小为  $k_{up} \times k_{up}$  的重组核, 其生成过程可表示为:

$$W_{l'} = \psi(N(X_l, k_{encoder})) \quad (4)$$

其中  $N(X_l, k_{encoder})$  表示以位置  $l$  为中心的局部邻域特征,  $\psi(\cdot)$  为共享参数的编码网络。预测得到的重组核经归一化处理后, 用于对邻域特征进行加权重组, 上采样结果可表示为:

$$X'_{l'} = \sum_{n=-r}^r \sum_{m=-r}^r W_{l'}(n, m) \cdot X_{l+n, l+m} \quad (5)$$

其中  $r = \lfloor k_{up} / 2 \rfloor$ 。该重组方式使上采样过程能够显式感知输入特征的内容分布, 从而增强对边缘区域与细粒度结构的表达能力。

从计算效率角度分析, CARAFE 通过将权重预测与特征重组解耦, 在引入内容感知能力的同时保持了较低的计算复杂度。相较于反卷积算子, 其参数量与 FLOPs 均显著降低, 更适合轻量化解码器结构。

将 CARAFE 应用于 SwinUNet 解码阶段, 有助于在不显著增加模型开销的前提下, 提高医学图像分割结果在边界细节与小尺度目标区域的重建精度。

### 3. 实验

#### 3.1. 数据集

本文选用公开的 Synapse 多器官腹部 CT 数据集作为实验数据来源。该数据集涵盖 30 个临床病例的腹部 CT 扫描影像, 累计包含 3779 幅轴位切片图像, 分割目标涉及 8 类腹部解剖结构: 主动脉(Aorta)、脾脏(Spleen)、双侧肾脏(Kidney)、胆囊(Gallbladder)、胰腺(Pancreas)、肝脏(Liver)及胃(Stomach), 按照官方推荐的划分方式, 18 个病例用于模型训练, 12 个病例用于独立测试, 以此评估改进架构的分割性能[12]。

#### 3.2. 实验设置

##### 3.2.1. 实验环境和参数设置

本研究在 Python 3.6 环境下开展模型开发与训练, 深度学习框架选用 PyTorch。实验硬件平台采用搭载 NVIDIA RTX 4090 (24 GB 显存)的单卡 GPU 服务器, 并配置 CUDA 11.7.0 以支持并行加速计算。针对训练流程, 具体设置如下:

- (1) 数据处理策略: 对样本训练前进行数据增强操作, 包括随机旋转与镜像翻转等, 以提高模型对不同空间形变的适应性。所有输入图像被统一缩放至  $224 \times 224$  分辨率, 且 Patch Size 固定为 4。
- (2) 参数初始化: 模型权重通过在 ImageNet 数据集上训练得到的预训练参数进行初始化。
- (3) 优化方法: 采用 SGD 优化器, 初始学习率设置为 0.05, 批量大小为 24, 动量系数为 0.9。为缓解过拟合问题, 在优化与反向传播过程中引入权重衰减项, 其值设为  $1 \times 10^{-4}$ 。

##### 3.2.2. 损失函数与评价指标

为同步优化区域一致性与边界精确性, 本文采用 Dice 损失与交叉熵损失的联合形式作为训练目标[13], 其表达形式如式(14)所示。

$$L = L_{CE} + \lambda L_{Dice} \quad (6)$$

其中,  $L_{CE}$  表示交叉熵损失项,  $L_{Dice}$  表示 Dice 损失项,  $\lambda$  表示权重平衡因子。

模型评估采用 Dice 相似系数(Dice Similarity Coefficient, DSC)与 95% Hausdorff 距离(HD95)作为定量评价指标[14]。其中, DSC 用于衡量模型预测结果与真实标注之间的重叠程度, 其定义如式(15)所示, 其中  $P$  与  $G$  分别表示预测分割与真实标注集合。HD95 度量边界点集的空间偏离程度, 可敏感反映轮廓描绘质量。两指标协同使用能够全面刻画模型在医学影像分割场景下的综合表现。

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \quad (7)$$

### 3.3. 实验分析

#### 3.3.1. 对比实验

为系统验证所提出方法在多器官分割任务中的有效性, 本文在 Synapse 数据集上与多种经典及代表性分割模型进行了定量对比, 实验结果汇总于表 1。可以观察到, U-Net 与 Att-UNet 作为典型的 CNN 架构, 在整体分割性能上表现稳定, 其平均 Dice 系数分别为 76.85% 和 77.77%, 但对应的 HD 值均高于 36, 表明在复杂器官边界区域仍存在一定误分割问题。基于 Transformer 架构的 R50ViT 和 ViT 模型在全局建模方面具备优势, 但由于局部细节刻画能力受限, 其平均 DSC 均低于 72%, 整体分割精度相对不足。相比之下,

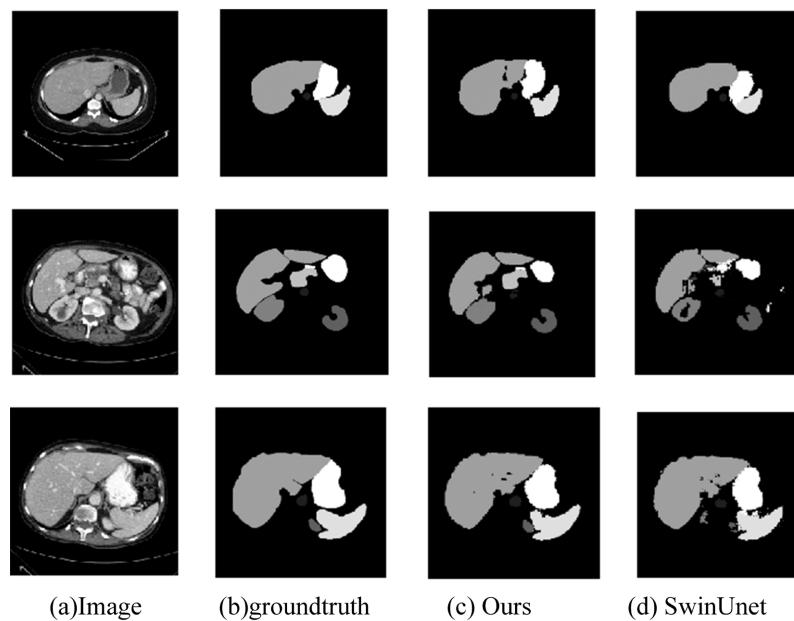
TransUNet 与 SwinUNet 通过融合卷积与 Transformer 结构，在全局语义理解与局部空间建模之间取得了更优平衡，其中 SwinUNet 的平均 DSC 提升至 79.13%，HD 下降至 21.55，显示出更强的多器官分割能力。

**Table 1.** The segmentation result data of different algorithms on the dataset  
**表 1.** 不同算法在数据集上的分割结果数据

Method <sup>a</sup>	Average		Aorta	Left kidney	Right kidney	Gallbladder	Pancreas	Liver	Spleen	Stomach
	DSC↑	HD↓								
U-Net	76.85	39.70	89.07	77.77	68.60	69.72	53.98	93.43	86.67	75.58
R50 Att-UNet	75.57	36.97	55.92	79.20	72.71	63.91	49.37	93.56	87.19	74.95
R50 ViT	71.29	32.87	73.73	75.80	72.20	55.13	45.99	91.51	81.99	73.95
Att-UNet	77.77	36.02	89.55	77.98	71.11	68.88	58.04	93.57	87.30	75.75
TransUNet	77.48	31.69	87.23	81.87	77.02	63.13	55.86	94.08	85.08	75.62
SwinUNet	79.13	21.55	85.47	83.28	79.61	66.53	56.58	94.29	90.66	76.60
<b>Ours</b>	<b>79.37</b>	<b>20.19</b>	<b>86.13</b>	<b>83.27</b>	<b>80.10</b>	<b>67.31</b>	<b>57.90</b>	<b>94.24</b>	<b>90.51</b>	<b>75.52</b>

**Table 2.** Comparison of model parameters, FLOPS, and inference time  
**表 2.** 各模型参数量、FLOPS、推理时间对比数据

Methods	Params/M↑	FLOPs/G	推理时间
U-Net	31.24	55.84	223
Att-UNet	34.88	66.57	235
TransUNet	105.28	24.64	246
SwinUNet	96.34	42.68	238
<b>Ours</b>	<b>50.13</b>	<b>25.61</b>	<b>205</b>



**Figure 4.** Visual comparison of different segmentation algorithms  
**图 4.** 不同算法分割效果图

在保证分割精度的同时，模型复杂度与推理效率同样是医学图像分割实际应用中的关键因素。由表 1 可知，本文方法在 Synapse 数据集上取得了 79.37% 的平均 DSC 和 20.19 的 HD，在分割精度和边界一致性方面均优于现有对比模型。进一步结合表 2 的复杂度分析可以发现，所提出模型的参数量为 50.13 M，较 SwinUNet 显著降低，同时 FLOPs 由 42.68 G 减少至 25.61 G，推理时间也缩短至 205 ms。上述结果表明，通过在跳跃连接中引入零参数 SimAM 注意力，并在解码阶段采用深度可分离卷积与 CARAFE 上采样策略，本文方法在有效压缩模型规模和计算开销的同时，仍保持了具有竞争力的分割性能，体现了良好的精度与效率权衡特性。其中图 4 展示了不同分割算法的视觉对比结果，本文方法(c 列)在边界精度和区域完整性上均优于对比方法(d 列)。

综上所述，所提出的改进模型不仅在整体分割精度上超越现有方法，还在器官细粒度层面展现出更强的泛化能力与稳定性，充分证明了其在医学图像多器官分割任务中的应用潜力。

### 3.3.2. 消融实验

为验证各改进模块对模型性能的具体贡献，本文在 Synapse 数据集上设计了渐进式消融实验，结果如表 3 所示。实验结果表明，在原始 SwinUNet 基础上引入 SimAM 零参数注意力后，分割精度提升至 79.59%，边界误差降至 19.97 mm，参数量同步减少 8%，验证了能量函数驱动的空间特征增强机制能够在无额外开销条件下改善跳跃连接的特征表达质量。进一步采用深度可分离卷积构建解码器后，参数量压缩至 48.32M，但 DSC 回落至 77.53%，HD95 上升至 23.16 mm，说明单纯的卷积分解虽实现高效轻量化，却在多尺度特征融合与细节恢复能力上存在性能折损。

引入 CARAFE 内容感知上采样模块后，完整架构性能得到全面补偿：DSC 恢复至 79.37%，HD95 显著优化至 20.19 mm，同时参数量与计算复杂度分别维持在 50.13 M 和 25.61 G。消融实验证明，SimAM 在前端提升特征质量，DSConv 在中端实现效率跃升，CARAFE 在后端保障边界刻画精度，使模型在满足临床分割需求的前提下具备边缘设备部署能力。

**Table 3.** Data from the melting experiment

**表 3. 消融实验结果数据**

Methods	DSC↑	HD↓	Params/M	FLOPs/G
SwinUNet	79.13	21.55	96.34	42.68
SwinUNet + SimAM	79.59	19.97	88.61	35.56
SwinUNet + SimAM + DSConv	77.53	23.16	48.32	30.12
SwinUNet + SimAM + DSConv + CARAFE	79.37	20.19	50.13	25.61

## 4. 结语

本文提出了一种名为 L-SwinUNet 的轻量化医学图像分割模型，旨在解决存在的参数冗余与计算复杂度高的问题。实验结果表明，与原始 Swin-UNet 相比，L-SwinUNet 在参数量降低约 48%、计算量减少约 40% 的同时，其 Dice 相似系数和 HD95 (95% Hausdorff 距离) 均表现更优。这进一步证实，所提出的轻量化策略不仅能显著降低模型资源消耗，而且能够通过结构优化提升模型表达能力，从而在保持甚至提升医学图像分割精度的前提下，实现有效且实用的轻量化设计。

## 基金项目

合肥综合性国家科学中心能源研究院(安徽省能源实验室)项目(21KZS202)。

## 参考文献

- [1] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) Imagenet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. <https://doi.org/10.1145/3065386>
- [2] Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., *et al.* (2022) Swin Transformer V2: Scaling up Capacity and Resolution. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, 18-24 June 2022, 12009-12019. <https://doi.org/10.1109/cvpr52688.2022.01170>
- [3] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., *et al.* (2023) Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky, L., Michaeli, T. and Nishino, K., Eds., *Lecture Notes in Computer Science*, Springer, 205-218. [https://doi.org/10.1007/978-3-031-25066-8\\_9](https://doi.org/10.1007/978-3-031-25066-8_9)
- [4] Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S., *et al.* (2023) Transformers in Medical Imaging: A Survey. *Medical Image Analysis*, **88**, Article 102802. <https://doi.org/10.1016/j.media.2023.102802>
- [5] Mehta, S. and Rastegari, M. (2021) Mobilevit: Lightweight, General-Purpose, and Mobile-Friendly Vision Transformer. arXiv:2110.02178.
- [6] Hinton, G., Vinyals, O. and Dean, J. (2015) Distilling the Knowledge in a Neural Network. arXiv:1503.02531.
- [7] Yang, L., Zhang, R.Y., Li, L., *et al.* (2021) SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. *Proceedings of the 38th International Conference on Machine Learning*, **139**, 11863-11874.
- [8] Howard, A.G., Zhu, M., Chen, B., *et al.* (2017) Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861.
- [9] Wang, J., Chen, K., Xu, R., Liu, Z., Loy, C.C. and Lin, D. (2019) CARAFE: Content-Aware Reassembly of Features. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, 27 October 2019-2 November 2019, 3007-3016. <https://doi.org/10.1109/iccv.2019.00310>
- [10] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. (2018) Mobilenetv2: Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, 18-23 June 2018, 4510-4520. <https://doi.org/10.1109/cvpr.2018.00474>
- [11] Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Feedforward Networks. *Deep Learning*, **1**, 161-217.
- [12] 康家荣, 邵鹏飞, 王元. 基于 Swin-Unet 改进的医学图像分割算法[J]. 人工智能与机器人研究, 2024, 13(2): 354-362.
- [13] 张文豪, 瞿绍军, 颜美丽. 基于深度学习的视网膜血管分割研究进展[J]. 计算机应用研究, 2025, 42(5): 1299-1311.
- [14] 任怡璇, 崔容宇. 人工智能深度学习在单光子计算机断层显像中的研究进展[J]. 新医学, 2024, 55(3): 159-164.