

CB-SwinUNet: 一种融合多模块注意力机制的乳腺超声肿瘤分割算法

周丰艺 杨喆 马苑 王军

摘要 目的 针对乳腺超声图像中强散斑噪声、边界模糊及低对比度等问题,现有方法在处理不规则肿瘤与复杂背景时,难以兼顾全局语义与局部细节。为此,本文提出融合 Swin Transformer 与 U-Net 优势的 CB-SwinUNet 模型,旨在实现复杂场景下的高精度分割。**方法** 模型构建基于 Swin Transformer 的 U 型架构,利用移动窗口自注意力捕捉长距离依赖。在此基础上引入两个增强模块:在编码阶段并行引入通道注意力与空间增强模块(channel attention and spatial enhancement, CASE),联合建模空间与通道特征,自适应抑制背景噪声并增强病灶表达;在跳跃连接处设计边界细节增强模块(boundary-aware detail module, BADM),利用轻量级卷积与多尺度池化聚合浅层高频信息,修正深层特征的边缘丢失。此外,结合复合损失函数与患者级五折交叉验证策略确保评估客观性。**结果** 在 BUSI(647 例)与自建数据库(500 例)上的实验显示,CB-SwinUNet 的 Dice 系数分别达到 0.859 和 0.842, IoU 与 Hausdorff 距离等指标均优于 U-Net、TransFuse 等主流模型。消融实验证实,CASE 显著提升了低对比度病灶定位能力,BADM 有效修正了边缘刻画精度。模型收敛迅速且未出现明显过拟合。**结论** CB-SwinUNet 通过深度嵌入式融合策略,实现了全局结构与局部细节的协同建模,有效解决了噪声干扰与边界模糊难题。实验结果充分验证了该方法在提升分割精度与泛化能力方面的优势,为临床辅助诊断提供了鲁棒的自动化方案。

关键词 乳腺超声图像;图像分割;深度学习;注意力机制;Swin Transformer

DOI: 10.3969/j.issn.1002-3208.2026.04.007.

中图分类号 R318 **文献标志码** A **文章编号** 1002-3208(2026)04-0376-10

本文著录格式 周丰艺,杨喆,马苑,等. CB-SwinUNet: 一种融合多模块注意力机制的乳腺超声肿瘤分割算法[J]. 北京生物医学工程, 2026, 45(4): 376-385. ZHOU Fengyi, YANG Zhe, MA Yuan, et al. CB-SwinUNet: a multi-module attention fusion algorithm for breast ultrasound tumor segmentation[J]. Beijing Biomedical Engineering, 2026, 45(4): 376-385.

CB-SwinUNet: a multi-module attention fusion algorithm for breast ultrasound tumor segmentation

ZHOU Fengyi, YANG Zhe, MA Yuan, WANG Jun

Peking University First Hospital Taiyuan Branch (Taiyuan Central Hospital of Shanxi Medical University),
Taiyuan 030009

Corresponding author: WANG Jun (E-mail: wangjun8998@163.com)

【Abstract】 Objective Existing methods often struggle to balance global semantic understanding with local detail recovery when dealing with irregular tumor shapes and complex backgrounds, due to prevalent issues in breast ultrasound images such as severe speckle noise, blurred boundaries, and low contrast. To address these challenges, this paper proposes a CB-SwinUNet model that integrates the advantages of Swin

Transformer and U-Net, aiming to achieve high-precision segmentation in complex clinical scenarios. **Methods** The model constructs a U-shaped architecture based on the Swin Transformer, utilizing shifted window self-attention mechanisms to capture long-range dependencies. Building on this, two key enhancement modules are introduced. First, a channel attention and spatial enhancement (CASE) module is incorporated in parallel during the encoding stage to jointly model spatial and channel features, adaptively suppressing background noise and enhancing lesion representation. Second, a boundary-aware detail module (BADM) is designed at the skip connections, leveraging lightweight convolutions and multi-scale pooling to aggregate shallow high-frequency information, thereby correcting the loss of edge details in deep features. Furthermore, a composite loss function and a patient-level 5-fold cross-validation strategy are employed to ensure objective evaluation. **Results** Experiments on the BUSI dataset (647 cases) and a self-built database (500 cases) demonstrate that CB-SwinUNet achieves Dice coefficients of 0.859 and 0.842, respectively. Metrics such as intersection over union (IoU) and Hausdorff distance are superior to those of mainstream models like U-Net and TransFuse. Ablation studies confirm that CASE significantly improves the localization capability for low-contrast lesions, while BADM effectively refines edge delineation accuracy. The model exhibits rapid convergence without significant overfitting. **Conclusions** Through a deep embedded fusion strategy, CB-SwinUNet achieves synergistic modeling of global structures and local details, effectively resolving challenges related to noise interference and blurred boundaries. The experimental results fully validate the advantages of the proposed method in enhancing segmentation accuracy and generalization ability, providing a robust automated solution for clinical computer-aided diagnosis.

【Keywords】 breast ultrasound image; image segmentation; deep learning; attention mechanism; Swin Transformer

0 引言

乳腺癌作为全球女性最常见的恶性肿瘤之一,其早期精准诊断是提升患者生存率和改善预后的关键^[1]。临床诊断中,超声成像凭借无辐射、实时性强及成本效益高等优势,已成为筛查的首选技术^[2]。然而,受成像机制限制,超声图像(见图1)常伴随散斑噪声、组织边界模糊及低对比度等伪影干扰^[3],导致诊断结果高度依赖医生经验,客观性与复现性不足。因此,开发高效的计算机辅助诊断(computer-aided diagnosis, CAD)系统^[4],特别是鲁棒的肿瘤分割算法,对于辅助临床精准诊疗至关重要。

传统的图像分割方法(如主动轮廓模型)因过分依赖浅层特征,在复杂超声环境中表现欠佳。近年来,以U-Net^[5]为代表的卷积神经网络(convolutional neural network, CNN)虽凭借局部特征提取能力在

医学分割领域取得突破,但其受限的感受野难以捕捉全局上下文信息^[6],在面对形态不规则或低对比度病灶时精度受限。引入自注意力机制的Transformer架构虽在全局建模上展现出巨大潜力^[7],但直接应用于高分辨率图像时面临计算开销大且局部细节(如微小边缘)捕捉不足的挑战。因此,如何有效融合CNN的局部感知优势与Transformer的全局建模能力,并在抑制噪声的同时实现对模糊边界的精准刻画,是当前亟需解决的科学问题^[8]。

针对上述挑战,本文提出一种融合多模块注意力机制的新型分割架构——CB-SwinUNet。该方法旨在利用Swin Transformer的层级化全局建模能力弥补CNN的局限,并创新性地引入通道注意力与空间增强模块(channel attention and spatial enhancement, CASE)及边界细节增强模块(boundary-aware detail module, BADM),分别针对性地解决超声图像中病灶与背景对比度低及边缘模糊的难题,以期在复杂临床场景下实现更精准、鲁棒的肿瘤分割。

基金项目:太原市科技项目计划(202233)资助

作者单位:北京大学第一医院太原医院/太原市中心医院(太原 030009)

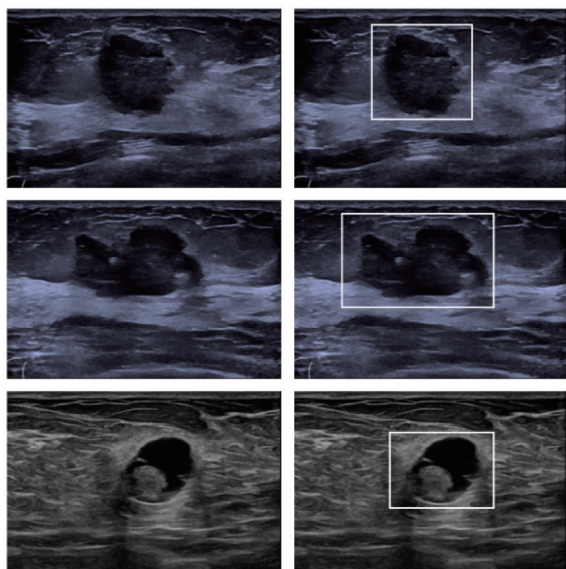
通信作者:王军。E-mail:wangjun8998@163.com

1 相关工作

乳腺超声图像分割技术主要分为传统方法、基于 CNN 的方法及基于 Transformer 的方法。

1.1 传统分割方法

早期研究主要依赖图像的灰度、纹理或边缘特征。代表性算法包括 Otsu 自适应阈值分割^[9]、基于统计特征的区域增长算法^[10]、以及以 Level Set 为代表的活动轮廓模型^[11]。尽管这些方法在简单场景下有效,但其对散斑噪声和灰度不均匀高度敏感,且依赖人工参数设置,难以适应临床图像的多样性与复杂性。



左列为原始图像;右列为白色框标注肿瘤所在区域

图 1 乳腺病灶超声图像示例

Figure 1 Example of breast lesion ultrasound images

1.2 CNN 方法

卷积神经网络凭借强大的特征提取能力成为主流,U-Net^[5]奠定了医学图像分割的基础。随后的研究致力于改进其结构缺陷;Zhou 等^[12]提出的 UNet++通过密集跳跃连接缩减了语义差距;Woo 等^[13]提出 CBAM 模块,通过融合通道与空间注意力机制增强了模型对关键区域的表征能力;后续研究^[14]进一步结合残差结构、多尺度池化及上下文融合模块,在一定程度上提升了分割精度。然而,受限于卷积核的局部感受野,CNN 难以捕捉长距离依赖关系,在处理低对比度或结构复杂的病灶时仍存

在精度瓶颈^[15-16]。

1.3 Transformer 方法

受自然语言处理领域启发,Transformer 利用自注意力机制实现了全局信息建模。为解决纯 Transformer [如视觉转换器(vision Transformer, ViT)]在局部细节捕捉上的不足,混合架构应运而生。Liu 等^[17]提出 Swin Transformer,通过移位窗口机制实现了高效的分层特征建模;Hatamizadeh 等^[18]提出 UNETR,将 Transformer 编码器引入医学图像分割框架。近期,Sclera-TransFuse^[19]和 DAMAF^[20]等研究尝试采用并行分支或双重注意力模块来进一步融合多模态特征。尽管如此,如何在抑制噪声的同时高效恢复微小边缘细节,并平衡计算开销,仍是当前亟待解决的难题。

综上所述,本文提出 CB-SwinUNet,旨在融合 Transformer 的全局建模优势与 CNN 的细节恢复能力,通过引入通道注意力与空间增强模块及边界细节增强模块,重点解决超声图像中的背景干扰与边缘模糊问题,实现兼顾精度与效率的鲁棒分割。

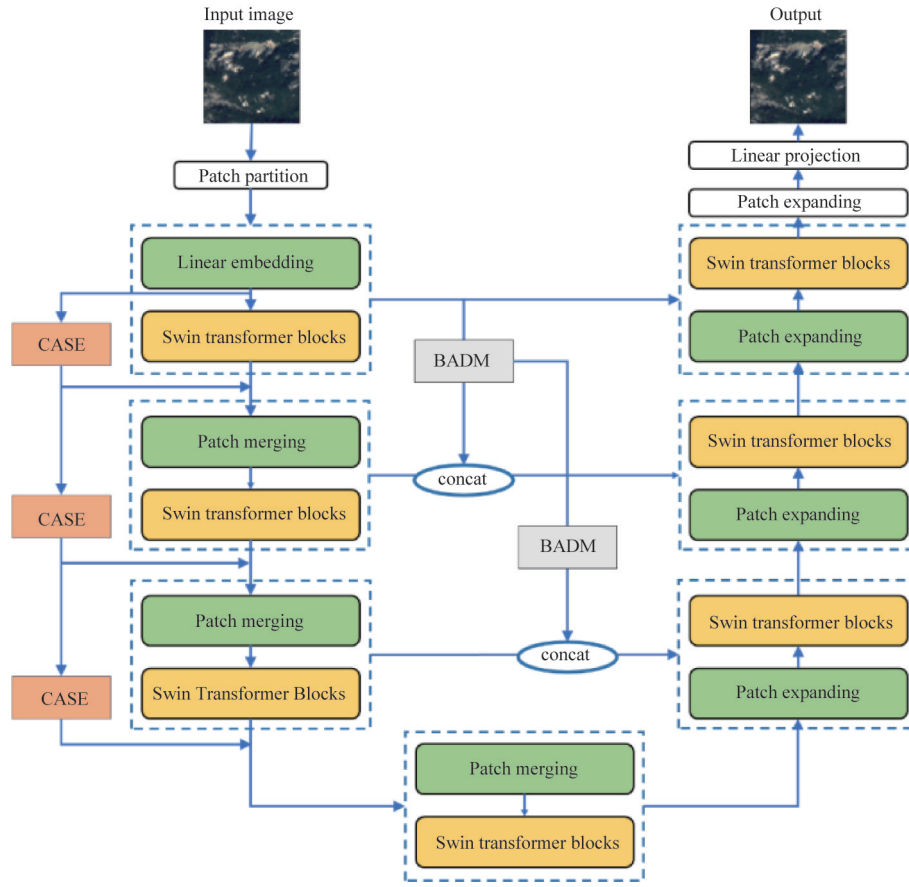
2 方法

2.1 整体架构

本文提出的 CB-SwinUNet 基于 Swin Transformer 构建对称 U 型编解码架构,旨在融合全局上下文建模与局部细节恢复能力。如图 2 所示,模型由 SwinUNet 主干、通道注意力与空间增强模块及边界细节增强模块构成。

模型输入为 256×256 灰度图像,首先经 Patch Partition(4×4)与线性嵌入映射为高维特征。编码器包含三个阶段,利用 Patch Merging 和 Swin Transformer Block 逐级下采样,实现分辨率减半与通道倍增,以提取多尺度语义特征。为增强特征表达,各阶段并行引入 CASE 模块,通过联合通道与空间注意力加权,抑制背景干扰并增强病灶判别力。编码器末端连接由两个 Swin Transformer Block 构成的 Bottleneck 模块,进一步整合深层特征并缓解窗口机制导致的信息割裂。

解码器通过 Patch Expansion 逐级恢复分辨率。为弥补下采样造成的细节丢失,在跳跃连接处引入 BADM 模块。该模块对编码器浅层特征进行卷积增强与多尺度融合,并将增强后的特征传输至解码



CASE表示通道注意力与空间增强模块;BADM表示边界细节增强模块

图2 CB-SwinUNet的体系结构

Figure 2 Architecture of CB SwinUNet

器,以强化对肿瘤边界的感知。最终,模型经线性映射层输出与原图尺寸一致的预测图,完成端到端的肿瘤分割。

2.2 Swin Transformer 模块

Swin Transformer 采用分层局部注意力机制,有效平衡了计算效率与长距离上下文建模能力。每个模块由两个连续的 Transformer Block 组成,分别基于窗口多头自注意力(window multi-head self-attention, W-MSA)和移位窗口多头自注意力(shifted window multi-head self-attention, SW-MSA)。W-MSA 将计算限制在非重叠窗口内以提取局部特征,而 SW-MSA 通过窗口偏移机制实现跨窗口信息交互,增强全局建模能力。两者结合多层感知器(multilayer perceptron, MLP)、层归一化(LayerNorm)及残差连接构成基本单元(结构如图3所示),连续两个 Block 的计算过程如式(1)所示。

$$\begin{aligned}\hat{z}^l &= W - \text{MSA}(\text{LN}(z^{l-1})) + z^{l-1} \\ z^l &= \text{MLP}(\text{LN}(\hat{z}^l)) + \hat{z}^l \\ \hat{z}^{l+1} &= \text{SW} - \text{MSA}(\text{LN}(z^l)) + z^l \\ z^{l+1} &= \text{MLP}(\text{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1}\end{aligned}\quad (1)$$

式中: z^{l-1} 和 z^{l+1} 分别表示该模块的输入和输出特征。

相较于标准 ViT 的平方级计算复杂度, Swin Transformer 通过局部窗口机制实现了与图像分辨率呈线性关系的复杂度。这一特性不仅降低了计算开销,更结合多层次特征抽象,有效兼顾了浅层空间细节与深层语义结构,使其能精准捕捉乳腺肿瘤复杂的多尺度边界特征。

2.3 通道注意力与空间增强模块

针对乳腺超声图像对比度低及背景伪影(如导管、脂肪)干扰严重的问题,本文在主干网络的编码阶段并行引入通道注意力与空间增强模块。如图4

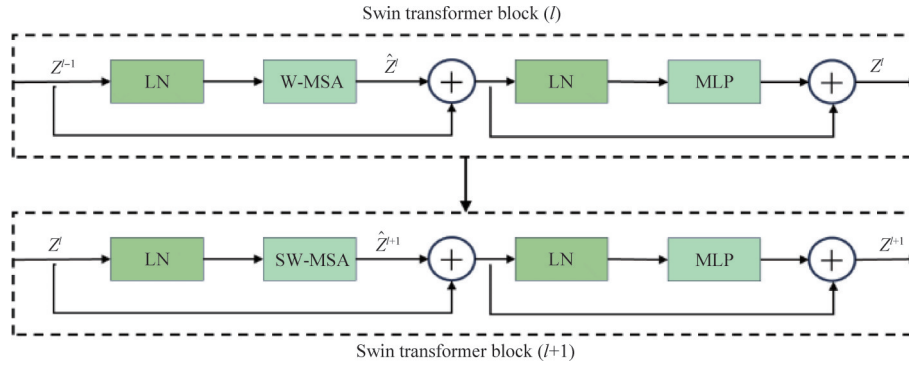


图3 Swin Transformer模块结构示意图

Figure 3 Schematic diagram of Swin Transformer module structure

所示,该模块旨在从通道和空间双维度联合建模特征显著性,通过抑制背景噪声并增强病灶聚焦,解决图像空间非均匀性导致的分割难题。CASE由通道注意力(coordinate attention, CA)与空间增强(squeeze-and-excitation, SE)两个子模块构成。

CASE模块的输入与主干网络对应阶段的输出一致,初始形式为二维序列特征 $F \in \mathbf{R}^{L \times C}$, 其中 $L=H \times W$ 表示非重叠 patch 的总数, C 为每个 token 的通道数。为了适配空间维度的建模需求,该输入需进行重塑(reshape)操作,转换为三维张量 $F_{\text{CASE}} \in \mathbf{R}^{H \times W \times C}$ 。随后, F_l 分别作为两个路径的输入:一方面施加 2×2 平均池化用于下一级CASE模块输入;另一方面进入 1×1 卷积块压缩通道至 C ,生成空间与通道注意力子模块共享的输入特征,其计算如式(2)所示:

$$F_1 = \text{ReLU}\left(\text{BN}\left(\text{Conv}_{3 \times 3}(F_{\text{CASE}})\right)\right) \quad (2)$$

式中: $F_1 \in \mathbf{R}^{H \times W \times 2C}$ 。一方面, F_1 经过 2×2 平均池化后作为下一阶段CASE模块的输入;另一方面输入 1×1 卷积块进行通道压缩,生成供CA与SE子模块共享的特征图 $V \in \mathbf{R}^{H \times W \times C}$,其计算如式(3)所示:

$$V = \text{ReLU}\left(\text{BN}\left(\text{Conv}_{1 \times 1}(F_1)\right)\right) \quad (3)$$

空间增强子模块通过方向性平均池化提取行列方向的特征响应,具体计算如式(4)所示:

$$R = \text{AvgRow}(V) \in \mathbf{R}^{H \times 1 \times C}, T = \text{AvgCol}(V) \in \mathbf{R}^{1 \times W \times C} \quad (4)$$

随后,采用广播机制将 R 与 T 相乘,并通过Sigmoid激活生成空间注意力图 S ,其定义如式(5)所示:

$$S = \sigma(R \times T) \in \mathbf{R}^{H \times W \times C} \quad (5)$$

通道注意力子模块对共享特征 V 施加全局平均池化并通过激活函数生成通道注意力图 $D \in \mathbf{R}^{1 \times 1 \times C}$ 。最终,将空间注意力 S 与通道注意力 D 融合,生成交叉注意力权重图 $M \in \mathbf{R}^{H \times W \times C}$,其计算如式(6)所示:

$$M = \sigma(\text{GAP}(V)) \odot \sigma(R \times T) \quad (6)$$

该注意力图被用于对主干输出特征 F 进行逐元素加权校准,具体操作如式(7)所示:

$$F' = F_{\text{CASE}} \otimes M \quad (7)$$

通过上述机制,CASE模块在编码阶段动态增强了病灶区域的语义特征,有效抑制了非均匀纹理与背景伪影干扰,为后续解码提供高判别力的特征支撑。

2.4 边界细节增强模块

为了解决深层网络下采样导致边缘细节丢失及轮廓模糊的难题,本文在编码器浅层与解码器之间的跳跃连接处引入边界细节增强模块。如图5所示,该模块旨在利用浅层特征保留的高频空间信息,补充解码阶段损失的细节,引导模型精准恢复病灶轮廓。

设来自编码器浅层的输入特征图为 $B \in \mathbf{R}^{H \times W \times C}$,该特征图包含丰富的空间纹理信息但通道容量有限,难以建模复杂边缘结构。BADM首先对其施加 1×1 卷积进行通道扩展至 $2C$,并依次施加 BatchNorm 与 ReLU 激活,其操作定义如式(8)所示:

$$U = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(B))), U \in \mathbf{R}^{H \times W \times 2C} \quad (8)$$

随后,模块引入一层 3×3 卷积和 LeakyReLU 激活构成边缘增强子模块,利用扩大感受野聚焦于边

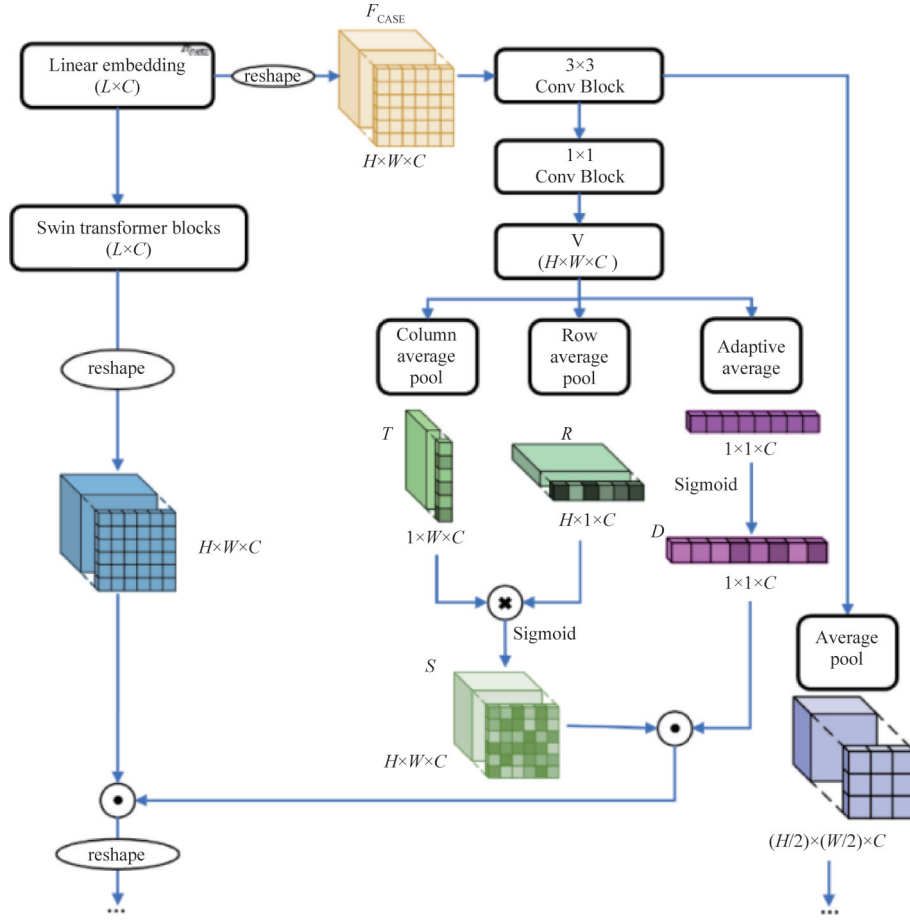


图4 通道注意力与空间增强模块结构

Figure 4 Structure of channel attention and spatial enhancement module

缘变化区域,有效提取肿瘤轮廓的局部结构特征,输出仍保持尺寸为 $H \times W \times 2C$,其过程如式(9)所示:

$$U_{\text{CONV1}} = \text{LeakyReLU}(\text{Conv}_{3 \times 3}(U)), U_{\text{CONV1}} \in \mathbb{R}^{H \times W \times 2C} \quad (9)$$

为降低纹理噪声干扰并提升边界响应的一致性,模块采用 3×3 平均池化操作(stride=2)对特征图进行空间压缩同时实现平滑和降噪,得到下采样特征,其计算如式(10)所示:

$$Z = \text{AvgPool}_{3 \times 3}(U_{\text{CONV1}}), Z \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 2C} \quad (10)$$

由于解码器期望输入为通道数 C 的特征图,BADM随后使用一个 1×1 卷积对 Z 进行通道压缩,并配合BN与ReLU激活,输出最终边界增强特征,其定义如式(11)所示:

$$B' = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(Z))), B' \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times C} \quad (11)$$

最终,边界增强特征 B' 与当前解码器主干输出特征 $D \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times C}$ 进行通道维度拼接操作,融合形成

新的解码输入特征 D' ,该操作如式(12)所示:

$$D' = \text{Concat}(D, B'), D' \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 2C} \quad (12)$$

BADM通过结构化的特征提取与反馈,在低参数量代价下有效提升了网络对微小病灶边缘的感知与恢复能力。

2.5 损失函数

针对乳腺超声图像中前景背景类别极度不平衡及病灶面积微小的特点,本文采用由二元交叉熵损失(BCE)与Dice损失加权融合的复合损失函数,以兼顾像素级分类精度与区域结构重建的完整性。

2.5.1 二元交叉熵损失(BCE Loss)

BCE损失用于度量每个像素预测值与真实标签之间的差异,其定义如式(13)所示:

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \lg(\hat{y}_i) + (1 - y_i) \lg(1 - \hat{y}_i)] \quad (13)$$

式中: $\hat{y}_i$ 表示模型对第 i 个像素预测为肿瘤的概率

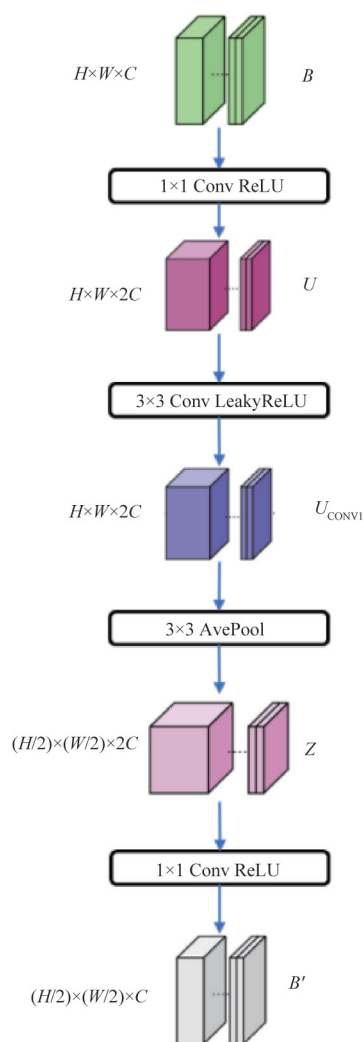


图 5 边界细节增强模块的结构

Figure 5 Structure of boundary-aware detail module

率; y_i 为对应的真实标签; N 为总像素数。

2.5.2 Dice 损失 (Dice Loss)

Dice 损失衡量预测区域与真实区域的重叠程度, 其定义如式(14)所示:

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i + \varepsilon}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i + \varepsilon} \quad (14)$$

式中: ε 为防止分母为零的平滑因子。

2.5.3 复合损失函数

为兼顾分类精度与结构重建效果, 本文将 BCE 与 Dice 损失加权组合, 最终训练目标函数如式(15)所示:

$$L_{\text{total}} = \lambda_1 L_{\text{BCE}} + \lambda_2 L_{\text{Dice}} \quad (15)$$

式中: λ_1 和 λ_2 为损失函数的权重系数, 实验中设定 $\lambda_1 = \lambda_2 = 0.5$ 时可取得良好平衡。通过联合优化, 该函数在保证训练稳定性的同时, 显著提升了模型对模糊边界与微小病灶的分割性能。

3 结果与分析

3.1 实验材料

3.1.1 BUSI 数据集

BUSI (Breast Ultrasound Images Dataset) 是领域内广泛应用的公开基准数据集。本研究从中选取全部良性与恶性肿瘤图像共 647 张, 每张图像均配有医学专家精确勾画的二值分割掩膜。该数据集图像分辨率不一, 且普遍存在散斑噪声、边界模糊、低对比度及声影等典型伪影, 肿瘤形态差异显著, 能有效评估模型在复杂临床场景下的鲁棒性与泛化能力。

3.1.2 自建数据库

自建数据库源自太原市中心医院超声科, 包含 500 张乳腺肿瘤 B 超图像, 所有数据均经伦理批准并脱敏处理。数据标注由两名资深主治医师进行独立“双盲”勾画, 并由一名主任医师审核确认, 确保了真值的高一致性。该数据集涵盖多台主流设备采集的影像, 肿瘤轮廓复杂且边界不规则, 难度高于公开数据集。本研究采用严格的“患者级五折交叉验证”策略, 即强制同一患者的所有图像仅被划分至同一折 (Fold) 中, 从而在保证评估客观性的同时, 真实反映模型对未见患者数据的泛化能力。

3.1.3 预处理与数据增强

为确保模型训练的一致性与效率, 所有图像统一缩放至 256×256 像素。为提升模型泛化能力, 训练阶段应用了多种数据增强策略, 具体包括水平翻转、随机旋转 ($\pm 15^\circ$)、亮度/对比度调整及高斯模糊等操作, 增强后的数据集如表 1 所示。

3.2 实验设置

实验基于 PyTorch 2.0 框架, 运行环境为

表 1 实验所用数据集

Table 1 Dataset used in the experiment

数据集	数据量	训练集	增强训练集
BUSI	647	518	2 590
自建数据库	500	400	2 000

Ubuntu 22.04 工作站,配置 NVIDIA RTX 3090 GPU (24 GB)、Intel Core i9-13900K CPU 及 128 GB RAM。

为了全面且客观地评估模型性能,本研究在实验过程中严格保证了 BUSI 数据集与自建数据库的独立性,分别进行训练与评估。具体的数据集划分策略及样本分布如表 2 所示。BUSI 数据集按 8:2 随机划分为训练集与测试集,并从训练集中抽取 10% 作为验证集;自建数据库采用患者级五折交叉验证,每轮取 1 折测试、4 折训练(其中 10% 用于验证)。最终实验结果取 5 次验证的平均值。

表 2 数据集划分策略及样本分布说明

Table 2 Dataset partitioning strategy and sample distribution description

数据集	总样本数	训练集	验证集	测试集
BUSI	647	466	52	129
自建数据库	500	360	40	100

模型训练采用 Adam 优化器,初始学习率设 1×10^{-4} 并使用 StepLR 策略(每 20 epoch 衰减 0.1 倍),批次大小为 8,最大迭代 100 epoch。损失函数由 BCE 与 Dice 加权组合(权重均为 0.5)。为防止过拟合,设置早停策略(耐心值为 20 epoch),并保存验证集表现最优的模型参数。

3.3 评估指标

为全面量化模型性能,本文选取全局准确率(global accuracy, Acc)、精确率(precision)、召回率(recall)、交并比(intersection over union, IoU)及 Dice 相似系数(Dice coefficient, Dice)作为基础指标,主要衡量像素分类精度与区域重叠度。此外,引入 95% 豪斯多夫距离(95% Hausdorff distance, HD95)

以量化预测轮廓与真实边界的贴合程度。在上述指标中,除 HD95 数值越低代表边界误差越小外,其余指标数值均越高表示分割性能越优。

3.4 比较结果

3.4.1 与对照方法的比较

本研究在统一实验框架下复现了 U-Net^[13]、U-Net++^[13]、Att-UNet^[13] 及 TransFuse^[13] 进行对比。所有模型均采用一致的数据划分、增强策略及输入尺寸,并配置相同的 Adam 优化器、复合损失函数及早停机制,以确保对比公平性。

为更直观地比较分割性能,从每个数据集的测试集中选取代表性样本。如图 6 所示,CB-SwinUNet 能更有效地抑制背景干扰并生成光滑边界,即便在低对比度或模糊边缘等复杂场景下,其分割结果仍高度贴合专家标注。

详细的定量对比结果如表 3 和表 4 所示:CB-SwinUNet 在所有数据集上均取得最佳性能。在 BUSI 与自建数据库上,其 Dice 系数分别达到 0.859 和 0.842,较次优模型分别提升 3.0% 和 3.2%。

3.4.2 消融研究

为验证 CASE 与 BADM 模块的有效性,本研究以 SwinUNet 为基线进行消融实验,详情见表 5 和表 6。移除 CASE 模块后,模型对病灶区域的整体判别能力显著减弱。具体表现为:在 BUSI 数据集上,Dice 系数从 0.859 大幅降至 0.798,HD95 由 8.2 恶化至 9.8;在自建数据库上,HD95 亦从 2.5 增至 3.1。这一结果有力证实了 CASE 模块在增强特征显著性及抑制背景噪声方面的关键作用;移除 BADM 模块后,模型对边缘细节的捕捉能力明显下降。具体表现为:在 BUSI 数据集上,召回率从 0.858 降至 0.830,HD95 由 8.2 增至 9.5;在自建数

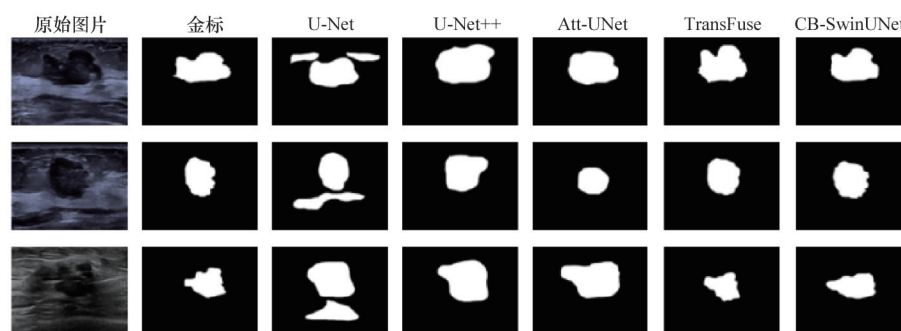


图 6 不同分割方法在乳腺超声图像上的定性比较

Figure 6 Qualitative comparison of different segmentation methods on breast ultrasound images

据库上,HD95 同样从 2.5 增至 2.9。这一结果充分验证了 BADM 利用浅层特征修正深层细节的有效性,对提升边缘刻画精度至关重要;完整 CB-SwinUNet 在绝大多数关键指标上均表现最优,表明双模块的协同作用显著提升了模型在低对比度及复杂形态场景下的分割鲁棒性。

4 讨论与结论

本研究针对乳腺超声图像噪声强、对比度低及边界模糊的难题,提出融合 Swin Transformer 全局建模与 U-Net 细节恢复优势的 CB-SwinUNet。实验表

明,该算法在 BUSI 及自建数据集上的 Dice 系数分别达 0.859 与 0.842,优于 U-Net 及 TransFuse 等主流方法。

模型优异的性能归因于针对性引入的关键模块。首先,CASE 模块通过联合通道与空间注意力加权,自适应抑制背景响应并增强病灶特征,有效解决“前景-背景不平衡”问题并显著提升 IoU 指标。其次,BADM 模块利用编码器浅层高频特征补充深层信息丢失,显著降低 Hausdorff 距离(HD95),有效修正不规则肿瘤轮廓,使分割结果更贴合临床金标准。

表 3 BUSI 数据集上不同方法的分割性能比较

Table 3 Performance comparison of different methods on the BUSI dataset

模型	训练集 Acc	测试集 Acc	交并比	Dice 系数	精确率	召回率
U-Net	0.971	0.919	0.638	0.704	0.710	0.691
U-Net++	0.987	0.922	0.696	0.729	0.717	0.743
Att-UNet	0.973	0.931	0.716	0.763	0.756	0.770
TransFuse	0.988	0.943	0.744	0.829	0.828	0.830
CB-SwinUNet	0.998	0.964	0.756	0.859	0.860	0.858

表 4 自建数据库上不同方法的分割性能比较

Table 4 Performance comparison of different methods on the self-built database

模型	训练集 Acc	测试集 Acc	交并比	Dice 系数	精确率	召回率
U-Net	0.979	0.915	0.569	0.661	0.681	0.642
U-Net++	0.982	0.930	0.625	0.730	0.720	0.741
Att-UNet	0.975	0.931	0.657	0.744	0.732	0.757
TransFuse	0.979	0.952	0.691	0.810	0.796	0.824
CB-SwinUNet	0.994	0.960	0.740	0.842	0.846	0.839

表 5 BUSI 数据集上的消融研究

Table 5 Ablation study on the BUSI dataset

方法	训练集 Acc	测试集 Acc	交并比	Dice 系数	精确率	召回率	HD95
SwinUNet(基线)	0.990	0.948	0.702	0.817	0.815	0.820	10.2
SwinUNet 基线+BADM	0.991	0.950	0.710	0.798	0.820	0.825	9.8
SwinUNet 基线+BADM	0.992	0.953	0.715	0.805	0.825	0.830	9.5
CB-SwinUNet(全模块)	0.998	0.964	0.756	0.859	0.860	0.858	8.2

表 6 自建数据库上的消融研究

Table 6 Ablation study on the self-built database

方法	训练集 Acc	测试集 Acc	交并比	Dice 系数	精确率	召回率	HD95
SwinUNet(基线)	0.992	0.960	0.720	0.832	0.835	0.830	3.3
SwinUNet 基线+BADM	0.993	0.963	0.725	0.837	0.840	0.835	3.1
SwinUNet 基线+CASE	0.994	0.965	0.730	0.843	0.845	0.842	2.9
CB-SwinUNet(全模块)	0.994	0.960	0.740	0.842	0.846	0.839	2.5

尽管表现优异,本算法仍存在局限:一是引入Transformer增加了计算复杂度,实时性略逊于轻量级CNN,未来需在模型轻量化蒸馏方面优化;二是针对直径小于5 mm或伴有严重钙化的微小病灶漏检率略高,后续拟引入加权损失或针对性数据增强加以改善。

综上所述,CB-SwinUNet通过架构创新与多模块融合,成功解决了复杂超声环境下的精准定位与边缘刻画难题,具有较高的临床辅助诊断价值。

参考文献

- [1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2021, 71(3): 209–249.
- [2] Iacob R, Iacob ER, Stoicescu ER, et al. Evaluating the role of breast ultrasound in early detection of breast cancer in low-and middle-income countries: a comprehensive narrative review[J]. Bioengineering, 2024, 11(3): 262.
- [3] Abbott JG, Thurstone FL. Acoustic speckle: theory and experimental analysis[J]. Ultrasonic imaging, 1979, 1(4): 303–324.
- [4] Doi K, Giger ML, Nishikawa RM, et al. Computer-aided diagnosis of breast cancer on mammograms[J]. Breast Cancer, 1997, 4(4): 228–233.
- [5] Siddique N, Paheding S, Elkin CP, et al. U-net and its variants for medical image segmentation: a review of theory and applications[J]. IEEE Access, 2021, 9: 82031–82057.
- [6] Liu LL, Wang Y, Chang J, et al. LLRHNet: multiple lesions segmentation using local-long range features[J]. Frontiers in Neuroinformatics, 2022, 16: 859973.
- [7] Feng JL, Dong XX, Liu XJ, et al. HTBE-Net: a hybrid transformer network based on boundary enhancement for breast ultrasound image segmentation[J]. Displays, 2024, 84: 102753.
- [8] Shamshad F, Khan S, Zamir SW, et al. Transformers in medical imaging: a survey[J]. Medical Image Analysis, 2023, 88: 102802.
- [9] Otsu N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62–66.
- [10] Adams R, Bischof L. Seeded region growing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(6): 641–647.
- [11] Kass M, Witkin A, Terzopoulos D. Snakes: active contour models[J]. International Journal of Computer Vision, 1988, 1: 321–331.
- [12] Zhou ZW, Siddiquee MMR, 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(2018). Granada, Spain: DLMIA, 2018, 11045: 3–11.
- [13] Woo S, Park J, Lee JY, et al. Cbam: convolutional block attention module[C]// European Conference on Computer Vision (ECCV) 2018. Munich, Germany: Springer, Cham, 2018: 3–19.
- [14] Jha D, Smedsrud PH, Riegler MA, et al. ResUNet++: an advanced architecture for medical image segmentation[C]// 2019 IEEE International Symposium on Multimedia (ISM). San Diego, CA, USA: IEEE Press; 2019: 225–2255.
- [15] Punns NS, Agarwal S. RCA-IUnet: a residual cross-spatial attention-guided inception U-Net model for tumor segmentation in breast ultrasound imaging[J]. Machine Vision and Applications, 2022, 33: 27.
- [16] Lou M, Meng J, Qi YL, et al. MCRNet: multi-level context refinement network for semantic segmentation in breast ultrasound imaging[J]. Neurocomputing, 2022, 470: 154–169.
- [17] Liu Z, Lin YT, Cao Y, et al. Swin Transformer: hierarchical vision transformer using shifted windows[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE Press, 2021: 9992–10002.
- [18] Hatamizadeh A, Tang YC, Nath V, et al. UNETR: Transformers for 3D medical image segmentation[C]// 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE Press, 2022: 1748–1758.
- [19] Wang C, Li H, Zhang Y, et al. Sclera-TransFuse: Fusing vision transformer and CNN for accurate sclera segmentation and recognition[J]. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2024, 6(4): 575–590.
- [20] Pan Y, Chen Q, Fang X. DAMAF: dual attention network with multi-level adaptive complementary fusion for medical image segmentation[J]. The Visual Computer, 2025, 41: 2409–2424.

(2025-10-05收稿,2026-02-03修回)