

# 分层 Transformer 和多尺度并行聚合的乳腺肿块分割方法

夏平<sup>1,2</sup>, 王雨蝶<sup>1,2</sup>, 雷帮军<sup>1,2\*</sup>, 彭程<sup>1,2</sup>, 张光一<sup>1,2</sup>, 唐庭龙<sup>1,2</sup>

<sup>1</sup>三峡大学水电工程智能视觉监测湖北省重点实验室, 湖北 宜昌 443002;

<sup>2</sup>三峡大学计算机与信息学院, 湖北 宜昌 443002

**摘要** 针对乳腺超声图像分割中肿块与正常组织间对比度低、边界模糊、肿块形状和位置复杂、图像噪声大等问题, 提出了一种分层 Transformer 与多尺度并行聚合的网络用于乳腺肿块分割。编码器采用 MiT-B2, 建立长距离特征依赖关系, 有效提取不同分辨率下的特征信息; 编码器与解码器间的跳跃连接构建多层分支感受野与置换注意力机制的级联模块; 构建多层分支感受野用以捕获肿块多尺度局部信息, 解决肿块与周围正常组织间高度相似的问题; 置换注意力机制准确识别、定位肿块, 并抑制噪声干扰; 解码端构建聚合模块逐层融合并行分支特征, 提升分割精度。在 BUSI 数据集上实验结果表明, 相较于基线 TransFuse 网络, 本文模型的 Dice、交并比指标分别提升了 3.21%、3.19%。在其他 2 个数据集上, 本文模型的分割性能也有优良的表现。

**关键词** 深度学习; 乳腺肿块分割; 分层 Transformer; 多尺度并行聚合模块

中图分类号 TP183; TP391.41

文献标志码 A

DOI: 10.3788/LOP240836

## Hierarchical Transformer with Multi-Scale Parallel Aggregation for Breast Tumor Segmentation

Xia Ping<sup>1,2</sup>, Wang Yudie<sup>1,2</sup>, Lei Bangjun<sup>1,2\*</sup>, Peng Cheng<sup>1,2</sup>, Zhang Guangyi<sup>1,2</sup>, Tang Tinglong<sup>1,2</sup>

<sup>1</sup>Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering,

Three Gorges University, Yichang 443002, Hubei, China;

<sup>2</sup>College of Computer and Information Technology, Three Gorges University, Yichang 443002, Hubei, China

**Abstract** The problems of breast tumor segmentation from ultrasound images, such as low contrast between the tumor and the normal tissue, blurred boundaries, complex shapes and positions of tumors, and high noise in images, are a concern for researchers. This paper presents a hierarchical transformer with a multiscale parallel aggregation network for breast tumor segmentation. The encoder uses MiT-B2 to establish long-range dependencies and effectively extract features at different resolutions. At the skip connection between the encoder and the decoder, a cascaded module incorporating a multi-scale receptive field block and shuffle attention (SA) mechanism is constructed. receptive field block is used to capture multi-scale local information of the tumor, addressing the problem of high similarity between the lesion and surrounding normal tissue. The SA mechanism accurately identifies and localizes tumors while suppressing noise interference. In the decoder, an aggregation module is constructed to progressively fuse features from parallel branches to enhance segmentation accuracy. The experimental results on the BUSI dataset show that, compared to TransFuse, the proposed model achieves improvements of 3.21% and 3.19% in the Dice and intersection over union metrics, respectively. The model also shows excellent results for the other two datasets.

**Key words** deep learning; breast tumor segmentation; hierarchical Transformer; multi-scale parallel aggregation module

## 1 引言

最新的全球癌症调查报告表明<sup>[1]</sup>, 乳腺癌已成为

全球第一大恶性肿瘤, 是威胁女性健康最常见的癌症。乳腺癌的发病率仅次于肺癌<sup>[2-3]</sup>, 发病率和死亡率呈上升趋势, 早期及时发现和诊断治疗乳腺癌是提高生存

收稿日期: 2024-03-06; 修回日期: 2024-04-17; 录用日期: 2024-04-25; 网络首发日期: 2024-05-06

基金项目: 国家自然科学基金(U1401252)、湖北省重点实验室开放基金(2018SDSJ07)

通信作者: \*Bangjun.Lei@ieee.org

率的关键<sup>[4]</sup>。超声成像已广泛应用于乳腺肿块的临床检测<sup>[5-6]</sup>,但对超声图像肿块的标记分析是一项复杂繁重的工作,仅依靠医生会存在因主观因素导致误检和漏检。随着人工智能技术的发展,计算机辅助诊断为医生提高对乳腺癌的筛查和诊断提供了可能<sup>[7-8]</sup>。乳腺肿块的形态、大小是乳腺癌诊断的主要依据,通过对乳腺肿块的分割,从复杂背景中精确标注肿块区域,有利于提高医生诊断的准确性<sup>[9-10]</sup>。受到超声设备的因素和影响,乳腺超声图像包含固有的斑点噪声,且乳腺肿块形态和位置复杂多样,病变组织与正常组织间对比度较低、边界模糊,准确识别存在一定难度。因而,对超声乳腺肿块准确分割是一项极具挑战的任务。

传统的乳腺肿块分割方法由于肿块位置、形状和大小各异,分割效果并不理想<sup>[11-13]</sup>。深度学习技术被广泛应用于医学图像领域,已取得优异的表现。基于对称结构的 U-Net<sup>[14]</sup>网络作为医学影像分割的主流原型模型,是医学影像分割模型改进和创新的基础。采用编解码思想的 U-Net 网络及其改进网络已成为医学图像分割领域的研究热点<sup>[15]</sup>。Almajalid 等<sup>[16]</sup>基于 U-Net 模型构建了乳腺超声图像分割框架,结合预处理技术提高乳腺图像分割的质量。UNet++<sup>[17]</sup>在 U-Net 基础上嵌套了一系列的密集跳跃路径,形成深度监督的网络。但网络有限的感受野获取的语义信息受限,难以捕捉长距离依赖关系,缺乏对乳腺图像全局信息的建模,导致网络分割性能提升有限。注意力机制聚焦需要重点关注的目标区域,过滤无关信息,是提高网络分割性能的有效手段<sup>[18-19]</sup>。Attention U-Net 网络<sup>[20]</sup>结合 U-Net 和注意力机制,在跳跃连接处设计注意力门控机制(AG),关注显著特征区域的同时抑制了无关区域中的特征。Zhuang 等<sup>[21]</sup>提出了残差注意力网络(RDAU-Net),使用残差模块代替普通卷积模块,克服因网络深度导致的分割性能下降的问题,使用注意力门模块抑制背景信息,增强特征学习能力。Dosovitskiy 等<sup>[22]</sup>使用 Transformer 代替卷积神经网络(CNN),建立了用于图像分类的视觉 Transformer(ViT),并取得了与使用卷积相当的性能。但仅采用 Transformer 结构,缺乏空间归纳偏差,在获取局部细节信息方面存在缺陷,因此,将 Transformer 的全局特征表示能力和卷积的局部特征表示能力结合<sup>[23]</sup>,所提出的 TransUNet 网络<sup>[24]</sup>在肝脏、脑部、前列腺和心脏等医学影像分割中均取得较好的效果。TransFuse 网络<sup>[25]</sup>以并行的方式结合 Transformer 和 CNN,使用 BiFusion 模块融合来自不同分支的多层次特征,在息肉、皮肤病变等医学图像分割中取得了优良的表现。

针对乳腺超声图像的特点,本文提出一种融合分层 Transformer 和多尺度并行聚合的乳腺超声图像肿块分割方法,主要工作:1) 构建非对称结构的乳腺图像分割网络,模型编码采用分层 Transformer,以 MiT-B2 网络<sup>[26]</sup>架构作为编码器,通过 Transformer 与 CNN

的结合,逐层提取乳腺肿块的语义信息和空间细节信息。解码部分构建卷积网络,采用并行分支逐步融合的方式恢复乳腺影像的特征,以减少各尺度特征的语义差异,提升网络的分割效果。2) 跳跃连接处构建多尺度感受野模块(RFB)<sup>[27]</sup>,以减少信息传递中的丢失,同时增强网络分割的泛化能力。3) 为准确定位病变区域,抑制无关信息对分割效果的影响,在多层分支感受野模块后引入置换注意力机制(SA)<sup>[28]</sup>,增强网络对肿块关键特征信息的关注。4) 为有效融合深层特征和浅层特征,在解码端构建聚合模块<sup>[29]</sup>,逐层融合不同尺度的特征,减少特征间的语义差异,提升网络分割精度。

## 2 网络结构

分层 Transformer 和多尺度并行聚合网络主要由三部分组成,编码器采用 MiT-B2 网络架构,建立长距离特征依赖关系,有效提取乳腺超声图像中高分辨率的细小特征和低分辨率的粗特征。编码器与解码器间的跳跃连接构建了 RFB 与 SA 的级联模块;构建 RFB,利用不同分辨率的感受野获取肿块多尺度局部信息;引入 SA,减少肿块分割过程中噪声信息的干扰,在空间和通道层面增强目标区域特征的特征能力。最后,解码端构建聚合模块逐层融合并行分支特征,细化边缘细节,建立局部与全局的依赖关系,获取最终分割预测图。如图 1 所示。

### 2.1 分层 Transformer

为解决乳腺肿块组织与相邻正常组织间对比度低的问题,编码网络采用四阶段的分层 Transformer<sup>[30]</sup>。每阶段均由重叠图块合并层、高效自注意力层、混合前馈网络(Mix-FFN)构成。第 1 阶段生成详细的外观信息,第 2~4 阶段获取乳腺图像中高级的语义信息。与 ViT 相比,分层 Transformer 可接收任意分辨率的乳腺输入图像,利用其分层结构输出乳腺图像的精特征和粗特征,以获取肿块丰富的语义信息和空间位置。同时,Transformer 可实现肿块全局上下文特征信息的建模,有效避免局部细节丢失。采用无位置编码的方式,可适应乳腺图像中测试集与训练集图像分辨率不同的情况,以降低乳腺图像分割误差。如图 2 所示。

通过自注意力层捕捉乳腺图像中像素间局部或全局的依赖关系,从而使模型更好地理解图像特征,提高模型的表现能力。输入相同维度的查询矩阵  $Q$ 、键矩阵  $K$  和价值矩阵  $V$ ,通过矩阵运算实现对乳腺超声图像的形状重塑,提高模型对关键结构的关注度,同时有效过滤图像中的非语义信息和噪声信息,特别是涉及病变区域跨越较大的图像,高效自注意力层能够捕获图像中存在的长距离关联和依赖关系,从而显著提高分割的准确性。在进行多头注意力操作前,先降低  $K$  和  $V$  的空间尺度,采用缩减比例  $R$  降低注意力机制带来的计算量。

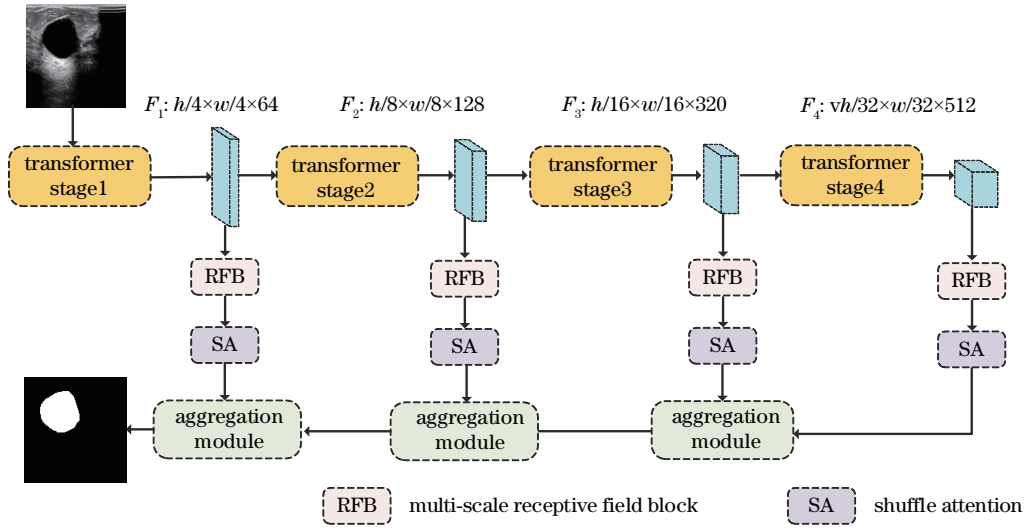


图 1 乳腺肿块分割网络结构  
Fig. 1 Structure of breast tumor segmentation network

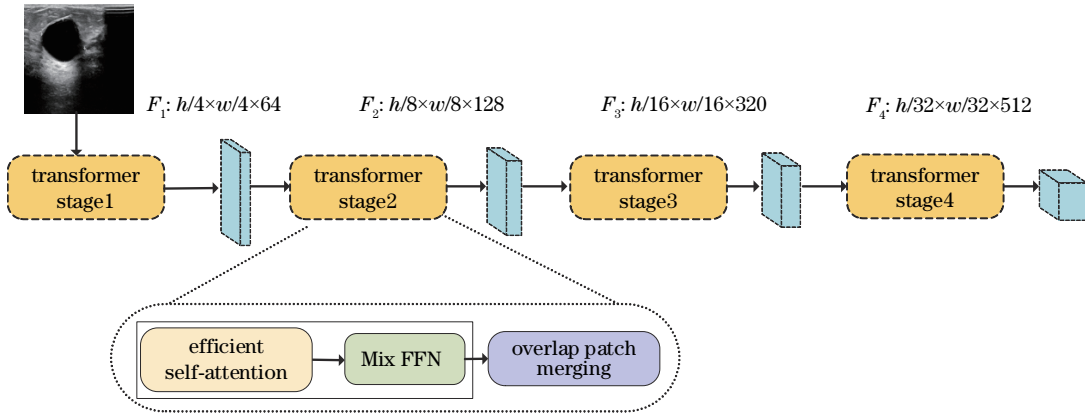


图 2 分层 Transformer 结构  
Fig. 2 Structure of hierarchical Transformer

多头注意力模块表示为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Soft max} \left( \frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_{\text{head}}}} \right) \mathbf{V} \quad (1)$$

式中:  $d_{\text{head}}$  为输入序列长度; 计算复杂度为  $O(N^2)$ , 高效自注意力层使用缩减比例  $R$  后的计算复杂度为  $O(N^2/R)$ , 从而适合于大分辨率的输入图像。

Transformer 编码器中每个注意力头都会关注输入乳腺图像中像素位置, 并自适应地学习重要的像素位置。采用位置编码的方式获取位置信息, 使训练阶段的位置编码固定, 当乳腺测试图像分辨率与训练图像分辨率不一致时, 利用采样操作使分辨率一致, 考虑到语义分割任务是像素级, 位置编码对像素级别的任务影响不大, 因而位置编码在语义分割中并非必需。为此, MiT 避免位置编码, 通过在前馈网络 (FFN) 中采用  $3 \times 3$  的卷积传递位置信息, 组合成 Mix-FFN, 如图 3 所示。高效自注意力层的输出经过混合前馈网络中多个线性层的非线性变换处理, 有效描述图像中的高级特征, 包括形状、纹理、边缘等信息, 提高本文网络

对乳腺超声图像的分割性能。

Mix-FFN 公式为

$$x_{\text{out}} = \text{Linear} \left\{ \text{GELU} \left\{ \text{Conv}_{3 \times 3} \left[ \text{Linear} (x_{\text{in}}) \right] \right\} \right\} \quad (2)$$

式中:  $x_{\text{in}}$  为高效自注意力层输出; Linear 为线性层;  $\text{Conv}_{3 \times 3}$  表示  $3 \times 3$  卷积; GELU 表示激活函数;  $x_{\text{out}}$  为 Mix-FFN 输出。

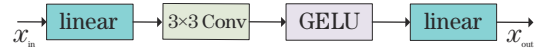


图 3 MiX-FFN 结构  
Fig. 3 Structure of Mix-FFN

重叠图块合并层通过零填充卷积来实现重叠图块的合并, 采用步长为  $S$ 、卷积核大小为  $2S-1$ 、零填充为  $S-1$ 、卷积核个数为  $C_1$  的卷积, 将  $H \times W \times C$  大小的乳腺图像转化为  $(H/S) \times (W/S) \times C_1$  大小的图像, 实现对乳腺影像局部特征连续性建模。采用重叠划分的策略, 旨在促使模型全面理解乳腺结构的时空关系, 更有效地捕捉乳腺的细节信息和局部结构, 增强模型局

部感知能力。通过重叠图块合并,提高模型的感受野,促进模型对乳腺图像上下文信息的理解,有效避免局部细节丢失的问题,提升乳腺图像的分割性能。如图 4 所示。

### 2.2 跳跃连接

乳腺超声图像中肿块形状和位置差异较大,且肿

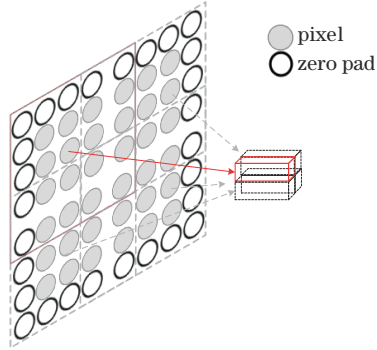


图 4 重叠图块合并结构

Fig. 4 Structure of overlap patch merging

块区域与正常组织间对比度低,若采用单一尺度的特征学习容易忽略乳腺图像不同特征信息间的联系,缺乏在多尺度特征间探索肿块区域,无法建立局部与全局信息的依赖关系。为此,本文在跳跃连接模块中构建多尺度 RFB 学习乳腺肿块的特征信息,增强网络对乳腺肿块的特征提取能力。

RFB 包含四个并行分支,每个分支均包含一个  $1 \times 1$  卷积层,用于减少通道数并将其限制在 64,以控制网络的参数量;后三个分支在  $1 \times 1$  卷积层后,先采用大小为  $1 \times k$  的卷积核进行卷积操作,再采用大小分别为  $k \times 1$  的卷积核进行卷积操作,提取乳腺肿块特征信息;其中,  $k$  分别取 3、5、7;经过三个膨胀系数分别为 3、5、7 的膨胀卷积后,将四个分支提取的肿块多尺度特征信息进行串联拼接,保证整个肿块的信息间充分交互。本文的多层分支结构有效地捕获不同尺度间肿块的特征信息,膨胀卷积在保证网络参数量不变的情况下能有效地增大卷积核的感受野,减少了肿块特征信息的丢失。如图 5 所示。

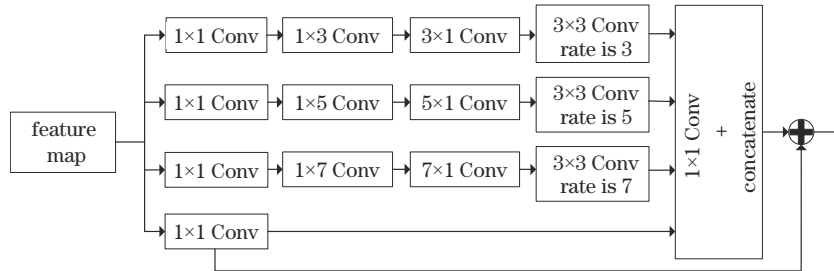


图 5 多尺度感受野模块结构

Fig. 5 Structure of multi-scale receptive field block

乳腺超声图像中噪声的影响,易导致肿块定位错误以及边缘分割不准确。为此,本文跳跃连接在 RFB 后串入 SA,强化不同分支间的信息交互,使网络聚焦于目标肿块,弱化噪声。如图 6 所示。相比于混合注意力机制、瓶颈注意力模块(BAM)<sup>[31]</sup>和卷积注意力模

块(CBAM)<sup>[32]</sup>,SA 的特点在于轻量级和通用性,可有效捕获空间和通道特征的相关性信息,去除非重要信息,以降低计算成本,同时增强共性特征激活,使模型更好地识别、理解、处理这些共性特征,从而提升模型的泛化能力和分割效果。

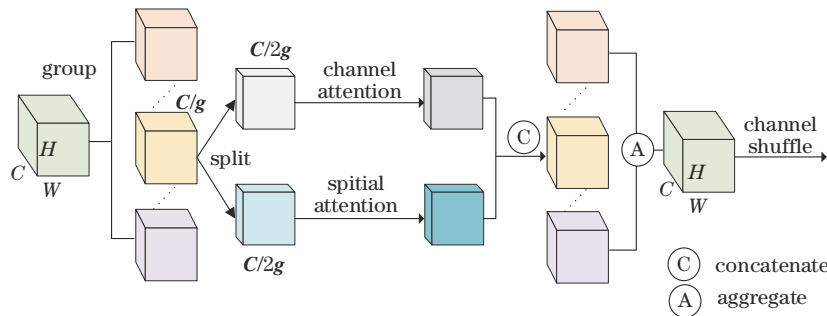


图 6 置换注意力机制结构

Fig. 6 Structure of shuttle attention mechanism

SA 先将输入的乳腺图像沿通道维度进行分组,再将每一组特征图划分为两组,对这两组特征图分别在通道与空间维度上进行注意力操作,同时实现对位置信息和语义信息的关注。其中,通道注意力对输入的

乳腺特征图进行全局平均池化操作,后将其与创建的通道注意力机制的权重和偏置参数相乘,通过 Sigmoid 激活函数进行通道注意力计算,最后将其应用到输入特征图上,以产生加权的通道特征,实现通道注意力操

作。空间注意力是对输入乳腺特征图进行归一化处理,后将其与创建的空间注意力机制的权重和偏置参数相乘,通过 Sigmoid 激活函数进行空间注意力计算,最后将其应用到输入特征图上,以产生加权的空间特征。将通道特征和空间特征进行串联拼接生成每组的乳腺图像特征图输出,聚合每组输出特征图,采用通道混洗<sup>[33]</sup>操作,沿通道维度实现跨组信息交互。

### 2.3 解码聚合模块

低分辨率特征图包含肿块丰富的语义信息,有利

于识别肿块外观细节,高分辨率特征图包含乳腺肿块丰富的空间信息,有利于实现肿块分割的精准定位。在解码端构建聚合模块融合来自分支不同尺度的乳腺特征图,不仅弱化各特征间的语义差距,而且进一步细化肿块的边缘细节,保证肿块区域分割的连续性,提升了乳腺分割的精度。聚合模块采用简单而有效的设计不仅确保模型的计算效率和参数量的合理性,同时也利于降低过拟合风险,提高模型的泛化能力和稳定性。如图 7 所示。

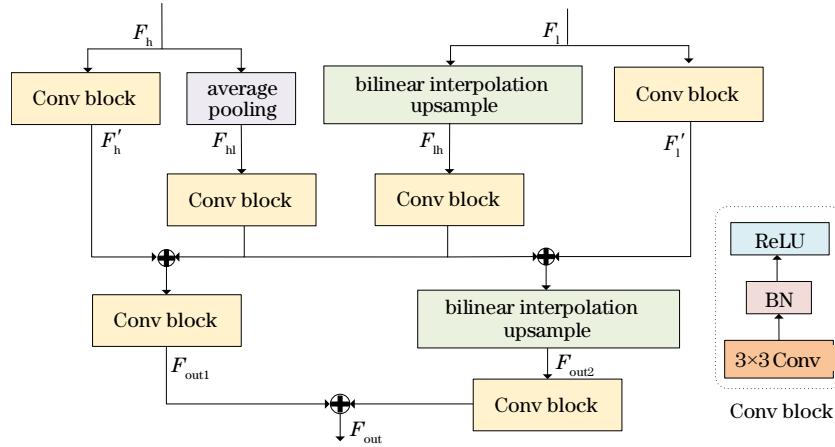


图 7 聚合模块结构

Fig. 7 Structure of aggregation module

具体地:对高分辨率特征  $F_h$  分别进行平均池化下采样和卷积操作,得到特征图  $F_{hl}$  和  $F'_h$ ;对低分辨率特征图  $F_l$  分别进行双线性插值上采样和卷积操作,得到特征图  $F_{lh}$  和特征图  $F'_l$ 。通过平均池化下采样和双线性插值上采样操作,可以在不同分辨率下有效地提取乳腺肿块的特征信息,保留低分辨率特征图的语义信息和高分辨率特征图的空间信息。然后特征图  $F_{hl}$  卷积运算后与特征图  $F'_h$  进行逐元素加法操作,再通过卷积运算得到特征图  $F_{out1}$ ;特征图  $F_{lh}$  卷积运算后与特征图  $F'_l$  进行逐元素加法操作,其结果再经过双线性插值上采样和卷积操作得到  $F_{out2}$ ;特征图  $F_{out1}$  和特征图  $F_{out2}$  逐元素加法运算后输出特征图  $F_{out}$ 。通过逐元素加法操作将不同尺度的特征图进行融合,弥补特征之间的差距,进而实现对乳腺分割任务的精准定位和连续性分割。

## 3 实验结果分析与讨论

### 3.1 实验环境及数据集

实验环境:NVIDIA GeForce RTX3090,显存为 24 GB,Ubuntu20.04 操作系统,编程为 Python3.8。使用 Adam 优化器进行参数优化,初始学习率为  $1 \times 10^{-4}$ ,batch-size 设置为 4。

本文使用 3 个公开数据集来评估网络的分割效果。BUSI 数据集<sup>[34]</sup>收集了 600 例 25~75 岁女性患者 780 幅乳腺图像,其中,良性肿块 437 例,恶性肿块 210 例,正常肿块 133 例。BUSI 数据集是最具挑战性

的数据集之一,因为其中的乳腺肿块形状大小各异且边界复杂<sup>[35]</sup>。UDIAT 数据集<sup>[36]</sup>收集了 163 幅乳腺肿块图像,其中,恶性病例 53 幅,良性病例 110 幅。中山大学癌症中心的 BUS 数据集<sup>[37]</sup>中,患者的平均年龄为  $46.6 \pm 14.2$  岁,有 320 幅乳腺肿块图像,其中,良性肿块 160 例,恶性肿块 160 例。由于数据集中图像分辨率大小不一,为方便训练和测试,预处理时将其调整为  $256 \times 256$ 。选用先进的 TransFuse 网络为基线。

### 3.2 损失函数

为更有效地捕捉乳腺中肿块的边界和形状信息,解决分割中出现误分割、分割不足的问题,在网络中构建二元交叉熵损失  $L_{bce}$  和交并比损失  $L_{iou}$  组成的混合分割损失函数。乳腺超声图像肿块分割属于二分类问题,而二元交叉熵损失常用于像素级别的二分类问题,可帮助网络准确地分类病变和正常组织,从而提高分割的准确性。交并比损失能够更好地关注目标的边界和形状信息,有助于处理边界模糊的乳腺肿块,从而进一步提高分割的质量。

二元交叉熵损失函数  $L_{bce}$  表示为

$$L_{bce} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (3)$$

式中: $N$  表示样本的像素数目; $y_i$  为真实标签; $p_i$  表示模型的预测结果。

混合损失函数  $L$  表示为

$$L = L_{bce} + L_{iou} \quad (4)$$

### 3.3 评估指标

采用 Dice 相似性系数、交并比 (IoU)、召回率、精确度、特异性和准确度等 6 个指标对模型分割性能进行定量分析。

$$D_{\text{Dice}} = \frac{2P_{\text{TP}}}{2P_{\text{TP}} + P_{\text{FP}} + P_{\text{FN}}} \quad (5)$$

$$U_{\text{IoU}} = \frac{P_{\text{TP}}}{P_{\text{TP}} + P_{\text{FP}} + P_{\text{FN}}} \quad (6)$$

$$R_{\text{recall}} = \frac{P_{\text{TP}}}{P_{\text{TP}} + P_{\text{FN}}} \quad (7)$$

$$P_{\text{precision}} = \frac{P_{\text{TP}}}{P_{\text{TP}} + P_{\text{FP}}} \quad (8)$$

$$S_{\text{specificity}} = \frac{P_{\text{TN}}}{P_{\text{TN}} + P_{\text{FP}}} \quad (9)$$

$$A_{\text{accuracy}} = \frac{P_{\text{TP}} + P_{\text{TN}}}{P_{\text{TP}} + P_{\text{TN}} + P_{\text{FP}} + P_{\text{FN}}} \quad (10)$$

式中:  $P_{\text{TP}}$ 、 $P_{\text{FP}}$ 、 $P_{\text{TN}}$ 、 $P_{\text{FN}}$  分别表示真阳性、假阳性、真阴性、假阴性。

### 3.4 实验结果与分析

#### 3.4.1 定量评价

为验证模型的分割性能,在相同实验环境下将本文模型与 U-Net<sup>[14]</sup>、UNet++<sup>[17]</sup>、Attention U-Net<sup>[20]</sup>、

TransUNet<sup>[24]</sup>,以及 TransFuse<sup>[25]</sup>等 5 个深度学习模型的乳腺分割结果进行比较。

本文将 Dice 和 IoU 作为评估模型分割性能的主要指标。6 种分割模型在 3 个公开数据集上的定量评价结果如表 1 所示。表 1 结果表明,在 BUSI 数据集上,本文分割方法在 Dice、IoU、召回率、精确度和准确度等指标上均取得最好的结果,与基线 TransFuse 网络相比, Dice 提升了 3.21 百分点, IoU 提高了 3.19 百分点。在 UDIAT 数据集上,相较于 TransFuse 网络, Dice 和 IoU 分别提高了 1.58 百分点和 1.71 百分点。在 BUS 数据集上, Dice 和 IoU 比 TransFuse 网络分别高出 1.93 百分点和 3.29 百分点。在 BUSI 和 UDIAT 数据集中,基于 CNN 的结构 U-Net 和 Attention U-Net 网络的 Specificity 指标略高于本文网络,主要由于从局部特征中提取肿块的相关信息较易捕获到有助于提高特异性的特征。在 UDIAT 数据集中, TransUNet 网络的 Precision 指标高于本文网络,主要是由于在训练时 TransUNet 网络对 UDIAT 数据集中的特征分布具备更好的适应性,从而提高了分割指标 Precision。综合考虑说明,本文模型在乳腺超声图像肿块分割中,相较于其他 5 种分割网络具有优势,能有效提升乳腺肿块的分割精度。

表 1 不同分割方法在三个数据集上的分割结果

Table 1 Segmentation results of different segmentation methods on three datasets

Dataset	Metric	U-Net	Attention U-Net	UNet++	TransUNet	TransFuse	Proposed
BUSI	Dice	0.7209	0.7453	0.7737	0.8032	0.8193	<b>0.8514</b>
	IoU	0.6284	0.6547	0.6863	0.7181	0.7435	<b>0.7754</b>
	Recall	0.7507	0.8314	0.8358	0.8502	0.8361	<b>0.8680</b>
	Precision	0.7250	0.7291	0.7694	0.7977	0.8246	<b>0.8505</b>
	Specificity	<b>0.9812</b>	0.9745	0.9744	0.9675	0.9675	0.9802
	Accuracy	0.9512	0.9504	0.9574	0.9658	0.9658	<b>0.9710</b>
UDIAT	Dice	0.7433	0.7913	0.7836	0.8310	0.8260	<b>0.8418</b>
	IoU	0.6336	0.7029	0.6723	0.7411	0.7313	<b>0.7484</b>
	Recall	0.7775	0.7953	0.8439	0.8349	0.8668	<b>0.8981</b>
	Precision	0.7554	0.8156	0.7724	<b>0.8634</b>	0.8214	0.8295
	Specificity	0.9915	<b>0.9944</b>	0.9916	0.9940	0.9918	0.9910
	Accuracy	0.9785	0.9856	0.9810	0.9846	0.9865	<b>0.9865</b>
BUS	Dice	0.9271	0.9315	0.9292	0.9371	0.9324	<b>0.9517</b>
	IoU	0.8676	0.8745	0.8705	0.8840	0.8752	<b>0.9081</b>
	Recall	0.9503	0.9382	0.9579	0.9664	0.9405	<b>0.9591</b>
	Precision	0.9117	0.9312	0.9080	0.9143	0.8988	<b>0.9514</b>
	Specificity	0.9317	0.9478	0.9267	0.9326	0.9162	<b>0.9604</b>
	Accuracy	0.9380	0.9428	0.9392	0.9461	0.9405	<b>0.9584</b>

#### 3.4.2 定性评价

图 8 为不同网络对三个乳腺超声图像数据集的分割可视化结果。第 1~3 行病例为 BUSI 数据集<sup>[34]</sup>中 6 个不同网络的分割结果;第 4~6 行病例为 UDIAT 数据集<sup>[36]</sup>中不同网络的分割结果;第 7、8 行病例是在

BUS 数据集<sup>[37]</sup>中不同网络的分割结果。

BUSI 数据集中,乳腺超声图像肿块组织与正常组织对比度较低,因而,网络特征提取能力不足时,容易出现因缺乏学习细节信息而导致误分割的问题。由图 8 可知,本文模型不仅对大区域乳腺肿块具有较好的分割

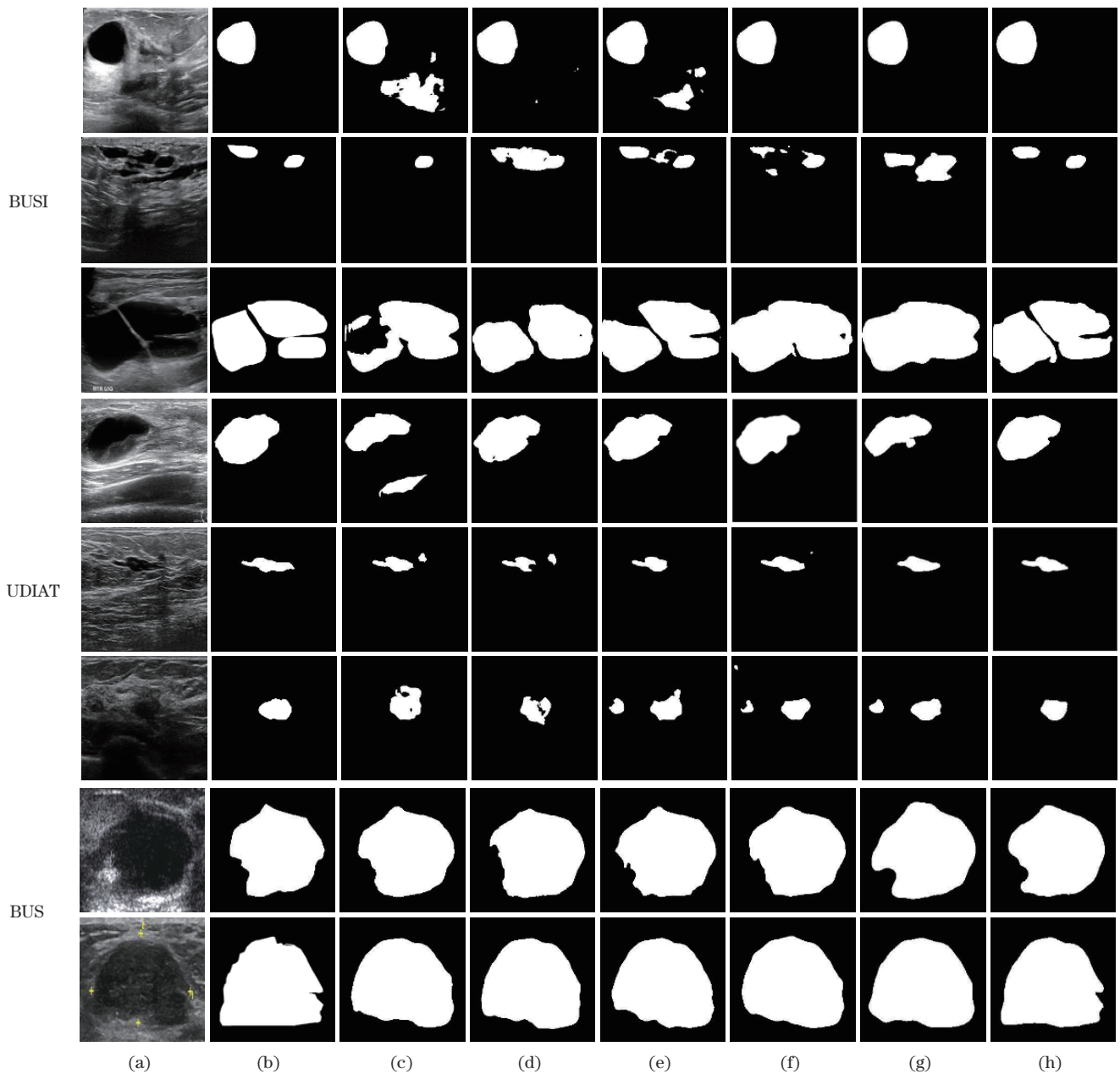


图 8 三个数据集在不同方法下的分割结果可视化。(a)原图;(b)标签;(c) U-Net;(d) Attention U-Net;(e) UNet++;(f) TransUNet;(g) TransFuse;(h) 本文算法

Fig. 8 Visualization of the segmentation results for different methods on three datasets. (a) Original images; (b) mask; (c) U-Net; (d) Attention U-Net; (e) UNet++; (f) TransUNet; (g) TransFuse; (h) proposed

效果,如病例1、病例3,且对多目标肿块也表现出良好的分割效果,如病例2。基线 TransFuse 网络针对大病变区域分割的效果亦较好,但在病变边界的精确分割方面仍存在缺陷;其他4个网络的分割结果中均存在一定程度的误分割、分割过度或分割不足以及分割不连贯的问题,将正常组织识别为病变区域,分割效果不理想。

UDIAT 数据集中,乳腺超声图像包含较多微小病变,由于噪声的影响,有效精准定位微小病变的目标区域存在一定的困难。由图8的病例4~6可知,相比于其他网络,本文方法在肿块边界对比度低的区域仍能较好地识别并处理病变边界信息,在高噪声图像

中表现出较强的鲁棒性,对小目标乳腺肿块分割精准。来源于本文模型通过分层 Transformer 提取不同分辨率的乳腺图像肿块特征,在其中采用 Mix-FFN 传递病变组织位置信息,RFB 减少了肿块特征信息的丢失,SA 实现对肿块病变组织语义信息和位置信息的关注,最后通过解码聚合结构提升网络分割微小病变的能力。

由图8的病例7、8可知,BUS 数据集<sup>[37]</sup>中的乳腺超声图像病变区域较大且对比度明显,这一特征利于病变区域的准确定位,因此所有6个模型均取得较好的分割效果。但相较于其他网络,本文模型可以更好

地识别病变边缘细节,实现边界的精准分割。

为进一步反映不同模型对对比度低的大目标乳腺肿块的分割效果,在 BUSI 数据集<sup>[34]</sup>选取 3 幅典型大目标且对比度低的病例,分割结果如图 9 所示。可以看出,本文方法实现了准确的目标定位和分割,分割结果最接近真实标签,而其他方法均存在误判、欠分割、分割不连续的问题。

针对对比度低的小目标乳腺肿块,不同方法在

UDIAT 数据集<sup>[36]</sup>的分割结果如图 10 所示。在病例 1 和病例 3 中,UNet 网络<sup>[14]</sup>未能准确定位病变区域,Attention U-Net<sup>[20]</sup>、UNet++<sup>[17]</sup>、TransUNet<sup>[24]</sup>则存在误分割、分割不连续的问题。在病例 2 中,U-Net 网络未能分割出目标,TransUNet 和 TransFuse<sup>[25]</sup>同样存在误分割的问题。因此,在处理对比度低、边缘模糊的乳腺图像时,本文模型能够更精准地识别肿块的轮廓,展现更为优良的分割性能。

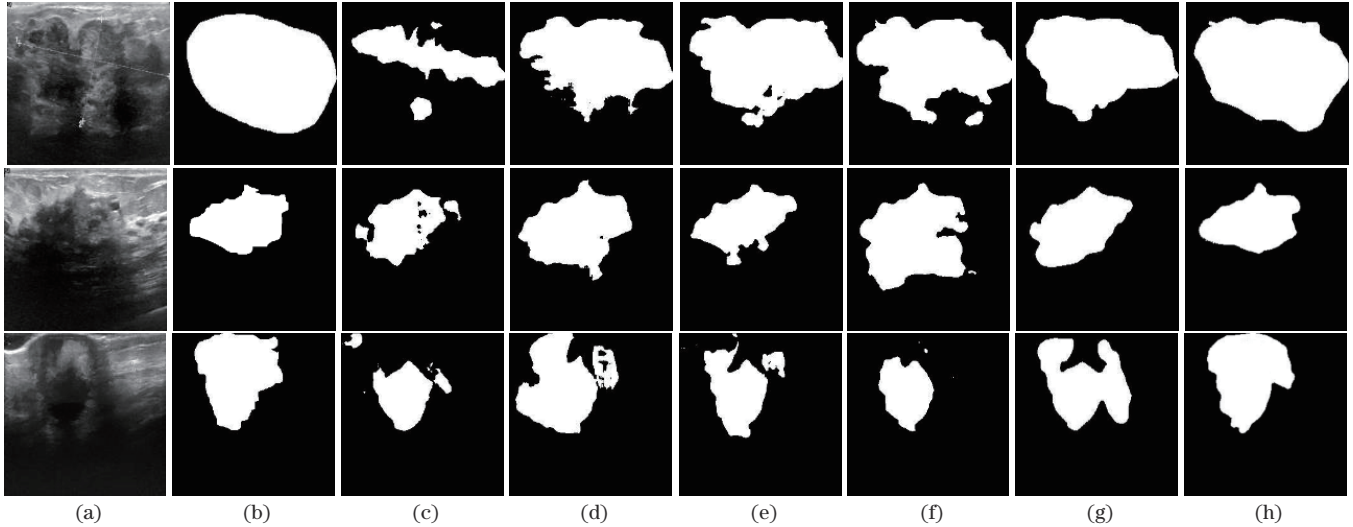


图 9 BUSI 数据集中对比度低的大目标乳腺肿块在不同分割方法下的分割结果可视化。(a)原图;(b)标签;(c) U-Net;(d) Attention U-Net;(e) UNet++;(f) TransUNet;(g) TransFuse;(h)本文算法

Fig. 9 Visualization of segmentation results for cases with large targets and low contrast in BUSI using different methods. (a) Original images; (b)mask; (c) U-Net;(d) Attention U-Net;(e) UNet++;(f) TransUNet;(g) TransFuse; (h) proposed

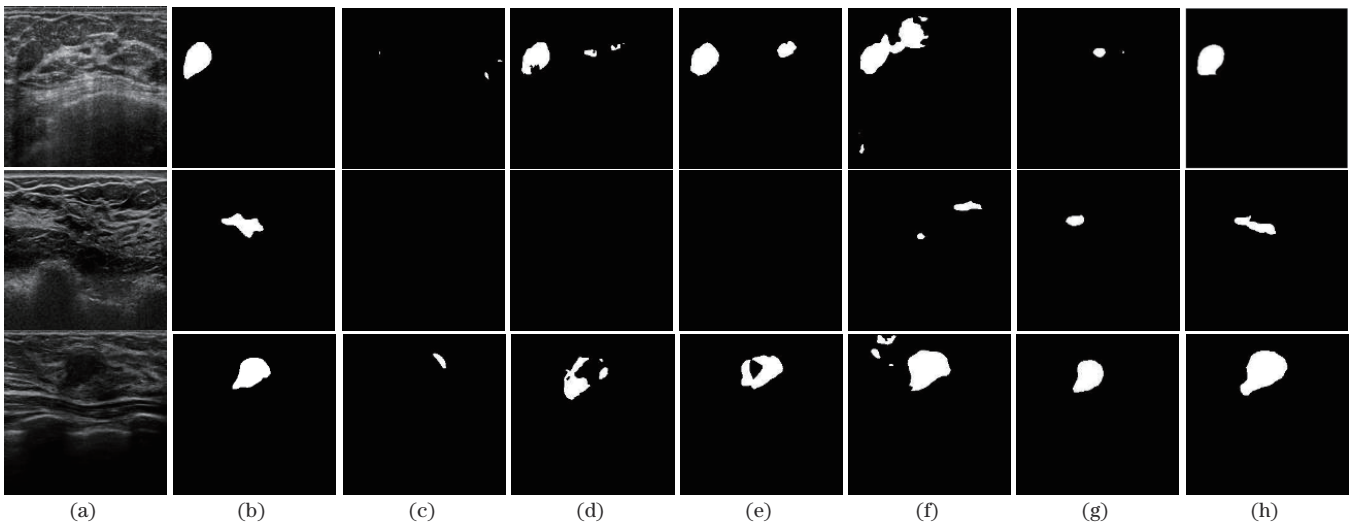


图 10 UDIAT 数据集中对比度低的小目标乳腺肿块在不同分割方法下的分割结果可视化。(a)原图;(b)标签;(c) U-Net;(d) Attention U-Net;(e) UNet++;(f) TransUNet;(g) TransFuse;(h) 本文算法

Fig. 10 Visualization of the segmentation results for cases with large targets and low contrast in the UDIAT using different methods. (a) Original images; (b)mask; (c) U-Net;(d) Attention U-Net;(e) UNet++;(f) TransUNet;(g) TransFuse; (h) proposed

6 种网络在 BUSI 数据集、UDIAT 数据集和 BUS 数据集上分割的 Dice 和 IoU 指标比较如图 11 所示。BUSI 数据集中的乳腺肿块对比度较低,UDIAT 数据集以微小乳腺肿块为主,而 BUS 数据集以较大的病变

肿块为主。由图 11 可知,本文模型在三个数据集上均取得了最高的 Dice 和 IoU 值,表明本文模型在处理对比度低、微小病变和大病变的复杂场景下,相对于其他方法更具优势。



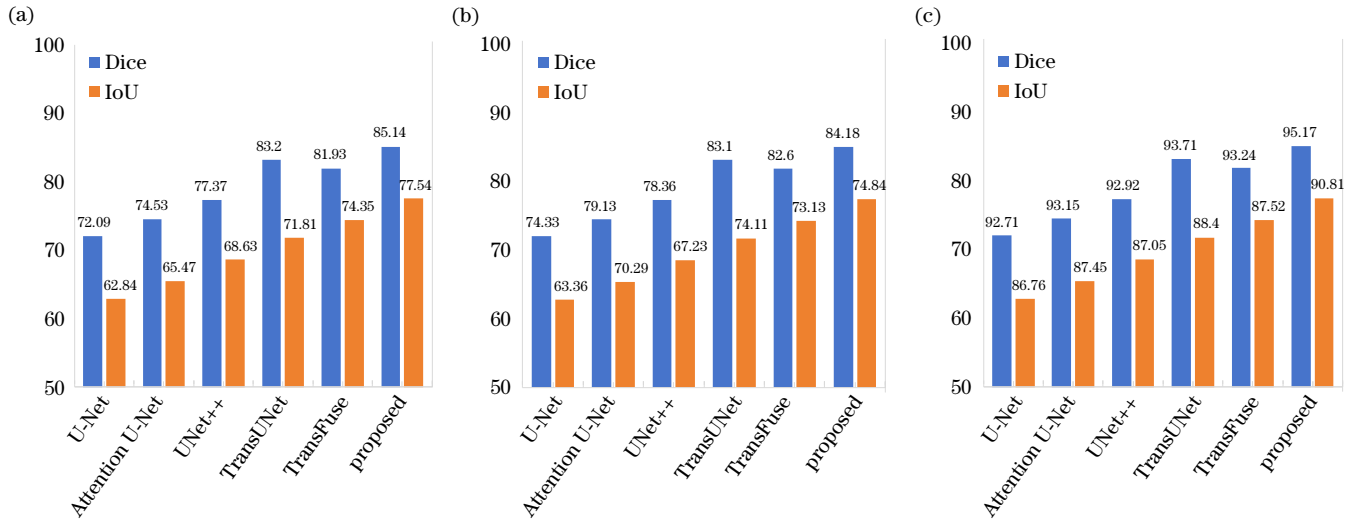


图 11 三种数据集在不同分割方法下Dice 和 IoU 指标的直观对比。(a) BUSI数据集;(b) UDIAT数据集;(c) BUS数据集  
Fig. 11 Intuitive comparison of Dice and IoU metrics with different segmentation methods on three datasets. (a) BUSI dataset; (b) UDIAT dataset; (c) BUS dataset

综上,对乳腺超声图像肿块的分割,本文方法相较于 U-Net、UNet++、Attention U-Net、TransUNet、TransFuse 模型等 5 种模型不仅能很好地分割出对比度低的病变区域,且具有优良的微小病变的检测能力,具有更好的分割效果。

### 3.5 消融实验

为验证本文网络中 MiT、RFB、SA、Aggregation Block 模块的有效性,采用 U-Net 作为基线,在 BUSI、UDIAT 数据集上进行消融实验,不同组合下的分割结果如表 2 所示。在 UDIAT 数据集上消融实验分割结果可视化如图 12 所示。

在 BUSI 数据集上,基线 U-Net 网络分割的 Dice 和 IoU 指标分别为 72.09% 和 62.84%,在 UDIAT 数据集上分别为 74.33% 和 63.36%。将基线的编码器替换为 MiT, MiT 模块建模了远程像素间的依赖关

系,模型的各项指标均有不同程度提升。与基线相比, BUSI 数据集上分割结果的 Dice 和 IoU 分别提高了 6.57 个百分点和 6.79 个百分点, UDIAT 数据集上分别提高了 4.65 个百分点和 5.89 个百分点。由图 13 可知,长距离依赖关系的建模在乳腺超声图像的分割中取得更好的效果。其次,通过对比 MiT+RFB 和基线,在 BUSI 数据集上分割结果的 Dice 和 IoU 分别提高了 11.05 个百分点和 12.28 个百分点, UDIAT 数据集上分别提高了 7.72 个百分点和 9.04 个百分点。由图 12 中 MiT+RFB 的可视化结果可知,模型的分割效果有了一定程度的提升。说明利用 RFB 不同的感受野捕获到更丰富的信息,使模型获取更多的语义信息和空间细节信息。通过对比 MiT+SA 和基线, BUSI 数据集上分割结果的 Dice 和 IoU 分别提高了 10.29 个百分点和 11.72 个百分点, UDIAT 数据集上分别提高了 6.78 个百分点和 8.09 个百分点。图 12

表 2 在 BUSI 数据集和 UDIAT 数据集上消融实验的结果  
Table 2 Results of ablation experiments on BUSI and UDIAT datasets

Dataset	Method				Metric					
	MiT	RFB	SA	Aggregation Block	Dice	IoU	Recall	Precision	Specificity	Accuracy
BUSI					0.7209	0.6284	0.7507	0.7250	0.9812	0.9512
	✓				0.7866	0.6963	0.8079	0.8311	<b>0.9833</b>	0.9538
	✓	✓			0.8314	0.7512	0.8360	<b>0.8523</b>	0.9830	0.9660
	✓		✓		0.8238	0.7456	0.8456	0.8317	0.9775	0.9639
	✓	✓	✓		0.8365	0.7533	0.8638	0.8421	0.9783	0.9663
	✓	✓	✓	✓	<b>0.8514</b>	<b>0.7754</b>	<b>0.8680</b>	0.8505	0.9802	<b>0.9710</b>
UDIAT					0.7433	0.6336	0.7775	0.7554	0.9915	0.9785
	✓				0.7898	0.6925	0.8256	0.8209	<b>0.9931</b>	0.9833
	✓	✓			0.8205	0.7240	0.8439	<b>0.8369</b>	0.9927	0.9853
	✓		✓		0.8111	0.7145	0.8540	0.7824	0.9906	0.9850
	✓	✓	✓		0.8291	0.7315	0.8896	0.8264	0.9922	0.9855
	✓	✓	✓	✓	<b>0.8418</b>	<b>0.7484</b>	<b>0.8981</b>	0.8295	0.9910	<b>0.9865</b>

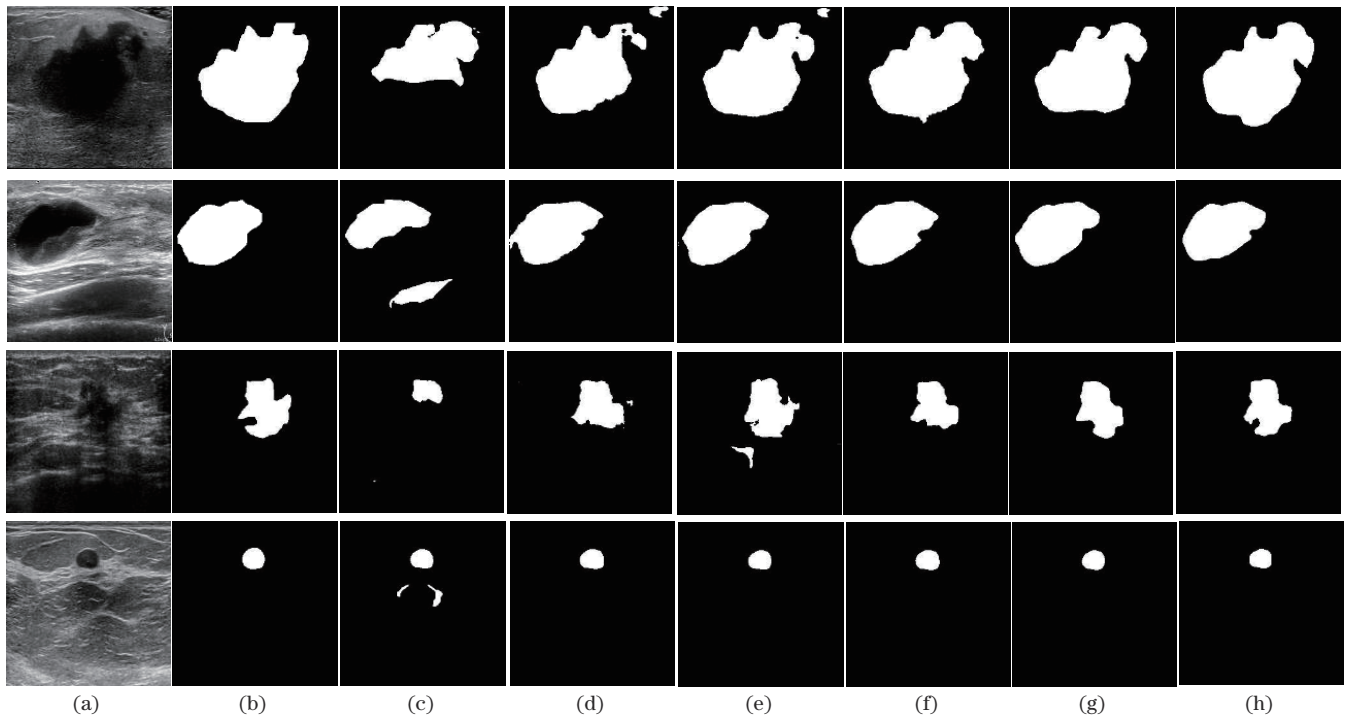


图 12 UDIAT 数据集上消融研究的分割结果可视化。(a)原图;(b)标签;(c) U-Net;(d) MiT;(e) MiT+RFB;(f) MiT+SA;(g) MiT+RFB+SA;(h) RBF+SA+Aggregation Block

Fig. 12 Visualization of segmentation results of ablation studies on UDIAT dataset. (a) Original images; (b) mask; (c) U-Net; (d) MiT; (e) MiT+RFB; (f) MiT+SA; (g) MiT+RFB+SA; (h) RBF+SA+Aggregation Block

中 MiT+SA 的分割结果未出现伪影的情况,能准确定位待分割区域。实验结果说明,置换注意力机制加强空间和通道相关性信息的表征能力,增强了模型获取乳腺肿块关键特征的能力,同时抑制了背景信息的响应。最后,在 MiT 中引入 RBF+SA+Aggregation Block,与 MiT+RFB+SA 相比,BUSI 数据集上的 Dice 和 IoU 分别提高了 1.49 个百分点和 2.21 个百分点,UDIAT 数据集上分别提高了 1.27 个百分点和 1.69 个百分点。对比图 12 中 MiT+RFB+SA 与 RBF+SA+Aggregation Block 的分割结果,RBF+SA+Aggregation Block 的分割结果更加接近真实标签,其边缘细节更为完整,且没有出现误分割和分割不连续的情况。实验结果表明,聚合模块实现深层、浅层特征信息的交互融合,减少语义鸿沟,有效捕获乳腺肿块边缘结构信息,建立局部与全局的依赖关系,从而实现更好的乳腺图像分割性能。

## 4 结 论

本文提出了分层 Transformer 与多尺度并行聚合网络用于乳腺超声图像肿块的分割,有效解决乳腺超声图像中病灶区域和周围正常组织对比度低、微小肿块误分割等问题。模型利用分层 Transformer 结构提取乳腺肿块的外观信息和高层次语义信息,实现了信息间的交互并建立长距离信息的依赖关系;在跳跃连接处构建多尺度感受野模块和置换注意力机制,获取丰富的多尺度信息利于识别乳腺肿块中不同大小的病

变目标,同时实现目标区域的精准定位,有效抑制无关信息的影响;最后解码端构建聚合模块通过逐层融合并行分支特征,有效融合深层信息、浅层信息和全局上下文信息,减少语义鸿沟,构建了局部与全局的联系,得到乳腺病变区域预测分割图。实验结果表明,本文模型整体分割效果优于 U-Net、UNet++、Attention U-Net、TransUNet、TransFuse 等现有方法,具有更好的鲁棒性。

理论建模与实验结果表明,本文模型应用于乳腺超声图像的分割,有助于医生对乳腺肿块的准确诊断和后续治疗。课题组后续进一步优化模型,增强模型抗阴影干扰的能力;此外,正着手研究和探索基于通用分割模型的乳腺超声图像分割方法,进一步提升对乳腺图像分割性能。

## 参 考 文 献

- [1] Sung H, Ferlay J, Siegel R L, 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] Siegel R L, Miller K D, Fuchs H E, et al. *Cancer statistics, 2022*[J]. CA: Cancer Journal for Clinicians, 2022, 72(1): 7-33.
- [3] Zheng R S, Zhang S W, Zeng H M, et al. *Cancer incidence and mortality in China in 2016*[J]. Journal of the National Cancer Center, 2022, 2(1): 1-9.

- [4] Bleicher R J, Ruth K, Sigurdson E R, et al. Time to surgery and breast cancer survival in the United States[J]. *JAMA Oncology*, 2016, 2(3): 330-339.
- [5] Ginsburg O, Yip C H, Brooks A, et al. Early detection of breast cancer: a phased approach to implementation[J]. *Cancer*, 2020, 126(Suppl 10): 2379-2393.
- [6] Sood R, Rositch A F, Shakoor D, et al. Ultrasonography for breast cancer detection: a systematic review and meta-analysis[J]. *Journal of Global Oncology*, 2019, 5: 1-17.
- [7] Yang S L, Gao X C, Liu L W, et al. Performance and reading time for automated breast US with and without computer-aided detection[J]. *Radiology*, 2019, 292(3): 540-549.
- [8] Huang Q H, Zhang F, Li X L. Machine learning in ultrasound computer-aided diagnostic systems: a survey[J]. *BioMed Res International*, 2018, 2018:5137904.
- [9] Hupse R, Samulski M, Lobbes M B, et al. Computer-aided detection of masses on mammography: interactive decision support versus prompts[J]. *Radiology*, 2013, 266(1): 123-129.
- [10] Sanders D R, Vukich J A, Doney K, et al. U.S. food and drug administration clinical trial of implantable contact lenses for moderate-to-high myopia[J]. *Ophthalmology*, 2003, 110(2): 255-266.
- [11] Karunanayake N, Aimmanee P, Lohitvisate W, et al. Particle method for the segmentation of breast tumors in ultrasound images[J]. *Mathematics and Computers in Simulation*, 2020, 170:257-284.
- [12] Mughal B, Muhammad N, Sharif M. Adaptive hysteresis thresholding segmentation technique for localizing breast masses in the curve-stitching domain[J]. *International Journal of Medical Informatics*, 2019, 126:26-34.
- [13] Chan T F, Vese L. Active contour and segmentation models using geometric PDE's for medical imaging[M]//Malladi R. *Geometric methods in biomedical image processing*. Mathematics and Visualization. Berlin: Springer, 2002: 63-75.
- [14] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. *Medical image computing and computer-assisted intervention MICCAI 2015*. *Lunk Notes in Computer Science*. Cham: Springer, 2015, 9351: 234-241.
- [15] 张欢, 仇大伟, 冯毅博, 等. U-Net模型改进及其在医学图像分割上的研究综述[J]. *激光与光电子学进展*, 2022, 59(2): 0200005.
- Zhang H, Qiu D W, Feng Y B, et al. Improved U-Net models and their applications in medical image segmentation: a review[J]. *Laser Optoelectronics Progress*. 2022; 59(2): 0200005.
- [16] Almajalid R, Shan J, Du Y D, et al. Development of a deep learning-based method for breast ultrasound image segmentation[C]//2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), December 17-20, 2018, Orlando, FL, USA. New York: IEEE Press, 2018: 1103-1108.
- [17] Zhou Z W, Siddiquee M R, Tajbakhsh N, et al. UNet++: nested U-Net architecture for medical image segmentation[M]//Stoyanov D, Taylor Z, Carneiro G, et al. *Deep learning in medical image analysis and multimodal learning for clinical decision support*. *Lunk Notes in Computer Science*. Cham: Springer, 2018, 11045: 3-11.
- [18] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention[EB/OL]. (2014-06-24) [2024-04-10]. <https://arxiv.org/abs/1406.6247>.
- [19] 赵凤, 钟蓓蓓, 刘汉强. 基于注意力机制的多尺度残差U-Net眼底血管分割[J]. *激光与光电子学进展*, 2022, 59(18): 1810002.
- Zhao F, Zhong B B, Liu H Q. Multi-scale residual U-Net fundus blood vessel segmentation based on attention mechanism[J]. *Laser Optoelectronics Progress* 2022, 59(18), 1810002.
- [20] Oktay O, Schlemper J, le Folgoc L, et al. Attention U-Net: learning where to locate the pancreas[EB/OL]. (2018-04-11)[2024-04-10]. <https://arxiv.org/abs/1804.03999>.
- [21] Zhuang Z M, Li N, Raj A N J, et al. RDAU-NET model for lesion segmentation in breast ultrasound images[J]. *PLoS One*, 2019, 14(8): e0221535.
- [22] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image worth 16 × 16 words: transformers for image recognition at scale[EB/OL]. (2020-10-22) [2024-04-10]. <https://arxiv.org/abs/2010.11929>.
- [23] 郝文月, 蔡怀宇, 左廷涛, 等. 融合Transformer分支和拓扑强化的血管内超声图像分割方法[J]. *激光与光电子学进展*, 2024, 61(12): 1237008.
- Hao W Y, Cai H Y, Zuo T T, et al. Intravascular ultrasound image segmentation using transformer branches and topology enforcement[J]. *Laser Optoelectronics Progress* 2024, 61(12), 1237008.
- [24] Chen J N, Lu Y Y, Yu Q H, et al. TransUNet: transformers are strong encoders used for medical image segmentation[EB/OL]. (2021-02-08)[2024-04-10]. <https://arxiv.org/abs/2102.04306>.
- [25] Zhang Y D, Liu H Y, Hu Q. TransFuse: fusing transformers and CNNs for medical image segmentation [EB/OL]. (2021-02-16)[2024-04-10]. <https://arxiv.org/abs/2102.08005>.
- [26] Zhou B L, Zhao H, Puig X, et al. Semantic understanding of scenes using ADE20K dataset[J]. *International Journal of Computer Vision*, 2019, 127(3): 302-321.
- [27] He K M, Zhang X Y, Zhang, Ren S Q, et al. Deep residual learning for image recognition[EB/OL]. (2015-12-10)[2024-04-10]. <https://arxiv.org/abs/1512.03385>.
- [28] Zhang Q L, Yang YB. SA-Net: shuffle attention for deep convolutional neural networks[C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), June 6-11, 2021, Toronto, ON, Canada. New York: IEEE Press, 2021: 2235-2239.
- [29] Qiu Z H, Wang Z C, Zhang M, et al. BDG-Net: boundary distribution-guided network for accurate polyp segmentation[EB/OL]. (2022-04-17)[2021-04-10]. <https://arxiv.org/abs/2201.00767>.
- [30] Xie E Z, Wang W H, et al. SegFormer: a simple and

- efficient design for semantic segmentation with transformers[J]. *Advances in Neural Information Processing Systems*, 2021, 34:12077-12090.
- [31] Park J, Woo S, Lee J Y, et al. **Bam: bottleneck attention module**[EB/OL]. (2018-07-18) [2024-04-10]. <https://arxiv.org/abs/2010.11929>.
- [32] Woo S, Park J, Lee J Y, et al. **CBAM: convolutional block attention module**[EB/OL]. (2018-07-18)[2024-04-10]. <https://arxiv.org/abs/1807.06521>.
- [33] Ma N N, Zhang X Y, Zheng H T, et al. **ShuffleNet V2: practical guidelines for efficient CNN architectural design** [EB/OL]. (2018-07-30)[2024-04-10]. <https://arxiv.org/abs/1807.11164>.
- [34] Al-Dhabyani W, Gomaa M, Khaled H, et al. **Breast ultrasonography image dataset** [J]. *Data in Brief*, 2020, 28: 104863.
- [35] Xu M, Huang K, Chen Q, et al. **Mssa-Net: multi-scale self-attention network for breast ultrasound image segmentation**[C]//2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), April 13-16, 2021, Good, France. New York: IEEE Press, 2021: 827-831.
- [36] Yap M H, Goyal M, Osman F, et al. **Breast ultrasound region of interest detection and lesion localization** [J]. *Artificial Intelligence in Medicine*, 2020, 107:101880.
- [37] Huang Q H, Huang Y H, Luo Y Z, et al. **Segmentation of breast-ultrasound images with semantic classification of superpixels**[J]. *Medical Image Analysis*, 2020, 61: 101657.