

基于多尺度特征提取与多特征融合的甲状腺结节 超声影像分割

闫家奕 魏国辉

摘要 目的 针对甲状腺结节超声影像分割中结节尺寸差异显著、图像低对比度及强斑点噪声干扰的问题,提出一种基于多尺度特征提取与全局多特征融合的精准分割方法,旨在提升不同尺度结节的识别能力与复杂背景下的泛化性能。**方法** 采用经典的编码器-解码器网络架构作为基础框架。在此基础上,集成多尺度残差连接模块(multi-scale residual connection module, MRCM),用于捕获传统固定感受野卷积核难以同时捕获到的多尺度结节特征。同时,引入改进的全局金字塔引导模块(improved global pyramid guidance, IGPG),用于完成高效且去噪的全局多特征融合。使用本文所构建的方法在TN3K和DDTI两个公开数据集上进行了验证。**结果** MRCM通过密集堆叠3个级联的3×3卷积核,实现感受野的连续扩展,使其能够自适应地捕捉从精细局部纹理到更大范围上下文信息的不同尺度特征,有效应对结节尺寸的剧烈变化。IGPG摒弃空洞卷积,转而利用MRCM进行跨层特征融合,并以深层语义信息作为空间注意力向导,有效筛选并整合多阶段特征,从而显著抑制噪声干扰,增强判别性特征的表达能力。在TN3K数据集上,本文算法在交并比(intersection over union, IoU)、95%豪斯多夫距离(Hausdorff distance-95%, HD95)和F1分数上分别达到了72.21%、18.49和80.97%,相较于U-Net基准, IoU提升了4.0%, F1提升了3.5%。在DDTI数据集上,本文算法的IoU、HD95和F1分数分别为63.40%、18.61和74.97%,相较于U-Net基准, IoU提升了3.0%, F1提升了2.4%。结果均显著优于FCN、U-Net、MultiResUNet、SGUNet和DC-UNet等现有主流方法,并在消融实验中证明了MRCM和IGPG模块的有效性和协同作用。**结论** 基于MRCM和IGPG的分割方法通过协同作用,可有效解决甲状腺结节超声影像分割中的尺度变异和噪声干扰问题,显著提升分割精度和鲁棒性,为甲状腺结节计算机辅助诊断系统提供新的技术方案。

关键词 甲状腺结节;超声影像;图像分割;多尺度特征提取;多特征融合

DOI: 10.3969/j.issn.1002-3208.2026.03.002.

中图分类号 R318.04 **文献标志码** A **文章编号** 1002-3208(2026)03-0229-10

本文著录格式 闫家奕,魏国辉. 基于多尺度特征提取与多特征融合的甲状腺结节超声影像分割[J]. 北京生物医学工程, 2026, 45(3): 229-238. YAN Jiayi, WEI Guohui. Ultrasound image segmenation of thyroid nodules based on multi-scale feature extraction and multi-feature fusion [J]. Beijing Biomedical Engineering, 2026, 45(3): 229-238.

Ultrasound image segmenation of thyroid nodules based on multi-scale feature extraction and multi-feature fusion

YAN Jiayi, WEI Guohui

Shandong University of Traditional Chinese Medicine, Jinan 250355

Corresponding author: WEI Guohui (E-mail: bmie530@163.com)

[Abstract] **Objective** To address the challenges of significant variations in nodule size, low image

contrast, and strong speckle noise in ultrasound image segmentation of thyroid nodules, this paper proposes a precise segmentation method based on multi-scale feature extraction and global multi-feature fusion, aiming to enhance the recognition capability of features at different scales and improve generalization performance under complex backgrounds. **Methods** A classical encoder-decoder network architecture is adopted as the foundational framework. Building upon this, a Multi-scale Residual Connection Module (MRCM) is integrated to capture multi-scale nodule features that traditional fixed-receptive-field convolution kernels struggle to capture simultaneously. Concurrently, an Improved Global Pyramid Guidance (IGPG) module is introduced to achieve efficient and denoised global multi-feature fusion. The developed method was validated on two public datasets, TN3K and DDTI. **Results** The MRCM achieves continuous expansion of the receptive field by densely stacking three cascaded 3×3 convolution kernels, enabling it to adaptively capture features at different scales—from fine local textures to broader contextual information—effectively addressing drastic variations in nodule size. The IGPG module abandons dilated convolutions, instead utilizing MRCM for cross-layer feature fusion, and employs deep semantic information as a spatial attention guide to effectively filter and integrate multi-stage features, thereby significantly suppressing noise interference and enhancing the expression ability of discriminative features. On the TN3K dataset, the algorithm achieved IoU, HD95, and *F1* scores of 72.21%, 18.49, and 80.97% respectively. Compared to the U-Net baseline, IoU increased by 4.0% and *F1* by 3.5%. On the DDTI dataset, the algorithm's IoU, HD95, and *F1* scores were 63.40%, 18.61, and 74.97% respectively. Compared to the U-Net baseline, IoU increased by 3.0% and *F1* by 2.4%. The results significantly outperformed existing mainstream methods such as FCN, U-Net, MultiResUNet, SGUNet, and DC-UNet, and the effectiveness and synergistic effect of the MRCM and IGPG modules were demonstrated in ablation experiments. **Conclusions** By synergistically integrating the MRCM and IGPG modules, the proposed segmentation method effectively addresses the issues of scale variation and noise interference in ultrasound image segmentation of thyroid nodules, markedly improving segmentation accuracy and robustness, thereby offering a novel technical solution for computer-aided diagnosis systems of thyroid nodules.

【Keywords】 thyroid nodule; ultrasound imaging; image segmentation; multi-scale feature extraction; multi-feature fusion

0 引言

甲状腺是重要的人体内分泌器官,通过分泌甲状腺激素调节机体代谢活动。甲状腺结节是由甲状腺组织细胞异常增殖导致的离散性病变,是内分泌系统最常见的疾病之一^[1]。依据病理质地,甲状腺结节通常可分为实性结节、囊性结节以及混合性结节^[2]。临床研究表明,无论实性或囊性结节均存

在恶性转化风险。国家癌症中心 2017 年发布的数据显示,我国甲状腺癌发病率已位居恶性肿瘤第 7 位,且呈逐年上升趋势^[3]。因此,实现甲状腺结节的早期精准发现、准确诊断与及时有效治疗,对于改善患者预后、降低疾病负担具有极其重要的临床意义和现实需求。

当前,电子计算机断层扫描(computed tomography, CT)、磁共振成像(magnetic resonance imaging, MRI)以及超声成像等技术广泛应用于甲状腺疾病的影像学诊断^[4]。在这些成像方法中超声成像因具有无创性、经济实用、实时动态观察等优势,已成为甲状腺结节筛查与诊断的首选影像学方法^[5]。然而,临床实践中约 90% 的甲状腺结节为良性病变^[6],恶性结节的鉴别诊断高度依赖医师的临

基金项目:山东省自然科学基金面上项目(ZR2022MH203)、山东省研究生教育优质课程和专业学位研究生教学案例库立项项目(SDYAL20050)、山东中医药大学大学生创新训练计划项目(2025227)资助

作者单位:山东中医药大学医学信息工程学院(济南 250355)

通信作者:魏国辉。E-mail:bmie530@163.com

床经验。更具挑战性的是,超声图像固有的低对比度、强斑点噪声及伪影干扰等问题,极大地增加了诊断的复杂性,使得经验不足的医师易出现误诊或漏诊,甚至延误最佳治疗时机^[7]。鉴于此,研发高精度、高鲁棒性的计算机辅助诊断(computer-aided diagnosis, CAD)系统,以期提升甲状腺结节良恶性诊断的客观性、准确性及效率,是当前医学影像分析领域亟待解决的关键科学问题。

深度学习技术在医学影像分割领域取得了突破性进展。在通用语义分割领域,Song等^[8]提出的全卷积网络(fully convolutional networks, FCN)开创了端到端图像分割的先河;Lafraxo等^[9]则系统阐述了空洞卷积(atrous convolution)在扩展感受野、平衡分辨率与感受野关系方面的优势,为后续医学影像分割方法提供了重要的技术基础。Haribabu等^[10]提出的U-Net网络采用编码器-解码器架构,通过跳跃连接融合浅层细节特征与深层语义特征,在医学影像分割任务中取得了里程碑式的成功,并成为后续众多医学图像分割方法的基础。然而,U-Net在处理边界模糊、形态多变的病灶时,仍易出现分割边界不清晰的问题。针对该局限性,Liu等^[11]提出的TransUNet网络,将Transformer^[12-13]引入到U-Net架构,利用自注意力机制建模长距离依赖关系,以增强全局上下文信息的捕获能力。在甲状腺结节分割领域,现有研究已取得重要进展:曹玉珍等^[14]提出双分支融合网络,并行构建基于Transformer的全局特征提取分支与基于卷积神经网络的局部特征提取分支,但其采用固定尺度的卷积核难以动态适应结节尺寸的剧烈变化;Liu等^[15]将柯尔莫哥洛夫-阿诺德卷积(Kolmogorov-Arnold convolution, KAC)模块与卷积神经网络(convolutional neural network, CNN)和Transformer集成用于甲状腺超声图像分割;Bi等^[16]提出了一种用于超声甲状腺结节分割的新型保边界组装TransformerUNet(BPAT-UNet),增强边界特征并构建自适应多尺度特征融合模块,但仍局限于通道维度融合;传统U-Net的跳跃连接直接传递浅层细节,在超声低信噪比条件下易将背景噪声引入深层网络。Deng等^[17]虽结合Swin Transformer与CNN,但未解决跨层噪声传递问题。可借鉴门控跳跃连接(gated skip connection)或注意力引导的特征筛选机制,在融合前对浅层特征进行

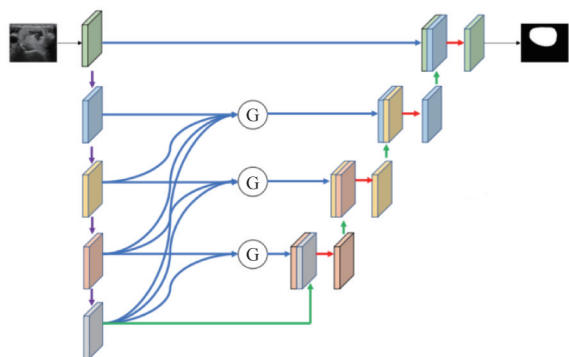
噪声抑制与信息筛选,仅保留对分割任务有益的边界细节,避免误分割与边界模糊。TransUNet等纯Transformer方法虽能捕获长距离依赖,但自注意力的平滑效应可能导致细粒度边界信息丢失;而纯CNN方法感受野受限,难以建立全局语义关联。现有研究表明,在医学影像分割中,基于卷积神经网络的深度优化以及与Transformer等新型架构的融合,仍然是提升模型性能的关键方向。

针对上述科学问题,本文提出一种基于多尺度特征提取与全局多特征融合的甲状腺结节超声图像分割方法。本研究的核心创新在于,针对甲状腺结节显著的尺度变异性导致的多尺度特征表征不足问题,设计了多尺度残差连接模块(multi-scale residual connection module, MRCM)。该模块通过级联小卷积核的密集堆叠,实现了感受野的连续扩展,能够自适应地捕获从微小结节到巨大病灶的有效特征,克服了传统固定感受野卷积核的局限性。为解决超声影像低信噪比导致的跨层特征融合噪声敏感性问题,提出了改进的全局金字塔引导模块(improved global pyramid guidance, IGPG)。IGPG摒弃了传统空洞卷积易引入信息不连续性的缺点,转而利用MRCM进行跨层特征融合,并以深层语义信息作为空间注意力向导,有效筛选并整合多阶段特征,从而显著抑制噪声干扰,增强判别性特征的表达能力。

本文方法在保留精细边界信息的同时,有效整合了全局语义上下文,实现了全局信息与局部细节的平衡,从而在超声图像质量受限条件下,显著提升了甲状腺结节的精准分割能力和鲁棒性。旨在有效解决甲状腺结节超声影像分割中存在的尺度变异和噪声干扰问题,为临床甲状腺结节计算机辅助诊断系统提供可靠的技术方案。

1 方法

本文以编码器-解码器网络为框架,采用多尺度特征提取模块替代传统网络各阶段的普通卷积操作,以实现多尺度特征的有效提取,同时引入全局多特征融合策略改进跳跃连接机制,增强跨层特征融合能力并抑制噪声干扰。网络整体结构如图1所示。



紫色箭头表示多尺度特征提取与下采样;绿色箭头表示上采样;红色箭头表示多尺度特征提取;G为改进的全局金字塔引导模块(IGPG);编码器阶段1至阶段5的特征图尺寸分别为输入图像的1、1/2、1/4、1/8、1/16;输入为甲状腺超声图像,输出为结节分割掩膜

图1 基于多尺度特征提取与全局多特征融合的甲状腺结节分割网络结构

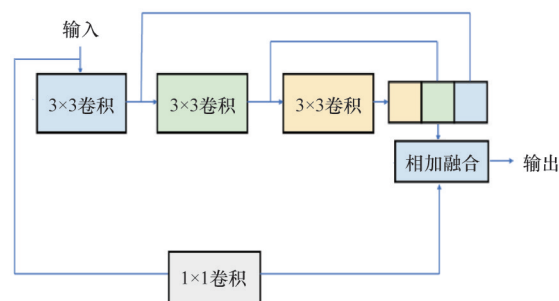
Figure 1 Network structure of thyroid nodule segmentation based on multi-scale feature extraction and global multi-feature fusion

1.1 多尺度特征提取模块

甲状腺结节在超声影像中呈现显著的尺度差异性,结节直径可从数毫米至数厘米不等,这种尺度变异对特征提取的适应性提出了严峻挑战。传统固定尺寸的卷积核难以同时捕获微小结节与巨大病灶的有效特征。为克服这一瓶颈,本文创新性地采用了多尺度残差连接模块^[18]实现自适应多尺度特征提取。

MRCM模块由3个级联的 3×3 卷积层、一个 1×1 卷积层、批量归一化层(batch normalization, BN)^[19]及线性整流函数层(rectified linear units, ReLu)^[20]构成。与直接使用大卷积核或并行多分支空洞卷积不同,MRCM的核心创新在于通过3个级联的 3×3 卷积层实现感受野的连续扩展。根据文献^[21]的理论分析,2个 3×3 卷积层的堆叠等效于 5×5 卷积核的感受野,3个 3×3 卷积层则等效于 7×7 卷积核。这种级联策略在显著降低参数量与计算开销的同时,确保了感受野的连续性和特征的层级精炼。浅层卷积(3×3)能够捕捉精细的局部纹理和边缘信息,中层卷积(等效 5×5)捕获中等尺度的结构特征,而深层卷积(等效 7×7)则能感知更大范围的上下文信息。将不同感受野的特征在通道维度拼接,并与输入特征进行残差连接^[22],形成互补且丰富的多尺度特征表示。这种设计使得MRCM能够自适应地捕

获不同尺度的结节形态特征,尤其对微小结节的识别能力有显著提升,模块结构如图2所示。



3个级联的 3×3 卷积层其等效感受野分别为 3×3 、 5×5 、 7×7 ;Concat表示通道拼接操作;相加融合表示逐元素相加

图2 多尺度残差连接模块结构

Figure 2 Structure of multi-scale residual connection module

1.2 全局多特征融合模块

在原始的U-Net中,特征信息从浅层到深层传递时,会失去一些全局的上下文信息,而跳跃连接虽然能够结合本层的语义信息,但有时也可能会引入一些无关的噪声。为解决传统跳跃连接在传递细节时易引入背景噪声,以及原始GPG模块采用空洞卷积导致信息不连续性的问题,本文提出了一种改进的全局金字塔引导模块^[23],以实现高效且去噪的全局多特征融合。

下面以第3阶段的IGPG模块为例说明其结构,如图3所示。

首先,将第3阶段(当前阶段)及更深阶段(第4、第5阶段)的特征图分别经过 3×3 卷积调整通道数至相同维度。其次,将第4阶段和第5阶段的特征图用双线性插值法上采样至与第3阶段相同的空间尺寸。最后,将各阶段处理后的特征图在通道维度拼接,并创新性地通过本文提出的MRCM提取多尺度特征,得到最终的融合特征图。

IGPG的核心创新在于摒弃了原始GPG模块中采用的空洞卷积。原始GPG模块为了从不同层次的特征图中提取全局上下文信息,采用了多个并行的空洞卷积层。然而,空洞卷积会导致卷积核的不连续性,使得卷积操作在像素间产生跳跃和不连续的覆盖,进而引发信息的连续性损失,特别是在处理小目标物体时,这种损失可能会更加明显。本文通过用MRCM替代空洞卷积,成功避免了上述问

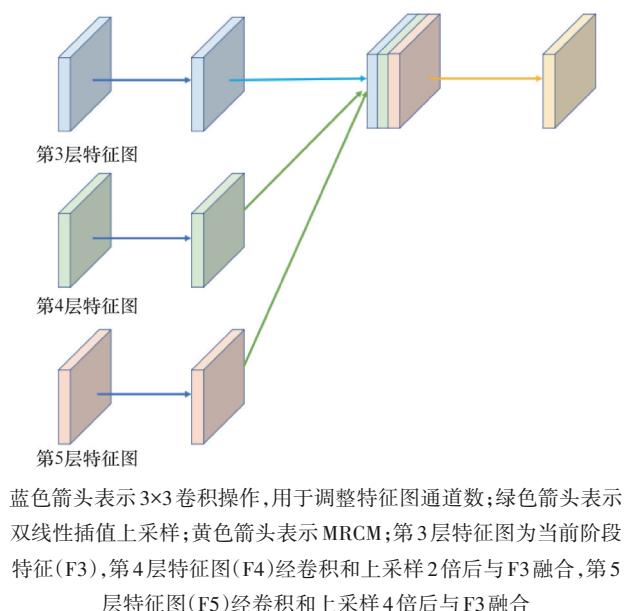


图3 第3阶段的GPG模块示意图

Figure 3 Schematic diagram of the GPG module of the third stage

题。MRCM通过其密集的小卷积核级联,保持了感受野的连续扩展,确保了特征提取的完整性,从而在融合多阶段特征时,能够更有效地抑制噪声干扰,并以深层语义信息作为空间注意力向导,筛选并增强判别性特征的表达能力。这种设计使得IGPG能够将高级的特征信息引导至不同阶段,同时有效过滤掉无关的噪声,进一步增强了模型的分割性能以及泛化性能。

不同阶段的GPG模块可按式表示:

$$G_m = \text{MRCM} \left(\text{Concat}_{n=m}^5 \left[\text{Up}_{2^{n-m}} \left(\text{Conv}_{3 \times 3} (F_n) \right) \right] \right)$$

式中: G_m 为第 m 阶段GPG模块输出的融合特征; F_n 为编码器第 n 阶段特征($n=m, m+1, \dots, 5$), $\text{Conv}_{3 \times 3}$ 为3×3卷积(用于调整通道数); $\text{Up}_{2^{n-m}}$ 为倍率为 2^{n-m} 的双线性上采样(当 $n=m$ 时,上采样倍率为1,即不进行上采样操作);Concat为通道拼接操作;MRCM为多尺度残差连接模块。

为了节省计算资源,本文设计的网络只包含了3个GPG模块(对应编码器第2、3、4阶段)。通过在编码器和解码器之间嵌入多个GPG模块,可以将高级的特征信息引导至不同阶段,同时过滤掉无关的噪声,进一步增强了模型的分割性能以及泛化性能。

2 实验

2.1 实验环境与训练策略

实验基于Ubuntu 20.04操作系统,采用Python 3.8编程语言与PyTorch深度学习框架,硬件环境为NVIDIA Tesla V100 16GB GPU。优化器选用自适应矩估计(adaptive moment estimation, Adam)^[24],初始学习率设为0.0001,批大小为16,训练迭代200个epoch。学习率调度采用余弦退火策略(cosine annealing learning rate scheduler),实现学习率的平滑衰减。实验损失函数采用二元交叉熵损失(binary cross entropy, BCE)和戴斯相似系数(Dice)损失的加权组合,参考文献[25],设置 $\alpha=0.5, \beta=1$ 。

$$\mathcal{L} = \alpha \mathcal{L}_{\text{BCE}} + \beta \mathcal{L}_{\text{Dice}}$$

2.2 数据集与预处理

实验数据集采用两个公开甲状腺结节超声影像数据集:TN3K^[26]与DDTI^[27]。TN3K数据集包含3493张甲状腺结节超声图像。由两名具有5年以上甲状腺超声诊断经验的医师独立勾画结节轮廓,对于分歧病例引入第3名高年资医师(10年以上经验)仲裁,最终经多数投票确定金标准。数据集划分:2879张用于训练,614张用于测试。DDTI数据集包含637张甲状腺结节超声图像,由单一专家标注后经另一位专家审核修正,存在争议的标注经科室讨论确定。数据集划分:511张用于训练,126张用于测试。图4和图5中的金标准图均为上述专家手工标注流程产生的分割掩膜,作为模型训练和定量评价的真值参考。

需要说明的是,由于甲状腺超声图像固有的低对比度、斑点噪声以及结节形态的复杂性和边缘模糊性,即使是经验丰富的专家在手动勾画结节轮廓时,也可能存在一定的主观性和不确定性,导致金标准与肉眼观察的原图在形态上存在细微偏差。这种偏差是医学图像标注的常见挑战,也是本研究致力于通过算法提升分割精度和客观性的重要驱动力。

所有图像经双线性插值统一缩放至256×256像素,保持纵横比不变,不足部分采用边缘像素镜填充。数据增强策略包括:随机水平翻转(概率0.5)、随机垂直翻转(概率0.5)、随机旋转(在 $[-20^\circ, +20^\circ]$ 范围内均匀采样旋转角度,采用双线性

性插值进行图像旋转,标注图采用最近邻插值保持标签值不变)、随机亮度调整(对比度因子[0.8, 1.2])。

2.3 评价指标

为了定量评估本文所提出的模型的性能,本文采用交并比(intersection over union, IoU),豪斯多夫距离(Hausdorff distance, HD)以及 $F1$ 分数作为模型的性能评估指标,其中 95% 豪斯多夫距离(HD95)可以减轻异常值的影响,故选择 HD95 作为评价指标。3 种评价指标的定义如下所示:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

$$\text{HD}_{95}(X, Y) = P_{95}(\{d(x, Y) | x \in X\} \cup \{d(y, X) | y \in Y\})$$

式中: X 和 Y 分别为预测边界点集和真实边界点集; $d(x, Y) = \min_{y \in Y} d(x, y)$, 为点 x 到集合 Y 的最短距离; $d(y, X) = \min_{x \in X} d(y, x)$, 为点 y 到集合 X 的最短距离。为抑制异常值影响,实际采用 95% 豪斯多夫距离(HD95),即取所有距离的第 95 百分位数。

$$F1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

式中, TP(true positive, 真阳性)为实际为正类并且模型预测为正类的样本数量; FP(false positive, 假阳性)为实际为负类而模型预测为正类的样本数量; FN(false negative, 假阴性)为实际为正类而模型预测为负类的样本数量; X 和 Y 分别为真实值像素和分割像素。

2.4 实验结果与分析

为了全面评估本文所提出的算法性能,选取 5 种具有代表性的分割方法进行对比实验,包括 FCN^[8]、U-Net^[10]、MultiResUNet^[18]、SGUNet^[28]及 DC-UNet^[29],所有对比方法均采用官方开源代码,在相同实验条件下重新训练:相同数据集划分(TN3K: 2879/614; DDTI: 511/126)、相同输入尺寸(256×256)、相同数据增强策略、相同训练轮次(200 epoch)。FCN 与 U-Net 采用原始论文实现, MultiResUNet、SGUNet、DC-UNet 采用开源代码,并根据本文数据集进行参数调优至最优性能。表 1 和表 2 分别展示了各算法在 TN3K 数据集和 DDTI 数据集上的定量分割结果。本文算法在 IoU、HD95 和 $F1$ 三个指标上均取得最优表现:在 TN3K 数据集上分别为 72.21%、18.49 和 80.97%,在 DDTI 数据集上分别为 63.40%、18.61 和 74.97%。相较于 U-Net 基

准,本文算法在 TN3K 上 IoU 提升 4.0%、 $F1$ 提升 3.5%,在 DDTI 上 IoU 提升 3.0%、 $F1$ 提升 2.4%,验证了多尺度特征提取与全局特征融合策略的有效性。

表 1 不同方法在 TN3K 数据集上的分割性能对比
Table 1 Comparison of segmentation performance of different methods on TN3K dataset

方法	IoU/%	HD95	$F1$ /%
FCN	70.61	21.76	79.92
U-Net	68.22	23.34	77.48
MultiResUNet	61.37	25.76	72.28
SGUNet	70.86	22.66	80.09
DC-UNet	69.17	25.27	78.73
本文算法	72.21	18.49	80.97

表 2 不同方法在 DDTI 数据集上的分割性能对比
Table 2 Comparison of segmentation performance of different methods on DDTI dataset

方法	IoU/%	HD95	$F1$ /%
FCN	58.60	21.28	71.18
U-Net	60.46	22.54	72.60
MultiResUNet	59.12	26.61	71.59
SGUNet	58.57	21.22	71.31
DC-UNet	62.34	20.91	72.98
本文算法	63.40	18.61	74.97

表 1 显示 SGUNet 在 TN3K 上 IoU 达 70.86%(仅次于本文方法),但图 4 第 1 行显示其出现严重过分割(将非结节区域误判)。经分析,SGUNet 对低对比度图像敏感,在清晰边界案例(测试集多数)表现良好,但在图 4 第 1、2 行等模糊边界极端案例中稳定性不足。这提示仅依赖平均指标可能掩盖模型鲁棒性缺陷,本文方法在困难案例上的稳定性更优。

本文方法在 TN3K (IoU 72.21%) 与 DDTI (IoU 63.40%) 上存在 8.8% 的性能差距,主要是由于 TN3K 采集自高分辨率超声设备(平均约 600×400 像素),DDTI 分辨率较低(平均约 400×300 像素)且斑点噪声更严重;TN3K 采用双人标注+仲裁机制,金标准可靠性更高;DDTI 为单人标注+审核,部分边界模糊结节的标注存在不确定性;TN3K 微小结节(< 10 mm)占比更高,本文多尺度方法对此类目标更具

优势;DDTI大结节占比高但对比度低,边界识别困难。

甲状腺结节分割结果对比如图4所示。每行展示同一张超声影像的不同模型分割结果。

典型失效案例分析如下。第1行(单结节低对比度):FCN、U-Net、MultiResUNet、SGUNet均将周围肌肉组织误判为结节。根本原因在于缺乏全局上下文约束,无法区分灰度相似的解剖结构。DC-UNet虽引入注意力机制,但局部注意力在强噪声下仍易失效。第2行(双结节):对比方法出现小结节漏检或边界粘连,原因是固定感受野难以同时适应大小结节。第4行(多结节):仅本文方法完整检出所有结节,其余方法因未能有效提取多尺度特征而漏检小结节。

2.5 消融实验

为验证各模块对分割性能的贡献,设计以下消融实验:(1)Baseline,标准U-Net,无GPG模块,使用普通跳跃连接;(2)w/GPG(Original),引入原始GPG模块(采用空洞卷积);(3)w/IGPG(Ours),引入本文改进GPG模块(采用MRCM替代空洞卷积)。实验结果如表3和表4所示。

由表3和表4的分割结果可以看出,相比

表3 GPG模块在TN3K数据集上的消融实验

Table 3 Ablation study of the GPG module on the TN3K dataset

方法	IoU/%	HD95	F1/%
无GPG模块	68.22	21.15	78.97
原始GPG模块	70.13	20.32	79.48
本文GPG模块	72.21	18.49	80.97

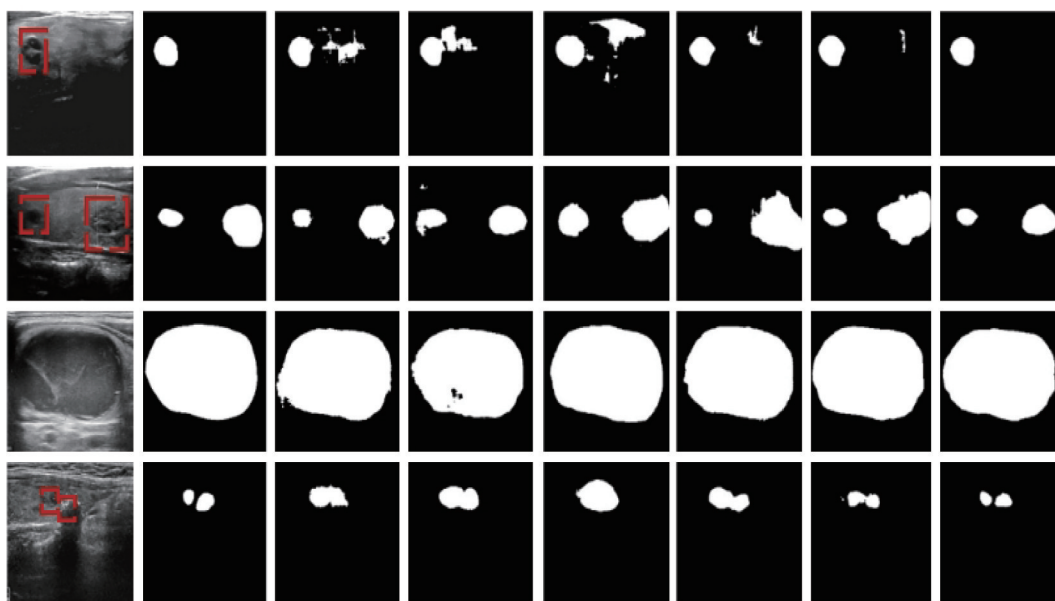
表4 GPG模块在DDTI数据集上的消融实验

Table 4 Ablation study of the GPG module on the DDTI dataset

方法	IoU/%	HD95	F1/%
无GPG模块	60.46	20.74	72.35
原始GPG模块	61.95	20.73	72.79
本文GPG模块	63.40	18.61	74.97

Baseline,添加GPG模块后IoU在TN3K上从68.22%提升至72.21%(+3.99%),在DDTI上从60.46%提升至63.40%(+2.94%),F1分数相应提升,HD95显著降低,证明全局多特征融合对分割性能至关重要。

本文IGPG模块相比原始GPG模块,在TN3K上IoU提升2.08%(72.21% vs 70.13%),在DDTI上提

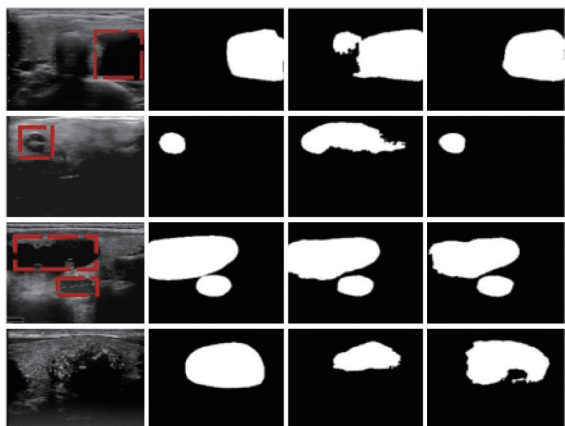


第1列为原图(图中红色虚线框标记为甲状腺结节区域),第2列为金标准(专家手工标注),第3~8列分别为FCN、U-Net、MultiResUNet、SGUNet、DC-UNet及本文算法结果

图4 不同网络的甲状腺结节分割结果对比

Figure 4 Comparison of thyroid nodules segmentation results in different networks

升 1.45% (63.40% vs 61.95%)。这验证了多尺度残差连接相比空洞卷积在超声图像上的优势——空洞卷积的核不连续性导致小目标信息损失,而 MRCM 通过密集小卷积核保持感受野连续扩展,避免信息断层。消融实验分割结果对比如图 5 所示,每行展示同一张影像的不同配置结果。



第 1 列为原图(图中红色虚线框标记为甲状腺结节区域),第 2 列为金标准,第 3 列为无 GPG 模块(Baseline),第 4 列为本文 IGPG 模块

图 5 消融实验分割结果对比

Figure 5 Comparison of the segmentation results of the ablation study

无 GPG 模块时,模型在低质量图像上出现明显过分割(第 1 行)与边界模糊(第 2 行);引入 IGPG 后,通过跨阶段特征融合有效过滤噪声,分割精度显著提升。图 5 第 3 行展示多结节案例,无 GPG 模块漏检小结节,而 IGPG 模块完整检出,验证了多尺度特征融合的有效性。MRCM 与 IGPG 联合使用产生显著互补效应(TN3K: +3.99% IoU; DDTI: +2.94% IoU),远超单独模块提升之和,表明多尺度特征提取与全局特征融合存在协同优化机制。

3 讨论

本文提出的基于多尺度特征提取与全局多特征融合的甲状腺结节超声影像分割方法,旨在系统性解决超声图像中普遍存在的尺度变异性和噪声干扰两大核心挑战。

针对传统 U-Net 固定感受野难以捕获多尺度结节特征的问题, MRCM 通过级联 3×3 卷积核实现感受野的连续扩展和特征分层精炼。与 MultiResUNet 等并行策略不同, MRCM 的级联结构在降低参数量

的同时,确保了从精细边缘到抽象语义特征的全面捕获。实验证明, MRCM 显著提升了模型对微小结节的识别能力,在 TN3K 数据集上 IoU 提升 4.0%,有效解决了多尺度特征表征不足的难题。为解决 U-Net 跳跃连接易引入背景噪声的问题, IGPG 引入跨层特征筛选机制,以深层语义信息作为“注意力向导”抑制无关特征响应。 IGPG 利用 MRCM 进行跨层特征融合,更适合超声图像的细粒度分割。消融实验显示,在 DDTI 数据集上, HD95 从 20.73 像素显著降至 18.61 像素,表明其在噪声抑制和边界定位精度方面表现卓越,有效提升了模型的鲁棒性。 MRCM 和 IGPG 协同作用,实现了全局上下文与局部细节的有效平衡,避免了纯 Transformer 模型可能损失关键边界细节,也克服了纯卷积操作感受野受限的局限。

相较于 FCN 和 U-Net 等经典方法,本文通过显式的 MRCM 和 IGPG 设计,有效解决了其固定感受野的局限性,在处理不同尺寸结节和复杂背景方面展现出更优异的性能。例如, SGUNet 在低对比度图像中易出现严重过分割,凸显了其对于噪声的敏感性和对局部信息或单一先验的过度依赖。此外,与曹玉珍等^[14]提出的 Transformer-CNN 双分支网络相比,本文 IGPG 模块以轻量化的跨层注意力机制实现了类似全局上下文建模的功能,在保持优越性能的同时,显著降低了模型复杂度,更适合临床实时应用。

尽管本文方法在超声图像特有的尺度变异和噪声干扰方面表现出显著优势,但基于 Transformer 的分割网络在医学图像领域展现出的强大潜力^[30],其强大的全局建模能力在某些任务上已展现出 SOTA 性能。纯 Transformer 模型虽擅长全局建模,但在甲状腺超声结节分割这类细粒度任务中,可能面临计算资源需求高、训练数据依赖大以及局部细节捕捉不足的挑战。本文当前阶段聚焦于卷积神经网络架构的深度优化,通过 MRCM 的级联小卷积核设计有效保留了精细局部纹理和边缘信息,实现了全局与局部信息的有效平衡,更符合临床实时应用需求。

为此,仍对 Transformer 的强大潜力保持高度关注。未来的研究将着重于探索如何将本文提出的高效多尺度特征提取和去噪全局特征融合机制与最新的 Transformer 架构进行有策略的深度融合。

例如,研究将MRCM作为Transformer编码器中的局部特征提取单元,以增强其对局部纹理和边缘的感知能力;或将IGPG作为Transformer解码器中的跨层信息聚合器,在利用Transformer全局建模优势的同时,有效过滤噪声并精炼特征。这种混合架构有望在结合两者优势的同时,进一步提升模型在复杂超声图像中的分割精度和泛化能力,从而为构建更可靠的甲状腺结节计算机辅助诊断系统提供新的技术方案。

此外,本研究仍存在一些值得深入探讨的局限性。在模型复杂度与部署优化方面,本文网络包含3个IGPG模块,相较于原始U-Net,其参数量和计算量有所增加,这在一定程度上限制了其在边缘计算设备上的高效部署。未来将探索模型剪枝、量化或知识蒸馏等技术以优化模型结构,降低计算成本的同时保持高性能。对于极端微小结节(如直径小于2 mm)的自适应能力仍有提升空间,未来的研究可引入更精细的多尺度建模策略,例如结合超分辨率技术或更具针对性的微小目标检测机制,以进一步提升对这类挑战性目标的识别精度。泛化性能需在大规模、多中心、多设备的临床数据集上进行更广泛的验证。同时,数据标准化与采集协议差异对模型性能的影响也需系统分析,并探索开发能够适应不同采集协议的自适应分割模型,以提高模型的跨设备迁移能力。

4 结论

本文针对甲状腺结节超声影像分割中普遍存在的尺度变异性和噪声干扰两大挑战,提出了一种创新的多尺度特征提取与全局多特征融合分割网络。核心贡献在于引入了多尺度残差连接模块和改进的全局金字塔引导模块。MRCM通过级联小卷积核实现了感受野的连续扩展,有效捕获了不同尺寸结节的精细特征;IGPG则通过跨层特征融合机制,巧妙地抑制了背景噪声并增强了全局上下文建模能力。在TN3K和DDTI两个公开数据集上的大量实验结果表明,本文方法在交并比(IoU)、95%豪斯多夫距离(HD95)和F1分数等关键指标上均显著优于FCN、U-Net、MultiResUNet、SGUNet和DC-UNet等主流对比方法。本研究为构建高可靠性的甲状腺结节良恶性计算机辅助诊断系统提供了新的技

术方案和坚实的理论依据,具有重要的临床应用价值。

参考文献

- [1] 中华医学会内分泌学分会,中华医学会外科学分会甲状腺及代谢外科学组,中国抗癌协会头颈肿瘤专业委员会,等.甲状腺结节和分化型甲状腺癌诊治指南(第二版)[J].中华内分泌代谢杂志,2023,39(3):181-226.
- [2] 中华医学会超声医学分会浅表器官和血管学组,中国甲状腺与乳腺超声人工智能联盟.2020甲状腺结节超声恶性危险分层中国指南:C-TIRADS[J].中华超声影像学杂志,2021,30(3):185-200.
- [3] Chen WQ, Zheng RS, Zhang SW, et al. Cancer incidence and mortality in China in 2013: an analysis based on urbanization level [J]. Chinese Journal of Cancer Research, 2017, 29 (1): 1-10.
- [4] Chen JY, You HJ, Li K. A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images [J]. Computer Methods and Programs in Biomedicine, 2020, 185: 105329.
- [5] Tao Z, Dang H, Shi YT, et al. Local and context-attention adaptive LCA-Net for thyroid nodule segmentation in ultrasound images [J]. Sensors, 2022, 22(16): 5984.
- [6] Durante C, Grani G, Lamartina L, et al. The diagnosis and management of thyroid nodules: a review [J]. JAMA, 2018, 319 (9): 914-924.
- [7] Chi JN, Li ZL, Sun ZY, et al. Hybrid transformer UNet for thyroid segmentation from ultrasound scans [J]. Computers in Biology and Medicine, 2022, 153: 106453.
- [8] Song PF, Li JJ, Fan H, et al. TGDAUNet: Transformer and GCNN based dual-branch attention UNet for medical image segmentation [J]. Computers in Biology and Medicine, 2023, 167: 107583.
- [9] Lafraxo S, El Ansari M, Koutti L, et al. SEDARU-net: a squeeze-excitation dilated based residual U-Net with attention mechanism for automatic melanoma lesion segmentation [J]. Multimedia Tools and Applications, 2025, 84 (21): 23935-23962.
- [10] Haribabu K, Prasath R, Ir P. MLRT-UNet: An Efficient Multi-Level Relation Transformer Based U-Net for Thyroid Nodule Segmentation [J]. Computer Modeling in Engineering & Sciences, 2025, 143(1): 413-448.
- [11] Liu Z, Lin YT, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//IEEE/CVF (Eds.). Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2021: 10012-10022.

- [12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]// 31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, CA, USA: Curran Associates, 2017: 6000–6010.
- [13] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [J]. arXiv:2010.11929.
- [14] 曹玉珍, 郑洁, 余辉, 等. 基于超声影像的甲状腺结节智能分割算法[J]. 天津大学学报, 2022, 55(7): 674–681.
Cao YZ, Zheng J, Yu H, Wang F, Zhang J. Intelligent segmentation of thyroid nodules based on ultrasound images [J]. Journal of Tianjin University (Science and Technology), 2022, 55(7): 674–681.
- [15] Liu J, Peng YL, Pu MS, et al. KTUNet: A hybrid CNN-Transformer with kolmogorov-arnold convolution for thyroid nodule segmentation in ultrasound images [J]. Digital Medicine, 2025, 11(2): e24–00019.
- [16] Bi H, Cai CJ, Sun JW, et al. BPAT-UNet: Boundary preserving assembled transformer UNet for ultrasound thyroid nodule segmentation [J]. Computer Methods and Programs in Biomedicine, 2023, 238: 107614.
- [17] Deng XY, Dang ZY, Pan LH. STU3Net: An improved U-net with swin transformer fusion for thyroid nodule segmentation [J]. International Journal of Imaging Systems and Technology, 2024, 34(5): 23160.
- [18] Ibtehaz N, Rahman MS. MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation [J]. Neural Networks, 2020, 121: 74–87.
- [19] Zhu SQ, Yu CR, Hu JJ. Regularizing deep neural networks for medical image analysis with augmented batch normalization [J]. Applied Soft Computing, 2024, 154: 111337.
- [20] Wang HC, Fu TF, Du YQ, et al. Scientific discovery in the age of artificial intelligence [J]. Nature, 2023, 620: 47–60.
- [21] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]//IEEE eds. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 2818–2826.
- [22] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]//AAAI eds. Proceedings of the AAAI Conference on Artificial Intelligence. San Francisco, CA, USA: AAAI Press, 2017.
- [23] Feng SL, Zhao HM, Shi F, et al. CPFNet: Context pyramid fusion network for medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2020, 39(10): 3008–3018.
- [24] Chen XN, Liang C, Huang D, et al. Symbolic discovery of optimization algorithms [J]. Advances in Neural Information Processing Systems, 2023, 36: 49205–49233.
- [25] Valanarasu JMJ, Patel VM. Unext: Mlp-based rapid medical image segmentation network [C]//MICCAI eds. International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature, 2022: 23–33.
- [26] Gong HF, Chen GQ, Wang RR, et al. Multi-task learning for thyroid nodule segmentation with thyroid region prior [C]//IEEE eds. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). Nice, France: IEEE, 2021: 257–261.
- [27] Pedraza L, Vargas C, Narváez F, et al. An open access thyroid ultrasound image database [C]//SPIE eds. 10th International Symposium on Medical Information Processing and Analysis. Bogota, Colombia: SPIE, 2015: 188–193.
- [28] Pan HT, Zhou Q, Latecki LJ. Sgunet: Semantic guided unet for thyroid nodule segmentation [C]//IEEE eds. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). Nice, France: IEEE, 2021: 630–634.
- [29] Lou AG, Guan SY, Loew M. DC-UNet: rethinking the U-Net architecture with dual channel efficient CNN for medical image segmentation [C]//SPIE eds. Medical Imaging 2021: Image Processing San Diego, CA, USA: SPIE, 2021: 758–768.
- [30] Cao H, Wang YY, Chen J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation [C]//ECCV eds. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 205–218.

(2026-01-05 收稿, 2026-03-25 修回)