

引用格式: 赵欣, 黎红豆, 王洪凯. 融合多尺度特征和注意力机制的超声甲状腺结节分割[J]. 声学技术, 2024, 43(5): 668-676. [ZHAO Xin, LI Hongdou, WANG Hongkai. Ultrasound image segmentation of thyroid nodule combining multi-scale features and attention mechanism[J]. Technical Acoustics, 2024, 43(5): 668-676.] DOI: 10.16300/j.cnki.1000-3630.2024.05.009

融合多尺度特征和注意力机制的 超声甲状腺结节分割

赵 欣¹, 黎红豆¹, 王洪凯²

(1. 大连大学信息工程学院, 辽宁大连 116622; 2. 大连理工大学生物医学工程学院, 辽宁大连 116024)

摘要: 针对目前超声影像下甲状腺结节分割不够精准的问题, 提出一种融合多尺度特征和注意力机制的超声甲状腺结节分割方法。该模型编码设计了多感受野通道选择模块, 通过核心选择注意力对多个不同感受野的特征进行自适应加权组合, 使包含目标的感受野通道占据主导。同时, 设计自适应全局上下文模块自适应地提取瓶颈层多个尺度的全局上下文特征, 以实现瓶颈层高级语义的有效编码。此外, 设计双注意力引导模块增强编解码器对等层之间的特征融合, 以减少上采样过程中的信息损失。在公开的超声甲状腺结节数据集上进行实验, 结果表明, 文中所提方法优于其他对比网络, 能更加精准地分割出甲状腺结节, 有效提升了甲状腺结节的分割性能。

关键词: 深度学习; 甲状腺结节; 超声图像分割; 多尺度特征提取; 注意力机制

中图分类号: TP391

文献标志码: A

文章编号: 1000-3630(2024)-05-0668-09

Ultrasound image segmentation of thyroid nodule combining multi-scale features and attention mechanism

ZHAO Xin¹, LI Hongdou¹, WANG Hongkai²

(1. School of Information Engineering, Dalian University, Dalian 116622, Liaoning, China;

2. School of Biomedical Engineering, Dalian University of Technology, Dalian 116604, Liaoning, China)

Abstract: A new ultrasound thyroid nodule segmentation model with multi-scale features and attention mechanisms is proposed to address the problem of inaccurate segmentation of thyroid nodules in ultrasound images. In the model, a multi receptive field channel selection module designed in the encoding phase is used to adaptively weight and combine features from different receptive fields with a focus on channels containing the target object. Additionally, an adaptive global context module is designed to extract global context features from multiple scales of the bottleneck layer to effectively encode high-level semantic information. Furthermore, a dual attention-guided module is used to enhance feature fusion between peer layers of the encoder and decoder for reducing information loss during upsampling. Experimental results on a publicly available ultrasound thyroid nodule dataset show that the proposed method outperforms other state-of-the-art networks and achieves more accurate segmentation of thyroid nodules, effectively improving segmentation performance.

Key words: deep learning; thyroid nodule; ultrasound image segmentation; multi-scale feature extraction; attention mechanism

0 引 言

近年来甲状腺癌的发病率逐年上升, 已经成为威胁人类健康的十大癌症之一。甲状腺结节是甲状腺内的肿块, 有一定概率会恶化成甲状腺癌, 因此

甲状腺结节的早期发现与诊断是降低甲状腺癌发病率和死亡率的关键环节。超声是筛查甲状腺结节的常用检测手段, 具有实时成像、检测方便、无创等优点。然而, 超声成像具有分辨率低、易受噪声干扰、伪影较多等特性, 加之医生个人经验的局限和长时间工作产生的视觉疲劳, 导致临床检查时甲状腺结节区域不易精确判别, 容易产生漏诊和误诊。因此开发出一种准确快速的分割方法帮助医生对结节进行良恶性评估是十分必要的。首先, 甲状腺结节与周围组织特征相似, 易出现将背景区域误认为是病灶区域的现象。其次, 甲状腺结节形状不规

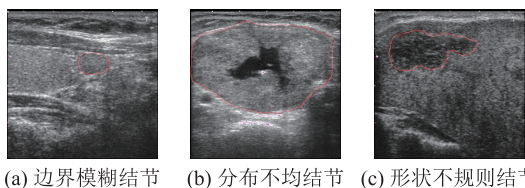
收稿日期: 2023-02-16; 修回日期: 2023-03-03

基金项目: 国家自然科学基金(81971693)。

作者简介: 赵欣(1974—), 女, 辽宁锦州人, 博士, 副教授, 研究方向为人工智能, 医学图像处理。

通信作者: 赵欣, E-mail: zhaoxin@dlu.edu.cn

则,边界不清,导致对结节边缘轮廓的分割不精准。甲状腺结节超声图像如图1所示。因此,从超声图像中准确分割出甲状腺结节存在一定困难。



(a) 边界模糊结节 (b) 分布不均结节 (c) 形状不规则结节

图1 甲状腺结节超声图像

Fig.1 Ultrasound image of thyroid nodules

基于人工智能技术的超声甲状腺结节自动分割和识别技术可为临床检测甲状腺结节提供有效的辅助手段,对提升甲状腺结节的诊断效率和准确率,降低医生工作强度具有积极意义。早期的超声甲状腺结节自动分割方法主要采用传统的机器学习方法。例如,Iakovidis等^[1]提出一种新的甲状腺超声图像分割框架,该方法利用可变背景区域消除图像强度不均匀性并加入遗传算法的参数调整机制自动搜索最佳参数。Kollorz等^[2]使用独立于结节类型的功率分水岭方法对甲状腺结节进行半自动分割。郑斌等^[3]提出结合形态学闭运算的超声图像甲状腺结节分割方法。邵蒙恩等^[4]建立CV-RSF模型,该模型将基于主动轮廓分割模型(chan-veese, CV)的全局信息产生的粗分割结果作为区域可变灰度拟合(region scalable fitting, RSF)模型的初始轮廓,降低了初始轮廓的敏感性,提高了分割的准确率。

传统的分割方法过于依赖人为的预定义特性,分割精度有待提高。随着深度学习的出现及其在图像处理领域的发展,越来越多的学者将深度学习方法用于图像分割中。近两年,基于深度学习的超声甲状腺结节分割方法主要通过改进U-Net网络来提高分割精度。一些学者通过延长和拓展网络结构来提升特征提取能力。例如,2021年,Ouahabi等^[5]提出了一种全卷积密集扩张网络来提高甲状腺结节分割效率该网络采用扩张卷积扩大感受野,使用密集连接实现多尺度特征的融合,并且在网络中引用因子化滤波器以进一步提高模型的效率。Pan等^[6]提出一种高级语义引导的U-Net模型,该模型从解码阶段的高级特征中提取单通道的语义特征图,并以此形成对低级特征的语义引导,以获得精确的结节分割。2022年,Peng等^[7]提出使用两个解码器来增强特征表达能力,并采用金字塔池化提取多尺度特征。另有学者采用多任务学习的方式提升甲状腺结节分割效果。2022年,Kang等^[8]通过提出的任务内和任务间一致性学习实现超声图像中甲状腺结节的分割和分类,有效消除了分割和分类任务的不一

致性,从而提高了甲状腺结节分割性能。

还有学者通过数据增强改进分割效果。例如,2021年,Chu等^[9]采用U-Net网络结构,保留在数据采集时结节长轴和短轴的四个节点信息来辅助甲状腺结节分割,但是这种方法不适用于没有标记的甲状腺结节分割,具有一定的局限性。

随着注意力机制在图像分割中的普遍采用,有些研究开始关注通过注意力机制改进超声甲状腺结节的分割。例如,2022年,王波等^[10]提出了基于残差多尺度特征和注意力机制的甲状腺结节分割网络,该网络使用混合注意力解决U-Net模型中的跳跃长连接在传递关键细节特征时将无用信息或噪声同时传入解码层的问题。Nie等^[11]为超声甲状腺结节分割设计了一种密集全卷积神经网络N-Net,并在跳跃连接之前使用注意力模块降低特征提取时背景噪声的干扰。Nguyen等^[12]采用双U-Net的顺序嵌套结构来提高甲状腺结节的分割性能,其中一个是使用较少的卷积核和网络层数的小型U-Net,用于图像的粗分割,分割结果传输给另一个带有注意力的标准U-Net网络,但该方法的分割效果并不理想,且网络训练的计算代价较高。

近两年,基于Transformer的自注意力^[13]在深度学习领域受到广泛关注,在超声图像分割中也被尝试使用。例如,2023年,Chi等^[14]将Transformer用于分割超声图像中甲状腺部位,提出了一种混合Transformer U-Net,能以较高的准确率分割出超声图像中甲状腺所在区域。而Li等^[15]将Transformer用于超声甲状腺结节分割,提出了一种BTNet模型,采用边界注意力和Transformer来增强网络对目标边界的学习能力和网络的整体学习能力。虽然在模型中融入Transformer可提升性能,但模型的计算成本和训练成本较高,同时需要大量的训练数据才能达到较好的分割效果。

上述方法为甲状腺结节分割研究提供了有意义的参考,但现有研究工作的甲状腺结节分割精度仍有待提高,特别是对形状不规则、边界不清晰的结节,实现其精准分割的难度仍然较大。为此,本文提出一种融合多尺度和注意力的超声甲状腺结节分割网络。本文的主要研究内容如下:

(1)将注意力引入到卷积神经网络(convolutional neural network, CNN)的多感受野特征提取中,设计了多感受野通道选择模块,用于取代传统U-Net架构中编码层的单一尺度卷积,使不同大小的目标不再只是通过单一尺度的感受野提取特征,而是通过多个不同大小感受野的特征通道的自适应加权组合提取特征。

(2)为了对瓶颈层高级语义深入解读,本文设计了一种采用注意力机制的金字塔池化多尺度特征学习模块,也称自适应全局上下文模块。模块采用通道注意力机制对融合后的多尺度特征通道赋予不同权重,以此来构建全局多尺度特征各通道之间的线性关系,从而有效提取瓶颈层的多尺度特征。

(3)为了缓解跳跃连接两端不同深度特征间的语义差异,本文设计了双注意力跨层引导模块,通过注意力机制抑制浅层特征中的背景噪声,聚焦目标区域,从而使浅层特征和深层特征更好地融合。

1 本文方法

本文以传统U-Net网络为基础架构,设计了多感受野通道选择模块(multi receptive field channel selection module, MRFCSM)、自适应全局上下文模块(adaptive global context module, AGCM)和双注意力跨层引导模块(dual attention cross-layer guidance module, DACGM),并将这些应用到U-Net网络中提升网络的整体性能。模型的总体结构如图2所示。在模型的编码部分,每一编码层使用MRFCSM模块代替原有U-Net的普通卷积操作进行特征提取,以提高编码层特征的提取能力。在瓶颈层,利用AGCM模块替换原始U-Net瓶颈层的卷积

操作,使模型更好地把握全局上下文信息。在跳跃连接处采用DACGM模块将浅层特征与深层特征进行融合,目的是通过注意力抑制浅层特征中的背景噪声,以缓解深浅层信息间的语义差异,促进跳跃连接两端更好地融合。

1.1 多感受野通道选择模块

2019年,选择性内核网络(selective kernel net, SKNet)中将注意力机制应用于两个不同大小的感受野特征通道,通过训练,让不同大小的目标不再只是通过单一尺度的感受野提取特征,而是通过两个不同大小感受野的特征通道的自适应加权组合提取特征^[16]。受SKNet的启发,本文设计了多感受野通道选择模块,将注意力引入到CNN的多感受野特征提取中,可对不同大小的目标自适应地改变感受野,其结构如图3所示。本文将MRFCSM模块用于U-Net架构编码部分的每一层,用以取代原始U-Net模型编码层的单一尺度卷积。MRFCSM模块通过三个不同扩张率的扩张卷积(dilated convolution, DC)可捕获三个不同大小感受野的特征,这些特征通过像素级相加融合后通过通道注意力对不同感受野生成的特征通道分别进行量化筛选,最后,对量化后的通道特征重新进行像素级相加融合。该模块可获取多个视野下的卷积特征,并通过通道注意力

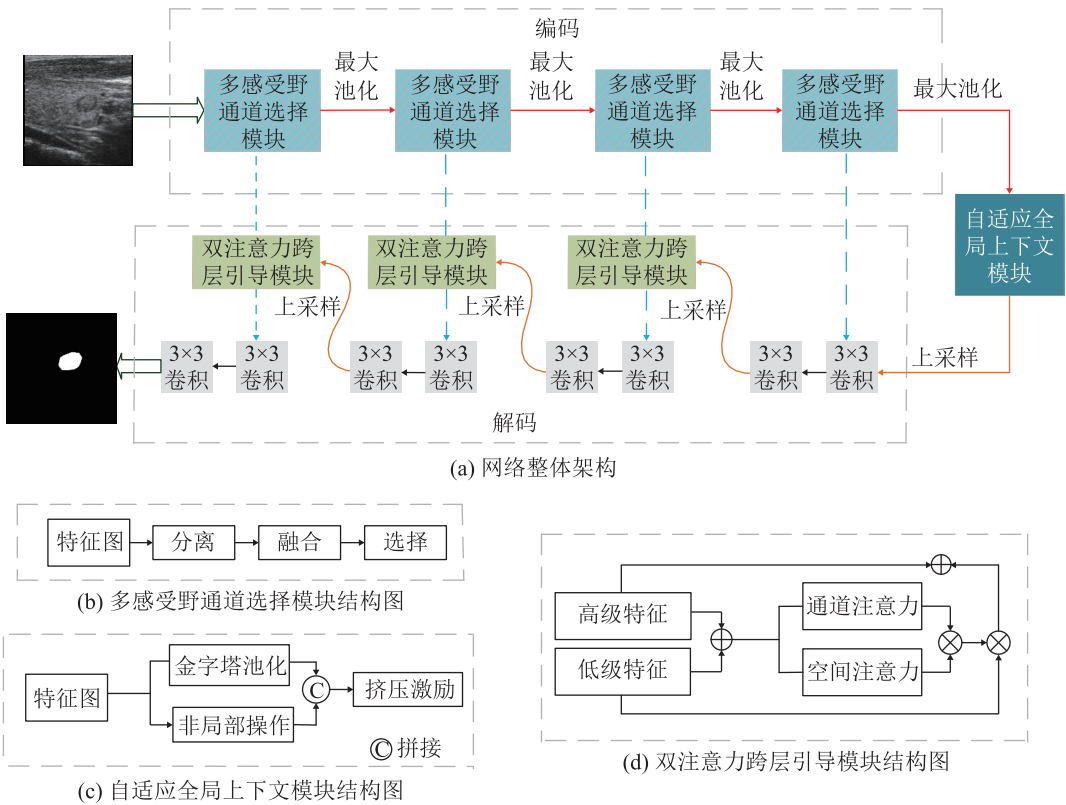


图2 网络整体架构及模块结构图

Fig.2 Diagrams of overall network architecture and module structure

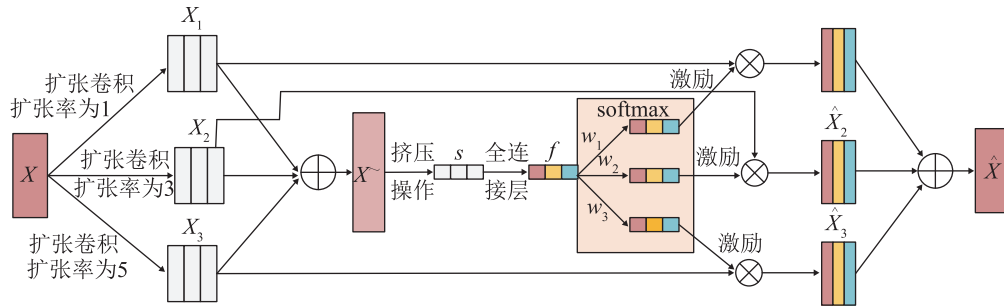


图3 多感受野通道选择模块结构图

Fig.3 Structure diagram of multi receptive field channel selection module

筛选，使包含目标特征的通道占据主导。

MRFCSM模块的具体步骤为：(1)通过扩张率为1、3、5的卷积核进行多尺度的特征融合。(2)通过挤压操作和全连接层处理后得到一个一维特征向量 f ，对该向量进行特征映射得到三个通道特征图分支，随后在三个特征图的同一位置应用softmax函数，产生针对不同卷积视野的通道注意力。(3)分别与三条卷积支路的输出加权相乘，从而对通道进行筛选和关注。MRFCSM模块通过引入不同尺寸大小的卷积核，让网络能够在不同的通道维度自适应地选择感受野，同时根据不同特征的重要程度对通道进行筛选并赋予不同的权重，提高目标的分割精度。

1.2 自适应全局上下文模块

编码阶段的深层特征属于高级语义，在进行特征提取时，相比多感受野的特征提取，对特征进行全局性的多尺度观察更加重要。金字塔池化^[17]通过并联多个具有不同池化窗口大小的自适应池化通路提取多尺度全局上下文特征，然后利用 1×1 卷积对各个通道特征进行线性融合，是处理深层特征多尺度观察的合适选择。但是，传统金字塔池化仅仅利用了线性函数，不足以刻画多尺度特征各通道之间的非线性关系。为此，本文设计了一种采用注意力机制的金字塔池化多尺度特征学习模块，被命名为自适应全局上下文模块，结构如图4所示。将其用于编码阶段的瓶颈层，目的是对编码最深层的全局上下文信息进行进一步的多尺度特征提取，以实现

对瓶颈层高级语义的深入解读。

图4中输入特征首先通过四条并行支路对特征进行多尺度观察，其中三条支路以不同感受野大小的池化操作从不同尺度观察全局上下文信息，一条支路捕捉每个位置特征的全局相关性。接着使用 1×1 卷积把通道缩减为原来的四分之一，并将特征上采样到与输入相同的空间大小。全局相关性分支采用Non-local操作^[18]，目的是获取每个特征点与距离较远的非相邻特征点间的长距离依赖关系，从而通过全局依赖映射进一步优化特征，使特征输出能统观全局信息分布。为了提高特征的利用效率，利用注意力机制增加目标区域的权重，避免背景像素的干扰，使模型更好地把握全局上下文信息。将并行各支路的输出在同一通道维度上进行拼接，然后经过挤压与激励模块^[19] (squeeze and excitation, SE)产生通道注意力，从而自适应地重新校准通道特征响应，并将生成的通道注意力作用到拼接后的输出上，以实现目标特征的进一步关注。AGCM模块通过改进的金字塔池化和Non-local操作充分聚合了上下文信息，并利用注意力机制使网络专注于目标区域，从而增强了全局特征的表达能力。

1.3 双注意力跨层引导模块

虽然高级特征包含丰富的语义信息，但缺乏必要的空间信息；而浅层特征包含丰富的空间信息，但缺乏准确的语义信息。可以看出，两者所包含的信息是互补的。然而，传统的U-Net网络只是简单

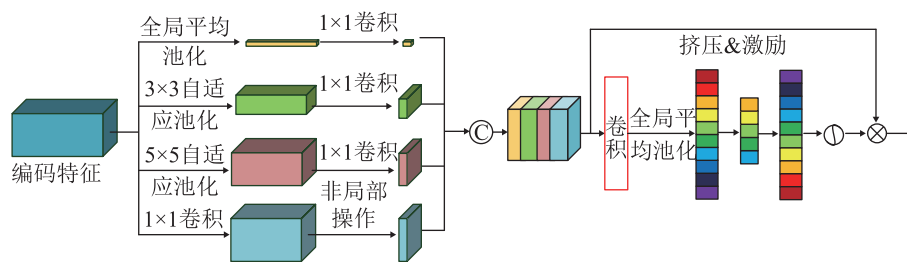


图4 自适应全局上下文模块结构图

Fig.4 Structure diagram of adaptive global context module

地将低级特征与上采样得到的高级特征拼接起来再做卷积处理，没有充分利用高级特征与低级特征的互补性对信息融合进行有效的引导，从而造成信息

冗余和一些重要特征的丢失。为此，本文设计了双注意力跨层引导模块，并将其用于编解码对等层之间的跳跃连接中，其结构如图5所示。

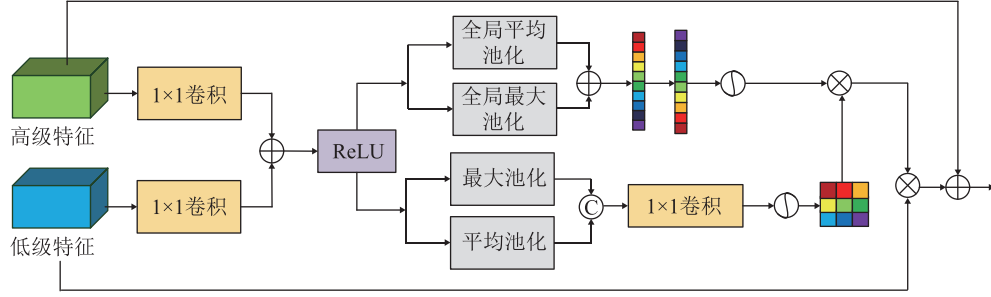


图5 双注意力跨层引导模块结构图

Fig.5 Structure diagram of dual attention cross-layer guidance module

DACGM模块同时接收来自跳跃连接两端的编码器的浅层特征和解码器的深层特征；然后对两个输入执行 1×1 卷积操作，以减少特征图的通道数；接着分别基于输出的特征生成通道注意力和空间注意力。具体操作如下：一方面对输出的特征分别进行全局平均池化和全局最大池化，将生成的两个一维向量相加后经全连接网络和sigmoid激活函数得到通道注意力。另一方面，对输出的特征分别进行通道维度上的最大池化和平均池化生成两个独立的单通道空间上下文信息，然后执行 1×1 卷积操作和sigmoid激活获得空间权重矩阵，即空间注意力图。最后，将得到的空间注意力和通道注意力相乘得到的注意系数与输入模块的浅层特征相乘，再与输入模块的深层特征相加，最终得到深浅层的融合特征。该模块的优点在于，其产生的注意力是通道和空间两个维度上的双注意力，同时充分考虑了高级特征与低级特征的互补性。因而关注度更加精确。此外，将生成的注意力作用在浅层特征上，可抑制浅层特征中不必要的背景噪声干扰，保留更重要的空间上下文信息，以缓解深层次信息间的语义差异，促进跳跃连接两端更好地融合。

1.4 损失函数

本文采用了由二元交叉熵(binary cross entropy, BCE)损失和骰子(dice)损失组成的混合损失函数。BCE损失函数通过对比前背景像素的平均加权，加速细节信息的收敛，dice损失函数由于能够减少类不平衡问题而被广泛应用于图像分割的训练当中。使用混合损失函数能够更好地分割显著性对象区域。混合损失函数定义为

$$L = L_{\text{BCE}} + L_{\text{Dice}} \quad (1)$$

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \ln(p_i) + (1-y_i) \ln(1-p_i)] \quad (2)$$

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N y_i p_i + \varepsilon}{\sum_{i=1}^N (y_i + p_i) + \varepsilon} \quad (3)$$

其中： L_{BCE} 与 L_{Dice} 分别定义为BCE损失和dice损失。 N 表示输入图像的像素总数。 $y_i \in \{0, 1\}$ 表示第 i 个像素的真实标签，其中 $y_i = 1$ 表示正类对应的像素， $y_i = 0$ 表示负类。 $p_i \in \{0, 1\}$ 表示像素 i 预测为正类的概率。为了防止分母无效，把 ε 设置为1。

2 实验结果分析与讨论

2.1 实验数据及预处理

实验使用的数据集是哥伦比亚国立大学创建的甲状腺结节分割公共数据集(digital database thyroid image, DDTI)^[20]。DDTI数据集包含299名患者的超声甲状腺图像，每一例均有原始超声图像和记录各种信息的XML文件，包括专家放射科医师对可疑甲状腺病变的完整注释和诊断描述。在进行结节分割工作之前，我们采用Labelme软件并对照XML文件的坐标信息对结节的边界进行标注。然后，我们去掉了损坏的图像和每个XML文件中未完成或标注不完整的图像，此外，我们还将含有两个结节的超声图像分成了两个样本，最终得到594张图像用于实验。实验将所有数据随机划分为5份，使用五折交叉验证，其中四倍(80%的图像)用于训练，一倍(20%的图像)用于测试，在训练期间对15%的训练图像进行验证。由于甲状腺超声图像噪声严重且存在较多标记信息，我们对图像进行了预处理。首先，去除图像周围无效的信息，保留超声图像中间区域的有效信息。其次，对所有数据进行高斯滤波和直方图均衡化处理，以去除噪声、增强图像的对比度，使图像更加清晰。

2.2 实验环境与参数设置

实验的硬件设备包括 Intel Xeon(R) CPU E5-2680 v4 @ 2.40 GHz x56 处理器, 两张 NVIDIA GeForce RTX 2080Ti GPU 显卡, GPU 显存为 11 GB。软件环境为 Ubuntu18.04.6 LTS 64 位操作系统, 开发工具为 Python3.7 和 Keras 框架。实验采用自适应矩估计(adaptive moment estimation, Adam)优化器进行训练, Adam 优化器的学习率设为 1×10^{-4} 。实验将所有图像的大小缩放为 256×256 , 训练中将批处理样本数设置为 4, 迭代次数设置为 200。训练采用早停策略, 当验证集上的损失函数在 20 个迭代周期内没有下降时终止训练。

2.3 评价指标

为了定量地衡量模型的性能, 实验采用 dice 系数 C_{DS} , 召回率(Recall) R_{REC} , 精确度(Precision) P 和准确率(Accuracy) R_{AC} 作为分割甲状腺结节的评价指标。 C_{DS} 反映分割结果与真实标签之间的重叠相似度, 取值在 0~1 之间, 且越接近 1, 表示分割效果越好。召回率又称为查全率, 表示甲状腺结节被正确分割出的比例, 用来衡量甲状腺结节分割的完整性。精确度又称为查准率, 表示分割的目标中实际为甲状腺结节的比例, 用来衡量病灶分割的精确度。准确率是对模型分割结果的全面评价, 具体公式可表示为

$$C_{DS} = \frac{2P_T}{N_F + P_F + 2P_T} \quad (4)$$

$$R_{REC} = \frac{P_T}{P_F + N_F} \quad (5)$$

$$P = \frac{P_T}{P_T + P_F} \quad (6)$$

$$R_{AC} = \frac{P_T + N_T}{P_T + P_F + N_T + N_F} \quad (7)$$

其中: P_T 为真阳性, 表示被正确分割的甲状腺结节像素数; P_F 为假阳性, 表示被错误分割为甲状腺结节像素数; N_T 为真阴性, 表示被正确分割的背景像素的数量; N_F 为假阴性, 表示被错误标记为背景像素的结节像素。

2.4 实验结果及分析

2.4.1 模块可行性验证

为了验证本文设计模块的合理性及有效性, 设计了 4 个消融对比实验, 以 U-Net 网络为基础网络, 依次叠加 MRFCSCM 模块、DACGM 模块和 AGCM 模块。所有实验均使用相同的参数, 实验结果如表 1 所示。

表 1 模块可行性实验验证结果

Table 1 Verification results of module feasibility experiment

模型	dice 系数	召回率	精确度	准确率
U-Net	0.652	0.685	0.691	0.853
U-Net+MRFCSCM	0.697	0.713	0.776	0.865
U-Net+MRFCSCM+DACGM	0.734	0.757	0.746	0.870
U-Net+MRFCSCM+DACGM+AGCM	0.768	0.853	0.806	0.877

从表 1 中可以看出, 在添加 MRFCSCM 模块后, 分割结果的评价指标相较原始 U-Net 的结果均有提高, dice 系数、召回率、精确度和准确率分别提高了 4.5 个百分点、2.8 个百分点、8.5 个百分点和 1.2 个百分点。这说明用 MRFCSCM 模块替换 U-Net 的标准卷积是有效的。在此基础上继续添加 DACGM 模块, 虽然精确度有所下降, 但其他指标均进一步提高。dice 系数、召回率和准确率在添加 MRFCSCM 模块的基础上分别提高 3.7 个百分点、4.4 个百分点和 0.5 个百分点。其中召回率提高显著, 说明 DACGM 模块通过使用注意力机制引导浅层特征和深层特征融合, 增强了对深层特征的关注, 抑制了浅层特征无关背景噪声, 对病灶的定位更加准确, 找出了更多的病变组织。最后, 在添加前两个模块的基础上添加 AGCM 模块, 模型的分割结果达到最优, 说明 AGCM 模块通过捕捉全局的依赖关系和上下文信息, 有效提升了网络的整体性能。通过消融实验, 可以看到本文设计的每一模块都能提升网络性能, 且将这些模块组合在一起效果最佳。

2.4.2 与经典分割模型的对比实验

为了验证本文所提方法的性能, 本文基于 DDTI 公共数据集, 分别在 U-Net 网络^[21]、U-Net++ 网络^[22]、ResUNet 网络^[23]和 AttentionUNet 网络^[24]四种经典模型上实现超声甲状腺结节分割, 并将分割结果与本文所提方法进行对比, 结果如表 2 所示。从表 2 中可以看出, 本文方法在 dice 系数、召回率、精确度和准确率上均优于上述各网络效果, 说明本文提出的改进模型在性能上是优于经典模型的。

表 2 本文模型与其他分割模型的评价指标对比

Table 2 Comparison of evaluation indicators between the model in this paper and other segmentation models

模型	dice 系数	召回率	精确度	准确率
U-Net	0.652	0.685	0.691	0.853
U-Net++	0.666	0.707	0.722	0.862
ResUNet	0.655	0.685	0.690	0.854
AttentionUNet	0.686	0.689	0.739	0.866
本文方法	0.768	0.853	0.806	0.877

图6是本文方法与U-Net、U-Net++、ResUNet以及AttentionUNet模型在超声甲状腺结节分割上的视觉效果对比图。从图6中可以看出，对于对比度低、边缘模糊的结节图像，如图6(a)第1行中的图像，本文的分割结果在形态上最接近真实标签，而U-Net++存在过分割现象，U-Net和ResUNet均存在漏分割情况。AttentionUNet虽然分割出了结节，但对边缘轮廓的分割不如本文方法精确。对于轮廓比较清晰的结节图像，如图6(a)第2、3行中的图像，虽然其他方法基本分割出结节所在区域，但本文方法的分割结果更为精确。对于结节与周围背景组织相近的超声图像，如图6(a)第4行中的图像，本文方法可以准确定位结节的位置，而U-Net和U-Net++均存在将背景区域误认为是目标区域的情况。这说明，本文方法由于引入了多感受野通道选择模块、自适应全局上下文模块和跨层双注意力引导模块，使网络能充分提取结节特征并更好地区分前背景信息，对于结节边界定位更加准确。综上所述，对比其他网络，本文方法的分割结果更加接近真实标签，并在边缘细节方面优于其他方法，能较好地解决由于超声图像对比度低和结节边界模糊造成甲状腺结节分割不准确的问题，可以更准确地分割出甲状腺结节。

2.4.3 模型的收敛性及复杂度分析

图7展示了本文方法与U-Net、U-Net++、ResUNet以及AttentionUNet的训练损失函数变化曲线

图。横轴为网络训练迭代次数，纵轴为训练集损失函数值。从图中可以看到，U-Net和U-Net++的收敛速度缓慢，迭代次数多，ResUNet和AttentionUNet的收敛速度较慢，迭代次数较多，而本文模型的损失曲线下下降速度快，在第40次左右趋于平稳，在第76次左右结束训练。这说明在实验一致的情况下，本文模型的收敛速度最快，收敛后达到的损失函数值最低，因此，具有良好的收敛性。

本文评估了不同网络的参数量(params)、计算量(FLOPs)、训练时间(training times)三个参数，params表示模型中需要训练的参数总数，FLOPs指浮点运算次数，用来衡量模型的复杂度，training times表示模型完成训练的时间，不同网络的复杂度对比如表3所示。从表3中可以看出，本文模型的参数量和计算量较高，但完成训练的总体时间是最短的。由于本文模型结构中添加了三个新的模块，使得参数量增加，同时也增加了模型训练中对计算资源的消耗。但是本文模型的收敛速度快、训练轮次少，这说明网络设计和各项设定比较合理，收敛速度快这一点在图7中也得到了验证。

2.4.4 在同一数据集上不同算法的对比实验

此外，本文还与使用同一甲状腺结节公共数据集DDTI的文献的分割结果进行对比，这些方法的指标如表4所示。SGUNet^[6]在每个解码步骤中进一步从高级特征中提取单通道像素语义图，用作对低级特征的高级语义指导，但它在分割结节细节上并

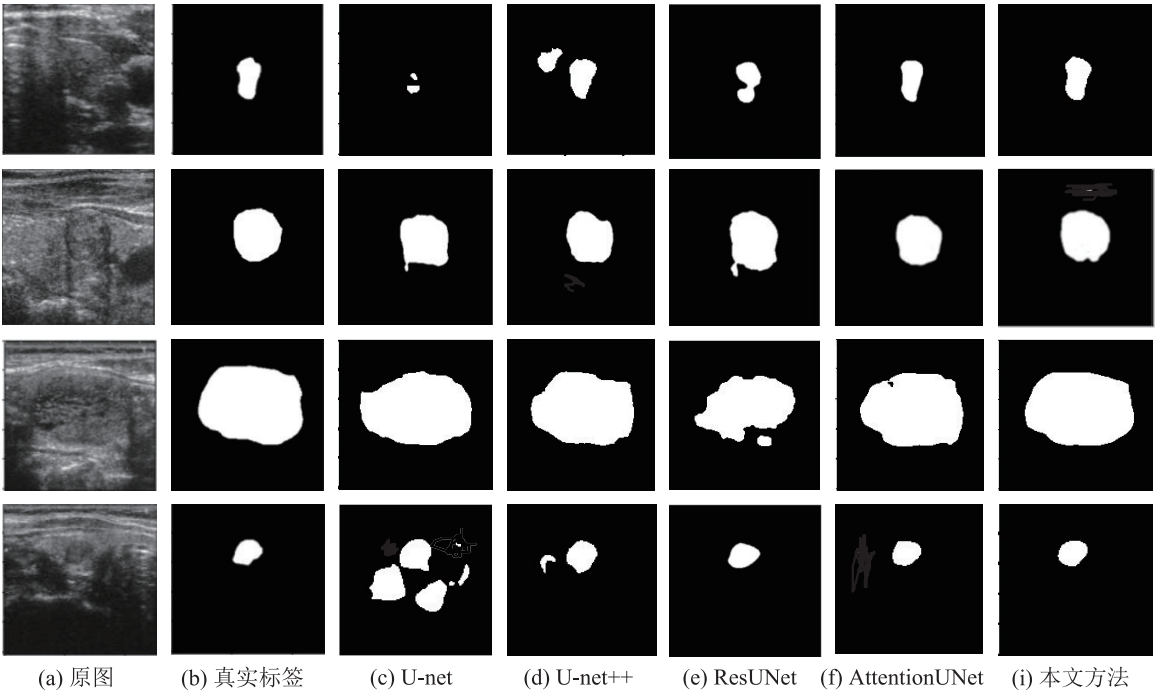


图6 甲状腺结节分割结果对比图
Fig.6 Comparison of segmentation results of thyroid nodules

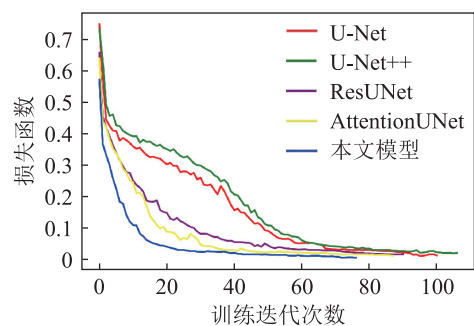


图 7 不同模型的训练损失曲线
Fig.7 Training loss curves for different models

表 3 不同模型的复杂度对比 Table 3 Complexity comparison of different models			
模型	参数量	计算量	训练时间/s
U-Net	6.75×10^6	21.5×10^9	1 052
U-Net++	9.04×10^6	59.6×10^9	1 349
ResUNet	4.72×10^6	19.8×10^9	1 012
AttentionUNet	8.90×10^6	34.4×10^9	1 066
本文方法	9.73×10^6	71.9×10^9	905

不精准，dice 系数、召回率和精确度三项指标均低于本文方法。MShNet^[7]使用两个解码器来获得更强的特征表达能力，但其各项指标均低于本文方法所得指标。Nguyen 等^[12]使用双 U-Net 网络的顺序嵌套结构来提高甲状腺结节的分割性能，但其分割效果不佳，dice 系数明显低于本文方法。BTNet^[15]采用边界注意力和 Transformer 来增强网络对目标边界的学习能力和网络的整体学习能力，但其训练时间较长，分割效果的提升并不明显，dice 系数不如本文方法。从整体上看，本文方法的 dice 系数、召回率和精确度在所有对比方法中是最高的。这说明本文方法对结节细节特征的把握更好，对病灶的定位能力更强，具有较高的分割精度。

表 4 在 DDTI 数据集上本文模型与其他文献模型的分割结果对比

Table 4 Comparison of segmentation results between the model in this paper and the models in the other literatures on the DDTI dataset				
方法	dice 系数	召回率	精确度	准确率
SGUNET ^[6] (2021)	0.729	0.783	0.753	0.931
MShNet ^[7] (2022)	0.758	0.778	—	—
Nguyen 等 ^[12] (2022)	0.612	—	—	—
BTNet ^[15] (2022)	0.757	—	—	—
本文模型	0.768	0.853	0.806	0.877

2.4.5 本文模型在 TN3k 数据集上的实验分析

为验证本文模型的实用性，我们选择 TN3k 甲状腺结节数据集^[25]进行验证。TN3k 数据集和 DDTI 数据集分别来自两个不同机构，具有不同的数据源

和数据量大小，该数据集包括来自 2 421 名患者的 3 493 张超声图像。我们把训练好的模型在 TN3k 数据集上进行测试，得到的实验结果为 DSC 系数为 0.765、召回率为 0.849、精确度为 0.796、准确率为 0.874。从指标中可以看出，本文模型在 TN3k 数据集上也取得了较好的效果。TN3k 数据集上的可视化分割结果如图 8 所示。从图 8 中可以看出，本文模型对甲状腺结节的分割较为完整，如图 8 第 1 行中的图像。对小结节和边缘模糊的结节也可以准确定位，如图 8 第 3 行中的图像。但是，对形状不规则的甲状腺结节，本文模型的分割结果仍存在不准确或边缘粗糙的现象，如图 8 第 2 行和第 4 行中的图像。今后将在模型改进时针对这个问题进一步优化，以提升甲状腺结节的分割效果。

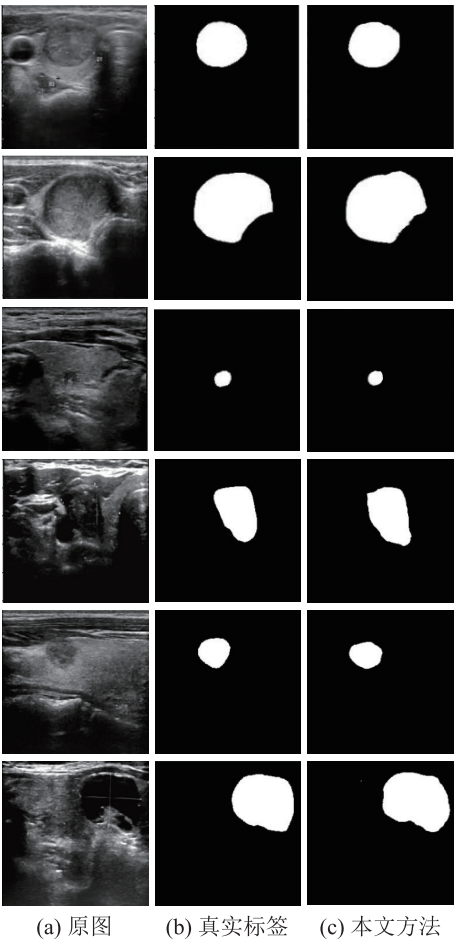


图 8 本文模型在 TN3k 数据集上的分割结果
Fig.8 Segmentation results of the model proposed in this paper on TN3k dataset

3 结 论

针对现有医学分割算法对超声甲状腺结节分割不准确的问题，本文基于 U-Net 网络，提出了具有多尺度特征提取和注意力机制的超声甲状腺结节分

割网络。首先,在网络的编码部分提出多感受野通道选择模块,使网络具有多尺度特征提取的能力。其次,在瓶颈层提出自适应全局上下文模块使其对高级语义信息进一步深入关注。最后,提出双注意力跨层引导模块替代编解码的跳跃连接结构,使浅层特征的细节信息和深层特征的全局信息得到充分融合,提高了模型的整体性能。实验结果表明,改进后的网络能够有效提升对甲状腺结节超声图像的分割精度。本文方法的分割结果在主观视觉上与真实标签更接近,对甲状腺结节有较好的分割效果,可辅助医生做出高效的诊断。如何降低模型复杂度、设计更高效的网络模型以及将这种分割框架应用于其他医学图像分割任务是下一步的研究方向。

参 考 文 献

- [1] IAKOVIDIS D K, SAVELONAS M A, KARKANIS S A, et al. A genetically optimized level set approach to segmentation of thyroid ultrasound images[J]. *Applied Intelligence*, 2007, **27** (3): 193-203.
- [2] KOLLORZ E, ANGELOPOULOU E, BECK M, et al. Using Power Watersheds to Segment Benign Thyroid Nodules in Ultrasound Image Data[M]//HANDELS H, EHRHARDT J, DESERNO T, et al. *Bildverarbeitung für die Medizin 2011*. Berlin, Heidelberg: Springer, 2011: 124-128.
- [3] 郑斌,徐峰,郭进祥,等.基于Chan-Vese模型的甲状腺结节超声图像分割方法[J]. *电脑与电信*, 2018(6): 4-7.
ZHENG Bin, XU Feng, GUO Jinxiang, et al. Thyroid nodule ultrasound image segmentation method based on Chan-ve-se model[J]. *Computer & Telecommunication*, 2018(6): 4-7.
- [4] 邵蒙恩,严加勇,崔峭峭,等.基于CV-RSF模型的甲状腺结节超声图像分割算法[J]. *生物医学工程研究*, 2019, **38**(3): 336-340.
SHAO Meng'en, YAN Jiayong, CUI Xiaoyao, et al. Ultrasound image segmentation of thyroid nodule based on CV-RSF algorithm[J]. *Journal of Biomedical Engineering Research*, 2019, **38**(3): 336-340.
- [5] OUAHABI A, TALEB-AHMED A. RETRACTED: Deep learning for real-time semantic segmentation: application in ultrasound imaging[J]. *Pattern Recognition Letters*, 2021, **144**: 27-34.
- [6] PAN H T, ZHOU Q, LATECKI L J. SGUNET: semantic guided UNET for thyroid nodule segmentation[C]//2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). Nice, France. IEEE, 2021: 630-634.
- [7] PENG Y J, YU D, GUO Y F. MShNet: Multi-scale feature combined with h-network for medical image segmentation[J]. *Biomedical Signal Processing and Control*, 2023, **79**: 104167.
- [8] KANG Q B, LAO Q C, LI Y Y, et al. Thyroid nodule segmentation and classification in ultrasound images through intra- and inter-task consistent learning[J]. *Medical Image Analysis*, 2022, **79**: 102443.
- [9] CHU C, ZHENG J H, ZHOU Y. Ultrasonic thyroid nodule detection method based on U-Net network[J]. *Computer Methods and Programs in Biomedicine*, 2021, **199**: 105906.
- [10] 王波,李梦翔,刘侠.基于改进U-Net网络的甲状腺结节超声图像分割方法[J]. *电子与信息学报*, 2022, **44**(2): 514-522.
- WANG Bo, LI Mengxiang, LIU Xia. Ultrasound image segmentation method of thyroid nodules based on the improved U-net network[J]. *Journal of Electronics & Information Technology*, 2022, **44**(2): 514-522.
- [11] NIE X Q, ZHOU X G, TONG T, et al. N-Net: a novel dense fully convolutional neural network for thyroid nodule segmentation[J]. *Frontiers in Neuroscience*, 2022, **16**: 872601.
- [12] TIEN N D, JIHO C, RYOUNG P K. Thyroid nodule segmentation in ultrasound image based on information fusion of suggestion and enhancement networks[J]. *Mathematics*, 2022, **10** (19): 3484.
- [13] STRUDEL R, GARCIA R, LAPTEV I, et al. Segmenter: transformer for semantic segmentation[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2022: 7242-7252.
- [14] CHI J N, LI Z L, SUN Z Y, et al. Hybrid transformer UNet for thyroid segmentation from ultrasound scans[J]. *Computers in Biology and Medicine*, 2023, **153**: 106453.
- [15] LI C F, DU R Q, LUO Q Y, et al. A novel model of thyroid nodule segmentation for ultrasound images[J]. *Ultrasound in Medicine & Biology*, 2023, **49**(2): 489-496.
- [16] LI X, WANG W H, HU X L, et al. Selective kernel networks [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2020: 510-519.
- [17] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 6230-6239.
- [18] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 7794-7803.
- [19] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **42**(8): 2011-2023.
- [20] PEDRAZA L, VARGAS C, NARVÁEZ F, et al. An open access thyroid ultrasound image database[C]//10th International Symposium on Medical Information Processing and Analysis. SPIE, Cartagena de Indias, Colombia, 2015, 9287: 188-193.
- [21] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[EB/OL]. 2015: arXiv: 1505.04597. <https://arxiv.org/abs/1505.04597>
- [22] ZHOU Z, SIDDIQUEE M M R, TAJBAKHS N, et al. UNet++: a nested U-net architecture for medical image segmentation[J]. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, **11045**: 3-11.
- [23] 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.
- [24] SCHLEMPER J, OKTAY O, SCHAAP M, et al. Attention gated networks: learning to leverage salient regions in medical images[J]. *Medical Image Analysis*, 2019, **53**: 197-207.
- [25] GONG H F, CHEN G Q, WANG R R, et al. Multi-task learning for thyroid nodule segmentation with thyroid region prior [C]//2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). Nice, France. IEEE, 2021: 257-261.