

DOI: 10.13382/j.jemi.B2306509

基于注意力机制的甲状腺超声图像 感兴趣区域定位方法^{*}

罗亦铭¹ 王建林¹ 田 艳² 张 波² 随恩光¹ 韩思齐³

(1. 北京化工大学信息科学与技术学院 北京 100029; 2. 中日友好医院超声医学科 北京 100029;
3. 首都医科大学燕京医学院 北京 101300)

摘 要:针对甲状腺超声图像中背景干扰及数据集规模受限的问题,提出了基于注意力机制的甲状腺超声图像感兴趣区域定位方法。采用跨尺度注意力交互策略,改进定位模型的特征网络,提高不同尺度下各层级特征的融合效率;通过知识蒸馏实现特征网络特征提取能力的强化,解决数据规模不足引起的网络过拟合问题;依据解剖学甲状腺形态统计分布设计t掩码,联合注意力掩码计算特征损失,引导网络对甲状腺超声图像关键通道和像素信息的学习,实现对甲状腺超声图像感兴趣区域的定位。实验结果表明,当IoU阈值为0.5时,甲状腺超声图像感兴趣区域定位AP达到92.7%,对辅助医生进行甲状腺疾病的诊断具有临床意义和价值。

关键词: 甲状腺超声图像;注意力机制;感兴趣区域;区域定位模型;特征网络
中图分类号: TP391;TN247 **文献标识码:** A **国家标准学科分类代码:** 510.8060

Location of regions of interest in thyroid ultrasound images based on attention mechanism

Luo Yiming¹ Wang Jianlin¹ Tian Yan² Zhang Bo² Sui Enguang¹ Han Siqu³

(1. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing
100029, China; 2. Ultrasound Medical Department, China Japan Friendship Hospital, Beijing 100029, China;
3. Yanjing Medical College, Capital Medical University, Beijing 101300, China)

Abstract: To address the problems of background interference and limited dataset size, we propose a method for locating the region of interest in thyroid ultrasound images. The method utilizes an attention mechanism based on cross-scale attention interaction strategy to improve the fusion efficiency of hierarchical features in the localization model. The feature network of the localization model is enhanced through knowledge distillation to solve the problem of overfitting. A t-mask is designed based on the statistical distribution of anatomical thyroid morphology, and a joint attention mask is calculated to guide the network in learning key channels and pixel information of thyroid ultrasound images, thereby achieving the localization of the region of interest. Experimental results demonstrate that the average precision (AP) for thyroid ultrasound image region of interest localization reaches 92.7% when the IoU threshold is set to 0.5, which is clinically significant and valuable for assisting doctors in diagnosing thyroid diseases.

Keywords: thyroid ultrasound image; attention mechanism; region of interest (ROI); region localization model; feature networks

0 引 言

在医学成像技术领域中,二维超声检查具有无创、安全、低成本以及能提供即时信息的特点,在表征结节性甲状腺疾病方面得到广泛应用^[1-2]。对甲状腺超声图像中的感兴趣区域(region of interest, ROI)定位是诊断甲状腺疾病的重要步骤,ROI为包含甲状腺腺体和特定背景的

矩形选框区域,对其准确定位能够降低甲状腺疾病诊断的错误率^[3]。传统的ROI定位方法是医生依据病理知识对超声图像进行手工标注,不仅增加了医生的工作负担,且当超声图像中的甲状腺区域与周边组织特征差异较小、边界模糊时,ROI的提取将严重依赖医生的主观认知和临床经验^[4]。智能超声影像诊断相比于医师阅片诊断,效率更高、客观性更强^[5],能够实现对甲状腺感兴趣区域的定位,已成为提高甲状腺疾病诊断准确率的重要

途径^[6-7]。

现有的甲状腺感兴趣区域智能定位利用卷积神经网络(convolutional neural network, CNN)良好的分层特征提取及高层语义表示能力,将甲状腺结节作为 ROI 进行单独提取^[8-10],Zheng 等^[11]首次将多阶段目标检测网络应用于甲状腺结节的定位任务,提出了改进的 Cascade CNN 级联检测器用于结节区域的定位,针对感兴趣区域目标尺寸不同的情况进行多尺度特征的融合,但其在计算过程中首先进行降维,丢失了部分通道层面信息。Ma 等^[12]在 Cascade CNN 的基础上提出了一种包含级联 CNN 结构的模型,在分割网络生成概率图后采取二值化、膨胀和侵蚀算子获得连通区域,再串联另一 CNN 对甲状腺 ROI 区域进行定位。Liu 等^[13]为了增强定位模型的抗干扰能力和表示能力,引入了注意力机制模块以实现自适应特征细化,同时以一种联合训练注释稀释的方法对数据集进行处理,利用结节周围伪前景区域作为空间上的先验约束条件,实现了对甲状腺结节的定位,然而其仅有相同层级特征进行了自注意力的计算,而未考虑不同层级特征之间的信息流动。Abdolali 等^[14]在 Mask R-CNN 的基础上,重新设计了带有正则化的损失函数,引入迁移学习对定位模型进行微调,实现了对甲状腺结节的定位,但其固有的自上而下的特征交互路径并未实现整体跨尺度的注意力交互。Liu 等^[15]建立了基于多尺度区域建议的检测模型,结合真实结节尺寸和形状分布的先验知识,在不同层级的特征图上实现了不同大小结节的精确定位,但该方法的特征网络在归一化组合各个不同尺度特征的过程中,统一固定了每个层级特征的分辨率,丢失了较多语义和像素级的信息。Wang 等^[16]改进 YOLO 检测器并引入迁移学习和数据增强实现对小尺寸结节的定位。Shi 等^[17]在改进特征提取网络残差连接方式的同时引入先验约束条件实现了甲状腺超声图像 ROI 区域的定位,在临床知识的约束下应用深度模型改善数据分部的系数情况提升定位准确率。这些方法在加深网络结构以追求更高定位准确率时疏于对医学图像数据集规模问题的关注,采用一般的数据增强算法不能从本质上解决模型过拟合和训练样本的多样性缺失的问题,使得大规模模型在数据较少的医学图像领域得不到最佳效果。知识蒸馏通过优化策略来提高定位模型的性能^[18],实现模型压缩的同时对模型进行增强。

本文提出一种基于注意力机制的甲状腺超声图像感兴趣区域定位方法。为解决跨尺度特征交互信息丢失及背景干扰问题,在特征金字塔网络的基础上,使用自注意力机制进行引导,改进跨层级特征的数据交互,对语义及像素信息差异较大的不同尺度特征进行整合并构建感兴趣区域定位模型;为解决医学图像数据集规模受限导致的模型冗余问题,通过前景背景分离的分部蒸馏和全局

蒸馏方法,引入乳腺超声图像进行迁移训练,构建增强的甲状腺超声图像感兴趣区域定位模型,降低计算量的同时提升精度。

1 基于注意力机制的甲状腺超声图像感兴趣区域定位方法

本文通过构建跨尺度注意力交互特征网络,增强特征网络的特征融合能力并改善信息流,减少特征网络不同层级特征交互时信息丢失,降低甲状腺超声图像背景干扰带来的影响,在同尺度自注意力交互部分,利用三重注意力机制对每个层级的特征进行处理,在通道数压缩前,对不同层级特征的通道注意力进行计算,构建特征金字塔输入至后续跨尺度空间注意力交互部分,依据空间注意力自适应地设计卷积核融合不同层级特征;在模型强化阶段引入解剖学中甲状腺形态统计数据设计 t 分部掩码,引导损失函数计算使特征更贴合真实甲状腺图像,采取分部知识蒸馏和全局知识蒸馏,利用注意力赋予前景背景不同关注度且不丢失全局信息,实现对甲状腺感兴趣区域定位模型的强化,降低医学超声图像数据规模受限问题对定位精度的影响。

1.1 跨尺度注意力交互特征网络

在同尺度注意力交互部分,由于骨干网络特征输入 FPN 前需要进行 1×1 的卷积操作以保持不同层级特征的通道数一致,这导致特征由于卷积核固定感受野的限制而缺失全局信息,引入三重注意力模块(triplet attention module, TAM)对每个层级的特征进行自注意力处理,有效捕获上下文信息,其架构如图 1 所示。

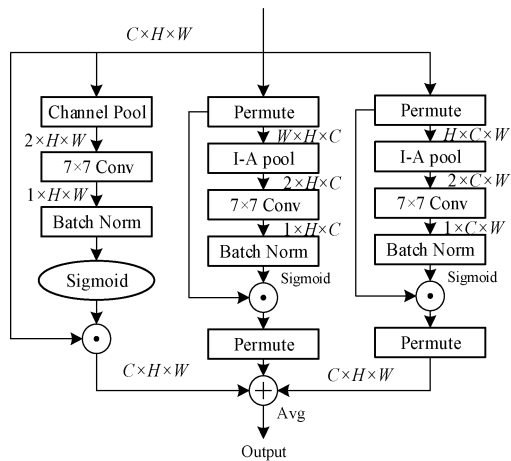


图 1 TAM 注意力模块结构

Fig. 1 TAM-triplet attention network structure

TAM 中两条分支通过在两个空间维度方向上对特征进行旋转,捕获通道和空间的相关性;另一条分支对空

间注意力单独计算。通过残差连接并取均值对其进行聚合。在前两支路当中输入特征张量 x 分别沿 H 和 W 轴逆时针旋转 90° 得到 x_1 及 x_2 , 而后通过拼接重要性池化与平均池化的 I-A pool 池化层压缩为 x_1^* 、 x_2^* , 而 x_3 为通过通道池化得到的张量。则对于特征提取网络输出特征 $x \in \mathbb{R}^{C \times H \times W}$ 有:

$$\hat{x} = \text{Avg}(x_1 \sigma(\phi_1(x_1^*)) + x_2 \sigma(\phi_2(x_2^*)) + x \sigma(\phi_3(x_3))) \quad (1)$$

式中: σ 表示 sigmoid 函数; ϕ_1 、 ϕ_2 、 ϕ_3 分别代表 3 条分支当中卷积核为 7×7 的卷积运算; Avg 为 3 条支路的均值运算。则在特征网络当中, 同尺度注意力机制的交互可以表示为:

$$C_i = \frac{F_{i1}^r \varphi_1 + F_{i2}^r \varphi_2 + F_i^r \varphi_3}{3} \quad (2)$$

式中: F_i^r 表示第 i 层特征; F_{i1}^r 和 F_{i2}^r 表示该层特征沿 H 和 W 轴旋转所得特征; φ_1 、 φ_2 和 φ_3 分别为 3 条分支中计算得到的注意力权重。

跨尺度通道注意力交互旨在获取各个尺度特征所包含的不同通道信息来优化各层级特征。该模块主要有两种工作方式: 分别为用高层级信息渲染低层级信息, 它采用自下而上的方式, 输出与较高级别的特征图具有相同的通道数的张量; 以及用低层级信息丰富高层级信息, 采用自上而下的方式, 输出与低层级的特征图具有同样通道数的张量。工作流程如图 2 所示, 在与高层特征进行交互时, 首先对高层特征进行反卷积以获得相同的尺寸

和通道数, 而后进行通道注意力的加权计算。整个流程表示为:

$$C'_{i+1} = \text{deconv}(C_{i+1}) \quad (3)$$

$$P_{(i,n-i)} = C_i \otimes \sigma(\text{Conv}(\text{GAP}(C'_{i+1}))) \oplus C'_{i+1} \quad (4)$$

式中: GAP 表示全局池化; 获得 C_i 与 C_{i+1} 间对应跨尺度通道注意力交互后的特征 $P_{(i,n-i)}$; 在完成每层特征与其他层级特征的交互之后, 对同一尺度的特征进行加权拼接, 而后输入到卷积层将它们调整为原始通道数。

模块第 i 层的输出为:

$$P_i^e = \text{Conv}(\psi(P_{i1}^e, P_{i2}^e, \dots, P_{in}^e)) \quad (5)$$

在跨尺度空间注意力交互中, 引入 CARAFE 模块对高层特征进行上采样, 相较于仅利用亚像素邻域的上采样方法具有更大的视野, 使得网络可以在更大的感受野内聚合上下文信息, 其具备内容感知处理, 能够动态生成自适应内核。在空间层面的跨尺度注意力交互当中将自适应卷积核生成结构加入到自上而下的特征融合通路中, 聚合相邻特征之间的互补信息以指导内容感知上采样, 并采用空间注意力来增强相邻特征融合时的像素信息一致性, 其结构如图 2 所示。通过上一通道注意力交互结构后生成的特征记为 $\{P_0^e, P_1^e, \dots, P_n^e\}$, 以 $P_i \in \mathbb{R}^{C \times H \times W}$ 的和其上方特征 $P_{i+1}^e \in \mathbb{R}^{C \times H/s \times W/s}$ 为例, 根据输入特征的内容生成具体的重组内核, 经过的 $k_{\text{encoder}} \times k_{\text{encoder}}$ 卷积得到尺寸为 $\sigma H \times \sigma W \times k_{\text{up}}^2$ 的上采样核, 而后对上采样和进行归一化处理, 如式 (6); 最终得到的新内核 $W_{i'}$ 与 $P_{(i,i+1)}^e$ 进行重新组合运算式:

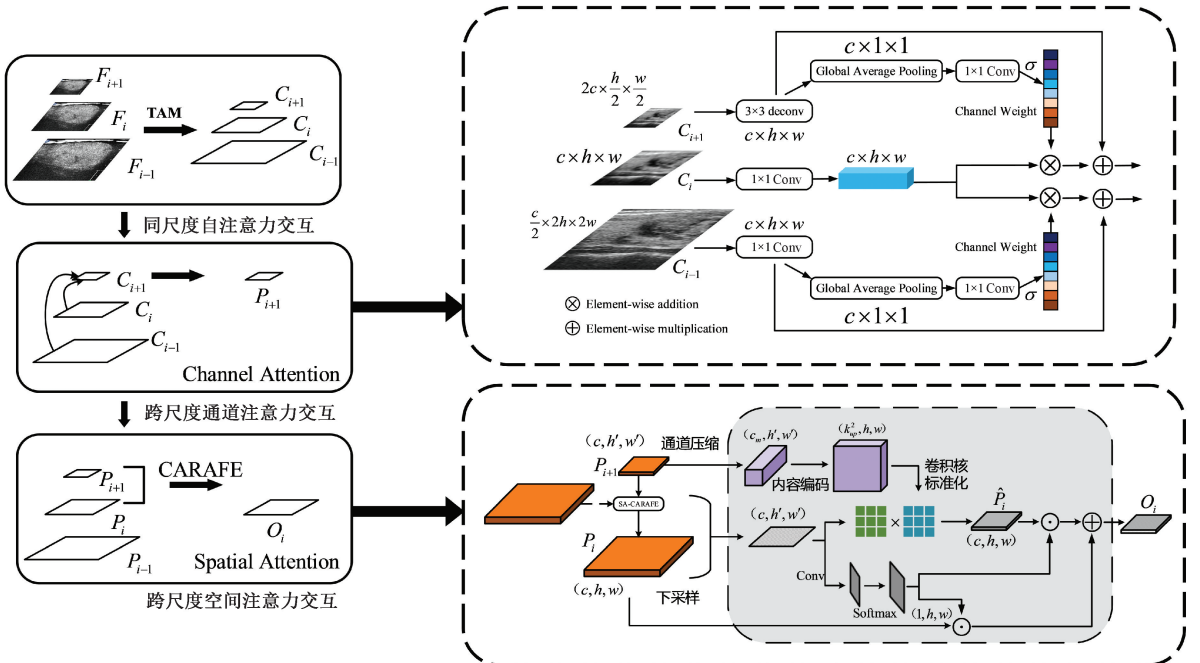


图2 跨尺度注意力交互特征网络

Fig. 2 Cross-scale attentional interaction feature network

$$W_{(i+1)'} = \psi(N(P_{i+1}, k_{\text{encoder}})) \quad (6)$$

$$\hat{P}_i = \phi(N(P_{(i,i+1)}, k_{\text{up}}), W_{(i+1)'}) \quad (7)$$

式中: ψ 和 δ 分别代表核预测模块和上下文感知重组模块; $N(F_l, k)$ 为以位置为中心的特征 F 的子区域, 即 F_l 的邻域; 在完成上采样过程后根据 $P_{(i,i+1)}^c$ 特征计算空间注意力权重, 分配给上采样后的特征 $\hat{P}_i$ 和原始的底层特征 P_i ; 最后对其进行元素相加并通过参数 α, β 调节融合过程, 表示为:

$$O_i = \alpha(M_s \odot \hat{P}_i^c) + \beta(M_s \odot P_i^c) \quad (8)$$

1.2 知识蒸馏的感兴趣区域定位模型强化

本文所提模型强化方法结构如图 3 所示, 由注意力掩码联合 Student-t 掩码引导下的分部蒸馏和基于 Gram 矩阵的全局蒸馏构成。

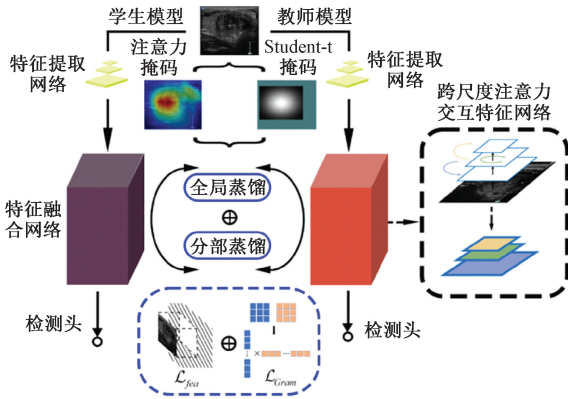


图 3 甲状腺感兴趣区域定位模型增强方法结构

Fig. 3 Thyroid localization model augmentation method

在特征信息蒸馏阶段, 为解决前景背景不平衡的问题及标注框对真实甲状腺组织的不契合问题, 通过引入类高斯分布的 student-t 分布掩码来突出前景当中的甲状腺腺体部分, 抑制标注框内其他背景组织的干扰。二维 student-t 掩码定义为:

$$t_1 = \frac{x - x_0}{\lambda_1 / \sqrt{w}}; t_2 = \frac{y - y_0}{\lambda_2 / \sqrt{h}} \quad (9)$$

$$\theta = \left[1 + \frac{t_1^2}{g_1} + \frac{t_2^2}{g_2} \right]^{-\frac{g_1 + g_2 + 2}{4}} \quad (10)$$

$$T = \begin{cases} \frac{\Gamma[(g_1 + g_2)/4 + 1/2]}{\Gamma[(g_1 + g_2)/4]} \theta & (x, y) \in GT \\ 0, & \text{其他} \end{cases} \quad (11)$$

式中: Γ 是伽马函数; t_1 和 t_2 对应两个维度上的参数比率; g_1 和 g_2 分别为纵横两个方向的衰减系数, 遵循医学统计规律进行设置以更贴合对象目标的形状特点^[19-20]。该掩码相较于二值掩码和高斯掩码能更好地凸显甲状腺

腺体组织。真实图像和掩码的对应情况如图 4 所示。

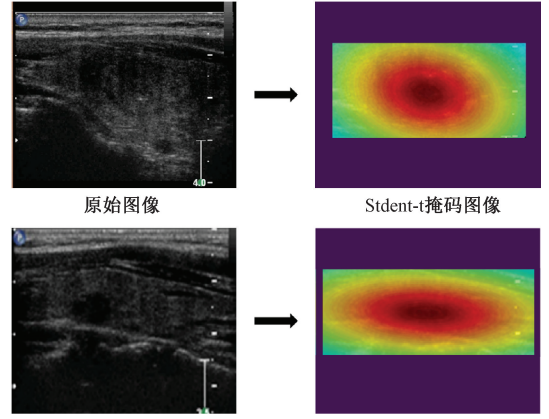


图 4 真实标注框与对应的 t 掩码对比

Fig. 4 Comparison of real image and student-t mask

在前景背景分部蒸馏过程中, 分别引入通道和像素上的注意力机制来分别对前景及背景区域进行注意力掩码的计算。对于通道注意力掩码的计算, ECA-Net 使用了一种局域的跨通道交互算法, 仅靠一维卷积就可实现通道注意力的计算, 降低网络负担。通道注意力的计算公式为:

$$C_{i,j} = C(\sigma(W_2(\text{ReLU}(\frac{W_1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W |F_{i,j}|))) / T) \quad (12)$$

式中: H, W, C 分别表示特征图的高宽以及通道数; T 是定义的蒸馏温度超参数; 采用的 ECA-Net 结构计算过程中将全连接层函数替换为了卷积核尺寸为 K 的卷积操作, 对应 W_2 和 W_1 ; 卷积核尺寸 K 能够根据通道情况自适应地进行调整, 计算公式如下:

$$K = \phi(C) = \left\lfloor \frac{\log_2(C)}{\lambda} + \frac{r}{\lambda} \right\rfloor_{\text{odd}} \quad (13)$$

式中: λ 和 r 为进行映射的参数; odd 表示取最邻近的奇数; 至此完成通道层面注意力掩码的计算, 在像素的层面的注意力计算公式为:

$$S_{i,j} = H \cdot W (\sigma(\frac{1}{C} \cdot \sum_{k=1}^C |F_c| / T)) \quad (14)$$

利用掩码处理后的教师特征与学生特征间的损失引导学生网络模型的训练, 综合上述的 student-t 和空间、通道层面的掩码, 提出基于特征的蒸馏联合损失函数:

$$L_{\text{Hint}} = \lambda_1 \sum_{k=1}^C \sum_{i=1}^H \sum_{j=1}^W M_{i,j} T_{i,j} C_{i,j} S_{i,j} (T_{c,h,w} - f(S_{c,h,w}))^2 + \lambda_2 (1 - M_{i,j}) (1 - T_{i,j}) C_{i,j} S_{i,j} (T_{c,h,w} - f(S_{c,h,w}))^2 \quad (15)$$

式中: λ_1 和 λ_2 为平衡甲状腺腺体区域和由其他组织构成的背景之间损失的超参数; $T_{c,h,w}$ 和 $S_{c,h,w}$ 分别为教师模型和学生模型所输出的特征; $M_{i,j}$ 表示由真实标注框获取的二值掩码; $T_{i,j}$ 为 t 分布掩码。

格拉姆矩阵 (Gram matrix) 为多维欧氏空间当中所选任意个向量之间的内积所组成的矩阵,表征了纹理信息,在灰阶超声图像中,甲状腺腺体与周围组织区别可由纹理特征的差异表示。

在全局信息蒸馏部分,采取 GC-Net 当中的全局注意力池构建方法,引入 GCT 当中的高斯激活函数对归一化后的全局上下文信息进行注意力激活,如下所示:

$$z = e^{-\frac{\text{norm}(\sigma(W_c x))^2}{2\gamma^2}} \quad (16)$$

式中: W_c 为卷积层; norm 表示标准化层; γ 为调节超参数。得到最终的格拉姆损失函数为:

$$L_{\text{Gram}} = \alpha \cdot \sum (\mathcal{G}(z^T) - \mathcal{G}(z^S))^2 \quad (17)$$

式中: $\mathcal{G}$ 为格拉姆矩阵运算, z^T 和 z^S 分别代表获取全局上下文信息之后教师和学生的特征。最终基于二阶段检测器得到学生检测器的联合损失函数如下:

$$L_{\text{Overall}} = L_{\text{cls}} + L_{\text{reg}} + L_{\text{RPN}} + L_{\text{Gram}} + L_{\text{Hint}} \quad (18)$$

2 甲状腺超声图像感兴趣区域定位算法

2.1 定位算法流程

图5为基于跨尺度注意力及模型增强的甲状腺灰阶超声图像感兴趣区域定位方法的算法流程,包括含有数据集构建、跨尺度注意力交互特征网络的教师模型训练和基于知识蒸馏的学生模型增强3个步骤。

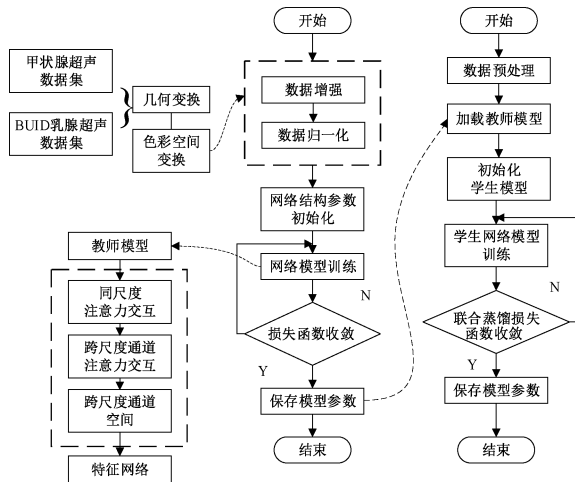


图5 基于注意力的甲状腺超声图像感兴趣区域定位模型增强方法算法流程

Fig. 5 Flow chart of the proposed algorithm

首先构建由甲状腺超声和乳腺超声图像组成的数据集并进行基于几何变换和色彩空间变换的数据增强。而后采取归一化、标准化对数据进行预处理。

其次,设置教师网络和训练参数。由于二阶段检测

器对于医学图像的定位更具优势,采用 Grid R-CNN 检测头和 ResNet-101 骨干网络,结合文中设计的特征网络构成教师模型。初始化网络后构建损失函数,对模型进行训练,迭代网络参数至损失函数收敛后获得基于改进特征网络的甲状腺超声图像感兴趣区域定位模型。

最后,搭建学生模型并初始化,其由 Grid R-CNN 检测头及 ResNet-50 骨干网络组成,加载训练完成的教师模型,采用文章的特征蒸馏方法使用甲状腺灰阶超声图像对学生模型进行训练。最终得到基于跨尺度注意力及模型增强的甲状腺灰阶超声图像感兴趣区域定位模型。

2.2 定位算法实现步骤

步骤1)通过几何变换 (geometric transformations) 及颜色空间变换 (color space transformations) 对采集切片得到的甲状腺超声图像及乳腺超声图像进行增强;

步骤2)将输入图像划分网格,生成共2000组锚框,采用跨尺度注意力交互特征网络提取特征,通过卷积操作对选框回归进行预测,得到分类概率 cls 及边界框 $bbox = [a_x, a_y, a_w, a_h]$;

步骤3)将数据集的训练集输入到网络模型,并对训练参数,如优化函数、初始学习率、epoch、迭代次数等进行设置,开始迭代训练教师网络,直到网络收敛;

步骤4)将上步训练得到的教师模型网络权重 θ_{bias} 固定,重新构建学生网络模型,输入仅由甲状腺超声图像构成的训练集,在知识蒸馏损失函数的引导下对学生模型即强化模型进行训练,直到网络收敛。

步骤5)得到甲状腺感兴趣区域定位模型后采用非极大值抑制方法 (non-maximum suppression, NMS) 对概率最大的预测目标进行选择,最终获取甲状腺感兴趣区域即甲状腺腺体的位置坐标,得到定位结果。

3 实验及结果分析

3.1 实验环境

本节实验均在 Ubuntu 18.04 系统下进行,硬件配置为 CPU Intel (R) Xeon (R) Gold 5217 3.00 GHz 64 GB; GPU Nvidia RTX 2080T 12.00 GB。采用 Pytorch 深度学习框架及 MMCV 算法库,实验代码基于 Python 3.7 编程语言实现。

3.2 实验数据集与评价指标

实验数据集选用甲状腺灰阶超声 CEUI-ID 数据集及乳腺超声 BUID^[21] 数据集。CEUS-ID 采集自北京市某三甲医院自2020年10月~2021年8月23日的真实病例的甲状腺超声造影数据,包含169个良性病例和331个恶性病例的横纵切面甲状腺超声图像共1000张。为构建适用于感兴趣区域定位任务的实验数据集,对每个视

频的 10~30 s 区间内随机选取一帧切片输出尺寸为 800×600 的图像作为样本。用于对教师模型进行补充训练的数据集为乳腺超声图像数据集 (BUID), 由 600 名女性的 780 张乳腺灰阶超声图像组成。

采用平均精度 (average precision, AP) 作为性能评价指标, 评价所提出的基于注意力机制的甲状腺超声图像感兴趣区域定位算法的定位性能。实验中, 预测框和真实甲状腺标注框间的 IOU 大于设定阈值时记为 TP, 预测框和目标框并集以外的部分为 TN, 没有被检测出来的目标框记为 FN, 被错误检测出来的部分记为 FP。根据 TP、TN、FP、FN 对精度和召回率进行计算:

$$P = precision = \frac{TP}{TP + FN}$$

(19)

$$R = recall = \frac{TP}{TP + FN}$$

(20)

绘制 Precision-Recall 曲线即 P-R 曲线并积分得到 AP 值, 计算过程可以表示为:

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) p_{interp}(r_{i+1})$$

(21)

3.3 特征网络实验

训练过程采用随机梯度下降 (stochastic gradient descent, SGD) 优化算法迭代更新网络节点权值; 训练动量选为 0.9, 权值衰减系数为 1×10^{-4} 。采用余弦衰减学习率调整策略, 初始学习率 lr 设置为 2×10^{-3} , 同时为防止初始学习率较大导致模型陷入局部最优, 采用 warmup 策略, 线性调整学习率至第 5 个 epoch 达到初始 lr , 未特殊注明情况下, 训练采用 60 个 epoch。

为分析所提特征网络中每个模块的重要性, 将同尺度的自注意力交互模块和跨尺度的注意力交互模块分别应用于定位模型以验证其有效性; 此外, 将不同模块进行组合, 以讨论上述模块单独作用和级联后对模型性能的影响情况, 实验结果如表 1 所示。结果表明, 每个模块均有效提高了 baseline 的性能, 具体而言同尺度注意力交互模块使得 AP 提升了 23.7%, 得益于该模块在相同分辨率的特征图上, 使每个独立像素均包含特征整体信息, 减少了语义丢失, 最终检测头能更有效地关注甲状腺腺体组织区域; 同样, 跨尺度注意力交互模块使得 AP 提升 37.0%, 表明该模块能有效从通道和空间层面利用尺度不同的张量, 丰富金字塔各层级特征包含的信息, 注意力引导的统筹策略降低了不同层级特征融合带来的差异和冗余信息。将两个模块组合后 baseline 的性能进一步提升, 主要体现在指标 AP_{75} 有 60.2% 的提升上, 表明在更严厉的评判准则中模型能有更好的表现, 对于更加精细的特征有更好的识别效果。

表 1 各交互模块消融实验结果对比

Table 1 Ablation study of different interactive module

同尺度注意力交互模块	跨尺度注意力交互模块	AP	AP ₅₀	AP ₇₅
✓	Baseline	0.316	0.806	0.211
		0.391	0.848	0.284
	✓	0.433	0.908	0.254
✓	✓	0.493	0.931	0.478

跨尺度注意力交互网络是一种改进的特征网络。为了验证方法的有效性, 将所提出的模型架构与目标检测领域先进的特征网络进行对比, 检测器均采用 Grid R-CNN, 结果如表 2 及图 6 所示。

表 2 各特征网络实验结果

Table 2 Comparison of different feature network

方法	AP	AP ₅₀	AP ₇₅
FPN	0.416	0.876	0.381
PA-FPN	0.428	0.903	0.397
Bi-FPN	0.430	0.916	0.402
NAS-FPN	0.453	0.922	0.433
BFP	0.449	0.921	0.410
本文	0.493	0.931	0.478

结果表明, 本文所提方法在 CEUS-ID 甲状腺灰阶超声测试集上感兴趣区域定位 $AP = 0.479$, $AP_{50} = 0.927$, $AP_{75} = 0.474$, 在同等实验条件下相较于基础 FPN 取得了 15.14% 的 AP 提升。相较于 PA-FPN、Bi-FPN、NAS-FPN 和 BFP 分别取得了 11.92%、11.40%、5.74% 和 6.68% 的 AP 提升。

3.4 模型强化实验

模型强化实验与其他两种用于目标检测的注意力引导蒸馏方法进行比较。在本次实验当中, 采用 ResNet-50 作为学生检测器模型的骨干网络, 选择具有 ResNet-101 的相同检测器作为教师模型, 采用基于锚框的二阶段检测器 Grid R-CNN 作为基础 baseline。针对 t 掩码根据甲状腺形态统计学数据进行配置, 对应的超参数设置为 $\alpha = 1 \times 10^{-4}$ 、 $\beta = 5 \times 10^{-5}$; $\lambda = 1 \times 10^{-6}$, 统计数据如表 3 所示。

表 3 正常成人甲状腺检查结果尺寸统计信息

Table 3 Statistical information on the size of thyroid

	长度/cm	宽度/cm	厚度/cm
甲状腺左叶	5.19±0.66	1.29±0.17	1.74±0.18
甲状腺右叶	5.56±0.73	1.34±0.18	1.82±0.18

实验结果如表 4 所示, 根据结果可以得到所提蒸馏方法在两种不同类别的检测器上都有显著的 AP 提升。由于教师模型过深, 在训练数据不足的情况下有较为严重的过拟合现象, 相对于学生模型均有不同程度的 AP

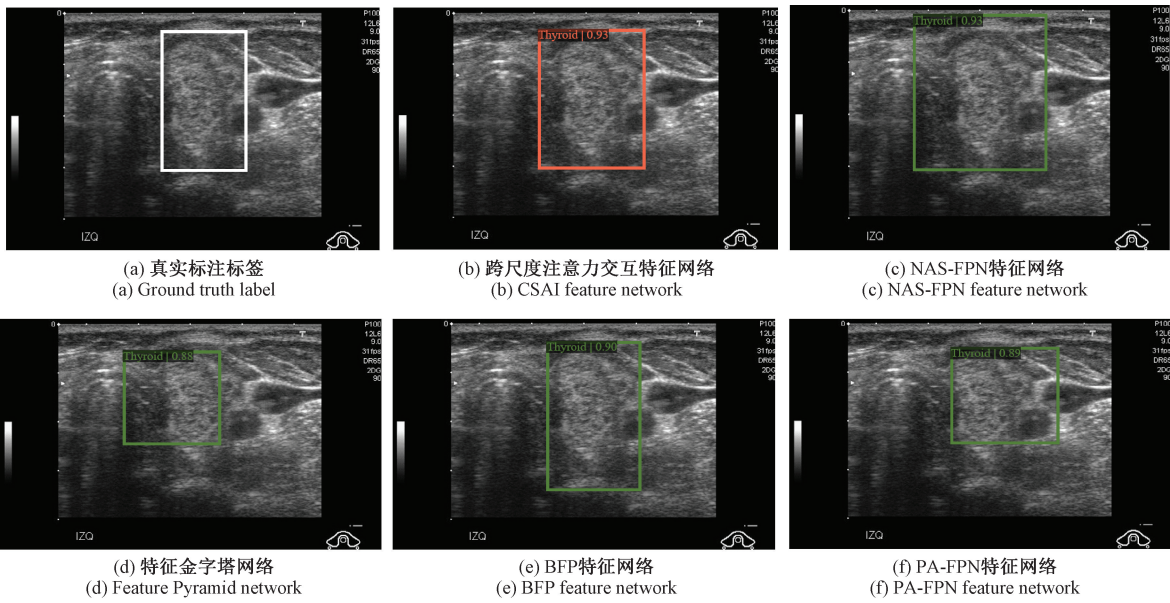


图 6 不同特征网络方法对应甲状腺灰阶超声数据集上定位效果

Fig. 6 Comparison of localization effects of different feature network methods

下降。在引入乳腺超声数据并对学生模型蒸馏之后,基于双阶段检测器的学生模型相较于蒸馏前模型 AP 分别提升了 9.78% 和 6.83%,表明更深的检测器能够从知识蒸馏当中受益更多。同时能有效改善复杂模型因训练数据集不足造成的过拟合,能够进一步加深网络结构,提升模型的性能。相较于 AGD 和 FGD 方法,本文蒸馏方法在二阶段检测器上的 AP 提升了 9.31% 及 3.36%,与 AGD 和 FGD 模型强化方法的 Grad CAM 热图及定位结果对比如图 7 所示。

表 4 模型强化实验结果

方法	AP
<i>Two-stage</i>	
Grid R-CNN-Res101(T)	0.431
Grid R-CNN-Res50(S)	0.449
AGD	0.451
FGD	0.477
本文	0.493

3.5 甲状腺超声图像感兴趣区域定位方法实验

除上述各部分的独立实验外,本文还将所提出的特征网络集成至 Grid R-CNN 检测器进行增强之后当中与目前先进的目标检测及感兴趣区域定位模型在甲状腺灰阶超声数据集上进行对比实验,实现对应检测器所提出的特征网络的配套作为基线参考模型。包括一阶段和二阶段检测器、基于锚框和 anchor free 的检测器等,最终结果如表 5 所示。相较一阶段检测器,二阶段检测器在精度上更具有优势,对比 anchor free 方法在训练轮次更少

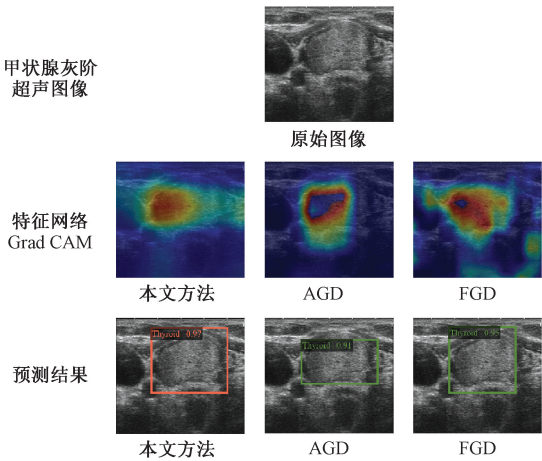


图 7 本文方法与其他模型蒸馏方法的 Grad CAM 热图及甲状腺超声图像定位结果对比

Fig. 7 Comparison of Grad CAM heat map and positioning results between different model distillation methods

的情况下能达到更高的 AP 精度,在 IoU 阈值设定为 0.5 的条件下定位模型的 AP 达到 93.1%。

4 结 论

本文针对甲状腺超声图像背景干扰和医学图像数据规模限制导致对甲状腺超声图像感兴趣区域定位效果不佳的问题,提出了基于注意力的甲状腺超声图像感兴趣区域定位模型增强方法。通过引入三重注意力模块并解耦空间与通道注意力,强调特征网络当中更具价值的

表 5 在甲状腺超声图像上与各先进方法定位实验结果对比

Table 5 Comparison of localization experimental results with advanced methods				
方法	Backbone	AP	AP ₅₀	AP ₇₅
<i>Two-stage</i>				
Faster R-CNN	ResNet-50+FPN	0.407	0.892	0.275
Mask R-CNN	ResNet-50+FPN	0.410	0.904	0.319
Cascade R-CNN	ResNet-50+PAFPN	0.429	0.925	0.321
Grid R-CNN	ResNet-50+BFP	0.449	0.921	0.410
<i>One-stage</i>				
YOLOv3	Darknet-19+YOLO NECK	0.347	0.891	0.198
SSD 512	ResNet-50	0.334	0.892	0.196
RetinaNet	ResNet-50+FPN	0.439	0.900	0.367
<i>Anchor free</i>				
FCOS	ResNet-50+FPN	0.418	0.900	0.325
FSAF	ResNet-50	0.415	0.909	0.329
本文	ResNet-50+CSAI	0.493	0.931	0.478

信息抑制冗余信息;其次,引入先验知识和注意力引导的分部蒸馏结合格拉姆矩阵全局蒸馏,通过引入同数据域乳腺超声图像对模型进行训练使网络获得更强的特征提取能力,最终对定位模型进行增强。基于自建甲状腺灰阶超声图像数据集的对比实验结果表明,经过特征网络改进并强化后的模型在检测精度上相较于现有文献方法展现了明显的优势。本文方法可以为医生诊断及计算机辅助诊断算法提供有效信息,满足临床对感兴趣区域定位的准确性要求。

参考文献

[1] FILETTI S, DURANTE C, HARTL D, et al. Thyroid cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up [J]. Annals of Oncology, 2019,30(12): 1856-1883.

[2] ROSSI E D, PANTANOWITZ L, RAFFAELLI M, et al. Overview of the ultrasound classification systems in the field of thyroid cytology [J]. Cancers, 2021 (13): 3133.

[3] GULAME M B, DIXIT V V, SURESH M. Thyroid nodules segmentation methods in clinical ultrasound images: A review [J]. Materials Today: Proceedings, 2021: 2270-2276.

[4] XIE F, LUO Y K, LAN Y, et al. Differential diagnosis and feature visualization for thyroid nodules using computer-aided ultrasonic diagnosis system: Initial clinical assessment [J]. BMC Medical Imaging, 2023, 23(1): 1-7.

[5] 邸拴虎, 杨文瀚, 廖苗, 等. 基于 RA-Unet 的 CT 图像肝脏肿瘤分割 [J]. 仪器仪表学报, 2022, 43(8): 65-72.

DI SH H, YANG W H, LIAO M, et al. Liver tumor segmentation from CT images based on RA-Unet [J].

Chinese Journal of Scientific Instrument, 2022, 43(8): 65-72.

[6] CHEN H, YU M A, CHEN C, et al. FDE-net: Frequency-domain enhancement network using dynamic-scale dilated convolution for thyroid nodule segmentation [J]. Computers in Biology and Medicine, 2023: 106514.

[7] GONG H, CHEN J, CHEN G, et al. Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules [J]. Computers in Biology and Medicine, 2022: 106389.

[8] YU R, LIU K, WEI X, et al. Localization of thyroid nodules in ultrasonic images [C]. International Conference on Wireless Algorithms, Systems, and Applications, 2018: 635-646.

[9] HARO A R, TOALOMBO J C, GALARZA E E, et al. Detection of thyroid nodules through neural networks and processing of echographic images [C]. International Conference on Computational Science and Its Applications, 2020: 166-178.

[10] CHEN L, ZHENG W, HU W. MTN-Net: A multi-task network for detection and segmentation of thyroid nodules in ultrasound images [C]. International Conference on Knowledge Science, Engineering and Management, 2022: 219-232.

[11] ZHENG Y, QIN L, QIU T, et al. Automated detection and recognition of thyroid nodules in ultrasound images using improve cascade mask R-CNN [J]. Multimedia Tools and Applications, 2022(10): 13253-13273.

[12] MA J, WU F, JIANG T A, et al. Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images [J]. Medical Physics, 2017(5): 1678-1691.

[13] LIU R, ZHOU S, GUO Y, et al. Nodule localization in thyroid ultrasound images with a joint-training convolutional neural network [J]. Journal of Digital Imaging, 2020(5): 1266-1279.

[14] ABDOLALI F, KAPUR J, JAREMKO J L, et al. Automated thyroid nodule detection from ultrasound imaging using deep convolutional neural networks [J]. Computers in Biology and Medicine, 2020: 103871.

[15] LIU T, GUO Q, LIAN C, et al. Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural networks [J]. Medical Image Analysis, 2019: 101555.

[16] WANG X, ZHANG N, DING M, et al. YOLO-TN: Ultrasound diagnosis of thyroid nodules based on transfer learning and convolutional neural networks [J]. International Journal of Adaptive and Innovative Systems,

2022(2): 87-99.

- [17] SHI M, DING J, ZHAO S, et al. Automatic thyroid ultrasound image detection and classification with priori knowledge [C]. The 5th International Conference on Computer Science and Application Engineering, 2021: 1-6.
- [18] 彭继慎, 孙礼鑫, 王凯, 等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J]. 仪器仪表学报, 2021, 42(10): 161-170.
- PENG J SH, SUN L X, WANG K, et al. ED-YOLO power inspection UAV obstacle avoidance target detection algorithm based on model compression [J]. Chinese Journal of Scientific Instrument, 2021, 42 (10): 161-170.
- [19] KAPRAL N, KHOT R. Thyroid anatomy and ultrasound evaluation [J]. Techniques in Vascular and Interventional Radiology, 2022: 100818.
- [20] VIDUETSKY A, HERREJON C L. Sonographic evaluation of thyroid size: A review of important measurement parameters [J]. Journal of Diagnostic Medical Sonography, 2019(3): 206-210.
- [21] AL-DHABYANI W, GOMAA M, KHALED H, et al. Dataset of breast ultrasound images [J]. Data in Brief, 2020: 104863.

作者简介



罗亦铭, 2020 年于北京化工大学获得学士学位, 现为北京化工大学硕士研究生, 主要研究方向为医学图像处理。

E-mail: 18970875409@163.com

Luo Yiming received his B. Sc. degree from Beijing University of Chemical Technology in 2020. He is currently a M. Sc. candidate at Beijing University of Chemical Technology. His main research interest includes medical image processing.



王建林 (通信作者), 1993 年于天津大学获得硕士学位, 1997 年于天津大学获得博士学位, 现为北京化工大学教授, 主要研究方向为智能医学影像处理、智能检测与传感技术。

E-mail: wangjl@mail.buct.edu.cn

Wang Jianlin (Corresponding author) received his M. Sc. degree from Tianjin University in 1993 and Ph. D. degree from Tianjin University in 1997, respectively. He is currently a professor in Beijing University of Chemical Technology. His main research interests include visual detection technology, intelligent detection and sensing technology.