基于深度学习算法 Mask R－CNN 的甲状腺结节检测模型研究*

王 杰[1] 王至诚[1] 娄 帅[2] 董建成[2] 曹新志[1]

([1] 江苏省中西医结合医院信息中心 南京 210028 [2] 江苏中康软件有限责任公司 南通 226001)

〔摘要〕 **目的/意义** 采用基于区域卷积神经网络的目标掩码分割算法(mask region－based convolutional neural network, Mask R－CNN)建立目标检测模型,智能识别甲状腺超声图像结节位置,为超声医生决策提供参考。**方法/过程** 收集超声结节图像 1 650 张,使用 labelme 工具进行结节位置标注。对 Mask R－CNN 的主干网络分别采用 MobileNetV3、ResNet50、ResNet101 和 ResNet152 进行替换,并引入特征金字塔和感兴趣区域对齐,采用迁移学习训练策略训练模型,比较不同网络下目标检测效果。**结果/结论** 主干网络采用 ResNet101 训练的模型平均精确度为 86.8%,平均召回率为 95.3%,平均 *F*1 分数为 90.6%,优于其他主干网络,能更精确地检测甲状腺结节,具有一定临床应用价值。

〔关键词〕 甲状腺结节;Mask R－CNN;目标检测;神经网络

〔中图分类号〕R－058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2025.03.015

Study on Thyroid Nodule Detection Model Based on Deep Learning Algorithm Mask R－CNN

WANG Jie[1], *WANG Zhicheng*[1], *LOU Shuai*[2], *DONG Jiancheng*[2], *CAO Xinzhi*[1]

[1] *Department of Information, Jiangsu Province Hospital on Integration of Chinese and Western Medicine, Nanjing 210028, China*; [2] *Jiangsu Zhongkang Software Co. Ltd., Nantong 226001, China*

〔**Abstract**〕 **Purpose/Significance** To establish an object detection model through mask region－based convolutional neural network (Mask R－CNN), so as to intelligently identify the nodule location in thyroid ultrasound images and provide references for the decision－making of ultrasound doctors. **Method/Process** 1 650 ultrasound nodule images are collected, and the labelme tool is used to label the nodule locations. The backbone network of Mask R－CNN is replaced by MobileNetV3, ResNet50, ResNet101 and ResNet152 respectively, feature pyramid network and region of interest align are introduced. The model is trained using transfer learning training strategy and the object detection performance is compared under different networks. **Result/Conclusion** The model trained with ResNet101 for the backbone network has an average accuracy of 86.8%, an average recall rate of 95.3%, and an average *F*1 score of 90.6%, which is superior to other backbone networks and can detect thyroid nodules more accurately, and has certain clinical application value.

〔**Keywords**〕 thyroid nodule; Mask R－CNN; object detection; neural network

〔修回日期〕 2024－11－26

〔作者简介〕 王杰,工程师,发表论文 2 篇;通信作者:曹新志,博士,高级工程师。

〔基金项目〕 国家自然科学基金资助项目(项目编号:81971708)。

1 引言

甲状腺结节是甲状腺腺体组织内的一种异常增生，可以是单发或多发，常见症状包括颈部肿块或肿胀感、吞咽困难、声音嘶哑、甲状腺功能异常等。甲状腺结节大多数为良性，恶性率为 5% ~ 7%[1]。目前，超声检查因简单方便、经济易行、检查时间短等特点成为临床首选的甲状腺结节检测手段[2]。超声医师根据超声图像判断结节的数量和位置，其准确性对于患者后续诊疗至关重要。而不同超声医师对结节的判断存在主观差异，容易造成误诊或漏诊。

基于人工智能深度学习算法自动检测甲状腺结节位置可有效避免观察者偏差[3]，提高诊断效率和准确率。近年来已有不少自动化或半自动化甲状腺结节检测方法。Zhang L 等[4]优化改进基于 YOLOV3 的算法模型，用于检测甲状腺结节。YOLOV3 是一种单阶段目标检测算法，使用单个网络同时产生候选区域并预测物体类别和位置，检测速度较快，但准确度和对不同尺寸的物体适应性较差。刘明坤等[5]和吴雯娟等[6]使用基于区域卷积神经网络的目标掩码分割算法（mask region - based convolutional neural network，Mask R - CNN），能够较好地分割或检测出甲状腺结节。郑天雷等[7]提出基于改进算法 Faster R - CNN 实现对甲状腺结节超声图像目标检测，其主干网络采用残差网络（residual network，ResNet）替换原有的 VGG16 网络，提高对目标结节的检测能力。以上实验模型均采用单一主干网络进行训练，无法知晓不同主干网络对模型的影响。

本研究提出一种基于 Mask R - CNN 的卷积神经网络模型实现甲状腺结节检测，采用 4 种不同的主干网络（MobileNetV3、ResNet50、ResNet101、ResNet152）进行特征提取，引入特征金字塔（feature pyramid network，FPN）和感兴趣区域对齐（region of interest align，ROIAlign），前者能够处理不同尺寸的目标，后者通过更精细的坐标计算，提供更准确的特征提取，提高小目标检测精度[7]。对比不同主干网络甲状腺结节检测效果，选择最优检测模型，为超声医师提供可靠、准确的结节检测位置参考。

2 实验与方法

2.1 数据预处理

采用甲状腺图像数字数据库（digital database thyroid image，DDTI）和甲状腺结节（thyroid nodule 3 thousand，TN3K）公开数据集训练和验证模型，DDTI 数据集包含来自单一设备的 637 张带有像素级标签的超声甲状腺成像，TN3K 数据集包括 3 493 张超声图像，来自 2 421 名患者[8]。从两个数据集中选取 1 650 张结节图像，先进行预处理裁剪，去除边框黑色阴影区域，并使用 labelme[9]工具对甲状腺图像结节区域进行标注，用矩形框框定结节位置；然后使用 labelme2voc. py 脚本生成 PASCAL VOC 数据集格式文件；最终研究数据集包含甲状腺结节超声图像、标注图像以及对应的 XML 格式标签文件。以 8∶1∶1 的比例随机划分为训练集、验证集和测试集，分别用于模型训练、参数调整和模型评估。

在深度学习图像领域，为了增加训练样本数量，常常采用数据增广，对训练集进行随机水平翻转、色调饱和度调整、随机旋转缩放平移等，本研究采用随机水平翻转对图像进行预处理，对应的 XML 标签中结节坐标信息也做翻转处理，扩充后的数据集能够避免模型过拟合[10]。

2.2 Mask R - CNN 网络模型构建

主干网络分别采用轻量级网络 MobileNetV3 和深度 ResNet 系列架构进行特征提取，比较不同网络的平均精确度、平均召回率和平均 *F*1 分数。引入 FPN 和 ROIAlign，并采用非极大值抑制（non - maximum suppression，NMS）保留图像上得分最高的检测框。

2.2.1 Mask R - CNN Mask R - CNN 是一个多任务深度学习模型，在 Faster R - CNN 算法的基础上发展而来，具备强大的目标检测能力，是目前流行的目标检测算法[11]。通过其 Mask 分支可完成分割

任务，使用了若干不同的网络模型，包括 ResNet、FPN、区域候选网络（region proposal network，RPN）、分类和回归等模型。基本结构，见图 1。

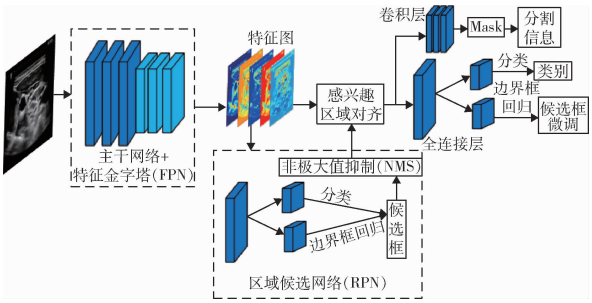


图 1 Mask R－CNN 基本结构

2.2.2 特征提取网络选取 特征提取网络由多个神经网络组成，用于图像的特征提取。本研究使用 ResNet 和 MobileNetV3 作为主干网络，ResNet 是一种深度神经网络结构，广泛应用于图像分类、目标检测和语义分割等计算机视觉任务，根据层数不同分为浅层网络和深层网络，其中 ResNet50、ResNet101 和 ResNet152 属于深层网络，采用批量标准化技术加速训练，解决梯度消失或梯度爆炸问题[12]，从而提高网络特征提取能力。然而某些应用场景如移动或嵌入式设备，在内存不高的条件下，难以应用如此复杂的模型。MobileNetV3 经过 V1 和 V2 两个版本的积累，性能和速度表现优异，具有参数量少、速度快、深度适中等优点[13]，可作为主干网络应用于目标检测任务。本研究采用迁移学习方法进行模型训练，通过复用预训练网络参数替代随机初始化参数的策略，在小样本数据条件下仍能实现理想的建模效果。该方法不仅显著降低训练成本[14]，同时提升模型收敛效率，并有效抑制过拟合风险。选取 ResNet50、ResNet101、ResNet152 和 MobileNetV3 进行检测效果对比，结合 FPN 提升目标检测性能。网络结构，见图 2。

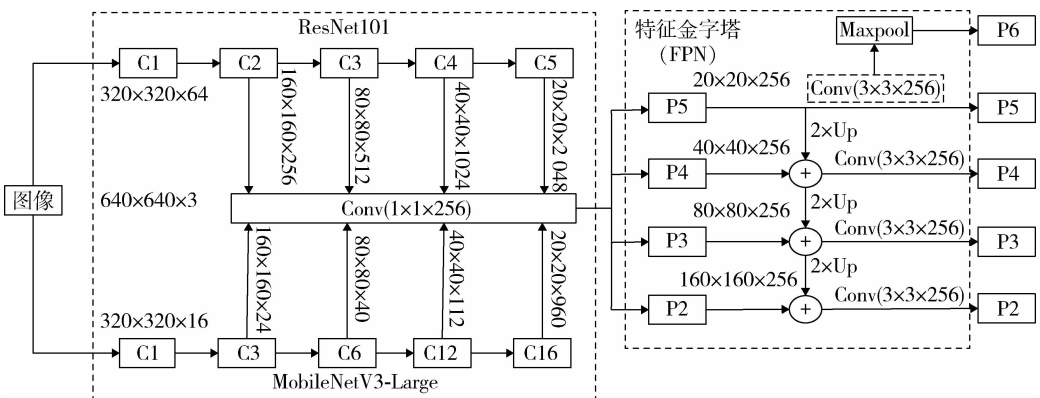


图 2 FPN 网络结构

ResNet 深层网络以 ResNet101 + FPN 为例，包括自下而上的 ResNet101 支路、自上而下的上采样支路和横向连接 3 个部分。自下而上的路径为卷积层 C1 到 C5 的前向特征提取过程，图像经过这些层时尺寸和通道数会发生变化，如经过第 5 层时 C5 的特征图长宽和通道数为 20、20 和 2 048，每层的横向连接中会将最后的输出特征图降维处理，进行卷积核为 1×1、通道数为 256 的卷积运算，不改变图像长宽并将特征图通道数都置为 256，得到特征图 P2、P3、P4、P5；然后与上采样的特征图进行加和操作，在融合之后再采用 3×3 的卷积核对已经融合的特征进行处理，目的是消除上采样的混叠效应[6]，得到最终输出特征。MobileNetV3 是轻量化网络模型之一，采用轻量级的深度可分离卷积和残差块等结构，减少计算成本和训练时间[15]。采用 MobileNetV3－Large 版本，网络一共 20 层，选取 4 层进行特征提取，从索引 0 开始，依次选取索引为 3、6、12、16 的 4 层，对应 4、8、16、32 倍下采样率，经卷积处理后生成的特征图尺寸分别标记为 C3、C6、C12、C16，FPN 网络将含有不同尺寸的特征图进行融合，使最终的特征图在各尺寸上都能含有较为丰富的空间信息和语义信息，这种网络模

型的应用在速度和精度之间取得了平衡[16]。

2.2.3 目标检测框生成 在 FPN 输出的 5 个特征图（P2—P6）上分别对应设置 5 个像素大小不同（32、64、128、256、512）的候选框，并设置 3 种长宽比（0.5、1.0 和 2.0），即每个特征图的每个像素点生成 3 个候选框，其中尺寸较大的特征图设置尺寸较小的候选框，能够更好地检测小目标物体，而较小的特征图上的单元划分较稀疏，每个单元设置的先验框尺寸较大，以适应大目标的检测。通过利用不同尺寸的特征图来检测不同大小的物体，提高目标检测的准确性和效率[6,17]。在经过分类和回归修正后，RPN 会筛选出高质量的 ROI，这些区域包含了检测目标的分类、Mask 信息和边框修正信息。接着基于 ROIAlign 进行更高精度的特征提取，再输入后续网络中，进行目标的分类、边界框的精确回归，完成对目标物体的检测。

2.3 评价指标

为了比较不同主干网络下模型的检测效果，将交并比（intersection over union，IoU）的阈值设置为 0.5，如果 IoU 大于 0.5 则认为甲状腺结节被正确检测到，使用平均精确度（precision，P）、平均召回率（recall，R）和平均 *F*1 分数作为检测评价指标。其中真阳性（true positive，TP）表示正确检测出甲状腺结节位置，假阳性（false positive，FP）表示模型将一个不存在甲状腺结节的区域错误地判断为有结节，假阴性（false negative，FN）表示模型没有检测出一个实际含有甲状腺结节的区域。

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{3}$$

3 研究结果

3.1 不同主干网络实验结果对比

实验采用以深层网络 ResNet 和 MobileNetV3 - Large 为主干网络的 Mask R - CNN 模型，对含有甲状腺结节的图像数据集进行训练，其中批处理大小参数（batch size）设置为 4，迭代次数（epoch）设置为 100，采用学习率衰减策略，初始学习率设置为 0.004，衰减的倍率因子为 0.1。以 ResNet101 为主干网络时 Mask R - CNN 训练过程中的损失和学习率变化情况，见图 3，当迭代到第 16 轮时学习率衰减到 0.000 4，到第 22 轮时衰减到 0.000 04，适当的学习率衰减使模型在不同阶段达到更好的训练效果和稳定性。

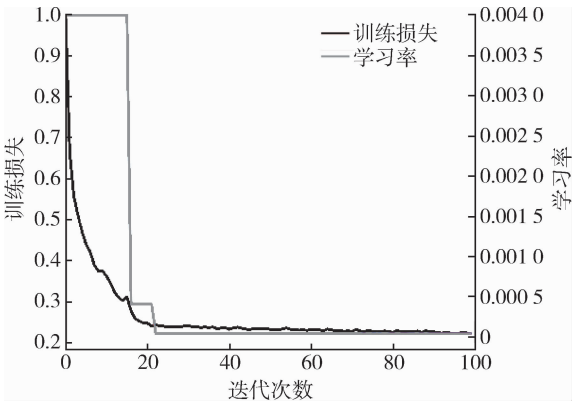


图 3 训练损失和学习率变化曲线

4 种不同主干网络模型经过 100 轮次迭代训练，甲状腺结节检测平均精确度、平均召回率、平均 *F*1 分数和耗时对比结果，见表 1。

表 1 不同主干网络实验结果对比

主干网络	平均精确度	平均召回率	平均 *F*1 分数	每次迭代耗时（秒）
MobileNetV3	0.802	0.962	0.875	0.26
ResNet50	0.803	0.923	0.855	0.33
ResNet101	0.868	0.953	0.906	1.45
ResNet152	0.844	0.942	0.887	6.48

以轻型网络 MobileNetV3 为主干网络的 Mask R - CNN 训练速度快，耗时少，但是精确度比深层网络 ResNet101 和 ResNet152 低。ResNet152 与 ResNet101 精确度相差较小，耗时却是 ResNet101 的 4 倍之多，训练速度较慢。随着迭代次数增多，不同网络模型平均精确度不断上升，后趋于稳定，其中 ResNet101 网络平均精确度最高，见图 4。

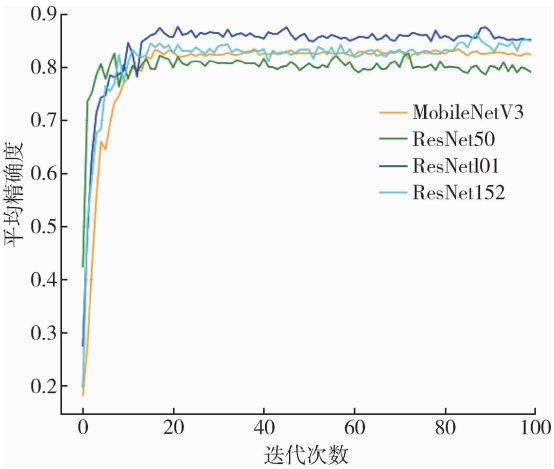


图 4　不同主干网络平均精确度曲线

3.2　消融实验

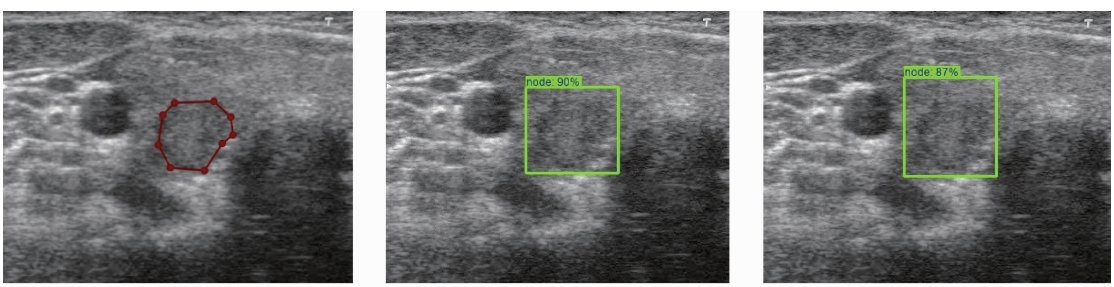
为了验证各模块对整体模型的有效性，基于 Mask R－CNN 算法，以 ResNet101 为主干网络，引入 FPN 结构并采用迁移学习策略构成不同网络模型进行消融实验[18]，见表 2。模型引入 FPN 结构并采用迁移学习策略后性能得到明显提升。

表 2　模型性能对比结果

主干网络	学习方式	平均精确度	平均召回率	平均 *F*1 分数
ResNet101	迁移学习	0.789	0.886	0.834
	非迁移学习	0.762	0.851	0.804
ResNet101＋FPN	迁移学习	0.868	0.953	0.906
	非迁移学习	0.835	0.867	0.850

3.3　模型预测结果展示

以最优网络 ResNet101 和轻型网络 MobileNetV3 为主干网络，引入 FPN 结构并采用迁移学习策略训练模型，预测甲状腺结节，见图 5。两种主干网络均能较准确地预测出结节位置，且 ResNet101 网络结构模型预测一张甲状腺结节图片耗时约 1.03 秒，MobileNetV3 网络结构模型预测一张图片耗时约 0.56 秒，速度更快；ResNet101 网络结构预测出的结节得分较高（90%），预测框标注的位置更精确。



（a）人工标注区域　（b）ResNet101网络预测区域　（c）MobileNetV3网络预测区域

图 5　模型检测效果展示

4　讨论

随着医疗技术的不断发展，甲状腺结节患者超声检查检出率越来越高[19]，如何快速准确地识别结节位置，帮助超声医师提高诊断能力和工作效率，是本实验研究的关注点。基于 Mask R－CNN 深度学习框架构建甲状腺结节检测模型，通过迁移学习策略初始化模型参数，并集成 FPN 结构，实现多尺寸特征的自适应融合，有效提升甲状腺结节检测精度。其主干网络采用轻型网络 MobileNetV3 和深层网络 ResNet 进行模型训练，通过实验可知，MobileNetV3 网络训练和预测阶段速度均更快，更适用于资源有限的移动或嵌入式设备，且其精确度与 ResNet50 网络相差无几，能够较准确地预测出甲状腺结节的位置。而深度更深的 ResNet101 和 ResNet152 网络在模型预测精确度方面表现更好，ResNet101 具有更多的层数，每一层处理的通道数相对较少，ResNet152 层数更多，有更大的容量来学习更复杂的特征，但每一层处理的通道数相对较多[20]。在实际应用中，模型性能受训练过程、优化策略和计算资源的影响，训练一个相对较浅的网络模型如 Res-

Net101，如果训练得当，即使网络不是最深的，也能达到超越更深网络的效果。

5 结语

使用深度学习算法 Mask R－CNN 构建的检测模型，可以辅助临床决策，提高对甲状腺结节患者的诊断效率。本研究共收集甲状腺结节图像 1 650 张，采用显存为 12 GB 的 GPU 进行模型训练，迭代 100 次，虽然能够准确地定位结节位置，但训练集样本数量偏少，模型训练时自我学习更新的优势没有得到充分发挥[21]，未来将继续收集大量甲状腺结节超声图像，引入注意力机制，进一步提高模型的准确性和效率。

作者贡献：王杰负责研究设计、模型构建与训练、论文撰写；王至诚负责数据收集、清洗、预处理与标注；娄帅负责实验结果分析与可视化展示；董建成负责研究设计、提供指导；曹新志负责提供指导、论文审核。
利益声明：所有作者均声明不存在利益冲突。

参考文献

1 THEODORA P, SARA A, ATHANASIOS B, et al. Thyroid nodule shape independently predicts risk of malignancy [J]. The journal of clinical endocrinology & metabolism, 2022, 107 (7): 1865－1870.
2 HAN D, IBRAHIM N, LU H. Automatic detection of thyroid nodule characteristics from 2D ultrasound images [J]. Ultrasonic imaging: an international journal, 2024, 46 (1): 41－55.
3 龚黎，李霞，方晗，等. 基于优化卷积网络 Faster R－CNN 自动检测甲状腺结节超声图像的研究 [J]. 中国超声医学杂志，2023，39 (2)：209－213.
4 ZHANG L, ZHUANG Y, HUA Z, et al. Automated location of thyroid nodules in ultrasound images with improved YOLOV3 network [J]. Journal of x－ray science and technology, 2021, 29 (1): 75－90.
5 刘明坤，张俊华，李宗桂. 改进 Mask R－CNN 的甲状腺结节超声图像分割方法 [J]. 计算机工程与应用，2022，58 (16)：219－225.
6 吴雯娟，戚琪，邓梓杨，等. 基于改进 Mask R－CNN 的多标签甲状腺结节检测模型 [J]. 南昌大学学报（理科版），2023，47 (2)：189－194.
7 郑天雷，杨娜，耿诗，等. 一种基于 Faster R－CNN 的甲状腺结节超声图像目标检测改进算法 [J]. 四川大学学报（医学版），2023，54 (5)：915－922.
8 谢紫薇，鲁大营，李志琦，等. 基于扩张卷积与注意力的甲状腺超声分割方法 [J]. 计算机技术与发展，2023，33 (3)：71－77.
9 吴星瑾，缪传鹏，李鹏，等. 基于 Mask R－CNN 的舌体分割方法 [J]. 中国卫生信息管理杂志，2021，18 (6)：843－848.
10 MA X, SUN B, TIAN C Z. Tnseg: adversarial networks with multi－scale joint loss for thyroid nodule segmentation [J]. Journal of supercomputing, 2024, 80 (5): 6093－6118.
11 LEI Y, HE X, YAO J, et al. Breast tumor segmentation in 3D automatic breast ultrasound using mask scoring R－CNN [J]. Medical physics, 2020, 48 (1): 204－214.
12 SAHIN M E, ULUTAS H, YUCE E, et al. Detection and classification of COVID－19 by using faster R－CNN and mask R－CNN on CT images [J]. Neural computing & applications, 2023, 35 (18): 13597－13611.
13 陈铭，梅雪，朱文俊，等. 一种新型 MobileUnet 网络的肺结节图像分割方法 [J]. 南京工业大学学报（自然科学版），2022，44 (1)：76－81，91.
14 MUNEEB M, FENG S, HENSCHEL A. Transfer learning for genotype－phenotype prediction using deep learning models [J]. BMC bioinformatics, 2022, 23 (1): 1－22.
15 GANG L, HAIXUAN Z, LINNING E, et al. Recognition of honeycomb lung in CT images based on improved MobileNet model [J]. Medical physics, 2021, 48 (8): 4304－4315.
16 杨靖祎，陈隆鑫，杨建凯，等. ARU－Net：基于残差注意力机制的胸腔积液图像分割模型 [J]. 医学信息学杂志，2024，45 (4)：85－90.
17 HU H, LIU A, ZHOU Q, et al. An adaptive learning method of anchor shape priors for biological cells detection and segmentation [J]. Computer methods and programs in biomedicine, 2021, 208 (3): 106260.
18 敬红燕，彭静，吴锡，等. 基于 Mask R－CNN 卷积神经网络的虹膜分割 [J]. 计算机系统应用，2023，32 (2)：83－93.
19 MARUF F, SUANJAYA M, MURTALA B, et al. Combination of thyroid ultrasound examination (TIRADS) and survivin gene mRNA expression to determine the type of thyroid nodule [J]. Bali medical journal, 2022, 11 (2): 1030－1034.
20 BALAMURUGAN A G, SRINIVASAN S, PREETHI D, et al. Robust brain tumor classification by fusion of deep learning and channel－wise attention mode approach [J]. BMC medical imaging, 2024, 24 (1): 147.
21 ZHAO D, JING Y, LIN X, et al. The value of color Doppler ultrasound in the diagnosis of thyroid nodules: a systematic review and meta－analysis [J]. Gland surgery, 2021, 10 (12): 3369－3377.