DOI: 10.13382/j.jemi.B2508188

基于轻量级改进 RT-DETR 的内窥镜息肉检测*

武 涛 魏利胜 邵自强
(安徽工程大学电气工程学院 芜湖 241000)

摘 要:针对息肉检测任务中存在息肉尺度差异显著、肠道环境复杂,以及医疗诊断设备资源有限影响检测精度的问题,提出一种基于 RT-DETR(real-time detection transformer)改进的轻量级息肉检测模型。首先,采用 FasterNet 作为 RT-DETR 模型主干网络,重构 FasterNet Block 模块,分流冗余特征的同时提升对息肉的关注度;其次,设计了新模块,在内尺度特征交互(attention-based intrascale feature interaction, AIFI)内部引入 HiLo(H-AIFI)高低频分离机制,分离局部高频细节和低频全局结构,聚焦复杂背景下的关键病灶点;最后,设计选择性边界聚合-特征金字塔网络(SBA-FPN)重校准特征融合网络替换跨尺度特征融合模块(cross-scale feature fusion module, CCFM),促进不同分辨率特征之间的双向融合,提升多尺度特征融合效果。实验结果表明,在公开的内窥镜息肉组合数据集上,与原始 RT-DETR 模型相比,改进模型 mAP@0.5 和 mAP@0.5:0.95 值分别提高 2.3%和 3.0%,参数量和计算量分别减少 44.4%、48.6%。在 Br35H 脑肿瘤数据集上,改进模型 mAP@0.5 提高 1.3%。由此可知,改进模型不仅满足息肉自动检测需求,而且满足医疗场景下泛化病灶的高精度检测。

关键词:息肉检测;RT-DETR;HiLo 高低频分离机制;FasterNet
中图分类号:TP391;TN911.73 **文献标识码**:A **国家标准学科分类代码**:520.2060

Endoscopic polyp detection based on lightweight improved RT-DETR

Wu Tao Wei Lisheng Shao Ziqiang
(School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China)

Abstract: Aiming at the problems of significant differences in polyp size, complex intestinal environment, and limited medical diagnostic equipment resources affecting detection accuracy in polyp detection tasks, a lightweight polyp detection model based on RT-DETR improvement was proposed. Firstly, FasterNet is used as the backbone network of the RT-DETR model to reconstruct the FasterNet Block module to divert redundant features while increasing attention to polyps. Secondly, the new module is designed to introduce HiLo high and low frequency separation mechanism into the attention-based intrascale feature interaction (AIFI) to separate local high frequency details and low frequency global structures, and focus on key lesions in complex backgrounds. Finally, an SBA-FPN recalibration feature fusion network is designed to replace the cross-scale feature fusion module (CCFM) to promote two-way fusion between features with different resolutions and improve the multi-scale feature fusion effect. The experimental results show that compared with the original RT-DETR model, the mAP@0.5 and mAP@0.5:0.95 values of the improved model are increased by 2.3% and 3.0% respectively, and the amount of parameters and calculations is reduced by 44.4% and 48.6% respectively. On the Br35H brain tumor dataset, the mAP@0.5 of the improved model increased by 1.3%. It can be seen that the improved model not only meets the needs of automatic polyp detection, but also meets the high-precision detection of generalized lesions in medical scenarios.

Keywords: polyp detection; RT-DETR; HiLo high-low frequency separation mechanism; FasterNet

收稿日期:2025-02-25 Received Date: 2025-02-25
*基金项目:安徽省教育厅重大项目(KJ2020ZD39)、安徽省高等学校省级质量工程项目(2023cxtd057)资助

0　引　言

结直肠癌已成为世界上第二大常见的恶性肿瘤,其主要来源于结直肠内息肉,通过内窥镜检查并切除息肉可以降低结直肠癌的发病率[1]。针对息肉检测过程中存在背景干扰且易受人工经验性、疲劳度等主观因素的影响,导致出现漏诊和误诊的问题。为提高息肉检测的准确度,保证其实时性,计算机辅助诊断系统被广泛应用于息肉检测。因此,如何更好地实现智能化的息肉检测具有重要的研究意义。

内窥镜息肉检测的研究工作,以机器学习方法和深度学习方法为主要研究方向。传统的机器学习方法通过人工提取特征并采用合适的分类器进行目标检测[2]。机器学习方法过分依赖人工经验,效率较低。随着深度学习的兴起[3],卷积神经网络(convolutional neural network, CNN)逐渐成为息肉检测任务的主流[4]。常见的深度学习息肉检测方法分为有锚点框方法和无锚点框方法。为了精确定位肿瘤并辨别息肉和腺瘤,杨昆等[5]对双阶段检测算法 Faster R-CNN 进行改进,通过数据集图像增强,采用改进网络进行训练,实现了更高的息肉检测精度。但模型依然存在对不同病灶区分度低的问题。为此,He 等[6]针对 Faster R-CNN 进一步优化,通过增加批量归一化卷积层,构建混合损失函数,采用预训练卷积结构和随机梯度下降法测试最优特征提取网络。然而,双阶段检测算法相对单阶段检测算法速度较慢,检测成本更高,因此在实际应用中单阶段检测算法更加常见。针对结肠病变检测中出现漏检、误检的问题,Gao 等[7]基于 YOLOv5 设计 YOLOv5x-CG 架构,采用 Mosaic 数据增强提高小目标息肉检出率,引入 CA 注意力机制实现病理特征有效提取,加快模型检测速度。但模型计算资源消耗较大,无法做到实时检测效果。于是设计轻量级 YOLOv7 模型,采用 CNeB 模块提取病灶区域特征,加入 SPD-Conv 缓解步进卷积引起的细粒度损伤,引入 SIoU 损失函数实现快速收敛,有效提高检测精度和效率[8]。杨昆等[9]在 YOLOv4 主干网络集成卷积块注意力模块(CBAM),剪枝特征融合层并优化网络结构,在不影响检测性能的情况下,模型计算复杂度和参数量进一步降低。

随着 Transformer[10]成为自然语言处理的基准范式,研究者们开始探索更多基于无锚点框的检测方法。DETR(detection transformer)[11]作为该领域代表,采用端到端检测方式,无需候选框生成和非极大值抑制(non-maximum suppression, NMS)等后处理步骤,简化了检测流程。Deformable DETR 作为息肉检测框架,采用不同损失函数和测试时间增强进行分析,实验表明所提方法的有效性,但模型依然存在参数量较大的问题[12]。为此,刘亚蒙等[13]通过在 RT-DETR (real-time detection transformer)的主干网络中集成轻量级 FasterNet Block 模块,引入 SimAM 注意力机制,采用 MPDIoU 损失函数加快收敛速度,缓解轻量化造成的算法精度下降。可见,具有端到端训练优势的 Transformer 架构能够有效应用于息肉检测。然而,内窥镜息肉区域存在多尺度问题,模型无法有效捕捉病变的全局特征和局部细节。此外,息肉与正常组织之间存在边界模糊,模型检测性能可能因此下降。同时,大多数模型对计算资源消耗较大,无法实现较好的实时检测效果。

针对以上研究仍然存在的问题,本文以兼顾高效混合编码器和多尺度特征融合的 RT-DETR[14]为基底模型,探究一种更加轻量化,检测效果更优的端到端息肉检测模型。

针对息肉多尺度分布且模型参数量较大,研究采用更加轻量化的 FasterNet[15]作为主干网络,重构 FasterNet Block 模块,采用多层部分卷积(partial convolution, PConv)进行特征提取,降低模型参数量的同时增强对关键病灶特征的敏感度,提高检测准确率。

针对复杂背景下无法聚焦关键病灶点,内尺度特征交互(attention-based intrascale feature interation, AIFI)模块内部引入 HiLo 高低频分离机制[16]分离特征图的高频与低频信息,弱化复杂背景对息肉检测的干扰,增强模型对小息肉和复杂特征的提取能力。

针对息肉与正常组织之间存在边界模糊,设计选择性边界聚合-特征金字塔网络(SBA-FPN)重校准特征融合网络,通过高分辨率特征与低分辨率特征之间双向融合机制,增强特征保留能力,提高息肉检测精度。

1　改进 RT-DETR 算法检测原理

RT-DETR 作为端到端目标检测框架,摒弃了传统目标检测算法的 NMS 处理,检测速度和精度均优于现有的实时检测器。研究采用 RT-DETR-r18 作为基准模型,其主要由主干网络、混合编码器、交并比(IoU)感知查询选择以及预测解码器 4 部分组成。主干网络 ResNet18 进行特征提取,捕捉图像中局部和全局信息,输出 S3、S4、S5 不同层级特征作为编码器输入;混合编码器通过 AIFI 和跨尺度特征融合(CCFM)将多尺度特征转化为结合上下文信息的丰富图像特征表示;IoU 感知查询选择通过优化查询选择机制,从编码器输出序列选择一定数量图像特征作为预测解码器初始查询对象,提高预测分析的准确性;预测解码器通过辅助预测头迭代优化对象查询,生成边框和置信度分数,将特征映射转换为最终检测结果。

尽管 RT-DETR 算法在多个具有挑战性数据集表现

出色，实现了高的实时性与准确性。然而，基于Transformer架构对计算资源和时间的需求较为庞大。因此，本文基于RT-DETR设计一种轻量级内窥镜息肉检测模型。如图1所示，首先，改进模型采用FasterNet作为主干网络，通过重构FasterNet Block模块，简化模型结构并消除干扰特征的冗余，提高息肉检测的效率与准确性；其次，由于研究选择4种公开的内窥镜息肉数据集，部分图像存在复杂背景且息肉之间具有高度相似性，模型在AIFI内部引入HiLo高低频分离机制，组成H-AIFI模块，分离特征图的高频和低频信息，弱化复杂背景对检测任务的影响，增强模型对息肉的聚焦能力；最后，设计SBA-FPN重校准特征融合网络，选择性聚合边界信息与语义信息，进一步提升多尺度特征融合效果，提高息肉的检测精度。

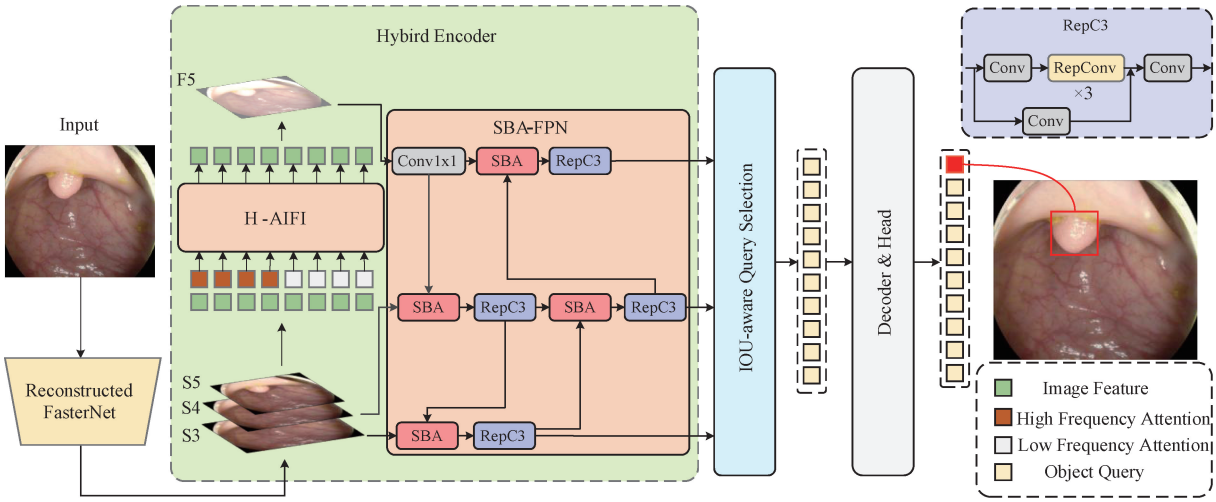


图1　轻量化RT-DETR模型整体结构

Fig. 1　The overall structure of the lightweight RT-DETR model

1.1　重构FasterNet主干网络

主干网络作为模型中的关键部分，其设计和结构直接决定模型的整体性能。RT-DETR采用ResNet系列作为主干网络，通过深度残差结构提取图像中关键空间特征信息。然而，ResNet的深层网络架构虽具有强大表征能力，但其复杂层级设计在处理高分辨率医学图像时存在计算冗余，制约了模型实际部署效率。

为减少医疗诊断设备所需的高昂计算成本，研究采用FasterNet网络作为主干网络进行图像特征提取。FasterNet通过部分卷积操作，对重要局部病灶特征具有高敏感度。同时，网络的轻量级架构使其拥有较少的计算资源消耗。具体来说，网络主要包含4个层级，每一层级前都包含嵌入层(embedding)或合并层(merging)。首先，网络通过嵌入层和合并层进行空间下采样和通道拓展。嵌入层从输入图像提取初始特征，为后续检测提供丰富的病灶信息。合并层整合不同层级的语义信息，提高模型对关键病灶点的识别能力。其次，FasterNet Block中PConv和逐点卷积(pointwise convolution，PWConv)提取局部特征并处理全局信息，增强特征表达能力，实现特征的深度融合。最后，网络通过全局平均池化、1×1卷积和全连接层进行特征转换并分类，输出最终检测结果。

针对FasterNet进行深入分析，FasterNet Block仅使用一个PConv 3×3模块，无法有效提取输入的病理特征，容易忽略特征的空间维度关系，其限制模型学习更复杂特征。因此，本文在第1个PConv 3×3后加入新的PConv 3×3扩大模型的感受野范围，进一步对已提取特征进行加工，挖掘更深层次的空间信息。在保留较多原始信息的基础上，对特征进行两次局部的细化处理，更好地识别息肉的形状和纹理。同时引入残差连接，避免多层卷积过程中丢失重要细节，促进模型学习有效特征表示。重构模型如图2所示，该结构在增加较少参数的前提下，有效融合了所有通道信息，优化了模型在更大范围内对关键病灶特征的聚焦能力，在部分卷积操作后，通过添加批归一化BN层和ReLU激活层，保持特征多样性并减少特征延迟，为高精度的息肉检测奠定基础。

FasterNet网路中PConv通过保持其他特征通道不变的情况下，只在部分输入通道应用常规卷积进行特征提取。PConv的浮点运算(floating-pointing operations，FLOPs)低于Conv，可以更好利用设备计算能力，有效提取空间特征。因此，在主干特征提取网络中集成PConv算子可以显著减少计算需求和内存访问，实现模型的轻量化并提升推理速度。

标准3×3卷积和PConv的FLOPs分别表示为：

$$F_{Conv} = H \times W \times K^2 \times C^2 \tag{1}$$

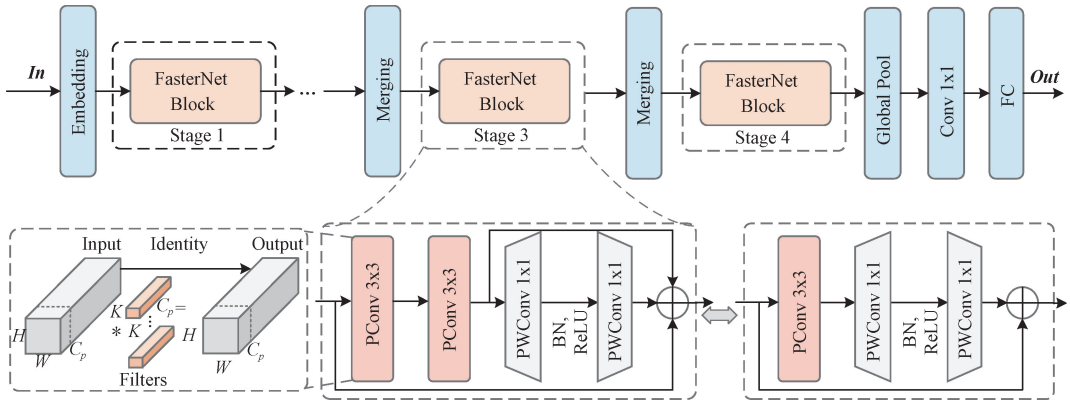


图 2　重构的 FasterNet 结构

Fig. 2　Reconstructed FasterNet structure

$$F_{PConv} = H \times W \times K^2 \times C_p^{\ 2} \tag{2}$$

式中：W 和 H 分别是输出特征图的高度和宽度；K 是卷积核的大小；C 是每个 3×3 卷积内核的通道数；C_p 是每个 PConv 卷积内核的通道数。每个 PConv 仅需对 1/4 通道进行卷积操作，较原来的 3×3 卷积参数量更小，使其浮点计算量仅为 3×3 卷积的 1/16。

1.2　H-AIFI 模块

AIFI 模块作为 RT-DETR 实现高效实时目标检测的核心组件，采用层级化特征交互机制对 S5 级特征进行优化。AIFI 中的多头自注意力（multi-head self-attention，MSA）机制通过并行处理架构，使各注意力头能够提取不同表示子空间的特征信息，捕获全局依赖关系。然而，面对内窥镜下的息肉检测任务，肠道内部环境复杂，息肉与正常组织颜色高度相似，MSA 通常只保持对所有图像块的全局关注，忽视了特征存在的不同基础频率，无法有效区分高频的息肉边缘信息和低频的图像背景信息。由于 MSA 缺乏有效的解耦策略，这种局限在检测小息肉时更加明显，同时也带来了较高的计算资源消耗。为克服这一限制，提高模型在肠道内部复杂环境下对息肉的识别效果，本研究引入 HiLo 高低频分离机制代替 MSA，设计 H-AIFI 模块，改进结构如图 3 所示。

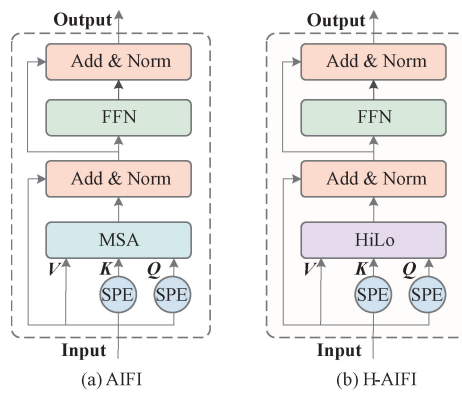


图 3　H-AIFI 模块

Fig. 3　H-AIFI module

H-AIFI 能够同时捕捉息肉图像中高频细节特征和低频结构特征，进一步优化特征交互和提取过程，从而增强模型对医疗复杂场景的处理能力。HiLo 由低频注意力（low-frequency attention，Lo-Fi）和高频注意力（high-frequency attention，Hi-Fi）两部分组成，结构如图 4 所示。首先，Hi-Fi 通过局部窗口自注意力（如 2×2 窗口）提取息肉的边缘和轮廓，Lo-Fi 对每个窗口应用平均池化捕获内窥镜图像背景等低频特征信息，并将池化后特征映射到键（key，K）和值（value，V）中，同时 Lo-Fi 的查询（query，Q）仍然来自原始特征图；其次，Lo-Fi 和 Hi-Fi 中不同局部窗口自注意力结果通过缩放点积注意力（scaled dot-product attention，SDPA）后进行拼接，拼接特征信息经过线性投影层（projection）进一步整合并转换融合；最后，高频和低频注意力输出特征再次拼接组成更加全面的特征表示。

HiLo 注意力机制将多头自注意力中相同数量的磁头分为两组，分组比为 α。其中 $(1-\alpha)N_h$ 磁头用于 Hi-Fi，其余 αN_h 磁头用于 Lo-Fi。Hi-Fi 利用局部窗口注意力提取特征图高频细节，Lo-Fi 通过平均池化简化计算，处理大尺度背景和全局特征。同时，由于每个注意力的复杂度都低于标准多头自注意力，因此确保了 HiLo 整体框架的低复杂度，并且保证了模型在 GPU 上的高吞吐量。

H-AIFI 的计算公式如下：

$$\boldsymbol{Q} = \boldsymbol{K} = \boldsymbol{V} = Flatten(\boldsymbol{S}_5) \tag{3}$$

$$H\text{-}AIFI(x) = [Hi\text{-}Fi(x) \cdot Lo\text{-}Fi(x)] \tag{4}$$

$$\boldsymbol{F}_5 = reshape(H\text{-}AIFI(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})) \tag{5}$$

式中：*Flatten* 表示将 $\boldsymbol{S}_5$ 特征展平并重新排列；$[\cdot]$ 表示 concat 拼接操作；*reshape* 表示将处理后特征图恢复原始空间维度和通道数；*Hi-Fi* 表示高频注意力操作；*Lo-Fi* 表示低频注意力操作。

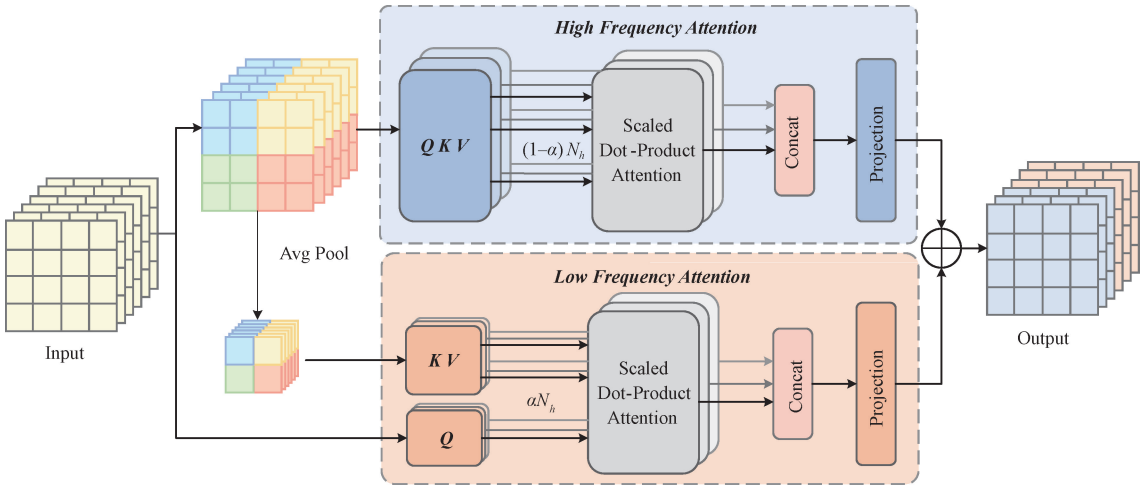


图4　HiLo注意力机制结构

Fig. 4　HiLo attention mechanism structure

1.3　SBA-FPN重校准特征融合网络

在RT-DETR的颈部网络中,CCFM能够融合不同层级特征,并迭代细化特征表达提高检测的精度与稳定性。然而,对于内窥镜息肉检测任务,息肉尺度变化较大,模型在处理多尺度特征融合时存在融合不充分,容易忽视高分辨率特征与低分辨率特征之间关系。具体来说,浅层特征层语义信息较少,主要包含息肉的边缘、纹理等细节信息,有更明显的边界和较少的失真。深层特征层蕴含丰富语义信息,更利于理解息肉全局结构,并结合上下文信息提高检测精度。因此,直接融合低级特征和高级特征可能导致信息的冗余与特征表达不一致。为了更好结合深层特征图高维信息与浅层特征图特征,参考文献[17],采用SBA模块对不同层级特征进行双向融合,挖掘深层特征与浅层特征之间互补性,进一步提高多尺度特征融合效果。

SBA模块结构如图5所示,SBA通过聚合关键边界和语义信息,增强多尺度目标检测性能,从而实现息肉边缘轮廓的亚像素级定位精度提升。具体来说,为了更精细融合不同层级特征,重校准注意力单元(re-calibration attention unit, RAU)在特征融合之前,通过自适应机制对来自编码器深层语义信息和主干网络浅层边界细节信息的两个输入特征($\boldsymbol{F}_s$,$\boldsymbol{F}_b$)进行协同重校准,构建具有判别性的多尺度特征表示,从而提取不同层级特征的互补信息。

RAU单元结构如图6所示,针对高层特征边界丢失细节与低层特征语义上下文信息不足的问题,通过差异化的重校准注意力单元处理机制,实现多层次特征的自适应优化。

为融合经RAU增强后的特征,两个重校准注意力单元输出会经过3×3卷积层进行整合。RAU函数公式

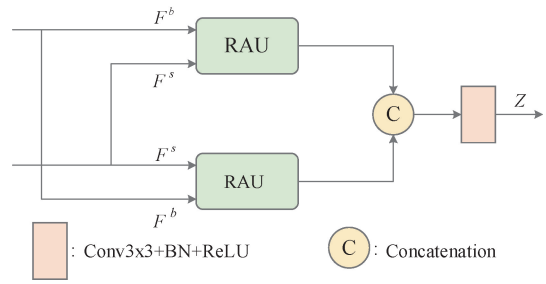


图5　SBA模块结构

Fig. 5　SBA module structure

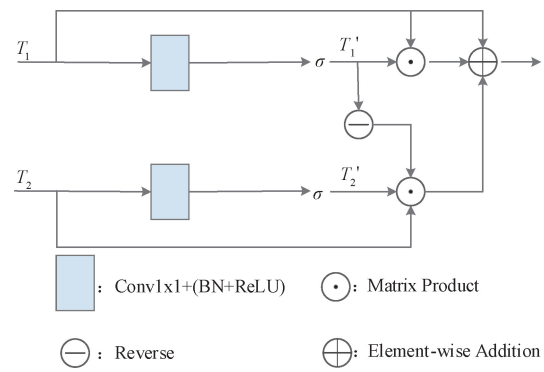


图6　RAU单元结构

Fig. 6　RAU unit structure

如下:

$$\boldsymbol{T}_1' = W_\theta(\boldsymbol{T}_1),\boldsymbol{T}_2' = W_\phi(\boldsymbol{T}_2) \tag{6}$$

$$RAU(\boldsymbol{T}_1,\boldsymbol{T}_2) = \boldsymbol{T}'_1 \odot \boldsymbol{T}_1 + \boldsymbol{T}'_2 \odot \boldsymbol{T}_2 \odot (\ominus(\boldsymbol{T}'_1)) + \boldsymbol{T}_1 \tag{7}$$

式中:$\boldsymbol{T}_1$和$\boldsymbol{T}_2$表示通过$W_\theta(\cdot)$和$W_\phi(\cdot)$处理的输入特征,其使用1×1卷积将通道维度减少到32,生成$\boldsymbol{T}'_1$和$\boldsymbol{T}'_2$;运算符号⊙表示逐元素相乘,⊖(·)表示$\boldsymbol{T}'_1$的补集。

SBA 的模块流程如下：

$$\boldsymbol{Z} = C_{3\times3}(\mathrm{Concat}(RAU(\boldsymbol{F}^s,\boldsymbol{F}^b),RAU(\boldsymbol{F}^b,\boldsymbol{F}^s))) \tag{8}$$

式中：$C_{3\times3}(\cdot)$ 表示带有批归一化 BN 层和 ReLU 激活层的 3×3 卷积；$\boldsymbol{F}^s \in R^{\frac{H}{8}\times\frac{W}{8}\times32}$ 包含编码器生成的深层语义信息；$\boldsymbol{F}^b \in R^{\frac{H}{4}\times\frac{W}{4}\times32}$ 包含主干网络的丰富边界信息；Concat(·)表示沿通道维度进行拼接，最终生成 SBA 模块的输出 $\boldsymbol{Z} \in R^{\frac{H}{4}\times\frac{W}{4}\times32}$。

SBA-FPN 重校准特征融合网络继承了 SBA 模块的双向融合优势，通过高分辨率与低分辨率特征之间的交互增强，实现更充分的信息传递与特征互补，进一步增强模型对关键病灶特征的关注度。同时，借助自适应注意力机制，模型能够根据特征图的分辨率与内容动态调整特征权重，更精准地捕捉息肉多尺度特征，进一步提升模型对复杂医疗场景的适应能力。过程表示如下：

$$F_{Output} = SBA - FPN(\{\boldsymbol{S}_3,\boldsymbol{S}_4,\boldsymbol{F}_5\}) \tag{9}$$

式中：$\boldsymbol{S}_3$ 表示浅层级特征图；$\boldsymbol{S}_4$ 表示中层级特征图；$\boldsymbol{F}_5$ 表示深层级特征图。

2　实验与分析

2.1　实验数据集

本文实验从 Kvasir-SEG[18]、CVC-ClinicDB[19]、CVC-ColonDB[20] 和 ETIS-LaribPolypDB[21] 4 种公开数据集中获取内窥镜息肉病灶图片作为研究数据集。Kvasir-SEG 数据集包含 1 000 张息肉图像以及对应的息肉标注信息。其中，息肉在大小、形状和颜色上各异。CVC-ClinicDB 数据集包含从西班牙巴塞罗那医院提供的 29 个结肠镜检查视频中提取的 612 帧息肉病灶图像，其涵盖了多种清晰度和光照条件的图像，更加接近实际临床场景。CVC-ColonDB 由来自 15 个不同视频的 380 个结肠息肉图像组成，包含不同形态的息肉，包括扁平息肉和带蒂息肉。ETIS-LaribPolypDB 数据集包含从 34 个内窥镜视频采集的 196 张息肉图像，该数据集中息肉尺寸较小，也常被用作测试集使用。

根据其标注信息通过 python 编写代码得到每张图像的边框标签用于训练。将数据集统一为病灶(lesion)一个大类，总共有 2 188 张图像，按照 7∶1∶2 随机划分为训练集、验证集和测试集。其中训练集 1 531 张，验证集 219 张，测试集 438 张。同时，在训练过程中对数据集进行数据增强，提高模型的泛化性，防止过拟合。图 7 所示为上述 4 组公开数据集的精选息肉样本和对应的边框注释标签，包含不同尺寸大小以及复杂环境(反光、肠道填充物、气泡和褶皱等)下的内窥镜息肉图像，直观呈现了每个数据集的不同风格特征。

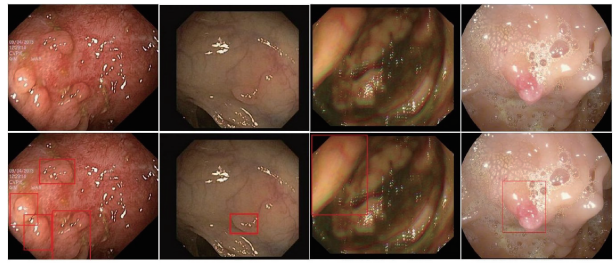


图 7　组合数据集的息肉样本图像

Fig. 7　Polyp sample images of the combined dataset

2.2　实验配置与测评标准

本文实验采用的硬件配置为 i5-12490F 处理器，NVIDIA GeForce RTX 3060 12 G 显卡；操作系统为 Windows 10，软件平台为 PyCharm；使用深度学习框架为 Pytorch2.3.1，Python 版本为 3.9，在 CUDA11.8 上进行加速训练。实验参数的设置如下：batch size 为 8；训练迭代次数为 200；采用 AdamW 作为模型优化器；采用余弦退火学习衰减方案；初始学习率为 0.000 1，权重衰减设为 0.000 1。

目标检测算法主要评价指标分为模型复杂度和检测精度。模型复杂度由计算量、参数量(Params)体现，指标数值越大代表模型复杂程度越高。模型检测精度由准确率(precision，P)、召回率(recall，R)、平均准确率(mean average precision，mAP)和 F1 分数衡量。准确率是所有预测为真样本结果中预测正确的概率，召回率是根据所有实际正样本中正确预测为正样本的比例计算，F1 分数是准确率与召回率的调和平均，定义分别如下：

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{12}$$

式中：TP 为真正例；FN 为假反例；FP 为假正例。

实验所用 mAP 指标分为 mAP@0.5 和 mAP@0.5∶0.95 两种，mAP@0.5 表示 IoU 阈值为 0.5 时的平均检测精度，mAP@0.5∶0.95 表示 IoU 阈值为 0.5 和 0.95 间的平均检测精度，计算公式如下：

$$AP = \int_0^1 P(R)\,\mathrm{d}R \tag{13}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_n \tag{14}$$

式中：n 表示内窥镜息肉类别数量。

此外，模型的实时性能通过每秒处理的图像数

(FPS)进行衡量，*FPS* 的计算公式如下：

$$FPS = \frac{framNum}{elapsedTime} \tag{15}$$

式中：*framNum* 表示固定时间内处理的图像数量；*elapsedTime* 表示处理单张图像采用的时间。

2.3　消融实验

为探究不同模块结构对内窥镜息肉检测任务的效果，获取最佳模型结构，在公开内窥镜息肉组合数据集上对不同模块进行消融实验，实验结果如表1所示。

表1　消融实验结果

Table 1　Ablation experimental results

方法	组1	组2	组3	组4	组5	组6
原始模型	√	√	√	√	√	√
FasterNet		√				
重构主干			√	√	√	√
H-AIFI				√		√
SBA-FPN					√	√
P/%	89.5	93.6	94.6	95.3	95.5	**95.6**
mAP@0.5/%	95.2	95.8	96.2	96.4	97.2	**97.5**
mAP@0.5:0.95/%	73.4	74.2	75.2	75.6	75.3	**76.4**
Params/($\times10^6$)	19.8	**10.8**	11.0	11.0	11.0	11.0
浮点计算量/GFLOPS	56.9	**28.6**	29.1	29.2	29.1	29.2
帧率/fps	64.2	84.4	80.5	89.9	80.8	**90.4**

对比实验组1和实验组2可知，将RT-DETR的主干网络替换为FasterNet进行特征提取后，模型精确率、mAP@0.5和mAP@0.5:0.95相较于原模型分别提升4.1%、0.6%和0.8%；对比实验组2和实验组3可知，在采用FasterNet主干网络基础上重构FasterNet Block进行特征提取，虽然模型的推理性能略微下降，但是模型的检测性能进一步提高，模型精确率、mAP@0.5和mAP@0.5:0.95相较原模型分别提升5.1%、1.0%和1.8%，说明重构的FasterNet能够缓解干扰特征的冗余，进一步提高息肉检测的准确性；在此基础上，采用HiLo高低频分离机制进一步挖掘特征图的深层次相关性，通过局部窗口自注意力捕捉细节，高效全局注意力处理全局结构，与原模型相比，改进模型的*P*、mAP@0.5和mAP@0.5:0.95分别提升5.8%、1.2%和0.9%，验证了HiLo解耦特征图高频和低频信息，弱化复杂背景对息肉检测任务影响的积极作用；对比实验组4和实验组5可知，采用SBA-FPN重校准特征融合网络后，模型的*P*、mAP@0.5分别提升0.2%和0.8%，说明SBA-FPN通过促进高分辨率特征与低分辨率特征之间的双向融合，进一步提高息肉检测的准确性；对比实验组1和实验组6可知，相较于原始模型，改进模型的计算量和参数量分别下降48.6%和44.4%，mAP@0.5和mAP@0.5:0.95值分别提高2.3%和3.0%，帧率达到了90.4 fps，保证了息肉检测的准确度和实时性。

为了更清晰地展示改进模型对不同内窥镜息肉的检测效果，如图8(a)所示，输入五张背景各异，不同形状息肉图像以充分验证改进模型对病灶特征提取的有效性。对比原模型(图8(b))和改进模型(图8(c))的热力图可知，原模型对息肉的聚焦不够专注，尚未提取到丰富的病灶特征，对息肉的识别精度不足。而改进模型对息肉的特征提取效果明显增强，能够捕捉病灶集中区域，更直观展示了关键病灶点，便于快速识别和定位息肉。

2.4　对比实验

将改进模型与Faster R-CNN[22]、SSD[23]、YOLOv7、YOLOv8s、YOLOv10m、YOLOv11m和RT-DETR模型在公开数据集上进行对比实验，结果如表2所示。

表2　对比实验结果

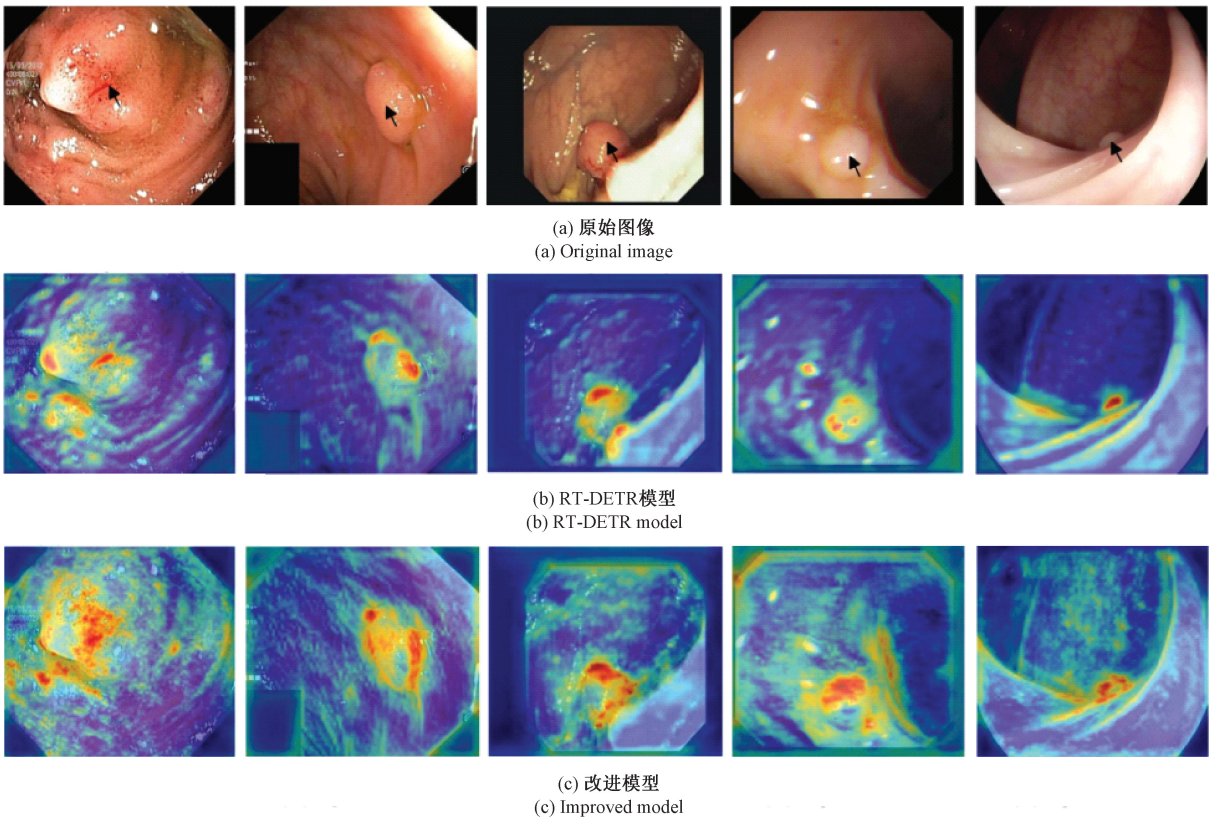
Table 2　Comparative experimental results

模型	*P*/%	mAP@0.5/%	mAP@0.5:0.95/%	浮点计算量/GFLOPs	Params/($\times10^6$)
Faster R-CNN[22]	92.1	94.6	-	-	-
SSD[23]	93.4	95.4	-	60.7	21.6
YOLOv7	93.2	94.4	72.2	106.4	37.6
YOLOv8s	93.6	96.1	73.4	**28.6**	11.1
YOLOv10m	91.6	94.3	72.6	59.1	15.3
YOLOv11m	95.0	95.7	75.2	68.2	20.0
RT-DETR	89.5	95.2	73.4	56.9	19.8
本文	**95.6**	**97.5**	**76.4**	29.2	**11.0**

由表2可知，在满足边缘设备上部署的前提下，改进模型相比其他模型检测效果更优；由于SSD和Faster R-CNN的网络层数较深，网络在特征提取过程中，随着感受野的逐层扩大，低层特征图中息肉的边缘、纹理等局部细节信息逐渐丢失，可能导致对微小息肉的检测效果不佳；虽然YOLOv8s对计算资源消耗较少，但是对息肉的检测精度还有待提高。而改进模型不仅检测精度得到提升，同时在轻量化方面也做出了优化。改进模型指标*P*、mAP@0.5、mAP@0.5:0.95和参数量达到最优，分别为95.6%、97.5%、76.4%和11.0×10^6。与原始RT-DETR模型相比，分别优化了6.1%、2.3%、3.0%和8.8×10^6，表明改进模型在确保高精度检测的同时兼顾计算效率。

图9所示为改进RT-DETR模型与其他部分主流目标检测算法的mAP@0.5曲线对比。可以看出改进RT-DETR模型的mAP@0.5曲线最高，表明改进模型的检测性能更好。

此外，为了更全面评估不同模型的检测效果，如图10所示，将改进模型与上述主流目标检测算法进行可视化对比。可以看出，其他模型在息肉不明显和复杂肠道环境中的检测精度较低，易出现漏诊和误诊，且在多尺度



(a) 原始图像
(a) Original image

(b) RT-DETR模型
(b) RT-DETR model

(c) 改进模型
(c) Improved model

图 8　特征提取效果对比

Fig. 8　Comparison of feature extraction effect

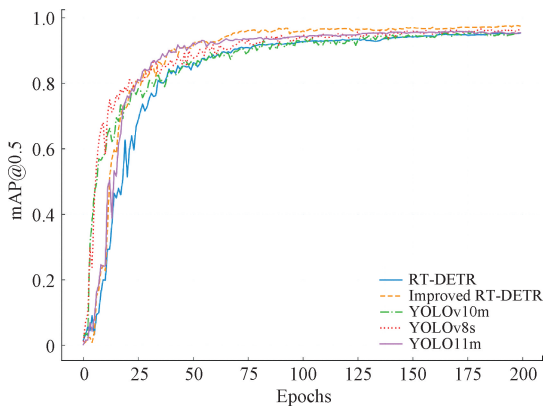


图 9　不同模型的 mAP@0.5 曲线对比

Fig. 9　Comparison of mAP@0.5 curves of different models

息肉检测方面表现不佳。而本模型能够精准定位不同环境下多个大小不一的息肉,漏诊和误诊数量较少。结果表明,所提模型在息肉检测方面具有明显优势。

2.5　跨数据集评估

由于医疗成像设备和采集规程的多样性,所获得的医学影像常常存在孤岛效应,这限制了模型的泛化能力,可能会影响模型在不同特征分布图像的检测效果。为了检验模型的泛化能力,采用 Kvasir-SEG、CVC-ClinicDB 和 CVC-ColonDB 3 种公开数据集进行模型训练,在 ETIS-LaribPolypDB 数据集上进行测试评估。此外,与当前领域的一些研究成果进行对比分析,结果如表 3 所示。

表 3　跨数据集性能评估

Table 3　Performance evaluation across datasets (%)

模型	Test Dataset	*P*	*R*	*F*1
文献[24]	ETIS-LaribPolypDB	83.2	71.6	77.0
文献[25]	ETIS-LaribPolypDB	83.2	71.6	77.0
文献[26]	ETIS-LaribPolypDB	63.9	81.7	71.7
文献[27]	ETIS-LaribPolypDB	77.8	87.5	82.4
文献[28]	ETIS-LaribPolypDB	82.3	**91.8**	**86.8**
YOLOv8s	ETIS-LaribPolypDB	74.9	69.4	72.0
YOLOv10m	ETIS-LaribPolypDB	71.6	55.1	62.0
YOLOv11m	ETIS-LaribPolypDB	80.6	70.3	74.2
RT-DETR	ETIS-LaribPolypDB	79.6	64.8	72.0
本文	ETIS-LaribPolypDB	**85.2**	71.7	78.6

由表 3 可知,模型在 ETIS-LaribPolypDB 数据集上的精确率、召回率和 F1 分数分别达到 85.2%、71.7% 和 78.6%,相比原始模型分别提高 5.6%、6.9%、6.6%。虽

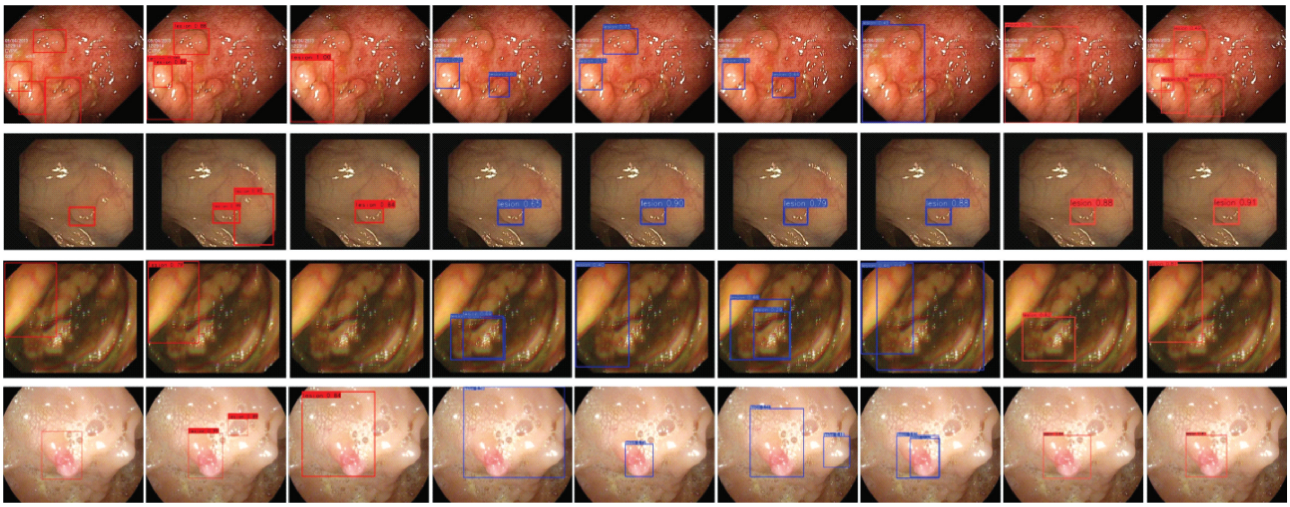


图 10　不同算法检测效果对比

Fig. 10　Comparison of detection effects of different algorithms

然改进模型相比其他研究召回率和 F1 分数略低,原因可能是训练数据集过少导致模型学习到的特征信息不够广泛,但是模型在精确率上有较明显的优势。同时,与其他主流目标检测模型相比,改进模型的各项指标优势较明显,表明改进模型有较好的检测性能和泛化能力,有效缓解了孤岛效应带来的不利影响。

2.6　泛化性实验

为了进一步验证改进模型的有效性,选取公开的 Br35H[29] 脑肿瘤检测数据集进行泛化性实验。该数据集包含 701 张图像,其中 500 张用于训练和 201 张用于测试。将改进模型与 Faster R-CNN、SSD、YOLOv7、YOLOv8s、YOLOv10m、YOLOv11m 和 RT-DETR 模型进行对比实验,实验结果如表 4 所示。

表 4　泛化性实验结果

Table 4　Generalization experimental results

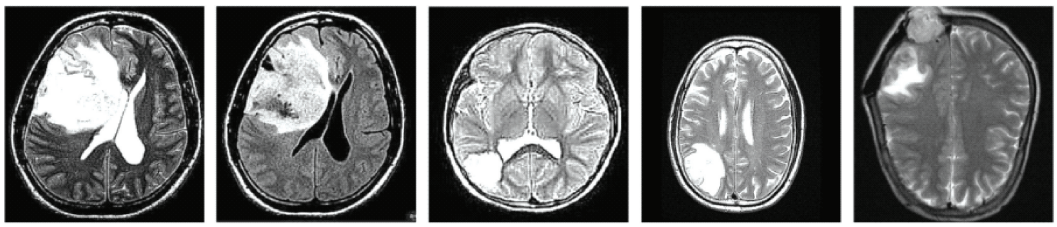
模型	P/%	mAP@0.5/%	计算量/GFLOPs	Params/($\times10^6$)
Faster R-CNN	93.0	93.9	–	–
SSD	92.5	91.5	60.7	21.6
YOLOv7	92.0	94.0	106.4	37.6
YOLOv8s	92.6	92.6	28.8	11.1
YOLOv10m	92.8	93.5	59.1	15.3
YOLOv11m	92.2	93.9	68.0	20.0
RT-DETR	91.8	93.2	56.8	19.8
本文	**93.3**	**94.5**	29.2	**11.0**

实验结果表明,改进模型 P 达到 93.3%,mAP@0.5 提升至 94.5%,与其他主流目标检测算法相比效果更优,相较于原始 RT-DETR 模型分别提高 1.5% 和 1.3%。同时,模型兼顾轻量化和实时性,验证了改进模型在病灶检测方面的优越性。

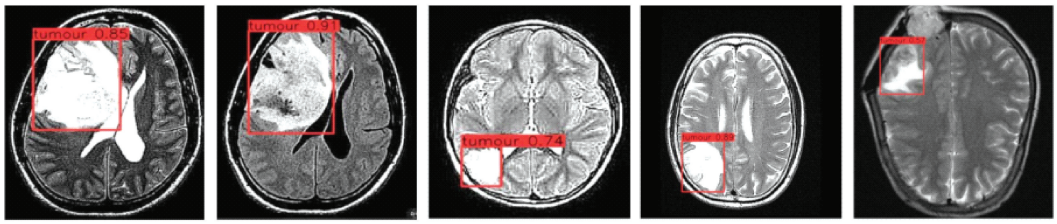
图 11 所示为 5 张不同的脑部磁共振成像(magnetic resonance imaging, MRI)。对比原模型(图 11(b))和改进模型(图 11(c))的肿瘤检测图可知,改进模型对脑肿瘤的检测优势明显,通过优化算法,模型不仅能够有效识别和定位视觉上难以辨识的息肉,还能够应用到脑肿瘤的检测任务中,提高肿瘤的检测精度。由于内窥镜图像和 MRI 属于两种不同的模态,因此模型实现了跨模态的双向提升。

3　结　论

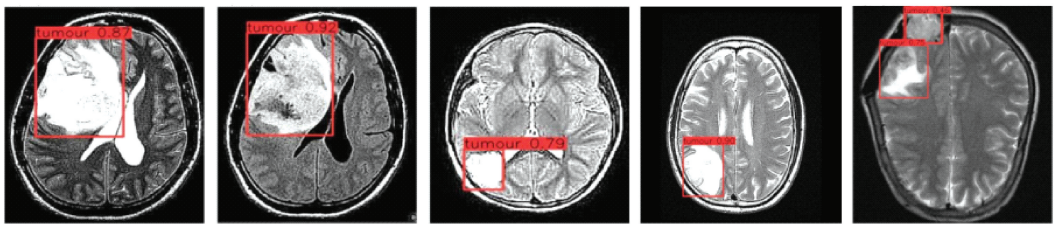
为解决内窥镜息肉检测中模型复杂度高、检测精度低的问题,研究一种轻量级改进 RT-DETR 算法。首先,通过引入 FasterNet 作为主干网络,重构 FasterNet Block 模块,有效减少了特征的冗余,进一步增强模型对病灶区域的关注度。其次,在内尺度特征交互 AIFI 内部结合 HiLo 高低频分离机制,通过区分高频和低频特征信息,使模型能够聚焦复杂背景下的关键病灶点,提升息肉检测的准确性。随后,设计 SBA-FPN 重校准特征融合网络,实现高分辨率特征与低分辨率特征之间的双向融合,增强多尺度特征融合效果,进一步提高模型的检测性能。最后,在公开数据集上进行消融实验和对比实验。实验结果表明,与其他主流算法相比,改进模型在检测性能和模型大小上均表现优异,便于实际应用中的部署。未来的研究将致力于进一步轻量化模型,同时维持其高精度的检测能力。



(a) 原始图像
(a) Original image



(b) RT-DETR模型
(b) RT-DETR model



(c) 改进模型
(c) Improved model

图 11　肿瘤检测效果对比

Fig. 11　Comparison of tumour detection effects

参考文献

[1] FREEDMAN D, BLAU Y, KATZIR L, et al. Detecting deficient coverage in colonoscopies [J]. IEEE Transactions on Medical Imaging, 2020, 39(11): 3451-3462.

[2] 赵佰亭,张晨,贾晓芬. ECC-YOLO:一种改进的钢材表面缺陷检测方法[J]. 电子测量与仪器学报,2024,38(4):108-116.
ZHAO B T, ZHANG CH, JIA X F. ECC-YOLO: An improved method for detecting surface defects in steel[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(4): 108-116.

[3] ZHU S B, WEI L S. PEDNet: A proposal enhancement dynamic network for fine-grained ship detection in optical remote sensing images[J]. IEEE Access, 2024, 12: 129813-129825.

[4] 薛钦原,胡珊珊,胡新军,等. 改进 YOLOv7 的结直肠息肉检测算法[J]. 计算机工程与应用,2025,61(1):243-251.
XUE Q Y, HU SH SH, HU X J, et al. Improved YOLOv7 algorithm for colorectal polyp detection[J]. Computer Engineering and Applications, 2025, 61(1): 243-251.

[5] 杨昆,原嘉成,高聪,等. 基于改进的 Faster R-CNN 的息肉目标检测和分类方法[J]. 河北大学学报(自然科学版),2023,43(1): 103-112.
YANG K, YUAN J CH, GAO C, et al. Object detection and classification of polyps based on improved Faster R-CNN[J]. Journal of Hebei University (Natural Science Edition), 2023, 43(1): 103-112.

[6] HE J, LIU T, LI L, et al. MFaster R-CNN for maize leaf diseases detection based on machine vision[J]. Arabian Journal for Science and Engineering, 2023, 48(2): 1437-1449.

[7] GAO J B, XIONG Q L, YU C, et al. White-light endoscopic colorectal lesion detection based on improved YOLOv5[J]. Computational and Mathematical Methods in Medicine, 2022, 2022(1): 9508004.

[8] GAO J B, LIANG J R, LI J L, et al. White-light endoscopic colorectal lesion detection based on improved

YOLOv7[J]. Biomedical Signal Processing and Control, 2024, 90: 105897.

[9] 杨昆,孙宇锋,汪世伟,等. YOLOF-CBAM:一种新的结直肠息肉实时分类与检测方法[J]. 电子测量技术, 2023,46(16):138-147.
YANG K, SUN Y F, WANG SH W, et al. YOLOF-CBAM: A new real-time classification and detection method for colorectal polyps[J]. Electric Measurement Technology, 2024, 46(16): 138-147.

[10] ASHISH V. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 6000-6010.

[11] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.

[12] MOHAMMAD F, SHARMA V, DAS P K. Polyp detection in colonoscopy images using improved deformable DETR[C]. IEEE Region 10 International Conference TENCON, 2022: 1-6.

[13] 刘亚蒙,赵友全,孙振涛,等. 构建改进 RT-DETR 算法检测隐形眼镜环状波纹缺陷[J]. 电子测量与仪器学报,2024,38(5):1-9.
LIU Y M, ZHAO Y Q, SUN ZH T, et al. Constructing an enhanced RT-DETR algorithm for detecting annular ripple defects in contact lenses[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(5): 1-9.

[14] ZHAO Y, LV W Y, XU S L, et al. Detrs beat YOLOs on real-time object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.

[15] CHEN J R, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.

[16] PAN Z Z, CAI J F, ZHUANG B H. Fast vision transformers with hilo attention[J]. Advances in Neural Information Processing Systems, 2022, 35: 14541-14554.

[17] TANG F, XU Z, HUANG Q, et al. DuAT: Dual-aggregation transformer network for medical image segmentation[C]. Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Singapore: Springer Nature Singapore, 2023: 343-356.

[18] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-SEG: A segmented Polyp dataset[C]. 26th International Conference on MultiMedia Modeling. Springer, 2019: 451-462.

[19] BERNAL J, SÁNCHEZ F J, FERNÁNDEZ-ESPARRACH G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians[J]. Computerized Medical Imaging and Graphics, 2015, 43: 99-111.

[20] TAJBAKHSH N, GURUDU S R, LIANG J M. Automated polyp detection in colonoscopy videos using shape and context information[J]. IEEE Transactions on Medical Imaging, 2015, 35(2): 630-644.

[21] SILVA J, HISTACE A, ROMAIN O, et al. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer[J]. International Journal of Computer Assisted Radiology and Surgery, 2014, 9: 283-293.

[22] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.

[23] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Proceedings of the 14th European Conference on Computer Vision. Cham: Springrer, 2016: 21-37.

[24] SORNAPUDI S, MENG F, YI S. Region-based automated localization of colonoscopy and wireless capsule endoscopy polyps[J]. Applied Sciences, 2019, 9(12): 2404.

[25] XU J W, ZHAO R, YU Y Z, et al. Real-time automatic polyp detection in colonoscopy using feature enhancement module and spatiotemporal similarity correlation unit[J]. Biomedical Signal Processing and Control, 2021, 66: 102503.

[26] JIA X, MAI X C, CUI Y, et al. Automatic polyp recognition in colonoscopy images using deep learning and two-stage pyramidal feature prediction[J]. IEEE Transactions on Automation Science and Engineering, 2020, 17(3): 1570-1584.

[27] LIU X Y, GUO X Q, LIU Y J, et al. Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images[J]. Medical Image Analysis, 2021, 71: 102052.

[28] YANG K, CHANG S L, TIAN Z X, et al. Automatic polyp detection and segmentation using shuffle efficient channel attention network[J]. Alexandria Engineering

Journal, 2022, 61(1): 917-926.

[29] AMRAN G A, ALSHARAM M S, BLAJAM A O A, et al. Brain tumor classification and detection using hybrid deep tumor network [J]. Electronics, 2022, 11(21): 3457.

作者简介

武涛,2023 年于安徽工程大学获得学士学位,现为安徽工程大学硕士研究生,主要研究方向为深度学习与医学图像处理。E-mail: 2230342267@ stu. ahpu. edu. cn

Wu Tao received his B. Sc. degree from Anhui Polytechnic University in 2023. Now he is a M. Sc. candidate at Anhui Polytechnic University. His main research interests include deep learning and medical image processing.

魏利胜(通信作者),2001 年于安徽工程大学获得学士学位,2004 年于中国航天科工集团 061 基地获得硕士学位,2009 年于上海大学获得博士学位,现为安徽工程大学教授,硕士生导师,主要研究方向为图像识别与应用、智能化网路控制系统与仿真。E-mail: lshwei_11@ 163. com

Wei Lisheng (Corresponding author) received his B. Sc. degree from Anhui Polytechnic University in 2001, M. Sc. degree from China Aerospace Science and Industry Corporation 061 Base in 2004, and Ph. D. degree from Shanghai University in 2009. Now he is a professor and master's supervisor at Anhui Polytechnic University. His main research interests include image recognition and application, intelligent network control system and simulation.