

一种基于改进型YOLOv5s的结直肠息肉检测算法QB-YOLO

张子健, 徐建宇, 杨 欢

(华南师范大学软件学院, 广东 佛山 528225)

摘 要: 在医学影像领域, 结直肠息肉的早期检测对预防结直肠癌等疾病至关重要。在实际医疗操作中, 自动化检测结直肠息肉的准确率受制于多种特殊条件。为此, 提出基于改进型YOLOv5s的结直肠息肉检测模型QB-YOLO。首先, 在原骨干网络中引入一种局部上下文信息增强模块——通道注意力机制(CAM), 替换原模型中的空间金字塔池化SPPF模块, 以增强模型对结直肠息肉的目标关注度; 其次, 在骨干网络中添加大核分离卷积注意力模块(LSKA模块), 加强模型捕捉结直肠息肉图像中局部细节的能力; 最后, 将软非极大值抑制Soft-NMS引入模型以应对部分结直肠息肉可能会密集分布的情况, 使模型更高效地处理重叠目标和密集目标。实验表明, 改进后模型的准确率、召回率、平均精度相较于原始模型提升4.1%、8.5%、3.9%。

关键词: YOLOv5s; 结直肠息肉; 目标检测; 局部上下文信息增强; 大核分离卷积注意力; 软非极大值抑制

DOI: 10.11907/rjtk.241413

开放科学(资源服务)标识码(OSID):

中图分类号: TP391.4

文献标识码: A

文章编号: 1672-7800(2025)006-0041-08



QB-YOLO: Colorectal Polyp Detection Algorithm Based on Improved YOLOv5s

ZHANG Zijian, XU Jianyu, YANG Huan

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: In the field of medical imaging, early detection of colorectal polyps is crucial for preventing diseases such as colorectal cancer. In practical medical operations, the accuracy of automated detection of colorectal polyps is limited by many special conditions. Therefore, a colon polyp detection model QB-YOLO based on improved YOLOv5s is proposed. Firstly, a local context information enhancement module (CAM module) is introduced into the original backbone network to replace the spatial pyramid pooling SPPF module in the original model, in order to enhance the model's target attention to colorectal polyps; Secondly, adding a Large Kernel Separation Convolutional Attention Module (LSKA module) to the backbone network enhances the model's ability to capture local details in colorectal polyp images; Finally, Soft NMS with soft non maximum suppression is introduced into the model to address the possibility of dense distribution of some colorectal polyps, making the model more efficient in handling overlapping and dense targets. The experiment shows that the accuracy, recall, and average precision of the improved model have increased by 4.1%, 8.5%, and 3.9% compared to the original model.

Key Words: YOLOv5s; colorectal polyps; object detection; local context augmentation; large separable kernel attention; Soft-NMS

0 引言

结直肠息肉是一种常见的肠道疾病, 通常指在结肠与直肠内形成的小突出物或肿块。尽管, 许多结直肠息肉为

良性, 但一些结直肠息肉有可能会继续恶化成结直肠癌, 因此早期发现和治疗结肠息肉至关重要^[1]。据统计, 结直肠癌是世界范围内最常见的癌症之一, 其发病率与致死率在人类癌症中排名前三, 结直肠息肉是结直肠癌的前期病变, 因此研究出一种精准可靠的结直肠息肉检测方法, 对

收稿日期: 2024-04-30

扫描二维码阅读全文:



基金项目: 广东省基础与应用基础研究基金项目(2020A1515110783)

作者简介: 张子健(2001-), 男, 华南师范大学软件学院硕士研究生, 研究方向为目标检测; 徐建宇(2001-), 男, 华南师范大学软件学院硕士研究生, 研究方向为目标检测; 杨欢(1986-), 男, 博士, CCF会员, 华南师范大学软件学院副研究员、硕士生导师, 研究方向为目标检测。本文通讯作者: 杨欢。

提升患者生存率和生活质量具有重要意义^[2,3]。

然而,高效地实现结直肠息肉的目标检测任务并不容易,具体包括:①通常需要内窥镜深入人体结直肠道内部,体内环境相对阴暗且结直肠息肉的大小和形状也可能存在较大差异;②结直肠息肉有可能在图像中密集分布,意味着模型不仅要能准确检测大目标,还要能处理密集微小目标;③由于人体结直肠道内生理环境复杂,镜检下可能会出现气泡、食物残渣等杂质影响模型识别精度;④不同患者的结直肠内部生理状态不完全相同,结直肠息肉位置与大小存在随机性,即使是同一患者的相同病灶也可能因内窥镜所处角度及位置不同在图像上发生形变;⑤结直肠息肉附着在消化道内壁,通常与内壁颜色差异较小,对应图像的边界较为模糊。综上,各种因素无疑加大了结直肠镜检下息肉检测的难度。

1 相关工作

当前,基于深度学习的通用目标检测算法主要分为两阶段检测和单阶段检测。两阶段检测方法包括 R-CNN 系列和 Faster R-CNN 等,首次提出先提取候选区域,再对候选区域进行分类和边界框回归^[4,5]。单阶段检测方法包括 YOLO 系列、SSD 等,将目标检测任务视为回归问题,通过神经网络直接预测目标类别和位置^[6-10]。在应用传统计算机视觉方法解决结直肠道息肉检测问题时,通常会采用一系列经典的图像处理和分析技术。在利用特征提取捕获病变区域的特征方面,可利用灰度共生矩阵(Gray Level co-occurrence Matrix, GLCM)描述病变区域的纹理特征,使用 Gabor 滤波器捕获病变区域的频域信息;在分类方面,通过设计合适的特征描述符与选择合适的分类器对病变进行分类,常见的分类器包括支持向量机、决策树、人工神经网络等,可从提取的特征中学习并建立病变与正常组织之间的分类模型^[11-13]。分类完成后还需进行后处理和结果分析,并与医生的人工诊断结果进行比较和验证。

然而,应用传统计算机视觉方法检测结直肠息肉时,准确率依然会受到一些限制。由于结直肠息肉形状、大小和颜色具有多样性,传统方法可能无法有效地捕捉到这些复杂的特征。例如,基于阈值分割或边缘检测的方法可能对不同光照条件或组织结构变化较为敏感,导致分割结果不稳定或不准确。同时,传统方法往往需要手工设计特征提取器和分类器,受限于专家经验和先验知识,可能无法充分挖掘数据中的潜在信息。此外,传统方法通常基于手工设计的规则和算法进行操作,可能在处理大规模数据和复杂情况下存在计算复杂度高、运行速度慢等问题,无疑都限制了其在结直肠息肉领域的应用。

近年来,研究人员开始转向使用深度学习和端到端的卷积神经网络等方法,以期更好地解决目标检测问题^[14,15]。端到端的卷积神经网络能直接从原始数据中学

习特征表示和训练目标识别模型,避免了手工设计特征提取器和分类器的过程,能更好地捕获图像中的关键特征和信息,准确定位息肉位置以辅助医师在内窥镜下执行切除操作。同时,卷积神经网络具有良好的层级结构和参数共享机制,能有效处理图像中的空间关系和局部模式,通过反向传播算法进行端到端的训练还能提升模型性能和泛化能力。范珊珊等^[16]使用深度学习框架,结合数据增强技术与迁移学习策略识别小肠息肉,比较了不同深度卷积神经网络模型的效果,在提升识别率的同时降低了训练时间和硬件要求。孙雪华等^[17]提出一种基于 Faster R-CNN 的肠道息肉检测方法,结合中值滤波的非线性滤波特性和数据增强方法,利用 ResNet-50 提取多尺度特征,实现了高准确率和召回率的息肉检测。Li 等^[18]提出一种改进的基于深度神经网络的结直肠镜息肉检测算法,采用改进的 Faster R-CNN 并融合了不同层级的特征图,实验表明该算法在 CVC-ClinicDB 数据集上准确率为 97.13%。任莉莉等^[19]提出一种新型的胃肠息肉分割网络 GLIA-Net,在 U-Net 基础上加入全局与局部交互式注意力融合模块提取丰富的多尺度语义信息,提升了胃肠息肉分割的精确率。

针对目标检测领域,Lau 等^[20]提出一种大核分离卷积注意力(Large Separable Kernel Attention, LSKA)模块,旨在提升模型提取输入图像中关键特征的能力。在引入局部敏感大内核及一维可分离卷积内核后,LSKA 模块能更有效地捕获结直肠内窥镜图像的全局和局部信息,从而提升检测准确率。Xiao 等^[21]提出一种局部上下文信息增强模块——通道注意力机制(Channel Attention Mechanism, CAM),提取模型对小目标图像的特征能力,并通过自适应方式对不同位置的特征进行加权,以确保模型更关注小目标的关键特征,从而提升目标检测任务的性能。Bodla 等^[22]提出一种基于软非极大值抑制(Soft Non Maximum Suppression, Soft-NMS)的目标检测方法,在 NMS 过程中引入衰减函数,根据检测框之间的重叠度来逐渐降低检测框得分,而非直接移除置零,既更好地保留了密集目标,又提升了模型检测召回率和定位精度。综上,本文以 YOLOv5s 为基础网络模型,引入 CAM 模块、LSKA 模块、Soft-NMS 模块加强原模型捕捉结直肠息肉图像中局部细节和特征的能力,从而提升模型对不同结直肠息肉目标的检测效果。

2 YOLOv5s 网络结构

YOLOv5 是一种高性能、轻量级、基于深度学习的目标检测模型,具有实时检测、高精度的特点。YOLOv5 具有 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 等版本,本文以 YOLOv5s 6.0 模型为基础结构进行改进,具体结构如图 1 所示。由此可见,YOLOv5s 主要包含 Conv、C3、SPPF、Concat 和 Upsample 模块,主干网络结构包括一系列卷积层、C3 模块的堆叠及最后的空间金字塔池化模块。

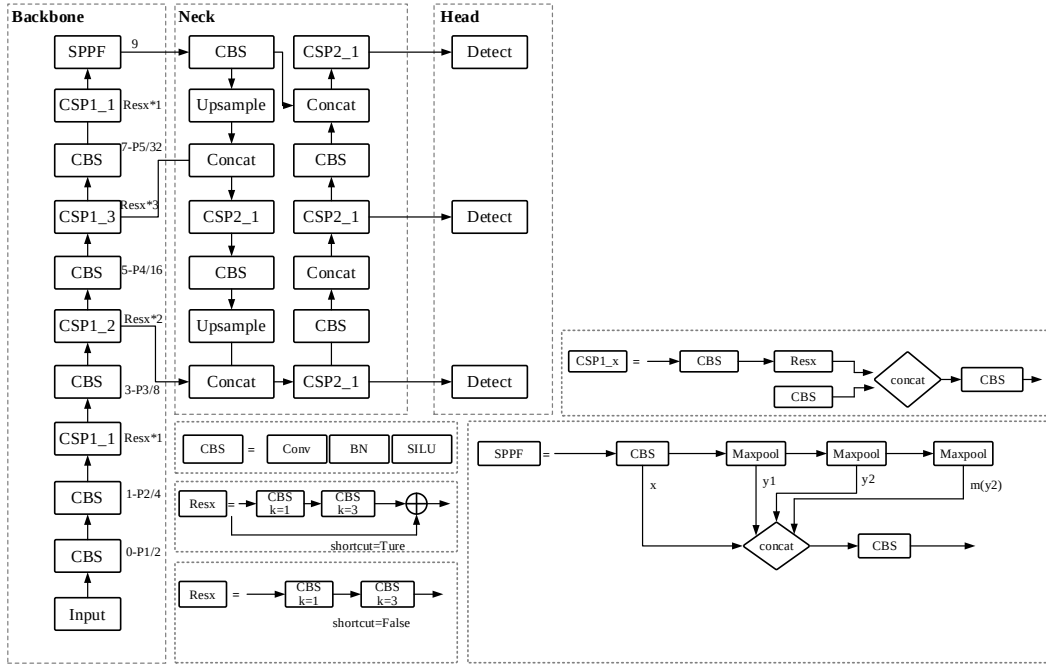


Fig. 1 YOLOv5s structure

图1 YOLOv5s 结构

Backbone 主干部分由 Conv、C3 和 SPPF 等 3 个模块组成。其中,Conv 表示卷积层,C3 模块由标准卷积模块和瓶颈模块组成。首先,从输入开始通过一系列卷积操作逐步提取图像特征;其次,在模型中穿插多个 C3 模块,每个块的输出通道数为 128,用于进一步提取特征;最后,通过 SPPF 模块对特征图进行空间金字塔池化,以获取多尺度的特征表示。

YOLOv5s 的 Neck 部分由主要包含 4 个特征融合层,通过卷积和上采样等操作,融合 Backbone 和 Head 的不同层级特征图。在第 10 层,Conv 模块通过卷积操作保持特征图尺寸不变,但通道数减半。第 11 层使用 Upsample 模块对提取到的特征图片进行上采样操作,使其恢复到 16 倍。第 12 层的 Concat 模块将第 6 层与第 11 层的特征图融合,通道数为两者相加之和。第 13 层的 C3 模块用于进一步提取图片的特征信息,这一循环操作持续进行,从第 10~13 层中得到第 14~17 层的特征图。在第 18 层 Conv 模块保持通道数不变,特征图尺寸减半,此时将进行 16 倍下采样操作。第 19 层的 Concat 模块融合了第 14 层与第 18 层的特征图。第 20 层使用 C3 模块再次提取特征信息,循环重复直至产生第 21~23 层的特征图,此时将进行 32 倍下采样尺寸。此外,YOLOv5s 中的 Head 部分由 Detect 模块组成,先将第 17、20 和 23 层的特征图作为输入,再对这些特征图进行处理和整合。

YOLOv5s 结构改进为 QB-YOLO 结构,如图 2 所示。

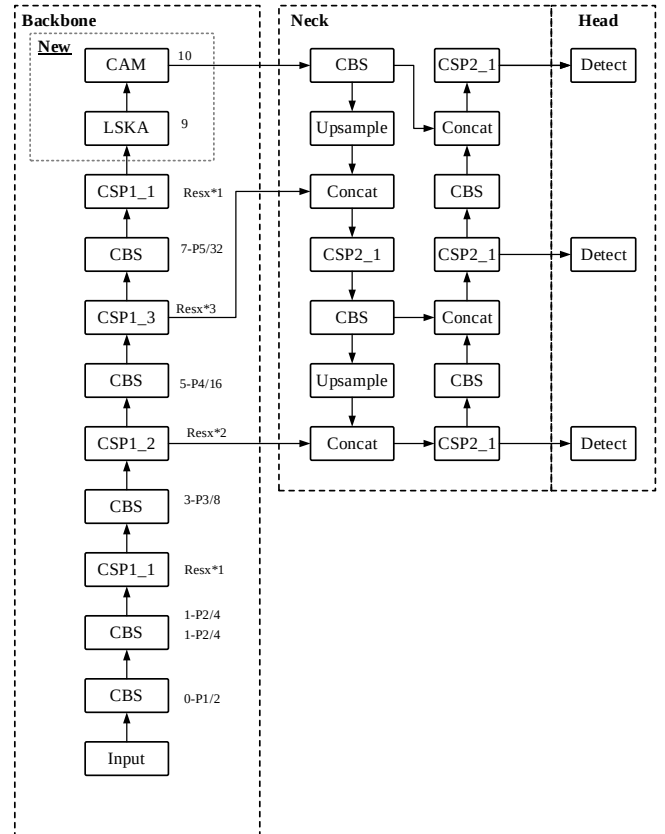


Fig. 2 QB-YOLO structure

图2 QB-YOLO 结构

3 本文算法

3.1 算法结构

本文针对结直肠镜下息肉检测面临的实际问题,将

QB-YOLO 的改进之处包括:

(1)引入 LSKA 模块。在 Backbone 的第 8 层,即 C3 模块后引入 LSKA 模块。LSKA 模块将 2D 卷积内核分解为级

联水平和垂直 1-D 内核,以缓解传统大内核设计所带来的计算量和内存开销较大的问题。该设计不仅扩大了模型感受野,还有效捕获了目标物体的上下文信息。同时,引入 LSKA 模块能使 QB-YOLO 更好地识别结直肠息肉图像中的微小目标和细微特征,使模型在检测任务中能更精确定位、识别目标。

(2) 更换原 SPPF 模块。在 Backbone 第 9 层引入 CAM 模块替换原 SPPF 模块,以显著提升模型检测部分微小结直肠息肉目标的速度。CAM 模块通过自适应的通道注意力权重,引导特征学习更好地捕获图像全局和局部信息。这种自适应加权方式有助于模型更准确地定位结直肠息肉,并提升模型目标检测精度和鲁棒性。

(3) 添加 Soft-NMS。引入 Soft-NMS 优化模型处理检测框之间的重叠关系,有助于提升模型在训练过程中的稳定性与泛化能力。相较于传统 NMS 算法,Soft-NMS 能降低与已选框重叠的其他检测框的得分,以提升模型对密集目标和重叠目标的检测能力,使 QB-YOLO 在面对复杂场景和密集结直肠息肉图像时检测准确性更高。

3.2 大核分离卷积注意力模块

为了增强模型检测能力,在原模型中引入 LSKA 模块,允许在注意力模块中使用较大的卷积核。LSKA 模块将 2D 卷积内核分解为级联水平和垂直 1-D 内核,以覆盖更大的感受野,从而有效地捕获目标物体的上下文信息^[20]。具体地,LSKA 模块由 LKA 模块改良而来,具体结构如图 3 所示。

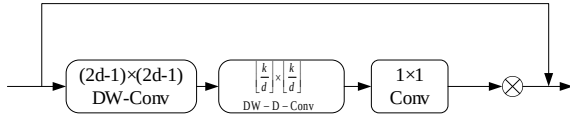


Fig. 3 LKA module structure

图 3 LKA 模块结构

LKA 模块的输出流程如式(1)一式(3)所示

$$Z^C = \sum_{H,W} W_{k \times k}^C * F^C \quad (1)$$

$$A^C = W_{1 \times 1} * Z^C \quad (2)$$

$$\bar{F}^C = A^C \otimes F^C \quad (3)$$

式中: $*$ 代表卷积; $\otimes$ 代表哈达玛积; Z^C 为将大小为 $k \times k$ 的卷积核 W 与输入特征图 F 进行卷积获得的深度卷积的输出; F 中每个通道 C 均与卷积核 W 中的相应通道进行卷积; k 表示卷积核 W 的最大感受野,通过 1×1 卷积获得注意力图 A^C ;LKA 的 $\bar{F}^C$ 为注意力图 A^C 与特征提取图 F^C 的哈达玛积。

然而,LKA 模块内部存在的一些大核深度卷积计算成本较高。本文将深度卷积分解为具有较小内核的深度卷积与具有超大内核的扩展深度卷积,经过处理后的 LKA 模块输出为:

$$\bar{Z}^C = \sum_{H,W} W_{(2d-1) \times (2d-1)}^C * F^C \quad (4)$$

$$Z^C = \sum_{H,W} W_{\left[\frac{k}{d}\right] \times \left[\frac{k}{d}\right]}^C * \bar{Z}^C \quad (5)$$

$$A^C = W_{1 \times 1} * Z^C \quad (6)$$

$$\bar{F}^C = A^C \otimes F^C \quad (7)$$

进一步,将大核深度卷积分解为小核深度卷积,将具有相同大核的膨胀深度卷积进行分解,以减轻深度卷积单独使用大核卷积造成计算成本呈平方增加的问题。同时,将深度卷积和深度扩展卷积的二维权重核拆分为两个级联的一维可分离权重核,得到 LSKA 模块。由文献[23]可知,LSKA 的输出如式(8)一式(11)所示。经过改进后的 LSKA 的结构如图 4 所示。

$$\bar{Z}^C = \sum_{H,W} W^C_{(2d-1) \times 1} * \left(\sum_{H,W} W^C_{1 \times (2d-1)} * F^C \right) \quad (8)$$

$$Z^C = \sum_{H,W} W_{\left[\frac{k}{d}\right] \times 1}^C * \left(\sum_{H,W} W_{1 \times \left[\frac{k}{d}\right]}^C * \bar{Z}^C \right) \quad (9)$$

$$A^C = W_{1 \times 1} * Z^C \quad (10)$$

$$\bar{F}^C = A^C \otimes F^C \quad (11)$$

式中: d 为指空洞率; $\bar{Z}^C$ 表示使用卷积核大小为 $(2d-1) \times (2d-1)$ 的深度卷积输出,该深度卷积不仅能捕获局部空间信息,还能补偿空洞深度卷积所带来的格栅效应;空洞深度卷积则负责捕获深度卷积输出 $\bar{Z}^C$ 的全局空间信息。

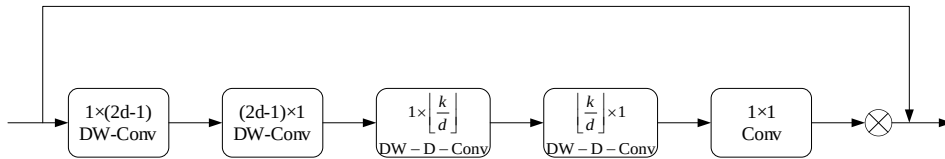


Fig. 4 Structure of the LSKA module

图 4 LSKA 模块的结构

3.3 CAM

为了增强模型检测密集结直肠息肉小目标的能力,在 YOLOV5s 第 9 层将 SPPF 模块更换为 CAM 模块。针对传统小目标检测存在的问题,文献[21]提出的 CAM 模块可生成自适应的通道注意力权重,进而引导特征在更重要的方向进行学习。首先,通过该模块在空间维度输入压缩特征

图,聚合全局特征信息,并结合自适应的平均池化和最大池化获取更精细的全局特征;其次,通过权重进行空间维度上的加和,经过 sigmoid 操作生成通道自适应权重,引导特征学习更重要的通道信息。CAM 模块能使模型更好地关注重要的通道信息,提升了模型的特征表征能力与目标检测性能。

CAM 模块的结构如图 5 所示。由此可见,该特征结构分别以 1、3 和 5 的比例,通过扩展卷积对图像进行处理,并融合不同感受野特征来获取语境信息。首先,在 C5 上使用不同的扩张卷积率进行扩张卷积,即在经过 3 个卷积核大小为 3×3,扩张卷积率为 1、3、5 的卷积;其次,融合 3 个卷积的输出得到不同感受野下的特征,既丰富了上下文信息,又获得了更好的特征提取效果。

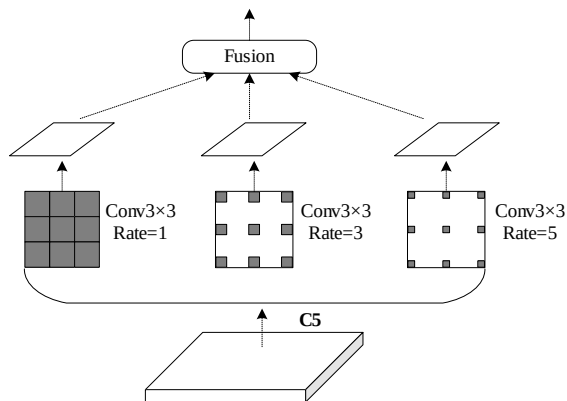


Fig. 5 CAM module structure

图 5 CAM 模块结构

3.4 Soft-NMS

非极大值抑制是目标检测的重要组成部分,原 YOLOv5s 模型针对重叠的待处理框,若 IoU 值大于设定阈值 N_t 则 NMS 算法将其剔除。然而,这种二元抑制方式可能会导致一些真实目标被误判为负例,特别是当检测框之间有部分重叠的情况^[22]。为此,本文利用 Soft-NMS 降低与已选框重叠的其他检测框的得分,而非直接将其置零,从而提升模型对密集目标的检测精度和检测效率。在传统 NMS 算法中,修剪步骤可写为重评分函数:

$$S_i = \begin{cases} S_i & \text{IoU}(M, b_i) < N_t \\ 0 & \text{IoU}(M, b_i) \geq N_t \end{cases} \quad (12)$$

式中: S_i 指各个建议框的得分; M 为当前得分最高框; N_t 指手动设置的阈值; b_i 指待处理框。

对于已选框重叠超过预定义阈值的其他检测框,算法将其得分归零,以便在后续处理进行剔除,只保留最相关的检测结果。当存在重叠框时,使用高 NMS 就会让模型在重叠阈值范围内的精度下降,同时有可能出现假阳性。为此,在 Soft-NMS 中引入高斯惩罚函数更新剪枝步骤(见式 13)。改进后的修剪步骤的重评分函数如式(14)所示。

$$S_i = S_{ie} \cdot \frac{\text{IoU}(M, b_i)^2}{\sigma}, \forall b_i \notin D \quad (13)$$

$$S_i = \begin{cases} S_i & \text{IoU}(M, b_i) < N_t \\ S_i \cdot (1 - \text{IoU}(M, b_i)) & \text{IoU}(M, b_i) \geq N_t \end{cases} \quad (14)$$

4 实验结果与分析

4.1 实验环境

本文所有实验均在 AutoDL 云平台上进行,实验环境

为 Pytorch,操作系统为 Linux 系统, GPU 型号为 NVIDIA RTX 4090,运行内存为 24 GB,编程语言为 Python,解释器为 Pycharm, CUDA 版本为 11.8。参数设置方面, Batch_size=8, Epochs=100,其他参数保持默认值。

4.2 实验数据集与预处理

本文数据集为 2020 年发布的 Kvasir-SEG 数据集,该数据集是一个针对结直肠息肉像素级分割的内窥镜分割数据集,共有 1 000 张不同的结直肠息肉图像^[24]。实验开始前,采用 labellmg 软件将数据集全部重新标注为 YOLO 格式,以满足目标检测需求。在标注时,用矩形框标记息肉病灶区域,命名为“polyp”,标签序号为 0。处理后的数据集中,训练集、验证集分别为 800、200 张,部分标注后的数据如图 6 所示。

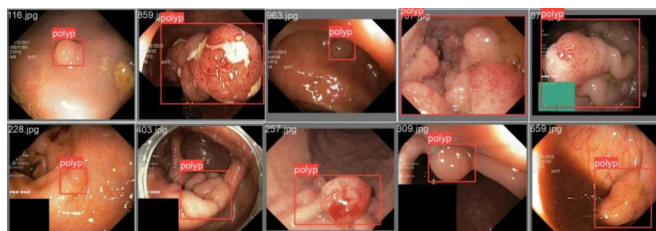


Fig. 6 Selected dataset images after labeling

图 6 经过标记后的部分数据集图像

4.3 实验评价指标

为了测试 QB-YOLO 的检测能力,采用平均精度(mAP)、召回率(Recall)、准确率(Precision)来综合评估其在结直肠息肉检测中的性能。召回率是指模型正确检测正样本数量与数据集中所有正样本数量的比例,即模型的查全率。召回率越高,表示模型漏检的正样本越少。

$$\text{precision} = \frac{TP}{TP + FP} \quad (15)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (16)$$

$$\text{mAP} = \frac{\sum_{i=1}^k \text{AP}_i}{k} \quad (17)$$

其中, TP(True Positive)代表真正例,即模型正确检测到的目标; FP(False Positive)代表假正例,即模型错误将背景或非目标区域误判为目标区域; FN(False Negative)代表假负例,即模型未能正确检测到的目标。

4.4 比较实验

为了客观评价 QB-YOLO 模型的性能,将其与 YOLOv3-tiny、YOLOv7、RT-DETR^[25]、YOLOv8、YOLOv9^[26] 模型进行比较。所有实验均在相同环境下进行 100 次迭代,实验结果如表 1 所示。由此可知,该模型的精度及准确率均超过上述模型,具有优良的性能。

4.5 消融实验

为了验证 QB-YOLO 各改进模块的有效性,共设计了 8 组消融实验,如表 2 所示。其中,模型 E0 为使用原 YOLOv5s 作为消融实验的基准模型;模型 E1 为只将 SPPF 模块替换为 CAM 模块;模型 E2 为只在第 8 层后添加 LSKA 模

Table 1 Comparative experimental results**表1 比较实验结果**

(%)

Models	Precision	Recall	mAP
YOLOv3-tiny	75.9	79.9	80.0
YOLOv7	75.0	84.0	81.6
YOLOv8n	87.5	84.2	89.3
RT-DETR	86.9	84.4	87.5
YOLOv9-c	90.9	87.0	92.5
YOLOv9-e	90.2	85.5	91.2
QB-YOLO	91.0	94.0	93.4

Table 2 Results of ablation experiments**表2 消融实验结果**

Models	CAM	LSKA	Soft-NMS	Precision/%	Recall/%	Map/%
E0				87.1	85.5	89.5
E1	√			90.3	86.4	90.2
E2		√		85.3	88.6	89.4
E3			√	87.9	92.0	90.4
E4	√	√		91.1	86.7	91.8
E5	√		√	90.2	92.0	92.8
E6		√	√	88.0	92.7	92.2
E7	√	√	√	91.0	94.0	93.4

图7为将CAM模块单独添加到原始的YOLOv5s模型中的实验结果。由此可见,CAM模块增强了模型对小目标图像特征提取的能力,通过自适应的通道注意力权重引导特征学习,可更好地捕获图像的全局和局部信息,提升模型检测精度和鲁棒性。

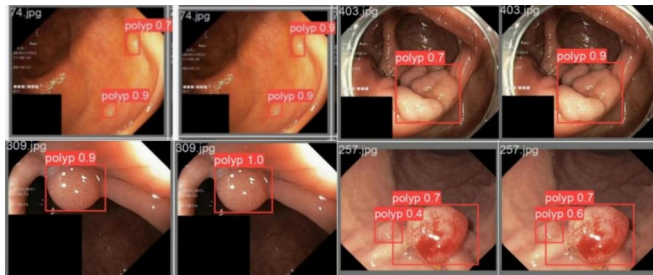
**Fig. 7 Experimental results after adding CAM module****图7 加入CAM模块后的实验结果**

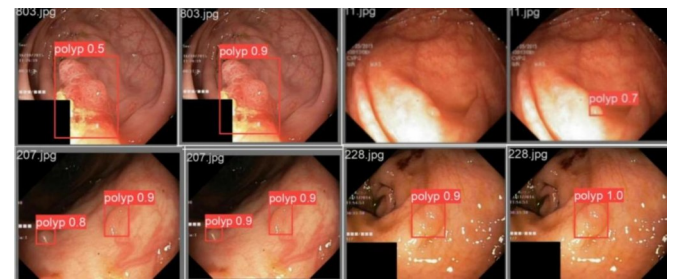
图8为将LSKA模块单独添加到原始的YOLOv5s模型中的实验结果。LSKA模块将2D卷积内核分解为级联水平和垂直1-D内核,使模型可有效捕获目标物体的上下文信息,既扩大了模型感受野,又提升了模型对微小目标和细微特征的识别能力,还增强了模型对空间和深度信息的捕获。

由图8可见,E2模型成功检测到了一些原模型未能发现的结肠息肉病灶,这表明LSKA模块增强了模型的感知能力和检测性能。

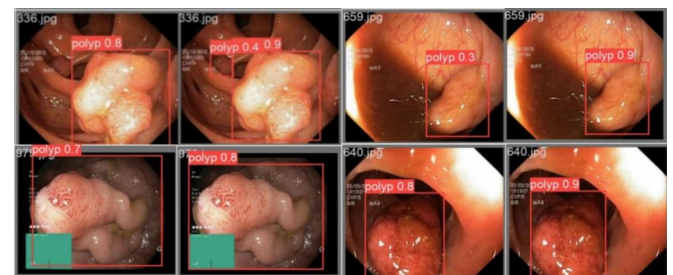
图9为将Soft-NMS模块单独添加到原始的YOLOv5s模型中的实验结果。引入Soft-NMS有效增强了模型对密集重叠结肠息肉目标的检测能力,相较于传统NMS算法并非直接将重叠框的得分置零,而通过高斯惩罚函数降低

块;模型E3为只引入Soft-NMS;模型E4为在模型E1的基础上加入LSKA模块;模型E5为在模型E1的基础上引入Soft-NMS;模型E6为在模型E2的基础上引入Soft-NMS;模型E7(QB-YOLO)为加入全部3个模块。

由表2可知,在只更换CAM模块的情况下,E1模型的准确率、召回率、mAP相较于E0分别提升3.2%、0.9%、0.7%;在更换CAM模块并引入LSKA模块后,E4模型的准确率、召回率、mAP相较于E1提升0.8%、0.3%、1.6%;继续引入Soft-NMS后,模型的准确率、召回率、mAP相较于基准模型共提升3.9%、8.5%、3.9%,提升效果十分明显。

**Fig. 8 Experimental results after adding LSKA module****图8 加入LSKA模块后的实验结果**

与已选框重叠的其他检测框的得分,以提升模型面对密集目标的检测精度。

**Fig. 9 Experimental results after adding Soft NMS module****图9 加入Soft-NMS模块后的实验结果**

由图9可见,在部分密集结肠息肉小目标图像中,E3模型成功检测出了原YOLOv5s模型未能检测出的部分病灶,还提升了已检出的病灶精准度。

4.6 实验结果与分析

图10、图11展示了改进后QB-YOLO模型与原YOLOv5s模型在100轮训练下的mAP和Recall。其中,虚线代表原YOLOv5s模型,实线代表改进后的QB-YOLO模型。

由此可见,改进后的QB-YOLO在相同的训练轮数下拥有更高的平均精度与召回率。

在平均精度方面,原始YOLOv5s模型在大约40轮训练后已接近收敛状态,QB-YOLO模型在相同训练轮数下收敛速度更快、性能更高;在召回率方面,在大约50轮训练后QB-YOLO的召回率、收敛速度开始明显优于原YOLOv5s模型,原因为所提模型引入了LSKA模块、CAM模块和Soft-NMS模块。

具体地,LSKA模块将2D卷积内核分解为级联水平和垂直1-D内核,扩大了模型感受野,既能使模型更有效地捕获目标物体的上下文信息,又能更准确地检测到微小目标和细微特征,从而提升了模型的平均精度和召回率;CAM模块通过自适应通道注意力权重来引导特征学习,增强了模型对小目标图像的特征提取能力,提升了模型的平均精度与召回率;Soft-NMS有效增强了模型检测密集重叠结直肠息肉目标的能力,减少了重叠框之间的竞争,使QB-YOLO模型能更好地捕获目标。

综上,上述模块使QB-YOLO模型在相同的训练轮数下能更快地学习与提取特征信息,从而达到更高的平均精度和召回率,为模型训练和性能提升提供了强有力的支撑。

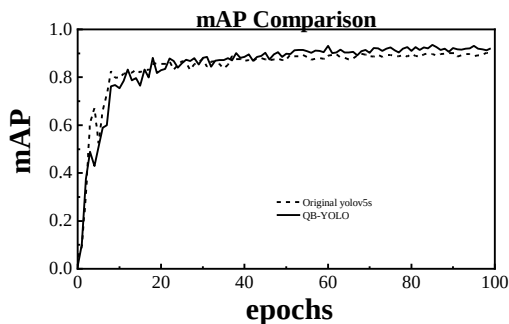


Fig. 10 mAP curves before and after improvement

图10 改进前后的mAP曲线

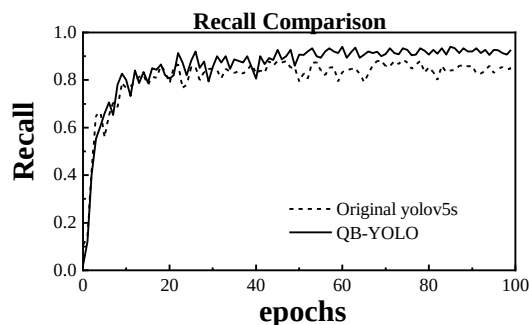


Fig. 11 Recall curves before and after improvement

图11 改进前后的Recall曲线

此外,本文将部分验证集上的实验结果进行可视化展示,如图12所示。由此可见,无论是大目标还是小目标的结直肠息肉病灶,QB-YOLO的检测精度均相较于原始模型更优秀。

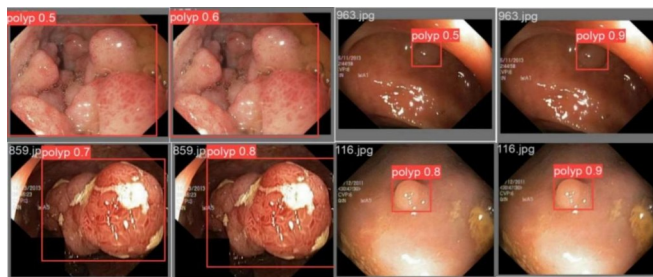


Fig. 12 Experimental results of models before and after improvement

图12 改进前后模型的实验效果

5 结语

本文提出一种基于改进型YOLOv5s的结直肠息肉检测模型QB-YOLO。通过引入LSKA模块、CAM模块和Soft-NMS模块,解决了结直肠息肉检测任务中微小目标检测和密集目标检测中存在的问题。

首先,LSKA模块和CAM模块显著提升了模型在检测微小目标方面的性能,使其能更准确地识别结直肠息肉图像中的微小目标和细微特征;其次,Soft-NMS有效解决了密集目标检测中的重叠问题,提升了目标检测的准确性和鲁棒性。实验表明,QB-YOLO相较于原始YOLOv5s模型在准确率、召回率和平均精度等关键指标上均取得了显著提升,充分验证了改进方法的优越性。

参考文献:

- [1] BERNAL J, TAJKBAKSH N, SAJKBKASH F J, et al. Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge [J]. IEEE Transactions on Medical Imaging, 2017, 36(6): 1231-1249.
- [2] GIACOSA A, FRASCIO F, MUNIZZI F. Epidemiology of colorectal polyps[J]. Techniques in Coloproctology, 2004, 8: 243-247.
- [3] ZHANG Y M. Application of multicolor fluorescence in situ hybridization to early detection of chromosome aberrations in colorectal cancer[D]. Beijing: Chinese Academy of Medical Sciences & Peking Union Medical College, 2009.
张月明. M-FISH应用于检测结直肠癌早期染色体改变的研究[D]. 北京:北京协和医学院;中国医学科学院,2009.
- [4] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [5] REN S, HE K, 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, 2017, 39(6): 1137-1149.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [7] REDMON J, FARHADI A. Yolov3: an incremental improvement [DB/OL]. <https://arxiv.org/abs/1804.02767>.
- [8] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-

- of-freebies sets new state-of-the-art for real-time object detectors [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2023: 7464–7475.
- [9] REIS D, KUPEC J, HONG J, et al. Real-time flying object detection with YOLOv8[DB/OL]. <https://arxiv.org/abs/2305.09972>.
- [10] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916.
- [11] KARARGYRIS A, BOURBAKIS N. Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos [J]. IEEE Transactions on Biomedical Engineering, 2011, 58(10): 2777–2786.
- [12] ESKANDARI H, TALEBPOUR A, ALIZADEH M, et al. Polyp detection in wireless capsule endoscopy images by using region-based active contour model[C]// 19th Iranian Conference of Biomedical Engineering, 2012: 305–308.
- [13] SILVA A, GOMBOLAY M, KILLIAN T, et al. Optimization methods for interpretable differentiable decision trees applied to reinforcement learning [C]// International Conference on Artificial Intelligence and Statistics, 2020: 1855–1865.
- [14] LI H R, LIU K, CHANG S L, et al. PET super-resolution reconstruction method based on residual mixed domain attention network[J]. Electronic Measurement Technology, 2021, 44(14): 103–110.
李浩然, 刘琨, 常世龙, 等. 基于残差混合域注意力网络的PET超分辨率重建方法[J]. 电子测量技术, 2021, 44(14): 103–110.
- [15] WANG G T, LIN Z Z, FU Q S, et al. Joint generative adversarial network model for classification of benign and malignant pulmonary nodules [J]. Chinese Journal of Scientific Instrument, 2020, 41(11): 188–197.
王桂棠, 林楨哲, 符秦沈, 等. 联合生成对抗网络的肺结节良恶性分类模型[J]. 仪器仪表学报, 2020, 41(11): 188–197.
- [16] FAN S H, LIU S C, CAO E, et al. Small intestinal polyp detection in wireless capsule endoscopy images[J]. Chinese Journal of Biomedical Engineering, 2019, 38(5): 522–532.
范姍慧, 刘士臣, 曹鄂, 等. 无线胶囊内镜图像小肠息肉的自动识别[J]. 中国生物医学工程学报, 2019, 38(5): 522–532.
- [17] SUN X H, PAN X Y. Faster R-CNN based polyps detection in endoscopic videos [J]. Journal of Xi'an University of Posts and Telecommunications, 2020, 25(2): 29–34.
孙雪华, 潘晓英. Faster R-CNN 内窥镜息肉检测[J]. 西安邮电大学学报, 2020, 25(2): 29–34.
- [18] LI J Y, ZHANG J, CHANG D D, et al. Computer-assisted detection of colonic polyps using improved faster R-CNN[J]. Chinese Journal of Electronics, 2019, 28(4): 718–724.
- [19] REN L L, BIAN X, WANG G L, et al. Polyp segmentation network GLIA-Net based on deep learning [J]. Computer Engineering, 2022, 48(12): 248–254.
任莉莉, 边璇, 王光磊, 等. 基于深度学习的息肉分割网络 GLIA-Net [J]. 计算机工程, 2022, 48(12): 248–254.
- [20] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. Expert Systems with Applications, 2024, 236: 121352.
- [21] XIAO J, ZHAO T, YAO Y, et al. Context augmentation and feature refinement network for tiny object detection [EB/OL]. <https://openreview.net/pdf?id=q2ZaVU6bEsT>.
- [22] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS-improving object detection with one line of code [C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 5561–5569.
- [23] GUO M H, LU C Z, LIU Z N, et al. Visual attention network [J]. Computational Visual Media, 2023, 9(4): 733–752.
- [24] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-seg: a segmented polyp dataset [DB/OL]. <https://arxiv.org/abs/1911.07069>.
- [25] ZHAO Y, LV W, XU S, et al. DETRs beat Yolos on real-time object detection [DB/OL]. <https://arxiv.org/abs/2304.08069>.
- [26] WANG C Y, YE H I, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information [DB/OL]. <https://arxiv.org/abs/2402.13616>.

(责任编辑:刘嘉文)