

# 改进 YOLOv4 框架的胃息肉检测<sup>①</sup>



吴宇杰, 肖满生, 范明凯, 胡一凡

(湖南工业大学 计算机学院, 株洲 412007)

通信作者: 肖满生, E-mail: xiaomansheng@hut.edu.cn

**摘要:** 在计算机视觉的内窥镜胃部息肉检测中, 高效提取小型息肉图像特征是设计深度学习的计算机视觉模型一个难点. 针对该问题, 提出了一种 YOLOv4 改进的 YOLOv4-polyp 检测模型. 首先在 YOLOv4 的基础上, 引入 CBAM 卷积注意力模块增强模型在复杂环境的特征提取能力; 其次设计出轻量级 CSPDarknet-49 网络模型, 在降低模型复杂度的同时提高检测精度和检测速度; 最后根据胃息肉数据集的特点, 采用 K-means++ 聚类算法对胃息肉数据集进行聚类分析, 得到优化后的锚框. 实验对比结果表明, YOLOv4-polyp 对于经典 YOLOv4 模型在保持检测速率不变的同时, 在两个数据集中平均检测精度分别提升了 5.21% 和 2.05%, 表现出良好的检测性能.

**关键词:** YOLOv4; 注意力机制; K-means++; 目标检测

引用格式: 吴宇杰, 肖满生, 范明凯, 胡一凡. 改进 YOLOv4 框架的胃息肉检测. 计算机系统应用, 2023, 32(2): 250-257. <http://www.c-s-a.org.cn/1003-3254/8931.html>

## Improved YOLOv4 Framework for Gastric Polyp Detection

WU Yu-Jie, XIAO Man-Sheng, FAN Ming-Kai, HU Yi-Fan

(School of Computer Science, Hunan University of Technology, Zhuzhou 412007, China)

**Abstract:** In endoscopic gastric polyp detection based on computer vision, efficiently extracting the features of images of small polyps is a difficulty in the design of a deep learning-based computer vision model. To solve this problem, this study proposes a detection model based on an improved you only look once version 4 (YOLOv4), namely YOLOv4-polyp. Specifically, on the basis of YOLOv4, this study adds a convolutional block attention module (CBAM) to enhance the feature extraction capability of the model in complex environments. Then, a lightweight CSPDarknet-49 network model is designed to both reduce the complexity of the model and improve its detection accuracy and detection speed. Finally, considering the characteristics of the gastric polyp datasets, the K-means++ clustering algorithm is used for the cluster analysis of the gastric polyp datasets and the attainment of the optimized anchor box. The experimental comparison results show that compared with the classical YOLOv4 model, the proposed YOLOv4-polyp achieves favorable detection performance on the two datasets as it improves the average detection accuracy by 5.21% and 2.05%, respectively, without compromising the detection speed.

**Key words:** YOLOv4; attention mechanism; K-means++; target detection

随着人工智能与目标检测应用技术的发展, 使用计算机视觉辅助检测胃部息肉成为一种趋势. 利用内窥镜采集的图像进行息肉的检测与识别是计算机视觉辅助的重要内容. 息肉检测过程中由于光照变化、

息肉边缘不明显、视角遮挡、分辨率等问题使得准确识别困难, 因此对于计算机视觉辅助检测的准确性和稳定性有很高的需求. 根据不同的检测手段, 将计算机视觉辅助检测分为传统方法<sup>[1,2]</sup>和基于深度学习的目

① 基金项目: 湖南省自然科学基金 (2021JJ50049, 2022JJ50077); 湖南省教育厅重点项目 (21A0607)  
收稿时间: 2022-06-17; 修改时间: 2022-07-18; 采用时间: 2022-08-09; csa 在线出版时间: 2022-11-04  
CNKI 网络首发时间: 2022-11-15

标检测方法。在传统方法中,医生会通过仪器捕捉处理过的画面来快速判断息肉是否存在<sup>[3]</sup>。如果画面中息肉特征不明显或有许多干扰,传统方法则无法快速准确判断息肉。基于深度学习的目标检测方法分为一阶段检测算法和二阶段检测算法。一阶段检测算法代表有 SSD 系列<sup>[4]</sup>和 YOLO 系列<sup>[5]</sup>等。二阶段检测算法代表有 R-CNN 系列<sup>[6]</sup>和 SPPNet<sup>[7]</sup>等。

一阶段目标检测算法着重于实时检测方面进行优化。如马燕<sup>[8]</sup>基于 SSD 进行改进,在动态肠镜息肉图像实时定位算法中,使用多种 inception 模块代替 SSD 基础特征提取网络结构中的传统卷积层来增加神经网络深度提取能力,最终平均检测精度 (*mAP*) 达到 96%;王博等人<sup>[9]</sup>在 M2det 基础上进行改进,在提出的基于 M2det 内窥镜息肉检测方法中,他们通过 FFMs 模块融合主干网络特征,使图像特征得到了充分利用,以及在 SFAM 模块中增加 SENet 注意力机制来保留有效特征,最终 *mAP* 达到了 98.25%;Laddha 等人<sup>[10]</sup>在针对息肉大小不一的基础上,使用 YOLOv3 和 YOLOv3-tiny 两种不同规模大小的模型,最终实验得到了 91% 和 82% 的准确率;Cao 等人<sup>[11]</sup>在 YOLOv4 的基础上,设计出一个特征提取和融合模块,并将其与 YOLOv4 网络相结合,最终在准确率方面达到了 91.6%。Yu 等人<sup>[12]</sup>在 Scaled-YOLOv4 的基础上,设计出 ITH 模块,ITH 模块与 YOLO 检测头共享权重以进行快速特征提取,实验结果表明,该模块对比原先提升了 30% 的识别速度。二阶段目标检测算法注重于目标检测精度方面进行提升。如 Taş 等人<sup>[13]</sup>基于 Faster R-CNN 进行改进,首先使用 ResNet-101 替代 VGG16,然后使用基于卷积神经网络的超分辨率 (SRCNN) 方法提高结肠镜图像的分辨率,最终提升模型的准确率;Cheng 等人<sup>[14]</sup>基于 Mask R-CNN 进行改进。首先提出了一种可学习的定向导数网络 (LOD-Net),计算息肉每个像素的 8 个定向导数,然后用这些导数值构成息肉的候选框,最终提高息肉分割边界预测的准确性。不过二阶段目标检测算法在实时性方面比较差。在实例分割领域,任莉莉等人提出了基于 U-Net 网络改进的 GLIA-NET 网络<sup>[15]</sup>,主要特点是在双层卷积后加入交互式注意力卷积模块,引入后检测精度提升 6.7%。

这些方法在针对常规息肉检测准确率或实时性上有良好的改进。由于胃部环境复杂,小型息肉具有分辨率低、特征信息不足问题,常规特征提取到的小型息肉特征不足,导致漏检现象发生。针对小型息肉漏检问

题,本文在 YOLOv4 的基础上进行了以下改进:(1)对 CSPDarknet-53 基础网络上进行改进,设计出一种精简的 CSPDarknet-49 网络,减少网络深度,使得网络能够保留更多的小目标特征信息,同时增加准确率;(2)利用息肉数据集的特点,使用 K-means++ 聚类算法对数据集进行聚类分析,得到 9 组不同比例大小的锚框,提升模型对小目标的检测精度;(3)在 YOLOv4 的骨干网络中引入卷积注意力模块,使得网络能够在空间和通道两个维度进行特征提取,从而学习到更深层次的特征信息,从而提高小目标的检测能力。

## 1 YOLOv4 网络介绍

YOLOv4 网络<sup>[16]</sup>结构如图 1 所示。YOLOv4 由 CSPDarknet53 主干特征提取网络、SPP、PANet 特征融合模块和检测头 YOLO head 组成。CSPDarknet53 由 CBM 和 CSP 模块组成,CSP 模块内部残差网络可以降低网络复杂结构,同时对残差边的信息进行少量处理后与原先提取特征进行堆叠,保留深层次特征同时过滤背景信息。经过主干网络处理后得到输出特征。经过 SPP 最大池化操作后,PANet 模块的上采样,卷积堆叠等操作对于不同层次的特征信息进行特征融合,最后将特征信息传给 YOLO Head 进行预测判断,从而实现目标检测。

## 2 YOLOv4 改进相关工作

前面介绍了在 YOLOv4 网络模型。由于息肉检测环境复杂,小型息肉分辨率不足和特征信息过少,YOLOv4 模型在检测小型息肉时容易发生漏检现象。本文针对上述问题,改进出 YOLOv4-polyp 网络模型。本文从锚框优化,主干网络结构和注意力机制 3 方面对 YOLOv4 模型<sup>[16]</sup>进行改进,改进后的 YOLOv4-polyp 模型如图 2 所示。其中橙色部分为对比图 1 中 YOLOv4 模型改进点。首先结合息肉数据集特点,使用 K-means++ 聚类算法对息肉数据集得到合适的锚选框。其次为了提高小型息肉的检测能力,在主干网络中引入 CBAM 卷积注意力模块,使得 YOLOv4-polyp 模型能够从通道和空间两个注意力机制增强小型息肉特征提取能力,从而避免发生漏检。为了降低 CBAM 卷积注意力模块的增加导致检测速率的降低,对主干网络 Res 模块层数进行精简,将部分 CBM 替换为 CBL 简化模型复杂度。在保证准确率的同时降低模型复杂度,从而提升模型的检测速率。

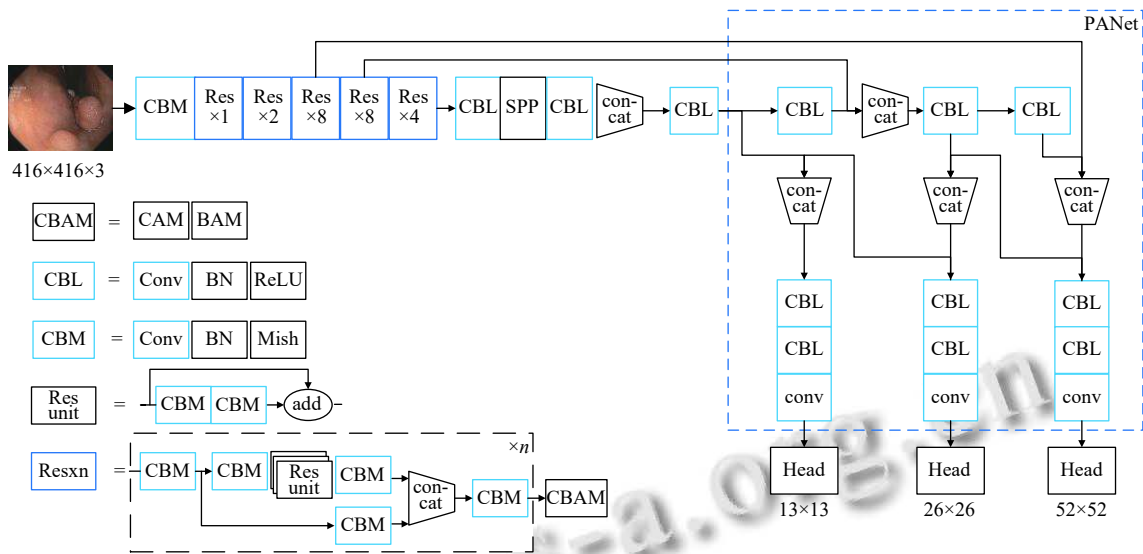


图1 YOLOv4网络结构

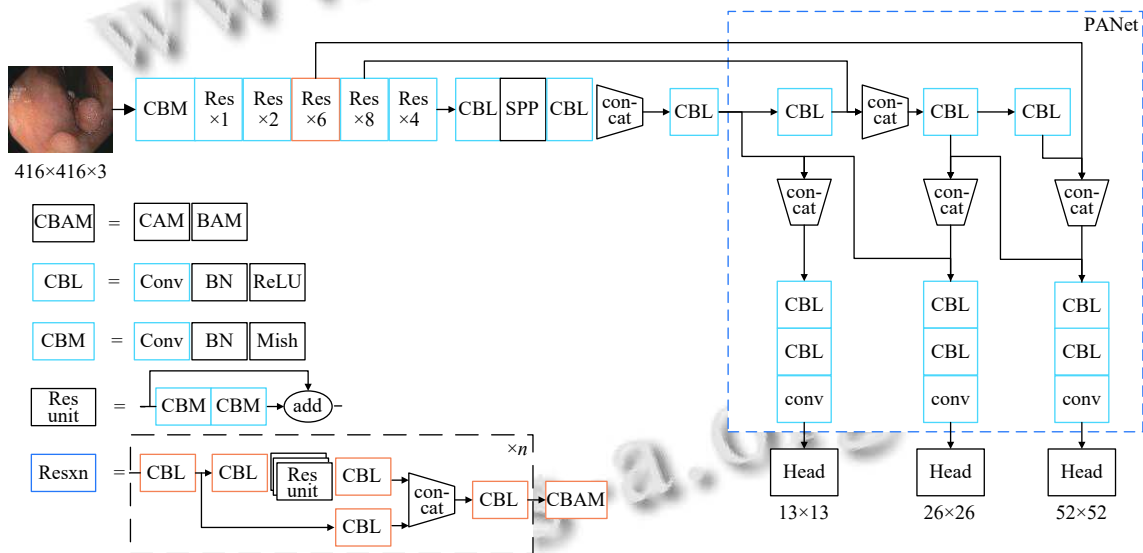


图2 YOLOv4-polyp网络结构

YOLOv4-polyp网络结构处理流程如图2所示, 首先网络输入图片大小为  $416 \times 416 \times 3$ . 该网络主要由 CBAM, CBL, CBM, Res unit 和 Res 组成, CBAM 卷积注意力模块由 CAM 和 BAM 组成, CBL 代表卷积层, 正则化层和 LeakyReLU 激活函数组成, CBM 与 CBL 不同的是激活函数使用了 Mish 激活函数. Res 层是由残差网络, 卷积层和 CBAM 模块组成. Res 层在经过特定次数卷积后, 会在尾部经过一次 CBAM 模块, 即经过 Res 层处理后的特征图会依次与通道注意图和空间注意图进行特征相乘, 抑制无关背景信息干扰, 提升模

型的自适应特征提取能力, 最终提升胃息肉检测精度. 由于降低了模型深度以及加入了 CBAM 卷积注意力机制, 在经过多次卷积操作后依旧保留小型目标的特征信息, 避免了深层次卷积网络在卷积过程中损失过多小型息肉特征信息, 从而加强网络的小型息肉特征提取能力.

### 2.1 CSPDarknet-49 网络

YOLOv4 通过 CSPDarknet-53 网络结构提取目标特征, 输出 3 种特征图进行多尺度检测. 3 种特征图的大小分别为  $13 \times 13$ ,  $26 \times 26$ ,  $52 \times 52$ . 其中  $52 \times 52$  的尺度

用于小型目标检测. 由于模型深度问题, 小型息肉在特征提取过程中会丢失部分特征信息, 最终导致检测效果不佳. 为了提高YOLOv4小型息肉检测能力, 通过实验得到一种简化的CSPDarknet-49网络结构, 在能够提高小型息肉检测能力情况下保持检测精度. 如表1所示, CSPDarknet-49网络在CSPDarknet-53网络基础上, 缩减CSPnet层数数量达到降低模型深度效果, 同时将部分Mish激活函数替换成LeakyReLU激活函数, 提升模型对小型目标检测精度并提升检测速度. 尺寸为416×416的图像在经过CSPDarknet-49网络处理后, 最终得到52×52, 26×26, 13×13三种不同尺度的特征图.

表1 CSPDarknet-49网络结构

Type	Filters	Size	Output	Nums
convolutional	32	3×3	416×416	—
convolutional	64	3×3/2	208×208	—
convolutional	32	1×1	—	—
convolutional	64	3×3	—	×1
residual	—	—	208×208	—
convolutional	128	3×3/2	104×104	—
convolutional	64	1×1	—	—
convolutional	128	3×3	—	×2
residual	—	—	104×104	—
convolutional	256	3×3/2	52×52	—
convolutional	128	1×1	—	—
convolutional	256	3×3	—	×6
residual	—	—	52×52	—
convolutional	512	3×3/2	26×26	—
convolutional	256	1×1	—	—
convolutional	512	3×3	—	×8
residual	—	—	26×26	—
convolutional	1024	3×3/2	13×13	—
convolutional	256	1×1	—	—
convolutional	512	3×3	—	×4
residual	—	—	13×13	—

## 2.2 K-means++与锚框优化

在实时目标检测网络中, 多数目标检测算法是根据实验来预设锚框尺寸和宽高比, 如SSD, YOLO系列模型. 但根据实验得到的锚框尺寸, 只适用于常规情况下的目标尺寸. 息肉数据集与常规数据集不同, 大多真实框与预设的锚框差距过大, 从而容易发生错检, 漏检现象. 因此本文采用K-means++<sup>[17]</sup>聚类算法对数据集生成合适的锚框, 从而提升检测精度.

传统的K-means++聚类算法在计算每个样本与聚类中心的最短距离是指欧式距离, 然而这种距离计算方式不利于区分大型边框与真实框和小型边框与真实框的误差, 因此在计算距离方面替换成利用预测的边

框与真实框的交并比IOU(intersection over union)来解决误差问题, 如式(1)所示:

$$Distance(box, cluster) = 1 - IOU(box, cluster) \quad (1)$$

其中,  $box$  表示真实框的集合,  $cluster$  表示各聚类中心集合,  $IOU(box, cluster)$  表示真实框与聚类中心交集与并集的比例. IOU值越高代表真实框与聚类中心越接近, 即两者相关度越高. 假设聚类框与 $box$ 具有相同坐标系, 则计算交集如式(2)所示:

$$intersection = \min(c_{la}, b_{oxa}) \times \min(c_{lb}, b_{oxb}) \quad (2)$$

聚类框与 $box$ 框面积计算方式为:

$$area1 = c_{la} \times c_{lb} \quad (3)$$

$$area1 = c_{la} \times c_{lb} \quad (4)$$

$$area2 = b_{oxa} \times b_{oxb} \quad (5)$$

因此IOU计算公式如式(5)所示:

$$IOU = \frac{intersection}{area1 + area2 - intersection} \quad (6)$$

利用K-means++聚类算法对数据集进行聚类分析, 最终得到9个不同比例的锚框参数, 将这些锚框参数按照从小到大分成3组, 不同尺度的使用对应的锚框来进行预测, 最终结果如表2所示.

表2 使用K-means++聚类分析优化的锚框尺寸

特征图大小	预设锚框尺寸	边界框数
13×13	(148, 145), (182, 223), (273, 303)	3
26×26	(81, 83), (104, 121), (117, 198)	3
52×52	(29, 34), (47, 62), (67, 118)	3

## 2.3 卷积注意力模块机制

针对解决息肉数据集中息肉大多偏小、边缘不明显导致漏检问题, 本文在YOLOv4网络模型的主干网络中引入CBAM注意力模块. CBAM注意力模块<sup>[18]</sup>是一种用于前馈卷积神经网络的简单而有效的注意力模块. 与SENet(squeeze-and-excitation networks)<sup>[19]</sup>注意力模块不同在于, SENet模块仅在通道上进行特征提取, 而CBAM卷积注意力模块结合通道和空间两个方向进行特征提取, 使得模型能更有效提取息肉边缘特征, 抑制背景信息干扰并丰富高层次特征信息, 从而减少胃息肉漏检现象.

如图3所示, 设输入特征 $F$ 大小为 $H \times W \times C$ , 在通道注意力模块(CAM)中, 输入特征 $F$ 分别进行最大化操作和平均池化操作得到两个大小为 $1 \times 1 \times C$ 特征

图. 接着两个特征图分别经过共享 MLP 层卷积处理后得到两个大小为  $1 \times 1 \times C$  特征图. MLP 层包含 2 层卷积层和一个 ReLU 激活函数, 第 1 个卷积层的输出通道数为  $C/r$ , 其中  $r$  为缩减系数, 第 2 个卷积层的输出通道数为  $C$ . 经过 MLP 层卷积处理后的两个特征层进行按位相加操作后, 经过 Sigmoid 激活函数最终得到权重系数  $M_c$ .  $M_c$  与输入特征  $F$  进行按位相乘后得到通

道注意力特征. 在空间注意力模块中, 首先对通道注意力特征分别进行最大池化和平均池化操作后得到两个大小为  $H \times W \times 1$  的特征图, 并对这两个特征图按通道进行拼接后得到大小为  $H \times W \times 2$  的特征图. 然后经过大小为  $7 \times 7$  卷积层进行卷积操作后, 再经过 Sigmoid 激活函数后得到权重系数  $M_s$ .  $M_s$  与通道注意力特征图进行按位相乘操作得到输出特征  $F'$ .

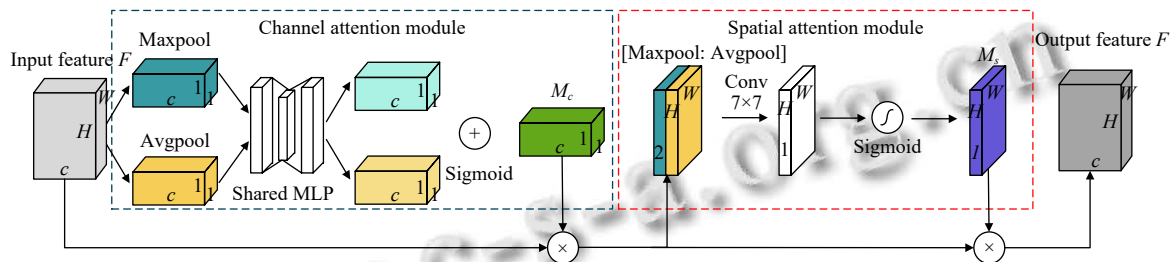


图3 CBAM 模块结构

### 3 实验和结果分析

本文的实验环境为 OS: Ubuntu 12.04; CPU: Xeon (R) CPU E5-2678 v3; RAM: 16 GB; GPU: NVIDIA GeForce RTX 2080 Ti, GPU 加速软件 CUDA 11.1, 使用 PyTorch 1.8 版本作为深度学习框架.

#### 3.1 数据集和训练设置

本次实验使用两个数据集, 分别是 Hyper-Kvasir<sup>[20]</sup> 和 CVC-ClinicDB<sup>[21]</sup> 数据集. Hyper-Kvasir 数据集是一个肠胃道图像和视频数据集, 总共包含了 11 万张图像和 374 个视频. 这些数据是在挪威的 Baerum 医院进行的胃镜和结肠镜检查期间收集的, 其中部分图像由经验丰富的肠胃内镜医师标记. 本次实验将采用 1000 张带“polyp”标签的胃息肉图像做实验. 该数据集息肉图片尺寸较大, 通常在  $622 \times 530$  像素之间. 息肉目标尺寸差别较大且不均匀, 其中多为中等大小目标. 小型息肉的图片数量偏少.

数据集中部分息肉标注丢失导致部分数据无法使用, 对数据集进行数据清洗后得到 994 张标注准确的图片用于数据集训练. 随机抽取 636 张作为训练集, 剩下的 200 张图片和 158 张图片分别作为测试集和验证集. 训练验证集和测试集的比例为 8:2.

CVC-ClinicDB 数据集是一个开源的数据集, 包含来自 29 个结肠镜序列的 612 张静态图像. 每张图片都带有手动注释息肉位置, 可用于医学目标图像检测. 该数据集图片尺寸较小, 通常在  $388 \times 288$  像素之间, 相比

Hyper-Kvasir 数据集每张图片带有更多的息肉目标.

对数据集进行数据清洗后得到 610 张标注准确的图片用于训练. 随机抽取 390 张图片作为训练集, 剩下的 122 张图像和 98 张图像分别作为测试集和验证集. 训练验证集和测试集的比例为 8:2.

#### 3.2 模型训练与参数设置

实验采用 PyTorch 深度学习框架搭建网络. 在训练图像送入模型前, 需要将图像缩放至  $416 \times 416$  像素大小. 在训练时采用迁移学习方法, 使用在 Imagenet 数据集中训练好的 CSPDarknet-53 预训练权重. 因为 YOLOv4-polyp 模型更改了 CSPDarknet-53 主干网络结构, 在导入前需要手动调整预训练权重参数以便预训练权重导入. 由于 Hyper-Kvasir 数据集和 CVC-ClinicDB 数据集属于小型数据集, 在数据增广上采用常规的旋转, 缩放, 对称等常见数据增强操作. 同时图像预处理采用了 Mosaic 数据增强, 即随机抽取 4 张图像, 进行随机裁剪缩放后在拼接到一张图像上. 这样不仅对于单次训练能够增加模型训练量, 在随机裁剪过程中也会随机出现小型目标, 进而增强模型的小型息肉特征提取能力. 训练参数设置如表 3 所示. 训练至模型停止收敛时结束训练.

#### 3.3 评价指标

实验采用准确率 (precision,  $P$ ), 召回率 (recall,  $R$ ), 平均  $AP$  值 ( $mAP$ ) 和  $F$  值作为评价评价指标. 在目标检测中, 根据交并比  $IOU$  来判断目标是否被成功检测

到. 对于数据集中的某类目标, 假设阈值为  $\alpha$ , 当预测框和真实框的  $IOU$  大于  $\alpha$  时, 表示模型正确预测, 相反则表示模型错误预测. 混淆矩阵如表 4 所示.

表 3 模型训练参数

参数名称	数值取值	备注
epoch	100	训练轮次
batch_size	8	单次批处理大小
learning_rate	0.001	初始学习率
momentum	0.9	动量
weight_decay	0.93	权重衰减系数
learning_rate_end	0.00001	结束时学习率

表 4 混淆矩阵定义

Real	Prediction	
	Positive	Negative
True	$TP$	$FN$
False	$FP$	$TN$

在表 4 中,  $TP$  表示正确预测的阳性样本数,  $FP$  表示错误预测的阴性样本数,  $FN$  表示错误预测的阳性样本数,  $TN$  表示正确预测的阴性样本数. 准确率和召回率的计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$AP$  值通常作为目标检测性能的评价指标.  $AP$  值是 P-R 曲线下的面积, 其中召回率取 X 轴, 精度取 Y 轴,  $N$  表示类别的个数.  $AP$  代表特定类别中模型的准确性.  $mAP$  表示所有类别的平均精度, 并可测量所有类别中网络模型的性能.  $mAP50$  表示预测框和真实框的  $IOU$  大于 0.5.  $mAP$  的计算公式如下:

$$mAP = \frac{1}{N} \sum_{i=1}^N \int_0^1 PdR \quad (9)$$

Precision 和 Recall 的指标大多数时候是互相排斥的, 为了能够综合考虑这两个指标, 常见的处理方法是使用  $F$  值指标.  $F$  值的计算公式如下:

$$F = \frac{(a^2 + 1)P \times R}{a^2(P + R)} \quad (10)$$

当  $a=1$  时,  $F1$  的计算公式为:

$$F1 = \frac{2PR}{P + R} \quad (11)$$

### 3.4 实验结果与分析

为了验证 YOLOv4-polyp 算法的合理性, 本文将

与现有的目标检测模型 SSD, RetinaNet, YOLOv3, 及 YOLOv4 分别在 Hyper-Kvasir 数据集和 CVC-ClinicDB 数据集进行实验对比, 通过对比数据集中的  $mAP$  指标,  $F$  值和平均检测速率 (time) 来进行对比, 由表 5 可见, YOLOv4-polyp 与原 YOLOv4 相比, 平均检测速率增加了 1 ms, 但是  $mAP$  提高了 5.21%, 在所有模型中 YOLOv4-polyp 的检测精度是最高的. 由表 6 可见, YOLOv4-polyp 与原 YOLOv4 相比, 平均检测速率增加 1 ms, 但是  $mAP$  提高了 2.05%, 在所有模型中 YOLOv4-polyp 依然是所有模型中检测精度最高的. 结合表 5、表 6 可知, YOLOv4 相比较于 YOLOv3, 平均检测速度增加了 2 ms, 但是在检测精度上均优于 YOLOv3. SSD 和 RetinaNet 在检测精度和平均检测速率均表现较差, SSD 由于先验框固定大小, 导致在检测小息肉方面会存在漏检现象, 导致检测精度下降. RetinaNet 在检测精度方面优于 SSD, 但是平均检测速率远不如 SSD, 无法满足实时检测需求, 因此 YOLOv4-polyp 在保证平均检测速率相较于 YOLOv4 几乎不变的情况下, 在检测精度达到了更好的效果, 最终检测性能更好.

表 5 不同模型在 Hyper-Kvasir 数据集下的表现

方法	$P$ (%)	$R$ (%)	$mAP$ (%)	$F1$	Time (s)
SSD	85.33	75.12	83.77	0.80	0.040
RetinaNet	85.56	76.56	85.09	0.81	0.092
YOLOv3	84.90	76.17	85.06	0.80	0.032
YOLOv4	88.78	81.25	86.82	0.85	0.030
YOLOv4-polyp	93.73	85.23	92.03	0.89	0.031

表 6 不同模型在 CVC-ClinicDB 数据集下的表现

方法	$P$ (%)	$R$ (%)	$mAP$ (%)	$F1$	Time (s)
SSD	87.71	78.11	86.52	0.83	0.040
RetinaNet	88.56	80.18	88.52	0.84	0.091
YOLOv3	92.76	90.71	90.96	0.92	0.032
YOLOv4	92.86	90.70	92.97	0.92	0.030
YOLOv4-polyp	95.97	92.25	95.02	0.94	0.031

图 4 为在 Hyper-Kvasir 数据集中针对常规和小型息肉检测试验对比, 图 4(a) 为 YOLOv4 算法检测结果图像, 图 4(b) 为 YOLOv4-polyp 检测结果图像.

可以看出, 当目标尺寸较小、特征不明显时, YOLOv4 模型难以检测出目标位置, 从而发生漏检现象. YOLOv4-polyp 通过改进模型深度, 引入 CBAM 模块来一定程度上缓解模型过深导致小型目标特征信息丢失, 提高模型特征提取能力, 从而可以提升模型在小型息肉和常规息肉检测能力.

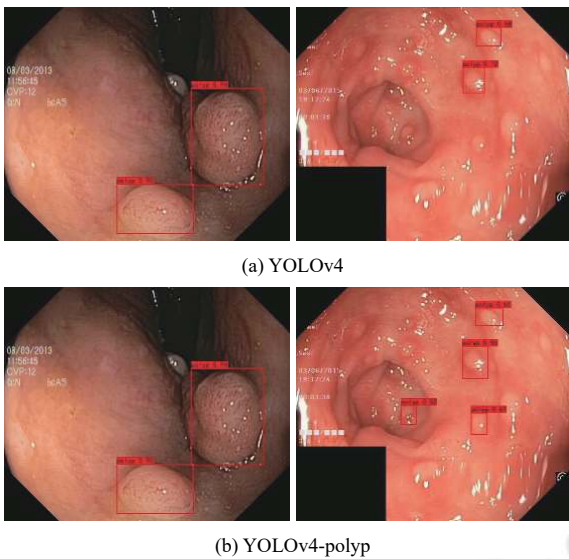


图4 不同模型在 Hyper-Kvasir 数据集下检测对比

图5为在CVC-ClinicDB数据集中小型息肉检测对比,图5(a)为YOLOv4算法检测结果图像,图5(b)为YOLOv4-polyp检测结果图像.可以看出YOLOv4在针对形状缺失,轮廓不完整的情况下容易丢失目标,容易发生漏检现象.YOLOv4-polyp通过引入CBAM注意力机制,提升模型的特征提取能力,可以准确检测在轮廓缺失或形状不规则下的小型息肉.

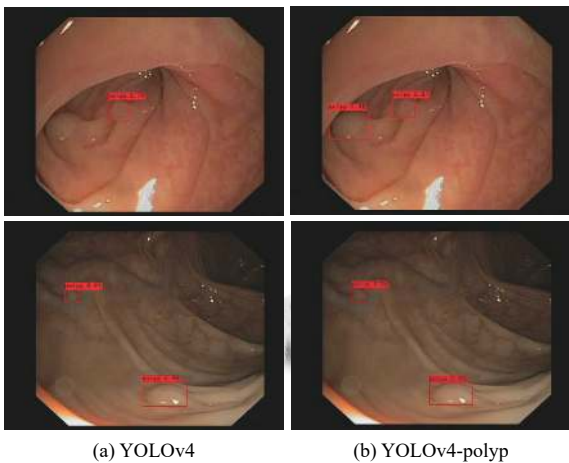


图5 不同模型在 CVC-ClinicDB 数据集下检测对比

### 3.5 消融实验

表7、表8记录了不同数据集在不同改进下算法的预测率,召回率,  $mAP$  和平均检测时间.对YOLOv4使用CSPDarknet-49, CBAM模块进行优化.从表7得知,YOLOv4模型使用改进的CSPDarknet-49在准确度上分别提升了1.3%,平均检测时间减少2ms.这是

因为减少了CSPDarknet的卷积层数可以缓解小目标在多次卷积后边缘信息的丢失,同时降低了网络的参数量和训练量,加快模型的收敛速度,并提升模型的准确精度.对比表7,表8可知,在YOLOv4模型嵌入CBAM模块可以使准确率分别提升4.48%和1.11%.这是因为YOLOv4加入了注意力模块后,可以更丰富的提取特征层的特征,增强特征融合并提高检测精度.综合上述改进点后,相比于原YOLOv4模型,本文提出的YOLOv4-polyp模型在保持检测速度的同时分别提高了5.21%和2.05%,说明了改进的YOLOv4-polyp模型的有效性.

表7 在 Hyper-Kvasir 数据集下的消融实验

模型	CSPDarknet-49	CBAM	P (%)	R (%)	$mAP$ (%)	平均检测速度 (s)
YOLOv4	×	×	88.78	81.25	86.82	0.030
YOLOv4+	√	×	91.84	80.36	88.12	0.028
CSPDarknet-49	×	√	94.88	81.61	91.30	0.033
YOLOv4-polyp	√	√	93.73	85.23	92.03	0.031

表8 在 CVC-ClinicDB 数据集下的消融实验

模型	CSPDarknet-49	CBAM	P (%)	R (%)	$mAP$ (%)	平均检测速度 (s)
YOLOv4	×	×	92.86	90.70	92.97	0.030
YOLOv4+	√	×	93.60	90.70	93.73	0.029
CSPDarknet-49	×	√	95.90	90.70	94.08	0.032
YOLOv4-polyp	√	√	95.97	92.25	95.02	0.031

## 4 结论与展望

针对息肉实时检测在面对小型息肉检测困难、检测精度低等问题,本文对此进行了研究,综述了目前息肉检测的方法,提出了YOLOv4-polyp胃息肉检测算法,并将其应用到息肉检测中.本文算法使用YOLOv4作为基础结构,在主干网络上使用CSPDarknet-49作为特征提取网络,在提高检测速度的同时降低网络模型参数;结合了CBAM卷积注意力模块,使网络可以在空间和通道两方面更加注意目标特征信息;根据息肉数据集特点,使用K-means++聚类算法对息肉数据集进行聚类分析得到合适的锚框,使网络不仅适合检测常规大小胃息肉,也适合检测小型胃息肉.相对原YOLOv4模型,在保持检测速率的同时提高了对小型息肉的准确率.在之后的工作中,将进一步改善检测效果,优化网络模型检测速度,在保证精度的同时提高检测速度.

## 参考文献

- 1 Bernal J, Tajkbaksh N, Sánchez FJ, *et al.* Comparative validation of polyp detection methods in video colonoscopy: Results from the MICCAI 2015 endoscopic vision challenge. *IEEE Transactions on Medical Imaging*, 2017, 36(6): 1231–1249. [doi: [10.1109/TMI.2017.2664042](https://doi.org/10.1109/TMI.2017.2664042)]
- 2 Felzenszwalb PF, Girshick RB, Mcallester D, *et al.* Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627–1645. [doi: [10.1109/TPAMI.2009.167](https://doi.org/10.1109/TPAMI.2009.167)]
- 3 van der Post RS, van Dieren J, Grelack A, *et al.* Outcomes of screening gastroscopy in first-degree relatives of patients fulfilling hereditary diffuse gastric cancer criteria. *Gastrointestinal Endoscopy*, 2018, 87(2): 397–404.e2. [doi: [10.1016/j.gie.2017.04.016](https://doi.org/10.1016/j.gie.2017.04.016)]
- 4 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016. 21–37.
- 5 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 779–788.
- 6 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014. 580–587.
- 7 He KM, Zhang XY, Ren SQ, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904–1916. [doi: [10.1109/TPAMI.2015.2389824](https://doi.org/10.1109/TPAMI.2015.2389824)]
- 8 马燕. 动态肠镜息肉图像实时定位算法研究与实现 [硕士学位论文]. 郑州: 郑州大学, 2020.
- 9 王博, 张丽媛, 师为礼, 等. 改进的 M2det 内窥镜息肉检测方法. *计算机工程与应用*, 2022, 58(2): 193–200. [doi: [10.3778/j.issn.1002-8331.2008-0084](https://doi.org/10.3778/j.issn.1002-8331.2008-0084)]
- 10 Laddha M, Jindal S, Wojciechowski J. Gastric polyp detection using deep convolutional neural network. *Proceedings of the 2019 4th International Conference on Biomedical Imaging, Signal Processing*. Nagoya: ACM, 2019. 55–59.
- 11 Cao CT, Wang RL, Yu Y, *et al.* Gastric polyp detection in gastroscopic images using deep neural network. *PLoS One*, 2021, 16(4): e0250632. [doi: [10.1371/journal.pone.0250632](https://doi.org/10.1371/journal.pone.0250632)]
- 12 Yu T, Lin N, Zhang X, *et al.* An end-to-end tracking method for polyp detectors in colonoscopy videos. *Artificial Intelligence in Medicine*, 2022, 131: 102363. [doi: [10.1016/j.artmed.2022.102363](https://doi.org/10.1016/j.artmed.2022.102363)]
- 13 Taş M, Yılmaz B. Super resolution convolutional neural network based pre-processing for automatic polyp detection in colonoscopy images. *Computers & Electrical Engineering*, 2021, 90: 106959.
- 14 Cheng MJ, Kong ZS, Song GL, *et al.* Learnable oriented-derivative network for polyp segmentation. *Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Strasbourg: Springer, 2021: 720–730.
- 15 任莉莉, 边璇, 王光磊, 等. GLIA-NET: 基于深度学习的息肉分割网络. *计算机工程*, 2022: 1–12. [doi: [10.19678/j.issn.1000-3428.0062860](https://doi.org/10.19678/j.issn.1000-3428.0062860)]
- 16 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. *arXiv:2004.10934*, 2020.
- 17 Arthur D, Vassilvitskii S. K-means++: The advantages of careful seeding. *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*. New Orleans: Society for Industrial and Applied Mathematics, 2007. 1027–1035.
- 18 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 3–19.
- 19 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 20 Borgli H, Thambawita V, Smedsrud PH, *et al.* HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific Data*, 2020, 7(1): 283. [doi: [10.1038/s41597-020-00622-y](https://doi.org/10.1038/s41597-020-00622-y)]
- 21 Bernal J, Sánchez FJ, Fernández-Esparrach G, *et al.* WMDOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 2015, 43: 99–111. [doi: [10.1016/j.compmedimag.2015.02.007](https://doi.org/10.1016/j.compmedimag.2015.02.007)]

(校对责编: 牛欣悦)