

基于 Mask R-CNN 的胃肠息肉图像增强检测

胡一凡,肖满生,范明凯,吴宇杰
(湖南工业大学 计算机学院,湖南 株洲 412007)

摘要:对胃肠息肉图像进行检测,是医学影像识别中的一个重要部分,在胃肠息肉图像中检测不显著息肉和小目标息肉更是一个难点。为了提升胃肠息肉图像的检测率,减少对息肉病理的误判,提出了一种基于 Mask R-CNN 的胃肠息肉增强检测模型。模型网络结构上采用残差网络和特征金字塔网络对图像进行多尺度特征提取,在残差网络的卷积层中改进了使用了可变形卷积以扩大模型的采样范围,采样后的特征图进一步输入区域候选网络中对息肉区域筛选。模型的预测输出部分设计了一个增强图像边缘强度的检测模块,通过滑动窗口二次检测的方式,增强对不显著息肉和小目标息肉的检测。模型能同时进行息肉图像检测和分割,并在 Kvasir-SEG 与 CVC-ClinicDB 两个息肉数据集上进行了验证,实验结果表明提出的方法能够在原有模型的基础上提升对息肉病灶的检测精确度和分割精确度,检测精确度分别达到了 94.3% 和 98.6%,分割精确度分别达到了 93.7% 和 98.7%,优于其他的对比模型。

关键词:Mask R-CNN;可变形卷积;边缘强度;息肉图像检测;图像分割

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2023)03-0173-07

doi:10.3969/j.issn.1673-629X.2023.03.026

Mask R-CNN Based Image Enhancement Detection of Gastrointestinal Polyps

HU Yi-fan, XIAO Man-sheng, FAN Ming-kai, WU Yu-jie
(School of Computer Science, Hunan University of Technology, Zhuzhou 412007, China)

Abstract: The detection of gastrointestinal polyps is an important part of medical image recognition, and the detection of non-salient polyps and small target polyps in gastrointestinal polyp images is even more difficult. In order to improve the detection rate of gastrointestinal polyp images and reduce the misclassification of polyp pathology, a Mask R-CNN based enhanced detection model for gastrointestinal polyps is proposed. The model uses a residual network and a feature pyramid network in the network structure for multi-scale feature extraction, and a deformable convolution is improved in the convolution layer of the residual network to extend the sampling range of the model, and the sampled feature maps are further input into the region proposal network for polyp region screening. The predictive output part of the model is designed to enhance the detection of non-salient polyps and small target polyps by means of a sliding window secondary detection module with enhanced image edge strength. The model was validated on two polyp datasets, Kvasir-SEG and CVC-ClinicDB, and the experimental results showed that the proposed method could improve the detection and segmentation precision of polyp lesions based on the original model, the detection precision reached 94.3% and 98.6%, and segmentation precision reached 93.7% and 98.7%, respectively, outperforming other compared models and validating the effectiveness of the model.

Key words: Mask R-CNN; deformable convolution; edge strength; polyp image detection; image segmentation

0 引言

内窥镜检查是耗费精力的一个项目,息肉检查通常也通过内窥镜实现,医生人为的内窥镜检查需要高度的手眼协调,耗费时间与精力,通过计算机技术辅助并减轻人为负担是一个发展的趋势。近几年深度学习技术给计算机辅助诊断带来了新的契机,目前一些在

通用检测领域使用的深度学习模型,如:Faster R-CNN^[1]、SSD^[2]、YOLO^[3]、RetinaNet^[4]等,都在目标检测上有着优良的表现,能够很好地迁移到医学领域。

目前,息肉图像检测的方法分为传统图像处理方法和深度学习方法。传统的图像处理方法需要对采集的息肉图像做一定的预处理排除干扰,同时需要进行

收稿日期:2022-06-07

修回日期:2022-10-09

基金项目:湖南省自然科学基金项目(2021JJ50049,2022JJ50077)

作者简介:胡一凡(1997-),男,硕士研究生,研究方向为图像处理和计算机视觉;通讯作者:肖满生(1968-),男,教授,研究方向为智能计算和智能信息处理。

手工特征提取以找出息肉特征,最后采用分类器进行分类。传统图像处理中提取的特征通常有 SIFT 特征^[5]、HOG 特征^[6]等,这些特征提取方法需要依靠人为的先验知识进行设计,缺乏通用性,鲁棒性差,难以实现方法的迁移以及端到端学习。深度学习方法目前是以 CNN(卷积神经网络)为基础进行拓展的目标检测网络,深度学习的优点在于不需要人工设计算法去提取图像的特征,通过 CNN 多次卷积和池化能够提取出图像不同的特征,CNN 的深度决定了能否提取更多的特征,目前在图像的分类^[7]、检测^[8]和分割^[9]领域都有许多成熟的应用。同时近年来出现的以 Vision Transformer^[10]为主的方法,将 Transformer^[11]从自然语言处理领域迁移到了计算机视觉领域,提出了一种新的检测思路,具有代表性的检测方法有 DETR^[12],也是一种极具发展潜力的目标检测模型。

在医学息肉的目标检测方面,Mo X 等人^[13]使用了 Faster R-CNN 在内窥镜息肉数据集上进行检测,获得了较高的精确度,实验结果对检测中等和大面积的息肉效果较好,但是对小目标息肉的检测效果不佳,并且在对比度不高和含有镜面反光的区域中也无法正确地识别息肉。Ohrenstein DC 等^[14]提出了一种 SSD 联合 GAN 的检测方法,GAN^[15]用于提升数据集中小目标的分辨率,在保证 SSD 检测的速度上,提升了检测小目标息肉的精确度,但依旧有可提升的空间。王博等人^[16]提出了一种改进的 M2det 的内窥镜息肉检测方法,使用了 FFM 模块融合主干网络特征,充分利用了图像特征,同时增加了 scSENet 注意力机制保留了有效特征,使得检测效率优于许多息肉检测方法,但是

模型对一些对比度低且形状不规则的息肉图像依然存在误检。Shen Zhiqiang 等人^[17]使用了一种 COTR 方法,将 CNN 和 Transformer 结合用于息肉检测,该方法促进了特征重建并且加速了模型的收敛,获得了与 CNN 检测相当的性能,但这种方法检测固着性息肉时产生的置信度较低,可能会导致模型漏检。

依据上述的研究现状,提出了一种基于 Mask R-CNN 胃肠息肉增强检测方法。在 Mask R-CNN 的特征提取网络上提出替换部分卷积采用可变形卷积从而自适应特征采样的范围,模型之外提出了一种增强图像边缘强度的增强检测模块(CLAHE Enhanced Detection, CED),通过滑动窗口的方式对原图进行切分和对对比度增强,有效地对息肉图像中的中小目标和不显著目标进行检测。

1 Mask R-CNN 模型介绍

Mask R-CNN^[18]是一个能进行多任务的图像检测模型^[19],模型能同时进行目标检测任务和实例分割^[20]任务。Mask R-CNN 主要部分由特征提取网络、RPN(区域候选网络)、ROI Align 和全连接网络四个部分组成,网络结构如图 1 所示。Mask R-CNN 从整体上可以看作 Faster R-CNN 与 FCN(全卷积网络)^[21]的组合,模型先对图像做目标检测,图像经过特征提取网络进行多尺度特征提取,并通过 RPN 确定目标的种类和候选框,得到每个息肉的具体位置与特征图,在此基础上通过 FCN 执行语义分割,使每一个检测框内对应生成当前类别的目标分割结果。

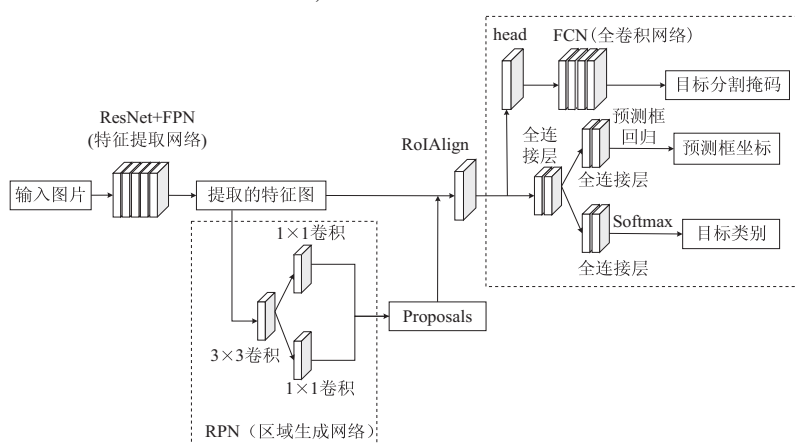


图 1 Mask R-CNN 模型结构

特征提取网络主要使用 CNN 进行多层的卷积获取输入图像的特征图。Mask R-CNN 采用了 ResNet(残差网络)^[22]和 FPN(特征金字塔网络)^[23]组成特征提取网络,FPN 能够输出不同尺度的特征,提供了多尺度的特征图,ResNet 具有更好的语义特征提取性能,能够保证一定网络深度的同时避免网络退化,同时

ResNet 中 Residual block(残差块)的简易结构,使得 ResNet 能够很好地与其他特征提取模块结合。ResNet 和 FPN 的模块组合,使得模型在图像中够提取到更加丰富的特征,对各类目标的检测率都有提升。

RPN 是由全卷积组成的网络,它为图像生成一系列目标候选框,以便后续进一步筛选,保证模型检测精

确度。RPN 预先定义一系列的候选框,通过锚点滑动窗口的方式,在特征图的每一个区域上滑过,并由特征图映射回原图,在原图上生成候选框,并判断哪些候选框中包含目标,这样能较精准地确定图片中目标的位置。RPN 默认定义了3种尺度的候选框,每一种尺度的候选框设置了3种比例,能够生成9种不同比例尺度的候选框,能够覆盖到绝大多数的目标。

ROIAAlign 是 Mask R-CNN 模型对 ROI Pooling 结构的改进, RPN 在特征图上生成候选框后通过 ROI Pooling 将对应的候选框位置池化为固定大小的特征图,再输入后续的全连接层进行目标分类和边框坐标回归, ROI Pooling 在池化操作中采取了取整的步骤,使得池化后的特征图丢失了部分细节,对 Mask R-CNN 的分割操作有很大的影响, ROIAAlign 通过双线性插值的方法算出特征图对应的像素值,避免了 ROI Pooling 的取整步骤,优化了模型结构。

2 可变形卷积 Mask R-CNN

可变形卷积(Deformable Convolution)^[24]是应用在特征提取网络中的一种特殊卷积,它增加了 CNN 的几何变换能力。通常 CNN 中卷积的结构是固定的,卷积在图像中滑动采样的过程中,受限于标准卷积的结构,模型的特征提取范围存在局限性,对未知目标的变化适应性差,泛化能力弱。可变形卷积在标准卷积的结构上引入了偏移量来增加空间采样范围,在原卷积核的每一个元素上增加了一个方向参数,使得卷积核在模型训练时自适应地扩大采样范围。

二维卷积在特征图上作用分两步:(1)在输入的特征图上使用标准的网格 R 进行采样;(2)将采样的结果乘上对应的权重 w 并进行求和。一个标准 3×3 的二维卷积如图 2(a)所示,网格 R 代表了二维卷积的采样范围, R 为 $\{(-1, -1), (-1, 0), \dots, (0, 1), (1,$

$1)\}$, 对于特征图 Y 上每个位置 P_0 , 都由公式(1)表示, 其中 P_n 对应了二维卷积中 R 每一个位置。可变形卷积在公式(1)的基础上增加了偏移量 $\{\Delta P_n | n = 1, 2, \dots, N\}$, 且 $N = |R|$, 如公式(2)所示, 原二维卷积的采样范围中增加了不规则的偏移 ΔP_n , 随着偏移量的取值和方向不同, 使得二维卷积的采样范围随给定的偏移量变化。

$$Y(P_0) = \sum_{P_n \in R} w(P_n) \cdot x(P_0 + P_n) \quad (1)$$

$$Y(P_0) = \sum_{P_n \in R} w(P_n) \cdot x(P_0 + P_n + \Delta P_n) \quad (2)$$

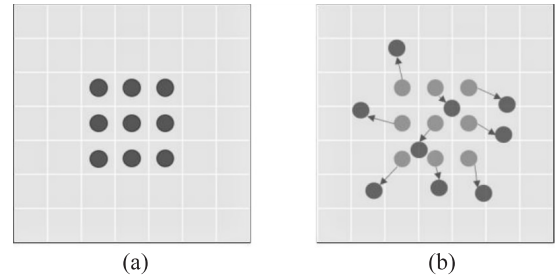


图2 特征图中二维卷积的采样范围(a)和可变形卷积的其中一种情况,通过引入偏移量 ΔP_n 扩大了采样范围(b)

文中 Mask R-CNN 使用了 ResNet50 作为特征提取网络, ResNet50 被分为5个部分,每个部分由多个不同通道数的 Residual block 叠加,如图3所示。模型改进使用可变形卷积来替换原网络中的标准卷积,由于可变形卷积引入的偏移量 ΔP_n 需要通过特征提取网络反向传播学习来确定,引入了一定的计算量,因而不适合将所有的标准卷积进行替换,经过实验验证,最后将 ResNet50 中后三个卷积组中的标准卷积替换成可变形卷积能获得比较好的效果。替换的可变形卷积增强了模型本身适应尺度变换的能力,模型可以更好地提取图像中目标形状的完整特征。

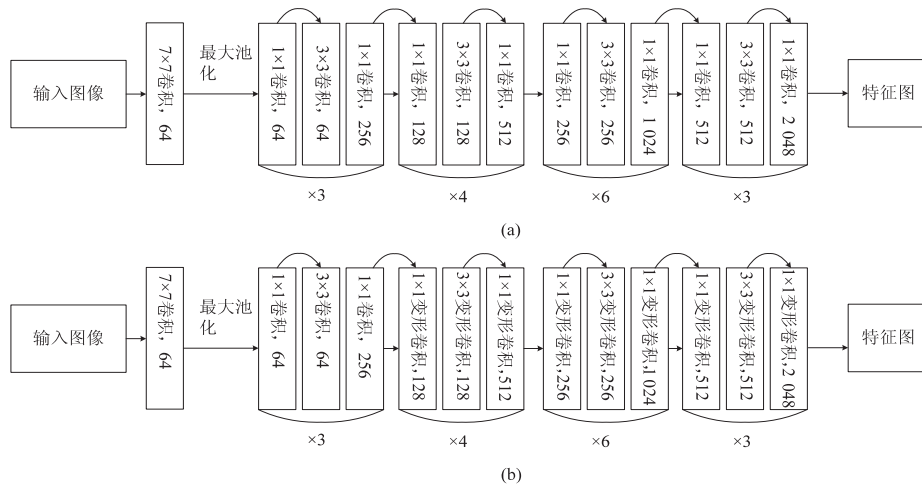


图3 ResNet50 特征提取网络(a)和替换了可变形卷积的 ResNet50 分别在后三个卷积组上使用了可变形卷积(b)

3 增强检测模块

3.1 对比度限制的自适应直方图均衡化 (CLAHE)

一些胃肠息肉图像中,检测目标与背景有极其相似的特征,这类目标称为不显著目标,目视检测的方法很难察觉到这些特征,对于通用的目标检测模型而言检测此类目标也有一定的难度。因此采取调整对比度的方法增强不显著目标与背景的边缘强度,扩大不显著目标与背景的差异,能有效提高模型的检测率。

直方图均衡化是一种提升对比度的常用方法,通过灰度直方图的分布得到映射曲线,再通过灰度概率分布实现图像的空间映射。对于离散的一幅灰度图像 gray,计算图像中灰度值为 i 的像素出现的概率有公式(3),其中 L 是图像所有的灰度数, N 是图像中所有的像素数, n_i 为像素值为 i 的像素数,计算所得 $P_{\text{gray}}(i)$ 是像素值为 i 的图像直方图。将 P_{gray} 应用于累积分布函数可得公式(4),即为灰度图像的累计归一化直方图。

$$p_{\text{gray}}(i) = p(\text{gray} = i) = \frac{n_i}{N}, \quad 0 \leq i \leq L \quad (3)$$

$$\text{CDF}_{\text{gray}}(i) = \sum_{j=0}^i p_{\text{gray}}(j) \quad (4)$$

将公式(4)的直方图运用于 RGB 三个图像通道上,可对彩色图像进行直方图均衡化,能很好地提高图像整体的对比度。

在直方图均衡化的基础上,CLAHE^[25]是自适应直方图均衡方法上的改进,CLAHE 在应用累积分布函数前对直方图进行了预定义值的裁剪,从而限制了高频噪声区域的局部信息增强,从而抑制噪声。对息肉图像而言, R 通道的像素与彩色图像原图接近,所以使用 CLAHE 对图像的 G 、 B 二通道进行增强,使目标的颜色发生一定偏移,从而减小息肉目标与背景的相似度。图 4 是使用 CLAHE 方法对部分息肉图像的增强效果,对一些不显著的息肉目标可以起到很好的增强边缘效果。

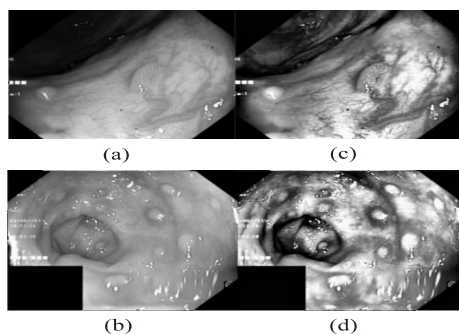


图 4 息肉数据集中原图像(a)(b)和使用了 CLAHE 增强后的息肉图像(c)(d)

3.2 增强边缘强度的检测模块 (CED)

在目标检测任务中小目标的检测是一个难点^[26],

目标检测的定义中小目标是指像素面积小于 32×32 个像素点,难点在于小目标所占的像素点非常少,传统模型的检测框通常针对中等目标和大型目标设计,无法很好地定位,且一些小目标在图像上的特征表现过于稀少且不显著,检测模型在定位的过程中很难判断是目标还是背景,导致检测阈值过低而被模型过滤。针对这个问题,在 Mask R-CNN 模型的基础上提出了一种结合 CLAHE 的滑动窗口增强检测模块,模型首先对图像全局进行检测,找出中等、大型息肉,在此基础上对图像进行滑动切片操作,将图像分成若干区域,再对若干区域进行双线性插值。双线性插值是一种图像上采样的方法,该方法通过图像已知的像素点进行插值计算确定出未知的像素点,目的是为了放大小目标特征。图像的双线性插值操作如下:已知图像中四个点的像素值 $f(P_{11})$, $f(P_{12})$, $f(P_{21})$, $f(P_{22})$ 和位置 $P_{11} = (x_1, y_1)$, $P_{12} = (x_1, y_2)$, $P_{21} = (x_2, y_1)$, $P_{22} = (x_2, y_2)$ 要计算上述四个点所包含的未知点 $Q(x, y)$ 的像素值。首先通过公式(5)和公式(6)对 Q 的 x 方向进行插值,算出与 Q 点在相同 x 坐标下的 Q_1 、 Q_2 点的像素值,再通过公式(7)对 Q 的 y 方向进行插值,即可算出 Q 点位置上的像素值。

$$f(Q_1) \approx \frac{x_2 - x}{x_2 - x_1} f(P_{11}) + \frac{x - x_1}{x_2 - x_1} f(P_{21}) \quad (5)$$

$$f(Q_2) \approx \frac{x_2 - x}{x_2 - x_1} f(P_{12}) + \frac{x - x_1}{x_2 - x_1} f(P_{22}) \quad (6)$$

$$f(Q) \approx \frac{y_2 - y}{y_2 - y_1} f(Q_1) + \frac{y - y_1}{y_2 - y_1} f(Q_2) \quad (7)$$

双线性插值上采样后像素的填充可能导致图像的边缘强度减弱,通过 CLAHE 提升对比度同时对图像边缘进行增强,从而能够使模型更好地检测小目标和不显著目标。综上所述,增强检测模块的具体流程(见图 5)如下:

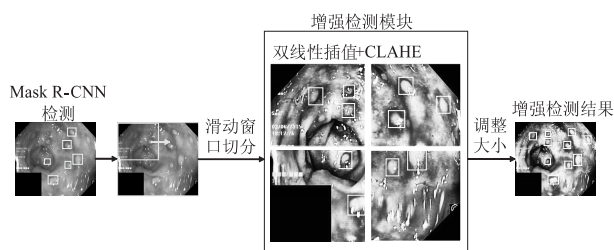


图 5 CED 增强检测模块的过程

(1) 检测图像先输入 Mask R-CNN 模型进行初步识别,得到部分检测结果。

(2) 根据输入图像的大小,确定合适的滑动窗口大小 n ,使用 $n \times n$ 的滑动窗口对原图像进行切片操作。

(3) 对切片的图像采用双线性插值进行适当上采样,上采样后的图片进行 CLAHE 增强后分别送入模

型进行增强检测。

(4)将步骤(3)增强检测的结果与步骤(1)初步检测的结果融合,过滤检测框重叠的结果,取置信度较高的检测框,调整大小最后输出增强检测结果。

4 实验及结果

4.1 数据集

该文采用了 Kvasir-SEG^[27] 数据集与 CVC-ClinicDB^[28] 2 个胃肠息肉图像数据集进行实验。Kvasir-SEG 数据集包含 1 000 张息肉类的原始图像并提供分割掩码图片和边界框标签,由经验丰富的胃肠病学家手动注释和验证,适用于图像逐像素分割与目标检测。CVC-ClinicDB 是一个公开的内窥镜图像数据集,包含来自 31 个结肠镜序列的 612 张分辨率为 384×288 的图像,图像提取来自结肠镜息肉检测视频,并提供原图和息肉相应区域的掩码信息,此数据集常用于医学图像分割,也是 MICCAI 2015 Sub-Challenge on Automatic Polyp Detection Challenge in Colonoscopy Videos 使用的官方数据集。

4.2 数据预处理

Kvasir-SEG 和 CVC-ClinicDB 数据集是从人体的腹腔中采集而来,腹腔中存在着大量粘液,在采集过程中由于光照会产生反光,反光现象一定程度上遮挡了图像的部分特征,引入了过多噪声。为了改善这种现象,使用了一种图像去反光的算法^[29]进行预处理,对一些具有强烈反光影响的息肉图像修复效果较好。该算法主要步骤是检测息肉图像中具有强烈反光的部分,然后计算周边非反光部分的像素值,并将这些像素值进行加权平均对反光部分进行填充。

4.3 评价指标

改进后的 Mask R-CNN 的实验结果包含目标检测与实例分割两个部分,且数据集训练和实验的标签转换成了 COCO 数据格式,因此采样相同的评价指标^[30],该指标能更加全面地评价模型。目标检测任务下,采用 AP^{bbox} (边界框平均精确度)和 AR^{bbox} (边界框平均召回率)作为模型性能评价指标, AP_{50}^{bbox} 代表 IOU (交并比) 阈值取 0.5 时模型的边界框平均精确度, AP_s^{bbox} 、 AP_m^{bbox} 、 AP_l^{bbox} 分别代表模型在小目标(像素面积小于 32×32)、中目标(像素面积 32×32 到 96×96)、大目标(像素面积大于 96×96)上的边界框平均精确度。实例分割任务下,采用 AP^{mask} (掩码平均精确度)和 AR^{mask} (掩码平均召回率)来评价模型, AP_{50}^{mask} 为 IOU 阈值为 0.5 时的掩码平均精确度, AP_s^{mask} 、 AP_m^{mask} 、 AP_l^{mask} 则分别对应小、中、大息肉目标的掩码平均精确度。

4.4 实验环境与训练过程

实验环境为 Ubuntu20.04, Intel Xeon E5-2678 v3

(6)@ 3.300 GHz, GPU NVIDIA RTX 2080Ti, CUDA 10.2, 深度学习框架为 Pytorch, 在 Kvasir-SEG 数据集上训练了 4 800 个迭代数, 在 CVC-ClinicDB 数据集上训练了 2 400 个迭代数, 初始学习率设为 0.02, 随训练缓慢递减。

改进后 Mask R-CNN 模型在两个数据集上训练情况曲线记录在图 6, 模型在两个训练集上的 loss 值随着训练迭代数的增加由急剧下降过渡到逐渐平缓, 表明模型在不断的拟合训练集样本, 有利于提升模型对测试集样本的检测精确度。

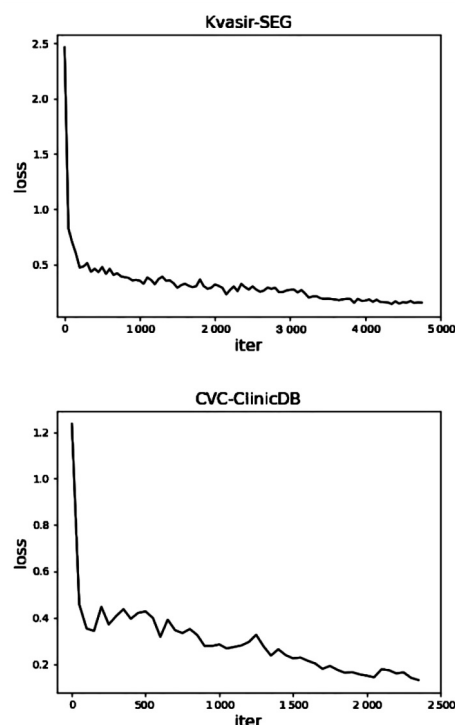


图6 改进后的 Mask R-CNN 在数据集上的训练损失值

4.5 实验结果与分析

模型对 Kvasir-SEG 与 CVC-ClinicDB 数据集分别实验, 训练集和测试集比例设为 8 : 2, 在 Kvasir-SEG 数据集的实验中随机选取了 800 张图片作为训练集, 200 张图片作为测试集, 实验中由于 Kvasir-SEG 数据集中小目标的数量较少, 评测代码无法对 AP_s 进行计算。在 CVC-ClinicDB 数据集中随机选择了 489 张图片作为训练集, 122 张图片作为测试集。对于目标检测任务, 选择了 5 个其他检测模型进行评测对比, 同时还做了三个基于 Mask R-CNN 的改进对比实验。对于实例分割任务只进行了 Mask R-CNN 改进对比实验, 目标检测和实例分割的试验结果分别由表 1、表 2 所示。

由表 1 所示, 在目标检测任务下, 改进后的模型 Mask R-CNN(DCN+CED) 在大部分指标下均优于所列出的其他 5 个目标检测模型, 对比改进前的 Mask R-CNN 模型, 可以发现可变形卷积和 CED 模块对原模

型的平均精确度进行了有效的提升,同时从 CVC-ClinicDB 数据集的实验指标 AP_s^{bbox} 值中可看出,Mask R-CNN(DCN+CED) 模型在小目标检测上的表现更

好,均优于提出的 5 个其他的模型和改进前模型,达到了 75% 的平均精确度。

表 1 目标检测实验结果

Model	Backbone	Kvasir-SEG(200 Images)					CVC-ClinicDB(122 Images)				
		AP_{50}^{bbox}	AR^{bbox}	AP_s^{bbox}	AP_m^{bbox}	AP_l^{bbox}	AP_{50}^{bbox}	AR^{bbox}	AP_s^{bbox}	AP_m^{bbox}	AP_l^{bbox}
Faster R-CNN	ResNet-50	90.3	76.6	—	70.6	65.9	95.6	73.2	67.0	70.5	61.1
SSD	VGG-16	88.0	66.0	—	45.2	56.6	98.3	69.3	67.4	66.2	61.8
YOLOv3	DarkNet-53	87.1	62.8	—	54.7	54.7	97.4	67.9	52.3	63.9	58.7
YOLOX	CSPDarkNet-53	90.4	76.6	—	74.0	67.1	98.7	80.1	71.4	77.8	77.6
Dynamic R-CNN	ResNet-50	90.1	72.3	—	68.5	63.1	97.8	73.4	62.4	73.0	66.5
Mask R-CNN	ResNet-50	91.0	77.6	—	76.9	67.7	97.0	76.1	60.8	69.9	71.1
Mask R-CNN(DCN)	ResNet-50	93.6	77.1	—	76.5	68.9	97.7	81.3	72.4	75.8	79.4
Mask R-CNN(DCN+CED)	ResNet-50	94.3	78.7	—	81.0	70.8	98.6	81.3	75.0	75.9	79.7

表 2 实例分割实验结果

Model	Backbone	Kvasir-SEG(200 Images)					CVC-ClinicDB(122 Images)				
		AP_{50}^{mask}	AR^{mask}	AP_s^{mask}	AP_m^{mask}	AP_l^{mask}	AP_{50}^{mask}	AR^{mask}	AP_s^{mask}	AP_m^{mask}	AP_l^{mask}
Mask R-CNN	ResNet-50	90.2	79.1	—	68.1	71.4	97.7	80.6	62.9	76.7	76.2
Mask R-CNN(DCN)	ResNet-50	93.4	80.3	—	74.4	73.3	98.8	83.6	72.6	80.6	82.5
Mask R-CNN(DCN+CED)	ResNet-50	93.7	80.9	—	76.3	74.3	98.7	84.4	74.1	81.0	82.2

表 2 主要在实例分割任务下对 Mask R-CNN 改进前后进行了对比,Mask R-CNN(DCN) 模型相较于原 Mask R-CNN 模型提升较大,在两个数据集的所有指标上提升较为明显,Mask R-CNN(DCN+CED) 相较于 Mask R-CNN(DCN) 仅在 CVC-ClinicDB 的实验指标 AP_s^{mask} 上略有提升, AP_s^{mask} 值由 72.6% 提升到了 74.1%,在其他指标上,CED 模块带来的提升不是特别明显,但相比于原 Mask R-CNN 模型仍有较好提升。

图 7 可视化了 Mask R-CNN(DCN+CED) 在不同 IOU 阈值下的 PR 曲线,可以看出在 IOU 值为 0.5 和 0.75 时模型在两个数据集上的表现良好,在 IOU 值为 0.9 时,模型表现欠佳,可以说明 Mask R-CNN(DCN+CED) 在较严格的测试条件下对两个数据集仍有较好的检测性能。

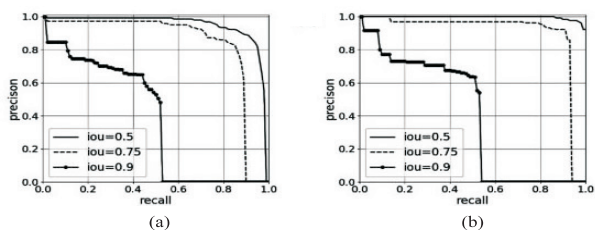


图 7 Kvasir-SEG 数据集在 IOU 分别为 0.5、0.75、0.9 时的 PR 曲线(a)和 CVC-ClinicDB 数据集在 IOU 分别为 0.5、0.75、0.9 时的 PR 曲线(b)

图 8 展示了 Mask R-CNN(DCN+CED) 在两个数

据集上的检测结果的可视化图,上三幅图来源于 Kvasir-SEG 数据集,下三幅图来源于 CVC-ClinicDB 数据集。从可视化结果可以看出,模型对不同形状大小的息肉检测效果良好,且对一些具有反光干扰下的息肉图片也有较好的检测效果,凸显了改进的有效性。

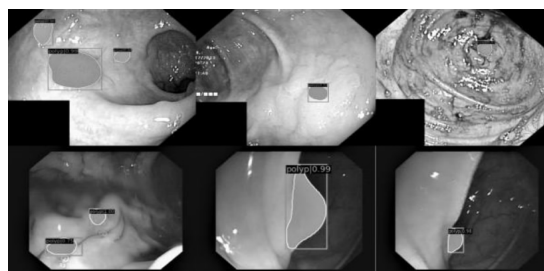


图 8 Kvasir-SEG 和 CVC-ClinicDB 部分测试集检测结果

5 结束语

该文提出了一种基于 Mask R-CNN 的增强检测模型用于更好地对息肉图像进行检测和分割,在特征提取网络中引入了可变形卷积提高了模型对图片的特征提取范围,同时设计了一种增强检测模块,在两个数据集上的验证实验和模型对比实验均表明可变形卷积和所提出的增强检测模块对小面积息肉和不显著息肉的检测有明显的提升效果。另外提出的增强检测模块在理论上能够迁移到其他检测模型,需要根据不同任务做出一定的参数和方法调整,增强检测需要花费多

余的时间进行切分和检测,后续将针对检测时间较长的问题进行一定的改进,同时也将继续研究将模块进一步融入模型以获得更好的兼容性。

参考文献:

- [1] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *Advances in Neural Information Processing Systems*, 2015, 28: 91-99.
- [2] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//*European conference on computer vision*. [s. l.]: Springer, 2016: 21-37.
- [3] 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*. Las Vegas: IEEE, 2016: 779-788.
- [4] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//*Proceedings of the IEEE international conference on computer vision*. Venice: IEEE, 2017: 2980-2988.
- [5] 王孝义, 邢素霞, 申楠, 等. 基于 SIFT 算法的乳腺肿块分类 [J]. *计算机仿真*, 2022, 39(1): 446-450.
- [6] 林克正, 张元铭, 李昊天. 信息熵加权的 HOG 特征提取算法研究 [J]. *计算机工程与应用*, 2020, 56(6): 147-152.
- [7] 张珂, 冯晓晗, 郭玉荣, 等. 图像分类的深度卷积神经网络模型综述 [J]. *中国图象图形学报*, 2021, 26(10): 2305-2325.
- [8] 谢富, 朱定局. 深度学习目标检测方法综述 [J]. *计算机系统应用*, 2022, 31(2): 1-12.
- [9] 张鑫, 姚庆安, 赵健, 等. 全卷积神经网络图像语义分割方法综述 [J]. *计算机工程与应用*, 2022, 58(8): 45-57.
- [10] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale [J]. *arXiv*: 2010. 11929, 2020.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//*Advances in neural information processing systems*. Long Beach: [s. n.], 2017: 5998-6008.
- [12] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]//*European conference on computer vision*. Glasgow: Springer, 2020: 213-229.
- [13] MO X, TAO K, WANG Q, et al. An efficient approach for polyps detection in endoscopic videos based on faster R-CNN [C]//*2018 24th international conference on pattern recognition (ICPR)*. Beijing: IEEE, 2018: 3929-3934.
- [14] OHRENSTEIN D C, BRANDAO P, TOTH D, et al. Detecting small polyps using a dynamic SSD-GAN [J]. *arXiv*: 2010. 15937, 2020.
- [15] 李祥霞, 谢嫻, 李彬, 等. 生成对抗网络在医学图像处理中的应用 [J]. *计算机工程与应用*, 2021, 57(18): 24-37.
- [16] 王博, 张丽媛, 师为礼, 等. 改进的 M2det 内窥镜息肉检测方法 [J]. *计算机工程与应用*, 2022, 58(2): 193-200.
- [17] SHEN Z, LIN C, ZHENG S. COTR: convolution in transformer network for end to end polyp detection [J]. *arXiv*: 2105. 10925, 2021.
- [18] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//*Proceedings of the IEEE international conference on computer vision*. Cambridge: IEEE, 2017: 2961-2969.
- [19] WANG X, KONG T, SHEN C, et al. SOLO: segmenting objects by locations [C]//*European conference on computer vision*. Glasgow: Springer, 2020: 649-665.
- [20] 李晓筱, 胡晓光, 王梓强, 等. 基于深度学习的实例分割研究进展 [J]. *计算机工程与应用*, 2021, 57(9): 60-67.
- [21] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. Boston: IEEE, 2015: 3431-3440.
- [22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas: IEEE, 2016: 770-778.
- [23] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. Honolulu: IEEE, 2017: 2117-2125.
- [24] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks [C]//*Proceedings of the IEEE international conference on computer vision*. Venice: IEEE, 2017: 764-773.
- [25] LI C, ZHAO X, WAN Y. The camouflage color target detection with deep networks [C]//*International conference on neural information processing*. Guangzhou: Springer, 2017: 476-485.
- [26] VAN ET TEN A. You only look twice: Rapid multi-scale object detection in satellite imagery [J]. *arXiv*: 1805. 09512, 2018.
- [27] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-seg: a segmented polyp dataset [C]//*International conference on multimedia modeling*. Daejeon: Springer, 2020: 451-462.
- [28] 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.
- [29] ARNOLD M, GHOSH A, AMELING S, et al. Automatic segmentation and inpainting of specular highlights for endoscopic imaging [J]. *EURASIP Journal on Image and Video Processing*, 2010, 2010: 814319.
- [30] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context [C]//*European conference on computer vision*. Zurich: Springer, 2014: 740-755.