DOI:10.19651/j.cnki.emt.2211929

YOLOF-CBAM:一种新的结直肠息肉实时分类与检测方法*

杨 昆[1,2,3] 孙宇锋[4] 汪世伟[1] 路宇飞[1] 薛林雁[1,2,3]
(1.河北大学质量技术监督学院 保定 071002;2.计量仪器与系统国家地方联合工程研究中心 保定 071002;
3.河北省新能源汽车动力系统轻量化技术创新中心 保定 071002;4.河北大学电子信息工程学院 保定 071002)

摘 要:针对目前常见的计算机辅助检测系统对结肠镜图像中息肉的分类与检测准确性和实时性不足的问题,提出了一种以 YOLOv4 为基本框架,结合空间注意力机制与改进特征融合层的 YOLOF-CBAM 模型,可对白光和窄带成像双模态内镜图像中的增生性息肉与腺瘤性息肉进行实时分类与检测。为了使息肉的特征提取更准确,在 YOLOv4 的主干网中增加 CBAM 模块,使网络特征提取层关注到更加重要的空间以及通道信息,抑制不必要特征向下传递;在此基础上,通过对特征融合层 PANet 进行剪枝操作优化网络结构,以此减少网络参数量,进一步提高模型的检测速度。为了对改进后的模型进行训练和测试,从河北大学附属医院收集了 2 988 张包含了白光和 NBI 的内镜图像,并按照 9∶1 的划分比例划分为训练集和测试集。实验结果表明,YOLOF-CBAM 在测试集上的 mAP 值为 86.44%,识别增生性息肉和腺瘤性息肉的召回率分别为 89.62%和 85.64%,精确率分别为 91.35%和 85.19%,且实时分类速度达到 47 FPS,证明所提出的模型具有潜在的临床应用价值。

关键词:结直肠息肉;实时分类与检测;注意力机制;YOLOv4;PANet
中图分类号:TP391;TH7 **文献标识码**:A **国家标准学科分类代码**:520.60

YOLOF-CBAM: A new real-time classification and detection method for colorectal polyps

Yang Kun[1,2,3] Sun Yufeng[4] Wang Shiwei[1] Lu Yufei[1] Xue Linyan[1,2,3]
(1. College of Quality and Technical Supervision, Hebei University, Baoding 071002, China;
2. National and Local Joint Engineering Research Center for Measuring Instruments and Systems, Baoding 071002, China;
3. Hebei New Energy Vehicle Powertrain Lightweight Technology Innovation Center, Baoding 071002, China;
4. College of Electronic Information Engineering, Hebei University, Baoding 071002, China)

Abstract: Aiming at the problem that the classification and detection of colorectal polyps by common computer-aided detection systems are not accurate and real-time, a YOLF-CBAM model combined with spatial attention mechanism (CBAM) and improved feature fusion layer based on YOLOv4 is proposed, which can classify and detect hyperplastic polyps and adenomatous polyps in dual modal of white light and NBI endoscopic images in real time. In order to make the feature extraction of polyps more accurate, a CBAM module is integrated to the backbone of YOLOv4, so that the network feature extraction layer pays attention to more important spatial and channel information, and inhibits the downward transmission of unnecessary features. On this basis, the network structure is optimized by pruning the feature fusion layer PANet to reduce the amount of network parameters and further improve the detection speed of the model. In order to train and test the improved model, 2 988 white light and NBI endoscopic images are collected from the Affiliated Hospital of Hebei University, and are divided into training set and test set at a ratio of 9 ∶ 1. Experimental results show that our proposed YOLOF-CBAM achieves a mAP of 86.44%, recalls of 89.62% and 85.64% for identifying hyperplastic and adenomatous polyps respectively, accuracies of 91.35% and 85.19% for identifying hyperplastic and adenomatous polyps respectively, and a classification speed of 47 FPS on the test set, which proves that the proposed model has potential clinical application value.

Keywords: colorectal polyps; real-time classification and detection; attention mechanisms; YOLOv4; PANet

0 引 言

根据世界卫生组织 2020 年发布的全球癌症报告[1],结直肠癌的发病率和死亡率在所有恶性肿瘤中高居第二位。结直肠最常用的检测手段是消化道内窥镜检查,通过息肉活检获取的病理报告是结肠癌筛查的金标准。结直肠癌通

收稿日期:2022-11-03
* 基金项目:河北大学多学科交叉研究项目(DXK201914)、河北大学校长科研基金(XZJJ201914)、大学生创新创业训练计划创新训练项目(2022373)资助

常起源于结肠和直肠的腺瘤性息肉,约60%结直肠癌导致的死亡可以通过筛查腺瘤性息肉进行预防[2]。有效提高腺瘤检出率可显著降低结直肠癌的风险,腺瘤检出率每增加1%,间期结直肠癌的发病率就会降低3%～6%[3]。然而,传统的内镜下息肉的诊断方法存在主观性强、工作量大以及专业水平要求高等限制,导致了大量漏检、误判等问题。因此,实时、准确地进行息肉病理分型预测对于提高腺瘤检出率、降低结直肠癌发病率及改善结直肠癌预后具有重要的临床意义。

随着计算机技术的飞速发展,基于机器学习的计算机辅助诊断技术以其精度高、速度快和成本低的特点在内镜图像自动检测中得到极大关注。传统的计算机视觉算法使用的是分治法,将图像识别问题分解为特征提取、特征分类等简单、可控且清晰的若干子问题。在进行特征提取时,通常以人类的先验经验为基础,采用加速鲁棒性特征(speeded up robust features, SURF)、定向梯度直方图(histogram of oriented gradients, HOG)等算法[4-6]手动提取息肉的颜色、纹理形状等特征,然后通过小波变换、边缘检测、支持向量机(support vector machine,SVM)等特征分类方法区分息肉非息肉区域[7-9]。然而,这类机器学习算法受限于先验特征的影响,并不适合息肉图像检测的临床应用。在此基础上,Tamaki等[10]提出了一种新的局部特征和采样方案的组合,通过使用诸如高斯差和网格采样之类的采样方案来提取多个局部特征,来弥补先验经验的缺失。不过分步解决子问题时,尽管可以在子问题上得到最优解,但并不意味着就能得到全局问题的最优解。与之相反的是端到端的学习方式,使整个学习的流程并不进行人为的子问题划分,而是完全交给深度学习模型直接学习从原始数据到期望输出的映射。

随着计算机硬件与人工智能技术的发展,基于端到端卷积神经网络(convolutional neural network, CNN)可以从数据集中自动学习丰富特征并进行分类与检测[11-14]。范姗慧等[15]对比分析不同的深度卷积神经网络模型(AlexNet、VGGNet和GoogLeNet)对息肉的识别效果,与传统手工提取图像特征方法相比,结合迁移学习策略的深度学习方法为小肠息肉的自动识别提供有效的解决方案。Faster R-CNN网络被广泛应用在医学图像检测任务中,Chen等[16]通过区分前景和背景来增强输入图像的对比度,以提高息肉区域的显著性,并添加了一个注意模块来关注有用的特征通道,弱化无用的特征通道的贡献。Li等[17]提出了一种新的基于深度神经网络的可扩展检测算法,通过增加不同级别的特征图的融合对Faster R-CNN进行改进,解决了连续级联的下采样导致息肉特征丢失的问题。孙雪华等[18]在数据预处理阶段,利用中值滤波的非线性滤波特性去除图像反光区域,并通过更快的区域神经网络对息肉候选区域进行训练,使Faster R-CNN可以更快速的收敛。

准确的描绘出息肉轮廓可以辅助医生进行切除,Mask R-CNN以其强大的像素级别的检测能力广泛应用于这一领域。Wittenberg等[19]采用Mask R-CNN网络框架,对息肉目标进行检测以及分割,在测试数据集上测试结果召回率达到了0.92,精确度达到了0.86,F1达到了0.89。同样以这个模型为基础,河北大学的杨昆团队提出了一种新的不降维的高效通道注意力网络,可以利用这个网络来有效地学习有效的通道注意力,并获得跨通道的互动,获得了94.9%的精度、96.9%的召回率、95.9%的F1评分和96.5%的F2评分[20]。然而Mask R-CNN受限于模型的庞大参数运算,网络的检测速度通常难以匹配临床结肠镜检查的视频的帧数。任莉莉等[21]提出了名为GLIA-Net的低复杂度高性能网络模型对内窥镜图像中的息肉进行分割,以U-Net为网络框架,在其双层卷积后加入交互式注意力融合模块,使得网络可以兼顾局部信息与全局信息,在通道与空间两个维度上进行注意力机制的应用,从而具备对空间与通道、局部与全局语义信息的处理能力,交互式注意力融合模块具有引入参数少,计算量小的优点,在保证网络计算效率的同时网络分割精度得到进一步提升。Pacal等[22]使用基于单阶段检测架构的YOLOv4算法,利用单阶段探测器将候选框的生成与判断过程一体化的优势,提升了检测的速度,Gao等[23]引入协调注意机制,关注到网络中的通道和位置权重信息,在白光的条件下实现了息肉、腺瘤与癌症三种病理特征的有效提取。

对于结肠镜图像,不同患者的结肠息肉形态、位置以及肠道环境都存在偏差,即使同一病灶的形态也会随着结肠镜拍摄角度以及光照条件不同而发生改变。此外,息肉与肠道内壁颜色差异较小,导致图像的对比度较低,边界模糊;同时肠道中存在的气泡、食物残渣等杂质也增加了计算机辅助诊断系统对内镜图像中的息肉进行自动诊断的难度。尽管目前算法的研究在息肉的分类、分割和检测上都有一定的进展,但在临床上依然缺乏一个可靠的检测手段,能够基于白光和窄带成像(narrow band imaging, NBI)图像,对结肠镜检查进行实时的息肉定位与分类。

为解决上述问题,本文构建了一种基于卷积注意力机制(convolutional block attention module,CBAM)和优化特征融合层的多分类实时目标检测网络YOLOF-CBAM(you look only once+one feature+CBAM),以提高内镜图像中息肉的检测精确度与分类准确度。为了模拟肠道环境,本文在数据预处理中引入了模拟遮挡、亮度调节、拼接等数据增强手段提高网络模型的鲁棒性;在网络的主干网络中加入了轻量化的注意力机制,建立了丰富的空间以及通道上的上下文依赖关系,抑制不必要的特征向下传播,增加了定位与分类的准确性;在网络的特征融合层PANet中[24],本文通过对主干网络特征的输出进行消融性实验对比,选取了单输入多输出的特征组合通过减少特征融合次数来达到减少运算的参数量的目的减少特征融合层的计算量,提升

网络运算性能。

1　网络模型的构建

1.1　YOLOv4 算法介绍

目前基于深度学习的主流目标检测算法分为两个架构，分别为基于候选框进行提取判断的双阶段检测器模型（如 Mask RCNN[25]，Faster RCNN[26]）和直接对位置进行预测的单阶段检测器模型（如 YOLO 系列[27-30]，SSD[31]，RetinaNet[32]）。单阶段探测器将候选框的生成与判断过程一体化，极大的提升了检测速度。为保证网络的实时性，本文使用的增生性息肉与腺瘤性息肉分类网络基于单阶段网络 YOLOv4 进行选取并改进。

YOLOv4 网络主要包含 4 部分：输入（Input）、骨干网络（Backbone）、颈部（Neck）、头部（Head），如图 1 所示。骨干网络采用 CSPDarknet53 网络进行特征提取，CSP 结构可以最大化梯度联合的差异。其使用梯度流截断的手段避免不同卷积层学习到重复的梯度信息，能够有效的减少重复的梯度学习[33]。颈部网络采用空间金字塔池化（space pyramid pool，SPP）模块和路径聚合网络（path aggregation network，PAN）模块进行串联。SPP 层可有效避免对图像区域裁剪、缩放操作导致的图像失真等问题。PANet 最大的贡献是提出了一个自顶向下和自底向上的双向融合骨干网络，通过添加跳过连接（short-cut）缩短层之间的路径，其中自适应特征池化可以用于聚合不同层之间的特征，保证特征的完整性和多样性[24]。最后，通过 Yolo 对图像特征进行预测，生成边界框并预测类别。

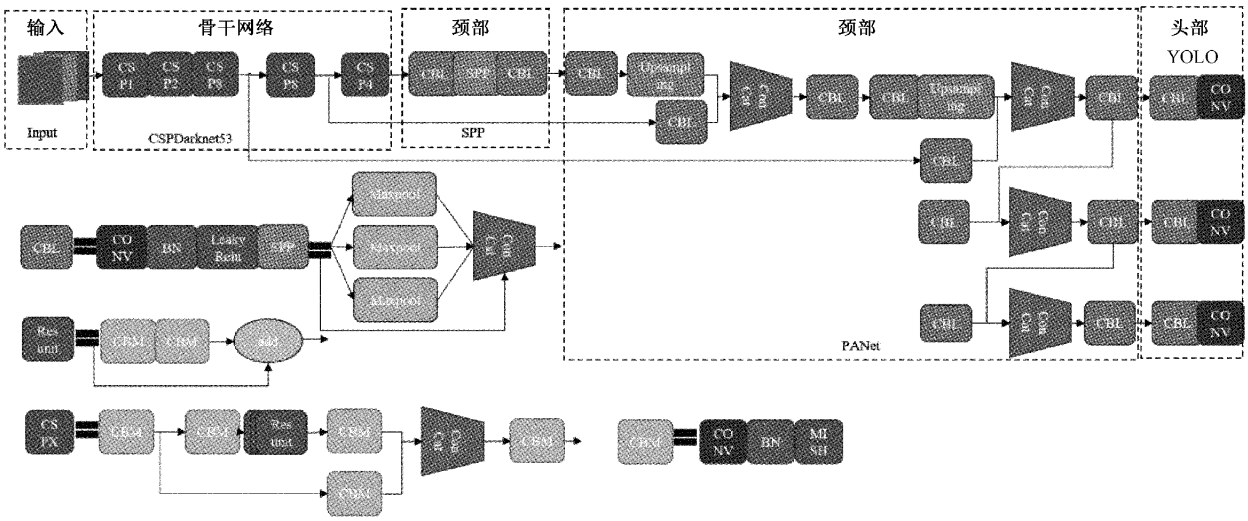


图 1　YOLOv4 网络的模型框架

尽管 PANet 特征融合层对小目标的检测精度提升很大，但是冗杂的特征级联方式在一定程度上降低了网络的实时性。如果在骨干网络中采用更有效的特征提取方式提取更多的息肉特征，就可以通过对特征融合层进行剪枝优化操作，来达到减少计算量提升网络实时性的目的。

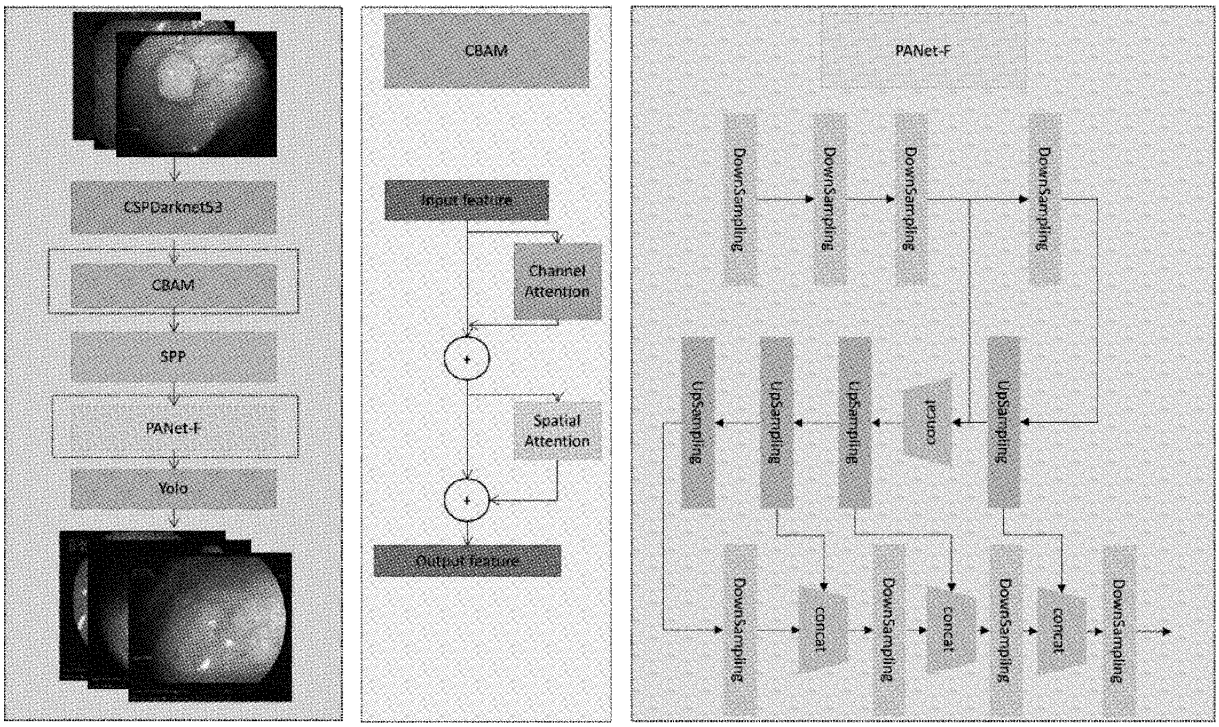
1.2　改进的 YOLOF-CBAM 网络

针对肠镜检测中对息肉目标分类的检测精度低、实时性差等问题，本文提出了 YOLOF-CBAM 算法，其网络结构图如图 2(a)所示。与 YOLOv4 基础网络相比，主要改进之处在于：1）通过在主干网络 CSPDarknet53 后添加 CBAM 卷积注意力模块来重新校准特征图，以增强特征表示能力，同时生成高质量的区域建议，通过强调信息特征通道，来达到弱化不太重要通道的显著性的目的。CBAM 模块的结构如图 2(b)所示。2）为解决模型参数量庞大，计算速度慢的缺陷，本文对 YOLOv4 的特征融合层 PANet 的第一阶段进行剪枝操作，并通过消融实验找到第一阶段的最佳单次采样输出，用单次采样的结果进行反卷积来代替多次采样的融合计算，达到降低模型参数量与计算量的目的，并在 PANet 后半段使用 concat 对分层上采样结果进行级联，来继续保持图像中目标多尺度特征学习的能力。改进后的 PANet-F 模块结构如图 2(c)所示。

1.3　CBAM 卷积注意力机制

近年来注意力机制广泛用于图像处理领域，其形式与人类视觉注意力相似。人类通过观察图像全局信息，获得其中的重要目标区域并将注意力集中于此，以获取更多细节信息。2018 年，Jie 等[34]开创性的提出了 SENet 通道注意力机制，SENet 通过损失函数去学习不同特征通道的特征图的重要程度，然后依此为依据给各个特征通道重新赋予权重值，从而让神经网络去重点关注高权重的特征层，忽略无效特征，使模型达到更好的效果。

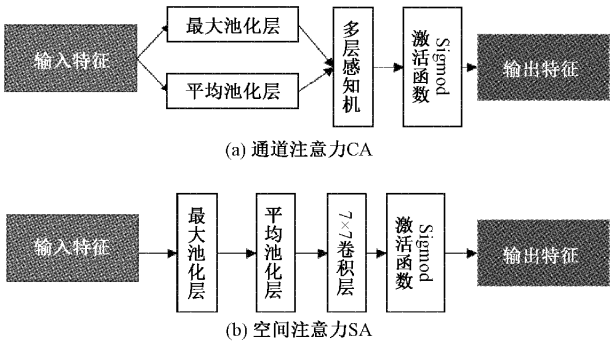
同年 Woo 等[35]提出 CBAM 卷积注意力机制，CBAM 是一种结合通道注意力（channel attention，CA）和空间注意力（spatial attention，SA）的轻量型注意力机制。由于息肉目标在图像中所占像素较少，骨干网络对特征提取的



(a) 整体网络结构　(b) CBAM卷积注意力结构　(c) 改进后PANet-F结构

图 2　改进后的 YOLOF-CBAM 网络

有效信息有限,故本文引入 CBAM 卷积注意力模块以增强小目标的通道信息和空间信息,从而提高息肉检测精度。CA 和 SA 注意模块的结构分别如图 3(a)和(b)所示。



(a) 通道注意力CA

(b) 空间注意力SA

图 3　通道注意和空间注意模块的结构

CA 模块在空间维度上对特征图进行压缩,基于输入的特征图的宽和高进行最大池化(Max pooling)和平均池化(Average pooling)生成两个不同的空间信息,再将特征聚合特征图的空间信息输入到一个由多层感知器(multilayer perceptron, MLP)组成的共享网络,通过元素求和输出通道注意力的特征。平均池化可以对特征图上的每一个像素点进行运算,最大池化只关注在反向传播中,特征图局部像素最大值。通道注意力机制数学模型可以表达为:

$$M_c(F)=\sigma(MLP(AvgPool(F))+MLP(MaxPool(F))) \tag{1}$$

式中:σ 是 sigmoid 操作,F 为特征图。

不同于 CA,SA 更加注重图像的位置信息,为了获得在空间维度的注意力特征,经通道注意力输出的特征图同样基于特征图的宽度和高度进行全局最大池化和全局平均池化,将特征维度由 $\boldsymbol{H}\times\boldsymbol{W}$ 转变成 1×1,接着经过卷积核为 7×7 的卷积和 Relu 激活函数后降低特征图的维度,然后在经过一次卷积后提升为原来的维度,最后将经过 Sigmoid 激活函数标准化处理后的特征图与通道注意力输出的特征图进行合并,从而在空间和通道两个维度上完成对特征图的重标定:

$$M_S(F)=\sigma(f^{7\times7}[AvgPool(F);MaxPool(F)]) \tag{2}$$

式中:7×7 表示卷积核大小。

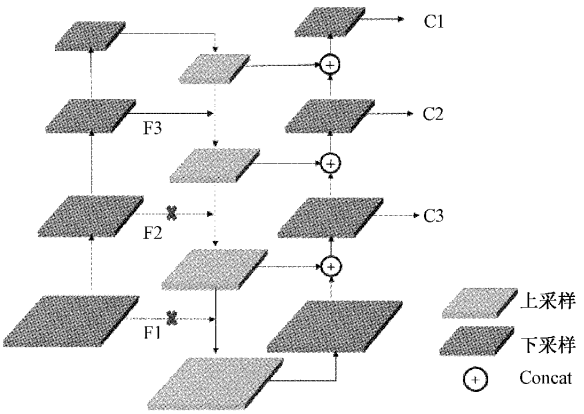
1.4　特征融合层剪枝

融合不同尺度的特征是提高检测性能的一个重要手段。低层特征分辨率高,包含更多位置、细节信息,由于经过的卷积较少,其语义性更低,噪声更多。高层特征则在不断地卷积池化过程中获得更强的语义信息,更大的感受野与全局信息,对于体积较小的目标来说很容易漏检。将不同层次的语义特征从图像中提取出来,合并成一个更具有判别能力的特征向下传递,是改善检测模型的关键[36]。

在常见的目标检测网络中,例如 Faster R-CNN 中的 RPN 层是利用单个高层特征图进行物体的分类和边界框的回归去预测目标,虽然计算量较小,但是容易在下采样过程中丢失像素信息较少的小目标;SSD 网络利用高级语义信息构建了多尺度的特征图,但是忽视了高分辨率的低

级语义特征图的空间信息；FPN 网络通过将每一级的语义信息与下一级的语义信息融合叠加并单独预测，解决了低级语义信息利用不充分的问题[37]。

随着卷积的进行高级别特征逐渐泛化，单个特征覆盖面大，为了解决高级别特征难以访问准确地定位信息的问题，PANet 网络被提出。如图 4 所示，图中 F1，F2，F3 分别代表了 PANet 第 1 个下采样阶段的三层采样输出，PANet 第 2 阶段通过层层上采样并与第一阶段输出的特征级联，特征融合后将会获得更丰富的语义信息。PANet 第 3 阶段再经过层层下采样与第 2 阶段的采样输出叠加获得更准确的位置信息，相当于在 FPN 的基础上，创建了一个自底向上的路径增强和聚合，使低层信息更容易传播。尽管特征层的级联融合可以获得更加丰富的图像细节，但是也带来了计算量大，内存占用过多的问题。



注：PANet-F 在 PANet 的基础上舍弃前两层（F1，F2）的输出，只保留 F3 输出向下传播。

图 4　三次下采样的输出

Chen 等[38]通过对特征金字塔网络的研究发现，特征金字塔网络取得成功的原因在于其将目标检测问题分而治之，而不是因为多尺度特征融合。因此本文结合内镜图像实时检测的要求对 PANet 的第 1、2 阶段的多尺度特征融合部分进行剪枝操作，如图 4 所示，舍弃前两层的采样输出，只保留第 3 次的采样后的特征层。经过三次下采样输出的特征图具有更高级的语义信息以及更大的感受野，因此并不会因为没有进行特征融合而损失过多的有用信息。同时，由于精简了特征融合层数量，减少了特征融合过程中的计算量，从而提升了网络的推理速度。感受野的计算公式如下：

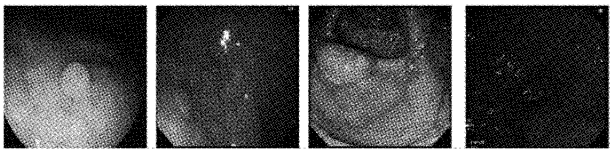
$$r_{i+1} = r_i + (k_{i+1} - 1) \times \prod_{n=0}^{i} s_n \tag{3}$$

式中：r 为感受野，s 为步长，k 为卷积核的尺寸，i 代表层数。

2　实验方法

2.1　数据集生成与标注

本研究从河北大学附属医院收集了 714 个病人的 6 100 张结肠镜检查图片，并按照如下标准进行清洗：图像清晰且无抖动痕迹；图像内至少包含一个完整清晰的息肉；肠道环境干净无明显异物；图像光线亮度适中；NBI 图像与白光图像都不限。图像清洗后最终筛选出 2 988 张图片，包含了腺瘤性息肉和增生性息肉的白光图像和 NBI 窄带结肠镜图像，如图 5 所示。



(a) 增生性息肉白光图像　(b) 增生性息肉 NBI图像　(c) 腺瘤性息肉白光图像　(d) 腺瘤性息肉 NBI图像

图 5　数据集图像示例

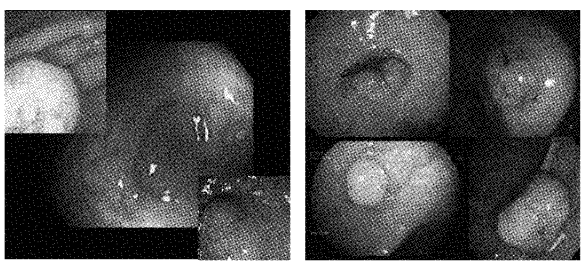
在进行数据标注时，以病理活检报告为金标准，由 1 名主任医师带领其研究生团队参照病理报告中息肉的位置和病理类型（增生性息肉或腺瘤性息肉）进行标注。使用 labelImg 软件作为标注工具，标注的矩形框要完全包含息肉，并且与息肉边缘相切。标注后的图片以 9∶1 的比例分为训练集和测试集。

2.2　数据增强

深度学习方法需要充足的样本量，样本量决定了网络模型的训练效果与泛化能力。然后在实际情况中，样本量往往无法达到预期，因此需要对数据进行扩充与增强。常用的数据增强有畸变和遮挡等。

畸变包括改变图像亮度、对比度、饱和度的光照畸变与随机缩放、剪裁、翻转的几何畸变。这两种都属于像素级别的调整，使图像的几何位置、尺寸、形状、方位等发生改变。旨在减弱背景（或噪音）因子的权重，使模型面对缺失值不敏感，从而可以产生更好的学习效果，增加模型稳定性。

遮挡包括 CutMix[39]以及马赛克数据增强等方法。如图 6(a)，CutMix 是通过从一个图像中切割一部分并将其粘贴到增强图像上来组合图像，通过要求模型从局部视图识别对象，对切割区域中添加其他样本信息，进一步增强模型的定位能力。由于填充的图像是从数据集中随机切割的，在训练的过程中不会出现非信息像素，从而能提高训练效率。如图 6(b)，马赛克数据增强将 4 张训练图像按一定比例组合成 1 张，这使得模型能够学习如何识别比正常尺寸小的物体。



(a) CutMix效果图　(b) 马赛克数据增强效果图

图 6　使用不同方法对息肉图像进行数据增强

2.3　评价指标

本文的评价指标采用精确率（Precision）、召回率（Recall）、F1 分数（F1-score）、平均精度（Average Precision，AP）、平均精度均值（mean average precision，mAP）、画面每秒传输帧数（frames per second，FPS）6 项性能指标评判网络性能，并辅以梯度加权类激活映射方法（gradient-weighted class activation mapping，Grad-CAM）来可视化特征权重。其中 Precision、Recall、F1-score、AP、mAP 的计算方法如下：

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{6}$$

$$AP = \frac{\sum Precision}{N} \tag{7}$$

$$mAP = \frac{\sum AP}{N_{class}} \tag{8}$$

式中：TP、FP 和 FN 分别表示模型正确检测到的目标数量、模型错误检测到的目标数量以及错误检测的数量，N_{class} 表示分类的类别数量。本文设置交并比 IoU 为 0.5，即预测框与实际框之间交集的面积并集的部分大于 50% 的时候，认为网络预测到了目标。

2.4　实验参数

本文实验环境配置如下：CPU 为 Intel（R）Xeon（R）Gold 6240 CPU@2.60 GHz；系统版本为：Ubuntu 18.04.5 LTS PC；GPU 为 NVIDIA RTX 2080Ti GPU × 2，CUDA11.1；Python 版本为 3.7.11；深度学习框架版本为 Pytorch 1.8.1。

在网络模型选取中，分别使用了 SSD、RetinaNet、YOLOv5、YOLOv4，以及添加了 CBAM 的 YOLOv4 进行训练，为了加快模型的训练速度，提高模型的稳定性和泛化能力，本文各个网络模型全部使用 ImageNet 数据集上学习到的权重作为初始网络参数，注意力模块的参数则进行了随机初始化。通过随机梯度下降（stochastic gradient descent，SGD）算法进行网络参数优化，权重衰减为 1×10^{-5}，学习动量为 0.9，batch size 大小为 128、交并比 IoU 设置为 0.5。每个模型训练 100 个 epoch，其中初始学习率设置为 0.01，并每隔 20 个 epoch 学习率乘以 0.1。

3　实验结果分析

3.1　通过 CBAM 提升精确度

本文在 CSPDarknet53 主干网络后面添加了一个 CBAM 卷积注意力模块。与原始网络的特征提取层相比，CBAM 卷积注意力模块聚合了特征图的空间与通道权重，权重参数的调整有利于模型识别性能的提升。如表 1 所示，相对于 YOLOv4 原始模型，引入卷积注意力模块后的网络 mAP 提升了 1.88%，其中增生性息肉的召回率上升了 4.13%，精确度上升了 4.14%，F1 综合评价指标上升了 0.04；腺瘤性息肉的精确度上升了 3.61%。因此，CBAM 模块参与训练后通过加深网络模型深度的方式改变了原始网络的参数结构，更加适合在息肉检测任务中区分息肉与背景的特征。

表 1　增加注意力机制前后对比

Net	mAP/%	AP/%		F1		Recall/%		Precision/%	FPS	
		增生性息肉	腺瘤性息肉	增生性息肉	腺瘤性息肉	增生性息肉	腺瘤性息肉	增生性息肉	腺瘤性息肉	
YOLOv4	85.69	88.45	82.94	0.88	0.84	87.38	85.49	88.24	82.91	41
YOLOv4+CBAM	87.57	92.54	82.59	0.92	0.84	91.51	81.91	92.38	86.52	40

为了对网络特征权重进行可视化，直观地对比改进前后的网络进行特征提取的感兴趣区域，利用 Grad-CAM 方法，使用特征图导数来分别获取 CSPDarknet53 和 CBAM 注意力机制提取的特征激活图[40]，图像颜色越深的地方表示网络对这个部位的关注度越高。

本文分别绘制了 YOLOv4 主干网络 CSPDarknet53 和在其后加入 CBAM 卷积注意力后的 Grad-CAM 热力图。选取了四张图像，其中 1＃和 3＃分别为白光下的增生性息肉与腺瘤性息肉图像，2＃和 4＃分别为 NBI 下的增生性息肉与腺瘤性息肉图像。每张图像中至少包含 1 个息肉（2＃包含 3 个），如图 7(a)中黑色箭头所指示。图 7(b)和(c)分别展示了不加 CBAM 注意力机制和加入 CBAM 注意力机制后网络的注意区域分布。对于图 1＃和图 4＃，增加了 CBAM 卷积注意力之后，网络对息肉部分及周边区域有了更加明显的关注，有助于网络在接下来传播过程中的学习与收敛；对于图 3＃，增加 CBAM 后对一副图上的多个目标的关注得到了显著的提升，可以有效的避免漏检的情况发生；对于图 3＃，YOLOv4 原网络对右上角非息肉区域产生了高关注度，在增加注意力后减少了对无关区域的注意力，有效的避免了错检。由此可以看出，针对同一张图片，加入 CBAM 注意力后，无论是白光还是蓝光的场景下，都有效的区分了背景和息肉的图像区域，可以更好的帮助网络学习与收敛，从而增加算法分类与检测的准确率。

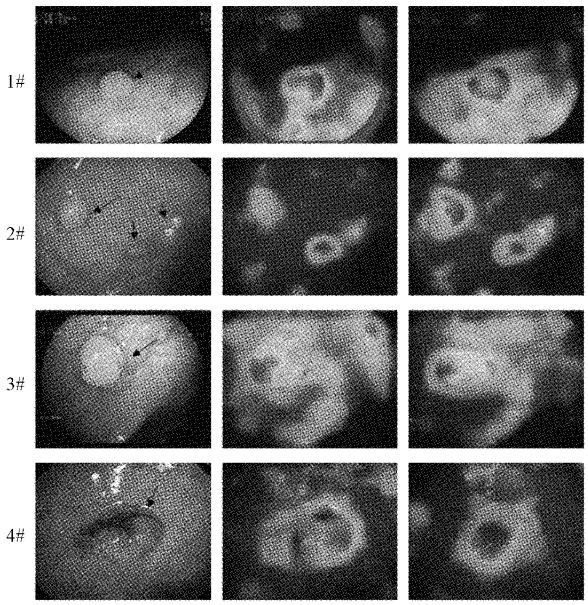


图 7　增加 CBAM 注意力前后 Grad-CAM 的对比

3.2　通过剪枝提升检测速度

通过消融实验验证特征融合层的不同层采样输入对网络性能的影响，以及算法设计合理性。本文分别设计了采用不同层次的采样信息与注意力机制的组合实验。除 1.4 部分提到的使用 F3 层采样结果，还可以结合使用 F2、F1 的低层次语义信息去验证。

在同一训练集和测试集下，以 YOLOv4＋CBAM 为对照网络，依次测试 F1、F2、F3 采样层单独输出对网络性能的影响。实验结果如表 2 所示，由于特征融合过程的取消，造成了部分网络性能指标（mAP）下降 1%（87.57%～86.44%）左右，与此同时，伴随着特征融合计算过程的消失，网络的推理速度提升了 20%（40～47 FPS）。由此可知，具有高级语义的 F3 层相对于 F1 层和 F2 层携带更多的息肉特征信息，可以起到替代特征融合的作用。与此同时，在数据量不充足、样本类别分布不平均的情况下，过深的网络层次会导致网络对样本数偏多的类别产生过拟合现象。更深的网络层次将获得更大的感受野，特征图中具有更丰富的全局信息，通过对特征融合层的剪枝操作来降低局部先验信息的权重有助于防止过拟合现象的发生。

表 2　使用单一采样信息与特征融合做对比

YOLOv4＋CBAM＋分支	mAP/%	AP/%		F1		Recall/%		Precision/%	FPS	
		增生性息肉	腺瘤性息肉	增生性息肉	腺瘤性息肉	增生性息肉	腺瘤性息肉	增生性息肉	腺瘤性息肉	
特征融合	87.57	92.54	82.59	0.92	0.84	91.51	81.91	92.38	86.52	40
F1	85.12	87.13	**83.12**	0.9	0.81	84.47	78.76	**95.6**	83.98	**48**
F2	85.53	**91.46**	79.61	0.88	0.83	89.62	80.32	85.59	84.43	47
F3	**86.44**	91.18	81.71	**0.9**	**0.85**	**89.62**	**85.64**	91.35	**85.19**	47

因此本文选择 YOLOv4＋CBAM＋F3（YOLOF-CBAM）作为最终的网络模型。

YOLOv4 与本文的检测结果对比如图 8 所示，图 8(a)～(d)分别为白光下腺瘤性息肉、窄带蓝光下腺瘤性息肉、白光下增生性息肉、窄带蓝光下增生性息肉由图可知，对于占据图像大块像素值的不规则形状息肉图 8(a)、(b)，本文网络预测出了准确度更高的检测框；在息肉本身与背景颜色接近且存在类似形状的粪便残渣的图 8(c)以及息肉像素占比较少并存在曝光干扰的图 8(d)，本文网络依然做出了准确的预测。因此得出结论，相比于 YOLOv4 原网络，本文网络在不同场景下对息肉的预测定位精度都有提升，有效的减少了错检、漏检率。

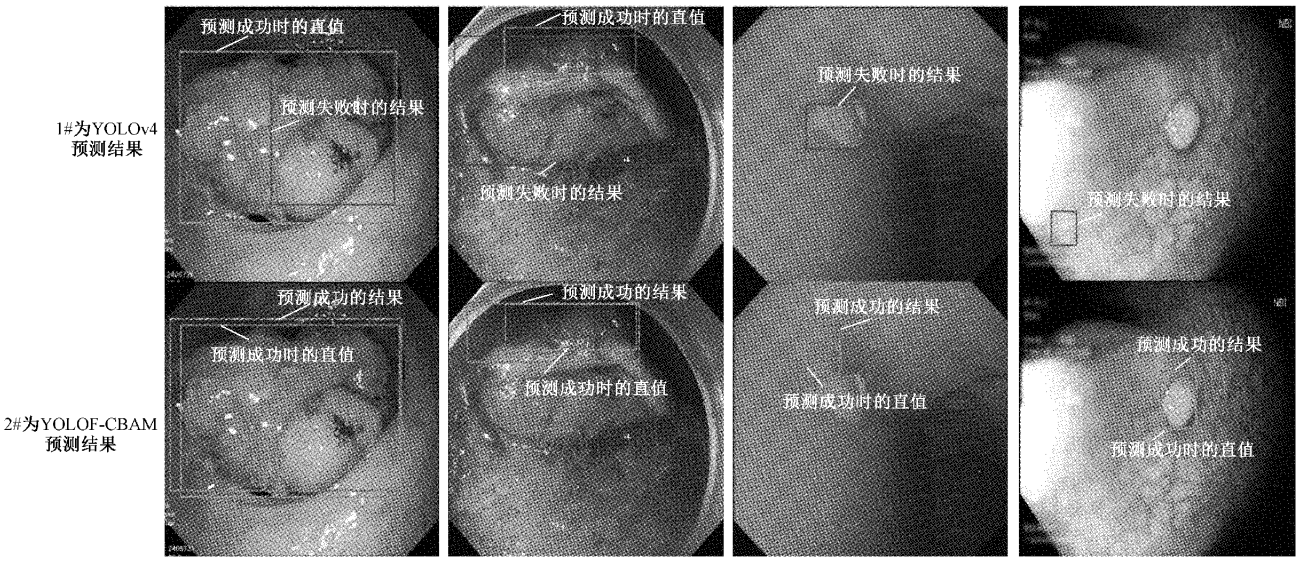
3.3　与其他网络作对比

如表 3 所示，将最优的指标进行加粗标注，与包括原网络在内的其他网络相比，YOLOF-CBAM 的网络性能得到了全面的提升，大部分评价指标处于第一名的位置，仅在腺瘤性息肉 AP 值和增生性息肉的精确度检测上分别落后于 YOLOv4 1%（81.71%～82.94%）和 RetinaNet 2%（91.35%～93.48%），而 RetinaNet 在增生性息肉的检测方面。在实时性指标上，FPS 上小幅落后于 YOLOv5 模型（47～55），然而在腺瘤性息肉的召回率上 YOLOv5 只达到了 66.87%，远远不能达到临床的要求。在临床应用中，单纯的追求 P、R 的提升并没有太大的意义，F1 是结合正负样本比而提出的综合指标，在这一指标上，YOLOF-CBAM 在所有对比网络中做到了最好。

结合以上分析得出结论，与其他经典的网络模型进行对比，本文的网络模型 YOLOF-CBAM 在检测的精度与速度上做到了更好的平衡。

3.4　YOLOF-CBAM 检测应用

为了判断 YOLOF-CBAM 算法在实际病例检测中与活检金标准是否存在结果差异，本文随机抽取了 3 例志愿者，使用该算法对其息肉视频图像中出现的息肉数量与类别进行检测。



(a) 分别为白光下腺瘤性息肉　(b) 窄带蓝光下腺瘤性息肉　(c) 白光下增生性息肉　(d) 窄带蓝光下增生性息肉

图 8　改进前后检测结果对比

表 3　本文网络与基础网络对比

Net	mAP/%	AP/%		F1		Recall/%		Precision/%		FPS
		增生性息肉	腺瘤性息肉	增生性息肉	腺瘤性息肉	增生性息肉	腺瘤性息肉	增生性息肉	腺瘤性息肉	
SSD	82.06	90.29	73.82	0.85	0.71	83.02	66.49	87.13	75.30	21
YOLOv5	84.20	89.07	79.32	0.87	0.73	84.04	66.87	89.53	81.44	55
RetinaNet	84.88	88.75	81.02	0.87	0.75	81.13	69.98	93.48	80.81	28
YOLOv4	85.69	88.45	82.94	0.88	0.84	87.38	85.49	88.24	82.91	41
YOLOF-CBAM	86.44	91.18	81.71	0.90	0.85	89.62	85.64	91.35	85.19	47

由表 4 数据可知，进过随机抽的 3 例患者进行结果对比，其中志愿者 1 与志愿者 2 的结果算法与金标准对息肉的数量与类别检测完全一致，在第 3 例中，算法检测与金标准相比也无漏检情况出现。由此可得，本文提出的 YOLOF-CBAM 算法与金标准结果并无显著性差异。

表 4　息肉类别检测　　（个）

方法	类别	志愿者 1	志愿者 2	志愿者 3
活检	腺瘤	1	0	1
	增生	3	2	2
算法	腺瘤	1	0	1
	增生	3	2	3

4　结　　论

本文提出了一个结肠镜下息肉实时检测方案，其中包括数据增强与扩充、特征提取和特征融合层优化 3 个阶段，完成了息肉的定位、增生性息肉与腺瘤性息肉二分类等两个主要功能。在这 3 个阶段针对性的进行设计、改进，使各阶段都能更加高效地发挥出其应有的性能。

在数据增强与扩充阶段，考虑到数据样本较少、数据集样本分布不均衡等特点，本文在数据预处理阶段使用了剪切、缩放、翻转等数据扩充手段，同时使用了明暗调节、遮挡等数据增强手段来模拟肠道环境，对于被遮挡的息肉有更强的识别能力。在特征提取阶段，本文增加了卷积注意力模块，将两个独立维度的注意力特征与输入特征进行自适应特征优化，可以更好地将息肉与背景区分，并关注到更多的息肉信息。本文基于 PANet 模块的特征融合层进行剪枝，克服了特征融合层参数多、计算量大以及内存占用多的缺点。优化后的网络对于息肉的检测与定位结果速度更快，实现了检测速度与精度的双提高。

参考文献

[1] SUNG H, FERLAY J, SIEGEL RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA: A Cancer Journal for Clinicians, 2020, 71(3): 209-249.

[2] WINAWER S, ZAUBER A, HO M, et al. Prevention of

colorectal cancer by colonoscopic polypectomy. The national polyp study workgroup[J]. New Englnd Journal of Medicine, 1993, 329(27):1977-1981.

[3] 赵胜兵，王树玲，方军,等. 国内外结直肠癌早诊早治现状[J]. 中华消化内镜杂志，2019，36(2)：143-147.

[4] SIDIBÉ D, SADEK I, MÉRIAUDEAU F. Discrimination of retinal images containing bright lesions using sparse coded features and SVM[J]. Computers in Biology & Medicine, 2015, 62: 175-184.

[5] SADEK I, SIDIBÉ D, MÉRIAUDEAU F. Automatic discrimination of color retinal images using the bag of words approach[J]. International Society for Optics and Photonics, 2016: 9414-9414J.

[6] VERAS R, SILVA R, ARAUJO F, et al. SURF descriptor an pattern recognition techniques in automatic identification of pathological retinas[C]. 2015 Brazilian Conference on Intelligent Systems (BRACIS), IEEE, 2015:316-321.

[7] WIMMER G, TAMAKI T, TISCHENDORF J J W, et al. Directional wavelet based features for colonic polyp classification[J]. Medical Image Analysis, 2016, 31:16-36.

[8] HWANG S, TAVANAPONG W, et al. Polyp detection in colonoscopy video using elliptical shape feature[C]. IEEE International Conference on Image Processing, IEEE, 2007: 465-468.

[9] YAQOOB M K, ALI SF, BILAL M, et al. Resnet based deep features and random forest classifier for diabetic retinopathy detection[J]. Sensors, 2021, 21: 1-14.

[10] TAMAKI T, YOSHIMUTA J, KAWAKAMI M, et al. Computer-aided colorectal tumor classification in NBI endoscopy using local features[J]. Medical Image Analysis, 2013, 17(1): 78-100.

[11] 喻殿智，张欣，迟杏. 基于 CA-DenseNet 的乳腺癌病理图像识别[J]. 国外电子测量技术，2022，41(5)：137-143.

[12] 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.

[13] 何晓云，许江淳，陈文绪. 基于改进 U-Net 网络的眼底血管图像分割研究[J]. 电子测量与仪器报，2021，35(10)：202-208.

[14] 王桂棠，林桢哲，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.

[15] 范姗慧，刘士臣，曹鹗，等.无线胶囊内窥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.

[16] CHEN B L, WAN J J, CHEN T Y, et al. A self-attention based faster R-CNN for polyp detection from colonoscopy images[J]. Biomedical Signal Processing and Control, 2021, 70: 103019.

[17] 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.

[18] 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.

[19] WITTENBERG T, ZOBEL P, RATHKE M, et al. Computer aided detection of polyps in whitelight-colonoscopy images using deep neural networks[J]. Current Directions in Biomedical Engineering, 2019, 5(1):231-234.

[20] 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.

[21] 任莉莉，边璇，王光磊，等. GLIA-NET：基于深度学习的息肉分割网络[J]. 计算机工程，2022，48(12)：248-254.

[22] PACAL I, KARABOGA D. A robust real-time deep learning based automatic polyp detection system[J]. Computers in Biology and Medicine, 2021, 134: 104519.

[23] 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: 1-11.

[24] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]. Conference on Computer Vision and Pattern Recognition (CVPR) M IEEE, 2018: 8759-8768.

[25] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. Proceedings of the IEEE International Conference on Computer Vision, ICVV, 2017: 2980-2988.

[26] 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.

[27] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detecton[C]. Conference on Computer Vision and Pattern

Recognition(CVPR), IEEE, 2016: 779-788.

[28] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017: 6517-6525.

[29] CHANDRIKA A, SUMAN D. YOLOv3 based real time social distance violation detection in public places[C]. 2021 International Conference on Computational Performance Evaluation(ComPE), 2021: 625-630.

[30] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv Preprint, 2020, ArXiv:2004.10934

[31] LIU W, ANGUELOV D, ERHAN, et al. SSD: Single shot MultiBox detector[J]. Lecture Notes in Computer Science, 2016:21-37.

[32] LIN T, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017: 2999-3007.

[33] WANG CY, LIAO H, WU YH, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2020: 1571-1580.

[34] JIE H, LI S, GANG, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017: 7132-7141.

[35] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[J]. European Conference on Computer Vision, 2018,1:3-19.

[36] YU W, YANG K, YAO H, et al. Exploiting the complementary strengths of multi-layer CNN features for image retrieval[J]. Neurocomputing, 2016, 237(MAY10): 235-241.

[37] LIN T, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Conference on Computer Vision and Pattern Recognition(CVPR), IEEE, 2017:936-944.

[38] CHEN Q, WANG Y, YANG T, et al. You only look one-level feature[C]. Conference on Computer Vision and Pattern Recognition(CVPR), 2021: 13034-13043.

[39] YUN S, HAN D, CHUN S, et al. CutMix: Regularization strategy to train strong classifiers with localizable features[C]. International Conference on Computer Vision(ICCV), 2019: 6022-6031.

[40] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization [J]. International Journal of Computer Vision, 2020, 128(2):336-359.

作者简介

薛林雁(通信作者),博士,副教授,主要从事基于人工智能的医学影像处理;人工视觉基础理论及应用研究。

E-mail:lyxue@hbu.edu.cn

杨昆,博士,教授,主要研究方向为生物医学工程技术及医学图像处理技术。

E-mail:yangkun@hbu.edu.cn