

改进 YOLOv5s 的明渠漂浮垃圾实时检测方法

石露露¹, 廖光忠²

(1. 武汉科技大学 计算机科学与技术学院, 湖北 武汉 430065;
2. 武汉科技大学 智能信息处理与实时工业系统湖北省重点实验室, 湖北 武汉 430065)

摘要:针对航拍图像上明渠漂浮垃圾尺寸小,且易受水面倒影、强光等因素干扰而造成漂浮垃圾漏检和误检的问题,提出了一种基于改进 YOLOv5s 的明渠漂浮垃圾实时检测方法。首先,通过数据增强的方式对数据集进行扩充,避免数据量过少产生过拟合现象;然后,结合加权双向特征融合网络(BiFPN),对 YOLOv5s 结构的特征融合过程进行修改,以提高检测精度和速度;并且在 Neck 和 Head 部分之间添加3个改进的三维 CBAM 注意力机制模块,加强网络信息的提取和定位能力,能够有效地降低检测的漏检率和误检率;最后,增大网络输入的分辨率,使图像具有更加丰富的细节信息和更加精确的位置信息,有利于小目标特征信息的提取。实验结果显示改进的 YOLOv5s 算法检测的平均精度达到了 89.9%,与原 YOLOv5s 算法相比提高了 9.5%,而且与其他目标检测算法对比,能够提高明渠漂浮垃圾检测的精度,确保检测的实时性。

关键词:YOLOv5s;实时检测;加权双向特征融合;注意力机制;小目标

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2023)09-0083-08

doi:10.3969/j.issn.1673-629X.2023.09.013

Real-time Detection Method of Floating Garbage in Open Channels Based on Improved YOLOv5s

SHI Lu-lu¹, LIAO Guang-zhong²

(1. School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China;
2. Hubei Provincial Key Laboratory of Intelligent Information Processing and Real-time Industrial System,
Wuhan University of Science and Technology, Wuhan 430065, China)

Abstract: Aiming at the problem that floating garbage in open channels is small in size and easily disturbed by factors such as reflection on the water surface and strong light, resulting in missed detection and false detection of floating garbage, a real-time detection method of floating garbage in open channels based on improved YOLOv5s is proposed. Firstly, the data set is expanded by means of data enhancement to avoid overfitting caused by too little data. Then, combined with the weighted bidirectional feature pyramid network (BiFPN), the feature fusion process of the YOLOv5s structure is modified to improve the detection accuracy and speed. Next, three improved 3D CBAM attention mechanism modules are added between the Neck and Head parts to enhance the extraction and positioning capabilities of network information, which can effectively reduce the missed detection rate and false detection rate of detection. Finally, the network input resolution is increased, so that the image had richer detailed information and more accurate position information, which is conducive to the extraction of small target feature information. The experimental results show that the average detection accuracy of the improved YOLOv5s algorithm reaches 89.9%, which is 9.5% higher than that of the original YOLOv5s algorithm, and compared with other target detection algorithms, it can improve the detection accuracy and ensure the real-time detection.

Key words: YOLOv5s; real-time detection; weighted bidirectional feature fusion; attention mechanism; small objects

0 引言

南水北调工程是目前解决国内北方地区水资源严重短缺问题的特大型基础设施项目。确保调水水质是

其成败的关键,但是这项工程的沿途都是明渠,水面难免会有漂浮的垃圾,传统处理明渠漂浮垃圾的方法主要是依靠人力,这种方式不仅耗时耗力,而且无法做到

收稿日期:2022-11-14

修回日期:2023-03-16

基金项目:国家自然科学基金项目(61502359)

作者简介:石露露(1997-),女,硕士研究生,CCF 会员(K9475G),通讯作者,研究方向为计算机视觉、物联网技术;廖光忠(1969-),男,硕士,副教授,CCF 会员(E4235M),研究方向为物联网技术、信息安全。

实时监测。随着无人机航拍和无人船航拍技术的兴起,人工巡检逐渐被取代,如何检测航拍图像上明渠漂浮垃圾成为了亟待解决的问题。

近年来,随着深度学习的快速发展,基于深度学习的目标检测算法成为了检测航拍图像上明渠漂浮垃圾的主流方法。基于深度学习的目标检测算法主要分为双阶段目标检测和单阶段目标检测。双阶段目标检测算法虽然检测的准确率高,但是速度较慢,无法实现实时检测。单阶段目标检测算法是直接对输入的图片进行卷积特征提取,不需要使用候选框,可以极大地提高检测速度。因此,越来越多的学者开始将单阶段目标检测算法应用于河道漂浮物检测领域。Li 等^[1]基于改进的 YOLOv3^[2]算法实现水面漂浮垃圾检测,将三尺度检测转换成两尺度检测以提高检测速度,并通过调整先前的锚框提高检测精度。唐小敏等^[3]基于 ResNet-101 特征提取网络的 SSD 网络模型对河道漂浮物进行目标检测,有效地提高了网络对不同尺度漂浮物的识别精度。李德鑫等^[4]对 YOLOv5s 算法进行改进实现河道漂浮垃圾分类检测,降低各类别目标检测精度之间的差异。以上单阶段目标检测算法虽然在精度上有了一定的提升,但是小目标检测精度仍不高。针对小目标检测问题,国内外学者也提出了许多算法。Benjumea 等^[5]提出了 YOLO-z,即通过 BiFPN^[6](Bi-directional Feature Pyramid Network,加权双向特征融合网络)代替 YOLOv5 中的 PANet^[7](Path Aggregation Network,路径聚合网络),扩大 Neck 层等一系列方法促进中浅层的特征融合,从而提升小目标的检测性能。Zhu 等^[8]提出了 TPH-YOLOv5,在 YOLOv5 的基础上增加一个预测头用于检测不同尺度的目标,还通过集成 CBAM^[9](Convolutional Block Attention Module)注意力机制来寻找对象密集场景下的注意力区域,有效地提升了小目标的检测性能。目前现有的大部分小目标检测算法应用领域比较局限,一般只在大型公共数据集上进行验证,而现实场景错综复杂,现有算法仍然存在检测精度低、难以准确检测到小目标、速度不满足日常需要、漏检和误检等问题。

在明渠漂浮垃圾检测的场景中,航拍图像中的漂浮垃圾大部分尺寸都很小,占用的图像区域都比较小,导致这些漂浮垃圾缺乏外观信息,很难与背景区分开来,从而增加了准确检测明渠漂浮垃圾的难度。而且水面的强光反射可能会导致图像过度曝光,使水面漂浮垃圾不可见,增加了明渠漂浮垃圾检测的漏检率。此外,水面的倒影会对漂浮垃圾检测造成干扰,容易产生误检。

为了解决以上问题,该文采用的方法如下:

(1)通过随机剪裁、水平翻转、加入高斯噪声、改

变对比度等图像处理操作进行数据增强来扩充数据集;

(2)借鉴 BiFPN 的结构对 YOLOv5s 原有的 Neck 部分进行改进,修改特征融合的过程,提高小目标检测检测精度和检测速度;

(3)进行模型优化,在 Neck 和 Head 部分之间添加 3 个改进的 CBAM 注意力机制,增强信息的特征提取和定位;同时增大网络的输入分辨率,使图像信息更加丰富和精确。

1 YOLOv5s 基本模型

2020 年 6 月,Ultralytics 团队 Glenn 等人提出了 YOLOv5 模型,虽然在性能方面与 YOLOv4^[10]不分伯仲,但是推理速度有非常大的提升,可以满足实时检测的要求。YOLOv5 包括 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 四种网络结构^[11]。该文主要是基于 YOLOv5s 进行改进的。为了可以在提高检测精度的同时,也能保证检测速率,以实现实时检测,YOLOv5s 模型主要由 Input、Backbone、Neck、Head 四部分组成。

Input 部分主要包括自适应计算锚框、Mosaic 数据增强和自适应缩放图片三个部分。Backbone 部分包括 CSPNet^[12](Cross Stage Partial Network,跨阶段局部网络)模块和 SPPF 模块,这部分的作用是为了处理输入图像,将图像转换成特征。YOLOv5s 6.0 为了避免 Focus 切片下采样会丢失部分图片原始信息,将 Focus 换成了普通的卷积块。此外,YOLOv5s 6.0 还用 SPPF 替代了原先的 SPP(Spatial Pyramid Pooling,空间金字塔池化),SPP 的 3 个多尺度最大池化层是并联连接的,而 SPPF 在此基础上将 3 个多尺度最大池化层串联起来,保留 SPP 优势的同时,也可以减少图片信息的丢失。CSPNet 可以缓解梯度消失的问题,避免梯度的重复计算,从而减少计算量,提升网络的计算能力以及减少内存的损耗。Neck 部分包括 FPN^[13](Feature Pyramid Networks,特征金字塔)和 PANet,这部分主要是为了充分地提取特征。Head 部分主要是对图像的特征进行预测,然后生成边界框、预测目标的种类。同时采用了三种损失函数分别计算分类损失、定位损失和置信度损失,并通过 NMS(Non-Maximum Suppression,非极大值抑制)提高网络预测的准确度。

2 改进的 YOLOv5s 模型

针对航拍明渠漂浮垃圾图像存在的问题,主要从数据增强、多尺度特征融合、引入注意力机制、增加网络的输入尺度这四个方面对 YOLOv5s 模型进行改进和优化,改进的 YOLOv5s 的网络结构如图 1 所示。

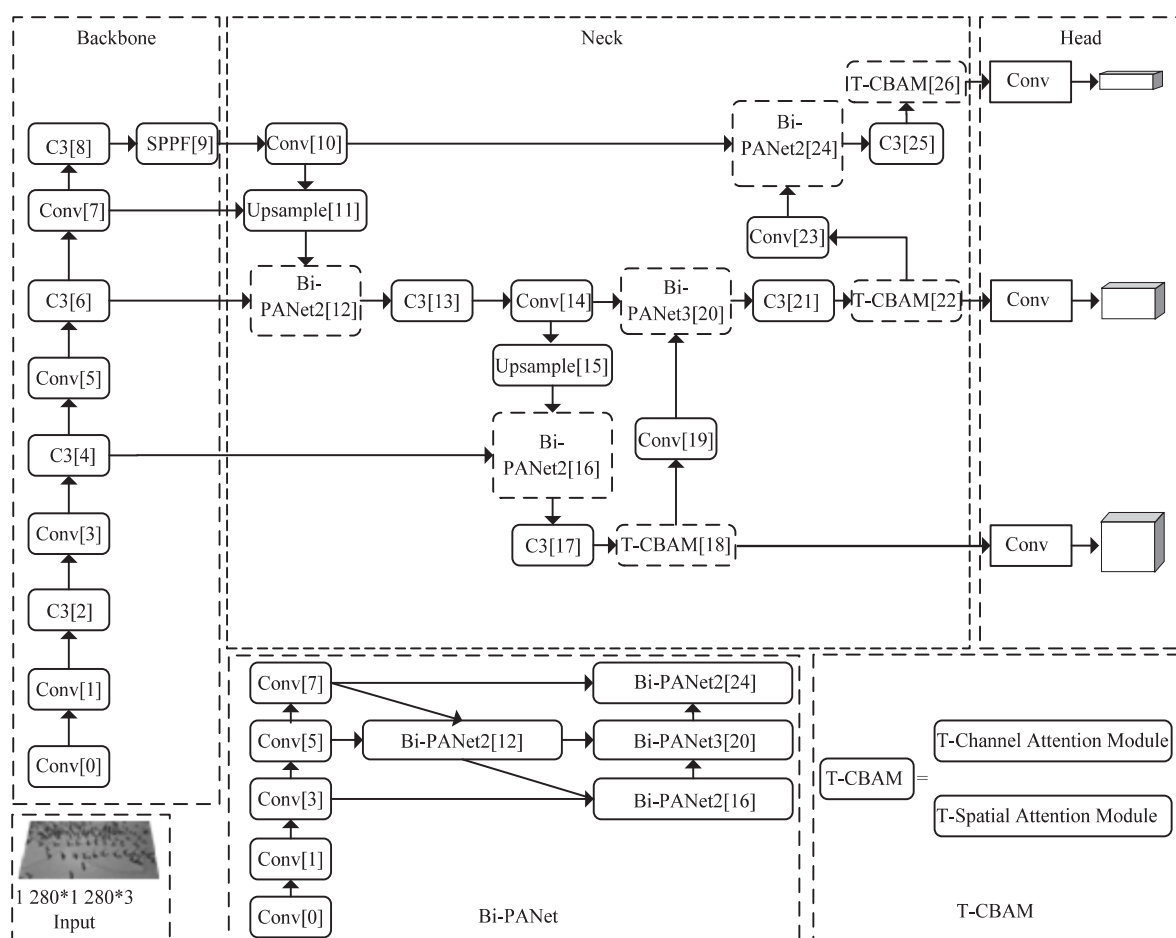
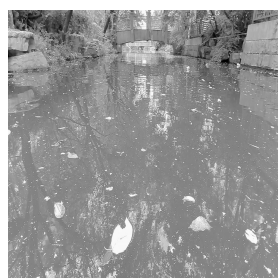


图1 改进的 YOLOv5s 的网络结构

2.1 数据增强

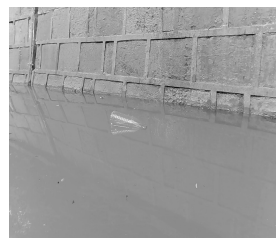
由于目前没有用于河道漂浮检测的大型公共数据集,该文使用了来自 FloW^[14] 的图像子数据集 FloW-Img 子集和在线收集的 2 000 张数据。FloW 数据集是全球第一个无人船视角的水面漂浮垃圾检测数据集。该数据集是使用无人船在真实场景下进行拍摄收集的,但是 FloW-Img 数据集中的漂浮垃圾主要包括不同外观的塑料瓶、饮料瓶、草和落叶,数据种类较少。于是在 FloW-Img 数据集中选取 1 000 张优质图片,剩余 1 000 张则是在线收集合适的水面漂浮垃圾图像,确保数据集具有多样性。最后数据集包括落叶、塑料袋、瓶子、牛奶盒、塑料垃圾、草、树枝、球 8 个类别,如图 2 所示。



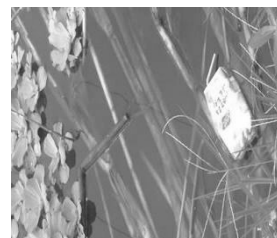
(a)落叶



(b)塑料袋



(c)瓶子



(d)牛奶盒



(e)塑料垃圾



(f)草



(g)树枝



(h)球

图2 水面漂浮垃圾类别

为了避免数据集太小而出现过拟合问题,将通过局部剪裁、随机翻转、加入高斯噪声、改变对比度等方式进行图像处理操作,让每张原始图像生成 5 张新图像,总共得到 10 000 张,但是经过数据增强之后,有的图片质量变化较为严重,需要将质量较差的图像进行人工筛选和剔除,最终数据集为 8 400 张。在图像处理时随机使用 1 到 3 种方式进行变换,可以增加数据集中图片的数量,提高图片信息的多样性,使模型在训练时可以学习到更加全面的水面漂浮垃圾特征。之后,通过数据标注工具 Labelimg 对所有图片进行标注,并将数据集所有图像按照 7 : 2 : 1 的比例划分成三个数据集,即训练数据集、验证数据集和测试数据集,图片数量分别为 5 880 张、1 680 张、840 张。最后,将 VOC 格式的数据标注文件转换为 YOLO 格式。

2.2 多尺度特征融合

在卷积神经网络的多层结构中,浅层网络更关注局部细节信息,如图像边缘、轮廓、纹理等,有利于对目标进行定位;深层网络则侧重于提取抽象信息,有利于对目标进行分类,但是细节信息比较缺乏。因此,将浅层网络特征与深层网络特征进行融合可以提高目标检测的性能。通常,构造特征金字塔 FPN 是一种直接有效的融合方式,即建立一条自顶向下的通路进行特征融合,然后用融合后具有高语义信息的特征层进行预测,但是这种结构受单向信息流的限制。为了解决这个问题,PANet 结构应运而生,PANet 结构是在 FPN 结构的基础上再添加一条自底向上的通路,这条通路主要用于将底层的位置信息也传到预测特征层中,使预测特征层既有丰富的语义信息也有精确的位置信息,不仅可以提高目标检测的精度,而且有利于小目标检测,如图 3(a) 所示。

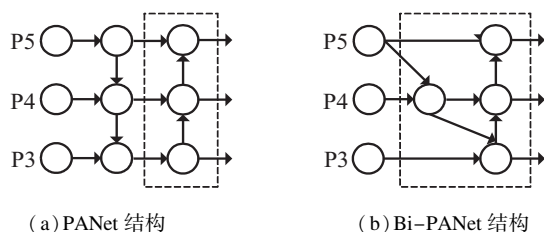


图 3 PANet 结构和 Bi-PANet 结构

虽然 PANet 可以提高目标检测的精度,但是需要更多的参数和更大的计算量。于是,借鉴 BiFPN 加权双向特征融合的思想将原先 PANet 的特征融合过程进行修改,如图 3(b) 所示,称为 Bi-PANet。首先在 PANet 的基础上去掉 P5、P3 中没有进行特征融合的不必要连接,使融合更多特征的同时消耗较少的成本,减少计算量;然后将图 3(b) 虚线方框里的双向路径作为一个特征网络层,重复叠加相同的特征网络层,以实现更高层次的特征融合;最后,给每个输入特征增加一

个额外的权重,让网络了解它们的重要性。其中,采用的加权特征融合方法是快速归一化融合,每个归一化权重的值都是在 0 到 1 之间,可以提高计算速度。快速归一化融合的公式如下所示:

$$O = \sum_i \frac{\omega_i}{\varepsilon + \sum_j \omega_j} \cdot I_i \quad (1)$$

其中, ω_i 是可学习的权重,用激活函数 ReLu 来确保每个 $\omega_i \geq 0$,该权重的值可由网络训练得到, I_i 是指输入的特征, O 是指输出特征, $\varepsilon = 0.0001$ 是避免数值不稳定的一个小值。

Bi-PANet 将主干网络的 P3、P4、P5 作为特征的输入。以节点 P4 为例,其形成 2 个融合特征的过程如下所示:

$$P_4^{\text{id}} = \text{Conv} \left(\frac{\omega_1 \cdot P_4^{\text{in}} + \omega_2 \cdot \text{Resize}(P_5^{\text{in}})}{\omega_1 + \omega_2 + \varepsilon} \right) \quad (2)$$

$$P_4^{\text{out}} = \text{Conv} \left(\frac{\omega'_1 \cdot P_4^{\text{id}} + \omega'_2 \cdot \text{Resize}(P_3^{\text{out}})}{\omega'_1 + \omega'_2 + \varepsilon} \right) \quad (3)$$

其中, P_4^{id} 是自顶向下路径 P4 的中间特征, P_4^{out} 是自底向上路径 P4 的输出特征, Resize 是上采样或下采样操作, Conv 是进行卷积操作。

将 YOLOv5s 结构上的特征融合方式结合 BiFPN 进行修改不仅可以加强特征融合,提高小目标检测的精度,解决航拍图像中的明渠漂浮垃圾大部分尺寸都很小的问题,还可以提高检测速度,确保明渠漂浮垃圾检测的实时性。

2.3 引入注意力机制

为了关注图像中待检测的漂浮垃圾,减少潜在信息的丢失,避免水面倒影、强光反射等对漂浮垃圾检测的干扰,加强漂浮垃圾特征的提取,该文借鉴 CBAM 的通道注意力机制和空间注意力机制的结构,通过保留跨三维的信息减少信息的丢失,同时使用两个卷积层促进空间信息融合从而使网络可以获得更多信息。CBAM 是一个轻量级的通用注意力机制模块,可以直接添加到目标检测网络中进行训练,而且带来的计算开销也极小,不会影响实时目标检测。CBAM 注意力机制模块包含两个子模块,一个是通道注意力机制模块,另一个是空间注意力机制模块,其结构如图 4 所示,图中 \otimes 表示元素级乘法。

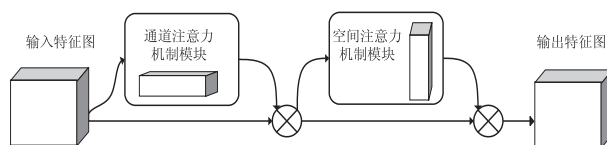


图 4 CBAM 注意力机制结构

虽然 CBAM 可以使特征图能够更多地包含和表现有效信息,但是它忽略了空间和通道之间的相互作用,会造成跨维信息的丢失。于是,对 CBAM 的通道

注意力机制和空间注意力机制进行修改,称为 T-CBAM。首先在通道注意力机制模块中使用 3D 排列^[15]来保留跨三维的信息,从而利用每一对三维通道、空间宽度和空间高度之间的注意力权重来减少信息的丢失,提高检测精度;然后在空间注意力机制模块中使用两个卷积层实现空间信息融合,从而可以关注到更多的空间信息。其具体实现过程如下:

主干网络的特征进行特征融合之后,进入通道注意力机制,先将输入的特征图在通道、空间宽度和空间高度的 3 个维度上实现全局平均池化和最大池化,并将池化得到的权重矩阵传入权重共享的多层感知器 MLP(Multi-Layer Perception)中进行特征相加,最后通过 Sigmoid 激活函数得到通道注意力特征图 $M_c(F)$ 。T-CBAM 的通道注意力机制如图 5 所示。

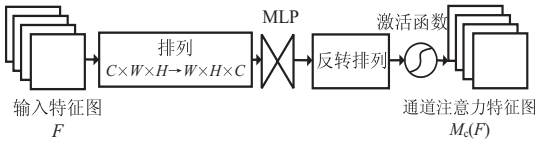


图 5 T-CBAM 的通道注意力机制

之后,通道注意力机制输出的特征图将作为输入的特征图进入空间注意力机制,同样在通道、空间宽度和空间高度的 3 个维度上实现全局平均池化和最大池化,并将结果串联连接起来,然后采用两个 7×7 的卷积核进行卷积操作,保留更多空间信息,最后通过 Sigmoid 激活函数得到空间注意力特征图 $M_s(F)$,实现特征细化。T-CBAM 的空间注意力机制如图 6 所示,其中 r 是通道注意力机制中 MLP 使用的压缩比。

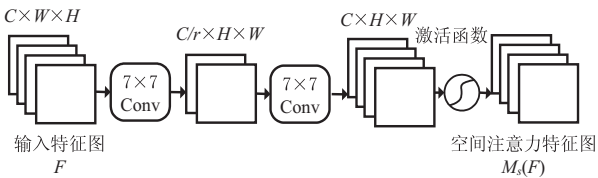


图 6 T-CBAM 的空间注意力机制

特征融合之后引入 T-CBAM 注意力机制的总体过程可以概括为:

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F'' &= M_s(F') \otimes F' \end{aligned} \quad (4)$$

其中, \otimes 表示元素级乘法, F' 表示经过通道注意力机制输出的特征图, F'' 表示经过通道注意力机制和空间注意力机制最终输出的精化特征图。

该文经过 Neck 部分的特征融合之后引入 3 个 T-CBAM, 这样会减少潜在信息的丢失, 使特征表现更加充分, 加强特征提取, 抑制无用信息, 从而更好地提高检测性能。引入 T-CBAM 注意力机制可以实现集中对水面漂浮垃圾特征的提取, 避免水面倒影及无害漂浮物干扰水面漂浮垃圾检测, 而且加强特征提取也会

减少水面强光反射对明渠漂浮垃圾检测的影响。

2.4 增加网络的输入尺度

小目标本身像素点少, 自身携带的特征信息少, 有时很难被发现, 导致位置信息获取困难, 最直接的解决办法就是增大网络输入分辨率, 这样可以使图像具有更加丰富的细节信息和更加精确的位置信息, 有利于小目标特征信息的提取。于是, 该文将输入尺度从 640×640 增大到 $1\,280 \times 1\,280$, 增大后的预测分支分辨率则变为 $32 \times 32, 64 \times 64, 128 \times 128$ 。这样不仅可以提高目标检测的检测精度, 还可以提高小目标的检测精度, 也有利于解决航拍图像中的明渠漂浮垃圾大部分尺寸都很小的问题。

3 实验与分析

3.1 实验环境及参数设置

为了训练和评估所提出的模型, 实验使用的是 CentOS Linux 7.9 系统, 并采用深度学习框架 Pytorch。所有模型都是在 GPU 上进行训练的, 其中 GPU 型号为 Tesla V100-SXM2-16GB。实验环境为 Python3.8, Pytorch 1.8.1, CUDA10.1, Cudnn7.6.5。实验使用的数据集是 FloW-Img 数据集的扩充数据集, 即通过剪裁、翻转、加入高斯噪声、改变对比度等方式进行图像处理从而增加数据集的数量。实验采用的是 Adam (Adaptive moment estimation) 优化器, 总共迭代 300 次, 其中批量大小设置为 16, 初始学习率设为 0.01, 动量为 0.937, 权重衰减系数为 0.000 5。

3.2 评价指标

为了评价模型的性能, 主要采用精确度 P (Precision)、召回率 R (Recall)、平均精度均值 mAP (mean Average Precision)、运行时间 (Times)、检测速度 FPS (Frames Per Second)、浮点运算量 GFLOPs (Giga Floating-point Operations Per Second) 和参数量 Params (Parameters) 作为模型性能的评价指标。

精确度表示正确分类的样本占总样本的比例, 如式(5), 精确度越大表示检测越精确, 在一定程度上也可以反映误检越少。召回率表示正确的正样本占总正样本的比例, 如式(6), 召回率越大表示检测越全面, 反之, 说明了漏检越少。mAP 表示所有类别的平均精度 AP (Average Precision), 如式(7), mAP 越大表示模型检测精度越大。其中 AP 是指在不同召回率下精确度的均值, 也就是以召回率为横坐标, 精确度为纵坐标的 P-R 曲线与坐标轴围成的面积, 如式(8)。mAP_0.5 表示 IoU 阈值为 0.5 时所有目标类别的平均检测精度。IoU (Intersection over Union, 交并比) 计算的是预测边框与真实边框的重叠率。运行时间表示模型进行检测消耗的时间, 单位为毫秒 (ms)。FPS 表示模型

每秒能检测多少张图像,单位为 $\text{f} \cdot \text{s}^{-1}$,用于衡量模型的实时性,FPS 越大表示检测速度越快,即模型越具有实时性。GFLOPs 表示每秒 10 亿次的浮点运算,用于衡量训练模型时的计算复杂度,GFLOPs 越大表示模型越复杂。参数量表示模型训练需要的参数总数,参数量越大所需的内存越大。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{mAP} = \frac{\sum_{n=1}^N \text{AP}(n)}{N} \quad (7)$$

$$\text{AP} = \int_0^1 p(r) dr \quad (8)$$

式(5)和式(6)中,TP (True Positives)表示明渠漂浮垃圾被正确检测为漂浮垃圾的个数;FN (False

Negatives)表示明渠漂浮垃圾没有被检测出来的个数,即漏检的个数;FP (False Positives)表示将其他的物体检测成漂浮垃圾的个数,即误检的个数。公式(7)中, n 表示类别, N 表示总类别。公式(8)中, $p(r)$ 表示以召回率为横坐标、精确度为纵坐标的 P-R 曲线。

3.3 对比实验

3.3.1 添加 CBAM 模块与 T-CBAM 模块对比

将直接添加 CBAM 模块与添加 T-CBAM 模块的检测效果进行对比。它们的检测效果如表 1 所示,其中 YOLOv5s-CBAM 表示添加 CBAM 模块,YOLOv5s-T-CBAM 表示添加 T-CBAM 模块。在复杂度和参数方面虽然不是最佳,但是相差不大,而且从模型的检测精度和检测速度两个方面分析可知,YOLOv5s-T-CBAM 的检测效果更好,精确度、召回率和平均精度均值都是最高的,且检测速度仍可以达到 $89 \text{ f} \cdot \text{s}^{-1}$,可以保持较好的实时性。

表 1 添加 CBAM 模块与 T-CBAM 模块结果对比

Methods	Precision	Recall	mAP_0.5	Times	FPS	GFLOPs	Params
YOLOv5s	0.802	0.798	0.826	11.1	90	15.8	7 022 326
YOLOv5s-CBAM	0.820	0.833	0.846	11.8	85	15.9	7 046 583
YOLOv5s-T-CBAM	0.847	0.842	0.878	11.2	89	16	7 124 369

3.3.2 与其他目标检测算法的对比

为了验证改进后算法的优越性,还将改进的 YOLOv5s 算法与 YOLOv3、YOLOv4、SSD、YOLOX、DETR、YOLOv7 等目标检测算法进行对比。由表 2 分析可知,在检测精度方面,改进的 YOLOv5s 算法明显优于其他算法,mAP_0.5 可以达到 0.899,而且还兼顾高精度和高召回率,既保证了检测的准确性,还减少

了检测的误检和漏检;在运行时间和检测速度方面,改进的 YOLOv5s 算法的检测速度达到了 $96 \text{ f} \cdot \text{s}^{-1}$,且运行时间最短,满足实时性的要求;在复杂度和参数方面,改进的 YOLOv5s 算法仅次于 YOLOv5s,且相差不大,以最小的牺牲提高了目标检测的精度和速度。因此,改进的 YOLOv5s 算法优于其他目标检测算法。

表 2 改进后算法与其他目标检测算法的对比

Methods	Precision	Recall	mAP_0.5	Times	FPS	GFLOPs	Params
YOLOv3	0.651	0.849	0.826	22.7	44	155.08	61 523 734
YOLOv4	0.702	0.853	0.841	24.6	41	141.44	63 937 686
SSD	0.753	0.634	0.712	19.2	52	123.08	25 221 643
DETR ^[16]	—	—	0.647	49.8	20	85.6	16 016 754
YOLOX ^[17]	—	—	0.862	35.7	28	42.2	33 707 164
YOLOv7 ^[18]	0.824	0.834	0.858	21.5	47	103.2	37 218 327
YOLOv5s	0.802	0.798	0.826	11.1	90	15.8	7 022 326
YOLO-Z ^[5]	—	—	0.873	11.7	85	—	—
TPH-YOLOv5 ^[8]	0.832	0.856	0.864	16.9	59	38.4	10 252 595
改进 YOLOv5s	0.864	0.860	0.899	10.4	96	16.1	7 132 128

3.4 消融实验

为了验证各个改进模块优化的作用,该文进行了消融实验,一共设计了 5 组实验,其实验结果如表 3 所示。其中原 YOLOv5s 表示输入图像的尺寸为 640×640 ,YOLOv5s 表示输入图像的尺寸为 $1\,280 \times 1\,280$ 。

由表 3 分析可知,在检测精度和速度方面,增加网络的输入尺度提升了检测精度,但是检测速度会有所下降;修改特征融合过程不仅可以大大提升检测精度、精确度和召回率,还可以提高检测速度,弥补增加网络输入尺度时造成的速度损失;引入 T-CBAM 提升了检测精

度和精确度,但是检测速度会降低;在增加网络的输入尺度的同时,既修改网络融合过程,又增加 T-CBAM 模块,即提出的改进算法,它不仅最大程度地提升了检测精度,使精确度和召回率达到最高,而且检测速度也有了一定提升。在复杂度和参数方面,增加网络的输入尺度不会改变模型的复杂度,参数量也不会增加,但是修改特征融合过程和引入 T-CBAM 都会影响模型复杂度,增加参数量,通过具体数据对比分析可知,改进 YOLOv5s 算法的 GFLOPs 仅比原 YOLOv5s 算法增加了 0.3,且参数量变化不大,以最小的牺牲提升了模型的检测性能。由图 7 分析可知,改进 YOLOv5s 的平均精度每次都比原 YOLOv5s 高,保持实时性的同时在精度方面有很大的提升。

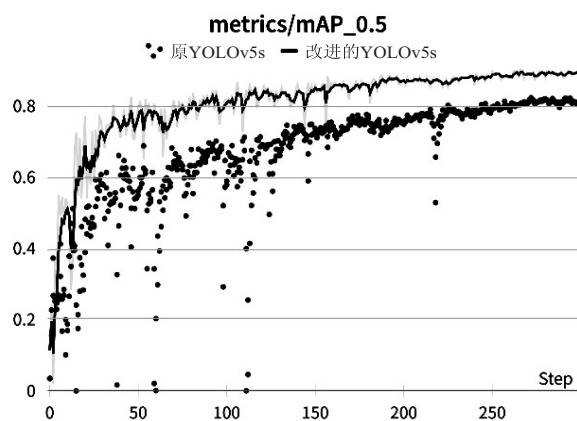


图7 原 YOLOv5s 与改进 YOLOv5s 的 mAP_0.5 对比

表3 消融实验

Methods	Precision	Recall	mAP_0.5	Times	FPS	GFLOPs	Params
原 YOLOv5s	0.803	0.731	0.804	10.6	94	15.8	7 022 326
YOLOv5s	0.802	0.798	0.826	11.1	90	15.8	7 022 326
YOLOv5s+Bi-PANet	0.844	0.856	0.880	10.2	98	16.1	7 087 871
YOLOv5s+T-CBAM	0.847	0.842	0.878	11.2	89	16	7 124 369
YOLOv5s+Bi-PANet+T-CBAM	0.864	0.860	0.899	10.4	96	16.1	7 132 128

3.5 检测结果分析

为了验证改进的 YOLOv5s 算法的可行性,分别使用原 YOLOv5s 算法和改进的算法对测试集上航拍图像的漂浮垃圾进行检测,以检测瓶子为例,如图 8 为原 YOLOv5s 算法和改进的 YOLOv5s 算法在不同干扰场景下的检测结果对比图,其中,左图为原 YOLOv5s 算法的检测结果,右图为改进 YOLOv5s 算法的检测结果。图 8(a) 表示水面漂浮垃圾较小的检测,可以看到左图会漏检掉远处较小的漂浮垃圾,而右图可以检测到远处的小漂浮垃圾;(b) 表示水面倒影干扰的检测,

左图将岸边物体在水面的倒影误检成漂浮垃圾,右图则检测正常;(c) 表示水面强光干扰的检测,左图受强光干扰,存在漏检,右图不仅将每个漂浮垃圾检测都出来了,而且检测的精度比左图高。综上,改进的 YOLOv5s 算法更优于原 YOLOv5s 算法,原 YOLOv5s 算法在复杂检测场景的干扰下检测性能较差,容易出现漏检和误检的情况,并且相较于改进的 YOLOv5s 算法,其目标定位精度偏低。因此,改进的 YOLOv5s 性能更优越,定位精度更准确。



(a) 水面漂浮垃圾较小的检测



(b) 水面有倒影干扰的检测



(c)水面有强光干扰的检测

图 8 不同干扰场景下算法检测结果对比

4 结束语

针对明渠漂浮垃圾检测存在的一系列问题,如:相关数据集较少,航拍图像中的明渠漂浮垃圾尺寸较小,明渠漂浮垃圾检测易受水面倒影和强光反射影响,基于原始的 YOLOv5s 进行改进优化。首先,通过数据增强进行数据集扩充,避免数据集太小出现过拟合的问题;然后,在借鉴 BiFPN 加权双向特征融合的思想修改特征融合过程,提高小目标的检测精度和速度;并且在 Neck 和 Head 部分之间添加 3 个改进的 CBAM 注意力机制,能够有效降低检测的漏检率和误检率,提高检测精度;最后,增大网络输入的分辨率,使图像具有更加精确的位置信息和更加丰富的细节,有利于小目标特征信息的提取。由实验结果可知,所提方法在保持实时性的同时,检测的平均精度达到了 89.9%,具有较好的应用价值。但是,如何将算法部署到嵌入式平台使无人机和无人船能够对视野中的目标进行实时识别是未来亟待解决的问题。

参考文献:

- [1] LI X, TIAN M, KONG S, et al. A modified YOLOv3 detection method for vision-based water surface garbage capture robot[J]. International Journal of Advanced Robotic Systems, 2020, 17(3): 1729881420932715.
- [2] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv:1804.02767, 2018.
- [3] 唐小敏, 舒远仲, 刘文祥, 等. 基于 SSD 深度网络的河道漂浮物检测技术研究[J]. 计算机技术与发展, 2020, 30(9): 154-158.
- [4] 李德鑫, 闫志刚, 孙久运. 基于无人机视觉的河道漂浮垃圾分类检测技术研究[J]. 金属矿山, 2021, 50(9): 199-205.
- [5] BENJUMEA A, TEETI I, CUZZOLIN F, et al. YOLO-Z: improving small object detection in YOLOv5 for autonomous vehicles[J]. arXiv:2112.11798, 2021.
- [6] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 10781-10790.
- [7] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 8759-8768.
- [8] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the IEEE/CVF international conference on computer vision. Montreal: IEEE, 2021: 2778-2788.
- [9] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). Munich: Springer, 2018: 3-19.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.
- [11] 王玲敏, 段 军, 辛立伟. 引入注意力机制的 YOLOv5 安全帽佩戴检测方法[J]. 计算机工程与应用, 2022, 58(9): 303-312.
- [12] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. Seattle: IEEE, 2020: 390-391.
- [13] 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.
- [14] CHENG Y, ZHU J, JIANG M, et al. FloW: a dataset and benchmark for floating waste detection in inland waters[C]//2021 IEEE/CVF international conference on computer vision (ICCV). Montreal: IEEE, 2021: 10933-10942.
- [15] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions[J]. arXiv:2112.05561, 2021.
- [16] 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.
- [17] GE Z, LIU S, WANG F, et al. YOLOX: exceeding yolo series in 2021[J]. arXiv:2107.08430, 2021.
- [18] 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[J]. arXiv:2207.02696, 2022.