

•智能交叉科学与工程•

DOI:10.12454/j.jsuese.202400031



本刊网刊

基于深度学习的河道航拍影像检测算法研究

邓业发^{1,2,3}, 郝志红^{1,3}, 吴鑫森^{1,3*}

(1. 河北农业大学 城乡建设学院, 河北 保定 071001; 2. 国家北方山区农业工程技术研究中心, 河北 保定 071000;
3. 河北省河湖长制研究院, 河北 保定 071001)

摘要:利用先进手段快速、精准地发现河湖四乱和水质等问题是提高河湖管理及河湖长制工作效率的重要途径之一,通过无人机巡视结合图像识别技术有望有效解决传统人工巡查河道观测范围受限、效率低下和涉河问题响应慢等问题,但目前河道水面环境复杂,河道漂浮垃圾形状不规则且形态多样、尺寸不一,蓝藻分布较广,非法采砂行为发生地点较为隐蔽且易受遮挡,依靠传统的图像识别算法进行检测存在较大的挑战性。针对这一挑战,本文提出一种基于改进 YOLO v5s 的河道巡查图像识别(YOLO v5s-CDF)模型。首先,使用 FocalNext 模块替代 Backbone 骨干网络中的 C3 模块,通过引入深度可分离卷积和空洞卷积,增强网络对小目标的特征提取能力;接着,在输出部分添加 Context Aggregation 注意力机制,通过引入上下文聚合机制来调整输入数据的权重,使模型能够更加关注图像中的关键信息;同时,使用解耦头(Decouple Head)替换原有的耦合检测头,将特征提取和任务预测分开,加快了网络的收敛速度,进一步提升了模型对小目标的检测能力。实验结果表明,YOLO v5s-CDF 模型的平均精度均值(mAP)为 86.7%,比 YOLO v5s 模型高了 4.1 个百分点,对比 YOLO v7-tiny 和 YOLO X-s 模型,平均精度均值提高了 20.4 个百分点和 8.3 个百分点。实例分析结果说明本文提出的 YOLO v5s-CDF 模型是实现河湖问题精准识别的可靠且有效的方法。

关键词:YOLO v5s; 小目标检测; 河道垃圾; FocalNext 模块; 注意力机制

中图分类号:TN911.73; TP391.4

文献标志码:A

文章编号:2096-3246(2026)01-0334-11

河道作为水的重要载体,其健康状况直接关系到人类生存和生态环境的稳定^[1]。随着城镇化的加快和人类活动的增加,河道环境正承受着多重压力。城市垃圾和污水处理不当,以及农药、化肥的过度施用导致河道漂浮物堆积和蓝藻爆发^[2];此外,城镇化进程中对砂石的巨大需求也催生了河道区域的非法采砂活动^[3-4],导致河道景观和水生态平衡被破坏。在全面推行河长制的背景下,提高河道巡查和监管的效能显得尤为关键^[5]。传统的人工巡查方式在效率、范围和频次上存在明显的限制,难以满足实时、全面的河道保护需求^[6]。近年来,航拍影像和目标检测技术的发展为河道巡查提供了新的方向。通过这些技术不仅能够快速获取河道区域的详细信息,还能在短时间内识别出河道中的垃圾、蓝藻和疑似非法采砂行为,为环境保护提供有力的技术支持。

随着机器视觉和深度学习的不断发展,目标检测技术由基于传统机器学习的方法逐渐过渡到基于深度学习的方法。传统的目标检测技术主要使用特征提取和分类器来检测目标,但特征选择过程往往需要人工参与,而且难以应对复杂和变化的目标。左建军等^[7]通过监控摄像机搜集河道水面影像,采用背景差法分割出包含水面漂浮物的样本,然后利用反向传播算法(BP)对水面漂浮物样本特征进行训练,建立漂浮物分类器,并使用分类器识别水面漂浮物。王敏等^[8]提出了一种水上物体检测分割算法,该算法基于水体和水面漂浮物饱和度的差异对水上物体进行特征提取,然后利用区域一致性的平均值梯度(GOA)进行边缘检测和定位。江杰等^[9]提出了基于混合高斯模型的背景差分法和帧差法来监测水面漂浮物。汤伟等^[10]使用基于色差灰度模型与改进的最大类间方差(OTSU)法

收稿日期:2024-01-11 修回日期:2024-02-19 网络出版日期:2024-05-07

基金项目:河北省水利科研和推广计划项目(2023-22);国家自然科学基金青年基金项目(51709278)

作者简介:邓业发,男,硕士生。研究方向:水利信息学。Email:z008qq@163.com

*通信作者:吴鑫森,教授,Email:wuxinmiao001@163.com

分割出水面目标,通过计算目标对象的质心坐标和最小外界矩,对目标进行标记,为机器人的动作提供依据。传统水面漂浮物检测方法通常依赖手工特征提取,不具备自我学习和自我优化的能力。对于复杂环境下水面漂浮物的检测和识别能力有限,容易受到光照变化、水面波动和水面反射等因素的影响,泛化能力有限。而基于深度学习的检测方法能够自动从原始数据中习得复杂的特征表示,对不同尺度和复杂度的漂浮物都能有较好的检测效果。

基于深度学习的目标检测技术主要利用神经网络来自动学习特征并进行分类和回归,一般分为2种类型:单阶段目标检测算法和双阶段目标检测算法。单阶段目标检测算法的典型代表是YOLO^[11-12]和SSD(single shot multibox detector)^[13],双阶段目标检测算法的典型代表是R-CNN系列(包括R-CNN^[14]、Fast R-CNN^[15]、Faster R-CNN^[16])。与双阶段目标检测算法相比,单阶段目标检测算法不需要生成候选区域,而是直接在一次前向传播中进行分类和定位,因此速度较快。

唐小敏等^[17]利用无人机采集河流视频影像,每隔90帧截取河流视频图像,用于训练基于SSD和Faster R-CNN的模型,实验结果显示,基于ResNet-101的SSD模型的检测精度更高。李国进等^[18]提出一种基于Faster R-CNN的改进识别与定位方法,该方法采用Faster R-CNN算法对水面漂浮物进行初次识别和定位,对输出的识别结果与定位框采用类别激活网络(CA)去除边界框,运用像素点来标注目标位置。李德鑫^[19]利用无人机采集河道漂浮垃圾图片,构建基于YOLO v5s的河道漂浮垃圾检测模型,并将检测结果在WebGIS平台进行可视化表达,提高河道管理的效率。周宇浩^[20]结合语义分割算法、MobileNetV3和算法和YOLO算法,实现了河岸线检测以及非法捕鱼现象识别。李科^[21]利用内河漂浮垃圾数据集Flow构建基于YOLO v7的河道小漂浮物检测模型,通过引入特征增强模块和卷积块注意力模块(CBAM),同时提出一种基于尺度惩罚交并比的方法,提高了算法对水面小目标漂浮物的检测性能,随后使用量化工具对模型进行优化,减少对计算资源的消耗以便在移动设备上部署。陈任飞等^[22]提出了一种基于SSD的水面漂浮物检测方法,该算法选择轻量化MobileNetV2网络作为骨干网络,在预测层中使用深度可分离卷积,同时使用动态特征金字塔网络(DyFPN)提高模型的多尺度检测性能,最后引入模型量化方法降低算法的运算成本。包学才等^[23]提出一种基于改进DeeplabV3+的水面多类型漂浮物识别的语义分割算法,该算法选择

Xception网络作为主干网络,同时引入注意力机制为不同特征层分配不同权重系数,利用全连接条件随机场平衡局部与全局语义信息,整体提高多类型漂浮物的识别率与检测精度。

将目标检测技术用于河道巡查的研究较少且主要用于识别河道漂浮物,如垃圾袋、塑料瓶、树枝等,结合无人机航拍影像对河道中的蓝藻和疑似非法采砂行为的检测研究也较少,不能满足全面推行河湖长制背景下河道巡查治理的需要。本研究将YOLO v5s模型应用于河道巡查领域,通过引用FocalNext模块^[24]、Context Aggregation注意力机制^[25]和解耦头(Decouple Head)^[26]对YOLO v5s模型进行优化,提升对小目标的检测能力,实现对河道垃圾、蓝藻和疑似非法采砂行为的检测,提升模型的识别精度。

1 数据采集与预处理

1.1 数据采集

实验所用数据由大疆Air 2s无人机对河北省境内多条河流倾斜摄影所得,为了让数据更具广泛性,使用无人机在不同时段、不同高度以及不同角度对多条河流进行摄影作业,增加影像的随机性,进而提高算法训练的准确度。数据集包含3类河湖问题影像,分别为河道垃圾、蓝藻和疑似非法采砂行为。运用标注工具Labelimg对影像进行标注,制作成Pascal VOC^[27]格式数据集,并按照8:2的比例划分为训练集和验证集。无人机航拍3类河湖问题数据集图片示例如图1所示。

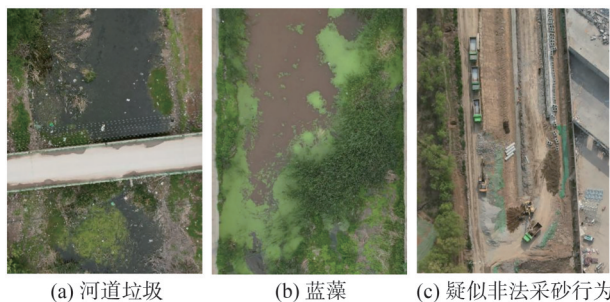


图1 无人机航拍3类河湖问题数据集图片示例

Fig. 1 Sample images from the UAV aerial photography dataset of three categories of river and lake issues

1.2 数据预处理

为了避免模型过拟合,提高算法的鲁棒性,对训练集中的图片进行数据增强。使用Imgaug图像增强库对图像同时进行翻转、颜色变换和仿射变换,数据增强前后效果如图2所示。

经过数据预处理后,包含河道垃圾、蓝藻和疑似非法采砂行为影像的数据集共包括1 308幅图像。考虑到北方许多排沥河道除汛期外,基本没有径流,所



(a) 图像增强前效果 (b) 图像增强后效果

图2 数据增强前后效果

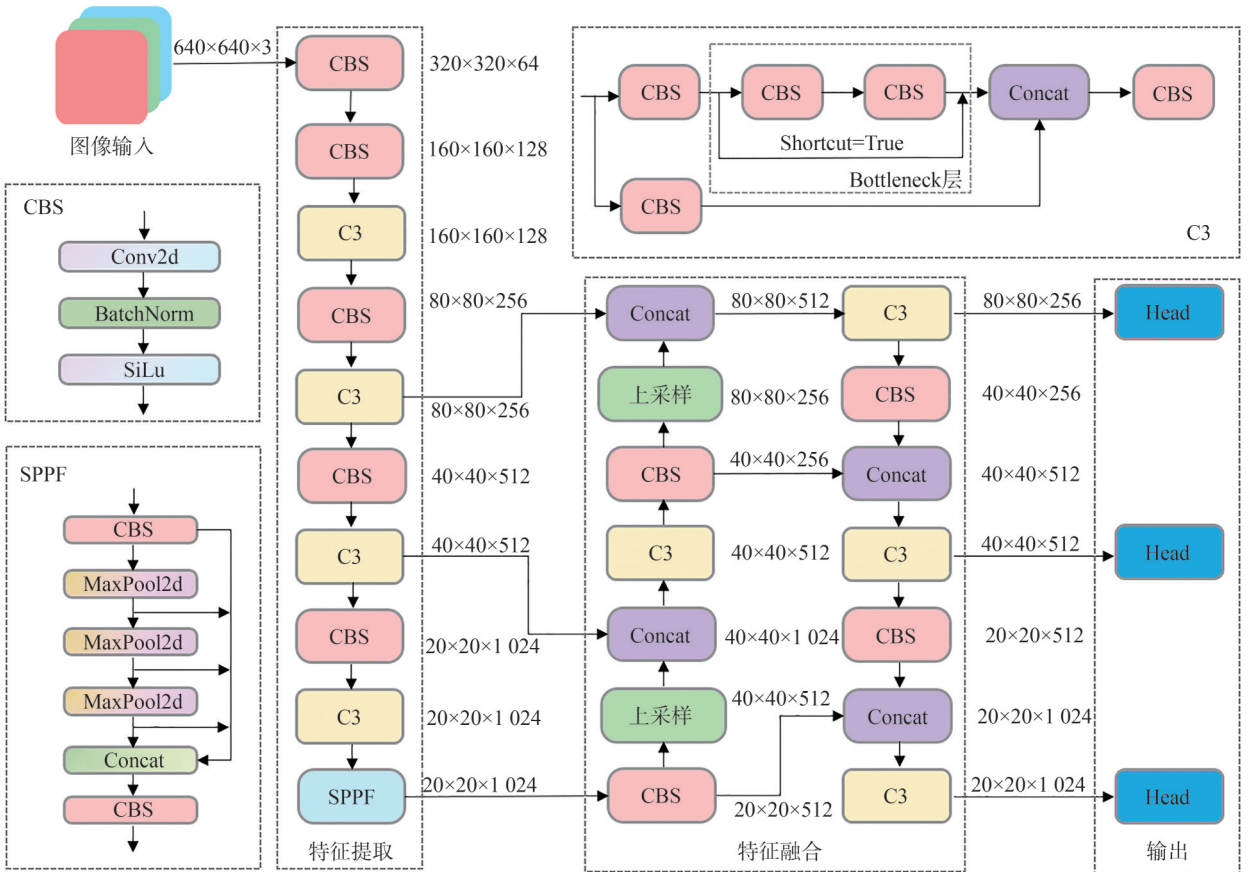
Fig. 2 Comparison of effects before and after data augmentation

以河道垃圾的图像样本由两部分组成：一部分是有水河道的图像，另一部分则是无水河道的图像。对河道疑似非法采砂行为的检测是通过检测出现在河道管理范围内的挖掘机、铲车、自卸运输车等工程车辆来实现的，非河道整治工程期间在河道管理范围内出现工程车辆则列为疑似非法采砂行为。

2 YOLO v5s 检测算法及其改进

YOLO v5s 网络包含 4 部分，分别为输入端、主干

网络 (Backbone)、颈部网络 (Neck) 和检测头 (Head)，各部分的主要作用分别为图像输入、特征提取、特征融合和输出。输入端主要对输入图像进行预处理，其核心功能包括 Mosaic 图像增强、自适应锚框计算及自适应图片缩放等。Backbone 主要用于提取特征和缩小特征图，由 CBS、C3 模块和快速空间金字塔池化 (SPPF) 组成。Neck 采用特征金字塔网络 (FPN) 加路径聚合网络 (PAN) 的结构：FPN 通过上采样的方式将深层网络高层次的语义信息融合到稍浅层的网络；PAN 通过下采样的方式将浅层的位置信息传递到深层网络，进行多尺度的特征融合。Head 主要用于检测目标，由 3 个 1×1 的卷积层组成，分别对 Neck 输出的 3 张特征图降维，然后生成 20×20 、 40×40 和 80×80 的特征图，通道数为 $3 \times (5+X)$ 。其中，3 表示每个网格预设的 3 个不同大小的锚框 (Anchor)，5 分别对应的是预测框的中心点横坐标、纵坐标、宽度、高度和置信度， X 为类别数。图 3 为 YOLO v5s 模型结构。



注：Conv2d 为普通卷积，BatchNorm 为批归一化，SiLu 为激活函数，MaxPool2d 为最大池化，Concat 为特征拼接，Shortcut=True 表示进行残差连接。

图3 YOLO v5s 模型结构

Fig. 3 Structure of the YOLO v5s model

2.1 改进的 YOLO v5s-CDF 检测算法

针对 YOLO v5s 算法在检测小目标和类别不平衡目标方面的不足，本实验基于 YOLOv5-7.0 算法进行改进，Backbone 引入 FocalNext 模块替换原有网络的

C3 模块 (改进 1)，在 Neck 和 Head 之间插入 Context Aggregation 模块 (改进 2)，同时使用 YOLO X 的 Decouple Head (解耦头) 替换默认的 YOLO 检测头 (改进 3)。图 4 为改进后的 YOLO v5s-CDF 模型结构。

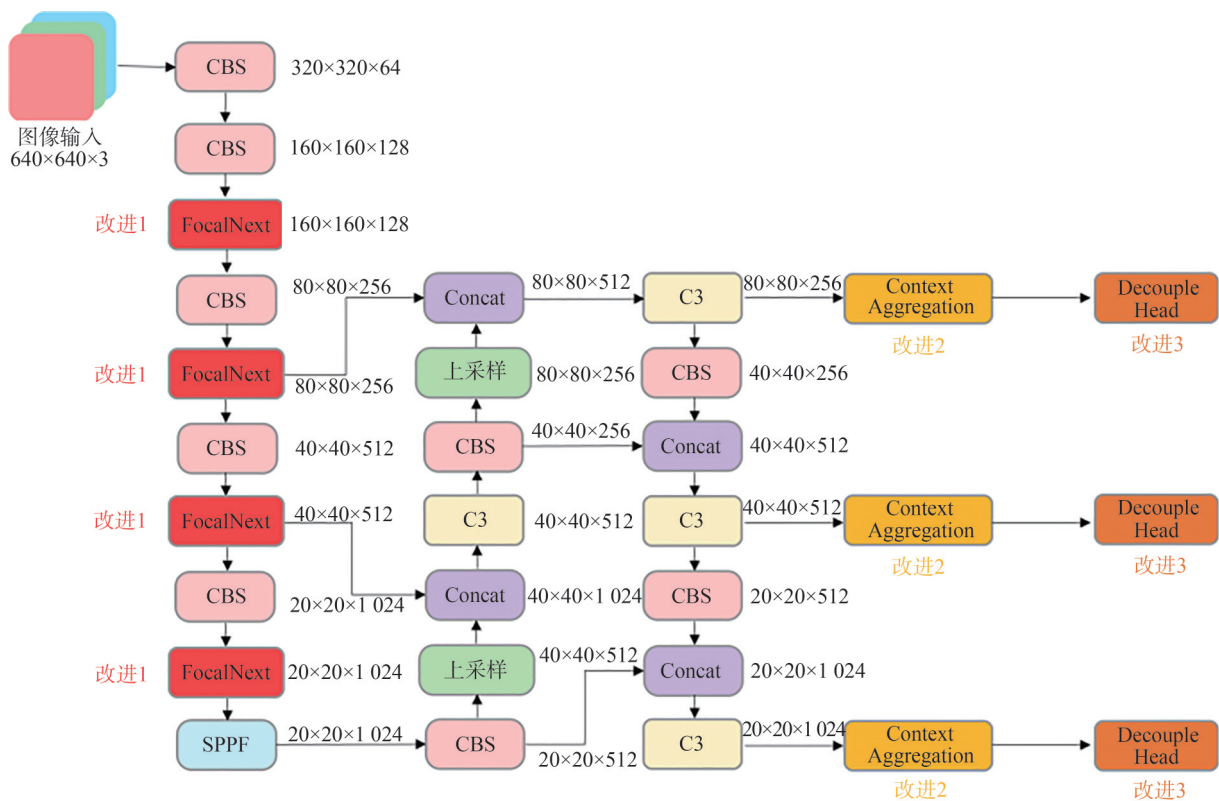


图 4 YOLO v5s-CDF 模型结构

Fig. 4 Structure of the YOLO v5s-CDF model

2.1.1 FocalNext 模块

在数据集中,蓝藻与另两类目标的尺度相差较大。为了提高算法对多尺度目标的检测能力以及优化小目标检测性能,使用FocalNext模块替换Backbone骨干网络中的C3模块。FocalNext模块在ConvNext模块^[28-29]的基础上引入包含空洞卷积的深度可分离卷积层和两次残差连接,深度可分离卷积由深度卷积与逐点卷积两部分组成。其中,深度卷积在每个输入通道上应用独立的卷积核以进行逐通道特征提取,实现了参数和计算量的显著减少,但由于各通道互不交互,因此无法利用通道间相关性,且不能改变特征图的通道维度。为此,在深度卷积之后引入逐点卷积,用于在深度方向上高效地融合不同通道的信息,并通过调整其卷积核数量灵活地实现特征图通道维度的升降。在此基础上,模块中引入的空洞卷积能够在不增加参数和计算量的前提下,有效扩大感受野,从而捕捉更广泛的上下文信息,极大地提升了模型的整体性能^[30]。

将空洞卷积嵌入深度卷积并与逐点卷积协同,使FocalNext可以融合细粒度的局部特征和粗粒度的全局特征,进一步提升网络的特征提取和多尺度信息融合能力,同时保持较低计算成本。图5为ConvNext和FocalNext模块结构。图5中, n 为特征图的通道数, $d7 \times$

7 表示 7×7 的深度卷积, r 为空洞卷积的膨胀率, 1×1 表示 1×1 的逐点卷积,GELU为一种激活函数。图5(b)中:橙色部分为FocalNext模块在ConvNext模块基础上的改进;每个 $d7 \times 7$ 后面都跟有一个层归一化(LayerNorm, LN)和一个GELU激活函数,为简洁未画出。FocalNext模块首先使用两个 7×7 深度卷积,在第二个 7×7 深度卷积上应用扩张率为3的空洞卷积,两次卷积

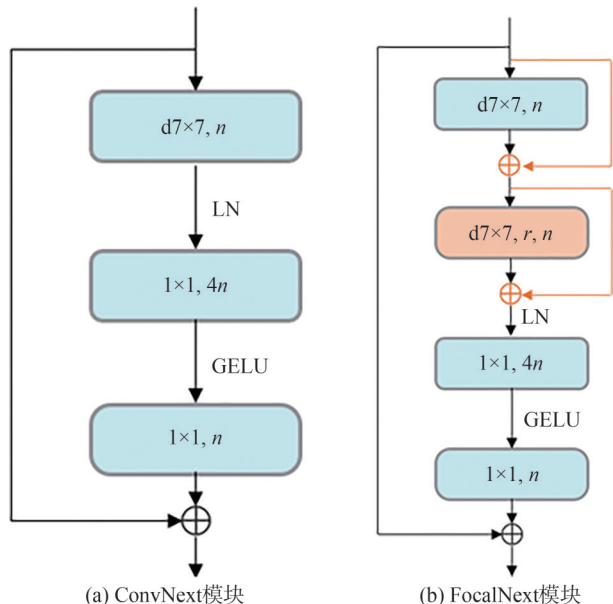


图 5 ConvNext 和 FocalNext 模块结构

Fig. 5 Structure of the ConvNext and the FocalNext modules

的结果再辅以残差连接;然后通过两个 1×1 的卷积层和一个GELU激活函数。在经过第一个 1×1 卷积层后,特征图通道数扩大为原来的4倍;在经过第二个卷积层后,压缩到原来的通道数。通过先扩大通道数(维度)再缩小,配合GELU非线性激活函数,可以增强模型的非线性,最终模块输出通过残差连接与模块的原始输入相加。在后续消融实验中,FocalNext模块可以明显提升模型的检测性能。

2.1.2 Context Aggregation 注意力机制

注意力机制通过给输入数据的每一部分赋予不同的权重,使模型从众多输入数据中选取最关键的信息。Context Aggregation通过全局上下文聚合和自适应空间加权两种机制来增强输入特征的判别力,具体包含4个卷积层,分别用于处理输入数据的不同方面。其中,卷积层A和K用于生成注意力权重,卷积层V则用于产生与注意力权重相关联的值张量。随后,卷积层M将中间处理层的通道数恢复为原始输入通道数。通过一系列的矩阵运算,最终得到一个精细调整后的特征表示,并将其与原始输入相加以产生最终的输出。Context Aggregation模块结构如图6所示。图6中, H 、 W 、 C 、 C' 分别为特征图的高度、宽度、通道数和降维后的通道数,Softmax和Sigmoid代表两个激活函数。通过引入上下文聚合机制来调整输入数据的权重,从而提取更加精细的特征表示,并将其嵌入更大的神经网络,以提升整体模型的性能。

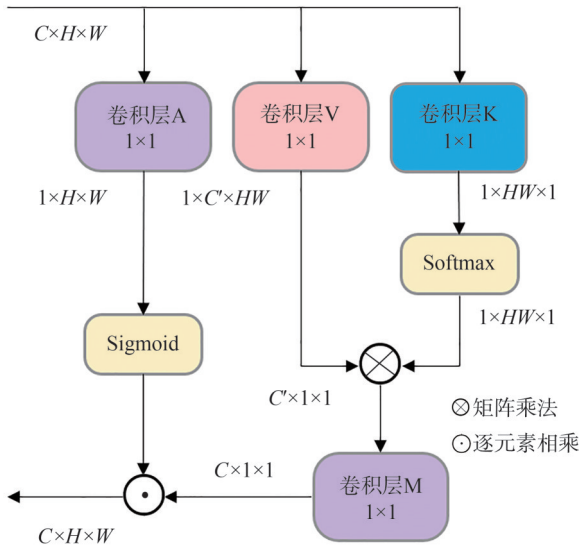


图6 Context Aggregation 模块结构

Fig. 6 Structure of the Context Aggregation module

2.1.3 解耦头

在目标检测任务中,检测头主要负责对象的分类和定位任务,YOLO v5s 检测头采用耦合头结构,即将对象分类和边界框回归任务耦合到一个共享的网络

头中,但是分类任务和回归任务往往存在冲突,因为两者所关注的特征性质不同。分类任务更加关注所提取的特征与哪个已有分类相似度更高,回归任务更关注锚框(anchor box)与真实框(ground-truth box)之间的位置关系,通过学习两者的相对坐标偏移量来修正初始框,从而生成最终的预测框,实现精确定位。因此,本文采用解耦头。

解耦头首先使用1个 1×1 的卷积层降低通道数,然后在分类和回归的分支后面分别接入两个串行的 3×3 卷积层以提取不同的特征,最后分别生成类别(Cls.)和边界框回归参数(Reg.)以及目标置信度(IoU)。解耦头的结构如图7所示。

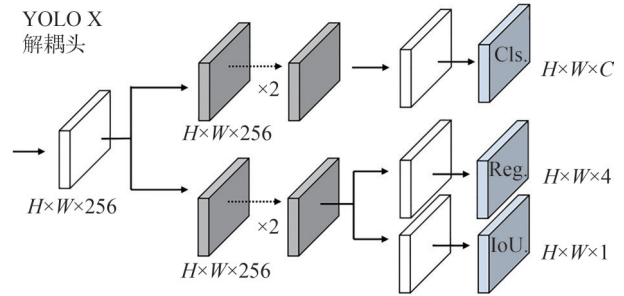


图7 解耦头的结构

Fig. 7 Structure of the Decouple Head

2.2 实验参数设置

本实验所用的CPU为Intel® Xeon® Platinum 8255C CPU @ 2.50 GHz, GPU为RTX 3080(显存为10GB), RAM为40GB, 操作系统为Ubuntu 20.04, Python版本为3.8, Pytorch版本为1.10, CUDA版本为11.3.1。实验所设置的超参数为:训练图像尺寸为 640×640 , 批量大小(batch-size)为8, 训练周期(epoch)为300, 线程数(num workers)为8, 初始学习率(initial learning rate)为0.01, 动量因子(momentum)为0.937, 权重衰减系数为0.0005。

2.3 评价指标

为评估模型性能,选择精确率(precision)、召回率(recall)、F1值、平均精度(AP)和平均精度均值(mAP)作为评价指标,计算公式分别如下。

$$a = \frac{f}{f+g} \quad (1)$$

$$b = \frac{f}{f+h} \quad (2)$$

$$c = 2 \times \frac{a \times b}{a + b} \quad (3)$$

$$d = \int_0^1 a(b) db \quad (4)$$

$$e = \frac{1}{m} \sum_{i=1}^m d_i \quad (5)$$

式(1)~(5)中: a 为精确率,是指在所有被预测为正样

本的样本中实际为正样本的比例; b 为召回率,是指在所有实际正样本中,被正确预测为正样本的比例; c 为精确率和召回率的调和平均值,即F1值; d 为单个类别下的平均精度; $a(b)$ 为 a 关于 b 的函数; m 代表检测样本类别数,本文中为3; e 为所有类别 d 的平均值,它表示模型在多个类别上的平均检测性能; f 为真正例(true positives),表示正确地将正样本预测为正样本的数量; g 为假正例(false positives),表示错误地将负样本预测为正样本的数量; h 为假负例(false negatives),表示错误地将正样本预测为负样本的数量。

其中,正样本是正确分类出的类别所对应的样本,如需要识别图像中的河道垃圾,则正样本为河道垃圾,其他的目标对象如树木则为负样本。通过绘制精确率-

表1 消融实验结果对比

Tab. 1 Comparison of ablation experiment results

模型	参数量/ 10^6	浮点运算量/ 10^9	精确率/%	召回率/%	平均精度均值/%
YOLO v5s	7.0	15.8	85.9	76.4	82.6
YOLO v5s+FocalNext 模块	6.8	15.5	85.3	78.6	84.4
YOLO v5s+解耦头	8.7	21.3	87.5	78.3	85.0
YOLO v5s+FocalNext 模块+解耦头	8.6	21.0	87.3	80.6	86.3
YOLO v5s-CDF	9.3	21.9	87.0	80.8	86.7

将YOLO v5s模型Backbone中的C3模块替换为FocalNext模块,相比初始模型,平均精度均值和召回率分别提升了1.8个百分点和2.2个百分点。通过引入空洞卷积和深度可分离卷积,增大了网络感受野,提升了网络的特征提取能力,同时参数量和运算量也有小幅下降。

将YOLO v5s模型的耦合检测头替换为解耦头,相比原始检测头,平均精度均值、召回率和精确率分别提升了2.4个百分点、1.9个百分点和1.6个百分点。通过将分类任务和回归任务分开处理,改善了网络的性能和精度,但模型参数量和运算量有相应的增加。

将FocalNext模块和解耦头同时集成到YOLO v5s模型中,与YOLO v5s模型相比,平均精度均值、召回率和精确率分别提升3.7个百分点、4.2个百分点

表2 YOLO v5s-CDF模型与YOLO v5s模型的实验结果对比

Tab. 2 Comparison of experimental results between YOLO v5s-CDF model and YOLO v5s model

模型	类别	精确率/%	召回率/%	c	平均精度/%	平均精度均值/%
YOLO v5s	河道垃圾	88.0	64.0	0.80	74.8	82.6
	疑似非法采砂行为	88.7	87.2		90.2	
	蓝藻	81.0	77.9		82.8	
YOLO v5s-CDF	河道垃圾	90.3	71.3	0.83	80.9	86.7
	疑似非法采砂行为	86.7	89.8		92.2	
	蓝藻	84.0	81.3		86.9	

召回率曲线并计算曲线下的面积得出平均精度(AP),使用交并比(IoU)判定预测框和真实框之间的重叠程度,当预测框与真实框之间的交并比大于0.5时,将其视为正确的检测结果。同时,使用参数量(parameters)和浮点运算量(FLOPs)表示模型的复杂程度。

3 实验结果及分析

3.1 消融实验

为了验证本文不同改进方案对网络模型的影响,在自制的3类河湖问题数据集上进行多次实验和消融实验,详细分析FocalNext模块、解耦头以及Context Aggregation注意力机制的可行性和有效性。表1为消融实验结果对比。

和1.4个百分点,但参数量和运算量比单独引入解耦头要低。

本文提出的YOLO v5s-CDF模型在引入FocalNext模块和解耦头的同时,将Context Aggregation模块插入Neck和Head之间,与YOLO v5s相比,平均精度均值、召回率和精确率分别提升4.1个百分点、4.4个百分点和1.1个百分点。

3.2 YOLO v5s模型改进前后对比

采用本文提出的YOLO v5s-CDF模型在测试集进行性能测试。表2为YOLO v5s-CDF模型与YOLO v5s模型的实验结果对比。由表2可见,YOLO v5s-CDF模型的性能参数明显优于YOLO v5s模型,改进的YOLO v5s-CDF模型的平均精度均值为86.7%,其中河道垃圾、疑似非法采砂行为和蓝藻的平均精度分别为80.9%、92.2%和

- ing management[D]. Zhengzhou: North China University of Water Resources and Electric Power,2020.[潘科. 漳河非法采砂治理研究[D]. 郑州:华北水利水电大学,2020.]
- [4] Wang Haijing, Xu Xiaohua, Liu Yisheng, et al. Research on the effectiveness of sand mining and transportation supervision in rivers and lakes in Jiangxi Province[J]. Technical Supervision in Water Resources, 2023, 31(4): 9–12.[王海菁, 许小华, 刘毅生, 等. 江西省河湖采运砂监管成效研究[J]. 水利技术监督, 2023, 31(4): 9–12.]
- [5] Zhang Yu. Design and implementation of intelligent river patrol system based on specific object detection[D]. Nanchang: Nanchang Hangkong University, 2022.[张宇. 基于河道目标检测的智慧巡河系统设计与实现[D]. 南昌:南昌航空大学, 2022.]
- [6] Yao Yalan. Awareness and reflections on strengthening smart regulation of rural water systems[J]. Gansu Agriculture, 2022(2): 99–102.[姚亚兰. 加强农村水系智慧监管的认识与思考[J]. 甘肃农业, 2022(2): 99–102.]
- [7] Zuo Jianjun, Wu Youfu. An intelligence monitoring technique for floater surface of water[J]. Software Guide, 2013, 12(4): 150–152.[左建军, 吴有富. 水面漂浮物智能监控技术[J]. 软件导刊, 2013, 12(4): 150–152.]
- [8] Wang Min, Zhou Shudao. Static water object detection segmentation[J]. Research and Exploration in Laboratory, 2010, 29(6): 30–32.[王敏, 周树道. 静态水上物体检测分割算法[J]. 实验室研究与探索, 2010, 29(6): 30–32.]
- [9] Jiang Jie, Li Gang. Research on automatic monitoring methods for floating objects in rivers [J]. Yellow River, 2010, 32(11): 47–48.[江杰, 李刚. 河流漂浮物的自动监测方法研究[J]. 人民黄河, 2010, 32(11): 47–48.]
- [10] Tang Wei, Liu Siyang, Gao Han, et al. A target detection algorithm for surface cleaning robot based on machine vision[J]. Science Technology and Engineering, 2019, 19(3): 136–141.[汤伟, 刘思洋, 高涵, 等. 基于视觉的水面垃圾清理机器人目标检测算法[J]. 科学技术与工程, 2019, 19(3): 136–141.]
- [11] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 779–788.
- [12] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger [C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6517–6525.
- [13] Liu Wei, Anguelov D, Erhan D, et al. SSD: Single shot multi-box detector[C]//Computer Vision—ECCV 2016. Cham: Springer, 2016: 21–37.
- [14] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580–587.
- [15] Girshick R. Fast R-CNN[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 1440–1448.
- [16] Ren Shaoqing, He Kaiming, 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.
- [17] Tang Xiaomin, Shu Yuanzhong, Liu Wenxiang, et al. Research on river floating object detection technology based on SSD deep network[J]. Computer Technology and Development, 2020, 30(9): 154–158.[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.]
- [18] Li Guojin, Yao Dongyi, Ai Jiaoyan, et al. Detection and localization of floating objects *via* improved faster R-CNN [J]. Journal of Xinyang Normal University (Natural Science Edition), 2021, 34(2): 292–299.[李国进, 姚冬宜, 艾娇燕, 等. 基于改进 Faster R-CNN 的水面漂浮物识别与定位[J]. 信阳师范学院学报(自然科学版), 2021, 34(2): 292–299.]
- [19] Li Dexin. Research and application of river floating garbage detection based on YOLOv5s[D]. Xuzhou: China University of Mining and Technology, 2021.[李德鑫. 基于 YOLOv5s 的河道漂浮垃圾检测研究与应用[D]. 徐州: 中国矿业大学, 2021.]
- [20] Zhou Yuhao. Research on river image detection algorithm based on deep learning using UAV[D]. Hangzhou: China University of Metrology, 2021.[周宇浩. 基于深度学习的无人机河道图像检测算法研究[D]. 杭州: 中国计量大学, 2021.]
- [21] Li Ke. Research on deep learning based algorithm for detection of tiny objects on water surface[D]. Chongqing: Southwest University, 2023.[李科. 基于深度学习的水面小物体检测算法研究[D]. 重庆: 西南大学, 2023.]
- [22] Chen Renfei, Peng Yong, Wu Jian, et al. Intelligent detection of floating objects on water surface based on deep learning [J]. Advanced Engineering Sciences, 2023, 55(3): 165–174.[陈任飞, 彭勇, 吴剑, 等. 基于深度学习的水面漂浮物智能检测方法[J]. 工程科学与技术, 2023, 55(3): 165–174.]
- [23] Bao Xuecai, Liu Feiyan, Nie Jugen, et al. Research on segmentation method of multiple types of floating objects on water surface based on improved DeeplabV³⁺[J]. Water Resources and Hydropower Engineering, 2024, 55(4): 163–175.[包学才, 刘飞燕, 聂菊根, 等. 基于改进 DeeplabV³⁺的

- 水面多类型漂浮物分割方法研究[J].水利水电技术(中英文),2024,55(4):163–175.]
- [24] Zhang Gang, Li Ziyi, Tang Chufeng, et al. CEDNet: A cascade encoder-decoder network for dense prediction[J]. Pattern Recognition, 2025, 158: 111072.
- [25] Liu Ye, Li Huifang, Hu Chao, et al. Learning to aggregate multi-scale context for instance segmentation in remote sensing images[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(1): 595–609.
- [26] Ge Zheng, Liu Songtao, Wang Feng, et al. YOLOX: Exceeding YOLO series in 2021 [EB/OL]. (2021–08–06)[2024–01–11]. <http://arxiv.org/abs/2107.08430>.
- [27] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303–338.
- [28] Liu Zhuang, Mao Hanzhi, Wu Chaoyuan, et al. A ConvNet for the 2020s[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 11966–11976.
- [29] Woo S, Debnath S, Hu Ronghang, et al. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 16133–16142.
- [30] Chen L C, Zhu Yukun, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision—ECCV 2018. Cham: Springer, 2018: 833–851.
- [31] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 7464–7475.

Research on River Channel Aerial Image Detection Algorithm Based on Deep Learning

DENG Yefa^{1,2,3}, QIE Zhihong^{1,3}, WU Xinmiao^{1,3*}

(1. College of Urban and Rural Construction, Agricultural University of Hebei, Baoding 071001, China;

2. National Engineering Research Center for Agricultural in Northern Mountainous Area, Baoding 071000, China;

3. River and Lake Chief System Research Institute of Hebei Province, Baoding 071001, China)

Abstract:

Objective Utilizing advanced methods to rapidly and accurately identify issues such as river and lake disorder and water quality represents a key approach to improving the efficiency of river and lake management and the river chief system. This approach aims to effectively address problems related to limited observation range, low efficiency, and delayed response to river-related issues that occur with traditional manual patrols employing unmanned aerial vehicle (UAV) patrols integrated with image recognition technology. However, current challenges persist, including complex river surface environments, irregular shapes, and diverse forms of floating debris in rivers of varying sizes, widespread distribution of algal blooms, and concealed or easily obstructed locations of illegal sand mining activities. These factors make it difficult for traditional image recognition algorithms to accurately detect such issues. Therefore, this study proposes a river patrol image recognition algorithm based on an improved YOLO v5s, referred to as the YOLO v5s–CDF algorithm.

Methods The YOLO v5s–CDF algorithm introduced several key improvements to enhance the detection performance of the original YOLO v5s model. First, the FocalNext module replaced the C3 module in the backbone network, incorporating depth-wise separable convolutions and dilated convolutions to improve feature extraction capabilities for small objects. The depth-wise separable convolution applied different convolutional kernels to each input channel, extracting important features from multiple channels, while the dilated convolutions increased the receptive field without adding parameters or computational cost, capturing broader contextual information. Second, the Context Aggregation attention mechanism was added between the neck and head structures to adjust the weights of input data, enabling the model to focus on key image information. This attention mechanism combined channel attention and spatial attention to refine the feature representation, enhancing the model's ability to capture critical details. Lastly, the Decoupled Head replaced the original coupled detection head, separating feature extraction and task prediction to accelerate network convergence and further enhance small object detection. The dataset used in this study consisted of aerial images captured by DJI Air 2s drones, covering various rivers in Hebei Province, China. The images were annotated using the Labeling tool and divided into training and validation sets in an 8:2 ratio. Data augmentation techniques, such as flipping, color transformation, and affine transformation, were applied to the training set to improve the model's robustness and generalization ability. The study evaluated the model's performance using metrics such as precision, recall, F1-score, Average Precision (AP), and mean Average Precision (mAP). The number of parameters and floating-point operations per second (FLOPs) were utilized to represent the model's complexity.

Results and Discussions The experimental results demonstrated that the YOLO v5s–CDF model achieved a mean Average Precision (mAP) of

86.7%, surpassing the original YOLO v5s model by 4.1%. The improved model exhibited significant enhancements in both precision and recall. The precision increased from 85.9% in the original YOLO v5s model to 87.0% in the YOLO v5s-CDF model, while the recall improved from 76.4% to 80.8%, indicating a substantial reduction in missed detections. When compared to other models, such as YOLO v7-tiny and YOLO X-s, the YOLO v5s-CDF model outperformed them with mAP improvements of 20.4 percentage points and 8.3 percentage points, respectively. Among all target categories, the detection performance for river garbage showed the most significant improvement, with the average precision increasing by 6.1 percentage points, rising from 74.8% to 80.9%, and the precision also increasing by 2.3 percentage points, from 88% to 90.3%. In addition, the recall demonstrated an improvement of 7.3 percentage points, increasing from 64.0% to 71.3%, highlighting the YOLO v5s-CDF model's remarkable enhancement in detecting small targets. For targets related to suspected illegal sand mining activities, the average precision increased from 90.2% to 92.2%; however, the precision experienced a decline from 88.7% to 86.7%, while the recall improved from 87.2% to 89.8%. Regarding green algae targets, the average precision rose from 82.8% to 86.9%, accompanied by an increase in precision from 81.0% to 84.0% and an improvement in recall from 77.9% to 81.3%. These improvements demonstrated the YOLO v5s-CDF model's effectiveness in detecting several river issues across different object categories. The visual analysis of detection results further confirmed the superior performance of the YOLO v5s-CDF model. When compared to the original YOLO v5s model, the improved model exhibited enhanced robustness and generalization capabilities in detecting river issues under complex environments, such as varied illumination conditions, water surface fluctuations, and reflections. The YOLO v5s-CDF model successfully identified and localized a higher number of objects, particularly small and irregularly shaped floating garbage, which were often missed by the original model. The visual results aligned with the quantitative improvements observed in precision and recall. Ablation studies revealed the individual contributions of each introduced module, with the FocalNext module improving mAP by 1.8 percentage points, the Decoupled Head by 2.4 percentage points, and the combination of both modules by 3.7 percentage points. The addition of the Context Aggregation attention mechanism on top of the improvements from the FocalNext module and the Decoupled Head further improved the mAP by 4.1 percentage points compared to the original YOLO v5s model, while only slightly increasing the model's parameters and computational complexity. However, the model's performance in detecting submerged objects and illegal sand mining activities occurring at night required further improvement. Future research will focus on enhancing the model's applicability and accuracy under various environmental conditions, such as incorporating river topography changes as detection targets using 3D drone imagery and improving the collection of submerged object data.

Conclusions The YOLO v5s-CDF algorithm presents a viable technical approach to addressing challenges in river and lake supervision by integrating drone-based remote sensing technology with advanced image recognition methods. The enhanced model, which incorporates the FocalNext module, Context Aggregation attention mechanism, and Decoupled Head, demonstrates superior accuracy and robustness in detecting river debris, algal blooms, and potential illegal sand mining activities. It highlights the broad application potential of UAV remote sensing and target detection technologies in ecological environment management. In addition, integrating the YOLO v5s-CDF algorithm with complementary technologies, such as water quality sensors, hydrological models, and geographic information systems (GIS), can provide a comprehensive framework for river health evaluation and management.

Key words: YOLO v5s; Small object detection; River garbage; FocalNext module; Attention mechanism

(编辑 陈雪)

引用格式: Deng Yefa, Qie Zhihong, Wu Xinmiao. Research on river channel aerial image detection algorithm based on deep learning[J]. *Advanced Engineering Sciences*, 2026, 58(1): 334-344. [邓业发, 郟志红, 吴鑫淼. 基于深度学习的河道航拍影像检测算法研究[J]. *工程科学与技术*, 2026, 58(1): 334-344.]