

## 多层特征图信息融合的海滩小目标检测\*

肖智阳<sup>1</sup>, 林坚普<sup>1,2\*\*</sup>, 张永爱<sup>1,2</sup>, 林志贤<sup>1,2</sup>

(1. 福州大学先进制造学院, 福建 泉州 362200; 2. 中国福建光电信息科学与技术创新实验室, 福州 350116)

**摘 要:** 提出一种多层特征图信息融合的海滩小目标检测方法, 从上下文信息与强化特征图信息融合的角度提升小目标游客的检出率。首先, 透过更全面、有效的GAM注意力机制思想结合CSP结构提出GCSAM结构, 用于增强检测YOLOv5模型中主干网络跨纬度感受区, 聚焦小目标特征学习; 其次, 在颈部融合端使用BIFPN结构优化YOLOv5网络中PANet结构, 补全跨层特征信息之间的传递, 使得特征图包含更多的上下文信息; 最后, 采用幂变换改进YOLOv5网络中CI-OU\_Loss为Alpha-CIOU\_Loss, 有效提升预测框的回归精度。实验证明, 在满足实时性要求的前提下, 相比于原始YOLOv5网络, 文中方法在海滩小目标游客检测上查准率提升2.00%, 查全率提升5.33%, 平均精度均值提升4.36%。文中方法在海滩小目标游客密集、遮挡、目标更小的情况下具有更好的鲁棒性。

**关键词:** 深度学习; 小目标检测; 注意力机制; 特征图

**中图分类号:** TP394.1 **文献标志码:** A **文章编号:** 1005-488X(2023)02-0142-08

## Detection of Beach Small Object Based on Multi-layer Feature Map Information Fusion

XIAO Zhiyang, LIN Jianpu, ZHANG Yongai, LIN Zhixian

(1. School of Advanced Manufacturing, Fuzhou University, Quanzhou Fujian 362200, CHN; 2. Fujian Science & Technology Innovation Laboratory for Optoelectronic Information of China, Fuzhou 350116, CHN)

**Abstract:** Aiming at the problem of insufficient detection accuracy of complex beach small object tourists, a beach small object tourist detection method based on multi-layer feature map information fusion was proposed, and the context and feature map information was used to improve the detection rate of small object tourists. Firstly, GCSAM structure was proposed by combining the more comprehensive and effective GAM attention mechanism idea with CSP structure. The GCSAM structure was

**收稿日期:** 2022-12-07

\* **基金项目:** 国家重点研发计划 (No.2021YFB3600603); 福建省自然科学基金 (No.2019J01221, No.2020J01468); 福建省教育厅中青年教师教育科研项目 (No.JAT210030)

**作者简介:** 肖智阳 (1998—), 男, 主要研究方向为数字图像处理技术; (E-mail: 812381505@qq.com);

林坚普 (1989—), 男, 博士, 讲师, 主要研究方向为数字图像处理技术、人工智能技术、3D立体显示技术、新型显示技术; (E-mail: ljp@fzu.edu.cn);

张永爱 (1977—), 男, 博士, 研究员, 主要研究方向为场致发射显示器、数字图像处理技术、3D显示技术等; 林志贤 (1975—), 男, 博士, 教授, 研究方向为图像处理以及平板显示驱动技术。

\*\* **通讯作者**

used to enhance the cross-latitude receptive field of YOLOv5 backbone. The backbone network was focused on small object feature learning. Secondly, the PANet structure was replaced by BIFPN structure in YOLOv5 network to complete transmission of feature information across layers at the neck network, more context was included in the feature map. Finally, the power transform was used to improve CIOU\_Loss to Alpha-CIOU\_Loss in YOLOv5 network, and the regression accuracy of prediction frame was effectively improved. Experimental results showed that in comparison with original YOLOv5 network, the precision was improved by 2.00%, the recall was improved by 5.33%, and the mean average precision was improved by 4.36% with real-time requirements. Moreover, this method had better robustness in the case of dense tourists, occlusion, and smaller objects.

**Key words:** deep learning; small object detection; attention mechanism; feature map

## 引 言

滨海旅游业是中国海洋经济发展的重要支柱产业,2021年滨海旅游业占主要海洋产业增值构成的44.9%<sup>[1]</sup>。但海滩游客的人身安全问题也日渐凸显,2001~2019年共发生海滩游客安全事故619起,造成1141人伤亡,事故发生频率逐年上升<sup>[2]</sup>。为应对可能出现的海滩游客安全事故,各沿海地区积极推动海滩安全系统建设,现阶段的海滩安全系统,主要依靠人工把守在监控前预防事故发生。随着人工智能技术飞速发展,智能监控被广泛运用在城市安防系统。海滩监控由于地理环境限制无法架设在海滩附近,导致在监控画面中获取的单个游客目标占整个画面的比重很小。当前的智能检测方案在此类复杂场景下小目标检测精度不足,因此提出一种针对海滩小目标游客检测方法以保证游客的人身安全具有重要的实际意义。

海滩游客目标人口密度大、细节特征少,监控设备捕获的图像可能存在少许噪声,现有的小目标检测方法<sup>[3-9]</sup>在针对此类具体场景时会出现各式各样的问题。例如,Hu和Ramanan<sup>[10]</sup>提出一种凹视觉描述符检测方法,此方法有效地证明特征图的上下文信息对小目标检测起到重要作用,但由于该方法高昂的计算代价使得无法部署在海滩小目标检测此类对实时性要求较高的场景中。Kisantal等人<sup>[11]</sup>提出一种“Copy-Pasting strategies”的策略来增加小目标在图像中出现的频率,使网络更好地捕捉小目标信息从而提高小目标的检出率,考虑到海滩游客数量密度足够,过度的数据增强可能导致网络训练出现过拟合现象。Alsubaei等人<sup>[12]</sup>设计的DLSODC-GWM检测方法用于小目标垃圾的检测与分类中,巧妙地设计超参数优化方法进行参数调

优从而获得很好的检测效果,但自适应的超参数调节对噪声鲁棒性较差,海滩环境下使用会受到限制。

综上所述,为更好的满足海滩小目标游客检测场景需求,兼顾小目标游客检测的精确性与实时性,文章从利用上下文信息以及强化特征图信息融合<sup>[13-15]</sup>的角度提出一种基于YOLO的多层特征图信息融合(Feature Map Information Fusion YOLO, FMIF-YOLO)小目标检测方法,在保证实时性前提下增强海滩小目标游客的检出率。

## 1 多层特征图信息融合的海滩小目标检测算法

### 1.1 网络整体结构

YOLOv5作为当前检测领域检测性能卓越、工程部署率较高的目标检测模型之一,网络结构主要由4部分组成:输入端、主干网络端、颈部融合端以及预测输出端。但该网络结构针对海滩小目标游客检测仍然存在一些不足:主干网络端所采用的跨阶段局部(Cross Stage Partial, CSP)结构与卷积交替堆叠,使得主干网络端对小目标游客的特征提取能力欠佳。颈部融合端采用PANet(Path Aggregation Networks)结构,由于其结构为单层单向信息传递方式,因此会造成一定的游客特征信息损失,从而降低检测精度。预测输出端采用完全交并比损失(Complete Intersection Over Union Loss, CIOU\_Loss)作为损失函数结合非极大值抑制处理因遮挡导致的检测准确率下降问题,但其会受到噪声等因素影响造成定位偏差降低检出率,无法适应海滩监控设备所采集的图像噪声。

针对上述问题,文中在主干网络端与颈部融合端根据全注意力机制(Global Attention Mechanism, GAM)思想<sup>[16]</sup>提出一种全局跨阶段注意力机制(Global Cross Stage Attention Mechanism, GC-SAM)结构,帮助主干网络更加细致地提取小目标游客特征信息,再将YOLOv5网络颈部融合端中PANet结构调整为双向特征金字塔网络(Bidirectional Feature Pyramid Network, BIFPN)结构融合

多尺度特征图上下文信息,使得送入检测端的小目标游客特征信息得到最大程度的保留,最后采用Alpha-CIOU\_Loss优化YOLOv5网络预测输出端中的CIOU\_Loss使检测框定位回归更加准确,解决海滩监控采集图像中的噪声所带来的定位精度损失问题,网络整体结构如图1所示。

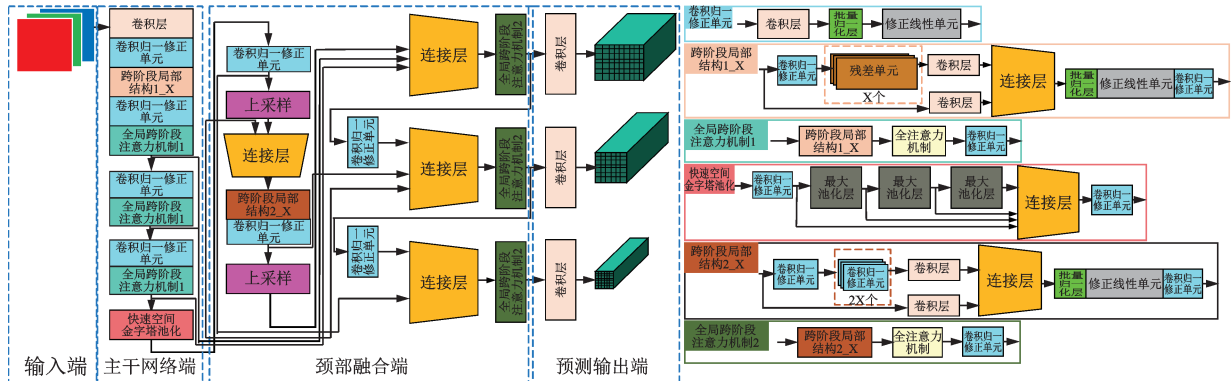


图1 FMIF-YOLO网络整体结构图

Fig.1 The structure of FMIF-YOLO network

1.2 主干网络端改进

考虑到海滩游客目标小、信息少的特点,强化已知信息的提取至关重要,注意力机制通过找寻原有数据之间的关联性,忽略无关信息的同时注重关键特征信息的提取。因此,引入注意力机制解决YOLOv5主干网络的CSP与卷积交叠结构提取小目标游客特征信息能力不足的问题。文章提出GC-SAM结构如图2所示,增加YOLOv5主干网络端跨纬度的感受区,针对小目标信息进行更加细致提

取,采用在尽可能减少信息弥散的情况下放大全局维度交互特征的设计思路,相比于如今主流的小目标检测注意力机制更好地从通道、空间宽度和空间高度三个维度整合所有的重要特征,增强跨维度的信息,提升抑制非重要像素信息的效率。对于一个给定的特征图 $F_1 \in \mathbb{R}^{C \times H \times W}$ ,通过通道注意力模块及空间注意力模块分别生成的特征图 $F_2 = M_c(F_1) \otimes F_1$ 、 $F_3 = M_s(F_2) \otimes F_2$ , $M_c$ 与 $M_s$ 分别表示通道注意力操作及空间注意力操作。

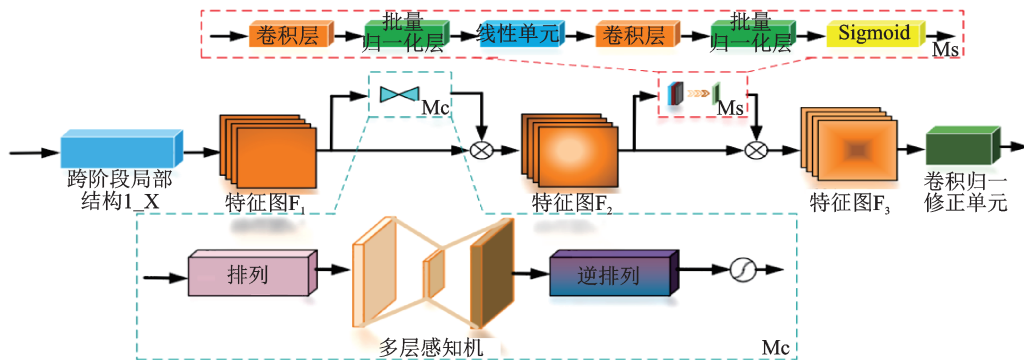


图2 GC-SAM模块结构图

Fig.2 Structural schematic of GC-SAM module

卷积块注意力模块(Convolutional Block Attention Module, CBAM)<sup>[17]</sup>在三维通道、空间高度、空间宽度之间只选取两个方向进行注意力权重施加,忽略三个方向间的相互作用,从而丢失跨维信息。

GC-SAM在通道注意力操作上直接采用3维排列(Permutation)的方式保留三个方向上的特征信息,通过多层感知机放大跨维通道及空间的依赖性。相比于挤压-激发(Squeeze-and-Excitation layer, SE-

layer)<sup>[18]</sup>注意力模块通过池化操作造成非重要像素信息抑制失败导致效率不佳的问题,GCSAM在空间注意力操作上通过删除挤压池化操作来保留空间特性映射获得更多的空间信息,再通过两个卷积层进行空间信息融合,提高抑制非重要像素信息的效率。

最后,不仅将GAM思想与主干网络端相结合,而且将其与颈部融合端相结合(如图1中GCSAM2模块所示,融合特征图经过CSP2\_X模块后,再进入GAM模块,最后通过卷积模块进行维度调整),一方面使得特征图更加关注到需要被检测的小目标,另一方面可以有效忽略特征融合后特征图中不需要被重点关注的区域,便于后续预测输出层进行游客目标及其位置的预测。

### 1.3 颈部融合端改进

原始YOLOv5在颈部融合端采用PANet<sup>[19]</sup>结构,通过FPN自顶向下传达强语义特征,再通过PAN自底向上传达强定位特征,达到不同主干层对不同检测层进行参数聚合的目的。但其结构为单层单向信息传递,由于特征图在缩放的过程中必定存在信息的丢失,单层传递的方式必然会缺失部分跨层的语义特征与定位特征。对此文章引入BIFPN结构<sup>[20]</sup>进入YOLOv5的颈部融合端(具体结构见图1颈部融合端)中调整PANet结构。PANet结构与BIFPN结构对比如图3所示。

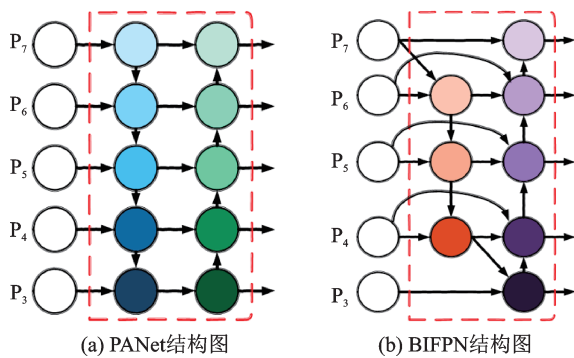


图3 PANet结构与BIFPN结构对比

Fig.3 Structural comparison between PANet and BIFPN

以 $P_6$ 层输出特征图为例,PANet的计算公式为:

$$P_6^{\text{out}} = \text{Conv}(P_6^{\text{in}} + \text{Resize}(P_7^{\text{in}}) + \text{Resize}(P_5^{\text{out}})) \quad (1)$$

而BIFPN的计算公式为:

$$P_6^{\text{id}} = \text{Conv}\left(\frac{\omega_1 \cdot P_6^{\text{in}} + \omega_2 \cdot \text{Resize}(P_7^{\text{in}})}{\omega_1 + \omega_2 + \epsilon}\right) \quad (2)$$

$$P_6^{\text{out}} = \text{Conv}\left(\frac{\omega_1' \cdot P_6^{\text{in}} + \omega_2' \cdot P_6^{\text{id}} + \omega_3' \cdot \text{Resize}(P_5^{\text{out}})}{\omega_1' + \omega_2' + \omega_3' + \epsilon}\right) \quad (3)$$

其中 $P_x^{\text{in}}$ 表示第 $X$ 层的输入特征图, $P_x^{\text{id}}$ 表示第 $X$ 层的中间特征图, $P_x^{\text{out}}$ 表示第 $X$ 层的输出特征图, $\omega_i$ 表示针对融合的各个尺度特征图增加一个权重信息调节每个尺度的贡献度, $\epsilon$ 用于防止数据溢出。PANet结构中 $P_6$ 层仅仅将不同尺寸的特征图调整为相同的大小后求和,实际中不同分辨率的输入特征对输出端的作用是不相等的。而BIFPN结构采用跨尺度连接的方式,最后送入预测端的特征图针对不同输入特征设置可调节的权重信息,让网络自适应学习调整权重。BIFPN促进不同层之间的特征融合,使得最后送给预测输出端的特征图包含更多的上下文信息,可以更好地帮助小目标游客传递语义及定位信息。

### 1.4 激活函数改进

边界框的定位回归关乎到最终能不能精确给出目标检测结果。YOLOv5预测输出端采用CI-OU\_Loss将重叠面积、中心点距离以及纵横比作为约束项加入损失函数中,图4为CI-OU\_Loss示意图,其表达式为:

$$\text{CI-OU\_Loss} = 1 - \text{IOU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \beta v \quad (4)$$

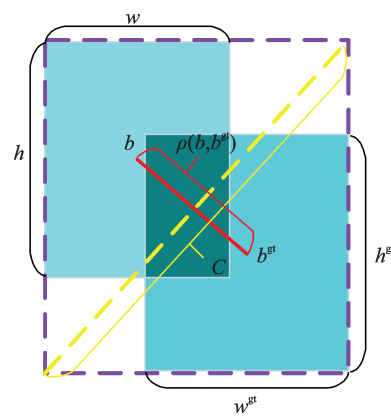


图4 CI-OU\_Loss示意图

Fig.4 Schematic of CI-OU\_Loss

其中 $b, b^{\text{gt}}$ 表示预测框与真实框的中心点坐标。 $\rho(\cdot)$ 表示计算两个框之间中心点的欧氏距离。 $c$ 表示将预测框与真实框完全封闭最小外接矩形的对角线距离。 $v$ 用于描述预测框与真实框纵横比的一致性, $\beta$ 为权重系数, $v$ 与 $\beta$ 的表达式为:

$$v = \frac{4}{\pi} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (5)$$

$$\beta = \frac{v}{(1 - \text{IOU}) + v} \quad (6)$$

其中  $w^{gt}$ 、 $h^{gt}$  和  $w$ 、 $h$  表示真实框与预测框的长宽。

尽管 CIOU\_Loss 可以解决许多 IOU 无法规避的问题,但 CIOU\_Loss 在一些小数据集和噪声情况下受到数据集或者图像质量的限制会出现预测框的回归精度下降问题。原因在于不同目标间的相似性或者噪声引入带来的错误信息导致有些目标 CIOU\_Loss 计算时正确的类别预测框与错误的类别预测框差距不明显。为消除在海滩监控下可能出现的噪声以及图像质量较差所导致的检测率下降问题同时避免引入过多算法使得模型复杂,文章在 CIOU\_Loss 的基础上引入幂变换得到 Alpha-CIOU\_Loss<sup>[21]</sup>,其表达式为:

$$\text{Alpha-CIOU\_Loss} = 1 - \text{IOU}^\alpha + \frac{\rho^{2\alpha}(b, b^{gt})}{c^{2\alpha}} + (\beta v)^\alpha, \quad \alpha > 0 \quad (7)$$

其中  $\alpha$  为幂变换约束项,由于幂变换不会改变 CIOU\_Loss 关键性质,而且幂变换使得其他约束项随 IOU 一同进行单调变化,改变  $\alpha$  的大小可以帮助模型关注更多高 IOU 数值的候选框,放大相似信息之间的差异性,提高网络定位和分类的性能。实验表明  $0 < \alpha < 1$  时倾向于降低最终性能。减少高 IOU 目

标的损失和梯度,最终会产生更多定位较差的目标; $\alpha > 1$  时引入一段绝对损失量,可以很好地优化所有检测层的候选目标,同时由于  $\alpha$  对高 IOU 数值候选框的绝对梯度变化,可以加快网络对这些候选框的学习。因此文章通过 Alpha-CIOU\_Loss 实现抗干扰能力达到对海滩小目标模型不同检测层预测框的回归精度提升。

## 2 实验与分析

### 2.1 实验数据集及实验环境

由于当前海滩安全建设仍处于发展阶段,可用公开数据信息不足,文章选择在自建海滩监控场景下小目标游客数据集上验证算法的有效性,同时在 Visdrone 公开数据集辅助验证。数据集通过采集监控视频画面制作,数据分辨率为  $1\,920 \times 1\,080$ ,帧率为 25 fps,共计 1227 张图像。从中选取 85% 图像作为训练集,15% 图像作为测试数据集用于网络模型性能指标的评价。为了提高模型的泛化能力,对原始数据集进行 YOLOv5 输入端数据增强操作,进一步扩充小目标数据。通过图像标注工具 LabelImg 到对应文件名的 xml 文件,“.xml”文件中记录对应标注框的坐标信息以及目标类型等信息。部分标注过程及海滩监控场景如图 5 所示。



图 5 部分标注过程及海滩监控场景

Fig.5 Some labeling processes and beach monitoring scenes

文章实验配置环境如下:Ubuntu18.04 操作系统、CPU 为 Intel Core i9-8950、GPU 为 NVIDIA GeForce GTX 1080(8GB 显存),内存 32GB。基于 Pytorch 架构使用 Python3.8 环境和通过 GPU 加速平台 CUDA 10.1 实现文章方法。模型训练过程中图像尺寸统一为 640,总迭代次数 200,初始学习率设为 0.0100,采用随机梯度下降优化方式,Alpha-CIOU\_Loss 中  $\alpha$  根据实验测试取 3 为最优。

### 2.2 公共评价指标

针对海滩场景下小目标游客的检测任务,采用

查准率(Precision)、查全率(Recall)、平均精度均值(mean Average Precision, mAP)以及每秒传输帧数(Frames Per Second, FPS)等公共评价指标对实验网络模型进行全面评价。

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

$$\text{mAP} = \frac{\sum_0^1 P(R) dR}{N} \times 100\% \quad (10)$$

其中  $TP$  表示预测正确的正样本,  $FP$  表示负样本预测为正样本,  $FN$  表示正样本预测为负样本,  $N$  表示需要检测的所有类别数, 文章中  $N=1$ , IOU 阈值设置为模型评估通用的 0.5。一般地, 检测速度不小于实时监控传输帧率(25 fps)即可认为该算法模型满足实时性要求。

### 2.3 消融实验

为验证文章所提的改进对海滩小目标游客检测性能的有效性, 设计一组消融实验, 对比分析了

(1) 原 YOLOv5 模型; (2) 在原模型上仅使用 Alpha-CIOU\_Loss 损失函数; (3) 在原模型上仅使用 BIFPN 连接结构; (4) 在原模型上仅使用 GCSAM 模块; (5) 在原模型上同时使用 Alpha-CIOU\_Loss 损失函数与 BIFPN 连接结构; (6) 在原模型上同时使用 Alpha-CIOU\_Loss 损失函数与 GCSAM 模块; (7) 在原模型上同时使用 GCSAM 模块与 BIFPN 连接结构; (8) 在原模型上同时使用 Alpha-CIOU\_Loss 损失函数、BIFPN 连接结构与 GCSAM 模块。具体性能如表 1 所示。

表 1 不同改进策略及其组合使用网络模型实验测试结果

Tab.1 Experimental results of network model with different improvement strategies and their combinations

网络结构	查准率/(%)	查全率/(%)	平均精度均值/(%)	每秒传输帧数
YOLOv5	89.49	76.70	83.20	41.49
YOLOv5+Alpha-CIOU <sub>Loss</sub>	86.85	78.09	83.83	42.37
YOLOv5+BIFPN	88.38	76.76	82.70	40.00
YOLOv5+GCSAM	90.89	78.90	86.06	28.57
YOLOv5+BIFPN+Alpha-CIOU <sub>Loss</sub>	87.57	78.14	84.30	42.55
YOLOv5+GCSAM +Alpha-CIOU <sub>Loss</sub>	90.52	81.28	87.22	28.49
YOLOv5+GCSAM+BIFPN	91.02	78.98	86.12	28.24
YOLOv5+GCSAM+BIFPN+Alpha-CIOU <sub>Loss</sub>	91.49	82.03	87.56	28.09

根据表中实验数据可以看出: 当单独使用 Alpha-CIOU\_Loss 时相比于原 YOLOv5 模型结构查全率提升 1.39%, 查准率下降 2.64%, 平均精度均值提升 0.63%, 说明 Alpha-CIOU\_Loss 可以有效降低将小目标游客错分为背景信息的情况, 提升小目标的漏检率。

当 BIFPN 结构与其他模块组合使用时均比单独使用某一方法有提升, Alpha-CIOU\_Loss 结合 BIFPN 时查全率进一步提升 0.05%, 有效改善查准率下降, 提升整体模型检测效果; GCSAM 结合 BIFPN 同样进一步提升检测效果, 有效证明 BIFPN 对比 PANet 结构有更佳效果。

采用 GCSAM 结构能很大程度提升原始 YOLOv5 的模型检测效果, 查准率提升 1.40%, 查全率提升 2.20%, 平均精度均值提升 2.68%, 充分说明文章使用 GAM 思想所提出的 GCSAM 结构可以更进一步聚焦到海滩小目标游客的信息提取上, 使得网络可以更加精确地关注和学习到小目标的特征, 提高模型的检测效果。

将 GCSAM 结构结合 Alpha-CIOU\_Loss 和 BIFPN 所实现的方法进一步提升了单独使用 GCSAM 情况下的模型检测性能, 查准率提升 2.00%, 查全率提升 5.33%, 平均精度均值提升 4.36%。充

分说明 GCSAM 结构的引入所带来的小目标特征图信息相比原始 YOLOv5 更加细致从而使得 BIFPN 可以弥补跨层采样所损失的特征信息, 而更细致的特征图信息包含更精确的定位信息, 促进 Alpha-CIOU\_Loss 进行预测框的回归。

### 2.4 文中算法性能测试及对比试验

为进一步验证改进的 YOLOv5 海滩小目标游客检测算法的有效性, 采用当前主流的小目标检测方法如: SSD 方法<sup>[6]</sup>、CBAM 方法<sup>[27]</sup>、SElayer 方法<sup>[28]</sup>以及原始 YOLOv5 与文中方法进行客观指标比较, 在相同硬件条件下复现其他方法并与文中方法进行对比测试, 实验测试结果如表 2 所示, 实验 PR 曲线如图 6 所示。

表 2 不同方法实验测试对比结果

Tab.2 The results of different methods

算法模型	查准率/ (%)	查全率/ (%)	平均精度 均值/(%)	每秒传输 帧数
SSD	62.56	68.33	65.23	32.01
原始 YOLOv5	89.49	76.70	83.20	41.49
YOLOv5+SElayer	87.28	71.91	80.02	32.78
YOLOv5+CBAM	89.18	75.59	82.34	34.84
文章方法	91.49	82.03	87.56	28.09

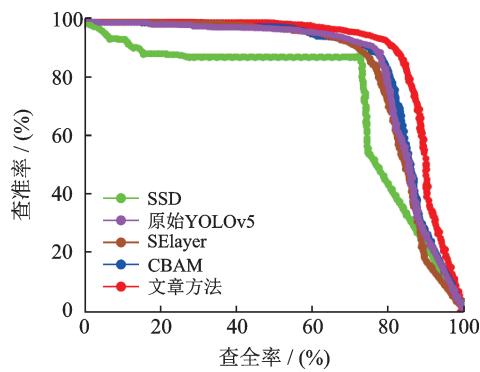


图6 不同方法实验测试PR曲线

Fig.6 Experimental PR curves of different methods

可以看出当训练模型趋于稳定后,在查准率方面,文章算法与SSD方法相比平均提升28.93%,与原始YOLOv5相比平均提升2.00%,与SElayer方法相比平均提升4.21%,与CBAM方法相比平均提升2.31%,说明文章方法在海滩小目标游客检测上可以取得更低的误检率。在查全率方面,文中算法与SSD方法相比平均提升13.70%,与原始YOLOv5相比平均提升5.33%,与SElayer方法相比平均提升10.12%,与CBAM方法相比平均提升6.44%,说明提出的方法在海滩小目标游客检测上可以取得更低的漏检率。在平均精度均值方面,文中算法与SSD方法相比平均提升22.33%,与原始

YOLOv5相比平均提升4.36%,与SElayer方法相比平均提升7.54%,与CBAM方法相比平均提升5.22%,说明文章方法整体网络模型检测性能优秀。为进一步说明文中算法效果,随机抽取一张检测结果可视化进行主观分析,不同网络检测效果如图7所示。

从实验结果图可以看出,实验图像中水中的游客目标实际共有10名,海滩游客特征信息获取难度随着海水对身体的遮挡提高,文中方法相比于其他方法取得5名游客信息检出的最好效果。在图像中间的密集游客区域共有11名游客信息,同样文中方法取得8名游客信息检测的最好结果,原始YOLOv5与SElayer方法都出现大面积的漏检情况,前者只检出3名游客信息而后者只检出4名。SSD方法检测效率严重不足,选取的两个部分只有1名游客信息,其他部分的游客信息检测效果也不及文中方法以及其他方法。所提的方法可以较大程度的检测出绝大多数水中的小目标游客以及密集区域小目标游客,进一步说明了文章方法相对于其他方法可以更加细致地通过部分特征进行预测。与其他方法相比,文章所提的一种改进的YOLOv5海滩小目标游客检测方法在保证实时性的前提下可以取得更精确的小目标游客检测结果,在密集、遮挡、目标尺寸更小的情况下都具有更好的鲁棒性。



图7 不同网络检测结果图

Fig.7 Detection results of different networks

## 2.5 Visdrone数据集对比试验

为进一步证明文章所提方法的有效性,在Visdrone数据集<sup>[22]</sup>上进行实验验证。Visdrone数据集是由天津大学等团队开源的一个大型无人机视角的数据集,基准数据集包括288个视频片段,由261908帧和10209幅静态图像组成,该数据集以小

目标、高密度、遮挡性、背景复杂为特点。在相同实验环境下对比文章所提算法与现有模型实验结果如表3所示。

根据表中数据可以看出:文中算法相对比SSD模型平均精度均值提升19.10%,相比于原始YOLOv5算法平均精度均值提升3.17%,与SElayer方

表3 Visdrone数据集对比实验结果

Tab.3 The results of different methods on the Visdrone dataset

算法模型	平均精度均值/(%)
SSD	15.24
原始YOLOv5	31.17
YOLOv5+SElayer	30.70
YOLOv5+CBAM	33.12
文章方法	34.34

法比较平均精度均值提升3.64%,与CBAM方法相比平均精度均值有着1.22%的提升。从公开数据集的表现可以看出,所提算法在复杂环境下的小目标检测任务中具有更好的检测性能。

### 3 结束语

提出一种基于YOLO的多层特征图信息融合(FMIF-YOLO)的检测方法用于海滩环境下小目标游客检测,有效解决了当前智能监控检测方案无法满足此类复杂场景下小目标游客的检测及其后续准确操作的难题。文中方法可实时检测海滩游客目标,有效保障危险区域中游客的人身安全。通过设计GCSAM结构强化主干网络端提取小目标特征的能力,所得到的特征图更加聚焦于小目标。将原有的PANet结构调整至BIFPN结构进行跨层间的特征图信息交流,使得最后送入预测输出端的特征图包含更多的上下文信息,提高在遮挡条件下的检出率。通过幂变换的特性提升不同检测层预测框,使用更加精确的预测框回归损失函数Alpha-CIOU\_Loss,有效改善一些特定情况下回归预测框精度下降的问题。最后通过实际的海滩环境进行改进检测网络的有效性实验。相比于原始YOLOv5模型,文中方法在实时性满足要求的前提下查准率提升2.00%,查全率提升5.33%,平均精度均值提升4.36%,并且获得比其他主流方法更好的主观视觉评价。

### 参 考 文 献

- [1] 自然资源部海洋战略规划与经济司. 2021年中国海洋经济统计公报[R]. 2022.
- [2] 陆旭,张弛,时健,等. 我国海滩游客安全事故数据库和事故特征分析[J]. 海洋开发与管理, 2021, 38(6): 3-11.
- [3] 谷雨,刘俊,沈宏海,等. 基于改进多尺度分形特征的红外图像弱小目标检测[J]. 光学精密工程, 2020, 28(6): 1375-1386.
- [4] 朱威,王立凯,靳作宝,等. 引入注意力机制的轻量级小目标检测网络[J]. 光学精密工程, 2022, 30(8): 998-1010.
- [5] Huang S Q, Liu Q. Addressing scale imbalance for small object detection with dense detector [J]. Neurocomputing, 2022, 473: 68-78.
- [6] Zhan W, Sun C F, Wang M C, et al. An improved Yolov5 real-time detection method for small objects captured by UAV [J]. Soft Computing, 2022, 26(1): 361-373.
- [7] 李正周,曹雷,邵万兴,等. 基于空时混沌分析的海面小弱目标检测[J]. 光学精密工程, 2018, 26(1): 193-199.
- [8] Lian J, Yin Y H, Wang Z H, et al. Small object detection in traffic scenes based on attention feature fusion [J]. Sensors, 2021, 21(9): 3031.
- [9] 鞠默然,罗海波,刘广琦,等. 采用空间注意力机制的红外弱小目标检测网络[J]. 光学精密工程, 2021, 29(4): 843-853.
- [10] Hu P Y, Ramanan D. Finding tiny faces [C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 1522-1530.
- [11] Kisantal M, Wojna Z, Murawski J, et al. Augmentation for small object detection [C]. 9th International Conference on Advances in Computing and Information Technology, Sydney, Australia, 2019: 119-133.
- [12] Alsubaei F, Al-wesabi F, Hilal A. Deep learning-based small object detection and classification model for garbage waste management in smart cities and IOT environment [J]. Applied Sciences, 2022, 12(5): 2281.
- [13] Lim J, Astrid M, Yoon H, et al. Small object detection using context and attention [C]. 2021 International Conference on Artificial Intelligence in Information and Communication, Jeju Island, Korea, 2021: 181-186.
- [14] Leng J X, Ren Y H, Jiang W, et al. Realize your surroundings: Exploiting context information for small object detection [J]. Neurocomputing, 2021, 433(8): 287-299.
- [15] Chen J Y, Liu S P, Zhao L, et al. Small object detection combining attention mechanism and a novel FPN [J]. Journal of Intelligent & Fuzzy Systems, 2022, 42(4): 3691-3703.
- [16] Liu Y C, Shao Z G, Hoffmann N. Global attention mechanism: Retain information to enhance channel-spatial interactions [R/OL]. <http://arXiv.org/pdf/2112.05561.pdf>: (2021-12-10) [2022-10-28]
- [17] Woo S, Park J, Lee J, et al. CBAM: Convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision, Munich, Germany, 2018: 3-19.
- [18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7132-7141.
- [19] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8759-8768.
- [20] Tan M X, Pang R M, Le Q. Efficient Det: Scalable and efficient object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 10781-10790.
- [21] He J B, Erfani S, Ma X J, et al. Alpha-IOU: A family of power intersection over union losses for bounding box regression [J]. Advances in Neural Information Processing Systems, 2021, 34: 20230-20242.
- [22] Zhu P, Wen L, Du D, et al. Detection and tracking meet drones challenge [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11): 7380-7399.