

文章编号: 2096-1618(2023)01-0037-07

# 基于 YOLOv3 的船只检测算法研究

李 静, 鲜 林, 王海江  
(成都信息工程大学, 四川 成都 610225)

**摘要:**针对长江流域错综复杂的生态环境以及执法部门人员短缺对长江 10 年禁渔令实施的限制情况;通过智能视频监控系统对长江流域过往船只目标检测,对于判别船只有无违法捕捞行为具有重要意义。当前,传统的目标检测算法早已被检测效率更高、算法复杂度更低的深度学习方法替代;基于智能视频监控对于实时性的要求,采用 YOLOv3 作为目标检测模型,在兼顾检测精度的同时检测速度也更高。YOLOv3 算法中,先验框作为目标检测算法的重要机制,影响着预测框的定位性能。在  $K$ -means 聚类算法上进行改进,通过改变  $K$  值初始化随机选择不能获取全局最优解的情况,对  $K$  值选择时应用轮盘法,选择距离已经形成的聚类中心尽可能远的值作为新的  $K$  值,使各个聚类中心相对距离尽可能大,从而尽可能获得全局最优的聚类结果。实验结果表明, $K$ -means 优化后获得的先验框训练模型让船只目标检测性能更加优异,整体 mAP 提升了 9.31%。

**关 键 词:**目标检测;深度学习;先验框; $K$ -means 聚类

**中图分类号:**TP391.4

**文献标志码:**A

**doi:**10.16836/j.cnki.jcuit.2023.01.006

## 0 引言

近年来长江流域在粗放经济的发展模式下,长江生态环境遭到严重破坏,针对这一现状,农业农村部宣布从 2020 年起实施为期 10 年的长江渔业“禁渔令”。然而禁渔令颁布至今,仍有许多人在利益的驱使下,在长江流域进行违法捕捞,致使长江的生态环境不仅没有得到改善,反而日趋恶化。由于违法捕捞由明转暗,担负长江流域监管责任的渔政部门人员不足、装备落后等,导致渔政执法更加困难,难以及时阻止和抓捕违法捕捞人员。长江流域禁渔令不能沦为一纸空文,为提高渔政执法的效率,减轻执法人员的工作量,提出利用数字图像处理和视频图像分析的方法,自动检测视频场景中运动目标,并判断目标行为,对异常情况进行预警,从而指导和规划相应的处置行动。对于错综复杂的内河流域环境,以及各类渔船、货船、客船多个船只种类,准确对船只进行检测分类,并对异常船只进行监控,对于监控违法捕捞行为具有重要作用。

传统意义上的目标检测算法建立在感兴趣区域选择的基础上,将整个输入样本图片作为扫描范围,利用滑动窗口的思路遍历整个图片,但存在算法时间复杂度较高的缺点;现实背景中目标是形态多样的,难以提取到鲁棒性较好的目标。大批研究学者为解决这一问题进行了大量研究。安建成<sup>[1]</sup>通过融合 LBP(local binary pattern)算子和 Canny 边缘实现对图像分割,提取图像纹理特征;彭棉珠<sup>[2]</sup>利用卷积网络融合尺度不变特征算法 SIFT,利用该算法搜索空间尺度极值点和提取相关位

置、尺度和旋转不变量等信息的特性实现图像自动标注研究;2001 年 Paul 等在人脸检测领域首次提出基于 Haar<sup>[3]</sup>特征并调用 Adaboost 算法<sup>[4]</sup>、“瀑布”级联分类器检测方法,检测速度得到显著提升;Felzenszwalb 等<sup>[5]</sup>提出 DPM(deformable part based model)将目标检测问题拆分成输入样本中各个小目标的检测问题,真正解决了三维目标对象在不同角度下的实时检测问题,也将早期经典特征检测算法推向了发展的顶峰阶段。

鉴于传统的目标检测算法冗余、鲁棒性差,而深度卷积网络具有优秀的鲁棒性能以及有效的特征表述能力,使目标检测速度得到大幅提升。胡葵等<sup>[6]</sup>将深度卷积网络 CNN 与目标检测相融合,实现了多尺度行人的实时检测。随着卷积神经网络迅速发展,应用范围也逐渐扩展,基于区域卷积网络的 R-CNN(regions with CNN features)<sup>[7-10]</sup>被提出,在输入样本中提取目标候选框,对所有的候选框采用数据归一化到同一尺寸,将特征向量作为 CNN 网络的输入,提取特征向量并传入分类器 SVM 进行目标分类。虽然 R-CNN 相比传统的目标检测有很大的进步,但是其本身训练分多个阶段,存在耗时等缺陷;Ding H 等<sup>[11]</sup>,Herk 等<sup>[12]</sup>,焦李成<sup>[13]</sup>在 SPP Net 模型提出后,对其进行了深入的研究和应用。其与 CNN 模型的区别主要在卷积网络的结构上,CNN 模型中的全连接层需要保证输入的尺寸,而 SPP Net 卷积层后的输出特征向量经过 SPP 层可以实现任意大小的区域特征提取,优化了候选区域计算冗余的问题,在保留 R-CNN 精度的情况下提升了检测的速度。

基于前两者的研究成果,2015 年相继提出了两个深度学习检测算法 Fast-RCNN 和 Faster-RCNN<sup>[14-15]</sup>。

前者的特点在于实现了多任务的学习方式,不需要分多阶段训练,将目标分类和 B Box 回归问题同阶段进行训练,实现了目标检测的快速性、实时性,但是仍然建立在预先提取候选框基础上;而后者真正地实现了“端到端”的目标检测,最大的创新在于候选区域生成网络 RPN 的提出,将原先依赖外部算法提取目标候选框的过程融入了整个检测网络框架中,同时检测精度和速度也得到了很大的提升。

随着目标检测算法的不断发展,研究出一体化卷积网络来进行目标检测,即通过单次的检测可以得到最终的检测结果,可以直观地得到输入输出映射关系。YOLO<sup>[16-17]</sup>(you only look once)算法作为目标检测主线中第一个单阶段检测算法,相较于基于目标候选框的两阶段检测算法,最大的提升在于速度性能指标上,但其精度相对下降;除了 YOLO 算法,单阶段目标检测算法还有 SSD<sup>[18-19]</sup>算法,在保留 YOLO 算法速度性能优势的情况下,在不同尺度的特征图进行多尺度的检测与边界框回归,使检测精度的 mAP 与 Faster-RCNN 同一水平。

1 YOLOv3 目标检测算法及改进

1.1 YOLOv3 算法

YOLO 算法是 2016 年由 Joseph Redmon 等提出的一种单阶段目标检测网络,检测速度能达到 45 帧每秒,

能够轻松实现实时目标检测。它的核心思想是利用回归方式实现目标检测,将图片作为网络输入,通过一个卷积网络,就得到目标预测框的位置以及类别。随着对算法的不断研究,YOLOv2 采用 darknet-19 作为特征提取网络,引入先验框机制,在速度和精度方面都得到了提升,弥补了 YOLOv1 召回率和定位精度低的情况。

YOLOv3 在 YOLOv2 基础上进行了大量修改,在网络结构上,舍去神经网络深度较低的 darknet-19,采用网络层次更深的 darknet-53 作为特征提取的主干网络,对更高层的图像信息更加敏感。YOLOv3 采用全卷积网络,借助残差网络的思想,通过跳层连接的方法,让深层次的图像特征与低层次图像特征结合,提升图像特征提取精度;YOLOv3 引入特征金字塔网络中的多尺度融合,得到 3 个不同尺度特征图对应不同大小的先验框,将高层次的特征信息经上采样堆叠到低层次特征上,有效加强了低层次的语义信息,提高了对不同尺度的目标检测能力。特征融合过程:首先网络经过 5 次降采样后分别得到 13×13、26×26、52×52 共 3 个有效特征图,将 13×13 的特征图 3 经过上采样后与上一层采样结果进行特征融合得到 26×26 的特征图 2;同理,将特征图 2 经过上采样与特征图 1 融合后得到 52×52 的特征图。具体网络结构如图 1 所示。

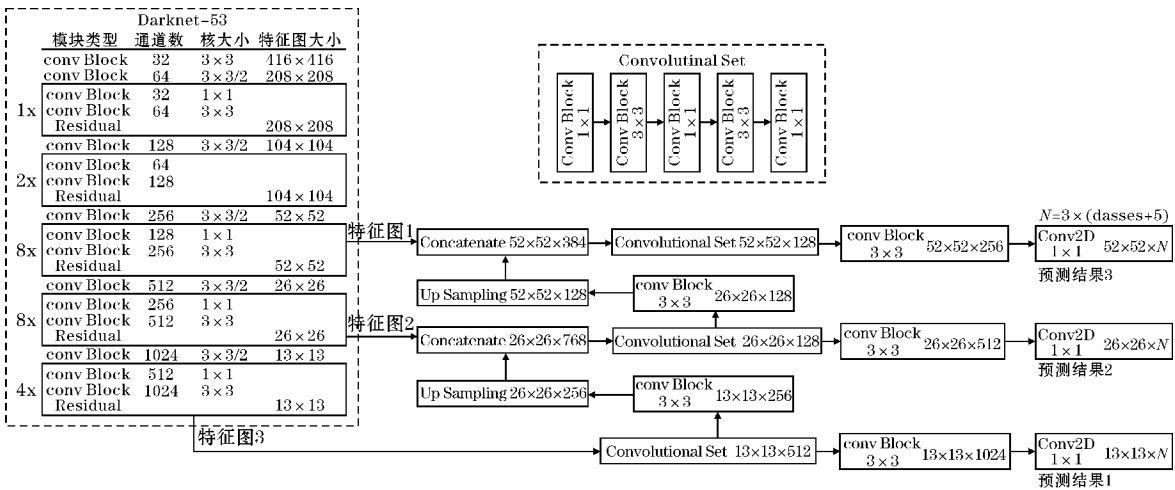


图 1 YOLOv3 网络结构

1.2 改进的 YOLOv3 算法

先验框是 YOLOv3 目标检测网络的一个重要组成部分,原来 YOLO 网络是直接对目标的宽高进行预测,得到的检测结果定位性能很差,直到 YOLOv2 借鉴 Faster R-CNN 的先验框机制来利用回归方式预测位置,其定位效果才有所提高。先验框能给目标检测网络提供一定先验信息让网络更快速地确定目标位置,

同时它也影响着模型训练过程中损失函数的计算,在目标检测网络中起着至关重要的作用。

YOLOv3 采用 3 种不同尺度的特征图来预测目标,每个尺度对应 3 种不同的先验框,随着特征图的数量和尺度的变化,先验框的尺寸也需要相应的调整。YOLOv3 采用 K-means 聚类方法,在 COCO 数据上聚类出 9 个先验框,如表 1 所示。



表 1 K-means 聚类的先验框

特征图大小	先验框大小	预测目标
13×13	(116×90); (156×198); (373×326)	大
26×26	(30×61); (62×45); (59×119)	中
52×52	(10×13); (16×30); (33×23)	小

COCO 数据集上包含的 80 类目标,各类目标大小差距过大,如果将该数据集生成的先验框应用到内核船只数据集上,其中部分先验框大小就不太合理了。

K-means 算法核心是随机选择初始聚类中心,计算出样本与 K 个聚类中心的距离,样本属于最近的聚类中心的所属类别。具体实现步骤:(1)选定聚类的个数 K;(2)计算每个样本到各个聚类中心的距离,将其归类于距离最近的样本中心;(3)每个聚类中心被分配到新的样本都要重新均值算出新的聚类中心;(4)判断新的聚类中心是否与前一个聚类中心相同或者接近;(5)若相同或者接近,则算法终止,否则更新中心点,循环(2)~(4)步,直到算法终止。

K-means 聚类随机初始化取值导致聚类效果严重依赖聚类中心的随机初始化,从而结果可能会趋向于局部最优化而非全局最优化,导致对小目标的漏检率和识别准确率下降。在河道运行场景下,远离镜头的船只属于小目标,为提高对该类目标物的检出率,采取对 K-mean++算法对 K-means 进行优化。K-means++算法的核心是选取聚类中心时,每个聚类中心的相对距离应该尽可能得远,保证每个聚类中心不会在同一个簇内进行迭代,尽可能获得全局最优的聚类效果,其具体步骤如下所示。

K-means++算法步骤:

- (1)从获取的所有 bounding box 的宽、高数据中,随机选择一个 anchor box 作为聚类中心  $O_1$ ;
- (2)计算数据集中每个 bounding box 与已经初始化的聚类中心的 IOU 值,将样本框  $X_n$  到聚类中心  $O_j$  之间的距离记作  $d=(X_n,O_j)$ ;传统的聚类方法使用欧氏距离来衡量差异,但是应用到比较面积较大的 boun-

ding box 时,误差比较大,所以引入了 IOU 值,用样本框和先验框的交并比代替原来的欧氏距离作为目标函数,由于 IOU 值往往越大越好,但是目标函数值越小,聚类结果越好,所以其公式为

$$D=\min_{\text{box}=0} \sum_{\text{cen}=0}^n \sum_k^k [1-\text{IOU}_{\text{cen}}^{\text{box}}]$$

(1)

式中 box 为样本标签的目标框,cen 为聚类中心,n 为样本数,k 为类别数。

(3)从样本集中选取下一个 bounding box 作为聚类中心的概率为

$$P=\frac{d^2(x_n,0)}{\sum_{j=1}^m d^2(x_j,0)}$$

(2)

计算每一个样本框到最近的聚类中心的距离  $D(x)$ 。

(4)按照轮盘法则,选取  $D(x)$  最大的样本框作为下一个聚类中心。

(5)重复步骤(3)、(4),直到选择出 K 个聚类中心。

针对内河船只数据集,运用 K-means++算法得到 9 个先验框尺寸为如表 2 所示。

表 2 K-means++聚类的先验框

特征图大小	先验框大小	预测目标
13×13	(324×70); (77×1007); (189×1020)	大
25×26	(106×25); (154×35); (222×46)	中
52×52	(50×13); (77×20); (120×14)	小

2 数据处理及模型训练

2.1 实验数据集

本文数据集来源于重庆嘉陵江岸设置的海康威视 DS-2CD3T86FWDV2-I5S 高清摄像机录制的江面船只视频,通过截取视频帧中货船、渔船以及客船图像,以及通过爬虫程序从互联网上下载的内河船只图片制作成共 7859 个内河船只数据集(图 2)。

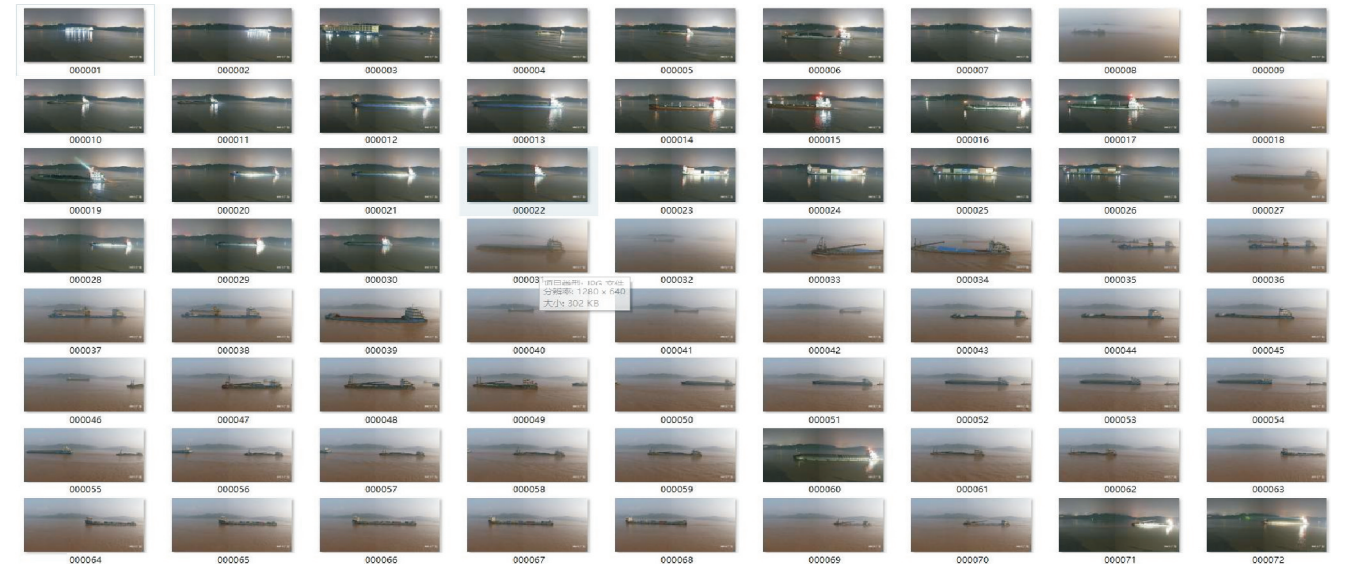


图 2 内河船只数据集

## 2.2 数据集处理

由于摄像机为广角摄像,视频截取的船只图片分辨率为  $5120 \times 1400$ ,为便于后续对图片进行特征提取,首先对图片进行预处理,将图像分辨率降到  $1980 \times 1080$ ,然后筛选并删除图像不清晰、船身不完整等数据集,再通过 LabelImg 标记软件对数据集进行船只目标框标记,标注船只图片中的位置信息以及类别信息,用于后续的模型训练。深度学习所使用的数据格式大部分采用 VOC 格式,将存储船只位置信息的 .xml 文件保存在 annotation 文件中,所有的数据集 .jpg 图片存入 JPEGImages 文件中。数据集处理流程如图 3 所示,数据集 ground truth 标记如图 4 所示。

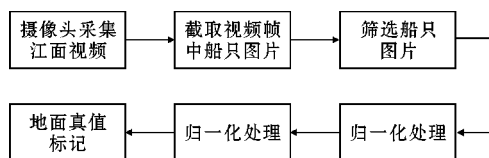


图3 数据集处理流程

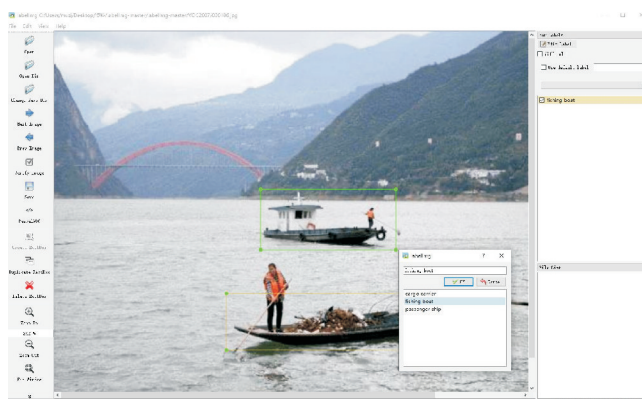


图4 数据集 ground truth 标记

## 2.3 模型训练

本文在 tensorflow-gpu1.14 版本的深度学习框架下采用 YOLOv3 网络实现内河船只检测,使用 python3.6 编译器,操作系统为 windows10,CPU 型号为 intel i9-10920x, GPU 型号为 NVIDIA GeForce RTX 2060,配置 CUDA 11.4.136 对 GPU 进行加速。模型训练分为两个阶段,第一阶段设置 batch\_size=4,动量=0.9,学习率=0.001,优化算法采用 Adam 算法,共训练 60 个 epochs;第二阶段使用第一阶段训练生成的权重文件对数据集二次训练,将 batch\_size 改为 3,训练 60 个 epoches,直至损失函数区域稳定,得到最后模型.h5 权重文件。YOLOv3 在内河船只数据集上模型训练时间大约花费 15 h,共 120 个 epochs。

## 3 实验结果及分析

### 3.1 评估标准

在目标检测中,一张图片可能包含多个不同类别的目标,而且还要得到每个目标的位置信息,图像分类中的精度指标不适用于模型的评测。本文采用平均精度度 (AP),引入了精准率 (precision) 和召回率 (recall),以及 COCO 指标中的 mAP (mean average precision) 值来评估模型的优劣。

$$\text{precision} = \frac{TP}{TP+FP} = \frac{TP}{\text{all detections}} \quad (3)$$

$$\text{recall} = \frac{TP}{TP+FN} = \frac{TP}{\text{all ground truths}} \quad (4)$$

式中,TP 表示检测器结果中被正确分类的个数,FP 表示检测器结果中被错误分类的个数,FN 表示 ground truth 中未被找到的个数。精准率表示模型给出的所有预测结果中正确预测样本所占的比例,召回率表示预测结果中预测正确样本数量占所有样本数量的比例。通过精准率和召回率可以绘出 P-R 曲线,P-R 曲线与坐标轴围成的面积就是单个目标类别的 AP 值,一般 AP 值越大表示模型对该类别预测结果越好。

$$AP = \int_0^1 P_{\text{smooth}}(r) dr \quad (5)$$

式中, $P_{\text{smooth}}(r)$  表示做平滑处理后的 P-R 曲线。由于内河河面出现的船只类别较多,如果只用 AP 值来逐个比较,难以比较模型的好坏,为更好地评价检测器在各个船只类别上的性能,采用 COCO 数据集中采用类别求平均的方法:先确定一个 IOU 阈值,在这个阈值下分别求船只各类别的 AP,将所有船只类别的 AP 值取平均得到 mAP,以此来衡量整个检测算法的性能。

$$mAP = \frac{\sum_0^k AP}{k} \quad (6)$$

式中, $k$  表示船只类别个数。

### 3.2 实验结果对比分析

对内河船只数据集分别使用 K-means 聚类的 YOLOv3 模型,以及优化的 k-means++ 聚类的 YOLOv3 模型进行对比试验。分别使用上述两种算法生成的先验框训练得到的权重文件,对同一批测试图片进行检测,得到的各船只类别 AP 值,P-R 曲线以及 mAP 值如表 3 和图 5 所示。

表 3 各船只类别实验结果参数对比 单位: %

算法模型	Precision 值精准率			Recall 值召回率		
	货船	渔船	客船	货船	渔船	客船
YOLOv3-K-means	81.70	63.10	80.00	95.42	79.20	92.31
YOLOv3-K-means++	91.04	91.50	96.23	93.13	93.96	98.08

由表 3 可以看出,优化后模型在精准率上货船的精准率提高了9.7%,渔船提高了28.4%,客船提高了

16.23%;在召回率上优化后模型在渔船和客船的召回率都明显优于未优化的模型结果,其中渔船召回率提高了14.76%,客船召回率提高了5.77%。根据数据分析可以看出,优化后的模型在目标较小的客船和渔船的召回率和精准率上都有较大幅度的提升,平均精确率提高了18.92%,平均召回率提高了6.52%,验证了 K-means++算法的优越性。

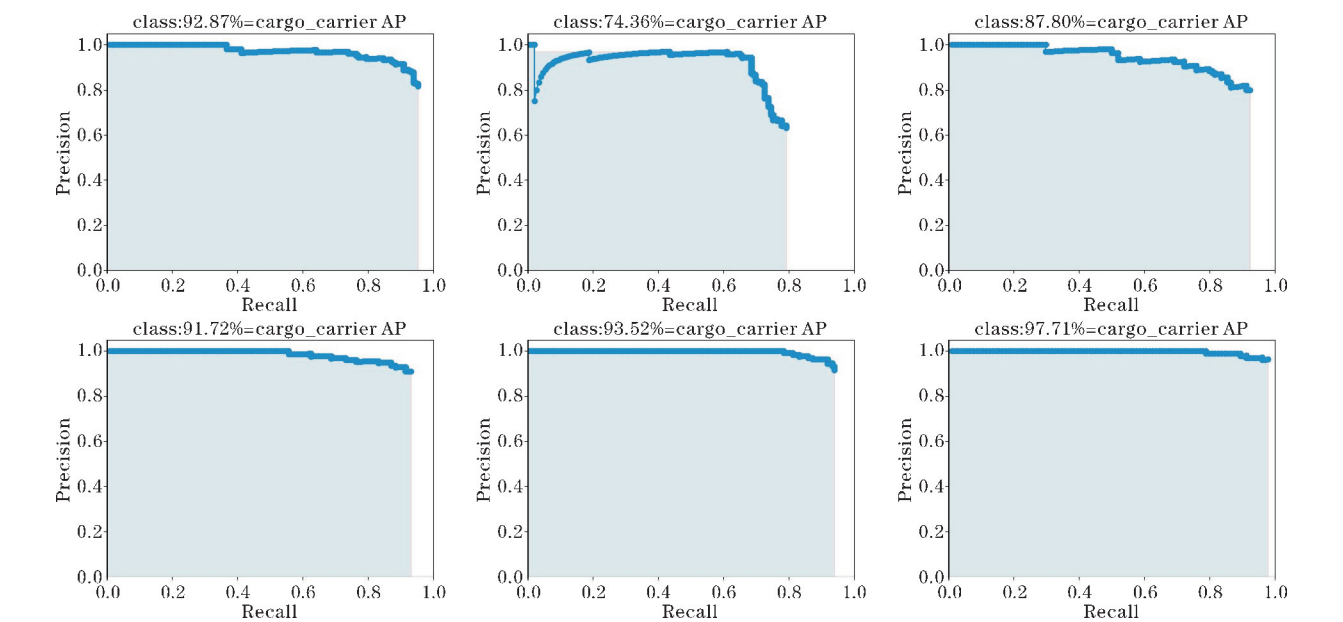


图 5 K-means 聚类各类船只 P-R 曲线及优化后 K-means 聚类各类船只 P-R 曲线

图 5 为两种模型行的各船只类别的 P-R 曲线对比,P-R 曲线下和坐标轴之间的面积表示各类船只的 AP 值,第二列、第三列为未优化模型和优化后模型对渔船和客船检测后绘出的 P-R 曲线图,可以看出优化后模型召回率明显高于未优化模型,而且后者的 P-R 曲线能够包含前者的 P-R 曲线,并且优化后模型渔船的 AP 值相较于未优化的模型提升了19.16%,客船 AP 值提升了9.91%,有效增强了检测渔船、客船等较

小目标物体的检测率。表明优化后模型相较于前者更加优异。

K-means 聚类的 YOLOv3 模型 mAP 值为85.01%,优化后 K-means 后 K-means++ 聚类的 YOLOv3 模型 mAP 值为94.32% (见图 6)。相较于 K-means 聚类模型效果,优化后模型效果提升了9.31%。由此得出优化后的算法效果在内河船只数据集上效果更优秀。

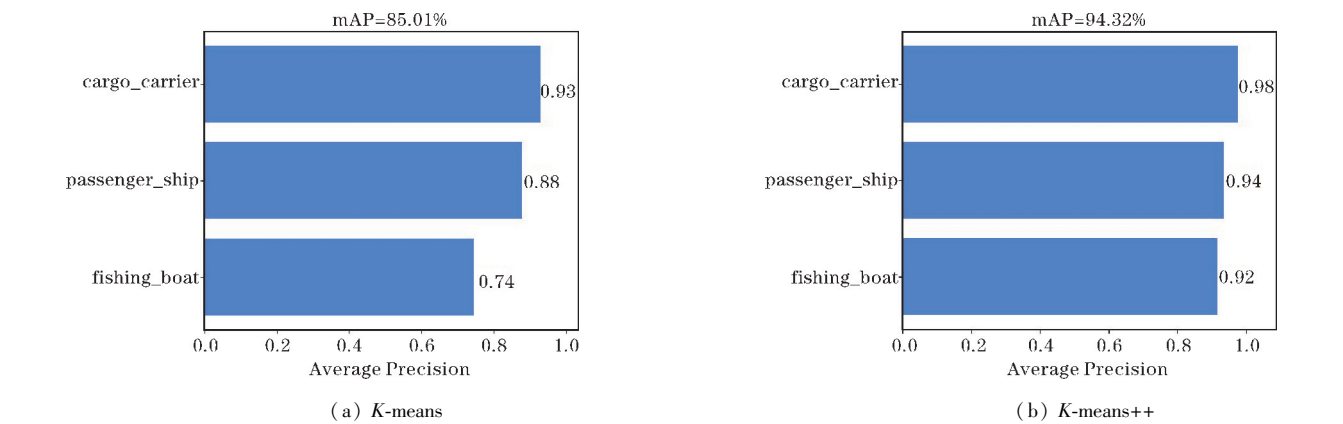


图 6 K-means 和 K-means++聚类的 YOLOv3 模型 mAP 值



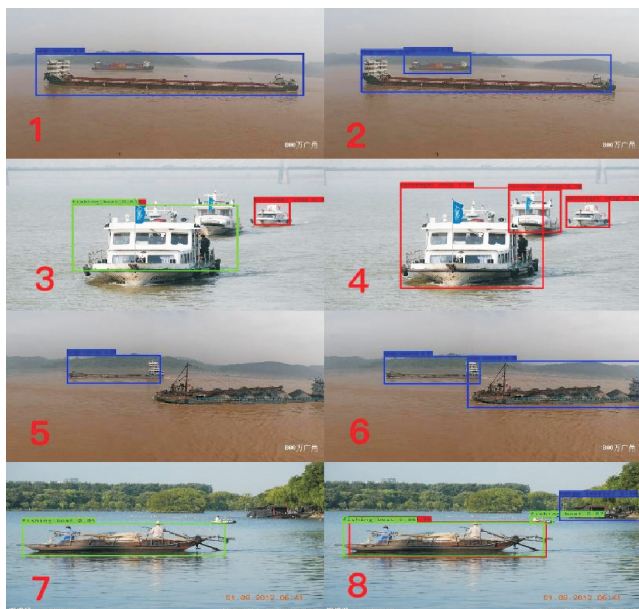


图7 两种算法模型船只检测效果对比

图7中1、3、5、7为K-means聚类模型检测结果；2、4、6、8为K-means++聚类模型聚类效果。由1、2对比和5、6对比可知，优化前的模型对于船只的定位和捕捉能力较低，很明显的目标也有被漏检的情况发生，优化后模型在召回率和定位准确率上有了明显的提升；7和8的对比，可以明显看到优化前湖面左上方的渔船完全没被识别定位，在优化后的模型中，不仅小渔船被识别出来，它后面的船只也被定位成功，但是被错误分类为货船，由此得出，优化后模型对于小目标的召回率高，定位准确率高，但是在目标识别准确率上仍有待提高。由3、4对比可知，优化前模型将第一只客船错误识别为渔船，第二只船只未被识别出来，在优化后模型中，被略微遮挡的3只客船全部被识别成功，且被准确定位。总体来说，K-means++优化后模型对于目标检测的召回率有了明显提升，漏检率降低，识别准确率相对于优化前有略微提升，但仍有识别错误的情况发生，这是未来模型优化的方向之一。

## 4 结束语

针对内河流域船只目标检测问题进行研究，在YOLOv3目标检测网络的基础上，建立内河船只数据集，针对K-means聚类算法无法得到全局最优化的结果，对K-means算法进行优化，让初始化的聚类中心尽可能的远。采用两种算法模型对内河船只数据集进行聚类，实验结果表明优化后的YOLOv3算法模型，能够更好地检测船只目标，有效解决了船只目标精准率和召回率较低的问题，优化后模型的mAP值提高了

9.31%，体现了优化后模型的优越性。

## 参考文献：

- [1] 安建成. 融合模糊LBP和Canny边缘的图像分割[J]. 计算机工程与设计, 2019, 40(12): 3533-3537.
- [2] 彭棉珠. 卷积网络和SIFT特征融合的图像自动标注研究[J]. 福建电脑, 2021, 37(10): 12-16.
- [3] 贺瑜飞. 基于Haar特征和改进的AdaBoost算法的人脸图像识别[J]. 榆林学院学报, 2019, 29(6): 69-70.
- [4] 兰胜坤. 基于Adaboost算法的人脸检测实现[J]. 电脑与信息技术, 2021, 29(2): 16-19.
- [5] Felzenszwalb P F, Mcallester D A, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]. 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008.
- [6] 胡葵, 章东平, 杨力. 卷积神经网络的多尺度行人检测[J]. 中国计量大学学报, 2017, 28(4): 472-477.
- [7] Wen junyu, Sumi Kim, Jeong-Hyu Lee, et al. Animal Detection in Highly Cluttered Natural Scenes by using Faster R-CNN[J]. International Journal of Recent Technology and Engineering (IJRTE), 2019, 8.
- [8] Nirmala, S Arivalagan, R Arunkumar. An Efficient and Robust Multi-Object Recognition and Tracking Algorithm using Mask Region based Convolution Neural Network (R-CNN) [J]. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019, 8(9).
- [9] K G Shreyas Dixit, Mahima Girish Chadaga, Sinchana S Savalgimath, et al. Evaluation and Evolution of Object Detection Techniques YOLO and R-CNN[J]. International Journal of Recent Technology and Engineering (IJRTE), 2019, 8.
- [10] 龚强. 基于Mask R-CNN的无人驾驶汽车道路前方目标检测的研究[D]. 南昌: 南昌大学, 2020.
- [11] Ding H, Tian Y, Peng C, et al. Inference attacks on genomic privacy with an improved HMM and an RCNN model for unrelated individuals[J]. Information Sciences, 2020, 512: 207-218.
- [12] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual

- Recognition[C]. IEEE Transactions on Pattern Analysis & Machine Intelligence. 2014:1904–1916.
- [13] 焦李成. 基于全局—局部 SPP Net 的高分辨率 SAR 图像变化检测方法[D]. 西安:西安电子科技大学,2019.
- [14] Zeghiche Oussama. 基于深度 Faster-RCNN 的车牌识别算法研究[D]. 西安:西安电子科技大学,2020.
- [15] 于晓倩. 基于改进 Faster RCNN 的行人检测研究[D]. 长春:吉林大学,2020.
- [16] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection [C]. Computer Vision & Pattern Recognition, 2016.
- [17] Shafiee M J, Chywl B, Li F, et al. Fast YOLO: A Fast You Only Look Once System for Real-time Embedded Object Detection in Video[J]. Journal of Computational Vision and Imaging Systems, 2017, 3(1).
- [18] 王婷. 基于 SSD 的航拍图像中绝缘子识别与定位研究[D]. 北京:华北电力大学,2021.
- [19] 姜敏, 王力, 王冬冬. 改进的 SSD 行人检测算法[J]. 软件, 2020, 41(2): 57–61.
- [16] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection

## Research on Ship Detection Algorithm based on YOLOv3

LI Jing, XIAN Lin, WANG Haijiang

(Chengdu University of Information Technology, Chengdu 610225, China)

**Abstract:** In view of the intricate ecological environment of the Yangtze River Basin and the limitation of the shortage of law enforcement personnel on the implementation of the ten-year fishing ban in the Yangtze River; in this paper, the intelligent video surveillance system is used to detect the target of passing ships in the Yangtze River Basin, which is of great significance for judging whether there is illegal fishing behavior. At present, traditional target detection algorithms have long been replaced by deep learning methods with higher detection efficiency and lower algorithm complexity. This paper considers the real-time requirements of intelligent video surveillance, and adopts the YOLOv3 algorithm under comprehensive consideration, which takes into account the detection accuracy and the detection speed is also higher. In the YOLOv3 algorithm, the prior frame is an important mechanism of the target detection algorithm, which affects the positioning performance of the prediction frame. In this paper, the  $K$ -means clustering algorithm is improved. By changing the  $K$  value initialization, the random selection the global cannot obtain the optimal solution, the roulette method is applied to the selection of the  $K$  value, and the value that is as far as possible from the cluster center that has been formed is selected as the new  $K$  value, so that the relative distance of each cluster center is as large as possible, so as to obtain the globally optimal cluster as much as possible. The experimental results show that the  $K$ -means optimized prior frame training model makes the ship target detection performance more excellent, and the overall mAP is increased by 9.31%.

**Keywords:** target detection; deep learning; a priori box;  $K$ -means