

基于 SPMYOLOv3 的水面垃圾目标检测^①



王一早^{1,2}, 马纪颖^{1,2}, 罗 星^{1,2}, 王书哲^{1,2}

¹(沈阳化工大学 计算机科学与技术学院, 沈阳 110142)

²(辽宁省化工过程工业智能化技术重点实验室, 沈阳 110142)

通信作者: 马纪颖, E-mail: 83001421@qq.com

摘 要: 为解决水面垃圾检测中存在目标形状尺度差异大, 难以区分背景以及目标偏小的问题, 本文提出了一种 SPMYOLOv3 目标检测算法来实现对水面垃圾的检测. 首先, 对收集到的水面垃圾数据集进行标注, 使用改进的 K-means 算法对数据集重新聚类, 得到与数据集更匹配的先验框. 其次, 在 YOLOv3 的主干网络后添加 SE-PPM 模块, 加强目标的特征信息, 保证目标尺度不变且保留全局信息. 再使用多向金字塔网络对不同尺度的特征图进行融合, 获得携带更加丰富的上下文信息的特征图. 最后使用在损失函数中使用 focal loss 计算负样本的置信度损失, 抑制了 YOLOv3 中正负样本不均衡问题. 改进后的算法在水面垃圾数据集上的实验结果表明, 相比于原 YOLOv3 算法检测精度提升了 3.96%.

关键词: 水面垃圾检测; YOLOv3 模型; 特征融合; SE-PPM 网络; focal loss

引用格式: 王一早, 马纪颖, 罗星, 王书哲. 基于 SPMYOLOv3 的水面垃圾目标检测. 计算机系统应用, 2023, 32(3): 163-170. <http://www.c-s-a.org.cn/1003-3254/9001.html>

Target Detection of Water Surface Garbage Based on SPMYOLOv3

WANG Yi-Zao^{1,2}, MA Ji-Ying^{1,2}, LUO Xing^{1,2}, WANG Shu-Zhe^{1,2}

¹(College of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang 110142, China)

²(Liaoning Key Laboratory of Intelligent Technology for Chemical Process Industry, Shenyang 110142, China)

Abstract: In water surface garbage detection, large differences occur in target shape and scale, and it is difficult to distinguish the background and the small target. Thus, this study proposes the SPMYOLOv3 detection algorithm to identify surface garbage. Firstly, massive surface garbage datasets are collected and annotated, and an improved K-means clustering method is applied to generate the priori boxes that better match the datasets. Secondly, the SE-PPM module is added after the backbone network of YOLOv3 for strengthening the feature information of the target, ensuring that the target scale remains unchanged and the global information is preserved. The multidirectional FPN is then applied to fuse the feature maps of different scales so that the feature maps after fusion contain richer context information. Finally, the Focal Loss is adopted to compute the confidence loss of negative samples, which alleviates the imbalance of positive and negative samples in YOLOv3. The modified algorithm is tested on the water surface garbage dataset, and the results show that the accuracy of the modified algorithm is 3.96% higher than that of the original YOLOv3 algorithm.

Key words: water surface garbage detection; YOLOv3 model; feature fusion; SE-PPM network; focal loss

随着社会经济的发展和旅游业的兴旺, 由于一小部分人的环境保护意识薄弱, 导致景区内水体质量每

况愈下, 水面垃圾的清理成为景区水体保护的新课题. 目前景区内保护水面环境主要通过人工驾驶小船的方

① 收稿时间: 2022-08-15; 修改时间: 2022-09-15; 采用时间: 2022-09-27; csa 在线出版时间: 2022-12-06

CNKI 网络首发时间: 2022-12-07

式清除水面垃圾.若安排人员定时清除,垃圾不多时浪费人力物力,人口高峰时垃圾增加过快,清除时已造成严重的水体污染.因此景区内水面垃圾的清除时间安排成为难题.随着计算机技术的不断发展,机器视觉成为水体监测的一个主要发展方向,使用机器视觉技术可对水面垃圾量进行有效的监测.因此,基于机器视觉的水面垃圾监测能有效反映水面的环境质量,能够在保持水面环境的前提下尽量减少清理次数从而节省相应的人力物力.同时,水面垃圾目标检测算法也是实现水面垃圾清除机器人智能化的关键核心技术.

目前,对于水面目标检测主要分为传统水面垃圾目标检测算法^[1]和基于深度学习的水面垃圾检测算法^[2].传统的水面垃圾检测算法主要是通过统计目标物体图像的颜色、纹理和边缘等信息建立模板,通过统计的模板使用相应的分类器进行类别识别.王斌^[3]先使用局部峰值奇异性进行图像预处理,再采用基于亮度区域划分的目标检测机制识别水面物体,但受到水面环境复杂多变和水面目标众多的干扰,导致算法识别率低下.方晶等人^[4]在视觉注意模型的基础上,通过小波变换,高斯金字塔和归一化算子等方法融合各类特征,提升了算法的抗干扰能力,但水面气候复杂、目标众多会影响识别精度.魏建荣^[5]结合图像的背景纹理和噪声优化算法,有效的提升水面目标的检测效率,但该方法易受水面建筑物和岸体轮廓的干扰,影响水面垃圾识别.汤伟等人^[6]通过结合 Meanshift 算法和基于色差灰度模型改进的最大类间方差法,有效抑制水面波纹,光照不均和水面倒影问题,在精度有所提升,但实时性不高,不利于移动硬件部署.总体而言,传统水面垃圾检测算法普遍存在检测速度不快和精度不高以及鲁棒性差等问题.

随着计算能力的提高,图形处理单元处理器的发展以及大数据技术的成熟,深度学习技术在目标检测任务中取得了相当大的性能提升.目前,基于深度学习的目标检测算法有两个解决方向:第1种是基于特征生成特征候选区域再执行分类和回归任务的目标检测算法^[7-13];第2种是基于回归方法在图像上直接划分候选区的目标检测算法^[13-18].基于特征生成候选区的主流算法有区域卷积神经网络(R-CNN)^[7],快速区域卷积神经网络(Fast R-CNN)^[8]和基于 Region 的全卷积神经网络(R-FCN)^[9]等.基于回归的算法主要的代表有 YOLO 系列^[10], SSD^[11], RetinaNet^[12]等.很多算法都是

基于主流算法的改进.如李佳琪等人^[13]基于注意力机制,生成对抗网络与 Mask R-CNN^[14]结合,解决遥感影响检测中的问题.王志鹏等人^[15]提出一种结合帧差法思想的 Faster R-CNN 算法^[16]实现对施工的智能化监管.王林等人^[17]通过在 YOLOv3 主干网络后添加空间金字塔池化,提升了对火焰和烟雾的检测精度.孙备等人^[18]基于多尺度卷积融合和空间注意力改进 SSD 检测算法,对水面微小目标的检测精度有很大提升.总体而言,深度学习的算法比起传统算法有很大的进步,然而针对水面垃圾数据集目标与背景差异颜色小,形状尺度多变以及小目标检测精度不高的问题深度学习的方法还存在很大挑战.

本文针对水面垃圾数据集面临的挑战,基于 YOLOv3 算法,提出一种结合注意力机制的多尺度池化模块和融合多层特征图丰富边缘信息的多向特征金字塔网络分别解决水面垃圾检测中的目标与背景颜色差异小,形状差异大和小目标检测问题.同时改进 K-means 聚类算法,使聚类得到的先验框更匹配数据集,加快算法损失收敛速度,提高算法检测精度.最后为缓解 YOLOv3 中正负样本失衡问题,使用 focal loss 替换计算负样本置信度的交叉熵.实验表明,本文基于 YOLOv3 的改进能够大大改善水面垃圾检测的精度.

1 改进的 YOLOv3 算法

本文基于 YOLOv3 的目标检测框架,添加了注意力多尺度池化模块和多向金字塔网络,来解决水面垃圾目标中的目标与背景差异小,形状尺度多变和小目标问题,在提升水面垃圾检测准确度的同时,不损失太多 YOLOv3 的检测速度.检测框架如图 1 所示.

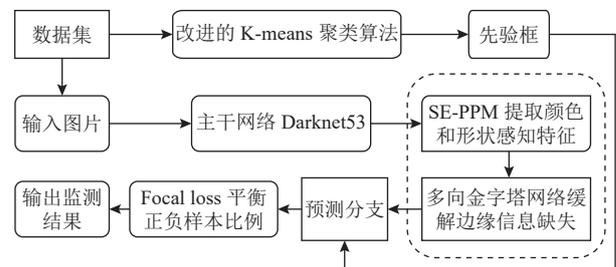


图 1 基于 YOLOv3 的水面垃圾检测框架整体流程

该框架的具体流程如下.

1) 通过数据集使用改进的 K-means 算法聚类得到先验框.

- 2) 输入待检测的图像.
- 3) 使用主干网络 Darknet53 提取卷积的特征, 即主干特征.
- 4) 把主干特征输入 SE-PPM 模块加强目标颜色和形状特征.
- 5) 使用多向金字塔网络缓解加强后的特征图在上采样过程中边缘信息缺失.
- 6) 在预测分支中通过聚类得到的先验框对目标进行预测.
- 7) 使用 focal loss 替换交叉熵计算负样本的置信度, 并进行梯度更新.
- 8) 输出当前图像目标检测结果, 即该图像中所有目标的分类准确度和定位框坐标信息.

1.1 改进的 K-means 聚类方法

因为本文中水面垃圾数据集与 CoCo 等公开数据集有很大差异, 需要使用聚类算法重新聚类先验框. YOLOv3 算法使用 K-means^[19] 聚类生成先验框, 其聚类的衡量指标为 $d=1-IoU$, 使用 d 作为衡量指标会造成无法正确判断聚类中心点的情况, 如图 2 所示.

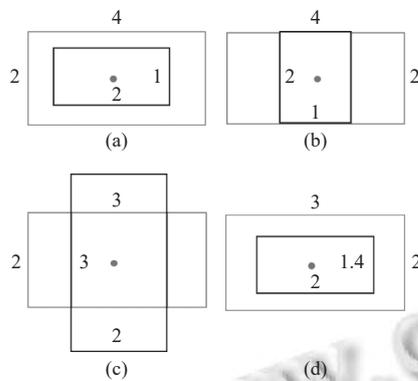


图 2 先验框生成示意图

图 2(a) 的 $IoU=0.5$, 图 2(b) 的 $IoU=0.5$, 使得 $d_a=d_b$, 无法判断该目标框属于哪一个聚类中心. 同理, 图 2(c) 的 $IoU=0.5$, 图 2(d) 的 $IoU=0.47$, $d_a < d_b$, 该目标框被划分给了图 2(c) 聚类中心, 而事实上该目标框划分给图 2(d) 聚类中心更加适合. 因此提出一种新的衡量指标 $WIoU$, 将 K-means 聚类中的距离衡量公式替换为 d_1 :

$$d_1 = 1 - WIoU_{(A,B)} \quad (1)$$

$$WIoU_{(A,B)} = IoU_{(A,B)} - \frac{|C \setminus (A \cup B)|}{|C|} - \alpha v \quad (2)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w_i}{h_i} \right)^2 \quad (3)$$

其中, C 为包含 A 与 B 的最小边界框; $C \setminus (A \cup B)$ 为 C 中去除 A 与 B 后的区域; α 是权重函数, 经过实验 $\alpha=1$ 时效果最好, v 用来度量长宽比的相似性.

使用 $WIoU$ 作为衡量指标时, 图 2(a) 的 $WIoU=0.5$, 图 2(b) 的 $WIoU=0.33$, $d_a=0.5 < d_b=0.67$, 目标被划分给更合适的聚类中心图 2(a); 图 2(c) 的 $WIoU=0.33$, 图 2(d) 的 $WIoU=0.469$, $d_a=0.67 > d_b=0.531$, 目标框划分给更匹配的聚类中心图 2(d). 由此可知, 使用 $WIoU$ 计算出来的先验框与真实标注框的贴合程度最高.

1.2 SE-PPM 模块

在水面垃圾数据集, 发现两个很大的挑战: 1) 不同的分类的目标框形状尺度差别很大; 2) 塑料垃圾和瓶子垃圾与背景颜色差异很小. 如图 3 所示.

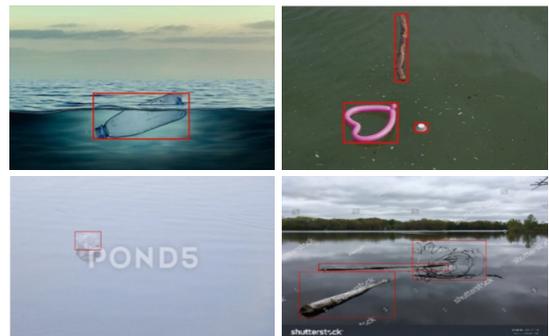


图 3 数据集的尺度形状以及背景差异

图 3 数据集的尺度形状以及背景差异

针对这两个挑战, 本文提出一种新的解决方法, 即带注意力机制的多尺度池化模块 (SE-PPM), 如图 4 所示.

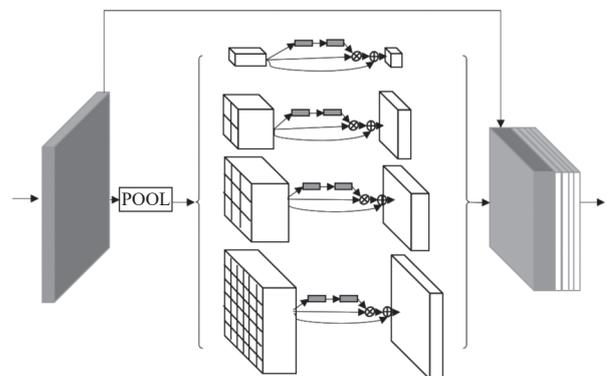


图 4 SE-PPM 网络结构

通过图4可以看出, 首先把主干网络提取的特征池化成4个不同尺度的特征图, 分别为 $1 \times 1 \times C$, $2 \times 2 \times C$, $3 \times 3 \times C$ 和 $6 \times 6 \times C$. 不同尺度的池化模块PPM^[20]的使用, 首先能将不同尺度的特征图转化为固定尺寸的特征图, 提高了图像目标的尺寸不变性, 缓解了图像中目标形状大小差异问题; 其次, 大尺度的池化能够提升感受野, 感知目标的整体信息, 小尺度的池化能关注目标的细节信息, 然而过大尺度的池化必然会导致细节信息丢失, 因此通过主干网络特征和经过不同尺度池化后特征图进行通道融合. 特征融合后的特征图使用一个卷积核为 1×1 的卷积层对通道进行调整, 最终与输入的特征图通道保持一致.

在经过池化后的特征图, 添加进带残差结构的通道注意力机制(SEnet)^[21]结构中, 经过squeeze和excitation后得到各个通道相应的权值再与原特征图对应通道的所有元素相乘, 这样的好处是能够加强图像中目标的特征, 降低检测与背景颜色差异小的目标的难度. 但注意力机制往往是聚焦于重要信息, 忽略相对不重要的信息, 造成了细节信息的丢失, 使小目标的检测变得更加困难, 为了保留细节信息不被丢弃, 借用残差网络的思想, 把输入的特征图与经过通道注意力机制提取的特征图相加.

实验证明, 本文设计的带注意力机制的多尺度池化模块(SE-PPM)不仅能增强对多尺度目标信息变化的关注, 而且一定程度上缓解了待检测目标和背景颜色差异小而漏检的问题.

1.3 多向金字塔

目标检测任务中, 随着卷积网络深度增加能获得更好的语义信息的特征图, 但深度增加的同时, 特征图的尺寸也在指数倍减小, 很多小目标的位置信息在此过程中被当成边缘信息被忽略掉, 导致小目标在检测任务中经常出现漏检和误检的情况. YOLOv3算法虽然通过最邻近元插值法上采样使3个不同尺度的特征图进行融合, 在一定程度上缓解了小目标位置信息的缺失, 但是对浅层的目标信息利用不够充分, 降低了对小目标的检测精度. 针对上述的挑战, 本文提出了一种新的基于特征融合的改进方法, 其网络结构如图5所示.

多向金字塔网络(multidirection FPN)的具体流程为: 首先通过自底向上的方式把携带丰富特征语义信息的深层上采样与携带准确位置信息的浅层特征融合,

在原有的基础上上采样融合特征到尺度为 112×112 的特征图. 融合后的特征图经过5个基础卷积模块后进入检测头部, 特征图在检测头部中先使用一个基础卷积块(图5中的DBL)调整通道, 再使用自顶向下的方式把融合后携带丰富上下文信息的特征图通过下采样与深层进行融合, 进一步增强检测小尺度的检测头中的上下文信息, 最后通过一个卷积核为 1×1 的卷积层(图5中的conv1)调整通道数后进行预测. 实验表明, 多向金字塔网络获得了携带更丰富上下文信息的特征图, 提高了检测精度.

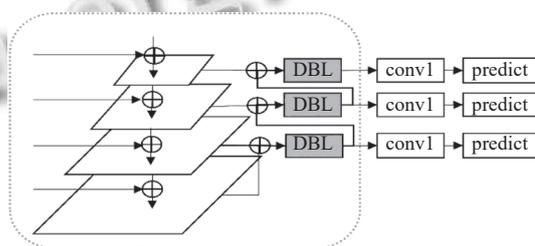


图5 多向金字塔网络结构

1.4 改进的损失函数

YOLOv3算法是基于网格进行预测的, 针对每个网格都会使用3个先验框去试探, 从而选出置信度最高的先验框帮助模型进行预测得到预测框, 再根据预测框的得分和位置等信息计算模型的总损失. 检测目标的总损失是由检测框损失, 置信度损失和类别损失加权求和得到. 其计算公式如下:

$$l_{box} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{m=0}^B I_{im}^{obj} (2 - w_i \times h_i) [(x_i - x'_i)^2 + (y_i - y'_i)^2 + (w_i - w'_i)^2 + (h_i - h'_i)^2] \quad (4)$$

$$l_{cls} = \lambda_{class} \sum_{i=0}^{S^2} I_{im}^{obj} \sum_{c \in classes} [(\hat{p}_i(c) \log(p_i(c)) + (1 - \hat{p}_i(c)) \log(1 - p_i(c)))] \quad (5)$$

$$l_{obj} = \sum_{i=0}^{S^2} \sum_{m=0}^B I_{im}^{obj} (c_i - c'_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{m=0}^B I_{ij}^{noobj} (c_i - c'_i)^2 \quad (6)$$

$$loss = l_{box} + l_{cls} + l_{obj} \quad (7)$$

事实上在目标检测任务中, 包含目标的网格在待检测的特征图中占比是很小的, 又因为YOLOv3使用的是多尺度的预测方法, 每一次检测时会输入 14×14 ,

28×28, 56×56 这 3 个待检测的特征图像, 这意味每次会输出 12 102 个检测框, 导致正负样本的比例不均衡的问题更突出了. Focal loss^[12] 从样本难易分类的角度出发解决了样本不均衡的问题, 提升了检测精度. 原因有二: 1) 根据统计学可知样本越少的类别分类的难度越大; 2) 模型在训练过程中会更倾向于难分类的目标图像, 更有助于小目标准确度的提升. 本文在通过 focal loss 替换置信度损失公式中负样本的交叉熵公式, 使模型在训练过程中让正样本占据更大权重, 使模型的梯度向正样本方向倾斜, 有效信息得到多次学习, 检测的精度因此也得到提升.

在 YOLOv3 模型中使用交叉熵公式来计算负样本置信度损失, 交叉熵如式 (8) 所示:

$$L_{ce} = -y \log \hat{p} - (1-y) \log(1-\hat{p})$$

$$= \begin{cases} -\log \hat{p}, & y = 1 \\ -\log(1-\hat{p}), & y = 0 \end{cases} \quad (8)$$

其中, y 表示真值, 1 为正样本, 0 为负样本; \hat{p} 表示检测值, 取值范围在 0 到 1 之间. 从式 (8) 可知, 正样本检测的概率越大或负样本的检测概率越小, L_{ce} 就越小. 因此, 在迭代训练中, 大量的负分类样本输入会导致模型偏差.

Focal loss^[16] 是基于交叉熵公式上, 通过调整置信度损失中正负样本比例, 使正负样本达到平衡. Focal loss 公式如下:

$$L_{fl} = -\alpha y(1-\hat{p})^\gamma \log \hat{p} - (1-\alpha)(1-y)\hat{p}^\gamma \log(1-\hat{p})$$

$$= \begin{cases} -\alpha y(1-\hat{p})^\gamma \log \hat{p}, & y = 1 \\ -(1-\alpha)\hat{p}^\gamma \log(1-\hat{p}), & y = 0 \end{cases} \quad (9)$$

如式 (9) 所示, focal loss 在交叉熵公式基础上添加了一个权重因子 γ . 当 $\gamma=0$ 时, focal loss 就为交叉熵公式. 当 $\gamma>0$ 时, 损失函数会减少易分类样本的计算, 增加对难分类样本的计算. 此外, 加入了平衡因子 α , 用来调整正负样本比例. 实验发现当 $\gamma=2$ 和 $\alpha=0.8$ 时, 实验效果最好.

2 实验结果与分析

为了验证改进的 YOLOv3 水面垃圾检测算法的优越性, 使用自制的水面垃圾检测数据集进行了以下实验: (1) 使用改进前后的 K-means 聚类算法的实验对比; (2) 本文 SPMYOLOv3 算法与其他目标检测算法的对比; (3) 多尺度对比实验分析; (4) 不同改进策略的消融

实验对比.

2.1 实验环境

算法的训练和测试环境如表 1 所示.

表 1 网络模型训练与测试环境

名称	版本
操作系统	Windows 10 专业版
CPU	AMD Ryzen7 5800H
GPU	NVIDIA GeForce RTX3070
框架	PyTorch 1.10.0

2.2 实验数据集

为验证改进 SPMYOLOv3 算法的优越性, 必须具备现实中复杂环境下的水面垃圾数据集. 数据集的来源有两种: 第 1 部分来源于政府部门; 第 2 部分来源于网络搜集水面垃圾图像. 数据集使用 Labeling 标注工具按照 voc 数据集格式标注了共 2 928 张水面垃圾检测图片, 共包含 5 814 个目标, 定义 32×32 尺寸以下为小目标, 96×96 尺寸以下为中目标, 其余尺寸为大目标, 其尺度分布如图 6 所示.

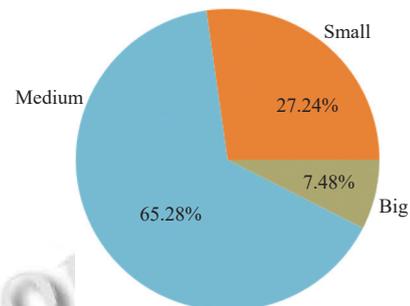


图 6 水面垃圾数据集尺度分布

从图 6 可知, 水面垃圾数据集内中小目标共占 92.52%. 数据集总共包含了 bottle, grass, branch, milk-box, plastic-bag, plastic-gar-bage, ball 和 leaf 这 8 个类别, 同时数据集中包含丰富的现实场景, 有远景, 近景, 单目标, 多目标等, 符合实际情况, 如图 7 所示.

2.3 改进前后的 K-means 聚类算法实验分析

在 YOLOv3 算法中有 3 个检测头, 每个检测头中包含 3 个先验框, 因此使用 K-means 算法聚类时设定 $k=9$. 当 $k=9$ 时, 比较 K-means 聚类算法改进前后聚类得到 $M_{IoU}(\%)$ 和 $mAP(\%)$ 的具体数据如表 2 所示.

改进后的先验框与数据集的检测框的 M_{IoU} 提升了约 13.20%, 聚类后的锚选框加入到 YOLOv3 原始模型中进行训练, 实验结果发现 mAP 提升了 0.43%. 由

此可知由改进的 M_IoU 指标聚类获得的 9 个先验框与水面垃圾数据集匹配程度更好。



(a) 单目标 (b) 多目标 (c) 近景 (d) 远景

图 7 水面垃圾数据集

2.4 检测算法对比实验分析

本文为验证改进后的模型的性能,对改进 YOLOv3

检测算法与几大主流目标检测算法进行两个方面比较:一是通过比较实验检测的效果图;二是比较实验检测的结果。

为了测试各模型对水面垃圾检测效果,选取了测试集中几张有代表性的图像进行对比,如图 8 所示。

表 2 K-means 聚类算法改进前后尺寸和精度对比

方法	检测头	先验框	M_IoU (%)	mAP (%)
改进前 K-means	14×14	(25.51, 25.92), (34.41, 52.74), (51.22, 88.25)	68.10	69.36
	28×28	(73.27, 202.23), (67.63, 41.26), (88.91, 111.59)		
	56×56	(143.96, 59.53), (255.38, 299.09), (167.95, 162.34)		
改进后 K-means	14×14	(22.74, 28.67), (39.50, 73.91), (44.46, 38.80)	81.30	69.79
	28×28	(90.67, 63.74), (109.92, 29.82), (71.09, 110.15)		
	56×56	(73.27, 202.24), (259.22, 281.56), (155.53, 149.62)		

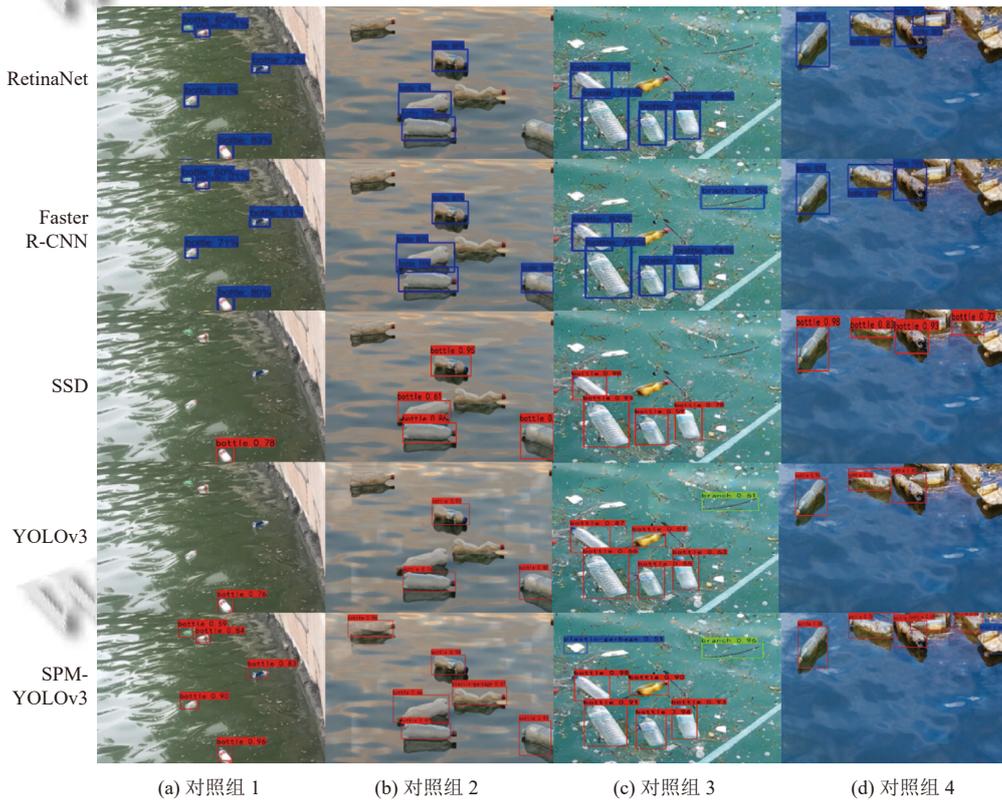


图 8 5 种算法的检测效果图

从图 8 的检测结果可以看出,改进的 YOLOv3 算法在检测效果图上表现最好,从图 8(a)-图 8(d) 这 4 组对照实验,其他算法均出现漏检和误检的情况,而

改进后的算法不仅没有出现误检和漏检,同时检测精度均高于其他算法。

实验检测结果采用目标检测领域常用的评价指标

平均精度均值 (mAP) 和每秒检测帧数 (FPS) 作为改进的 YOLOv3 检测算法和其他目标检测算法的衡量标准。

其中, mAP 的计算公式为:

$$mAP = \frac{\sum_{i=1}^N AP_{(i)}}{N} \quad (10)$$

其中, N 代表类别数; $AP_{(i)}$ 代表每一类的准确度的均值, 其计算公式如下:

$$AP_{(i)} = \int_0^1 P_{(i)}(R_{(i)}) dR_{(i)} \quad (11)$$

使用相同的训练集和测试集通过随机梯度下降 (SGD) 训练 150 个 epochs. 其实验结果如表 3 所示。

表 3 不同算法检测结果对比

算法	基础网络	mAP (%)	FPS	检测精度 AP (%)							
				瓶子	草	树枝	牛奶盒	塑料袋	塑料垃圾	球	树叶
RetinaNet	ResNet50	66.39	42.83	77.67	35.87	46.74	86.45	81.24	57.31	83.88	61.99
Faster R-CNN	ResNet50	68.10	13.38	72.10	48.29	47.50	80.87	82.90	51.17	87.51	74.7
SSD300	VGG-16	61.94	48.63	68.88	50.51	35.00	82.68	79.03	50.30	75.36	53.77
YOLOv3	Darknet53	69.79	62.33	79.72	56.10	45.01	89.87	76.30	58.97	90.04	62.37
SPMYOLOv3	Darknet53	73.32	58.66	80.81	60.11	46.09	91.64	86.40	69.05	91.48	75.52

从表 3 的检测结果对比中可知, 改进的 YOLOv3 算法在所有模型中的检测精度最高, 比二阶段的 Faster R-CNN 和一阶段的 SSD300 以及 RetinaNet 分别提高了 5.22%, 11.38%, 6.93%。从单个类别的检测精度上比较, 本文改进的算法在大多数类别中准确度最高。同时检测速度相比于最好的 YOLOv3 也没有太多的下降。改进后的算法既能满足对水面垃圾目标的检测精度要求, 也能满足对水面垃圾目标检测的实时性要求。

2.5 多尺度对比实验分析

在本节中, 为了深入探索本文算法对不同尺度的目标检测精度的影响, 分别对不同尺度的目标进行了实验, 实验结果如表 4 所示。

表 4 各尺度目标检测精度 mAP 对比 (%)

方法	整体	小尺度	中尺度	大尺度
YOLOv3	69.36	55.40	73.33	85.60
SPMYOLOv3	73.32	61.09	76.84	87.30
对比	↑3.96	↑5.59	↑3.51	↑1.70

从表 4 可知, 本文在大尺度目标的检测精度上只提升了 1.70%, 效果比较一般。但本文算法改进对于中小目标的检测精度提升较大, 分别为 3.51% 和 5.59%, 极大地改善了对于水面垃圾小目标检测精度不高的问题, 这也验证了本文算法针对中小目标检测精度不高的改进是有效的。

2.6 消融实验及分析

在本节中, 为了深入研究每个模块的改进对原始 YOLOv3 网络的影响, 本文在进行消融实验时逐个加入各个改进的模块, 并记录其效果。表 5 中√表示选择

使用该模块, 最终得到的结果如表 5 所示。其中, 方案 1 代表原始 YOLOv3 模型的精度; 方案 2 为使用 K-means 算法后, YOLOv3 模型的精度; 方案 3 是在 K-means 算法后, 使用 SE-PPM 模块改进 YOLOv3 的模型精度; 方案 4 在方案 3 的基础上, 加入 MdFPN 模块后的模型精度; 方案 5 在方案 4 的基础上使用 Focal loss 损失函数替换原有的损失函数得到模型的精度。

表 5 Darknet53 下不同改进方案消融实验的 mAP 对比

方案	K-means	SE-PPM	MdFPN	Focal loss	mAP (%)	对比 (%)
1	—	—	—	—	69.36	—
2	√	—	—	—	69.79	↑0.43
3	√	√	—	—	71.39	↑2.03
4	√	√	√	—	72.34	↑2.98
5	√	√	√	√	73.32	↑3.96

从表 5 可以看出, K-means 算法的改进在精度的提升了 0.43%, 且先验框和真实框的 M_{IoU} 提升了 13.20%, 使得检测框和先验框得到更大的交并比, 帮助模型更好的预测检测结果。

添加 SE-PPM 模块对 YOLOv3 网络的影响效果明显, 精度提升了 1.6%, 因为该模块通过多尺度池化缓解水面垃圾目标形状多变的难题, 且保留全局信息, 再通过其中带残差的注意力机制使得重要特征更容易被检测到, 提升了检测精度; 加入 SE-PPM 和 MdFPN 两个模块, 精度提升了 2.55%, 原因有二: 1) 多向特征融合解决了上下文信息利用不充分的问题, 提升了模型的检测精度; 2) SE-PPM 突出重要的局部信息, 保留了全局信息, 对于多向特征融合网络有更好的搭配作用, 让特征信息得到更充分的利用。

使用 focal loss 替换原模型中二元交叉熵函数也起到较为明显的效果,虽然在添加 focal loss 后精度只提升了 0.98%,但 focal loss 的使用解决了 YOLOv3 模型中负样本失衡的问题,同时让模型向正样本方向进行梯度更新,加快了模型的收敛速度,是很好的一个改进策略。

3 总结

水面垃圾检测在水体保护领域有重大意义,本文针对目前水面垃圾目标与背景差异颜色小,形状尺度多变以及小目标检测精度不高的问题提出一种 SPMYOLOv3 水面垃圾目标检测方法。通过先验框的优化,带注意力的多尺度池化网络,特征融合网络改进和交叉熵损失函数替换等措施提升了改进算法在水面垃圾数据集上的检测效果。后续工作将考虑使用轻量化网络替换基础网络,在确保精度不损失情况下,实现算法在下游移动端设备上的应用。

参考文献

- 1 Viola P, Jones MJ. Robust real-time face detection. *International Journal of Computer Vision*, 2004, 57(2): 137–154. [doi: [10.1023/B:VISI.0000013087.49260.fb](https://doi.org/10.1023/B:VISI.0000013087.49260.fb)]
- 2 Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection. *Proceedings of the 26th International Conference on Neural Information Processing Systems*. Lake Tahoe: Curran Associates Inc., 2013. 2553–2561.
- 3 王斌. 海面目标红外检测方法研究 [博士学位论文]. 大连: 大连海事大学, 2018.
- 4 方晶, 冯顺山, 冯源. 水面航行体对舰船目标的图像检测方法. *北京理工大学学报*, 2017, 37(12): 1235–1240.
- 5 魏建荣. 背景纹理模型在海面舰船目标检测中的应用研究. *舰船科学技术*, 2017, 39(10A): 159–161.
- 6 汤伟, 刘思洋, 陈景霞, 等. 基于计算机视觉的水面目标分割与识别方法. 2018 中国自动化大会 (CAC2018) 论文集. 西安: 中国自动化学会, 2018. 532–537.
- 7 Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014. 580–587.
- 8 Girshick R. Fast R-CNN. *Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015. 1440–1448.
- 9 Dai JF, Li Y, He KM, et al. R-FCN: Object detection via region-based fully convolutional networks. *Proceedings of the 30th International Conference on Neural Information Processing Systems*. Barcelona: Curran Associates Inc., 2016. 379–387.
- 10 Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 779–788.
- 11 Fu CY, Liu W, Ranga A, et al. DSSD: Deconvolutional single shot detector. arXiv:1701.06659, 2017.
- 12 Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. *Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 2999–3007.
- 13 李佳琪, 吴湘宁, 邓中港, 等. 基于注意力及生成对抗网络的遥感影像目标检测方法. 中国, CN113989612A. 2022-01-28.
- 14 He KM, Gkioxari G, Dollár P, et al. Mask R-CNN. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 2980–2988.
- 15 王志鹏, 王涛. 基于 Faster RCNN 的穿越围栏违规行为检测. *计算机系统应用*, 2022, 31(4): 346–351. [doi: [10.15888/j.cnki.csa.008423](https://doi.org/10.15888/j.cnki.csa.008423)]
- 16 Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2015. 91–99.
- 17 王林, 赵红. 改进 YOLOv3 的火灾检测. *计算机系统应用*, 2022, 31(4): 143–153. [doi: [10.15888/j.cnki.csa.008404](https://doi.org/10.15888/j.cnki.csa.008404)]
- 18 孙备, 左震, 吴鹏, 等. 面向无人艇环境感知的改进型 SSD 目标检测方法. *仪器仪表学报*, 2021, 42(9): 52–61. [doi: [10.19650/j.cnki.cjsi.J2108000](https://doi.org/10.19650/j.cnki.cjsi.J2108000)]
- 19 Hartigan JA, Wong MA. Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1979, 28(1): 100–108.
- 20 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 21 Zhao HS, Shi JP, Qi XJ, et al. Pyramid scene parsing network. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 6230–6239.

(校对责编: 牛欣悦)