

# 协助水下机器人作业的目标识别算法设计

胡睿康<sup>1</sup> 张紫辉<sup>2</sup> 李玉寒<sup>2</sup>

(1. 桂林电子科技大学 信息与通信学院, 广西 桂林 541004; 2. 桂林电子科技大学 机电工程学院, 广西 桂林 541004)

**摘要:** 目前, 目标识别技术广泛应用于各个领域, 文章提出一种基于 YOLOv5 的改进目标识别算法, 提升算法准确度, 达到水下目标识别的目的。使用自动多尺度视网膜与颜色恢复算法对图像进行预处理, 将头部解耦与卷积块注意模块引入 YOLOv5 目标识别模型中得到优化 YOLOv5 模型, 对优化 YOLOv5 模型进行稀疏、剪枝与微调 3 个策略, 得到改进 YOLOv5 模型。实验表明, 改进算法在准确度上有所提升, 同时保证了算法的低耗时。

**关键词:** 水下图像; 图像增强; 目标识别; YOLOv5

**中图分类号:** TP242 **文献标识码:** A **文章编号:** 1003-9767 (2023) 07-181-04

## Design of Target Recognition Algorithm to Assist Underwater Vehicle Operation

HU Ruikang<sup>1</sup>, ZHANG zihui<sup>2</sup>, LI Yuhang<sup>2</sup>

(1. School of Information and Communication, Guilin University of Electronic Technology, Guilin Guangxi 541004, China; 2. School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin Guangxi 541004, China)

**Abstract:** At present, target recognition technology is widely used in various fields, and the paper proposes an improved target recognition algorithm based on YOLOv5 to improve the accuracy of the algorithm and achieve the purpose of underwater target recognition. The images are preprocessed using automatic multiscale retina and color recovery algorithm, the head decoupling and convolutional block attention modules are introduced into the YOLOv5 target recognition model to obtain the optimized YOLOv5 model, and three strategies of sparse, pruning and fine-tuning are applied to the optimized YOLOv5 model to obtain the improved YOLOv5 model. Experiments show that the improved algorithm improves in accuracy while ensuring the low time consumption of the algorithm.

**Keywords:** underwater image; image enhancement; object recognition; YOLOv5

### 0 引言

水下作业常因为水流湍急、能见度低、环境复杂等因素, 导致作业开展的危险系数高、效率低、耗费人力物力等问题。针对以上问题, 遥控无人潜水器 (Remote Operated Vehicle, ROV) 在众多领域已经得到应用。其中, 在水产养殖业中, 水下机器人可以进行大范围养殖水质监测、水产品生长状态评估与行为研究<sup>[1]</sup>。在热液探测中, 水下机器人可以进行地形地貌探测、水体环境探测

和样品采集等工作<sup>[2]</sup>。在水利工程中, 水下机器人可以进行大坝缺陷和渗漏检测、混凝土结构水下检测和金属结构水下检测等工作<sup>[3]</sup>。可见, 水下机器人应用范围广, 但是水中杂物较多, 操作者定位所搜寻物体较为困难, 对水下目标识别算法研究有重大意义。

### 1 水下目标识别算法研究

#### 1.1 水下目标识别算法国内研究现状

对于水下目标识别算法, 国内已经有一定进展。徐

**收稿日期:** 2023-02-28

**基金项目:** 2022 年国家级大学生创新创业项目“基于水下目标识别技术的江河紧急搜救机器人”(项目编号: 202210595064)。

**作者简介:** 胡睿康 (2002—), 男, 江西宜春人, 本科在读。研究方向: 深度学习、信号处理。

承等人水声目标识别模型中引入对比学习无监督特征与多维自注意力机制,加强数据的特征表达与模型特征抽取能力,显著提升识别效果<sup>[4]</sup>。郭雨青等人通过对图像进行基于水下光衰减先验算法与 CLAHE 算法预处理提高图像对比度再将图像用于训练目标识别模型,识别平均精度提升了 33%<sup>[5]</sup>。徐建华等人基于 YOLOv3 算法的基础上引入多级融合、降采样重组、重新定义损失函数等方式改进 YOLOv3 模型,使得改进 YOLOv3 模型平均准确度能达到 75.1%,识别速度为 15 frames/s<sup>[6]</sup>。

## 1.2 改进 YOLOv5 模型研究

文章通过自动多尺度视网膜与颜色恢复 (Automatic Multi-Scale Retinex with Color Restoration, AMSRCR) 算法预处理数据集,将数据集集中的图像进行色彩矫正、对比度提升,提升数据集的质量。以 YOLOv5 为基础目标识别模型,引入头部解耦和卷积块注意力模块 (Convolutional Block Attention Module, CBAM) 注意力机制模块以提高 YOLOv5 目标识别模型的性能。将优化 YOLOv5 模型进行稀疏训练、剪枝、微调,不影响优化 YOLOv5 模型性能的同时压缩模型数据量与加快目标识别速度,得到改进 YOLOv5 模型。改进 YOLOv5 模型在识别准确度指标上得到有效提升,而压缩算法使得改进 YOLOv5 模型的识别速度能够达到 60 frames/s,能够很好地满足操作者实时进行水下作业的要求。改进算法整体框图如图 1 所示。

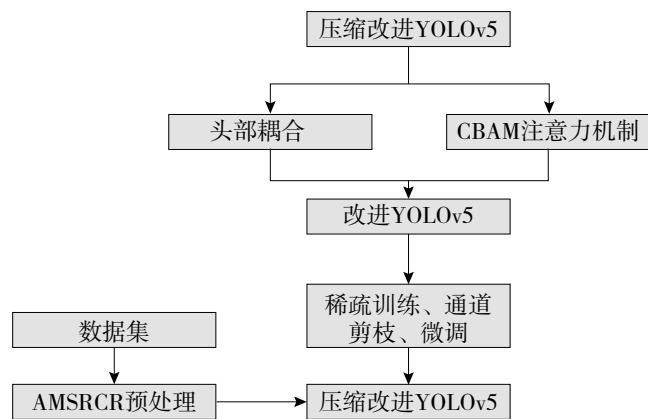


图 1 改进 YOLOv5 模型整体框图

## 2 图像预处理

Retinex 理论指出,图像是由反射光和照射光组成,而图像中物体的反射光是其固有属性,因此通过移除照射光,即可得到物体的真实图像。计算物体真实图像的数学表达式为

$$r(x,y)=\log S(x,y)-\log L(x,y) \quad (1)$$

式中:  $r$  为物体反射光图像;  $S$  为原始图像;  $L$  为照射光分量。

单尺度 Retinex 算法 (SSR) 基于 Retinex 理论,提出一种获得照射光分量的有效算法。获得光照分量的数学表达式与高斯模糊滤波器的数学表达式分别为

$$L=F*S \quad (2)$$

$$F=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{r^2}{\sigma^2}\right) \quad (3)$$

式中:  $F$  为高斯模糊滤波器;  $*$  为卷积符号;  $r$  为图像像素平方和;  $\sigma$  为尺度参数。

尺度参数较小时可以较好地保留图像细节但是色彩会有较大偏差。尺度参数较大时,图像可以有较好的色彩平衡,但是细节损失严重。针对以上问题,提出多尺度 Retinex 算法 (MSR),该算法将高、中、低 3 个尺度参数下得到的照射光分量进行算术平均,使得细节与色彩得到平衡。MSR 算法表达式为

$$r(x,y)=\sum_k^K w_k \left\{ \log S(x,y) - \log [F_k(x,y) \bullet S(x,y)] \right\} \quad (4)$$

式中:  $k$  为高斯模糊滤波器个数;  $w_k$  为算术平均系数。

虽然使用高斯尺度参数估计照射光分量计算速度快,但是存在强光与阴影过渡区出现光晕和色彩保持能力较弱的问题。多尺度视网膜与颜色恢复 (Multi-Scale Retinex with Color Restoration, MSRCR) 算法针对这两个问题提出基于原图像的色彩恢复因子,以此来调节 3 个通道的颜色比例。MSRCR 算法表达式为

$$R_{MSRCR_i}(x,y)=C_i(x,y)R_{MSR_i}(x,y) \quad (5)$$

式中:  $C_i$  表示第  $i$  个通道的色彩恢复因子。

MSRCR 算法中的参数依赖于原图像,从而导致算法对不同环境下的图像处理效果存在差异。因此,文章提出 AMSRCR 算法,参数从 MSRCR 增强后的图像中获得,使算法的适应性更强。由于水下环境颗粒多、光线弱,较为复杂,使用图像处理提升数据集的质量可以明显影响目标识别算法的性能<sup>[7]</sup>。AMSCRCR 算法基于 3 个算法的改进,在细节保留、色彩平衡与适应性等方面表现优秀,因此采用 AMSRCR 算法作为预处理算法,以提速算法的性能。

## 3 YOLOv5 算法原理及改进

### 3.1 YOLOv5 目标识别模型

YOLOv5 算法以其高精度与低耗时在目标检测算法中表现突出。而 YOLOv5s 是 YOLOv5 算法中计算量最小、耗时最短的模型,为达到水下机器人作业的实时性要求,选择其作为基础模型进行改进。其中, YOLOv5 目标识别模型的输入端由 Mosaic 数据增强与锚框构成,主干网络由跨阶段的平行网络 (Cross Stage

Partial Network, CSPNet) 构成, 颈部由特征金字塔网络 (Feature Pyramid Network, FPN) 和像素聚合网络 (Pixel Aggregation Network, PAN) 金字塔构成, 预测端由 GLOU\_loss 构成。

### 3.2 头部解耦

为降低定位与分类之间的干扰, 将预测层的任务分解为定位和分类两个子任务。头部解耦即将头部的卷积分解为两个子卷积, 两个子卷积分别进行定位子任务与分类子任务。在执行定位子任务的卷积中, 卷积更注重物体的边界信息与细粒度, 以保证定位框的召回率。在执行分类任务的卷积中, 卷积更注重图像中的信息, 物体包含的特征与上下文的语义关系。头部解耦可以让 YOLOv5 目标识别模型定位框定位更准确, 降低分类的错误率, 使得模型的性能得到提升。

### 3.3 优化 YOLOv5 模型压缩

对优化 YOLOv5 模型进行稀疏训练、通道减枝、微调操作, 降低模型计算量与耗时的同时保证性能<sup>[8]</sup>。稀疏训练是对模型进行稀疏, 将缩放因子引入每个通道中, 从而对通道重要程度进行区分。通道剪枝则是分析每个通道缩放因子, 剪除缩放因子小的通道, 保留缩放因子大的通道。通道剪枝后的模型的网络会减小, 网络间的连接性变弱, 结构整体变得稀疏、耦合性差, 导致模型的识别性能下降。微调则是对剪切后的模型进行调整, 通过微调批量样本的归一化 (Batch Normalization, BN) 层的参数使网络进行简单的移动, 模型重新变得紧凑, 网络性能回升。

## 4 试验分析

### 4.1 实验环境

实验环境如下: AMD® Ryzen 5 4600h with radeon graphics × 12; 内存为 16 GB, NVIDIA GeForce GTX 1650 4 GB; 操作系统为 Ubuntu 18.04.6 LTS; 软件环境

为 Python3.8; 框架为 PyTorch1.12.1。

数据集采用全国水下机器人 (湛江) 大赛提供的 URPC opticalimage dataset 数据集。该数据集中包含 11 087 张图像, 按 8 : 1 的比例将数据集分为训练集和验证集, 并在 YOLOv5s 的基础上分别训练 300 轮。

### 4.2 改进 YOLOv5 模型评估

在对 YOLO 模型优化后, 需要通过目标识别指标对模型进行客观评价与验证不同策略对模型改进的有效性。对 YOLOv5 目标识别模型、采用头部解耦改进的 YOLOv5 模型、采用卷积块注意模块改进的 YOLOv5 模型以及采用头部解耦和卷积块注意模块的 YOLOv5 模型 (优化 YOLOv5 模型模型) 进行消融实验。实验中的客观指标包括精确度、召回率、 $mAP^{val}0.5$ 、 $mAP^{val}0.5 : 0.95$ 、GPU 运行下的帧率、内存量。其中,  $mAP^{val}0.5$  表示在交并比阈值 0.5 下的均值平均精度值,  $mAP^{val}0.5 : 0.95$  表示在不同交并比阈值 (从 0.5 到 0.95, 步长 0.05) 下的均值平均精度值消融。实验结果如表 1 所示。

从表 1 可以得出, 在加入头部解耦后, YOLOv5 目标识别模型的召回率提升, 但内存量是原模型的 2 倍多。在加入 CBAM 注意力机制后精确度提升, 但是  $mAP$  略有下降。将头部解耦和 CBAM 改进策略同时应用在 YOLOv5 目标识别模型中, 其精确度与召回率均有明显提升, 但是其余指标下降。对优化 YOLOv5 模型进行稀疏、剪枝与微调进行压缩, 得到改进 YOLOv5 模型。

对优化 YOLOv5 模型使用压缩策略, 对比压缩前后模型的实验结果, 如表 2 所示。

从表 2 可以得出, 压缩算法将优化 YOLOv5 模型内存量压缩 5.8MB, 压缩率为 20%, 使得改进 YOLOv5 模型计算量降低提升帧率。相比于 YOLOV5 目标识别模型, 改进 YOLOv5 模型在精确度与召回率指标上均有提升, 而除内存量增加外, 其余性能指标均无明显变化。

表 1 消融实验

算法名称	精确度 /%	召回率 /%	$mAP^{val}0.5$	$mAP^{val}0.5 : 0.95$	GPU 运行下的帧率 / (frames/s)	内存量 /MB
YOLOv5	64.3	47.0%	0.513	0.26	60	14.4
YOLOv5+ 头部解耦	64.5%	48.7%	0.511	0.258	52.72	29.1
YOLOv5+CBAM	65.5%	46.5%	0.502	0.255	61.84	14.5
YOLOv5+ 头部解耦 +CBAM	65.7%	48.3%	0.507	0.253	57.01	29.2

表 2 优化 YOLOv5 模型与改进 Yolov5 模型对比实验

模型名称	准确度 /%	召回率 /%	$mAP^{val}0.5$	$mAP^{val}0.5 : 0.95$	GPU 运行下的帧率 / (frames/s)	内存量 /MB
优化 YOLOv5 模型	65.7%	48.3%	0.507	0.253	57.01	29.2
改进 YOLOv5 模型	66.0%	48.1%	0.505	0.251	62.42	23.4

(下转第 188 页)



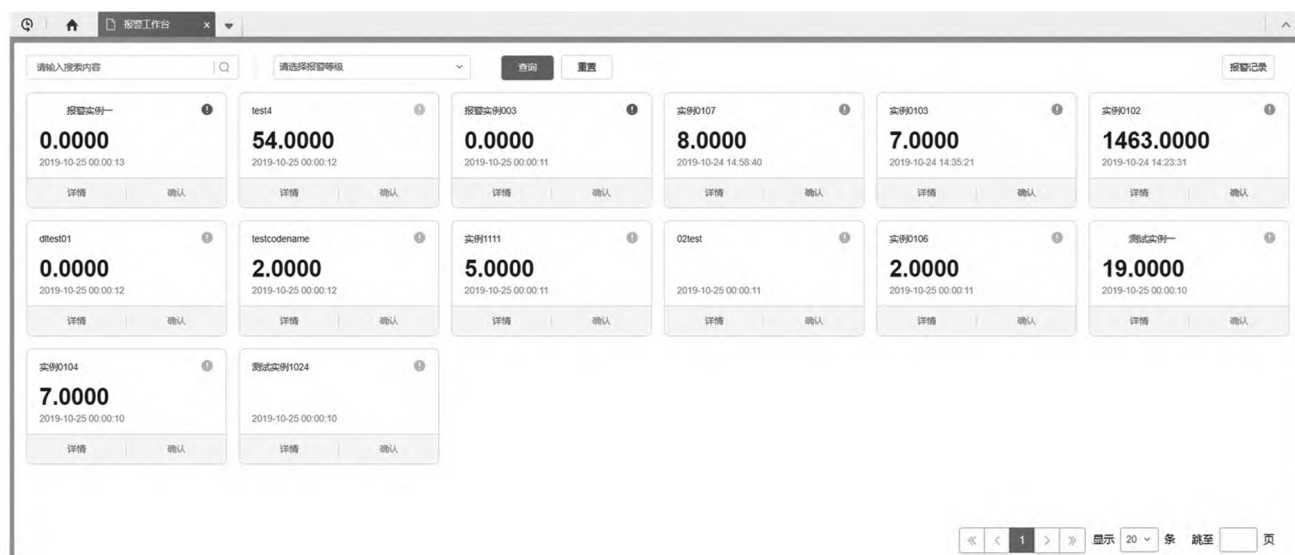


图6 看板报警示意图

警功能,为我国智慧环保管理平台建设的安全稳定运行和未来优化升级提供一定参考和启迪。

### 参考文献

- [1] 范晓鹏. 基于智慧环保的互联网+环保管家模式探讨[J]. 环境与发展, 2019(12):202-203.
- [2] 刘士清. 环保大数据在智慧环保监管领域的应用探

究[J]. 中国资源综合利用, 2020,38(6):28-30

- [3] 孙传国. 智慧城市背景下的智慧环保建设探索[J]. 智能城市, 2019(22):119-120.
- [4] 李晓韩. 基于智慧环保的区域环境管理[J]. 化工管理, 2020(28):88-89.
- [5] 胥彦玲, 李纯, 闫润生. 中国智慧环保产业发展趋势及建议[J]. 环境保护, 2018(7):119-123.

(上接第183页)

## 5 结语

针对缺乏水下图像数据集与水下图像质量低的问题,文章提出一种以YOLOv5为基础模型的改进YOLOv5模型。首先采用在细节保留、色彩平衡与曝光等方面均表现良好的AMSRRCR算法对数据集进行预处理,使得数据集质量提升,被识别目标的细节更容易被提取。其次,将分离定位任务与分类任务的头部解耦模块与融合空间注意力机制和通道注意力机制的CBAM注意力机制引入YOLOv5目标识别模型中,使得模型的精确度与召回率得到提升。最后,针对优化YOLOv5模型耗时增加、内存量增大的问题,对模型使用稀疏训练、通道剪枝与微调一系列压缩模型算法对模型进行压缩,使得内存量与耗时降低,整体性能提升。

### 参考文献

- [1] 夏英凯,朱明,曾鑫,等.水产养殖水下机器人研究进展[J]. 华中农业大学学报, 2021,40(3):85-97.
- [2] 蔡巍,陶春辉,王渊,等. 自主水下机器人海底热液

区应用综述[J/OL]. 机器人:1-13[2023-03-22].<https://doi.org/10.13973/j.cnki.robot.220100>.

- [3] 卢普光,刘莎. 水下机器人在水利工程检测中的应用现状[C]//中国水利学会2021学术年会论文集第三分册, 2021:321-323.
- [4] 徐承,李勇,张梦,等. 基于特征融合和自注意力机制的水下目标识别[J]. 移动通信, 2022,46(6):91-98.
- [5] 郭雨青,曾庆军,夏楠,等. 图像增强水下自主机器人目标识别研究[J]. 中国测试, 2021,47(11):47-52.
- [6] 徐建华,豆毅庚,郑亚山. 一种基于YOLO-V3算法的水下目标识别跟踪方法[J]. 中国惯性技术学报, 2020,28(1):129-133.
- [7] 吴宇,蔡永斌,汤荣华. 水下视觉图像处理和识别技术研究[J]. 舰船电子工程, 2019,39(5):93-96.
- [8] LIU Z, LI J, SHEN Z, et al. Learning efficient convolutional networks through network slimming[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017:2736-2744.