



湖北大学学报(自然科学版)
Journal of Hubei University(Natural Science)
ISSN 1000-2375,CN 42-1212/N

《湖北大学学报(自然科学版)》网络首发论文

题目: 基于改进 YOLOv5 的鱼群小目标检测优化算法
作者: 汪沛洁, 谌雨章, 王诗琦, 周雯
收稿日期: 2022-11-14
网络首发日期: 2023-03-03
引用格式: 汪沛洁, 谌雨章, 王诗琦, 周雯. 基于改进 YOLOv5 的鱼群小目标检测优化算法[J/OL]. 湖北大学学报(自然科学版).
<https://kns.cnki.net/kcms/detail/42.1212.N.20230302.1647.006.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

文章编号:1000-2375(2023)05-0000-00

基于改进 YOLOv5 的鱼群小目标检测优化算法

汪沛洁, 谌雨章, 王诗琦, 周雯

(湖北大学计算机与信息工程学院, 湖北 武汉 430062)

摘要:随着深度学习技术的发展,水下图像检测近年来受到广泛的关注,为了克服在复杂水下环境下传统小鱼群的误检、漏检和识别准确率低等问题,提出一种改进 YOLOv5 的目标检测方法(INV-YOLOv5).该方法包括将 YOLOv5m 中的 Focus 模块替换为卷积模块,提高网络精度;在主干网络(Backbone)中添加多头自注意力机制,增大网络特征提取视野;最后,在网络中引入了内卷算子和加权的特征融合,降低网络的参数量,提高检测精度.在实验阶段,使用 Labeled Fishes in the Wild 数据集和 WildFish 数据集验证,该方法的平均精度(mAP)分别为 81.7% 和 83.6%,与 YOLOv5m 网络相比分别提升了 6% 和 14.5%,不仅拥有较高的识别率并且更加轻量化,而且模型大小与 YOLOv5m 网络相比减少了 6 M (Mega)左右,验证了所提出的改进方法具有较好的效果.

关键词:深度学习;YOLOv5m;多头自注意力;内卷算子;鱼群检测

中图分类号:TP391.4;TP183 **文献标志码:**A **DOI:**10.3969/j.issn.1000-2375.2023.00.009

著录信息:汪沛洁,谌雨章,王诗琦,等. 基于改进 YOLOv5 的鱼群小目标检测优化算法[J]. 湖北大学学报(自然科学版),2023,45(5). DOI:10.3969/j.issn.1000-2375.2023.00.009.
WANG P J, CHEN Y Z, WANG S Q, et al. Optimization algorithm of fish shoal small target detection based on improved YOLOv5 [J]. Journal of Hubei University(Natural Science), 2023, 45(5). DOI:10.3969/j.issn.1000-2375.2023.00.009.

Optimization algorithm of fish shoal small target detection based on improved YOLOv5

WANG Peijie, CHEN Yuzhang, WANG Shiqi, ZHOU Wen

(School of Computer Science and Information Engineering, HuBei University, Wuhan 430062, China)

Abstract: With the development of deep learning technology, underwater image detection has received extensive attention in recent years. In order to overcome the problems of false detection, missing detection and low detection accuracy of traditional small fish groups in complex underwater environments, an improved YOLOv5 target detection method (INV-YOLOv5) was proposed. The method included replacing the focus module in YOLOv5m with the convolution module to improve the network accuracy. A multi-head self attention mechanism was added to the backbone to increase the view of network feature extraction. Finally, the feature fusion of involution operator and weighting was introduced into the network to reduce the amount of network parameters and improve the detection accuracy. The experimental verification on Labeled Fishes in the Wild dataset and WildFish dataset shows that the average accuracy (mAP) of this method is 81.7% and 83.6% respectively, which is increased by 6% and 14.5% respectively compared with YOLOv5m network. It not only has a higher recognition rate, but also is more lightweight. Compared with YOLOv5m network, the model size is reduced by about 6m (Mega), which verifies that the proposed improved method has a better effect.

收稿日期:2022-11-14
基金项目:教育部产学合作协同育人项目(202101142041)和大学生创新创业训练计划项目(X202110512062, X202110512069, X202110512100)资助
作者简介:汪沛洁(1998-),女,硕士生;谌雨章,通信作者,副教授,博士,主要研究方向为光电探测、图像处理,E-mail:hubucyz@foxmail.com

Key words: deep learning; YOLOv5m; multi-head self attention; involution; small fish detection

0 引言

随着世界各国对海洋资源利用水平的不断提升,世界范围内的学者更加关注水下图像检测问题.如何利用深度学习方法替代鲁棒性差、检测精度低的传统目标检测算法,从而达到更高的检测速度和检测精度,逐渐成为研究热点.目前,通用的目标检测技术虽然比较成熟,但在不同的水下场景中,鱼类目标的种类、形状、大小、数量等存在差异,对检测精度和速度的要求也不同,它给水下环境中的目标检测任务带来了挑战.目前,基于深度学习的鱼群目标检测已经取得了较高的准确率,然而,它在复杂的水下环境中还存在着很多挑战,主要有以下几个方面:

(1) 现有的目标检测算法是针对中/大目标而设计的,对于水下小目标的检测的效果较差.

(2) 小目标易受环境干扰,水下的条件复杂,光照不均匀导致小目标的特征信息难以被网络检测到.

(3) 各鱼群尺寸大小不一,小目标鱼群的可利用像素较小,网络很难提取到较好的特征信息,并且随着网络层数的增加,小目标特征信息及位置信息极易丢失,从而导致小目标鱼群目标的漏检.

近年来研究人员在水下小目标鱼群检测方面做了大量的工作,并取得了一定的成果.在实际自然场景中,由于光线、水质、水下环境等自然条件的原因,水下采集的图片分辨率低,且鱼群在图像中容易与背景颜色融为一体,在这样的条件下检测鱼群是一个较大的挑战.当前,基于深度学习的目标检测方法已经吸引了大量研究学者在此方向上进行研究,并在很多方面得到了进一步的扩展、完善和应用.基于深度学习的目标检测方法主要分为两类:两阶段(Two-stage)目标检测算法和一阶段(One-stage)目标检测算法.

两阶段目标检测算法是将检测问题划分为两个阶段,首先产生候选区域,然后对候选区域进行特征提取、域分类和位置调整.这类算法的代表有 R-CNN^[1] 及其变体 Fast R-CNN^[2]、Faster R-CNN^[3] 和 R-FCN^[4] 等.例如,袁红春等^[5]将 MSRCR 算法与 Faster R-CNN 二次迁移学习算法引入到小规模水下鱼类识别中,实现了鱼类目标的检测. Li 等^[6]通过改进区域卷积网络卷积层的结构,构建了检测鱼类目标的神经网络,在公开数据集上获得了较高的平均精度.一般来说两阶段算法具有较高的检测精度,但检测速度较慢.一阶段目标检测算法实现了检测和分类同时进行,通过回归直接输出物体的类别概率和位置坐标值. YOLOv1^[7]、YOLOv2^[8]、YOLOv3^[9] 和 SSD^[10] 等.例如, Xu 等^[11]使用 YOLOv3 模型检测三种不同的鱼类数据集,实现了仅使用两个数据集进行训练,对所有三个数据集进行测试,获得了平均 0.539 2 的识别准确率; Sung 等^[12]基于 YOLO 的架构实现了一个包含 24 个卷积层和两个全连接层的 CNN 结构,对自定义鱼群数据集进行了检测;王文成等^[13]提出了一种改进 SSD 的鱼类目标检测算法,具有较高的准确率.一阶段目标检测虽然检测速度较快,但是检测精度通常低于两阶段算法.

基于深度学习的鱼类目标检测算法具有更高的精度和鲁棒性,但检测精度和速度仍有待提高.因此,结合深度学习目标检测方法研究小鱼群目标检测算法,对提高鱼群图像检测的精度和速度具有重要的研究价值.因此,本研究针对水下小目标鱼群检测任务,展开关于在情况复杂的水下的小目标鱼群检测研究,以实现水下小目标鱼群的检测.

1 基于多头自注意力机制和内卷算子的 YOLOv5 算法

1.1 YOLOv5 算法 YOLOv5^[14]于 2020 年 5 月提出,是一个高性能、通用的目标检测算法. YOLOv5 包含四个部分: Input 部分、Backbone 部分、Neck 部分以及 Output 部分. YOLOv5 结构图如图 1 所示.

1.1.1 Input 部分介绍 Input 部分对输入的图片进行了 Mosaic 数据增强、图片尺寸处理、自适应锚框设计系列操作. Mosaic 数据增强技术丰富了样本的多样性,大大降低了模型进行多样性学习的难度,该方法非常适用于小目标检测,满足本文水下小鱼群检测的需要^[15]. 图片尺寸处理操作是将输入图片调

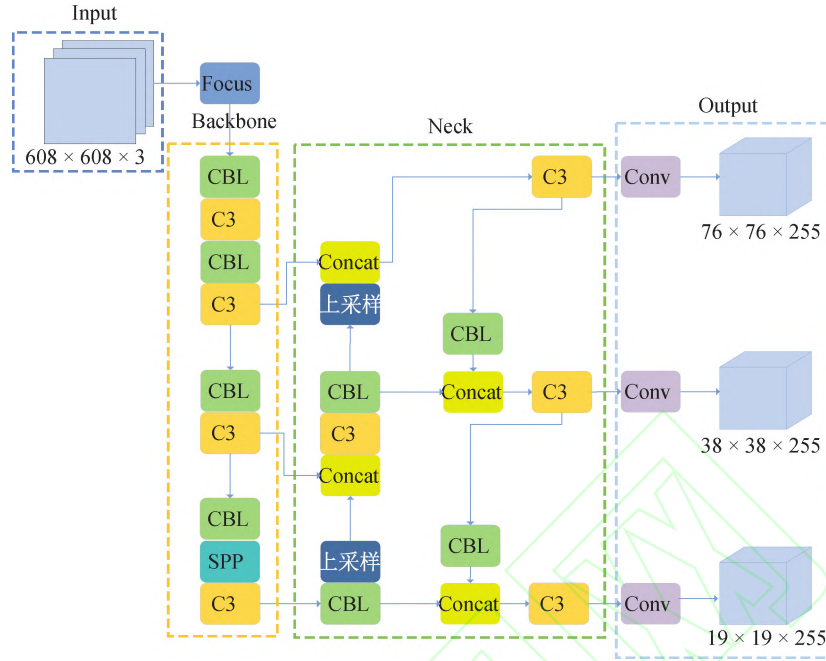


图 1 YOLOv5 结构图

整为网络接受的标准尺寸,再输入进网络模型中进行训练.自适应锚框设计是指网络在初始锚框的基础上输出预测框并与真实框进行比较,计算两者差距并反向更新,不断迭代调整网络参数.

1.1.2 Backbone 部分介绍 Backbone 部分是主干网络,其作用是提取特征,由 Focus 模块和 CSP 结构组成.其中,Focus 模块为 YOLOv5 独创的结构,其作用为对进入网络的图片进行切片操作,提高卷积速度.如图 2 所示,图像的像素尺寸与输入网络前相比缩小了一半,减少了卷积的计算量,提高了卷积运算速度,最重要的是同时保留了图像完整的信息,但降低了网络的检测精度.本研究将 Focus 模块替换为 Conv 模块,以提高网络的检测精度.

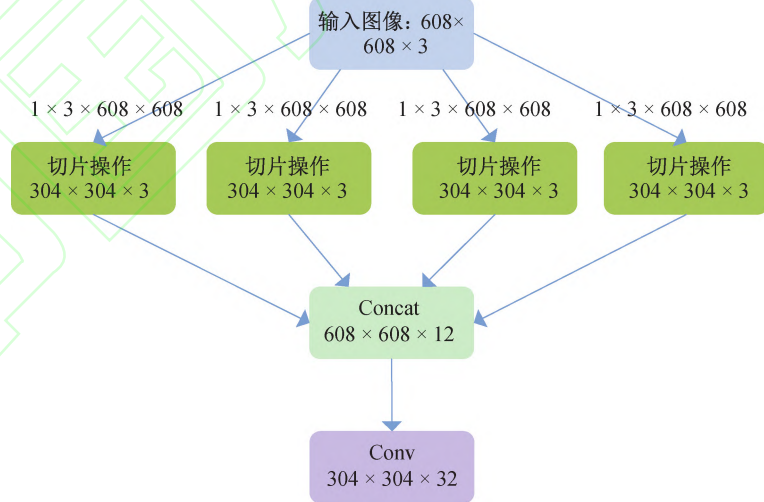


图 2 Focus 模块结构示意图

YOLOv5 版本使用

了两种 CSP 结构,分别是 CSP1_X 和 CSP2_X. Backbone 网络中主要使用了 CSP1_X 结构,Neck 结构中主要使用 CSP2_X 结构,新版 YOLOv5 中,将 Bottleneck CSP(瓶颈层)模块转变为了 C3 模块,结构作用基本相同均为 CSP 架构,是对残差特征进行学习的主要模块,其结构分为两支,如图 3 所示.

1.1.3 Neck 部分介绍 Neck 部分是网络中间的衔接部分,起到传递获取的特征融合信息的作用. Neck 部分是 FPN+PAN 的结构,其结构如图 4 所示. FPN 结构是自顶向下的,通过上采样和融合低层信息的方式得到预测的特征图. PAN 是采用自底向上的对 FPN 进行补充的特征金字塔.

1.1.4 Output 部分介绍 Output 部分是输出,用来输出小目标识别的结果. YOLOv5 中使用 GIOU-loss 作为损失函数来替换普通的 IoU 计算方法,有效解决了边界框不重合的问题. YOLOv5 在性能上略低于

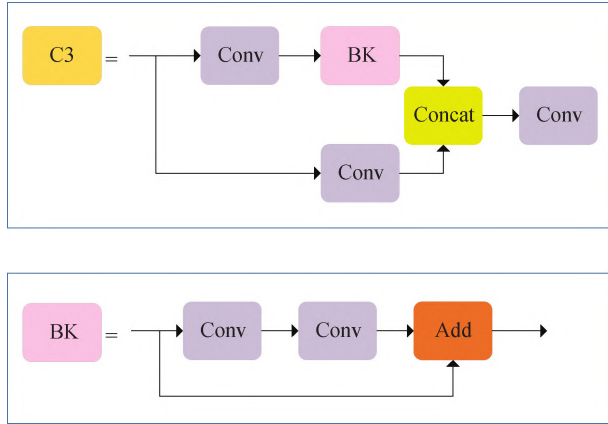


图 3 C3 模块结构示意图

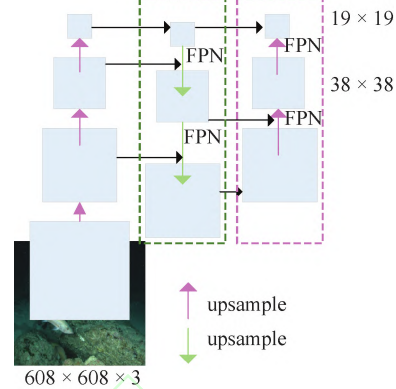


图 4 Neck 部分结构示意图

YOLOv4, 但比 YOLOv4 更灵活、更快, 所以它有一个优势在模型部署.

1.2 多头自注意力机制 注意力机制 (attention mechanism) 模仿了生物观察行为的过程^[16], 注意力机制可以快速提取系数数据的重要特征, 通过对输入图像数据采用注意力机制, 可以突出图片中目标的关键特征并抑制次要信息, 进而提高整个神经网络模型的检测的效率和精度, 同时也会促使模型学习中再次出现类似特征时将注意力放在该图像特征上, 提高模型的学习及泛化能力^[15]. 注意力机制中存在查询矩阵 Q 、键矩阵 K 和值矩阵 V . 而自注意力机制 (self-attention mechanism) 是注意力机制的改进, 其令 $Q=K=V$, 抑制了网络对外部信息的依赖, 更关注数据或特征内部之间的联系^[16], 能够处理比常规卷积更大的感受野, 提高模型的学习能力, 并增强了神经网络的可解释性.

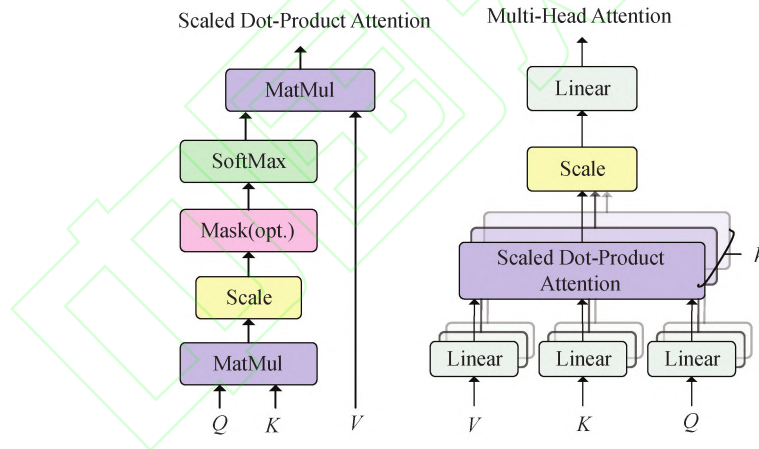


图 5 缩放点积注意力和多头自注意力模块结构示意图

多头注意力机制 (multi-head attention) 的本质是在参数量总体不变的情况下, 将查询 Q 、键 K 、值 V 三个参数进行 h 组拆分, 每组拆分将参数映射到高维空间的不同子空间中执行注意力函数运算计算注意力权重, 从而能够从多个维度提炼特征信息^[16], 得到最终的输出, 其结构如图 5 所示. 多头自注意力以缩放点积注意力 (scaled dot-product attention, SDA) 为原理, 如式 (1) 所示:

$$SDA(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (1)$$

式中: Q, K, V 分别为计算的自注意力的查询、键、值矩阵; QK^T 为注意力矩阵, 对 V 矩阵加权; d_k 表示键的维数; $\sqrt{d_k}$ 是将注意力矩阵变为标准的正态分布, 使结果更稳定.

并行计算各个子空间的 SDA, 如式 (2) 所示:

$$\text{head}_i = SDA(QW_1^Q, KW_1^K, VW_1^V) \quad (2)$$

融合所有子空间的注意力信息, 如式 (3) 所示:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}^0 \quad (3)$$

其中, (X, Y) , $\mathbf{W}_1^Q, \mathbf{W}_1^K, \mathbf{W}_1^V$ 为线性变换时的参数矩阵, head_i 表示多头自注意力模块中的第 i 个头, Concat 表示拼接操作, MultiHead 表示最终的结果. 可见多头自注意力机制可以捕获空间上更远距离的信息, 因此将多头自注意力机制引入到 C3 模块中, 提高网络的检测能力.

1.3 内卷算子 经典的卷积核具有空间不变性和通道独立性, 但刘文婷等^[16]提出的新的神经网络算子 Involution (内卷算子) 却与之相反.

Involution 的示意图如图 6 所示.

Involution 核 $H \in R^{H \times W \times K \times K \times G}$ 包含 Spatial 和 Channel 的反向特征变换. 是专门为对应坐标 (i, j) 的像素 $X_{i,j} \in R^C$ 定制的, 但在通道上共享, 其中 C 表示计算通道, G 表示每个组共享相同的 involution 核的组数, 卷积核的大小取决于输入特征映射的大小. 利用 involution 核对输入进行乘加运算, 得到 involution 的输出特征映射:

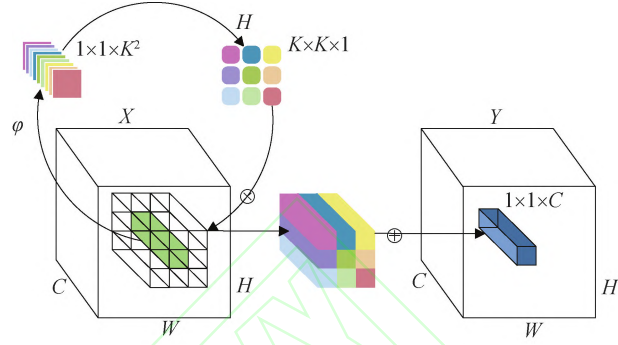


图 6 Involution 示意图

$$Y_{i,j,k} = \sum_{(u,v) \in \Delta K} H_{i,j,u+\lfloor K/2 \rfloor, v+\lfloor K/2 \rfloor, \lceil kG/C \rceil} X_{i+u, j+v, k} \quad (4)$$

Involution 核 H 的形状取决于输入特征映射 X 的形状, 生成以原始输入张量为条件的 Involution 核, 使输出核与输入核对齐. 这里将核生成函数符号为 ϕ , 并将每个位置 (i, j) 的函数映射抽象为:

$$H_{i,j} = \phi(X_{\psi_{i,j}}) \quad (5)$$

从单个像素 $X_{i,j}$ 跨越每个 Involution 内核 $H_{i,j}$. 每个像素 $\psi_{i,j}$ 有核生成函数 ϕ , 其形式如式 (6) 所示:

$$H_{i,j} = \phi(X_{i,j}) = W_{1\sigma}(W_0 X_{i,j}) \quad (6)$$

其中, W_0 和 W_1 代表两个线性变换, 共同构成 Bottleneck 结构, σ 表示批处理归一化后对两个线性变换的非线性激活函数 BN 和 ReLU. 基于以上优点, 本文将部分 C3 模块中引入内卷算子即 C3Inv 模块来替代部分原 YOLOv5 中的 C3 模块, 这有利于捕获空间长距离信息和减少网络参数的数量.

2 改进算法的结构

YOLOv5 目标检测算法本身是一个比较优秀的检测算法, 目前国内外研究仍有很大的改进空间. 参考前面提到的算法理论研究对 YOLOv5 算法进行了改进, 得到改进的 YOLOv5 算法——INV-YOLOv5. 其结构如图 7 所示, 网络整体结构如表 1 所示.

考虑到图像中存在占像素信息少的目标, 更容易受到背景的影响, YOLOv5m 模型算法在使用卷积网络检测样本时容易漏掉小目标的部分特征信息, 从而致使小目标的检测精度不高, 因此在主干模块中, 将 Focus 模块替换为 Conv 模块, 目的是提高网络的检测精度. C3 模块被 Involution 参与的 C3Inv 模块所取代, 部分 C3 模块被多头注意力机制参与的 C3TR 模块所取代, 这有利于特征提取, 可以捕获空间上更远距离的信息, 告诉模型需要注意哪些位置和信息, 进一步提高检测的准确率, 可以减少网络参数量. 同时, 将 Concat 修改为加权的 Weightconcat (加权的 Concat), 给特征图不同的权重然后再合并在一起, 如式 (7) 所示:

$$x = \text{conv}(\delta(\sum_{i=0}^1 W_i X_i)) \quad (7)$$

其中, W_0 和 W_1 是进行归一化处理后得到的, 网络输出为 w'_0 和 w'_1 , 如式 (8) 所示:

$$W_0 = \frac{w'_0}{w'_0 + w'_1}; W_1 = \frac{w'_1}{w'_0 + w'_1} \quad (8)$$

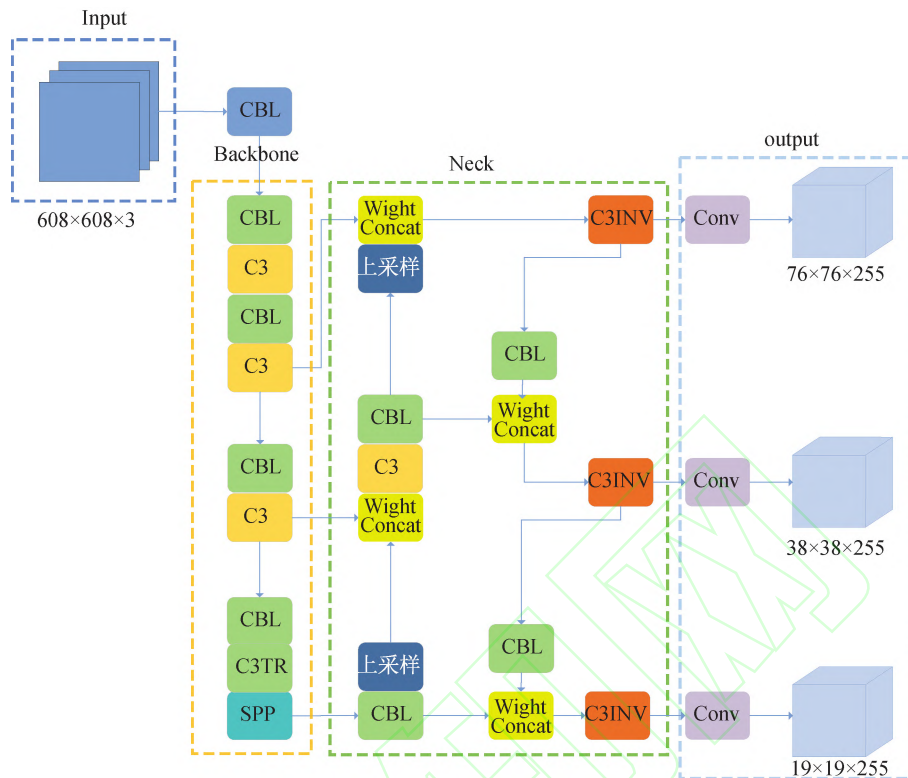


图 7 改进 YOLOv5 网络框架示意图

表 1 改进 YOLOv5 网络整体结构

序号	输入	参数	子模块	模块参数
0	-1	5280	Conv	[48, 6, 2, 2]
1	-1	41664	Conv	[96, 3, 2]
2	-1	65280	C3	[96, 96, 2]
3	-1	166272	Conv	[192, 3, 2]
4	-1	629760	C3	[192, 192, 6]
5	-1	664320	Conv	[384, 3, 2]
6	-1	2512896	C3	[384, 384, 6]
7	-1	2655744	Conv	[384, 768, 3, 2]
8	-1	3987840	C3TR	[768, 768, 2, False]
9	-1	1476864	SPP	[768, 768, [5, 9, 13]]
10	-1	295680	Conv	[768, 384, 1, 1]
11	-1	0	Upsample	[None, 2]
12	[-1, 6]	148229	Weightconcat	[384, 384]
13	-1	1035264	C3	[384, 384, 2, False]
14	-1	74112	Conv	[384, 192, 1, 1]
15	-1	0	Upsample	[None, 2]
16	[-1, 4]	37253	Weightconcat	[192, 192]
17	-1	119640	C3Inv	[192, 192, 2, False]
18	-1	332160	Conv	[192, 192, 3, 2]
19	[-1, 14]	37253	Weightconcat	[192, 192]
20	[-1, 6]	0	Concat	[1]
21	-1	548592	C3Inv	[576, 384, 2, False]
22	-1	1327872	Conv	[384, 384, 3, 2]
23	[-1, 10]	148229	Weightconcat	[384, 384]
24	-1	1597152	C3Inv	[384, 768, 2, False]

3 实验结果与分析

3.1 数据集的选取与建立

3.1.1 数据集介绍 本文中使用的公开数据集有 Labeled Fishes in the Wild 数据集^[17]和 WildFish 数据集^[18]. Labeled Fishes in the Wild 数据集是从部署在遥控车辆(ROV)上的摄像机系统记录的水下视频中提取的,它包括两类(鱼和非鱼)和 1 329 张图像. WildFish 是鱼类检测和识别领域最大的数据集. WildFish 由 1 000 种鱼类和 54 459 张图像组成,图片来自 FishBase^[19]、专业网站(如海洋中心和博物馆)和谷歌图像搜索. 图片经过筛选、清洗以及标注等步骤,增加了许多小目标和密集目标的数量,可以提高对小目标和密集目标的检测率. 图 8 所示为 Labeled Fishes in the Wild 数据集的分析可视化结果图示例,图 8(a)中横纵坐标表示中心点 x, y 坐标的位置,图 8(b)中横坐标 width 物体的宽,纵坐标 height 代表物体的高. 最终得到的数据集总共有 6 765 张图片,将数据集按照按照 8 : 2 的比例分为两部分:训练集和测试集. 其中训练集 5 412 张图片,测试集 1 353 张图片. 图 9 是来自两数据集的示例图像.

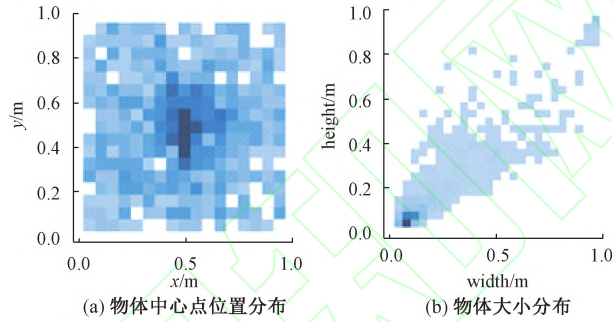


图 8 数据集分析

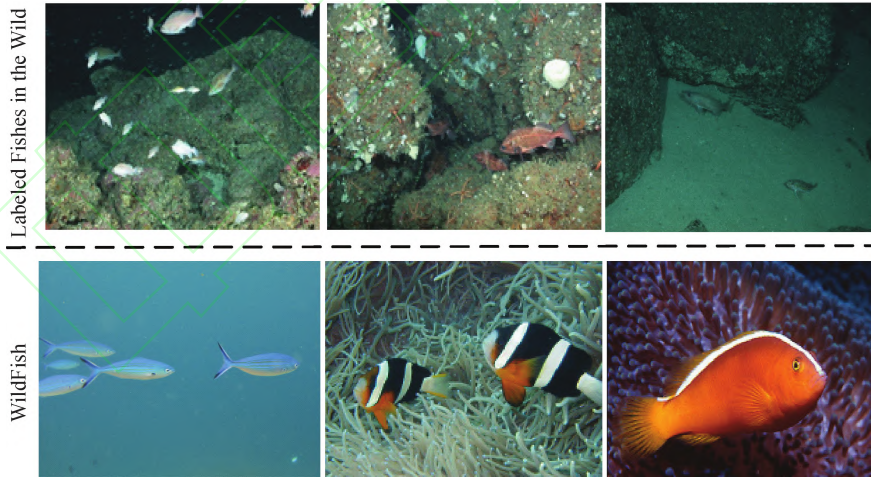


图 9 来自数据集的示例图像

3.1.2 数据集标记 经过 LabelImg 标注的数据集是 PASCAL VOC 格式的 .xml 文件,由于本研究使用了算法 YOLOv5m,所以需要将其转换为 YOLO 格式的 .txt 格式文件,其公式如下:

$$\begin{aligned} x &= (x_{\max} + x_{\min})/d_w \\ y &= (y_{\max} + y_{\min})/d_h \\ w &= (x_{\max} - x_{\min})/d_w \\ h &= (y_{\max} - y_{\min})/d_h \end{aligned} \quad (9)$$

.xml 文件中物体位置坐标的左上角坐标和右下角坐标为 $x_{\max}, x_{\min}, y_{\max}, y_{\min}$. d_w, d_h 表示图片的高和宽. 数据集标注示例如图 10 所示.

3.2 实验配置 本实验在 Windows10 操作系统中进行,采用 Python 语言编程实现,如表 2 所示,列举了

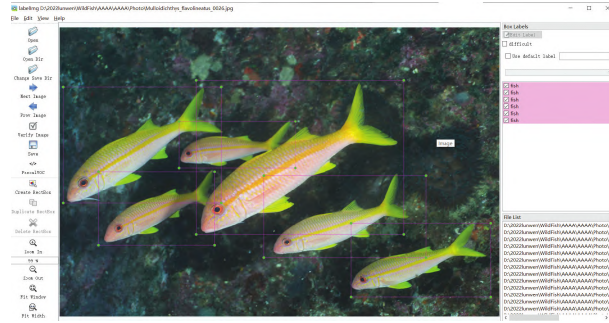


图 10 数据集标注示例

实验硬件配置和所搭建的程序环境.

表 2 改进 YOLOv5 识别小鱼程序环境

环境搭建	
操作系统	Windows10
CPU	R7-4800H
显卡	RTX2060(115W)
内存	16 GB
Python	Python3.7

3.3 评价指标的选取 本实验在 Windows10 操作系统中进行,采用 Python 语言编程实现,如表 2 所示,列举了实验硬件配置和所搭建的程序环境.

本实验使用精确率(Precision)、召回率(Recall)、平均精度 AP(average precision)、平均精度均值 mAP(mean average precision)来评价模型.其中,ture positives(TP)是指将鱼群的图片识别成正样本;ture negatives(TN)是指背景的图片没有被识别出来,或者是背景被正确分类成背景;false positives(FP)是指将背景图片识别成了鱼群;false negatives(FN)是指将鱼群的图片识别为背景^[20].

精确率(Precision)表示所有被识别出来的鱼群中正确识别的鱼群所占的比例:

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{n} \quad (10)$$

召回率(Recall)是被正确识别出来的鱼群数量与测试集中所有的鱼群数量的比值:

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

平均精度 AP(average precision)指的是以召回率 Recall 为横轴以精确率 Precision 为纵轴组成的曲线围成的面积.公式如下:

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) p(r_{i+1}) \quad (12)$$

实验中平均精度均值 mAP(mean average precisin)是指数据集中所有类别的平均精度的均值.公式如下:

$$mAP = \frac{1}{m} \sum_{i=1}^m AP_i \quad (13)$$

其中, m 表示测试集中的样本个数.

3.4 结果分析 为了更好地展现该改进模型的优势,实验将改进的 YOLOv5m 算法即 INV-YOLOv5 算法与 YOLOv5m、YOLOv4、SSD、CenterNet 在 Labeled Fishes in the Wild 和 WildFish 两种数据集上进行对比,模型使用相同实验环境来进行训练、验证,并将实验结果进行分析.

表 3 所示为各个模型在数据集 Labeled Fishes in the Wild 上在 mAP、Recall 和实时检测 FPS(frames per second)的对比结果.从表 3 的实验结果可以看出,在 Labeled Fishes in the Wild 这个数据集上,INV-YOLOv5 与 YOLOv5m 相比, mAP 高出了 6%,与 YOLOv4、SSD、CenterNet 相比, mAP 提升了 9.1%、

10.8% 和 13.6%.

表 3 数据集 Labeled Fishes in the Wild 实验数据

方法	mAP@ 0.5/%	Recall/%	FPS
CenterNet	68.1	82.1	11.33
SSD	70.9	85.9	18.25
YOLOv4	72.6	91.1	7.21
YOLOv5m	75.7	95	20.47
INV-YOLOv5	81.7	94	15.69

表 4 所示为各个模型在数据集 WildFish 上在 mAP、Recall 和实时检测 FPS 的对比结果. 从表 4 的实验结果可以看出, 在 Labeled Fishes in the Wild 这个数据集上, INV-YOLOv5 与 YOLOv5m 相比, mAP 提升了 14.5%, 与 YOLOv4、SSD、CenterNet 相比, mAP 均有提升, 说明改进的 YOLOv5 算法 INV-YOLOv5 检测准确率优于目前先进算法, 并且 FPS 相差不大. 同时处理速度有较大提升.

表 4 数据集 WildFish 实验数据

方法	mAP@ 0.5/%	Recall/%	FPS
CenterNet	63.6	92.2	10.55
SSD	65.8	94	15.64
YOLOv4	68.5	93	10.15
YOLOv5m	69.1	96	18.96
INV-YOLOv5	83.6	97	16.83

表 5 所示为各个模型在数据集 Labeled Fishes in the Wild 上复杂度对比的对比结果. 从表 5 的实验结果可以看出, 在数据集 Labeled Fishes in the wild 上 YOLOv5 检测算法中的 YOLOv5m 模型与一阶段检测模型 SSD 和 YOLOv4 还有 CenterNet 相比较, YOLOv5m 属于轻量级的模型. 本文提出的方法在 YOLOv5m 轻量级的基础上更减少了模型的参数量, 由原本的 41.9 MB 减少到了 35.41 MB.

表 5 不同网络复杂度对比

方法	参数量	模型大小/MB
CenterNet	32 719 832	124
SSD	26 285 486	100.2
YOLOv4	64 040 001	244.29
YOLOv5m	21 056 406	41.9
INV-YOLOv5	17 931 602	35.41

为了更直观地分析检测结果, 图 11 可视化了 YOLOv5m 与 INV-YOLOv5 在 Labeled Fishes in the Wild 数据集上的检测结果. 图 12 可视化了 YOLOv5m 与 INV-YOLOv5 在 WildFish 数据集上的检测结果. 综合模型大小、mAP、FPS 及可视化结果来看, 本文中方法 INV-YOLOv5 在鱼群小目标检测上具有较好的性能.

3.5 消融实验 为了进一步验证改进算法的有效性, 对算法所提的替换 Focus 模块、使用 C3Inv、使用 C3TR 和使用 Weightconcat 模块等方法在 WildFish 数据集上设计了 5 组消融实验, 每组实验所使用的环境和训练技巧相同. 实验结果可从表 6 得到. 其中 YOLOv5m-A 表示在原始 YOLOv5m 的主干网络中将 Focus 模块替换为 Conv 模块; YOLOv5m-B 表示在 YOLOv5m-A 基础上将网络中一部分的 C3 模块替换为有 Involution 参与的 C3Inv 模块; YOLOv5m-C 表示在 YOLOv5m-B 的基础上将网络中的另一部分 C3 模块替换为被多头注意力机制参与的 C3TR 模块; INV-YOLOv5 表示本研究所提出的模型, 在 YOLOv5m 网络的基础上同时替换 Focus 模块、使用 C3Inv、使用 C3TR 和使用 Weightconcat 模块. 通过对比平均精度可以看出, 在原始 YOLOv5m 的主干网络中将 Focus 模块替换为 Conv 模块提高了模型 4.8% 的 mAp, 证明 Conv 模块能够提高网络的检测精度. 提出的 C3Inv 模块让模型的 mAp 提升 2.3%. C3TR 模块对于特征提取有显著效果, mAP 提高 5.7%, 证明该模块能够增强特征的表达能力, 有利于特征提取捕获空间上更远距离的信息, 告诉模型更需要注意哪些位置和信息, 进一步提高检测的准确率. Concat 修改为加

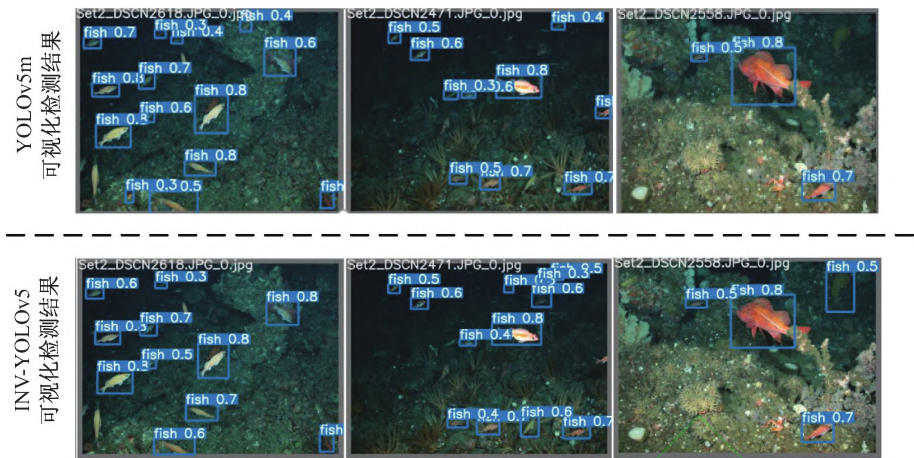


图 11 Labeled Fishes in the Wild 数据集可视化结果

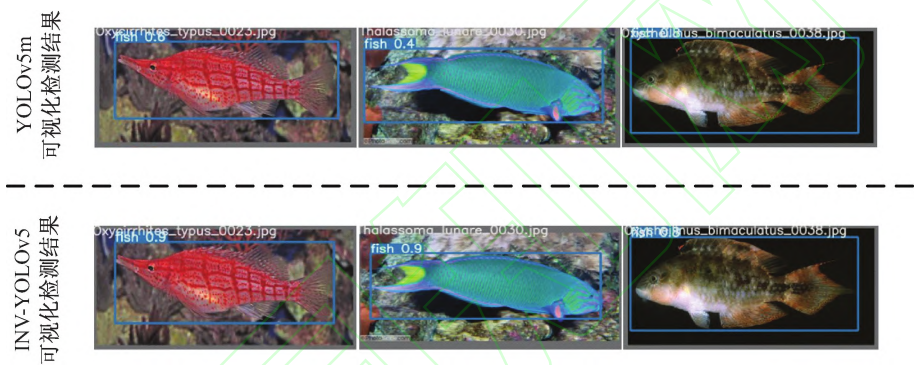


图 12 WildFish 数据集可视化结果

权的 Weightconcat(加权的 Concat),给特征图不同的权重然后再合并在一起,能够丰富特征图的信息,对于模型检测的 mAp 提高 1.7%.实验表明,对于模型检测小目标鱼群,所提方法都改善了对小目标的检测效果.

表 6 INV-YOLOv5 消融实验

方法	Conv	Involution	多头注意力机制	mAp/%
YOLOv5m	×	×	×	69.1
YOLOv5m-A(YOLOv5m+Conv)	√	×	×	73.9
YOLOv5m-B(YOLOv5m-A+C3Inv)	√	√	×	76.2
YOLOv5m-C(YOLOv5m-B+C3TR)	√	√	√	81.9
INV-YOLOv5(YOLOv5m+Conv+C3Inv+C3TR+Weightconcat)	√	√	√	83.6

4 结束语

INV-YOLOv5 模型在 Labeled Fishes in the Wild 数据集和 WildFish 数据集上实验验证表明,该方法的平均精度(mAP)分别为 81.7%和 83.6%,通过综合分析认为可以实现对小鱼群目标的检测,不仅拥有较高的识别率并且更加轻量化,模型大小与 YOLOv5m 网络相比减少了 6 M(Mega)左右,达到了在复杂环境下对小目标检测的要求.根据分析结果,今后可在水下获取更多小目标鱼群影像,扩充数据集,进行补充训练,进一步提高识别率.

5 参考文献

[1] ZHANG N, DONAHUE J, GIRSHICK R, et al. Part-based R-CNNs for fine-grained category detection[C]//European conference on computer vision. Springer, Cham, 2014: 834-849.

- [2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
- [3] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 39(6): 1137-1149.
- [4] DAI J, LI Y, HE K, et al. R-fcn: object detection via region-based fully convolutional networks[J]. Advances in neural information processing systems, 2016, 29.
- [5] 袁红春, 张硕. 基于 Faster R-CNN 和图像增强的水下鱼类目标检测方法[J]. 大连海洋大学学报, 2020, 35(4): 612-619.
- [6] LI X, TANG Y, GAO T. Deep but lightweight neural networks for fish detection[C]//OCEANS 2017-Aberdeen. IEEE, 2017: 1-5.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
- [8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
- [9] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.
- [11] XU W, MATZNER S. Underwater fish detection using deep learning for water power applications[C]//2018 International conference on computational science and computational intelligence (CSCI). IEEE, 2018: 313-318.
- [12] SUNG M, YU S C, GIRDHAR Y. Vision based real-time fish detection using convolutional neural network[C]//OCEANS 2017-Aberdeen. IEEE, 2017: 1-6.
- [13] 王文成, 蒋慧, 乔倩, 等. 基于深度学习的鱼类识别与检测的算法研究[J]. 信息技术与网络安全, 2020, 39(8): 57-61.
- [14] WU D, LV S, JIANG M, et al. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments[J]. Computers and Electronics in Agriculture, 2020, 178: 105742.
- [15] 谈世磊, 别雄波, 卢功林, 等. 基于 YOLOv5 网络模型的人员口罩佩戴实时检测[J]. 激光杂志, 2021, 42(2): 147-150.
- [16] 刘文婷, 卢新明. 基于计算机视觉的 Transformer 研究进展[J]. 计算机工程与应用, 2022, 58(6): 1-16.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv:1706.03762v5.
- [18] LI D, HU J, WANG C, et al. Involution: Inverting the inheritance of convolution for visual recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 12321-12330.
- [19] CUTTER G, STIERHOFF K, ZENG J. Automated detection of rockfish in unconstrained underwater videos using Haar cascades and a new image dataset: labeled fishes in the wild[C]//2015 IEEE Winter Applications and Computer Vision Workshops. IEEE, 2015: 57-62.
- [20] 沈军宇. 基于深度学习的鱼群检测方法研究[D]. 苏州科技大学, 2019.
- [21] ZHUANG P, WANG Y, QIAO Y. Wildfish: a large benchmark for fish recognition in the wild[C]//Proceedings of the 26th ACM international conference on Multimedia, 2018: 1301-1309.

(责任编辑 江津)