

基于图像增强与 GC-YOLO v5s 的水下环境 河蟹识别轻量化模型研究

张 铮 鲁 祥 胡庆松

(上海海洋大学工程学院, 上海 201306)

摘要: 利用机器视觉技术识别水下河蟹目标是实现河蟹养殖装备智能化的有效途径之一。针对水下环境目标识别困难、河蟹包含特征信息少、主流的目标检测模型复杂度高等问题,在 YOLO v5s 的基础上提出了一种适用于水下环境的轻量级河蟹识别模型 GC-YOLO v5s (GhostNetV2-CBAM-YOLO v5s)。利用改进的图像增强算法对水下河蟹图像进行预处理以改善其质量;为降低模型复杂度,提出了基于 GhostNetV2 的 G3 模块以改进模型的特征提取网络,并利用幻影卷积进一步轻量化模型;为了优化模型的河蟹特征学习能力,在 Neck 层和 Head 层之间引入卷积块注意力模块 (Convolution block attention module, CBAM)。实验结果表明,该模型测试集的平均精度均值 (Mean average precision, mAP)、召回率和精确率分别为 95.61%、97.03% 和 96.94%,较 YOLO v5s 分别提升 2.80、2.25、2.28 个百分点;而 GC-YOLO v5s 的参数数量、浮点运算量和模型内存占用量仅为 YOLO v5s 的 69.1%、56.3% 和 58.3%。通过实验对比,该模型在识别精度和模型复杂度上优于其他主流目标检测模型;识别速度仅次于 YOLO v5s,可达到 104 f/s。

关键词: 水产养殖; 河蟹识别模型; 图像增强; YOLO v5s; 轻量化

中图分类号: TN98

文献标识码: A

文章编号: 1000-1298(2024)11-0124-08

OSID:



Lightweight Model for River Crab Detection Based on Image Enhancement and Improved YOLO v5s

ZHANG Zheng LU Xiang HU Qingsong

(College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China)

Abstract: Using machine vision technology to identify underwater crab targets is one of the effective ways to achieve intelligent crab farming equipment. However, river crab detection methods face challenges in the difficulty of target detection in underwater environments, limited feature information and high complexity of mainstream target detection models. To solve these challenges, a lightweight river crab detection model GC-YOLO v5s (GhostNetV2-CBAM-YOLO v5s) was proposed. These specific enhancements were as follows: an improved image enhancement algorithm was used to preprocess underwater crab images to improve the detection accuracy; in order to reduce model complexity, a G3 module based on GhostNetV2 was proposed to improve the feature extraction network of the model, and Ghost convolution was used to further lightweight the model; the convolution block attention module (CBAM) was introduced to solve the challenge of extracting deep features within underwater environments, which were integrated into the feature extraction network. The experimental results demonstrated the improved model's mAP₅₀, recall, and precision on the test set, reaching 95.61%, 97.03% and 96.94%, respectively. These metrics displayed enhancements of 2.80 percentage points, 2.25 percentage points and 2.28 percentage points compared with the baseline. Moreover, GC-YOLO v5s' parameters, computations, and model size were only 69.1%, 56.3%, and 58.3% of YOLO v5s respectively. Comparative trials against mainstream object detection algorithms showcased the superiority in accuracy and model complexity. While slightly trailing YOLO v5s in detect speed, GC-YOLO achieved 104 f/s.

Key words: aquaculture; river crab detection model; image enhancement; YOLO v5s; lightweight

收稿日期: 2024-01-11 修回日期: 2024-03-06

基金项目: 上海市崇明区农业科创项目(2021CNKC-05-06)和上海市水产动物良种创制与绿色养殖协同创新中心项目(2021科技02-12)

作者简介: 张铮(1981—),男,副教授,博士,主要从事物联网工程 and 智能仪器设计研究, E-mail: z-zhang@shou.edu.cn

通信作者: 胡庆松(1979—),男,教授,博士,主要从事渔业工程装备研究, E-mail: qshu@shou.edu.cn

0 引言

目前,我国河蟹养殖产量约 8×10^5 t,总产值将近千亿元,已成为我国水产养殖中发展迅速、具有前景的支柱产业^[1]。智能化装备可有效降低河蟹养殖的劳动强度,利用机器视觉技术识别水下养殖目标对推动养殖装备的智能化具有重要意义^[2-4]。

赵德安等^[5]为探测河蟹在养殖塘中的数量及分布情况,提出一种基于计算机视觉的水下河蟹识别方法。该方法通过优化的 Retinex 算法提高图像的对比度,并利用 YOLO v3 模型对增强后的图像进行识别,在测试集上的平均精度为 86.42%,识别速度达到 10.67 f/s。CAO 等^[6]应用形态成分分析、Retinex 算法、暗通道先验算法和自适应直方图均衡化方法对水下图像进行预处理,利用 MobileNetV2 作为 SSD 的主干网络并在特征提取层中引入了特征金字塔结构,提升对小目标河蟹的识别能力,mAP 和 F1 值分别为 99.01% 和 98.94%。熊海涛等^[7]提出基于 Faster R-CNN 的水下海参目标检测算法,采用 Swin Transformer 作为特征提取网络,同时融入多尺度特征提取层和实例分割功能,提高算法的自适应特征融合能力。JI 等^[8]针对水下河蟹图像模糊、亮度低的特点,提出基于多尺度金字塔融合图像增强和 MobileCenterNet 模型的河蟹目标检测方法,采用改进的 MobileNetv2 主干网络和坐标注意力模块进行特征提取,设计特征融合模块提取多尺度特征图信息,并引入空洞空间卷积池化金字塔模块来融合不同感受野的上下文信息。改进后模型的 mAP 和 F1 值分别为 97.86% 和 97.94%,模型内存占用量仅为 24.46 MB,检测速度达到 48.18 f/s。

上述河蟹识别方法均有一定的应用价值,但未考虑水下环境可能存在河蟹特征相似物等问题,另外,其模型参数量和模型内存占用量大,难以部署到资源受限的边缘设备。YOLO 是一种实时目标识别模型,通过全局推理来获得更多的上下文信息,实现了识别速度和检测精度的良好平衡^[9-13]。本研究提出一种适用于水下环境的轻量级河蟹识别模型 GC-YOLO v5s。利用改进的图像增强算法对水下河蟹图像进行预处理。在 YOLO v5s 的基础上,引入 G3 模块和利用幻影卷积实现模型轻量化。针对水下环境中存在类似河蟹的背景干扰物,在 Neck 层与 Head 层之间嵌入 CBAM 模块增强模型学习重要河蟹特征的位置与信息的能力,以提高对河蟹的识别准确度。

1 数据集构建

1.1 数据采集

水下河蟹图像采集于上海崇明裕安养殖基地。采集时间为 2022 年 7—9 月,采集设备为 SK5Z-P5X10 型水下摄像机,帧率为 60 f/s,分辨率为 1 920 像素×1 080 像素。为提升数据集的多样性,增强模型泛化能力和鲁棒性,本文选取 70% 的真实水下河蟹图像和 30% 的网络河蟹图像组成原始数据集,共 1 200 幅图像。

1.2 图像增强

1.2.1 图像增强原理

JAFFE^[14]提出的水下光学成像模型表明,在水下相机捕获到的光源中,前向散射分量导致水下河蟹图像细节特征模糊,后向散射分量则使采集图像对比度低。水下光线波长越长,衰减速度越快,图像色偏越严重。图像增强通过提高原始图像质量来提升识别精度^[15-16],且不增加模型大小。针对水下河蟹图像存在色偏、对比度低和细节模糊等问题,本文提出一种改进的水下图像增强方法。其预处理步骤为:针对存在色偏的水下河蟹图像采用白平衡,使图像色彩得到还原;再通过 Gamma 校正,调整图像对比度和亮度。而此时图像缺少细节特征,因此利用暗通道先验的方法实现去雾。

1.2.2 图像增强结果

本文将图像按照顺序图像增强处理,对比主流图像增强算法——MSR 算法^[17]、SSR 算法^[18]。各算法图像增强效果如图 1 所示。由图 1a~1d 可以看出,本文提出的顺序图像增强算法的图像质量得到了较大的提升。由图 1e 可以看出,经过 SSR 算法

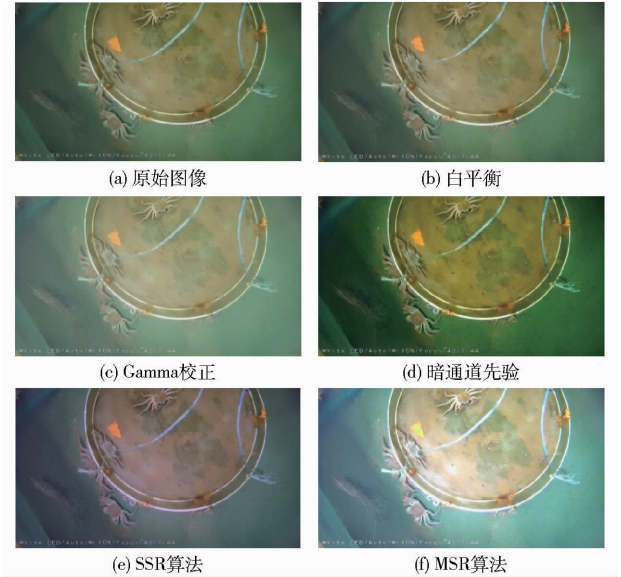


图 1 不同图像增强方法对比

Fig. 1 Comparison of different image enhancement methods

处理的图像,整体对比度提升但存在严重的色偏。由图 1f 观察得到,经过 MSR 算法处理的原始图像,色偏和清晰度得到改善但细节特征丢失。

为了客观评价图像增强后的效果,引入峰值信噪比 (PSNR)^[19] 和水下图像质量评价指标 (UIQM)^[20]。PSNR 是用来衡量图像失真情况的指标,PSNR 值越高表示图像失真越小。UIQM 指标会综合对图像色彩、清晰度和对比度做出评价 (计算得到的指标值越大说明图像质量越好)。不同图像增强方法评价结果如表 1 所示,经过顺序图像增强后,PSNR 和 UIQM 有提升,说明该方法的有效性,且增强后图像评价指标优于 SSR 算法和 MSR 算法。

表 1 不同图像增强方法评价结果

Tab.1 Evaluation of results from different image enhancement methods

图像增强方法	PSNR/dB	UIQM
原始图像	22.456 1	2.065 9
白平衡	23.949 8	2.326 8
Gamma 校正	24.782 7	2.945 7
暗通道先验	26.825 0	3.326 5
SSR 算法	25.925 0	3.122 5
MSR 算法	26.011 0	3.133 1

1.3 数据集构建与标注

将图像增强后的原始数据集划分为训练集 (80%)、验证集 (10%) 和测试集 (10%)。数据增强可以提升模型鲁棒性和泛化能力,随机对划分好的数据集进行旋转、平移、加适量噪声、翻转等方式进行数据增强。数据增强后的训练集、验证集和测试集分别为 2 880、360、360 幅。本文使用 Labellmg 将图像中的河蟹目标标注为 crab。但是河蟹具有 1 对蟹螯和 4 对蟹脚,如果全部标注导致标注区域包含较多背景,背景中存在的大量无用特征会影响模型性能。因此,本文对于图像中的河蟹标注原则为:如果河蟹的足、螯张开导致标注区域内包括大量无关背景,则只标注躯体部分的外接矩形。其他情况根据河蟹足和螯的外接矩形进行标注,标注文件为 json 格式。YOLO v5 所需训练文件格式为 .txt,另外编写格式转换程序将标注文件格式进行转换。

2 轻量化河蟹识别模型

2.1 改进 YOLO v5 网络结构

YOLO v5 根据模型深度和宽度分为 YOLO v5n、YOLO v5s、YOLO v5m、YOLO v5l 和 YOLO v5x。为了选取合适的目标识别模型,分别使用 2 880 幅河蟹图像在 YOLO v5 的 5 个版本上进行训练。由训练结果可知 YOLO v5s 表现出轻量级的模型复杂

度、良好的识别精度和较快的识别速度。因此本文选用 YOLO v5s 作为基础模型进行如下改进:在 YOLO v5s 的 Neck 和 Head 之间引入 CBAM 注意力模块。设计基于 GhostNetV2 的 G3 模块以改进模型的特征提取网络,并利用幻影卷积进一步降低模型复杂度,以便于部署到边缘设备。GC - YOLO v5s 网络结构如图 2 所示。

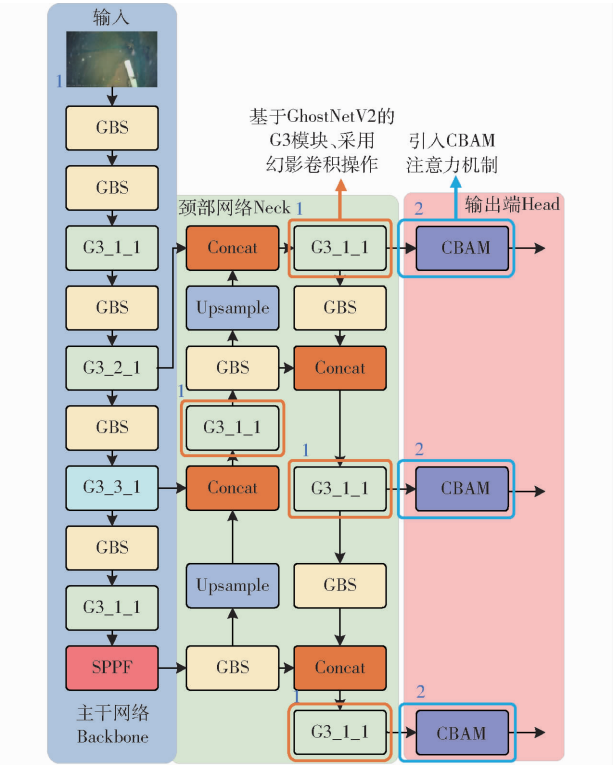


图 2 GC - YOLO v5s 网络结构图
Fig.2 GC - YOLO v5s structure diagram

2.1.1 CBAM 注意力模块

如图 3 所示,CBAM^[21] 由通道注意力模块 (Channel attention module,CAM)^[22] 和空间注意力模块 (Spatial attention module,SAM)^[23] 组成,旨在增强卷积神经网络对特征的学习能力。CAM 利用平均池化和最大池化来聚合河蟹特征图中的有用空间信息,再将得到的特征图输入到多层感知机 (MLP) 生成关键特征,最后利用激活函数获得通道注意力权重。计算公式为

$$M_c(F') = \sigma(MLP(AvgPool(F))) + MLP(MaxPool(F))$$

式中 $M_c(F')$ ——通道注意力权重
 F' ——中间特征图
 σ ——Sigmoid 函数
AvgPool——平均池化
MaxPool——最大池化
MLP——多层感知机
 F ——输入特征图

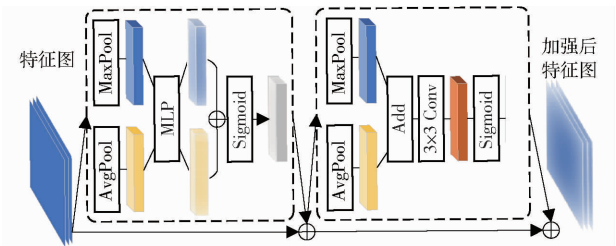


图 3 CBAM 注意力模块

Fig.3 CBAM attention module

SAM 将获得的通道注意力图利用池化来突出河蟹有用特征的信息区域,并通过 3×3 的卷积运算将两个特征进行融合处理得到空间注意力特征图,最后利用 Sigmoid 得到空间注意力权重。计算公式为

$$M_s(F'') = \sigma(f([\text{AvgPool}(F'); \text{MaxPool}(F')])) \quad (2)$$

式中 $M_s(F'')$ ——空间注意力权重
 f —— 3×3 的卷积核
 F'' ——输出特征图

因此为了提高模型在水下环境中能学习到河蟹有用特征的通道信息和空间信息,本文在模型不同位置引入 CBAM 注意力模块,并比较其综合评价指标,实验结果如表 2 所示。由表 2 可知,在 YOLO v5s 的 Neck 和 Head 之间嵌入 CBAM 模块得到的实验结果最佳。

表 2 CBAM 模块嵌入不同位置的实验结果

Tab.2 Experimental results of CBAM module at different locations

模型	mAP ₅₀ /%	浮点运算量	内存占用量/MB	参数量
1	93.02	1.58×10^{11}	13.9	7.04×10^6
2	93.19	1.58×10^{11}	14.1	6.98×10^6
3	93.84	8.90×10^{10}	14.9	7.34×10^6
4	94.96	1.58×10^{11}	14.5	7.05×10^6

注:模型 1 为 CBAM 替换 YOLO v5s 主干网络的常规卷积操作;模型 2 为 CBAM 替换 YOLO v5s 的 G3 模块;模型 3 为 CBAM 替换 YOLO v5s 颈部网络的 G3 模块;模型 4 为 CBAM 分别嵌入 YOLO v5s 的 3 个 Neck 和 Head 之间。

2.1.2 模型轻量化

为降低模型复杂度,基于 GhostNetV2^[24] 的 G3 模块改进模型主干网络 Backbone 和颈部网络 Neck,并利用幻影卷积进一步轻量化模型。GhostNetV2 对比 GhostNet^[25] 和 MobileNet 等轻量化卷积神经网络,采用幻影卷积模块和 DFC 注意力模块,在保证轻量化的前提下产生更多冗余特征图,学习特征图上的空间信息。幻影卷积模块通过常规卷积减少输入特征图的通道数,再采用分组线性操作生成大量冗余特征图,最后将常规

卷积生成的特征图和冗余特征图合并,幻影卷积的结构如图 4 所示。

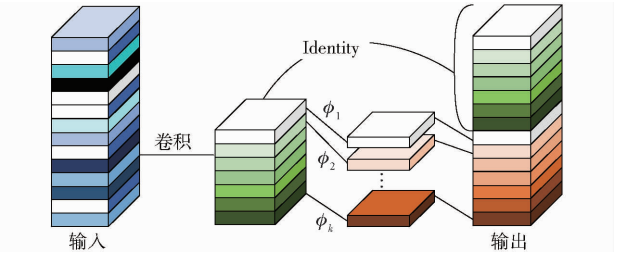


图 4 幻影卷积模块结构图

Fig.4 Ghost module

采用幻影卷积操作可以有效降低模型复杂度,但是模型的特征表示能力由于卷积操作仅能学习局部特征信息而下降。由图 5a 可知,DFC 注意力模块最初通过对输入图像进行下采样,目的是为了减少计算复杂度,当下采样为输入图像的一半尺寸时可以节省 75% 的计算量;其次采用 1×1 卷积来对下采样后的特征图进行通道扩展,为了匹配 Ghost 模块生成的冗余特征图;DFC 注意力模块将全连接层分解为水平全连接层和垂直全连接层,并分别赋予一个变换权重,再通过水平和垂直方向上对二维特征图进行特征聚合,生成更大的全局感受野;最后使用 Sigmoid 函数将生成的注意力特征图映射到 $(0,1)$ 之间。

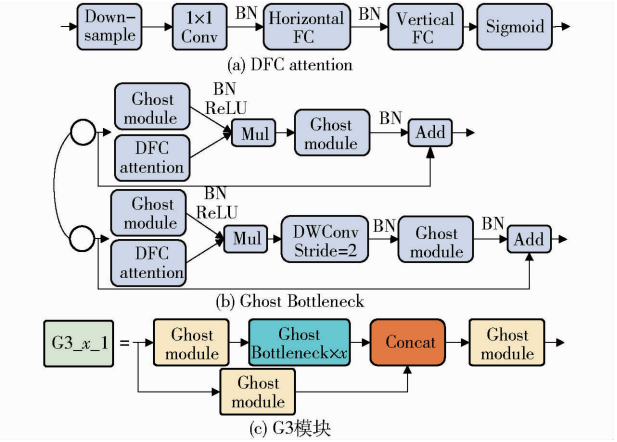


图 5 模型轻量化改进结构图

Fig.5 Model lightweight improvement structure diagram

基于 GhostNetV2 和 CSPNet^[26] 结构设计 Ghost Bottleneck 模块和 G3 模块,并引入幻影卷积和深度可分离卷积来降低计算量,Ghost Bottleneck 模块和 G3 模块如图 5b、5c 所示。G3 模块通过并行 Ghost 模块和 DFC 注意力模块来降低模型参数量、计算量和提升模型全局特征表示能力,将 Ghost 模块和 DFC 注意力模块这两个分支进行点乘,再采用一个 Ghost 模块来降低网络通道数,匹配短连接。如图 5c 可知,当 G3 模块步长为 2 时,引入一个步长为 2 的深度可分离卷积来提取特征图。为了加快模型

学习速度和缓解梯度消失,在 GhostNetV2 模块引入 BN 层。

2.2 评价指标

针对模型识别精度采用 3 种评价指标:精确率 (Precision, P)、召回率 (Recall, R)、平均精度 (mAP_{50});针对识别速度采用指标帧率 (Frame per second),帧率达到 25 f/s 可以实时识别;衡量模型轻量级程度采用参数量、浮点运算量、模型内存占用量这 3 种指标。

2.3 训练环境与超参数设置

实验软硬件配置为:CPU 为 Intel(R) Xeon(R)-W2223@ 3.60 GHz,GPU 为 NVIDIA GTX2080Ti,11 GB

显存容量,操作系统为 Win 10,深度学习框架为 Pytorch 1.13.1,Python 版本为 3.8,CUDA 版本为 11.7,Cudnn 版本为 8.0.5。超参数配置为:输入图像分辨率为 640 像素×640 像素,批处理大小为 16,最大学习率为 0.01,并逐渐下降,动量设置为 0.937,优化器为随机梯度下降,训练轮数为 500 轮。

3 识别结果与分析

3.1 模型消融实验

为了验证改进模型性能,对每一个改进策略做消融实验。在同一数据集、训练环境、超参数下进行,消融实验结果如表 3 所示。

表 3 消融实验结果

Tab.3 Ablation experiment results

模型	参数量	浮点运算量	模型内存占用量/MB	精确率/%	召回率/%	$mAP_{50}/\%$
YOLO v5s	7.01×10^6	1.58×10^{11}	14.4	94.66	94.78	92.81
YOLO v5s-CBAM	7.05×10^6	1.58×10^{11}	14.5	96.04	95.23	94.96
YOLO v5s-G3	4.80×10^6	8.90×10^{10}	8.3	95.83	95.14	93.72
GC-YOLO v5s	4.85×10^6	8.90×10^{10}	8.4	96.85	96.19	94.89

由表 3 可知,在 YOLO v5s 的 Neck 和 Head 之间嵌入 CBAM 注意力模块在保证模型复杂度基本不变的情况下, P 、 R 、 $mAP50$ 分别提升 1.38、0.45、2.15 个百分点。在 YOLO v5s 上嵌入 CBAM 取得检测精度提升的因素是:CBAM 作为一种混合注意力模块,其继承了 CAM 和 SAM 二者的优势,学习特征图的空间信息和通道信息。因此改进后的模型可以在水下环境中学习到河蟹目标的重要特征并抑制背景这类无用特征的提取。虽然 CBAM 是一种轻量级的模块,但是也会增加模型的参数量和模型内存占用量。本文引入 G3 模块和幻影卷积,这一策略使得模型识别精度有所上升的同时,模型参数量、浮点运算量和模型内存占用量大幅度下降,主要原因是 GhostNetV2 通过二次轻量级的卷积操作产生更多的特征图,且并行采用 DFC 注意力模块,在不增加计算开销的前提下获得比纯卷积操作更大的感受野。

综合消融实验结果可知,采用的 CBAM 和 GhostNetV2 这两处改进对模型的识别精度和轻量化产生积极作用。在模型识别精度上 GC-YOLO v5s 较 YOLO v5s 在 P 、 R 、 $mAP50$ 分别提升 2.19、1.41、2.08 个百分点;在模型复杂度方面,参数量、浮点运算量和模型内存占用量分别仅为 YOLO v5s 的 69.1%、56.3% 和 58.3%。因此,GC-YOLO 是一个既能精确识别河蟹目标又轻量化的模型。

3.2 模型改进前后识别结果对比

为了验证 GC-YOLO 对不同场景的适应力,除利用原有测试集外,还利用在浦东滨海养殖基地采

集的视频数据进行性能验证。图 6 为 YOLO v5s 和 GC-YOLO v5s 对两种场景下的河蟹识别结果。其中,图 6 的第 1 行图是 GC-YOLO v5s 的识别结果,第 2 行是 YOLO v5s 的识别结果。图 6a~6c、图 6e~6g 为原有测试集的识别结果,图 6d、6h 为滨海养殖基地视频数据的识别结果。从识别结果可知,YOLO v5s 对两种场景的河蟹识别存在明显的漏检和误检(漏检用紫色框标注,误检用黑色框标注),预测框的置信度不高,最高为图 6f 中的 0.85,最低为图 6a 中的 0.49。图 6 中 GC-YOLO v5s 对两种场景下的河蟹的平均置信度高于 0.9,预测框可以准确地定位水下河蟹目标。这是因为 GC-YOLO v5s 采用 G3 模块和幻影卷积操作,使得模型可以获取丰富的冗余特征图供 GC-YOLO v5s 进行特征提取。DFC 注意力模块使得模型在丰富冗余特征图上学习到更多的全局信息。此外,在 GC-YOLO v5s 的颈部网络和头部网络之间引入 CBAM 注意力模块,通过增强特征图中有用河蟹特征,抑制无关干扰物特征,使得模型更加关注河蟹的位置和形状,改善了干扰物的影响,提升了 GC-YOLO v5s 在不同场景下的河蟹识别能力。

3.3 图像增强算法有效性实验

为了进一步验证本文所设计的图像增强算法对 GC-YOLO v5s 的有效性,选取两个场景的图像,分别进行图像增强前后的模型识别对比实验。对比实验结果如表 4 所示,YOLO v5s 和 GC-YOLO v5s 的 mAP_{50} 分别提升 0.71 个百分点和 0.72 个百分点,召

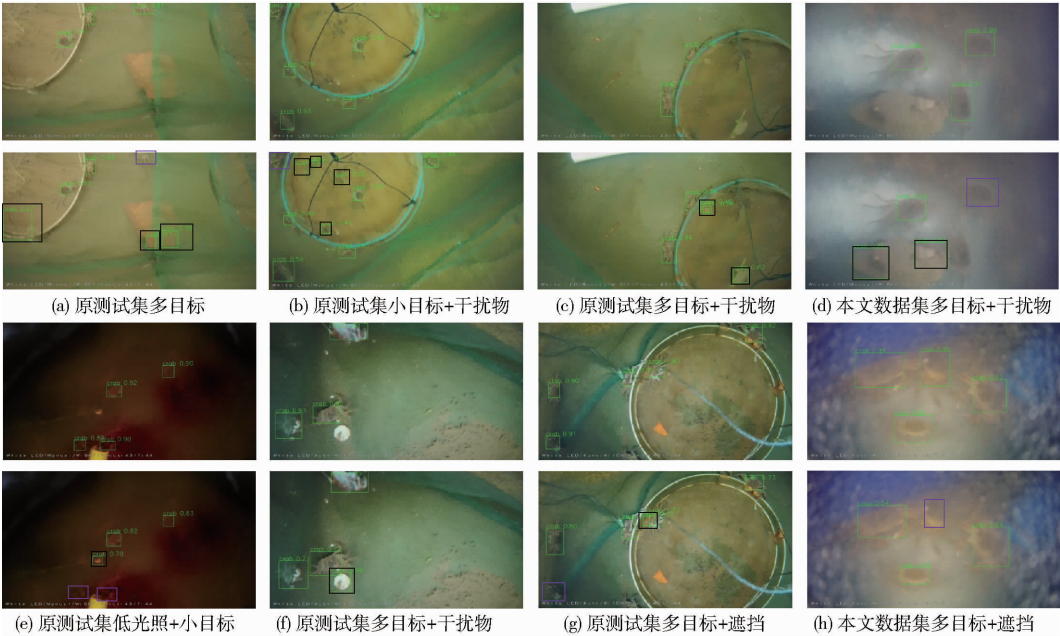


图 6 模型识别结果可视化

Fig. 6 Visualization of model detect results

回率分别提升 0.69 个百分点和 0.84 个百分点,精确率均略有提升。

表 4 图像增强对比实验结果

模型	<i>P</i>	<i>R</i>	<i>mAP</i> ₅₀
YOLO v5s(无图像增强)	94.66	94.78	92.81
YOLO v5s(有图像增强)	94.83	95.47	93.52
GC-YOLO v5s(无图像增强)	96.85	96.19	94.89
GC-YOLO v5s(有图像增强)	96.94	97.03	95.61

识别结果如图 7 所示,其中第 1 行是图像增强前识别结果,第 2 行是图像增强后识别结果。在图 7a 上幅图中,YOLO v5s 对原始图像中的水下河蟹识别分别存在 1 处漏检和 1 处误检(漏检用紫色框标注,误检用黑色框标注),置信度不高,而图像增强后的置信度达到 0.76 和 0.89。在图 7b 中,图像增强前置信度分别为 0.88 和 0.87,而经过图像增强后 GC-YOLO v5s 的置信度达到 0.93 和 0.94。结果表明该方法能够有效提高识别模型的性能。

3.4 不同注意力模块对比实验

为了对比 CBAM 注意力模块、通道注意力模块和空间注意力模块在养殖场景下的性能,随机挑选测试集中的一幅河蟹图像,通过 Grad-CAM 生成三者的可视化热力图,如图 8 所示。颜色越接近蓝色表示特征占比越低;颜色越红表示特征占比越高。特征占比越高表明对河蟹识别发挥着更重要的作用。图 8 中使用 CBAM 注意力模块红色占比更高,且精准位于河蟹的背部特征处,因为 CBAM 实现了通道注意力模块和空间注意力模块

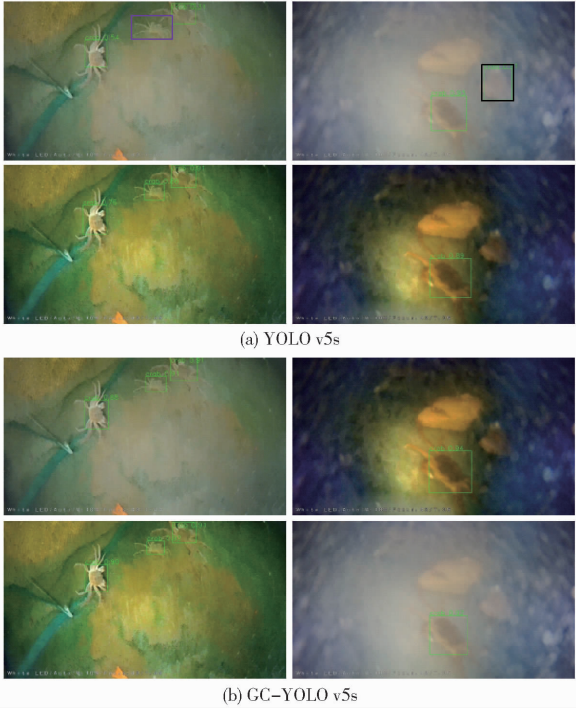


图 7 图像增强后模型识别结果对比

Fig. 7 Comparison of detection results of models after image enhancement

互补,通过学习通道和空间注意力权重,提高了水下环境中有用河蟹细节特征的学习能力。

为了验证其他注意力模块在河蟹识别中的可行性,本文对采用主流注意力模块 SE、CA、ECA 和 CBAM 后的 YOLO v5s 与 YOLO v5s 进行对比,对比实验结果如表 5 所示。由表 5 可知,YOLO v5s 在采用不同注意力模块后,每个模型性能均有一定程度的提高,其中 CBAM 注意力模块检测精度提升最

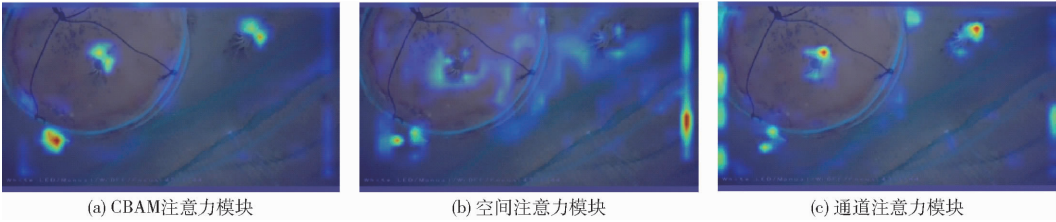


图 8 热力图对比

Fig. 8 Comparison of heat map effect

表 5 不同注意力模块对比实验结果

Tab. 5 Comparative experimental results of different attention mechanisms			%
模型	<i>P</i>	<i>R</i>	mAP ₅₀
YOLO v5s	94. 66	94. 78	92. 81
YOLO v5s - CA	95. 01	95. 12	93. 21
YOLO v5s - SE	95. 74	95. 09	93. 95
YOLO v5s - ECA	95. 73	95. 11	94. 23
YOLO v5s - CBAM	96. 04	95. 23	94. 96

高。由于 SE 注重特征图中通道间的重要性,ECA 和 CA 更侧重于有用通道特征内的相关性。这三者忽略了特征图中的空间特征,导致对水下环境不同姿态河蟹的识别能力较差,CBAM 注意力模块能够学习到不同尺度的河蟹特征。

综上所述,在 YOLO v5s 中引入 CBAM 注意力模块是一种改善水下环境河蟹目标识别精度的最有效改进策略。

3.5 不同模型对比实验

为了进一步验证 GC - YOLO v5s 的先进性,选取 SSD、YOLO v4、YOLO v7 - tiny 和 Faster R - CNN 算法进行对比实验,结果如表 6 所示。Faster R -

CNN 在识别精度、模型复杂度和识别速度方面表现的并不突出,特别是浮点运算量高,达 9.47×10^{11} ,且 mAP₅₀最低。SSD 算法和 YOLO v4 较 Faster R - CNN 虽然在平均精度上有一定的提升,但是这两者模型复杂度过高,很难在资源受限的嵌入式设备上 进行部署。YOLO v7 - tiny 较 YOLO v5s 在参数量、浮点运算量和模型内存占用量有一定幅度的下降,但是识别精度和识别速度还不够平衡。而 GC - YOLO v5s 参数量、浮点运算量和模型内存占用量均为最低,分别为 4.85×10^6 、 8.9×10^9 和 8.4 MB,更易于部署到边缘计算端。此外,由表 6 可知,GC - YOLO v5s 识别速度仅次于 YOLO v5s。GC - YOLO v5s 在平均精度指标上优于对其它模型,其平均精度可以达到 95.61%,相较于 SSD、Faster R - CNN、YOLO v4、YOLO v7 - tiny,分别提高 16.64、22.27、24.32、16.13 个百分点。因为 GhostNetV2 可以通过两阶段的卷积操作学习到更多特征图和全局信息,在降低模型复杂度的同时提升识别精度。另外,引入 CBAM 注意力模块可以快速定位河蟹有用特征的位置区域和深层信息。

表 6 先进模型对比结果

Tab. 6 Comparison of SOTA models results

模型	参数量	浮点运算量	模型内存占用量/MB	mAP ₅₀ /%	帧率/(f·s ⁻¹)
YOLO v4	5.25×10^7	1.20×10^{11}	2.563×10^4	71. 29	55
YOLO v5s	7.01×10^6	1.58×10^{10}	14. 4	93. 52	114
Faster R - CNN	2.81×10^7	9.47×10^{11}	108	73. 34	12
GC - YOLO v5s	4.85×10^6	8.90×10^9	8. 4	95. 61	104
SSD	2.36×10^7	2.74×10^{11}	90. 6	78. 97	39
YOLO v7 - tiny	6.02×10^6	1.30×10^{10}	11. 5	79. 48	88

不同模型识别河蟹可视化结果如图 9 所示。上面 2 行为原有测试集识别结果,下面 2 行为浦东滨海养殖基地视频数据的识别结果。由结果可知,GC - YOLO v5s 模型置信度优于其它模型,最高达到 0.94。在有被遮挡的河蟹和卵石等类似背景干扰物的情况下,未出现漏检和误检问题。YOLO v5s 模型虽然识别置信度较高,但存在漏检误检的情况,总共存在 6 处误检和 1 处漏检(漏检用紫色框标注,误检用黑色框标注)。这是因为 YOLO v5s 判别相似特

征的能力较差,未能有效地学习水下河蟹特征图中的多尺度特征。Faster R - CNN 的水下河蟹识别共存在 3 处误检和 5 处漏检,而 YOLO v4 和 SSD 存在多处的误检和漏检问题,说明它们易受类似河蟹的背景干扰物影响,造成模型的识别精度下降。YOLO v7 - tiny 虽然可以较准确地定位水下河蟹位置,但置信度不高,最高为 0.7,并且容易将类似背景特征的河蟹和被遮挡的河蟹漏检,其对水下环境中河蟹特征的学习能力较差。

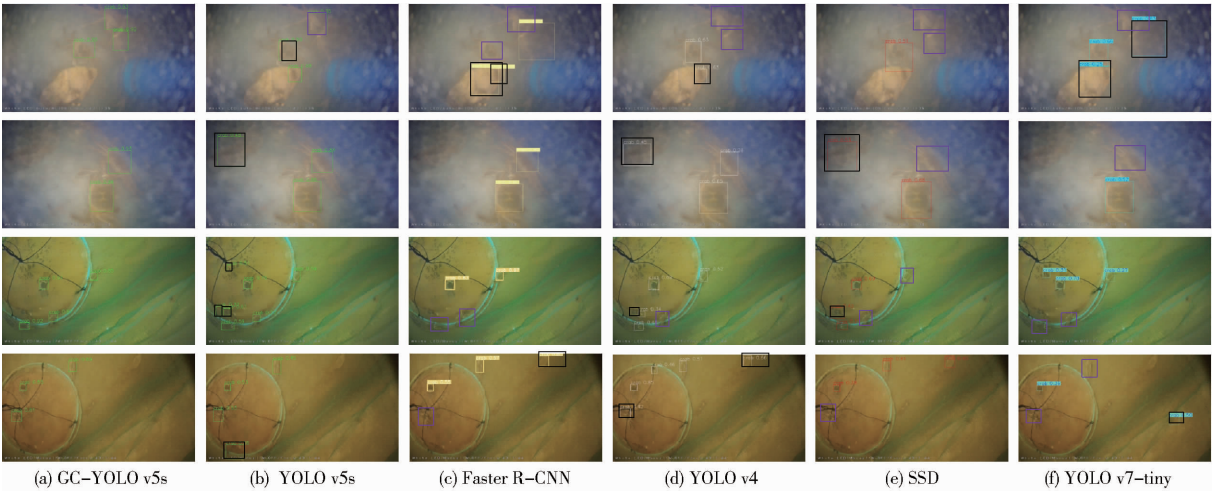


图 9 不同模型识别结果可视化

Fig.9 Visualization of different model detect results

4 边缘设备部署

为验证 GC-YOLO v5s 模型是否满足部署在边缘设备上实时识别的性能要求,将 GC-YOLO v5s 模型训练后的 best. pt 文件转换为 onnx 模型,再转换为适用于边缘设备的 RKNN 模型,对转换后的模型进行量化处理,为了降低模型识别时需要的计算量和内存,会降低一定的精度损失,最后将量化后的 RKNN 模型部署到边缘设备上。LubanCat-4 是基于 RockChip RK3588s 设计的一款低功耗、高性能的边缘设备,内存容量 8 GB,存储器 128 GB。将 GC-YOLO v5s 和 YOLO v5s 都部署在 LubanCat-4 上,并利用浦东滨海养殖基地的视频数据进行测试,测试结果如表 7 所示。经过量化处理后的 YOLO v5s 和 GC-YOLO v5s 模型内存占用量分别为 8.1 MB 和 2.7 MB,但是识别精度有一定的下降。GC-YOLO v5s 识别速度为 28 f/s,说明该模型中基于 GhostNetV2 设计的 G3 模块优化了特征提取网络,改善了识别的实时性能。由于 CBAM 注意力模块的引入,GC-YOLO v5s 能够精准定位有效河蟹特征,其在边缘设备上的 mAP₅₀较 YOLO v5s 有明显提高。

表 7 边缘设备部署对比实验结果

Tab.7 Edge device deployment comparison experiment results

模型	模型内存占用量/MB	mAP ₅₀ /%	帧率/(f·s ⁻¹)
YOLO v5s	8.1	81.14	19
GC-YOLO v5s	2.7	88.63	28

GC-YOLO v5s 和 YOLO v5s 部署在 LubanCat-4 上的识别结果如图 10 所示。GC-YOLO v5s 面对水下环境河蟹识别置信度明显提升,达到 0.96。在面对类似蟹形干扰物、模糊背景等情况下均有良好

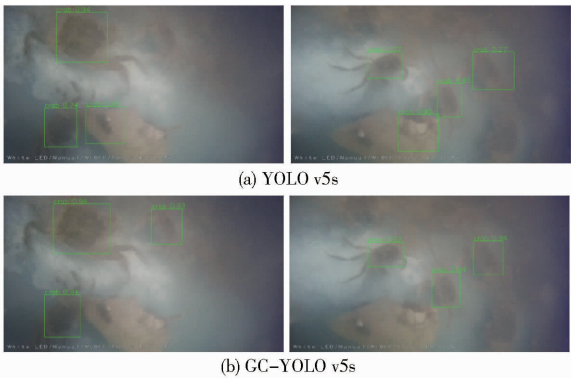


图 10 改进前后模型在 LubanCat-4 上的识别结果对比
Fig.10 Comparison of detection results of improved model on LubanCat-4

的识别效果。

5 结论

(1)对比改进前后的模型,GC-YOLO v5s 平均精度提升 2.80 个百分点,且其参数量、浮点运算量和模型内存占用量分别仅为 YOLO v5s 的 69.1%、56.3% 和 58.3%。

(2)对比实验结果表明,GC-YOLO v5s 在识别精度和模型轻量化上均优于 SSD、Faster R-CNN、YOLO 系列,识别速度达到 104 f/s。

(3)为了进一步验证本文所提出的方法在不同场景下的适应力,除利用原有测试集外,还利用其它水下场景采集的河蟹视频数据进行测试。GC-YOLO v5s 在两种场景下均优于对比模型。

(4)利用采集的水下河蟹视频数据进行边缘设备端模型部署测试。由测试结果可知,GC-YOLO v5s 识别速度达到 28 f/s,满足实时识别条件,经过量化后的 RKNN 模型内存占用量仅为 2.7 MB, mAP₅₀较 YOLO v5s 提升 7.49 百分点。

coefficient and decomposition[J]. Resources and Environment in the Yangtze Basin, 2023, 32(2): 235 – 249. (in Chinese)

[40] 张爱国. 基于分组数据基尼系数的一种估算方法[J]. 统计与决策, 2022, 38(2): 10 – 15.
ZHANG Aiguo. An estimation method for Gini coefficient based on grouped data[J]. Statistics and Decision, 2022, 38(2): 10 – 15. (in Chinese)

[41] 刘精明. 教育扩张与分布型教育不平等——复合教育基尼系数的演化性质及其检验[J]. 社会学研究, 2023, 38(1): 68 – 95.
LIU Jingming. Education expansion and distributive inequality in education—the evolutionary properties and empirical tests of the composite Gini coefficient of education[J]. Sociological Studies, 2023, 38(1): 68 – 95. (in Chinese)

[42] 王金萍, 方波. 世界基尼系数计算方法的理论比较与实证[J]. 统计与决策, 2023, 39(23): 5 – 10.
WANG Jinping, FANG Bo. Theoretical comparison and empirical research of calculation methods of the world Gini coefficient [J]. Statistics and Decision, 2023, 39(23): 5 – 10. (in Chinese)

(上接第 131 页)

参 考 文 献

[1] 农业农村部渔业渔政管理局, 全国水产技术推广总站, 中国水产学会. 2022 中国渔业统计年鉴[M]. 北京: 中国农业出版社, 2022.

[2] 赵德安, 曹硕, 孙月平, 等. 基于联动扩展神经网络的水下自由活蟹检测器研究[J]. 农业机械学报, 2020, 51(9): 163 – 174.
ZHAO Dean, CAO Shuo, SUN Yueping, et al. Small-sized efficient detector for underwater freely live crabs based on compound scaling neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(9): 163 – 174. (in Chinese)

[3] ZHANG L, LI B, SUN X, et al. Intelligent fish feeding based on machine vision: a review[J]. Biosystems Engineering, 2023, 231: 133 – 164.

[4] LIU C, WANG Z, LI Y, et al. Research progress of computer vision technology in abnormal fish detection[J]. Aquacultural Engineering, 2023, 103: 102350.

[5] 赵德安, 刘晓洋, 孙月平, 等. 基于机器视觉的水下河蟹识别方法[J]. 农业机械学报, 2019, 50(3): 151 – 158.
ZHAO Dean, LIU Xiaoyang, SUN Yueping, et al. Detection of underwater crabs based on machine vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(3): 151 – 158. (in Chinese)

[6] CAO S, ZHAO D, LIU X, et al. Real-time robust detector for underwater live crabs based on deep learning[J]. Computers and Electronics in Agriculture, 2020, 172: 105339.

[7] 熊海涛, 林琪, 宣魁, 等. 基于改进 Faster R – CNN 的海参目标检测算法[J]. 农业机械学报, 2022, 53(增刊 2): 204 – 209.
XIONG Haitao, LIN Qi, XUAN Kui, et al. Sea cucumber object detection algorithm based on improved Faster R – CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(Supp. 2): 204 – 209. (in Chinese)

[8] JI W, PENG J, XU B, et al. Real-time detection of underwater river crab based on multi-scale pyramid fusion image enhancement and MobileCenterNet model[J]. Computers and Electronics in Agriculture, 2023, 204: 107522.

[9] 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.

[10] REDMON J, FARHADI A. YOLO v3: an incremental improvement[J]. arXiv Preprint, arXiv:1804.02767, 2018.

[11] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLO v4: optimal speed and accuracy of object detection[J]. arXiv Preprint, arXiv:2004.10934, 2020.

[12] LI C, LI L, JIANG H, et al. YOLO v6: a single-stage object detection framework for industrial applications[J]. arXiv Preprint, arXiv:2209.02976, 2022.

[13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLO v7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464 – 7475.

[14] JAFFE J S. Computer modeling and the design of optimal underwater imaging systems[J]. IEEE Journal of Oceanic Engineering, 1990, 15(2): 101 – 111.

[15] LI C, ANWAR S, HOU J, et al. Underwater image enhancement via medium transmission-guided multi-color space embedding[J]. IEEE Transactions on Image Processing, 2021, 30: 4985 – 5000.

[16] 覃学标, 黄冬梅, 宋巍, 等. 模糊水下图像多增强与输出混合的鱼类检测方法[J]. 农业机械学报, 2022, 53(7): 243 – 249.
QIN Xuebiao, HUANG Dongmei, SONG Wei, et al. Fish detection method of multiple enhanced and outputs blend for blurred underwater images[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(7): 243 – 249. (in Chinese)

[17] ZHANG S, WANG T, DONG J, et al. Underwater image enhancement via extended multi-scale Retinex[J]. Neurocomputing, 2017, 245: 1 – 9.

[18] GIANINI G, RIZZI A, DAMIANI E. A retinex model based on absorbing Markov chains[J]. Information Sciences, 2016, 327: 149 – 174.

[19] ATAIE R, EMRANI Z A A, SAFAEI M Y. An efficient inexact full adder cell design in CNFET technology with high-PSNR for image processing[J]. International Journal of Electronics, 2019, 106(6): 928 – 944.

[20] PANETTA K, GAO C, AGAIAN S. Human-visual-system-inspired underwater image quality measures[J]. IEEE Journal of Oceanic Engineering, 2016, 41(3): 541 – 551.

[21] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 – 19.

[22] FU J, ZHENG H, MEI T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4438 – 4446.

[23] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011 – 2023.

[24] TANG Y, HAN K, GUO J, et al. GhostNetv2: enhance cheap operation with long-range attention[J]. Advances in Neural Information Processing Systems, 2022, 35: 9969 – 9982.

[25] HAN K, WANG Y, TIAN Q, et al. Ghostnet: more features from cheap operations[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580 – 1589.

[26] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390 – 391.