

基于SPD-Conv结构和NAM注意力机制的鱼群小目标检测

谌雨章, 王诗琦, 周雯, 周婉婷

引用本文

谌雨章, 王诗琦, 周雯, 周婉婷. 基于SPD-Conv结构和NAM注意力机制的鱼群小目标检测[J]. 计算机科学, 2024, 51(6A): 230500176-7.

CHEN Yuzhang, WANG Shiqi, ZHOU Wen, ZHOU Wanting. Small Object Detection for Fish Based on SPD-Conv and NAM Attention Module [J]. Computer Science, 2024, 51(6A): 230500176-7.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[面向产线AI质检的少样本评测方法研究和验证](#)

Study and Verification on Few-shot Evaluation Methods for AI-based Quality Inspection in Production Lines

计算机科学, 2024, 51(6A): 230700086-8. <https://doi.org/10.11896/jsjcx.230700086>

[融合注意力机制与线激光辅助的输送带缺陷检测网络](#)

Conveyor Belt Defect Detection Network Combining Attention Mechanism with Line Laser Assistance

计算机科学, 2024, 51(6A): 230800115-6. <https://doi.org/10.11896/jsjcx.230800115>

[卷烟厂卷包车间工人违规作业行为检测方法](#)

Detection Method for Workers' Illegal Operation Behavior in Packaging Workshop of Cigarette Factory

计算机科学, 2024, 51(6A): 230700123-8. <https://doi.org/10.11896/jsjcx.230700123>

[基于改进FCOS的遥感图像舰船目标检测](#)

Ships Detection in Remote Sensing Images Based on Improved FCOS

计算机科学, 2024, 51(6A): 230700166-7. <https://doi.org/10.11896/jsjcx.230700166>

[感受野扩展与多分支聚合的目标检测方法](#)

Object Detection with Receptive Field Expansion and Multi-branch Aggregation

计算机科学, 2024, 51(6A): 230600151-6. <https://doi.org/10.11896/jsjcx.230600151>

基于 SPD-Conv 结构和 NAM 注意力机制的鱼群小目标检测

谌雨章 王诗琦 周雯 周婉婷

湖北大学计算机与信息工程学院 武汉 430062

(hubucyz@foxmail.com)

摘要 为解决因水下成像环境退化导致图像分辨率较低,以及因鱼群目标较小等因素导致的检测精度不高的问题,提出了一种结合 SPD-Conv 结构和 NAM 注意力机制的改进 YOLOv7 检测算法。首先,采用 Space-to-Depth(SPD)结构改进头部网络,取代了网络中原有的跨步卷积结构,保留了更多的细粒度信息,提升了特征学习的效率,提高了网络对低分辨率图像的检测效果。然后在网络中引入 Normalization-based Attention Module(NAM)注意力机制,采用 CBAM 的模块集成方式,使用 BN 缩放因子来计算注意力权重,抑制了不显著的特征,提升了小目标检测的准确率。最后针对水下成像退化,对检测图片做反卷积预处理,减小了水下成像退化因素对检测造成的影响。实验结果显示,在 WildFish 数据集上模型的整体精度达到 97.2%,与 YOLOv7 算法相比提升了 7.6%,准确率提升了 8.5%,召回率提升了 9.8%,与 Efficientdet,SSD,YOLOv5 及 YOLOv8 算法相比,所提模型精度分别提升了 12.6%,17.8%,4%及 2.9%。在 Aquarium 数据集上模型的整体精度达到 80.5%,相比 Efficientdet,SSD,YOLOv5,YOLOv7 及 YOLOv8 分别提升了 18.4%,11.6%,6.9%,2.0%及 2.7%,可以满足水下鱼群识别的需求。

关键词: SPD-Conv 结构; NAM 注意力机制; YOLOv7 算法; 鱼群检测; 目标检测

中图分类号 TP391.41; TP18

Small Object Detection for Fish Based on SPD-Conv and NAM Attention Module

CHEN Yuzhang, WANG Shiqi, ZHOU Wen and ZHOU Wanting

School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China

Abstract In order to solve the problem of low image resolution due to the degradation of underwater imaging environment and low detection accuracy caused by small fish targets, an improved YOLOv7 detection algorithm combining SPD-Conv structure and NAM attention mechanism is proposed. Firstly, the space-to-depth(SPD) structure is used to improve the head network, which replaces the original straddle convolution structure in the network, retains more fine-grained information, improves the efficiency of feature learning, and improves the detection effect of the network on low-resolution images. Then, the normalization-based attention module(NAM) attention mechanism is introduced into the network, and the module integration method of CBAM is adopted, and the BN scaling factor is used to calculate the attention weight, which suppresses the insignificant features and improves the accuracy of small target detection. Finally, for underwater imaging degradation, the detection image is deconvolved and preprocessed, which reduces the impact of underwater imaging degradation factors on detection. Experimental results show that in the WildFish dataset, the overall accuracy of the model reaches 97.2%, which is 7.6% higher than that of the YOLOv7 algorithm, the accuracy rate is increased by 8.5%, and the recall rate is increased by 9.8%, compared with the Efficientdet, SSD, YOLOv5 and YOLOv8 algorithms, the accuracy of the proposed model is improved by 12.6%, 17.8%, 4% and 2.9%, respectively. The overall accuracy of the model reaches 80.5%, which is 18.4%, 11.6%, 6.9%, 2.0% and 2.7% higher than that of Efficientdet, SSD, YOLOv5, YOLOv7 and YOLOv8, respectively, which can meet the needs of underwater fish identification.

Keywords Space-to-Depth Conv(SPD Conv), Normalization-based attention module(NAM), YOLOv7, Fish detection, Object detection

1 引言

随着近年来发达国家开展智能化、信息化的渔业生产模式,挪威等欧盟国家逐渐超越中国成为鱼类产品领先市场。而中国尚处于由传统向智能养殖的过渡阶段,整体水产养殖模式从传统的粗放型逐渐转变为集约型,生产结构不断调整升级。因此,利用现代化技术建设中国智慧渔业养殖是一项迫在眉睫的任务。现代化渔业发展的最终目标即实现工厂化养殖,而鱼群目标检测是工业化精准养殖的基础^[1]。时至

今日,研究人员已经在鱼群检测方面做了大量工作。

传统的鱼群检测方法常利用声呐法或轮廓检测法对鱼群进行检测,利用声呐探测收集水下鱼群图像具有一定的抗干扰能力,但在复杂场景下生成图片分辨率较低,导致检测准确度不高,利用轮廓检测法得到的检测精度较高,但鱼群重叠后造成的图片轮廓变化会使检测效果大幅降低^[2]。在实际应用场景中,由于光线、水质、水下环境等原因,水下采集的图片具有很低的分辨率,且鱼群在图像中与背景颜色较为相似,图片构成相对复杂,在这种情况下,相比传统的鱼群检测方法,基

基金项目:教育部产学合作协同育人项目(202101142041)

This work was supported by the Industry-University Cooperation and Education Program of the Ministry of Education(202101142041).

通信作者:王诗琦(wangshiqi888@foxmail.com)

于深度学习的检测方法在经过不断训练后对水下环境表现出的适应性会更强,具有更高的精度和鲁棒性。

当前,基于深度学习的目标检测方法主要分为两类,分别是以 R-CNN^[53] 为代表的双阶段目标检测及以 SSD^[4] 和 YOLO 系列^[5-7] 为代表的一阶段目标检测。双阶段算法检测精度较高,但其检测速度较慢,不能满足实时性的要求,单阶段算法检测和分类同时进行,检测速度较快。这两类算法均已在鱼类检测方向有所应用。例如, Tseng 等^[8] 使用改进的 mask R-CNN 在像素级自动检测和分割背景中的鱼类,时间阈值去除假阳性检测,距离阈值避免重复计数鱼。Zeng 等^[9] 将对抗性遮挡网络(AON)添加到标准的 F-RCNN 中, F-RCNN 与 AON 网络共同对抗学习,提高了算法在水下检测中的鲁棒性。Zhao 等^[10] 在 iFaster RCNN 的基础上,将特征金字塔与原有的 FRCNN 相结合,解决了目标检测中的多尺度问题,另一方面,采用 Diou 代替 IoU,提高检测精度和速度。Shen 等^[11] 在 SSD 的基础上,提出了一种基于特征金字塔的 SSD 网络模型,同时利用 SSD 及特征金字塔进行特征信息的提取,提升了各卷积层语义特征的表达能力。Zhang 等^[12] 提出了一种基于多尺度融合与无锚点 YOLOv3 的鱼群计数方法,采用基于多尺度融合的方法提取鱼群图像特征。Muksit 等^[13] 在 YOLOv3 的基础上,通过优化上采样步长及添加空间金字塔池化模块提出了 YOLO-Fish 算法,减小了检测过程中小鱼的漏检比例及复杂环境中的整体漏检比例。Zhao 等^[14] 在 YOLOv4 的基础上使用 MobileNetv3 代替特征提取主干网络减少参数量,并引入可变形卷积提高了算法的检测速度,实现了死鱼的实时监测。Zhang 等^[15] 基于 YOLOX 算法提出了水下目标检测算法 AX-YOLO,优化了其优化器及数据增强方法。Krishnan 等^[16] 设计了一种深度卷积神经网络混合模型 HDCNN-UODT,将 RetinaNet 与 EfficientNet 模型混合,使用 3 个基准数据集进行模拟,提高了模型的检测精度。

本文提出了一种基于 YOLOv7 的水下鱼群检测方法,首先根据水下成像退化的特点,采用反卷积对图像进行预处理,减少了水下成像退化因素对目标检测的干扰,随后在 YOLOv7 中引入了 SPD-Conv 模块,取代了跨卷积层和池化层,消除了它们的交替,减少了细粒度的丢失,提高了特征表示的学习效率,并在主干网络中加入了 NAM 注意力机制,将权重稀疏惩罚应用于注意力模块,使模型在保持相似性能的同时更具计算效率。

2 理论与方法

2.1 YOLOv7

YOLOv7^[17] 是目前 YOLO 系列的最新算法,其中提供了 YOLOv7, YOLOv7x 等 7 种不同大小的模型以适应不同检测场景的需求。YOLOv7 的网络由输入网络(Input)、主干网络(Backbone)、头部网络(Neck)及预测网络(Prediction) 4 个部分组成。

输入网络将输入图像缩放调整成固定的像素大小,以适应主干网络的输入要求。主干网络中 CBL 卷积层由卷积层、批量归一化层(Batch Normalization, BN)、LeakyReLU 激活函数组成,用于特征提取;E-ELAN 卷积层在原有 ELAN 结构上使用组卷积来增加添加特征的基数,并以混洗和合并基数的方式组合不同组的特征,在不干扰原有梯度路径的情况下提高网络的学习能力;MP 卷积层在原 CBL 卷积层的基础上增加了 Maxpool 结构,其包含两个分支,一个分支通过

Maxpool 使图像长宽减小为原来的一半,通过 CBL 使图像通道减少为原来的一半,另一个分支通过两个 CBL 分别使图像的长宽及通道减为原来的一半,最后使用 cat 操作将上下分支特征融合。头部网络由路径聚合特征金字塔网络(Path Aggregation Feature Pyramid Network, PAFPN) 组成,实现不同层次特征的融合。预测网络将头部网络输出的 3 层不同大小的特征图经过 REP 结构输出置信度、类别和锚框的预测结果。

尽管 YOLOv7 在 COCO 数据集上的检测速度及精度都已具有较好的表现,但将其直接应用于水下场景还是会出现以下问题:1)由于光线、水质等原因,水下采集的图片通常具有较低的分辨率,整体颜色通常偏蓝或偏绿,鱼类与环境颜色相差不大,具有较高的检测难度;2)在自然水体采集的图片中,鱼群通常在图像中占比较小,对鱼群小目标进行特征提取较为困难。

2.2 SPD-Conv 模块

在目标检测任务中,卷积神经网络(CNN)发挥了巨大的作用,但当检测图像的分辨率较低或检测目标较小时,网络的性能往往会降低。

这是由于 CNN 结构中使用了跨步卷积或池化层,这种架构会导致细粒度的丢失及低效特征的学习,而 SPD-Conv (Space-to-Depth Conv)是一种新的 CNN 块,可以取代网络中原有的跨步卷积及池化层,从而提升网络对低分辨率图像及小目标的性能^[18]。

SPD 组件将图像转换技术推广到对 CNN 网络进行特征下采样中,设尺寸为 $S \times S \times C_1$ 的任何中间特征图为 X ,其一系列子特征图切片公式为:

$$f_{0,0} = X[0:S:scale, 0:S:scale], f_{1,0} = X[1:S:scale, 0:S:scale], \dots, f_{scale-1,0} = X[scale-1:S:scale, 0:S:scale] \quad (1)$$

$$f_{0,1} = X[0:S:scale, 1:S:scale], f_{1,1}, \dots, f_{scale-1,1} = X[scale-1:S:scale, 1:S:scale] \quad (2)$$

$$f_{0,scale-1} = X[0:S:scale, scale-1:S:scale], f_{1,scale-1}, \dots, f_{scale-1,scale-1} = X[scale-1:S:scale, scale-1:S:scale] \quad (3)$$

通常给定特征图 X 后,子特征图 $f_{x,y}$ 由所有 $X(i,j)$ 组成,其中 $i+x$ 和 $i+y$ 可以被 $scale$ 整除。因此,每个子特征图可以按 $scale$ 向下采样 X ,图 1(a)~图 1(c)给出了 $scale=2$ 时的示例,下采样得到 4 个形状为 $\left(\frac{S}{2}, \frac{S}{2}, C_1\right)$ 的子特征图 $f_{0,0}, f_{1,0}, f_{0,1}, f_{1,1}$,下采样 X 的因子为 2。接着按照通道将 4 个子特征图连接得到特征图 X' ,其空间维度减小,而通道维数增加。在 SPD 特征变换层后,添加了一个具有 C_2 滤波器的非跨步卷积层,其中 $C_2 < scale^2 C_1$,并进一步完成变换:

$$X' \left(\frac{S}{scale}, \frac{S}{scale}, scale^2 C_1 \right) \rightarrow X'' \left(\frac{S}{scale}, \frac{S}{scale}, C_2 \right) \quad (4)$$

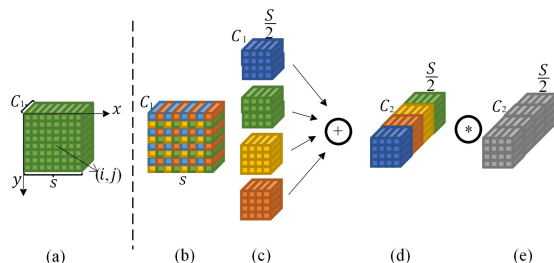


图 1 $scale=2$ 时 SPD-Conv 的结构

Fig. 1 Structure of SPD-Conv with $scale=2$

本文算法使用 SPD-Conv 模块替换 YOLOv7 网络中原有的跨步卷积结构,提升了对低分辨率图像及小目标进行检测的效果。

2.3 NAM 注意力机制

在水下目标检测中常伴有检测目标较小的问题,注意力机制即插即用,能有效增强信息的特点,并能很好地增强算法的准确性。

NAM 注意力机制 (Normalization-based Attention Module)^[19] 是一种轻量型的高效注意力机制,其采用了 CBAM 的模块集成方式,重新设计了通道注意力机制和空间注意力子模块,对于通道注意力子模块,该模块使用了批量归一化 (BN) 中的缩放因子,如式 (5) 所示。缩放因子即 BN 中的方差,方差越大表示该通道的变化越大,包含的信息越丰富,重要性也越高,因此缩放因子能反映出各通道的大小变化及通道重要性。

$$B_{out} = BN(B_{in}) = \gamma \frac{B_{in} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \quad (5)$$

其中, μ_B 和 σ_B 分别为小批量 B 的平均值和标准偏差, γ 和 β 是可训练的仿射变换参数。通道注意力子模块如图 2 和式 (6) 所示,其中 M_c 表示输出特征, γ 表示每个通道的缩放因子,其权重为 $W_\gamma = \gamma_i / \sum_{j=0} \gamma_j$ 。该算法还使用了像素归一化,即将 BN 的缩放因子应用于空间维度以确定像素的权重。相应的空间注意力子模块如图 3 和式 (7) 所示。其中, M_s 表示输出, γ 表示缩放因子,其权重为 $W_\lambda = \lambda_i / \sum_{j=0} \lambda_j$ 。同时,损失函数中添加了正则化项以抑制不突出的权重,如式 (8) 所示。其中, x 表示输入, y 表示输出, W 表示网络权重, $l(\cdot)$ 表示损失函数,

$g(\cdot)$ 表示 l_1 范式惩罚函数, p 是平衡 $g(\gamma)$ 和 $g(\lambda)$ 的惩罚值。

$$M_c = \text{sigmoid}(W_\gamma (BN(F_1))) \quad (6)$$

$$M_s = \text{sigmoid}(W_\lambda (BN(S(F_2)))) \quad (7)$$

$$Loss = \sum_{(x,y)} l(f(x,W), y) + p \sum g(\gamma) + p \sum g(\lambda) \quad (8)$$

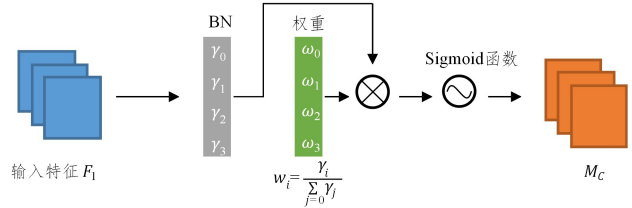


图 2 通道注意力模块

Fig. 2 Channel attention module

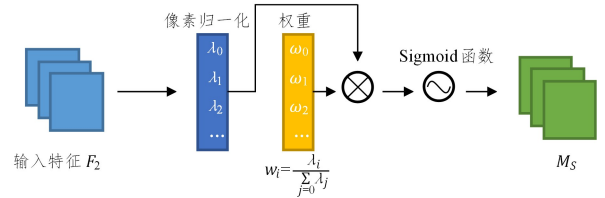


图 3 空间注意力模块

Fig. 3 Spatial attention module

本文将 NAM 注意力模块嵌入 YOLOv7 算法 Neck 网络中的 3 个 ELAN 模块之后,减少全连接、卷积等额外的计算和参数,直接使用 BN 缩放因子来计算注意力权重,有效抑制了不显著的特征,使算法在水下鱼群检测中获得更高的准确率。

2.4 算法结构

图 4 给出了改进的 YOLOv7 鱼群检测算法的结构。

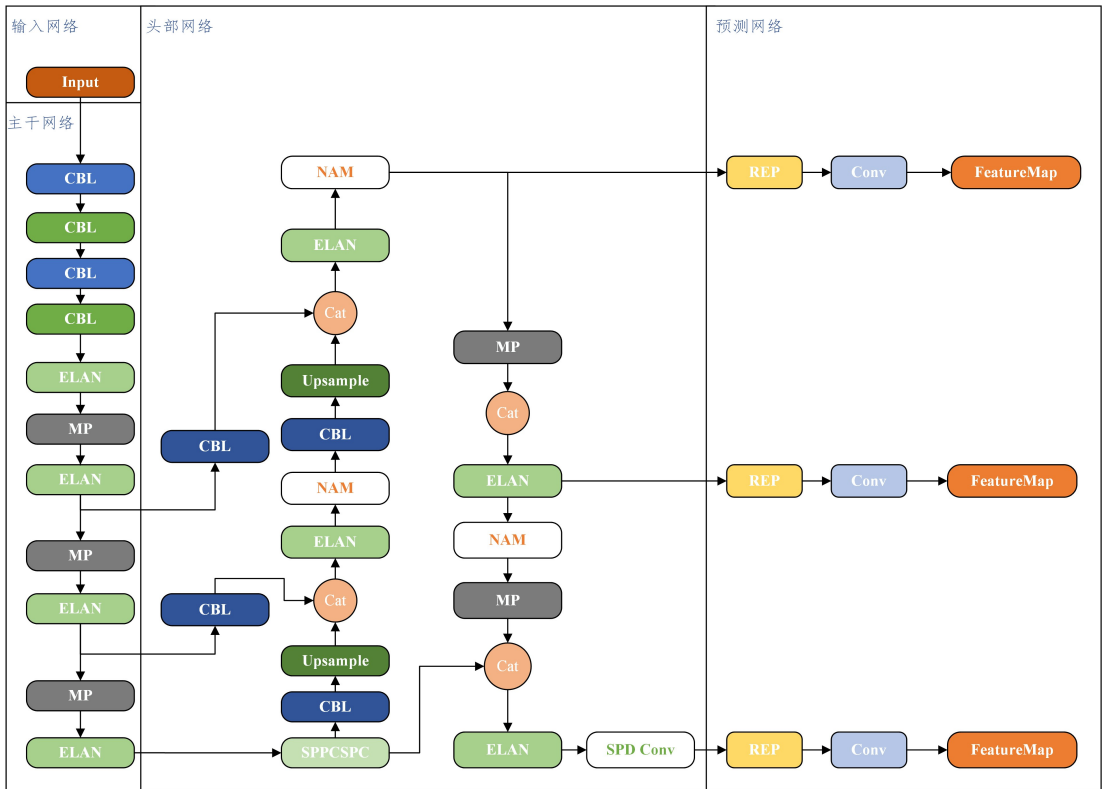


图 4 改进后的 YOLOv7 算法结构

Fig. 4 Structure of the improved YOLOv7 algorithm

针对水下图像普遍分辨率较低的问题,在 YOLOv7 主干网络中引入了 SPD-Conv 结构,其是由 Space-to-Depth 结构和

非跨步卷积层组成的,SPD 结构对特征图 X 进行切片得到子特征图 $f_{x,y}$,后,按照通道将子特征图相连接得到新的特征图

X' , 在 SPD 特征变换层后, 添加了一个具有 C_2 滤波器的非跨步卷积层, 其尺寸由 $S \times S \times C_1$ 变为 $\frac{S}{scale}, \frac{S}{scale}, scale^2 C_1$, 其空间维度减少, 通道数增加, 避免了图像细粒度信息丢失的问题, 提高了对特征表达的学习效率, 提升了对低分辨率图像的检测精度。

针对图像中鱼类目标较小的问题, 在网络中加入 NAM 注意力机制, 在通道注意力子模块中, 应用 BN 缩放因子反应各通道的大小变化及通道重要性。在空间注意力子模块中, 将 BN 缩放因子应用于空间维度以确定像素的权重 W_A , 避免了添加 SE, BAM 和 CBAM 中使用的全连接和卷积层, 抑制了不重要的通道及像素, 增强了特征提取的效果, 提升了模型对小目标检测的精确度。

2.5 水下成像退化处理

根据 Jaffe-McGlamery 等^[20]对水下光学成像的研究, 摄像机的感光单元捕捉的光信号主要来源于 3 个部分: 衰减前的直射光信号、前向散射光信号、后向散射光信号。光信号能量守恒的表达式如下:

$$I_o = I_A + I_B + I_C \quad (9)$$

其中, I_o 表示相机接收到的总光信号能量, I_A 表示直射光信号能量, I_B 表示前向散射光信号能量, I_C 表示后向散射光信号能量。

水下光学退化过程可以被视为一个时移不变的线性系统, 水下光学退化图像 $g(x, y)$ 是水下目标原始图像 $f(x, y)$ 与成像过程中系统退化函数 $h(x, y)$ 的卷积再加上成像系统噪声 $n(x, y)$ 共同作用的结果。水下目标光学成像退化的计算式如下:

$$\begin{aligned} g(x, y) &= \iint_{-\infty}^{+\infty} f(x_i, y_i) h(x-x_i, y-y_i) dx_i dy_i + \\ & n(x, y) \\ &= f(x, y) * h(x, y) + n(x, y) \end{aligned} \quad (10)$$

针对水下成像的退化过程, 本文对检测图像进行了反卷积的图像复原预处理, 减少了水下成像退化对鱼群检测的干扰。预处理公式如下:

$$H(\omega) = \int_{-\infty}^{\infty} PSF(\theta) \exp(-j2\pi\theta) d\theta \quad (11)$$

其中, PSF 是水下成像系统点扩散函数。

3 数据集

3.1 数据集标注

本文使用的公开数据集为 WildFish 数据集^[21]及 Aquarium 数据集^[22]。WildFish 是鱼类检测和识别领域最大的数据集, 由 1000 种鱼类和 54459 张图片组成, 图片来源包括 Fish-Base、海洋中心和博物馆的网站以及谷歌图像搜索。Aquarium 为 roboflow 开源数据集, 包含 7 种海洋生物, 共计 638 张原始图片, 数据增强后共计 4670 张图片。为确保图像质量, 对 WildFish 数据集中的图片进行了筛选及标注, 以获得用于模型训练的有效鱼群图像数据集。筛选后的图片包含鱼类重叠及鱼类隐藏在复杂背景中的状况, 使数据样本更贴近真实状况。最终得到的数据集包括 1077 张图片, 按照 8:2 比例将数据集分为训练集、测试集。其中, 训练集有 835 张图片, 测试集有 242 张图片。

数据集图像的标注使用 LabelImg 图像标注工具进行, 得到 PASCAL VOC 格式的 xml 文件, 存储的主要信息包含图片的实际宽高及标注框左上角与右下角的坐标值, 而 YOLO 格式的标签中存放的是标注框的中心坐标值与标注框的宽高。因此, 将 VOC 格式的标注文件转换为 YOLO 格式的标注文件, 需要对坐标值进行归一化运算, 计算式如下:

$$x = (x_{\min} + x_{\max}) / 2width \quad (12)$$

$$y = (y_{\min} + y_{\max}) / 2height \quad (13)$$

$$\omega = (x_{\max} - x_{\min}) / width \quad (14)$$

$$h = (y_{\max} - y_{\min}) / height \quad (15)$$

其中, $width, height, depth, x_{\min}, x_{\max}, y_{\min}, y_{\max}$ 分别为图片的宽, 高和标注框的坐标值, x, y, ω, h 为 YOLO 格式下标注框的中心值坐标及标注框的宽和高。图 5 给出了数据集标注图片示例。

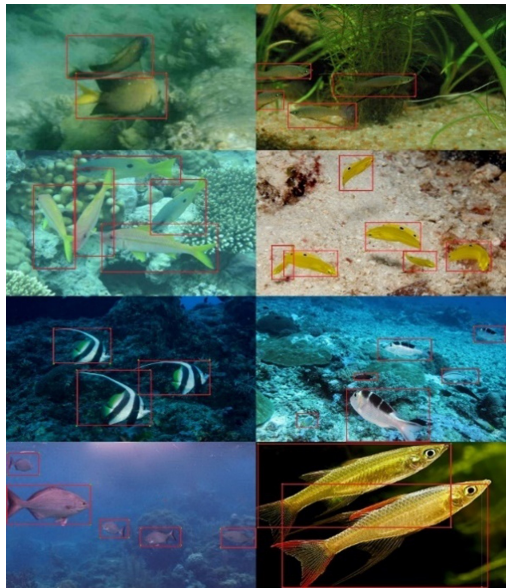


图 5 数据集标注图片示例

Fig. 5 Example of dataset annotated images

本文算法的运行环境为 Windows10 操作系统, 编程语言为 Python3.8, 算法训练使用 GPU P100。

3.2 图像预处理

本文针对水下成像退化模型, 对检测图像进行了反卷积处理。数据集优化前后的效果对比图如图 6 所示。



图 6 数据集优化前后的对比

Fig. 6 Comparison before and after dataset optimization

表 1 图像处理前后评价指标的对比

Table 1 Comparison of evaluation indicators before and after image processing

PSNR	SSIM	MSE
31.93	0.71	20.35

表 2 图像处理前后检测精度的对比

Table 2 Comparison of detection accuracy before and after image processing

	mAP@.5/%	Precision/%	Recall/%
原图	96.7	94.5	90.3
反卷积处理后	97.2	97.7	91.8

4 实验结果分析

4.1 评价指标

为了客观评价本文算法在改进前后的差异,本文使用准确率(Precision)、召回率(Recall)、平均精度均值 mAP(mean Average Precision)3个指标来评价模型的检测效果。为了更加全面地比较本文算法与 Efficientdet, SSD, YOLOv5, YOLOv7 及 YOLOv8 算法之间的差异,选取准确率、召回率、平均精度均值 mAP、参数量及模型大小 5 个指标分别作为检测精度及网络复杂度的评价指标。

4.2 消融实验

为了验证 SPD-Conv 和 NAM 注意力机制对本文算法的有效性,本文以 YOLOv7 为基准结合改进模块进行了消融实验的设计。实验结果如表 3 所列。从表 3 可以看出,在 YOLOv7 头部网络加入 NAM 注意力机制后,模型的特征提

取能力得到了提升,相比 YOLOv7 模型,召回率由 81.4% 提升到了 86.8%,同时 mAP@0.5 的值提升了 3.0%。在 YOLOv7 头部网络加入 SPD-Conv 模块后,提升了模型对低分辨率图像及小目标的检测效果,在准确率、召回率及 mAP 值上都有明显的提升,其中 mAP@0.5 的值提升了 6.3%,准确率提升了 8.4%,召回率提升了 6.3%。本文算法在此基础上加入 SPD-Conv 模块后,模型的整体精度上升至 97.2%,相比原始 YOLOv7 算法提升了 7.6%,模型准确率达到 97.7%,相比原始算法提升 8.5%,模型召回率达到 91.2%,提升了 9.8%。图 7 给出了本文算法与 YOLOv7 算法的训练效果对比。可以看出改进后的本文算法在模型整体精度、准确率及召回率上的表现均优于原始的 YOLOv7 算法,且模型收敛速度相比原始算法更快。

表 3 消融实验结果的对比分析

Table 3 Comparative analysis of ablation experimental results

实验	SPD-Conv	NAM	mAP@.5/%	Precision/%	Recall/%
1			89.6	89.2	81.4
2		✓	92.6	88.7	86.8
3	✓		95.9	97.6	87.7
4	✓	✓	97.2	97.7	91.2

分析消融实验结果可知,在 YOLOv7 模型上添加不同模块带来的性能提升并不是单一数值的叠加,而是各模块间相互影响共同作用,在准确率、召回率与平均精度均值中,本文算法均取得了更好的表现,证明了添加各模块的有效性。在实际训练过程中,本文算法的模型收敛速度也得到了提升,证明了本文算法对多尺度的目标检测具有稳定性。

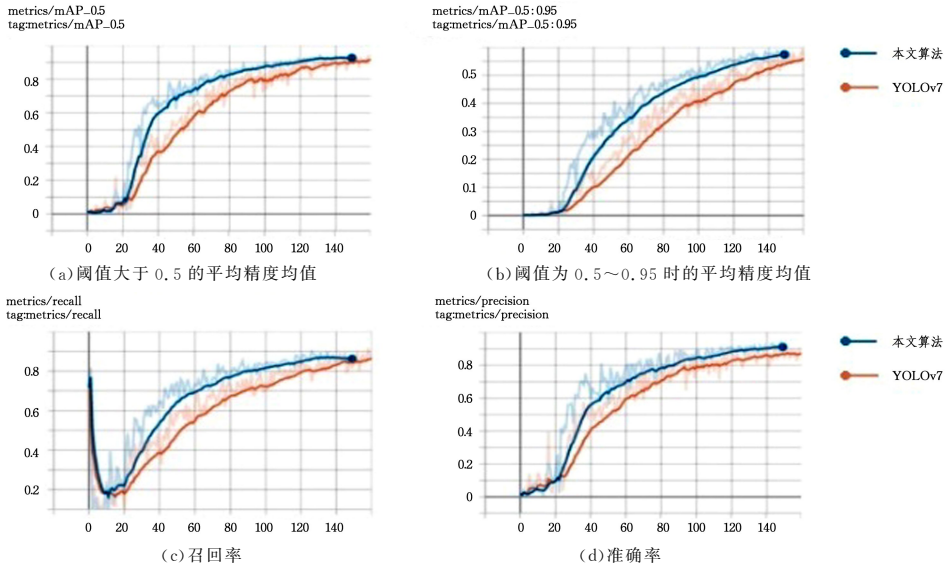


图 7 训练效果对比

Fig. 7 Comparison of training effects

4.3 对比实验

为了客观验证本文改进算法在鱼群检测中的有效性,将本文方法与 YOLOv5, YOLOv7, SWIPNet, SSD 在检测性能上进行对比,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 4、表 5 列出了各个模型在 WildFish 数据集及 Aquarium 数据集上的检测效果对比。从表 4 可见,本文算法在 WildFish 数据集中取得了最高的平均精度均值、准确率及召回率,其具体值分别为 97.2%,97.7% 及 91.2%。且 FPS 值达到 124.6 帧每秒,可以满足实时监测的需求。在

召回率方面,本文算法相比对比算法的提升均达到 4% 以上。

表 4 数据集 WildFish 的实验数据

Table 4 WildFish dataset experimental data

方法	mAP@0.5/%	Precision/%	Recall/%	FPS/(帧每秒)
Efficientdet	84.6	89.7	83.5	78.9
SSD	79.4	75.7	73.9	57.6
YOLOv5	93.2	88.8	78.8	96.3
YOLOv7	89.6	89.2	81.4	125.7
YOLOv8	94.3	93.7	90.5	101.3
本文算法	97.2	97.7	91.2	124.6

从表 5 可见,在 Aquarium 数据集中,本文模型的 mAP 为 80.5%,较 Efficientdet,SSD,YOLOv5,YOLOv7 及 YOLOv8 分别提升了 18.4%,11.6%,6.9%,2.0% 及 2.7%。在 5 个类别的精度均值中,本文算法达到最高。这说明本文算法针对水下鱼群目标较小、图像较不清晰而导致的检测效率低的

问题做出的改进是有效的。

图 8、图 9 为各算法在 WildFish 数据集与 Aquarium 数据集上的鱼群检测效果图。通过对比可见,本文算法能够正确检测出更多低分辨率小目标图像,显示了在检测鱼群目标方面的优势,验证了本文算法的有效性。

表 5 数据集 Aquarium 实验数据

Table 5 Aquarium data set experimental data

方法	mAP@0.5/%	Fish	Jellyfish	Penguin	Puffin	Shark	Starfish	Stingray
Efficientdet	62.1	44.7	65.9	49.8	52.6	60.0	84.6	76.3
SSD	68.9	76.5	83.7	65.2	53.1	60.2	72.3	71.3
YOLOv5	73.6	80.7	90.0	69.7	60.3	64.5	73.2	76.8
YOLOv7	78.5	82.0	92.1	71.9	63.9	71.4	82.4	85.8
YOLOv8	77.8	79.6	93.6	73.6	64.1	72.0	78.5	83.2
本文算法	80.5	83.1	94.9	74.8	63.2	73.4	83.1	90.7

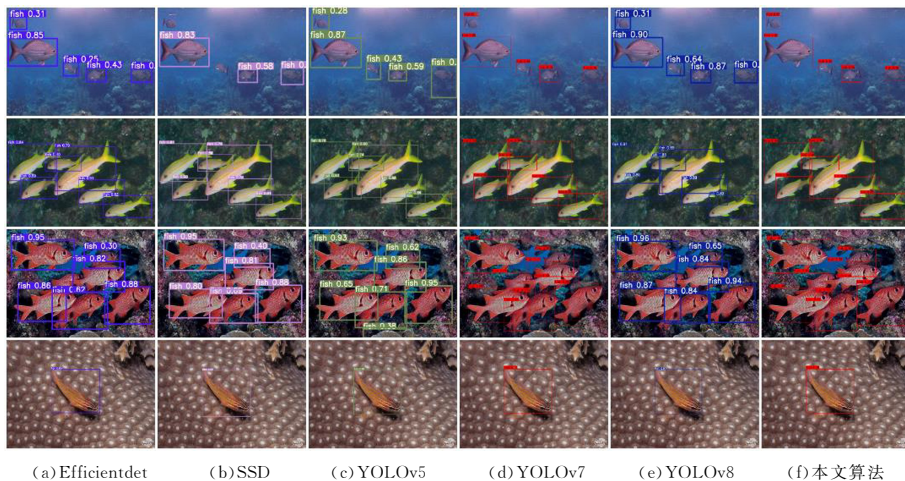


图 8 WildFish 数据集检测效果对比图

Fig. 8 Comparison of detection effects of WildFish dataset

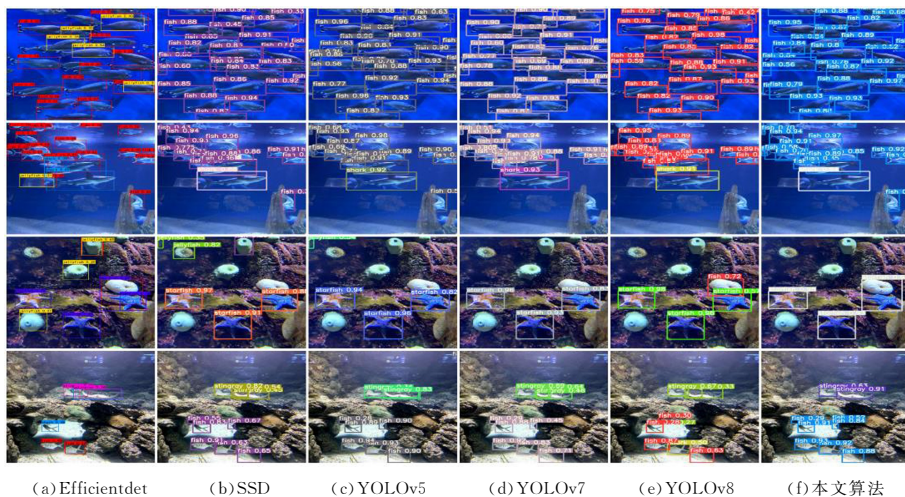


图 9 Aquarium 数据集检测效果对比图

Fig. 9 Comparison of detection effects of Aquarium dataset

结束语 本文基于 YOLOv7 提出了一种用于鱼群检测的改进算法,针对水下图像普遍分辨率较低的问题,在 YOLOv7 主干网络中引入了 SPD-Conv 结构。SPD-Conv 由 Space-to-Depth 结构和非跨步卷积层组成,SPD 层对特征图进行下采样,保留了信道维度中的所有信息,因此避免了图像细粒度信息丢失的问题,提高了对特征表达的学习效率,提升了对低分辨率图像的检测精度。针对图像中鱼类目标较小的

问题,在网络中引入 NAM 注意力机制,减少了网络中的全连接和卷积层,抑制了不重要的通道及像素,增强了特征提取的效果,提升了模型对小目标检测的精确度。针对水下成像退化,对数据集图像进行了反卷积预处理,减小了退化因素对检测图像造成的影响。

实验结果表明,在 WildFish 数据集上本文模型的平均精度均值(mAP)为 97.2%,在 Aquarium 数据集上本文模型的

平均精度均值为 80.5%,与先进的目标检测模型进行比较后得出,本文算法拥有较高的识别率并且更加轻量化,满足了在复杂水下环境中对鱼群检测的要求。

参 考 文 献

- [1] ID L, MIAO Z, PENG F, et al. Automatic counting methods in aquaculture; a review[J]. Journal of the World Aquaculture Society, 2021, 52(2): 269-283.
- [2] FAN L Z, LIU Y. Automate fry counting using computer vision and multi-class least squares support vector machine[J]. Aquaculture, 2013, 380 / 381 / 382 / 383: 91-98.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. IEEE, 2014: 580-587.
- [4] LIU W, ANGUELOV D, ERHAND, et al. Ssd: Single shot multibox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [5] 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.
- [6] REDMON J, FARHADIA, YOLO9000: better, faster, stronger [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 7263-7271.
- [7] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv:1804. 02767, 2018.
- [8] TSENG C H, KUO C H. Detecting and counting harvested fish and identifying fish types in electronic monitoring system videos using deep convolutional neural networks[J]. ICES Journal of Marine Science, 2020, 77(4): 1367-1378.
- [9] ZENG L C, SUN B, ZHU D Q. Underwater target detection based on Faster R-CNN and adversarial occlusion network[J]. Engineering Applications of Artificial Intelligence, 2021, 100: 104190.
- [10] ZHAO D, YANG B, DOU Y, et al. Underwater fish detection in sonar image based on an improved Faster RCNN[C]//2022 9th International Forum on Electrical Engineering and Automation (IFEEA). Zhuhai, China, 2022: 358-363.
- [11] SHEN J Y, LI L Y, DAI Y L, et al. A fish stock detection method based on feature fusion SSD[J]. Computer Simulation, 2020, 37(11): 422-426, 469.
- [12] ZHANG L, HUANG L, LI B B, et al. Fish counting method based on multi-scale fusion and anchorless YOLO v3[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(S1): 237-244.
- [13] ABDULLAH A M, FAKHRUL H, MD F H B E, et al. Fahad Hasan Bhuiyan EMON, et al. YOLO-Fish: A robust fish detection model to detect fish in realistic underwater environment [J]. Ecological Informatics, 2022, 72: 101847.
- [14] ZHAO S L, ZHANG S, LU J M, et al. A lightweight dead fish detection method based on deformable convolution and YOLOV4[J]. Computers and Electronics in Agriculture, 2022, 198: 107098.
- [15] ZHANG Y S, XU W X, YANG S S, et al. Improved YOLOX detection algorithm for contraband in X-ray images[J]. Applied Optics, 2022, 61: 6297-6310.
- [16] VIJIYAKUMAR K, GOVINDASAMY V, AKILA G. Hybridization of Deep Convolutional Neural Network for Underwater Object Detection and Tracking Model[J]. Microprocessors and Microsystems, 2022, 94: 104628.
- [17] WANG C Y, BOCHKOVSKIY A I, LIAO H Y. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv:2207. 02696, 2022.
- [18] SUNKARA R, LUO T. No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects[J]. arXiv:2208. 03641, 2022.
- [19] LIU Y, SHAO Z, TENG Y, et al. NAM: Normalization-based Attention Module[J]. arXiv, abs/2111. 12419, 2021.
- [20] MIAO Y. Underwater image adaptive restoration and analysis by turbulence model[C]//2012 World Congress on Information and Communication Technologies. IEEE, 2012: 1182-1187.
- [21] ZHUANG P, WANG Y, QIAO Y Y. Wildfish: A large benchmark for fish recognition in the wild[C]//Proceedings of the 26th ACM international conference on Multimedia. 2018: 1301-1309.
- [22] Roboflow. Aquarium Combined Image Dataset [EB/OL]. <https://roboflow.com>



CHEN Yuzhang, born in 1984, Ph.D, associate professor. His main research interests include photoelectric detection and image processing.



WANG Shiqi, born in 1999, postgraduate. Her main research interests include deep learning and neural networks.