

文章编号: 1672-8785(2025)02-0049-08

基于改进 YOLOx-nano 的海上红外目标检测算法

张 俊 位 门 吕 璐

(中国电子科技集团公司第五十八研究所, 江苏 无锡 430070)

摘 要: 提出了一种基于改进 YOLOx-nano 的海上红外目标检测算法。通过对检测头的分类与定位任务进行解耦, 并引入改进的特征金字塔网络 (Feature Pyramid Network, FPN) 结构, 不仅提升了模型的精度和收敛速度, 而且提高了红外大目标检测能力。将改进的压缩和激励网络 (Squeeze-and-Excitation Network, SENet) 通道注意力机制模块加入到模型中, 增强了模型的非线性表达能力, 同时提高了有效特征学习能力。为了加快嵌入式平台模型的前向推理速度, 引入剪枝技术来实现模型剪枝, 在保证召回率不降低的情况下减少模型参数。通过测试集对本文算法进行了验证。结果表明, 该算法的平均精度 (Average Precision, AP) 比原始 YOLOx-nano 算法提高了 1.35%, 达到了 93.92%。本文算法平衡了模型精度与耗时的矛盾关系, 在提升性能的同时, 保证了模型检测的速度。

关键词: 红外目标检测; YOLOx-nano; 模型剪枝; 注意力机制

中图分类号: TP751 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2025.02.006

Marine Infrared Target Detection Algorithm Based on Improved YOLOx-nano

ZHANG Jun, WEI Men, LV Lu

(The 58th Research Institute of CETC, Wuxi 430070, China)

Abstract: A marine infrared target detection algorithm based on improved YOLOx-nano is proposed. By decoupling the classification and positioning tasks of the detection head and introducing an improved feature pyramid network (FPN) structure, not only the accuracy and convergence speed of the model are improved, but also the infrared large target detection capability is improved. The improved squeeze-and-excitation network (SENet) channel attention mechanism module is added to the model to enhance the nonlinear expression ability of the model and improve the effective feature learning ability. In order to speed up the forward reasoning speed of the embedded platform model, the pruning technology is introduced to implement model pruning, and the model parameters are reduced without reducing the recall rate. The algorithm in this paper is verified by the test set. The results show that the average precision (AP) of the algorithm is 1.35% higher than that of the original YOLOx-nano algorithm, reaching 93.92%. The algorithm in this paper balances the contradictory relationship between model accuracy and time consumption, and ensures the speed of model detection while

收稿日期: 2024-09-24

作者简介: 张俊(1995-), 男, 湖北武汉人, 工程师, 主要研究方向为图像处理、基于深度学习的目标检测。
E-mail: ZHANG_JUN@cumt.edu.cn

improving performance.

Key words: infrared target detection; YOLOx-nano; model pruning; attention mechanism

0 引言

在海防安全监测中,海上目标的准确检测有着非常重要的作用。近远海防御通过检测目标来推断目标位置,从而增强海防预警能力^[1]。搭载由可见光/红外相机和云台构成的目标监控系统可以实现更远的视场并及时发现目标。该系统由于其检测优势在海防领域取得了广泛的使用^[2]。

随着人工智能技术的不断发展,越来越多的深度学习算法得到了实际应用。在目标检测领域,伴随着大数据的爆发式发展,基于深度学习的目标检测算法相对于传统的目标检测算法表现出更大的优势。目前主流的目标检测算法大致分为一阶段与两阶段两种。两阶段算法的设计比一阶段算法复杂。具体说来就是两阶段算法首先将输入图像经过候选区域生成网络来生成候选框,再通过分类网络对候选框进行分类,从而获得目标的类别和位置信息。其中有代表性的经典算法有 R-CNN^[3]、Fast R-CNN^[4]、Faster R-CNN^[5]等。而一阶段算法只需将输入图像通过一个主干网络就可以在输出层中直接获得目标的类别和位置信息。其中有代表性的经典算法有 SSD^[6]和 YOLO^[7-9]系列。两阶段算法的检测精度比一阶段算法高,但是计算量更大且运算速度慢,所以一阶段算法在追求实时性的工业界得到了较为广泛的应用。

为解决传统检测方法提取的浅层特征在复杂的海面场景泛化能力差而导致检出率和识别率较低的问题,杜晨锡等人提出了一种改进的基于 YOLOv2 的检测算法^[10]。基于深度学习的目标检测技术涉及大量的卷积运算,对部署设备算力要求高,翁佳昊等人就此提出一种基于多目标优化检测模型的剪枝算法^[11]。针对海面复杂场景下目标存在多个尺度的问题,高明星等人提出了一种基于改进 YOLOv5 的复

杂场景下路面病害检测方法^[12]。

YOLO 系列算法在目标检测中取得了较多的应用。本文基于 YOLOx-nano^[13]提出一种改进的海上红外目标检测算法。该算法通过改进模型解耦头提升了模型的精度和收敛速度,并将改进的 SENet 的注意力机制模块加入到模型中,增加了模型的抗非线性能力。为设计一个轻量化模型,通过引入剪枝技术实现模型剪枝。在不降低召回率的情况下建立了轻量化模型,再通过引入改进 FPN 模块来改变网络结构,提高了红外大目标检测能力。通过对模型的实验验证,本文算法的有效性得到了证明。

1 算法原理

1.1 轻量级红外目标检测框架

YOLOx 是 YOLO 系列目标检测算法中的一员,其主干网络是在 YOLOv3^[14]的基础上进行改进的,主要创新点是在 YOLO 输出层中使用了 Anchor-free 机制。其参数量是 Anchor-Based 参数量的 2/3,可在一定程度上减少模型的参数量。在训练过程中使用 Mosaic、Mixup 两种数据增强方法,可以有效提高训练数据的丰富度,从而增强模型的泛化性。与之前的经典 Sinkhorn-Knopp 算法不同,该算法需要通过用多次迭代求得最优解来分配标签。采用精简版 SimOTA 算法进行标签分配可以减少 25% 的训练时间。

为了将 YOLOx 算法部署在嵌入式平台且能实时进行目标检测,设计了轻量级目标检测网络——YOLOx-nano。YOLOx-nano 的网络结构分为输入部分、主干网络、颈部、输出层四个部分。其中,输入部分的作用是将输入图像统一缩放到模型输入的 416×416 像素。主干部分是特征提取器,由 Focus 模块、残差模块、SPP 空间金字塔模块等构成。Focus 模块对输入的三通道图像进行切片处理,获得 4 个

分开的特征层。它们分别经过卷积后进行特征层堆叠, 此时就将宽高维度上的信息转换到了通道维度, 再通过卷积的方式提取不同的特征, 在下采样的同时可以减少计算量。残差模块使用 1×1 和 3×3 的卷积核进行特征提取, 经过 Batch Normalization 后使用 SiLU 进行激活。空间金字塔池化模块使用不同大小的池化核为特征层进行最大池化操作, 融合不同感受野的特征层可有效增强特征提取能力。颈部则是将提取的语义特征与位置特征进行融合, 同时对深层特征与浅层特征进行堆叠, 结合不同尺度的特征可使模型得到更加丰富的特征。输出层与之前的 YOLO 系列有所不同。改进的 YOLOx-nano 为了解决分类与回归任务之间的冲突, 通过检测头对目标类别、位置、交并比

进行解耦。解耦后的检测头结构如图 1 所示。实验结果表明, 解耦后不仅在精度上有所提高, 而且网络的收敛速度也加快了。

1.2 注意力机制增强结构

注意力机制指的是网络学习局部目标信息的机制, 来源于人类的注意力方式, 可以提高感兴趣的目标区域的影响力, 从而实现对背景信息的抑制。注意力机制可以分为通道注意力机制和特征图注意力机制。其中, 通道注意力机制可以将输入数据的全局信息传达到模型的尾部, 使得有效特征图通道的权重变大、无效特征图通道的权重变小, 从而提高整个模型对目标的召回率。因此, 本文采取通道注意力机制来增强特征提取。图 2 所示为本文所用的通道注意力机制结构。

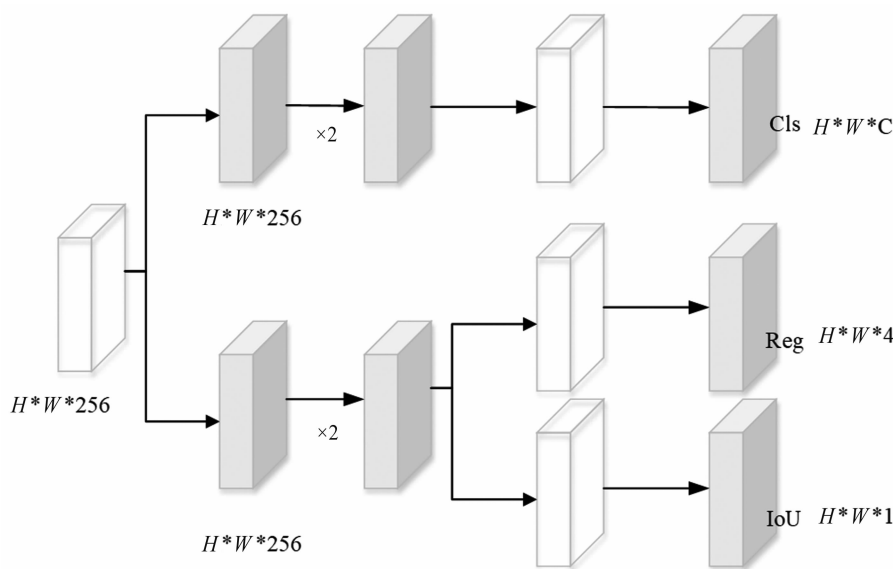


图 1 解耦检测头

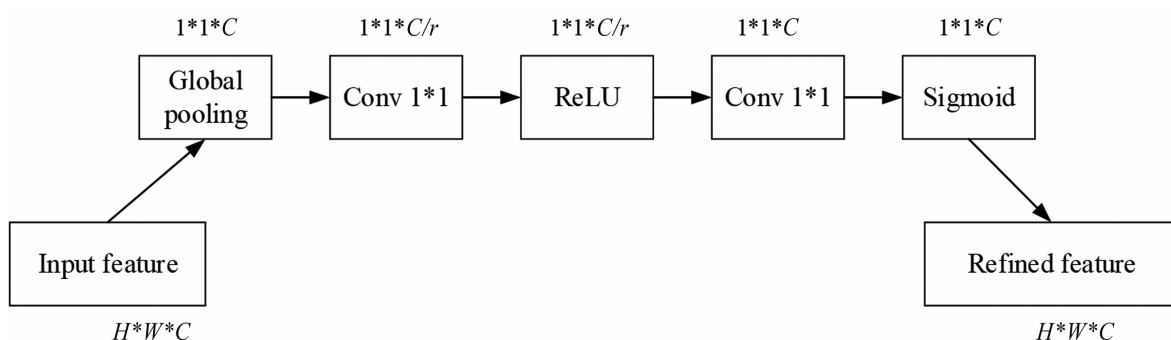


图 2 改进的 SENet 注意力机制

将 SENet 注意力机制的 FC 全连接层换成 1×1 的卷积层, 可有效减少参数量, 同时在进行通道数变为 C/r 的降维和将通道数变为 C 的升维过程中能够加强跨通道信息交互, 在不改变特征图尺寸的条件下增强模型的非线性表达能力。

1.3 网络模型剪枝

深度学习网络模型从卷积层到全连接层存在大量冗余的参数, 大量神经元激活值趋近于 0。将这些神经元去除后可以表现出同样的模型表达能力。这种情况被称为过参数化, 而对应的技术则被称为模型剪枝^[15]。剪枝就是通过去除网络中冗余的通道、卷积核、卷积层以得到一个更轻量级的网络, 同时不影响性能。

本文利用 BN 层缩放因子作为依据对网络进行剪枝, 可以在不引入新参数的同时提升剪枝效率。BN 层所作的变换为

$$\begin{cases} \hat{Z} = \frac{Z_{in} + \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \\ Z_{out} = \gamma \hat{Z} + \beta \end{cases} \quad (1)$$

式中, Z_{in} 和 Z_{out} 表示 BN 层的输入和输出, μ_B 和 σ_B 表示激活值的平均值和标准差, γ 和 β 是训练的参数。本文将 γ 与特征通道输出的乘积作为判定通道重要程度的依据。结果较小则表明此通道重要程度低。通过去除不重要的通道来减少模型参数^[16]。评估通道重要程度的目标函数为

$$L = \sum_{(x,y)} l(f(x,M), y) + \varphi \sum g(\gamma) \quad (2)$$

式中, x 、 y 分别是输入与输出, M 为网络权重, $g(\gamma)$ 是缩放因子的惩罚项, φ 是用来约束模型损失和惩罚项的超参数。本文采用不同的剪枝率对模型进行剪枝。试验结果表明, 将剪枝率设置为 0.2 时综合性能最好。相比于原模型 yolox_nano, 剪枝后模型参数量从 9×10^5 降低为 5.6×10^5 , 降幅为 38%, 但是性能并没有下降, AP 甚至从 91.84% 上升到 93.92%。在置信度都设置为 0.4 时, 船类目标的召回率从 91.35% 上升为 92.31%, 精度从 86.36% 上

升到 86.61%。

1.4 网络模型的输出层改进

在卷积神经网络中, 深层特征层表达的是语义特征, 浅层特征层表达的是几何特征, 目标检测输出层不仅要输出目标位置还要输出目标类别, 所以需要将浅层的几何信息和深层的语义信息进行融合。YOLOx-nano 采用 FPN 网络结构(见图 3)进行信息融合。

FPN 网络对主干网络的特征层构建两组特征金字塔^[17]。首先构建自下而上的金字塔, 对应图中的 1、2、3 三个特征层。它们的尺寸分别是模型输入尺寸的 $1/8$ 、 $1/16$ 、 $1/32$ 。然后将特征层 3 进行复制, 生成特征层 4。对特征层 4 进行上采样操作, 使其尺寸与特征层 2 一致。再用 1×1 卷积对特征层 2 进行降维处理。接着将两者对应元素相加, 就得到了特征层 5。将特征层 5 进行复制后对特征层 5 进行上采样操作, 使其与特征层 1 的尺寸一致。再用 1×1 卷积对特征层 1 进行降维处理, 然后将两者对应元素相加就得到了特征层 6, 也就构成了自上而下的金字塔。最后将融合后的特征层 4、5、6 经过几个 3×3 卷积操作就可以得到最终的输出。

FPN 网络中包含三个感受野不同的输出层, 分别用来检测不同大小的目标。相同距离下, 在图像上成像尺寸较小的目标可用步长为 8 的 yolo 层进行检测。该层分辨率较大, 保留的小尺寸目标特征信息较多, 可有效地对小尺寸目标进行探测。考虑到船类目标靠近光电系统时在视场上的占比较大, 本文在原有 FPN 的基础上增加了步长为 64 的 yolo 层检测。这一层的感受野最大, 检测大目标的能力也最强。改进后的 FPN 网络结构如图 4 所示。

2 算法实现

2.1 数据集构建

本文按照 8:1:1 将海上红外船舶数据集分为训练集、验证集、测试集。在训练的过程中, 预处理运用亮度调整、随机加入噪声、旋转、裁剪等数据增强方法, 增加数据的泛化性,

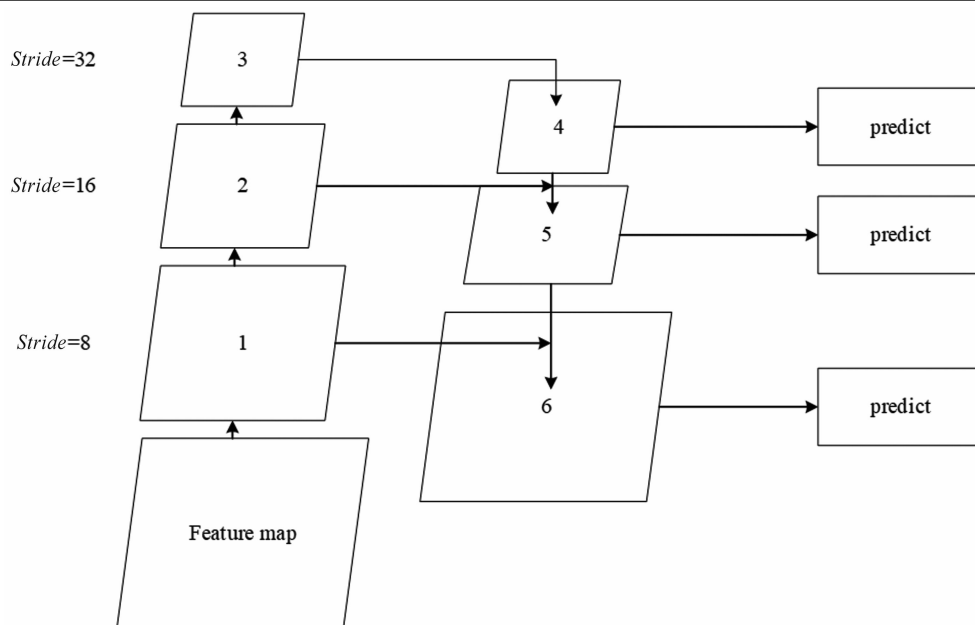


图 3 FPN 网络结构

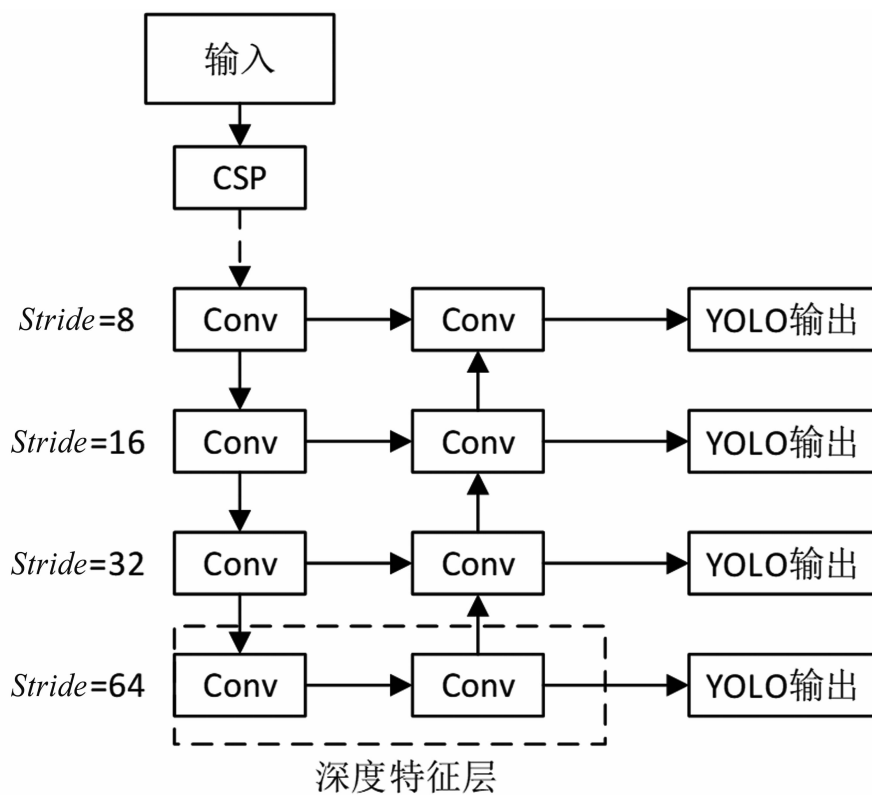


图 4 改进的 FPN 网络结构

从而提高算法的性能。

2.2 评价指标

精确率-召回率 (Precision-Recall, PR)^[18] 曲线代表的是精准率与召回率之间的关系。一般情况下, 分别将召回率和精确率设置为横坐

标和纵坐标。其计算公式为

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

式中, TP 为检测识别正确的目标框数量, FP 为误检的目标框数量, FN 为漏检的目标框数量。判断目标被正确检测识别的依据如下: 算法输出的检测框类别预测与实际标注框相交, 且两框的交并比大于等于 0.35。

2.3 模型训练

本文中训练模型的系统为 Ubuntu18.04, 使用 PyTorch 开源深度学习框架; 软件环境为 Python3.7、CUDA11.4、CuDnn8, 显卡为 NVIDIA GeForce RTX3060。表 1 列出了本文所提算法训练过程设置的部分网络参数。

表 1 参数配置

参数名	参数值
输入尺寸	416×416
学习率	0.0013
批大小	16
学习率衰减因子	0.1
非极大值抑制	0.5
动量	0.949

2.4 模型部署

算法移植的嵌入式平台是海思 3559。HI3559A 芯片包含双核 ARM Cortex A73@1.8 GHz、双核 ARM Cortex A53@1.8 GHz、单核 ARM Cortex A53@1.8 GHz、四核 DSP@700 MHz、双核 NNIE@840 MHz 神经网络加速引擎、双核 ARM Mali G71@900 MHz。系统支持 2 路独立 ISP 处理、8 路 sensor 输入、6 路视频拼接, 同时集成了专用的神经网络加速引擎, 为目标检测识别深度学习算法的运行提供了强大的硬件环境。将训练好的权重文件通过 onnx 工具包转换为通用格式 onnx, 然后利用 MNN 框架将 onnx 转换为 MNN 框架所需要的权重文件格式, 最后通过 MNN 框架进行网络模型解析与前向推理。

3 实验结果分析

本文训练过程中的损失变化曲线和 PR 曲线分别如图 5 和图 6 所示。随着迭代次数的增加, 损失值逐渐减小至 0.95 左右(浮动), 整个过程中没有出现明显的震荡。模型在经过

120 次迭代后逐步达到收敛状态。从图 6 中可以看出, 在虚警率为 15% 的情况下, 本文算法的召回率大于 95%, AP 值达到 93.92%, 充分表明本文算法具有较强的检测性能。图 7 所示为 YOLOx-tiny 算法、YOLOx-nano 算法和本文算法在不同场景下的红外船舶检测结果。

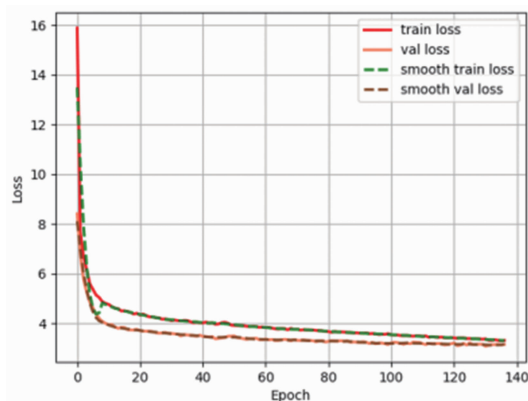


图 5 损失变化曲线

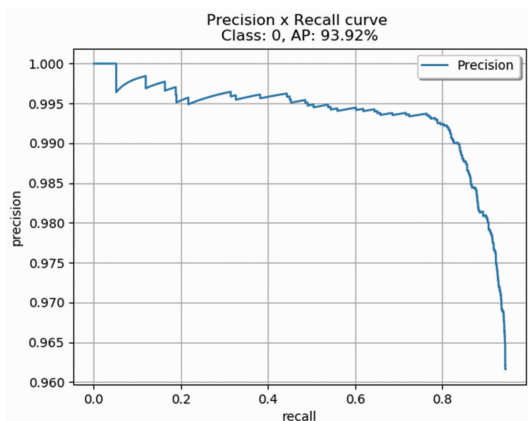


图 6 PR 曲线

由图 7 中第一列的检测结果可以看出, 在海面背景比较简单的情况下, 虽然红外船舶目标占整个视场的区域较小, 但三种算法均能检测出目标, 本文算法检测出目标的置信度达到 0.96。图 7 中的第二列反映了海岸背景比较复杂的情况, 包含 5 个目标。YOLOx-tiny 算法、YOLOx-nano 算法和本文算法检测出的目标数分别为 4、4、5。只有本文算法能将所有目标正确检测出来, 在参数量最少的情况下召回率依旧很高。由图 7 中第三列的检测结果可以看出, 当目标占整个视场的区域较大时, YOLOx-

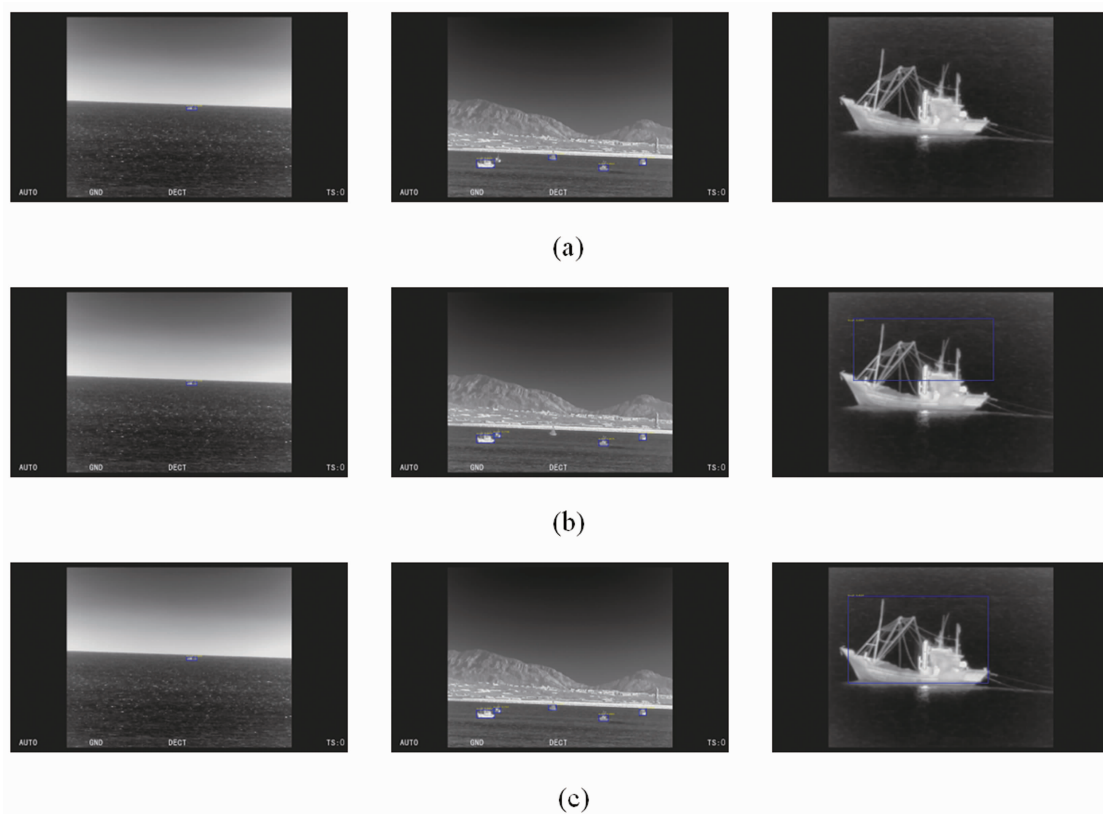


图 7 不同场景下的红外船舶目标图像检测结果:(a) YOLOx-tiny;(b) YOLOx-nano;(c)本文算法

表 2 算法性能对比

算法	参数量	精确率	召回率	耗时
YOLOx-tiny	5.06MB	87.27%	91.49%	509 ms
YOLOx-nano	0.91MB	86.36%	91.35%	176 ms
本文算法	0.56MB	86.61%	92.31%	68 ms

tiny 算法没有检测出目标, Yolox-nano 算法虽然能检测出目标但是定位不准确, 表明本文改进的 FPN 结构能够有效提高大目标检测能力。测试集共有 1000 张测试数据, 包含 1469 个船舶目标。表 2 列出了通过三个模型分别进行前向推理后的统计结果。

由表 2 可以看出, 本文算法通过剪枝操作将模型参数量优化到最小值(仅有 0.56MB), 在测试集上的精度比 YOLOx-nano 算法提高了 0.25%, 召回率更是提高了 0.96%。YOLOx-tiny 算法虽然在精度上是最高的, 但是召回率比本文算法低 0.82%, 同时在嵌入式上的耗时也是最大的。综合比较精度、召回率、耗时三个指标后发现, 本文算法的检测性能较好, 处理速度最快。

4 结束语

本文通过对检测头进行解耦, 并引入改进的 SENet 注意力机制和改进的 FPN 检测头, 提高了模型检测精确率和大目标召回率; 使用模型剪枝技术对训练好的模型进行剪枝以减少模型参数量, 对于嵌入式平台进行实时的红外目标检测识别具有重要指导意义。目前, 海上红外目标检测识别还存在红外目标特征不明显、有效红外样本难以获取等问题。针对红外目标特征不明显的问题, 后续可采用对图像进行细节增强后再检测的两阶段检测识别方法。针对有效红外样本难以获取的问题, 可以通过数据增强或迁移学习等技术进行融合, 从而提高红外目标的检测识别召回率和精确率。

参考文献

- [1] 周明, 马亮, 王宁. 面向海面目标检测的陆海分离和海面分区算法研究 [J]. 雷达学报, 2019, 8(3): 366–372.
- [2] 张泽琨, 谭震彪, 余坤, 等. 特征图强化网络: 利用特征图强化船舶检测模型训练的网络结构 [J]. 遥感学报, 2023, 27(12): 2697–2705.
- [3] 李宗民, 邢敏敏, 刘玉杰, 等. 结合 Faster RCNN 和相似性度量的行人目标检测 [J]. 图学学报, 2018, 39(5): 901–908.
- [4] 卢志鹏, 黄凯华, 刘思明, 等. 基于 Faster-RCNN 的船舶焊缝 X 射线缺陷图像检测技术应用 [J]. 无损检测, 2023, 45(7): 36–40.
- [5] 冯小雨, 梅卫, 胡大帅. 基于改进 Faster R-CNN 的空中目标检测 [J]. 光学学报, 2018, 38(6): 250–258.
- [6] Son Y J, Choi O. Image-based hand pose classification using faster R-CNN [C]. Jeju: 17th International Conference on Control, Automation and Systems (ICCAS), 2017.
- [7] Prathap B R, Kumar K P, Chowdary C R, et al. AI-Based Yolo V4 Intelligent Traffic Light Control System [J]. Journal of Automation, Mobile Robotics and Intelligent Systems, 2022, 16(4): 53–61.
- [8] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger [C]. Honolulu: 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [9] Adiono T, Ramadhan R M, Lin M C H. Fast and scalable multicore YOLOv3-tiny accelerator using input stationary systolic architecture [J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2023, 31(11): 1774–1787.
- [10] 杜晨锡, 严云洋, 刘以安, 等. 基于 YOLOv2 的视频火焰检测方法 [J]. 计算机科学, 2019, 46(6): 301–304.
- [11] 翁佳昊, 秦永法, 唐晓峰, 等. 基于多目标优化的车道线检测模型剪枝算法 [J]. 传感器与微系统, 2023, 42(7): 125–127.
- [12] 高明星, 关雪峰, 范井丽, 等. 基于改进 YOLOv5-DeepSORT 算法的公路路面病害智能识别 [J]. 森林工程, 2023, 39(5): 161–174.
- [13] Song C Y, Zhang F, Li J S, et al. Detection of maize tassels for UAV remote sensing image with an improved YOLOX model [J]. Journal of Integrative Agriculture, 2023, 22(6): 1671–1683.
- [14] 于子涵, 李凯峰, 黄静颖, 等. 基于 FPGA 的高分辨率红外目标检测技术研究 [J]. 激光与红外, 2024, 54(1): 130–135.
- [15] 张良, 张增, 舒伟华, 等. 基于 YOLOv3 的卷积层结构化剪枝 [J]. 计算机工程与应用, 2021, 57(6): 131–137.
- [16] 刘宇, 雷雪梅. 融合 MobileNetV3 特征的结构化剪枝方法 [J]. 上海交通大学学报, 2023, 57(9): 1203–1213.
- [17] 曾宝繁, 潘祖望, 秦坚贵, 等. 基于 YoloX-BTF-PN 网络的输送带表面损伤检测方法 [J]. 港口装卸, 2023, 30(6): 46–49.
- [18] 吕璐, 程虎, 朱鸿泰, 等. 基于深度学习的目标检测研究与应用综述 [J]. 电子与封装, 2022, 22(1): 010307.