

基于轻量化与注意力机制的船舶除漆机器人 实时目标检测*

袁小芳^{1,2}, 李潘¹, 孙荣武³, 许浩志¹

1. 湖南大学电气与信息工程学院, 湖南长沙 410082
2. 机器人视觉感知与控制技术国家工程研究中心, 湖南长沙 410082
3. 湖南星邦智能装备股份有限公司, 湖南长沙 410600

摘要: 自动巡航船舶除漆机器人目标检测受外部干扰时, 存在算法检测精度下降、难以达到实时性要求等问题。为了解决这些问题, 首先将重参深度可分离移动网络模块(Repvit-MobileNet block)引入到YOLOV5的主干网络中, 提高检测速度。其次, 在骨干网络每个阶段后增加位置注意力机制, 扩大模型的全局感受野, 提升模型的目标定位及抗干扰能力。然后, 将卷积块注意力模块(CBMA)引入到颈部网络中, 通过融合CBMA模块增强特征提取能力, 提高网络模型的检测性能。最后, 提出了一种Refine-Loss损失函数, 通过优化预测框和真实框的几何关系、兼顾IOU的权重和置信度信息, 提高对机器人目标位置的检测精度。在船舶机器人实验数据集中进行测试与验证, 结果表明: 融合Repvit-MobileNet block与注意力机制的YOLOV5轻量化网络平均检测精度达到了84.1%, 在边缘设备上的推理运算速度达到了26.6 f/s, 满足船舶除漆机器人目标检测工业应用的需求。

关键词: 除漆机器人; 轻量化; 注意力机制; 目标检测

中图分类号: TP242.6 **文献标志码:** A **文章编号:** 2097-0137(2026)01-0013-10

Ship paint-removal robots real-time object detection based on lightweight and attention mechanism

YUAN Xiaofang^{1,2}, LI Pan¹, SUN Rongwu³, XU Haozhi¹

1. College of Electrical and Information Engineering, Hunan University, Changsha 410082, China
2. National Engineering Research Center of RVC, Changsha 410082, China
3. Hunan Sinoboom Intelligent Equipment Company Limited, Changsha 410600, China

Abstract: When the automatic ship paint-removal robot encounters external interference, existing algorithms suffering from performance degradation and insufficient real-time processing capability. To address these challenges, the Repvit-MobileNet block is integrated into the backbone network of YOLOV5 to enhance detection speed. Additionally, the positional attention mechanism has been incorporated after each stage of the backbone network, broadening the model's global receptive field and improving both target localization and interference resistance. Then, a convolutional block attention module (CBAM) is implemented in the neck network, and the feature extraction ability is enhanced by integrating the CBMA module to improve the detection performance of the network model. Lastly, a Refine-Loss loss function is proposed to optimize the geometric relationship between

* 收稿日期: 2024-12-07

录用日期: 2024-12-25

网络首发日期: 2025-04-07

基金项目: 国家自然科学基金(62473140)

作者简介: 袁小芳(1979年生), 男; 研究方向: 机器视觉感知; E-mail: yuanxiaofang@hnu.edu.cn

全文阅读



ZR20240344

the predicted bounding box and the true bounding box which also balances weight and confidence information related to IOU, leading to improved accuracy in detecting the robot's target position. Subsequent experiments from ship robotic datasets show that the lightweight YOLOV5 network combining Repvit-MobileNet block and attention mechanism can reach 84.1% in the experiment with average precision, and the inference speed on the edge device reaches 26.6 f/s, which meets the need of industrial applications for object detection of ship paint-removal robots.

Key words: paint-removal robot; light weight; attention mechanism; object detection

船舶除漆机器人在船厂的应用越来越多(李磊等, 2023; 姜泽等, 2023), 但目前船舶除漆机器人自动化和智能化程度不高, 且其作业运动控制由人工远程遥控作业, 存在操控难度大、误操控和疲劳作业等情况, 导致船舶表面除漆作业质量不稳定和工作效率低等问题, 影响工程按时交付。因此, 将自动化控制与智能化检测技术应用到船舶除漆机器人中, 可以提高生产效率和作业质量, 降低工人劳动强度, 并促进船舶行业向智能化的方向发展。

随着机器视觉和深度学习技术的发展, 自动化控制与智能化检测技术在工业中的应用越来越多。基于深度学习的目标检测方法(Ding et al., 2021)已在各种场景中广泛应用, 特别是在工业自动化(Li et al., 2024), 自动驾驶(Haris et al., 2020)、卫星遥感(Lv et al., 2020)、高空安全搜救(Lei et al., 2012)等领域。其中, 基于视觉的目标检测算法最适合目标识别与定位任务, 其能够同时获得目标的精确位置和分类信息, 从而解决船舶除漆机器人的目标识别与定位难题。促进船舶除漆机器人自动化与智能检测系统的开发与应用。

目前, 研究人员在船舶除漆机器人定位检测领域开展了大量的工作并提出了多种检测方法。方璇等(2023)提出并实现了一种基于外部 RGB-GD 相机和惯性测量单元组合的爬壁机器人自主定位方法。该方法使爬壁机器人位置定位误差 < 0.02 m, 姿态估计航向角误差 $< 2.5^\circ$, 横滚角误差 $< 2.5^\circ$, 俯仰角误差 $< 1.5^\circ$, 有效提高了爬壁机器人定位精度。张文等(2022)针对大型钢结构外表面巡检、涂装等作业, 提出了一种基于即时定位与地图构建的自主移动机器人。李希平等(2024)提出了一种除漆爬壁机器人结构, 并采用视觉辅助导航技术, 实现了机器人的自主移动。尽管这些方法在常规数据集上取得了不错的结果, 但复杂环境下的船舶机器人检测还是存在误报和漏检的问题, 导致检测可靠性和鲁棒性比较低。

自动巡航船舶除漆机器人需要实时精准定位,

才能完成对除漆机器人运动轨迹的决策与控制。而船舶表面背景复杂, 除漆机器人在运动过程中姿态变化多样, 受某些特定的工况背景和光照条件的影响, 很容易出现漏检和误检测的情况。只有避免目标的漏检和错检, 提高实时检测效率, 才能保障除漆机器人安全、高效、稳定地工作。YOLOV5 是一种单阶段目标检测网络, 具有出色的性能, 实现了端到端的训练而无需中间过程, 并且检测速度快, 可以满足一般目标的检测要求。但是在处理大规模、高分辨率图像时, 因为扩张卷积模块的设计需要大量的计算资源, 增加了算法的计算时间和能源成本。直接将 YOLOV5 算法应用于船舶机器人目标检测时, 检测效率相对较慢, 检测精度和实时性也不能满足工业应用的要求。本文提出了一种具有轻量化模块和多重注意力机制的 YOLOV5 改进网络。该方法提升了检测速度和稳定性, 将其部署到船舶机器人的视觉智能检测系统中, 在模拟平台上取得了良好的实验和测试效果。而且, 其提高了检测效率和可靠性, 满足了船舶领域对智能化和自动化的急迫需求。

1 YOLOV5 算法改进

1.1 YOLOV5 算法原理

YOLOV5 网络结构模型由骨干网络、颈部网络和检测头组成。使用 CSPDarknet53 作为骨干网络, 它由 Conv、C3、SPPF 模块组成。Conv 模块是一种标准的卷积模块, 用于对特征图进行下采样、升维和降维、归一化和非线性等。C3 由 Conv 模块和 CSP 结构组成, 是一个重要的特征提取模块。先降维是为了更好地理解特征信息, 再升维能提取到更多更详细的特征, 最后利用残差结构, 将输入与输出相加, 避免梯度消失的问题。SPPF 的主要功能是融合多尺度特征, 将不同尺度的特征图, 转化为统一尺度。

颈部网络 Neck 主要实现图像特征的融合, 采用路径聚合网络(PANet)模型, 结合特征金字塔网络

(FPN)(Lin et al., 2017)和像素聚合网络(PAN)组成,FPN结构从上到下提取特征信息,提升目标检测效果。PAN结构从下往上提取特征层信息,以获得位置信息,将浅层图形特征和深层语义特征结合,增强了定位能力。Head层为检测网络的检测模块。同时,YOLOV5采用马赛克数据增强方法,通过翻转、随机裁剪、亮度变化等手段拼接4张图像,以丰富图像信息并增强网络的鲁棒性,是应用最广泛的单阶段目标检测网络之一。

1.2 改进的YOLOV5模型

由于YOLOV5主干特征提取网络采用C3结构,有较大的参数量。在某些实际应用场景中,由于模型过于庞大,硬件面临着计算资源不足的问题,导致检测速度慢或者检测系统响应速度不够快,而且如此大而复杂的模型难以被部署到移动或者嵌入式设备上。如果除漆机器人视觉检测系统检测速度慢,会影响机器人目标决策与运动控制。虽然未来硬件的性能会越来越强,但小而高效的卷

积神经网络(CNN, convolutional neural network)模型在这些移动设备中非常有应用价值。

本文提出的YOLO-RMA网络分为Backbone、Neck、Head三个部分,如图1所示。检测图像输入到Backbone部分后,输出F2、F3和F4三个特征层映射到Neck中。由于机器人目标移动后尺度会发生变化,目标的全局信息对检测精度有较大影响,为了让检测网络在不增加计算量的情况下保持较高的检测精度,在Neck网络中引入位置注意力(CA)和卷积块注意力(CBMA)模块。CA模块的注意力操作能区分空间方向(坐标),并且生成坐标感知的特征图,提升了目标在长距离空间信息下的特征表示能力。CBMA模块通过融合局部感受野内的空间信息和通道信息来提取特征,可以更加关注检测目标自身的感兴趣区域,提取有价值的目标特征信息。融合上述模块的YOLO-RMA网络,不仅实现了网络模型的轻量化,而且达到了检测精度与速度的平衡。

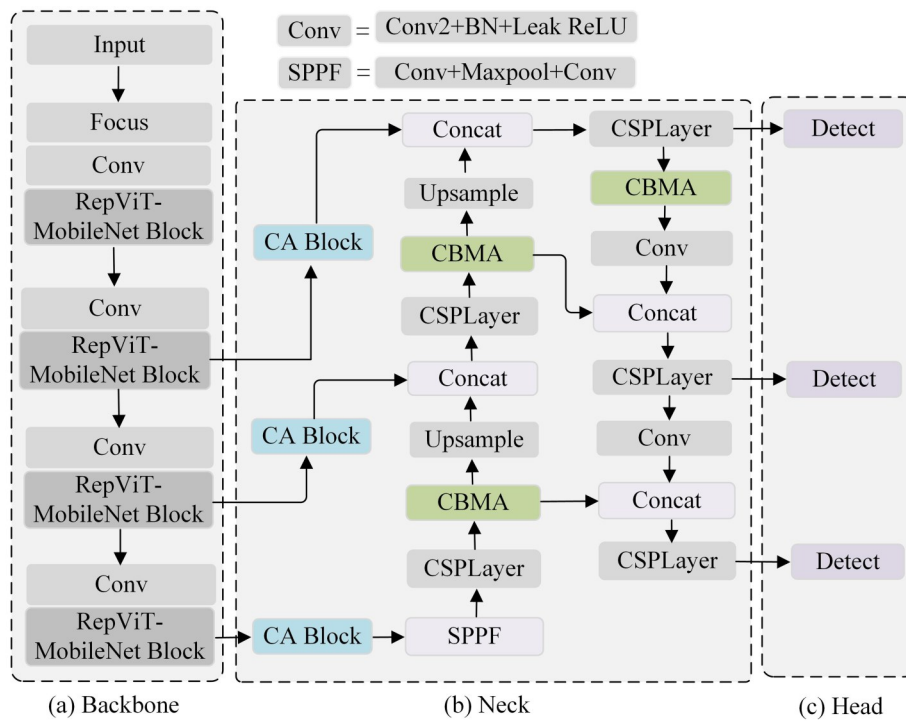


图1 YOLO-RMA网络结构

Fig. 1 Network structure of YOLO-RMA

1.3 MobileNet轻量化模型

MobileNet系列网络是谷歌提出的轻量级网络模型,具有模型和参数量小、计算速度快等特点,经常被使用到移动端或者嵌入式的系统中。MobileNet V1(Howard et al., 2018)采用深度可分离卷积减少参数量和计算量,在保持较高精度的同

时,提升了模型的效率和速度。MobileNetV2(Sandler et al., 2018)网络设计了一个具有线性瓶颈的倒残差块,先将特征图扩展到高维,使用深度可分离卷积提取特征,然后映射回低维表示,以便更好地学习如何扩展和压缩数据。MobileNet V3(Howard et al., 2019)通过神经网络架构搜索算法

(NAS)搜索到最优的模型结构,并引入压缩和扩展模块(SE)通道注意力机制(Hu et al., 2018),性能和速度都表现优异。MobileV3 Block首先执行扩张操作(使用深度可分离卷积),然后是挤压操作(通过瓶颈层,例如 1×1 卷积)。扩张操作旨在提取包含多尺度信息的重要特征,随后挤压操作从中提取必要信息用于目标检测任务。这有助于网络关注到未充分表示的特征,提高目标检测的准确性。因为扩展层和挤压层都可以学习参数,所以整个网络结构可以学习到如何更好地扩展数据和重新压缩数据。MobileV3 Block中的卷积不会带来计算量大的问题,是因为深度可分离卷积降低了计算量。

如何通过不增加计算复杂度来提高特征表达能力,还能提高检测的准确率和推理速度。RepVit-block(Wang et al., 2024)作为一个即插即用模块,其主要作用是通过自注意力和卷积操作,在不显著增加计算开销的前提下,提升模型在这些任务中的表现。重参深度可分离移动模块(RepVit-MobileNet block)的网络结构是基于Inception模块

结构(Szegedy et al., 2015),通过组合多个不同大小的卷积核(例如 $1 \times 1, 3 \times 3$)操作融合感受野,来实现提取不同尺度的特征。

如图2所示,重参深度可分离移动模块的输入层接收原始图像作为输入,输入图像首先经过一个卷积(Convolution)、批量归一化(BN, batch normalization)和ReLU激活函数组成的输入块处理,以初始化特征提取过程。先用 1×1 卷积的深度可分离卷积执行扩张操作得到特征 T_1 ,再使用 3×3 可分离模块、 1×1 可分离模块、恒等模块对输入特征进行运算操作的模块称为RepVit block,融合后得到中间特征 T_2 ,然后使用压缩和扩展模块(SE)对特征 T_2 进行卷积处理,得到另一个融合特征 T_3 。为了保留原始特征信息,通常还会包含一个残差连接,即对输入 T_3 特征通过一个 1×1 卷积处理,经过归一化和激活后与 T_3 特征相加,形成最终的输出特征 T_4 。这种结构在骨干网络中提取、增强并保留全面的特征细节,同时考虑计算成本约束,确保了高性能。而每个阶段包含多个分支、叠加不同感受野的深度可分离卷积来捕获多尺度特征。

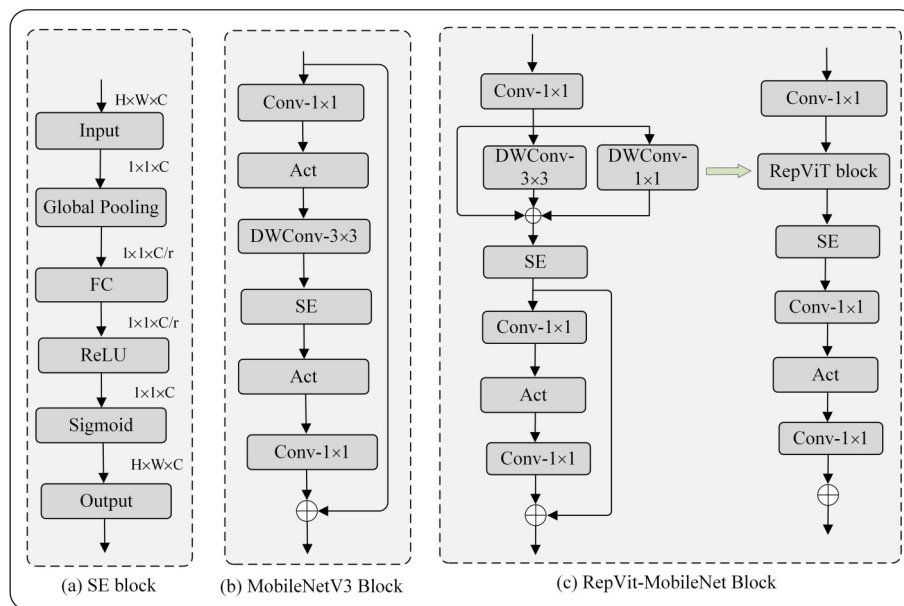


图2 重参深度可分离移动模块

Fig. 2 RepVit-MobileNet block

YOLOV5模型的主干网络中的CSPDarknet结构相对复杂,网络层数较深,参数数量和计算量较大,运行效率较低。RepVit-MobileNet block是一种轻量级的模块,不仅降低了计算开销,同时也维持了比较高的检测性能。将YOLOV5的骨干网络中的CSPLayer结构换成了梯度流更丰富的RepVit-

MobileNet block,并对不同尺度模型调整不同的通道数。实施步骤如下:

- 1)第一个卷积层的kernel设置为 3×3 ;
- 2)骨干网络中所有的CSPLayer模块换成RepVit-MobileNet block,结构如图1(a)所示;
- 3)骨干网络中RepVit-MobileNet block在每个

阶段的模块数量为 3-6-6-3。因此本文提出的基于 RepVit-MobileNet block 的轻量化 YOLO 骨干网络, 不仅降低模型的参数量, 而且提高了模型的检测效率和实时性, 并能在移动设备上应用和部署, 达到了检测速度与精度有效的平衡。

1.4 融合注意力机制的颈部网络

CBAM(Woo et al., 2018)由通道注意力模块和空间注意力模块两个模块组成, 通道注意力模块(CAM, channel attention module)关注每个通道的特征图的重要程度。如图 3 所示, 空间注意力模块(SAM, spatial attention module)关注特征图中每个部分(像素)的重要程度, 同时兼顾了通道与空间特征的表达, 相比传统的卷积层参数更少、效果更好。CAM使用并行的平均池化和最大池化, 每个池化分别经过两个卷积模块, 然后经过 Sigmoid 得到注意力机制概率图, 即

$$M_c(F) = \sigma \left(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F)) \right) \\ = \sigma \left(W_1 \left(W_0(F_{\text{avg}}^c) \right) + W_1 \left(W_0(F_{\text{max}}^c) \right) \right).$$

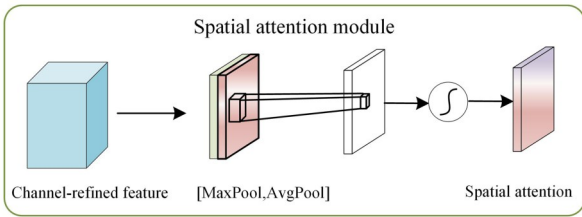


图 3 空间注意力模块

Fig. 3 Spatial attention module

SAM判断特征图中每个部分的重要程度, 需要综合每个部分所有通道的特征。如图 4 所示, 图中相同颜色的部分有一个标记重要程度的权重, 其计算公式为

$$M_s(F) = \sigma \left\{ f^{7 \times 7} \left(\left[\text{AvgPool}(F); \text{MaxPool}(F) \right] \right) \right\} \\ = \sigma \left(f^{7 \times 7} \left(\left[F_{\text{avg}}^s; F_{\text{max}}^s \right] \right) \right),$$

σ 表示 Sigmoid 函数, $f^{7 \times 7}$ 表示大小为 7×7 的卷积运算。利用平均池化(Average Pooling)以降低特征图的分辨率, 逐步下采样以捕获更高级别的特征。

CBAM是一个轻量级、通用的模块, 它可以无缝地集成到任何 CNN 架构中, 而开销可以忽略不计, 并且可以与基础 CNN 一起进行端到端训练, 有助于分离跨通道相关性和空间相关性, 从而提取更有意义的特征。为了使模型更加关注通道与空间的信息以提高检测精度, 本文在 PANET 中使用 CBAM 注意力机制(如图 5 所示), 让

颈部特征提取网络可以更加关注感兴趣区域, 提取有价值的目标特征。

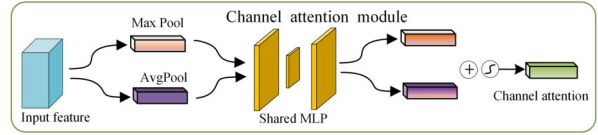


图 4 通道注意力模块

Fig. 4 Channel attention module

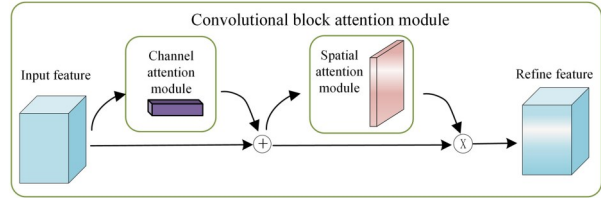


图 5 CBMA 模块

Fig. 5 CBMA module

虽然通道注意可以让模型的性能取得了很好的效果, 但是也削弱了通道注意机制内的位置信息, 位置信息在生成空间选择性注意力中是非常重要的。因此在骨干网络和颈部之间的跳连中加入了位置注意力(CA)模块(Hou et al., 2021), 目的是在每个阶段的输出特征上更加有效地传输全局信息, 使特征融合模块能够访问来自多个抽象级别(不同尺度阶段)的特征。所以骨干网络在各个阶段特征后连接 CA 自注意力机制, 可以更好地捕获长距离依赖性。虽然在骨干网络的每个阶段引入了位置注意力机制, 但得益于底层阶段的深度可分离卷积, 整体计算复杂度仍然保持在可控范围内, 确保了模型的实时性。

融合通道注意力和位置注意力的双路注意力机制颈部网络(DA-PAFPN), 如图 1(b)所示。DA-PAFPN 作为 PAFPN 框架下的一个变体, 在骨干网络输出的特征采用深度可分离卷积作为令牌混合器, 有效降低了计算复杂度, 并保持了良好的性能; 结合深度可分离卷积和位置注意力机制的优势, 使得 YOLO-RMA 网络在目标检测任务上实现了更高的准确率, 为实际应用提供了可靠保障。

1.5 损失函数优化

YOLOV5 的损失函数包含框回归损失, 置信度预测损失, 类别预测损失。总体的损失即为以上三者的加权相加, 通过改变权值可以调整对三者损失的关注度。损失函数

$$\text{Loss} = \omega_{\text{box}} L_{\text{box}} + \omega_{\text{obj}} L_{\text{obj}} + \omega_{\text{cls}} L_{\text{cls}},$$

其中包围框回归损失使用 CIOU, 即

$$L_{\text{CIOU}}(b, b^g) = 1 - \text{IOU}(b, b^g) + \frac{\rho^2(b, b^g)}{c^2} + \alpha v,$$

$$\alpha = \frac{v}{1 - \text{IOU} + v},$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^g}{h^g} - \arctan \frac{w}{h} \right)^2.$$

由于边界框的预测需要着重于该框与真实框的重叠区域面积, 且重叠区域的面积与两者并集区域的面积之比越大越好, 只使用平方损失无法很好地衡量这一点。为了解决这一问题, 衍生出多个基于 IOU 的损失计算方式, 其中 CIOU 通过更多的维度来考虑预测框与真实框的差异, 效果更好, 被广泛采用。

在实际样本中会存在低质量的样本, 而造成损失值剧烈震荡的问题, 将 CIOU 的回归分支中加入 Focal Loss, 因此 Focal-CIOU 损失函数

$$L_{\text{Focal-CIOU}}(b, b^g) = \text{IOU}^\gamma L_{\text{CIOU}}.$$

通过整合 CIOU 损失函数(CIOU Loss)和 Focal 损失函数(Focal Loss), 得到了最终的 Focal-CIOU 损失函数, 其中 γ 是一个用于控制曲线弧度的超参。分类损失函数和置信度损失函数均采用二元交叉熵损失函数(BCE Loss), 计算公式分别为

$$L_{\text{obj}}(p_o, p_{\text{iou}}) = \text{BCE}_{\text{obj}}^{\text{sig}}(p_o, p_{\text{iou}}),$$

$$L_{\text{cls}}(c_p, c_{gt}) = \text{BCE}_{\text{cls}}^{\text{sig}}(c_p, c_{gt}).$$

本文在 YOLOV5 损失函数的基础上, 采用新的边界框回归损失函数 Focal-CIOU Loss, 通过调整损失函数之间的关联信息, 优化预测框与真实框(GT box)之间的几何关系, 在模型训练过程中动态地为不同质量的锚框分配权重。该损失函数(Refine-Loss)由框回归损失, 置信度预测损失, 类别预测损失组成。Refine-Loss 损失函数为

$$\text{Loss} = \lambda \omega_{\text{box}} L_{\text{box}} + \omega_{\text{obj}} L_{\text{obj}} + \omega_{\text{cls}} L_{\text{cls}},$$

ω 作为损失函数对应项的权重参数, 默认设置为

$$\omega_{\text{box}} = 0.4, \quad \omega_{\text{obj}} = 0.3, \quad \omega_{\text{cls}} = 0.3,$$

$\lambda = [\lambda_1, \lambda_2, \lambda_3]$ 用于平衡每个尺度的输出特征图的边界框的权重。 λ 的默认值为 [0.1, 0.8, 0.1], 对应三种尺度的特征图。

2 模型训练与移动端部署

2.1 算法实验及应用平台

算法应用平台基于 Ubuntu 20.04 系统, 算法实

验采用 PyTorch 1.12.0+MMYOLO 0.5.0 的深度学习框架, 编程语言采用 Python 3.8.16, 软硬件配置如表 1 所示。

表 1 算法实验的软硬件配置

Table 1 Software and hardware configuration for experiment

名称	配置
操作系统	Ubuntu 20.04
编程语言	Python 3.8.16
深度学习框架	PyTorch 1.12.0
CPU	Intel Core i5-12600KF
GPU	NVIDIA RTX 4080
CUDA	CUDA 11.3
平台	PyCharm

2.2 实验数据

实验平台搭建与数据采集由湖南星邦智能装备股份有限公司提供。采集设备是海康威视网络摄像头, 数据集采集时覆盖多种拍摄角度、采集不同曝光程度的数据, 充分丰富数据集, 让网络模型学习到多种特征, 提升泛化能力。采集的图片经过裁剪处理后, 图像分辨率为 1 920×1 080, 图像总数为 8 764 张。

采用微软通用目标检测(MS COCO)数据集格式对本实验数据集的数据进行标注, 使用 Labellmg 软件人工标注除漆机器人的外形区域。数据集中包含 2 种尺寸、规格不同的除漆机器人设备。本文提出的 YOLO-RAM 算法在机器人数据集上进行训练、验证和测试, 图片和标签都按照 70%、10% 和 20% 的比例进行划分。

2.3 优化策略与训练过程

本文采用 3 组优化器分别对卷积神经网络的权重参数、偏置参数和批量归一化参数进行优化。在热身训练阶段, 不同组的优化器采用不同的学习率进行参数更新和优化, 同时采用基于迭代次数(Iter)的更新策略。在度过热身训练阶段后, 训练过程趋于稳定, 此时采用基于迭代次数(Epoch)的更新策略。训练时采用指数移动平均(EMA, exponential moving average)策略平滑模型参数, 并开启自动混合精度训练(AMP, automatic mixed-precision training)加速模型训练过程。

首先, 整个模型在数据集上训练 300 个 Epoch, 其中特征和检测网络由第一阶段学习到的权值初始化, 使用随机梯度下降(SGD)优化器进行初始化, 学习率初始化为 0.01, 并进行线性预热, 从第 11

个 Epoch 开始,学习率降低到初始值的 1/10。其次,使用 SGD 和初始学习率为 0.002 的余弦学习计划进行模型训练。批量大小、权重衰减和 SGD 动量分别设置为 8、0.000 5 和 0.9,具体参数如表 2 所示。

表 2 训练参数设置
Table 2 Training parameter settings

配置	参数
优化器(Optimizer)	SGD
初始学习率(Base lr)	0.001
权重衰减系数(Weight decay)	0.000 5
优化器动能系数(Optimizer momentum)	0.95
批量大小(Batch size)	4
学习率策略(Learning rate schedule)	Cos
训练迭代次数(Training epochs)	300

2.4 评价指标

实验评估和评价指标主要有标准精度率(P)、召回率 R 、平均精度(AP)、平均精度均值(mAP)、F1

分数、浮点运算数(FLOPs)和每秒传输帧数(FPS)。具体来说,FPS表示模型的推理速度,FLOPs表示模型占用的内存空间。AP是单个类别的平均精度,综合了不同召回率下的精确率表现。mAP是对每个类别,计算不同IOU阈值(如0.5~0.95,步长0.05)下的AP。精度率 P 表示预测的正确样本在所有样本中的比例,召回率 R 表示预测的正确样本在实际正确样本中的比例。

3 实验结果及分析

3.1 消融实验

为了验证各个改进模块的有效性,进行了消融实验。实验中依次加了各个模块,基准模型YOLOV5网络的mAP为83.3%。在骨干网络中添加Repvit-MobileNet block后,mAP下降至81.0%。颈部网络中添加CBMA与CA模块后,mAP提升至83.5%,检测速度有了显著提升。修改为Refine-Loss后,检测精度也提升到了84.1%,实验结果如表3所示。

表 3 消融实验结果

Table 3 Results of ablation experiments

算法	mAP[0.5,0.95]	mAP[0.5]	Param/ 10^6	FLOPs/ 10^6	FPS
YOLOV5s	83.3	98.9	28.7	30.9	12.1
+ RepVit-MobileNet block	81.0 ↓ 2.3	98.6 ↓ 0.2	11.3 ↓ 17.4	12.1 ↓ 18.8	26.9 ↑ 14.8
+ CA&CBAM	83.5 ↑ 0.2	99.5 ↑ 0.6	12.1 ↓ 16.6	13.4 ↓ 17.5	26.5 ↑ 14.4
+ Refine-Loss	84.1 ↑ 0.8	99.5 ↑ 0.6	12.1 ↓ 16.6	13.5 ↓ 17.4	26.6 ↑ 14.5

3.2 对比实验

将基于Repvit-MobileNet block改进的YOLOV5s网络与其他常见的目标检测网络进行对比,结果如表4所示。从表4可以发现,相较于其他算法,改进后的YOLOV5s的平均检测精度最高,分别比YOLOV5s、MobileNet-V2、SSDVGG

300、Resnet50高出0.8%、9.2%、4.7%和1.6%。改进的YOLOV5s在FPS指标上取得了最高的性能;在推理速度方面,改进的YOLOV5s更有优势。综合而言,改进YOLOV5s检测的mAP和FPS分别为84.1%和26.6 f/s,优于其他目标检测算法,满足工业检测的要求。

表 4 实验结果对比

Table 4 Comparison of experimental results

算法	Backbone	mAP[0.5,0.95]	mAP[0.5]	Param/ 10^6	FLOPs/ 10^6	FPS
YOLOV5s	Our	84.1	99.5	12.1	13.5	26.6
	CSPDarknet	83.3	99.4	28.7	30.9	12.1
Retinanet	Mobilenetv2	78.9	95.8	2.9	2.8	20.8
	SSDVGG 300	80.4	96.8	31.8	178	6.2
	Resnet 50	82.5	98.6	68.6	76.5	7.1

3.3 算法训练过程及结果

损失变化曲线反映了神经网络模型学习和拟合的整个过程和趋势。目标检测算法常用的损失函数有 CIOU(预测框位置损失), Obj(预测框置信度损失), Cls(分类损失)。选择 mAP 指标最优的一

轮作为最终的训练结果,各项指标如图 6-7 所示。从图 6 可以发现,在训练和验证的初始阶段,损失函数下降速度非常快;随着训练的推进,下降趋势逐渐平缓,曲线开始起伏;当训练超过 300 次后,曲线趋于平稳,损失函数逐步收敛完成,训练完成。

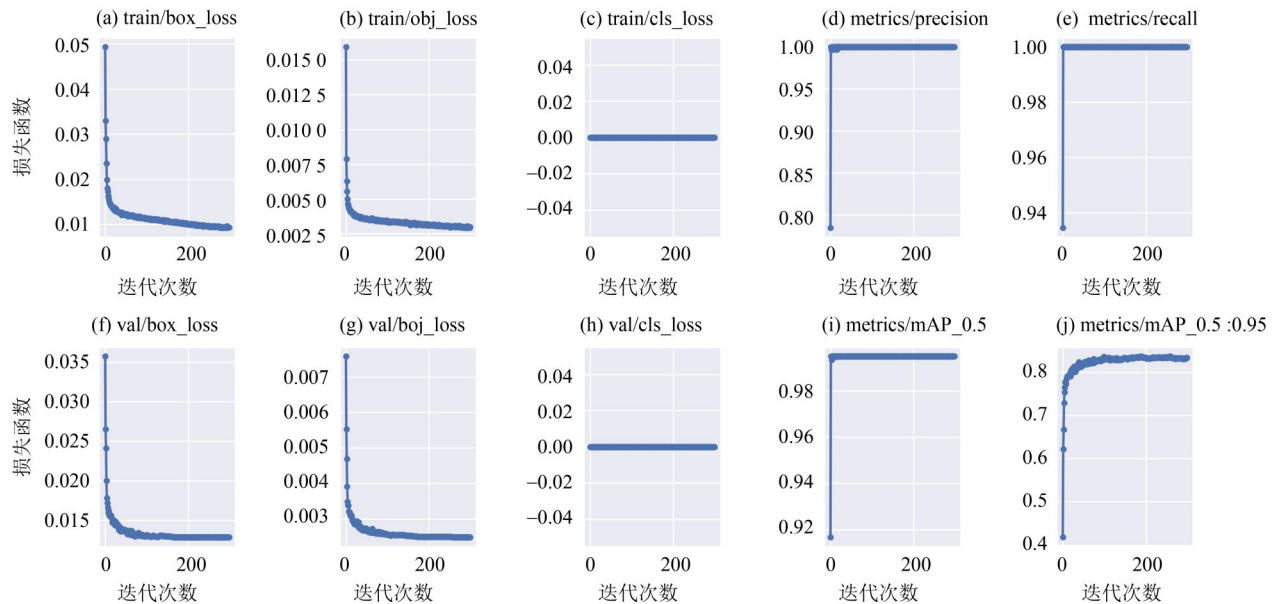


图 6 训练过程及结果

Fig. 6 Training process and results

3.4 模型转化与推理实验结果

模型训练主要是获得层与层之间的权重参数,以及一个性能优异的模型,关注点是模型的准确度、精度等。推理则不一样,其没有了训练中的反向迭代过程,只是针对新的数据进行预测。相较于训练,推理更关注的是部署简单、处理速度快、吞吐率高和资源消耗少。TensorRT 是一个高性能的深度学习推理优化器,可以为深度学习应用提供低延迟、高吞吐率的部署推理。将 Pytorch 训练好的模型(xx.pt)用 TensorRT 推理加速,先将模型由 Pytorch 保存的模型文件转换成 ONNX 模型,然后再将 ONNX 模型转换成 TensorRT 推理引擎(engine)文件,实现对模型的量化加速。将其部署和应用到边缘终端设备上,就可以快速地调用它来执行模型的加速推理,推理测试结果如图 8 所示。

4 结论

本文提出了一种改进的 YOLOV5 网络,通过融合 Repvit-MobileNet Block 与注意力机制模块,实现复杂场景下船舶除漆机器人的目标检测。其主要优势如下:

1) 基于 RepVit-MobileNet block 的骨干网络,运算参数显著下降,浮点计算量只有原模型的 30.9%,提高了检测运算速度。此外,RepVit-MobileNet block 不仅扩大了感受野,还获得了丰富、多层次特征表达的特征图,维持了较高的检测性能。

2) 由于运动过程中机器人的姿态变化,除漆机器人的形状会发生变化,这可能导致模型准确性的下降。通过加入位置注意力模块来增强机器人在全局信息中的特征提取,并忽略背景噪声,以提高检测准确性。再通过 CBMA 模块来增加关键重要特征的权重,增强了对目标信息的学习,减少了不重要特征的权重,并忽略了一些干扰信息,使网络更加专注于目标位置的识别,有效提高了目标的检测性能。

3) 为了更精准地拟合真实框和加快训练收敛速度,在网络损失函数中加入一种新的 Refine-IOU 损失函数,以增强模型对目标位置检测的稳定性。

4) 在船舶除漆机器人数据集上进行实验与验证。结果表明,所提出的方法在数据集上的平均精度为 84.1%,硬件部署后的推理速度达到了 26.6 f/s。检测精度和速度指标都满足应用要求。

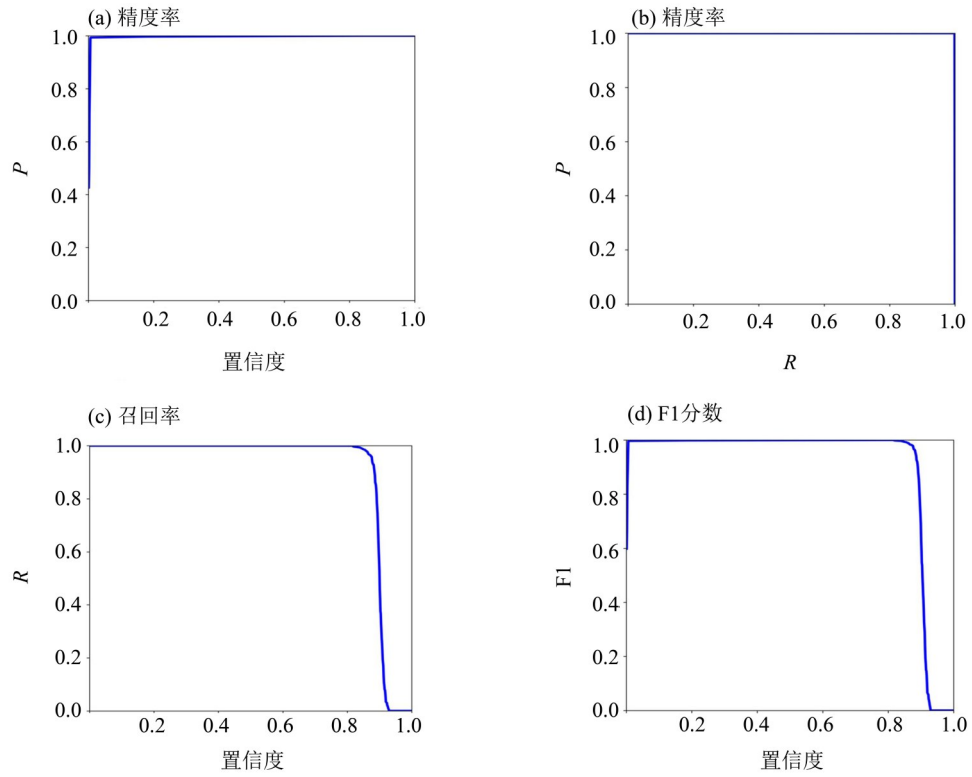


图 7 本文模型的性能曲线

Fig. 7 Performance curve of model in this paper

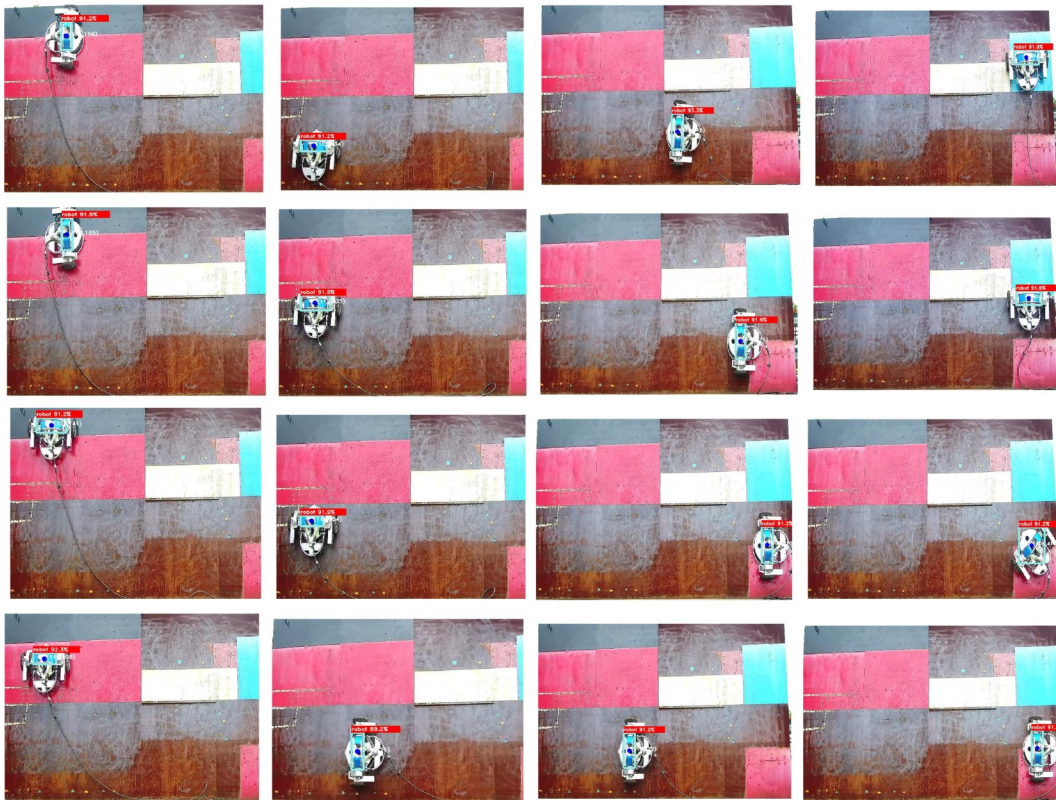


图 8 测试图片

Fig. 8 Test pictures

参考文献:

- 方璇, 刘俊锋, 陈勇, 等, 2023. 基于SLAM的爬壁机器人自主移动研究[J]. 制造业自动化, 45(6):85-88.
- 姜泽, 王珉, 赵哲, 等, 2023. 爬壁机器人发展现状与关键技术研究综述[J]. 包装工程, 44(12):29-38.
- 李磊, 杨幸, 秦绪杰, 等, 2023. 爬壁机器人研究现状及发展趋势[J]. 信息对抗技术, 1:1-6.
- 李希平, 舒勇, 魏莹吉, 等, 2024. 基于视觉辅助导航的船用除漆爬壁机器人设计[J]. 工程机械, 36(11):35-40.
- 张文, 丁雨林, 陈咏华, 等, 2022. 基于外部视觉与机载IMU组合的爬壁机器人自主定位方法[J]. 清华大学学报(自然科学版), 62(9):1524-1531.
- DING X, ZHANG X, MA X Y, et al, 2021. RepVGG: Making VGG-style ConvNets great again [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: 13728-13737.
- HARIS M, HOU J, 2020. Obstacle detection and safely navigate the autonomous vehicle from unexpected obstacles on the driving lane[J]. Sensors, 20(17):4719.
- HOWARD A, SANDLER M, CHEN B, et al, 2019. MobileNetV3: Searching for MobileNetV3 [C]//IEEE/CVF International Conference on Computer Vision. Seoul, Korea: 1314-1324.
- HOWARD A G, ZHU M L, BO C, et al, 2018. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. arXiv:1704.04861.
- HOU Q B, ZHOU D Q, FENG J S, et al, 2021. Coordinate attention for efficient mobile network design [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual:13708-13717.
- HU J, LI S, SUN G, et al, 2018. Squeeze-and-excitation networks [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA:7132-7141.
- LEI R, SHI C, RAN X, 2012. Target detection of maritime search and rescue: Saliency accumulation method [C]// International Conference on Fuzzy Systems and Knowledge Discovery. Chongqing, China:1972-1976.
- LI C Y, LI L L, JIANG H L, et al, 2024. YOLOV6: A single-stage object detection framework for industrial applications[J]. arXiv:2209.02976.
- LIN T Y, DOLLÁR P, GIRSHICK R, et al, 2017. Feature pyramid networks for object detection [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA:936-944.
- LV S D, ZHU L, WANG W W, et al, 2020. Improving SSD for detecting small target in remote sensing image [C]// Chinese Automation Congress. Shanghai, China: 567-571.
- SANDLER M, HOWARD A, ZHU M, et al, 2018. MobileNetV2: Inverted residuals and linear bottlenecks [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: 4510-4520.
- SZEGEDY C, LIU W, JIA Y, et al, 2015. Going deeper with convolutions [C]//IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA:1-9.
- WANG A, CHEN H, L Z J, et al, 2024. Rep ViT: Revisiting MobileNet CNN from ViT perspective [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA:15909-15920.
- WOO S, PARK J, LEE J Y, et al, 2018. CBAM: Convolutional block attention module [C]//Computer Vision—ECCV 2018. Munich, Germany: 3-19.

(责任编辑 王海蓉)