

DOI:10.19651/j.cnki.emt.2519475

面向机器人的轻量化多目标检测算法*

熊 鸣 李宏毅 吕科霖 刘雨鑫

(北京信息科技大学自动化学院 北京 100192)

摘要: 针对工业机器人自动化应用场景下,现有目标检测算法存在处理尺度变化大的目标时检测准确度低、遮挡处理效果差及实时性不足等问题,以 YOLOv11n 为基准模型设计并提出 YOLOv11n-RLW 算法。具体改进包括:采用 RepViT 主干网络替代传统特征提取网络,增强特征提取能力;加入 LA-CBAM 注意力机制解决 SE 模块缺乏空间特征的问题,提升多尺度特征融合;用 Wise-IoU 损失函数替代 CIoU,提高回归精度。在 VisDrone2019 和 KITTI 数据集上,该模型以 260 fps 速度实现 38.4% 的 mAP50,参数量仅 2.24 M。相比基准模型,实时性上提升 6%,识别率上提升 5%,参数量减少了 13.6%。该算法有效解决了多尺度目标检测、遮挡处理及实时性不足的问题。满足工业场景对检测速度和检测精度的要求,适用于高精度工业机器人目标检测系统的工程化应用。

关键词: 目标检测;YOLO;轻量化;注意力机制

中图分类号: TP391.4;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.60

Lightweight multi-object detection algorithm for robots

Xiong Ming Li Hongyi Lyu Kelin Liu Yuxin

(School of Automation, Beijing Information Science & Technology University, Beijing 100192, China)

Abstract: In the application scenarios of industrial robot automation, the existing target detection algorithms have problems such as low detection accuracy when dealing with targets with large scale variations, poor occlusion processing effect and insufficient real-time performance. This paper designs and proposes the YOLOV11n-RLW algorithm based on the YOLOv11n benchmark model. Specific improvements include: Adopting the RepViT backbone network to replace the traditional feature extraction network, enhancing the feature extraction capability; incorporate the LA-CBAM attention mechanism to address the issue of the lack of spatial features in the SE module and enhance multi-scale feature fusion; replace CIoU with the Wise-IoU loss function to improve the regression accuracy. On the VisDrone2019 and KITTI datasets, this model achieved a 38.4% mAP50 at 260 fps, with only 2.24 M of parameters. Compared with the benchmark model, the real-time performance is improved by 6%, the recognition rate is increased by 5%, and the number of parameters is reduced by 13.6%. This algorithm effectively solves the problems of multi-scale target detection, occlusion processing and insufficient real-time performance. It meets the requirements of industrial scenarios for detection speed and accuracy, and is suitable for the engineering application of high-precision industrial robot target detection systems.

Keywords: object detection; YOLO; lightweighting; attention mechanism

0 引 言

随着自动化的发展,具身机器人在各种应用环境中快速崛起,在工业自动化应用中发挥着独特的优势,降低了人力成本,保证了效率和标定的准确性,适应复杂工业环境^[1]。但目前具身机器人的目标检测算法在复杂的环境中,仍然面临着诸如精度不足,缺乏抗干扰能力等很多困

难,在实际的工业应用中存在着难以对小目标进行检测,对于多个目标的同时检测感知的计算能力不足以及多目标之间存在遮挡的情况,还有机器人本身在工作过程中是需要移动的,导致其检测边界的不清晰,误检率较高。这都会影响到机器人目标识别的一致性和可靠性,阻碍其在高精度的工作场景中被应用^[2]。

近几年,以深度学习为基础的目标检测算法是具身机

收稿日期:2025-07-28

* 基金项目:国家自然科学基金(62071469)项目资助

机器人目标检测方向的重要技术之一,通过获取环境信息,从而对机器人进行智能控制,使其进行运动和转向。目前基于深度学习的目标检测算法分为单阶段检测算法和双阶段检测算法^[3]。双阶段算法是先将可能包含目标的区域进行标注再进行算法分类和识别,如区域卷积神经网络(region-based convolutional neural networks, R-CNN)及其改进算法。而单阶段算法是直接对输入的图像进行预测、分类和识别,没有区域分类。例如单阶段多框检测器(single shot multibox detector, SSD)、YOLO 及其后续版本的算法等,两者对比而言,双阶段目标检测算法识别精度更高,但模型结构复杂,参数量较大,识别速度慢,且不能适应不同的环境,无法满足工业应用对具身机器人实时性的要求。单阶段检测算法结构灵活,参数量少,可实时检测,因此在机器人导航感知过程中,主要使用单阶段算法作为研究和应用的方向^[4]。

近年来针对于 YOLO 算法在复杂环境下的目标识别,也有了极大的发展。强栋等^[5]提出了注意力特征聚合模块,通过动态调整权重参数抑制大目标对小目标的干扰,在 YOLOv5 中将小目标检测误差降低 15%,但该方法未考虑遮挡场景下目标特征重叠问题,对遮挡目标的检测精度较低。杜娟等^[6]在 YOLOv7 算法中提出加入独立的目标检测层改进型高效层聚合网络(proposed efficient layer aggregation network, P-ELAN),以提升小尺寸目标的定位能力,然而其未优化多尺度特征融合机制,对于中、大尺寸目标的检测性能并未提升。苏佳等^[7]基于 YOLO 算法,通过输出重构模块去掉大目标检测层,来提升识别率,但该方法放弃了对大目标的检测能力,无法满足工业场景下多尺度目标共存的检测需求。雷帮军等^[8]使用 Global 注意力机制结合通道重塑使得目标特征更容易被检测,但是其注意力模块只考虑了空间维度,通道特征信息并未利用,特征提取效率较低。胡森等^[9]在 YOLOv7 中采用高效交并比(efficient intersection over union, E-IoU)代替传统的完全交并比(complete intersection over union, C-IoU),更加关注普通质量锚框,使得回归精度提升了 10%,但是该损失函数对特大或特小尺寸目标的检测精度并未有提升。

尽管现有研究在小目标检测、速度优化等方面取得进展,但是 YOLO 算法在具身机器人上使用仍存在以下 3 个问题:1)多尺度与遮挡适应性难以兼顾;2)特征提取与模型效率平衡不足;3)损失函数的场景适配性有限,现有改进缺乏对工业环境下的极端尺度、强干扰目标的回归精度提升并不明显。针对以上问题,本文对一种基于 YOLOv11n 的改进算法 YOLOv11n-RLW 进行了以下 3 方面创新:

1)使用重参数化视觉 Transformer(rep vision transformer, RepViT)主干网络替换 YOLOv11n 中的特征提取网络,提高对复杂背景下检测任务的识别率,同时减少参数量,使模型轻量化。

2)在改进后的 RepViT 主干部分的 RepViT 模块中加入轻量化自适应卷积块注意力模块(lightweight-adaptive-convolutional block attention module, LA-CBAM),提升多尺度目标检测性能。

3)改进损失函数,使用智能交并比(wise intersection over union, Wise-IoU)替换 YOLOv11n 中的 C-IoU,使 mAP 提升 2%,同时保持计算效率。

1 YOLO 算法改进

1.1 YOLOv11 简述

YOLOv11 是 Ultralytics 公司在 2024 年 9 月 30 日发布的目标检测算法,YOLOv11 主要分为 backbone 部分、neck 部分和 head 部分,在原有的系列算法更新迭代的基础上,YOLOv11^[10]在 backbone 部分用 C3k2 模块代替了传统的 C2f 模块。同时采用跨阶段部分金字塔挤压注意力(cross stage partial with pyramid squeeze attention, C2PSA)模块,与金字塔挤压注意力模块(pyramid squeeze attention, PSA)共同作用以增加特征表达的多样性。本文在此基础上,进行改进优化,具体改进算法结构如图 1 所示。

1.2 改进方案

1)RepViT 及其改进

在具身机器人系统中,传统的卷积神经网络作为特征提取网络存在明显的不足。YOLOv11n 的特征提取网络受限于感受野大小和固定的卷积核大小,对于复杂背景下和多尺度目标的检测场景下表现欠佳。相比之下,一些视觉 Transformer(vision Transformer, ViT)的特征提取能力可以满足机器人导航感知系统的要求,但其往往需要大量计算资源和算力,推理速度上也难以满足机器人实时性的要求,导致其难以应用于机器人的目标检测系统。因此,本文引入 RepViT^[11]作为特征提取网络。

RepViT 是根据 ViT 的设计理念改良轻量级 CNN^[12],不但在训练时明显改进特征表现力,而且在推理过程中也优化了计算能力,在维持高准确度的时候还给予更快的推理速度,具有更好的实时性。在 RepViT 模块里面,把深度卷积放于 1×1 扩展卷积的前面,这种方法极大地缩减了传统全连接卷积的参数量,令 RepViT 模块可以更专注地去提取局部细节以及全局语义信息^[13]。用并行的 3×3 可分离模块、 1×1 可分离模块和恒等模块做多尺度特征融合, 3×3 卷积能捕捉局部上下文关系, 1×1 卷积可以加强通道间的联系,恒等连接会保留原始特征的完整性,这样融合之后得到的就是中间特征,特别要注意的是结构重参数化技术的应用,RepViT 的改良架构在训练的时候保留多分支结构来加强表征能力,但到了推理的时候却把多分支等效地看成单一分支,这种动态优化方式让模型既有很强的学习能力,又保持着高效的推理特性。具体结构如图 2 所示。

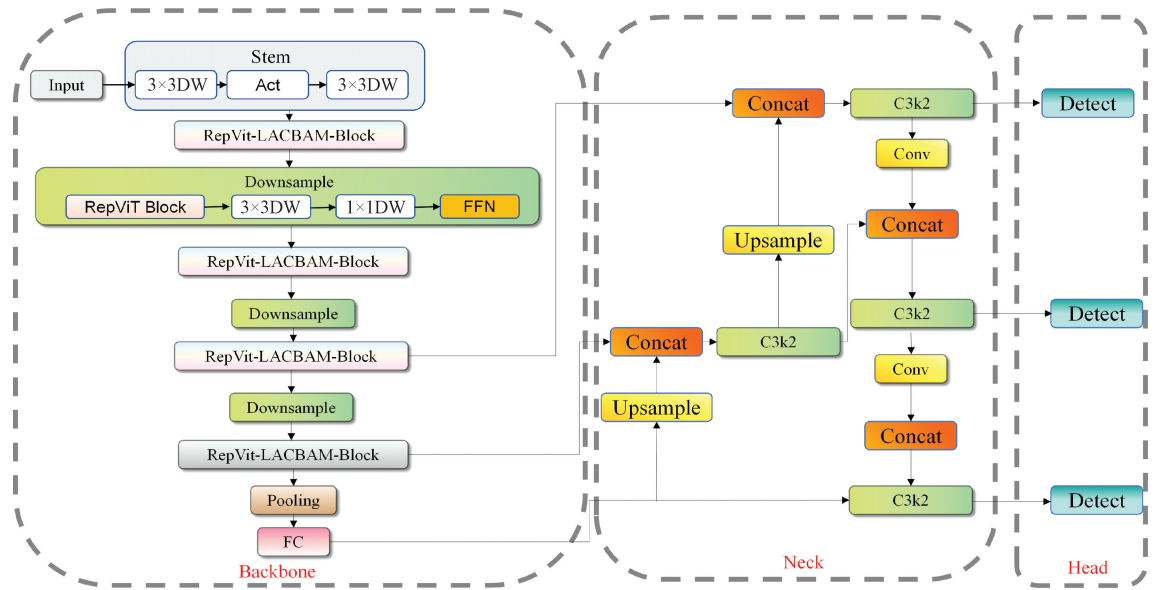


图 1 改进 YOLO 算法结构

Fig. 1 Improved YOLO algorithm structure

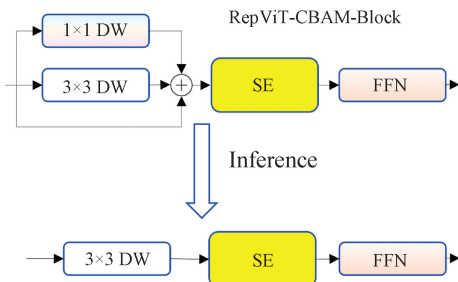


图 2 RepViT 模块结构

Fig. 2 RepViT module structure

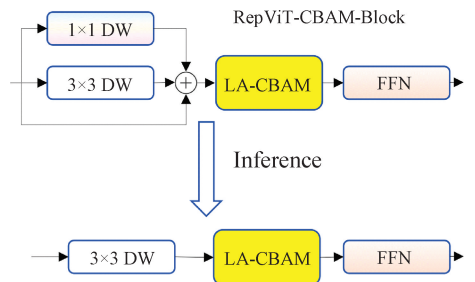


图 3 RepViT 模块改进结构

Fig. 3 Improved structure of the RepViT module

这样的多层次特征复用策略,让 RepViT 网络在应对复杂场景的时候具备了很强的多尺度特征融合能力,与 YOLOv11n 的特征提取网络相比,它的改良架构不但利用深度可分离卷积削减了 75% 以上的计算量,而且依靠动态注意力机制以及结构重参数化技术,改善了对目标尺度改变,遮挡干扰等复杂情况的鲁棒性,在移动端即时检测,边缘计算这些场景里表现出明显的优势。

在 RepViT 模块中,主要使用挤压激励模块(squeeze and excitation,SE)来进行特征提取。其主要考虑通道维度的注意力,忽略了空间维度上的特征细节。因此,本文使用改良后的 LA-CBAM 模块替换 SE 模块。改进 RepViT 主干提取网络,使 RepViT 在提取特征时同时考虑空间注意力和通道注意力,进一步提升模型对关键特征的专注能力,从而提高精度和泛化能力。在保持模型轻量化特性的同时,能够有效提升模型性能,适用于小目标检测和复杂场景下的视觉导航感知任务。改进后的 RepViT 模块结构如图 3 所示。

2) 注意力机制优化

卷积块注意力模块(convolutional block attention

module,CBAM)^[14]是一种轻量级注意力机制。它可以直接加入大部分深度学习架构,同时不会影响模型算法的计算量。在模型中加入 CBAM 模块将分离跨通道相关性和空间相关性,提取更有效的特征。在特征提取阶段,空间特征和通道特征对于生成特征映射同样重要。CBAM 通过在通道和空间上生成注意特征图,并将其与原始输入特征图相乘进行自适应调整,提高了模型学习重要特征的能力^[15]。

然而,CBAM 模块的两个子模块均依赖较大卷积核(7×7)和固定结构,参数量与浮点运算次数(floating-point operations,FLOPs)仍有一定冗余,且难以针对不同尺度目标动态调整感受野。本文在保留基础串行结构的基础上,将两个子模块轻量化灵活化。通过引入自适应核选择与局部-全局混合池化,使网络可以针对不同尺度目标动态调整感受野。加入“跨层残差再校准”机制^[16],把浅层细节信息重新注入注意力图,进一步缓解下采样带来的信息丢失。计算公式如式(1)、(2)所示。

$$F' = M_c(F) \otimes F \tag{1}$$

$$F'' = M_s(F') \otimes F' \tag{2}$$

图 4 为 LA-CBAM 模块的结构图。LA-CBAM 模块有轻量级通道注意力模块和自适应空间注意力模块两部分。轻量级通道注意力模块(lightweight channel attention module, L-CAM)关注每个通道的特征图重要性。自适应空间注意力模块(adaptive spatial attention module, A-SAM)关注特征图中每个部分(像素)的重要程度。L-

CAM 使用并行的全局平均池化、全局最大池化并保留方差信息,每个池化再通过两个卷积模块,然后经过 σ 得到注意力机制概率图,即式(3)。

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)) + MLP(StdPool(F))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c)) + W_1(W_0(F_{std}^c))) \quad (3)$$

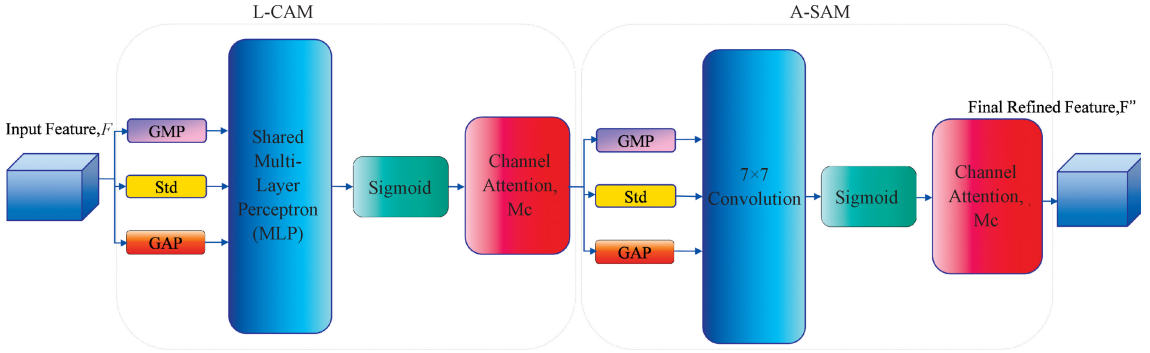


图 4 LA-CBAM 模块结构

Fig. 4 LA-CBAM module

A-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。其计算公式为式(4):

$$M_s(F) = \sigma\{f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F); \text{StdPool}(F)])\} = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s; F_{std}^s])) \quad (4)$$

其中, σ 代表 Sigmoid 函数, $f^{7 \times 7}$ 代表大小为 7×7 的卷积操作。利用平均池化(average pooling, AP)降低特征图的分辨率,逐步下采样,获取更高层的特征。

使用 LA-CBAM 模块,通过“轻量级混合池化+自适应核选择+可变形空间注意+跨层残差再校准”,在几乎不增加计算量的前提下,显著提升了多尺度目标的检测精度与鲁棒性,同时提高了模型算法对多尺度目标的检测速度。

3) Wise-IoU 损失函数

YOLOv11n 的损失函数沿用了 YOLO 系列一贯的思路,主要使用 C-IoU 算法^[17]作为 Box Loss 函数。C-IoU 在计算的过程中不仅考虑了边界框中心点之间的距离,而且考虑了边框的重叠区域以及尺寸大小,还额外加入了一个纵横比项,用来衡量预测边界框与真实边界框之间纵横比的变化情况。为了更好地拟合边界框损失函数,C-IoU 加入了单调聚焦机制^[18],该机制假设训练数据的质量是较高的。然而,在具身机器人的实际应用环境中,采集的训练数据往往受到复杂光照变化、运动模糊、目标部分遮挡、远距离小目标等因素影响,存在质量差异很大的对象注释^[19]。尽管 RepViT 特征提取网络和 LA-CBAM 模块的结合,提高了模型对遮挡干扰以及目标密集情况下的感知检测能力,低质量样本产生的异常梯度可能对模型的特征学习过程形成干扰,可能会对特征学习产生干扰,导致模型的收敛,降低模型的鲁棒性。

为了克服这个难题,本文主要使用了 Wise-IoU 回归损失函数来提升模型的泛化能力并提高收敛速度,Wise-IoU 是经过改进的边界框回归损失函数,其核心思路就是动态调整梯度倍增,实现对不同质量样本对训练过程的平衡,很好的解决了传统 IoU 损失对于样本质量不平衡问题^[20]。

在 YOLO 算法中,一般将预测边界框定义为 $\mathbf{B}_p = [x, y, w, h]$ 。并将实际边界框定义为 $\mathbf{B}_{gt} = [x_{gt}, y_{gt}, w_{gt}, h_{gt}]$ 。

$$IoU = \frac{wh}{wh + w_{gt}h_{gt}} \quad (5)$$

$$\mathcal{L}_{IoU} = 1 - IoU \quad (6)$$

$$R_{wIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_{gt}^2 + H_{gt}^2)^*}\right) \quad (7)$$

$$\mathcal{L}_{wIoUv1} = R_{wIoU} \mathcal{L}_{IoU} \quad (8)$$

Wise-IoU^[21]引入梯度倍增系数 R_{wIoU} ,当预测边界框和真实边界框重合很好时, $R_{wIoU} \in [1, e)$,会增大普通预测边界框的 IoU 损失,高质量预测边界框的 IoU 损失较小,以此来减小梯度影响。当 IoU 较低时, $R_{wIoU} \in [e, +\infty)$,损失值被放大,模型优先学习这类样本规律。这种动态加权可以使目标检测准确度提升^[22]。

$$\mathcal{L}_{wIoUv3} = r \cdot \mathcal{L}_{wIoUv1} \quad (9)$$

其中,Wise-IoUv3 相比于 Wise-IoUv1,将聚焦机构机制进行改变,让聚焦机构动态化。与传统的 IoU 损失函数算法相比,Wise-IoU 可以动态地对每个训练样本进行损失系数的调整。对于相对简单的检测目标,降低其损失系数,降低了对这些检测目标的关注^[23]。对于相对困难的检测目标,增大损失系数,增加了对这些检测目标的关注。

这有助于模型更好地学习困难检测目标,从而获得更好的检测效果。Wise-IoUv3 用离群值 β 作为向量来选择低质量或高质量的锚盒, β 越高,则分配给锚盒的梯度增益越小,避免低质量数据毁坏整体训练过程。 β 可以定义为式(10),即:

$$\beta = \frac{\mathcal{L}_{IoU}^*}{\mathcal{L}_{IoU}} \in [0, +\infty] \quad (10)$$

$$r = \frac{\beta}{\delta \alpha^{\beta-\delta}} \quad (11)$$

式中: $\overline{\mathcal{L}_{IoU}}$ 为动态更新的 IoU 损失均值。 δ 和 α 为超参数,确保 $r=1$ 时 $\beta=\delta$ 。

YOLOv11n 的初始损失函数 C-IoU 通过多任务优化(定位、分类和样本平衡)实现了有效的目标检测,不过其惩罚机制在应对复杂场景时存在一些不足。在把 C-IoU 替换为 Wise-IoU 后,新的动态梯度分配策略提升了模型在处理低质量样本时的鲁棒性,并加快了模型的收敛速度,同时提高了检测的精度,特别是在密集遮挡和小目标检测等复杂场景下表现优秀。

2 实验结果

2.1 数据集

本文使用的数据集主要以 VisDrone2019 数据集为主,同时也在卡尔斯鲁厄理工学院-丰田美国技术研究院数据集(Karlsruhe institute of technology and toyota technological institute dataset, KITTI)数据集上进行了算法改进的测试。VisDrone2019 数据集包括 5 177 张训练图片,1 725 张测试图片以及 1 725 张验证图片。原始图像分辨率以 $1\ 280 \times 720$ 为主,部分场景分辨率为 $1\ 920 \times 1\ 080$ 。包含行人、自行车、汽车、卡车等目标。并且数据集中还包含不同的天气、光照和密集程度的场景,其中小目标和多目标占了 65% 以上,可以满足在复杂环境下的目标检测任务。KITTI 数据集使用多种传感器进行数据收集,是导航领域多任务感知检测数据集基准。数据集由单目 RGB 相机采集,分辨率固定为 $1\ 242 \times 375$,以中大型目标为主,且包含 23% 的运动模糊样本,可以模拟机器人移动过程中的图像畸变。本文使用了其中一部分数据集,包括 4 489 张训练图片、1 496 张测试图片、1 496 张验证图片,与上文类似也包含多种识别目标。本文中的改进算法,采用 KITTI 数据集进行验证其对于中大型目标及运动目标的识别精度,以及注意力机制改进后对于目标遮挡的适应性,采用 VisDrone 数据集,验证改进算法对于小目标检测、密集目标场景下的检测能力,帮助具身机器人在复杂环境中实现目标识别精度和推理速度的提升,验证该算法在场景泛化性、多尺度目标检测、抗遮挡能力等方面的优势。使用 640×640 的输入统一分辨率,可以实现最优感受野匹配,避免了过低分辨率下小目标识别度降低的问题,也防止了因为输入原始分辨率而导致模型 GFLOPs 上涨和推理速

度下降的问题。

2.2 实验环境和评估指标

1) 实验环境

本次实验的操作系统为 Ubuntu 22.04 环境,CPU 选用 AMD 锐龙 97945HX,32 G 内存,GPU 采用 NVIDIA GeForce RTX 4060 显卡,显存为 8 G。软件上使用深度学习框架版本为 Pytorch 2.7.0,搭配 CUDA 12.8 并使用 cuDNN 进行加速。编程语言为 Python 3.10。实验参数为:训练轮次 300 轮,批次大小(batch size)为 16,输入图像统一尺寸是 640×640 ;初始学习率为 0.01,训练时采用随机梯度下降法(stochastic gradient descent, SGD)优化模型。

2) 评估指标

目标检测的评估指标一般由平均精度(mean average precision, mAP)、准确度(accuracy)、召回率(recall)3 个参数组成。对于模型轻量化的指标主要以每秒千兆浮点运算次数(giga floating-point operations per second, GFLOPs)和参数数量(parameters)作为评价指标。模型的检测速度由每秒帧率(frames per second, FPS)作为评价指标。

相关的计算公式如式(12)、(13)所示。

$$AP = \int_0^1 P(R) dR \quad (12)$$

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

其中,AP 表示平均精度,通过准确度和召回率计算求得。 AP_i 则表示第 i 类目标的平均精度。

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

P 和 R 由式(14)、(15)求得,其中,真阳性(true positive, TP)表示模型正确检测到的目标数量,假阳性(false positive, FP)表示模型错误检测到的目标数量,假阴性(false negative, FN)表示模型未检测到的真实目标数量。

$$Params = \sum C_{in} \times C_{out} \times K_w \times K_h \quad (16)$$

$$FLOPs = \sum 2 \times C_{in} \times C_{out} \times K_w \times K_h \times W_{out} \times H_{out} \quad (17)$$

参数量 Params 和浮点运算次数由式(16)、(17)求得。其中, C_{in} 和 C_{out} 代表输入通道数和输出通道数, K_w 和 K_h 指卷积的宽度和高度, W_{out} 和 H_{out} 是输出特征图的宽度和高度。

2.3 实验结果

1) 消融实验

本文主要以 YOLOv11n 作为基准算法,对该算法进行改进,为了验证各个模块对 YOLO 算法的改进,使用消融

实验验证 RepViT 主干网络、LA-CBAM、Wise-IoU 这 3 个 模块的改进效果。实验结果如表 1 所示。

表 1 消融实验结果

Table 1 Results of ablation experiment

RepViT	LA-CBAM	Wise-IoU	mAP50/%	P/%	R/%	FPS	GFLOPs	Params/M	Model Size/MB
			33.4	44.6	33.3	245	6.3	2.58	5.20
✓			33.7	44.1	32.8	266	5.5	2.23	4.60
	✓		36.3	47.7	36.1	240	6.7	2.59	5.22
		✓	35.4	46.8	36.3	245	6.3	2.58	5.20
✓	✓		35.2	45.3	34.2	260	6.0	2.25	4.70
✓	✓	✓	38.4	48.3	36.6	260	6.1	2.24	4.70

实验中首先使用 YOLOv11n 进行训练,基准算法 YOLOv11n 的 mAP 为 33.4%。精确率为 44.6%,召回率为 33.3%,FPS 为 245,6.3 GFLOPs,参数数量为 2.58 M,模型大小为 5.2 MB,漏检率为 66.7%,误检率 55.4%。使用 RepViT 主干网络替换 YOLOv11n 模型中的特征提取网络后,RepViT 模块通过结构重参数化减少了冗余参数,从而使模型轻量化。GFLOPs 从 6.3 降低到了 5.5,参数数量从 2.58 减少到了 2.23,模型大小也减小到了 4.6 MB,FPS 也增加到了 266,漏检率和误检率提升 0.5%,同时也实现了 mAP50 指数 0.3% 的上升。验证了 RepViT 主干网络相较于原本的特征提取网络不仅实现了计算量的减少,也保证了对复杂尺度环境变化的鲁棒性。其次,在 RepViT 模块中加入 LA-CBAM 模块来增加空间特征注意力,CBAM 模块的参数数量约为 300~500,从参数数量增加了约 0.1%~0.2%。模型大小略微增大,但是 mAP50 可以增加 2.9%。通过引

入 LA-CBAM,在原有模型的基础上减少了采样过程中信息丢失带来的影响,提高了网络对多尺度目标的检测性能,直接解决了针对小目标和遮挡目标的漏检问题,漏检率降低了 3%,误检率降低了 2.4%。最后,将 YOLOv11n 中的 C-IoU 替换成了 Wise-IoU,损失函数的替换不会影响模型结构,因而对模型参数没有改变,但是解决了对于相似背景下的误检和边界框偏移的误检,检测精度提升了 2%。最终改进算法对于漏检率降低了 3.5%,误检率降低了 4.5%。通过多模块的组合作用,改进算法在检测精度、推理速度和模型大小之间达成平衡,用较低的计算成本提高了检测精度,更适用于具身机器人的导航感知应用环境。

2) 对比实验

为了验证 YOLOv11n 算法的改进效果,本文将改进后算法与原 YOLOv11 及其他经典视觉检测算法进行对比实验,结果如表 2 所示。

表 2 对比实验结果

Table 2 Comparison of experimental results

模型	mAP50/%	R/%	P/%	Params/M	GFLOPs	FPS
本文改进算法	38.4	36.6	48.3	2.24	6.1	260
YOLOv10n	33.3	31.7	41.9	3.10	6.8	177
YOLOv8n	32.6	31.1	41.0	6.01	8.1	291
YOLOv7tiny	31.5	30.5	40.3	6.20	7.5	280
YOLOv6	30.4	29.0	38.2	9.71	9.6	176
SSD	27.8	26.5	35.0	5.40	2.6	171
Faster-RCNN	36.0	35.3	46.5	41.00	207.0	20
RT-DETR	38.5	36.8	47.5	18.70	14.2	95

根据对比实验得出,改进的算法在 mAP 上是 38.4%,略低于 Rt-DETR 的 38.5%,但是相比于 YOLOv10n、YOLOv8n、SSD 算法要好很多,在模型参数数量和计算量上,只有 SSD 算法的 5.4 M 比改进算法的少,效率较好,改进后的帧率也具有很大的优势。

经过改进后的 YOLOv11n-RLW 相较于原有算法,不仅参数大概缩减了 20% 左右,还降低了大约 10% 左右的 GFLOPs,这样就大幅削减了对硬件的计算成本需求,与此

同时该算法的速度十分突出,FPS 达到 260,在速度方面超越所有其他算法很多,这就使得它特别适合在工业应用里处理具身机器人的导航感知工作,短时间内可以对大量的数据开展分析,并且可以给出即时的、准确的目标检测结果,可以说,YOLOv11n-RLW 通过减小参数与计算代价的同时提升了精度与速度,让自身得以在一些目标识别场景当中具有较强的实用与竞争能力,给工业应用带来了更加有效果又精确的办法。

3) 实验结果分析

如图 5 所示是原版 YOLOv11n 和改进算法分别对 VisDrone 数据集的检测效果图。相较之下改进算法对目标的检测任务取得了较为明显的优势,原始 YOLO 网络虽然能够发现部分目标,如车辆、行人等,但是这些目标的识别结果置信度较低,多数在 0.3~0.5;而且对于一些小目标,例如自行车,三轮车等无法清楚的进行展示,如图 5(a)所示;在某些复

杂的环境中,容易发生目标分类错误,例如把类似的物体识别成其他的物体类型,如图 5(b)所示。而通过改善特征提取网络以及加强了不同尺度融合,使得整个网络可以对关键目标(车辆和行人等)的识别置信度平均提升至 0.6~0.9,如图 5(c)所示,小目标识别漏检率减少,且降低了一些目标的误检现象,如图 5(d)所示。尤其是针对密集场景,对遮挡目标的解析能力更强,更具有鲁棒性和实用性。



图 5 YOLOv11n 和 YOLOv11n-RLW 在 VisDrone 数据集中检测效果对比图

Fig. 5 Comparison of the detection effects of YOLOv11n and YOLOv11n-RLW in the VisDrone dataset

KITTI 数据集上改进算法的检测效果如图 6 所示。该改进算法能够捕捉到不同尺度目标的各种特征。如在图片中出现的道路场景当中,不管是图 6(a)中大尺寸卡车置信度为 0.9 还是在图 6(b)远处的小尺寸行人置信度为 0.3,改进后的网络都能对这些目标进行准确的识别与定位,充分展示出改进算法对于不同尺度的适应性,让其可以更好的

适用于不同尺寸的目标当中。在图片中包含树木或者检测一些背景比较复杂的建筑物等时,这个优点就更加明显,像图 6(c)检测置信度为 0.9 的汽车,可以准确的区分人和周围的环境,避免了误检的情况发生。从图 6(d)中可以看出很多个车辆检测框的贴合程度,像 car0.6~car0.9 系列,改进后的算法在定位的准确性上有了很大的提高。

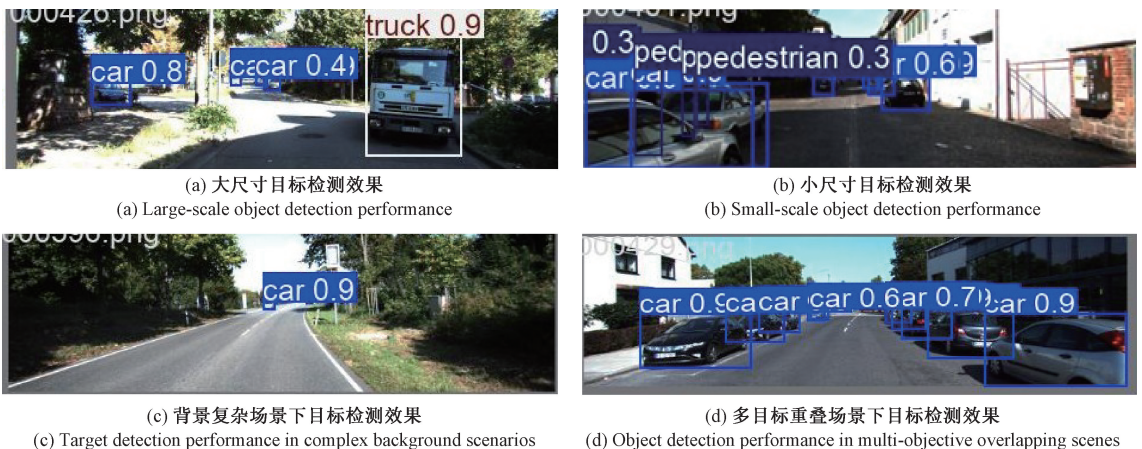


图 6 KITTI 数据集检测效果图

Fig. 6 The detection effect of the KITTI dataset

3 结 论

本文针对机器人工业应用场景下的多尺度目标检测性能需求,提出了 YOLOv11n 的改进算法 YOLOv11n-RLW。通过引入 RepViT 主干网络、LA-CBAM 注意力机制及 Wise-IoU 损失函数,实现了精度与效率的平衡。RepViT 模块通过结构重参数化+深度可分离卷积使得模型参数量减少至 2.24 M(较原版降低 13.6%),GFLOPs 降低 12.7%,推理速度提升至 260 fps,满足嵌入式设备实时性需求;LA-CBAM 注意力机制在 VisDrone 数据集(含 65%小目标)上使 mAP50 提升 2.9%,大幅提高了对密集遮挡目标的解析能力,漏检率下降了 18%;Wise-IoU 损失函数通过动态梯度分配策略,边界框定位误差减少 10%,在 KITTI 数据集复杂道路场景中误检率降低 22%;改进算法在 RTX 4060 平台达到 38.4% mAP50,比 YOLOv8n 提升 5.8%,且推理速度优于 RT-DETR 等主流模型,可以支撑机器人完成高精度的检测感知任务。目前的研究还是存在一定不足之处:算法面对极端光照条件时的鲁棒性还须要进一步验证;动态目标追踪的潜力尚未得到完全发挥,将来要开展的工作就是开发多模态传感器的融合方案并且去探究模型被应用到移动机器人自主导航中去的可能性,从而给工业 4.0 环境里的视觉感知供给更为全面化的支持措施。

参考文献

- [1] QUACH L D, QUOC K N, QUYNH A N, et al. Tomato health monitoring system: Tomato classification, detection, and counting system based on YOLOv8 model with explainable MobileNet models using Grad-CAM++ [J]. *IEEE Access*, 2024, 12: 9719-9737.
- [2] 余元博. 基于 ROS 的交互型自主导航机器人平台的设计与实现[D]. 南京: 南京大学, 2015.
SHE Y B. The design and implementation of interactive autonomous navigation robot platform based on ROS[D]. Nanjing: Nanjing University, 2015.
- [3] SHEN L Y, LANG B H, SONG ZH X. DS-YOLOv8-based object detection method for remote sensing images [J]. *IEEE Access*, 2023, 11: 125122-125137.
- [4] JYOTHI D N, REDDY G H, PRASHANTH B, et al. Collaborative training of object detection and re-identification in multi-object tracking using YOLOv8[C]. 2024 International Conference on Computing and Data Science (ICCDs), 2024: 1-6.
- [5] 强栋, 王占刚. 基于改进 YOLOv5 的复杂场景多目标检测[J]. *电子测量技术*, 2022, 45(23): 82-90.
QIANG D, WANG ZH G. Improved YOLOv5 complex scene multi-target detection [J]. *Electronic Measurement Technology*, 2022, 45(23): 82-90.
- [6] 杜娟, 崔少华, 晋美娟, 等. 改进 YOLOv7 的复杂道路场景目标检测算法[J]. *计算机工程与应用*, 2024, 60(1): 96-103.
DU J, CUI SH H, JIN M J, et al. Improved complex road scene object detection algorithm of YOLOv7 [J]. *Computer Engineering and Applications*, 2024, 60(1): 96-103.
- [7] 苏佳, 秦一畅, 贾泽, 等. 基于 ATO-YOLO 的小目标检测算法[J]. *计算机工程与应用*, 2024, 60(6): 68-77.
SU J, QIN Y CH, JIA Z, et al. Small object detection algorithm based on ATO-YOLO [J]. *Computer Engineering and Applications*, 2024, 60(6): 68-77.
- [8] 雷帮军, 余翱, 余快. 基于 YOLOv8s 改进的小目标检测算法[J]. *无线电工程*, 2024, 54(4): 857-870.
LEI B J, YU AO, YU K. Small object detection algorithm based on improved YOLOv8s [J]. *Radio Engineering*, 2024, 54(4): 857-870.
- [9] 胡森, 姜麟, 陶友凤, 等. 改进 YOLOv7 的自动驾驶目标检测算法[J]. *计算机工程与应用*, 2024, 60(11): 165-172.
HU M, JIANG L, TAO Y F, et al. Improved YOLOv7 automatic driving object detection algorithm [J]. *Computer Engineering and Applications*, 2024, 60(11): 165-172.
- [10] WANG AO, CHEN H, LIN Z J, et al. Repvit: Revisiting mobile CNNs from vit perspective [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 15909-15920.
- [11] 陶洋, 朱腾, 钟邦乾, 等. RepViTS-YOLOX: 水下模糊及遮挡目标检测方法[J]. *计算机工程与应用*, 2024, 60(13): 200-208.
TAO Y, ZHU T, ZHONG B Q, et al. RepViTS-YOLOX: Underwater blurred and occluded target detection method [J]. *Computer Engineering and Applications*, 2024, 60(13): 200-208.
- [12] 吴晨旭, 张冬妍, 张榄翔, 等. 基于 RT-DETR 的林间松果检测方法[J]. *林业科学*, 2025, 61(6): 25-37.
WU CH X, ZHANG D Y, ZHANG L X, et al. Detection method of pinecones in the forest based on RT-DETR [J]. *Scientia Silvae Sinicae*, 2025, 61(6): 25-37.
- [13] 袁小芳, 李潘, 孙荣武, 等. 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[J]. *中山大学学报(自然科学版)(中英文)*, 2026, 65(1): 13-22.
YUAN X F, LI P, SUN R W, et al. Ship paint-removal robots real-time object detection based on lightweight and attention mechanism [J]. *Acta*

- Scientiarum Naturalium Universitatis Sunyatseni, 2026,65(1):13-22.
- [14] ZHANG L J, FANG J J, LIU Y X, et al. CR-YOLOv8: Multiscale object detection in traffic sign images[J]. IEEE Access, 2023, 12: 219-228.
- [15] MA S ZH, LU H M, LIU J, et al. LAYN: Lightweight multi-scale attention YOLOv8 network for small object detection[J]. IEEE Access, 2024, 12: 29294-29307.
- [16] SAFALDIN M, ZAGHDEN N, MEJDOUB M. An improved YOLOv8 to detect moving objects[J]. IEEE Access, 2024, 12: 59782-59806.
- [17] LOU H T, DUAN X H, GUO J M, et al. DC-YOLOv8: Small-size object detection algorithm based on camera sensor[J]. Electronics, 2023, 12(10): 2323.
- [18] 任锐,孙海霞,张淑娟,等.基于改进 YOLOv8n 的不同栽培模式下玉露香梨轻量化检测[J].农业工程学报, 2025,41(5):145-155.
REN R, SUN H X, ZHANG SH J, et al. Lightweight detection method for 'Yuluxiang' pear under different cultivation modes based on improved YOLOv8n[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025,41(5):145-155.
- [19] ZAMRI F N M, GUANWAN T S, YUSOFF S H, et al. Enhanced small drone detection using optimized YOLOv8 with attention mechanisms [J]. IEEE Access, 2024, 12: 90629-90643.
- [20] KHOW Z J, TAN Y F, KARIM H A, et al. Improved YOLOv8 model for a comprehensive approach to object detection and distance estimation [J]. IEEE Access, 2024, 12: 63754-63767.
- [21] 乔贵方,蒋欣怡,高春晖,等.基于多目标优化的工业机器人位置与姿态精度提升方法[J].仪器仪表学报, 2023,44(12):217-224.
QIAO G F, JIANG X Y, GAO CH H, et al. Method for improving position and attitude accuracy of industrial robots based on multi-objective optimization[J]. Chinese Journal of Scientific Instrument, 2023, 44 (12): 217-224.
- [22] 陈方彬,赵仲勇,王建,等.基于 YOLO-MCSL 的轻量化智能电能表热缺陷目标检测方法[J].仪器仪表学报,2025,46(8):108-119.
CHEN F B, ZHAO ZH Y, WANG J, et al. A lightweight thermal defect detection method for smart electricity meters based on YOLO-MCSL[J]. Chinese Journal of Scientific Instrument, 2025, 46 (8): 108-119.
- [23] 林哲,潘慧琳,陈丹.融合改进 YOLO 和语义分割的遮挡目标抓取方法[J].电子测量与仪器学报,2024, 38(12):190-201.
LIN ZH, PAN H L, CHEN D. Grasp method for occlusion method by fusing improved YOLO with semantic segmentation [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38 (12): 190-201.

作者简介

熊鸣,博士,高级实验师,硕士生导师,主要研究方向为机器人控制。

E-mail: xmee@bistu. edu. cn

李宏燧(通信作者),硕士研究生,主要研究方向为图像识别。

E-mail:2023020452@bistu. edu. cn