

DOI:10.19651/j.cnki.emt.2416205

基于改进 YOLOv5 的仓储物体检测研究*

刘庭得¹ 王飞^{1,2} 刘旭¹ 蔡晨文¹

(1. 贵阳人文科技学院大数据与信息工程学院 贵阳 550025; 2. 贵州省模式识别与智能系统重点实验室 贵阳 550025)

摘要: 针对仓储复杂多样的环境以及对仓储物体传统检测模型性能较低的问题,本文基于 PaddlePaddle 框架,提出一种改进 YOLOv5 的仓储物体检测模型 YOLOv5-CE。首先,为提高在复杂多样环境下对仓储物体检测,采用 ConvNeXt 网络替换原 YOLOv5 的主干网络,以提高对中小型仓储物体的特征提取能力;其次,为提高模型的收敛速度和物体的检测精度,采用 EIoU Loss 代替原模型的损失函数;最后,通过利用自制仓储训练集进行多模型对比实验。实验结果表明,在对货物(cargo)、托盘(tray)、叉车(forklift)进行检测时,改进后的模型平均检测精度均值(mAP@0.5:0.95)达到 89.8%,相比原 YOLOv5 提升 1.1%,其中在小尺度仓储物体上,提升 4.2%;在中、大尺度仓储物体检测上,皆提升 1%;对于小仓储物体检测的平均召回率由 61.1%提升至 66.8%。与 YOLOv6、YOLOX、YOLOv7、Faster R-CNN 等其他模型相比,YOLOv5-CE 均体现出更好的精度。同时,相比上述模型,YOLOv5-CE 在模型参数量、检测速度以及检测精度上,也取得良好的平衡,能够较好地满足对不同大小、不同种类仓储物体的精准检测。

关键词: 仓储物体检测;YOLOv5;PaddlePaddle;ConvNeXt;EIoU Loss**中图分类号:** TP391.4;TN919.81 **文献标识码:** A **国家标准学科分类代码:** 520.60

Research on warehouse object detection based on improved YOLOv5

Liu Tingxun¹ Wang Fei^{1,2} Liu Xu¹ Cai Chenwen¹

(1. College of Big Data and Information Engineering, Guiyang Institute of Humanities and Technology, Guiyang 550025, China;

2. Key Laboratory of Pattern Recognition and Intelligent System, Guiyang 550025, China)

Abstract: In order to solve the problem of complex and diverse warehouse environment and the low performance of traditional warehouse object detection models, this paper proposes an improved YOLOv5 (You Only Look Once version 5) warehouse object detection model YOLOv5-CE (YOLOv5-ConvNeXt EIoU) which based on the PaddlePaddle framework. Firstly, to improve the detection of warehouse objects in complex and diverse environments, the ConvNeXt network is used to replace the original YOLOv5 backbone network to improve the feature extraction ability of small and medium-sized warehouse objects. Secondly, in order to improve the convergence speed of the model and the detection accuracy of objects, EIoU Loss (efficient intersection over union loss) is used to replace the loss function of the original model. Finally, by using the self-made warehousing training set to carry out multi-model comparison experiments. The experimental results show that when detecting cargo, tray and forklift, the average detection accuracy of the improved model (mAP@0.5:0.95, mean average precision@0.5:0.95) reaches 89.8%, which is 1.1 percentage points higher than the original YOLOv5, of which 4.2 percentage points is increased in small-scale warehousing objects; in the detection of medium and large-scale warehouse objects, it increased by 1 percentage point. The average recall rate for small warehouse objects increased from 61.1% to 66.8%. Compared with other models such as YOLOv6, YOLOX, YOLOv7, and Faster R-CNN, YOLOv5-CE all shows better accuracy. At the same time, in view of the above model, YOLOv5-CE also achieves a good balance in the number of model parameters, detection speed and detection accuracy, which can better meet the precise detection of warehouse objects of different sizes and types.

Keywords: warehouse object detection; YOLOv5; PaddlePaddle; ConvNeXt; EIoU Loss

0 引言

随着线上购物需求量的增加和智能化时代的到来,仓

储物流行业得到了快速发展。仓储物流的整个过程包含拣货、扫描包裹、配货、发货等各个环节,每个环节都需要大量的人工和自动化设备。近年来快速发展的计算机视觉极大

收稿日期:2024-06-07

* 基金项目:教育部产学研合作协同育人项目(221001766110209)、贵州省教育厅青年科技人才成长项目(黔教合 KY 字[2021]115)、贵阳人文科技学院科研基金项目(2022rwjs034,2023rwx038)、贵阳人文科技学院教学改革研究项目(20211365116)资助

程度地缓解了各环节对人工和自动化设备的依赖。通过计算机视觉与仓储环境管理相结合,对仓储物体的入库、出库、盘库等操作进行有效检测和比对,在降低劳动成本的同时,极大地促进了仓储环境智能管理的实现。

随着深度学习技术的快速发展,目标检测领域的大多数算法都是基于深度学习来实现的。目前,基于深度学习的目标检测算法主要分为单阶段目标检测算法和双阶段目标检测算法两大类。其中,单阶段目标检测算法主要以 Redmon 等^[1-3]提出的 YOLO(you only look once)系列,以及 Liu 等^[4]提出的单次多框检测器(single shot multibox detector,SSD)系列为主的算法,该类算法舍弃候选框的生成,只需要通过一次图像处理即可获得所需目标的位置和类别信息,可有效运用于实时性高的场景中。例如,王新等^[5]针对目标检测器检测跌倒时过于依赖卷积网络分类效果等问题,设计一种基于 YOLOv5s 和改进质心跟踪的跌倒检测模型。首先对 YOLOv5s 进行轻量化,同时将 MobileNetV3 网络中的 SE 模块替换为更高效的 ECA 模块;其次,引入哈希感知算法改进质心跟踪,增加目标关联的依据,提高跌倒检测的准确性。实验结果显示,改进 YOLOv5s 模型大小下降 52.2%,计算量下降 51.8%,精度高达 90.3%。双阶段目标检测算法主要以 Girshick 等^[6]提出的基于区域卷积神经网络(region with convolutional neural network,R-CNN)和 Ren 等^[7]提出的基于更快速的区域卷积神经网络(faster R-CNN)为主的基于候选区域的目标检测算法。例如,王飞等^[8]提出一种基于融合特征金字塔网络和 Faster R-CNN 的行人检测算法,该算法在 Caltech 数据集、KITTI 数据集和 ETC 数据集上的平均精确度均值(mean average precision, mAP)分别达到 69.72%,69.76%和 89.74%。与 Faster R-CNN 相比,该算法不仅提高了行人检测精度,而且在多尺度行人检测的问题上也获得了较为满意的检测效果。

当前,基于深度学习的仓储物体检测逐渐成为一个新兴的研究领域,但现有研究中针对仓储多物体检测的研究还相对较少。在国内,李映松等^[9]在 Faster R-CNN 模型的基础上,将 Resnet50 改进为 Swin Transformer 模型,平均准确率达到 90.6%,平均召回率达到 93.3%,平均检测速度达到 8.9 fps,较好地实现了仓储箱体的准确检测。陈亮杰等^[10]提出一种基于 SSD 的仓储物体检测算法,较好地实现了对仓储环境下的物体智能检测,但对于较小的仓储物体仍会出现漏检的情况,算法的鲁棒性不足。金秋等^[11]利用以 ZFNet 网络为改进后的 Faster R-CNN 物体检测算法基本实现了对仓储物体的粗定位,但由于 ZFNet 网络整体特征提取能力较弱,导致检测精度不足。在国外,Zaccaria 等^[12]提出一种基于卷积神经网络(convolutional neural networks,CNN)并结合决策制定步骤的方法,以解决使用单个 RGB 相机在工业环境中自动检测托盘的问题。Rutinowski 等^[13]利用深度学习技术实现对仓库中物流实

体的多视图姿态估计识别,以达到跟踪和定位的功能等研究。国外对仓储物体的检测更多偏向于仓储物体的瑕疵检测、与其他检测技术融合应用于仓储环境以及多目标的计数和识别等方面,对仓储物体的精准识别和对模型性能的研究还有所不足。

针对仓储复杂多样的环境以及对仓储物体传统检测模型性能较低的问题,经过测试发现 YOLOv7^[14]以及 YOLOv8 等高版本 YOLO 系列模型在本文自制的仓储物体数据集上所表现出的综合性能与 YOLOv5 差异不大,甚至低于 YOLOv5 模型。基于此,本文提出一种基于改进 YOLOv5 的仓储物体检测算法。首先,通过更换 YOLOv5 的主干网络,提升算法在复杂环境下对仓储物体特征的提取能力。其次,替换新的定位损失函数,提高模型的收敛精度,以得到鲁棒性更好的模型。

1 YOLOv5 算法

YOLOv5 是一种端到端的目标检测算法,网络结构主要由输入端(Input)、主干网络(Backbone)、颈部网络(Neck)和检测头(Head)四部分组成。

输入端:YOLOv5 模型采用 Mosaic 数据增强,通过将 4 张图像进行随机裁剪、缩放等方式最后拼合在一起,进一步提升了对不同类别和小物体的检测效果。相较于前期的 YOLO 系列模型,YOLOv5 模型不同于采用固定长宽比的锚框值,而是考虑到不同数据集的不同特点而顺应性调整计算出不同训练集中的最佳锚框值,利用此方法,相比于大多数更为先进的网络模型,其训练和推理速度都更具优势。

主干网络:YOLOv5 采用 CSPDarknet 作为其主干网络,其中的 CSP 模块被设计为 CSP1_X 和 CSP2_X 两种结构,两者分别作用于 YOLOv5 模型的主干网络以及颈部网络中。首先,CSP1_X 结构将特征图分为两部分,一部分直接进行卷积操作,另一部分先经过一系列变换后再与第一部分合并进行卷积,这种部分前向传播的方式有助于增加模型的宽度和深度,同时减少计算量。其次,CSP1_X 结构通过调整残差组件的数量以及卷积层数量,进一步优化了 Backbone 网络的特征提取能力。而 CSP2_X 结构与 CSP1_X 结构类似,也采用了部分前向传播的策略,利用进一步整合和优化从 Backbone 传递下来的特征图,并通过调整 Neck 部分的网络结构,提高了特征融合的效率 and 准确性。

颈部网络和检测头:由特征金字塔网络(Feature Pyramid Network,FPN)和路径聚集网络(Path Aggregation Network,PAN)组成模型的颈部网络,其中 FPN 和 PAN 分别自顶而下和自底而上对特征层进行上采样和下采样,并进行特征融合,最终将不同尺度的特征值融合,在预测层得到用于检测大、中、小 3 个类别尺度的物体预测结果。YOLOv5 网络结构如图 1 所示。

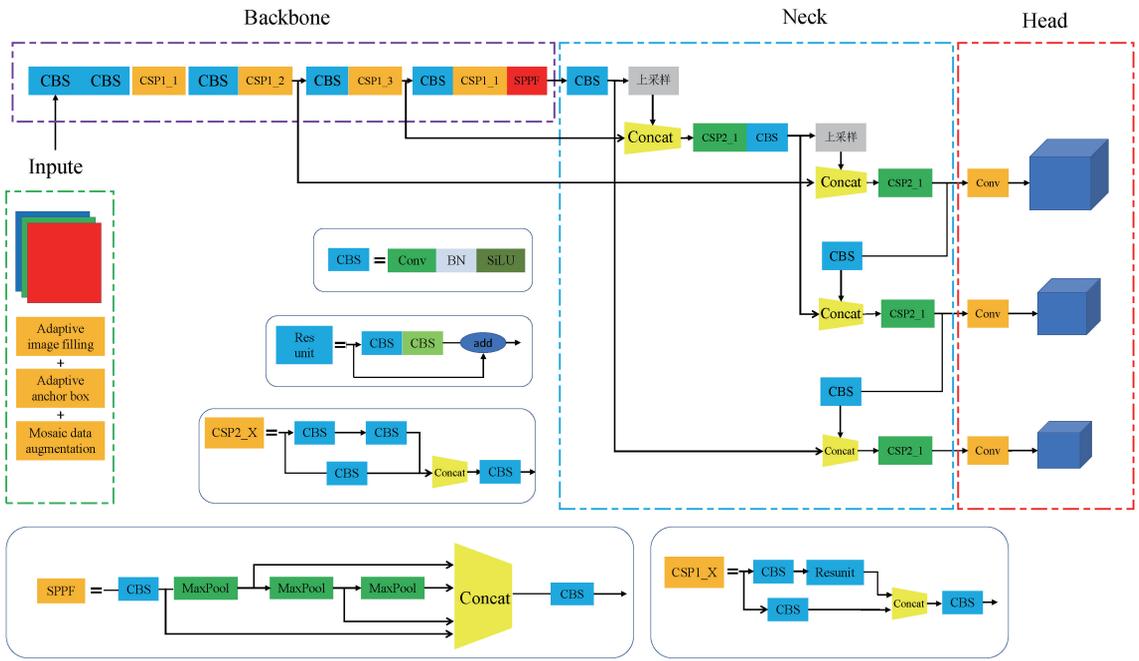


图1 YOLOv5网络结构
Fig.1 YOLOv5 network structure

2 改进YOLOv5的仓储物体检测算法

本文基于重构后的YOLOv5模型即百度飞桨PaddlePaddle的基础框架更换YOLOv5的主干网络为ConvNeXt^[15],提高图像特征的提取能力,并利用替换后的高效交并比损失^[16](efficient intersection over union loss, EIou Loss)作为新的定位损失函数对模型进行优化调整。

2.1 主干网络ConvNeXt

自2020年以来,Transformer凭借其自有的自注意力机制提取图像信息的能力在目标检测领域有着优异的表现,但考虑到Transformer网络不具备特征局部性以及平移不变性等CNN网络特点,导致其在训练中小规模数据集后的检测效果显著下降,结合本文研究的仓储物体数据集量较少的特点,设计使用卷积神经网络ConvNeXt作为模型的主干网络。

与传统的Swin-Transformer(Swin-T)相比,ConvNeXt纯卷积神经网络在经过多次实验对比后证实,在相同的计算成本下,ConvNeXt纯卷积神经网络拥有更为理想的推理速度和检测精度。ConvNeXt模型相比其他主干网络,因其存在独特的倒置瓶颈层结构且在每次下采样前后和全局均值池化之后的位置都添加了一个LN层,这些结构不仅有效地防止中小型仓储物体的特征信息在降采样过程中的丢失,也使得模型的稳定性得以提升。除了采用归一化策略,还因为减少了激活函数和BN层的使用,ConvNeXt结构不仅通过借鉴Swin-T的结构而使其提升对物体检测的准确率,也使网络降低了浮点运算量。

在ConvNeXt网络中,通过一个卷积核尺寸等于步长

的卷积层与LN层来构建ConvNeXt的Stem层,再通过一个卷积核尺寸等于步长的卷积层与级联LN层构成下采样层;且相比ResNet模块,4个阶段的Block数值比由(3, 4, 6, 3)修改成(3, 3, 9, 3)以依次堆叠倒瓶颈层和下采样层;对于倒瓶颈层设计,在网络提取到特征之后,先后利用1×1卷积分别进行升维和降维,以此来降低对高维信息的丢失程度。改进后的YOLOv5网络结构如图2所示。

2.2 损失函数改进

将原PaddlePaddle框架下的YOLOv5损失函数,完全交并比损失(complete intersection over union loss, CIoU Loss)替换为EIou Loss。针对一个好的目标框回归损失应该考虑3个重要的几何要素:重叠面积、中心点距离、长宽比等因素,CIoU在距离交并比(distance intersection over union, DIoU)的基础上,考虑了边界框Box长宽比的计算,其中衡量长宽比一致性的参数*v*如式(1)所示。这使得目标框回归更加稳定,尽可能地保证了预测框和真实框的宽和高的纵横比更为接近。

$$v = \frac{4}{\pi^2} \left(\arctan\left(\frac{w^{gt}}{h^{gt}}\right) - \arctan\left(\frac{w_p}{h_p}\right) \right)^2 \quad (1)$$

其中, w_p 和 h_p 分别表示预测框的宽和高, w^{gt} 和 h^{gt} 分别表示真实框的宽度和高度。CIoU损失函数完整定义如式(2)所示。

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + av \quad (2)$$

但是,CIoU损失,虽然在DIoU损失的基础上考虑了衡量预测框和真实框纵横比*v*,在一定程度上可以加快预测框的回归速度,但是存在的问题是在预测框回归的过程

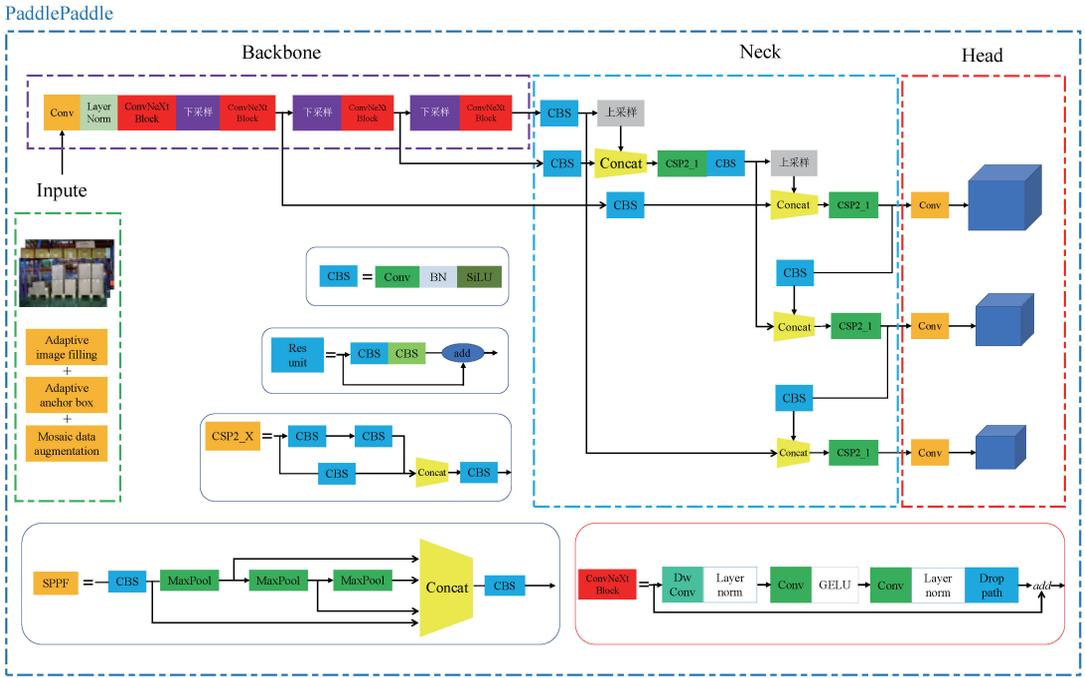


图 2 改进后的 YOLOv5 网络结构

Fig. 2 The improved YOLOv5 network structure

中,一旦预测框和真实框的宽高纵横比呈现线性比例时,预测框 w 和 h 不能同时增加或者减少,就不能继续进行回归优化。针对这一问题,文献[16]提出在 CIoU 的基础上将纵横比拆开,构建 EIoU Loss。

EIoU 的惩罚项是在 CIoU 的基础上分别计算目标框和锚框的长和宽,惩罚项表达式如式(3)所示。

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} \quad (3)$$

该损失函数包含 3 个部分:预测框和真实框的重叠损失 L_{IoU} , 预测框和真实框的中心距离损失 L_{dis} 以及预测框和真实框的宽高损失 L_{asp} 。相比 CIoU 损失,EIoU 损失新增的“宽高损失”通过调整目标框与锚框的尺寸,以最小化宽高差异的方式,使得模型收敛速度更快,同时引入焦点损失(Focal Loss)解决在仓储物体检测重难易样本不平衡的问题。EIoU 损失函数惩罚项展开如式(4)所示。

$$L_{EIoU} = 1 - EIoU = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \quad (4)$$

其中, b, b^{gt} 分别表示预测框和真实框的中心点, ρ^2 表示的两个中心点间的欧氏距离, c 表示能同时包含预测框和真实框的最小外接矩形的对角线距离, c_w 和 c_h 分别表示外接矩形的宽和高, w^{gt} 和 h^{gt} 表示真实框的宽和高。

3 实验结果与分析

3.1 实验数据集及实验环境

由于目前尚未有公开的仓储环境下的公开数据集,于

是本文在实际仓库环境中进行图像采集。首先,为丰富图像数据的种类,本文在不同光线环境下利用摄像头得到在距目标 2~3 m 的数据图像;其次,为提高数据集的代表性,本文采集不同种类、不同大小尺度和角度的 3 种仓储物体。摄像头采集的图像分辨率为 1 920×1 080。同时,本文为减少实验时间,加快模型收敛,将数据集分辨率调整为 640×360,部分仓储训练集如图 3 所示。



图 3 部分仓储训练集图像

Fig. 3 Part of the warehouse training set images

利用开源数据标注工具 LabelImg 进行数据标注,标注时将仓储物体的最小外接矩形框作为真实框,生成对应的可扩展标记语言(extensible markup language, XML)文本文件后,再对数据集进行检查清洗,剔除标注出错的数据图像,最后转化为 COCO 数据格式。本文使用的实验数据集为 10 382 张图像,标注物体实例共 161 832 个,包含托盘(tray)、叉车(forklift)和货物(cargo)3 个仓储物体类别,其中,货物一类包含仓储环境下绝大多数的物流物体种类,可有效满足此类环境下的仓储物流物体多样性的要

求。所含数据集中叉车图像约有2500余张,包含货物和托盘的图像共7800余张。数据集按照8:1:1比例划分为训练集、验证集和测试集。

本文实验的软硬件平台配置参数如表1所示。

表1 软硬件平台配置参数

Table 1 Software and hardware platform configuration parameters

配置	实验环境
GPU	NVIDIA GeForce RTX4070
CPU	Intel Core i7-13700F
Operating system	Ubuntu 22.04.3
CUDA	cuda 11.7
Framework	PaddlePaddle 2.6.0
Python	Python 3.11.8

3.2 评价指标

在目标检测领域中,常用的评价指标主要有召回率、精确率和准确度等。

召回率(recall, R)表示正确检测物体的个数占物体总数的百分比,其表达式如式(5)所示。

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

精确率(precision, P)表示正确检测物体的个数占总检测物体个数的百分比,其表达式如式(6)所示。

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

准确率(accuracy, Acc)表示正确分类的样本数占全部样本数的百分比,其表达式如式(7)所示。准确率越高,分类的效果越好。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

式(5)~(7)中,TP、FN、FP、TN分别表示真实情况为正时预测结果为正、真实情况为负时预测结果为负、真实情况为负时预测结果为正、真实情况为正时预测结果为负。

此外,如果以召回率为横坐标,准确率为纵坐标,且在一定阈值的基础之上构成的曲线称之为P-R曲线。P-R曲线与横纵坐标围成的面积即为平均精确度(average precision, AP),若对多个类别的AP值求平均,即可得到mAP值。COCO格式数据集下计算mAP的方式是先计算所有类别在交并比(intersection over union, IoU)从0.5到0.95,以步长为0.05下取10个IoU值的平均精确度AP值后再求其均值。

本文实验采用mAP等12个常用指标用以评价模型的泛化性能,如表2所示。同时,还采用模型参数量(parameters, Params)、帧率(frame per second, FPS)以及模型计算量(floating point operations, FLOPs)等来评价模

型的综合性能。

表2 常用的评价指标

Table 2 Common evaluation indexes

指标	注释
mAP	IoU=.50:.05:.95时的平均精度
mAP_50	IoU=.50时的平均精度
mAP_75	IoU=.75时的平均精度
mAP_s	小物体(面积<32 ²)的平均精度
mAP_m	中等物体(32 ² <面积<96 ²)的平均精度
mAP_l	大物体(面积>96 ²)的平均精度
AR_max_1	每张图像预测框数量阈值为1的平均召回率
AR_max_10	每张图像预测框数量阈值为10的平均召回率
AR_max_100	每张图像预测框数量阈值为100次的平均召回率
AR_s	小物体(面积<32 ²)的平均召回率
AR_m	中等物体(32 ² <面积<96 ²)的平均召回率
AR_l	大物体(面积>96 ²)的平均召回率

其中,在图像数据集的分辨率为640×360下,结合评价指标对小尺度物体的分辨率定义为像素面积<32²,即对应实际情况中,在2m距离时,小于0.123m为边长的正方形物体;大尺度物体的分辨率定义为像素面积>96²,即对应实际情况中,在2m距离时,大于0.4m为边长的正方形物体;处在0.123m与0.4m边长之间的正方形物体为中等尺度大小物体。不同尺度物体如图4所示。



图4 不同尺度物体数据集

Fig. 4 Object data sets of different scales

3.3 实验分析

为验证不同定位损失函数,如广义交并比损失函数(generalized intersection over union loss, GIoU Loss)、CIoU、DIoU和本文研究的ConvNeXt主干网络、EIoU损失函数对检测仓储物体效果的适用性,本文采用不同的模型改进方法来训练,利用消融实验对比分析模型的改进效

果。从而评估每个改进处理对改进后的 YOLOv5 模型的影响。本文以飞桨 PaddlePaddle 框架下的 YOLOv5s 模型作为基准,各模型结构设置如表 3 所示。

表 3 各模型结构设置

Table 3 The setting of each model structure

模型	定位损失函数	主干网络
YOLOv5s	CIoU Loss	CSPDarknet
YOLOv5s_I	IoU Loss	CSPDarknet
YOLOv5s_G	GIoU Loss	CSPDarknet
YOLOv5s_D	DIoU Loss	CSPDarknet
YOLOv5s_E	EIoU Loss	CSPDarknet
YOLOv5-CE	EIoU Loss	ConvNeXt

考虑到原 PaddlePaddle 框架下默认配置是以 8 卡 GPU 训练,且考虑到本文实验使用自定义训练集进行训练,所以实验采用对应模型训练好的 COCO 权重作为预训练权重以加速收敛,降低了网络迭代的轮数要求,本文实验的 epoch 仅设置为 100, batch_size 设置为 8,学习率为 0.001(原默认配置的 10%),其他超参数采用默认数值。

经过实验,改进后的各模型实验结果如表 4 所示。

表 4 消融实验结果(指标:mAP@[IoU=0.50:0.95])

Table 4 Results of ablation experiment
(Index: mAP@[IoU=0.50:0.95])

模型	mAP/ %	AP_c/ %	AP_tray/ %	AP_f/ %
YOLOv5s	88.7	86.8	83.6	95.9
YOLOv5s_I	88.5	86.8	83.7	95.1
YOLOv5s_G	88.7	86.6	83.9	95.7
YOLOv5s_D	88.7	86.8	83.7	95.7
YOLOv5_E	89.0	87.0	84.1	96.1
YOLOv5-CE	89.8	87.8	84.7	96.8

由表 4 可以看出,相比 YOLOv5 模型,模型 YOLOv5_E 结合 EIoU Loss 后,货物的检测精度(AP_c)以及叉车的检测精度(AP_f)都得到了小幅度的提高,且托盘检测精度

(AP_tray)提升了 0.5%,模型的 mAP 整体提升了 0.3%。由此可见,在基准模型检测精度较高的基础上,EIoU Loss 的改进依然有利于得到精度更高的模型。而替换了其他损失函数的模型在精度方面未得到显著提升。YOLOv5-CE 在结合 ConvNeXt 网络后,3 种仓储物体类型的平均精度都得到了较大提升,其中对货物的检测精度(AP_c)提升了 0.8%,模型整体相比 YOLOv5s_E 提升了 0.8%;相比原 YOLOv5s 模型(以 CIoU 为损失函数),mAP@[IoU=0.50:0.95]提升至 89.8%,表明 ConvNeXt 网络的特征提取能力显著高于 CSPDarknet 网络,能够较好地提取中小型仓储物体,如中小尺度货物(cargo)的特征。

同时,本文也对比了 100 训练轮数内不同模型的 mAP 值变化,如图 5 所示。

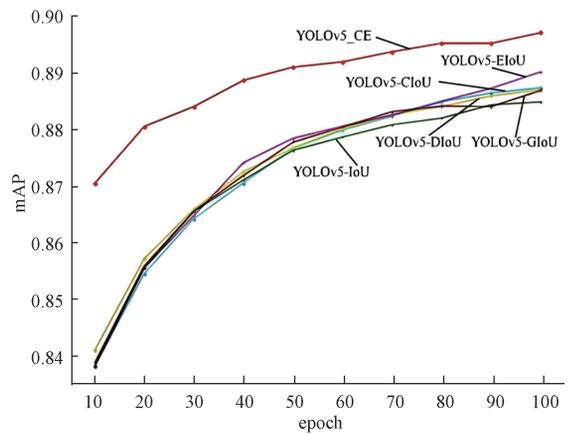


图 5 不同模型的 mAP 值对比

Fig. 5 Comparison of mAP values of different models

从图 5 可以看出,相比其他 IoU 改进的模型,以 EIoU 为损失函数的模型能够较快收敛,并在平均精度方面也有所提升。而在此基础上,再更换 YOLOv5 的主干网络得到的模型,相比原 YOLOv5 模型,不仅能迅速收敛趋于稳定、没有出现拟合和欠拟合现象的同时,平均精度也有显著提升。

此外,本文对比了不同尺度、不同 IoU 阈值下的模型检测精度,实验结果如表 5 所示。

表 5 平均精度对比

Table 5 Average precision comparison

模型	mAP_s/%	mAP_m/%	mAP_l/%	mAP_50/%	mAP_75/%	mAP/%
YOLOv5s	39.5	85.2	92.1	98.8	98.4	88.7
YOLOv5s_I	41.0	85.2	91.9	98.8	98.4	88.5
YOLOv5s_G	38.5	85.3	91.8	98.8	98.4	88.7
YOLOv5s_D	47.6	85.3	91.9	98.9	98.5	88.7
YOLOv5_E	40.7	85.5	92.4	98.8	98.4	89.0
YOLOv5-CE	43.7	86.2	93.1	98.9	98.2	89.8

从表5可以看出,随着各个损失函数的替换,各个模型对中小尺度物体的检测精度皆有不同程度的提升,但对大尺度物体的检测精度在损失函数替换为IoU、GIoU、DIoU后,精度有所下降,而在替换为EIoU后,提升精度达到了92.4%。同时,再更换主干网络之后,模型改进对小尺度物体的检测精度提升最大,提升了4.2%,对其他尺度

的检测精度皆有不同程度的提升,继而表明ConvNeXt主干网络以及EIoU的替换改进可以有效提高模型对中小尺度物体的特征提取能力,更精准检测出较小的仓储物体(cargo)。

本文也研究了各模型平均召回率的变化,得出实验结果如表6所示。

表6 平均召回率对比

Table 6 Comparison of average recall rates

模型	AR_max_1/%	AR_max_10/%	AR_max_100/%	AR_s/%	AR_m/%	AR_l/%
YOLOv5s	43.9	77.1	91.7	61.1	89.0	94.2
YOLOv5s_I	43.7	77.0	91.6	66.2	89.0	94.2
YOLOv5s_G	43.8	77.1	91.6	56.2	89.0	94.1
YOLOv5s_D	43.8	77.0	91.6	61.3	88.9	94.1
YOLOv5_E	43.9	77.1	91.7	61.3	89.0	94.3
YOLOv5-CE	44.1	77.7	92.3	66.8	89.7	94.8

从表6可以看出,在平均精度提升的同时,各大小尺度物体的平均召回率在损失函数替换为IoU、GIoU、DIoU后,有不同程度的降低,只有损失函数替换为EIoU后,小尺度物体的平均召回率才有小幅度提升。YOLOv5-CE相比YOLOv5_E提升了5.5%,说明伴随着检测精度的提升,改进后的模型对小尺度物体的检测也更为全面、完整,对其他尺度物体的平均召回率也有不同程度的提升。

为进一步证明本文模型改进的有效性和实用性,本文对YOLOv5与YOLOv5-CE模型的收敛速度进行对比分析,在同时训练100个周期后,得到两个模型的Loss值收敛曲线(平滑处理后)如图6所示。从图6可以看出,改进后的模型,其损失值的收敛速度在训练初期显著快于原YOLOv5模型,且经过充分训练后,最终趋于稳定,收敛结果也明显优于原模型。

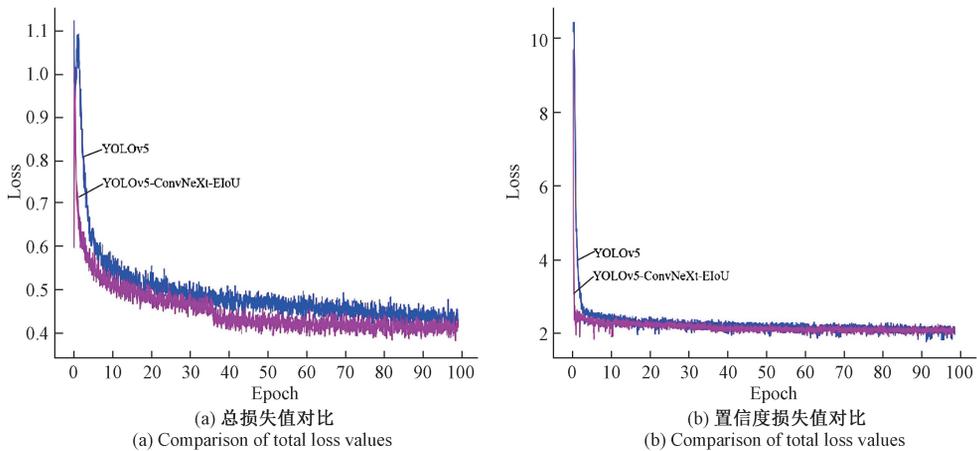


图6 损失值对比

Fig. 6 Loss value comparison

通过消融实验以及损失值对比,说明本文提出的改进算法能有效提高YOLOv5模型对中小尺度仓储物体的检测性能,更好地检测多尺度的仓储物体。为客观评估本文模型的检测性能,本文模型与YOLOv3、YOLOX、YOLOv7、SDD等模型进行对比分析,所有模型均采用COCO数据集下的权重作为预训练权重以加快模型收敛,除YOLO系列模型epoch值调整为100外,其他模型的超参数不变,得到YOLO系列和其他算法的对比结果如表7所示。

从表7可以看出,改进后的YOLOv5模型比绝大多数模型都显得更为出色,改进后YOLOv5模型的mAP@[IoU = 0.50 : 0.95],相比YOLOv7-tiny模型高出3.39%,相比YOLOv8高出0.9%,并且模型在IoU阈值设置为0.75得到的mAP值相比YOLOv7l,甚至YOLOv8s都有更好的表现。同时,与YOLOv7l、YOLOv3等模型相比,同样GPU下的图像检测速度(FPS)也更高,达到了76.71 fps。在参数数量的比较上,改进后模型的参数

表 7 YOLOv5-CE、YOLO 系列及其他算法的对比

Table 7 Comparison of YOLOv5-CE, YOLO series and other algorithms

网络模型	主干网络	mAP/ %	mAP_50/ %	mAP_75/ %	FPS/ $f \cdot s^{-1}$	Params/ M	FLOPs/ G
YOLOv3	DarkNet53	80.9	99.3	97.8	56.61	61.53	69.91
YOLOv3	MobileNet(v1)	77.7	99.1	95.2	114.99	24.16	21.69
SSD	MobileNet(v1)	83.8	99.0	97.6	145.75	5.61	1.16
SSD	ResNet	85.5	99.1	98.2	40.80	12.52	13.88
YOLOv6n ^[17]	EfficientRep	89.7	99.0	98.7	119.04	5.07	6.10
YOLOv7-tiny	ELANNet	86.4	98.5	97.6	88.29	6.02	6.56
YOLOv7l	ELANNet	85.6	97.7	96.8	58.40	37.21	52.48
YOLOv8s	CSPdarknet	87.9	98.6	97.5	92.25	11.14	14.30
Faster R-CNN	ResNet50	88.8	99.2	98.7	6.65	33.06	—
YOLOX-tiny	CSPDarkNet	84.1	99.1	98.2	164.41	5.06	6.45
YOLOX ^[18]	CSPDarkNet	89.5	99.3	99.1	85.60	9.00	26.80
YOLOv5-CE	ConvNeXt	89.8	98.9	98.2	76.71	34.23	8.49

量小于 YOLOv7l,但相比大部分模型,参数量偏大。而在浮点运算次数(FLOPs)的比较上,改进后的模型皆显著小于其他大部分模型,仅有 YOLOv3 的 39.14%,YOLOv7l 的 16.18%,YOLOv8s 的 59.18%。

经过对比实验结果,说明在仓储数据集下,以 YOLOv5s 为基础改进的模型相比大多数模型都更具有优

势,在保持较高的检测精度的同时,在浮点运算次数、检测速度以及各 IoU 阈值下的平均检测精度上相比其他传统检测模型取得了更为良好的平衡,能较好地满足对不同仓储物体的精准检测。

同样,本文也研究了模型平均召回率的变化,得出不同模型平均精度与平均召回率的实验结果如表 8 所示。

表 8 YOLOv5-CE、YOLO 系列及其他算法的对比

Table 8 Comparison of YOLOv5-CE, YOLO series and other algorithms

网络模型	AR_s/%	AR_m/%	AR_l/%	mAP/%	mAP_s/%	mAP_m/%	mAP_l/%
YOLOv3 (DarkNet53)	64.0	85.1	86.7	80.9	61.1	80.6	82.3
YOLOv3 (MobileNet)	52.9	82.7	83.6	77.7	49.9	78.0	79.3
SSD (MobileNet)	51.2	85.8	90.0	83.8	47.4	81.7	86.6
SSD (ResNet)	59.3	88.0	91.1	85.5	57.0	84.2	87.2
YOLOv6n	61.5	89.9	95.1	89.7	59.1	86.1	93.1
YOLOv7-tiny	55.2	87.6	93.1	86.4	41.4	82.9	89.6
YOLOv7l	59.9	86.8	93.0	85.6	34.6	81.0	89.8
YOLOv8s	59.9	87.8	93.8	87.9	39.1	83.8	91.6
Faster R-CNN	55.9	89.0	94.4	88.8	53.3	85.5	91.6
YOLOX-tiny	64.5	86.3	90.3	84.1	45.5	82.0	86.8
YOLOX	66.3	89.9	94.6	89.5	64.3	86.6	92.8
YOLOv5-CE	66.8	89.7	94.8	89.8	43.7	86.2	93.1

从表 8 可以看出,改进后的 YOLOv5 的平均精确度在中、大尺度物体上都较高。在中尺度物体上,本文模型相比 YOLOv8 高出 2.4%;在大尺度物体上,相比 YOLOv7l 提高了 3.3%,而仓储训练集中的大部分实例又皆以中尺度为准,说明改进后的模型较为适合用于精准检测仓储环境下的多种物体。

此外,在平均精确度提升的同时,小、中、大尺度物体

上的平均召回率也有所提升,其中 YOLOX 模型在中小尺度物体上的平均精度皆优于本文模型,但在小、大尺度物体上的平均召回率,相比本文的改进模型偏低,说明 YOLOX 模型相较于改进后的 YOLOv5,未能检测到所有实际存在的物体。而对于其他模型,本文模型在多尺度上的平均精度和平均召回率都展示了较好的优势。例如,在中尺度物体上的平均精度和平均召回率比 YOLOv8 高出

2.4%和1.9%,而在大尺度物体上的平均精度和平均召回率比YOLOv8高出1.5%和1.0%。

YOLOv5-CE模型对仓储物体的检测效果如图7所示。



图7 YOLOv5-CE模型检测效果

Fig.7 YOLOv5-CE model test results

4 结 论

考虑到复杂环境下仓储物体传统视觉检测模型性能以及检测精度较低的问题,本文在飞桨PaddlePaddle的框架下,提出一种基于YOLOv5的改进算法,即YOLOv5-CE,并进行对比实验。实验结果表明,本文提出的算法在模型参数量、检测速度与检测精度上取得了良好的平衡,但本文算法所提升的精度还不是很,还需从扩充数据集以及增加注意力机制方向进行深入研究,以进一步提升仓储物体检测精度。

参考文献

- [1] 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.
- [2] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [3] FARHADI A, REDMON J. YOLOv3: An incremental improvement [C]. Computer Vision and Pattern Recognition. Berlin/Heidelberg, Germany: Springer, 2018, 1804: 1-6.
- [4] LIU W, ANGELOV D, ERHAN D, et al. SSD: Single shot multiBox detector [C]. European Conference on Computer Vision. Cham: Springer Press, 2016: 21-37.
- [5] 王新, 杨秀梅. 基于YOLOv5s和改进质心跟踪的人员跌倒检测[J]. 电子测量技术, 2023, 46(24): 172-178.
WANG X, YANG X M. Personnel fall detection based on YOLOv5s and improved centroid tracking[J]. Electronic Measurement Technology, 2023, 46(24): 172-178.

- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 142-158.
- [7] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2015: 91-99.
- [8] 王飞, 王林, 张儒良, 等. 基于融合FPN和Faster R-CNN的行人检测算法[J]. 数据采集与处理, 2019, 34(3): 530-537.
WANG F, WANG L, ZHANG R L, et al. Pedestrian Detection Algorithm Based on Fusion FPN and Faster R-CNN [J]. Journal of Data Acquisition and Processing, 2019, 34(3): 530-537.
- [9] 李映松, 杨爱英, 刘轩, 等. 基于Transformer改进的Faster-Rcnn仓储箱体检测算法[J]. 自动化与仪器仪表, 2022(8): 1-6.
LI Y S, YANG A Y, LIU X, et al. Storage Box Detection Method Based on Transformer Improved Faster-Rcnn [J]. Automation and Instrumentation, 2022(8): 1-6.
- [10] 陈亮杰, 王飞, 王梨, 等. 基于SSD的仓储物体检测算法研究[J]. 软件导刊, 2019, 18(4): 28-31.
CHEN L J, WANG F, WANG L, et al. Research on Warehouse Object Detection Algorithm Based on SSD [J]. Software Guide, 2019, 18(4): 28-31.
- [11] 金秋, 李天剑. 仓储环境下基于深度学习的物体识别方法研究[J]. 北京信息科技大学学报(自然科学版), 2018, 33(1): 60-65.
JIN Q, LI T J. Object recognition method based on deep learning in storage environment [J]. Journal of

- Beijing Information Science and Technology, 2018, 33(1): 60-65.
- [12] ZACCARIA M, MONICA R, ALEOTTI J. A comparison of deep learning models for pallet detection in industrial warehouses [C]. 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing(ICCP). IEEE, 2020: 417-422.
- [13] RUTINOWSKI J, YOUSSEF H, GOUDA A, et al. The potential of deep learning based computer vision in warehousing logistics [J]. Logistics Journal: Proceedings, 2022 (18), DOI: 10.2195/lj _ proc _ rutinowski_en_202211_01.
- [14] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: 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.
- [15] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11976-11986.
- [16] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [17] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. ArXiv preprint arXiv: 2209.02976, 2022.
- [18] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021 [J]. ArXiv preprint arXiv: 2107.08430, 2021.

作者简介

刘庭浔, 本科, 主要研究方向为图像处理、目标检测。

E-mail: 1932106902@qq.com

王飞(通信作者), 硕士, 讲师, 主要研究方向为计算机视觉、图像处理、模式识别。

E-mail: wangfei10248@163.com

刘旭, 本科, 主要研究方向为机器学习、目标检测。

E-mail: 1431296403@qq.com

蔡晨文, 本科, 主要研究方向为图像处理、机器学习。

E-mail: 2364664762@qq.com