

YOLO 系列目标检测算法研究进展

王琳毅¹, 白 静^{1,2}, 李文静¹, 蒋金哲¹

1. 北方民族大学 计算机科学与工程学院, 银川 750021

2. 国家民委图形图像智能处理实验室, 银川 750021

摘 要: YOLO 算法是目标检测中研究的热点方向之一。近几年, 随着 YOLO 系列算法及其改进模型的不提出, 使其在目标检测领域取得了优异的成绩, 被广泛应用于现实中各个领域。针对 YOLO 系列目标检测算法, 整理了目标检测典型数据集及评价指标; 回顾了 YOLO 整体框架以及 YOLOv1~YOLOv7 目标检测算法的发展历程; 总结了在输入、特征提取和预测这三个阶段下的数据增强、轻量化网络构建和 IOU 损失优化等八个改进方向的模型及性能; 介绍了 YOLO 算法应用领域; 结合目标检测目前存在的实际问题, 总结并展望了 YOLO 算法的发展方向。

关键词: 计算机视觉; 目标检测; YOLO; 改进模型

文献标志码: A **中图分类号:** TP391.41 **doi:** 10.3778/j.issn.1002-8331.2301-0081

Research Progress of YOLO Series Target Detection Algorithms

WANG Linyi¹, BAI Jing^{1,2}, LI Wenjing¹, JIANG Jinzhe¹

1. School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China

2. The Key Laboratory of Images & Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China

Abstract: The YOLO-based algorithm is one of the hot research directions in target detection. In recent years, with the continuous proposition of YOLO series algorithms and their improved models, the YOLO-based algorithm has achieved excellent results in the field of target detection and has been widely used in various fields in reality. This article first introduces the typical datasets and evaluation index for target detection and reviews the overall YOLO framework and the development of the target detection algorithm of YOLOv1~YOLOv7. Then, models and their performance are summarized across eight improvement directions, such as data augmentation, lightweight network construction, and IOU loss optimization, at the three stages of input, feature extraction, and prediction. Afterwards, the application fields of YOLO algorithm are introduced. Finally, combined with the actual problems of target detection, it summarizes and prospects the development direction of the YOLO-based algorithm.

Key words: computer vision; object detection; YOLO; improved model

计算机视觉近年来发展迅猛, 在国内外都引起了广泛的关注。作为其核心问题之一的目标检测, 旨在找出图像中特定目标的类别和位置, 现已广泛应用于各个领域, 如自动驾驶、遥感图像、视频监控以及医疗检测等。

如图 1 所示, 自 2014 年 RCNN^[1] 被提出之后, 基于深度学习的目标检测方法取代传统方法成为主流。起初, 基于深度学习的目标检测采用双阶段(two-stage)策略,

即显式地将检测过程分为候选区域选取和目标区域判断两个阶段, 其检测精确度较高, 但检测速度较慢。2016 年, 以 YOLOv1^[2] 为代表的单阶段(one-stage)目标检测方法出现了。这类方法不需要预先提取候选区域, 而是直接预测输出目标物体的类别概率和位置, 极大地减少了计算资源的消耗, 提高了检测的速度, 因而受到了更多的关注。纵观单阶段目标检测方法的发展历程

基金项目: 国家自然科学基金(62162001, 61762003); 宁夏自然科学基金(2022AAC02041); 宁夏优秀人才支持计划; 北方民族大学创新项目(YCX22181)。

作者简介: 王琳毅(1997—), 男, 硕士研究生, 研究方向为图像处理与计算机视觉、深度表征学习; 白静(1982—), 通信作者, 女, 博士, 教授, CCF 高级会员, 研究方向为机器学习、深度表征学习、计算机视觉应用, E-mail: baijing@nun.edu.cn; 李文静(1998—), 女, 硕士研究生, 研究方向为图像处理与计算机视觉、深度表征学习; 蒋金哲(1998—), 女, 硕士研究生, 研究方向为图像处理与计算机视觉、深度表征学习。

收稿日期: 2023-01-09 **修回日期:** 2023-04-19 **文章编号:** 1002-8331(2023)14-0015-15

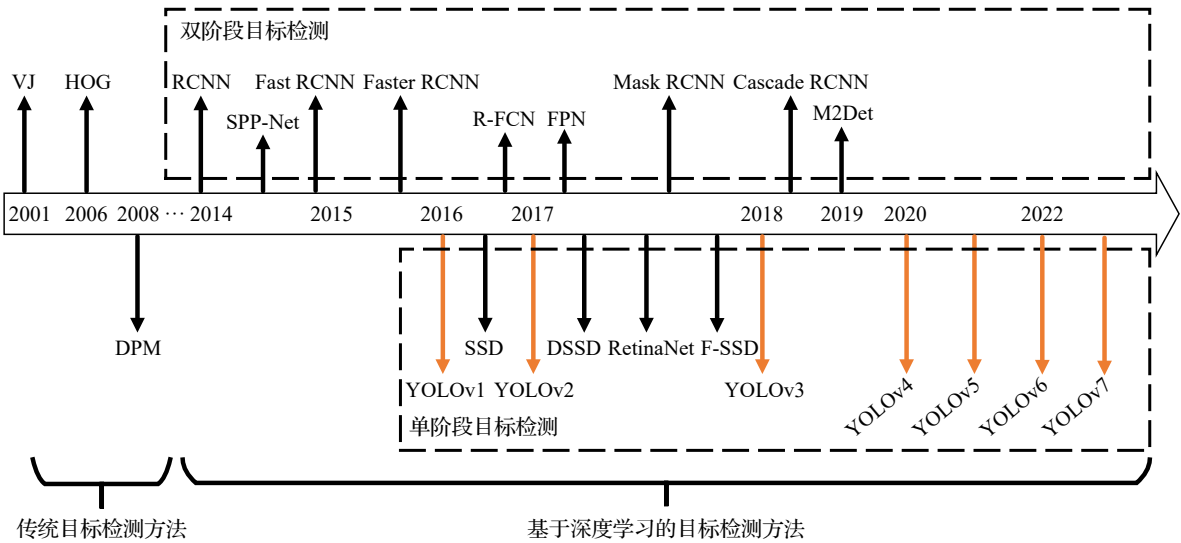


图1 目标检测发展历程

Fig.1 Development history of target detection

可以发现,从首个单阶段目标检测方法YOLOv1提出至2022年的YOLOv7^[3],YOLO系列的目标检测方法伴随着单阶段目标检测的发展而发展,已经成为one-stage方法的典型代表。

事实上,鉴于YOLO算法检测效率高、框架设计模块化、易改进等优点,在单阶段目标检测方法中不仅出现了YOLOv1至YOLOv7这7个版本的基础模型,还出现了大量的基于YOLO的改进模型,且被应用到各个领域。如,为了改善水下目标检测存在图像不清晰,水下环境复杂导致的漏检和误检问题,叶赵兵等人^[4]提出YOLOv3-SPP水下目标检测算法;为了解决安全帽检测方法模型参数量大、检测设备部署较难以及小目标检测效果差等问题,王建波等人^[5]提出改进YOLOv4-tiny的安全帽佩戴检测算法;为了提高人工检测地下污水管道缺陷的效率,避免主观判断管道缺陷等问题,Tan等人^[6]提出改进YOLOv3的污水管道缺陷检测算法。总体而言,如今,目标检测领域涌现出许多YOLO及其改进工作,这些方法已经自成体系且极大地推进了单阶段目标检测的发展。与此同时,现有综述主要针对基于深度学习的目标检测,缺乏对YOLO及其相关改进方法的综

述。本文以总结YOLO发展历程和改进方法为目标,为研究者改进YOLO模型提供参考方向。

1 目标检测典型数据集及评价指标

1.1 目标检测典型数据集

深度学习算法之所以得到广泛发展,很大一方面原因得益于大规模数据集的出现;目标检测算法能发展到现在,同样也离不开数据集的重要贡献。本节对目标检测任务的典型数据集,PASCAL VOC^[7-8]、MS COCO^[9]、ILSVRC^[10]、Open Image^[11]以及DOTA^[12]进行介绍。表1汇总了各个数据集的基本信息。

(1)PASCAL VOC

PASCAL VOC包含了目标检测、分类、分割以及行为识别等领域的一系列数据集。在目标检测领域,目前应用最广的是PASCAL VOC07和PASCAL VOC12,它们均包括20个类别。其中,PASCAL VOC07有9 963个图像和超过12 000个标记对象,PASCAL VOC12有大约22 531个图像和超过27 000个标记对象,训练集测试集的划分详见表1。

表1 目标检测典型数据集对比

Table 1 Comparison of typical target detection data sets

数据集	类别	图像数量	优点	缺点	适用场景
PASCAL VOC07	20	9 963	数据集规模较小,对硬件环境要求较低	部分类别相似性较高	图像分类、检测和分割
PASCAL VOC12	20	22 531			
MS COCO17	80	162 888	数据集中目标尺寸多,图片与目标实例丰富	部分目标存在遮挡	图像识别、分割、检测
ILSVRC	200	542 188	数据集的种类信息极其丰富,包含大多目标对象的底层特征	参数量较大,对硬件资源消耗较大	场景、图像分类、目标定位、目标检测以及视频目标检测
Open Image	600	1 910 000	数据集中目标对象位置标注清晰	图像背景较为复杂,容易干扰目标对象	目标检测、视觉关系检测、图像分割和图像描述
DOTA	15	2 806	数据集中样本丰富,包含各种尺度、方向和形状的目标物体;实例标记详细	数据集图像尺寸较大,对一般检测模型不友好	航空遥感检测

(2)MS COCO

MS COCO包括了物体检测、实例分割、无固定形状分割、全景分割以及人体关键点检测等领域的系列数据集。其中,MS COCO17为目标检测领域常用的数据集,其包含91个类别(检测任务使用80类),训练、验证和测试图像分别为117 266个、4 952个和40 670个,包含的标记对象超过89 000个,详细信息见表1。

(3)ILSVRC

ILSVRC数据集是基于ImageNet数据集构建的子集,其可用于场景分类、图像分类、目标定位、目标检测以及视频目标检测。该数据集包含200个类别以及超过50万张图像组成,详细信息见表1。

(4)Open Image

Open Image数据集常用于图像分割、图像分类以及对象检测任务。该数据集包含了191万张图片、600个种类,1 540万个标注框,每张图像标记得非常清晰,是实例标记规模最大的数据集,详细信息见表1。

(5)DOTA

DOTA是航空遥感检测任务的卫星航拍数据集,常用于检测和评估航拍图像中的目标对象。该数据集包含15个目标类别,共2 806张卫星航拍图像以及188 282个标记对象,详细信息见表1。

其中,PASCAL VOC是真实环境下的数据集,该数据集中个别样本数量差异较大,但数据集格式规范。与PASCAL VOC相比,MS COCO规模更大,数据集中包含了更多的真实环境图像,图像背景更加复杂,目标样本数量更多,目标尺寸差异更大。因此,MS COCO数据集上的任务更难,也更能衡量检测模型的优劣。

1.2 评价指标

在目标检测领域建立了一整套评价指标,可全面评价目标检测的速度和精度,包括每秒处理帧数(frame per second, FPS)、精确度(precision, P)、召回率(recall rate, R)、平均精度(average precision, AP)、平均精度均值(mean average precision, mAP)、交并比(intersection over union, IOU),这些指标通常适用于基于边界框的目标检测算法。

为方便定义各个指标,将检测目标分为正样本和负样本,当正样本被正确识别为正样本时称为真阳性(true positives, TP),当正样本被错误识别为负样本时称为假阴性(false negatives, FN),当负样本被正确识别为负样本称为真阴性(true negatives, TN),当负样本被错误识别为正样本时称为假阳性(false positives, FP),则各个指标的定义如下^[13]。

(1)FPS,每秒检测的图像帧数。

(2)P,精确率,是指预测出的类别中是正确类别的概率,计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (1)$$

(3)R,召回率,是指预测出的正确类别在该正样本中的概率,计算公式如下:

$$R = \frac{TP}{TP + FN} \quad (2)$$

(4)AP,平均精确度,是PR曲线下面的面积,也是预测类别的概率。

(5)mAP,平均AP值,是对多个预测类别的AP值求平均AP值。

(6)IOU,交并比,是指模型生成的预测框A和真实框B之间交集与并集的比例。

以上指标数值越大,模型性能越好,其中:FPS是评价算法的速度度量指标;精确度P和召回率R形成的PR曲线围成的面积为平均精度AP,反映了单类别的准确度;当检测多个类时,多个类别的AP求平均精度均值mAP反映了整个模型的检测效果;IOU体现了定位的准确性,比值越高说明预测的结果越准确。在YOLO算法中通常使用FPS和mAP来评价算法的实际性能^[14]:FPS越高代表检测的图像数量越多,算法的运行速度越快。mAP值越高代表算法的准确率高,算法性能好。

然而,现有的评价指标也存在一些缺陷。例如,评价目标检测准确性的IOU作为阈值来判断检测结果是否正确,这种方法对于不同大小和形状的目标可能不够敏感,也不能够反映出算法的检测能力是否足够鲁棒,难以精确评估多目标检测的性能;使用mAP评价算法的精确度时,只考虑了每个类别的精度,而没有考虑不同类别之间的关系,在一些特定的场景下,这可能会导致误判或者漏检;而评价算法速度的FPS则存在无法反映算法复杂度,并且受硬件影响较大等问题。

2 YOLO系列模型

为解决Two-Stage算法提取特征区域重复、检测速度慢等问题,2016年Redmon等人将目标检测看成回归问题,提出了基于端到端的YOLOv1算法,直接获取目标的类别和位置;此后相继出现了YOLO多个版本的进阶模型。本章将从YOLO模型的整体框架和发展历程两个方面对其进行介绍。

2.1 YOLO整体框架

如图2所示,YOLO模型整体可以划分为输入、特征提取以及预测三个阶段。其中:输入阶段对应Input层,旨在接收图像并对其进行数据增强,不同YOLO模型输入尺寸有所不同;特征提取阶段往往利用预训练的卷积神经网络作为Backbone,提取图像中的重要特征。在YOLOv2^[15]之后,学者们在Backbone之后设计新增了Neck层,通过特征金字塔等方法融合Backbone中的多层级特征,进一步增强网络的表达能力;预测阶段对应Head层,其根据特征提取阶段输出的特征图生成不同尺寸的锚框,再利用非极大值抑制(non-maximum

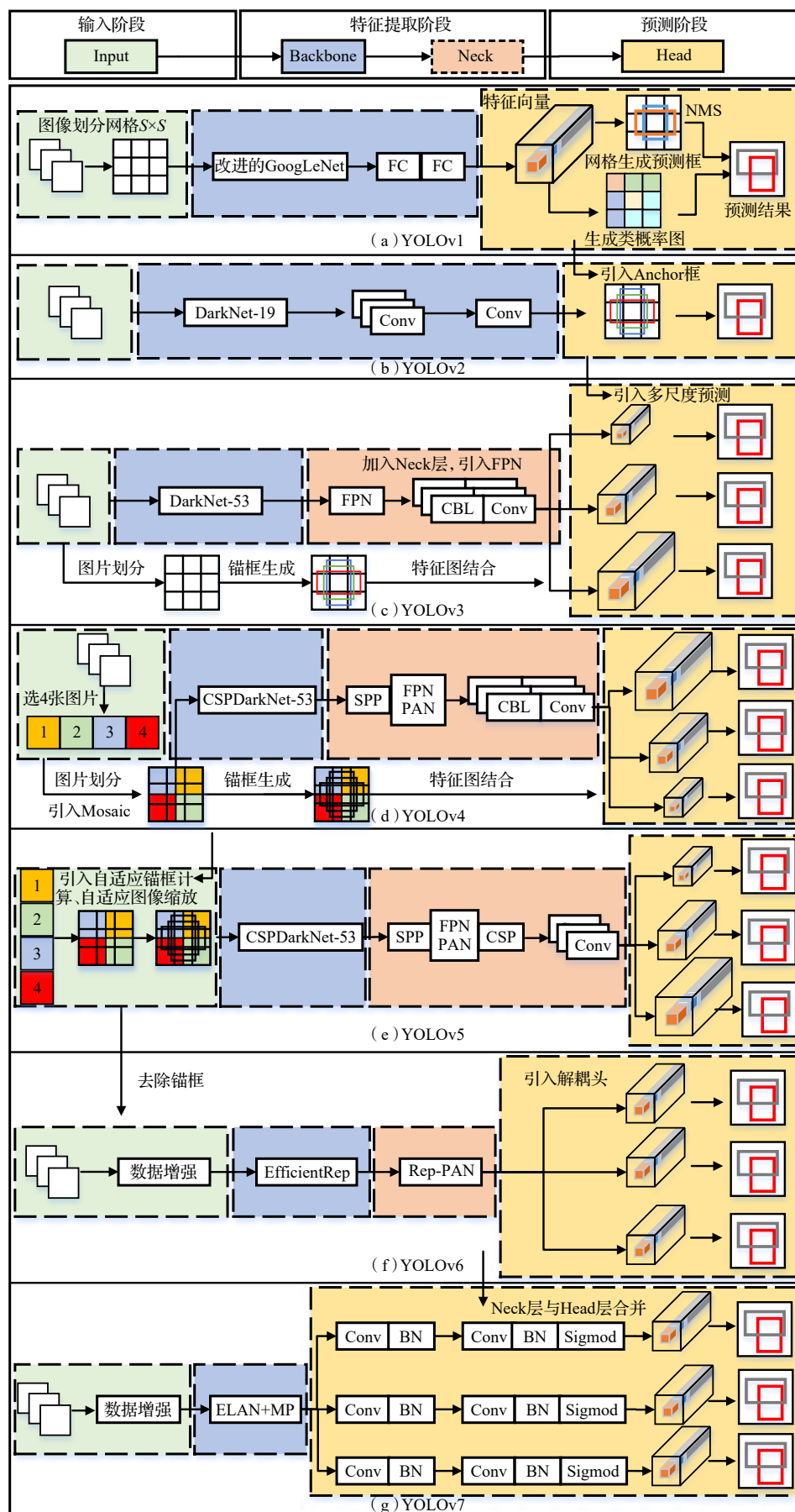


图2 YOLOv1~v7系列算法网络结构简略图

Fig.2 Schematic diagram of YOLOv1~v7 series algorithm network structure

suppression, NMS)^[16]去除冗余的预测框,并生成目标的类别。

2.2 YOLO发展历程

如图2(a)所示,2016年Redmon等人提出了YOLO系列的首个网络YOLOv1,检测流程如下:在输入阶段将图像分成 $s \times s$ 互不重叠的网格区域;在特征提取阶段利用改进的GoogLeNet为主干网络提取图像特征,再经过两个全连接层得到 $s \times s$ 个特征向量,其中每个特征向量包含 b 个矩形预测框和其分属 k 个类别的概率,每个矩形预测框包含矩形框参数 (x, y, w, h) 及其置信度 c ,其中 x 和 y 为预测框的中心点位置, w 和 h 为预测框的宽高, c 为是否包含目标物体的置信度。在预测阶段,每个特征向量要通过NMS过滤低置信度的框,只保留一个置信度最高的预测框,生成最终预测结果。YOLOv1通过直接回归的方式使FPS提升至45,且在PASCAL VOC07和PASCAL VOC12上取得了较高的检测准确率,mAP值分别为63.40%和57.90%。

由于YOLOv1划分的网格规定只能预测两个目标,存在一个区域包含多个目标时无法准确预测小目标以及物体定位不够准确的问题。为此,2017年Redmon等人提出检测与分类联合训练的YOLOv2(由于其可以检测9 000多种类别,也称YOLO9000)。如图2(b)所示,YOLOv2在特征提取阶段改为使用计算量和参数量小的DarkNet-19;在预测阶段进一步优化了损失函数,并在特征图上引入锚框提前预测目标位置;采用多尺度训练方法,以适应不同尺寸的图像。在PASCAL VOC07和PASCAL VOC12上的检测结果得到了显著提升,mAP值分别提高了15.2个百分点和15.5个百分点。

鉴于YOLOv2预测阶段只能进行单一尺度的特征图预测,导致对小目标检测性能差,近距离物体准确率较低的问题,次年,Redmon等人^[17]又提出速度与精度均衡的YOLOv3。如图2(c)所示,该算法在特征提取阶段设计了Darknet-53替换Darknet-19,加深了网络深度;同时,在主干网络和预测层之间加入特征金字塔网络(feature pyramid network, FPN)形成Neck层来融合主干网络提取到的特征;并建立多尺度预测模块,同一时间可检测3个不同尺度的目标对象。相比较而言,YOLOv3保持了YOLO系列检测速度快的特点,并且检测的准确率明显提升,在PASCAL VOC12和MS COCO17上分别为79.30%和31.00%。

2020年,Bochkovskiy等人^[18]在原有YOLOv3的基础上结合深度学习领域的新工作建立了YOLOv4。如图2(d)所示,该方法在输入阶段引入Mosaic、跨小批量归一化(cross mini-batch normal, CmBN)以及自对抗训练(self-adversarial-training, SAT)数据增强方式;在特征提取阶段将主干网络替换为具有跨阶段部分连接(cross-

stage partial connections, CSP)模块的CSPDarknet53,并加入了Mish激活函数和Dropblock正则化,解决了其梯度信息重复问题;在Neck层加入空间金字塔池化(spatial pyramid pooling, SPP)模块,并采用FPN+路径聚合网络(path aggregation network, PAN)的模块取代原有的FPN模块;在预测层优化了训练时的损失函数,使得预测框回归的速度和精度更高一些,在MS COCO17上AP达到了41.20%。

同年,继YOLOv4, YOLOv5^[19]被提出。如图2(e)所示,该方法在输入阶段引入Mosaic数据增强、自适应锚框计算以及自适应图像缩放;在主干网络中融合Focus、CSP模块,在Neck层也引入CSP模块;在预测层改进了训练时的损失函数,其检测效率比先前工作得到了明显提升,在MS COCO17上AP值提升了4.5个百分点,达到了45.70%。

2022年, Li等人^[20]提出了不再使用锚框辅助的YOLOv6。该算法在输入阶段通过数据增强将图像转换成统一尺寸;在特征提取阶段基于RepVGStyle设计了可重参数化、更高效的主干网络EfficientRep和Neck层Rep-PAN;对检测阶段进行了解耦,设计了更简洁有效的Efficient Decoupled检测头,并且采用SimOTA标签分配策略以及SIOU边界框回归损失来进一步提高检测精度, YOLOv6-N与YOLOv5-N相比, mAP值提高了5.5个百分点,达到了51.20%。

同年,为了将YOLO更好地应用于移动GPU设备, Alexey等人设计了YOLOv7。同样不再使用锚框辅助,在输入阶段采用了新的数据增强方法,如Mixup、Copy-Paste、Paste-In等;将特征提取阶段的Backbone替换为BConv层、E-ELAN层以及MPConv层;将Neck层与Head层合并为Head层,组合使用SPPCPC层、BConv层、MPConv层以及RepVGG block层进行特征图预测。该算法在移动设备保持检测速度的情况下明显提高了检测精度, AP达到了69.70%。

表2列举了YOLO系列算法在PASCAL VOC与MS COCO数据集上的检测效果。最初, YOLO系列算法在PASCAL VOC07和PASCAL VOC12上进行训练,随着YOLO算法的发展,其精度和速度明显提升。直到规模更大的MS COCO数据集出现, YOLO系列算法开始转向MS COCO数据集,一开始YOLOv3算法在目标更多、背景更复杂的MS COCO数据集上性能较低,但伴随着YOLO算法不断发展,精度和检测速度不同程度得到提升,并且YOLO算法也发展出不同尺寸的模型,其适用范围和性能也各不相同。

总体来说, YOLO算法发展至今,一方面其整体架构保持不变;另一方面,为提高检测效率和精度,每个方法都针对上一阶段方法的缺点并结合当前新技术进行了改进,取得了算法性能的提升显著。

表2 YOLOv1~v7 算法性能对比

Table 2 Performance comparison of YOLOv1~v7 algorithms

模型	主干网络	尺寸	FPS	PASCAL VOC07(mAP)/%	PASCAL VOC12(mAP)/%	MS COCO17(AP)/%
YOLOv1	改造的GoogLeNet	448	45	63.40	57.90	—
YOLOv2	DarkNet-19	544	40	78.60	73.40	—
YOLOv3	DarkNet-53	416	35	—	79.30	31.00
YOLOv4	CSPDarkNet-53	416	38	—	—	41.20
YOLOv5-N/S/M/L	CSPDarkNet-53	640	400	—	—	45.70
			227			56.80
			147			64.10
			85			67.30
			802			51.20
YOLOv6-N/S/M/L	EfficientRep	640	358	—	—	60.40
			179			66.80
			98			70.00
YOLOv7	ELAN+MP	640	161	—	—	69.70

3 YOLO改进模型

以YOLO系列目标检测网络为基础,针对各类需求,出现了一系列改进工作。如图3,根据YOLO的整体

结构和相关文献[21]的分析,同时考虑卷积神经网络的层次结构特点,将这些改进工作按照YOLO目标检测算法发生阶段,即输入、特征提取和预测,分为三大类,并

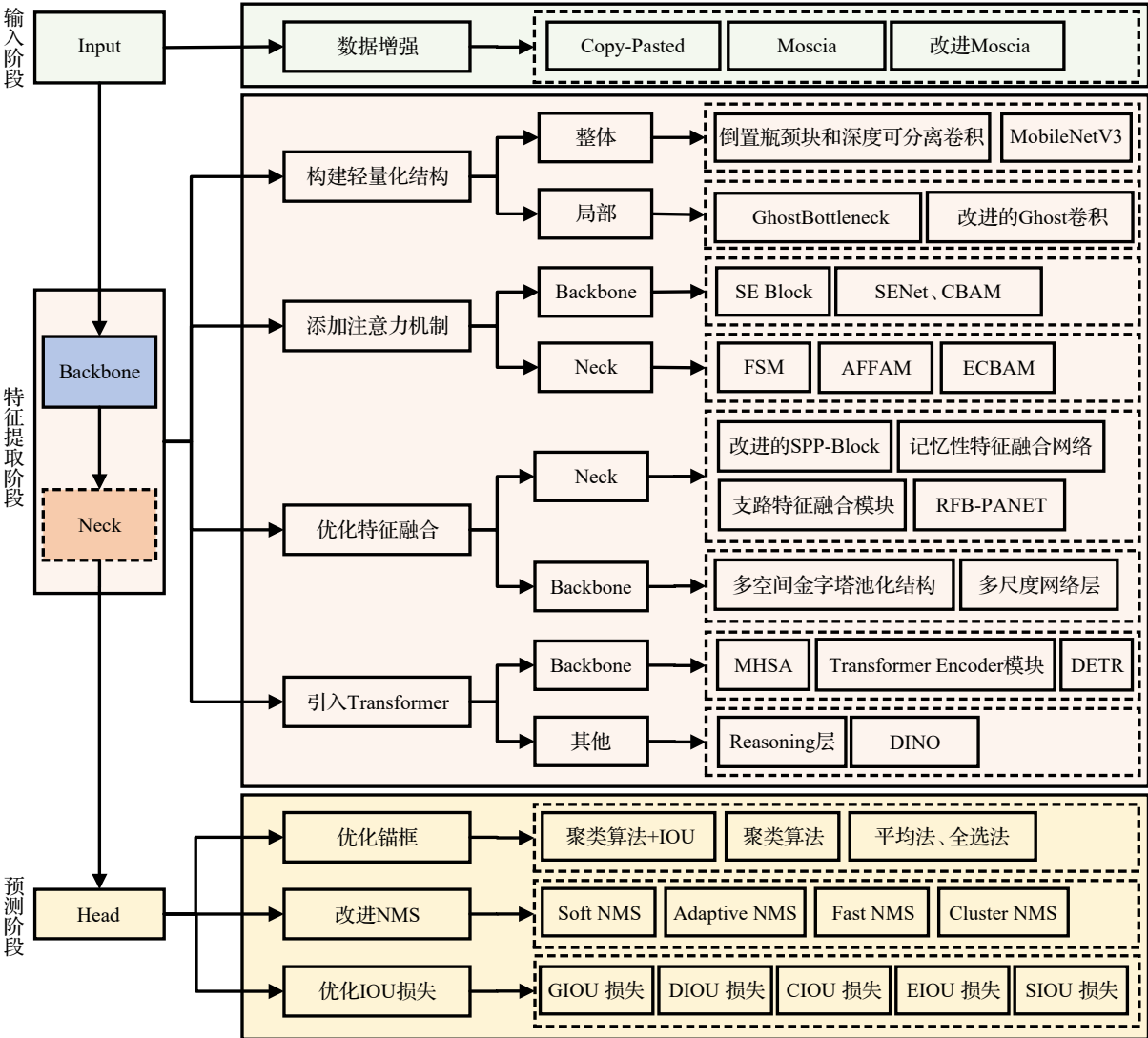


图3 基于YOLO改进的目标检测方法

Fig.3 Improved object detection method based on YOLO

以此为基础按照其改进方向细分为八个子类。

3.1 输入阶段的改进

深度模型是一种数据驱动的学习模型,基于深度学习的YOLO及其改进模型的训练效果直接受输入数据的影响。为此,学者将不同的数据增强方法引入YOLO的输入阶段,通过有限的数据产生更多等价数据,即生成与现有数据潜在分布相同的新数据,以解决数据量少和数据形式单一的问题来提高模型泛化能力。

2022年冷坤等人^[22]针对目前特定数据集数量较少,数据集包含的样本数量不满足于网络模型的训练要求等问题,通过引入Copy-Pasted数据增强方法,将前景目标粘贴到背景图片中生成新的图片,使得样本标签的数量增加。除此之外,2022年张欣怡等人^[23]引入Mosaic数据增强,通过随机缩放、裁剪、排布的方式进行拼接生成新的样本来扩充数据集,该方法提升了小目标检测的能力,并增强了网络的鲁棒性。同年,郭明镇等人^[24]对Mosaic数据增强进行改进,先进行Mosaic操作来拼接图像,然后再对图像进行传统的数据增强操作,该方法减少了进行数据增强的时间,提高了模型的训练效率。

总体而言,学者们根据实际问题的数据缺陷引入相应数据增强方法,解决了数据集样本数量少、泛化性低的问题。只是当前来看,生成数据的质量往往受限于原始数据的数量和质量;且现有的数据增强方法往往独立于目标检测任务,由学者们根据自己的经验预设完成。如何突破原始数据的限制,根据目标检测任务自适应地完成数据增强,以进一步提高目标检测任务仍然是值得研究的问题。

3.2 特征提取阶段的改进

特征提取阶段旨在提取输入图像的基本特征,进而支持后续的目标区域预测,其特征提取的好坏直接影响了网络模型的检测性能。针对其网络规模、运行速度及特征提取能力,学者们提出了一系列改进工作,以提高图像的特征提取能力和特征融合能力,以获取更多的特征信息。

根据改进位置的不同,这些方法可以分为两类:(1)对Backbone的改进,往往利用浅层网络的低级特征和深层网络的高级特征形成互补,使感受野较小的浅层网络捕获更多的细粒度特征,使深层网络获得更多的语义信息,进而提高特征提取能力;(2)对Neck层的改进,通过改进有效融合不同层级的特征图,提高最终特征的表征能力。根据改进角度的不同,这类方法又可以分为构建轻量化结构、引入注意力机制、优化特征融合和引入Transformer模块四个方向,下面将分别介绍。

3.2.1 构建轻量化结构

为提高YOLO的实时检测性能,使其更加适用于移动嵌入设备,2021年Zhang等人^[25]提出Smart-YOLO,其基本思想为将YOLOv3的Backbone改为倒置瓶颈块和

深度可分离卷积,使得结构轻量化;相比YOLOv3模型,其mAP值提升了12.67个百分点,FPS提高了53,模型尺寸(Model Size)减小了206.10 MB,更适合于需要更高速度和效率的场景。2022年何自芬等人^[26]直接引入轻量化的MobileNetV3为Backbone,在网络中添加空间金字塔池化模块与更小感受野的检测层,增强网络输出特征图的表征能力;与基础网络相比,其mAP值略微降低(这是构建轻量化结构遇到的普遍问题),但是FPS提高了19.05。

除对Backbone整体替换外,也有学者针对Backbone的局部引入了轻量化的模块来代替计算量代价大的模块。如,2021年Wu等人^[27]提出YOLOv5-Ghost,该模型将原有Backbone中的BottleneckCSP替换为GhostBottleneck,使模型计算复杂度降低,FPS上升了1.08,Model Size减少了234.30 MB,准确率增长了2个百分点,更适合嵌入式设备。2022年赵凤等人^[28]将YOLOv4-tiny中Backbone的普通卷积替换为改进的Ghost卷积,采用Ghost卷积进一步减少网络参数,Model Size减少了16.1个百分点,mAP值增加了8.3个百分点,提升了准确率和推理速度。

构建轻量化结构减少了模型的参数量和网络模型的大小,可更好地将改进模型移植到移动设备,提升目标检测的效率;但是随着网络深度或者宽度的减小,其检测精度可能会有所下降,如何在精度与速度之间寻找平衡一直是学者们关注的问题。

3.2.2 添加注意力机制

目标检测网络中,有的信息包含目标区域或其上下文信息,格外重要;有的信息仅仅为简单的背景区域,可有可无。为去除冗余,关注核心信息,以提高特征提取能力及目标检测的性能,学者们将注意力机制加入到目标检测网络的特征提取阶段,即,Backbone和Neck层的不同位置,下面分别予以介绍。

针对Backbone,2021年Yan等人^[29]将SE Block模块嵌入至YOLOv5中,以突出特征图中的重要特征,抑制不相关特征,使mAP值提升了1.44个百分点,从而提高网络检测的性能。同年,Li等人^[30]分别将SENet(squeeze-and-excitation networks)和CBAM(convolutional block attention module)引入YOLOv3的Backbone中,自主学习每个通道的权重,增强关键特征,抑制冗余特征。与YOLOv3相比,引入SENet和CBAM使mAP值分别提升了8和8.5个百分点。

针对Neck层,有一部分工作引入了通道注意力。如,2021年Ma等人^[31]提出将特征选择模块(feature select module,FSM)引入Neck层,对每个通道进行评分,过滤掉不相关的特征,关注高贡献的特征,避免特征融合阶段引入过多的不相关特征。该模块分别对YOLOv3、YOLOv4以及YOLOv5-L进行改进,提升效果分别为

0.6、1.1和1.5个百分点。另有一部分工作同时将空间和通道注意力进入至Neck层。如,2020年Ju等人^[32]利用全局注意机制和空间位置注意机制,将基于注意力机制的自适应特征融合方法(adaptive feature fusion with attention mechanism, AFFAM)与YOLOv3结合,同时自适应的学习通道特征的相关性和空间特征在不同尺度上的重要性。该方法在PASCAL VOC07、KITTI和Smart UVM数据集上分别获得了5.08、2.3和7.41个百分点的mAP值提升。2023年陈思雨等人^[33]在YOLOv5的Neck层引入设计的ECBAM注意力模块,通过使生成的特征图得到通道和空间维度上的权重,来提高特征在通道和空间维度上的联系,增强了模型的特征提取能力。在PLAD数据集上mAP值提升了1.6个百分点。

总体而言,在Backbone中引入注意力机制,可使网络关注目标所在的通道和区域,提高特征提取的效度;在Neck层引入注意力机制,可选择性地、自适应地完成特征融合,改善特征融合的有效性。只是,无论在哪个阶段引入注意力机制,都会带来网络复杂度的增加,如何有效平衡代价及其性能增益,如果针对目标检测的特定任务构建更加针对性的注意力机制都还值得进一步研究。

3.2.3 优化特征融合

鉴于目标物体大小不一、结构多变,为确保目标检测率,在特征提取阶段需充分融合不同感受野的多尺度特征;与此同时,背景区域的复杂性导致特征中包含一定的冗余信息,为特征融合带来了困难。针对以上问题,以YOLO系列目标检测网络为基础,学者们针对特征提取阶段的Backbone和Neck层提出了多样化的特征融合方式,下面分别予以说明。

2020年Huang等人^[34]提出DC-SPP-YOLO模型,该方法在Neck层引入改进的SPP-Block对多尺度局部区域特征进行池化和拼接,使网络能够更全面地学习不同尺度的对象特征,提高了网络的小尺度检测能力,在UA-DETRAC和PASCAL VOC12数据集上mAP值分别获得了2.25和1.2个百分点的提升。2022年钱伍等人^[35]在Neck层引入设计的记忆性特征融合网络,通过将顶层的高级语义信息和底层的位置信息进行融合,充分利用了特征信息。该方法在BDD100K数据集和Bosch数据集上与YOLOv5算法比较,分别提升了11和9.3个百分点。同年,王志欣等人^[36]在Neck层上增添了一条特征融合线路,以获取更大尺度的融合特征,并且在Neck层构建多空间金字塔池化结构以扩充网络感受野,增加主干特征接收范围,实现不同尺度的特征融合,使mAP值提升了0.93个百分点。2023年杨锦辉等人^[37]提出多分支的RFB-PANET特征融合方法,每个分支通过不同感受野的空洞卷积提取特征信息,并融合特征信息。该方法在不增加网络深度的情况下,通过较少的参数量获

得了更大的感受野,在PASCAL VOC数据集上,mAP值提升了2.2个百分点。

此外,2021年Yang等人^[38]提出构建与Backbone并行的特征融合模块组成支路网络有效提高网络层之间特征信息的传输和利用,特别是在网络浅层特征信息和深层信息之间的融合交互;进而添加了SPP模块对主网络和支路网络的输出特征进行多尺度聚合,以增强特征图的表征能力,mAP值提高了1.9个百分点。同年,宋艳艳等人^[39]借鉴残差网络的思想将Backbone的深层特征与浅层特征融合生成不同尺度的检测层,提高了网络对不同尺度目标的检测效果,在MOT16数据集上mAP值提升为80.54%。

总体来看,以上方法融合的特征往往来自Backbone或者Neck层的不同层级,这些特征之间感受野大小、语义层次存在差异,且彼此之间既存在互补又存在冗余,如何建立有效的融合手段,去除冗余且综合差异性信息仍然是该类方法面对的难点问题。

3.2.4 引入Transformer模块

针对CNN长距离建模能力较弱的问题,有学者在使用CNN从图像中提取局部信息的基础上,通过在Backbone、Neck层引入Transformer结构使模型更好地聚焦上下文环境进而选择性地聚焦输入图像的有效部分,提升模型目标特征学习的针对性。

2021年Zhang等人^[40]提出VIT-YOLO模型,该方法将多头自注意力层(multi-head self-attention, MHSA)加入Backbone形成MHSA-Darknet来保留足够的全局上下文信息,并通过多头自我注意机制提取更多差异化特征来进行检测,使AP值提高了4.21个百分点。同年,Zhu等人^[41]提出TPH-YOLOv5,该方法将Transformer Block模块引入Backbone底层来捕获全局信息和上下文信息,其mAP值增长明显,增长了10.3个百分点。2022年汤寓麟等人^[42]将DETR(end-to-end object detection with transformers)与YOLOv5的Backbone结合,提出了DETR-YOLO模型,有选择地聚焦输入图像的关键区域,提升模型目标特征学习的准确性。其FPS略有下降,mAP值提高了2.72个百分点。

除上述方法外,还有学者在Backbone与Neck层之间引入Transformer模块。如,2022年Aksoy等人^[43]提出的YOLO-Reasoning将基于注意力的多头推理层集成在Neck层上形成Reasoning层,从而利用更有意义的、细粒度的、增强的特征映射来提取关于不同图像区域之间关系的推理信息。该方法使用不同的Reasoning层使mAP值分别提高了1.27和2.5个百分点。也有学者将YOLO算法与其他算法相互结合形成双阶段检测方法。如,2022年Ouyang^[44]将作为第一阶段的YOLOv5(PRE-DEYO)与作为第二阶段的DINO(POST-DEYO)进行结合,提出基于渐进式的DEYO算法。该方法结合

了两个方法各自的优点,加快了收敛速度,缓解了后期NMS带来的性能问题。

总体而言,引入Transformer可获得长距离关联信息,提高了目标检测的准确率,但这类方法会将检测模型参数增加并且需要大规模的数据集支持,对小型数据集检测效果不明显。

3.3 预测阶段的改进

预测阶段旨在以特征提取阶段输出的特征信息为基础,对目标区域及其类别进行预测。围绕这部分的改进工作主要包括三大类:一类是对引入的锚框机制进行优化处理,以更好地生成合适数目的锚框,更加高效地对预测框进行修正;一类是对NMS机制进行优化,以精准去除多余的预测框;还有一类则对IOU损失,即在真实框与预测框之间的约束机制进行优化,进一步提高预测的精度,下面分别介绍。

3.3.1 优化锚框

为避免盲目搜索,YOLOv2在特征提取阶段生成特征图后引入了锚框,作为目标检测的候选框,在后续阶段针对这些候选框进行分类和微调找到目标中最接近的真实框,实现目标检测。显然,大量的锚框会造成高昂的计算代价,而不足的锚框可能造成漏检、误检等问题。优化锚框的数量和质量,以支持不同尺寸目标的合理检测,并有效平衡检测性能成为学者改进的主要方向。

为了一定程度减少锚框的数量而又不出现漏检等情况,有学者引入 K -means 和 K -means++ 算法对锚框进行聚类。如,2019年Ju等人^[45]将IOU与 K -means 聚类算法结合,使用基于IOU的数学推导方法来选择合适尺寸的锚框,提高了不同尺寸检测的性能。但是 K -means 聚类算法使用的平均重叠度量不满足于常用距离度量公式,2021年姜文志等人^[46]改进度量距离公式,重新设置了锚框的个数与相应参数,提高了模型的整体性能。而对于 K -means 聚类算法可能会造成距离偏差等问题,2022年Ying等人^[47]引入 K -means++ 聚类算法对初始聚类中心进行优化以获得合适的锚框,在保证速度的情况下提高了超小目标的检测精度。除此之外,2020年Liu等人^[48]提出平均法和全选法来选择合适的正锚框数量:平均法为每个目标匹配3个正锚框,输出的每个特征图

有一个正锚框;全选法增加了正锚框的数量,将每个目标匹配9个正锚框,输出的每个特征图有3个正锚框。该方法保证了锚框的数量和质量,提升了检测精度,减少了漏检现象。

3.3.2 改进NMS

NMS,即非极大值抑制,旨在去除冗余的预测框,使目标对象只保留一个预测框。围绕NMS的改进工作可分为提高预测精度和速度两大类。

在精度上,针对原始NMS在处理预测框设置阈值过小时,被遮挡的预测框容易被抑制;阈值过大时,容易造成误检,抑制效果不明显等问题,2017年Bodla等人^[49]提出了Soft NMS,不直接将低分框给过滤掉,而是将低分框的得分降低,避免了阈值设置大小的问题,解决了多个候选框重叠时过滤掉必要候选框的问题。2019年Liu等人^[50]对Soft-NMS进行了优化,提出Adaptive NMS,在目标分布稀疏时使用小阈值,保证尽可能多地去除冗余框,在目标分布密集时采用大阈值,避免漏检。

为提高NMS选取预测框的速度,2019年Bolya等人^[51]利用IOU矩阵为对称矩阵的特性提出Fast NMS,将矩阵的下三角与对角线都设置为零来减少运算量。为弥补因Fast NMS性能下降导致比原NMS抑制更多预测框的问题,2020年Zheng等人^[52]提出Cluster NMS。该算法在通过每次迭代后,忽略了Fast NMS算法在上一次迭代中冗余框对其他框的影响,使得算法可达到原始NMS的准确度。

3.3.3 优化IOU损失

IOU损失表示目标位置预测值和真实值之间的差距,可修正预测框的位置坐标。但最初的IOU预测框和真实框不相交时,不能反映出两个框距离的远近以及预测框和真实框无法反映重合度大小。所以许多学者对IOU进行了一系列改进来提高预测框的准确率,图4为IOU损失发展图。

2019年Rezatofighi等人^[53]提出GIOU损失,其在IOU损失的基础上引入预测框和真实框的最小外接矩形来反映重叠区域的面积以及计算非重叠区域的比例。该方法能更好地反映真实框和预测框的重合程度和远近距离。但由于GIOU损失预测框与真实框包含

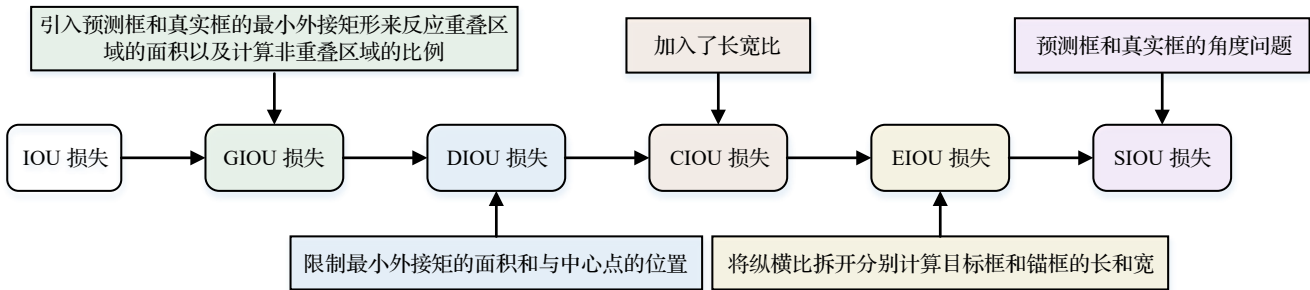


图4 IOU损失发展图

Fig.4 IOU loss development chart

时无法很好地衡量其相对的位置关系,并且预测框在水平或垂直方向优化困难,收敛速度慢。所以,2020年Zheng等人^[54]提出DIOU损失,在预测框与真实框包含时,限制最小外接矩的面积和与中心点的位置,使网络更倾向于移动预测框的位置来减少损失。虽然Zheng等人又加入了长宽比改进为CIOU损失。但由于其反映的是纵横比的差异,而不是宽高分别与其置信度的真实差异,2021年Zhang等人^[55]在CIOU损失的基础上提出EIOU损失,将纵横比拆开分别计算目标框和锚框的长和宽,并且加入Focal损失聚焦优质的锚框,使得收敛速度更快。2022年Gevorgyan等人^[56]考虑预测框和真实框的角度问题,提出SIOU损失。从GIOU损失到SIOU损失,IOU损失的优化迭代不断提升着预测框的检测精度。

总体来看,在特征提取阶段的改进中学者们常用的优化锚框方法还是聚类算法,通过聚类算法生成尺寸更合适的锚框;改进NMS都是从精度与速度的角度来进行,提高NMS的效率;不断优化IOU损失将NMS处理后的预测框与真实框的误差尽量降到最小,提高预测框的准确率,提升算法的性能。

3.4 不同改进方法对比

本文列举了不同类型的YOLO改进方法,表3对YOLO输入阶段、特征提取阶段以及预测阶段改进方法的机制、优势、局限性、适用场景进行了提炼分析和对比。三个阶段中对特征提取阶段的改进方法最多,其次是预测阶段,输入阶段改进方法相对较少,说明改进特征提取阶段和预测阶段对YOLO算法的检测效果明显,并且改进方法丰富可以适用于多种领域场景。表3可为学者研究特定应用领域的YOLO算法提供八种不同的改进思路,每种思路存在不同的机制、优势以及局限性。学者可通过单独使用或组合不同的改进方法来提高YOLO的准确率和检测速度。

4 YOLO 算法应用领域

鉴于YOLO算法检测速度快、目标识别效果较好、检测实时性好等优点,该算法在各个领域广泛应用于目标检测任务。特别是在工业、行人、农业、医学以及遥感等领域^[57],相对于传统方法,YOLO算法取得了显著的效果。

(1)工业检测

在工业领域YOLO算法被广泛应用于安全检测、零件检测和缺陷检测等任务。例如,2021年金雨芳等人^[58]为提高安全帽佩戴检测算法的性能,提出了一种基于改进YOLOv4的安全帽佩戴检测算法;2022年冯晨光等人^[59]对YOLOv4进行了剪枝和优化,提出了一种用于检测控制箱零件的轻量级算法SlimYOLO;而2023年,胡欣等人^[60]提出了一种改进YOLOv5的螺纹钢表面缺陷检测算法。

总体上来说,这些算法的提出和实践,进一步促进了工业生产过程的自动化和智能化,为工业检测应用提供了更加可靠和高效的解决方案。但是工业场景数据集公开较少,因此可能不适用于特定的检测任务,这是需要继续发展和完善的领域。此外,由于工业场景的背景比较复杂,这也容易影响检测效果,并且工业级应用对检测速度和精度要求较高,这些问题都是工业检测研究的重点。因此,为了进一步推动工业检测算法的发展和应用,需要不断探索和创新,在数据集的建立、算法优化、检测精度和速度等方面进行深入研究,以满足工业应用的实际需求。

(2)行人检测

行人检测广泛应用于自动驾驶的行人预警系统、街道监控和交通检测等领域。例如,2021年邓杰等人^[61]提出了一种基于YOLOv3的密集行人检测算法Crowd-

表3 YOLO改进方法对比
Table 3 Comparison of YOLO improvement methods

阶段	改进方法	机制	优势	局限性	适用场景
输入阶段	数据增强	通过一系列方法增加图像的数量	扩充了数据集数量,增加了数据集样本多样性	扩充后的数据集与原数据集不是完全不同,数据集质量有待提高	缺少数据集,泛化性、鲁棒性不足
	构建轻量化结构	减小网络的深度与宽度以及模型的参数量	网络模型小、速度快、实时性高,更适用于嵌入式设备	适合于需要更高速度和效率的场景,准确率与普通算法相比较低一些	实时目标检测
特征提取阶段	添加注意力机制	增强网络模型的特征提取能力	可以缓解网络模型梯度消失问题	可能会加深网络模型的深度,增加参数量	密集目标检测
	优化特征融合	融合浅层与深层结构的特征信息	更好地融合语义和尺度不一致的特征,并提高准确性	存在把无效的特征信息也融合进去,参数量变大	实时目标检测
	引入Transformer	提取有效的特征信息	可以使网络提取到感兴趣部分的特征信息	增加了网络深度和参数量,导致计算时间增加	小目标检测
	优化锚框	人为或自适应设置锚框尺寸	可以得到合适的锚框	人为设置的锚框存在偏差	目标对象尺寸与原锚框设定尺寸差异较大
预测阶段	改进NMS	选取尺寸更适合的预测框	降低漏检与错检的概率	锚框的选取精度和速度无法同时取得最优结果	存在漏检与错检
	优化IOU损失	通过定义的公式减小预测框与真实框的误差	提高了模型的定位能力	优化IOU可能导致计算复杂	模型精确度低

YOLO;2022年常青等人^[62]则基于YOLOv4-tiny提出了一种改进的行人检测算法,以应对当前行人检测方法计算量大、检测精度低等问题;此外,向南等人^[63]也提出了一种基于YOLO的遮挡行人检测算法YOLO-SSC-p。

行人检测技术在未来的交通安全和智慧城市建设中将会发挥越来越重要的作用。但目前行人检测仍然存在一些明显的问题。例如,行人之间可能存在密集遮挡,远处的行人难以清晰判断,当路上周围环境复杂时可能会产生误判等问题。这些问题严重影响了行人检测的准确性和可靠性,需要通过持续的研究和技术进步来解决。

(3) 农业检测

YOLO算法在农业检测应用上被广泛应用于植物叶片检测、果实检测、害虫识别等方面。例如,2023年张帆等人^[64]提出基于YOLOv3的玉米叶片气孔检测算法;2023年郝鹏飞等人^[65]提出果园遮挡果实检测模型YOLO-RD-Apple;2023年冯娟等人^[66]提出基于深度学习的单环刺螈洞口识别方法。

虽然YOLO算法在农业检测中取得了显著的成果,但农业场景的特殊性使得该算法在实际应用中仍然存在一些问题,如不同作物间的差异性和变异性、天气和光照条件对检测效果的影响等。因此,继续对该算法进行改进和优化,以满足农业场景的需求,仍然是研究重点。

(4) 医学检测

YOLO医学检测方法被广泛应用于心脏、白细胞和血细胞等诊断中。例如,2020年,Zhuang等人^[67]为了实现心脏液体运动的视觉分析,根据心脏血流超声图像的特点,提出了一种基于YOLO的心脏血流图分析与评价方法;同年,Sharif等人^[68]提出了一种基于YOLOv2的白细胞定位方法;而在2022年,王榆锋等人^[69]为了解决传统目标检测算法在血细胞检测任务中出现的检测精度偏低、错检及漏检等问题,提出了一种基于YOLO的目标检测算法,即YOLO-Att。

可以看出YOLO算法在医学领域的应用正逐步扩展,并取得了一定的进展。但是医学检测仍然存在所需数据集不足的问题,这些数据集需要进行病理学分析,并且需要专业医生对其进行有效标注。只有将大量的医学图像转化为有效的数据集信息,才能更好地促进医学检测的发展。

(5) 遥感检测

如今,遥感检测已广泛应用于城市规划、地质灾害探测、海洋检测等民用或军用领域。例如,2021年陈静等人^[70]提出一种尺度和密度约束下基于YOLOv3模型的风电塔架遥感检测方法;2022年肖振久等人^[71]提出基于YOLOv4的遥感图像目标检测方法,该方法在保持较高检测精度的同时,减少了参数量,提高了检测速度;同年,闫钧华等人^[72]为解决遥感图像地面弱小目标检测中

弱小目标信息量少、信息真假混杂的难题,提出了遥感图像地面弱小目标检测算法CC-YOLO。

总体来说,改进后的遥感检测算法有效地提高了检测遥感图像的准确率。但是由于遥感图像具有的特点,处理这些图像需要大量计算资源。此外,由于遥感图像的背景比较复杂,因此提取有效信息较为困难。所以构建性能良好的网络模型是遥感检测领域发展的重要方向。

5 总结与展望

本文总结了YOLO系列算法发展历程及改进研究。首先介绍了基于深度学习的目标检测并引出YOLO算法;其次介绍了YOLO典型数据集以及评价指标;然后详细地阐述了YOLO整体框架和发展历程;通过对现有工作的分析,分别从输入阶段、特征提取阶段以及预测阶段三个方向,总结出常用的改进方法;最后对YOLO算法应用领域进行阐述。

随着计算能力和数据集规模的不断增加,YOLO的性能得到进一步提升,也衍生出许多性能优越的改进算法,但将其已应用于实际场景时仍然面临一些问题^[73-74],如复杂场景下的密集遮挡问题、特殊场景下的无监督数据问题以及实时场景下的检测效率问题。本文将围绕这些实际需求进行分析展望。

(1)密集遮挡目标检测。在行人检测、人脸检测以及工业零件检测等领域中存在密集遮挡可能会造成YOLO算法漏检和错检的问题,给YOLO算法带来了巨大的挑战。遮挡部分阻碍了对目标对象进行特征提取,导致无法对目标对象进行准确定位和精准识别。今后在密集遮挡目标检测时可适当减少使用下采样方法,减少特征信息的丢失;对不同领域的检测目标设置合适的锚框尺寸;在主干网络增加注意力模块,将提取到的有效部分的特征与被遮挡目标周围相同的目标特征进行弥补和融合,充分利用上下文信息来进行检测。

(2)无监督目标检测。现阶段YOLO算法是依靠人工标注的强监督数据集来进行学习,无法脱离标记数据。在YOLO算法应用的领域中存在各种各样的数据集,不同数据集的图像数量和样本数量各不相同。图像和样本数量越多,标注工作量就会越大,非常耗时,可能还存在标注错误,重要的是标注的质量决定模型训练的程度。文献[75]专注于在没有人工标记的自然图像中进行无监督目标检测,以减少对人工标记数据的依赖,提升检测效率。发展无监督目标检测必将是目标检测的新方向。

(3)实时目标检测。在保证检测精度的情况下,自动驾驶、视频检测以及火灾检测对检测的实时性有很高的要求。YOLO算法自身的特点就是比其他方法检测速度快,但是在移动设备或者设备硬件较差的环境下,如何能保证准确率和速度之间的平衡。并且基于YOLO

算法改进的大多数网络结构为增强特征提取和融合,网络结构设计十分复杂,每个阶段的参数计算量大,网络的训练周期长,嵌入到移动设备不友好。可在今后改进中采用更轻量化的网络结构或者减少不必要的模块来提高在移动设备的兼容性,使运行反应速度和目标检测度达到相对的平衡。

此外,要想进一步推进 YOLO 算法的发展,数据集的构建也是必不可少的。当前 YOLO 广泛应用于各大领域,但是现有目标检测数据集缺少多样性,大部分为自然背景下标记的数据集,无法适用于特定场景,面向的都是固定的检测场景。今后工作可针对不同目标检测需求构建特定数据集,弥补不同领域的数据集缺陷,使数据集具有多样性,能够让 YOLO 算法更好地应用在各个领域实时目标检测。

6 结束语

综上所述,YOLO 算法在经过不断优化改进之后,检测性能大幅度提升,已经成为单阶段目标检测的主流方法。但 YOLO 算法应用到更加复杂的现实场景,还存在漏检、误检以及实时性和准确率低等问题。因此,针对不同应用场景,需结合实际问题采用不同的改进方法,进一步提升 YOLO 算法的检测性能,才能将 YOLO 算法扩展到更多应用领域。

参考文献:

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [2] 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.
- [3] WANG C Y, BOCHKOVSKIY A, LIAO H. 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.
- [4] 叶赵兵, 段先华, 赵楚. 改进 YOLOv3-SPP 水下目标检测研究[J]. 计算机工程与应用, 2023, 59(6): 231-240.
YE Z B, DUAN X H, ZHAO C. Research on underwater target detection by improved YOLOv3-SPP[J]. Computer Engineering and Applications, 2023, 59(6): 231-240.
- [5] 王建波, 武友新. 改进 YOLOv4-tiny 的安全帽佩戴检测算法[J]. 计算机工程与应用, 2023, 59(4): 183-190.
WANG J B, WU Y X. Helmet wearing detection algorithm of improved YOLOv4-tiny[J]. Computer Engineering and Applications, 2023, 59(4): 183-190.
- [6] TAN Y, CAI R, LI J, et al. Automatic detection of sewer defects based on improved you only look once algorithm[J]. Automation in Construction, 2021, 131(6): 103912-103928.
- [7] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes(VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [8] EVERINGHAM M, ESLAMI S M A, VAN GOOL L, et al. The pascal visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.
- [9] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//European Conference on Computer Vision. Cham: Springer, 2014: 740-755.
- [10] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [11] KUZNETSOVA A, ROM H, ALLDRIN N, et al. The open images dataset v4[J]. International Journal of Computer Vision, 2020, 128(7): 1956-1981.
- [12] XIA G S, BAI X, DING J, et al. Dota: a large-scale dataset for object detection in aerial images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3974-3983.
- [13] ZAIDI S, ANSARI M S, ASLAM A, et al. A survey of modern deep learning based object detection models[J]. Digital Signal Processing, 2022, 126: 103514-103530.
- [14] ZOU Z X, SHI Z W, GUO Y H, et al. Object detection in 20 years: a survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.
- [15] REDMON J, FARHADI A. Yolo9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [16] NEUBECK A, GOOL L. Efficient non-maximum suppression[C]//International Conference on Pattern Recognition, 2006: 850-855.
- [17] REDMON J, FARHADI A. Yolo v3: an incremental improvement[J]. arXiv: 1804.02767, 2018.
- [18] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolo v4: optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.
- [19] JOCHER G. Yolo v5[EB/OL]. [2023-03-20]. <https://github.com/ultralytics/yolov5>.
- [20] LI C Y, LI L L, JIANG H L, et al. Yolo v6: a single-stage object detection framework for industrial applications[J]. arXiv: 2209.02976, 2022.
- [21] 王鑫鹏, 王晓强, 林浩, 等. 深度学习典型目标检测算法的改进综述[J]. 计算机工程与应用, 2022, 58(6): 42-57.
WANG X P, WANG X Q, LIN H, et al. Review on improvement of typical object detection algorithms in deep learn-

- ing[J].Computer Engineering and Applications,2022,58(6):42-57.
- [22] 冷坤,秦伦明,王悉.基于CA-ASFF-YOLOv4的交通标志识别研究[J/OL].计算机工程与应用(2022-12-09)[2022-12-30].<https://kns.cnki.net/kcms/detail//11.2127.TP.20221208.1745.004.html>.
- LENG K, QIN L M, WANG X. Research on traffic sign recognition based on CA-ASFF-YOLOv4[J/OL].Computer Engineering and Applications(2022-12-09)[2022-12-30].<https://kns.cnki.net/kcms/detail//11.2127.TP.20221208.1745.004.html>.
- [23] 张欣怡,张飞,郝斌,等.基于改进YOLOv5的口罩佩戴检测算法[J/OL].计算机工程(2022-12-09)[2022-12-30].<https://doi.org/10.19678/j.issn.1000-3428.0065701>.
- ZHANG X Y, ZHANG F, HAO B, et al. Improved YOLOv5s in mask wearing detection algorithm[J/OL].Computer Engineering(2022-12-09)[2022-12-30].<https://doi.org/10.19678/j.issn.1000-3428.0065701>.
- [24] 郭明镇,汪威,申红婷,等.改进型YOLOv4-tiny的轻量级目标检测算法[J/OL].计算机工程与应用(2022-11-26)[2022-12-30].<https://kns.cnki.net/kcms/detail//11.2127.TP.20221125.1132.016.html>.
- GUO M Z, WANG W, SHEN H T, et al. Improved light-weight target detection algorithm for YOLOv4-tiny[J/OL].Computer Engineering and Applications(2022-11-26)[2022-12-30].<https://kns.cnki.net/kcms/detail//11.2127.TP.20221125.1132.016.html>.
- [25] ZHANG D Y, CHEN X Y, REN Y M, et al. Smart-YOLO: a light-weight real-time object detection network[J].Journal of Physics:Conference Series,2021,1757(1):012096.
- [26] 何自芬,陈光晨,陈俊松,等.多尺度特征融合轻量化夜间红外行人实时检测[J].中国激光,2022,49(17):130-139.
- HE Z F, CHEN G C, CHEN J S, et al. Multi-scale feature fusion lightweight real-time infrared pedestrian detection at night[J].Chinese Journal of Lasers,2022,49(17):130-139.
- [27] WU T H, WANG T W, LIU Y Q. Real-time vehicle and distance detection based on improved YOLOv5 network[C]//2021 IEEE World Symposium on Artificial Intelligence(WSAI),2021:24-28.
- [28] 赵凤,李永恒,李晶,等.基于改进YOLOv4-tiny的轻量化室内人员目标检测算法[J].电子与信息学报,2022,44(11):3815-3824.
- ZHAO F, LI Y H, LI J, et al. Lightweight indoor personnel detection algorithm based on improved YOLOv4-tiny[J].Journal of Electronics & Information Technology,2022,44(11):3815-3824.
- [29] YAN F, XU Y. Improved target detection algorithm based on YOLO[C]//2021 IEEE International Conference on Robotics, Control and Automation Engineering(RCAE),2021:21-25.
- [30] LI J C, WANG H Z, XU Y, et al. Road object detection of YOLO algorithm with attention mechanism[J].Frontiers in Signal Processing,2021,5(1):9-16.
- [31] MA Y J, ZHANG S H. Feature selection module for CNN based object detector[J].IEEE Access,2021,9:69456-69466.
- [32] JU M R, LUO J N, WANG Z B, et al. Adaptive feature fusion with attention mechanism for multi-scale target detection[J].Neural Computing and Applications,2020,33(7):2769-2781.
- [33] 陈思雨,付章杰.融合高效注意力的多尺度输电线路部件检测[J/OL].计算机工程与应用(2023-01-03)[2023-01-18].<https://kns.cnki.net/kcms/detail//11.2127.TP.20230103.1221.002.html>.
- CHEN S Y, FU Z J, et al. Multi-scale transmission line component detection incorporating efficient attention[J/OL].Computer Engineering and Applications(2023-01-03)[2023-01-18].<https://kns.cnki.net/kcms/detail//11.2127.TP.20230103.1221.002.html>.
- [34] HUANG Z C, WANG J L, FU X S, et al. DC-SPP-YOLO: dense connection and spatial pyramid pooling based YOLO for object detection[J].Information Sciences,2020,522:241-258.
- [35] 钱伍,王国中,李国平.改进YOLOv5的交通灯实时检测鲁棒算法[J].计算机科学与探索,2022,16(1):231-241.
- QIAN W, WANG G Z, LI G P. Improved YOLOv5 traffic light real-time detection robust algorithm[J].Journal of Frontiers of Computer Science and Technology,2022,16(1):231-241.
- [36] 王志欣,万绍俊,马晓莹.改进锚点框与融合多尺度特征的光学遥感目标检测[J].无线电工程,2021,51(9):915-920.
- WANG Z X, WAN S J, MA X Y. Optical remote sensing target detection based on improved anchor frames and fused multi-scale features[J].Radio Engineering,2021,51(9):915-920.
- [37] 杨锦辉,李鸿,杜芸彦,等.基于改进YOLOv5s的轻量化目标检测算法[J].电光与控制,2023,30(2):24-30.
- YANG J H, LI H, DU Y Y, et al. Lightweight object detection algorithm based on improved YOLOv5s[J].Electronics Optics & Control,2023,30(2):24-30.
- [38] YANG Y H, LI B. Water area object detection based on YOLO-fusion network[J].International Core Journal of Engineering,2021,7(5):100-107.
- [39] 宋艳艳,谭励,马子豪,等.改进YOLOV3算法的视频目标检测[J].计算机科学与探索,2021,15(1):163-172.
- SONG Y Y, TAN L, MA Z H, et al. Video target detection based on improved YOLOV3 algorithm[J].Journal of Frontiers of Computer Science and Technology,2021,15(1):163-172.

- [40] ZHANG Z, LU X, CAO G, et al. ViT-YOLO: transformer based YOLO for object detection[C]//2021 IEEE International Conference on Computer Vision (ICCV), 2021: 2799-2808.
- [41] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021: 2778-2788.
- [42] 汤寓麟, 李厚朴, 张卫东, 等. 侧扫声纳检测沉船目标的轻量化 DETR-YOLO 法[J]. 系统工程与电子技术, 2022, 44(8): 2427-2436.
- TANG Y L, LI H P, ZHANG W D, et al. Lightweight DETR-YOLO method for detecting shipwreck target in side-scan sonar[J]. Systems Engineering and Electronics, 2022, 44(8): 2427-2436.
- [43] AKSOY T, HALICI U. Analysis of visual reasoning on one-stage object detection[J]. arXiv:2202.13115, 2022.
- [44] OUYANG H. DEYO: DETR with YOLO for step-by-step object detection[J]. arXiv:2211.06588, 2022.
- [45] JU M R, LUO H B, WANG Z B, et al. The application of improved YOLOv3 in multi-scale target detection[J]. Applied Sciences, 2019, 9(18): 3775-3788.
- [46] 姜文志, 李炳臻, 顾佼佼, 等. 基于改进 YOLO V3 的舰船目标检测算法[J]. 电光与控制, 2021, 28(6): 52-56.
- JIANG W Z, LI B Z, GU J J, et al. A ship target detection algorithm based on improved YOLO V3[J]. Electronics Optics & Control, 2021, 28(6): 52-56.
- [47] YING Z P, LIN Z T, WU Z Y, et al. A modified-YOLOv5s model for detection of wire braided hose defects[J]. Measurement, 2022, 190: 110683-110693.
- [48] LIU T, PANG B, AI S M, et al. Study on visual detection algorithm of sea surface targets based on improved YOLOv3[J]. Sensors, 2020, 20(24): 7263-7276.
- [49] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS-improving object detection with one line of code[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 5561-5569.
- [50] LIU S, HUANG D, WANG Y. Adaptive NMS: refining pedestrian detection in a crowd[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 6459-6468.
- [51] BOLYA D, ZHOU C, XIAO F, et al. YOLACT: real-time instance segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9157-9166.
- [52] ZHENG Z, WANG P, REN D, et al. Enhancing geo-metric factors in model learning and inference for object detection and instance segmentation[J]. arXiv:2005.03572, 2020.
- [53] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.
- [54] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 12993-13000.
- [55] 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.
- [56] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression[J]. arXiv:2205.12740, 2022.
- [57] 许德刚, 王露, 李凡. 深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用, 2021, 57(8): 10-25.
- XU D G, WANG L, LI F. Review of typical object detection algorithms for deep learning[J]. Computer Engineering and Applications, 2021, 57(8): 10-25.
- [58] 金雨芳, 吴祥, 董辉, 等. 基于改进 YOLOv4 的安全帽佩戴检测算法[J]. 计算机科学, 2021, 48(11): 268-275.
- JIN Y F, WU X, DONG H, et al. Improved YOLOv4 algorithm for safety helmet wearing detection[J]. Computer Science, 2021, 48(11): 268-275.
- [59] 冯晨光, 魏巍, 陈灯, 等. 基于 SlimYOLO 的控制箱零件检测方法[J]. 电子测量技术, 2022, 45(17): 120-126.
- FENG C G, WEI W, CHEN D, et al. Detection method of electrical cabinet parts based on SlimYOLO[J]. Electronic Measurement Technology, 2022, 45(17): 120-126.
- [60] 胡欣, 周运强, 肖剑, 等. 基于改进 YOLOv5 的螺纹钢表面缺陷检测[J/OL]. 图学学报(2023-01-06)[2023-03-17]. <https://kns.cnki.net/kcms/detail/10.1034.T.20230106.1212.003.html>.
- HU X, ZHOU Y Q, XIAO J, et al. Surface defect detection of threaded steel based on improved YOLOv5[J/OL]. Journal of Graphics (2023-01-06) [2023-03-17]. <https://kns.cnki.net/kcms/detail/10.1034.T.20230106.1212.003.html>.
- [61] 邓杰, 万旺根. 基于改进 YOLOv3 的密集行人检测[J]. 电子测量技术, 2021, 44(11): 90-95.
- DENG J, WAN W G. Dense pedestrian detection based on improved YOLOv3[J]. Electronic Measurement Technology, 2021, 44(11): 90-95.
- [62] 常青, 韩文, 王清华, 等. 改进 YOLO 轻量化网络的行人检测算法[J]. 光学技术, 2022, 48(1): 80-85.
- CHANG Q, HAN W, WANG Q H, et al. Pedestrian detection algorithm based on improved YOLO lightweight network[J]. Optical Technique, 2022, 48(1): 80-85.
- [63] 向南, 王璐, 贾崇柳, 等. 改进 YOLO 的遮挡行人检测仿真[J]. 系统仿真学报, 2023, 35(2): 286-299.
- XIANG N, WANG L, JIA C L, et al. Simulation of occluded pedestrian detection based on improved YOLO[J]. Journal of System Simulation, 2023, 35(2): 286-299.

- [64] 张帆,郭思媛,任方涛,等.基于改进YOLO v3的玉米叶片气孔自动识别与测量方法[J].农业机械学报,2023,54(2):216-222.
ZHANG F, GUO S Y, REN F T, et al. Automatic identification and measurement of maize leaves stomata based on YOLO v3[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 216-222.
- [65] 郝鹏飞,刘立群,顾任远. YOLO-RD-Apple 果园异源图像遮挡果实检测模型[J/OL]. 图学学报(2023-02-01)[2023-03-17]. <http://kns.cnki.net/kcms/detail/10.1034.T.20230201.1105.001.html>.
HAO P F, LIU L Q, GU R Y. YOLO-RD-Apple orchard heterogenous image obscured fruit detection model[J/OL]. Journal of Graphics(2023-02-01)[2023-03-17]. <http://kns.cnki.net/kcms/detail/10.1034.T.20230201.1105.001.html>.
- [66] 冯娟,梁翔宇,曾立华,等.基于改进YOLO v4的单环刺蛾洞口识别方法[J].农业机械学报,2023,54(2):265-274.
FENG J, LIANG X Y, ZENG L H, et al. Urechis unicinctus burrows recognition method based on improved YOLO v4[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 265-274.
- [67] ZHUANG Z, LIU G, DING W, et al. Cardiac VFM visualization and analysis based on YOLO deep learning model and modified 2D continuity equation[J]. Computerized Medical Imaging and Graphics, 2020, 82: 101732-101743.
- [68] SHARIF M, AMIN J, SIDDIQA A, et al. Recognition of different types of leukocytes using YOLOv2 and optimized bag-of-features[J]. IEEE Access, 2020, 8: 167448-167459.
- [69] 王榆锋,李大海.改进YOLO框架的血细胞检测算法[J].计算机工程与应用,2022,58(12):191-198.
WANG Y F, LI D H. Improved YOLO framework blood cell detection algorithm[J]. Computer Engineering and Applications, 2022, 58(12): 191-198.
- [70] 陈静,陈静波,孟瑜,等.尺度和密度约束下基于YOLOv3的风电塔架遥感检测方法[J].自然资源遥感,2021,33(3):54-62.
CHEN J, CHEN J B, MENG Y, et al. Detection of wind turbine towers in remote sensing based on YOLOv3 model under scale and density constraints[J]. Remote Sensing for Natural Resources, 2021, 33(3): 54-62.
- [71] 肖振久,杨玥莹,孔祥旭.基于改进YOLOv4的遥感图像目标检测方法[J].激光与光电子学进展,2023,60(6):407-415.
XIAO Z J, YANG Y Y, KONG X X. Object detection method based on improved YOLOv4 network for remote sensing images[J]. Laser & Optoelectronics Progress, 2023, 60(6): 407-415.
- [72] 闫钧华,张琨,施天俊,等.融合多层次特征的遥感图像地面弱小目标检测[J].仪器仪表学报,2022,43(3):221-229.
YAN J H, ZHANG K, SHI T J, et al. Multi-level feature fusion based dim small ground target detection in remote sensing images[J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 221-229.
- [73] 邵延华,张铎,楚红雨,等.基于深度学习的YOLO目标检测综述[J].电子与信息学报,2022,44(10):3697-3708.
SHAO Y H, ZHANG D, CHU H Y, et al. A review of YOLO object detection based on deep learning[J]. Journal of Electronics & Information Technology, 2022, 44(10): 3697-3708.
- [74] 李科岑,王晓强,林浩,等.深度学习中的单阶段小目标检测方法综述[J].计算机科学与探索,2022,16(1):41-58.
LI K C, WANG X Q, LIN H, et al. Survey of one-stage small object detection methods in deep learning[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(1): 41-58.
- [75] WANG Y, SHEN X, HU S X, et al. Self-supervised transformers for unsupervised object discovery using normalized cut[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 14523-14533.