

YOLO 算法在目标检测中的研究进展

耿 创^{1,2}, 宋品德^{1,2}, 曹立佳^{2,3,4}

- (1. 四川轻化工大学 自动化与信息工程学院, 四川 宜宾 644000;
2. 桥梁无损检测与工程计算四川省高校重点实验室, 四川 宜宾 644000;
3. 四川轻化工大学 计算机科学与工程学院, 四川 宜宾 644000;
4. 人工智能四川省重点实验室, 四川 宜宾 644000)

摘要:YOLO 系列算法以其优异的检测性能成为目标检测领域的热点问题。回顾了目标检测领域的发展历程,综述了基于深度学习的两类目标检测算法。分别以 YOLO 系列 3 个常用算法(YOLO V2、YOLO V3 和 YOLO V4),在复杂气象条件、复杂背景或复杂姿态下的目标检测任务为切入点,介绍了针对 YOLO 做出的改进,如:改进特征图、调整分辨率、优化 NMS、替换目标函数、融合多模态、修改检测框等措施,并分析了改进算法与原算法的性能。分析研究了目标检测领域常用的数据集,总结了当前 YOLO 算法在目标检测领域存在的问题以及未来的研究发展趋势。

关键词:YOLO; 目标检测; 深度学习; 改进算法; 数据集

本文引用格式:耿创,宋品德,曹立佳. YOLO 算法在目标检测中的研究进展[J]. 兵器装备工程学报,2022,43(09): 162-173.

Citation format:GENG Chuang, SONG Pinde, CAO Lijia. Research progress of YOLO algorithm in target detection[J]. Journal of Ordnance Equipment Engineering,2022,43(09):162-173.

中图分类号:TP75

文献标识码:A

文章编号:2096-2304(2022)09-0162-12

Research progress of YOLO algorithm in target detection

GENG Chuang^{1,2}, SONG Pinde^{1,2}, CAO Lijia^{2,3,4}

- (1. School of Automation & Information Engineering, Sichuan University of Science & Engineering, Yibin 644000, China;
2. Sichuan Province University Key Laboratory of Bridge Non-destruction Detecting and Engineering Computing, Yibin 644000, China; 3. School of Computer Science & Engineering, Sichuan University of Science & Engineering, Yibin 644000, China;
4. Artificial Intelligence Key Laboratory of Sichuan Province, Yibin 644000, China)

Abstract:YOLO series algorithms have become a hot issue in the field of target detection with their excellent detection performance. We reviewed the development process of the target detection field, and systematically reviewed two types of target detection algorithms based on deep learning. From the three commonly used algorithms of the YOLO series (YOLO V2, YOLO V3 and YOLO V4) in the field of target detection, the target detection tasks under complex weather conditions, complex backgrounds, or complex attitudes were proposed. We introduced the improvement measures made by the researchers to improve the feature map, adjust the resolution, optimize the NMS algorithm, replaced the objective function, fuse multi-modality, modify the detection frame, etc. and analyzed the performance of the improved algorithm

收稿日期:2021-11-12;修回日期:2021-12-08

基金项目:四川省科技计划项目(2020YJ0368);自贡市科技计划重点项目(2019YJ03);桥梁无损检测与工程计算四川省高校重点实验室开放课题基金(2021QZJ01);研究生创新创业基金项目(y2021064)

作者简介:耿创(1998—),男,硕士研究生, E-mail:geng_c2222@163.com。

通信作者:曹立佳(1982—),男,副教授,硕士生导师, E-mail:caolj@suse.edu.cn。

and the original algorithm. Then we summarized the commonly used data sets in the field of target detection, and summarized the current problems of the YOLO algorithm in the field of target detection, in the end, looked forward to the future research and development.

Key words: YOLO; target detection; deep learning; improved algorithm; dataset

1 引言

目标检测是计算机视觉领域重要的研究方向之一,主要任务是根据用户感兴趣目标信息找出输入图像中所有符合要求的目标,并对目标所属类别和所处位置进行分析。在实际生活中,输入图像往往因为目标在外观、颜色、姿态上的差异,加上采集图像数据时光照、遮挡、抖动、辐射等外界因素的干扰或拍照要求的限制,导致成像质量不佳,严重影响目标检测算法的定位精度,使得定位精度问题一直是目标检测领域极具挑战性的焦点问题。

传统的目标检测方法借助机器学习算法,通过不同尺度的滑动窗口确定目标候选区域,对区域内目标特征如梯度直方图特征(histogram of oriented gradients, HOG)^[1]、尺度不变特征(scale invariant feature transform, SIFT)^[2]、局部二值化特征(local binary pattern, LBP)^[3]、哈尔特征(haar-like features, Haar)^[4]等进行提取,接着将提取到的图像特征输入到如支持向量机(support vector machine, SVM)^[5]、K最近邻(K-Nearest-Neighbor, KNN)、决策树(Decision Tree)、迭代器(Ada-Boost)^[6]等的分类器中对候选目标进行分类。基于传统的目标检测方法有其固有缺陷:基于滑窗的区域选择算法对不同特征目标的针对性弱,时间复杂度高且窗口大量冗余;所提取到的图像特征本质上是一种手工设计的特征,针对不同的图像检测任务,特征好坏对检测性能有着重大直接的影响,需要研究人员针对所要解决的任务领域设计针对性的特征。此外,基于机器学习的目标检测方法从图像底层特征提取到表现特征提取的流程相对复杂,对人工特征的设计依赖性强。因此,传统基于机器学习的目标检测方法遇到瓶颈,需要一种更加高效精确的方法完成图像特征的提取。

与传统的机器学习相比,基于深度学习的目标检测算法通过多层神经网络直接对图像进行卷积和池化操作来提取图像的本质特征,在图像识别领域优势明显。

深度学习的发展^[7],特别是基于深度学习的卷积神经网络^[8](convolutional neural network, CNN)的出现,为图像特征提取带来全新的解决方法。CNN是局部连接网络,在CNN中,选取图像中小块区域作为底层输入数据,信息通过前向传播经过网络中各个层,每一层都由过滤器构成,以便能获得观测数据的一些显著特征,其层次间的紧密联系和空间信息使其特别适用于图像特征提取工作。

在实际工程应用中,部分场景对于模型精度要求非常高,部分场景对于模型检测速度要求非常高,但更多的是对速度和精度都有很高要求的情况。所以亟需一种能够在满足检测精度要求的同时达到实时检测的检测算法。

2 目标检测算法

2014年, Girshick等^[9]在CVPR(IEEE conference on computer vision and pattern recognition)大会上提出了R-CNN(region with CNN feature)网络, He K等^[10]在网络的全连接层前加了一个空间金字塔(spatial pyramid pooling)层,使得该网络在全连接层输入不需要归一化图像尺寸,提取的特征有更好的尺度不变性,降低了过拟合的可能,在VOC(visual object classes)2012数据集上,将平均准确率(mean average precision, mAP)提升到53.3%。2015年Girshick等^[11]和Ren等^[12]又相继提出了Fast R-CNN和Faster R-CNN模型,在提高模型检测精度的同时优化了检测速度。2016年Redmon等^[13]提出的YOLO(You Only Look Once)V1以45FPS(frames per second)真正达到了实时检测视频的速度,这为实时检测视频运动目标指明了方向;Liu等^[14]提出了SSD(Single Shot MultiBox Detector)物体检测模型,在实时性和准确率都高于YOLO V1,在VOC数据集测试, mAP达到了74.3%,检测速度也达到了46FPS;随后,Redmon等^[15]又提出了YOLO V2,在VOC数据集上测试,将mAP提升至78.6%,并且大幅提高了检测速度,达到了67FPS的实时检测效果。同时,基于区域建议的CNN目标检测^[16]和基于耦合CNN的弱监督学习^[17]方法也被提出来,使得标记成本大幅降低。

姚群力等^[18]系统综述了深度学习在目标检测中的研究进展,着重介绍了基于R-CNN算法的改进优化算法,对无区域建议的基于回归的YOLO算法和SSD算法等进行了简单介绍,提出了一些深度学习在目标检测领域遇到的难点问题以及研究趋势。肖雨晴等^[19]就非机动车、机动车和行人3个方面对目标检测在交通场景中的应用做了系统综述,介绍了几个目标检测算法的优化方向与经典算法。员娇娇等^[20]针对目标检测领域遇到的小目标检测问题做了系统综述,分别就多尺度、提高特征分辨率、上下文语义信息、数据增强和优化网络这五个方面的多种算法进行介绍,并对算法的特点和网络进行了对比研究。

基于区域建议的两步检测算法在精度上有很好的表现,但是在检测速度上无法达到实时检测的要求。基于回归的算法与基于区域建议的算法相比,在精度上做了让步,但是在检测速度上却有很大提高,完全可以达到实时检测的要求。YOLO算法相比SSD算法,在YOLO V3^[21]时期与SSD算法性能不相上下,但是随后推出的YOLO V4^[22]和YOLO V5^[23],已在性能上远超SSD算法。

3 对 YOLO V2 算法的改进算法

YOLO V2 算法的基础模型为 Darknet-19, 包含 19 个卷积层和 5 个最大值池化层。其中, 卷积层主要采用 3×3 卷积, 2×2 池化层, 可以将特征图中的通道数增加两倍。最后由全局平均池化层进行结果预测, 并在 3×3 卷积之间添加 1×1 卷积对特征图通道进行压缩, 以降低模型计算量和参数。YOLO V2 的 Darknet-19 模型相比 YOLO V1 模型, 主要在计算量上减少了约 33%, Darknet-19 结构如图 1 所示^[15]。

Type	Filters	Size/Stride	Output
Convolutional	32	3×3	224×224
Maxpool		$2 \times 2/2$	112×112
Convolutional	64	3×3	112×112
Maxpool		$2 \times 2/2$	56×56
Convolutional	128	3×3	56×56
Convolutional	64	1×1	56×56
Convolutional	128	3×3	56×56
Maxpool		$2 \times 2/2$	28×28
Convolutional	256	3×3	28×28
Convolutional	128	1×1	28×28
Convolutional	256	3×3	28×28
Maxpool		$2 \times 2/2$	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Maxpool		$2 \times 2/2$	7×7
Convolutional	1 024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1 024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1 024	3×3	7×7
Convolutional	1 000	1×1	7×7
Avgpool		Global	1 000
Softmax			

图 1 Darknet-19 网络结构

Fig. 1 Darknet-19 net structure

3.1 基于多尺度分辨率优化网络的小目标检测

海上船舶目标的检测问题在海事搜救、交通管理和无人船只等领域具有重大意义, 船舶目标具有目标小, 特征点少和易受环境影响等特点。周薇娜等^[24]针对海上船舶目标检测问题提出了一种改进的网络结构, 结合 SELU^[25] (scaled exponential linear units) 激活函数, 有效提高了 YOLO V2 算法的检测精度和 YOLO V3 算法训练过程的冗余问题, 但是在检测精度上还有待提高。

针对船舶目标检测任务, 于洋等^[26]基于 YOLO V2 网络, 使用 K-Means 算法对自制船舶数据集中的目标框重聚类, 优化船舶检测问题的 anchor 数据。对输入图像的尺寸进行调整, 使之每 10 epoch 改变一次图像的尺寸, 提高网络对不同尺度目标的检测能力。考虑到直通层可以将相邻特征

堆叠到不同通道, 即把高分辨率的特征映射成低分辨率的特征映射, 通过将特征细粒度化, 可以达到增强小目标定位问题的特性。于洋等在原 YOLO V2 的基础上又增加了一个直通层, 以增强对小目标物体的定位能力, 同时删减了网络中冗余的卷积层和池化层, 提升网络的检测速度。在与原 YOLO V2 做对比试验发现结果如表 1 所示^[26]。

表 1 改进算法与原算法结果

Table 1 Comparing the improved algorithm with the original one

模型	mAP	FPS
YOLO V2	76.8	67
New YOLO V2	79.1	64

尽管模型的检测速度在不影响实时性的基础上略有降低, 但是检测精度相比原 YOLO V2 算法有所提高。而且改进网络在船舶的小目标定位问题上精度更高, 针对同一张图片中出现不同分辨率大小的船舶时, 改进网络的检测效果更好。

3.2 基于改进 NMS 的复杂条件下目标检测

复杂气象条件下采集到的船舶目标信息具有多种复杂的噪声, 周薇娜等^[24]从新加坡海事数据库^[27] (singapore maritime dataset, SMD) 数据集中选取了传感器噪声、干扰物噪声、大尺度变化噪声和恶劣天气背景噪声 4 种极具代表性的噪声作为检测对象, 验证改进算法的有效性。试验结果表明改进算法针对复杂气象条件下的船舶目标检测有一定效果, 但是检测精度还有待提高。

针对同样复杂气象条件下的目标检测任务中采集到的图像数据不清晰等问题, 熊咏平等^[28]以 YOLO V2 网络模型为基础, 针对复杂气象条件下海上舰船目标的检测任务, 对非极大值抑制算法进行改进, 通过将暗通道先验去雾算法^[29]的基本物理特征和图像亮度值的差异相结合形成的显著性特征对图像是否清晰进行判别, 不清晰图像采用去雾算法处理后再进行目标检测任务。考虑到传统非极大值抑制算法存在只能对同类目标进行抑制, 非同类目标不抑制的问题, 以及试验所用数据集中检测对象分布稀疏的特点, 熊咏平等对所有检测到的类别进行统一 NMS 处理, 不再单独处理, 以提高模型检测精度。针对检测框与真实区域重叠度低的问题, 熊咏平等采用基于上下文感知的显著性算法对检测结果中显著性特征进行提取, 使用自适应阈值分割方法使得检测框大小与真实区域更为接近。在 2017 年中国计算机学会大数据与计算智能大赛使用的数据集上与 Faster R-CNN 和原 YOLO V2 算法目标检测结果如表 2 所示^[28]。

根据表 2 可以看出, 经过熊咏平等的改进之后, 模型的检测速度略微降低, 但检测精度上大幅提高。为复杂气象条件下遥感图像目标检测任务提出一种可行性方案。

表2 复杂气象条件下目标检测结果

Table 2 Comparison of target detection of performance under complex weather conditions

模型	mAP/%	FPS
YOLOV2	69.3	21.7
Faster R-CNN	72.6	2.5
YOLOV2 + 清晰化	73.1	21.7
Faster R-CNN + 清晰化	76.2	2.5
YOLOV2 + 清晰化 + 显著性	77.4	21.3
Faster R-CNN + 清晰化 + 显著性	79.3	2.0
Improved YOLOV2	84.5	19.8

3.3 基于位置关系的骑行者目标检测

YOLO 网络中的 anchor 参数是在 VOC 数据集上聚类得到的,识别目标种类多但是针对特种对象的目标检测效果不佳,在面对群簇小目标检测时即便加入了细粒度特征依旧存在检测效果不佳的问题^[30]。

胡超超等^[31]针对道路上自行车与骑行者群簇小目标的目标检测问题,在 YOLO V2 的基础上使用了细粒度特征和残差网络对网络进行改进,提出了 YOLO-R 网络模型。并使用聚类算法中类内相似度最大差异和类间相似度最大相似度最小的原则对聚类过程进行优化,提出了基于 IOU 的评价函数 F ^[31]。

$$F = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (1 - IOU(box, centroid_i))}{\sum_{i=1}^k n_i (1 - IOU(centroid_i, abox))}$$

式中, $\sum_{i=1}^k \sum_{j=1}^{n_i} (1 - IOU(box, centroid_i))$ 表示为类内方差;
 $\sum_{i=1}^k n_i (1 - IOU(centroid_i, abox))$ 为类间方差,矩形框 $abox$ 的长、宽为真实框长、宽的均值,评价函数 F 的值越小说明聚类的效果越好。

在区分行人与骑行者时,胡超超等通过比较行人框与自行车框的中心坐标距离与位置关系判断是否为骑行者。实现多目标跟踪则是根据 Kalman 滤波与匈牙利匹配算法相结合实现。在同一个行人与骑行者数据集上与原 YOLO V2 算法进行对比试验发现结果如表 3 所示^[31]。

表3 YOLO V2 与 YOLO-R 对比试验结果

Table 3 Test results of YOLO V2 and YOLO-R

模型	mAP	FPS
YOLO V2	76.9	22.2
YOLO-R	80.3	25.0

根据表 3 不难看出,改进后的 YOLO-R 算法不管是检测精度还是检测速度上都相比较原算法有了很大提高,而且为

多物体综合检测问题提出了一种基于位置关系的可行性方案。

4 基于 YOLO V3 的改进算法

YOLO V3 算法相比 YOLO V2,主要有分类层、多尺度融合和骨干网络架构的改进。这 3 个方面的改进使得 YOLO V3 算法相比较 YOLO V2 算法在不影响速度的情况下,大幅度提高了检测精度,同时在不提高输入图片分辨率的情况下大大提高了对于小目标物体的检测性能。

YOLO V2 的网络结构在分类时采用的是单标签多分类的 Softmax 层,而 YOLO V3 则改进为适用于多标签多分类的 logistic 层。主要是针对实际生后中复杂场景下,同一张图片中有多个类别物体重叠在一起,影响检测网络的性能。同时因为 Softmax 层可由多个 logistic 分类器取代且不会对模型分类准确率产生影响。

多尺度方面的改进则是主要借鉴了 SSD 算法,采用 Multi-Scales 策略,通过 3 个不同尺度的特征层进行预测,大大提高了对尺度变换特征明显的图片中小目标对象的识别,提高了在遥感目标检测领域的适用范围。

骨干网络的改进主要有加深网络层,使用残差网络模块,卷积代替池化,使用金字塔池化。网络层数的加深可以大大提高模型的检测精度;残差模块的加入可以防止因网络层数过深引起的梯度下降;步长为 2 的卷积核代替原本的池化操作可以防止因池化导致的信息丢失;金字塔池化的使用则可以实现模型的多尺寸输入和相同尺寸的输出。YOLO V3 的骨干网络 Darknet-53 结构如图 2 所示^[21]。

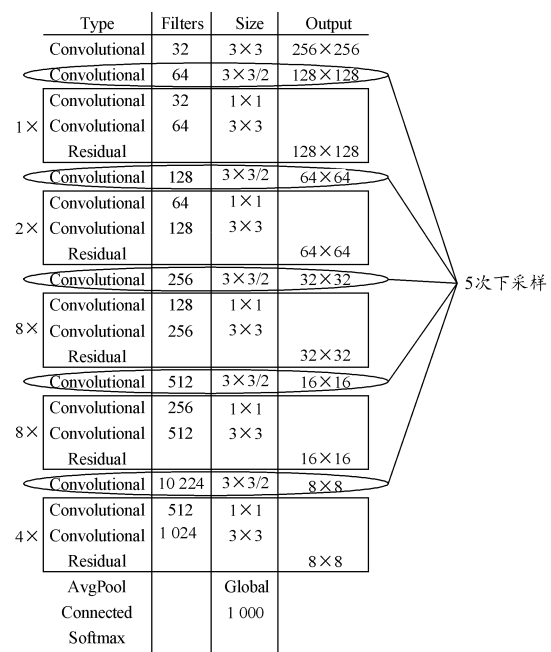


图2 Darknet-53 网络结构示意图

Fig. 2 Darknet-53 net structure

4.1 基于改进目标函数的安全帽检测

YOLO V3 算法中目标函数是基于误差平方和建立,存在目标函数和评价函数不统一的问题,容易出现目标局部最优不一定是 IOU 局部最优的问题,Rezatofighi 等^[32]提出了 GIOU 指标,为解决该问题提供了新的可行性方案。但是针对在模型训练开始时出现真实框与预测框不相交的问题,原 YOLO V3 算法进行了优化,随着训练过程的进行,这种情况的次数会不断降低,而 Rezatofighi 等提出的 GIOU-YOLO V3 并未考虑这种情况,直接对目标函数进行了替换,这会对模型的训练过程造成一定影响。

王兵等^[33]基于 YOLO V3 网络,融合 IOU 与 GIOU 的优点,提出一种将包含真实框和预测框的最小框减去两框相交区域的目标函数,同时修改模型的 IOU 计算方法,降低预测框中负样本对训练过程的影响。在 VOC2007 和自制施工现场安全帽佩戴检测数据集中试验结果如表 4 所示^[33]。

表 4 改进算法结果

Table 4 Performance comparison of improved algorithms

模型	mAP	
	VOC2007	安全帽数据集
改进 YOLO V3	82.29	94.06
GIOU-YOLO V3	81.15	93.92
YOLO V3	80.22	92.01
SSD	73.58	83.65

根据表 4 可知,改进后的 YOLO V3 算法相比较 GIOU-YOLO V3、原 YOLO V3 算法和 SSD 算法在检测精度上最佳。王兵等自己采集制作的安全帽数据集对安全帽目标的标注在精度上很高,针对真实情况下安全帽检测任务有很好的效果。

4.2 基于背景过滤的目标检测

YOLO V3 算法属于单结构回归密集型检测,每个特征图上都预设有一定数量的锚框,每个锚框设定一个训练样本,根据单结构网络回归对预测锚框所属类别和相对于真实框的偏移量进行预测。但是锚框与真实框匹配后,数据集中属于背景的负样本数量就会远大于属于检测对象的正样本数量,导致正负样本比例不平衡。针对这种情况,基于回归的 SSD 算法在训练过程中选择只取一些损失比较大的负样本加入训练,但是对于 YOLO V3 算法,Lin 等^[34]为了控制正负样本在损失函数中的权重使用改进的交叉熵损失函数(Focal Loss)代替原损失函数以减少数据集中负样本所占权重,但是试验发现模型的精度反而出现下降的情况。

赵琼等^[35]基于 YOLO V3 网络,在原 YOLO V3 网络中增加卷积模块对聚类结果锚框进行粗调,之后使用原预测网络进行细调。为了解决正负样本不平衡问题,在 YOLO V3 特

征融合层后增加卷积模块对样本中的背景样本进行分类,设置阈值生成 mask 标签,对预测为负样本且高于阈值的区域,mask 值设为 0,其余为 1。然后将其映射到预测层上,mask 值为 0 的区域样本不参与训练和预测,解决了正负样本不平衡问题。同时对损失函数进行了部分改进,在其基础上添加了针对预测锚框的宽高损失和针对前景背景的分类损失两部分。在 VOC 数据集上与原算法试验结果如表 5 所示^[35]。

表 5 算法结果

Table 5 Algorithmic performance comparison

模型	mAP
YOLO V3	79.4
改进 YOLO V3	80.2

可见,改进后的算法相较原算法在精度上略有提高,并未出现精度下降问题。

4.3 基于修改特征图分辨率的遥感建筑物检测

在实际工程应用中针对遥感图像建筑物的目标检测很难达到实时检测的任务要求,而且遥感图像中目标建筑物小也对检测任务造成很大困难。张曼等^[36]基于 YOLO V2 目标检测算法,对算法的特征维度和卷积过程进行改进,略微提升了检测精度,但是在检测速度上有所降低。同时提供了一种大幅遥感图像的通用处理框架,有效解决了遥感图像分辨率过大导致检测网络无法识别的问题。

董彪等^[37]基于 YOLO V3 网络,对算法的特征提取网络方面进行了改进,提高了模型的性能。考虑到原 YOLO V3 算法中采取的 3 个尺度(13×13 、 26×26 、 52×52)融合的方法对遥感图像中小目标检测的不利影响,董彪等将其修改为 26×26 和 52×52 两个较大尺度的分辨率,并对特征图进行修改。将原网络的 26×26 特征图作为第一尺度,将 61 层结果进行 5 次卷积操作后使用 1×1 卷积操作降维,通过上采样与 36 层融合,结果用 3×3 卷积核消除混叠得到 52×52 的特征图作为第二尺度特征。考虑到基于原 YOLO V3 算法中的聚类结果是在 COCO 数据上得到的,对于特种目标检测任务效果较差,于是使用 K-Means 算法对遥感建筑物数据集进行重聚类。与原 YOLO V3 和 SSD 算法在自制遥感建筑物数据集上试验结果如表 6 所示^[37]。

表 6 算法结果

Table 6 Performance comparison

模型	mAP
YOLO V3	85.94
SSD	80.49
改进 YOLO V3	91.29

改进后的 YOLO V3 算法相比原 YOLO V3 算法和 SSD 算法在精度上有大幅度提高。

4.4 基于多尺度预测的小目标预测方法

目前目标检测算法中,利用高层特征的语义信息和底层特征的细节信息最多也是比较好的方式是 FPN 的架构方式。YOLO V3 和 RetinaNet^[38]采用的是 concatenation 和 element-wise 的直接衔接或者相加的架构方式,这种方式对 FPN 的多层不同尺度的特征的利用不够完善,而且多层特征尺度与待检测目标之间存在不一致的问题,比如大特征图识别小目标,小特征图识别大目标。同一个目标在不同特征层可能会出现分类完全相反的情况,这样的矛盾信息对于目标检测任务很不利。所以 SongTao Liu 等^[39]基于 FPN 架构,通过更新权重参数的方式将 FPN 的多层特征图融合在一起得到新的特征图进行最终预测,提出了一种基于自适应空间特征的融合方法。将 ASFF 这种架构方式应用到 YOLO V3 中验证 ASFF 的有效性,结果如表 7 所示^[38]。

根据表 7 可知,改进后的算法在不影响实时性的情况下在检测精度上有大幅提高。

表 7 ASFF 算法结果

Table 7 ASFF algorithm comparison

YOLOV3	mAP	FPS
Baseline	42.9	50
Baseline + Concat	43.8	42
Baseline + Sum	43.5	48
Baseline + ASFF	45.8	46

4.5 基于改进聚类算法的小目标检测

YOLO V3 算法中使用 K-Means 算法作为聚类算法,该算法对噪声点和孤立点敏感,容易造成聚类结果出现偏差,进而影响模型训练效果。岳晓新等^[40]针对该聚类算法存在的问题,提出一种基于融合 DBSCAN^[41](density-based spatial clustering of applications with noise)算法和 K-Means 算法的聚类算法,DBSCAN 算法可以减弱噪声点和孤立点数据对试验的影响,为 K-Means 算法提供准确的 K 值或 Anchor Box 个数,提高 K-Means 算法的聚类性能,达到优化小目标检测的 Anchor Box 的目的。岳晓新等还针对 Sigmoid 激活函数在输出较大时可能存在梯度消失,造成网络收敛速度变慢的情况,使用 Focal Loss 损失函数替代原 YOLO V3 算法中的损失函数。对 KITTI 数据集中行人类别识别进行消融试验结果如表 8 所示^[40]。

根据表 8 可知,改进聚类算法和优化损失函数都可以提高网络的检测精度,同时使用 2 种优化方法得到的试验结果最佳。

表 8 算法结果

Table 8 Algorithm performance comparison

YOLOV3	mAP	FPS
baseline	90.07	42.17
baseline + DBSCAN	91.23	44.31
baseline + Focal Loss	91.29	43.08
baseline + DBSCAN + Focal Loss	92.43	44.52

4.6 基于多层特征图的轨道扣件缺陷检测

韦若禹等^[42]发现目前轨道线路养护工作自动化程度不高,大多依靠人工巡检的方式进行,准确率低而且效率不高。基于此问题,韦若禹等用高速工业相机采集了某地铁线路轨道的 DTVI2 型扣件的图像数据制作了轨道扣件缺陷检测专用数据集,并通过数据增强技术对缺陷样本数据进行扩展,解决了缺陷样本稀缺的问题。然后对 YOLO V3 网络的聚类算法和特征选取方面进行了改进,减少了算法的运算量,提高了检测速度,改进后的 YOLO V3 网络结构如图 3 所示^[42]。

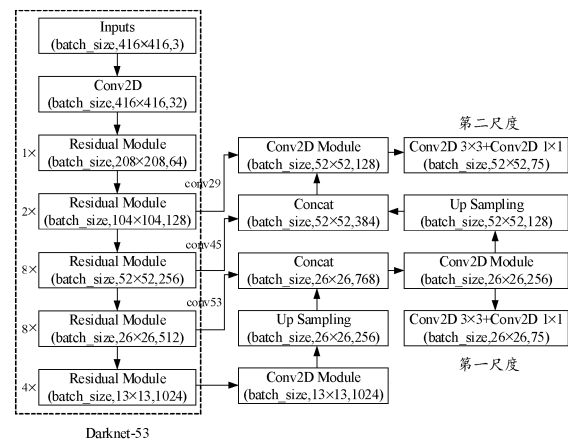


图 3 改进 YOLO V3 网络结构框图

Fig. 3 Improved YOLO V3 net structure

与 Faster R-CNN、SSD 和原 YOLO V3 算法试验结果如表 9 所示^[42]。

表 9 改进 YOLO V3 算法试验结果

Table 9 Comparing test results of improved YOLO V3 algorithm

模型	mAP	FPS
Faster R-CNN	80.61	8.7
SSD	69.73	26.5
YOLOV3	74.98	32.0
Improved YOLOV3	95.62	58.4

根据表 9 可知,改进后的 YOLO V3 算法在检测精度和检测速度上均有大幅度提高。韦若禹等还针对轨道扣件检

测任务制作了专用的数据集,大大提高了该领域目标检测任务的精度。

5 基于 YOLO V4 的改进算法

YOLO V4 模型的输入端借鉴了 Sangdoo Yun 等^[43]提出的 CutMix 数据增强方式,提出了 Mosaic 数据增强方法。Alexey Bochkovskiy 等^[22]提出的 Mosaic 数据增强方式在 CutMix 的使用两张图片进行拼接的基础上,使用了四张图片以随机的缩放、裁切和排列等方式进行拼接,大大增强了模型针对大、中、小样本目标不均衡问题的鲁棒性。而且 Mosaic 数据增强提高了单 GPU 训练的效果,降低了训练成本。

YOLO V4 主网络由 YOLO V3 的 Darknet-53 升级为 CSP-Darknet-53,借鉴了 Chien-Yao Wang 等^[44]提出的 CSPNet, CSPNet 主要解决了模型卷积过程中出现的重复问题,减少了模型的参数和计算量。激活函数也由 ReLU 替换为 Mish, Mish 相比 ReLU 具有更加平滑的梯度,可以更好地保持模型准确度。

特征提取则采用了 SPP 和 PANet 两大改进,分别利用 4 个不同尺度的最大池化层对特征进行处理,最大化增大感受野面积,使图像上下文特征更加显著。YOLO V4 的整体网络结构如图 4 所示。

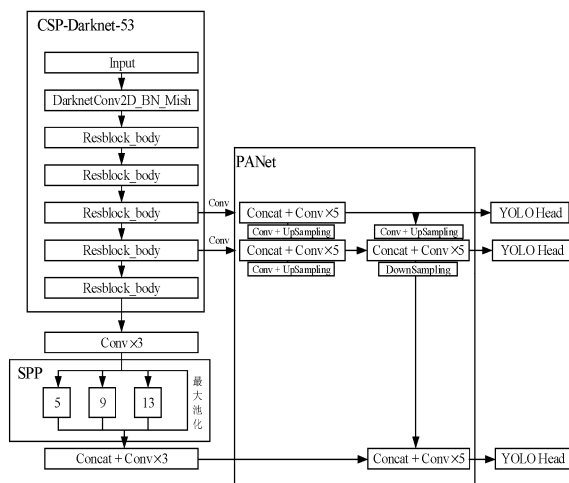


图 4 YOLO V4 整体网络结构框图

Fig. 4 YOLO V4 overall structure

5.1 基于区域位置的安全帽佩戴检测

Li Kang 等^[45]提出一种基于 YOLO V4 算法的安全帽佩戴检测方法,通过对运动对象目标进行图像分割定位人所在区域,根据人身体比例的五分之一判断头部所在位置。Aziz, Kheireddine 等^[46]根据图像处理,筛选出工人肤色区域,通过设定骨架点对头部区域与身体夹角进行计算,识别出头部所在区域。上述 2 种方法只能检测直立状态下的施工人员,而且基于传统机器学习方法的目标检测算法在鲁棒性和适用

性上不及基于深度学习的目标检测算法。

针对实际工程应用中施工人员在复杂姿态下的是否佩戴安全帽的目标检测问题,王雨生等^[47]基于 YOLO V4 网络,提出一种融合肤色特征提取和区域位置判断的安全帽佩戴检测方法。通过图像色彩变换、形态学处理筛选出疑似头部区域,与算法头部区域检测结果进行交叉验证确定头部区域。因为疑似头部区域的判断是通过肤色区域面积判定,考虑到存在工人手持安全帽的情况,王雨生等根据肤色区域空间位置判断头部所在区域,通过头部区域和安全帽中心点坐标的距离计算,判断工人是否安全佩戴安全帽,为安全帽佩戴检测提出一种基于位置的可行性方案。通过 Li Kang 和 Aziz Kheireddine 的安全帽检测,结果如表 10 所示^[47]。

表 10 算法结果

Table 10 Performance comparison

模型	Precision/%
Li Kang	82.9
Aziz Kheireddine	84.6
M-YOLO V4	94.3

由表 10 可知,王雨生等的改进算法在安全帽佩戴检测上的精度远超 Li Kang 和 Aziz Kheireddine 的检测算法。而且 Li Kang 和 Aziz Kheireddine 的检测算法在面对施工人员复杂姿态下安全帽佩戴问题上存在严重的漏检情况。

5.2 基于多尺度的遥感图像飞机检测

小目标检测是遥感图像目标检测中的难点问题,针对小目标检测问题,目前常用的解决方法是采用多尺度特征融合的方法,提高对小目标对象检测的语义特征,进而提高小目标对象的检测精度。王银等^[48]以 YOLO V4 网络为基础,融合 PANet-D^[49]特征网络提取到的 4 个尺度特征图,获取小目标的语义特征,增强对小目标对象的检测能力。

徐佰祺等^[50]基于 YOLO V4 网络模型提出一种基于多尺度特征和融合注意力机制的 M-YOLO V4 模型。考虑到图像特征金字塔底层特征图包含大量目标的特征信息,但是感受野小、缺乏目标语义信息的问题。徐佰祺等将 104×104 尺度的特征图通过下采样与 52×52 的特征图融合,对底层特征图的语义信息进行补充,将融合后的结果经过 7 次卷积后送入到注意力机制模块中达到更好的融合效果。考虑过多的背景信息会降低模型运行速度,徐佰祺等在模型中添加了注意力机制。SSD、YOLO V3、YOLO V4 算法在 RSOD-Dataset 数据集下的试验结果如表 11 所示^[50]。

根据表 11 可知,改进后的 M-YOLO V4 算法在准确率和检测速度上均超越同类基于回归的检测算法。

表 11 算法结果

Table 11 Algorithm performance comparison

模型	mAP	FPS
SSD	77.80	7.3
YOLO V3	79.52	7.9
YOLO V4	82.32	9.0
M-YOLO V4	83.95	12.5

5.3 基于多尺度的红外弱目标检测

通常,把信噪比小于 3 的目标视为弱目标,空间领域的红外弱目标检测技术是重要的国防科技项目,目前国内针对弱目标的检测任务不多。因为弱目标信噪比低,导致目标纹理特征和结构形状模糊,给目标检测任务带来很大难度。相比可见光状态下的目标检测任务,红外图像的纹理单一、分辨率低,特征提取困难,而且红外图像下的目标和周围背景对比度低,几何轮廓模糊,难以提取有效的边缘信息。王国刚等^[51]针对红外弱小目标在高层感受野信息丢失问题提出了一种基于多通道多尺度特征融合的红外弱小目标检测算法(J-MSF)。

刘杨帆等^[52]基于 YOLO V4 网络,对原算法进行重聚类,使其适应红外弱目标检测任务。考虑到弱目标尺寸过小,特征金字塔中底层特征包含少量语义信息,上层特征包含特征信息过少,所以对 52×52 尺度特征进行上采样与 104×104 特征融合经过 5 次卷积后输出,即扩展原算法的 3 个尺度为 4 个尺度(13×13 、 26×26 、 52×52 、 104×104),提出一种基于多尺度融合的 MY-YOLO V4 模型。YOLO V3 与 YOLO V4 在 COCO 数据集和自制红外数据集上试验结果如表 12 所示^[52]。

表 12 算法结果

Table 12 Performance comparison

模型	数据集	mAP	FPS
YOLO V3	COCO 数据集	82.25	6.83
	自制红外数据集	84.55	6.89
YOLO V4	COCO 数据集	88.25	28.12
	自制红外数据集	90.59	30.46
MY-YOLO V4	COCO 数据集	89.13	24.07
	自制红外数据集	93.25	25.65

由表 12 可知,改进后的算法在满足实时性的前提下,准确度相比原算法有所提高。

6 目标检测数据集

通常,为了提高模型的性能,一般选择调整模型的网络

结构、参数或者修改替换模型的某些计算公式,但是一个好的数据集也会对模型的性能产生一定影响。一个好的数据集中图像噪声不会太大,不存在冗余数据或者脏数据,对于训练结束后分析模型的收敛曲线等参数有很大帮助。下面介绍一些目标检测常用的数据集。

Pascal VOC^[53]:PASCAL VOC 挑战赛(The PASCAL Visual Object Classes)专用数据集,属于公开数据集,全称 Pattern Analysis, Statical Modeling and Computational Learning。自 2005 年起举办,2012 年为最后一届,每年举办一次,最初只有 4 个类别,到 2007 年增至 20 个类别,目前使用最广的是 VOC2007 和 VOC2012。

COCO^[54]:由微软团队发布,常用于目标检测、图像分割、关键点识别等任务,包含大量的小目标和日常场景中的常见物体类别,并提供像素级的实例标注,共计 91 类目标,328 000 张图像和 250 万个标注框。

Open Images^[55]:由谷歌团队发布,常用于目标检测、语义分割等任务,包含 600 个类别,190 万张标注图片,平均每张图片有 8 个标签,标签框由专业注释人员手动标注,确保标签的一致性和准确度,是现有的包含位置标注信息的最大数据集。

ImageNet^[56]:目前世界上最大的图像识别数据库。由专门的团队维护,常用于目标检测、分类和场景识别等方面,包含 2 万多个类别,约 1 420 万张图像。其中超过百万图像有精确类别和位置标注信息,并有 200 个类别可用于目标检测任务。

DOTA^[57]:常用于遥感图像目标检测,包含 15 个类别,2 806 张航空遥感图像,188 282 张标注框,图像尺寸为 $4k \times 4k$ 。标注格式为四点确定的任意形状和方向的四边形。特点是图像的尺度变化性大,包含大量小目标。

NWPU-VHR-10^[58]:由西北工业大学发布,常用于遥感目标检测等方面,包括飞机、轮船、车辆等 10 个类别,800 张高分辨率卫星拍摄图像,图像取自 Google Earth 和 Vaihingen,专家手动标注,背景图 150 张。

UCAS-AOD^[59]:由中国科学院大学发布,常用于遥感图像的目标检测等方面,只包含汽车、飞机 2 类目标以及背景负样本。共有 2 个版本,其中包含 1 000 张飞机图像,7 482 个飞机样本,以及 510 张汽车图像,7 114 个汽车样本,和 910 张背景负样本图像。

RSOD-Dataset^[60]:由武汉大学发布,常用于航空遥感等方面,包含飞机、操场、立交桥、油桶 4 类目标。有 976 张图像,6 950 个标签,飞机类别最多。

Tsinghua-Tencent100K^[61]:由清华大学和腾讯公司合作发布,常用于交通标志的目标检测任务,包含 3 万个交通标志类别,10 万张交通标志的图像。

URPC 2018^[62]:水下图像的数据集,常用于小目标检测

和重叠目标检测,包含海参、海胆、海星和扇贝 4 个类别,2901 张训练图像和 800 张测试图像。

Tiny Person^[62]:由中国科学院大学发布,常用于小目标检测等方面。该数据集在海上和海滩场景图像中人分辨率很小,包含 sea person 和 earth person 两个类别。在船上、躺在水里或者身体的一半以上在水里的为 sea person,其他的是 earth person。图像中人的识别有密集人群和水中倒影的影响,为长距离和大量背景的微小物体检测开辟了一个新的方向,包含 794 张训练图像和 816 张测试图像。

7 研究分析

YOLO 目标检测算法是目标检测领域不可替代的经典算法,但是随着目标检测算法的发展,YOLO 目标检测算法不可避免地有一些方面落后于其他算法,对此提出一些未来的研究趋势:

数学理论尚待完善进步。只有充分的数学理论支撑,才能研究出更加先进,更为优化的 YOLO 算法模型,同时因为缺乏一定的数学理论支撑,导致 YOLO 目标检测算法框架模型无法进一步提高性能。数学理论的发展可以为 YOLO 算法模型的改进提供原理性的指导,避免在改进模型时仅能依靠过往的经验和虚无缥缈的直觉。

无监督式的学习。因为随着社会的发展,人力成本越发昂贵,导致数据集的标注工作成为一个极度浪费人力、财力和时间的项目,而且数据的标注工作非常枯燥,标注过程中易出现误差或错误。神经网络模型是以人为模板进行研究的,而人的学习过程以无监督式学习为主导的,所以研究出融合无监督式学习的 YOLO 算法模型,对降低标注成本,丰富目标检测数据集有很好的效果。

融合图像细节信息和语义信息的结构。对于小目标检测而言,仅细节特征信息很难将检测精度提高到极限,但是加上图像语义信息,比如上下文关联信息和场景信息,结合细节信息构建出新的特征图,那么检测精度必定会有极大提高。目前已经有很多学者针对改进 YOLO 算法提取特征的过程进行改进,融合上下层语义信息和细节信息来提高 YOLO 算法针对小目标问题的检测精度。但是,目前关于这个方向的改进并没有很大的提高,亟需一种更好的特征融合提取结构来提高 YOLO 算法的性能。

提高复杂气象条件下 YOLO 目标检测算法的性能。随着时代的发展,在极端条件下与复杂气象条件下的人员搜救与目标检测逐渐成为一个新的研究方向。极端条件下的风沙遮挡,灰尘覆盖,海上飓风以及复杂气象条件雨雾,阴云和雷暴等的种种不利条件,都会严重影响目标检测算法的性能。这些不利条件会导致数据采集过程中相机的记录失真或难以记录有效场景信息。YOLO 算法还没有很好的改进

措施来解决复杂气象条件下的目标检测任务。所以,针对复杂气象条件下的改进 YOLO 算法需要进一步创新发展。

更加先进的融合目标检测算法。比如说针对夜间的目标检测任务使用可见光与红外光融合的多模态检测,以提高模型检测精度;针对特定对象的目标检测任务使用融合注意力机制,可以减少计算量,提高检测速度。目前通过多类别传感器协同作用完成目标检测任务的例子很少,仿生的算法未来尚还需要努力。

8 结论

目标检测是计算机视觉热点问题,也是未来人工智能发展不可或缺的一部分。从传统基于机器学习的目标检测算法人工设计特征到基于深度学习的卷积神经网络自动选择特征,对于目标检测算法的要求始终都是为了提高检测算法的性能,提高精度、速度以及更加轻量化。基于区域的目标检测算法尽管精度很高,但是检测速度无法达到实时性要求,很难适应实际的社会需求,基于回归的目标检测算法则在满足实时性要求的前提下拥有足够的精度,但是未来的改进依旧是更高的精度以及实时性要求,YOLO 算法自其提出以来不断,融合各种不同网络结构,不断提高检测性能,相信它未来的发展更好。

参考文献:

- [1] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection; 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) [C]//San Diego, CA, USA, 2005.
- [2] LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints [J]. International Journal of Computer Vision, 2004, 60(02): 91-110.
- [3] OJALA T, PIETIKAINEN M, MAENPAA T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(07): 971-987.
- [4] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features [C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Kauai, HI, USA, 2001.
- [5] HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines [J]. IEEE Intelligent Systems and their Applications, 1998, 13(04): 18-28.
- [6] LIENHART R, MAYDT J. An extended set of Haar-like

- features for rapid object detection[C]//Proceedings International Conference on Image Processing. Rochester, NY, USA, 2002.
- [7] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436–444.
- [8] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, OH, USA, 2014.
- [9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA, 2014.
- [10] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(09): 1904–1916.
- [11] R. G. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile, 2015.
- [12] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(06): 1137–1149.
- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA, 2016.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[J]. arXiv e-prints, 2015: 1512–2325.
- [15] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA, 2017.
- [16] 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, 2016, 38(01): 142–158.
- [17] ZHANG F, DU B, ZHANG L, et al. Weakly Supervised Learning Based on Coupled Convolutional Neural Networks for Aircraft Detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(09): 5553–5563.
- [18] 姚群力, 胡显, 雷宏. 深度卷积神经网络在目标检测中的研究进展[J]. 计算机工程与应用, 2018, 54(17): 1–9.
- Yao Q L, Hu X, Lei H. Application of deep convolutional neural network in object detection[J]. Computer Engineering and Applications, 2018, 54(17): 1–9.
- [19] 肖雨晴, 杨慧敏. 目标检测算法在交通场景中应用综述[J]. 计算机工程与应用, 2021, 57(06): 30–41.
- Xiao Y Q, Yang H M. Research on Application of Object Detection Algorithm in Traffic Scene[J]. Computer Engineering and Applications, 2021, 57(06): 30–41.
- [20] 员娇娇, 胡永利, 孙艳丰, 等. 基于深度学习的小目标检测方法综述[J]. 北京工业大学学报, 2021, 47(03): 293–302.
- Yuan J J, Hu Y L, Sun Y F, et al. Survey of Small Object Detection Methods Based on Deep Learning[J]. Journal of Beijing University of Technology, 2021, 47(03): 293–302.
- [21] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018: 1804–2767.
- [22] BOCHKOVSKIY A, WANG C, LIAO H M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. arXiv e-prints, 2020: 2004–10934.
- [23] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios[J]. arXiv e-prints, 2021: 2108–11539.
- [24] 周薇娜, 丁豪文, 周颖. 一种海上弱小运动船舶实时检测方法[J]. 合肥工业大学学报(自然科学版), 2021, 44(09): 1187–1192.
- Zhou W N, Ding H W, Zhou Y. A Real-Time Detection Method for Dim and Small Moving Ships at Sea[J]. Journal of Hefei University of Technology (Natural Science), 2021, 44(09): 1187–1192.
- [25] KLAMBAUER G, UNTERTHINER T, MAYR A, et al. Self-Normalizing Neural Networks[J]. arXiv e-prints, 2017: 1706–2515.
- [26] 于洋, 李世杰, 陈亮, 等. 基于改进 YOLOv2 的船舶目标检测方法[J]. 计算机科学, 2019, 46(08): 332–336.
- Yu Y, Li S J, Chen L, et al. Ship Target Detection Based on Improved YOLOv2[J]. Computer Science, 2019, 46(08): 332–336.
- [27] D. K P, D. R, L. R, et al. Video Processing From Electro-Optical Sensors for Object Detection and Tracking in a Maritime Environment: A Survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(08): 1993–2016.
- [28] 熊咏平, 丁胜, 邓春华, 等. 基于深度学习的复杂气象条件下海上船只检测[J]. 计算机应用, 2018, 38(12): 3631–3637.
- Xiong Y P, Ding S, Deng C H, et al. Ship detection under

- complex sea and weather conditions based on deep learning [J]. Journal of Computer Applications, 2018, 38(12): 3631-3637.
- [29] HE K, SUN J, TANG X. Single Image Haze Removal Using Dark Channel Prior[J]. IEEE Trans Pattern Anal Mach Intell, 2011, 33(12): 2341-2353.
- [30] WANG L, LI W, ZHANG Y, et al. Pedestrian detection based on YOLOv2 with skip structure in underground coal mine [C]//2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC). Chongqing, 2017.
- [31] 胡超超, 刘军, 张凯, 等. 基于深度学习的行人和骑行者目标检测及跟踪算法研究[J]. 汽车技术, 2019(07): 19-23.
- Hu C C, Liu J, Zhang K et al. Research on Target Detection and Tracking of Pedestrian and Cyclist Based on Deep Learning[J]. Automobile Technology, 2019(07): 19-23.
- [32] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA, 2019.
- [33] 王兵, 李文璟, 唐欢. 改进 YOLOv3 算法及其在安全帽检测中的应用[J]. 计算机工程与应用, 2020, 56(09): 33-40.
- Wang B, Li W J, Tang H. Improved YOLOv3 Algorithm and Its Application in Helmet Detection [J]. Computer Engineering and Applications, 2020, 56(09): 33-40.
- [34] T. L. P. G., R. G., et al. Focal Loss for Dense Object Detection [C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy, 2017.
- [35] 赵琼, 李宝清, 李唐薇. 基于改进 YOLOv3 的目标检测算法[J]. 激光与光电子学进展, 2020, 57(12): 313-321.
- Zhao Q, Li B Q, Li T W. Target Detection Algorithm Based on Improved YOLO v3 [J]. Laser & Optoelectronics Progress, 2020, 57(12): 313-321.
- [36] 张曼, 李杰, 丁荣莉, 等. 基于改进 YOLO-V2 算法的遥感图像目标检测技术研究[J]. 计算机科学, 2020, 47(S1): 176-180.
- Zhang M, Li J, Ding R L. et al. Remote Sensing Image Object Detection Technology Based on Improved YOLO-V2 Algorithm [J]. Computer Science, 2020, 47(S1): 176-180.
- [37] 董彪, 熊风光, 韩变, 等. 基于改进 Yolo v3 算法的遥感建筑物检测研究[J]. 计算机工程与应用, 2020, 56(18): 209-213.
- Dong B, Xiong F G, Han X et al. Research on Remote Sensing Building Detection Based on Improved Yolo v3 Algorithm [J]. Computer Engineering and Applications, 2020, 56(18): 209-213.
- [38] LIN T, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [39] LIU S, HUANG D, WANG Y. Learning Spatial Fusion for Single-Shot Object Detection [J]. arXiv e-prints, 2019: 1911-9516.
- [40] 岳晓新, 贾君霞, 陈喜东, 等. 改进 YOLOv3 的道路小目标检测 [J]. 计算机工程与应用, 2020, 56(21): 218-223.
- Yue X X, Jia J X, Chen X D et al. Road Small Target Detection Algorithm Based on Improved YOLOv3 [J]. Computer Engineering and Applications, 2020, 56(21): 218-223.
- [41] ESTER M, KRIEGLER H, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise [C]//Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. Portland, Oregon, 1996.
- [42] 韦若禹, 李舒婷, 吴松荣, 等. 基于改进 YOLOv3 算法的轨道扣件缺陷检测 [J]. 铁道标准设计, 2020, 64(12): 30-36.
- Wei R Y, Li S T, Wu S R. Defect Detection of Track Fastener Based on Improved YOLOv3 Algorithm [J]. Railway Standard Design, 2020, 64(12): 30-36.
- [43] SANGDOO Y, DONGYOON H, SANGHYUK C, et al. Cut-Mix: Regularization Strategy to Train Strong Classifiers With Localizable Features [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South), 2019.
- [44] CHIEN-YAO W, HONG-YUAN M L, YUEH-HUA W, et al. CSPNet: A New Backbone that can Enhance Learning Capability of CNN [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA, 2020.
- [45] LI K, ZHAO X, BIAN J, et al. Automatic Safety Helmet Wearing Detection [J]. arXiv e-prints, 2018: 1264-1802.
- [46] AZIZ K, MERAD D, IGUERNAISSI R, et al. Head detection based on skeleton graph method for counting people in crowded environments [J]. Journal of Electronic Imaging, 2016, 25(1): 13012.
- [47] 王雨生, 顾玉宛, 庄丽华, 等. 复杂姿态下的安全帽佩戴检测方法研究 [J]. 计算机工程与应用, 2020: 1-11.

- Wang Y S, Gu Y W, Zhuang L H. Research on Detection Method of Helmet Wearing in Complex Posture [J]. Computer Engineering and Applications, 2020; 1 – 11.
- [48] 王银, 王飞翔, 孙前来. 多尺度特征融合车辆检测方法[J]. 系统仿真学报, 2021; 1 – 12.
- Wang Y, Wang F X, Sun Q L. Vehicle Detection Method Based on Multi Scale Feature Fusion [J]. Journal of System Simulation, 2021; 1 – 12.
- [49] LIU S, QI L, QIN H, et al. Path Aggregation Network for Instance Segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018.
- [50] 徐佰祺, 江刚武, 刘建辉, 等. 联合多尺度特征和注意力机制的遥感影像飞机目标检测[J]. 测绘科学技术学报, 2020, 37(04): 398 – 403.
- Xu B Q, Jiang G W, Liu J H et al. Aircraft Target Detection in Remote Sensing Image Based on Multi-Scale and Attention Mechanism [J]. Journal of Geomatics Science and Technology, 2020, 37(04): 398 – 403.
- [51] 王国刚, 孙召进, 刘云鹏. J-MSF: 一种新的多通道多尺度红外弱小目标检测算法[J]. 红外与激光工程, 2021; 1 – 10.
- Wang G G, Sun Z J, Liu Y P. J-MSF: A New Infrared Dim Target Detection Algorithm Based on Multi-Channel and Multi-Scale [J]. Infrared and Laser Engineering, 2021; 1 – 10.
- [52] 刘杨帆, 曹立华, 李宁, 等. 基于 YOLOv4 的空间红外弱小目标检测[J]. 液晶与显示, 2021, 36(04): 615 – 623.
- Liu Y F, Cao L H, Li N et al. Detection of space infrared weak target based on YOLOv4 [J]. Chinese Journal of Liquid Crystals and Displays, 2021, 36(04): 615 – 623.
- [53] SHETTY S. Application of Convolutional Neural Network for Image Classification on Pascal VOC Challenge 2012 dataset [J]. arXiv e-prints, 2016; 1607 – 3785.
- [54] LIN T, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common Objects in Context [C]//ECCV 2014. Zurich, Switzerland, 2014. Cham, Springer.
- [55] KUZNETSOVA A, ROM H, ALLDRIN N, et al. The Open Images Dataset V4 [J]. International Journal of Computer Vision, 2020, 128(07): 1956 – 1981.
- [56] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge [J]. International Journal of Computer Vision, 2015, 115(03): 211 – 252.
- [57] XIA G, BAI X, DING J, et al. DOTA: A Large-scale Dataset for Object Detection in Aerial Images [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, 2018.
- [58] CHENG G, HAN J, ZHOU P, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98 (dec.): 119 – 132.
- [59] ZHU H, CHEN X, DAI W, et al. Orientation robust object detection in aerial images using deep convolutional neural network [C]//2015 IEEE International Conference on Image Processing (ICIP). Quebec City, QC, Canada, 2015.
- [60] XIAO Z, LIU Q, TANG G, et al. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images [J]. International Journal of Remote Sensing, 2015, 36(02): 618 – 644.
- [61] ZHU Z, LIANG D, ZHANG S, et al. Traffic-Sign Detection and Classification in the Wild [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA, 2016.
- [62] ZHANG L, YANG X, LIU Z, et al. Single Shot Feature Aggregation Network for Underwater Object Detection [C]//2018 24th International Conference on Pattern Recognition (ICPR). Beijing, 2018.
- [63] YU X, GONG Y, JIANG N, et al. Scale Match for Tiny Person Detection [C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass, CO, USA, 2020.

科学编辑 齐乃新 博士 (火箭军工程大学)
责任编辑 唐定国