

本文引用格式: 刘伟鑫,林邦演,黄汉亿,等.基于 MBE-YOLOv5 的轻量化化工袋目标检测方法[J].自动化与信息工程,2024,45(2):34-40.

LIU Weixin, LIN Banyan, HUANG Hanyi, et al. Lightweight chemical bag target detection method based on MBE-YOLOv5[J]. Automation & Information Engineering, 2024,45(2):34-40.

基于 MBE-YOLOv5 的轻量化化工袋目标检测方法

刘伟鑫 林邦演 黄汉亿 李旻龙

(东莞市新一代人工智能产业技术研究院, 广东 东莞 523867)

摘要: 针对化工厂中化工袋种类繁多、遮挡干扰、放置复杂等因素,导致化工袋识别模型的定位效果较差、实时性不佳等问题,提出一种基于 MBE-YOLOv5 的轻量化化工袋目标检测方法。首先,用 MobileNetV3 网络替换 YOLOv5 的主干网络,降低模型的参数和运算量,提高模型的检测速度;然后,在 YOLOv5 的颈部网络引入双向特征金字塔网络结构进行多尺度特征融合,提高模型的识别准确率;最后,采用 EIoU 函数优化损失,提高模型的定位精度。实验结果表明,MBE-YOLOv5 模型相比 YOLOv5 模型,参数量下降了 37.7%,运算量降低了 58.1%,检测速度提升了 9.5%,mAP@0.5 提高了 0.7%;在检测速度和检测精度之间取得较好的平衡,能满足化工袋在线检测识别定位的要求。

关键词: YOLOv5 模型; MobileNetV3 网络; 双向特征金字塔网络; EIoU 函数; 化工袋目标检测

中图分类号: TP391.41

文献标志码: A

文章编号: 1674-2605(2024)02-0006-07

DOI: 10.3969/j.issn.1674-2605.2024.02.006

Lightweight Chemical Bag Target Detection Method Based on MBE-YOLOv5

LIU Weixin LIN Banyan HUANG Hanyi LI Minlong

(Dongguan Institute of New Generation Artificial Intelligence Industry Technology, Dongguan 523867, China)

Abstract: A lightweight chemical bag target detection method based on MBE-YOLOv5 is proposed to address the issues of poor positioning and real-time performance of chemical bag recognition models in chemical plants, which are caused by various types of chemical bags, occlusion interference, and complex placement. Firstly, replace the backbone network of YOLOv5 with MobileNetV3 network to reduce model parameters and computational complexity, and improve the detection speed of the model; Then, a bidirectional feature pyramid network structure is introduced into the neck network of YOLOv5 for multi-scale feature fusion to improve the recognition accuracy of the model; Finally, the EIoU function is used to optimize the loss and improve the positioning accuracy of the model. The experimental results show that compared to the YOLOv5 model, the MBE-YOLOv5 model reduces the number of parameters by 37.7%, the computational complexity by 58.1%, and the detection speed by 9.5%, mAP@0.5 Improved by 0.7%; Achieving a good balance between detection speed and accuracy can meet the requirements of online detection, recognition, and positioning of chemical bags.

Keywords: YOLOv5 model; MobileNetV3 network; bidirectional feature pyramid network; EIoU function; chemical bag target detection

0 引言

化工产业是我国国民经济的支柱产业和基础产业^[1]。化工袋作为化工产业的关键物流单元,其高效、安全地搬运对生产流程至关重要。目前,化工厂中化

工袋搬运主要依赖人工操作(叉车、吊车等),不仅效率低,且容易引发安全事故。此外,化工袋可能含有腐蚀性、有毒、有害物质,影响工人健康^[2-4]。因此,迫切需求开发一种自动、安全、高效的化工袋自动搬

运设备。其中,目标识别定位技术是该设备研发的关键点和难点。

采用传统的图像处理算法对化工袋进行目标识别定位^[5-6]时,因化工袋种类繁多、摆放不整齐、相互遮挡等问题,导致定位效果较差,无法满足实际生产需求。

随着深度学习目标检测技术的快速发展,其在自动驾驶、农业、食品、新能源、化工、半导体、工业、消费等领域均取得较好的应用效果^[7-9]。文献[10]采用 BiFPN 结构构建车辆检测模型的 Neck 网络,利用解耦头实现车辆目标的预测,在 UA-DETRAC 测试集上的平均精确率超过 99%。文献[11]提出一种改进的快速区域卷积神经网络 (faster region-based convolutional neural network, Faster-RCNN) 人脸检测算法,采用 ResNet50 替换主干网络,利用多尺度融合技术提高人脸目标检测精度,在 Wider Face 数据集上平均准确率达到 89%。文献[12]针对番茄目标检测,提出一种改进的 Mask R-CNN 模型,通过改造输入层实现 RGB 与深度图像的融合,采用多类预测子网络解耦预测,提高果实识别率,识别准确率达到 93.76%。文献[13]针对猕猴桃目标识别的应用场景,提出基于 Faster-RCNN 的机器人视觉系统,识别精度达到了 87.61%,且具有较好的鲁棒性。文献[14]提出一种船舶目标检测算法,在 Faster-RCNN 结构的基础上,设计一种具备特征放大和多尺度特征跳跃的连接结构,提高了船舶的检测精度,平均检测精度达到了 89.9%。然而,上述深度学习目标检测模型存在实时性差、参数量大、运算量大等问题,难以满足化工厂中化工袋快速、准确识别定位的要求。

为此,本文提出一种 MBE-YOLOv5 轻量化模型,在线快速识别定位化工厂中化工袋。首先,将 MobileNetV3 网络作为 YOLOv5 的骨干网络,以降低模型的参数量、计算量,加快推理速度;然后,模型颈部网络采用 BiFPN 结构,并融合多尺度特征信息,提高模型识别准确率;最后,采用 EIoU 函数作为边框回归

损失函数,提高模型定位精度。

1 MBE-YOLOv5 模型

1.1 YOLOv5 模型

YOLOv5 作为一种单阶段目标检测模型,继承了 YOLO 系列算法的核心思想,在保证检测精度的同时,进一步优化了模型的大小和推理速度。YOLOv5 模型的网络结构主要包括输入(Input)、主干(Backbone)、颈部(Neck)和检测头(Head)4部分。其中,在输入部分,模型对输入图像进行尺寸调整、归一化等预处理操作,以便模型能够更好地提取图像特征;主干作为模型的主体部分,采用跨阶段局部网络(cross stage partial network, CSPNet)结构,有效地减少了模型的计算量,同时增强了特征的提取能力,使模型更准确地识别目标;颈部通过路径聚合网络(path aggregation network, PANet)结构进行特征融合,使模型更好地利用不同尺度的特征信息,提高模型对不同大小目标的检测性能,从而更准确地定位目标;检测头部分负责将模型提取的特征转化为具体的目标检测结果,包括目标的位置、大小和类别等。

本文用 MobileNetV3 替换 YOLOv5 的主干网络;在 YOLOv5 的颈部网络引入双向特征金字塔网络(bidirectional feature pyramid network, BiFPN)结构;采用 EIoU 函数优化损失,实现 YOLOv5 模型轻量化的同时保证检测效果,使其更适用于化工袋的在线识别定位。MBE-YOLOv5 网络结构如图 1 所示。

1.2 MobileNetV3

MobileNetV3 是 Google 团队于 2019 年发布的一种轻量级的深度学习模型^[15],其保持较高检测精度的同时,降低了计算复杂度和模型大小。

MobileNetV3 中的 block 结构是核心部分。每个 block 采用倒残差结构,整体包括深度可分离卷积、SE 模块、线性瓶颈层和 h-swish 激活函数等部分,结构示意图如图 2 所示。

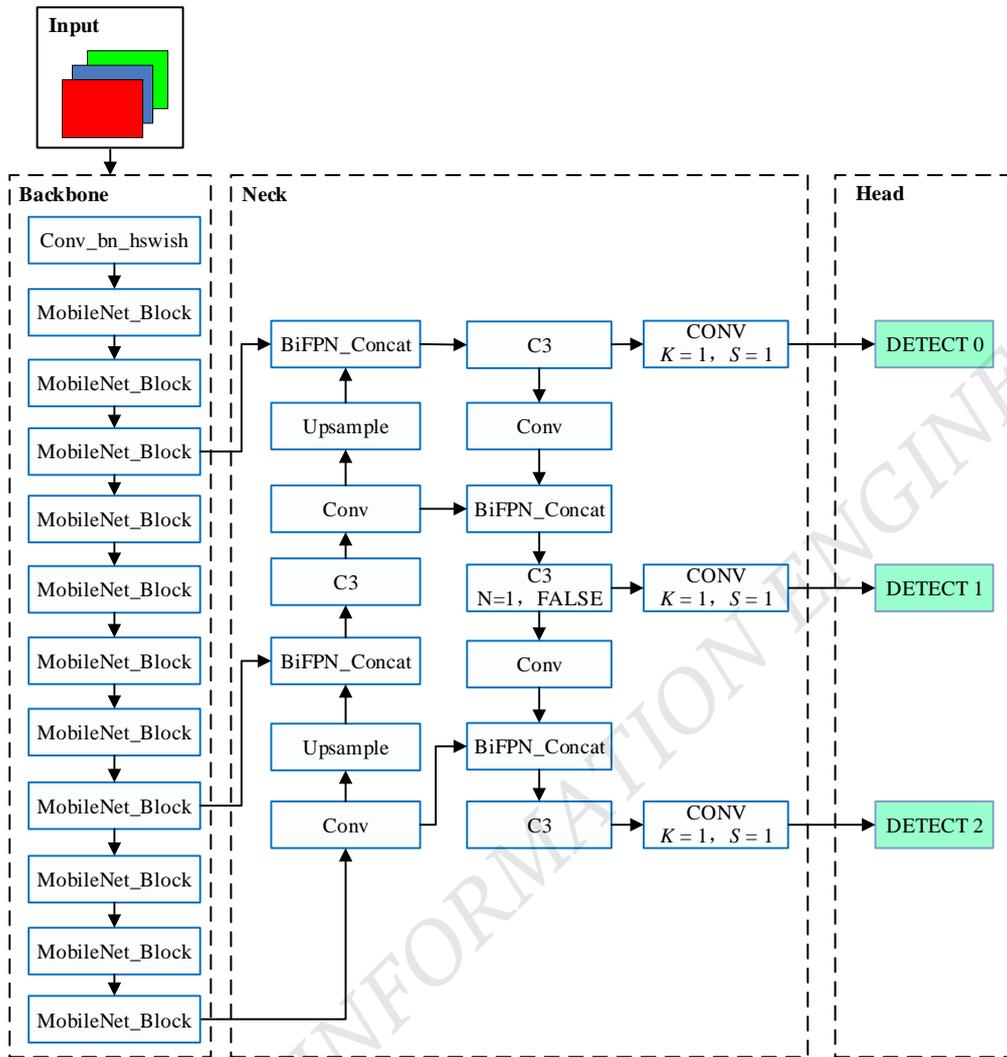


图1 MBE-YOLOv5 网络结构

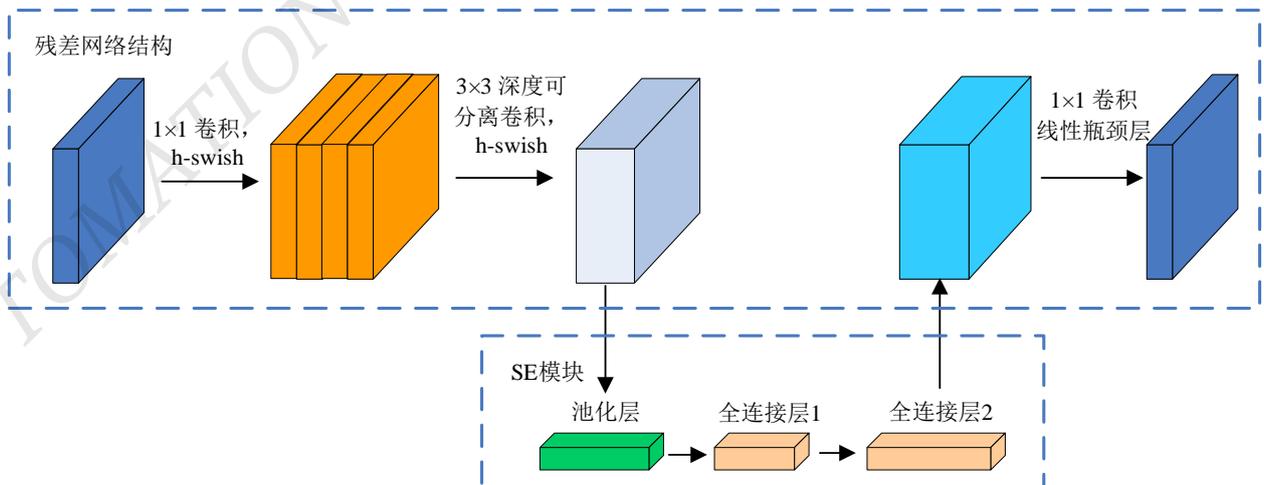


图2 block 结构示意图

深度可分离卷积操作包含两步: 第一步, 深度卷积, 每个输入通道特征图独立应用单个卷积核, 降低计算复杂度; 第二步, 逐点卷积, 采用 1×1 的卷积核对深度卷积后的特征图进行卷积操作, 整合各通道信息并调整通道数, 实现模型轻量化与性能优化。

压缩-激励 (squeeze-and-excitation, SE) 模块是 MobileNetV3 引入的一个轻量级注意力模块, 主要由 Squeeze 和 Excitation 子模块组成, 结构如图 3 所示。

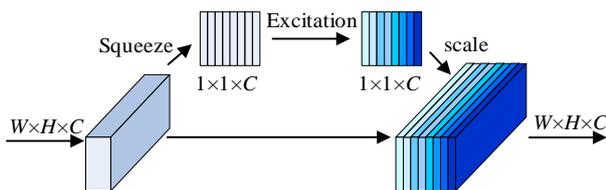


图 3 SE 模块结构

Squeeze 子模块通过对输入特征图进行通道维度上的压缩, 捕捉全局空间信息。利用全局平均池化将每个通道的特征图转换为一个实数, 代表该通道的全局响应。Excitation 子模块主要学习每个通道的重要性, 并根据这些重要性重新调整通道特征。Squeeze 子模块压缩后的特征向量经过 Excitation 子模块的 2 个全连接层: 第一个全连接层先将 Squeeze 的输出降低到较小维度, 以减少计算量, 再使用 ReLU 激活函数增加非线性; 第二个全连接层将维度恢复到与输入特征图通道数相同, 并使用 Sigmoid 激活函数生成每个通道的权重。

线性瓶颈层通过 1×1 的卷积核降低特征图维度, 从而减少模型的大小和计算量。

h-swish 激活函数在 swish 函数的基础上进行了改进, 通过减少乘法运算, 并允许更多的激活值位于 ReLU 的有效范围内, 提高了计算效率。

1.3 BiFPN 结构

YOLOv5s 颈部网络采用特征金字塔网络 (feature pyramid network, FPN) + 路径聚合网络 (path aggregation network, PAN) 的结构, 实现浅层和深层不同尺度的特征图融合, 以增强特征的语义表达能力。但在网络融合过程中, PAN 的输入信息为 FPN 处理后的信息, 导致融合后的特征没有充分地利用原始特

征信息。为提高特征融合效果, 本文采用 BiFPN 结构替代颈部网络的 (FPN+PAN) 结构, 增强相同网络层之间的特征信息融合能力和不同网络层之间的特征信息传递能力, 从而提高模型的感知和表达能力。

BiFPN 结构是一种多尺度特征融合的网络结构^[16-19], 如图 4 所示。

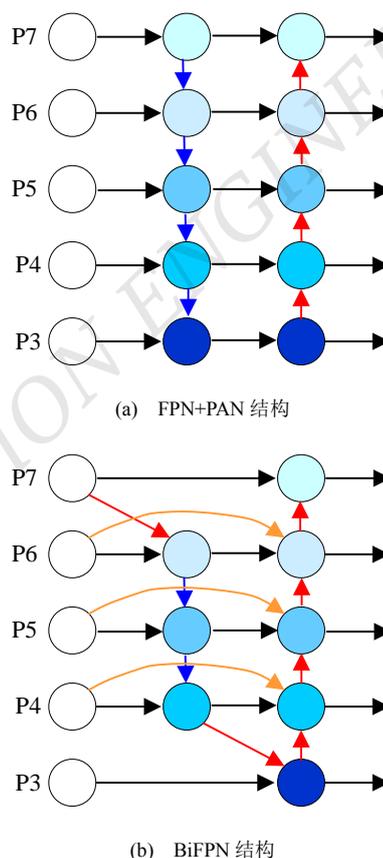


图 4 颈部网络结构对比

BiFPN 结构通过双向传播的方式将高层和低层的特征信息融合。BiFPN 结构在高层、低层特征图之间进行多次上下采样和融合, 使低层高分辨率特征与高层语义信息结合, 有效地捕捉目标的细节信息和跨尺度特征, 从而提高模型检测的准确性和鲁棒性。

1.4 EIou 损失函数

边界框损失函数对模型的目标定位和识别准确率至关重要。YOLOv5 采用 CIoU 损失函数, 其考虑了边界框的位置、形状和大小之间的关系。但 CIoU 损失函数面对目标尺度变化或长宽比较大时, 无法准确地衡量边界框的重叠程度, 影响目标检测的准确性^[20]。

为此，本文引入 EIoU 目标框损失函数^[21]。EIoU 目标框损失函数主要包含重叠程度、中心距离损失、宽高损失，能真实地反映预测框和真实框的高度和宽度。EIoU 目标框损失函数在 CIoU 损失函数的基础上分别计算宽高的差异值，提高预测框的回归速度，聚焦于高质量锚框，提高目标预测精度。EIoU 目标框损失函数的计算公式为

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp}$$

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

式中： w^c 和 h^c 分别为预测框和真实框外包最小矩形的宽和高， L_{IoU} 为 IoU 损失， L_{dis} 为距离损失， L_{asp} 为边长损失， b 和 b^{gt} 分别为预测框和真实框的

中心点， w 和 h 分别为预测框的宽和高， w^{gt} 和 h^{gt} 分别为真实框的宽和高， ρ 为欧几里德距离。

2 实验测试与分析

2.1 实验环境

软件环境：深度学习框架 pytorch 1.12、集成开发环境 spyder 5.0、第三方函数库 Anaconda3-5.2.0 以及操作系统 Windows10。

硬件环境：酷睿 i7-11700、英伟达 RTX3060、16 GB 内存、1 TB 硬盘。

2.2 实验数据集

本实验数据集来自某化工厂中 50 种不同规格型号的化工袋，共约 1000 幅图像，每种化工袋约有 20 幅图像。其中，750 幅图像用于训练，250 幅图像用于测试。部分化工袋图像样本如图 5 所示。



图 5 部分化工袋图像样本

2.3 评价指标

本文采用每秒检测帧数 FPS、运算量来评价 MBE-YOLOv5 模型的检测速度; 采用模型参数量来评价 MBE-YOLOv5 模型的复杂度; 采用 mAP@0.5 来评价 MBE-YOLOv5 模型的检测精度, mAP@0.5 是指当 IoU 为 0.5 时的 mAP, 计算公式为

$$AP @ 0.5 = \frac{1}{x} \sum_{i=1}^x P_i \quad (2)$$

$$mAP @ 0.5 = \frac{1}{n} \sum_{j=1}^n AP_{0.5_j} \quad (3)$$

式中: x 为某一类别正样本预测框的数量; P_i 为每个候选框对应的查准率; AP 为某一类别的所有查准率求和取平均值; mAP@0.5 为检测目标中所有类

别的 AP 平均值, 能直观地反映模型的性能; n 为类别数。

2.4 实验设计与分析

2.4.1 消融实验

为验证本文方法的有效性, 在实验数据集上设计消融实验, 具体如下:

- 1) YOLOv5s 模型作为消融实验的基准线;
- 2) 用 MobileNetV3 替换 YOLOv5s 的主干网络, 实现模型轻量化;
- 3) YOLOv5s 的颈部网络引入 BiFPN 结构, 提高模型的预测准确率;
- 4) CIoU 损失函数改为 EIoU 目标框损失函数, 提高目标预测精度。

消融实验结果如表 1 所示。

表 1 消融实验的结果

序号	MobileNetV3	BiFPN	EIoU	mAP@0.5	FPS/(f/s)	参数量/MB	运算量/GFLOPs
1				0.961	84.7	7.07	16.5
2	✓			0.936	94.8	3.4	6.4
3	✓	✓		0.955	91.9	4.4	6.9
4	✓	✓	✓	0.968	92.8	4.4	6.9

由表 1 可以看出: 用 MobileNetV3 替换 YOLOv5s 的主干网络后, 模型参数量下降了 51.9%, 运算量下降了 61.2%, 检测速度从 84.7 f/s 提升到 94.8 f/s, 但 mAP@0.5 下降了 2.5%; 颈部网络引入 BiFPN 结构后, 模型参数量增加约 1 MB, 运算量增加 0.5 GFLOPs, 检测速度下降约 3 f/s, mAP@0.5 提升到 0.955, 但比原模型低 0.6%; CIoU 损失函数改为 EIoU 目标框损失函数后, 模型参数量下降了 37.7%, 运算量下降了 58%, 检测速度提升了 9.5%, mAP@0.5 提高了 0.7%, 达到 0.968。

2.4.2 不同模型对比实验

为验证本文模型的有效性、优越性, 在实验数据集和相同测试环境下, 将本文模型 (MBE-YOLOv5) 与 YOLOv6n、YOLOv5s、YOLOv4-tiny、YOLOv7-tiny、YOLOv8n、YOLOv6s 等主流模型进行对比实验,

实验测试结果如表 2 所示。

表 2 7 种模型对比测试结果

模型	mAP@0.5	FPS/(f/s)	参数量/MB	运算量/GFLOPs
YOLOv6n	0.895	107.9	4.7	11.5
YOLOv5s	0.961	84.7	7.07	16.5
YOLOv4-tiny	0.832	125.5	8.9	5.6
YOLOv7-tiny	0.889	108.4	6.3	13.8
YOLOv8n	0.897	110.6	3.2	8.7
YOLOv6s	0.962	69.6	18.5	45.2
MBE-YOLOv5	0.968	92.8	4.4	6.9

由表 2 可以看出: 相比 YOLOv5s 模型, MBE-YOLOv5 模型参数量下降了 37.7%, 运算量下降了 58.1%, 检测速度提升了 9.6%, mAP@0.5 提高了 0.7%; YOLOv6n、YOLOv4-tiny、YOLOv7-tiny、YOLOv8n 模型的检测速度比 MBE-YOLOv5 模型快 10%~20%,

但这 4 个模型的 mAP@0.5 都没达到 0.9, 检测效果较差; 相比于 YOLOv6s 模型, MBE-YOLOv5 模型的 mAP@0.5 略高 0.6%, 但参数量、运算量、检测速度都有明显优势。

3 结论

本文针对化工袋目标检测任务, 提出一种 MBE-YOLOv5 模型。MBE-YOLOv5 模型采用 MobileNetV3 网络替换 YOLOv5s 主干网络, 减少模型的参数量、计算量; 同时引入 BiFPN 结构、EIoU 目标框损失函数来提高模型的目标检测精度。实验结果表明, 本文提出的 MBE-YOLOv5 模型在实现模型轻量化的同时, 获得较好的化工袋检测效果, 能满足化工厂中化工袋在线识别定位的要求。

参考文献

- [1] 张瑶, 崔燕, 李媛. 化工安全生产与环境保护的方法研究[J]. 化工管理, 2023(33):61-64. DOI:10.19900/j.cnki.ISSN1008-4800.2023.33.01
- [2] 马传海. 化工行业中工艺流程设计对安全生产的影响[J]. 现代盐化工, 2023, 50(6): 55-57.
- [3] CHEN C, RENIERS G. Chemical industry in China: The current status, safety problems, and pathways for future sustainable development[J]. Safety Science, 2020, 128:104741.
- [4] SONG J, HAN B. Green chemistry: A tool for the sustainable development of the chemical industry[J]. National Science Review, 2015, 2(3):255-256.
- [5] 刘华冠. 基于机器视觉的袋装物料位姿自动识别研究[D]. 济南: 济南大学, 2011.
- [6] 陈州尧, 徐敏, 苏鹭梅. 一种码垛视觉系统中物料袋的识别与定位方法[J]. 制造业自动化, 2015, 37(3):47-49;59.
- [7] LE Q, MIRALLES-PECHUÁN L, KULKARNI S, et al. An overview of deep learning in industry[J]. Data Analytics and AI, 2020:65-98.
- [8] WANG J, ZHANG T, CHENG Y, et al. Deep learning for object detection: A survey[J]. Computer Systems Science & Engineering, 2021, 38(2):165-182.
- [9] LIU L, OUYANG W, WANG X, et al. Deep learning for generic object detection: A survey[J]. International Journal of Computer Vision, 2020, 128(2):261-318.
- [10] SONG Y, HONG S, HU C, et al. MEB-YOLO: An efficient vehicle detection method in complex traffic road scenes[J]. Computers, Materials & Continua, 2023, 75(3):5761-5784.
- [11] 李祥兵, 陈炼. 基于改进 Faster-RCNN 的自然场景人脸检测[J]. 计算机工程, 2021, 47(1):7.
- [12] XU P, FANG N, LIU N, et al. Visual recognition of cherrytomatoes in plant factory based on improved deep instance segmentation[J]. Computers and Electronics in Agriculture, 2022, 197:106991.
- [13] SONG Z Z, FU L S, WU J Z, et al. Kiwifruit detection infield images using Faster R-CNN with VGG16[J]. IFAC-Papers on Line, 2019, 52(30):76-81.
- [14] LI Y, ZHANG S, WANG W Q. A lightweight faster R-CNN for ship detection in SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19:1-5.
- [15] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019:1314-1324.
- [16] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [17] 来春庆, 黄勇, 朱喆, 等. 基于特征融合多尺度卷积网络的光伏组件红外图像故障诊断[J]. 机电工程技术, 2023, 52(9): 145-151.
- [18] 郎永存, 李积元, 郑佳昕. 基于机器学习机床机械加工特征信息与加工材料关联性研究[J]. 机床与液压, 2022, 50(16): 194-199.
- [19] 杨俊杰, 谢维成, 曹倩. 级联 H 桥逆变器的多特征融合 CNN 故障诊断[J]. 中国测试, 2020, 46(7):8-17.
- [20] GEVORGYAN Z. SIOU loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv:2205.12740, 2022.
- [21] 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.

作者简介:

刘伟鑫, 男, 1992 年生, 硕士研究生, 工程师, 主要研究方向: 机器视觉与人工智能应用。E-mail: 15989156033@139.com