

基于多尺度特征增强的列车摘钩关键部件检测

潘培玉¹ 刘宏州¹ 孙欣¹ 李瑞² 石磊²
PAN Peiyu LIU Hongzhou SUN Xin LI Rui SHI Lei

摘要

针对列车摘钩关键部件风管和折角塞门的小尺寸目标检测问题,使用多尺度特征增强改进 YOLOv5s 检测算法。首先在主干网络增加一个多分支卷积模块 C2f 输出一个大尺寸特征图;然后在颈部结构采用残差特征增强方式 RFA 提升特征金字塔的多尺度特征融合表征能力;最后检测头输出 4 种尺度的检测特征图,其中新增特征图的尺寸最大,用于增强小尺寸目标的检测能力。实验结果表明,改进 YOLOv5s 在平均检测精度 mAP 上提升了 1.4%,优于 SSD,提升了小尺寸风管和折角塞门的检测能力。

关键词

铁路货运;列车摘钩部件;目标检测;多尺度特征;残差特征增强

doi: 10.3969/j.issn.1672-9528.2024.01.026

0 引言

摘钩是货运列车业务中的一项重要作业,概况为一关前二关后三提风管四提钩,涉及的部件有风管、折角塞门和上下作用式提杆。随着机器人和智能视觉技术的发展,自动摘钩机器人^[1]正逐步替代安全风险高、效率低的人工摘钩作业,在设计摘钩机器人检测系统时,需重点关注折角塞门与风管的检测,其检测精度直接影响摘钩作业的完成度。

风管和折角塞门的检测属于目标检测(object detection),包括目标识别与定位任务。目前,深度学习目标检测算法为主流方法,按检测流程可分为两阶段和一阶段检测算法。两阶段检测算法,如 Faster-RCNN^[2],首先候选建议框 proposals 进行目标粗略定位,然后使用检测头对目标进行精细定位并分类。一阶段目标检测算法无需候选建议框,直接地回归目标物体的位置及其类别,典型算法包括 YOLOv5^[3],SSD^[4]。基于深度学习的目标检测算法逐步在铁路列车行业开展应用,如电务轨旁的设备巡检中的铁道塞钉自动定位技术^[5],基于神经网络结构搜索的铁路施工人员安全帽佩戴检测等^[6]。郭忠峰等人^[7]使用 K-means 聚类算法对预设锚框的个数及其宽高值调优,改进 YOLOv2^[8]检测摘钩作业中的上下作用式提风管。

风管和折角塞门的检测具有一定的难度。折角塞门属于小目标,它的尺寸不到整幅图像的 1%。风管和折角塞门的

形状不规则,其中风管的标注框会产生镂空现象。此外,折角塞门与周围部件由相同的材质构造,与其它部件具有相似的纹理,容易引起误检。目前,基于深度学习目标检测算法进行风管和折角塞门检测的文献较少。鉴于 YOLOv5 可用性强,在工业界广受欢迎的特点,本文通过改进 YOLOv5 的结构,增强多尺度特征提取能力,提升风管和折角塞门的检测精度。

1 改进型 YOLOv5 目标检测算法

1.1 YOLOv5 算法

YOLOv5 检测算法由 Ultralytics 团队在 2020 年提出,在工业界广受欢迎^[9-10],在开源网站 GitHub 上已获得超 1.5 万次的项目优化 fork。YOLOv5 的结构可分为图像输入端、主干网络 backbone、neck 和检测头 head 输出端等 4 个部分。图像输入端对图像进行增强操作。目前版本 YOLOv5(v6.0)在主干网络采用 C3 模块,即组合 3 个 CBS 模块的方式进行特征提取。neck 部分进行多尺度特征融合,主要采用特征金字塔模式 FPN^[11](feature pyramid network)和路径聚合网络 PAN^[12](path aggregation network)。检测头标示目标位置并输出类别概率。本文将改进 YOLOv5 的主干网络增强小目标物体的特征表征能力,并在 neck 部分进行残差特征增强。

1.2 主干网络的改进

图 1 中虚线框显示改进部分,包括主干网络新增一个 C2f 模块和大尺寸特征图 160×160 输出分支,neck 部分使用 RFA 改进特征金字塔 FPN 以增强多尺度特征融合能力,检测头部分新增一个大尺寸特征图 160×160×21 提升小目标检测能力。

1. 陕西榆横铁路有限责任公司 陕西榆林 719000

2. 四川国软科技集团有限公司 四川成都 610031

[基金项目] 四川省经济和信息化厅-中国制造 2025 四川行动资金项目, CJXDZ-Y2021-003; 陕西煤业化工集团有限责任公司 2021 年度板块级科研项目, 2021SMHKJ-BK-J-52

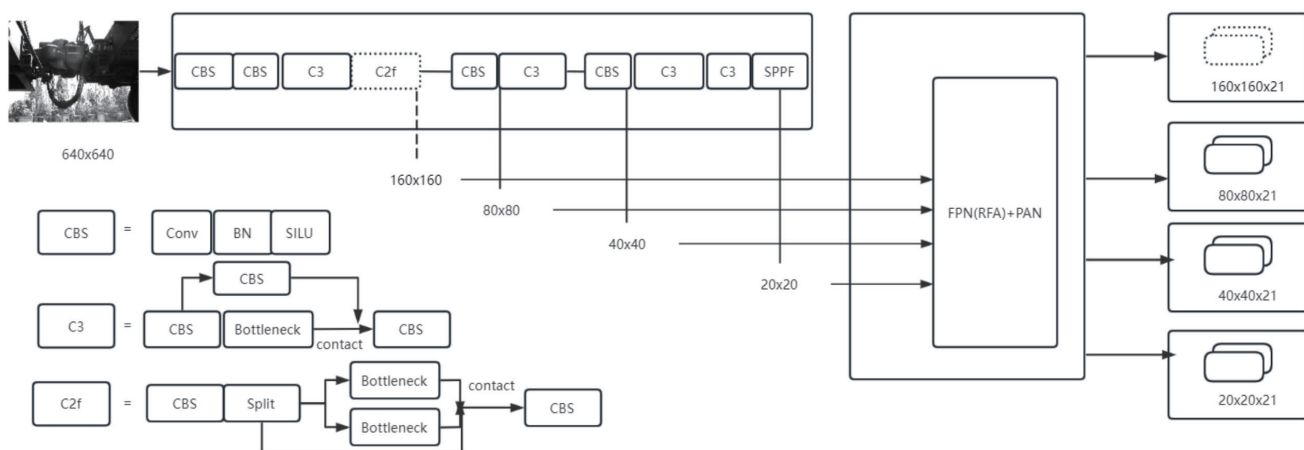


图1 改进后的YOLOV5结构

主干网络 backbone 用于提取图片特征，供后续的颈部模块和检测头使用。为提升小目标的检测能力，本文在主干网络增加了一个大尺寸的输出分支，并使用 C2f 模块丰富该分支的梯度流信息。

多尺度特征^[11-17]是小目标检测常用的方法之一，其中大尺度特征图用于检测小目标。原因在于大尺度特征图下采样的倍数较小，每个特征点对应的尺寸范围也较小。假设输入图像为 $640 \times 640 \times 3$ ，原 YOLOv5 输出特征图的大小分别为 $80 \times 80 \times N \times (5+C)$ 、 $40 \times 40 \times N \times (5+C)$ 和 $20 \times 20 \times N \times (5+C)$ ，其中 N 为每个特征点的锚框数量， C 为类别数量。最大尺度特征图 $80 \times 80 \times N \times (5+C)$ 负责小目标检测，特征图单元的检测分辨率下限约为 $(640/80) \times (640/80) = 8 \times 8$ 。同理， $40 \times 40 \times N \times (5+C)$ 和 $20 \times 20 \times N \times (5+C)$ 的特征单元的检测分辨率下限约为 16×16 和 32×32 。在 YOLOv5 中， 80×80 特征图的预设锚框的大小分别为 10×13 、 16×30 和 33×23 。这只是理论值，由于网络提取特征的能力随着网络深度的增加逐步地损失，实际能检测的目标尺寸远大于预设锚框值，因而导致小目标检测效果不理想。

为了检测折角塞门这一小目标，本文在主干网络增加了一个尺寸为 160×160 的大尺度特征图输出分支（图1虚框所示）。新增特征图单元的检测分辨率下限约为 4×4 ，为原 YOLOv5 的检测下限的 $1/4$ ，因而比原模型具有更强的小目标检测能力。从图1可以看出，新增特征图是浅层特征。为提升新增特征图的特征表达能力，本文不直接从 C3 模块后提取特征，而是新增了一个 C2f 模块用于生成新增特征图，并将该特征图作为一个分支输出到 neck 和 head。如图1所示，C2f 模块结合了流行的梯度分流和残差结构模块。经过 CBS 后，其中 CBS=卷积(Conv)+

块归一化(BN)+SiLU 激活函数，C2f 使用 Split 函数将特征图分成相等的 2 部分，其中一半的通道分配给 2 个级联的 Bottleneck 进行特征提取，每个 Bottleneck 的输出与原有特征进行合并操作，生成了 2 倍原通道数的特征图输入 CBS 模块提取新增的特征。除了 C2f，C3 模块也被使用在网络中。SPPF (spatial pyramid pooling-fast) 模块对于检测性能提升至关重要，也被用在了改进后的网络中。它将 SPP 的三种并联最大池化改成了只有一种 5×5 卷积核的级联最大池化，提升了运算速度。

1.3 多尺度残差特征增强

为进一步提升特征融合能力，本文将改进 neck 的特征金字塔 FPN。FPN 融合本文提出的 4 类不同尺度特征图。最高层特征图 20×20 自顶向下地传递上下文信息给其它尺度的特征图。最高层特征图在融合前使用 1×1 卷积降低通道数，这造成了信息损失。本文通过特征残差增强模块 Residual feature augmentation (RFA)^[18] 为顶层特征图增加多样化的上下文信息，弥补特征信息丢失的损失。RFA 的结构如图2的虚线框所示，依次执行自适应池化，双线性上采样插值和特征图的相加求平均值等 3 个步骤。增强后的特征图与原有顶层特征图相加，然后自顶向下地与其它尺度特征进行融合。RFA 基于残差的思想将多样化的空间上下文信息注入原有顶层特征图，有利于减少特征信息损失，并提高特征金字塔性能。除了使用 FPN 自顶向下地融合多尺度特征，PAN 模块也在 neck 结构中被使用，实现自下而上地传递图形信息到高层语义特征中，以达到提升模型目标定位的能力。改进后的 FPN+PAN 结构如图2所示，图2中虚线框显示本文所做的改进，residual feature augmentation (RFA) 为特征残差增强模块。虚线框展示了 RFA 详图，a1、a2、a3 是自适应池化的系数。

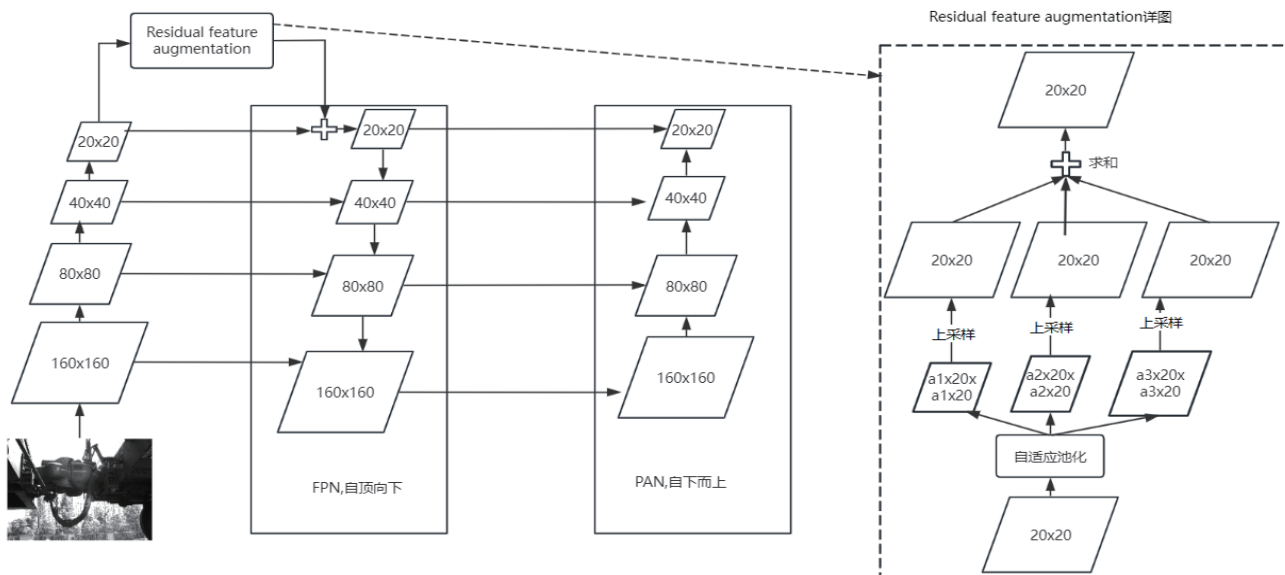


图2 改进后的FPN+PAN结构

多尺度残差特征增强的步骤如下。

(1) 输入图像经过主干网络提取特征后,得到4个尺度不同的特征图。

(2) 对最高层特征图 20×20 进行特征残差增强RFA,增强后的特征与原顶层特征图相加。

(3) 对高层的小尺度特征图,进行上采样操作,使其尺寸与较大尺度的特征图相同。

(4) 使用加法操作依次对 40×40 、 80×80 、 160×160 的特征图进行自顶而下地融合。

(5) 使用PAN进一步地对特征图进行融合增强,提升检测算法的定位能力。

1.4 检测头的改进

改进后的检测头含有 $160 \times 160 \times 21$ 、 $80 \times 80 \times 21$ 、 $40 \times 40 \times 21$ 和 $20 \times 20 \times 21$ 等4个特征图。新增大尺度特征图的大小为 $160 \times 160 \times 3 \times (5+2) = 160 \times 160 \times 21$,其中 160×160 表明待测图像被分割后的总单元数目,3表示每个特征单元格使用3个锚框作为预测框,5=4+1表示目标的左上和右下坐标 (x_1, y_1) 、 (x_2, y_2) 和有无物体的概率 p ,2表示风管与折角塞门2个类别。每个锚框输出7维特征,拼接3个锚框的特征获得了21维特征。对于一张待检测图像,检测头的4个分支都会输出不同尺寸大小的目标框,经过如概率 p 的限制,非最大值抑制等筛选机制后,得到最终的检测图。

2 实验结果与分析

2.1 实验设置

本实验的主要硬件设备是XEON 6138 CPU, 32 GB内存和一张Nvidia GeForce RTX3090Ti 24 GB GPU卡,实验软件

平台为PyTorch。数据集由2340张尺寸为 $1920 \times 1080 \times 3$ 的RGB图片构成,涵盖了不同拍摄距离、视角、光照下的不同场景图片。Labelimg用于标定风管和折角塞门,类别名称分别为fengguan、zhejiao。训练和测试集图像数目分别为1950张和390张。训练数据增强和训练策略使用了Ultralytics的默认方式。评估指标采用类别平均精度均值 P (mean average precision),定义为:

$$P = (P_1 + P_2) / 2 \quad (1)$$

式中: P_1 和 P_2 分别为风管和折角塞门的平均检测精度。mAP50和mAP@50:5:95,是COCO目标检测数据集上的典型指标,前者设置IOU=0.5,后者的IOU值从0.5到0.95,步长为0.05。由于采用了更大的IOU阈值,意味着需要更大的重叠范围,因而采用mAP@50:5:95进行评估的检测更难,其值小于mAP50。

2.2 实验结果

本次实验与工业界常用的目标检测算法进行对比,包括YOLOv5s、YOLOv5x、SSD以及两阶段目标检测器Faster R-CNN。表1显示改进后的YOLOv5s比改进前的YOLOv5s平均提升了约1.4%的mAP@50:5:95,这得益于新增了一个大尺度特征,并使用RFA进行多尺度特征融合。YOLOv5x是YOLOv5系列中最宽、最深、参数最多的网络。然而,由于本文的训练数据集有限,YOLOv5x的测试性能与YOLOv5s相同。这表明在工业应用上,应根据场景选择合适的检测器。在IOU=0.5的mAP50的性能比较上,除了SSD,所有方法都取得了大于99.0%的检性能,本文方法取得了最好的检测值99.5%。

表 1 对比实验结果

检测模型	mAP@50:5:95/%	mAP50/%	每张图片平均检测时间 /ms
YOLOV5s	91.0	99.5	7
本文改进 YOLOV5s	92.4	99.5	7.7
YOLOV5x	91.0	99.5	14
Faster R-CNN	93.0	99.3	50
SSD	46.5	89	35

表 1 结果显示,改进方法在时间性能上达到了约 130 帧/s,即检测一张图片大约需要 7.7 ms,远快于两阶段目标检测器 Faster-RCNN 的 20 帧/s。

3 结论

本文基于多尺度特征增强的思想改进 YOLOv5s 的主干网络,颈部结构和检测头。首先主干网络使用多梯度流分支模块增加了一个大尺度的特征输出分支;然后颈部结构使用残差特征增强提升特征金字塔的特征融合能力;最后在检测头部分形成了 4 个特征图输出用于检测尺寸小,形状不规范的风管和折角塞门。实验结果表明,改进 YOLOv5s 算法达到了 92.4% 的平均检测精度,优于工业界常用的目标检测算法。后续工作将研究本文检测算法与 3D 检测算法的融合,以进一步增强摘钩机器人的识别能力。

参考文献:

[1] 马波,裴正定. 驼峰摘钩机械手最优控制仿真及机器视觉分析[J]. 铁道学报, 1998(3):6-11.

[2] REN S, HE K, GIRSHICK R B, et a. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 39: 1137-1149.

[3] YOLOv5 官网 [EB/OL].[2023-08-01].<https://github.com/ultralytics/YOLOv5>.

[4] LIU W, ANGUELOV W, ERHAN D, et al. SSD: single shot multi-box detector[J].European Conference on Computer Vision (ECCV), 2016(9):132-139.

[5] 杜馨瑜. 基于计算机视觉技术的铁道塞钉自动定位算法[J]. 铁道通信信号, 2016, 52(9):68-72.

[6] 张榕,周瑾,焦雯,等. 基于神经网络架构搜索的安全帽佩戴检测算法研究[J]. 铁道通信信号, 2022,58(4):43-47.

[7] 郭忠峰,张渊博,王赫莹,等. 深度学习目标检测算法在货运列车车钩识别中的应用[J]. 铁道科学与工程学报, 2020,17(10):2479-2484.

[8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//IEEE Conference on Computer Vision Pattern Recognition. Piscataway:IEEE, 2017:6517-6525.

[9] 罗伟,刘思远,徐健祥,等. 基于改进 YOLOv5s 的太阳能电池缺陷检测算法[J]. 计算机与现代化, 2023(7):119-126.

[10] 王艺成,张国良,张自杰. 基于改进 YOLOv5 的小目标检测方法[J]. 计算机与现代化, 2023(5):100-105.

[11] LIN L, DOLLÁR P, GIRSHICK R B, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2016: 936-944.

[12] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2018: 21-26.

[13] 朱柏松,王燕妮. 基于多尺度跨层特征融合的轻量化小目标检测算法[J]. 探测与控制学报, 2023,45(6):77-86.

[14] 张浩,董锴龙,高尚兵等. 融合非临近跳连与多尺度残差结构的小目标车辆检测[J]. 中国图象图形学报, 2023, 28(12):3797-3809.

[15] 秦振,李学伟,刘宏哲. 基于改进 SSD 的鲁棒小目标检测算法[J]. 东北师大学报(自然科学版), 2023,55(4):59-66.

[16] 冷佳旭,莫梦竞成,周应华等. 无人机视角下的目标检测研究进展[J]. 中国图象图形学报, 2023,28(9):2563-2586.

[17] 郑东宏,何志权. 基于多尺度特征增强的小样本计数[J]. 生物学工程进展, 2023,44(4):405-411.

[18] GUO C, FAN B, ZHANG Q, et al. AugFPN: improving multi-scale feature learning for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020:12592-12601.

【作者简介】

潘培玉(1987—),男,安徽贵池人,硕士研究生,经济师,研究方向:智能化铁路运输、机器视觉。

刘宏州(1982—),男,陕西周至人,本科,工程师,研究方向:铁路运输。

孙欣(1988—),男,陕西富平人,本科,工程师,研究方向:铁路运输。

李瑞(1994—),男,四川南部人,本科,工程师,研究方向:软件工程。

石磊(1980—),通信作者,男,四川宜宾人,本科,工程师,研究方向:机电一体化、智能化铁路。

(收稿日期: 2023-11-03)