

文章编号: 2095-2163(2022)10-0227-08

中图分类号: U279.3

文献标志码: A

基于改进 YOLOv4 的地铁车辆螺栓状态检测方法

徐宝康¹, 郑树彬^{1,2}, 戚玮玮^{1,2}, 李立明^{1,2}, 丁亚琦³

(1 上海工程技术大学 城市轨道交通学院, 上海 201620; 2 上海工程技术大学 上海市轨道交通振动与噪声控制技术工程研究中心, 上海 201620; 3 上海地铁维护保障有限公司车辆分公司, 上海 200235)

摘要: 针对地铁车辆螺栓数量多、目标小导致的高误检率问题, 提出一种基于改进 YOLOv4 的地铁车辆螺栓状态检测方法。采用级联的策略分 2 步检测实现螺栓目标的定位与分类, 从而降低误检率。两步检测算法均采用改进的 YOLOv4: 首先利用聚类算法获取先验框的尺寸, 并通过改变先验框生成的初始位置, 优化预测框位置回归的策略, 以提升网络训练与检测速度, 其次重构了特征融合机制, 用自适应空间特征融合的方法对 PANet 模块的输出进行操作, 改善了特征的比例不变性, 提高了对螺栓的检测精度。实验结果表明, 提出的检测方法 *mAP*、召回率分别可达 99.5%、99.8%, 可更高效地分类与识别螺栓小目标。

关键词: 地铁车辆; 螺栓状态检测; 深度学习; 特征提取; YOLOv4

Improved YOLOv4 detection algorithm for metro vehicles bolt condition

XU Baokang¹, ZHENG Shubin^{1,2}, QI Weiwei^{1,2}, LI Liming^{1,2}, DING Yaqi³

(1 School of Urban Rail Transportation, Shanghai University of Engineering Science, Shanghai 201620, China; 2 Shanghai Engineering Research Centre of Vibration and Noise Control Technologies for Rail Transit, Shanghai University of Engineering Science, Shanghai 201620, China; 3 The Vehicle Branch, Shanghai Metro Maintenance Guarantee Co., Ltd., Shanghai 200235, China)

[Abstract] Aiming at the problem of high error detection rate caused by large number of bolts and small target in metro vehicles, a detection method for the status of bolts in metro vehicles based on improved YOLOv4 is proposed. In order to reduce the error detection rate, a cascade detection strategy is adopted to realize the positioning and classification of bolt targets in two steps. Improved YOLOv4 is adopted for both detection algorithms: First, clustering algorithms are used to obtain a priori box size, and by changing the prior box to generate the initial position. After that, the box position regression forecasting strategy is optimized to enhance the speed of network training and testing, and the feature fusion mechanism is reconstructed. Based on the above, by the method of adaptive spatial feature fusion PANet module output, the characteristics of scale invariance is improved, the detection accuracy of bolts is improved. The experimental results show that the proposed detection method *mAP* and recall rate can reach 99.5% and 99.8% respectively, which can classify and identify bolt small targets more efficiently.

[Key words] metro vehicle; bolt state testing; deep learning; feature extraction; YOLOv4

0 引言

轨道车辆的螺栓状态检测是保障列车安全运营的一项重要检修工作, 地铁车辆段主要由简单直接的人工巡检完成, 但工人长时间工作造成的视觉疲劳和精力下降会导致效率低下、误检率高等问题。目前已有地铁维保公司安装了基于机器视觉的检测设备, 但采用的检测算法误检率过高, 检测算法的性能亟需提升。

近年来国内外对利用机器视觉实现螺栓状态检测技术进行了研究。目前主要检测方法是利用图像

处理技术对螺栓特征进行识别检测, 如文献[1]采用混合高斯模型剔除正常螺栓, 利用 SIFT 提取螺栓边缘和纹理等特征。而传统图像处理方法提取的是浅层特征信息, 存在定位不准、误漏率高的缺陷。有学者采用基于深度学习来建立目标检测模型, 如文献[2]提出利用 Faster Rcnm 级联的输电塔螺栓缺失检测方法。文献[3]利用 SSD 算法定位螺栓并切割, 再利用 YOLOv3 检测缺陷螺栓。深度学习方法虽较前述方法优势突出, 但针对特定小目标检测仍存在不足。

目标检测算法按检测策略分为 2 类。一类是检

基金项目: 国家自然科学基金(51975347, 51907117)。

作者简介: 徐宝康(1993-), 男, 硕士研究生, 主要研究方向: 基于深度学习的目标检测; 郑树彬(1979-), 男, 博士, 教授, 主要研究方向: 轨道交通车辆/轨道设备状态检测技术与理论。

通讯作者: 郑树彬 Email: shubin.zheng@sues.edu.cn

收稿日期: 2021-12-07

测和定位分2次完成的两阶段检测算法,例如 Faster Rcnm^[4]、Mask Rcnm^[5],此类算法精度高但检测速度较慢。另一类是直接一次完成检测定位的单阶段检测算法,例如 SSD^[6]、YOLO^[7-9]系列,此类方法检测速度快,但精准度偏低。其中,YOLOv4 对检测精度与时效性较均衡,但实际工程检测中需根据数据集具体特征进行优化。针对螺栓检测的具体问题特征来改进 YOLOv4^[10],采用自适应空间特征融合(ASFF)^[11]技术对多尺度特征地图进行融合,改

善各特征比例固定的性质,提升检测性能。将改进 YOLOv4 算法在实验数据集上训练,以期缩短训练时长,改进检测精度及速度,并在验证集上验证、对比检测结果。

1 YOLOv4 算法简介

YOLOv4 算法由 Alexey 等人提出^[10],可分为主干网络、尺寸变换模块和检测器预测模块三部分,具体结构及相应模块如图1所示。

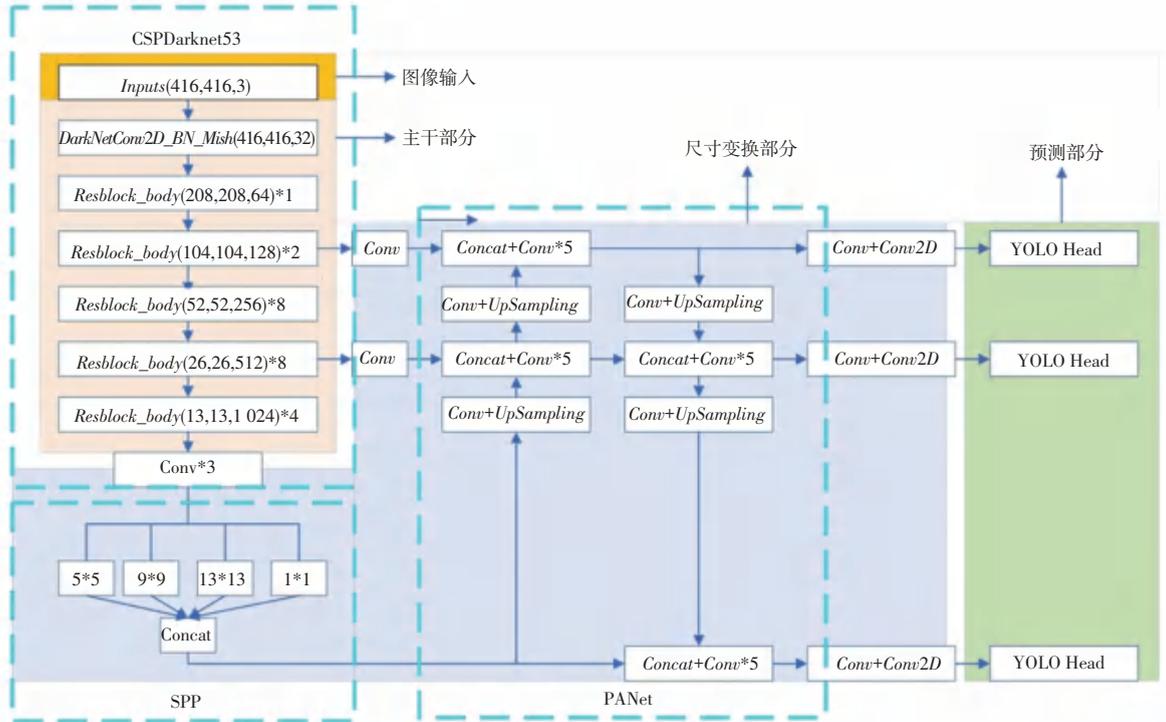


图1 YOLOv4 结构图

Fig. 1 YOLOv4 structure diagram

主干网络的结构部分融入 CSPNet^[12]的设计思路,在 ResNet^[13]残差结构基础上添加 CSP 结构构成 CSPX 基本组件,并采用 Mish 函数保证较好的信息流动性。整个网络共 52 层,通过卷积、下采样、特征合并、残差的堆积复用,在 DarkNet53 基础上构成新主干网络结构。

尺寸变换模块主要由 SPPNet^[14]、FPN^[15] + PANet^[16]构成,如图1所示。经 SPPNet 并行池化再合并所产生的固定大小特征更有效地扩张主干特征的提取范围,见图1中 SPP 模块。FPN+PANet 结构中,FPN 输出更多的强语义特征,PANet 输出更多的强定位特征,2种网络将2种特征通过不同主干与检测层实行参数合并达到融合目的,高底层间过长的信息传输路径也因此缩短,减少信息衰减,见图1中 PANet 模块。

检测器预测模块是对输出特征进行解码。首先

使用 $n \times n$ 网格划分图片,以网格为预测单位确认网格中目标,对存在目标的网格生成先验框,根据特征信息得出目标中心点对网格左上角的偏移量,再加上对应网格的位置信息,其中预测置信度和预测类别概率的乘积为预测结果的得分,通过 NMS 排除冗余预测框,设定预测阈值,在逐层筛选后,最终得到预测结果。

2 基于改进 YOLOv4 的地铁车辆螺栓状态检测方法

检测模型的改进分为2个方向:模型的检测策略改进和 YOLOv4 网络改进。整体的检测策略使用二级级联的方法,分2步实现螺栓目标的定位与分类;YOLOv4 网络改进是对算法和结构的优化,上述的二级级联的2步检测均采用此改进后的 YOLOv4

作为检测算法。

2.1 二级 YOLOv4 网络的级联

转向架侧梁螺栓的识别特征较为简单, 地铁车身在日常运行过程中会产生脏污斑点, 其外形和螺栓目标相似, 且车身图像内容复杂, 更容易导致误检率增加的问题。观察地铁车辆侧身图像, 可发现螺栓目标所处位置较集中, 构成规则的矩形阵列, 因此可将整体看作为一个目标。本次研究中, 就可以采用级联结构来检测螺栓矩形阵列, 对螺栓的目标检

测由粗到细地进行定位和分类。检测系统的二级级联示意如图 2 所示。由图 2 可知, Stage 1 中首先使用下述改进后的 YOLOv4 算法对整体图像进行粗定位, 得到螺栓矩形阵列的目标区域, 输入图像调整为 900×900, 由于螺栓矩形阵列面积大且特征明显, 所以容易识别。该输出作为 Stage 2 的输入再次对螺栓目标进行定位和分类, 因此在 Stage 2 中基于改进 YOLOv4 的定位分类体系结构被级联。通过此级联可降低螺栓目标的误检率, 提高检测精度和召回率。

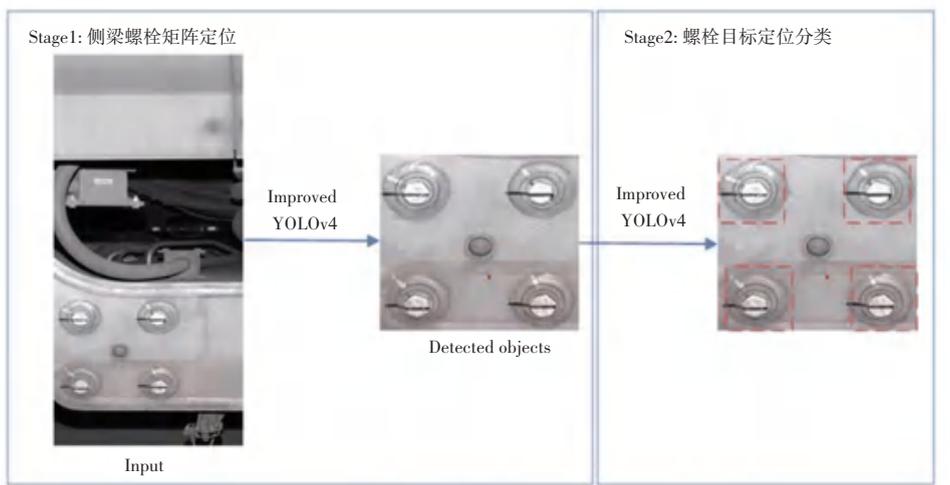


图 2 检测系统的二级级联示意图

Fig. 2 Schematic diagram of two level cascade of detection system

2.2 YOLOv4 网络改进

2.2.1 先验框的聚类与设定

根据数据集的特征先将先验框尺度聚类为 2 类, 再按尺寸大小依次分配给 3 个特征层。由于本实验的数据集目标尺寸变化率极小, 统计出螺栓所占像素为 250×260 左右, 面积占原图 2 000×4 096 的 0.79%, 可知螺栓尺寸单一且归属于小目标。原算法聚类得到的先验框是为了覆盖各种尺寸目标, 不适用于本实验数据集。改进的算法先根据 SSE - k 的畸变程度曲线, 如图 3 所示。由图 3 可见, 依据肘部法可得最佳 K 值为 2, 经 K-means++ 对基于输入尺寸 416×416 的螺栓尺寸数据进行聚类, 得到尺寸为 (30.4, 3.1) 和 (34.7, 42.4), 为了计算简便取整数值, 将先验框 0~5 设为 (30, 34), 6~8 设为 (35, 42)。先验框的设定影响训练过程的速度, 依据数据集的特点获取尺寸可加快模型回归速度。

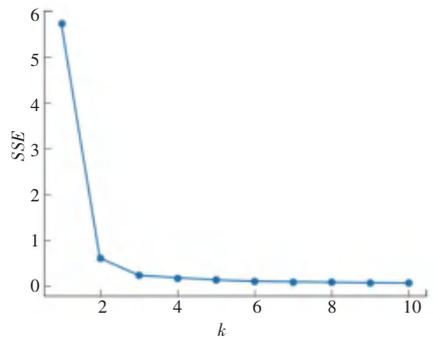


图 3 SSE-k 畸变程度曲线

Fig. 3 SSE-k distortion degree curve

2.2.2 先验框的回归改进

自适应生成先验框位置如图 4 所示。原算法是以每网格的左上角生成先验框, 目标由其中中心点所处的网格负责预测。而目标中心点位置随机, 常会出现先验框中心起点坐标远离目标中心点坐标, 这种情况下则需要网络大量负反馈调整预测框位置降低 Loss 值以提高交并比, 因此影响了回归效率。改

进网络提出根据目标中心点的具体坐标,在算法中通过四舍五入函数 $round(x,y)$, 选取最靠近中心点的网格角点作为生成先验框起点中心坐标,再通过特征层的调整策略进行网络参数的回归更新。此时网格的4个角点均可为先验框的生成位置,边界框的位置计算方式依据4种生成位置选择其计算方式,以图4为例先验框生成位置为右下角,其对应的预测框实际位置计算参见式(1):

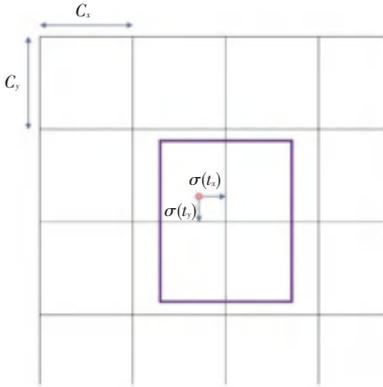


图4 自适应生成先验框位置

Fig. 4 Adaptively generating priori box position

$$\begin{cases} b_x = 2c_x - \sigma(t_x) \\ b_y = 2c_y - \sigma(t_y) \end{cases} \quad (1)$$

其中, c_x 、 c_y 为网格单位长度; t_x 、 t_y 为坐标变换系数; b_x 、 b_y 为预测框的实际位置。

该改进智能选择生成先验框的初始位置,有效提高网络的训练效率。

2.2.3 网络结构改进—多尺度特征融合模块

对于处理特征融合方式,原 YOLOv4 网络在 FPN 基础上融入了 PANet, 缩短了高低层间特征传递路径,保证强语义及强定位信息的融合。但单级检测模型多直接使用 *concat* 或 *add* 的方式融合特征,而各特征尺度是不一致的,上述方式并不能充分利用各尺度的不同特征,一定程度限制了特征融合的效能。改进的算法在此基础上将网络的输出特征进行 ASFF 操作,通过学习不同尺度特征层之间的联系参数来处理上述问题,在空间层面,冲突信息被限制传递,梯度反传时使其不一致性得以抑制,各特征之间的比例问题得到改善,同时降低推理消耗,提高了网络对螺栓检测的精度^[15]。

图5描述了由不同特征尺寸融合的方式,由 $level - 1 \sim level - 3$ 输出,3层特征分别与权重参数相乘再求和,得到新的融合特征。其中,权重参数是通过网络自适应地学习获得,只保留有用的信息进行组合,可看作是代表每个位置上不同层次特征的导入程度,有效地避免不同尺度特征融合时出现的空间矛盾问题,从而增强了特征表示,此处需要用到的数学公式依次可写为:

$$\gamma_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l} \quad (2)$$

$$\alpha_{ij}^l = \frac{e^{\lambda_{\alpha_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \quad (3)$$

$$\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1 \quad (4)$$

其中, α_{ij}^l 、 β_{ij}^l 、 γ_{ij}^l 分别指3个不同级别尺寸的特征空间权重,是由网络自适应学习获得; $x_{ij}^{1 \rightarrow l}$ 、 $x_{ij}^{2 \rightarrow l}$ 、 $x_{ij}^{3 \rightarrow l}$ 分别指来自 $level - 1 \sim level - 3$ 三层的特征。

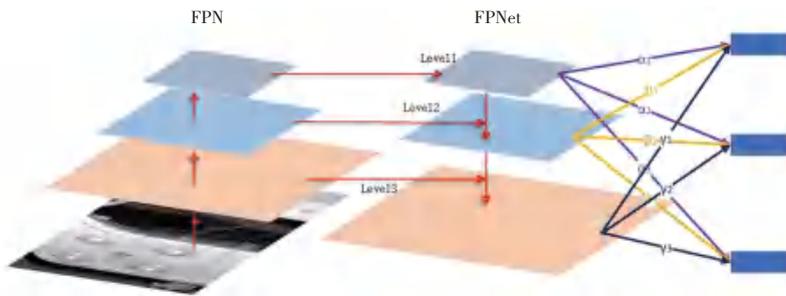


图5 自适应空间特征融合示意图

Fig. 5 Schematic diagram of adaptive spatial feature fusion

式(3)、式(4)为空间权重的限制条件,此处的 $\lambda_{\alpha_{ij}}^l$ 、 $\lambda_{\beta_{ij}}^l$ 、 $\lambda_{\gamma_{ij}}^l$ 是激活函数中的控制参数,并且 α_{ij}^l 、 β_{ij}^l 、 $\gamma_{ij}^l \in [0,1]$,这些条件共同定义 α_{ij}^l 、 β_{ij}^l 、 γ_{ij}^l 的取

值。公式(2)中各层特征分别与权重参数相乘后再求和,需要保证各相加项的尺度和通道数一致,根据具体需求对各特征层进行上采样或下采样操作,并

将各特征层通道数调为一致以满足条件。例如将 $level - 1$ 特征图调整至 $level - 2$ 的特征图尺寸。对于上采样, 先用 1×1 卷积操作调整通道数, 再通过插值调整尺寸。对于下采样, 则是通过 3×3 大小、步长为 2 的卷积操作完成。另外, 对于 $level - 3 \sim level - 1$, 则是在此基础上加上最大池化层完成跨级下采样。对于权重参数, 则是通过尺寸调整后的特征图经过 1×1 卷积获取, 因而可以通过基本反向传播来学习及更新参数。利用此方法, 在每个尺度特征输出上自适应地对各个级别的特征进行聚集融合, 各通道的输出分别用于 YOLOv4 的对应通道的对象检测。

3 实验设置

3.1 数据集准备

实验使用图像由上海地铁九号线提供, 正常及丢失螺栓这 2 个检测目标分别为正负样本。先将数据集通过图 6 流程实现扩充, 再使用黑色涂抹螺栓区域模拟丢失螺栓以解决负样本稀少问题。图像由定点相机拍摄, 转向架在图像中位置不变且上下部位均是背景区域, 在训练过程中将此区域裁剪, 原图的尺寸大小变为 $2\ 000 \times 1\ 400$, 裁剪后可减少冗余计算。共采集图像 1 269 张, 其中正常螺栓 4 155 个, 丢失螺栓 311 个; 经过数据处理后达 3 500 张, 其中正常螺栓 14 938 个, 丢失螺栓 1 526 个。

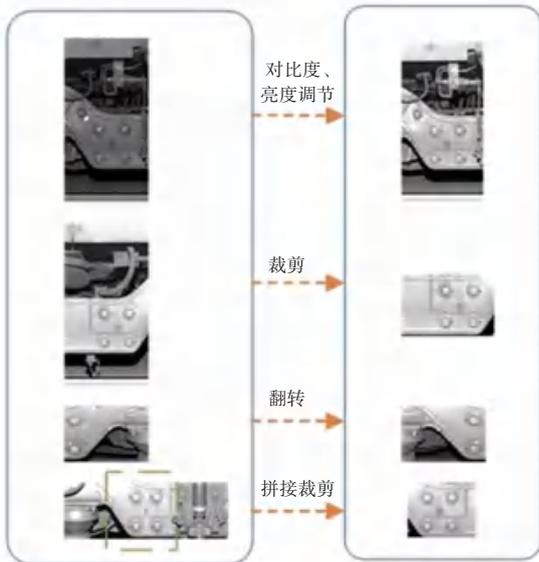


图 6 数据集扩增与预处理

Fig. 6 Dataset amplification and pretreatment

3.2 实验环境与网络训练

硬件配置: CPU 为 Core i7-8700, GPU 为 RTX 2080, RAM 32 G; 软件环境为 python 3.6, 训练框架为 Keras 2.1.5 + Tensorflow 1.3.1, OS 为 Windows 10, 结合

OpenCV 等第三方库, CUDA 10.0 加速。

输入尺寸为 416×416 , 设置 150 次 $epoch$, 采用分阶段迁移训练策略, 先冻结前层参数完成 70 个 $epoch$ 训练, 解冻再训练剩余 $epoch$, $batch\ size$ 、初始学习率、动量分别设置为 6、0.000 1 及 0.9。采用批梯度下降策略调整学习率, 权重衰减值为 0.000 01。数据集按 9:1 的比例划分为训练集、测试集, 另留 589 张图像作后续测试集, 包含正常螺栓 1 495 个, 丢失螺栓 434 个。

3.3 评价指标

采用平均精确度、平均精度均值、召回率、预测时长、每秒传输帧数, 即 FPS 作为性能评价指标, 计算公式具体如下:

$$P = TP / (TP + FP) \quad (5)$$

$$R = TP / (TP + FN) \quad (6)$$

$$AP = \int_0^1 P(r) dr \quad (7)$$

$$mAP = \frac{1}{m} \sum_1^m AP \quad (8)$$

采用训练时长作为模型训练开销的评价指标, 时长越短说明模型训练越高效。

4 实验对比与分析

图 7 为改进 YOLOv4 算法训练过程中 $Loss$ 变化曲线。设置迭代 150 轮 $epoch$, 第一轮训练结束损失值近 680, 随后剧烈下降, 在 20 ~ 80 世代缓慢下降。冻结前层网络经过 70 个 $epoch$ 后, 解冻训练 $Loss$ 值再次下降, 最终稳定于 10, 观察曲线可知 $epoch$ 训练收敛理想。

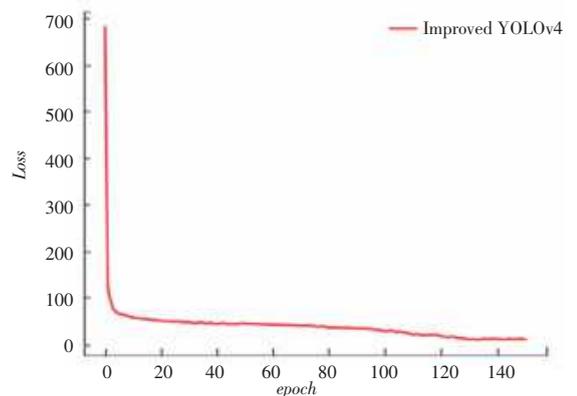


图 7 $Loss$ 值收敛情况

Fig. 7 Convergence of the $Loss$ value

为剖析对比改进措施对原算法的影响状况, 采用消融实验方法将原算法结合不同改进措施验证性能, 裁剪成 4 组算法分别进行训练及测试, 结果见表 1。表 1 中, “√”表示相应改进措施被采用。

表1 不同算法改进措施的对比

Tab. 1 Comparison of improvement measures of different algorithms

算法改进	YOLOv4	先验框聚类 与设定	先验框的 回归改进	多尺度特征 融合模块	二级级 联检测	mAP / %	$recall$ / %	FPS / 帧	训练时长/h
原算法	√					95.4	97.7	26	5.5
改进算法1	√	√				95.8	98.6	37	4.9
改进算法2	√	√	√			95.5	98.3	39	4.5
改进算法3	√	√	√	√		99.1	98.8	31	5.3
改进算法4	√	√	√	√	√	99.5	99.8	25	6.5

由表1可知,改进算法1先根据 $SSE - k$ 确定 k 值,再通过聚类算出尺寸,其训练时长和 FPS 有明显改善,时长缩短超 0.5 h, FPS 从 26 帧加速到 37 帧,结果说明依据检测目标尺寸调整先验框对算法速度有提升效果。改进算法2改进先验框回归策略,进一步加快训练和预测速度, mAP 和 $recall$ 值无明显变化。算法3加入 ASFF 模块,针对不同尺度特征的不一致性自适应地学习获取融合所需用的权重参数,以解决空间矛盾问题, mAP 提升了 3.7%, $recall$ 值提升了 1.1%,证明 ASFF 模块的加入对检测效果有重大影响, FPS 也有小幅提升,训练时长有小幅缩短,在算法体量增加的情况下速度仍有较好表现,有效提升检测效果。算法4采用级联的方法分2步检测螺栓目标, mAP 提升较小, $recall$ 值再次得到 1% 的提升,误检率得以降低。

进一步验证改进算法的性能,分别对原算法和改进算法进行不同 $epoch$ 轮数和不同数量训练集的

训练及验证,结果见图8。图8中,柱状图代表原算法 YOLOv4,折线图代表改进 YOLOv4,迭代轮数划分为 70、90、110、130、150,训练集划分为 700、1 000、1 750、3 500。图8(a)中,随着迭代轮数递增,2种算法 mAP 值均在提高,总体上改进的 YOLOv4 算法 mAP 均高于原算法,且在早期第 70 轮 $epoch$ 时已达到 90% 以上的 mAP 。可见算法的改进能够更高效地提取及融合特征,加快特征学习,即在较少迭代的情况下也能较快达到比原算法更高的检测效果。图8(b)与图8(a)趋势相似,随着迭代轮数递增, mAP 值皆随之递增,且改进算法 mAP 均高于原算法,对于小样本训练集的学习,相较于原算法也具有更好的检测结果,对 700 个图像的学习 mAP 已能达到 87.6%,高出原算法 7.9%,再次证明改进算法对图像特征学习的高效性及强大的泛化能力,综上说明算法的改进具有重要的性能影响。

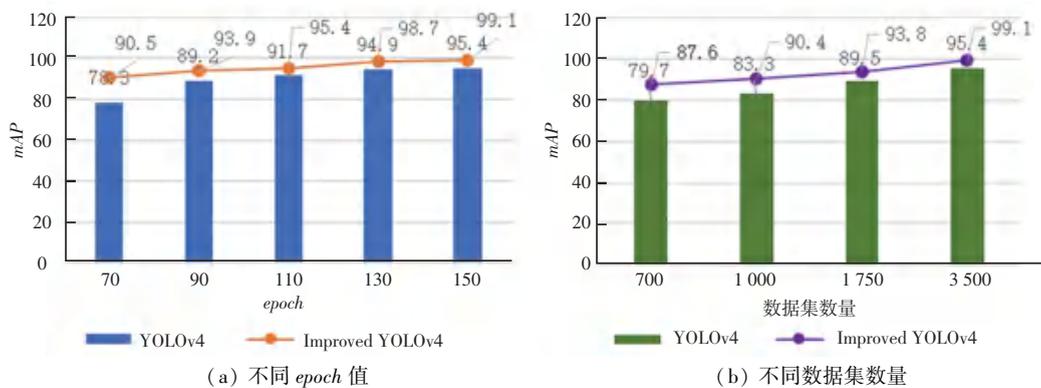


图8 原算法与改进算法的检测结果对比

Fig. 8 The detection results of the original algorithm and the improved algorithm

近年来存在的热门目标检测模型较多,针对上述验证集本文将改进的 YOLOv4 算法与其他主流算法进行了对比,结果详见表2。由表2可知,算法 Faster-RCNN 和改进 YOLOv4 的 mAP 与 $Recall$ 值相较于其他检测算法都表现出明显优势,分别达到了 99.6% 和 99.1%,但前者的检测速度最慢,而改进的算法对比原算法快出 5 帧/s,超出了 Faster-RCNN

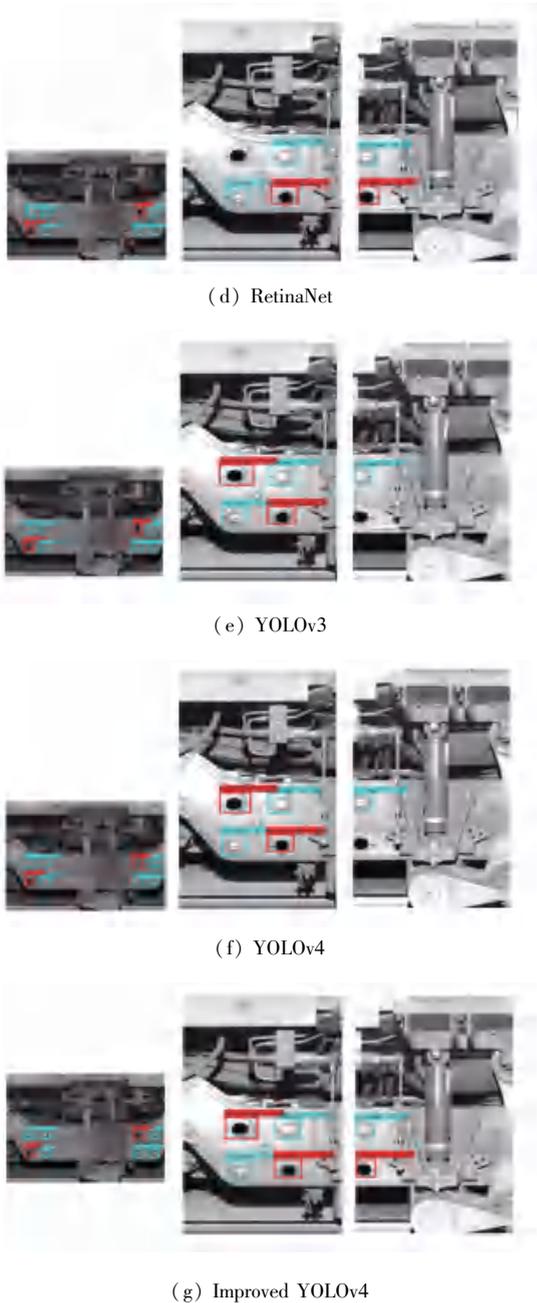
的 4 倍,检测速度表现突出,其训练时长及算法存储量相较于其他算法也处于适中水平。SSD 训练所需时长较短,但其算法存储量与 RetinaNet^[17] 都相对较大,检测性能分别低于改进 YOLOv4 的 6.5%、5.8%。观察表2数据可知,改进 YOLOv4 综合水平在原算法的基础上提升明显,其检测效果出色,性能水平较高,所以算法改进的优势进一步得以验证。

表 2 各主流算法检测结果

Tab. 2 Detection results of various mainstream algorithms

算法	mAP/ %	Recall/ %	FPS/ 帧	训练时长/h	预测时长/ms	算法存储量/KB
SSD	92.6	94.5	15	3.8	85	1 746
Faster-RCNN	99.6	99.7	7	6.2	160	675
RetinaNet	93.3	96.8	17	4.7	78	2 202
YOLO-v3	91.3	97.6	28	5.5	35	1 041
YOLO-v4	95.4	97.7	26	5.5	38	719
改进 YOLO-v4	99.5	99.8	25	6.5	38	835

图 9 展示了各主流算法的实测对比图,首列为人工手动图像标注。算法在训练时模型学习能力弱、对图像特征提取不够充分等性能问题都会导致目标误漏检测情况。参见图 9(b)可以看到,SSD 和 RetinaNet 存在丢失螺栓的漏检问题,参见图 9(c)中可知,螺栓显示不完整,这也分别表明 YOLOv3 和 YOLOv4 出现了漏检问题。观察改进 YOLOv4 算法表现,其目标检测、且定位精准,不完整目标也可得到正常高置信度识别,整体检测性能优于其他检测算法。



(d) RetinaNet

(e) YOLOv3

(f) YOLOv4

(g) Improved YOLOv4

(a) 人工标注图

(b) SSD

(c) Faster-RCNN

图 9 各主流算法实测效果对比

Fig. 9 Comparison of results of various mainstream algorithms (下转封三)