

YOLOv5改进算法在机械零件中的识别与应用

张浩洋¹, 何仕荣^{1*}, 孟冬平^{1,2}

¹上海理工大学机械工程学院, 上海

²上海理工大学理学院, 上海

收稿日期: 2022年11月22日; 录用日期: 2022年12月22日; 发布日期: 2022年12月31日

摘要

针对智能生产制造过程中, 传统的目标检测算法对机械零件识别率不高, 识别速度慢等问题, 结合深度学习与现有算法, 提出一种基于YOLOv5改进的目标检测算法。该算法在YOLOv5骨干网络中加入注意力机制, 以用来改善原始YOLOv5算法对相似背景下相似零件识别率低的问题; 其次, 通过引入CIou损失函数, 使得改进后的YOLOv5算法能够更快地收敛并具有更好的性能; 最后, 将改进前后的算法模型分别训练后对机械零件进行识别并对比分析, 证实改进后的YOLOv5算法具有高的识别精度和鲁棒性。

关键词

深度学习, 骨干网络, YOLOv5, 机械零件识别

Identification and Application of YOLOv5 Improved Algorithm in Mechanical Parts

Haoyang Zhang¹, Shirong He^{1*}, Dongping Meng^{1,2}

¹School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

²College of Science, University of Shanghai for Science and Technology, Shanghai

Received: Nov. 22nd, 2022; accepted: Dec. 22nd, 2022; published: Dec. 31st, 2022

Abstract

Aiming at the problems such as low recognition rate and slow recognition speed of mechanical parts in the process of intelligent production and manufacturing, an improved object detection algorithm based on YOLOv5 was proposed by combining deep learning and existing algorithms. This algorithm adds an attention mechanism into the YOLOv5 backbone network to improve the low recognition rate of similar parts under similar backgrounds of the original YOLOv5 algorithm;

*通讯作者。

Secondly, by introducing Clou loss function, the improved YOLOv5 algorithm can converge faster and have better performance; Finally, the algorithm models before and after the improvement are trained respectively to identify the mechanical parts and make a comparative analysis. It is confirmed that the improved YOLOv5 algorithm has high recognition accuracy and robustness.

Keywords

Deep Learning, Backbone Network, YOLOv5, Machine Part Identification

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

传统制造业往往利用人工进行机械零件的装配，一方面人工劳动强度大，企业成本高；另一方面因为工人参差不齐的水平、长时间工作后产生的视觉疲劳会造成机械零件的误识别，这严重影响了制造速度与装配精度[1]。随着社会的发展，智能制造已经成为中国制造强国战略的主攻方向，智能制造过程中往往存在大量的机械零件装配过程[2]。机械零件的快速识别与正确定位是产品装配质量的根本保障。智能制造过程中利用机械臂与计算机视觉系统相结合的方法能够显著提高装配速度，降低制造成本，决定识别速度与准确率的关键是机械零件目标检测算法，为此必须设计一种精度与速度并存的目标检测算法。

现有的目标检测算法一般分为两类，一类是以方向梯度直方图(Histogram of Oriented Gradient, HOG)算法[3]、尺度不变特征变换(Scale-invariant feature transform, SIFT)算法[4]、局部二值模式(Local Binary Patterns, LBP)算法[5]为代表且使用支持向量机[6] (Support Vector Machine, SVM)、迭代[7] (AdaBoost)分类器进行回归预测的算法；一类是利用多层卷积神经网络对图片信息进行特征提取的基于深度学习的目标检测算法，它以 R-CNN [8]、Fast R-CNN [9]、Faster R-CNN [10]、YOLO [11]，SSD [12]算法为代表。例如，Yan [13]等人将 YOLOv5 用于苹果采摘识别，通过替换 YOLOv5 算法中的 BottleneckCSP 模块并改进 YOLOv5 的初始锚框，改进后的模型能够快速识别被树叶遮挡或被其它苹果遮挡的可摘苹果，提高了算法对苹果的识别精度，但该改进措施只能用于识别苹果，局限性较大；Xi [14]等人将 YOLOv5 应用到齿轮点蚀的测量系统，通过结合 YOLOv5 和 Deeplabv3+，实现了实时在线检测齿轮点蚀率的精准测量，但该算法参数量太大，造成训练时间过长；杨琳[15]等人将 Canny 边缘提取算子加入 YOLOv4 网络结构，使原有预测边界矩形框改进为凸包和最小外接矩形框，提高了机械零件定位精度，但该模型训练后的权重文件比较大，不利于模型在终端的部署；余永维[16]等人将 Inception 网络结构引入到 SSD 网络结构中，并使用标准化模块(BN)与残差结构连接，将算法的鲁棒性与检测精度提高，但却牺牲了检测速度。

针对以上问题，本文通过对现有的 YOLOv5 算法进行改进，并在自己制作的机械零件数据集完成相关训练与测试，实现目标的初步识别。经过不断地训练与调试，保留最优权重，实现该算法对机械零件轮廓与位置的精准捕获与识别。

2. YOLOv5 原理及模型架构

YOLOv5 算法无需复杂的设计过程，可以直接选择从输入图片中提取的整个特征图进行模型的训练，从而更好地实现对目标和背景区域的精准划分并能够同时预测多个目标检测框的位置及其类别。此外，在 YOLOv5 的 4 个模型中，如表 1 所示，YOLOv5s 网络结构较简单、推理及运算时间最小、易于在各种

移动设备上部署、且具有较高的定位精度以及最快的检测速度。因此，为兼顾推理时间与推理精度，本文选用 YOLOv5s 作为目标检测的主要网络。

YOLOv5s 单阶段目标检测网络是在经历了 YOLO、YOLOv2、YOLOv3、YOLOv4 四个阶段的不断优化得到的，其网络结构如图 1 所示。在 Backbone 结构中，YOLOv5 替换了 YOLOv3 的主干网络 Dark-net53，将尺寸为 $640 \times 640 \times 3$ 的图像作为输入，经过切片、拼接函数(Concat)拼接，提高特征提取能力，YOLOv5 设计 SPP 模块，该模块借鉴了来自 2018 年 CVPR 上主要应用于图像分割的(Path Aggregation Network, PANet)。其中，(Feature Pyramid Networks, FPN)层自顶向下传达强语义特征，而特征金字塔则自底向上传达强定位特征，两者结合，从不同的主干层对不同的检测层进行参数聚合，进一步提高了特征提取的能力，SPP 模块为空间池化金字塔模块，其采用分别采用 5/9/13 的最大池化操作，再进行 Concat 融合，提高感受野；在 Neck 结构中，YOLOv5 设计 CSP2 结构，提高了算法模型的特征融合能力；在 Prediction 结构中，采用 3 种不同的感受野来识别不同尺寸的目标。

Table 1. Comparison of four YOLOv5 models
表 1. YOLOv5 四种模型对比

Model	Size	AP ₅₀	FPS	Params
YOLOv5x	640	68.7	167	87.7M
YOLOv5l	640	66.4	264	47.0M
YOLOv5m	640	63.1	345	21.4M
YOLOv5s	640	55.6	455	7.3M

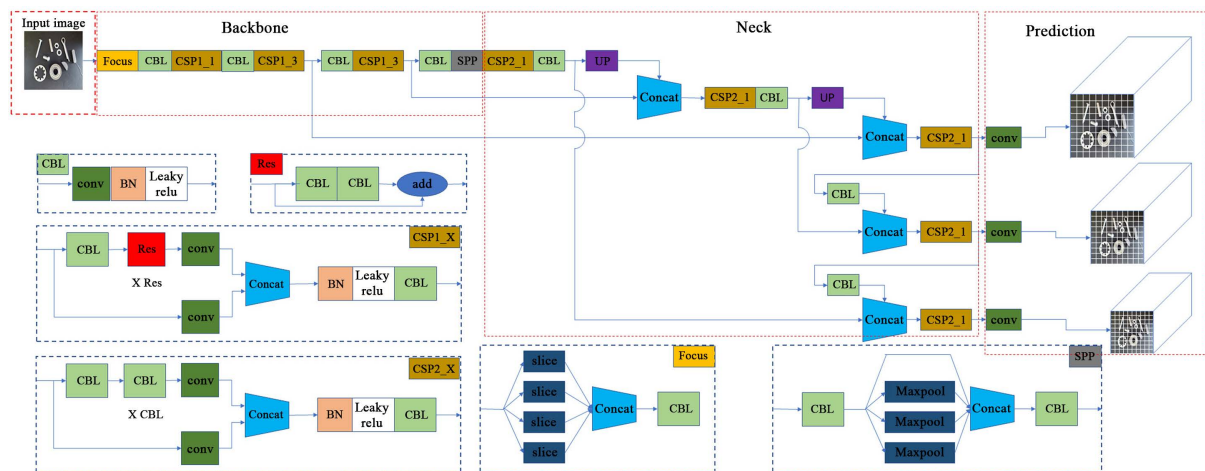


Figure 1. YOLOv5s network structure
图 1. YOLOv5s 网络结构图

3. 改进 YOLOv5 识别模型

3.1. Backbone (骨干网络)添加注意力机制

由于不同种类的机械零件形状相似，造成原始 YOLOv5 在实际检测过程中极容易出现漏检、误检等问题。为了改善原始算法的不足，提出在 YOLOv5 骨干网络部分添加(Convolutional Block Attention Module, CBAM)注意力机制[17]的策略来改进以上不足，CBAM 的原理为在给定的中间特征图的前提下，将特征通道与特征空间两个维度相结合构成注意力机制，然后将注意力聚焦的范围信息与输入的特征图相乘，

进行自适应的特征细化，其结构如图 2 所示。

CBAM 通过通道模块和空间模块两个顺序子模块能够让每个卷积块自适应的细化提取到的中间特征图，以便每个分支都能够分别学习到在通道和空间上要关注的“是什么”和“在哪里”的信息，并能够通过此模块可以学习要强调和抑制哪些信息，有效的帮助信息在网络机构中传递。CBAM 的总体计算过程可以用下式概括：

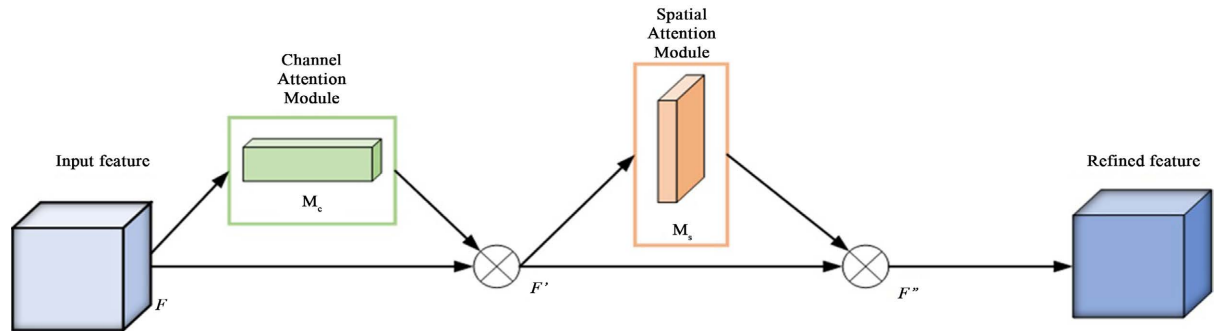


Figure 2. CBAM structure
图 2. CBAM 结构

$$F' = M_c(F) \otimes F \quad (1)$$

$$F'' = M_s(F') \otimes F' \quad (2)$$

式中： F 指的是中间特征图， $F \in R^{C \times H \times W}$ 且 F 作为输入； M_c 指的是 1 维通道注意力特征图，且 $M_c \in R^{C \times 1 \times 1}$ ； M_s 指的是 2 维空间注意力特征图，且 $M_s \in R^{1 \times H \times W}$ ； \otimes 指的是两个矩阵对应位置元素进行乘积，在相乘期间，通道注意力值可以沿着空间维度进行传播，空间维度也可以沿着注意力值进行传播； F' ， F'' 分别为中间输出和最终输出。本文改进后的 Backbone 结构如图 3 所示。

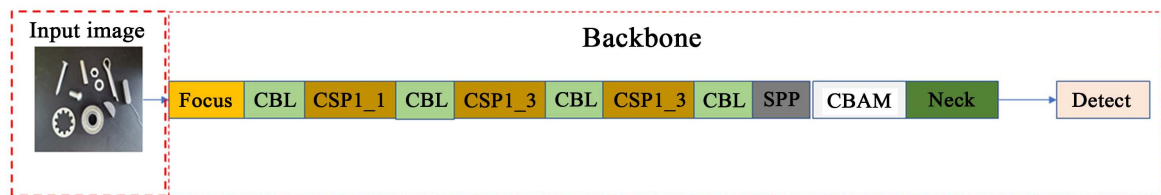


Figure 3. Backbone network structure with CBAM attention mechanism embedded
图 3. 嵌入 CBAM 注意力机制后的骨干网络结构

3.2. 构建边界损失函数

为了使机械零件检测模型有更快的收敛和更好的性能，此模型边界损失函数采用 $CIoU$ 损失函数， $CIoU$ 损失函数在 $DIoU$ 损失函数的基础上将宽高比吸纳进来。 $CIoU$ 损失包括重叠面积、中心点距离和宽高比三项，其定义如下：

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (3)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (4)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (5)$$

式中: α , v 是控制预测检测框的宽高且能够尽可能快速地与真实检测框的宽高接近的影响因子; w^{gt} , h^{gt} 分别代表真实检测框的宽和高; w , h 分别代表预测检测框的宽和高。

$CIoU$ 损失函数计算过程如图 4 所示, 其中左下角虚线框代表预测检测框, 大实线框代表真实检测框, 同时也是最小边界矩形框; 最小边界矩形框的对角线长度为 A , 虚线框与实线框之间的欧氏距离为 B 。

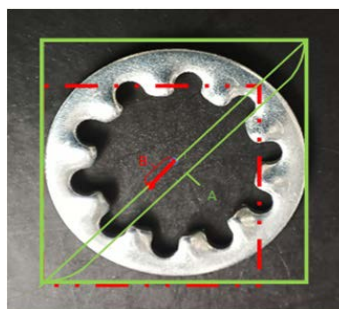


Figure 4. Normalized distance between real box and prediction box

图 4. 真实框与预测框之间的归一化距离

4. 相关实验设置

4.1. 机械零件数据集制作及其划分

机械零件目标检测模型属于深度学习网络, 欲使该检测模型具有较高的精度与鲁棒性, 必须有充足的数据样本作为支撑。而目前在机械零件识别方面并没有如 ImageNet、CoCo、Pascal-Voc 等公共数据集。因此, 必须建立一套用于机械零件检测的数据集, 本文所建立的数据集包括 Bearing (轴承)、Bolt (螺栓)、Flat-key (平键)、Lock-washe (锁紧垫片)、Nut (螺母)、Round-pin (圆柱销)、Screw (螺钉)、Shim (垫片)、split-pin (开口销)、woodruff-key (半圆键) 10 类机械零件、共 9000 张零件图像, 20,000 个目标零件(共 10 类零件, 每类 2000 个), 随机选取 5400 张图片作为训练集; 随机选取 1800 张图片作为验证集; 随机选取 1800 张图片作为测试集; 其中一张图像包含的目标种类有以下四种情况, 如图 5 所示: a 有一类零件, b 有两类零件, c 有 5 类零件, d 有 10 类零件。通过不同位置、不同姿态、不同角度模拟实际生产、装配、分拣过程中的零件情况。



Figure 5. Machine parts data set

图 5. 机械零件数据集

4.2. 数据集标注

为将图片送入算法模型中进行训练，需要对图片中的零件进行位置和种类的标注，为兼顾标注速度与准确性，采用 labelImg 软件进行零件的标注，标注后的图片如图 6 所示。



Figure 6. Annotated partial images
图 6. 标注后的部分图片

4.3. 实验平台配置

机械零件目标检测模型的实验平台配置如表 2 所示。

Table 2. Experimental platform configuration
表 2. 实验平台配置

名称	配置参数
CPU	Intel [®] Core [™] i5-10400F CPU @2.90GHz
内存	16G
GPU	NVIDIA GeForce RTX3060
GPU 加速库	CUDA 11.3
深度学习框架	Pytorch
编程环境	Python

4.4. 模型训练参数及评估指标

4.4.1. 模型训练参数

为了训练模型，并提高模型的准确率及鲁棒性，采用表 3 所示的参数进行网络训练。

Table 3. Training parameters of the model
表 3. 模型的训练参数

参数名称	设置数值
Learningrate	0.001
Momentum	0.937
Weightdacay	0.0005
Batch size	16

4.4.2. 模型评估指标

为评估修改后的 YOLOv5 算法模型的识别结果, 本文引入以下几个参数, 分别为每类平均准确率 (Average Precision, AP)、总类别平均准确率 (mean Average Precision, mAP)、精确度 (Precision, P)、召回率 (Recall, R), 其计算如下所示:

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

$$AP = \int_0^1 P(R) dR \quad (8)$$

$$mAP = \frac{1}{|Q_R|} \sum_{q=Q_R} AP(q) \quad (9)$$

式中: (Truth Positive, TP) 为正确预测的正样本; (False Positive, FP) 为错误预测的负样本; (False Negative, FN) 为错误预测的负样本; q 为零件种类数, 此时 $q = Q_R = 10$ 。

5. 实验结果分析

损失函数 (Loss Function) 的大小与收敛速度代表了模型在训练后是否具有高性能, Loss 值越低, 趋于收敛值越快, 说明网络模型训练的越好。修改后的 YOLOv5 模型与原始 YOLOv5 模型训练过程中的 Loss 值变化趋势如图 7 所示。从图中可以看出, 修改后的 YOLOv5 模型在训练开始阶段已经低于原始模型, 差值为 0.03, 随着训练 epoch 的增加, 新模型相比原始模型能够快速趋于稳定且最终的 Loss 收敛值为 0.008, 趋于稳定的 Loss 值比原始模型小 0.002, 证明了修改损失函数能够增加模型的性能。

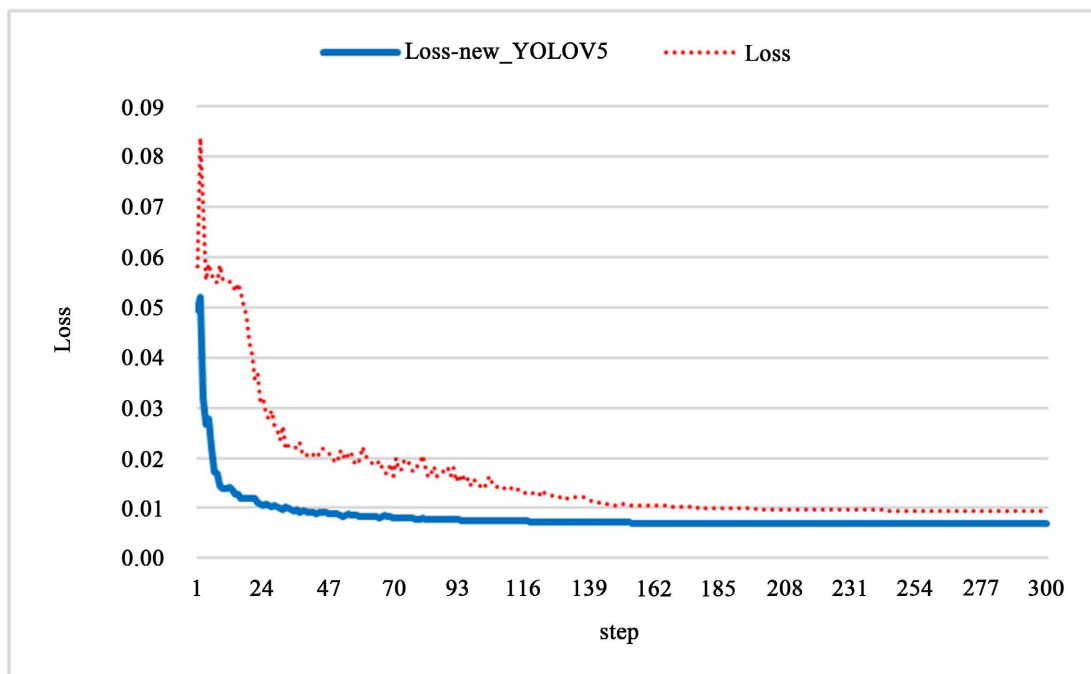


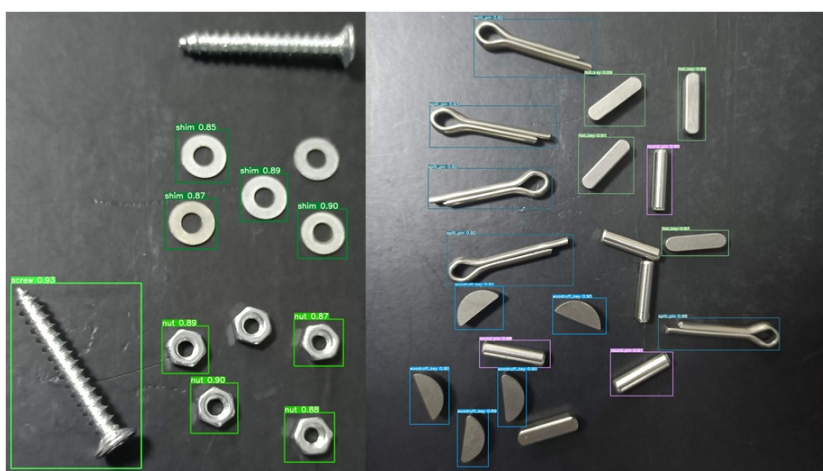
Figure 7. Change of loss function of different models during training

图 7. 训练过程不同模型损失函数变化

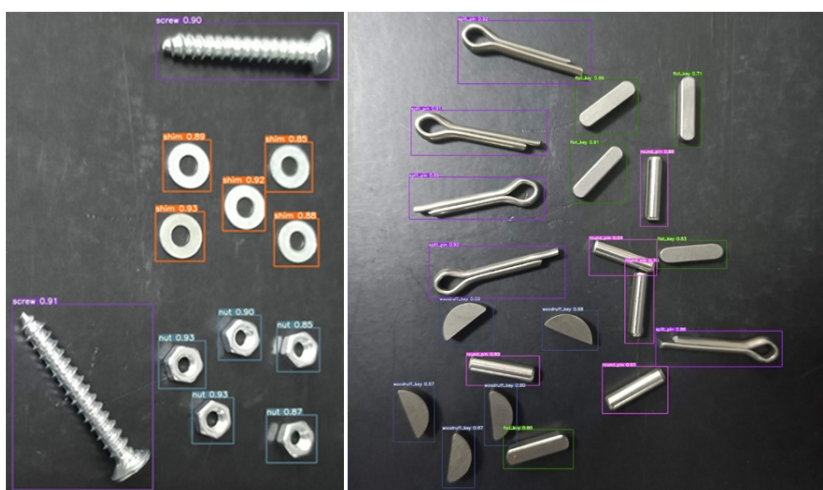
检测精度 AP 、 mAP 能够直观的评估模型检测效果的好坏，其值越高，检测效果越好，两种模型的检测结果如表 4 所示。

Table 4. Comparison of identification accuracy
表 4. 识别精度对比

模型算法	AP (%)										mAP (%)
	Bear-ing	Bolt	Flat-key	Lock-washe	Nut	Round-pin	Screw	Shim	Split-pin	Woodruff-key	
原始 YOLOv5	92.8	77.9	85.2	93.2	77.8	80.2	82.2	83.4	88.0	84.3	84.5
改进 YOLOv5	92.8	78.8	87.2	93.2	82.6	83.1	86.2	86.3	88.8	84.3	86.3
增加值	0	0.9	2	0	4.8	2.9	4	2.9	0.8	0	1.8



(a)



(b)

Figure 8. Comparison of model detection effects of the two algorithms. (a) Detection effect of the original YOLOv5 algorithm; (b) Detection effect of the improved YOLOv5 algorithm

图 8. 两种算法模型检测效果对比。(a) 原始 YOLOv5 算法检测效果；(b) 改进 YOLOv5 算法检测效果

从表 4 可以看出,改进后的 YOLOv5 模型相较于原始模型在 10 类机械零件的检测与识别中均有不同程度的提升,其中新模型在 Nut、Screw 类中提升最大,分别提高了 4.8%、4%;由于此类零件外形相似,因此原始 YOLOv5 无法提取更多的特征细节。在引入 CBAM 注意力机制并修改损失函数的改进算法中能够提取到更多的关于 Nut、Screw 的纹理及深层次的特征图,因此 AP 及 mAP 均有提升,更进一步证实了改进 YOLOv5 算法的有效性。两种算法模型实际检测机械零件的效果对比如图 8 所示:对比(a)(b)可以明显看出,原始 YOLOv5 识别结果中 Screw、Shim、Nut、Flat-key、Round-pin 未被检测出;而改进后的 YOLOv5 能够准确识别出(a)图中漏检部分,再次证明了改进算法的先进性与有效性。

6. 结论

本文针对智能制造图像识别中传统识别算法存在识别准确率低、鲁棒性差的问题,改进了 YOLOv5 原始算法并设计了新的算法模型。改进的思路主要有以下部分:首先在原始检测算法模型的骨干部分加入 CBAM 注意力机制,以改善原始模型提取冗余特征造成的识别率低等问题,使得特征提取模块能够自动学习并更多地关注对准识别零件有用的特征,提高改进算法对机械零件的识别率;其次,将原始模型的 DIoU 损失函数替换为 CIoU 损失函数,进一步优化了锚框框选目标的准确性。实验对比表明,修改后的 YOLOv5 算法较原始模型在识别率及模型鲁棒性方面均有较高提升。

参考文献

- [1] 尹旭悦, 范秀敏, 顾岩, 等. 动态视觉手势识别下手工装配时序控制的智能防错方法[J]. 计算机集成制造系统, 2017, 23(7): 1457-1468.
- [2] 周济. 智能制造是“中国制造 2025”主攻方向[J]. 企业观察家, 2019(11): 54-55.
- [3] Dalal, N. and Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, San Diego, 20-26 June 2005, 886-893.
- [4] Lowe, D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- [5] Ojala, T., Pietikainen, M. and Maenpaa, T. (2002) Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 971-987. <https://doi.org/10.1109/TPAMI.2002.1017623>
- [6] Bin, G., et al. (2015) Incremental Support Vector Learning for Ordinal Regression. *IEEE Transactions on Neural Networks & Learning Systems*, **26**, 1403-1416. <https://doi.org/10.1109/TNNLS.2014.2342533>
- [7] Wen, X., Ling, S., Yu, X., et al. (2015) A Rapid Learning Algorithm for Vehicle Classification. *Information Sciences: An International Journal*, **295**, 395-406. <https://doi.org/10.1016/j.ins.2014.10.040>
- [8] Girshick, R., Donahue, J., Darrell, T., et al. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation Tech Report (v5). 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [9] Girshick, R. (2015) Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 11-18 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [10] Ren, S., He, K., Girshick, R., et al. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [11] Redmon, J., Divvala, S., Girshick, R., et al. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [12] Liu, W., Anguelov, D., Erhan, D., et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV*, Springer, Cham, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [13] Yan, B., Fan, P., Lei, X., et al. (2021) A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sensing*, **13**, 1619-1642. <https://doi.org/10.3390/rs13091619>
- [14] Xi, D., Qin, Y. and Wang, S. (2021) YDRSNet: An Integrated Yolov5-Deeplabv3+ Real-Time Segmentation Network

for Gear Pitting Measurement. *Journal of Intelligent Manufacturing*, 1-15.

<https://doi.org/10.1007/s10845-021-01876-y>

- [15] 杨琳, 陈赛旋, 崔国华, 等. 基于改进 YOLOv4 算法的零件识别与定位[J]. 组合机床与自动化加工技术, 2021(10): 28-32+37.
- [16] 余永维, 韩鑫, 杜柳青. 基于 Inception-SSD 算法的零件识别[J]. 光学精密工程, 2020, 28(8): 1799-1809.
- [17] Woo, S., Park, J., Lee, J., *et al.* (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., 2018 *European Conference on Computer Vision (ECCV)*, Springer, Cham, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1