

改进特征金字塔网络的Mask RCNN研究综述

李改俊¹, 韩建枫²

¹天津商业大学理学院, 天津

²天津商业大学信息工程学院, 天津

收稿日期: 2022年9月20日; 录用日期: 2022年10月18日; 发布日期: 2022年10月27日

摘要

随着计算机视觉的发展, 目标检测技术的精度成为科研人员的重要研究内容之一。目标检测分为one-stage和two-stage两种检测方法, 其中YOLO, SSD属于一阶段检测, R-CNN (Fast RCNN, Faster RCNN, Mask RCNN)属于两阶段检测。目标检测的精度依赖于特征提取的好坏, 特征金字塔是用于检测不同尺度的对象的识别系统中的基本组件。Mask RCNN是基于分割掩码区域建议卷积神经网络的两阶段目标检测算法, 精度相对较高, 本文主要从它的特征金字塔网络出发, 对近几年特征金字塔网络的改进算法进行研究。研究发现: 基于原特征金字塔网络引入新的自底向上或自顶向下侧边连接的特征融合路径可以实现对底层信息的充分利用; 通过双向、分层跳连等融合方法可以提高小目标检测的正确率。改进算法有效地提高了目标检测的精度。

关键词

目标检测, Mask RCNN, 特征金字塔, 特征融合

A Survey of Mask RCNN Research with Improved Feature Pyramid Network

Gaijun Li¹, Jianfeng Han²

¹School of Science, Tianjin University of Commerce, Tianjin

²School of Information Engineering, Tianjin University of Commerce, Tianjin

Received: Sep. 20th, 2022; accepted: Oct. 18th, 2022; published: Oct. 27th, 2022

Abstract

With the development of computer vision, the accuracy of target detection technology has become one of the important research contents of researchers. Target detection is divided into one-stage and two-stage detection methods. Among them, YOLO and SSD belong to one-stage detection, and R-CNN (Fast RCNN, Faster RCNN, Mask RCNN) belongs to two-stage detection. The accuracy of ob-

ject detection depends on the quality of feature extraction, and feature pyramid is a basic component in a recognition system for detecting objects of different scales. Mask RCNN is a two-stage target detection algorithm based on the segmentation mask area proposal convolutional neural network, with relatively high accuracy. This paper mainly starts from its feature pyramid network, and studies the improved algorithm of feature pyramid network in recent years. The research found that: based on the original feature pyramid network, the introduction of a new bottom-up or top-down side-connected feature fusion path can fully utilize the underlying information; the fusion methods such as bidirectional and hierarchical jump connections can improve accuracy rate of small target detection. The improved algorithm effectively improves the accuracy of target detection.

Keywords

Target Detection, Mask RCNN, Feature Pyramid, Feature Fusion

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

计算机视觉的快速发展, 行业对目标检测和分割的精度要求越来越高。R. Girshick 等人[1]提出基于区域的卷积神经网络(R-CNN), 在目标检测中, 将深度学习机制应用其中, 利用深度卷积网络对目标建议进行分类, 获得良好的目标检测精度, 实现目标的自适应检测。工程上大部分应用速度快的一阶段的YOLO, SSD 等算法, R-CNN 系列作为两阶段检测, 精度相对较高。但是对于 R-CNN 算法流程来说, 训练是一个多阶段的流水线, 训练过程复杂, 消耗了时间和存储空间。R. Girshick [2]提出了快速区域卷积神经网络(Fast R-CNN)方法, 引入特征金字塔网络(FPN)融合多尺度特征映射图, 优化特征提取过程。HE K M 等人[3]针对图像重复区域, 边缘粗糙的问题, 增加了 Mask 部分, 提出一种基于分割掩码区域卷积神经网络(Mask R-CNN)的算法, 且改进了 Faster R-CNN [4]中 RoI Pooling, 取消取整操作, 使用双线性插值来精确找到每一块对应的特征, 有效提高了分割的精度。

图像金字塔(也称为特征化图像金字塔) [5]是连续的像素阵列, 它是用于图像增强和对象图案化的高效工具。在最初通过手工设计来提取特征的时代, 占据重要地位。图像金字塔不是计算多尺度特征金字塔的唯一方法, 图像金字塔由多个层级, 如果去特征化它的每一个层级, 会明显受到限制而且也大大增加了推理时间。单一特征图[6]是经过卷积神经网络得到的单一尺度特征映射图, 作为预测阶段的唯一输入, 大大缩短了检测时间。金字塔特征层次结构[7]是重复使用 ConvNet 计算的金字塔特征层次结构, 类似于一个特征化的图像金字塔。特征金字塔网络是利用卷积网络特征层次结构的金字塔形状, 通过生成一个能够在任意尺度上都有较强语义信息的特征金字塔, 缩短了模型训练的时间, 可以应用于以任何图像尺度计算的特征。

2. 特征金字塔网络

Tsung-Yi Lin 等人[8]利用多尺度深度卷积网络的固有金字塔层次结构, 以边际额外成本构建特征金字塔, 开发了一种具有横向连接的自上而下架构, 用于构建所有尺度的高级语义特征图, 该架构成为特征金字塔网络(FPN), 如图 1 所示。输入图像经过卷积网络生成 {C2, C3, C4, C5} 特征映射图, 特征图尺寸分别为原图像的(1/4, 1/8, 1/16, 1/32), {P2, P3, P4, P5} 为自上而下通过横向连接特征融合得到。该过程特

征融合是简单的求和操作(红色虚线框内)。

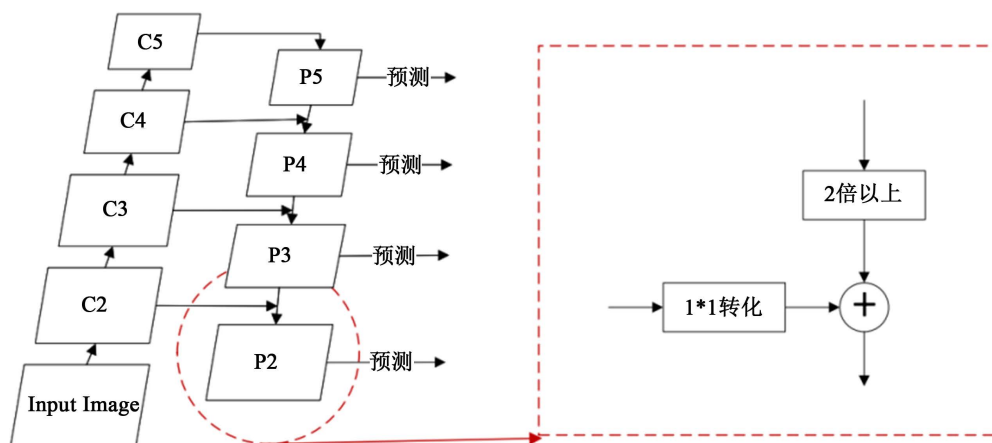


Figure 1. Feature pyramid network (FPN) structure diagram
图 1. 特征金字塔网络(FPN)结构图

3. 特征融合

在计算机识别中, 特征融合用于将多尺度特征图通过融合得到更多图像信息。输入图像在经过骨干网络生成不同尺度的特征映射图。对于底层特征图来说, 它的分辨率高, 包含丰富的细节信息, 而高层特征由于经过了多次卷积得到的, 语义信息鲜明, 图像特征点更明显。为了提高检测和分割的精确度, 人们将高层与底层特征通过融合的方式结合起来, 利用丰富的图像信息, 实现目的。值得说明的是, 根据融合和预测顺序的先后, 将融合类别分为早融合(Concat, Add)和晚融合(FPN, SSD, Densent 等)。

4. 特征金字塔网络改进研究

4.1. 基于新增侧边连接的改进

Mask RCNN 特征提取网络是由骨干网络(VGG, ResNet 等)和特征金字塔网络(FPN)组成。卷积层特征提取, 中间输出自下而上不同尺度的特征映射图, 通过特征融合与 FPN 自上而下的特征图进行横向连接, 传入 RPN 层的特征映射图则具备了底层特征的细节信息和高层特征的语义信息, 进行边框类别预测。基于此想法, 研究人员通过另外引入正向(反向)侧边连接路径来加强底层信息的利用率, 改进 Mask RCNN 算法。

1) 在 FPN 网络中新增一条自下而上的侧边连接特征融合路径[9]-[17]成为改进策略的热门选择。YOLOv4 [18]作为集成精度与检测速度一体的强大目标检测网络, 将 FPN 网络引入 YOLO 系列。YOLOv4 网络中的 PANet [19]是在 Mask RCNN 上做了多处改进, 其中最重要的就是对于特征提取过程的改进, 充分利用了特征融合。对于图像提取来说, 底层信息特征往往体现了目标的边缘形状特征, 做实例分割时, 底层信息特征非常重要。新引入侧边自下而上的横向连接特征融合路径如图 2 所示, $\{C2, C3, C4, C5\}$ 是多尺度特征映射图, 由输入图像通过卷积神经网络生成, $\{P2, P3, P4, P5\}$ 则是上一阶段的中间输出层经过卷积, 上采样, 融合得到的特征图。该路径生成的 $\{N2, N3, N4, N5\}$ 特征图由 $\{P2, P3, P4, P5\}$ 经过卷积, 下采样, 融合得到。通过新增加的路径使得 $N5$ 特征映射图充分利用 $C2$ 底层特征图上的信息, 提高图像的信息利用率。

2) 任之俊, 蔺素珍等人[20]为了提高目标的边缘检测精度, 提出了新增一条自下而上和一条自上而下的侧边连接特征融合路径的改进策略, 如图 3 所示。

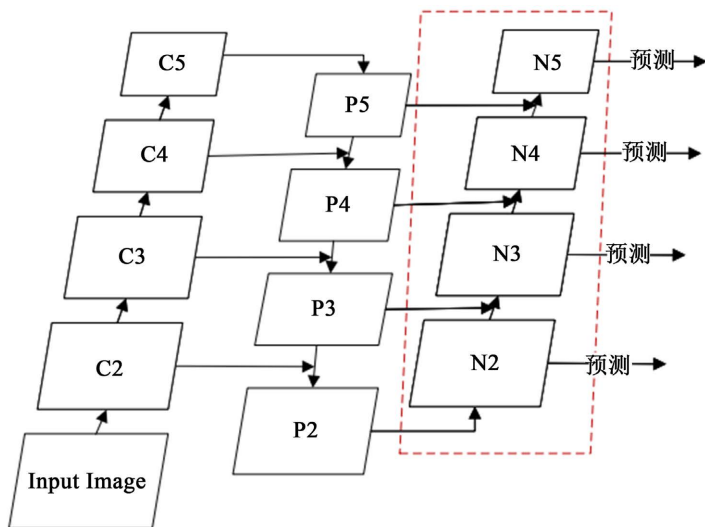


Figure 2. Add a new path to improve the FPN structure
 图 2. 新增一条路径的 FPN 改进结构图

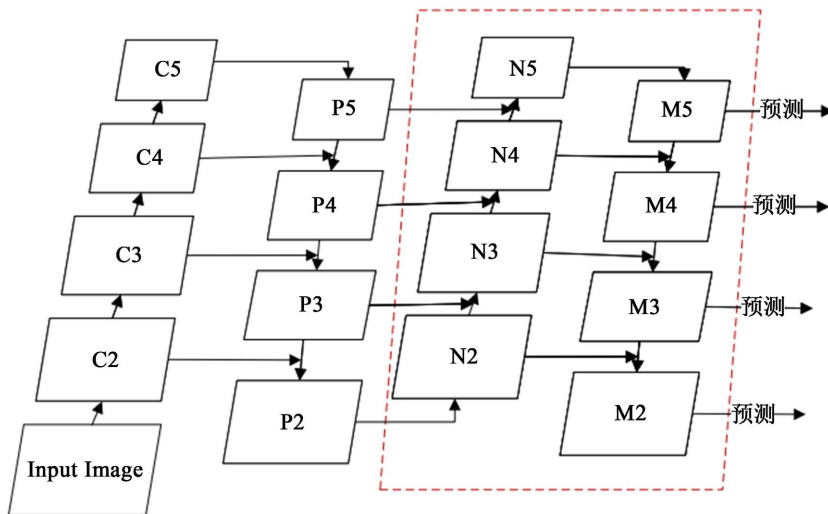


Figure 3. Add two new paths to improve the FPN structure
 图 3. 新增两条路径的 FPN 改进结构图

改进后的算法在目标识别的准确率上较原网络提高了 2.4%，在目标框检测的准确率上提高了 3.8%。由于增加两条特征融合路径，新的算法网络结构复杂，FPN 层产生了大量的冗余信息，检测速率在一定程度上有所延迟。而且新网络集中于提高对底层信息的使用，可以更好的检测到了图像中小目标，但对大目标的检测能力并没有很大提升。

4.2. 基于增加串联特征金字塔网络的改进

在进行实例分割任务时，Mask 分支主要用于处理自然场景中目标出现的重叠，遮挡，复杂等问题。但是特征提取过程中底层信息的丢失是的分割精度不高，边缘粗糙，音松等人[21]在骨干网络后面新增一个串联特征金字塔网络(CFPN)模块。CFPN 得到的特征映射图跳过 RPN 网络，直接经过 ROI Align 操作后作为 Mask 分支的输入，进行分割任务，如图 4 所示。

该改进方法得到更多的底层细节信息，使得分割对于边缘区域更敏感，准确率更高。

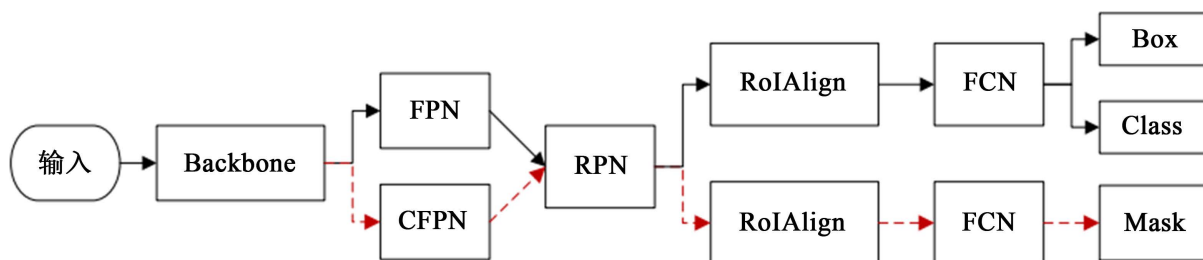


Figure 4. Flow chart of improved algorithm of Mask RCNN in series FPN

图 4. 串联 FPN 的 Mask RCNN 改进算法流程图

4.3. 基于双向融合的特征金字塔网络的改进

遥感图像具有图像清晰度低, 视野广, 内容物丰富的特点, 因此对遥感图像做目标检测和分割时, 效果往往不好。余慧明, 周志祥等人[22]提出了双向融合 FPN 层多尺度特征图的方法, 用 BiFPN 代替 FPN 如图 5 所示。图像经过骨干卷积神经网络产生 7 个多尺度特征图 {C1, C2, C3, C4, C5, C6, C7}。中间路径的 L6 特征图是由 L7 经过卷积, 上采样生成, L5 特征图由 C5 和 L6 融合产生, L4 同理。FPN 中的 {P3, P4, P5, P6, P7} 特征图由两支路径双向融合得到。

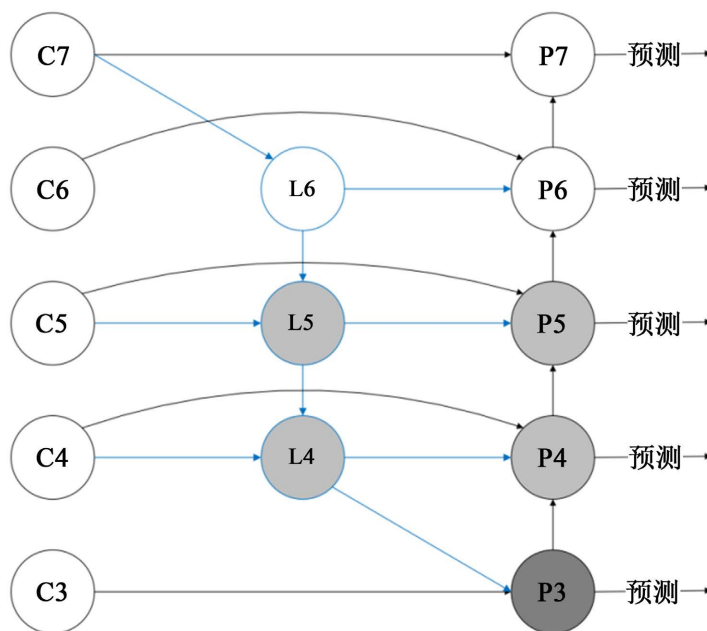


Figure 5. The structure of BiFPN

图 5. BiFPN 结构

该改进算法有效结合底层与高层信息的联系, 网络结构相对简单, 尽可能的使用经过卷积神经网络得到的多尺度特征映射图, 减少 FPN 层的冗余信息, 在各个评价指标上有一定的优越性。对于仅仅依靠增加若干条侧边连接特征融合路径来实现高低层信息结合来说, 更有效。

4.4. 基于分层跳连融合的改进

李森森等人[23]为了改善 Mask RCNN 算法在遥感图像目标检测和分割中误检率, 漏检率高的问题, 提出了分层跳连融合方式将多尺度特征图进行融合, 如图 6 所示。其中 {C1, C2, C3, C4, C5, C6} (黄色虚

线框)由{C1, C2, C3, C4, C5, C6}通过分层跳连融合所得, 彼此之间信息不进行传递。对于{P2, P3, P4, P5, P6}特征映射图来说, P2 由 C1 和 C3 经过分层跳连融合方式与 P3 进行融合所得。P3, P4 同理。P5 只经过 C4 和 C6 分层跳连融合所得。C6 与 P6 仅仅通过 $1 * 1$ 的卷积得到。

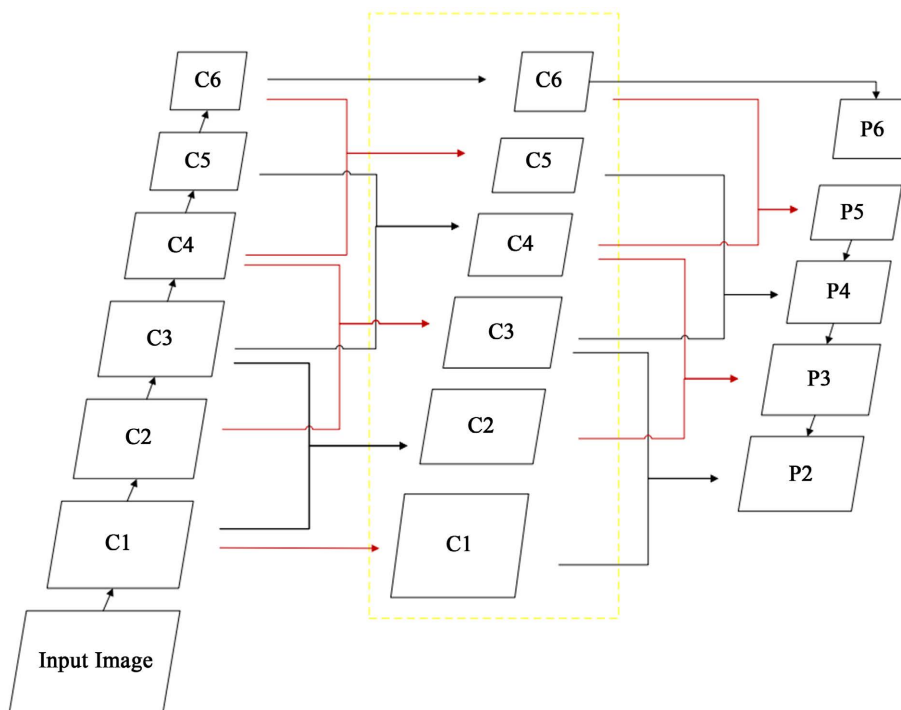


Figure 6. Hierarchical skip-connection fusion feature extraction network
图 6. 分层跳连融合特征提取网络

通过分层跳连的方式将底层特征传入高层特征, 在进行遥感图像检测和分割时, 能够在保证大目标检测的准确率同时, 充分识别到小目标。而且在平均正确率提升了 3.32% 时, 时间仅仅提升了 0.065 s。

5. 结论

特征金字塔网络在特征提取中具有重要作用, 通过侧边横向连接融合多尺度特征映射图。本文基于实例分割 Mask RCNN 的改进算法对特征金字塔网络(FPN)的发展现状进行分析。增加 FPN 侧边连接网络结构的改进, 是 FPN 改进的第一选择, 通过增加特征提取网络的复杂度, 从根本上提升了检测的精度。对于融合方式的改进来说, 主要从结构层次去设计融合路径。设计背景大部分是在对遥感、工业缺陷等复杂图像的检测上。因为对这类图像来说, 目标物分辨率低, 图像内容丰富, 依赖较精准的检测模型。融合和增加侧边横向连接等方式对特征金字塔网络的改进, 使底层特征的细节信息与高层特征的语义信息充分结合, 在小目标检测和分割的准确率方面有着很大的提升。在未来, 通过改进 FPN 的方式来优化 Mask RCNN 算法依旧是主流技术, 但是随着网络结构复杂程度的增加, 产生大量的冗余信息, 检测和分割的时间大, 这也是继续研究和改善的关键。

参考文献

- [1] Girshick, R., Donahue, J., Darrell, T., Malik, J., et al. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>

- [2] Girshick, R. (2015) Fast R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [3] He, K.M., Gkioxari, G., Dollár, P., et al. (2017) Mask R-CNN. *Proceedings of 2017 IEEE International Conference on Computer Vision ICCV*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/ICCV.2017.322>
- [4] Ren, S.Q., He, K.M., Girshick, R., et al. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Proceedings of Advances in Neural Information Processing Systems*, Montreal, 7-12 December 2015, 1-9.
- [5] Adelson, E.H., Anderson, C.H., Bergen, J.R., Burt, P.J. and Ogden, J.M. (1984) Pyramid Methods in Image Processing. *RCA Engineer*, **29**, 33-41.
- [6] He, K., Zhang, X., Ren, S. and Sun, J. (2014) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *13th European Conference Proceedings*, Zurich, 6-12 September 2014, 346-361. https://doi.org/10.1007/978-3-319-10578-9_23
- [7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C. and Reed, S. (2016) SSD: Single Shot Multibox Detector. *14th European Conference*, Amsterdam, 11-14 October 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [8] Lin, T.-Y., Dollár, P. and Girshick, R. (2017) Feature Pyramid Networks for Object Detection. <https://arxiv.org/abs/1612.03144>
- [9] 温尧乐, 李林燕, 尚欣茹, 胡伏原. 一种改进的 Mask RCNN 特征融合实例分割方法[J]. 计算机应用与软件, 2019, 36(10): 130-133.
- [10] 李梁, 董旭彬, 赵清华. 改进 Mask R-CNN 在航拍灾害检测的应用研究[J]. 计算机工程与应用, 2019, 55(21): 167-176.
- [11] 王海云, 王剑平, 张果, 欧阳鑫, 罗付华. 改进 FPN 的 Mask R-CNN 工业表面缺陷检测[J]. 制造业自动化, 2020, 42(12): 35-40+97.
- [12] 陈敏, 王君, 董明利, 燕必希, 贾欣雨. 改进的 Mask R-CNN 多尺度实例分割算法研究[J]. 激光杂志, 2020, 41(5): 40-44.
- [13] 朱繁, 王洪元, 张继. 基于改进的 Mask R-CNN 的行人细粒度检测算法[J]. 计算机应用, 2019, 39(11): 3210-3215.
- [14] 宣锦昭, 徐超, 冯博, 闪文章. 一种改进的 Mask R-CNN 图像篡改检测模型[J]. 小型微型计算机系统, 2020, 41(11): 2333-2339.
- [15] 江昆鹏, 闫洪涛, 杨红卫, 张庆辉. 改进 Mask R-CNN 的细粒度车型识别算法[J]. 软件, 2020, 41(3): 1-5.
- [16] 喻丽春, 刘金清. 基于改进 Mask R-CNN 的火焰图像识别算法[J]. 计算机工程与应用, 2020, 56(21): 194-198.
- [17] 张超, 文传博. 基于改进 Mask R-CNN 的风机叶片缺陷检测[J]. 可再生能源, 2020, 38(9): 1181-1186.
- [18] Wang, C.-Y., Bochkovskiy, A. and Liao, H.Y.M. (2021) Scaled-yolov4: Scaling Cross Stage Partial Network. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 13029-13038.
- [19] Liu, S., Qi, L., Qin, H.F., Shi, J.P. and Jia, J.Y. (2018) Path Aggregation Network for Instance Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, 18-23 June 2018, 8759-8768. <https://doi.org/10.1109/CVPR.2018.00913>
- [20] 任之俊, 蔺素珍, 李大威, 王丽芳, 左健宏. 基于改进特征金字塔的 Mask R-CNN 目标检测方法[J]. 激光与光电子学进展, 2019, 56(4): 174-179.
- [21] 音松, 陈雪云, 贝学宇. 改进 Mask RCNN 算法及其在行人实例分割中的应用[J]. 计算机工程, 2021, 47(6): 271-276+283
- [22] 余慧明, 周志祥, 彭杨, 崔志斌. 一种基于改进 Mask R-CNN 模型的遥感图像目标识别方法[J]. 信息技术与网络安全, 2021, 40(3): 38-42+47.
- [23] 李森森, 吴清. 改进 Mask R-CNN 的遥感图像多目标检测与分割[J]. 计算机工程与应用, 2020, 56(14): 183-190.