

文章编号: 1671-7449(2024)01-0054-06

基于改进YOLOv4的遥感图像目标检测算法

刘敏, 樊永生*

(中北大学 计算机科学与技术学院, 山西 太原 030051)

摘要: 针对遥感卫星图像目标尺寸较小、背景较为复杂并且分布密集的问题, 在YOLOv4算法基础上, 通过使用K-means重新聚类, 优化先验锚框的数量和尺寸, 减少网络的计算; 引入动态激活函数, 自适应网络输入特征, 提高模型的泛化能力; 改进PANet结构, 扩展了浅层特征的特征融合。实验结果表明, 改进的YOLOv4算法对遥感图像中密集排列的小目标图像识别的准确率、检测速度都有所提高, 满足了对于遥感图像小目标检测的要求。

关键词: YOLOv4算法; 遥感图像; 小目标检测

中图分类号: TP391.41

文献标识码: A

doi: 10.3969/j.issn.1671-7449.2024.01.008

引用格式: 刘敏, 樊永生. 基于改进YOLOv4的遥感图像目标检测算法[J]. 测试技术学报, 2024, 38(1): 54-59.

LIU Min, FAN Yongsheng. Research on Remote sensing image target detection algorithm based on improved YOLOv4[J]. Journal of Test and Measurement Technology, 2024, 38(1): 54-59.

Remote Sensing Image Target Detection Algorithm Based on Improved YOLOv4

LIU Min, FAN Yongsheng*

(College of Computer Science and Technology, North University of China, Taiyuan 030051, China)

Abstract: Based on the YOLOv4 algorithm, this paper optimizes the number and size of the a priori anchor frames by using K-means re-clustering to reduce the computation of the network; introduces a dynamic activation function to adapt to the input features of the network and improve the generalization ability of the model; improves the PANet structure and extends the shallow features of feature fusion. The experimental results show that the improved YOLOv4 algorithm has improved the accuracy and detection speed for the recognition of densely arranged target images in remote sensing images, which meeting the requirements for small target detection in remote sensing images.

Key words: YOLOv4 algorithm; remote sensing image; object detection of small target

0 引言

遥感图像目标检测广泛应用于智能监控、城市规划、港口管理和军事侦察等。与自然场景图

像不同, 遥感图像通常是从鸟瞰图中捕获, 这给检测任务带来了额外的挑战: 如图像由复杂的背景和许多密集的物体组成、位置分布不平衡、范围广而目标较小等问题, 使遥感图像难以检测

收稿日期: 2023-03-21

作者简介: 刘敏(1998-), 女, 硕士生, 主要从事图像处理、深度学习等研究。E-mail: 849785647@qq.com。

* 通信作者: 樊永生(1967-), 男, 教授, 硕士生导师, 主要从事图形图像与信息处理、机器学习、仿真与可视化、嵌入式系统、深度学习等研究。E-mail: fanys67@163.com。

识别。

早期的目标检测模型是由人工制作的特征提取器组成，如 Viola-Jones 检测器^[1]、方向梯度直方图 (Histogram of Oriented Gradient, HOG)^[2] 等，这些模型速度慢、准确率低，应用环境改变后效果不佳。卷积神经网络 (CNNs) 的提出改变了视觉感知。AlexNet 在 2012 年 ImageNet 大规模视觉识别挑战赛 (ILSVRC) 中的应用^[3]，激发了 CNNs 在计算机视觉中应用的进一步研究。如今，目标检测应用在自动驾驶、人脸识别等多领域^[4]。

长期以来，各行各业的技术人员都试图通过计算机模拟遥感专家，进行遥感图像识别的复杂过程^[5]。然而，他们未能实现足够高的信息提取精度和准确性。由于遥感图像的复杂性，最成功的遥感图像识别方法为视觉判读，而海量遥感数据与人类有限的识别能力之间的矛盾，制约了目前遥感服务于发展中社会潜力的发挥。

YOLOv4 算法由于其准确性和速度优势被广泛用于目标检测，有数据显示，YOLOv4 是

YOLO 系列中最佳的目标检测框架。检测遥感图像的困难在于目标较小、背景复杂、分布密集，所以应该考虑其检测速度和泛用性，本文将基于深度学习的 YOLOv4 检测算法应用于遥感目标的检测，通过对数据集进行重新聚类、重新选择激活函数并且对 PANet 结构进行优化，改进对遥感图像中小目标的识别准确率和识别效率。

1 YOLOv4 网络结构分析

1.1 YOLOv4 网络结构

YOLOv4 算法主要由三部分组成：用于特征提取的 Backbone 结构、用于提取特征语义表示的 Neck 结构和用于预测的 Head 结构，如图 1 所示。主干采用 CSPDarknet53 结构，并且应用 SPP 的思想增加感受野，采用 PANet 中的路径聚合模块作为 Neck 部分。在主干网络中使用 Mish 激活函数，之后的网络中使用的激活函数都是 Leaky_relu 激活函数。

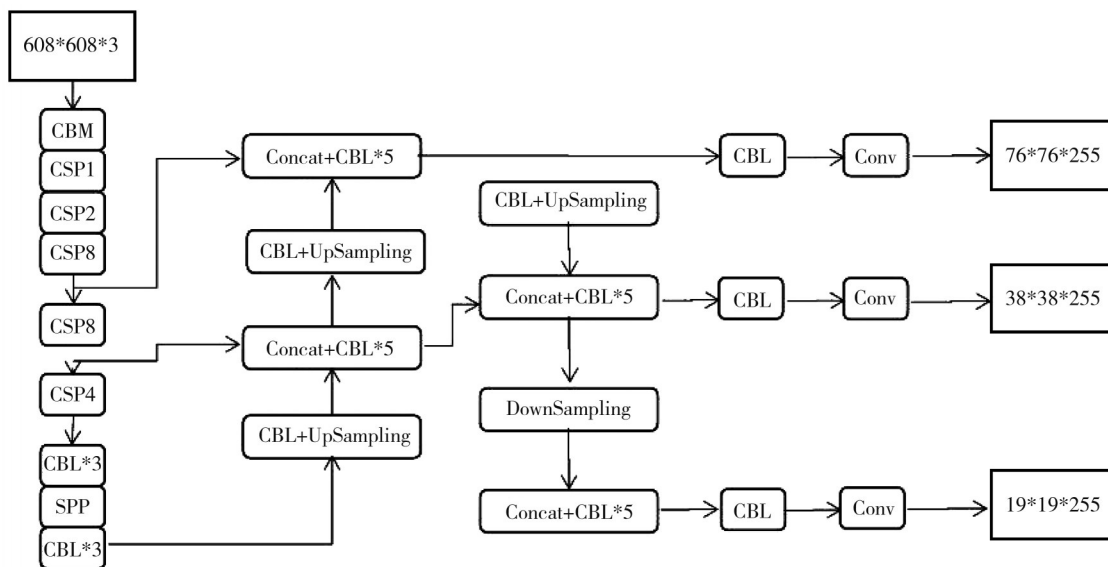


图 1 YOLOv4 网络结构图

Fig. 1 Network structure diagram of YOLOv4

1.2 YOLOv4 检测过程

YOLOv4 算法将输入图像大小调整为 608x608x3，然后放入主干网络 CSPDarknet53 结构中训练进行特征提取，得到 3 个有效特征层，大小分别为 76x76，38x38，19x19，再进入 Neck 进行处理，在 SPP 模块中进行多尺度融合，3 张特征图就是 YOLOv4 算法的输出检测结果。

YOLOv4 使用 K-means 聚类得到的先验框适用

于自然界中一般图像的目标检测，但对遥感图像小目标的检测效果不佳，因此，本文对数据集重新聚类；遥感图像的背景复杂、分布密集，而 YOLOv4 在主干中使用的 Mish 函数属于静态激活函数，不能自适应不同的输入特征，所以本文把 Mish 函数改换为动态激活函数，自适应地学习激活函数的结构，决定是否激活神经元；遥感图像的目标较小，所以特征融合方法对本文至为重要。因此，本文对 YOLOv4 算法中的 PANet 进行改进，以加强浅层特

征提取,提高对小目标的检测精度。

2 改进的YOLOv4

YOLOv4 算法是在 Microsoft Common Objects in Context (MS COCO) 数据集上训练来的,并不适用于遥感图像目标检测^[6]。所以,根据遥感图像目标的特点,对YOLOv4算法进行了以下改进,以提高遥感图像目标检测的能力。

1) 采用K-means方法对锚框数量和尺寸进行优化,以提升检测精度和速度;

2) 将YOLOv4算法主干网络中的Mish激活函数改为动态激活函数,自适应网络输入特征,以提高模型的泛化能力;

3) 对YOLOv4的Neck部分的PANet结构进行改进,丰富浅层网络,提高对小目标的检测能力。

2.1 优化锚框

YOLOv4使用的K-means聚类算法,将所有样本分成K个聚类,以欧几里得距离计算,产生有效的数据挖掘结果^[7]。距离计算为

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}, \quad (1)$$

式中: x_n 为边界框中心的标准水平轴; y_n 为边界框中心的标准数轴。不同的数据集和锚机制可能会对检测器性能产生很大影响。YOLOv4算法中使用的锚框(anchor box)尺寸主要适用于大自然中一般目标,不适用于遥感图像目标检测。因此,本文对遥感图像目标重新聚类出合适的参数,提升检测精度和速度。

IoU(Intersection over Union)值是指先验框和真实框的重合区域^[8],定义为

$$IoU = \left| \frac{A \cap B}{A \cup B} \right|. \quad (2)$$

从式(2)中可以看出,当先验框和预测框重合度越高时,IoU值越大,说明模型预测越准确;当重合度越低时,模型性能越差。如图2所示,将K的区间设置为1~9,锚框数目为3时是K的拐点,平衡了avg IoU和模型复杂度,因此,聚类中心的数量设置为3,并且在此基础上获得的增强边界框的大小分别为(11, 11)、(10, 51)、(56, 10)。

原始YOLOv4算法为3个特征图分别分配3个边界框,用于预测的尺度分别为 52×52 、 26×26 和 13×13 。 26×26 和 13×13 是语义丰富信息的深层特征,具有更大的感受野。研究表

明,有限的感受野很难检测到小目标^[9]。因此,本研究舍弃了小尺度特征图,保留了 52×52 的特征图,这也降低了网络的复杂性。

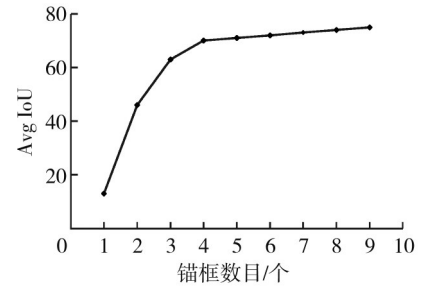


图2 不同锚框数目所对应的Avg IoU

Fig. 2 Avg IoU for different number of anchor frames

2.2 设计激活函数

原YOLOv4算法中使用的激活函数是Mish函数,属于静态激活函数,无法通过调整激活函数以适应不同输入特征来改善对复杂数据的处理。为了解决这个问题,本文引入了一个动态激活函数,通过网络自适应学习激活函数结构,决定是否激活神经元。

$$f(x) = S_{\alpha}(\epsilon_1 x, \epsilon_2 x). \quad (3)$$

网络可以通过学习 ϵ_1 和 ϵ_2 自适应调整激活函数。平滑因子 α 控制是否激活神经元。

$$\alpha = \lambda W_1 W_2 \sum_{h=1}^H \sum_{\sigma=1}^W x_{c,h,w}, \quad (4)$$

式中: W_1 和 W_2 为2个卷积层的参数。通过此方式形成新的激活层,替换原YOLOv4算法主干网络中的激活层,可以自适应不同的输入特征,使在面对背景复杂、分布密集的遥感小目标时,可以更好分开正负样本,提高了模型泛化能力^[10]。

2.3 PANet结构优化

缺乏浅层特征提取会导致小目标漏检并降低检测精度,因此,改进YOLOv4算法中的PANet以保留浅层特征信息。深度语义丰富的特征可以增强浅层特征信息^[11]。特征融合是多个不同特征信息的融合,既可以获得深层特征信息,也可以获得浅层特征信息。所以,特征融合方法在本文中至为重要。

图3为YOLOv4算法中的PANet结构,包含了浅层特征信息和深层特征信息。然而,该结构对于小目标检测有一些缺陷,例如, 26×26 直接向上采样到 104×104 可能会丢失用于预测的信息,信道数量从256减少到128也可能丢失信息。

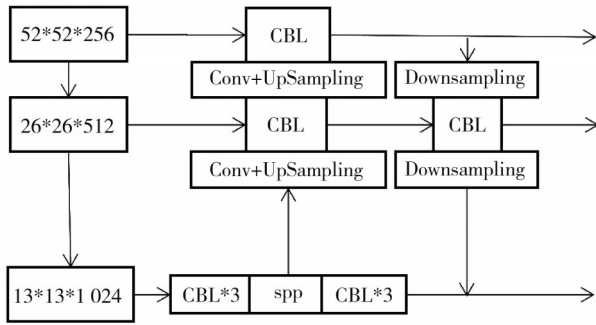


图 3 YOLOv4的PANet结构

Fig. 3 PANet structure for YOLOv4

改进的PANet结构如图 4 所示。直接将 26×26 上采样到 104×104，在 SPP 之后，分别对具有 512 个通道的 26×26、具有 128 个通道的 52×52 和具有 128 个通道的 104×104 的每个特征图进行特征融合。

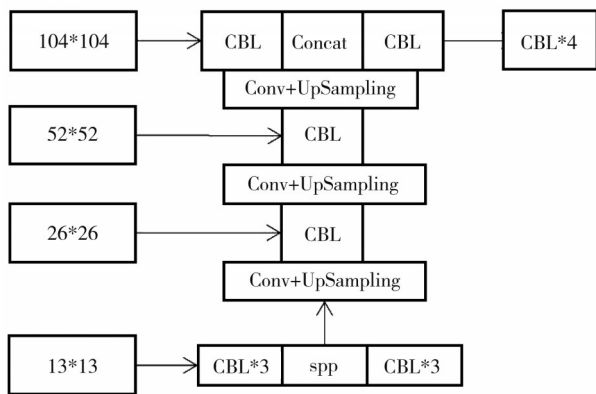


图 4 改进的PANet结构

Fig. 4 Improved PANet structure

3 实验

3.1 实验环境设置

本实验所使用的实验环境如表 1 所示。

表 1 实验环境

Tab. 1 Experimental setting

配置	参数
操作系统	Windows 10
CPU	Intel Core i5-10400F
GPU	NVIDIA GeForce RTX 2060 SUPPER
开发平台	Pycharm
学习框架	Pytorch

3.2 数据集

为评估实验方法的有效性，本文使用 2 个数据集进行实验。

1) DOTA 数据集：用于目标检测的大规模数据集，每个图像的大小约为 4 k×4 k 像素，包

含不同尺度、方向和形状的目标。完全注释的 DOTA 数据集包含 188 282 个实例，涉及 15 个常见目标类别：飞机、棒球场(BD)、桥梁、场地跑道(GFT)、小型车辆(SV)、大型车辆(LV)、网球场(TC)、篮球场(BC)、储罐(ST)、足球场(SBF)、环形交叉路口(RA)、游泳池(SP)、以及港口和直升机(HC)。数据集随机选择一半的原始图像作为训练集，1/6 作为验证集，1/3 作为测试集^[10]。根据高度将所有目标分为 3 个部分：10~50 像素范围为小，50~300 像素范围为中，300 像素以上为大。因此，将高度范围为 10~50 像素的目标定义为小对象。图 5 为训练集和验证集中不同类别的小目标的占比。

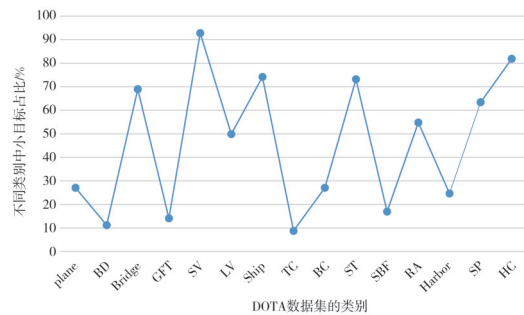


图 5 DATO 数据集中不同类别的小目标占比

Fig. 5 Percentage of small targets in different categories in the DATO dataset

2) DIOR-R：它包含 192 518 个实例和 23 463 个图像，分辨率从 0.5~30 m 不等。DIOR-R 共有 20 个常见类别，包括飞机、机场、棒球场、篮球场等。图像被分成训练和测试数据集，比例为 9:1。原始数据集使用数据增强方法进行了扩展，包括几何变换和马赛克方法。马赛克数据增强基于 Cut-Mix 理论。不同的是，在 Cut-Mix 中拼接了 2 幅图像，但马赛克使用了 4 幅图像。使用马赛克的优点是它丰富了图像数据集的背景^[12]。首先，使用镜像和旋转将数据集扩展到原始大小的 8 倍；然后，使用马赛克方法合并图像。

3.3 评价指标

本文使用真实数据(Ground Truth)和预测边界框(IOU)之间的平均精度(AP)和平均精度均值(mAP)、准确率(Precision)、召回率(Recall)来评估目标检测性能^[13]。计算公式为

$$Precision = \frac{TP}{TP + FP} \times 100\%, \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \times 100\%, \quad (7)$$

$$mAP = \frac{\sum_{i=1}^K AP_i}{K} \quad (8)$$

3.4 实验结果

根据上式验证优化的特征融合网络的结果,本研究在DIOR-R数据集和DOTA数据集上进行了实验。

表2展示了本文方法和其他方法在DOTA数据集上的*mAP*比较,实验证明,本文方法在DOTA数据集上的*mAP*较其他方法都有所提高,证明了本文方法对于遥感小目标图像的有效性。

表2 本文方法和其他方法在DOTA数据集上的*mAP*比较
Tab.2 Comparison of the *mAP* of the method in this paper with other methods on the DOTA dataset

	<i>mAP</i> /%				
	R-FCN	FPN	YOLOv3	SSD	本文方法
Plane	81.01	88.70	79.00	57.85	89.20
BD	58.96	75.10	77.10	32.79	79.21
Bridge	31.64	52.60	33.90	16.14	57.35
GFT	58.97	59.20	68.10	18.67	70.10
SV	49.77	69.40	52.80	0.05	80.25
LV	45.04	78.80	52.20	36.93	81.14
Ship	49.29	84.50	49.80	24.74	88.32
TC	68.99	90.60	89.90	81.16	91.13
BC	52.07	81.30	74.80	25.10	82.20
ST	67.42	82.60	59.20	47.47	83.00
SBF	41.83	52.50	55.50	11.22	56.32
RA	51.44	62.10	49.00	31.53	63.12
Harbor	45.15	76.70	61.50	14.12	77.63
SP	53.30	66.30	55.90	9.09	68.51
HC	33.89	60.10	41.70	0.01	63.20
Avg	52.59	72.23	60.03	27.12	75.39

表3为YOLOv4和本文方法在DIOR-R数据集的性能比较。

表3 YOLOv4和改进YOLOv4比较

Tab.3 Comparison of YOLOv4 and improved YOLOv4

	<i>Precision</i>	<i>Recall</i>	<i>mAP</i> /%
YOLOv4	0.79	0.82	82.59
YOLOv3	0.75	0.87	84.12
SSD	0.85	0.72	79.50
本文方法	0.89	0.83	86.76

如表3所示, *mAP*从82.59%增加到86.76%,根据之前的锚框的选择,改进的YOLOv4的IoU低于YOLOv4。改进的YOLOv4

具有更好的性能,并且平衡了精度和速度。

表格中数据显示,本文方法在DOTA数据集的各个类别上都优于其他方法,证明本文方法的有效性。

3.5 消融实验

为了验证本文方法的有效性,在DIOR-R数据集进行了消融实验对比。实验结果如表4所示,实验包含了5种模型:1)原始的YOLOv4模型;2)优化YOLOv4模型中的锚框;3)改进YOLOv4模型中的激活函数;4)改进YOLOv4中的PANet模块;5)本文方法。

表4 消融实验结果对比

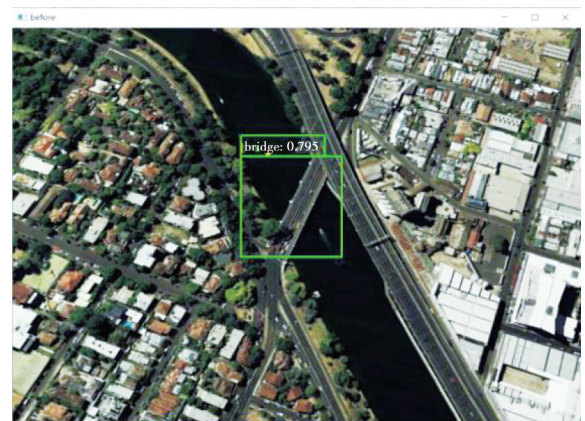
Tab.4 Comparison of results of ablation experiments

检测算法	<i>mAP</i> /%	<i>Speed</i> /fps
YOLOv4	82.6	26.3
YOLOv4+优化锚框	83.1	24.1
YOLOv4+动态激活函数	84.9	25.3
YOLOv4+改进PANet	85.4	25.9
本文方法	86.3	23.5

方法1)和方法2)对比表明,K-means对遥感图像小目标重新聚类,对锚框的数量和尺寸进行优化,有效提高了实验的精度和速度;方法1)和方法3)对比表明,使用动态激活函数自适应不同的输入特征,提高了YOLOv4算法的性能;YOLOv4和本文方法对比,*mAP*提高了3.7%,检测速度达到了23.5 fps,对比结果证明了本文改进算法的合理性和有效性。

3.6 检测结果对比

实际检测效果对比如图6所示。其中图6(a)为原版YOLOv4的检测效果,图6(b)为本文改进后的YOLOv4的检测效果。



(a) YOLOv4检测效果图



(b) 改进YOLOv4检测效果图

图 6 改进前后的YOLOv4算法检测效果对比

Fig. 6 Comparison of detection effects before and after improving YOLOv4 algorithm

由图 6 可以直观看出：原版的YOLOv4算法对目标的检测效果不佳，改进后的YOLOv4算法提高了特征提取能力，检测效果有了很大的提高。

4 结论

针对遥感图像小目标检测效果不佳的问题，对YOLOv4算法的K-means重新聚类，优化了锚框的数量和尺寸，提升检测的精度和速度；将激活函数改为了动态激活函数，自适应输入特征，提高了模型的泛化能力；同时，对PANet结构进行改进，丰富了浅层网络，更好地提高对小目标的检测能力。实验结果表明，改进的YOLOv4在检测遥感小目标时，在参数数量、检测精度方面优于原始的YOLOv4模型。

参考文献：

- [1] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001: 511-518.
- [2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'05), 2005: 1063-6919.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Image Net classification with deep convolutional neural networks[J]. Advances in Neural Information Process-

ing Systems, 2012, 25(2): 84-90.

- [4] 侯越千, 张丽红. 基于Transformer的多尺度物体检测[J]. 测试技术学报, 2023, 37(4): 342-347.
- HOU Yueqian, ZHANG Lihong. Multi-scale object detection based on Transformer[J]. Journal of Test and Measurement Technology, 2023, 37(4): 342-347. (in Chinese)
- [5] 高倩, 潘杨, 朱磊, 等. 基于SIOU函数的改进YOLOv5遥感目标检测方法[J]. 长江信息通信, 2022, 35(11): 5-8.
- GAO Qian, PAN Yang, ZHU Lei, et al. An improved YOLOv5 remote sensing target detection method based on SIOU function[J]. Changjiang Information & Communication, 2022, 35(11): 5-8. (in Chinese)
- [6] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: Optimal speed and accuracy of object detection [J]. arXiv. org. 2020. DOI: 10.48550/arXiv.2004.10934.
- [7] LIU C, FENG L, LIU G, WANG H B, et al. Bottom-up broadcast neural network for music genre classification[J]. Multimedia Tools and Applications, 2020(1): 1-19.
- [8] GORBAN A N, MIRKES E M, TUKIN I Y. How deep should be the depth of convolutional neural networks: a backyard dog case study[J]. Cognitive Computation, 2020, 12(1): 388-397.
- [9] YI T. Exponential stability of pseudo almost periodic solutions for fuzzy cellular neural networks with time-varying delays[J]. Neural Processing Letters, 2019: 49(15): 1-11.
- [10] ARIF I, ASLAM W, HWANG Y. Barriers in adoption of internet banking: a structural equation modeling neural network approach[J]. Technology in Society, 2020, 61: 101231.
- [11] LIU M. Uav-yolo: small object detection on unmanned aerial vehicle perspective [J]. Sensors, 2020, 20(8): 2238.
- [12] XIA G S, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//IEEE Conferent on Computer Vision an Pattern Recognition, 2018: 3974-3983.
- [13] LÜ X, DUAN F, JIANG J J, et al. Deep metallic surface defect detection: the new benchmark and detection network[J]. Sensors, 2020, 20(6): 1562.