

基于 MM-YOLOv4 的无人机目标检测算法

程千顷, 王红军*

(国防科技大学电子对抗学院, 合肥, 230037)

摘要 针对当前无人机目标图像检测方法精度较低和检测速度过慢的问题, 提出一种结合轻量级网络和改进多尺度结构的目标检测算法。首先采用 MobileNetV3 轻量级网络替换 YOLOv4 的主干网络, 减少模型复杂度, 提升检测速度; 其次, 引入改进多尺度结构的 PANet 网络, 增强高维图像特征和低维定位特征的流动叠加, 提升对小目标的分类和定位精度; 最后, 利用 K-means 方法对目标锚框进行参数优化, 提升检测效率。同时结合公开数据集和自主拍摄方式构建一个新的无人机目标图像数据集 Drone-dataset, 并基于数据增强的方法开展算法性能实验。实验结果表明, 该算法的 mAP 达到了 91.58%, FPS 达到了 55 帧/s, 参数量为 44.39 M 仅是 YOLOv4 算法的 1/6, 优于主流的 SSD、YOLO 系列算法和 Faster R-CNN 算法, 实现了对多尺度无人机目标的快速检测。

关键词 无人机; 目标检测; 轻量级网络; 改进多尺度结构

DOI 10.3969/j.issn.2097-1915.2022.05.014

中图分类号 TP391 **文献标志码** A **文章编号** 2097-1915(2022)05-0090-06

Research on UAV Target Detection Algorithm Based on MM-YOLOv4

CHENG Qianqing, WANG Hongjun*

(Electronic Engineering Institute of National University of Defense Technology, Hefei 230037, China)

Abstract Aimed at the problems that the target image detection methods are low in accuracy and slow at detection speed of current UAV (Unmanned Aerial Vehicles, UAV), a target detection algorithm in combination with the lightweight network and the improved multi-scale structure is proposed. Firstly, MobileNetV3 lightweight network is used to replace the backbone network of YOLOv4, reducing the model complexity and improving the detection speed. Secondly, the improved multi-scale PANet network is introduced to enhance the flow superposition of high-dimensional image features and low-dimensional location features, and improve the classification and location accuracy of small targets. Finally, the K-means method is introduced to optimize the parameters of the target anchor frame to improve the detection efficiency. Meanwhile, a new UAV target image dataset Drone-dataset is constructed by combining with the open dataset and the self-shot images. The results show that the mAP and FPS of the proposed algorithm reach 91.58% and 55 f/s, and the parameter number of 44.39M is only 1/6 of the YOLOv4 algorithm and is superior to the mainstream SSD, the YOLO series and the Faster R-CNN algorithms.

Key words unmanned aerial vehicle (UAV); target detection; lightweight network; improved multi-scale

收稿日期: 2022-05-23

作者简介: 程千顷(1997—), 男, 山东临沂人, 硕士生, 研究方向为图像识别、目标检测。E-mail: qianqing_cheng@163.com

通讯作者: 王红军(1968—), 男, 江苏镇江人, 博士, 教授, 研究方向为分布式网络、目标检测。E-mail: hongjun-wang@163.com

引用格式: 程千顷, 王红军. 基于 MM-YOLOv4 的无人机目标检测算法研究[J]. 空军工程大学学报, 2022, 23(5): 90-95. CHENG Qianqing, WANG Hongjun. Research on UAV Target Detection Algorithm Based on MM-YOLOv4 [J]. Journal of Air Force Engineering University, 2022, 23(5): 90-95.

structure

无人机类目标有飞行高度低、慢速以及小型化的特点^[1],这使得传统的雷达、无线电方法对其检测难度巨大,性价比不高^[2-3],而基于声音的检测手段存在受噪声干扰大的缺点^[4]。因此一种新的无人机目标检测方法的研究需十分迫切。

随着深度学习技术的成熟以及图像运算设备的更新换代,基于图像的目标检测方法逐渐成为目标检测领域的主流热点^[5]。其中具有代表性的算法有基于二阶段检测机制的 Faster R-CNN 和基于一阶段检测机制的 SSD(single shot multibox detector, SSD)以及 YOLO(you only look once, YOLO)系列等算法。YOLOv4^[6]算法是开创性地使用了 CSP-DarkNet-53 网络作为主干网络提取图像特征, PANet(path aggregation network, PANet)作为特征融合结构以及使用 SPP(spatial pyramid pooling, SPP)结构加强特征提取^[7],使得一阶段检测算法在精度和速度上赶超了大部分二阶段算法。

虽然目标检测算法凭借良好的性能深受广大研究者喜爱,但较大的模型体量和计算开销使得快速检测成为一个难题。文献[8]针对航拍目标检测问题,通过对 YOLOv4 的主干网络进行轻量化改造,精简模型结构,提升了 0.008 s 的检测速度以及减少了 13 M 的参数量,但算法对模型改造不彻底,提升效果并不明显。文献[9]中使用基于迁移学习的 YOLOv3 算法在自建无人机数据集上实现了精准

检测,平均准确率达到到了 88.9%,但未对 YOLOv3 算法进行过多改进,准确率仍有提升空间,实时性也没有体现。文献[10]使用卷积神经网络对无人机进行检测,在自建数据集上达到了 93% 的识别准确率,但实验使用早期的支持向量机和 K 最近邻方法,难以体现算法的先进性。文献[11]针对检测模型内存占用大、检测实时性差的问题,提出了一种改进双尺度的 YOLOv4 算法,通过裁剪通道和稀疏训练的方法,减少了 60% 的内存占用,提升了 35% 的 FPS,达到了 58 帧/s,但 mAP 精度有所下降。

针对算法参数量过大以及小目标检测精度低的问题,本文提出了一种结合轻量级网络 MobileNet V3 和改进多尺度结构 PANet 的无人机目标检测算法 MM-YOLOv4 (mobilenetV3-mutlscale-YOLOv4, MM-YOLOv4)。

1 MM-YOLOv4 算法模型

1.1 模型架构

MM-YOLOv4 是在 YOLOv4 基础上进行改进的算法。算法模型的图像输入尺寸为 $416 \times 416 \times 3$,输出为 3 种尺度大小的预测结果,其网络架构主要由主干网络 MBNNet16 和特征融合网络 Mutlscale-PANet 以及预测模块 Yolohead 3 部分组成。见图 1。

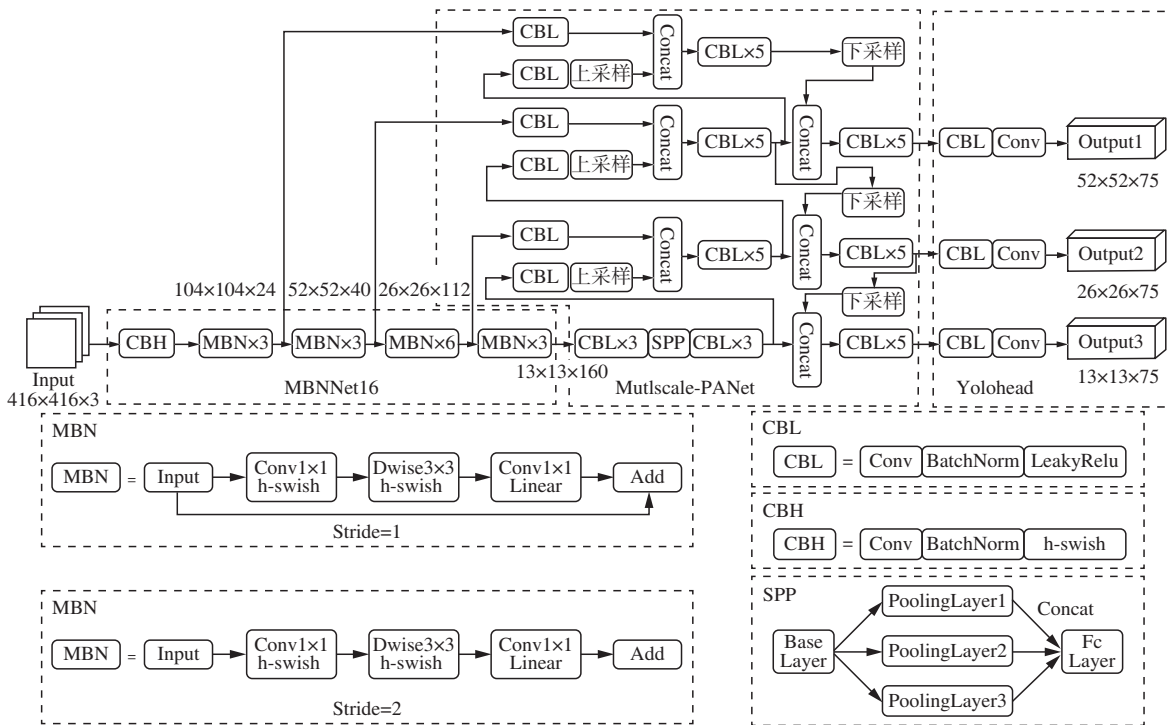


图 1 MM-YOLOv4 网络模型架构图

MBNNet16 由 CBH (conv + batchnorm + h-wish, CBH) 模块和 MBN (mobile bottleneck, MBN) 模块堆叠构成, 共计 16 层。其作用是输出不同尺度的卷积特征图, 以适应不同大小目标的检测。在图 1 中 MBNNet16 以输出特征图的节点为分割点, 将 MBN 模块进行叠加表示, 例如 $MBN \times 3$, 表示为 3 个 MBN 模块叠加。Mutlscale-PANet 主要由 CBL (conv + batchnorm + leakyrelu, CBL) 模块以及 SPP 模块搭建而成。其作用是对输入的特征图进行融合叠加, 丰富特征图信息。Yolohead 由 CBL 模块和卷积层组成, 其作用是检测目标的位置和类别并输出。

1.2 模型轻量化改进

轻量级网络以其精简的网络结构和快速提取图像特征的优点被广泛应用。

MobileNetV3 网络是 Andrew Howard 等学者提出的第 3 个版本的轻量级网络^[12]。它结合了 V1 版本的深度可分离卷积和 V2 版本的线性瓶颈逆残差结构, 并引入 h-swish 激活函数以节省计算开销, 因此算法使用 MobileNetV3 作为 YOLOv4 轻量化改进的主干网络^[13], 来减少模型复杂度和提升检测速度。

主干网络的替换主要是使用剪枝后的 MobileNetV3 网络作为模型的起始部分, 对输入图像进行卷积运算, 并输出 4 种尺度特征图, 其大小为 104×104 , 52×52 , 26×26 以及 13×13 , 兼顾了大、中、小 3 种尺度目标的全局和局部特征, 可以为后续的分类和定位预测提供了充足的信息。MobileNetV3 网络在充当主干时剪除了最后的 Classifier (分类层), 由输出第 4 尺度特征图的 MBN 层作为主干结构的最后一层, MBNNet16 具体结构见表 1。operator 表示本层所用模块及卷积核大小; exp size 表示在 MBN 模块中增加的通道数; out size 表示输出的通道数; SE 表示是否在深度可分离卷积块中增加 SE (squeeze-and-excite, SE) 结构; s 表示步长。

表 1 MBNNet16 网络结构

floor	Input	operator	exp size	out size	SE	s
1	$416^2 \times 3$	CBH, 3×3	-	16	-	2
2	$208^2 \times 16$	MBN, 3×3	16	16	-	1
3	$208^2 \times 16$	MBN, 3×3	64	24	-	2
4	$104^2 \times 24$	MBN, 3×3	72	24	-	1
5	$104^2 \times 24$	MBN, 5×5	72	40	✓	2
6	$52^2 \times 40$	MBN, 5×5	120	40	✓	1
7	$52^2 \times 40$	MBN, 5×5	120	40	✓	1
8	$52^2 \times 40$	MBN, 3×3	240	80	-	2

表 1 MBNNet16 网络结构 续表

floor	Input	operator	exp size	out size	SE	s
9	$26^2 \times 80$	MBN, 3×3	200	80	-	1
10	$26^2 \times 80$	MBN, 3×3	184	80	-	1
11	$26^2 \times 80$	MBN, 3×3	184	80	-	1
12	$26^2 \times 80$	MBN, 3×3	480	112	✓	1
13	$26^2 \times 112$	MBN, 3×3	672	112	✓	1
14	$26^2 \times 112$	MBN, 5×5	672	160	✓	2
15	$13^2 \times 160$	MBN, 5×5	960	160	✓	1
16	$13^2 \times 160$	MBN, 5×5	960	160	✓	1

1.3 模型多尺度结构改进

PANet 网络是用于实例分割的路径聚合网络^[6]。部分结构示意图见图 2。

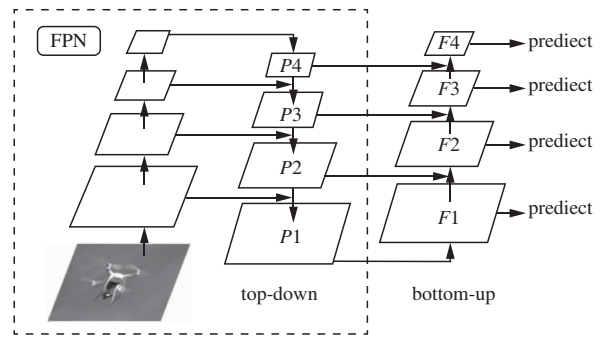


图 2 PANet 网络部分结构图

YOLOv4 算法的特征融合网络应用了 PANet 网络的部分结构, 在进行特征融合时使用了 3 个尺度的特征图, 这种方式增强了算法对图像特征的提取和整合能力, 但在一定程度上损失了部分小目标的特征信息。本文为了加强对小尺度目标的定位和图像特征提取能力, 通过改进 PANet 网络作为颈部特征融合结构的功能, 增加一个特征输出分支, 来丰富浅层网络提取的目标特征信息。改进后的具体结构如图 1 中 Mutlscale-PANet 模块所示。

1.4 K-means 算法锚框改进

锚框是目标检测算法用来遍历图像中感兴趣区域并进行预测识别的像素框。MM-YOLOv4 使用 K-means 算法对数据集进行聚类分析, 以得到适合目标数据集训练的锚框大小, 自适应能力强且可以针对不同的数据集进行不同的锚框设计。

K-means 算法是通过选取距离度量和准则函数不断进行迭代最终选取 k 个聚类中心的聚类算法^[16]。论文选用真实框和预测锚框之间的面积交并比 IoU 作为度量进行聚类。公式为:

$$f = \arg \max \frac{\sum_{i=1}^k \sum_{j=1}^{n_k} I_{IoU}(B, A)}{N} \quad (1)$$

式中: B 为真实框; A 为预测锚框; k 为聚类中心 (预测锚框) 的个数; n_k 为第 k 个聚类中心中的真实框的

数量; N 为真实框的总个数; $I_{IoU}(B, A)$ 为真实框和预测锚框的面积交并比; i 代表聚类中心序号; j 表示聚类中心中真实框的序号。

表 2 为 YOLOv4 算法的默认锚框和 MM-YOLOv4 算法使用 K-means 算法对 Drone-dataset 数据集进行锚框聚类得出的最佳锚框对比表。

表 2 默认锚框和 K-means 算法锚框

目标类型	默认锚框	K-means 方法锚框
小尺度	(12,16)(19,36) (40,28)	(10,9)(15,17) (27,26)
中尺度	(36,75)(76,55) (72,146)	(57,48)(112,89) (181,155)
大尺度	(142,110)(192,243) (459,401)	(320,189)(254,260) (369,313)

2 数据集与数据增强

本文通过结合文献[15]中的公开数据集和自主拍摄无人机目标图像的方式构建了数据集 Drone-dataset。Drone-dataset 中包含了大、中、小 3 种尺度的无人机目标,图像数量为 3 698 张,其中 2 664 张为原公开数据集图像,其余 1 034 张为后续扩充图像。表 3 为 Drone-dataset 数据集的详细信息以及训练集、验证集和测试集划分情况。

表 3 Drone-dataset 数据集信息表

类别	图像数量	大尺度目标	中尺度目标	小尺度目标
训练集	2 662	1 621	547	509
验证集	296	176	60	63
测试集	740	434	171	138
总计	3 698	2 231	778	710

为了测试 MM-YOLOv4 算法的可行性,文中还选用了 PASCAL VOC2012 公开数据集作为实验的检测基准。

3 实验

3.1 实验环境与配置

实验环境基于深度学习框架 PyTorch 且配有 CUDA10.2 平台作为并行运算结构,计算机环境为 Windows 10 操作系统 Intel(R) Core(TM) i9-10900F CPU @ 2.80 GHz 2.81 GHz,内存为 64 G,GPU 为 NVIDIA Quadro P4000,8 G 显存。

模型训练的超参数设置如下:训练的初始学习率为 0.01,最低学习率为 0.000 1,并使用余弦退火的学习率衰减方法,模型的图像输入大小为 416 ×

416,Epoch 次数为 200,Batch size 为 16。

3.2 评价指标

实验选取目标检测精度 mAP(mean average precision,mAP)、检测速度 FPS(frames per second,FPS)和模型参数量作为评价指标,来检验模型的整体性能。

mAP 可以全方面地评价模型对多类别和多目标任务的定位和分类效果。计算 mAP 需要计算识别任务中每个类别的 average precision (AP) 然后取其平均值,其公式为:

$$mAP = \frac{\sum_{i=1}^C AP_i}{C} \quad (2)$$

式中: C 为总类别数目; AP_i 为第 i 类的 AP 值。

AP 由精确率(precision)和召回率(recall)计算得来:

$$precision = \frac{TP}{TP + FP} \quad (3)$$

$$recall = \frac{TP}{TP + FN} \quad (4)$$

$$AP = \int_0^1 P(R) dR \quad (5)$$

式中:TP(true positive)即实际输入为正样本,预测结果也为正样本;FP(false positive)即实际输入为负样本,预测结果为正样本;FN(false negative)即实际输入为正样本,预测结果为负样本;TN(true negative)即实际输入为负样本,预测结果为负样本;P 代表 precision,R 代表 recall。

3.3 实验结果与分析

3.3.1 可行性与先进性实验

为验证 MM-YOLOv4 算法的可行性,本实验使用 Faster R-CNN、SSD、YOLOv3、YOLOv4、MobileNetV3-YOLOv4 和 MM-YOLOv4 算法在 PASCAL VOC2012 数据集上进行对比实验,PASCAL VOC2012 数据集包含 20 个类别的大中小尺度目标,能综合反映算法的检测性能。表 4 为各算法在 PASCAL VOC2012 数据集上各性能指标的检测结果对比。

表 4 结果表明,在检测精度方面,MM-YOLOv4 在众多算法中的 mAP 指标最高,达到了 90.69%;在检测速度 FPS 和模型参数量方面,MM-YOLOv4 的性能略差于 MobileNetV3-YOLOv4,但优于其它算法,FPS 指标达到了 55 帧/s,参数值仅有 44.39 M,略优于 SSD 算法 46 帧/s 的 FPS 和 99.76 M 的参数值。此外相较于 YOLOv4、Faster R-CNN 等算法不到 30 帧/s 的检测速度和 200 M 以上的参数量,MM-YOLOv4 算法则更占优势。

综上,MM-YOLOv4 算法在 YOLOv4 的基础

上结合 MobileNetV3 轻量级网络和改进多尺度 PANet 结构,大幅度缩减了参数量,加快了检测速度,在 PASCAL VOC 2012 数据集得到了不错的检测结果,相较于 YOLOv3、YOLOv4 和 Faster R-CNN 等二阶段检测算法,MM-YOLOv4 算法表现出更大优势,体现出了算法的可行性以及有效性。

表 4 PASCAL VOC2012 数据集各算法检测结果对比

模型	主干网络	mAP/%	FPS/ (帧/s)	参数 量/M
Faster R-CNN	ResNet-50	79.42	17	522.91
SSD	VggNet-16	70.72	46	99.76
YOLOv3	DarkNet-53	81.50	35	235.08
YOLOv4	CSPDarkNet-53	85.81	27	244.29
MobileNetV3- YOLOv4	MobileNetV3	87.64	65	43.51
MM-YOLOv4	MobileNetV3	90.69	55	44.39

为验证算法的先进性,实验基于 Drone-dataset 数据集开展对比实验。该数据集包含了大、中、小 3 种尺度的无人机目标,可以全面检验算法对多尺度无人机目标检测的适应能力。表 5 为各算法的训练验证结果对比,实验指标使用 COCO 检测指标,其中检测精度 mAP 特指在 IoU=0.5 下的值,单位为%,其余的 AP_{small}、AP_{medium}、AP_{large} 分别是 3 种尺度目标的平均精度值。

结果表明,MM-YOLOv4 算法在 IOU=0.5 下的 mAP 精度指标达到了 91.58%,优于除 Faster R-CNN 以外的其他算法。此外,在 AP 精度指标上 MM-YOLOv4 表现良好,由小尺度目标到大尺度目标的精度值分别为 22.92%、37.43% 和 56.00%。尤其是对小尺度目标检测的 AP 精度,相较于 YOLOv4 和 Mobilev3-YOLOv4 分别提升了 4.4% 和 5.15%,体现出改进多尺度结构对小目标检测的提升效果。

表 6 基于 Drone-dataset 数据集的消融实验

模型	CSPDarkNet-53	MobileNetV3	PANet	Mutiscale- PANet	K-means	mAP/%	FPS/ (帧/s)	参数量/M
YOLOv4	✓		✓			91.06	27	244.29
YOLOv4+MutilScale	✓			✓		93.00	12	247.93
YOLOv4+MobileNetV3		✓	✓			89.83	65	43.51
YOLOv4+Kmeans	✓		✓		✓	91.92	28	244.29
MM-YOLOv4		✓		✓	✓	91.58	55	44.39

4 结语

针对无人机目标的快速检测问题,论文在 YOLOv4 算法的基础上采取多种改进策略,提出了

表 5 Drone-dataset 数据集各算法检测结果对比 单位:%

模型	mAP	AP _{small}	AP _{medium}	AP _{large}
Faster R-CNN	93.07	36.68	42.86	62.92
SSD	90.22	9.99	43.79	64.12
YOLOv3	85.87	18.30	34.60	62.37
YOLOv4	91.06	18.52	36.93	62.45
MobileNetV3- YOLOv4	89.83	17.77	35.84	56.06
MM-YOLOv4	91.58	22.92	37.43	56.00

3.3.2 消融实验

为了体现 MM-YOLOv4 的改进策略相比于 YOLOv4 的提升,对算法开展消融实验,实验通过使用 Drone-dataset 数据集对 MM-YOLOv4 算法的每一部分改进策略进行训练并记录其 mAP(IoU=0.5)、FPS 以及参数量指标,查看其改进效果。表 6 为各部分改进策略在 Drone-dataset 数据集上的指标对比。

由表 6 可知,本文算法在 3 个指标的性能上均有不同程度的提升,其中基于主干结构的轻量级网络改进对模型性能的提升效果比较明显,在 mAP 精度仅下降 0.23% 的代价下,将 FPS 提升到了 65 帧/s,并且参数量降到了 43.51 M;基于 PANet 改进的多尺度结构表现尚可,提升了模型 1.94% 的 mAP,但是 FPS 下降了 15 帧/s,同时参数量增加了 3.64 M;基于 K-means 的锚框聚类方法提升了模型 0.86% 的 mAP 精度以及 1 帧/s 的检测速度。最终,结合了这 3 种改进策略的 MM-YOLOv4 算法 mAP 精度达到了最高的 91.58%,提高了 0.52%;FPS 值达到了 55 帧/s,提高了 28 帧/s;参数量降到 44.39 M 仅为 YOLOv4 算法的 1/6,这体现出 MM-YOLOv4 算法在检测精度、检测速度以及模型参数量上的高效平衡以及针对无人机目标检测任务的优异性能。

MM-YOLOv4 无人机目标检测算法。首先,使用 MobileNetV3 轻量级网络作为 YOLOv4 的主干特征提取模块,精简模型结构,提升检测速度;然后采用改进多尺度结构的 PANet 网络作为特征融合模块,充分利用浅层网络提取的目标特征信息,增强对

小尺度目标的鉴别能力;最后采用 K-means 聚类方法对目标数据集进行聚类分析,优化锚框参数,增强检测效率。实验结果表明,MM-YOLOv4 算法在 Drone-dataset 和 PASCAL VOC2012 这两类目标复杂的数据集上,表现出比常见算法更高的检测精度和更快的检测速度,为无人机目标的快速检测提供了一种实用可行的研究思路。下一步的研究将针对 PANet 网络改进造成的检测速度下降问题,选取合适的特征金字塔结构,优化特征融合通道,以适应更快的无人机目标检测需求。

参考文献

- [1] 蒋镛圻,白若楷,彭月平. 低慢小无人机目标探测技术综述[J]. 飞航导弹, 2020(9): 100-105.
- [2] MENDIS G J, RANDENY T, JIN W, et al. Deep Learning Based Doppler Radar for Micro UAS Detection and Classification[C]//Military Communications Conference. [S.l.]: IEEE, 2016: 924-929.
- [3] BISIO I, GARIBOTTO C, LAVAGETTO F, et al. Unauthorized Amateur UAV Detection Based on WiFi Statistical Fingerprint Analysis[J]. IEEE Communications Magazine, 2018, 56(4): 106-111.
- [4] BOUGAIOV N, DANIK Y. Hough Transform for UAV's Acoustic Signals Detection[J]. The Advanced Science Journal, 2015(6): 65-68.
- [5] 谢富,朱定局. 深度学习目标检测方法综述[J]. 计算机系统应用, 2022, 31(2): 1-12.
- [6] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2020. ArXiv: 2004. 10934v1 [cs.CV].
- [7] LIU S, QI L, QIN H, et al. Path Aggregation Network for Instance Segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. doi: <http://doi.org/10.48550/arXiv.1803.01534>.
- [8] 赵玉卿,贾金露,公维军,等. 基于 pro-YOLOv4 的多尺度航拍图像目标检测算法[J]. 计算机应用研究, 2021, 38(11): 3466-3471.
- [9] XUN D T W, LIM Y L, SRIGRAROM S. Drone Detection Using YOLOv3 with Transfer Learning on NVIDIA Jetson TX2[C]//2021 Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP). [S.l.]: IEEE, 2021: 1-6.
- [10] MAHDAVI F, RAJABI R. Drone Detection Using Convolutional Neural Networks[C]//2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS). [S.l.]: IEEE, 2020: 1-5.
- [11] 田港,张鹏,邹金霖,等. 基于改进 YOLOv4 的无人机目标检测方法[J]. 空军工程大学学报(自然科学版), 2021, 22(4): 9-14.
- [12] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), [S.l.]: IEEE, 2019: 1314-1324.
- [13] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 4510-4520.
- [14] ZHANG X, ZOU J, HE K, et al. Accelerating Very Deep Convolutional Networks for Classification and Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(10): 1943-1955.
- [15] HAO Y J, TECK L K, XIANG C Y, et al. Fast Drone Detection Using SSD and YoloV3 [C]//2021 21st International Conference on Control, Automation and Systems (ICCAS). [S.l.]: IEEE, 2021: 1172-1179.

(编辑:徐敏)