

李博豪, 李智. 基于改进 YOLOv8n 的无人机航拍小目标检测算法[J]. 智能计算机与应用, 2025, 15(5): 173–179. DOI: 10.20169/j. issn. 2095–2163. 24120902

基于改进 YOLOv8n 的无人机航拍小目标检测算法

李博豪, 李智

(四川大学 电子信息学院, 成都 610065)

摘要: 随着低空经济的迅速发展, 无人机在环境监测、农业植保、应急救援等多个领域展现出广阔的应用前景。然而, 无人机航拍图像中的小目标检测问题一直是限制其广泛应用的技术瓶颈之一。针对无人机航拍图像视角广、目标小且背景复杂的特点, 提出了一种改进的 YOLOv8n 目标检测算法, 以提高无人机视角下小目标的检测精度。首先, 为了更好地捕捉无人机视角下不同视距的目标的形状特征, 引入了可变形卷积核。其次, 增加了 P2 小目标检测头, 并在 Neck 部分的 C2f 模块中引入了空间通道注意力机制, 提升模型对不同区域目标的区分能力。最后, 提出了自适应边界损失函数, 使其在处理小目标时生成更优的边界框。实验结果表明, 在 MDMT (Multi-Drone Multi-Target) 数据集上, 该算法在小目标检测精度上显著优于基线模型原生 YOLOv8n、YOLOv5 和 Faster-RCNN 等模型, $mAP50$ 达到了 61.0%, 分别提升了 13.7%、14.0%、50.8%。

关键词: YOLOv8n; 小目标检测; 无人机视角; 注意力机制; 可变形卷积

中图分类号: TP391

文献标志码: A

文章编号: 2095–2163(2025)05–0173–07

Research on small target detection algorithm for drone aerial photography based on improved YOLOv8n

LI Bohao, LI Zhi

(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: With the rapid development of the low-altitude economy, Unmanned Aerial Vehicles (UAVs) have demonstrated broad application prospects in multiple fields such as environmental monitoring, agricultural plant protection, and emergency rescue. However, the problem of small target detection in UAV aerial images has always been one of the technical bottlenecks restricting their wide application. In response to the characteristics of UAV aerial images, which feature wide angles of view, small targets, and complex backgrounds, an improved YOLOv8n target detection algorithm is proposed to enhance the detection accuracy of small targets from the perspective of UAVs. Firstly, deformable convolution kernels are introduced to better capture the shape features of targets at different viewing distances from the UAV's perspective. Secondly, a P2 small target detection head is added, and a spatial-channel attention mechanism is introduced into the C2f module in the Neck part to improve the model's ability to distinguish targets in different regions. Finally, an adaptive bounding box loss function is proposed to enable the generation of better bounding boxes when dealing with small targets. Experimental results show that on the MDMT (Multi-Drone Multi-Target) dataset, this algorithm is significantly superior to baseline models such as the native YOLOv8n, YOLOv5, and Faster-RCNN in terms of small target detection accuracy. The $mAP50$ reaches 61.0%, with improvements of 13.7%, 14.0%, and 50.8% respectively.

Key words: YOLOv8n; small target detection; UAV perspective; attention mechanism; deformable convolution

0 引言

目前, 低空经济迎来了快速发展, 其核心内容为利用 500 m 以下的空域进行各种商业活动。由于无人机技术的进步, 已在道路交通监控^[1]、关键区域巡视^[2]、航空摄影^[3]和检查服务^[4]等领域得到广泛应用。出现了无人机监控、航空摄影和检查服务等

应用。这些创新不仅提高了运营效率, 还为物流、农业和应急响应等行业开辟了新的商机。而实现这些商业任务离不开基于计算机视觉的识别、追踪等任务, 目前主流的方案是利用 YOLO 系列的深度学习模型, 但在实际当中由于飞行高度过高、分辨率过大、画面目标存在遮掩等问题, 原生的 YOLO 模型往往很难达到应用目的, 往往需要进行针对性的改进。

作者简介: 李博豪(1999—), 男, 硕士研究生, 主要研究方向: 多视角无人机航拍小目标联合检测算法研究。

通信作者: 李智(1974—), 男, 博士, 教授, 主要研究方向: 物联网与边缘计算, 压缩感知与频谱感知, 群体智能。Email: 286086522@qq.com。

收稿日期: 2024–12–09

哈尔滨工业大学主办 ◆ 专题设计与应用

由于无人机高度高、视距大、视角倾斜验证,导致有的目标例如行人、汽车等在画面中占据的像素位置很少,这一现象给目标检测任务带来了严峻挑战,也促使相关研究聚焦于多方面的优化策略。谢椿辉等学者^[5]提出一种 Drone-YOLO 算法,有效提高无人机航拍影像的小目标实时检测精度。韩玉洁等学者^[6]通过数据增强以及对原网络激活函数和损失函数的改进,有效提高了检测准确率。Liang 等学者^[7]提出了一种新颖的轻量级跨层三分支并行融合网络(CTPF net),增强了低层次原始特征和深层次语义特征之间的跨尺度转移融合,有效提高了无人机图像中小目标的实时检测精度。Ye 等学者^[8]提出了一种全局-局部特征增强网络(GLF-net)和多尺度特征融合模块,解决无人机图像中大量小目标的问题。Chen 等学者^[9]提出的基于模拟数据迁移学习与自适应融合的综合方法,进一步增强了小目标检测性能,为解决当前无人机目标检测面临的挑战提供了新思路。戚玲珑等学者^[10]针对小目标检测容易出现漏检问题,设计了一种改进的 YOLOv7 目标检测模型,通过对 YOLOv7 网络模型中的 MPConv 模块进行改进,以减少网络特征处理过程造成的特征损失,使得漏检情况得到明显改善。Zhou 等学者^[11]为提升复杂背景下红外小目标的检测能力,基于 YOLO 检测框架与 SRGAN 网络提出了一种改进的检测算法 YOLO-SASE。Hui 等学者^[12]提出了一种名为 STF-YOLO 的无人机遥感图像小目标检测算法,提出了一种名为 STRCN 的新型卷积结构,小目标检测性能取得显著提升。韩强^[13]通过引入可变形卷积、注意力机制和修改损失函数提出了一种改进 YOLOv8 的小目标检测算法。

综上所述,本研究提出了一种基于改进 YOLOv8n 的无人机航拍小目标检测算法,该算法以 YOLOv8n 为底座,引入了可变形卷积和空间通道注意力机制,增加了一个 P2 小目标检测头,同时对其中的 C2f 模块进行改进,引入 CBAM 模块进行特征交互融合,最后设计了一种考虑真实边界框与预测边界框横纵向长度比例的损失函数 $C'IoU$,提升了模型对小目标物体框的位置和大小的回归性能。

1 YOLOv8n 模型改进方法

1.1 改进 YOLOv8n 网络结构

改进 YOLOv8n 网络结构如图 1 所示。首先输入图像通过一系列卷积层(Conv)和可变形卷积层(DCNv4)进行初步处理。骨干网络使用一系列卷

积层,并穿插了 C2f-CBAM 模块,CBAM 通过自适应地细化特征图来增强和交互融合特征表示。网络的颈部部分设计用于融合来自骨干网络不同层次的特征。通过使用上采样(UP)和连接(Concat)操作来结合不同尺度的特征(P2-P5),这对于多尺度目标检测至关重要。网络的头部处理融合后的特征以生成预测。每个检测模块(Detect)会输出不同尺度(P2、P3、P4、P5)的物体的类别概率和边界框坐标。

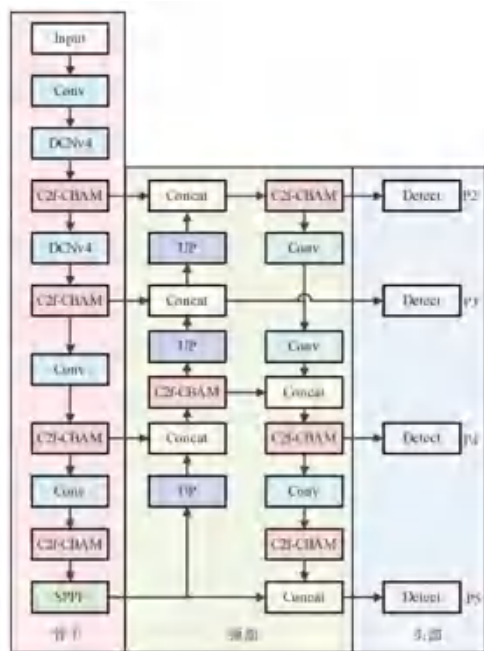


图1 改进 YOLOv8n 网络模型

Fig. 1 Improved YOLOv8n network model

1.2 可变形卷积

可变形卷积网络(Deformable Convolution Network, DCN)^[14]系列算法的提出旨在增强模型学习复杂目标不变性的能力。原生 YOLOv8n 模型仅使用 3×3 这样的单一卷积核进行特征提取,无法对不同形状的目标特征进行有效的特征学习。因此,为解决模型对目标形状特征提取能力的不足,本研究在原网络结构中引入可变形卷积核模块 DCNv4^[15]。该模块可以自适应地调整卷积核大小、形状以及位置等信息,从而更精准地贴合目标的实际特征进行特征提取操作。同时 DCNv4 相较于早期版本的可变形卷积实现方式,减少了不必要的冗余计算环节,从而整体提升了运算速度,有力保障了检测的实时性。

本文将 DCNv4 加入至 YOLOv8n 的骨干网络,如图 2 所示:本文将骨干网络的第 1 个、第 2 个 C2f 模块之前的 Conv 模块替换为 DCNv4。即骨干网络对应 P2、P3 检测层的卷积替换为可变形 DCNv4,用来提升模型对中小目标的关注。在增加

可变形卷积模块后,因为小目标所在位置的感受野自适应地发生了改变,在预测框回归时,模型能够更

好地调整预测框的回归参数,使得模型增加对小目标的关注度,进而提升模型的整体性能。

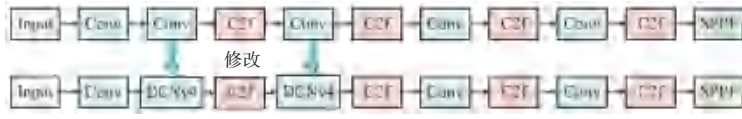


图2 YOLOv8n 骨干网络修改结构图

Fig. 2 Modified structure diagram of the backbone network of YOLOv8n

1.3 P2 小目标检测头

在 YOLOv8n 模型的特征提取阶段,随着网络深度的增加,特征图的通道数逐渐增多,而特征图的空间尺寸则逐渐减小。这一过程实现了对输入图像局部信息的映射和压缩。随着网络层的深入,特征图的空间分辨率降低,单个像素点所覆盖的原始图像区域变得更大,这可能导致模型在处理较小目标时的性能下降。为了捕捉小目标的细节信息,需要保留那些感受野与小目标尺寸相匹配的特征图,以便用于后续的分类和边界框回归任务。

因此,本研究提出将原网络结构的第一个 C2f 模块的输出特征图纳入检测头,添加检测头 P2,并与其他具有较大感受野的特征图进行有效融合。这种融合策略旨在增强模型对小尺寸目标的检测能力,提高模型的整体鲁棒性。

1.4 C2f-CBAM 模块

注意力机制旨在模拟人类视觉系统的焦点选择功能,以识别和强调复杂场景中的关键特征。在目标检测任务中,尤其是在面对无人机图像中小目标检测这种复杂场景时,有效捕捉关键特征对于提升检测精度至关重要。

CBAM (Convolutional Block Attention Module)^[16]卷积块注意力模块,是一个极具灵活性的即插即用模块。主要由通道注意力模块和空间注意力模块两部分构成。其中,通道注意力模块能够依据特征图各个通道的信息分布,自动学习每个通道的重要程度权重。空间注意力模块则聚焦于特征图的空间维度,通过分析空间位置上特征的相关性,确定不同空间位置的重要性权重,这样一来,网络能够更精准地定位到目标在图像中的具体位置,即使目标较小且处于复杂背景中,也能突出目标区域的特征,抑制无关背景区域的干扰。

本研究在 C2f 模块拼接操作前添加了 CBAM 模块,改进过程如图 3 所示。在传统的 C2f 模块中,这 2 部分处理后的特征图直接进行 Concat 操作并由底部的 Conv 层输出最终特征图。然而,这种方式可能

无法充分挖掘特征图中的关键信息,对于小目标特征的提取和利用不够高效。而通过在 Concat 操作之前添加 CBAM 模块,CBAM 模块能够自适应地对特征图进行加权,在通道和空间维度上调整特征的权重分布。这样一来,在后续的 Concat 操作及最终输出时,网络能够更好地关注重要的特征,提高网络的性能和特征表示能力。

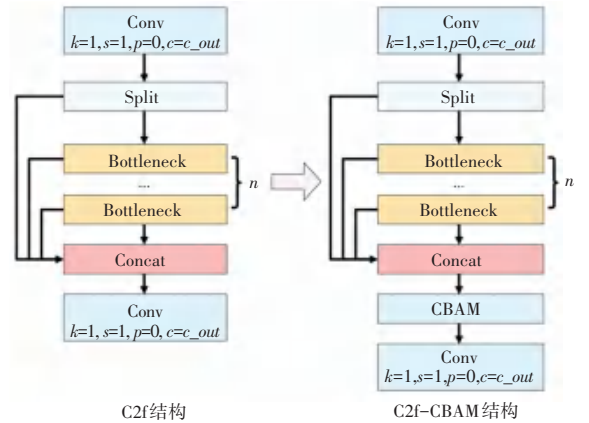


图3 C2f 结构到 C2f-CBAM 结构的改进过程

Fig. 3 Improvement process from C2f structure to C2f-CBAM structure

1.5 改进损失函数 $C'IoU$

YOLOv8n 原生网络边界框损失函数采用的是 $CIoU$ (Complete Intersection over Union, $CIoU$) 损失,旨在解决 IoU 在某些情况下的局限性。 $CIoU$ 损失不仅考虑了边界框之间的重叠区域,还考虑了预测框和真实框中心点之间的距离以及长宽比的差异,从而使得模型在训练过程中能够更有效地学习到目标的位置和尺度信息。计算公式如下:

$$CIoU = IoU - \frac{\rho^2(b_{pred}, b_{gt})}{C^2} - \alpha V \quad (1)$$

$$IoU = \frac{b_{gt} \cap b_{pred}}{b_{gt} \cup b_{pred}} \quad (2)$$

$$V = \frac{4}{\pi^2} \left(\arctan \frac{w_{gt}}{h_{gt}} - \arctan \frac{w_{pred}}{h_{pred}} \right)^2 \quad (3)$$

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

根据式(1)~式(4)可以看出其侧重于预测框和真实框的重叠部分,忽略了边界框的大小回归是否正确。而边界框的大小对于小目标物体来说,相对于边界框位置更加重要。因此将 $CIoU$ 修改为 $C'IoU$, 添加预测框和真实框之间的横纵向比例的修正来对 IoU 进行加权, 数学公式具体如下:

$$V' = \left(2 - \left(\frac{w^{gt}}{w^{pred}} + \frac{h^{gt}}{h^{pred}} \right) \right)^2 \tag{5}$$

$$C'IoU = IoU - \frac{\rho^2(b_{pred}, b_{gt})}{C^2} - \alpha(V + V') \tag{6}$$

在 $C'IoU$ 中,增加了预测边界框和真实边界框横纵向的比例作为损失的一部分。通过这种方式, $C'IoU$ 能够更好地评估检测结果,并且在存在大小物体不平衡的情况下也能给出更准确的评价。

2 实验结果与分析

为了验证改进后的 YOLOv8n 网络有效性,本文将该算法与部分目标检测网络 YOLOv5^[17]、Faster-RCNN^[18]、YOLOv7^[19]、YOLOv8n 进行横向对比,使用的数据集均相同。

2.1 实验数据集

Multi-Drone Multi-Target Tracking (MDMT) 数据集^[20]由 2 架 DJI PHANTOM 4 Pro 无人机在不同场景(如广场、城市道路、乡村道路、交叉路口、立交桥和停车场)和光照条件下采集,包含 88 组(44 对)视频序列,共 39 678 帧,分辨率为 1920×1080。MDMT 数据集示例如图 4 所示。



图 4 MDMT 数据集示例

Fig. 4 Examples of the MDMT dataset

2.2 环境与参数设置

工作站配置如下:操作系统为 Windows10 系统,处理器为 AMD Ryzen 5600 G, GPU 为 NVIDIA GeForce GTX1080Ti, 软件环境为 Python 3. 11. 9, CUDA12. 2。

本文训练时使用 SGD 优化器优化网络,初始学习率 $lr = 0. 01$, 权重衰减为 $5e-5$;开启 Mosaic 数据

增强。经过多次对比实验后,将训练次数设定为 200 轮次,批量大小设置为 8,每训练 1 次保存 1 次模型参数文件,最后在网络内使用最优的参数文件对测试集进行预测。

2.3 实验评价指标

本文将采用平均精度 (mAP)、精确率 ($Precision$) 和召回率 ($Recall$) 等来对算法检测效果进行评估。其中, $Precision$ 表示模型预测为正类别的样本中真正为正类别的比例, $Recall$ 表示所有实际为正类别的样本中被正确预测的比例。推得的公式如下:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

所有类别的检测平均精度 (Average Precision, AP) 的公式定义为:

$$AP = \int_0^1 P(R) dR \tag{9}$$

$mAP50$ 意味着当 IoU 阈值设置为 0. 5 时的平均精度。 $mAP95$ 指的是当 IoU 阈值设置为 0. 95 时的平均精度。综合所有类别,得到 mAP , 计算公式为:

$$mAP = \frac{1}{N} \sum_{i=0}^n AP_i \tag{10}$$

mAP 越大越好,说明模型效果更好。

2.4 算法有效性验证

本文针对数据集中小目标的显著性特征,提出了改进 YOLOv8n 的目标检测方法。为定性分析该方法的有效性和合理性,将本文方法与主流深度学习目标检测网络 Faster-RCNN、YOLOv5、YOLOv7-tiny 以及 YOLOv8n 相比较。将本文所改进的网络与上述网络用同一数据集进行训练,并将所有训练完成的网络使用同一测试集进行测试,将检测后的结果使用相同的方法进行评价,结果见表 1。图 5 为本文算法的训练过程可视化结果。

表 1 不同网络结果对比

Table 1 Comparison of results of different networks %				
算法	Precision	Recall	mAP50	mAP95
Faster-RCNN	—	—	10. 2	7. 3
YOLOv5	68. 3	43. 3	47. 0	25. 6
YOLOv7-tiny	53. 5	42. 4	40. 5	19. 8
YOLOv8n	68. 6	43. 9	47. 3	26. 8
本文算法	73. 8	54. 7	61. 0	33. 7

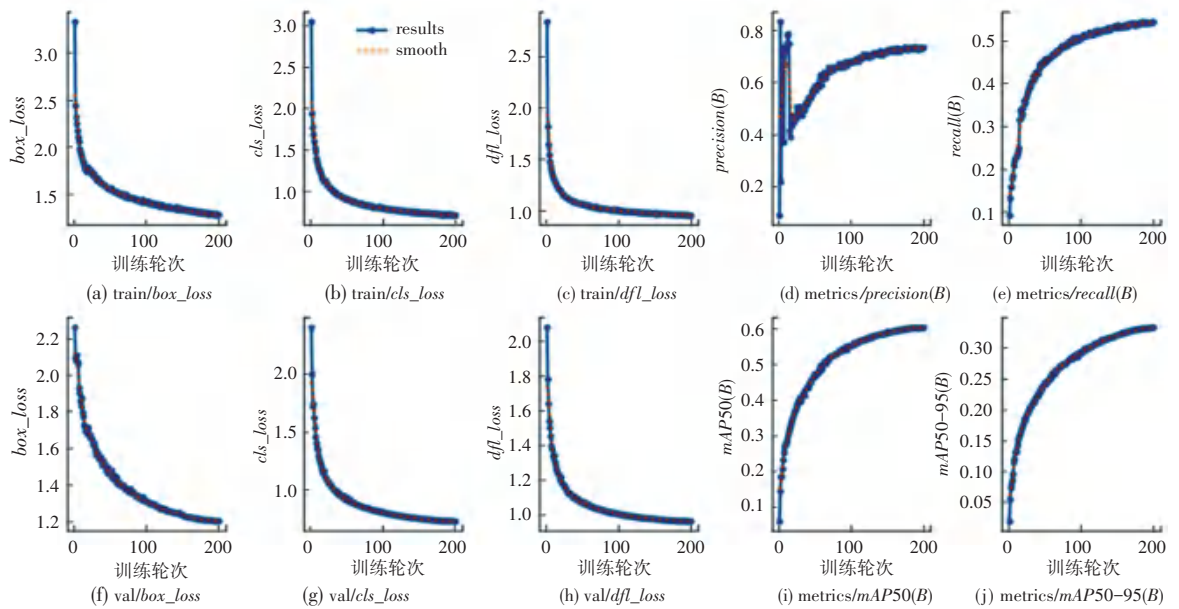


图 5 模型训练可视化结果

Fig. 5 Visualization results of model training

从图 5 中可以看到完整训练过程中,各种指标在训练集和测试集上的变化情况。训练集和验证集上的各类损失随迭代次数增加而逐渐减小。同时,模型的精度、召回率以及不同阈值下的平均精确度

随迭代次数增加而逐渐提高,表明模型性能在训练过程中不断得到优化。图 6 展示了各种模型在对行人小目标检测时的表现性能。



图 6 不同模型在小目标(行人)上的检测性能

Fig. 6 Detection performance of different models on small targets (pedestrians)

原图中,行人小目标在 1920×1080 的分辨率下仅为十几个像素的黑点,且与背景融合较深,一些较大的目标,例如车辆等与之混合。可以看到,所有模型对大目标(车辆)都具有较好的检测性能。但是

YOLOv5、YOLOv7-tiny 和 YOLOv8n 对行人小目标存在大量的漏检。在增加 P2 检测混合层和 DCNv4 之后,大部分小目标都被检测到,但是存在检测框错位和重叠的情况。通过引入改进损失函数,本文模型有效地降低了边界框的重叠和误框,提升了检测性能。

2.5 不同模块的消融实验

为了分析本文算法中提出的各改进模块对小目标检测结果的影响,设计了消融实验来评估。消融实

验共分为 5 组。第 1 组实验使用原生 YOLOv8n 网络,第 2 组实验在 YOLOv8n 网络的基础上添加 P2 小目标检测层,第 3 组实验在 YOLOv8n 网络的基础上添加 P2 和 DCNv4,第 4 组实验在 YOLOv8n 网络的基础上添加 P2、DCNv4 和 C2f-CBAM,第 5 组则是在 YOLOv8n 网络的基础上添加所有模块,包括 P2、DCNv4、C2f-CBAM 和 $C'IoU$ 损失函数,实验所使用的训练参数一致。消融实验具体结果见表 2。

表 2 不同模块的消融实验
Table 2 Ablation experiments of different modules

算法	Precision	Recall	mAP50	mAP95
YOLOv8n	68.6	43.9	47.3	26.8
YOLOv8n+P2	69.7	49.4	57.4	31.5
YOLOv8n+P2+DCNv4	71.1	53.6	59.1	33.1
YOLOv8n+P2+DCNv4+C2f-CBAM	72.3	54.1	60.2	33.4
YOLOv8n+P2+DCNv4+C2f-CBAM+Loss(本文算法)	73.8	54.7	61.0	33.7

由表 2 实验结果可知,通过逐步加入 P2 层、DCNv4、CBAM 和优化的损失函数 $C'IoU$,模型的性能在 *Precision*、*Recall*、*mAP50* 和 *mAP95* 等多个指标上都有了显著的提升。

3 结束语

实验结果表明,上述改进措施显著提升了模型的性能。在 MDMT 数据集上,改进后的 YOLOv8n 模型相比基线模型原生 YOLOv8n,在小目标检测精度上有了明显的提升,*mAP50* 达到了 61.0%,相比原版提升了 13.7%,*mAP95* 也从 26.8% 提升至 33.7%。这些结果不仅展示了算法在技术上的显著优势,也为无人机技术在低空经济中的深入应用提供了有力支持。

然而,尽管本文提出的改进 YOLOv8n 算法在无人机视角下的小目标检测任务中取得了显著的成果,但仍存在一些局限性和可以改进的方面。本文的实验主要集中在无人机航拍图像上的小目标检测。然而,算法在其他类型图像数据集上的适用性尚未得到充分验证。未来的研究可以进一步扩展到其他类型的图像数据集上,如医学影像、遥感图像和工业检测图像等,以验证算法的泛化能力和适用性。这将有助于确定算法在不同应用场景中的有效性和局限性。

参考文献

[1] KANISTRAS K, MARTINS G, RUTHERFORD M J, et al. A

Survey of unmanned aerial vehicles (UAVs) for traffic monitoring [J]. Handbook of UAVs, 2015, 34(2): 2643-2666.
[2] KORKI M, SHANKAR N D, SHAH R N, et al. Automatic fault detection of power lines using unmanned aerial vehicle (UAV) [C]//Proceedings of the 1st International Conference on Unmanned Vehicle Systems - Oman (UVS). Piscataway, NJ: IEEE, 2019:1-6.
[3] 郭辰阳,敖万忠,吕宜宏. 低空经济与通用航空、无人机、UAM 的关系分析[J]. 财经界, 2023(28): 30-32.
[4] 黄巧龙,蔡雪雄. 低空经济产业:发展现状、问题与政策建议[J]. 发展研究, 2024, 41(5): 58-64.
[5] 谢椿辉,吴金明,徐怀宇. 改进 YOLOv5 的无人机影像小目标检测算法[J]. 计算机工程与应用,2023,59(9):198-206.
[6] 韩玉洁,曹杰,刘琨,等. 基于改进 YOLO 的无人机对地多目标检测[J]. 电子测量技术,2020,43(21):19-24.
[7] LIANG B, SU J, FENG K, et al. Cross-layer triple-branch parallel fusion network for small object detection in UAV images [J]. IEEE Access, 2023, 11: 39738.
[8] YE T, QIN W, LI Y, et al. Dense and small object detection in UAV-vision based on a global-local feature enhanced network [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1.
[9] CHEN Rui, GUO Youwei, ZHENG Huafei, et al. A comprehensive approach for UAV small object detection with simulation-based transfer learning and adaptive fusion[J]. arXiv preprint arXiv, 2109.01800, 2021.
[10] 戚玲珑,高建瓴. 基于改进 YOLOv7 的小目标检测[J]. 计算机工程,2023,49(1):41-48.
[11] ZHOU Xiao, JIANG Lang, HU Caixia, et al. YOLO-SASE: An improved YOLO algorithm for the small targets detection in complex backgrounds[J]. Sensors, 2022, 22(12):4600.
[12] HUI Yangming, WANG Jue, LI Bo. STF-YOLO: A small target detection algorithm for UAV remote sensing images based on improved Swin Transformer and class weighted classification

decoupling head[J]. Measurement,2024,224:113936.

[13] 韩强. 面向小目标检测的改进 YOLOv8 算法研究[D]. 长春: 吉林大学,2023.

[14] DAI Jifeng, QI Haozhi, XIONG Yuwen, et al. Deformable convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ; IEEE,2017: 764–773 .

[15] XIONG Yuwen, LI Zhiqi, CHEN Yuntao, et al. Efficient deformable convnets; Rethinking dynamic and sparse operator for vision applications[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ; IEEE, 2024;5652.

[16] MA Ke, ZHAN C, YANG Feng. Multi – classification of arrhythmias using ResNet with CBAM on CWGAN–GP augmented ECG Gramian angular summation field [J]. Biomedical Signal Processing and Control, 2022, 77:103684.

[17] JOCHER G, CHAURASIA A, STOKEN A, et al. ultralytics/ YOLOv5: v6. 2 – YOLOv5 classification models, apple m1, reproducibility, clearml and deci. ai integrations [EB/OL]. https://DOI:10.5281/zenodo.7347926.

[18] SUN Xudong, WU Pengcheng, STEVEN C H. Face detection using deep learning: An improved faster RCNN approach [J]. Neurocomputing, 2018, 299: 42.

[19] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag–of–freebies sets new state–of–the–art for real–time object detectors [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ; IEEE, 2023; 7464–7475.

[20] LIU Zhihao, SHANG Yuanyuan, LI Timing, et al. Robust multi–drone multi – target tracking to resolve target occlusion; A benchmark [J]. IEEE Transactions on Multimedia, 2023, 25: 1462.