

# 面向空中小目标检测任务的 YOLOv7 改进模型



董凤禹, 魏振忠

北京航空航天大学, 北京 100191

**摘要:** 针对空中小目标检测任务, 本文进行了基于改进的 YOLOv7 模型的空中小目标检测算法的研究, 解决了当前空中小目标检测算法推理速度慢和检测精度不足的问题。首先, 建立了涵盖多种目标尺度、姿态和天气条件的飞机目标基准数据集; 其次, 在 YOLOv7 基准模型的基础上, 提出了一种基于广义特征金字塔网络和 Wasserstein 度量的目标检测方法; 最后, 在公开数据集与自建数据集上对该方法与其他主流算法进行了对比试验, 结果表明, 相较于原始 YOLOv7 模型, 改进模型在自建数据集上对小目标检测的平均精确率提高了 7.3%, 并且推理速度高于大部分主流检测算法。本文研究为空中小目标检测任务提供了一种快速且高精度的检测算法, 对于相关算法在航空航天领域的进一步工程应用具有重要的推动作用。

**关键词:** 空中目标; 目标检测; 计算机视觉; 深度学习; 损失函数

中图分类号: V219

文献标识码: A

DOI: 10.19452/j.issn1007-5453.2023.12.013

空中小目标检测任务通常是指在航空或航天等领域中, 对飞行器或卫星等载具进行实时监控, 以实现空中交通管制、地质探测、天气预测、空中加油<sup>[1]</sup>、飞行器姿态测量<sup>[2]</sup>等应用。相较于高分辨率的图像, 空中目标往往相对尺寸小或尺度变化极大, 且目标速度快、动态性强, 因此对检测算法的精度和实时性都有很高的要求。在中小目标检测任务中, 常需要处理复杂的背景干扰、快速移动的目标和目标尺度变化等问题。同时, 由于这些目标通常是在远距离或高空下进行检测, 所以还需要考虑光照、气象等因素的影响。因此, 针对这些问题, 需要使用高精度、高鲁棒性的目标检测算法, 并且在算法的设计上需要考虑到实时性和计算复杂度等因素。

当前, 随着深度学习技术在计算机视觉领域的应用, 目标检测器的精度和速度都有了巨大的提升。目前主流的目标检测算法包括 R-CNN、Fast R-CNN<sup>[3]</sup>、Faster R-CNN<sup>[4]</sup>、YOLOv7<sup>[5]</sup>、RetinaNet<sup>[6]</sup>、CornerNet<sup>[7]</sup>、CenterNet<sup>[8]</sup> 等。尽管目标检测方法发展迅速, 但小目标检测仍是当前亟待解决的难题。根据国际光学工程学会的定义, 小目标的成像尺寸小于整个成像区域的 0.12%。小目标在图像中通常只占

少数像素, 携带较少的目标特征, 因此通用目标检测方法难以对小目标实现较好的检测效果。针对小目标检测任务, 相关研究人员已经进行了一些研究。B. Zoph 等<sup>[9]</sup>尝试通过设计有效的数据增强策略解决小目标检测数据不足的问题。Li 等<sup>[10]</sup>使用生成对抗网络(GAN)将低分辨率特征作为 GAN 的输入来生成高分辨率特征, 但这种方法的检测速度较慢, 并且对用于训练的数据集有较高的要求。Yang 等<sup>[11]</sup>采用了一种查询机制来提高基于特征金字塔的对象检测器的推理速度, 提高了小目标检测器的性能。

综上所述, 这些小目标检测方法都是为了在中小目标检测场景中提高检测精度和速度而被提出。它们的具体思路和技术实现方式不同, 但都采用了一些特殊的技巧来解决小目标检测问题。然而, 这些方法仍然存在一些局限性和缺陷, 如模型复杂性增加、速度变慢等, 需要进一步研究和改进。本文针对空中小目标检测任务, 在 YOLOv7 基准模型的基础上进行了改进。具体来说, 采用 GFPN 结构替换 YOLOv7 原有的 PANet, 并且设计新的损失函数以提高模型在处理小目标检测和目标尺度变化较大的检测任务时的表现。通过改进, 本文方法在中小目标检测和目标尺度变化较大

收稿日期: 2023-06-30; 退修日期: 2023-10-09; 录用日期: 2023-11-06

基金项目: 航空科学基金(201946051002); 国家自然科学基金(52127809)

引用格式: Dong Fengyu, Wei Zhenzhong. An improved YOLOv7 model for small aerial object detection [J]. Aeronautical Science & Technology, 2023, 34(12): 111-117. 董凤禹, 魏振忠. 面向空中小目标检测任务的 YOLOv7 改进模型 [J]. 航空科学技术, 2023, 34(12): 111-117.

的数据集上取得了较好的表现,比YOLOv7原始模型有更高的检测精度和较快的检测速度,适合在实际应用中使用。

度比YOLOX快180%(FPS),比Dual-Swin-T快1200%(FPS),比ConvNext快550%(FPS),比Swin-L快500%(FPS)。此外,即使在精度为56.8%AP的情况下,模型的检测速度仍可达到30帧/s以上。YOLOv7的出色性能使其可用于实时目标检测和跟踪任务中,其结构如图1所示。

### 1 YOLOv7目标检测模型

YOLOv7是一种单步法目标检测模型,YOLOv7的速

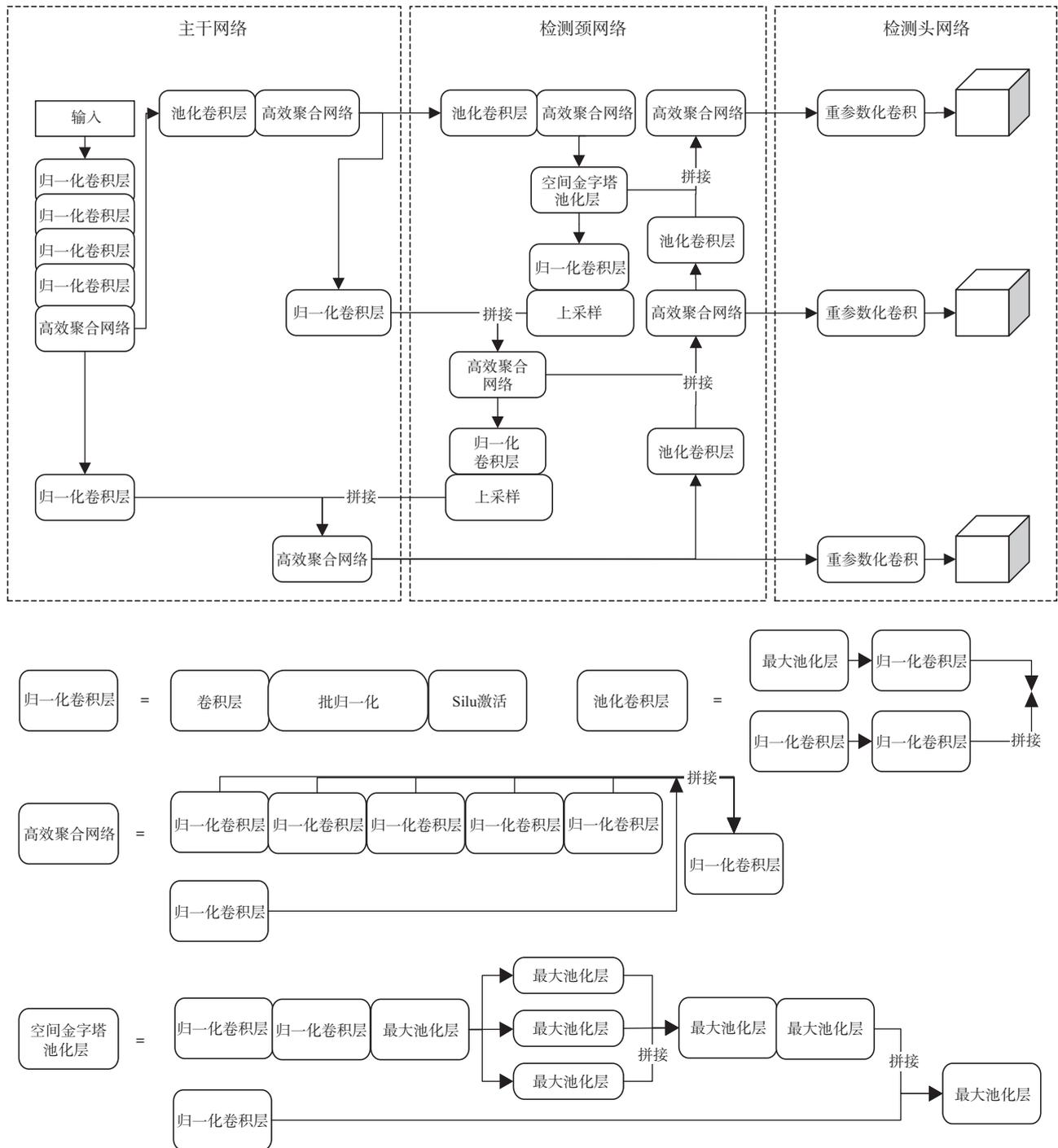


图1 YOLOv7模型网络结构  
Fig.1 YOLOv7 network structure

YOLOv7将模型重参数化的思想引入网络架构中,通过对模型参数进行适当的重组,可以在不影响模型性能的前提下,减少模型参数数量,从而降低模型计算复杂度。YOLOv7采用了YOLOv5的跨网格搜索和YOLOX的匹配策略,对标签分配进行了优化,提高了模型的准确性。此外,YOLOv7提供了一种新的高效网络架构ELAN,该架构可以在保持准确性的前提下,大幅减少计算和内存开销,并通过引入辅助头,可以在不影响推理时间的前提下,提高模型的训练效率和准确性。尽管YOLOv7通过模型重参数化、标签分配优化、高效网络架构设计和辅助头训练方法等多方面的改进,提高了目标检测模型的性能和效率,但其在小目标检测方面仍有较大提升空间。

## 2 基于中间层改进与 Wasserstein 散度的YOLOv7模型

### 2.1 基于广义特征金字塔网络的YOLOv7中间层改进

特征金字塔网络(FPN)的目的是解决CNN骨干网络在提取不同分辨率、多尺度特征时存在的融合问题。除了最初的FPN, PANet和BiFPN等也常用于目标检测模型。尽管这些FPN结构注重特征融合,但它们缺乏块内链接,可能会导致在小目标检测任务中无法提取关键特征。而广义特征金字塔网络(GFPN)<sup>[12]</sup>采用一种新的路径融合方法,包含跳层和跨尺度连接,能够更好地提取和融合特征,相比于YOLOv7原有的PANet结构,GFPN结构在处理小目标检测和目标尺度变化较大的场景下具有更好的表现。GFPN结构能够更好地保留小目标的细节信息,并通过融合不同尺度的特征信息来提高检测精度。GFPN具体结构如图2所示。

GFPN中含有两种特征链接方法:dense-link和 $\log_2^n$ -

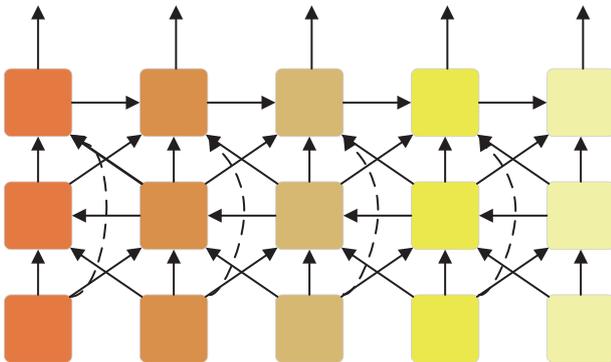


图2 GFPN结构示意图

Fig.2 Schematic diagram of GFPN structure

link,两种链接分别用以下公式表示

$$P_k^l = \text{Conv}(\text{Concat}(P_k^0, \dots, P_k^{l-1})) \quad (1)$$

$$P_k^l = \text{Conv}(\text{Concat}(P_k^{l-2^n}, \dots, P_k^{l-2^i}, P_k^{l-2^0})) \quad (2)$$

式中, $P_k^l$ 表示第 $l$ 层特征图在第 $k$ 级的尺度特征,Concat()表示将前面所有层生成的特征图串联在一起,Conv()表示使用 $3 \times 3$ 卷积对这些特征图进行处理。

式(1)表示dense-link的链接方式,在dense-link中第 $l$ 层接收所有前面层的特征图;式(2)表示 $\log_2^n$ -link的链接方式, $\log_2^n$ -link在每个级别 $k$ 中,第 $l$ 层从最多 $\log_2^l + 1$ 个前面的层接收特征图, $\log_2^n$ -link结构只会在反向传播时将层间距离从1增加到 $1 + \log_2^l$ ,因此可以被扩展到更深的神经网络中。

### 2.2 基于Wasserstein散度的损失函数

YOLOv7采用交并比损失(IoU)作为模型的损失函数,其计算公式如式(3)~式(5)所示

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

$$v = \frac{4}{\pi^2} (\arctan \frac{w_{\text{gt}}}{h_{\text{gt}}} - \arctan \frac{w}{h})^2 \quad (4)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (5)$$

式中, $b$ 为预测框; $b_{\text{gt}}$ 为真实框; $w, h$ 分别为预测框的宽和高; $w_{\text{gt}}, h_{\text{gt}}$ 分别为真实框的宽和高; $c$ 表示能够同时包含预测框和真实框的最小包围区域的对角线距离; $\alpha$ 为平衡参数; $v$ 用来衡量长宽比是否一致。

就常规目标检测任务而言,对于微小物体检测器来说IoU并不适用。因为在预测边界框与真实边界框之间没有重叠或者预测边界框完全包含真实边界框等情况下,IoU无法为优化网络提供梯度,而这种情况在微小物体检测中比较常见。根据式(3)~式(5),虽然YOLOv7采用的CIoU可以在一定程度上处理这些情况,但是它仍对微小物体的位置偏差比较敏感<sup>[13]</sup>。

为了解决这个问题,本文采用基于最优传输理论的损失函数来衡量目标框坐标的准确程度,通过Wasserstein散度来代替标准的IoU。Wasserstein散度能够一致地反映分布之间的距离,即使它们没有重叠。对于微小的物体,边界框中往往会有一些背景像素,因为大多数真实物体不是严格的矩形。因此,边界框可以建模为二元高斯分布,其中边界框的中心像素具有最高的权重,像素的重要性从中心到边界递减。边界框之间的相似度可以转换为两个高斯分布之间的分布距离。

具体而言,目标边界框 $R = (c_x, c_y, w, h)$ 可以建模为二

元高斯分布  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , 其中

$$\boldsymbol{\mu} = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \quad (6)$$

在此基础上, 将边界框 A 和 B 之间的相似度转换为高斯分布之间的分布距离并通过 Wasserstein 距离计算。对于两个二元高斯分布  $\mu_1 = N(m_1, \Sigma_1)$  和  $\mu_2 = N(m_2, \Sigma_2)$ ,  $\mu_1$  和  $\mu_2$  之间的二阶 Wasserstein 距离定义为

$$W_2^2(\mu_1, \mu_2) = \|m_1 - m_2\|_2^2 + \|\Sigma_1^{1/2} - \Sigma_2^{1/2}\|_F^2 \quad (7)$$

式中,  $\|\cdot\|_F$  是 Frobenius 范数。

此外, 对于从边界框  $A = (cx_a, cy_a, w_a, h_a)$  和  $B = (cx_b, cy_b, w_b, h_b)$  建模的高斯分布  $N_a$  和  $N_b$ , 式(7)可进一步简化为

$$W_2^2(N_a, N_b) = \left\| \left( \begin{bmatrix} cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right) \right\|_2^2 \quad (8)$$

但是,  $(N_a, N_b)$  是距离度量, 不能直接用作相似度度量, 因此用它的指数形式归一化, 得到损失函数

$$L(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}\right) \quad (9)$$

### 3 测试试验

为了评估本方法的性能, 本文在 VisDrone 基准数据集上进行了试验, VisDrone 是一个大规模的计算机视觉基准数据集, 包含 288 个视频片段和 10209 张静态图像, 共计 261908 帧, 图像内容涉及多个方面, 如不同的地理位置、环境条件、物体种类、目标密度等。此外, 该数据集还包含在不同天气和光照条件下拍摄的图像, 为数据集增加了多样性。为了提高数据集的可用性, 对所涉及的目标物体手动注释了超过 260 万个边界框, 并提供了其他属性, 如场景可见度、目标类别和遮挡情况。

此外, 为了验证改进前后的网络模型对于飞机目标的检测性能, 本文提出了一个新的飞机目标基准数据集, 该数据集包含 2800 张带有目标矩形框标注的图像, 采集自飞机试飞现场真实视频。数据集被划分为训练集、验证集和测试集, 分别包含 1700 张、500 张和 600 张带标注图像。如图 3 所示, 该数据集涵盖了多种天气条件、目标姿态和大小变化, 包括晴天、阴天、雾霾等多种天气条件, 实例尺寸覆盖 9~1521px, 其中小目标超过实例总数的 70%, 可以有效地评估目标检测算法的鲁棒性。



(a) 天气变化 (b) 姿态变化 (c) 尺度变化

图 3 空中目标检测基准数据集中多样化的图像

Fig.3 Diverse images in aerial object detection dataset

本文试验软件环境为 Ubuntu18.04, python3.8.13, pytorch1.11.0; GPU 为 NVIDIA GeForce 2080Ti; 图片大小设置为  $1280 \times 1280$ , 迭代次数 100 次。在基准数据集上检测网络进行性能测试, 试验中, 遵循 VisDrone2021 数据集的使用规则, 在 VisDrone 2021 中的训练集上进行网络的训练, 在 VisDrone2021 中的验证集上进行评估。训练过程中, 使用随机梯度下降法对网络参数进行优化, 使用 warm-up 来训练策略, warm-up epoch 设置为第三次迭代, 初始学习率取 0.01, 优化器权重衰减为 0.0005。

通过对比同样试验环境下改进前后的网络模型对不同类型图像的检测结果, 可以评估漏检和误检情况。评估方法主要选取准确率—召回率(P-R)曲线、平均准确率(AP)和平均精度均值(mAP)三个指标。

(1) 准确率—召回率(P-R)曲线

$$P = \frac{T_{TP}}{T_{TP} + F_{FP}} \times 100\% \quad (10)$$

$$R = \frac{T_{TP}}{T_{TP} + F_{FN}} \times 100\% \quad (11)$$

P-R 曲线是一个以准确率为纵轴、召回率为横轴的曲线图。其中, 准确率 P 是指在所有被分类为正样本中, 真正为正样本的比例; 而召回率 R 是指在所有实际为正样本中, 被分类为正样本的比例。对于不同阈值的分类器, 绘制其 P-R 曲线可以帮助我们选择最优的阈值, 以获得更好的检测效果。

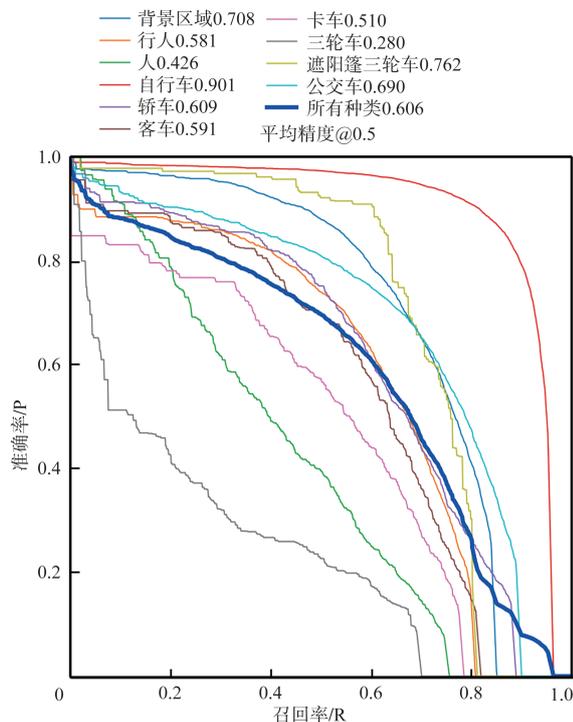
(2) 平均准确率(AP)

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P_{inter}(r_{i+1}) \quad (12)$$

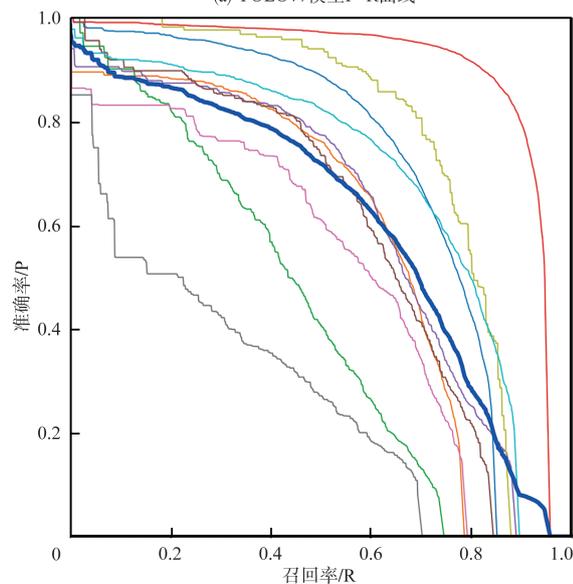
AP 是用于评估分类模型精度的指标之一, 它是 P-R 曲线下面积的平均值。通过计算 P-R 曲线下的面积可以得到 AP 的值, AP 值越高说明模型的检测效果越好。

## (3) 平均精度均值(mAP)

mAP是所有类别AP的平均值,是评估整个模型性能的一个综合指标。它可以很好地反映模型在不同类别图像上的检测效果。本文所提出的模型与YOLOv7基准模型在VisDrone数据集上的检测结果(P-R曲线)如图4所示,使用



(a) YOLOv7模型P-R曲线



(b) 本文模型P-R曲线

图4 YOLOv7模型与本文改进后的模型P-R曲线对比  
Fig.4 Comparison between the YOLOv7 model and the improved model P-R curve in this paper

P-R曲线来展示模型的性能,其中曲线下的面积AP值是反映模型综合性能的指标。试验结果表明,本文提出的模型在VisDrone数据集上表现出色,相较于YOLOv7基准模型,该模型的检测精度更高。

为了进一步验证本文提出的模型在VisDrone数据集上的性能,本文对其他常用目标检测算法进行了试验,并将试验结果列于表1中。根据测试结果,本文提出的模型在VisDrone数据集上具有更高的检测精度,相比于其他常用目标检测算法表现更出色。

表1 本文模型与现有代表性算法在VisDrone数据集上的试验结果比较

Table 1 Comparison of experimental results between the model in this paper and existing representative algorithms on the VisDrone dataset

检测方法	mAP/%	AP50/%
RetinaNet	11.81	21.37
RefineDet	14.90	28.76
DetNet59	15.26	29.23
Cascade-RCNN	16.09	31.91
FPN	16.51	32.20
Light-RCNN	16.53	32.78
CornetNet	17.41	34.12
DPNet-ensemble	29.62	54.00
SMPNet	35.98	59.53
YOLOv7	35.82	58.90
Ours	37.57	60.57

此外,本文在自己建立的空中目标基准数据集上进行了试验,以验证本文模型在小目标和中等大小的目标检测方面的效果,并将试验结果列于表2中。与通用的单步法检测器YOLOv3、YOLOv4等相比,在推理速度相当的情况下,本文提出的模型在中小目标检测方面具有更高的精度。与两步法检测器Cascade R-CNN等相比,本文提出的模型在精度方面与其相当,但在推理速度上有很大优势。

#### 4 结束语

本文针对空中小目标检测任务,对YOLOv7基准模型进行了改进,中间层引入GFPN结构,使用Wasserstein散度替换原有的交并比损失,并建立了一个涵盖各种天气条件下多种姿态和不同大小飞机目标的基准数据集。在该基准数据集和公开数据集上对模型的性能进行测试,结果表明,本文提出的模型在公开数据集与自建数据集上都有着较好的表现,具有较快的检测速度和较高的检测精度,对于空中小目标具有良好的检测性能,为解决空中小目标检测任务

表2 本文模型与现有代表性算法在基准数据集上的试验结果比较

Table 2 Comparison of experimental results between the model in this paper and existing representative algorithms on benchmark datasets

检测方法	mAP/%	mAP <sup>s</sup> /%	mAP <sup>m</sup> /%	时间/ms
Cascade R-CNN	98.6	89.4	99.2	105
Faster R-CNN	96.9	76.4	97.2	96
Libra R-CNN	89.1	76.4	89.3	95
RetinaNet	95.0	83.2	95.8	79
FCOS	90.1	83.3	92.5	164
YOLOv3	83.6	78.8	83.8	33
YOLOv7	97.0	82.2	87.6	21
Ours	98.1	89.5	98.5	41

的实际问题提供了有效的方法和思路。

AST

## 参考文献

- [1] 阎峰,刘震,基于立体视觉的无人机位姿测量方法[J].航空科学技术,2023,34(4):72-78.  
Yan Feng, Liu Zhen. UAV pose measurement method based on stereo vision[J]. Aeronautical Science & Technology, 2023, 34(4):72-78.(in Chinese)
- [2] 张西林,何亚坤,张格易,等.面向自主空中加油任务的目标检测技术研究[J].航空科学技术,2023,34(2):64-71.  
Zhang Xilin, He Yakun, Zhang Keyi, et al. Object detection technology for autonomous air to air refueling[J]. Aeronautical Science & Technology, 2023,34(2):64-71.(in Chinese)
- [3] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [4] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [5] 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, 2023: 7464-7475.
- [6] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [7] Law H, Deng J. Cornernet: Detecting objects as paired keypoints[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 734-750.
- [8] Duan K, Bai S, Xie L, et al. Centernet: Keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.
- [9] Zoph B, Cubuk E D, Ghiasi G, et al. Learning data augmentation strategies for object detection[C].Computer Vision-ECCV 2020: 16th European Conference, 2020: 566-583.
- [10] Li J, Liang X, Wei Y, et al. Perceptual generative adversarial networks for small object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1222-1230.
- [11] Yang C, Huang Z, Wang N. QueryDet: Cascaded sparse query for accelerating high-resolution small object detection[C]//Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, 2022: 13668-13677.
- [12] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.
- [13] Xu C, Wang J, Yang W, et al. Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 190: 79-93.

---

## An Improved YOLOv7 Model for Small Aerial Object Detection

Dong Fengyu, Wei Zhenzhong

*Beihang University, Beijing 100191, China*

**Abstract:** Addressing the challenges of slow inference speed and limited detection accuracy in small airborne target detection tasks, this paper investigated an enhanced detection algorithm grounded in the YOLOv7 model. Firstly, an encompassing benchmark dataset for aircraft targets was established, encapsulating varying scales, orientations, and weather conditions. Secondly, a novel detection approach was introduced, leveraging a generalized feature pyramid network and Wasserstein distance metric within the YOLOv7 framework. Finally, comparative evaluations encompassing public and self-constructed datasets validated the method against mainstream algorithms. The enhanced model exhibited a 7.3% boost in small target detection precision on self-constructed datasets, alongside accelerated inference speed surpassing mainstream counterparts. This research delivers a swift and highly precise detection algorithm tailored for small airborne target detection, contributing significantly to advancing the application of related algorithms in aerospace engineering.

**Key Words:** aerial object; object detection; computer vision; deep learning; loss function

---

**Received:** 2023-06-30; **Revised:** 2023-10-09; **Accepted:** 2023-11-06

**Foundation item:** Aeronautical Science Foundation of China(201946051002);National Natural Science Foundation of China(52127809)