

基于改进YOLOv8算法的密集型小目标检测

张弛, 游浩, 高谦, 张光明, 谈栋华, 彭菊红

(湖北大学人工智能学院, 湖北 武汉 430062)

摘要: 为了解决无人机视角下场景复杂多变、目标尺度变化大且大多数为小目标等问题, 提出一种改进的YOLOv8算法。首先, 将传统的FPN替换为BiFPN, 通过其双向信息流和加权特征融合机制, 能够更有效地整合来自网络不同层级的信息; 其次, 将检测头替换为Dynamic Head, 通过动态融合机制适应不同尺度的特征, 从而提升对各种尺寸目标的检测能力; 最后, 引入Soft-NMS, 通过降低重叠框的置信度, 更有效地区分和保留紧密排列的目标, 进而提升模型对密集型小目标的检测性能。结果表明, YOLOv8-BDS算法在VisDrone2019数据集上的平均精度值($mAP@IoU=0.5$)达到了54.2%, 相比原基准网络提升了6.8个百分点。该算法能够更好地满足密集型小目标的检测需求。

关键词: 双向特征金字塔网络; YOLOv8; Dynamic Head; 密集型小目标检测; Soft-NMS

DOI: 10.11907/rjdk.241524

中图分类号: TP391

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)008-0196-05



Dense Small Object Detection Based on Improved YOLOv8 Algorithm

ZHANG Chi, YOU Hao, GAO Qian, ZHANG Guangming, TAN Donghua, PENG Juhong

(School of Artificial Intelligence, Hubei University, Wuhan 430062, China)

Abstract: In order to solve the problems of complex and variable scenes from the perspective of drones, large changes in target scale from their perspective, and mostly small targets, an improved YOLOv8 algorithm is proposed. Firstly, changing the traditional FPN to BiFPN, through its bidirectional information flow and weighted feature fusion mechanism, can more effectively integrate information from different levels of the network; Secondly, the detection head is replaced with Dynamic Head, which adapts to features of different scales through dynamic fusion mechanism, improving the detection ability for targets of various sizes; Finally, by introducing Soft-NMS and reducing the confidence of overlapping boxes, it is possible to more effectively distinguish and preserve these tightly arranged targets, thereby improving the detection performance of the model for dense small targets. The results indicate that the YOLOv8-BDS algorithm has an average accuracy value on the VisDrone2019 dataset ($mAP@IoU=0.5$) achieved 54.2%, which is 6.8 percentage points higher than the original benchmark network. This algorithm can better meet the requirements of dense small object detection.

Key Words: BiFPN; YOLOv8; Dynamic Head; dense small target detection; Soft-NMS

0 引言

随着现代社会信息化和工业化的融合发展, 无人机航拍图像中目标的精确检测是实现无人机智能感知在多个实际场景被广泛应用的关键技术之一。文献[1]介绍了由于拍摄距离远、视野宽广, 导致拍摄的图像具有检测目标小、分布密集以及背景复杂等特点。因此, 实现复杂背景

下无人机图像中小目标的准确检测具有一定的挑战性。传统目标检测算法受人工设计的干扰严重, 导致算法泛化性和鲁棒性不足, 难以在小尺度上获得高精度的定位, 且容易受到噪声的干扰, 无法在复杂多变的场景中有效提取小目标的特征。随着深度学习技术的飞速发展, 研究人员已经提出了许多小目标检测算法。文献[2-7]介绍了包括基于双阶段检测的RCNN系列算法、基于单阶段检测器的SSD系列算法和YOLO系列算法。其中, 基于区域候选网

收稿日期: 2024-06-11

扫描二维码阅读全文:



作者简介: 张弛(2000-), 男, 湖北大学人工智能学院硕士研究生, 研究方向为人工智能算法设计; 游浩(1999-), 男, 湖北大学人工智能学院硕士研究生, 研究方向为计算机视觉; 高谦(2001-), 女, 湖北大学人工智能学院硕士研究生, 研究方向为人工智能; 张光明(1999-), 男, 湖北大学人工智能学院硕士研究生, 研究方向为自然语言处理; 谈栋华(2001-), 男, 湖北大学人工智能学院硕士研究生, 研究方向为语义分割; 彭菊红(1978-), 女, 博士, 湖北大学人工智能学院副教授, 研究方向为信号处理、电路建模与分析、人工智能算法。本文通讯作者: 彭菊红。

络的RCNN系列算法产出一系列的稀疏矩阵的候选框,容易引起无人机图像中密集小目标特征的丢失。单阶段检测器的SSD算法有着端到端的性能优势,但用低级feature去检测小目标,存在小目标特征提取不充分、检测精度有限的问题。无人机视角下场景复杂多变,目标尺度变化大,小目标数量多,传统的目标检测框架难以取得理想的检测效果。为此,本文对YOLO系列算法进行了改进,在YOLOv8的基础上提出YOLOv8-BDS(YOLOv8-Bidirectional Feature Pyramid Network, Dynamic Head and Soft-Non-Maximum Suppression)。

1 改进的YOLOv8算法

本文从检测速度和检测精确度两个方面综合考虑,使用YOLOv8s来进行密集小目标检测。改进后的模型命名为YOLOv8-BDS。改进算法在原始算法的基础上,提出BiFPN模块替代原模型中的FPN模块。文献[8]介绍了相较于旧模块,新模块可以使低层细节特征向高层传递,从而有利于小目标的检测。使用基于注意力机制的目标检测头DYHea代替原始的检测头,文献[9]介绍了该方法通过在尺度感知的特征层、空间感知的空间位置以及任务感知的输出通道内连贯地结合多头self-attention注意力机制,显著提高了目标检测头的表示能力。文献[10]介绍了在后处理部分引入Soft-NMS方法,减少漏检,并增强密集目标检测的能力。YOLOv8-BDS模型的结构如图1所示。

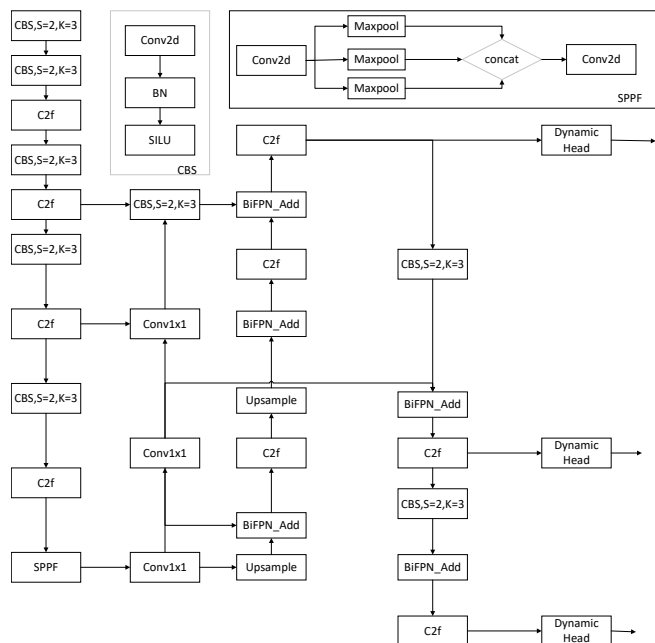


Fig. 1 YOLOv8-BDS structure

图1 YOLOv8-BDS结构图

1.1 重复加权双向特征金字塔网络

在YOLOv8网络中,文献[11]介绍了传统的FPN主要采用自顶向下的信息流,这意味着高层语义丰富的特征被

逐层传递到低层。尽管这种方法能够增强低分辨率特征图的语义信息,但它忽视了低层细节特征向高层传递的可能性,可能对细节丰富的小目标检测效果有限。针对这一问题,本文在YOLOv8中引入BiFPN代替传统的FPN。BiFPN采用一个多层次的结构,其中包括自上而下和自下而上的信息流。这种双向结构使得特征层之间可以互相传递和加强信息,从而充分利用高层次的语义信息和低层次的细节信息。在自上而下的传递过程中,高层特征首先通过上采样和 1×1 卷积进行处理,然后与相应的底层特征相融合;在自下而上的传递过程中,融合后的特征通过下采样和 1×1 卷积,再与下一层的高层特征融合。同时,BiFPN引入了可学习的权重因子,用于调整不同特征层在融合过程中的贡献程度。这些权重通过训练过程自动学习,使模型能够根据不同任务和特性调整特征的重要性。在每次特征融合点,BiFPN会计算来自不同路径(如上采样路径和水平连接)的特征图的加权和。权重是通过softmax函数归一化,保证不同特征图贡献的平衡。BiFPN通过双向信息流和加权融合机制,比传统FPN更有效地利用了多尺度的信息。这使得BiFPN在处理具有复杂尺度变化的图像时能更好地保留细节和语义信息。加权特征融合机制允许模型根据不同的数据和任务需求动态调整特征层之间的相互作用,提高了模型的适应性和鲁棒性。尽管BiFPN引入了更多的特征交互,但其设计优化了计算路径和参数使用,使得整体计算效率得到提升。BiFPN结构图如图2所示。

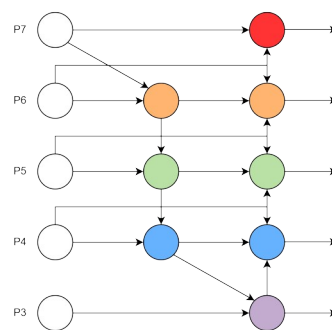


Fig. 2 BiFPN structure

图2 BiFPN结构图

1.2 基于注意力机制的目标检测头

在传统的目标检测网络中,多尺度特征融合通常是静态的,每个尺度的特征处理方式是固定的。普通检测头通常在特征金字塔的较低分辨率层级上进行预测,这对于大型目标是足够的,但对于小目标,这样的特征图缺乏足够的细节,导致小目标的特征难以捕捉和正确识别。针对这一问题,本文提出DYHea代替原始的检测头。DYHea设计了一种动态融合机制,可以根据输入特征的不同尺度动态调整融合策略。这种机制允许模型通过学习到的权重参数来自适应地调整不同尺度特征的融合方式,优化特征的组合。融合机制通常通过一个可学习的权重矩阵来实现,该矩阵针对每种特征组合有不同的权重配置,使得模

型能够在训练过程中自主学习如何最有效地结合来自不同层次的特征。同时, DyHead引入了注意力模块, 特别是自注意力机制, 用于加强模型对特征中重要部分的关注。这些注意力模块可以在不同的特征尺度上独立运作, 进一步提升了特征表达的质量和相关性。注意力模块通常包括多个头(head), 每个头负责计算特征图中不同部分的注意力得分。通过这种方式, 模型不仅能够突出重要的特征, 还能抑制不相关的信息, 从而在复杂的场景中保持较高的识别精度。DyHead结构图如图3所示。

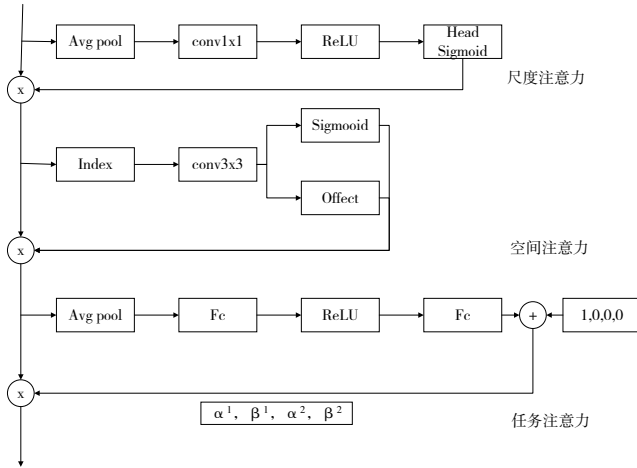


Fig. 3 DyHead structure

图3 DyHead 结构

1.3 平滑非最大抑制算法

在密集小目标环境中, 目标之间的重叠更为常见。YOLOv8中传统的NMS可能会错误地剔除有效目标, 因为一旦两个边界框的交并比(IoU)超过了阈值, 置信度较低的框就会被完全移除。针对这一问题, 本文提出的Soft-NMS通过降低重叠框的置信度而不是完全移除它们, 这样即使框之间存在较大的重叠, 也能够保留更多的真实目标。在处理密集小目标时, 这一点尤其重要, 因为每一个目标都可能非常关键, 而且目标的丢失更容易影响整体性能。在高密度的小目标场景中, 传统的NMS可能将相邻的、重叠的目标视为一个单一目标而将其剔除。Soft-NMS通过逐渐调整置信度, 可以更有效地区分和保留这些紧密排列的目标从而提升模型对密集型小目标的检测性能。Soft-NMS计算降低置信度的权重常用两种方法: 线性法和高斯法。公式(1)为原始NMS计算置信度分数的方法。

$$s_i = \begin{cases} s_i, & \text{IoU}(M, b_i) < N_i \\ 0, & \text{IoU}(M, b_i) \geq N_i \end{cases} \quad (1)$$

式中: M 代表当前得分最高的预测框; b_i 表示待处理的预测框; N_i 是一个预先设定的IoU阈值。当 b_i 与 M 的IoU大于等于 N_i 时, 认定 b_i 与 M 重叠过多, 将其置信度分数置零, 即剔除 b_i ; 当 b_i 和 M 的IoU小于 N_i , 保持其分数不变。公式(2)为Soft-NMS采用的线性惩罚函数。

$$s_i = \begin{cases} s_i, & \text{IoU}(M, b_i) < N_i \\ s_i(1 - \text{IoU}(M, b_i)), & \text{IoU}(M, b_i) \geq N_i \end{cases} \quad (2)$$

Soft-NMS对原始的NMS进行了显著改进。在处理与最高得分预测框交并比大于等于阈值的候选框 b_i 时, 并不直接将其置信度分数置为零, 而是根据交并比的大小线性降低其置信度分数。这意味着, 与最高得分预测框 M 的交并比越大的候选框 b_i 将经历越明显的置信度下降, 受到越强的抑制。然而, 这种线性惩罚函数在阈值附近是不连续的, 可能导致处理后的预测框集合中出现分数断层。为了解决这个问题, Soft-NMS引入了高斯惩罚函数, 如公式(3)所示。

$$s_i = s_i e^{-\frac{\text{IoU}(M, b_i)^2}{\sigma}}, \forall b_i \notin D \quad (3)$$

式中: D 表示用来存放最终 box 的集合, 当 b_i 与 M 的交并比较大时, 惩罚就较大; 交并比较小时, 惩罚也逐步降低, 当交并比为0时, 惩罚为零。该函数通过平滑的曲线来降低那些与 M 的交并比接近阈值的预测框的分数, 使分数的变化更加连续, 避免出现断层现象。

2 实验结果分析

2.1 数据集介绍

文献[12]介绍了本文使用的数据集是VisDrone2019, 由天津大学 AISKEYE 团队收集的无人机航拍图像数据集。该数据集包含10个不同的类别, 包括行人、人、汽车、面包车、公共汽车、卡车、货车、自行车、遮阳三轮车和三轮车。该数据集包含了261 908个目标实例样本, 分为6 471张训练集图像、548张验证集图像和3 190张测试集图像。行人和远处物体的标注非常小且密集, 通过对数据集的分析, 可以观察到它主要是由小对象组成, 因此该数据集非常适合进行小目标检测算法的研究。

2.2 环境配置

本文实验设备为NVIDIA GeForce RTX3060ti, 显存为12 GB, 实验环境选择基于PyTorch。相关参数设置如epochs为200, batch size为32, 输入尺寸为1 600x1 600等, 利用官方提供的预训练模型开始训练。

2.3 评价指标

本文模型除了具备良好的准确度, 还应兼顾模型检测速度要求。因此, 文献[13]介绍了采用平均精度(average precision, AP)作为每个缺陷类别的评估指标; 平均精度均值(mean average precision, mAP)评价整个网络模型的性能; 刷新率(frames per second, FPS)作为模型检测速度评价。精确度、召回率和模型参数量作为辅助性指标。

2.4 模型对比评估

2.4.1 消融实验

为了验证本文改进算法在VisDrone2019数据集上的有效性, 本文设计了6组实验。第一组为原始的YOLOv8s

算法;第二组加入 BiFPN(结果命名为 YOLOv8-B);第三组在原模型的基础上单独加入 DyHead;第4组在原模型的基础上单独加入 Soft-NMS。前四组实验旨在验证单独添加本文不同改进方式的有效性。为进一步验证不同改进方式组合对算法有效性的影响,本文在第二组实验的基础上加入 DyHead 检测头作为第五组实验(结果命名为 YOLOv8-BD);在第五组实验的基础上引入 Soft-NMS(结果命名为 YOLOv8-BDS)。如表 1 所示,在原模型的基础上加入 BiFPN 后,mAP0.5 值提升了 0.8 个百分点,证明使用 BiFPN 代替原有的 FPN 可以更有效检测小目标。在此基础上,DyHead 检测头通过动态融合机制适应不同尺度的特征,从而可以提升对小目标的检测性能。通过数据分析,在第二组实验的基础上加入 DyHead 检测头后,mAP0.5 值提升了 1.2 个百分点。而在第五组实验的基础上引入了 Soft-NMS,其通过降低重叠框的置信度而非完全移除它们的思想,即使在小目标很密集的情况下仍能更多地保留有效目标,从而提升检测精度。数据分析显示,mAP0.5 值提升了 4.8 个百分点,表明 Soft-NMS 的引入显著提升了对小目标的检测性能。

Table 1 Ablation experiment results (VisDrone2019 dataset)

表 1 消融实验结果 (VisDrone2019 数据集)

Modle	Parameters/M	GFLOPs	mAP0.5	mAP0.5-	FPS/S
			/%	0.95/%	
YOLOv8s	11.10	179.20	47.4	29.0	140
yolov8s+BiFPN	7.58	166.56	48.2	29.1	113
yolov8s+Dyhead	13.07	205.96	48.7	29.6	77
yolov8s+Soft-NMS	11.10	28.70	52.4	34.5	22
+BiFPN+Dyhead	10.26	199.95	49.4	30.0	67
+BiFPN+Dyhead+S	10.26	199.95	54.2	35.7	18

2.4.2 VisDrone2019 数据集对比实验

为了验证本文改进算法的优越性和有效性,在 Vis-Drone2019 数据集上将本文算法与主流目标检测算法及其他优秀模型的改进算法进行了对比实验。从表 2 可以看出,传统的目标检测方法如 SSD、Faster-RCNN、YOLOv3 等不仅检测效果不理想,而且参数量过高,不适合小目标的研究。而在最新的改进算法中,YOLOv7 系列虽然改进精度有所提升,但与 YOLOv8 算法相比,其参数量仍然较高。本文提出的基于 YOLOv8 的改进算法相较于其他算法在检测精度上显著优于其他算法,同时参数量也较其他系列算法有所降低。但由于修改图片的尺寸会增加算法的计算量,其检测速度会有一定的下降。综上所述,本文所提出的基于 YOLOv8 的改进算法在密集小目标检测中表现出色。

2.4.3 可视化结果

图 4 展示了本文改进模型与基线模型在检测结果上的对比。上方两张图呈现了基线模型的检测效果,下方两张图则展示了改进模型的检测结果。通过观察图中用白色椭圆标注的区域,可以明显看出,改进后的模型能够识别出更多的目标,这表明其检测性能得到了显著提升。

Table 2 Comparison of different algorithms on VisDrone2019 test set

表 2 不同算法在 VisDrone2019 测试集对比

Modle	Parameters/M	mAP0.5/%	mAP0.5-0.95/%	FPS/S
SSD	24.5	24.1	18.8	21
Faster-RCNN	41.3	33.8	21.4	15
YOLOv3	63.1	36.0	19.4	33
YOLOv5s	7.0	34.1	18.6	66
YOLOv7 ^[14]	30.4	46.4	27.1	38
YOLOv7 ^[15]	34.2	45.8	26.3	/
YOLOv8n	1.9	38.2	22.9	81
YOLO8s ^[16]	7.4	48.8	29.5	62
RT-DETR 网络 ^[17]	28.1	47.1	28.8	/
YOLOv7 ^[18]	/	52.2	30.3	/
YOLOv8s ^[19]	7.6	50.1	31.1	/
YOLOv8s	11.1	47.4	29.0	140
YOLOv8-BDS	10.3	54.2	35.7	18



Fig. 4 Comparison of visualization results

图 4 可视化结果对比

3 结语

本文提出了一种基于 YOLOv8 改进的小目标检测算法,旨在提高小目标的检测性能。首先,采用 BiFPN 模块代替基准网络中的 FPN 模块,极大减少了采样过程中的信息丢失;其次,引入基于注意力机制的目标检测头 DY-Head,通过在尺度感知的特征层、空间感知的空间位置以及任务感知的输出通道内连贯地结合多头 self-attention 注意机制,显著提高了目标检测头的表示能力;最后,利用 Soft-NMS 方法进行后处理,进一步提升密集目标检测能力,减少了漏检。实验结果表明,本文提出的小目标检测模型在 VisDrone2019 数据集上提高检测精度的同时也保证了检测的实时性,能够满足实际应用场景的需求。此

外,在实验阶段,文献[20]提出了通过生成对抗网络生成超分辨率图像的方法,但由于参数量过大,无法满足实时性要求。因此,未来可以考虑采用更轻量化的骨干网络和卷积模块进行替代,并进一步研究如何在尽量减少网络参数的同时最大化检测精度。

参考文献:

- [1] PENG C, ZHU M, REN H, et al. Small object detection method based on weighted feature fusion and CSMA attention module [J]. Electronics, 2022, 11(16): 25-46.
- [2] SUN S Y, YANG Z. Target detection algorithm based on improved Faster RCNN[J]. Computer and Digital Engineering, 2022, 50(12): 2654-2659.
孙顺远, 杨镇. 基于改进Faster RCNN的目标检测算法[J]. 计算机与数字工程, 2022, 50(12): 2654-2659.
- [3] CHEN D H, SUN S R, WANG Y C, et al. Study on small target detection with improved SSD algorithm[J]. Sensors and Microsystems, 2023, 42(3): 65-68, 72.
陈德海, 孙仕儒, 王昱朝, 等. 改进SSD算法的小目标检测研究[J]. 传感器与微系统, 2023, 42(3): 65-68, 72.
- [4] REDMON J, FARHADI A. YOLOv3: an incremental improvement[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1-6.
- [5] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: optimal speed and accuracy of object detection[DB/OL]. <https://arxiv.org/abs/2004.10934>.
- [6] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications [DB/OL]. <https://arxiv.org/abs/2209.02976>.
- [7] 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 [DB/OL]. <https://arxiv.org/abs/2207.02696>.
- [8] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection[C]//Seattle 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [9] DAI X, CHEN Y, XIAO B, et al. Dynamic head: unifying object detection heads with attentions[C]//Kuala Lumpur: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [10] GAO G, SHUAI C, WANG S, et al. Using improved YOLO V5s to recognize tomatoes in a continuous working environment[J]. Signal, Image and Video Processing, 2024, 18(5): 4019-4028.
- [11] WANG S, LI Y, QIAO S. ALF-YOLO: enhanced YOLOv8 based on multi-scale attention feature fusion for ship detection[J]. Ocean Engineering, 2024(8): 118-233.
- [12] SHEN S R, ZHANG X, YAN W, et al. An improved UAV target detection algorithm based on ASFF-YOLOv5s[J]. Mathematical Biosciences and Engineering: MBE, 2023, 20(6): 10773-10789.
- [13] KUN L, NANA L, AIMEI L, et al. A new self-reference image de-composition algorithm for strip steel surface defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(7): 4732-4741.
- [14] CUI L Q, CAO H W. Improving YOLOv7 for aerial image object detection [J/OL]. Computer Engineering and Applications, 2024-05-28. <https://link.cnki.net/urlid/11.2127.TP.20240527.1251.008>.
崔丽群, 曹华维. 改进YOLOv7的航拍图像目标检测[J/OL]. 计算机工程与应用, 2024-05-28. <https://link.cnki.net/urlid/11.2127.TP.20240527.1251.008>.
- [15] NIE Y, LAI H C, GAO G X. Improving YOLOv7+Bytetrack for small object detection and tracking [J/OL]. Computer Engineering and Applications, 2024-03-10. <https://link.cnki.net/urlid/11.2127.TP.20240308.1703.014>.
聂源, 赖惠成, 高古学. 改进YOLOv7+Bytetrack的小目标检测与追踪[J/OL]. 计算机工程与应用, 2024-03-10. <https://link.cnki.net/urlid/11.2127.TP.20240308.1703.014>.
- [16] FU J Y, ZHANG Z J, SUN W, et al. Improved YOLOv8 algorithm for detecting small targets in aerial images [J]. Computer Engineering and Applications, 2024, 60(6): 100-109.
付锦燧, 张自嘉, 孙伟, 等. 改进YOLOv8的航拍图像小目标检测算法[J]. 计算机工程与应用, 2024, 60(6): 100-109.
- [17] SUN F, HE N, LI R, et al. GD-PAN: a multiscale fusion architecture applied to object detection in UAV aerial images [J]. Multimedia Systems, 2024, 30: 143.
- [18] XU H Z, GU X N. Optimization of traffic small target image detection algorithm from the perspective of drones [J/OL]. Computer Engineering and Applications, 2024-06-06. <https://link.cnki.net/urlid/11.2127.TP.20240605.1427.008>.
徐慧智, 古旭楠. 无人机视角下交通小目标图像检测算法优化研究[J/OL]. 计算机工程与应用, 2024-06-06. <https://link.cnki.net/urlid/11.2127.TP.20240605.1427.008>.
- [19] JIANG W, WANG W H, YANG J J. AEM-YOLOv8s: small object detection in drone aerial images [J/OL]. Computer Engineering and Applications, 2024-06-06. <https://link.cnki.net/urlid/11.2127.tp.20240605.1040.004>.
蒋伟, 王万虎, 杨俊杰. AEM-YOLOv8s: 无人机航拍图像的小目标检测[J/OL]. 计算机工程与应用, 2024-06-06. <https://link.cnki.net/urlid/11.2127.tp.20240605.1040.004>.
- [20] HU F Y, WANG Q L, ZHAO Q R. Color blindness test graph generation based on generative adversarial networks [J/OL]. Computer Simulation, 2024-06-07. <https://link.cnki.net/urlid/11.3724.TP.20240607.1010.004>.
胡飞宇, 王青龙, 赵倩茹. 基于生成对抗网络的色盲测试图生成[J/OL]. 计算机仿真, 2024-06-07. <https://link.cnki.net/urlid/11.3724.TP.20240607.1010.004>.

(责任编辑: 陈维捷)