

DOI:10.19651/j.cnki.emt.2518342

改进 YOLOv8n 的航拍多尺度目标检测模型^{*}

贾亮 陈茂辉 王琪 徐城

(沈阳航空航天大学电子信息工程学院 沈阳 110136)

摘要: 针对无人机航拍图像中目标小而密集,且极易出现漏检和误检等情况,文章提出一种改进 YOLOv8n 的无人机航拍复杂背景下多尺度目标检测模型 UCM-YOLOv8。首先设计一种集聚与扩散特性于一体的金字塔式网络结构,让每个尺度的特征都具有详细的上下文信息。其次,提出任务动态对齐检测头,从多个卷积层中学习交互特征,提高检测精度。此外,将卷积加性自注意力机制与 C2f 模块有效融合,进一步增强特征表达能力。最后,为抑制低分辨率图像产生的有害梯度,运用 Wise-Inner 损失函数替换原 CIou 损失函数。在 VisDrone2019 数据集上进行对比实验和消融实验,mAP50 值较基线模型提升了 10.8%,参数量减少了 9.6%。证明本模型在无人机航拍视角下的小目标检测效果优异,适合无人机航拍图像的应用。

关键词: 无人机航拍;目标检测;损失函数;空间金字塔;解耦检测头;多尺度

中图分类号: TP391.41;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Improved multi-scale target detection model for aerial photography based on YOLOv8n

Jia Liang Chen Maohui Wang Qi Xu Cheng

(School of Electronic Information Engineering, Shenyang Aerospace University, Shenyang 110136, China)

Abstract: To address the challenges of small and densely packed targets in drone aerial images, which are prone to missed and false detections, this paper proposes an improved multi-scale target detection model, UCM-YOLOv8, based on YOLOv8n, for complex backgrounds in drone aerial photography. Initially, a pyramid network structure that integrates aggregation and diffusion mechanisms is designed, enabling features at each scale to capture detailed contextual information. Second, a task dynamic alignment detection head is introduced to learn interactive features from multiple convolutional layers, enhancing detection precision. Furthermore, the effective integration of the convolutional additive self-attention mechanism with the C2f module further strengthens the network's feature representation capacity. Finally, the Wise-Inner loss function is employed to replace the original CIou loss function, suppressing harmful gradients caused by low-resolution images. The proposed model was validated through comparative and ablation experiments on the VisDrone2019 dataset. Results show a 10.8% improvement in mAP50 over the baseline model and a 9.6% reduction in parameters. These findings demonstrate the model's superior performance in detecting small targets from drone perspectives, making it well-suited for drone aerial image applications.

Keywords: drone aerial photography; target detection; loss function; spatial pyramid; decoupled detection head; multi-scale

0 引言

无人机是一种成本低、体积小、灵活度高、便于携带且操作简易的设备,能够在各种复杂或危险的场景中开展任务^[1]。近年来,无人机技术迅速发展,无人机航拍图像检测技术已广泛应用于交通监测^[2]、农业^[3]、灾害救援^[4]等生活

领域,以及海上船舶检测等军事领域^[5]。

目前,基于深度学习的目标检测算法主要有两大类:二阶段目标检测算法和一阶段目标检测算法。二阶段算法以区域卷积神经网络(region-based convolutional neural network, R-CNN)^[6]系列算法为代表,虽然检测精度较高,但是其检测速度较慢,模型占用的内存量大,无法满足实时

收稿日期:2025-03-14

* 基金项目:国家自然科学基金(61671310)项目资助

检测的需求。一阶段目标检测算法主要有 YOLO (you only look once, YOLO)^[7] 系列和单次多边框检测 (single shot multibox detector, SSD)^[8] 等, YOLO 系列算法检测速度较快。因此, 诸多研究者选择基于一阶段算法进行无人机航拍目标检测, 并展开了一系列研究。

刘涛等^[9]解决了由于小目标图像背景复杂以及分布密集而导致的误检和漏检的问题, 在 YOLOv5 模型的基础上, 改进了锚框的聚类算法, 提高了召回率, 从而提升了小目标的捕捉能力。张徐等^[10]改进了 YOLOv7 算法, 通过在主干中融入 Swin Transformer 模块, 显著提升了模型提取小目标特征的能力, 改进后的算法能较好的检测目标。钟师等^[11]在 YOLOv8 上通过添加动态检测头, 在特征的水平、空间和通道 3 个维度上部署注意力机制, 以降低远景小目标检测时的漏检、误检率, 提高检测性能。Sui 等^[12]以及范博淦等^[13]将 YOLOv8 颈部 FPN+PAN 金字塔网络替换为双向加权特征金字塔网络, 以增强特征融合, 提升了小目标检测的精度, 然而, 其特征融合策略主要依赖同层级特征交互, 缺乏对跨层级语义关联的充分建模, 导致上下文信息利用率不足, 限制了复杂场景下的检测效果提升。梁秀满等^[14]在 YOLOv8 颈部网络中加入了小目标检测层, 并在主干网络中加入了 LSK 注意力机制^[15], 丰富了特征提取, 但在实际应用中这种方式容易导致参数量以及计算量过大, 难以在边缘计算设备等资源受限场景中推广。

综上所述, 尽管现有的航拍目标检测算法在提高检测精度方面提供了切实可行的改进策略, 然而, 现阶段对于无人机航拍小目标检测精度方面仍然面临着巨大挑战, 不难发现由于无人机在飞行过程中通常需要在较高的空中拍摄小目标, 这会导致拍摄的图像分辨率较低, 使得目标检测算法难以从低质量图像中提取出有效的特征并进行特征融

合。为了提高无人机航拍图像小目标的检测精度, 提出了一种基于 YOLOv8n 改进的无人机航拍图像多尺度目标检测模型 UCM-YOLOv8, 在 Visdrone2019 数据集上验证模型改进后的性能。研究作出的贡献如下:

1) 提出一种聚合扩散机制, 并设计基于此机制的聚合扩散金字塔网络结构 (aggregation and diffusion pyramid network, ADPN), 旨在加强多尺度的特征融合, 提高模型对小目标的感知能力和准确性。

2) 提出一种任务动态对齐检测头 (task dynamic alignment detection head, TDADH), 弥补原检测层提取小目标特征能力的不足, 显著提高微小目标的检测精确率。采用共享卷积, 具有多尺度感知能力, 能有效地减少参数量以便于实际应用部署。

3) 引入 AdditiveBlock 改进 C2f 模块, 增强了对小目标的检测能力, 使特征提取更为灵活。

4) 设计 Wise-Inner-IoU 替换原 CIoU 损失函数, 有效降低了高质量样本和低质量样本的梯度增益, 进一步提高了目标定位的准确性。

1 算法改进

1.1 YOLOv8 算法

YOLO 系列是单阶段检测算法, 经历了几次升级迭代后, 具有检测速度快、实时性好的优点。YOLOv8 算法主要由输入层、主干部分、颈部网络以及检测头等主要部分组成, 分为 5 个大小不同的模型, YOLOv8n 网络模型是这五个模型中参数量最少并且复杂度最小的模型, 具有卓越的检测速度和较大的精度提升空间, 这使得它成为部署在无人机等资源有限环境中的理想选择。YOLOv8 网络结构图如图 1 所示。

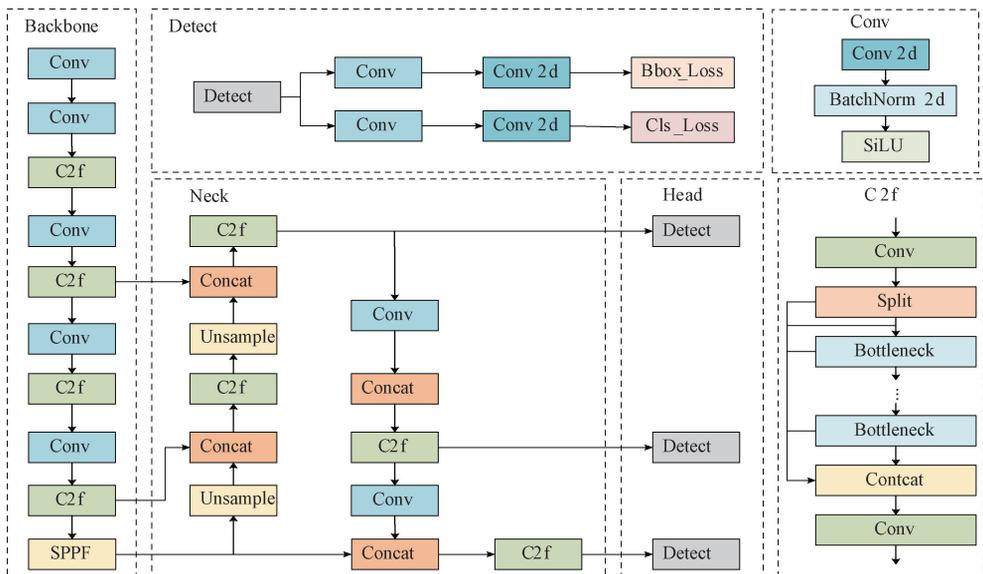


图 1 YOLOv8 网络结构图

Fig. 1 YOLOv8 network structure diagram

1.2 改进 YOLOv8n 算法

针对传统无人机航拍目标检测算法存在漏检高、精度低等问题,本文提出了一种航拍图像多尺度目标检测模型 UCM-YOLOv8。该模型在 YOLOv8n 的基础上进行改进,首先,设计 ADPN 金字塔网络结构,提高模型对输入数据的多尺度特征提取。其次,设计 TDADH 检测头

替换掉原有的检测头,在降低模型参数数量的同时提升了检测精度,使模型更具轻量化。此外,引入 AdditiveBlock 到 C2f 模块中,重构了 C2f 模块。最后,改进损失函数得到 Wise-Inner-IoU,优化了原损失函数存在的缺点,提高了模型对小目标的友好性。UCM-YOLOv8 网络结构如图 2 所示。

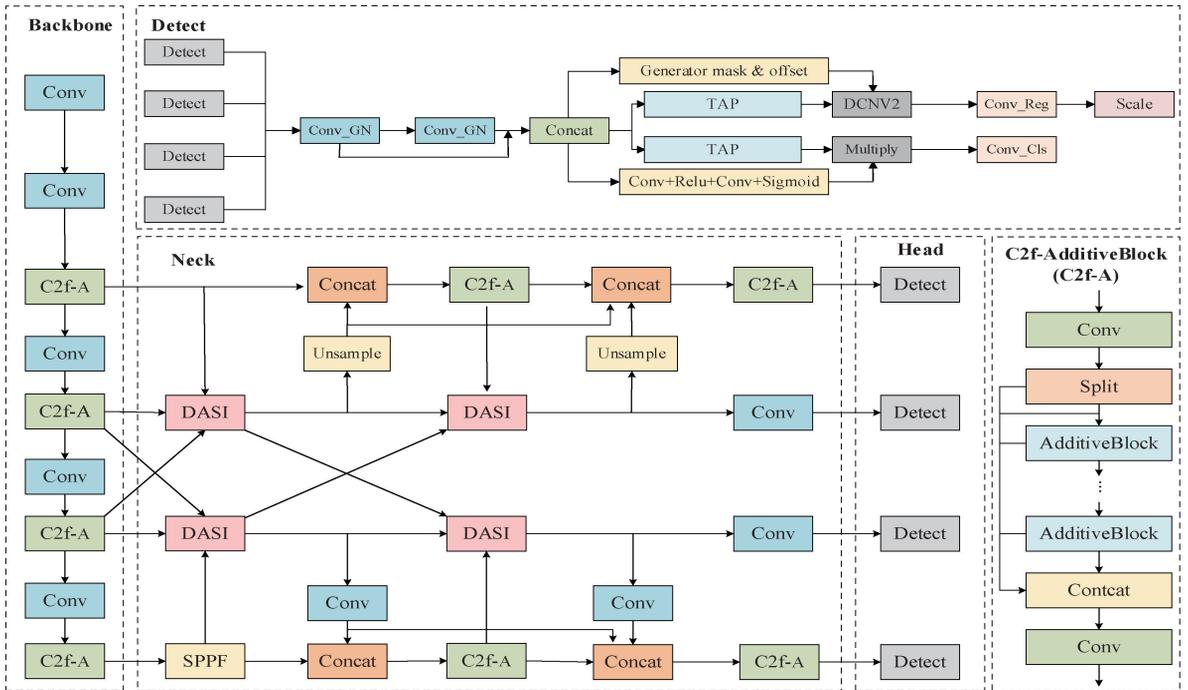


图 2 UCM-YOLOv8 网络结构图

Fig. 2 UCM-YOLOv8 network structure diagram

1.3 ADPN 特征聚合扩散金字塔网络

在 YOLOv8 原算法模型中,特征融合网络采用了路径聚合网络 PANet^[16] 结构,如图 3 所示。尽管在 YOLOv8 网络中,这种结构的机制可以更好地进行特征融合,但以顺序的方式进行融合的特征会更多的关注本层以及上层分辨率,较少的关注到其他分辨率,从而导致其他层的特征在融合过程中稀释,不利于目标检测中对于浅层特征的需要。

为了减少在特征融合过程中导致的特征丢失,本文根据 HCF-Net^[17] 网络提出一种聚合扩散机制,基于此机制设计 ADPN 金字塔网络结构,并加入了小目标检测层,旨在加强多尺度的特征融合,提高模型对小目标的感知能力和准确性。ADPN 结构图如图 4 所示。

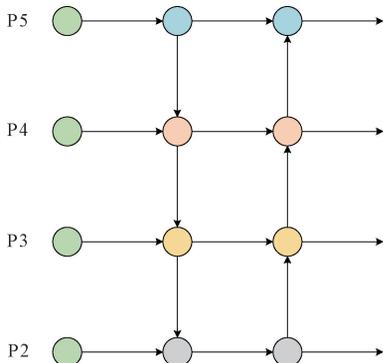


图 3 PANet 结构图

Fig. 3 Structure diagram of PANet

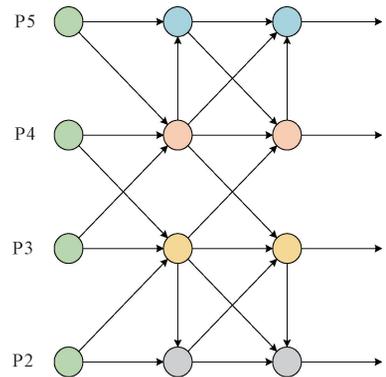


图 4 ADPN 结构图

Fig. 4 Structure diagram of ADPN

ADPN 的具体融合过程如下:以 P2、P3、P4 层为例,将 P2、P3、P4 层中的原始语义特征通过维度感知选择性集成

模块(dimension-aware selective integration, DASI)进行第一次聚合成 P_3^{id} 。随后通过扩散机制,将富含上下文信息的 P_3^{id} 跳跃扩散到 P2 层和 P4 层,扩散机制包括一次上采样操作,这种机制缓解了原网络中因连续上下采样导致的特征丢失问题。再将 P_2^{id} 、 P_3^{id} 、 P_4^{id} 通过 DASI 模块进行第二次聚合成 P_3^{out} 。P3、P4、P5 层聚合扩散过程与其相似,唯一不同的是将扩散过程中的采样操作换为卷积操作。

在特征融合过程中,ADPN 的聚合功能通过 DASI 进行实现。DASI 模块通过一种新颖的信道分区选择机制,使其能够根据物体的大小和特征自适应地选择合适的特征进行融合。具体来说,DASI 通过卷积和双线性插值操作,将高维特征 $F_h \in \mathbb{R}^{H_h \times W_h \times C_h}$ 和低维特征 $F_l \in \mathbb{R}^{H_l \times W_l \times C_l}$ 性插值与当前层的特征 $F_u \in \mathbb{R}^{H \times W \times C}$ 对齐。然后,将这些特征在通道维度 C 上划分成 4 个相等的部分,得到 $(h_i)_{i=1}^4 \in \mathbb{R}^{H \times W \times \frac{C}{4}}$, $(l_i)_{i=1}^4 \in \mathbb{R}^{H \times W \times \frac{C}{4}}$, $(u_i)_{i=1}^4 \in \mathbb{R}^{H \times W \times \frac{C}{4}}$, 其中 h_i 、 l_i 、 u_i 、分别表示高维、低维以及当前层特征的第 i 个分区的特征。分区后的特征通过以下公式计算:

$$\alpha = \text{sigmoid}(u_i) \quad (1)$$

$$u'_i = \alpha l_i + (1 - \alpha)h_i \quad (2)$$

$$F'u = [u'_1, u'_2, u'_3, u'_4] \quad (3)$$

$$\hat{F}u = \delta(\beta(\text{Conv}(F'u))) \quad (4)$$

其中, $\alpha \in \mathbb{R}^{H \times W \times \frac{C}{4}}$ 表示通过 sigmoid 激活函数得到的融合权重,当 $\alpha > 0.5$ 时,模型会优先考虑细粒度特征;当 $\alpha < 0.5$ 时,模型则更侧重于上下文特征。 $u'_i \in \mathbb{R}^{H \times W \times \frac{C}{4}}$, 表示每个分区选择性聚合的结果。然后把所有的 u_i 在通道维度上进行合并,得到 $F'u$,再通过卷积($*$)、批量归一化处理(β)、以及修正线性单元 $ReLU$ (δ)操作,得到最终结果 $\hat{F}u$ 。DASI 模块结构图如图 5 所示。

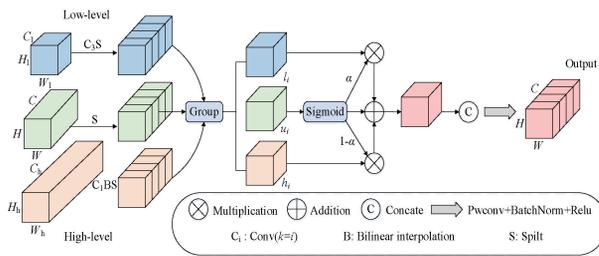


图 5 DASI 模块结构图

Fig. 5 DASI module structure diagram

1.4 任务动态对齐检测头

YOLOv8 的检测头采用 Decoupled Head,每个检测层都通过独立的卷积模块进行特征提取。以颈部网络的输出作为输入,采用解耦头设计分为分类分支和回归分支,其参数量几乎占据了整个模型的 1/5。其次由于航拍图像背景复杂且分辨率低的特性,导致小目标与背景难以区分,在一定程度上影响了检测精度。

针对于此,本文设计了 TDADH 检测头。首先采用参数共享的思想降低了检测头的参数量,这使得模型更轻便,特别是在无人机这类资源受限的设备上。其次,现有的目标检测器头部通常使用独立的分类和定位分支,这会导致两个任务之间缺乏交互,本文参照 TOOD^[18]的思想,在检测头上设计定制任务对齐的结构,实现了两者的相互交互。TDADH 检测头结构如图 6 所示。

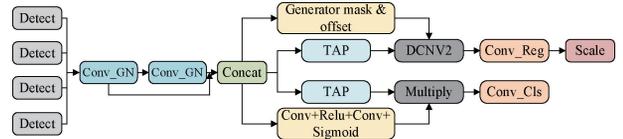


图 6 TDADH 结构图

Fig. 6 Structure diagram of TDADH

具体来说,输入特征图首先通过共享卷积以及归一化处理,利用残差的方式进行拼接,随后分别进入两个定制的任务对齐预测器(task-aligned predictor, TAP),通过其独特的曾注意力机制再层级计算不同任务特定的特征从而进行任务分解,TAP 结构如图 7 所示。对于分类和定位,分别计算其 task-specific 特征:

$$X_k^{task} = \omega_k \cdot X_k^{inter}, k \in \{1, 2, \dots, N\} \quad (5)$$

$$\omega_k = \sigma(fc_2(\delta(fc_1(x^{inter})))) \quad (6)$$

其中, ω_k 是学习到的层注意力权重 $\omega \in \mathbb{R}^N$ 的第 k 个元素, fc_1 和 fc_2 是两个全连接层, σ 是 sigmoid 激活函数, x^{inter} 是对 X_k^{inter} 应用平均池化得到的。随后分类和定位的预测结果根据每个任务的 X^{task} 按以下公式得到。

$$Z^{task} = \text{conv}_2(\delta(\text{conv}_1(X^{task}))) \quad (7)$$

其中, X^{task} 是对 X_k^{task} 进行 concatenate 得到的, Z^{task} 经过 sigmoid 转换为分类得分。最后定位分支再使用 DCNV2^[19] 和 Scale 操作,动态调整目标尺度,以应对每个检测头所检测目标尺寸不一致的问题;分类分支则使用 Multiply 操作进行交互特征以完成动态特征选择。

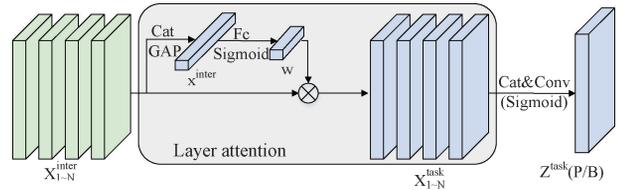


图 7 TAP 模块结构图

Fig. 7 TAP module structure diagram

1.5 C2f-AdditiveBlock 模块

YOLOv8 采用 C2f 模块通过残差方式来集成上下文信息,从而获得更丰富的梯度信息,但由于航拍图像分辨率较低,目标过小并且与背景相似等特点,使得关注点过于分散,导致特征提取难度较大。为解决上述问题,本文提出一种新的特征融合模块命名为 C2f-AdditiveBlock。

卷积加性自注意力机制(conv additive self-attention,

CASA)是一种由 Zhang 等^[20]提出的新型注意力机制,通过简单有效的操作将多种信息进行交互,这种注意力创新性的将相似度函数定义为 $Q \in R^{N \times d}$ 和 $K \in R^{N \times d}$ 的上下文分数之和,即:

$$Sim(Q, K) = \Phi(Q) + \Phi(K), \Phi(Q) = C(S(Q)) \quad (8)$$

其中, $\Phi(\cdot)$ 表示上下文的映射函数,具体化为基于 Sigmoid 函数的通道注意力 $C(\cdot) \in R^{N \times d}$ 和空间注意力 $S(\cdot) \in R^{N \times d}$ 。自注意力输出表示为:

$$O = \Gamma(\Phi(Q) + \Phi(K)) \cdot V \quad (9)$$

其中, $\Gamma(\cdot) \in R^{N \times d}$ 其表示用于整合上下文信息的线性变换。CASA 示意图如图 8 所示。

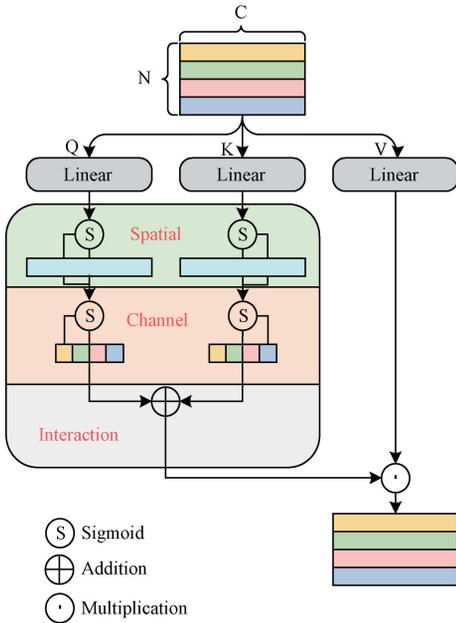


图8 CASA 示意图

Fig.8 Schematic diagram of CASA

C2f-AdditiveBlock 通过将卷积加性自注意力机制结合前馈网络融合到 C2f 模块中,降低背景无关信息的干扰,使得特征提取模块可以更有效地关注到目标本身,提高了检测精度。C2f-AdditiveBlock 结构图如图 9 所示。

1.6 优化损失函数

YOLOv8 结合了 CIoU Loss 和 DFL Loss 来计算边界框回归损失。CIoU 考虑了重叠面积和长宽比等因素计算预测框与真实框的差异,但其没有考虑难易的平衡,当以小目标作为对象时,较小的位置偏移就会导致 IoU 的较大变动。

基于动态非单调机制设计的 Wise-IoUv3^[21] 则弥补了这一缺点。它使用一种合理的梯度增益分配,极大地较少了样本中出现的有害梯度,从而提升了网络模型的泛化能力和整体性能。

Inner-IoU^[22] 通过辅助边界框来计算 IoU 损失,针对不

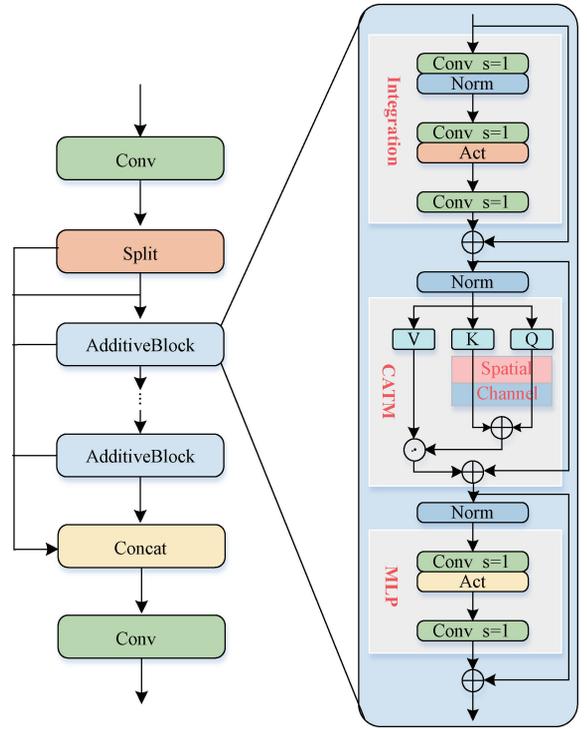


图9 C2f-AdditiveBlock 结构图

Fig.9 Structure diagram of C2f-AdditiveBlock

同的数据集与检测器,通过引入一个尺度因子 ratio 控制辅助边框的尺度大小用于计算损失,从而加速收敛实现更快和更有效的回归效果。Inner-IoU 计算示意图如图 10 所示。

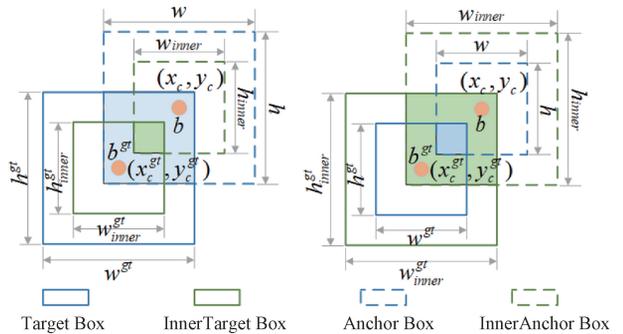


图10 Inner-IoU 计算示意图

Fig.10 Inner-IoU calculation diagram

综上所述,本文将 Inner-IoU 和 Wise-IoUv3 相结合,提出 Inner-WIoU 损失函数,通过调节尺度因子 ratio 生成较大尺度的预测边界框,从而更好地检测小目标。相关公式如下:

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} * ratio}{2} \quad (10)$$

$$b_r^{gt} = x_c^{gt} + \frac{w^{gt} * ratio}{2} \quad (11)$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} * ratio}{2} \quad (12)$$

$$b_b^{gt} = y_c^{gt} + \frac{h^{gt} * ratio}{2} \quad (13)$$

$$b_l = x_c - \frac{w * ratio}{2} \quad (14)$$

$$b_r = x_c + \frac{w * ratio}{2} \quad (15)$$

$$b_t = y_c - \frac{h * ratio}{2} \quad (16)$$

$$b_b = y_c + \frac{h * ratio}{2} \quad (17)$$

$$inter = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l) * \min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)) \quad (18)$$

$$union = (w^{gt} * h^{gt}) * (ratio)^2 + (w * h) * (ratio)^2 - inter \quad (19)$$

$$IoU^{inter} = \frac{inter}{union} \quad (20)$$

$$L_{lmer-WIoU} = L_{WIoUv3} + IoU - IoU^{inter} \quad (21)$$

其中, GT 框和预测框分别表示为 b^{gt} 和 b , 而 b_l, b_r, b_t, b_b 分别表示预测框的上下左右 4 个边界的坐标, x_c, y_c 以及 x_c^{gt}, y_c^{gt} 分别表示预测框和真实框的中心点坐标, $ratio$ 为调节尺度因子, L_{WIoUv3} 则表示 $WIoUv3$ 损失函数。

2 实验结果与分析

2.1 数据集介绍

实验采用公开无人机航拍 VisDrone2019 数据集, 该数据集由天津大学机器学习和数据挖掘实验室 AISKEYEYE 团队收集。使用各种型号的无人机在中国 14 个不同城市捕获, 覆盖范围广泛, 拍摄环境复杂, 目标数量庞大, 类别丰富。由 6 471 张训练图像, 548 张验证图像和 3 190 张测试图像组成。数据集包含 10 类检测对象, 分别为行人、人、自行车、汽车、面包车、卡车、三轮车、遮阳篷三轮车、公交车和摩托车。

2.2 实验环境与参数设置

本次实验平台使用 Ubuntu22.04 操作系统, CPU 型号为 Intel (R) Xeon (R) Platinum 8352V CPU, 主频

2.10 GHz, GPU 采用 RTX 4090 (24 GB)。运行环境为 PyTorch 1.13.0, Cuda 11.7, Python 3.8。实验参数设置如表 1 所示。

表 1 实验参数设置

Table 1 Experiment parameter settings

| epoch | imgsz | batch | lr | momentum |
|-------|-------|-------|------|----------|
| 300 | 640 | 8 | 0.01 | 0.937 |

2.3 评价指标

本次实验采用召回率 (recall, R)、准确率 (precision, P)、平均精度 (mean average precision, mAP) 以及参数量 (parameters, Params) 作为模型的评价指标。预测结果分为正样本和负样本, 真实值为正样本且预测为正样本定义为 TP, 真实值为负样本但预测为正样本定义为 FP, 真实值为负样本且预测为负样本定义为 FN。相关公式如下:

$$P = \frac{TP}{TP + FP} \quad (22)$$

$$R = \frac{TP}{TP + FN} \quad (23)$$

$$AP = \int_0^1 P(R) dR \quad (24)$$

$$mAP = \frac{\sum_{i=0}^n AP_i}{N} \quad (25)$$

其中, 准确率 P 指的是在所有预测结果中正确预测数量所占比例, 召回率 R 表示数据集目标中实际检测出的比例, AP 的数值常为 $P-R$ 图中曲线下方的面积, mAP 是对多类别目标的平均精度求均值, mAP_{50} 为 $IoU=0.5$ 时的 mAP , $mAP_{50:95}$ 指 IoU 阈值取 $0.5-0.95$ 时的 mAP 平均值。

2.4 消融实验

为验证每个改进部分的有效性, 本文使用 YOLOv8n 作为基线模型, 采用 VisDrone2019 作为数据集进行消融实验, 实验结果如表 2 所示。

表 2 消融实验结果

Table 2 Ablation experiment results

| 方法 | C2f-AdditiveBlock | ADPN | TDADH | Inner-WIoU | P/% | R/% | Params/ 10^6 | $mAP_{50}/\%$ | $mAP_{50:95}/\%$ |
|----|-------------------|------|-------|------------|-------------|-------------|----------------|---------------|------------------|
| 1 | | | | | 44.3 | 34.0 | 3.01 | 33.9 | 19.8 |
| 2 | ✓ | | | | 46.3 | 34.0 | 2.89 | 34.5 | 20.1 |
| 3 | ✓ | ✓ | | | 51.8 | 40.2 | 3.15 | 41.7 | 24.9 |
| 4 | ✓ | ✓ | ✓ | | 51.5 | 42.4 | 2.72 | 43.8 | 26.6 |
| 5 | ✓ | ✓ | ✓ | ✓ | 53.9 | 43.4 | 2.72 | 44.7 | 27.1 |

从消融实验的结果来看, 每个改进模块相对于基线模型各项指标都有不同程度的提升。方法 1 为 YOLOv8n 基

线模型; 方法 2 是将卷积加性自注意力与 C2f 模块融合, 使模型降低无关信息的干扰, 更准确地关注有效信息, 从而

使平均精度分别提升了 0.6% 和 0.3%;方法 3 在方法 2 的基础上改进了颈部网络,加强多尺度的特征融合,提高了模型对小目标的感知能力和准确性,与基线网络相比,平均精度分别提升了 7.8% 和 5.1%,召回率也提升了 7.5%,但由于网络结构的变化,使得参数量也增加了 4.6%;方法 4 在以上方法的基础上增加了任务动态对齐检测头 TDADH,加入了共享卷积以及可变形卷积,与基线网络相比,参数量降低了 9.6%,平均精度分别提升了 9.9% 和 6.8%;方法 5 是将原损失函数替换为 Inner-WIoU 损失函数,并将上述 3 个改进整合到一块,可以看出各个指标都有一定程度的提升,召回率提升了 9.6%,参数量降低了 9.6%,平均精度分别提升了 10.8% 和 7.3%,充分说明了改进后模型对小目标检测的有效性。

2.5 损失函数对比试验

为了进一步验证 Inner-WIoU 相对于其他边界框损失函数的优越性,在其他训练条件一致的情况下,进行了不同损失函数的对比实验,包括 DIoU^[23]、EIoU^[24] 以及 WIoU。本次实验均在 YOLOv8n 模型上进行对比,具体实验结果如表 3 所示。

表 3 损失函数对比实验结果

Table 3 Comparison of experimental results on loss functions

| 损失函数 | P/% | R/% | mAP50/% | mAP50:95/% |
|------------|-------------|-------------|-------------|-------------|
| CIoU | 44.8 | 33.6 | 33.9 | 19.8 |
| DIoU | 44.4 | 33.9 | 33.8 | 19.7 |
| EIoU | 44.8 | 33.1 | 33.5 | 19.7 |
| WIoUv3 | 45.0 | 33.9 | 34.0 | 19.6 |
| Inner-WIoU | 45.0 | 34.0 | 34.2 | 19.9 |

从表 3 中可以看出,在将 CIoU 损失换成 Inner-WIoU 损失后,在召回率以及各个检测精度上均高于其他的 IoU 损失函数,证明了 Inner-WIoU 损失函数的有效性。

2.6 主流算法对比试验

为了进一步验证本文算法在航拍图像目标检测中的性能,将改进后的模型与不同尺寸的 YOLO 系列模型以及当前主流的目标检测模型进行对比试验,实验依旧采用 VisDrone2019 数据集,实验结果如表 4 所示。

表 4 各算法在 VisDrone2019 数据集上的对比实验结果

Table 4 Comparison experiment results of various algorithms on the VisDrone2019 dataset

| 模型 | mAP50/% | mAP50:95/% | Params/10 ⁶ | GFLOPs |
|---|-------------|-------------|------------------------|------------|
| Faster-RCNN+ResNeXt101 ^[25] | 40.2 | 22.6 | 108.60 | 233.0 |
| Cascade-RCNN+ResNeXt101 ^[25] | 41.2 | 24.4 | 156.60 | 276.2 |
| Swin Transformer ^[26] | 35.6 | 20.6 | 34.20 | 44.5 |
| RT-DETR ^[27] | 44.7 | 27.0 | 20.00 | 57.3 |
| YOLOv5n | 25.7 | 12.9 | 2.65 | 7.7 |
| YOLOv5s | 32.9 | 17.8 | 9.14 | 24.0 |
| YOLOv8n | 33.9 | 19.8 | 3.01 | 8.1 |
| YOLOv8s | 40.1 | 23.9 | 11.14 | 28.6 |
| YOLOv11n ^[28] | 33.3 | 19.5 | 2.62 | 6.5 |
| YOLOv11s | 39.4 | 23.6 | 9.42 | 21.3 |
| UCM-YOLOv8 | 44.7 | 27.1 | 2.72 | 26.2 |

Faster-RCNN+ResNeXt101^[25]、Swin Transformer^[26] 作为代表性的双阶段以及多阶段算法,展现了其较高的检测精度,但其参数量以及计算量过高。与主流的实时检测模型 RT-DETR^[27] 相比,两者虽说在精度指标上不相上下,但由于 RT-DETR 是基于 Transformer 架构实时检测模型,同样导致其参数量和计算量过大,难以满足在无人机这种有限资源的设备上稳定运行。YOLOv8 使用了 C3 改进而来的 C2f,从而带来了更加丰富的梯度信息加强了特征提取能力,使得其综合性能优于 YOLOv5,而最新推出的 YOLOv11^[28] 有了 C3K2 模块以及深度可分离卷积的引入,使其在参数量和计算量方面有一定优势,但不可避免的导致了精度的下降。UCM-YOLOv8 相较于过去具有代

表性的 YOLO 模型,虽说在计算量方面有所提升,但在精度上都远超于以上模型的 n 版本以及 s 版本,并且在模型大小方面都与其对应版本逼近甚至超越。改进后的算法在各项指标上都完全超越了更大规模的 YOLOv8s 模型,充分证明了本文算法改进的有效性。

为进一步体现算法先进性,选择近期基于 YOLOv8 的改进模型与本文算法进行对比实验。实验结果如表 5 所示。文献[29]基于 YOLOv8s 采用动态检测头提升检测精度并结合 BiFPN 融合小目标检测层降低其参数量使其各项指标均略优于原 YOLOv8 算法。文献[30]则将可变形卷积融合进 C2f 模块中从而引导卷积操作聚焦于图像中特定区域,以增强特征提取能力。SM-YOLOv8n^[31] 通过设

计 SOEP 模块采用多个分支使用多种注意力机制来捕捉上下文信息,增强特征提取以及融合能力,使精度有所提升,但由于大卷积核的使用难免导致参数量有所增加。MECW-YOLO^[32]采用轻量化主干网络 MobileViT 来替换原主干网络同时借鉴 CCFM 结构思想,促进多尺度特征融合,检测性能有所提升,但参数量略小于 UCM-YOLOv8。实验结果表明,UCM-YOLOv8 在检测精度方面均优于同级别算法,虽说在参数量方面略高于文献[30]以及 MECW-YOLO,但综合来看,其突出的检测精度更适用于对识别准确率要求较高的无人机航拍任务。

2.7 可视化分析

为直观评估改进后算法在航拍领域中的优势,本文从选取了白天密集场景,晚间车流模糊场景以及高空俯视角

表 5 改进模型与各种新模型的对比实验结果

Table 5 Comparison of experimental results between the improved model and various new models

| 模型 | mAP50/% | mAP50:95/% | Params/10 ⁶ |
|----------------------------|-------------|-------------|------------------------|
| 文献[29] | 41.7 | — | 9.2 |
| 文献[30] | 40.4 | 24.4 | 2.3 |
| SM-YOLOv8n ^[31] | 37.4 | — | 3.3 |
| MECW-YOLO ^[32] | 39.1 | 23.3 | 2.4 |
| UCM-YOLOv8 | 44.7 | 27.1 | 2.72 |

拍摄 3 个具有代表性的场景进行可视化实验,图 11 分别展示了原图、YOLOv8n 与 UCM-YOLOv8 的检测效果对比图。

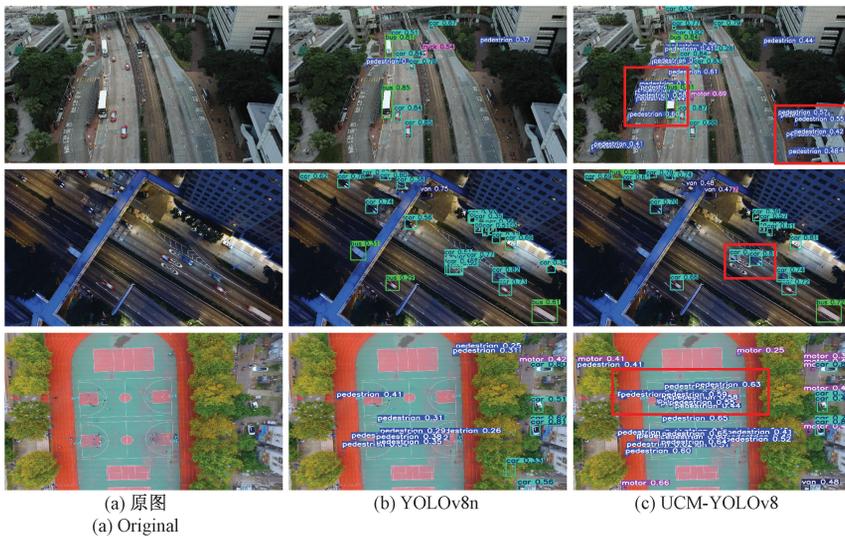


图 11 检测效果对比图

Fig. 11 Comparison of detection results

由上图可以看出,在第 1 行图片中原算法在图片左右两侧均出现了对等待公交车人群的漏检情况,第 3 行图片中 YOLOv8n 算法同样出现了对俯拍人群的漏检情况,而改进后的算法则在很大程度上减少了漏检;第 2 行图片中 YOLOv8n 算法则出现了大量对路标误检成车辆的情况,而改进后的 UCM-YOLOv8 模型则改善了误检这一情形。由此可以看出,本文算法更适合应用于航拍场景下的小目标检测。

3 结 论

本文针对无人机航拍图像中目标小而密集,且极易出现漏检和误检等问题,提出了 UCM-YOLOv8 目标检测算法。首先,设计特征聚合扩散金字塔网络,以增强浅层与深层的特征融合;其次,设计任务动态对齐检测头,在降低参数量的同时将分类和定位分支相互交互,增强多尺度特征的提取能力,提高训练和推理的效率;再者,引入卷积加

性注意力机制到 C2f 模块中,提高特征提取能力;最后,将 Inner-IoU 和 Wise-IoUv3 相结合,提出 Inner-WIoU 损失函数替代原损失函数,扩大回归的有效范围,从而更好地检测小目标。在 VisDrone2019 数据集上的实验结果表明,该算法在检测精度以及轻量化方面相较于当前主流算法均有一定优势,证明本模型在无人机航拍视角下的小目标检测效果优异,适合无人机航拍图像的应用。

参考文献

[1] MOHSAN S A H, KHAN M A, NOOR F, et al. Towards the unmanned aerial vehicles (UAVs): A comprehensive review[J]. Drones, 2022, 6(6): 147.
 [2] ZHANG G. 6G enabled UAV traffic management models using deep learning algorithms[J]. Wireless Networks, 2024, 30(8): 6709-6719.
 [3] TSOUROS D C, BIBI S, SARIGIANNIDIS P G. A review on UAV-based applications for precision

- agriculture[J]. Information, 2019, 10(11): 349.
- [4] ALSAMHI S H, SHVETSOV A V, KUMAR S, et al. UAV computing-assisted search and rescue mission framework for disaster and harsh environment mitigation[J]. Drones, 2022, 6(7): 154.
- [5] KANELLAKIS C, NIKOLAKOPOULOS G. Survey on computer vision for UAVs: Current developments and trends [J]. Journal of Intelligent & Robotic Systems, 2017, 87(1): 141-168.
- [6] 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 and Machine Intelligence, 2016, 39(6): 1137-1149.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 779-788.
- [8] LIU W, ANGELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. 14th European Conference on Computer Vision(ECCV), Amsterdam, The Netherlands. Springer, 2016: 21-37.
- [9] 刘涛,高一萌,柴蕊,等.改进YOLOv5s的无人机视角下小目标检测算法[J].计算机工程与应用,2024,60(1):110-121.
- LIU T, GAO Y M, CHAI R, et al. Small object detection algorithm from UAV perspective based on improved YOLOv5s [J]. Computer Engineering and Applications, 2024, 60(1): 110-121.
- [10] 张徐,朱正为,郭玉英,等.基于cosSTR-YOLOv7的多尺度遥感小目标检测[J].电光与控制,2024,31(4): 28-34.
- ZHANG X, ZHU ZH W, GUO Y Y, et al. Multi-scale small object detection in remote sensing images based on cosSTR-YOLOv7[J]. Electronics Optics & Control, 2024, 31(4): 28-34.
- [11] 钟帅,王丽萍.改进YOLOv8的无人机航拍图像目标检测方法[J].航空兵器,2025,32(3): 91-101.
- ZHONG SH, WANG L P. Target detection method for UAV aerial images based on improved YOLOv8[J]. Aero Weaponry, 2025, 32(3): 91-101.
- [12] SUI J, CHAN D, ZHENG X, et al. A new algorithm for small target detection from the perspective of unmanned aerial vehicles[J]. IEEE Access, 2024, 12: 29690-29697.
- [13] 范博滢,王淑青,陈开元.基于改进YOLOv8的航拍无人机小目标检测模型[J].计算机应用,2025,45(7): 2342-2350.
- FAN B G, WANG SH Q, CHEN K Y. Aerial UAV small object detection model based on improved YOLOv8[J]. Computer Applications, 2025, 45(7): 2342-2350.
- [14] 梁秀满,贾梓涵,刘振东,等.改进YOLOv8n的无人机航拍图像检测算法[J].电光与控制,2025,32(1):34-40,67.
- LIANG X M, JIA Z H, LIU ZH D, et al. UAV aerial image detection algorithm based on improved YOLOv8n[J]. Electronics Optics & Control, 2025, 32(1): 34-40,67.
- [15] LI Y, HOU Q, ZHANG Z, et al. Large selective kernel network for remote sensing object detection[C]. IEEE/CVF International Conference on Computer Vision, 2023: 16794-16805.
- [16] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [17] XU S, ZHENG S, XU W, et al. Hcf-net: Hierarchical context fusion network for infrared small object detection [C]. 2024 IEEE International Conference on Multimedia and Expo(ICME). IEEE, 2024: 1-6.
- [18] FENG C, ZHONG Y, GAO Y, et al. Toody: Task-aligned one-stage object detection [C]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE Computer Society, 2021: 3490-3499.
- [19] ZHU X, HU H, LIN S, et al. Deformable convnets v2: More deformable, better results[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9308-9316.
- [20] ZHANG T, LI L, ZHOU Y, et al. Cas-vit: Convolutional additive self-attention vision transformers for efficient mobile applications [J]. ArXiv preprint arXiv:2408.03703, 2024.
- [21] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism [J]. ArXiv preprint arXiv: 2301.10051, 2023.
- [22] ZHANG H, XU C, ZHANG S. Inner-IoU: More effective intersection over union loss with auxiliary boundingbox [J]. ArXiv preprint arXiv: 2311.02877, 2023.
- [23] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]. AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

- [24] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*, 2022, 506: 146-157.
- [25] TANG W, SUN J, WANG G. Horizontal feature pyramid network for object detection in UAV images[C]. 2021 China Automation Congress (CAC). IEEE, 2021: 7746-7750.
- [26] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. *IEEE/CVF International Conference on Computer Vision*, 2021: 10012-10022.
- [27] ZHAO Y, LYU W, XU S, et al. DETRs beat YOLOs on real-time object detection[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 16965-16974.
- [28] KANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements [J]. *ArXiv preprint arXiv:2410.17725*, 2024.
- [29] 曲晨阳, 吕进, 卫策. 基于改进 YOLOv8s 的无人机目标检测算法[J]. *国外电子测量技术*, 2024, 43(7): 14-23.
- QU CH Y, LYU J, WEI C. UAV target detection algorithm based on improved YOLOv8s[J]. *Journal of Foreign Electronic Measurement Technology*, 2024, 43(7): 14-23.
- [30] 刘付刚, 刘巾瑞, 祝永涛. 改进 YOLOv8 的无人机小目标检测方法[J]. *黑龙江科技大学学报*, 2024, 34(6): 985-989.
- LIU F G, LIU J R, ZHU Y T. Small object detection method for UAV based on improved YOLOv8 [J]. *Journal of Heilongjiang University of Science and Technology*, 2024, 34(6): 985-989.
- [31] 向征, 张佳浩. 基于 SM-YOLOv8n 的无人机航拍目标检测[J]. *海军航空大学学报*, 2025, 40(2): 321-328.
- XIANG ZH, ZHANG J H. Target detection in UAV aerial imagery based on SM-YOLOv8n[J]. *Journal of Naval Aviation University*, 2025, 40(2): 321-328.
- [32] 项家灏, 徐慧英, 徐广平, 等. MECW-YOLO: 基于改进 YOLOv8 的无人机视角小目标检测算法[J/OL]. *计算机工程与科学*, 1-12[2025-05-15]. <http://kns.cnki.net/kcms/detail/43.1258.TP.20241225.1106.008.html>.
- XIANG J H, XU H Y, XU G P, et al. MECW-YOLO: Small object detection algorithm from UAV perspective based on improved YOLOv8 [J/OL]. *Computer Engineering and Science*, 2025: 1-12[2025-05-15]. <http://kns.cnki.net/kcms/detail/43.1258.TP.20241225.1106.008.html>.

作者简介

贾亮, 硕士, 副教授, 主要研究方向为信息获取与处理。

E-mail: jialiang@sau.edu.cn

陈茂辉(通信作者), 硕士研究生, 主要研究方向为图像处理、目标检测。

E-mail: 547840282@qq.com