

文章编号:1007-2780(2026)02-0253-14

## 面向无人机影像小目标检测的轻量化算法

罗可心<sup>1</sup>, 李松江<sup>1,2</sup>, 王 鹏<sup>1,3</sup>, 杨华民<sup>1\*</sup>

(1. 长春理工大学 计算机科学技术学院, 吉林 长春 130022;

2. 吉林省大数据科学与工程联合重点实验室, 吉林 长春 130022;

3. 吉林省网络数据库应用软件科技创新中心, 吉林 长春 130022)

**摘要:**针对现有检测模型对无人机影像中目标、特别是小目标检测不充分的问题,本文提出一种轻量化小目标检测模型 PRSU-YOLO。首先,设计风车结构卷积自适应模块,增强对细微特征的提取能力;其次,构建重参数化空间-通道卷积模块,通过动态特征重建优化多尺度表征;再次,在颈部网络中嵌入小目标检测增强分支,构建高分辨率细节特征的强化路径;最后,引入尺度动态交并比损失函数,使模型能够自适应地调整边界框回归策略。本文模型在计算复杂度仅增加 14.6 GFLOPS 的基础上,在 VisDrone2019 数据集上的 mAP@0.5 达到 37.4%,较基线模型提升 4.4%;在 TinyPerson 数据集上的精确率达到 34.5%,提升 3.5%。实验结果表明,该模型在显著提升检测能力的同时,有效控制了计算成本,为无人机对地观测场景下的目标检测任务提供了有效的解决方案。

**关键词:**计算机视觉;无人机影像;小目标检测;特征增强

**中图分类号:**TP391.4 **文献标识码:**A **doi:**10.37188/CJLCD.2025-0250 **CSTR:**32172.14.CJLCD.2025-0250

## Lightweight algorithm for small object detection in UAV images

LUO Kexin<sup>1</sup>, LI Songjiang<sup>1,2</sup>, WANG Peng<sup>1,3</sup>, YANG Huamin<sup>1\*</sup>

(1. School of Computer Science and Technology, Changchun University of Science and Technology,  
Changchun 130022, China;

2. Jilin Provincial Joint Key Laboratory of Big Data Science and Engineering, Changchun 130022, China;

3. Jilin Provincial Science and Technology Innovation Center of Network Database Application Software,  
Changchun 130022, China)

**Abstract:** To address the insufficient feature extraction of existing detection models for tiny targets, this paper proposes a lightweight small object detection model named PRSU-YOLO. First, a Pinwheel-shaped Convolutional Adaptive Module is designed to enhance the directional extraction capability of subtle features. Second, a Reparameterized Spatial-Channel Convolution Module is constructed to optimize multi-scale representation through dynamic feature reconstruction. Third, a Small Object Detection Branch is embedded in the neck network to establish an enhancement pathway for high-resolution detail

**收稿日期:**2025-12-16;**修订日期:**2026-01-21.

**基金项目:**吉林省科技创新平台建设项目(No.YDZJ202302CXJD027)

Supported by Jilin Province's Science and Technology Innovation Platform Construction Project (No. YDZJ202302CXJD027)

\*通信联系人, E-mail: 2024200207@mails.cust.edu.cn

features. Finally, a Scale-based Dynamic Intersection over Union loss function is introduced, enabling the model to adaptively adjust the bounding-box regression strategy. With only a 14.6 GFLOPS increase in computational complexity, the proposed model achieves an mAP@0.5 of 37.4% on the VisDrone2019, representing a significant improvement of 4.4% over the baseline. On the TinyPerson, it attained a precision of 34.5%, which is an increase of 3.5% compared to the baseline. The experimental results demonstrate that the model significantly enhances detection capability while effectively controlling computational cost, providing an effective solution for small object detection tasks in UAV-based ground observation scenarios.

**Key words:** computer vision; unmanned aerial vehicle imagery; small object detection; feature enhancement

## 1 引 言

随着我国低空智能感知技术的不断发展,无人机已成为对地观测、公共安全等领域不可或缺的感知工具<sup>[1-2]</sup>。无人机从高空俯视获取的影像,具有覆盖范围广、响应速度快等优势,但同时也存在着目标尺度微小、背景信息复杂等一系列严峻的技术挑战<sup>[3-6]</sup>。特别是在智慧城市管理、边境巡逻等无人机对地观测的实际应用场景中,对影像中行人、车辆等小目标的精准检测,是实现高层次场景理解与决策支持的基础。因此,研究适用于无人机影像的小目标检测技术,对提升国产遥感平台的智能解译能力、服务国家相关战略需求具有重要的理论价值与现实意义。

当前,基于深度学习的目标检测方法主要分为两阶段与单阶段两类。两阶段目标检测方法以 R-CNN<sup>[7]</sup>为开端,其首次将候选框送入卷积网络提取特征,再用支持向量机精修边界框。随后, Fast R-CNN<sup>[8]</sup>提出感兴趣区域池化实现端到端训练。而 Faster R-CNN<sup>[9]</sup>引入了区域提议网络,使得候选框生成与检测共享卷积特征,奠定了两阶段框架的基础。这类方法的检测精度高但计算复杂,难以满足无人机平台的实时处理需求。单阶段方法将检测视为一次回归问题,直接在特征图上预测类别与位置。以 YOLO 系列<sup>[10-13]</sup>方法为例,此类方法将输入图像划分为网格,利用全卷积网络一次性输出检测框与类别,显著提升了速度。此后, YOLO 系列通过不断的改进,大幅提高了单阶段目标检测模型的性能。然而,这些研究多针对通用场景或地面视角,其网络结构对无人机影像中存在的小目标检测仍显不足。

近年来,面向无人机影像小目标检测的算法研究主要分为网络结构轻量化设计与检测精度专项提升两方面。在网络结构设计方面, Huo<sup>[14]</sup>等人提出了可变形特征聚合网络。谭海英<sup>[15]</sup>等人设计了轻量化三元注意力网络。这些方法在降低计算复杂度方面有所贡献,但对在复杂场景下的小目标特征保持能力仍有不足。在检测精度提升方面, Zhang<sup>[16]</sup>等人通过重构颈部网络结构,强化了跨尺度特征的提取与融合能力。Xu<sup>[17]</sup>等人在 YOLOv8 中引入高效多尺度注意力机制,提升模型对小目标检测的精度。Doloriel<sup>[18]</sup>等人设计了一种复合损失函数,以提升模型在复杂场景中对小目标的检测精度。

然而,现有研究仍面临两方面突出局限。一方面,多数方法为提升性能而引入的复杂结构,往往导致模型复杂度与计算成本显著上升,难以满足无人机实时性的要求。另一方面,现有网络架构对无人机影像中小目标检测问题缺乏针对性设计,浅层细节特征在向深层传递过程中易受稀释,而深层语义特征又无法充分反馈小目标位置信息。这些机制上的不足导致现有模型在应对小目标的特征信息匮乏、定位困难等挑战时,鲁棒性和精确性受限。

针对上述问题,本文提出一种面向无人机影像的轻量化小目标检测模型——PRSU-YOLO。该模型以 YOLOv9s 为基准架构,摒弃了单纯增加模块的传统思路,转而采用“特征增强-信息融合-损失引导”的协同结构,在保持较低计算复杂度的同时,增强对小目标的检测能力。本文具体贡献如下:

(1) 在特征增强阶段,设计了风车结构卷积自适应模块(Pinwheel-shaped Convolutional Adaptive

Module, PCAM), 其通过多方向卷积与通道-空间自适应权重的协同作用, 显著提升了网络对小目标初级空间特征的定向提取与聚焦能力;

(2) 在信息融合阶段, 构建了重参数化空间-通道卷积模块 (Reparameterized Spatial-Channel Convolution Module, RepSCCM) 与小目标检测增强分支, 前者实现多尺度特征的自适应重建与冗余抑制, 后者则为高分辨率细节特征建立强化传输路径, 两者协同工作, 确保了从浅层细节到深层语义的高效融合;

(3) 在优化引导阶段, 结合了尺度动态交并比损失函数, 其能够根据目标的绝对尺度与当前特征图分辨率动态调整边界框回归重点, 从而与

前端增强的特征提取能力形成精准的闭环反馈与协同。

## 2 模型结构

基于 YOLOv9s, 本文提出了轻量化小目标检测模型 PRSU-YOLO。该模型首先在主干网络引入风车卷积自适应模块, 以多方向卷积增强对特征的感知能力; 其次在颈部构建重参数化空间-通道卷积模块, 通过动态特征重建优化多尺度表征; 同时添加小目标检测增强分支, 构建高分辨率细节信息的强化通路; 最后采用 SDIoU<sup>[19]</sup>, 实现基于目标尺度的自适应回归。PRSU-YOLO 的结构如图 1 所示。

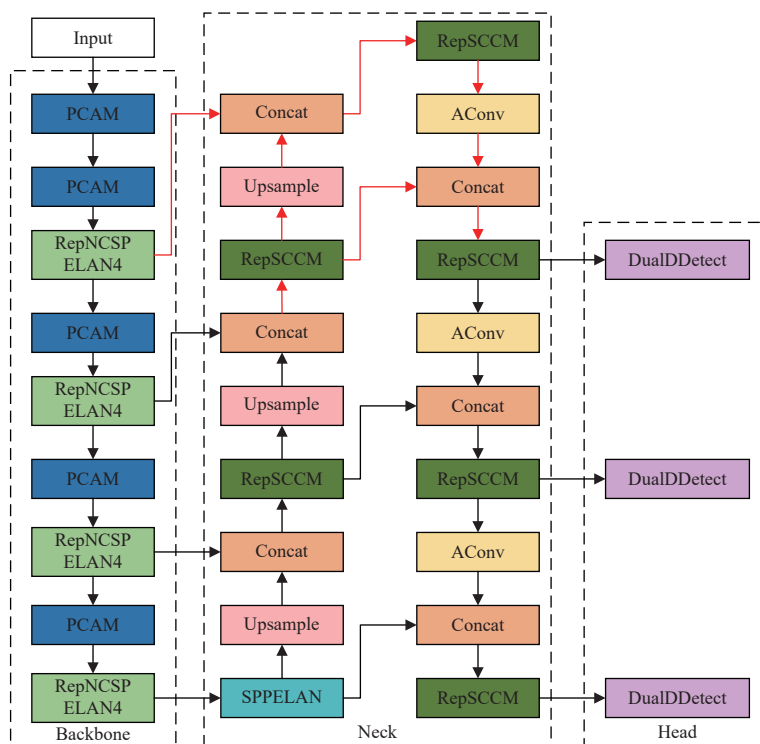


图 1 PRSU-YOLO 模型结构图

Fig. 1 Structure of PRSU-YOLO model

### 2.1 风车结构卷积自适应模块

为应对小目标在无人机影像中像素少、纹理弱的问题, 我们设计了 PCAM, 如图 2 所示。该模块由风车卷积块和通道-空间自适应块两部分构成。风车卷积块采用多方向非对称卷积, 模拟小目标的高斯空间分布; 通道-空间自适应块动态地校准特征图的通道与空间权重, 聚焦与特征

的关键信息。

具体来说, 设输入特征为  $X_{in} \in \mathbb{R}^{N \times C_m \times H \times W}$ , 该模块首先采用 4 个不同填充方向的非对称卷积核并行处理输入特征, 以捕捉不同方向的边缘与纹理信息。具体公式如式 (1) 所示:

$$X_k = \text{SiLU}(\text{BN}(\text{Conv}_{(k)}(X_{in}; W_k))), k=1, 2, 3, 4, \quad (1)$$

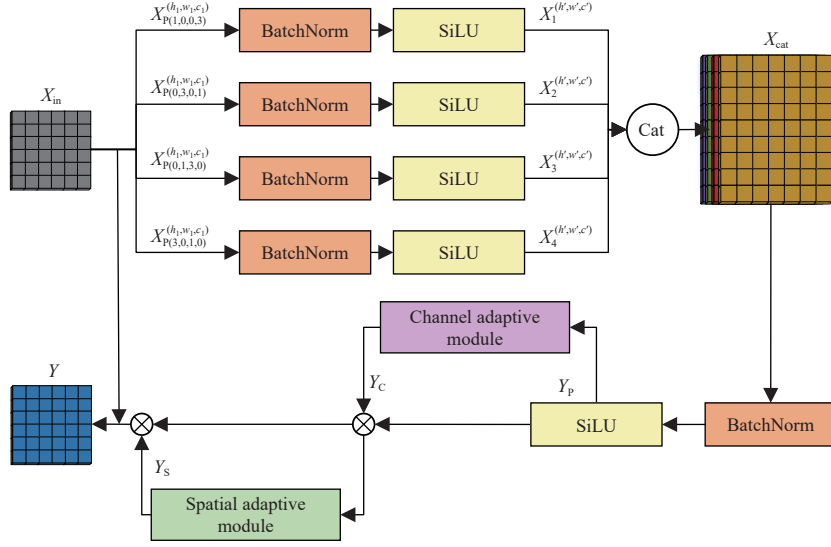


图2 风车结构卷积自适应模块结构图

Fig. 2 Structure of PCAM

其中:  $\text{Conv}_{(k)}$  表示第  $k$  个方向的卷积操作, 其卷积核  $W_k$  在水平方向尺寸大小为  $1 \times 3$ 、垂直方向则相反。

模块为整合多方向特征信息, 将得到的 4 个方向的输出特征在通道维度上进行拼接, 具体如式(2)所示:

$$X_{\text{cat}} = \text{Concat}(X_1, X_2, X_3, X_4), \quad (2)$$

其中:  $X_k \in \mathbb{R}^{N \times C' \times H' \times W'}$ ,  $k=1, 2, 3, 4$  为第  $k$  个方向的输出; 特征图  $X_{\text{cat}} \in \mathbb{R}^{N \times 4C' \times H' \times W'}$ ,  $C'=C_{\text{out}}/4$ ,  $C_{\text{out}}$  为该模块的总输出通道数;  $H' \times W'$  为经过卷积核填充后的特征图尺寸, 计算公式如式(3)所示:

$$H' = \left\lfloor \frac{H + \text{padding}_{\text{top}} + \text{padding}_{\text{bottom}} - k_h}{s} \right\rfloor + 1, \\ W' = \left\lfloor \frac{W + \text{padding}_{\text{left}} + \text{padding}_{\text{right}} - k_w}{s} \right\rfloor + 1, \quad (3)$$

其中:  $k_h, k_w$  为卷积核的宽和高;  $s$  为卷积步长;  $\text{padding}_{\text{direction}}$  为各方向的填充数, 具体填充数见图 2。

通过一个  $2 \times 2$  卷积整合拼接后的多方向特征, 并进行通道降维, 具体如式(4)所示:

$$Y_p = \text{SiLU}(\text{BN}(f_{\text{conv}}^{2 \times 2}(X_{\text{cat}}; W_p))) \in \mathbb{R}^{N \times C_{\text{in}} \times H' \times W'}. \quad (4)$$

为了增强关键通道特征并抑制相关性较低的特征, 将特征  $Y_p$  输入到通道自适应子块中。

首先, 通道自适应子块对输入特征  $Y_p$  进行池化操作, 然后将结果传递到多层感知机  $\text{MLP}(\cdot)$

中, 接着对其输出结果进行求和, 并利用 Sigmoid 函数处理得到特征  $Y_c$ 。该子块的具体公式如式(5)所示:

$$Y_c = \text{Sigmoid}(\text{MLP}(\text{AvgPool}(Y_p)) + \text{MLP}(\text{MaxPool}(Y_p))). \quad (5)$$

随后  $Y_c$  输入到空间自适应子块中, 其通过分配空间注意力权重, 探索特征内的空间关系和关键区域。

与通道自适应子块一致, 该子块将特征  $Y_p$  和  $Y_c$  相乘后, 进行平均池化和最大池化。随后, 将拼接特征通过卷积和 Sigmoid 函数生成特征  $Y_s$ 。相应的公式如式(6)所示:

$$Y_s = \text{Sigmoid}(f_{\text{conv}}^{7 \times 7}([\text{AvgPool}(Y_p \times Y_c); \text{MaxPool}(Y_p \times Y_c)])). \quad (6)$$

最后, 为保持梯度流动的稳定性和特征的不稳定性, 模块将输入特征  $X_{\text{in}}$  与特征  $Y_p, Y_c$  和  $Y_s$  进行残差连接:

$$Y = f_{\text{conv}}^{1 \times 1}(X_{\text{in}}) + Y_p \times Y_c \times Y_s. \quad (7)$$

## 2.2 重参数化空间-通道卷积模块

为提高无人机影像中小目标特征的提取和融合, 本研究提出了 RepSCCM。该模块通过空间重建单元和通道重建单元, 显式地挖掘特征并压缩冗余信息, 进而提升模型对小目标的检测能力。其具体结构如图 3 所示。

具体来说, 输入特征张量  $X_i \in \mathbb{R}^{N \times C_i \times H \times W}$  首先经过一个  $1 \times 1$  卷积层进行通道数调整, 得到特征

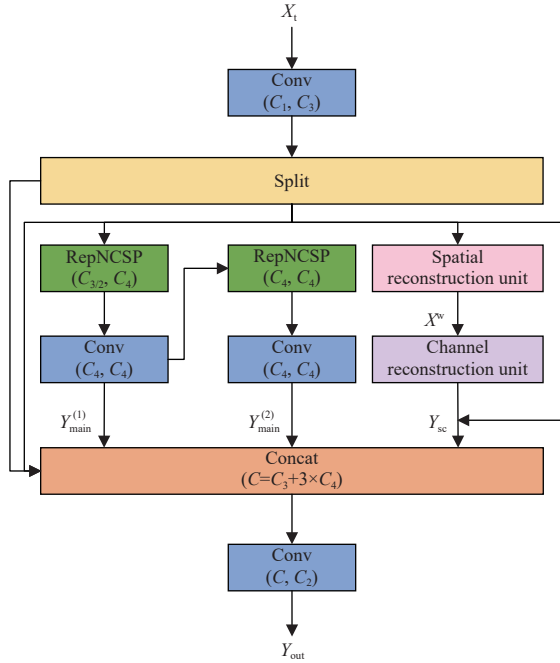


图 3 重参数化空间-通道卷积模块结构图

Fig. 3 Structure of RepSCCM

$X_{in} = f_{conv}^{1 \times 1}(C_1, C_3)(X_t)$ , 随后  $X_{in}$  沿通道维度均等分割为两个分支, 即主分支特征  $X_{main} \in \mathbb{R}^{N \times (C_3/2) \times H \times W}$  和通道-空间卷积分支特征  $X_{sc} \in \mathbb{R}^{N \times (C_3/2) \times H \times W}$ 。

主分支旨在继承 YOLOv9s 中高效的多尺度特征融合和重用能力, 其中包括两个级联的重参数化跨阶段部分 (Reparameterized Nested Cross Stage Partial, RepNCSP) 块, 每个块的具体操作如下:

首先,  $X_{main}$  通过一个  $1 \times 1$  卷积将通道数调整为  $C_4$ :

$$X'_{main} = f_{conv}^{1 \times 1}(C_3/2, C_4)(X_{main}). \quad (8)$$

接着, 顺序通过两个 RepNCSP 块, 每个块在训练阶段采用多分支结构; 而在推理阶段, 通过结构重参数化技术可等效为单个  $3 \times 3$  卷积层。第  $i$  个 RepNCSP 块的操作可表示为:

$$Y_{RepNCSP}^{(i)} = R(f_{conv}^{3 \times 3}(Y^{(i-1)})) + Identity(Y^{(i-1)}), \quad i = 1, 2, \quad (9)$$

其中:  $R(\cdot)$  表示重参数化过程,  $Identity(\cdot)$  表示残差连接。

每个 RepNCSP 块后接一个  $3 \times 3$  卷积层对特征进行精炼:

$$Y_{main}^{(i)} = f_{conv}^{3 \times 3}(C_4, C_4)(Y_{RepNCSP}^{(i)}). \quad (10)$$

经过两个级联块处理后, 得到主分支的输出

特征  $Y_{main} \in \mathbb{R}^{N \times C_4 \times H \times W}$ 。

通道-空间卷积分支由两个互补的组件组成, 即增强空间特征的空间重建单元 (Spatial Reconstruction Unit, SRU) 和调整通道特征的通道重建单元 (Channel Reconstruction Unit, CRU), 旨在显式地挖掘信息丰富的特征并压缩冗余信息。

空间重建单元通过基于归一化参数的动态权重计算与交叉重建机制, 实现对特征图空间维度的筛选和重构, 有效增强信息丰富区域的特征响应, 同时抑制冗余信息的干扰。该单元结构如图 4 所示。

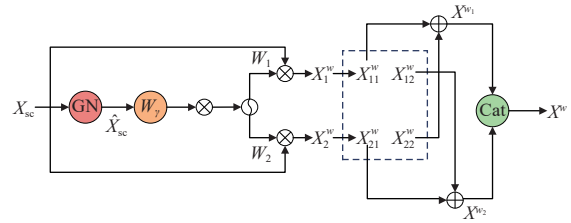


图 4 空间重建单元结构图

Fig. 4 Structure of SRU

SRU 首先对输入  $X_{sc}$  进行组归一化处理, 以稳定训练过程并加速收敛, 其公式如式 (11) 所示:

$$\hat{X}_{sc} = GN(X_{sc}) = \gamma \frac{X_{sc} - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta, \quad (11)$$

其中:  $\mu, \sigma$  为特征的均值和标准差,  $\gamma, \beta$  为可学习的仿射参数,  $\epsilon$  为数值稳定常数。

同时, 对于第  $j$  个通道, 计算其重要性权重  $w_j$ , 进而得到归一化的权重向量  $W_j$ , 具体公式如式 (12) 所示:

$$W_j = \{w_j\} = \frac{\gamma_j}{\sum_{k=1}^{\frac{C_3}{2}} \gamma_k}, \quad j = 1, 2, \dots, \frac{C_3}{2}, \quad (12)$$

其中,  $\gamma$  为可学习阈值参数。

其次, 将  $W_j$  与归一化特征相结合, 通过 Sigmoid 函数与门控机制生成空间掩码, 计算公式如式 (13) 所示:

$$W = Gate(M) = Gate(\text{sigmoid}(W_j(GN(\hat{X}_{sc})))) = \begin{cases} 1 & \text{if } M > \gamma \\ 0 & \text{其他} \end{cases}. \quad (13)$$

由此得到两个互补的二进制掩码, 即信息区域掩码和冗余区域掩码, 其公式如式 (14) 所示:

$$\begin{cases} W_1 = W \\ W_2 = 1 - W_1 \end{cases}, \quad (14)$$

其中, 1 表示全矩阵。

接着, 将原始输入特征分别与两个掩码逐元素相乘, 得到信息特征  $X_1^w$  与冗余特征  $X_2^w$ :

$$\begin{cases} X_1^w = W_1 \otimes X_{sc} \\ X_2^w = W_2 \otimes X_{sc} \end{cases}, \quad (15)$$

其中,  $\otimes$  表示逐元素相乘。

然后, 为了进一步精细化处理, 将  $X_1^w$  与  $X_2^w$  分别沿空间维度均匀划分为  $2 \times 2$  的子网络, 得到 4 个子特征块, 利用子特征块生成两个重建特征块, 具体公式如式 (16) 所示:

$$\begin{cases} X^{w_1} = X_{11}^w \oplus X_{22}^w \\ X^{w_2} = X_{21}^w \oplus X_{12}^w \end{cases}, \quad (16)$$

其中,  $\oplus$  表示逐元素相加。

最后, 将两个重建特征块在通道维度拼接, 并通过一个  $3 \times 3$  卷积层进行融合, 得到空间精炼后的特征  $X^w$ :

$$X^w = f_{\text{conv}}^{3 \times 3}(\text{concat}(X^{w_1}, X^{w_2})). \quad (17)$$

通道重建单元通过自适应通道分割与分组-逐点卷积混合策略, 实现对关键通道特征的强化提取与非关键通道特征的轻量化处理。通道重建单元结构如图 5 所示。

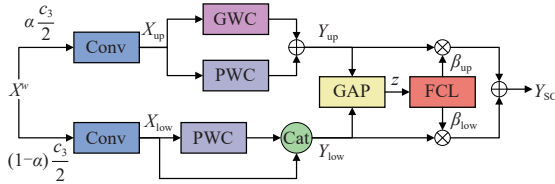


图 5 通道重建单元结构

Fig. 5 Structure of CRU

CRU 首先将  $X^w$  按  $\alpha = 0.75$  的比例分为上下两部分, 即  $X_{up} \in \mathbb{R}^{N \times \alpha(C_3/2) \times H \times W}$  和  $X_{low} \in \mathbb{R}^{N \times (1-\alpha)(C_3/2) \times H \times W}$ 。

其次, 对  $X_{up}$  采用分组卷积 (Group-Wise Convolution, GWC) 与逐点卷积 (Point-Wise Convolution, PWC) 并行的混合结构进行处理, 得到特征  $Y_{up}$ , 具体公式如式 (18) 所示:

$$Y_{up} = M^G f_{\text{conv}}(X_{up}) + M^P f_{\text{conv}}(X_{up}), \quad (18)$$

其中,  $M^G$  和  $M^P$  分别为分组卷积和逐点卷积的权重矩阵。

同时, 对  $X_{low}$  采用轻量化的逐点卷积处理,

具体公式如式 (19) 所示:

$$Y_{low} = M_{low}^P X_{low}, \quad (19)$$

其中  $M_{low}^P$  为逐点卷积的权重矩阵。

然后, 对拼接后的特征图  $Y_{up}$  和  $Y_{low}$  进行全局平均池化, 得到  $z = \text{GAP}([Y_{up}, Y_{low}]) \in \mathbb{R}^{N \times 2C_3}$ , 再通过全连接层 (Fully Connected Layer, FCL) 学习通道重要性, 具体公式如式 (20) 所示:

$$\beta = \text{FCL}(z) = \text{Softmax}(\text{FC}_2(\text{ReLU}(\text{FC}_1(z))))). \quad (20)$$

接着, 将  $\beta$  拆分为对应上下分支的两部分  $[\beta_{up}, \beta_{low}]$ , 进而得到分支输出特征  $Y_{sc}$ :

$$Y_{sc} = \beta_{up} Y_{up} + \beta_{low} Y_{low}. \quad (21)$$

最后, 将  $Y_{main}$  与  $Y_{sc}$  在通道维度上进行拼接, 再通过一个  $1 \times 1$  卷积进行特征融合与通道调整, 得到最终输出  $Y_{out}$ , 具体公式如式 (22) 所示:

$$Y_{out} = f_{\text{conv}}^{1 \times 1}(\text{concat}(Y_{main}, Y_{sc})). \quad (22)$$

### 2.3 小目标检测增强分支

在无人机高空影像中, 小目标所占像素极少, 其丰富的细节与纹理信息主要存在于网络浅层的高分辨率特征图中。然而, 在标准的检测架构中, 这些浅层特征在经由深层骨干网络提取及颈部网络的多级下采样与融合过程中, 极易被稀释或淹没, 导致专属于小目标的判别性特征显著流失。针对这一问题, 本文在模型的颈部中引入一个轻量化的小目标检测增强分支, 确保对检测小目标至关重要的高分辨率细节信息能够被有效地传递至检测头, 从而显著提升模型的小目标感知能力。

具体来说, 该分支首先选取来自主干网络第二阶段的特征图, 记为  $P_2 \in \mathbb{R}^{N \times C_2 \times H_2 \times W_2}$ 。其次, 将  $P_2$  与颈部中已经完成上采样的特征  $F_{neck} \in \mathbb{R}^{N \times C_n \times H_n \times W_n}$  在通道维度上进行拼接:

$$F_{cat1} = \text{concat}(P_2, F_{neck}). \quad (23)$$

接着, 将  $F_{cat1}$  输入到重参数化空间-通道卷积模块进行深度处理:

$$F_{sc} = \text{RepSCCM}(F_{cat1}). \quad (24)$$

然后, 使用一个自适应卷积层对提炼后的特征进行调整与压缩, 得到  $F_{arconv}$ , 并与  $F_{neck}$  进行拼接:

$$F_{cat2} = \text{concat}(F_{arconv}, F_{neck}). \quad (25)$$

最终, 融合后的特征  $F_{cat2}$  被直接输送至检测头, 参与目标位置回归与类别预测。

## 2.4 模型损失函数

PRSU-YOLO模型作为单阶段目标检测器,其总体训练损失函数 $L_{total}$ 由边界框回归损失、置信度损失和分类损失3个部分组成:

$$L_{total} = L_{SDIoU} + \lambda_{conf} L_{conf} + \lambda_{cls} L_{cls}. \quad (26)$$

传统的边界框回归损失在处理无人机影像小目标时存在固有缺陷,如CIoU通常用于边界框回归,但其中的宽高比惩罚项无法根据目标尺寸自适应调整。为解决这一问题,本文引入了SDIoU,旨在使损失函数能够感知目标的绝对尺寸和当前特征图的分辨率,并动态调整中心点定位与尺度匹配两项子损失的优化权重。

SDIoU由两项加权子损失构成,具体公式如式(27)所示:

$$L_{SDIoU} = \beta_{L_{BS}} L_{BS} + \beta_{L_{BL}} L_{BL}, \quad (27)$$

其中 $L_{BS}$ 表示尺度匹配损失, $L_{BL}$ 表示位置定位损失,两个损失的计算公式如式(28)所示:

$$\begin{cases} L_{BS} = 1 - \text{IoU} + \alpha v \\ L_{BL} = \frac{\rho^2(b_p, b_{gt})}{c^2} \end{cases}. \quad (28)$$

此外,权重系数 $\beta_{L_{BS}}$ 和 $\beta_{L_{BL}}$ 的计算公式如式(29)所示。

$$\begin{cases} \beta_{L_{BS}} = 1 - \delta + \beta_B \\ \beta_{L_{BL}} = 1 + \delta - \beta_B \end{cases}, \quad (29)$$

其中: $\delta$ 表示目标尺寸因子,用于衡量目标在原始图像中的相对大小; $\beta_B$ 为尺度自适应基础权重,其计算公式如式(30)所示:

$$\beta_B = \min\left(\frac{\text{Area}(B_{gt})}{A_{max}} \times R_{oc} \times \delta, \delta\right), \quad (30)$$

其中, $R_{oc}$ 为分辨率补偿因子,该因子反应了目标在特征图上的密度,具体公式如式(31)所示:

$$R_{oc} = \frac{w_o \times h_o}{w_f \times h_f}, \quad (31)$$

其中: $(w_o, h_o)$ 是模型输入尺寸, $(w_f, h_f)$ 是当前计算损失所在特征图的尺寸。

## 3 实验结果分析

### 3.1 环境及参数设置

实验在Linux操作系统下进行,使用的CPU为Intel Xeon Gold 6300,GPU为NVIDIA RTX 6000 Ada,CUDA版本为12.4。采用高性能GPU旨在完成模型的训练,并确保与各类主流模型进

行公平对比。模型训练采用随机梯度下降(SGD)优化器进行优化,设置初始学习率为0.01,批大小为16,训练轮次为200。详细的实验环境及模型参数如表1和表2所示。

表1 实验环境

环境	型号
操作系统	Linux
CPU	Inter Xeon Gold 6300
GPU	RTX 6000 Ada
CUDA	12.4

表2 模型参数

参数名称	取值
优化器	SGD
学习率	0.01
批大小	16
训练轮次	200

### 3.2 数据集

本文选用无人机影像目标检测数据集Vis-Drone2019<sup>[20]</sup>进行实验。该数据集采集自多样化的现实场景,标注了行人、车辆等共10类目标,图像总计8 629张。数据集中目标尺度差异显著,且包含大量像素占比极小的目标,紧密贴合无人机高空观测的实际挑战。数据集已预先划分为训练集(6 471张)、验证集(548张)和测试集(1 610张),本研究严格遵循此划分进行评估。

同时,为验证模型在小目标检测任务上的泛化能力,本研究在TinyPerson<sup>[21]</sup>数据集上进行了实验。该数据集采集自高空俯视视角的真实海岸线、广场等场景,标注的行人目标平均像素不足 $20 \times 20$ ,共包含1 610张图像。本研究遵循该数据集的原始划分,使用训练集(794张)进行模型训练,并在测试集(816张)上评估性能。

### 3.3 评价指标

为定量评估模型性能,本研究采用目标检测领域通用的精确率(Precision)、召回率(Recall)与平均精度(Average Precision, AP)、全类平均精度(mean Average Precision, mAP)作为评价指标,其计算公式如下:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (32)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (33)$$

$$\text{AP} = \int_0^1 P(R) dR, \quad (34)$$

$$\text{mAP} = \frac{1}{C} \sum_{i=1}^c \text{AP}_i, \quad (35)$$

其中: TP 为正确检测到的目标数量, FP 为错误检测的目标数量, FN 为未被检测到的真实目标

数量, C 为待检测的目标类别总数。

### 3.4 消融实验

为验证本文所提出的各模块的有效性, 本小节在 VisDrone2019 测试集上设计并进行了一系列消融实验。实验结果如表 3 所示, 其中“P”代表精确率, “R”代表召回率。“mAP@0.5”表示交并比阈值为 0.5 时的 mAP, “mAP@0.5:0.95”表示交并比阈值从 0.5 以 0.05 为步长递增至 0.95 时, 计算各阈值下的 mAP 的平均值。

表 3 模型消融实验结果

Tab. 3 Results of the model ablation experiment

YOLOv9s	PCAM	RepSCCM	SODB	SDIoU	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%
✓					47.5	33.9	33.0	18.9
✓	✓				46.4	36.4	34.7	20.5
✓	✓	✓			47.1	36.4	35.0	20.8
✓	✓	✓	✓		49.8	38.0	37.1	22.4
✓	✓	✓	✓	✓	49.9	37.9	37.4	22.4

从表 3 的实验结果可知, 不同模块对模型性能的影响呈现递进增强趋势。具体而言, 当仅加入 PCAM 模块时, 模型在召回率、mAP@0.5 及 mAP@0.5:0.95 三个指标上分别较基线模型提升 2.5%、1.7% 和 1.6%。这表明该模块使得模型有效捕获了更多微小目标的特征, 显著扩大了检测覆盖范围。尽管其精确率因初步特征融合可能引入的噪声而微降 1.1%, 但整体检测性能获得明显改善。

在此基础上, 加入 RepSCCM 模块使模型精确率回升至 47.1%, 同时 mAP@0.5 与 mAP@0.5:0.95 也分别提升至 35.0% 与 20.8%。该模块有效抑制了特征冗余并丰富了特征多样性。

随后, SODB 的集成使模型精确率与召回率分别大幅提升至 49.8% 和 38.0%, mAP@0.5 与 mAP@0.5:0.95 也跃升至 37.1% 与 22.4%。这一结果强有力地证明, SODB 分支是解决小目标特征在深层网络中流失问题的关键。

最后, 采用 SDIoU 损失函数的模型在召回率仅下降 0.1% 的情况下, 精确率和 mAP@0.5 分别达到了 49.9% 和 37.4%, mAP@0.5:0.95 则维持在最高水平。这验证了 SDIoU 所具备的尺度自适应动态调整机制能够更精细地指导边界框回归过程, 在有效抑制误检的同时进一步优化

了定位精度。

综上所述, 改进的模型通过各模块的协同作用实现了性能上的显著提升, 充分证实了本研究整体改进方案的必要性与合理性。

### 3.5 对比实验

#### 3.5.1 与主流模型的对比

为全面评估本文模型的检测性能, 本研究将其与当前主流的目标检测方法在 VisDrone2019 数据集上对综合性能进行对比。参与对比的模型包括两阶段检测器 Faster R-CNN, 基于 Transformer 编解码结构的 RT-DETR<sup>[22]</sup> 以及 YOLO 系列中不同版本与复杂度的模型: YOLOv9s/c、YOLOv10s/m 和 YOLOv11s/m。各模型在相同实验设置下进行评估, 具体结果如表 4 所示。表中各评价指标的最优结果以加粗标出。

由表 4 数据可知, 在准确率和召回率指标上, PRSU-YOLO 取得多项最优水平。其精确率达到 49.9%, 与结构复杂、计算量最大的 YOLOv9c 模型持平; 召回率达到 37.9%, 为所有模型中最高, 较基础模型 YOLOv9s 高 4%。这表明本文模型在保证高检测置信度的同时, 能够更有效地覆盖真实目标, 尤其是在包含大量小目标的复杂场景中。

在综合性能指标上, PRSU-YOLO 模型的 mAP@0.5 达到 37.4%, 相较于基线模型 YOLOv9s

表4 模型对比实验结果

Tab. 4 Results of the model comparison experiment

方法	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPS
Faster R-CNN	35.0	17.5	22.1	9.2	—
RT-DETR	36.1	24.0	21.0	10.6	125.7
YOLOv9s	47.5	33.9	33.0	18.9	38.8
YOLOv9c	49.9	37.4	36.7	21.8	236.7
YOLOv10s	42.6	33.7	31.2	18.0	24.5
YOLOv10m	47.9	36.0	34.8	20.5	63.5
YOLOv11s	44.3	33.4	31.5	18.3	21.3
YOLOv11m	47.0	37.0	34.9	20.7	67.7
本文模型	49.9	37.9	37.4	22.4	53.4

提升了4.4%,相对提升幅度达到13.3%。同时,与计算量是其4.4倍的YOLOv9c相比,也领先0.7%。此外,在mAP@0.5:0.95指标上,PRSU-YOLO模型达到了22.4%,同样超越了YOLOv9c的21.8%和YOLOv11m的20.7%。

在计算效率方面,PRSU-YOLO模型的计算复杂度仅为53.4 GFLOPS,显著低于基于Transformer的RT-DETR模型(125.7 GFLOPS)。与基线模型YOLOv9s相比,本模型在仅增加14.6 GFLOPS的情况下,mAP@0.5等综合性能指标均实现了显著提升。同时,与计算量高达其4.4倍的YOLOv9c模型相比,该模型在多项核心指标上均实现了更优的结果。此外,与YOLOv10m及YOLOv11m相比,PRSU-YOLO模型在计算量相近甚至更低的情况下,仍保持了全面的性能领先。

综上所述,本文模型在表4中取得的全面性能优势,主要由多方向特征提取、高分辨率细节路径与尺度自适应回归3项关键改进共同驱动。这些改进从特征感知、信息保留到定位回归实现了系统性增强,最终在有限计算开销下实现了检测性能的显著提升。

为进一步评估PRSU-YOLO模型在复杂多类别目标检测场景下的综合性能与泛化能力,本节对其在VisDrone2019测试数据集10个不同目标类别上的平均精度进行了详细分析,并与多个主流检测模型进行了对比。具体对比结果如图6所示。

从图6可以看出,与YOLOv9s相比,PRSU-YOLO模型在所有检测类别上均实现了提升。在行人类别上,该模型的AP值达到30.1%,相较

于YOLOv9s提升了6.2%;在摩托车类别上,其AP值也从27.9%提升至33.0%。这充分证明,本文模型有效增强了对各类目标,尤其是小目标的检测能力。

与结构更复杂的YOLOv9c模型相比,PRSU-YOLO在多数类别上检测性能突出。在厢式货车检测上,PRSU-YOLO的AP值为45.5%,高于YOLOv9c的43.4%;在公共汽车检测上,其AP值达到63.8%,亦超越后者的61.6%。这表明,本文通过针对性的网络结构优化,能够以更低的计算成本实现可比甚至更优的类别判别精度。

与基于Transformer架构的RT-DETR模型相比,PRSU-YOLO在小目标检测上展现出显著优势。在行人和自行车等典型小目标类别上,PRSU-YOLO的AP值分别为30.1%和12.8%,远超RT-DETR的18.1%和1.2%。这反映出在小目标场景下,本文模型相比依赖全局注意力机制的Transformer架构,更能精准捕捉微小目标的特征。

与两阶段检测器Faster R-CNN相比,PRSU-YOLO在小汽车类别上的平均精度值高达75.8%,显著优于Faster R-CNN的51.7%。这证明单阶段检测框架结合本文的增强机制,在处理小目标时具有更高的效率与准确性。

最后,与YOLOv10及YOLOv11系列的模型相比,PRSU-YOLO模型同样保持了全面的性能领先。其整体mAP@50分别比YOLOv10s、YOLOv10m、YOLOv11s及YOLOv11m高出6.2%、2.6%、5.9%和2.5%,进一步验证了其改进方案在精度提升上的有效性与先进性。

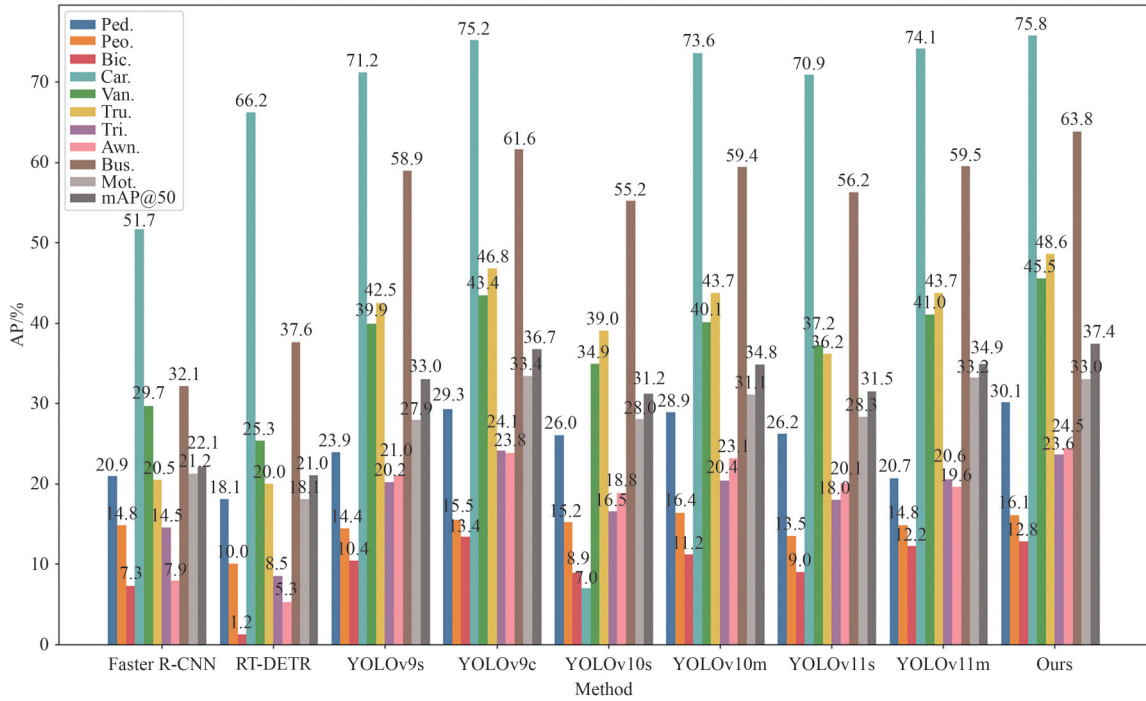


图 6 模型在各目标类别上的平均精度对比

Fig. 6 Comparison of AP of the models on each object category

3.5.2 损失函数对比

为验证 SDIoU 在 PRSU-YOLO 模型中的有效性,本研究在保持其他网络结构与训练设置完全一致的情况下,将 SDIoU 与当前主流的边界框回归损失函数进行了对比实验。该实验旨在评估不同损失函数对模型检测性能,具体对比结果如表 5 所示。

表 5 损失函数在 VisDrone 测试集上的对比

Tab.5 Comparison of loss functions on the VisDrone testset

损失函数	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%
CIoU	49.8	38.0	37.1	22.4
GIoU	48.8	38.1	37.1	22.3
DIoU	49.0	38.2	37.2	22.4
MDPIoU	49.9	37.7	37.2	22.4
SDIoU	49.9	37.9	37.4	22.4

在精确率方面,使用 SDIoU 或 MPDIoU 的模型精确率为 49.9%,优于使用其他损失函数的模型。这表明 SDIoU 在抑制误检、提升检测置信度方面具有与最先进方法相当的能力。

在召回率方面,使用 CIoU 的模型以 38.0% 的表现最优,而使用 SDIoU 的模型为 37.9%,略低 0.1%,但较 MPDIoU 高 0.2%。这一微小差距表明,SDIoU 在动态调整优化目标时,为追求更高的定位精度和抑制误报,对极端困难的正样本覆盖做出了一定权衡,但其召回率依然保持在较高水平。

在 mAP@0.5 上,使用 SDIoU 的模型达到 37.4%,超越了所有对比损失函数。SDIoU 的领先证明了其尺度动态加权机制能够有效引导模型学习更高质量的边界框回归。

在对定位精度要求更严格的 mAP@0.5:0.95 指标上,除使用 GIoU 的模型外,其余模型均达到 22.4%,表现相近。这说明在更高的阈值下,各损失函数引导的模型均能达到相近的定位精度上限。

综合分析表明,SDIoU 实现了更优的精度-召回率平衡,验证了 SDIoU 作为 PRSU-YOLO 模型损失函数的合理性与先进性。

为进一步探究不同损失函数对模型优化过程的影响,本研究对比了各损失函数在完整训练周期内模型的精确率、召回率及平均精度均值的变化趋势,其结果如图 7 所示。

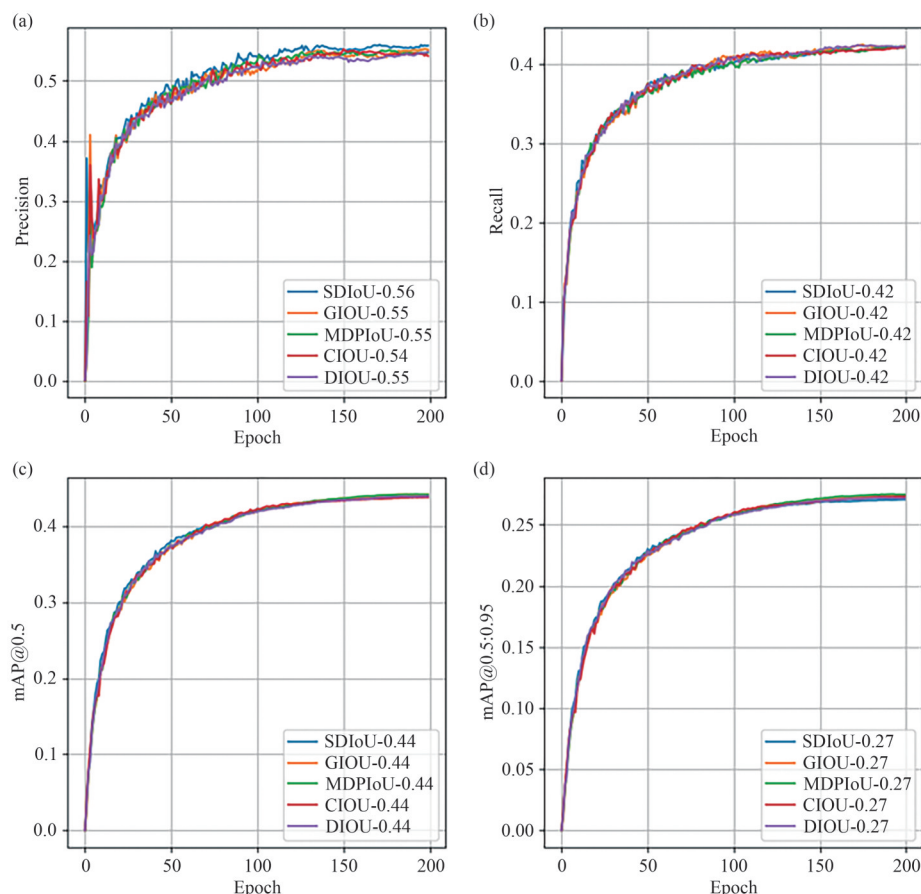


图7 损失函数在模型训练过程中的性能指标变化对比。(a)精确率;(b)召回率;(c) mAP@0.5;(d) mAP@0.5:0.95。  
Fig.7 Comparison of performance metrics of the loss functions during model training. (a) Precision; (b) Recall; (c) mAP@0.5; (d) mAP@0.5:0.95.

从图7可以看出,在召回率与平均精度均值的收敛趋势上,SDIoU与其他对比损失函数最终达到的数值高度相近,表明不同损失函数在引导模型覆盖真实样本与实现整体定位精度方面的能力基本相当。然而,在精确率指标上,训练结束时,使用SDIoU的模型的精确率达到56%,分别优于GIoU与MPDIoU的55%、CIoU的54%以及DIOU的55%。尽管提升幅度微小,但其意味着SDIoU能够在保持同等水平召回率与mAP的前提下,进一步抑制误检的产生。

综合训练过程与最终性能来看,SDIoU不仅能够关键指标上达到或超越现有方法,更重要的是,其提供了一种更优的精度-召回平衡策略,尤其在应对背景杂乱、小目标密集的无人机检测场景时,展现出更强的鲁棒性与实用潜力。

### 3.6 可视化分析

为直观呈现 PRSU-YOLO 模型在检测小目

标方面的优越性能,本节从 VisDrone 测试集中选取了多幅具有代表性的复杂场景图像,将其检测结果分别与 YOLOv9s、RT-DETR 及 YOLOv11 模型进行可视化对比。

#### 3.6.1 与基准模型 YOLOv9s 的可视化对比

从图8可以看出,在第1个场景中,YOLOv9s对画面中距离较远、尺寸更小的车辆目标存在明显漏检,且对部分已检出的小目标边界框的定位不够精确;而 PRSU-YOLO 则能更全面地覆盖这些小目标。在目标聚集分布的第2个场景中,YOLOv9s 表现欠佳,而 PRSU-YOLO 能够更准确地识别出密集小目标的边界框。在第3个场景中,PRSU-YOLO 对在边缘区域的小目标检出率与定位精度均优于 YOLOv9s。在第4个场景中,PRSU-YOLO 显著超越了 YOLOv9s,能够识别出更多的行人目标,且边界框划分更精细。

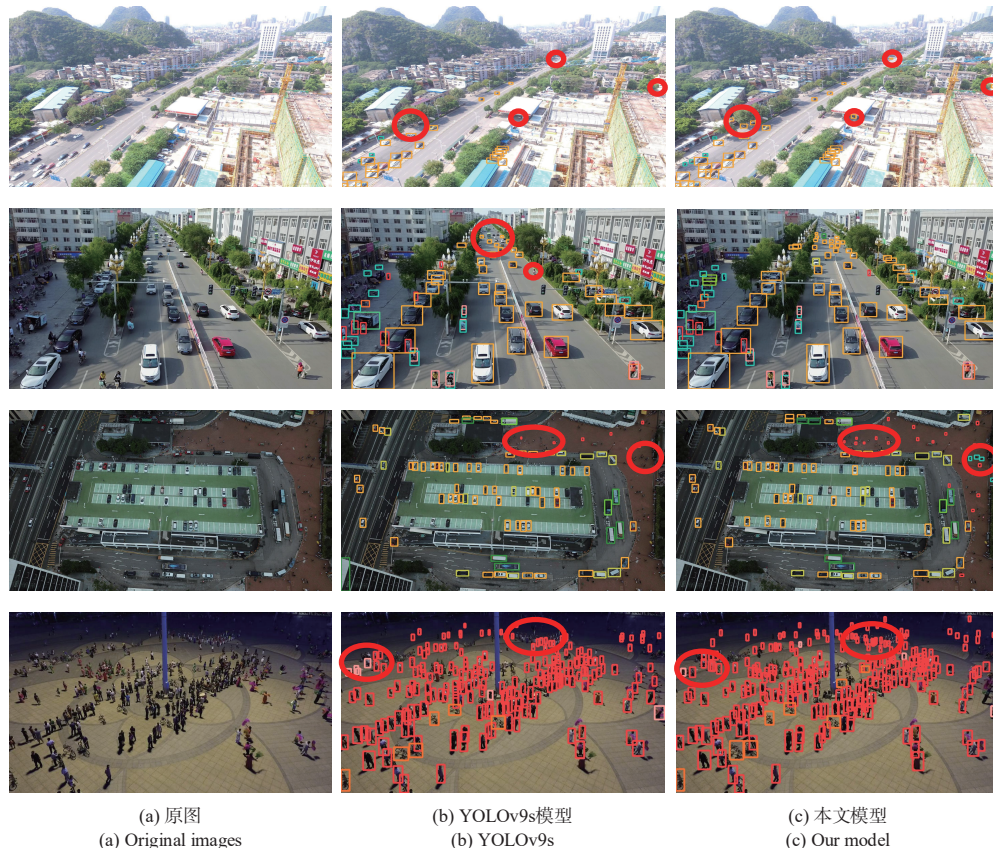


图 8 本文模型与 YOLOv9s 的可视化对比图

Fig. 8 Visualization comparison of the proposed model with YOLOv9s

### 3.6.2 与其他模型的可视化对比

从图 9 可以看出, RT-DETR 模型虽然能够检测到较多的目标类别,但其检测结果的精确性存在明显不足。例如,在第 1 个场景中,其对行人的检测框存在位置不准、多个框体堆叠的问题;在第 2 个场景中,将高层建筑上的窗户错误识别为车辆,产生了严重的误检。YOLOv11 模型在

检测准确性上表现良好,但存在大量的漏检问题。在第 3 个场景中,前景中的车辆未被检出;在第 4 个场景中,右下角区域的大部分小尺寸行人目标也未被发现。相比之下, PRSU-YOLO 模型不仅能够准确识别出目标类别,还能有效地检测出图像中存在的更多目标,尤其是被其他模型忽略的微小或密集目标。

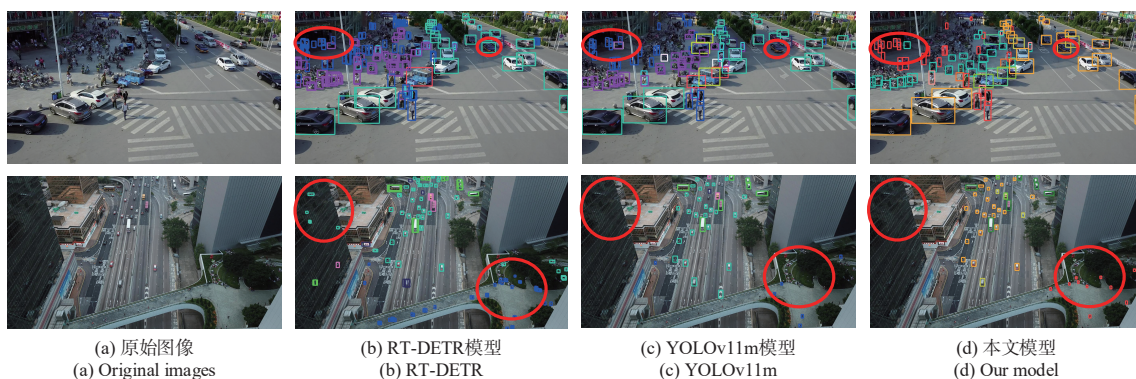


图 9 本文模型与主流模型的可视化对比图

Fig. 9 Visualization comparison of the proposed model with the mainstream models

### 3.7 泛化实验

为验证 PRSU-YOLO 模型的泛化能力,本研究在 TinyPerson 数据集上进行了验证,其具体结果如表 6 所示。

表 6 模型在 TinyPerson 数据集上的对比  
Tab. 6 Comparison on TinyPerson dataset

模型	P/%	R/%	mAP@	
			0.5/%	0.5:0.95/%
YOLOv9s	31.0	14.3	11.5	4.1
本文模型	34.5	15.6	13.6	4.9

从表 6 可以看出,PRSU-YOLO 模型相较于基础模型,其各项指标均取得了显著提升。其中,精确率从 31% 提升至 34.5%,召回率从 14.3% 提升至 15.6%。同时,综合性能指标 mAP@0.5 从 11.5% 提升至 13.6%,定位精度要求更严格的 mAP@0.5:0.95 指标也从 4.1% 提升至 4.9%。这一结果表明,PRSU-YOLO 模型并非针对特定数据集的过拟合优化,而是从根本上提升了模型对小目标本质特征的检测能力。

## 4 结 论

本文针对无人机遥感影像目标检测任务中存在的小目标特征信息匮乏与检测困难等核心挑战,基于 YOLOv9s 模型提出了 PRSU-YOLO 模型。通过引入风车结构卷积自适应模块、重参数化空间-通道卷积模块、小目标检测增强分支以及尺度动态交并比损失函数,构建了一个从特征提取、融合到损失优化的完整增强框架。在 Vis-Drone2019 数据集上的综合实验表明,所提模型在 mAP@0.5 指标上达到 37.4%,相较于基准模型 YOLOv9s 提升 4.4%。在 TinyPerson 数据集上的进一步验证表明,该模型对小目标的泛化能力同样优于基线模型。同时,其计算复杂度相较于基础模型仅增加 14.6 GFLOPS。综上所述,本文提出的 PRSU-YOLO 模型为无人机影像小目标检测提供了一种在精度与效率间取得良好平衡的解决方案。该模型的轻量化特性使其具备向嵌入式设备部署的潜力,未来将在此基础上,集中于在更复杂的实际条件下验证模型的鲁棒性,并探索其在边缘计算平台上的高效部署。

## 参 考 文 献:

- [1] 李大湘,辛嘉妮,刘颖. 位置敏感 Transformer 航拍图像目标检测模型[J]. 光学精密工程,2024,32(5):727-739.  
LI D X, XIN J N, LIU Y. Position-sensitive Transformer aerial image object detection model [J]. *Optics and Precision Engineering*, 2024, 32(5): 727-739. (in Chinese)
- [2] 吕学涵,李富,祁铭瑞,等. 基于改进 YOLOv11s 的无人机小目标检测算法[J]. 液晶与显示,2025,40(11):1744-1756.  
LV X H, LI F, QI M R, et al. Target detection algorithm based on improved YOLOv11s UAV aerial image [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(11): 1744-1756. (in Chinese)
- [3] ZHANG H X, ZHANG H, LIU K, et al. UAV-DETR: efficient end-to-end object detection for unmanned aerial vehicle imagery [C]. 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hangzhou, China: IEEE, 2025: 15143-15149.
- [4] 张志豪,厉小润,陈淑涵. 基于改进 YOLO11 的无人机航拍图像小目标检测算法[J]. 液晶与显示,2025,40(6): 915-930.  
ZHANG Z H, LI X R, CHEN S H. Small object detection algorithm in UAV aerial images based on improved YOLO11 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(6): 915-930. (in Chinese)
- [5] 林珊玲,张雪,陈燕,等. 融合全局信息与双域注意力机制的光学遥感飞机目标检测[J]. 光学精密工程,2024,32(20): 3085-3098.  
LIN S L, ZHANG X, CHEN Y, et al. Integrating global information and dual-domain attention mechanism for optical remote sensing aircraft target detection [J]. *Optics and Precision Engineering*, 2024, 32(20): 3085-3098. (in Chinese)
- [6] 杨路霞,刘泽凯,张红瑞,等. 基于上下文协同感知的航拍小目标检测算法[J]. 液晶与显示,2025,40(9):1308-1320.  
YANG L X, LIU Z K, ZHANG H R, et al. Aerial small target detection algorithm based on context collaborative perception [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(9): 1308-1320. (in Chinese)

- [7] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580-587.
- [8] GIRSHICK R. Fast R-CNN [C]. 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 1440-1448.
- [9] REN S Q, HE K M, 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*, 2017, 39(6): 1137-1149.
- [10] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 779-788.
- [11] WANG C Y, YE H I, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information [C]//*Proceedings of 18th European Conference on Computer Vision—ECCV 2024*. Milan: Springer, 2025: 1-21.
- [12] WANG A, CHEN H, LIU L H, *et al.* YOLOv10: real-time end-to-end object detection [C]//*Proceedings of the 38th International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2024: 3429.
- [13] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements [J/OL]. *arXiv*, 2024: 2410.17725.
- [14] HUO T X, LIU Z Q, ZHANG S C, *et al.* CDNet: object detection based on cross-level aggregation and deformable attention for UAV aerial images [J]. *The Visual Computer*, 2025, 41(7): 4603-4621.
- [15] 谭海英, 杨军. 面向遥感影像的轻量级卷积神经网络目标检测[J]. *遥感技术与应用*, 2025, 40(1): 167-176.
- TAN H Y, YANG J. Lightweight convolutional neural network object detection for remote sensing images [J]. *Remote Sensing Technology and Application*, 2025, 40(1): 167-176. (in Chinese)
- [16] ZHANG H Y, XIAO P, YAO F F, *et al.* Fusion of multi-scale attention for aerial images small-target detection model based on PARE-YOLO [J]. *Scientific Reports*, 2025, 15(1): 4753.
- [17] XU W Y, CUI C, JI Y C, *et al.* YOLOv8-MPEB small target detection algorithm based on UAV images [J]. *Heliyon*, 2024, 10(8): e29501.
- [18] DOLORIEL C T C, CAJOTE R D. Improving the detection of small oriented objects in aerial images [C]. 2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW). Waikoloa: IEEE, 2023: 176-185.
- [19] YANG J N, LIU S L, WU J J, *et al.* Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection [C]. Thirty-Ninth AAAI Conference on Artificial Intelligence. Philadelphia: AAAI Press, 2025: 9202-9210.
- [20] DU D W, ZHU P F, WEN L Y, *et al.* VisDrone-DET2019: the vision meets drone object detection in image challenge results [C]. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul: IEEE, 2019: 213-226.
- [21] YU X H, GONG Y Q, JIANG N, *et al.* Scale match for tiny person detection [C]. 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass: IEEE, 2020: 1246-1254.
- [22] ZHAO Y A, LV W Y, XU S L, *et al.* DETRs beat YOLOs on real-time object detection [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2024: 16965-16974.

#### 作者简介:



罗可心,女,博士研究生,2024年于长春理工大学获得硕士学位,主要从事计算机视觉及无人机影像目标感知方面的研究。E-mail: luokexin0201@163.com



杨华民,男,博士,教授,2007年于长春理工大学获得博士学位,主要从事人工智能、计算机仿真与虚拟现实方面的研究。E-mail: 2024200207@mails.cust.edu.cn