

# 结合重参数化与多层次特征融合的航拍图像小目标检测<sup>①</sup>



曹洁<sup>1,2</sup>, 李立晶<sup>1</sup>, 梁浩鹏<sup>1</sup>

<sup>1</sup>(兰州理工大学 计算机与通信学院, 兰州 733050)

<sup>2</sup>(兰州城市学院 信息工程学院, 兰州 733070)

通信作者: 曹洁, E-mail: [haop1115@163.com](mailto:haop1115@163.com)

**摘要:** 针对无人机航拍图像目标尺寸小、分布密集且被遮挡造成误检漏检等问题, 提出一种结合重参数化思想与多层次特征融合的航拍图像小目标检测算法. 首先, 利用重参数化思想设计了重参数化卷积模块 (reparameterized convolution module, RCM), 与 C2f 模块结合设计了 C2f-RCM 模块, 通过扩大感受野有效绘制上下文信息, 更好地提取图像中的细微特征. 其次, 为解决颈部网络在特征融合部分造成的信息丢失问题, 提出一种多层次特征融合模块 (multi-level feature fusion module, MFFM), 该模块利用跨层次间的信息融合, 有效减少了在遮挡情况下的漏检现象, 使得网络在检测大、中、小目标时能够显著提升准确度. 最后, 提出一种 Inner-Shape IoU 边界框回归损失函数, 通过构建辅助边框和关注边界框的自身形状, 以增强模型的收敛速度. 实验结果表明, 与基线模型相比, 本文方法在 VisDrone2019 中, *mAP@0.5*、*Precision*、*Recall* 分别提高了 5.7%、5.7%、2.4%, 在 AI-TOD 中, *mAP@0.5*、*Precision*、*Recall* 提升了 3.7%、3.9%、5.3%, 验证了本文方法在航拍图像小目标检测方面的有效性.

**关键词:** 小目标检测; 重参数化; 特征融合; 损失函数

引用格式: 曹洁, 李立晶, 梁浩鹏. 结合重参数化与多层次特征融合的航拍图像小目标检测. 计算机系统应用, 2025, 34(4): 218–227. <http://www.c-s-a.org.cn/1003-3254/9811.html>

## Small Target Detection for Aerial Image Combining Reparameterization and Multi-level Feature Fusion

CAO Jie<sup>1,2</sup>, LI Li-Jing<sup>1</sup>, LIANG Hao-Peng<sup>1</sup>

<sup>1</sup>(School of Computing and Communications, Lanzhou University of Science and Technology, Lanzhou 733050, China)

<sup>2</sup>(School of Information Engineering, Lanzhou City University, Lanzhou 733070, China)

**Abstract:** To address the problems of small target size, dense distribution, and occlusion caused false detection and missed detection in unmanned aerial vehicle (UAV) aerial images, this study proposes a small target detection algorithm for aerial images which combines reparameterization and multi-level feature fusion. Firstly, the reparameterized convolution module (RCM) is designed by using the idea of reparameterization, and the C2f-RCM module is designed by combining the RCM with the C2f module, which can effectively draw contextual information by enlarging the sensory field and better extract the subtle features in the images. Secondly, to solve the problem of information loss caused by the neck network in the feature fusion part, this study proposes a multi-level feature fusion module (MFFM), which utilizes cross-level information fusion to effectively reduce the missed detection phenomenon in the case of occlusion, so that the network is able to detect large, medium, and small targets with a significant improved accuracy. Finally, an Inner-Shape IoU bounding box regression loss function is proposed to enhance the convergence speed of the model by constructing

① 基金项目: 甘肃省重点研发计划 (22YF7GA130)

收稿时间: 2024-09-14; 修改时间: 2024-10-30; 采用时间: 2024-11-07; csa 在线出版时间: 2025-02-25

CNKI 网络首发时间: 2025-02-26

auxiliary borders and focusing on the shape of the bounding box. Compared with the baseline model, the proposed method improves  $mAP@0.5$ , *Precision*, and *Recall* by 5.7%, 5.7%, and 2.4% in VisDrone2019 and 3.7%, 3.9%, and 5.3% in AI-TOD, respectively, which verifies that the proposed method is effective in detecting small targets in aerial images.

**Key words:** small target detection; reparameterization; feature fusion; loss function

无人机图像的智能化分析处理可以快速高效地提取地物信息,在工业巡检、交通管理、应急救援、安防等领域发挥着重要的作用<sup>[1]</sup>。无人机航拍图像通常是俯视角拍摄,目标尺寸小、密集且容易受到环境干扰,除此之外,受拍摄角度限制会导致目标被遮挡,这些问题都会导致检测效果不佳。因此,有效地从无人机图像中检测和识别小目标成为一项具有挑战性的任务。

当前,主流目标检测算法根据网络结构主要分为一阶段目标检测算法和两阶段目标检测算法。两阶段目标检测算法如 R-CNN<sup>[2]</sup>、Mask R-CNN<sup>[3]</sup>、Cascade R-CNN<sup>[4]</sup>和 Faster R-CNN<sup>[5]</sup>等,在检测精度上表现出色,但处理速度较慢,不适合应用于无人机实时检测场景。相比之下,一阶段目标检测算法,如 CenterNet<sup>[6]</sup>、RetinaNet<sup>[7]</sup>、SSD<sup>[8]</sup>系列、YOLO<sup>[9]</sup>系列具有实时性好的特点,但对于小目标检测效果仍然不佳。为确保快速检测的同时提高小目标检测精度,过往的研究工作主要集中在优化算法。首先,研究如何有效提取小目标关键特征信息。例如,文献[10]将骨干网络中的普通卷积模块替换为深度可分离卷积模块;文献[11]设计并在骨干网络中嵌入一个注意力模块 SMSE,增强对复杂背景下无人机小目标的关注度;文献[12]将特征提取网络中的激活函数 LeakyReLU 替换为 SiLU,以提升模型训练时的收敛速度与模型泛化性。文献[13]通过在原骨干网络中结合三维注意力机制,帮助网络更好地聚焦于图像中的关键信息;文献[14]通过在骨干网络中加入空间金字塔池(SPP)、卷积块注意模块(CBAM)、沙钟特征提取模块(SCFEM)、Ghost 模块等卷积层,提高网络的精度。以上研究通过替换卷积模块、添加注意力机制等方式提高骨干网络对有效信息的关注度。其次,探索如何增强特征表达,文献[15]提出一种稀疏连接渐近特征金字塔网络(SCAFPN)的方法,旨在消除特征融合过程中的层间干扰,从而提高模型性能。文献[16]设计了一种 PrFuFPN 的颈部结构来改进整个网络结构,同时在检测头部分增加一个小目标检测头,以实现更高级的多尺度融合。文献[17]提出一种双向密集

特征金字塔网络(BDFPN)的网络结构,其目的是将浅层特征图中的细粒度信息与深层特征图中的语义信息进行整合,通过扩大特征金字塔的层次结构,并结合跳跃连接,实现对多尺度特征的优化融合。以上文献通过设计不同的特征融合策略,使网络能够更有效地融合多种特征,从而增强网络检测小目标的能力。尽管上述算法根据无人机图像的特点从不同角度入手进行了一定的优化,但这些方法没有考虑到检测目标重叠或被遮挡时,模型会因提取不到遮挡物体的空间信息而导致候选框定位不准确和漏检误检,以及随着网络层数的增加而丢失小目标信息的问题。

针对此情况,需要更加丰富的空间位置信息才能更好地定位被遮挡目标。为提高无人机图像小目标检测的精确度,本文提出结合重参数化与多层次特征融合的航拍图像小目标检测算法,首先提出一种重参数化卷积模块(RCM),并与 C2f 模块结合设计了 C2f-RCM 模块,在不增加网络深度和参数量前提下,扩大骨干网络的感受野,丰富小目标特征。其次,提出一种多尺度卷积注意力特征融合结构(MFFM),实现了跨层次间的信息融合,改善基于自顶向下思想的特征融合网络造成的信息损失。最后,提出一种损失函数 Inner-Shape IoU,在构造辅助边框前提下,更加关注边界框本身的形状和尺度来计算损失,有效提高了模型的收敛速度。

## 1 YOLOv8s 算法

面向无人机图像的小目标检测中,对算法的计算复杂度和参数量有很大限制,本文选用 YOLOv8s 为基线模型。YOLOv8s 网络结构包括骨干网络、颈部网络和检测头。骨干网络的主要构成是基本卷积层、C2f 模块和空间金字塔池化模块。颈部网络采用路径聚合网络与特征金字塔网络(path aggregation network and feature pyramid network, PAN-FPN)的思想,将 FPN 构建的特征金字塔与 PAN 的上下采样和横向连接相结合,对骨干网络的输出直接进行上采样操作,并结合 C2f 模块提取特征。检测头使用了解耦头结构,包含 3 个检

测头, 每个头负责不同尺寸的目标检测. 但是直接将 YOLOv8s 算法应用于无人机图像检测存在以下问题: (1) 无人机图像目标尺寸小且目标密集, 骨干网络对密集目标特征提取能力较弱, 导致检测精度不高; (2) PAN-FPN 间隔层之间信息融合过程中会造成信息丢失.

## 2 本文方法

无人机图像目标尺寸小、分布密集, 且受拍摄视角限制存在遮挡的情况, 会导致漏检误检的问题. 另一方面, YOLOv8s 存在小目标检测精度低的问题. 本文提出了结合重参数化和多层次特征融合航拍图像小

目标检测算法, 网络结构如图 1 所示, 整体网络模型包含基于 C2f-RCM 模块的骨干网络、基于多层次特征融合模块 MFFM 的颈部网络和检测头. 基于 C2f-RCM 模块的骨干网络在高层扩张感受野, 提取丰富的语义信息. 基于多层次特征融合模块的颈部网络利用浅层特征融合模块 (shallow feature fusion module, SFFM) 聚合不同尺寸的特征图, 为后续检测提供丰富的小目标局部信息, 接着利用深层特征融合模块 (deep feature fusion module, DFFM) 聚集深层特征图, 以提高大目标和中目标的准确率. 检测头作为最终预测阶段, 利用 3 种不同尺寸的特征图来检测不同大小的目标.

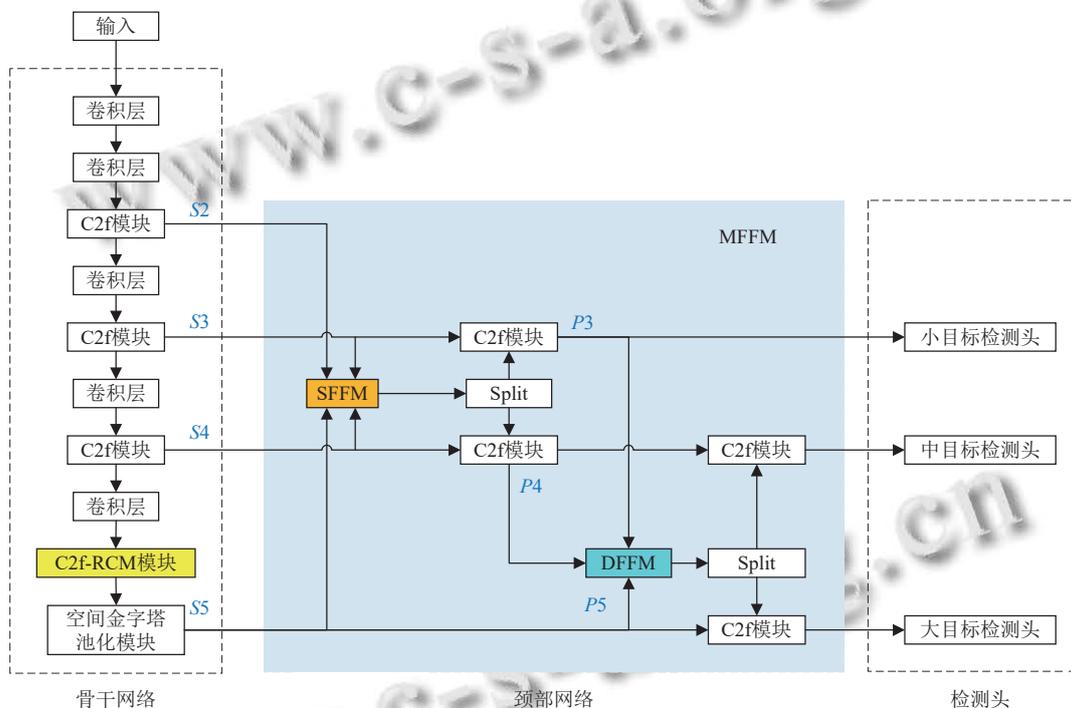


图 1 本文方法网络结构图

### 2.1 C2f-RCM 模块

通常情况下, 无人机图像中背景信息远超于目标信息, 使用针对大、中目标设计的骨干网络会提取到大量无效背景信息, 对小目标检测而言并不有利. 除此之外原始算法中下采样操作导致 C2f 模块对小目标检测效果不佳, 且该模块参数量较大, 本文设计了 C2f-RCM 模块, 该模块通过扩张感受野有效绘制上下文信息, 提高了小目标检测精度, 同时降低了参数量. C2f-RCM 模块的设计思路是把 C2f 模块中 Bottleneck 层的标准卷积层替换为重参数化卷积模块, 可以捕捉更复杂的细节信息, 有效提升得对小目标检测效果. 关于重

参数化卷积模块和 C2f-RCM 模块的详细介绍见第 2.1.1 节和第 2.1.2 节.

#### 2.1.1 重参数化卷积模块 RCM

Wei 等人<sup>[18]</sup>使用空洞卷积改变了网络的感受野, 强调了感受野大小对于特征提取效率的重要性, 但该方法可能导致局部信息丢失和远距离信息不相关等问题. 在此基础上, 本研究设计了一种重参数化卷积模块 RCM, 旨在有效捕捉图像中的广阔上下文信息, 同时保持计算效率. 该模块集成了 3 种不同尺寸的卷积核: 一个标准的 3×3 卷积核, 以及两个重参数化的 5×5 和 7×7 卷积核. 3×3 卷积核因其较低的计算复杂度

而被广泛使用,但它可能无法有效捕获较大范围的特征.相比之下,5×5和7×7卷积核能够覆盖更广的图像区域,从而捕捉到更丰富的上下文信息,但其较高的计算复杂度限制了其在深度学习模型中的直接应用.为克服这一挑战,本文借鉴Ding等人<sup>[19]</sup>提出的采用扩张的小核心卷积层来增强非扩张的大核心层的思想,在不显著增加计算负担的情况下,有效地利用较大的卷积核.通过重参数化,可以将大尺寸卷积核转换为等效的、计算效率更高的形式,从而在保持模型性能的同时,降低计算成本.最后,为进一步优化信息流并防止梯度消失,在RCM中引入了残差连接.残差连接有助于保留输入特征的原始信息,使得网络能够更有效地进行训练和特征传递.

该模块融入骨干网络C2f模块,替换其中的Bottleneck层.重参数化卷积模块结构如图2(a)所示,重参数化5×5卷积和重参数化7×7卷积可视化如图2(b)与图2(c)所示,重参数化5×5卷积是由卷积核 $K=(5, 3, 3)$ ,空洞率 $r=(1, 1, 2)$ 的3个卷积并行经过归一化后相加;重参数化7×7卷积是由卷积核 $K=(7, 5, 3, 3)$ ,空洞率 $r=(1, 1, 2, 3)$ 的4个卷积并行经过归一化后相加.重参数化卷积使用一个非膨胀的小卷积核和多个膨胀的小卷积核层来增强非膨胀的中等卷积核,从参数角度看等同于一个具有更大稀疏核心的非扩张卷积层,从而在保持可学习参数数量和计算效率的同时,增强网络对空间信息的捕获能力,同时为卷积网络提供更广泛的感受野,且不会增加模型的深度.

假设输入为通道数 $2C$ 的特征图 $T \in R^{2C \times H \times W}$ ,输出特征图为 $F_{\text{output}} \in R^{2C \times H \times W}$ .首先,对输入特征图使用3×3的标准卷积,结合批处理归一化层和ReLU层提取特征,整体输出为几个需要细化的小空间区域的集合 $\{T_A, T_B, T_C\} \in R^{3C \times H \times W}$ ,如式(1)所示:

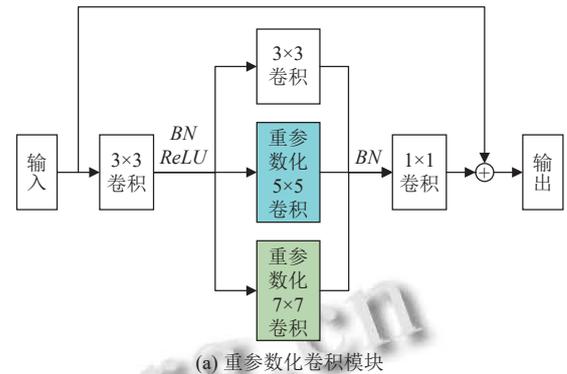
$$\{T_A, T_B, T_C\} = \text{ReLU}(\text{BN}(\text{Conv}(T))) \quad (1)$$

其次,分别将细小区域集合中的每一个元素 $T_A \in R^{C \times H \times W}$ ,  $T_B \in R^{C \times H \times W}$ ,  $T_C \in R^{C \times H \times W}$ 通过并行的卷积模块. $T_A \in R^{C \times H \times W}$ 通过第1个分支3×3卷积,如式(2)所示:

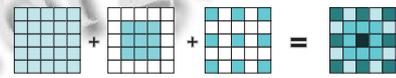
$$T_{A'} = \text{Conv}3 \times 3(T_A) \in R^{C \times H \times W} \quad (2)$$

$T_B \in R^{C \times H \times W}$ 通过第2个分支重参数化5×5卷积,如式(3)所示.重参数化5×5卷积处理过程如图2(b)所示,使用两个扩张率分别为1和2的3×3卷积增强普通5×5卷积.

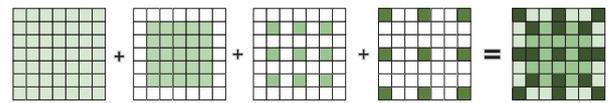
$$T_B' = \text{ReparamConv}5 \times 5(T_B) \in R^{C \times H \times W} \quad (3)$$



(a) 重参数化卷积模块



(b) 重参数化5×5卷积可视化



(c) 重参数化7×7卷积可视化

图2 重参数化卷积结构

$T_C \in R^{C \times H \times W}$ 通过重参数化7×7卷积,如式(4)所示.重参数化7×7卷积处理过程如图2(c)所示,使用一个扩张率为5×5卷积和两个扩张率分别为2和3的3×3卷积增强普通7×7卷积.

$$T_{C'} = \text{ReparamConv}7 \times 7(T_C) \in R^{2C \times H \times W} \quad (4)$$

然后将 $T_{A'} \in R^{C \times H \times W}$ ,  $T_B' \in R^{C \times H \times W}$ ,  $T_{C'} \in R^{2C \times H \times W}$ 这3个输出特征通过1×1卷积对通道降维,如式(5)所示:

$$F = \text{Conv}1 \times 1(\text{BN}(T_{A'}, T_B', T_{C'})) \in R^{2C \times H \times W} \quad (5)$$

最后,将输出特征 $F \in R^{2C \times H \times W}$ 与输入特征 $T \in R^{2C \times H \times W}$ 进行像素级相加,如式(6)所示:

$$F_{\text{output}} = (F + T) \in R^{2C \times H \times W} \quad (6)$$

### 2.1.2 C2f-RCM 模块结构

图3为C2f-RCM模块的结构图,将C2f模块中的Bottleneck标准卷积层替换为重参数化卷积模块RCM.C2f-RCM模块首先通过一个卷积层,接受输入特征图并生成中间特征图.然后将生成的中间特征图通过Split操作拆分为两部分,一部分直接传递拼接模块,另一部分输入多个串行的RCM进行处理,对输出特征图直接进行拼接,使模型可以综合利用多尺度、多层次的信息.最后将拼接后的特征图输入一个最终的卷积层进行处理,生成最终输出特征图.

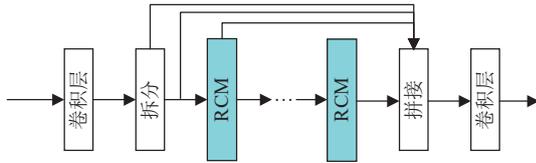


图3 C2f-RCM模块

### 2.2 多层次特征融合模块 MFFM

原始颈部网络结合了FPN构建的特征金字塔与PAN的上下采样和横向连接,但对于非连续层特征,仅能通过“递归”方式获取,这导致了算法运行速度慢、特征交叉性不足以及信息丢失等问题,限制了信息融合的整体效果<sup>[20]</sup>。针对该问题,受聚集分发机制思想的启发,在颈部网络中设计了一个多层次特征融合模块MFFM。该模块不仅能有效获取对小目标检测有益的浅层信息,还能促进浅层特征与深层特征的融合。MFFM主要由浅层特征融合模块SFFM与深层特征融合模块DFFM组成,第2.2.1、2.2.2和2.2.3节分别对SFFM、DFFM和MFFM整体结构进行详细介绍。

#### 2.2.1 浅层特征融合模块 SFFM

浅层特征图对于无人机图像中小目标检测来说具有丰富的有效信息。然而,在进行全局相关性特征融合时,会带来较大的计算量和显存消耗。采用卷积神经网络计算局部特征可以有效减少这些问题。通过构建卷积块来形成浅层特征融合模块SFFM,可以融合不同尺度的特征图,从而增强浅层特征图中的语义信息,并为后续检测提供富含小目标信息的浅层特征图。SFFM的具体结构如图4所示。

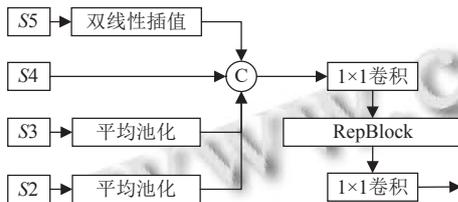


图4 浅层特征融合模块

SFFM首先对骨干网络每一阶段输出的不同层级不同大小的特征进行拼接,选择以S4输出尺寸大小为基准,对S2和S3大特征图通过平均池化进行下采样,对S5小特征图通过双线性插值进行上采样,以实现特征图大小统一。其次通过Concat操作将特征图合并,合并后的特征 $F_C$ 计算方式可表示为式(7):

$$F_C = \text{Concat}(\text{Avgpool}(S2), \text{Avgpool}(S3), S4, \text{Bilinear}(S5)) \quad (7)$$

然后处理输出的特征图,先使用 $1 \times 1$ 卷积调整通道数,使用RepBlock模块进行特征融合,之后再次使用 $1 \times 1$ 卷积调整通道数,此时输出特征图计算公式如式(8):

$$F_F = \text{Conv}1 \times 1(\text{RepBlock}(\text{Conv}1 \times 1(F_C))) \quad (8)$$

最后,将通过SFFM获得的具有局部信息的特征图进行Split操作拆分为两部分注入C2f模块。

#### 2.2.2 深层特征融合模块 DFFM

浅层特征图富含局部细节信息,而深层特征图则包含丰富的语义信息。为了提高小目标检测的精度,需要同时利用这两类信息。因此,我们设计了深层特征融合模块DFFM,它采用多头注意力机制来生成全局信息,从而为后续的检测分支提供高效的特征表示。DFFM的具体结构如图5所示。

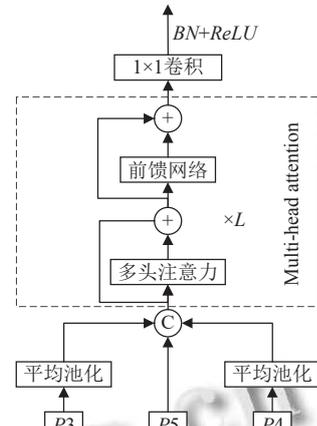


图5 深层特征融合模块

首先以P5为基准,使用平均池化操作对P3、P4降维到统一尺寸,其次通过Concat操作将特征图合并,如式(9)所示:

$$F_c = \text{Concat}(\text{Avgpool}(P3), \text{Avgpool}(P4), P5) \quad (9)$$

然后将合并后的特征图通过Multi-head attention模块进行特征融合,如式(10)所示:

$$F_f = \text{MultiHead}(F_c) \quad (10)$$

Multi-head attention模块有L层,每一层由1个多头注意力模块、1个前馈网络以及残差连接组成,通过统一的模块对各个尺度的信息进行采集和融合,然后将融合够的特征注入颈部网络,既避免了传统FPN结构固有的信息丢失,同时增强了颈部网络的信息融合能力。

#### 2.2.3 多层次特征融合模块 MFFM

MFFM结构如图6所示,首先通过SFFM融合并

提取骨干网络每一层输出特征图中浅层的特征信息,其次对输出特征进行 Split 操作拆分为两部分,分别传入两个 C2f 模块,分别与特征图 S3 和 S4 进行融合.然后使用 DFFM 将两个 C2f 模块输出与骨干网络最后一层输出 S5 进行特征融合,最后使用 Split 操作分别传入两个 C2f 模块,输入检测头进行目标检测.

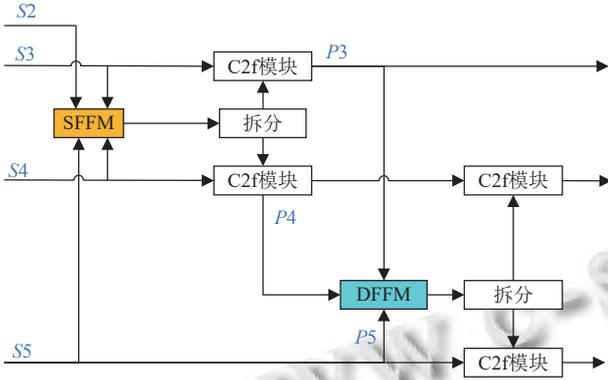


图6 多层次特征融合模块

### 2.3 Inner-Shape IoU 损失函数

传统的边界框回归方法普遍采用交并比 ( $IoU$ ) 损失函数来量化预测边界框与真实边界框的重叠面积,然而,这种做法未能充分考虑边界框的形状和尺度特性,这可能会影响回归结果的精确度.为了克服这一缺陷,本文提出了一种新的 Inner-Shape IoU 损失函数,旨在优化传统  $IoU$  损失函数的不足.该方法通过构建辅助边界框,专注于分析边界框的形状和尺度特征.通过这种方式,Inner-Shape IoU 损失函数能够更精确地捕捉目标形状和尺度的变化,进而对回归结果产生更准确的影响,从而提升边界框回归的精度.

Inner-IoU<sup>[21]</sup>通过构建辅助边界框计算  $IoU$ ,引入比例因子  $ratio$  控制辅助边界框的比例大小,Inner-IoU 的表达式如式 (11)–式 (14) 所示:

$$L_{Inner-IoU} = 1 - IoU^{inner} \quad (11)$$

$$IoU^{inner} = \frac{inter}{union} \quad (12)$$

$$inter = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) \times (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)) \quad (13)$$

$$union = (w^{gt} \times h^{gt}) \times ratio^2 + (w \times h) \times ratio^2 - inter \quad (14)$$

Shape-IoU<sup>[22]</sup>关注边界框的形状和尺度对边界框回归的影响,Shape-IoU 的表达式如式 (15)–式 (17):

$$L_{Shape-IoU} = 1 - IoU + D^{shape} + 0.5 \times \Omega^{shape} \quad (15)$$

$$D^{shape} = hh \times (x_c - x_c^{gt})^2 / c^2 + ww \times (y_c - y_c^{gt})^2 / c^2 \quad (16)$$

$$\Omega^{shape} = \sum_{t=w,h} (1 - e^{-w_t})^\theta, \theta = 4 \quad (17)$$

其中,  $ww$  与  $hh$  分别表示水平方向与垂直方向上的权重系数,其值与真实框的形状有关.

本文利用 Inner-IoU 构建辅助边框损失的思想改造 Shape-IoU,替换了 Shape-IoU 中  $IoU$  的计算部分,  $L_{Inner-Shape}$  表达式如式 (18):

$$L_{Inner-Shape} = IoU^{inner} - D^{shape} - 0.5 \times \Omega^{shape} \quad (18)$$

$L_{Inner-Shape}$  由 3 部分组成:  $IoU^{inner}$  是基于辅助边界框计算的  $IoU$ ,  $D^{shape}$  是真实框与预测框中心点距离的惩罚项,  $\Omega^{shape}$  是真实框与预测框形状的惩罚项.通过构建 Inner-Shape IoU 损失函数提高了模型的收敛速度,同时提升了在目标检测任务中的泛化能力.

## 3 实验准备

### 3.1 数据集

本文采用的数据集是由天津大学机器学习和数据挖掘实验室 AISKYEYE 团队收集的 VisDrone2019<sup>[23]</sup>.共包含 10 209 张静态图像,其中 6 417 张图像用于训练,548 张图像用于验证,3 190 张图像用于测试.该数据集是由各种无人机摄像头捕获,覆盖范围广泛,提供了 11 个类别,分别是“pedestrian”“people”“bicycle”“car”“van”“truck”“tricycle”“awning-tricycle”“bus”“motor”和“others”,其中“others”是非有效目标区域,在实验过程中予以忽略.

此外,使用小目标丰富的遥感数据集 AI-TOD 验证本文方法的有效性.AI-TOD 是用于航空图像微小物体检测的数据集,包括 28 036 张航拍图像,提供 700 621 个对象实例,涵盖 8 个类别.通过使用这样一个大规模、多样化的数据集进行验证,可以更充分地评估模型在检测小目标方面的表现,并且有助于确保模型的泛化能力和有效性.

### 3.2 实验配置

实验环境运行在 Linux 系统,使用 2 张 NVIDIA GeForce RTX3090 GPU 进行训练,实验软件环境是 Python 3.10.12、Torch 1.12.1、CUDA 10.1 等,所有实验均在相同实验环境和超参数下完成的.训练轮数设

置为 200 epoch, 学习速率为 0.01, 优化方式为随机梯度下降法 (SGD).

### 3.3 实验评价指标

本研究选取了目标检测常用的 5 种评价指标: 准确率 (*Precision*)、召回率 (*Recall*)、平均精确度均值 (*mAP*)、浮点运算次数 (GFLOPs) 和推理时间 (FPS) 来评估本文改进算法的有效性. 接下来将介绍每个指标的具体含义.

#### (1) 精确度

精确度是指被模型正确预测的正样本数量与所有预测为正样本的样本数量之比. 计算公式如式 (19):

$$Precision = \frac{TP}{TP + FP} \quad (19)$$

其中, *TP* 表示模型正确预测为正类别的样本数量, *FP* 表示模型错误预测为正类别的样本数量. 在目标检测中, 精确度高表示模型在判断为正样本时较为准确.

#### (2) 召回率

召回率是指被模型正确预测的正样本数量与所有实际正样本数量之比. 计算公式如式 (20):

$$Recall = \frac{TP}{TP + FN} \quad (20)$$

其中, *FN* 是模型错误预测为负类别的样本数量. 召回率衡量了模型对于实际正类别样本的覆盖程度, 即模型能够正确检测出多少正类别样本.

#### (3) 平均精确度均值

平均精确度均值通常用于多类别目标检测任务, 是对所有类别检测结果进行综合评价的指标, 在不同类别的检测结果上, 平均精确度均值可以反映算法的整体性能. 计算平均精确度均值需要先计算每个类别的平均精确度, 再取所有类别的平均精确度的平均值. 计算公式如式 (21):

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i \quad (21)$$

其中, *C* 表示类别的数量, *AP<sub>i</sub>* 表示每个类别的 *AP* 值. 平均精确度越高表示算法在各个类别下的检测效果越好, 但平均精确度均值并不能反映出算法的速度和计算资源的消耗, 因此需要结合浮点运算次数和推理时间等指标进行综合评价.

#### (4) 浮点运算次数

浮点数运算次数 (GFLOPs) 表示模型在测试或推理过程中执行的浮点运算次数, 通常用于衡量模型的

计算复杂度.

#### (5) 推理时间

推理时间 (FPS) 表示模型每秒处理的图像帧数. 它是一个衡量模型推理速度的指标, 用于评估模型在实时应用中的性能.

## 4 实验结果分析

### 4.1 C2f-RCM 对检测效果的影响

为深入探究 C2f-RCM 模块在骨干网络中的不同位置和数量对检测精度的影响, 进行了一系列实验, 通过将骨干网络中的 C2f 模块替换为 C2f-RCM 模块来实现. 鉴于 C2f-RCM 模块在深层特征提取方面的优势, 主要关注在骨干网络的最后两层进行模块替换, 并设计了 3 种不同方案来考察其效果: 方案 1: 仅替换第 3 阶段 C2f 模块为 C2f-RCM 模块; 方案 2: 仅替换第 4 阶段 C2f 模块为 C2f-RCM 模块; 方案 3: 替换第 3、4 阶段 C2f 模块为 C2f-RCM 模块. 实验结果如表 1 所示, 通过分析表中数据可以看出, 在 3 种方案中, 仅将第 4 阶段的 C2f 模块替换为 C2f-RCM 模块 (方案 2) 时, 检测准确度相较于其他方案有最显著的提升, 达到了 3.3%. 此外, 选择方案 2 时, 模型的参数量也有所减少, 这表明在这种配置下, 模型在保持甚至提高检测准确度的同时, 还实现了更高的效率. 这些实验结果揭示了 C2f-RCM 模块在骨干网络中的最佳放置位置, 即在网络的深层阶段进行替换, 可以有效提升检测精度, 并且不会引入过多的参数负担.

表 1 不同位置 C2f-RCM 对检测结果的影响

方案	Params (M)	<i>mAP@0.5</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	GFLOPs	FPS
方案1	13.7	41.4	53.0	39.7	30.5	115
方案2	10.7	42.2	55.3	40.7	28.2	130.7
方案3	13.6	41.1	51.9	38.6	27.8	99.8

### 4.2 比例因子消融实验

为了确定 Inner-Shape IoU 中比例因子 *ratio* 的最佳取值, 进行了 *ratio* 消融实验. 在实验中设置, 并观察了检测结果的变化. 当 *ratio*=1 时, Inner-Shape IoU 损失函数实际上等同于 Inner-IoU 损失函数. 特别是对于小目标而言, 由于其尺寸较小, 相对于高 *IoU* 的样本, 其需要的回归精度可能较低. 通过分析表 2 中的实验结果, 在本文方法中, 当 *ratio*=0.7 时, 实验结果较好. 将 *ratio* 设置为较小的值有助于加快具有高 *IoU* 的样本的

回归速度,从而提高检测速度.然而,并非所有情况下 *ratio* 取值越小越好,具体的最优取值需要根据具体的实验设置进行调整.

表2 *ratio* 消融实验 (%)

<i>ratio</i>	<i>mAP@0.5</i>	<i>Precision</i>	<i>Recall</i>
0.5	38.4	48.7	37.9
0.6	39.7	50.2	39.3
0.7	<b>39.8</b>	<b>52.3</b>	<b>39.6</b>
0.8	38.5	49.6	37.7
0.9	38.5	49.7	38.5
1.0	39.6	51.6	39.8
1.1	38.5	49.7	37.8
1.2	38.5	49.7	37.8
1.3	38.5	49.8	37.8
1.4	38.4	49.7	37.9
1.5	38.4	49.7	37.9

### 4.3 消除实验

为了验证本文方法的改进效果,选取了小目标丰富的 VisDrone2019 数据集进行测试,在 YOLOv8s 骨干网络中加入 C2f-RCM、颈部网络加入 MFFM 模块、替换损失函数为 Inner-Shape IoU. 实验结果如表 3 所示. 表 3 中第 1 组实验为 YOLOv8s 原模型实验结果,作为基准参考. 第 2 组实验是使 C2f-RCM, 根据实验数据可以看出, C2f-RCM 模块在减少参数量和计算量的同时,将 *mAP@0.5* 的指标提升了 3.3%,证明了该模块的有效性. 第 3 组实验是对比加入 MCAFF 模块的效果,根据实验结果可以看出,与原始基于 PAN-FPN 思想的颈部网络相比,改进后的颈部网络有效避免了跨层次间信息融合时的信息丢失,提升了检测精度. 第 4 组实验则是针对损失函数的改进,将传统的 CIoU 替换为 Inner-Shape IoU. 实验结果显示,这种替换显著提升了检测精度,验证了新的损失函数在目标检测任务中的优越性. 第 5-8 组实验则是将不同的改进点进行组合,以验证本文提出的改进算法的整体有效性. 这些组合实验的结果一致表明,经过改进的算法在小目标检测上取得了显著的性能提升. 综上,本文所提出的改进策略不仅各自有效,而且组合使用时能够相互增强,从而在小目标检测任务中实现了更加精确和鲁棒的性能. 这些实验结果充分证明了本文改进算法的有效性和实用价值.

### 4.4 VisDrone 数据集对比实验

本文在 VisDrone2019 数据集上进行了对比实验,以 YOLOv8s 作为基准模型,并与近年来的一系列先进方法进行了全面对比. 这些方法包括 Faster R-CNN、

RetinaNet、YOLOv5s、YOLOv7-tiny、YOLOv8s、YOLOv10s、PrFu-YOLO、SOD-YOLO<sup>[24]</sup>、文献[25]中提出的方法以及 EA-YOLO<sup>[26]</sup>等. 根据表 4 数据可知,在关键的性能指标 *mAP@0.5*、*Precision*、*Recall* 上,本文算法达到了 44.6%、56.3%、41.3%,相较于 YOLOv8s 提高了 5.7%、6.3%、2.4%;相较于复杂度最低的 YOLOv7-tiny 提高了 7.4%、8.8%、1.5%;相较于最新的无人机图像小目标检测模型 PrFu-YOLO、SOD-YOLO<sup>[24]</sup>、文献[25]和 EA-YOLO<sup>[26]</sup>关键指标 *mAP@0.5* 有明显的提升. 通过对比以上指标证明了本文所提出的模型有效提升了小目标的检测精度. 此外,较高的复杂度会影响网络的实时性,本文所提算法的复杂度为 30.3,相较于基础模型有所提升,相比于复杂度最低的 YOLOv7-tiny 增加了 2.2 倍. 但综合其他评价指标,实时性仍满足无人机实时部署的基本要求. 综上,本文提出的算法在保持较高检测精度的同时,也兼顾了实时性,使其成为适合无人机应用场景的有效解决方案.

表3 消融实验

编号	C2f-RCM	MFFM	Inner-Shape IoU	Params (M)	<i>mAP@0.5</i> (%)	GFLOPs
1	—	—	—	11.1	38.9	28.8
2	√	—	—	10.7	42.2	28.1
3	—	√	—	13.6	41.8	30.3
4	—	—	√	11.2	39.8	28.9
5	√	√	—	13.1	43.6	29.7
6	√	—	√	10.6	42.9	28.1
7	—	√	√	13.6	42.7	29.9
8	√	√	√	<b>13.1</b>	<b>44.6</b>	<b>30.3</b>

表4 VisDrone 数据集对比实验

模型	Params (M)	<i>mAP@0.5</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	GFLOPs	FPS
Faster R-CNN <sup>[5]</sup>	43	29.2	40.5	28.8	207.1	15.2
RetinaNet <sup>[7]</sup>	15.3	28.7	39.5	27.6	26.6	40
YOLOv5s	7.2	38.2	50.4	37.3	16.8	115
YOLOv7-tiny <sup>[27]</sup>	6.3	37.2	47.5	39.8	13.4	102
YOLOv8s	11.1	38.9	50.6	38.9	28.6	130
YOLOv10s <sup>[28]</sup>	8.1	39.1	49.8	37.9	24.8	122.4
PrFu-YOLO <sup>[16]</sup>	10.3	43.7	—	—	—	67.13
SOD-YOLO <sup>[24]</sup>	6.9	39.2	—	—	12.9	—
文献[25]	—	38.2	66.5	59.1	—	23
EA-YOLO <sup>[26]</sup>	—	39.9	50.9	38.7	20.9	—
本文方法	<b>13.1</b>	<b>44.6</b>	<b>56.3</b>	<b>41.3</b>	<b>30.3</b>	<b>86</b>

### 4.5 AI-TOD 数据集实验结果

选择 AI-TOD 数据集进一步验证本文方法在小目

标检测任务中的有效性. 所有算法输入图片的大小统一设置为  $800 \times 800$  像素. 选取了 YOLOv8s、SS-YOLO 和文献[29]这 3 个模型进行对比, 根据表 5 实验结果可以看出, 针对更小尺寸的目标, 本文方法在  $mAP@0.5$  这一关键指标上有提升, 证明了本文算法在小目标检测方面的有效性. 具体来说, 检测精度的提升意味着在处理小目标时, 本文方法能够准确识别和定位目标, 减少漏检和误检的情况.

表 5 AI-TOD 数据集对比实验

模型	Params (M)	Precision (%)	Recall (%)	$mAP@0.5$ (%)
YOLOv8s	11.1	50.3	40.6	40.2
SS-YOLO <sup>[30]</sup>	—	—	—	42.6
文献[29]	—	—	—	36.8
本文方法	13.1	54.5	44.2	43.9

#### 4.6 可视化分析

图 7、图 8 分别展示了在相同实验条件和参数设置下, YOLOv8s 原始模型与本文方法在无人机图像目标检测任务中的表现. 具体来说, 图 7 对比了两种模型在目标稀疏场景下的检测效果, 稀疏与目标密集的目标检测效果. 在图 7 中, 左侧为 YOLOv8s 模型的检测效果图, 右侧为本文方法的检测效果图. 通过对比可以看出, 在目标稀疏的情况下, 本文方法显著降低了漏检和误检的发生, 显示出更好的检测性能. 同样地, 在图 8 的对比中, 目标密集的场景下, 本文方法同样有效地减少了漏检现象. 这些实验结果一致表明, 本文方法在小目标检测方面取得了显著提升, 有效提高了检测准确率和鲁棒性.

## 5 结论

本文是以 YOLOv8s 为基础模型, 聚焦于现有目标检测器对小目标检测不佳、误检漏检等问题, 提出结合重参数化与多层次特征融合的航拍图像小目标检测算法. 首先, 设计了 C2f-RCM 模块, 通过扩大感受野获取有效的上下文信息. 然后, 在颈部网络中设计了多层次特征融合模块, 该模块有效地整合了网络中的浅层和深层特征, 增强了特征的表达能力, 有助于更准确地定位小目标. 最后, 提出了 Inner-Shape IoU 损失函数, 以增强模型的收敛速度. 在 VisDrone2019 数据集上进行了详尽的实验验证, 结果表明, 相比于主流检测器, 本文方法在 VisDrone2019 数据集中检测精度达到了 44.6%, 保证实时性的情况下, 能够尽可能多检测到小目标, 减少了误检漏检的情况.

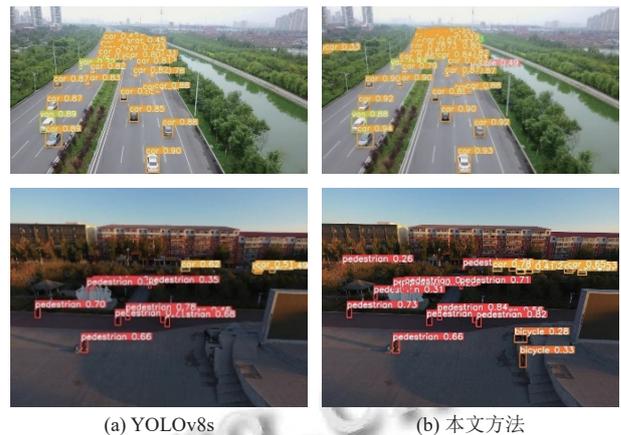


图 7 稀疏目标检测效果对比图

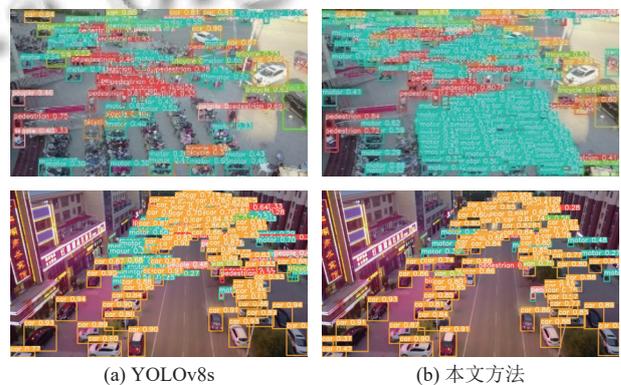


图 8 密集目标检测效果对比图

## 参考文献

- 江波, 屈若锟, 李彦冬, 等. 基于深度学习的无人机航拍目标检测研究综述. 航空学报, 2021, 42(4): 524519.
- Chen CY, Liu MY, Tuzel O, et al. R-CNN for small object detection. Proceedings of the 13th Asian Conference on Computer Vision. Taipei: Springer, 2016. 214–230.
- He KM, Gkioxari G, Dollár P, et al. Mask R-CNN. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980–2988.
- Cai ZW, Vasconcelos N. Cascade R-CNN: High quality object detection and instance segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1483–1498. [doi: 10.1109/TPAMI.2019.2956516]
- Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137–1149.
- Duan KW, Bai S, Xie LX, et al. CenterNet: Keypoint triplets for object detection. Proceedings of the 2019 IEEE/CVF

- International Conference on Computer Vision. Seoul: IEEE, 2019. 6568–6577.
- 7 Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017. 2980–2988.
  - 8 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
  - 9 Zeng S, Yang WZ, Jiao YY, *et al.* SCA-YOLO: A new small object detection model for UAV images. The Visual Computer, 2024, 40(3): 1787–1803. [doi: [10.1007/s00371-023-02886-y](https://doi.org/10.1007/s00371-023-02886-y)]
  - 10 陈卫彪, 贾小军, 朱响斌, 等. 基于 DSM-YOLO v5 的无人机航拍图像目标检测. 计算机工程与应用, 2023, 59(18): 226–233. [doi: [10.3778/j.issn.1002-8331.2302-0324](https://doi.org/10.3778/j.issn.1002-8331.2302-0324)]
  - 11 薛珊, 安宏宇, 吕琼莹, 等. 复杂背景下基于 YOLOv7-tiny 的图像目标检测算法. 红外与激光工程, 2024, 53(1): 20230472.
  - 12 张光华, 李聪发, 李钢硬, 等. 基于改进 YOLOv7-tiny 的无人机航拍图像小目标检测算法. 工程科学与技术. <https://link.cnki.net/urlid/51.1773.TB.20231211.1812.001>, (2023-12-12) [2024-09-14].
  - 13 邹振涛, 李泽平. 改进 YOLOv7 的航拍图像目标检测. 计算机工程与应用, 2024, 60(8): 173–181. [doi: [10.3778/j.issn.1002-8331.2305-0264](https://doi.org/10.3778/j.issn.1002-8331.2305-0264)]
  - 14 Sangaiah AK, Yu FN, Lin YB, *et al.* UAV T-YOLO-rice: An enhanced tiny YOLO networks for rice leaves diseases detection in paddy agronomy. IEEE Transactions on Network Science and Engineering, 2024, 11(6): 5201–5216. [doi: [10.1109/TNSE.2024.3350640](https://doi.org/10.1109/TNSE.2024.3350640)]
  - 15 Jia YJ, Fu K, Lan H, *et al.* Maize tassel detection with CA-YOLO for UAV images in complex field environments. Computers and Electronics in Agriculture, 2024, 217: 108562. [doi: [10.1016/j.compag.2023.108562](https://doi.org/10.1016/j.compag.2023.108562)]
  - 16 Tian ZJ, Liu HS, Wu JQ, *et al.* PrFu-YOLO: A lightweight network model for UAV-assisted real-time vehicle detection toward an IoT underlayer. IEEE Internet of Things Journal, 2024, 11(23): 37536–37549. [doi: [10.1109/JIOT.2024.3435130](https://doi.org/10.1109/JIOT.2024.3435130)]
  - 17 Jiang LJ, Yuan BX, Du JW, *et al.* MFFSODNet: Multiscale feature fusion small object detection network for UAV aerial images. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 5015214.
  - 18 Wei HR, Liu X, Xu SC, *et al.* DWRSeg: Rethinking efficient acquisition of multi-scale contextual information for real-time semantic segmentation. arXiv:2212.01173, 2022.
  - 19 Ding XH, Zhang YY, Ge YX, *et al.* UniRepLKNNet: A universal perception large-kernel ConvNet for audio, video, point cloud, time-series and image recognition. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 5513–5524.
  - 20 Wang C, He W, Nie Y, *et al.* Gold-YOLO: Efficient object detector via gather-and-distribute mechanism. arXiv:2309.11331, 2023.
  - 21 Zhang H, Xu C, Zhang SJ. Inner-IoU: More effective intersection over union loss with auxiliary bounding box. arXiv:2311.02877, 2023.
  - 22 Zhang H, Zhang SJ. Shape-IoU: More accurate metric considering bounding box shape and scale. arXiv:2312.17663, 2023.
  - 23 Du DW, Zhu PF, Wen LY, *et al.* VisDrone-DET2019: The vision meets drone object detection in image challenge results. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshops. Seoul: IEEE, 2019. 213–226.
  - 24 He ZQ, Cao LJ. SOD-YOLO: Small object detection network for UAV aerial images. IEEE Transactions on Electrical and Electronic Engineering, 2024.
  - 25 Zhang YZ, Wu CY, Zhang T, *et al.* Self-attention guidance and multiscale feature fusion-based UAV image object detection. IEEE Geoscience and Remote Sensing Letters, 2023, 20: 6004305.
  - 26 Dong DH, Li JZ, Liu HY, *et al.* EA-YOLO: An efficient and accurate UAV image object detection algorithm. IEEE Transactions on Electrical and Electronic Engineering, 2024.
  - 27 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475.
  - 28 Wang A, Chen H, Liu LH, *et al.* YOLOv10: Real-time end-to-end object detection. arXiv:2405.14458, 2024.
  - 29 Guo GQ, Chen PF, Yu XH, *et al.* Save the tiny, save the all: Hierarchical activation network for tiny object detection. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(1): 221–234. [doi: [10.1109/TCSVT.2023.3284161](https://doi.org/10.1109/TCSVT.2023.3284161)]
  - 30 胡清翔, 饶文碧, 熊盛武. 面向无人机遥感场景的轻量级小目标检测算法. 计算机工程, 2023, 49(12): 169–177.

(校对责编: 王欣欣)