Computer Science and Application 计算机科学与应用, 2026, 16(2), 522-531
Published Online February 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.162081

Hans汉斯

基于改进YOLOv11s的无人机航拍图像小目标检测模型

李云龙

江西理工大学理学院，江西 赣州

收稿日期：2026年1月21日；录用日期：2026年2月20日；发布日期：2026年2月26日

摘　要

无人机在巡查、农业监测等工程应用中具有重要的工程价值，但是无人机航拍图像目标尺寸小，分布密集、数量多，实际检测工作难度大，因此提高无人机航拍图像的小目标检测性能是当前重要而迫切的技术问题。本文提出了一种基于改进YOLOv11s的小目标检测模型，命名为RB-YOLOv11s。首先，设计了重参数化幽灵跨阶段高效聚合网络(RepGhostCSPELAN Net，简称RGNet)，该网络能够增强模型的表征能力，并且整合多层级特征，降低模型参数量和计算量。此外，以BiFPN-GLSA网络替换原有的路径聚合网络(PANet)，使得主干网络和颈部网络产生的特征层相互融合，增强模型对全局及局部空间信息的感知能力。在VisDrone2019无人机航拍图像数据集上的实验表明，RB-YOLOv11s模型展现出卓越性能，与原模型YOLOv11s相比，RB-YOLOv11s模型的计算量虽然小幅增加，但参数量下降了25.5%，检测精度提高了1.9%。本文模型有效解决了无人机航拍图像中小目标检测精度低的问题。

关键词

无人机，YOLOv11s，小目标检测，检测精度

A Small Object Detection Model for UAV Aerial Images Based on an Improved YOLOv11s Approach

Yunlong Li

School of Science, Jiangxi University of Science and Technology, Ganzhou Jiangxi

Received: January 21, 2026; accepted: February 20, 2026; published: February 26, 2026

Abstract

Unmanned aerial vehicles (UAVs) hold significant engineering value in applications such as patrols

文章引用: 李云龙. 基于改进 YOLOv11s 的无人机航拍图像小目标检测模型[J]. 计算机科学与应用, 2026, 16(2): 522-531. DOI: 10.12677/csa.2026.162081

and agricultural monitoring. However, UAV aerial images feature small target dimensions, dense distributions, and large quantities, making actual detection challenging. Therefore, improving small object detection performance in UAV aerial images is a critical and urgent technical issue. This paper proposes a small object detection model based on an improved YOLOv11s architecture, named RB-YOLOv11s. First, we designed the Reparameterized Ghost Cross-Stage Efficient Polymerization Network (RepGhostCSPELAN Net, abbreviated as RGNet). This network enhances the model's representational capacity while integrating multi-level features, thereby reducing both the number of model parameters and computational requirements. Additionally, the original Path Aggregation Network (PANet) is replaced with the BiFPN-GLSA network. This enables the fusion of feature layers generated by the backbone network and the neck network, enhancing the model's perception of both global and local spatial information. Experiments on the VisDrone2019 UAV aerial image dataset demonstrate that the RB-YOLOv11s model exhibits outstanding performance. Compared to the original YOLOv11s model, RB-YOLOv11s exhibits a slight increase in computational complexity but achieves a 25.5% reduction in parameters while improving detection accuracy by 1.9%. This model effectively addresses the issue of low detection accuracy for small objects in aerial drone imagery.

Keywords

UAV, YOLOv11s, Small Object Detection, Detection Accuracy

1. 引言

基于计算机视觉的无人机航拍图像目标检测主要用于识别与定位检测画面中的车辆、行人等目标[1]，已经应用到城市安防[2]、交通监控[3]、农业观察[4]、灾情监测[5]等多个场景中，但由于其目标尺寸较小，背景干扰较多且环境条件复杂，提高了特征提取的难度，对检测精度的提升带来了严峻挑战。

无人机航拍图像的目标检测方法，从特征的生成方式上可分为两类，一类是依据人工特征设计的传统模型，另一类是基于深度学习的模型。以方向梯度直方图特征和支持向量机分类器(HOG + SVM)为代表的传统模型在无人机高空、多角度拍摄的情况下，特征表示能力很差，容易受复杂背景影响，易在检测任务中出现漏检、误检的情况[6]。

而深度学习方法直接从数据中学习特征表示，能够自动提取比传统方法更具表达力的特征[7]。根据检测阶段数量的不同，深度学习模型主要分为单阶段模型与两阶段模型。

两阶段检测模型的过程为：先生成大量候选区域，然后提取候选区域进行特征提取，利用卷积神经网络进行分类；最后通过处理方式实现目标边界框的定位，该类模型包括 R-CNN [8]、Faster R-CNN [9]、Masked R-CNN [10]等。

两阶段模型在检测精度和定位精度上已经达到较高的程度，但还是无法满足实际应用的实时性要求。

YOLO 系列模型近年来得到了广泛的关注，YOLO 系列模型设计的目的就是将分类和定位任务融合在一起，模型基于回归思想解决检测问题，可以通过一次前向传播就实现输入图像到输出边界框坐标及对应类别的全过程。

虽然第一代 YOLOv1 模型[11]检测速度快于 SSD 模型，但对于密集排列的目标和小的目标群难以检测。

YOLOv2 [12]对检测框架进行了改进，添加针对小目标的锚框机制，以实现对小尺寸目标的识别。

YOLOv3 [13]的框架采用 YOLOv2 的框架，主要改进是采用金字塔网络，将多层次的特征进行融合，大大提升了模型的小尺度目标检测能力。YOLOv4 [14]延续了 YOLOv3 检测头的设计思路并进行了架构改进。它使得模型在保持小目标精度的同时，提高了对不同尺寸目标的识别能力。YOLOv5 [15]技术进一步优化，使用 GIoU 作为边界框回归损失函数，并加入 Adam 优化器等改进方式，使模型在密集遮挡的场景下检测效率更高；YOLOv6 [16]和 YOLOv7 [17]在此以后，主要针对模型层面的改进进行研究，如网络结构和训练方法等。

YOLOv8 引入了支持更广泛任务(如分割和追踪)的功能，并采用了无锚点检测机制，显著增强了其在各类数据集上实现泛化的能力。

相较于之前 YOLO 系列模型，YOLOv11 是演进路线的最新一代[18]，使用了一些新的架构增强方法，包括注意力、提取层和无锚框，以适应复杂场景下的小目标、遮挡目标和高速运动车辆等识别任务，同时保持模型的高效性。

研究人员基于目前现有的模型检测精度不足的问题进行改进，提出一系列改进措施。

Pham 等人[19]在 YOLOv3 的基础上提出一种改进的目标检测框架，可以对融合层中的不同特征进行检测，提高模型检测精度。Gu 等人[20]在 YOLOv5 架构中使用 SPD 卷积模块替换原有结构中的标准卷积与池化层，这一改进有效缓解了特征丢失问题，使模型能够更好地保留输入图像的细节信息，同时增强了网络的特征表征能力。Bangbang Chen 等人[21]在 YOLOv8 中引入 BiFPN-GLSA 网络替代颈部网络中的路径聚合网络(PANet)，优化主干网络与颈部网络特征层的融合，增强模型捕捉全局与局部空间特征的能力。

尽管现有的改进模型在目标检测的精度都有所提高，但是无人机航拍图像的小目标由于其像素占比低、特征表达差，无法被常规模型所识别，而最新的 YOLOv11 在保持较高检测速度与精度的同时，具备了优秀的特征提取与多尺度预测能力。因此，本文以最新的检测模型 YOLOv11s 为基线模型，提出了改进 YOLOv11s 的小目标检测模型 RB-YOLOv11s，以有效提高小目标的检测性能。

2. 模型介绍

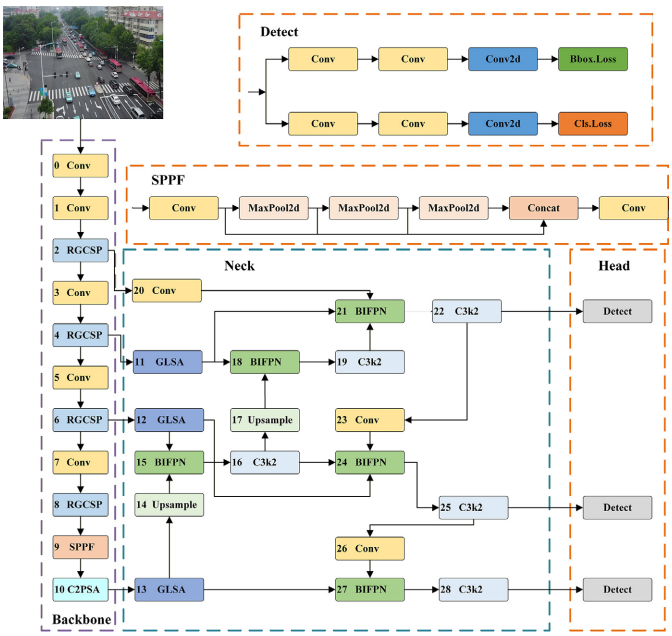


Figure 1. RB-YOLOv11s network architecture diagram
图 1. RB-YOLOv11s 网络结构图

如图 1 所示，为本文提出的 RB-YOLOv11s 模型的网络结构。本文的创新为设计了重参数化幽灵跨

阶段高效聚合网络(RepGhostCSPELAN Net，简称 RGNet)，该网络能够增强特征提取和梯度流通的能力，并且采用廉价的操作生成一部分冗余特征图，以此来降低计算量和参数量。此外，本文在颈部结构设计中引入 BiFPN-GLSA 网络，以替代原有的路径聚合网络(PANet)，旨在优化主干与颈部网络特征层的融合过程，进而增强模型捕获全局与局部空间信息的能力。

2.1. 重参数化幽灵跨阶段高效聚合网络 RGNet

本文设计了重参数化幽灵跨阶段高效聚合网络(RGNet)，RGNet 网络由 RepGhostCSPELAN、GhostNCSP 以及 GhostNBottleneck 模块组成，其结构如图 2 所示。这种网络的设计能够减小参数规模、提高计算效率以及强化特征表示能力，从而可以在资源有限的设备上实现高效部署，同时保持高检测能力，以达到模型轻量化和性能表现的平衡。

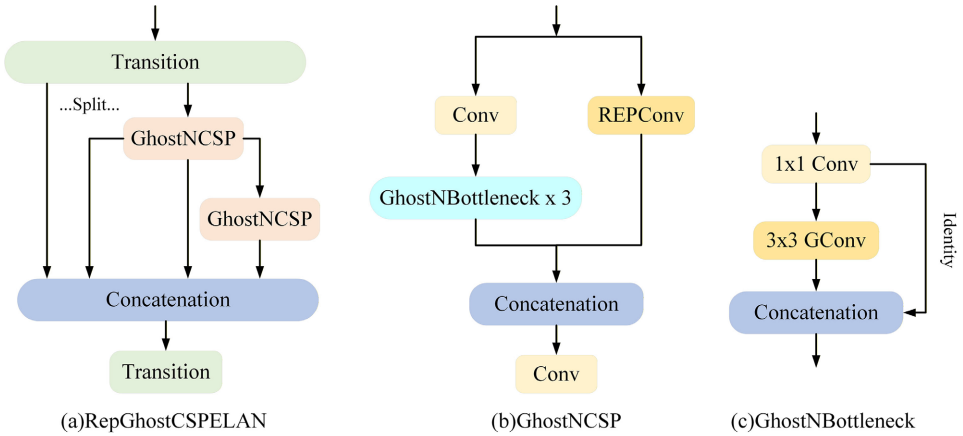


Figure 2. Block diagram of RepGhostCSPELAN, GhostNCSP, and GhostNBottleneck modules
图 2. RepGhostCSPELAN、GhostNCSP 以及 GhostNBottleneck 模块结构图

如图 2(a)所示，RepGhostCSPELAN 通过集成多层级卷积和 GhostNCSP 的输出，可以增强对上下文信息和细节特征的提取能力，改善模型的特征表示能力，从而更好地获取多尺度目标信息，提升模型鲁棒性，最终提高无人机航拍图像中小目标的检测精度。

如图 2(b)所示，GhostNCSP 通过双分支结构处理输入特征图，以增强多尺度特征的捕获能力。每个分支在减少参数量的同时，有效降低了整体模型的计算复杂度。由于其中的一个 GhostNBottleneck 分支未采用瓶颈连接，为补偿由此可能带来的性能损失，另一分支引入了重参数化卷积。该设计在训练阶段通过多路径梯度反馈促进更丰富的特征学习；在推理时则可将多个卷积层融合为单次运算，从而显著提升计算效率。最终，两个分支提取的差异化特征进行融合，使模型能够整合更为丰富的语义信息，进而提升对小目标的检测性能。

如图 2(c)所示，GhostNBottleneck 模块通过组卷积和线性变换构建轻量化特征生成方法，能够快速创建更多的特征图，进一步展示更深层的特征，从而增强模型表征能力和最终检测效果，还可以降低计算开销，提高推理效率，更加适合模型实际使用。

2.2. BIFPN-GLSA 网络

在 YOLOv11 的颈部设计中，基础架构采用了路径聚合网络(PANet) [22]。相比特征金字塔网络(FPN) [23]等传统设计，PANet 通过增加自底向上的传播路径构建了双向融合机制。这一改进既强化了深层特征的语义表征，又有效维护了浅层特征的细节与定位信息，从而减轻了 FPN 因单向信息流所造成的特征衰减问题。

然而，当处理背景复杂、目标尺寸较小的无人机航拍图像时，PANet 对于细微特征的提取与增强能力仍然有限。这使其难以有效提升小目标特征的显著性，从而可能对复杂环境中的检测精度造成影响。

为提升特征融合效果，本研究采用 BiFPN-GLSA 网络对 YOLOv11s 的颈部结构进行重构。该设计通过融合骨干网络与颈部输出的多层级特征，有效增强了模型捕捉全局与局部空间特征的能力。在 BiFPN 结构中，其第 10 至 12 层的卷积块由全局－局部空间聚合模块(Global-to-local spatial aggregation，简称 GLSA) [24]所取代，如图 3 所示，该模块同时集成了全局空间注意力(GSA)与局部空间注意力(LSA)组件，从而兼顾了非局部与局部的空间建模能力。

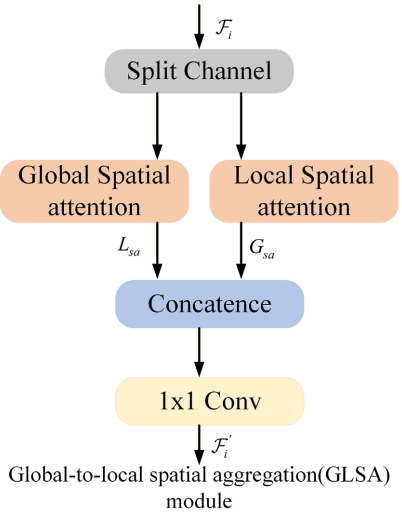


Figure 3. GLSA module structure diagram
图 3. GLSA 模块结构图

具体而言，64 通道的特征映射 $\left\{F\big|_i\, i\in(2,3,4)\right\}$ 被划分为两个子集 $F_1^i, F_2^i\,(i\in(2,3,4))$，并分别输入至 GSA 模块与 LSA 模块。这两个模块的输出结果随后通过 1 × 1 卷积进行处理，最终在通道维度上进行拼接融合。此过程由公式(1)与公式(2)描述。

$$\mathcal{F}_i^1, \mathcal{F}_i^2 = \text{Split}\left(\mathcal{F}_\text{i}\right) \tag{1}$$

$$\mathcal{F}_i' = C_{1\times1}\left(\text{Concat}\left(G_{sa}\left(\mathcal{F}_i^1\right), L_{sa}\left(\mathcal{F}_i^2\right)\right)\right) \tag{2}$$

其中，G_{sa} 表示全局空间注意力，L_{sa} 表示局部空间注意力，而 $\mathcal{F}_i' \in \mathbb{R}^{\frac{H}{8}\times\frac{W}{8}\times32}$ 为输出的特征。

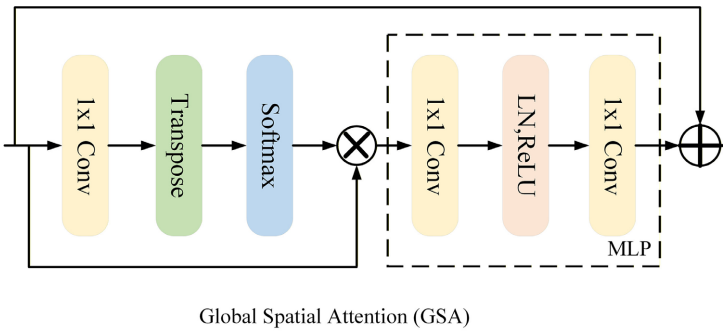
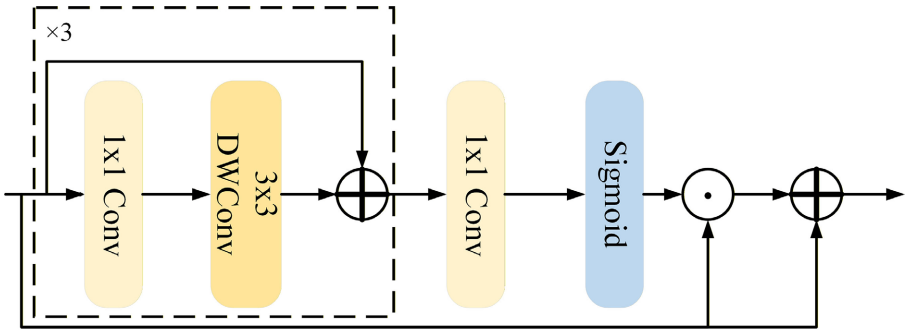


Figure 4. GSA module structure diagram
图 4. GSA 模块结构图

如图 4 所示，GSA 模块通过关注像素之间的长距离依赖关系来强化特征表征，并与局部空间注意力机制相互补充。该模块基于输入 $\mathcal{F}_i'$ 生成全局空间注意力图 $G_{sa} \in \mathbb{R}^{\frac{H}{8}\times\frac{W}{8}\times 32}$，其具体计算过程参见公式(3)与公式(4)。



Local Spatial ATtention(LSA)

Figure 5. LSA module structure diagram
图 5. LSA 模块结构图

$$\mathrm{Att}_G\left(F_i^1\right)=\mathrm{Softmax}\left(\mathrm{Transpose}\left(C_{1\times1}\left(F_i^1\right)\right)\right) \tag{3}$$

$$G_{sa}\left(F_i^1\right)=\mathrm{MLP}\left(\mathrm{Att}_G\left(F_i^1\right)\otimes F_i^1\right)+F_i^1 \tag{4}$$

在公式中，$\mathrm{Att}_G(\cdot)$ 代表注意力计算操作，$C_{1\times1}$ 表示 1×1 卷积，符号 $\otimes$ 指代矩阵乘法。

$\mathrm{MLP}(\cdot)$ 采用两层全连接结构，其中包含 ReLU 非线性激活函数及归一化处理。第一层将输入映射至高维空间，其维度扩展比为 2；第二层则把特征投影回原始尺寸，从而实现维度恢复。

$$Att_L\left(F_1^2\right)=\sigma\left(C_{1\times1}\left(F_c\left(F_i^2\right)+F_i^2\right)\right) \tag{5}$$

$$L_{sa}=\mathrm{Att}_L\left(F_i^2\right)\odot F_i^2+F_i^2 \tag{6}$$

如图 5 所示，局部空间注意力(LSA)模块能够从输入特征图中，沿空间维度高效聚焦并提取关键的局部特征信息，这对于小目标检测尤为重要。如图 5 所示，该模块以局部空间注意力响应作为输入，其具体计算流程由公式(5)与公式(6)详细定义。其中，$F_c(\cdot)$ 的结构由三个级联的 1×1 卷积层与一个 3×3 的深度卷积层共同构成。在 F_c 中，通道数被调整为 32。$\mathrm{Att}_L(\cdot)$ 指局部注意力操作，$\sigma(\cdot)$ 是 Sigmoid 函数，符号 $\odot$ 表示逐点乘法。该结构能够在参数受限的条件下，高效实现局部空间信息的整合。

3. 实验设计

3.1. 数据集与评估指标

本文选用公开数据集 VisDrone2019 [25]进行模型验证，其采集于多样化的真实道路场景，包含行人、车辆等共 10 类无人机航拍检测目标。该数据集已划分为训练集(6471 张)、验证集(548 张)及测试集(1610 张)，适用于无人机视角下的性能评估。实验在统一环境中进行，通过对比改进前后的模型表现，以精确率、召回率、mAP、参数量、计算量以及 FPS 等指标综合评价性能。

3.2. 实验环境

实验在 Ubuntu 20.04 系统下采用 Python 3.8 与 Pytorch 2.0.0 环境，设置 batch size 为 8，训练 300

epoch，输入图像尺寸为 640 × 640，具体硬件与参数配置见表 1。

Table 1. Training parameter settings
表 1. 训练参数设置

Types	Configuration	Type	Value
GPU	RTX 4090	Learning rate	1×10^{-4}
CPU	Intel (R) 5418Y	momentum	0.9
CUDA	11.8	optimizer	AdamW
CuDNN	8.7.0	batch	8

4. 结果分析

4.1. 消融实验

将设计重参数化幽灵跨阶段高效聚合网络(RGNet)记为改进方案 A，BIFPN-GLSA 网络替换原 YOLOv11s 模型的颈部网络 PANet 记为改进方案 B，下表 2 中 YOLOv11s + A 为只加入改进方案 A，YOLOv11s + B 为只加入改进方案 B，YOLOv11s + A + B 为将改进方案 A 和 B 都加入，也就是本文提出的模型 RB-YOLOv11s。

Table 2. Various improved ablation experiments
表 2. 各个改进的消融实验

Model	Precise (%)	Recall (%)	mAP50 (%)	mAP50:95 (%)	P (M)	G (G)	FPS
YOLOv11s	46.3	33.9	32.7	18.7	9.4	21.3	556
YOLOv11s + A	45.7	35.3	33.7	19.1	8.1	20.9	526
YOLOv11s + B	46.1	35.3	33.7	19.5	7.7	23.3	455
YOLOv11s + A + B	46.8	35.7	34.6	20.0	7.1	24.3	385

如表 2 所示，加入改进方案 A 后，YOLOv11s+A 模型的 mAP50 和 mAP50:95 分别为 33.7%和 19.1%，相比 YOLOv11s，分别提升了 1%和 0.4%，并且参数量和计算量分别下降了 1.3 M 和 0.4 G，说明所设计的重参数化幽灵跨阶段高效聚合网络(RGNet)有效增强了小目标的特征表征能力，在实现模型轻量化的同时，显著提升了小目标检测精度。加入改进方案 B 后，YOLOv11s + B 模型的 mAP50 和 mAP50:95 分别为 33.7%和 19.5%，相比 YOLOv11s，分别提升了 1%和 0.8%，并且参数量下降了 1.7 M，说明 BIFPN-GLSA 网络通过增强多尺度特征融合能力，有效改善了小目标的检测性能，并在提升精度的同时实现了模型结构的精简。将改进方案 A 和 B 一起加入到 YOLOv11s 后，YOLOv11s + A + B 模型的 mAP50 和 mAP50:95 分别为 34.6%和 20.0%，相比 YOLOv11s，分别提升了 1.9%和 1.3%，参数量下降了 2.3 M，说明重参数化幽灵跨阶段高效聚合网络(RGNet)与 BIFPN-GLSA 网络通过协同增强小目标的特征表征能力与多尺度特征融合，有效提升了检测精度，并实现了模型参数的显著精简。

4.2. 对比实验

如表 3 所示，RB-YOLOv11s 在 VisDrone2019 测试集上的 mAP50 达到 34.6%，其检测精度优于原 YOLOv11s 模型及 YOLO 与 DETR 系列的其他对比模型。该模型的精确率为 46.8%，召回率为 35.7%，分别高于原模型的 46.3%和 33.9%，表明其在提升小目标检测准确性的同时，有效降低了误检与漏检。在

模型复杂度方面，RB-YOLOv11s 的参数量和计算量分别为 7.1 M 和 24.3 G，虽然计算量较原模型略有增加，但参数量仍低于同级别的 YOLOv8s。在实时性方面，RB-YOLOv11s 的 FPS 为 285，虽略低于原模型，但仍可满足实时检测要求。

Table 3. Comparative experiments of different VisDrone models
表 3. VisDrone 不同模型的对比实验

Model	Precise (%)	Recall (%)	mAP50 (%)	mAP50:95 (%)	P (M)	G (G)	FPS
YOLOv5s	38.7	31.0	27.6	14.4	7.0	15.8	204
YOLOv8s	44.8	35.0	32.8	18.7	11.1	28.5	625
YOLOv9s	49.1	35.5	34.9	20.6	9.6	38.8	278
YOLOv10s	45.4	34.6	33.0	18.6	7.2	21.4	526
YOLOv11s	**46.3**	**33.9**	**32.7**	**18.7**	**9.4**	**21.3**	**556**
TOOD	——	——	33.6	20.4	32.03	199	46
Faster-RCNN	——	——	32.6	21.7	41.4	208	51
Atss_r50_fpn	——	——	33.8	20.4	38.91	110	47
Retinanet_r50_fpn	——	——	27.6	16.4	36.517	210	58
Cascade-Mask-Rcnn	——	——	32.6	19.7	69.29	236	49
DETR	——	——	33.2	17.4	60	187	21
Deformable-DETR	——	——	30.7	16.9	40.101	193	30
RB-YOLOv11s	**46.8**	**35.7**	**34.6**	**20.0**	**7.1**	**24.3**	**385**

4.3. 检测结果可视化对比分析

Figure 6. Detection performance comparison between YOLOv11s and RB-YOLOv11s models
图 6. YOLOv11s 与 RB-YOLOv11s 模型的检测效果对比图

如图 6 所示，在第一列图像中，YOLOv11s 错误地将左下角的背景目标识别为卡车，而 RB-YOLOv11s 则未产生此误检。在第二列图像中，YOLOv11s 漏检了右上角的卡车，RB-YOLOv11s 则能准确识别。上述对比结果表明，RB-YOLOv11s 模型有效增强了对小目标的特征辨别能力，显著减少了复杂背景下的误检与漏检。

5. 结论

本文提出的 RB-YOLOv11s 模型是一种基于 YOLOv11s 的无人机航拍图像小目标检测模型，可为农业病虫害监测、交通流量分析与灾害应急评估等应用提供关键的技术支持。

在 VisDrone2019 数据集上的实验表明，RB-YOLOv11s 模型通过重参数化幽灵跨阶段高效聚合网络 RGNet 和 BiFPN-GLSA 网络增强了小目标特征表征与多尺度融合，相比原模型 YOLOv11s，RB-YOLOv11s 模型的 mAP50、mAP50:95 分别提升了 1.9%与 1.3%。为适应无人机平台的部署要求，后续将通过轻量化技术降低模型的参数量和计算量。

参考文献

[1] Minaeian, S., Liu, J. and Son, Y.J. (2015) Vision-Based Target Detection and Localization via a Team of Cooperative UAV and UGVs. *IEEE Transactions on Systems, Man, and Cybernetics*: *Systems*, **46**, 1005-1016. https://doi.org/10.1109/TSMC.2015.2491878

[2] Biehle, T. (2025) Urban Dimension of U-Space: Local Planning Considerations for Drone Integration. *Drones*, **9**, Article 744. https://doi.org/10.3390/drones9110744

[3] Bisio, I., Garibotto, C., Haleem, H., Lavagetto, F. and Sciarrone, A. (2022) A Systematic Review of Drone Based Road Traffic Monitoring System. *IEEE Access*, **10**, 101537-101555. https://doi.org/10.1109/access.2022.3207282

[4] Hama, A., Tanaka, K., Chen, B. and Kondoh, A. (2021) Examination of Appropriate Observation Time and Correction of Vegetation Index for Drone-Based Crop Monitoring. *Journal of Agricultural Meteorology*, **77**, 200-209. https://doi.org/10.2480/agrmet.d-20-00047

[5] Kucharczyk, M. and Hugenholtz, C.H. (2021) Remote Sensing of Natural Hazard-Related Disasters with Small Drones: Global Trends, Biases, and Research Opportunities. *Remote Sensing of Environment*, **264**, Article ID: 112577. https://doi.org/10.1016/j.rse.2021.112577

[6] Cazzato, D., Cimarelli, C., Sanchez-Lopez, J.L., Voos, H. and Leo, M. (2020) A Survey of Computer Vision Methods for 2D Object Detection from Unmanned Aerial Vehicles. *Journal of Imaging*, **6**, Article 78. https://doi.org/10.3390/jimaging6080078

[7] Ahmed, S.F., Alam, M.S.B., Hassan, M., Rozbu, M.R., Ishtiak, T., Rafa, N., *et al.* (2023) Deep Learning Modelling Techniques: Current Progress, Applications, Advantages, and Challenges. *Artificial Intelligence Review*, **56**, 13521-13617. https://doi.org/10.1007/s10462-023-10466-8

[8] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. https://doi.org/10.1109/cvpr.2014.81

[9] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/tpami.2016.2577031

[10] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2980-2988. https://doi.org/10.1109/iccv.2017.322

[11] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/cvpr.2016.91

[12] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 6517-6525. https://doi.org/10.1109/cvpr.2017.690

[13] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. arXiv: 1804.02767.

[14] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv: 2004.10934.

[15] Zhang, Y., Guo, Z., Wu, J., Tian, Y., Tang, H. and Guo, X. (2022) Real-Time Vehicle Detection Based on Improved

YOLO V5. *Sustainability*, **14**, Article 12274. https://doi.org/10.3390/su141912274

[16] Li, C., Li, L., Jiang, H., *et al.* (2022) YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. arXiv: 2209.02976.

[17] Wang, C., Bochkovskiy, A. and Liao, H.M. (2023) YOLOv7: Trainable Bag-Of-Freebies Sets New State-Of-The-Art for Real-Time Object Detectors. 2023 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, 17-24 June 2023, 7464-7475. https://doi.org/10.1109/cvpr52729.2023.00721

[18] Khanam, R. and Hussain, M. (2024) YOLOv11: An Overview of the Key Architectural Enhancements. arXiv: 2410.17725.

[19] Liu, Y., Zhang, H., Wen, F., Wang, X., Wang, L. and Cheng, X. (2025) Object Detection in Remote Sensing Images Based on Knowledge Graph and Visual Attention. 2025 *International Joint Conference on Neural Networks* (*IJCNN*), Rome, 30 June-5 July 2025, 1-8. https://doi.org/10.1109/ijcnn64981.2025.11229269

[20] Gu, Z., Zhu, K. and You, S. (2023) YOLO-SSFS: A Method Combining SPD-Conv/STDL/IM-FPN/SIoU for Outdoor Small Target Vehicle Detection. *Electronics*, **12**, Article 3744. https://doi.org/10.3390/electronics12183744

[21] Chen, B., Tan, K., Li, K., Ma, B. and Liu, X. (2025) Research on Detection and Counting Method of Green Walnut Based on YOLOv8n-RBP. *IEEE Access*, **13**, 39275-39288. https://doi.org/10.1109/access.2025.3546313

[22] Liu, S., Qi, L., Qin, H., Shi, J. and Jia, J. (2018) Path Aggregation Network for Instance Segmentation. 2018 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8759-8768. https://doi.org/10.1109/cvpr.2018.00913

[23] Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 936-944. https://doi.org/10.1109/cvpr.2017.106

[24] Tang, F., Xu, Z., Huang, Q., Wang, J., Hou, X., Su, J., *et al.* (2023) Duat: Dual-Aggregation Transformer Network for Medical Image Segmentation. In: Liu, Q., *et al.*, Eds., *Pattern Recognition and Computer Vision*, Springer, 343-356. https://doi.org/10.1007/978-981-99-8469-5_27

[25] Zhou, Z., Yu, X. and Wang, X. (2024) Object Detection in Drone Video Based on Recurrent Motion Attention. *Pattern Recognition Letters*, **183**, 56-63. https://doi.org/10.1016/j.patrec.2024.04.029