计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目：　改进 YOLOv5s 的无人机视角下小目标检测算法
作者：　吴明杰，云利军，陈载清，钟天泽
网络首发日期：　2023-09-21
引用格式：　吴明杰，云利军，陈载清，钟天泽．改进 YOLOv5s 的无人机视角下小目标检测算法[J/OL]．计算机工程与应用．
https://link.cnki.net/urlid/11.2127.tp.20230920.1219.042

网络首发时间：2023-09-21 17:42:55
网络首发地址：https://link.cnki.net/urlid/11.2127.tp.20230920.1219.042

改进 YOLOv5s 的无人机视角下小目标检测算法

吴明杰[1,2]， 云利军[1,2*]， 陈载清[1,2]， 钟天泽[1,2]

1.云南师范大学 信息学院，昆明 650500
2.云南省教育厅计算机视觉与智能控制技术工程研究中心，昆明 650500

摘要：针对无人机飞行时与目标距离较远，被拍摄的目标大小有明显的差异且存在被物体遮挡等问题，提出一种基于 YOLOv5s 的无人机视角下小目标检测改进算法 BD-YOLO。在特征融合网络中采用双层路由注意力(Bi-level Routing Attention, BRA)，其以动态稀疏的方式过滤特征图中最不相关的特征，保留部分重要区域特征，从而提高模型特征提取的能力；由于特征图经过多次下采样后会丢失大量位置信息和特征信息，因此采用一种结合注意力机制的动态目标检测头 DyHead(Dynamic Head)，该检测头通过尺度感知、空间感知和任务感知的三者统一，以实现更强的特征表达能力；使用 Focal-EIoU 损失函数，来解决 YOLOv5s 中 CIoU Loss 计算回归结果不准确的问题，从而提高模型对小型目标的检测精度。实验结果表明，在 VisDrone2019-DET 数据集上，BD-YOLO 模型较 YOLOv5s 模型在平均精度(mAP@0.5)指标上提高了 6.2%，对比其他主流模型对于小目标的检测都有更好的效果。

关键词：无人机视角；YOLOv5s；小目标；注意力机制；损失函数
文献标志码:A **中图分类号**:TP391 doi：10.3778/j.issn.1002-8331.2307-0223

Improved YOLOv5s small object detection algorithm in UAV view

WU Mingjie[1,2], YUN Lijun[1,2*], CHEN Zaiqing[1,2], ZHONG Tianze[1,2]

1. School of Information, Yunnan Normal University, Kunming 650500, China
2. Yunnan Provincial Department of Education Computer Vision and Intelligent Control Technology Engineering Research Center, Kunming 650500, China

Abstract: Aiming at the problems such as the long distance between UAV and object in flight, the obvious difference in the size of the photographed object and the existence of object occlusion, an improved algorithm BD-YOLO based on YOLOv5s for small object detection under UAV perspective was proposed. In the feature fusion network, Bi-level Routing Attention (BRA) is used to filter the least relevant features in the feature map in a dynamic sparse way, and retain some important regional features, so as to improve the feature extraction ability of the model. Since the feature map will lose a lot of location and feature information after multiple subsampled, a Dynamic object detection Head(DyHead) combining attention mechanism is adopted. The DyHead integrates scale perception, space perception and task perception to achieve stronger feature representation capability. Focal-EIoU Loss function was used to solve the problem of inaccurate regression results of CIoU Loss calculation in YOLOv5s, so as to improve the detection accuracy of the model for small object. The experimental results show that on the VisDrone2019-DET dataset, the BD-YOLO model has increased the mean Average Precision (mAP) index by 6.2% compared with the YOLOv5s model, and has better results for small object detection than other mainstream models.

基金项目：云南省教育厅科学研究基金项目（2023Y0533）。
作者简介：吴明杰(1996－)，男，硕士研究生，研究方向为目标检测；云利军(1973－)，通信作者，男，博士，教授，CCF 会员，研究方向为物联网技术、视频图像处理，E-mail: yunlijun@ynnu.edu.en；陈载清(1978－)，男，博士，教授，研究方向为颜色与图像视觉；钟天泽(1998－)，男，硕士研究生，研究方向为计算机视觉、深度学习。

Key words：unmanned aerial vehicle perspective; YOLOv5s; small object; attention mechanism; loss function

随着基于深度学习的目标检测技术快速发展，结合无人机航拍进行目标检测的方法越来越普遍。如森林防火、楼道巡检、农业监测等。无人机在检测任务中提供了高空视角，但是高空视角的图像会带来目标尺度变化大、目标之间相互遮挡的问题，这将导致使用常规的目标检测算法在进行检测任务时出现误检漏检问题。因此研究出一个针对无人机视角下的小目标检测算法模型成为关键的研究内容之一。

近几年以来，伴随深度学习技术的快速发展，国内外的研究学者逐渐将深度学习技术应用于目标检测。韩玉洁等人[1]在 YOLO 上进行数据增强，修改激活函数，添加 CIoU，修改后模型精度有所提升，但模型内存占用过大；丁田等人[2]加入注意力以及 CIoU，加快了模型收敛的速度，增加检测准确率，但也增加了计算成本；冒国韬等人[3]对 YOLOv5s 算法引入多尺度分割注意力，以应对小目标背景复杂及特征提取困难等问题，虽然检测精度有一定提升，但是牺牲了参数量；Zhu 等人[4]在 YOLOv5 模型中引入 CBAM 注意力机制以解决航拍图像目标模糊的问题，但是改进后的模型对硬件性能要求较高，不易实现；Yang[5]等人在 YOLOv5 网络的颈部增加上采样，形成用于收集小目标特征的特征图，增强了算法的小目标检测能力，但改进后的算法提取到的小目标特征信息较少，且检测速度较慢，实时性不足以满足实际需求。

为了增强 YOLOv5s 模型对小目标特征捕获的灵活度、缓解特征图经过多次降采样后信息的丢失、提高模型检测头的表示能力、解决回归计算时误差太大的问题，本文提出了一种基于 YOLOv5s 的无人机航拍小目标改进算法 BD-YOLO。主要贡献如下：

（1）在模型的 11 层与 12 层之间加入双向路由注意力（Bi-level Routing Attention, BRA）[6]，在不增加过多参数量的同时，提升模型在特征提取时对小目标区域的关注度和精确度。

（2）将原模型的目标检测头部替换成带自注意力机制的检测头部（Dynamic Head, DyHead）[7]，提高模型检测层的表示能力，以应对特征图进行多次下采样后特征信息严重丢失的问题。

（3）使用 Focal-EIoU[8]优化 CIoU 在进行预测框回归计算时误差较大的问题，提高模型在背景复杂的图像中对小目标检测的鲁棒性。

1 相关工作

1.1 小目标检测

目前针对“小目标”的定义主要有两种：一种是绝对尺寸，尺寸小于 32×32 的目标被认为是小目标；另一种是相对尺寸，根据国际光电工程学会定义，小目标为 256×256 像素的图像中成像面积小于 80 像素的目标，即目标的尺寸小于原图的 0.12%则可被认为是小目标[9]。小目标的检测一直是目标检测中一个具有挑战性的难题，对于图像特征的深刻理解是提升小目标检测效果的前提。近年来产生了许多有用的方法来提高小目标检测的性能。

针对小目标检测的难点，Chen 等人[10]提出了 Stitcher 方法，采用损失函数作为反馈，当小目标贡献过小时，则在下一次迭代中通过图片拼贴的方式提高小目标占比以提高小目标训练效果。Kisantal 等人[11]通过复制粘贴小目标来提高在数据集中所占比例，从而提高小目标对网络的贡献，提升模型对小目标的检测效果。由于小目标在图像中的占比较小，而数据增强中简单的拼贴可以帮助小目标增加其在图像中占比，可以在一定程度上提高模型对小目标的检测效果，但是数据增强只是简单增加目标的比例，并不会提高对深层语义信息的利用。

由于小目标的尺寸较小且利用信息少，因此可以利用上下文信息的方式来增强模型的检测能力。李青援等人[12]在 SSD 模型中引入一条自深向浅的递归反向路径，通过特征增强模块将深层包含上下文信息的语义特征增强到浅层。但是，并不是所有的上下文信息都是有效的，当图像中缺少与目标关联较高的信息时，会产生冗余的信息噪声。

1.2 目标检测算法概述

目标检测从早期的传统方法到目前基于深度学习的方法，发展已经有 21 年。当前基于深度学习的目标检测算法有两种：第一种是以 Faster RCNN[13]为代表的二阶段检测算法等，该类模型首先利用算法生成预选框，再使用深度卷积网络对预选框进行检测类别的分类，二阶段检测算法获得的精度更高，但速度较慢，不能满足实时性要求较高的场合；第二种是以 YOLO[14]、YOLOv3[15]、YOLOv4[16]、YOLOv5 以及 SSD[17]为代表的单阶段检测算法，通过将检测框的定位和分类任务结合到一起，以达到快速检测出目标位

置的效果，通过适当的改进可同时具有更好的实时性与检测精度。

1.3 YOLOv5s 概述

YOLOv5 属于一种单阶段的目标检测算法，可以实现端到端目标检测，运行速度快，但是在检测精度上相较于二阶段的 RCNN 算法略低。YOLOv5 一共有 4 个版本：YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x，四个版本的区别在于网络的深度和宽度，YOLOv5s 网络结构是 YOLOv5 系列中深度最小且特征图的宽度最小的。另外三种都是在 YOLOv5s 基础上不断加深、不断加宽。YOLOv5s 的网络结构较简洁，运行速度也最快，对于小目标应用场景的考虑，本研究选择使用 YOLOv5s 模型。其网络结构如图 1 所示。

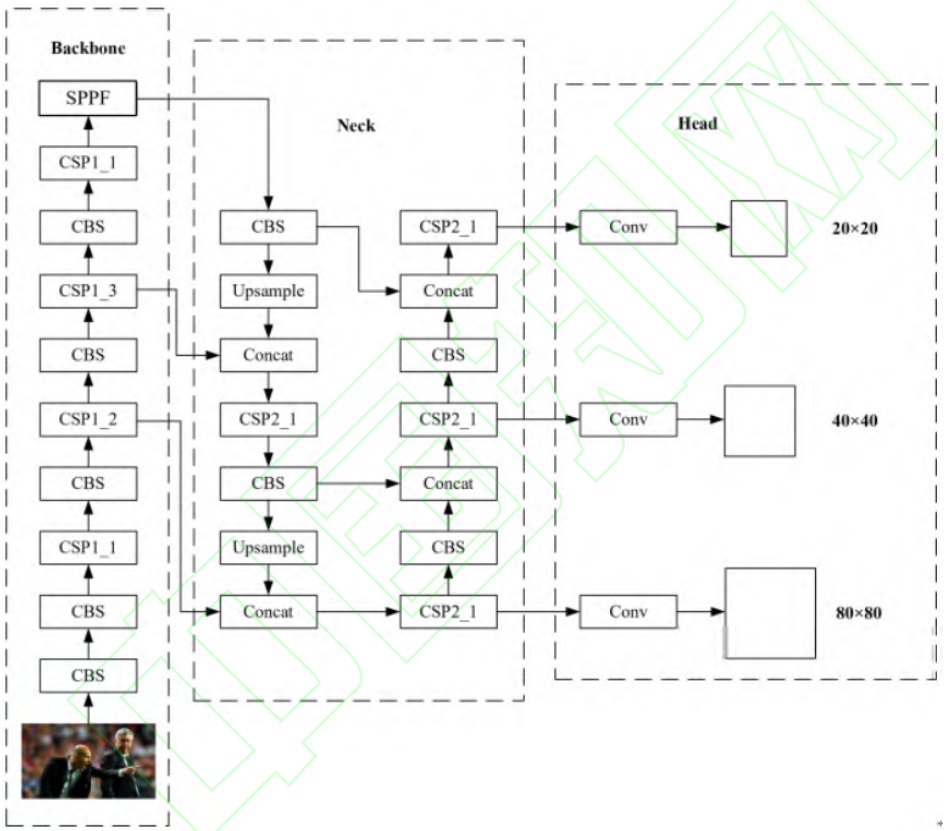


图 1 YOLOv5s 结构

Fig.1 YOLOv5s structure

YOLOv5s 模型的网络结构中，可分为输入端、Backbone 网络、Neck 网络、Head 输出层 4 个部分。

（1）输入端：输入端对输入进来的图像进行一个 Mosaic 数据增强处理，首先对几张图像进行随机缩放、剪裁、排布，然后将图像进行随机拼接，如图 2 所示。通过 Mosaic 处理后不仅可以丰富数据集，还可以提升模型训练的速度。YOLOv5s 算法会针对不同类型的数据集自适应计算出最佳的锚点框。

图 2 Mosaic 数据增强

Fig.2 Mosaic data enhancement

（2）Backbone 网络：YOLOv5s 的 Backbone 网络中由 CSP[18]、CBS 和 SPPF 模块组成，分别为特征提取模块、卷积模块和空间金字塔池化模块。CSP 结构分为两类：CSP1_X、CSP2_1，其中，Resx 模块中有两种模式，一种是卷积之后使用 shortcut 与初始输入相加后输出，另一种是卷积之后直接输出，如图 3 所示。CBS 由 Conv、BN 归一层和 Leaky relu 激活函数组成，如图 3 所示。该版本的 SPPF 模块进行多个 5×5 的 MaxPool2d 操作，不再使用原先的 1×1，5×

5，9×9，13×13 的最大池化，这样可以提高模型的推理速度。

（3）Neck 网络：Neck 由特征金字塔网络（Feature Pyramid Network, FPN）[19]和路径聚合网络（Path Aggregation Network, PAN）[20]组成。FPN 自顶向下传达强化语义特征，PAN 自底向上传达强化位置特征，FPN+PAN 的结构将不同阶段的特征图进行特征融合，提高了模型对小目标检测的精度。

（4）Head 输出层：Head 的主体为三个 Detect 检测器。当输入的图像尺寸为 640×640 时，Head 层先对 Neck 层的三个输出进行卷积操作，再在三个尺度分别为 20×20、40×40、80×80 的特征图上生成对应的预测框。

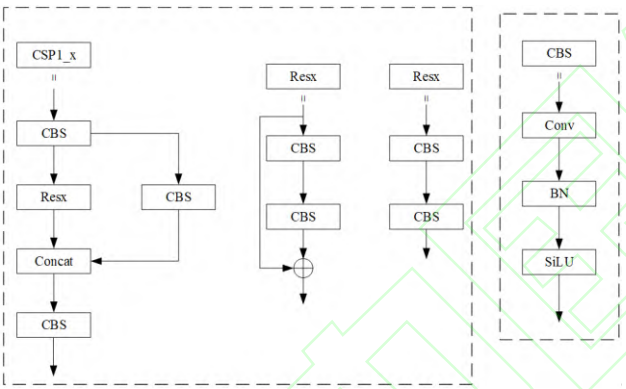


图 3 CSP 和 CBS 结构

Fig.3 Structure of CSP and CBS

目前 YOLOv5s 模型仍然在大多数目标检测任务中发挥巨大的优势，但是由于 YOLOv5s 使用过多的超参数来优化模型训练，因此模型训练的时间相对较慢。在检测小目标时，由于 YOLOv5s 特征提取网络使用了较大的感受野，可能导致小目标信息的丢失。在应用场景中，YOLOv5s 模型相对复杂的网络结构需要更高的计算资源，在结合硬件环境部署模型时，将对硬件提出更高的要求。

计算复杂度是指模型进行一次前向传播所需要的浮点运算次数，可以通过计算模型每层的计算量，然后求和得到整个模型的计算复杂度。对于 YOLOv5s 算法的复杂度，需要结合参数量、计算量、内存需求和推理速度这几个方面进行考量。YOLOv5 算法中 4 个不同版本的参数如表 1 所示。通常来说，一个模型的参数量越多模型越复杂。从表 1 可以看出，YOLOv5s 相较于其他版本是最轻量级的模型。

表 1 YOLOv5 算法不同版本的参数

Table 1 YOLOv5 algorithm different versions of parameters

模型	推理速度(ms) V100 b1	参数量(M)	FLOPs(B)
YOLOv5s	6.4	7.2	16.5
YOLOv5m	8.2	21.2	49.0
YOLOv5l	10.1	46.5	109.1
YOLOv5x	12.1	86.7	205.7

2 BD-YOLO 算法

2.1 BD-YOLO 算法结构

根据小目标检测任务的特点，提出了针对小目标检测的 BD-YOLO 模型。引入动态稀疏注意力机制 BRA 来提高小目标检测精度，使用带有注意力机制的检测头 DyHead 来提高模型的表达能力，以及采用更适合小目标检测的损失函数 Focal-EIoU Loss 来提高模型的预测精度。BD-YOLO 算法结构如图 4 所示。

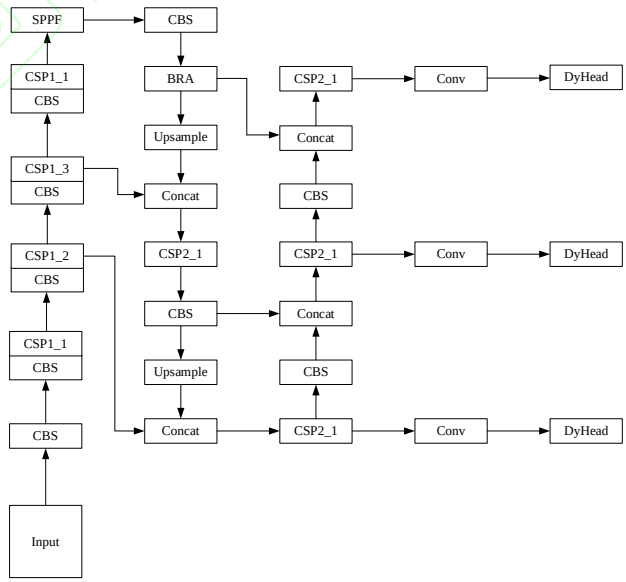


图 4 BD-YOLO 结构

Fig.4 BD-YOLO structure

2.2 Bi-level Routing Attention 注意力

目前常用的注意力机制 SE[21]、CA[22]、ECA[23]、CBAM[24]等，都是在一个全局范围获取重点关注的目标，将会导致计算复杂度较高以及消耗大量的内存。为了提高模型特征提取能力，同时在不增加过多网络计算复杂度的情况下，在模型的 11 层与 12 层之间添加一种动态稀疏注意力机制（Bi-level Routing Attention, BRA），以实现更强大的特征提取能力，更灵活的

计算分配和内容感知。BRA 注意力机制的原理如图 5 所示。BRA 先将输入进来的一张特征图 $X \in R^{H\times W\times C}$，划分为 $S\times S$ 个不同的区域，每个区域包含 $\frac{WH}{S^2}$ 个特征向量，即可将 X 变为 $X^r \in R^{S^2\times\frac{HW}{S^2}\times C}$。通过线性映射获得 $Q,K,V \in R^{S^2\times\frac{HW}{S^2}\times C}$。$Q,K,V$ 的具体表达式如公式(1)-(3)所示:

$$Q = X^r W^q \tag{1}$$

$$K = X^r W^k \tag{2}$$

$$V = X^r W^v \tag{3}$$

其中，$W^q,W^k,W^v \in R^{C\times C}$，分别属于 query、key、value 的投影权重。然后通过构造一个有向图找到不同键值对对应的参与关系，最后应用细粒度的 token-to-token 注意力操作，计算公式(4)如下:

$$O = Attention(Q,K^g,V^g) + LCE(V) \tag{4}$$

其中，K^g 和 V^g 是聚合后 key 和 value 的 tensor，函数 $LCE(\cdot)$ 使用深度卷积参数化，在 BD-YOLO 模型的设计中所使用的参数值为 5。BRA 通过稀疏性操作直接省略最不相关区域的计算，以实现计算有效分配的目的。全局注意力与 BRA 注意力实现效果对比如图 6 所示。

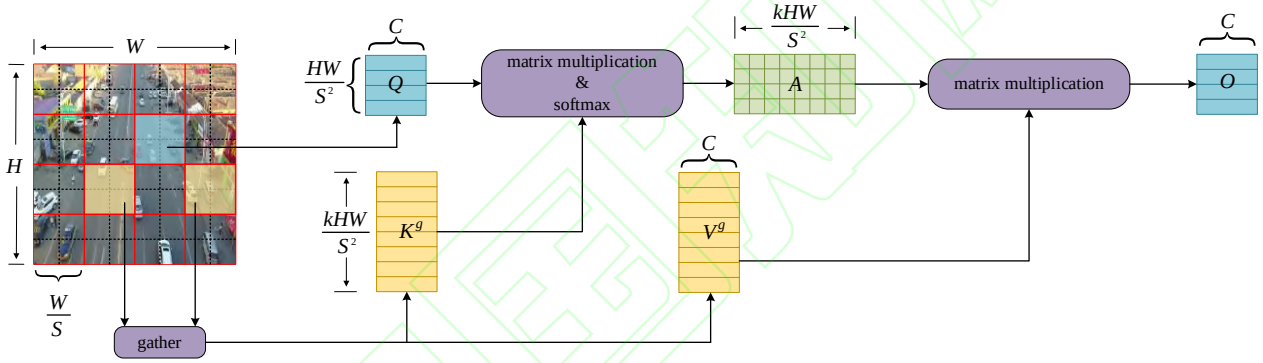
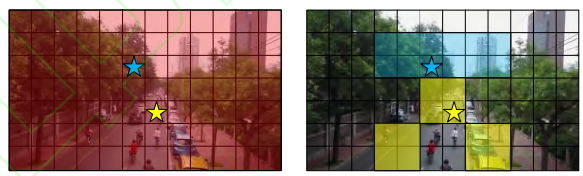


图 5 BRA 注意力原理

Fig.5 BRA attention principle



(a)全局注意力效果 (b)BRA 注意力效果

图 6 注意力实现效果对比

Fig.6 Attention achieves effect contrast

2.3 自注意力检测头

YOLOv5s 模型只有三个检测头，当该模型对小目标进行检测时，对较小的目标可能存在漏检的现象。目前，许多研究学者会通过在原模型三层检测层的基础上增加到四层。俞军等人[25]在 YOLO 模型中增加了一层专门针对于小目标的检测层，使得由更浅层的特征图融合而来的特征图具有更强大的语义信息和精确的位置信息。

在 YOLOv5s 模型中，主干网络的输出是一个三维张量，其维度为水平 × 空间 × 通道。因此，将 YOLOv5s 模型的检测头部替换为一种可同时实现尺度感知注意力、空间感知注意力和任务感知注意力统一的动态检测头 DyHead（Dynamic Head），即在特征张量的每个特定维度上添加注意力机制。在检测层上给定三维特征张量 $F \in R^{L\times S\times C}$，该注意力函数计算公式(5)如下所示:

$$W(F) = \pi_C(\pi_S(\pi_L(F)\cdot F)\cdot F)\cdot F \tag{5}$$

其中，$\pi_L(\cdot)$、$\pi_S(\cdot)$、$\pi_C(\cdot)$ 分别是应用在维度 L、S、C 上的三个不同的注意力函数，这三种注意力顺序应用于检测头部，可以多次叠加使用。在 BD-YOLO 模型的设计中，使用了四组 $\pi_L(\cdot)$、$\pi_S(\cdot)$ 和 $\pi_C(\cdot)$ 模块依

次叠加，让检测头具备更强的表示能力，从而提升算法对小目标的检测效果。DyHead 结构如图 7 所示。

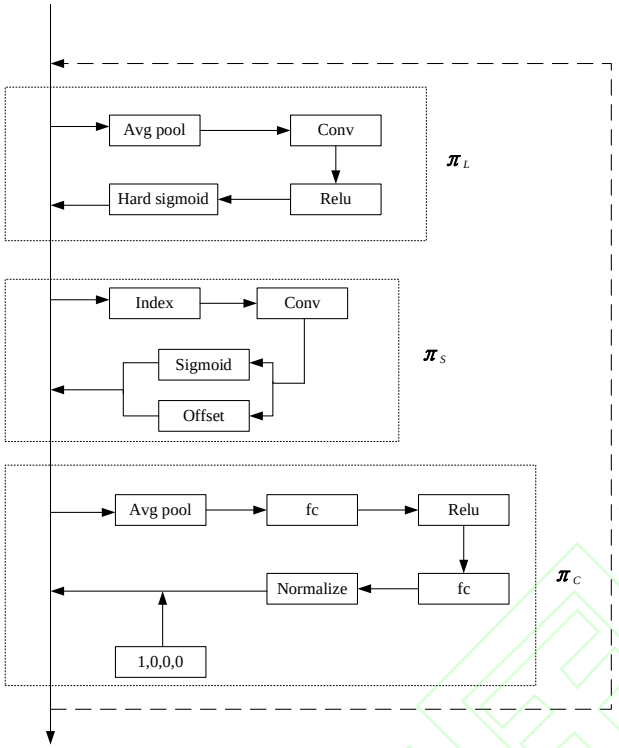


图 7 DyHead 结构

Fig.7 DyHead structure

2.4 损失函数的改进

YOLOv5s 模型使用 CIoU Loss 计算矩形框损失，其主要由三部分组成：预测矩形框位置的损失 (L_{bbox})、置信度的损失 (L_{obj})、分类损失 (L_{cls})，CIoU Loss 具体的计算公式如公式(6)所示：

$$L = L_{obj} + L_{cls} + L_{bbox} \tag{6}$$

CIoU Loss 将重叠的面积、中心距离和矩形框的宽高比同时加入计算，提高了模型训练的稳定性和收敛速度。但是，CIoU Loss 并未真正表示出矩形框的高宽与其置信度的真实差异，这将会导致回归预测的结果不够精准。

针对 CIoU Loss 的问题，将预测框的宽高分别考虑，使用了 Focal-EIoU Loss，该损失函数由 Focal Loss 和 EIoU Loss 组合而成。Focal Loss 将预测框的宽高拆分，分别与最小外界框的宽高作差值运算。EIoU Loss 通过减小预测框和真实框宽高上的差异，使得收敛速度更快且有更好的定位结果，其具体的计算公式如式(7)所示：

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} \tag{7}$$

其中，L_{IoU}、L_{dis}、L_{asp} 分别为 IoU 损失、距离损失、高宽损失。在一张样本图片中，回归误差小的锚框数量远远少于误差大的锚框数量，质量较差的锚框会产生较大的梯度，这将会直接影响模型的训练效果。因此，在 EIoU Loss 基础上添加 Focal Loss，把高质量的锚框与低质量的锚框分开，计算公式如公式(8)所示：

$$L_{Focal-EIoU} = IoU^{\gamma} L_{EIoU} \tag{8}$$

其中 γ 是用于控制曲线弧度的超参数。

损失函数 Focal-EIoU Loss 弱化了易回归样本的权重，使模型更专注于预测框与真实框重叠低的样本，从而实现提高回归精度的效果。

3 实验结果与分析

3.1 数据集

本文所使用的是公开数据集 VisDrone2019[26]，该数据集一共包含 8599 张由无人机位于高空拍摄的静态图像，其中 6471 张用于训练，548 张用于验证，1580 张用于测试。图像类别包括行人、人、自行车、汽车、面包车、卡车、三轮车、遮阳篷-三轮车、公共汽车和摩托车，一共 260 万个标注信息。其中训练集实例数量分布如图 8 所示。

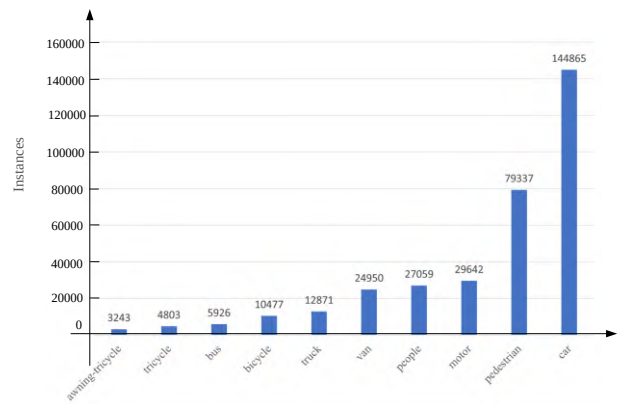


图 8 训练集实例数量分布

Fig.8 The instance distribution of the train dataset

VisDrone2019 数据集中的图片尺寸有 960×540 和 1360×765 两种，各个实例的尺寸大小分布如图 9 所示。从图 9 中可以看出大部分目标的尺寸的长宽比

例小于整张图像的 0.1 倍，满足小目标的相对尺寸定义。

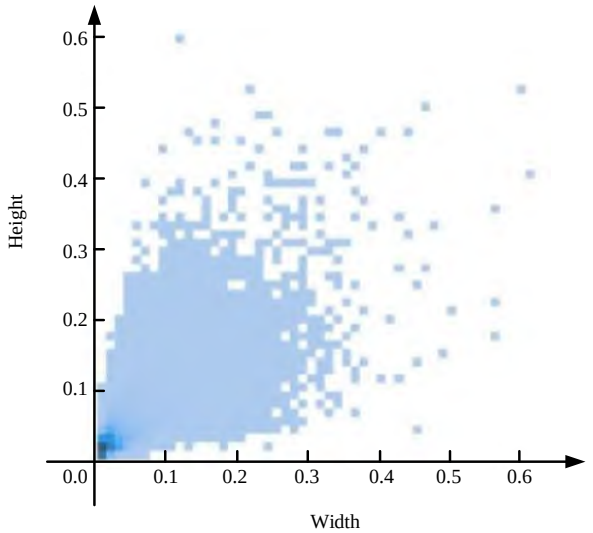


图 9 目标尺寸大小分布图

Fig.9 Object size distribution diagram

3.2 实验环境及参数设置

实验所使用的预训练权重是由 COCO 数据集上进行训练得到。训练模型过程中常用的优化器有 SGD、Adam、RMSProp 等，优化器的性能会影响训练的收敛速度和稳定性。实验中使用了 SGD 优化器，并使用了表 3 中的训练参数，以加快模型的收敛速度。为了实验的公平性，每次实验轮数设置为 300epoch。实验软硬件环境如表 2 所示，训练参数如表 3 所示。

表 2 实验配置

Table 2 Experimental configuration

名称	参数
操作系统	Ubuntu 20.04
显卡	NVIDIA RTX 4090
CUDA	11.8
深度学习框架	Pytorch 2.0.0
语言	Python 3.8

表 3 训练参数

Table 3 Training parameters

参数名称	参数信息
学习率(lr0)	0.01
余弦退火参数(lrf)	0.01
学习率动量(momentum)	0.937
权重衰减系数(weight_decay)	0.0005
批量大小(batchsize)	32
图片尺寸(imgsz)	640×640

3.3 模型评价指标

在模型检测小目标时，通常出现误检漏检问题，因此评价一个模型的检测效果是否精准，通常使用 mAP@0.5、mAP@0.5:0.95，mAP 指标综合了不同类别的精准率（*Precision*，*P*）和召回率（*Recall*，*R*），是一个更加全面的评价指标。在考虑模型检测效果好坏的同时也需要考虑模型的大小。在评估模型大小时，通常使用模型的参数量和 GFLOPs 指标。

（1）平均精度均值（Mean of Average Precision, *mAP*）：mAP@0.5 是所有类别的 IoU 阈值在 0.5 时的平均检测精度；mAP@0.5:0.95 是以步长为 0.05，计算 IoU 阈值在 0.5~0.95 之间的所有 IoU 阈值下的平均检测精度。在目标检测中，mAP 值越高，说明模型检测效果越好。公式为：

$$mAP=\frac{\sum AveragePrecision(c)}{Num(cls)} \tag{9}$$

其中 $AveragePrecision(\cdot)$ 为某个类的平均精度，$Num(\cdot)$ 为数据集所有类别的数量。

（2）精确率：指在所有检测到的目标中，真实的目标数量与总检测目标数量之比。精确率公式为：

$$P=\frac{TP}{(TP+FP)} \tag{10}$$

其中，*TP* 表示真正例，即在检测的结果中正确检测出的目标数量；*FP* 表示假正例，即在检测的结果中被错误检测的目标数量。

（3）召回率：指在所有真实的目标中，被检测到的目标数量与总真实目标数量之比。召回率公式为：

$$R=\frac{TP}{(TP+FN)} \tag{11}$$

其中，*FN* 表示假反例，即在正确的目标中未被检测出的目标数量。

（4）参数量：可以用来评价模型大小和复杂度，是对每一层的权重参数数量进行求和得到。当参数量较小时，表示模型属于轻量模型；当模型参数量较大时，意味着能更好地捕获特征，但是也消耗更多的存储空间和计算资源。

（5）GFLOPs（Floating Point Operations Per Second）：代表模型在推理过程中一秒钟内执行的浮点运算次数，可以用来评估模型的计算复杂度和性能。

3.4 实验与结果分析

3.4.1 注意力机制对比实验

在 BRA 注意力机制中使用了局部上下文增强项（Local Context Enhancement, LCE）[27]，函数 $LCE(\cdot)$ 使用了深度卷积进行参数化。本文设计了实验以探究

不同参数值对模型检测性能的影响。实验数据为VisDrone2019数据集，基线模型为YOLOv5s。实验结果如表4所示。

表4 LCE函数参数值对BD-YOLO性能的影响

Table 4 Effect of LCE function parameter values on BD-YOLO performance

参数值	mAP@0.5	mAP@0.5:0.95	R
1	0.260	0.141	0.39
3	0.260	0.142	0.37
5	**0.271**	**0.147**	**0.39**
7	0.267	0.145	0.39
9	0.263	0.141	0.39

由表4可知，当参数值设置为5时，模型的检测性能可达到最优的效果。当参数值取1和3时，模型性能没有提升的趋势；当参数值取5时，较基础参数1的mAP@0.5和mAP@0.5:0.95指标分别提升了1.1和0.6个百分点。当参数值大于5时，精度值呈现出下降的趋势。综上实验所得的结果，本文将函数$LCE(\cdot)$的参数值设置为5。

为了验证BRA注意力机制对小目标检测效果的有效性，将设计实验使用参数值为5时的BRA与不同注意力进行对比。基线模型为YOLOv5s，数据集为VisDrone2019，训练次数为300轮。实验结果如表5所示。

表5 注意力机制对比试验

Table 5 Comparative experiment of attention mechanism

模型	mAP@0.5	mAP@0.5:0.95
Baseline	0.264	0.142
+BRA(5)	**0.271(+0.007)**	**0.147(+0.005)**
+CA	0.262(-0.002)	0.143(+0.001)
+SE	0.263(-0.001)	0.142(-)
+EMA	0.261(-0.003)	0.139(-0.003)
+CBAM	0.260(-0.004)	0.141(-0.001)

由表5可知，BRA注意力机制凭借其灵活的特征感知能力，使YOLOv5s模型的mAP@0.5指标提升了0.7%，对模型精度的影响明显优于其他注意力机制。CA、SE和CBAM注意力高度依赖通道内的特征信息，由于小目标在通道内的信息相对较少，导致这类注意力很难准确捕获小目标的特征。EMA注意力[28]通过平滑模型的注意力来减少噪声，但是因为这种平滑性使得模型很难定位小目标的位置。

3.4.2 DyHead检测头性能对比实验

通过控制不同数量的DyHead块进行叠加，探究其对模型性能和计算成本的影响。实验基线模型使用YOLOv5s，对基线模型分别叠加1、2、4、6、8、10个DyHead块。其中，叠加个数为0的是基线模型。实验结果如表6所示。

表6 DyHead叠加个数对模型性能的影响

Table 6 Effect of the number of DyHead superposition on model performance

个数	mAP@0.5	R	FPS	参数量	GFLOPs
0	0.264	0.39	91.74	7.03	15.8
1	0.266	0.39	74.63(84.3%)	+0.16	+0.4
2	0.271	0.36	51.55(56.2%)	+0.29	+0.7
4	**0.292**	**0.43**	**67.58(73.7%)**	**+0.56**	**+0.96**
6	0.297	0.44	41.15(44.9%)	+0.82	+1.80
8	0.318	0.48	58.47(63.7%)	+1.08	+2.40
10	0.346	0.54	27.03(29.5%)	+1.35	+2.90

从表6可看出，动态检测头DyHead随着叠加个数的增加，精度也随之增加，但是计算成本和参数量也有小幅度增加。本文将算法精度、复杂度和推理速度进行综合考虑，选择将4个DyHead块叠加集成入算法中，此时模型的平均精度达到29.2%，参数量增加了0.56M，FPS仅降低了26.3%。

3.4.3 损失函数对比实验

在BD-YOLO模型中所使用的损失函数是Focal-EIoU，为了验证该损失函数对模型检测小目标的精度具有更好的提升效果，将对使用不同损失函数后的相同模型进行对比。以YOLOv5s添加注意力机制BRA及修改了带注意力的检测头为基础的模型进行对比实验，训练300轮。实验结果如表7所示。

表7 损失函数对比实验

Table 7 Loss function comparison experiments

方法	mAP@0.5	R	GFLOPs	权重文件大小
CIoU	27.9	51	28.0	16.7
DIoU	26.6	40	26.3	15.0
EIoU	29.2	53	28.0	16.7
Focal	28.8	53	28.2	15.9
Focal-EIoU	**29.6**	**53**	**28.2**	**16.7**

由表7可知，在与其他损失函数相比，Focal-EIoU虽然在GFLOPs和权重文件的大小上都有小幅度增加，但是对模型的检测效果提升最大，说明损失函数Focal-EIoU更适合小目标的检测。

3.4.4 综合对比实验

为了体现BD-YOLO模型的优越性，使用了目前

较为流行的目标检测模型进行对比，在相同的实验环境配置与参数的情况下对 VisDrone2019 数据集进行检测，模型对比实验结果如表 8 所示。

在检测精度上，BD-YOLO 模型在 mAP@0.5、mAP@0.5:0.95 和 R 上都具有较大的优势。mAP@0.5、mAP@0.5:0.95 和 R 指标分别为 32.6%、17.9%、53%。与近年来最新的检测模型相比，BD-YOLO 的检测精度优于 YOLOv5n、TPH-YOLOv5、VA-YOLO 和 YOLOv8n，但不如 YOLOv3 和 YOLOv8s 这类相对较大的模型。

在算法复杂度上，YOLOv3 的 GFLOPs 是 BD-YOLO 的 5.6 倍，参数量多了 53.68M。YOLOv8s 的精度虽然比 BD-YOLO 高，但是 GFLOPs、参数量、权重文件大小分别比 BD-YOLO 大了 0.9G、3.27 百万、6.4MB。因此，BD-YOLO 模型相对较低的计算复杂度和参数量，更有利于存储和部署在边缘设备上。

从表 9 可以看出，与其他模型相比，BD-YOLO 在行人和人的目标上的检测精度是最高的，分别达到了 42.6%和 34.9%。在自行车、汽车、面包车、三轮车和摩托车类别上的检测精度仅次于 YOLOv3 和 YOLOv8s。

综上，BD-YOLO 模型对比基线模型 YOLOv5s 有了明显提升。与其他模型相比，在精度和复杂度上具有较大的优势。因此，说明本文的改进是有效的。

表 8 模型对比实验

Table 8 Model comparison experiments

模型	mAP@0.5	mAP@0.5:0.95	R	GFLOPs	参数量	权重文件大小
YOLOv3-tiny	0.135	0.058	0.31	12.9	8.68	17.5
YOLOv3	0.333	0.185	0.55	154.7	61.54	123.6
YOLOv5s	0.264	0.142	0.39	15.8	7.03	14.4
YOLOv5n	0.204	0.104	0.32	4.2	1.77	3.8
YOLOv5s6	0.276	0.140	0.50	16.2	12.34	25.1
TPH-YOLOv5[29]	0.299	0.169	0.41	22.5	8.41	17.8
VA-YOLO[29]	0.236	0.126	0.36	16.5	6.56	13.6
YOLOv8n	0.266	0.155	0.48	8.1	3.01	6.2
YOLOv8s	0.394	0.234	0.60	28.5	11.13	22.5
BD-YOLO	0.326	0.179	0.53	27.6	7.86	16.1

表 9 类别对比实验结果

Table 9 Category comparison experiment results

模型	mAP	行人	人	自行车	汽车	面包车	卡车	三轮车	遮阳棚-三轮车	公共汽车	摩托车
YOLOv3-tiny	13.5	8.7	7.3	2.5	48.2	13.7	9.2	3.4	2.7	30.5	9.1
YOLOv3	33.3	31.8	20.0	10.7	74.0	34.8	39.8	19.5	15.6	57.5	29.3
YOLOv5s	24.4	14.7	11.7	6.5	64.4	30.1	22.3	9.6	10.9	53.0	20.6
YOLOv5n	20.4	28.3	18.8	5.7	54.6	19.0	9.1	8.0	6.3	30.8	23.0
YOLOv5s6	27.6	24.0	16.4	7.8	68.1	30.3	28.8	13.3	13.1	50.8	22.9
TPH-YOLOv5	29.9	40.1	26.7	12.9	63.6	33.9	19.0	13.5	9.6	47.4	32.8
VA-YOLO	23.6	31.5	24.1	6.8	59.1	24.6	13.9	9.8	8.8	30.1	27.6
YOLOv8n	26.6	19.5	11.0	5.8	66.1	31.6	27.0	12.6	15.8	53.4	23.0
YOLOv8s	39.4	41.7	32.6	13.1	79.4	44.6	36.7	29.5	15.4	56.1	44.3
BD-YOLO	32.6	**42.6**	**34.9**	12.6	70.5	34.7	24.9	16.8	9.4	40.4	39.1

3.4.5 消融实验

为了证明本文提出的每个改进模块对于模型检测能力的提升，将对不同模块对模型检测的效果做消融实验进行评估。以 YOLOv5s 为基线模型，在该模型上

逐个添加改进的模块，首先添加注意力机制 BRA，然后添加带注意力的检测头 DyHead，最后改用 Focal-EIoU 损失函数。数据集选用 VisDrone2019，图片大小设置为 640×640，使用预训练权重加速训练，共训练 300epoch。

由表 10 可知，所改进的模块对模型检测小目标的准确度均有提升。Baseline 模型为无改进的基线模型 YOLOv5s，在 Baseline 上添加注意力机制 BRA 后，虽然参数量和 GFLOPs 分别增加了 0.27M、10.6G，但是 mAP@0.5 提升了 0.7%，说明 BRA 过滤掉最不相关的区域并提高对有价值区域的关注，增强了模型定位小目标的效果；将 Baseline 上的检测头改为 DyHead，在 mAP@0.5 和 mAP@0.5:0.95 指标上的提升最大，分别提高了 2.8%、1.6%，说明检测层带有注意力机制将会大幅提升模型检测小目标的效果；使用 Focal-EIoU Loss 损失函数，将高质量的锚框和低质量的锚框分开，会使得模型的检测精度得到提升；最后，将 BRA、DyHead 和 Focal-EIoU Loss 同时作用于基线模型，在 mAP@0.5 和 mAP@0.5:0.95 指标上较基线模型提升了 6.2 个百分点和 3.7 个百分点，虽然参数量和 GFLOPs 有所增加，但是属于合理范围。

表 10 消融实验

Table 10 Ablation experiment

模型	mAP@0.5	mAP@0.5:0.95	参数量	GFLOPs
Baseline	0.264	0.142	7.03	15.8
+BRA	0.271	0.147	7.30	26.4
+DyHead	0.292	0.158	7.59	17.0
+Focal-EIoU	0.277	0.155	7.03	15.8
+BRA+DyHead	0.285	0.158	7.86	26.5
+BRA+DyHead+ Focal-EIoU	0.326	0.179	7.86	27.6

YOLOv5s 算法和 BD-YOLO 算法的检测结果如图 10 所示，(a)和(c)图为改进前 YOLOv5s 算法检测结果，(b)和(d)图为改进后的 BD-YOLO 算法的检测结果。

(a)YOLOv5s 算法检测结果

(b)BD-YOLO 算法检测结果

(c) YOLOv5s 算法检测结果

(d) BD-YOLO 算法检测结果

图 10 检测效果对比

Fig.10 Comparison of detection effects

4 结束语

针对 YOLOv5s 模型对无人机高空航拍图像的小目标进行检测出现漏检误检的问题，本文基于 YOLOv5s 模型提出了改进的小目标检测模型 BD-YOLO。首先，在网络结构中加入注意力机制 BRA，提高对小目标检测区域的关注度；其次，将 YOLOv5s 中普通的检测层改为使用带有自注意力的检测层 DyHead，以改进模型对小目标的表示能力；最后将 YOLOv5s 中的损失函数 CIoU Loss 替换成 Focal-EIoU Loss，从而进一步提升模型的检测效果。在数据集 VisDrone2019 的实验结果表明，BD-YOLO 模型对于小目标的检测效果都优于其他的主流模型，在没有添加过多参数量的情况下检测精度达到了 32.6%，可以说明本文改进后的模型在无人机视角下的小目标检测效果具有优势。

参考文献:

[1] 韩玉洁,曹杰,刘琨,等.基于改进 YOLO 的无人机对地多目标检测[J].电子测量技术,2020,43(21):19-24.
HAN Y J, CAO J, LIU K, et al. UAV ground multi-target detection based on improved YOLO[J]. Electronic Measurement Technology, 2020,43(21): 19-24.

[2] 丁田,陈向阳,周强,等.基于改进 YOLOX 的安全帽佩戴实时检测[J].电子测量技术,2022,45(17):72-78.
DING T, CHEN X Y, ZHOU Q, et al. Real-time detection of helmet waring based on improved YOLOX[J]. Electronic Measurement Technology, 2022,45(17): 72-78.

[3] 冒国韬,邓天民,于楠晶.基于多尺度分割注意力的无人机航拍图像目标检测算法[J].航空学报,2023,44(05):273-283.
MAO G T, DENG T M, YU N J. Object detection in UAV images based on multi-scale split attention[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(05):273-283.

[4] Zhu X, Lyu S, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios[C]//Pro- ceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 2778-2788

[5] YANG Y Z. Drone-view object detection based on the improved YOLOv5[C]//Proceedings of the IEEE International Conference on Electrical Engineering, Big Data and Algorithms. Changchun: IEEE, 2022: 612–617

[6] Zhu L, Wang X, Ke Z, et al. BiFormer: Vision Transformer with Bi-Level Routing Attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 10323-10333.

[7] Dai X, Chen Y, Xiao B, et al. Dynamic head: Unifying object detection heads with attentions[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 7373-7382.

[8] 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.

[9] 李红光,于若男,丁文锐.基于深度学习的小目标检测研究进展[J].航空学报,2021,42(07):107-125.
LI H G, YU R N, DING W R. Research development of small object traching based on deep learning[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(7) :107-125.

[10] Chen Y, Zhang P, Li Z, et al. Stitcher: Feedback-driven data provider for object detection[J]. arXiv preprint arXiv: 2004.12432, 2020, 2(7): 12.

[11] Kisantal M, Wojna Z, Murawski J, et al. Augmentation for small object detection[J]. arXiv preprint arXiv:1902.07296, 2019.

[12] 李青援,邓赵红,罗晓清,等.注意力与跨尺度融合的 SSD 目标检测算法[J].计算机科学与探索, 2022, 16(11): 2575-2586.
LI Q Y, DENG Z H, LUO X Q, et al. SSD Object Detection Algorithm with Attention and Cross-Scale Fusion[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(11): 2575-2586.

[13] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-time Object Detection with Region Prposal Networks[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(6): 1137–1149.

[14] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once:Unified,Real-time Object Detetion[C]//2016 IEEE Conference on Computer Vision and Pattern Re- cognition.Las Vegas:IEEE,2016:779-788.

[15] REDMON J, FARHADI A.YOLOv3:An Incremental Improvement[J/OL].(2018-04-08)[2023-01-11].https://arxiv.org/abs/1804.02767.

[16] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal Speed and Accuracy of Object Detection[J/OL].(2020-04-23)[2023-01-11].https://arxiv.org/abs/2004.10934.

[17] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]//European Conference on Computer Vision (ECCV). Amsterdam: Springer, 2016: 21-37.

[18] WANG CY, LIAO HY M, WU YH, et al. CSPNet: A New Backbone that Can Enhance Learning Capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE, 2020:1571-1580.

[19] LIN TY, DOLLAR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition.Honolulu:IEEE,2017: 936-944.

[20] LIU S, QI L, QIN H F, et al. Path Aggregation Network for Instance Segmentation[C]//2018 IEEE/CVF Con- ference on Computer Vision and Pattern Recognition. Salt Lake City:IEEE,2018:8759-8768.

[21] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[22] Hou Q, Zhou D, Feng J. Coordinate attention for efficient

mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.

[23] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.

[24] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[25] 俞军,贾银山.改进 YOLOv5 的小目标检测算法[J].计算机工程与应用,2023,59(12):201-207.
YU J, JIA Y S. Improved YOLOv5 for Small Object Detection Algorithm[J]. Computer Engineering and Applications, 2023,59(12):201-207.

[26] 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. Seoul: IEEE, 2019: 213–22

[27] Ren S, Zhou D, He S, et al. Shunted self-attention via multi-scale token aggregation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 10853-10862.

[28] Ouyang D, He S, Zhang G, et al. Efficient Multi-Scale Attention Module with Cross-Spatial Learning[C]// ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[29] 刘展威,陈慈发,董方敏.基于 YOLOv5s 的航拍小目标检测改进算法研究[J/OL].无线电工程: 1-10 [2023-08-20]. http://kns.cnki.net/kcms/detail/13.1097.TN.20230411.1645.026.html.
LIU Z W, CHEN C F, DONG F M. Improved aerial small object detection algorithm based on YOLOv5s[J/OL]. Radio Engineering,1-10[2023-08-20]. http://kns.cnki.net/kcms/detail/13.1097.TN.20230411.1645.026.html.