

doi:10.3969/j.issn.1003-3106.2023.10.008

引用格式: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]. 无线电工程,2023,53(10):2286-2294. [LIU Zhanwei, CHEN Cifa, DONG Fangmin. Improved Aerial Small Object Detection Algorithm Based on YOLOv5s [J]. Radio Engineering, 2023, 53(10): 2286-2294.]

## 基于 YOLOv5s 的航拍小目标检测改进算法研究

刘展威<sup>1,2</sup>, 陈慈发<sup>1,2\*</sup>, 董方敏<sup>1,2</sup>

(1. 三峡大学 计算机与信息学院, 湖北 宜昌 443002;

2. 三峡大学 湖北省建筑质量检测装备工程技术研究中心, 湖北 宜昌 443002)

**摘要:** 针对无人机航拍时拍摄的对象大小不一、种类繁多且容易被建筑遮挡等问题, 提出了一种基于 YOLOv5s 的无人机目标检测改进算法 VA-YOLO。在已有的主干网络中添加 CA 注意力机制模块, 扩大检测区域, 获得更准确的位置信息; 针对检测小目标时尺度不一导致语义丢失的问题, 添加小目标检测层与 BiFPN 结构, 加深浅层语义与深层语义结合, 以此丰富对检测目标的语义信息; 使用损失函数 Varifocal loss 与 EIoU, 改善模型对小目标检测的准确性。实验结果表明, 在 VisDrone2019-DET 数据集上, 该算法的平均检测精度 (mean Average Precision, mAP) 达到了 39.01%, 相比 YOLOv5s 提高了 6.26%。

**关键词:** 无人机航拍; YOLOv5s; 小目标; Varifocal loss

**中图分类号:** TP391.4

**文献标志码:** A

**开放科学(资源服务)标志码(OSID):**

**文章编号:** 1003-3106(2023)10-2286-09



## Improved Aerial Small Object Detection Algorithm Based on YOLOv5s

LIU Zhanwei<sup>1,2</sup>, CHEN Cifa<sup>1,2\*</sup>, DONG Fangmin<sup>1,2</sup>

(1. College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China;

2. Hubei Province Engineering Technology Research Center for Construction Quality Testing Equipment, China Three Gorges University, Yichang 443002, China)

**Abstract:** To solve the problems of different sizes, various types and easy to be blocked by buildings in aerial photography of UAV, an improved UAV target detection algorithm VA-YOLO based on YOLOv5s is proposed. Firstly, add CA attention mechanism module to the existing backbone network to expand the detection area and obtain more accurate location information; secondly, in order to solve the problem of semantic loss caused by different scales of small target detection, the small target detection layer and BiFPN structure are added to deepen the combination of shallow semantics and deep semantics, so as to enrich the semantic information of the detection target; finally, the loss function Varifocal loss and EIoU are used to improve the accuracy of the model for small target detection. The experimental results show that the mean Average Precision (mAP) of the algorithm is 39.01% on the VisDrone2019-DET dataset, which is 6.26% higher than YOLOv5s.

**Keywords:** UAV aerial photography; YOLOv5s; small target; Varifocal loss

### 0 引言

随着基于深度学习的目标检测技术的快速发展,“无人机航拍+目标检测”方法的使用越来越普遍,如森林防火、电路巡检和农业监管等。目前针对“小目标”的定义主要有 2 种:一种是绝对尺寸,尺寸小于 32 pixel×32 pixel 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 pixel×256 pixel 的图像中成像面积小于 80 pixel 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<sup>[1]</sup>,因此研究出一个针对航拍小目标检测的算法模型成为关键的研究内容之一。

目标检测从早期的传统方法到如今基于深度学习的方法,发展已有 21 年。当前基于深度学习的目标检测算法有 2 种:第 1 种是以 Faster R-CNN<sup>[2]</sup>为代表的双阶段检测算法等,利用算法生成预选框,再

收稿日期:2023-04-07

基金项目:国家自然科学基金新疆联合基金重点项目(U1703261)

Foundation Item: Key Program of Xinjiang Joint Fund of National Natural Science Foundation of China (U1703261)

使用深度卷积网络对预选框进行检测类别的分类,双阶段检测算法的精度更高,但速度过慢,不能满足实时性要求较高的场合;第2种是以YOLO<sup>[3]</sup>、YOLOv3<sup>[4]</sup>、YOLOv4<sup>[5]</sup>、YOLOv5以及SSD<sup>[6]</sup>为代表的单阶段检测算法,将检测框的定位和分类任务结合,能够快速检测出目标位置,通过适当地改进可同时具有更好的实时性与检测精度。韩玉洁等<sup>[7]</sup>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<sup>[8]</sup>加入注意力以及CIoU,加快了模型收敛速度,增加检测准确率,但同时也增加了计算成本。肖繁俊等<sup>[9]</sup>加入注意力机制,并引入自适应特征模块,增强了网络特征提取能力,但精度提升幅度较小且漏检率过大。目前改进方法主要通过添加模块来提升检测精度,但同时增加了计算成本。因此本文不局限于模块添加,将在网络结构上进一步改进。

目前无人机航拍检测任务主要有以下难题:

① 不同于普通目标检测,航拍所检测对象大多数是小目标,可以提取到有用的特征过少,绝大多数算法都存在对小目标检测的漏检、误检问题。

② 现实场景中目标繁多,图片中背景复杂,常常混杂许多其他干扰性较大的目标。

③ 由于飞行高度问题,各个目标之间大小不同,存在着尺度不一致的问题,导致检测效果不佳。

为提高无人机航拍目标的检测精度,结合应用场景的限制,本文提出一种基于YOLOv5s的航拍小目标改进算法VA-YOLO。该算法主要改进与创新有:

① 融入模块来重新构建主干网络,提升对小目标的检测精度。

② 通过添加小目标检测层与加权双向金字塔(Bidirectional Feature Pyramid Network, BiFPN)<sup>[10]</sup>将深层语义跟浅层语义进行多尺度特征融合,解决尺度不一、特征信息利用不充分的问题。

③ 将损失函数改为Varifocal loss与EIoU,解决检测精度低与背景干扰过大的问题,使得算法对小目标取得更好的检测效果。

## 1 YOLOv5s 算法

YOLOv5是一种单阶段的目标检测算法,相比较前一代,它汲取了许多优秀网络结构的优点,精度更高、速度也更快,已经能够做到实时地检测目标。YOLOv5一共有4个版本:YOLOv5s、YOLOv5m、YOLOv5l和YOLOv5x四种,其网络深度和宽度逐渐加大。考虑到应用场景,相比其他3个版本,YOLOv5s的网络结构最简洁,运行速度最快,计算资源消耗最少,同时也更容易移植到其他平台,本研究选择使用模型深度与宽度最浅的YOLOv5s模型。其网络结构如图1所示。

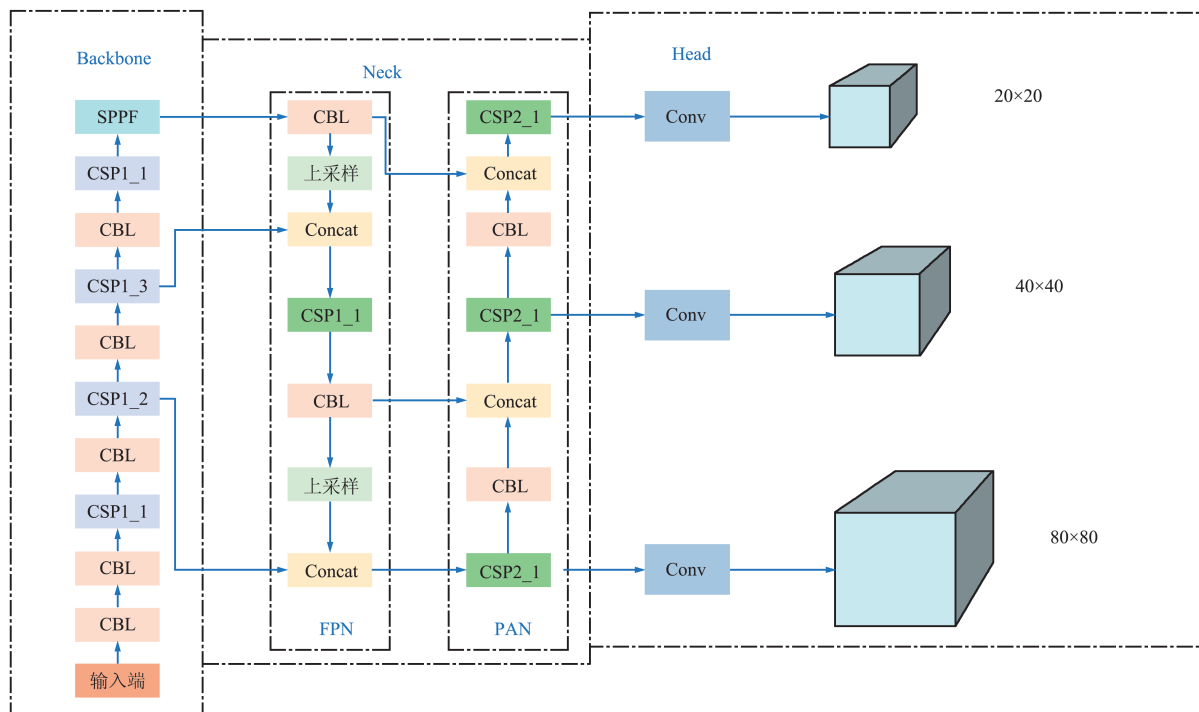


图1 YOLOv5s 结构

Fig.1 YOLOv5s structure

YOLOv5s 模型主要包括图片输入端、Backbone 网络、Neck 网络和 Head 输出层四部分。

输入端会对输入的图片进行预处理。首先,对几张图片进行随机裁剪拼接,不仅可以充分利用数据集,还能提升训练速度;其次,在模型训练中,YOLOv5s 算法根据数据集自适应地计算出最佳锚点框,考虑到无人机航拍检测对象绝大部分为小目标,本文将锚框调整为 $[3, 4, 4, 8, 6, 6]$ 、 $[6, 12, 12, 9, 11, 17]$ 、 $[21, 11, 17, 22, 36, 18]$ 或 $[26, 34, 54, 41, 100, 85]$ ;最后,将图片调整到统一大小后输入到

Backbone 网络中。

YOLOv5s(6.0)<sup>[11]</sup>的 Backbone 网络由特征提取模块 CSP<sup>[12]</sup>、卷积模块 CBL 与空间金字塔池化(Spatial Pyramid Pooling-Fast, SPPF)模块组成。CSP 结构分为 2 类,如图 2 所示。SPPF 模块不再使用 $1 \times 1$ 、 $5 \times 5$ 、 $9 \times 9$ 、 $13 \times 13$ 的最大池化,而是通过多个 $5 \times 5$ 的 MaxPool2d 取代原有的池化层,例如 2 个 $5 \times 5$ 卷积操作代替一个 $9 \times 9$ 的卷积操作,用以提高效率。CBL 由 Conv、BN 归一层和 Leaky ReLU 激活函数组成,如图 2 所示。

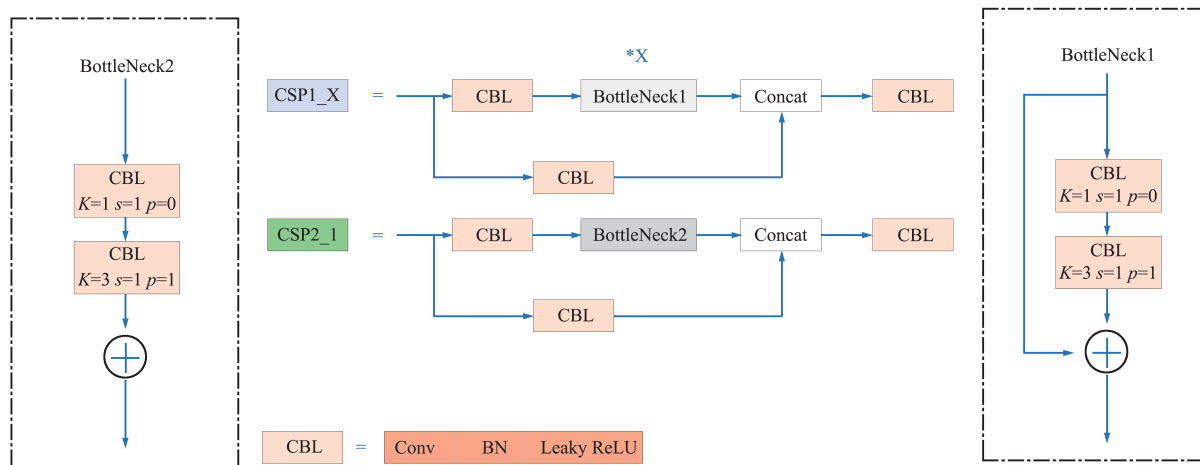


图2 CSP 和 CBL 结构

Fig.2 Structure of CSP and CBL

Neck 由特征金字塔网络 (Feature Pyramid Networks, FPN)<sup>[13]</sup> 和路径聚合网络 (Path Aggregation Network, PAN)<sup>[14]</sup> 构成。FPN 自上而下用来强化语义特征, PAN 自下而上用来强化位置特征,二者将不同阶段特征图进行特征融合,提高对主干网络输出特征的利用率,从而提升对小目标检测精度。

Head 输出层主要是对 Neck 层的 3 个输出进行卷积操作,并在大小为 $20 \times 20$ 、 $40 \times 40$ 、 $80 \times 80$ 的特征图上生成对应的预测边界框。

## 2 VA-YOLO 算法

YOLOv5s 具有较好的检测效果,但其平均检测精度 (mean Average Precision, mAP) 与 YOLOv3、ClusDet<sup>[15]</sup> 等主流算法相比仍有较大差距,因此,本文将从 Backbone 网络、Neck 网络与 Head 输出层、损失函数三方面对 YOLOv5s 进行改进,力求在不增加运算资源消耗的同时提高精度。

### 2.1 Backbone 网络的改进

YOLOv5s 在进行航拍目标检测时,往往检测的对象过小,检测过程中容易丢失特征信息,导致检测

效果不佳。因此本文加入 Coordinate Attention (CA)<sup>[16]</sup> 注意力机制,可以让算法模型集中在需要关注的区域,有效提升对小目标检测的精度。与 Squeeze-and-Excitation (SE)<sup>[17]</sup> 注意力机制使用 2D 全局池化转变成单个特征向量不同的是,CA 将其分成了 2 个不同方向聚合特征的 ID 编码,既可以获得空间方向的长程依赖,同时也不会缺少位置信息。CA 结构如图 3 所示。

首先在输入端使用 2 个池化核向水平方向和竖直方向进行池化,得到高度为  $h$ , 宽度为  $w$  的输出。然后将输出的特征图进行融合拼接后使用 $1 \times 1$ 大小的卷积核进行卷积,生成过渡特征图。其次,将过渡特征图分解为单独的特征图,再使用 $1 \times 1$ 大小的卷积核进行卷积,得到与输入一致的通道数量。最后使用激活函数,将结果与输入进行相乘输出,从而达到强化特征的目的。

为同时确保检测网络的大小与精度,经过实验,选择在 Backbone 网络中将原有的一个 CSP 模块替换为 CA 模块,不仅计算资源开销更低效果也更为明显,实验对比结果如图 4 所示。图 4(a) 为未更改

的 YOLOv5s 检测情况,图 4(b)为将 CSP 替换为 CA 的检测效果图。可以看到图片下方不仅检测出多个漏掉的行人以及摩托目标,而且准确率也有所提升。

实验结果表明,加入注意力机制后,在检测小目标上模型效果得到显著提升的同时也具有较好的鲁棒性。

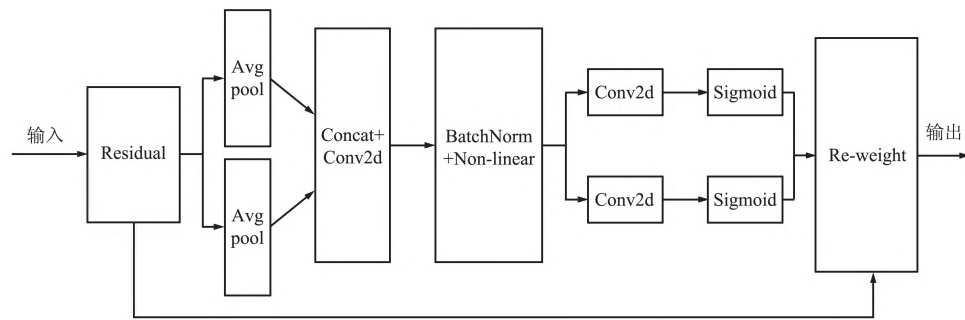


图 3 CA 结构  
Fig.3 CA structure



(a) 未更改的 YOLOv5s 检测效果



(b) 替换为 CA 的检测效果

图 4 检测效果对比

Fig.4 Comparison of detection effects

## 2.2 Neck 网络与 Head 输出层的改进

通常来讲,无人机航拍图像检测都是小目标,目标的类别以及位置尤为重要,对于较浅层分辨率较高的特征图,往往能够保留更加全面的小目标位置信息;较深层的分辨率较低的特征图,往往保留大目标特征复杂的语义信息。因此,本文在 YOLOv5s 原

有基础上,通过添加小目标检测层来继续上采样扩大特征图,进而增强多尺度之间的特征图的关联,让更为浅层的位置信息与深层的语义信息进一步结合,提升对小目标信息的利用率。

YOLOv5s 中 Neck 网络部分使用的是 FPN + PAN 结构,主要是为了解决检测小目标精度不佳的情况,通过 FPN 结构将浅层特征图的位置信息进行向下的递增,再使用 PAN 结构做向上的语义信息的递增,达到融合多层特征信息特征图的效果。但这种方式只能进行相同尺寸特征图之间的融合,未利用到不同尺寸的特征图信息。因此,本文引入了新的 BiFPN 特征图融合网络结构。首先,将单个特征图输入的节点去除,由于该节点融合的特征信息过少,去掉可以避免不必要的运算;其次,融合同一尺度但更浅层的特征图,强化了特征图的丰富性;最后,不限于自顶向下或者自底向上的路径,将删除节点的输入特征图输出到下一级特征图中,形成新的 BiFPN 结构,如图 5 所示。P1、P2、P3、P4 为经过主干网络的特征图,F1、F2、F3、F4 为经过 BiFPN 处理后的特征图。

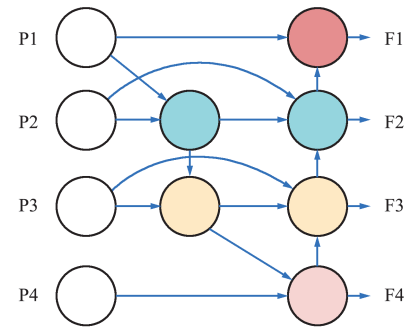


图 5 BiFPN 结构  
Fig.5 BiFPN structure

BiFPN 使用加权融合,通过权值来控制对不同特征信息的学习程度,这里 BiFPN 使用快速归一化法,即:

$$Out = \sum_i \frac{\omega_i}{\varepsilon + \sum_i \omega_i} I_i, \quad (1)$$

式中: $\omega_i$  为经过网络训练的权重,由激活函数确保其大于等于零, $\varepsilon$  用来保持整体结果的稳定性,为常数, $I_i$  表示输入图片的特征, $Out$  表示融合结果。经过改进后,BiFPN 从 Backbone 网络中提取不同尺度的特征图进行加权特征融合。

增加小目标检测层后,其对应检测层输出也比原有多了一个,分别为  $160 \times 160$ 、 $80 \times 80$ 、 $40 \times 40$ 、 $20 \times 20$ ,其中最大的特征图为添加的小目标检测层。例如,在 17 层过后,增加上采样的同时,在 20 层与 Backbone 层中的第 2 层特征图进行拼接融合,由此获得更大的特征图用以进行小目标检测。

### 2.3 损失函数的改进

YOLOv5s 的损失计算主要有矩形框损失 (box\_loss)、置信度损失 (obj\_loss) 和分类损失 (cls\_loss) 三种,YOLOv5s 使用 CIOU\_LOSS<sup>[18]</sup> 作为衡量矩形框的损失。矩形框损失函数为:

$$L_{CIOU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v, \quad (2)$$

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^p}{h^p} \right)^2, \quad (3)$$

式中: $IoU$  为预测的物体框和真实的物体框的交集的面积与并集的面积之比, $b$ 、 $b^{gt}$  分别表示预测框和真实框的中心点, $\rho^2(b, b^{gt})$  为计算 2 个中心点的欧式距离, $c$  为能同时包含预测框和真实框的最小包围区域的对角线距离, $\alpha$  为权重函数, $v$  用来度量长宽比的相似性, $w$  为宽, $h$  为高, $gt$  为真实框, $p$  为预测框。

CIOU 同时考虑了边界框的重叠面积、中心点距离以及长宽比,但长宽比  $v$  是一个模糊值,并不精确,并且 CIOU 未能很好地平衡正负样本的问题。所以本文选用 EIou<sup>[19]</sup> 作为矩形框回归的损失函数,EIoU\_Loss 函数为:

$$L_{EIou} = L_{IoU} + L_{dis} + L_{asp} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{C^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2}, \quad (4)$$

式中:EIoU Loss 总共包含 3 个损失,即重叠损失、中心距离损失和宽高损失,前半部分参数与 CIOU 一致, $C_w$  与  $C_h$  为覆盖 2 个边界框的最小外接框的宽

度与长度。EIou 在 CIOU 基础上将宽高比进一步细化,解决了宽高比过于模糊的问题,有效提升了回归框的精度。

YOLOv5s 的置信度与分类损失使用 BCEWithLogitsLoss 损失函数,该损失函数将 sigmoid 与 BCELoss 整合到一起,在输入之前自动进行 sigmoid 激活操作映射到  $(0, 1)$ ,然后再对输出和 target 进行 BCELoss 处理。其 BCEWithLogitsLoss 计算公式为:

$$loss = \frac{1}{N} \sum_{n=1}^N l_n, \quad (5)$$

$$l_n = -\omega [y_n \cdot \ln x_n + (1 - y_n) \cdot \ln(1 - x_n)], \quad (6)$$

式中: $N$  为样本数量, $l_n$  为第  $n$  个样本对应的 loss, $x_n$  为对应样本的预测概率,由 sigmoid 激活处理, $y_n$  为对应样本类别的真实概率,取 0 或 1, $\omega$  为超参数,表示样本真实类别。sigmoid 函数为:

$$sigmoid = \frac{1}{1 + \exp(-x)}. \quad (7)$$

BCEWithLogitsLoss 通常用于检测较为明显的样本,但对于无人机航拍的小目标样本而言,存在着大量负样本,类别极其不均衡,且占比大容易对训练结果产生影响,从而降低模型检测目标的精度。因此,本文引入置信度与定位精度的 IoU 感知分类评分 (IoU-Aware Classification Score, IACS),该评分可以同时表示物体样本置信度与位置精度,从而在密集小目标中生成更为优秀的检测排序。在此基础上引入了 Varifocal loss,该函数由 Focal Loss<sup>[20]</sup> 改进而来,Varifocal loss 函数定义为:

$$VFL(p, q) = \begin{cases} -q(q \ln(p) + (1-q) \ln(1-p))^\gamma, & q > 0 \\ -\alpha p^\gamma \ln(1-p), & q = 0 \end{cases}, \quad (8)$$

式中: $p$  为前景类的预测概率, $q$  为目标 IoU 得分。在训练的正样本中,将  $p$  设置为生成的预测框和真实框之间的 IoU;而在训练的负样本中,所有类别的训练目标  $q$  均为 0。

本文借鉴了 Focal Loss 利用加权的方法解决样本不平衡的问题,Focal Loss 选择平等地对待正负样本,但本文选择有所重点地对待它们。Varifocal loss 首先预测 IACS 评分,再通过  $p$  的  $\gamma$  次方人为地降低负样本的权重,正样本保持不变。再通过  $q$  的 IoU 得分提升正样本的权重,由此实现将训练重点放在高质量的样本上。同时,为了让正负样本的比例更加合理,在上述公式中添加了  $\alpha$  因子来进行平衡。

### 2.4 VA-YOLO 网络结构

根据上述改进方法,可以得到本文的 VA-YOLO 算法,其网络结构如表 1 所示。



表 1 VA-YOLO 网络结构  
Tab.1 VA-YOLO network structure

序号	连接模块	参数量	模块	模块参数
0	-1	3 520	Conv	[3,32,6,2,2]
1	-1	18 560	Conv	[32,64,3,2]
2	-1	18 816	CSP	[64,64,1]
3	-1	73 984	Conv	[64,128,3,2]
4	-1	115 712	CSP	[128,128,2]
5	-1	295 424	Conv	[128,256,3,2]
6	-1	625 152	CSP	[256,256,3]
7	-1	1 180 672	Conv	[256,512,3,2]
8	-1	25 648	CoorAtt	[512,512]
9	-1	656 896	SPPF	[512,512,5]
10	-1	131 584	Conv	[512,256,1,1]
11	-1	0	Upsample	[None,2]
12	[-1,6]	0	Concat	[1]
13	-1	361 894	CSP	[512,256,1,False]
14	-1	33 024	Conv	[256,128,1,1]
15	-1	0	Upsample	[None,2]
16	[-1,4]	0	Concat	[1]
17	-1	90 880	CSP	[256,128,1,False]
18	-1	8 320	Conv	[128,64,1,1]
19	-1	0	Upsample	[None,2]
20	[-1,2]	0	Concat	[1]
21	-1	22 912	CSP	[128,64,3,2]
22	-1	36 992	Conv	[64,64,3,2]
23	[-1,18,4]	0	Concat	[1]
24	-1	90 880	CSP	[256,128,1,False]
25	-1	147 712	Conv	[128,128,3,2]
26	[-1,14,6]	0	Concat	[1]
27	-1	361 984	CSP	[512,256,1,False]
28	-1	590 336	Conv	[256,256,3,2]
29	[-1,10]	0	Concat	[1]
30	-1	1 182 720	CSP	[512,512,False]

3 实验与结果分析

实验环境配置:CPU 为 AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz, GPU 为 NVIDIA RTX3060 6 GB 显存,操作系统为 Windows 11,学习框架为 Pytorch1.9.0。

3.1 数据集

本文所使用的数据集是 VisDrone2019-DET<sup>[21]</sup>,共包含 8 599 张静态图像,其中 6 471 张用于训练,548 张用于验证,1 580 张用于测试。类别主要包含行人、人、汽车、面包车、巴士、卡车、摩托车、自行车、遮阳棚三轮车和三轮车。

3.2 消融实验

使用单类别平均精度(Average Precision, AP)、多类别 mAP、精准率(Precision)、模型权重和检测速度(Frame per Second, FPS)作为评价指标。共设置了以下几组不同情况下的实验:YOLOv5s、替换为 CA 模块的 YOLOv5s、网络结构更改后的 YOLOv5s、损失函数改进后的 YOLOv5s、VA-YOLO,以此来验证各个模块的提升效果。实验结果如表 2 所示。

表 2 消融实验结果  
Tab.2 Results of ablation experiments

算法	mAP/%	精准率/%	模型权重/MB	FPS
YOLOv5s	32.75	47.98	13.7	85
CA-YOLOv5s	34.89	47.99	11.5	89
结构更改	37.91	51.25	14.6	75
Varifocal-YOLOv5s	34.53	47.11	13.7	88
ElIoU-YOLOv5s	35.75	46.44	13.7	83
VA-YOLO	39.01	51.63	12.4	75

表 2 中的结果表明,加入 CA 注意力机制后,mAP 提升了 2.14%,模型权重减少 2.2 MB,这是因为将 CSP 替换为更轻量化的 CA 后,整体网络层数减少了 13 层,从而训练生成的模型权重也略微降低;结构更改后 mAP 提升了 5.16%,但模型权重增加 0.9 MB;Varifocal loss 的引入带来精度的小幅度提升,同时不会导致模型体积的增加;损失函数的改进在保证模型权重不变的情况下,mAP 分别提高了 1.78%、3%。综上所述,VA-YOLO 的绝大多数指标有优势,其中 mAP 提升了 6.26%,且模型权重也有所下降,仅只有 FPS 低于其他模型。由此可见,改进后的算法对网络性能的提升起到了明显作用。

3.3 对比试验

为了验证本文算法较其他算法的优势,本文与各种主流算法进行对比,结果如表 3 所示。其中 mAP 相比 Mixed YOLOv3-LITE 提升了 10.5%,相比次高位提升了 6.2%。而在行人、人、汽车、面包车、巴士、摩托车和自行车这 7 个类别的 AP 都达到了最优检测效果,各自取得了 48.2%、37.4%、81.2%、40.8%、51.4%、44.8%、14.8%。本文算法经过改进后对小

目标关注度更多,提升了对其特征信息的利用,同时融合多个尺度的特征图,减少特征信息的丢失,提升了对小目标的检测效果。根据实验可知,本文算法

在检测种类繁多、检测对象过小时,依然能够充分利用特征图中的小目标信息,在进行航拍图像小目标检测时拥有更大优势。

表 3 对比试验结果  
Tab.3 Comparative experiment results

算法	精度/%										
	mAP	行人	人	汽车	面包车	巴士	卡车	摩托车	自行车	遮阳棚三轮车	三轮车
Faster R-CNN <sup>[22]</sup>	21.7	21.4	15.6	51.7	29.5	31.4	19.0	20.7	6.7	7.7	13.1
DetNet59 <sup>[21]</sup>	15.3	15.3	4.1	36.1	17.3	26.0	20.9	10.9	3.1	10.5	13.5
RetinaNet <sup>[22]</sup>	13.9	13.0	7.9	45.5	19.9	17.8	11.5	11.8	1.4	4.2	6.3
RefineDet <sup>[21]</sup>	14.9	14.9	3.7	30.1	16.3	21.9	18.1	8.38	2.02	10.3	9.0
Cascade-RCNN <sup>[22]</sup>	23.2	22.2	14.8	54.6	31.5	34.9	21.6	21.4	7.6	8.6	14.8
CenterNet <sup>[23]</sup>	26.0	28.0	11.6	51.0	36.5	37.7	27.9	21.0	9.0	19.9	20.1
Light-RCNN <sup>[21]</sup>	16.5	17.0	4.8	32.3	22.1	29.0	18.4	11.9	5.7	11.9	16.6
RRNet <sup>[24]</sup>	29.1	30.4	14.9	51.4	36.1	44.2	35.2	25.9	13.7	19.0	28.0
CornerNet <sup>[21]</sup>	17.4	20.4	6.6	40.9	20.2	24.4	20.5	12.1	4.6	9.3	14.0
Mixed YOLOv3-LITE <sup>[25]</sup>	28.5	34.5	23.4	70.8	31.3	40.9	21.9	32.7	7.9	6.2	15.3
YOLOv5s	32.8	38.7	30.5	71.6	35.2	44.9	30.0	37.8	10.9	9.6	18.4
VA-YOLO	39.0	48.2	37.4	81.2	40.8	51.4	33.0	44.8	14.8	12.4	24.7

3.4 算法效果与分析

为了进一步验证本文算法在实际场景中的效果,选取测试集中具有挑战性的图片进行可视化对比,类型分别为:密集人群小目标、复杂干扰背景小目标、高空车辆小目标、远距离多尺度小目标。检测结果如图 6 所示,左侧为本文算法实验效果,右侧为未经更改的 YOLOv5s 实验效果。经过改进后的算法融合了不同尺度下小目标的语义信息,强化了对

小目标的特征信息利用,同时也减少对目标的误检、漏检问题。例如在夜间密集人群环境下,本文算法能够检测更多行人目标与车辆目标,且受到光照等背景因素影响的可能性也更小,有效降低了误检率与漏检率。综上所述,本文改进后的算法对比未经更改的 YOLOv5s 算法具有显著的优势,检测准确率更高、抗干扰更强,对于常见环境下密集小目标检测的特征信息提取均有明显提升。



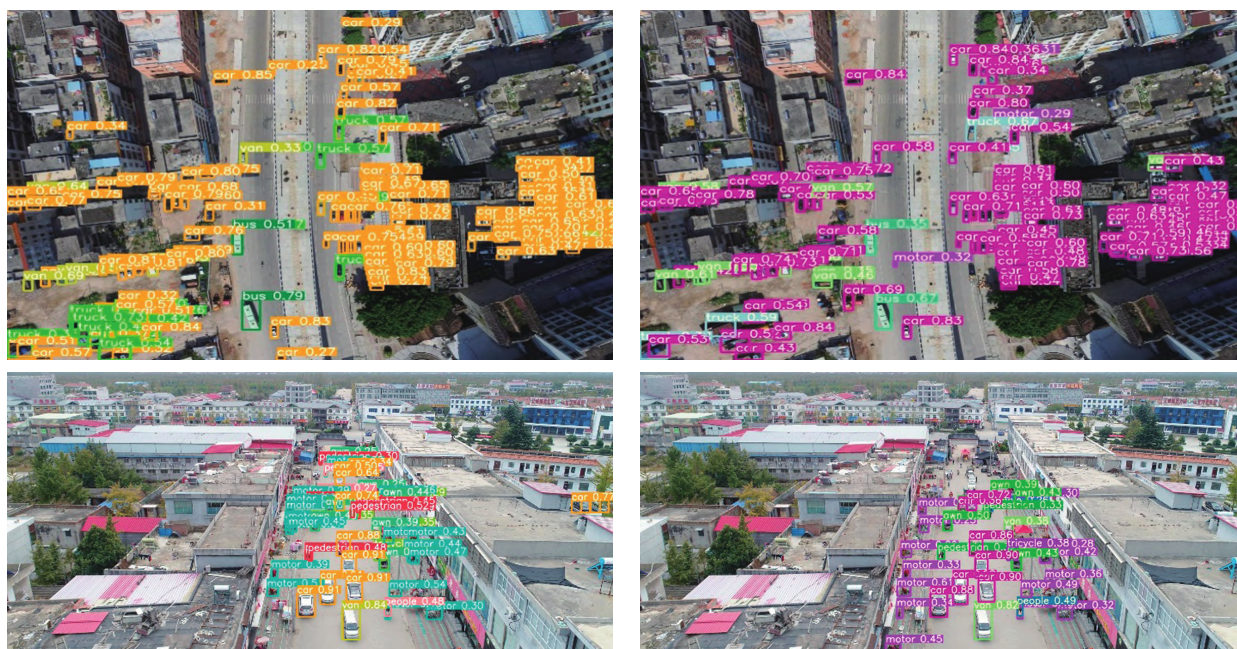


图6 检测效果对比

Fig.6 Comparison of detection effects

## 4 结束语

本文提出的基于 YOLOv5s 的无人机目标检测算法,针对航拍目标对象大小不一致、类别繁多且易受背景干扰的问题,首先用 CA 模块替换掉 Backbone 网络中的一个 CSP 模块,在能够小幅提升检测精度同时,还能降低模型参数量与模型权重,并对密集小目标更有兴趣。其次更改网络结构,将 YOLOv5s 中未被利用到的浅层特征图与已有的深层特征图进行融合,并添加小目标检测层来对小目标进行针对性的检测,有效减少语义与位置信息的丢失。接着改进损失函数,引入了 Varifocal loss 与 EIoU 函数,降低负样本权重并提高正样本权重的同时提升回归框的准确性。最后同几种主流算法进行对比实验,实验证明,本文算法的精度在 Vis-Drone2019-DET 数据集上同其他算法相比具有一定优势,但由于网络结构的更改,一定程度上增加了训练之后模型推理的时间开销,因此后续工作重点将优化模型参数,在保证速度的同时使其能够更好地应用于不同环境。



## 参考文献

- [1] 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(7):107-125.
- [2] 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.
- [3] 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. Las Vegas; IEEE, 2016: 779-788.
- [4] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement [J/OL]. (2018-04-08) [2023-01-11]. <https://arxiv.org/abs/1804.02767>.
- [5] 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>.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 肖繁俊,潘睿志,李超,等. 基于改进 YOLOv5s 绝缘子缺陷检测技术研究 [J/OL]. 电子测量技术: 1-7 [2023-01-11]. <http://kns.cnki.net/kcms/detail/11.2175.TN.20220815.1031.004.html>.



- [10] TAN M, PANG R, LE Q V. EfficientDet: Scalable and Efficient Object Detection [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10778–10787.
- [11] JOCHER G, NISHIMURA K, MINEEVA T, et al. YOLOv5 (Minor Version 6.0) [EB/OL]. (2021–10–12) [2023–01–11]. <https://github.com/ultralytics/yolov5/releases/tag/v6.0>.
- [12] WANG C Y, LIAO H Y M, WU Y H, 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.
- [13] LIN T Y, 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.
- [14] LIU S, QI L, QIN H F, et al. Path Aggregation Network for Instance Segmentation [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759–8768.
- [15] WU X, LI W, HONG D F, et al. Deep Learning for UAV-based Object Detection and Tracking: A Survey [J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10 (1): 91–124.
- [16] HOU Q B, ZHOU D Q, FENG J S. Coordinate Attention for Efficient Mobile Network Design [C] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13713–13722.
- [17] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42 (8): 2011–2023.
- [18] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression [C] // Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 12993–13000.
- [19] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and Efficient IOU Loss for Accurate Bounding Box Regression [J/OL]. (2021–01–20) [2023–01–11]. <https://arxiv.org/abs/2101.08158>.
- [20] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection [C] // 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2999–3007.
- [21] 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: 213–226.
- [22] YU W P, YANG T J N, CHEN C. Towards Resolving the Challenge of Long-tail Distribution in UAV Images for Object Detection [C] // 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021: 3258–3267.
- [23] ZHOU X, WANG D, KRAHENBUHL P. Objects as Points [J/OL]. (2019–04–16) [2023–01–11]. <https://arxiv.org/abs/1904.07850>.
- [24] CHEN C R, ZHANG Y, LYU Q X, et al. RRNet: A Hybrid Detector for Object Detection in Drone-captured Images [C] // 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul: IEEE, 2019: 100–108.
- [25] ZHAO H P, ZHOU Y, ZHANG L, et al. Mixed YOLOv3-LITE: A Lightweight Real-time Object Detection Method [J]. Sensors, 2020, 20 (7): 1861–1878.

#### 作者简介

刘展威 男, (1998—), 硕士研究生。主要研究方向: 目标检测。

(\* 通信作者) 陈慈发 男, (1967—), 教授, 硕士生导师。主要研究方向: 嵌入式系统、物联网和计算机测控系统。

董方敏 男, (1965—), 博士, 教授, 博士生导师。主要研究方向: 计算机图形图像处理、计算机视觉和智能信息处理等。