

# 一种无人机视角下的小目标检测算法

李 杨<sup>1</sup>, 武连全<sup>1</sup>, 杨海涛<sup>2</sup>, 牛瑾琳<sup>2</sup>, 楚宪腾<sup>2</sup>, 王华朋<sup>2</sup>, 邹清龙<sup>1</sup>

(1. 中国刑事警察学院 警察技能战术训练部, 辽宁 沈阳 110854;

2. 中国刑事警察学院 公安信息技术与情报学院, 辽宁 沈阳 110854)

**摘要:** 使用无人机对场景区域中的人、车、物、事等小目标进行实时有效监测有利于维护公共安全。针对无人机视角下小目标存在的目标遮挡、重叠、复杂环境干扰等问题, 提出一种无人机视角下的小目标检测算法, 该算法使用 You Only Look Once X (YOLOX) 网络作为基线系统, 首先在 Neck 网络部分增大输出特征图减小感受野提高网络的细节表现能力, 删除小尺寸特征图的检测头提高小目标的检出率; 其次使用 Anchor Free 的关联机制, 降低真值标签中噪声的影响并同时减少参数设置加快网络运行; 最后提出一种小目标真实占比系数来计算小目标的位置损失, 该系数增大对小目标误判的惩罚使网络对小目标更加敏感。使用该算法在 VisDrone2021 数据集上进行实验, mAP 值较基线系统提高了 4.56%, 参数量减少 29.4%, 运算量减少 32.5%, 检测速度提升 19.7%, 较其他主流算法也具有优势。

**关键词:** YOLOX; VisDrone2021; 小目标; 无人机航拍

**中图分类号:** TP391.4      **文献标志码:** A      **文章编号:** 1001-8891(2023)09-0925-07

## A Small Target Detection Algorithm from UAV Perspective

LI Yang<sup>1</sup>, WU Lianquan<sup>1</sup>, YANG Haitao<sup>2</sup>, NIU Jinlin<sup>2</sup>, CHU Xianteng<sup>2</sup>, WANG Huapeng<sup>2</sup>, ZOU Qinglong<sup>1</sup>

(1. Department of Police Skills and Tactics Training, Criminal Investigation Police University of China, Shenyang 110854, China;

2. Video and Audio Material Examination Department, Criminal Investigation Police University of China, Shenyang 110854, China)

**Abstract:** The use of unmanned aerial vehicles (UAVs) for effective real-time monitoring of small targets, such as people, cars, and objects in the scene area, can help maintain public security. To address the problems of small-target occlusion, overlapping, and interference of complex environments in UAV images, a small-target detection algorithm is proposed from the UAV perspective. The algorithm uses the YOLOX network as the baseline system. First, the neck part of the network increases the output feature graph to reduce the receptive field, thereby improving the performance of the network details, and the detection head of the small-sized feature graph is deleted to improve the detection rate of small targets. Second, the anchor-free association mechanism is used to reduce the influence of noise in the truth tag while simultaneously reducing the parameter setting to speed up network operations. Finally, a true proportion coefficient is proposed for small targets to calculate position loss, thereby increasing the penalty for misjudging small targets, which makes the network more sensitive to small targets. Experiments on the VisDrone2021 dataset using this algorithm showed that the mAP value increased by 4.56%; the number of parameters decreased by 29.4%; the amount of computation decreased by 32.5%; and the detection speed increased by 19.7% compared with those of the baseline system, which is an advantage over other mainstream algorithms.

**Key words:** YOLOX, VisDrone2021, small object, drone aerial photography

## 0 引言

无人机航拍在维护公共安全方面具有广泛应用前景, 通过对场景区域中的人、车、物、事等小目标

进行实时有效监测能够提高公安机关执法办案效率, 有效预防、打击违法犯罪行为<sup>[1]</sup>。随着目标检测的发展, 对于大、中目标的检测取得了不错的成果, 但是目前无人机视角下的小目标检测的效果并不理想<sup>[2]</sup>:

收稿日期: 2022-07-02; 修订日期: 2023-01-18.

作者简介: 李杨 (1987-) 男, 硕士, 讲师, 主要从事警务指挥与战术、反恐处置与大数据应用研究。E-mail: 369352485@qq.com。

基金项目: “公共安全风险防控与应急技术装备”国家重点专项 2018 年度项目 (2018YFC0810102)。

一方面由于现实场景复杂,目标之间存在遮挡,重叠等情况,并且无人机航拍距离远、高速低空飞行产生运动模糊以至成像图片中的目标物体较小,所能够被利用特征较少,定位要求高;另一方面由于使用的神经网络存在一定的局限性不能很好地捕捉小目标的有效特征并且用于训练网络模型的数据正负样本不平衡。

以上情况都会导致检测效果变差,为此学者们在不同方面提出优化策略以期提高检测精度。研究表明<sup>[3]</sup>对前端数据进行增强可以提高小目标的特征利用率,能够增强网络的鲁棒性及泛化能力,但是它一定程度上增加了数据量给计算增加难度,更加考验网络模型的性能。现阶段目标检测网络模型可分为双阶段算法和单阶段算法<sup>[4]</sup>。双阶段代表算法是 Region-CNN (R-CNN)<sup>[5]</sup>, Chen 等<sup>[6]</sup>在 R-CNN 网络中引入上下文模型使得网络模型对小目标感知更为灵敏,相较于基线系统的 R-CNN 网络模型平均精准率均值提高了 29.8%,但是双阶段算法通过牺牲检测速度来换取准确率,实际应用场景受限。单阶段算法则以 YOLO 系列网络为代表<sup>[7]</sup>,此系列网络检测速率高,准确率逐代增加是现阶段进行目标检测的主流模型。Cao 等<sup>[8]</sup>在 YOLOv4 的主干网络加入注意力机制提升网络对小目标的感知能力取得了不错的检测效果。奉志强等<sup>[9]</sup>在 YOLOv5 主干网络加入 Transformer 结构并将空间注意力和通道注意力结合最后进行多尺度特征融合提高了检测精度。Akyon<sup>[10]</sup>团队则提出一种切片辅助超推理的 Finetune 策略,能够有效提升检测精度并且方便部署。以上算法在小目标检测上都取得了不错的效果,但是仍然存在一些问题:①主流算法多使用 Anchor-Based 方法,锚框的设定需要引入大量超参,可能带来冗余框并且不能很好地应对复杂背景的干扰,加大网络训练难度。②对无人机视角下的图像背景复杂存在遮挡、重叠,运动模糊等情况考虑不完善,容易造成漏检、误检。③网络结构下采样过多会破坏特征图的分辨率和小目标的特征。YOLOX<sup>[11]</sup>是以 YOLOv3 为基线系统进行的改进,较 YOLOv5 性能方面也有自身的优势,因此在各种目标检测任务中也有所应用<sup>[12]</sup>。本文以 YOLOX 为基线系统,针对小目标检测任务的特点结合基线系统特性,提出一种无人机视角下的小目标检测算法,该算法主要改进为:

- 1) 使用 YOLOX 的 Anchor Free 无锚机制直接预测小目标的边框,减少真值标签中的噪声提升性能,减少网络超参的设定降低网络的复杂程度加快运算;
- 2) 改进多尺度学习,通过增大输出特征图减小感受野提高网络的细节表现能力,同时只使用一个检测

头提高小目标的检出率,加快网络运行;

- 3) 优化损失函数,使用一种新的损失函数计算方法提高小目标对损失的贡献,加大网络误判惩罚。

## 1 无人机视角下的小目标检测

### 1.1 骨干网络

骨干网络在小目标检测过程中随着网络层次的加深会造成小目标特征的破坏,但是网络过浅则对信息提取不充分,因此需要选取特征提取能力强大而又不会破坏目标特征的主干网络。CSPDarknet53 网络是在 Darknet53<sup>[13]</sup>中引入 CSPNet<sup>[14]</sup>结构,该结构将输入数据分成两部分,然后通过 Cross-stage hierarchy 跨阶段分层结构进行组合,将上一单元的数据进行下采样的同时保留自身的特征能够丰富不同梯度的信息,减少因下采样而造成的特征损失也避免了学习重复的信息,丰富了梯度组合,也减少了运算量,能够充分提取小目标信息而不降低分辨率破坏小目标特征,骨干网络整体网络框架如图 1 所示。

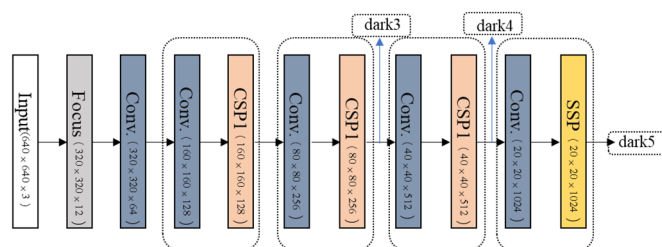


图 1 骨干网络整体网络框架

Fig.1 Backbone network structure

CSPDarknet53 由 3 个 CSP 模块和 1 个 Focus 模块组成,将  $640 \times 640 \times 3$  ( $C, H, W$ ) 的无人机遥感图输入网络首先通过 Focus 模块进行切片,Focus 将图片按间隔像素取值将图片的宽度、高度信息转换到通道,通道数扩充 4 倍这样下采样之后不损失特征信息,切片后再拼接特征图的大小为  $320 \times 320 \times 12$ 。经卷积下采样后输入到第一个跨阶段局部网络 (Cross Stage Partial, CSP) 结构中,依次为卷积层和 CSP 模块,最后一个模块为空间金字塔池化 Spatial Pyramid Pooling, SPP),该结构可以忽略输入特征图的大小尺寸,对物体变形具有鲁棒性。CSP、Focus、SSP 结构分别如图 2 中(a)、(b)、(c)所示,其中的 CBS 表示卷积层,归一化层 (Batch Normalization, BN) 和激活函数 SiLU, Res 表示残差网络。

骨干网络中卷积层的激活函数使用 SiLU 计算公式如式(1)所示,能够提升网络泛化能力加快网络收敛。

$$\text{SiLU}(x) = x \times \text{sigmoid}(x) \quad (1)$$

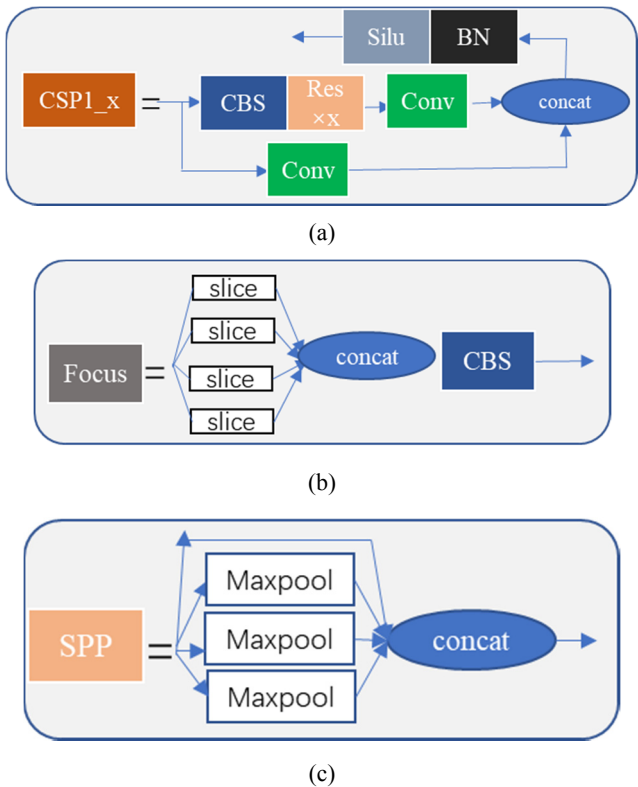


图2 骨干网络内模块结构图 (a) CSP 结构; (b) Focus 结构; (c) SPP 结构

Fig.2 Model structure diagram in backbone network (a)CSP Structure; (b) Focus structure; (c) SPP structure

1.2 Neck 多尺度特征融合

在目标检测中，浅层网络具有的感受野小语义信息较弱但是细节表达能力强，网络的感受野随着网络加深逐渐变大细节表征能力也逐渐减弱，网络过深或过浅都会影响目标检测的准确率。多尺度特征融合是平衡这一矛盾的有效策略，YOLOX 的 Neck 部分通

常使用的特征金字塔 (Feature Pyramid Networks, FPN),通过浅层特征图与深层特征图的融合使得新生成的特征图同时兼具细节和语义表达能力从而提高目标检测的准确率，该结构最后输出 3 个特征图的大小分别为  $20\times20\times1024$ ,  $40\times40\times512$ ,  $80\times80\times256$ , 对应下采样倍数为 32、16、8，分别有利于大、中、小型目标的检测。该方法在尺寸不一的数据集上表现良好，能够提升系统的鲁棒性和泛化能力。

图3所示为本文所用数据集标签大小分布，(a)为 1000 数据量标签分布情况，(b)为 5000 数据量标签分布情况，(c)为全部数据标签分布情况，横坐标表示宽，纵坐标表示高，可以发现随着数据量的增长在左下角的数据更加密集，说明小目标的数量远超中、大目标。小目标的检测框很小，当下采样倍数为 16 时大部分的小目标仅剩几个像素值，特征图将无法有效表征小目标的位置细节信息，到 32 倍时所包含的信息量则更少。

考虑到数据集集中的小目标数量多并且浅层网络中小目标的细节信息保存较好，本文改变原来的多尺度特征融合方式，不将下采样倍数多的  $20\times20\times1024$  和  $40\times40\times512$  两种尺度的特征图输入检测头，这两种尺度特征图所包含小目标细节信息量低，对于小目标检测准确率贡献度低并且增加了计算量， $80\times80\times256$  的特征图为浅层网络的输出，细节表征能力强更适合小目标检测，最后本文使用双线性插值将该尺寸的特征图上采样至  $160\times160\times256$ ，放大小目标的特征信息进一步提高图像的分辨率。改进后的 Neck 结构如图4所示。

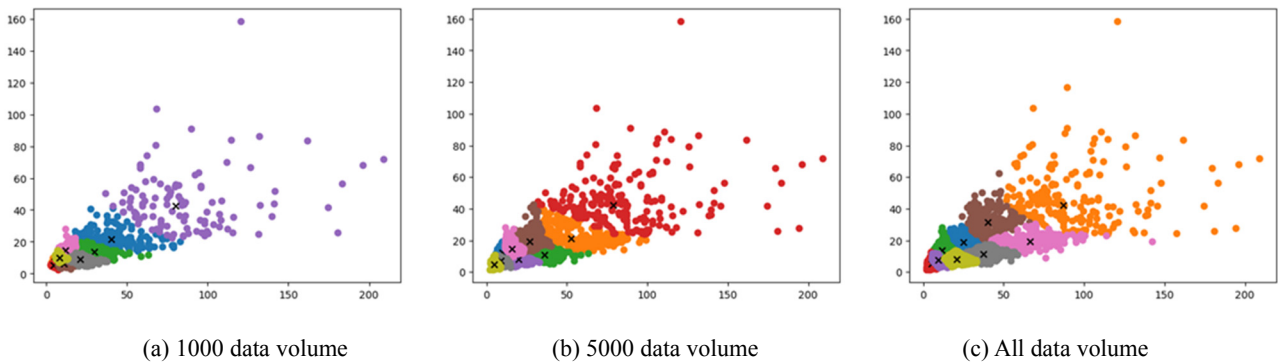


图3 数据集标签分布情况

Fig.3 Dataset label distribution

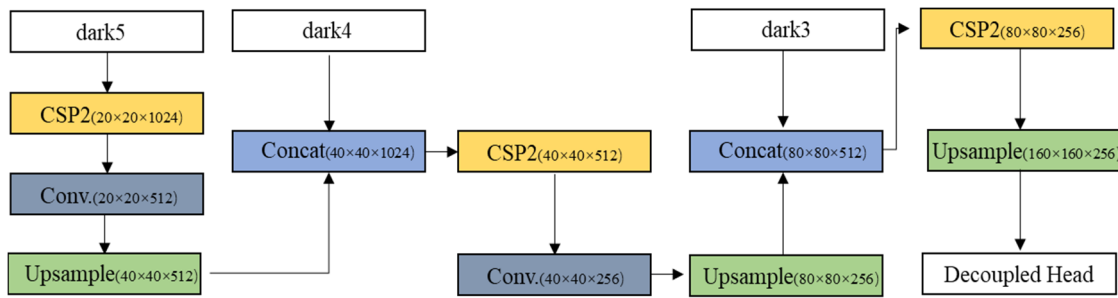


图4 改进后的 Neck 结构

Fig.4 Improved Neck structure

### 1.3 Head 检测头

Head 检测头使用的是 Decoupled Head 解耦头, 相较于传统 YOLO 的检测头 Decoupled Head 对于目标信息表达能力更强、精度更高、收敛速度更快, 其结构如图 5 所示。将 Neck 部分输出的  $160 \times 160 \times 256$  特征图输入检测头, 首先进行  $1 \times 1$  的卷积操作而后分支解耦, 分别进行两次  $3 \times 3$  的卷积操作后分开进行回归和分类任务, 并在回归任务中增加一个分支来判断其前景或是背景。

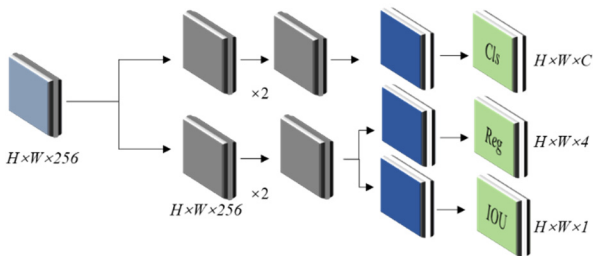


图5 解耦头结构

Fig.5 Decoupled Head structure

其中的分类分支判断检测框的类别经过激活函数 sigmoid 输出特征图的大小为  $160 \times 160 \times 10$ , 其中 10 代表的是数据集类别; 置信度分支经激活函数 sigmoid 处理后得出置信分数, 判断的是前后景的置信度大小, 其最后输出大小为  $160 \times 160 \times 1$ ; 位置分支包含检测框的中心点位置  $(x,y)$  以及框的宽  $w$  高  $h$ , 其大小为  $160 \times 160 \times 4$ 。3 个分支经 Concat 得到  $160 \times 160 \times 15$ , 然后 Reshape 得到特征为  $25600 \times 15$ , 最后经 Transpose 得到二维向量  $15 \times 25600$ , 15 包含了预测框的类别, 置信及位置信息, 25600 是预测框的数量。

接着网络需要 Anchor 规则将预测框与实际框进行关联, 在 Anchor Based 的方法中往往需要对小目标框预先进行聚类分析, 如果预设的框大于小样本则很容易造成正负样本不平衡使得模型无法有效学习到小目标的特征, 并且一个特征图中有 3 个大小不同的锚框会对网络计算造成负担。YOLOX 使用 Anchor

Free 将目标中心点作为正样本, 由原来预测的 3 个位置降为 1 个, 其计算量是 Anchor Based 方法的 1/3, 它直接预测框的宽、高和两个偏移量。

经过 Anchor Free 关联预测框与实际框后模型开始计算两者的误差, 并依次计算损失, YOLOX 的损失函数主要包括位置损失 ( $\text{Loss}_{\text{IOU}}$ )、分类损失 ( $\text{Loss}_{\text{cls}}$ ) 和目标损失 ( $\text{Loss}_{\text{obj}}$ )。位置损失表征的是模型预测框大小位置与真实框大小位置之间的吻合程度, 其计算公式如式(2)~(3):

$$\text{Loss}_{\text{IOU}} = 1 - \text{IOU}^2 \quad (2)$$

$$\text{IOU} = \frac{A \cap B}{A \cup B} \quad (3)$$

式中:  $A$  为预测框;  $B$  为真实框。

分类损失是计算目标预测类别与真实类别之间的交叉熵损失; 目标损失是表征预测框含有目标的概率, 置信度区间为  $[0,1]$  数值越大含有目标的概率越高。分类损失与目标损失均使用  $\text{BCE}_{\text{Loss}}$ , 其计算公式如式(4)~(5):

$$\text{BCE}_{\text{Loss}} = \{l_1, \dots, l_N\} \quad (4)$$

$$l_n = -[y_n \cdot \log(\sigma(x_n)) + (1 - y_n) \cdot \log(1 - \sigma(x_n))] \quad (5)$$

式中: 输入的张量为  $(x,y)$ ;  $N$  为 batch size, 每个批次预测  $n$  个标签;  $c$  为检测物品类别;  $\sigma$  表示 sigmoid 函数。

小目标检测中往往由于对目标定位不准确、聚集等问题对检测准确率产生较大影响, 为提升网络模型对小目标的感知能力加大网络对误判的惩罚, 本文对位置损失函数进行改进, 在原有的计算方法前引入小目标真实占比系数如式(6)所示:

$$\text{ratio} = 1 - \left( \frac{g_t}{640^2} \right)^k, \quad k = 1, 2, 3 \quad (6)$$

式中:  $g_t$  表示小目标真实框面积; 系数  $k$  能够控制该函数的曲率,  $k$  越大该系数越接近于 1。该系数能够表征单个小目标真实框在整张  $(640 \times 640)$  尺寸图片上的占比, 当目标真实尺寸较大时位置损失相对减少,



以降低网络因小目标尺寸不一所带来的影响提升网络的稳定性；当目标真实尺寸较小时位置损失则相对增大，增大对小目标误判的惩罚使网络对小目标更加敏感，改进后的位置损失计算公式如式(7)所示：

$$\text{Loss}_{\text{IOU}} = \text{ratio}(1 - \text{IOU}^2)$$

(7)

2 实验与结果分析

2.1 实验环境

本文实验配置如表 1 所示。

表 1 实验配置

Table 1 Experiment configuration

Configuration	Version
Operating systems	Windows10
CPU	12th Gen Intel(R) Core(TM) i7-12700KF
GPU	NVIDIA GeForce RTX 3090
Framework	Pytorch1.7.1
Software	Pycharm2020.312

2.2 数据集

实验所用数据集为目前无人机小目标检测挑战通用数据集 VisDrone2021<sup>[15]</sup>，该数据集由天津大学 AISKEYEYE 团队使用无人机采集于中国 14 个不同的城市，包括不同的环境、对象和密集度，共有 400 个由 265228 帧和 10209 幅静态图像组成的视频，有行人 (pedestrian)、汽车 (car)、自行车 (bicycle)、摩托 (motor)、三轮车 (tricycle)、遮阳篷三轮车 (awningtricycle)、卡车 (truck)、巴士 (bus)、面包车 (van) 10 类对象。该数据集采集于多个实际场景与本文的研究任务中常遇到的小目标相符，能够很好反应实际情况。本文将 10209 张图片按 8:1:1 的比例划分为训练集、验证集、测试集，。

2.3 网络参数设置及评价指标

模型训练采用迁移学习策略进行预训练，使用在数据集 Pascal voc2007 上得到的预训练权重作为初始权重，迁移学习能够提高网络模型的性能。IOU 阈值默认为 0.5，一共设置了 110 个 epoch，前 50 个 epoch 为冻结训练，batchsize 为 16，学习率设置为 0.001，衰减率为 0.92，优化器使用 Adam，后 60 个 epoch 为解冻训练，batchsize 为 8，学习率设置为 0.0001，衰减率为 0.92，优化器使用 Adam。选取平均精准率均值 (mean Average Precision, mAP)，参数量 Params，计算量 GFLOPs 以及检测速率 (Frame Per Second, FPS) 作为评价指标。

2.4 消融实验

为探究改进算法的有效性及时性，本文选取

YOLOXs 作为基线系统 Baseline，在 Neck 部分设置了 4 组改进实验分别是：N1：输入检测头的特征图大小分别为 80×80 和 40×40；N2：输入检测头的特征图大小为 80×80；N3：输入检测头的特征图大小为 160×160；N4：输入检测头的特征图大小为 80×80 和 160×160。在损失函数计算过程中小目标真实占比系数 K 分别取 1，2，3，对应的实验组为 L1,L2,L3。最后对最优的改进策略进行整合，实验组为 Ours，实验结果如表 2 所示。

表 2 消融实验结果

Table 2 Ablation experiment results

Models	mAP/%	Params	GFLOPs	FPS/(frame/s)
Baseline	35.28	9.00M	26.8	117
N1	35.39	7.75M	20.2	121
N2	35.68	6.32M	17.9	128
N3	36.21	6.35M	18.1	132
N4	36.11	8.33M	23.0	123
L1	37.00	9.00M	26.8	119
L2	37.73	9.00M	26.8	118
L3	37.38	9.00M	26.8	117
Ours	39.84	6.35M	18.1	140

由表 2 可以看出，在相同的环境配置下对检测头进行删减提高了小目标检测的 mAP 值，当删除 20×20 的检测头时实验结果 mAP 值较基线系统有所提高，参数量也随之减少，删除 40×40 的检测头时结果进一步优化，使用双线性插值扩大特征图到 160×160 的尺寸，mAP 较基线系统提高了 2.63%，参数量、计算量都大幅度下降，检测速率逐渐提升，保留 80×80 和 160×160 的检测头，实验结果较 N3 组略差；在对损失函数改进过程中，为确定小目标真实框占比，最优值设置其系数分别为 1，2，3，可以看出 3 组结果较基线系统都有所提升，当 K 取值为 2 时结果最佳，mAP 值较基线系统带来 2.1%的增益。使用以上两种最优的改进策略最终使得 Ours 组实验结果取得最佳：mAP 值提高了 4.56%，参数量减少 29.4%，运算量减少 32.5%，检测速度提升 19.7%。以上实验结果说明本文改进算法在降低参数量、计算量的情况下提高了小目标检测的效率，保证了系统的有效性和实时性，能够应用于小目标的实时检测。

2.5 比较实验

为验证本文算法的优越性，选取了其他的经典 SOTA (state-of-the-art) 模型来进行比较，包括经典的 YOLOV3-V5 系列网络，Anchor Free 算法 CenterNet，双阶段算法 Faster-RCNN 以及现阶段的一些改进算法<sup>[16-18]</sup>，主要比较检测的精度和速度，结

果如表 3 所示。

表 3 不同算法实验结果对比

Table 3 Comparison of experimental results of different algorithms

Models	mAP/(%)	FPS/(frame/s)
Vistronger Det <sup>[16]</sup>	33.85	-
TPH-YOLOv5 <sup>[17]</sup>	39.18	-
Sineglazov's <sup>[18]</sup>	37.20	-
CenterNet	22.92	81
Faster-RCNN	25.16	29
YOLOv3	27.38	61
YOLOv4	31.26	86
YOLOv5s	35.13	105
Ours	39.84	140

由表 3 可以看出本文提出算法较现阶段的 SOTA 模型在检测精度及检测速度方面具有一定的优势，较 3 种经典 YOLO 模型，mAP 值分别高了 12.46%、8.58%、4.71%，FPS 值分别高了 79 frame/s、54 frame/s、35 frame/s；较 Anchor Free 算法 CenterNet 的 mAP 和 FPS 分别高了 16.92%、59 frame/s；较双阶段算法 Faster-RCNN 的 mAP 和 FPS 分别高了 14.68%，111 frame/s；

本文算法相较于 3 种改进算法，在 mAP 指标中表现最优。

2.6 检测结果可视化

为更加直观地显示所提算法对目标检测的有效性，本文选取了数据集中 4 组具有代表性的图片进行可视化检测，结果如图 6 所示。每组图片左边为改进后算法检测效果，右边为基线系统 YOLOXs 检测效果，红框为漏检目标区域。

其中(a)组为实际道路密集车辆图片，该场景下小目标近景远景分布不均，由检测效果图可以看出所提算法对于远景小目标有较好的效果，能够有效应对小目标近景远景分布不均的问题；(b)组为实际停车场图片，车辆存在重叠和遮挡的情况，效果图中所提算法对图左被遮挡的摩托、左下角位置使用保护罩罩住的汽车、右上角重叠的面包车以及右下角骑摩托的人和车都能很好地检出；(c)组为光线环境较暗的场景，基线系统对于下方及右上角的行人存在漏检的情况，而所提算法能够适应较暗的环境将环境内的目标全部检出；(d)组显示了所提算法的局限性，由于在 Neck 网络部分删除了较小的检测头，导致网络对近景中较大目标的语义信息把握不准造成了对画面中较大卡车的漏检。

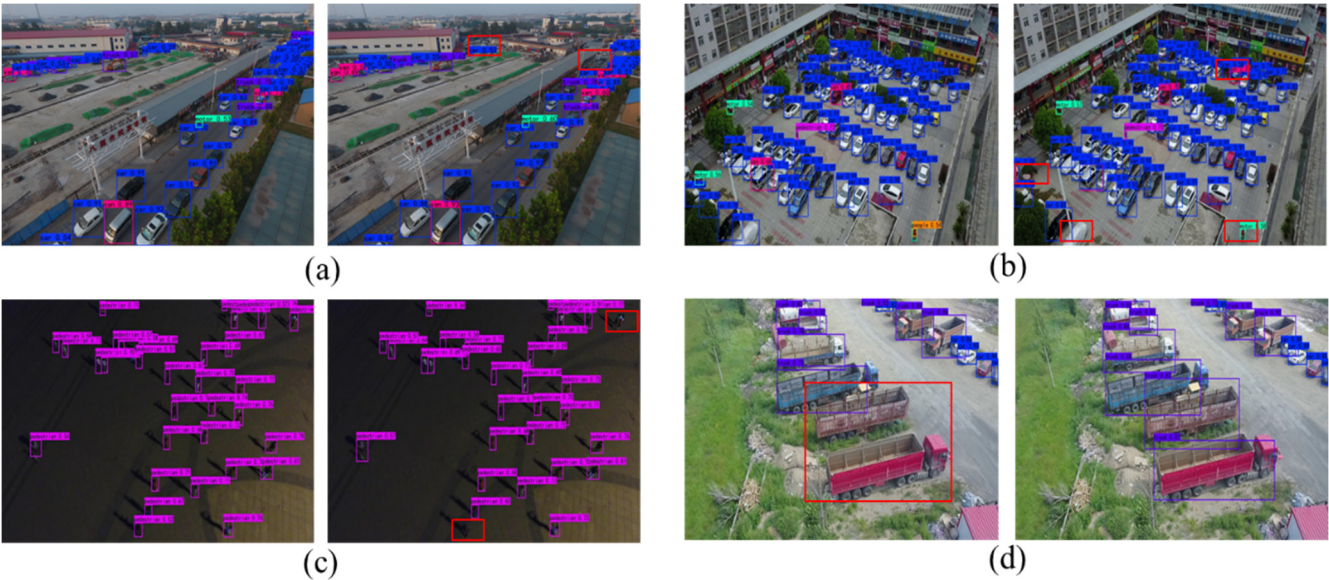


图 6 不同场景下检测效果图 (a)道路交通场景；(b)静态车辆场景；(c)暗场景；(d)近景

Fig.6 Detection renderings in different scenarios (a) Road traffic scenarios; (b) Static vehicle scenarios; (c) Dark scenes;(d) Close shot

3 结论

为应对实际环境中无人机小目标检测所遇到的问题，提出了一种基于 YOLOX 的改进算法，首先针对数据集中小目标数量多删除了尺寸较小的检测头

以提高检测精度和速度，再对回归损失函数进行改进，引入了小目标真实占比系数增大对小目标误判的惩罚。实验使用 VisDrone2021 数据集，对比了目前的主流算法，结果表明所提算法具有较高的检测精度和速度，能够很好地应对实际环境中的小目标检测，具

有有效性和实时性。最后需要指出的是所提算法由于未利用深层次的特征图,因而不能很好地对中大目标进行检测,这也是未来工作中需要改善之处。

## 参考文献:

- [1] 武连全, 邹清龙. 警用无人机执法应用现状、问题与对策[J]. 北京警察学院学报, 2021(4): 36-43. DOI:10.16478/j.cnki.jbjpc.20210507.002.
- WU L Q, ZOU Q L. Application status, problems and countermeasures of police UAV In law enforcement[J]. *Journal of Beijing Police College*, 2021(4): 36-43. DOI:10.16478/j.cnki.jbjpc.20210507.002.
- [2] TONG K, WU Y, ZHOU F. Recent advances in small object detection based on deep learning: a review[J]. *Image and Vision Computing*, 2020, **97**: 103910.
- [3] Kisantal M, Wojna Z, Murawski J, et al. Augmentation for small object detection[J/OL]. *arXiv preprint arXiv*: 1902.07296, 2019. <http://export.arxiv.org/abs/1902.07296>.
- [4] Nguyen N D, Do T, Ngo T D, et al. An evaluation of deep learning methods for small object detection[J]. *Journal of Electrical and Computer Engineering*, 2020, **2020**: 1-18.
- [5] Bharati P, Pramanik A. Deep learning techniques—R-CNN to mask R-CNN: a survey[J]. *Computational Intelligence in Pattern Recognition*, 2019, **999**: 657-668.
- [6] CHEN C, LIU M Y, Tuzel O, et al. R-CNN for small object detection[C]// *Asian Conference on Computer Vision*, 2016: 214-230.
- [7] Keles M C, Salmanoglu B, Guzel M S, et al. Evaluation of YOLO models with sliced inference for small object detection[J/OL]. *arXiv preprint arXiv*:2203.04799, 2022. <https://arxiv.org/abs/2203.04799>.
- [8] Ziming C, HAN Y, Lingjun K, et al. Multi-scene small object detection with modified YOLOv4[C]//*Journal of Physics: Conference Series*, 2022, **2253**(1): 012027.
- [9] 奉志强, 谢志军, 包正伟, 等. 基于改进 YOLOv5 的无人机实时密集小目标检测算法 [J/OL]. 航空学报: 1-15. [2022-07-01]. <http://kns.cnki.net/kcms/detail/11.1929.V.20220509.2316.010.html>.
- FENG Z Q, XIE Z J, BAO Z W, et al. Real-time dense small object detection algorithm for UAV based on improved YOLOv5 [J/OL]. *Acta Aeronautica et Astronautica Sinica*: 1-15. [2022-07-01]. <http://kns.cnki.net/kcms/detail/11.1929.V.20220509.2316.010.html>
- [10] Akyon F C, Onur Altinuc S, Temizel A. Slicing aided hyper inference and fine-tuning for small object detection[C]// *IEEE International Conference on Image Processing (ICIP)*, 2022: 966-970, Doi: 10.1109/ICIP46576.2022.9897990.
- [11] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[J/OL]. *arXiv preprint arXiv*:2107.08430, 2021. <https://doi.org/10.48550/arXiv.2107.08430>
- [12] 武连全, 楚宪腾, 杨海涛, 等. 基于改进 YOLOX 的 X 射线违禁物品检测[J]. 红外技术, 2023, **45**(4): 427-435.
- WU L Q, CHU X T, YANG H T, et al. X-ray detection of prohibited items based on improved YOLOX[J]. *Infrared Technology*, 2023, **45**(4): 427-435.
- [13] JIANG P, Ergu D, LIU F, et al. A Review of YOLO algorithm developments[J]. *Procedia Computer Science*, 2022, **199**: 1066-1073.
- [14] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020: 390-391.
- [15] ZHU P, WEN L, DU D, et al. Detection and tracking meet drones challenge[J/OL]. *arXiv preprint arXiv*: 2001.06303, 2020. <https://doi.org/10.48550/arXiv.2001.06303>
- [16] WAN J, ZHANG B, ZHAO Y, et al. Vistrongerdet: stronger visual information for object detection in VisDrone images[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 2820-2829.
- [17] 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]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 2778-2788.
- [18] Sineglazov V, Kalmykov V. Image processing from unmanned aerial vehicle using modified YOLO detector[J]. *Electronics and Control Systems*, 2021, **3**(69): 37-42.