

基于 NDM-YOLOv8 的无人机图像小目标检测

程期浩^{1,2}, 陈东方^{1,2}, 王晓峰^{1,2}

(1. 武汉科技大学 计算机科学与技术学院, 湖北 武汉 430065;

2. 智能信息处理与实时工业系统湖北省重点实验室, 湖北 武汉 430065)

摘要:针对无人机航拍图像中小目标实例多、目标之间存在遮挡的现象,容易造成漏检、误检等问题,提出一种新的基于非跨步动态多头结构的小目标检测算法(Non-strided Dynamic Multihead YOLOv8, NDM-YOLOv8)。首先,设计了 SConv 模块,融合了非跨步卷积,尽可能地保留输入数据的判别特征信息,以降低小目标特征的细粒度信息的丢失;其次,设计了 C2f-LSK 模块,通过采用选择机制对空间特征进行有效加权,动态地调整感受野,灵活地捕捉不同尺度的特征和上下文信息,提高模型对小目标的关注度;最后,设计了 P2 小目标检测头,并和高层网络进行残差连接,减少小目标特征丢失,以强化算法对小目标特征的提取能力。实验表明,NDM-YOLOv8 有效提高了对无人机图像中小目标检测精度。在公开数据集 VisDrone2019 上,NDM-YOLOv8 比 YOLOv8n 在 $mAP_{0.5}$ 提高了 5.3 个百分点, $mAP_{0.5,0.95}$ 上提高了 3.3 个百分点,对比其他模型,也取得了较优的检测效果,能够有效地完成无人机航拍图像中小目标检测任务。

关键词:无人机;小目标检测;YOLOv8;感受野;特征提取

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2024)09-0063-07

doi:10.20165/j.cnki.ISSN1673-629X.2024.0177

Small Target Detection in UAV Images Based on NDM-YOLOv8

CHENG Qi-hao^{1,2}, CHEN Dong-fang^{1,2}, WANG Xiao-feng^{1,2}

(1. School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China;

2. Hubei Provincial Key Laboratory of Intelligent Information Processing and Real-time Industrial Systems, Wuhan 430065, China)

Abstract: In view of the phenomenon that there are many instances of small targets in UAV aerial images and there are occlusions between targets, which can easily cause problems such as missed detection and false detection, a new small target detection algorithm based on a non-strided dynamic multihead structure, NDM-YOLOv8 (Non-strided Dynamic Multihead YOLOv8), is proposed. Firstly, the SConv module was designed, which integrated non-strided convolution and preserved the discriminative feature information of the input data as much as possible to reduce the loss of fine-grained information of small target features. Secondly, the C2f-LSK module was designed to effectively weight spatial features through a selection mechanism, dynamically adjust receptive fields, flexibly capture features and contextual information of different scales, and improve the model's attention to small targets. Finally, a P2 small target detection head was designed and residual connected to the high-level network to reduce the loss of small object features and enhance the algorithm's ability to extract small target features. Experiments show that the NDM-YOLOv8 effectively improves the detection accuracy of small targets in drone images. On the public dataset VisDrone2019, the NDM-YOLOv8 improved by 5.3 percentage points on $mAP_{0.5}$ and 3.3 percentage points on $mAP_{0.5,0.95}$ compared with YOLOv8n. Compared with other models, it also achieved better detection performance and can more effectively complete the task of small target detection in UAV aerial images.

Key words: unmanned aerial vehicle; small target detection; YOLOv8; receptive field; feature extraction

0 引言

近些年来,无人机航拍检测技术得到迅猛发展,在航空勘探、作物监测、海上搜救等领域得到了广泛应用。在无人机图像目标检测任务中,由于飞行高度的

限制,拍摄的图像中包含许多小尺寸目标,并且这些目标之间存在相互遮挡的情况。这种情况降低了无人机目标检测的准确度。因此,在无人机航拍视角下,设计一种小目标检测算法十分重要。

收稿日期:2023-12-21

修回日期:2024-04-23

基金项目:湖北省教育科学研究计划重点项目(D20211106)

作者简介:程期浩(2000-),男,硕士研究生,CCF会员(R9185G),研究方向为图像处理、计算机视觉;通讯作者:陈东方(1967-),男,博士,教授,研究方向为图形学与虚拟现实、机器视觉。

目前,基于深度学习的目标检测算法可以分为两类:二阶段目标检测算法和一阶段目标检测算法。二阶段目标检测算法,如 SPP-Net^[1]、R-CNN^[2-3] 系列等,通过生成一系列候选框作为样本,然后利用卷积神经网络对这些样本进行分类。一阶段目标检测算法,如 SSD^[4]、YOLO^[5-7] 系列等,不需要产生候选框,而是直接将目标框的定位问题转化为回归问题进行处理。由于其网络结构复杂且需要多个阶段的处理,二阶段目标检测算法在计算量上具有较高的复杂性,需要更多的计算资源和时间来完成检测任务。一阶段目标检测算法的模型结构简单,能够在计算资源有限的情况下有效地进行实时目标检测,且它在检测精度和计算效率之间拥有较好的平衡性,符合该文针对无人机图像小目标检测的需求。

为了提高无人机图像小目标的检测精度,许多研究人员提出了一些有效方法。Zhang 等人^[8] 在 YOLOv8 主干中,采用 RepVGG (Re-parameterization VGG)^[9] 模块作为下采样层,强化了网络学习多尺度特征的能力,从而有效提升了对无人机小目标检测精度。Zhong 等人^[10] 在 YOLOv8 中引入 SE (Squeeze and Excitation) 注意力机制,能够自适应地提高模型从小目标中提取特征的能力,并且用 GSConv (Group Shuffle Convolution)^[11] 模块代替普通卷积,提高了模型对无人机目标的检测能力。Pan 等人^[12] 在 YOLOv8 中引入可分离卷积思想,改进空间金字塔池化层,融合不同层级的特征,并且改进 MPDIoU (Minimum Point Distance Based IoU)^[13] 来代替原来的损失函数,增强了模型对小目标的特征提取能力。Hao 等人^[14] 在 YOLOv7 骨干网络中引入 CBAM (Convolutional Block Attention Module)^[15] 注意力机制并且使用 FReLU 激活函数替代 SiLU 激活函数,有效地降低了原始算法对小目标的误检漏检率。Zhang 等人^[16] 在 YOLOv7-tiny 中使用 ConvMixer^[17] 检测头,通过其中的深度卷积和逐点卷积来获得特征信息的空间和通道关系,提升对小目标的检测能力。Pan 等人^[18] 在 YOLOX 中融合了一种通道-空间注意力机制模块 (Channel-Spatial Attention Module, CSA),通过最优的权重分配使得网络关注特征图中小目标密集的区域,并且使用一种带位置引导的标签分配策略,提高整体预测框的质量,从而强化算法对小目标的特征提取能力。

综上所述,前人的工作虽然有效地提高了对无人机小目标的检测能力,但是仍然存在漏检、误检等问题。为了解决上述问题,该文提出了一种基于改进 YOLOv8 的无人机图像小目标检测算法 NDM-YOLOv8。改进工作如下:(1)设计 SConv 模块,减少小目标细粒度的损失,避免了低像素导致的错误或缺

失检测。(2)设计 C2f-LSK 模块,动态调整目标的感受野,更好地捕捉输入图像的上下文特征信息。(3)增加 P2 小目标检测层,与其他三个检测层相结合捕获更多小目标特征。

1 相关工作

1.1 小目标检测

当前对于小目标的定义分为基于相对尺度与基于绝对尺度两类。基于相对尺度定义,即从目标与图像的相对比例这一角度考虑,如:Chen 等人^[19] 对小目标定义为同一类别中所有目标实例的相对面积,即边界框的面积与图像面积之比的中位数在 0.08% ~ 0.58% 之间。基于绝对尺度定义,则从目标绝对像素大小这一角度考虑,如:MS COCO 数据集将小目标定义为分辨率小于 32×32 像素的目标。

针对小目标检测的难点,Cubuk 等人^[20] 提出了 AutoAugment 方法,自动地为数据集找到数据增强策略,提高算法对小目标的检测精度。Qiao 等人^[21] 提出了递归特征金字塔 (Recursive Feature Pyramid, RFP) 方法,利用高层网络提取的语义信息和低层网络提取的细节信息来预测多尺度的目标,提升模型对小目标的检测效果。Haris 等人^[22] 提出一种端到端训练的超分辨率方法,对分辨率低的区域进行超分辨率的处理,提升了模型对小目标检测的性能。

1.2 YOLOv8 概述

YOLOv8 作为一阶段目标检测算法,在检测精度和速度上具有较好的平衡性。有 n、s、m、l、x 五个版本,其网络模型大致相同,不同在其网络深度和宽度。其中 YOLOv8n 网络的深度和宽度最小。考虑到无人机实际应用场景,由于 YOLOv8n 的网络层次最为简单,所需计算资源最少,运行速度最快,具有最佳的移植性,该文选择对 YOLOv8n 模型进行进一步改进。YOLOv8n 模型的网络结构包括输入端 (Input)、骨干网络 (Backbone)、颈部网络 (Neck) 以及头部网络 (Head)。

(1)输入端:对输入图像进行 Mosaic 数据增强来丰富数据集;此外,采用了无锚框 (Anchor Free) 策略来减少预测框的数量,从而加速了非极大值抑制 (Non-Maximum Suppression, NMS) 过程。

(2)骨干网络:包含 Conv、C2f、SPPF (Spatial Pyramid Pooling-Fast) 等模块。Conv 模块负责对输入图像进行卷积、批量归一化以及 SiLU 激活等操作;C2f 模块能在轻量化的同时获得更为丰富的梯度流信息;SPPF 模块能在不同尺度下对图像进行特征提取和编码。

(3)颈部网络:包括特征金字塔网络 (Feature

Pyramid Network, FPN) 和路径聚合网络 (Path Aggregation Network, PAN)。FPN 通过自顶向下的传递方式强化语义特征, PAN 通过自底向上的传递方式强化位置特征。FPN 与 PAN 相结合使得不同阶段的特征图进行了有效的特征融合。

(4)头部网络:采用解耦头 (Decoupled Head) 策略,将分类头和检测头分开。根据三个不同尺度的特征图来获取目标的类别和位置信息。

2 NDM-YOLOv8 算法

该文以 YOLOv8n 为基线模型设计了改进后的无人机图像小目标检测算法 NDM-YOLOv8,具体工作为:(1)由于神经网络在卷积和池化操作过程中会导致一定程度的细粒度信息丢失,为使得模型降低无人机图像中小目标的误检及漏检率,通过在骨干网络中设计 SConv 模块代替原有的 Conv 模块;(2)考虑到无人机图像具有背景复杂、视角独特以及检测目标尺寸较小的特点,为了更有效地调整感受野,获得不同尺度的上下文特征信息,在颈部网络中设计 C2f-LSK 模块代替原有的 C2f 模块;(3)考虑到小尺寸目标需要较浅的特征图,在头部网络中引入一个额外的 P2 小目标检测头,以更有效地捕捉细节信息和局部特征。具体结构如图 1 所示。

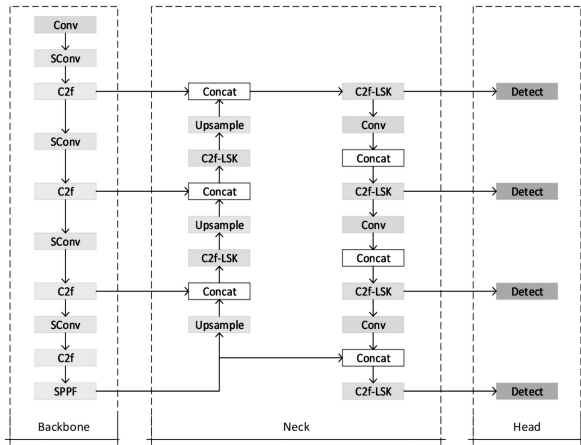


图 1 NDM-YOLOv8 网络结构

2.1 SConv 模块

YOLOv8 骨干网络中多次应用 Conv 层过滤冗余信息,从而造成无人机航拍图像中丢失小目标的细粒度信息。为了解决上述问题,该文在 SUNKARA 等人^[23]提出的 SPD-Conv 模块的基础上重新构建了 YOLOv8 中 Conv 网络结构,设计了 SConv 模块。

如图 2 所示,SPD-Conv 由一个空间到深度 (space-to-space, SPD) 层和一个非跨步卷积 (Non-Strided Convolution, Conv) 层组成。在 SPD 层,部分假定原始特征图 X 的尺寸为 $s \times s \times c_1$,其中 s 是长和宽, c_1 是通

道数。通过切片操作得到 4 个子图,尺寸都是 $s/2 \times s/2 \times c_1$ 。每个子图都按照比例因子为 2 的比例对特征图 X 下采样。将这 4 张子图进行在通道维度上的拼接,得到尺寸为 $s/2 \times s/2 \times 4c_1$ 的特征图 X' 。在 Conv 层,为了最大程度地保留其特征信息,采用了一个无卷积步长层,步长设置为 1,特征图 X' 尺寸变为 $s/2 \times s/2 \times c_2$,其中 $c_2 < 4c_1$ 。

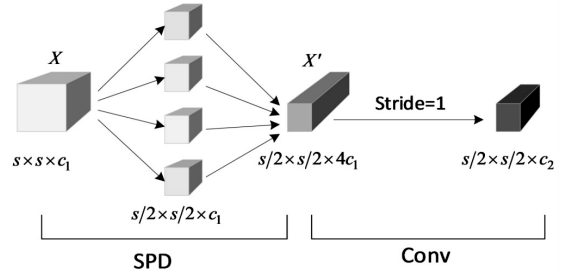


图 2 SPD-Conv 模块结构

图 3 是重新构建的 SConv 模块。该文把骨干网络 Conv 模块中的 Conv2d 层替换为 SPD-Conv,应用其 SPD 层的切片分子图原理,将小目标全局空间信息映射到通道维度,以达到保留无人机航拍场景下更多小目标特征信息的目的,并且在 SPD 层后,进行了非跨步卷积操作,对通道维度进行调整。

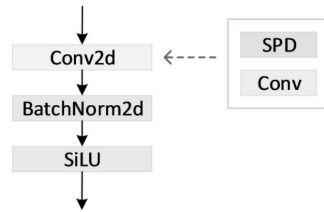


图 3 SConv 模块结构

2.2 C2f-LSK 模块

无人机航拍图像通常是以鸟瞰视角拍摄的,图像中的大多数物体是小尺寸的。当感受野较小时,有可能无法有效地捕捉到小目标的细节信息,从而降低了检测性能。为了解决上述问题,该文在 Li 等人^[24]提出的 LSKNet (Large Selective Kernel Network) 模块的基础上,对 YOLOv8 中 C2f 模块进行了重新构建,设计了 C2f-LSK 模块。这一设计旨在调整目标的感受野,以更有效地检测到无人机航拍图像中的小目标。

如图 4 所示,每个 LSKNet 块由大核选择 (LK Selection) 子块和前馈网络 (FFN) 子块组成。LK Selection 子块动态地调整目标的感受野。FFN 子块被应用于通道混合和特征细化,包括全连接层、深度卷积层和 GELU 激活函数。LSK Module 子块作为核心模块被嵌入到 LK Selection 子块中,它包括全连接层、一系列的大卷积核和空间内核选择机制。在空间内核选择机制中,对一系列的大卷积核所处理的特征进行有效加权,随后在空间上将它们合并。这些核的权重是

根据输入动态确定的,使得模型能够自适应地选择不同的卷积核,并根据需要调整每个目标在空间中的感受野。

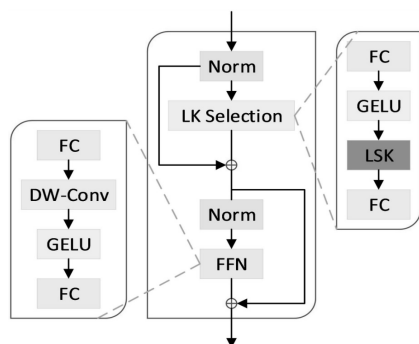


图 4 LSKNet 模块结构

图 5 中展示了根据 LSKNet 模块构建的 Bottleneck-LSK,由两层 LSKNet 模块堆叠而成。图 6 是 C2f-LSK 模块,使用新构建的 Bottleneck-LSK 模块代替 C2f 模块中的 Bottleneck 模块,由 CBS 层、分离层和 Bottleneck-LSK 层组成。这一改进动态地调整了目标的感受野,有效地捕获了不同尺度的上下文特征信息。



图 5 Bottleneck-LSK 模块结构

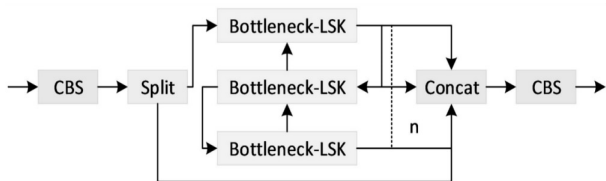


图 6 C2f-LSK 模块结构

2.3 P2 小目标检测层

当输入图像分辨率为 640×640 时,原始的 YOLOv8 算法对输入图像分别进行了 8 倍,16 倍和 32 倍的下采样,生成 $80 \times 80, 40 \times 40, 20 \times 20$ 的预测特征图,用于检测 $8 \times 8, 16 \times 16, 32 \times 32$ 以上的目标。然而,在无人机航拍图像中,小目标的尺寸较小,目标之间存在遮挡并且 YOLOv8 多次进行下采样过程,特征图中的目标信息变得非常稀少,较深的特征图比较难学习到小目标的特征信息,导致 YOLOv8 原模型对小目标的检测精度较低。

因此,该文在保持其他特征图的尺度不变的情况下,在网络原有的三个检测层的基础上新引入了一个 160×160 的 P2 小目标检测层。该层的高分辨率预测特征图中保留了更多关于小目标的位置信息和更丰富的细节特征,可以更好地获取小目标的类别和位置信息。首先,将骨干网络中的第七层的 80×80 尺度特征层与颈部网络中的上采样特征层进行堆叠。通过上采样操作,可以将低分辨率的特征图增加到与高分辨率

特征图相同的尺度,从而保留小目标的细节信息。接下来,将堆叠后的特征层经过 C2f-LSK 层和上采样处理,得到深层语义特征层。C2f-LSK 操作是调整目标特征的感受野以及对通道维度进行融合,以整合不同层级的特征表示。通过插值或反卷积操作进行上采样,将特征图的尺度恢复到原始大小,以更深入地提取小目标的语义信息。然后,将深层语义特征层与骨干网络中的第三层的浅层位置特征层进行叠加。这样可以完善 160×160 尺度的融合特征层,以更好获得小目标的语义特征和位置信息。最后,将融合特征层再次经过 C2f-LSK 层送入到头部网络中的一个额外的解耦头。

3 实验与分析

3.1 数据集

该文采用公开数据集 VisDrone2019,该数据集由天津大学的 AISKYEYE 团队收集。由各种无人机摄像头捕获,是在不同天气、位置以及光照条件下收集的。数据集由 288 个视频片段、261 908 帧和 10 209 幅静态图像组成,其中训练集 6 471 张、验证集 548 张、测试集 3 190 张。数据集图像类别分别为行人、人、自行车、汽车、面包车、卡车、三轮车、遮阳篷三轮车、公共汽车和摩托车这十类。数据集中待检测的目标数量多并且大多数是以小目标情况存在,符合实验要求。

3.2 实验环境与参数

实验是基于 Ubuntu20.04 操作系统,实验环境为 PyTorch2.0.0、Python3.8、CUDA11.8。CPU 是 12 vCPU Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10 GHz,GPU 是 NVIDIA GeForce RTX 4090,24 GB 显存。相关参数设置:训练轮数为 200 次;批量大小为 16;输入分辨率为 640×640 ;初始学习率为 0.01;余弦退火参数为 0.01;优化器为 SGD;动量为 0.937;权重衰减系数为 0.005。

3.3 模型评价指标

所用到的评价指标包括精准率 (Precision)、召回率 (Recall)、平均精度均值 (mAP)、模型参数量 (Params)、浮点运算次数 (GFLOPs)。

(1) 精准率 (Precision): 表示在所有被模型预测为正例的样本中真正例的比例。公式如下:

$$P = \frac{TP}{TP + FP} \quad (1)$$

(2) 召回率 (Recall): 表示在所有实际正例样本中,模型成功预测为正例的比例。公式如下:

$$R = \frac{TP}{TP + FN} \quad (2)$$

(3) 平均精度均值 (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。公式如下:

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \quad (3)$$

$$AP = \int_0^1 P(R) dR \quad (4)$$

(4) 模型参数量 (Params): 通常用于评估模型的复杂度。较多的参数量可能意味着模型具有更强的表示能力, 可以更好地适应训练数据, 但也可能增加过拟合的风险。相反, 较少的参数量可能导致模型的表示

能力不足, 难以捕捉数据中的复杂关系。

(5) 浮点运算次数 (GFLOPs): 用于衡量模型在推理或训练过程中执行的浮点运算的总数量。它是计算模型复杂度的指标之一, 用于评估模型的计算资源需求和效率。

3.4 消融实验

为了证明提出的每个改进模块对无人机图像小目标检测效果的有效提升, 做消融实验进行评估。实验结果如表 1 所示。

表 1 消融实验

Method	SConv	C2f-LSK	P2	$mAP_{0.5}/\%$	$mAP_{0.5,0.9}/\%$	Params/MB	GFLOPs/G
基线				26.6	0.145	3.01	8.20
改进 1	√			29.1	0.161	3.27	11.7
改进 2		√		27.4	0.151	3.18	8.50
改进 3			√	29.5	0.163	2.93	12.4
改进 4	√	√		29.4	0.163	3.44	12.0
改进 5	√		√	31.5	0.176	3.19	15.9
改进 6		√	√	30.1	0.167	3.10	12.9
改进 7	√	√	√	31.9	0.178	3.36	16.4

从表 1 消融实验结果可以看出, 该文提出的改进方法对模型检测小目标的精度均得到一定程度的提升。改进 1 在骨干网络中设计 SConv 模块来替代原有的 Conv 模块, 减少小目标细粒度信息的丢失和降低学习效率较低的特征表示, $mAP_{0.5}$ 提高了 2.5 百分点, 参数量和计算量较少提升在可接受范围内, 验证了改进 1 的有效性。改进 2 在颈部网络中设计 C2f-LSK 模块来代替原有的 C2f 模块, 动态调整小目标的感受野, 捕获更多的上下文特征信息, $mAP_{0.5}$ 提高了 0.8 百分点, 参数量和浮点运算数小幅度提升, 验证了改进 2 的有效性。改进 3 在头部网络中添加一个 P2 小目标检测头, 有效地获取了小目标的位置以及类别信息, $mAP_{0.5}$ 提升了 2.9 百分点, 模型的复杂度与计算量增加在可接受范围内, 验证了改进 3 的有效性。改进 4 是在改进 1 的基础上引入 C2f-LSK 模块, 检测精度有所提升, 验证了改进 4 的有效性。改进 5 是在改进 1 基础上添加了 P2 小目标检测头, 检测精度也有所提升。改进 6 是在改进 2 基础上添加了 P2 小目标检测头, 精度相比改进 2 提升 4.1 百分点。改进 7 是融合该文提出的三个改进方法, 精度较基线有所提升。综上所述, 综合三个改进后的模型 NDM-YOLOv8 与基准模型 YOLOv8n 对比, 在模型参数量和浮点运算数较少的提高的同时, 其 $mAP_{0.5}$ 和 $mAP_{0.5,0.95}$ 分别提高了 5.3 百分点和 3.3 百分点, 证明了改进后的模型能够有效完成对无人机航拍图像小目标的检测任务。

3.5 对比实验

为进一步证明 NDM-YOLOv8 算法的有效性, 与当前主流算法 YOLOv5、YOLOv7-tiny、YOLOv8、Drone-YOLO^[8]、LW-YOLOv8^[10]、TPH-YOLOv5s^[25] 在公开数据集 VisDrone2019 上进行对比实验, 实验结果如表 2 所示。

表 2 对比实验

Method	$mAP_{0.5}/\%$	Params/MB	GFLOPs/G
YOLOv5n	24.6	1.78	4.30
YOLOv5s	25.7	7.05	16.0
YOLOv7-tiny	29.5	6.04	13.3
YOLOv8n	26.6	3.01	8.20
YOLOv8s	31.5	11.14	18.7
Drone-YOLO	31.0	3.05	10.7
LW-YOLOv8	30.8	2.83	9.20
TPH-YOLOv5s	28.3	7.47	37.0
文中算法	31.9	3.36	12.4

从表 2 可以看出, YOLOv5、YOLOv7-tiny 和 YOLOv8 在检测性能上都明显低于文中算法。Drone-YOLO^[8] 与 LW-YOLOv8^[10] 的 $mAP_{0.5}$ 比文中算法分别低 0.9 百分点和 1.1 百分点, 并且参数量和计算量与文中算法接近, 证明了文中算法性能更好。TPH-YOLOv5s^[25] 的 $mAP_{0.5}$ 低于文中算法, 参数量和计算量高于文中算法, 证明了文中算法性能更好。

综上所述,NDM-YOLOv8 算法虽然在参数量和计算量方面有较小幅度的增长,但是在可接受范围内,能够较好应用于无人机的实时场景。并且 NDM-YOLOv8 在检测精度方面较其他算法有所优势, $mAP_{0.5}$ 达到了 31.9%。因此进一步证明了该算法的有效性。

3.6 可视化分析

为了验证 NDM-YOLOv8 算法在实际场景的有效性,在 VisDrone2019 测试集里面选取了密集、黑暗以及高空场景下代表性图片,图片中目标都属于小目标,检测效果如图 7 所示。图 7(a)是密集场景下检测效

果图,可以看出图中存在大量行人目标且彼此之间重叠,相互遮挡,但是文中算法仍然检测出行人这一目标。图 7(b)是黑暗场景下检测效果图,可以看出图中光线昏暗,细节难以分辨,目标缺乏明显的视觉特征,但是文中算法仍然全面检测到图中目标。图 7(c)是高空场景下检测效果图,图中目标位于较远的距离,目标尺寸较小,但是文中算法仍然检测到了图中汽车、行人以及摩托车等目标。通过图 7 的检测结果可知,文中算法具有较好的鲁棒性,可以有效地应用于实际场景。

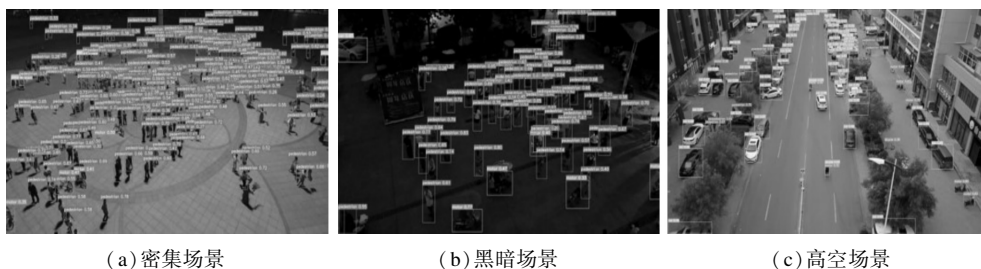


图 7 不同场景下的检测效果

为了验证 NDM-YOLOv8 算法的有效性,随机从 VisDrone2019 测试集中选取部分图片,图片中目标都符合小目标定义。将基线算法 YOLOv8 与文中算法进行对比,检测效果如图 8 所示。图 8(a)是基线算法的检测图,图 8(b)是文中算法的检测图。为了易于发现改进前后的差异,将不同的地方用矩形框标注。从上半部分可以看出,由于目标尺寸较小、目标之间相互遮挡,基线算法漏检了大量汽车以及行人目标,而文中算法有效地检测到了汽车与行人目标。但是由于背景

过于复杂、目标之间过于密集遮挡,导致漏检了极少量的汽车目标,文中算法明显降低了漏检、误检率。从下半部分可以看出,由于树木遮挡原因,基线算法将路灯误检为汽车,而文中算法没有误检。同时基线算法漏检了行人以及三轮车目标,而文中算法能够有效地检测。通过图 8 的对比检测图可知,NDM-YOLOv8 与 YOLOv8 相比,减少了漏检、误检情况的发生,提高了模型对小目标的检测精度。由此可知,证明了该算法的有效性。

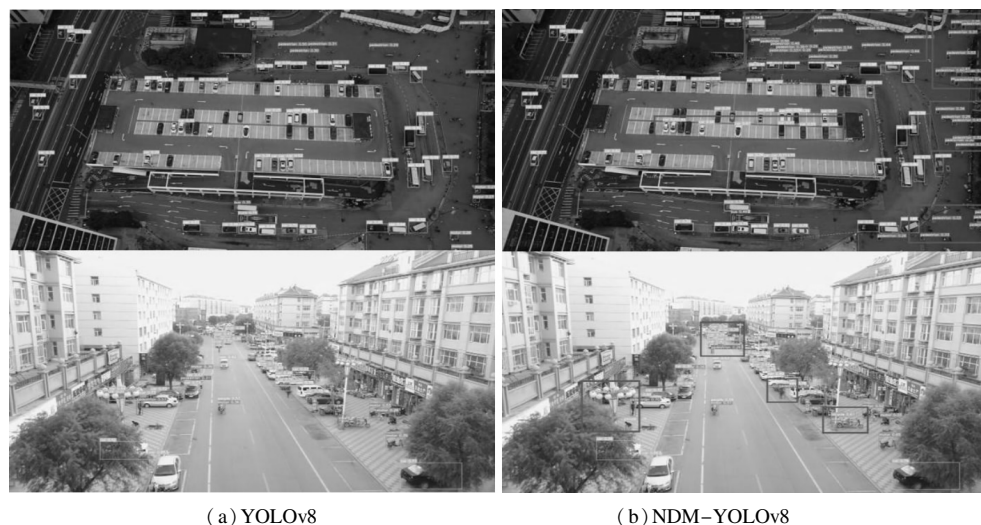


图 8 改进前后检测效果对比

4 结束语

针对无人机图像中小目标多,目标之间存在遮挡,容易出现漏检、误检等问题,提出了一种基于改进

YOLOv8 的无人机图像小目标检测算法 NDM-YOLOv8,主要工作和结论如下:

首先,在骨干网络中设计 SConv 模块代替 Conv 模块,使用 SPD 层和非跨步卷积层来减少小目标细粒

度信息的丢失,降低漏检、误检情况。其次,在颈部网络中设计 C2f-LSK 模块来代替 C2f 模块,动态地调整目标的感受野,获取到丰富的上下文特征信息。最后,在头部网络中增加较浅的 P2 小目标检测头,在特征提取阶段更有效地获取小目标的细节信息,以增强模型对小目标的检测性能。

在 VisDrone2019 数据集上的结果表明,NDM-YOLOv8 的 $mAP_{0.5}$ 的值为 31.9%,与基线模型相比提高了 5.3 个百分点。对比其它的目标检测算法也取得了较好的检测精度,参数量与计算量较小,能满足无人机航拍实时检测的需求。接下来将对降低模型参数量和计算量方面进行进一步研究,使模型更加轻量化。

参考文献:

- [1] PURKAIT P, ZHAO C, ZACH C. SPP-Net: deep absolute pose regression with synthetic views [J]. arXiv: 1712.03452, 2017.
- [2] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE international conference on computer vision. Santiago: IEEE, 2015: 1440-1448.
- [3] REN S, HE K, 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.
- [4] LIU W, ANGUOLOV D, ERHAND, et al. SSD: single shot multibox detector [C]//Proc of IEEE conference on computer vision and pattern recognition. Berlin: Springer, 2016: 21-37.
- [5] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. arXiv: 1804.02767, 2018.
- [6] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: optimal speed and accuracy of object detection [J]. arXiv: 2004.10934, 2020.
- [7] LI C Y, LI L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications [J]. arXiv: 2209.02976, 2022.
- [8] ZHANG Z. Drone-YOLO: an efficient neural network method for target detection in drone images [J]. Drones, 2023, 7(8): 526-549.
- [9] DING X, ZHANG X, MA N, et al. Repvgg: making vgg-style convnets great again [C]//2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Nashville: IEEE, 2021: 13733-13742.
- [10] ZHONG M, HUANG F, LI S. Lightweight you only look once v8: an upgraded you only look once v8 algorithm for small object identification in unmanned aerial vehicle images [J]. Applied Sciences, 2023, 13(22): 12369-12382.
- [11] LI H, LI J, WEI H, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles [J]. arXiv: 2206.02424, 2022.
- [12] 潘 玮, 韦 超, 钱春雨, 等. 面向无人机视角下小目标检测的 YOLOv8s 改进模型 [J]. 计算机工程与应用: 1-10 [2024-03-27]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240308.1652.008.html>.
- [13] VMA S, XU Y. MPDIoU: a loss for efficient and accurate bounding box regression. [J]. arXiv: 2307.07662, 2023.
- [14] 郝紫霄, 王 琦. 基于 YOLO-v7 的无人机航拍图像小目标检测改进算法 [J]. 软件导刊, 2024, 23(1): 167-172.
- [15] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module [C]//Proceedings of the European conference on computer vision. Berlin: Springer, 2018: 3-19.
- [16] 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 [J]. 工程科学与技术: 1-14 [2024-03-27]. <https://doi.org/10.15961/j.jsuese.202300593>.
- [17] TROCKMAN A, KOLTER J Z. Patches are all you need? [J]. arXiv: 2201.09792, 2022.
- [18] 潘 翔, 陈前斌, 黄 昂, 等. 基于改进 YOLOX 的无人机航拍图像小目标检测算法 [J]. 南京邮电大学学报: 自然科学版, 2024, 44(1): 90-100.
- [19] CHEN C, LIU M Y, TUZEL O, et al. R-CNN for small object detection [C]//Asian conference on computer vision (ACCV). Switzerland: Springer, 2016: 10115.
- [20] CUBUK E D, ZOPH B, MANE D, et al. Autoaugment: learning augmentation policies from data [J]. arXiv: 1805.09501, 2018.
- [21] QIAO S, CHEN L, YUILLE A. Detectors: detecting objects with recursive feature pyramid and switchable atrous convolution [C]//2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Nashville: IEEE, 2021: 10213-10224.
- [22] HARIS M, SHAKHAROVICH G, UKITA N. Task-driven super resolution: object detection in low-resolution images [C]//Neural information processing: 28th international conference. Switzerland: Springer, 2021: 387-395.
- [23] SUNKARA R, LUO T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects [J]. arXiv: 2208.03641, 2022.
- [24] LI Y X, HOU Q B, ZHENG Z H, et al. Large selective kernel network for remote sensing object detection [J]. arXiv: 2303.09030, 2023.
- [25] ZHU X K, LYU S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]//2021 IEEE/CVF international conference on computer vision workshops. Montreal: IEEE, 2021: 2778-2788.