



基于改进的Yolov5的无人机图像小目标检测

何宇豪, 易明发, 周先存, 王冠凌

引用本文:

何宇豪, 易明发, 周先存, 王冠凌. 基于改进的Yolov5的无人机图像小目标检测[J]. 智能系统学报, 2024, 19(3): 635–645.

HE Yuhao, YI Mingfa, ZHOU Xiancun, et al. UAV image small-target detection based on improved Yolov5[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(3): 635–645.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202210032>

您可能感兴趣的其他文章

基于改进的Faster RCNN面部表情检测算法

Facial expression recognition based on improved Faster RCNN

智能系统学报. 2021, 16(2): 210–217 <https://dx.doi.org/10.11992/tis.201910020>

隔级融合特征金字塔与CornerNet相结合的小目标检测

Small target detection based on a combination of feature pyramid and CornerNet

智能系统学报. 2021, 16(1): 108–116 <https://dx.doi.org/10.11992/tis.202004033>

双向特征融合与注意力机制结合的目标检测

Target detection based on bidirectional feature fusion and an attention mechanism

智能系统学报. 2021, 16(6): 1098–1105 <https://dx.doi.org/10.11992/tis.202012029>

基于改进FCOS的拥挤行人检测算法

Crowded pedestrian detection algorithm based on improved FCOS

智能系统学报. 2021, 16(4): 811–818 <https://dx.doi.org/10.11992/tis.202010012>

基于跳跃连接金字塔模型的小目标检测

Skip feature pyramid network with a global receptive field for small object detection

智能系统学报. 2019, 14(6): 1144–1151 <https://dx.doi.org/10.11992/tis.201905041>

多层卷积特征的真实场景下行人检测研究

Research on pedestrian detection based on multi-layer convolution feature in real scene

智能系统学报. 2019, 14(2): 306–315 <https://dx.doi.org/10.11992/tis.201710019>

DOI: 10.11992/tis.202210032

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20231204.1028.002>

基于改进的 YOLOv5 的无人机图像小目标检测

何宇豪¹, 易明发¹, 周先存², 王冠凌¹

(1. 安徽工程大学 电气工程学院, 安徽 芜湖 241000; 2. 皖西学院 电子与信息工程学院, 安徽 六安 237012)

摘要: 为了解决无人机航拍图像小目标检测算法检测速度与精度无法兼顾的问题, 在 YOLOv5 的基础上, 提出了针对于无人机图像小目标检测的 YOLOv5_GBCS 算法。在新的算法中, 添加一个额外的检测头, 以便增强对小目标的特征融合效果; 在主干网络中分别采用 GhostConv 卷积模块、GhostBottleneckC3 模块替换部分 Conv 模块和 C3 模块用以提取丰富特征和冗余特征以提高模型效率; 引入加权双向特征金字塔网络 (bidirectional feature pyramid network, BiFPN) 结构, 用以提高对小目标的检测精度; 在主干网络和颈部网络中引入轻量化的卷积块注意力模块 (convolutional block attention module, CBAM), 关注重要特征并抑制不必要的特征, 增强小目标特征表达能力; 使用 Soft-NMS 算法来替换 NMS, 因此降低了小目标在密集场景下的漏检率。通过在 VisDrone2019 数据集上的实验结果表明, 集成了所有改进的方法后的 YOLOv5_GBCS 算法, 不仅提高了检测精度, 而且有效地提高了检测速度, 模型的 mAP 从 38.5% 提高到 43.2%, 检测速度也从 53 f/s 提高到 59 f/s。YOLOv5_GBCS 算法可以有效地实现无人机航拍图像中小目标识别。

关键词: 图像处理; GhostConv 卷积模块; 双向特征金字塔网络; 卷积块注意力模块; Soft 双向特征金字塔网络; 轻量化模型; 小目标检测; VisDrone 数据集

中图分类号: TP391.4; V19 文献标志码: A 文章编号: 1673-4785(2024)03-0635-11

中文引用格式: 何宇豪, 易明发, 周先存, 等. 基于改进的 YOLOv5 的无人机图像小目标检测 [J]. 智能系统学报, 2024, 19(3): 635-645.

英文引用格式: HE Yuhao, YI Mingfa, ZHOU Xiancun, et al. UAV image small-target detection based on improved YOLOv5[J]. CAAI transactions on intelligent systems, 2024, 19(3): 635-645.

UAV image small-target detection based on improved YOLOv5

HE Yuhao¹, YI Mingfa¹, ZHOU Xiancun², WANG Guanling¹

(1. College of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China; 2. The Department of Electronic and Information Engineering, West Anhui University, Lu'an 237012, China)

Abstract: The detection speed and accuracy of small targets captured by UAV aerial photography cannot be considered at the same time. To address this problem, a new algorithm based on the YOLOv5 algorithm, namely YOLOv5_GBCS, is proposed for small-target detection of UAV-captured images. In the new algorithm, an additional detector head is added to enhance the feature fusion effect of small targets. In the backbone network, the GhostConv convolution module and GhostBottleneckC3 module are used to replace some original Conv modules and C3 modules for extracting rich and redundant features, respectively, to improve the model efficiency. The weighted bidirectional feature pyramid network structure is introduced to enhance the detection accuracy of small targets. The lightweight convolutional block attention module is introduced into the backbone and neck networks to focus on important features and suppress unnecessary features to boost the ability of small-target feature expression. The Soft-NMS algorithm is used to replace the NMS for reducing the miss detection rate of small targets in dense scenes. Experimental results on the VisDrone 2019 dataset show that the YOLOv5_GBCS algorithm integrating all improved methods enhances the detection accuracy and effectively improves the detection speed. The mAP of the subject model has been increased from 38.5% to 43.2%, and the detection speed from 53 f/s to 59 f/s. Therefore, the YOLOv5_GBCS algorithm can effectively recognize small targets in the image captured by UAV aerial photography.

Keywords: image processing; GhostConv convolution module; bidirectional feature pyramid network; convolutional block attention module; soft bidirectional feature pyramid network; lightweight model; small-target detection; VisDrone dataset

收稿日期: 2022-10-25. 网络出版日期: 2023-12-04.

基金项目: 国家自然科学基金项目 (61572366); 安徽高校自然科学研究重大项目 (J2021ZD0116); 皖江高端装备制造协同创新中心开放基金项目 (GCKJ2018013).

通信作者: 王冠凌. E-mail: ahpu_405@163.com.

©《智能系统学报》编辑部版权所有

随着网络技术的快速发展, 无人机图像目标检测技术具有广泛的应用范围, 包括森林防火、农业信息、电力线检测、道路和桥梁损伤评估等。相较于自然环境视角下拍摄的图像, 无人机

航拍图像具有巨大的高空视角优势。但无人机拍摄图像却存在背景复杂度高、目标尺寸小、外观模糊等缺陷。同时,基于传统方法的目标检测技术对小目标检测精度低,对复杂背景下目标的适应性弱,容易发生漏检和误检。随着深度学习的发展,基于卷积神经网络的目标检测方法近年来迅速发展起来^[1],在检测精度和速度方面都超过了传统的目标检测方法,然而当前大多数目标检测模型仅仅适用于处理自然环境图像,如果采用同样的算法用于处理无人机航拍图像就会存在一些问题。首先,目前大多数目标检测模型无法满足实时检测需求,关键是在提高检测精度的同时减少计算量和参数量;其次,无人机航拍图像中含有大量小目标,其形态表征能力较弱,严重影响小目标的检测精度。

近些年来,Yolo 系列检测模型因其速度快,精度高而被广泛引用于无人机航拍图像目标检测。Liu 等^[2]提出 MTI-YOLO 用于无人机电力线路检查中的绝缘子等目标,在 YOLOv3Tiny 的基础上添加特征融合结构和 SPP 模块来扩展颈部网络,并且添加了主干网络的输出层。发现,这种模型在颈部网络的改进是多余的,需要优化网络结构。为了提高小目标检测精度,Sahin 等^[3]扩展了 YOLOv3 的主干输出层,以检测图像中不同尺度的物体,将原来的 3 个检测层增加到 5 个,这样的结构在特征融合部分起到一定的作用,但这也导致了检测模型参数量极大增加,从而增加了训练和计算成本。Wang 等^[4]提出了 LDS-YOLO 检测模型,考虑到无人机图像中小目标特征难以提取的问题,该模型在 YOLOv5 的基础上构造了一个新的特征提取模块,在 SPP 模块中引入 SoftPool 结构,并采用深度可分离卷积替代传统卷积。该模型的优点在于参数量较小,但由于训练时样本不足,很容易使模型无法学习到目标特征。为了解决电力巡检无人机检测速率与精度无法兼顾的问题,提出 ED-YOLO 目标检测算法,在 YOLOv4 算法上引入注意力机制和深度可分离卷积,使得模型检测速度大大提升,但伴随着精度的下降^[5]。为了提高无人机小目标检测精度,陈蕊等^[6]提出了 AF-YOLO 模型,引入了自适应融合机制,在检测精度超过了原模型,但是精度的提升是在牺牲检测性能的情况下获得的。

综上所述,基于无人机平台目标检测算法的性能问题,提出了两个指导思想:1)在不降低检测精度的同时,通过轻量化设计理念提高检测性能;2)提升小目标检测算法的检测精度。

于是对原有的 YOLOv5s 模型采用了 5 种改进方法:1)添加 P_2 特征级别;2)在主干网络部分采

用 GhostConv 卷积模块替换 baseline 中的部分 Conv 模块,在 backbone 网络部分采用 GhostBottleneckC3 模块替换 baseline 中的部分 C3 模块;3)在颈部网络采用 BiFPN(bidirectional feature pyramid network)特征提取结构,弥补下采样特征信息丢失,以便快速地进行多尺度特征融合;4)在主干网络和颈部网络中采用了卷积注意力模块(convolutional block attention module, CBAM)^[7],用于提取重要特征,提升模型检测精度;5)在预测网络中引入 Soft-NMS 算法^[8],解决了 NMS 对密集物体检测的检测效果欠佳的问题。

1 YOLOv5 网络改进

1.1 YOLOv5s 网络结构改进

本研究基于 YOLOv5 提出了改进网络 YOLOv5_GBCS。为了确保提升小目标检测算法的检测精度的同时,提高检测效率,在主干网络采用 GhostConv 取代后 3 个普通卷积,再采用由 GhostBottleneck 结果组成的 GhostBottleneckC3 取代主干网络的前 3 个 C3 模块,极大地减少了参数量。同时,分别在主干网络和颈部网络引入 CBAM,增强了网络的特征提取能力的同时保持精度;添加 P_2 特征级别,提高了网络的深度和小目标检测的准确性;在颈部网络采用 BiFPN 结构,快速的进行多尺度特征融合。改进后的网络结果如图 1 所示。

1.2 引入 CBAM

在无人机航拍图像中含有大量非目标区域,可以通过引入注意机制使模型更好地关注目标区域的信息。注意力机制从大量的信息中选择少量的重要信息。改进模型中使用的 CBAM 是一种结合空间和通道的混合注意机制模块,图 2 为 CBAM 的整体流程结构,它由通道注意力模块和空间注意力模块两大模块构成。与只关注通道本身的 SENet^[9]相比,CBAM 整合了通道注意力(见图 2(a))与空间注意力(见图 2(b))映射过程,可以保留更多的特征信息,有效地提高训练出来的网络模型的检测速度,同时提高网络的特征提取能力和检测精度。CBAM 注意力模块的结构如图 2(c)所示。

图 2(a)所示的通道注意力模块首先分别对输入特征图的宽度和高度维度进行全局最大池化和全局平均池化,以聚合特征图的空间信息。然后将结果传递到多层感知机中,以便使用共享的全连接层进行处理。最后,通过 Sigmoid 激活函数,生成最终的通道注意力特征图 $M_c(F)$ 。该特征图的结果与输入特征图加权后,作为空间注意力模块的输入特征。

通道注意特征图 $M_c(F)$ 的数学表达式:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (1)$$

式中: σ 为 Sigmoid 激活函数, MLP 为层间的连接权值操作, AvgPool 和 MaxPool 分别表示最大池化操作和平均池化操作。

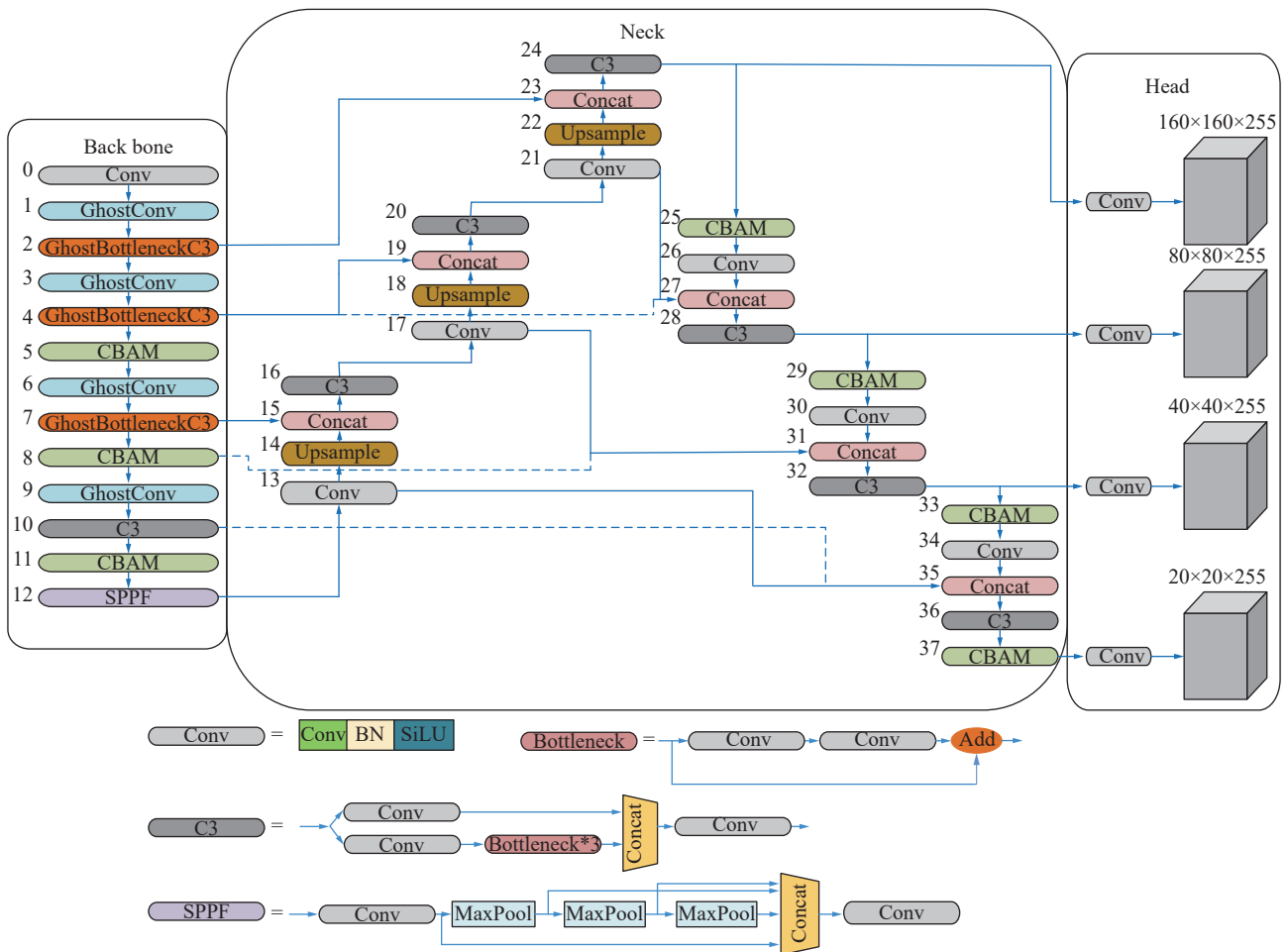


图 1 YOLOv5_GBCS 结构

Fig. 1 YOLOv5_GBCS structure diagram

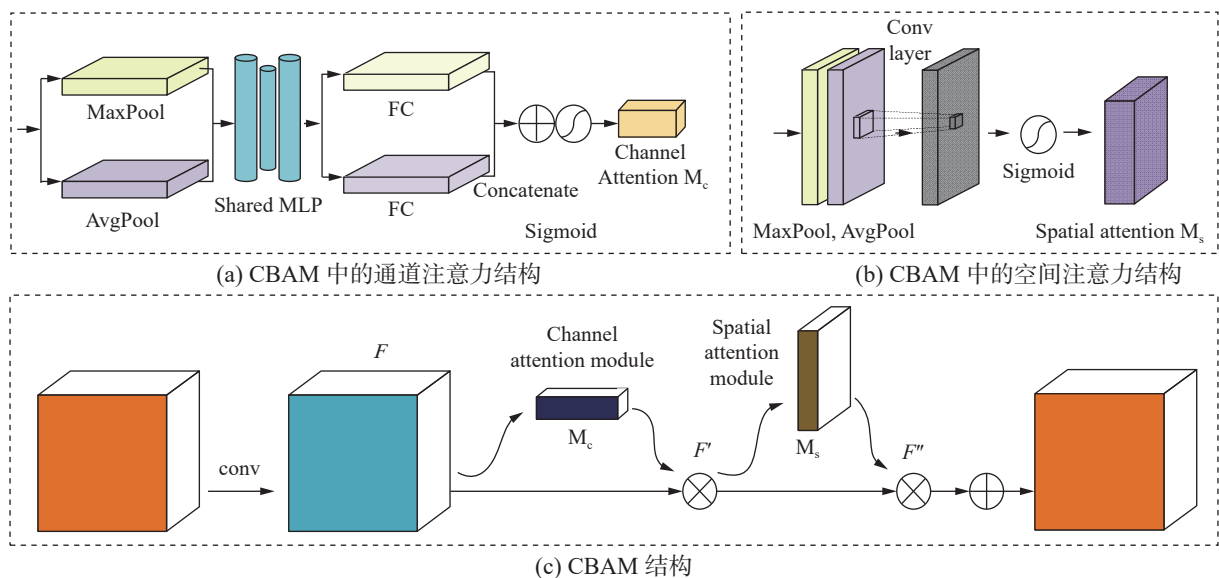


图 2 CBAM 模块以及子模块结构

Fig. 2 Structure of CBAM module and sub modules

在通道注意力模块之后,引入了如图 2(b) 所示的空间注意力模块。与通道注意力类似,空间注意力机制压缩通道,并分别在通道维度上进行均值池化和最大池化。首先,输入特征层在每个特征点的通道上取最大值和平均值。之后将两个结果叠加。最后通过 Sigmoid 函数生成空间注意力特征图 $M_s(F)$, 其数学表达式为

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (2)$$

式中: 7×7 表示卷积核的大小,即对特征图进行 7×7 的卷积操作。经验表明, 7×7 的卷积核优于 3×3 的卷积核。在主干网络和颈部网络引入了 CBAM。

从图 3 中可以看出,在添加了 CBAM 后,改进的算法相对于原 YOLOv5s 能够更多地关注目标区域,而忽略杂乱背景中的不相关因素,有利于提高目标特征的学习能力,提高检测精度。

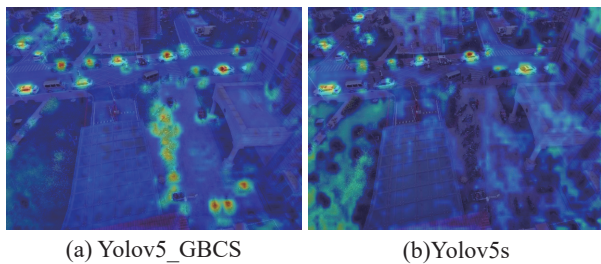


图 3 热力图可视化结果对比

Fig. 3 Comparison of visualization results of thermal map

1.3 基于 BiFPN 的特征融合网络

无人机的拍摄视角更高,视野更广可以捕捉到更多的物体,但是这也考验了检测算法对小目标识别的精度,所以如何表示和处理多尺度特征是一个难题。图 4 给出了多尺度特征处理结构。其中,图 4(a) 为 FPN 的结构,该方法将多尺度特征相结合,使用复合特征层和更多的语义信息进行预测。根据这一想法,图 4(b) 的 PANet^[10] 继续在 FPN 之上添加一个自下而上的路径聚合网络,同时考虑了顶层的语义信息和底层的位置信息。YOLOv5 框架将多尺度特性与 FPN+PAN 结构集成在一起,但由于输入特性的分辨率不同,FPN+PAN 结构对融合输出特性的贡献往往是不均匀的,不同尺度之间的特性不能得到充分利用。因此,在 YOLOv5 框架中引入了一种简单、高效的 BiFPN 网络结构,以提高检测精度。

如图 4(c) 所示, BiFPN 网络加强了更高层次的特征融合,将每个双向路径作为特征网络层进行处理,并在同一层上重复多次。通过加权特征的融合,学习不同输入特征的重要性,并对不同的特征进行区分融合。

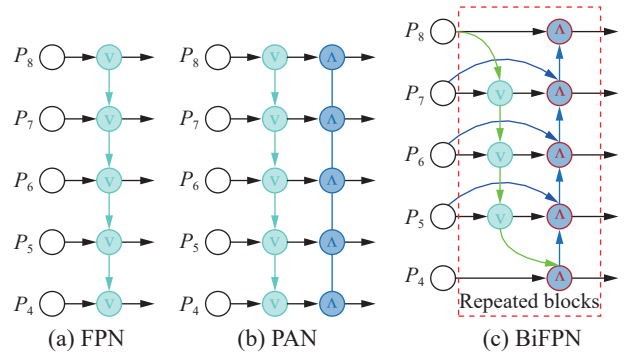


图 4 不同特征融合网络结构

Fig. 4 Structure of different feature fusion networks

BiFPN 使用快速归一化来融合加权特征,其定义为

$$O = \sum_i \frac{W_i}{\varepsilon + \sum_j W_j} \cdot I_i \quad (3)$$

式中:学习权重 W_i 使用 ReLU 激活函数,其值为 $\varepsilon = 0.0001$,以加强该值的稳定性。为了进一步提高深度网络学习模型的检测效率, BiFPN 使用了可分离的卷积融合特征,并在每次卷积后添加了批处理归一化和激活函数操作。以第 6 层为例,描述 2 个融合特征的定义如下:

$$P_6^{\text{td}} = \text{Conv} \left(\frac{W_1 \cdot P_6^{\text{in}} + W_2 \cdot R_{\text{esize}}(P_7^{\text{in}})}{W_1 + W_2 + \varepsilon} \right) \quad (4)$$

$$P_6^{\text{out}} = \text{Conv} \left(\frac{W'_1 \cdot P_6^{\text{in}} + W'_2 \cdot P_6^{\text{td}} + W'_3 \cdot R_{\text{esize}}(P_5^{\text{out}})}{W'_1 + W'_2 + W'_3 + \varepsilon} \right) \quad (5)$$

其中: P_6^{td} 从上到下表示第 6 层的中间特征, P_6^{out} 是第 6 层从下到上的输出特征。通过不同层间的互连和融合,最终实现了 BiFPN 的双向交叉尺度连接和快速归一化融合。基于上述优点,引入 BiFPN,改进特征金字塔结构,增强多尺度特征融合,提高模型检测精度。

1.4 针对小目标增加预测头

YOLOv5_GBCS 骨干网络包含连续的下采样卷积层,因此在特征信息提取过程中,特征图的尺寸随着网络的加深而减小。特征图的小尺寸影响了对图像中的小物体的检测。然而,无人机图像中包含了大量的小物体。为了增强小物体的特征融合效果,于是添加了一个额外的检测头。改进的模型如图 1 所示。在第 20 层之后,继续对特征图进行上采样,以确保特征图继续扩展,在第 23 层,将尺寸为 160×160 的特征图与主干网络中第 2 层的特征图进行融合,以获得小目标的更大特征图输入至第 24 层后的检测头中。将高层、低层的特征信息进行特征融合,使添加

的检测头对小物体更加敏感。最后, 检测头的 4 种不同尺度特征更能适应无人机图像物体尺度的剧烈变化。

1.5 网络结构轻量化设计

GhostNet 提出了一个创新的模块 Ghost, 它用更少的参数与计算量得到更多的特征图。Ghost 模块一部分是普通卷积, 另一部分是线性运算。Ghost 的工作原理如图 5 所示。

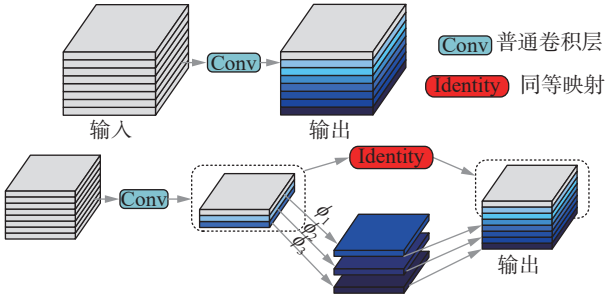


图 5 Ghost 卷积与普通卷积的比较

Fig. 5 Comparison of Ghost convolution and general convolution

普通卷积层的操作为

$$Y = X \cdot f + b \quad (6)$$

Ghost 模块首先通过普通卷积生成少量的特征图; 然后对第 1 步得到的特征图进行计算量较小的线性操作, 生成 Ghost 特征图。最后, 对两组特征图进行通道拼接, 生成足够的特征图来匹配

给定的输出通道数量。

Ghost 卷积采用了逐步进行的策略, 计算方法为

$$Y' = X \cdot f' \quad (7)$$

$$Y_{ij} = \phi_{ij} \cdot Y'_i, i[1, m], j[1, s] \quad (8)$$

式中: $Y \in \mathbf{R}^{W' \times H' \times m}$ 表示输入 X 在经过普通卷积 $f' \in \mathbf{R}^{k \times k \times C \times m}$ 生成的 m 个特征图 ($m \ll n$); 然后将 m 个特征图逐个进行线性操作, 每个特征图生成 1 个特征图, 共生成 $n = m \times s$ 个特征图。特征图 ϕ_{ij} 表示对第 1 步卷积生成的第 j 个特征图, Y'_i 表示第 i 次线性运算。为了保证 CPU 或 GPU 的效率和实用性, 将每个线性操作的卷积核的大小设置为 $d \times d$; 那么普通卷积和 Ghost 卷积的速度比可以通过下式计算得出:

$$R_s = \frac{W' \cdot H' \cdot n \cdot k \cdot k \cdot C}{W' \cdot H' \cdot m \cdot k \cdot k \cdot C + W' \cdot H' \cdot (n - m) \cdot d \cdot d \cdot C} = \frac{W' \cdot H' \cdot n \cdot k \cdot k \cdot C}{W' \cdot H' \cdot (n/s) \cdot k \cdot k \cdot C + W' \cdot H' \cdot (s - 1) \cdot d \cdot d \cdot C} = \frac{k \cdot k \cdot C}{(1/s) \cdot k \cdot k \cdot C + (s - 1) \cdot d \cdot d} \approx \frac{s \cdot C}{s + C - 1} \approx s \quad (9)$$

Ghost 卷积是一个更轻、更快的模块。在此基础上, 主干网络采用了 GhostConv 以及基于 Ghost 模块的 GhostBottleneckC3 模块, 具体结构如图 6 所示。

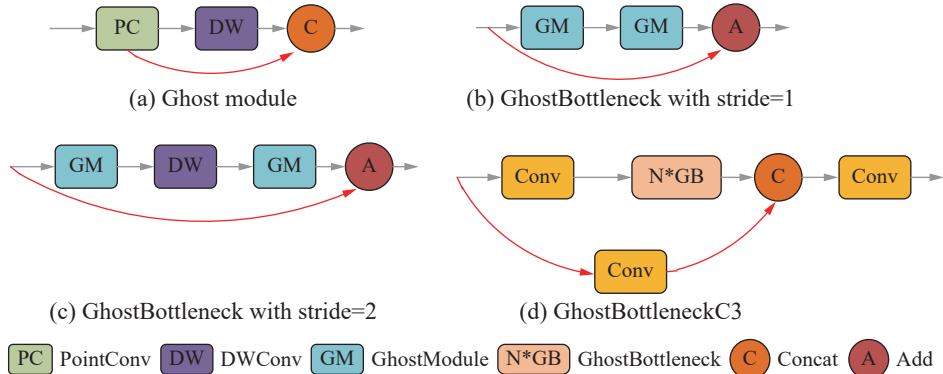


图 6 Ghost module 模块、GhostBottleneck 模块以及 GhostBottleneckC3 模块结构

Fig. 6 Structure of Ghost module, GhostBottleneck module, and GhostBottleneckC3 module

由于 Bottleneck 中使用了较多 3×3 标准卷积, 因此用 GhostBottleneck 结构替换 C3 模块中的 Bottleneck 模块能够将模型的体积和计算量进一步压缩。用图 6(b) 的 GhostBottleneck 结构替换 YOLOv5s 中 C3 的 Bottleneck 模块得到新的 GhostBottleneckC3 结构, 如图 6(d) 所示。GhostBottleneckC3 结构可以减少模型的计算量并压缩体积。

1.6 NMS 的改进

NMS 被应用于大多数最先进的检测器, 因为它能够显著地减少了假阳性的数量。NMS 算法的流程图如图 7 所示。首先, 列表中的所有检测框都根据它们的置信度得分进行排序。其次, 将得分最高的检测框 B_m 移动到最终的检测列表 D 中, 并为剩余的检测框分配一个唯一的标识符 B_i 。第三, 去除任何与的重叠面积大于一定阈值

N_i 的预测框 B_i 。对其余的框 B_i 重复此过程, 直到初始列表被置空为止。然而, 由于无人机视角的特

殊性, 会导致大量小目标重叠, 于是会导致 NMS 算法的检测失误。

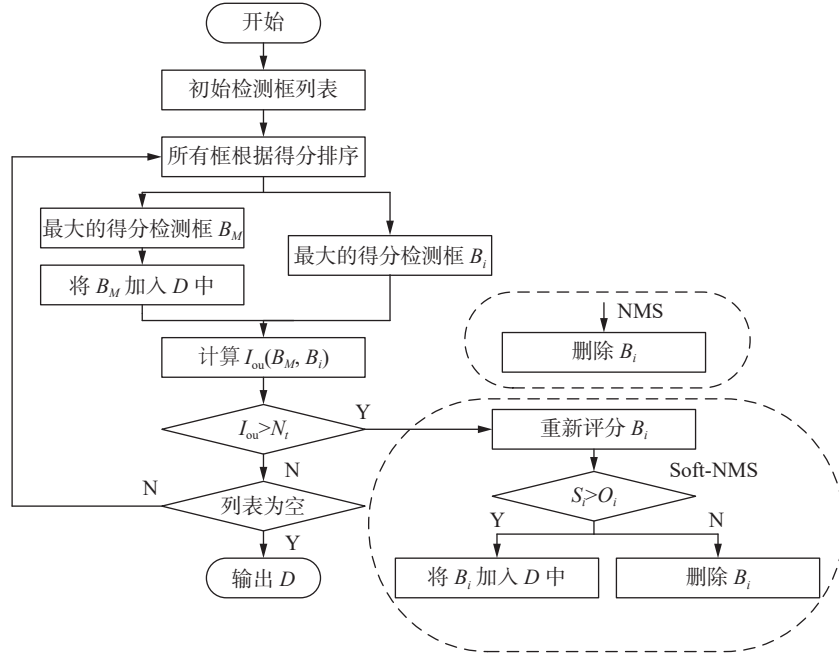


图 7 NMS 和 Soft-NMS 的算法流程

Fig. 7 The algorithm flow of NMS and Soft-NMS

接下来, 对于无人机的密集目标的场景, 本身属于两个目标的检测框可能 I_{ou} 很高, 如果使用 NMS 算法, 会将置信度较小的那个检测框删除掉, 有可能造成漏检, 降低了模型的召回率。因此, 引入了 Soft-NMS 而不是 NMS 来处理检测结果。

Soft-NMS 的核心思想是使用一个惩罚函数来减弱与 B_M 重叠的预测框的分数, 而不是将这些分数设置为零。Soft-NMS 的 S_i 的计算方式为

$$S_i = \begin{cases} S_i, & I_{ou}(B_M, B_i) < N_t \\ S_i \cdot \exp\left(-\frac{I_{ou}(B_M, B_i)^2}{\sigma}\right), & I_{ou}(B_M, B_i) \geq N_t \end{cases} \quad (10)$$

式中: $\exp\left(-\frac{I_{ou}(B_M, B_i)^2}{\sigma}\right)$ 是一个高斯惩罚函数, 而 σ 是根据经验选择的超参数。很明显, 与 B_M 重叠较大的预测框的得分会大大降低, 而远离 B_M 的检测框则不会受到影响。如果预测框的得分仍然高于惩罚后的置信阈值 O_i 。然后, 预测框将被保留, 而不是被丢弃。在目标密集的场景中, 使用 Soft-NMS 来处理小目标检测问题, 可以大大降低误检率。

2 实验结果与分析

2.1 数据集及实验环境

为了评价改进后的 Yolov5s 方法, 使用 Vis-

Drone2019 数据集对该算法进行检验。VisDrone2019 数据集中共计包含 6 471 张训练集图像、548 张验证集图像和 1 610 张测试集图像。数据集中包含各种交通场景, 包括高速公路、十字路口、丁字路口, 不同的环境背景, 包括白天、夜晚、雾霾天、雨天。该类数据集满足实验训练的需求, 能够满足无人机对地面小目标的检测效果验证的需要。本实验中的所有算法均在训练集上进行训练, 并在验证集上进行评估。

实验平台的硬件环境为 CPU: Intel(R) Xeon(R) Silver 4 114 CPU, GPU: TeslaV100S-PCIE 32 GB, 内存: 64 GB, 软件环境为 CUDA10.1+cuDnn 7.6.5+Python-3.9.12+Pytorch1.8.1。

2.2 实验步骤

Visdrone2019 数据集和 VOC 数据集的目标尺度与宽高比大不相同, 先验框不再合适本次实验, 需要重新计算。因此, 在训练步骤之前, 使用 k-means^[11] 聚类算法在 Visdrone2019 数据集上计算新的先验框。

对所有实验学习率均采用 Warmup^[12] 训练, 保持模型的深度稳定性。在热身训练阶段, 学习率从 0 上升到设置大小的 0.01。在预热阶段结束后, 使用余弦退火算法更新学习速率, 在模型训练中采用 Adam 优化器。网络模型的权值进行更新和优化。具体参数设置为: batch size 大小为

64, 学习率为 0.01, 动量参数为 0.937, 权重衰减因子为 0.0005, 训练轮数设置为 300。

2.3 评价指标

在本实验中, 选择 mAP@0.5、参数量、每秒千兆浮点运算量 (gigabit floating-point operations per second, GFLOPS) 和检测速率来作为目标检测模型的评价指标。其中 mAP@0.5 表示在 I_{ou} 阈值为 0.5 时的平均检测精度 P_A 。

参数量和计算量是衡量深度学习算法的重要指标, 计算量对应的是算法的时间复杂度, 参数量对应的是算法的空间复杂度, mAP@0.5 主要用于体现模型的识别能力, 检测速率为每秒检测图像的数量, 值越大检测速度越快, 效率越高。

上述各指标的计算公式如下:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (11)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (12)$$

$$P_A = \int_0^1 PdR \quad (13)$$

$$P_m = \frac{\sum_{i=1}^N P_{Ai}}{N} \quad (14)$$

$$v = \frac{N_{Figure}}{T_{Total}} \quad (15)$$

式中: N_{TP} 、 N_{FP} 和 N_{FN} 分别表示物体被正确检测的数量、物体被错误检测的数量和没有被检测到的物体的数量, P_{Ai} 为第 i 类物体的平均检测精度, N 为物体的类别数量, N_{Figure} 表示检测图片的总数, T_{Total} 表示总检测需要消耗的时间。

2.4 对比实验

在 VisDrone2019 数据集上测试了 Yolov5_GBCS 模型, 并将结果与其他具有代表性的模型进行了比较。结果如表 1 所示。证明了其对无人机图像的小目标检测的有效性。这些结果表明, 该模型可以在保持目标检测性能的同时, 提高了小目标检测能力。

表 1 不同网络模型下不同目标的平均检测精度与 mAP@0.5

Table 1 Average precision and mAP@0.5 of different targets under different network models

模型	不同目标的 P_A										mAP@0.5
	Pedestrian	People	Bicycle	Car	Van	Trunk	Tricycle	Awing-t	Bus	Motor	
Yolov5s	0.542	0.474	0.282	0.645	0.528	0.525	0.429	0.304	0.627	0.505	0.385
CenterNet ^[13]	0.173	0.105	0.056	0.483	0.263	0.185	0.084	0.055	0.309	0.155	0.336
RetinaNet ^[14]	0.084	0.031	0.002	0.414	0.154	0.100	0.040	0.030	0.134	0.074	0.191
Cascade-RCNN ^[15]	0.199	0.123	0.084	0.541	0.353	0.264	0.174	0.092	0.422	0.196	0.390
Fcos ^[16]	0.173	0.093	0.033	0.513	0.267	0.225	0.086	0.070	0.341	0.096	0.319
Tridentnet ^[17]	0.169	0.105	0.059	0.508	0.288	0.224	0.144	0.071	0.330	0.168	0.353
Faster-RCNN ^[18]	0.176	0.120	0.072	0.505	0.301	0.233	0.144	0.089	0.372	0.182	0.371
FSAF ^[19]	0.221	0.141	0.056	0.538	0.293	0.215	0.089	0.058	0.309	0.165	0.364
ATSS ^[20]	0.197	0.065	0.074	0.544	0.310	0.245	0.142	0.087	0.370	0.181	0.363
DDOD	0.219	0.119	0.074	0.553	0.314	0.254	0.145	0.086	0.372	0.198	0.382
TOOD ^[21]	0.219	0.130	0.086	0.562	0.330	0.260	0.161	0.091	0.388	0.214	0.398
VFNet ^[22]	0.206	0.091	0.067	0.553	0.323	0.253	0.147	0.083	0.390	0.192	0.373
LDS-YOLO	0.557	0.459	0.335	0.692	0.541	0.519	0.430	0.326	0.688	0.521	0.415
M-YOLO	0.581	0.492	0.366	0.652	0.557	0.521	0.443	0.314	0.741	0.533	0.424
Yolov5_GBCS	0.574	0.540	0.350	0.683	0.569	0.528	0.447	0.334	0.736	0.546	0.432

从表 1 中可以看出, 与 Yolov5s 相比, 提出的模型无论各个物体类别的精度还是 mAP@0.5 超过了 Yolov5s。Pedestrian 的 P_A 提升了 3.2%, People 的 P_A 提升了 6.6%, Bicycle 的 P_A 提升了 6.8%, Car 的 P_A 提升了 3.8%, Van 的 P_A 提升了 4.1%, Tricycle 的 P_A 提升了 13.9%, Awing-Tricycle 的 P_A 提升了 3.0%, Bus 的 P_A 提升了 10.9%, Motor 的 P_A 提升了 4.1%, 其中 Trunk 的 P_A 提升的并不大, 只有 0.3%。最后, 改进后的模型 Yolov5_GBCS 的 mAP@0.5 比

Yolov5s 提升了 5.1%, 与 LDS-YOLO 和 M-YOLO 这两个同样基于改进的 Yolov5 的小目标检测算法相比 mAP@0.5 分别提升了 1.7% 和 0.8%。由此可以得出结论, 在针对无人机图像小目标检测方面, Yolov5_GBCS 模型比原 Yolov5s 更加优秀。

图 8 中给出了来自 VisDrone2019 数据集上的 Yolov5_GBCS 的一些具有代表性的检测结果。可以看出相对于 Yolov5s, Yolov5_GBCS 可以检测出更多的小目标。

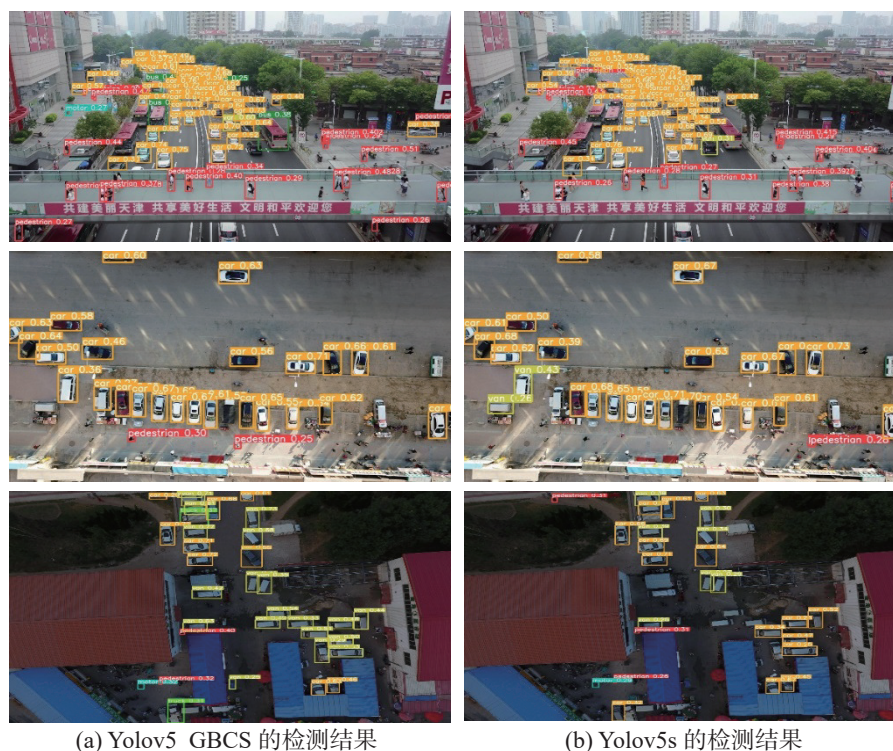


图 8 VisDrone2019 数据集上的检测结果的可视化

Fig. 8 Visualization of detection results on the VisDrone2019 dataset

2.5 消融实验

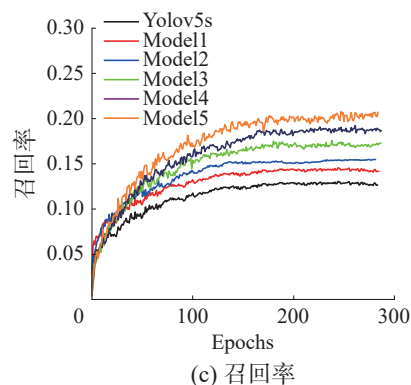
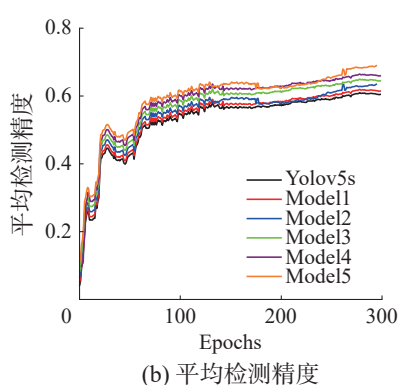
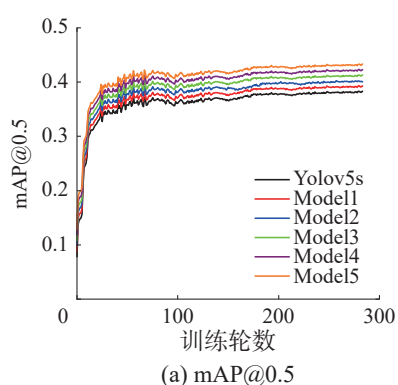
为了验证 5 种改进方法的有效性, 于是基于 VisDrone 2019 数据集进行了消融实验。在表 2 和图 9 中 Model1 为在 Yolov5s 中增加 P_2 特征级别检测头, Model2 为在 Model1 的基础上将主干网络分

别采用 GhostConv 卷积模块、GhostBottleneckC3 模块替换 baseline 中的部分 Conv 模块和 C3 模块, Model3 为在 Model2 的基础上引入 BiFPN 结构, Model4 为在 Model3 的基础上加入 Soft-NMS 算法, Model5 为在 Model4 的基础上加入 CBAM 注意力模块。

表 2 VisDrone 数据集上 6 种模型的消融研究

Table 2 Ablation study of six models on VisDrone dataset

模型	P_2	Ghost	BiFPN	CBAM	Soft-NMS	参数量/ 10^6	GFLOPS	检测速率/(f/s)	mAP@0.5/%
Yolov5s	—	—	—	—	—	7.04	16.0	53	38.5
Model1	√	—	—	—	—	7.27	19.3	45	40.2
Model2	√	√	—	—	—	5.96	14.6	59	41.3
Model3	√	√	√	—	—	5.96	14.6	59	42.1
Model4	√	√	√	√	—	5.97	14.6	59	42.8
Model5	√	√	√	√	√	5.97	14.6	59	43.2



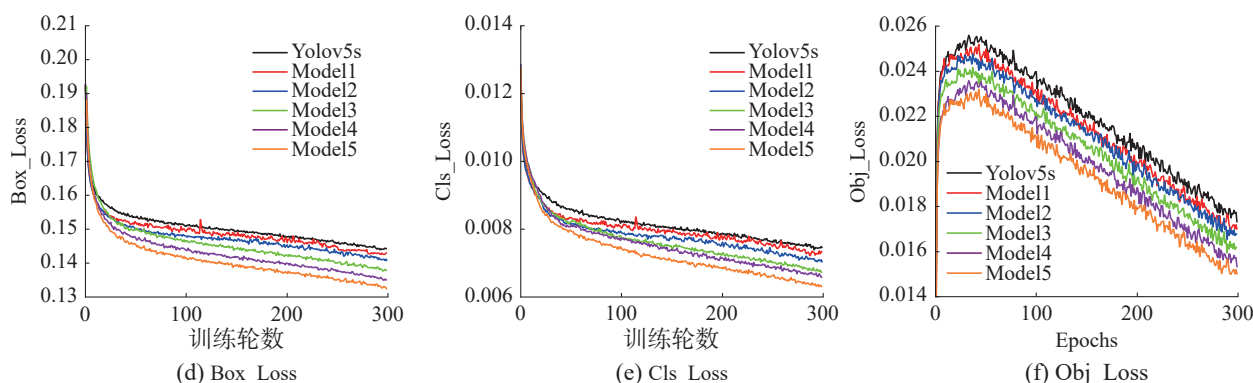


图9 比较6个模型的训练结果

Fig. 9 Compare the training results of the six models

如图9所示,在训练过程中,损失函数值呈下降趋势,Adam算法对网络进行优化,网络权值等参数不断更新。在训练到40轮之前,损失函数值迅速下降,平均值测精度、召回率和 $mAP@0.5$ 值迅速提高。当训练到100轮左右时,损失函数值的下降逐渐减慢。同样,平均值测精度、召回率和 $mAP@0.5$ 的提升也有所放缓。当训练期到250轮时,训练的损失曲线几乎没有下降趋势,其他指标值也趋于稳定,网络模型基本达到收敛状态,并在训练结束时得到最优的网络权值。图9(a)显示,Model5模型经过约200轮训练后, $mAP@0.5$ 达到了42%,并逐渐稳定,最大值达到43.5%。图9(b)显示,当Model5模型训练到230轮时,平均检测精度达到63.82%,并继续增长到68.88%。图9(c)显示,Model5模型的召回率在40轮之前快速上涨,然后继续增长到20.3%的最高值。整体的模型性能已经达到了预期水平,甚至超出了预期水平。该模型的具体损失函数如图9(d)、(e)、(f)所示。从图9中可以看出,随着训练轮数的逐渐增加,Model5模型算法曲线逐渐收敛,损失值越来越小。当模型迭代230次时,损失值基本稳定,网络基本收敛。与其他模型相比,Model5模型对无人机小目标具有更好的检测性能和识别效果,回归速度更快、更准确,证明了该模型的有效性。消融实验试验结果表明,每种改进的方法都比原来的Yolov5s模型有一定的改进。

通过比较表2中Yolov5s和Model1的数据,发现增加 P_2 特征级别检测头可以使模型的 $mAP@0.5$ 从38.5%增加到40.2%,这表明针对小目标数据集增加尺度更大的检测头,保留了更丰富的小目标特征信息,在图像中的错误检测和遗漏检测得到改进,使得目标框和先验框都能更好地匹配并帮助损失函数更好地收敛,但是参数量和GFLOPS分别增加了3.3%和20.6%,直接导致塑

料明显下降。对比表2中Model1与Model2的数据可以看出,在引入了GhostConv模块和Ghost-BottleneckC3模块后参数量、GFLOPS分别下降了18.0%和24.3%, $mAP@0.5$ 上涨了0.9%,检测速率也由45 f/s上涨至59 f/s。这表面对网络结构进行轻量化处理后,在避免了网络模型过于复杂从而造成的实时性能损失的同时,检测精度也有所提升,达到速度与精度的平衡。对比表中Model2和Model3的数据,参数量和GFLOPS都没有增加的情况下, $mAP@0.5$ 上涨了0.8%,这表明在颈部网络采用BiFPN后,弥补下采样特征信息丢失,增强多尺度特征融合,提高了模型检测精度。CBAM注意力模块旨在提高网络提取重要特征的能力,从表中可以看出,Model4相对与Model3的 $mAP@0.5$ 在检测速率、参数量和GFLOPS几乎不变的情况下,上涨了0.7%,可以说明CBAM机制的增加了网络模型的深度,在保证速度不变的同时,提高了小目标的检测效果。用Soft-NMS算法取代NMS算法后,模型有效地克服了待检测对象相互重叠时漏检的问题,模型的检测精度得到提高。Yolov5_GBCS相对于Yolov5s, $mAP@0.5$ 方面提高到4.7%;在参数量、GFLOPS方面,分别下降了15.2%和8.7%;检测速率从53 f/s提升到59 f/s。基于上述可以得知,Yolov5_GBCS无论是在检测效率方面还是检测精度方面都远远优于Yolov5s。

3 结束语

无人机的飞行高度、视角和速度均会导致无人机图像中的目标较小,背景复杂,遮挡量大,给无人机图像中的目标检测带来了困难。为了解决上述问题,提出了一种改进的无人机航拍图像目标检测模型Yolov5_GBCS算法。通过增加一个 P_2 特征级别的检测头,算法在具有复杂背景的无

人机图像中检测小目标的能力得到了提高。引入 Ghost 模块和 GhostBottleneckC3 模块取代 YOLOv5 中的 Conv 和 C3 模块,减小了模型的参数量与运算量。添加 BiFPN 来进行多尺度特征融合,提高特征提取性能。将注意机制 CBAM 融入算法中,提高复杂背景下对象区域的显著性,有效提高检测器的精度。此外,Soft-NMS 有效的解决多目标重叠的问题,降低了误检和漏检的概率。在 Visdrone2019 数据集上的实验表明, YOLOv5_GBCS 算法能够有效地实时检测无人机视觉中的小物体,并且在 mAP@0.5 方面超过 YOLOv5s 以及其他大多数检测模型,证明了该方法的有效性。

后续计划在 YOLOv5_GBCS 算法的卷积层上进行通道剪枝,以减少网络参数的数量,加快网络计算速度。同时,将采用知识蒸馏来补偿剪枝模型的精度损失。轻量级设计使计算速度和精度之间的平衡,以满足部署无人机等嵌入式设备的需求。显然,该算法对环境监测、地质灾害探测、精准农业、城市规划等各种民用和军事任务具有较高的实用价值。

参考文献:

- [1] 曹家乐, 李亚利, 孙汉卿, 等. 基于深度学习的视觉目标检测技术综述 [J]. 中国图象图形学报, 2022, 27(6): 1697-1722.
CAO Jiale, LI Yali, SUN Hanqing, et al. A survey on deep learning based visual object detection[J]. Journal of image and graphics, 2022, 27(6): 1697-1722.
- [2] LIU Chuanyang, WU Yiquan, LIU Jingjing, et al. MTI-YOLO: a light-weight and real-time deep neural network for insulator detection in complex aerial images[J]. *Energies*, 2021, 14(5): 1426.
- [3] SAHIN O, OZER S. YOLODrone: improved YOLO architecture for object detection in drone images[C]//2021 44th International Conference on Telecommunications and Signal Processing. Brno: IEEE, 2021: 361-365.
- [4] WANG Xuwen, ZHAO Qingzhan, JIANG Ping, et al. LDS-YOLO: A lightweight small object detection method for dead trees from shelter forest[J]. *Computers and electronics in agriculture*, 2022, 198: 107035.
- [5] 彭继慎, 孙礼鑫, 王凯, 等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法 [J]. 仪器仪表学报, 2021, 42(10): 161-170.
PENG Jishen, SUN Lixin, WANG Kai, et al. ED-YOLO power inspection UAV obstacle avoidance target detection algorithm based on model compression[J]. *Chinese journal of scientific instrument*, 2021, 42(10): 161-170.
- [6] 陈蕊, 郑华飞, 蒋鸿宇, 等. 结合仿真迁移学习和自适应融合的无人机小目标检测 [J/OL]. 小型微型计算机系统, 2022: 1-10. (2022-05-10). <https://kns.cnki.net/kcms/detail/21.1106.TP.20220509.1930.032.html>.
- [7] CHEN Rui, ZHENG Huafei, JIANG Hongyu, et al. Combination of simulation-based transfer learning and adaptive fusion for UAV small object detection[J/OL]. *Journal of Chinese computer systems*, 2022: 1-10. (2022-05-10). <https://kns.cnki.net/kcms/detail/21.1106.TP.20220509.1930.032.html>.
- [7] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision. Cham: Springer, 2018: 3-19.
- [8] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS-improving object detection with one line of code[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 5562-5570.
- [9] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132-7141.
- [10] LIU Shu, QI Lu, QIN Haifang, 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.
- [11] KRISHNA K, NARASIMHA MURTY M. Genetic K-means algorithm[J]. *IEEE transactions on systems, man, and cybernetics, part B (cybernetics)*, 1999, 29(3): 433-439.
- [12] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.
- [13] DUAN Kaiwen, BAI Song, XIE Lingxi, et al. CenterNet: keypoint triplets for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 6568-6577.
- [14] WANG Haotong, LIU Bailin, LIU Zhiping, et al. LSS target threat estimation and capture prediction trajectory simulation[C]//2020 7th International Conference on Information Science and Control Engineering. Changsha: IEEE, 2020: 806-813.
- [15] CAI Zhaowei, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6154-6162.
- [16] TIAN Zhi, SHEN Chunhua, CHEN Hao, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9626-9635.

- [17] LI Yanghao, CHEN Yuntao, WANG Naiyan, et al. Scale-aware trident networks for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 6053–6062.
- [18] REN Shaoqing, HE Kaiming, 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.
- [19] ZHU Chenchen, HE Yihui, SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 840–849.
- [20] ZHANG Shifeng, CHI Cheng, YAO Yongqiang, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 9756–9765.
- [21] FENG Chengjian, ZHONG Yujie, GAO Yu, et al. TOOD: task-aligned one-stage object detection[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 3490–3499.
- [22] AHMED A, TANGRI P, PANDA A, et al. VFNet: a convolutional architecture for accent classification[C]//2019

IEEE 16th India Council International Conference. Rajkot: IEEE, 2019: 1–4.

作者简介:



何宇豪, 硕士研究生, 主要研究方向为目标检测。E-mail: 1772307157@qq.com。



易明发, 硕士研究生, 主要研究方向为目标检测。E-mail: 1781987848@qq.com。



王冠凌, 教授, 主要研究方向为嵌入式开发。主持省教育厅科研重点项目 1 项, 获省科技进步奖一等奖 1 项, 获发明专利授权 4 项, 发表学术论文 20 余篇。E-mail: ahpu_405@163.com。

2024 全球人工智能技术大会 The Global Artificial Intelligence Technology Conference (GAITC 2024)

2024 年 6 月 22—24 日, 由中国人工智能学会主办的 2024 全球人工智能技术大会(GAITC 2024)将在杭州盛大举行, 大会汇聚来自多国的人工智能专家和技术大咖, 探讨人工智能技术的最新发展、未来趋势以及其对各行各业的影响。

本着高端化、国际化的办会理念, 大会将邀请由中外院士领衔的, 来自高校、研究院、企业的数百位国内外专家学者同场交流探讨, 以“会、展、赛”三位一体的内容形式, 通过 1 场主论坛、30 余场专题论坛, 以及 2024 全球人工智能技术创新大赛、全球人工智能技术博览会、2024 IEEE“一带一路”人工智能可持续发展大会等多个重磅国际性同期活动, 分享学术成果和技术创新, 呈现应用实践和产业趋势, 共谋安全防范与社会治理, 促进人工智能界与各行各业的跨域交流与协同创新, 为我国“人工智能+”行动添能蓄力。

更多详细信息请前往官网 <https://gaitc.caai.cn/>。