



DOI:10.12404/j.issn.1671-1815.2403707

引用格式:白俊卿,王梦婷,沈守婷.基于YOLOv5的无人机遥感图像车辆小目标检测算法[J].科学技术与工程,2025,25(12):5110-5118.
Bai Junqing, Wang Mengting, Shen Shouting. Vehicle small target detection algorithm for UAV remote sensing images based on YOLOv5[J].
Science Technology and Engineering, 2025, 25(12): 5110-5118.

基于 YOLOv5 的无人机遥感图像车辆 小目标检测算法

白俊卿,王梦婷*,沈守婷
(西安石油大学计算机学院,西安 710065)

摘要 遥感图像具有尺度多样、密集排列和目标尺寸小等特点,针对遥感图像中背景噪声多,车辆目标较小难以获取的问题,提出一种基于改善特征融合方法的车辆目标检测算法 Atiny-YOLO。首先,在 YOLOv5 的 Neck 层中引入针对小目标的额外检测层,从而生成更大规模的特征图,有效识别小物体的细节特征;其次,向 C3 模块中添加 split 操作以复用图像特征信息,进一步优化 Swin Transformer 模块提高有效信息的使用率;最后,通过改善特征融合通道,提升检测精度的同时减少模型参数。在无人机视角数据集(aerial universal autonomous inspection and recognition, AU-AIR)数据集上验证 Atiny-YOLO 模型的有效性。实验结果表明:Atiny-YOLO 算法相较于基线算法的平均检测精度提高了约 2.9%。达到 95.5%,检测速度达到 234 帧/s。这些结果验证了 Atiny-YOLO 算法在满足实时性的同时,模型检测精度大幅提升。

关键词 遥感图像; Swin Transformer; 车辆检测; AU-AIR; Atiny-YOLO

中图分类号 TP753; 文献标志码 A

Vehicle Small Target Detection Algorithm for UAV Remote Sensing Images Based on YOLOv5

BAI Jun-qing, WANG Meng-ting*, SHEN Shou-ting
(School of Computer Science, Xi'an Petroleum University, Xi'an 710065, China)

[Abstract] Remote sensing images are characterized by diverse scales, dense arrangement and small target sizes, etc. Aiming at the problem that there is much background noise in remote sensing images and vehicle targets are small and difficult to be acquired. A vehicle target detection algorithm based on improved feature fusion method, Atiny-YOLO was proposed. Firstly, an additional detection layer for small targets was introduced into the Neck layer of YOLOv5 so as to generated a small target detection algorithm for drone remote sensing images. Neck layer to introduce an additional detection layer for small targets, so as to generated a larger-scale feature map and effectively identified the detailed features of small objects. Secondly, a split operation was added to the C3 module to reuse the image feature information, and the Swin Transformer module was further optimized to improve the usage rate of the effective information. Lastly, by improving the feature fusion channel, the detection accuracy was improved while the model parameters were reducing the model parameters. The Atiny-YOLO algorithm was tested on the AU-AIR (aerial universal autonomous inspection and recognition) dataset. The experimental results show that the average detection accuracy of the Atiny-YOLO algorithm compared to the baseline algorithm is improved by about 2.9%. It reaches 95.5% and the detection speed reaches 234 frames/s. These results verify that the Atiny-YOLO algorithm meets the real-time performance while the model detection accuracy is greatly improved.

[Keywords] remote sensing images; Swin Transformer; vehicle detection; AU-AIR; Atiny-YOLO

无人机遥感技术^[1]具有系统简单、成本较低和快速、宏观、动态的显著特点,很好地弥补了传统遥感技术的应急缺陷^[2]。由于无人机拍摄得到的遥感图像中信息含量较多,细节更为复杂,实时性较

强,且会受到天气因素干扰,因此对遥感图像中的车辆目标进行检测面临着以下困难。由于拍摄距离和角度的变化,不同类型的车辆在图像中呈现出不同的尺寸和形状,存在显著差异;遥感图像的背景

收稿日期:2024-05-19; 修订日期:2025-01-28

基金项目:陕西省自然科学基金基础研究计划项目(2023-JC-YB-601);西安市科技计划高校院所人才服务企业项目(23GXFW0077);西安石油大学研究生精品课程建设项目(2023-X-YKC-003)

第一作者:白俊卿(1983—),女,汉族,河南商丘人,博士,副教授,硕士研究生导师。研究方向:机器学习、人工智能。E-mail:13636804262@qq.com。

*通信作者:王梦婷(2000—),女,汉族,河南驻马店人,硕士研究生。研究方向:目标检测。E-mail:2570046402@qq.com。

投稿网址:www.stae.com.cn

景包含了大量与车辆目标相似或其他易受干扰的信息,如建筑物阴影、道路纹理和树木遮挡等,这些背景干扰会降低车辆目标的显著性;小目标的检测具有挑战性,因为在遥感图像中,车辆目标通常只有几个像素大小,与背景的对比度较低,难以捕捉到车辆的局部细节特征^[3]。

车辆目标检测方法基于神经网络主要分为两种:双阶段检测算法和单阶段检测算法。双阶段检测算法首先提取候选框,然后对这些候选框中的目标进行识别。然而,由于其结构限制,双阶段检测算法通常需要较长的检测时间,难以满足实时性需求。代表性的算法包括R-CNN(region-based convolutional neural networks)系列的R-CNN^[4]、Fast R-CNN^[5]和Faster R-CNN^[6]。相比之下,单阶段算法省略了候选框提取阶段,能够在—个阶段内完成目标的确定和定位。其中非常具有代表性的单阶段算法包括YOLO(you only look once)^[7-9]系列和SSD(single shot multibox detector)^[10]系列。

近年来,为了解决遥感图像中车辆目标检测精度低的问题,中外学者针对目标检测方法进行了相关改进,文献[11]通过替换YOLOv3的特征提取网络,优化了卷积网络结构,同时引入密集连接提高了检测精度,但不能满足实时性的要求。文献[12]通过引入SPP模块优化YOLOv3算法,同时采用4倍、8倍、16倍下采样对更多小目标信息抓取,实现了检测精度与速度的提升,较YOLOv3精度提高了13.29%,检测速度达到39FPS,但是对于多尺度小目标的检测效果不佳。文献[13]对轻量级车辆和行人检测网络进行了深入研究,尽管网络参数量大幅下降,但模型精度仍有待提高。文献[14]在YOLOv5的颈部引入C3Ghost和Ghost模块,同时在Backbone层中引入CBAM注意力模块^[15],以增强对车辆检测任务中重要信息的提取能力,采用完全交并比损失函数(CIOU Loss)^[16],以提高算法的定位精度。文献[17]将空洞卷积融入到YOLOv5的空间池化金字塔模块,增强了网络的特征提取能力,提高了网络对小目标的检测精度。文献[18]通过添加特征识别模块,成功提升了YOLOv5s的检测精确度,实现了车辆的多目标跟踪,满足对图像检测实时性的需求。

上述方法在检测精度上都有了不同程度的提升,但是大多数算法在优化模型Backbone时,没有充分利用Backbone提取的多尺度特征图,同时未考虑不同尺度特征对融合后特征的贡献度的问题,导致融合后特征的信息冗杂,削弱了小目标的特征信息;此外,遥感图像中的目标面临着尺度变化巨大

的挑战,然而目前只有少数算法能够同时检测到小尺寸目标和大尺寸目标,网络对目标尺度的泛化能力仍需进一步提高。

针对上述问题,为了解决车辆小目标检测精度低且常漏检的问题,提出一种基于特征融合方法的检测方法Atiny-YOLO。在Neck层中添加小目标检测层,将三尺度检测变为四尺度检测,加强了对小尺度目标的学习能力,降低车辆小目标的漏检率;用自定义的myC3模块替换Backbone和head层中C3模块,并进一步提高Swin Transformer的运算速率,减小算法运算量,同时使用自定义的myWinTR来解决在实际检测过程中遇到的负面因素,如天色昏暗、目标物距离较远或体积较小等;修改特征融合的网络层,对较靠前的网络层输出图像额外进行拼接识别操作,这是因为未经过多层卷积的特征图像尺度更大,具有更清晰的特征,有助于提高目标识别的精度。

1 YOLOv5 模型

YOLOv5是一种目标检测模型,它继承了YOLO系列的特点,并在多个方面进行了改进和优化。同时,YOLOv5以其优异的性能、易用性和广泛的适用性,在多个领域得到了广泛的应用,不仅可以实时监控交通流量,检测违章行为,还可以作为城市安防系统和公共设施的监控工具,提高城市的安全管理水平。此外,YOLOv5还被应用于农业和工业中的产品检测和缺陷识别,大大提高了生产效率。

YOLO基于端到端的理念,将目标识别任务视为一个回归问题,并通过整个图像的全部像素直接获得边界框的坐标、图像所包含的物体、可靠性和类别概率。在此之前的对象检测方法主要依靠大量的潜在边界框,利用潜在边界框来框取候选区域标识的对象,从而判断对象类别的概率或可靠性。YOLOv5模型主要包括5个处理步骤。

(1)图像预处理。输入的图像都会经过预处理步骤,包括大小调整、归一化、增强等,以便于后续模型处理。

(2)特征提取。预处理后图像会被传输到模型中经过Backbone层,该网络可以有效地从图像中提取特征信息。

(3)特征图处理。在Backbone层的基础上,YOLOv5通过添加一系列卷积层和池化层,将特征图进行进一步处理,以提取更加高级的特征信息。

(4)目标检测。通过在特征图上运行卷积操作,YOLOv5可以同时预测图像中多个目标的位置

和类别。它将图像分成不同的网格,并在每个网格上预测边界框(bounding box)和相应的类别概率。

(5)后处理。在目标检测完成后,YOLOv5 会进行一些后处理步骤,如非极大值抑制(NMS),以去除重叠较多的边界框,并保留置信度较高的边界框。

YOLO 系列模型的特点是速度快、体积小、结构简单,在保持高检测速度的同时大大提高了检测精度,其中 YOLOv5 结合了 YOLOv3-SPP 和 YOLOv4 的一些功能,性能上有了更高的提升。故选用 YOLOv5s 作为车辆小目标检测模型进行实验。

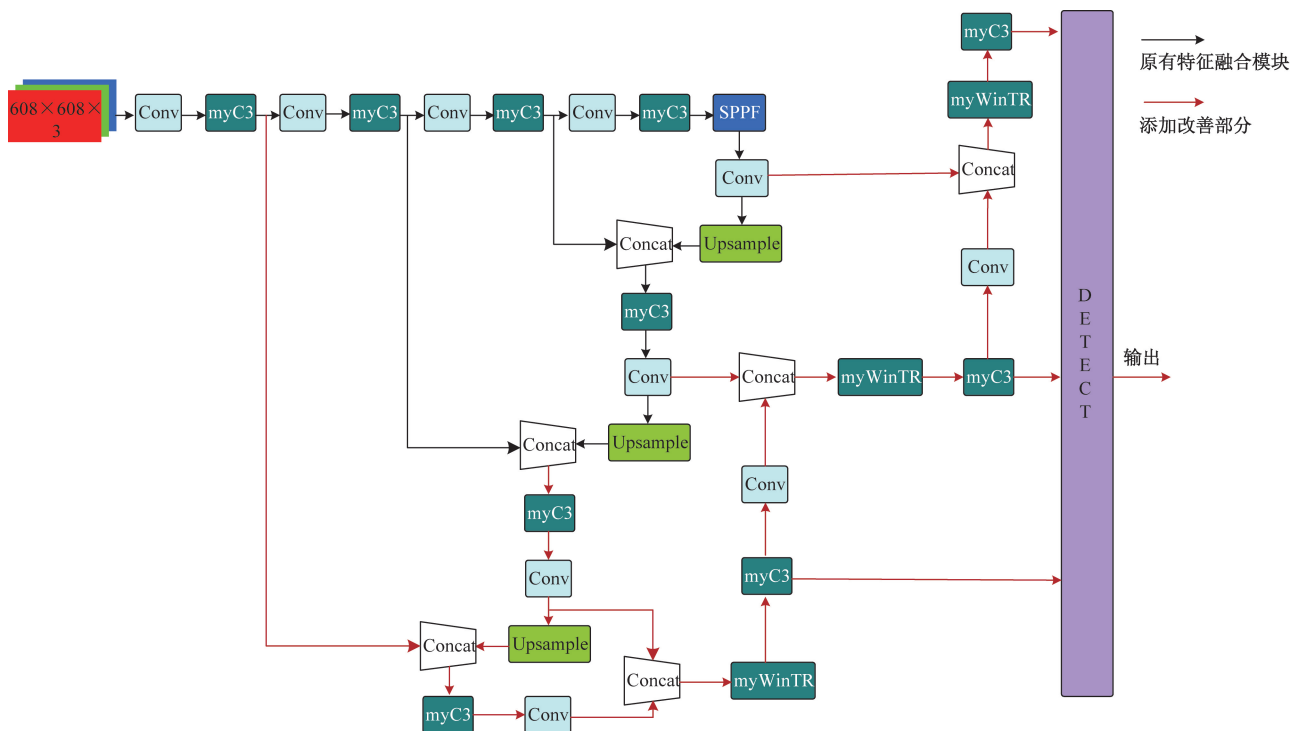
2 Atiny-YOLO 模型

在实际检测过程中,当有两个或更多数量的类似车辆目标出现在同一图像中时,YOLOv5 网络往往会误判它们属于同一类别,为了提高模型的检测精度,提出一种改善特征融合的车辆目标检测算法 Atiny-YOLO。首先,输入端经过 4 倍、8 倍、16 倍、32 倍下采样后,分别得到了 4 种包含不同尺度空间信息和语义信息的特征图,并在 Neck 层进行特征融合,增强了网络对多尺度目标的检测能力;在 Backbone 层中,Atiny-YOLO 嵌入所提出的 myC3 和 myWinRT 模块,其中,myC3 模块通过卷积操作有效地提取前景特征信息,同时抑制背景特征信息,而 myWinRT 模块则利用窗函数技术,进一步增强前景特

征的表达能力,从而提升了模型对目标的辨识能力;Atiny-YOLO 在网络层的设计上进行了调整,使得结果在输出前能够与初始特征图进行拼接。由于初始特征图包含大量的底层信息,这样可以有效地减少特征信息在传递过程中的损失,在输出结果中保留更多的细节信息,提高了检测结果的精确度和鲁棒性。Atiny-YOLO 模型结构如图 1 所示。

2.1 小目标检测层

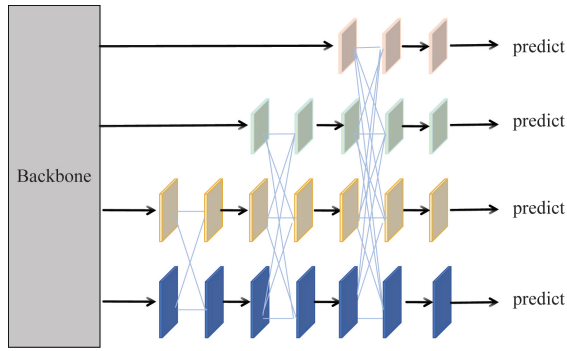
为了提高模型对小目标特征信息的捕捉能力,在 YOLOv5s 的 Neck 层中增加一个小目标检测层。原始网络有 3 个检测头,分别为 20×20 (小目标)、 40×40 (中目标)、 80×80 (小目标),要在 80×80 上层增加小目标检测层,也就是增加 160×160 尺寸。由于原始 Neck 网络中没有 160×160 的特征图,需要对 80×80 的特征图进行上采样得到 160×160 的特征图,随后将这个特征图与 Backbone 层的第三层输出进行融合,确保它们在通道数和大小上匹配。这样合并的特征图连同其他检测层的输出一同传入 Head 层进行后续的分类和检测处理。由于 Neck 层生成了多个不同尺度的特征图,每个特征图对应不同大小的锚点,使得模型在检测过程中能够在较大的特征图上检测较小的物体,在较小的特征图上检测较大的物体,从而增强了对图像中不同尺寸物体的表示能力。小目标检测层流程如图 2 所示。



Conv 为卷积;myC3 为本文改进 C3 模块;Concat 为相加模块;Upsample 为上采样模块;myWinTR 为本文改进 Transformer 模块;SPPF 为空间金字塔快速池化模块;DETECT 为检测层

图 1 Atiny-YOLO 模型结构

Fig. 1 Atiny-YOLO model structure



Backbone 为主干网络; predict 为预测输出
图 2 小目标检测层

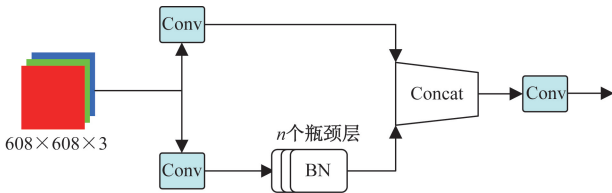
Fig. 2 Small target detection layer

2.2 C3 模块

在 YOLOv5 网络结构中,作用在 Backbone 层和 Head 层的 C3 模块取代了传统的 Bottleneck CSP 模块, C3 模块中包含 3 个标准卷积层以及多个 Bottleneck 模块,在学习残差特征过程中, C3 模块会分为两个分支进行,一个使用标准卷积层和叠加起来的 Bottleneck,另一个仅通过基本卷积模块,最后将两个分支整合拼接得到目标特征图。C3 模块结构图如图 3 所示。

本文算法在原基础上加入 split 方法,将卷积层后输入的 n 个 BN 层进行拆分,使其分别输出图像并添加一次拼接操作,将改进后的 C3 定义为 myC3。myC3 模块结构图如图 4 所示。

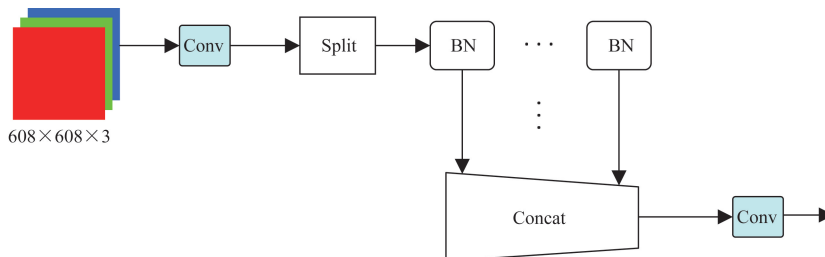
myC3 模块对批 BN 层输出的所有图像进行拼接操作,以实现细节的融合,从而增加模型接收到的特征信息的丰富性,并显著提升车辆识别的准确率。



Conv 为卷积; BN 为瓶颈层; concat 为相加模块

图 3 C3 模块结构图

Fig. 3 Structure of C3 module



Conv 为卷积; Split 为分割操作; BN 为瓶颈层; Concat 为相加模块;

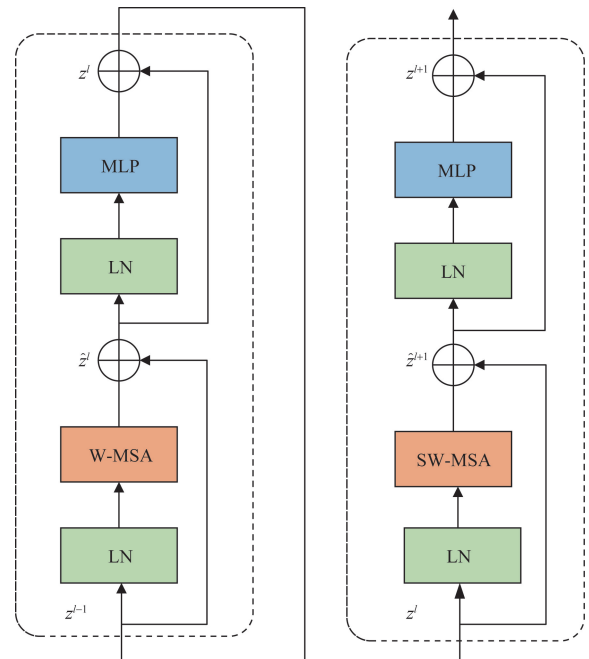
图 4 myC3 模块结构图

Fig. 4 Structure of myC3 module

2.3 Transformer 模块

传统的 Transformer 框架在目标检测过程中主要有两个问题,一是当不同场景的视觉对象发生显著变化时,Transformer 的性能会下降;二是当图像分辨率较高、像素点的数量较大时,Transformer 所使用的基于全局自注意力的计算方式会使计算量增大。文献[19]在 Transformer 的基础上提出了一种融合滑动操作且具有分层设计的 Swin Transformer, Swin Transformer 模块如图 5 所示。

Swin Transformer 在每一次卷积完成后选取 2×2 组特征向量进行融合和压缩。本文算法在 Swin Transformer block 中嵌入 3 层卷积网络,并将其定义为 myWinTR,目的是提高网络结构的稳定性和泛化

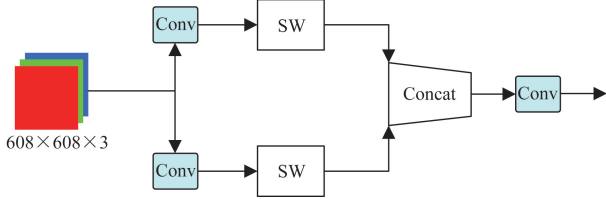


MLP (multi-layer perceptron) 为多层感知机,用于提取特征; LN (layer normalization) 为正则化方法,用于标准化输出; W-MSA (Window based multi-head self-attention) 和 SW-MSA (Shifted Window based multi-head self attention) 均为多头自注意力模块; z^l 和 z^{l+1} 分别为第 l 个 block 的 MLP 模块输出特征和 W-MSA 模块输出特征

图 5 Swin Transformer 模块结构图

Fig. 5 Structure of Swin Transformer module

能力,避免图片过度拟合,同时减少权重参数在运算中的占比,降低计算量,从而提高运算速率。my-WinTR 模块结构图如图 6 所示。



Conv 为卷积;SW 为多头自注意力模块;Concat 为相加模块

图 6 myWinTR 模块结构图

Fig. 6 Structure of myWinTR module

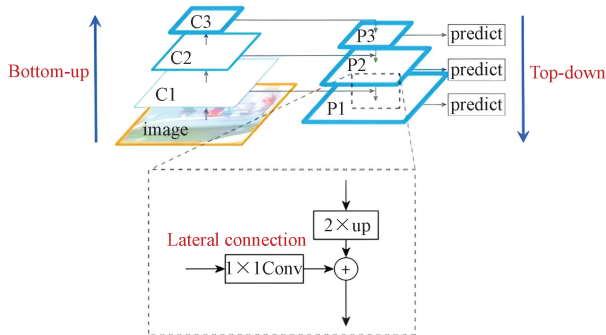
当待检测图片中出现距离较远或目标物体体积较小等情况时,myWinTR 能够更准确地识别和捕捉到目标物,对车辆特征的提取有很大帮助。

2.4 特征融合

在 YOLOv5 中通常使用 FPN 结构对 Backbone 所提取的特征图像进行增强融合处理,使后期对目标物体的提取更加准确。FPN 结构如图 7 所示,FPN 使用自上而下的特征融合方法来解决目标识别过程中的多尺度变化问题,但 FPN 在特征融合的过程中通常会遗漏许多小目标特征信息。

本实验在原有的 FPN 结构上进行添加改进,使用 myC3 模块替换 FPN 框架中的 C3 模块,并在后续的网络框架中加入 myWinTR,最终能够有效提高对距离较远和体积较小的难检测物体的识别效率。

将图片输入层看作第零层,紧邻的卷积为第一层,以此类推。当第 4 层的 myC3 输出的图片与第 15 层的上采样特征图完成拼接操作后,模型会将输出结果重新输入新添加的 myC3 层中,继续添加卷积神经网络并进行上采样。此时网络框架已经叠加到 19 层,将得到的结果重新与第 2 层的 myC3 输出结果进行拼接操作,得到的输出再一次经历 myC3



Bottom-up 为从下向上提取图像信息;Top-down 为从上向下合并特征图信息;Lateral connection 为横向连接特征信息;C1、C2、C3 为 3 个卷积层,用于特征提取;P1、P2、P3 为 3 个预测层,用于生成预测结果

图 7 FPN 结构图

Fig. 7 FPN structure

层和卷积神经网络,之后加入 myWinTR 层,最终需要将各层相互卷积和拼接得到的结果进行整理。

在添加的网络框架中,第一次拼接和第 2 层的输出是同时进行的,这是因为第 2 层 myC3 输出的特征图大小为 160×160 ,它更贴合输入图片的特征信息,未经过多层卷积神经网络的特征图尺度更大,特征更清晰,有助于更高效的识别目标物体。最后加入的 Detect 层,相当于给每个要输出的 myC3 层增加一个改变通道数的卷积层,它能使对应的输出通道变为 $3 \times (n + 5)$,即使 myC3 层的层数不同,也能保持不同层输出的图片通道数相同,方便后续实验操作。

3 实验结果及分析

3.1 实验配置及评价指标

3.1.1 实验数据集

选择 AU-AIR 数据集作为实验数据集,AU-AIR (aerial universal autonomous inspection and recognition) 数据集是第一个用于目标检测的多模态无人机数据集,它满足无人机视觉和机器人技术的要求。实验中使用的图像均采集自 AU-AIR 原始视频中的截图,通过人工筛选规格为 640×640 的图像并进行人工标注,随后对图像进行批量旋转,扩充训练数据集,增加样本的多样性和数量,通过旋转可模拟不同角度和视角下的图像变化,使模型能够更好地适应不同的旋转变换。最终得到 4 980 张图片作为数据集,实验中设定每次随机选择 360 张图片进行测试。数据集如图 8 所示。



图 8 实验数据集

Fig. 8 Experimental data set

3.1.2 实验环境及参数

本实验所使用的操作系统为 Windows 11,深度学习框架为 Pytorch 和 TorchVision,GPU 的型号是 RTX4090(24 GB),CPU 的型号是 AMD EPYC 7T83。实验选择随机梯度下降算法(SGD 算法)更新权值,

学习衰减策略采用 \cosin , 网络的求解参数设置如表 1 所示。

表 1 实验参数设置(SGD 优化器)

Table 1 Experimental parameter settings (SGD optimizer)

参数	数值
初始化学率	0.01
动量	0.937
权重衰减	0.0005
迭代次数	50
Batch size	20

3.1.3 评价指标

针对遥感图像中车辆目标检测任务,使用如下 4 种性能评价指标来全面评估模型在遥感图像中的表现。

(1) 准确率 P 。衡量模型检测到的车辆目标中有多少是真正的车辆目标,准确率越高,说明检测过程中误检现象少。通过计算真正例(TP)与所有预测为正确例(TP + FP)的比值得出。

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

(2) 召回率 R 。衡量所有正确的车辆目标中有多少被正确检测的比例,召回率越高,说明正例越少被漏检。通过计算真正例(TP)与所有正例(TP + FN)的比值得出。

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

(3) 平均精度 mAP。综合评估模型在多个类别上的性能。mAP 计算每个类别的 AP,并取所有类别 AP 的平均值,能够反映不同类别的检测结果。

$$AP = \int_0^1 P(R) dR \quad (3)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k(\text{classes})} \quad (4)$$

式中: AP_i 为第 i 张图片的检测精度; $k(\text{classes})$ 为类别的总数。

(4) IoU 阈值 mAP(mAP@0.5 和 mAP@0.5-0.95)。由于目标检测任务中使用交并比(intersection over union, IoU)来评估预测结果的正确性,因此可以在不同的 IoU 阈值下计算 mAP。其中,mAP@0.5 表示在 IoU 阈值为 0.5 时的平均精度,mAP@0.5-0.95 表示在 0.5 到 0.95 的 IoU 阈值上的平均 mAP 值。

(5) 参数量(Params)。Params 衡量模型大小和复杂度,较低的参数量有助于模型轻量化。

(6) 每秒传输帧数(FPS)。FPS 衡量算法处理速度的重要指标,表示模型每秒可以处理多少帧图

像,即图像的刷新频率。FPS 越高,模型才能满足实时检测的需求。

3.2 实验结果及分析

3.2.1 消融实验

为了更好地说明改进算法的优良性,同时探究本文算法 Atiny-YOLO 中每个改进策略对算法性能的贡献和作用,设计以下消融实验,实验结果如表 2 所示。

表 2 消融实验结果对比

Table 2 Comparison of ablation experiment results

检测层	myC3	mywinTR	特征融合	mAP@0.5/ %	Params/ M	FPS/ (帧·s ⁻¹)
				92.6	7.0	352
✓				94.5(+1.9)	15.1	120
✓	✓			94.6(+0.1)	16.5	134
✓	✓	✓		95.1(+0.5)	15.9	150
✓	✓	✓	✓	95.5(+0.4)	13.5	234

注:✓表示采纳当前改进策略。

由表 2 可知,Head 层中额外增加一个小目标检测层后,由于更多低层特征的语义信息得到保留,有利于确定小目标车辆的位置,相较于原始算法,mAP 提升了 0.9%,参数量增加导致模型在检测速度方面有一些损失,达到 120 帧/s;将 C3 模块替换成 myC3 后,模型获得更丰富的梯度信息且推理速度得到加快,在 mAP 提升 0.1% 的同时,检测速度也有所提升;得益于 myWinTR 提高了在复杂情况下对目标的检测率,增加 myWinTR 模块后,检测精度得到了提升;通过改善特征融合通道,模型的检测速度有了很大提升,最终模型的 mAP 较基线模型提升了 2.9 个百分点。

3.2.2 对比实验

为了验证 Atiny-YOLO 模型检测性能的优越性,将其与 YOLOv5 系列基准模型 YOLOv5s、YOLOv5l、YOLOv5n、YOLOv5x 进行对比。同时,目前主流的一些深度学习的方法在目标检测领域也取得了显著的成绩,选择单阶段模型 SSD、双阶段模型 Faster R-CNN、轻量模型 EfficientDet-D0、MobileNet V3L 和 TPH-YOLOv5 模型与 Atiny-YOLO 进行对比实验。对上述算法从 Precise、Recall、mAP@0.5、mAP@0.5-0.95(%)、参数量(Params)和每秒传输帧数(FPS)6 个评价指标进行对比,实验结果如表 3 所示。

由表 3 可知,Atiny-YOLO 的准确率达到 97.2%,对比其他基线模型,准确率均有不同程度的提升,这意味着 Atiny-YOLO 在识别车辆目标时更加精准,这得益于 Atiny-YOLO 模型中额外的小目标检测层,减少了误判的可能性;从实验结果可以看出,额外增加小目标层使得 Atiny-YOLO 的参数量

表3 不同模型性能对比

Table 3 Comparison of the performance of different models

模型	P/%	R/%	mAP@0.5/%	mAP@0.5-0.95/%	Params/M	FPS/(帧·s ⁻¹)
SSD	92.3	78.2	74.1	47.8	24.1	212
Faster R-CNN	63.5	90.8	76.5	49.5	15.7	73
EfficientDet-D0	75.3	85.0	71.2	46.7	3.9	29
MobileNet V3L	88.1	84.4	80.3	48.8	4.2	89
TPH-YOLOv5	91.0	83.7	89.5	32.1	45.4	208
YOLOv5s	94.3	91.9	92.6	59.5	7.0	352
YOLOv5l	94.6	92.6	93.5	62.8	46.2	176
YOLOv5n	94.7	90.4	93.2	55.1	1.8	256
YOLOv5x	95.4	92.0	94.7	63.6	86.2	291
YOLOv6	96.5	93.0	94.4	59.8	4.1	180
Atiny-YOLO	97.2	92.1	95.5	59.9	13.5	234

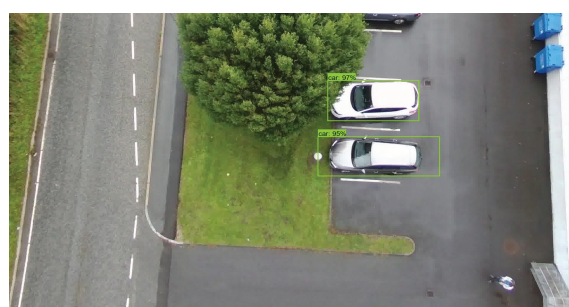
(Params)有了小幅度的提升,但是 YOLOv5s 作为轻量模型尽管增加了一定的模型复杂度,检测速度 FPS 依旧能够满足检测需求,在增加少量参数的同时能够获得显著的性能提升,这说明 Atiny-YOLO 在参数量和检测性能中实现了平衡,模型的改进是有效的;同时,Atiny-YOLO 的 mAP 达到了 95.5%,较 YOLOv5s、YOLOv6、Faster R-CNN 和 MobileNet V3L 分别提升了 2.9 个百分点、1.1 个百分点、19 个百分点和 15.2 个百分点,mAP 值提升明显,从侧面验证算法能够降低误检,证明了本文算法的有效性;其中 TPH-YOLOv5 模型是专门针对无人机场景目标信息捕获的算法,本文算法相较于 TPH-YOLOv5 模型在检测精度和检测速度上均有不同程度的提升。实验过程中还发现双阶段检测模型 Faster R-CNN 在验证集上的检测效果较差,这是因为 Faster R-CNN 对小目标数据特征的捕捉能力较差,导致训练过程中欠拟合。

3.2.3 检测效果对比

为了能更直观地展现所改进的算法的性能,选择数据集中有代表性的 3 张图像,图像(a)包含清晰车辆目标,同时图像边缘有少量车辆目标信息;图像(b)包含远距离目标,受天气原因车辆目标不清晰;图像(c)天气状况较好,图像右上角存在大量目标密集,同时目标尺寸较小,对目标检测算法要求最高。图 9~图 11 展示了不同算法在检测过程中的表现。

Faster R-CNN 模型能够针对无人机遥感图像中的车辆目标实现检测识别,但针对目标信息缺失,目标展示不完全的情况,Faster R-CNN 只能检测到部分目标物,依旧存在大量目标物漏检的情况,同时模型反馈的准确率数值有所下降。

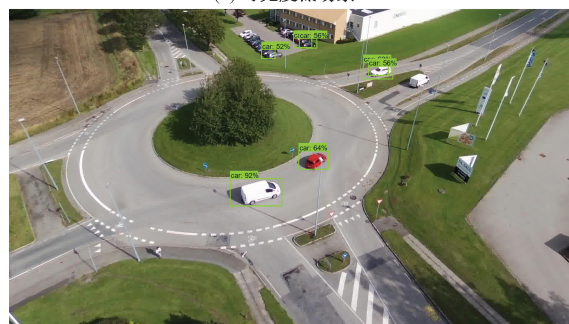
与 Faster R-CNN 模型相比,在目标物较小、距离较远、天色较昏暗、目标物被不完全遮挡等负面



(a) 边缘目标场景



(b) 可见度低场景

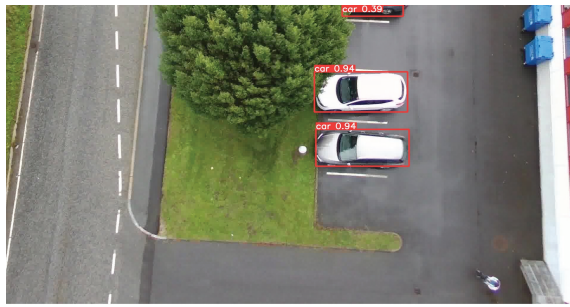


(c) 小目标场景

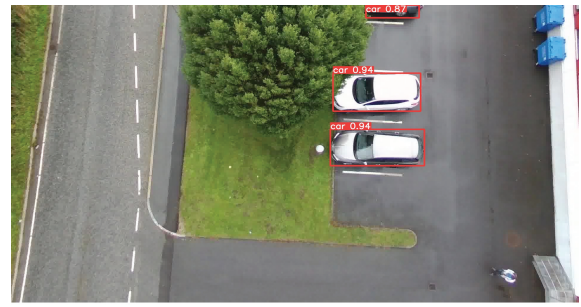
图9 Faster R-CNN 模型检测结果

Fig. 9 Faster R-CNN model detection results

因素影响之下,YOLOv5s 模型检测精度和检测效果都得到了显著增强。在 Faster R-CNN 模型实验中目标漏检情况相比,YOLOv5s 模型实验中的这种问题得到了明显的改善,不易检测的目标物精确度系



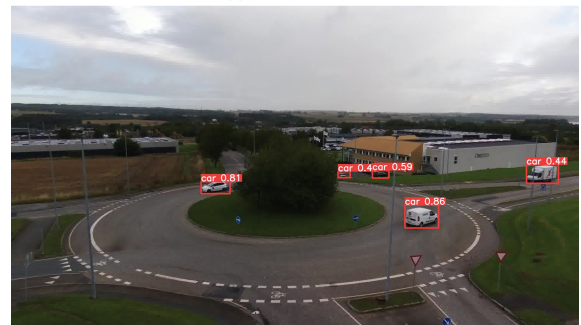
(a) 边缘目标场景



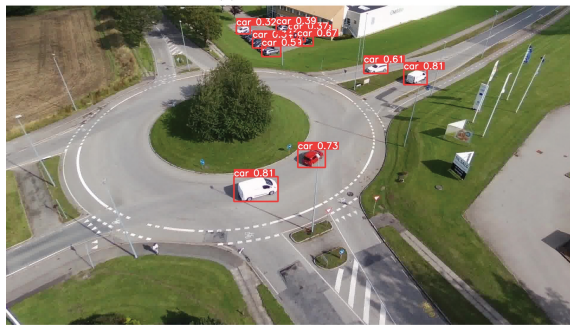
(a) 边缘目标场景



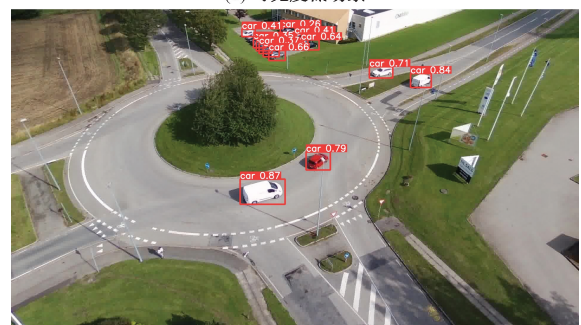
(b) 可见度低场景



(b) 可见度低场景



(c) 小目标场景



(c) 小目标场景

图 10 YOLOv5s 模型检测结果

Fig. 10 YOLOv5s model detection results

图 11 Atiny-YOLO 模型检测结果

Fig. 11 Atiny-YOLO model detection results

数得到提升,但当图像中存在远距离目标密集时, YOLOv5s 同样存在目标漏检情况。

Atiny-YOLO 模型在检测过程中对汽车的识别和检测的精确度得到了进一步提升,对于图像边缘的车辆信息捕捉能力较强,尤其是在目标密集场景中表现出色,检测精度较前两种模型均有不同程度的改善,这表明本文改进算法相较于原始算法具有较好的检测能力。

针对图 11(c) 中存在的不同类型车辆目标,包括厢式货车、轿车和电动代步车等, Atiny-YOLO 模型均能进行检测,两个厢式货车的检测置信度分别为 0.87 和 0.84,可见 Atiny-YOLO 模型对特征信息的抓取能力强,能够在训练过程中学习到车辆目标的有效特征,使之在训练集之外的未知数据上也能表现出色,即模型泛化能力较强。

为了检验 Atiny-YOLO 模型对动态特性的融合能力,选取一条 18 s 无人机拍摄视频传入模型进行检测,实验结果如图 12 中视频截图所示。

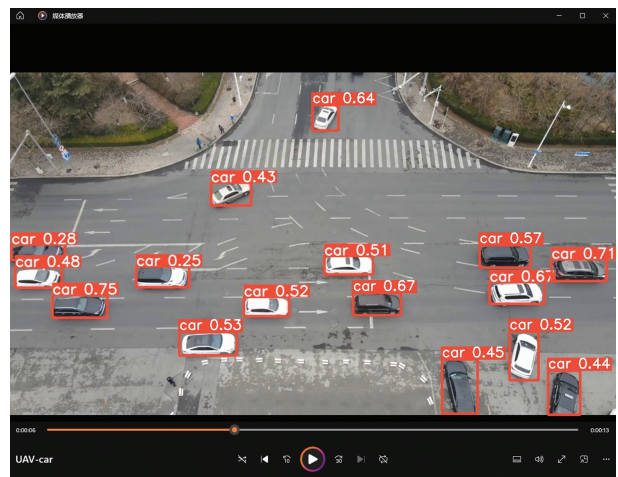


图 12 视频流检测结果

Fig. 12 Video stream detection results

由图 12 可知,模型在检测视频过程中没有出现误检和漏检情况,画面中出现的车辆在无遮挡,特征信息显示完全的情况下均能正常检测,由于视频

中车辆目标变化快,模型对车辆检测的置信度相对图像检测时较低,但可以基本满足持续追踪视频中车辆目标的实际要求,这说明 Atiny-YOLO 模型对车辆目标的检测能力较强,在图像和视频流中均有良好表现。

4 结论

提出一种改进 YOLOv5s 算法的车辆目标检测算法,有效提升了无人机遥感图像中车辆小目标的检测精度,同时降低了误检和漏检的概率。通过实验对比发现,所改进的 Atiny-YOLO 模型在实际检测场景中检测效果更好,检测精度相较原始算法有明显提升,同时对检测速度未造成显著影响,基本满足实际检测过程中的要求。Atiny-YOLO 模型的泛化性和鲁棒性较好,具有一定的实用价值和工程意义。在后续的研究中,需要对现有的深度神经网络做更多训练,提高车辆目标检测的颗粒度,以此更好的分析交通网络现状。

参 考 文 献

- [1] 欧阳凯. 基于测绘工程测量中无人机遥感技术运用[J]. 工程建设与设计, 2022(22): 96-98.
Ouyang Kai. Application of UAV remote sensing technology in surveying and mapping based engineering measurement[J]. Engineering Construction and Design, 2022(22): 96-98.
- [2] 胡义强, 杨骥, 荆文龙. 基于无人机遥感的海岸带生态环境监测研究综述[J]. 测绘报, 2022(6): 18-24.
Hu Yiqiang, Yang Ji, Jing Wenlong. A review of coastal zone ecological environment monitoring research based on UAV remote sensing[J]. Surveying and Mapping Journal, 2022(6): 18-24.
- [3] Qin J X, Yang W J, Wu T, et al. Incremental road network update method with trajectory data and UAV remote sensing imagery[J]. ISPRS International Journal of Geo-Information, 2022, 11(10): 502.
- [4] Li J, Wei X. Research on efficient detection network method for remote sensing images based on self attention mechanism[J]. Image and Vision Computing, 2024, 142: 104884.
- [5] Nan H H, Tianyi Z, Tung C Y, et al. Image segmentation using transfer learning and Fast R-CNN for diabetic foot wound treatments [J]. Frontiers in Public Health, 2022, 10: 969846.
- [6] Yang Y. Vehicle target detection algorithm based on improved faster R-CNN for remote sensing images[J]. Journal of Artificial Intelligence Practice, 2024, 7(1): 7958815.
- [7] Yang X, Xiu J, Liu X. Research on improved YOLOv5 vehicle target detection algorithm in aerial images[J]. Drones, 2024, 8(5): 103390.
- [8] Zhao X, Wang Q, Zhang M, et al. CSFF-YOLOv5: improved YOLOv5 based on channel split and feature fusion in femoral neck fracture detection[J]. Internet of Things, 2024, 26: 101190.
- [9] Yuan H, Lu Z, Zhang R, et al. An effective graph embedded YOLOv5 model for forest fire detection[J]. Computational Intelligence, 2024, 40(2): DOI: 10.1111/coin. 12640.
- [10] Wang H L, Qi H M, Feng S, et al. L-SSD: lightweight SSD target detection based on depth-separable convolution[J]. Journal of Real-Time Image Processing, 2024, 21(2): 2547-2561.
- [11] 袁小平, 马绪起, 刘赛. 改进 YOLOv3 的行人车辆目标检测算法[J]. 科学技术与工程, 2021, 21(8): 3192-3198.
Yuan Xiaoping, Ma Xuqi, Liu Sai. Improved pedestrian-vehicle target detection algorithm for YOLOv3 [J]. Science, Technology and Engineering, 2021, 21(8): 3192-3198.
- [12] 姜宗文, 金立左. Vehicle-YOLO——一种基于航拍影像的车辆检测模型[J]. 微型电脑应用, 2023, 39(9): 134-137.
Jiang Congwen, Jin Lizuo. Vehicle-YOLO: a vehicle detection model based on aerial images [J]. Microcomputer Applications, 2023, 39(9): 134-137.
- [13] 涂媛雅, 汤国放, 张建勋. Lite-YOLOv3 轻量级行人与车辆检测网络[J]. 小型微型计算机系统, 2023, 44(1): 211-217.
Tu Yuanya, Tang Guofang, Zhang Jianxun. Lite-YOLOv3 lightweight pedestrian and vehicle detection network[J]. Small Microcomputer Systems, 2023, 44(1): 211-217.
- [14] Dong X, Tong F X, Yang G, et al. A detection method of spangle defects on zinc-coated steel surfaces based on improved YOLOv5 [J]. The International Journal of Advanced Manufacturing Technology, 2023, 128(1/2): 937-951.
- [15] Yan Y, Liu Z, Xu J, et al. A temperature-decoupled impedance-based mass sensing using CBAM-CNN and adaptive weighted average preprocessing with high accuracy[J]. Mechanical Systems and Signal Processing, 2024, 21: 3111347.
- [16] Jin Y F, Gong D X, Zhao S Y, et al. A real-time fire detection method from video for electric vehicle-charging stations based on improved YOLOX-tiny[J]. Journal of Real-Time Image Processing, 2023, 20(3): 101007.
- [17] Xue L S, Du S H, Wu H T, et al. Defect signal intelligent recognition of weld radiographs based on YOLOv5-improvement [J]. Journal of Manufacturing Processes, 2023, 99: 373-381.
- [18] 蒲玲玲, 杨柳. 改进 YOLOv5 的多车辆目标实时检测及跟踪算法[J]. 科学技术与工程, 2023, 23(28): 12159-12167.
Pu Lingling, Yang Liu. Improved YOLOv5 real-time multi-vehicle target detection and tracking algorithm [J]. Science Technology and Engineering, 2023, 23(28): 12159-12167.
- [19] 王璨, 武新慧, 张燕青, 等. 基于移位窗口 Transformer 网络的玉米田间场景下杂草识别[J]. 农业工程学报, 2022, 38(15): 133-142.
Wang Can, Wu Xinhui, Zhang Yanqing, et al. Weed recognition in corn field scenarios based on shift-window transformer network[J]. Journal of Agricultural Engineering, 2022, 38(15): 133-142.