

文章编号:1001-5078(2023)12-1877-08

· 红外技术及应用 ·

# 航空机载红外图像的车辆目标自主检测识别

杨 雪<sup>1,2</sup>, 修吉宏<sup>1</sup>, 刘小嘉<sup>1,2</sup>, 罗 宁<sup>1,2</sup>

(1. 中国科学院长春光学精密机械与物理研究所, 吉林 长春 130031; 2. 中国科学院大学 北京 100190)

**摘要:** 红外光学成像载荷利用目标的热辐射强度特性成像, 具有一定的揭伪能力, 可规避可见光成像装备无法在夜间和恶劣气象条件下成像的限制, 但红外图像对比度低、边缘不清晰, 大大降低了成像目标识别的准确率。本文提出一种基于 YOLOv5 的红外车辆目标检测算法, 在浅层特征层引入 RFBs 模块, 以提高小目标的感受野及检测效果, 在颈部网络(Neck)部分, 使用 BiFPN 结构, 实现对底层特征的再次利用, 以融合更多的特征, 并添加 CBAM 注意力机制以提升检测精度。实验结果表明: 在 DroneVehicle 数据集上的检测效果要优于原始网络, 精确率(Precision)提升 2.8%, 召回率(Recall)提升 16%, 平均精度(mAP)提升 2.3%。结论: 可有效应用于航空红外图像的车辆自主检测识别。

**关键词:** 目标检测; YOLOv5; 红外图像; RFBs; BiFPN; 注意力机制

**中图分类号:** TP394.1.4; TN219 **文献标识码:** A **DOI:** 10.3969/j.issn.1001-5078.2023.12.014

## Autonomous vehicle target detection and recognition from airborne infrared imagery

YANG Xue<sup>1,2</sup>, XIU Ji-hong<sup>1</sup>, LIU Xiao-jia<sup>1,2</sup>, LUO Ning<sup>1,2</sup>

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130031, China;

2. University of Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** Infrared optical imaging payloads make use of the thermal radiation intensity characteristics of the target for imaging, which has a certain degree of falsification capability and can circumvent the limitations of visible light imaging equipment that cannot be imaged at night and under adverse meteorological conditions. However, the infrared image has low contrast and unclear edges, which greatly reduces the accuracy of imaging target recognition. In this paper, an infrared vehicle target detection method is proposed based on YOLOv5. Firstly, the RFBs module in the shallow feature layer is introduced to improve the receptive field and detection effect of small targets. Secondly, the BiFPN structure is used to realize the re-use of the underlying features in order to fuse more features, and the CBAM attention mechanism is added to improve the detection accuracy in the neck network(Neck) part. The experimental results show that the detection effect on the Drone Vehicle dataset is better than that of the original network, the precision rate(Precision) is improved by 2.8%, the recall rate(Recall) is increased by 16%, and the average precision(mAP) is increased by 2.3%. It is concluded that it can be effectively applied to the autonomous detection and recognition of vehicles in aerial infrared images.

**Keywords:** target detection; YOLOv5; infrared image; RFBs; BiFPN; attention mechanism

**基金项目:** 粤港关键领域重点突破项目(No. 2004A10403021); 广东省攻关项目(No. 2006A10401006)资助。

**作者简介:** 杨 雪(1997-), 女, 硕士研究生, 主要从事深度学习、图像处理方面的研究。E-mail: yangxue211@mails.ucas.ac.cn

**通讯作者:** 修吉宏(1976-), 女, 研究员, 博士生导师, 主要从事航空成像中的动态成像补偿、目标定位算法技术及相关图像处理技术研究。E-mail: xiujihong0809@163.com

**收稿日期:** 2023-03-17

## 1 引言

航空成像具有高时效、低成本等突出优势,通过有人或无人机搭载对地/海面目标成像,是实现侦察与监视的重要手段。红外光学成像载荷工作波长较长,具有较强的穿透特性,具备全天时成像探测能力。但与可见光图像相比,红外图像信噪比低、灰度范围窄,且由于航空机载红外成像的照相距离较远,其成像目标尤其是汽车等小目标在红外图像中所占有效像素较小,常常淹没在复杂背景杂波中,大大增加了红外目标的检测难度。此外,车辆目标容易受到场景变化影响,特别是随着光照强度、气候条件、地形等变化,目标在复杂背景下的显著性差异较大。如何快速有效地在复杂背景条件下实现红外图像中汽车等小目标的检测与识别具有重要意义。

传统的目标检测算法以手工设计特征为主,在复杂场景中的泛化能力较弱,具有一定的局限性,如 Sobel 边缘检测特征、DPM 等。2012 年的 AlexNet 使得目标检测开启了新的发展方向<sup>[1]</sup>。随着深度学习的发展,目标检测算法主要分为单阶段检测和双阶段检测两类,双阶段算法首先产生区域候选框,再对每个候选框提取特征,如 R-CNN<sup>[2]</sup>、SPPNet<sup>[3]</sup>、Faster R-CNN<sup>[4]</sup>等,精度高但检测速度较慢,存在大量冗余的候选框。单阶段检测算法产生候选框并进行分类和边界框检测,如 SSD<sup>[5]</sup>、RetinaNet<sup>[6]</sup>、YOLO 系列等,在速度快的同时,随着算法不断地优化改进,检测性能及精确度也在不断地提高,被广泛应用于目标检测、目标追踪等任务中。

航空红外图像目标检测的重点是提高图像中物体的检测精度,降低误检率及漏检率,且考虑到机载光电成像系统对目标信息获取和处理均有一定的时间要求,目标检测算法的执行时间尽可能短,YOLOv5 采用 CSP 结构,因而在模型部署的速度上有较大优势<sup>[7]</sup>,在自动驾驶等方面得到了广泛的应用,但其在提取特征时对特征图进行多次下采样,使得特征信息丢失过多,对红外条件下车辆小目标检测的效果并不理想。在此背景下,本文提出一种基于 YOLOv5 的 RBC-YOLOv5 检测模型,主要贡献如下:(1)针对 YOLOv5 检测小目标效果差的缺点,在浅层特征层 P3 引入感受野模块,丰富语义信息,增强感受野;(2)针对 YOLOv5

没有有效利用底层特征的缺点,通过重复双向交叉连接与加权的特征融合机制,加强对底层特征的利用;(3)通过改进颈部网络,增强对融合后特征的提取能力,引入注意力机制,将有限的资源分配给 YOLOv5 中重要的部分。

## 2 基本原理

RBC-YOLOv5 检测模型基于 YOLOv5 进行改进,本节主要对 YOLOv5 的基本原理及引入的各个模块结合航空红外图像特点进行介绍。根据其网络结构顺序,首先在浅层特征引入 RFBs 模块,提升对车辆小目标的检测精度;再添加 CBAM 模块,以提升模型对红外图像中模糊特征的注意力;最后使用 BiFPN 结构以实现更高层次的特征融合。

### 2.1 YOLOv5 基本原理

YOLOv5<sup>[8]</sup>是一个高性能单阶段目标检测器,作为 YOLO 系列的延伸,代表了对人工智能开源方向的研究,在组成的各个部分基于 YOLOv3 均有改进,使得达到实时性的同时,对红外车辆目标的检测效果也得到了显著提升<sup>[9]</sup>。

对输入图像进行基于 Cutmix 的数据增强方式 - 马赛克 (Mosaic) 增强<sup>[10]</sup>,增强背景的丰富度,扩充数据集中小样本的数量,使得小样本更小,在复杂的红外背景下,提高对车辆小目标的检测效果。YOLOv5 通过自适应图片操作,将不同大小的红外车辆输入图像处理为统一尺寸,相比于针对不同数据集都需要先通过聚类提取先验框尺度并设定特定长宽的 YOLOv4<sup>[11]</sup>,YOLOv5 自适应计算出最佳的锚框。

YOLOv5 在主干网络引入了 Focus 结构,采用切片操作,直接处理输入的图像,再经过拼接 (Concat) 与卷积 (Conv) 操作,最后在通道 (channel) 维度进行拼接,用来减少浮点运算计算量 (FLOPs) 并增加计算速度,达到对航空图像实时检测的效果。主要使用 SPPF 模块增大感受野,对特征图进行多尺度特征提取,提高模型精度,原理类似于空间金字塔池化 (Spatial Pyramid Pooling, SPP)<sup>[12]</sup>,可以忽略输入红外图像的尺寸并产生固定大小的输出,用不同的尺寸去提取特征,且不会带来运算速度的减少。

输出端 (Head) 主要由颈部 (Neck) 特征融合结构与检测部分 (Detect) 组成。特征金字塔 (Feature Pyramid Network, FPN)<sup>[13]</sup>与路径聚合结构 (Path

Aggregation Network, PAN)<sup>[14]</sup>相结合,将FPN自顶向下的高层的强语义信息传递给低层,再通过自底向上的PAN结构传达强定位特征,从而达到对不同的检测层进行参数聚合,有助于提升对航拍图像中不同大小车辆目标的检测效果。

2.2 RFBs 模块

该RFBs<sup>[15]</sup>模块模拟人类视觉神经的特点,参考了Inception的并行思想,使用空洞率(rate)为1、3、5的空洞卷积来调节不同大小卷积核,使之能够完成轻量化任务,提高了网络的特征提取能力。相较于原始RFB模块,既减少了参数量,又有利于捕捉更小的感受野,能够有效保持YOLOv5的轻量级网络的速度,又能够提升检测精度。RFBs模块的结构如图1所示,使用3×3、1×3、1×1的不同卷积组合,分别等价于3×3、5×5、7×7的卷积,再引入rate为1、3、5的空洞卷积提高感受野,通过拼接(Concat)对不同的感受野进行特征融合,最后使用1×1卷积提升模型的检测速度。

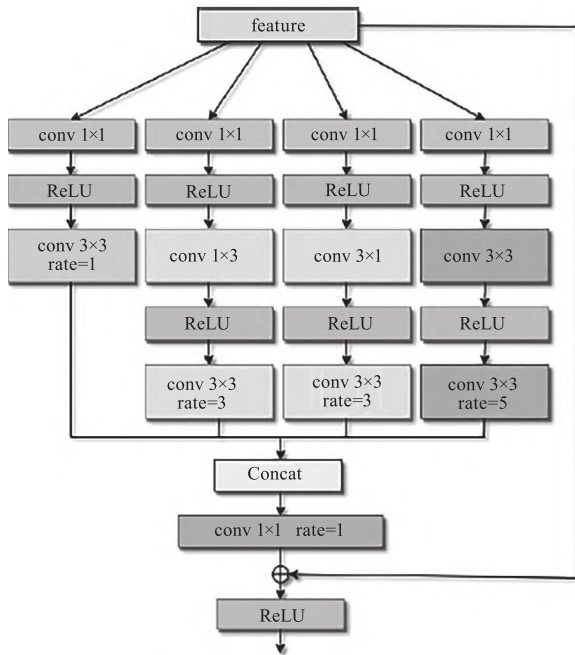


图1 RFBs 模块 Fig.1 RFBs module

2.3 CBAM 模块

通道注意力机制自适应的关注通道,对通道生成掩码 mask,每个通道都是特征图上每个特征点的特征浓缩,更加关注更高语义信息的特征,使其权重更高,如SENet<sup>[16]</sup>等。空间注意力机制自适应分辨空间中的区域,对空间进行 mask 的生成,网络自适

应的去关注位置,使关注点的权重更大,如 Spatial Attention Module。

CBAM 是一种轻量化的注意力模块,同时对通道和空间注意力进行打分,依次应用通道和空间注意力模块,进行自适应特征细化。CBAM 注意力模块的结构如图2所示,能够增强对目标特征的提取能力,使得网络有效关注提取区域,提高检测性能,并且可广泛应用于语音识别、自言语言处理等领域当中<sup>[17]</sup>。

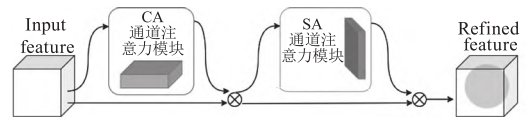


图2 CBAM 注意力模块 Fig.2 CBAM attention module

2.4 BiFPN

随着深度学习的发展,特征融合的方法在不断地改进。近年来使用较多的是PANet,特征图同时具备高语义信息和位置信息,有效提高目标检测等任务的精度,NAS-FPN<sup>[18]</sup>结构的检测效果好,但网络结构不规则且运行需要消耗大量计算资源,M2det<sup>[19]</sup>兼顾效率和速率,但参数过多。BiFPN<sup>[20]</sup>是一种高效的特征融合方法,具备自顶向下与自底向上的双向加权特征金字塔结构,同时传递高层的语义信息与底层的位置信息,并可学习不同输入特征的重要程度,重复的跳跃连接融合,进一步提升了检测精度。可看做是加强版的PANet,BiFPN结构图如图3所示。

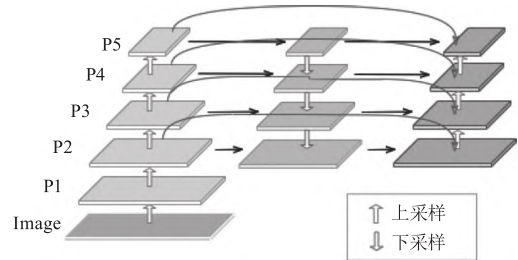


图3 BiFPN 结构 Fig.3 BiFPN structure

3 RBC-YOLOv5 模型

由YOLOv5包含YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l、YOLO5x五个模型变体,由左至右模型复杂度逐渐增加,参数量 depth\_multiple 与 width\_multiple 也在逐渐增加,检测精度越高,但检测速度越慢。本文选用YOLOv5s作为基础模型,节省

了一定的计算成本,检测速度快的同时兼具较高的检测精度。

针对像素点小、边缘特征不明显的航空机载红外图像,对 YOLOv5 进行改进,提出了 RBC-YOLOv5 模型结构,在浅层特征层引入 RFBs 感受野模块,提升小尺寸目标的检测精度,使用 BiFPN 结构实现更高层次的特征融合,添加 CBAM 注意力模块有效提升检测精度,使其更适用于对比度低和边缘不清晰的红外车辆小目标检测,其网络组件及网络结构如图 4 所示。

### 3.1 输入端(Input)改进

YOLOv5 使用 Mosaic 数据增强方法,本文在此基础上,引入 mixup 数据增强方法,将输入的两张图

像先进行一定的翻转、缩放,并按照一定比例进行混合,形成一张新的图像传入到网络当中,通过样本与标签的线性插值得到新的训练样本,可以用来提升模型的鲁棒性和泛化性,能够降低样本中存在的噪声,减轻过拟合现象。假设  $x_{i,j}$  与  $y_{i,j}$  为随机样本,  $\lambda$  的值为  $[0-1]$ , 则核心公式如(1)、(2)所示:

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \tag{1}$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \tag{2}$$

将图片预处理成  $640 \times 640$  大小传入网络中,并进行归一化等操作,采用改进后的数据增强方式提升网络计算速度及训练精度,利用自适应图片缩放操作及自适应锚框计算操作,将图像通过缩放比例缩放到适合于模型的大小并自适应计算出最佳的锚点框。

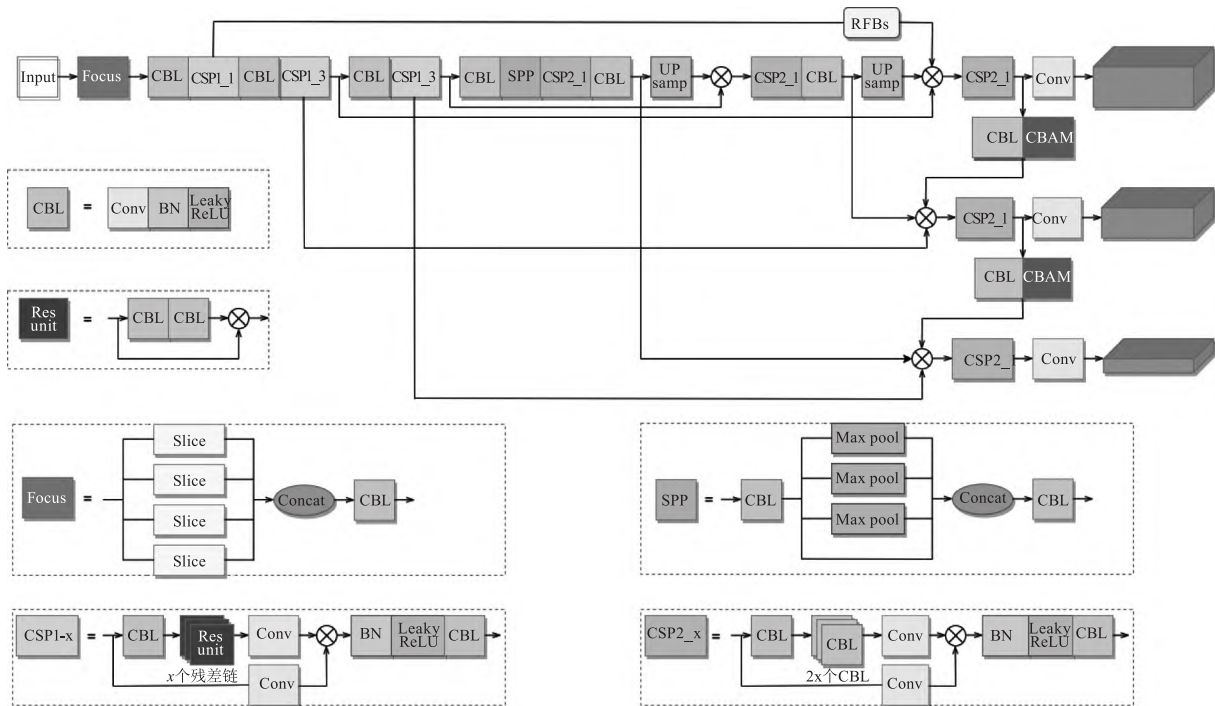


图 4 RBC-YOLOv5 模型结构图

Fig. 4 RBC-YOLOv5 model structure diagram

### 3.2 主干网络(backbone)改进

backbone 使用 CSPDarknet53 及 Focus 结构,主要使用 CBL 模块与 SPPF 模块,有效提升模型推理速度,在 YOLOv5 的多尺度目标检测算法中,共进行 5 次下采样,得到特征层  $P_1 \sim P_5$ ,  $P_{i-1}$  层为  $P_i$  层图像大小的  $1/2$ ,将  $P_3 \sim P_5$  传入 Neck 进行特征融合,但随着特征尺度的减小,目标经多次下采样的分辨率也在逐渐变小,进一步增加了红外车辆目标的检测

难度。

浅层特征的分辨率较大,可用于检测小目标,但语义信息不够丰富、感受野较小,为了进一步提升航拍红外图像中汽车等小目标的检测精度,在 backbone 的  $P_3$  层与融合后的  $P_{3,1}$  层之间引入 RFBs 模块,增强语义信息、增大感受野,提升 YOLOv5 对图像中车辆小目标的检测效果。RBC-YOLOv5 模型简化结构如图 5 所示。

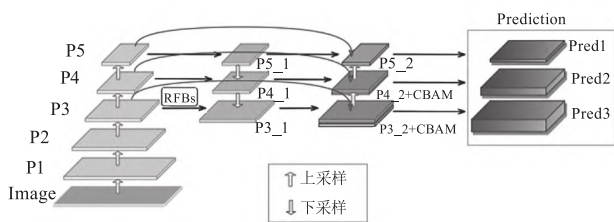


图5 RBC-YOLOv5 模型简化结构图

Fig. 5 RBC-YOLOv5 model simplified structure diagram

3.3 颈部网络(Neck)改进

在航拍红外车辆目标检测的场景下,虽然红外图像具备一定的揭伪能力,能够规避可见光图像无法在夜间成像的限制,但由于照相机距离地面较远,所拍摄物体所占像素较小,对于背景复杂且边缘模糊的目标检测具有一定难度。

使用原始 FPN + PAN 进行检测时自顶向下与自底向上相结合的融合计算量较大,并且没有融合 backbone 中所含有的原始特征。为尽可能的保留原始特征,并多次利用同一层,因此本文采用双向加权的 BiFPN,通过跨连接将  $P_i$  层特征传递给  $P_{i-2}$  层 ( $i = 3, 4, 5$ ),在不增加过多计算量的同时,提升对特征的重复有效利用,以实现更高层次的特征融合。

由于小尺寸目标只能在底层  $P_3$ 、 $P_4$  层被检测出来,为提升对小目标的检测效果和精度,根据 Hu<sup>[21]</sup> 等人所述,通过注意力机制模块增加小目标的上下文信息,可提升网络对小目标的检测率,因此在颈部网络融合后的特征中引入两个 CBAM 注意力模块,先通过通道注意力机制的全局平均池化、采用 ReLU 与 Sigmoid 激活函数的全连接,调整更具意义的通道信息的权重,提取出更加重要的通道特征,再经过空间注意力机制,在特征图的每一个点取最大值和平均值,经过 sigmoid,获得了每个特征点的比重,与输入进来的特征层相乘,也就是完成了空间注意力机制,使得在目标尺寸小、特征不明显的条件下,改善网络特征提取能力及特征融合能力。

4 实验分析

4.1 数据集

DroneVehicle 数据集<sup>[22]</sup>是 RGB/红外跨模式车辆检测数据集,涵盖不同的物体和场景,如城市道路、农村、住宅区、停车场和高速公路等,不同的高度、角度和光照条件来模拟现实中的数据分布。在不同高度获取的图像包含不同尺度的物体,在不同角度获取的图像显示出各种形状的物体,在不同光照条件下获

取的图像包含更多的阴影遮挡和亮度变化,各种复杂条件使得 DroneVehicle 数据集更具挑战性。

本文从中抽取部分红外图像,包括汽车(car)、卡车(truck)、公交车(bus)、货车(freight\_car)四类,共 28439 个样本,训练集(train)中包含 17990 张图像,测试集(test)中包含 8980 张图像,验证集(val)中包含 1469 张图像,数据可视化结果如图 6 所示。

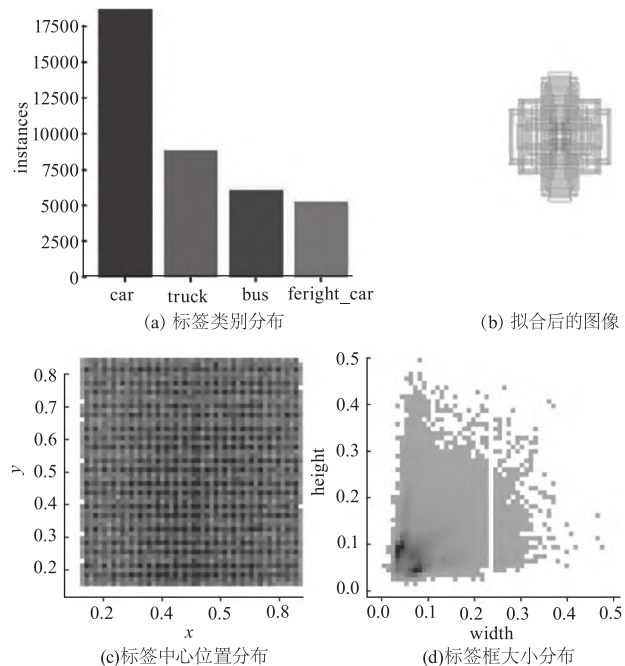


图6 数据可视化结果

Fig. 6 Data visualization results

4.2 网络设置与训练

本文所使用的配置为 NVIDIA GeForce GTX 1660,显存 6 GB,处理器(CPU)为 Intel Core i7-4770 @ 3.40GHz 八核,Python 版本 3.9.7,PyTorch 版本 1.7.1。

对训练数据进行数据增强,包含 Mix-up、Mosaic、Flip left-right、Hsv 等数据增强方法,以提高模型的泛化能力,具体参数如表 1 所示。

表 1 数据增强参数

Tab. 1 Data augmentation parameters

参数名称	大小
HSV-Hue	0.0103
HSV-Saturation	0.5562
HSV-Value	0.2826
Translation	0.0458
Scale	0.7371
Flipleft-right	0.5
Mosaic	0.8716
Mix-up	0.0429

使用 YOLOv5s 模型结构,设置深度倍数 depth\_multiple 为 0.33,宽度倍数 width\_multiple 为 0.5,改进后的 RBC-YOLOv5 模型的训练参数如表 2 所示。

表 2 模型训练参数

Tab. 2 Model training parameters

参数名称	大小
输入尺寸	[640,640]
批次大小	64
初始学习率	0.0033
最终学习率	0.1518
训练轮数	150
IoU 训练阈值	0.20
优化器	SGD
优化器权重衰减	0.00025

4.3 评价指标

为验证模型的性能,本文采用平均精度(AP),平均精度均值(mAP),精确率(P),召回率(R)为评价指标来评估模型性能,公式如(3)、(4)所示:

$$P = \frac{TP}{TP + FP} \cdot 100\% \quad (3)$$

$$R = \frac{TP}{TP + FN} \cdot 100\% \quad (4)$$

其中,TP 为真样本被正确识别;TN 为假样本被正确识别;FP 为假样本被错误识别;FN 为真样本被错误识别;准确率 P 为 TP 所占比例,召回率 R 为被正确识别为真样本的比例。通过 P、R 绘制 P-R 曲线。

AP 为 P-R 曲线下的面积;mAP 是多个类别 AP 的平均值,公式见(5)和(6):

$$AP = \int_0^1 P(R) dR \quad (5)$$

$$mAP = \frac{\sum P_A}{N_c} \quad (6)$$

4.4 实验结果分析

将改进的 RBC-YOLOv5 模型训练 150 个 epochs,与原始 YOLOv5 在验证集上进行测试,边界框回归损失(box\_loss)通过预测框和真实框的 IoU 值来反应预测结果的好坏,目标置信度损失(obj\_loss)由正样本匹配得到的样本对计算损失,类别损失(cls\_loss)通过预测框的类别分数和目标框类别的独热编码表现来计算类别损失,观察各类损失的变化,可见 RBC-YOLOv5 模型的收敛速度更快,各损失更低,box\_loss 稳定在 0.0148 左右不再降低,obj\_loss 稳定在 0.0149 左右,cls\_loss 稳定在 0.0027 左右,具体如图 7 所示。

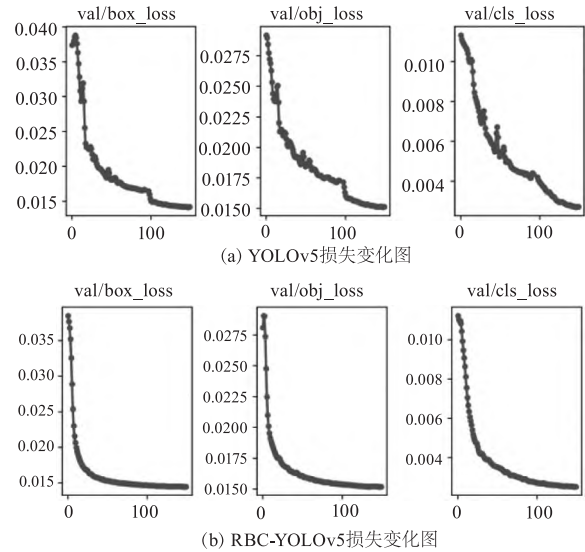


图 7 损失变化图

Fig. 7 Loss change graph

在迭代 150 个 epochs 后,P 提升 2.8%,R 提升 16%,mAP 提升 2.3%,具体数据如表 3。

表 3 YOLOv5 与 RBC-YOLOv5 性能指标对比

Tab. 3 Comparison of YOLOv5 and RBC-YOLOv5

评估指标	YOLOv5	RBC-YOLOv5
P	0.824	0.852
R	0.82	0.98
mAP@0.5	0.818	0.841

在验证集上测试,观察 P、R 与 mAP 的变化,如图 8 所示,RBC-YOLOv5 模型的性能整体优于 YOLOv5。

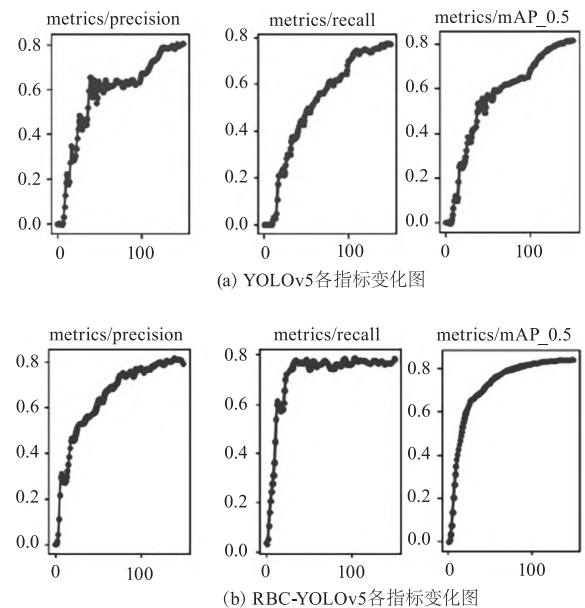


图 8 指标变化图

Fig. 8 Index change chart

消融实验是系统分析的最直接方式,通过消融实验观察不同改进带来的变化,改进内容包括感受野模块 RFBs、卷积注意力模块 CBAM、加权双向特征金字塔 BiFPN,从表 4 可以观察到,BiFPN 使精确度提升了 2.8%,RFBs 与 CBAM 对召回率其关键作用,分别提升了 11% 和 3%,RFBs 与 CBAM 使平均精度均值分别提升了 1% 和 0.7%。

表 4 RBC-YOLOv5 消融实验结果

Tab.4 RBC-YOLOv5 ablation experiment results

模型	<i>P</i>	<i>R</i>	mAP@0.5
YOLOv5	0.824	0.82	0.818
YOLOv5 + RFBs	0.824	0.93	0.828
YOLOv5 + RFBs + CBAM	0.824	0.96	0.835
YOLOv5 + RFBs + CBAM + BiFPN	0.852	0.98	0.841

RBC-YOLOv5 在 DroneVehicle 数据集中各个类别的测试数据如表 5 所示。

表 5 测试集 RBC-YOLOv5 性能指标

Tab.5 Test set RBC-YOLOv5 performance indicators

模型	<i>P</i>	<i>R</i>	mAP@0.5
All	0.852	0.98	0.841
Car	0.690	0.900	0.934
Feright_car	0.750	0.630	0.710
Truck	0.590	0.730	0.762
Bus	0.860	0.930	0.957

4.5 实验结果展示

图 9 为 RBC-YOLOv5 在 DroneVehicle 数据集上的一些检测结果。



检测结果

Fig.9 Test result

图 10 为将 RBC-YOLOv5 模型应用于某型航空相机实际航拍图像的检测结果。左侧为 YOLOv5 检测结果,右侧为 RBC-YOLOv5 模型检测结果。如图 10

(a)与 10(b),对于遮挡及漏检的汽车,原 YOLOv5 不能检测出目标,改进模型能检测出目标,降低漏检率,其检测精度分别为 0.65 和 0.66,如图 10(c),原 YOLOv5 将斑马线误检为目标,改进后的模型有效识别出真实目标,降低了误检率,如图 10(d)、(e)与 (f),RBC-YOLOv5 模型能够有效提升 1%~9% 的检测精度,进一步验证了 RBC-YOLOv5 的检测性能。

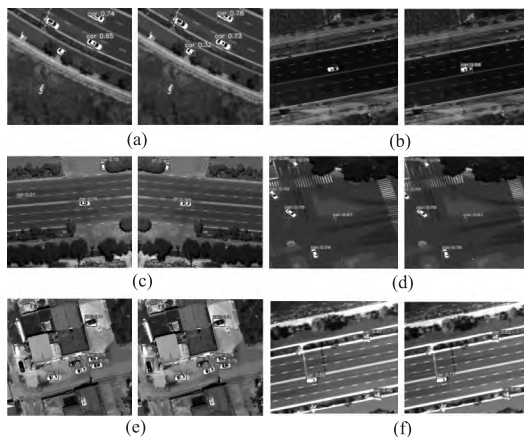


图 10 航空机载红外图像检测结果

Fig.10 AirborneInfrared Image Inspection Results

5 结论与展望

针对机载红外图像车辆目标的检测问题,本文提出基于 YOLOv5 的改进模型,在浅层特征层引入感受野模块,提高对小目标的检测效果,将 PANet 替换为 BiFPN,强化底层特征,引入 CBAM 模块,使模型更加关注重要特征,有效提升对特征的提取能力,本文在 DroneVehicle 数据集与实拍航空红外图像中做了多组对比实验,实验结果表明,检测的精确率提升了 2.8%、召回率提升了 16% 与平均精度提升了 2.3%,降低了漏检率及误检率,证明了本文方法的有效性,不过本文所用的加权双向融合方法增加了模型的大小,后续工作致力于提升模型精度的同时减小模型大小。

参考文献:

[1] Chen H C, Widodo A M, Wisnujati A, et al. AlexNet convolutional neural network for disease detection and classification of tomato leaf[J]. Electronics,2022,11(6):951.  
 [2] Xu X,Zhao M,Shi P,et al. Crack detection and comparison study based on faster R-CNN and mask R-CNN[J]. Sensors,2022,22(3):1215.  
 [3] Wang X,Wang S,Cao J,et al. Data-driven based tiny-YOLOv3 method for front vehicle detection inducing SPP-

- net[J]. *IEEE Access*,2020,8:110227 – 110236.
- [4] Mansour R F, Escorcia-Gutierrez J, Gamarra M, et al. Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model[J]. *Image and Vision Computing*,2021,112:104229.
- [5] Zhai S, Shang D, Wang S, et al. DF-SSD: an improved SSD object detection algorithm based on DenseNet and feature fusion[J]. *IEEE Access*,2020,8:24344 – 24357.
- [6] Lin TY, 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.
- [7] Wang Xiaohong, Chen Zheqi. Research on pedestrian detection in infrared image based on YOLOv5 algorithm[J]. *Laser & Infrared*,2023,53(1):57 – 63. (in Chinese)  
王晓红,陈哲奇. 基于 YOLOv5 算法的红外图像行人检测研究[J]. *激光与红外*,2023,53(1):57 – 63.
- [8] Zhu X K, Lyu S C, 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 International Conference on Computer Vision Online;IEEE,2021,2778 – 2788.
- [9] Wang Henghui, Cao Dong, Zhao Yang, et al. Survey of infrared dim small target detection algorithm based on deep learning [J]. *Laser & Infrared*, 2022, 52 (9): 1274 – 1279. (in Chinese)  
王恒慧,曹东,赵杨,等. 基于深度学习的红外弱小目标检测算法研究综述[J]. *激光与红外*,2022,52(9): 1274 – 1279.
- [10] Yun S, Han D, Oh S J, et al. Cutmix: regularization strategy to train strong classifiers with localizable features [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision,2019:6023 – 6032.
- [11] Zakria Z, Deng J, Kumar R, et al. Multiscale and direction target detecting in remote sensing images via modified YOLO-v4[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*,2022,15:1039 – 1048.
- [12] Huang Z, Wang J, Fu X, et al. DC-SPP-YOLO: dense connection and spatial pyramid pooling based YOLO for object detection [J]. *Information Sciences*, 2020, 522: 241 – 258.
- [13] Zhang T, Zhang X, Ke X. Quad-FPN: a novel quad feature pyramid network for SAR ship detection [J]. *Remote Sensing*,2021,13(14):2771.
- [14] Liu S, Qi L, Qin HF, 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] Liu S, Huang D. Receptive field block net for accurate and fast object detection[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 385 – 400.
- [16] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,2020,42(8):2001 – 2023.
- [17] Wang S H, Fernandes S L, Zhu Z, et al. AVNC: attention-based VGG-style network for COVID-19 diagnosis by CBAM [J]. *IEEE Sensors Journal*,2021,22(18):17431 – 17438.
- [18] Ghiasi G, Lin TY, Le QV. NAS-FPN: learning scalable feature pyramid architecture for object detection [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach;IEEE,2019. 7029 – 7038.
- [19] Zhao Q J, Sheng T, Wang Y T, et al. M2Det: a single-shot object detector based on multi-level feature pyramid network [C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu;AAAI,2019. 9259 – 9266.
- [20] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle;IEEE,2020. 10778 – 10787.
- [21] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:7132 – 7141.
- [22] Y Sun, B Cao, P Zhu, et al. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning[J]. *IEEE Transactions on Circuits and Systems for Video Technology*,2022,32(10):6700 – 6713.