

基于改进 YOLOv5s 的航拍红外图像目标识别方法

王 悠¹, 韩立祥¹, 付 贵^{1,2}

(1. 中国民用航空飞行学院, 四川 广汉 618307; 2. 四川省通用航空器维修工程技术研究中心, 四川 广汉 618307)

摘要: 为了提高无人机在黑暗条件下的识别效率, 降低在复杂环境及路况方面存在漏检及延时效果等问题, 本文提出了一种改进的 YOLOv5s-GN-CB 红外图像识别方法, 该方法可以提高无人机红外航拍图像对车、人等多类目标识别效率。本文对 YOLOv5s 的主要改进包括以下 3 个方面: 将 Ghost 引入到 YOLOv5s 主干网络中, 并将 NWD loss 损失函数融入至 Ghost 中; 添加注意力机制 CA; 添加加权双向特征金字塔 BiFPN。经验证, 改进的 YOLOv5s-GN-CB 检测模型在 InfiRay 红外航拍人车检测数据集下目标识别平均精度均值 (mAP@0.5) 达到 95.1%, FPS 提高至 75.188 帧/s。相较于 YOLOv5 原始模型的平均精度均值和 FPS 分别提高了 4.2% 和 12.02%。在对同一场景中无人机航拍红外图像目标识别的检测精度有明显提升, 延时率有所下降。

关键词: 红外目标检测; 改进 YOLOv5s; Ghost 网络; 注意力机制

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1001-8891(2024)07-0775-07

Aerial Infrared Image Target Recognition Method Based on Improved YOLOv5s

WANG You¹, HAN Lixiang¹, FU Gui^{1,2}

(1. Civil Aviation Flight University of China, Guanghan 618307, China;

2. General Aircraft Maintenance Engineering Technology Research Center of Sichuan, Guanghan 618307, China)

Abstract: To enhance the recognition efficiency of UAVs in dark conditions and reduce missed detections and delays in complex environments and road conditions, this study proposes an improved YOLOv5s-GN-CB infrared image recognition method. This method enhances the efficiency of UAV infrared aerial images for detecting vehicles, people, and other types of targets. The main improvements to YOLOv5s achieved in this study include the following three aspects: 1) introducing the Ghost module into the YOLOv5s backbone network and incorporating NWD loss into Ghost; 2) adding the coordinate attention (CA) mechanism; 3) incorporating a weighted bidirectional feature pyramid network (BiFPN). The improved YOLOv5s-GN-CB detection model achieves an average accuracy of 95.1% (mAP@0.5) on the InfiRay infrared aerial photography man-vehicle detection dataset, with the FPS increased to 75.188 frames per second. Compared with the original YOLOv5 model, the average accuracy and FPS are improved by 4.2% and 12.02%, respectively. In the same scenario, the detection accuracy of UAV aerial photography infrared image target recognition has been significantly improved, and the delay rate has decreased.

Keywords: infrared object detection, improved YOLOv5s, ghost network, attention mechanism

0 引言

目标识别技术是计算机视觉领域的重要研究方向之一。但是在可见光效果较差的环境下, 传统的可见光目标检测技术往往难以达到良好的效果, 使用红外成像在夜间或低光环境下比可见光成像效果

更好^[1], 因此红外目标检测在安防、军事等领域中有着广泛的应用。

随着在机器视觉领域引入深度学习, 目标检测技术取得了快速发展, 基于深度学习的目标检测方法解决了传统机器学习检测精度低、漏检率高、检测速率低的问题。基于深度学习的目标检测算法按

收稿日期: 2023-04-17; 修订日期: 2023-05-17.

作者简介: 王悠 (1978-), 女, 广西陆川人, 高级实验师, 硕士生导师, 研究方向: 计算机技术协同应用研究。Email: yoyo_wang@126.com。

通信作者: 付贵 (1990-), 男, 博士研究生, 讲师, 主要从事深度学习、图像处理、无人机视觉伺服控制等研究。Email: abyfugui@163.com。

基金项目: 中央高校基本科研业务费基金项目 (J2022-024); 四川省通用航空器维修工程技术研究中心资助课题 (GAMRC2021ZD01)。

其网络体系结构，主要分为单阶段检测算法及双阶段网络算法，但经大量研究证实了双阶段目标检测算法可以在红外图像下进行目标检测，但过程繁琐，且需要大量重复性实验。随着YOLOv3^[2]、YOLOv4^[3]、YOLOv5的先后提出，YOLO^[4]系列算法在精度和速度上对比同单阶段算法有着较大的优势，YOLOv5s算法作为YOLO系列算法比较成熟且模型体积较小的目标检测算法，对红外图像进行目标识别有着较好的表现，但因红外图像的图像特性，在检测速率、精度上尚有改进空间。

本文基于YOLOv5s^[5]算法，提出了以下几点改进：

1) 引入了Ghost网络，并将NWD loss (Normalized Wasserstein Distance loss) 损失函数融入至Ghost网络中，Ghost模型减少模型参数的同时可以提高YOLOv5s的性能，将损失函数替换为NWD loss，在不影响模型参数数量的同时提高了模型的精度。

2) 添加注意力机制CA，CA不仅考虑了通道信息，还考虑了方向相关的位置信息，且能够简单地插入到轻量级网络的核心模块中。

3) 添加了加权双向特征金字塔BiFPN，BiFPN具有高效的双向跨尺度连接和加权特征融合，可以实现更多的特征融合。

1 算法理论分析

1.1 YOLOv5 目标检测算法

YOLOv5 算法是一种轻量化的目标检测算法，其网络结构由输入端、主干网络、颈部和检测网络组成。YOLOv5 算法具有 4 个不同深度和宽度的网络，分别为YOLOv5s、YOLOv5m、YOLOv5l 和YOLOv5x，其中YOLOv5s是深度最小、特征图宽度最小且训练速度最快的网络。该算法具有高准确性和较快的训练速度，适用于移动端和嵌入式设备的实时目标检测应用。

1.2 Ghost Bottleneck 模块

Ghost Bottleneck^[6]是一种轻量化的卷积神经网络结构，旨在减少计算量和参数数量的同时，提高模型的精度和鲁棒性。如图 1 为 Ghost Bottleneck 模块原理图，Ghost Bottleneck 主要由两个 Ghost 模块叠加而成，第一个模块用于增加 t 通道数量，而第二个模块则用于减少通道数量，Ghost Bottleneck 包含 stride = 1 和 stride = 2 两个版本，stride = 1 时不进行下采样，直接进行两个 Ghost 卷积操作；stride = 2 时 shortcut 路径下进行下采样，并在 Ghost 模块中加入

一个步长为 2 的深度卷积操作，通过这种方式，Ghost Bottleneck 可以在保证特征提取效果的同时，大幅减少参数数量和计算量。其中 Ghost module 模块原理图如图 2 所示。

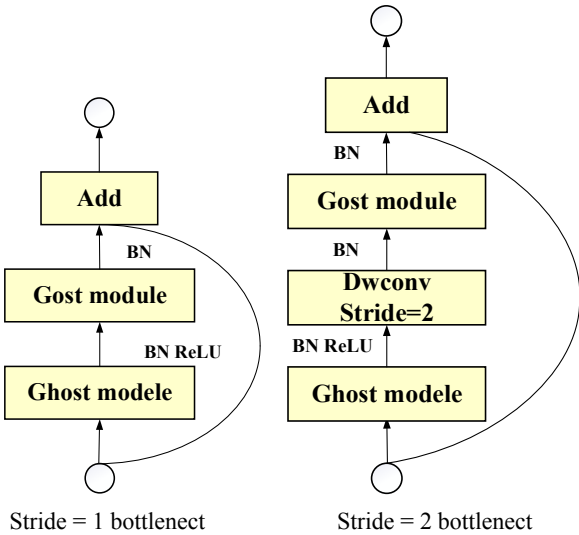


图 1 Ghost Bottleneck 模块原理图

Fig.1 Module schematic of the Ghost Bottleneck

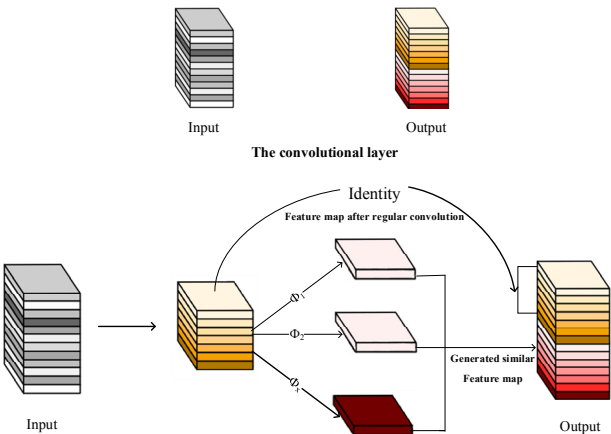


图 2 Ghost module 模块原理图

Fig.2 Module schematic of the Ghost module

1.3 NWD Loss

NWD loss^[7]是一种用于目标检测任务的损失函数。NWD loss 是基于 Wasserstein 距离的思想，旨在降低不同尺度物体之间的特征差异，提高目标检测的准确性。相比于传统的均方误差 (Mean Squared Error) 损失函数，NWD loss 可以更好地适应物体检测任务中不同尺度物体的特征差异，从而更好地优化模型。在 NWD loss 中，Wasserstein 距离是在两个分布之间计算的，这里的分布指的是特征图上的物体分布。通过计算分布之间的 Wasserstein 距离，可以度量两个分布之间的相似度，从而用作损失函数的衡量标准。

NWD:

$$\text{NWD}(N_p, N_g) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}\right) \quad (1)$$

NWD loss:

$$L_{\text{NWD}} = 1 - \text{NWD}(N_p, N_g) \quad (2)$$

式中: N_p 是预测框的高斯分布; N_g 是 GT 框的高斯分布; C 是一个常数, 和数据集有关。

1.4 注意力机制 CA

Coordinate Attention^[8] (CA) 是一种新颖的移动网络注意力机制, CA 注意力很简单, 可以灵活地插入到经典的移动网络中, 而且几乎没有计算开销。CA 框图如图 3 所示。

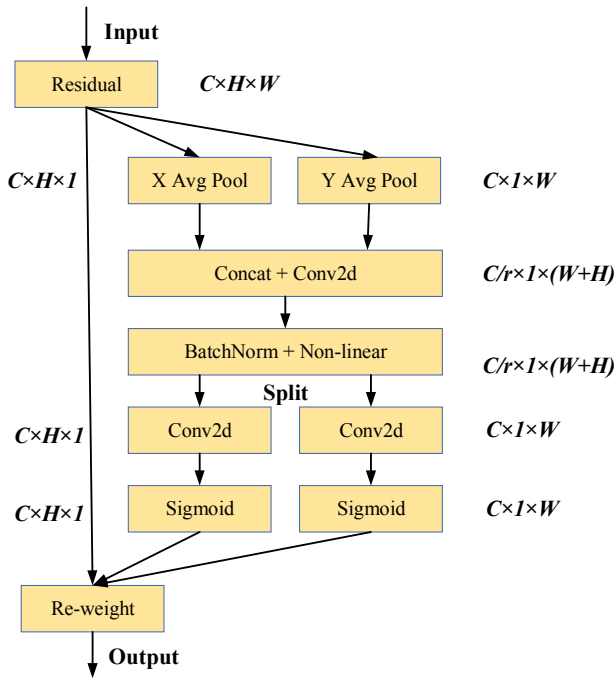


图 3 CA 注意力模块

Fig.3 CA attention module

为了使注意力模块能够捕捉具有精确位置信息的远程空间交互, CA 按照以下公式分解了全局池化。

转化为一对一维特征编码操作:

$$z_c = 1 - \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (3)$$

高度为 h 的第 c 通道的输出:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (4)$$

宽度为 w 的第 c 通道的输出:

$$z_c^h(h) = \frac{1}{H} \sum_{0 \leq i < H} x_c(j, w) \quad (5)$$

对尺寸为 $C \times H \times W$ 输入特征图 Input 分别按 X

方向和 Y 方向进行池化, 分别生成尺寸 $C \times H \times 1$ 和 $C \times 1 \times W$ 的特征图, 如图 4 所示。

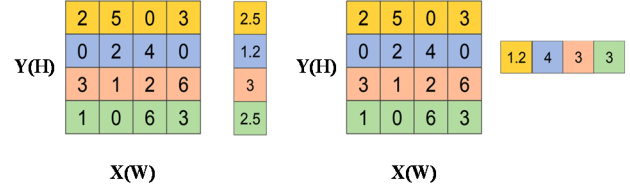


图 4 $C \times H \times 1$ 和 $C \times 1 \times W$ 的特征图

Fig.4 $C \times H \times 1$ and $C \times 1 \times W$ characteristic diagram

将生成的 $C \times 1 \times W$ 的特征图进行变换, 然后进行 Concat 操作:

$$f = \delta(F_1([z^h, z^w])) \quad (6)$$

Coordinate Attention Block 的输出 Y :

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (7)$$

1.5 加权双向特征金字塔 BiFPN

Bipartite Feedforward Networks (BiFPN) 是一种新型的神经网络结构, 它可以应用于计算机视觉任务中的特征提取。BiFPN^[9] 不同于其他的 FPN 结构 (不同分辨率的特征融合时直接相加), 在 PANet^[10] 的基础上, 若输入和输出都是同一水平, 则添加一条额外的边。EfficientDet^[11] 为解决因不同的特征具有不同的分辨率, 对特征融合的贡献不平等的问题, 提出在特征融合期间为每个输入添加一个额外的权重, 让网络去学习每个输入特征的重要性。如图 5 所示, BiFPN 做了两个方面的改进:

1) PANet 进行特征融合时, 是通过 Concat 的, 一般高层和低层的特征贡献程度相同, BiFPN 在特征融合时, 通过一组可学习的归一化权重参数调整各层贡献程度。

2) BiFPN 堆叠多层实现更高维度的特征融合。

2 网络模型优化

2.1 骨干网络模块替换

使用 Ghost Bottleneck 模块替代原始 YOLOv5s 网络结构中的 C3 模块以及 Conv 模块。YOLOv5s 改进模型算法网络结构示意图和模块替换示意图如图 6 所示。具体而言, 第三层 C3 模块用 step=1 的 Ghost Bottleneck 模块替代; 第二层、第四层、第六层和第八层的 Conv 模块用 step=2 的 Ghost Bottleneck 模块替代, 以降低模型参数数量; 第五层、第七层及第九层的 C3 模块则用 3 个 step=1 的 Ghost Bottleneck 模块替换。将本次改进后的模型 YOLOV5-Ghost 称为 YOLOv5-CG。

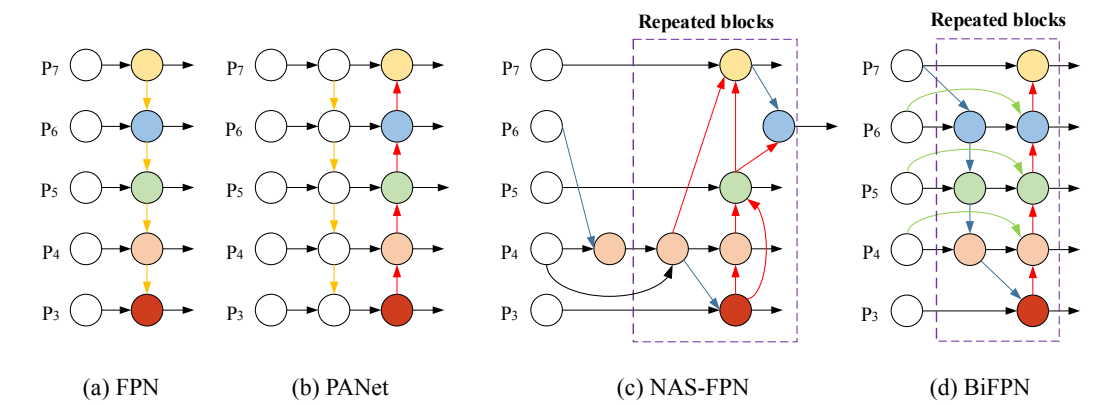


图5 BiFPN 注意力模块
Fig.5 BiFPN attention module

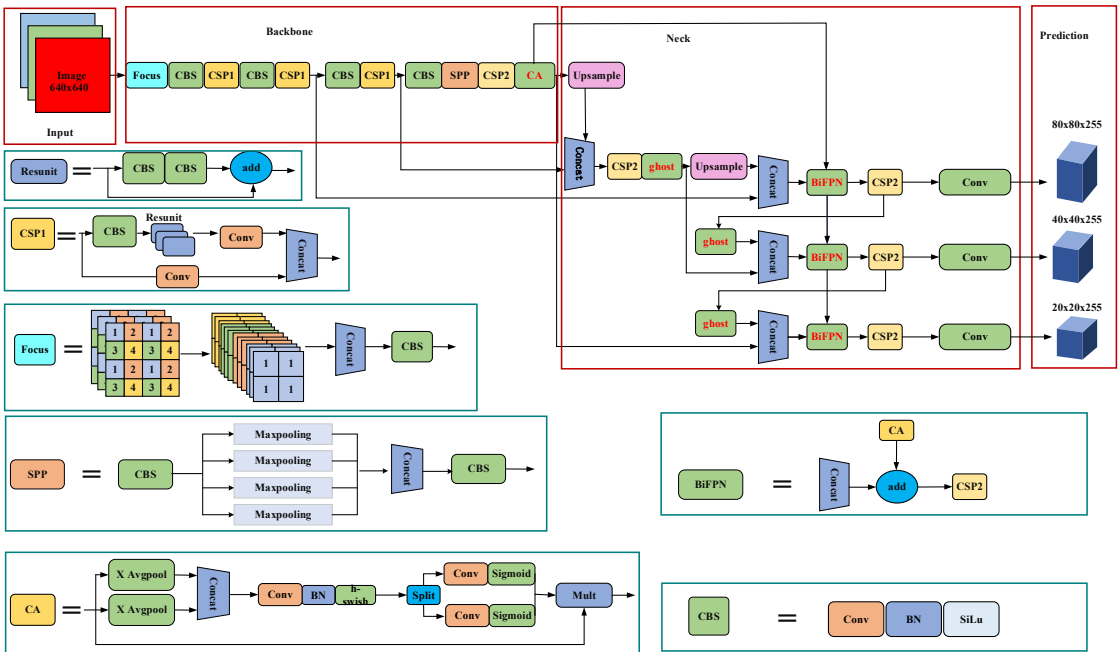


图6 改进后 YOLOv5s 网络结构
Fig.6 Improved YOLOv5s network structure

2.2 损失函数替换

CIOU loss^[12]无法对目标的长宽比进行很好的处理，导致对于形状较为特殊的目标，检测精度可能不如 NWD loss，且 CIOU 损失函数未考虑难易样本的平衡问题。NWD loss 可以对目标之间的相似性进行更准确的度量，并且可以自适应地调整样本的难易程度，从而提高模型的泛化能力。

因此将 Ghost 网络模块回归的损失由原始 CIOU Loss 替换为 NWD Loss，将替换后的模型称为 YOLOv5-GN，通过实验证明，模型表现较好。

2.3 注意力模块 CA 添加

为提升模型训练过程中对训练图像不同区域信息学习效率，因此在不同尺寸的特征图后添加注意力 CA 模块，将 CA 放在 Backbone 部分的最末端，即在第十层的输出端添加 CA 模块，这样可以使注意

力机制看到整个 Backbone 部分的特征图，使得特征图的深度加权平均，提升网络模型的精度，同时较好地提升模型学习效率。

2.4 引入双向特征金字塔网络 BiFPN

将 BiFPN 添加到 Neck 部分的第十层，BiFPN 引入了跳跃连接，即在相同尺度的输入节点到输出节点中间再增加 1 个跳跃连接，因为运算处于相同层，该方法可在参数较少的同时结合更多特征。BiFPN 将每一条双向路径视作 1 个特征网络层，并对同一层参数进行多次计算，以实现更多的特征融合。

3 实验验证及性能风险

3.1 实验平台环境

本文实验是基于 Windows10 操作系统下进行实验验证。训练平台采用 Nvidia GeForce RTX 3090（24

GB/微星)，13th Gen Intel(R) Core(TM) i7-13700K CPU @ 3.40 GHz 处理器，语言为 Python3.7，加速环境为 CUDA11.6，深度学习网络框架为 Pytorch。测试部署平台采用 Jetson AGX Orin 32 GB，该开发板具备支持多个并发 AI 推理管线的 200TOPS 算力，外形小巧，性能出色。该开发板具备 8 核 Arm® Cortex®-A78AE v8.2 64 位 CPU 2 MB L2+4 MB L3，GPU 采用搭载 56 个 Tensor Core 的 1792 核 NVIDIA Ampere 架构 GPU。

3.2 实验数据及参数说明

本文训练验证采用数据集为 InfRay^[13]提供的红外航拍人车检测数据集，数据库使用 person、car、bus、cyclist、bike、truck 分别作为行人、小汽车、公交车、骑自行车的人、自行车、卡车的标签。验证训练集 11045 张、验证集 2000 张、测试集 550 张。网络模型部分重要训练参数设置如下：训练轮次为 300，批尺寸为 16，初始学习率为 0.01，周期学习率为 0.01，学习率动量为 0.937，权重衰减系数为 0.0005。

3.3 网络模型的评价指标

本文对模型训练的评价指标主要从以下几方面分析：精确率（Precision）：是指预测为正例中真正为正例的样本数占有所有预测正例的样本数的比例；召回率（Recall）：是指真正为正例中预测为正例的样本数占有所有真正为正例的样本数的比例；平均精度^[14]（Mean Average Precision, mAP）：是对所有类别的 AP 取平均得到的指标；F1 分数（F1-score）：是衡量模型准确率和召回率之间平衡的指标，F1 分数越高，代表模型综合表现越好。帧率（FPS）：每秒传输帧数，YOLOv5 的帧率计算主要为以下 3 个参数：图像预处理时间（pre-process）、推理速度^[15]、后处理时间，FPS 即 1000 ms 除以这 3 个时间之和。其中 TP 表示被模型预测为正类的正样本数量，FN 表示被模型预测为负类的正样本数量，AP 表示平均准确度，N 表示总的类别数。即预测为正例且正确的样本数除以预测为正例的样本数。

$$\text{Precision} = \frac{TP}{TP+FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{9}$$

$$\text{mAP} = \frac{\sum_{i=1}^N \text{AP}}{N} \tag{10}$$

$$\text{F1-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{11}$$

$$\text{FPS} = \frac{1000}{\text{pre-process} + \text{ms inference} + \text{NMS}} \tag{12}$$

3.4 实验结果及性能分析

将改进后的模型训练所得的 results.txt 文件，使用 Matlab 软件提取有效信息，并对关键数据进行可视化展示，消融实验分两组对比，其中一组对比各模型的精准率、召回率、mAP@0.5、mAP@0.5:0.95。另一组对比各模型的 mAP@0.5、FPS、权重文件大小。为了更直观地感受检测效果，在两个场景下对各模型的检测效果做展示。

3.4.1 消融实验

为验证本文改进算法的有效性，在相同数据集上进行了消融实验。在 YOLOv5s 原模型的基础上，依次进行修改：引入 Ghost 主干网络（YOLOv5s-GC）；引入替换 NWD loss 的 Ghost（YOLOv5s-GN）；引入 CA 和 BiPFN 注意力机制（YOLOv5s-CB）。消融实验结果如表 1 所示。从表 1 中实验结果可以看出，YOLOv5s-GC mAP@0.5 提高了 3%，FPS 增加了 8.1；YOLOv5s-GN mAP@0.5 上升了 3.2%，FPS 增加了 7.51；YOLOv5s-CB mAP@0.5 上升 3.3%，F1 提升了 3，但 FPS 下降了 3.2。对比 YOLOv5s-GN-CB 与 YOLOv5s 原模型，mAP@0.5 上升 4.2%，F1 提升了 4，FPS 上升了 8.1，综上所述证明，本文所改进方法的有效性，在提高模型平均精度的同时，提高了模型计算速度，减少耗时，提高了检测实时性。

表 1 消融实验结果对比

Table 1 Comparison of ablation experimental results			
Models	mAP@0.5	F1	FPS
YOLOv5s	90.9	87	67.114
YOLOv5s- GC	93.9	89	75.188
YOLOv5s- GN	94.1	89	74.627
YOLOv5s- CB	94.2	90	44.053
YOLOv5s- GN-CB	95.1	91	75.188

3.4.2 部分检测结果示例

为了更直观地验证改进后的 YOLOv5s-GN-CB 模型的检测效果，本文用各改进算法分别对两个不同场景进行了红外人车检测验证，场景 a 包括了多个分类，场景 b 下虽分类不多但框图多，分别对两种场景 10 组检测结果进行了对比，检测结果图像如图 7 所示，每组检测结果中，第一列为原 YOLOv5s 检测结果，最后一列为改进的 YOLOv5s-GN-CB 检测结果。通过检测结果可以看出 YOLOv5s-GN-CB 检测精度更高，特征提取能力有所增强，在不同场景下的检测均有一定提高。

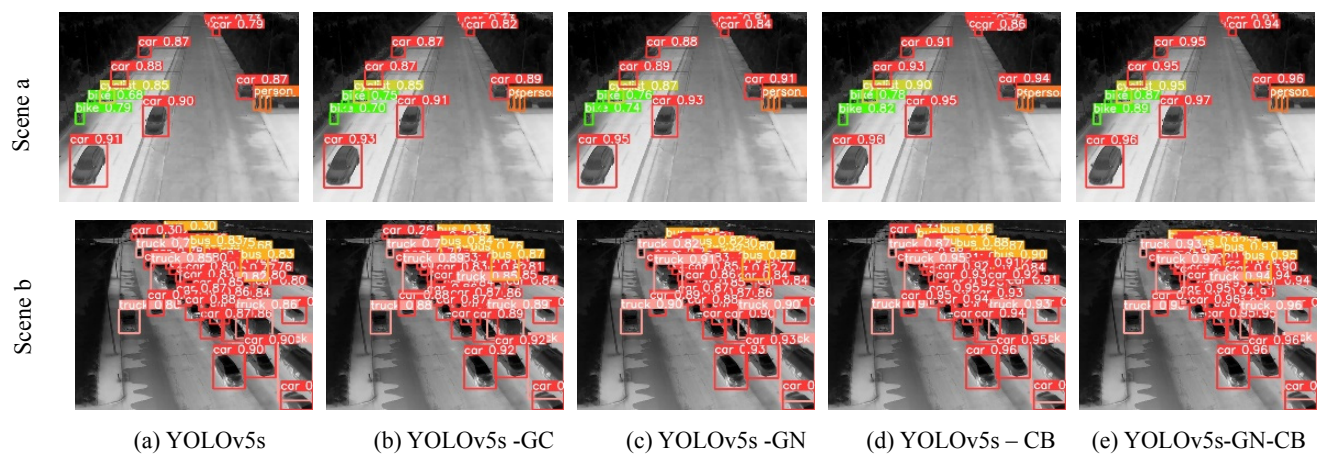


图 7 检测结果示例：两个场景下各改进各算法的检测效果

Fig.7 Example of detection results: The detection effect of each algorithm is improved in two scenarios

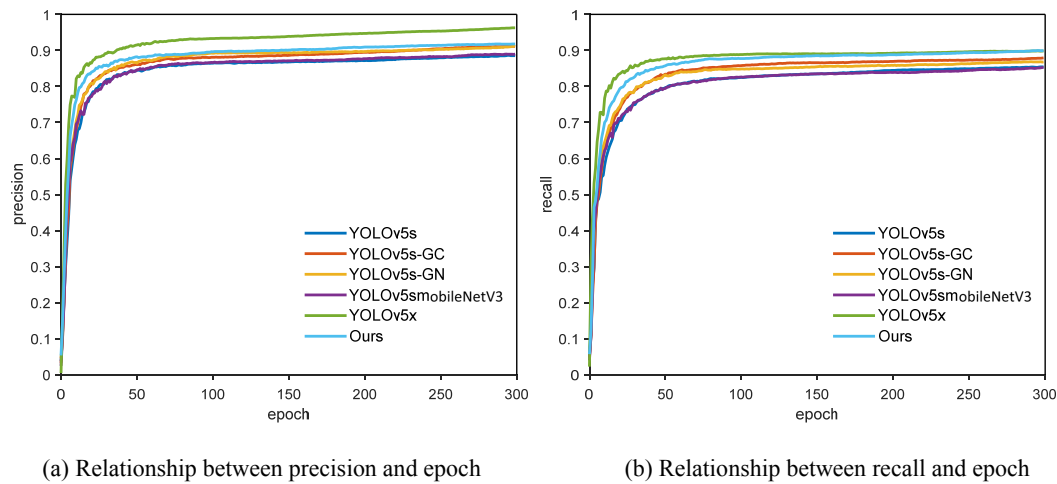
3.4.3 对比实验

为了更加客观地评价本章所改进的 YOLOv5s 模型在红外场景检测中的效果，本文将改进后的 YOLOv5s 模型与其它 YOLOv5 主流算法进行了对比，包括原 YOLOv5s、YOLOv5s-Ghost (GC) [16]、YOLOv5s-GN、YOLOv5s-MobileNetV3[17]、YOLOv5x，实验过程遵循控制变量原则，实验软硬件环境保持一致。评价指标采用平均准确率 (mAP)、每秒帧数 (FPS)、权重文件大小 (MB)。实验结果如表 2 所示，通过实验结果可以看出，Ours 相对于 YOLOv5s mAP@0.5 提高了 3.2%，FPS 增加了 8.1；相对于 YOLOv5s-Ghost mAP@0.5 上升了 2.2%；相对于 YOLOv5s-GN mAP@0.5 上升 1%；相对于 YOLOv5s-MobileNetV3 mAP@0.5 上升 3.1%，FPS 上升了 19；虽然 mAP@0.5 对比 YOLOv5x 略有下降，

但在 FPS 及权重文件大小上有较大优势。图 8 显示了与其它 YOLOv5 主流算法的参数变化对比，通过观察图表可以看出本文改进的算法在平均精度、帧率及权重大小上具有显著优势。

表 2 主流算法对比实验结果

Models	mAP@0.5/ (%)	FPS/ (frame/s)	Weight coefficient/MB
YOLOv5s	90.9	67.114	3.69
YOLOv5s-Ghost	93.9	75.188	7.44
YOLOv5s-GN	94.1	75.188	11.5
YOLOv5s-MobileNetV3	92.0	56.180	7.31
YOLOv5x	97.7	14.164	171
Ours	95.1	75.188	11.6



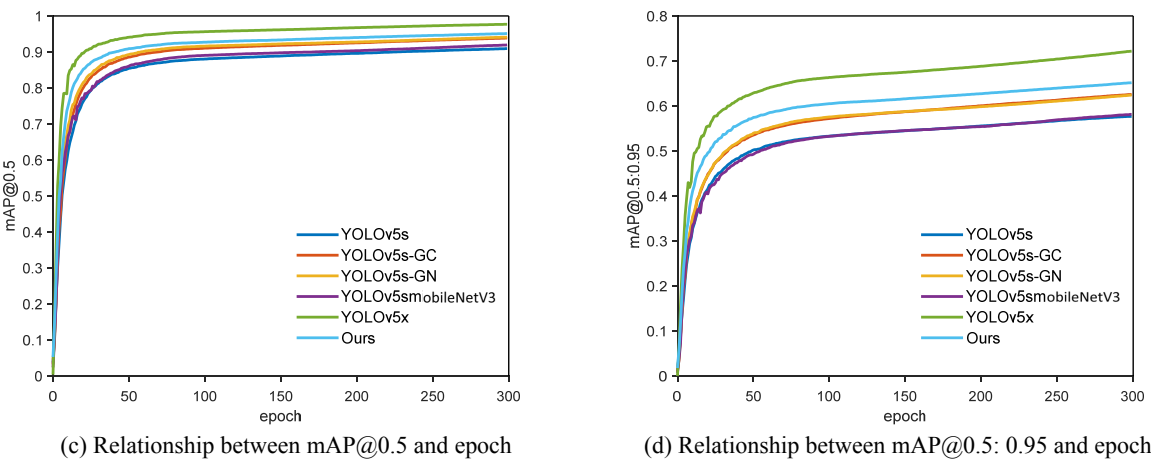


图 8 改进前后不同模型的参数变化对比

Fig.8 Comparison of parameter variations of different models before and after improvement

4 结语

本文提出一种基于改进 YOLOv5s 的红外人车目标检测的优化算法，改进后 YOLOv5s-GN-CB 模型应用于红外场景下对人车等多种分类进行检测分类。实验将模型部署至 Jetson AGX Orin 平台经测试模型，使用改进后算法对红外场景下的航拍人车进行目标识别 mAP@0.5 可达到 95.1%，FPS 为 75.188 帧/s，模型产生的权重文件为 11.6 MB。实验表明，相对于原 YOLOv5s 模型，改进后的模型在利于便携式嵌入式设备的情况下可有效提高平均精度和 FPS。此研究对增强无人机在夜间飞行过程中对周围环境的检测能力，推动无人机的发展具有重要意义。

参考文献：

[1] 张志强, 王萍, 于旭东, 等. 高精度红外热成像测温技术研究[J]. 仪器仪表学报, 2020, 41(5): 10-18.
ZHANG Zhiqiang, WANG Ping, YU Xudong, et al. Research on high-precision infrared thermal imaging temperature measurement technology [J]. *Chinese Journal of Scientific Instrument*, 2020, 41(5): 10-18.

[2] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. 2018-04-08, [2022-08-12]. <https://arxiv.org/abs/1804.02767>.

[3] LIU W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Computer Vision-ECCV*, 2016: 21-37.

[4] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 779-788.

[5] ZHAO Z Y, YANG X X, ZHOU Y CH, et al. Real-time detection of particleboard surface defects based on improved YOLOV5 target detection[J]. *Scientific Reports*, 2021, 11(1): 1-15.

[6] WANG J, CHEN Y, GAO M, et al. Improved YOLOv5 network for real-

time multi-scale traffic sign detection[J/OL]. arXiv:2112.08782, <https://arxiv.org/abs/2112.08782>

[7] WEI Yi, LI Xiaofei, LIN Lihui, et al. Causal discovery on discrete data via weighted normalized wasserstein distance[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(4): 4911-4923. Doi: 10.1109/TNNLS.2022.3213641

[8] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13713-13722.

[9] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10781-10790.

[10] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 8759-8768.

[11] Nawaz Marriam, Nazir Tahira, Baili Jamel, et al. CXray-EffDet: chest disease detection and classification from X-ray images using the efficientdet model[J]. *Diagnostics*, 2023, 13(2): 248-248.

[12] HAO Hongyun, FANG Peng, DUAN Enze, et al. A dead broiler inspection system for large-scale breeding farms based on deep learning[J]. *Agriculture*, 2022, 12(8): 1176-1176.

[13] 郭吉朋. 基于 YOLOv3 模型的实时目标检测算法的 FPGA 实现[D]. 武汉: 华中科技大学, 2020. DOI:10.27157/d.cnki.ghzku.2020.004780.
GUO Jipeng. FPGA implementation of real-time object detection algorithm based on YOLOv3 model[D]. Wuhan: Huazhong University of Science and Technology, 2020. DOI:10.27157/d.cnki.ghzku.2020.004780.

[14] Aliyi Salih, Dese Kokeb, Raj Hakkins. Detectio of gastrointestinal tract disorders using deep learning methods from colonoscopy images and videos[J]. *Scientific African*, 2023, 20: e01628.

(下转第 801 页)

- image synthesis[C]//*Proc. CVPR*, 2021: 12873-12883. DOI: 10.1109/CVPR46437.2021.01268.
- [16] JIANG Y, CHANG S, WANG Z. Transgan: two transformers can make one strong GAN[J/OL]. *Computer Vision and Pattern Recognition*, 2021, **34**: <https://arxiv.org/abs/2102.07074>.
- [17] ZHANG K, LIANG J, Van Gool L, et al. Designing a practical degradation model for deep blind image super-resolution[C]//*Proc. ICCV*, 2021: 4771-4780. DOI: 10.1109/ICCV48922.2021.00475.
- [18] WANG X, XIE L, DONG C, et al. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data[C]//*Proc. ICCVW*, 2021: 1905-1914. DOI: 10.1109/ICCVW54120.2021.00217
- [19] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//*Proc. CVPR*, 2017: 4681-4690. DOI: 10.1109/CVPR.2017.19.
- [20] ZHOU Y, WU G, FU Y, et al. Cross-MPI: cross-scale stereo for image super-resolution using multiplane images[C]//*Proc. CVPR*, 2021: 14842-14851. DOI: 10.1109/CVPR46437.2021.01460.
- [21] 烟台艾睿光电科技有限公司. 艾睿光电红外开源数据库[EB/OL]. [2023-02-26]. http://openai.raytrontek.com/apply/Super_resolution.html/.
IRay Technology Co., Ltd. IRay Optoelectronic Infrared Open Source Database[EB/OL]. [2023-02-26]. http://openai.raytrontek.com/apply/Super_resolution.html/.
- [22] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]//*Proc. ECCV*, 2018: 286-301. DOI: 10.1007/978-3-030-01234-2_18.
- [23] NIU B, WEN W, REN W, et al. Single image super-resolution via a holistic attention network[C]//*Proc. ECCV*, 2020, **12357**: 191-207.
- [24] WANG X, YU K, WU S, et al. ESRGAN: Enhanced super-resolution generative adversarial networks[C]// *Proc. ECCV*, 2019: 63-79. DOI: 10.1007/978-3-030-11021-5_5.
- [25] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//*Proc. ICCV*, 2021: 9992-10002. DOI: 10.1109/ICCV48922.2021.00986.
- [26] WANG Y, WANG L, WANG H, et al. Resolution-aware network for image super-resolution[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, **29**(5): 1259-1269.
- [27] Andreas Lugmayr, Martin Danelljan, Radu Timofte. Unsupervised learning for real-world super-resolution[J]. *ICCV Workshop*, 2019, **2**(3): 3408-3416.
- [28] DAI T, CAI J, ZHANG Y, et al. Second-order attention network for single image super-resolution[C]//*Proc. CVPR*, 2019: 11057-11066. DOI: 10.1109/CVPR.2019.01132.
- [29] Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: Dataset and study[C]// *Proc. CVPRW*, 2017: 126-135. DOI: 10.1109/CVPRW.2017.150.
- [30] LIANG J, CAO J, SUN G, et al. SwinIR: Image restoration using swin transformer[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 1833-1844. DOI: 10.1109/ICCVW54120.2021.00210.
- [31] Blau Y, Mechrez R, Timofte R, et al. The 2018 PIRM challenge on perceptual image super-resolution[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, DOI: 10.1007/978-3-030-11021-5_21
- [32] MA C, YANG C Y, YANG X, et al. Learning a no-reference quality metric for single-image super-resolution[J]. *Computer Vision and Image Understanding*, 2017, **158**: 1-16.

(上接第 781 页)

- [15] 王新彦, 易政洋. 基于改进 YOLOv5 的割草机器人工作环境障碍物检测方法研究[J]. *中国农机化学报*, 2023, **44**(3): 171-176. DOI: 10.13733/j.jcam.issn.2095-5553.2023.03.024.
WANG Xinyan, YI Zhengyang. Research on obstacle detection method of mowing robot working environment based on improved YOLOv5[J]. *Journal of Chinese Agricultural Mechanization*, 2023, **44**(3): 171-176. DOI: 10.13733/j.jcam.issn.2095-5553.2023.03.024.
- [16] 徐正军, 张强, 许亮. 一种基于改进 YOLOv5s-Ghost 网络的交通标志识别方法[J]. *光电子·激光*, 2023, **34**(1): 52-61. DOI:10.16136/j.joel.2023.01.0216.
XU Zhengjun, ZHANG Qiang, XU Liang. Research on obstacle detection method for working environment of lawn mower robot based on improved YOLOv5[J]. *Journal of Optoelectronics Laser*, 2023, **34**(1): 52-61. DOI:10.16136/j.joel.2023.01.0216.
- [17] 党宏社, 党晨, 张选德. 基于改进 YOLOv5s 的交通标志识别算法[J]. *实验技术与管理*, 2022, **39**(9): 97-102. DOI: 10.16791/j.cnki.sjg.2022.09.016.
DANG Hongshe, DANG Chen, ZHANG Xuande. Traffic sign recognition algorithm based on improved YOLOv5s[J]. *Experimental Technology and Management*, 2022, **39**(9): 97-102. DOI: 10.16791/j.cnki.sjg.2022.09.016.