

DOI:10.19651/j.cnki.emt.2417018

改进 YOLOv8n 的轻量级遥感图像军用飞机检测算法^{*}杨瑞君¹ 张浩¹ 叶璟²

(1. 上海应用技术大学计算机科学与信息工程学院 上海 201418; 2. 上海应用技术大学化学与环境工程学院 上海 201418)

摘要: 针对现有的轻量级目标检测算法在应用于遥感图像军用飞机目标检测任务所面临的模型参数大、检测速度慢的情况,提出一种基于 YOLOv8n 的轻量级遥感图像军用飞机目标检测算法:LeYOLO-MARs。采用了优化后的倒置瓶颈模块替换原始主干网络中的经典瓶颈模块,更换高效的骨干网络特征提取模式,在保持特征提取能力的同时,有效降低了计算需求并提升了计算速度;颈部网络中引入了快速金字塔架构网络,减少了卷积层数并提高了语义信息共享的效率,减少了锁定和等待时间,同时考虑了有限的并行化机会和架构的复杂性;使用轻量级解耦网络头,通过逐点卷积简化检测头结构;使用 Inner-SIoU 作为新的定位回归损失函数,提升对小目标样本的学习能力并加快回归边界框的收敛;加入了轻量级金字塔压缩注意力机制模块,有效整合局部注意力和全局注意力,以建立 long-range 通道依赖关系。实验结果表明,改进的算法在保证检测速度的同时取得了 95.7% 的检测精度,比基线模型高 0.4%,模型参数缩小 43%,计算量减少 63%,较主流算法在检测效果上有一定的提升,能够对军用飞机目标进行高质量实时检测。

关键词: 目标检测;军用飞机;YOLOv8;倒置瓶颈;注意力模块;特征金字塔

中图分类号: TP391.4;TN919.4 **文献标识码:** A **国家标准学科分类代码:** 520.20

Improved lightweight military aircraft detection algorithm for remote sensing images with YOLOv8n

Yang Ruijun¹ Zhang Hao¹ Ye Jing²

(1. School of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai 201418, China;

2. School of Ecological Technology and Engineering, Shanghai Institute of Technology, Shanghai 201418, China)

Abstract: Aiming at the large model parameters and slow detection speed encountered by current lightweight target detection algorithms when applied to the task of detecting military aircraft in remote sensing images, this study proposes a lightweight detection algorithm for military aircraft targets based on YOLOv8n, named LeYOLO-MARs. The algorithm introduces an optimized inverted bottleneck module to replace the traditional bottleneck in the backbone network, reducing computational requirements while maintaining feature extraction capabilities and improving processing speed. In the neck network, a fast pyramid architecture is integrated to reduce the number of convolutional layers, enhance the efficiency of semantic information sharing, and decrease lock and wait times, while also considering limited parallelization opportunities and architectural complexity. A lightweight decoupled detection head, simplified through pointwise convolution, is employed, alongside the use of Inner-SIoU as the new localization regression loss function, which enhances the ability to learn from small target samples and accelerates the convergence of bounding box regression. Moreover, the algorithm incorporates a lightweight pyramid compression attention mechanism, effectively combining local and global attention to establish long-range channel dependencies. Experimental results demonstrate that the improved algorithm achieves a detection accuracy of 95.7%, 0.4% higher than the baseline model, while reducing model parameters by 43% and computational load by 63%, marking a notable improvement in detection performance compared to mainstream algorithms and enabling high-quality real-time detection of military aircraft targets.

Keywords: object detection; military aircraft; YOLOv8; inverted bottleneck; attention module; feature pyramid network

0 引言

在全球安全问题不断变化和航空技术迅速发展的时

代,建立一个可靠的即时军用飞机识别系统显得尤为紧迫。遥感图像是通过飞机或卫星对地球表面进行观测而获取的图像,这些图像通常包含丰富的背景信息和多样的地貌特

收稿日期:2024-09-29

* 基金项目:上海应用技术大学协同创新基金-跨学科、多领域合作研究专项(XTCX2024-03)资助

征。由于遥感图像的广泛采集范围和目标像素在整体图像中的占比相对较小,军用飞机的特征往往表现得不够明显。这种情况使得军用飞机目标检测任务面临更大的挑战,特别是军用飞机类型的识别,这是一种细粒度特征,增加了任务的难度^[1]。随着军工武器装备的不断升级,军用飞机种类不断增多。军用飞机作为最具威胁性的战略目标之一,其位置和型号的识别对战场决策有着重要影响。

近年来,计算机视觉发展迅速,目标检测作为其核心问题之一,已广泛应用于军事领域。自 2014 年 RCNN 提出以来,基于深度学习的目标检测方法逐渐取代传统方法成为主流,主要分为双阶段检测算法和单阶段检测算法。双阶段检测算法包括 Faster RCNN^[2]、FPN^[3]、Mask RCNN^[4]和 Cascade RCNN^[5]等,因其对候选区域进行细粒度处理,通常具有较高的检测精度,并且在复杂场景下表现出较强的鲁棒性。然而,此类方法计算复杂度高,实时性较差。相比之下,YOLO^[6]、SSD^[7]、RetinaNet^[8]、EfficientDet^[9]和 DETR^[10]等单阶段检测算法则直接预测输出目标的类别概率和位置,显著减少了计算资源的消耗,提高了检测速度,更适合对实时性要求较高的场景。因此,研究人员在遥感军用飞机检测领域主要采用单阶段检测技术。单慧琳等^[11]引入了自适应特征增强模块,并设计了特征融合上采样模块,提出了一种基于可辨别特征提取和上下文感知的遥感图像飞机目标检测方法,飞机目标检测精度提高了 3.7%~18%。张武等^[12]对 YOLOv5 网络进行重构,加入小目标检测层,引入极化自注意力机制,改进后的算法对飞机目标的检测精度达到 90.1%。尽管上述算法在某种程度上提高了检测精度,但同时也导致了模型整体参数和计算量的增加,无法解决嵌入式设备在性能稳定性和计算资源有限方面的问题。

针对上述问题,学者们进行了模型轻量化设计。杨鑫等^[13]提出轻量化特征提取网络的改进型 Faster R-CNN 飞机检测算法,删去对检测小目标冗余的深层特征层,网络参数量减少 38.4%,mF1 分数提升 3.5%。王杰等^[14]提出的 YOLO-MAR 模型相比于原 YOLOv5s,模型体积减少至 3.95 MB,同时实现了 91.7%的平均检测精度。刘丽等^[15]从网络设计和模型压缩两个方面对 YOLOv8n 进行了轻量化改进,通过优化网络结构,使用基于 LAMP 分数的通道剪枝对重新设计的模型进行了压缩。改进后的模型体积仅有 0.7 MB。然而,由于删除了原始主干网络的大目标检测层,模型在不同场景下的泛化性可能受限。综上,关于模型轻量化的研究主要集中在替换轻量化模块和模型压缩两个方向,但尚未能在模型复杂度与检测精度之间找到平衡。

本文关注军用飞机检测算法在模型参数过大和移动端部署困难的挑战。基于 YOLOv8n 作为基线模型,提出了一种轻量高效的军用飞机检测算法 LeYOLO-MARs。该算法的主要贡献如下:

1)使用优化的倒置瓶颈模块替换原始主干网络中的经典瓶颈模块,更换高效的骨干网络特征提取模式,通过优化通道数量有效减少计算需求,在保持特征提取能力的同时提升计算速度。

2)提出了一种快速金字塔架构网络,其具有更少的卷积层数、通道数和更高效的语义信息共享,简化了颈部网络,减少了锁定和等待时间,同时考虑了有限的并行化机会和架构的复杂性。

3)提出了 DNIN 解耦网络头,通过逐点卷积简化检测头,将其作为逐像素滑动的多层感知器解决方案,为每个网络预测提供独立的分类和回归操作。

4)引入 Inner-SIoU 作为新的回归损失函数,利用较大的辅助边界框加速网络对低 IoU 样本的回归,从而提升了对小目标的定位精度。

5)使用了一种名为金字塔压缩注意力机制的模块,具有轻量化特性,有效整合了局部注意力和全局注意力,显著提升了模型性能,并弥补了模型精度的下降。

1 LeYOLO-MARs 算法

1.1 YOLOv8 模型

YOLOv8 是由 Ultralytics 发布的 YOLO 模型,是在 YOLOv5 的基础上提出,并结合了其他 YOLO 框架^[6]的优点。YOLOv8 通过结合多尺度特征融合、自注意力机制、CSPDarknet53 结构等技术,以及多尺度训练和数据增强技术,显著改善了遥感图像中军用飞机目标分布不均的问题。YOLOv8 网络结构如图 1 所示。

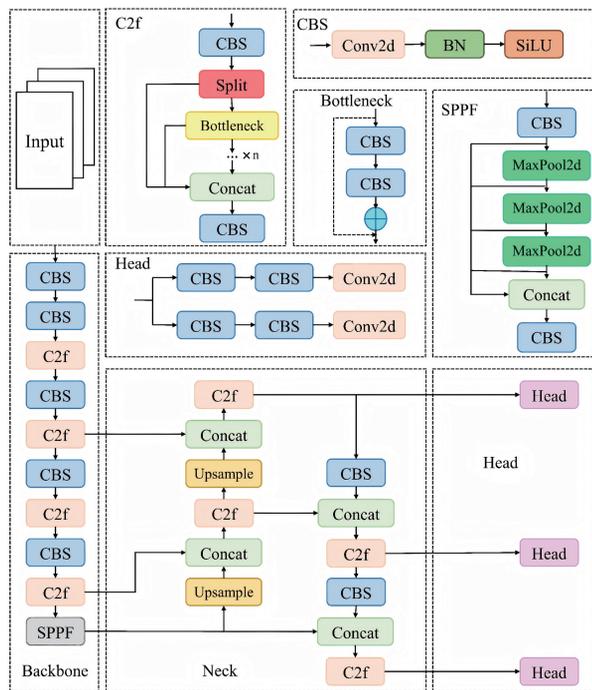


图 1 YOLOv8 网络结构

Fig. 1 YOLOv8 Network structure

YOLOv8 有 5 个版本,模型容量各不相同:纳米(n)、小型(s)、中型(m)、大型(l)和超大型(x),分别称为 YOLOv8n/s/m/l/x。这 5 种模型对应的通道宽度、深度和最大数量如表 1 所示。其中 n 模型具有较低的计算量和参数,为在嵌入式设备上的部署提供了便利,本文选择 YOLOv8n 作为基准模型。

表 1 不同大小的 YOLOv8 参数

Table 1 YOLOv8 parameters for different sizes

模型	深度	宽度	最大通道数
YOLOv8n	0.33	0.25	1 024
YOLOv8s	0.33	0.50	1 024
YOLOv8m	0.67	0.75	768
YOLOv8l	1.00	1.00	512
YOLOv8x	1.00	1.25	512

1.2 LeYOLO-MARs 网络结构

为了解决军用飞机检测算法在模型参数和计算量方面较大的问题,对 YOLOv8n 算法进行了改进,新的模型命名为 LeYOLO-MARs。LeYOLO-MARs 的网络架构由主干网络(Backbone)、颈部网络(Neck)和检测头(Head)三部分组成,其模型结构如图 2 所示。在 Backbone 网络设计中,640×640×3 的原始图像经过 STEM 层的逐点卷积与标准卷积操作,输出尺寸为 160×160×16 的特征图。然后使用改进的倒置瓶颈模块作为特征提取器,分别经过 2 次、5 次和 4 次特征提取后得到 P3、P4、P5 层级的特征图,依次为 80×80×32、40×40×64、20×20×96。将(pyramid squeeze attention,PSA)注意力机制添加在 Backbone 网络中的 SPPF 模块之前,优化计算资源的同时提升了多尺度特征融合效果。在 Neck 层,沿用快速金字塔架构网络,通过更少的卷积层和更低的通道数量,实现更高效地共享语义信息。通过从不同的主干层对不同的检测层进行特征聚合,实现了不同尺寸目标检测间的平衡。在 Head 层,使用 DNiN 检测头代替 YOLOv8 原始的检测头,进一步实现轻量化。在计算损失阶段采用 Inner-SIoU 损失函数,提高了对小目标的定位精度。

1.3 优化主干网络

1)改进瓶颈模块

经典残差瓶颈块(classical bottleneck)由 ResNet^[16] 系列引入,通过引入残差连接来解决深层神经网络中的梯度消失问题,并提高模型的泛化能力,但这种瓶颈模块需要较大的计算资源,YOLOv8 中使用了这种模块。倒置残差瓶颈块(inverted bottleneck)由 MobileNetV2^[17] 引入,主要目的是在移动和嵌入式设备上实现高效计算。其设计思路是与经典瓶颈块相反的:从低维到高维再回到低维。

倒置残差瓶颈块的核心思想是,通过在高维空间进行特征处理,可以有效增加特征表示能力,而使用深度卷积大

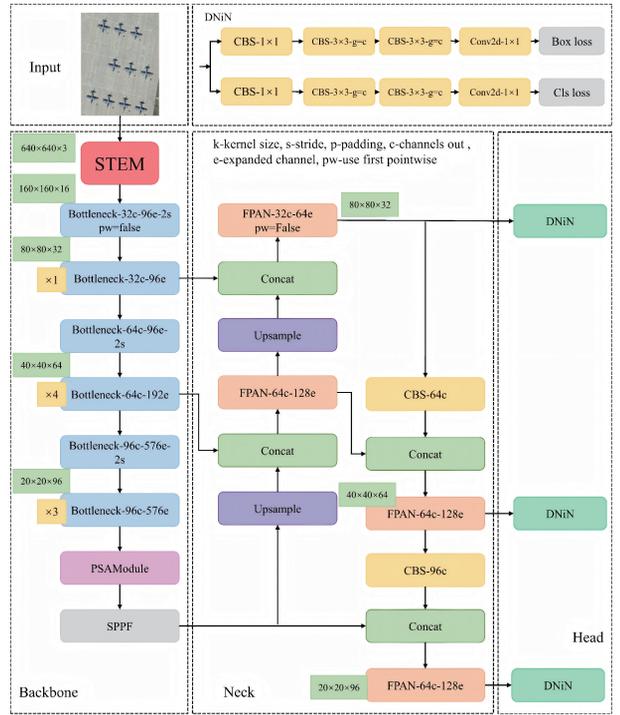


图 2 LeYOLO-MARs 网络结构

Fig. 2 Network structure of LeYOLO-MARs

幅减少计算量。这种设计非常适合计算资源受限的设备,同时也保持了较好的模型性能。

本文在倒置瓶颈的基础上发现,通过优化通道数量,尤其是在较大空间特征图尺寸下,可以有效减少计算需求。如果模块的扩展比例为 1,或由于串联效应输入通道数 C_{in} 等于扩展层数 C_{mid} ,则不需要使用第一个逐点卷积。模块要始终保留残差连接,即使省略了第一个逐点卷积,只要输入的 C_{in} 和输出的 C_{out} 张量相等。图 3 显示了经典瓶颈、倒置瓶颈和改进的倒置瓶颈的区别。

用 \otimes 表示两个值之间的卷积。对于 $F_{in} \in \mathbb{R}^{1,1,C_{in},C_{mid}}$, $F_{out} \in \mathbb{R}^{1,1,C_{mid},C_{out}}$ 和 $F_{mid} \in \mathbb{R}^{k,k,1,C_{mid}}$,其卷积中 C_{in} 是输入通道数, C_{mid} 是反向瓶颈中扩展通道的数量, C_{out} 是输出通道数。卷积过程如式(1)所示。

$$y = \begin{cases} F_{out} \otimes [F_{mid} \otimes (F_{in} \otimes x)], C_{in} \neq C_{mid} \\ F_{out} \otimes [F_{mid} \otimes (F_{in} \otimes x)], C_{in} = C_{mid}, F_{in} = True \\ F_{out} \otimes [F_{mid} \otimes (x)], C_{in} = C_{mid}, F_{in} = False \end{cases} \quad (1)$$

如图 4 所示,改进后的瓶颈结构借鉴了倒置瓶颈的思想,设计从低维到高维再回到低维。通过优化扩展通道数量,有效减少计算需求。此外,是否使用第一层的逐点卷积取决于输入通道数是否等于扩展通道数。这样的设计不仅提高了网络的计算效率,还保留了足够的特征表达能力。

2)改进 STEM

“STEM”一词用来描述神经网络中最初处理输入图像

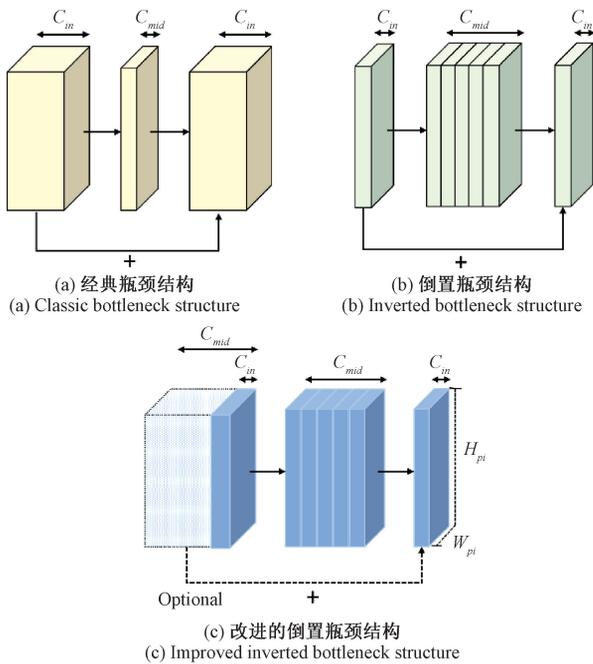


图 3 经典瓶颈、倒置瓶颈和改进的倒置瓶颈的区别
Fig. 3 Distinctions between classical bottlenecks, inverted bottlenecks and modified inverted bottlenecks

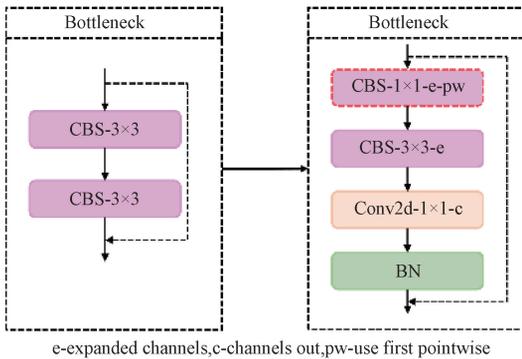


图 4 bottleneck 的改进
Fig. 4 Improvement of bottleneck

和低语义信息的第一层。这些层的主要功能是快速且高效地减少图像的空间尺寸,同时将初始信息通道数增加到更高的通道数。这样做的主要优势是降低计算成本,因为如果处理层的空间尺寸过大,目标检测的计算成本会迅速增加。

在现有的 YOLO 模型中,只有 YOLOv7 没有在第一层使用步长卷积,从而保持了较大的初始空间尺寸 P_0 ,这也带来了较高的计算成本。YOLOv6 和 YOLOv8 这两个模型在将通道数和层数缩减至 $\times 0.25$ 时,其总计算成本为 0.32 GFLOP。它们通过快速将特征图大小缩减到 160×160 来补偿在高空间尺寸上滑动卷积的高成本。

为了更进一步在较大特征图尺寸上高效地使用卷积,本文在整个 STEM 阶段使用逐点卷积与标准卷积相结合

的方法,同时限制通道数的增长。从初始的 $P_0(640 \times 640)$ 到后续的 $P_2(160 \times 160)$,都保持了较低的通道数,如表 2 所示。这种设计不仅减少了计算量,还能有效保持模型的精度和性能。

表 2 STEM 的结构概述

Table 2 Overview of the structure of STEM

模块结构	通道数	卷积核	步幅	激活函数	输出
Conv+bn+act	16	3	2	SiLU	320×320
Conv+bn+act	16	1	1	SiLU	320×320
Depthwise+bn+act	16	3	2	SiLU	160×160
Conv+bn	16	1	1	SiLU	160×160

3) 高效骨干特征提取模式

目标检测器通常用于辅助分类任务,因此一些模型使用经典的分类模型作为特征提取器。在设计中,本文首先选择了倒置瓶颈模块,因其具有优异的计算效率和成本-准确率比。在层数选择方面,神经网络的最新趋势显示: P_4 或类似层级上的层重复次数显著增加。使用类似 NAS(神经架构搜索)算法选择层数或重复次数时,这一现象也得到了验证,表明 P_4 层级在网络结构中的重要性超过其他层级。

这一结论不仅适用于 YOLO 模型,还适用于其他使用卷积的分类模型、自注意力机制模型,以及基于 ResNet 的目标检测器 DETR^[10],这些模型都在 P_4 层级或相应的中间层中投入了更多的注意力和计算资源。因此,为了创建一个低计算成本的神经网络,不能过于依赖 P_3 层级来提取特征。在本文的主干网络设计中,包含了表 3 中显示的层重复次数,以优化网络的计算效率和性能。

表 3 其他模型的层重复情况与 LeYOLO-MARs 相比

Table 3 Layer repeats of other models compared to LeYOLO-MARs

模型	P3	P4	P5
YOLOv3	$\times 8$	$\times 8$	$\times 4$
YOLOv4	$\times 8$	$\times 8$	$\times 4$
YOLOv5	$\times 6$	$\times 9$	$\times 3$
YOLOv6	$\times 12$	$\times 12$	$\times 6$
LeYOLO-MARs	$\times 2$	$\times 5$	$\times 4$

1.4 快速金字塔架构网络

在目标检测中,Neck 是模型中聚合多个层次语义信息的部分,负责从深层到浅层共享提取的特征。研究人员通常使用 PANet^[18] 或 FPN^[3] 高效共享特征图,通过将多个语义信息连接到 PANet 及其相应输出,实现多个检测层,如图 5(a) 所示。本文比较了两种聚合网络:BiFPN^[9] 和

YOLOF 的 SiSO^[19]。BiFPN 的特点是使用计算成本低的层(如拼接、加法、深度卷积和逐点卷积)。然而,BiFPN 需要大量语义信息和阻塞状态(等待前一层、复杂的图结构),难以保持快速执行速度。另一方面,SiSO 在目标检测方法上颇具创新。YOLOF 的作者选择为模型的 Neck 部分采用单一输入和输出。与 YOLOF 论文中提出的其他解决方案相比,具有多个输出的 Neck 和具有单一输出的 Neck 之间存在显著差异。本文特别关注的是 SiMO 的潜在效率研究,证明通过仅使用一个丰富的输入来优化语义信息流动,可以改进 YOLO 模型 Neck 前几层的性能。

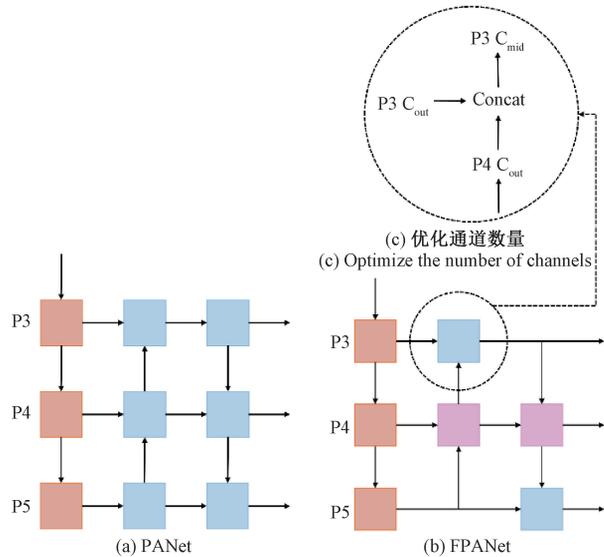


图 5 PANet 和 FPNNet 的计算区别

Fig. 5 Computational distinction between classic PANet and FPNNet

本文提出快速 PANet(FPNNet),其特点是卷积层较少、通道数量较低,并更高效地共享语义信息。其设计与 YOLOv8 的 Neck 概念相似。通过减少主干网络后 P3 到 P5 之间的计算流量,并在进入头部之前直接增强 P4 的语义信息层级,如图 5(b)所示。此外,FPANet 简化了 Neck 结构,减少了锁定和等待时间,考虑到架构的有限并行化机会和复杂性。如图 5(c)所示,通过优化通道数量减少 P3 的计算量。初始的逐点卷积步骤是不必要的,因为来自 P4 的自底向上路径和主干网络中的 P3 信息的拼接与 PAN 中 P3 的倒置瓶颈扩展通道数相匹配,即 $C_{in}P4 + C_{in}P3 = C_{mid}P3$ 。

在 YOLOv8n 的颈部网络中,通过 c2f 模块双向融合低级特征和高级特征,从而提高对不同尺度目标的检测能力。而 LeYOLO-MARs 使用如图 6 所示的 FPN 结构来实现这一功能。FPAN 结构由 3 个改进的瓶颈网络组成,具有较少的卷积层和较低的通道数量,但能更高效地共享语义信息。

1.5 DNiN 检测头

自 YOLOv6 以来,模型头被划分为两个部分:一个用

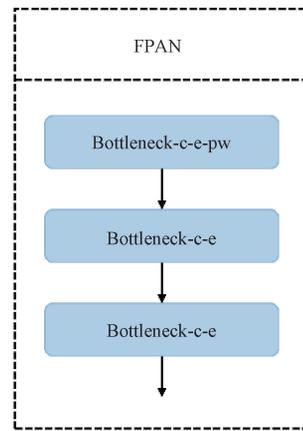


图 6 FPN 模块
Fig. 6 FPN module

于分类,另一个用于目标回归。虽然这种设计非常高效,但几乎将计算成本翻倍,因为需要分别进行分类和检测的卷积运算。本文认为,没有必要额外添加空间信息,只需通过轻量级的深度卷积来精细化主干和颈部提取的特征即可。历史上,YOLO 模型通过网格方式工作,为每个网格像素提供由锚框调整的分类建议。锚框提供了几种可能的检测尺寸,而不仅仅是逐像素的检测。

YOLO 的逐点卷积使简化检测头成为可能。通过使用逐点卷积作为逐像素的滑动多层感知器解决方案,类似于为每个像素提出分类建议。多个深度卷积仅用于空间操作,细化两次逐点卷积之间的空间关系,实现逐像素的分类和回归。

本文提出的 DNiN(解耦网络中网络头),如图 7 所示,这是一种以逐点操作为中心的方法。每个网络提案分别进行两次逐点操作:分类和回归(边界框)。逐点操作在目标检测中至关重要,作为网络中网络框架内的逐像素分类器和回归器。深度卷积分为两个 3×3 卷积,以降低相对于单个 5×5 卷积的总体成本。

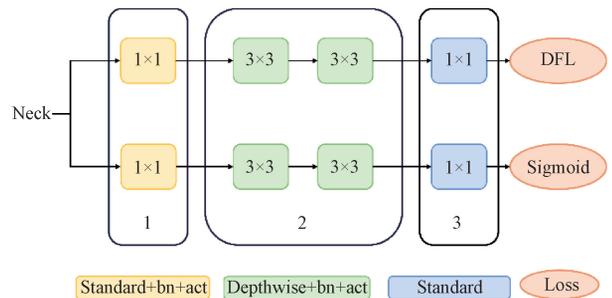


图 7 DNiN 检测头网络结构
Fig. 7 Network structure of DNiN head

在 DNiN 中,本文进行了两个独立的逐点卷积:一个用于分类,另一个用于回归。分类和边界框提取的不同需求导致了这种区别。因此,每个像素都代表一个潜在的预测。使用 1×1 卷积可以追溯到神经网络的起源,尤其是

NiN^[20]模型,其中逐点卷积取代了传统的多层感知器。

1.6 优化损失函数

边界框回归在目标检测器的定位精度中扮演着至关重要的角色。YOLOv8 选择了 CiOU 作为其损失函数,这是因为 CiOU 同时考虑了预测框与真实框之间的重叠面积、中心点的距离以及框的纵横比等多个因素。与传统的 IoU 指标相比,CiOU 提供了一个更为全面的评估标准,从而显著提升了边界框回归的精度。这种改进不仅增强了模型的定位能力,也为复杂场景下的目标检测提供了更为可靠的支持。

SIoU^[21]在边界框回归中引入了角度损失,以全面考虑 Anchor Box 与真 GT Box 之间的角度影响,从而提高定位精度。该损失函数由 4 个主要部分组成:形状损失、距离损失、角度损失和 IoU 损失,其定义如下:

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (2)$$

式中: IoU 为预测框与真实框的交并比, Δ 为距离损失, Ω 为形状损失。

$$\begin{cases} \Delta = \frac{1}{2} \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) \\ \rho_x = \left(\frac{b_{c_x} - b_{c_x}^{gt}}{c_w} \right)^2, \rho_y = \left(\frac{b_{c_y} - b_{c_y}^{gt}}{c_h} \right)^2 \\ \Lambda = \sin \left(2 \sin^{-1} \frac{\min(|x_c^{gt} - x_c|, |y_c^{gt} - y_c|)}{\sqrt{(x_c^{gt} - x_c)^2 + (y_c^{gt} - y_c)^2 + \varepsilon}} \right) \\ \gamma = 2 - \Lambda \end{cases} \quad (3)$$

式中: C_w 和 C_h 为 Pred Box 与 GT Box 的最小外接矩形的宽和高,角度损失 Λ 表示 GT Box 中心点与 Anchor Box 连接之间的最小角度,形状损失 Ω 主要描述了 GT Box 和 Anchor Box 的形状差异,其中, $\theta \in [2, 6]$, 该值决定了形状成本的重要性。

$$\begin{cases} \Omega = \frac{1}{2} \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \\ \omega_w = \frac{|\tau - \tau_{gt}|}{\max(\tau, \tau_{gt})}, \omega_h = \frac{|h - h_{gt}|}{\max(h, h_{gt})} \end{cases} \quad (4)$$

尽管 SIoU 在提升检测性能方面表现出色,但在处理军用飞机 MAR20 数据集中小尺寸飞机样本时,其效果并不理想。为了应对这一挑战,本文引入了 Inner-IoU^[22] 的理念,通过生成辅助边界框来计算 IoU 损失,增强模型对小样本的定位精度,同时加速预测边界框的回归过程。辅助边界框的引入使得模型能够更好地捕捉小尺寸目标的特征。Inner-SIoU 示意图如图 8 所示。Inner-SIoU 的定义如下:

$$L_{Inner-SIoU} = 1 - IoU_{inner} + \frac{\Delta + \Omega}{2} \quad (5)$$

式中:真实框(GT Box)和预测框(Pred Box)分别表示为 b^{gt} 和 b 。GT Box 和 Inner GT Box 的中心点表示为 (x_c^{gt}, y_c^{gt}) , 而 (x_c, y_c) 代表 Pred Box 和 Inner Pred Box 的中心

点。GT Box 的宽度和高度分别表示为 τ^{gt} 和 h^{gt} , Pred Box 的宽度和高度分别表示为 τ 和 h 。变量 $ratio$ 对应于比例因子,范围通常在 0.5~1.5 内。

$$\begin{cases} b_l^{gt} = x_c^{gt} - \frac{\tau^{gt} \times ratio}{2}, b_r^{gt} = x_c^{gt} + \frac{\tau^{gt} \times ratio}{2} \\ b_t^{gt} = y_c^{gt} - \frac{h^{gt} \times ratio}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} \times ratio}{2} \\ b_l = x_c - \frac{\tau \times ratio}{2}, b_r = x_c + \frac{\tau \times ratio}{2} \\ b_t = y_c - \frac{h \times ratio}{2}, b_b = y_c + \frac{h \times ratio}{2} \end{cases} \quad (6)$$

$$inter = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) \times (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)) \quad (7)$$

$$union = (\tau^{gt} \times h^{gt}) \times (ratio)^2 + (\tau \times h) \times (ratio)^2 - inter \quad (8)$$

$$IoU_{inner} = \frac{inter}{union} \quad (9)$$

本实验中取 $ratio$ 为 1.1,通过创建更大的辅助边界框,扩展回归的有效范围,从而提升模型对低 IoU 样本的回归效果,进而提高对军用飞机中小目标的定位精度。

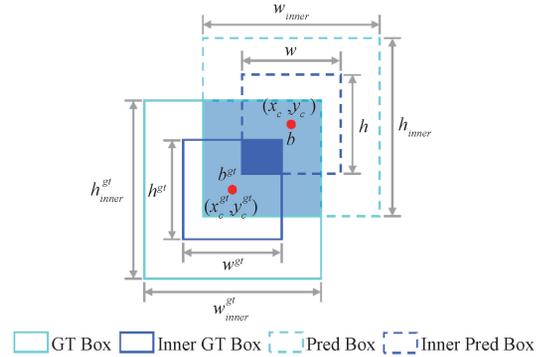


图 8 Inner-SIoU 示意图

Fig. 8 Inner-SIoU diagram

1.7 金字塔压缩注意力机制

在光学遥感图像中,目标与背景往往难以清晰区分,模型难以从整幅图像中准确提取所需的感兴趣区域。为了解决这个问题,利用通道注意力和空间注意力可以显著提升模型性能。本文引入一种低成本、高性能的新型注意力模块——金字塔压缩注意力(PSA)^[23]。该模块通过较低的计算复杂度来学习注意力权重,从而有效整合局部和全局注意力,以构建长距离的通道依赖关系。这一过程使得 PSA 模块能够灵活处理不同尺度的空间信息,并建立多尺度通道注意力的长期依赖关系。图 9 为 PSA 模块的工作原理。

PSA 模块通过 4 个步骤实现:首先,利用 SPC 模块对通道进行切分,SPC 模块工作原理如图 10 所示。

针对每个分割后的通道特征图,通过多尺度卷积核分组卷积的方法,提取不同尺度特征图中的空间信息,其中, Group 与 Kernel size 的关系式为: $G = 2^{\frac{K-1}{2}}$, 因此多尺度

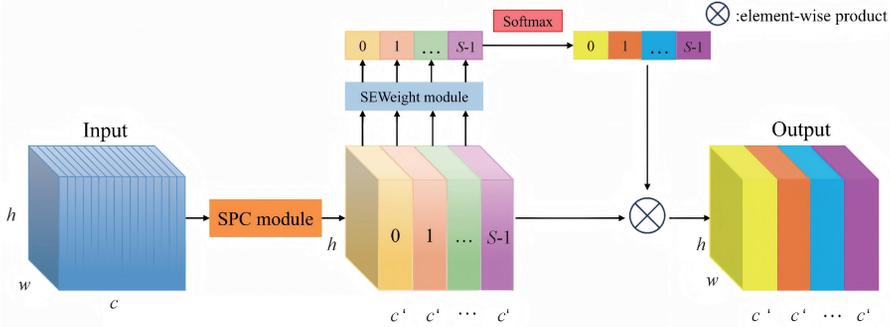


图 9 PSA 模块概念图

Fig. 9 Conceptual diagram of the PSA module

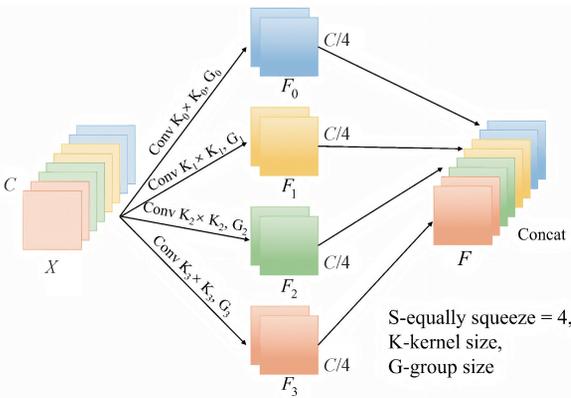


图 10 SPC 模块概念图

Fig. 10 Conceptual diagram of the SPC module

特征提取的过程如式(10)所示。

$$\begin{cases} F_i = Conv(k_i \times k_i, G_i)(X), i = 0, 1, 2 \dots S - 1 \\ K_i = 2 \times (i + 1) + 1 \\ F_i \in R^{C \times H \times W} \\ F = Cat([F_0, F_1, \dots, F_{S-1}]) \end{cases} \quad (10)$$

第二,通过 SEWeight 模块提取多尺度特征图的通道注意力,进而获得每个尺度的通道注意力向量。这一过程使得模型能够关注不同尺度下特征的重要性。SEWeight 模块工作原理如图 11 所示。

$$\begin{cases} Z_i = SEWeight(F_i), i = 0, 1, 2, \dots, S - 1 \\ Z = Z_0 \oplus Z_1 \oplus \dots \oplus Z_{S-1} \end{cases} \quad (11)$$

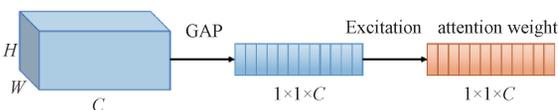


图 11 SEWeight 模块

Fig. 11 SEWeight module

第三,使用 Softmax 函数对这些向量进行归一化处理,从而生成新的多尺度通道注意力权重。

$$att_i = Softmax(Z_i) = \frac{\exp(Z_i)}{\sum_{i=1}^{S-1} \exp(Z_i)} \quad (12)$$

第四,将校准后的权重与相应的特征图逐元素相乘,输出经过注意力加权的特征图,使得模型能够更有效地捕捉到关键特征。

$$\begin{cases} Y_i = F_i \odot att_i, i = 1, 2, 3, \dots, S - 1 \\ Out = Cat([Y_0, Y_1, \dots, Y_{S-1}]) \end{cases} \quad (13)$$

2 实验与结果分析

2.1 实验环境与数据集

为评估改进的轻量级算法在遥感图像中对军用飞机目标的检测性能,实验所用的硬件环境包括 Intel(R) Xeon (R) Silver 4214R CPU@2.40 GHz,90 GB 内存和 Nvidia GeForce RTX 3080 Ti(12 G)显卡,软件环境为 Ubuntu 18.04 操作系统,深度学习框架为 Pytorch 1.9.0+cu111。实验参数设置如表 4 所示。

表 4 实验参数设置

Table 4 Experimental parameter setting

参数	设置
输入图像大小	640×640
迭代次数	250
批量大小	32
动量因子	0.937
学习率	0.01
权重衰减	0.0005
优化器	SGD

本文选用 MAR20^[1]数据集验证改进模型的有效性。该数据集是目前最大的军用飞机目标识别数据集,涵盖了从全球 60 个军用机场通过 Google Earth 收集的 3 842 张高分辨率遥感图像。这些图像中包含了 20 种不同型号的军用飞机,共计 22 341 个实例,图像的主要尺寸为 800×800 像素。为了确保模型的训练和评估能够科学有效地进行,该数据集被按 7:1:2 的比例划分为训练集、验证集和测

试集。

2.2 评价指标

为了验证轻量化改进模型在目标检测任务中的性能,本文选择了一系列关键的评价指标,包括精度、召回率、mAP0.5、模型参数量、模型大小以及每秒检测帧数(FPS)。在计算这些指标时,本文利用了真实检测结果中的真正例(TP)、假正例(FP)和假负例(FN)数据。这些指标共同为理解轻量化模型在实际应用中的效果提供了重要依据。

精度是指模型预测的阳性样本数与所有检测出的样本总数之间的比值,计算公式如式(14)所示。

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

召回率则表示模型正确预测的阳性样本数占实际阳性样本总数的比例。召回率的计算公式如式(15)所示。

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

平均准确度均值(mAP)是一个综合性指标,通过对各类样本的平均精度(AP)值进行加权平均,以衡量模型在不同类别上的检测性能,其计算方法是通过精度-召回率曲线下的面积来量化。mAP的高低直接反映了模型在处理复杂场景时的有效性和可靠性,mAP的计算公式如式(16)所示。

$$AP = \int_0^1 Precision(Recall)d(Recall) \quad (16)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

式中: AP_i 表示类别索引值为 i 的 AP 值, N 表示训练数据集中样本的类别数。mAP0.5 表示检测模型的 IoU 设置为 0.5 时的平均准确率。

2.3 消融实验

在主干网络进行特征整合时,本文引入了注意力机制以改善模型的训练过程,并对多种注意力模块对整体性能的影响进行了验证。实验结果如表 5 所示。引入 SKA^[24]

的模型参数量最大,达到 8.63 M,检测精度也最高,达到 96%,但参数量增加了 1.87 倍,偏离了轻量化设计理念。加入 ParNet^[25] 后,模型参数量为 3.74 M,检测精度为 95.7%,但 FPS 值只有 93 fps,检测速度较低。Double 模块的参数量最小,但检测精度和速度基本没有提升。PSA 模块在实现最高精度(95.8%)的同时,通过在瓶颈块中替换卷积,构建了高效的金字塔挤压注意力块,具有较小的参数量和最快的速度,本文最终选择 PSA 模块。

表 5 不同注意力模块性能

Table 5 Performance of different attention modules

模块	Params/M	FLOPs/G	AP50/%	FPS
Baseline	3.01	8.2	95.3	93
+SKA	8.63	12.7	96	89
+ParNet	3.74	8.8	95.7	93
+Double	3.08	8.3	95.3	96
+PSA	3.45	8.6	95.8	101

本文通过多组消融实验,利用 MAR20 数据集对 LeYOLO-MARs 模型的改进效果进行了系统性检验,实验结果如表 6 所示。表 6 中展示了以原始 YOLOv8n 模型为基线进行的对照实验。在实验 2 中,加入了改进后的倒置瓶颈模块 Im-bottleneck,结果表明该模块显著提升了模型的轻量化和计算效率。实验 3 引入 DNiN 模块替换原有的检测头,对比了两者的性能差异。实验 4 中,本文使用 Inner-SIoU 作为新的回归损失函数,展示了其相较于原 CIoU 的优势。实验 5 将 PSA 模块集成到模型的骨干网络中,以评估其在轻量化和特征提取方面的性能。实验 6 评估了 Im-bottleneck 和 DNiN 模块对检测性能的综合影响。实验 7 整合了实验 2、3、4 的结果,分析了 Im-bottleneck、DNiN 模块以及 Inner-SIoU 损失函数对检测性能的综合作用。最后,PSA 模块在轻量化和特征提取方面展现出了突出优势,增强了模型的整体能力。

表 6 MAR20 数据集消融实验结果

Table 6 ablation experiment results for the MAR20 dataset

实验	Im-bottleneck	DNiN	Inner-IoU	PSA	Params/M	FLOPs/G	AP50/%	模型尺寸/MB	FPS
1	×	×	×	×	3.01	8.2	95.3	5.99	93
2	√	×	×	×	1.37	3.8	92.9	2.98	73
3	×	√	×	×	2.34	5.6	95.4	5.45	96
4	×	×	√	×	3.01	8.2	95.6	5.98	98
5	×	×	×	√	3.45	8.6	95.8	7.16	101
6	√	√	×	×	1.09	2.5	93.5	2.45	76
7	√	√	√	×	1.09	2.5	94.3	2.46	76
8	√	√	√	√	1.71	3.0	95.7	3.65	78

这 8 组消融实验的结果充分验证了本文方法在提高遥感图像中军用飞机目标检测性能方面的有效性。实验 2 中通过使用改进的倒置瓶颈模块 Im-bottleneck 并修改网络结构,参数量减少了 54.4%,计算量降低了 53.7%,模型大小仅有 2.98 MB。在实验 3 中,引入轻量化的 DNiN 模块后,模型参数减少了 22.3%,计算量降低了 31.7%,mAP 较原模型提高了 0.1%。实验 4 中,仅更换了损失函数为 Inner-IoU,mAP 提升了 0.3%。在实验 5 中,通过引入 PSA 注意力模块,模型能够更准确地关注位置信息,并更加高效地提取目标特征,mAP 提升了 0.5%。实验 6~8 展示了整合多个模块的效果,使最终模型的准确率(mAP)提升了 0.4%。同时,相比于原始模型,计算量减少了 5.2 G,模型参数降低了 1.71 M,模型大小仅有 3.65 MB,符合轻量级网络设计的要求。

为了验证 Inner-SIoU 的有效性,对比改进算法与基准算法在训练过程中的损失,如图 12 所示。结果表明,改进算法的 box_loss 相比于改进前有所降低,并且收敛速度更快。

2.4 对比试验

为进一步验证本文算法的优势,首先将 LeYOLO-MARs 与其他 YOLO 系列算法进行比较,包括 YOLOv3-tiny、YOLOv5n、YOLOv5s、YOLOv7-tiny 和 YOLOv10n。由表 7 可知,在轻量化方面,LeYOLO-MARs 具有最低的参数量和计算量,其参数量仅为 YOLOv3-tiny、YOLOv5s 和 YOLOv7-tiny 的 19.6%、24.2% 和 28.1%,计算量则分

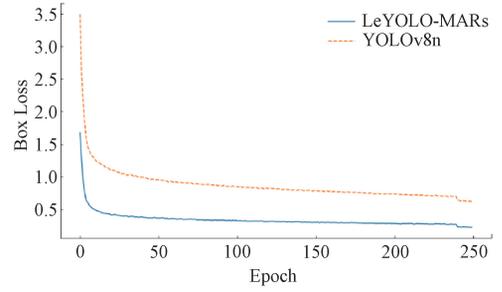


图 12 Box_loss 对比

Fig. 12 Box_loss Comparison

别为它们的 22.9%、18.6% 和 22.6%。在准确率方面,与同量级的 YOLOv5n 和 YOLOv8n 相比,LeYOLO-MARs 表现出最高的检测精度,且参数量最低。虽然 YOLOv10n 模型具有较高的检测精度,但其参数量和计算量较 LeYOLO-MARs 分别高出 0.6 倍和 1.8 倍。此外,本文还与其他学者改进的优秀算法进行了对比。YOLO-MAR^[14] 算法的 FPS 较高,但其计算量高达 11.3 GFLOPs,mAP 仅为 91.7%,与本文改进模型相比存在较大差距。FasterRCNN^[2] 和 PETDet^[26] 算法在计算量和参数量上过大,不符合轻量化要求,且检测精度不如本文算法。实验结果证明,LeYOLO-MARs 算法在遥感场景下能够实现精准检测,并且在移动端部署中表现优异,综合精度和轻量化方面达到了整体最优表现,适合实际应用。

表 7 不同模型的对比实验

Table 7 Comparison experiment of different models

算法	Params/M	FLOPs/G	AP50/%	AP50-90/%	FPS
YOLOv8n	3.01	8.2	95.3	72.3	93
YOLOv3-tiny	8.71	13.1	94.6	69.7	262
YOLOv5n	1.79	4.3	95.2	71.6	125
YOLOv5s	7.07	16.1	96.1	73.4	109
YOLOv7-tiny	6.07	13.3	96.3	72.8	152
YOLOv10n	2.71	8.4	96.2	73.2	73
FasterRCNN ^[2]	40.8	176	92.5	72.2	—
YOLO-MAR ^[14]	—	11.3	91.7	—	147
PETDet ^[26]	46.8	143	85.9	61.5	—
本文	1.71	3.0	95.7	89.1	78

为了验证本文算法训练得到的模型的适应性,本文在 DIOR 数据集上选取了不同分辨率与背景的遥感飞机图片进行测试,实验结果如表 8 所示。对比分析结果显示,尽管改进算法在 DIOR-plane 单类数据集上的精度提升并不显著,但与 YOLOv8n 相比,其参数规模和计算复杂度却显著降低。与 YOLOv3-tiny 相比,改进算法在精度上展现出明显的优势,同时保持了更轻量化的特性,使其更适合实时应用。虽然与 YOLOv5n 相比,二者在参数和模型大

小上相近,但改进算法在精度表现上更为出色。此外,尽管与 YOLOv7-tiny、YOLOv5s 和 YOLOv10n 的 mAP 值没有显著提高,改进算法的参数量和计算复杂度却大幅降低,这验证了其在实际应用中的实用性和有效性,尤其是在追求高效性与精确性的场景中。

2.5 可视化分析

梯度加权类别激活映射 (Grad-CAM) 被用于生成 YOLOv8n 和 LeYOLO-MARs 的热力图,旨在揭示模型

表 8 DIOR-Plane 数据集实验结果

Table 8 Experimental result on the DIOR-Plane dataset

算法	Params/	FLOPs/	AP50/	AP50-90/
	M	G	%	%
YOLOv8n	3.01	8.2	94.7	66.2
YOLOv3-tiny	8.67	13.0	94.1	62.8
YOLOv5n	1.77	4.2	94.5	65.6
YOLOv5s	7.02	15.9	95.3	68.0
YOLOv7-tiny	6.01	13.2	95.4	68.3
YOLOv10n	2.71	8.4	95.3	67.8
本文	1.71	3.0	94.8	66.7

在进行类别预测时所关注的具体区域。通过反向传播模型输出的类别置信度,可以提取出梯度值,进而生成热力图。在热力图中,高温区域表示模型对这些区域的关注程度较高,而低温区域则表明关注较少。实验结果如图 13 所示,YOLOv8n 对小型军用飞机的关注度较低,同时对稀疏区域内飞机的敏感性不足。相比之下,改进后的模型在处理背景噪声方面表现得更加优越,对军用飞机中心位置的关注更加精准,这一改进显著提升了边界框的定位准确性和整体检测效果。

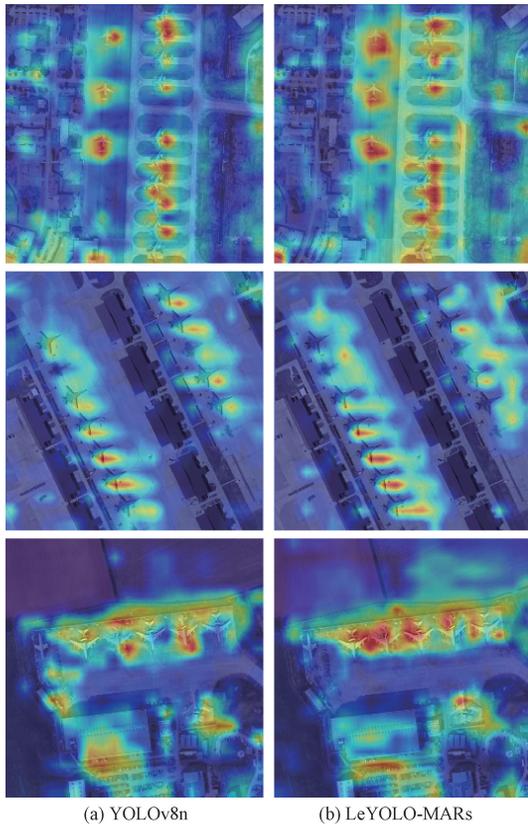


图 13 热力图对比

Fig. 13 Heat map comparison

为便于直观呈现本文算法的改进效果,对测试集中的每一张图片都进行了逐个检测,选取了具有代表性的复杂

背景数据进行可视化实验。这些数据包括尺度差异大、密集分布以及与背景区分度较小的情况,并将 YOLOv8n 与 LeYOLO-MARs 进行了直观对比。图 14 呈现了军用飞机在复杂背景条件下的检测效果。

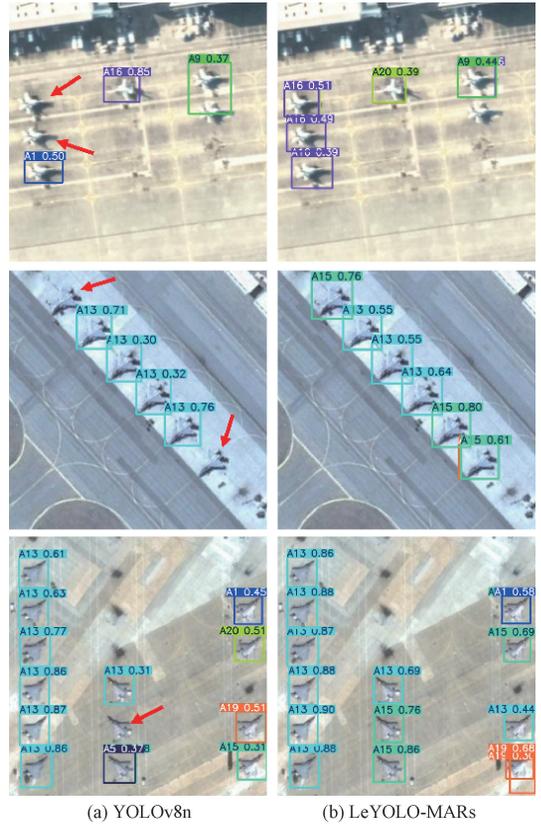


图 14 复杂背景军用飞机检测对比图

Fig. 14 Comparison chart of complex background military aircraft detection

热力图中的矩形框代表了模型的预测结果,而红色箭头则指示了漏检的位置。实验结果表明,基础的 YOLOv8n 模型在复杂背景下对军用飞机目标的检测效果不尽如人意,尤其在不同类型的军用飞机聚集时,漏检和误检问题更加突出。相比之下,改进后的模型显著减少了漏检现象,能够更加准确地定位光学遥感图像中的军用飞机目标,并有效地抑制了干扰信息,从而显著提高了检测精度。

3 结 论

本文提出了一种基于 YOLOv8n 的轻量级军用飞机目标检测算法,在满足检测精度的同时,显著减少了模型参数,优化了整体性能。具体而言,本文采用了优化后的倒置瓶颈模块替换原始主干网络中的经典瓶颈模块,更换高效的骨干网络特征提取模式,有效降低了计算需求并提升了计算速度。在颈部网络中引入了快速金字塔架构网络,减少了卷积层数并提高了语义信息共享的效率,从而简化

并减少了锁定和等待时间。引入轻量级解耦网络头,通过逐点卷积简化检测头结构,使每个网络预测能够独立完成分类和回归操作。使用 Inner-SIoU 替代原有的 CIoU 损失函数,提高了小目标的定位精度,并加快了网络的收敛速度。加入了金字塔压缩注意力机制模块,具有轻量化特性,弥补了模型精度的损失。实验结果表明,LeYOLO-MARs 在模型大小与计算效率方面均有显著提升,大幅降低了其在边缘设备上的运行难度。未来研究将致力于进一步提升该模型在实时设备上的探测速度,并增强其在复杂空域和特殊场景下的检测能力。

参考文献

- [1] 禹文奇,程攀,王美君,等. MAR20:遥感图像军用飞机目标识别数据集[J]. 遥感学报, 2023, 27(12): 2688-2696.
YU W Q, CHENG G, WANG M J, et al. MAR20: A benchmark for military aircraft recognition in remote sensing images[J]. National Remote Sensing Bulletin, 2023, 27(12): 2688-2696.
- [2] REN SH Q, HE K M, 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, 2016, 39(6): 1137-1149.
- [3] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [4] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]. IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [5] CAI ZH W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [6] JIANG P Y, ERGU D, LIU F Y, et al. A review of YOLO algorithm developments [J]. Procedia Computer Science, 2022, 199: 1066-1073.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, 2016: 21-37.
- [8] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [9] TAN M X, PANG R M, LE Q V. Efficientdet: Scalable and efficient object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [10] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision, 2020: 213-229.
- [11] 单慧琳,吕奎奎,付相为,等. 基于 DFECANet 的遥感图像飞机目标检测方法[J]. 电子测量与仪器学报, 2024, 38(2): 19-29.
SHAN H L, LYU Z K, FU X W, et al. Aircraft target detection in remote sensing images based on DFECANet [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(2): 19-29.
- [12] 张武,刘秀清. 基于改进 YOLOv5 的 SAR 图像飞机目标细粒度识别[J]. 国外电子测量技术, 2024, 43(6): 143-151.
ZHANG W, LIU X Q. Fine-grained recognition of aircraft targets in SAR images based on improved YOLOv5 [J]. Foreign Electronic Measurement Technology, 2024, 43(6): 143-151.
- [13] 杨鑫,王琼,姚亚洲,等. 基于 Faster R-CNN 改进的光学遥感图像飞机检测[J]. 激光与光电子学进展, 2023, 60(12): 427-437.
YANG X, WANG Q, YAO Y ZH, et al. Improved aircraft detection of optical remote sensing image based on Faster R-CNN [J]. Laser & Optoelectronics Progress, 2023, 60(12): 427-437.
- [14] 王杰,张上,张岳,等. 改进 YOLOv5 的军事飞机检测算法[J]. 无线电工程, 2024, 54(3): 589-596.
WANG J, ZHANG SH, ZHANG Y, et al. Military aircraft detection algorithm based on improved YOLOv5 [J]. Radio Engineering, 2024, 54(3): 589-596.
- [15] 刘丽,张硕,白宇昂,等. 改进 YOLOv8 的轻量级军事飞机检测算法[J]. 计算机工程与应用, 2024, 18: 114-125.
LIU L, ZHANG SH, BAI Y ANG, et al. Improved lightweight military aircraft detection algorithm of YOLOv8 [J]. Computer Engineering and Applications, 2024, 18: 114-125.
- [16] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [17] SANDLER M, HOWARD A, ZHU M L, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [18] ZHAO H SH, SHI J P, QI X J, et al. Pyramid scene parsing network [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.

- [19] CHEN Q, WANG Y M, YANG T, et al. You only look one-level feature[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13039-13048.
- [20] LIN M, CHEN Q, YAN SH CH. Network in network[J]. ArXiv preprint arXiv: 1312.4400, 2013.
- [21] GEVORGYAN Z. SIoU loss: More powerful learning for bounding box regression [J]. ArXiv preprint arXiv: 2205.12740, 2022.
- [22] ZHANG H, XU C, ZHANG SH J. Inner-IoU: More effective intersection over union loss with auxiliary bounding box[J]. ArXiv Preprint arXiv: 2311.02877, 2023.
- [23] ZHANG H, ZU K K, LU J, et al. EPSANet: An efficient pyramid squeeze attention block on convolutional neural network[C]. Asian Conference on Computer Vision, 2022: 1161-1177.
- [24] LI X, WANG W H, HU X L, et al. Selective kernel networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510-519.
- [25] GOYAL A, BOCHKOVSKIY A, DENG J, et al. Non-deep networks [J]. Advances in Neural Information Processing Systems, 2022, 35: 6789-6801.
- [26] LI W T, ZHAO D P, YUAN B, et al. PETDet: Proposal enhancement for two-stage fine-grained object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-14.

作者简介

杨瑞君, 博士, 副教授, 主要研究方向为计算机视觉、机器嗅觉等。

E-mail: yangruijun@sit.edu.cn

张浩(通信作者), 硕士研究生, 主要研究方向为目标检测。

E-mail: 17671241205@163.com

叶璟, 博士, 副教授, 主要研究方向为环境科学、水质基准和风险评估。

E-mail: yejinganna@sit.edu.cn