

文章编号 1004-924X(2024)20-3085-14

融合全局信息与双域注意力机制的 光学遥感飞机目标检测

林珊玲^{1,2}, 张 雪^{1,2}, 陈 燕^{1,2}, 林坚普^{1,2*}, 吕珊红^{1,2}, 林志贤^{1,2,3}, 郭太良^{2,3}

(1. 福州大学 先进制造学院, 福建 泉州 362251;

2. 中国福建光电信息科学与技术实验室, 福建 福州 350116;

3. 福州大学 物理与信息工程学院, 福建 福州 350116)

摘要:针对光学遥感图像中飞机目标检测算法因背景复杂、受测飞机目标较小以及飞机外观差异较小导致检测精度不足的问题,本文基于 YOLOv8n 模型,提出一种融合全局信息与双域注意力机制的光学遥感图像飞机目标检测算法。首先,设计了 SPPF_Global 模块,通过全局最大池化层提供全局特征概览,帮助模型在复杂环境中更好地区分目标与背景;其次,提出了双域注意力机制,通过空间域和通道域的信息引导,提高模型对机翼形状等重要区域的关注度,增强对不同飞机型号的细微差别的区分能力;最后,采用并行路径改进下采样模块并引入 Powerful-IoU 损失函数,通过自适应惩罚因子加速模型收敛,提高对小目标飞机的识别能力和预测框的回归效率。实验结果表明:与原始 YOLOv8n 相比,改进后的模型在公开数据集 MAR20 上的精确率、召回率、mAP50 以及 mAP50-95 分别提高了 3.3%, 2.6%, 3.2% 和 2.6%, 在 NWPU VHR-10 数据集上分别提高了 5%, 5.1%, 2.5% 和 0.3%;同时,参数量和运算量分别降低了 6.6% 和 3.7%。证明了本文模型的高效性和优越性,有效提升了光学遥感图像中飞机目标检测算法的应用价值。

关键词: 光学遥感图像; 飞机目标检测; YOLOv8; 注意力机制; Powerful-IoU

中图分类号: TP394.1 **文献标识码:** A **doi:** 10.37188/OPE.20243220.3085

Integrating global information and dual-domain attention mechanism for optical remote sensing aircraft target detection

LIN Shanling^{1,2}, ZHANG Xue^{1,2}, CHEN Yan^{1,2}, LIN Jianpu^{1,2*}, LÜ Shanhong^{1,2},

LIN Zhixian^{1,2,3}, GUO Tailiang^{2,3}

(1. School of Advanced Manufacturing, Fuzhou University, Quanzhou 362251, China;

2. Fujian Science and Technology Innovation Laboratory for Photoelectric Information,

Fuzhou 350116, China;

3. College of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China)

* Corresponding author, E-mail: ljp@fzu.edu.cn

Abstract: To address the problem of insufficient detection accuracy of aircraft targets in optical remote sensing images due to complex backgrounds, small targets, and similar appearances among aircraft, an aircraft target detection algorithm was proposed in this paper based on the YOLOv8n model that integrated

收稿日期: 2024-05-22; 修订日期: 2024-07-28.

基金项目: 国家重点研发计划 (No. 2022YFB3603705); 国家自然科学基金青年基金 (No. 62101132)

the global information and the dual-domain attention mechanism in optical remote sensing images. Firstly, the SPPF_Global module was designed to provide a global feature overview through the global maximum pooling layer, which helped the model better distinguish objects from the background in complex environments. Secondly, a dual-domain attention mechanism was proposed to improve the attention to important areas such as wing shape and other distinctive structures through the information guidance of space domain and channel domain, and enhanced the ability to distinguish the nuances of different aircraft models. Finally, the parallel path downsampling method and the Powerful-IoU loss function was introduced, and the adaptive penalty factor was used to accelerate the convergence of the model, which improved the recognition ability of the model for small target aircraft and the regression efficiency of the prediction frame. The experimental results show that compared with the original YOLOv8n, the accuracy rate, recall rate, mAP50 and MAP50-95 of the proposed model on the open data set MAR20 are increased by 3.3%, 2.6%, 3.2% and 2.6% respectively. On the NWPU VHR-10 dataset, the parameters are increased by 5%, 5.1%, 2.5% and 0.3% respectively, while the number of parameters and the calculation amount are decreased by 6.6% and 3.7% respectively, which proves the efficiency and superiority of the proposed model, and effectively improves the application value of the aircraft target detection algorithm in optical remote sensing images.

Key words: optical remote sensing image; aircraft target detection; YOLOv8; attention mechanism; Powerful-IoU

1 引言

随着科技的进步和传感器技术的发展,光学遥感图像的高分辨率、实时性和高可读性,使其在当前国防安全、城市规划以及军事等领域应用广泛,人们可以通过光学遥感图像获得清晰且丰富的地面目标信息^[1-2]。在遥感图像的众多目标中,飞机是现代战争中争夺制空权的关键武器,其出现的位置及型号时刻影响着战场的决策,因此,面向光学遥感图像的飞机目标检测具有重要的研究意义。

传统的目标检测方法多以人工设计特征为主,在面对海量数据和复杂背景等情况时,该方法费时费力且准确性、鲁棒性大大降低。随着深度学习的迅速发展,国内外学者基于深度学习的方法对遥感图像飞机目标检测做了大量研究。当前,基于深度学习的目标检测算法主要可分为两阶段和一阶段。两阶段的目标检测算法如 Region-based Convolutional Neural Networks (R-CNN)系列,利用区域生成网络(Region Proposal Network, RPN)生成潜在的目标候选区域,然后再将这些候选区域进行分类和精细调整,能够有效处理各种尺寸和形状的目标。而一阶段的目

标检测算法如 You Only Look Once (YOLO)系列和 Single Shot multibox Detector (SSD),具有快速检测速度和简单的网络结构,在前向传播过程中预测目标的边界框和类别,直接从图像中一次性预测结果^[3-5]。

这些算法在自然环境下能够较好地完成检测任务,但遥感图像的拍摄视角单一,目标尺寸小,以及飞机目标特征相似,使用通用目标检测算法会出现漏检、误检的问题。因此,研究人员提出了一系列针对遥感图像飞机目标检测的改进方法。Wu 等人^[6]通过自校准卷积和空洞卷积优化 Mask R-CNN,丰富了高层特征信息,以提高高分辨率遥感图像中密集目标和复杂背景的检测效果。Wang 等人^[7]在 Faster R-CNN 的基础上提出了飞机目标区域提议网络,用滑动窗口扫描特征图的方法,根据飞机尺寸与基座尺寸的比值生成对应尺寸的飞机尺寸的候选框,解决了不同飞机之间的尺度差异对模型检测效果的影响。Zhang^[8]等人模仿人工检测过程,提出了一个级联的三视图网络,三个视图分别为定位机场区域、检测飞机目标以及提取机场跑道,减少了检测范围和虚警,提高了检测精度。Yang 等人^[9]在 YOLOv4 的基础上使用深度可分离卷积替换传统卷

积,并在每个CSP连接单元中引入SE注意力机制,有效提高了飞机目标的检测精度。Zhang等人^[10]利用Efficient-RepGFPN改进YOLOv5的特征融合结构,将不同尺度特征图的信息联系得更加紧密,使有效特征得到更好的表达,有利于对图像中的小目标进行检测。Zhao等人^[11]通过设计多尺度策略用于飞机卫星图像目标检测任务,通过检测不同尺度下的目标来提高检测精度和覆盖率。

综合而言,光学遥感图像的飞机目标检测与在深度学习的引领下取得了显著进展,然而仍存在如下问题:(1)尽管部分方法针对高分辨率遥感图像中的小目标进行了优化,但仍然存在漏检和误检的问题,特别是在飞机目标尺寸较小时,检测算法的准确性和鲁棒性仍有待提高;(2)当前方法中,全局信息的利用程度有限,尤其是对于光学遥感图像中复杂背景下的飞机目标,算法往往缺乏全局上下文的有效利用,导致检测结果的不稳定性;(3)一些方法虽然在提高检测准确性方面取得了进展,但往往牺牲了模型的运行效率,导致算法在实际应用中可能存在计算资源消耗过大的问题。基于上述问题,本文基于更为轻量的一阶段目标检测算法进行研究,在YOLOv8n模型的基础上,提出一种融合全局信息与双域注意力机制的光学遥感飞机目标检测算法,在不增加模型复杂度的基础上,提高模型对小目标的识别能力、全局信息捕获能力以及对飞机细节特征的提取能力,从而提高飞机目标检测的准确性和鲁棒性。

2 融合全局信息与双域注意力机制的光学遥感飞机目标检测算法

2.1 网络整体结构

YOLOv8^[12]采用了CSPDarkNet结构作为特征提取模块(Backbone),并引入了梯度流更丰富的C2f模块替代了C3模块。同时,针对不同尺度的模型进行了通道数的调整以降低模型计算量,提升了收敛速度和效果;在特征融合模块(Neck)采用PAN-FPN结构,通过上采样、通道融合,最终将三个输出分支送入检测模块;在检测头模块(Head)采用解耦头结构进行检测,将回归分支和

预测分支进行分离,加速了模型的收敛;采用了TAL正样本分配策略计算损失,能够根据分类与回归的分数加权的分数选择正样本。

YOLOv8模型虽然在通用目标检测任务中表现优异,但是针对光学遥感图像的飞机目标检测仍对小目标、相似外观飞机的识别精度低、背景复杂情况下的检测困难等问题。YOLOv8提供了n/s/m/l/x 5个版本,为了更好地满足光学遥感图像飞机目标检测任务对计算资源、快速响应和轻量级部署的需求,本文基于体积最小的YOLOv8n模型进行改进,首先设计了SP-PF_Global模块,以增强全局信息捕获能力,减少复杂背景对模型检测效果的影响;其次在backbone与neck之间加入双域注意力机制(Dual-Domain Attention, DDA),从而增强模型对图像中细微特征的感知能力,并帮助模型更好地理解图像中的重要区域;然后引入并行下采样模块(Parallel Downsampling, PDS),帮助模型保留小目标的细节信息;最后通过Powerful-IoU损失函数给模型带来更快的收敛速度和更高的检测精度。

2.2 SPPF_Global模块

在快速空间金字塔池化模块^[13](Spatial Pyramid Pooling - Fast, SPPF)中,通过连续三次卷积核为 5×5 的最大池化,将池化前和每次池化后的结果进行拼接,以提取特征并将其转换为固定大小的特征向量。尽管SPPF模块在一定程度上能够捕获多尺度的特征,但其固定的池化窗口大小限制了模型对不同尺寸输入的适应性,这使得SPPF模块主要关注局部特征,对整体背景信息的感知能力相对较弱。因此,在复杂环境下,背景与目标之间的区分变得困难。

为了改进这一点,增强模型的全局信息捕获能力,减少复杂背景对模型检测效果的影响,本文在原始的SPPF模块中引入了全局最大池化层,提出了基于全局信息的快速空间金字塔池化模块(SPPF_Global)。

SPPF_Global模块的结构如图2所示,通过将两个自适应最大池化层(AdaptiveMaxPool2d)的输出大小设置为(1,1),使得池化窗口的大小与输入特征图的大小相匹配,从而实现对整个特征图的全局最大池化操作。无论输入图像的大小是多少,SPPF_Global模块都能够捕获整个图

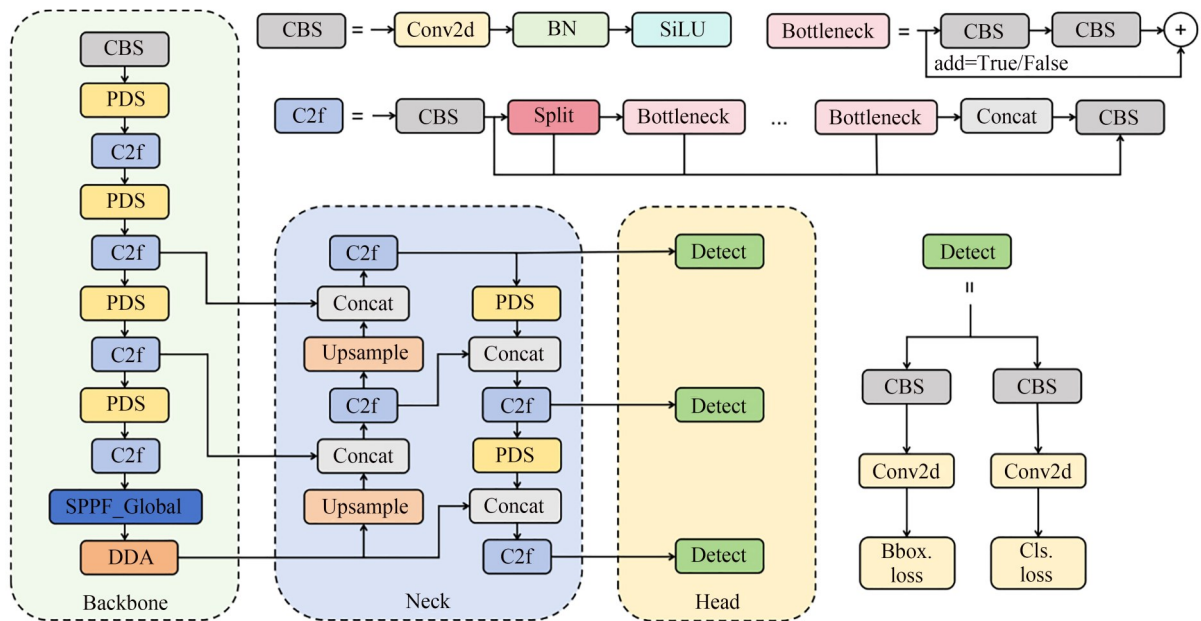


图1 网络模型整体结构图

Fig. 1 Overall structure of the network model

像的全局特征。综合考虑背景信息,在整个图像中寻找最显著的特征,并且这些特征往往与飞机目标的关键部位相关联,有助于减少背景干扰对飞机目标检测的影响,从而提高了飞机目标检测的准确性和鲁棒性。

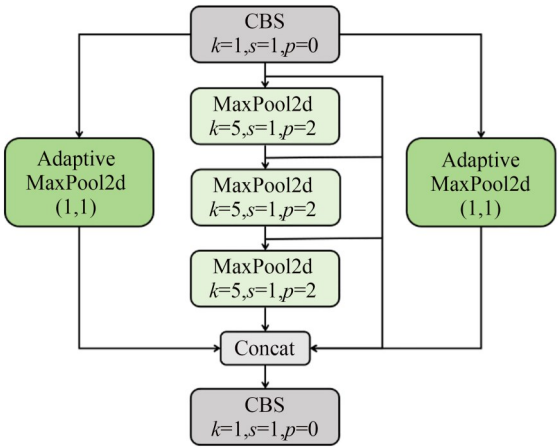


图2 SPPF_Global模块结构

Fig. 2 Structure of SPPF_Global module

2.3 双域注意力机制

由于各类飞机的整体特征高度相似,不同飞机型号的区别主要体现在局部细节的不同,而模型难以区分这些细小差距,导致飞机型号误检率高。因此,为了提高模型对不同型号飞机的检测精度,增强模型捕获和利用图像中重要的细节特

征信息的能力,本文提出了一种双域注意力机制(Dual-Domain Attention, DDA),通过空间域与通道域的信息引导,帮助模型关注图像中的重要区域,增强不同通道特征之间的关联性,从而提高模型对不同机型的检测精度。

DDA 融合了高效通道注意力^[14](Efficient Channel Attention, ECA)和空间注意力^[15](Spatial Attention, SA), DDA 模块的示意图如图 3 所示。

首先输入特征图会经过全局平均池化(Global Average Pooling, GAP),将特征图从大小为 $W \times H \times C$ 的矩阵变成大小为 $1 \times 1 \times C$ 的向量。然后,通过一维卷积操作得到每个通道的权重,完成跨通道间的信息交互。一维卷积的卷积核大小 k 由一个自适应函数来控制, k 的公式如式(1)所示:

$$k = \left\lceil \frac{\left| \log_2(c) + b \right|}{\gamma} \right\rceil_{\text{odd}}, \quad (1)$$

其中: $\lceil \cdot \rceil_{\text{odd}}$ 表示 k 取距离最近的奇数,式中的 γ 和 b 用于改变通道数 C 与卷积核大小 k 的比例,将其分别设置为 2 和 1^[14], k 的取值会根据通道数改变,能够充分融合部分通道间的交互。最后,通过 Sigmoid 激活函数得到归一化后的权重,将原始特征图与权重通过逐点元素相乘,得到具有通道维度上加权特征的输出并传递给 SA 模块。

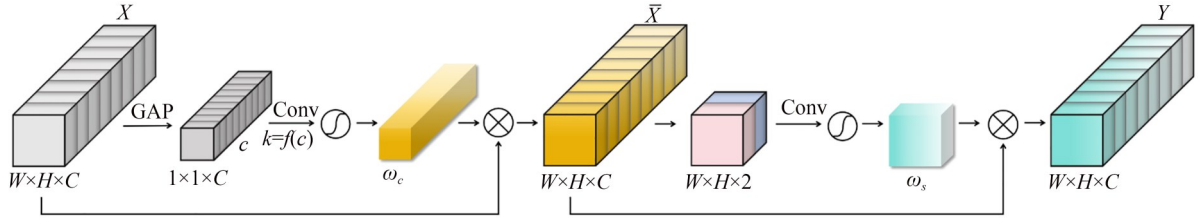


图3 DDA模块结构示意图

Fig. 3 Schematic of the structure of DDA module.

SA模块会将此时的特征图沿着通道维度分别进行求均值和求最大值的操作,得到两个特征图。然后将这两个特征图连接起来,通过卷积操作及Sigmoid激活函数得到注意力权重。最后,将权重应用于原始输入特征图上,得到加权后的特征图作为模型的输出。

通过DDA模块,模型能够同时关注到图像中的通道特征和空间位置,帮助模型更全面地理解图像,捕获并利用图像中重要的细节特征信息。这种细粒度的特征学习和关注机制有助于降低模型对于不同飞机型号之间细微差别的误判,从而提高飞机目标检测的检测精度。

2.4 下采样方式的改进

在计算机视觉任务中,下采样是指通过降低输入数据的空间维度来减少数据量。下采样操作会使特征图的空间尺寸减小,但却能增大每个神经元的感受野,这有助于模型捕捉更广泛的上下文信息,提取图像的重要特征,帮助模型更好地理解图像内容。然而,光学遥感图像中的飞机目标的尺寸普遍较小,在YOLOv8模型中仅单独采用步幅为2的卷积层进行下采样,这个过程中模型会偏向于学习图像中的大型目标,导致小目标的细节信息丢失,降低了模型对小目标的识别能力。

为了解决这个问题,本文采用了并行下采样模块^[16](Parallel Downsampling, PDS),通过两个并行的路径,同时使用最大池化和卷积操作实现下采样。为了在提取丰富特征和高效融合之间取得平衡,在Backbone和Neck阶段设置不同的通道数,图4为并行下采样模块结构图,展示了PDS在Backbone和Neck的结构及维度变化情况,其中 B 表示批次大小, C 表示通道数, H 表示高度, W 表示宽度。

首先,左边的分支采用最大池化操作,通过

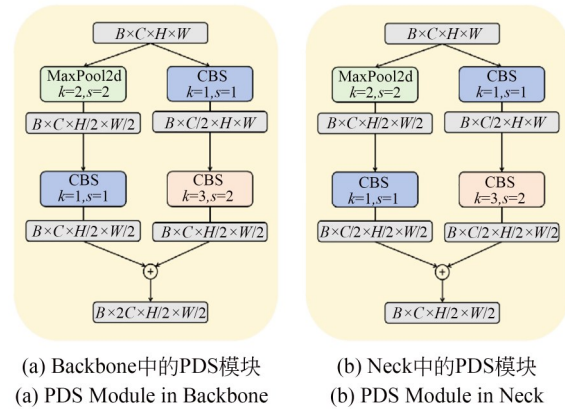
(a) Backbone中的PDS模块
(a) PDS Module in Backbone(b) Neck中的PDS模块
(b) PDS Module in Neck

图4 PDS模块结构

Fig. 4 Structure of PDS module

将输入特征图的空间尺寸减半来减小特征图的大小。随后,在Neck部分,通过一个 1×1 的卷积操作,进一步减少通道数以降低计算量,提高计算效率。在Backbone部分则通道数保持不变,使得模型能够更好地保留输入特征中的信息。右边的分支首先通过 1×1 卷积操作压缩通道,再使用步幅为2的 3×3 卷积操作减小特征图的尺寸,并且调整通道数以匹配左侧分支的输出通道数。最后,将左右两个分支的结果进行合并,完成下采样操作。

这样的下采样方式兼顾了最大池化和卷积操作,能够在不同尺度上提取特征,有助于保留小目标的细节信息。同时,并行策略可以更充分地捕捉特征信息,弥补了YOLOv8中下采样操作在光学遥感目标检测任务中存在的缺陷,提高模型对小目标的识别能力和检测精度。此外,由于最大池化操作不需要额外的参数,并且PDS模块中的第一个卷积操作能够将输入特征图的通道数减半,降低了模型的参数量和运算量。因此,PDS模块不仅能够提高模型的性能,还能够有效地减少模型的参数量和运算量,使得模型更加轻量化。

2.5 损失函数的改进

在目标检测中,一个高效的损失函数对于实现良好的定位性能至关重要。许多学者基于 IoU 的损失函数做了大量改进工作,这些改进工作都是围绕惩罚项进行的。在 YOLOv8 中采用了 CIoU^[17] 作为损失函数,其表达式如式(2):

$$L_{\text{CIoU}} = L_{\text{IoU}} + \frac{d^2}{c^2} + \alpha\vartheta, \quad (2)$$

其中: d 是预测框中心点和真实框中心点之间的距离, c 是预测框和真实框的最小外部框的对角线长度, α 是权重系数, ϑ 用来衡量长宽比的一致性, α 和 ϑ 的表达式如式(3):

$$\alpha = \frac{\vartheta}{1 - \text{IoU} + \vartheta}, \quad (3)$$

$$\vartheta = \frac{4}{\pi^2} \left(\arctan \frac{w_{\text{gt}}}{h_{\text{gt}}} - \arctan \frac{w}{h} \right)^2, \quad (4)$$

其中: w_{gt} 和 h_{gt} 分别表示真实框的宽和高, w 和 h 表示预测框的宽和高。当预测框与真实框不重叠时,单纯增大预测框的尺寸会降低 L_{CIoU} ,导致回归过程中预测框的扩大,从而影响收敛速度。此外,虽然 CIoU 能够反映中心点距离和纵横比差异,但是没有直接捕捉预测框与真实框之间的形状差异,也不能反映真实框大小的变化。而光学遥感图像中的飞机目标具有多样的形状和尺寸,并且会受到观测角度和距离的影响,这使得 CIoU 损失函数未能有效地适应飞机目标检测任务的需求。

因此,本文引入一种具有尺寸自适应性惩罚因子的边界框回归损失函数 Powerful-IoU^[18] (PIoU),帮助预测框更加直接有效地回归。PIoU 的表达式如下:

$$L_{\text{PIoU}} = L_{\text{IoU}} + 1 - e^{-P^2}, \quad (5)$$

其中, P 表示惩罚因子, P 的表达式如式(6):

$$P = \frac{\frac{dw_1}{w_{\text{gt}}} + \frac{dw_2}{w_{\text{gt}}} + \frac{dh_1}{h_{\text{gt}}} + \frac{dh_2}{h_{\text{gt}}}}{4}, \quad (6)$$

其中: dw_1, dw_2, dh_1, dh_2 是预测框与真实框对应边的距离绝对值, w_{gt} 和 h_{gt} 是真实框的宽和高。CIoU 和 PIoU 表达式中所涉及的参数如图 5 所示。

由于 P 值与预测框和真实框的最小外接框的大小无关,因此扩大预测框不会改变 P 值,这有效解决了预测框放大问题,加快了收敛速度。

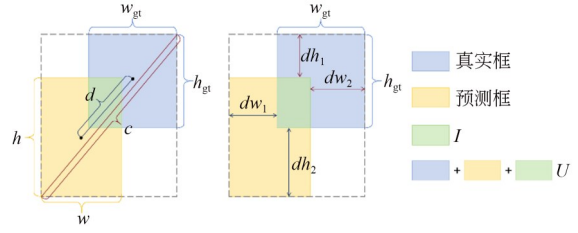


图 5 CIoU、PIoU 参数示意图

Fig. 5 Schematic of CIoU and PIoU Parameters

此外,PIoU 的惩罚函数,即 $f(P) = 1 - e^{-P^2}$,可以根据预测框的质量自适应地调整梯度幅度,当惩罚因子 P 较大(>2)时,表示预测框和真实框之间存在较大差异, $f'(P)$ 取较小的值,从而抑制了来自低质量锚框的有害梯度。当 P 接近 1 时,表示预测框接近真实框,较大的 $f'(P)$ 会加快回归速度。 $f'(P)$ 随着预测框质量的提升而逐渐减小,直至稳定地优化到与真实框完全对齐,给飞机目标检测带来更快的收敛速度和更高的检测精度。

3 实验结果及分析

3.1 实验环境与参数设置

本文实验的硬件环境配置如下:操作系统为 Ubuntu 20.04.5 LTS, CPU 为 AMD EPYC 9754 128-Core, GPU 为 NVIDIA GeForce RTX 3090,显存容量为 24 GB。本文方法基于 Python 3.8.18 和 Pytorch 1.13.1 深度学习框架实现。模型训练总迭代轮数为 300 轮,初始学习率设定为 0.01,使用随机梯度下降 SGD 优化器,学习动量为 0.937。输入训练样本大小设置为 640×640 。

3.2 数据集

MAR20 数据集^[19]是当前公开规模最大的用于军用飞机目标识别的光学遥感图像数据集之一。数据来自 Google Earth,包含 3 842 张图像,涵盖了 20 种不同型号的军用飞机,编号为 A1 到 A20。这些图像大多尺寸为 800×800 pixels。由于数据集中的 20 个细粒度类别都属于同一个飞机大类,不同型号的飞机具有相似的外观特征,因此类间相似性较高。数据集按照不同机场包含的各飞机型号的目标数量划分为训练集和验证集。其中,训练集包括 1 331 张图像和 7 870 个

目标实例,验证集包含 2 511 张图像和 14 471 个目标实例。

NWPU VHR-10 数据集^[20]包含 800 张光学遥感图像,其中 650 张为有目标图像,150 张为背景图像,共 3 651 个实例。数据集的类别为飞机、船、储罐、棒球场、网球场、篮球场、田径场、港口、桥和车辆共 10 个类别。数据集按照 6:2:2 的比例随机划分为训练集、验证集和测试集。

3.3 公共评价指标

为验证本文算法性能,采用精确率(Precision, P)、召回率(Recall, R)、平均精度均值(Mean Average Precision, mAP)、参数量(Parameters)、浮点运算次数(Floating-point Operations Per Second, FLOPs)、在 GPU 条件下每张图的推理时间(Latency)作为评价指标。上述指标的计算表达式分别为:

$$Precision = \frac{TP}{TP + FP}, \quad (7)$$

$$Recall = \frac{TP}{TP + FN}, \quad (8)$$

$$mAP = \frac{\sum_{i=1}^n AP_i}{n}, \quad (9)$$

其中:TP, FP 和 FN 分别表示正确检测到目标框、误检框和漏检框的数量;AP 表示平均精度, mAP 是对所有类别的 AP 求平均值所得, mAP50 和 mAP50-95 分别表示不同 IoU 阈值下的平均精度均值; n 表示检测的类别数。

3.4 不同空间金字塔池化模块的实验

为了验证本文提出的 SPPF_Global 模块在光学遥感图像飞机目标检测任务上的有效性,本文在 MAR20 数据集上基于 YOLOv8n 模型对比了 SPPF_Global 模块与其他主流空间金字塔池化模块(SPP^[21], SPPF, ASPP^[22], RFB^[23], SPPC-SPC^[16]和 SimCSPSPPF^[24])对模型大小和精度的影响。实验结果如表 1 所示。

从表 1 的实验结果中可以看出,不同空间金字塔池化模块在模型性能上存在一定差异。其中, SPPF_Global 模块在精确率、召回率、mAP50 和 mAP50-95 指标上均表现优异。相比于原始模型使用的 SPPF 模块, SPPF_Global 模块在 mAP50 上提升了 1%, 在 mAP50-95 上提升了 0.8%, 有效提高了模型的检测精度。此外, SPPF_Global 模块在参数量和计算量方面与 SPPF 模块相近, 保持了较低的资源消耗水平。而推理速度也与 SPPF 模块相当, 保持在 0.88 ms。这表明 SPPF_Global 模块在保持资源消耗较低的同时, 成功提升了模型的检测性能。对比其他主流空间金字塔池化模块, SPPF_Global 模块表现出更高的性能和更低的资源消耗。这些结果表明了 SPPF_Global 模块的有效性, 它在保持了模型速度的同时, 提升了检测的准确性, 为光学遥感图像飞机目标检测任务提供了一个有效的选择。

表 1 不同空间金字塔池化模块的性能对比

Tab. 1 Comparison of performance of different spatial pyramid pooling modules

Module	Precision	Recall	mAP50	mAP50-95	Parameters/M	FLOPs/B	Latency/ms
SPP	78.3%	77.2%	83.2%	62.2%	3.01	8.1	0.89
SPPF	81.0%	78.7%	84.6%	63.2%	3.01	8.1	0.88
ASPP	81.5%	77.4%	84.5%	63.2%	3.83	8.7	0.91
RFB	81.1%	77.0%	84.0%	62.8%	3.18	8.2	0.89
SPPCSPC	81.3%	78.1%	84.2%	63.1%	4.62	9.4	0.97
SimCSPSPPF	81.0%	78.1%	84.9%	63.4%	3.36	8.4	0.90
SPPF_Global	82.2%	78.5%	85.6%	64.0%	3.06	8.0	0.88

3.5 不同注意力机制的实验

为了评估 DDA 模块对模型大小和精度方面的影响,在 MAR20 数据集上进行了实验,与其他注意力机制 Convolutional Block Attention Module^[25](CBAM), Coordinate Attention^[26](CA),

Channel prior convolutional attention^[27](CPCA), Deformable Large Kernel Attention^[28](DLKA), Efficient Multi-Scale Attention^[29](EMA), Large Separable Kernel Attention^[30](LSKA), Shuffle Attention^[31](SA), A Simple, Parameter-Free At-

tention Module^[32] (SimAM), Triplet Attention^[33] (TA)共 9 种进行了性能对比,表 2 为不同注意力机制的性能对比结果。

从表中可以看出,在原始 YOLOv8n 模型上加入 DDA 模块之后,模型的精确率、召回率、mAP50 和 mAP50-95 分别提高了 0.7%,0.8%,1.3%,0.9%。说明 DDA 模块通过空间域和通道域的信息引导,有效地提升了模型对图像中重要细节特征的捕获和利用能力,降低了模型在不同飞机型号之间的误判率。值得注意的是,添加 DDA 模块后,模型的参数量、运算量和推理时间保持不变。这表明 DDA 在提升模型性能的同

时,没有增加资源损耗。与其他注意力机制进行性能比较,DDA 模块表现出了显著的优势。其中添加 SimAM 后模型的精确率最高,但是也导致了召回率的下降,这说明此时模型更专注于准确识别那些确定为飞机的区域,但可能会错过一些细小的飞机部分或者与飞机相似的目标,从而导致漏检。DLKA 的精度略高于 DDA,但是其需要更高的计算成本,不利于资源受限或实时推理的场景下的应用。

因此,综合考虑检测精度与模型大小,本文提出的 DDA 模块都优于其他注意力机制,更加适用于光学遥感图像飞机目标检测任务。

表 2 不同注意力机制的性能对比

Tab. 2 Comparison of performance of different attention mechanisms

Model	Precision	Recall	mAP50	mAP50-95	Parameters/M	FLOPs/B	Latency/ms
YOLOv8n	81.0%	78.7%	84.6%	63.2%	3.01	8.1	0.88
+CBAM	80.7%	79.0%	84.6%	63.1%	3.08	8.2	0.89
+CA	81.0%	78.4%	84.8%	63.6%	3.02	8.1	0.95
+CPCA	81.5%	78.8%	85.0%	63.6%	3.14	8.3	1.21
+DLKA	82.4%	78.9%	86.0%	64.2%	4.64	9.4	0.90
+EMA	81.9%	76.5%	84.6%	63.3%	3.02	8.2	0.90
+LSKA	80.6%	78.0%	84.8%	63.5%	3.08	8.2	0.89
+SA	80.6%	77.7%	84.8%	63.4%	3.01	8.1	0.89
+SimAM	82.7%	77.3%	85.1%	63.6%	3.01	8.1	0.89
+TA	80.4%	79.1%	85.1%	63.7%	3.01	8.1	0.90
+DDA	81.7%	79.5%	85.9%	64.1%	3.01	8.0	0.88

3.6 消融实验

为了验证文中所提出的各模块及模块组合的有效性,本文在 MAR20 数据集上展开一组消融实验,

其中“√”表示在 YOLOv8n 中加入该模块,“×”表示没有加入该模块。实验结果如表 3 所示。

由表中数据可知,采用 SPPF_Global 模块后

表 3 消融实验

Tab. 3 Ablation experiments

SPPF_Global	DDA	PDS	PIoU	Precision	Recall	mAP50	mAP50-95	Parameters/M	FLOPs/B	Latency/ms
×	×	×	×	81.0%	78.7%	84.6%	63.2%	3.01	8.1	0.88
√	×	×	×	82.2%	78.5%	85.6%	64.0%	3.06	8.0	0.88
×	√	×	×	81.7%	79.5%	85.9%	64.1%	3.01	8.1	0.88
×	×	√	×	83.1%	78.5%	86.0%	64.6%	2.74	7.8	1.04
√	√	×	×	82.6%	79.7%	86.5%	64.6%	3.06	8.0	0.89
√	√	√	×	82.7%	80.8%	87.0%	65.6%	2.81	7.8	1.06
√	√	√	√	84.3%	81.3%	87.8%	65.8%	2.81	7.8	1.06

精确率、mAP50和mAP50-95分别提高了1.2%,1%和0.8%。这表明SPPF_Global模块能够有效地提高模型的检测准确性,减少了误检情况,为模型带来了明显的性能提升。

尽管单独使用DDA模块也能提高检测精度,但是与SPPF_Global结合后,模型的总体性能进一步得到了提升,这说明SPPF_Global模块和DDA注意力机制具有互补性,它们的组合能够综合利用全局信息和细粒度的注意力机制,提高了模型对目标的检测准确性和鲁棒性。

使用PDS模块改进YOLOv8n的下采样方式虽然因为引入分支和拼接操作略微增加了0.16 ms的推理时间,但却也成功降低了参数量和运算量,有效地改善了模型对小目标的检测能

力,使得模型能够更好地捕获细节信息并提高检测精度。将PDS模块引入SPPF_Global模块和DDA注意力机制组合的模型后,模型的精确率、召回率、mAP50和mAP50-95较原始模型分别提高1.7%,2.1%,2.4%,2.4%。

采用PIoU作为损失函数后,精确率和召回率分别提高了1.6%,0.5%,mAP50和mAP50-95分别提高了0.8%,0.2%,说明PIoU能够帮助预测框更加直接有效地回归。

最后,综合以上改进,相较于原模型,改进后的模型精确率、召回率、mAP50和mAP50-95分别提高了3.3%,2.6%,3.2%,2.6%,参数量和运算量分别降低6.6%,3.7%,更加满足光学遥感图像飞机目标检测的需求。

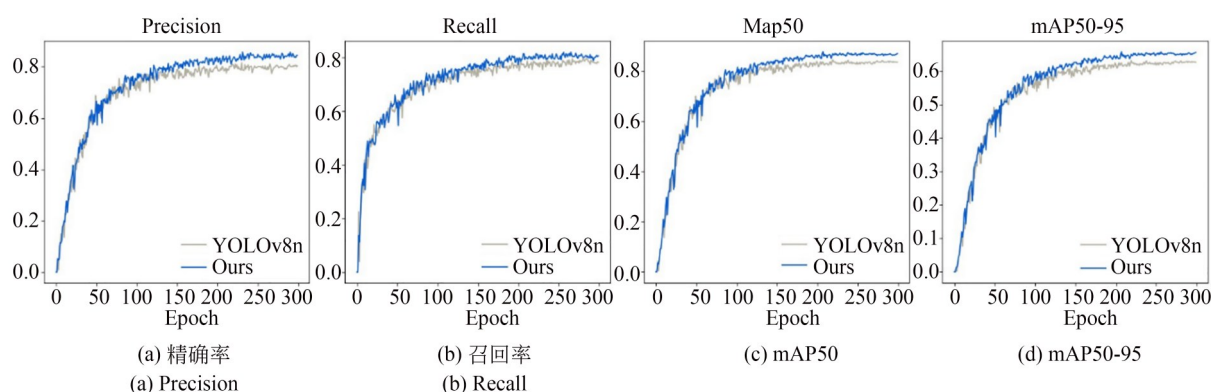


图6 训练曲线图

Fig. 6 Training curve

图6(a)~6(d)分别为模型改进前后精确率、召回率、mAP50和mAP50-95的训练曲线,能够清晰地观测到改进后的有效性。

3.7 对比实验

为了更好地验证本文改进的模型在光学遥感图像飞机目标检测任务上的先进性,本文在MAR20数据集上与Faster R-CNN^[34], RT-DETR^[35], Gold-YOLO-n^[36], YOLOv5n^[13], YOLOv5s, YOLOv6n^[24], YOLOv7-tiny^[16]和YOLOv8n共8种模型进行性能对比,对比结果如表4所示,其中加粗表示最优值,下划线表示次优值。

从表中可以看到,本文模型在mAP50和mAP50-95这两项指标上表现最为优异,这表明本文模型在飞机目标检测任务上精确性和鲁棒性都有很好的表现。虽然在精确率和召回率上

略低于RT-DETR,但是RT-DETR需要大量的计算资源,从其高参数量、运算量和较长的推理时间可以看出。相比之下,本文模型在保持较高性能的同时,在资源消耗和推理时间方面表现得更为高效。在模型体积方面,YOLOv5n的体积最小,但其精度相对较低,这说明虽然较小的模型可能会节省资源,但会以降低精度为代价。因此,本文模型在平衡精度和资源消耗方面表现出更好的效果。

综上所述,本文模型的性能最佳。不仅在精度和鲁棒性上优于其他模型,而且保持较低的资源消耗和较快的推理速度,在光学遥感图像飞机目标检测任务上具有较高的实用性和先进性。

图7~图8分别展示了不同模型在复杂环境及小目标环境下的可视化结果,每种飞机型号

表 4 MAR20 数据集对比实验
Tab. 4 Comparative experiments on MAR20 dataset

Model	Precision	Recall	mAP50	mAP50-95	Parameters/M	FLOPs/B	Latency/ms
Faster R-CNN	66.1%	61.1%	80.9%	58.7%	41.75	91.23	8.74
RT-DETR	87.7%	83.4%	83.0%	61.9%	19.9	57.0	4.34
Gold-YOLO-n	78.2%	75.7%	81.8%	59.8%	5.61	12.07	1.35
YOLOv5n	80.6%	75.4%	81.9%	59.6%	1.79	4.2	0.72
YOLOv5s	82.6%	79.5%	<u>85.9%</u>	<u>63.5%</u>	7.06	15.9	1.53
YOLOv6n	79.7%	74.3%	81.9%	59.4%	4.63	11.35	0.99
YOLOv7-tiny	79.9%	75.5%	82.5%	59.5%	6.06	13.2	<u>0.85</u>
YOLOv8n	81.0%	78.7%	84.6%	63.2%	3.01	8.1	0.88
Ours	<u>84.3%</u>	<u>81.3%</u>	87.8%	65.8%	<u>2.81</u>	<u>7.8</u>	1.06

用特定颜色框表示(彩图见期刊电子版)。可以发现,RT-DETR 和 Gold-YOLO-n 虽然在这两种环境下都检测出了飞机目标,但它们过分关注图像中的细节,将一些无关的背景噪声和纹理误判为目标;YOLOv5s 和 YOLOv7-tiny 虽然没有受到复杂环境的影响,但是缺乏对小目标的适应性,YOLOv7-tiny 不仅将地面上的飞机痕迹误判为目标,而且还出现了漏检和错检的情况,对于小目标的检测能力较弱;YOLOv6n 和 YOLOv8n 则在复杂环境和小目标环境下都未

能准确定位和区分飞机目标。相比于其他模型,本文模型不仅成功准确地找到了飞机的位置,并进行了正确分类,在定位和分类方面表现出色,主要原因在于本文模型采用了 SP-PF_Global 模块,增强了模型对整体图像情况的理解能力,从而提高了对复杂背景的抵抗能力。另外,双域注意力机制增强了模型对于关键区域的捕获能力,提高了对飞机机型的识别准确性。并行下采样模块进一步改进了模型的尺度感知能力,帮助小目标更好地被检测和定位。

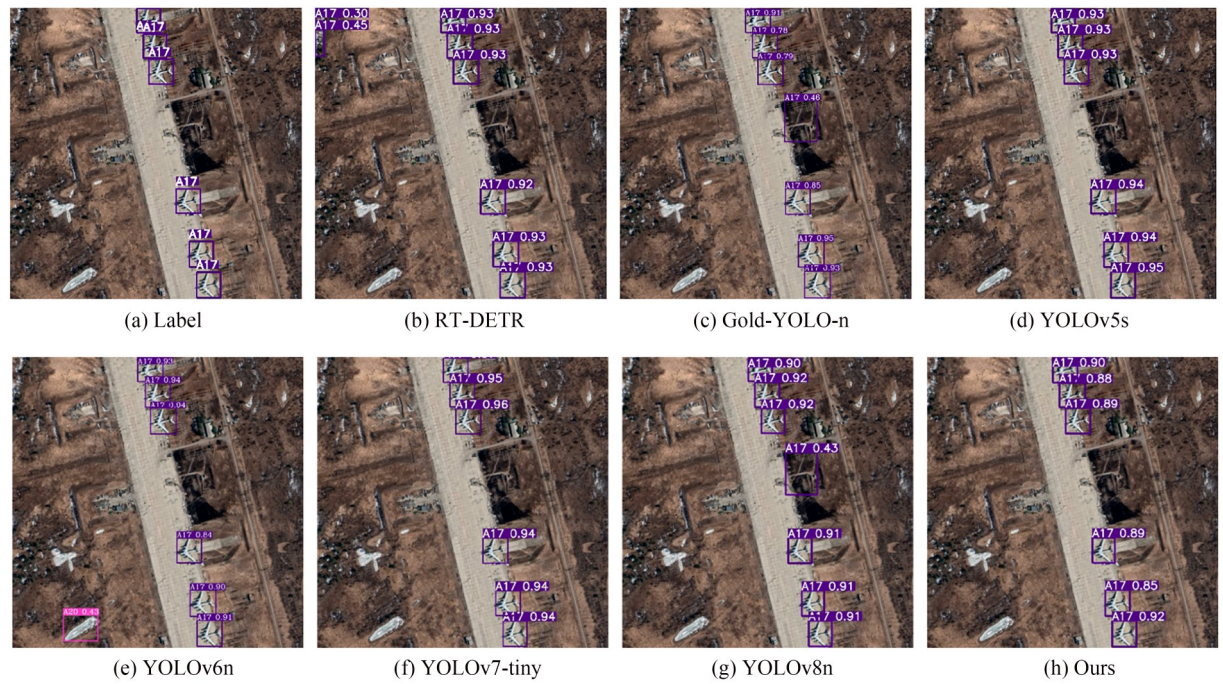


图 7 复杂环境下不同模型的可视化结果

Fig. 7 Visualization results of different models in complex environments

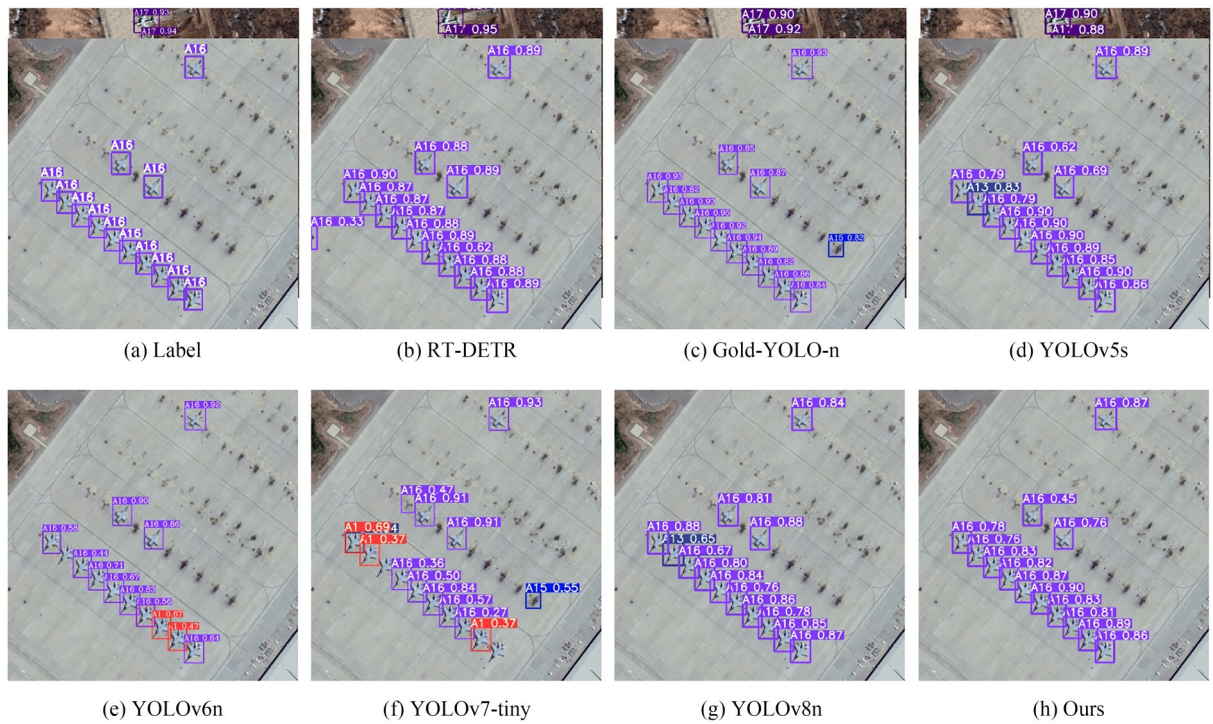


图 8 小目标环境下不同模型的可视化结果
Fig. 8 Visualization results of different models with small targets

最后,为了验证改进后模型在不同数据集中的泛化性,在NWPU VHR-10数据集上进行了对比实验,表5为对比实验结果。根据表中数据可以看出,改进后模型在NWPU VHR-10数据集上的性能优越。精确率、召回率、mAP50和mAP50-95分别达到了91.1%,87.2%,90.2%,54.4%,较原始模型提升了5%,5.1%,2.5%,

0.3%。其中改进后模型的召回率和mAP50达到是表中最高的,高召回率意味着模型能够找到更多的真实目标,较高的mAP50表示模型在较低的IoU阈值下具有卓越的精确度,这两个指标的提升表明改进后模型不仅在检测准确度上取得了明显的进步,而且在整体的目标覆盖率和检出性能上都得到了显著改善。

表 5 NWPU VHR-10数据集对比实验
Tab. 5 Comparative experiments on NWPU VHR-10 dataset

Model	Precision	Recall	mAP50	mAP50-95	Parameters/M	FLOPs/B	Latency/ms
RT-DETR	87.6%	<u>86.3%</u>	88.9%	56.1%	19.88	57.0	4.34
Gold-YOLO	89.9%	79.0%	86.0%	48.2%	5.61	12.06	2.07
YOLOv5n	90.2%	78.6%	86.1%	46.8%	1.77	4.2	1.20
YOLOv5s	92.5%	85.3%	<u>89.6%</u>	51.6%	7.04	15.8	1.79
YOLOv6n	85.2%	80.9%	85.6%	47.7%	4.63	11.34	1.55
YOLOv7-tiny	87.2%	87.0%	89.4%	49.5%	6.03	13.1	2.29
YOLOv8n	86.1%	82.1%	87.7%	54.1%	3.01	8.1	0.87
Ours	<u>91.1%</u>	87.2%	90.2%	<u>54.4%</u>	<u>2.80</u>	<u>7.8</u>	<u>1.05</u>

这些优势共同表明改进后模型在该数据集上的泛化能力和实用性大幅增强。此外,改进后模型在资源效率和推理速度方面表现得也非常

出色,远低于其他表现较好的模型,这意味着在保持高性能的同时,模型的复杂度和计算资源需求相对较低,改进后模型的预测延迟仅为

1.05 ms, 显著快于其他模型。综合来看, 本文模型在 NWPU VHR-10 数据集上的性能表现优异, 既保持了较高的精度和召回率, 同时在资源效率和推理速度方面也表现得非常出色, 进一步验证了本文模型在不同数据集中的泛化能力。

4 结 论

本文基于 YOLOv8n 模型提出了一种融合全局信息与双域注意力机制的光学遥感图像飞机目标检测算法。首先, 设计了 SPPF_Global 模块, 引入全局最大池化层以增强对整个图像全局特征的捕获能力, 减少背景干扰, 提高检测准确性。其次, 通过融合空间域和通道域信息, 提出 DDA 注意力机制, 加强模型对图像中重要区域和关键部件的关注, 有助于更好地区分不同飞机型号的细微差别。随后, 采用并行路径改进下采

样模块, 提升了小目标识别能力。最后, 通过引入了 PIoU 损失函数, 以自适应惩罚因子帮助预测框更直接有效地回归, 提高了模型的收敛速度和检测精度。本文在公开数据集 MAR20 和 NWPU VHR-10 上进行了实验, 结果表明, 与原始 YOLOv8n 相比, 改进后的模型在 MAR20 数据集上精确率、召回率和 mAP50 分别提高了 3.3%, 2.6%, 3.2% 和 2.6%, 在 NWPU VHR-10 数据集上分别提高了 5%, 5.1%, 2.5% 和 0.3%。同时, 模型的参数量和运算量分别降低了 6.6% 和 3.7%, 这些改进综合提升了模型在光学遥感图像飞机目标检测任务中的性能, 并展现了在不同数据集上的泛化性能, 为高效、准确地检测飞机目标提供了可靠的解决方案。在未来的研究中, 将探索将图像分割技术与目标检测相结合, 以进一步提升遥感图像中飞机目标检测算法的性能。

参考文献:

- [1] 黄泽贤, 吴凡路, 傅瑶, 等. 基于深度学习的遥感图像舰船目标检测算法综述[J]. 光学精密工程, 2023, 31(15): 2295-2318.
HUANG Z X, WU F L, FU Y, *et al.* Review of deep learning-based algorithms for ship target detection from remote sensing images[J]. *Opt. Precision Eng.*, 2023, 31(15): 2295-2318. (in Chinese)
- [2] 刘忻伟, 朴永杰, 郑亮亮, 等. 面向航天光学遥感复杂场景图像的舰船检测[J]. 光学精密工程, 2023, 31(6): 892-904.
LIU X W, PIAO Y J, ZHENG L L, *et al.* Ship detection for complex scene images of space optical remote sensing[J]. *Opt. Precision Eng.*, 2023, 31(6): 892-904. (in Chinese)
- [3] WANG Z, LI P, CUI Y C, *et al.* Automatic detection of individual trees in forests based on airborne LiDAR data with a tree region-based convolutional neural network (RCNN) [J]. *Remote Sensing*, 2023, 15(4): 1024.
- [4] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: single shot multibox detector[C]. *European Conference on Computer Vision*. Cham: Springer, 2016: 21-37.
- [5] 林珊玲, 彭雪玲, 王栋, 等. 多尺度增强特征融合的钢表面缺陷目标检测[J]. 光学精密工程, 2024, 32(7): 1075-1086.
- [6] LIN S L, PENG X L, WANG D, *et al.* Object detection of steel surface defect based on multi-scale enhanced feature fusion[J]. *Opt. Precision Eng.*, 2024, 32(7): 1075-1086. (in Chinese)
- [7] WU Q F, FENG D Q, CAO C Q, *et al.* Improved mask R-CNN for aircraft detection in remote sensing images[J]. *Sensors*, 2021, 21(8): 2618.
- [8] WANG B, ZHOU Y, ZHANG H N, *et al.* An aircraft target detection method based on regional convolutional neural network for remote sensing images [C]. 2019 *IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*. Beijing, China. IEEE, 2019: 474-478.
- [9] ZHANG L B, LI C Y, ZHAO L J, *et al.* A cascaded three-look network for aircraft detection in SAR images[J]. *Remote Sensing Letters*, 2020, 11(1): 57-65.
- [10] YANG Y K, XIE G R, QU Y. Real-time detection of aircraft objects in remote sensing images based on improved YOLOv4[C]. 2021 *IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. Chongqing, China. IEEE, 2021: 1156-1164.
- [11] ZHANG D Y, ZHAO Z H, HUANG S H. Investigation of aircraft target detection of remote sens-

- ing images based on the improved YOLOv5[C]. 2023 *International Conference on Computer Applications Technology (CCAT)*. Guiyang, China. IEEE, 2023: 266-270.
- [11] ZHAO T J, QIAO N. Research on target detection technology of aircraft satellite images based on improved YOLOv5 model[C]. 2023 *4th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*. Nanjing, China. IEEE, 2023: 89-94.
- [12] JOCHER G, CHAURASIA A, QIU J. ULTRALYTICS YOLOv8[EB/OL]. (2024-1-10)[2024-1-11]. <https://github.com/ultralytics/ultralytics>.
- [13] JOCHER G. ULTRALYTICS. YOLOv5[EB/OL]. (2022-11-22) [2023-08-31]. <https://github.com/ultralytics/yolov5>.
- [14] WANG Q L, WU B G, ZHU P F, *et al.* ECA-net: efficient channel attention for deep convolutional neural networks[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA. IEEE, 2020: 11531-11539.
- [15] ZHU X Z, CHENG D Z, ZHANG Z, *et al.* An empirical study of spatial attention mechanisms in deep networks[C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South). IEEE, 2019: 6687-6696.
- [16] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada. IEEE, 2023: 7464-7475.
- [17] ZHENG Z H, WANG P, LIU W, *et al.* Distance-IoU loss: faster and better learning for bounding box regression[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34 (7) : 12993-13000.
- [18] LIU C, WANG K G, LI Q, *et al.* Powerful-IoU: more straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism[J]. *Neural Networks*, 2024, 170: 276-284.
- [19] 禹文奇, 程臻, 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. (in Chinese)
- [20] CHENG G, ZHOU P C, HAN J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(12): 7405-7415.
- [21] HE K M, ZHANG X Y, REN S Q, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37 (9): 1904-1916.
- [22] CHEN L C, PAPANDREOU G, KOKKINOS I, *et al.* DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.
- [23] LIU S, HUANG D, WANG Y. Receptive field block net for accurate and fast object detection[C]. *Proceedings of the European Conference on Computer Vision (ECCV)*. September 8-14, 2018. Munich, Germany. Springer-Verlag, 2018: 404-419.
- [24] LI C Y, LI L L, GENG Y F, *et al.* YOLOv6 v3.0: a full-scale reloading[EB/OL]. 2023: *arXiv*: 2301.05586. <http://arxiv.org/abs/2301.05586>
- [25] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional block attention module[C]. *Proceedings of the European Conference on Computer Vision (ECCV)*. September 8-14, 2018. Munich, Germany. Springer-Verlag, 2018: 3-19.
- [26] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA. IEEE, 2021: 13708-13717.
- [27] HUANG H L, CHEN Z, ZOU Y, *et al.* Channel Prior Convolutional Attention for Medical Image Segmentation [EB/OL]. 2023: *arXiv*, 2306.05196. <https://arxiv.org/abs/2306.05196>
- [28] AZAD R, NIGGEMEIER L, HÜTTEMANN M, *et al.* Beyond self-attention: deformable large kernel attention for medical image segmentation [C]. 2024 *IEEE/CVF Winter Conference on Ap-*

- plications of Computer Vision (WACV)*. Waikoloa, HI, USA. IEEE, 2024: 1276-1286.
- [29] OUYANG D L, HE S, ZHANG G Z, *et al*. Efficient multi-scale attention module with cross-spatial learning[C]. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes Island, Greece. IEEE, 2023: 1-5.
- [30] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. *Expert Systems with Applications*, 2024, 236: 121352.
- [31] ZHANG Q L, YANG Y B. SA-net: Shuffle attention for deep convolutional neural networks[C]. *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Toronto, ON, Canada. IEEE, 2021: 2235-2239.
- [32] YANG L, ZHANG R Y, LI L, *et al*. SimAM: a simple, parameter-free attention module for convolutional neural networks [C]. *Proceedings of the 38th International Conference on Machine Learning*. July 18-24, 2021 Vienna, Austria. PMLR, 2021: 11863-11874.
- [33] MISRA D, NALAMADA T, ARASANIPALAI A U, *et al*. Rotate to attend: convolutional triplet attention module[C]. *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA. IEEE, 2021: 3138-3147.
- [34] REN S 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*, 2017, 39(6): 1137-1149.
- [35] OUYANG H. DEYOv3: DETR with YOLO for real-time object detection[EB/OL]. 2023: arXiv: 2309.11851. <http://arxiv.org/abs/2309.11851>
- [36] WANG C C, HE W, NIE Y, *et al*. Gold-YOLO: efficient object detector via gather-and-distribute mechanism [EB/OL]. 2023: arXiv: 2309.11331. <http://arxiv.org/abs/2309.11331>

作者简介:



林珊玲(1991—),女,福建泉州人,博士,讲师,硕士研究生导师,2014年、2020年于福州大学分别获得学士、博士学位,主要从事显示驱动、图像处理等方面研究。E-mail: sllin@fzu.edu.cn

通讯作者:



林坚普(1989—),男,福建泉州人,博士,讲师,硕士研究生导师,2012年、2019年于福州大学分别获得学士、博士学位,主要从事图像处理、深度学习、液晶透镜器件、裸眼3D显示技术等方面的研究。E-mail: ljp@fzu.edu.cn