

doi: 10.16104/j.issn.1673-1891.2023.02.011

基于改进 YOLOv5 的遥感图像飞机目标检测

孙廉尧^{1,2}, 侯秀丽¹, 罗青青¹

(1. 安徽商贸职业技术学院信智学院, 安徽 芜湖 241000; 2. 芜湖市物联网智慧交通工程技术研究中心, 安徽 芜湖 241000)

摘要: 针对遥感图像中飞机检测尺寸大小不一、背景复杂导致的难以识别问题, 提出一种基于 YOLOv5 网络模型的改进方法。首先, 在 YOLOv5 网络模型中融入 Swin-Transformer 模块, 使网络全局建模并使全维度信息交互, 以提升网络的特征提取能力; 然后, 对损失函数进行优化, 引入 SIOU 损失函数以考虑真实框和预测框之间的向量角度问题。对比实验结果表明: 改进后的网络模型召回率为 91.2%, 比改进前提升 0.6 个百分点; 改进后平均检测精度 $mAP_{0.5}$ 为 95.7%, 比改进前提升 0.2 个百分点。改进后的 YOLOv5 网络模型能在一定程度上提升遥感图像中飞机目标检测性能。

关键词: 遥感图像; YOLOv5; Swin-Transformer; SIOU

中图分类号: TP751 **文献标志码:** A **文章编号:** 1673-1891(2023)02-0066-06

Aircraft Target Detection Based on Improved YOLOv5 for Remote Sensing Images

SUN Xieyao^{1,2}, HOU Xiuli¹, LUO Qingqing¹

(1. Xinzhi College, Anhui Vocational & Technical College of Business & Trade, Wuhu, Anhui 241000, China;
2. Wuhu Internet of Things Smart Transportation Engineering Technology Research Center, Wuhu, Anhui 241000, China)

Abstract: Aiming at the difficult recognition problem caused by the different sizes of aircraft detection and complex backgrounds in remote sensing images, an improving method based on the YOLOv5 network model is proposed. Firstly, the Swin-Transformer module is incorporated into the YOLOv5 network model to model the network globally and make the full-dimensional information interactive to improve the feature extraction capability of the network; then, the loss function is optimized and the SIOU loss function is introduced to consider the vector angle problem between the real frame and the predicted frame. The comparison experiments' results show that the recall rate of the improved network model is 91.2%, which is 0.6 percentage points higher than the pre-improvement rate, and the average detection precision $mAP_{0.5}$ is 95.7%, which is 0.2 percentage points higher than the pre-improvement rate. The improved YOLOv5 network model can improve the performance of aircraft target detection in remote sensing images to a certain extent.

Keywords: remote sensing images; YOLOv5; Swin-Transformer; SIOU

0 引言

随着人工智能技术的快速发展, 目标检测技术在各领域中有广泛的应用。由于遥感图像中蕴含的信息丰富且复杂, 遥感图像目标检测可以广泛应用在国防军事、地质灾害、海洋监测以及智慧城市中。通过提取图像中的信息达到某种应用效果。

近年来, 很多研究者提出了多种遥感目标检测方法。李惠惠等^[1]提出了一种改进型的 YOLOv5 网络模型检测算法, 算法先对锚框进行了优化, 其次增加了卷积注意力机制以提升特征提取的能力, 最后通过增加浅层特征图来提高小目标的识别精度; 曾伦杰等^[2]提出了一种二阶段锚框和类均衡损失的遥感图像目标检测算法, 通过 k 均值聚类(k-means)

收稿日期: 2023-02-06

基金项目: 安徽省高校自然科学研究重点项目(2022AH052740); 安徽省职教 2022 教育科研规划课题(Azcyj2022128); 安徽商贸职业技术学院自然科学重点项目(2022KZZ05)。

作者简介: 孙廉尧(1993—), 男, 河南鹤壁人, 讲师, 硕士, 主要研究方向: 计算机技术。

和k-means生成的标签作为第2阶段的中心,生成预设锚框,以提升检测精度,并构建类别平衡损失(class equalization loss, CEQL)提高对少数类别标签的关注程度;张云佐等^[3]提出的多尺度融合与注意力机制的遥感图像目标检测,改进了金字塔池化模块以提高不同图像的感受野,通过注意力模块来学习特征、提升算法特征提取能力,并引入加权双向特征金字塔网络与主干网络结合来促进特征融合;Kala等^[4]提出了一种用FastCNN-NB进行遥感图像提取的物体检测新框架;Hou等^[5]提出了一种用于在高分辨率遥感图像中进行任意导向的目标检测R-YOLO。但是以上方式有的在检测精度或检测速度上还有所缺陷。本文将在YOLOv5算法基础上进行优化改进,在自构建的含有飞机目标的遥感图像

数据集中进行训练,通过添加Swin-Transformer(shifted windows transformer)结构提升特征提取能力,通过对原始模型损失函数改进以提升遥感图像目标检测的精度。

1 YOLOv5 概述

YOLO经历了多种版本的迭代,作为单阶段算法,相比较其他算法具有检测速度快、可以直接完成目标定位的特点。YOLOv5针对不同大小的网络出了不同的版本,分为YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l和YOLOv5x,区别在于采用的深度与宽度不同。YOLOv5网络中模型包含了Input、Backbone和Neck等3个部分^[6]。YOLOv5网络模型结构如图1所示。

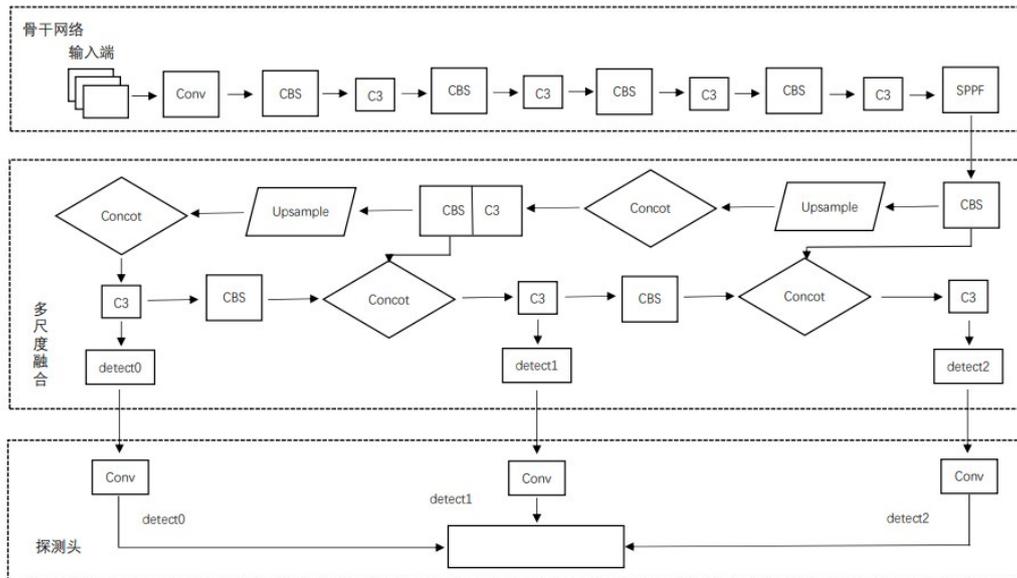


图1 YOLOv5网络结构图

上述包含卷积层Conv、拼接层Concat、上采样层Upsample、跨阶段局部网络SPPF、CBS(由Conv层+Bn层+激活函数三者组成)、模块C3层和探测模块detect。

2 YOLOv5 算法改进

2.1 引入Swin-Transformer结构

由于YOLOv5主干特征提取网络为CNN,并且CNN网络具有平移不变性和局部性等问题,缺乏全局建模的能力。2021年Liu等^[7]提出了一种在计算机视觉领域通用的骨干网络Swin-Transformer,由于数据集中飞机属于小目标,YOLOv5原始网络对小目标检测能力有限,本文引入Swin-Transformer结

构,将Swin-Transformer融合到YOLOv5网络结构中,改进Backbone中的C3模块,以此来提高全局检测能力,提高目标检测效果。

Swin-Transformer结构采用移位窗口的方式,通过将自注意力的计算限制在不重叠的局部窗口中,同时也允许跨窗口连接,这样不仅有了更高的效率,同时通过移动操作,能够让相邻的2个窗口之间有了交互,上下层窗口也有了交叉连接,从而达到一种全局建模的能力。Swin-Transformer包含了2个模块,一个是W-MSA(Window-Multi Self Attention)窗口注意力机制模块,另一个是SW-MSA(Shifted Window-Multi Self Attention)移位窗口自注意力机制模块,如图2所示。其中,MLP为人工神经

网络, LN 为 LayerNorm, W-MSA 为窗口注意力机制模块, SW-MSA 为移位窗口自注意力机制模块。

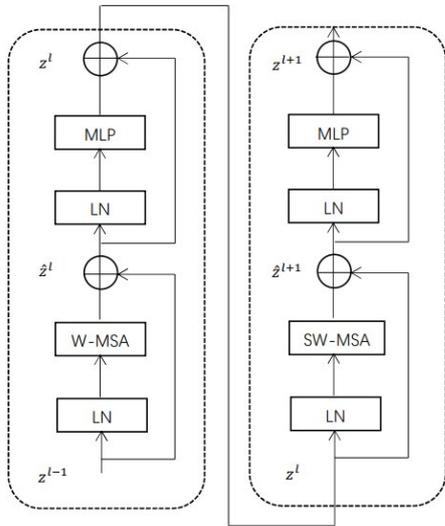


图 2 Swin-Transformer 架构流程^[7]

2.1.1 W-MSA 窗口注意力机制

在原始 MSA 模块中,特征图内每个像素需要与所有像素进行关联计算得到注意力值 Attention(用 A_i 表示),MSA 会对输入特征图的每一个像素进行求解 Q, K, V , 每一个像素求得的 Q 会和特征图上每一个像素的 K 进行匹配。然后再进行一系列的操作。Attention 数值计算公式如式(1)所示^[7]。

$$A_i = (Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \quad (1)$$

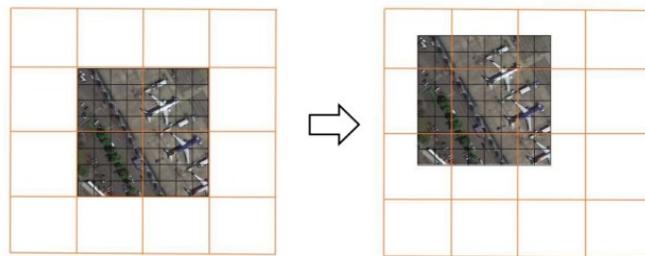


图 3 移位窗口

通过图 3 的移位过程看可以发现,窗口和窗口间有信息交互,但是出现了新的问题——窗口变多且计算量又增加了,所以提出了另一种更加高效的计算方法。对划分区域进行平移,如图 4 所示:将图 4(a)中 A、B、C 区域移动到最底部,得到图 4(b);再将图 4(b)中 D、G、A 区域移动到最右边,得到 4(c);最后重新划分为 4 个区域,得到图 4(d)。这样就确保了和原来一样是 4 个窗口,计算量是一样的。

式中: Q 和 K 是计算 A_t 权重的特征向量; V 为输入特征的向量; T 为转置符号; d 为 Q/K 的维度; B 为相对位置偏差。

假设 Q, K, V 的向量长度与特征图的深度 C 一致, h 和 w 分别为特征图的高和宽, 令 $X = Q^{hw \times C}, Y = A_i^{hw \times C}, Z = W_q^{C \times C}$ 那么所有像素 Q 产生的过程如式(2)所示。

$$X = Y \times Z \quad (2)$$

式中: Y 为将所有像素拼接在一起得到的矩阵; Z 为生成 q 的变换矩阵; X 为所有像素通过 Z 得到的 q 拼接后的矩阵。根据矩阵运算的计算量公式分别算出 Q, V 的计算量, 所以最终可得原始 MSA 模块计算总量如式(3)所示。

$$4hwC^2 + 2(hw)^2C \quad (3)$$

而 W-MSA 窗口注意力机制首先会对特征图进行分割处理, 分割为一个个的小窗口, 然后在每一个窗口内部执行 MSA, 以此来减少计算量。故使用 W-MSA 的计算量为 $4hwC^2 + 2M^2hwC$ (M 是特征图的尺寸)。

2.1.2 SW-MSA 移位窗口自注意力机制

由上述分析可知, 虽然 W-MSA 窗口注意力机制解决了计算复杂度的问题, 但是窗口和窗口之间还没有交互, 这样就达不到全局建模, 所以提出了移位窗口的方式。每做一次 W-MSA 窗口注意力, 就做一次 SW-MSA 移位窗口自注意力, 通过窗口移位达到窗口和窗口之间的交互。SW-MSA 移位窗口示意图如图 3 所示。

2.2 损失函数改进

目前损失函数有很多, GIoU (Generalized-IoU) 考虑了重叠面积, 基于交并比 (intersection over union, IoU) 解决边界框不相交时 loss 等于 0 的问题; DIoU (Distance-IoU) 考虑了重叠面积和中心点距离, 基于 IoU 解决 GIoU 收敛慢的问题; CIoU (Complete-IoU) 考虑了重叠面积、中心点距离、纵横比, 基于 DIoU 提升回归精确度。YOLOv5 采用 CIoU

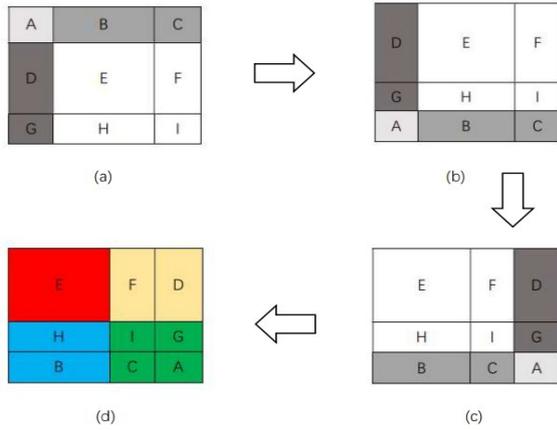


图 4 移位窗口过程示意图

损失函数,不管是 GIoU、DIoU 还是 CIoU 都没有考虑到真实框与预测框之间的方向,导致某种程度上收敛速度还是较慢。对此,2022 年 Gevorgyan^[8]提出了一种损失函数 SIOU (Smoothed-IoU),其引入真实框和预测框之间的向量角度重新定义相关损失函数,具体包含了角度损失、距离损失、形状损失和 IOU 损失。其中,角度损失 Λ 定义如式(4)所示^[8]。

$$\begin{aligned}\Lambda &= 1 - 2 \times \sin^2\left(\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right) \\ &= \cos\left(2 \times \left(\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right)\right)\end{aligned}\quad (4)$$

式中: c_h 为两目标框中心点的纵向距离; σ 为两中心点对角线距离。

距离损失 Δ 定义如式(5)所示^[8]。

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma p_t}) = 2 - e^{-\gamma p_x} - e^{-\gamma p_y}\quad (5)$$

式中: t 为一个参数位置; x, y 分别为最小外接矩形左上坐标和右下坐标; p_x 为真实框和预测框中心点的宽度差与最小外接矩形宽之比的平方; p_y 为真实框和预测框中心点的高度差与最小外接矩形高之比的平方。其中, $\gamma = 2 - \Lambda$ 。

形状损失 Ω 定义如式(6)所示^[8]。

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^\theta = (1 - e^{-w_x})^\theta + (1 - e^{-w_h})^\theta\quad (6)$$

式中: w_w 为预测框宽减去真实框宽的绝对值与预测框宽和真实框宽中的最大值的比值果; w_h 为预测框高减去真实框高的绝对值与预测框高和真实框高中的最大值的比值。其中,

$$w_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}, w_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}\quad (7)$$

式中: (h, w) 和 (h^{gt}, w^{gt}) 分别为预测框和真实框的高和宽; θ 值表示对形状损失的关注程度。如果 θ 的值

设置为 1,它将立即优化一个形状,从而损害形状的自由移动。

最终,SIOU 损失函数(用 S_i 表示)定义如式(8)所示^[8]。

$$S_i = 1 - I_o + \frac{\Delta + \Omega}{2}\quad (8)$$

式中: I_o 为齐交比。

在遥感图像飞机目标检测中,采用 SIOU 损失函数替换原始算法 CIoU 损失函数来加快收敛速度。

3 实验与分析

3.1 实验数据

由于硬件设备限制,采用自构建的小样本量数据集,并默认使用 YOLOv5 自带的增强方式进行测试。为了获取遥感图像的多样性,自构建含有飞机的遥感图像数据集从 DOTA、WHU-RS19、RSD46-WHU 等数据集中筛选而来,统一标准处理后,共计 2 602 张遥感图片。训练集:测试集:验证集 = 6:2:2。

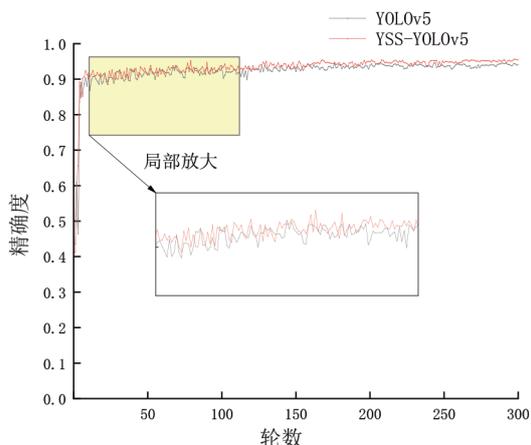
3.2 实验结果分析

实验采用 Pytorch 深度学习框架,训练轮数 epochs 设置为 300。改进后算法命名为 YSS-YOLOv5。由表 1 可知,在 YOLOv5 网络模型基础上添加 Swin-Transformer 结构、改进损失函数为 SIOU 后,在保持原有检测精度的同时,召回率 R 有所提升,从 90.6% 提升为 91.2%,上涨 0.6 个百分点;平均检测精度 $mAP_{0.5}$ 由 95.5% 提升到 95.7%,上涨 0.2 个百分点。同时参数量 Parameters 由原来的 7 012 822 降低为 4 821 600,能有效提高推理速度。对比实验结果表明,改进后算法模型召回率和平均检测精度均有所提升,有助于识别遥感图像中的飞机目标。

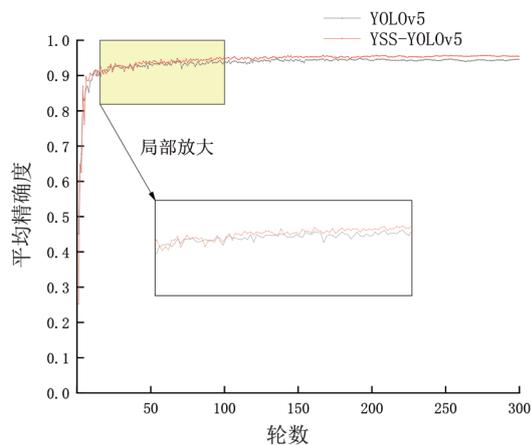
表 1 实验结果

算法	Parameters	P	R	$mAP_{0.5}$
YOLOv5	7 012 822	0.953	0.906	0.955
YSS-YOLOv5	4 821 600	0.953	0.912	0.957

从算法模型其他对比结果看,改进后算法检测性能有了一定的提升。算法改进前后精确度 P 对比曲线和平均精确度 $mAP_{0.5}$ 对比曲线如图 5 所示,召回率 R 对比曲线如图 6 所示。虽然最终结果显示改进前后算法检测精度相同,都为 0.953,但从精度曲线来看,大概在训练 100 轮后有所提升,实际检测效果也有所提高(图 5)。



(a) 检测精度 P 对比曲线



(b) 平均检测精度 $mAP_{0.5}$ 对比曲线

图 5 算法改进前后精确度 P 与平均检测精度 $mAP_{0.5}$ 对比曲线

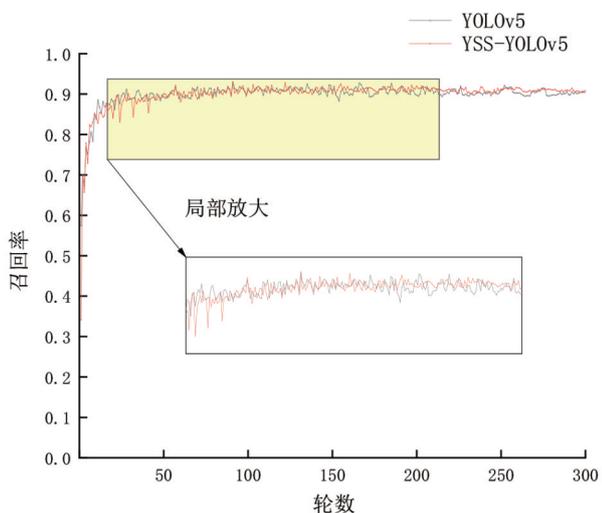


图 6 算法改进前后召回率 R 对比曲线

为了更直观展示检测效果,算法改进前后检测

效果对比如图 7 所示。由图 7 可知,此遥感图像中飞机尺寸大小不一,图像背景目标类别复杂。左侧目标飞机在原算法中置信度为 0.93,在改进后算法中置信度为 0.94,提升 1%;右侧目标飞机在原算法中置信度为 0.90,在改进后算法中置信度为 0.92,提升 2%;上方小飞机在原算法中置信度为 0.30,在改进后算法中置信度为 0.49,提升 19%。由此可知,改进后算法模型能在一定程度上提升检测效果。

4 结语

本文对遥感图像飞机目标检测中背景复杂、数量多且密集等问题,在 YOLOv5 网络模型基础上进行了优化。优化后网络模型通过引入 Swin-Transformer 结构来提升全局建模能力和全局检测能力;通过将 CIOU 损失函数替换为 SIOU 函数,提高算



图 7 算法改进前后检测效果图

法在背景复杂、数量多且密的情况下的定位能力。平均检测精度提升了 0.2%。表明优化后网络模型对比实验结果表明,优化后网络模型算法更适应遥感图像飞机目标检测的特点,召回率提升了 0.6%,算法有一定的提升效果。

参考文献:

- [1] 李惠惠,范军芳,陈启丽.改进YOLOv5的遥感图像目标检测[J].弹箭与制导学报,2022,42(4):17-23.
- [2] 曾伦杰,储珺,陈昭俊.二阶段锚框和类均衡损失的遥感图像目标检测[J].图学学报,2023,44(2):249-259.
- [3] 张云佐,郭威,蔡昭权,等.联合多尺度与注意力机制的遥感图像目标检测[J].浙江大学学报(工学版),2022,56(11):2215-2223.
- [4] KALA K, PADMASINI N, KAPALI B, et al. A new framework for object detection using fastcnn- Naïve Bayes classifier for remote sensing image extraction[J]. Earth Science Informatics, 2022, 15(3):1779-1787.
- [5] HOU Y J, SHI G, ZHAO Y X, et al. R-Yolo: a Yolo-based method for arbitrary-oriented target detection in high-resolution remote sensing images[J]. Sensors, 2022, 22(15):5716-5716.
- [6] 张学志,赵红东,刘伟娜,等.基于改进YOLOv5的红外车辆检测方法[J/OL].红外与激光工程:1-11[2023-06-08].<http://kns.cnki.net/kcms/detail/12.1261.TN.20230605.1630.008.html>.
- [7] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision, October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2021: 9992-10002.
- [8] GEVORGYAN Z. SIOU loss: more powerful learning for bounding box regression [EB/OL]. (2022-05-25) [2022-10-02]. <https://arxiv.org/abs/2205.12740>.