

中图法分类号: TP751 文献标识码: A 文章编号: 1006-8961(2025)03-0683-13

论文引用格式: Xia B, Xue W T, Zhou X Y and Huang H. 2025. Remote sensing object detection via local parameter-free attention and combined loss. Journal of Image and Graphics, 30(3):0683-0695(夏波, 薛卫涛, 周新尧, 黄鸿. 2025. 局部无参注意力和联合损失的遥感目标检测. 中国图象图形学报, 30(3):0683-0695)[DOI:10.11834/jig.240316]

局部无参注意力和联合损失的遥感目标检测

夏波¹, 薛卫涛², 周新尧¹, 黄鸿^{1*}

1. 重庆大学光电技术与系统教育部重点实验室, 重庆 400044; 2. 北京空间机电研究所产品检测中心, 北京 100094

摘要: 目的 遥感目标检测技术在遥感测绘、智慧城市、乡村振兴和国防军事等领域具有广泛应用, 但遥感影像存在背景信息复杂、待检目标小且多等特点, 导致目标特征随着网络加深淹没在背景信息中, 不利于后续检测任务。针对此问题, 基于YOLOv5s(you only look once)网络, 提出一种局部无参注意力和联合损失的遥感目标检测方法。**方法** 首先提出一种局部无参注意力机制, 能够根据当前特征提高局部区域内的目标关注度, 而不引入训练参数, 以此构建无参注意力残差块, 并将其嵌入到骨干网络的不同阶段, 实现多尺度目标特征增强和背景信息抑制。在此基础上, 利用最优传输距离度量边界框的相似性, 构造W-CIoU(Wassertein-complete intersection of union)联合度量方法和相应损失函数, 以缓解锚框与真实框差异过大导致的标签误分配问题, 降低小目标漏检率。**结果** 在RSOD(remote sensing object detection)和RSSOD(remote sensing super-resolution object detection)数据集上的实验结果表明, 该方法在保证模型参数量和复杂度基本不变的前提下, 其平均精度均值(mean average precision, mAP)分别达到98.2%及87.4%, 相较两种主要对比方法DAB-DETR(dynamic anchor boxes for detection Transformer)和YOLOv8s, 在RSOD数据集上分别提升0.8%和0.3%, 在RSSOD数据集上分别提升2.9%和2.6%。**结论** 本文方法能够抑制背景信息并增强目标特征, 提高检测性能, 相较对比方法能更好地满足复杂场景下小目标检测需要。

关键词: 遥感影像; 目标检测; 局部无参注意力(LPFA); 最优传输距离; 联合损失函数

Remote sensing object detection via local parameter-free attention and combined loss

Xia Bo¹, Xue Weitao², Zhou Xinyao¹, Huang Hong^{1*}

1. Key Laboratory of Optoelectronic Technology and Systems of the Ministry of Education of China, Chongqing University, Chongqing 400044, China; 2. Product Testing Center, Beijing Institute of Space Machinery and Electronics, Beijing 100094, China

Abstract: **Objective** Remote sensing object detection technology has been widely used in various fields, including remote sensing mapping, smart cities, rural revitalization, resource exploration, national security, and military affairs. Particularly with the completion and improvement of notable Chinese projects in high-resolution Earth observation systems, strong data support and development opportunities are available for remote sensing object detection. Traditional detection algorithms have relatively weak generalization capabilities and are easily affected by noise, data distribution, and other factors. Consequently, challenges are encountered in overcoming the diverse scales and directions of remote sensing objects, as

收稿日期: 2024-06-13; 修回日期: 2024-08-13; 预印本日期: 2024-08-20

* 通信作者: 黄鸿 hhuang@cqu.edu.cn

基金项目: 国家自然科学基金项目(42071302); 先进光学遥感技术北京市重点实验室开放基金(AORS202315)

Supported by: National Natural Science Foundation of China(42071302); Beijing Key Laboratory of Advanced Optical Remote Sensing Technology Fund (AORS202315)

well as complex data distribution. In recent years, the strong representation learning capability and generalization capacity of deep learning have led to its widespread use in the field of object detection. Current detection algorithms based on deep learning can be simply divided into three categories: region-, pixel-, and query-based methods. Region-based methods typically offer high detection accuracy but require high computation and exhibit low efficiency. Pixel-based methods are usually single-stage detectors with low computation and high efficiency, but they often struggle with lower detection accuracy, particularly in small object detection tasks. Query-based methods require a large amount of data and have low efficiency. However, the object features extracted using the aforementioned methods become overshadowed by the background information as the network deepens due to the complexity of background information and the many small objects in remote sensing images. This condition is not conducive to subsequent detection tasks, leading to limited final detection performance. In response to the issue of insufficient small object detection performance under complex backgrounds, this work proposes a local parameter-free attention YOLO (you only look once) network (LPFA-YOLO) based on YOLOv5s.

Method First, a local parameter-free attention (LPFA) mechanism is proposed. This mechanism can improve the attention for objects within a local region based on current features, without introducing any trainable parameters, thus constructing a bottleneck with parameter-free attention (BPFA). The attention mechanism in this block only assigns corresponding weights to the residuals and does not have a notable impact on the main weights. This approach helps avoid the problem of gradient disappearance and accelerates convergence by utilizing pre-trained weights. The C3 with attention module (C3A) constructed based on BPFA is then embedded into different stages of the backbone network to cater to objects of different scales. The shallow stage enhances the features of small objects, whereas the deep stage improves the features of medium and large objects, thereby realizing multiscale object feature enhancement and background information suppression, which addresses the issue of redundant background information. On this basis, the Wasserstein distance is utilized to measure the similarity of bounding boxes. A combined measurement method called Wasserstein-complete intersection of union (W-CIoU) and the related loss function are developed. This method can alleviate the sensitivity of small objects to position deviation and enable the separation of objects of different scales. Consequently, this method alleviates the issue of label misallocation due to the substantial difference between the anchor box and ground truth, reducing the missed detection rate of small objects.

Result Experiments are conducted on two datasets and compared with seven advanced algorithms, including EfficientNet, YOLOv4, detection Transformer (DETR), Swin Transformer, detecting objects with recursive feature pyramid and switchable atrous convolution (DetecORS), dynamic anchor boxes for DETR (DAB-DETR), and YOLOv8s. On the remote sensing object detection (RSOD) dataset, the mean average precision (mAP) reaches 98.2%, which is 0.9% and 0.3% higher than the baseline and YOLOv8s, respectively. The average precision for small objects (AP_s) reaches 42.7%, showing a 2.4% improvement compared to the baseline. The average precision for medium (AP_m) and large (AP_l) objects reach 67.7% and 82.1%, respectively, both demonstrating improvements compared to the baseline. On the remote sensing super-resolution object detection (RSSOD) dataset, the mAP achieves 87.4%, which is 2.7% and 2.6% higher than the baseline and YOLOv8s, respectively. The APs increases to 38.0%, revealing a 2.7% rise compared to the baseline. The AP_m and AP_l achieve 52.7% and 46.4%, respectively, both showing improvements compared to the baseline. Simultaneously, the model has the fewest number of parameters and computation compared to contrast algorithms. Experiments are conducted using two external remote sensing images obtained from Google Earth to assess the generalization of the algorithm. The results indicate that the algorithm detects all objects of interest, demonstrating good generalization capabilities.

Conclusion In this work, a YOLO network model based on local parameter-free attention is introduced. The experimental results demonstrate that this model can effectively address the requirements of small object detection in complex scenes compared to several existing methods.

Key words: remote sensing images; object detection; local parameter-free attention (LPFA); Wasserstein distance; combined loss function

0 引言

遥感目标检测技术在遥感测绘、智慧城市、乡村振兴、资源勘探和国防军事等领域得到广泛应用,尤其是我国高分辨率对地观测系统重大专项的建成与完善,为其提供了有力的数据支撑和发展机遇。然而,相比自然图像,遥感影像中的目标通常缺乏细节信息,特别是对于一些小目标,背景信息会淹没其有用特征(石争浩等,2023)。因此,遥感目标检测亟需解决以下难点:1)复杂背景信息冗余,干扰检测任务;2)小目标数量多且排列密集,漏检率较高(袁翔等,2023)。

传统算法如模板匹配方法(Cheng和Han,2016)、基于超像素的方法(Blaschke,2010)以及尺度不变特征匹配(Sedaghat等,2011)、方向梯度直方图(Dalal和Triggs,2005)等需要人工设计算子来提取特征,泛化能力较弱,易受到噪声、数据分布等因素的影响,难以克服遥感影像中的背景噪声较多、目标尺度与方向多样和数据分布复杂等问题,检测性能无法保证。

由于深度学习的表征学习能力和泛化能力强,其被广泛应用于目标检测领域。目前深度学习检测算法可以简单分为3类,即基于区域的方法、基于像素的方法和基于查询的方法。基于区域的方法(Ren等,2017;He等,2015;Lin等,2017),如Ren等人(2017)提出的Faster R-CNN(faster regions with convolutional neural network),检测精度较高,但计算效率和速度较低。基于像素的方法(Redmon等,2016;Liu等,2017)通常是单阶段检测器,如Redmon等人(2016)提出的YOLO(you only look once),其计算效率和速度高,但检测精度低,在小目标检测任务上性能不足。Carion等人(2020)提出的DETR(detection Transformer)在目标检测中引入Transformer,启发了后续基于查询的方法(Liu等,2022),该类方法使用直接集预测,舍弃了锚框设计、非极大抑制等步骤,但数据量需求较大,计算效率低。

对于遥感影像,其背景信息复杂,待检目标小、数量多且排列密集,导致自然图像处理方法的检测性能受限,一些学者精心设计了高质量的大规模基准模型。Hong等人(2024)创建了一个通用遥感基准模型,使用3D生成式预训练Transformer对光谱遥

感影像进行处理,在单/多标签场景分类、语义分割和变化检测任务中取得了很大的成功。Li等人(2023a)提出了一种新的高光谱异常检测基准网络,通过结合低秩表示模型和深度学习方法来增强网络的表示性能。针对目标检测任务,赵文清等人(2021)将长短期记忆网络(long short-term memory, LSTM)引入到单次多框检测器(single shot multibox detector, SSD)中,以改善检测性能。Zhang等人(2023)将自适应多粒度机制和双路径编码应用到金字塔视觉Transformer中,解决训练样本效率低和知识学习困难的问题。Yang等人(2021b)提出的R³Det(refined rotation RetinaNet)通过旋转框检测目标,提高了模型检测性能。Li等人(2023b)探索了大卷积核在遥感场景中独特先验知识的作用,通过挖掘不同目标与背景信息的关联程度,减少误检情况。

尽管现有方法在处理遥感影像方面取得了重大进展,遥感影像中的复杂背景信息与大量小目标仍然阻碍了它们的进一步应用。此外,一些提高检测性能的方法通常会带来较多的额外参数和计算复杂度。

针对复杂背景下小目标检测性能不足和计算复杂度难以兼顾的问题,本文基于YOLOv5s网络提出一种局部无参注意力YOLO(local parameter-free attention YOLO, LPFA-YOLO)网络,在保证参数量与计算复杂度基本不变的前提下,降低小目标漏检率的同时减少误检情况,改善检测性能。

1 LPFA-YOLO网络

本文提出的LPFA-YOLO遥感目标检测方法如图1所示。首先提出一种局部无参注意力(local parameter-free attention, LPFA)机制,能够根据当前特征提高局部区域内的目标关注度,而不引入训练参数,以此构建无参注意力残差块(bottleneck with parameter-free attention, BPFA);然后将以BPFA为基础构造的C3A(C3 with attention)模块嵌入到骨干网络的不同阶段,以实现多尺度目标特征增强和背景信息抑制的目的,解决背景信息冗余问题;在此基础上,引入最优化传输距离,设计最优传输—完全交并比(Wassertein-complete intersection of union, W-CIoU)的联合度量方法及相应损失函数,以缓解锚框与真实框差异过大导致的标签误分配问题,有效

提高小目标检测性能。

1.1 LPFA 局部无参注意力

遥感影像中的复杂背景在检测任务中容易淹没小目标的特征信息,导致出现误检与漏检,而注意力机制可以从大量背景中筛选出目标区域。Yang 等人(2021a)提出的 SimAM 机制 (simple attention mechanism) 是一种通过能量值施加注意力的机制,但对于输入特征 $F_i \in \mathbf{R}^{C \times H \times W}$,该方法只计算一次通道内所有神经元的均值和方差,导致对特征信息的利用不足。此外,对于遥感影像,背景占据绝大多数区域,由此计算的均值和方差通常表现为背景信息,不能有效突出小目标特征信息。因此,本文提出一种 LPFA 机制,通过计算局部区域内神经元的均值与方差,得到每个神经元 x_i 的能量值 e 与相应权重,

具体过程如图 2 所示,相应的计算为

$$e = \frac{4(\sigma^2 + \lambda)}{(t - \mu)^2 + 2\sigma^2 + 2\lambda} \quad (1)$$

式中, t 表示当前神经元,正则化项 $\lambda=10^{-4}$ 。

$$\mu = \frac{1}{K \times K} \sum_{i=1}^{K \times K} x_i \quad (2)$$

$$\sigma^2 = \frac{1}{K \times K - 1} \sum_{i=1}^{K \times K} (x_i - \mu)^2$$

式中, K 表示局部区域大小,考虑到感受野与计算量的平衡,令 $K=5$ 。能量值 e 越小,对周围神经元的抑制作用越强,则赋予更大权重,故权重设为 $1/e$ 。每个局部区域内单独计算完毕后,对注意力矩阵在空间上进行拼接,图 2 输出特征中的深色块即表示目标区域。若输入特征尺寸无法被整除,对其进行填充,计算后取原始大小,以此解决维度不匹配问题。

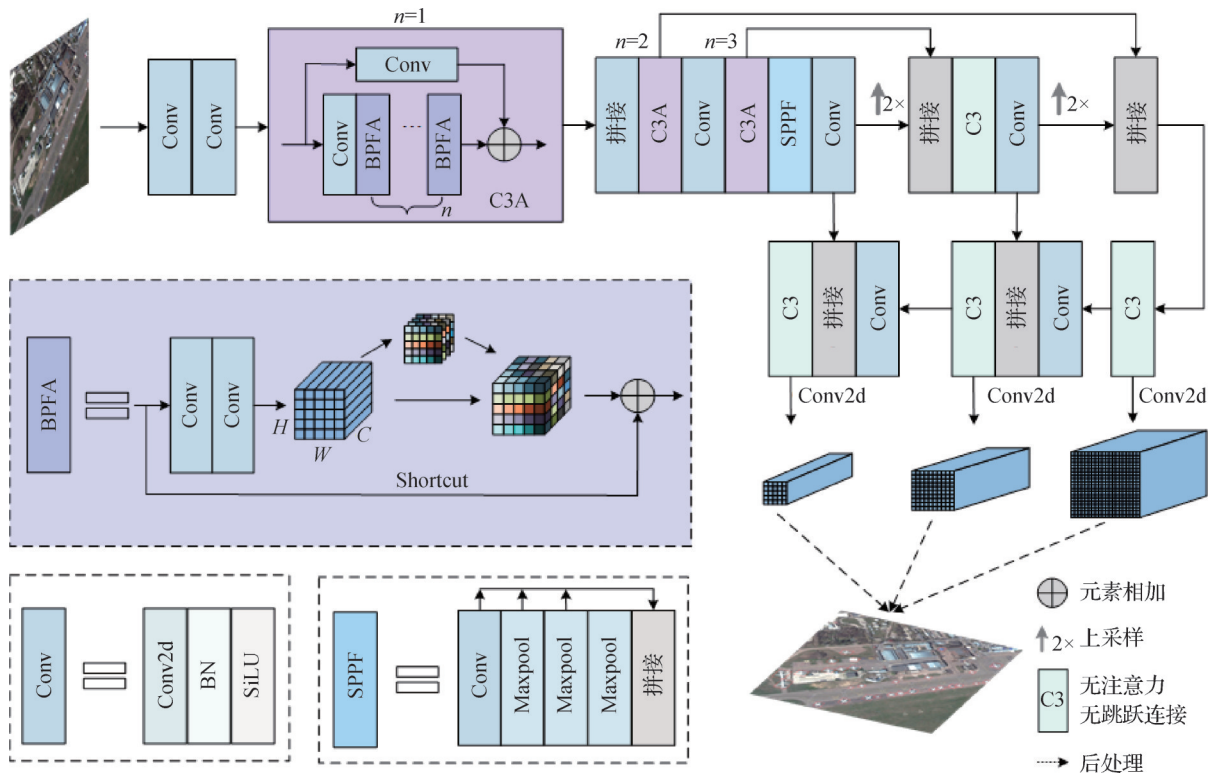


图 1 LPFA-YOLO 网络整体框架

Fig. 1 Overall framework of LPFA-YOLO network

为成功添加注意力,将 LPFA 模块嵌入到残差块中,得到 BPFA 残差块,对于输入特征 $F_i \in \mathbf{R}^{C \times H \times W}$,输出特征 F_o 可以表示为

$$F_o = F_i + Conv(Conv(F_i)) \odot \delta\left(\frac{1}{E}\right) \quad (3)$$

式中, $Conv$ 为卷积操作, \odot 为 Hadamard 积, δ 是 sigmoid 激活函数, E 表示能量值矩阵。括号内卷积对

输入降维以减少计算量,括号外卷积进行特征提取并升维,之后与注意力权重点积运算;然后与输入相加得到最终输出。由于只对残差赋予相应权重,而对主要权重不产生较大影响,既缓解了梯度消失的问题,又能够利用预训练权重,加快收敛。

将 BPFA 引入到 YOLOv5s 网络的 C3 模块中,得到 C3A 模块,嵌入到骨干网络的不同阶段,以对应不

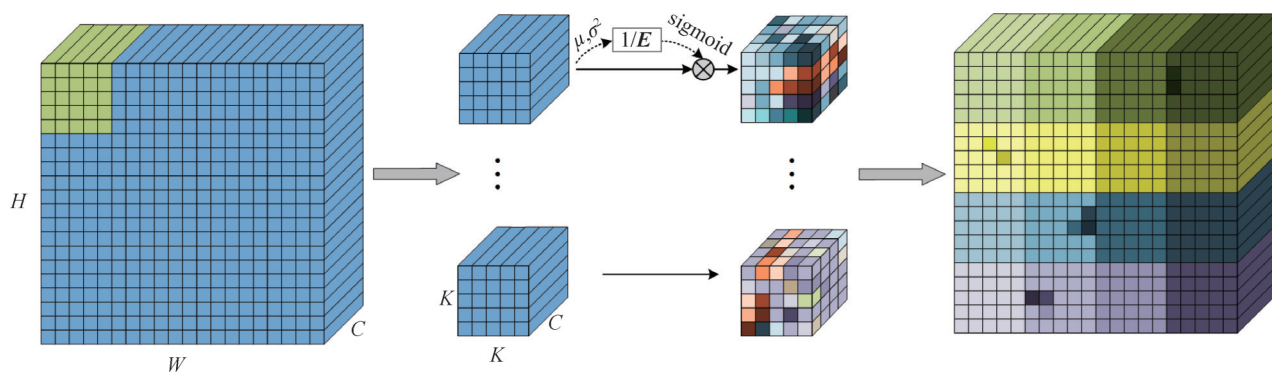


图2 LPFA模块示意图

Fig. 2 Schematic diagram of LPFA module

同尺度的目标,浅层实现对小目标的特征增强,中、深层实现对中、大型目标的增强,从而达到多尺度目标特征增强和背景信息抑制的目的。

1.2 W-CIoU联合损失

传统的基于交并比(intersection of union, IoU)的损失(如 DIoU (distance IoU)、CIoU (complete IoU)、Focal-EIoU (focal efficient IoU)、SIoU (shape IoU)、WIoU (wise IoU)等)对不同尺度目标的灵敏度变化大,小位置偏移导致样本的正负标签反转,难以合理分配样本,网络收敛困难,检测精度较低。为解决上述问题,Wang等人(2021)提出归一化最优传输距离(normalized Wasserstein distance, NWD)计算边界框的相似性。对于预测框 $P = (x_p, y_p, w_p, h_p)$ 与真实框 $G = (x_g, y_g, w_g, h_g)$,二者的NWD度量值 W 具体为

$$W = \exp \left(- \sqrt{\left\| \left[\begin{matrix} x_p, y_p, \frac{w_p}{2}, \frac{h_p}{2} \end{matrix} \right]^T, \left[\begin{matrix} x_g, y_g, \frac{w_g}{2}, \frac{h_g}{2} \end{matrix} \right]^T \right\|_2^2} \times \frac{1}{D} \right) \quad (4)$$

式中, x, y 代表边界框的中心坐标, w, h 代表边界框的宽和高, D 为数据集相关常数。尽管NWD改善了小目标的尺度敏感性,但会使得中大型目标的相似性度量与小目标一致,不同尺度目标在相同偏移情况下的优劣程度无法有效区分。图3表示不同尺度下各方法的偏移曲线比较,其中图3(a)模拟两框之间的偏移情况,图3(b)~(d)分别表示不同大小的框在不同偏移情况下得到的CIoU、NWD和W-CIoU度量值(偏移距离为水平或垂直距离)。

如图3(b)(c)所示,这种情况可能带来无法有效发掘困难样本,导致网络无法收敛的问题。在相同偏移情况下,各尺度目标的相似性度量应具备以

下特点:1)小($< 32 \times 32$)、中($32 \times 32 \sim 96 \times 96$)、大($> 96 \times 96$)型目标的相似性度量应呈现递增趋势,具有一定可分性。2)小目标的相似性度量随着位置变化应较快,而中、大型目标变化应较慢。为此,提出了W-CIoU联合度量方法,具体为

$$W_c = \alpha W + (1 - \alpha)C, \alpha \in [0, 1] \quad (5)$$

式中, W_c 表示W-CIoU值, α 表示NWD所占权重,默认值为0.5, C 为CIoU值。由图3(d)可知,在相同偏移下,W-CIoU对于各尺度目标具备可分性。在有重叠情况下,小尺度目标的W-CIoU值变化相对较快,而中、大型目标变化相对较慢,符合上述特性。此外,在无重叠情况时,各尺度目标仍具有一定区分性。最终,本文提出的回归损失 L_{reg} 为

$$L_{reg} = \alpha(1 - W) + (1 - \alpha)(1 - C) \quad (6)$$

该损失有效缓解了小尺度目标对位置偏移的敏感性,也使不同尺度目标具备可分性,相比现有损失具备优越性,2.5.2节展示了各损失的详细对比结果。

2 实验与分析

2.1 实验环境与评价指标

实验在Linux系统中进行,Python版本为3.8,深度学习框架PyTorch版本为1.12。主要超参数设置如下:最大学习率=0.01,最小学习率=最大学习率 $\times 0.01$,动量=0.937, batchsize=8,权重衰减系数=0.0005,优化器为SGD (stochastic gradient descent),学习率衰减策略采用余弦退火,共训练300个epoch。

实验采用平均准确率(average precision, AP)、平均准确率均值(mean average precision, mAP)及各尺度平均准确率(AP_s (小尺度), AP_m (中尺度), AP_l

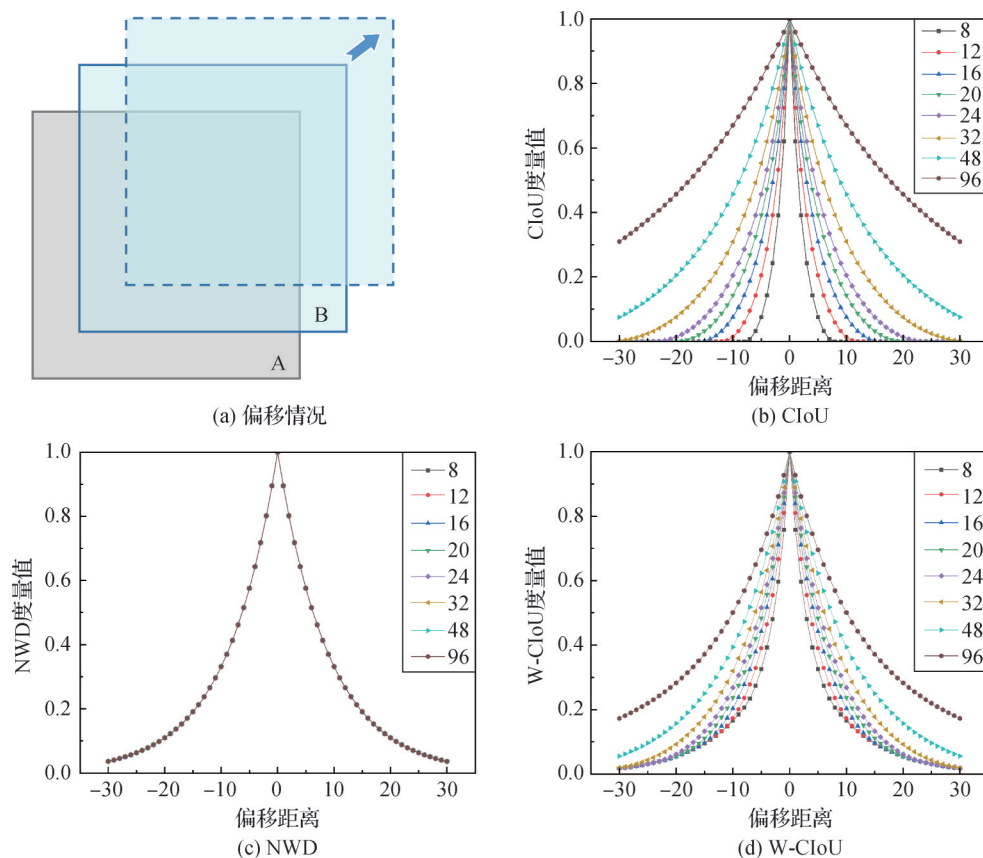


图3 不同尺度下各方法的偏移曲线比较

Fig. 3 Comparison of deviation curves of each method in different scales ((a) offset condition; (b) CIoU; (c) NWD; (d) W-CIoU)

(大尺度))评估算法性能。其中AP通过查准率和召回率计算得到,mAP是所有类别AP的平均值。

2.2 数据集介绍

实验采用RSOD(remote sensing object detection)(Long等,2017)和RSSOD(remote sensing super-resolution object detection)(Wang等,2022)数据集进行验证。RSOD数据集中小目标数量较多,包括976幅遥感影像,共4类目标,其中立交桥通常为特大型目标,因此本文选取飞机、油罐和操场3类目标进行验证。实验中训练集、验证集和测试集之比为6:2:2。RSSOD数据集中背景相对复杂,共1759幅图像,实验选取其中车辆、树和船舶3类目标,按照原始子集比例(7:2:1)进行划分。图4第1行为RSOD数据集实例图像,第2行为RSSOD数据集实例图像。

2.3 参数敏感性分析

为探究W-CIoU损失函数中相关参数对模型性能的影响,在两个数据集上进行参数敏感性实验。首先采用默认参数 $\alpha = 0.5$ 探究D的最佳值,再将D的最佳值代入,进一步探究 α 的最佳值,实验结果如表1所示。由表1可知,针对RSOD数据集,D的最优

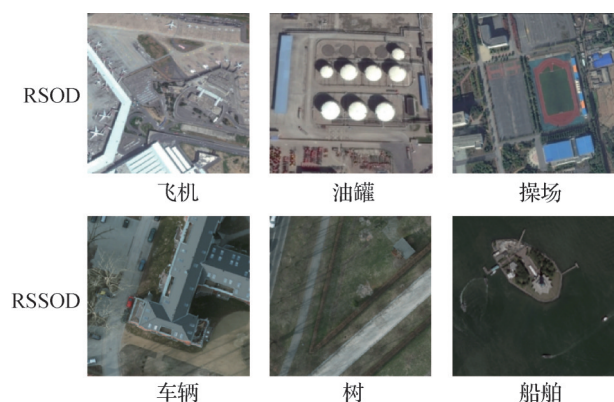


图4 RSOD和RSSOD数据集图像实例

Fig. 4 Image instances of RSOD and RSSOD datasets

值为16, α 的最优值为0.3。同理,在RSSOD数据集中,D的最优值为20, α 的最优值为0.5。D和 α 两个参数的最优取值在两个数据集上有一定差异,但都在一定范围内波动,具有较好的稳定性。

2.4 实验结果

为验证本文算法相比其他算法的优越性,选取7种先进算法EfficientNet(Tan和Le,2019)、YOLOv4(Bochkovskiy等,2020)、DETR(Carion等,2020)、Liu

表1 两个数据集上参数敏感性实验结果

Table 1 Results of parameter sensitivity experiment on two datasets

D	mAP/%		α	mAP/%	
	RSOD	RSSOD		RSOD	RSSOD
8	97.7	85.3	0.1	97.8	85.6
10	97.7	84.8	0.2	97.7	85.7
12	97.4	85.4	0.3	97.9	85.0
14	97.7	85.6	0.4	97.7	85.6
16	97.8	85.7	0.5	97.8	86.1
20	97.8	86.1	0.6	97.7	85.1
24	97.6	85.9	0.7	97.6	84.7
32	97.6	85.8	0.8	97.6	84.6
64	97.7	85.1	0.9	97.5	83.1
96	97.5	84.2	1.0	97.6	82.9

注:加粗字体表示各列最优结果。

等人(2021)提出的SwinT(Swin Transformer)、Qiao等人(2021)提出的DetecoRS(detecting objects with

recursive feature pyramid and switchable atrous convolution)、Liu等人(2022)提出的DAB-DETR(dynamic anchor boxes for DETR)和YOLOv8s(Jocher等,2023)进行对比实验,表2为本文算法与其他算法在RSOD和RSSOD数据集上的对比结果。各对比算法均使用官方预训练权重,输入图像尺寸选择默认尺寸,其中EfficientNet、DETR、SwinT、DetecoRS、DAB-DETR输入图像尺寸为1 333 × 800像素,其余算法输入图像尺寸均为640 × 640像素。

由表2可知,本文算法在检测精度和模型大小上取得最佳效果。1)在检测精度上,该算法在两个数据集中的mAP值均高于当前算法,在RSOD数据集上相比两种代表算法DAB-DETR和YOLOv8s,其mAP分别提升0.8%和0.3%,在RSSOD数据集上相比DAB-DETR和YOLOv8s,mAP分别提升2.9%和2.6%;2)在模型大小与计算复杂度上,本文算法相较对比算法达到当前最低,相比总体效果最优的YOLOv8s算法,本文算法模型参数量与计算复杂度分别降低4.1 M和12.9 G。

表2 RSOD和RSSOD数据集上各算法检测结果对比

Table 2 Comparison of detection results of each algorithm on RSOD and RSSOD datasets

算法	参数量/M	复杂度/G	RSOD数据集				RSSOD数据集			
			飞机	油罐	操场	mAP/%	车辆	树	船舶	mAP/%
EfficientNet(Tan和Le,2019)	20.0	68.2	93.8	98.9	97.7	96.8	94.7	65.2	90.3	83.4
YOLOv4(Bochkovskiy等,2020)	64.0	60.0	88.7	99.1	99.8	95.9	93.3	61.9	81.6	78.9
DETR(Carion等,2020)	36.7	73.6	92.6	98.5	99.9	97.0	96.6	66.7	84.0	82.4
SwinT(Liu等,2021)	66.1	127.2	95.6	96.7	97.4	96.6	94.2	59.7	85.1	80.0
DetecoRS(Qiao等,2021)	123.2	117.2	94.3	96.7	99.5	96.8	93.3	48.2	77.8	73.1
DAB-DETR(Liu等,2022)	42.9	43.7	93.7	98.6	99.9	97.4	95.2	66.1	92.2	84.5
YOLOv8s(Jocher等,2023)	11.1	28.7	97.1	98.9	97.8	97.9	97.1	62.0	95.4	84.8
LPFA-YOLO(本文)	7.0	15.8	97.7	99.3	97.5	98.2	97.2	66.9	98.0	87.4

注:加粗字体表示各列最优结果。

图5为本文算法与Baseline、DAB-DETR和YOLOv8s在RSOD数据集的可视化结果。在图5第1行图像中,DAB-DETR与YOLOv8s的检测效果较差,前者依赖于大规模数据,且对小目标不友好,后者对于小目标的优化也不足。LPFA-YOLO检出的小目标更多,在小目标检测中更具优势。图5第2行图像中,DAB-DETR存在误检情况,YOLOv8s和LPFA-YOLO均没有误检情况。图5第3行图像中,

Baseline存在误检案例,其余算法均无误检。

图6为本文算法与Baseline、DAB-DETR、YOLOv8s在RSSOD数据集的可视化结果(第3行图像进行了长度压缩)。图6第1行图像中,LPFA-YOLO与DAB-DETR检出隐藏的目标,这得益于二者都应用了注意力机制。图6第2行图像中,场景相对复杂,DAB-DETR与YOLOv8s均存在漏检与误检情况,而LPFA-YOLO检测效果明显优于两者,表明其

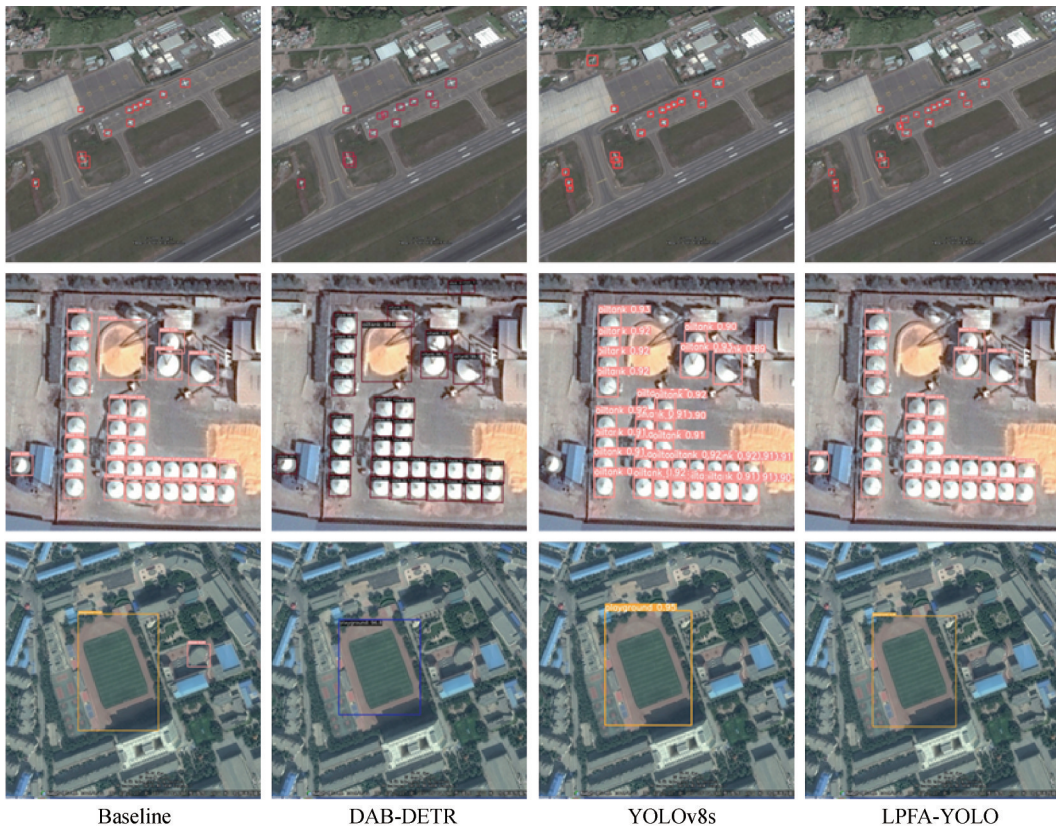


图5 各算法在RSOD数据集的可视化检测结果

Fig. 5 Visual detection results of each algorithm on RSOD dataset

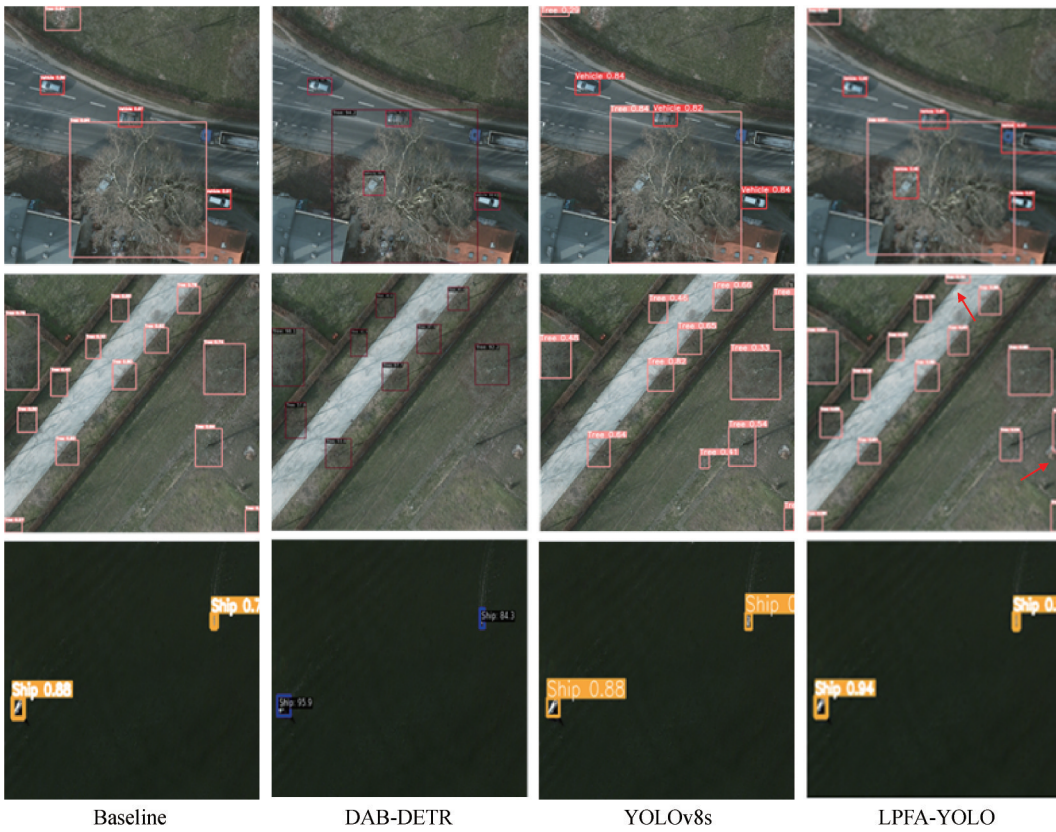


图6 各算法在RSSOD数据集的可视化检测结果

Fig. 6 Visual detection results of each algorithm on RSSOD dataset

具有更好的前景—背景区分能力。此外,图6第2行图像中,LPFA-YOLO在检测不完全的目标时具有良好表现(红色箭头部分)。图6第3行图像中,所有算法均检出两个目标,相对而言LPFA-YOLO置信度更高。总体上,LPFA-YOLO置信度高于其他算法,尽管DAB-DETR在少数目标中的置信度较高,但其对于小目标以及背景较为复杂的目标的置信度低于LPFA-YOLO。

本文算法存在进一步优化空间,图5第3行图像中,LPFA-YOLO精度与YOLOv8s齐平,低于DAB-DETR。DAB-DETR利用自注意力机制和可学习对象关系建模,能够从全局理解图像,并不依赖预定义锚框,因此在中、大型目标中具有良好表现。

2.5 消融实验

表3和表4分别为随机种子固定条件下本文算法在两个数据集上的消融实验结果。对于RSOD数

据集,本文算法在YOLOv5s基础上进一步提升了检测性能。其中飞机提升0.6%,油罐提升0.6%,操场提升1.4%,mAP提升0.9%,小目标指标 AP_s 提升2.4%,中、大型目标分别达到67.7%和82.1%,这表明该算法能有效增强多尺度目标的特征表示。对于RSSOD数据集,改进方法检测结果显示,车辆提升0.4%,树提升2.8%,这说明模型在复杂背景中的检测性能得到提升。此外,船舶提升4.9%,mAP提升2.7%,针对小目标的指标 AP_s 提升2.7%,中、大型目标分别达到52.7%和46.4%,这也表明模型对多尺度目标的表示能力得到增强。

由图5和图6可视化结果可知,改进方法检出的目标数量与置信度均得到提升。结果表明,本文提出的LPFA-YOLO算法在抑制背景信息干扰、提升小目标检测性能方面是有效的,具有较好的可靠性。

表3 RSOD数据集消融实验结果

Table 3 Results of ablation experiments on RSOD dataset

BPFA	W-CIoU	飞机	油罐	操场	mAP	AP_s	AP_m	AP_L
-	-	97.1	98.7	96.1	97.3	40.3	66.6	80.5
√	-	97.3	98.9	97.4	97.8	41.5	67.8	81.8
-	√	97.6	99.2	96.8	97.9	42.1	67.7	81.1
√	√	97.7	99.3	97.5	98.2	42.7	67.7	82.1

注:加粗字体表示各列最优结果。“√”为采用,“-”为未采用。

表4 RSSOD数据集消融实验结果

Table 4 Results of ablation experiments on RSSOD dataset

BPFA	W-CIoU	车辆	树	船舶	mAP	AP_s	AP_m	AP_L
-	-	96.8	64.1	93.1	84.7	35.3	51.3	42.0
√	-	97.0	67.4	94.1	86.2	37.2	52.1	43.1
-	√	97.1	65.2	95.9	86.1	37.7	51.5	45.0
√	√	97.2	66.9	98.0	87.4	38.0	52.7	46.4

注:加粗字体表示各列最优结果。“√”为采用,“-”为未采用。

2.5.1 LPFA有效性验证

为客观验证LPFA的有效性,对输入图像的提取特征进行可视化,结果如图7所示。其中,图7(a)为第2阶段(第1个C3模块)的特征图,图7(b)为第4阶段(第2个C3模块)的特征图,图7(c)为最终输

出层的热力图,图7(d)为图7(c)的局部细节图。由图7(a)可知,虽然LPFA在浅层中优势较小,也可以使得目标与背景信息进行一定程度分离(红框中添加LPFA后,模型对于目标特征更加聚焦,背景与目标混淆程度减小)。随着网络加深,LPFA体现出其

显著优势。对于较深层的低分辨率特征,LPFA能够从当前的局部区域内找到具有显著抑制作用的神经元(即目标特征),赋予其更大的权重,使得目标与背景分离,如图7(b)所示。由图7(c)可知,引入LPFA后,在最终输出阶段,网络对各尺度目标的关注度得到提升,针对一些与背景相似的目标,如图7(d)中隐藏在白色场地背景中的飞机(红框中目标),模型能够有效突出其有用特征。同时由表3、表4定量分析可知,在添加LPFA后,模型总体检测性能得到改善。因此,本文提出的LPFA机制能够有效抑制背景信息,增强目标特征,有利于后续检测任务。

2.5.2 损失函数对比实验

为验证设计的损失函数的优越性,选取部分先进损失函数进行对比实验。表5为各损失函数在RSOD数据集上的评估结果。由表5定量分析可知,相比其他损失函数,W-CIoU具有明显提升,针对飞机类目标,AP提高了0.5%,表明NWD能更高效地处理小目标检测任务,同时在其他类别目标中的检

测精度也有一定提升。图8为训练过程中各损失函数的损失曲线。由图8可知,W-CIoU损失初始值低于其他损失,且收敛时仍处在最低水平,这表明使用W-CIoU计算所得度量值更高,使得模型能更有效地处理正负样本的分配问题。此外,在训练过程中,W-CIoU的曲线斜率高于其他损失函数,表明使用W-CIoU能够进一步加快网络的收敛速度。

表5 各损失函数在RSOD数据集上的评估结果
Table 5 Evaluation results of each loss function on RSOD dataset

损失	飞机	油罐	操场	mAP	AP _s	AP _M	AP _L	/%
CIoU	97.1	98.7	96.1	97.3	40.3	66.6	80.5	
SIoU	97.4	98.5	96.7	97.5	41.2	66.6	80.9	
EIoU	97.5	99.2	96.6	97.8	40.3	67.5	79.9	
WIoU	97.2	98.9	95.7	97.3	41.5	67.1	78.9	
W-CIoU	97.6	99.2	96.8	97.9	42.1	67.7	81.1	

注:加粗字体表示各列最优结果。

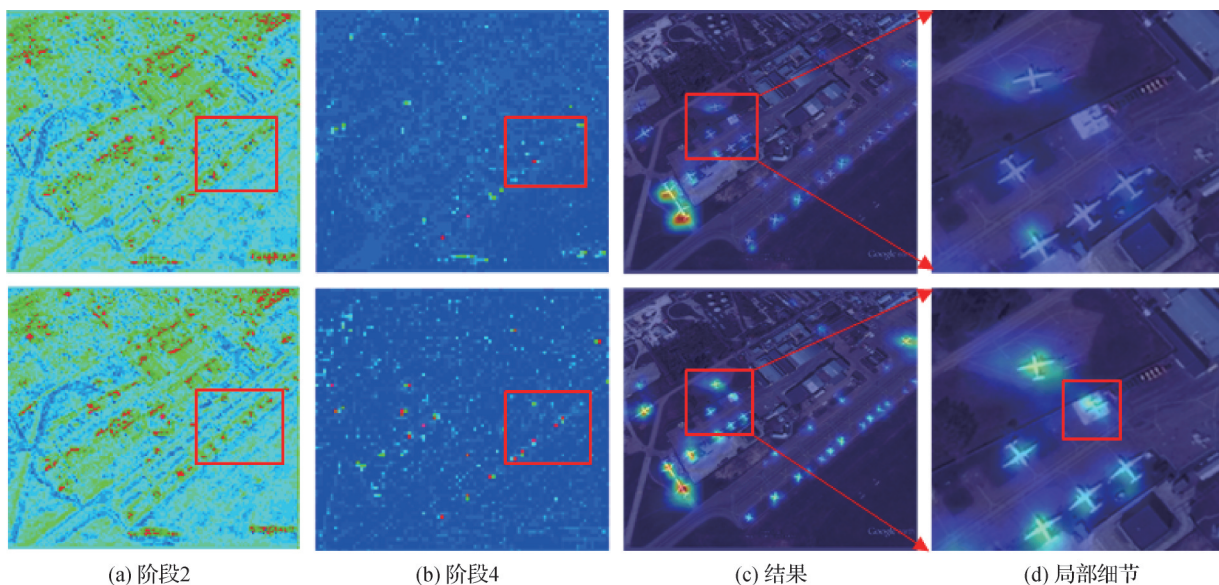


图7 添加LPFA前后的可视化热力图
Fig. 7 Visual heatmaps before and after adding LPFA ((a) stage 2; (b) stage 4; (c) results; (d) local detail)

2.6 外部验证

为进一步评估本文算法的泛化性,使用外部数据对算法进行了验证。本文从Google Earth上截取了两幅遥感影像(尺寸大小为1491×971像素),由于对整幅图像进行检测相对困难,因此裁剪出部分区域送入到模型中进行测试(输入大小经预处理后

与训练一致,为640×640像素),得到最终检测结果,如图9所示。1)图9(b)中,本文算法在目标区域中表现良好,成功检出其中的多个目标;2)图9(d)中,对于密集目标,本文算法也具有较好的检测效果。结果表明,本文算法在外部数据中展示出较好的检测性能,具有良好的泛化性。

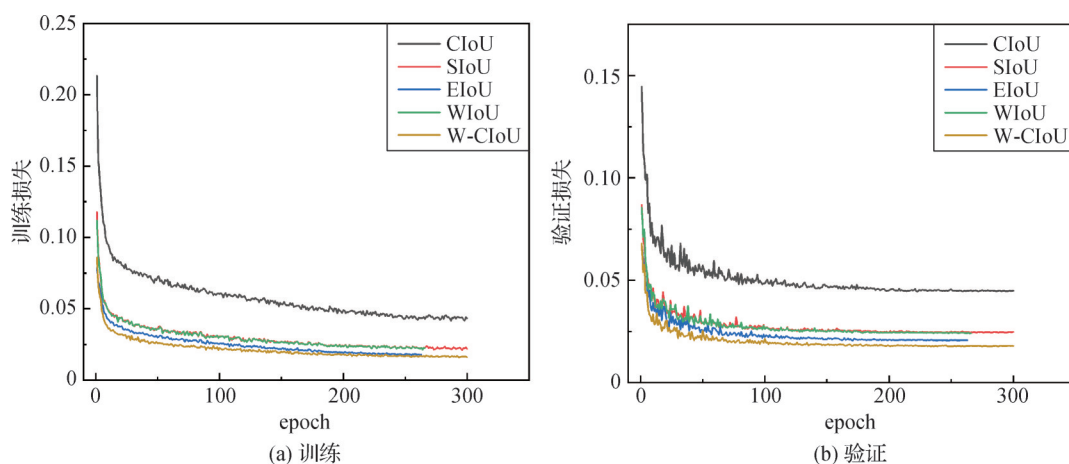


图8 各损失函数损失曲线

Fig. 8 Loss curves of each loss function ((a) training; (b) validating)

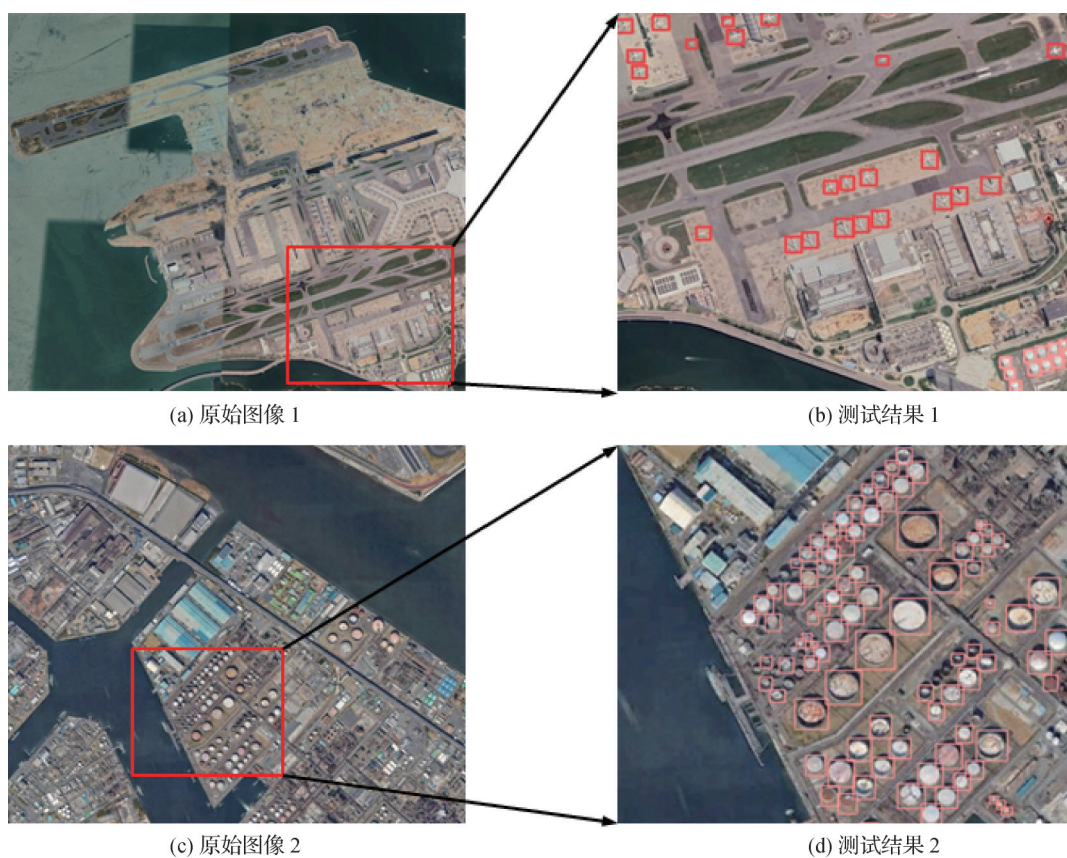


图9 外部验证结果

Fig. 9 External verification results ((a) original image 1; (b) test result 1; (c) original image 2; (d) test result 2)

3 结论

本文针对遥感目标检测任务存在的背景干扰、待检目标小且多等问题展开研究,提出了一种局部无参注意力YOLO网络检测算法。骨干网络中嵌入的局部无参注意力模块,使骨干网络能够抑制背景

信息并增强多尺度目标特征,提高其特征提取能力。同时,本文探索了更高效的相似性度量方法,利用最优传输距离设计了W-ClIoU联合度量和损失函数,缓解小目标锚框与真实框差异过大导致的标签误分配问题,提高小目标的检测性能。在RSOD和RSSOD数据集上,该算法的检测性能相比大多现有算法得到提升,且模型参数量和复杂度较小,并在外部数据

中对算法的泛化性进行了验证。

由于该算法仍采用单解耦头,导致分类与回归任务共用同一输入特征,而通常分类任务更关注深层语义信息,回归任务更关注浅层细节信息,这限制了检测性能的进一步提升,因此后续工作将探索双分支解耦头,针对不同任务设计特定的解耦头,以减小不同任务间可能存在的干扰。同时考虑到模型部署问题,可以采用知识蒸馏方法对模型进行轻量化,将大型教师模型学习到的知识(标签、中间特征等)传递给小型学生模型,在降低模型复杂度和减少计算资源需求的情况下进一步提升其检测性能。

参考文献 (References)

- Blaschke T. 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1): 2-16 [DOI: 10.1016/j.isprsjprs.2009.06.004]
- Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. [2024-06-13]. <https://arxiv.org/pdf/1512.03012.pdf>
- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with Transformers//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 213-229 [DOI: 10.1007/978-3-030-58452-8_13]
- Cheng G and Han J W. 2016. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117: 11-28 [DOI: 10.1016/j.isprsjprs.2016.03.014]
- Dalal N and Triggs B. 2005. Histograms of oriented gradients for human detection//Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE: 886-893 [DOI: 10.1109/CVPR.2005.177]
- He K M, Zhang X Y, Ren S Q and Sun J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9): 1904-1916 [DOI: 10.1109/TPAMI.2015.2389824]
- Hong D F, Zhang B, Li X Y, Li Y X, Li C Y, Yao J, Yokoya N, Li H, Ghamisi P, Jia X P, Plaza A, Gamba P, Benediktsson J A and Chanussot J. 2024. SpectralGPT: spectral remote sensing foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8): 5227-5244 [DOI: 10.1109/TPAMI.2024.3362475]
- Joher G, Chaurasia A and Qiu J. 2023. Ultralytics YOLO (Version 8.0). <https://github.com/ultralytics/ultralytics>.
- Li C Y, Zhang B, Hong D F, Yao J and Chanussot J. 2023a. LRR-Net: an interpretable deep unfolding network for hyperspectral anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5513412 [DOI: 10.1109/TGRS.2023.3279834]
- Li Y X, Hou Q B, Zheng Z H, Cheng M M, Yang J and Li X. 2023b. Large selective kernel network for remote sensing object detection//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 16748-16759 [DOI: 10.1109/ICCV51070.2023.01540]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Liu S L, Li F, Zhang H, Yang X, Qi X B, Su H, Zhu J and Zhang L. 2022. DAB-DETR: dynamic anchor boxes are better queries for DETR [EB/OL]. [2024-06-13]. <https://arxiv.org/pdf/2201.12329.pdf>
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2017. SSD: single shot multibox detector//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021. Swin Transformer: hierarchical vision Transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]
- Long Y, Gong Y P, Xiao Z F and Liu Q. 2017. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5): 2486-2498 [DOI: 10.1109/TGRS.2016.2645610]
- Qiao S Y, Chen L C and Yuille A. 2021. DetectorRS: detecting objects with recursive feature pyramid and switchable atrous convolution//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 10208-10219 [DOI: 10.1109/CVPR46437.2021.01008]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Sedaghat A, Mokhtarzade M and Ebadi H. 2011. Uniform robust scale-invariant feature matching for optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11): 4516-4527 [DOI: 10.1109/TGRS.2011.2144607]
- Shi Z H, Wu C W, Li C J, You Z Z, Wang Q and Ma C C. 2023. Object detection techniques based on deep learning for aerial remote sens-

- ing images: a survey. *Journal of Image and Graphics*, 28(9): 2616-2643 (石争浩, 仵晨伟, 李成建, 尤珍臻, 王泉, 马城城). 2023. 航空遥感图像深度学习目标检测技术研究进展. *中国图象图形学报*, 28(9): 2616-2643 [DOI: 10.11834/jig.221085]
- Tan M X and Le Q. 2019. EfficientNet: rethinking model scaling for convolutional neural networks//*Proceedings of the 36th International Conference on Machine Learning*. Long Beach, USA: [s. n.]: 6105-6114
- Wang J W, Yang W, Guo H W, Zhang R X and Xia G S. 2021. Tiny object detection in aerial images//*Proceedings of the 25th International Conference on Pattern Recognition*. Milan, Italy: IEEE: 3791-3798 [DOI: 10.1109/ICPR48806.2021.9413340]
- Wang Y, Bashir S M A, Khan M, Ullah Q, Wang R, Song Y L, Guo Z and Niu Y L. 2022. Remote sensing image super-resolution and object detection: benchmark and state of the art. *Expert Systems with Applications*, 197: #116793 [DOI: 10.1016/j.eswa.2022.116793]
- Yang L X, Zhang R Y, Li L D and Xie X H. 2021a. SimAM: a simple, parameter-free attention module for convolutional neural networks//*Proceedings of the 38th International Conference on Machine Learning*. Virtual Event: [s. n.]: 11863-11874
- Yang X, Yan J C, Feng Z M and He T. 2021b. R3Det: refined single-stage detector with feature refinement for rotating object//*Proceedings of the 35th AAAI Conference on Artificial Intelligence*. Virtually: AAAI: 3163-3171 [DOI: 10.1609/aaai.v35i4.16426]
- Yuan X, Cheng G, Li G, Dai W, Yin W X, Feng Y C, Yao X W, Huang Z L, Sun X and Han J W. 2023. Progress in small object detection for remote sensing images. *Journal of Image and Graphics*, 28(6): 1662-1684 (袁翔, 程焱, 李戈, 戴威, 尹文昕, 冯焱超, 姚西文, 黄钟冷, 孙显, 韩军伟). 2023. 遥感影像小目标检测研究进展. *中国图象图形学报*, 28(6): 1662-1684 [DOI: 10.11834/jig.221202]
- Zhang C, Su J R, Ju Y K, Lam K M and Wang Q. 2023. Efficient inductive vision Transformer for oriented object detection in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5616320 [DOI: 10.1109/TGRS.2023.3292418]
- Zhao W Q, Kong Z X, Zhou Z D and Zhao Z B. 2021. Target detection algorithm of aerial remote sensing based on feature enhancement technology. *Journal of Image and Graphics*, 26(3): 644-653 (赵文清, 孔子旭, 周震东, 赵振兵). 2021. 增强小目标特征的航空遥感目标检测. *中国图象图形学报*, 26(3): 644-653 [DOI: 10.11834/jig.190612]

作者简介

夏波,男,硕士研究生,主要研究方向为遥感影像处理和机器学习。E-mail:20192504@stu.cqu.edu.cn

黄鸿,通信作者,男,教授,主要研究方向为流形学习、模式识别和遥感影像智能化处理。E-mail:hhuang@cqu.edu.cn

薛卫涛,男,高级技师,主要研究方向为航天产品几何量测量。E-mail:516228117@qq.com

周新尧,男,博士研究生,主要研究方向为遥感影像处理和多模态融合。E-mail:zhouxinyao@stu.cqu.edu.cn