

一种改进的 Faster R-CNN 遥感图像多目标检测模型研究

苗茹^{1,2}, 李祎^{1,2,3}, 周珂^{1,2,3*}, 张俨娜¹, 常然然^{1,2,3}, 孟更^{1,2,3}

(1. 河南大学计算机与信息工程学院, 河南 开封 475004; 2. 河南省空间信息处理工程研究中心, 河南 开封 475004;

3. 河南省时空大数据技术创新中心, 河南 开封 475004)

摘要: 针对遥感图像背景复杂、目标种类多和尺度差异大所造成的目标漏检和误检问题, 提出一种改进 Faster R-CNN 多目标检测模型。首先, 采用 Swin Transformer 来替代 ResNet 50 骨干网络, 增强模型特征提取能力; 其次, 添加平衡特征金字塔(BFP)模块融合浅层和高层语义信息, 进一步加强特征融合效果; 最后, 在分类和回归分支中, 添加动态权重机制, 促进网络在训练过程中更关注高质量候选框, 提高目标定位和分类的精确度。在 RSOD 数据集上的实验结果表明, 所提模型相较于 Faster R-CNN 模型每秒浮点运算次数(FLOPs)大幅度减少, 并且模型的 mAP@0.5:0.95 提高了 10.7 个百分点, 平均召回率提高 10.6 个百分点。相较于其他主流检测模型, 所提模型在降低漏检率的同时, 取得了更高的精度, 能显著提高复杂背景下遥感图像的检测精度。

关键词: 遥感图像; 多目标检测; Faster R-CNN; Swin Transformer 模块; 平衡特征金字塔; 动态权重机制

源代码链接: <https://github.com/lee0019/myown.git>

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0068856

A Study on Improved Faster R-CNN Model for Multi-Object Detection in Remote Sensing Images

MIAO Ru^{1,2}, LI Yi^{1,2,3}, ZHOU Ke^{1,2,3*}, ZHANG Yanna¹, CHANG Ranran^{1,2,3}, MENG Geng^{1,2,3}

(1. College of Computer and Information Engineering, Henan University, Kaifeng 475004, Henan, China;

2. Henan Province Engineering Research Center of Spatial Information Processing, Kaifeng 475004, Henan, China;

3. Henan Provincial Spatio-Temporal Big Data Technology Innovation Center, Kaifeng 475004, Henan, China)

【Abstract】 The complex backgrounds, diverse target types, and significant scale variations in remote sensing images lead to target omission and false detection. To address these issues, this study proposes an improved Faster R-CNN multi-object detection model. First, the ResNet 50 backbone network is replaced with the Swin Transformer to enhance the model's feature extraction capability. Second, a Balanced Feature Pyramid (BFP) module is introduced to fuse shallow and deep semantic information, further strengthening the feature fusion effect. Finally, in the classification and regression branches, a dynamic weighting mechanism is incorporated to encourage the network to focus more on high-quality candidate boxes during training, thereby improving the precision of target localization and classification. The experimental results on the RSOD dataset show that the proposed model significantly reduces the number of Floating-Point Operations per second (FLOPs) compared to the Faster R-CNN model. The proposed model achieves 10.7 percentage points improvement in mAP@0.5:0.95 and 10.6 percentage points increase in Average Recall (AR). Compared to other mainstream detection models, the proposed model achieves higher accuracy while reducing the false detection rate. These results indicate that the proposed model significantly enhances detection accuracy in remote sensing images with complex backgrounds.

【Key words】 remote sensing images; multi-object detection; Faster R-CNN; Swin Transformer module; Balanced Feature Pyramid(BFP); dynamic weighting mechanism

0 引言

在民用领域, 遥感图像目标检测应用于城市规划^[1]、环境监测^[2]和自然资源管理^[3]等方面。而在军事领域, 有助于完成情报收集和目標识别等任务。遥感图像目标检测往往具有复杂的背景, 而目标在

尺度和形状上存在着巨大的差异。因遥感图像的特殊性, 导致目标漏检和误检。提高遥感图像多目标检测的准确性和效率已成为亟待解决的问题。

随着深度学习的不断发展, 其在遥感图像领域展现出卓越的潜力, 众多学者都是通过改进网络特征来提升模型检测性能。文献[4]提出改进的特征金字塔

收稿日期: 2023-11-16 修回日期: 2024-04-01

基金项目: 高分辨率对地观测系统国家科技重大专项(民用部分)科研项目(80-Y50G19-9001-22/23); 河南省科技攻关项目(222102210061)。

通信作者 E-mail: *zhouke@henu.edu.cn

网络(FPN)通过拆分空间维度和通道维度的相关性,减少卷积计算所需的参数个数,提升卷积核参数的使用效率,使用大卷积高效直观地增加感受野,减少采样过程带来的特征,但检测头限制了模型对小目标的特征提取能力。文献[5]借助多层级融合结构,将深层次的语义特征与浅层次的细节特征相结合,生成多种尺度的既具有精确的位置信息又具有深层次的语义特征的特征图,但是在提高目标检测精度的同时降低了目标检测速率。文献[6]提出将双向尺度特征融合网络作为特征融合器,采用自顶向下和自底向上的特征融合策略,有效促进不同层次特征之间信息的交互。该方法提升了模型检测精度,但应用场景单一。文献[7]结合特征分离合并思想,对网络模型中的MPConv模块进行改进,减少网络特征处理过程造成的特征损失,但受检测范围的限制。

此外,研究人员从回归损失函数角度提升模型检测精度。文献[8]将边框回归损失函数改进为Wise-IoU,有效提高了中、大尺度遥感目标在高交并比(IoU)阈值情况下的均值平均精度,但对微小目标的检测优化效果不明显。文献[9]将坐标注意力机制融合到YOLOv5的骨干网络、颈部和输出端3个位置,以提高检测器的特征提取能力。采用CIoU-loss作为损失函数,改善目标检测框的定位精度,在提升精度的同时增大了参数量。文献[10]引入SIoU损失函数,快速稳定边框回归,重新定义惩罚指标,并加快整体网络的收敛速度,在一定程度上增加了算法的参数量,导致检测速度下降。

综上所述,在遥感图像中受背景复杂的影响,导致目标的尺寸较小,所占比例降低,往往难以获得足够的上下文信息对检测器进行识别和定位。此外,目标经过层层降维后,细节信息丢失,进一步增加了目标检测的难度。目标尺度变化剧烈也是一个重要的挑战。在遥感图像中,受视角高度变化的影响,目标尺度可能会发生显著变化。对于大部分目标,其尺度可能较小,提供的信息有限。同时,随着网络深度增加,计算量和参数量增加,目标细节信息逐渐丢失,因此需要网络结构对目标具有敏感性。

本文提出一种改进的Faster R-CNN^[11]网络结构FSBD R-CNN,专门针对遥感图像中的多目标检测问题进行优化。遥感图像常面临背景复杂、目标多样、尺度差异大的挑战,所提的模型有效缓解了目标漏检和误检的问题。本文的主要贡献如下:

1)为了增强特征提取能力,将传统的ResNet 50骨干网络替换为Swin Transformer。Swin Transformer能够有效捕捉遥感图像中的长距离依赖

关系,并提升多尺度信息的提取能力,从而更好地处理遥感图像的复杂场景问题。

2)通过引入平衡特征金字塔(BFP)模块优化不同层次特征之间的融合,能够保留更多的语义信息,特别是在低层次和高层次特征之间的平衡,从而提升模型在多尺度目标上的表现,尤其在复杂场景下有效减少信息丢失。

3)在分类和回归分支中加入动态权重机制,动态调整训练过程中对高质量候选框的关注,优化目标的定位精度和分类效果。该机制根据训练进展调整损失函数,进一步提高模型的精度。

1 总体流程

本文所设计的遥感图像多目标检测总体流程如图1所示。首先通过裁剪和数据集划分等操作对遥感图像数据集进行预处理,制作出可以直接输入到模型中进行训练、验证和测试的数据集;然后将训练集和验证集中的数据增强后,按固定批大小输入到改进后的Faster R-CNN模型中训练;最后训练完成后使用评估指标对模型进行评估;在改进后的模型上进行模块消融实验、不同模型对比实验,通过评价指标和可视化检测结果评估改进模型的可行性。

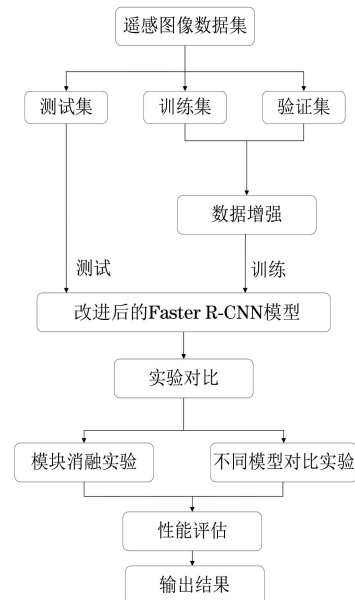


图1 总体技术流程

Fig.1 Procedure of overall technical

2 FSBD R-CNN 模型

许多研究^[12]已经证明了使用ResNet 50^[13]和FPN^[14]的结合对于目标检测任务的有效性。这些结构在广泛的应用领域中已被验证,并且具有较好的性能。本文采用Faster R-CNN+ResNet 50+

FPN 为基线模型进行优化,以满足对遥感图像多目标检测的需求。

基于 Faster R-CNN 改进的 FSBD R-CNN 遥感图像多目标检测算法的整体架构如图 2 所示,该流程主要有以下步骤:1)Swin Transformer^[15] 作为骨干网络(Backbone),负责从输入图像中提取 4 个尺度特征的特征图 $\{C_2, C_3, C_4, C_5\}$; 2)特征融合模块(Neck)和 FPN 联合平衡特征金字塔^[16],对特征图进行重整、精炼等进一步加工特征图,输出 $\{P_2, P_3, P_4, P_5\}$ 4 个尺度的特征图,以提供更丰富的语

义信息;3)候选区域提出网络(RPN)使用这些特征图生成一系列可能包含检测目标的候选目标区域,这些生成感兴趣的候选区域(RoI),通过 RoI 池化层被转换成具有固定大小的特征图;4)这些 RoI 池化后的特征图被传递给头部网络(Head),在头部网络添加动态权重机制^[17],分别用于目标分类和边界框回归,应用动态标签分配提高候选框质量,应用回归损失函数度量目标分类和边界框回归的准确性,有助于训练模型;5)为了消除候选框之间的重叠现象,采用非极大值抑制策略获取最终的目标检测结果。

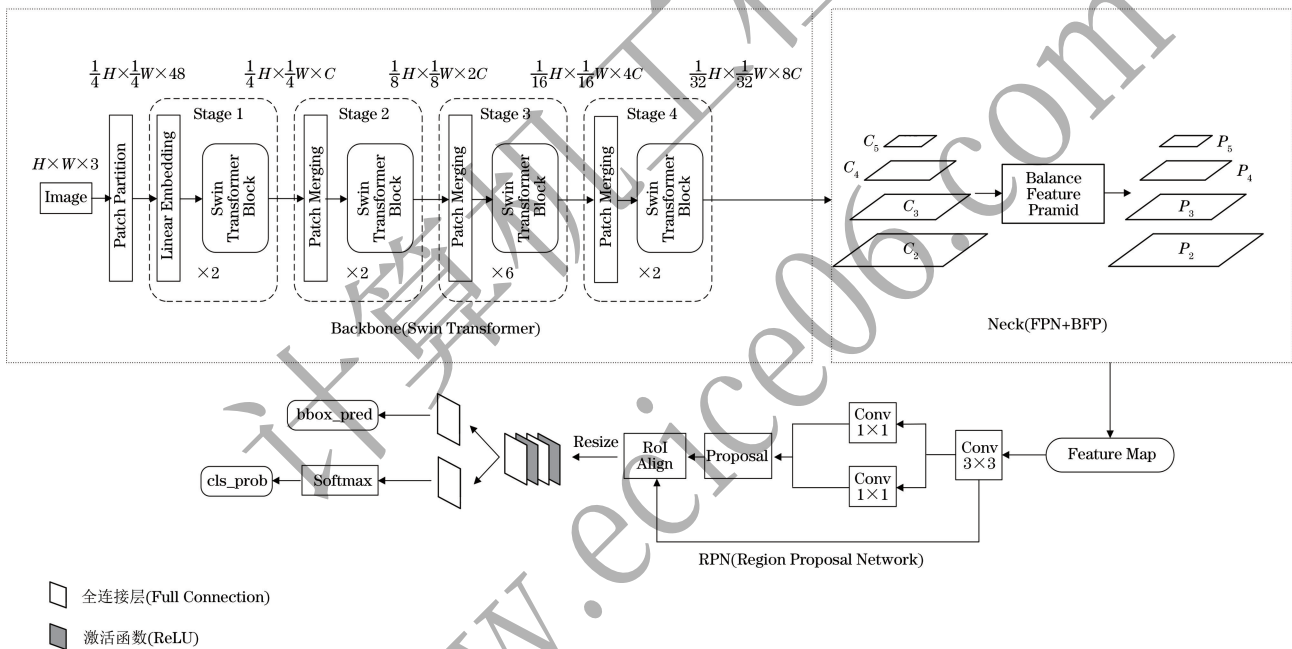


图 2 FSBD R-CNN 网络架构
Fig.2 Architecture of FSBD R-CNN network

2.1 骨干网络

ResNet 50 有效解决了深度神经网络中的梯度消失问题,但卷积操作中固定感受野限制了其在遥感图像捕捉大尺度上下文信息方面的能力。

本文引入 Swin Transformer 作为主干网络来提取特征,如图 2 所示(Backbone),有效地捕捉长距离依赖关系和全局上下文信息。此外, Svin Transformer 在局部区域内做自注意力计算,降低了整体计算复杂度。Swin Transformer 相较于 ResNet 50 减少了计算量,还能在目标检测任务中更灵活地捕捉多尺度信息,提高检测器在复杂场景下的性能。Swin Transformer 模块如图 3 所示。由 1 个基于窗口的多头自注意力(W-MSA)模块和 1 个基于移动窗口的多头自注意力(SW-MSA)模块组成,中间由 1 个带有 GeLU 非线性的 2 层多层感知器(MLP)组成,在每个 MSA 模块和每个 MLP 模块之前应用 1 个归一化层(LN),在每个模块之后应

用 1 个残差连接。

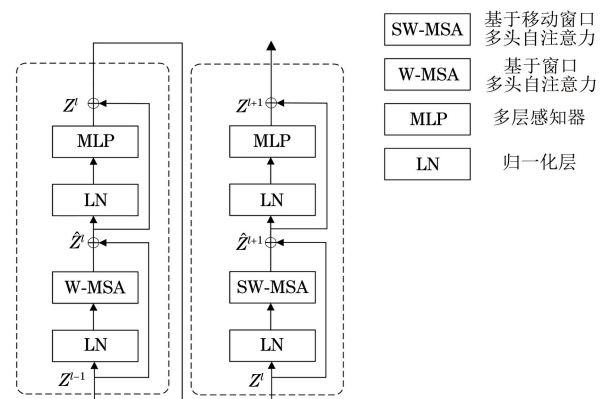


图 3 Swin Transformer 模块
Fig.3 Swin Transformer module

基于窗口的多头自注意力模块缺乏窗口间的连接,限制了模型的建模能力。为了引入窗口间的连接,同时保证非重叠窗口的高效划分方法,基于移动窗口的第 1 个模块采用常规窗口划分策略,从左上角像素

开始,将 $n \times n$ 特征图均匀分为 $m \times m$ 的 $\frac{n}{m} \times \frac{n}{m}$ 个窗口。下一个模块采用一种窗口配置,与前一层不同,从定期划分的窗口平移 $\lfloor \frac{n}{m} \times \frac{n}{m} \rfloor$ 个像素。通过位移窗口划分方法,连续的 Swin Transformer 模块计算式如下:

$$\hat{z}^l = \text{W-MSA}(\text{LN}(z^{l-1})) + z^{l-1} \quad (1)$$

$$z^l = \text{MLP}(\text{LN}(\hat{z}^l)) + \hat{z}^l \quad (2)$$

$$\hat{z}^{l+1} = \text{SW-MSA}(\text{LN}(z^l)) + z^l \quad (3)$$

$$z^{l+1} = \text{MLP}(\text{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1} \quad (4)$$

式中: $\hat{z}^l$ 和 z^l 分别表示第 l 个 Swin Transformer 模块中 W-MSA 和 MLP 的输出特征。基于移动窗口的多头自注意力操作如图 4 所示(彩色效果见《计算机工程》官网 HTML 版,下同)。

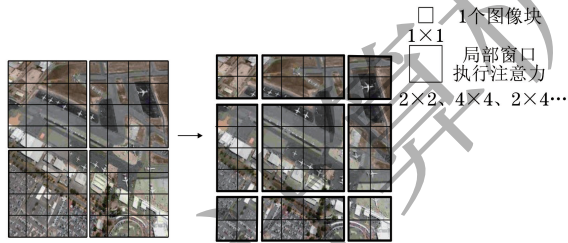


图 4 基于移动窗口的多头自注意力操作
Fig. 4 Multi-head self-attention operation based on shifted window

输入的整个图像建模后窗口以非重叠的方式排列,每个窗口包含 $n \times n$ 个图像块。对于 1 个大小为 $h \times w$ 的图像,全局多尺度注意力模块和基于窗口的多尺度注意力模块的复杂度计算式如下:

$$\Omega(\text{MSA}) = 4hwc^2 + 2(hw)^2C \quad (5)$$

$$\Omega(\text{W-MSA}) = 4hwc^2 + 2M^2hwc \quad (6)$$

式(5)中复杂度的计算与 $h \times w$ 的图像呈二次关系,当式(6)中的 M 固定时(默认为 $M=m$, m 为划分图像块的边长)呈线性关系,对于较大的 $h \times w$ 图像,全局注意力计算量过大,而基于窗口的自注意力具有可伸缩性。

2.2 特征融合模块

为了在遥感图像中提高目标检测的准确性,特征金字塔等方法通过横向连接将不同层次特征进行融合,以获得更丰富和全面的特征表示。然而,在特征金字塔中,信息整合方式对遥感图像目标检测的性能至关重要。

特征融合应该在每个分辨率层次包含均衡的信息,以便检测器能够适应不同尺度和分辨率的目标。这种特征不平衡的融合会对遥感图像目标检测的性能产生不利影响,因此本文添加平衡特征金字塔模

块,提高目标检测的精度和鲁棒性。BFP 模块通过整合平衡的语义特征来增强原始特征,以确保金字塔中的每个分辨率都获得相等的信息。平衡特征金字塔模块如图 5 所示。

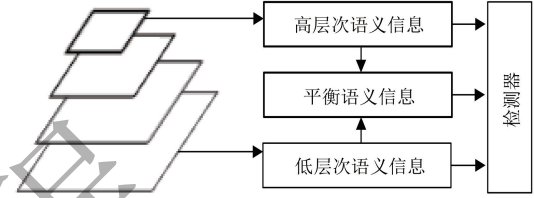


图 5 平衡特征金字塔模块

Fig. 5 Balanced feature pyramid module

这种平衡信息流的方式使得特征更富有判别性,从而提高目标检测的性能。与之前的方法不同,该方法采用重整、整合、精炼和强化步骤,集成平衡语义特征来增强多层次特征。这种方法有助于确保每个分辨率层次都能得到充分的语义信息。

分辨率水平为 l 的特征记为 C_l ,多层次特征的个数记为 L ,最底层和最高层的指标为 $l_{\min}$ 和 $l_{\max}$,在图中, C_2 具有高分辨率。为了整合多层次的特征并同时保持语义层次,首先将多层次特征的大小 $\{C_2, C_3, C_4, C_5\}$ 调整到中间大小,即与 C_4 相同的大小。这个过程包括插值和池化等操作,一旦特征被重新缩放,通过求平均得到平衡语义特征。

$$C = \frac{1}{L} \sum_{l=l_{\min}}^{l_{\max}} C_l \quad (7)$$

平衡特征金字塔与特征金字塔的过程相反,重新缩放获得特征,以加强原始特征,在此过程中,每个分辨率从其他分辨率获得相等的信息。平衡特征金字塔中的特征整合部分是非参数方法,不涉及任何参数。这种非参数方法的改进能有效促进特征图之间的融合。为了进一步提高特征的质量和判别性,采用高斯非局部注意力作为默认值,采用精炼步骤能增强集成的特性,进一步改善结果,提高目标的检测精度。在特征融合的 FPN 后添加平衡特征金字塔模块,能更好地融合从低级到高级的特征,输出 $\{P_2, P_3, P_4, P_5\}$ 用于目标检测,遵循特征金字塔的流程。特征金字塔联合平衡特征金字塔的结构如图 6 所示。在添加平衡特征金字塔模块后,其检测结果在小、中、大尺度上都得到了改善,平衡了低级层次和高级层次的语义信息。

2.3 动态权重机制

2.3.1 动态标签分配

标签分配问题是如何有效地为提出的目标区域分配正确标签,特别是在分类训练过程中。遥感场

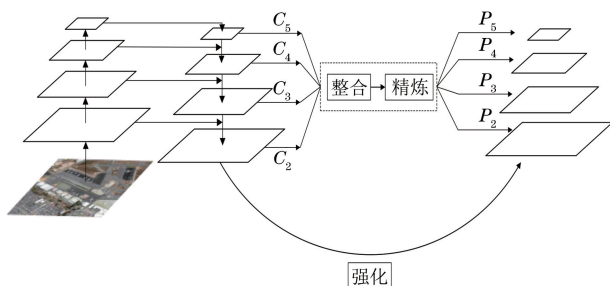


图 6 特征金字塔联合平衡特征金字塔的结构

Fig.6 Structure of feature pyramid combined with balanced feature pyramid

景的背景复杂,目标之间边界通常模糊不清,使得为目标区域分配正样本标签变得复杂。本文添加动态标签分配(DLA)提高候选框质量。

在基线模型中,通常采用 IoU 阈值来确定如何为检测框与真实目标之间分配标签。该过程以二元分类的规范化表示如下:

$$L_{\text{Label}} = \begin{cases} 1, & \text{if } \max I_{\text{IoU}}(b, G) \geq T_+ \\ 0, & \text{if } \max I_{\text{IoU}}(b, G) < T_- \\ -1, & \text{otherwise} \end{cases} \quad (8)$$

式中: b 表示边界框; G 表示真实目标; T_+ 和 T_- 表示 IoU 正负阈值, 1, 0, -1 表示正、负、忽略样本; I_{IoU} 表示交并比。在基线模型的第 2 个阶段, $T_+ = 0.5$, $T_- = 0.5$, 即正负样本的定义基本是手工设置。

分类器的关键任务是对正负样本进行区分, 采用不同的 IoU 阈值进行训练会对分类器的性能产生显著影响。为了实现高质量目标检测, 必须对具有高 IoU 阈值的分类器进行训练。随着训练的进行, 候选目标区域的质量逐渐提高, 这种特性可以通过渐进式训练策略来实现。初始阶段, RPN 无法生成高质量候选目标区域, 因此适用于较低的 IoU 阈值, 更好适应这些质量较差的候选目标区域。随着训练不断进行, 候选目标区域质量逐渐提高, 逐渐出现了高质量候选区域, 同时, 通过不断增加 IoU 阈值, 更好地利用这些高质量候选区域, 提高检测器性能, 尤其在高 IoU 阈值条件下, 其判别性能更加显著。本文添加的动态标签分配模块如图 7 所示。

根据式(8), 动态标签分配模块可以表示为:

$$L_{\text{Label}} = \begin{cases} 1, & \text{if } \max I_{\text{IoU}} \geq T_{\text{now}} \\ 0, & \text{if } \max I_{\text{IoU}} < T_{\text{now}} \end{cases} \quad (9)$$

式中: T_{now} 表示当前的 IoU 阈值。考虑到训练过程中的动态因素, 目标候选区域的分布是随时间变化的, 动态标签分配模块会根据目标候选区域的统计数据自动更新 T_{now} , 以适应这种分布变化。

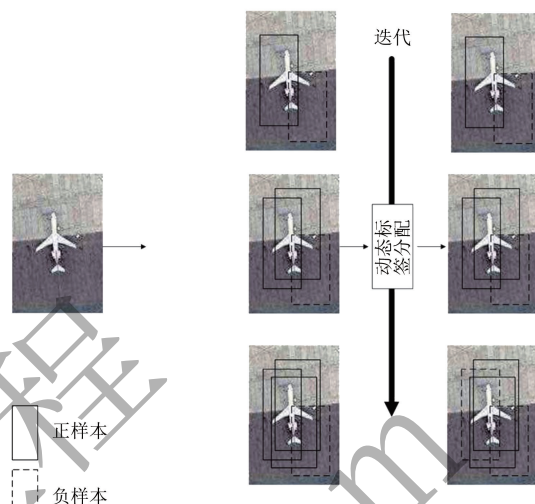


图 7 动态标签分配流程

Fig.7 Procedure of dynamic label assignment

2.3.2 回归损失函数

回归器与分类器具有类似的问题。在训练过程中, 尽管 DLA 提高了候选目标区域的质量, 但损失函数的固定性质使得对高质量候选目标区域的训练不足。本文引入动态 SmoothL1 损失函数代替 L1Loss 用于回归分支, 更好地适应回归标签分布的变化, 确保高质量样本对整体训练的贡献。基线模型所使用的 L1Loss 表示为:

$$L_1(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (10)$$

式中: y_i 代表样本 i 的真实标签; $\hat{y}_i$ 表示模型对样本的预测标签; N 代表样本个数。L1Loss 直接计算预测值与实际值之间的差异, 但在中心点不能求导。当损失值较小时, 得到的梯度相对较大, 可能导致模型震荡, 不利于收敛。回归误差与 L1Loss 的曲线如图 8 所示。

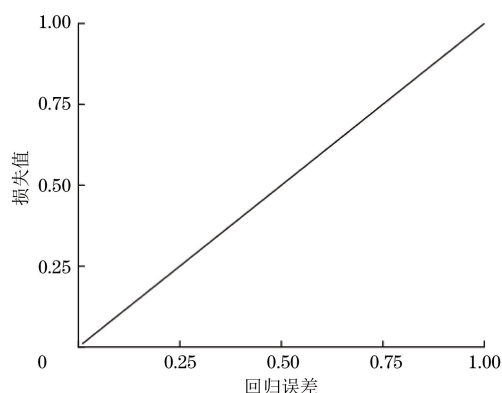


图 8 L1Loss 与回归误差曲线

Fig.8 Curve of L1Loss and regression error

在遥感图像中目标尺度差异较大, SmoothL1 损失函数在回归标签值比较大时, 梯度为常数; 当回归标签值较小时, 梯度会动态减小。本文在回归分

支中采用 SmoothL1 损失函数,可表示为:

$$\text{SmoothL1}(x, \beta) = \begin{cases} 0.5|x|^2/\beta, & \text{if } |x| < \beta \\ |x| - 0.5\beta, & \text{otherwise} \end{cases} \quad (11)$$

式中: x 代表回归标签; β 为超参数,用于控制在哪个范围内使用较为平缓的损失函数。图 9 所示为不同 β 值的 SmoothL1 损失与回归误差曲线。考虑到训练的鲁棒性, $\beta = 1.0$ 加速回归误差的收敛,防止在早期训练阶段因网络性能较差而导致损失过大。

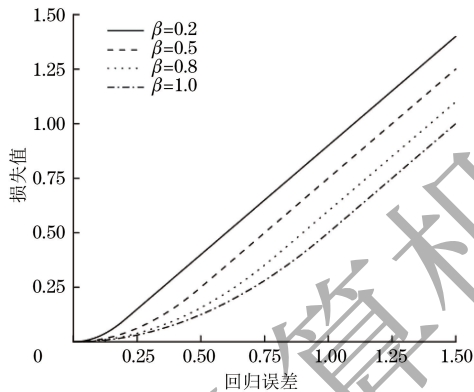


图 9 不同 β 值的 SmoothL1 损失与回归误差曲线
Fig.9 Curves of SmoothL1 loss and regression error for different β values

为了应对标签分布变化以及高质量样本的影响,本文引入动态 SmoothL1 Loss(DSL, L_{DSL}),逐渐调整损失函数的形式。这个调整使损失函数在训练过程中更加关注高质量样本:

$$L_{\text{DSL}}(x, \beta_{\text{now}}) = \begin{cases} 0.5|x|^2\beta_{\text{now}}, & \text{if } |x| < \beta_{\text{now}} \\ |x| - 0.5\beta_{\text{now}}, & \text{otherwise} \end{cases} \quad (12)$$

与动态标签分配相似,DSL 会根据回归标签的统计信息变化来调整 β_{now} 的值,反映目标位置的精度。具体而言,先获取候选提议框及它们对应目标真实标签之间的回归标签 E ,然后在这些回归标签中选择第 K_{β} 个最小值,将其用于更新 SmoothL1 Loss 中 β_{now} 。 K_{β} 值是在每 C 个迭代周期中,通过计算每个批次中回归标签第 K_{β} 最小标签的中位数确定的。选择中位数的原因在于回归标签可能存在一些异常值,DSL 通过动态调整的方式能够自动选择适当的参数,以更好地训练样本,从而训练出高质量的回归器。

3 实验结果与分析

3.1 实验环境与参数设置

网络实验环境为 Windows 11、Python 3.8、PyTorch 2.8.2 和 MMDetection 2.26.0,相关硬件配置信息如表 1 所示。

表 1 实验相关硬件配置

Table 1 Experimental related hardware configuration

名称	配置
GPU	GTX 1660 Ti
CPU	i7-9750H
CUDA	11.3
CuDNN	8.3.2
RAM/GB	16
Hard disk/TB	1

在实验过程中,所有模型采用统一训练策略,将固定步长衰减作为学习率衰减策略,其他具体参数设置如表 2 所示。

表 2 实验参数设置

Table 2 Experimental parameter settings

名称	配置
Batch Size	32
Epoch	50
初始学习率	0.0001
优化器	AdamW
权重衰减	0.01

3.2 实验数据集

本实验采用由武汉大学标注的遥感图像目标检测 RSOD 数据集^[18-19]。数据集包含飞机、油箱、立交桥和运动场,如图 10 所示。数据集按照 7:1:2 划分训练集、测试集和验证集。

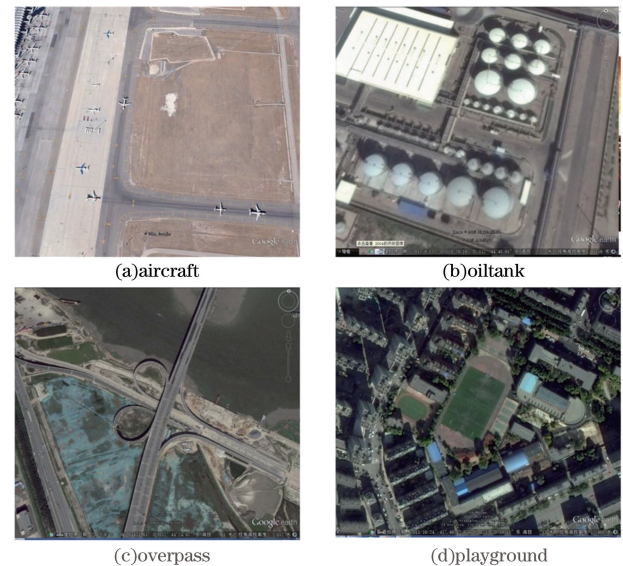


图 10 RSOD 数据集示例

Fig.10 Examples of RSOD dataset

数据集中目标标签数量如图 11 所示。

遥感图像通常具有有限的标记和广泛的多样性数据的独特性。在训练过程中,本文采用一系列数据增强方式,包括随机翻转、多尺度训练、像素填充和重复训练。数据增强策略如表 3 所示。

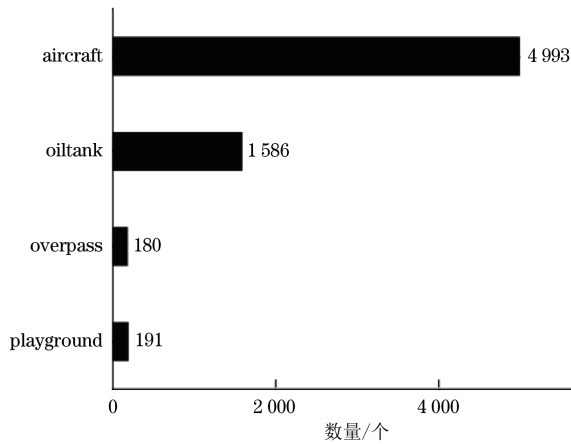


图 11 目标种类标签数量

Fig. 11 Number of object class labels

表 3 数据增强策略

Table 3 Strategy of data augmentation

增强方式	增强效果
随机翻转	50%概率对样本随机水平翻转
多尺度训练	保持样本纵横比,将样本尺度缩放 (1 333,640),(1 333,800)指定尺度
像素填充	对样本进行填充,确保尺寸可以被 32 整除
重复训练	对数据集样本重复训练 3 次

数据增强的目的是通过多样性变换和噪声的引入来丰富训练数据,增加数据的多样性,在提升模型鲁棒性的同时解决数据量过少的问题。通过数据增强可以减少过拟合风险,每次输入到模型的图像都是增强后的图像,从而使得模型在面对陌生遥感场景时表现出更好的泛化能力。

3.3 模型评价指标

为了验证改进后模型的性能,本文采用多个广泛使用的评价指标,这些指标包括召回率(R)、平均精度(AP, A_{AP})、平均精度均值(mAP, m_{mAP})和平均召回率(AR, A_{AR})^[20],计算式如下:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (13)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (14)$$

$$A_{AP} = \int_0^1 P(R) dR \quad (15)$$

表 4 骨干网络对模型性能的影响

Table 4 Impact of backbone network on model performance

模型	骨干网络	参数量/ 10^6	FLOPs/ 10^9	FPS/(帧· s^{-1})	mAP/%
Faster R-CNN	ResNet 50	41.14	203.43	8.7	93.7
	Swin Transformer	44.76	103.00	6.3	96.6

从表 4 可以看出,在 Faster R-CNN 采用

$$m_{mAP} = \frac{\sum_{n=1}^N A_{AP}(n)}{N} \quad (16)$$

$$A_{AR} = 2 \int_{0.5}^1 R(o) do = \frac{2}{n} \sum_{i=1}^n \max(I_{IoU}(G_{GT_i}) - 0.5, 0) \quad (17)$$

式中: N_{TP} 表示模型正确检测到的真实目标数量; N_{FP} 表示模型错误报告了存在目标的情况,实际上没有目标; N_{FN} 表示模型没有检测到目标,而实际上存在目标; o 为 IoU 重叠部分; n 为每张图片中所有标注框与 G_{GT} 接近的检测结果框的重叠部分数量,即 COCO 指标中的最大检测数(maxDets),本文若无特殊说明,默认 maxDets 为 1 000; P 为精确率。式(17)表示 IoU 阈值为 0.5~1 之间的平均召回率。

除此之外,还遵循标准的 COCO 评估指标,包括 $mAP@0.5$ (IoU 阈值为 0.5)、 $mAP@0.75$ (IoU 阈值为 0.75)和 $mAP@0.5:0.95$ (不同 IoU 阈值下的综合表现)。本文若无特殊说明,mAP 默认为 $mAP@0.5$ 。为了更好地验证改进效果,还引入了 $mAP@s$ (面积小于 32×32 像素)、 $mAP@m$ (面积大于 32×32 像素,小于 96×96 像素)和 $mAP@l$ (面积大于 96×96 像素)。

此外,为了验证改进前后模型的大小、计算量和推理时间,综合判别模型性能,本文引入参数量(Params)衡量模型的大小,每秒浮点运算次数(FLOPs)衡量模型的复杂度,FPS 即每秒内可以处理的图片数量来评估检测速度。

3.4 结果分析

3.4.1 Swin Transformer 改进实验

为了验证 Swin Transformer 替代 ResNet 50 作为骨干网络能更好地适应复杂遥感图像场景特征提取能力,降低误检、漏检风险,本文采用 Swin Transformer 和 ResNet 50 分别作为骨干网络进行对比实验,比较 mAP、参数量、FPS 和 FLOPs 4 个指标。其深度和阶段采用{2,2,6,2}配置,数字表示每阶段的堆叠块数量,窗口大小默认设置为 7,多层感知机的膨胀层 $\alpha=4$,嵌入维度设置为 96。不同骨干网络对模型性能的影响如表 4 所示。

Swin Transformer 作为骨干网络后,参数量增加

3. 62×10^6 , 但是 FLOPs 大幅度减少, FPS 减少 2.4, mAP 提高 2.9 百分点。分析认为有 3 个原因: 1) Swin Transformer 使用多层 Transformer 模块, 每个模块包含多头注意力机制和前馈神经网络, 在增加网络复杂性的同时增大了参数量; 2) Swin Transformer 使用自注意力机制来捕捉图像中的长距离依赖关系, 自注意力只考虑像素之间的相关性, ResNet 50 只在整个区域进行操作,

前者相较于后者计算更加高效; 3) 在遥感图像复杂场景中, Swin Transformer 的自注意力机制对于捕捉关键信息非常有帮助, 而 ResNet 50 对长距离依赖关系的建模能力较弱, 从而导致目标错检或漏检。

3.4.2 特征融合网络改进实验

相较于单独的 FPN, 添加 BFP 作为增强特征融合的补充, 添加模块后模型的性能如表 5 所示。

表 5 平衡特征金字塔对模型性能的影响

Table 5 Impact of balanced feature pyramid on model performance

模型	特征融合网络	参数量/ 10^6	FLOPs/ 10^9	FPS/(帧 $\cdot$ s $^{-1}$)	mAP/%
Faster R-CNN	FPN	41.14	203.43	8.7	93.7
	FPN+BFP	41.40	204.53	8.3	93.9

从表 5 可以看出, 在加入 BFP 模块后 mAP 提高了 0.2 百分点, 证明该模块有助于减少高层和底层之间信息的丢失, 保留更多有用信息。基于此, 本文通过对比数据集中各类目标的 AP 值验证 BFP 模块的有效性, 如表 6 所示。

从表 6 可以看出, 在加入平衡特征金字塔模块

后, 飞机、油桶和操场的 AP 值平均提高约 0.7 百分点, 立交桥的 AP 值提高了 6.2 百分点。立交桥具有多尺度、复杂结构的特点, 无法捕捉全部特征。添加平衡特征金字塔模块来增强语义信息, 有助于更准确地区分背景和目标, 减少尺度差异引起的误检和漏检问题。

表 6 平衡特征金字塔模块对各类目标平均精度的影响

Table 6 Impact of balanced feature pyramid module on average precision for various objects

模型	特征融合网络	AP				mAP@0.5:0.95/%
		飞机	油桶	立交桥	操场	
Faster R-CNN	FPN	60.6	67.8	13.5	69.9	61.6
	FPN+BFP	61.4	68.7	19.7	70.3	62.4

3.4.3 动态权重机制改进实验

在模型训练过程中固定网络设置极大地影响了性能, 固定标签分配策略和回归损失函数不能适应候选框分布变化, 不利于训练高质量的模型。添加动态权重机制可以动态调整标签分配和回归函数, 从而更好地利用训练样本, 使得模型适应更多高

IoU 阈值的样本, 如表 7 所示。

从表 7 可以看出, 加入动态权重机制后, mAP@0.5 反而下降 0.2 百分点。为了验证动态权重机制在训练过程中更关注高质量候选框, 本文引入 IoU 阈值为 0.5、0.75 和 0.5:0.95 3 个 mAP 指标进行验证, 如表 8 所示。

表 7 动态权重机制在训练过程中对模型性能的影响

Table 7 The impact of dynamic weighting mechanism on model performance during training

模型	动态权重机制	参数量/ 10^6	FLOPs/ 10^9	FPS/(帧 $\cdot$ s $^{-1}$)	mAP@0.5/%
Faster R-CNN		41.14	203.43	8.7	93.7
	✓	41.14	203.43	9.0	93.5

表 8 动态权重机制在不同 IoU 阈值的平均精度

Table 8 The average precision of dynamic weighting mechanism at different IoU thresholds

模型	动态权重机制	mAP@0.5	mAP@0.75	mAP@0.5:0.95
Faster R-CNN		93.7	71.9	61.6
	✓	93.5	74.6	63.0

从表 8 可以看出, 引入动态权重机制后, 随着 IoU 阈值不断增加 模型在高 IoU 阈值下的 mAP 指标也在增加。引入动态权重机制的 mAP@0.5:0.95 提高了 1.4 百分点。

在同一参数设置下, 使用 DSL 和 L1Loss 两种不同回归损失函数, 对 Faster R-CNN 的总损失函数收敛性进行验证。两种 Loss 函数随着迭代次数的变化

曲线如图 12 所示。其中两条曲线分别代表模型在不同回归损失函数 DSL 和 L1Loss 的总损失情况。

从图 12 可以看出,随着迭代次数的增加,DSL 和 L1Loss 最终都处于收敛状态。训练初期,DSL 相较于 L1Loss 下降得较为平缓,而且损失值更小。因此,使用 DSL 作为本文数据集的回归损失函数更有利于训练样本。

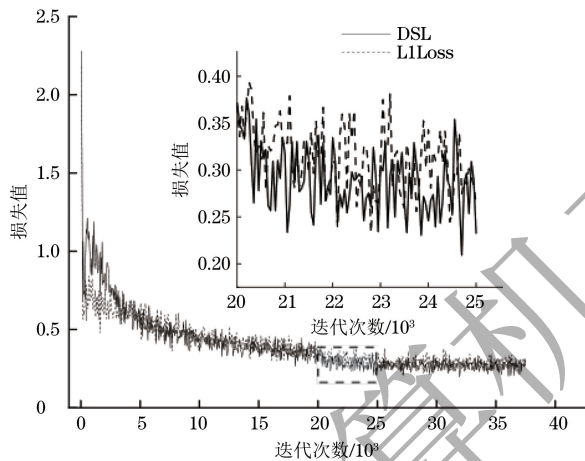


图 12 Loss 函数随着迭代次数的变化曲线

Fig. 12 The change curves of the Loss function with the number of iterations

3.4.4 模块消融实验

为验证各个模块之间的有效性,本文在公开遥

感图像 RSOD 数据集上进行消融实验。本文在 Faster R-CNN 基础上依次改进骨干网络 Swin Transformer、添加 BFP、动态权重机制(Dynamic)。消融实验结果如表 9 所示。

从表 9 可以看出,在 Faster R-CNN 基础上依次加入 Swin Transformer、BFP、Dynamic 与 Faster R-CNN 相比 AR 提高了 10.6 百分点。本文所提的 FSBD R-CNN 在不同 IoU 阈值召回率的性能较优。小目标的 mAP@s 在经过 BFP 模块和 Dynamic 后与 Faster R-CNN 相比下降了 1.2 百分点,说明在添加 BFP 模块后特征融合依然会丢失部分语义信息。动态权重机制对于模型的复杂度和计算量没有很大的影响,说明该模块关注模型训练过程。在 IoU 阈值较高的情况下,本文所提的模型性能较好,与 Faster R-CNN 相比 mAP@0.5:0.95 提高了 10.7 百分点。与 Faster R-CNN 模型相比,FLOPs 降低了约 100×10^9 ,主要归因于 Swin Transformer 对参数效率的改进以及在处理长距离依赖关系的图像区域时具有卓越性能。但是,在输入数据相同情况下,本文提出模型的 FPS 下降约 2.5。主要因为 Swin Transformer 基于注意力机制,与传统卷积网络相比需要更多的内存与计算能力,导致计算复杂度增大。

表 9 各模块消融实验结果

Table 9 Results of ablation experiments for each module

Faster R-CNN	Swin Transformer	BFP	Dynamic	AR/%	mAP@s/%	mAP@m/%	mAP@l/%	mAP@0.5/%	mAP@0.75/%	mAP@0.5:0.95/%	参数量/ 10^6	FLOPs/ 10^9	FPS/(帧·s ⁻¹)
✓				67.7	37.2	62.1	67.0	93.7	71.9	61.6	41.14	203.43	8.7
✓	✓			73.9	46.6	67.3	77.4	96.6	84.0	69.3	44.76	103.00	6.3
✓		✓		68.0	36.2	63.0	68.1	93.9	73.3	62.4	41.40	204.53	8.3
✓			✓	68.2	33.1	64.0	68.1	93.5	74.6	63.0	41.14	203.43	9.0
✓	✓	✓		74.0	45.2	67.9	78.1	96.6	84.3	69.7	45.02	104.11	6.4
✓	✓		✓	73.9	44.5	68.6	78.4	95.5	85.1	70.1	44.76	103.00	6.6
✓		✓	✓	68.7	36.0	64.4	69.5	93.1	75.5	64.1	41.40	204.53	8.3
✓	✓	✓	✓	78.3	46.4	70.8	80.3	96.5	87.1	72.3	45.02	104.11	6.2

3.4.5 数据增强实验

本文所采用数据增强将训练集重复 3 次输入到模型中,在数值上不会增加数据样本量,在特征上增加了数据量,在采用数据增强操作时,输入图像会依次执行数据增强函数内的序列操作,其中有些操作是随机的,按照一定概率执行或不执行,执行完数据增强后输出的仍是一张图像。训练集、测试集和验证集的比例仍保持 7:2:1。数据增强实验结果如表 10 所示。

表 10 数据增强实验结果

Table 10 Experimental results of data augmentation %

模型	数据增强	mAP@0.5	mAP@0.5:0.95
FSBD R-CNN	✓	96.5	72.3
		93.7	69.5

从表 10 可以看出,FSBD R-CNN 模型使用数据增强前后 mAP@0.5 提高了 2.8 百分点,说明数据增强对模型在遥感图像目标识别场景下具有良好的泛化性。

3.5 对比实验

3.5.1 预测结果分析

本文以 Faster R-CNN 为基线模型,在 RSOD 数据集上进行实验,验证所提方法的有效性,并与其他模型进行对比。本文对 YOLO 系列模型采用 200 Epoch 进行训练。本文所涉及的对比模型为 ATTS^[21]、Cascade R-CNN^[22]、Deformable DETR^[23]、FCOS^[24]、PAA^[25]、VarifocalNet^[26]、

YOLOF^[27]、YOLOX^[28]、DINO^[29]、Align-DETR^[30]。表 11 显示了不同模型在 RSOD 数据集上各类别的检测精度结果,加粗表示最优数据。本文所提的模型在各个类别中具有最佳的 mAP。相比基线模型,本文模型在飞机类别上的 mAP@0.5 提高 8.6 百分点,油桶的 mAP@0.5 提高 12.2 百分点,立交桥的 mAP@0.5 提高 24.1 百分点,运动场的 mAP@0.5 提高 17.6 百分点。

表 11 在 RSOD 数据集上各类别目标模型的评价指标对比

Table 11 Comparison of evaluation indicators for various category target models on the RSOD dataset						%
模型	mAP@0.5				mAP@0.5:0.95	
	飞机	油桶	立交桥	运动场		
ATTS	46.6	51.8	17.0	67.6	47.7	
Cascade R-CNN	55.8	60.9	2.4	56.5	55.7	
Deformable DETR	39.3	54.5	16.6	59.1	45.3	
FCOS	42.2	50.0	17.1	61.9	43.8	
PAA	45.5	52.2	15.9	70.6	46.8	
VarifocalNet	43.1	50.4	20.5	62.8	45.0	
YOLOF	25.3	40.4	14.5	57.6	34.5	
YOLOX	60.7	61.4	19.6	56.2	49.5	
DINO	32.6	34.9	14.8	59.9	35.5	
Align-DETR	54.2	59.6	15.0	77.4	51.4	
Faster R-CNN	60.6	67.8	13.5	69.9	61.6	
FSBD R-CNN	69.2	80.0	37.6	87.5	72.3	

本文所提模型在 RSOD 数据集 4 个类别中的检测精度都达到了最佳水平。在密集场景中,如飞机和油桶,FSBD R-CNN 的评价指标取得了最佳性能。此外,立交桥、运动场中大型目标上与其他模型

相比,FSBD R-CNN 的评价指标也有着显著提升。通过对比实验结果,进一步验证了本文所提模型的有效性和鲁棒性。在 RSOD 数据集上不同模型的评价指标对比如表 12 所示。

表 12 在 RSOD 数据集上不同模型的评价指标对比

Table 12 Comparison of evaluation metrics for different models on the RSOD dataset								
模型	mAP@0.5/	mAP@0.75/	mAP@0.5:0.95/	mAP@s/	mAP@m/	mAP@l/	参数量/	FLOPs/
	%	%	%	%	%	%	10 ⁶	10 ⁹
ATTS	84.1	49.2	47.7	12.9	49.3	54.1	31.89	206.30
Cascade R-CNN	88.3	63.4	55.7	29.2	57.4	59.2	68.89	200.47
Deformable DETR	80.3	40.0	45.3	8.4	43.0	54.9	39.82	112.61
FCOS	82.7	40.4	43.8	7.6	45.6	50.7	31.84	206.67
PAA	83.2	47.4	46.8	11.6	48.6	54.3	31.89	206.30
VarifocalNet	81.4	44.9	45.0	6.3	46.6	52.0	32.49	198.59
YOLOF	70.7	27.6	34.5	1.9	26.0	42.8	42.13	103.14
YOLOX	85.9	55.8	49.5	36.0	61.3	53.3	13.32	8.94
DINO	64.9	34.3	35.5	11.0	34.0	39.4	47.55	215.04
Align-DETR	80.1	57.2	51.4	28.3	53.9	57.0	47.49	215.04
Faster R-CNN	93.7	71.9	61.6	37.2	62.1	67.0	41.14	203.43
FSBD R-CNN	96.5	87.1	72.3	46.4	70.8	80.3	45.02	104.11

在 RSOD 数据集上,FSBD R-CNN 在 mAP 中取得了最高的性能,mAP@0.5 和 mAP@0.75 获

得较优的检测效果,参数量相较于基线模型小幅度增长,mAP@s、mAP@m 和 mAP@l 与 Faster

R-CNN 相比分别提高 9.2、8.7 和 13.3 百分点。与其他模型相比,FSBD R-CNN 的参数量与 FLOPs 没有达到理想的效果。

3.5.2 基线模型与 FSBD R-CNN 的对比

改进前后网络模型对于遥感图像目标检测的 P - R 曲线如图 13 所示。 P - R 曲线与坐标轴围成的面积(AUC- P - R)表示模型在 IoU 阈值为 0.5 的综合性能。从图 13 可以明显看出,FSBD R-CNN 在 RSOD 数据集的检测中取得较好的性能,AUC- P - R 明显高于基线模型。

为了直观验证所提模型的有效性,本文给出了各类场景的检测结果。本文模型能够有效地检测大尺度和语义信息不明显的目标,并有效解决基线模

型在定位准确性和错检问题上,如图 14 所示。

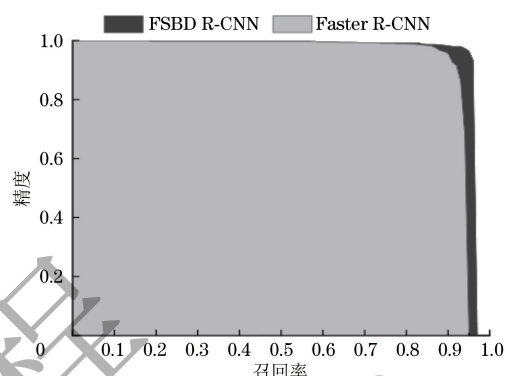


图 13 改进前后网络模型 P - R 曲线对比

Fig. 13 Comparison of P - R curves of network models before and after improvement



图 14 不同模型对语义信息不明显物体的检测结果

Fig. 14 The detection results of objects with unclear semantic information by different models

目标漏检错检结果对比如图 15 所示。从图 15(b)可以看出,基线模型虽然检测出目标物体,但未能检测出图像边缘物体,并将非飞机目标打标签,出现了错检和漏检情况,检测操场时

同样出现漏检。从图 15(c)可以看出,本文所提模型能够准确地检测出物体并检测到在遥感图像边缘仅表现出部分特征的物体,并优化了目标漏检情况。

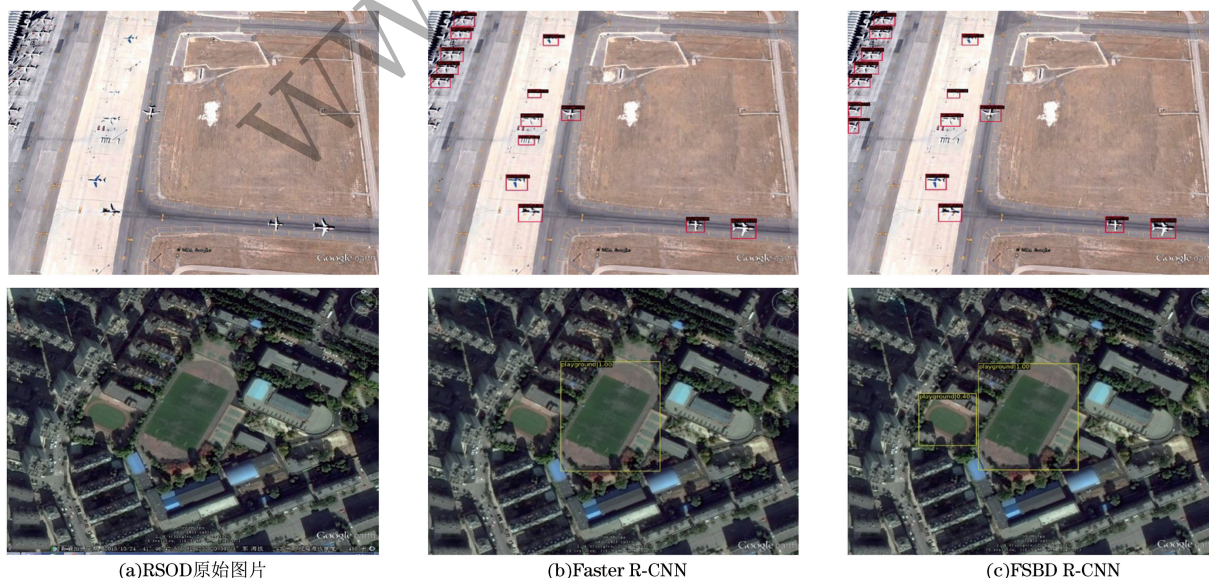


图 15 目标漏检错检结果对比

Fig. 15 Comparison of target missed detection and false detection results

本文所提出的模型能够检测到更多复杂背景物体,而且在检测小物体方面明显优于基线模型。复

杂背景检测结果对比如图 16 所示。从图 16(b)可以看出,基线模型检测复杂背景出现漏检、错检。从

图 16(c)看出,本文所提出的模型缓解了基线模型 错检和漏检问题。

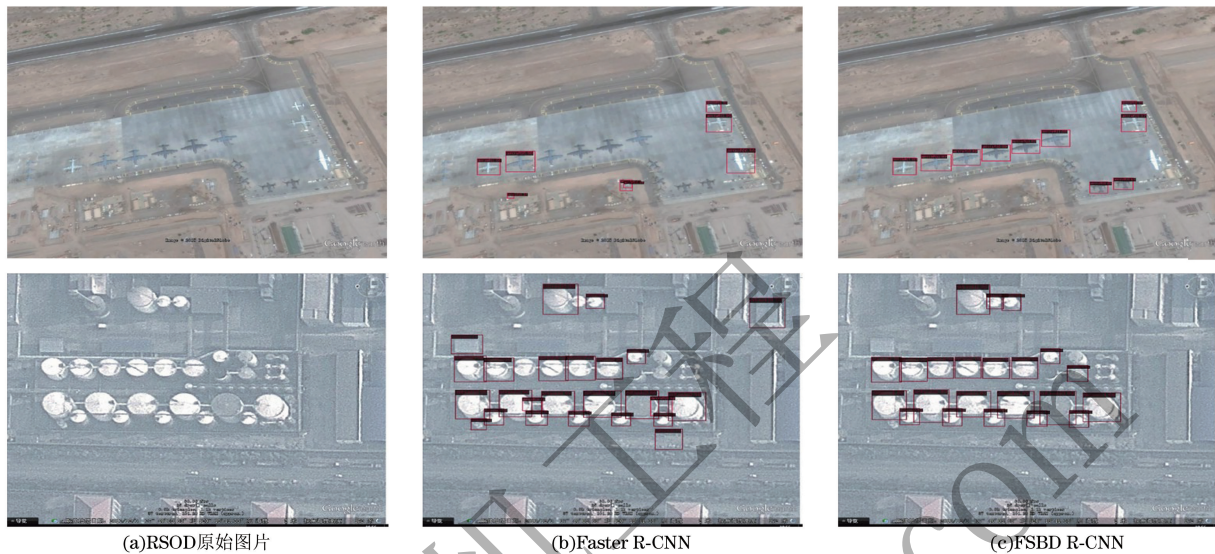


图 16 复杂背景检测结果对比

Fig. 16 Comparison of detection results in complex backgrounds

4 结束语

针对遥感图像目标检测中漏检、误检问题,本文提出一种改进的 Faster R-CNN 多类别遥感图像目标检测方法。以 Faster R-CNN 为基线模型,通过将原网络中骨干网络 ResNet 50 替换为 Swin Transformer,增强模型处理输入图像中目标不同位置之间的关联,帮助模型理解不同目标之间的语义关系。在分类和回归分支添加动态权重机制,增强模型对高质量候选框的关注。实验结果表明,与基线模型相比,本文模型的 mAP@0.5:0.95 提升了 10.7 百分点,平均召回率提升了 10.6 百分点,有效缓解了基线模型漏检和错检问题。在公开 RSOD 数据集上与其他模型相比,本文所提的 FSBD R-CNN 平均精度较高,在复杂场景下具有较优的鲁棒性。本文所提出的 FSBD R-CNN 模型优化了遥感图像中多目标检测出现的漏检和错检相关问题,但是参数量较大,在资源受限的场景下难以高效运行,除此之外,在排列密集的目标中,受检测框重叠的影响,导致输出后的检测结果可读性较差。下一步将模型整体轻量化,并在模型中添加旋转框,在减少规模参数量的同时保持较好性能,解决模型在资源受限环境下高效运行以及检测结果可读性相关问题。

参考文献

- [1] 李振鲁,黄威,孙错. 复杂环境下的轻量化道路目标识别算法研究[J]. 计算机工程, 2024, 50(4): 219-227.
LI Z L, HUANG W, SUN K. Research on lightweight road-target-recognition algorithm in complex environment [J]. Computer Engineering, 2024, 50(4): 219-227. (in Chinese)
- [2] 周金涛,高迪驹,刘志全. 基于全景视觉的无人船水面障碍物检测方法[J]. 计算机工程, 2024, 50(2): 113-121.
ZHOU J T, GAO D J, LIU Z Q. Detection method of water-surface obstacles for unmanned ships based on panoramic vision [J]. Computer Engineering, 2024, 50(2): 113-121. (in Chinese)
- [3] 蒋心璐,陈天恩,王聪,等. 大田环境下的农业害虫图像小目标检测算法[J]. 计算机工程, 2024, 50(1): 232-241.
JIANG X L, CHEN T E, WANG C, et al. Small object detection algorithm for agricultural pest images in field environments [J]. Computer Engineering, 2024, 50(1): 232-241. (in Chinese)
- [4] 管嘉程,任红卫,周宋佳. 基于 YOLOv5 改进的轻量化目标检测[J]. 计算机系统应用, 2023, 32(9): 132-142.
GUAN J C, REN H W, ZHOU S J. Lightweight object detection based on YOLOv5 improvement [J]. Computer System Applications, 2023, 32(9): 132-142. (in Chinese)
- [5] 沙苗苗,李宇,李安. 改进 Faster R-CNN 的遥感图像多尺度飞机目标检测[J]. 遥感学报, 2022, 26(8): 1624-1635.
SHA M M, LI Y, LI A. Improving Faster R-CNN for multi-scale aircraft target detection in remote sensing images [J]. Journal of Remote Sensing, 2022, 26(8): 1624-1635. (in Chinese)
- [6] 曲海成,王蒙,柴蕊. 双向多尺度特征融合的高效遥感图像车辆检测[J]. 计算机工程与应用, 2024, 60(12): 346-356.
QU H C, WANG M, CHAI R. Efficient vehicle detection in remote sensing images with bidirectional multi-scale feature fusion [J]. Computer Engineering and Applications, 2024, 60(12): 346-356. (in Chinese)
- [7] 戚玲珑,高建瓴. 基于改进 YOLOv7 的小目标检测[J]. 计算机工程, 2023, 49(1): 41-48.
QI L L, GAO J L. Small object detection based on improved YOLOv7 [J]. Computer Engineering, 2023, 49(1): 41-48. (in Chinese)
- [8] 梁嘉杰,李星星. 特定任务上下文解耦的遥感图像目标检测方法[J]. 计算机工程与应用, 2025, 61(2): 293-303.
LIANG J J, LI X X. Task-specific context decoupling object detection method for remote images [J]. Computer Engineering and Applications, 2025, 61(2): 293-303. (in Chinese)

- [9] 王龙博, 刘建辉, 张贝贝, 等. 利用注意力机制融合的 YOLOv5 遥感图像目标检测[J]. 信息工程大学学报, 2023, 24(4): 438-446.
- WANG L B, LIU J H, ZHANG B B, et al. Object detection in YOLOv5 remote sensing image using attention mechanism fusion [J]. Journal of University of Information Engineering, 2023, 24 (4): 438-446. (in Chinese)
- [10] 左露, 牛晓伟, 朱春惠, 等. 基于改进 ConvNeXt 的遥感图像目标检测算法[J]. 电光与控制, 2024, 31(2): 46-51, 91.
- ZUO L, NIU X W, ZHU C H, et al. Remote sensing image object detection algorithm based on improved ConvNeXt [J]. Electronics Optics and Control, 2024, 31 (2): 46-51, 91. (in Chinese)
- [11] 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.
- [12] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of IEEE International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2017: 2980-2988.
- [13] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2016: 770-778.
- [14] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2017: 936-944.
- [15] LIU Z. Swin Transformer: hierarchical vision Transformer using shifted windows [C]// Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2021: 9992-10002.
- [16] PANG J M, CHEN K, SHI J P, et al. Libra R-CNN: towards balanced learning for object detection [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2019: 821-830.
- [17] ZHANG H K, CHANG H, MA B P, et al. Dynamic R-CNN: towards high quality object detection via dynamic training [C]// Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 260-275.
- [18] LONG L, GONG Y P, XIAO Z F, et al. Accurate object localization in remote sensing images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5): 2486-2498.
- [19] XIAO Z F, LIU W, TANG G F, et al. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images[J]. International Journal of Remote Sensing, 2015, 36(2): 618-644.
- [20] HOSANG J, BENENSON R, DOLLÁR P, et al. What makes for effective detection proposals? [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(4): 814-830.
- [21] ZHANG S F, C CHI, YAO Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2020: 9756-9765.
- [22] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 6154-6162.
- [23] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable Transformers for end-to-end object detection[EB/OL]. [2023-10-10]. <https://arxiv.org/abs/2010.04159?context=cs>.
- [24] TIAN Z, SHEN C H, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2019: 9626-9635.
- [25] KIM K, LEE H S. Probabilistic anchor assignment with IoU prediction for object detection [EB/OL]. [2023-10-10]. <https://arxiv.org/pdf/2007.08103>.
- [26] ZHANG H Y, WANG Y, DAYOUB F, et al. VarifocalNet: an IoU-aware dense object detector [C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2021: 8510-8519.
- [27] CHEN Q, WANG Y M, YANG T, et al. You only look one-level feature[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2021: 13034-13043.
- [28] ZHENG G, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. [2023-10-10]. <https://arxiv.org/pdf/2107.08430>.
- [29] ZHANG H, LI F, LIU S L, et al. DINO: DETR with improved denoising anchor boxes for end-to-end object detection[EB/OL]. [2023-10-10]. <https://arxiv.org/pdf/2203.03605>.
- [30] CAI Z, LIU S T, WANG G D, et al. Align-DETR: improving DETR with simple IoU-aware BCE loss[EB/OL]. [2023-10-10]. <https://arxiv.org/abs/2304.07527?context=cs>.

编辑 薛晋栋