

DOI:10.19651/j.cnki.emt.2211618

基于改进 YOLOX 的航拍遥感图像检测模型^{*}

左 露 牛晓伟 朱春惠

(重庆三峡学院电子与信息工程学院 重庆 404100)

摘要: 针对遥感图像中小目标众多,目标尺度变化剧烈,背景复杂所造成的目标检测精度低的问题,提出了一种基于改进 YOLOX 的目标检测算法,在 YOLOX 的基础上,首先在主干网络中加入注意力机制提高网络对遥感图像中小目标的感知能力,丰富语义信息;其次在特征融合部分中加入 MSCE-R 多尺度信息融合模块,通过融合不同尺寸的特征图,减少遥感图像因为尺度变化造成的图像细节信息损失;最后通过引入 CIoU 损失函数加快网络收敛速度,使其满足实时性的需求。本文将提出的检测算法在 RSOD 遥感数据集进行实验,平均检测准确率为 95.12%,相比于未做改进的 YOLOX,检测精度提高了 8.69%。实验结果证明,所提方法具有更高的检测精度。

关键词: 注意力机制;YOLOX;小目标;特征融合

中图分类号: TP751.1 **文献标识码:** A **国家标准学科分类代码:** 510.4

Aerial remote sensing image detection model based on improved YOLOX

Zuo Lu Niu Xiaowei Zhu Chunhui

(College of Electronics and Information Engineering, Chongqing Three Gorges College, Chongqing 404100, China)

Abstract: Aiming at the problem of low target detection accuracy caused by numerous small targets in remote sensing images, drastic changes in target scales and complex backgrounds, a target detection algorithm based on improved YOLOX is proposed. On the basis of YOLOX, firstly, an attention mechanism is added to the backbone network to improve the network's ability to perceive small targets in remote sensing images and enrich the semantic information; secondly, the feature fusion part is added to the MSCE-R multi-scale information fusion module in the feature fusion part to reduce the loss of image detail information caused by scale changes in remote sensing images through feature maps of different sizes; finally, the convergence speed of the network is accelerated by introducing CIoU loss function to make it meet the demand of real-time. In this paper, the proposed detection algorithm is experimented on the RSOD remote sensing dataset, and the average detection accuracy is 95.12%, which is 8.69% higher than that of the unimproved YOLOX. The experimental results prove that the proposed method has higher detection accuracy.

Keywords: attention mechanism;YOLOX;small target;feature fusion

0 引 言

随着目标检测技术不断地突破与改进,所涉及到的领域也越来越丰富,遥感图像目标检测作为其中的一项任务,因为图像的内容包括了飞机,车辆,建筑物,桥梁等,所以被广泛的应用在军事,灾害控制,城市建设规划等领域。遥感的图像种类繁多,并且数据量庞大,如何提高对遥感图像中检测目标的检测精度并投入到应用中,这是一个亟待解决的问题。

目前,基于深度学习的目标检测算法主要分为两类,一

类是基于目标候选框检测和卷积神经网络结合的双阶段目标检测算法,其中具有代表性的是 R-CNN 系列算法^[1-3]、Cascade R-CNN^[4]等;另一类是单阶段目标检测算法,比较具有代表性的是 YOLO 系列^[5],这类算法的特点是不需要计算候选框区域,直接通过主干网络得到目标分类和位置信息,并且图片只需要经过一次处理,检测速度相比于双阶段目标检测算法提升显著。这些算法均是基于卷积神经网络进行特征提取,而遥感图像的一个特点是小目标众多,图像经过多次的卷积和池化之后,虽然提取到的语义信息更为丰富,但是容易造成大量的小目标细节信息的丢失,在检

收稿日期:2022-10-07

^{*} 基金项目:国家重点研发计划课题(2021YFB3901405)、科技部专项课题(2021YFB3901400)、重庆市重点实验室开放基金(ZD2020A0301)项目资助

测的过程中容易出现错检、漏检等情况。为了提高对图像中小目标的检测精度, Lin 等^[6]提出了 FPN (feature pyramid networks), 对不同尺度的特征图进行特征融合, 最后再进行预测, 使小目标检测的精度有所提升; Pang 等^[7]则是通过提取四个不同尺度的特征图, 将其集成为顶层特征图后再与各不同尺寸特征图进行融合, 虽然达到提升小目标检测精度的目的, 但是结构复杂, 并且增加了计算量; 冯号等^[8]对 YOLOv3 算法提出改进, 通过优化置信度损失函数, 引入注意力机制加强模型对图像位置信息的特征提取, 提升对小目标的检测精度, 但是面对遥感图像中拥有复杂背景的目标时, 检测精度仍然不理想; 王成龙等^[9]针对目标检测算法参数量大, 检测速度慢等问题提出了一种实时的遥感图像检测算法, 结合轻量化卷积, 虽然有效的降低了模型参数量, 但是检测精度仍有待提高。

为了解决遥感图像中小目标较众多, 边缘特征难以提取, 且因多尺度变化更加剧烈, 前景和背景更加复杂导致的检测困难问题, 进一步提升模型的检测精度, 本文在基于 YOLOX-s 模型架构上做出改进, 主要改进点在于:

1) 在主干网络 CSPDarknet53 中引入注意力机制, 在减少信息弥散的情况下, 放大 3 个维度的交互作用, 提高对遥感图像中小目标的检测精度;

2) 改进模型中的 FPN 部分, 在特征金字塔中加入多尺度信息融合模块 (MSCE-R), 进一步改善遥感图像因尺度

变换所带来的影响, 提高对遥感目标的检测精度。

3) 对损失函数做出改进, 引入 CIoU 损失函数替换模型中 IoU 损失函数, 快速稳定边框回归, 加快收敛, 使模型满足实时性的需求。

1 YOLOX

YOLOX^[10]是旷视科技于 2021 年 8 月推出的一个新的单阶段目标检测算法, 比起之前的 YOLO 系列算法, YOLOX 做出的改进主要有 3 点, 第 1 点是使用了解耦头 (decoupled head), 在预测部分即输出端引入 Decoupled head 结构来加快网络收敛, 提高收敛速度; 第 2 点是在输入端使用 Mosaic 和 Mixup 的数据增强策略增加了数据的多样性, 丰富了图片的背景; 第 3 点是无锚点 (anchor free), 无锚机制所带来的效果丝毫不逊色于基于锚的目标检测算法, 并且有效的减小了参数量, 做出的改进让检测精度和检测速度都取得了不错的效果。YOLOX 整体结构可以分为 4 个部分, 分别是输入端, 主干网络, 特征融合网络以及预测端。在输入端部分, YOLOX 使用 Mosaic 和 Mixup 的数据增强策略对输入的图像进行预处理, 数据增强的效果如图 1 所示; 主干网络部分采用 CSPDarknet53 进行特征提取; 特征融合部分采用 FPN 结构; 在预测部分即输出端引入 3 个 Decoupled head 结构来加快网络收敛。

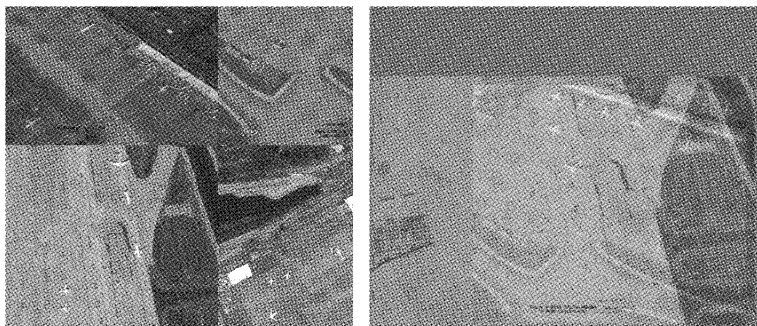


图 1 数据增强

2 改进策略

2.1 主干网络

作为在整个模型中极其重要的一环, 主干网络起到的作用是提取主要特征, 在 YOLOX 中使用 CSPDarknet53 作为模型的主干网络。CSPDarknet53 主要是由 1 个 Focus 模块跟 4 个 CSP 模块构成, CSP 模块将输入分成两个分支, 分别进行卷积操作, 其中一个分支进行多次卷积后跟另一部分进行拼接作为输出, 最大程度的让模型捕捉到更多的特征。遥感图像中小目标众多, 目标排列紧密, 因此很难提取到主要特征信息, 为了使主干网络在提取特征的过程中更加关注遥感图像中的目标, 提升检测精度, 在主干网络中加入一种全新的注意力机制, 全局注意力机

制 (global attention mechanism, GAM)^[11], 该注意力机制保留了通道和空间两个方面的信息, 通过减少信息和放大全局的交互作用来提高深度神经网络的性能, 具体算法流程如图 2 所示。

GAM 模块由通道注意力子模块与空间注意力子模块组成, 在对遥感目标进行感知时在通道和空间两个维度中生成相应的权重, 计算公式如下:

$$\begin{cases} F_2 = M_c(F_1) \otimes F_1 \\ F_3 = M_s(F_2) \otimes F_2 \end{cases} \quad (1)$$

第 1 个 stage 中的卷积模块输出的特征图为 $F_1 \in \mathbb{R}^{C \times H \times W}$, C 为特征图像的通道, H 及 W 为特征图中每个通道的高和宽, 特征图输入到 GAM 模块中, 首先在通道注意力模块中进行前向传播, 放大跨纬度空间依赖性并生成向

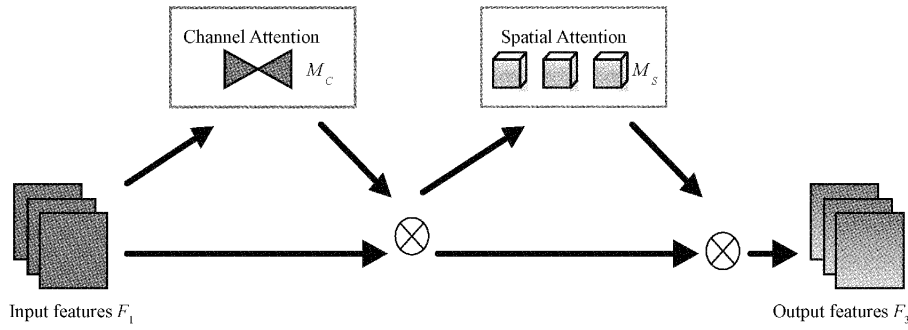


图2 注意力机制模块

量 M_c , 然后将该通道注意力向量与特征图 F_1 相乘生成通道权重特征 F_2 , 再将 F_2 输入到空间注意力模块进行空间维度计算生成空间矩阵 M_s , 最后将 F_2 与 M_s 相乘获取空间注意力权重并最终得到注意力映射 F_3 。

2.2 多尺度特征融合

在神经网络中,更深层次网络能提取到的语义信息往往更为丰富,但同时也会损失一部分的细节特征。为了弥补这部分损失,在此前很多的研究都会涉及到不同尺度的图像特征融合,以此来提高检测的精度。

在YOLOX中,特征融合部分采用的是特征金字塔(feature pyramid networks, FPN),如果输入为 640×640 , 则第1个特征图输出的是大小为 80×80 , 第2个是 40×40 , 最后一个为 20×20 , 由3个不同尺寸的特征图进行特征融合,使得最后进入预测端的特征图具有更丰富的语义信息,而遥感图像中很多目标的尺度变化相对剧烈,在经过主干网络中不断的卷积与池化之后,造成这类目标的细节与位置信息损失更为严重,并且因为遥感图像的背景复杂等问题,容易导致错检和漏检等情况发生,为了使模型更加适合遥感图像的检测,本文提出一种多尺度信息融合模块MSCE-R,在本文模型的Neck部分原有的FPN结构中加入这一模块,通过对不同尺寸的特征图进行信息融合,进一步的保留不同尺寸中目标的位置信息与细节特征,从而加强在FPN结构上的特征融合。MSCE-R模块的主要结构如图3所示。在MSCE-R模块中,输入的特征图首先分别由 1×1 卷积层进行处理以获得两个特征图。一个特征映射用传统的 3×3 卷积进行细化,而另一个特征映射被送入3个独立的并行非对称卷积分支进行多尺度特征提取。3个不对称卷积分支中使用的卷积核大小分别为3、5和7。该过程可表述为式(2)、(3):

$$f_1 = \text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(f)) \quad (2)$$

$$f_2 = A\text{Conv}_{3 \times 3}(\tilde{f}) + A\text{Conv}_{5 \times 5}(\tilde{f}) + A\text{Conv}_{7 \times 7}(\tilde{f}),$$

$$\tilde{f} = \text{Conv}_{1 \times 1}(f) \quad (3)$$

其中, f 表示输入特征映射, $\text{Conv}_{1 \times 1}$ 表示 1×1 卷积层, $\text{Conv}_{3 \times 3}$ 表示 3×3 卷积层, $A\text{Conv}_{i \times i}$ 是非对称卷积,其中 i 表示卷积核的大小。所有卷积层之后是批量归一化和

ReLU激活。除此之外还有一个通道加权分支,在该分支中,通过最大池化层和平均池化层的操作将输入特征映射压缩为一维向量。然后将这个一维向量送入由 1×1 卷积实现的全连接层,并进一步组合以生成注意力向量,该过程可由公式化表示为式(4):

$$\text{Atten} = \sigma(\text{fc}(\text{avgpool}(f)) + \text{fc}(\text{maxpool}(f))) \quad (4)$$

这里的 σ 表示sigmoid函数, fc 表示全连接层, avgpool 和 maxpool 分别是空间维度上的平均池和最大池。最后MSCE-R模块中的过程可以公式化表示为式(5):

$$F = \text{Atten} \odot (\text{Cat}(\text{Conv}_{1 \times 1}(f_1), \text{Conv}_{1 \times 1}(f_2))) \quad (5)$$

2.3 损失函数

损失函数反应的是在模型训练过程中,真实值与预测值之间的差异程度并对其做成分析与评价,YOLOX中所采用的损失函数主要包括3种:IoU损失(LossIoU),类别损失(Losscls),置信度损失(Lossobj)。其中IoU指的是真实框跟预测框之间的交并比,其计算公式如下所示。

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (6)$$

$$\text{Loss}_{\text{IoU}} = 1 - \text{IoU} \quad (7)$$

式中: A 为预测框, B 为真实框, $A \cap B$ 代表 A 与 B 的交集面积, $A \cup B$ 代表 A 与 B 的并集面积。IoU所考虑的是预测框与真实框之间的交并比,但是如果预测框与真实框直接之间没有交集,那IoU的计算结果就会一直为0,而且当2个框中一个在另一个内部时,如果框的大小不变,那么计算出的IoU值也不会发生变化,这样模型将难以优化。本文选用CIoU作为IoU损失,CIoU权衡了目标回归框中比较重要的3个点,分别是重叠面积、中心点距离和长宽比,从而具有更好的稳定的性能,由于考虑了检测目标与Anchor Box之间的距离、重叠率以及尺度等因素,使得目标框的回归更加稳定。CIoU损失函数的计算公式如下:

$$\text{CIoU} = \text{IoU} - \frac{P^2(b, b^{\text{gt}})}{c^2} - av \quad (8)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (9)$$

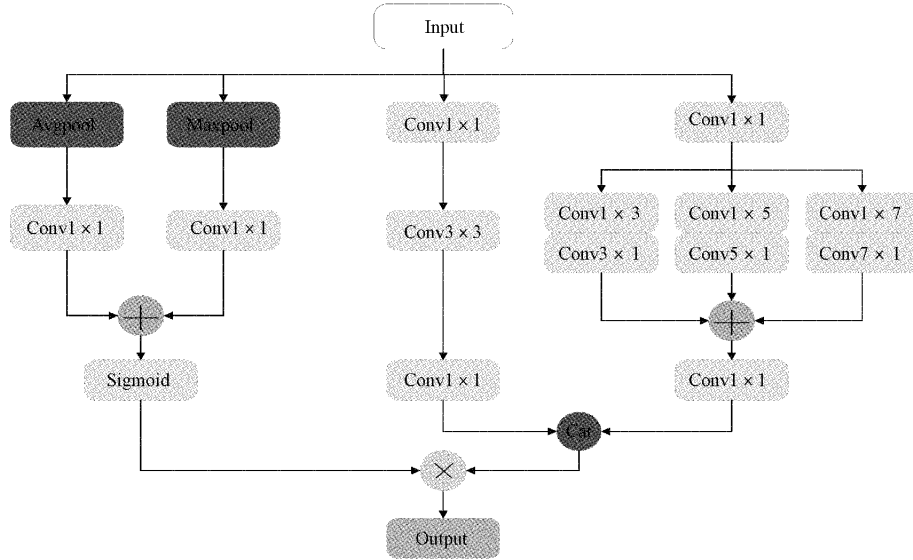


图 3 MSCE-R 多尺度信息融合模块

$$v = \frac{4}{\pi^2} (\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h})^2 \quad (10)$$

$$LossCIoU = 1 - CIoU \quad (11)$$

式中： b 和 b^{gt} 分别代表预测框和真实框的中心点， $d = \rho^2(b, b^{gt})$ 代表预测框和真实框的中心点的欧式距离， c 代表预测框与真实框的最小外接矩形的对角线距离， ω 和 ω^{gt} 分别代表预测框与真实框的宽度， h 和 h^{gt} 分别代表预测框与真实框的高度。

与 IoU 相比，CIoU 损失函数具有对称性与随机性，并且收敛速度更快；同时 CIoU 考虑了两个框之间没有交集的部分，能够更好地反映出 2 个 Anchor 框重叠的方式。

2.4 改进后的 YOLOX-s 模型结构

为了解决遥感图像中小目标多，目标排列紧密，目标尺度变化剧烈等问题，在 YOLOX-s 模型的基础上提出了 3 改进点，首先在 Backbone 中引入 GAM 注意力机制，保留图像信息以放大全局交互特征，提高模型对遥感图像中小目标的感知能力，其次在 Neck 部分的特征金字塔中加入多尺度信息融合模块 MSCE-R，通过对不同尺寸的特征图进行信息融合，增强特征图所具有的细节特征和位置信息，并以此提高对目标的检测精度；最后使用 CIoU 损失函数作为边框损失函数，提高模型的收敛速度，使检测效果更进一步，改进后的网络结构如图 4 所示。

3 实验结果分析

3.1 数据集介绍

本文实验所采用的数据集为由武汉大学于 2017 年发布的 RSOD 数据集，在 RSOD 数据集中原有 976 张光学遥感图像，提取自 Google Earth 和 Tianditu，以 PASCAL VOC 数据集的格式进行标注，数据集由飞机 (aircraft)，油箱 (oiltank)，立交桥 (overpass) 和操场 (playground) 4 个对

象类组成，共计 6 950 个目标。为了增加样本的数量，以旋转、缩小、放大的方式对数据集进行扩充，将数据集的数量增加到了 2 326 张。

3.2 实验环境及参数设置

实验操作系统为 Windows10，搭载的 CPU 版本为 12 核 Intel(R) Xeon(R) Gold 5320 CPU@2.20 GHz，GPU 为 NVIDIA GeForce RTX A4000，深度学习框架为 PyTorch 1.10.0。考虑实际硬件性能与对比实验，将超参数进行如下设置：epoch=100，batch-size=8，learning rate=0.01，学习衰减率为 5×10^{-4} ，优化器使用 SGD。

3.3 评价指标

本文选取了平均精准率 (average precision, AP)，平均精准率均值 (mean average precision, mAP) 以及检测速率 (frame per second, FPS) 作为评价指标。AP, mAP 这两个评价指标是对精准率 (Precision)，召回率 (Recall) 的综合考量，可以评估模型的有效性；FPS 指的是每秒检测的帧数，可以用来衡量模型的实时性。

精准率是指被正确预测为某类遥感目标占全部预测为该遥感目标的比例；召回率是指正确预测为某类遥感目标占实际上全部为该遥感目标的比例。两者表达式如式 (12) 和 (13)，其中 TP 指正确预测为正例；FP 指错误预测为正例；FN 指错误预测为负例。

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

平均精准率可以用来评估某一类遥感目标的预测效果，它是精准率-召回率曲线 (P-R) 曲线在区间 [0, 1] 内精准率对召回率的积分，其公式为式 (14)，其中 P 为 Precision, R 为 Recall。平均精准率均值能够衡量本文探

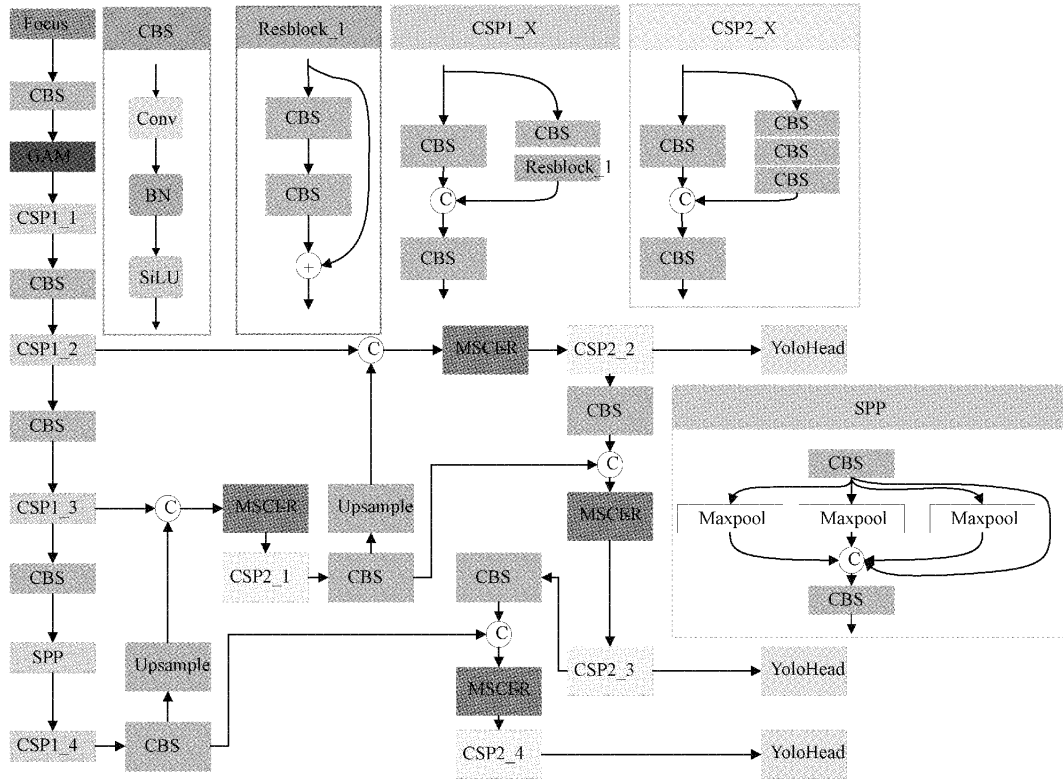


图4 YOLOX-s改进

讨的所有类遥感目标的检测效果,指的是4类遥感目标平均精准率的平均值。

$$AP = \int_0^1 p(r)dr \quad (14)$$

3.4 消融实验

为了在训练中提高对目标的检测精度,对特征提取网络做出改进,在特征提取网络中的某个位置添加注意力机制,为了探究位置的合适性,本文在特征提取网络中的4个CSP结构前添加注意力机制,根据其表现来确定在何位置添置GAM注意力机制。

由表1可以知道,注意力机制添加在第一个CSP结构前面对各类遥感目标检测的平均精度均取得最佳。相比于将注意力机制放在后面的3个位置,放在第一个CSP结构的前面能够使得特征提取网络对图像中的目标更加关注,并且随着特征提取网络的加深,语义信息特征不断被强化,放大注意力机制所带来的增益效果,经实验表明,将GAM模块添加在第一个CSP位置前比较好。

为探讨以上3个改进策略对网络模型的贡献程度,以YOLOX-s为基线在RSOD数据集上进行消融实验,实验参数设置及实验环境保持一致,结果如表2所示。

表1 不同位置的GAM添加的结果

| 位置 | 飞机/% | 油箱/% | 立交桥/% | 操场/% | 平均精准率均值/% |
|-------|-------|-------|-------|-------|-----------|
| CSP_1 | 90.52 | 99.13 | 88.42 | 100.0 | 94.51 |
| CSP_2 | 90.44 | 99.38 | 81.03 | 100.0 | 92.71 |
| CSP_3 | 90.48 | 99.01 | 74.79 | 88.39 | 88.17 |
| CSP_4 | 90.43 | 99.02 | 87.58 | 100.0 | 94.25 |

表2 改进策略的消融实验

| GAM | MSCE | Loss | 飞机/% | 油箱/% | 立交桥/% | 操场/% | 平均精准率均值/% |
|-----|------|------|-------|-------|-------|-------|-----------|
| - | - | - | 89.33 | 90.91 | 65.47 | 100.0 | 86.43 |
| ✓ | - | - | 90.52 | 99.13 | 88.42 | 100.0 | 94.51 |
| - | ✓ | - | 90.51 | 99.54 | 88.71 | 100.0 | 94.69 |
| - | - | ✓ | 90.41 | 99.18 | 86.82 | 100.0 | 94.10 |
| ✓ | ✓ | ✓ | 90.49 | 99.46 | 90.51 | 100.0 | 95.12 |

由表 2 可知,采用 GAM 注意力机制较 YOLOX-s 的 mAP 增加了 8.08%,由此说明 GAM 模块可以使网络更加聚焦在遥感目标上,使整个模型的性能得到提升。改进多尺度特征融合模块为模型带来了 8.26%的增益,其原因采取了 MSCE-R 结构进行了不同尺度的信息提取,网络模型的细节表现能力更强了。采用 CIoU 损失函数使得 mAP 提高了 7.67%,CIoU 损失函数加快了模型的收敛速度,从而提高网络模型的性能。使用以上改进策略最终为模型带来了 8.69%的增益,证明了 3 种改进策略的有效性。

3.5 对比实验

现阶段的目标检测模型很多,比较好的改进策略也不少,为了验证本文所提方法在遥感图像检测中的优异表现,本文选择目前目标检测领域主流模型进行比较,选取的模型有 Faster R-CNN、SSD、DETR、YOLOX-s、TOOD 对比实验采取的指标为 mAP、FPS,在 RSOD 数据集上的对比结果如表 3 所示。

由对比实验结果所示,所提改进模型对比现阶段目标检测领域的主流模型的性能都有所提高。从实验结果分析来看,Faster-R-CNN、SSD、DETR、TOOD 等目标检测算法虽然能在 RSOD 数据集上取的较好的检测结果,但是相较于本文所提模型的 mAP 分别低了 7.41%、12.94%、

表 3 各算法检测精度与速度对比

| 模型 | 平均精准率均值/% | 检测速率/(img/s) |
|--------------|-----------|--------------|
| Faster R-CNN | 87.71 | 31.5 |
| SSD | 82.18 | 33.9 |
| DETR | 82.52 | 14.3 |
| YOLOX-s | 86.43 | 38.5 |
| TOOD | 92.47 | 22.5 |
| Ours | 95.12 | 34.4 |

12.60%、2.65%,在检测速度上也略低于本文所采用的模型;但是可以看见未做改进的 YOLOX-s 的 mAP 虽然低于本文模型,但是检测速度是在对比实验中最快的,高于其他算法。

3.6 可视化分析

图 5 展示 Faster R-CNN、SSD、DETR、YOLOX-s、TOOD 以及本文所采用模型的可视化结果。ROSD 数据集中的飞机目标的检测结果显示在不同的彩色框中。由于遥感图像的背景复杂多样,因此在检测的过程中经常出现漏检和误检,随着检测精度的提高,这种现象将减少。图 6 P-R 曲线图精确地展现了用于对比模型对每类遥感目标检出的准确率,其中横坐标 Recall 表示召回率,纵坐标 Precision 表示精准率。

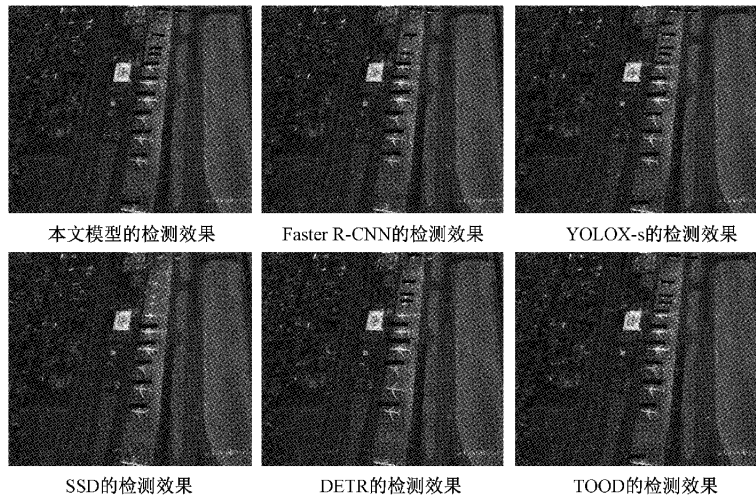
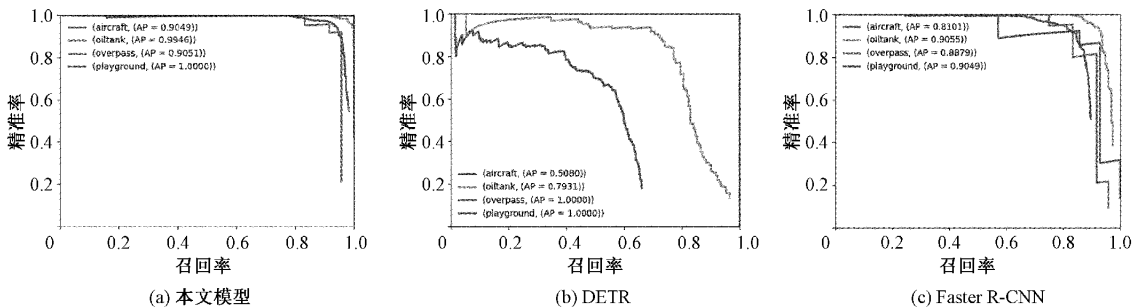


图 5 各类模型可视化对比



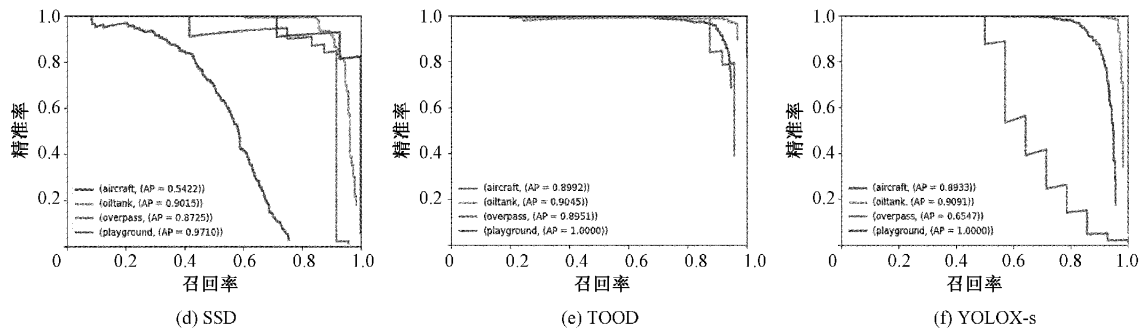


图6 对比模型各类遥感目标 P-R 曲线

4 结 论

为了提高在遥感图像中目标检测的精度,解决遥感图像因尺度变化导致的检测困难,提出一种改进模型,在YOLOX-s网络模型的基础上,在骨干网络中首先引入注意力机制,其次在特征金字塔部分加入多尺度信息融合模块,最后更改损失函数。实验在RSOD公共遥感数据集上面展开,对比了5种现阶段主流的目标检测模型。结果表明:所提方法表现最佳,3种改进策略可以有效的解决遥感图像小目标众多,目标排列紧密以及在特征提取过程中因尺度变换所导致的检测困难等问题,所提方法具有检测精度高等特点。但是本文所采用的多尺度信息融合模块MSCE-R所带来的影响是增加了模型的参数量,在对比实验中,在检测速度方面,仅能满足实时检测的要求,所以在今后的研究中,可以从减少参数、提升检测精度的角度开始研究。

参考文献

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [2] GIRSHICK R. Fast r-cnn [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [3] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [4] CAI Z, VASCONCELOS N. Cascade r-cnn: Delving into high quality object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object

detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

- [6] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [7] PANG J, CHEN K, SHI J, et al. Libra r-cnn: Towards balanced learning for object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 821-830.
- [8] 冯号,黄朝兵,文元桥. 基于改进YOLOv3的遥感图像小目标检测 [J]. 计算机应用, 2022, 42(12): 3723-3732.
- [9] 王成龙,赵倩,赵琰,等. 基于深度可分离卷积的实时遥感目标检测算法 [J]. 电光与控制, 2022, 29(8): 45-49.
- [10] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021 [J]. ArXiv Preprint, 2021, ArXiv:2107.08430.
- [11] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions [J]. ArXiv Preprint, 2021, ArXiv:2112.05561.
- [12] CAO Y S, NIU X, DOU Y. Region-based convolutional neural networks for object detection in very high resolution remote sensing images [C]. 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), IEEE, 2016: 548-554.
- [13] 常鹏飞,段云龙. Faster R-CNN模型在遥感图像飞机目标检测中的应用 [J]. 无线电工程, 2019, 49(10): 925-929.
- [14] LIN Y, SUN H, LIU N, et al. A lightweight multi-scale context network for salient object detection in optical remote sensing images [J]. ArXiv Preprint, 2022, ArXiv:2205.08959.

- [15] ZHU X, SU W, LU L, et al. Deformable detr: Deformable transformers for end-to-end object detection [J]. ArXiv Preprint, 2020, ArXiv: 2010.04159.
- [16] 薛雅丽,孙瑜,马瀚融. 航空遥感影像中的轻量级小目标检测[J]. 电光与控制,2022,29(6):11-15.
- [17] 李玉峰,顾曼璇,赵亮. 采用改进 Faster R-CNN 的遥感图像目标检测方法[J]. 信号处理,2020,36(8): 1363-1373.
- [18] DENG Z, SUN H, ZHOU S, et al. Multi-scale object detection in remote sensing imagery with convolutional neural networks [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 145: 3-22.
- [19] FENG C, ZHONG Y, GAO Y, et al. Toood: Task-aligned one-stage object detection [C]. 2021 IEEE/ CVF International Conference on Computer Vision (ICCV), IEEE Computer Society, 2021: 3490-3499.
- [20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [21] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [22] 雷帮军,耿红彬,吴正平. 基于自适应校准和多分支注意力的遥感目标检测[J]. 电子测量技术,2022, 45(22):106-111.

作者简介

左露,硕士研究生,主要从事深度学习与目标检测方面的研究。

E-mail:1278588531@qq.com

朱春惠,硕士研究生,主要从事深度学习与医学图像处理方面的研究。

牛晓伟,硕士,副教授,硕士生导师,主要从事智能信号处理方面的研究。

E-mail:luxiangxiang21@gmail.com