

基于多尺度下采样的遥感图像目标检测算法

周华平, 刘伟东

安徽理工大学 计算机科学与工程学院, 安徽 淮南 232001

摘要:目的 针对现有遥感图像目标检测算法在目标尺度变化大以及存在复杂背景信息干扰的场景中检测精度不佳的问题,提出一种基于多尺度下采样的遥感图像目标检测算法 MSD-YOLO(Multi-Scale Downsampling-YOLO)。方法 首先设计一种多尺度下采样(Multi-Scale Downsampling, MSD)模块,通过 3 个并行下采样分支在进行下采样的同时实现多尺度特征提取,避免了现有模型在经过多次下采样后,小目标特征信息丢失严重的问题,并引入具有自适应激活特性的 ACON(Activate or Not)激活函数,提升模型泛化能力;其次对三重注意力(Triplet Attention)机制进行改进,提出 ITA(Improved Triplet Attention)机制,通过捕获跨维度交互并强调空间注意力自适应地调整特征图的权重分配,提升模型在复杂背景信息干扰场景中的检测性能。结果 实验结果表明:MSD-YOLO 在 NWPU VHR-10 及 RSOD 数据集上的平均精度(mAP)分别达到 94.9%及 96.8%,相较于基线网络 YOLOv7,均提升了 1.5%,并且精度优于其他经典网络模型。结论 提出的 MSD-YOLO 算法可以有效提升在尺度变化大以及存在复杂背景信息干扰等场景中的检测精度,在遥感图像目标检测场景中有着一定的应用价值。

关键词:计算机视觉;遥感图像目标检测;MSD-YOLO;YOLOv7;Triplet Attention

中图分类号:TP 391.4 **文献标识码:**A **doi:**10.16055/j.issn.1672-058X.2025.0004.001

Remote Sensing Image Object Detection Algorithm Based on Multi-scale Downsampling

ZHOU Huaping, LIU Weidong

School of Computer Science and Engineering, Anhui University of Science and Technology, Anhui Huainan 232001, China

Abstract: Objective Aiming at the problem of poor detection accuracy of existing remote sensing image object detection algorithms in scenes with large variations of object's scales and complex background information interference, a remote sensing image object detection algorithm, MSD-YOLO (Multi-Scale Downsampling-YOLO) based on multi-scale downsampling, was proposed. **Methods** Firstly, a multi-scale downsampling (MSD) module was designed, which employed three parallel downsampling branches to extract multi-scale features simultaneously during downsampling. This approach avoided the problem of severe loss of small target feature information after many times of downsampling in existing models. Additionally, an ACON (Activate or Not) activation function with adaptive activation characteristics was introduced to enhance the model's generalization ability. Secondly, the Triplet Attention mechanism was improved, and the Improved Triplet Attention (ITA) mechanism was proposed to adaptively adjust the weight allocation of feature maps by capturing cross-dimensional interactions and emphasizing spatial attention. This mechanism improved the detection performance of the model in scenes with complex background information interference. **Results** The experimental results

收稿日期:2023-10-27 **修回日期:**2024-01-18 **文章编号:**1672-058X(2025)04-0001-08

基金项目:安徽省重点研发计划国际科技合作专项(202004B11020029)。

作者简介:周华平(1979—),女,河南唐河人,教授,博士,从事目标检测研究。

通信作者:刘伟东(1998—),男,安徽亳州人,硕士研究生,从事目标检测研究。Email:1512722690@qq.com。

引用格式:周华平,刘伟东. 基于多尺度下采样的遥感图像目标检测算法[J]. 重庆工商大学学报(自然科学版), 2025, 42(4): 1-8.

ZHOU Huaping, LIU Weidong. Remote sensing image object detection algorithm based on multi-scale downsampling[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(4): 1-8.

showed that the mean average precisions (mAP) of MSD-YOLO on NWPU VHR-10 and RSOD datasets reached 94.9% and 96.8%, respectively, which were both improved by 1.5% compared with the baseline network YOLOv7, and the precisions outperformed those of other classical network models. **Conclusion** The proposed MSD-YOLO algorithm effectively improves detection accuracy in scenarios with significant scale variations and complex background interference and has valuable applications in remote sensing image object detection scenarios.

Keywords: computer vision; remote sensing image object detection; MSD-YOLO; YOLOv7; Triplet Attention

1 引言

遥感图像广泛应用于土地利用分析、地籍测绘、城市规划以及自然灾害监测等任务。相较于传统的人工处理遥感图像,基于深度学习的遥感图像目标检测,通过深度学习算法将特征从遥感图像中提取出来,完成目标的识别和定位,能够更快速、更准确地识别大范围区域的目标,在提高检测效率的同时降低检测成本^[1-3]。

近年来,许多优秀的遥感图像目标检测算法不断涌现,如 Han 等^[4]提出一种多尺度残差块(Multiscale Residual Block, MRB),在级联残差块中引入扩展卷积来捕获多尺度上下文信息,并且提出一种多尺度感受野增强模块(Multiscale Receptive Field Enhancement Module, MRFEM),增强多尺度特征表示能力;Lei 等^[5]提出一种区域增强卷积神经网络(Region-Enhanced CNN, RECNN),引入注意力约束和多层融合策略到 CNN 模型中,提高遥感图像的目标检测精度;刘涛等^[6]在 YOLOv5 特征提取和特征融合网络的 C3 模块中引入轻量级通道注意力,并且增加一个融合浅层语义信息检测层来提高对小目标的检测效果,但是增加检测层的同时也增大了模型计算量,影响模型检测效率;Cheng 等^[7]针对光学遥感图像中尺度变化以及类间相似性带来的挑战,提出一个端到端的跨尺度特征融合(Cross-Scale Feature Fusion, CSFF)框架,使用特征金字塔网络获得多级特征,在顶部层插入 SE(Squeeze-and-Excitation, SE)模块,但忽略了多次下采样导致的小目标特征信息丢失问题。

尽管现有检测模型取得了大幅进展,不断推动着遥感图像目标检测技术的进步,但现有模型的主流研究方向是通过多尺度融合来提升检测精度,却忽略了遥感图像中大量小目标在经过多次下采样后特征信息丢失严重、复杂背景信息干扰严重等问题,导致现有模型检测精度不佳。针对上述问题,本文以 YOLOv7^[8]为基线网络进行改进,提出一种基于多尺度下采样的遥感图像目标检测算法 MSD-YOLO。首先设计一种多尺度下采样模块,通过并行的 3 分支设计在进行下采样

的同时实现多尺度特征提取,从而减少目标细节特征的丢失,并引入 ACON(Activate or Not)^[9]激活函数提升模型泛化能力;其次对三重注意力机制(Triplet Attention)^[10]进行改进,提出 ITA(Improved Triplet Attention, ITA)机制,通过强调空间注意力,自适应地调整特征图的权重分配,抑制背景噪声干扰,提高模型在复杂背景信息干扰场景中的检测性能。

2 网络系统结构设计

2.1 基线网络 YOLOv7 介绍

本文所选用的基线网络是单阶段目标检测网络 YOLOv7,该网络主要由 4 部分组成,第 1 部分为输入模块(Input),负责对输入的图片预处理,将输入图片调整为 640 * 640 尺寸的训练图片。第 2 部分为主干网络(Backbone),负责提取输入图像的特征,将提取到的特征输入到特征融合网络中。YOLOv7 主干网络主要由基础卷积模块(CBS)、高效层聚合网络模块(ELAN)以及最大池化模块(MP1)3 个部分组成。第 3 部分为特征融合网络(Neck),YOLOv7 的特征融合网络沿用了 YOLOv5 系列的路径聚合特征金字塔(PAFPN)架构,将 YOLOv5 中的 CSP 模块换为 ELAN-W 模块,特征融合网络用以融合不同尺度特征图,并将融合后的特征图输入头部网络。第 4 部分为头部网络(Head),头部网络由集合了 RepVGG^[11]重参数化思想的卷积模块(RepConv)以及检测头(IDetect)组成,头部网络负责完成类别预测与目标边界框的锚定,最终输出不同尺寸的预测结果。

2.2 MSD-YOLO 网络模型设计

针对遥感图像中目标尺度变化大且存在复杂背景信息干扰,导致现有模型检测精度不佳的问题,选取单阶段目标检测网络 YOLOv7 为基线网络进行改进,提出一种基于多尺度下采样的遥感图像目标检测网络 MSD-YOLO。MSD-YOLO 的网络模型结构如图 1 所示,其中改进模块主要包括多尺度下采样模块 MSD 以及改进三重注意力 ITA 机制,用红色线条标注,并使用不同的颜色来代表不同的改进模块。

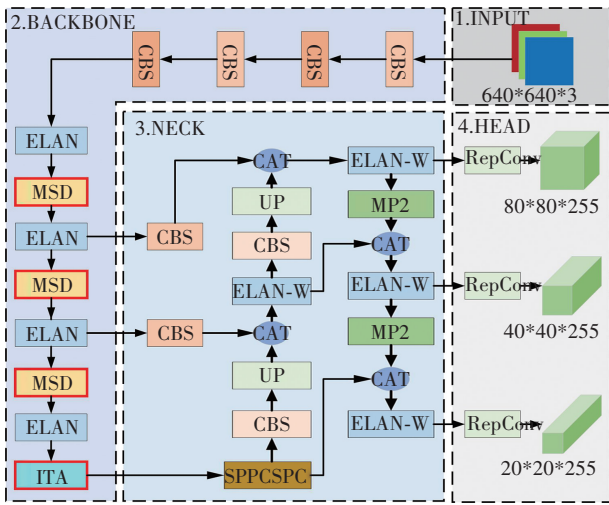


图 1 MSD-YOLO 网络模型结构图

Fig. 1 Structure diagram of MSD-YOLO network model

2.3 多尺度下采样模块

在遥感图像目标检测场景中目标尺度变化大,小目标占比高,原始 YOLOv7 网络在使用 MP1 模块进行多次下采样后会导致目标细节特征丢失严重,从而导致模型在应对这种尺度变化大的场景时表现欠佳。针对上述问题,本文设计了 MSD 模块,在进行下采样的同时捕捉不同尺度的特征,MSD 模块结构如图 2 所示。MSD 模块为并行的 3 分支设计,主要由 DWBA、CBA 以及 MaxPool 3 个部分组成,输入的特征经过 3 个并行分支,在进行下采样的同时实现了对不同尺度特征的特征提取操作,从而减少了目标细节特征的丢失。

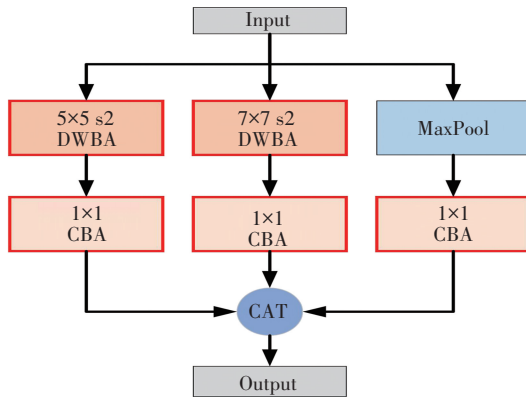


图 2 MSD 结构图

Fig. 2 Structure diagram of MSD

DWBA 以及 CBA 模块结构如图 3 所示。DWBA 模块由深度可分离卷积 (Depthwise Separable Convolution, DWConv)、归一化层 (Batch Normalization, BN) 以及 ACON 激活函数组成。遥感图像中小目标占比高,而且往往伴随着复杂的背景和多样的环境条件,通过增大感受野,模型可以更好地理解目标与周围环境的关系,提供丰富的上下文信息,从而实现更准确的目标定位,

有利于提升小目标的检测精度。在目标尺度变化大的场景中综合使用不同尺度的卷积核,可以提取不同尺度的特征,提高模型检测准确性和鲁棒性。因此本文在前两个分支中的 DWBA 模块中分别使用了卷积核大小为 5×5 以及 7×7 ,步距为 2 的深度可分离卷积实现下采样操作,在提供多尺度感受野提取多尺度特征信息的同时控制了模型的参数量。CBA 模块由卷积核大小为 1×1 的卷积、归一化层以及 ACON 激活函数组成,用以实现通道降维并增加非线性从而提高模型的效率。

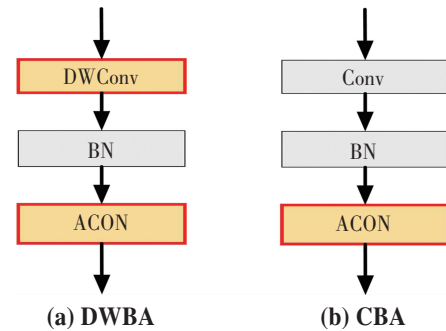


图 3 DWBA 及 CBA 模块结构图

Fig. 3 Structure diagram of DWBA and CBA modules

MSD 模块中引入的 ACON 激活函数具有自适应学习激活与否的特性,如图 4 所示 (黄色标注表示激活,白色表示未激活)。ACON 允许每个神经元根据输入样本的特征自适应地选择是否激活。相较于 YOLOv7 模型中使用的 SiLU 激活函数,ACON 激活函数具有减少过拟合风险、更好的稳定性、更快的收敛速度和适应不同任务等优势。这样的设计可以帮助提升模型泛化能力以及检测精度,更好地适应遥感图像目标检测任务中的尺度变化和复杂背景情况,以提升模型的性能以及稳定性。

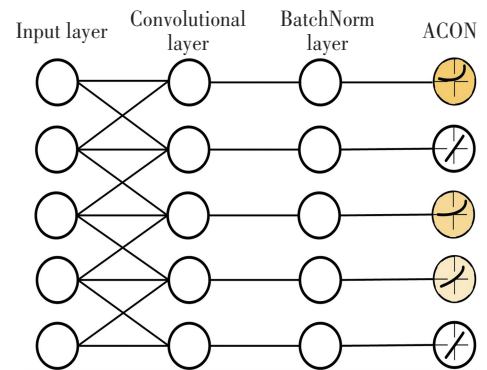


图 4 ACON 激活函数示意图

Fig. 4 Diagram of ACON activation function

2.4 改进的三重注意力机制

注意力机制在遥感图像目标检测中扮演着重要的角色,它能够建立通道或者空间位置之间的相互依赖

关系,从而使模型能够集中注意力于感兴趣的目标区域,同时忽略掉无关的背景信息。在遥感图像目标检测这样存在复杂背景干扰的场景中,引入注意力机制可以显著提升模型的检测性能。

近年来,众多优秀的注意力机制不断涌现,推动着目标检测技术的持续发展。其中,三重注意力机制 TA (Triple Attention) 是一种轻量级的注意力机制,通过利用 3 个并行分支结构捕获跨维度交互并计算注意权重。与 SE (Squeeze-and-Excitation, SE)^[12] 和 CBAM (Convolutional Block Attention Module, CBAM)^[13] 等注意力机制相比,三重注意力机制在实现出色性能的同时具有可忽略不计的计算开销,其结构如图 5 所示。

然而,在处理复杂背景干扰严重的遥感图像目标检

测场景时,三重注意力性能受限,为了改进三重注意力机制并解决这一问题,本文提出一种改进的三重注意力机制 ITA (Improved Triple Attention),其结构如图 6 所示。

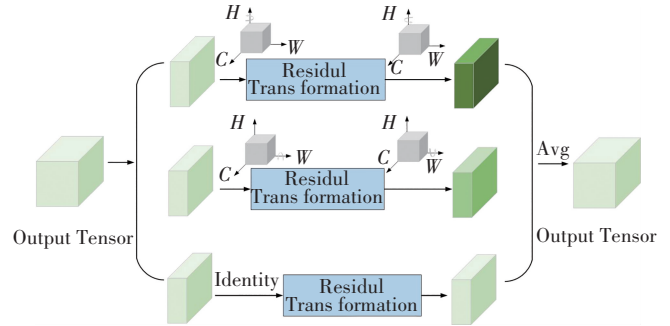


图 5 三重注意力机制结构图

Fig. 5 Structure diagram of triple attention mechanism

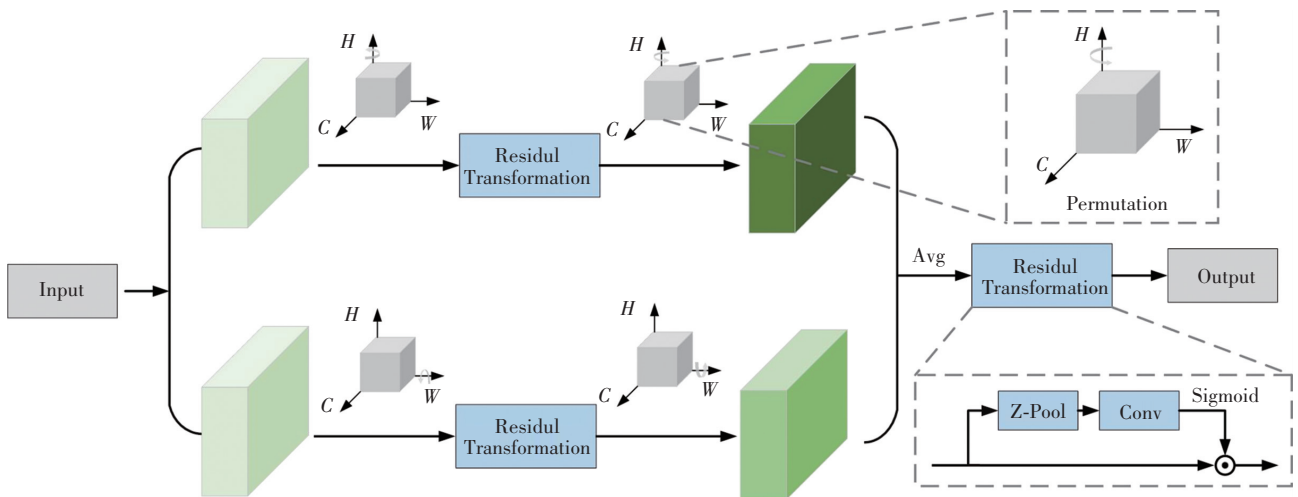


图 6 改进的三重注意力机制结构图

Fig. 6 Structure diagram of improved triple attention mechanism

相较于原始三重注意力机制通过 3 个并行分支结构来捕获跨维度交互,并计算注意权重的结构设计,ITA 机制首先通过两个并行分支对通道维度 C 与空间维度 H 或 W 之间的交互进行捕获,使网络能够更好地整合不同维度之间的信息,并提高特征的表达能;再使用平均池化操作对输出特征进行处理,减小特征图的尺寸,并将每个通道的信息进行整合;随后再经过 Residual Transformation 模块进行空间注意力,权重的计算,强调了空间注意力以提高模型在遥感图像目标检测场景中的准确性和稳健性。

Residual Transformation 模块中的 Z-pool 层的作用是将维度 C 的张量缩减为 2 维,并将平均池化特征和最大池化的特征连接起来,其计算公式为式(1):

$$Z\text{-pool}(X) = [\text{MaxPool } 0d(X), \text{AvgPool } 0d(X)] \quad (1)$$

式(1)中,MaxPool 0d 和 AvgPool 0d 分别代表在第 0 维度进行最大池化以及平均池化操作。输入张量 $X \in \mathbf{R}^{C \times H \times W}$,经过 TA 模块的操作计算公式为式(2):

$\mathbf{R}^{C \times H \times W}$,经过 TA 模块的操作计算公式为式(2):

$$y = \frac{1}{2} (\tilde{X}_1 \sigma(\psi_1(\tilde{X}_1^*)) + \tilde{X}_2 \sigma(\psi_2(\tilde{X}_2^*))) \sigma(\psi_3(\tilde{X}_3)) \quad (2)$$

式(2)中, σ 为 sigmoid 激活函数, ψ_1 、 ψ_2 和 ψ_3 表示三重注意 3 个分支中的标准二维卷积层。将式(2)化简后可得 y ,为式(3):

$$y = \frac{1}{2} (\overline{X_1 \omega_1} + \overline{X_2 \omega_2}) \omega_3 = \frac{1}{2} (\bar{y}_1 + \bar{y}_2) y_3 \quad (3)$$

式(3)中, ω_1 、 ω_2 和 ω_3 是 3 个跨维注意权重; $\bar{y}_1$ 和 $\bar{y}_2$ 表示 90°顺时针旋转,以保持输入张量 $X \in \mathbf{R}^{C \times H \times W}$ 的原始输入形状不变。

将 ITA 机制嵌入到 MSD-YOLO 特征提取网络中,通过强调空间注意力自适应地调整特征图的权重分配,对不同尺度的特征进行加权融合,并抑制背景噪声的干扰,能够帮助模型更好地满足复杂背景干扰场景

中的检测需求。

3 仿真实验与结果分析

3.1 实验数据集

本文实验采用的数据集是 NWPU VHR-10^[14] 以及 RSOD^[15]。NWPU VHR-10 数据集是由中国西北工业大学(NWPU)创建的一种用于遥感图像识别的公共数据集,该数据集共包含 10 个不同类别,一共包含 800 张图像,其中 650 张带有目标,150 张为背景图像。该数据集已经对 3 896 个目标实例进行了标注和验证。RSOD 数据集由武汉大学创建,一共包含 936 张图像,6 950 个实例样本。

在本文的实验中,分别将 NWPU VHR-10 以及 RSOD 数据集以 8 : 2 的比例随机划分为训练集和验证集,并使用 Mosaic 数据增强算法增加样本多样性,提高模型鲁棒性,减轻过拟合风险并提高训练效率,提高目标检测模型的性能和泛化能力。

3.2 实验设置和训练参数

实验中使用相同的超参数配置,包括 epochs 为 300、Batchsize 为 8、Img-size 为 640×640、lr0 为 0.01。实验采用的实验环境设置如表 1 所示。

表 1 实验环境设置

Table 1 Settings of experimental environment

名 称	配置信息
操作系统	Windows11
处理器	12th Gen Intel(R) Core(TM) i5-12400F
显卡	NVIDIA GeForce RTX 3060 (12G)
深度学习框架	Pytorch 2.0.0
开发环境	Python 3.10.9、CUDA 11.7

表 2 NWPU VHR-10 数据集消融实验结果

Table 2 Ablation results of NWPU VHR-10 dataset

组 别	改进方法			$f_{\text{Precision}}/\%$	$f_{\text{Recall}}/\%$	$f_{\text{mAP}}/\%$	参数量/M	浮点运算 次数/G
	MSD(SiLU)	MSD(AcoN)	ITA					
1				92.8	86.2	93.4	37.24	105.3
2	✓			93.9	89.3	94.1	35.40	98.6
3		✓		91	91.5	94.3	35.90	99.0
4			✓	93.4	88.7	94.5	37.24	105.4
5		✓	✓	93.9	89.5	94.9	35.90	99.1

根据表 2 的消融实验结果可知:在组别 2 中,相对于组别 1 添加了 MSD(SiLU)模块,使模型在进行下采样的同时实现了多尺度特征提取, f_{mAP} 提升了 0.7%,达到了 94.1%;在组别 3 中,相对于组别 1 添加了 MSD(ACON)模块,引入 ACON 激活函数提升模型泛化能力, f_{mAP} 提升了 0.9%,达到了 94.3%,相对于组别 2 使

3.3 实验评价指标

本文实验使用平均精度率(mean Average Precision, f_{mAP})、准确率(Precision, $f_{\text{Precision}}$)以及召回率(Recall, f_{Recall})作为评估指标对模型的性能进行评估。其计算公式为式(4)、式(5):

$$f_{\text{Precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \quad (4)$$

$$f_{\text{Recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \quad (5)$$

f_{AP} 和 f_{mAP} 是用来评价网络模型性能优劣的重要指标。在计算 f_{AP} 时,根据不同的阈值对模型的预测结果进行排序,并计算出在每个阈值下的准确率和召回率;然后,通过计算这些准确率和召回率下的面积来得到 f_{AP} 值;最后,将不同类别的 f_{AP} 值取平均得到 f_{mAP} ,以更全面地评估模型的整体性能。计算公式为式(6)、式(7):

$$f_{\text{AP}} = \int_0^1 f_{\text{Precision}}(f_{\text{Recall}}) df_{\text{Recall}} \quad (6)$$

$$f_{\text{mAP}} = \frac{\sum_{i=1}^N f_{\text{AP},i}}{N} \quad (7)$$

同时使用参数量(Parameters)衡量模型体积,浮点运算次数(FLOPs)衡量模型计算量或时间复杂度。

3.4 消融实验

本文在 NWPU VHR-10 数据集上通过消融实验对比基线网络 YOLOv7 与不同改进模块,以此来评估各个改进模块的有效性。NWPU VHR-10 数据集消融实验结果如表 2 所示,其中 MSD(SiLU)表示使用 SiLU 激活函数的 MSD 模块,MSD(ACON)表示引入 ACON 激活函数的 MSD 模块;“✓”表示采用对应的改进方法。

用的 SiLU 激活函数, f_{mAP} 提升了 0.2%,证明了 ACON 激活函数的有效性;在组别 4 中,相对于组别 1 嵌入了 ITA 机制,抑制背景干扰信息,突出目标特征, f_{mAP} 提升了 1.1%,达到了 94.5%;在组别 5 中,同时添加 MSD(ACON)模块以及 ITA 模块, f_{mAP} 提升了 1.5%,达到了 94.9%。上述消融实验的结果证明了各个改进模块的

有效性,从图 7 检测效果对比图可以直观地看出:MSD-YOLO 的检测效果明显的优于 YOLOv7 网络模型,显

著改善了模型的漏检以及误检问题。

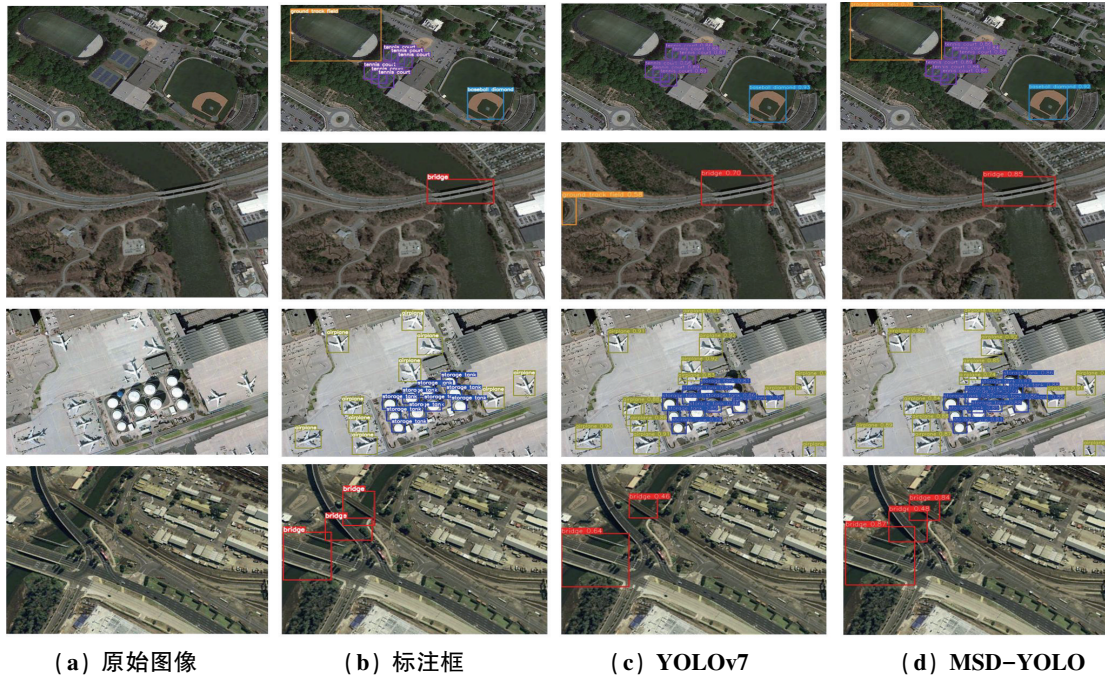


图 7 检测效果对比图

Fig. 7 Comparison of detection effects

3.5 热力图可视化对比

为了直观地探究 MSD-YOLO 网络模型在遥感图像目标检测场景中是否优于 YOLOv7 网络模型,本文使用 Grad-CAM^[16]对检测目标进行热力图可视化操作,通过对比 Grad-CAM 热力图辅助验证模型的有效性,理解模型决策过程以及改进模型性能。

Grad-CAM 热力图可视化对比如图 8 所示。可以直观地看出:相较于 YOLOv7 网络模型,MSD-YOLO 热力图中的高亮区域更加集中,这表明 MSD-YOLO 更精确地关注到了与目标相关的区域,在目标识别和定位方面表现优于 YOLOv7,证明了 MSD-YOLO 在遥感图像目标检测场景中优于 YOLOv7。

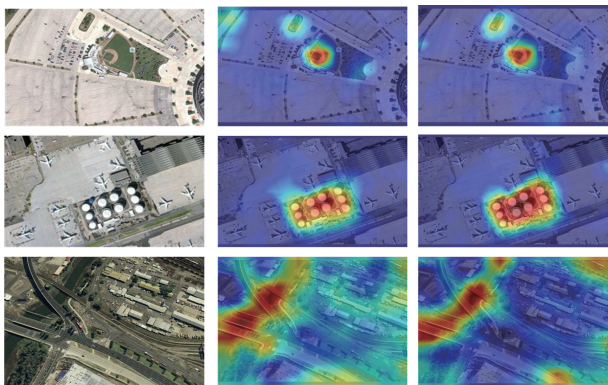


图 8 Grad-CAM 热力图可视化对比

Fig. 8 Visual comparison of Grad-CAM thermal maps

3.6 模型对比实验

本文在 NWPU VHR-10 数据集上进行对比实验。以 f_{mAP} 作为主要评价指标,将 MSD-YOLO 与基线网络 YOLOv7、二阶段目标检测网络 Faster R-CNN^[17]、单阶段目标检测网络 YOLOv5m 以及 YOLOv8m、实时端到端目标检测网络 RT-DETR-L、TDMD^[18] 以及 GLSANet^[19] 进行对比。表 3 展示了在 NWPU VHR-10 数据集上的对比实验结果。

根据表 3 的对比实验结果可知:MSD-YOLO 在 f_{mAP} 指标上相对于基线网络 YOLOv7 提升了 1.5%,达到了 94.9%;与 Faster R-CNN 以及 RT-DETR-L 相比,本文算法分别提升了 11.3% 以及 10.4%,实时端到端目标检测网络 RT-DETR-L 在本数据集上检测精度略低,该网络在小目标占比高且密集的场景中优势不明显;与 YOLOv5m、YOLOv8m、TDMD 以及 GLSANet 相比,MSD-YOLO 分别提升了 2.2%、2.7%、1.1% 以及 0.4%;相较于其他目标检测算法,对于诸如网球场、田径场以及汽车等尺度变化大的目标,MSD-YOLO 取得了更好的检测结果;对于诸如桥梁以及汽车等存在复杂背景信息干扰的目标,MSD-YOLO 的检测精度取得了明显的提升。上述对比实验结果表明:相较于其他目标检测算法,MSD-YOLO 取得了明显的精度提升,证明了 MSD-YOLO 的有效性。

表 3 NWPU VHR-10 数据集对比实验结果

Table 3 Comparison of experimental results on the NWPU VHR-10 dataset

/%

类别	f_{AP}							
	Faster R-CNN	YOLOv5m	YOLOv7	YOLOv8m	RT-DETR-L	TDMD	GLSANet	MSD-YOLO
飞机	95.4	99.5	99.5	99.4	98.4	99.7	99.9	99.5
舰船	81.6	91.3	93.1	93.0	88.4	90.9	95.8	93.2
油罐	70.3	91.0	96.1	84.2	73.7	91.6	97.1	93.0
棒球场	88.1	98.5	98.4	98.3	97.0	90.7	99.4	97.5
网球场	78.3	90.6	95.7	93.7	92.7	90.6	98.7	97.7
篮球场	91.7	83.9	91.7	81.6	72.7	98.6	86.1	92.4
田径场	96.6	98.3	97.3	99.0	89.3	96.5	99.4	98.9
港口	76.4	92.9	90.2	88.4	75.1	90.3	97.5	86.2
桥梁	74.3	90.9	81.0	92.3	76.7	90.8	84.8	96.7
汽车	85.1	89.9	91.0	92.5	80.6	97.9	86.7	94.4
$f_{mAP}/\%$	83.6	92.7	93.4	92.2	84.5	93.8	94.5	94.9

本文同时在 RSOD 数据集上进行对比实验,进一步验证 MSD-YOLO 的有效性。根据表 4 的对比实验结果可知:在 ROSD 数据集中,MSD-YOLO 的 f_{mAP} 指标相对于基线网络 YOLOv7 的 95.3% 提升 1.5%,达到 96.8%;尤其是在立交桥这类存在复杂背景信息干扰情况下目标检测精度相较于 YOLOv7 提升 6%,进一步证明了 MSD-YOLO 在复杂背景信息干扰检测场景中的有效性;此外,MSD-YOLO 的检测精度也不同程度地高于其他网络模型。

表 4 ROSD 数据集对比实验结果

Table 4 Comparison of experimental results on the ROSD dataset

/%

网络模型	f_{AP}				f_{mAP}
	飞机	油罐	立交桥	操场	
YOLOv7	93.5	98.4	90.4	98.8	95.3
YOLOv8m	93.2	97.2	88.6	99.0	94.5
GLSANet	96.8	98.4	91.4	97.0	95.9
MSD-YOLO	93.3	98.3	96.4	99.3	96.8

4 结 论

针对遥感图像目标检测场景中目标尺度变化大,并且存在复杂背景信息干扰,导致现有模型检测精度不佳的问题,提出一种基于多尺度下采样的遥感图像

目标检测算法 MSD-YOLO:首先设计一种多尺度下采样模块,通过并行的 3 分支设计,在进行下采样的同时实现多尺度特征提取,从而减少了目标细节特征的丢失,并引入 ACON 激活函数提升模型泛化能力;其次对三重注意力机制进行改进,提出 ITA 机制,通过强调空间注意力,自适应地调整特征图的权重分配,抑制背景噪声干扰,提升模型在复杂背景信息干扰场景中的检测性能。在公共数据集 NWPU VHR-10 以及 RSOD 上的实验结果表明:相较于基线网络 YOLOv7,改进后的算法在 f_{mAP} 指标上提高了 1.5%,分别达到了 94.9%及 96.8%,证明了 MSD-YOLO 在遥感图像目标检测场景中的有效性。

然而难以检测的微小目标仍然限制着模型精度的进一步提升,如何更加巧妙地融合浅层语义信息与深层语义信息是关键,后续将沿着深浅层语义信息融合方向进行模型改进工作。

参考文献(References):

[1] 廖育荣, 王海宁, 林存宝, 等. 基于深度学习的光学遥感图像目标检测研究进展[J]. 通信学报, 2022, 43(5): 190-203.

LIAO Yu-rong, WANG Hai-ning, LIN Cun-bao, et al. Research progress of deep learning-based object detection of optical remote sensing image[J]. Journal on Communications, 2022, 43(5): 190-203.

- [2] 汪强, 郭来功, 程伟涛. 基于轻量化 NDFEDet-SOLOv2 的遥感图像建筑物提取方法[J]. 重庆工商大学学报(自然科学版), 2024, 41(6): 20-29.
- WANG Qiang, GUO Lai-gong, CHENG Wei-tao. A method for extracting buildings from remote sensing images based on lightweight NDFEDet-SOLOv2 [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(6): 20-29.
- [3] 付涵, 范湘涛, 严珍珍, 等. 基于深度学习的遥感图像目标检测技术研究进展[J]. 遥感技术与应用, 2022, 37(2): 290-305.
- FU Han, FAN Xiang-tao, YAN Zhen-zhen, et al. Progress of object detection in remote sensing images based on deep learning [J]. Remote Sensing Technology and Application, 2022, 37(2): 290-305.
- [4] HAN W, KUERBAN A, YANG Y, et al. Multi-vision network for accurate and real-time small object detection in optical remote sensing images [J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
- [5] LEI J, LUO X, FANG L, et al. Region-enhanced convolutional neural network for object detection in remote sensing images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(8): 5693-5702.
- [6] 刘涛, 丁雪妍, 张冰冰, 等. 改进 YOLOv5 的遥感图像检测方法[J]. 计算机工程与应用, 2023, 59(10): 253-261.
- LIU Tao, DING Xue-yan, ZHANG Bing-bing, et al. Improved YOLOv5 for remote sensing image detection [J]. Computer Engineering and Applications, 2023, 59(10): 253-261.
- [7] CHENG G, SI Y, HONG H, et al. Cross-scale feature fusion for object detection in optical remote sensing images [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(3): 431-435.
- [8] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Rainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2023: 7464-7475.
- [9] MA N, ZHANG X, LIU M, et al. Activate or not: Learning customized activation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 8032-8042.
- [10] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: Convolutional triplet attention module[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE, 2021: 3139-3148.
- [11] DING X, ZHANG X, MA N, et al. RepVGG: Making VGG-style convNets great again[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 13733-13742.
- [12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 7132-7141.
- [13] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision. Cham: Springer International Publishing, 2018: 3-19.
- [14] CHENG G, HAN J. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117: 11-28.
- [15] XIAO Z F, LIU Q, 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.
- [16] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 618-626.
- [17] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.
- [18] ZHAO B, WANG Q, WU Y, et al. Target detection model distillation using feature transition and label registration for remote sensing imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 5416-5426.
- [19] GAO T, NIU Q, ZHANG J, et al. Global to local: A scale-aware network for remote sensing object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61(7): 1-14.

责任编辑:李翠薇