

基于 DETR 的轻量级遥感图像目标检测算法

周孟然^a, 王 澳^b

安徽理工大学 a. 电气与信息工程学院; b. 计算机科学与工程学院, 安徽 淮南 232000

摘要:目的 针对传统遥感图像目标检测模型在无人机、卫星等低算力场景下难以部署的问题, 实现保持检测精度的同时, 降低模型的参数量, 提出一种基于 DETR 的轻量级遥感图像目标检测算法。方法 该模型首先采用 EfficientViT 特征提取模块作为轻量级骨干网络, 用于图像特征提取和筛选; 同时, 设计了一个轻量级高效混合编码器, 旨在降低模型参数量和计算量的同时, 保持检测精度, 该编码器由 S-AIFI 模块和 MSFM 模块组成, 其中 S-AIFI 模块专注于处理深层特征, 以增强对特征信息的聚合能力。而 MSFM 模块通过多尺度特征融合提高模型在遥感图像中对不同大小目标的检测能力; 最后, 引入了 shape-IoU 损失函数, 以进一步提高模型的检测精度。结果 在 DOTA-v1 数据集和 SIMD 数据集上进行实验, 该模型的 mAP 达到了 75.5% 及 81.9%, 其参数量降低到了 10.3 M。结论 训练后的模型具有较小的内存占用和参数量, 适用于计算资源有限的遥感图像处理应用场景。

关键词: 遥感图像; 轻量级网络; efficientvit; 多特征融合

中图分类号: TP751 文献标识码: A doi: 10.16055/j.issn.1672-058X.2026.0002.007

Lightweight Remote Sensing Image Object Detection Algorithm Based on DETR

ZHOU Mengran^a, WANG Ao^b

a. School of Electrical and Information Engineering; b. School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232000, Anhui, China

Abstract: Objective To address the deployment challenge of traditional object detection models in low-computation scenarios (e.g., drones and satellites), this study proposes a lightweight remote sensing image object detection algorithm based on DETR (Detection Transformer), which reduces model complexity while preserving detection accuracy. **Methods** Firstly, the proposed model employed an EfficientViT feature extraction module as a lightweight backbone for image feature extraction and selection. Then, a lightweight and efficient hybrid encoder was designed to reduce the number of parameters and computational cost of the model while maintaining detection accuracy. This encoder comprised two key components: the S-AIFI (Slim-Attention-based Intrascala Feature Interaction) module, which focused on processing deep features to enhance contextual aggregation of feature information, and the MSFM (Multi-Scale Feature Fusion Module), which improved detection capability for objects of varying sizes in remote sensing images through effective multi-scale fusion. Furthermore, a shape-IoU loss function was incorporated to refine the detection precision of the model. **Results** Experiments on the DOTA-v1 and SIMD datasets showed that the model achieved mean average precision (mAP) scores of 75.5% and 81.9%, respectively, with its parameter count reduced to 10.3 M. **Conclusion** The trained model exhibits a small memory footprint and low parameter count, making it suitable for remote sensing image processing applications with limited computational resources.

Keywords: remote sensing image; lightweight network; EfficientViT; multi-feature fusion

收稿日期: 2024-03-07 修回日期: 2024-05-24 文章编号: 1672-058X(2026)02-0054-07

基金项目: 国家重点研发计划专项子课题(2018YFC0604503)。

作者简介: 周孟然(1965—), 男, 安徽淮南人, 博士, 教授, 从事矿山机电系统监测、光电信息处理与煤矿安全检测研究。

通信作者: 王澳(1998—), 男, 安徽宿州人, 硕士, 从事遥感图像处理研究。Email: 1318275750@qq.com。

引用格式: 周孟然, 王澳. 基于 DETR 的轻量级遥感图像目标检测算法[J]. 重庆工商大学学报(自然科学版), 2026, 43(2): 54-60.

ZHOU Mengran, WANG Ao. Lightweight remote sensing image object detection algorithm based on DETR[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(2): 54-60.

随着光学遥感卫星技术的不断进步和在轨运行的卫星数量的增多,获取的遥感图像质量得以显著提升,尤其是在图像分辨率方面。利用高清晰度的光学遥感影像进行物体的检测与精确定位,是遥感技术应用中一个关键的研究内容^[1]。这一技术在多个领域内具有关键意义,包括推动智能城市的发展、进行生态与地质资源的详细调查、实施灾害的实时监控以及加强军事侦察能力等。卷积神经网络因其能力,在进行卓越的非线性特征提取、利用局部连接和权重共享等方面表现突出,在处理遥感图像的目标识别任务中已成为核心技术。然而,许多现有的卷积网络模型结构较为复杂,进行推理时往往涉及庞大的计算量和参数,这通常需要依靠高性能的 GPU 来保障顺畅的运算过程。在资源有限的环境中,如野外工作站、无人机系统或卫星平台,这些传统模型难以得到有效部署,或者它们需要依赖远程云服务来实现及时的数据处理,这可能在通信不便的条件下显得尤为困难。因此,在遥感图像处理领域,如何设计一个轻量化模型,以适应各种环境和设备条件,保证高效地实时推理。这一领域的研究对于实现遥感技术在更广泛的应用中发挥更大作用具有重要意义。

卷积神经网络在可移动设备上的应用需求不断增加,轻量级网络结构的设计得到了极大的发展^[2]。汪强等^[3]添加 BiFPN(Bi-FPN-Former)特征融合模块形成双向水平路由注意特征金字塔网络结构,从而聚焦微小建筑物轮廓信息,以实现更高层次的特征融合。另一方面,苗茹等^[4]针对小目标检测设计了新的小目标检测层,从而获得更精细的特征信息,提升网络模型对于小目标的检测精度。华为诺亚提出了 GhostNet^[5],通过恒等映射和线性变换的方式成功减少了模型的参数量。Wan 等^[6]利用多层特征金字塔、多探测头策略和混合注意模块对 YOLOv5 网络进行改进,提高了光学遥感图像目标检测网络的检测精度。2020 年 Nicolas Carion 等^[7]在目标检测领域使用端到端的 Transformer,直接预测物体,不再依赖 anchor/proposal 进行近似,也不进行 NMS(Non-Maximum Suppression)后处理策略,从而提高了目标检测速度。Peize Sun 等^[8]提出了一种用于图像中目标检测的纯稀疏方法,通过提供一组固定的学习目标建议稀疏集,以通过动态头执行分类和定位。Wang 等^[9]基于 DETR(Detection Transformer)进行改进,提出了一种新的查询方式,用于对一个区域的多个目标进行检测,并使查询更具解释性,以便于优化。

上述研究虽然具有一定的价值,但在遥感图像的检测精度方面仍存在一些不足:

(1) 传统检测方法通常只能对遥感图像中的目标进行大致识别,其检测效果远未达到理想状态,无法满足高精度的需求。

(2) 采用卷积神经网络(CNN)的检测方法虽然在小目标识别上取得了改善,但是面对遥感图像中同一类别内目标尺寸差异显著的情况,CNN 仍旧难以准确识别。此外,这类方法往往涉及庞大的参数量和计算量,限制了它们在实际应用中的部署与使用。

(3) 基于 DETR 的检测方法通过直接预测物体而省去了 NMS(非极大值抑制)后处理步骤,显著减少了检测时间。然而,这种方法在小目标检测上的表现并不尽如人意,导致在遥感图像检测任务中的精度不高。

综上所述,对于遥感图像的检测精度方面仍存在一些不足,模型的参数量和计算量依然较大,这使得检测速度难以满足实时检测的要求。因此,为了在遥感图像检测任务中尽可能保持精度,同时降低模型参数量和计算负担,本文提出了一种基于 DETR^[10]的轻量级遥感图像目标检测算法模型。该模型采用 EfficientViT 特征提取模块^[11]作为轻量级骨干网络,进行图像特征提取和筛选。为了在保持检测精度的同时降低模型参数量,设计了一个轻量级高效混合编码器,该编码器由 S-AIFI 模块(Slim-Attention-based Intrascale Feature Interaction)和 MSFM 模块(Multi-scale Feature Fusion Module)组成,其中 S-AIFI 模块是一个轻量化的尺度内特征交互模块,用于处理空间信息和进行特征提取,最终输出具有更强语义信息的特征图,MSFM 模块则用于对多尺度特征进行融合,将其转化为一系列图像特征。随后,通过 IoU 感知查询选择编码器输出序列中的固定数量的图像特征作为解码器的初始对象查询;最后,使用 Shape-IoU 损失函数来优化原始损失函数。该损失函数能够通过聚焦边框自身形状与尺度计算损失,从而使得边框回归更为精确。通过在 Dota-v1 和 SIMD 数据集上进行消融实验和对比实验,本文方法在降低模型参数量的同时,保持了对于遥感图像目标的检测精度。

1 系统结构设计

针对现有的目标检测模型其参数量大、计算负担重,在大多数遥感图像处理应用场景中计算资源紧缺的问题,本文提出一个基于 DETR 架构的轻量级遥感图像目标检测算法模型,其系统总体结构如图 1 所示。本节将详细介绍改进网络的总体结构和各个模块的结构和功能。

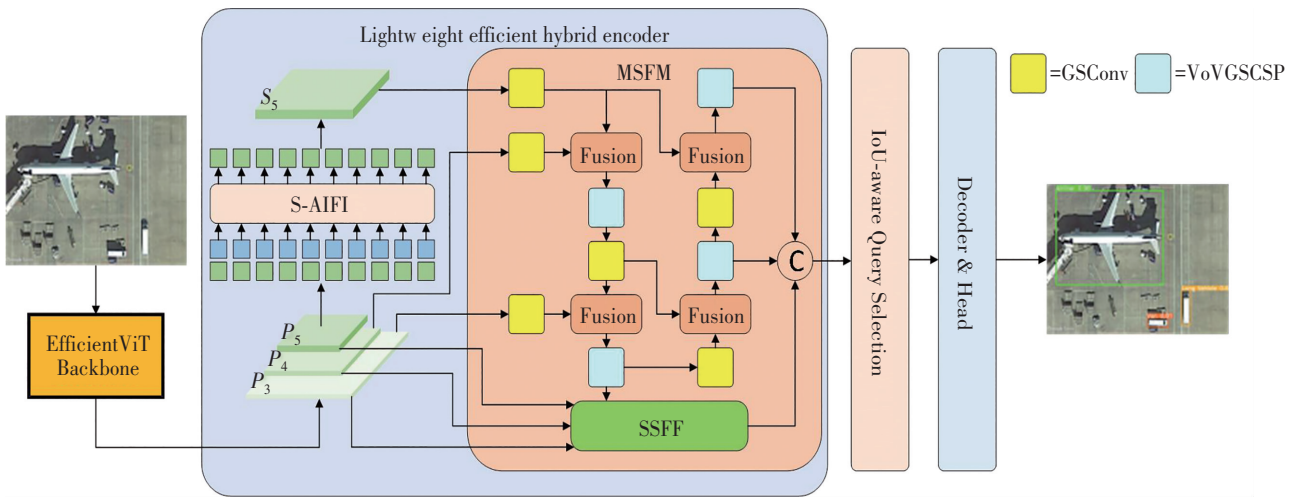


图 1 系统总体结构图

Fig. 1 Overall structure diagram of the system

1.1 系统总体结构

改进后的模型采用了 DETR 架构,整体结构包括特征提取模块 (Backbone)、轻量级高效混合编码器 (Lightweight Efficient Hybrid Encoder) 和解码器三个关键部分。在特征提取模块中,模型引入了 EfficientViT 特征提取网络,该网络以参数少和高效提取特征信息为特点。轻量级高效混合编码器由 S-AIFI 模块和 MSFM 模块组成,这两个模块能够有效地处理特征信息。随后,通过 IoU 感知查询选择进行进一步的优化,最终,解码器用于完成检测任务。这一结构的综合优势在于提高了模型的性能和效率,为检测任务的准确性和速度带来了显著的提升。

1.2 EfficientViT 特征提取网络

为在轻量化网络模型的同时,可以高效地提取图片特征,本文模型在 Backbone 部分采用了 EfficientViT 特征提取网络,用于对图片的特征进行提取和处理。EfficientViT 特征提取网络遵循标准的设计,包括一个 Stem 和四个 Stage,其中特征图大小逐渐减小,通道数逐渐增加。具体而言,Efficient 模块插入到第三和第四阶段。为了进行下采样,采用了带有步长为 2 的 MBConv。其结构如图 2 所示。

其中,Efficient 模块结合了轻量级多尺度注意力机制 (Lightweight MSA) 和 MBConv 块,以实现图像处理中的高效性能,其结构如图 3(a) 所示。轻量级多尺度注意力模块用于提取上下文信息,以协助模型理解图像中的全局信息。MBConv 块结构主要用于提取和利用局部信息,它包括以下关键部分:一个 1×1 的普通卷积、一个

$k \times k$ 的 Depthwise Conv 卷积(其中 k 的值主要有 3 和 5 两种情况)、一个 SE (Squeeze-and-Excitation) 模块、一个 1×1 的普通卷积以及一个 Dropout 层,其结构如图 3(b) 所示。全局感受野和多尺度学习对目标检测任务非常重要,可显著提高模型的性能。

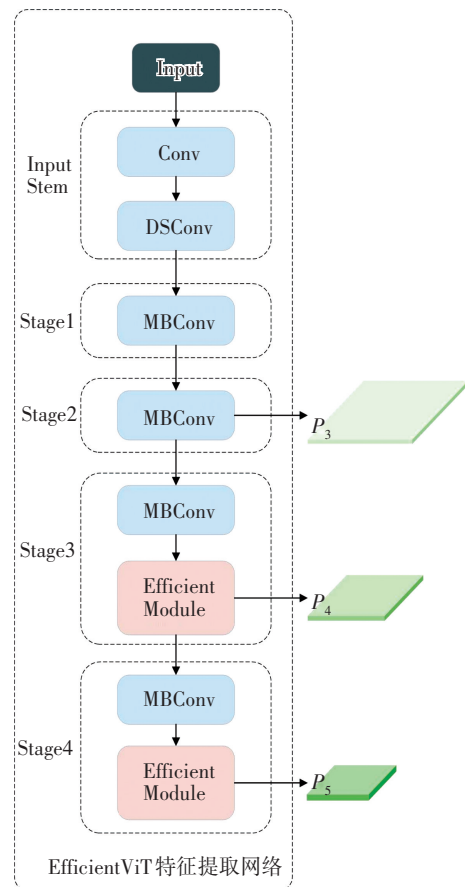


图 2 EfficientViT 特征提取网络结构图

Fig. 2 Structure diagram of EfficientViT feature extraction network

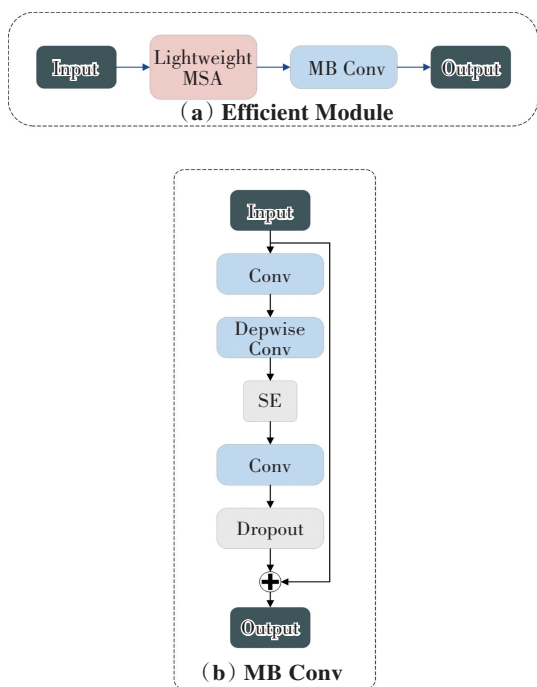


图 3 Efficient 模块结构图

Fig. 3 Structure diagram of Efficient module

1.3 轻量级高效混合编码器

为了在遥感图像目标检测中提升模型的精度并降低计算复杂性,设计了一个轻量级高效混合编码器 (Lightweight Efficient Hybrid Encoder),该编码器主要由 S-AIFI 模块和 MSFM 模块组成。

其中 S-AIFI 模块专门处理 P_5 特征图,其流程如下:首先,将二维的 P_5 特征拉伸成向量,然后经过一次多头自注意力处理。接下来,将处理后的结果输入一个全连接层进行进一步处理,最终将输出再次拉回到二维特征 S_5 。这一过程实现了一次尺度内的特征交互,从而提升了特征中包含的目标信息。S-AIFI 模块的具体数学表达式如式 (1) 和式 (2) 所示:

$$Q = K = V = \text{Flatten}(P_5) \quad (1)$$

$$S_5 = \text{Reshape}(\text{Attn}(Q, K, V)) \quad (2)$$

对于 MSFM 模块,其主要功能是对输入的特征进行多尺度特征融合,进一步整合目标特征,由此提升模型对目标检测精度,它由 GSConv 模块、VoVGSCSP 模块和 SSSF(Scale sequence feature fusion module) 模块组成,其结构如图 1 所示。

为了在降低计算量的同时有效地处理特征信息,该模块采用了轻量级卷积 GSConv 模块和 VoVGSCSP 模块。在 GSConv 模块中,特征图经过深度可分离卷积处理,随后通过采样随机化将由 SC 生成的信息(密集卷积操作)渗透到深度可分离卷积生成的信息的每个部分。这种方式使得 GSConv 能够在保留重要特征信息的同时减轻模型的复杂度。通过这种设计,可以进一步提高模型的速度和效率,同时保持较高的检测精度。使用一次性聚合方法来设计跨级部分网络

VoVGSCSP 模块。其结构如图 4 所示。

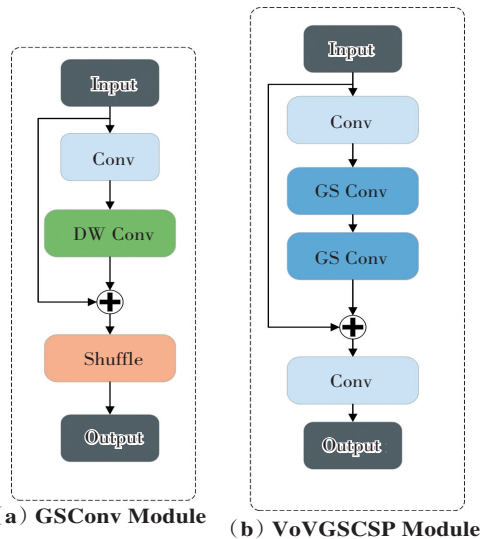


图 4 GSConv 和 VoVGSCSP 模块结构图

Fig. 4 Structure diagrams of GSConv and VoVGSCSP modules

在当前的特征融合方法中,存在对所有特征图之间相关性有效利用不足的问题。因此,在进行特征融合时,设计了一个称为 SSFF 模块,旨在更有效地将深层特征图的高维信息与浅层特征图的信息相结合,以优化 P_3, P_4 和 P_5 的特征图融合过程。在 SSFF 中,首先将 P_3, P_4 和 P_5 特征图归一化为相同大小,并进行上采样,然后将它们堆叠在一起,作为输入传递给 3D 卷积层,以便结合多尺度特征。最终,通过与经过处理的特征图进行一次拼接操作,输出的特征图能够捕获不同空间尺度下覆盖各种大小和形状的不同遥感目标类型。SSFF 模块的结构如图 5 所示。

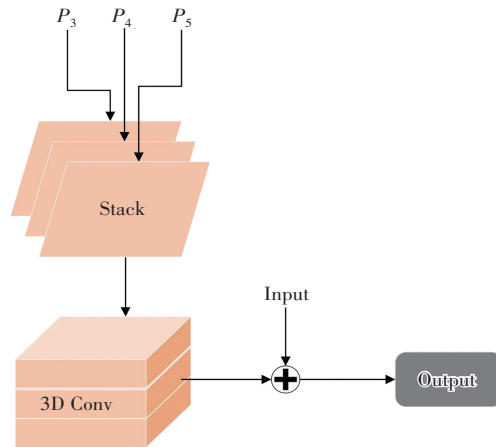


图 5 SSFF 模块结构图

Fig. 5 Structure diagram of SSFF module

1.4 Shape-IoU 损失函数

由于遥感图像中存在着目标尺度的显著差异以及大量小型目标的存在,目前的边框回归方法通常关注考虑实况框(GT 框)与预测框之间的几何关系,通过考虑边框之间的相对位置和形状等因素来计算损失。然而,这些方法忽视了边框本身的形状和尺度等固有属性对边框回归的潜在影响,因而在处理遥感图像时检

测精度较低。

为解决这一问题,引入了一种聚焦于边框自身形状和尺度的边框回归方法,即 Shape-IoU(形状-交并比)损失函数。该损失函数通过专注于边框的自身形状和尺度,计算损失,从而使得边框回归更加精准。其计算公式如下所示:

$$W_w = \frac{2 \times (w^{gt})^s}{(w^{gt})^s + (h^{gt})^s} \quad (3)$$

$$h_h = \frac{2 \times (h^{gt})^s}{(w^{gt})^s + (h^{gt})^s} \quad (4)$$

$$d^s = h_h \times \frac{(x_c - x_c^{gt})^2}{c^2} + w_w \times \frac{(y_c - y_c^{gt})^2}{c^2} \quad (5)$$

$$\Omega = \sum_{t=w,h} (1 - e^{-\theta t})^\theta, \theta = 4 \quad (6)$$

$$\begin{cases} \omega_w = h_h \times \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ \omega_h = w_w \times \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \quad (7)$$

其中, s 为尺度因子,与数据集中检测目标的尺度有关, w^{gt} 和 h^{gt} 是GT框的长宽, x_c^{gt} 和 y_c^{gt} 是GT框中心点的坐标, w 和 h 是先验框(Anchor)的长宽, x_c 和 y_c 是先验框中心点的坐标, w_w 与 h_h 分别为水平方向与垂直方向的权重系数,其取值与GT框的形状有关, θ 控制对形状损失的关注程度,为了避免过于关注形状损失而降低对预测框的移动,其取值通过遗传算法确定为4。其对应的边框回归损失如下:

$$L_{\text{Shape-IoU}} = 1 - f_{\text{IoU}} + d^s + 0.5 \times \Omega^s \quad (8)$$

其中,该损失函数由3个Cost函数组成:IoU损失函数(f_{IoU})、距离损失函数(d^s)和形状损失函数(Ω^s)。

2 仿真实验与结果分析

模型训练和性能评价试验在GPU服务器上完成,服务器硬件配置如下:CPU为Intel(R)Xeon(R)Platinum 8255C CPU,显卡型号为RTX 3090(24GB),软件环境为Python版本为3.11.4,torch版本为2.0.1,CUDA版本为12.2。为了对比实验的公平,实验设置统一的训练参数,具体参数:优化器采用的是标准

SGD,单次训练样本数量为16,迭代次数为300。

2.1 数据集介绍

本文模型在Dota-v1和SIMD两个数据集上进行实验。Dota-v1是由武汉大学遥感国家重点实验室与华中科技大学电信学院等机构共同创建的航拍图像数据集。这个数据集汇集了2806幅不同尺寸(范围从800×800像素到4000×4000像素)的遥感影像,涵盖了188282个标注实体。这些实体被划分进15个不同的类别,包括飞行器、船舶、油罐、棒球场、网球场、篮球场、田径跑道、港口、桥梁、大型运输工具、小型运输工具、直升机、足球场、立体交叉路口以及游泳池等。

SIMD数据集是由巴基斯坦国立科学技术大学提出遥感图像目标检测数据集,包含5000幅遥感图像(图片尺寸:1024×768)和45096个实例。数据集中的实例分为15类:汽车、卡车、厢货、长车、公共汽车、客机、螺旋桨飞机、教练机、包机、战斗机、其他、楼梯车、拖车、直升机和船。

2.2 评价指标

在本次研究中,为了衡量模型的效能,将使用以下四个关键指标进行评估:查准度 P (Precision)、召回率 R (Recall)、平均精确度均值(Mean Average Precision,简称mAP),以及模型参数量(Parameters)。通过这些指标,可以全面地评价模型在特定任务上的表现和效率:

$$P = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \quad (9)$$

$$R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \quad (10)$$

$$F_{\text{mAP}} = \frac{\sum_{i=1}^k i_{\text{AP}}}{k} \quad (11)$$

其中, N_{TP} 分别表示正确分类为目标的本数, N_{FP} 和 N_{FN} 分别表示被错误分类为目标和背景的本数, i_{AP} 表示第 i 个类别检测目标的平均精度。而 F_{mAP} 则是将所有类别的平均精度相加并除以类别总数得到的平均值。

2.3 消融实验

本文根据各个改进模块设置了不同的对比实验,表1是在Dota-v1数据集上进行消融实验得到的数据对比。

表1 本文方法在Dota-v1数据集上的消融实验

Table 1 The ablation experiment of the method in this paper performed on the Dota-v1 dataset

方法	EfficientViT	S-AIFI	MSFM	Shape-IoU	$P/\%$	$R/\%$	$F_{\text{mAP}}/\%$	M_{Par}/M
A	✓				72.1	73.9	72.3	10.7
B		✓			76.2	74.9	75.7	20.2
C	✓		✓		73.5	73.7	73.7	10.4
D	✓	✓	✓		74.0	75.2	74.6	10.4
E	✓	✓	✓	✓	76.3	74.0	75.5	10.3

表1中,对比了不同方法(A~E)在遥感图像目标检测中的性能表现。方法A中采用了EfficientViT特征提取网络,使得网络在保持检测精度的同时极大地减少了计算量和参数量。方法B中,使用了ResNet50作

为主干网络,并引入了S-AIFI模块,对深度特征实现了一次尺度内的特征交互,从而提升了特征中包含的目标信息,从而提升检测精度,但参数量显著提升。方法C中,采用了MSFM模块来对图像特征进行多尺度

特征融合,使得输出的特征图能够捕获不同空间尺度下覆盖各种大小和形状的不同遥感目标类型,这一改进相较于方法 A 使得 mAP 提升了 1.1%。方法 D 中,将 S-AIFI 模块和 MSFM 模块组成一个轻量级高效混合编码器,使得参数量降低到 10.4 M,mAP 提升了 0.9%。

方法 E 中,引入了 Shape-IoU,来提升网络对遥感图像中的不同目标的检测精度。实验结果显示,本文方法参数量降低到了 10.3 M,mAP 达到了 75.5%。

图 6 展示了原始图片和各个方法的检测结果。本方法具有更低的误检率,对于密集目标的检测效果更出色。

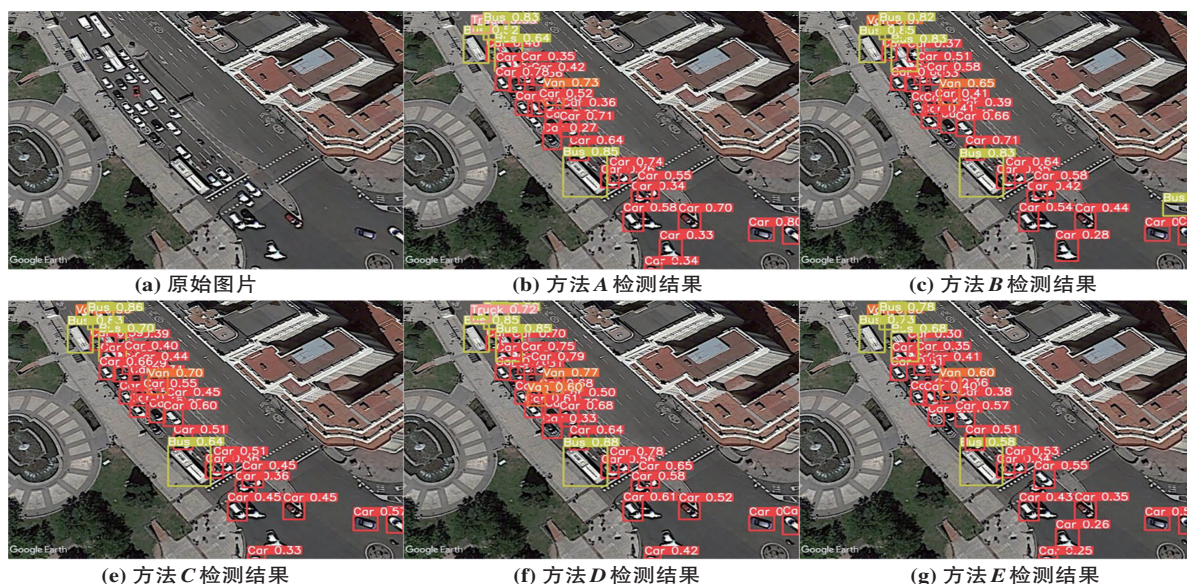


图 6 消融实验效果图

Fig. 6 Results of ablation experiment

2.4 对比实验

2.4.1 不同特征提取网络对比

本文方法采用了 EfficientViT 特征提取网络作为特征提取网络,通过替换不同的特征提取网络在 Dota-v1 数据集进行对比实验,最终选择了 EfficientViT 特征提取网络。对比实验如表 2 所示。

表 2 不同特征提取网络的对比实验结果

Table 2 Comparison of experimental results of different feature extraction networks

特征提取网络	P/%	R/%	F _{mAP} /%	M _{Par} /M
LSKNet ^[12]	74.0	75.2	74.8	15.7
ResNet50 ^[13]	76.2	74.9	75.7	20.1
Swin Transformer ^[14]	76.4	75.2	76.2	36.5
EfficientViT	76.3	74.0	75.5	10.3

2.4.2 Shape-IoU 不同参数对比

本文在网络模型中引入了 Shape-IoU 损失函数,在这个函数中存在一个 Scale 参数,其大小与数据集中目标的尺度有关,对不同的参数在 Dota-v1 数据集上进行对比实验,最终将 Scale 参数设置为 0.3。对比实验如表 3 所示。

表 3 shape-IoU 不同参数的对比实验

Table 3 Comparative experiment of different parameters of shape-IoU

Scale	0.0	0.1	0.2	0.3	0.4	0.5	1.0
mAP(%)	74.6	74.1	74.8	75.5	74.8	73.6	73.6

2.4.3 与其他算法对比

为了体现该改进方法的有效性,在 SIMD 数据集上进行与其他现有的算法的实验对比,为了实验的公平性,在训练时设置统一的参数,参与对比的算法有:SCRDet^[15]、RSDe^[16]、S2ANet^[17]、RVSA^[18]以及 YOLOv7^[19]。

经过实验对比,本文方法将参数量降低到了 10.3 M 的同时,其 mAP 达到了 81.9%,实现了在降低模型参数量的同时,保持了对于遥感图像目标的检测精度。对比结果如表 4 所示。

表 4 不同方法在 SIMD 数据集上实验结果的比较

Table 4 Comparison of experimental results of different methods on SIMD datasets

模 型	P/%	R/%	F _{mAP} /%	M _{Par} /M
YOLOv7	78.5	68.3	73.8	30.6
SCRDet	70.1	69.2	72.6	36.6
RSDe	75.2	70.1	76.6	26.2
S ² Anet	79.8	79.4	81.5	38.6
RVSA	80.5	80.1	82.1	41.9
本文方法	80.7	79.4	81.9	10.3

3 结论与展望

在遥感图像的应用领域,如野外勘察、无人机侦察和卫星监测等,常常面临计算资源有限且通信不畅的挑战。在这样的条件下,传统的检测模型往往难以直接部署,又不能依赖云服务来实现实时分析,故如何能

能够在保持检测精度的同时减少网络模型参数量和计算量的算法显得尤为重要,由此本文提出了一个基于 DETR 的轻量级遥感图像目标检测算法。首先采用 EfficientViT 特征提取模块作为轻量级骨干网络,用于进行图像特征提取和筛选;其次设计了一个轻量级高效混合编码器,该编码器由 S-AIFI 模块和 MSFM 模块组成,对多尺度特征进行融合,将其转化为一系列图像特征;最后,通过使用 Shape-IoU 损失函数来完善原始的损失函数,进一步提高了模型在遥感图像目标检测上的精度。在公共数据集 DOTA-v1 以及 SIMD 数据集集中的实验结果表明:改进后模型的参数量降低到了 10.3 M, mAP 分别达到了 75.5% 及 81.9%, 实现了在降低模型参数量的同时,保持了对于遥感图像目标的检测精度。然而由于遥感图像的特殊性,图像中包含大量的小型目标,如何处理小目标的特征信息,减少漏检和误检,将是后续模型改进工作的重点方向。

参考文献(References):

- [1] 李彦胜, 张永军. 耦合知识图谱和深度学习的新一代遥感影像解译范式[J]. 武汉大学学报(信息科学版), 2022, 47(8): 1176-1190.
LI Yan-sheng, ZHANG Yong-jun. A new paradigm of remote sensing image interpretation by coupling knowledge graph and deep learning[J]. Geomatics and Information Science of Wuhan University, 2022, 47(8): 1176-1190.
- [2] WU C, WEN W, AFZAL T, et al. A compactDNN: Approaching GoogLeNet-level accuracy of classification and domain adaptation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 761-770.
- [3] 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.
- [4] 苗茹, 岳明, 周珂, 等. 基于改进 YOLOv7 的遥感图像小目标检测方法[J]. 计算机工程与应用, 2024, 60(10): 246-255.
MIAO Ru, YUE Ming, ZHOU Ke, et al. Small target detection method in remote sensing images based on improved YOLOv7[J]. Computer Engineering and Applications, 2024, 60(10): 246-255.
- [5] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 1577-1586.
- [6] WAN D, LU R, WANG S, et al. YOLO-HR: improved YOLOv5 for object detection in high-resolution optical remote sensing images[J]. Remote Sensing, 2023, 15(3): 613-614.
- [7] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Computer Vision-ECCV 2020. Cham: Springer, 2020: 213-229.
- [8] SUN P, ZHANG R, JIANG Y, et al. Sparse R-CNN: end-to-end object detection with learnable proposals[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 14449-14458.
- [9] WANG Y, ZHANG X, YANG T, et al. Anchor DETR: query design for transformer-based detector[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(3): 2567-2575.
- [10] DAI X, CHEN Y, YANG J, et al. Dynamic DETR: end-to-end object detection with dynamic attention[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2022: 2968-2977.
- [11] LIU X, PENG H, ZHENG N, et al. EfficientViT: memory efficient vision transformer with cascaded group attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 14420-14430.
- [12] LI Y, HOU Q, ZHENG Z, et al. Large selective kernel network for remote sensing object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2024: 16748-16759.
- [13] THECKEDATH D, SEDAMKAR R R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks[J]. SN Computer Science, 2020, 1(2): 79.
- [14] LIU Z, NING J, CAO Y, et al. Video swin transformer[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 3192-3201.
- [15] YANG X, YANG J, YAN J, et al. SCRDet: towards more robust detection for small, cluttered and rotated objects[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2020: 8231-8240.
- [16] QIAN W, YANG X, PENG S, et al. Learning modulated loss for rotated object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(3): 2458-2466.
- [17] HAN J, DING J, LI J, et al. Align deep features for oriented object detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5602511.
- [18] WANG D, ZHANG Q, XU Y, et al. Advancing plain vision transformer toward remote sensing foundation model [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5607315.
- [19] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable 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. Piscataway: IEEE Press, 2023: 7464-7475.

责任编辑:陈 芳