

doi:10.13682/j.issn.2095-6533.2025.06.011

多元分支融合自注意力的遥感图像目标检测算法

亢红波,温家正,杨春杰,王文庆

(西安邮电大学人工智能学院、自动化学院,陕西西安710121)

摘要: 针对遥感拍摄目标角度变化导致图像中的检测目标尺度多样且密集分布难以准确检测的问题,提出一种多元分支融合自注意力(Multi Branch Fusion Self-Attention, MFS)的遥感图像目标检测算法。该算法先设计由卷积和自注意力机制组成的多分支模块,形成特征提取网络,再建立针对小物体的第4个检测头,旨在融合不同尺度的特征。同时,利用DepGraph剪枝方法进行剪枝,降低参数规模使其轻量化。实验结果表明,所提算法在航拍图像(Dataset for Object deTecton in Aerial Image, DOTA)数据集和NWPU VHR-10(Northwestern Polytechnical University Very High Resolution-10)数据集的平均准确率分别为77.7%和96.5%,优于同等参数规模的检测算法。特别是在剪枝后,参数规模仅有6.64 M的情况下,所提算法对DOTA数据集检测精度可以保持在72.9%。

关键词: 目标检测;遥感图像;多元分支;自注意力;轻量化

中图分类号: TP75

文献标志码: A

文章编号: 2095-6533(2025)06-0094-10

Multi-branch fusion self-attention object detection algorithm for remote sensing images

KANG Hongbo, WEN Jiazheng, YANG Chunjie, WANG Wenqing

(School of Artificial Intelligence, School of Automation, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

Abstract: To address the challenges of scale variation and dense object distribution in remote sensing imagery caused by varying imaging angles, a novel object detection algorithm is proposed based on multi-branch fusion self-attention (MFS). A multi-branch module that integrates convolutional and self-attention mechanisms is designed to build a feature extraction network, and the fourth detection head is built for small objects to facilitate multi-scale feature fusion. Meanwhile, the resulting model is pruned by the DepGraph method to achieve a lightweight architecture. Experiments on the DOTA and NWPU VHR-10 datasets demonstrate that the proposed algorithm achieves mean average precision (mAP) scores of 77.7% and 96.5% respectively, outperforming the peer detectors of similar algorithm complexity. Notably, the pruned version maintains a mAP of 72.9% on DOTA, with only 6.64 million parameters.

Keywords: target detection; remote sensing images; multivariate branching; self-attention; lightweight

遥感图像^[1]是通过不同波段记录各种地形地貌的图像,可为科学探索提供丰富的数据资源。目标检测作为遥感图像处理领域的基础任务之一^[2],具有重要的应用价值。面对丰富的遥感影像资源,依赖人工提取目标有效信息不仅效率低、成本高,且

在实际应用场景中无法快速准确获取目标信息,因此对目标检测算法提出了更高的技术要求。

基于深度学习的目标检测算法可分为两阶段目标检测算法和单阶段目标检测算法^[3]。两阶段的目标检测算法是将目标检测任务分为区域建议和

收稿日期: 2024-10-11

基金项目: 陕西省科技厅重点计划项目(2018ZDXM-GY-039)

引文格式: 亢红波,温家正,杨春杰,等.多元分支融合自注意力的遥感图像目标检测算法[J].西安邮电大学学报,2025,30(6):94-103.

KANG H B, WEN J Z, YANG C J, et al. Multi-branch fusion self-attention object detection algorithm for remote sensing images [J]. Journal of Xi'an University of Posts and Telecommunications, 2025, 30(6): 94-103.

检测两个阶段。如文献[4]采用 K-means 对遥感图像中的飞机数据进行聚类,并改进定位点,利用改进的 VGG16(Visual Geometry Group 16)提取出小飞机的定位特征。文献[5]提出一种基于脉冲卷积神经网络(Spatial Convolutional Neural Network, SCNN)稀疏表征的场景分类方法,去除遥感图像中与场景无关的冗余信息。文献[6]使用多尺度特征的动态融合,强调了图像的目标尺度对特征融合的影响。基于回归的单阶段目标检测算法最大的优点是检测速度快,可充分应用在实时性要求高的检测任务中。如文献[7]提出一种增强弱特征表达的一阶段轻量级小目标检测算法,提高了弱特征的解码能力,增强了小目标检测的性能。文献[8]提出基于单次多盒检测器的一种特征增强小目标检测算法,对小目标特征进行了进一步增强。文献[9]通过压缩 ResNeXt 获得轻量化主干,并通过自蒸馏头部提高了定位识别能力。文献[10-12]通过改进 YOLO(You Only Look Once)目标检测网络,提升了遥感图像目标检测的精度。文献[13]提出融合注意力机制的单次多边框检测(Single Shot Multi-Box Detector, SSD)网络,并结合空洞卷积得到更丰富的特征信息。

上述相关研究虽然在遥感图像的检测中取得了一定的成果,但是遥感图像因成像角度差异导致特征变化大,目标方向难以确定^[2]。此外,较大的参数规模和计算量,导致遥感图像目标检测网络延时较大。因此,本文拟提出一种多元分支融合自注意

力(Multi Branch Fusion Self-Attention, MFS)的遥感图像检测算法。通过构建多元分支结构级联的特征提取网络,设计多分支连接融合自注意力以达到充分提取特征信息的目的。添加针对小目标检测的第 4 个检测头,从而增强网络对小尺寸目标的检测能力。最后,使用结构化通道剪枝降低其参数量和计算量,得到轻量化的遥感图像目标检测算法。

1 遥感图像目标检测算法

基于 YOLOv7 目标检测算法,多元分支融合自注意力的遥感图像检测算法主要包括多元分支融合自注意力的特征提取网络、路径聚合网络和多尺度目标检测头,具体架构示意图如图 1 所示。输入的遥感图像,首先经过 MFS 模块和 MPConv(Max-pooling+Conv)级联结构进行特征提取,然后进入全连接空间金字塔池化卷积(Spatial Pyramid Pooling and Fully Connected Spatial Pyramid Convolution, SPPCSPC)模块进行信息聚合。特征提取网络将提取到的特征输入至特征融合网络,充分利用高效层聚合网络(Extended Efficient Layer Aggregation Networks, E-ELAN)模块^[14]进行特征提取,并使用自下而上和自上而下的路径聚合网络(Path Aggregation Network, PANet)结构^[15]聚合信息,将不同层的特征信息进行特征融合。最后,构建 4 个基于重参数化卷积^[16](Re-param Conv, Rep-Conv)的目标检测器进行目标检测。

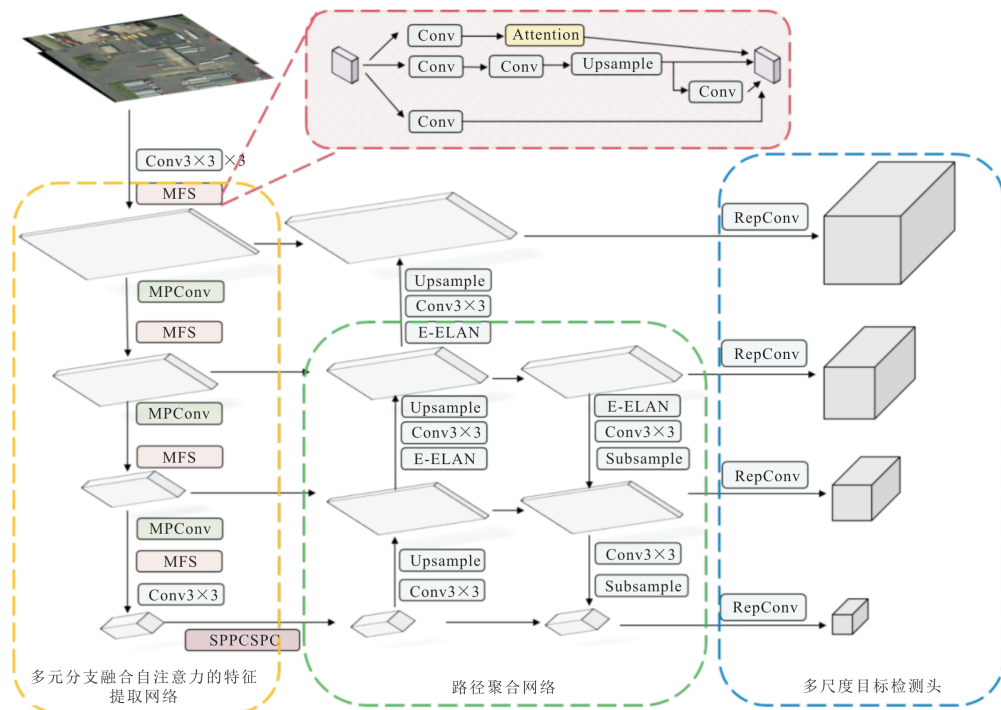


图 1 多元分支融合自注意力的遥感图像目标检测算法架构示意图

1.1 多元分支融合自注意力的特征提取网络

遥感图像中目标尺度变化大,单一分支结构的特征提取模块难以充分提取多尺度目标的信息特征,且仅使用单一卷积结构难以精确捕捉多尺度的目标特征。多元分支融合自注意力的特征提取网络结构是由3个MFS模块级联,再连接MPCConv作为输入,至特征提取网络的最后一层SPPCSPC层完成图像特征信息特征提取。

1.1.1 MFS 模块

MFS模块充分利用类残差式拼接多路分支的不同卷积特征提取结果,通过卷积的权值在不同分支的特征图上形成不同的感受野,为特征图带来不

可或缺的归纳偏差^[15],以增加模块对输入图像的多样性表示能力,从而更全面地捕捉多尺度目标的信息特征。在多路分支之中,融合嵌入自注意力机制,通过计算每个输入元素与其他元素之间的相似度,动态分配权重,在关注局部特征的同时,关注全局特征,捕捉最佳图像。不同区域的特征,通过评价目标像素和该像素周围的像素之间的相似度,动态调整注意力的分布,尽可能减少漏检或误检现象。

输入特征图 x 进行多分支特征提取,经过3个卷积核为 1×1 卷积,输出从上至下形成3个分支,对应输入为 x_1 、 x_2 和 x_3 。MFS模块的结构示意图如图2所示。

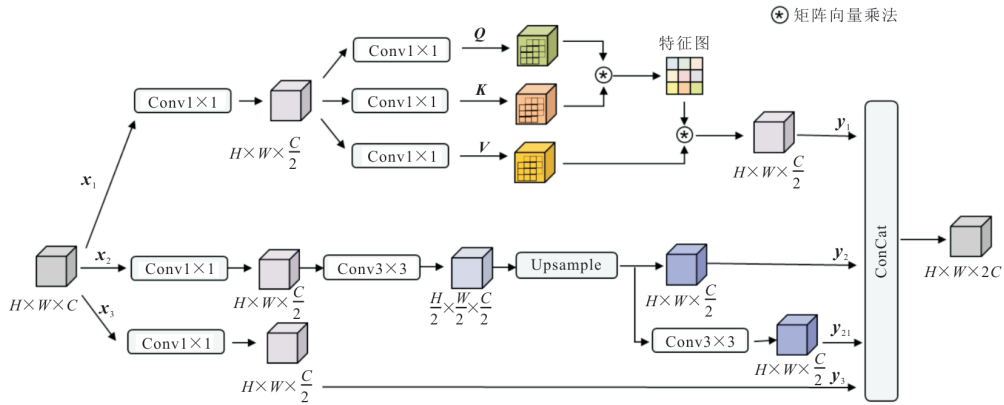


图2 MFS模块结构示意图

MFS模块的第一分支为融合自注意力机制的细粒特征提取通道。首先将输入特征图 x_1 的图像特征转换为 Q 和 K 两个特征矩阵用来计算注意力^[16],其表达式分别为

$$Q(x_1) = W_Q x_1 \quad (1)$$

$$K(x_1) = W_K x_1 \quad (2)$$

式中: W_Q 和 W_K 分别表示 Q 和 K 可学习的权值矩阵,以卷积核为 1×1 卷积实现。

将 $Q(x_1)$ 转置与 $K(x_1)$ 相乘得到注意力矩阵 s_{ij} ,反映各个像素之间的相关性,再将 s_{ij} 矩阵进行归一化(Softmax)处理,得到

$$\beta_{ij} = \text{Softmax}_{N_a(i,j)} \left(\frac{s_{ij}}{\sqrt{d}} \right) \quad (3)$$

式中: $N_a(i,j)$ 表示以 (i,j) 为中心,宽度为 a 的区域; d 为 $Q(x_1)$ 的特征维数。

细粒特征提取通道的最终输出的表达式为

$$y_1 = \sum_{N_a(i,j)} \beta_{ij} V(x_1) \quad (4)$$

其中,

$$V(x_1) = W_V x_1。$$

式中: W_V 为特征矩阵 V 可学习的权重矩阵,以卷积核为 1×1 卷积实现。

在第二分支中,对特征图 x_2 进行两次 3×3 卷积,获得更大更充分的感受野和权重共享,并引入更多非线性变换,进一步增加特征的提取能力。第二分支两个输出 y_{21} 和 y_2 的表达式分别为

$$y_{21} = \text{Up}(\text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(x_2))) \quad (5)$$

$$y_2 = \text{Conv}_{3 \times 3}(\text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(x_2))) \quad (6)$$

式中: $\text{Up}(\cdot)$ 表示 Upsample 上采样操作; $\text{Conv}(\cdot)$ 表示卷积操作。

第三分支 MFS 模块的输出 y 的表达式为

$$y = \text{Concat}(y_1, y_2, y_{21}, y_3) \quad (7)$$

其中,

$$y_3 = \text{Conv}_{1 \times 1}(x_3)。$$

式中: $\text{Concat}(\cdot)$ 表示通道拼接操作。

MFS模块嵌入自注意力机制,将局部特征与全局特征相结合,使模型具备更好的上下文感知能力,提升小尺度目标特征提取的准确性,同时利用卷积操作形成多分支结构,融合输入特征图中的不同层次特征,类残差结构组合特征信息,充分提高特征表达的综合能力。

1.1.2 特征提取网络

输入图像依次经过3个 3×3 卷积层(Conv2D,

Batch Normalization, Sigmoid-Weighted Linear Unit, Conv2D_BN_SiLU),旨在以较小的参数规模获取一定的感受野,得到宽、高、通道数为 $[160, 160, 128]$ 的特征图。将该特征图送入首个 MFS 模块,学习特征信息的分布规律,进行特征重新校准并聚焦于目标位置,输出尺寸为 $[160, 160, 256]$ 的特征图。

MFS 模块的输出紧接着与 1 个 MPCConv 模块级联,进行类下采样的操作,进一步整理特征图的通道并增大感受野。此过程中,上下分支通过 Concat 通道拼接,最终输出尺寸为 $[80, 80, 256]$ 的特征图。此级联过程被重复 3 次,其中 3 次 MFS 模块分别输出尺寸为 $[160, 160, 128]$ 、 $[80, 80, 512]$ 、 $[40, 40, 1024]$ 的特征图。最终,最后 1 次级联的结果再次通过 MFS 模块和额外的卷积操作后,被输入到 SPPCSPC 模块中。SPPCSPC 模块采用多种尺度的最大池化操作,旨在全面获取不同的感受野,以进

一步捕捉图像中不同尺寸目标的特征。

将第 2 次和第 3 次 MFS 模块经过卷积操作后的输出,以及 SPPCSPC 模块的最终输出,作为特征融合网络 PANet 的 3 个输入。鉴于遥感图像场景中普遍存在大量小尺寸目标,且这些目标在浅层特征提取网络中通常具有更高分辨率的特征图,同时包含更丰富的特征信息,在 PANet 原有基础上进行改进。具体而言,将第 1 次 MFS 模块的输出通过 1 个 3×3 卷积处理后,与 PANet 自下而上路径的最顶层输出 $[160, 160, 32]$ 特征图进行 Concat,生成尺寸为 $[160, 160, 64]$ 的特征图,作为新增的 1 个小目标检测头的输入,即第 4 个检测头。该新增检测头对小尺寸目标具有较高的敏感性,并且其结构设计能够有效缓解较大目标尺度差异对模型检测精度的不利影响。

多元分支融合自注意力的特征提取网络结构示意图如图 3 所示。

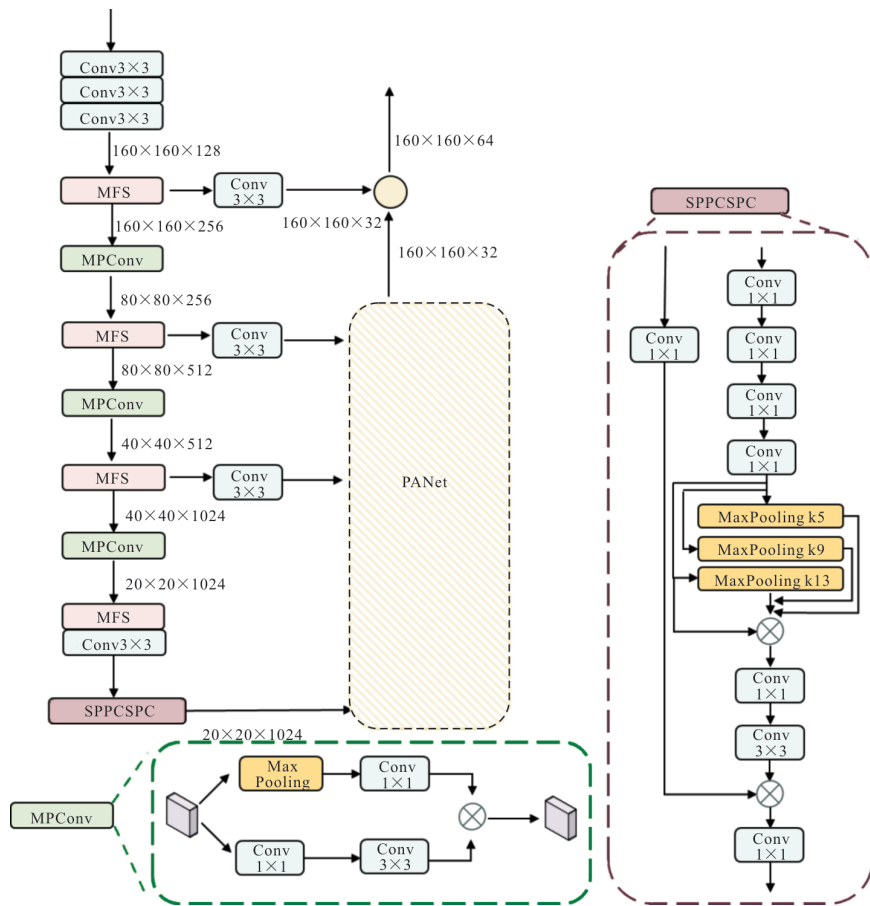


图 3 多元分支融合自注意力的特征提取网络结构示意图

1.2 基于辅助边框的 IoU 损失

现有 IoU (Intersection over Union) 损失仍集中在通过添加新的损失函数项以达到加速收敛的目的,并且泛化能力弱,因此,使用 Inner-IoU 损失^[17]结合 L_{CIoU} (Complete IoU Loss) 计算 IoU 损失。在

Inner-IoU 损失中,使用尺度比例因子控制辅助边界框的比例大小,由于辅助边界框与实际边界框之间的尺寸差异,回归过程中的 IoU 值变化趋势与实际边界框的 IoU 变化趋势一致,更加充分地反映实际边界框回归结果的质量,并且小规模辅助边界

框计算 IoU 损失,更容易使高 IoU 样本回归,实现加速收敛。

Inner-IoU 损失的原理示意图如图 4 所示,图中 b^{gt} 为真实框, b 为锚框,真实框和内部真实框的中心点为 (x_c^{gt}, y_c^{gt}) ,锚框和内部锚框的中心点为 (x_c, y_c) ,真实框的宽度和高度分别为 w^{gt} 和 h^{gt} ,锚框的宽度和高度分别为 w 和 h 。

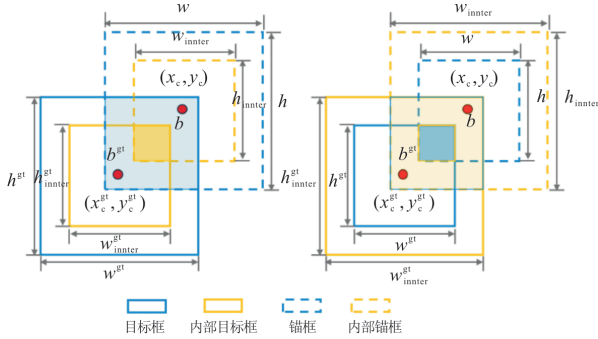


图 4 Inner-IoU 损失原理示意图

与 IoU 损失相比,当尺度因子小于 1 时,辅助框尺寸小于实际框,其回归的有效范围小于 IoU 损失,但其梯度绝对值大于 IoU 损失所得的梯度,能够加速高 IoU 样本的收敛。与之相反,当尺度因子大于 1 时,较大尺度的辅助边框扩大了回归的有效范围,对于低 IoU 的回归有所增益,因此选择尺度因子等于 1^[17]。

使用 Inner-IoU 结合 L_{CloU} ,边框回归损失函数的表达式为

$$L_{Inner-CIoU} = L_{CloU} + L_{IoU} + L_{Inner-CIoU} \quad (8)$$

式中: L_{IoU} 表示预测框与真实框之间的 IoU 损失; $L_{Inner-CIoU}$ 表示 Inner-IoU 损失。

1.3 基于 DepGraph 的结构化通道剪枝

为了平衡模型精度与检测效率之间的矛盾,使用基于 DepGraph 的结构化剪枝^[18]方法,降低模型的参数量和计算复杂度,从而得到轻量化的遥感图像目标检测算法。结构化通道剪枝原理示意图如图 5 所示,结构化通道剪枝通过先进行稀疏化训练,确定哪些通道“不重要”,即权重较小,再对其剪枝,从而在尽可能少损失检测精度的前提下,显著提升模型的检测速度,增强其在实际应用中的部署能力。

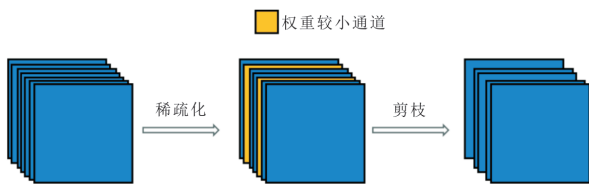


图 5 结构化通道剪枝原理图

向模型训练任务相关的损失函数 $l(f(x_i, \mathbf{W}), y_i)$

添加惩罚项,以获得稀疏化模型的表达式为

$$L = \sum_{i=1}^N l(f(x_i, \mathbf{W}), y_i) + \lambda R_s(\gamma) \quad (9)$$

式中: λ 为权重系数; γ 为可学习参数; $R_s(\cdot)$ 表示 L_1 范数计算。

批归一化(Batch Normalization, BN)层的可学习参数与滤波器的数量相等,可使用 BN 层作为通道剪枝的“门”,对通道重要性进行评估^[19],BN 层对输出原理的表达式为

$$z_{out} = \gamma \frac{z_{in} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \quad (10)$$

式中: z_{in} 和 z_{out} 分别为 BN 层的输入和输出; μ_B 和 σ_B 分别为当前 mini-batch 的平均值和标准差; ϵ 为防止分母为零的小量; γ 和 β 为可学习参数,分别表示尺度因子和位移。稀疏正则化算法使用 L_1 范数,有 $R_s(\gamma) = \|\gamma\|_1$,即给每个通道的 γ 尺度因子施加稀疏处理,可以将非重要通道的尺度因子趋于零,达到获得稀疏模型的目的。

2 实验结果与分析

2.1 实验环境搭建

实验操作平台为 Ubuntu 20.04 系统,采用 Pytorch1.10.0 深度学习框架,CPU 的型号为 AMD EPYC 7542 32-Core Processor,显卡的型号为 NVIDIA GeForce RTX A5000,显卡内存 24 GiB,编程语言为 Python。设置实验初始学习率为 0.001,采用随机梯度下降法(Stochastic Gradient Descent, SGD)更新网络参数,学习动量为 0.9,权重衰减率为 0.0005。

2.2 评价指标

采用平均精确率(Average Precision, AP)和平均精确率均值(Mean Average Precision, mAP)对多元分支融合自注意力的遥感图像目标检测模型的检测效果进行评估。AP 由召回率与查准率决定,以召回率为横坐标,每个召回率对应的查准率最大值为纵坐标,绘制出 Precision-Recall 曲线,对曲线取积分求曲线下面积即为平均精确率(Average Precision, AP)值。在得到多个单一类别值后,对其求平均值得到 mAP 值,mAP 值为所有类别平均精确率均值,通过 mAP 值衡量模型对所有类别的检测效果。最后,采用参数规模和浮点运算(Giga Floating Point Operations, GFLOPs)充分评估模型选取。

2.3 数据集

实验采用航拍图像(Dataset for Object deTec-tion in Aerial Image, DOTA)数据集^[20]和(North-western Polytechnical University Very High Reso-lution-10, NWPU VHR-10)数据集^[21]。DOTA 数据集一共有 2 806 张航拍图像,图像尺寸集中在 800 × 800 像素至 4 000 × 4 000 像素,将原图裁剪为多个 1 024 × 1 024 像素的子图,对于分辨率不足的图像,将其填充为 1 024 × 1 024 大小。数据集共包含 15 个类别,分别是飞机、船舶、储罐、棒球场、网球场、篮球场、跑道、港口、桥梁、大型车辆、小型车辆、直升机、环岛、足球场和游泳池。训练集、验证集和测试集按照 8 : 1 : 1 划分。桥梁、小型车辆、船舶、网球场、篮球场等类别的大小存在显著差异^[22],同

一物体表现出不同的大小,对精确检测造成了困难。

NWPU 数据集包含 800 张图像,数据集中分为一个具有标注的图像集和一个负图像集,负图像集共有 150 张图像,实验仅使用具有标注的 650 张图像,共 10 种类别,分别是飞机、船舶、储罐、棒球场、网球场、篮球场、跑道、港口、桥梁和车辆,将数据集随机划分为训练集 350 幅图像、验证集 150 幅图像和测试集 150 幅图像。

2.4 消融实验

采用消融实验在 DOTA 数据集上验证所提算法的性能,模型训练后使用测试集进行测试,共 4 组实验。以标准 YOLOv7 网络为基线模型,从网络的检测精度和参数规模等方面进行对比,消融实验结果如表 1 所示。

表 1 模型检测消融实验

实验	基线模型	Inner-IoU	MFS	第 4 个检测头	查准率/%	召回率/%	mAP/%	mAP _{0.5~0.95} /%	参数规模/M	GFLOPs
实验 1	✓				78.2	73.5	76.4	49.5	36.9	104.7
实验 2	✓	✓			78.6	73.5	76.6	50.3	36.9	104.7
实验 3	✓	✓	✓		78.6	74.3	77.3	50.0	39.1	134.3
实验 4	✓	✓	✓	✓	78.9	74.7	77.7	50.5	39.7	139.9

由表 1 可知,在实验 2 中,辅助边框 IoU 损失提升锚框精准度,显著提高港口目标检测精度,mAP_{0.5~0.95} 较 YOLOv7 提升 0.8%。实验 3 采用多元分支融合自注意力网络,增强特征提取,虽参数增加,但在跑道等场景达到最佳精度,召回率和 mAP 分别提高 0.8% 和 0.7%。实验 4 引入第 4 个检测头,小型车辆和储罐检测精度分别达 65.4% 和 83.9%,至少提升 1.2% 和 1.1%,且精确率、召回率和 mAP 最高,分别为 78.9%、74.7% 和 77.7%。因此,对比 YOLOv7 基线模型,所提算法在保证参数规模不大幅度上升的前提下,得到了优于 YOLOv7 的检测精度,验证了对于遥感目标检测的有效性。

2.5 剪枝实验

为更加准确地控制目标检测模型在稀疏化训练过程中的正则化强度,通过调整式(9)中的 λ 权重系数,以便获得同时兼具高稀疏度和高精度的检测模型。不同 λ 权重下模型检测精度如表 2 所示。

表 2 不同 λ 权重下模型检测精度/%

λ	查准率	召回率	mAP
0.001 00	72.8	63.0	65.4
0.000 75	73.4	68.7	70.0
0.000 50	76.9	70.1	72.9
0.000 25	76.6	67.9	71.4
0.000 10	71.3	66.9	67.2

由表 2 可知,当 $\lambda=0.001 00$ 时,正则化约束程度较大,模型不能够较好地保证其检测精度,逐步调整 λ 取值进行稀疏训练。当 $\lambda=0.000 50$ 时,查准率达到 76.9%,召回率达到 70.1%,mAP 达到 72.9%。因此最终选取 $\lambda=0.000 50$ 作为损失函数惩罚项的权重系数。

当 $\lambda=0.000 50$ 时,BN 层缩放因子 γ 分布情况如图 6 所示。

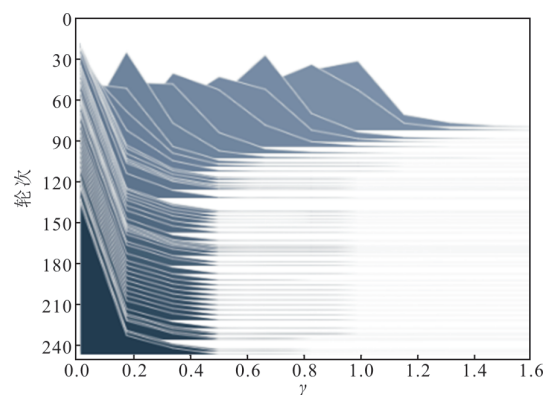


图 6 BN 层缩放因子 γ 分布情况

在原始模型中,也就是稀疏训练 0 次时, γ 接近正态分布,分布中心接近 1,此时无法判断通道的重要程度。随着稀疏训练的进行, γ 分布开始趋近于 0,在经过约 50 轮稀疏训练后,缩放因子 γ 大部分接近 0 并趋于稳定,此时,BN 层的 γ 越接近 0,代表此通道的重要程度越低,在剪枝过程中可以考虑被

剪去,达到模型压缩的目的。最终得到压缩后的剪枝 MFS-YOLO 的参数量为 6.64 M,计算量为 29.05 GFLOPs,mAP 为 72.9%。

2.6 对比实验

为了验证所提算法的优越性,并更全面地评价模型性能,在 DOTA 数据集上分别与 TPH-YOLOV5s^[23]

(Transformer Prediction Head YOLOV5 Small)、SSD^[24]、Retinanet^[25]、YOLOv8^[26]、YOLOv7^[14]、MPF^[27]、HyNet^[28] 和文献[29]算法进行对比实验,结果如表 3 所示。表中“*”表示可直接从论文中获得的结果;“—”表示在原始文本中没有提供相应值。

表 3 不同检测算法在 DOTA 数据集上的实验结果

算法	AP/%															mAP/%	参数规模/M
	飞机	棒球场	桥梁	跑道	小型车	大型车	船舶	网球场	篮球场	储罐	足球场	环岛	港口	游泳池	直升机		
TPH-YOLOv5	92.0	71.0	46.6	64.7	64.4	83.3	88.1	90.7	53.7	73.4	61.5	54.2	81.9	66.0	46.5	68.5	45.4
SSD	86.2	61.5	21.1	38.5	46.5	71.5	85.4	92.0	72.9	60.2	57.7	59.6	64.9	62.8	44.2	61.6	26.1
Retinanet	75.0	67.2	68.0	53.2	27.5	61.8	70.6	83.2	74.3	50.3	58.8	64.9	68.1	56.8	50.5	62.0	37.9
YOLOv8l	94.7	78.3	55.0	75.0	58.2	84.2	91.1	93.6	60.0	78.2	63.6	63.8	85.4	68.3	66.6	74.4	43.7
YOLOv7	95.2	79.0	57.5	76.4	64.0	86.6	90.2	94.3	70.9	82.8	74.6	64.2	85.1	70.5	54.0	76.4	36.9
文献[29]*	77.0	74.1	54.2	72.7	48.1	63.1	70.6	88.0	55.5	48.4	54.8	63.1	78.5	34.2	44.9	61.8	66.3
MPF*	92.8	76.9	50.9	68.0	66.9	86.4	88.1	94.5	69.9	73.0	55.5	64.5	85.2	66.9	59.1	73.3	—
HyNet*	86.9	58.7	43.7	54.0	64.4	80.5	87.8	85.1	53.8	60.4	41.8	47.6	76.6	48.7	40.3	62.0	—
所提算法	95.2	79.3	58.7	78.8	65.4	86.5	90.1	94.4	72.9	83.9	75.0	64.7	85.5	69.9	65.6	77.7	39.7

由表 3 可知,所提算法的 mAP 是 77.7%,为对比算法中的最优结果,证明了所提算法的有效性。TPH-YOLOv5 算法通过引入大量注意力机制提升了检测精度,尤其对小尺寸目标如小型车,mAP 为 64.4%,对比其他算法有一定检测能力,但多尺度目标,例如桥梁目标,检测能力仍有待提高。YOLOv8l 算法的直升机 mAP 为 66.6%,在所有对比算法上表现最优,其 43.7M 的较大参数规模和优秀的网络设计增强了目标与背景的区别能力,但训练成本显著增加。SSD 的整体 mAP 仅有 61.6%,算法因缺乏特征信息融合和特征图过小易导致信息丢失,检测能力较弱。Retinanet 算法的桥梁 mAP 为 68.0%,对于宽高比差异大的目标上检测精度高,但对小而密的目标,例如小型车 mAP 为 27.5%,检测能力最差。尽管文献[29]及 MPF、HyNet 等算法在解决遥感图像尺度变化问题上有所改进,但总体检测精度为 61.8%、73.3%、62.0%,对比其他算法仍精度较差。通过细化其特征提取网络的设计,进一步区分了对象信息和背景信息,使得所提算法在物体边界模糊的跑道和足球场类别上都达到了最佳的检测精度,分别为 78.8%和 75.0%。此外,所提算法对棒球场、储罐和港口的检测精度最高,准确率分别为 79.3%、83.9%和 85.5%。

为验证剪枝后的 MFS-YOLO 在同等规模模型的优越性,在 DOTA 数据集上分别与 YOLOV5m^[26]、

YOLOX-s^[30]、SDSDet^[31] (Self-Distillation Sampling for Object Detection)、文献[32]等不同轻量模型进行对比实验,结果如表 4 所示。表中“*”表示可直接从论文中获得的结果;“—”表示在原始文本中没有提供相应值。

表 4 不同轻量模型在 DOTA 数据集上的实验结果

模型	查准率/%	召回率/%	mAP/%	参数规模/M	GFLOPs
YOLOv5m	73.5	68.9	70.1	21.2	49.0
YOLOX-s	69.3	60.2	67.3	9.0	26.8
SDSDet*	—	—	67.5	4.8	11.3
文献[32]*	—	—	50.0	5.4	—
所提算法	74.7	72.3	72.9	6.6	29.0

由表 4 可知,剪枝后所提算法在参数规模仅有 6.64 M 的同时,mAP 达到 72.9%,为同规模对比算法中最优结果。通过剪枝获得的紧凑模型与通过轻量化网络结构获得的轻量模型相比,拥有更好的检测精度,证明了所提算法的有效性。

为了进一步验证所提算法在航拍数据集上的有效性和泛化性,在 NWPU VHR-10 数据集上分别与 TPH-YOLOv5^[23]、SSD^[24]、Retinanet^[25]、YOLOv8l^[26]、YOLOv7^[14]、FMSSD^[33] (Feature-Merged Single Shot MultiBox Detector) 和 YOLO-DCTI^[34] (You Only Look Once-Dynamic Convolutional Temporal Interpolation) 进行对比实验,结果如表 5 所示。表中“*”表示可直接从论文中获得的结果。

表 5 不同检测算法在 NWPU VHR-10 数据集上的实验结果

算法	AP/%										mAP/%
	飞机	船舶	储罐	棒球场	网球场	篮球场	跑道	港口	桥梁	车辆	
TPH-YOLOv5	99.0	85.1	99.1	97.8	88.0	66.0	98.5	93.2	80.1	76.9	88.4
SSD	99.9	90.8	98.0	99.1	85.0	89.0	98.5	93.2	85.7	88.3	92.7
Retinanet	99.9	79.3	85.3	92.0	94.1	93.7	94.6	97.4	89.7	75.9	90.8
YOLOv8l	99.4	88.6	92.0	97.3	97.7	97.1	99.5	96.2	89.8	89.2	94.7
YOLOv7	99.5	89.7	96.6	98.8	94.7	92.7	99.5	93.2	85.0	91.2	94.1
FMSSD*	99.7	89.9	90.3	98.2	86.0	96.8	99.6	75.6	80.1	88.2	90.4
YOLO-DCTI*	99.6	93.0	96.8	99.5	90.9	95.0	99.2	91.5	98.9	90.3	95.5
所提算法	99.3	93.5	94.9	99.3	98.3	98.3	99.5	95.7	92.3	94.3	96.5

由表 5 结果可知,FMSSD 使用空洞卷积以及多个平行层的卷积统一来自不同特征图的上下文信息,极易导致对特定尺度目标关注不足,如港口 mAP 为 75.6%、网球场 mAP 为 86.0%,尤其是当目标尺度与感受野不匹配时;YOLO-DCTI 设计 DCTI 结构旨在解耦分类和回归任务,检测精度达到 95.5%,但解耦不够有效,导致任务之间的信息共享不足。所提算法在 NWPU VHR-10 数据集上达到了 96.5%的最佳检测精度。

通过两组对比实验可知,所提算法在 DOTA 数据集和 NWPU VHR-10 数据集检测任务中,均达到最佳检测精度。

3 结语

多分支融合自注意力机制的遥感图像目标检测算法设计了一种融合卷积与自注意力机制的多分支模块,构建了高效的特征提取网络,并针对遥感图像中普遍存在的小目标检测难题,引入了第 4 个专用检测头以增强对小尺度物体的感知能力。实验结果表明,该算法在 DOTA 数据集上实现了 77.7%的 mAP,在 NWPU VHR-10 数据集上达到了 96.5%的 mAP,展现出优异的检测性能。为进一步提升模型效率,采用结构化剪枝技术对网络进行压缩,压缩后模型参数量仅为 6.64M,且在 DOTA 数据集上的 mAP 仍保持在 72.9%,显著优于同规模的其他检测模型,兼具高精度与轻量化优势。

参 考 文 献

[1] CHENG G, HAN J W. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117: 11-28.

[2] WANG H N, LI Y, FANG Y Q, et al. SRS-Net: Training object detectors from scratch for remote sensing images without pretraining[J]. Chinese Journal of Aeronautics, 2023, 36(8): 269-283.

[3] 黄泽贤, 吴凡路, 傅瑶, 等. 基于深度学习的遥感图像舰船目标检测算法综述[J]. 光学精密工程, 2023, 31(15): 2295-2318.

HUANG Z X, WU F L, FU Y, et al. Review of deep learning-based algorithms for ship target detection from remote sensing images[J]. Optics and Precision Engineering, 2023, 31(15): 2295-2318. (in Chinese)

[4] ZHANG Y, SONG C L, ZHANG D W. Small-scale aircraft detection in remote sensing images based on Faster-RCNN[J]. Multimedia Tools and Applications, 2022, 81(13): 18091-18103.

[5] 张哲益, 曹卫华, 朱蕊, 等. 基于脉冲卷积神经网络稀疏表征的高分辨率遥感图像场景分类方法[J]. 控制与决策, 2022, 37(9): 2305-2313.

ZHANG Z Y, CAO W H, ZHU R, et al. Sparse representation with spike convolutional neural networks for scene classification of remote sensing images of high resolution[J]. Control and Decision, 2022, 37(9): 2305-2313. (in Chinese)

[6] 谢星星, 程塔, 姚艳清, 等. 动态特征融合的遥感图像目标检测[J]. 计算机学报, 2022, 45(4): 735-747.

XIE X X, CHENG G, YAO Y Q, et al. Dynamic feature fusion for object detection in remote sensing images[J]. Chinese Journal of Computers, 2022, 45(4): 735-747. (in Chinese)

[7] 周葳楠, 吴治海, 张正道, 等. 基于弱特征增强的轻量化小目标检测方法[J]. 控制与决策, 2024, 39(2): 381-390.

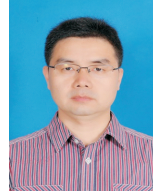
ZHOU W N, WU Z H, ZHANG Z D, et al. Lightweight small target detection method based on weak feature enhancement[J]. Control and Decision, 2024, 39(2): 381-390. (in Chinese)

[8] 严春满, 王铖. 基于特征增强的 SAR 图像舰船小目标

- 检测算法[J]. 控制与决策, 2023, 38(1): 239-247.
- YAN C M, WANG C. A ship small target detection algorithm based on feature enhancement in SAR image [J]. Control and Decision, 2023, 38(1): 239-247. (in Chinese)
- [9] 田中原. 遥感图像多尺度目标的轻量化检测方法[J]. 测绘科学, 2023, 48(6): 104-111.
- TIAN Z Y. Lightweight detection method for multi-scale objects in remote sensing images[J]. Science of Surveying and Mapping, 2023, 48(6): 104-111. (in Chinese)
- [10] AHMED M, EL-SHEIMY N, LEUNG H, et al. Enhancing object detection in remote sensing: A hybrid YOLOv7 and transformer approach with automatic model selection[J]. Remote Sensing, 2024, 16(1): 51-68.
- [11] 赵同祥, 张瑞全, 高树静, 等. 基于 DN-YOLOv5 遥感目标快速检测方法[J]. 电子设计工程, 2024, 32(4): 186-190.
- ZHAO T X, ZHANG R Q, GAO S J, et al. A fast detection method of remote sensing target based on DN-YOLOv5[J]. Electronic Design Engineering, 2024, 32(4): 186-190. (in Chinese)
- [12] 庄文华, 唐晓刚, 张斌权, 等. 基于改进 YOLOv5 的遥感图像旋转框目标检测[J]. 电子设计工程, 2023, 31(14): 137-141.
- ZHUANG W H, TANG X G, ZHANG B Q, et al. Remote sensing image rotatable bounding box object detection based on improved YOLOv5[J]. Electronic Design Engineering, 2023, 31(14): 137-141. (in Chinese)
- [13] 兰旭婷, 郭中华, 石甜甜, 等. 融合 SPP 与 FPN 的光学遥感图像飞机目标检测[J]. 光电与控制, 2023, 30(4): 6-11.
- LAN X T, GUO Z H, SHI T T, et al. Aircraft target detection in optical remote sensing images by fusing SPP and FPN [J]. Electronics Optics & Control, 2023, 30(4): 6-11. (in Chinese)
- [14] 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]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 7464-7475.
- [15] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759-8768.
- [16] DING X H, ZHANG X Y, MA N N, et al. RepVGG: Making VGG-style ConvNets great again [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13728-13737.
- [17] ZHANG H, XU C, ZHANG S J. Inner-IoU: More effective intersection over union loss with auxiliary bounding box[EB/OL]. [2024-10-13]. <https://arxiv.org/abs/2311.02877>.
- [18] FANG G F, MA X Y, SONG M L, et al. DepGraph: Towards any structural pruning[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 16091-16101.
- [19] LIU Z, LI J G, SHEN Z Q, et al. Learning efficient convolutional networks through network slimming [C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2755-2763.
- [20] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3974-3983.
- [21] CHENG G, HAN J W, ZHOU P C, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98: 119-132.
- [22] RAN Q, WANG Q, ZHAO B Y, et al. Lightweight oriented object detection using multiscale context and enhanced channel attention in remote sensing images [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 5786-5795.
- [23] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal: IEEE, 2021: 2778-2788.
- [24] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[C]//Computer Vision-ECCV 2016. Cham: Springer, 2016: 21-37.
- [25] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2999-3007.
- [26] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS[J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.

- [27] HUANGFU P P, DANG L X. A multi-scale pyramid feature fusion-based object detection method for remote sensing images [J]. *International Journal of Remote Sensing*, 2023, 44(24): 7790-7807.
- [28] ZHENG Z, ZHONG Y F, MA A L, et al. HyNet: Hyper-scale object detection network framework for multiple spatial resolution remote sensing imagery [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 166: 1-14.
- [29] 吕奕龙, 李敏, 吴肇青, 等. 稠密连接递归特征金字塔的遥感目标检测算法 [J]. *遥感学报*, 2024, 28(6): 1602-1614.
- LYU Y L, LI M, WU Z Q, et al. Object detection in remote sensing images using densely connected recursive feature Pyramids [J]. *National Remote Sensing Bulletin*, 2024, 28(6): 1602-1614. (in Chinese)
- [30] GE Z, LIU S T, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [EB/OL]. [2024-10-13]. <https://arxiv.org/abs/2107.08430>.
- [31] LIU J P, ZHENG K Y, LIU X Y, et al. SDSDet: A real-time object detector for small, dense, multi-scale remote sensing objects [J]. *Image and Vision Computing*, 2024, 142: 104898-104913.
- [32] 刘富宽, 罗素云, 何佳, 等. FVIT-YOLO v8: 基于多尺度融合注意机制的改进 YOLO v8 小目标检测 [J]. *红外技术*, 2024, 46(8): 912-922.
- LIU F K, LUO S Y, HE J, et al. FVIT-YOLOv8: Improved YOLOv8 small object detection based on multi-scale fusion attention mechanism [J]. *Infrared Technology*, 2024, 46(8): 912-922. (in Chinese)
- [33] WANG P J, SUN X, DIAO W H, et al. FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(5): 3377-3390.
- [34] MIN L T, FAN Z M, LV Q Y, et al. YOLO-DCTI: Small object detection in remote sensing base on contextual transformer enhancement [J]. *Remote Sensing*, 2023, 15(16): 3970-3988.

[作者简介]



亢红波(1974—),男,陕西凤翔人,硕士,西安邮电大学副教授,主要研究方向为智能控制与信息集成。E-mail: khb2000@xup t. edc. cn



温家正(1998—),男,陕西榆林人,西安邮电大学硕士研究生,主要研究方向为遥感图像智能处理。E-mail: wjz19980220 @163. com



杨春杰(1976—),女,甘肃白银人,硕士,西安邮电大学副教授,主要研究方向为智能控制、物联网技术、嵌入式系统开发等。E-mail: ycj@xupt. edu. cn



王文庆(1964—),男,北京房山人,博士后,西安邮电大学教授,主要研究方向为复杂系统结构分析与鲁棒控制、智能信息处理、信息系统分析等。E-mail: wwq@xupt. edu. cn

[责任编辑:祝 剑]