

DOI:10.20176/j.cnki.nxdz.000112

基于大核选择和形状自适应的遥感图像目标检测

赵子澳¹, 董爱华^{1,2*}, 黄 荣^{1,2}

(1. 东华大学 信息科学与技术学院, 上海 201620; 2. 东华大学 数字化纺织服装技术教育部工程研究中心, 上海 201620)

摘要: 光学遥感图像目标检测是遥感图像数据智能解译的关键技术。为了解决遥感图像目标检测时,目标尺度差异大,目标受背景因素干扰,目标形状各异的问题,提出了LMK(large multiscale kernel)网络。该网络通过大核卷积分解和多尺度注意力机制模块,能够动态调整空间感受野,从而更好地捕获遥感场景中物体的上下文信息。此外,设计了一种面向目标检测的形状自适应选择(SAS, shape-adaptive selection)标签分配策略。该策略将目标形状信息集中于长宽比,通过结合物体的形状信息和特征分布计算IoU(intersection over union)最优阈值。针对遥感图像目标姿态旋转定位难的问题,引入了KFIoU损失函数。实验结果表明,所提出的目标检测模型在HRSC2016、UCAS-AOD和DOTA数据集上的精度分别达到了96.73%、97.85%和77.26%。改进后的模型优于目前绝大多数目标检测算法。

关键词: 目标检测;深度学习;标签分配;多尺度注意力;大核网络

分类号: (中图)TP391.4

文献标志码: A

可见光遥感图像中自动定位与识别有价值目标(如飞机、舰船、桥梁等),是遥感图像处理领域一个重要的研究方向^[1]。当前,针对光学遥感图像的目标检测问题,大体可以分为传统检测算法和基于深度学习的检测算法。传统检测算法往往需要人工选择并设计合适的特征,这些方法通常计算量大、时间复杂度高,难以满足实时系统的应用需求^[2]。近年来,深度学习的快速发展给计算机视觉领域的图像目标识别和检测带来了巨大突破^[3],特别是基于卷积神经网络的目标检测方法在图像解译方面取得了显著成果,在许多目标检测任务中甚至可以达到与人工判读相媲美的高精度。因此,基于深度学习的目标检测方法也逐渐被引入光学遥感图像目标检测的应用领域。

最近的研究表明,设计具有大感受野的卷积网络可以显著提高图像识别任务的性能。例如,ConvNeXt^[4]在其主干中使用了7×7深度卷积,从而显著提高了下游任务的性能。此外,RepLkNet^[5]借助重新参数化技术,采用31×31的卷积内核,实现了颇具说服力的性能提升。随后,SLaK^[6]通过核分解和稀疏组技术,进一步将核大小扩展到51×51。姚宗亮

等^[7]通过多模态融合技术学习了不同模态、不同尺度的特征变化。类似地,SegNeXt^[8]和Conv2Former^[9]的研究表明,大核卷积在调制卷积特征以获取更丰富的上下文信息方面发挥着重要作用。尽管大核卷积在一般目标识别领域已经受到了广泛关注,但在具体的遥感检测领域,对其意义的研究仍相对较少。

近年来,人们提出了许多优秀的动态样本选择策略。例如,Meta Anchor^[10]是一种锚框函数,能够基于先验框生成自适应锚框。DAL^[11]根据定义的匹配度动态分配锚点,从而综合评估锚框的定位能力。Free Anchor^[12]是一种学习匹配方法,允许目标在最大似然原则下动态选择锚点。PAA^[13]是一种新颖的标签分配策略,它以概率方式自适应地将锚点划分为正样本和负样本,划分依据是锚点与真实边界框的匹配程度。尽管这些自适应策略实现了样本的动态选择,但大多忽略了目标形状信息与所选正样本之间的差异。

针对上述问题,Li等^[14]提出了大核选择性网络LMK Net(改进自LSK Net网络)。LMK Net能够动态调整其大空间感受野,使检测器可以更好地捕获遥感场景中物体的上下文信息,从而有效解决微小目标的漏检和误检问题。本文提出了一种形状

收稿日期:2024-11-03

基金项目:国家自然科学基金资助项目(62001099);中央高校基本科研业务费专项资金资助项目(2232023D-30)

作者简介:赵子澳(1999—),男,硕士研究生,主要从事遥感图像检测研究,(电子信箱)2549453530@qq.com。

*通信联系人:董爱华(1970—),女,副教授,博士,主要从事智能纺织服装研究,(电子信箱)dongaihua@dhu.edu.cn。

引用格式:赵子澳,董爱华,黄荣.基于大核选择和形状自适应的遥感图像目标检测[J].宁夏大学学报(自然科学版中英文),2026,47(1):33-41.

自适应(SAS, shape-adaptive selection)动态标签分配策略。该策略使检测模型能够根据目标的形状和特征分布选择正样本,从而解决了遥感目标高宽比差异大以及形状各异的问题。此外,本文引入了基于卡尔曼滤波的近似 SkewIoU 损失函数和 KFIoU 损失函数^[15]。这些损失函数的可微性,使得模型能够更好地处理基于角度信息的旋转目标检测问题,避免因边界不连续而导致性能下降。

1 基于大核选择和形状自适应的目标检测模型

首先,引入大核选择性网络(LMKNet),替换

YOLOv8骨干网络中的CBS模块。在遥感图像中,准确检测目标往往需要广泛的上下文信息,而不同目标类型所需的上下文信息范围差异很大。LMK网络可以针对目标调整感受野,从而更准确地提取遥感图像目标特征。其次,由于大多数标签分配策略忽略了物体的高宽比和形状信息,引入SAS动态选择标签分配策略。该策略根据目标的形状动态选择样本用于细化检测头,可以根据物体的高宽比和形状信息动态标记正样本,其中实线框和虚线框分别表示阳性和阴性样本。最后,为了解决遥感图像姿态旋转定位难的问题,引入KFIoU旋转损失函数。改进后的模型结构如图1所示。

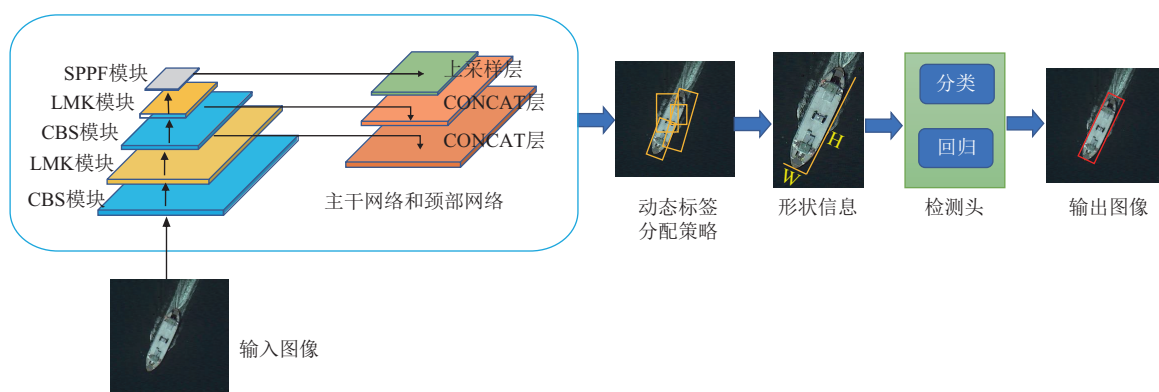


图1 基于大核网络和形状自适应标签分配的目标检测模型

Fig. 1 Object detection model based on large kernel network and shape-adaptive label assignment

1.1 大核选择性网络

在遥感图像目标检测任务中,遥感目标的尺度往往差异较大。对于微小的遥感目标,如果没有参考足够广泛的背景信息,可能会被错误地检测到。

而不同类型的目标所需的背景范围可能也有所不同。本文提出的大核选择性网络(LMKNet),其核心由大核卷积和多尺度注意力机制组成(见图2)。

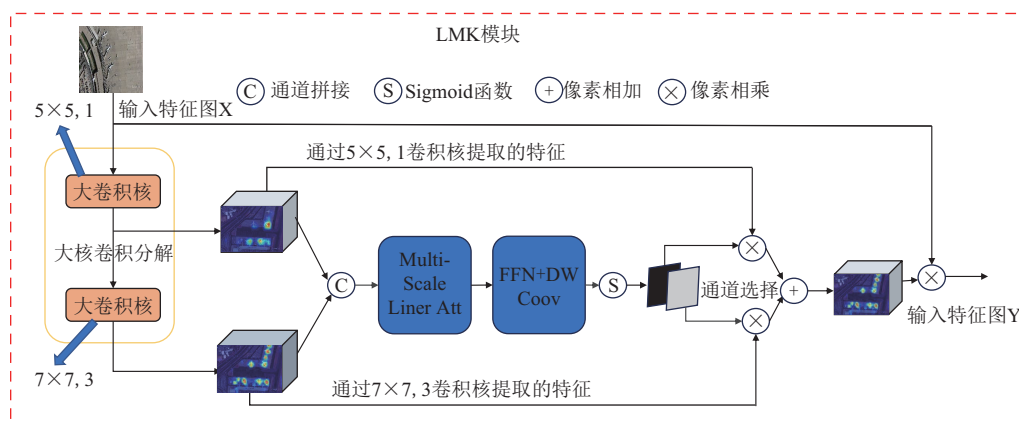


图2 LMKNet模型结构图

Fig. 2 Architecture diagram of the LMKNet model

在遥感图像中,不同目标类型所需的上下文信息范围差异较大。为了更有效地捕获广泛的上下文信息,可以构建一个更大的卷积核,并将其分解为具有大增长核和增加膨胀率的深度卷积序列。由于准确检测遥感目标往往需要广泛的上下文信息,核大

小和膨胀率的增加能够确保感受野的扩张速度足够快。具体而言,本文将一个 23×23 ,空洞率为1的大卷积核分解为一个 5×5 ,空洞率为3的卷积核和一个 7×7 ,空洞率为1的卷积核,其理论感受野为23。通过将这两个分解后的卷积核串联起来,输入特征

图 X 可以获得较大的感受野以及不同感受野的特征图。将得到的特征图在通道维度上拼接,然后分别通过多尺度注意力机制生成两张 $H \times W \times 1$ 的特征图,并再次在通道维度上拼接。接着,通过一个卷积层和 Sigmoid 函数,生成两张空间选择掩码。将这两个掩码与对应感受野的特征图逐像素相乘,再将相乘后的两组特征图逐像素相加,并通过卷积层融合获得多尺度注意力图。最后,将空间注意力图与输入特征图 X 逐像素相乘,从而实现对感兴趣区域的增强和对多余背景信息的抑制。

将大卷积核分解为两个卷积核(见表 1)。研究表明,过小或过大的感受野都会影响 LSKNet 的性

表 1 大卷积核的不同分解方法比较

Table 1 Comparison of different decomposition methods for large convolutional kernels

(k1, d1)	(k2, d2)	RF	FPS	mAP/%
(3,1)	(5,2)	11	23.1	80.83
(5,1)	(7,3)	23	21.8	80.96
(7,1)	(9,4)	39	22.3	80.85
(5,1)	(7,3)	23	19.7	81.34

能,而感受野大小约为 23 时效果最佳。此外,实验表明,将大卷积核分解为两个卷积核,分别为 5×5

的卷积核(空洞率为 1)和 7×7 的卷积核(空洞率为 3),这种方式更为有效。

所提出的设计具有两个优点。首先,它能够显式地生成多个具有较大感受野的特征,从而使得后续的内核选择更加容易。其次,顺序分解的方式比简单地应用一个大卷积核更为有效。在理论感受野相同的条件下,与标准大卷积核相比,文中提出的分解方法可显著减少模型参数量。为了从输入 X 中提取不同范围的丰富上下文特征,本文采用了一系列具有不同感受野的深度卷积分解策略,具体表示为

$$U_0 = X, U_{i+1} = F_i^{DW}(U_i) \quad (1)$$

其中: F_i^{DW} 是具有 $i \times i$ 大小和空洞率 DW 的卷积核。通过两个卷积核分别对特征 X 进行卷积操作后,将得到的特征进行拼接,允许每个空间特征向量的通道混合,进而提取多尺度的目标特征。在此基础上,本文提出了一种基于多尺度注意力机制的目标特征选择机制(见图 3)。该机制由一个多尺度线性注意力模块和一个深度卷积的前馈网络(FFN+DWConv)组成。其中,多尺度线性注意力模块用于捕获上下文信息,而 FN+DWConv 用于捕获局部信息。

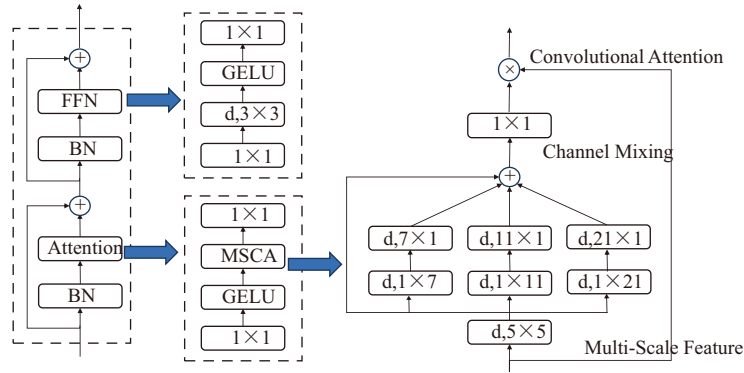


图 3 多尺度注意力机制

Fig. 3 Mechanism of multi-scale attention

如图 3 所示,多尺度通道注意力(multi-scale-channel-attention, MSCA)模块包含 3 个部分:用于聚合局部信息的深度卷积,用于捕获多尺度上下文的多分支深度卷积,以及用于模拟不同通道之间关系的 1×1 卷积。将 1×1 卷积的输出作为注意力权重,直接对 MSCA 的输入进行加权。在数学上, MSCA 可以表示为

$$Att = \text{Conv}_{1 \times 1} \left(\sum_{i=0}^3 \text{Scale}_i(\text{DW-Conv}(F)) \right), \quad (2)$$

$$\text{Out} = Att \otimes F. \quad (3)$$

其中: F 表示输入特征; Att 和 Out 分别表示注意图和

输出; \otimes 表示元素的矩阵乘法运算。DW-Conv 表示深度卷积, Scale_i 表示第 i 个分支。在每个分支中,使用两个深度分离卷积近似具有大核的标准深度分条卷积。这里,每个分支的卷积核大小分别设置为 7, 11 和 21。选择深度条形卷积的原因有两个:一方面,条形卷积是轻量级的,为了模拟核大小为 7×7 的标准 2D 卷积,只需要一对 7×1 和 1×7 的卷积;另一方面,检测场景中存在一些高宽比物体(如航母、货船等)。因此,条形卷积可以作为网格卷积的补充,有助于提取条形特征。为了展示 LMKNet 的效果,利用 Grad-CAM 对注意力分配进行了可视化(见图 4)。

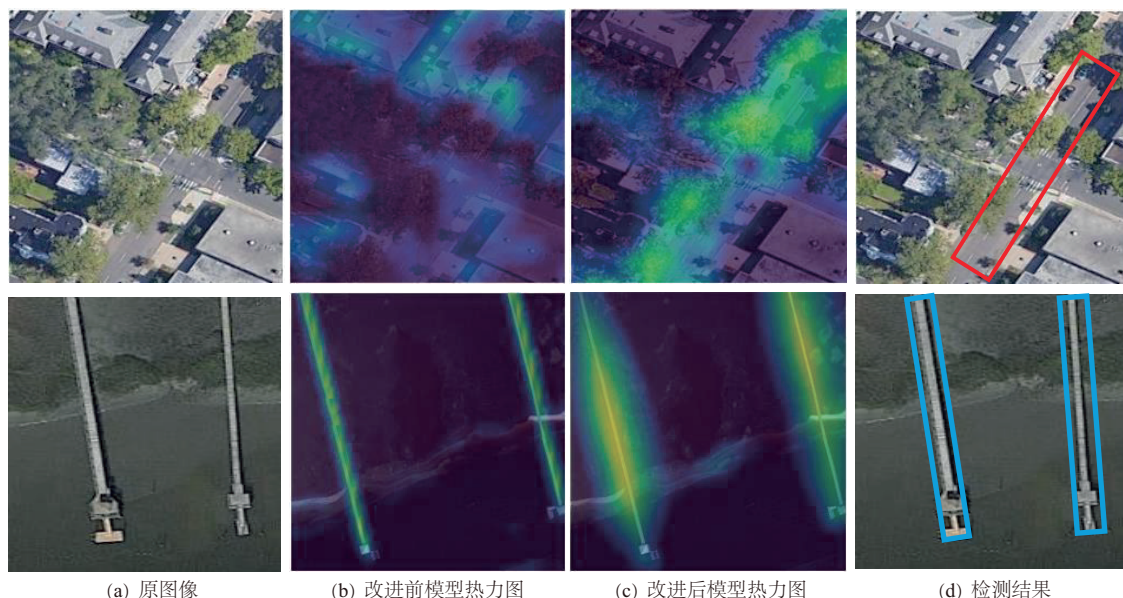


图 4 注意力可视化结果图

Fig. 4 Attention visualization results

由图 4 可知,文中提出的 LMKNet 能够对大范围上下文信息进行建模,进而在各类复杂场景下取得更优性能。未加入该模块前,网络无法准确识别十字路口;受路口周边树木遮挡的影响,该区域被错误判别,究其原因在于模型对特征的捕获能力不足,目标识别精度欠佳。引入大核网络模块后,模型可同时捕获目标的全局特征与局部特征,实现对目标的精准识别。在港口检测任务中,由于港口目标的检测需依赖较大感受野,未加入 LMK 模块时模型存在误检现象,全局特征提取能力存在明显短板;而加入该模块后,检测性能得到显著改善。

1.2 自适应动态标签分配策略 SAS

针对遥感图像中目标存在高宽比大、形状各异的问题,Zhang 等^[16]提出了 ATTS(adaptive training sample selection)标签分配策略,并指出对象的纵横比越大,IoU 阈值越低,性能越好。这可能是因为在不同形状下,IoU 值对定位误差的敏感性不同。传统的基于 IoU 的样本选择策略对所有对象使用相同的预定义 IoU 值,这在面对多种形状的对象时,难以有效挖掘高质量样本用于多类目标检测,尤其是当目标形状差异较大时。为此,本文提出了一种 SAS(shape-aware sample selection)动态标签分配策略。

在 YOLOv8 检测头中,采用了基于 IoU 的样本选择策略。然而,这种基于 IoU 的策略忽略了对象的形状,并使用相同的固定规则处理所有对象。虽然这种方法可能适用于大多数对象,但它忽略了一些具有特殊形状的对象。为了优化样本选择过程,文中提出了一种 SAS 策略,该策略根据目标的形状和特征分布自

适应调整 IoU 阈值,以实现高质量样本的精准筛选。选择样本的阈值 T_i^{IoU} 为

$$T_i^{\text{IoU}} = f(r_i) \times S^\alpha \times \mu^\beta. \quad (4)$$

其中: r_i 表示与预测框相对应的真实框的宽高比,其计算方法为长边与短边之比。根据前面的分析可知,权重应该随着长宽比的增加而减小,这样拉长的物体就会被分配一个较低的 IoU 阈值。因此,设计了加权系数的单调递减函数为

$$f(r_i) = e^{-\frac{r_i}{\omega}}. \quad (5)$$

其中: $f(r_i)$ 为目标的加权因子函数; ω 是一个加权参数,经验值默认为 4。当数据集包含大量拉长的对象时,较大的 ω 通常会获得更好的性能。使用公式对每个 GT(ground truth)框实例计算 anchor-level 的对齐程度: S 和 μ 分别为分类得分和 IoU 值, α 和 β 均为权重超参数(YOLOv8 默认值为 $\alpha = 0.5$ 、 $\beta = 6.0$)。将两者相乘即可衡量预测框与真实框的匹配程度,分类得分越高且 CIoU 越大, T_i^{IoU} (匹配程度)的值就越接近于 1,此时预测框与真实框的匹配效果越好,也越符合正样本的筛选标准。该标签分配策略的算法步骤如下。

动态标签分配算法

Input:输入图像 X ,锚框集 A 。

Output:更新后的训练标签 T_i^{IoU} 。

1. 计算预测类别分值: $\text{pred}_{\text{scores}} = \text{Forward}(X, A) \%$, $\text{pred}_{\text{scores}} = F_{\text{sigmoid}}(\text{pred}_{\text{scores}})$ 。
2. 去除重复分配:通过 mask 矩阵与 anchor 求和,找出分给多个 GT 框的 anchor,只保留 CIoU 值

最大的 anchor。

3. 获得每个样本对应的训练标签为

$$T_i^{\text{IoU}} = f(r_i) \times S^a \times \mu^b。$$

4. 返回更新的训练标签 T_i^{IoU} 。

实验结果见图 5。从图 5 可以看出,ATTS 动态标签分配策略在处理高宽比较大的目标时存在误

检和漏检问题。这是因为在原标签分配策略下,当目标的高宽比较大时,所需的 IoU 阈值较低,而给定的 IoU 阈值无法满足这一要求。因此,在引入 SAS 动态标签分配策略后, $f(r_i)$ 可以根据目标的形状动态计算所需的 IoU 阈值,从而更准确地筛选出正样本,进而提高检测精度。

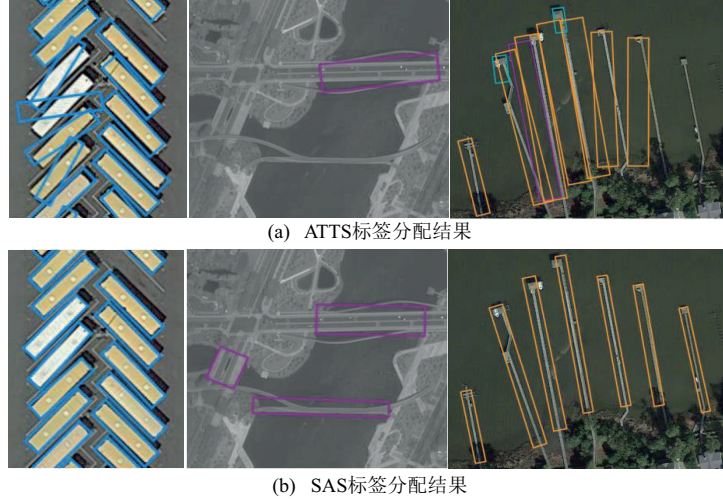


图 5 ATTS 标签分配和 SAS 标签分配对比结果

Fig. 5 Comparison results of ATTS label assignment and SAS label assignment

2 实验结果与分析

2.1 数据集

本文在 HRSC 2016, DOTA 和 UCAS-AOD 三个公共遥感数据集上进行了实验。这些数据集中的 GT 框均使用旋转边界框标注。HRSC 2016 是一个高分辨率遥感船舶检测数据集,包含 1 061 张图像,图像大小范围为 300 px×300 px 至 1 500 px×900 px。整个数据集被分为训练集、验证集和测试集,分别包含 436 张、181 张和 444 张图像。UCAS-AOD 是一个航空飞机和汽车检测数据集,包含 1 510 张来自 Google Earth 的图像,其中飞机图像 1 000 张、汽车图像 510 张。由于该数据集官方未提供划分,文中将其按 5:2:3 的比例随机分为训练集、验证集和测试集。在实验中,UCAS-AOD 中的所有图像均被调整为 800 px×800 px。DOTA 是遥感图像中定向目标检测的最大公开数据集,包含 2 806 张航空图像和 188 282 个注释实例,涵盖 15 个类别,分别为:飞机 (PL)、棒球场 (BD)、桥梁 (BR)、田径场 (GTF)、小型车辆 (SV)、大型车辆 (LV)、船舶 (SH)、网球场 (TC)、篮球场 (BC)、储油罐 (ST)、足球场 (SBF)、环岛 (RA)、港口 (HA)、游泳池 (SP) 和直升机 (HC)。由于 DOTA 中的图像尺寸较大,实验中将原始图像裁剪为 800 px×800 px 的块,步幅为 200,用于训练和测试。

2.2 模型训练环境

实验环境配置如下:CPU 为 Intel Core i7-10700K,主频 3.8 GHz(实际运行频率 3.79 GHz),内存为 96 GB,GPU 为 GeForce RTX 3090,显存为 24 GB。使用的深度学习框架为 PyTorch 2.1,编程语言为 Python 3.9。在 DOTA, HRSC2016 和 UCAS-AOD 数据集上分别进行了 250, 300 和 300 个训练轮数。训练过程中采用随机梯度下降法 (stochastic gradient descent, SGD),动量和权重衰减分别设置为 0.9 和 5×10^{-4} 。学习率采用余弦退火策略,计算公式为初始学习率 $l_r \times$ 批量大小 $\text{batchsize}/64$,其中初始学习率为 0.01。由于 DOTA 数据集中不同类别的实例分布极不均衡,因此将类别损失函数修改为焦点损失 (focal loss),以缓解类别不平衡带来的影响。

2.3 模型对比实验

在 HRSC2016, DOTA 和 UCAS-AOD 数据集上,分别探讨了多种先进的遥感图像目标检测方法的准确性。

为了进一步验证改进后 YOLOv8 模型的有效性,在 UCAS-AOD 数据集上进行了一系列实验,结果见表 2。基于改进 YOLOv8 的方法在两个类别上均取得了最高的平均精度 (AP) 值,分别为 99.09% 和 96.51%,并且在 mAP50 指标上优于目前所有先进方法,充分证明了所提方法的性能更优越。

表 2 不同检测方法在 UCAS-AOD 上的检测精度比较

Table 2 Comparison of detection accuracy of different detection methods on UCAS-AOD

方法	飞机	汽车	%
			mAP
R3Det ^[17]	98.20	94.14	96.17
RSdet ^[18]	98.04	94.97	96.50
PolarDet ^[19]	99.08	96.26	97.02
DAL ^[11]	89.25	90.49	89.87
Ours	99.09	96.51	97.80

对 HRSC 2016 数据集的实验结果采用 mAP50 指标进行评估,以便公平地比较不同方法的性能。如表 3 所示,改进的 YOLOv8 模型在 mAP50 指标下达到了 96.34% 的精度,优于 R3Det 等现有的旋转目标检测方法。更重要的是,在该模型下,所提方法在 RTX 3090 GPU 上实现了 42 FPS 的快速检测,这充分说明所提方法在面对大纵横比的目标时仍具有显著优势。通过上述结果可以看出,改进的 YOLOv8 模型能够在 HRSC 2016 数据集中实现准确且高效的旋转目标检测。

DOTA 是目前公开的最全面的用于目标检测的

表 3 不同检测方法在 HRSC2016 上的检测精度

Table 3 Detection accuracy of different detection methods on HRSC2016

方法	骨干网络	FPS	mAP/%
R3Det ^[17]	ResNet101	12.0	96.01
R3Det-DCL ^[20]	ResNet101	—	96.41
S2ANet ^[21]	ResNet101	12.7	95.01
AOPG ^[22]	ResNet101	—	96.22
DAL ^[11]	ResNet101	—	89.77
Ours	YOLOv8	42	96.34

遥感图像数据集,不仅包含车辆、船只等小目标,还包含桥梁等大长宽比的目标。这为检验旋转目标检测算法的性能提供了良好的基础。为了评估本文方法的性能,实验中所有方法均使用相同的数据作为训练集,实验结果来源于 DOTA 评估服务器的测试结果。在 DOTA v1.0 数据集上的实验结果见表 4。由表 4 的检测结果可知,改进的 YOLOv8 模型在 DOTA v1.0 上取得了 76.13% 的检测精度,优于目前绝大多数旋转目标检测器。这充分说明改进的 YOLOv8 模型能够为遥感目标分配更合适的标签,从而提升目标检测的精度。其在 DOTA v1.0 上的检测结果见图 6。

表 4 不同检测方法在 DOTA 上的检测精度

Table 4 Detection accuracy of different detection methods on DOTA

方法																%
	PL	DB	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
DRN ^[23]	88.91	80.22	43.52	63.35	73.48	70.69	84.94	90.14	83.85	84.11	50.12	58.41	67.62	68.60	52.50	70.70
R3Det ^[17]	88.76	83.09	50.91	67.27	76.23	80.39	86.72	90.78	84.68	83.24	61.98	61.35	66.91	70.63	53.94	73.79
DAL ^[11]	88.68	76.55	45.08	66.80	67.00	76.76	79.74	90.84	79.54	78.45	57.71	62.27	69.05	73.14	60.11	71.44
AOPG ^[22]	89.14	82.74	51.87	69.28	77.65	82.42	88.08	90.89	86.26	85.13	60.60	66.30	74.05	67.76	58.77	75.39
S2A ^[21]	89.30	80.11	50.97	73.91	78.59	77.34	86.38	90.91	85.14	84.84	60.45	66.94	66.78	68.55	51.65	74.13
SASM ^[24]	86.42	78.97	52.47	69.84	77.30	75.99	86.72	90.89	82.63	85.66	60.13	68.25	73.98	72.22	62.37	74.92
Ours	89.08	85.41	52.41	75.02	76.28	78.11	89.21	84.47	80.52	86.79	61.23	71.23	72.85	75.81	63.53	76.13



图 6 在 DOTA 数据集上的检测结果

Fig. 6 Detection results on the DOTA dataset

2.4 消融实验

为了验证改进的 YOLOv8 模型中各模块的有效性,设计了消融实验。在基准模型的基础上,逐步加入改进算法模块,并依据模型评价标准对比不同算法模块的优化效果。实验结果见表 5。

表 5 消融实验对比结果

Table 5 Comparison results of ablation experiments

方法	HRSC 2016	DOTA	UCAS-AOD
Baseline	95.56	74.01	95.14
Baseline+KFIoU	95.73	74.57	96.11
Baseline+LMKNet	95.81	75.66	97.46
Baseline+SAS	96.11	75.91	96.77
Baseline+LMK+SAS	96.34	76.13	97.80
Ours	96.73	77.26	97.85

首先,验证了 LMKNet 模块的有效性。该模块在提取遥感图像的固有特征时,需要更广泛且适应性强的上下文理解能力。通过调整其较大的空间感受野,LMKNet 可以有效捕捉不同目标类型的上下文细微差别。因此,在包含微小目标的遥感场景

检测任务中,LMKNet 表现出显著的优势。实验结果显示,在 DOTA 和 UCAS-AOD 数据集上,模型性能得到了显著提升。

其次,在 HRSC 2016 数据集上对所提出的形状自适应标签分配策略(SAS)进行了实验。该数据集包含大量高长宽比的目标,实验结果证明了 SAS 策略的有效性。

最后,引入了 KFIoU 损失函数,以解决损失函数不可微的问题。实验结果表明,目标检测器的性能得到了稳定提升。

此外,将改进的 YOLOv8 算法与原 YOLOv8 算法的部分图像检测效果进行了直观对比,结果见图 7。由图 7 可知,原 YOLOv8 算法对小目标、大长宽比目标和旋转目标的误检率较高。其原因在于原 YOLOv8 算法对目标特征的提取能力不足,无法准确提取目标特征。同时,改进前的标签分配策略未能适配目标的形状特征,针对大长宽比目标容易出现误检和漏检问题。此外,由于改进前的损失函数不可微,在检测旋转目标时检测框回归不准确。相比之下,改进的 YOLOv8 算法显著降低了漏检和误检率,并且对旋转目标的检测能力有了一定提升。

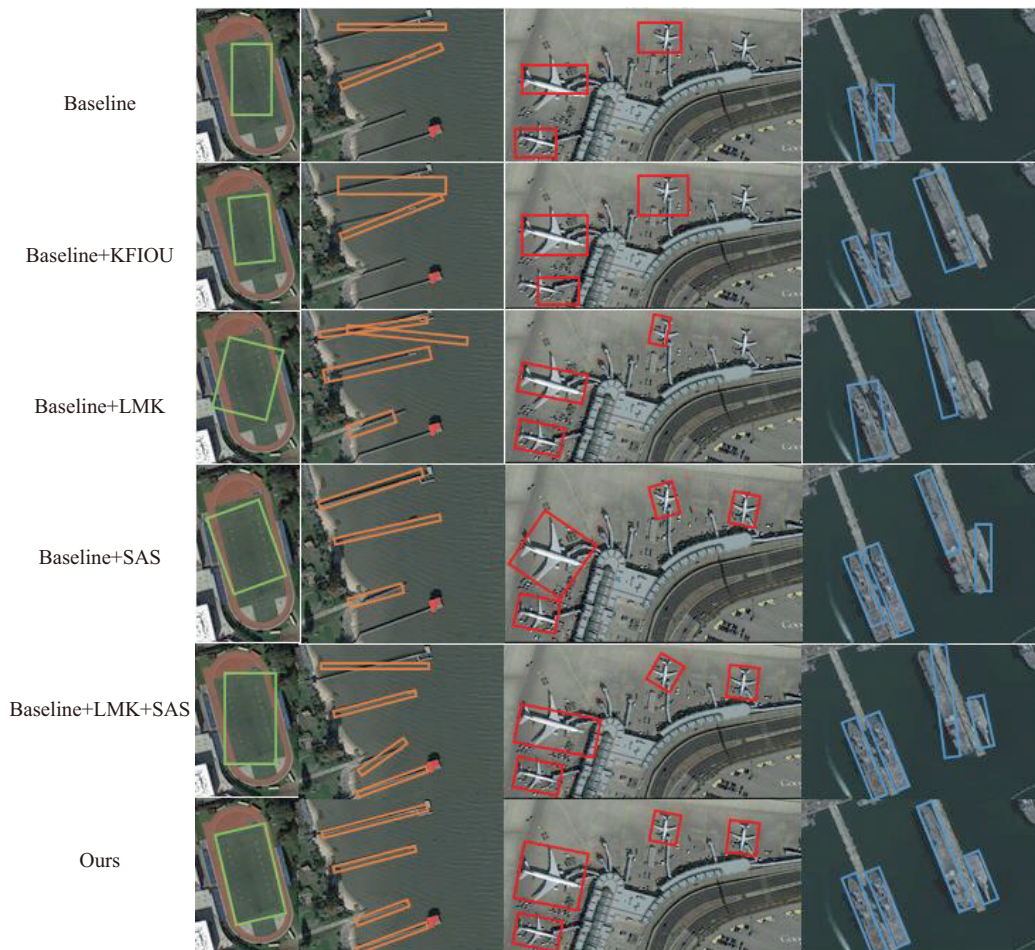


图 7 消融实验可视化结果图

Fig. 7 Visualization results of ablation experiments

3 总结

为提高遥感图像目标检测算法的性能,提出了改进的YOLOv8算法。为了使模型能够针对不同目标调整感受野,从而更加准确地提取遥感图像中的目标特征,设计了大核选择性网络(LMKNet)。为了使模型对形状各异、高宽比大的物体不误检,设计了形状自适应标签分配策略(SAS),有效适应了物体的形状,提高了检测精度。此外,为了提高模型对旋转目标的检测能力,引入了KFIoU损失函数,以适应遥感图像中目标物体的形变和旋转。

在HRSC 2016, DOTA和RSOD数据集上,相对于原YOLOv8算法,改进的YOLOv8算法在mAP50指标下分别提升0.78%、2.12%和2.66%。未来的工作将主要聚焦于进一步优化网络模型,以提高遥感图像目标检测的性能。

参考文献:

- [1] CHENG Gong, HAN Junwei. A survey on object detection in optical remote sensing images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 117: 11-28.
- [2] 张园林. 基于高分遥感影像的深度学习目标检测方法研究[D]. 西安:中国科学院大学(中国科学院西安光学精密机械研究所),2020.
- [3] REN Shaqing, HE Kaiming, 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.
- [4] LIU Zhuang, MAO Hanzhi, WU Chaoyuan, et al. A ConvNet for the 2020s [C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA:IEEE, 2022: 11966-11976.
- [5] DING Xiaohan, ZHANG Xiangyu, HAN Jungong, et al. Scaling up your kernels to 31×31 : Revisiting large kernel design in CNNs [C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA:IEEE, 2022: 11953-11965.
- [6] LIU Shiwei, CHEN Tianlong, CHEN Xiaohan, et al. More ConvNets in the 2020s: Scaling up kernels beyond 51×51 using sparsity[EB/OL]. (2023-03-03) [2024-11-03]. <https://doi.org/10.48550/arXiv.2207.03620>.
- [7] 姚宗亮,黄荣,董爱华,等. 基于多模态融合和自适应剪枝Transformer的脑肿瘤图像分割算法[J]. *宁夏大学学报(自然科学版)*,2024,45(1):16-24.
- [8] GUO Menghao, LU Chengze, HOU Qibin, et al. Seg-next: Rethinking convolutional attention design for semantic segmentation[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 1140-1156.
- [9] HOU Qibin, LU Chengze, CHENG Mingming, et al. Conv2Former: A simple transformer-style ConvNet for visual recognition [EB/OL]. (2022-11-22) [2024-11-03]. <https://doi.org/10.48550/arXiv.2211.11943>.
- [10] YANG Tong, ZHANG Xianyu, LI Zeming, et al. MetaAnchor: Learning to detect objects with customized anchors [EB/OL]. (2018-07-03) [2024-11-03]. <https://doi.org/10.48550/arXiv.1807.00980>.
- [11] MING Qi, ZHOU Zhiqiang, MIAO Lingjuan, et al. Dynamic anchor learning for arbitrary-oriented object detection[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(3): 2355-2363.
- [12] TIAN Zhi, SHEN Chunhua, CHEN Hao, et al. FCOS: Fully convolutional one-stage object detection [C]// *IEEE/CVF International Conference on Computer Vision (ICCV)*. Republic of Korea: IEEE, 2020:9626-9635. DOI:10.1109/ICCV.2019.00962.
- [13] KIM K, LEE H S. Probabilistic anchor assignment with IoU prediction for object detection [M]// VEDALDI A, BISCHOF H, BROX T, et al. *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2020: 355-371.
- [14] LI Yuxuan, HOU Qibin, ZHENG Zhaohui, et al. Large selective kernel network for remote sensing object detection [C]//*IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France: IEEE, 2023: 16748-16759.
- [15] YANG Xue, ZHOU Yue, ZHANG Gefan, et al. The KFIoU loss for rotated object detection [EB/OL]. (2023-02-06) [2024-11-03]. <https://doi.org/10.48550/arXiv.2201.12558>.
- [16] ZHANG Shifeng, CHI Cheng, YAO Yongqiang, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA:IEEE, 2020: 9759-9768.
- [17] YANG Xue, YAN Junchi, FENG Ziming, et al. R3Det: Refined single-stage detector with feature refinement for rotating object[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(4): 3163-3171.
- [18] QIAN Wen, YANG Xue, PENG Silong, et al. RSDet: Point-based modulated loss for more accurate rotated object detection [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(11): 7869-7879.

- [19] ZHAO Pengbo, QU Zhenshen, BU Yingjia, et al. PolarDet: A fast, more precise detector for rotated target in aerial images [J]. *International Journal of Remote Sensing*, 2021, 42(15): 5831-5861.
- [20] YANG Xue, HOU Liping, ZHOU Yue, et al. Dense label encoding for boundary discontinuity free rotation detection [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA:IEEE, 2021:15819-15829.
- [21] HAN Jiaming, DING Jian, LI Jie, et al. Align deep features for oriented object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 5602511. DOI: 10.1109/TGRS.2021.3111234.
- [22] CHENG Gong, WANG Jiabo, LI Ke, et al. Anchor-free oriented proposal generator for object detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5625411. DOI: 10.1109/TGRS.2022.3183022.
- [23] PAN Xingjia, REN Yuqiang, SHENG Kekai, et al. Dynamic refinement network for oriented and densely packed object detection [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA:IEEE, 2020: 11207-11216.
- [24] HOU Liping, LU Ke, XUE Jian, et al. Shape-adaptive selection and measurement for oriented object detection[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(1): 923-932.

Target Detection in Remote Sensing Images Based on Large Kernel Selection and Shape Adaptation

ZHAO Ziao¹, DONG Aihua^{1,2*}, HUANG Rong^{1,2}

(1. College of Information Science and Technology, Donghua University, Shanghai 201620, China;

2. Engineering Research Center of Digitized Textile and Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, China)

Abstract: Target detection in optical remote sensing images is a key technology for the intelligent interpretation of remote sensing data. To address the challenges posed by of significant scale variations, background interference and the diversity of target shapes in remote sensing image detection, this paper proposes the Large Multi-scale Kernel (LMK) network. This network utilized large kernel convolution and a multi-scale attention mechanism to dynamically adjust the spatial receptive field, thereby enhancing the capture of contextual information about objects in remote sensing scenes. Furthermore, a Shape-Adaptive Selection (SAS) label allocation strategy was designed for target detection, which focused on the aspect ratio of the target shapes and calculated an optimal IoU threshold based on the shape information and feature distribution. To address the difficulty of target orientation and positioning in remote sensing images, this paper introduced the KFIoU loss function. Experimental results show that the proposed target detection model achieves accuracies of 96.73%, 97.85%, and 77.26% on the HRSC 2016, UCAS-AOD, and DOTA datasets, respectively, outperforming most existing target detection algorithms.

Key words: target detection; deep learning; label allocation; multi-scale attention; large kernel network

(责任编辑 张 娣)