

DOI:10.19651/j.cnki.emt.2416615

# 改进 Oriented R-CNN 的遥感图像定向 目标检测算法\*

王雷雨 王正勇 陈洪刚 何小海  
(四川大学电子信息学院 成都 610065)

**摘要:** 近年来遥感目标检测的研究主要集中在改进边界框的表示方法,而忽略了遥感场景中独特的先验知识。为了进一步提高双阶段模型的检测精度,同时保持推理复杂度,本文以大核卷积构建的特征提取器 LSKNet 为基线,在特征表示和训练策略上进一步做出了改进。首先,通过 RFA 提取比例不变的上下文信息,以缓解 LSKNet 引入的背景噪声、提高模型对噪声的鲁棒性;然后,通过构建 CS 进一步提出 CS 策略来缩小不同尺度特征之间的语义鸿沟、使模型具备多尺度能力的同时更专注于小目标。本文的方法在几乎没有增加推理复杂度的同时,在大型遥感图像数据集 DOTA 上的单尺度结果达到了 79.03% mAP<sub>50</sub>,证明了提出方案的有效性。

**关键词:** 目标检测;遥感图像;小目标;残差特征增强;一致性监督

**中图分类号:** TP391; TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.2040; 430.40

## Augment Oriented R-CNN for remote sensing object detection

Wang Leiyu Wang Zhengyong Chen Honggang He Xiaohai

(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

**Abstract:** In recent years, research on remote sensing object detection has mainly focused on improving the representation methods for bounding boxes, while overlooking the unique prior knowledge present in remote sensing scenes. To further enhance the detection accuracy of two-stage models while maintaining inference complexity, this paper presents improvements in feature representation and training strategies based on the feature extractor LSKNet constructed with large kernel convolutions. First, the RFA module is introduced to extract scale-invariant contextual information, alleviating the background noise introduced by LSK and enhancing the model's robustness to noise. Then, the CS loss is proposed to implement a consistent supervision training strategy that reduces the semantic gap between features of different scales, enabling the model to possess multi-scale capabilities while focusing more on small objects. The proposed method achieves a single-scale result of 79.03% mAP<sub>50</sub> on the large remote sensing image dataset DOTA, demonstrating the effectiveness of the proposed approach with almost no increase in inference complexity.

**Keywords:** object detection; remote sensing images; small targets; residual feature augmentation; consistent supervision

## 0 引言

遥感图像中的目标检测近年来获得了极大的关注。该任务致力于识别遥感图像(remote sensing image, RSI)中特定对象的存在,并随后确定其类别和精确位置<sup>[1-3]</sup>。与一般产生水平边界框的普通目标检测不同,RSI 目标检测旨在产生与对象的方向精确对齐的边界框。近年来,许多的努力致力于提出各种定向边界框(oriented bounding box,

OBB)检测器<sup>[4-25]</sup>并提高 OBBs 的角度预测精度。在分类上,和普通视景下的目标检测类似,遥感定向目标检测器也可大致分为基于 DETR (detection transformer)的检测器<sup>[4-6]</sup>、单阶段检测器<sup>[7-13]</sup>和双阶段检测器<sup>[14-25]</sup>三大类。

基于 DETR 的定向检测器<sup>[4-6]</sup>真正实现了端到端的训练,不需要传统方法中的复杂预处理和后处理步骤,如锚点生成、特征金字塔、非极大值抑制等,大大简化了检测的流程,缺点是具有较高的计算复杂度。Dai 等<sup>[4]</sup>通过引入定

收稿日期:2024-08-05

\* 基金项目:国家自然科学基金(62271336, 62211530110)、四川省科技计划资助(2024YFHZ0212)、中央高校基本科研业务费专项资金(SCU2023D062, 2022CDSN-15-SCU)项目资助

向提议生成机制、自适应定向提议细化模块和旋转感知集匹配损失提出了 AO<sup>2</sup>-DETR,大幅简化了检测流程并显著提升了检测器的性能。Ma 等<sup>[5]</sup>通过应用深度可分离卷积提出了 O<sup>2</sup>-DETR,显著降低了在原始 Transformer 中使用多尺度特征的内存和计算成本,在遥感数据集上取得了极高的表现。Zeng 等<sup>[6]</sup>提出了 ARS-DETR 网络,通过纵横比感知圆平滑标签(aspect ratio aware circle smooth label, AR-CSL)更合理的平滑角度标签,并丢弃先前的工作中引入的超参数,在高精度定向目标检测任务上取得了有竞争力的性能。然而,这些方法由于基于 DETR 的原因均具有较大的参数量和较高的计算复杂度。

单阶段定向检测器<sup>[7-13]</sup>具有较低的参数量和较快的推理速度,劣势则是相对于双阶段检测器来说精度较低。Hou 等<sup>[7]</sup>通过提出定向对象检测的柔性形状自适应选择(shape-adaptive selection, SA-S)和形状自适应测量(shape-adaptive measurement, SA-M)策略进而提出 SASM 网络,在遥感数据集上取得了优异的结果。Yang 等<sup>[8]</sup>通过提出一种近似的 SkewIoU 损失(Skew intersection over union, SkewIoU),进一步提出了一种端到端精细化的单级旋转检测器 S3Det,通过使用从粗粒度到细粒度的逐步回归方法,实现快速准确的目标检测。Yang 等<sup>[9]</sup>通过一种新的基于高斯 Wasserstein 距离(Gaussian wasserstein distance, GWD)的回归损失提出了 R3Det-GWD 网络,缓解了边界不连续性及其与最终检测度量的不一致性问题。还有类似的研究方向,Yang 等<sup>[10]</sup>将旋转的边界框转换为二维高斯分布,然后计算高斯分布之间的 Kullback-Leibler 散度(Kullback-Leibler divergence, KLD)作为回归损失,进而提出了 R3Det-KLD 网络,提高了检测器在具有大纵横比的目标上的精度。Li 等<sup>[11]</sup>提出了 O-RepPoints 网络,用自适应点表示捕获任意方向实例的几何信息,同时引入空间约束来惩罚自适应学习中的异常点,其在遥感定向数据集上也取得了优异的结果。Tian 等<sup>[12]</sup>提出了 FCOS 网络,通过直接预测每个位置的类别和边界框并引入中心度分支提高了检测精度。Han 等<sup>[13]</sup>提出了 S<sup>2</sup>Anet 网络,在该网络中使用了特征对准模块(feature alignment module, FAM)和定向检测模块(oriented detection module, ODM),根据锚点框自适应地对齐卷积特征,同时采用主动旋转滤波器对方位信息进行编码,然后产生方位敏感和方位不变的特征,缓解了分类得分和定位精度之间的一致性问题。

双阶段定向检测器<sup>[14-22]</sup>尽管具有推理速度较慢的缺陷,然而其在 RSI 上的精度表现更加优异。Yang 等<sup>[14]</sup>设计了一种采样融合网络 SCRDet,将多层特征与有效的锚采样融合在一起,提高了对小物体的灵敏度。同时,将 IoU(intersection over union, IoU)常数因子添加到平滑 L1 损失中,解决了旋转边界框的边界问题。Xu 等<sup>[15]</sup>提出的 G. V. 检测器通过滑动水平边界框的每个对应边上的顶点准确地描述多方向对象,其通过回归四个长度比避免了定

向对象的顺序标签点的混淆问题,在定向检测基准上去除了优越的性能。Long 等<sup>[16]</sup>提出的 CenterMap 检测器通过结合在线地图引擎、开源地理数据和公共地理数据库实现了高效、准确的遥感目标检测效果。Han 等<sup>[17]</sup>提出的 ReDet 网络显式地编码旋转等变和旋转不变性,不仅仅准确率上优于基线方法,更是相对于基线将参数数量减少了 60%(313 Mb 对 121 Mb)。同样的,通过优化检测头的编码方法提升 RSI 检测效果的双阶段方法还有很多,比如 Ding 等<sup>[18]</sup>在面向 OBB 边界框注释的监督下对 RoI 应用空间变换并学习变换参数,进而提出了 Roi Trans 网络。Ren 等<sup>[19]</sup>也是通过修改边界框的编码方式将 Faster R-CNN 适配为了定向双阶段检测器 Rotated Faster R-CNN。还有效果十分惊艳的 Oriented R-CNN, Xie 等<sup>[20]</sup>通过提出了一个面向区域的提案网络,以几乎无成本的方式直接生成高质量的面向区域的方案;随后使用定向 R-CNN 头进一步细化定向兴趣区域并识别它们,达到了至今也名列前茅的精度。Feng 等<sup>[21]</sup>结合三角函数损失,提出了 CADLA 网络,通过一种基于类别反馈的动态标签分配法提高了负样本选择代表性,在多个数据集上也达到了优异的精度。Xie 等<sup>[22]</sup>创新的提出了一种物质激活网络 OAN,通过判断物体的存在性,显著减少了检测器的计算负担。

由于双阶段模型 Oriented R-CNN 结构在精度方面具有极高的精度表现,近年来很多模型基于 Oriented R-CNN 做出改进并不断刷新领域内权威数据集的最优精度<sup>[23-25]</sup>。Li 等<sup>[23]</sup>提出的 LSKNet 通过将大核卷积和空洞卷积合并到骨干网络中来选择性地扩大较大物体的空间感受野,以捕获更多的场景上下文信息。Pu 等<sup>[24]</sup>提出的 ARC 通过一种自适应旋转卷积根据图像中不同方向的目标自适应旋转以提取特征,并引入了高效的条件计算机制来应对目标在单张图片中的大范围方向变化。Cai 等<sup>[25]</sup>提出的 PKINet,通过无膨胀的多尺度卷积核和上下文锚注意力缓解了特征表示的稀疏问题,达到了较高的检测精度。然而,这些方法依旧存在一定的局限性,一方面使用大核卷积可能会引入大量的背景噪声,这可能对某些目标的准确检测有害;另一方面这些改进仅仅在特征提取骨干层面上,在遥感场景下的特征金字塔生成的多尺度特征之间依然存在不可忽视的语义鸿沟。

为了应对上述挑战,同时使模型具备高精度、低参数量的优势,本文选取双阶段网络 LSKNet 作为基线,首先利用该强大而轻量的特征提取骨干 LSK-S 为对象提供足够的上下文信息;其次通过构建残差特征增强(residual feature augmentation, RFA)模块,缓解了大核卷积引入背景噪声而导致弱小目标检测效果下降的问题;最后,本文引入了一致性监督(consistent supervision, CS)策略,进一步通过 CS 损失约束多尺度特征金字塔训练过程,缩小特征融合前不同尺度特征之间的语义差距。进一步,本文提出了改进 Oriented R-CNN 的遥感图像定向目标检测算法。其综合

了 LSKNet 的核选择、RFA 模块对噪声的高鲁棒性以及 CS 策略在多尺度特征上的指导,保证轻量的同时也保持了双阶段定向检测器的高精度优势。其在 DOTA<sup>[26]</sup> 数据集上的单尺度精度超越了多数先进旋转检测算法,达到单尺度 79.03% mAP<sub>50</sub>。

### 1 改进型 Oriented R-CNN

Oriented R-CNN 是一种通用双阶段定向目标检测器。其网络结构如图 1 所示。其第 1 阶段通过定向区域提议网络生成带有方向的区域提议,第 2 阶段通过定向 R-CNN 头

部对区域提议进行分类并进一步细化其空间位置。Oriented R-CNN 已经在 RSI 上取得了优越的表现,然而,其研究主要集中在改进边界框的表示方法上,而忽略了 RSI 中独特的先验知识,并不能捕捉充分的场景上下文信息;尽管 LSKNet 通过可选择性大核卷积优化了该缺陷,然而依旧具有局限性。一方面使用大核卷积可能会引入大量的背景噪声,这可能对某些小目标的准确检测有影响;另一方面这些改进仅仅在特征提取骨干层面上,在遥感场景下的特征金字塔生成的多尺度特征之间依然存在不可忽视的语义鸿沟。

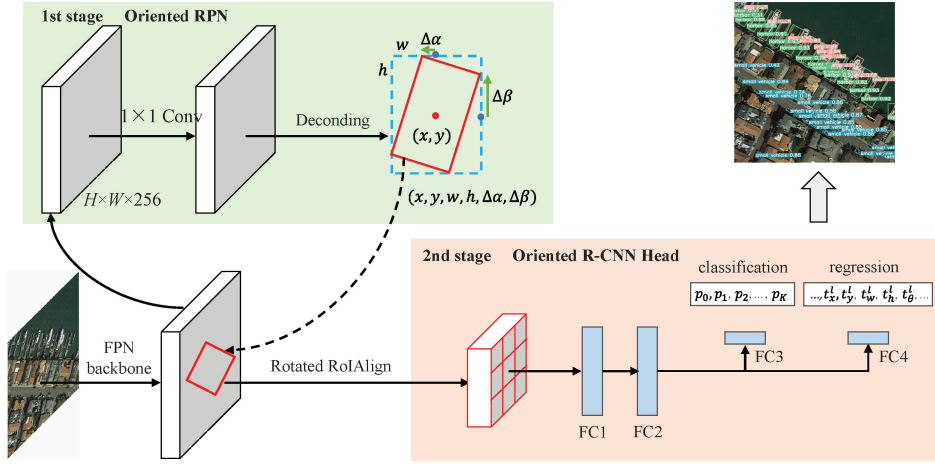


图 1 Oriented R-CNN 的网络结构示意图  
Fig. 1 Illustration of the Oriented R-CNN

为了缓解上述问题,本文以 LSKNet 为基线方法,提出了改进 Oriented R-CNN 的遥感图像定向目标检测算法 (augmented oriented R-CNN, AO-RCNN)。如图 2 所示,首先基于强大而轻量的特征提取骨干网络 LSK-S(图 2 左侧),依靠大核或空洞卷积来扩展感受野,为对象提供足够的上下文

信息;其次通过构建 RFA 模块(图 2 顶部)来缓解大核卷积引入背景噪声而导致部分弱小目标检测效果下降的问题;最后,使用了 CS 策略(图 2 中间部分),进一步通过 CS 损失指导多尺度特征金字塔的训练过程,缩小特征融合前不同尺度特征之间的语义差距,使网络进一步适应 RSI 的弱小目标特性。

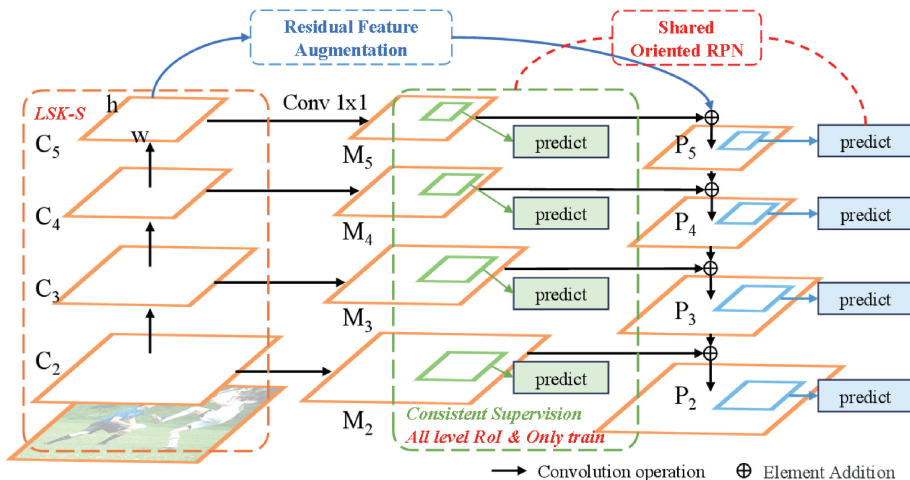


图 2 改进 Oriented R-CNN 的遥感图像定向目标检测算法示意图  
Fig. 2 AO-RCNN for remote sensing object detection algorithm

### 1.1 可选择性大核卷积网络 LSK-S

可选择性大核卷积网络 LSKNet 是考虑到不同类型的遥感目标所需的背景信息量不同而提出的一种内部核大小可选择大核轻量级特征提取骨干网络,如图 3 所示。本文使用了 LSKNet 的 S 版本即 LSK-S 作为特征提取器,其网络结构如图 3(a)所示。LSK-S 分为 4 个阶段来进行

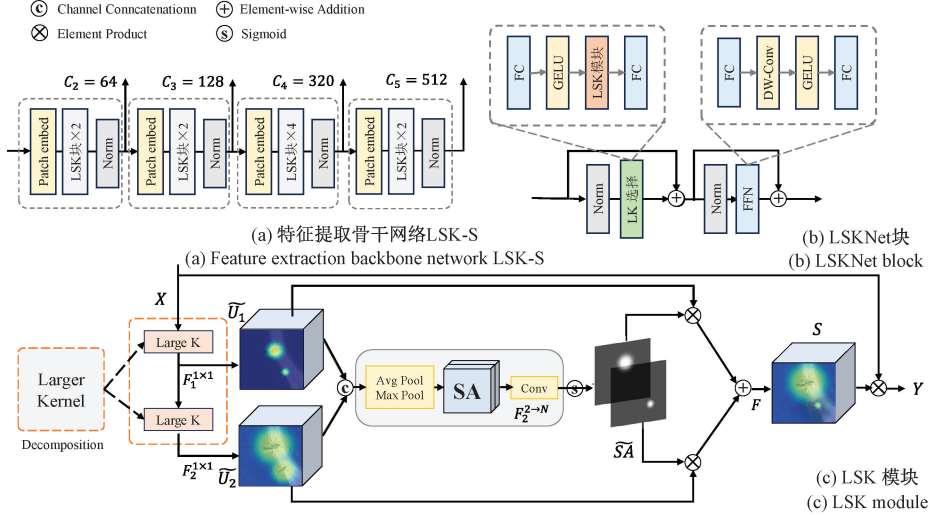


图 3 可选择性大核卷积网络 LSK-S 示意图

Fig. 3 S version of selective large kernel convolution network

LSKNet 块的示意图如图 3(b)所示。其为骨干网络中的一个重复块。每一个 LSKNet 块由两个残差子块组成:大内核选择子块(主要包含 LSK 模块)和前馈网络(FFN)子块。大内核选择子块根据需求动态调整网络的感受野。FFN 子块主要用于通道混合和特征细化,其由 1 个全连接层、1 个深度可分离卷积、一个 GELU<sup>[27]</sup> 激活层和第 2 个全连接层的序列组成。

如图 3(c)所示的 LSK 模块是整个特征提取骨干的核心。LSK 模块将一个更大的卷积显示的分解为具有增长核(核的大小逐渐变大)的深度膨胀卷积序列。

具体的,第  $i$  个深度卷积感受野  $RF_i$  的定义如下:

$$k_{i-1} \leq k_i; d_1 = 1, d_{i-1} < d_i \leq RF_{i-1} \quad (1)$$

$$RF_1 = k_1, RF_i = d_i(k_i - 1) + RF_{i-1} \quad (2)$$

其中,  $k$  为核大小,  $d$  为膨胀率。核大小和扩张率的增加保证了感受野的扩张速度足够快,同时通过设置上限来保证膨胀卷积不会在特征映射中引入间隙。

例如表 1,可以将一个大的内核分解为 2 或 3 个深度卷积。如此将原始的大核卷积分解为具有增长核的卷积序列可以显式的提取具有各大接受域的多个特征,便于后续内核选择;其次,卷积序列比简单的应用一个更大的核更有效(更低的参数量,更低的计算复杂度)。

LSK 模块首先通过一系列具有不同感受野的深度卷积分解输入特征,该过程可以描述为:

$$U_0 = X, U_{i+1} = F_i^{dw}(U_i) \quad (3)$$

提取 4 个尺度的多尺度特征  $\{C_2, C_3, C_4, C_5\}$ 。每个阶段首先分别经过一个 Patch embed 模块(包含一个大核卷积和一个归一化层)进行特征嵌入,将特征通道分别调整为  $\{64, 128, 320, 512\}$ ; 随后各个阶段分别经过  $\{2, 2, 4, 2\}$  个 LSKNet 块及归一化层进行多尺度特征  $\{C_2, C_3, C_4, C_5\}$ 。

表 1 大核卷积的显式分解示例

Table 1 Explicit decomposition example of large kernel convolution

示例	原始卷积	解耦卷积
1	$RF = 23, (23, 1)$	$(5, 1) \rightarrow (7, 3)$
2	$RF = 29, (29, 1)$	$(3, 1) \rightarrow (5, 2) \rightarrow (7, 3)$

其中,  $F_i^{dw}(\cdot)$  表示具有卷积核大小为  $k_i$  和膨胀率为  $d_i$  的深度膨胀卷积。随后通过  $1 \times 1$  卷积进行空间特征向量的通道交互,其可以描述为:

$$\tilde{U}_i = F_i^{1 \times 1}(U_i) \quad (4)$$

基于分解的具有不同感受野的特征,随后执行的是多尺度特征对不同目标动态选择核的空间选择机制。首先,将不同核获得的不同感受野的特征拼接:

$$\tilde{U} = [\tilde{U}_1; \dots; \tilde{U}_i] \quad (5)$$

然后,对  $\tilde{U}$  在通道层执行均值池化  $P_{avg}(\cdot)$  和最大值池化  $P_{max}(\cdot)$ :

$$SA_{avg} = P_{avg}(\tilde{U}), SA_{max} = P_{max}(\tilde{U}) \quad (6)$$

其中,  $SA_{avg}$  和  $SA_{max}$  分别为均值池化和最大值池化的空间描述特征。为了使不同的空间描述符进行信息交互,LSKNet 拼接了空间描述特征,并使用卷积层  $F_2^{2 \rightarrow N}$  将拼接后的空间描述特征(2 通道)通过下式转换为  $N$  个空间注意力图:

$$\widehat{SA} = F^{2 \rightarrow N}([\widehat{SA}_{avg}; \widehat{SA}_{max}]) \quad (7)$$

对于每个空间注意力图  $\widehat{SA}_i$ ，使用一个 sigmoid 激活函数来分别获得每个分解的大核单独的空间选择掩码：

$$\widetilde{SA} = \sigma(\widehat{SA}_i) \quad (8)$$

其中， $\sigma(\cdot)$  表示激活函数。然后对分解后的大核序列中的特征进行相应的空间选择蒙版加权，并且通过卷积层  $F(\cdot)$  进行融合，得到注意力特征：

$$S = F\left(\sum_{i=1}^N (\widetilde{SA}_i \cdot \widetilde{U}_i)\right) \quad (9)$$

LSK 模块最后输出特征是  $X$  和  $S$  之间的元素积：

$$Y = X \cdot S \quad (10)$$

本文在图 3(c) 中直观的描述了上述的自适应选择卷积核的过程。

### 1.2 残差特征增强模块

LSK-S 使模型充分利用了遥感目标的背景先验知识，使模型具备了较强的特征提取能力。然而，LSK-S 在利用 RSI 的背景先验知识时，并没有考虑到对背景噪声的引入的约束，这使得模型在利用背景先验时，易受背景噪声干扰。鉴于此，本文通过一个 RFA 模块提取比例不变的上下文信息，来缓解前者引入的背景噪声。在 Oriented RCNN 中使用的 FPN<sup>[28]</sup> 中，最高级别的特征图  $M_5$  以自上而下的路径传播，并逐渐与较低级别的特征图  $\{M_4, M_3, M_2\}$  融合。一方面，较低层次的特征映射和较高层次的语义信息

融合，获得了噪声减弱的、具有上下文信息的尺度特征。另一方面， $M_5$  就只包含与其他级别特征不兼容的单一尺度上下文信息。单一尺度特征信息很容易被噪声干扰，产生比较差的特征表示。这种现象在使用 LSK-S 这种大量提取背景先验的大核特征提取网络时尤其明显。本文希望利用更高级的空间上下文信息来净化  $M_5$  特征，增强其特征表示对噪声的鲁棒性。

为此，本文使用了如图 4 所示的 RFA。首先通过对尺度为  $S = h \times w$  的  $C_5$  执行比例不变的自适应池化 (ratio-invariant adaptive pooling, RAP) 生成多个不同尺度 ( $\alpha_1 \times S, \alpha_2 \times S, \alpha_3 \times S$ ) 的上下文信息。然后每个上下文特征分别经过  $1 \times 1$  的卷积层，将特征通道维数降至 256。最后，通过双线性差值将其上采样到  $S$  的尺度，用于后续的融合。考虑到插值产生的混叠效应，本文设计了自适应空间融合模块 (adaptive spatial fusion, ASF) 自适应地融合上述的上下文特征。具体地，ASF 使用上采样特征作为输入，并为每个特征生成一个空间权重。这些权重将上下文特征聚合成对噪声具有鲁棒性的  $M_6$ ，然后  $M_6$  通过求和与  $M_5$  结合传播，并与其他低级特征融合。最后，在每个特征上通过附加一个  $3 \times 3$  卷积构建特征金字塔  $\{P_2, P_3, P_4, P_5\}$ 。

本文提出的 RFA 模块，通过优化  $M_5$  的特征表示，在原本的尺度上融合了更高维的特征  $M_6$ ，最终增强了模型提取的特征对噪声的鲁棒性。

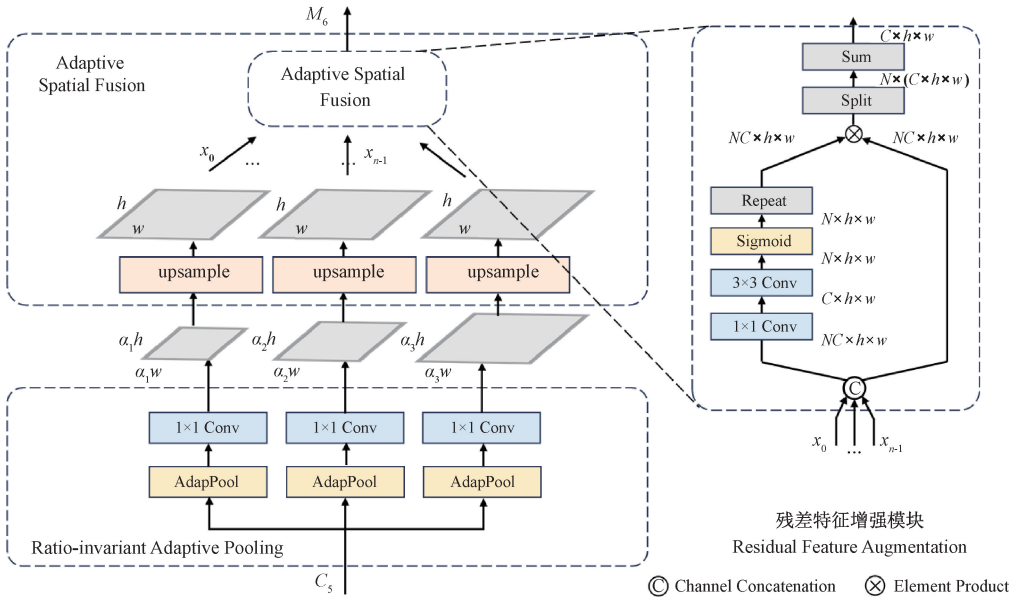


图 4 残差特征增强模块  
Fig. 4 Residual feature enhancement module

### 1.3 一致性监督损失

FPN 利用网络内的特征层次结构生成不同分辨率的特征映射，构建特征金字塔。然而，在遥感场景下，由于 RSI 覆盖广阔的地理范围，同一目标在不同位置可能因为传感器高度、视角或分辨率的不同而呈现出非常大的尺度

差异。尽管 FPN 采用自上而下的上采样和求和方式融合不同尺度的特征，但是它们之间依然存在较大的语义差距，直接将 FPN 应用于 RSI 目标检测来融合多个语义间隙较大的特征会导致次优特征金字塔。相当多的工作，仅仅从特征融合层面做出改进，这往往会增加模型的参数量

或降低模型的推理速度。这启发使用如图 5 所示的 CS 策略,即在融合前对多尺度特征施加相同的监督信号,以缩小它们之间的语义差距,使模型更加适用于尺度变化剧烈的遥感场景。CS 策略仅仅影响训练过程,完全没有对推理性能带来损耗。具体地,本文基于来自 LSK-S 的多尺度

特征  $\{C_2, C_3, C_4, C_5\}$  来进一步构建多尺度特征金字塔  $\{P_2, P_3, P_4, P_5\}$ 。多尺度特征  $\{P_2, P_3, P_4, P_5\}$  经过 Oriented RCNN 原来的定向 RPN 和定向 RCNN 头进行类别、位置预测,而多尺度特征  $\{C_2, C_3, C_4, C_5\}$  经过 CS 辅助头进行监督训练。

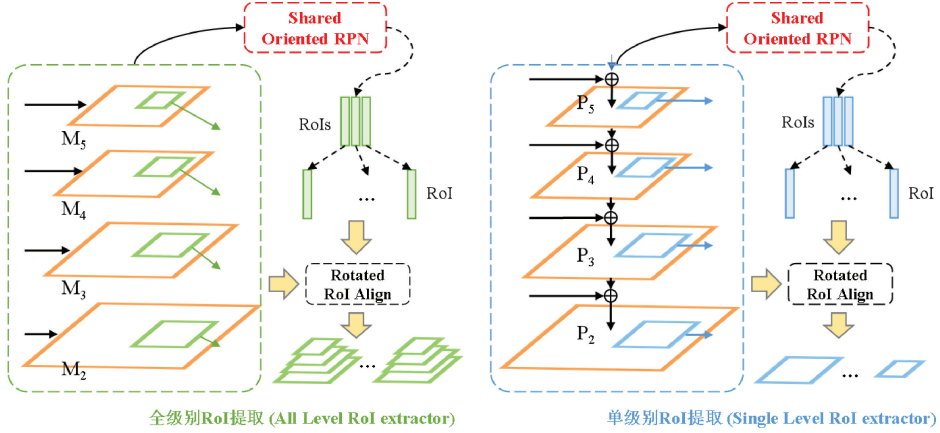


图 5 全级别特征提取和单级别特征提取示意图

Fig. 5 All level and single level feature extraction diagram

下面详细介绍辅助头的设计。首先,在多尺度特征  $\{M_2, M_3, M_4, M_5\}$  上应用和推理检测头共享的定向 RPN 头生成多个 RoI。除了和推理检测头共享定向 RPN 外,CS 辅助头还有一个显著性特点,如图 5 所示,每个 RoI 被映射到所有的特征层上进行全特征级别的 roi 提取(all level roi extractor, ALRE)而不是仅仅映射到对应级别的特征单特征级别的特征提取(single level roi extractor, SLRE),并且每个 RoI 通过 Oriented RCNN 中的 Rotated RoI-Align 从  $\{M_2, M_3, M_4, M_5\}$  上获得。最后,在这些特征上附加多个分类和回归头,产生 CS 损失。这些分类和回归头的参数跨不同级别共享,这进一步迫使不同的特征图学习除了相同的监督信号以外的相似语义信息。

为了更加稳定的训练,本文使用一个权值来平衡 CS 产生的辅助损失和原始损失。AO-RCNN 头部最终损失表示为:

$$L_{rcnn} = \lambda \sum_{M=2}^5 (L_{cls,M}(p_M, t^*) + \beta [t^* > 0]) L_{loc,M}(d_M, b^*) + \sum_{P=2}^5 (L_{cls,P}(p, t^*) + \beta [t^* > 0]) L_{loc,P}(d, b^*) \quad (11)$$

其中,  $L_{cls,M}(\cdot)$  和  $L_{loc,M}(\cdot)$  是附加在  $\{M_2, M_3, M_4, M_5\}$  上的辅助损失所对应的目标函数,  $L_{cls,P}(\cdot)$  和  $L_{loc,P}(\cdot)$  是特征金字塔  $\{P_2, P_3, P_4, P_5\}$  上的原始损失函数,  $p_M, d_M$  和  $p, d$  分别为中间层和最终金字塔层的预测值,  $t^*$  和  $b^*$  分别是真值类别标签和回归标签,  $\lambda$  是平衡 CS 损失和原检测损失的权重,  $\beta$  的用来平衡分类和定位损失的权重,  $[t^* > 0]$  定义为:

$$[t^* > 0] = \begin{cases} 1, & t^* > 0 \\ 0, & t^* = 0 \end{cases} \quad (12)$$

本文在 AO-RCNN 的测试推理阶段移除了 CS 分支,只使用特征金字塔后的分支进行最终预测。因此推理时不会给模型引入额外的参数和计算量。

## 2 实验及结果分析

在本节中,首先介绍采用的实验数据集及详细实验设置。然后,将所提出的 AO-RCNN 与近年来的 RSI 定向检测器进行公平比较以评估其性能表现。最后介绍如何通过消融实验,来证明提出方法的有效性。

### 2.1 实验设置

#### 1) 数据集

本文在遥感领域大型公开数据集 DOTA-v1.0 上评估所提出的方法。2018 年公开发布的 DOTA-v1.0 数据集是遥感领域权威数据集之一。其每张图像分辨率在  $800 \times 800$  到  $20\,000 \times 20\,000$  像素之间,包含不同尺度、方向和形状的目标。DOTA-v1.0 包含 2 806 张图像,188 282 个带有定向边界框注释的实例,包括以下 15 个类别:飞机(plane, PL),棒球场(baseball diamond, BD),桥梁(bridge, BR),田径场(ground track field, GTF),小型车(small vehicle, SV),大型车(large vehicle, LV),船舶(ship, SH),网球场(tennis court, TC),篮球场(basketball court, BC),储油罐(storage tank, ST),足球场(soccer-ball field, SBF),环形岛(roundabout, RA),海港(harbor, HA),游泳池(swimming pool, SP),直升机(helicopter, HC)。遵循现有训练方法,本文使用 200 的步幅处理数据集,并将训练集和验证集用于训练,测试集用于测试。将测试结果提交到 DOTA 评估服务器获得检测精度。

本文遵循主流评价指标  $mAP_{50}$  (预测框与真实框的

IoU 重叠阈值  $\geq 0.5$ ) 和模型参数量 Param 来验证 AO-RCNN 算法的优越性。

## 2) 训练细节

实验环境基于 Linux 系统 Ubuntu20.04、深度学习框架版本为 Pytorch1.12, GPU 加速工具版本 CUDA11.3.1。训练、测试和推理均使用单张 GeForce RTX3090。

实验使用 MMRotate<sup>[29]</sup> 算法框架。所有使用 Backbone(ResNet50<sup>[30]</sup>、LSK-S) 均在 ImageNet<sup>[31]</sup> 上经过 300 轮的预训练处理。随后正式的训练的过程中使用了水平翻转、垂直翻转和随机旋转 3 种数据增强方法; 使用 AdamW 优化器优化整个网络, 初始学习率为 0.0001, 动量为 0.9, 权衰减为 0.05; 使用批量 2 的输入 ( $1024 \times 1024$ ) 训练模型 12 轮。所有实验数据、特征可视化、检测可视化等均得自第 12 轮的权重文件。

## 2.2 对比实验

在本节中, 本文在 DOTA-v1.0 数据集上比较了所提出的方法与最近提出的遥感图像定向目标检测器的性能, 如表 2 所示。表 2 比较了每个类别的 AP<sub>50</sub> 结果和所有类

别的 mAP<sub>50</sub> 结果, 并且最优和次优结果分别使用双下划线和下划线的加粗字体标记。在表 2 中所列的先进定向目标检测算法中, 提出的 AO-RCNN 旋转目标检测算法获得了 79.03% 的最佳单尺度 mAP<sub>50</sub> 结果。在 mAP<sub>50</sub> 指标上分别超越了 LSKNet-S、PKINet-S 0.89%、0.64% 的 mAP<sub>50</sub>。值得注意的是, AO-RCNN 相对于最低参数量的检测器 PKINet-S, 仅仅增加了 0.7 M 的参数量, 大大提高了检测效果。在 DOTA-v1.0 测试集的 15 个类别中, 本文方法在类别 GTF、LV、TC、BC、SP、HC 上达到了最优; 在类别 PL、SH、HA 上达到了次优; 在类别 BD、BR、SV、ST、SBF、RA 上也保持着较优的 AP<sub>50</sub> 结果。显而易见, 本文方法在大多数类别上达到最优或次优, 这体现了所提网络结构在遥感图像弱小目标定向检测方面的有效性。表 3 展示了几年来基于 Oriented R-CNN 框架做出改进的 SOTA 方法在 DOTA-v1.0 数据集上的对比结果。CADLA 是基于 Faster RCNN 做出的旋转框编码上的改进, 本质上并非基于 Oriented R-CNN, 因此未在表 3 中陈列。

表 2 在 DOTA-v1.0 数据集上单尺度训练和测试设置下的对比实验结果

Table 2 Comparison results of single-scale training and testing on DOTA-v1.0 dataset

方法	All categories in DOTA-v1.0															mAP <sub>50</sub>	
	#P↓	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP		HC
AO <sup>2</sup> -DETR	40.8 M	87.99	79.46	45.74	66.64	78.90	73.90	73.30	90.40	80.55	85.89	55.19	<b>83.62</b>	51.83	70.15	60.04	70.91
O <sup>2</sup> -DETR	—	86.01	75.92	46.02	66.65	79.70	79.93	<b>89.17</b>	90.44	81.19	76.00	56.91	62.45	64.22	65.80	58.96	72.15
ARS-DETR	41.6 M	86.61	77.26	48.84	66.76	78.38	78.96	87.40	90.61	82.76	82.19	54.02	62.61	72.64	72.80	64.96	73.79
R3Det	41.9 M	89.00	75.60	46.64	67.09	76.18	73.40	79.02	90.88	78.62	84.88	59.00	61.16	63.65	62.39	37.94	69.70
Rotated FCOS	31.9 M	88.52	77.54	47.06	63.78	<b>80.42</b>	80.50	87.34	90.39	77.83	84.13	55.45	65.84	66.02	72.77	49.17	72.45
S <sup>2</sup> Anet	38.5 M	89.11	82.84	48.37	71.11	78.11	78.39	87.25	90.83	84.90	85.64	60.36	62.60	65.26	69.13	57.94	74.12
SASM	36.6 M	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
O-RepPoints	36.6 M	87.02	83.17	54.13	71.16	80.18	78.40	87.28	90.90	85.97	86.25	59.90	70.49	73.53	72.27	58.97	75.97
R3Det-GWD	41.9 M	88.82	82.94	55.63	72.75	78.52	83.10	87.46	90.21	86.36	85.44	64.70	61.41	73.46	76.94	57.38	76.34
R3Det-KLD	41.9 M	88.90	84.17	55.80	69.35	78.72	84.08	87.00	89.75	84.32	85.73	64.74	61.80	76.62	<b>78.49</b>	<b>70.89</b>	77.36
CenterMap	41.1 M	89.02	80.56	49.41	61.98	77.99	74.19	83.74	89.44	78.01	83.52	47.64	65.93	63.68	67.07	61.59	71.59
SCRDet	41.9 M	<b>89.98</b>	80.65	52.09	68.36	68.36	60.32	72.41	90.85	<b>87.94</b>	<b>86.86</b>	65.02	66.68	66.25	68.24	65.21	72.61
Rotated Faster R-CNN	41.1 M	89.40	81.81	47.28	67.44	73.96	73.12	85.03	90.90	85.15	84.90	56.60	64.77	64.70	70.28	62.22	73.17
Roi Trans	55.1 M	89.01	77.48	51.64	72.07	74.43	77.55	87.76	90.81	79.71	85.27	58.36	64.11	76.50	71.99	54.06	74.05
G. V.	41.1 M	89.64	<b>85.00</b>	52.26	77.34	73.01	73.14	86.82	90.74	79.02	<b>86.81</b>	59.55	<b>70.91</b>	72.94	70.86	57.32	75.02
ReDet	31.6 M	88.79	82.64	53.97	74.00	78.13	84.06	88.04	90.89	87.78	85.75	61.76	60.39	75.96	68.07	63.59	76.25
O-RCNN	41.1 M	89.46	82.12	<b>54.78</b>	70.86	78.93	83.00	88.20	90.90	87.50	84.68	63.97	67.69	74.94	68.84	52.28	75.87
OAN	—	89.70	84.03	54.61	73.46	79.30	83.27	88.12	90.90	85.22	84.68	62.06	66.87	75.28	70.71	62.75	76.73
CADLA	—	89.54	83.14	55.32	71.56	80.09	83.58	88.20	90.90	87.93	85.77	65.69	66.30	74.80	71.29	63.72	77.19
ARC	74.4 M	89.40	82.48	55.33	73.88	79.37	84.05	88.06	90.90	86.44	84.83	63.63	70.32	74.29	71.91	65.43	77.35
LSKNet-S	<b>31.0</b> M	89.79	84.06	53.36	<b>78.79</b>	79.11	<b>84.93</b>	88.22	<b>90.90</b>	87.41	86.75	<b>66.68</b>	63.68	77.02	71.05	70.31	78.14
PKINet-S	<b>30.8</b> M	89.72	<b>84.20</b>	<b>55.81</b>	77.63	<b>80.25</b>	84.45	88.12	90.88	87.57	86.07	<b>66.86</b>	70.23	<b>77.47</b>	73.62	62.94	<b>78.39</b>
AO-RCNN (本文)	31.5 M	<b>89.88</b>	83.70	54.74	<b>78.87</b>	79.05	<b>84.97</b>	<b>88.42</b>	<b>90.90</b>	<b>88.10</b>	86.37	66.46	63.20	<b>77.37</b>	<b>79.53</b>	<b>73.87</b>	<b>79.03</b>

表 3 基于 Oriented R-CNN 框架做出改进的 SOTA 方法在 DOTA-v1.0 数据集上比较

Table 3 Comparison of sota methods based on Oriented R-CNN framework on the DOTA-v1.0 dataset

方法	All categories in DOTA-v1.0															AP <sub>50</sub>	
	Backbone	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP		HC
O-RCNN	R50	89.46	82.12	54.78	70.86	78.93	83.00	88.20	90.90	87.50	84.68	63.97	67.69	74.94	68.84	52.28	75.87
OAN	R50	89.70	84.03	54.61	73.46	79.30	83.27	88.12	90.90	85.22	84.68	62.06	66.87	75.28	70.71	62.75	76.73
ARC	R50	89.40	82.48	55.33	73.88	79.37	84.05	88.06	90.90	86.44	84.83	63.63	70.32	74.29	71.91	65.43	77.35
LSKNet-S	LSK-S	89.79	84.06	53.36	78.79	79.11	84.93	88.22	90.90	87.41	86.75	66.68	63.68	77.02	71.05	70.31	78.14
PKINet-S	PKINet-S	89.72	84.20	55.81	77.63	80.25	84.45	88.12	90.88	87.57	86.07	66.86	70.23	77.47	73.62	62.94	78.39
AO-RCNN (本文)	LSK-S	89.88	83.70	54.74	78.87	79.05	84.97	88.42	90.90	88.10	86.37	66.46	63.20	77.37	79.53	73.87	79.03

为了进一步展示本文方法在弱小目标检测上的卓越表现。本文对基线网络与所提方法进行了多尺度级别特征可视化及检测结果可视化,如图 6 所示。Oriented R-CNN 的低级特征图所表示的特征并不细腻,即存在较大范围的特征聚集,这并非底层特征所应该表现出的特性,并且它似乎更加关注实例边缘。Oriented R-CNN 的中级特征表现出了和低级特征类似的特性,并且较高级特征表示并不直观、有待改进。从检测效果上看,其在小目标

检测方面的检测效果也并不理想,存在大量目标的漏检。本文提出的方法通过可选择核大小的的大核卷积来扩展感受野,充分利用了 RSI 的上下文信息;同时通过一致性增强模块提高了模型对噪声的鲁棒性,改善了底层特征的特征表示(特征图更加细腻、并非仅仅关注实例边缘)。此外,CS 策略做到了降低不同尺度语义鸿沟(中级特征与较高级特征的具有较高的相似度)。本方法缓解了基线方法的漏检问题,在弱小目标检测上更加具有优势。

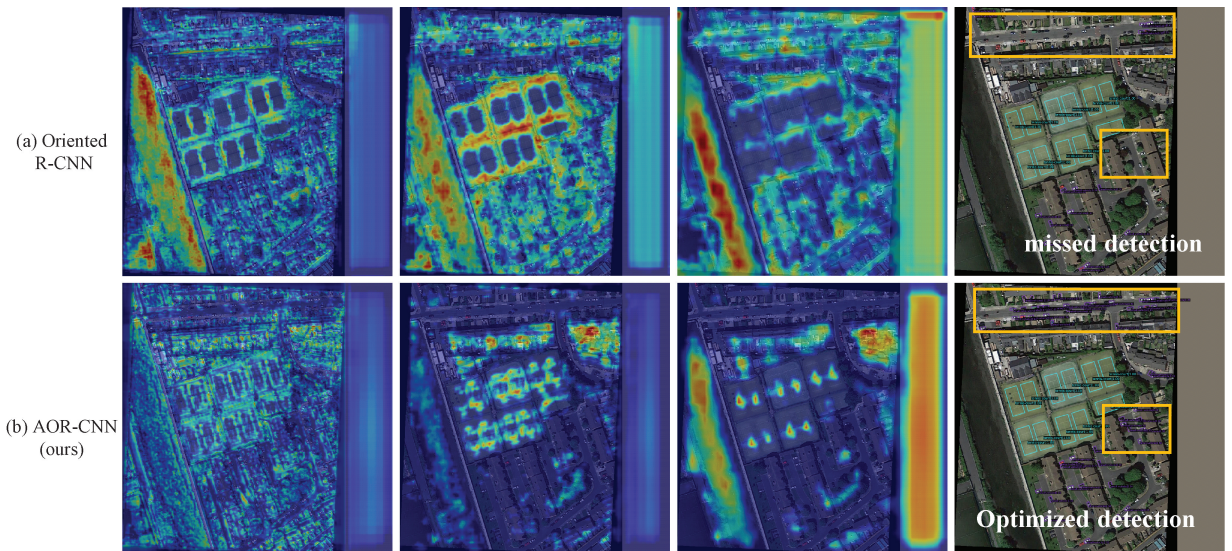


图 6 Oriented R-CNN 和 AO-RCNN 的多尺度级别特征可视化与检测结果可视化图(左:底层特征;右:较高级特征)

Fig. 6 Multi-Scale feature and detection results visualization for Oriented R-CNN and AO-RCNN (left: low-level; right: high-level)

本文给出了 AO R-CNN 和最新研究在 DOTA v1.0 测试集上的检测结果可视化,如图 7 所示。在背景复杂的遥感场景中,本文方法在弱小目标检测上更具优势(改善了目标的漏检、错检问题)。此外,为了显示提出方法在遥感定向检测场景中准确识别、检测多种弱小目标的优越性,本文广泛选取了 DOTA-v1.0 测试集中较为困难的检测场景,如排列时而稀疏时而密集、角度方向变化、存在背景噪声干扰、目标大小分布不均匀且尺度变化剧烈等场景,并进行了可视化。如图 8 所示,所提方法在弱小目标(如 SH、SV 或 LV)和复杂背景目标上的识别效果优越。

### 2.3 消融实验

为了探究 AO-RCNN 中各个模块的重要性,本文逐步

将 LSK-S、RFA 模块、CS 策略引入到 Oriented R-CNN 中来验证它们的相互影响。本节将详细介绍模块消融实验以及各模块的超参数调整实验。

首先,介绍不同模块组合的实验结果。如表 4 所示,以 Oriented R-CNN 的精度结果作为标准,首先复现了 LSK-S 在单尺度下的精度。然后分别集成 RFA、CS 单模块到 Oriented R-CNN 模型中,RFA 通过提高模型对噪声的鲁棒性带来了 0.83% 的 mAP<sub>50</sub> 增益,CS 通过降低不同尺度之间的尺度鸿沟带来了 0.43% 的 mAP<sub>50</sub> 增益,结果表明了单个模块有效性。然而由于深度学习的黑盒性质,不同模块之间可能并非是相互促进的关系。鉴于此,本文也进行了后续不同模块组合的消融实验。实验表明 RFA



图 7 AO R-CNN 和最新方法在 DOTA v1.0 测试集上的可视化图  
Fig. 7 AO R-CNN and latest methods visualization on DOTA v1.0 test

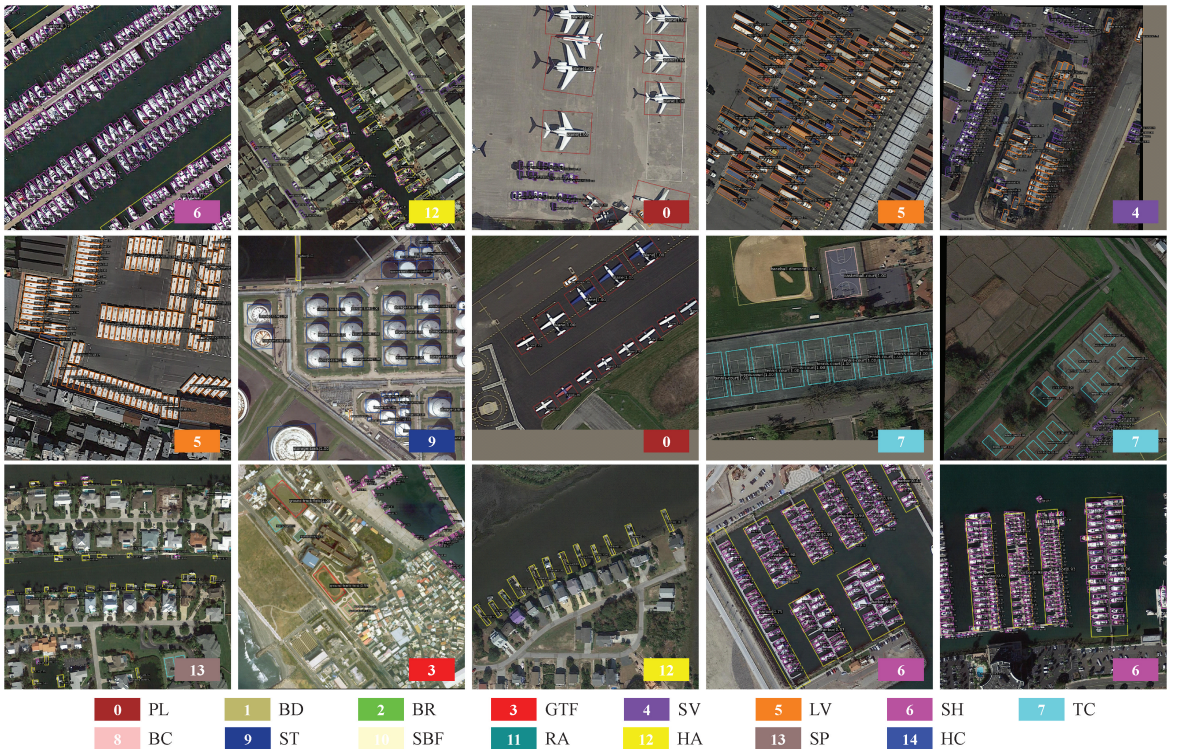


图 8 AO-RCNN 在 DOTA-v1.0 测试集上的结果可视化图  
Fig. 8 AO-RCNN results visualization on DOTA v1.0 test

表 4 AO-RCNN 设计方法的消融实验结果

Table 4 Ablation results for AO-RCNN design method

方法	组合模块	mAP <sub>50</sub>	mAP <sub>75</sub>	mAP	Param/M
O R-CNN	无	75.87	48.38	46.03	41.14
	LSK-S	78.14	51.62	48.09	30.98
	RFA	76.70	49.03	46.33	42.92
	CS	76.30	48.91	46.03	41.14
AO R-CNN(本文)	RFA+CS	76.69	49.73	46.78	56.83
	LSK-S+RFA	78.46	50.35	47.74	31.57
	LSK-S+CS	78.42	51.86	48.76	30.98
	LSK-S+RFA+CS	79.03	50.97	48.13	31.57

和 CS、LSK-S 和 RFA、LSK-S 和 CS 相互组合相对于基线分别带了 0.82%、2.59% 和 2.55% 的 mAP<sub>50</sub> 增益。最后,集成了 LSK-S、RFA、CS 3 个模块的 AO-RCNN 以 3.16% 的 mAP<sub>50</sub> 增益达到了最高的 79.03% 结果。这充分表明了所用方法的单模块有效性,且体现了各模块之间的正向

促进作用。

其次,介绍超参数实验结果。如表 5 所示,为了探究 RFA 中比例不变的自适应池化的池化因子的影响,本文设置  $\alpha_1$ 、 $\alpha_2$ 、 $\alpha_3$  的值分别为 (0.1, 0.4, 0.7)、(0.1, 0.3, 0.5)、(0.1, 0.2, 0.3)。

表 5 AO-RCNN 设计方法的超参数实验结果

Table 5 Hyperparameter results for AO-RCNN design method

组合模块	Settings	mAP <sub>50</sub>	mAP <sub>75</sub>	mAP
无	—	75.87	48.38	46.03
RFA	0.1, 0.4, 0.7	76.26	48.92	46.65
	0.1, 0.3, 0.5	76.45	49.07	46.02
	0.1, 0.2, 0.3	76.70	49.03	46.33
CS	0.50	75.20	48.46	45.43
	0.25	76.17	49.26	46.47
	0.125	<b>76.29</b>	48.90	46.03
LSK+RFA	0.1, 0.4, 0.7	78.12	51.58	47.57
	0.1, 0.3, 0.5	77.88	51.46	47.74
	0.1, 0.2, 0.3	<b>78.46</b>	50.35	47.74
LSK+RFA+CS	0.1, 0.2, 0.3, 0.25	77.80	51.78	48.01
	0.1, 0.2, 0.3, 0.125	78.45	52.07	48.63
	0.1, 0.2, 0.3, 0.062 5	<b>79.03</b>	50.97	48.13

本文认为这 3 组参数几乎涵盖了大、中、小的尺度分布。实验结果表明 RFA 模块凭借较小尺度因子 (0.1, 0.2, 0.3) 达到最优精度,这符合直觉。即,特征金字塔顶层特征通过和 RFA 产生的更高维度语义特征融合来强化自身对噪声的抗干扰性。此外,本文也调整辅助损失的权重  $\lambda$  为 0.50、0.25 和 0.125 来探究 CS 损失的最优表现。实验表明,辅助损失的权重值  $\lambda$  较小时,可以带来更好的辅助训练效果。最后,经过反复超参数实验,本文得到在比例不变的自适应池化的池化因子为 (0.1, 0.2, 0.3), CS 损失中的辅助损失权重  $\lambda$  为 0.062 5 时, AO-RCNN 以 3.16% 的增益达到了 79.03% AP<sub>50</sub> 的最优表现,超过了大多数旋转目标检测模型的单尺度精度结果。

### 3 结 论

本文以 LSKNet 作为基线,通过引入 RFA 模块和 CS 策略进而提出一种改进 Oriented R-CNN 的遥感图像定向目标检测算法,即 AO-RCNN。其基于特征提取骨干网络 LSK-S 的可选择性大核和空洞卷积来扩展感受野,为对象提供足够的上下文信息;其次通过构建的 RFA 模块,缓解了大核卷积引入背景噪声干扰的问题;最后,引入了 CS 策略,进一步通过 CS 损失约束多尺度特征金字塔的训练过程,缩小特征融合前不同尺度特征之间的语义差距。AO-RCNN 综合了 LSKNet 的核选择、RFA 模块对噪声的高鲁棒性以及 CS 策略在多尺度特征上的指导作用,保证自身

轻量的同时,也保持了双阶段定向检测器的高精度优势。其在大型遥感数据集 DOTA-v1.0 上的单尺度精度达到 79.03% AP<sub>50</sub> 的性能,超越了大多数先进旋转检测算法的单尺度结果。

尽管 CS 策略并没有影响模型的推理参数量和运算量,但需要更多的训练资源。在后续的工作中,将进一步优化训练算法,针对训练时的计算资源消耗等问题,通过采用更高效的计算策略和优化技术来降低训练成本,从而提高算法的整体效率。

## 参考文献

- [1] 过鑫炎,朱硕,孙佳豪,等. 基于注意力机制融合特征的车辆目标检测方法[J]. 电子测量技术,2024,47(9): 52-60.  
GUO X Y, ZHU SH, SUN J H, et al. A vehicle object detection method based on feature fusion with attention mechanism [J]. *Electronic Measurement Technology*, 2024, 47(9): 52-60.
- [2] 张慧敏,李锋,黄炜嘉,等. 基于 CAM-YOLOX 的大场景 SAR 图像近岸场景舰船目标检测[J]. 电子测量技术,2024,47(6):86-93.  
ZHANG H M, LI F, HUANG W J, et al. Ship object detection in coastal scenes of large-area SAR images based on CAM-YOLOX [J]. *Electronic Measurement Technology*, 2024, 47(6):86-93.
- [3] DING J, XUE N, XIA G S, et al. Object detection in aerial images: a large-scale benchmark and challenges[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(11): 7778-7796.
- [4] DAI L H, LIU H, TANG H, et al. AO<sup>2</sup>-DETR: Arbitrary-oriented object detection transformer[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 33(5): 2342-2356.
- [5] MA T L, MAO M Y, ZHENG H H, et al. Oriented object detection with transformer[EB/OL]. [2024-08-10]. <https://arxiv.org/abs/2106.03146>.
- [6] ZENG Y, CHEN Y SH, YANG X, et al. ARS-DETR: Aspect ratio sensitive oriented object detection with transformer[EB/OL]. [2024-08-10]. <https://arxiv.org/abs/2303.04989>.
- [7] HOU L P, LU K, XUE J, et al. Shape-adaptive selection and measurement for oriented object detection [C]. *AAAI Conference on Artificial Intelligence*, 2022, 36(1): 923-932.
- [8] YANG X, YAN J CH, FENG Z M, et al. R3Det: Refined single-stage detector with feature refinement for rotating object[C]. *AAAI Conference on Artificial Intelligence*, 2021, 35(4): 3163-3171.
- [9] YANG X, YAN J CH, MING Q, et al. Rethinking rotated object detection with gaussian wasserstein distance loss[C]. *International Conference on Machine Learning*, PMLR, 2021: 11830-11841.
- [10] YANG X, YANG X J, YANG J R, et al. Learning high-precision bounding box for rotated object detection via Kullback-Leibler divergence [J]. *Advances in Neural Information Processing Systems*, 2021, 34: 18381-18394.
- [11] LI W T, CHEN Y J, HU K X, et al. Oriented reppoints for aerial object detection[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 1829-1838.
- [12] TIAN ZH, SHEN CH H, CHEN H, et al. FCOS: A simple and strong anchor-free object detector [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 44(4): 1922-1933.
- [13] HAN J M, DING J, LI J, et al. Align deep features for oriented object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-11.
- [14] YANG X, YANG J R, YAN J CH, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects [C]. *IEEE/CVF International Conference on Computer Vision*, 2019: 8232-8241.
- [15] XU Y CH, FU M T, WANG Q M, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 43(4): 1452-1459.
- [16] LONG Y, XIA G S, LI SH Y, et al. On creating benchmark dataset for aerial image interpretation: reviews, guidances, and million-AID [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 4205-4230.
- [17] HAN J M, DING J, XUE N, et al. ReDet: A rotation-equivariant detector for aerial object detection[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 2786-2795.
- [18] DING J, XUE N, LONG Y, et al. Learning RoI transformer for oriented object detection in aerial images [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 2849-2858.
- [19] REN SH Q, HE K M, 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.
- [20] XIE X X, CHENG G, WANG J B, et al. Oriented R-

- CNN for object detection [C]. IEEE/CVF International Conference on Computer Vision, 2021: 3520-3529.
- [21] FENG M K, YU H CH, DANG X Y, et al. Category-aware dynamic label assignment with high-quality oriented proposal [EB/OL]. [2024-08-10]. <https://doi.org/10.48550/arXiv.2407.03205>.
- [22] XIE X X, CHENG G, LI Q Y, et al. Fewer is more: Efficient object detection in large aerial images [J]. Science China Information Sciences, 2024, 67(1): 112106.
- [23] LI Y X, HOU Q B, ZHENG ZH H, et al. Large selective kernel network for remote sensing object detection [C]. IEEE/CVF International Conference on Computer Vision, 2023: 16794-16805.
- [24] PU Y F, WANG Y R, XIA ZH F, et al. Adaptive rotated convolution for rotated object detection [C]. IEEE/CVF International Conference on Computer Vision, 2023: 6589-6600.
- [25] CAI X H, LAI Q X, WANG Y W, et al. Poly kernel inception network for remote sensing detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 27706-27716.
- [26] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3974-3983.
- [27] HENDRYCKS D, GIMPEL K. Gaussian error linear units (GELUs) [EB/OL]. [2024-08-10]. <https://doi.org/10.48550/arXiv.1606.08415>.
- [28] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [29] ZHOU Y, YANG X, ZHANG G F, et al. MMRotate: A rotated object detection benchmark using PyTorch [C]. 30th ACM International Conference on Multimedia, 2022: 7331-7334.
- [30] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [31] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.

### 作者简介

王雷雨, 硕士研究生, 主要研究方向为目标检测、计算机视觉。

E-mail: 18538331073@163.com

王正勇(通信作者), 硕士生导师, 主要研究方向为图像处理、模式识别、计算机视觉、智能系统。

E-mail: 690728634@sina.com

陈洪刚, 副研究员, 博士生导师, 主要研究方向为图像处理、计算机视觉、人工智能、多媒体通信等。

E-mail: honggang\_chen@scu.edu.cn

何小海, 教授, 博士生导师, 主要研究方向为图像处理、模式识别、计算机视觉、图像通信、软件工程等。

E-mail: hxh@scu.edu.cn