

## 基于特征重组的遥感图像有向目标检测

王友伟<sup>1,2</sup>, 郭颖<sup>1,2\*</sup>, 邵香迎<sup>1,2</sup>, 王季宇<sup>1,2</sup>, 鲍正位<sup>1,2</sup><sup>1</sup>南京信息工程大学江苏省大气环境与装备技术协同创新中心, 江苏 南京 210044;<sup>2</sup>南京信息工程大学自动化学院, 江苏 南京 210044

**摘要** 针对遥感有向目标存在的检测问题,设计了一个基于改进Rotated RPN的网络,设计特征重组机制,通过加权使网络关注有效目标区域。使用新的有向框标注方法,避免在临界角度出现错位等问题。在检测头前端使用极化注意力模块,改善因为分类和回归任务所需特征不一致导致的性能下降问题。实验结果表明,该模型可以提高多类目标的检测精度。相较于基准Rotated RPN,该模型在Dior-R数据集上精度提升4.95%,在HRSC2016数据集上精度提升11.75%。

**关键词** 遥感; 有向目标检测; 深度学习; 特征重组; 极化注意力

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/AOS230957

## 1 引言

光学遥感图像有向目标检测是指将给定的光学遥感数据集输入检测模型,给出带有目标定位框、目标类别以及置信度的过程,其无论是在民用领域还是在军用领域都有着较大的应用价值<sup>[1-2]</sup>。因此,遥感图像目标检测具有重要的研究意义和良好的应用前景。

由于拍摄角度和拍摄距离不同,遥感图像与自然图像相比具有小目标规模大、不同目标尺寸相差大、背景复杂等检测难点,从而影响检测性能。结合深度学习在检测领域取得的成绩,人们研究了许多方法来实现遥感图像中的目标检测。针对遥感图像定向目标检测框架,FFAM-YOLO<sup>[3]</sup>利用级联注意力机制解决遥感图像中小目标特征信息少、定位困难的问题。FFC-SSD<sup>[4]</sup>采用分组聚类获得更合适的样本尺寸分布,设计反池化高效多尺度特征融合模块增强特征提取能力。Suresh等<sup>[5]</sup>利用滑动窗口提取不同的特征,然后根据提前设置好的字典中的特征与目标进行匹配。Xu等<sup>[6]</sup>放弃使用IoU作为样本分配策略,提出一种基于高斯分布的标签分配策略,使用高斯感受野距离来度量真实值与预测值之间的距离。Nan等<sup>[7]</sup>为模型引入注意力机制,关注存在目标的高响应区域,提升模型的检测性能。

由于遥感图像的拍摄角度为俯视,目标的方向各异。传统的定向框难以高效地标目标,导致模型的检测性能受限。针对遥感图像有向目标检测框架,徐志京等<sup>[8]</sup>设计生成对抗网络DEGAN来增强小目标的

纹理特征,采用深度强化学习增强图像的颜色信息,设计自适应特征变换金字塔对特征进行提取。Zhao等<sup>[9]</sup>提出基于注意力机制的特征融合模块、可变形横向连接模块和基于锚框引导的检测模块,对飞机的特征进行细化调整。Yang等<sup>[10]</sup>使用分割网络来指导检测模型,设计了去噪模块,可以突出小目标的特征。Ding等<sup>[11]</sup>在RPN阶段只生成水平预选框,之后模型通过轻量的全连接层将水平候选框旋转为有向候选框,然后从有向候选框中获取旋转区域特征,对旋转框进行回归。为了解决在临界角度出现的不连续性和方形问题,Yang等<sup>[12]</sup>提出使用高斯距离计算回归损失。Zhang等<sup>[13]</sup>针对目标定位问题在前人的基础上进行优化,固定使用特征金字塔中的某两层输入RPN网络,解决模型对于不同层的选择问题。为了解决遥感图像背景复杂的问题,Li等<sup>[14]</sup>在多尺度的基础上引入视觉注意力机制,并在不同尺寸图上单独使用,从而削弱背景噪声的影响。Ma等<sup>[15]</sup>提出的Rotated RPN算法放弃在原有的水平框基础上加入角度参数作为有向框标注,直接生成任意方向的建议框。

本文面向遥感图像有向目标检测任务,针对通用模型对遥感有向目标检测性能不足的问题,提出一种基于特征重组和极化注意力机制的有向目标检测模型。针对遥感小目标特征难以提取的问题,设计基于特征重组的特征金字塔模块,保证提取特征的完整性。针对常用有向框标注法存在的临界角度问题,设计自适应建议区域生成模块,缓解角度不连续引起的损失变大问题。针对常见检测头在使用遥感目标特征信息

收稿日期: 2023-05-10; 修回日期: 2023-05-30; 录用日期: 2023-06-27; 网络首发日期: 2023-07-10

基金项目: 国家自然科学基金(61971229)

通信作者: \*yguo@nuist.edu.cn

时存在的不一致性问题,设计极化注意力检测头,区分回归和分类头所需的不同特征,提升检测精度。该模型在 Rotated RPN 模型的基准上进行优化,实验结果表明,改进后模型在有向遥感数据集上的检测精度明显提升。

## 2 PFR-Rotate 模型

### 2.1 基准 Rotated RPN 模型

Rotated RPN<sup>[15]</sup>原本是针对文本检测提出的,用于处理面向任意方向的文本检测任务。由于遥感图像目标检测任务的兴起,不少学者将其引入遥感图像检测任务中。Rotated RPN 主要工作有两点:1)提出了 RRPN 这一网络结构,这种结构可以为 RPN 直接生成带有方向参数的预测框,从而使预测框能更贴合有向目标,达到提升精度的目的;2)为了配合 RRPN,将生成

的任意方向的建议框映射到特征图上再进行池化,提出了 RROI。Rotated RPN 是在二阶段检测模型的基础上进行的改进,整体框架如图 1 所示。

Rotated RPN 在光学遥感图像有向目标检测任务中精度仍有提升空间,问题可能在于:1)遥感数据集中有向小目标规模大,特征难以提取完全,Rotated RPN 使用的传统主干网络以及特征金字塔网络对分布密集的小目标关注度不够高,导致模型无法对所有小目标精准定位;2)Rotated RPN 在每个锚点上生成多个角度的锚框,但仍然难以兼顾到所有角度,且导致计算量增加;3)Rotated RPN 使用的通用检测头将分类和回归任务耦合在一起,没有考虑到不同任务所需特征的不一致性。本文针对以上问题,在 Rotated RPN 的基础上进行优化改进,提出 PFR-Rotate 模型。

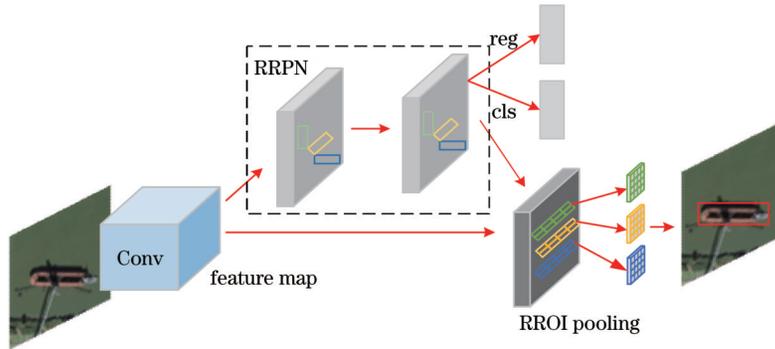


图 1 Rotated RPN 整体框架

Fig. 1 Framework of Rotated RPN

为了缓解遥感图像中有向目标存在小目标规模大、背景比例较大、目标密集等问题,设计了一种基于特征重组和极化注意力的有向目标检测模型 PFR-

Rotate,改进的模型增强了不同尺寸特征之间的互补性,提升了模型的检测能力。PFR-Rotate 模型结构如图 2 所示。

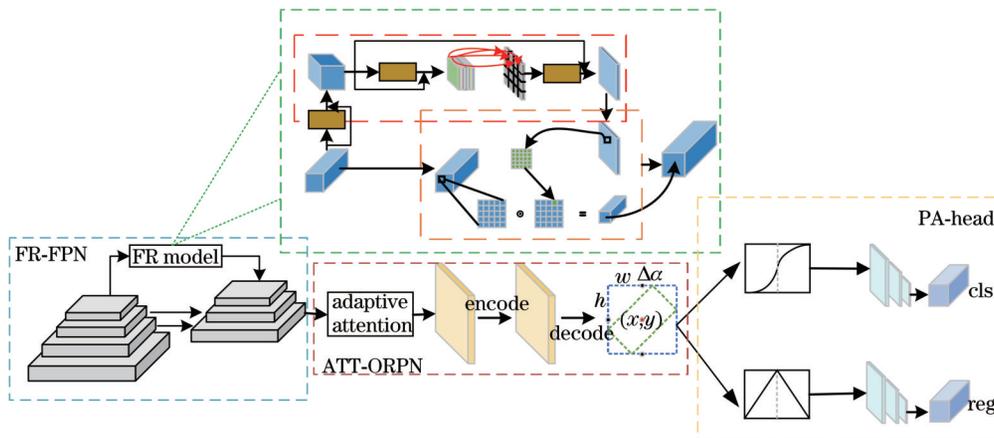


图 2 PFR-Rotate 模型结构图

Fig. 2 Framework of PFR-Rotate

### 2.2 基于特征重组的特征金字塔模块

遥感图像中小目标规模大、密集分布且背景占比较大比例,导致特征难以提取,影响检测性能。通用的

特征金字塔模块尽管考虑到不同尺寸之间的融合,但并未针对小目标的特征提取采取有效措施,因此在数据集存在密集小目标的情况下检测效果不佳。本文为

加强模型特征提取能力以及对密集小目标的检测性能,提出FR-FPN模块,在特征金字塔最深层使用FR模块作为横向连接。

FR模块包括两部分:第一部分是上采样核预测模块。首先,对输入特征图进行通道压缩,减少后续计算量。将上采样核 reshape 为  $1 \times C$  和  $C \times 1$  尺寸( $C$ 为卷积核通道数)并进行有机融合,得到新的融合特征间相关性的上采样核。reshape 部分结构如图 3 所示, reshape 可以增加特征各通道之间的关联性,保证后续模型能够使用更为完整的遥感目标特征信息。然后,使用上采样核对上采样部分进行预测,使用 softmax 对计算结果归一化。为了保证特征信息的完整性,将初

始特征作为信息补充送入后续各阶段。第二部分是特征重组模块。这一部分将上采样核预测模块得到的结果与映射到特征图相应位置中  $k \times k$  ( $k$ 为卷积核的尺寸)区域的值作点积得到最终结果。FR模块可以充分利用特征图的语义信息,在保证较大感受野的同时保持较少的计算量。图 4 为特征重组模块的结构图。

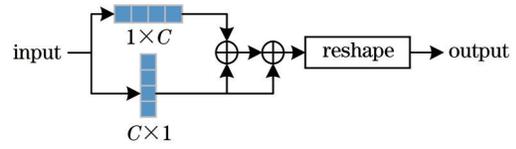


图 3 reshape 模块结构图  
Fig. 3 Framework of reshape model

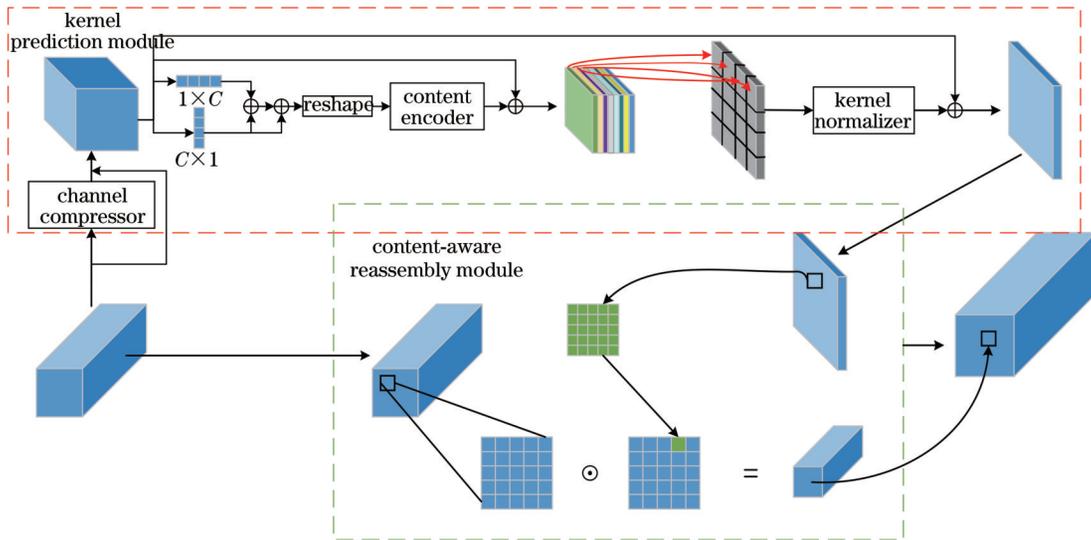


图 4 特征重组模块(FR)结构图  
Fig. 4 Framework of feature recombination model

### 2.3 基于特征自适应的建议区域生成模块

常见的有向框定义法有 opencv 定义法、长边定义法、八参数定义法等。opencv 表示法由于角度定义不连续,在临界角度时需要转动较大角度,回归损失较大。长边表示法定义较长边与 X 轴的夹角为旋转角,但预测框刚好是正方形时没有真正意义上的长边,表示方式不唯一。八参数定义法在临界角度时会出现顶点顺序交换导致损失变大的问题。Rotated RPN 为了避免这些临界问题,在生成预选框时直接生成多个多角度预测框,但计算量也因此增大,且生成的预测框不能兼顾到所有情况。

本文使用改进的中点偏移法<sup>[16]</sup>来定义有向框,可以有效解决传统表示方法中角度周期性、边顺序交换、方形问题。锚框使用 6 个参数来定义,其中  $x, y$  表示候选框外接矩形的中心点位置,  $w, h$  表示候选框外接矩形的宽度和高度,  $\Delta\alpha, \Delta\beta$  表示候选框顶点与外接矩形宽和高中点的偏移量。中点偏移定义法如图 5 所示。

在得到 6 个参数后,通过式(1)计算可以得到候选框 4 个顶点坐标  $(v_1, v_2, v_3, v_4)$ 。

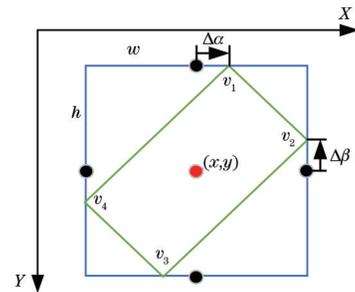


图 5 中点偏移定义法  
Fig. 5 Midpoint offset definition method

$$\begin{cases} v_1 = \left(x, y - \frac{h}{2}\right) + (\Delta\alpha, 0) \\ v_2 = \left(x + \frac{w}{2}, y\right) + (0, \Delta\beta) \\ v_3 = \left(x, y + \frac{h}{2}\right) + (-\Delta\alpha, 0) \\ v_4 = \left(x - \frac{w}{2}, y\right) + (0, -\Delta\beta) \end{cases} \quad (1)$$

ATT-ORPN 流程如图 6 所示。考虑到遥感目标特征表示不明显的问题,在 RPN 模块编码前使用自适应注意力机制来加强有效目标的表示能力。自适应注意力机制采用多分支并行的策略,一个分支使用自适应全局池化和全连接,另一分支使用 Conv+BN+siReLU 为特征赋予不同的权重。之后对有向框进行编码解码,获得有向框的表示。ATT-ORPN 可以解决常见有向框表示法存在的各种边界问题,并且提升检测精度。此外,为了配合 ATT-ORPN 的使用,本文还使用了 RROI Align。

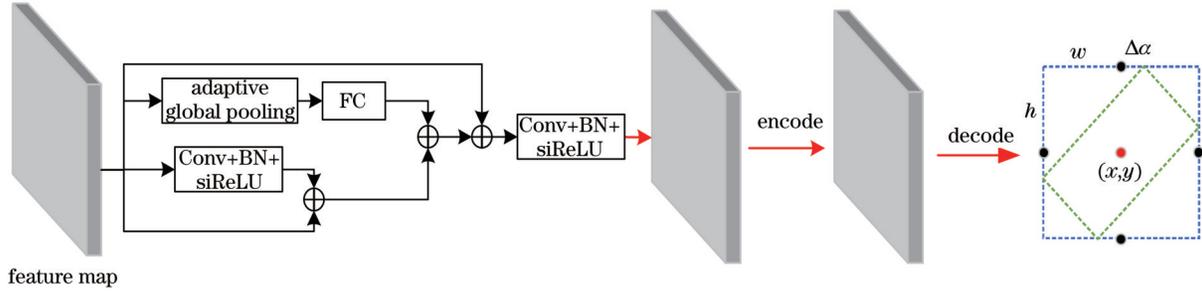


图 6 ATT-ORPN 流程图

Fig. 6 Flow chart of ATT-ORPN

### 2.4 极化注意力检测头模块

常见的检测头是耦合的,分类和回归使用的特征共享,忽视了不同任务之间所需特征的差异性,而在处理遥感任务时,模型对遥感目标特征提取能力不足,导致特征差异性对检测能力影响更大。分类任务的关注点应该在目标本身,需要对预测框内部有效信息高响应,而回归任务只需要关注目标边界区域,应该减少对内部信息的关注,减少干扰。本文在极化注意力<sup>[17]</sup>的基础上设计了改进的双分支极化注意力检测头,分类分支使用高响应的全局特征,以减少噪声干扰,回归分支关注边界特征,抑制不相关的高激活

区域对回归的影响。双分支极化注意力检测头结构如图 7 所示。

极化检测头模块有两个分支,两个分支在经过空间注意力和通道注意力并联的结构后分别使用不同的响应函数,使用注意力机制可以进一步让模型聚焦于不同任务所需的不同特征。分类分支为了选择高响应区域作为分类的依据而使用激活函数,回归分支为了抑制高响应区域并使模型关注边界区域帮助回归任务的完成而使用抑制函数。极化注意力检测头在每一步完成后都结合原始特征,保证特征信息在模型处理时不丢失。

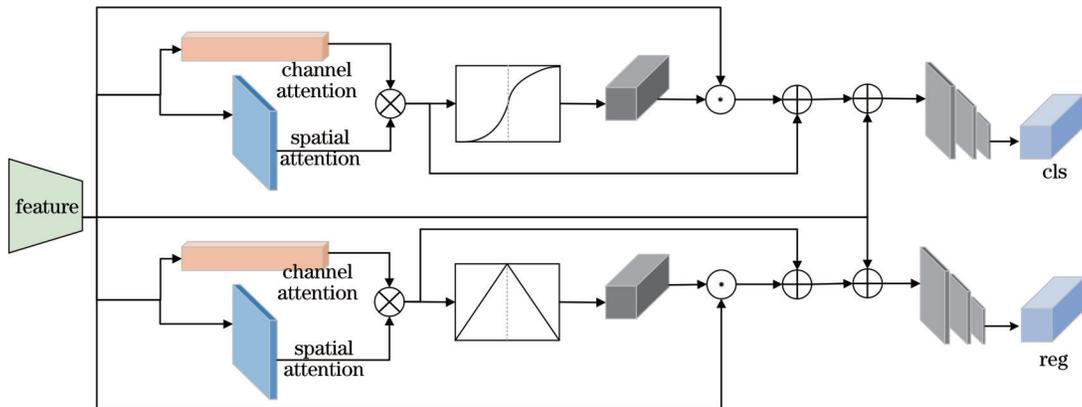


图 7 双分支极化注意力检测头结构图

Fig. 7 Structure diagram of double polarization attention head

### 2.5 训练策略

目前,目标检测模型在训练时使用的训练策略多为 SGD 方法,一般会选择最后一个 epoch 的权重,或是在验证集上取得最优效果的 epoch 权重作为最终模型的权重。但传统的 SGD 方法训练的模型存在局限性,会收敛到最优解的边界区域。本文使用 SWA 训练策略<sup>[18]</sup>,在 SGD 的基础上循环运行,即通过 SGD 方法额外训练多个 epoch 得到对应权重空间边界附近的多个

点,平均这些点的结果,找到一个接近最优解的结果权重。由于平均了多个点的结果,SWA 相比 SGD 具有更好的泛化性。

## 3 实验结果与分析

### 3.1 数据集

Dior-R 数据集<sup>[19]</sup>由西北工业大学韩军伟课题组于 2022 年提出,由 23463 幅图像和 190288 个目标实例组

成,共包含 20 个类别:飞机(APL)、机场(APO)、棒球场(BF)、篮球场(BC)、桥梁(BR)、烟囱(CH)、水坝(DAM)、高速公路服务区(ETS)、高速公路收费站(ESA)、高尔夫球场(GF)、田径场(GTF)、港口(HA)、立交桥(OP)、船只(SH)、体育场(STA)、储油罐(STO)、网球场(TC)、火车站(TS)、车辆(VE)和火车(WM)。由于成像条件、天气、季节不同,成像结果差异较大,并且具有较高的类间相似性和类内多样性,这些因素都增加了检测难度。

HRSC2016 数据集由西北工业大学于 2016 年发布,图像尺寸范围为 300~1500 pixel。训练集、验证集和测试集分别有 436 幅图像(包括 1207 个样本)、181 幅图像(包括 541 个样本)和 444 幅图像(包括 1228 个样本)。HRSC2016 与 Dior-R 数据集一样使用有向标注格式,尽管 HRSC2016 数据集只有一个大类,但其目标尺寸差异更大,背景颜色干扰更为严重,因此可以用作本文模型的泛化实验。

### 3.2 实验设置

本实验基于 mmrotate 框架<sup>[20]</sup>,实验环境为 Python 3.8、Pytorch 1.7.0、torchvision 0.7.0, batch size 为 2, 动量因子为 0.9,权重衰减系数为 0.0005,所有实验均采用 Resnet50<sup>[21]</sup>作为主干网络,以 50% 的概率随机翻转图像作为数据增强手段。在 Dior-R 数据集上训练 12 个 epoch,学习率初始值设置为 0.001,在迭代到第 9 和 11 个 epoch 后学习率分别下降到  $1 \times 10^{-4}$  和  $1 \times 10^{-5}$ 。在 HRSC2016 数据集上训练 40 个 epoch,学习率初始值为 0.005。硬件设备为 Intel® Core™ i9-10900X CPU、NVIDIA RTX3080Ti 显卡。

### 3.3 消融实验

FPN 层关系着模型对目标特征提取以及融合的能力,对检测性能有着重要的影响。为了验证在 FPN 深层引入特征重组模块的有效性,在 Dior-R 数据集上针对 FPN 设置了一组消融实验,将 reshape 思想融入 SPP、CBAM 等模块中,实验结果如表 1 所示,其中  $f$  表示帧率。

表 1 FPN 部分消融实验  
Table 1 Comparison of FPN ablation experimental results

Baseline	SPP	CBAM	SE	Carafe	reshape	FR	mAP / %	mAP variation / %	Recall / %	$f / (\text{frame} \cdot \text{s}^{-1})$
✓							59.54	—	—	—
✓	✓						62.26	↑ 2.72	71.07	56.6
✓	✓				✓		62.32	↑ 2.78	71.11	54.3
✓		✓					62.42	↑ 2.88	71.29	56.7
✓		✓			✓		62.49	↑ 2.95	71.31	51.9
✓			✓				62.77	↑ 3.23	70.18	56.6
✓			✓		✓		62.86	↑ 3.32	70.35	55.3
✓				✓			62.90	↑ 3.36	70.42	54.6
✓						✓	63.03	↑ 3.49	70.68	51.3

从表 1 可看出,SPP 模块融合了多尺度的特征信息,能够较好地解决遥感图像中目标尺寸变化较大的问题,而本文使用的数据集目标尺寸变化也较大,所以 SPP 模块能提升检测精度。CBAM 模块通过串联通道注意力和空间注意力模块在特征金字塔深层进一步提取语义信息,但模型较深时性能提升不大。从实验结果来看,针对稠密目标预测的 Carafe 模块提升效果是最好的,且速度下降在可接受范围内。各模块在使用 reshape 结构后检测性能均有小幅提升。本文通过改进基于 reshape 的 Carafe 模块,得到 FR-FPN。

检测头模块关系着模型针对特征的分析能力,影响最终的检测结果。为了验证极化注意力检测头中极化函数选择的有效性,针对 PA-head 中分类分支和回归分支使用的极化函数进行了一组消融实验,分别使用式(2)~(7)进行组合,挑选效果最好的式(4)和式(5)组合作为最终模型。极化函数消融实验结果如表 2 所示,✓表示使用该公式。

$$f_{\text{cls1}} = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}, \quad (2)$$

$$f_{\text{cls2}} = \frac{1}{1 + e^{(x-0.5)}}, \quad (3)$$

$$f_{\text{cls3}} = \frac{1}{1 + e^{-\eta(x-0.5)}}, \quad (4)$$

$$f_{\text{reg1}} = \begin{cases} x, & x < 0.5 \\ 1 - x, & \text{otherwise} \end{cases}, \quad (5)$$

$$f_{\text{reg2}} = -[(x - 0.5)^2 + C], \quad (6)$$

$$f_{\text{reg3}} = \begin{cases} \frac{1}{2x} - C, & x > 0 \\ 0, & x = 0 \\ -\frac{1}{2x} - C, & x < 0 \end{cases}. \quad (7)$$

针对分类分支中式(4)的超参数  $\eta$  进行了消融实验,实验结果如表 3 所示。根据实验结果,最终取  $\eta = 15$  作为后续实验的设置值。

为了验证本文模型各模块的有效性,在 Dior-R 数

表 2 极化函数消融实验

Table 2 Comparison of polarization function ablation experimental results

cls			reg			mAP / %
Eq. (2)	Eq. (3)	Eq. (4)	Eq. (5)	Eq. (6)	Eq. (7)	
✓			✓			61.38
✓				✓		60.87
✓					✓	61.03
	✓		✓			61.97
	✓			✓		61.52
	✓				✓	61.59
		✓	✓			62.58
		✓		✓		61.70
		✓			✓	61.96

表 3 式(4)中  $\eta$  取值的消融实验

Table 3 Ablation experiment for value selection about  $\eta$  in Eq. (4)

Models	$\eta=5$	$\eta=10$	$\eta=15$	$\eta=20$
mAP / %	62.13	62.37	62.58	62.55

数据集上设置了一组消融实验,以 Rotated RPN 作为基准模型,分别使用所改进的各部分模块在基准网络上进行实验,实验结果如表 4 所示,其中 ✓ 表示模型含有该模块。

表 4 消融实验结果对比

Table 4 Comparison of ablation experimental results

Baseline	FR-FPN	ATT-ORPN	PA-head	SWA	mAP / %	mAP variation / %	Recall / %	$f / (\text{frame} \cdot \text{s}^{-1})$
✓					59.54	—	—	—
✓	✓				63.03	↑ 3.49	70.68	51.3
✓		✓			62.36	↑ 2.82	69.90	29.8
✓			✓		62.58	↑ 3.04	70.31	42.6
✓	✓	✓	✓		63.83	↑ 4.29	71.16	18.9
✓	✓	✓	✓	✓	64.49	↑ 4.95	71.35	12.8

从表 4 可以看出,所提的各模块在遥感有向目标检测任务上性能均有提升,其中基于特征重组的特征金字塔贡献最大,其精度较基准网络提升了 3.49%。由于引入极化注意力机制,输入分类和回归分支的特征得到区分,检测头可以更好地进行检测任务,精度也提升了 3.04%。ATT-ORPN 模块由于需要生成较多的参数,因此速度下降较快,但其检测精度仍提升了 2.82%。将上述模块组合起来并使用 SWA 训练策略精度可以提升 4.95%,但模型参数的增加以及多次训

练的策略,导致速度下降较快,这可能是二阶段检测模型难以避免的缺点。

图 8 是基准模型与本文模型对 Dior-R 数据集中各类样本的检测精度对比。从图 8 可见,PFR-Rotate 对各类样本的检测精度较基准模型均有提升,其中对机场、水坝、体育场、火车站等较难区分类别的检测精度提升明显,说明本文模型能够有效增强对于遥感图像有向目标特征的提取能力。

图 9 所示为本文模型与基准模型在 Dior-R 数据集

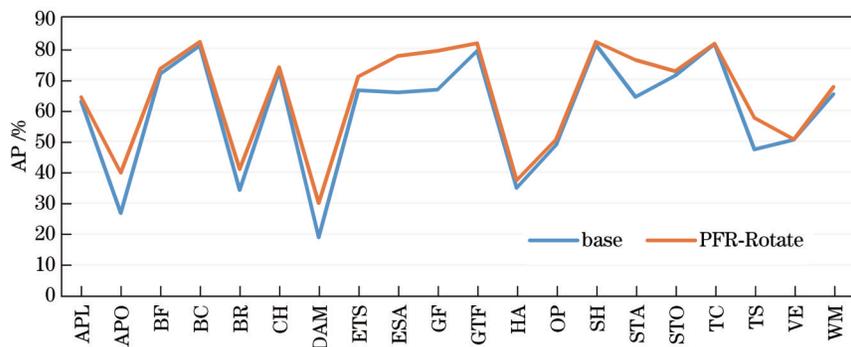


图 8 各实验对 Dior-R 数据集每类样本的检测精度

Fig. 8 Average precision (AP) for each category in Dior-R dataset for each experiment

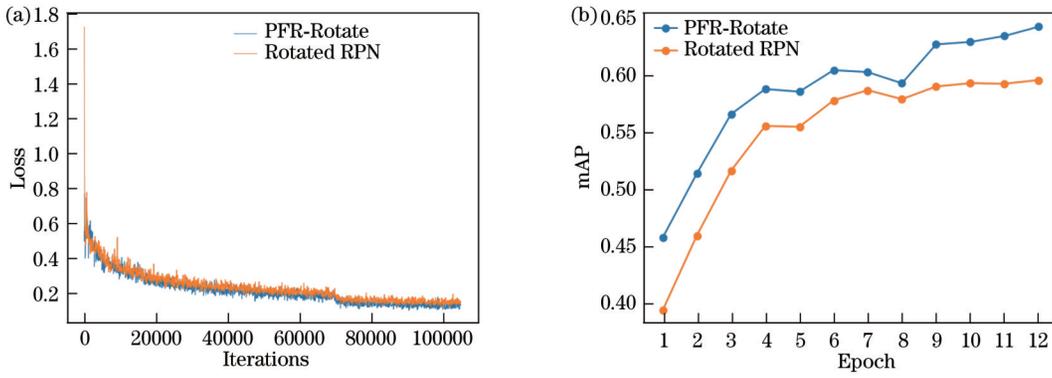


图 9 损失函数和精度对比。(a)损失函数;(b)mAP

Fig. 9 Loss function and mAP comparison. (a) Loss function; (b) mAP

上的损失函数和 mAP 曲线。从图 9 可以看出,在总损失经历短暂的振荡之后,本文模型和基准模型都能很快收敛,但本文模型的损失始终低于基准模型。从精度方面来看,各模型都有小幅振荡,但本文模型的精度始终优于基准模型。

### 3.4 模型对比实验

为了验证所提出的用于遥感有向目标检测模型 PFR-Rotate 的有效性,将本文模型与其他一些一阶段和二阶段代表性有向检测模型进行比较,结果如表 5 所示,其中 ms 表示多尺度训练,主干网络为

ResNet50。基准模型 Rotated RPN 在 Dior-R 数据集上的精度为 59.54%。Gliding Vertex 使用水平框近似来表示有向框,避免了边界预测不完整的问题,但并未完全解决水平标注法存在的问题,精度为 60.06%。ROI Transformer 通过定义新的有向框表示方法和新的标签处理过程将精度提升到 63.87%,但由于需要经过两次回归得到最终的结果,因此在速度上还是有些不足。本文模型一方面在提取特征时能够关注有效特征,避免大量背景对检测结果的干扰,另一方面为分类和回归分配不同特点的特征,提升网络的检测精度。

表 5 Dior-R 数据集上各模型检测精度对比

Table 5 Comparison of precision of different network models on Dior-R dataset

unit: %

Model	RetinaNet-O <sup>[22]</sup>	Rotated RPN <sup>[15]</sup>	Gliding Vertex <sup>[23]</sup>	ROI Transformer <sup>[11]</sup>	AOPG <sup>[19]</sup>	PFR-Rotate	PFR-Rotate (ms)
APL	61.49	62.79	<b>65.35</b>	63.34	62.39	64.25	64.54
APO	28.52	26.8	28.87	37.88	37.79	39.78	<b>41.2</b>
BF	73.57	71.72	74.96	71.78	71.62	73.39	<b>75.5</b>
BC	81.17	80.91	81.33	87.53	<b>87.63</b>	82.13	83.55
BR	23.98	34.2	33.88	40.68	40.9	40.98	<b>41.69</b>
CH	72.54	72.57	74.31	72.6	72.47	73.89	<b>75.54</b>
DAM	19.94	18.95	19.58	26.86	31.08	29.98	31.35
ETS	72.39	66.45	70.72	<b>78.71</b>	65.42	70.91	73.02
ESA	58.2	65.75	64.7	68.09	77.99	77.56	<b>78.82</b>
GF	69.25	66.63	72.3	68.96	73.2	79.18	<b>79.4</b>
GTF	79.54	79.24	78.68	82.74	81.94	81.66	83.55
HA	32.14	34.95	34.22	<b>47.71</b>	42.32	37.39	40.19
OP	44.87	48.79	74.64	<b>55.61</b>	54.45	50.62	53.56
SH	77.71	81.14	80.22	81.21	81.17	82.18	<b>83.38</b>
STA	67.57	64.34	69.26	<b>78.23</b>	72.69	76.24	78.21
STO	61.09	71.21	61.13	70.26	71.31	72.68	<b>74.56</b>
TC	81.46	81.44	81.49	81.61	81.49	81.53	<b>82.05</b>
TS	47.33	47.31	44.76	54.86	<b>60.04</b>	57.52	58.75
VE	38.01	50.46	47.71	43.27	<b>52.38</b>	50.57	51.64
WM	60.24	65.21	65.04	65.52	<b>69.99</b>	67.51	69.64
mAP	57.55	59.54	60.06	63.87	64.41	64.49	<b>65.97</b>

在配合新的有向框定义法以及训练策略的基础上精度达到 64.49%，使用多尺度训练可以达到 65.97%。可以看出，本文模型在多个类别取得最优的效果，说明了本文模型的有效性，不仅兼顾了不同尺寸、方向的目标，还适应了大部分场景，达到了在多尺度、多方向、复杂场景下的遥感图像有向目标检测性能优化的目的。

表 6 Dior-R 数据集定向和有向标注结果对比

Table 6 Comparison of directional and directed labeling results of Dior-R dataset

Directional labeling			Directed labeling		
Model	Backbone	mAP / %	Model	Backbone	mAP / %
SSD <sup>[24]</sup>	VGG16	58.6	Retinanet-O <sup>[22]</sup>	Resnet50	57.55
YOLO V3 <sup>[25]</sup>	Darknet53	57.1	Rotated RPN <sup>[15]</sup>	Resnet50	59.54
Cascade R-CNN <sup>[26]</sup>	Resnet50	60.5	Gliding Vertex <sup>[23]</sup>	Resnet50	60.06
SA-Cascade <sup>[27]</sup>	Resnet50	62.1	PRF-Rotate	Resnet50	64.49

为了验证本文模型在有向数据集上的通用性，将本文模型与其他模型在 HRSC2016 数据集上也进行对比实验，结果如表 7 所示，括号中的 07、12 沿用了 VOC 数据集的表述方式，代表两种不同的评价。从表 7 可以看出，本文模型在有向目标数据集上有较好的通用性。本文模型的检测精度相较于基准模型有较大的提升，主要原因在于基准模型对于小目标的检测能力不足，而 HRSC2016 数据集中只有船舰这一大类，该类别目标通常较小，本文模型可以较好地解决小目标的检测问题，所以效果有所提升。S<sup>2</sup>A-Net 提出特征对齐模块和

将几个代表性的定向检测模型和有向检测模型在 Dior-R 数据集上分别使用水平标注和有向标注进行实验，实验结果如表 6 所示。从表 6 可以看出，对同一个数据集使用水平标注格式和有向标注格式时检测精度相差较大，由于遥感图像中有向目标数量较多，水平框检测模型即使在使用更复杂结构的情况下效果依然不如定向目标检测框架。

面向检测模块，通过锚点细化网络生成高质量的锚点，并通过自适应对齐特征，又采用主动旋转滤波器对方向信息进行编码，提取旋转不变特征，从而有效解决分类性能与回归性能不一致的问题，在 HRSC2016 数据集上获得 90.17% 的检测精度。Oriented RepPoints 提出三种新的定向转换函数和样本分配策略，引入空间约束来惩罚自适应学习的异常之处，和 Oriented R-CNN 同样取得 90.4% 左右的检测精度。本文模型使用的特征重组方法能对密集目标做出有效检测，所以在 HRSC2016 数据集上取得最佳精度 90.83%。

表 7 HRSC2016 数据集上各模型结果对比

Table 7 Comparison of test results of different network models on HRSC2016 dataset

Model	Backbone	mAP(07) / %	mAP(12) / %
RetinaNet-O <sup>[22]</sup>	Resnet101	89.18	95.21
Rotated RPN <sup>[15]</sup>	Resnet50	79.08	85.64
Gliding Vertex <sup>[23]</sup>	Resnet101	88.20	—
ROI Transformer <sup>[11]</sup>	Resnet101	86.20	—
S <sup>2</sup> A-Net <sup>[25]</sup>	Resnet101	90.17	95.01
AOPG <sup>[19]</sup>	Resnet101	90.34	96.22
Oriented R-CNN <sup>[28]</sup>	Resnet50	90.40	96.50
Oriented RepPoints <sup>[29]</sup>	Resnet50	90.38	97.26
AFO-RPN <sup>[30]</sup>	Resnet101	90.45	—
PFR-Rotate	Resnet50	90.83	97.35

图 10 和图 11 所示分别为本文模型在 Dior-R 数据集和 HRSC2016 数据集上的可视化结果以及和基准模型对比的结果，可以看出，本文模型无论是检测精度还是漏检率和误检率都取得了较基准算法更好的表现。相比于基准模型，本文模型将图像中所有目标都检测出来了，并且没有误检框。上述两个对比说明了本文模型针对有向目标检测的有效性。

## 4 总 结

为了解决通用目标检测模型对遥感图像中有向目标检测能力不足的问题，在 Rotated RPN 的基础上进行改进，设计了一种基于特征重组和极化注意力的有向目标检测框架 PFR-Rotate。引入特征重组模块，将 FPN 深层特征进行重组加权，使模型能够更加关注有效特征而忽视无效特征对检测的干扰。使用新的有向



图 10 Dior-R数据集可视化结果对比

Fig. 10 Comparison of visualization results for Dior-R dataset

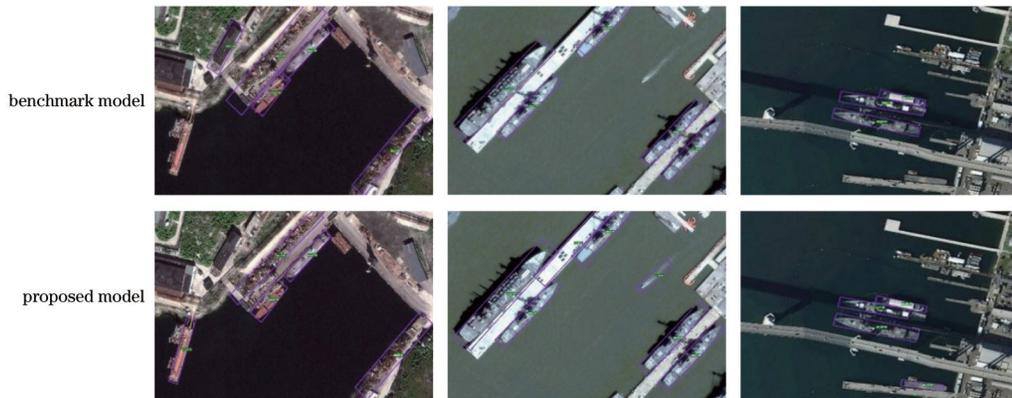


图 11 HRSC2016数据集可视化结果对比

Fig. 11 Comparison of visualization results for HRSC2016 dataset

框标注方法并且为RPN设计自适应注意力模块,增强特征表达。为检测头引入极化注意力模块,以减弱不同任务之间所需特征的差异性。从结果来看,本文模型在两个数据集中均取得优秀的表现,体现了模型的有效性。本文模型虽然提升了检测精度,但是参数量较大,不能很好地适应实时检测的需求。后续工作考虑在尽可能保证精度的前提下对模型进行轻量化处理。

参 考 文 献

[1] Yang F, Fan H, Chu P, et al. Clustered object detection in aerial images[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2020: 8310-8319.

[2] Cui Z Y, Li Q, Cao Z J, et al. Dense attention pyramid networks for multi-scale ship detection in SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8983-8997.

[3] 张寅, 朱桂熠, 施天俊, 等. 基于特征融合与注意力的遥感图

像小目标检测[J]. 光学学报, 2022, 42(24): 2415001.

Zhang Y, Zhu G Y, Shi T J, et al. Small object detection in remote sensing images based on feature fusion and attention[J]. Acta Optica Sinica, 2022, 42(24): 2415001.

[4] 薛俊达, 朱家佳, 张静, 等. 基于FFC-SSD模型的光学遥感图像目标检测[J]. 光学学报, 2022, 42(12): 1210002.

Xue J D, Zhu J J, Zhang J, et al. Object detection in optical remote sensing images based on FFC-SSD model[J]. Acta Optica Sinica, 2022, 42(12): 1210002.

[5] Suresh V, Janik P, Rezmer J, et al. Forecasting solar PV output using convolutional neural networks with a sliding window algorithm[J]. Energies, 2020, 13(3): 723.

[6] Xu C, Wang J W, Yang W, et al. RFLA: Gaussian receptive field based label assignment for tiny object detection[M]//Avidan S, Brostow G, Cissé M, et al. Computer vision - ECCV 2022. Lecture notes in computer science. Cham: Springer, 2022, 13669: 526-543.

[7] Nan Z X, Peng J Z, Jiang J J, et al. A joint object detection and semantic segmentation model with cross-attention and inner-attention mechanisms[J]. Neurocomputing, 2021, 463: 212-225.

[8] 徐志京, 柏雪. 基于双重特征增强的遥感舰船小目标检测[J]. 光学学报, 2022, 42(18): 1828002.

Xu Z J, Bai X. Small ship target detection method for remote

- sensing images based on dual feature enhancement[J]. Acta Optica Sinica, 2022, 42(18): 1828002.
- [9] Zhao Y, Zhao L J, Liu Z, et al. Attentional feature refinement and alignment network for aircraft detection in SAR imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 5220616.
- [10] Yang X, Yan J C, Liao W L, et al. SCRDet: detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2384-2399.
- [11] Ding J, Xue N, Long Y, et al. Learning RoI transformer for oriented object detection in aerial images[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 2844-2853.
- [12] Yang X, Yan J C, Ming Q, et al. Rethinking rotated object detection with Gaussian Wasserstein distance loss[EB/OL]. (2021-01-28)[2023-03-02]. <https://arxiv.org/abs/2101.11952>.
- [13] Zhang S, He G H, Chen H B, et al. Scale adaptive proposal network for object detection in remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(6): 864-868.
- [14] Li Q P, Mou L C, Jiang K Y, et al. Hierarchical region based convolution neural network for multiscale object detection in remote sensing images[C]//2018 IEEE International Geoscience and Remote Sensing Symposium, July 22-27, 2018, Valencia, Spain. New York: IEEE Press, 2018: 4355-4358.
- [15] Ma J Q, Shao W Y, Ye H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.
- [16] Xie X X, Cheng G, Wang J B, et al. Oriented R-CNN for object detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 3500-3509.
- [17] Ming Q, Miao L J, Zhou Z Q, et al. CFC-net: a critical feature capturing network for arbitrary-oriented object detection in remote-sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5605814.
- [18] Zhang H Y, Wang Y, Dayoub F, et al. SWA object detection [EB/OL]. (2020-12-23) [2023-03-02]. <https://arxiv.org/abs/2012.12645>.
- [19] Cheng G, Wang J B, Li K, et al. Anchor-free oriented proposal generator for object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5625411.
- [20] Zhou Y, Yang X, Zhang G F, et al. MMRotate: a rotated object detection benchmark using PyTorch[C]//Proceedings of the 30th ACM International Conference on Multimedia, October 10-14, 2022, Lisboa, Portugal. New York: ACM, 2022: 7331-7334.
- [21] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [22] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.
- [23] Xu Y C, 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, 2021, 43(4): 1452-1459.
- [24] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision - ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [25] Redmon J, Farhadi A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08) [2023-03-02]. <https://arxiv.org/abs/1804.02767>.
- [26] Cai Z W, Vasconcelos N. Cascade R-CNN: delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6154-6162.
- [27] 王友伟, 郭颖, 邵香迎. 基于改进级联算法的遥感图像目标检测[J]. 光学学报, 2022, 42(24): 2428004.  
Wang Y W, Guo Y, Shao X Y. Target detection in remote sensing images based on improved cascade algorithm[J]. Acta Optica Sinica, 2022, 42(24): 2428004.
- [28] Han J M, Ding J, Li J, et al. Align deep features for oriented object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5602511.
- [29] Li W T, Chen Y J, Hu K X, et al. Oriented RepPoints for aerial object detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 1819-1828.
- [30] Li J X, Tian Y, Xu Y P, et al. Oriented object detection in remote sensing images with anchor-free oriented region proposal network[J]. Remote Sensing, 2022, 14(5): 1246.

## Oriented Object Detection in Remote Sensing Images Based on Feature Recombination

Wang Youwei<sup>1,2</sup>, Guo Ying<sup>1,2\*</sup>, Shao Xiangying<sup>1,2</sup>, Wang Jiyu<sup>1,2</sup>, Bao Zhengwei<sup>1,2</sup>

<sup>1</sup>Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing 210044, Jiangsu, China;

<sup>2</sup>School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, Jiangsu, China

### Abstract

**Objective** Object detection of optical remote sensing images is the process of providing a given optical remote sensing image dataset with object positioning frame, object category, and confidence by model processing, and it is an important task in remote sensing image processing and has practical significance in both civil and military fields. In the civil field, it can be employed to analyze the situations of airport flights and ships in ports and thus facilitate timely adjustment and avoid

congestion. In the military field, enemies' military deployment is analyzed by the photographed images, and feasible plans are made to ensure successful military operations. Therefore, object detection of remote sensing images has research significance and application prospect. Compared with the traditional detection algorithms, the detection method based on the convolutional neural network has become the mainstream object detection of remote sensing images. The method based on deep learning can yield better accuracy than the traditional object detection methods of visible light remote sensing images, and it is unnecessary to manually design rules, which has a relatively unified standard and enhances the model robustness. However, there are still many defects in introducing the object detection model dealing with natural images directly into remote sensing tasks. Starting from the oriented object detection difficulties of remote sensing, we design an oriented object detection algorithm for optical remote sensing images to improve the feature extraction and feature recognition ability of multi-scale and multi-directional remote sensing small targets in complex backgrounds.

**Methods** Aiming at the poor performance of general algorithms for remote sensing oriented object detection, we propose an oriented object detection model based on SWA training strategy and feature recombination. The model is optimized based on the Rotated RPN algorithm. On the one hand, the feature recombination mechanism is introduced to make the model focus on effective features, which can reduce unnecessary computing resources and improve the model accuracy. On the other hand, based on RPN, the rotating RPN is introduced, and the position and angle parameters are regressed by the midpoint offset method to generate high-quality directed candidate frames. For the required feature inconsistency between classification and regression tasks, a polarized attention detector is employed, and the training strategy is improved. Meanwhile, the model is trained by cyclic mode to alleviate the problem that the traditional training strategy will converge to the boundary region of the optimal solution.

Specifically, we conduct the following improvements based on Rotated RPN. 1) Given the problems in the object detection tasks of remote sensing images, such as a large number of small targets, a large proportion of background, and a large change in target size, the feature pyramid can not extract effective information during extracting and fusing features, which degrades detection performance. Therefore, we consider making changes in the feature pyramid to strengthen the feature extraction ability of the feature pyramid and the ability to fully fuse information of various sizes. Additionally, the reshape module is designed and integrated into the Carafe model as a deep horizontal connection of FPN. 2) To solve the problems of angle discontinuity and edge order exchange in the critical angle of the common directed box representation, we introduce the midpoint offset method to define the directed box. An adaptive attention module is designed in front of the suggested area generation module to enhance the ability of effective feature representation and further strengthen the ability of feature extraction and characterization. 3) The features required for the classification task should have the same response to different angles, which is because the focus of the classification task should be on the target itself. Thus, it should be highly responsive to the effective information inside the prediction frame, while the features required for the regression task should be sensitive to the angle change. Meanwhile, more attention should be paid to the boundary area of the target and less attention is to the information inside the prediction frame for realizing accurate angle and position prediction and reducing interference. Therefore, to avoid feature interference between different tasks and extract key features, we introduce a polarization attention module to the shared convolution layer at the front end of the dual-branch detector and adopt different response functions to distinguish the representation ability of different features. The classification head and regression head employ an activation function and an inhibition function respectively. 4) In view of the limitation that the traditional training strategy may converge to the boundary region of the optimal solution, we introduce the SWA cyclic training strategy, obtain the corresponding weights by adopting the SGD method to train more epochs, and average these results to acquire results closest to the optimal solution.

**Results and Discussions** To verify the algorithm performance, we select two remote sensing oriented annotation datasets Dior-R and HRSC2016 to compare the algorithm performance. Several typical one-stage and two-stage oriented object detection models are selected and compared with this model. On the Dior-R dataset, our algorithm yields the best accuracy of 64.49%, 4.95% higher than that of the benchmark model (Table 5). On the HRSC2016 dataset, the proposed algorithm achieves the best accuracy of 90.83%, which is 11.75% higher than that of the benchmark model (Table 7). Additionally, we analyze the performance improvement after introducing the feature recombination module, focus shift method, adaptive attention module, polarized attention detector, and SWA training strategy respectively. The experimental results show that the algorithm has sound detection performance for remote sensing oriented objects in complex backgrounds.

**Conclusions** To improve the detection performance of oriented objects in remote sensing images, we propose an oriented object detection model based on feature recombination and polarized attention. The experimental results show that the algorithm can effectively detect oriented objects in remote sensing images, and has good performance in all kinds of scenes.

**Key words** remote sensing; oriented object detection; deep learning; feature recombination; polarized attention