特征重组和自注意力的遥感图像有向目标检测

闵令通[1]，范子满[1]，谢星星[2]，吕勤毅[1]

1. 西北工业大学 电子信息学院, 西安 710072;
2. 西北工业大学 自动化学院, 西安 7100072

摘 要：遥感图像有向目标检测是一项非常有挑战性的任务，受到了广泛的关注。随着深度学习的迅速发展，基于卷积神经网络（CNN）和自注意力网络（Transformer）的神经网络在有向目标检测方面取得了显著成果。然而，对于遥感图像中的有向目标，仍然存在对边界信息和显著特征信息的关注不足的问题。其中，不同方向目标的边界信息有限且难以提取，而显著特征的全局依赖关系相对稀疏。因此，本文提出了基于特征重组和自注意力的遥感图像有向目标检测方法。该方法主要包括空间通道重组的回归分支和自注意力分类分支。其中，回归分支通过在通道维度中重组空间信息，更加关注边界敏感信息，以实现对定位框的精确定位。分类分支依据带有位置信息的自注意力捕获目标根本判别性的特征，并增强特征的全局依赖性，从而实现准确分类。通过广泛的实验验证，证明了所提出模型的有效性和鲁棒性。在公开数据集DOTA、HRSC2016和SODA-A上表现优秀。

关键词：遥感图像，有向目标检测，检测头，特征重组，自注意力

中图分类号：TP701/P2

引用格式：**闵令通，范子满，谢星星，吕勤毅.2023.特征重组和自注意力的遥感图像有向目标检测.遥感学报，27(12)：2716-2725**

Min L T，Fan Z M，Xie X X and Lyu Q Y. 2023. Feature reassembly and self-attention for oriented object detection in remote sensing images. National Remote Sensing Bulletin，27(12)：2716-2725[DOI：10.11834/jrs.20233456]

1 引 言

现实世界中，随着卫星与成像技术的快速发展，每天都会产生大量高分辨率遥感图像。这些高分辨率图像往往包含大量信息。在交通管理、测绘等现实应用中对快速准确的从海量信息中检测出感兴趣目标的类别与位置提出了新的要求（李红光 等，2021）。

深度学习的发展为目标检测提供了新的思路，一系列基于卷积神经网络（CNN）的检测器不断提升了自然图像检测性能，即检测精度与推理速度（Xu等，2021）。与自然图像相比，高分辨率遥感图像目标往往以任意方向和视角排列并且像素有限、信息较少，准确快速的检测识别出其中的感兴趣目标是十分具有挑战性的工作（聂光涛和黄华，2021）。为了提升检测器对任意方向目标的检测性能，Xie等（2021）提出了有向区域提议网络（O-RCNN）模型。O-RCNN从水平框到有向框的转变可以的提升有向目标的检测性能，降低误检漏检。但是除了以任意方向排列，有向目标往往存在像素小、信息有限等特征不足的问题，旋转不变的有向框并没有对回归分支和分类分支特征信息给予更多关注，即O-RCNN检测头网络中回归分支对特征向量空间维度进行简单的展平处理难免会丢失部分边界信息，不利于目标特征信息的建模。而在分类分支中，简单的展平处理可能导致显著性特征全局依赖稀疏，不利于类别检测。

注意力机制是提取和捕获像素有限条件下有向目标特征信息的有效方法，尤其是自注意力（Transformer）在全局信息建模方面表现优秀。例如，Min等（2023）基于Transformer构建的检测网络提升了遥感目标信息的捕获能力，在检测精度的同时保证推理速度。Cheng等（2023a）基于

收稿日期：2023-11-01；**预印本**：2023-11-29

基金项目：国家自然科学基金(编号：62206221)

第一作者简介：闵令通，研究方向为人工智能、智能感知和模式识别。E-mail：minlingtong@nwpu.edu.cn

通信作者简介：吕勤毅，研究方向为人工智能、雷达信息处理和智能感知。E-mail：lvqinyi@nwpu.edu.cn

Transformer 提出了 SFPNet，在检测头部分加入 Transform 结构，提升了目标特征信息捕获能力。Carion 等（2020）使用 Transformer 构建了端到端对象检测（DETR），提升了检测器的性能。Zhang 等（2023）针对基于 Transformer 的网络做出了改进以提升检测性能。

为此，本文重新设计了 O-RCNN 的检测头部网络。该网络包括两部分：空间和通道信息重组回归分支 RBIR（Regression Branch of spatial channel Information Recombination）和自注意力分类分支 CLT（Classification Transformer），能够增强对有限条件下有向目标特征信息的捕获能力。其中，RBIR 模块将空间信息与通道信息重组建模来解决定位过程中边界区域特征提取中信息湮灭和丢失的问题。CLT 模块从全局信息入手来捕获不同目标的特征信息，提升有向目标类别的检测能力。最后，本文基于 O-RCNN 网络提出了特征重组和自注意力检测网络 FRTDNet（Feature Recombination and Transformer Detection Networks），并在公开数据集 DOTA（Xia 等，2018）、HRSC2016（Liu 等，2016）和 SODA-A（Cheng 等，2023b）中进行了实验，验证了 FRTDNet 的有效性。

2 研究方法或原理

FRTDNet 模型的总体框架如图 1 所示，基于 O-RCNN 框架，输入图像依次通过主干网络（Backbone）、特征金字塔（FPN）、区域建议网络（RPN）和感兴趣区域对齐 RoI Align（He 等，2017）一系列操作之后得到特征矩阵 $\boldsymbol{F} \in \mathbb{R}^{C \times H \times W}$。接着，检测头网络的输入 $\boldsymbol{F} \in \mathbb{R}^{C \times H \times W}$ 分别输入到空间和通道信息重组回归分支 RBIR 模块和自注意力分类分支 CLT 模块得到遥感有向目标定位框信息和类别信息。训练损失计算与 O-RCNN 保持一致，其中回归损失采用 Smooth L1，类别损失采用交叉熵损失（Cross Entropy Loss）。接下来分别对 RBIR 模块和 CLT 模块进行介绍。

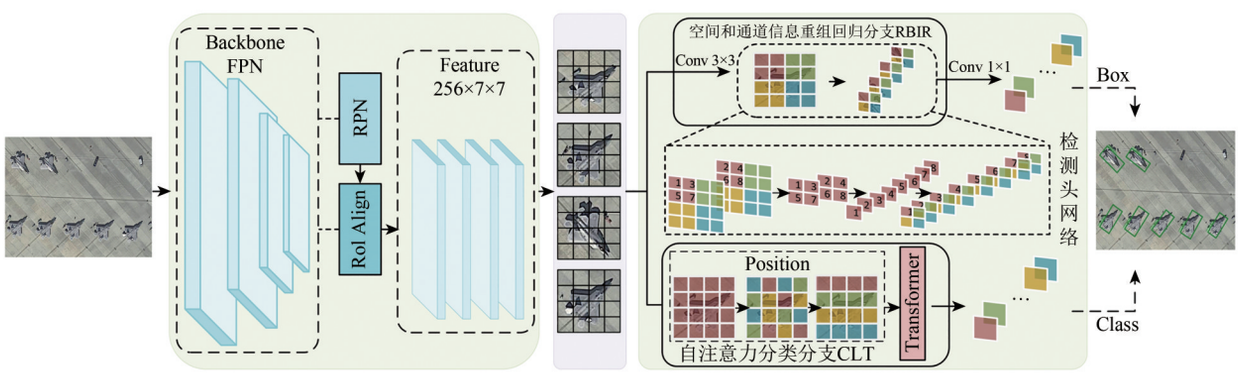


图 1 FRTDNet 的总体检测框架

Fig. 1 The overview detection framework of FRTDNet

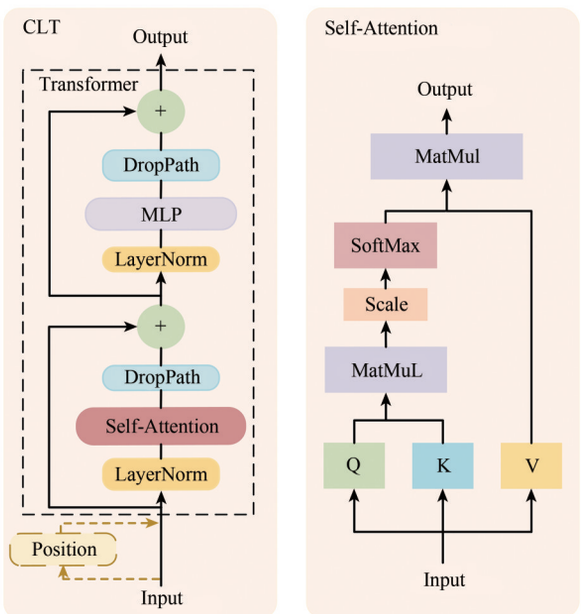
2.1 回归分支 RBIR

由于回归检测分支对特征的空间维度进行展平操作（对 H，W 所在维度进行展平处理）会忽视目标的空间信息，降低有向边界框的定位准确性。所以在回归检测分支采用空间向通道的转换模块是十分必要的操作。受到下采样模块（Jiang 等，2022）的启发，本文设计了空间和通道信息重组回归分支模块。首先，如图 1 中所示，输入特征 $\boldsymbol{F} \in \mathbb{R}^{C \times H \times W}$ 通过 3×3 的卷积核得到过程特征矩阵 $\boldsymbol{T} \in \mathbb{R}^{C \times H_1 \times W_1}$，$H_1 = \frac{1}{2}H$，$W_1 = \frac{1}{2}W$，可用如下公式表示：

$$t_{ij} = \sum_{m=1}^{3}\sum_{n=1}^{3} \omega_{mn} \cdot f_{(i-2+m)(j-2+n)} \tag{1}$$

式中，t_{ij} 表示过程矩阵 $\boldsymbol{T}$ 中的元素，ω_{mn} 表示卷积核，$f_{(i-2+m)(j-2+n)}$ 是输入特征 $\boldsymbol{F}$ 中对应的元素。为了方便表示，公式中省略了批次标准化（BN）和线性激活函数（ReLU）。之后，以图 1 中红色元素为例，将中间特征相邻的元素提取出来，将空间排列的方式转换为通道排列，依次将其他颜色的元素进行转换，得到特征 $\boldsymbol{P} \in \mathbb{R}^{(C \times s^2) \times (H_1/s) \times (W_1/s)}$，其中 s 为缩放因子，一般情况下取 2。空间向通道的转换增加了每个通道的上下文信息，扩大了感受野，对回归任务中关注目标边界信息有利。通过将临近像素的信息在通道维度上编码，定位分支

可以更好的理解和利用空间上下文之间的相关性。同时增强了定位分支的定位能力，在通道上编码空间信息，可以使得定位检测头捕捉到来自RoI Align数据中的结构和模式，提升对遥感有向目标图像构的建模能力。



（a）转换器类别模块　　（b）自我注意力

（a）CLT　　（b）Self-Attention

图2　转换器类别模块和自我注意力

Fig. 2　CLT and Self-Attention

最后为了改善空间向通道变化过程中的不稳定性，会通过1×1的卷积核得到输出$\boldsymbol{Y} \in \mathbb{R}^{(C \times s^2)}$。在此之后，一般会采用全连接层（FC）将输出$\boldsymbol{Y}$的维度转换为回归边界框的位置坐标，实现最后的定位效果。

2.2　分类分支CLT

分类任务更加关注图像的显著特征区域（Song等，2020），而对分类分支的简单空间维度展平忽视了类别显著性特征。因此在分类分支选择用带有位置信息（Position）的Transformer构建自注意力分类分支。首先，如图2中所示，输入特征矩阵$\boldsymbol{F}$通过较为常用的正弦编码方法（Vaswani等，2017）进行位置信息编码得到特征矩阵$\boldsymbol{M} \in \mathbb{R}^{C \times (H \times W)}$，之后特征矩阵$\boldsymbol{M}$会通过Transformer的第一个残差模块，即层归一化（LayerNorm）、自我注意力（Self-Attention）、DropPath得到中间特征$\widehat{\boldsymbol{M}} \in \mathbb{R}^{C \times (H \times W)}$。其中，Self-Attention操作会将第一个残差模块的输入特征矩阵$\boldsymbol{M}$会沿着通道维度分为N部分，每个部分都会经过嵌入矩阵$\left(\boldsymbol{W}_{\mathrm{q}}, \boldsymbol{W}_{\mathrm{k}}, \boldsymbol{W}_{\mathrm{v}}\right)$，如图2所示，分别将$\boldsymbol{M}$变换为查询$\boldsymbol{Q} = \boldsymbol{M}\boldsymbol{W}_{\mathrm{q}}$、键$\boldsymbol{K} = \boldsymbol{M}\boldsymbol{W}_{\mathrm{k}}$、值$\boldsymbol{V} = \boldsymbol{M}\boldsymbol{W}_{\mathrm{v}}$。需要注意的是，每个嵌入矩阵是指全连接的线性变换。之后通过一定缩放的点积运算得到特征矩阵可用式（2）表示为

$$\widehat{\boldsymbol{M}}_i = Soft\max\left(\frac{\boldsymbol{Q}_i \boldsymbol{K}_i^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V}_i \tag{2}$$

式中，$\widehat{\boldsymbol{M}}_i$表示由第$i$个部分导出的自相关特征矩阵，$d$为缩放参数。为了方便表示，式（2）中省略了LayerNorm和DropPath。经过Self-Attention操作后的特征向量$\widehat{\boldsymbol{M}}$和第一个残差模块的输入特征$\boldsymbol{M}$经过残差结构构成了第一个残差模块的输出特征向量$\boldsymbol{T}$。第二个残差模块由LayerNorm、多层感知机MLP（Multilayer Perceptron）和DropPath构成，MLP包含两个线性层和一个非线性激活函数，可以用式（3）表示：

$$\boldsymbol{T}_i' = f_l\left(\left(\boldsymbol{T}_i \boldsymbol{w}_1 + \boldsymbol{b}_1\right) \times \boldsymbol{w}_2 + \boldsymbol{b}_2\right) \tag{3}$$

式中，$\boldsymbol{T}_i'$表示第i个输入元素经由MLP模块的输出值，$\boldsymbol{w}_1$，$\boldsymbol{b}_1$，$\boldsymbol{w}_2$，$\boldsymbol{b}_2$分别表示表示第一个和第二个线性变换层的权重矩阵和偏置向量，$f_l()$表示非线性激活函数。经过MLP操作后的特征向量$\boldsymbol{T}'$和MLP输入的输入特征向量$\boldsymbol{T}$经过残差模块得到Transformer的输出$\widehat{\boldsymbol{T}} \in \mathbb{R}^{C \times (H \times W)}$，可用如下式（4）表示：

$$\widehat{\boldsymbol{T}} = \boldsymbol{T} + \boldsymbol{T}' \tag{4}$$

Self-Attention和MLP中的LayerNorm用于对每个Transformer的通道层进行归一化，减少每个特征之间的差异，使得每个层的输入分布更加稳定，提高Transformer的稳定性和收敛性，有助于在训练过程中更好的学习和传播信息。DropPath是一种正则化技术，可以在Transformer中的每个层中丢弃一部分连接，通过一定概率丢弃部分连接，可以减少Transformer的过拟合风险，增强鲁棒性。这两种策略可以在一定程度上提高Transformer的稳定性、收敛性和鲁棒性。

自注意力分类分支CLT聚焦于有向目标的类别显著性信息，从全局特征进行建模，有助于有向目标类别信息的捕获，提升了检测性能。

3 实验结果

为了验证FRTDNet算法的有效性，本文在公开数据集DOTA（Xia等，2018）、HRSC2016（Liu等，2016）和SODA-A（Cheng等，2023b）上进行了训练和测试。本文使用PyTorch 1.6.0、Python 3.8、Cuda 10.1构建实验环境，采用一块显存为32 GB的Tesla-V100。DOTA数据集的评价指标采用官方的AP50。SODA-A采用COCO（Lin等，2014）的评价指标，分别为AP、AP50、AP75、APeS、APrS、APgS、APN和AP0.5：0.95，其中AP和AP0.5：0.95一致。HRSC2016采用VOC（Everingham等，2010）作为评价指标，分别为mAP（07）和mAP（12）。3种数据集的目标一部分是本身有一定的方向性（飞机，轮船等），一部分是本身并不具有方向性（储罐等）。但由于遥感图像的拍摄角度并不固定，这两种目标都按照一定方向排列，因此认为这两种都是有向目标。

3.1 DOTA数据集不同算法的比较

为了验证本文算法的检测结果，并与其他模型进行对比，本文在表1中展示了骨干（Backbone）、桥梁（BR）、港口（HA）、船舶（SH）、飞机（PL）、直升机（HC）、小型车辆（SV）、大型车辆（LV）、棒球场（BD）、田径场（GTF）、网球场（TC）、篮球场（BC）、足球场（SBF）、环形交叉路口（RA）、游泳池（SP）、储罐（ST）以及AP50的对比情况。值得一提的是，黑体表示相应类别的最高值。从表中可以直观地观察到，本文的算法在大型车辆、船舶、网球场、港口和AP50方面取得了最优的效果。这些结果证明了特征重组检测头在分类和回归特征上对于有向目标具有更为细致的补充，从而减少了特征的丢失和湮灭。值得注意的是，在使用相同的骨干网络的情况下，本文的算法仍然表现出色，进一步证明了本文的空间通道特征重组和转换器检测头在大尺度遥感有向目标检测中具有出色的性能。

表1 DOTA上不同模型对比结果，加粗数字代表相应类别最好性能

Table 1 Different models on DOTAcompare the results，bold numbers represent the best performance of the corresponding category

方法	Backone	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	AP50
DRN(Pan等,2020)	H-104	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(Yang等,2021)	R-101-FPN	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
PIOU(Chen等,2020)	DLA-34	80.90	69.70	24.10	60.20	38.30	64.40	64.80	**90.90**	77.20	70.40	46.50	37.10	57.1	61.90	64.00	60.50
RSDet(Qian等,2021)	R-101-FPN	89.80	82.90	48.60	65.20	69.50	70.10	70.20	90.50	85.60	83.40	62.50	63.90	65.60	67.20	**68.00**	72.20
DAL(Ming等,2021)	R-50-FPN	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
S2Anet(Han等,2022)	R-50-FPN	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
ICN(Azimi等,2018)	R-101-FPN	81.36	74.30	47.70	70.32	64.89	67.82	69.98	90.76	79.06	78.20	53.64	62.90	67.02	64.17	50.23	68.16
CAD-Net(Zhang等,2019)	R-101-FPN	87.80	82.40	49.40	73.50	71.10	63.50	76.60	**90.90**	79.20	73.30	48.40	60.90	62.00	67.00	62.20	69.90
RoI Transformer(Ding等,2019)	R-101-FPN	88.64	78.52	43.44	75.92	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
SCRDet(Yang等,2019)	R-101-FPN	80.65	80.65	52.09	68.36	68.36	60.32	72.41	90.85	**87.94**	**86.86**	65.02	66.68	66.25	68.24	65.21	72.61
Gliding Vertex(Xu等,2021)	R-101-FPN	89.64	85.00	52.26	**77.34**	73.01	73.14	86.82	90.74	79.02	86.81	59.55	**70.91**	72.94	70.86	57.32	75.02
FAOD(Li等,2019)	R-101-FPN	**90.21**	79.58	45.49	76.41	73.18	68.27	79.56	90.83	83.40	84.68	53.4	65.42	74.17	69.69	64.86	73.28
CenterMap-Net(Wang等,2021)	R-50-FPN	88.88	81.24	53.15	60.65	78.62	66.55	78.10	88.83	77.80	83.61	49.36	66.19	72.10	72.36	58.7	71.74
FR-Est(Fu等,2021)	R-101-FPN	89.63	81.17	50.44	70.19	73.52	77.98	86.44	90.82	84.13	83.56	60.64	66.59	70.59	66.72	60.55	74.20
Mask OBB(Wang等,2019)	R-50-FPN	89.61	**85.09**	51.85	72.90	75.28	73.23	85.57	90.37	82.08	85.05	55.73	68.39	71.61	69.87	66.33	74.86
Oriented R-CNN(Xie等,2021)	R-50-FPN	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
Oriented R-CNN(Xie等,2021)	R-101-FPN	88.86	83.48	55.27	76.92	74.27	82.10	87.52	**90.90**	85.56	85.33	**65.51**	66.82	74.36	70.15	57.28	76.28
OrientedRepPoints(Li等,2022)	R-101-FPN	89.53	84.07	**59.86**	71.76	**79.95**	80.03	87.33	90.84	87.54	85.23	59.15	66.37	75.23	**73.75**	57.23	76.52
本文方法	R-101-FPN	89.45	83.88	55.9	73.66	78.87	**83.29**	**88.20**	**90.90**	86.72	86.06	62.90	70.33	**76.68**	71.89	53.69	**76.83**

注：加粗表示相应类别的最高值。

3.2 SODA-A数据集不同算法的比较

为了验证本文算法的检测结果，并与其他模型进行对比，本文在表2中展示了骨干（Backbone）、飞机（APL）、直升机（HC）、小型车辆（SV）、大型车辆（LV）、船舶（SH）、集装箱（CT）、储罐（ST）、游泳池（SP）和风车（WM）以及AP0.5:0.95的对比情况。值得一提的是，黑色加粗数字表示相应类别的最高值。

表2 SODA-A上不同模型具体类别对比结果，加粗数字代表相应类别最好性能

Table 2 Comparison results of specific categories of different models on SODA-A, the bold numbers represent the best performance of the corresponding category

方法	Bacbone	APL	HC	SV	LV	SH	CT	ST	SP	WM	AP0.5:0.95
Rotated Faster RCNN(Pan等,2020)	ResNet-50	49.4	18.1	33.4	19.6	43.5	29.8	42.8	34.1	21.9	32.5
Rotated RetnaNet(Lin等,2017)	ResNet-50	42.0	16.8	29.9	10.0	35.1	23.7	35.1	30.7	18.1	26.8
Gliding Vertex(Xu等,2021)	ResNet-50	46.7	12.8	33.3	21.9	43.4	29.8	**43.3**	31.2	**22.7**	31.7
Oriented-RCNN(Xie等,2021)	ResNet-50	**52.2**	20.2	34.4	**24.4**	45.2	32.1	43.1	36.3	22.2	34.4
S2A-Net(Han等,2022)	ResNet-50	41.5	20.4	31.2	14.0	36.7	26.1	29.6	33.8	21.6	28.3
DODet(Cheng等,2022)	ResNet-50	49.4	19.8	32.1	17.3	41.3	26.0	42.2	34.7	21.3	31.6
Oriented RepPoints(Li等,2022)	ResNet-50	51.7	8.5	30.3	2.6	28.0	19.6	40.3	33.2	21.9	26.3
DHRec(Nie和Huang,2023)	ResNet-50	45.5	17.2	31.0	15.6	38.5	28.5	38.8	34.5	20.9	30.1
本文方法	ResNet-50	50.6	**22.4**	**37.5**	24.1	**46.4**	**33.8**	42.6	**37.3**	22.0	**35.2**

注：加粗表示相应类别的最高值。

从表2中可以直观地观察到，本文的算法在直升机、小型车辆、船舶、集装箱、游泳池和AP0.5:0.95方面取得了最优的效果，其中AP0.5:0.95指标表明更为关注有向小目标。在相同的骨干网络下，Oriented-RCNN因为其有向旋转框的优势，成功解决了小目标在密集排列的情况下，水平框与水平框之间相互重叠的问题。但是值得注意的是，在对小目标特征信息提取方面仍然有提升空间。本文的FRTDNet模型在分类分支依赖Transformer捕获全局依赖，在回归分支创新的用空间通道特征重组代替原来的展平空间向量的操作，这使得对于有向小目标特征提取更为充分，同时也关注到了全局依赖。

为了进一步说明对有向小目标的关注度，在表3中给出mAP0.5:0.95的4个子集作为评价标准，分别为AP_{eS}，AP_{rS}，AP_{gS}，AP_{N}。从表3中可以看出，对小目标的检测性能有所提升，进一步验证了本文模型的有效性。

表3 SODA-A上不同模小目标评价指标，加粗数字代表相应类别最好性能

Table 3 The small object evaluation indicators of different models on SODA-A, the bold numbers represent the best performance of the corresponding category

方法	Bacbone	AP	AP50	AP75	AP_{eS}	AP_{rS}	AP_{gS}	AP_{N}
Rotated Faster RCNN(Pan等,2020)	ResNet-50	32.5	70.1	24.3	11.9	27.3	42.2	34.4
Rotated RetnaNet(Lin等,2017)	ResNet-50	26.8	63.4	16.2	9.1	22.0	35.4	28.2
Gliding Vertex(Xu等,2021)	ResNet-50	31.7	70.8	22.6	11.7	27.0	41.1	33.8
Oriented-RCNN(Xie等,2021)	ResNet-50	34.4	70.7	**28.6**	12.5	28.6	44.5	36.7
S2A-Net(Han等,2022)	ResNet-50	28.3	69.6	13.1	10.2	22.8	35.8	29.5
DODet(Cheng等,2022)	ResNet-50	31.6	68.1	23.4	11.3	26.3	41.0	33.5
Oriented RepPoints(Li等,2022)	ResNet-50	26.3	58.8	19.0	9.4	22.6	32.4	28.5
DHRec(Nie和Huang,2023)	ResNet-50	30.1	68.8	19.8	10.6	24.6	40.3	34.6
本文方法	ResNet-50	**35.2**	**73.2**	28.3	**13.6**	**29.9**	**45.5**	**37.4**

注：加粗表示相应类别的最高值。

3.3 HRSC2016数据集不同算法的比较

为了进一步验证本文的算法，在有向船舶数据集HRSC2016上进行了算法验证。如表4所示，在相同骨干下，本文模型相较于基准模型有了一定的提升。在VOC 2007和VOC 2012两种评价指标下，对于有向船舶的特征提取略优于O-RCNN，证明了FRTDNet模型的有效性。

表4 在HRSC2016数据集上对比结果

Table 4 Comparing the results on the HRSC2016 dataset

方法	Bacbone	mAP(07)	mAP(12)
PIoU(Chen等,2020)	DLA-34	89.20	—
DRN(Pan等,2020)	H-34	—	92.70
R3Det(Yang等,2021)	R-101-FPN	89.26	96.01
DAL(Ming等,2021)	R-101-FPN	89.77	—
S2Anet(Han等,2022)	R-101-FPN	90.17	95.01
Rotated RPN(Ma等,2018)	R-101	79.08	85.64
R2CNN(Jiang等,2017)	R-101	73.07	79.73
RoI Transformer (Ding等,2019)	R-101-FPN	86.20	—
Gilding Vetex (Xu等,2021)	R-101-FPN	88.20	—
CenterMap-Net (Wang等,2021)	R-50-FPN	—	92.80
Oriented R-CNN (Xie等,2021)	R-50-FPN	90.40	96.50
Oriented R-CNN (Xie等,2021)	R-101-FPN	90.50	97.60
本文方法	R-50-FPN	**90.56**	96.66
本文方法	R-101-FPN	**90.56**	**97.68**

注：加粗表示相应类别的最高值。

3.4 消融试验

为了进行消融实验并验证本文的模块的性能，单独引入了分类分支（CLT）和回归分支（RBIR）。我们选用了R-101-FPN作为骨干网络，并通过表5展示了在DOTA数据集上的实验结果，证明了CLT和RBIR模块的优秀性能。

基准模型在引入CLT后，AP50增长了0.12%。这表明本文的分类分支在全局信息建模和提取显著特征方面发挥了重要作用，有效提高了模型的鲁棒性和稳定性。另外，我们单独引入回归定位分支RBIR以提取有向目标边界信息并准确定位回归框。实验结果显示，RBIR分支使AP50提升了0.07%，证明了RBIR模块在遥感有向目标的边界定位回归方面具有显著效果。最后，本文将回归分支和定位分支都加入基准模型，AP50提升了0.55%，这证明了FRTDNet模型的有效性。

表5 消融试验

Table 5 Ablation experiment

模型	Bacbone	分类分支（CLT）	回归分支（RBIR）	AP50
Oriented R-CNN	R-101-FPN			76.28
FRTDNet(Ours)	R-101-FPN	√		76.40
	R-101-FPN		√	76.35
	R-101-FPN	√	√	76.83

此外，为了进一步验证对有向小目标的有效性，在表6展示了分别加入CLT和RBIR模块在AP50、AP75和mAP这3个不同评价指标上的性能对比情况。我们发现，在单独引入CLT模块后，尽管AP50有所增长，但在AP75和mAP上的表现并不出色。我们认为Transformer结构在常规大小目标上的改进效果明显，但在处理小目标时表现可能不如人意。然后，本文单独引入RBIR模块，可以清晰地看到在这3个指标上都有明显的增长，尤其是针对小目标的指标mAP和AP75分别增长了0.61%和1.1%。这比Oriented-R-CNN模型的表现更出色，也证明了RBIR模块在小目标和常规目标上都有出色的表现。最后，FRTDNet模型在AP50、AP75和mAP这3个评价指标上分别带来了0.55%、0.08%和0.06%的增长，证明我们的模型在常规目标和小目标上都表现出色。

表6 在AP50、AP75、mAP指标下的对比

Table 6 Comparison in terms of AP50，AP75，and mAP metrics

模型	Bacbone	AP50	AP75	mAP
Oriented R-CNN	R-101-FPN	76.28	51.73	47.75
+CLT	R-101-FPN	76.40	50.00	47.04
+RBIR	R-101-FPN	76.35	52.83	48.36
Ours	R-101-FPN	76.83	51.81	47.81

3.5 实验结果展示

本文在Dota数据集中选择了一些有向目标进行检测，如图3所示。从图中可以观察到遥感图像具有多变复杂的背景和过多的干扰信息，同时目标本身所包含的信息也十分有限。然而，FRTDNet模型成功地应对了这些问题，展现了较为明显的

检测效果。值得注意的是，尽管储罐（ST）等目标自身并不具有方向性，但是由于拍摄角度多样，储罐（ST）仍然是以不同方向排列的，因此本文认为这些本身不具有方向的目标在遥感图像中依然是有向目标。

FRTDNet模型在一定程度上能够有效缓解有向小目标密集排列情况下导致的模型漏检问题。如图4所示，红色框表示基准模型漏检的情况。可以直观地看到，在一定程度上减少了漏检现象的发生，但仍然可能有漏检的情况发生。

图3 检测结果展示

Fig. 3 Display of detect results

图4 漏检情况对比，第一行为基准模型Oriented R-CNN，第二行为FRTDNet（本文的）

Fig. 4 Comparison of missed detections, the first behavior is the benchmark model, and the second behavior is FRTDNet (Ours)

4 结 论

针对O-RCNN网络忽略了有向目标边界和显著性信息导致的漏检误检问题，本文重新设计了检测器头部网络。该网络包括两部分：空间和通道信息重组回归分支RBIR和自注意力分类分支CLT，能够增强对有限条件下有向目标特征信息的捕获能力。其中，RBIR模块将空间信息与通道信息重组建模来解决定位过程中边界区域特征提取中信息湮灭和丢失的问题。CLT模块从全局信息入手来捕获不同目标的特征信息，提升有向目类别的检测能力。最后，本文基于O-RCNN网络提出了特征重组和自注意力检测网络FRTDNet，并在公开数据集DOTA、HRSC2016和SODA-A中进行了实验，验证了FRTDNet的有效性。值得注意的是，遥感有向目标还存在类间差异小、类内差异大导致的误检漏检问题，我们将在未来的研究中对此问题进行深入探讨。

参考文献（References）

Azimi S M, Vig E, Bahmanyar R, Körner M and Reinartz P. 2018. Towards multi-class object detection in unconstrained remote sensing imagery//Proceedings of the 14th Asian Conference on Computer Vision. Perth: Springer: 150-165 [DOI: 10.1007/978-3-030-20893-6_10]

Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer: 213-229 [DOI: 10.1007/978-3-030-58452-8_13]

Chen Z M, Chen K A, Lin W Y, See J, Yu H, Ke Y and Yang Y. 2020. PIoU loss: towards accurate oriented object detection in complex environments//Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer: 195-211 [DOI: 10.1007/978-3-030-58558-7_12]

Cheng G, Li Q Y, Wang G X, Xie X X, Min L T and Han J W. 2023a. SFRNet: fine-grained oriented object recognition via separate feature refinement. IEEE Transactions on Geoscience and Remote Sensing, 61: 5610510 [DOI: 10.1109/TGRS.2023.3277626]

Cheng G, Yao Y Q, Li S Y, Li K, Xie X X, Wang J B, Yao X W and Han J W. 2022. Dual-aligned oriented detector. IEEE Transactions on Geoscience and Remote Sensing, 60: 5618111 [DOI: 10.1109/TGRS.2022.3149780]

Cheng G, Yuan X, Yao X W, Yan K B, Zeng Q H, Xie X X and Han J W. 2023b. Towards large-scale small object detection: survey and benchmarks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(11): 13467-13488 [DOI: 10.1109/TPAMI. 2023.3290594]

Ding J, Xue N, Long Y, Xia G S and Lu Q K. 2019. Learning RoI transformer for oriented object detection in aerial images//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE [DOI: 10.1109/CVPR.2019.00296]

Everingham M, Van Gool L, Williams C K I, Winn J and Zisserman A. 2010. The pascal visual object classes (VOC) challenge. International Journal of Computer Vision, 88(2): 303-338 [DOI: 10.1007/s11263-009-0275-4]

Fu K, Chang Z H, Zhang Y and Sun X. 2021. Point-based estimator for arbitrary-oriented object detection in aerial images. IEEE Transactions on Geoscience and Remote Sensing, 59(5): 4370-4387 [DOI: 10.1109/TGRS.2020.3020165]

Han J M, Ding J, Li J and Xia G S. 2022. Align deep features for oriented object detection. IEEE Transactions on Geoscience and Remote Sensing, 60: 5602511 [DOI: 10.1109/TGRS.2021.3062048]

He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE [DOI: 10.1109/ICCV.2017.322]

Jiang Y Q, Tan Z Y, Wang J Y, Sun X Y, Lin M and Li H. 2022. GiraffeDet: a heavy-neck paradigm for object detection. arXiv preprint arXiv: 2202.04256

Jiang Y Y, Zhu X Y, Wang X B, Yang S L, Li W, Wang H, Fu P and Luo Z B. 2017. R2CNN: rotational region CNN for orientation robust scene text detection. arXiv preprint arXiv: 1706.09579

Li C Z, Xu C Y, Cui Z, Wang D, Zhang T and Yang J. 2019. Feature-attentioned object detection in remote sensing imagery//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei, China: IEEE [DOI: 10.1109/ICIP.2019.8803521]

Li H G, Yu R N and Ding W R. 2021. Research development of small object traching based on deep learning. Acta Aeronauticaet Astronautica Sinica, 42(7): 024691 (李红光, 于若男, 丁文锐. 2021. 基于深度学习的小目标检测研究进展. 航空学报, 42(7): 024691) [DOI: 10.7527/S1000-6893.2020.24691]

Li W T, Chen Y J, Hu K X and Zhu J K. 2022. Oriented RepPoints for aerial object detection//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE [DOI: 10.1109/CVPR52688.2022.00187]

Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2017. Focal loss for dense object detection//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE [DOI: 10.1109/ICCV.2017.324]

Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]

Liu Z K, Wang H Z, Weng L B and Yang Y P. 2016. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. IEEE Geoscience and Remote Sensing Letters, 13(8): 1074-1078 [DOI: 10.1109/LGRS.2016.2565705]

Ma J Q, Shao W Y, Ye H, Wang L, Wang H, Zheng Y B and Xue X Y. 2018. Arbitrary-oriented scene text detection via rotation proposals. IEEE Transactions on Multimedia, 20(11): 3111-3122 [DOI: 10.1109/TMM.2018.2818020]

Min L T, Fan Z M, Lv Q Y, Reda M, Shen L H and Wang B L. 2023. YOLO-DCTI: small object detection in remote sensing base on contextual transformer enhancement. Remote Sensing, 15(16): 3970 [DOI: 10.3390/rs15163970]

Ming Q, Zhou Z Q, Miao L J, Zhang H W and Li L H. 2021. Dynamic anchor learning for arbitrary-oriented object detection//Proceedings of the 35th AAAI Conference on Artificial Intelligence. [s.l.]: AAAI Press [DOI: 10.1609/aaai.v35i3.16336]

Nie G T and Huang H. 2021. A survey of object detection in optical remote sensing images. Acta Automatica Sinica, 47(8): 1749-1768 (聂光涛, 黄华. 2021. 光学遥感图像目标检测算法综述. 自动化学报, 47(8): 1749-1768) [DOI: 10.16383/j.aas.c200596]

Nie G T and Huang H. 2023. Multi-oriented object detection in aerial images with double horizontal rectangles. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 4932-4944 [DOI: 10.1109/TPAMI.2022.3191753]

Pan X J, Ren Y Q, Sheng K K, Dong W M, Yuan H L, Guo X W, Ma C Y and Xu C S. 2020. Dynamic refinement network for oriented and densely packed object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE [DOI: 10.1109/CVPR42600.2020.01122]

Qian W, Yang X, Peng S L, Yan J C and Guo Y. 2021. Learning modulated loss for rotated object detection//Proceedings of the 35th AAAI Conference on Artificial Intelligence. [s.l.]: AAAI Press [DOI: 10.1609/aaai.v35i3.16347]

Song G L, Liu Y and Wang X G. 2020. Revisiting the sibling head in object detector//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE [DOI: 10.1109/CVPR42600.2020.01158]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones K, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc.: 6000-6010

Wang J W, Ding J, Guo H W, Cheng W S, Pan T and Yang W. 2019. Mask OBB: a semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images. Remote Sensing, 11(24): 2930 [DOI: 10.3390/rs11242930]

Wang J W, Yang W, Li H C, Zhang H J and Xia G S. 2021. Learning center probability map for detecting objects in aerial images. IEEE Transactions on Geoscience and Remote Sensing, 59(5): 4307-4323 [DOI: 10.1109/TGRS.2020.3010051]

Xia G S, Bai X, Ding J, Zhu Z, Belongie S, Luo J B, Datcu M, Pelillo M and Zhang L P. 2018. DOTA: a large-scale dataset for object detection in aerial images//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE [DOI: 10.1109/CVPR.2018.00418]

Xie X X, Cheng G, Wang J B, Yao X W and Han J W. 2021. Oriented R-CNN for object detection//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE [DOI: 10.1109/ICCV48922.2021.00350]

Xu Y C, Fu M T, Wang Q M, Wang Y K, Chen K, Xia G S and Bai X. 2021. Gliding vertex on the horizontal bounding box for multi-oriented object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(4): 1452-1459 [DOI: 10.1109/TPAMI.2020.2974745]

Yang X, Yan J C, Feng Z M and He T. 2021. R3Det: refined single-stage detector with feature refinement for rotating object//Proceedings of the AAAI 35th Conference on Artificial Intelligence. [s.l.]: AAAI Press [DOI: 10.1609/aaai.v35i4.16426]

Yang X, Yang J R, Yan J C, Zhang Y, Zhang T F, Guo Z, Sun X and Fu K. 2019. SCRDet: towards more robust detection for small, cluttered and rotated objects//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE [DOI: 10.1109/ICCV.2019.00832]

Zhang G J, Lu S J and Zhang W. 2019. CAD-Net: a context-aware detection network for objects in remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 57(12): 10015-10024 [DOI: 10.1109/TGRS.2019.2930982]

Zhang G J, Luo Z P, Tian Z C, Zhang J Y, Zhang X Q and Lu S J. 2023. Towards efficient use of multi-scale features in transformer-based object detectors//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE [DOI: 10.1109/CVPR52729.2023.00601]

Feature reassembly and self−attention for oriented object detection in remote sensing images

MIN Lingtong[1], FAN Ziman[1], XIE Xingxing[2], LYU Qinyi[1]

1.School of Electronic Information, Northwestern Polytechnical University, Xi'an, 710072, China;
2.School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China

Abstract: Oriented object detection in remote sensing images is an exceptionally challenging task that has elicited widespread attention.

With the rapid advancement of deep learning, neural networks based on convolutional neural networks and self-attention networks (e.g., Transformers) have achieved remarkable progress in oriented object detection. However, the focus on boundary and salient feature information in oriented objects in remote sensing images is lacking. Specifically, extracting boundary information for objects with varying orientations is difficult, and the global dependency of salient features is sparse. To address these issues, we propose a method of small-object detection in remote sensing images on the basis of feature reassembly and self-attention. This method consists of a regression branch that incorporates spatial channel reassembly and a self-attention classification branch. The regression branch reassembles spatial information along the channel dimension and emphasizes boundary-sensitive information to achieve accurate localization of bounding boxes. The classification branch leverages self-attention with positional information to capture fundamentally discriminative object features, thus enhancing global feature dependencies for precise classification. Extensive experiments demonstrate the effectiveness and robustness of the proposed model and showcase its excellent performance on publicly available datasets, such as DOTA, HRSC2016, and SODA-A.

Key words：remote sensing image, small object detection, detection head, feature reorganization, transformer

Supported by National Natural Science Foundation of China (No. 62206221)