

## 基于注意力机制与动态激活的空间目标检测算法

刘胜利<sup>1</sup>, 郭裕兰<sup>2</sup>, 王刚<sup>1\*</sup><sup>1</sup>空军工程大学防空反导学院, 陕西 西安 710051;<sup>2</sup>国防科技大学电子科学学院, 湖南 长沙 410073

**摘要** 基于图像的空间目标检测已成为保障在轨卫星运行安全的重要需求之一。已有基于深度学习的无锚框目标检测算法取得了良好进展,但是仍存在检测头结构简单、表征能力不足的问题。对此,提出了基于注意力机制与动态激活的空间目标检测算法。以无锚框目标检测算法的通用网络结构为基础,在检测头中使用基于通道与空间感知的残差注意力模块,以增强网络的特征表征能力;同时,在检测头中串联基于通道感知的动态激活模块,以提升网络在特定空间目标检测任务中的性能。在 SPARK 空间目标检测数据集上的实验结果表明,所提算法的 AP@IoU=0.50:0.95 指标达 77.1%,检测性能显著优于主流算法 Faster R-CNN、YOLOv3 及 FCOS。此外,所提算法在训练过程中采用动态样本匹配策略,进一步提升了对小目标的检测能力。

**关键词** 目标检测; 注意力机制; 动态激活; 空间目标

中图分类号 TP391.4 文献标志码 A

DOI: 10.3788/LOP202259.1415021

## Space Target Detection Algorithm Based on Attention Mechanism and Dynamic Activation

Liu Shengli<sup>1</sup>, Guo Yulan<sup>2</sup>, Wang Gang<sup>1\*</sup><sup>1</sup>College of Air and Missile Defense, Air Force Engineering University, Xi'an 710051, Shaanxi, China;<sup>2</sup>College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, Hunan, China

**Abstract** Image-based space target detection has become one of the crucial requirements to ensure the safety of in-orbit satellites. Existing anchor-free target detection algorithms based on deep learning have achieved outstanding results. However, their detection heads have a simple structure, resulting in insufficient representation ability. To overcome this challenge, we propose a space target detection algorithm based on attention mechanism and dynamic activation. Based on the anchor-free target detection algorithm's general network structure, the channel and spatial aware-based residual attention module is employed in the detection head to improve the network's feature representation ability. Meanwhile, the channel aware-based dynamic activation module is connected in series with the detection head to enhance the network's performance in a specific space target detection task. The experimental findings on the SPARK space target detection dataset demonstrate that the proposed algorithm achieves an AP@IoU=0.50:0.95 of 77.1%, and its detection performance is substantially better than the mainstream algorithms such as Faster R-CNN, YOLOv3, and FCOS. Additionally, to further enhance the detection ability for small targets, the dynamic label assignment approach is adopted in the training process.

**Key words** object detection; attention mechanism; dynamic activation; space target

## 1 引言

随着航天技术的进步,以卫星为基础设施的定位导航、卫星通信、遥感观测以及监视侦察等技术<sup>[1-3]</sup>得

到蓬勃发展,在民用与国防领域发挥着至关重要的作用。但是,数量不断增长的卫星,以及伴随产生的大量太空垃圾与废弃卫星,对在轨正常工作卫星的安全构成了严重威胁。因此,使卫星具有自动识别周围环境

收稿日期: 2022-03-31; 修回日期: 2022-04-26; 录用日期: 2022-05-18

基金项目: 国家自然科学基金(62106283)

通信作者: \*sharesunny123@163.com

中目标的能力,已成为最小化在轨卫星安全风险的迫切需求。在过去几年中,图像传感器被大量应用到卫星上,用于增强卫星对其周围环境的感知能力,这使基于图像的空间目标检测方法具有重大的研究意义和应用价值。

近年来,随着深度学习技术的不断发展,端到端的神经网络已成为主导目标检测模型发展的主流方法。根据是否使用预定义锚框,可将基于神经网络的目标检测模型分为两类:基于锚框的目标检测模型与无锚框的目标检测模型。基于锚框的目标检测模型包含两阶段目标检测模型(如 Faster R-CNN<sup>[4]</sup>、Mask R-CNN<sup>[5]</sup>等)和单阶段目标检测模型(如 RetinaNet<sup>[6]</sup>、SSD<sup>[7]</sup>、YOLO 系列<sup>[8-11]</sup>等)。两阶段目标检测模型需要使用 region proposal network(RPN)<sup>[4]</sup>生成包含目标的候选区域,模型结构与计算复杂度较高。为简化目标检测模型的网络结构,单阶段目标检测模型放弃了基于 RPN 的技术路线,直接在特征图上进行逐像素推理。但是,对于预定义锚框的设置,需要根据特定数据集进行聚类分析,而且锚框的引入增加了检测头的复杂度以及生成结果的数量,故基于锚框的目标检测模型的复杂度和计算代价依然难以令人满意。因此,针对上述不足,该领域的研究者提出了结构更加简单、训练成本更加低廉的无锚框目标检测模型,如 FCOS<sup>[12]</sup>、CenterNet<sup>[13]</sup>等。无锚框的目标检测模型不使用预定义锚框作为先验信息,在特征图上直接逐像素进行目标分类与边界框参数回归,取得了优越的性能表现,并大幅降低了模型的复杂度。但是,无锚框目标检测模型的检测头通常是由若干卷积层堆叠而成的,这种简单的结构设计导致网络的特征表征能力不足,从而限制了目标检测的性能。

针对空间目标检测任务,Chen 等<sup>[14]</sup>提出了首个基

于神经网络的空间目标检测模型,该模型在基于锚框的两阶段目标检测模型 Faster R-CNN 的基础上,使用可提取高分辨率特征图的 HRNet<sup>[15]</sup>作为骨干网络,以增强网络的特征提取能力,使模型具有良好的检测性能。但是,由于沿用了 Faster R-CNN 的模型框架,模型的参数量大、复杂度高,导致计算代价较大。对此,Hu 等<sup>[16]</sup>使用基于锚框的单阶段目标检测模型 YOLOv3<sup>[10]</sup>进行空间目标检测,在确保模型检测精度的同时,降低了模型的复杂度与计算代价。但是,上述方法皆是基于锚框的目标检测方法,与无锚框目标检测方法相比,在模型复杂度、计算代价等方面依然不具有优势。

针对上述问题,面向空间目标检测任务,本文提出了一种基于注意力机制与动态激活的无锚框空间目标检测模型。在无锚框目标检测模型的通用网络结构的基础上,利用通道与空间感知注意力机制改进检测头的结构,提升网络的特征表征能力,并结合残差连接构建残差注意力网络模块,使网络在训练过程中可稳定收敛。同时,为了提升网络在特定的空间目标检测任务中的性能表现,在检测头中引入通道感知动态激活函数,进一步提升模型的检测精度。最后,针对模型对小目标的检测能力不足的问题,在训练过程中使用动态样本匹配策略,增强了模型对小目标的检测能力。

## 2 空间目标检测算法模型

### 2.1 网络结构

提出的无锚框的空间目标检测模型的网络结构如图 1 所示。模型使用 ResNet50<sup>[17]</sup>作为主干网络进行特征提取,使用特征金字塔网络(FPN)<sup>[18]</sup>进行特征融合,在检测头部分使用注意力机制与动态激活函数。

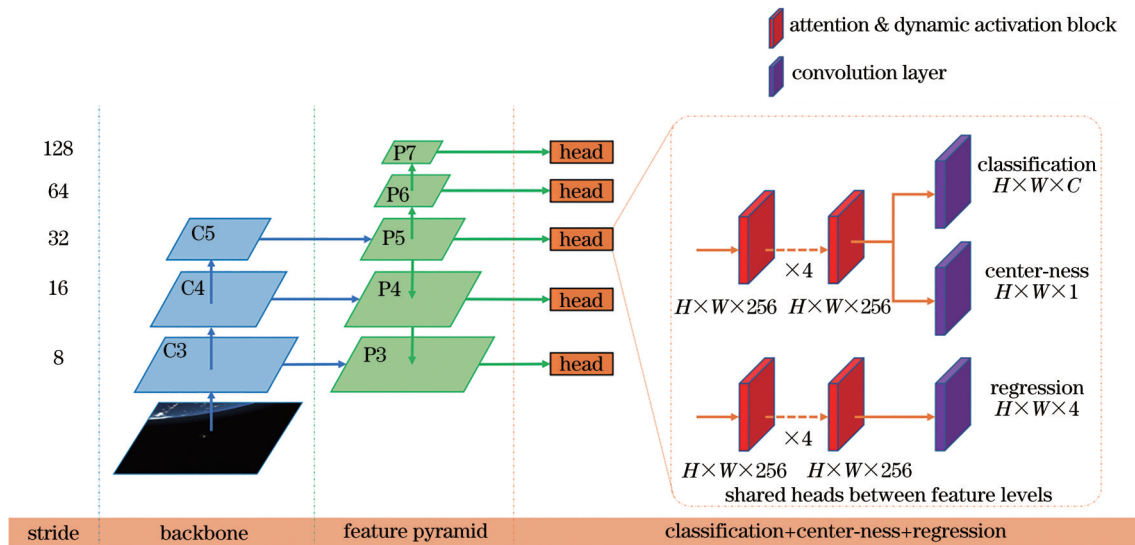


图 1 空间目标检测模型的网络结构

Fig. 1 Network structure of space target detection model

通过采用多尺度检测策略,FPN生成5种尺寸的特征图,分别为输入图像尺寸的1/8,1/16,1/32,1/64,1/128。这些生成的特征图将输入图像划分为与其尺寸对应的网格,检测头预测各网格内是否包含目标和所包含目标的类别、网格中心与目标的真值边界框中心的接近程度、目标边界框参数。检测头在各尺度特征

图上共享权重。

### 2.2 注意力与动态激活模块

空间目标检测模型的检测头部分包含4个依次连接的注意力与动态激活模块。该模块的具体网络结构如图2所示,包括基于通道与空间感知的残差注意力模块与基于通道感知的动态激活模块。

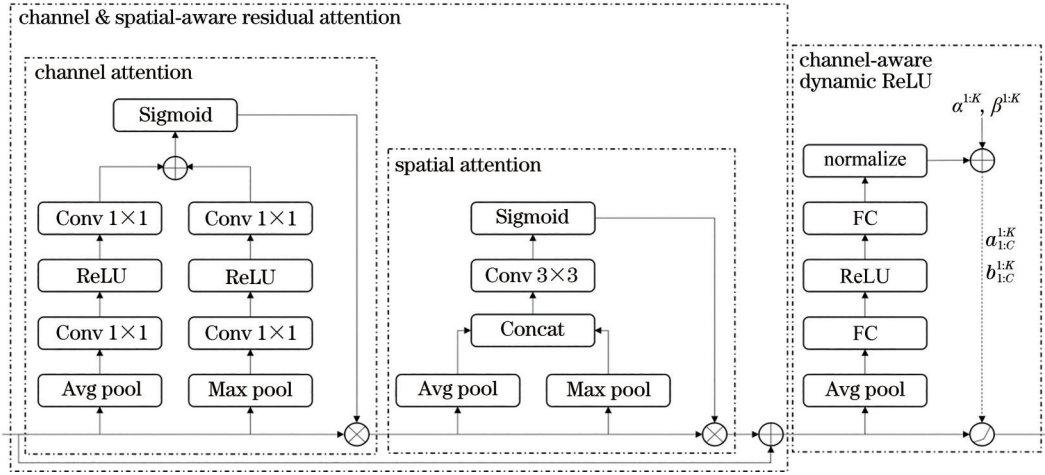


图2 注意力与动态激活模块的网络结构

Fig. 2 Network structure of attention and dynamic activation module

#### 2.2.1 基于通道与空间感知的残差注意力模块

注意力机制是一个动态选择的过程,其根据输入的重要性自适应地为特征分配权重,从而提升网络的特征表征能力。本文在目标检测模型的检测头部分使用了通道与空间注意力机制。在神经网络中,不同特征图的不同通道通常表示不同的目标,而通道注意力机制对每个通道的权重进行自适应校准,从而实现了目标的自适应选择,引导网络更多地关注重要的目标。空间注意力机制则是一种自适应的空间区域选择机制,引导网络更多地关注特征图上重要的空间区域。基于通道与空间感知的残差注意力模块联合了通道注意力与空间注意力的优势,引导网络自适应地选择重要的目标与区域,同时引入残差连接,使网络在训练时更易收敛。

基于通道与空间感知的残差注意力模块对卷积块注意力机制<sup>[19]</sup>与残差注意力网络<sup>[20]</sup>进行融合。首先,使用通道感知注意力机制,实现动态自适应的目标选择。设输入的特征图为 $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ ,其中 $C$ 、 $H$ 和 $W$ 分别为特征图的通道数、高度和宽度,则通道感知注意力机制的公式为

$$M_c(\mathbf{X}) = \sigma \{ \text{MLP}[\text{GAP}^s(\mathbf{X})] + \text{MLP}[\text{GMP}^s(\mathbf{X})] \} \mathbf{X}, \quad (1)$$

式中: $\text{GAP}^s(\cdot)$ 和 $\text{GMP}^s(\cdot)$ 分别为在空间域上的全局平均池化操作和全局最大池化操作; $\text{MLP}(\cdot)$ 为包含单个隐层的多层感知机; $\sigma$ 为Sigmoid激活函数。

然后,使用空间感知注意力机制,寻找特征图中重要的空间区域,与通道感知注意力机制形成互补。空

间感知注意力机制的公式为

$$M_s(\mathbf{X}) = \sigma \{ \text{Conv} \{ [\text{GAP}^c(\mathbf{X}), \text{GMP}^c(\mathbf{X})] \} \} \mathbf{X}, \quad (2)$$

式中: $\text{Conv}(\cdot)$ 表示卷积操作; $\text{GAP}^c(\cdot)$ 和 $\text{GMP}^c(\cdot)$ 分别为在通道域上的全局平均池化操作和全局最大池化操作; $[\cdot]$ 表示在通道域上的堆叠操作。

最后,结合残差网络结构,基于通道与空间感知的残差注意力机制可总结为

$$M(\mathbf{X}) = \mathbf{X} + M_s[M_c(\mathbf{X})]. \quad (3)$$

基于通道与空间感知的残差注意力模块通过挖掘特征图中重要空间区域的分布以及特征图中各通道的重要性,增强网络的特征表示能力,有助于提升目标检测模型的检测效果。

#### 2.2.2 基于通道感知的动态激活模块

修正线性单元(ReLU)<sup>[21]</sup>是在神经网络中被广泛使用的激活函数之一,它虽结构简单,但功能强大,可以提升神经网络的性能。但是,ReLU是一种静态激活函数,对不同的输入皆以相同的方式进行处理,这将导致ReLU的功能可能与特定输入、特定任务不匹配的问题,从而影响了网络在不同输入上、不同任务中的性能表现。因此,使用动态的激活函数将对提升网络性能起到积极作用。本文在目标检测模型的检测头部分使用了动态ReLU激活函数<sup>[22]</sup>,首先对输入张量的全局信息通过超函数进行编码,然后由超函数的输出自适应地确定动态ReLU激活函数的参数,这将显著提升网络的特征表征能力。同时,由于超函数结构简单,动态ReLU激活函数额外产生的计算代

价可忽略不计,从而保持网络的计算效率。

基于通道感知的动态激活模块使用动态 ReLU 作为动态激活函数,针对特征图的不同通道动态生成不同的激活函数。设输入张量为  $\mathbf{x}$ ,  $\mathbf{x}$  的第  $c$  通道为  $x_c$ , 动态 ReLU 的公式为

$$y_c = f_{\theta(\mathbf{x})}(x_c) = \max_{1 \leq k \leq K} [a_c^k x_c + b_c^k], \quad (4)$$

式中:  $\theta(\mathbf{x})$  为超函数。  $(a_c^k, b_c^k)$  为  $\theta(\mathbf{x})$  的输出, 即

$$\begin{bmatrix} a_1^1, \dots, a_c^1, \dots, a_1^K, \dots, a_c^K, b_1^1, \dots, \\ b_c^1, \dots, b_1^K, \dots, b_c^K \end{bmatrix}^T = \theta(\mathbf{x}), \quad (5)$$

式中:  $K$  为激活函数中分段函数的数量。

超函数的具体实现过程为首先使用全局平均池化操作对输入进行降维处理, 然后通过包含单个隐层的多层感知机输出  $2KC$  个参数, 即残差  $a_{1:c}^{1:K}$  和  $b_{1:c}^{1:K}$ , 记作  $\Delta a_{1:c}^{1:K}$  和  $\Delta b_{1:c}^{1:K}$ , 接着通过  $2\sigma(\mathbf{x}) - 1$  将残差标准化为  $-1$  到  $1$  之间的值, 其中  $\sigma(\mathbf{x})$  为 Sigmoid 激活函数, 最后输出残差与初始化值的和, 即

$$a_c^k = \alpha^k + \lambda_a \Delta a_c^k, \quad (6)$$

$$b_c^k = \beta^k + \lambda_b \Delta b_c^k, \quad (7)$$

式中:  $\alpha^k$  和  $\beta^k$  是  $a_c^k$  和  $b_c^k$  的初始化值;  $\lambda_a$  和  $\lambda_b$  是控制残差范围的标量;  $\alpha^k, \beta^k, \lambda_a$  和  $\lambda_b$  都是超参数。实验设置  $K = 2, \alpha^1 = 1, \alpha^2 = \beta^1 = \beta^2 = 0, \lambda_a = 1.0, \lambda_b = 0.5$ 。

### 3 实验与结果分析

#### 3.1 数据集

实验所使用的空间目标数据集为 SPARK 数据集<sup>[23-24]</sup>, 样例如图 3 所示。2021 年, 卢森堡大学针对空间目标识别与检测任务提出了 SPARK 数据集。该数据集包含 10 类卫星目标与 1 类太空垃圾目标, 共计 11 类空间目标,  $150 \times 10^3$  张 RGB 图像。由于 SPARK 数据集是利用 Unity3D 游戏引擎合成的图像数据集, 故为使合成图像更加逼近于真实场景, 卢森堡大学的研究者在充分考虑了外层空间光照、宇宙与地球背景、

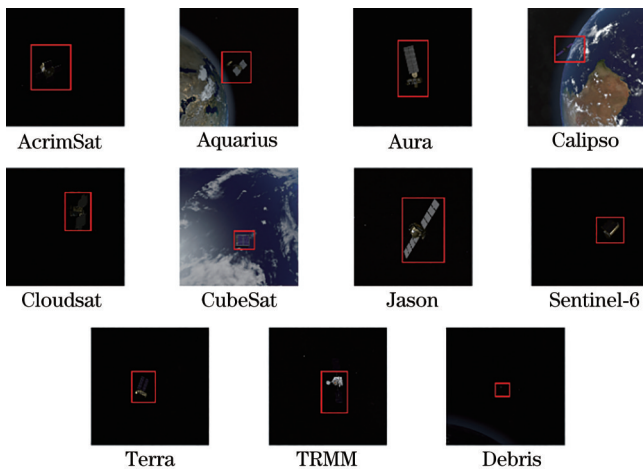


图 3 SPARK 数据集样例

Fig. 3 Examples of SPARK dataset

相机与目标间距离、光学传感器噪声等现实情况的基础上, 生成了大量背景丰富、光照条件不同、目标尺度多样且带有噪声的图像, 这将确保在 SPARK 数据集上训练的空间目标检测模型对真实场景有良好的泛化能力。

#### 3.2 评价指标

对于目标检测模型, 通常使用平均精度 (AP) 度量单目标检测效果, 使用平均精度均值 (mAP) 度量多目标检测效果<sup>[25]</sup>。mAP 为所有类别的 AP 的平均值, 通常 mAP 也被称为 AP。实验采用 COCO 检测挑战中的评价指标 AP 与 AP Across Scales<sup>[26]</sup> 度量目标检测模型的性能。在 50% 到 95% 的范围内, 以 5% 的步长均匀采样, 得到 10 个 intersection over union (IoU) 阈值, 计算在这些 IoU 阈值上的 AP 的平均值, 记作 AP@IoU = 0.50:0.95。这种对多个 IoU 上的 AP 求平均的策略可以更全面地度量检测模型的性能。AP Across Scales 为在三种目标尺寸条件下的 AP@IoU = 0.50:0.95, 分别为小目标 ( $S < 322$  pixel)、中目标 ( $322$  pixel  $< S < 962$  pixel) 和大目标 ( $S > 962$  pixel), 其中  $S$  为真值边界框面积。

为评价目标检测模型对各类目标的具体检测效果, 采用针对目标检测任务的标准化混淆矩阵进行量化分析。目标检测中的混淆矩阵的正负例是根据检测框与真值框的 IoU 进行判定的, 具体分为四种情况:

- 1) 当两个框的 IoU 大于阈值, 同时两个框对应的类别也相同时, 检测框被判定为正确的检测结果 (TP);
- 2) 当两个框的 IoU 大于阈值, 但两个框对应的类别不相同, 检测框被判定为误检 (FP);
- 3) 当两个框的 IoU 小于阈值时, 检测框被判定为背景的误检 (FP);
- 4) 当没有检测框时, 则被判定为漏检 (FN)。

因此, 由于混淆矩阵的行指标表示真值类别, 列指标表示预测类别, 故目标检测混淆矩阵中的对角线元素表示各类的 TP 数量, 最后一列元素表示各类的 FN 数量, 最后一行元素表示背景被误检的 FP 数量, 其余元素表示被误检为其他各类的 FP 数量。但是, 由于目标检测数据集中各类别的目标数目过多, 为了能更好地显示, 同时也为了能更直观地看出每个类别的检测率和误检率, 故进行归一化处理, 即对目标检测混淆矩阵的每一行中的数值都除以对应类别的目标总数, 以百分比来表示结果, 从而得到标准化混淆矩阵。在标准化混淆矩阵中, 对角线元素表示各类的检测率, 最后一列元素表示各类的漏检率, 最后一行元素表示背景误检率, 其余元素表示被误检为其他各类的误检率。

#### 3.3 实验环境与训练策略

实验使用的硬件为 Intel Core i9-10900X CPU、NVIDIA GeForce RTX 3090 24 G GPU、32 G 内存, 采用的软件环境为 Python 3.7、CUDA 11.3, 使用的

深度学习框架是 PyTorch 1.9。模型训练时使用 ResNet50 在 ImageNet 数据集<sup>[27]</sup>上的预训练模型初始化主干网络的参数,可以加快模型的训练速度;同时,使用冻结训练策略,将 ResNet50 的 stem 与第一个 stage 的权重固定,不进行训练。设置 epoch 数为 12, batch\_size 为 16, 初始学习率为 0.01, 采用阶梯式衰减策略调整学习率, 在第 8 个 epoch 与第 11 个 epoch 将学习率衰减为原学习率的 10%。

### 3.4 结果分析

为验证所提算法的有效性,将基于锚框的 Faster R-CNN 和 YOLOv3 算法、无锚框的 FCOS 算法作为基准算法进行比较,并进行消融实验,结果如表 1 所示,其中“√”表示含有该结构,“×”表示未含有该结构,Attn 表示在检测头中仅使用基于通道与空间感知的残差注意力模块的空间目标检测模型,DyReLU 表示在检测头中仅使用基于通道感知的动态激活模块的空间目标检测模型,Ours 表示所提空间目标检测模型。实验结果表明,所提算法的检测性能总体优于 Faster R-CNN、YOLOv3 以及 FCOS 算法, AP@IoU=

0.50:0.95 指标达 77.1%, AP@IoU=0.50 指标达 91.4%, AP@IoU=0.75 指标达 82.8%。作为一种无锚框空间目标检测算法,所提算法的各 AP 指标相较于性能较优的基于锚框的 Faster R-CNN 算法分别提升 1.7 个百分点、2.9 个百分点及 1.1 个百分点,相较于无锚框的 FCOS 算法分别提升 1.6 个百分点、3.1 个百分点及 2.2 个百分点。同时,同作为无锚框目标检测算法,所提算法对大目标与中目标的检测能力均优于 FCOS,但是对小目标的检测能力劣于 FCOS,针对这一不足,下文进行了分析与改进。

通过消融实验可以看出,Attn 的 AP 指标相较于 FCOS 均获得提升,证明基于通道与空间感知的残差注意力模块比卷积层具有更强的特征表征能力,使检测头的性能得到较大提升;DyReLU 的 AP 指标相较于 FCOS 有小幅度的提升,同时 Ours 的 AP 指标相较于 Attn 亦有小幅度的提升,证明在特征提取层后串联基于通道感知的动态激活模块可以有效提升检测头的性能。因此,两个模块在空间目标检测模型中对提升检测效果均起到了重要作用。

表 1 消融实验结果对比

Table 1 Comparison of ablation experimental results

Model	Attention	Dynamic ReLU	AP / %			AP Across Scales / %		
			IoU=0.50:0.95	IoU=0.50	IoU=0.75	Small	Medium	Large
Faster R-CNN	×	×	75.4	88.5	81.7	41.6	72.0	80.3
YOLOv3	×	×	64.1	87.6	73.1	29.8	63.6	67.8
FCOS	×	×	75.5	88.3	80.6	36.3	71.9	80.6
Attn	√	×	76.5	89.9	82.1	31.8	73.7	83.7
DyReLU	×	√	75.6	88.6	81.1	33.8	71.9	81.5
Ours	√	√	77.1	91.4	82.8	32.2	74.4	84.6

为评价所提算法在 SPARK 数据集上的检测效果,绘制 FCOS 与所提算法的标准化混淆矩阵 (IoU 阈值为 50%), 如图 4 所示。通过标准化混淆矩阵可以看出,除 Cloudsat 类外,所提算法对其他各类目标的检测能力均与 FCOS 相当,检测率皆在 89% 以上。而 Cloudsat 类是 SPARK 数据集中一类十分具有挑战性的目标,FCOS 与所提算法均没有实现对该类目标的检测,其中 FCOS 与所提算法的漏检率分别达 53% 和 52%。这是由于 Cloudsat 类目标是弱纹理目标,与背景的差异较小,所提算法与 FCOS 所采用的骨干网络与特征融合网络皆为 ResNet50 与 FPN,对弱纹理目标的表征学习能力较弱,所提取特征的辨识度不强,进而导致目标检测模型对该类目标的检测性能不佳,产生了较多漏检目标。此外,网络对 Debris 与 Terra 类目标的背景误检率分别为 28% 与 22%,产生该现象的原因是 Debris 类是由形状不规则的太空垃圾组成的,与背景中的噪声较为相似,故网络将背景中噪声误检为

Debris 类目标。而 Terra 类目标的主体结构是与背景纹理相似的太阳能板,从而导致较高的背景误检率。

针对所提算法对小目标检测能力不足的问题,通过对目标检测模型的全流程分析,发现训练中所采用的样本匹配策略是造成该问题的主要原因。实验中采用的样本匹配策略是最大 IoU 匹配,即将与原图像上真值边界框区域对应的特征图上的像素皆定义为正样本,而这将导致小目标的正样本数目较少,进而导致网络对小目标的检测能力较弱。因此,实验改进了样本匹配策略,采用 simOTA 样本匹配策略<sup>[28]</sup>动态地确定各目标的正样本数目,从而增加小目标的正样本数量,缓解样本不均衡的问题,实验结果如表 2 所示,其中模型 1 使用最大 IoU 匹配策略进行训练,模型 2 使用 simOTA 匹配策略进行训练。实验结果表明,使用 simOTA 样本匹配策略后,所提算法对小目标的检测能力得到显著提升,同时也保持了对其他尺寸目标的检测精度。

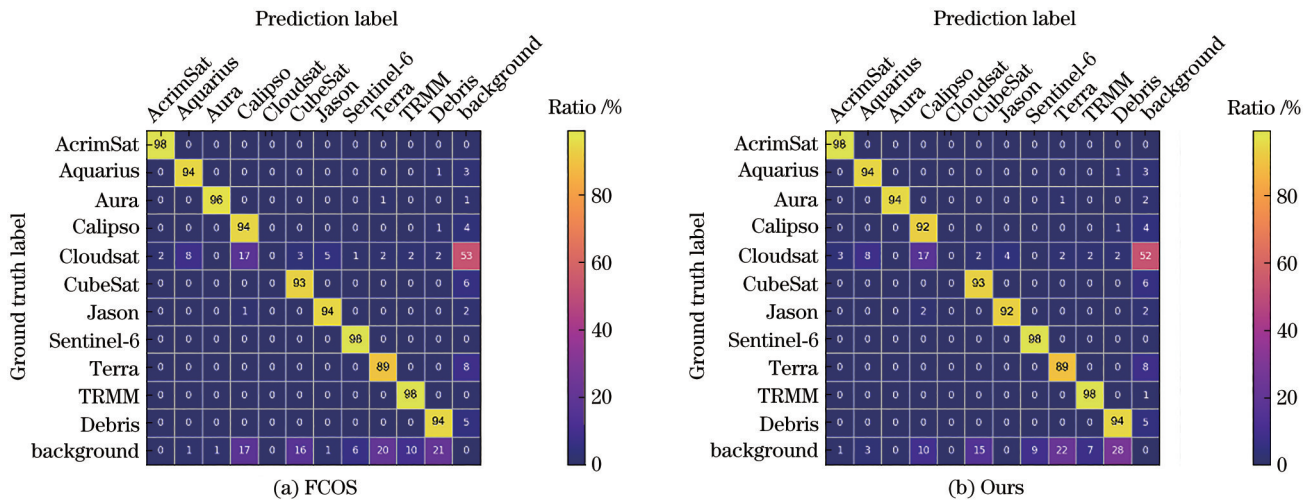


图 4 标准化混淆矩阵  
Fig. 4 Normalized confusion matrix

表 2 样本匹配策略对比  
Table 2 Comparison of label assignment strategy

Model	Label assigner	AP / %			AP Across Scales / %		
		IoU=0.50:0.95	IoU=0.50	IoU=0.75	Small	Medium	Large
FCOS	MaxIoU	75.5	88.3	80.6	36.3	71.9	80.6
Model 1	MaxIoU	77.1	91.4	82.8	32.2	74.4	84.6
Model 2	simOTA	76.7	90.5	82.3	37.7	73.8	83.5

## 4 结 论

由于保障卫星在轨工作安全的需要,基于图像的空间目标检测已成为十分重要的研究方向。针对目前无锚框目标检测模型的主流网络框架中检测头结构简单、表征能力不足的问题,结合空间目标检测任务,提出了一种基于注意力机制与动态激活函数的空间目标检测网络模型。在检测头中使用基于通道与空间感知的残差注意力网络模块,提升检测头的特征表征能力,并确保网络在训练时可以稳定收敛。同时,为了提升网络在空间目标检测任务中的检测性能,在检测头中引入了基于通道感知的动态 ReLU 激活函数,对特征图的不同通道动态生成不同的激活函数。在 SPARK 数据集上的实验结果表明:所提算法与基于锚框的 Faster R-CNN 相比,AP@IoU=0.50 指标提升了 2.9 个百分点,AP@IoU=0.75 指标提升了 1.1 个百分点,AP@IoU=0.50:0.95 指标提升了 1.7 个百分点;而与无锚框的 FCOS 相比,AP@IoU=0.50 指标提升了 3.1 个百分点,AP@IoU=0.75 指标提升了 2.2 个百分点,AP@IoU=0.50:0.95 指标提升了 1.6 个百分点,取得了良好的性能表现。另外,针对所提算法对小目标的检测能力不足的问题,通过使用动态样本匹配策略 simOTA,显著提升了对小目标的检测效果。尽管所提算法取得了较好的性能,但仍存在不足,例如对弱纹理目标的检测能力不足、对背景噪声误检率较高等,

后续工作可以使用表征学习能力更强的特征提取网络、跨尺度信息交互能力更强的特征融合机制、可深度挖掘目标自身相关性的多尺度自监督注意力机制等方法对无锚框空间目标检测算法进行进一步优化。

## 参 考 文 献

- [1] 汪鹏, 刘瑞, 辛雪静, 等. 基于残差网络的光学遥感图像场景分类算法[J]. 激光与光电子学进展, 2021, 58(2): 0210001.  
Wang P, Liu R, Xin X J, et al. Scene classification of optical remote sensing images based on residual networks [J]. Laser & Optoelectronics Progress, 2021, 58(2): 0210001.
- [2] 尚华哲, 胡斯勒图, 李明, 等. 基于被动遥感卫星可见至红外通道观测的云特性遥感[J]. 光学学报, 2022, 42(6): 0600003.  
Shang H Z, Hu S L T, Li M, et al. Remote sensing of cloud properties based on visible-to-infrared channel observation from passive remote sensing satellites[J]. Acta Optica Sinica, 2022, 42(6): 0600003.
- [3] 汪亚妮, 汪西莉. 基于注意力和特征融合的遥感图像目标检测模型[J]. 激光与光电子学进展, 2021, 58(2): 0228003.  
Wang Y N, Wang X L. Remote sensing image target detection model based on attention and feature fusion[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0228003.
- [4] Ren S 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 and

- Machine Intelligence, 2017, 39(6): 1137-1149.
- [5] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.
- [6] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.
- [7] 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.
- [8] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [9] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [10] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-04-03]. <https://arxiv.org/abs/1804.02767>.
- [11] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-05-04]. <https://arxiv.org/abs/2004.10934>.
- [12] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9626-9635.
- [13] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points [EB/OL]. (2019-04-16)[2021-04-05]. <https://arxiv.org/abs/1904.07850>.
- [14] Chen B, Cao J W, Parra A, et al. Satellite pose estimation with deep landmark regression and nonlinear pose refinement[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), October 27-28, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 2816-2824.
- [15] Wang J D, Sun K, Cheng T H, et al. Deep high-resolution representation learning for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3349-3364.
- [16] Hu Y L, Speierer S, Jakob W, et al. Wide-depth-range 6D object pose estimation in space[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 15865-15874.
- [17] He K M, Zhang X Y, Ren S Q, et al. Identity mappings in deep residual networks[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9908: 630-645.
- [18] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [19] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [20] Wang F, Jiang M Q, Qian C, et al. Residual attention network for image classification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6450-6458.
- [21] Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines[C]//Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel. Madison: Omnipress, 2010.
- [22] Chen Y P, Dai X Y, Liu M C, et al. Dynamic ReLU [M]//Vedaldi A, Bischof H, Brox T, et al. Computer Vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12364: 351-367.
- [23] Musallam M A, Ismael K A, Oyedotun O, et al. SPARK: SPAcecraft recognition leveraging knowledge of space environment[EB/OL]. (2021-04-13)[2021-08-07]. <https://arxiv.org/abs/2104.05978>.
- [24] Musallam M A, Gaudilliere V, Ghorbel E, et al. Spacecraft recognition leveraging knowledge of space environment: simulator, dataset, competition design and analysis[C]//2021 IEEE International Conference on Image Processing Challenges, September 19-22, 2021, AK, USA. New York: IEEE Press, 2021: 11-15.
- [25] Padilla R, Netto S L, da Silva E A B. A survey on performance metrics for object-detection algorithms[C]//2020 International Conference on Systems, Signals and Image Processing (IWSSIP), July 1-3, 2020, Niteroi, Brazil. New York: IEEE Press, 2020: 237-242.
- [26] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. Computer Vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8963: 740-755.
- [27] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL. New York: IEEE Press, 2009: 248-255.
- [28] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-06-18)[2021-04-05]. <https://arxiv.org/abs/2107.08430>.