基于残差自注意力和分离集合匹配的高效端到端航天器组件检测

陈明[1],牛燕菲[1],段莉[2*],高铁梁[3],楚杨阳[1],曹洁[1]

(1.郑州轻工业大学软件学院,河南 郑州 450000;2. 北京交通大学计算机与信息技术学院,北京 100000;
3.新乡学院商学院,河南 新乡 453000)

摘 要: 随着我国航天技术的迅猛发展,各种航天器相继发射,然而航天器在运行时将受到辐射、温度变化等不可控因素的影响,这会导致地面站无法精确测量和定位航天器的位置与姿态,从而对通信和航天器之间的对接或抓捕等空间在轨服务产生影响。为了解决上述问题,首先对包含检测、分割与部件识别的航天器数据集 SDDSP 中的部件进行人工标注,该数据集共包含 3 117 张航天器图片,标注后得到 11 001 个检测目标;然后提出一种空间在轨服务中基于残差自注意力(RS)和分离集合匹配(SSM)的高效端到端航天器组件检测模型,该模型在 Sparse DETR 模型的基础上引入残差自注意力机制解决了稀疏标记(token)导致的收敛速度降低并影响模型预测精度的问题,引入分离集合匹配机制解决了二分匹配过程中可能出现的不稳定性现象。实验结果表明,在 SDDSP 数据集上,该模型的平均精确率(AP)和收敛速度相比于基线 DETR 模型提升了 17.9 个百分点和 10 倍,相比于 Sparse DETR 模型提升了 3.1 个百分点和 20%。

关键词: 航天器组件检测;Sparse DETR 模型;残差自注意力;分离集合匹配;航天器数据集

源代码链接: https://github.com/304999183/RSSSM-DETR-master.git

中图分类号: TP301.6 **文献标志码:** A **DOI:** 10.19678/j.issn.1000-3428.0068524

Efficient End-to-End Spacecraft Component Detection Based on Residual Self-attention and Separated Set Matching

CHEN Ming[1], NIU Yanfei[1], DUAN Li[2*], GAO Tieliang[3], CHU Yangyang[1], CAO Jie[1]

(1. College of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, Henan, China;
2. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100000, China;
3. Bussiness School, Xinxiang University, Xinxiang 453000, Henan, China)

【Abstract】 The rapid development of space technology in China has led to a multitude of spacecraft launches. However, these spacecraft are expected to experience the influence of uncontrollable factors such as radiation and temperature changes during operation. These changes may impede the accurate measurement of spacecraft positions and behaviors by ground stations, thereby impacting on-orbit services such as communications and docking, as well as grappling between spacecraft. To solve these problems, the present study first annotates the SDDSP spacecraft dataset which encompasses detection, segmentation, and component recognition with 3 117 spacecraft images and 11 001 detection targets. An efficient end-to-end spacecraft component detection model is then proposed based on Residual Self-attention (RS) and Separated Set Matching (SSM) in space on-orbit services. The RS mechanism is introduced on the basis of the Sparse DEtection TRansformer (DETR) model to solve the problem of sparse tokens, which slows convergence and degrades the prediction accuracy of the model. Furthermore, SSM is deployed to address the phenomenon of instability that may occur in the process of dichotomous matching. The experimental results show that the Average Precision (AP) and convergence speed of the model are improved by 17.9 percentage points and 10 times, respectively, compared with those of the baseline DETR model, as well as 3.1 percentage points and 20%, respectively, compared with those of the Sparse DETR model.

【Key words】 spacecraft component detection; Sparse DERT model; Residual Self-attention(RS); Separated Set Matching(SSM); aircraft dataset

0 引言

随着科技进步,人类探索太空的能力得到突飞猛进的发展,各种卫星以及载人航天器相继发射,空间活动日益频繁。作为空间活动的重要环节[1-3],航天器的空间在轨服务旨在保障太空环境的安全与可持续性发展,其内容包括空间装配(如航天器、空间系统或空间结构的在轨连接、构建或

收稿日期: 2023-10-08 **修回日期:** 2023-11-22

基金项目: 国家自然科学基金(62072414);河南省重点研发与推广专项(212102210104, 162102210214)。

通信作者 E-mail: * duanli@bjtu.edu.cn

组装)、空间维护(如航天器的表面修补、部件替换)、空间服务(如失效航天器的回收、空间碎片的捕获)。在实现航天器的空间装配过程中,精准的目标航天器检测与识别是保障装配效率和成功的关键,它为自动化装配提供了保障。在航天器的空间维护中,通过目标航天器的快速识别与定位能够及时发现并解决表面损伤、部件老化等问题从而保障航天器的长期运行。此外,随着越来越多的航天器投入使用,失效航天器和空间碎片也逐渐成为威胁,因此对于失效航天器的回收与空间碎片的捕获也需要高效的目标航天器检测与识别技术的支持,以保持太空环境的清洁与安全。因此,空间在轨服务领域的发展需要不断提升目标航天器检测与识别技术的精度和可靠性,以适应不同的应用场景和任务需求,为人类探索太空和开展空间活动提供更加强有力的保障[4]。然而,传统目标航天器的检测与识别技术存在对遮挡和变形敏感、难以应对复杂背景和噪声等问题。因此,现代目标航天器检测与识别技术正逐渐与深度学习技术相结合,通过大量数据的训练和学习来提高航天器检测与识别在不同场景下的适应性和准确性[5]。

特别是近些年 Transformer[6]在自然语言处理领域取得了显著成果,这引起了研究者们的兴趣并将其逐渐应用于目标检测领域。作为 Transformer 在目标检测领域的开山之作,2020 年 Facebook 团队提出将 Transformer 与卷积神经网络(CNN)相结合的目标检测模型 DETR[7]。这种结合使得 DETR 拥有了 Transformer 的强大建模能力和 CNN 优秀的特征提取能力。相比于单阶段(one-stage)目标检测模型[如 YOLO V1[8]、YOLO V2[9]、YOLO V3[10]、单点多框检测器(SSD)[11]等]或两阶段(two-stage)目标检测模型[如区域卷积神经网络(R-CNN)[12]、快速区域卷积神经网络(Fast R-CNN)[13]、更快速区域卷积神经网络(Faster R-CNN)[14]、掩码区域卷积神经网络(Mask R-CNN)[15]等],DETR 不需要使用非极大值抑制(NMS)算法[16]、选择性搜索算法[17]以及其他启发式先验知识,而是直接通过二分集合匹配和 Transformer 编码器-解码器架构来预测图像中物体的类别和位置,这不但简化了检测流程而且使得 DETR 成为一个端到端的检测模型。然而,DETR 在计算注意力时需要先计算全局像素点,再从中筛选出稀疏的目标像素点,这导致解码器端在查询时需要付出高昂的计算代价,同时使用二分集合匹配会使模型在分配正负样本时将正样本数量降低,从而导致训练过程变慢。为了解决这个问题,ZHU 等[18]提出 Deformable DETR 模型,其通过使用稀疏注意力机制来降低解码器的计算复杂度,并使得其训练速度比 DETR 提升了 10 倍,但 Deformable DETR 使用了多尺度特征来提高检测性能,这导致编码器端的 token 数量比 DETR 增加了 20 倍,进而增加了编码器的计算复杂度。随后的研究表明,当进一步对 Deformable DETR 中编码器端的 token 进行稀疏化时,可以显著降低编码器端的计算复杂度。因此,Sparse DETR[19]模型应运而生。Sparse DETR 通过对密集 token 进行稀疏采样并保留与显著特征相关的 token,从而减少计算量。但由于编码器端的稀疏 token 使得模型在二分匹配过程中可能出现不稳定性现象,因此造成模型收敛速度降低并影响模型的预测精度。同时,在解码器端进行查询预测时,较少的 token 会减少解码器获取的输入信息,进而影响模型性能及降低模型预测精度。

基于以上分析和发现,本文提出空间在轨服务中基于残差自注意力(RS)和分离集合匹配(SSM)的高效端到端航天器组件检测模型 RSSSM-DETR。该模型的优势为:1)在 Sparse DETR 框架的基础上,通过引入残差注意力机制来增加编码器端输出的有效 token 数量,以解决 Sparse DETR 模型中稀疏 token 导致的收敛速度降低并影响模型预测精度的问题;2)在 Sparse DETR 模型后处理阶段,利用分离集合匹配代替传统的二分集合匹配,通过将一对一集合匹配和一对多集合匹配的优点进行并行操作来解决二分集合匹配会使正样本数量降低从而造成训练过程缓慢的问题。

1 RSSSM-DETR 模型

如图 1 所示,本文提出的 RSSSM-DETR 模型主要包括 3 个模块:1)Sparse DETR 模块,这是目标检测的基础模块,选用 Sparse DETR 的原因是希望通过稀疏操作来降低计算复杂度以提高模型收敛效率;2)残差自注意力模块,在 Sparse DETR 的基础上引入残差自注意力机制,目的是解决编码器稀疏化后模型精度降低的问题,通过增加编码器端输出的有效 token 来提高模型收敛的效率以及模型的泛化能力,使其更适用于大规模的目标检测任务;3)分离集合匹配模块,目的是解决在稀疏环境中二分匹配不稳定导致的精度降低以及模型收敛速度变慢等问题。

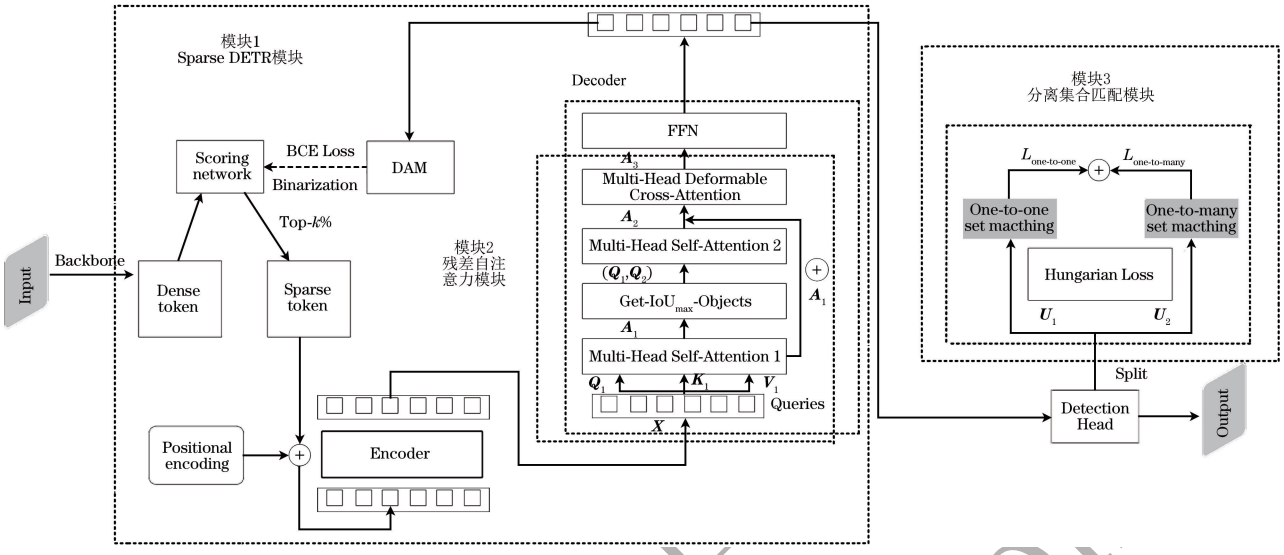


图 1 RSSSM-DETR 模型整体框架

Fig. 1 Overall framework of RSSSM-DETR model

1.1 Sparse DETR 模块

传统的目标检测模型往往存在计算量大、运算速度慢等问题，因此本文选用 Sparse DETR[18] 作为基础框架用于目标检测，Sparse DETR 通过使用评分网络（Scoring network）将含有大量语义信息的 token 进行稀疏选择，可以有效减少编码器端的计算复杂度，提高模型的正向推理速度。具体步骤如下：

1）先将输入图像通过骨干网络得到特征图，再将特征图转化为 token。

2）使用 Scoring network 对所有的输入 token 进行评分，即通过累加解码器中每一层的交叉注意力输出值得到解码器的交叉注意力图（DAM），然后对 DAM 使用二值化（Binarization），将二值化的 DAM 作用于 Scoring network 从而筛选出显著性较高的 token，在二值化的过程中用二进制交叉熵（BCE）Loss 进行监督。

3）将评分网络输出的前百分之 k 个 token（Top-k%）加上位置编码（Positional encoding）后传递至编码器进行编码。

1.2 残差自注意力模块

Sparse DETR 模型经过稀疏化操作后，由于稀疏化减少了 token 数量会使得编码器在对 token 进行特征提取、编码和捕捉相关信息时出现因信息不足而导致的内容缺失及信息不准确等问题。针对这一情况，在原有 Sparse DETR 模块的基础上添加了残差自注意力模块。残差自注意力模块中包含 Get-IoU$_{max}$-Objects 方法，该方法可以找出与原查询 $\boldsymbol{Q}_1$ 的预测框交并比（IoU）最接近的查询 $\boldsymbol{Q}_2$，通过将 $\boldsymbol{Q}_1$、$\boldsymbol{Q}_2$ 融合为（$\boldsymbol{Q}_1$,$\boldsymbol{Q}_2$）来变相提高编码器端得到的 token 质量，同时也增加了 token 数量，即原查询 $\boldsymbol{Q}_1$ 对应的 token 和（$\boldsymbol{Q}_1$,$\boldsymbol{Q}_2$）对应的 token。

在解码器端，当对输入的低质量稀疏化 token 进行预测时，会因上下文信息缺失导致预测准确率下降和模型性能下降等问题。此外，目标不同程度的遮挡也会导致模型预测精度下降。针对这一问题，使用残差连接的方式在解码器端将两个查询 $\boldsymbol{Q}_1$ 和（$\boldsymbol{Q}_1$,$\boldsymbol{Q}_2$）的自注意力机制输出值进行融合，并输出得到最终的交叉注意力机制输出值 $\boldsymbol{A}_3$，此机制不仅使模型获得了更好的学习输入细节和原始表示，减轻了梯度消失的问题，而且提高了被遮挡目标的检测效果。具体步骤如下：

1）Queries：首先自定义可学习权重矩阵 $\{\boldsymbol{W}^q,\boldsymbol{W}^k,\boldsymbol{W}^v\}$，$\boldsymbol{W}$、$\boldsymbol{q}$、$\boldsymbol{k}$、$\boldsymbol{v}$ 均为随机参数；然后将 Sparse DETR 模块的输出序列 $\boldsymbol{X}$ 投影到该权重矩阵上得到三元组 $\boldsymbol{Q}_1=\boldsymbol{X}\times\boldsymbol{W}^q$、$\boldsymbol{K}_1=\boldsymbol{X}\times\boldsymbol{W}^k$、$\boldsymbol{V}_1=\boldsymbol{X}\times\boldsymbol{W}^v$，$\boldsymbol{Q}_1$ 即为原查询。

2）Multi-Head Self-Attention 1：首先计算 $\boldsymbol{Q}_1$、$\boldsymbol{K}_1$ 的关联性并进行归一化处理，然后乘以 $\boldsymbol{V}_1$ 权重系数得到原查询 $\boldsymbol{Q}_1$ 的自注意力值 $\boldsymbol{A}_1$。具体计算如式（1）所示：

$$\boldsymbol{A}_1=\text{Attention}(\boldsymbol{Q}_1,\boldsymbol{K}_1,\boldsymbol{V}_1)=\text{Softmax}\left(\frac{\boldsymbol{Q}_1\boldsymbol{K}_1^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}_1 \tag{1}$$

式中：d_k 是 $\boldsymbol{Q}_1$、$\boldsymbol{K}_1$ 矩阵的列数，即向量的维度。

3）Get-IoU$_{max}$-Objects：首先找出 $\boldsymbol{Q}_1$ 的所有预测框集合，然后从中选取概率最大的预测框 a，接着找到与 a 预测框 IoU 最接近的 b 预测框，最后通过索引找到 b 对应的查询 $\boldsymbol{Q}_2$，$\boldsymbol{Q}_2$ 为与原查询 $\boldsymbol{Q}_1$ 的预

测框 IoU 最接近的查询。

4) Multi-Head Self-Attention 2：将 $\boldsymbol{Q}_1$、$\boldsymbol{Q}_2$、$\boldsymbol{K}_1$、$\boldsymbol{K}_2$、$\boldsymbol{V}_1$、$\boldsymbol{V}_2$ 进行拼接操作后传入自注意力机制中得到输出 $\boldsymbol{A}_2$。具体计算如式(2)所示：

$$\boldsymbol{A}_2 = \text{Attention}((\boldsymbol{Q}_1,\boldsymbol{Q}_2),(\boldsymbol{K}_1,\boldsymbol{K}_2),(\boldsymbol{V}_1,\boldsymbol{V}_2)) = \text{Softmax}\left(\frac{(\boldsymbol{Q}_1,\boldsymbol{Q}_2)(\boldsymbol{K}_1,\boldsymbol{K}_2)}{\sqrt{d_k}}\right)(\boldsymbol{V}_1,\boldsymbol{V}_2) \tag{2}$$

5) Multi-Head Deformable Cross-Attention：首先通过残差连接的方式将 $\boldsymbol{A}_1$、$\boldsymbol{A}_2$ 进行相加，然后通过交叉注意力机制得到输出 $\boldsymbol{A}_3$。

1.3 分离集合匹配模块

Sparse DETR 模型将数据从解码器端输出后采用一对一集合匹配计算总损失值，一对一集合匹配是成功消除后处理操作(如非极大值抑制)的关键设计，正是因为这种设计使得 Sparse DETR 成为真正意义上端到端的检测，但由于一对一集合匹配中被判定为正样本的查询数量太少，因此正样本的训练效率降低且模型收敛速度变得缓慢。一对多集合匹配能提高模型的性能以及训练收敛的速度，但一对多集合匹配的一个不良影响是产生了重复的预测，需要手工添加后处理操作将重复的预测去除，因此会产生较大的计算开销。基于上述原因，本文提出一种分离集合匹配机制，该机制利用一对一集合匹配和一对多集合匹配的优点进行并行操作，然后将输出值相加得到最终损失值。通过结合这两种匹配算法的输出融合操作，可以充分发挥它们各自的优势，从而提高匹配的准确性、鲁棒性。此外，结合一对一集合匹配和一对多集合匹配可以减少不必要的计算和匹配的搜索空间，从而提高匹配的效率，加速模型的收敛速度。具体步骤如下：

1)将编码器端的输出分离为 $\boldsymbol{U}_1$、$\boldsymbol{U}_2$。

2)对 $\boldsymbol{U}_1$、$\boldsymbol{U}_2$ 分别使用匈牙利(Hungarian)算法进行一对一集合匹配和一对多集合匹配后得到输出 $L_{\text{one-to-one}}$、$L_{\text{one-to-many}}$，Hungarian 算法为预测框与真实框之间建立最优的一对一匹配关系，避免了非极大值抑制的操作，使得模型可以准确学习检测任务并提高性能。

一对一集合匹配损失值如式(3)所示：

$$L_{\text{one-to-one}} = \sum_{i=1}^{I} \text{Hungarian}(\boldsymbol{A}) \tag{3}$$

式中：$\boldsymbol{A} = (M_1^i, G_{\text{T}})$，$M_1^i$ 代表由 $\boldsymbol{U}_1$ 在第 i 层编码器所预测的输出，G_{T} 代表只有唯一的一组标注信息集合 $\{G_{\text{T}}\}$；I 为编码器总层数。

由于 G_{T} 的数量唯一，因此可以进行一对一集合匹配。

一对多集合匹配损失值如式(4)所示：

$$L_{\text{one-to-many}} = \sum_{i=1}^{I} \text{Hungarian}(\overline{\boldsymbol{A}}) \tag{4}$$

式中：$\overline{\boldsymbol{A}} = (M_2^i, \overline{G}_{\text{T}})$，$M_2^i$ 代表由 $\boldsymbol{U}_2$ 在第 i 层编码器所预测的结果，$\overline{G}_{\text{T}}$ 代表多组标注信息集合，即为 $\{G_{\text{T},1}, G_{\text{T},2}, \cdots, G_{\text{T},n}\}$，$G_{\text{T},1}, G_{\text{T},2}, \cdots, G_{\text{T},n}$ 由 G_{T} 复制得到，n 为查询数量。

由于 G_{T} 数量与当前查询数量相同，因此可以进行一对多集合匹配。

3)将 $L_{\text{one-to-one}}$ 与 $L_{\text{one-to-many}}$ 进行相加，得到最终损失值。

2 实验对比

2.1 实验平台及实验数据

实验硬件环境配置为：CPU 14 核 CPU Intel® Xeon® Gold 6330 CPU @ 2.00 GHz，GPU NVDIA GeForce RTX 3090 24 GB，内存 80 GB。软件环境配置为：操作系统 Ubuntu 5.4.0，IDE PyCharm 2020.1 专业版。

实验数据集使用包含检测、分割与部件识别的航天器数据集 SDDSP[1]，该数据集共有 3 117 张卫星和空间站图像，分辨率统一设置为 1 280×720 像素。通过标注工具 LabelImg 进行手动标注得到 3 667 个航天器主体、7 334 个太阳帆板，共 11 001 个检测目标并且将整个数据集以接近 6∶4的比例随机划分为 2 516 张训练集图像和 600 张验证集图像。

2.2 实验结果及分析

2.2.1 评价指标

实验评价指标使用平均精确率(AP)、$\text{AP}_{0.5}$、$\text{AP}_{0.75}$、AP_{L}、平均召回率(AR)，其中，AP 指预测框和标注框之间 IoU 从 0.5 开始，每间隔 0.05 求一次 AP 值，一直取值至 0.95，然后求均值，AP 是实验对比中主要的评价指标。AP 在图像中的具体表示是精确率(Precision)-召回率(Recall)曲线下的面积。精确率的计算公式如下：

$$P = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \tag{5}$$

式中：P 表示精确率；N_{TP} 指预测框和标注框之间 IoU 大于 0.5 时的预测框数量；N_{FP} 指预测框和标注框之间 IoU 小于 0.5 时的预测框数量。

召回率的计算公式如下：

$$R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \tag{6}$$

式中：R 表示召回率；N_{FN} 指模型没有检测到标注框，即预测框和标注框之间 IoU 为 0 时的预测框数量。

$AP_{0.5}$、$AP_{0.75}$ 指预测框和标注框之间 IoU 阈值为 0.5、0.75 时的 AP 值。AP_L 指当面积大于 96×96 像素时的 AP 值，具体是指分割掩码中的像素数量。AR 指每个图像中检测到固定数量的最大召回率，在类别和 IoU 上取平均值。

2.2.2 实验参数确定

本节将详细介绍实验所采用的实验参数及其选择依据，并且通过此过程来验证实验假设并推动研究的可重复性。

1)初始学习率。

模型采用学习率衰减策略，如式(7)所示：

$$l = l_{start} \times r^{\frac{E_{now}}{E_{threshold}}} \tag{7}$$

式中：l 为当前学习率；l_{start} 为初始学习率；r 为学习衰减率；E_{now} 为当前迭代次数(Epoch)；$E_{threshold}$ 为学习率衰减阈值。

图 2 是初始学习率分别为 0.001、0.000 1、0.000 01 时的精度对比图。在学习率参数设置过程中，将学习率衰减阈值设置为 40，学习衰减率保持为 0.1，其他实验参数不变。

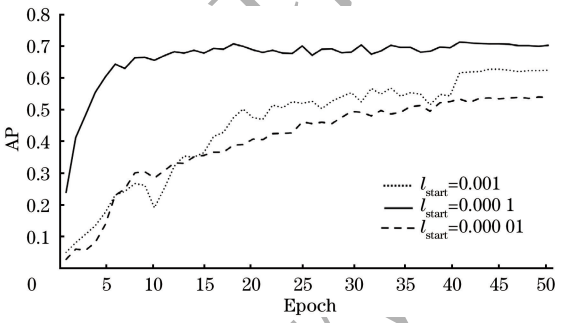


图 2 初始学习率对比结果

Fig.2 Comparative results of initial learning rate

由图 2 可知：当 l_{start}=0.001 时，参数更新步长较长，这将导致模型在训练过程中可能会跳过最优解所在区域，从而使得模型的性能无法达到最佳水平；当 l_{start}=0.000 01 时，参数更新步长较短，从而导致模型训练速度缓慢，需要更多的迭代次数才能收敛到最优解，因此模型无法在有限的时间内达到较高的精确率；当 l_{start}=0.000 1 时，AP 指标明显优于 l_{start}=0.001 或 l_{start}=0.000 01，且模型收敛速度更快。因此，l_{start} 的最佳实验数值为 0.000 1。

2)Epoch 阈值。

图 3 显示了 Epoch 阈值为 50、75、100 时的精度对比图，在 Epoch 阈值设置过程中，首先将 Epoch 阈值分别设置为 50、75、100，然后初始学习率设置为 0.000 1，学习率衰减阈值分别设置为 40、65、90，其他原始实验参数不变。

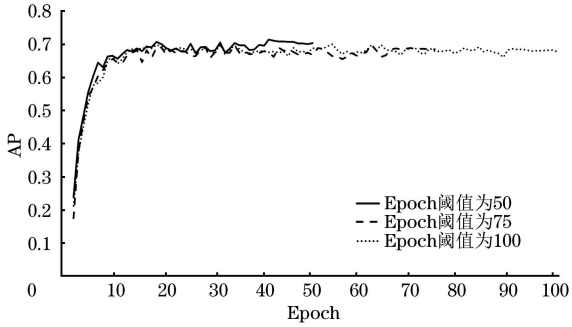


图 3 Epoch 阈值对比结果

Fig.3 Comparative results of Epoch threshold

由图 3 可知：当 Epoch 阈值选取为 50 时，模型预测精确率最高且收敛迅速；当 Epoch 阈值选取为 75、100 时，分别出现了过拟合现象，导致预测精确率降低。因此，Epoch 阈值的最佳实验数值为 50。

3)Dropout。

图 4 是 Dropout 分别为 0.1、0.2、0.3、0.4 时的精度对比图。在 Dropout 设置过程中，将 Dropout 进行改变，分别设置为 0.1、0.2、0.3、0.4，其他原始实验参数不变。

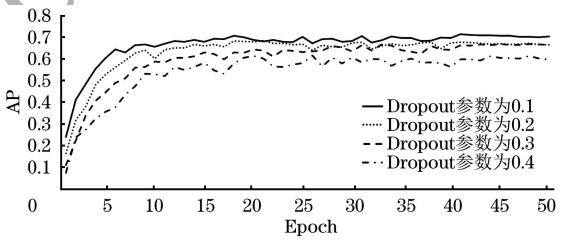


图 4 Dropout 对比结果

Fig.4 Comparative results of Dropout

由图 4 可知：当 Dropout 为 0.2、0.3、0.4 时，模型在训练过程中丢失了过多的神经元，造成信息丢失和模型的不稳定，这是因为模型过拟合训练数据，造成准确率降低；当 Dropout 为 0.1 时，AP 指标均优于其他实验参数。因此，Dropout 的最佳实验数值为 0.1。

4)k 参数。

图 5 是 k 参数分别为 0.3、0.4、0.5、0.6 时的精度对比图。在 Top-k%参数设置过程中，分别将 k 设置为 0.3、0.4、0.5、0.6，其他原始实验参数不变。

由图 5 可知：当 k=0.3、k=0.4 时，由于编码器端经过稀疏化操作后的 token 大幅降低，导致了准确率直接下降，因此与图中的其他曲线相比精确率整体降低；当 k=0.6 时，token 数目增多，但由于含有无用信息的 token 也随之增多，因此会导致性

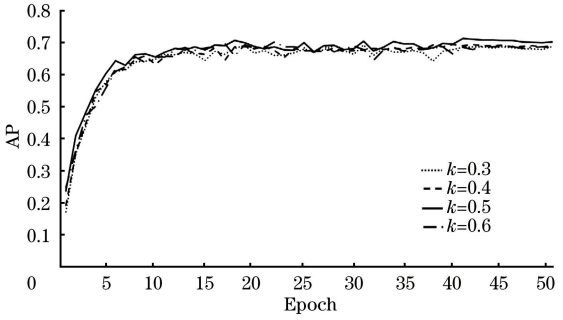


图 5 k 参数对比结果

Fig.5 Comparative results of k parameter

能不稳定及精确率下降；当 $k=0.5$ 时，编码器端 token 数目适中，且精确率保持稳定。因此，k 的最佳实验数值为 0.5。

通过以上对比实验可得，本文实验参数设置为：学习衰减率为 0.1，初始学习率为 0.000 1，当 Epoch 为 0～39 时学习率为 0.000 1，当 Epoch 为 40～49 时学习率为 0.000 01，Dropout 为 0.1，Top-k%中的 $k=0.5$，Epoch 为 50。每迭代一个 Epoch 保存一次模型，最终选取精度最高的模型。

2.2.3 消融实验

通过引入残差自注意力机制和分离集合匹配机制来提高 RSSSM-DETR 模型的性能。对于残差自注意力机制和分离集合匹配机制的不同效果，本节将通过消融实验来验证。由表 1 可知，当 Epoch 为 50 时，由于残差自注意力机制的加入使得编码器中含有大量目标信息的 token 增多，解码器端可以进行有效查询，因此加入残差自注意力机制的 Sparse DETR 中 AP 较 Sparse DETR 增加 1.2 个百分点，AR 增加 0.4 个百分点。在残差自注意力机制后引入分离集合匹配机制，能够将由解码器输出的含有丰富目标语义的查询输入集合匹配，从而进一步输出高精度的预测，即通过一对一集合匹配和一对多集合匹配的使用，可以将输入的查询分离后分别进行匹配，从而达到提高精度的作用。当 Sparse DETR＋RS 模型中加入分离集合匹配机制时，模型整体性能提升，AP 相比 Sparse DETR＋RS 增加 1.9 个百分点，AP_L 增加 2.0 个百分点，$AP_{0.75}$ 增加 0.9 个百分点，AR 增加 0.5 个百分点。

表 1 消融实验对比

Table 1 Comparison of ablation experiments

Model	Epoch	Backbone	AP/%	$AP_{0.5}$/%	$AP_{0.75}$/%	AP_L/%	AR/%
Sparse DETR	50	ResNet-50	68.8	89.9	74.7	70.7	49.5
Sparse DETR＋RS	50	ResNet-50	69.5	90.9	77.1	72.0	49.9
Sparse DETR＋RS＋SSM	50	ResNet-50	71.4	90.7	78.0	74.0	50.4

2.2.4 整体模型对比实验

RSSSM-DETR 模型通过引入稀疏化结构的残差自注意力机制模块使得编码器端中输出的有效 token 增加，进而使得解码器端在查询耦合时可以提取更多有效的检测目标信息，并提高遮挡物体的检测精度。同时，RSSSM-DETR 模型提出的分离集合匹配模块作用在稀疏化结构的残差自注意力机制模块的输出上，可以将分离后的输出同时进行一对一集合匹配和一对多集合匹配，使得模型在保留端到端检测的同时能分配到更多的有效查询。为了检验 RSSSM-DETR 模型的性能，将 RSSSM-DETR 模型与基线模型 DETR 以及其他主流模型 C-DETR[20]、SMCA-DETR[21]、UP-DETR[22]、Deformable DETR、Sparse DETR 在相同的实验环境下分别进行训练，这里需要注意的是 DETR 结构在注意力机制的计算中带来了庞大的计算量导致收敛缓慢，使得当 DETR 训练 50 个 Epoch 时无法收敛，只有当训练 500 个 Epoch 时才完全收敛，当 UP-DETR 训练 50 个 Epoch 时也未完全收敛，但其精确率均已超过 DETR，当训练 300 个 Epoch 时，UP-DETR 才完全收敛。实验结果对比如表 2 所示，其中最优指标值用加粗字体标示。由表 2 可知，当模型训练 50 个 Epoch 时，RSSSM-DETR 在对航天器部件进行目标检测时效果明显优于 DETR、C-DETR、SMCA-DETR、UP-DETR、Deformable DETR、Sparse DETR，其 AP 对比 Sparse DETR 提升 3.1 个百分点，$AP_{0.75}$ 提升 3.3 个百分点，AP_L 提升 3.3 个百分点，AR 提升 0.9 个百分点。当将其与基线模型 DETR 进行对比时，由于 DETR 在编码器端对输入的 token 并未做稀疏化操作，因此在编码时需要付出高昂的计算代价，且会将大量输出后的低质量 token 也一并传入解码器端，导致模型训练收敛速度降低及预测精度降低，RSSSM-DETR 中使用稀疏化 token 作为编码器的输入，并使用残差自注意力结构来确保编码器中输出的高质量 token，解码器端使用分离集合匹配结构使得模型在样本分配的过程中保持稳定。因此，RSSSM-DETR 相比于训练 500 个 Epoch 才收敛的 DETR 在各项指标上均有大幅提升，AP 值增长 17.9 个百分点且收敛速度提升了 10 倍；相比于基于无监督预训练的 UP-DETR，收敛速度提升了 6 倍，AP 值增长 5.3 个百分点，$AP_{0.5}$ 增长

1.1 个百分点，$AP_{0.75}$ 增长 4.0 个百分点，AR 增长了 1.2 个百分点。

综上。RSSSM-DETR 模型在各方面均优于对比模型，因此验证了 RSSSM-DETR 模型的可行性。

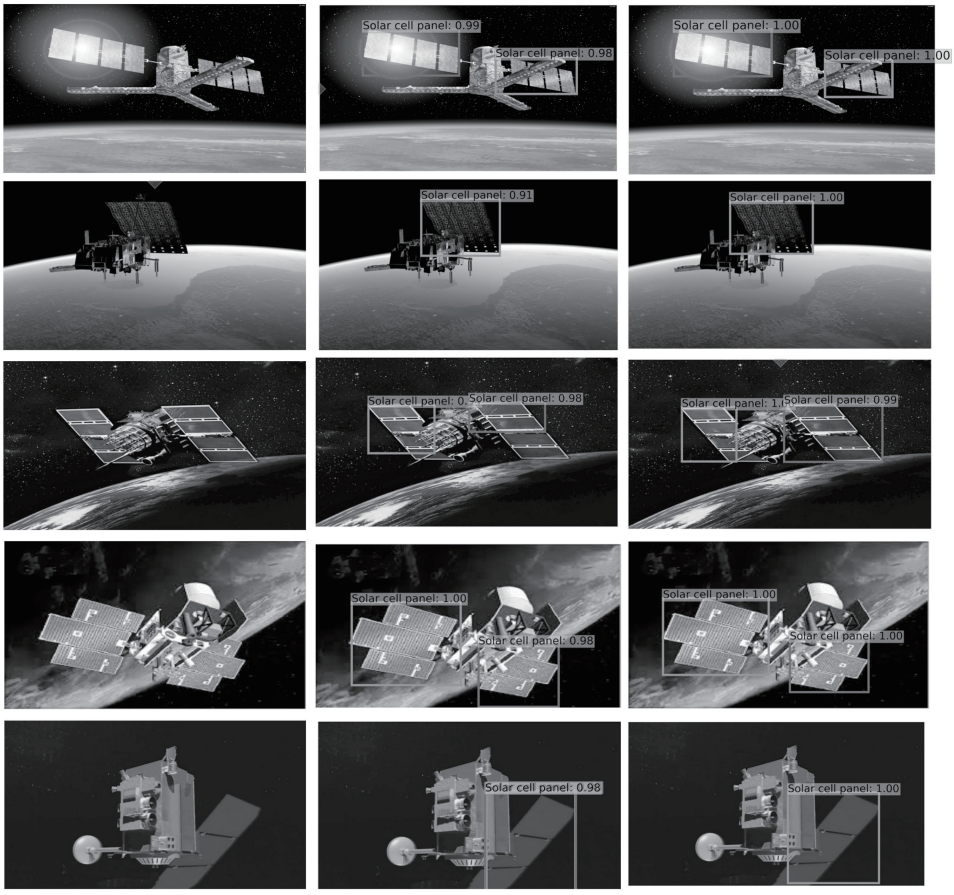
表 2 整体模型实验对比

Table 2 Comparison of overall model experiments

Model	Epoch	Backbone	AP/%	$AP_{0.5}$/%	$AP_{0.75}$/%	AP_L/%	AR/%
DETR	500	ResNet-50	53.5	83.9	58.9	57.3	43.0
UP-DETR	50	ResNet-50	58.7	88.3	65.3	63.2	46.0
UP-DETR	300	ResNet-50	66.1	89.6	74.0	70.0	49.2
C-DETR	50	ResNet-50	50.7	83.6	55.5	54.8	40.2
SMCA-DETR	50	ResNet-50	52.2	83.1	57.8	56.2	41.5
Deformable DETR	50	ResNet-50	67.8	88.6	74.4	70.8	49.4
Sparse DETR	50	ResNet-50	68.3	89.9	74.7	70.7	49.5
RSSSM-DETR	50	ResNet-50	**71.4**	**90.7**	**78.0**	**74.0**	**50.4**

2.2.5 目标遮挡对比实验

当航天器处于特定角度时，会使太阳板被遮挡。本节主要是对遮挡的太阳板进行检测来验证 RSSSM-DETR 模型的性能。首先，从数据集的验证集中随机选择 5 张图片作为本节实验的样本。然后，将其分别送入已训练好的 Sparse DETR 模型和 RSSSM-DETR 模型，通过输出结果的类别概率和定位信息可以看出 RSSSM-DETR 模型在目标定位效果上优于 Sparse DETR 模型，预测框坐标定位更加精确且类别概率也得到了提高。具体的实验结果对比如图 6 所示。由图 6 可知，当太阳板被遮挡时，Sparse DETR 模型预测的类别判断基本正确且类别概率相对较高，但 token 在编码器端经过稀疏化操作后，含有有效语义信息的 token 数量减少，导致了预测框的坐标定位精度降低。对于上述问题，在 RSSSM-DETR模型中通过加入残差自注意力机制



(a)原始图片 (b)Sparse DETR检测结果 (c)RSSSM-DETR检测结果

图 6 航天器组件遮挡对比结果

Fig.6 Comparative results of spacecraft component occlusion

使得编码器端输出的有效 token 数量增多，因此解码器端可以获得更多的上下文信息，从而增强被遮挡物体的坐标定位能力，同时也使得类别概率得到了提升。

2.2.6 训练过程分析

本节基于 SDDSP 数据集进行训练，共训练 50 个 Epoch，图 7 是 RSSSM-DETR 模型与 Sparse DETR 模型的 AP 变化对比曲线图，其中，横轴为 Epoch，纵轴为训练时的 AP。由图 7 可知，在 AP 变化曲线中，当 Epoch 为 40 时，RSSSM-DETR 模型收敛并达到全局峰值，且 AP 曲线变化平稳。然而，此时 Sparse DETR 模型的 AP 值仍在提高直至本次训练结束。上述结果证明了 RSSSM-DETR 模型加快了收敛速度。

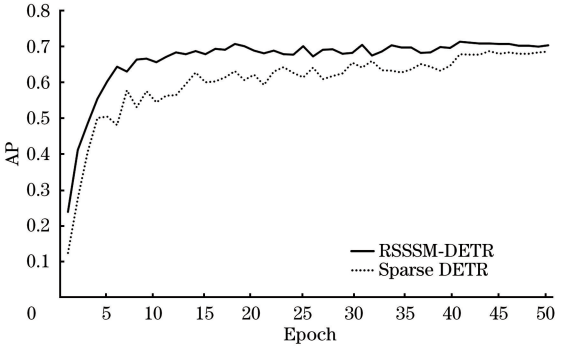


图 7 AP 变化对比曲线

Fig. 7 Comparative curves of AP change

图 8 是 Loss 变化对比曲线图。由图 8 可知：RSSSM-DETR 模型在前 40 个 Epoch 时，Loss 稳定下降，模型迅速收敛，且 Loss 低于 Sparse DETR；在 40 个 Epoch 后，Loss 趋于稳定，模型完成收敛，而此时 Sparse DETR 的 Loss 仍在下降，而 RSSSM-DETR 模型整体收敛过程稳定并且无过拟合或欠拟合现象，训练结果较理想且在 40 个 Epoch 时完成模型收敛，相比 Sparse DETR 收敛速度加快 20%。

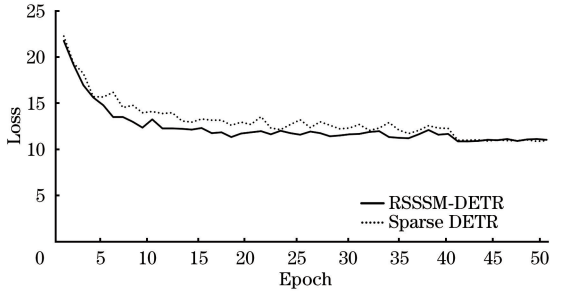


图 8 Loss 变化对比曲线

Fig. 8 Comparative curves of Loss change

3 结束语

针对空间在轨服务中目标航天器检测与识别问题，本文提出一种高效端到端航天器组件检测模型。该模型通过在 Transformer 中引入残差自注意力机制及分离集合匹配机制来解决由于稀疏化 token 所引起的精度下降及匹配不稳定等问题，并且有效地提高了模型的预测精度和收敛速度。实验结果证明，RSSSM-DETR 模型相比当前主流模型具有较好的检测效果，在平均精确率和收敛速度上比基线模型 DETR 提高了 17.9 个百分点和 10 倍，比 Sparse DETR 提高了 3.1 个百分点和 20%。但是，当 RSSSM-DETR 模型由于拍摄角度而出现部件严重遮挡时，会导致检测精度下降或者误检。因此，后续将研究生成式辅助模型，通过生成式辅助模型来生成遮挡部分的目标图像，以增强目标完整性，从而降低遮挡对模型性能的影响，进而提升模型对于部件严重遮挡情况的鲁棒性。

参考文献

[1] DUNG H A, CHEN B, CHIN T J. A spacecraft dataset for detection, segmentation and parts recognition [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Washington D.C., USA: IEEE Press, 2021: 2012-2019.

[2] FLORES-ABAD A, MA O, PHAM K, et al. A review of space robotics technologies for on-orbit servicing [J]. Progress in Aerospace Sciences, 2014, 68: 1-26.

[3] 逄晨曦，李文辉. 基于注意力改进的自适应空间特征融合目标检测算法[J]. 吉林大学学报(理学版)，2023，61(3)：557-566.
PANG C X, LI W H. Adaptive spatial feature fusion object detection algorithm based on attention improvement [J]. Journal of Jilin University (Science Edition), 2023, 61(3): 557-566. (in Chinese)

[4] KANG G H, ZHANG Q, WU J Q, et al. Pose estimation of a non-cooperative spacecraft without the detection and recognition of point cloud features[J]. Acta Astronautica, 2021, 179: 569-580.

[5] KOTHARI V, LIBERIS E, LANE N D. The final frontier: deep learning in space [C]// Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications. New York, USA: ACM Press, 2020: 45-49.

[6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. [2023-09-05]. http://arxiv.org/abs/1706.03762.

[7] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer International Publishing, 2020: 213-229.

[8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington D.C., USA: IEEE Press, 2016: 779-788.

[9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington D.C., USA: IEEE Press, 2017: 6517-6525.

[10] REDMON J, FARHADI A. YOLOv3: an incremental improvement [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D.C., USA: IEEE Press, 2018: 1-6.

[11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single

shot multibox detector[M]. Berlin, Germany: Springer International Publishing, 2016.
[12] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(1): 142-158.
[13] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2015: 1440-1448.
[14] 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.
[15] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.
[16] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
[17] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.
[18] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable Transformers for end-to-end object detection[EB/OL]. [2023-09-05]. http://arxiv.org/abs/2010.04159.
[19] ROH B, SHIN J, SHIN W, et al. Sparse DETR: efficient end-to-end object detection with learnable sparsity[EB/OL]. [2023-09-05]. http://arxiv.org/abs/2111.14330.
[20] MENG D P, CHEN X K, FAN Z J, et al. Conditional DETR for fast training convergence[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2021: 3631-3640.
[21] GAO P, ZHENG M H, WANG X G, et al. Fast convergence of DETR with spatially modulated co-attention[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2021: 3601-3610.
[22] DAI Z G, CAI B L, LIN Y G, et al. UP-DETR: unsupervised pre-training for object detection with Transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2021: 1601-1610.

编辑 陆燕菲