

# 基于 ISAR 成像的非合作航天器位姿识别网络<sup>①</sup>



谭启麟<sup>1</sup>, 吴 珏<sup>1</sup>, 李志辉<sup>2</sup>, 杨 雷<sup>1</sup>

<sup>1</sup>(西南科技大学 计算机科学与技术学院, 绵阳 621010)

<sup>2</sup>(北京流体动力科学研究中心, 北京 100191)

通信作者: 谭启麟, E-mail: [354077334@qq.com](mailto:354077334@qq.com)

**摘 要:** 非合作航天器缺乏合作信息, 无法直接利用传感器获得位姿数据, 提出一种基于 ISAR 图像的位姿识别网络. 相比于空间摄影卫星拍摄的图像以及仿真数据, 该图像更易获取、成本更低, 但存在分辨率低、面板成像不完整等问题. 因此, 该网络在图像预处理时, 通过对 YOLOX-tiny 的调整, 将其作为航天器裁剪网络, 避免图像中标记的数据影响后续网络的训练, 使网络仅关注航天器所在区域. 利用增强的 Lee 滤波滤除图像噪声, 提升图像的质量. 在骨干网络中, 加入 STN 模块, 使网络选择最相关的区域注意, 将 U-Net 网络设计成密集残差块结构并结合 CBAM 模块, 减少下采样期间的特征损失, 提高模型的准确性. 此外, 引入了多头自注意力来捕获更多的全局信息. 实验结果表明, 该模型最小、最大、平均误差较于目前的一些主流模型均有所提升, 误差缩小了 0.5-0.6, 从而证明该网络具有更好的位姿识别能力.

**关键词:** 非合作航天器; 逆合成孔径雷达; 位姿识别; 注意力机制; 卷积神经网络

引用格式: 谭启麟, 吴珏, 李志辉, 杨雷. 基于 ISAR 成像的非合作航天器位姿识别网络. 计算机系统应用, 2024, 33(10): 47-55. <http://www.c-s-a.org.cn/1003-3254/9670.html>

## Non-cooperative Spacecraft Pose Recognition Network Based on ISAR Imaging

TAN Qi-Lin<sup>1</sup>, WU Jue<sup>1</sup>, LI Zhi-Hui<sup>2</sup>, YANG Lei<sup>1</sup>

<sup>1</sup>(School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China)

<sup>2</sup>(Beijing Aerohydrodynamic Frontier Research Center, Beijing 100191, China)

**Abstract:** Due to the lack of cooperative information, non-cooperative spacecraft cannot obtain pose data directly from sensors. Therefore, a pose recognition network based on inverse synthetic aperture radar (ISAR) images is proposed. Compared with the images taken by space photography satellites and simulation data, this kind of image is easier to obtain and cheaper, but there are some problems such as low resolution ratio and incomplete panel image. Therefore, in image preprocessing, the network uses YOLOX-tiny as a spacecraft clipping network by adjusting it to avoid the data marked in the image affecting the subsequent network training, so that the network only focuses on the region where the spacecraft is located. The enhanced Lee filter is used to remove image noise and improve image quality. In the backbone network, the STN module is added to make the network select the most relevant region attention, and the U-Net is designed into a dense residual block structure and combined with the CBAM module to reduce the feature loss during sampling and improve the accuracy of the model. In addition, multi-head self-attention is introduced to capture more global information. The experimental results show that the minimum, maximum, and average errors of this model are improved compared with some mainstream models, and the errors are reduced by 0.5-0.6. All this proves that the network has a better pose recognition ability.

① 基金项目: 国家重点研发计划 (2022YFA1005104)

收稿时间: 2024-03-30; 修改时间: 2024-04-23; 采用时间: 2024-05-29; csa 在线出版时间: 2024-08-28

CNKI 网络首发时间: 2024-08-29

**Key words:** non-cooperative spacecraft; inverse synthetic aperture radar (ISAR); pose recognition; attention mechanism; convolutional neural network (CNN)

航天技术既需要各领域前沿科技的支撑,同时也可以促进这些技术的发展,是科学技术最前沿的领域之一. 独立的航天技术是衡量一个国家科技实力和综合国力的重要标志<sup>[1]</sup>,关系到国家的战略地位. 随着航天技术的进步以及航天任务的日益复杂,对非合作目标实时的位姿识别在航空航天领域发挥着至关重要的作用. 目前如何准确地识别非合作航天器的位姿基本上已经成为航空航天领域的议程<sup>[2]</sup>. 搭建一个准确的非合作航天器位姿识别模型不仅能够提高效率,也使得相关人员能够及时获取到位姿信息. 同时也对后续航天器交会对接<sup>[3]</sup>、在轨维修<sup>[4]</sup>、太空碎片清理<sup>[5]</sup>和废弃航天器的重返等太空任务中有着广泛的应用.

过去,航天器的位姿通常是通过高精度的传感器来估计的<sup>[6,7]</sup>,这些传感器有着较高的成本和功耗. 而非合作航天器无法直接利用传感器获取其位姿数据,只能通过额外的观察来完成. 目前,对非合作航天器的位姿识别的数据来源大致可分为两类<sup>[8]</sup>,第1类是利用其他专门的空间摄影卫星捕获的图像<sup>[9]</sup>和仿真数据<sup>[10]</sup>. 第2类是利用地面逆合成孔径雷达 (inverse synthetic aperture radar, ISAR) 技术对空间中的航天器进行成像<sup>[11]</sup>,从而无需依赖专门的卫星而仅依赖雷达成像,如图1所示. 受到测量条件以及成像原理等因素的影响,空间目标 ISAR 图像的特征提取相对困难<sup>[12]</sup>,主要表现在 ISAR 图像不同于传统的光学图像,通常更难理解,由于存在斑点噪声的因素,ISAR 图像的质量会出现不同程度的下降. 因此通过 ISAR 图像的位姿识别与地球和卫星图像上的目标位姿识别有较大的差异. 因此,目前主流的位姿估计方法无法直接用于非合作航天器的位姿估计,如 HRNe<sup>[13]</sup>, AggPose<sup>[14]</sup>.

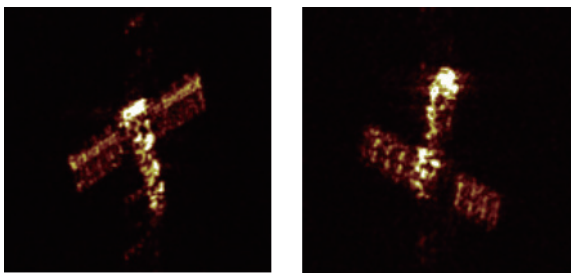


图1 航天器 ISAR 图像

近年来,人工智能尤其是深度学习的飞速发展,使得基于卷积神经网络 (CNN) 的方法能够被用于非合作空间目标的位姿识别,这成为近年来解决该问题的一个有吸引力的方向<sup>[9,15-18]</sup>. 当前基于深度学习的非合作采用仿真数据或者其他卫星拍摄的图像作为模型训练集. 尽管使用模拟数据训练的模型在测试中确实有效,但在实际应用场景中还面临诸多挑战. 利用更容易获取的航天器 ISAR 图像来进行非合作目标姿态估计,为了解决该问题的更好途径. ISAR 是目前获取远距离非合作空间目标结构信息的有效工具. 它通过发射宽带信号获得距离向高分辨,通过合成空间处理获得方位向高分辨<sup>[19]</sup>,从而对目标进行高分辨率成像,能够准确反映目标的结构特征以及散射中心分布,提供更加丰富的目标信息,例如: 航天器、卫星和空间站. ISAR 克服了航空相机、光学摄像机等设备受气候条件限制的障碍<sup>[20]</sup>. 但在成像过程中会产生斑点噪声,导致图像的质量下降,这会使得网络在进行特征提取时遇到困难. Zhang 等人<sup>[21]</sup>设计了一种增强的 Lee 滤波方法 (ELF) 来滤除 ISAR 图像中的斑点噪声,通过幂变换和归一化提高图像对比度增强图像质量. 想要获得大量的航天器 ISAR 图像相对困难,这就使得在训练网络时面临了小样本的问题,再加之图像特征稀疏,简单的语义和固定的结构,还包括轮廓边界模糊、面板图像不完整等特点. Gao 等人<sup>[11]</sup>提出了一种密集残差 U-Net 网络 (SU-Net). 同时, ISAR 图像中存在左上角已有的标记数据的干扰以及航天器在二维图像上的投影带来的图像变形等因素都会对后续航天器的位姿识别产生影响. 受以上工作的启发,并结合其他因素对位姿识别的影响,本文在 SU-Net 网络的结构上进行改进,得到了一种更适合 ISAR 图像中航天器位姿识别的网络结构. 其包含以下几个关键组件. 1) 使用航天器检测网络来裁剪出原始图像中仅存在航天器的固定大小正方形区域; 2) 应用增强的 Lee 滤波方法来滤除图像的斑点噪声,通过幂变换和归一化处理来增强图像对比度,提高图像质量; 3) 采用迁移学习的方法来克服非合作航天器 ISAR 图像数据有限的问题,使用预训练模型进行模型训练; 4) 利用 STN 空间变换模块使网络拥有空间不变

性, 针对图像质量的不同变化可以产生类似的结果, 提高模型鲁棒性; 5) 采用密集残差块 (DRB) 来允许重用先前层的特征图, 缓解了图像特征稀疏的问题, 避免下采样导致的特征丢失; 6) 添加了多头自注意块 (MHSA) 来捕获全局信息, 以及 CBAM 注意力模块突出重要特征. 本文的主要贡献总结如下.

(1) 一种用于 ISAR 图像中航天器检测的网络, 裁剪出非合作航天器的区域.

(2) 利用 STN 模块, 帮助网络拥有空间不变性, 提高模型鲁棒性.

(3) DRB 中融入 CBAM 轻量化注意力模块, 来强调在通道轴和空间轴这两个维度上有意义的特征.

## 1 相关介绍

### 1.1 U-Net

U-Net 网络<sup>[22]</sup>是一种常见的图像分割模型, 结构完全对称分布, 其中浅高层分辨率特征用来解决像素定位问题, 深低层分辨率特征图用来解决像素的分类问题, 并且采用跳跃连接的方式将特征图进行维度拼接, 极大地保留了更多的位置和特征信息. 其构建了一套完整的 encoder (编码器)-decoder (解码器) 结构. 它由一个收缩路径和一个扩张路径组成. 收缩路径用于提取图像上下文信息和高级语义特征, 采用典型的卷积网络结构, 通过池化操作逐步缩小特征图的空间大小, 增加特征通道数, 以提取抽象语义信息. 扩张路径用于精确定位, 它通过上采样和卷积逐步恢复特征图的空间细节, 同时结合来自收缩路径的上下文信息, 实现精确的语义定位. 为了增强 U-Net 网络实用性, 研究人员在后续也对模型进行了相应改进<sup>[20]</sup>. 深度学习的迅速发展使得神经网络的结构越来越复杂, 训练所需时间越来越长. 相比之下, U-Net 网络结构简单, 只需要较小的数据集就能达到比较好的效果. 根据 ISAR 图像中航天器的特点, 选择在 U-Net 网络架构上进行改进, 从而利用 U-Net 架构来实现航天器位姿识别. 同时 U-Net 主干特征提取网络常用 VGG<sup>[23]</sup>和 ResNet<sup>[24]</sup>网络, 能够极大地提高 U-Net 网络的学习能力.

### 1.2 STN 空间变换模块

如果网络能够对经过变换操作的图片得到与未经变换前相同的检测结果, 我们就说这个网络具有空间变换不变性. 卷积神经网络缺乏对输入数据的空间不变性, 会导致模型性能下降. CNN 中的池化层有助于

获得这种特性, 网络能够显式地学习平移不变性, 以及隐式的学习旋转不变性. 但与其让网络隐式的学习到某种能力, 不如其设计一个专门的处理模块, 以处理以上的各种变换. STN<sup>[25]</sup>可以使模型具有空间不变性, 通过学习变换矩阵, 根据变换矩阵找到映射关系最后根据位置映射计算像素值, 使神经网络能够在特征图本身的条件下自动地对特征进行空间变换. 当在网络中添加具有旋转不变性的模块时, 能够使网络更加关注目标所在的区域, 这样能够提取到图像中更多的关键特征信息<sup>[23]</sup>.

### 1.3 密集连接和残差连接

Huang 等人<sup>[26]</sup>提出了一种密集卷积网络, 密集连接将网络的每一层都通过前馈连接与其他每一层直接相连, 使每一层可以直接获取其他前面层的特征信息. 这种连接方式可以避免梯度消失, 并促进特征重用. 为了解决训练深度网络的困难, He 等人<sup>[24]</sup>引入了残差网络. 残差连接将后面层的特征通过残差形式加回前面层, 即前面层的输出为后面层输出的残差. 这种连接方式不仅可以防止梯度消失, 还可以降低模型复杂度以减轻过度拟合. 两者可以相结合形成更强大的网络结构, Wang 等人<sup>[27]</sup>就将残差连接和密集连接结合了起来, 按照特定任务定制了不同的残差网络结构.

### 1.4 注意力机制

注意力机制是一种常用的技术, 其可以捕捉图像或文本序列中的全局信息, 通过学习来自全局的权重或相关性, 定位到感兴趣的信息, 抑制无用信息, 使模型在处理数据时更加准确和有效. 在许多视觉任务中加入注意力机制可以提升模型性能, 如图像分类<sup>[28]</sup>、分割<sup>[29]</sup>、检测<sup>[30]</sup>等. 注意力机制通常是通过加权操作实现的, 即对输入序列中的每个位置进行打分, 得到每个位置的权重系数, 然后与输入特征进行加权结合, 强调重要特征.

## 2 具体方法

如图 2 所示, ISAR 图像的航天器位姿识别网络包括 6 个步骤.

为了检测出光流扰动效应, 设计了算法 1.

算法 1. ISAR 图像的航天器位姿识别网络

- 1) 通过航天器检测网络, 裁剪出仅包含航天器的固定大小正方形区域, 避免标记的数据对后续网络训练的干扰.
- 2) 利用图像增强, 提高图像对比度并减少噪声, 提高图像质量.
- 3) 为使网络具有空间不变性, 引入空间变换模块.

- 4) 通过采用预训练模型提取浅层特征, 克服数据集限制.
- 5) 通过密集残差 U-Net 网络提取深层特征.
- 6) 对航天器位姿结果进行回归, 得到最终结果.

### 2.1 航天器检测网络

通过航天器检测网络来自动找到 ISAR 图像中航天器的位置如图 3 所示. 由于在单张 ISAR 图像中只存

在一个航天器, 因此考虑到更小的模型消耗更少, 选择了 YOLOX-tiny 版本作为该网络, 并修改其中的某些参数值, 使其能够直接根据检测框裁剪固定大小的图片, 为确保边界框能包含整个航天器, 将裁剪图像的大小设置为 150×150 像素. 从而避免左上角标记数据对后续网络训练的干扰.

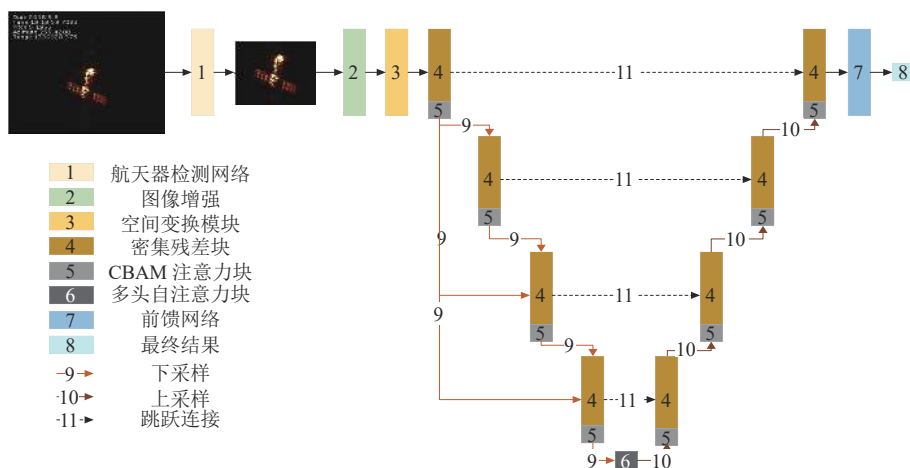


图 2 ISAR 图像的航天器位姿识别网络结构

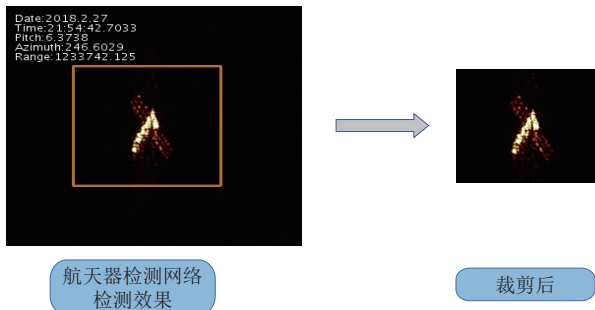


图 3 航天器检测网络裁剪示例

### 2.2 图像预处理

如图 4 所示, 由于 ISAR 图像中的散射单元包含随机分布的小散射点, 这会在图像中产生斑点噪声, 降低图像质量并影响后续识别. 为了提高图像质量, 选择增

强的 Lee 滤波算法来滤除噪声. 其可以有效保留边缘信息并平滑噪声点. 通过处理提高 ISAR 图像质量后, 可以进行更准确的图像识别和分析. ISAR 图像中的斑点噪声通常被认为是乘性噪声模型, 可以表示为:

$$I(t) = x(t) \times u(t) \tag{1}$$

其中,  $I(t)$  为噪声干扰后的图像,  $x(t)$  表示随机地面反射目标的散射截面,  $u(t)$  为噪声.

在 Lee 滤波算法中, 将乘性噪声模型近似为一个线性系统, 然后应用最小均方误差 (MSE) 准则来得到 Lee 滤波的滤波公式:

$$\omega = \bar{I} + k(I - \bar{I}) \tag{2}$$

其中,  $\bar{I}$  是滤波器窗口的平均像素强度,  $\omega$  是受噪声影响的图像估计,  $k$  是加权系数.

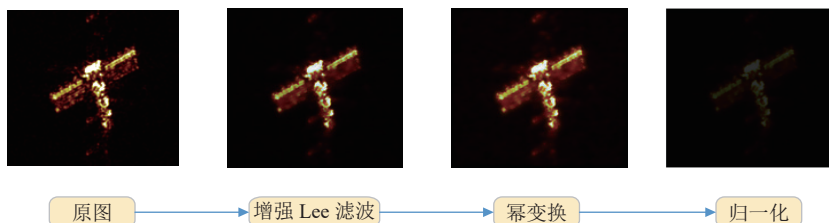


图 4 图像预处理过程

但 Lee 滤波器无法有效过滤未完全形成的斑点噪声. 为提高去噪效果, 改用增强的 Lee 滤波方法, 将 Lee 滤波器的表达式改为:

$$\begin{cases} \omega = \bar{I}, & C_I < C_{\min} \\ \omega = \bar{I} + k(I - \bar{I}), & C_{\min} < C_I < C_{\max} \\ \omega = I, & C_I > C_{\max} \end{cases} \quad (3)$$

其中,  $C_I$  表示图像的局部标准系数偏差,  $C_{\min} = C_n$ ,  $C_{\max} = \sqrt{1 + \frac{2}{L}}$ ,  $L$  表示等效视图数.

增强后的 Lee 滤波算法将图像区域分成 3 类, 每一类采取不同的处理方式, 如下.

$C_I < C_u$  ( $u$  表示噪声) 的区域代表同质区域, 滑动窗口内像素的平均值为该区域的中心值.

$C_u < C_I < C_{\max}$  的区域代表弱纹理区域, 该区域采用传统的 Lee 滤波器.

$C_I > C_{\max}$  的区域代表异质区域, 该区域直接保留原始值.

由于 ISAR 图像的像素值普遍都比较集中, 这就导致了图像的对比度不是很高. 为了提升 ISAR 图像的对比度, 引入了幂变换的方式. 幂变换的方式如下:

$$J(i, j) = [H(i, j)]^\alpha \quad (4)$$

通过调整参数的值就可以调整图像的对比度.

同时, ISAR 图像的散射强度随着距离变化, 会影响图像质量. 采用归一化处理, 通过不变矩找到不变量, 来减弱距离变化对图像质量的影响. 归一化的公式如下:

$$K(i, j) = \frac{J(i, j)}{\sqrt{\sum_i \sum_j |J(i, j)|^2}} \quad (5)$$

其中,  $K(i, j)$  表示归一化后的图像.

### 2.3 STN 空间变换模块

在 ISAR 图像中, 航天器不同位姿的图像差异很大. 即使是同一位姿, 也会由于图像质量的影响使得学习难度加大, 该模块能使模型学习到平移、缩放、旋转和更一般的扭曲不变性, 缩小图像差异带来的影响. 其是一种动态机制, STN 能动态生成转换, 对图像或特征图进行空间变换, 如缩放、剪切、旋转以及非刚性变形. 这使得网络不仅可以选图像中最相关的区域注意, 而且可以将这些区域转换成规范的、预期的位姿. 空间变换模块的结构如图 5 所示, 输入特征映射

$U$  被传递到一个定位网络, 该网络回归转换参数  $\theta$ . 将规则空间网格  $G/V$  转换为采样网格  $T_\theta(G)$ . 将采样网格应用于  $U$ , 产生扭曲的输出特征映射  $V$ . 定位网络和抽样机制的结合定义了一个空间变换块.

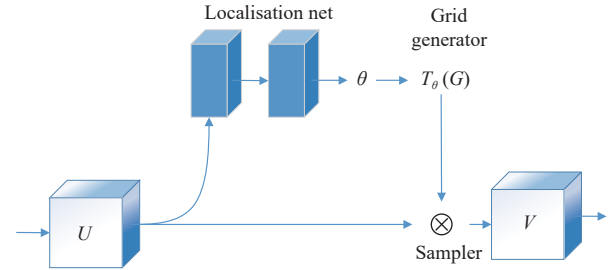


图 5 空间变换模块网络结构

### 2.4 预训练模型

由于已有的航天器 ISAR 图像数据量小、特征稀疏, 如果直接利用其训练模型的话相对困难. 采用迁移学习, 使用预训练模型来提取图像中的浅层特征. 从而加速训练. 预训练模型大多基于 ImageNet 数据集的, 数据集中的图像与航天器 ISAR 图像有较大的差异. 但都是由主要对象的彩色图像组成, 并在训练后调整原始权重. 同时, 为了保留原始 ISAR 图像特征, 将原始航天器 ISAR 图像特征  $X$  添加到增强特征  $X_e$  中, 输入到 ResNet34 预训练模型中, 得到浅层特征  $X_r$ , 如下:

$$X_r = \text{ResNet}(X + X_e) \quad (6)$$

### 2.5 密集残差 U-Net

密集残差网络是基于 U-Net 的网络结构改进的. 如图 2 所示. 在其中加入了 3 个关键组件: (1) 密集残差块 (DRB). (2) 一种轻量的注意力模块 CBAM. (3) 多头自注意力块 (MHSAB). 向网络中添加  $X$ 、 $X_e$ 、 $X_r$ .

令  $XE_i$  为编码器部分中第  $i$  ( $i=1, 2, 3, 4$ ) 层的输出特征,  $XD_i$  为解码器中每一层对应的输入特征. 编码器输出特征与对应解码器层特征融合.

$$\begin{cases} XE_0 = \text{STN}(X + X_e + X_r) \\ XE_1 = \text{CBAM}(\text{DRB}(XE_0)) \\ XE_i = \text{CBAM}(\text{DRB}(\text{DOWN}(XE_i)) + XE_{i-1}) \\ XD_4 = XE_4 + \text{UP}(\text{MHSAB}(XE_4)) \\ XD_i = XE_i + \text{UP}(XD_{i+1}) \end{cases} \quad (7)$$

其中,  $\text{STN}(\cdot)$  表示空间变换模块,  $\text{DRB}(\cdot)$  表示密集残差块,  $\text{CBAM}(\cdot)$  表示空间和通道上进行注意力机制,  $\text{MHSAB}(\cdot)$  表示多头自注意力块,  $\text{UP}(\cdot)$  表示上采样操作. 最后通过该网络获得深层特征  $X_d$ , 如下所示:

$$X_d = \text{DRB}(XD_1) \quad (8)$$

其中, 第1块编码器的输出被传递到所有后续编码器, 这种方法能够减少特征损失并提高模型的整体性能. 并且在每一块的编码器后加入 CBAM 注意力块, 及时调整感兴趣的信息. 这样能够网络在接下来的每一层捕获到更有意义的特征.

密集残差块是多个残差块的密集连接, 每个残差块中包含归一层、卷积、简单门和残差连接, 如图5所示. 给定输入特征为  $F$ , 通过 DRB 提取特征:

$$\begin{cases} F_d = \text{RB}(\text{RB}(F) + F) + \text{RB}(F) + F \\ \text{RB}(F) = F + C_3(\text{SG}(C_2(C_1(\text{LN}(F)))))) \\ \text{SG}(F) = F_{f1} \times F_{f2} \end{cases} \quad (9)$$

其中,  $C_1$ 和 $C_3$ 是  $1 \times 1$  的卷积,  $C_2$ 是  $3 \times 3$  的深度卷积, LN表示归一化层,  $F_{f1}$ 、 $F_{f2}$ 是将  $F$ 划分为通道维度得到的. 残差连接使位姿识别模块能够通过不同块之间的短连接来聚合各个级别的特征. 通过密集连接减少传输过程中的特征损失, 减轻由于航天器 ISAR 图像特点的稀疏性所带来的训练难度. 引入简单的门控机制, 该机制使用乘法而不是, 非线性激活函数, 使其更适合处理航天器 ISAR 图像的任务.

### 2.6 CBAM 注意力机制

Woo 等人<sup>[28]</sup>在 2018 年提出一种新的卷积注意力模块, 创新地提出了通道注意力与空间注意力融合的注意力机制, 如图6所示, 对前馈卷积神经网络是一个简单而有效的注意力模块. 卷积运算通过将跨通道信息和空间信息混合在一起来提取信息特征, 因此提出 CBAM 来强调通道轴和空间轴这两个维度上的有意义特征. 输入的特征图为  $F \in \mathbb{R}^{C \times H \times W}$ , 利用 CBAM 可以依次推导出二维通道注意力图  $M_C \in \mathbb{R}^{C \times 1 \times 1}$  和二维空间注意力图  $M_S \in \mathbb{R}^{1 \times H \times W}$ , 总的注意过程可以概括为:

$$\begin{cases} F' = M_C(F) \otimes F \\ F'' = M_S(F') \otimes F' \end{cases} \quad (10)$$

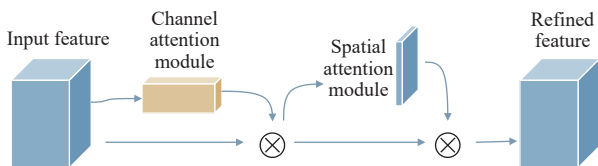


图6 CBAM 注意力块结构

在通道注意力中关注的是哪个通道上的特征是有意义的. 输入的特征图是  $H \times W \times C$ , 先分别进行一个全

局平均池化和全局最大池化得到两个  $1 \times 1 \times C$  的特征图, 将两个特征图分别送入两层的全神经网络, 对于这两个特征图, 这个两层的全连接神经网络是共享参数的. 然后将两个特征图相加, 通过 Sigmoid 函数得到 0-1 之间的权重系数, 将权重系数与输入的特征图相乘, 从而得到最终输出的特征图.

$$M_C(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (11)$$

通道注意力输出之后, 再引入空间注意力模块, 关注空间中哪部分的特征有意义. 输入的特征图分别进行一个通道维度的最大池化和平均池化得到两个  $H \times W \times 1$  的特征图. 然后将两个特征图拼接起来, 将  $H \times W \times 2$  特征图经过一个卷积层, 降为 1 个通道, 卷积核采用  $7 \times 7$ , 保持  $H$ 、 $W$  不变, 输出的特征图, 最后通过 Sigmoid 函数生成空间权重系数, 再与输入的特征图相乘得到最终的特征图.

$$M_S(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (12)$$

### 2.7 多头自注意力块

该模块能够捕获全局信息, 已经成为视觉任务中一个受欢迎的工具, 对其的应用也在逐渐增多. 计算方式如式(13)所示, 其中 MSA、 $h$ 、 $w$ 、 $C$ 表示多头自注意力、图像高度、宽度和通道.

$$\Omega_{\text{MSA}} = 4hwC^2 + 2(hw)^2C \quad (13)$$

将多头自注意力块 (MHSAB) 作为编码器和解码器之间的连接器, 利用全局自注意力来处理合并来自编码器的最后一层. 这种方式对处理大图像时特别有效, 因为卷积对空间进行下采样, 而注意力可以有效地处理较小的分辨率, 从而提高性能.

在 MHSAB 中应用  $1 \times 1$  逐点卷积来生成查询矩阵、键矩阵和值矩阵  $Q \in \mathbb{R}^{H \times W \times d}$ ,  $K \in \mathbb{R}^{H \times W \times d}$ ,  $V \in \mathbb{R}^{H \times W \times d}$  如下:

$$\begin{cases} Q = C_1(XE_4) \\ K = C_1(XE_4) \\ V = C_1(XE_4) \end{cases} \quad (14)$$

为了有效地关联位置感知对象之间的信息, 注意力机制考虑了内容信息和不同位置特征之间的相对距离. 将高度和宽度的相对距离位置编码分别表示为  $P_h \in \mathbb{R}^{H \times 1 \times d}$  和  $P_w \in \mathbb{R}^{1 \times W \times d}$ . 注意力计算为  $QK^T + QP^T$ , 其中,  $QK^T$ 表示内容-内容交互,  $QP^T$ 表示内容-位置

交互。

随后,注意力矩阵由注意力机制计算为:

$$Attention(Q, K, P, V) = Softmax(QK^T + QP^T)V \quad (15)$$

通过将注意力机制加入模型中,可以计算序列中不同位置之间的相关性,为每个位置分配权重,使模型能够全局关注序列并动态建模长距离依赖关系,则能够减少参数耦合,增强模型的表达力和性能。最后将注意力矩阵重塑回其原始尺寸 $H \times W \times d$ 。然后将所得到的输出添加到 $XE_4$ 并传递到解码器。

## 2.8 前馈网络

通过前馈网络对改进后的密集残差 U-Net 网络获得的深度特征 $X_d$ 进行回归来获得最终的位姿结果,如图 2 所示。前馈网络如图 7(c) 所示,由顺序方式连接的神经元组成。每个神经元仅与前一层的神经元连接,它接收该层的输出,同时将其输出提供给下一层。流程如下:

$$\hat{y} = W_1(W_2(X_d)) \quad (16)$$

其中, $W_i$ 是权重向量, $\hat{y}$ 表示位姿估计的结果值。

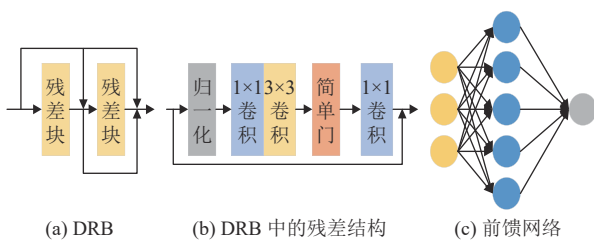


图 7 DRB、DRB 中的残差结构和前馈网络

## 3 实验结果与分析

本研究采用自制数据集进行训练和测试。在 Windows 11 操作系统下进行训练,采用 NVIDIA GeForce RTX 3060, 16 GB 显卡,开发语言是 Python 3.9.0,并运用 CUDA 11.8 版本加速训练,深度学习框架为 PyTorch 1.10.2, IDE 是 PyCharm。输入数据集图像的颜色通道为 RGB,图像初始尺寸为  $301 \times 301$ ,裁剪后大小为  $150 \times 150$ ,batch-size 设置为 16,学习率设置为 0.000 015,训练周期为 1000,其他参数值为默认值。

利用经过筛选后的特定航天器的 4 762 张 ISAR 图像数据,该图像数据为对 ISAR 成像转换后得到的实测数据。数据集按照 80% 的训练集数据和 20% 的验证集数据组成,划分后的训练集、验证集的图像数量分

别为 3 810、952。针对初始图像,先通过 OCR 文字识别方法提取图像中航天器位姿数据,并结合裁剪后的图像一起用于训练和验证。数据的分布情况如图 8 所示。由于 ISAR 成像技术只能在有限的范围内探测到航天器,因此所有图像的位姿角都在  $0^\circ$  和  $30^\circ$  之间,且高度集中在  $5^\circ$  和  $10^\circ$  之间。图 9 为模型训练的 Loss 曲线变化图,图 9 中横坐标为 epoch,纵坐标为训练损失。

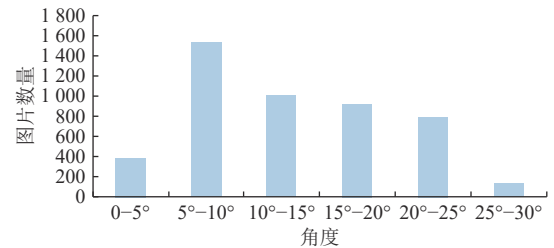


图 8 各区间位姿角度数量

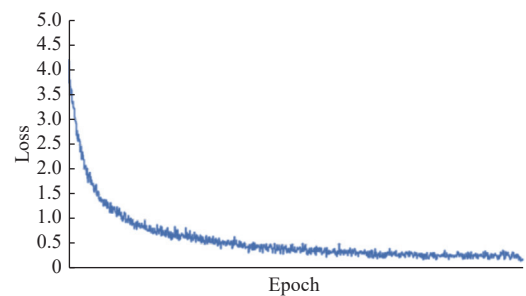


图 9 Loss 曲线

为了客观地验证改进后模型的性能,进行了对比测试、消融实验。系统地删除了模型中的不同组件,比较不同消融版本和完整模型之间的性能差异,对比它们所估计的位姿角与正确角度的误差大小。使用最小绝对误差 (MinAE)、最大绝对误差 (MaxAE) 和平均绝对误差 (MeanAE) 来判断。最小绝对误差是在通过模型预测一批位姿时其目标值和预测值之间的最小差异。平均绝对误差是在模型预测一批位姿时的平均误差。证明所提出的预处理方式,STN 和 CBAM 的有效性,确保它们能够提升模型的性能。

### 3.1 对比实验

为了评估模型改进后的性能,对目前的一些主流方法以及原模型进行了比较评估,如表 1 中所示。为了获得这些方法在该数据集上的结果,利用了论文公开的代码,并针对目前的数据集对模型进行了微调。对于实验参数设置为一致。

根据表 1,能够看出改进后的新模型能够更好地获

取航天器 ISAR 的图像特征并且能更加精确地识别航天器的位姿。

表 1 不同模型的对比实验

Model	MinAE	MaxAE	MeanAE
AlexNet	1.2612	1.5429	1.3274
ResNet	0.5827	0.8314	0.6983
HRNet	0.6475	0.7806	0.7374
AggPose	0.6212	0.7543	0.7128
SU-Net	0.1383	0.4859	0.3826
Ours	0.0865	0.3579	0.1826

### 3.2 消融实验

在这里,利用消融实验来分析模型中每个组件对模型性能的提升,如表 2 所示。

表 2 消融实验结果对比

指标	All	-IE	-PM	-STN	-CBAM
MinAE	0.0865	0.1168	0.0996	0.1084	0.1317
MeanAE	0.1862	0.2616	0.2123	0.2158	0.3627

随着图像增强方法的滤除,整个模型产生了更差的结果,位姿估计的误差增加。这就验证了该组件的有效性。

预训练模型可以减小对数据大小的依赖,并增强模型捕获更多特征信息的能力。在去掉了预训练模型后导致位姿识别的结果增大了误差。

空间变换模块使得网络不仅可以选择不相关区域,而且可以将这些区域转换成规范的、预期的姿态。在去掉该模块后,模型最后的预测结果明显增大,验证了该模块的有效性。

将跨通道信息和空间信息混合在一起提取信息特征,通过从模型中删除该注意力机制来证明它的有效性。

## 4 结论语

为使得针对 ISAR 图像中非合作航天器的位姿识别更加精确,本文在原有的一种密集残差 U-Net 网络结构上进行改进。通过航天器检测网络优先裁剪出包含航天器的完整区域,随着区域的裁剪从而避免标记数据对训练造成的影响,并使用增强的 Lee 滤波平滑噪声点,提升对比度从而提高图像质量,除此之外,为了克服图像中航天器不同程度的变形,引入 STN 空间变换模块,使模型学习到平移、缩放、旋转和更一般的扭曲不变性,为每个输入样本生成适当的转换,加入了通道注意力与空间注意力融合的注意力机制,强调

通道轴和空间轴这两个维度上的有意义特征。实验表明,改进后的网络拥有更加精确的检测精度,性能更加优良。

### 参考文献

- 包为民. 航天飞行器控制技术研究现状与发展趋势. 自动化学报, 2013, 39(6): 697-702.
- 梁斌, 杜晓东, 李成, 等. 空间机器人非合作航天器在轨服务研究进展. 机器人, 2012, 34(2): 242-256.
- Saleh JH, Lamassoure E, Hastings DE. Space systems flexibility provided by on-orbit servicing: Part 1. Journal of Spacecraft and Rockets, 2002, 39(4): 551-560. [doi: 10.2514/2.3844]
- Kassebom M. ROGER—An advanced solution for a geostationary service satellite. Proceedings of the 54th International Astronautical Congress of the International Astronautical Federation, the International Academy of Astronautics, and the International Institute of Space Law. Bremen: AIAA, 2003. U.1.02.
- Taylor B, Aglietti G, Fellowes S, *et al.* Remove debris mission, from concept to orbit. Proceedings of the 32nd Annual AIAA/USU Conference on Small Satellites. 2018. 1-28.
- Andert F, Ammann N, Maass B. Lidar-aided camera feature tracking and visual slam for spacecraft low-orbit navigation and planetary landing. Proceedings of the 3rd Advances in Aerospace Guidance, Navigation and Control. Toulouse: Springer, 2015. 605-623.
- Gaudet B, Furfaro R, Udrea B. Attitude estimation via LIDAR altimetry and a particle filter. Proceedings of the 24th AAS/AIAA Space Flight Mechanics Meeting. Orlando, 2014. 26-30.
- Pérez-Villar JIB, García-Martín A, Bescós J. Spacecraft pose estimation based on unsupervised domain adaptation and on a 3D-guided loss combination. Proceedings of the 2022 European Conference on Computer Vision. Tel Aviv: Springer, 2022. 37-52.
- Deng LW, Suo HF, Jia YQ, *et al.* Pose estimation method for non-cooperative target based on deep learning. Aerospace, 2022, 9(12): 770. [doi: 10.3390/aerospace9120770]
- Li KC, Zhang HP, Hu CY. Learning-based pose estimation of non-cooperative spacecrafts with uncertainty prediction. Aerospace, 2022, 9(10): 592. [doi: 10.3390/aerospace9100592]
- Gao H, Li ZH, Wang N, *et al.* SU-Net: Pose estimation network for non-cooperative spacecraft on-orbit. Scientific Reports, 2023, 13(1): 11780. [doi: 10.1038/s41598-023-



- 38974-1]
- 12 Xue B, Tong NN. DIOD: Fast and efficient weakly semi-supervised deep complex ISAR object detection. *IEEE Transactions on Cybernetics*, 2019, 49(11): 3991–4003. [doi: [10.1109/TCYB.2018.2856821](https://doi.org/10.1109/TCYB.2018.2856821)]
  - 13 Sun K, Xiao B, Liu D, *et al.* Deep high-resolution representation learning for human pose estimation. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 5686–5696.
  - 14 Cao X, Li XY, Ma LY, *et al.* AggPose: Deep aggregation vision transformer for infant pose estimation. *Proceedings of the 31st International Joint Conference on Artificial Intelligence*. Vienna: ijcai.org, 2022. 5045–5051.
  - 15 Huang HR, Zhao GP, Gu DQ, *et al.* Non-model-based monocular pose estimation network for uncooperative spacecraft using convolutional neural network. *IEEE Sensors Journal*, 2021, 21(21): 24579–24590. [doi: [10.1109/JSEN.2021.3115844](https://doi.org/10.1109/JSEN.2021.3115844)]
  - 16 Zhou YJ, Li WF, Ma Y, *et al.* Attitude estimation of on-orbit spacecraft based on the U-linked network. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 6518005.
  - 17 Cassinis LP, Fonod R, Gill E, *et al.* CNN-based pose estimation system for close-proximity operations around uncooperative spacecraft. *Proceedings of the AIAA Scitech 2020 Forum*. Orlando: AIAA, 2020. 1457.
  - 18 Oestreich C, Lim TW, Broussard R. On-orbit relative pose initialization via convolutional neural networks. *Proceedings of the AIAA Scitech 2020 Forum*. Orlando: AIAA, 2020. 0457.
  - 19 Li Y, Wu R, Xing M, *et al.* Inverse synthetic aperture radar imaging of ship target with complex motion. *IET Radar, Sonar & Navigation*, 2008, 2(6): 395–403.
  - 20 Chen VC, Martorella M. *Inverse Synthetic Aperture Radar Imaging: Principles, Algorithms and Applications*. Edison: SciTech Publishing, 2014. 56.
  - 21 Zhang Y, Yuan HX, Li HB, *et al.* Meta-learner-based stacking network on space target recognition for ISAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 12132–12148. [doi: [10.1109/JSTARS.2021.3128938](https://doi.org/10.1109/JSTARS.2021.3128938)]
  - 22 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 234–241.
  - 23 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *Proceedings of the 3rd International Conference on Learning Representations*. San Diego: ICLR, 2015.
  - 24 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 770–778.
  - 25 Jaderberg M, Simonyan K, Zisserman A, *et al.* Spatial transformer networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2015. 2017–2025.
  - 26 Huang G, Liu Z, Van Der Maaten L, *et al.* Densely connected convolutional networks. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 2261–2269.
  - 27 Wang XT, Yu K, Wu SX, *et al.* ESRGAN: Enhanced super-resolution generative adversarial networks. *Proceedings of the 2019 European Conference on Computer Vision*. Munich: Springer, 2019. 63–79.
  - 28 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 3–19.
  - 29 Fu J, Liu J, Tian HJ, *et al.* Dual attention network for scene segmentation. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 3141–3149.
  - 30 Wang XL, Girshick R, Gupta A, *et al.* Non-local neural networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7794–7803.

(校对责编: 孙君艳)