

# 基于信息表征增强的空间弱小目标检测方法

范铭楷<sup>1)</sup> 薛丹娜<sup>1)</sup> 闫庆森<sup>1)</sup> 朱 宇<sup>1)</sup> 孙瑾秋<sup>2)</sup> 张艳宁<sup>1)</sup>

<sup>1)</sup>(西北工业大学计算机学院 西安 710129)

<sup>2)</sup>(西北工业大学航天学院 西安 710072)

**摘要** 基于深度学习的目标检测技术近年来取得了显著进展,已被广泛应用于多个领域。然而,目前尚缺乏专门针对空间目标设计的深度学习检测方法。与自然图像目标检测相比,空间目标检测面临着独特的挑战:首先,由于拍摄距离极远,空间目标在图像中通常仅呈现为包含少量像素的弱小光斑,缺乏清晰的形状和颜色特征;另外,复杂的空间环境以及设备因素会导致图像中存在各种噪声,例如杂散光引起的明亮背景、相机因素导致的热像素噪声;最后,视场中密集的恒星会导致在空间图像中容易出现部分光斑相互粘连的情况。上述难点无疑增加了空间目标检测的难度。本文提出了一种基于信息表征增强的无锚框空间弱小目标检测方法,该方法通过设计端到端卷积神经网络模型,能够实现同时对空间弱小目标检测和质心定位。针对目标弱小、背景噪声干扰的问题,我们首先设计了跨空间-通道注意力模块和压缩-激励的多尺度特征融合模块,提升模型对含噪图像中弱小目标的有效信息关注度,从而提高噪声背景下目标的检测能力;在此基础上,为了进一步解决光斑粘连的问题,我们引入了基于密度图的损失函数,通过使模型更加直接地学习图像中目标的整体空间分布与数量信息,以此使模型掌握粘连光斑中目标的数量,从而更准确分辨粘连光斑中各个独立目标。此外,本文仿真了包含点状、条状目标和多种噪声背景的空间图像,并对目标的质心位置、边界框和像素坐标进行了精细标注。为验证方法的有效性,我们在本文的仿真图像与公开的实拍序列图像数据集 SpotGEO 中进行了实验验证。在仿真图像数据测试中,本文方法的 F1 分数为 95.34%,并实现了 0.4478 的亚像元级平均质心定位精度。同时,我们还对本文方法进行了不同噪声强度影响下质心定位精度测试以及不同硬件条件下处理效率测试,以更全面分析本文方法的能力。在公开的实拍序列图像数据集 SpotGEO 中,本文方法结合引入的序列后处理方法取得的 F1 分数为 93.08%。实验结果证明了本文方法对空间弱小目标准确检测和精确定位方面的优越性能。

**关键词** 空间弱小目标检测;质心定位;注意力机制;特征融合;数量信息

**中图法分类号** TP391      **DOI号** 10.11897/SP.J.1016.2025.00537

## Dim and Small Space Target Detection Method Based on Enhanced Information Representation

FAN Ming-Kai<sup>1)</sup> XUE Dan-Na<sup>1)</sup> YAN Qing-Sen<sup>1)</sup>

ZHU Yu<sup>1)</sup> SUN Jin-Qiu<sup>2)</sup> ZHANG Yan-Ning<sup>1)</sup>

<sup>1)</sup>(School of Computer Science, Northwestern Polytechnical University, Xi'an 710129)

<sup>2)</sup>(School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072)

**Abstract** In recent years, deep learning based object detection technologies have achieved significant advancements and have been widely utilized in various fields. However, there remains a lack of deep learning detection methods specifically designed for space targets. Compared to object detection in natural image, space target detection encounters unique challenges: One primary challenge arises

收稿日期:2024-07-04;在线发布日期:2025-01-08.本课题得到国家自然科学基金(62301432)、陕西省自然科学基础研究计划资助项目(2023-JC-QN-0685,QCYRCXM-2023-057)、空天地海一体化大数据应用技术国家工程实验室资助.范铭楷,博士研究生,中国计算机学会(CCF)学生会员,主要研究方向为空间目标检测、小目标检测.E-mail: 8jin@mail.nwpu.edu.cn.薛丹娜,博士研究生,主要研究方向为图像与视频分割.闫庆森,博士,教授,主要研究领域为高动态范围成像、深度学习、医学图像分析和计算机视觉.朱 宇,博士,副研究员,主要研究领域为图像去模糊、图像增强和图像超分辨率.孙瑾秋,博士,教授,主要研究领域为空间图像处理、计算机视觉、模式识别.张艳宁(通信作者),博士,教授,中国计算机学会(CCF)会士,主要研究领域为遥感图像分析、计算机视觉和模式识别等.E-mail: ynzhang@nwpu.edu.cn.

from the extreme shooting distances involved, which results in space targets appearing as mere faint spots in space images, comprising only a few pixels and lacking distinct shape and color features. Additionally, the complex space environment and device-related factors contribute to various types of noise within the images, such as bright backgrounds caused by stray light and hot pixel noise arising from camera issues. Moreover, some regions in space contain a large number of stars, the densely packed stars within the field of view frequently lead to the partial overlapping of light spots in space images, further complicating the detection task. These difficulties undeniably exacerbate the challenges of space target detection. In response to these challenges, this paper introduces an anchor-free detection method for small space objects, grounded in enhanced information representation. The proposed method employs a specifically designed end-to-end convolutional neural network (CNN) model, capable of concurrently detecting small space targets and precisely locating their centroids. To tackle the issues of weak object signals and background noise interference, we have developed a cross-space-channel attention module and a squeeze-excitation multi-scale feature fusion module. These components are designed to enhance the model's ability to focus on the pertinent information within noisy images, thereby significantly improving its capability to detect targets in the challenging noise-laden backgrounds. Building upon this, to further address the problem of overlapping light spots, we incorporate a density map-based loss function. This approach enables the model to more effectively learn the overall spatial distribution and quantity information of the targets within an image. As a result, the model accurately discerns the number of individual targets within overlapping clusters, facilitating more precise differentiation and identification of each distinct target. To comprehensively validate the effectiveness of the proposed method, we simulated space images containing point targets and streak targets with various noisy backgrounds. These images were meticulously annotated with detailed information regarding the centroid locations, bounding boxes, and pixel coordinates of the targets. Experimental validations were conducted using both the simulated dataset and the publicly available real image sequences dataset SpotGEO. In testing conducted on the simulated image data, our method achieved a *F1* score of 95.34% and a sub-pixel level average centroid localization accuracy of 0.4478. Additionally, we performed further tests to examine the impact of varying noise levels on centroid localization accuracy, as well as an analysis of processing efficiency under different hardware conditions to provide a comprehensive assessment of the method's capabilities. Furthermore, within the publicly available SpotGEO dataset, after integrating a sequential post processing method, our method yielded an *F1* score of 93.08%. Experimental results demonstrate the superior performance of our approach in accurately detecting and precisely localizing these small but crucial space objects.

**Keywords** dim and small space target detection; centroid localization; attention mechanism; feature fusion; quantity information

## 1 引 言

近年来,基于深度学习的目标检测方法在遥感目标检测、自动驾驶、故障与缺陷检测等领域中表现出卓越性能。相较而言,空间目标检测领域的深度学习方法较为稀缺。如何设计深度学习模型并充分发挥其出色的特征提取能力,以实现更全、更准的空

间目标检测,是值得探索的问题。空间目标检测的主要对象是太空环境中如任务碎片、碰撞碎片和解体碎片<sup>[1]</sup>等各种空间弱小目标。不同于一般目标检测,空间目标检测需要确定空间图像中目标的质心坐标,从而为后续空间位置计算和轨道确定等任务提供关键数据支撑。

空间图像中,目标、恒星均以光斑的形式呈现(见图1),在某些拍摄模式和参数设置情况下,目

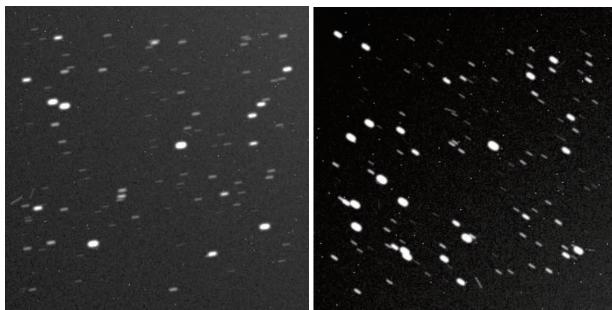


图 1 空间图像

标、恒星光斑及其相似,仅依靠单帧图像的信息极难区分,需要利用图像序列信息。因此,众多空间目标检测方法通常包括两个主要阶段:第一阶段在单帧空间图像中按相同类别检测出目标、恒星光斑作为候选,并结合质心定位算法计算光斑的质心坐标;第二阶段基于第一阶段结果,在序列图像中进行数据关联从而对目标、恒星光斑进一步区分。如果第一阶段对光斑的检测结果不佳,将极大影响最终结果。本文方法聚焦于第一阶段研究,在单帧图像中将目标、恒星光斑均作为候选目标进行检测,且在不结合质心定位算法的情况下,直接预测它们的质心坐标。

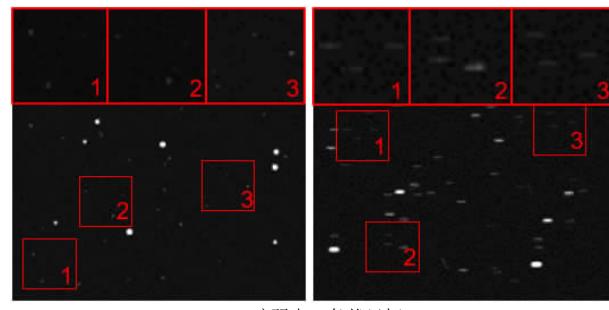
然而,拍摄空间图像时往往会影响到宇宙空间环境、成像距离以及相机自身元器件性能等因素的影响,从而提升了空间图像目标检测的难度,主要有以下几个方面:

(1) 目标暗弱:空间目标进行拍摄的拍摄距离极大,因此空间目标在图像中仅呈现为包含少量像素的点状或由于相对运动而形成的条状光斑、无自身形状、颜色等特征信息。其中众多目标尺寸极小,与背景对比度低(见图 2(a))。

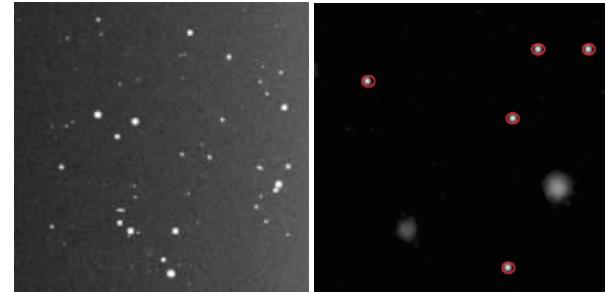
(2) 图像中存在各种噪声:由于空间环境和相机自身的元器件原因,拍摄的空间图像中存在杂散光、热像素等噪声(见图 2(b))。杂散光会极大改变空间图像背景的灰度分布,从而影响目标的灰度特征;热像素在图像中表现为点状的光斑,易造成混淆。

(3) 光斑粘连:当空间目标与恒星部分遮挡时,表现在图像中为碎片光斑与恒星光斑粘连,呈现为一块光斑(见图 2(c))。粘连光斑的大小、形状、灰度分布与其他非粘连的目标光斑存在较大差异,在此类粘连光斑上需要准确区分各独立目标。

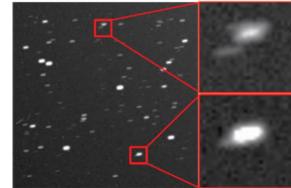
针对单帧空间图像的候选目标检测,目前主要采用传统方法,其中众多方法<sup>[2-4]</sup>对图像进行预处理后,通过设定全局或分区域阈值分割候选目标与背景,这样的处理方式容易将极暗弱的目标当作背景噪声滤除,并且会将粘连的两块光斑误判为一个候



(a) 暗弱点、条状目标



(b) 杂散光与热像素



(c) 光斑粘连现象

图 2 空间目标检测难点

选目标。此外,这些传统方法确定质心坐标的方式是对提取的光斑区域应用插值法、拟合法等质心定位算法计算质心坐标,导致质心坐标受到光斑区域提取误差和质心定位算法误差的双重影响。基于深度学习的方法在空间目标检测领域的研究成果较少<sup>[5-6]</sup>,通常是直接应用通用的深度学习检测模型,但空间目标与一般的小目标相比普遍尺寸更小,信息更加匮乏,直接应用通用检测模型会导致较多目标漏检,同时这些方法在质心定位方面也缺乏专门设计。另外,深度学习方法需要数据驱动,而空间图像需要专业的设备进行拍摄,成本较高,对拍摄到的图像进行准确标注也非常困难。

为了解决上述问题,本文提出了一种基于信息表征增强的无锚框空间碎片检测定位方法,直接输出目标的质心位置。首先,本文在模型的骨干网部分与特征融合部分分别设计了跨空间-通道注意力模块与压缩-激励的多尺度特征融合模块,提升模型对噪声背景中弱小光斑目标的检测能力;其次,本文引入了基于密度图的损失函数,提升模型对粘连光斑的处理能力。另外,本文仿真了包含多种目标形态的空间图像数据集进行模型训练与测试,对现有

的公开数据进行了补充。在本文构建的仿真图像数据集和实拍图像数据集 SpotGEO<sup>[7]</sup> 上本文方法取得了优秀的检测结果。

本文的工作是目前少有的对深度学习方法在空间目标检测领域应用的探索, 主要贡献如下:

(1) 本文提出了端到端的目标检测模型, 设计了质心坐标预测分支直接预测目标的质心位置, 避免检测和定位分两阶段的误差累加。

(2) 针对空间目标暗弱、易受背景噪声干扰的问题, 在骨干网中设计了跨空间-通道注意力模块, 充分挖掘和利用空间位置及通道之间的信息, 通过二者的联合表征, 优化模型对弱小目标有效信息的表达。在特征融合部分, 设计了压缩-激励的多尺度特征融合模块, 深入学习待融合特征图中所有对应通道特征间的相互关系, 调整各通道特征的重要程度。以上两模块共同提升模型对含噪图像中弱小光斑目标的有效信息关注度, 从而提高噪声背景下弱小光斑目标检测能力。

(3) 针对光斑粘连的问题, 引入基于密度图的损失函数, 在目标数量信息层面上对模型施加更直接的约束, 使模型掌握粘连光斑中的目标数量, 从而更准确地对粘连光斑中的各独立目标进行区分。

本文第 2 节介绍传统空间目标检测方法以及基于深度学习的空间目标检测方法的相关工作; 第 3 节介绍本文所提出的基于信息表征增强的空间目标检测定位方法; 第 4 节展示本文方法与其他方法的实验结果对比; 第 5 节为对全文进行总结。

## 2 相关工作

### 2.1 传统空间目标检测方法

空间目标检测的传统方法可以分为检测后跟踪 (Detect-Before-Track, DBT) 与检测前跟踪 (Track-Before-Detect, TBD)。DBT 方法先在每帧图像中检测目标, 再利用序列图像信息对单帧的目标检测结果进行滤除与补充。TBD 方法则在单帧预处理后保留大量光斑, 在序列图像中对保留的大量光斑进行轨迹关联, 从而实现目标判断。

#### (1) 基于 DBT 的空间目标检测方法

Xu 等人<sup>[8]</sup> 提出了一种加权迭代阈值算法进行空间目标检测, 该方法通过改变引入的两个权重系数, 获得能够极大降低后续目标质心定位误差的最优阈值。Sun 等人<sup>[2]</sup> 引入形态学算子首先抑制空间

图像噪声并进行图像恢复, 然后使用中值滤波消除视场恒星影响并进行二值化分割提取目标位置。Do 等人<sup>[9]</sup> 基于高斯过程回归对空间图像进行前背景分割提取候选目标, 然后通过序列图像配准移除恒星, 最后利用运动信息对剩余的目标进行虚警滤除与跟踪。Yao 等人<sup>[10]</sup> 利用恒定虚警率确定分割阈值进行单帧候选目标点提取, 之后在多帧中基于迭代距离与轨迹关联完成空间目标的检测。Sun 等人<sup>[11]</sup> 提出了由包含不同尺度和方向的各类滤波器组合提取特征训练支持向量机以实现像素级分类, 从而进行空间目标检测的方法, 提升了对低信噪比目标的检测结果。

#### (2) 基于 TBD 的空间目标检测方法

Huang 等人<sup>[12]</sup> 提出时空域联合滤波的空间目标检测方法, 首先对单帧图像进行空域上的背景抑制, 在此基础上对多帧图像进行时域相关与时域差分处理完成目标检测。Li<sup>[13]</sup> 提出了一种基于时空域管道的拟多级假设检验算法, 将光斑组成树形结构进行假设检验的精细检测, 提高了复杂情况下空间目标的检测精度和鲁棒性。同样基于假设检验, Xi 等人<sup>[14]</sup> 提出了时间索引多级假设检验方法进行空间目标检测。Liu<sup>[15]</sup> 通过设置双重假设检验条件进行空间目标的最终判断。Li 等人<sup>[16]</sup> 基于弱小目标的时空特性, 结合邻域聚类方法和轨迹连续性提出了一种基于时空假设检验的弱小运动目标轨迹检测算法。Ahmadi 等人<sup>[17]</sup> 提出了一种两级假设检验的分层跟踪系统。对于划分至低层中的每个点, 应用多目标粒子群优化技术以减少每个跟踪树中的分支数。在高层中, 通过迭代连接连续帧的最优轨迹。

以上方法极大推动了空间目标检测的发展。但是, 传统方法在单帧图像中提取候选目标时大多应用阈值分割, 易将极弱小目标误归为背景, 且会将相互粘连的候选目标分割为同一块光斑。TBD 方法在一定程度上提升了弱小目标检测能力, 但是此类方法计算量大。其中假设检验方法的结果还对设置的假设检验条件敏感。不同于上述方法, 本文基于神经网络自主学习数据的特征表示, 避免了对人工先验知识的依赖, 能自动提取并整合不同尺度的特征, 从而更好地表征弱小目标, 并能够对相互粘连的候选目标进行区分。

### 2.2 基于深度学习的目标检测方法

基于深度学习的目标检测算法包括直接预测目标边界框和类别的一阶段方法 (One-Stage), 如 (You Only Look Once, YOLO)<sup>[18]</sup> 系列, 以及通过

生成候选区域并进一步处理的两阶段方法(Two-Stage),如 Faster R-CNN 及其变体<sup>[19-20]</sup>。一阶段方法通常具有较快的推理速度,而两阶段方法通常具有更高的检测精度。在这些检测方法的基础上,引入注意力机制能够通过突出关键信息,抑制无关特征优化目标信息的表征。Hu 等人<sup>[21]</sup>提出 SENet(Squeeze-and-Excitation Networks)方法,通过“压缩-激励”过程来捕捉通道之间的相关性,自适应地重新加权各个通道特征提升模型表现。Woo 等人<sup>[22]</sup>提出了 CBAM(Convolutional Block Attention Module)方法,先后通过独立的两个模块分别调整各通道和空间位置的权重,进一步增强了特征表达的准确性。

近年来,基于 Transformer 的目标检测方法发展迅速,其本质上也是一种自注意力机制,DETR(Detection Transformer)<sup>[23]</sup>是这一方向的代表性工作。DETR 通过 Transformer 架构对全局图像信息进行建模,消除了检测方法中的锚框机制,同时能够建模物体间的长程依赖关系。但由于其计算复杂度较高,训练、推理速度较慢。Zhu 等人<sup>[24]</sup>提出 Deformable DETR,减少了全局注意力机制中冗余的计算,专注于特定位置的关键信息提取,效率相较 DETR 模型显著提高。DINO 方法<sup>[25]</sup>同样是对 DETR 的一项重要改进,通过设计混合查询选择方法并对真值进行去噪训练,极大提升了目标检测精度。Pan 等人<sup>[26]</sup>引入了 HiLo 自注意力机制,将注意力头分为两组,一组通过局部窗口自注意力编码高频信息,另一组通过全局注意力编码低频信息,不仅取得了良好的检测结果,还实现了更快的速度和更低的内存消耗。Zhang 等人<sup>[27]</sup>提出了密集不同查询方法,通过铺设密集查询并对不同查询进行一对一分配,解决了稀疏查询召回率低和密集查询优化困难的问题,使用 Swin 作为主干网,取得了很好的检测效果。Peng 等人<sup>[28]</sup>通过交互,融合卷积善于提取的局部特征和 Transformer 善于提取的全局特征,从而使卷积神经网络和 Transformer 的优势互补,以提升模型局部细节和全局特征的学习能力。

上述方法在目标检测领域取得了优良的效果,但由于模型中最终用于分类和定位的特征往往经过下采样,这对尺寸极小的空间目标不友好,容易导致特征丢失。基于 Transformer 的方法尽管具有全局建模能力,但全局自注意力机制会导致对局部特征关注不足。虽然有的方法融合了卷积和 Transformer 的优点,试图同时捕捉局部和全局特征,但局部特征在融合过程中依然易被削弱,这对于本就弱小的空

间目标影响极大。

### 2.3 基于深度学习的空间目标检测方法

在基于深度学习的小目标检测方面,目前已有大量研究,然而目前仅有少量基于深度学习的空间目标检测研究。Liu 等人<sup>[29]</sup>提出以 ResNet<sup>[30]</sup>为骨干网的模型,通过引入通道注意力与空间注意力模块,增强模型对空间目标的表征能力。Li 等人<sup>[31]</sup>将 Mask R-CNN 模型<sup>[32]</sup>应用于空间目标检测与识别任务中,并应用 R-FCN<sup>[33]</sup>与 Light-head R-CNN<sup>[34]</sup>的结构对其进行优化,提升了模型的推理速度。上述两种方法<sup>[29,31]</sup>所检测的对象主要是工作于近地轨道上、拍摄距离较近、包含像素数较多且具有明显形状特征的空间目标,信息更加丰富,与本文所研究的对象存在差异。Xue 等人<sup>[35]</sup>提出了一种基于 U-Net<sup>[36]</sup>的单帧空间目标分割模型,并引入权重图的方式解决正负样本不平衡的问题。但是,此方法是对图像进行分割,并不直接输出质心坐标,需要再对分割结果应用质心定位算法完成质心坐标的确定,这会导致最终的坐标误差来源于两个阶段。Dai 等人<sup>[5]</sup>提出在多帧中检测空间目标的方法,利用现有的 PP-Yolov2 模型<sup>[37]</sup>在单帧检测候选目标,然后设计非深度学习的候选目标过滤与补充方法实现最终的检测。Abay 等人<sup>[6]</sup>提出了使用 EfficientNet<sup>[38]</sup>作为骨干网结合特征金字塔网络(Feature Pyramid Network, FPN)<sup>[39]</sup>进行在单帧上进行候选目标检测,然后设计结合多帧信息的后处理算法对检测结果进行优化。但是,文献[5]与文献[6]方法都是直接利用通用检测模型进行空间目标检测,而空间目标尺寸远小于一般图像中的目标,上述方法也未根据空间目标的特性进行相应调整,检测结果仍然具有提升空间,且无专门负责质心坐标预测的模块。

不同于现有基于深度学习的空间目标检测方法,本文专门设计了增强模型对空间弱小目标的信息表征的模块,并且增加了质心预测头从而使模型能够直接输出质心坐标。最终共同实现空间目标的高精度检测定位。

## 3 方法论述

### 3.1 模型整体结构

本文所提出的检测模型整体结构如图 3 所示,主要包括骨干网部分、压缩-激励的多尺度特征融合模块以及输出质心坐标、分类结果和边界框坐标的检测头部分。

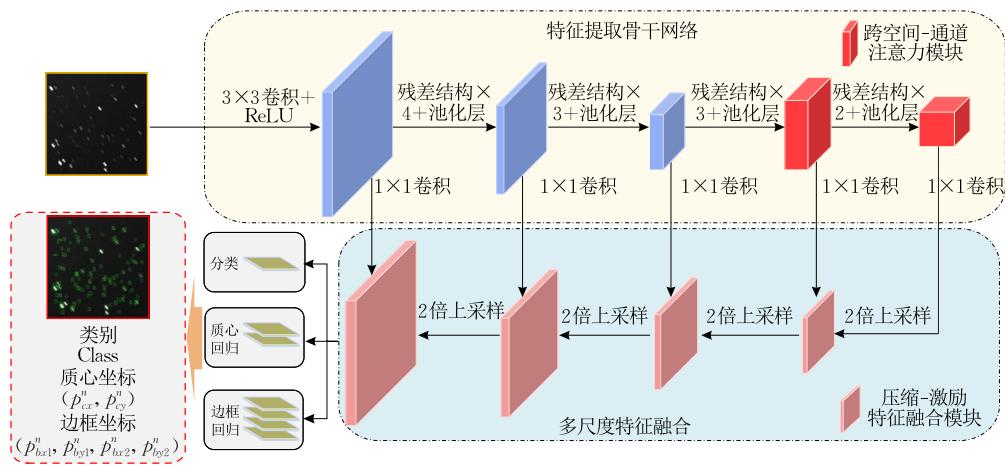


图 3 模型结构示意图

骨干网部分由残差模块<sup>[30]</sup>和跨空间-通道注意力模块共同组成,骨干网结构如表 1 所示。残差模块通过跨层连接的方式,有效缓解了深层网络中梯度消失的问题,残差结构如图 4 所示。为进一步提升模型对弱小目标的有效信息表达,本文设计了跨空间-通道注意力模块。该模块充分交互了空间和通道的信息,能够学习两个维度信息之间的潜在关系,从而对特征图进行进一步优化。

表 1 骨干网络结构

名称	具体结构
卷积层	$3 \times 3$ 卷积, 64 Batch Normalization(BN) ReLU
残差层 1	残差结构 $\times 4$ 池化(stride=2)
残差层 2	残差结构 $\times 3$ 池化(stride=2)
残差层 3	残差结构 $\times 3$ 池化(stride=2)
特征增强	跨空间-通道注意力模块
残差层 4	残差结构 $\times 2$ 池化(stride=2)
特征增强	跨空间-通道注意力模块

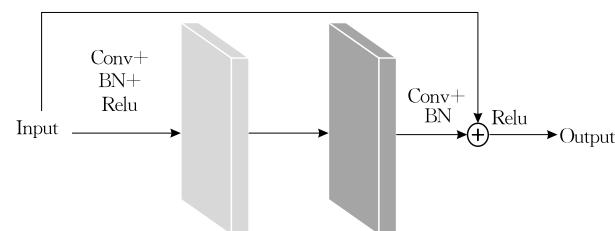


图 4 残差结构示意图

为使模型更关注待融合的特征图中对最终结果更有益的通道,本文在经典特征金字塔网络的基础上设计了压缩-激励的多尺度特征融合模块,对不同尺度的特征共进行四次融合。考虑到空间目标极小的尺寸,本文仅保留最后的融合模块输出的大分辨

率特征图输入至检测头部分,其大小对应骨干网的输入特征图大小。

为实现端对端目标检测与质心定位,本文设计了包含三个分支的检测头部分,分别是:类别分支,生成热力图,负责区分目标与背景;质心回归分支,输出双通道的特征图,分别代表目标质心与热力图中关键点的横纵坐标偏差;边框回归分支,负责对目标的边界框坐标进行预测。

### 3.2 跨空间-通道注意力模块

CBAM 方法<sup>[22]</sup>先后在空间维和通道维为特征图生成权重,从而对特征图进行调整。其使用串行结构(见图 5(a))依次生成通道注意力与空间注意力,尽管这种方式能够在通道和空间维度上调整特征图,但由于通道维度和空间维度的特征信息是在各自维度分别优化的,并未充分挖掘空间与通道维度之间的潜在关系,在一定程度上限制了模型的表达能力。为了解决这一问题,本文提出了跨空间-通道注意力模块。与 CBAM 结构不同,跨空间-通道注意力模块采用并行结构(见图 5(b)),使空间和通道信息在同一处理阶段进行交互,在学习各自维度内信息关系

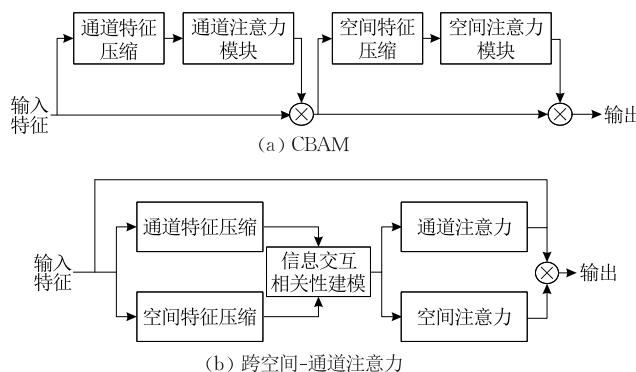


图 5 CBAM 与跨空间-通道注意力流程示意图

的基础上还学习了两个维度信息之间的潜在关系，确保特征图中每个位置的信息是不同维度的信息同时进行联合优化的结果，从而使模型能够更加一致和准确地表达含噪图像中的弱小目标的有效信息。

跨空间-通道注意力模块如图 6 所示，对于特征  $X \in R^{H \times W \times C}$ ，为了获取空间和通道维度更全面的特征表示，首先在空间与通道维度均使用平均池化与最大池化，这两种池化操作能够从不同角度提取特征信息：平均池化关注特征信息的均衡性，最大池化关注特征信息的显著性。两种池化结果进行元素级

相加，得到通道特征表示  $F_C \in R^{1 \times 1 \times C}$  与空间特征表示  $F_S \in R^{H \times W \times 1}$ 。 $F_C$  与  $F_S$  结合了两种池化的结果，确保模型既能捕捉到全局趋势，又不忽视关键的显著信息。将空间特征表示  $F_S \in R^{H \times W \times 1}$  的维度重塑为  $F'_S \in R^{1 \times 1 \times HW}$ ，以便后续与通道特征表示  $F_C \in R^{1 \times 1 \times C}$  拼接成  $F_{\text{Concat}} \in R^{1 \times 1 \times (HW+C)}$ ，拼接后的  $F_{\text{Concat}}$  包含了空间、通道两个方向的信息。

$$F_S = \text{reshape}(F_S^{H \times W \times 1}, (1, 1, H \times W)) \quad (1)$$

$$F_{\text{Concat}} = \text{Concat}(F_C, F_S)_{\text{dim}=3} \quad (2)$$

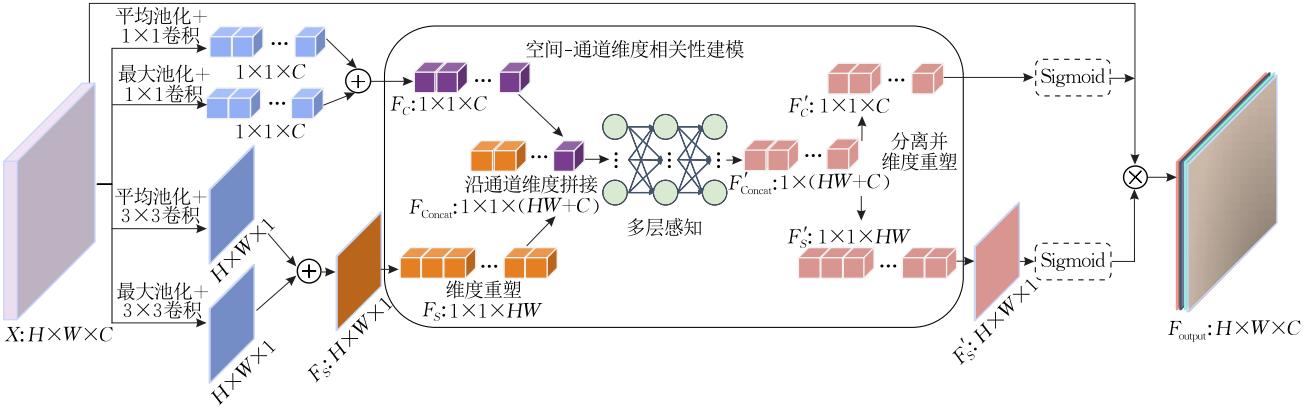


图 6 跨空间-通道注意力模块示意图

为了使  $F_{\text{Concat}} \in R^{1 \times 1 \times (HW+C)}$  中的每个位置均为空间与通道信息的共同表示，本文利用多层次感知器使  $F_{\text{Concat}} \in R^{1 \times 1 \times (HW+C)}$  中的空间、通道信息产生交互，其输出为  $F'_{\text{Concat}} \in R^{1 \times 1 \times (HW+C)}$ ， $F'_{\text{Concat}}$  中每个元素均为空间和通道维度的共同特征描述信息。最后基于  $F'_{\text{Concat}}$  中的信息，增强对特征图中蕴含重要信息的空间位置与通道的表达，具体实现方法是：将多层次感知机处理后的结果  $F'_{\text{Concat}}$  分离为  $F'_C \in R^{1 \times 1 \times C}$  与  $F'_S \in R^{H \times W \times 1}$ ，使用 Sigmoid 函数分别生成空间权重与通道权重，与原始输入特征相乘，得到最终的特征表示  $F_{\text{output}} \in R^{H \times W \times C}$ 。

$$F'_S, F'_C = \text{Split}(\text{MLP}(F_{\text{Concat}})) \quad (3)$$

$$F_{\text{output}} = \text{Sigmoid}(F'_S) \text{Sigmoid}(F'_C) X \quad (4)$$

### 3.3 压缩-激励的多尺度特征融合模块

经典特征金字塔网络的融合方式将所有信息“一视同仁”，而本文所考虑的是融合时模型应该更加关注两个特征图中对最终结果更有益的信息。受 SENet<sup>[21]</sup>启发（见图 7(a)），本文设计了压缩-激励的多尺度特征融合模块（见图 7(b)）。与 SENet 作用于单个特征图不同，此模块接收骨干网相邻两个层次产生的两个特征图作为输入，交互这两个特征图中所有通道特征，学习其相互关系。基于此，在融合时

向待融合的通道信息赋予对应其重要程度的权重。在权重生成阶段，采用 Softmax 函数代替 Sigmoid 函数，从而更倾向于待融合通道信息间的“相对重要性”，即两个待融合的通道之间哪一方对目标信息表征相对更有益。最终以各通道加权融合的方式优化融合特征。

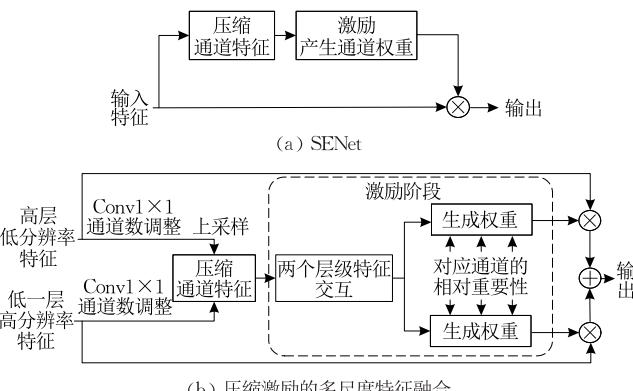


图 7 SENet 与压缩激励的多尺度特征流程示意图

压缩-激励的多尺度特征融合模块要对骨干网相邻两个层次产生的特征图进行融合，二者的维度不同。因此需要将两特征图调整为相同通道数，并且对高层的低分辨率特征图  $F_{lr}$  进行上采样，使其具有与低一层高分辨率特征图  $F_{hr}$  相同的分辨率。该

处理过程可表示为

$$F_1^{H \times W \times C} = \text{Conv}_{1 \times 1}(\text{upsample}(F_{lr})) \quad (5)$$

$$F_2^{H \times W \times C} = \text{Conv}_{1 \times 1}(F_{hr}) \quad (6)$$

维度处理完成后,此模块对维度均已调整为  $H \times W \times C$  的两个特征图  $F_1 \in R^{H \times W \times C}$  与  $F_2 \in R^{H \times W \times C}$  进行融合操作,如图 8 所示。首先以像素级元素相加的方式融合  $F_1, F_2$  的信息,得到  $F_{1,2} \in R^{H \times W \times C}$ 。为了

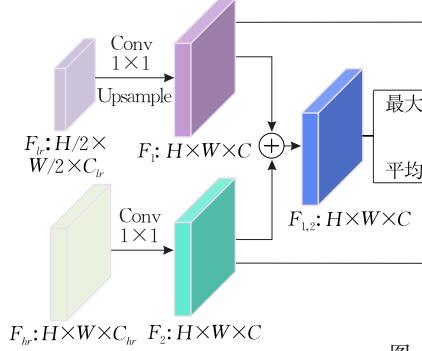


图 8 压缩-激励的多尺度特征融合模块示意图

$$F_{AM}^{1 \times 1 \times C} = F_{Avg} + F_{Max} \quad (7)$$

$$F_3^{1 \times 1 \times 2C} = \text{MLP}(F_{AM}) \quad (8)$$

将  $F_3$  的通道维度分离为  $F_{c1} \in R^{1 \times 1 \times C}$ ,  $F_{c2} \in R^{1 \times 1 \times C}$ , 对  $F_{c1}$  和  $F_{c2}$  的对应通道进行逐通道的 Softmax 处理, 得到各对应通道的权重  $W_{c1} \in R^{1 \times 1 \times C}$  与  $W_{c2} \in R^{1 \times 1 \times C}$ , 具体计算方式为

$$W_{c1}(1,1,i) = \frac{e^{F_{c1}(1,1,i)}}{e^{F_{c1}(1,1,i)} + e^{F_{c2}(1,1,i)}} \quad (9)$$

$$W_{c2}(1,1,i) = \frac{e^{F_{c2}(1,1,i)}}{e^{F_{c1}(1,1,i)} + e^{F_{c2}(1,1,i)}} \quad (10)$$

其中,  $i=1,2,\dots,C$ 。得到通道权重后, 原本两个待融合的特征图  $F_1, F_2$  分别与  $W_{c1}, W_{c2}$  相乘后再进行融合, 得到最终融合结果  $F_{\text{Fuse}} \in R^{H \times W \times C}$ :

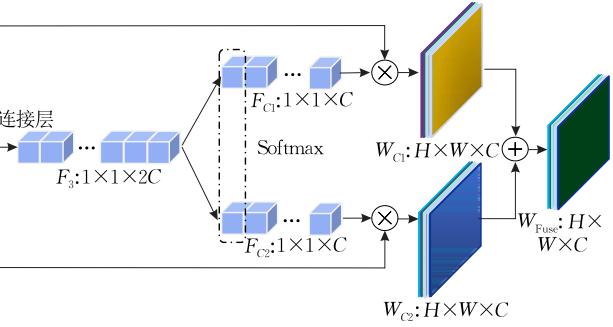
$$F_{\text{Fuse}} = W_{c1}F_1 + W_{c2}F_2 \quad (11)$$

### 3.4 密度图损失

密度图中所有元素总和对应于图像中的目标个数, 通常被用于目标计数任务中<sup>[40-41]</sup>。目标检测模型输出的热力图包含的是图像中目标的关键点信息, 而密度图包含有图像中目标的数量信息, 理想情况下, 热力图中目标关键点的数量应该与密度图的元素总和一致。故目前也有目标检测方法利用密度图所携带的目标整体数量信息处理密集遮挡的场景<sup>[42-43]</sup>。

为了进一步解决光斑粘连的问题, 本文在提升模型对重要信息关注度的基础上引入基于密度图的损失<sup>[27]</sup>。该损失提供了一个关于全局目标数量的约束, 使得模型能够更加直观地学习到图像中目标数量

聚合  $F_{1,2}$  的空间信息, 对  $F_{1,2}$  每个通道的特征分别使用平均池化和最大池化进行压缩, 生成  $F_{Avg} \in R^{1 \times 1 \times C}$  与  $F_{Max} \in R^{1 \times 1 \times C}$ , 将  $F_{Avg}$  与  $F_{Max}$  进行像素级相加, 得到  $F_{AM} \in R^{1 \times 1 \times C}$ 。 $F_{AM}$  目前包含待融合的两个特征图的通道信息描述, 但此时通道之间尚未产生交互, 因此使用全连接层交互  $F_{AM}$  所有通道的信息, 对其相关性进行建模, 输出  $F_3 \in R^{1 \times 1 \times 2C}$ 。



信息, 从而准确掌握粘连光斑的目标数量信息, 更加精准地分辨粘连光斑中的独立目标。基于密度图的损失具体定义为

$$L_{den} = \|U \otimes G_{\sigma_j} - D_{gt}\|_2^2 \quad (12)$$

其中,  $\|\cdot\|_2^2$  是均方误差损失,  $U$  是根据热力图生成的代表候选目标关键点的指示矩阵, 具体生成方法为: 基于类别热力图, 将其中大于目标置信度阈值的位置设置为 1, 其余位置设置为 0。 $G_{\sigma_j}$  是高斯核, 对生成的指示矩阵  $U$  进行高斯化。 $D_{gt}$  是密度图真值, 其计算方式如下

$$D_{gt}(p) = \sum_{j=1}^N \delta(p - p_j) \otimes G_{\sigma_j}(p) \quad (13)$$

其中,  $p$  代表位置坐标,  $p_j$  为目标中心位置坐标,  $D_{gt}$  包含目标数量信息, 将式(6)作为整体损失函数的一部分从而在整体目标数量层面对模型进行约束。

### 3.5 损失函数

整体损失函数包含四个部分, 分类损失  $L_{cls}$ 、质心回归损失  $L_{centroid}$ 、边界框回归损失  $L_{bbox}$  以及密度图约束  $L_{den}$ 。

分类损失  $L_{cls}$ :  $L_{cls}$  采用的是 Focal loss<sup>[44]</sup> 的变化形式, 它被应用在无锚框方法<sup>[45-46]</sup> 中, 具体表达式如下:

$$L_{cls} = \frac{1}{N} \sum_{ijc} \begin{cases} (1 - \hat{Y}_{ijc})^\alpha \log(\hat{Y}_{ijc}), & Y_{ijc} = 1 \\ (1 - Y_{ijc})^\beta (\hat{Y}_{ijc})^\alpha \log(1 - \hat{Y}_{ijc}), & \text{其他} \end{cases} \quad (14)$$

其中,  $\hat{Y}$  为预测的热力图,  $Y$  为真值热力图,  $i, j, c$  表示

特征图三个维度的坐标,  $\alpha, \beta$  为超参数, 分别为 2 和 4。

$Y$  按式(9)方式产生:

$$Y_{ijc} = \exp\left(-\frac{(x-p_{kx})^2 + (y-p_{ky})^2}{2\sigma_p^2}\right) \quad (15)$$

其中,  $p_{kx}, p_{ky}$  为真值关键点横纵坐标, 在本文构建的仿真图像数据集与 SpotGEO 数据集标签中均包含目标质心坐标信息, 本方法中的  $p_{kx}, p_{ky}$  由质心坐标真值  $p_c = (p_{cx}, p_{cy})$  向下取整得到。 $\sigma_p$  是适应目标尺度的标准差<sup>[45]</sup>。

质心回归损失  $L_{\text{centroid}}$ :  $L_{\text{centroid}}$  采用 L1 损失, 具体表达式如下:

$$L_{\text{centroid}} = \frac{1}{N} \sum_{k=1}^N |\tilde{o} - (p_c - p_k)| \quad (16)$$

其中,  $\tilde{o} = (\tilde{o}_x, \tilde{o}_y)$  为关键点坐标与质心坐标的预测偏差值,  $p_c = (p_{cx}, p_{cy})$  为质心坐标真值,  $p_k = (p_{kx}, p_{ky})$  为关键点坐标。

边界框回归损失  $L_{\text{bbox}}$ :  $L_{\text{bbox}}$  采用 L1 损失, 具体表达式如下:

$$L_{\text{bbox}} = \frac{1}{N} \sum_{k=1}^N |\tilde{s}_k - s_k| \quad (17)$$

其中,  $\tilde{s}_k$  为预测的边界框坐标与质心坐标的差值,  $s_k = (p_{cx}^k - x_1^k, x_2^k - p_{cx}^k, p_{cy}^k - y_1^k, y_2^k - p_{cy}^k)$ ,  $x_1^k, y_1^k, x_2^k, y_2^k$  为边界框真值坐标,  $p_{cx}^k, p_{cy}^k$  为质心真值坐标。

整体的损失函数  $L_{\text{total}}$ :

$$L_{\text{total}} = \frac{1}{e^{W_{\text{cls}}}} L_{\text{cls}} + \frac{1}{e^{W_{\text{cen}}}} L_{\text{centroid}} + \frac{1}{e^{W_b}} L_{\text{bbox}} + \frac{1}{e^{W_d}} L_{\text{den}} + W_{\text{cls}} + W_{\text{cen}} + W_b + W_d \quad (18)$$

其中,  $W_{\text{cls}}, W_{\text{cen}}, W_b, W_d$  为可训练参数<sup>[47]</sup>, 使模型训练过程各个损失项的权重能够根据反馈动态调整, 从而自适应地平衡不同损失项的贡献, 避免了人工调参的复杂性。 $W_{\text{cls}}, W_{\text{cen}}, W_b, W_d$  的求和部分作为一个成本项, 避免了在训练过程中通过无限增大超参数  $W_{\text{cls}}, W_{\text{cen}}, W_b, W_d$  来降低总损失函数值的情况。

## 4 实验

### 4.1 数据集与评价指标

#### 4.1.1 数据集

##### (1) 仿真图像数据集

实际进行空间目标拍摄时, 由于不同拍摄模式与参数设置, 目标与恒星在图像中的形态分布可能极其相似也可能具有明显差异。而实拍图像数据集 SpotGEO 仅包含目标与恒星存在明显形态差异的

图像, 并且仅标注了质心坐标。这对于研究多种拍摄模式下的空间目标检测具有一定的局限性。因此, 本文构建了仿真图像数据集, 其训练集与测试集的图像均包含不同大小点状、不同长度条状的目标与恒星以及不同强度的噪声。仿真的依据是对公开的实拍空间图像的数据分析, 实拍空间图像包括: 美国国家航空航天局(National Aeronautics and Space Administration, NASA)公开的空间图像<sup>①</sup>以及欧洲空间局(European Space Agency, ESA)和阿德莱德大学发布的 SpotGEO 数据集中的图像。

数据集中共包含 4000 张仿真图像(如图 9 所示), 分辨率为  $512 \times 512$ , 其中训练集包含图像 3200 张, 测试集包含图像 800 张。此仿真图像数据集主要验证模型在单帧上将目标与恒星均作为候选目标进行检测与定位的能力, 因此仿真图像数据集中将点状、条状的目标与恒星均归为同一类目标进行了标注, 标注的具体内容包括其质心坐标、边界框坐标、光斑的所有像素点坐标。

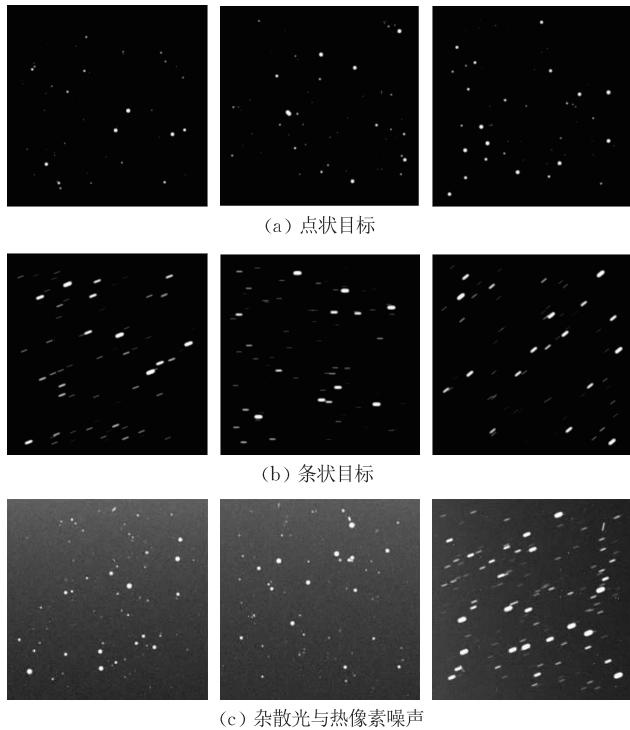


图 9 数据集中图像示例

#### (2) 实拍图像数据集 SpotGEO

SpotGEO 数据集是欧空局与阿德莱德大学在 2020 年组织的 SpotGEO 挑战赛中公开的空间图像序列数据集。其训练集和测试集分别包含 1280 个

<sup>①</sup> NASA image and video library, <https://images.nasa.gov/>

图像序列和 5120 个图像序列, 每个序列均包含 5 帧由地基望远镜连续拍摄的空间图像, 分辨率为 640 × 480。由于其拍摄模式, 图像中的恒星呈现为长拉线光斑, 目标呈现为点状光斑。此数据集标注了目标的质心坐标, 未对恒星进行任何标注。

#### 4.1.2 评价指标

评价指标包括对检测的准确性和全面性进行评价的精确率(precision)、召回率(recall)、F1 分数; 对质心定位精度进行评价的平均质心定位误差(mean error)、均方误差MSE(Mean Square Error)。本文的数据集评价指标为精确率、召回率、F1 分数和平均质心定位误差。SpotGEO 数据集使用官方评价指标 1-F1 与定位均方误差, 使用官方验证代码<sup>①</sup>进行指标计算。上述评价指标定义如下。

精确率定义为

$$P = \frac{\sum_i^N TP_i}{\sum_i^N (TP_i + FP_i)} \quad (19)$$

其中,  $TP_i$  为第  $i$  帧检测结果中真正例(True Positive)的个数,  $FP_i$  为第  $i$  帧检测结果中假正例(False Positive)的个数。

召回率定义为

$$R = \frac{\sum_i^N TP_i}{\sum_i^N (TP_i + FN_i)} \quad (20)$$

其中,  $TP_i$  为第  $i$  帧检测结果中真正例的个数,  $FN_i$  为第  $i$  帧检测结果中假负例的个数。

F1 分数定义为

$$F_1 = 2 \times \frac{P \times R}{P + R} \quad (21)$$

其中,  $P$  与  $R$  分别为精确率与召回率。

平均质心定位误差定义为

$$E_c = \frac{1}{K} \sum_i^K \sqrt{(p_{cx_i} - x_i)^2 + (p_{cy_i} - y_i)^2} \quad (22)$$

其中,  $p_{cx_i}$  与  $p_{cy_i}$  为第  $i$  个目标的质心坐标真值,  $x_i$  与  $y_i$  为第  $i$  个目标的质心坐标预测值,  $K$  真正例个数。

质心定位均方误差按照 SpotGEO 官方文档定义为

$$MSE = \frac{\sum_{k=1}^K \sum_{f=1}^5 SSE^f}{\sum_{k=1}^K (TP_k + FN_k + FP_K)} \quad (23)$$

其中,  $f$  为序列中的帧序号,  $K$  为序列数,  $SSE^f$  为  $f$

帧的质心定位误差, 定义为

$$SSE^f = \sum_{(i,j) \in TP^f} \pi(x_i^f, y_j^f) + \sum_{j \in FN^f} \tau^2 + \sum_{i \in TP^f} \tau^2 \quad (24)$$

$\pi(x_i^f, y_j^f)$  为

$$\pi(x_i^f, y_j^f) = \begin{cases} 0, & \|x_i^f - y_j^f\|_2 \leq \epsilon \\ \|x_i^f - y_j^f\|_2^2, & \text{其他} \end{cases} \quad (25)$$

$x_i^f$  与  $y_j^f$  分别为质心坐标的预测值和真值,  $\tau$  与  $\epsilon$  均为常数。

#### 4.2 实验实施细节

本文基于 Pytorch 框架编写代码以完成模型的构建, 模型训练与测试所采用 GPU 的型号为 NVIDIA A100-SXM4-40G。在构建的仿真图像数据集与实拍图像数据集 SpotGEO 上进行训练之前未进行任何的预训练, epoch 均设置为 100, 学习率设置为 0.0001, 采用 Adam 优化器提升模型的收敛速度, 训练与测试的 batch size 均设置为 8。

#### 4.3 实验结果

##### 4.3.1 仿真图像数据集检测结果

应用仿真图像数据集验证模型在单帧图像上的目标检测能力。对比方法包括文献[5]、文献[6]在单帧检测阶段分别使用的 PP-Yolov2 和 EfficientNet 结合 FPN 检测模型, 还包括其他性能优秀的目标检测方法<sup>[24-25, 27, 48-50]</sup>。同时, 由于空间目标检测领域众多方法是基于分割进行单帧目标提取, 因此, 对比方法中还包含分割方法, 具体为: 非深度学习的自适应局部阈值分割方法(Adaptive Local Threshold Segmentation, ALTS)<sup>[10]</sup>与基于 Unet 的空间目标分割方法 StarNet<sup>[35]</sup>。当某方法设计了多个模型时(如 Yolov10 包含 N、S、M、B、L、X 多个模型), 本文选择其中检测结果最佳的模型进行对比。另外, 本文在仿真图像数据集上设置了消融实验, 测试本文方法在移除 3.2 节与 3.3 节的模块后的检测结果与不使用 3.4 节的密度图损失的检测结果。为了比较各方法对不同形态目标的检测能力差异, 分别统计了目标中点状、条状目标的结果。检测结果如表 2 所示。

由表 2 可以看出, 本文方法在仿真图像数据集上取得的精确率为 96.95%、召回率为 93.78%、F1 分数为 95.34%, 各项指标均优于其他方法, 特别是在召回率方面优势更加明显, 这得益于本文方法针对空间目标的专门设计, 从而检测出了更多弱小

<sup>①</sup> SpotGEO Starter Kit. <https://doi.org/10.5281/zenodo.3874368>

表 2 本文方法与其他方法在仿真图像数据集的检测结果

方法	所有目标			点状目标			条状目标		
	精确率/%	召回率/%	F1 分数/%	精确率/%	召回率/%	F1 分数/%	精确率/%	召回率/%	F1 分数/%
ALTS	53.56	65.07	58.76	50.56	59.39	54.62	56.07	70.11	62.31
StarNet	92.04	86.78	89.33	93.77	87.49	90.52	90.54	86.15	88.29
EfficientNet+FPN	91.18	81.51	86.07	91.26	74.38	81.96	91.12	87.85	89.45
PP-Yolov2-Res101	93.10	82.88	87.69	93.14	76.23	83.84	93.07	88.79	90.88
Yolov10	95.22	85.17	89.91	<b>95.28</b>	79.02	86.39	95.18	90.71	92.89
DiffusionDet	96.06	86.94	91.27	93.56	82.04	87.42	98.44	91.63	94.92
RTMDet-L-Swin-B-P6	96.04	82.42	88.71	93.39	72.73	81.77	98.01	91.03	94.39
Deformable-DETR	95.55	81.33	87.87	92.84	81.77	86.95	98.13	80.93	88.70
DINO-SwinL	93.35	88.41	90.81	88.62	81.80	85.07	97.36	94.29	95.80
DDQ-DETR-SwinL	94.20	89.04	91.55	90.17	82.92	86.39	97.61	94.48	96.02
本文方法 (无3.2节、3.3节模块)	94.24	92.60	93.41	93.51	91.91	92.70	94.89	93.22	94.05
本文方法 (无密度图损失)	96.18	92.93	94.53	94.21	92.08	93.13	97.96	93.69	95.78
本文方法 (SENet替换3.2节模块)	96.22	92.87	94.51	94.23	91.93	93.06	98.02	93.72	95.82
本文方法 (CBAM替换3.2节模块)	96.39	93.26	94.80	94.36	92.39	93.36	98.23	94.03	96.08
本文方法	<b>96.95</b>	<b>93.78</b>	<b>95.34</b>	94.69	<b>92.65</b>	<b>93.66</b>	<b>99.01</b>	<b>94.78</b>	<b>96.85</b>

注：粗体数字表示最优结果；其他表同理。

的空间目标。而其他深度学习检测方法设计时的作用对象往往相较于空间目标尺寸更大，与背景对比更明显，导致此类方法对弱小的空间目标漏检较多。PP-Yolov2 和 EfficientNet+FPN 虽针对空间目标检测，但其方法是通用目标检测模型的模块组合，并未针对空间目标的性质进行相应的修改，因此同样存在较多弱小空间目标漏检。传统的区域自适应阈值分割方法 ALTS 取得的结果明显低于其他方法，这是由于本文的仿真图像数据集中的图像包含各种背景噪声，而传统方法往往需要额外的预处理算法先对背景噪声进行抑制。因此，直接处理包含杂光背景与热像素噪声图像的效果不佳。本文

还展示了不同形态目标的检测结果，本文方法除点状目标精确率以外的其他指标均为最优。点状目标的精确率为 94.69%，与 Yolov10 方法相近，召回率高于 Yolov10 近 14%。此外，本文方法在移除 3.2 节与 3.3 节设计的模块后，F1 分数下降了 1.93，在不使用密度图损失的情况下，F1 分数下降了 0.81，结果表明本文设计模块与引入密度图损失的有效性。

放大图像的局部区域对检测结果进行可视化。图 10、图 11 分别展示了稀疏区域与密集区域的检测结果，图中红色方框代表真值，绿色方框代表检测结果。对于分割的方法，在结果图中也标出了真值边界框，以便更加直观地展示漏检、误检的情况。

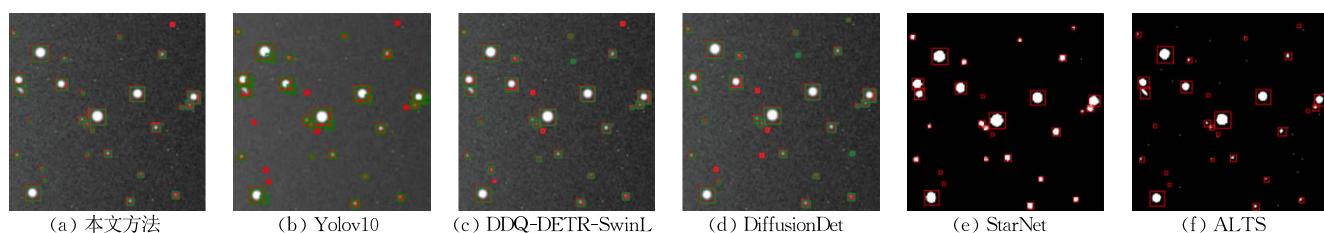


图 10 稀疏区域目标检测结果可视化

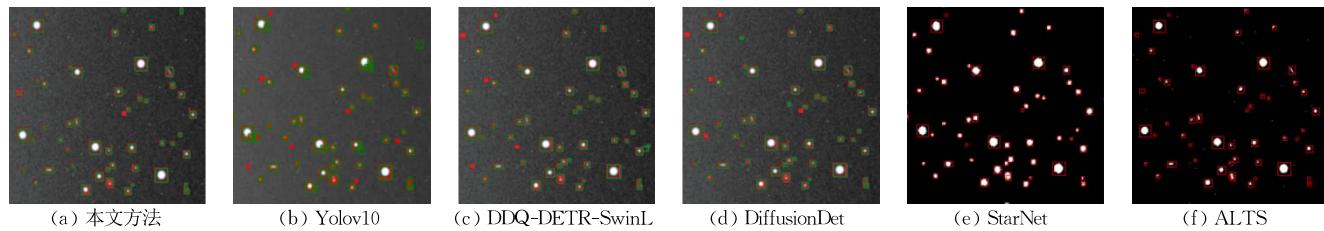


图 11 密集区域目标检测结果可视化

由可视化结果可以看出,所有方法对于较大的目标均取得了不错的检测效果。对于尺寸较小的目标,本文方法仍然检测出了其中的绝大多数,而其他深度学习方法漏检了较多弱小目标,ALTS 的方法在杂光背景下对很多弱小目标失效。

#### 4.3.2 仿真图像数据集质心定位结果

在质心定位方面,本文所提出的模型端对端直接输出目标质心坐标,但 StarNet 方法与 ALTS 方法还需要结合质心定位方法确定分割出的光斑的

质心坐标。因此,本文引入基于能量累加的质心定位方法(Energy Accumulation, EA)<sup>[51]</sup>对 StarNet, ALTS 的光斑进行质心定位,进而比较质心定位结果。同时,在 Spot & Line 数据集上设置了消融实验以验证本文所设计模块与引入密度图损失的有效性,测试本文方法在移除 3.2 节与 3.3 节的模块后的检测结果与不使用 3.4 节的密度图损失的质心定位结果。质心定位结果如表 3 所示。

表 3 本文方法与其他方法的质心定位精度

方法	所有目标			点状目标			条状目标		
	平均质心定位误差	定位误差标准差	真正例数	平均质心定位误差	定位误差标准差	真正例数	平均质心定位误差	定位误差标准差	真正例数
ALTS+EA	0.4776	0.8221	60363	<b>0.4133</b>	0.5015	25927	0.5260	1.0427	34436
StarNet+EA	0.4990	0.5294	80508	0.4528	0.4574	38194	0.5406	0.5776	42314
本文方法 (无 3.2 节、3.3 节模块)	0.4738	0.3525	85910	0.4638	0.2587	40123	0.4914	0.4324	45787
本文方法 (无密度图损失)	0.4658	0.3132	86215	0.4415	0.2295	40197	0.4870	0.3730	46018
本文方法	<b>0.4478</b>	<b>0.2951</b>	<b>86999</b>	0.4258	<b>0.2221</b>	<b>40446</b>	<b>0.4669</b>	<b>0.3460</b>	<b>46553</b>

由表 3 可知,所有的方法质心定位精度均达到了亚像元级,条状目标的定位误差普遍高于点状目标,这是条状目标更强的运动模糊造成的。本文方法对于所有目标的平均质心定位误差为 0.4478,优于其他方法。对于点状目标,本文方法质心定位误差略高于 ALTS 结合 EA 的方法,但是本文方法的检测结果中点状目标的真正例数为 40446 个,ALTS 方法为 25927 个,本文方法在多检测出了 14519 个目标的情况下平均质心定位误差仅高出约 0.0125 个像元。为了验证方法的稳定性,表 3 罗列了各方法的质心定位误差标准差。结果显示,本文方法的标准差最小。综上,本文提出的直接预测质心位置的模型在检测精度与稳定性方面具有显著优势。此外,消融实验结果显示,当移除本文设计模块或不使用密度图损失时,质心定位误差与标准差有所增加,证实了本文设计的模块与引入的密度图损失能够有效提升质心定位精度。

对粘连目标质心定位结果进行可视化。如图 12 所示,红色代表质心坐标真值,绿色代表本文方法得出的质心位置,蓝色和黑色分别代表 StarNet+EA 与 ALTS+EA 方法得到的质心坐标。

由图 12 可以看出,对于粘连的目标,本文方法直接输出了粘连目标各自的质心坐标。而分割的方法只输出了一个质心坐标结果,这是因为分割方法

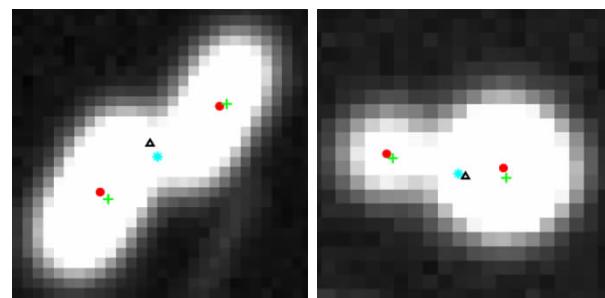


图 12 粘连目标质心定位结果可视化

将粘连目标分割成了一块光斑,从而使后续质心定位算法对整块粘连的光斑进行质心位置预测,这与实际情况严重不符。

为验证不同强度噪声对本文方法的质心定位精度影响,分别在轻度噪声、中度噪声和重度噪声的图像(见图 13)中进行实验,噪声类型包括背景噪声、杂散光、热像素。通过噪声与能表示的灰度上限的比值量化背景噪声与杂散光强度,通过数量对热像素进行量化,如表 4 所示。误差与标准差结果见表 5。



图 13 轻度至重度噪声干扰图像

表 4 各类噪声强度或数量统计

噪声强度	背景噪声 (与灰度上限 比值区间)	杂散光 (与灰度上限 比值区间)	热像素 (数量区间)
轻度噪声	[0.01, 0.02]	[0.05, 0.1]	数量: [70, 90]
中度噪声	[0.02, 0.03]	[0.1, 0.15]	数量: [90, 110]
重度噪声	[0.03, 0.04]	[0.15, 0.25]	数量: [110, 130]

表 5 不同强度噪声对质心定位精度影响

噪声强度	本文方法结果	
	定位误差	标准差
轻度噪声干扰	0.4269	0.2772
	0.4313	0.2854
中度噪声干扰	0.4558	0.3219
	0.4432	0.3304

中度噪声干扰的情况下,本文方法的质心定位误差及标准差相较于轻度噪声干扰的增幅不明显。当重度噪声干扰时,质心定位误差和标准差较明显增大,表明高强度的噪声会降低定位精度。具体对模型的影响在于重度噪声严重改变目标的分布特性,而训练时模型的学习是兼顾各种数据的过程,重度噪声干扰下模型难以极好地完成对数据的拟合。但即使在重度噪声环境下,本文方法的质心定位误差仍然维持在 0.5 像元以内,体现出本文方法对含噪图像中目标质心定位的优良能力与较强鲁棒性。

本文引入一种后处理方法<sup>[52]</sup>进一步提升质心定位精度与稳定性,该方法在本文模型输出的边界框和质心坐标基础上使用,后处理流程如图 14 所示,质心定位结果如表 6 所示。

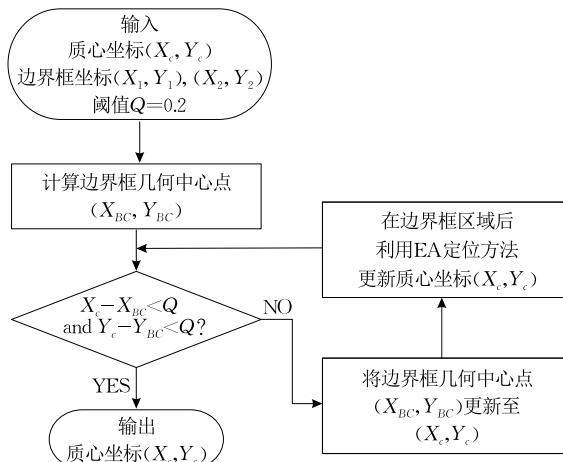


图 14 后处理方法流程图

由表 6 结果可知,后处理方法对本文原始结果精度有一定提升,说明引入的后处理方法能够有效优化本文方法的结果;同时也说明了本文方法对后处理方法具有良好的可扩展性。在实际应用中,可

以根据对精度和稳定性的具体要求,灵活决定是否采用后处理方法。

表 6 结合后处理方法质心定位结果

不同数据	本文方法结果	增加后处理
原仿真数据集	定位误差 0.4478	0.4012
	标准差 0.2951	0.2817
轻度噪声干扰	定位误差 0.4269	0.3926
	标准差 0.2772	0.2674
中度噪声干扰	定位误差 0.4313	0.4104
	标准差 0.2854	0.2807
重度噪声干扰	定位误差 0.4558	0.4432
	标准差 0.3219	0.3304

#### 4.3.3 计算效率分析

实验在配备 GPU 与无 GPU 的硬件环境下进行。在配备 GPU 的硬件环境中,对 4.3.1 节基于深度学习的代表性方法的模型参数量(Params)、浮点运算次数(FLOPs),以及处理速度(Speed)进行了统计,各项设置保持与 4.3.1 节中对应方法相同。此外,在不同计算能力的 NVIDIA GPU 上对本文方法的处理速度进行了统计。在无 GPU,仅配备 CPU 的硬件环境下测试了本文方法与传统方法的处理速度。从而全面评估在各种硬件配置下的方法性能,测试结果如表 7 所示。

表 7 本文方法与其他方法的计算效率

配备 NVIDIA GPU 硬件环境				
方法	硬件环境	Params/M	FLOPs/G	Speed/fps
StarNet	A100 SXM4	15.16	105.74	45
		47	112	53
		73	61	79
		31	170	87
EfficientNet+FPN	1660ti	40	126	27
PP-Yolov2-Res101		218	860	11
Yolov10-X				48
Deformable-DETR				12
DDQ-DETR-SwinL	2080ti	55	251	19
本文方法				28
本文方法				
本文方法				
无 GPU, 仅 CPU 硬件环境				
方法	硬件环境	Speed/fps		
ALTS+EA	Intel(R) i7 12700KF	11		
本文方法		2		

在配备 GPU 的硬件环境下,本文方法在额外完成质心定位功能的情况下仍然展现了良好的推理速度,最高每秒处理图像 48 帧。在算力较低的 1660ti GPU 上每秒处理图像数仍超过 10 帧。计算资源需求方面,地基处理平台往往不存在资源受限情况,可以提供与本文相当或更好的硬件资源。天基平台方面,目前的 NVIDIA Jetson 系列平台配备 GPU 并具备 CUDA 支持,其中部分型号的算力与本文硬件条件相当甚至更强大(如 NVIDIA Jetson

AGX Orin 系列)<sup>①</sup>,能够满足本文方法的硬件要求。在无 GPU 的 CPU 硬件环境下,传统方法因不涉及大量的张量运算,具备处理速度的优势。

跨空间-通道注意力模块与压缩-激励的多尺度特征融合模块的计算开销如表 8 所示。

表 8 本文模块的参数量与浮点运算次数统计表

模块	Params/M	FLOPs/G
跨空间-通道注意力模块	42.93	0.433
压缩-激励的多尺度特征融合模块	0.11	1.6216

跨空间-通道注意力模块的总参数量 42.93 M, 总 FLOPs 为 0.433G。参数量较大是由于其全连接层的输入为空间维度特征展平后与通道特征拼接而成,元素数较多,全连接层需要设置较大的参数矩阵完成元素间的交互。但是其计算是较简单的全局化的线性变换,这样整体性的计算不会使 FLOPs 极大增长。压缩-激励融合总参数量为 0.11 M,额外引入参数量较小。总 FLOPs 为 1.6216 G,这是由于该模块需要处理和融合来自不同尺度的特征图,涉及大尺度特征图与较多的卷积运算,即在特征图上局部重复地滑动小的卷积核、进行密集的计算。

实际模型选择时,建议根据具体的应用场景需求权衡性能和计算资源。例如,在地基平台这类几乎无计算资源限制的场景或者天基平台采用的硬件设备包含超过 1660ti 算力的 GPU 时,建议使用本文方法或其他深度学习方法。当不具备 GPU 条件时,建议选择传统的检测方法。

#### 4.3.4 SpotGEO 检测结果

根据 SpotGEO 官方文献[7]描述,一些目标在序列中的部分帧上由于被完全遮挡或过于暗弱而不可见,但标注中仍然具有这些目标的信息,这意味着需要检测出这些目标并且只能利用这类目标在序列中其他帧上的信息。SpotGEO 挑战赛排行榜中的方法在单帧候选目标提取后结合图像序列信息进行后处理得到最终的检测结果。因此,本文在 SpotGEO 数据集上验证时,同样在单帧检测后使用后处理方法<sup>[6]</sup>,通过运动轨迹的直线拟合对检测结果进行补充和过滤,从而与其他在 SpotGEO 数据集上验证的方法进行公平的对比。

SpotGEO 数据集中恒星为条状光斑,目标为点状光斑,具有明显形态差异。并且真值中无恒星标注,仅标注了目标的质心坐标。因此,在此数据集上移除了模型的边界框回归分支并重新训练模型,先在单帧图像中检测目标光斑,然后对结果进行后处

理优化。

本文在表 9 中列出了 SpotGEO 挑战赛官方排行榜的前 10 名与本文方法的 F1 分数和 MSE 结果。

表 9 SpotGEO 官方排行榜

排名	方法	F1 分数/%	MSE
1	AgeniumSPACE	94.83	33838.9931
2	POTLAB@BUAA	94.43	30541.73189
3	dwiuzila	92.89	41198.45863
4	Magpies	90.43	48919.9227
5	Mr_huangLTZaaa	88.42	62021.80923
6	francescodg	87.89	65772.46337
7	mhalford	87.70	69566.90857
8	PedroyAgus	86.61	70104.96654
9	elmihailol	86.11	83172.81408
10	Barebones	83.66	105518.4199
—	本文方法+后处理	93.08	40764.3107

如表 9 所示,本文方法结合后处理最终取得的 F1 分数为 93.08%,MSE 为 40 764.3107。在官方排行榜中排名第三。根据官方文献,排名第一的 AgeniumSPACE 团队采用了多阶段的处理流程:首先使用 L1-Spline 对图像进行预处理,消除复杂背景噪声的影响;之后结合手工设计的恒星特征描述和 RANSAC 算法估计恒星帧间位移,以实现恒星抑制;然后,训练了一个集成多 U-Net 的模型在抑制恒星后的图像上检测目标;最后,又通过线检测和轨迹填充进行后处理。排名第二的 POTLAB@BUAA 团队采用非学习型多阶段方法:首先,逐像素计算信噪比并选择高于人工设定的阈值的像素;其次,对选出的像素进行连通区域标记确定候选对象;之后,估计帧间恒星位移从而过滤掉候选对象中的恒星;然后,再过滤掉恒星的候选对象中进行帧间目标位移估计;最后,将满足此估计的候选对象确定为目标。尽管 AgeniumSPACE 与 POTLAB@BUAA 团队在指标上取得了更好的结果,但是本文方法不需要任何图像预处理或检测空间目标前抑制恒星的步骤,处理流程更加简洁直接的同时也取得了较好的检测结果。

对 SpotGEO 检测结果进行可视化,如图 15 所示。为更清楚展示结果,本文以质心坐标为圆心生成圆圈代表目标位置,红色圆圈表示目标真值,绿色圆圈表示本文方法的检测结果。图 15 清楚展示了本文方法在 SpotGEO 数据集的实拍空间图像中对于弱小目标、与恒星粘连目标以及高亮或极亮背景中的目标仍然具有良好的效果。

① <https://developer.nvidia.cn/embedded/jetson-modules>

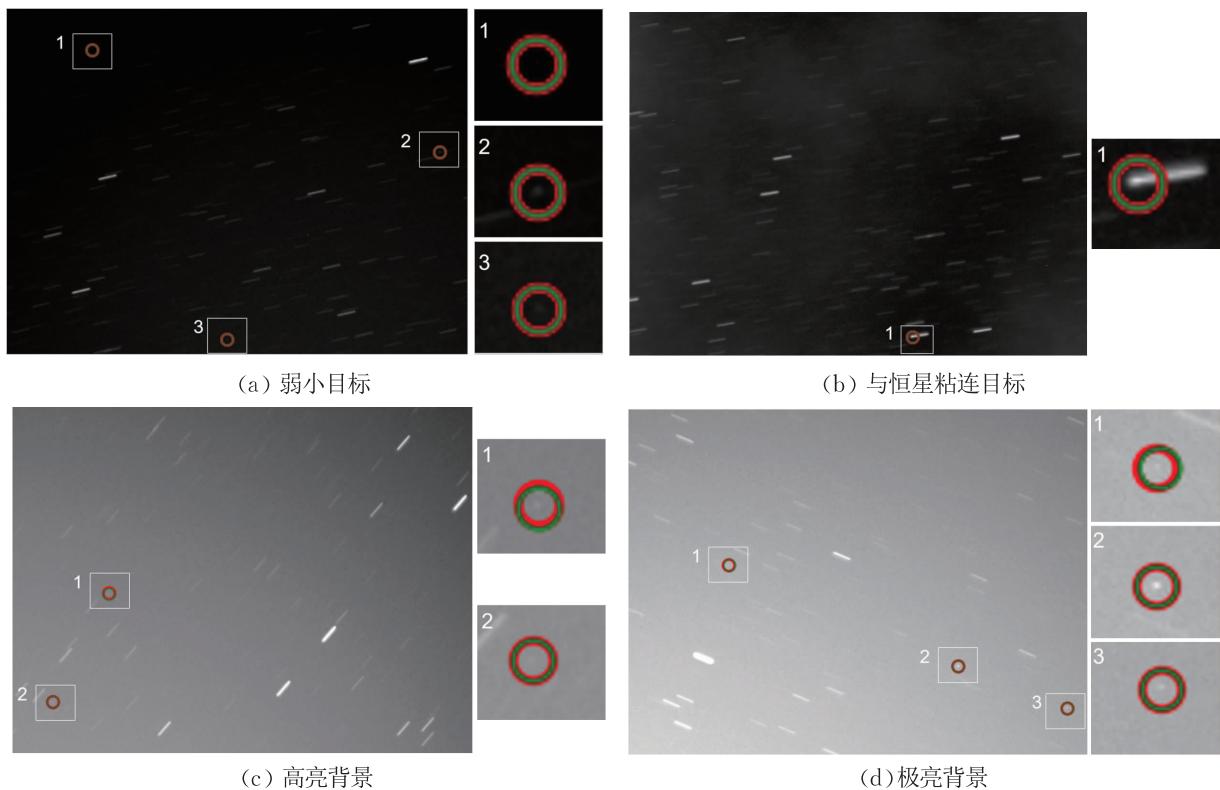


图 15 SpotGEO 数据集检测结果可视化

#### 4.4 参数讨论

##### 4.4.1 分类损失超参数

在仿真数据集和 SpotGEO 数据集测试了分类损失超参数  $\alpha, \beta$  的多组取值, 各组结果如表 10 和表 11 所示。

表 10 仿真数据集上不同超参数选择的结果

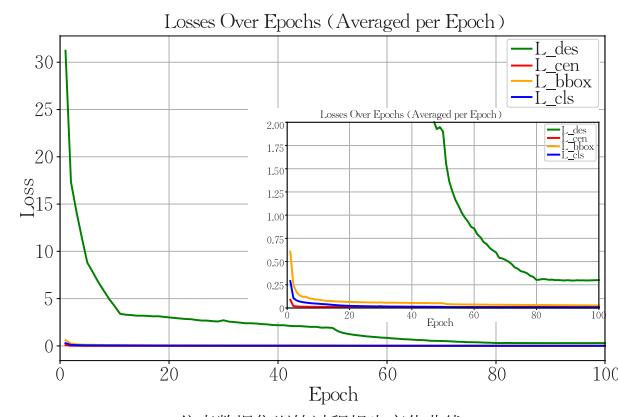
$\alpha$	$\beta$	F1 分数/%
1	2	94.37
1	3	94.58
1	4	94.76
1	5	94.61
2	4	<b>95.34</b>
3	4	94.88

表 11 SpotGEO 数据集上不同超参数选择的结果

$\alpha$	$\beta$	F1 分数/%
1	2	92.04
1	3	92.42
1	4	92.77
1	5	92.46
2	4	<b>93.08</b>
3	4	92.85

##### 4.4.2 损失函数变化情况

总损失函数中的超参数为可学习的参数, 图 16 和图 17 分别展示了在仿真数据集和 SpotGEO 数据集训练时, 各项损失及其权重的变化情况。



(a) 仿真数据集训练过程损失变化曲线

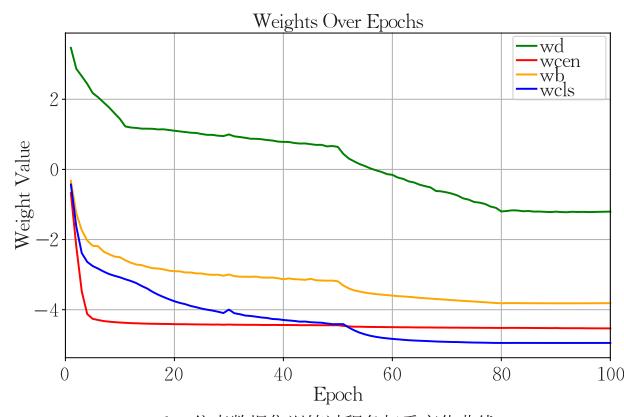
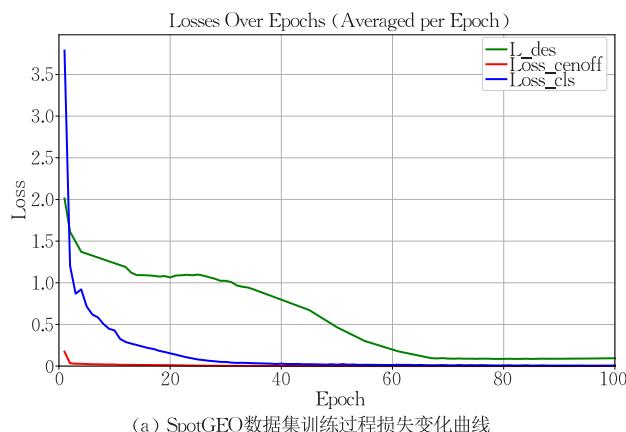
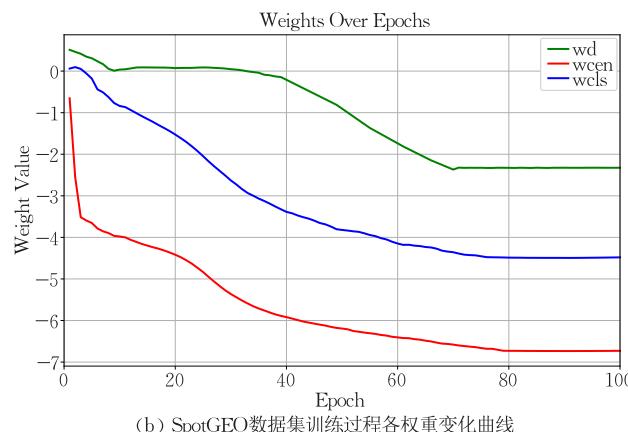


图 16 仿真数据集损失及其权重变化情况



(a) SpotGEO 数据集训练过程损失变化曲线



(b) SpotGEO 数据集训练过程各权重变化曲线

图 17 SpotGEO 数据集损失及其权重变化情况

由于密度图损失在训练开始阶段较大,直接与其他损失曲线共同展示难以明确看出其他损失的变化情况,因此在图 16(a)中放大了纵轴的 0 至 2 范围。SpotGEO 数据集中,由于移除了边界框回归部分,因此其只有三条曲线。由图 16、图 17 可以看出,各项损失最终均收敛于较小的值,说明模型已经很好地完成了训练过程。

## 5 结 论

本文提出了一种基于信息表征增强的无锚框空间弱小目标检测方法,相比传统方法和现有基于深度学习的检测方法,本文方法利用端到端卷积神经网络,实现了空间弱小目标检测和质心定位的同步处理,直接输出目标质心坐标,从而避免了检测与定位分阶段处理所带来的误差累积。通过设计跨空间-通道注意力模块和压缩-激励的多尺度特征融合模块,提升了模型对含噪图像中弱小目标有效信息的关注度。在此基础上,引入了基于密度图的损失在整体目标数量层面对模型进行约束,进一步解决了光斑粘连遮挡的难点问题。在本文仿真的空间图像与 SpotGEO 实拍图像数据集中均进行了实验验证,本文方法在仿真图像数据集上的 F1 分数、平均质心定位误差优于其他方法。在 SpotGEO 数据集上,本文方法在不进行预处理和恒星抑制的情况下,结合后处理方法取得了 93.08% 的 F1 分数。

本文关注的重点是单帧空间图像中检测与定位目标,并证实了在噪声影响的情况下仍然实现了高精度的质心定位。然而,在实际拍摄过程中,可能有部分目标完全被恒星遮挡或湮没于复杂背景,从而在某帧图像中不可见。为解决这一问题,目前本文

的工作是结合后处理方法,实现了对这些特殊目标的检测和定位。在未来的工作中,我们将致力于研发基于深度学习的端到端空间目标检测与跟踪模型,针对空间图像序列,通过有效整合单帧中目标的灰度、大小等静态信息与序列中的运动信息,不仅能够检测上述不可见目标,还能够对目标进行有效跟踪,实现“一步到位”的目标检测、定位与编号。这样的设计将无需再依赖后处理算法,提升模型的实用性和灵活性,并进一步提高模型的检测精度和鲁棒性。

## 参 考 文 献

- [1] Mejía-Kaiser M. The Geostationary Ring. Hague, The Netherlands: Brill Nijhoff, 2020
- [2] Sun R, Zhan J, Zhao C, et al. Algorithms and applications for detecting faint space debris in GEO. Acta Astronautica, 2015, 110: 9-17
- [3] Sun R Y, Zhao C Y. A new source extraction algorithm for optical space debris observation. Research in Astronomy and Astrophysics, 2013, 13(5): 604-614
- [4] Liu D Q, Chen B, Chin T J, et al. Topological sweep for multi-target detection of geostationary space objects. IEEE Transactions on Signal Processing, 2020, 68: 5166-5177
- [5] Dai Y, Zheng T, Xue C, et al. Effective multi-frame optical detection algorithm for GEO space objects. Applied Sciences, 2022, 12(9): 4610-4626
- [6] Abay R, Gupta K. GEO-FPN: A convolutional neural network for detecting GEO and near-GEO space objects from optical image//Proceedings of the 8th European Conference on Space Debris. Darmstadt, Germany, 2021: 123-129
- [7] Chen B, Liu D Q, Chin T-J, et al. Spot the GEO satellites: From dataset to Kelvins SpotGEO challenge//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Nashville, USA, 2021: 2086-2094