

在轨光学目标检测模型轻量化研究

吕晓宁¹, 夏玉立¹, 赵军锁^{1,2}, 乔鹏¹

1. 中国科学院软件研究所 天基综合信息系统重点实验室, 北京 100190;

2. 中国科学院大学, 北京 100049

摘要:目前应用广泛的神经网络模型结构复杂、参数量大,对星上有限的计算和存储资源占用较多。本文面向微纳卫星在轨计算平台,提出一种深度可分离卷积神经网络模型,该模型结合反向残差结构与通道注意力的思路改进在轨识别算法 YOLOv4 网络模型结构,改进网络模型的局部模块结构,降低整体网络结构的深度与复杂度;同时利用可分离卷积结构实现空间卷积,改进 SPP 模块与 PANet 模块,降低模型参数量;通过合并卷积层与 Batch Normalization 层,进一步加快前向推理速度;此外借鉴 Focal 损失函数的思想,改进目标检测网络的损失函数,减缓前景与背景样本比例不均衡问题。通过与原 YOLOv4 网络模型对比,在保证识别精度 94.09% 的前提下参数量降低约 7 倍, FLOPs 降低约 30 倍。同时与 YOLO 系列、SSD、MobileNet、CenterNet 等前沿网络模型对比的实验结果再一次验证了本文的算法性能,为实现在轨目标识别与过滤无用数据提供了理论支撑。

关键词: 遥感, 卷积神经网络, YOLOv4, 在轨目标检测, 反向残差结构

中图分类号: TP391.4/P2

引用格式: 吕晓宁,夏玉立,赵军锁,乔鹏.2024.在轨光学目标检测模型轻量化研究.遥感学报,28(4): 1041-1051

Lyu X N, Xia Y L, Zhao J S and Qiao P. 2024. Lightweight model for On-Orbit optical object detection. National Remote Sensing Bulletin, 28(4): 1041-1051 [DOI:10.11834/jrs.20221556]

1 引言

光学遥感图像中的目标检测 (Han 等, 2020) 是确定图像中是否包含一个或多个目标,进而确定每个目标的位置。遥感图像中的飞机作为重要的运输载体和军事目标,其检测在飞机救助、战争预警等领域具有重要意义 (Dong 等, 2020; Liu 等, 2013)。因此,基于光学遥感图像中的飞机检测成为一个热点问题。

然而飞机探测平台多为卫星与机载平台,计算能力和存储能力有限 (Hasan 等, 2010),无法支持高能耗的神经网络模型在轨目标检测,在轨检测准确率受限。目前,遥感目标检测主要由计算能力强大的地面处理器完成,但遥感图像在空间—地面传输链路带宽受限,从卫星载荷成像到飞机等目标检测过程中耗时较多,无法满足实际应用需求 (Yu 等, 2016)。

近年来,深度学习在特征提取方面 (Chen 等, 2017) 取得了较大成就,目标检测具有较强的鲁棒性,可以适应不同的情况。基于神经网络模型的目标检测主流算法分为 Two-Stage 和 One-stage (Tang 等, 2015) 两种类型。Two-Stage 具有代表性的算法有 Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren 等, 2017) 等,该算法第 1 阶段通过搜索遍历图像像素,提取预选的目标框,第 2 阶段基于当前目标框进一步特征提取,精细化目标特征,完成检测任务,这类算法检测精度虽高,但效率较低;One-Stage 具有代表性的算法有 SSD (Liu 等, 2016), YOLOv3 (Redmon 等, 2016), YOLOv4 (Bochkovskiy 等, 2020), YOLOv5 (Wang 等, 2021), CornerNet (Law 等, 2018), CenterNet (Zhou 等, 2019) 等,相较于 Two-Stage 系列算法,One-Stage 系列算法直接在图像中提取目标特征与目标区域,避免了区域候选阶段,同时兼顾检测

收稿日期: 2021-08-20; 预印本: 2022-04-11

基金项目: 国家自然科学基金(编号:62027801)

第一作者简介: 吕晓宁,研究方向为遥感图像处理。E-mail: 746166746@qq.com

通信作者简介: 夏玉立,研究方向为遥感图像处理。E-mail: xiayuli@iscas.ac.cn

速度与检测精度。而星上在轨检测飞机目标受计算能力、存储能力以及模型复杂度的限制导致实时检测效果不佳,为进一步提高检测速度并兼顾检测性能,需降低模型的参数与计算量。国内外研究者相继提出了 MobileNet-SSD (Biswas 等, 2019)、Yolov3-Tiny、Yolov4-Tiny、Yolov5s 等轻量级神经网络,通过优化计算结构来减少计算量。然而,上述网络模型大多针对数据量较低、目标较明显的可见光图像,例如 VOC 数据集等,而遥感图像数据量巨大,飞机目标尺寸相对于背景所占比例较小,模型对遥感目标的检测性能较低,无法满足在轨任务需求。考虑卫星在轨检测的效率与精度,本文借鉴反向残差结构结合 SENet (Roy 等, 2018) 的通道注意力思想,改进在轨识别算法 Yolov4 网络模型的主干网络,并调整网络

模块的层数,使用较少的参数来生成更多的特征,减少网络间特征冗余,压缩神经网络整体框架,基于深度可分离卷积结构实现空间卷积替换 SPP 结构与 PANet 结构的普通卷积,实现模型参数量减少,采用多尺度融合方式,有效检测遥感目标。

2 Yolov4 算法原理

Yolov4 算法主要包括 3 个部分: Backbone 主干网络结构、Neck 模块、Yolo Head 预测模块。当输入的图片分辨率为 608×608 时,具体的网络框架如图 1 所示。Backbone 网络结构主要使用 CSPDarkNet53 提取模型特征; Neck 模块主要包括 SPP 结构和 PANet 结构,搭建图像目标特征金字塔结构,进一步加强特征提取网络; Yolo Head 预测模块利用前面模块提取的特征完成目标检测任务。

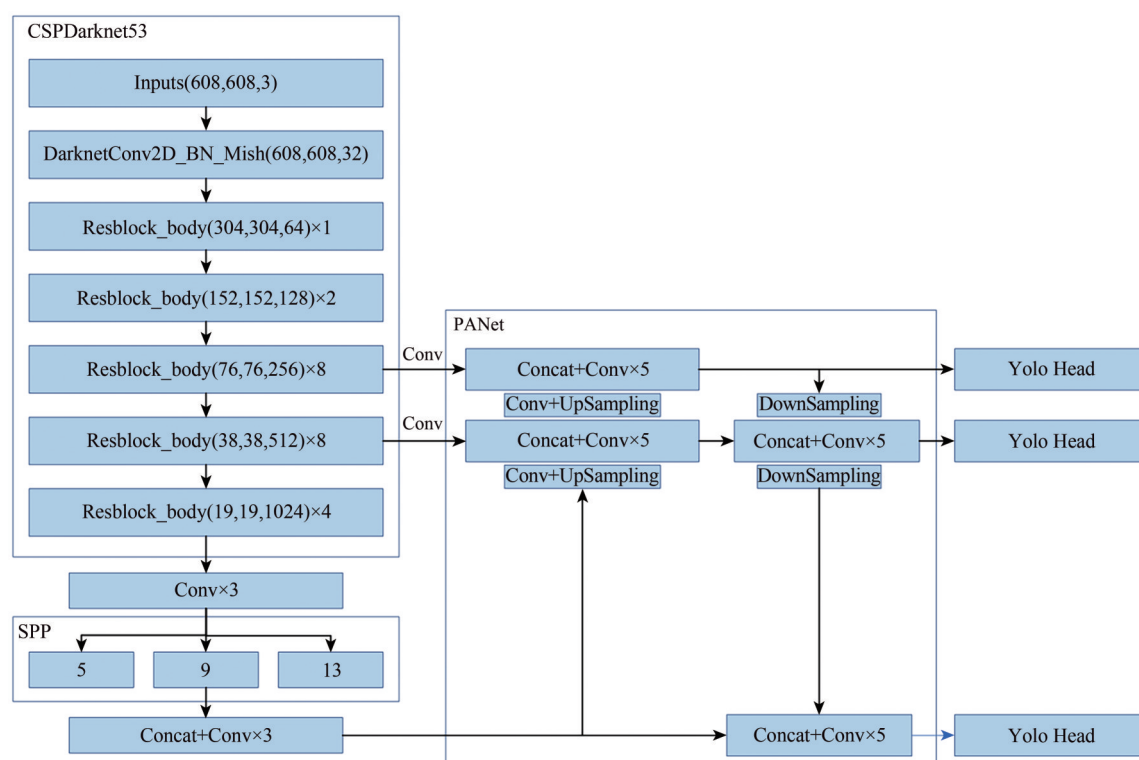


图1 Yolov4网络模型架构

Fig. 1 Yolov4 network model architecture

CSPDarkNet53 网络基于 Mish 激活函数与 CSPNet 结构对 DarkNet53 网络进行改进构建的,其中包含 53 个卷积层。输入图像分辨率为 608×608 时,首先经过一个 $3 \times 3 \times 32$ 大小的滤波器,再经过 5 组 ResBlock_body 模块进行特征提取,其中 5 组 ResBlock_body 单元,每个单元由单个卷积层和一系列重复执行的卷积层组成,重复次数分别为 1、

2、8、8、4。

ResBlock_body 模块为主干网络中重要组成模块,采用了 CSPNet 结构,主要提取图像的目标特征,同时降低计算瓶颈,构造残差边部分,避免模型出现梯度消失问题,具体内部 Module 架构如下图 2 所示。

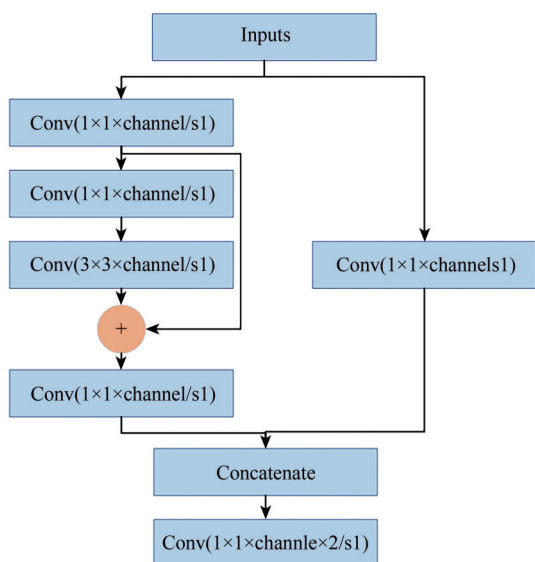


图2 ResBlock的module结构
Fig. 2 ResBlock module structure

3 轻量化在轨目标识别模型

3.1 模型框架改进

神经网络模型构建过程中需满足3个原则(Tan等, 2020): (1) 利用残差神经网络, 搭建较深的神经网络实现特征提取, 提升网络的深度; (2) 改变每一层提取的特征层数, 实现更多的特征提取, 提升网络的宽度; (3) 增大输入图片的分辨率, 可使网络学习与表达的特征更加丰富, 有利于提高精确度。

因此本次Yolov4的模型压缩中遵循上述特点, 基于深度可分离卷积, 借鉴反向残差结构的构建思路, 建立SwishBlcok bottleneck模块, 从3个方面同时扩充网络的特性, 替换ResBlock_body以作为Yolov4的主框架总体设计思路, 同时借鉴SENet的Channel Attention思想融入网络结构中, 对提取的特征图赋予不同的权重, 抽取更关键的特征信息, 同时不会替模型的计算与存储增加开销。

本文构建的SwishBlcok bottleneck模块实际架构为倒置的残差结构, 残差结构中间特征图通道数少而两边的特征图通道数多, 反向残差结构中间特征图通道数多而两边的特征图通道数少, 反向残差结构的架构设计被证实可以提高内存利用效率。

因此本文设计该模块在3×3可分离卷积结构前首先利用1×1卷积升维, 增加提取特征图的维度,

进一步提升图像的特征提取能力; 后面紧接着3×3的可分离卷积、BatchNorm与Swish激活函数, 高维度提取更多的图像纹理特征, 同时兼具可分离卷积降低模型参数的特性; 为了抽取更重要的特征信息, 在3×3网络结构后增加了一个通道注意力机制结构, 采用全局池化与Sigmoid激活函数训练每个通道的权重值, 将权重值非线性映射至0—1, 与图像特征图加权相乘; 再利用1×1卷积降维处理, 根据权重值保留更重要的特征向量; 最后增加一个大残差边, 避免网络发生梯度消失现象。具体的SwishBlcok bottleneck模块的架构设计如图3所示。

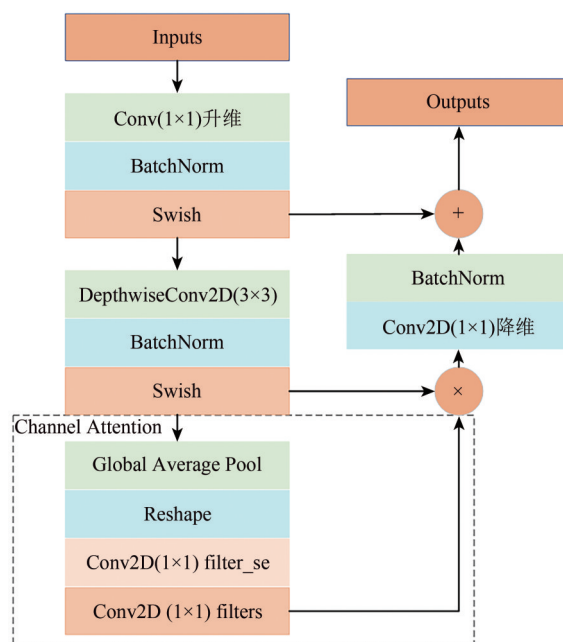


图3 SwishBlcok的module结构
Fig. 3 SwishBlock module structure

本小节遵循Yolov4模块中的SPP与PANet基本体系结构的优势, 利用SwishBlcok bottleneck替换Yolov4主框架中的ResBlock_body结构。利用SwishBlcok bottleneck结构构建神经网络模型, 同时为了减少整体网络的参数量, 调整原Yolov4的网络模型模块的数量, 将原Yolov4模型的特征图维度由608×608×32、304×304×64、152×152×128、76×76×256、38×38×512、19×19×1024降低到304×304×16、152×152×24、76×76×48、38×38×88、38×38×120、19×19×208、19×19×352, 大大减少参数量, 达到压缩Yolov4网络模型的目的, 改进后的网络结构如图4所示。

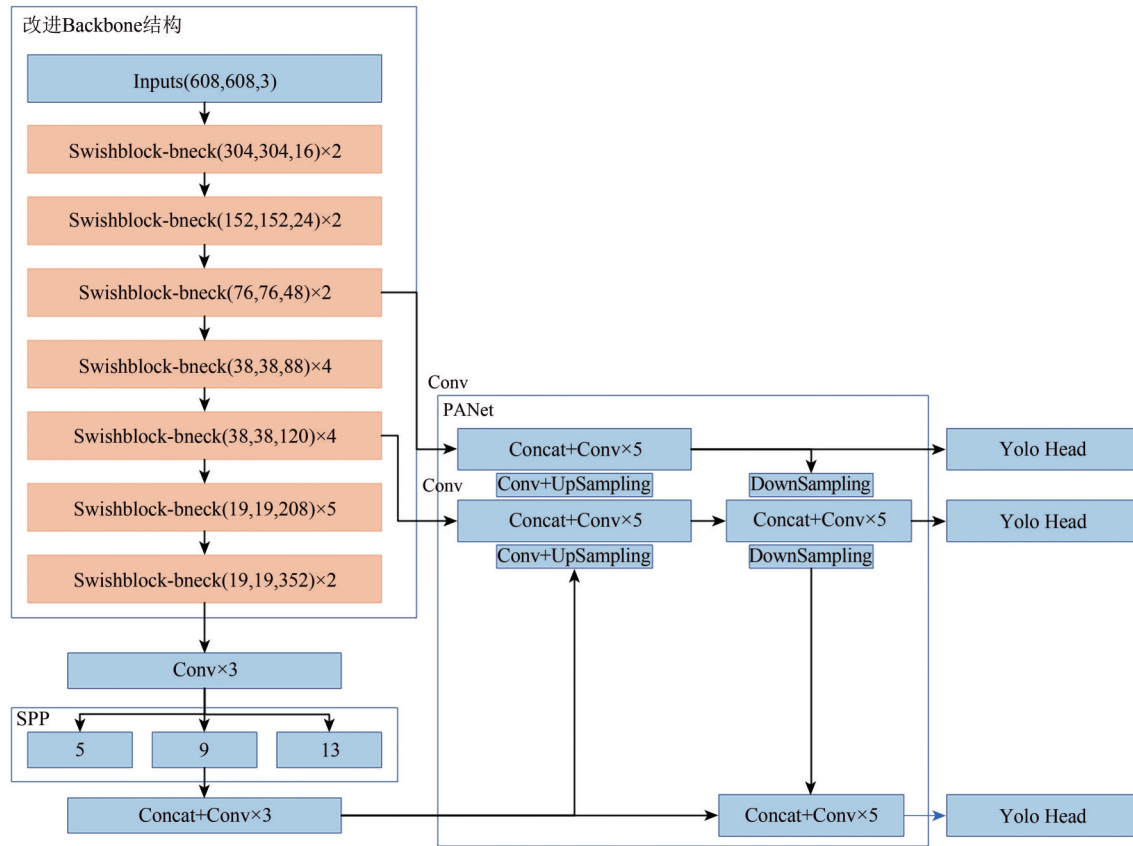


图4 YOLOv4网络模型压缩后网络架构

Fig. 4 YOLOv4 Network Model Compressed Network Architecture

3.2 SPP结构与PANet模块改进

YOLOv4网络模型中的PANet结构与SPP结构起到加强特征提取网络。分析SPP结构与PANet的结构，发现网络结构内含大量的 3×3 卷积层与连续5次卷积的结构，大大增加了模型的计算量。

深度可分离卷积由深度卷积与 1×1 的点卷积构成，如图5(b)(c)所示。图5(a)为常规卷积提取图像细节特征的过程，其中 M 为输入图像的维数， N 为卷积核滤波的通道数，卷积核大小为 3×3 ，则常规卷积提取的特征层维度数为 N ，则常规卷积的参数数量为 $M \times N \times 3 \times 3$ 。而深度可分离卷积采用逐通道卷积的策略，提取的特征图采用逐点卷积的形式同样得到维度数为 N 的特征图，如图5(b)(c)所示，输入图像的维数为 M ，可分离卷积核大小为 3×3 。因此可分离卷积操作的参数数量为 $M \times 3 \times 3 + 1 \times 1 \times M \times N$ 。

相较于常规卷积操作，采用可分离卷积提取图像的纹理特征，参数数量减少如式(1)：

$$\frac{M \times 3 \times 3 + 1 \times 1 \times M \times N}{M \times N \times 3 \times 3} = \frac{1}{N} + \frac{1}{9} \quad (1)$$

PANet结构与SPP结构含有连续五次卷积与连续三次卷积加强图像特征提取的过程，式(1)分析采用可分离卷积操作相较于常规卷积操作大大减少参数量。为了进一步压缩网络模型，本文在保持通道分离的前提下，基于可分离卷积结构实现空间卷积，替换SPP结构与PANet结构的普通卷积，实现模型参数量减少，内存依赖减少，具体替换如下图6所示。

3.3 模型推理速度提升

第3.2节中的SwishBlock bottleneck模块采用卷积层加Batch Normalization层的形式进行前向计算。Batch Normalization层能够加速网络的收敛，并且能够控制网络的过拟合，虽然训练时起到积极的作用，但在网络前向推理时增加了前向运算，占用更多的显存空间，影响处理速度。因此本节将SwishBlock bottleneck模块中的Batch Normalization层中的参数与卷积层进行合并，以提升模型前向推理速度。

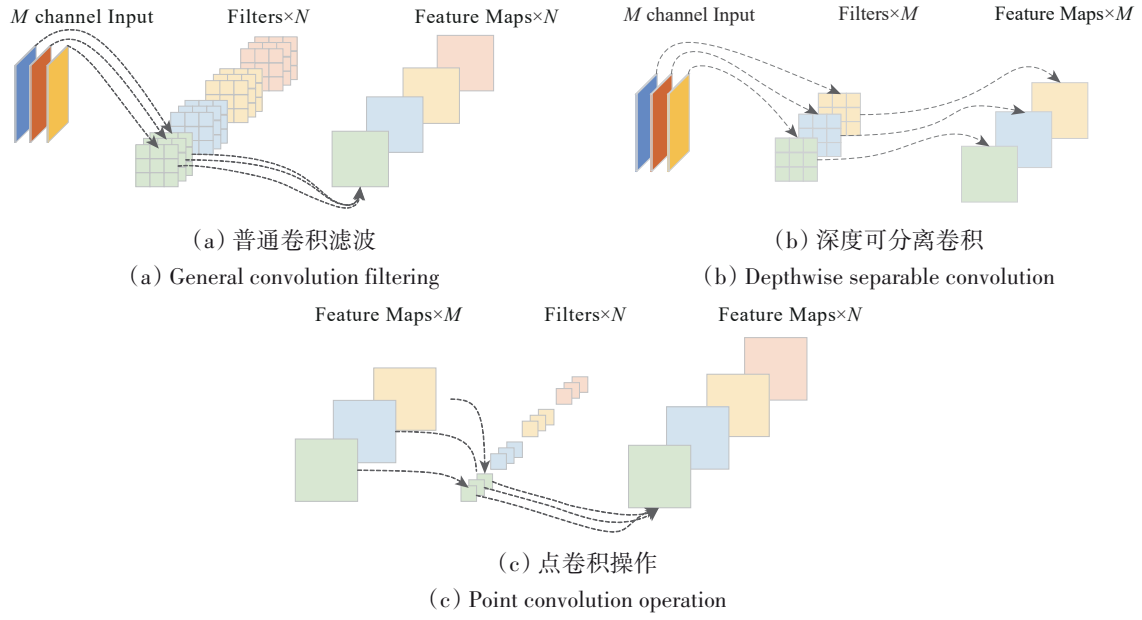


图5 深度可分离卷积与普通卷积

Fig. 5 Depthwise separable convolution versus ordinary convolution

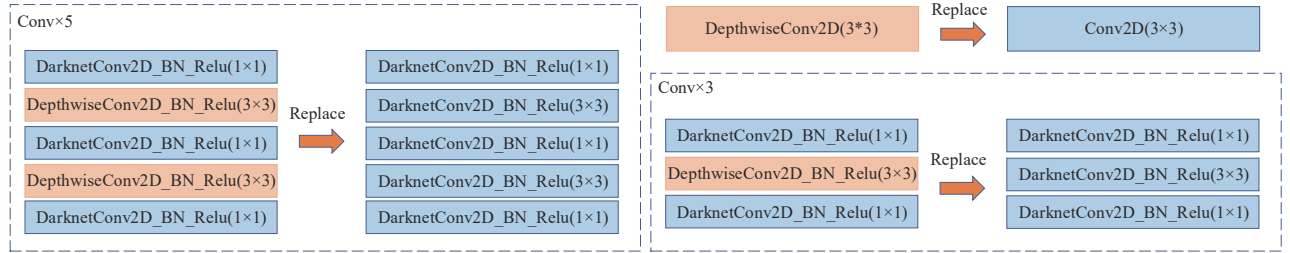


图6 深度可分离卷积替换普通卷积模块

Fig. 6 Depthwise separable convolution replaces ordinary convolution

原 Batch Normalization 层的表示方式为

$$\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i \quad (2)$$

$$\sigma_B^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (\mathbf{x}_i - \mu_B)^2 \quad (3)$$

$$\hat{\mathbf{x}}_i \leftarrow \frac{\mathbf{x}_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (4)$$

$$\mathbf{y}_i \leftarrow \gamma \hat{\mathbf{x}}_i + \beta \quad (5)$$

式中, μ_B 为数据集 batch 的均值, σ_B^2 为数据集 batch 的方差, $\hat{\mathbf{x}}_i$ 为数据集 batch 的归一化, $\mathbf{y}_i$ 为经过 Batch Normalization 层平移与缩放的输出结果。

卷积层的计算为

$$\text{out} = \sum_{i=1}^k w_i \mathbf{x}_i + b \quad (6)$$

BN 层经过卷积层的输出为

$$\text{BN} = \frac{\gamma^*(\text{out} - \mu_B)}{\sqrt{\sigma_B^2 + \varepsilon}} + \beta \quad (7)$$

将式 (6) 代入到式 (7) 中, 将式进行展开, 以

计算卷积层新的权重与偏置项。

$$\text{BN} = \frac{\gamma \times \left(\sum_{i=1}^k w_i \mathbf{x}_i + b - \mu_B \right)}{\sqrt{\sigma_B^2 + \varepsilon}} + \beta = \frac{\gamma \times \sum_{i=1}^k w_i \mathbf{x}_i + \gamma \times (b - \mu_B)}{\sqrt{\sigma_B^2 + \varepsilon}} + \beta \quad (8)$$

这样新的权重与偏置为

$$w_{\text{new}} = \frac{\gamma \times \sum_{i=1}^k w_i}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (9)$$

$$b_{\text{new}} = \frac{\gamma \times (b - \mu_B)}{\sqrt{\sigma_B^2 + \varepsilon}} + \beta \quad (10)$$

使用上述新的权重与偏置项来执行 SwishBleok bottleneck 模块的卷积层计算, 得到与原始卷积层加 Batch Normalization 层相同结果的同时减少模型前向推理速度。

3.4 损失函数的改进

Yolov4 网络训练时的损失函数由边界框回归损失 L_{ciou} 、置信度损失 L_{conf} 、分类损失 L_{class} 等3部分组成。

$$\begin{cases} L_{\text{conf}} = -\sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \left(\bar{C}_i^j \log(C_i^j) + (1 - \bar{C}_i^j) \log(1 - \bar{C}_i^j) \right) \\ \quad - \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} \left(\bar{C}_i^j \log(C_i^j) + (1 - \bar{C}_i^j) \log(1 - \bar{C}_i^j) \right) \\ L_{\text{class}} = -\sum_{i=0}^{S^2} I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} \left(\bar{P}_i^j \log(P_i^j) + (1 - \bar{P}_i^j) \log(1 - \bar{P}_i^j) \right) \\ \text{Loss} = L_{\text{ciou}} + L_{\text{conf}} + L_{\text{class}} \end{cases} \quad (11)$$

式中, S^2 、 B 为特征图尺度与先验框, λ_{noobj} 为权重系数, $I_{i,j}^{\text{obj}}$ 与 $I_{i,j}^{\text{noobj}}$ 为第 i 网格的第 j 先验框处有目标与没有目标, c 为预测框与实际框闭包区域的对角线距离, b 、 w 、 h 为预测框的中心坐标及宽高, b^{gt} 、 w^{gt} 、 h^{gt} 为实际框的中心坐标及宽高, C_i^j 、 $\bar{C}_i^j$ 为预测框与标注框的置信度、 P_i^j 、 $\bar{P}_i^j$ 为预测框与标注框的类别概率。

模型处理遥感影像在轨目标检测时, 为减缓前景与背景数据样本比例不均衡的问题, 本节借鉴 RetinaNet 网络 (Lin 等, 2020) 的 Focal Loss 损失函数, 给目标检测分类的交叉熵损失函数添加一个调制系数 α , 变成

$$L_{\text{class}} = -\sum_{i=0}^{S^2} I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} \left(\bar{P}_i^j \log(P_i^j) + (1 - \bar{P}_i^j)^\alpha \log(1 - \bar{P}_i^j) \right) \quad (12)$$

4 实验结果与分析

4.1 客观评价指标

基于深度卷积神经网络的在轨目标识别方法采用客观评价指标定量衡量, 可以多角度的获得衡量复原算法质量。目标检测性能公开的客观评价指标指标如 IOU、准确率 (Precision)、召回率 (Recall)、AP。交并比 (IOU) 是模型所预测的边界框和真实边界框的交集和并集之比, 同时被称作 Jaccard 指数。

平均精度 AP 是目标检测领域中用于评估算法精度的常用指标, 以 Recall 值为横轴, Precision 值

为纵轴, 这样便可得到 PR 曲线, 进而求取 AP 值, 简单来说就是对 PR 曲线上的 Precision 值求均值, AP 值的定义为

$$\text{AP} = \int_0^1 p(r) dr \quad (13)$$

式中, $p(r)$ 为 PR 曲线, 而曲线的构建是通过准确率与召回率绘制得出, 准确率 (Precision) 定义为

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (14)$$

式中, TP 为检测类别为正类预测成正类, FP 为检测类别为负类预测成负类, 预测正负类通过交并比 (IOU) 值超过阈值判断得到的。准确率即所有被预测为正类样本的识别正确个数。

召回率 (Recall) 定义为

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{P} \quad (15)$$

式中, FN 为检测类别为正类预测成负类, P 为测试数据集中正类样本个数, 召回率反应模型的漏检率。精确率和召回率是相互独立的, 若准确率很高意味误检率较低, 误检率低的情况下会导致漏检率偏高。

4.2 测试用例

本节采用的公开的遥感数据集 RSOD, RSOD 数据集在 2015 年公开, 它是一个用于遥感图像目标检测的开放数据集, 数据来源有 Tianditu 等平台, 采用 PASCAL VOC 的标注格式。数据集包括飞机、油箱、操场和立交桥 4 类对象, 共有 976 张图像, 每幅图像的大小约在 512×512 到 1083×923 像素之间, 实例对象具体分布如下表 1 所示。

表 1 RSOD 数据集实例对象分布

Table 1 Instance object distribution of RSOD dataset

参数	飞机	操场	立交桥	油箱
图像数量	446	189	176	165
实例数量	4993	191	180	1586

RSOD 数据集的飞机类别图像数量为 446, 数量较少, 不能全面体现模型的检测性能与泛化能力, 因此本节额外自制数据集进行神经网络模型的混合训练。自制数据集的图像来源于 Google Earth, 来自于多个城市, 分辨率为 2400×2400 左右, 飞机尺寸为 30×30 像素大小, 目标所占背景比例大概为 $\frac{1}{6400}$ 左右, 采用 PASCAL VOC 的标注格式, 因本节实际解决在轨检测飞机的需求, 数据集只标注

飞机的类别，一共有1863张，加上ROSD数据的飞机数量，一共2309张，飞机的实例数量大概为25623，采用中小目标为主，样本分布密集的构建原则。实际数据集的训练集、验证集与测试集划分数量如表2所示。

表 2 训练集、验证集与测试集数量

Table 2 Training set, verification set, and test set

数据集		分类			
		飞机	训练集	验证集	测试集
RSOD数据集	图像数量	446	268	60	118
	实例数量	4993	2994	672	1327
自制数据集	图像数量	1863	1118	325	420
	实例数量	20630	12522	3413	4695

图7为RSOD公开数据集的样本示例，看到图像分辨率尺度不一，数据集中民用客机也存在不同空间分辨率，分别为1—2 m不等，由图7仍可判断出飞机的轮廓与飞机的位置。图8为自制数据集的样本示例，同样民用客机的空间分辨率也互有差异，分别为2—4 m不等组成，从图像中隐约可判断飞机轮廓，目标极小且样本分布密集。



图7 RSOD公开数据集样本示例
Fig. 7 RSOD dataset samples

4.3 改进算法分模块的测试对比结果

为充分验证本文对Yolov4网络模型多方面改进的合理性，采用分步骤的形式开展验证实验，采用客观评价指标定性衡量复原算法质量与速度。目标检测性能公开的客观评价指标为AP、参数量、计算量FLOPs。测试数据集采用公开的RSOD数据集，测试平台为高性能GPU服务器。硬件条件为Intel I9-10900K、64 G DDR4内存、NVIDIA GeForce

RTX 2080 Super GPU，基于深度学习框架Pytorch 1.14。首先对模型分阶段的改进，具体的指标值如表3所示。



图8 自制数据集样本示例
Fig. 8 Samples of self-made data sets

表 3 不同算法采用RSOD数据集在轨检测目标的性能对比
Table 3 Comparison of performance of different algorithms using RSOD dataset for object detection

方法	AP/%	参数量/M	FLOPs/GMacs
Yolov4	95.5	64.00	63.92
3.1节改进	92.09	10.31	1.84
3.2节改进	91.89	8.22	0.69
3.3节改进	91.89	8.04	0.64
3.4节改进	94.09	8.04	0.64

可看到3.1节仅改进模型的主框架，相较于Yolov4网络模型，AP指标为Yolov4模型的96.43%，模型参数量减少83.89%，算法的计算量减少97.12%。3.2节基于深度可分离卷积结构实现空间卷积替换SPP结构与PANet结构的普通卷积的操作后，模型的参数由10.31减少到8.22，计算量从1.84减少到0.69。为提升模型前向推理速度，3.3节执行合并卷积层与Batch Normalization层中的参数操作，采用每秒处理的帧数来考量前向推理速度的有效性，并对比原Yolov4网络模型的处理速度，具体的帧/s的指标值如表4所示。

表 4 合并卷积层与BN层的性能对比
Table 4 Performance comparison of merge convolution layer and BN layer

方法	速度/(帧/s)	
	416×416	608×608
Yolov4	32	22
3.2节改进	59	51
3.3节改进	62	54

表4可看出,相较于原Yolov4网络模型,在不同尺寸下,改进的网络模型每秒处理的帧数均有一定的提高。而执行合并卷积层与Batch Normalization层中的参数操作后,在416×416的图像尺寸下处理速度由59帧/s提升到62帧/s,608×608图像尺寸下处理速度由51帧/s提升到54帧/s。

3.4节为减缓前景与背景数据样本比例不均衡的问题,改进损失函数的具体AP、模型参数、计算量由表3可看出,通过为模型的损失函数添加调制因子的操作,算法在测试数据集的AP指标由91.89%提升到94.09%。

同时由表3可看出Yolov4网络模型采用公开数据集RSOD压缩前的AP为95.5%,模型压缩后的AP值为94.09%,检测性能稍有下降。利用模型大小、计算量等评价指标对Yolov4模型压缩前后进行评测,Yolov4网络模型压缩后的图像检测性能AP为压缩前的98.52%,模型大小减少87.43%,算法的计算量减少98.99%。

4.4 前沿算法的测试对比结果

本文利用ROSD加自制数据集的形式进行训练测试,并与目标检测领域处于前沿水平的网络模型进行对比,并与前沿的轻量级神经网络进行对比,例如Yolov5、MobileNet-SSD、MobileNetv2-Yolov4、GhostNet-Yolov4对比的结果如表5所示。

表 5 不同算法在轨检测目标的性能对比
Table 5 Comparison of object detection performance of different algorithms

方法	Backbone	AP/%	参数量/M	FLOPs/GMacs
Yolov4	CSPDarknet	89.89	64.00	63.92
SSD	VGG-16	79.64	26.29	127.50
Yolov4-tiny	CSPDarknet	54.22	5.90	4.31
Yolov5s	CSPDarknet	67.32	7.30	17.00
Yolov5m	CSPDarknet	76.58	21.40	51.30
MobileNet-SSD	MobileNet-v1	69.28	8.85	12.40
Mobilev2-Yolov4	MobileNet-v2	71.82	12.13	8.25
GhostNet-Yolov4	GhostNet	78.98	11.07	7.07
CenterNet	ResNet-50	82.35	32.67	35.79
本文方法	CSPDarknet	88.87	8.04	0.64

由表5知,本文所提模型压缩算法在检测性能AP相较于高性能算法几乎没有损失,但模型大小与计算量大大减少。相较于速率较快的轻量级网络模型,模型大小指标几近相同,计算量FLOPs略有领先,但检测性能AP指标更优于经典轻量级网络模型。为了更直观对深度卷积神经网络的在轨目标识别方法的检测性能进行说明,本节从测试集挑选出一些飞机目标极小且样本分布密集的图片,取具有代表性的几个神经网络算法进行对比测试,检测结果如图9所示。

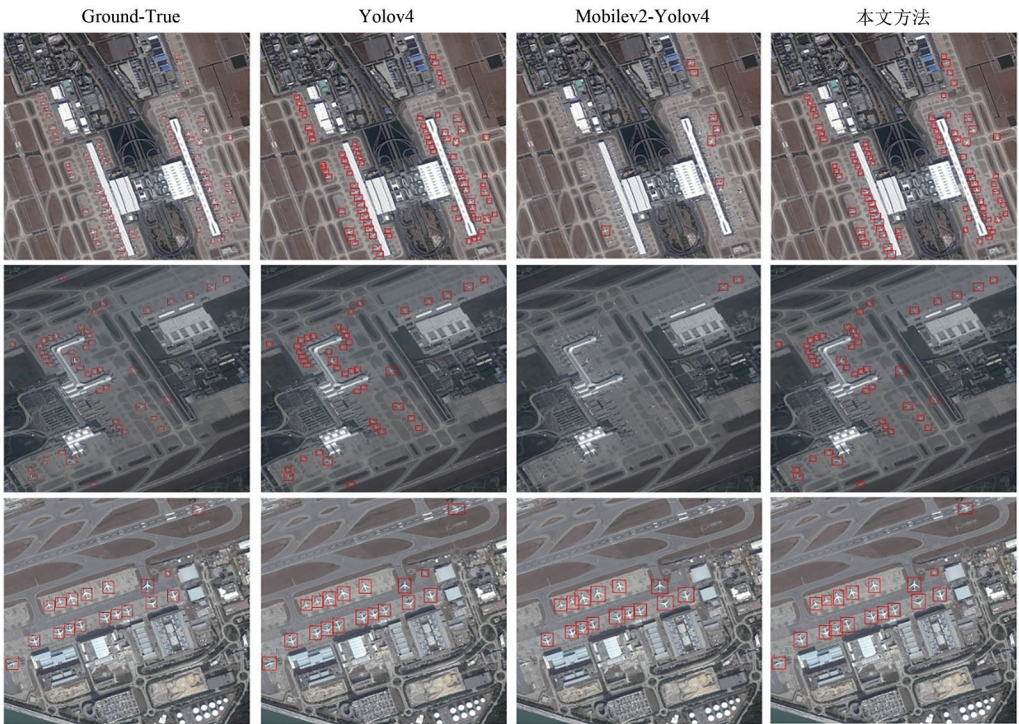




图9 不同网络模型的检测性能
Fig. 9 Detection performance of different network models

由图9可以看出，测试图像的飞机目标极小且分布密集，4张图像中飞机的分辨率由小到大递增。由4张图像可以看出，飞机像元尺寸较小（像元比例 $\frac{1}{6400}$ ）的测试图像（如第1张和第2张图像），Yolov4网络模型与本文压缩的网络模型检测效果几近相同，检测效果较好，而基于Mobilev2-Yolov4网络模型检测小目标效果较差；飞机像元尺寸较大（像元比例 $\frac{1}{144}$ ）的测试图像（如第3张和第4张图像），3种神经网络模型的准确率较高，但Mobilev2-Yolov4网络模型对图像内像元较小的飞机目标仍未识别。由此证明该网络同时兼顾检测速度与检测精度，有效的检测出遥感小目标。

4.5 嵌入式平台实验结果

为验证本文改进Yolov4算法模型在轨目标识别的检测速度，将训练好的模型部署到Jetson Xavier NX嵌入式平台上进行验证处理，该处理器外形小巧、功耗较低、计算性能强，可作为卫星超算平台的计算节点之一。分别从AP与FPS客观评价指标对比了Yolov4网络模型、Yolov4-tiny网络模型与本文改进模型的性能，如表8所示。

表8 不同算法在轨检测目标的性能对比

Table 8 Comparison of object detection performance of different algorithms				
方法	AP/%		速度/(帧/s)	
	416×416	608×608	416×416	608×608
Yolov4	86.86	89.89	8	5
Yolov4-tiny	46.16	54.22	41	31
Our	85.03	88.87	32	26

由表8可知，面向Jetson Xavier NX嵌入式平台，输入图像尺寸为416×416与608×608时，Yolov4网络模型因模型结构复杂每秒处理的帧数为8帧与5帧，未能满足星上在轨实时检测的需求，而Yolov4-tiny网络模型与本文改进的模型每

秒处理的帧数可达41、32与31、26帧，满足在轨实时检测，但Yolov4-tiny网络模型的AP指标为46.16，识别率较低，而本文改进的Yolov4网络模型较Yolov4-tiny网络模型精度高且处理速度快，满足在轨实时准确检测目标的需求。

4 结 论

本文基于深度卷积神经网络提出一种在轨目标检测模型，旨在解决星上计算资源和存储资源有限，无法支持高能耗的网络模型在轨检测目标的不足，结合反向残差结构与通道注意力思想改进在轨识别模型Yolov4网络模型的模块结构，加深加宽网络模型的模块结构；利用可分离卷积结构，改进SPP结构与PANet结构，实现空间卷积，减少模型参数量；为加快前向推理速度，同时合并卷积层与Batch Normalization层，重新计算卷积层参量；借鉴Focal损失函数改进目标检测分类的损失函数，解决前景与背景数据样本比例不均衡。实验从地面平台与嵌入式平台的结果证明：本文提出的在轨目标检测算法同时兼顾检测性能，能够有效检测遥感小目标。在未来研究中，仍需要扩展遥感数据集的规模，更全面提高模型应用场景的普适性。

参考文献(References)

Biswas D, Su H B, Wang C Y, Stevanovic A and Wang W M. 2019. An automatic traffic density estimation using single shot detection (SSD) and MobileNet-SSD. Physics and Chemistry of the Earth, Parts A/B/C, 110: 176-184 [DOI: 10.1016/j.pce.2018.12.001]
Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: optimal speed and accuracy of object detection. arXiv: 2004. (10934 [DOI: 10.48550/arXiv.2004.10934]
Chen Y H, Krishna T, Emer J S and Sze V. 2017. Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks. IEEE Journal of Solid-State Circuits, 52(1): 127-138

- [DOI: 10.1109/JSSC.2016.2616357]
- Demirel H and Anbarjafari G. 2010. Satellite image resolution enhancement using complex wavelet transform. *IEEE Geoscience and Remote Sensing Letters*, 7(1): 123-126 [DOI: 10.1109/LGRS.2009.2028440]
- Dong Y F, Zhang C T, Wang P and Feng Z. 2020. Airplane detection of optical remote sensing images based on deep learning. *Laser and Optoelectronics Progress*, 57(4): 041007 (董永峰, 仇长涛, 汪鹏, 冯哲. 2020. 基于深度学习的光学遥感图像飞机检测算法. *激光与光电子学进展*, 57(4): 041007) [DOI: 10.3788/LOP57.041007]
- Girshick R. 2015. Fast R-CNN//2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]
- Han X F, Jiang T, Zhao Z F and Lei Z T. 2020. Research on remote sensing image target recognition based on deep convolution neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 34(5): 2054015 [DOI: 10.1142/S0218001420540154]
- Law H and Deng J. 2020. CornerNet: detecting objects as paired keypoints. *International Journal of Computer Vision*, 128(3): 642-656 [DOI: 10.1007/s11263-019-01204-1]
- Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2020. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2): 318-327 [DOI: 10.1109/TPAMI.2018.2858826]
- Liu G, Sun X, Fu K and Wang H Q. 2013. Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior. *IEEE Geoscience and Remote Sensing Letters*, 10(3): 573-577 [DOI: 10.1109/LGRS.2012.2214022]
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY and Berg AC. 2016. SSD: single shot MultiBox detector//14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]
- Redmon J and Farhadi A. 2018. YOLOv3: an Incremental Improvement. arXiv: 1804. (02767 [DOI: 10.48550/arXiv.1804.02767])
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 137-149 [DOI: 10.1109/TPAMI.2016.2577031]
- Roy A G, Navab N and Wachinger C. 2018. Concurrent spatial and channel ‘squeeze and excitation’ in fully convolutional networks//21st International Conference on Medical Image Computing and Computer-Assisted Intervention. Granada, Spain: Springer: 421-429 [DOI: 10.1007/978-3-030-00928-1_48]
- Tan M X, Pang R M and Le Q V. 2020. EfficientDet: scalable and efficient object detection//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 10778-10787 [DOI: 10.1109/CVPR42600.2020.01079]
- Tang J X, Deng C W, Huang G B and Zhao B J. 2015. Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine. *IEEE Transactions on Geoscience and Remote Sensing*, 53(3): 1174-1185 [DOI: 10.1109/TGRS.2014.2335751]
- Wang C, Zhang L Q, Yan T P, Jiang T, Jia Y J and Lai Z H. 2021. Research on the state detection of the secondary panel of the switchgear based on the YOLOv5 network model. *Journal of Physics: Conference Series*, 1994(1): 012030 [DOI: 10.1088/1742-6596/1994/1/012030]
- Yu J Y, Huang D, Wang L Y, Guo J and Wang Y H. 2016. A real-time on-board ship targets detection method for optical remote sensing satellite//2016 IEEE 13th International Conference on Signal Processing (ICSP). Chengdu, China: IEEE: 204-208 [DOI: 10.1109/ICSP.2016.7877824]
- Zhou X Y, Zhuo J C and Krähenbühl P. 2019. Bottom-up object detection by grouping extreme and center points//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 850-859 [DOI: 10.1109/CVPR.2019.00094]

Lightweight model for On-Orbit optical object detection

LYU Xiaoning¹, XIA Yuli¹, ZHAO Junsuo^{1,2}, QIAO Peng¹

1. *Science & Technology on Integrated Information System Laboratory, Institute of Software Chinese Academy of Sciences, Beijing 100190, China;*

2. *University of Chinese Academy of Sciences, Beijing 100049, China*

Abstract: As an important transport carrier and military target, aircraft detection in remote sensing images is important for aircraft rescue, early warning, and other fields. At present, the widely used neural network model has a complex structure and requires a large number of parameters, which limits the computing and storage resources of aircraft detection satellites. The efficiency and accuracy of satellite in-orbit detection need to be studied, and the computational structure must be optimized. Using neural networks in lightweight operation can reduce the computational costs and compress the overall framework.

In this study, on the basis of a deep separable convolution neural network combined with deep separable convolution, the SwishBlok

bottleneck module was established by referring to the construction idea of a reverse residual structure. The characteristics of the network were simultaneously expanded in three aspects as follows: ResBlock_body was replaced with the overall design idea of the main framework of YOLO v4. Simultaneously, the channel attention mechanism of SENet was used for reference and integrated into the network structure. Different weights were given to the extracted feature maps and information. On the premise of maintaining channel separation, a separable convolution structure was used to improve the SPP structure and PANet structure; in this manner, both the number of model parameters and the memory dependence could be reduced. Moreover, the convolution layer and the batch normalization layer were merged to further accelerate forward reasoning. Drawing on the focal loss function, the loss function of object detection was improved to solve the imbalance between foreground and background data samples.

The quality of algorithm restoration necessitates verification. In this study, objective evaluation indices were used to measure the algorithm from multiple angles. The public RSOD dataset and an internally produced dataset were used to compare the high-performance network models for algorithm verification. In terms of the rationality of the various improvements in the network model, verification experiments were conducted to measure the quality and processing speed of the algorithm. Then, the trained model was deployed on an embedded platform to verify the detection speed of the improved YOLO v4 algorithm model for on-orbit object recognition. The number of parameters of the proposed scheme was reduced by sevenfold compared with that of the original method, and the number of FLOPs was reduced by approximately 30 at a recognition accuracy of 94.09%. Subsequently, the experimental results were compared with the findings for the YOLO series, SSD, MobileNet, CenterNet, and other cutting-edge network models. The proposed algorithm outperformed the other methods.

The proposed on-orbit object detection model can overcome the limitations of computing and storage resources, which traditionally cannot support high-precision complex models. The experimental results from ground and embedded platforms also prove that the proposed on-orbit object detection algorithm can effectively detect remote sensing targets based on detection performance. Future research may expand the scale of remote sensing datasets and improve the universality of model application scenarios

Key words: remote sensing, convolution neural network, Yolov4, optical remote object detection, reverse residual structure

Supported by National Natural Science Foundation of China (No. 62027801)