

基于 Faster R-CNN 改进的光学遥感图像飞机检测

杨鑫, 王琼, 姚亚洲, 唐振民*

南京理工大学计算机科学与工程学院, 江苏 南京 210094

摘要 针对光学遥感图像飞机检测任务中超大图像尺寸、小目标检测、复杂背景干扰等问题, 提出一种轻量化特征提取网络与注意力机制融合的改进型 Faster R-CNN 飞机检测算法。通过删去对检测小目标冗余的深层特征层, 所提算法对小目标的检测能力得到有效提升, 且网络参数量减少 38.4%, 实现了轻量化处理, 推理速度也有显著提升; 为强化特征提取能力、弱化背景干扰, 创造性地仅在特征提取网络的主干部分引入卷积块注意力模块, 有效增加模型对飞机目标的检测能力; 在测试推理阶段, 采用中线单帧预测后处理方式, 对重叠区域内的飞机目标进行单帧预测, 避免重复推理、预测结果不一致现象。实验证明, 改进后的算法在光学遥感数据集上的 mF1 分数比改进前算法提升 3.5%, 最终达到 88.97。

关键词 遥感; 飞机检测; 注意力机制; 轻量化; 遥感图像

中图分类号 TP753 文献标志码 A

DOI: 10.3788/LOP221679

Improved Aircraft Detection of Optical Remote Sensing Image Based on Faster R-CNN

Yang Xin, Wang Qiong, Yao Yazhou, Tang Zhenmin*

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China

Abstract To address the issues of extremely large image size, small target detection, and complex background interference in aircraft detection task of optical remote sensing image, an improved Faster R-CNN aircraft detection algorithm based on the fusion of lightweight feature extraction network and attention mechanism is proposed. The proposed algorithm's ability to detect small targets is significantly enhanced by removing the deep feature layer of small target detection redundancy, which also results in a 38.4% reduction in the number of network parameters, enabling lightweight processing and significantly improving the reasoning speed. To strengthen the feature extraction ability and weaken the background interference, convolutional block attention module is creatively presented only in the backbone of the feature extraction network, which successfully increases the detection ability of the model to aircraft targets. To avoid repeated reasoning and inconsistent prediction results, the midline single frame prediction post-processing mode is used in the test reasoning stage to predict the aircraft target in the overlapping area in a single frame. The experiment demonstrates that the improved algorithm achieves a final mF1 score of 88.97, which is 3.5% higher than the original algorithm on the optical remote sensing dataset.

Key words remote sensing; aircraft detection; attentional mechanism; lightweight; remote sensing image

1 引言

近年来,我国卫星发射数量稳步增加,各种波段的卫星传感器性能显著提高,为科学研究与应用提供了大量高成像质量的遥感图像。高质量的光学遥感图像能够直接反映地物目标细节信息,便于人眼观察,极大地推动了遥感图像解译任务的发展。

飞机既是重要的交通工具,也是重要的军事战略目标。因此,无论是民用领域还是军事领域,在遥感图像中解译出飞机的型号与位置,都有着重大的现实需

求^[1]。随着深度学习的快速发展,常规图像目标检测的研究和应用取得了快速发展。然而,遥感图像与常规图像相比有着其自身的特性,若直接将基于常规图像的目标检测算法应用于遥感图像目标检测,并不能取得理想的效果^[2-3]。

遥感图像目标检测面临诸多挑战:1)遥感图像具有超大图像尺寸,对于现阶段的计算硬件资源,针对常规图像的深度学习目标检测算法无法直接处理大尺寸的遥感图像;2)光学传感器成像质量易受天气影响,如云、雨、雾遮挡等,使得部分遥感目标外观模糊、边界不

收稿日期: 2022-05-24; 修回日期: 2022-06-22; 录用日期: 2022-07-04; 网络首发日期: 2022-07-14

通信作者: *tzm.cs@njjust.edu.cn

清晰,难以提取到细节特征,加大了检测识别难度;3)不同类型飞机的尺寸差异大,对检测网络的多尺度检测能力^[4]要求高;4)由于超高成像视角,遥感图像中大多数飞机以小目标形式呈现,飞机细节信息丢失严重,现有深度特征提取网络需要对输入图像进行多级下采样,深层网络很难提取到小目标的深层语义特征^[3-5],从而带来检测困难。

针对于光学遥感图像的特点与检测挑战,本文设计了一种基于Faster R-CNN^[6]的改进目标检测算法,对超大尺寸的光学遥感图像使用带重叠区域的滑动窗口切割法,将图片切割成尺寸合适的子图,使用轻量化的ResNet^[7]与特征金字塔网络(FPN)^[8]提取图像多尺度特征,并在ResNet的主干网络中引入卷积块注意力模块(CBAM)注意力机制^[9]提高网络对重要区域特征信息的关注,在推理阶段采用中线单帧预测的后处理方式解决重叠区域重复预测的问题。改进后的算法在光学遥感数据集上取得较大提升,展现出较大的优越性。

2 相关工作

目标检测包含2个子任务:1)对图像中检测的目标进行分类;2)以水平矩形包络框的形式对目标进行定位^[10]。早期的遥感图像飞机检测算法通常采用人工设计特征提取方式提取图像特征,Lin等^[11]根据模板匹配的思想,提出一种径向梯度角的旋转不变特征用于模板匹配,该算法在复杂背景下展现出一定的抗噪性。由于不同类型的飞机外观并不一样,单一特征难以表达所有类型飞机的特征,Grabner等^[12]将遥感图像的Hear特征、方向直方图和局部二值模式(LBP)特征进行线性级联检测飞机目标,在人工设计特征时取得了较好的效果。人工设计特征提取方式能够高效地实现特定任务的检测,但需要大量的专家知识,且提取的特征鲁棒性差、数据泛化能力较弱,不能适应现代多样化任务检测的需求。随着深度神经网络的快速发展,基于深度学习的遥感图像飞机检测算法逐渐成为主流。董永峰等^[13]提出一种以Mask R-CNN为基础框架的算法,利用特征重用技术,提升遥感图像中小目标检测精度。肖振久等^[14]针对遥感图像特点,在YOLOv4算法中引入轻量化网络,并在预测层串联自注意力机制,有效提高了算法检测速度与检测精度。基于深度学习的遥感图像飞机检测算法能够自动进行图像特征提取,避免人工设计特征,具有较好的鲁棒性和强大的检测性能。

目前,基于深度学习的目标检测方法有2种分类标准:1)依据是否需要感兴趣区域提取,将目标检测方法划分为两阶段检测算法和单阶段检测算法^[3,5]。两阶段目标检测算法需要进行感兴趣区域提取,相较于同时期的单阶段检测算法,有着较高的精度性能,但感兴趣区域的提取带来了计算量提升,限制了算法的检测推理速度,实时性能不如单阶段目标检测算法。两阶段目标检测算法的代表主要有R-CNN算法^[15],

Fast R-CNN算法^[16]、Faster R-CNN算法等。相较于两阶段目标检测算法,单阶段目标检测算法舍去感兴趣区域的提取过程,直接在图像特征图上进行目标分类和定位框的回归,有着较快的检测速度,但由于缺少感兴趣区域的提取,不能获得高质量的正样本,正负样本比例失衡现象严重,因此算法精度性能不如两阶段目标检测算法。单阶段的目标检测代表算法主要有YOLO算法^[17]、YOLO9000算法^[18]、YOLOv3算法^[17]、SSD算法^[19]等。2)依据是否需要预先设定锚点框(anchor),将目标检测方法划分为基于锚点框的检测算法和基于关键点的检测算法^[3]。基于锚点框检测会在预设的锚点框基础上进行目标边框回归,使得网络输出值是在锚点框基础上的相对值,值域偏移范围较小、网络收敛速度较快、检测精度性能较高,缺点是手动设计高质量的锚点框。基于锚点框检测的代表算法有Faster R-CNN算法、SSD算法、RetinaNet算法^[20]等。基于关键点检测的算法避免了超参数繁多的锚点框设计过程,同时也避免了检测过程中目标重合度计算过程,减少了运算量,因此检测速度快于基于锚点框检测算法。基于关键点检测的代表算法有CenterNet算法^[21]、CornerNet算法^[22]、FCOS算法^[23]等。

飞机作为重要的民用交通工具和军事战略目标,在光学遥感图像中准确地识别飞机目标具有十分重大的意义。综合考虑检测速度与检测性能,采用综合性能优越的Faster R-CNN算法作为光学遥感图像飞机检测的基本框架。

3 所提算法

3.1 Faster R-CNN算法基本原理

Faster R-CNN算法主要由特征提取子网络、区域建议网络(RPN)、感兴趣区域池化(ROI pooling)网络和检测子网络组成。近些年,随着基于深度学习的目标检测算法的深入研究,涌现出许多性能优异的插件模块,如FPN、ROI align^[23]、NAS-FPN^[24]等。结合光学遥感图像飞机检测的特点,引入FPN模块,并将ROI pooling网络替换为效果更好的ROI align网络^[24]作为基础的Faster R-CNN算法框架,引入基础模块后的算法框架如图1所示。

引入基础模块后的Faster R-CNN算法,首先将图片输入特征提取网络进行图像特征提取,生成多级不同尺寸的特征图。接着,将特征提取网络生成的多级不同尺寸的特征图送入FPN,实现不同尺度特征图之间的融合,使浅层特征图也能具有深层特征图的语义信息,从而提升检测性能。然后,将融合后的特征图送入RPN,RPN会在特征图上预设一系列不同长宽比和不同大小的锚点框。在训练阶段,根据锚点框与真值边界框的交并比大小区分锚点框包含的区域是背景还是目标前景,从而生成候选区域,并送入ROI align网络。在推理阶段,则根据网络权重进行前向传播计算,

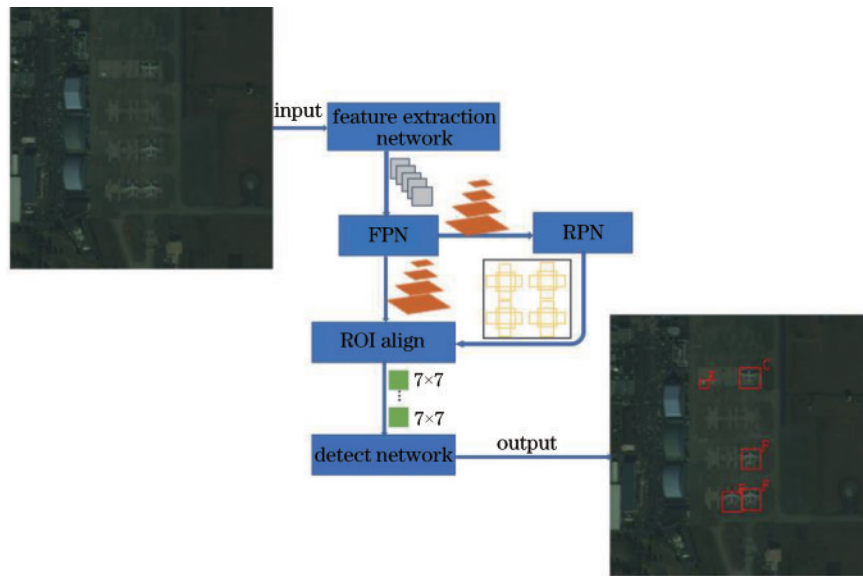


图1 引入基础模块后的 Faster R-CNN 算法框架

Fig. 1 Faster R-CNN algorithm framework after basic module introduced

选择得分最高的 2000 个锚点框作为候选区域,送入 ROI align 网络;ROI align 网络利用双线性插值算法,实现图像上候选区域到特征图上的映射,得到固定尺寸(7×7)的特征图。最后,将 ROI align 网络生成固定尺寸的特征图送入全连接检测网络,进行目标分类与边界框位置回归。

3.2 轻量化特征提取网络

轻量化是在保证原有模型性能不显著下降的情况

下,对原有模型进行修改删减的过程,轻量化后的模型具有结构简单、参数量少、计算效率高等特点。特征提取网络是 Faster R-CNN 算法的核心部分,这部分网络的参数量占据了模型参数量的大部分,因此对特征提取网络进行轻量化具有十分重大的意义。采用 ResNet50 作为特征提取网络提取图像特征,并针对遥感图像中飞机目标的特点,设计了一种高效的轻量化改动方法,实现模型的轻量化。ResNet50 基本结构如图 2 所示。

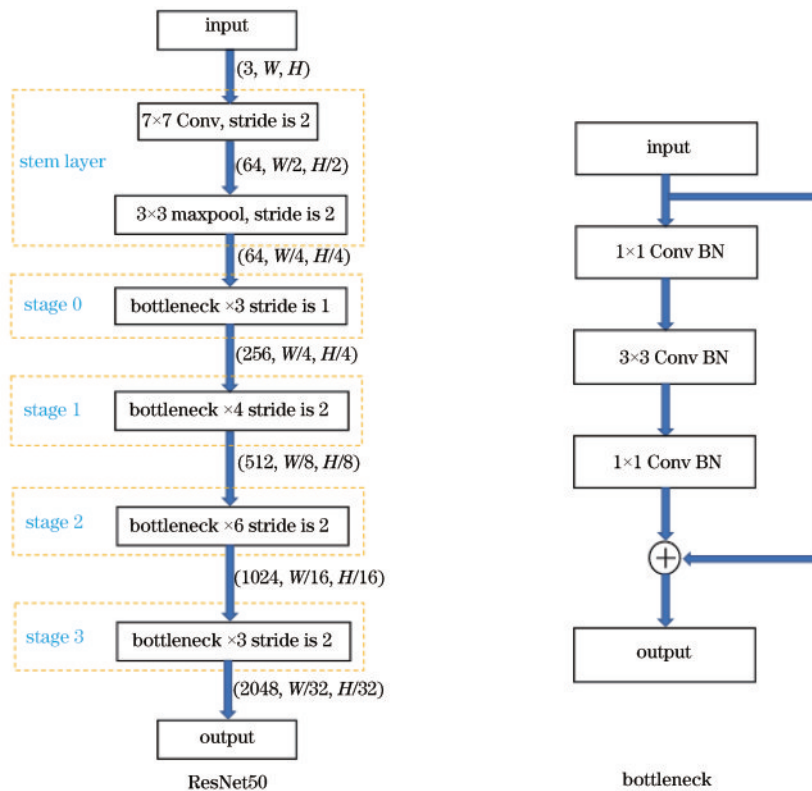


图2 ResNet50 结构

Fig. 2 Structure of ResNet50

整个网络由主干网络和 4 个层级 (stage) 组成, 每个层级是由不同数量的 bottleneck 模块堆叠起来的。最后 3 个层级中, 每个层级的第 1 个 bottleneck 使用步长为 2 的 3×3 卷积实现下采样, 降低特征图的分辨率。由此可计算出, ResNet50 最后 1 个层级输出特征图上的一个点对应于输入图像上 $32 \text{ pixel} \times 32 \text{ pixel}$ 的像素区域。

对于小于 $32 \text{ pixel} \times 32 \text{ pixel}$ 的小目标来说, 最后 1 个层级上的特征图会由于缺少小目标的细节信息, 难以获得小目标完整的语义信息, 从而降低网络对小目标的特征提取能力, 影响下游检测任务的性能。此外, 在 ROI align 网络中, 需要根据候选区域的大小, 将候选区域映射回相应分辨率的特征图上, 通常候选区域大小与特征图层级的映射关系如下:

$$k = k_0 + \left\lceil \log_2 \left(\frac{\sqrt{w \times h}}{224} \right) \right\rceil, \quad (1)$$

式中: k_0 是 FPN 输出的特征图总层级数; 224 指的是在 ImageNet 数据集上预训练的图像尺寸; w, h 分别是指 ROI 区域的宽与高。因此, 对于 ROI 区域面积大于 $448 \text{ pixel} \times 448 \text{ pixel}$ 的大目标, 使用最深层的特征图, 即在特征图 (256, $W/64, H/64$) 上进行训练与推理; 对于 ROI 区域面积小于 $56 \text{ pixel} \times 56 \text{ pixel}$ 的小目标, 使用特征图 (256, $W/4, H/4$) 进行训练与推理。

光学遥感图像以天基传感器对地面上的飞机进行拍摄成像, 超高的成像视角以及超广的成像覆盖地域

使得飞机在遥感图像上占比极小。以波音 737 客机为例, 其长度为 37.81 m、翼展为 28.45 m, 在分辨率为 0.5~1 m 的可见光光学卫星的拍摄成像中, 最大仅为 $76 \text{ pixel} \times 59 \text{ pixel}$ 。对于一些超大型运输机, 成像大小一般也是小于 $224 \text{ pixel} \times 224 \text{ pixel}$ 。而 ROI align 网络只有在输入的 ROI 区域大于 $448 \text{ pixel} \times 448 \text{ pixel}$ 时才会使用最深层的特征图。因此, 对于光学遥感图像飞机检测任务来说, 最后一层级的特征图是冗余的。此外, 根据感受野理论, ResNet50 第 41 层和第 50 层对应的感受野大小分别为 267×267 和 427×427 , 可见仅使用 ResNet50 的前 41, 感受野大小就足以覆盖遥感图像中飞机的大小。因此, 删除 ResNet50 的最后 1 个层级, 使网络总层数降为 41, 输出 3 种不同大小的分辨率特征图到 FPN 中, 命名为 ResNet41-3, 如图 3 所示。这是一个极其轻便和高效的轻量化方法, 此方法针对遥感图像飞机检测任务的特性, 直接删去 ResNet50 的最后 1 个层级, 可以直接加载预训练模型参数一步训练到位, 避免传统模型压缩算法中 finetune 的步骤, 提高训练效率。ResNet41-3 相比于 ResNet50, 模型参数量减少了 38.4%, 提升了检测速度。在光学遥感飞机图像数据集上进行的相应实验表明, 去除深层冗余特征图后, 网络的 mF1 评价指标分数由 85.98 分提升到 87.48 分, 同时网络结构得到轻量化, 参数量从 4.19×10^7 减少到 2.58×10^7 , 推理速度由 16.3 frame/s 提高到 18.2 frame/s。

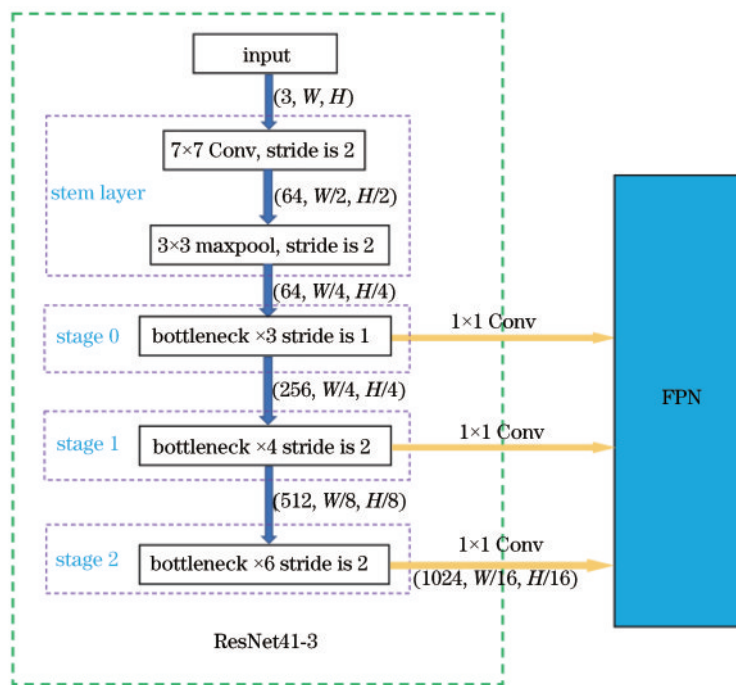


图 3 ResNet50 与 FPN 结构
Fig. 3 Structure of ResNet50 and FPN

3.3 注意力机制

文献[9]创造性地提出 CBAM 注意力模块, 其不

同于传统单模块的注意力模块, 仅使用通道注意力机制或空间注意力机制, CBAM 注意力模块则将通道

注意力机制和空间注意力机制同时引入计算机视觉任务中,模拟人类的视觉习惯,让特征提取网络对潜在重要目标区域赋予较大的权重,使网络更加聚焦重点区域的检测。因此,所提算法引入CBAM注意力模块,提高特征提取网络对遥感图像中飞机目标区域的关注度。

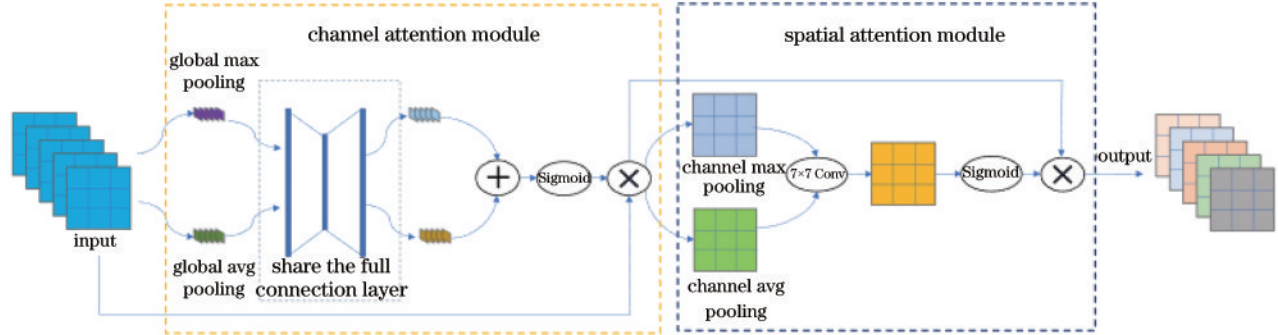


图4 CBAM结构

Fig. 4 Structure of CBAM

给定输入特征图 $F_{input} = (C, H, W)$, 其中, C 为输入特征图的通道数、 H 和 W 分别为输入特征图的高度和宽度。通道注意力机制分别对输入特征图进行全局最大池化和全局平均池化操作, 分别生成 $F_{c_maxpool} = (C, 1, 1)$ 和 $F_{c_avgpool} = (C, 1, 1)$ 序列; 2 个序列分别经过 1 个共享全连接层参数且只含 1 个隐含层的 3 层感知器, 生成 $F_{fc_maxpool} = (C, 1, 1)$ 和 $F_{fc_avgpool} = (C, 1, 1)$ 这 2 组特征向量; 将 2 组特征向量相加, 并使用 Sigmoid 函数激活, 得到通道加权特征向量; 将输入特征图 $F_{input} = (C, H, W)$ 与通道加权特征向量相乘得到通道加权特征图 $F_{c_output} = (C, H, W)$ 。空间注意力机制将通道加权特征图 $F_{c_output} = (C, H, W)$ 作为输入, 在通道维度上分别对特征图上的每一元素进行通道最大池化和通道平均池化, 得到特征图 $F_{s_maxpool} = (1, H, W)$ 和 $F_{s_avgpool} = (1, H, W)$, 将其拼接生成 $F_{s_pool} = (2, H, W)$; 使用步长为 1 的 7×7 卷积核进行降维操作, 并使用 Sigmoid 函数激活, 生成空间位置权重特征图 $F_{s_conv} = (1, H, W)$; 将输入特征图与空间权重特征图相乘, 得到空间加权特征图 $F_{s_output} = (C, H, W)$ 。

文献[9]创造性地提出CBAM注意力模块,将这种轻量模块加入ResNet50中的每一个bottleneck之后,使得网络检测性能得到进一步提升。文献[25-26]将CBAM模块加入特征提取网络与FPN之间的连接处,提升目标特征的权重。CBAM注意力模块虽然有效提升了网络的特征提取能力,但也带来了额外的运算量,严重限制了模型的检测速度。由于飞机目标在遥感图像中基本以小目标形式出现,而小目标的检测主要依赖于较为浅层的特征图,为了尽可能不增加网络参数数量和浮点数运算量、减少CBAM的插入数量,仅在ResNet41-3的主干部分插入一个CBAM注意力

模块,其包含两种类型的注意力机制,分别是通道注意力机制和空间注意力机制,CBAM注意力模块具体结构如图4所示。对于特征提取网络中的某一中间特征图,CBAM注意力模块在时间序列上先后对输入特征图进行通道注意力加权与空间注意力加权。

模块,即在ResNet41-3网络中的maxpool与stage 0之间插入CBAM注意力模块。同时,主干部分的特征图是浅层特征图,通道维数少,不适合较高的下采样率,为此改进了CBAM中通道注意力机制的下采样率。在光学遥感飞机图像数据集上进行的相应实验表明,在ResNet41-3的主干网络上引入CBAM注意力模块,并将通道注意力的下采样率设置为8倍后,网络的mF1评价指标分数由87.48分提升到88.78分,同时参数量仅增加 1×10^3 。

3.4 中线单帧预测

一般的遥感图像尺寸较大,而现有深度学习算法以及计算硬件资源难以实现大尺寸图像的计算,因此,需要将大尺寸的遥感图像切割成数张小尺寸图像。为提高图像的计算效率,使用 1024×1024 的窗口大小在遥感图像上以一定的步长进行滑动切图。对于训练集图像的切割,当滑动窗口的边界将飞机切分成两部分时,设置切割阈值为0.7。当被切割的标注框与原标注框的交并比大于0.7时,判定这是一个可以参与训练的标注框,并在子图标签文件中生成相应的标注框坐标,如图5中的2号滑动窗口;当被切割的标注框与原标注框的交并比小于0.7时,判定这不是一个可以检测的飞机,则不会在子图标签文件中赋予相应标注框,如图5中的1号和3号滑动窗口。为确保滑动切割过程中每个目标飞机在子图中至少有一幅子图能够完整包含该目标飞机,滑动窗口的重叠宽度应至少大于所有飞机目标中的最大尺寸。同时,使用两种滑动步长分别对遥感图像进行切割,一方面尽可能减少飞机目标被切割带来的影响,另一方面实现训练数据的扩充增强。

由于在滑动窗口切割的过程中,以小于窗口尺寸



图 5 滑动窗口切割过程

Fig. 5 Cutting process of sliding window

的步长进行滑动,滑动窗口的上一帧与下一帧会产生重叠区域,因此,在算法推理过程中对重叠区域的飞机目标会进行两次预测,如图 6 中的①②④号飞机。

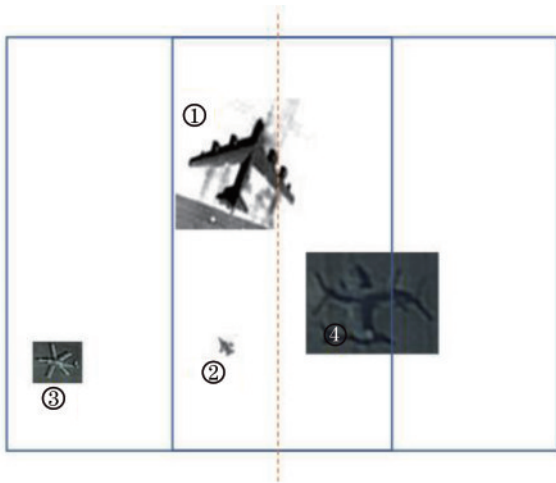


图 6 预测结果后处理

Fig. 6 Post-processing of prediction results

对于重叠区域内的飞机,其左右两侧的背景信息丰富度在上下两帧图像中是不一样的,如图 6 中的④号飞机,左侧帧图像未能包含飞机右侧的背景信息,同时也缺少完整的飞机信息,网络会因为看不到右侧机翼信息将其预测为直升机类,而右侧帧图像完整包含了整个飞机左右两侧信息,网络会将其正确预测为鱼鹰运输机类。同一架飞机在前后两帧图像中预测成不同的类别,这对算法的精确度性能影响是致命的。针对这个问题,设计了一种中线单帧预测的后处理方法:在重叠区域内,设置一条中线(如图 6 中的虚线);对于非边界子图的预测结果,当预测的飞机右侧边界框位于中线上或者位于中线左侧时,认定左侧帧图像的预测结果是正确的,同时删除右侧帧图像关于此目标的预测结果;当预测的飞机右侧边界框位于中线右侧时,认定右侧帧图像的预测结果是正确的,同时删除左侧帧图像关于此目标的预测结果。因此,对于图 6 中的飞机,②③号飞机归左侧帧图像预测,①④号飞机归右侧帧图像预测。在光学遥感飞机图像数据集上进行的相应实验表明,采用中线单帧预测的后处理方法,能有效避免重复预测结果不一致的现象。

4 实验与结果分析

4.1 数据集分析与预处理

本实验数据集来源于国产自主产权系列卫星在多种成像条件下的机场图像,命名为 RSAICP-plane。数据集影像的地面分辨率约为 0.5~1 m,可分辨机身长度大于 8 m 的飞机,尺寸为 4096×4096,共 430 幅图像。其中,训练集有 308 幅图像,测试集有 122 幅图像,一共包含 11 类飞机目标,9069 架飞机,飞机类别切片示例如图 7 所示。由于飞机型号众多,单一类别飞机可能由一个或多个型号组成,因此同一类别飞机外观可能会有不一致现象。

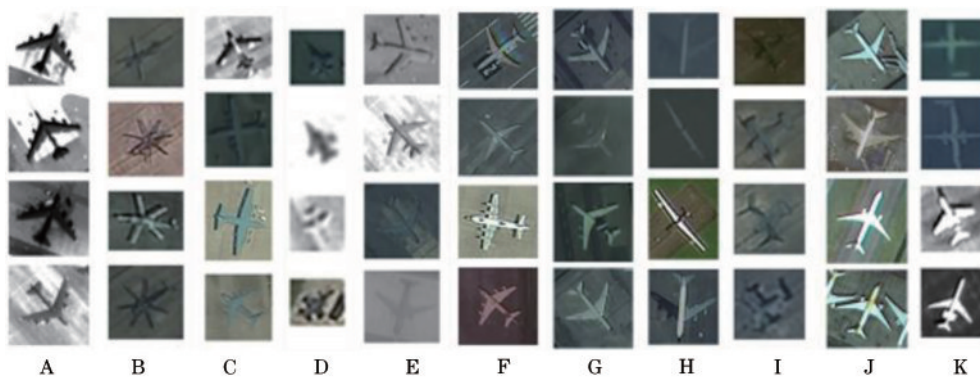


图 7 飞机类别示例

Fig. 7 Examples of aircraft categories

飞机各类别数量分布如图 8 所示,可以发现数据集中存在一定的类别不均衡现象,不同大小的飞机数量分布如图 9 所示。飞机大小定义遵循 COCO 数据

集^[27]中对目标大小的定义标准,即飞机标注框小于 32 pixel×32 pixel 的为小飞机,飞机标注框大于 32 pixel×32 pixel 且小于 96 pixel×96 pixel 的为中飞

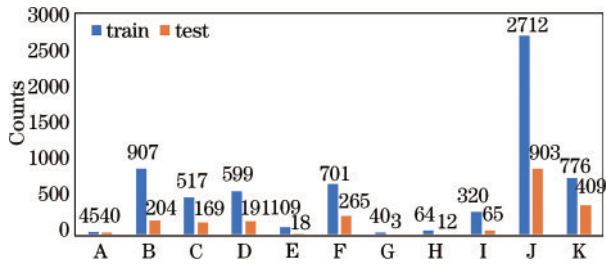


图 8 各类飞机数量分布

Fig. 8 Distribution of aircraft number by class

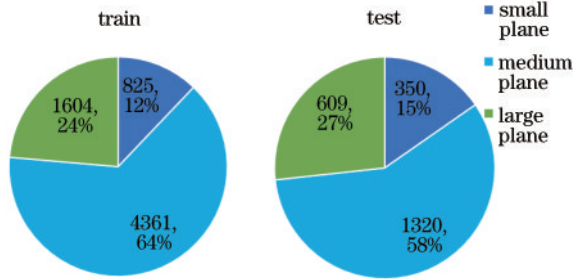


图 9 飞机大小分布

Fig. 9 Distribution of aircraft size

机,飞机标注框大于 $96 \text{ pixel} \times 96 \text{ pixel}$ 的为大飞机。从图 8 和图 9 可以发现,本数据集中小飞机数量占比超过 10%,给准确检测图像中的飞机带来巨大挑战。

在构建输入训练集与测试集时,为提高小目标的检测能力,会对原图使用双线性插值方法放大 1.5 倍,得到大小为 6144×6144 的图像。构建训练集时,以 1024×1024 的滑动窗口以及 900 和 644 的步长对图像进行切割处理,共生成 40040 张 1024×1024 大小的子图,去除所有不包含标注飞机的图像,剩下 7608 张子图,由此构成网络输入训练集。而对于测试集,由于 380 的重叠区域大小能够完全覆盖数据集中最大尺寸的飞机目标,同时为了减少测试时间消耗,仅使用 644 的步长进行滑动切割,共生成 9882 张 1024×1024 大小的子图。

4.2 实验环境与参数配置

所有实验在 Ubuntu18.04 LTS 系统下进行,使用深度学习框架为 PyTorch1.7.1,采用商汤科技开源的 mmdetection 目标检测框架搭建 Faster R-CNN 算法, mmdetection 版本为 0.6.0, GPU 为一个 16 GB 的 Nvidia Tesla V100。

训练图片和测试图片输入大小均为 1024×1024 , batch_size 默认设置为 6,每次训练 30 个 epoch,使用随机梯度下降法(SGD)对网络参数进行调整。初始学习率为 0.02,动量设置为 0.9,权重衰减率设置为 0.0001。数据增强策略为以 0.4 的概率对输入图像进行随机亮度变化、随机对比度变化和随机饱和度变化。

评价指标采用所有类别的算术平均 F1 (mF1) 分数作为模型的评价指标。mF1 的计算公式为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (2)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (3)$$

$$F_{F1} = 2 \times \frac{P \times R}{P + R}, \quad (4)$$

$$F_{mF1} = \frac{\sum_{i=1}^c F_{F1i}}{c} \times 100, \quad (5)$$

式中: P 为精确度; R 为召回率; N_{TP} 为测试结果中被正确识别的正样本数量; N_{FP} 为测试结果中被错误识别的正样本数量; N_{FN} 为测试结果中被错误识别的负样本数量; c 为类别总数。对于位置判定,交并比阈值设置为 0.5,当预测边界框与真实边界框的交并比大于 0.5 时,视为正确预测。F1 分数为精确度与召回率的调和平均数,mF1 分数为所有类别的算术平均数。

4.3 实验结果分析

4.3.1 轻量化特征提取网络

为验证对 ResNet50 特征提取网络进行轻量化处理的有效性,在 RSAICP-plane 数据集上进行相应的训练与测试,并在 Res2Net50^[28]与 ResNext101^[29]做同样的轻量化处理,实验结果如表 1 所示。其中,ResNet41-3、Res2Net41-3、ResNext92-3 分别为 ResNet50、Res2Net50、ResNext101 删去最后一个 stage 模块后得到的轻量网络。从表 1 可以看出,ResNet41-3、Res2Net41-3、ResNext92-3 的参数量相比原网络分别减少了 38.4%、38.7%、26.0%,mF1 分数分别提高了 1.7%、5.3%、29.2%,检测速度分别提高了 1.9 frame/s、1.6 frame/s、0.8 frame/s。同时为验证深层低分辨率的特征层对遥感飞机检测的影响,对 ResNet50 设计了一种只输出 4 种不同分辨率的 FPN 结构,即删去 FPN 中的 maxpool 层,使 FPN 只输出 4 种与 ResNet50 同样分辨率大小的特征图,记为 ResNet50+FPN-4。对于最大飞机尺寸不超过 $448 \text{ pixel} \times 448 \text{ pixel}$ 的检测任务来说,下采样率小于 1/32 的深层低分辨率特征图难以提取到小目标的特征信息,表明深层低分辨率特征图不利于小目标的检测,同时这部分特征图对网络结构来说是冗余的,会影响检测性能。

表 1 特征提取网络轻量化前后性能对比

Table 1 Performance comparison of feature extraction network before and after lightweight

Feature extraction network	mF1	Size / 10^6	FPS
ResNet50	85.9822	41.9	16.3
ResNet41-3	87.4802	25.8	18.2
Res2Net50	78.8559	41.6	14.1
Res2Net41-3	83.0056	25.5	15.7
ResNext101	59.8662	60.0	10.6
ResNext92-3	77.3568	44.4	11.4
ResNet50+FPN-4	86.5830	41.4	17.1

为验证所提轻量化方法在不同目标检测框架中的泛化性能,在 RSAICP-plane 数据集上对 Cascade R-CNN 目标检测算法与 FCOS 目标检测算法的特征提取网络做同样的轻量化处理,实验结果如表 2 所示。从表 2 可以发现,所提特征提取网络轻量化方法具有强大的适用性,在不同算法框架上同样有着优秀的性能,且 Faster R-CNN 目标检测算法与其他目标检测相比,在检测精度与检测速度上具有较好的平衡,综合性能优秀。综合分析,所提轻量化特征提取网络方法在光学遥感图像飞机检测任务中具有显著优势,使用更少参数量,实现更好的检测性能和更快的检测速度。

为进一步验证所提轻量化方法在不同目标检测任务的适用性,分别在 SAR-Ship 数据集与 CASIA-aircraft 数据集上进行相应对比实验,结果如表 3 所示。其中, SAR-ship 数据集中图像尺寸为 256 pixel \times 256 pixel,仅有 1 种 ship 类别;CASIA-aircraft 数据集中图像尺寸为 399 pixel \times 399 pixel,一共包含 62 种飞

表 3 不同目标检测任务轻量化前后性能对比

Table 3 Performance comparison of different target detection tasks before and after lightweight

Dataset	Method	mF1	Size /10 ⁶	FPS
RSAICP-plane	Faster R-CNN+ResNet50	85.9822	41.9	16.3
	Faster R-CNN+ResNet41-3	87.4802	25.8	18.2
SAR-ship	Faster R-CNN+ResNet50	90.2748	41.9	33.1
	Faster R-CNN+ResNet41-3	91.1617	25.8	37.9
CASIA-aircraft	Faster R-CNN+ResNet50	96.9330	42.1	25.5
	Faster R-CNN+ResNet41-3	97.3240	26.0	27.8

4.3.2 CBAM 注意力机制

注意力机制的本质是对特征图权重的一种加权平均过程,使网络更加注重目标区域。CBAM 注意力机制是一种组合式的注意力方法,包含通道注意力和空间注意力,通道注意力通过对通道进行加权,使检测网络偏重重要特征通道,空间注意力通过对特征图上的空间位置进行加权,使检测网络偏向关注目标区域特征。为验证 CBAM 注意力机制是否能够提升检测性能,同时尽可能减少额外参数的引入,将 CBAM 注意力模块加入 ResNet41-3 特征提取网络的 maxpool 层与 stage 0 之间。同时为验证 CBAM 中通道注意力的下采样率对检测性能的影响,分别设置 16 倍、8 倍和 4 倍的下采样率作为对照,在 RSAICP-plane 数据集上进行相关实验,结果如表 4 所示。实验结果表明,当 CBAM 模块加入 ResNet41-3 特征提取网络的 maxpool 层与 stage 0 之间时,特征提取网络能够更加聚焦飞机区域特征,使检测性能得到提升,且仅增加微小的额外参数量。当下采样率为 8 倍时,网络取得最好的检测性能,mF1 分数提升到 88.78 分。

将 ResNet41-3_CBAM-8 特征提取网络生成的特征图线性变换到 0~255 区间进行分析,得到图 10~

表 2 不同目标检测算法轻量化前后性能对比

Table 2 Performance comparison of different target detection algorithms before and after lightweight

Method	mF1	Size /10 ⁶	FPS
Faster R-CNN+ResNet50	85.9822	41.9	16.3
Faster R-CNN+ResNet41-3	87.4802	25.8	18.2
Cascade R-CNN+ResNet50	87.1740	69.2	10.7
Cascade R-CNN+ResNet41-3	88.5346	53.1	12.9
FCOS+ResNet50	74.1348	32.1	16.9
FCOS+ResNet41-3	78.8771	16.1	18.5

机型号类别。从表 3 可以发现,所提轻量化方法在不同遥感图像目标检测任务中同样具有优秀性能,用更少参数量,实现更高的检测精度与更快的检测速度。综合分析,所提轻量化特征提取网络方法在遥感图像目标检测任务中具有显著优势,有效减少模型冗余参数、提升模型计算效率,同时提高模型精确度。

表 4 CBAM 注意力机制性能对比

Table 4 Performance comparison of CBAM attention mechanism

Feature extraction network	mF1	Size /10 ⁶	FPS
ResNet41	87.4802	25.8	18.2
ResNet41-3_CBAM-16	87.7677	25.8005	15.3
ResNet41-3_CBAM-8	88.7803	25.801	15.3
ResNet41-3_CBAM-4	87.9745	25.802	15.3

12。从图 10~12 可以发现,ResNet41-3 主干网络上的特征图在经过通道注意力子模块后,对 64 个通道分别赋予不同权重,例如:图 10、图 11 中的 25 号、63 号通道,在经过通道注意力后,通道权重得到弱化;图 10、图 11 中 0 号、44 号通道,在经过通道注意力后,通道权重得到增强。综上可知,在通道维度上,每一个通道对网络检测性能的贡献不一样,通道注意力能够自动为特征判别性大的通道赋予大的权重。

从图 12 可以发现,空间注意力子模块具有指示网络更加聚焦目标区域的作用。图 12(b)为 ResNet41 主干网络未加入空间注意力的 14 号通道特征图,可以发现左下角三架被云雾遮挡的飞机特征与背景特征混合,无法区分两者。图 12(c)为 ResNet41 主干网络加

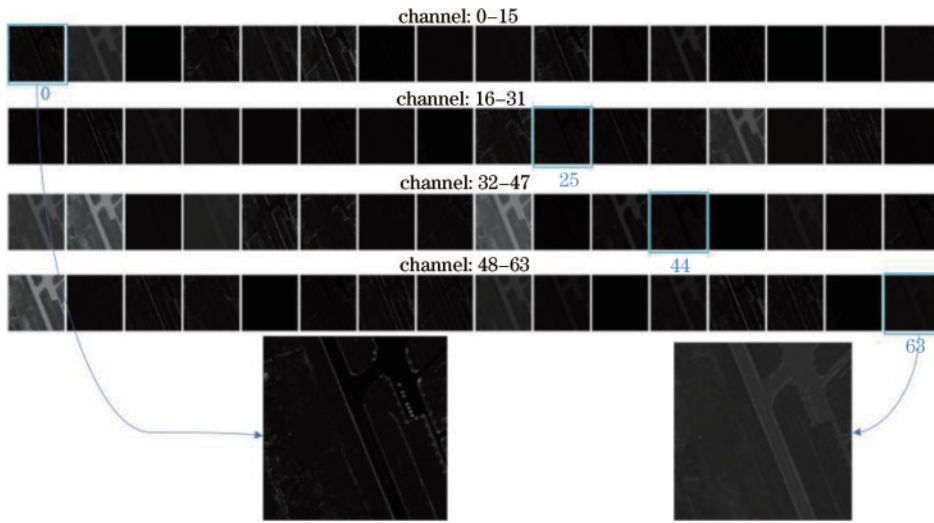


图 10 加入通道注意力前的特征图

Fig. 10 Feature map before adding channel attention

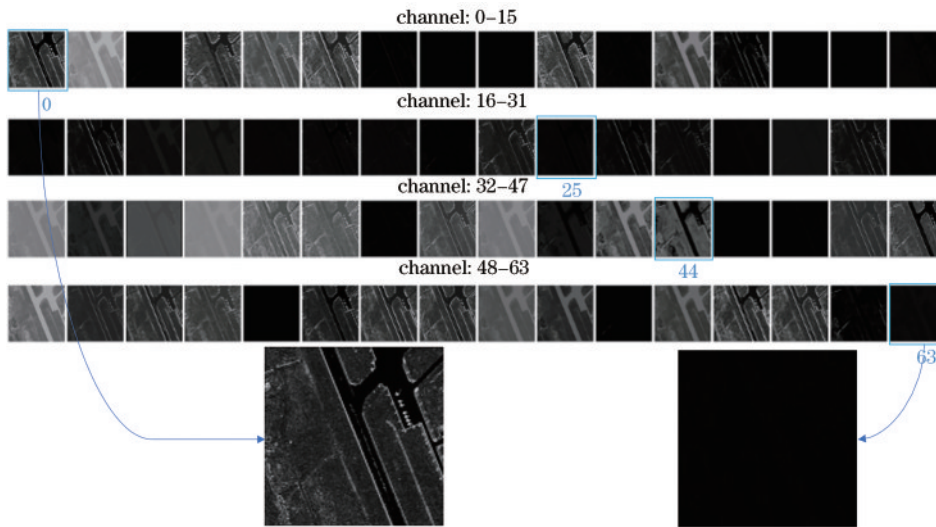


图 11 加入通道注意力后的特征图

Fig. 11 Feature map after adding channel attention

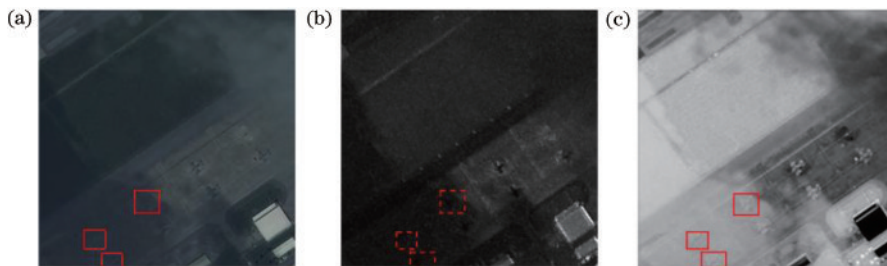


图 12 加入空间注意力前后特征图对比。(a)输入;(b)未加入CBAM的14号通道特征图;(c)加入CBAM后的14号通道特征图
 Fig. 12 Feature map comparison before and after adding spatial attention. (a) Input; (b) feature map of channel 14 that is not added to CBAM; (c) feature map of channel 14 after CBAM added

入空间注意力后的14号通道特征图,可以发现即使存在云雾遮挡,左下角三架飞机特征也能较好地与背景特征区分开。以上分析表明,空间注意力可以很好地定位飞机目标区域,有效克服云雾对飞机遮挡的影响,提升网络检测性能。

4. 3. 3 中线单帧预测后处理与多尺度训练

为避免滑动窗口切割过程中的重叠区域,同一个目标两次预测,前后两帧图像预测结果不同的现象发生,在推理过程中采用中线单帧预测的后处理方法。为使模型应对多尺度飞机目标更具鲁棒性,在训练过

程中随机使用 1024×1024 和 800×800 这 2 种尺度输入进行训练。在 RSAICP-plane 数据集上进行相关实验,结果如表 5 所示。实验结果表明,多尺度训练是一种有效提升模型多尺度泛化的训练技巧,同时中线单

帧预测的后处理方法能够有效避免重叠区域内的飞机两次预测结果不一致的现象。图 13 中,在使用中线单帧的后处理方法后左下角的飞机仅出现一次且正确的预测结果。

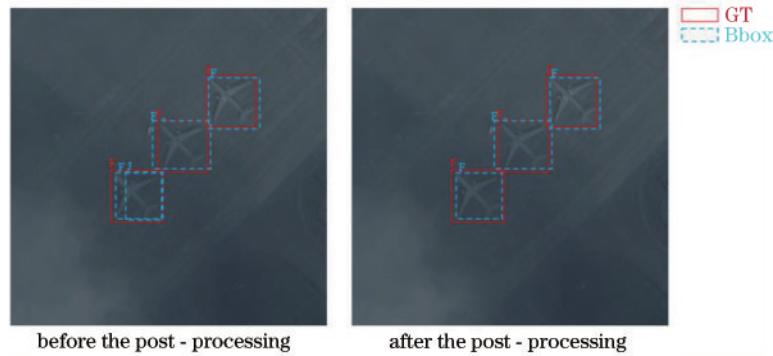


图 13 采用后处理前后对比

Fig. 13 Comparison before and after use of post-processing

表 5 中线单帧预测与多尺度训练实验结果

Table 5 Experiment results of midline single frame prediction and multi-scale training

Feature extraction network	mF1
ResNet41-3	87.4802
ResNet41-3_ms	88.9047
ResNet41-3_ms*	88.9465
ResNet41-3_CBAM-8	88.7803
ResNet41-3_CBAM-8_ms	88.7926
ResNet41-3_CBAM-8_ms*	88.9682

Notes: ms stands for multi-scale training and * stands for post-processing using midline single frame prediction.

4 结 论

以融合 FPN 和 ROI align 网络的 Faster R-CNN 目标检测算法为基线模型,针对光学遥感图像飞机检测的特点,采用带重叠区域的滑动窗口切割方法对大尺寸遥感图像进行切割处理,使模型训练得以正常进行。对 ResNet50 特征提取网络进行轻量化处理,并创造性地仅在 ResNet 主干网络中引入 CBAM 注意力模块对基线模型进行改进。在光学遥感数据集上进行的相关实验表明,轻量化后的特征提取网络 ResNet41-3 参数量比原网络减少了 38.4%,检测速度提高了 11.7%,mF1 分数比基线模型提高了 1.7%。CBAM 模块有效增强了网络对重要特征的关注,一定程度上克服了云雾遮挡对检测的影响,进一步将 ResNet41-3 模型的 mF1 分数提高了 1.5%。在后处理阶段,提出的中线单帧预测方式,有效解决了重叠区域内目标的两次预测结果不一致现象。采用图像多尺度输入训练的策略,使模型的 mF1 分数相比基线模型提高了 3.5%,达到 88.97 分,实现了光学遥感图像中飞机的准确识别与定位。

参 考 文 献

- [1] 张欣. 基于 YOLO 算法的遥感图像飞机目标检测技术研究[D]. 长春: 中国科学院长春光学精密机械与物理研究所, 2021: 1-2.
Zhang X. Research on aircraft target detection technology in remote sensing image based on YOLO algorithm[D]. Changchun: Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 2021: 1-2.
- [2] 张磊, 张永生, 于英, 等. 遥感图像倾斜边界框目标检测研究进展与展望[J]. 遥感学报, 2022, 26(9): 1723-1743.
Zhang L, Zhang Y S, Yu Y, et al. Research progress and prospect of target detection in inclined bounding box of remote sensing images[J]. National Remote Sensing Bulletin, 2022, 26(9): 1723-1743.
- [3] 聂光涛, 黄华. 光学遥感图像目标检测算法综述[J]. 自动化学报, 2021, 47(8): 1749-1768.
Nie G T, Huang H. A survey of object detection in optical remote sensing images[J]. Acta Automatica Sinica, 2021, 47(8): 1749-1768.
- [4] 沙苗苗, 李宇, 李安. 改进 Faster R-CNN 的遥感图像多尺度飞机目标检测[J]. 遥感学报, 2022, 26(8): 1624-1635.
Sha M M, Li Y, Li A. Multiscale aircraft detection in optical remote sensing imagery based on advanced Faster R-CNN[J]. National Remote Sensing Bulletin, 2022, 26(8): 1624-1635.
- [5] 赵佳琦, 张迪, 周勇, 等. 基于深度强化学习的遥感图像可解释目标检测方法[J]. 模式识别与人工智能, 2021, 34(9): 777-786.
Zhao J Q, Zhang D, Zhou Y, et al. Interpretable object detection method for remote sensing image based on deep reinforcement learning[J]. Pattern Recognition and Artificial Intelligence, 2021, 34(9): 777-786.
- [6] 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.
- [7] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [8] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [9] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018*. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [10] Liu L, Ouyang W L, Wang X G, et al. Deep learning for generic object detection: a survey[J]. *International Journal of Computer Vision*, 2020, 128(2): 261-318.
- [11] Lin Y D, He H J, Yin Z K, et al. Rotation-invariant object detection in remote sensing images based on radial-gradient angle[J]. *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(4): 746-750.
- [12] Grabner H, Nguyen T T, Gruber B, et al. On-line boosting-based car detection from aerial images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2008, 63(3): 382-396.
- [13] 董永峰, 仇长涛, 汪鹏, 等. 基于深度学习的光学遥感图像飞机检测算法[J]. *激光与光电子学进展*, 2020, 57(4): 041007.
Dong Y F, Zhang C T, Wang P, et al. Airplane detection of optical remote sensing images based on deep learning[J]. *Laser & Optoelectronics Progress*, 2020, 57(4): 041007.
- [14] 肖振久, 杨玥莹, 孔祥旭. 基于改进 YOLOv4 的遥感图像目标检测方法[J]. *激光与光电子学进展*, 2023, 60(6): 0628009.
Xiao Z J, Yang Y Y, Kong X X. Object detection method based on improved YOLOv4 network for remote sensing images[J]. *Laser & Optoelectronics Progress*, 2023, 60(6): 0628009.
- [15] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [16] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [17] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [18] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [19] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016*. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [20] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.
- [21] Duan K W, Bai S, Xie L X, et al. CenterNet: keypoint triplets for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6568-6577.
- [22] Law H, Deng J. CornerNet: detecting objects as paired keypoints[J]. *International Journal of Computer Vision*, 2020, 128(3): 642-656.[LinkOut]
- [23] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9626-9635.
- [24] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.
- [25] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [26] 高梦婷, 孙晗, 唐云祁, 等. 基于改进 YOLOv5 的指纹二级特征检测方法[J]. *激光与光电子学进展*, 2023, 60(10): 1010006.
Gao M T, Sun H, Tang Y Q, et al. Fingerprint second-order minutiae detection method based on improved YOLOv5[J]. *Laser & Optoelectronics Progress*, 2023, 60(10): 1010006.
- [27] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. *Computer vision-ECCV 2014*. Lecture notes in computer science. Cham: Springer, 2014, 8693: 740-755.
- [28] Gao S H, Cheng M M, Zhao K, et al. Res2Net: a new multi-scale backbone architecture[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(2): 652-662.
- [29] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5987-5995.