

基于深度学习的光学遥感图像目标检测综述

冯长峰,王春平,付强,张冬冬

(陆军工程大学石家庄校区电子与光学工程系,河北石家庄 050003)

摘要:目标检测是光学遥感图像解译的核心环节,广泛应用于情报侦察、土地利用、城市规划等领域。近年来,深度学习的发展成熟使光学遥感目标检测的精确度和效率得到大幅提升。本文首先综述了基于深度学习的通用目标检测算法;然后介绍了当前常用的光学遥感图像目标检测数据集并依据数据特点分析了数据集的发展趋势;接着根据遥感图像检测难点,从任意方向、多尺度、小目标、密集分布、复杂背景5个方面详细梳理了算法的优化方案;最后展望了光学遥感图像目标检测研究的发展方向。

关键词:目标检测;深度学习;光学遥感图像;数据集;机器视觉

中图分类号:TP391.41;TP73 **文献标识码:**A **DOI:**10.3969/j.issn.1001-5078.2023.09.002

Survey of object detection in optical remote sensing images based on deep learning

FENG Chang-feng, WANG Chun-ping, Fu Qiang, ZHANG Dong-dong

(Department of Electronic and Optical Engineering, People Liberation Army Engineering University-Shijiazhuang, Shijiazhuang 050003, China)

Abstract: Object detection is the core part of optical remote sensing image interpretation, which has wide applications in intelligence reconnaissance, land resource utilization, urban planning and other fields. In recent years, the maturation of deep learning has led to significant improvements in the accuracy and efficiency of optical remote sensing object detection. In this paper, the general target detection algorithms based on deep learning is reviewed first. Secondly, the current commonly used optical remote sensing image object detection data set is introduced and the development trend of the data set based on the data characteristics is analyzed. Then, according to the difficulties of remote sensing image detection, the optimization scheme of the algorithm is sorted out in detail from five aspects: arbitrary direction, multi-scale, small target, dense distribution, and complex background. Finally, the development direction of optical remote sensing image object detection research is prospected.

Keywords: object detection; deep learning; optical remote sensing image; dataset; machine vision

1 引言

光学遥感图像是指使用可见光波段传感器采集

的航空和卫星影像,其包含丰富的细节信息,能够直观反映地物目标的形状、颜色和纹理等特征^[1]。光

基金项目:军内科研项目(No. KYSZJWJK2236)资助。

作者简介:冯长峰(1999-),男,硕士研究生,主要从事图像处理、机器视觉与目标检测方面的研究。E-mail: fengcf1999@163.com

通讯作者:付强(1981-),男,博士,讲师,主要从事目标检测、自动控制、图像工程方面的研究。E-mail: Fu_Qiang@aeu.edu.cn

收稿日期:2022-10-25; **修订日期:**2022-12-12

学遥感图像目标检测的主要任务是对图像中的感兴趣目标(通常为飞机、船舶、建筑物等)进行分类和定位,作为航空和卫星图像分析领域的基础任务,其在情报侦察、目标监视、交通疏导、灾害救援等领域发挥着重要作用,同时也是开展目标跟踪、场景分类、图像分割等任务的前提。

传统的遥感图像目标检测算法可分为模板匹配方法和机器学习方法。模板匹配方法通过手工设计生成目标模板,随后以特定评价标准计算图像与模板匹配度来寻找目标。机器学习方法则先通过滑动窗口生成大量潜在目标区域,之后对每个区域进行特征提取,最后使用分类器分类以剔除虚假目标获得最终的检测结果。2012年 AlexNet 首次被应用于图像分类,取得了性能上的跨越性突破,此后学者开始将深度学习应用到目标检测任务并取得巨大成功,由此开启了深度学习目标检测算法时代^[2]。

近年来,随着传感器、卫星、航拍飞行器等技术的发展成熟,遥感数据量成指数级增长,图像的空间分辨率也由中等分辨率(4~70 m)和高分辨率(1~4 m)提升为超高分辨率(VHR)。超高分辨率的图像能够提供更多的空间和纹理信息,使许多以往无法察觉的小目标也具备了被检测的可能性。遥感技

术的进步催生出对智能高效检测算法的巨大需求,光学遥感图像目标检测算法性能的提升主要面临两点困难:(1)缺乏高质量标注的光学遥感数据;(2)遥感图像特点鲜明,包含诸多干扰因素。针对上述问题,许多研究团队构建了大型遥感图像目标检测数据集,并就遥感图像特点对通用检测模型进行了一系列改进,产生了大量基于深度学习的光学遥感图像检测算法^[3]。

本文分类总结了基于深度学习的通用目标检测算法,并就其中经典模型进行了详细介绍;整理归纳了当前应用广泛的光学遥感目标检测数据集;基于光学遥感图像目标的 5 个显著特点:任意方向、目标多尺度、小目标分布、目标密集分布以及复杂背景,对当前流行的遥感图像目标检测算法进行了详细梳理;对光学遥感图像目标检测研究的未来方向进行了展望。

2 通用目标检测算法

基于深度学习的通用目标检测算法可分为 3 类:基于锚框的目标检测方法、基于无锚框目标检测方法以及基于 Transformer 目标检测方法。这些方法构建了现今深度学习目标检测的主要框架,部分经典检测模型框架如图 1 所示。

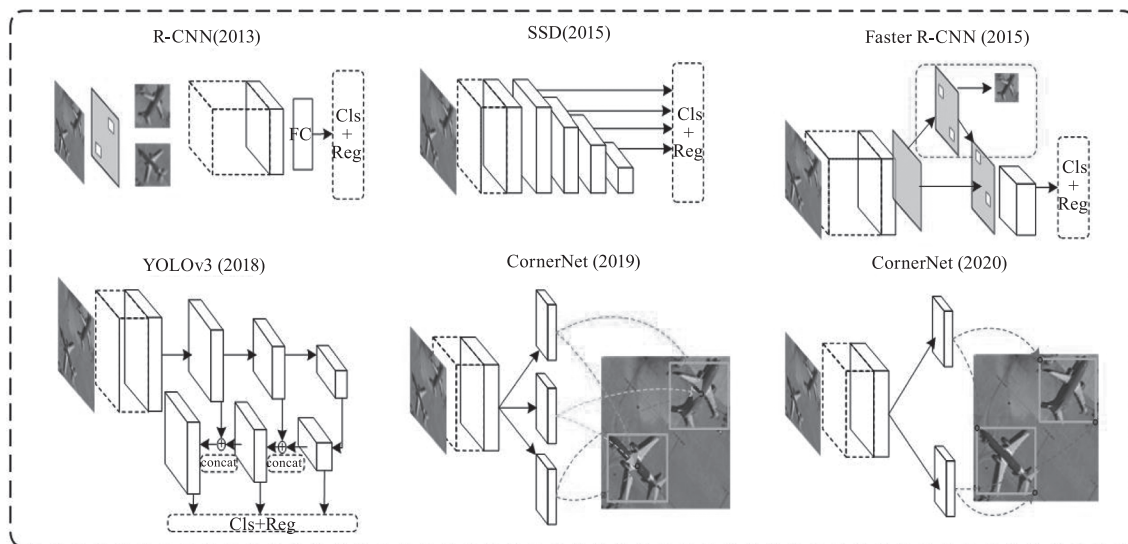


图 1 通用目标检测模型

Fig. 1 Generic object detection models

2.1 基于锚框的目标检测方法

基于锚框的目标检测方法通过显式或隐式的方法,创建一系列具有不同尺寸、长宽比的先验框(锚框),然后对先验框进行分类和回归。根据是否有

显式的区域建议,基于锚框的目标检测方法分为两阶段(Two-Stage)和单阶段(One-Stage)两种。

2.1.1 两阶段检测模型

2014年, Girshick 等^[4]提出 R-CNN,该算法首先

使用选择性搜索选取包含目标的候选区域,随后利用卷积神经网络(Convolutional Neural Network, CNN)对候选区域进行特征提取,最后通过分类器和回归器实现目标的分类和定位。R-CNN极大提升了目标检测任务的准确率,但是候选区域的特征提取过程存在大量的重复计算,严重影响了算法检测速度。鉴于此,He等^[5]提出了SPP-Net,与R-CNN相比只需对整幅图像进行一次特征映射,提升了网络的推理速度。

2015年,Ren等^[6]将区域候选网络(Region Proposal Network, RPN)融入Fast R-CNN网络中,提出了Faster R-CNN,该模型具备较强鉴别能力并可以借助GPU加速,极大提升了检测的精度和速度。以Faster R-CNN为基础,研究人员进行了大量的改进。Lin等^[7]在框架中引入特征金字塔网络(Feature Pyramid Network, FPN),利用深层网络的丰富语义信息来加强空间信息丰富的浅层网络特征,在多尺度检测方面取得重大进展。Mask R-CNN^[8]增添了一个额外的分割支路,通过像素级预测生成目标的分割掩膜,并引入感兴趣区域聚集(RoIAlign)层以解决量化误差问题。

2.1.2 单阶段检测模型

YOLO系列是最早应用于实践的单阶段目标检测算法。YOLO^[9]舍弃了繁琐的区域建议步骤,直接在特征图上对目标的类别、置信度和位置偏移进行预测,其检测速度比Faster R-CNN提高近十倍。YOLOv3^[10]借鉴ResNet构建了Darknet-53主干网络,在保持性能前提下大幅降低了计算量;在三种不同尺度的特征图上进行分类和回归,改善了多尺度目标的检测效果;在预测类别时使用Logistic回归代替Softmax方法,以支持多标签分类任务。YOLOv5在输入端利用Mosaic数据增强技术提升数据丰富度;在主干网中采用跨阶段局部网络结构,减少计算量的同时也保证了准确率;提供了不同规模的变体以满足不同的应用需求。

Liu等^[11]在2016年提出了SSD算法。该模型通过借鉴YOLO算法的网格划分和Faster R-CNN中的锚框机制,同时引入金字塔型特征层级充分挖掘卷积层的特征信息,在一定程度上克服了YOLO算法难以检测小目标、定位不准的缺点,并保证了推理速度。

2.2 基于无锚框目标检测方法

基于锚框的目标检测需要人为设定锚框的尺寸、长宽比等超参数,导致检测器对超参数敏感且泛化能力较差。鉴于上述问题,研究人员提出了无锚框目标检测方法,主要可分为两类:基于内部点的检测方法和基于关键点的检测方法。

基于内部点的检测算法通过预测物体内部点的边界偏移量及类别信息,达到目标检测效果,代表算法有CenterNet^[12]。CenterNet将目标建模为一个中心点,通过中心点处的特征来预测其他目标属性如大小、深度甚至姿态。同时,采用最大池化操作提取峰值点,舍弃了非极大值抑制(NMS)步骤,显著提高了检测速度。

基于关键点的目标检测方法则选择预测物体的多个关键点,并通过关键点集成实现对物体的最终检测,代表算法有CornerNet^[13]。CornerNet采用Hourglass网络提取的高分辨率特征图,分别预测同一对象类别内所有实例的左上角热图、右下角热图及角点嵌入特征向量,随后通过向量相似性匹配,将属于同一物体的角点关联生成最终检测框。

2.3 基于Transformer的目标检测方法

Transformer是一类起源于自然语言处理(Natural Language Processing, NLP)领域的新型深度神经网络,不同于传统网络经过不断卷积获得多层级信息,它主要通过自注意力机制提取数据的内在特征。近年来,基于Transformer的目标检测算法凭借出色检测性能引起了研究者的广泛关注。

Carion等提出的DETR^[14]算法属于较早将Transformer引入目标检测的尝试,其利用卷积神经网络提取特征并送入一个编解码器,直接预测物体位置及分类得分。DETR舍弃了锚框等手工设计组件和后处理步骤,将检测任务简化为集合预测问题,实现了真正意义上的端到端检测。还有研究者尝试利用Transformer作为检测器的主干网络。Dosovitskiy等^[15]首次提出将Transformer作为主干网络用于图像分类,并取得SOTA结果。随后,研究者致力于探索更为通用的框架设计,Liu等在2021年提出了Swin Transformer^[16]网络,在COCO数据集目标检测任务上超过了当时最先进算法的精度,展示了Transformer模型作为计算机视觉主干网络的巨大潜力。

表1总结了不同类别检测方法的原理以及优缺点

点。其中,双阶段检测模型凭借其优良的性能在光学遥感图像目标检测领域得到广泛应用^[17-19],单阶段方法相比双阶段相关研究较少,主要用于对检测速度有更高要求的快速检测场景^[20]。近年来,无锚

框方法逐渐受到研究者的青睐,但是还无法取代锚框方法^[21-22],两者仍属于共存关系。对基于 Transformer 的光学遥感图像目标检测^[23],目前也有一部分尝试,但仍处于初步阶段。

表 1 深度学习目标检测方法比较

Tab. 1 Comparison of object detection approach

Approach		Principle	Advantages	Disadvantages
Anchor-based	Two stage	Regression and classification of the regions of interest	High detection accuracy	Slow detection speed
	One stage	Direct classification and regression of the whole image	Real-time detection	Low detection accuracy
Anchor-free		Generate bounding boxes by searching the key points of the object (corner points or center points)	High adaptability to multiple targets	Slow model convergence
Transformer-based		Predict targets using encode-decode structure based on self-attention	Strong capability of feature extraction	High consumption of computing resources

3 光学遥感图像目标检测数据集

深度学习是一个数据驱动的概念,一个具有挑战性的数据集可以加速该领域的发展。其中最为典型的例子就是 ImageNet 和 MS COCO 促进了自然场景图像分类和目标检测算法的发展。本章首先介绍了一批具有代表性的光学遥感目标检测公开数据集,随后通过对其中图像特性的进一步分析,归纳总结出光学遥感图像区别于常规图像的显著特点。

3.1 光学遥感图像目标检测数据集

遥感技术的广泛应用加上深度学习的数据需要,使得过去十几年中涌现出一批高质量光学遥感目标检测数据集,其中又以 NWPU VHR-10^[24] 和 DOTA^[25] 使用较为广泛。

NWPU VHR-10 是西北工业大学于 2014 年发布的航天遥感目标检测数据集。该数据集包含 10 类目

标:飞机、棒球场、篮球场、桥梁、港口、田径场、船舶、油罐、网球场和车辆,由 715 张空间分辨率为 0.5 ~ 2 m 的 RGB 图像和 85 张分辨率为 0.08 m 的锐化彩色红外图像组成,采用水平框(HBB)标注。

DOTA 由武汉大学遥感国家重点实验室和华中科技大学联合提出,包含 15 类目标:飞机、船舶、储罐、棒球场、网球场、游泳池、田径场、港口、桥梁、大型车辆、小型车辆、直升机、环岛、足球场和篮球场。由包含 188282 个目标实例的 2806 张卫星图像组成,图像尺寸在 800 × 800 到 4000 × 4000 之间,采用旋转框(OBB)标注。图像来源于多个不同的光学传感器,空间分辨率在 0.1 ~ 1 m 之间。

表 2 选取了 7 个公开的光学遥感图像目标检测数据集,并就空间分辨率、实例数、标注种类(注:括号内代表细粒度标注种类)等内容进行对比。

表 2 公开光学遥感图像目标检测数据集

Tab. 2 Public object detection datasets of optical remote sensing image

Dataset	Year	Resolution/m	Image	Instance	Class	Annotation
FAIRIM ^[26]	2022	0.3 ~ 0.8	15 266	1 020 000	5 (37)	OBB
DIOR ^[27]	2020	—	23 463	192 472	20	OBB
RSD-GOD ^[28]	2019	—	18 187	40 990	5	HBB
LEVIR ^[29]	2018	0.2 ~ 1.0	22 000	11 000	3	OBB
DOTA ^[25]	2018	0.1 ~ 1.0	2 806	188 282	15	OBB
HRSC2016 ^[30]	2016	0.13	1 070	2 976	1	OBB
NWPU VHR-10 ^[24]	2014	0.08, 0.5 ~ 2.0	800	3 775	10	HBB

通过分析发现,这些光学遥感数据集的演进存在以下趋势:

(1) 来源更为多样:早期一些研究者通过访问政府机构数据库收集遥感图像。现今数据集大多选

择整合了航拍与卫星影像的商业化平台,如 Digital Globe 和 Google Earth 来获取高分辨率遥感影像;(2)内容更为丰富:DIOR^[27]、LEVIR^[29]等数据集都选择了多数据来源,收集了各种复杂场景下的多分辨率光学遥感图像,Weir 等^[31]提出了首个多视角遥感数据集 SpaceNet MVOI。相比于早期数据集,这些数据集更加贴合真实场景并缓解了由于传感器来源单一、观测区域重复等引起的数据偏差问题;(3)标注信息更为细致:近年来,许多遥感数据集采用旋转边界框来取代水平框,提供定位精度更高的空间位置标注。在类别标注上,一些数据集也对物体进行了细粒度的标注划分。FAIR1M 对目标进行了 5 种大类别、37 个细粒度的分类,并采用旋转框标注,是目前最大的光学遥感图像细粒度目标检测识别数据集。

3.2 光学遥感图像特性

光学遥感传感器往往位于高空或大气层外,对地面呈鸟瞰视角。这使得光学遥感图像展现出不同于常规自然图像的特殊性质,也给目标检测任务带来了新挑战,具体可总结为以下五点:

(1)目标方向随机:如图 2(a)所示,不同于自

然图像中目标的竖直排布,光学遥感图像中的目标方向呈多样随机分布,一般称之为旋转目标。

(2)目标多尺度:如图 2(b)所示,光学遥感目标有着更为显著的多尺度特征。目标类间、类内尺度差异以及空间分辨率变化使图像内目标尺度呈现多样化分布。

(3)小目标:如图 2(c)所示,由于空间分辨率限制,光学遥感图像存在数量可观的小目标分布。这些小目标往往只有几十甚至十几个像素,缺乏检测所必要的形状、颜色、纹理等信息,造成很大程度上的漏检。

(4)目标密集分布:如图 2(d)所示,在停车场、码头等特殊场景中,往往容易出现密集分布情形。稠密排布的目标在特征图上会出现出现边界模糊、高密度目标特征聚集现象,导致误检、定位不准确和目标分类错误。

(5)复杂背景:如图 2(e)所示,光学遥感图像中背景情况更为复杂。一方面,遥感图像背景像素占有率远高于前景,导致检测器降低了对目标的检测效率。另一方面遥感图像中往往包含多种地物,相似地物可能对目标检测造成干扰导致误检。

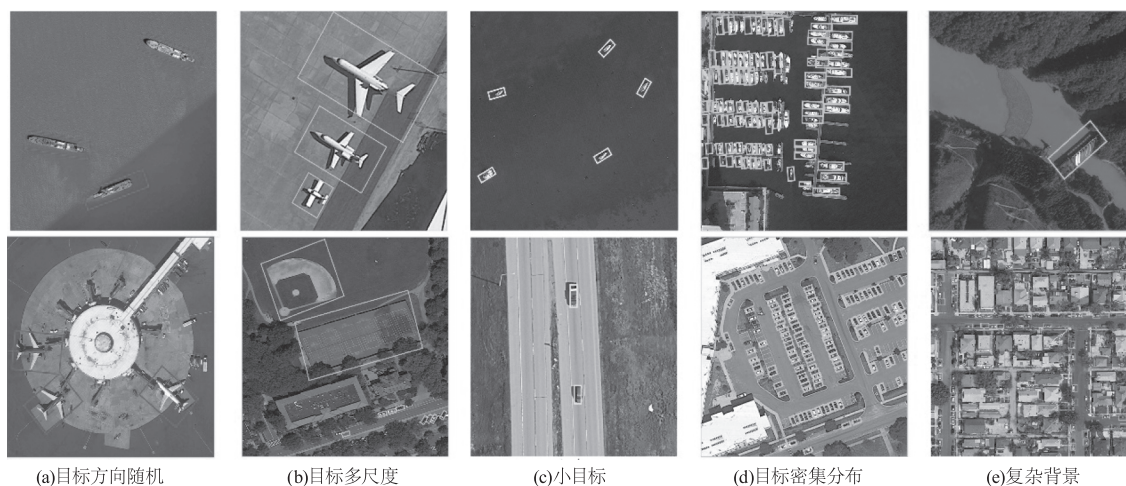


图 2 光学遥感图像特性

Fig. 2 Characteristic of optical remote sensing images

4 光学遥感图像目标检测算法

在常规图像上性能极佳的经典检测算法迁移后效果并不理想,究其原因光学遥感图像自身特性给检测器带来了新挑战。针对上述问题,研究者提出了一系列的改进算法。本章以目标方向随机、多尺度、小目标、密集分布、复杂背景 5 个方向为出发点,对上述改进算法进行梳理总结。

4.1 针对目标方向随机的改进研究

为解决目标方向随机问题,研究者主要采取提取目标旋转不变特征和旋转框检测两种思路。对于提取旋转不变特征,可通过数据增强和添加旋转不变子模块两种方式实现。数据增强就是在训练过程中对图像进行旋转变换扩充训练数据,增强网络对旋转目标的适应能力,从而达到间接提取旋转不变

特征的效果。但数据增强并未从根本上解决旋转敏感问题,更为有效的做法是在网络中添加旋转不变子模块。Cheng 等^[32]通过优化一个新目标函数,在 CNN 特征上加入旋转不变正则化和 Fisher 判别正则化,以实现旋转不变性。

水平框在面对旋转目标时往往会引入无用背景信息影响模型判断,这一现象在大长宽比物体检测中尤为明显。鉴于此,有研究者选择旋转框来检测目标。主流方法一般将生成旋转框作为回归任务,通过回归表征参数生成最终预测结果。常见的旋转框表示方案有五参数法^[33-34]和八参数法^[19],其中

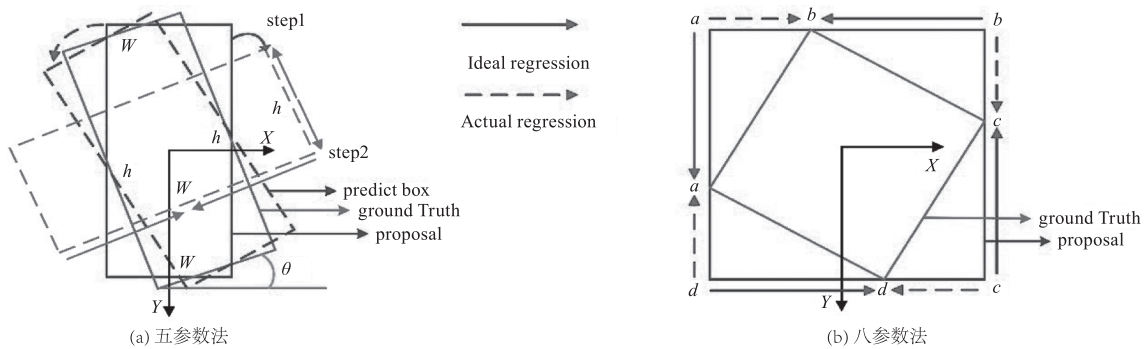


图 3 边界问题示意

Fig. 3 Illustration of boundary problem

近年来,利用实例分割^[36]和角度分类^[37]生成旋转框的方法逐渐流行。分割方法以 Mask R-CNN^[8]为原型,通过寻找目标类别掩膜上的最小外接矩形来生成旋转框,但额外的分割过程也增加了算法复杂度。角度分类方法与五参数法类似,但是将其中角度参数的预测由连续回归转化为离散分类,消除了造成边界问题的不连

续点。五参数法回归目标中心点的坐标以及宽度、高度和旋转角度;八参数法则回归旋转框四个顶点的水平和垂直坐标。

如图 3 所示,基于回归方法在运用中存在边界问题,即模型损失值在边界情况下突然增加。边界问题会造成模型收敛困难,并影响最终检测效果。针对边界问题,RSDet^[35]从改进损失函数的角度出发,弱化边界情况下训练样本的损失值,有效缓解了边界问题。Gliding vertex^[19]模型摒弃了直接回归顶点坐标的做法,通过回归比例因子使顶点在水平框侧边滑动,避免了顺序歧义。

续点。

4.2 针对目标多尺度的改进研究

针对遥感目标多尺度特点,研究者多采用特征融合方法使检测器兼具多尺度检测能力。目前常用的特征融合方法主要包括三种:多层次特征融合、特征金字塔融合以及跨尺度特征融合,其具体结构如图 4 所示。

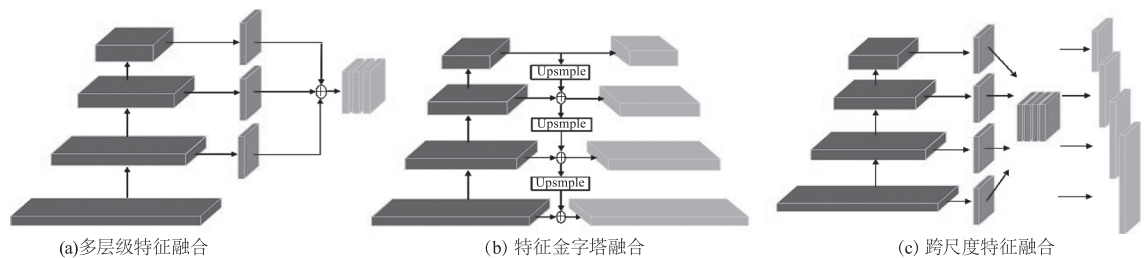


图 4 特征融合方法

Fig. 4 Feature fusion method

多层次特征融合结构将顶层特征与相邻层相结合,弥补了网络深化带来的空间信息缺失^[34,38]。这种方法连接结构最为简洁,但没有充分利用底层丰富的位置信息,导致目标定位效果差。特征金字塔借鉴 FPN 网络,采用横向连接模块将上采样后的深

层特征与相邻的浅层特征结合,直到最底层接收到融合信息反馈,从而形成特征金字塔用于多尺度预测^[39]。跨尺度融合方法结合上述两种策略,通过跨尺度融合模块将特征预先融合,随后生成不同规格特征图用于后续检测^[33,40]。融合后的特征有效兼

顾了空间和语义信息,但是对偏离中心层特征进行了较大跨度的采样,这会导致信息损失。特征金字塔与跨尺度融合相比于第一种方法,都使用了更为复杂的连接方式,在检测效果提升的同时也不可避免引入了大量无用信息。

除上述特征融合方法,现有大多数算法在训练过程中,采用图像金字塔进行多尺度训练,融合了更多图像信息从而加强了模型的特征学习能力,但也增加了大量计算开销。

通过对众多模型的分析发现,构建多尺度特征融合结构并结合一些优化策略如跳跃连接^[40]、高低频特征^[41],已成为一种趋势,可以有效提升遥感图像目标检测精度。

4.3 针对小目标检测的改进研究

小目标检测问题理论上属于多尺度范畴,因此上述多尺度改进方法对小目标检测有很大借鉴意义。除此之外,研究者主要通过增大特征图尺寸方法来提升小目标检测的精度。

增大特征图尺度的其中一种方法是上文提及的多尺度特征融合,通过拼接浅层的大尺寸特征图来增强对小目标的特征表示。另外一种方法是通过反卷积对深层特征上采样来恢复丢失的小目标信息,这里需要注意的是上采样只有在深层特征中仍存在小目标前提下才有意义。也有一部分工作将两种方法相结合,将浅层特征和反卷积输出融合实现互补,这种方法也被称为自下而上特征融合^[42]。

近年来,一些研究者将超分辨率技术应用于光学遥感图像,该技术可以重建物体的鉴别信息,从数据角度出发解决小目标没有足够语义信息的缺陷。当前研究主要围绕基于对抗生成网络(Generative Adversarial Network, GAN)的超分辨率重构展开。Mostafa 提出一个联合超分辨率遥感车辆检测器 Joint-SRVDNet^[43],利用多尺度生成对抗网络输出 2X 和 4X 的超分辨率图像,随后使用 YOLOv3^[10]对重建图像进行检测。CycleGAN 的提出将数据限制由配对图像扩展到图像域,解决了匹配数据不可用的问题。文献^[44]在检测网络汇总使用了改进的 CycleGAN 模型,增强了遥感图像中小物体的检测效果。

4.4 针对目标密集分布的改进研究

相比于自然图像中的密集场景如拥挤道路中的车辆检测,遥感目标由于视角原因不易出现遮

挡现象,其检测难点在于密集分布往往与小目标紧密联系。同时倾斜排列的密集目标采用水平框检测会出现高度重叠,导致非极大抑制时过滤正确目标引起漏检。因此,上述基于小目标和旋转目标的改进都在一定程度上缓解了密集场景。此外,还有许多研究通过特征增强方案来提升单个目标的区分度。

Yang 等^[20]通过设计一个精细化模块,利用逐像素插值将边界框位置信息编码到对应的特征点上,最大程度保证了目标特征和目标中心的对齐,从而提升了密集目标的定位精度。SCRDet++^[45]借助语义分割网络中间层特征来指导特征提取,间接使用注意力机制俩进行密集目标特征增强,从而提升密集目标的边界区分度。

4.5 针对复杂背景的改进研究

注意力机制被视为解决光学遥感图像复杂背景问题的有力方式,其本质是一种资源分配的方法,根据场景中对象的重要性对均匀分布的资源重分配。从原理考虑,注意力机制可以分为空间注意力、通道注意力和联合注意机制。

空间注意力机制能够捕捉像素与像素间关系,通过掩码赋权的方式提升关键区域的特征表达,当前许多研究采用在网络中嵌入自注意力模块的方式生成空间注意力掩码图。Wang 等^[46]将自注意力模块嵌入主干网络,以捕获不同区域相关性并获得更具鉴别性的特征。

通道注意力机制旨在通过学习的方式为不同通道赋予权重系数以强化重要特征。Wu 等^[47]在可变形卷积网络中引入了通道注意力,提出可变形卷积通道注意块,达到了抑制遥感图像背景噪声突出目标的效果。联合注意机制采用空间和通道双重注意机制,在图像和特征通道两次分配最初均匀分布的权重。

注意力机制较好地解决了遥感图像中复杂背景的问题,但是网络需要自主学习图像分布并在空间和通道分配权重,带来了额外的计算开销。

综上,针对光学遥感图像的 5 种显著特点,对不同的改进算法进行了细致的分类总结。其中,各算法验证大多采用了 NWPU VHR-10 和 DOTA 数据集,利用多个检测类别的平均精度值即 mAP 来评估在数据集上的整体性能。典型光学遥感图像目标检测算法性能对比如表 3 所示。

表 3 典型光学遥感图像目标检测算法性能对比

Tab. 3 Performance comparison of typical optical remote sensing image object detection algorithms

Dataset	Algorithm	Backbone	Arbitrarydirection	Scale diversity	Tiny objects	Dense objects	Complex background	mAP
DOTA	AFO-RPN ^[21]	Res-101	✓					80.68
	ADT-Det ^[23]	Res-152	✓		✓	✓		79.95
	OrientedRep ^[22]	Swin-T	✓					77.63
	DCL ^[48]	Res-152	✓	✓				77.37
	SCRDet ++ ^[45]	Res-101	✓		✓	✓	✓	76.81
	R ³ Det ^[20]	Res-152	✓			✓		76.50
	AFC-Net ^[33]	Res-101	✓		✓		✓	76.27
	CSL ^[37]	Res-152	✓		✓	✓		76.17
	Mask OBB ^[36]	ResNeXt	✓				✓	75.33
	Gliding Vertex ^[19]	Res-101	✓		✓		✓	75.02
	SPH-YOLOv5 ^[49]	CSPD-53			✓		✓	71.60
RoI Trans ^[50]	Res-101	✓		✓			69.56	
NWPU VHR-10	SPH-YOLOv5 ^[49]	CSPD-53		✓		✓		98.00
	LFPN ^[41]	Res-101			✓			93.23
	CANet ^[40]	Res-101		✓				92.20
	SCRDet ^[34]	Res-101	✓	✓	✓	✓	✓	91.75
	Sig-NMS ^[17]	VGG16		✓				82.90

结合上述分析和算法对比,可以得到以下结论:首先,不同特点之间具有很强的关联性,甚至是共生的。例如:小目标往往伴随有密集分布情形,其相比于其他尺度的物体也更容易受到复杂背景的干扰。其次,针对不同特点的改进很大程度上也是相通的,例如:注意力机制可以解决多个特殊问题;旋转框在一定程度上也降低了密集分布场景的检测难度。因此,基于光学遥感图像的改进方法应该相互借鉴,使得面向各类特点的多任务优化目标得以协同解决。

5 研究展望

近年来基于深度学习的光学遥感目标检测取得了长足的进步,但是仍有着巨大的发展潜力。我们认为光学遥感图像目标检测在以下 4 个方面值得进一步研究:

(1) 高性能网络结构。作为模型的基础,基础网络结构是提升算法性能的关键所在。近段时间以来遥感检测器性能提升速度有所放缓,表明现有方法已经达到了局限性,因此专门为遥感图像设计高性能网络结构,将更有效的应对遥感图像中的复杂目标,这无疑是一个值得的研究方向。

(2) 弱监督与半监督学习。不断增多的光学遥感载荷平台,为研究者提供了指数级增长的遥感数据,但同时目标检测数据集也存在标注困难、稀有目标缺乏等问题。鉴于此,发展弱监督和半监督学习,在遥感标注不明确或目标无标注情况下训练模型,将会是未来一项重要的研究内容。

(3) 轻量化参数模型。为了提取具有丰富信息表示的特征,检测网络往往使用深度结构,模型收敛需要优化大量参数,当前便携嵌入式设备的低算力也无法实现此类重型模型。如何在尽量保持性能前提下,借鉴剪枝、量化等、知识蒸馏等操作生成轻量化模型,对算法的落地有着重大意义。

(4) 多源遥感数据融合。当前使用的光学遥感图像主要使用 RGB 三个波段,限制了其在夜晚、微光、烟雾、伪装等条件下的应用。随着多光谱、红外、合成孔径雷达传感器的发展,未来天基遥感将包含多源异构数据,进一步发展出多源异构遥感数据为基础的目标检测。

参考文献:

- [1] Nie G T, Huang H. A survey of object detection in optical remote sensing image[J]. Acta Automatica Sinica, 2021,

- 47(8):1749–1768. (in Chinese)
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.
- [2] Wang Henghui, Cao Dong, Zhao Yang, et al. Survey of infrared dim small target detection algorithm based on deep learning[J]. *Laser & Infrared*, 2022, 52(9):1274–1279. (in Chinese)
王恒慧, 曹东, 赵杨, 等. 基于深度学习的红外弱小目标检测算法研究综述[J]. *激光与红外*, 2022, 52(9):1274–1279.
- [3] Zhu X X, Tuia D, Mou L, et al. Deep learning in remote sensing: a comprehensive review and list of resources[J]. *IEEE Geoscience and Remote Sensing Magazine*, 2017, 5(4):8–36.
- [4] 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 (CVPR), July 20–23, 2014, Columbus, USA. New York: IEEE Press, 2014: 580–587.
- [5] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9):1904–1916.
- [6] Ren S, He K, 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] 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 (CVPR), July 21–26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017:2117–2125.
- [8] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22–29, 2017 Venice, Italy. New York: IEEE, 2017:2961–2969.
- [9] 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 (CVPR), June 27–30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016:779–788.
- [10] Redmon J, Farhadi A. Yolov3: an incremental improvement[J]. *arXiv Preprint arXiv:1804.02767*, 2018.
- [11] 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.
- [12] Zhou X, Wang D, Krähenbühl P. Objects as points[J]. *arXiv Preprint arXiv:1904.07850*, 2019.
- [13] Law H, Deng J. Corner net: detecting objects as paired keypoints[J]. *International Journal of Computer Vision*, 2020, 128(3):642–656.
- [14] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[M]. *Computer vision-ECCV 2020. Lecture Notes in Computer Science*. Cham: Springer, 2020, 12346:213–229
- [15] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[J]. *arXiv Preprint arXiv:2010.11929*, 2020.
- [16] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10–17, 2021, Montreal, Canada. New York: IEEE, 2021:10012–10022.
- [17] Dong R, Xu D, Zhao J, et al. Sig-NMS-based faster R-CNN combining transfer learning for small target detection in VHR optical remote sensing imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(11):8534–8545.
- [18] Fu K, Chang Z, Zhang Y, et al. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 161:294–308.
- [19] Xu Y, Fu M, Wang Q, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(4):1452–1459.
- [20] Yang X, Yan J, Feng Z, et al. R3Det: Refined single-stage detector with feature refinement for rotating object[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(4):3163–3171.
- [21] Li J, Tian Y, Xu Y, et al. Oriented object detection in remote sensing images with anchor-free oriented region proposal network[J]. *Remote Sensing, Multidisciplinary Digital Publishing Institute*, 2022, 14(5):1246.
- [22] Li W, Chen Y, Hu K, et al. Oriented reppoints for aerial object detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 19–23, 2022, New Orleans, Louisiana, USA. New

- York: IEEE Press, 2022: 1829 – 1838
- [23] Zheng Y, Sun P, Zhou Z, et al. ADT-Det: Adaptive dynamic refined single-stage transformer detector for arbitrary-oriented object detection in satellite optical imagery [J]. *Remote Sensing*, 2021, 13 (13): 2623.
- [24] Cheng G, Han J, Zhou P, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 98: 119 – 132.
- [25] Xia G S, Bai X, Ding J, et al. DOTA: a large-scale dataset for object detection in aerial images [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18 – 23, 2018, Salt Lake City, UT, USA. New York: IEEE Press 2018: 3974 – 3983.
- [26] Sun X, Wang P, Yan Z, et al. FAIR1M: a benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 184: 116 – 130.
- [27] Zhang Y, Yuan Y, Feng Y, et al. Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57 (8): 5535 – 5548.
- [28] Zhuang S, Wang P, Jiang B, et al. A single shot framework with multi-scale feature fusion for geospatial object detection [J]. *Remote Sensing, Multidisciplinary Digital Publishing Institute*, 2019, 11 (5): 594.
- [29] Zou Z, Shi Z. Random access memories: a new paradigm for target detection in high resolution aerial remote sensing images [J]. *IEEE Transactions on Image Processing*, 2017, 27 (3): 1100 – 1111.
- [30] Liu Z, Wang H, Weng L, et al. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds [J]. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13 (8): 1074 – 1078.
- [31] Weir N, Lindenbaum D, Bastidas A, et al. Spacenet MVOI: a multi-view overhead imagery dataset [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15 – 20, 2019 Long Beach, CA, USA. New York: IEEE Press, 2019: 992 – 1001.
- [32] Cheng G, Zhou P, Han J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54 (12): 7405 – 7415.
- [33] Chen L, Liu C, Chang F, et al. Adaptive multi-level feature fusion and attention-based network for arbitrary-oriented object detection in remote sensing imagery [J]. *Neurocomputing*, 2021, 451: 67 – 80.
- [34] Yang X, Yang J, Yan J, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15 – 20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 8232 – 8241.
- [35] Qian W, Yang X, Peng S, et al. Learning modulated loss for rotated object detection [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35 (3): 2458 – 2466.
- [36] Wang J, Ding J, Guo H, et al. Mask OBB: a semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images [J]. *Remote Sensing*, 2019, 11 (24): 2930.
- [37] Yang X, Yan J. Arbitrary-oriented object detection with circular smooth label [J]. *Lecture Notes in Computer Science*, Cham, Springer, 2020, 12353: 677 – 694.
- [38] Han W, Kuerban A, Yang Y, et al. Multi-vision network for accurate and real-time small object detection in optical remote sensing images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19: 1 – 5.
- [39] Liu E, Zheng Y, Pan B, et al. DCL-Net: augmenting the capability of classification and localization for remote sensing object detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59 (9): 7933 – 7944.
- [40] Li Y, Huang Q, Pei X, et al. Cross-layer attention network for small object detection in remote sensing imagery [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 2148 – 2161.
- [41] Zhang W, Jiao L, Li Y, et al. Laplacian feature pyramid network for object detection in VHR optical remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1 – 14.
- [42] Dong Z, Lin B. BMF-CNN: An object detection method based on multi-scale feature fusion in VHR remote sensing images [J]. *Remote Sensing Letters*, 2020, 11 (3): 215 – 224.
- [43] Mostofa M, Ferdous S N, Riggan B S, et al. Joint-SRVD-Net: Joint super resolution and vehicle detection network [J]. *IEEE Access*, 2020, 8: 82306 – 82319.

- [44] Bashir S, Wang Y. Small object detection in remote sensing images with residual feature aggregation-based super-resolution and object detector network [J]. *Remote Sensing*, 2021, 13(9):1854.
- [45] Yang X, Yan J, Liao W, et al. SCRDet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing [J/OL]. <http://arxiv.org/pdf/2004.13316.pdf>.
- [46] Wang J, Wang Y, Wu Y, et al. FRPNet: A feature-reflowing pyramid network for object detection of remote sensing images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19:1–5.
- [47] Wu Y, Zhao W, Zhang R, et al. AMR-Net: Arbitrary-oriented ship detection using attention module, multi-scale feature fusion and rotation pseudo-label [J]. *IEEE Access*, 2021, 9:68208–68222.
- [48] Yang X, Hou L, Zhou Y, et al. Dense label encoding for boundary discontinuity free rotation detection [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New York: IEEE Press, 2021: 15819–15829.
- [49] Gong H, Mu T, Li Q, et al. Swin-transformer-enabled Yolov5 with attention mechanism for small object detection on satellite images [J]. *Remote Sensing*, 2022, 14(12):2861.
- [50] Ding J, Xue N, Long Y, et al. Learning RoI transformer for oriented object detection in aerial images [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15–20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019:2849–2858.