

光学遥感图像目标检测的深度学习算法研究进展

徐丹青，吴一全

南京航空航天大学 电子信息工程学院, 南京 211106

摘要：光学遥感图像目标检测在军事和民用方面都有很广泛的应用前景。本文综述了基于深度学习的光学遥感图像目标检测算法研究进展。首先，详细分析了光学遥感图像中的目标特点；其次，简要回顾了光学遥感图像目标检测算法的发展历程；再次，阐明了基于深度学习的光学遥感图像目标检测流程；最后，以深度学习为出发点，概述了经典的目标检测框架，并根据光学遥感图像目标的特点，分别针对尺度多样性、方向多样性、形状多样性、小尺寸、特征相似性、背景复杂性、分布密集性、弱特征的光学遥感图像目标检测问题，对各种改进算法进行了系统性的总结，此外，归纳了针对光学遥感图像目标检测的非全监督学习算法，介绍了常用的开源光学遥感图像数据集和目标检测的性能评估指标，指出了现阶段在光学遥感图像目标检测领域仍然面临的主要挑战和下一步的发展方向。

关键词：光学遥感图像，目标检测，深度学习，目标特点，目标检测流程，目标检测框架，非全监督学习，数据集，性能评估指标

中图分类号：TP751.1/TP391.41/P2

引用格式：徐丹青, 吴一全.2024.光学遥感图像目标检测的深度学习算法研究进展.遥感学报,28(12): 3045–3073

Xu D Q and Wu Y Q. 2024. Progress of research on deep learning algorithms for object detection in optical remote sensing images. National Remote Sensing Bulliten, 28(12):3045–3073[DOI:10.11834/jrs.20243166]

1 引言

遥感技术广泛应用于军事侦察、测绘工程、农林管理、气象预报、资源勘探、灾害预防、环境监控、城市规划、海洋监测等领域。按照成像方式不同，遥感图像可分为红外、多光谱、合成孔径雷达SAR（Synthetic Aperture Radar）和可见光图像。其中，红外遥感成像（乐应波等，2022）的工作波段在760—1000 nm，成像设备依靠温度差分辨不同物体，图像对比度较低，目标细节的分辨率不够高。多光谱成像（Zhou等，2009）通常包含若干较窄的波段，多光谱图像的空间分辨率较低且数据处理过程繁杂，设备的成本也较高，此外图像配准困难和图像的几何畸变也极大地限制了多光谱图像的应用范围。SAR成像是指由合成孔径雷达获取图像的技术，通常来说，SAR图像由于斑点噪声导致信噪比较低，视觉效果不佳。

可见光遥感成像的波段一般在380—760 nm，和上述几种遥感图像相比，可见光遥感图像能够更加直观地反映目标的轮廓、颜色和细节特征，因此更加方便人眼的直接观察。此外，可见光的成像方式和设备多样性也较高，能够被广泛使用。严格来说，光学遥感是一个大类，包括可见光、近红外以及短波红外遥感，在本文中特指可见光遥感。随着卫星和遥感技术的快速发展，先进的遥感卫星不断被投入使用，获得大量的高清遥感图像，在硬件上推动了光学遥感目标检测的发展。

早期的光学遥感图像目标检测技术一般基于传统方法，即：首先进行候选区域的提取；然后通过人工的方式设计特征；最后使用分类器输出目标的类别和坐标。在提取候选区域的过程中，往往需要手工设置大量的滑动窗口，极大消耗硬件和人力资源。此外，特征提取主要依赖目标的颜色、轮廓、纹理等视觉信息，有很大的主观性。

收稿日期：2023-05-26；预印本：2023-09-14

基金项目：国家自然科学基金(编号:61573183)

第一作者简介：徐丹青，研究方向为遥感图像目标检测。E-mail: xudanqing@nuaa.edu.cn

通信作者简介：吴一全，研究方向为遥感图像处理与理解、目标检测与识别、视频处理与智能分析、视觉检测与图像测量等。E-mail: nuaaimage@163.com

传统方法也难以有效提取目标特征，在应对复杂的光学遥感图像时鲁棒性不足。

进入到21世纪，深度学习取得突破性发展。深度学习可以通过不断地堆叠极深的网络来获取有用的信息，具有非常强的数据挖掘和分析能力，因此近年来备受研究者关注。深度学习不仅能够有效提取图像中的深层语义信息，而且在鲁棒性和应用范围方面都比传统算法更有优势，因此发展迅猛。由于各种类型图像中的目标特征各异，目标检测所面临的挑战也不尽相同。

光学遥感图像目标有以下3个特点：

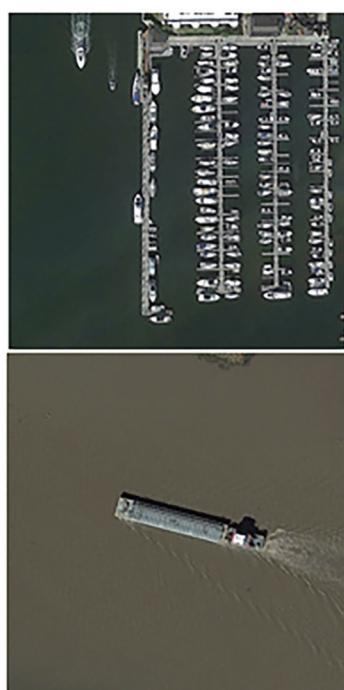
(1) 目标特征的复杂性：遥感设备的成像距离远，因而能够覆盖较大的范围，图像中可能存在尺度和形状变化较大的目标，如船舶和港口等（图1(a)）；遥感目标排列的方向往往不同（图1(b)）；大量的遥感目标具有尺寸小的特点，这些目标通常只有十几个像素，有的甚至只有几个像素（图1(c)）；有些类别目标之间的特征非常相似，如篮球场和网球场（图1(d)）。

(2) 背景复杂性：遥感图像中，背景往往会占据较大的面积，这会导致目标常常“淹没”在复杂的背景中，使得检测器难以有效地分辨出目标（图1(e)），背景的高度复杂性对目标检测算法提出了极高的要求。

(3) 分布密集性：遥感图像中，目标不仅体

现在尺寸小，方向多变等方面，有时候，遥感目标往往分布较为密集（图1(f)），这使得检测器难以将同一类别，不同目标的特征进行分离，这无疑给目标检测带来了巨大的挑战。

上述光学遥感图像中的目标具有有别于常规目标的特点，也正是其检测难点。为此，宋志娜等（2021）、张财广等（2020）对光学遥感图像中的水面船舶目标检测算法进行了重点归纳；祝文韬等（2020）总结了光学遥感图像中的飞机目标检测算法；李晓斌等（2019）主要对光学遥感图像目标检测的流程进行介绍，但缺乏对具体改进算法的总结；聂光涛和黄华等（2021）针对光学遥感图像目标检测中遇到的问题对各种改进算法进行梳理，但是缺乏对算法优缺点的对比；刘天颖等（2020）总结了光学遥感图像中，基于深度学习的目标检测算法，但是缺乏对光学遥感目标特点的分类，另外列举的光学遥感图像数据集也不够完整；廖育荣等（2022）主要整理了针对检测难点的改进算法，但是缺少对算法优缺点的总结和对光学遥感图像数据集的归纳；Li等（2022f）主要从网络的结构方面对各种光学遥感图像目标检测算法进行了介绍，但是缺乏对光学遥感图像目标检测问题的总结；Cheng和Han等（2016）主要对传统算法进行梳理，但是对深度学习算法的涉及较少。



(a) 尺度和形状多样性
(a) Scale and shape diversity



(b) 方向多样性
(b) Directional diversity



(c) 小尺寸
(c) Small size



图1 光学遥感图像中的典型实例
Fig. 1 Typical examples in optical remote sensing images

因此,本文以深度学习为主要出发点,分析和总结了光学遥感图像的各种目标检测算法,具体研究内容如下:梳理光学遥感图像目标检测算法的大体发展历程;针对深度学习算法,阐明光学遥感图像目标检测的大体流程;阐述经典的深度学习目标检测模型并针对光学遥感图像中的目标特点总结研究人员提出的一系列改进算法;介绍常用的光学遥感图像数据集和评价指标并对光学遥感图像目标检测算法进行了性能评估;指出了现有的问题,并对今后的发展方向进行展望并给出结论。

2 光学遥感图像目标检测算法的发展历程

针对光学遥感图像目标检测问题,最早期的研究方向是模板匹配,其算法的研究最早始于20世纪50年代。其实现过程如下:首先人工制成用于目标检测的模板;然后相似度计算用于搜索包含目标的潜在区域并计算其与模板之间的相似度从而实现对目标的定位。基于模板匹配的方法在原理上非常简洁,但是缺点也很明显,人工制作模板使得其效率很低,滑动窗口的计算也有较大的复杂度,此外,这种方法只能处理较简单的目标。

20世纪60年代到80年代,针对光学遥感图像目标检测问题,又先后发展出基于先验知识的方法和基于机器学习的方法。基于先验知识的方法首先利用几何知识对目标的外观信息进行建模,然后通过上下文知识在目标和背景之间创建了一个特殊的空间约束,将目标的隐式知识转换为显式检测规则,通过显式检测规则搜索目标。基于先验知识的方法的关键是先验知识的准确性,这很大程度上取决于人的主观性,因此,其检测具有很大的不稳定性。

机器学习在光学遥感图像目标检测中有较多应用。机器学习本质上是将目标检测看作一个分类问题,算法模型首先搜索图像中可能包含目标的区域,在提取方向梯度直方图HOG(Histogram of Oriented Gradient)特征、词袋BOW(Bag of Word)特征、上下文特征等之后,采用一个独立的分类器进行目标类别的区分并判定区域内是否包含目标。机器学习在分类和特征提取上有其固有缺陷,如难以进行端到端的训练、并且计算成本高、对目标的定位能力不足等,这些缺陷都极大地制约了光学遥感图像目标检测性能的提升。

20世纪90年代以来,深度学习DL(Deep Learning)得到快速发展,深度学习的核心是卷积

神经网络 CNN (Convolutional Neural Network), 从 21 世纪 10 年代开始, 光学遥感图像目标检测逐渐结合深度学习算法并慢慢取代传统算法。图 2 为光学遥感图像目标检测算法发展历程的总结。

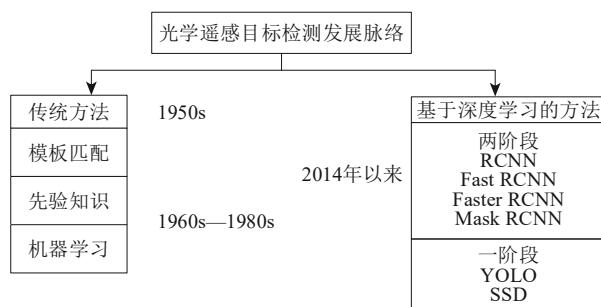


图 2 光学遥感图像目标检测发展历程

Fig. 2 Development of optical remote sensing object detection algorithms

3 基于深度学习的光学遥感图像目标检测流程

对于光学遥感图像目标检测来说, 深度学习的算法流程如下: (1) 数据预处理; (2) 特征提取; (3) 检测; (4) 后处理。遥感图像首先进行预处理以满足模型的输入要求, 然后通过特征提取网络提取目标特征, 再根据提取的特征对目标进行预测, 最后去除多余的检测框得到最终的检测结果。其中, 按照检测方法的不同, 可以分为基于锚框的检测、无锚框的检测和端到端检测, 基于锚框的检测按照检测步骤可以分为一阶段算法和两阶段算法, 无锚框的检测可以分为基于关键点算法和基于中心点算法。整个检测过程如图 3 所示。

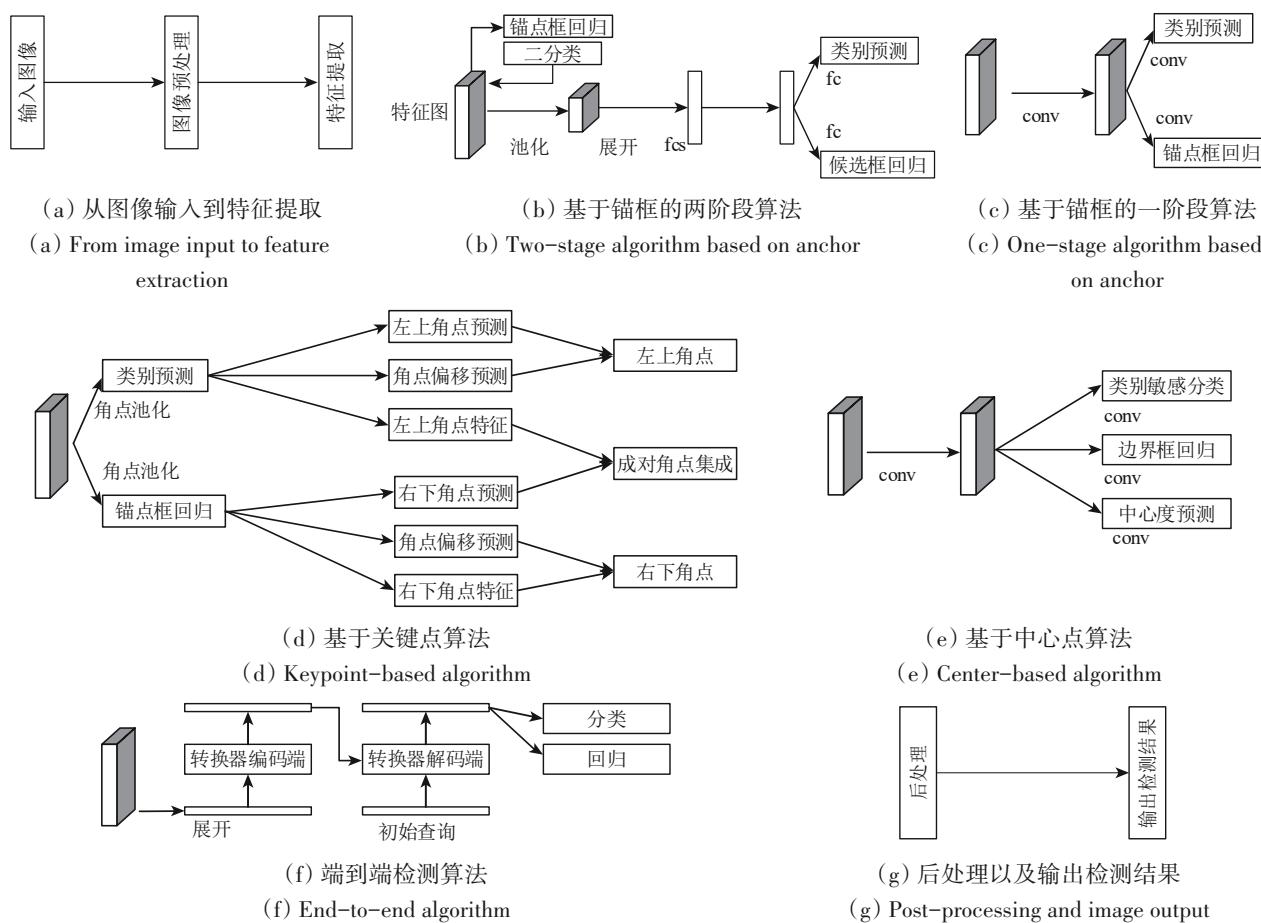


图 3 基于深度学习的光学遥感图像目标检测流程

Fig. 3 The process of optical remote sensing object detection based on DL

3.1 数据预处理

3.1.1 数据增强

和传统算法相比, 深度学习包含大量参数, 这些海量的参数需要足够的数据训练作为支撑。在实际应用中, 数据集往往存在数据量不充分或者不平衡的问题, 在不增加成像设备工作量的基础上, 数据增强是一个有效的办法, 几何变换、颜色变换和模糊变换等是最常用的数据增强手段。

几何变换主要有如下4种方法: (1) 平移变换。平移是一种简单的几何变换, 通过简单的平移, 目标会出现在图像中的不同位置, 从而达到了增加数据量的目的; (2) 旋转变换。旋转变换是将目标进行一定程度的旋转; (3) 翻转变换。翻转变换是让目标以某一坐标轴为基准进行翻转, 旋转变换和翻转变换可以丰富目标的角度信息, 使得网络能够学习到目标在各个方向上的特征; (4) 缩放变换。缩放变换是通过缩放因子对图像进行放大或者缩小, 在一定程度上可以增加目标的尺度多样性。

颜色变换也可以称为HSV (Hue, Saturation, Luminance) 变换, 由于光照、大气等因素的影响, 光学遥感图像的亮度会有所不同, 通过调整图像的色调、饱和度和亮度使得网络对不同环境下光学遥感图像的鲁棒性更强。

模糊变换通常用于降低图像中像素之间的差异性和图像中的噪声信息。常用的方法是高斯模糊, 即利用像素的正态分布形式加权平均达到平滑的效果。

3.1.2 图像裁剪

和常规图像不一样, 光学遥感图像由于成像的范围较广, 因此通常具有较大的尺寸, 现有的模型通常无法直接输入如此大的图像, 因此要将较大的图像裁剪为若干较小的图像。在具体实施的过程中, 需要保留一定程度的重叠区域以防止目标因处在边界区域导致被分割成两部分。在裁剪的过程中, 重叠率过大可较好地保留完整的目标, 也会导致生成过多的子图像, 而重叠率过小会生成较少的子图像, 也会使得目标不能被完整地保留, 因此如何合理地选择重叠率是关键。

3.2 特征提取

特征提取是光学遥感图像目标检测的关键, 在深度学习中, 特征提取网络也被称为骨干网络。AlexNet (Krizhevsky等, 2017) 是深度学习领域的里程碑, 随后研究人员不断提出新的骨干网络, VGGNet (Simonyan和Zisserman, 2015) 采用较小的卷积核和池化核减少了参数量, GoogleNet (Szegedy等, 2015) 采用并行的卷积核提升了网络性能, ResNet (He等, 2016) 利用残差结构原理, 缓解了梯度爆炸和梯度消失现象, DenseNet (Huang等, 2017) 和ResNet相比实现了特征信息的充分利用。此外, 研究人员为YOLO (You Only Look Once) 模型专门设计了DarkNet网络, 也取得了不错的效果。

3.3 检测

按照检测方法的不同, 主流的算法可以分为基于锚框的两阶段算法、基于锚框的一阶段算法、基于关键点算法、基于中心点算法和端到端算法。

3.3.1 基于锚框的检测算法

基于锚框的算法会设置多个不同大小和形状的矩形框, 尽可能地涵盖所有类别的目标, 以约束检测框的大小和形状。其中, 两阶段算法通常生成若干候选框, 然后利用卷积神经网络提取候选框特征, 最后对候选框进行分类和回归。和两阶段算法相比, 一阶段算法则省去了选取候选框的步骤, 直接利用神经网络获取目标的类别和坐标信息。

3.3.2 无锚框的检测算法

基于锚框的检测算法需要事先设置锚框的大小和形状, 因此, 其对于参数较为敏感, 泛用性较差, 基于此缺陷, 无锚框的检测算法应运而生。其中基于关键点的检测算法通常先预测目标的关键点 (如左上点和右下点等), 然后对角点的偏移进行预测以实现对目标的检测。基于中心点的算法通常通过预测目标中心点及其到边界的偏移量实现对目标的检测。

(3) 端到端检测算法: 在目标检测时, 通常会存在多个检测框检测一个目标的情况, 因此需要去除冗余框。而端到端检测算法直接为每个目标预测一个检测框, 最典型的端到端检测算法为

基于 Transformer 的 DETR (detection transformer) 算法 (Carion 等, 2020)。DETR 算法首先设置 K 个查询目标特征, 再将目标特征与转换器编码端输出的特征输入到转换器解码端生成 K 个目标特征, 最后利用预测网络进行分类和回归。总之, 该类算法可以端到端地预测每一个目标的检测框, 而无需进行去冗余之类的后处理过程, 因此, 效率更高。

3.4 后处理

在后处理部分, 冗余的检测框被去除, 通常采用非极大值抑制算法 NMS (Non-Maximum Suppression) 或者其改进算法。

4 基于深度学习的经典目标检测算法及针对光学遥感目标的改进算法

在引言中分析了光学遥感图像中的目标特点, 这一部分将梳理经典的深度学习目标检测框架, 再根据光学遥感图像中的目标特征, 分类综述各种改进算法。

4.1 基于深度学习的经典目标检测算法

2012年, Krizhevsky 基于卷积神经网络设计了 AlexNet 并于 2012 年在 ImageNet 大赛中击败了支持向量机 SVM (Support Vector Machine) 算法。此

后, 深度学习逐渐得到重视并被应用到很多领域, 目标检测是其中之一。基于深度学习的目标检测算法大体上有两种, 即一阶段算法和两阶段算法, 图 4 汇总了当前的经典深度学习目标检测模型。

4.1.1 两阶段检测算法

顾名思义, 两阶段算法由两个部分组成, 首先获取一系列可能包含目标的候选区域, 再通过主干网络对候选区域进行特征提取, 最终输出预测结果。图 5 展示了两阶段算法的检测过程。RCNN 是两阶段算法的代表, 由 Girshick 等 (2014) 提出, RCNN 在目标检测领域取得了较好的效果, 但是, 该算法需要单独训练分类器, 此外, 检测时, 每个输入图像都将分配将近 2000 个候选框, 使得网络的整体运行效率很差, 还有一个问题就是网络需要将图像大小固定以便输入到网络中, 这使得图像由于剪切或者长宽比扭曲导致图像失真或者丢失信息。

为了克服 RCNN 耗时的问题, He 等 (2015) 提出 SPPNet, SPPNet 最大的改进就是摒弃了对每个候选框进行复杂计算的过程, 而是对图像进行整体特征提取, 这种方式极大地提高了效率。此外, SPPNet 通过空间金字塔池提取每一个候选框的特征, 有效地避免了在调整候选框大小时造成的信息丢失问题。



图 4 经典的深度学习目标检测模型汇总

Fig. 4 Summary of classical object detection models based on deep learning

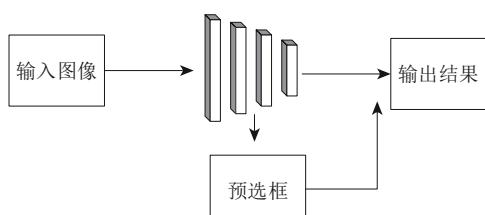


图 5 两阶段算法基本流程

Fig. 5 Basic flow of two-stage algorithms

Girshick (2015) 提出了改进型的 Fast RCNN 算法, Fast RCNN 在 RCNN 的基础上采用了新型的多任务损失函数和 SoftMax 分类器, 此外, Fast RCNN 采用了 ROI Pooling, 因此相比 RCNN, 运行速度有所提升。Faster RCNN (Ren 等, 2017) 改进了候选框选取方法, 区域推荐网络的使用使得效率大大提高。Anchor 机制的使用可以约束检测框的形状和

大小以进一步地提升检测精度。Mask RCNN (He 等, 2017) 第一次实现了检测和分割的双重任务。

Faster RCNN 及其衍生算法在目标检测领域被

广泛使用, 其最大的优势是具有较高的精度, 但是和一阶段检测算法相比, 实时性较差。表1为 RCNN 系列框架的技术总结。

表1 RCNN 系列框架的技术总结

Table 1 Technical summary of RCNN series frameworks

算法	发布年份	技术特点	缺陷	效果
RCNN	2014	采用SS算法提取目标可能区域	生成候选框耗时; 图像尺寸固定	精度在 VOC2010 数据集中达到 53.7%
Fast RCNN	2015	使用特征金字塔网络进行多尺度特征融合	速度精度不够理想	在 VOC2007 和 VOC20012 数据集中分别达到 70.0% 和 68.4%
Faster RCNN	2015	使用区域推荐网络; 锚框机制用于约束检测框长宽比	原始主干网络特征提取能力不足, ROI Pooling 取整操作会造成精度下降	在 VOC2007 和 VOC20012 数据集中分别达到 76.4% 和 73.8%
Mask RCNN	2017	融合了目标检测和图像分割功能	分割不够精确	精度在 COCO 数据集中达到 37.1%

4.1.2 一阶段检测算法

一阶段算法可以直接获取目标的坐标和类别信息, 因此算法结构简洁, 效率较高。图6展示了 一阶段算法的基本流程。

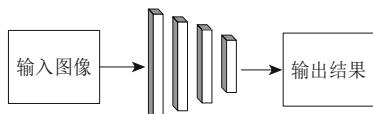


图6 一阶段算法基本流程

Fig. 6 Basic flow of one-stage algorithms

(1) YOLO 系列。Redmon 等 (2016) 提出了能够同时预测多个目标位置信息和对应类别信息的 YOLO V1 算法。YOLO V1 算法摒弃了 RPN 网络, 而是直接获取目标的位置和类别信息, 具体过程如下, 输入图像首先被划分为 $S \times S$ 个网格 (Grid Cell)。每一个网格都会输出若干边界框 (Bounding Box), 其包含的信息 (x, y, w, h, c), 分别为目标的坐标, 宽高和置信度, 因此, YOLO 的最终输出为 $S \times S \times (B \times 5 + C)$, 其中, C 代表类别数, B 代表每个网格输出的边界框的数量。YOLO V1 后续的改进型也基本继承了这一算法流程。YOLO V1 的最大优势在于其快速检测能力, 但是 YOLO V1 的缺陷也很明显: YOLO V1 的主干网络在最后连接有 FCN, 因此, 图像的输入尺寸固定; YOLO V1 的检测数量受限, 且精度较差。

表2为 YOLO 系列算法的技术总结。Redmon 和 Farhadi (2017) 在 YOLO V1 的基础上进行改进, 提出了 YOLO V2 (YOLO 9000) 算法, 在框架上, YOLO V2 的主干网络采用了 DarkNet19 网

络; YOLO V2 引入了锚框的概念, 利用 K-means 算法对数据集中的 Bounding Box 进行聚类分析, 从而得到一系列先验框, 使网络更加容易训练。YOLO V2 在提升精度的同时保持了 YOLO 算法检测速度较快的优点。

YOLO V3 (Redmon 和 Farhadi, 2018) 是最经典的一个版本, YOLO V3 使用残差网络 (ResNet) 原理构造的 DarkNet53 主干网络使得 YOLO V3 能够在极大加深网络的同时有效防止梯度消失和梯度爆炸的现象, 此外, 摒弃了池化层, DarkNet53 采用步长为 2 的卷积进行下采样, 在检测部分, YOLO V3 借鉴了 FPN 的原理对目标进行多尺度预测, 这有利于不同尺寸的目标检测。YOLO V3 不仅仅是一阶段检测算法的里程碑, 也是 YOLO 系列框架最经典的版本, 之后, 研究人员又陆续提出了很多版本的 YOLO 算法, 如 YOLO V4 (Bochkovskiy 等, 2020), YOLO V5 (Yuan 等, 2021), YOLO V6 (Li 等, 2022a), YOLO V7 (Wang 等, 2022a), YOLO V8 和基于 Anchor-Free 机制的 YOLOX (Ge 等, 2021) 等, 但是其框架和基本原理仍然是以 YOLO V3 为基础。

(2) SSD 系列。Liu 等 (2016) 提出了 SSD (Single Shot MultiBox Detector) 算法, SSD 算法采用 VGG16 作为其主干网络, SSD 算法将回归思想和锚框机制相结合, 实现了精度和推理速度的平衡。在 YOLO V1 中, 图像被划分为 $S \times S$ 个网格, 这不利于小目标的检测, SSD 算法去除了 YOLO V1 中的 FCN, 并用卷积层代替, 和 YOLO V3 以及以后的版本一样, SSD 也采用 FPN 进行多尺度目标检测。

表2 YOLO 系列框架的技术总结

Table 2 Technical summary of YOLO series frameworks

算法	发布年份	技术特点	缺陷	效果
YOLO V1	2015	端到端检测, 推理速度快	目标检测数量受限, 精度低	在 VOC2007 和 VOC20012 数据集中分别达到 63.4% 和 57.9%
YOLO V2	2016	DarkNet19 网络, K-means 聚类, 批归一化, 锚框机制	目标定位不准确	在 VOC2007 和 VOC20012 数据集中分别达到 78.6% 和 73.4%
YOLO V3	2018	DarkNet53 网络, FPN 多尺度预测	难以设计超参数, 正负样本不平衡	在 COCO 数据集上的精度达到 55.3%
YOLO V4	2020	CSPDarkNet53+SPP, FPN+PANet, 数据增强	和两阶段算法相比, 检测框定位不准, 召回率相对较低	在 COCO 数据集上的精度达到 65.7%
YOLO V5	2021	CSPDarkNet53(Fcos)+SPP, 推理速度快	检测精度略低	在 COCO 数据集上的精度达到 62.6%
YOLO V6	2022	EfficientRep 主干网络+Rep-PAN, 进一步提升检测效果, 优化检测速度	—	在 COCO 数据集上的精度达到 70.0%
YOLO V7	2022	高效的聚合网络+辅助训练模块, 大大减少了参数量的同时保持较高的精度	—	在 COCO 数据集上的精度达到 69.7%
YOLO V8	2023	InceptionNext 主干网络+基于分层注意力的 FasterViT+轻量级的 CloFormer, 实现精度和速度的提升	—	在 COCO 数据集上的精度达到 80.1%

Fu 等 (2017) 提出了一种改进型的 DSSD (Deconvolution SSD) 算法, 相比较于 SSD, DSSD 的主干部分采用了残差网络 ResNet101, 反卷积层的引入使得网络在卷积的过程中, 特征图在宽高的维度上形成了先宽再窄再到宽的结构, DSSD 提升了多尺度的目标检测能力。

SSD 无法进行多尺度特征融合, 针对这一缺陷, Li 等 (2024) 提出了 FSSD (Feature Fusion SSD)

算法, FSSD 将提取的不同尺度的特征图在同一尺度上进行 Concat 操作, 之后重新构建金字塔模型。这种方法使得网络能够对不同尺度的特征图进行特征融合, 其本质上是 SSD+FPN。

总的来说, SSD 算法是继 YOLO 算法之后又一种典型的一阶段检测算法, SSD 系列算法的技术路线如表 3 所示。

表3 SSD 系列框架的技术总结

Table 3 Technical summary of SSD series frameworks

算法	发布年份	技术特点	缺陷	效果
SSD	2016	锚框机制, 多尺度预测	对小目标检测效果不佳	在 VOC2007 和 VOC20012 数据集中分别达到 79.8% 和 78.5%
DSSD	2017	采用 ResNet101 网络, 反卷积, 提升小目标检测效果	反卷积使得网络无法获取更精细的信息, 更小的感受野可能造成全局信息丢失	在 VOC2007 和 VOC20012 数据集中分别达到 81.5% 和 80.0%
FSSD	2017	FPN, 特征融合	原始的 VGG 主干网络特征提取能力受限	在 VOC2007 和 VOC20012 数据集中分别达到 84.5% 和 82.2%

4.1.3 性能对比

VOC 数据集是测试目标检测模型性能常用的公开数据集, 由欧洲计算机视觉会议 ECCV (European Conference on Computer Vision) 开发, 其中 VOC2007 和 VOC2012 由于其数据量和数据类别足够而被广泛使用, 这两个数据集均包含 20 个类别目标, 其中 VOC2007 包含 9963 张图片, VOC2012 包含 11530 张图片。表 4 总结了部分一阶段和两阶段算法在 VOC (Visual Object Classes) 数

据集上的性能对比。可知, 一阶段算法在检测速度上明显占优势, 尽管两阶段算法如 Faster RCNN 使得检测速度从 Fast RCNN 的 0.5 FPS 提升到了 7 FPS, 但是总的来说, 两阶段检测算法的速度还是无法和一阶段算法相比。此外, 早期的一阶段算法如 YOLO 在精度上比两阶段算法要低很多, 但是随着更加先进的主干网络如 ResNet 和 DarkNet 等的引入, 一阶段算法在精度上能够达到甚至超过两阶段算法。

表4 VOC2007和VOC2012数据集部分深度学习目标检测算法的性能指标

Table 4 Performance of some object detection models based on deep learning on VOC2007 and VOC2012

	算法	主干网络	输入尺寸	FPS	mAP/%	
					VOC2007	VOC2012
两阶段算法	Fast RCNN	VGG16	1000×600	0.5	70.0	68.4
	Faster RCNN	VGG16	1000×600	7	73.2	70.4
	OHEM	VGG16	1000×600	7	74.6	71.9
	HyperNet	VGG16	1000×600	0.88	76.3	71.4
	Faster RCNN	ResNet101	1000×600	2.4	76.4	73.8
	ION	VGG16	1000×600	1.25	76.5	76.4
	MR CNN	VGG16	1000×600	0.03	78.2	73.9
一阶段算法	RFcn	ResNet101	1000×600	9	80.5	77.6
	CoupleNet	ResNet101	1000×600	8.2	82.7	80.4
	YOLO	GoogleNet	448×448	45	63.4	57.9
	RON384	VGG16	348×348	15	75.4	73.0
	SSD321	ResNet101	321×321	11.2	77.1	75.4
	SSD300	VGG16	300×300	46	77.2	75.8
	DSOD300	DS32-12-16-1	300×300	17.4	77.7	76.3
	YOLO V2	DarkNet19	544×544	40	78.6	73.4
	DSSD321	ResNet101	321×321	9.5	78.6	76.3
	SSD512	VGG16	512×512	19	79.8	78.5
	SSD513	ResNet101	513×513	6.8	80.6	79.4
	DSSD513	ResNet101	513×513	5.5	81.5	80.0

目标检测在实际中的应用非常广泛，随着应用的领域越来越广泛，场景越来越复杂，依靠大量人工操作的传统目标检测算法显然越来越无法满足当下的需求。深度学习在目标检测领域得到了快速发展，不仅如此，前文中提到的基于深度学习的经典模型到现在已经被应用到相当多的实际领域。然而，这些算法在公开的常规数据集上能够取得越来越好的效果，但是在面对遥感图像时，往往检测效果并不好，遥感目标的特点和常规目标相比有较大差别，因此，需要针对遥感目标的特征对模型进行相应的改进。

4.2 针对尺度多样性的光学遥感图像目标检测算法

相对于常规目标来说，遥感目标的尺度变化更大，这不仅仅体现在不同类别的目标之间，如：飞机、车辆等类型目标通常较小，而建筑、操场等类型目标通常较大，还体现在同一类型目标也会呈现出尺寸差异较大的现象，为此，当前研究的主流方向是多尺度特征的融合。

针对目标尺度多样性的算法比较见表5。Li等(2022c)提出了一种改进的R(3)Det算法，将特征金字塔网络结构替换为基于搜索架构的特征金字塔网络，使网络能够自适应地学习和选择特征组

合更新，丰富多尺度特征信息；此外，在上下文信息增强模块中加入浅层特征，对小目标语义信息进行补充，从而实现了对水面船舶目标的检测。Zhang等(2022a)提出了一种基于频率注意和特征融合增强的多尺度遥感图像目标检测方法，通过信道注意网络有效地学习多频分量提取的特征，实现各信道多元信息的交互；此外，采用交互式上下采样和跳跃连接设计了独特的特征金字塔网络以充分融合特征。Teng等(2022)提出了一种多尺度感知(MSP)模块，将其提取的全局上下文线索和Clip长短句记忆(Clip-lstm)编码的局部空间上下文相关性结合起来，进一步利用这些丰富的语义特征，设计了自适应锚(SA)模块，以缓解遥感图像目标的尺度变化。Huang等(2019)采用一个多尺度特征表示的特征金字塔结构融合可变形卷积模块来学习目标的多尺度特征。Zhou等(2022)将多尺度结构中的每个卷积层划分为大小相等的子层，通过多尺度子层获得超尺度特征，通过强化卷积层内部的信息传递改善多尺度目标检测效果。Li等(2020)提出了一种细化的特征金字塔网络，通过构造自顶向下的特征图实现多尺度检测。Han等(2021)提出了一种感受野增强模块并结合多尺度检测机制，以实现不同尺寸目标的检测。

表5 针对目标尺度多样性的算法比较

Table 5 Comparison of algorithms for object scale diversity

文献	方法	优点	缺陷
Li等(2022c)	基于搜索架构的特征金字塔网络	对复杂环境适应性较好	NAS FPN结构复杂,增大计算量
Zhang等(2022a)	多尺度特征融合	深层语义信息提取能力优秀	过程繁琐,不易于实现
Teng等(2022)	长短记忆编码融合尺度感知模块	能够有效提取语义信息	网络结构以及部署复杂,且实现过程繁杂
Huang等(2019)	多尺度特征融合+可变形卷积	适用于旋转目标检测	易造成信息丢失
Zhou(2022)	将多尺度卷积层划分为大小相等的子层,通过多尺度子层获得超尺度特征	对尺度变化较大目标有较好的适应性	实时性较差
Li等(2020)	构造细化的特征金字塔网络	适应复杂背景下的小目标检测	需要对输入图像及特征图进行多次 resize,使得网络运行效率较低
Han等(2021)	膨胀卷积融合感受野增强网络	适用于小目标检测	膨胀卷积易造成信息丢失

针对尺度多样性的光学遥感图像目标,主流的方向是提取主干网络中的不同尺度特征并进行多尺度特征融合,以适应不同大小的目标检测,特征金字塔的提出是突破性的,并被广泛应用,在光学遥感图像的多尺度检测中取得了比较好的效果。

4.3 针对方向多样性的光学遥感图像目标检测算法

由于遥感图像获取的角度通常是自上而下,因此,相对于通常目标,遥感目标排列的方向往往更加多样化,通常来说,卷积神经网络提取的特征对目标的方向并不敏感,即目标的方向不同,网络提取的特征具有较大的差异性,通常来说,解决这一问题有两种方法:数据扩增和改进网络。

在数据扩增方面,通常是采用多个角度,对数据集中的目标以及标注框进行旋转,以达到增强数据上角度多样性的目的。Yang等(2018b),He等(2018)和Fu等(2018)对数据集的角度进行一定程度的变换,之后将变换后的数据集和原数据集融合作为训练数据集。Li等(2018b)和Schilling等(2018)单纯对数据集中的图像进行随机角度变换。这种数据扩增的方法对处理角度多样性目标有一定作用,但是作用有限,并不能从根本上解决问题。

在改进网络结构方面,Li等(2018a)提出在传统RPN网络的多尺度、多纵横比锚点的基础上增加额外的多角度锚点方法来处理目标多角度问题。Zhang等(2019b)将多个全连接层特征叠加在一起以提升网络对目标旋转和缩放的鲁棒性。Liu等(2018)设计了基于方向角信息的目标检测网络,并将角度信息融合到边界框回归中,从而精确地拟合了边界框和目标区域。Shi等(2021a)

采用了几何变换模块来处理目标的多角度特征。

遥感目标的方向多样性不仅会影响网络的特征提取,也会对目标位置信息的获取造成困难,像船舶这样长宽比较大的目标,水平框难以实现精确检测。Li等(2018c)采用旋转边界框回归模型实现了对水面船舶目标的定位。Liu等(2017)采用了旋转检测机制实现目标的定位。Yang等(2018a)采用旋转包围盒预测、舰船方向预测和旋转非极大抑制的方法来定位船舶的方向。Cui等(2021)针对船舶目标中心关键点和形态尺寸(长宽和旋转角度)进行建模,并提出了一种正交池化模块以提高中心关键点和形态尺寸的预测精度。

此外,通过优化目标函数也是一个方向。Cheng和Han(2016)通过施加正则化约束用以改进目标函数。Liu等(2021)为了解决目标角度误差的影响,提出了一种横纵比加权角度损失函数,该方法根据物体的长宽比对角度损失进行惩罚,提升了对细长目标的检测精度。

针对方向多样性的光学遥感目标检测,通过数据扩增的方法虽然易于实现,但是效果不明显,最先进的算法倾向于利用改进的网络实现性能的改善,主流的方法是采用旋转检测框,精度提升较为明显,不过对网络的改动较为复杂。此外,部分研究对目标函数进行改进,此种方法较为简洁,不过效果提升不稳定。表6给出了针对这一问题的算法的比较。

4.4 针对形状多样性的光学遥感图像目标检测算法

和常规目标不同的是,光学遥感图像目标在类间的形状差异性较大,例如飞机、油桶、操场、车辆等目标在长宽上差别不大,而像桥梁、道路、舰船等目标在长宽比例上类似于“长条形”,这样

的形状或者长宽比多样性的目标容易因为预设锚框不匹配造成较大的误检或者漏检等问题。目前

主流的方法有3种: 增加锚框数量、改进网络和关键点检测。

表6 针对目标方向多样性的算法比较

Table 6 Comparison of algorithms for object direction diversity

文献	方法	优点	缺陷
Yang等(2018b)	旋转后与原数据融合	易于实现	效果有限
He等(2018)	旋转后与原数据融合	易于实现	效果有限
Fu等(2018)	旋转后与原数据融合	易于实现	效果有限
Li等(2018b)	随机角度旋转	易于实现	效果有限
Schilling等(2018)	随机角度旋转	易于实现	效果有限
Li等(2018a)	RPN+多角度锚点	对多尺度和模糊目标效果好	检测效率较低
Zhang等(2019b)	多尺度特征+全连接层	适用于数据集不足的情况	全连接层的计算量较大, 浪费硬件资源
Shi等(2021a)	融合几何变换模块	实时性好	Darknet53特征提取能力受限
Liu等(2018)	角度信息融合边界框回归	网络结构简洁	精度提升有限
Li等(2018c)	旋转边界框回归	定位精确	适用于舰船检测, 泛用性不佳
Liu等(2017)	基于旋转框检测的网络	定位精确	需要进行大量调参, 实现过程较繁杂
Yang等(2018a)	旋转包围盒预测、舰船方向预测和旋转非极大抑制	能够融合不同层的特征	步骤复杂, 运行效率较低
Cui等(2021)	正交池化模块	精度提升明显	针对船舶目标, 泛用性有待提高
Cheng等(2019)	使用正则化约束改进的函数	易于实现	检测效果提升不明显
Liu等(2022)	使用横纵比加权损失函数	易于实现	对权重较为敏感, 使得性能不稳定

文献(Yang等, 2018a; Azimi等, 2019; Wang等, 2019b; Ding等, 2019)通过采用长宽比更加多样性的锚框并增加锚框的数量来增强网络对目标形状的敏感性。这种方法实现简单, 但是效果有限, 对于长宽比非常大的目标, 方向极小的变化就会引起检测框匹配性能的下降。此外, 加大锚框的数量增加了计算资源的消耗。

除了加大锚框的数量之外, 改进网络是一个可行的方向, 主要通过可变形卷积和注意力机制实现。Xu等(2017)将常规卷积替换为可变形卷积, 以可变形卷积为基础构建网络, 以实现光学遥感图像中不同形状目标的检测。Ren等(2018)在Faster RCNN的基础上使用可变形卷积代替标准卷积构造主干网络最后一层, 并在此基础上对目标进行预测。Li等(2019a)针对目标具有不同横纵比的情况, 引入了自适应多感受野注意力机制。Qiu等(2019)提出了一种端到端的自适应横纵比多尺度网络来适应不同形状的目标检测。可变形卷积的引入并没有使得效果有明显地提升, 而且增加了训练难度。而注意力机制的使用也不能很好地解决锚框不匹配的问题。

基于关键点检测的算法是一种新的目标检测

方式, 和广泛使用的Anchor-Based检测方式相比, 基于关键点的检测算法直接舍弃了锚框和目标真实框匹配这一繁琐的过程, 由于不需要对锚框进行匹配, 因此可以在一定程度上缓解目标形状差异的问题。Lin等(2021)提出了一种基于像素点预测的检测算法, 通过融合分类信息和目标框回归信息对遥感目标进行预测。Wei等(2020)提出了一种用中线来表示有方向目标的算法, 通过对中线进行编码以实现遥感目标的检测。Zhu等(2020)通过两个二维周期向量表示角度信息, 简化了检测框角度的计算, 并且针对目标长宽比较大的问题, 提出了一种新的交并比计算方法以适应不同形状目标的检测。

针对形状多样性的光学遥感目标检测, 通常可以采用增加锚框数量和改进网络实现, 增加锚框数量的方法较为简单, 但是效果有限。在改进网络方面, 可变形卷积和注意力机制的使用是两种可行的方法, 不过精度提升较低, 最先进的方法使用无锚框机制的关键点检测方法, 在提高检测效果的同时, 也大大提升了网络运行的效率。表7给出了针对这一问题的算法的比较。

表7 针对目标形状多样性的算法比较

Table 7 Comparison of algorithms for object shape diversity

文献	方法	优点	缺陷
Yang等(2018a)	增加锚框数量	仅需增加Anchor数量, 实现简单	效果有限, 增加计算资源消耗
Azimi等(2019)	增加锚框数量	仅需增加Anchor数量, 实现简单	效果有限, 增加计算资源消耗
Wang等(2019b)	增加锚框数量	仅需增加Anchor数量, 实现简单	效果有限, 增加计算资源消耗
Ding等(2019)	增加锚框数量	仅需增加Anchor数量, 实现简单	效果有限, 增加计算资源消耗
Xu等(2017)	可变形卷积	无需数据增强, 精度提升明显	训练复杂度高
Ren等(2018)	可变形卷积	实现简单, 网络简洁	无法实现实时检测
Li等(2019a)	注意力机制+感受野扩增	位置信息预测准确	仍然存在漏检问题
Qiu等(2019)	注意力机制+特征金字塔网络	精度较高	在主干网络中添加了多种模块, 使得网络运行效率下降
Lin等(2021)	基于像素点预测+无锚框	对目标形状的适应性好	精度提升有限
Wei等(2020)	以中线表示目标方向+无锚框	易于实现	精度和基于锚框的算法相比无优势
Zhu(2020)	自适应周期嵌入法	步骤简洁, 适应性好	精度提升有限

4.5 针对小尺寸的光学遥感图像目标检测算法

卷积神经网络在不断地下采样的过程中, 可能会导致小尺度目标特征的丢失, 进而造成小目标在一定程度上的漏检。针对这一问题, 主流的研究认为融合浅层特征是一种有效的办法, 此外, 通过特征图尺寸扩增的方法也有一定的研究。

在浅层特征的融合方面, Wang等(2019a)采用多层特征融合的方式来检测如飞机这样的小尺寸目标。Qing等(2021)采用改进型的FPN和PANet模块集成不同层的特征映射, 在多个尺度上组合上下文信息, 提升小目标的检测效果。Wang等(2022a)使用循环特征金字塔网络来加强特征表示, 实现了小目标的高精度检测。Li等(2022d)提出了一种改进型的特征融合模块用以改善浅层特征表示来适应小目标的检测。Fu等(2020)通过自顶向下的路径对浅层特征进行语义表示, 再通过自底向上的路径将上层特征映射与底层信息融合, 以实现小目标检测。Li等(2021)通过双向融合深层和浅层特征以及跳跃连接来获得更丰富的上下文信息以及小目标的非局部关联。

特征图尺寸扩增主要是以反卷积的形式实现。Zhang等(2018)提出了一种改进型的Faster RCNN算法, 通过在原有的主干网络末端添加了一个反卷积层网络以适应小尺寸目标的检测。Shi等(2021b)将反卷积和坐标注意力机制结合在一起, 提高了飞机的检测精度。反卷积的引入在提高了小目标检测效果的同时也带来了参数量的增加, 为了减少计算负载, Zhang等(2022b)设计了一种融合膨胀卷积和残差结构的特征增强网络在挖

掘更深层次语义信息的基础上减少了参数量。Wang等(2021a)同样引入了膨胀卷积, 融合了不同尺度的特征并在增加了感受野的同时控制了学习参数数量。Wang等(2019a)提出了一个感受野增强模块, 通过整合从3个具有不同膨胀率的平行膨胀卷积中获得的特征图来提取上下文信息。膨胀卷积在减少参数量的同时也会在一定程度上造成局部信息丢失的问题, 为此, Ying等(2019)提出了一种多尺度、多注意目标检测方法, 缓解了膨胀卷积的这一缺陷。

不同于多尺度特征融合, 光学遥感图像中的小目标检测更倾向于采用浅层特征融合。此外, 反卷积和膨胀卷积的使用也能够有效地检测光学遥感图像中的小目标。表8给出了针对这一问题的算法的比较。

4.6 针对特征相似的光学遥感图像目标检测算法

在光学遥感图像中, 通常会包含有特征较相似的目标, 这给目标检测带来了一些困难, 针对这一问题, 研究人员主要通过注意力机制和采用辅助网络的方式解决。

Cheng等(2021b)提出了一种端到端的多尺度特征融合模块, 利用特征金字塔融合SE模块获取多级特征映射。Li等(2022b)在DarkNet53网络的基础上添加了一个辅助网络, 通过辅助网络提取浅层语义信息, 并与主干网络提取的深层语义特征融合以增强网络对目标特征的识别能力。目前, 关于这方面的研究还较少。表9给出了针对这一问题的算法的比较。

表8 针对小尺寸目标的算法比较

Table 8 Comparison of algorithms for small objects

文献	方法	优势	缺陷
Wang等(2019a)	多层特征融合	检测速度块	小型网络特征提取能力有限
Qing等(2021)	改进型FPN网络和PANet对骨干网输出的特征进行再处理	不同尺寸特征融合能力强	改进型算法的网络运行效率有较大幅度的下降
Wang等(2022a)	改进型特征金字塔网络	小目标检测率高	运行效率较低
Li等(2022d)	特征融合+特征超分辨率网络	小目标检测精度高	适用于汽车目标,泛用性不足
Fu等(2020)	深层信息和浅层信息融合	适用于多方向目标检测	实时性较差
Li等(2021)	双向融合深层和浅层特征以及跳跃连接来获得更丰富的上下文信息以及小目标的非局部关联	对小目标检测性能强	对数据量的要求较大
Zhang等(2018)	Faster RCNN+反卷积	提升了网络的特征表达能力,大大减少了参数量	反卷积层在主干网络尾部,无法放大过小的目标
Shi(2021b)	反卷积+坐标注意力机制	大大提升了运行效率,整体精度较高	无法检测过小的目标
Zhang(2022b)	融合膨胀卷积和残差结构	感受野扩增同时减少参数	膨胀卷积丢失局部信息
Wang等(2021a)	膨胀卷积	感受野扩增同时减少参数	膨胀卷积丢失局部信息
Wang等(2019a)	感受野增强模块+膨胀卷积	感受野扩增同时减少参数	膨胀卷积丢失局部信息
Ying等(2019)	全局空间关注模块	深层特征位置表达能力强	对密集分布目标的定位能力不足

表9 针对特征相似性的算法比较

Table 9 Comparison of algorithms for complex background

文献	方法	优势	缺陷
Cheng等(2021b)	特征金字塔+注意力机制	端到端快速检测	SE模块需要更多计算资源
Li等(2022b)	引入辅助特征提取网络	易于实现	主干网络特征提取能力不理想

4.7 针对背景复杂性的光学遥感图像目标检测算法

遥感图像由于其覆盖范围广,因而包含了大量各种类别的物体,而标注的待检测目标类别相对较少,导致背景占整个图像的绝大多数面积,遥感目标常常被包围在复杂的背景中,这给目标检测带来了一些困难。为此,研究人员通常会在增强目标特征、减少背景干扰方面进行改进。

Wang等(2022a)提出了一种改进型的级联算法SA-Cascade,利用基于可学习锚的RPN网络对遥感目标进行精确定位。Cheng等(2021a)提出了一种基于卷积神经网络的多样化上下文信息融合框架(DCIFF-CNN),利用结构化的对象级关系来提高复杂背景下的目标检测精度。杨钧智等(2021)提出了一种多尺度圆频滤波器(MSCFF)和卷积神经网络融合的飞机目标自动检测算法,利用多尺度圆频滤波器对图像中的噪声进行滤除,提取不同尺度上的飞机目标候选区域。Zhang等(2022a)针对复杂场景中强耦合的语义关系,设计了一种复合语义特征增强方法生成有效的语义描述来增强以像素为单位的目标中心点解释。Ye

等(2020)为了抑制背景杂波的干扰,通过分层的形式融合不同层的特征,再将融合后的特征通过特征过滤模块投影到低维子空间中,以获取更相关、更精确的上下文信息。Lei等(2020)通过提取显著图并利用所提出的模型来指导训练,并引入多层融合策略连接不同的卷积层,以有效增强目标区域同时抑制背景噪声。Bai等(2023)融合了全局上下文感知和局部上下文感知模型,构建了全局—局部—全局上下文感知网络。

此外,一些研究认为注意力机制能够有效检测一些背景干扰严重的目标。Wang等(2019a)采用语义注意网络提供语义特征,并利用初始横向连接网络对目标进行特征增强,以有效地从杂乱的背景中识别出目标。兰旭婷等(2021)提出了一种基于ResNet50的SSD300网络,引入了注意力机制CBAM模块和FPN模块以有效检测复杂背景下的飞机目标。Wang等(2022c)用MobileNet作为主干网络提取遥感图像的多层次特征,采用多尺度特征集中注意力模块MFCA抑制背景噪声,增强有效特征重用。Cheng和Zhang(2022)提出了一

种集成上采样特征增强模块和注意力机制的遥感图像目标检测网络 (C-2-YOLO)，设计了一种新的特征融合模块，然后集成了坐标注意模块，对复杂场景中的目标进行检测。Wang 等 (2021a) 针对光学遥感图像中的背景干扰，基于注意力机制原理提出了一种背景滤波模块。Chen 等 (2021) 提出了一种由自注意模块和空间注意模块组成的

级联注意网络 (CA-Net) 以抑制背景噪声。

面对复杂背景下的光学遥感图像目标检测，主流的方法考虑增强目标特征和减少背景干扰。通常利用在网络中添加特征增强模块或者采用注意力机制的方法实现，此外，利用显著性约束和频域滤波器的方法也有一定研究。采用表 10 给出了针对这一问题的算法的比较。

表 10 针对复杂背景的算法比较

Table 10 Comparison of algorithms for complex background

文献	方法	优点	缺陷
Wang 等(2022a)	特征自适应模块+特征融合模块	对小目标检测效果好	RPN 网络导致推理速度慢
Cheng 等(2021a)	多尺度网络+局部上下文 RPN	目标检测识别精度较高	实时性较差
杨钧智等(2021)	常规方法频率滤波器融合深度学习	定位准确	针对飞机目标,泛用性不足
Zhang 等(2022a)	复合语义特征增强	定位精确	网络较复杂,网络实时性较差
Wang 等(2019a)	特征增强型金字塔网络	易于实现,鲁棒性好	无法实时检测
Lei 等(2020)	显著性约束和多层次融合策略	背景抑制效果好	训练复杂,实时性不足
兰旭婷等(2021)	CBAM 模块+FPN 模块	查全率较高	针对飞机目标,泛用性不足
Wang 等(2022c)	注意力机制融合特征金字塔网络	对多类型目标适应性强	MobileNet 特征提取能力不足
Cheng 和 Zhang (2022)	改进损失函数	对小目标的检测效果较好	整体的精度提升有限
Wang 等(2021a)	多分支特征提取模块	有效抑制背景干扰	网络对注意力机制模块添加的位置较为敏感
Ye 等(2020)	通过分层的形式融合多尺度特征	适用于多尺度目标检测	—
Chen 等(2021)	自注意机制和空间注意机制	适用于多方向多尺度检测	实时性不足

4.8 针对分布密集性的光学遥感图像目标检测算法

在光学遥感图像中，一些目标如汽车、飞机、舰船等常常呈现密集分布的特点，使得网络难以将同一类的不同目标进行区分，这也给检测增加了一些困难。目前的主流方法主要是通过增强目标特征和增强目标位置信息精度的方式解决。

在增强目标特征方面，Zhou 等 (2022) 在 CSPDarkNet53 的基础上增加了一些卷积层，并且提出了一种多尺度跨层检测器以检测密集分布目标。Hou 等 (2022) 通过在网络中融合 CBAM 注意力模块来改善密集目标的检测效果。Wang 等 (2021b) 设计了一种多分支特征提取 (MBFE) 模块，以更好地捕获密集分布目标中的特征信息。Long 等 (2019) 结合了基于图像模型的结构和新设计的特征融合深度网络来处理小尺寸、密集分布的目标。Yang 等 (2019) 结合监督像素注意网络和通道注意网络，在抑制背景干扰的同时突出

目标特征，实现了小而密集的目标检测。

在增强目标位置信息精度方面，Cheng 和 Zhang (2022) 通过在损失函数中加入角度损失来提升密集目标的旋转角度精度。Long 等 (2017) 提出了一种基于分数的边界框回归算法，并结合 NMS 算法来优化目标区域的边界框以实现密集分布目标的检测。Yang 等 (2021) 采用新的编码机制：设计两个密集编码标签取代现有基于分类的检测器中的稀疏编码标签进行角度分类，以对密集目标进行精确定位。Gao 等 (2023) 设计了一种定向回归损失函数来增强目标的定位。

密集分布的目标给检测任务带来了新的困难，在改进网络方面，主流的方法采用添加特征增强模块的方法实现，在检测性能上提升较大，不过对网络过多的改动也提升了网络训练的复杂度。通过改进损失函数来增强目标位置信息也是一种简单可行的方法，不过和改进网络结构相比，整体表现不佳。表 11 给出了针对这一问题的算法的比较。

表 11 针对目标密集分布的算法比较
Table 11 Comparison of algorithms for densely distributed objects

文献	方法	优点	缺陷
Zhou 等(2022)	CSPDarkNet53+多尺度跨层检测	检测精度较高	召回率较低
Hou 等(2022)	采用 CBAM 注意力机制	特征提取能力较强	小目标检测能力不足
Cheng 和 Zhang(2022)	在损失函数中加入角度损失	易于实现	检测效果提升有限
Wang 等(2021b)	多分支特征提取模块	优化遮挡、阴影干扰严重的目标检测	网络结构较复杂, 实现困难
Long 等(2019)	基于图形模型和深度特征融合网络	抗背景干扰能力较强	实现过程繁琐
Yang 等(2019)	通道注意网络	能够有效抑制背景干扰	步骤繁琐, 实时性不佳
Long 等(2017)	边界框回归算法结合 NMS 算法	定位精度高	采用 RPN 网络, 无法实时检测
Yang 等(2021)	设计新的密集编码标签	定位精度高	—
Gao(2023)	定向回归损失函数	易于实现, 定位精度高	—

4.9 针对弱特征的光学遥感图像目标检测算法

弱特征的光学遥感图像目标并不是指目标的尺寸小, 而是在成像的过程中, 光学遥感图像受到雨、雾、雪或者成像设备分辨率等因素的影响, 造成目标部分被遮挡或者目标模糊的问题, 这对网络提取的特征质量造成不小负面影响。针对这一问题, 目前的研究方向集中在特征强化方向。

Wang 等 (2022d) 在 RetinaNet 的基础上提出

了一种单阶段特征增强网络用以检测弱特征的光学遥感图像目标。史文旭等 (2020) 改进 SSD 模型, 通过在主干网络和检测层中添加浅层和深层特征增强模块来检测光学遥感弱目标。徐志京和柏雪 (2022) 提出了一种基于双重特征增强的目标检测网络, 利用改进型金字塔网络对目标进行感受野增强。表 12 给出了针对这一问题的算法的比较。

表 12 针对弱特征目标的算法比较
Table 12 Comparison of algorithms for weak objects

文献	方法	优点	缺陷
Wang 等(2022d)	RetinaNet+特征增强模块	能够较好地适应小目标检测	检测过程生成大量无用的 anchor, 浪费计算资源
史文旭等(2020)	SSD+特征增强模块	网络结构简洁, 实时性好	无法实现多尺度融合
徐志京和柏雪(2022)	改进型特征金字塔网络	适用于小目标检测	双重网络使得其无法进行实时性检测

现有方法大多通过特征增强的方式对检测性能进行改善, 在一定程度上可以缓解弱特征目标检测难的问题, 但是要从根本上解决此问题还是要从硬件上着手, 提升图像的成像质量。

4.10 针对非全监督形式的光学遥感图像目标检测算法

在深度学习算法中, 针对光学遥感目标检测, 主流的方式为全监督学习, 即目标的标注具有明确的坐标和类别信息。前面的小节针对光学遥感图像目标检测所遇到的问题总结的方法均为全监督学习。但是面对海量的图像和目标, 人工进行精确的标记本身就是一种费时费力的工作, 基于此问题, 研究人员开发出非全监督学习的方法来简化标注的工作量。非全监督学习的形式大体上有半监督和弱监督两种。半监督学习和弱监督学

习的示意图如图 7 所示。其中半监督学习指训练集只包含一部分精确标注的目标, 如图 7 (a) 所示, 图中只有部分油桶目标被精确标注, 而另外一部分油桶未被标注; 而弱监督学习表示训练集只能提供粗略的标注信息, 如图 7 (b) 所示, 图中的油桶目标只包含有标签, 而没有标注框。

4.10.1 基于半监督的光学遥感图像目标检测算法

基于全监督学习的光学遥感图像目标检测算法对于数据集的标注有较大的依赖, 这种方式无疑大大增加了时间成本和工作量, 半监督学习能够大大缓解这一问题。

陈国炜等 (2020) 结合传统的卷积神经网络和 GAN, 通过生成器生成假样本, 判别器生成真样本并学习类别信息以实现数据集中标记的区分。Zhong 等 (2020) 提出了由实例细化循环、离线学

习和在线表示组成的循环数据驱动框架，通过网络预测，选择得分较高的 Bounding Box 进行主动学习和标签更新。Wu 等 (2021) 构造了一个半监督标签生成模块，通过逐步提高伪标签的阈值，由

易到难训练检测器，以降低网络对标注标签的依赖。此外，Cheng 等 (2021b) 和 Li 等 (2021) 采用小样本学习，通过从数据中大量的已知类别中学习元知识的方式来学习样本中的未知类别。

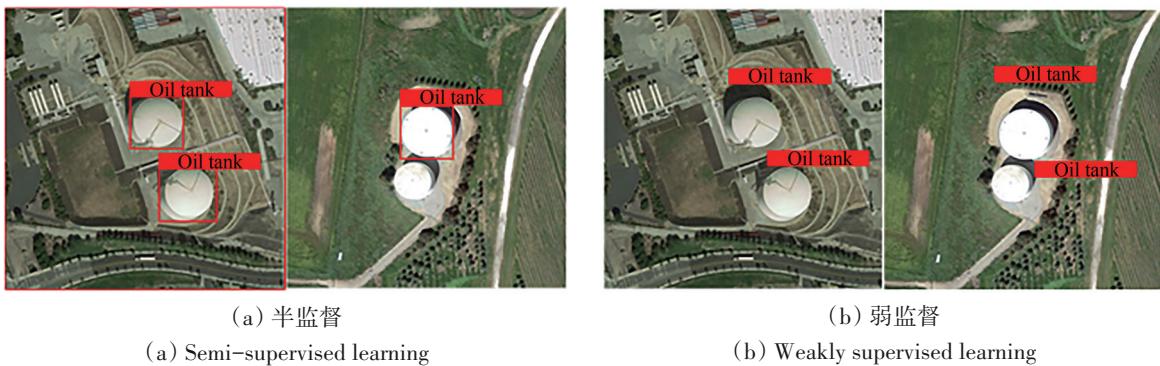


图7 半监督学习和弱监督学习的对比示意图

Fig. 7 The comparison diagram of semi-supervised learning and weakly supervised learning

大多数光学遥感目标检测算法追求高性能，半监督学习由于标签数量较少，其检测精度和全监督相比仍然存在一定差距，半监督学习的优势主要体现在其对数据标注的依赖较少，因此，半监督学习在数据方面的优势使其作为一个小分支被研究人员持续关注。

4.10.2 基于弱监督的光学遥感图像目标检测算法

目标边界的精确标注是一项非常耗费时间的工作，而弱监督学习只需要提供粗粒度标签，减轻了标注的困难。基于弱监督学习的光学遥感图像目标检测大致可以分为3类：基于多实例学习的方法、基于分割的方法和基于其他方法。

基于多实例学习的方法将图像视为一个包，将图像中的目标视为一个实例，将包视为未标记实例的集合。网络通过接受一组标记的包来训练和预测包中的未标记实例。Bilen 和 Vedaldi (2016) 第一次提出了基于多实例学习的弱监督学习算法。随后，Tang 等 (2017) 和 Wan 等 (2019) 采用伪标签机制进行监督回归，该机制在后续研究中被不断改进并被应用在遥感目标图像检测中：针对检测模型可能将多个相邻目标判定为一个目标的缺陷，Feng 等 (2021) 设计了全局上下文感知增强模块来激活整个目标的特征，并进一步开发了双局部上下文残差模块，利用局部上下文的语义差异捕获图像中的目标。Yao 等 (2021) 通过对检测器由易到难进行训练，将动态课程学习的思想引入遥

目标检测中。Wang 等 (2021a) 采用了伪标签生成算法，利用目标的定位信息生成目标的类别标签，设计了一种有效的自适应阈值选择策略，以消除不同类别和形状差异对计算样本伪标签的影响。

基于分割的方法首先生成目标激活图，并通过提取图上的区域来检测目标。Li 等 (2018b) 利用场景的像素级标签来学习特定类别的激活权值，并设计了一种多尺度场景投票策略来计算特定类别对象的激活图。Wu 等 (2020) 基于 AlexNet 通过反向加权生成热图来定位对象。

除了基于多实例学习的方法、基于分割的方法之外，研究人员还提出了其他的一些基于弱监督的光学遥感图像目标检测算法。Yang 等 (2022) 设计了一种边界感知模块，通过边界标签生成模块生成的高置信度目标标签显式监督使模型能够学习目标结构的特征，从而提升目标边界定位精度。Chen 等 (2022a) 设计了一种结构感知扩展模块，通过有效利用结构特征恢复目标结构，并提出了一种边缘结构感知损失函数限制恢复结构的范围。Wu 等 (2022) 采用深度双边滤波，通过附近特征值的非线性组合来过滤深度特征，并且通过扩大边缘两侧特征之间的距离来区分边界，从而很好地保留了边界信息。Wang 等 (2022c) 提出了基于空间图的投票机制，通过构建基于外观图的实例挖掘模型，根据外观相似性传播标签信息来挖掘具有相同类别的所有可能的实例。

虽然弱监督学习算法已经得到了广泛的发展，

但没有精确边界框标注,使得弱监督学习算法的性能远远落后于全监督学习算法。此外,弱监督学习算法没有边界回归过程,导致检测结果中存在目标定位不准确的问题。

总的来说,非全监督学习的算法能够大大减轻标注的工作量,不过检测精度较低一直是一个待解决的问题。表13给出了非全监督算法的比较。

表13 针对非全监督的算法比较

Table 13 Comparison of non-strongly supervised learning-based algorithms

文献	方法	优点	缺陷
陈国炜等 (2020)	传统卷积神经网络+GAN	融合了传统CNN和GAN使得其精度相对较高	训练不稳定,易造成梯度消失的问题
半监督 学习算法	Cheng等(2021) 原型引导RPN网络+Faster RCNN	和其他的半监督算法相比,精度较高	实时性较差
	Li等(2021) 小样本学习+多尺度检测	对不同尺寸的目标检测性能较好	对参数设置的依赖较大
	Zhong等(2020) GAN+循环数据驱动框架	缓解在大尺度遥感图像中标注训练样例非常耗时的缺点	精度较低
弱监督 学习算法	Feng等(2021) 全局上下文感知增强模块融合双局部上下文残差模块	对密集分布的目标具有较好的检测效果	网络结构较复杂,实现较困难
	Yao等(2021) 伪标签机制+动态学习	精度在半监督学习算法中较高	训练过程繁琐复杂
	Wang等(2021a) 伪标签生成算法+自适应阈值选择策略	可较好地适应不同形状的目标检测	自适应阈值算法选择阈值不稳定
	Li等(2018b) 通过多尺度场景投票策略来计算特定类别对象的激活图	精度较高	
	Wu等(2020) 通过反向加权生成热图来定位目标	网络结构简洁,实时性较好	AlexNet的特征提取能力有限
	Yang等(2022) 边界感知模块+边界标签生成模块	能够有效抑制背景干扰	实时性较差
	Chen等(2022a) 结构感知扩展模块+改进型边缘损失函数	网络结构简洁,实用性好	和优秀的全监督算法相比检测性能不够高
	Wu等(2022) 采用深度双边滤波并扩增边关特征距离	减少目标边缘信息的损失	
	Wang等(2022c) 基于空间图的投票机制构建基于外观图的实例挖掘模型	能够进行多尺度目标检测	精度较低

4.11 其他先进算法

4.1节—4.9节着重于发现光学遥感图像目标检测过程中遇到的困难并系统性地总结相应的算法,4.10节针对非全监督学习,总结了相关算法在光学遥感图像目标检测中的应用。近年来,其他的一些先进方法,如:Transformer、迁移学习、知识图谱和先验知识也被研究人员引入到计算机视觉领域,本节将重点汇总近年来上述的一些先进算法在光学遥感图像目标检测中的应用。

4.11.1 基于Transformer的光学遥感图像目标检测算法

Transformer首先被用于序列转导任务,采用基于自注意的多编码器—解码器结构,目前已成为自然语言处理中的主导模型。编码器结构主要由多头自注意层和前馈神经网络层组成。前者关注子空间在不同位置的特征表示,后者对Q, K, V空间的多组特征进行滤波,并将其送入解码器。

Transformer模型庞大复杂的特点增加了训练资源的消耗,但多个编码器和解码器可以并行处理,充分利用计算资源,另外该模型可以在词向量之间建立全局关系,甚至是那些特别遥远的词向量,因此,基于Transformer的算法具有很大的发展潜力。目前,Transformer被扩展到计算机视觉的各个领域,DETR(detection transformer)算法(Carion等,2020)首次将Transformer引入到目标检测中。

在光学遥感图像目标检测任务中,Zheng等(2021)将Transformer融合到一个轻量级FPN中,通过连接不同层次的特征来增强特征映射的语义信息。Xu等(2021)改进了Swin Transformer网络并将其应用到光学遥感目标检测中,可以有效地获取目标的局部感知,提升小目标的检测性能。针对当前被广泛关注的YOLO模型,Zhu等(2021)将Transformer结构结合YOLO V5模型,以更好地捕获全局信息和丰富的上下文信息。编码器结构

也被插入到预测头中，以减少昂贵的计算和存储成本。针对旋转目标检测，Ma等（2021）首次采用Transformer，并在位置预测头中增加了方向维度，此外，该算法还创造性地将Transformer的自关注结构分解为深度可分离卷积，极大地减少了原有的计算负担。

目前，Transformer在目标检测任务中被大量

使用。然而，在遥感领域，该算法的研究还只是起步阶段。遥感图像的复杂性使得Transformer的性能不太令人满意。此外，该模型在面对小目标时的性能较差，而在遥感图像中，这个问题变得更加突出，因此，将Transformer应用到光学遥感图像目标检测领域仍然是一个难点。表14给出了基于Transformer的算法的比较。

表14 基于Transformer的光学遥感图像目标检测算法比较

Table 14 Comparison of object detection algorithms in optical remote sensing images based on Transformer

文献	方法	优点	缺陷
Zheng(2021)	Transformer融合轻量级FPN	可丰富不同层次的语义信息，	轻量级主干网络的特征提取能力不足
Xu等(2021)	局部感知增强模块+Swin Transformer	小目标检测精度相对较高	网络结构较复杂，不易实现
Zhu等(2021)	Transformer融合YOLO V5	可较好地捕获全局信息和丰富的上下文信息	以YOLO V5为基本框架，检测精度相对于其他先进算法无优势
Ma等(2021)	Transformer+旋转框检测+可分离卷积	大大减少了计算负载	和其他基于Transformer的算法相比，精度较低

4.11.2 基于迁移学习的光学遥感图像目标检测算法

在深度学习中，迁移学习是指将在一个领域或任务中学习到的信息应用到另一个相关领域解决相应问题的策略。源领域和迁移领域往往具有共同性，这使得在源领域训练的网络在迁移到目标领域时会产生促进效应，从而迁移学习成为遥感目标检测领域的有效工具。迁移学习可以根据目标域的数据分布对模型进行迁移和微调，大大减少了从头开始训练模型的时间。此外，迁移学习还可以减少目标检测对样本的需求。

迁移学习在面对小样本时非常有效。Dong等（2019）为了解决小目标检测的问题，将训练好的检测器转移到遥感图像，实现了对小目标的自动

标注。Li等（2019b）利用CNN传递模型，解决了遥感数据不足导致的过拟合问题。Chen等（2022b）将迁移学习策略和Faster RCNN网络结合，设计并选择了两种域自适应结构作为域间相似性度量准则。采用对抗性训练来缓解领域转移问题。此外，迁移学习还可以提升网络的训练效率。Zhong等（2018）和Chen等（2019）使用预训练的网络来加速模型的训练过程。

总的来说，迁移学习在光学遥感图像目标检测领域有广泛的应用，该策略不仅可以大大提升网络的训练速度，在面对遥感图像中部分类别目标样本不足时，其也能大大改善训练和检测效果。表15给出了基于迁移学习的算法的比较。

表15 基于迁移学习的光学遥感图像目标检测算法对比

Table 15 Comparison of object detection algorithms in optical remote sensing images based on transfer learning

文献	方法	优点	缺陷
Dong等(2019)	预训练+自动标注算法	能够较好地适应小目标检测	目标自动标注的精度不够高
Li等(2019b)	采用卷积神经网络传递模型	可有效防止过拟合	精度相对其他先进算法不高
Chen等(2022b)	Faster RCNN+迁移学习	精度较高	实时性较差
Zhong等(2018)	坐标敏感平衡算法	模型训练收敛速度较快	算法实现过程较复杂
Chen等(2019)	构建解码器网络预测+压缩迁移学习	模型训练收敛速度较快	网络结构较复杂

4.11.3 基于知识图谱的光学遥感图像目标检测算法

知识图谱是一种知识库，它可以展示真实世界中不同物体之间相互联系的结构。知识图谱可以通过自底向上的形式，提取已知信息中的相关概念，通过选取部分信息并添加到相关知识库中，

进而构造不同概念之间的关系连接。近年来，知识图谱被应用到以目标检测为代表的计算机视觉任务中，其可以通过图像中的背景信息辅助目标检测。从信息论的角度看，额外信息的引入可以使目标检测效果更加优越。而在遥感目标检测中，基于知识图谱的算法也被广泛研究。

针对大多数基于深度学习的光学遥感目标检测算法忽略了上下文信息的缺陷, Tian等(2020)通过建立一个知识库来揭示不同类别目标之间的关系,并利用目标和目标之间的关系来探索上下文信息。Zhang等(2023)通过构建知识库对每个区域进行相应的嵌入编码,并在区域之间构建图结构来传播和学习关键的空间和语义关系,通过关注不同关系的影响程度,提高目标特征表示的

可靠性和鲁棒性。

目前,基于知识图谱的算法在图像分割和目标检测领域的应用非常广泛,在遥感目标检测领域,其研究主要集中在SAR遥感图像以及红外遥感图像的目标检测,对光学遥感图像目标检测方面的研究较少。未来,基于知识图谱的光学遥感图像目标检测算法急需被重点研究。表16给出了基于知识图谱的算法的比较。

表16 基于知识图谱的光学遥感图像目标检测算法对比

Table 16 Comparison of object detection algorithms in optical remote sensing images based on knowledge graph

文献	方法	优点	缺陷
Tian等(2020)	知识库+目标上下文信息	检测精度较高	适用于场景较简单的目标检测,泛用性不足
Zhang等(2023)	构建知识库进行编码	精度较高,鲁棒性强	实现过程较复杂

4.11.4 基于先验知识的光学遥感图像目标检测算法

先验知识源于人类通过积累以往的经验来快速获取的知识。在光学遥感目标检测任务中,先验知识的应用有着很重要的意义,通过引入先验知识,判定目标可能所处的位置和可能存在目标的区域,并结合上下文,进一步推进网络的训练和检测。

Wei等(2023)通过设计3个模块将空间先验知识融入到检测过程,为了加快收敛速度和进一步提高检测精度,提出了一种端到端的检测器从模板中提取先验知识以更好地收敛,并以并行的方式丰富模板特征,实现了对遥感目标的高精度

检测。为了降低对手工设计先验知识的依赖,Lin等(2023)将建筑边缘作为先验知识,通过解耦建筑体与边缘来学习建筑边缘特征,并结合多目标损失函数进一步优化网络,增强建筑边缘特征提取的针对性,从而有效提高了对建筑物提取的精度。Tang和Dong等(2022)提出了一种新型的Bounding Box转换模块,该模块依赖于桥梁的相关先验知识,并结合空间注意力机制和可变形卷积,有效提升了水面上桥梁目标检测精度。

先验知识在光学遥感图像目标检测中有很重要的应用,通过引入目标或者场景的先验知识,可以有效加快网络的收敛速度,提升目标检测性能。表17给出了基于知识图谱的算法的比较。

表17 基于先验知识的光学遥感图像目标检测算法对比

Table 17 Comparison of object detection algorithms in optical remote sensing images based on prior knowledge

文献	方法	优点	缺陷
Wei等(2023)	检测过程融入空间先验知识	检测精度较高	网络结构较复杂
Lin等(2023)	解耦目标与边缘来学习目标特征	鲁棒性强	仅仅针对建筑物目标的检测,泛用性不足
Tang和Dong(2022)	空间注意力机制+可变形卷积	对桥梁目标的检测精度较好	适用于大目标的检测,泛用性不足

4.12 小结

本节对基于深度学习的光学遥感目标检测所遇到的困难和挑战归纳总结了相对应的解决方案,对于这些解决方案,需要注意的是:在光学遥感图像中,目标检测问题并不是孤立和单一的,有些问题往往是共存的,如:光学遥感小目标有时候会处在复杂的环境中,有时候小目标也会呈现密集分布的特点等。其次,针对各种问题的解决

方案有时候往往是相通的,如针对多尺度目标检测所提出的算法往往也能够同时处理小尺寸目标的检测;注意力机制的引入也能够解决遥感目标检测所遇到的很多问题等。

同时,光学遥感目标图像检测的难点在于其目标和背景的复杂性。目标的复杂性一方面在于其特征的多样性,如:尺度、方向、形状的多样性;另一方面在于其特征难以检测,如:尺寸过小和特征相似。而背景的复杂性又给检测器的特

征提取造成了不小的干扰。虽然解决各种问题的方法有其共同性，但是也有其独特性，如：多尺度目标检测更加倾向于采用多尺度特征融合，而小尺寸目标更倾向于采用浅层特征融合；方向多样性的目标和密集分布的目标有其共同性，但是前者多采用旋转检测机制，而后者在采用旋转检测机之外，需要采用特征增强模块对不同的目标特征进行分离。

因此，在探究光学遥感图像目标检测问题时，不应把各种问题单独考虑，也不应把各种方法割裂开，解决各种问题和采用各种改进方案应该是可以相互借鉴的一个整体。但是，由于目标以及光学遥感图像特征的多样性，在面对不同的问题

时，又需要针对性地提出不同的解决方案。

5 常用的光学遥感数据集及目标检测性能评估

近年来，随着高性能的遥感卫星相继投入使用，研究人员获取了大量高质量的光学遥感图像，各研究机构纷纷发布了大量的公开数据集，本节将介绍几种典型的光学遥感数据集以及目标检测领域常用性能评价指标并比较各种改进型算法的性能。

5.1 常用光学遥感图像数据集

表18列出了当前常用的光学遥感图像数据集，包含图像数量、实例数、标注方式、目标类别数和下载地址。

表18 常用光学遥感数据集参数

Table 18 Parameters of commonly used optical remote sensing datasets

数据集	图像数量	实例数	标注方式	目标类别	下载地址
NWPU VHR-10	800	3651	水平标注	10	http://jiong.tea.ac.cn/people/JunweiHan/NWPUVHR10dataset.html
DOTA	2806	188282	水平标注,旋转标注	15	https://captainwhu.github.io/DOAI2019/dataset.html
DIOR	23463	192472	水平标注,旋转标注	20	https://drive.google.com/drive/folders/1UdlgHk49iu6WpcJ5467iT-UqNPpx_CC
UCAS-AOD	910	6029	水平标注,旋转标注	2	http://www.ucassdl.cn/resource.asp
HRSC2016	1070	2917	水平标注,旋转标注	1	https://sites.google.com/site/hrsc2016/
RSOD	976	6950	水平标注	4	https://github.com/RSIA-LIESMARS-WHU/RSOD-Dataset
LEVIR	3791	11028	水平标注	3	https://justchenhao.github.io/LEVIR/
HRRSD	26722	55740	水平标注	13	https://aistudio.baidu.com/aistudio/datasetdetail/53186
AI-TOD	28036	700621	水平标注	8	http://m6z.cn/5MjlYk
VEDAI	1210	—	水平标注	9	https://downloads.greyc.fr/vedai/
TAS	30	1319	水平标注	1	http://ai.stanford.edu/~gaheitz/Research/TAS/
ITCVD	173	29088	旋转标注	1	https://eostore.itc.utwente.nl:5001/fsdownload/zZYfgbB2X/ITCVD
COWC	12729	32716	旋转标注	1	https://gdo152.llnl.gov/cowc/

5.2 目标检测的评价指标

目前，在目标检测领域，公认的评价指标主要如表19所示。

预测结果按照预测值和实际值可被分为4类，其混淆矩阵如表20所示。在表19中，交并比IOU (Intersection Of Union) 被用于衡量检测框 (proposal) 和标注框 (ground truth) 交集和并集之比；TP+FN代表所有框的总数，而TP+FP代表检测框的总数。单独考虑Precision和Recall都太片面，F-measure 将两种评判标准结合起来，F-measure 计算公式中的 β 为参数，当 $\beta = 1$ 时，即为最常用的F1-measure；AP表示召回率在0—1的平均准确

率，具体来说，AP为根据每一个召回率和其对应的最大的精确度，以Recall为纵轴，Precision为横轴所绘制曲线之下的面积，而平均精度值表示所有C个类别的AP的平均值。

5.3 光学遥感目标检测性能评估

为了比较各个算法在光学遥感图像目标检测方面的表现，本节选用了NWPU-VHR10和DOTA数据集，其中，NWPU-VHR10数据集的标注的类别较少，因此挑战性较小，是早期深度学习算法研究的常用数据集。DOTA数据集类别更多，场景更加复杂，因此对于目标检测算法来说更加具有

挑战性。由于传统方法常常针对特定目标进行处理,泛用性以及对场景的适应性较差,因此不在本章算法的评估范围内。对于针对特定类别的深度学习模型也不在评估范围内。表21给出了针对光学遥感图像目标检测的改进算法在两个数据集上的性能测试结果。

从表21可知:(1)面对NWPU-VHR10这样较简单的数据集,现有的深度学习模型在精度上已经达到很高的水平,当前先进的算法在平均精度值上能够达到90%以上,继续提升精度的空间有限。当面对DOTA这样更加复杂的数据集时,各算法的精度在整体上有明显的下降,后续仍然有较大的提升空间。

(2)和定向框相比,旋转框能够输出旋转角度信息,现在研究人员对旋转框检测的关注越来越多,旋转框算法的精度比定向框稍低,这表明旋转框的检测难度更高。

表21 各算法的改进型模型在NWPU-VHR10和DOTA上的性能

Table 21 Performance of improved models on NWPU-VHR10 and DOTA datasets

模型	作者及年份	数据集:NWPU-VHR10解决问题									
		多尺度	多方向	小尺寸	特征相似	背景复杂	密集分布	多形状	弱特征	标注方式	mAP/%
FSODM	Li等(2022d)					√			√	定向框	65.00
DFSSD	Qu等(2020)				√	√	√			定向框	65.35
HSF-Net	Li等(2018a)	√		√						定向框	81.15
	Li等(2018b)	√	√							定向框	87.12
TRD	Li等(2022e)	√		√			√			定向框	87.90
AFF SSD	Lu等(2021)			√						定向框	88.70
CAD-Net	Zhang等(2019a)	√				√				定向框	91.50
SCRDet	Yang等(2019)	√	√	√			√			定向框	91.75
CANet	Feng等(2021)	√			√			√		定向框	92.20
GLNet	Teng等(2022)	√								定向框	94.80
MSNet	Han等(2021)	√								定向框	95.40
HyNet	Zheng等(2020)	√		√						定向框	98.52
模型	作者及年份	数据集:DOTA解决问题									mAP
		多尺度	多方向	小尺寸	特征相似	背景复杂	密集分布	多形状	弱特征	标注方式	
AF-SSD	Yin等(2020)			√		√				定向框	52.60
Lo-det	Huang等(2022)	√							√	旋转框	66.17
RADet	ResNet101	√			√		√			定向框	69.09
GLNet	Teng等(2022)	√								定向框	70.80
FMSSD	Wang等(2020)	√		√				√		定向框	72.43
FSoD Net	Wang等(2022c)	√			√	√				定向框	75.33
AMFFANet	Chen等(2021)	√								定向框	76.27
R-YOLO	Hou等(2022)	√		√		√				定向框	77.01
SARA	Hou等(2021)	√		√				√		旋转框	79.91
SKNet	Cui等(2021)		√			√				旋转框	83.90
—	Zhu等(2023)	√				√				定向框	90.88

表19 目标检测评价指标

Table 19 Object detection evaluation indexes

评价指标	公式
交并比	$IOU = \frac{\text{area}(\text{Proposal} \cap \text{Ground Truth})}{\text{area}(\text{Proposal} \cup \text{Ground Truth})}$
精度	$Precision = \frac{TP}{TP + FP}$
召回率	$Recall = \frac{TP}{TP + FN}$
F-measure	$F_{\beta} = \frac{(1 + \beta^2)PR}{\beta^2 P + R}$
AP	$AP = \int_0^1 P(R)dR$
平均精度值	$mAP = \frac{1}{C} \sum_{i=1}^C \int_0^1 P_i(R_i)dR_i$

表20 混淆矩阵

Table 20 Confusion matrix

预测结果	实际结果为正样本	实际结果为负样本
预测结果为正样本	TP	FN
预测结果为负样本	FP	TN

(3) 利用特征金字塔进行多尺度融合现已经成为各先进算法的主流方法，可以有效检测多尺度目标。(4) 针对特征类间相似的相关研究并不多，目前仍然缺乏有效方案。

6 现存问题和发展方向

当前，针对光学遥感图像中的目标检测，一些问题依然没有被解决：

(1) 针对尺寸特别巨大的光学遥感图像目标，现有的模型仍然不能够直接将整幅图像输入进行全局处理，通过分割的方式会造成计算资源的消耗，此外，人为地分割图像有可能导致目标信息的丢失。

(2) 针对光学遥感图像中目标类间相似性的研究还比较少，目前仍然没有系统性的解决办法。

(3) 现有的深度学习模型更加关注遥感目标本身的特征，对全局和上下文信息的把握较少，一部分研究人员在算法中引入了全局信息和上下文信息，不过对于深层语义信息的解释性不足。

(4) 各个机构相继发布了自己的光学遥感数据集，并由专业人员进行标注。大规模的遥感数据集满足了深度学习的需求，不过受限于硬件条件，大量小尺寸、模糊、遮挡的目标仍然没有标注，这对现有模型性能的发挥也是一个限制。

(5) 由于卷积神经网络对于方向信息不敏感，因此，水平框会造成位置信息不准确等问题，当前，旋转检测框在面对密集目标时比水平检测框更加具有优势，但是，相比之下，旋转检测框仍然存在精度下降的缺点。

(6) 当前的光学遥感数据集在数据量上越来越大，导致整个预处理和训练过程变得繁琐耗时，此外，精度和实时性的平衡也是一个问题。

未来的主要发展方向有：

(1) 深度学习网络的改进。当前的光学遥感图像目标检测模型在精度上提升缓慢，性能上已趋于饱和，因此，如何进一步提升网络性能是目前急需解决的问题。作为特征提取的重要部分，主干网络在结构上的突破是关键，针对光学遥感图像目标的特征而设计的主干网络将会是一个研究重点。

(2) 轻量化网络的研究。卷积神经网络能够提取目标深层语义信息，为了加强特征提取能力，主流网络通常具有极深的网络结构，这无疑增加

了计算资源的消耗，而在实际的工程应用中，这些网络结构难以满足便携式设备的需求。因此，如何减少模型大小和参数量，使得网络在轻量化的同时保持较高的性能是一个很有潜力的研究方向。

(3) 弱监督学习。在处理超大规模数据集时，精确标注目标是一项极费人力的事情，更不要说在光学遥感数据集中存在大量小而密集、遮挡的情况，而弱监督训练则缓解了这一困难。然而，弱监督学习在性能上受限，这限制了其应用范围，此外，由于缺乏目标的边界信息，针对旋转框检测的弱监督学习算法仍然缺乏系统的研究，而水平框不能很好地适应复杂的光学遥感图像，因此在弱监督的条件下采用旋转框检测机制将会是一个很有潜力的发展方向。

(4) 小目标问题。在光学遥感图像中，小目标检测是一个重要的研究方向，由于目标的尺寸小，其占用像素少，且特征难以提取，目前各种改进算法的表现仍然不令人满意，主要的困难体现在以下几个问题：1) 样本不平衡：在光学遥感图像中，小目标所占的比例不大，因此对于各种先进的算法来说，小目标容易漏检；2) 损失不平衡：小目标对损失函数的影响较小，这也会导致其检测性能不佳；3) 匹配不平衡的问题：Bounding Box 和小目标之间轻微的偏差都会导致 IOU 的巨大变化，从而导致匹配失败。上述几点也是光学遥感图像目标检测领域需要系统性研究的方向。

(5) 改进旋转检测机制。目前，基于带有方向信息的旋转框检测机制已经越来越受到研究人员重视，大部分研究将目标的旋转角度设置在 0° — 180° ，而没有考虑到首尾的方向性，如：船舶目标并没有区分船首和船尾。此外，目前针对旋转目标的 IOU 计算存在一定的局限性，检测框和标注框角度的少许变化会导致 IOU 的急剧下降。因此上述两点都是下一步值得系统研究的方向。

7 结语

本文以深度学习为出发点，研究了光学遥感图像中的目标检测算法。首先分析了光学遥感图像中目标的特点，包括尺度多样性、方向多样性、形状多样性、小尺寸、特征相似、背景复杂、分布密集和特征弱，并总结了光学遥感图像目标检

测算法经历了从传统方法，包括模板匹配、先验知识和机器学习到基于深度学习算法的发展历程；其次，阐明了光学遥感图像目标检测的具体步骤，包括数据预处理、特征提取、生成 Bounding Box 和检测及后处理；随后，概述了经典的深度学习目标检测算法；重点针对光学遥感图像目标的特点总结了对应的改进算法；然后介绍了常用的光学遥感图像数据集以及评价指标并对一系列的算法进行性能比较；最后，针对目前仍然存在的问题，如光学遥感图像尺寸过大、目标类间相似性、深层语义信息的解释性较差、数据集标注困难、旋转框检测、精度和实时性平衡问题进行总结，并对未来的发展方向，如主干网络改进、轻量化网络研究、弱监督学习、小目标检测、旋转检测机制进行展望。

参考文献(References)

- Azimi S M, Vig E, Bahmanyar R, Körner M and Reinartz P. 2019. Towards multi-class object detection in unconstrained remote sensing imagery//14th Asian Conference on Computer Vision (ACCV). Perth: Springer: 150-165 [DOI: 10.1007/978-3-030-20893-6_10]
- Bai Z, Li G Y and Liu Z. 2023. Global-local-global context-aware network for salient object detection in optical remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 198: 184-196 [DOI: 10.1016/j.isprsjprs.2023.03.013]
- Bilen H and Vedaldi A. 2016. Weakly supervised deep detection networks//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE: 2846-2854 [DOI: 10.1109/CVPR.2016.311]
- Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: optimal speed and accuracy of object detection. arXiv Preprint arXiv: 2004.10934 [DOI: 10.48550/arXiv.2004.10934]
- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer: 213-229 [DOI: 10.1007/978-3-030-58452-8_13]
- Chen G W, Liu L, Guo J Y, Pan Z X and Hu W L. 2020. Semi-supervised airplane detection in remote sensing images using generative adversarial networks. Journal of University of Chinese Academy of Sciences, 37(4): 539-546 (陈国炜, 刘磊, 郭嘉逸, 潘宗序, 胡文龙. 2020. 基于生成对抗网络的半监督遥感图像飞机检测. 中国科学院大学学报, 37(4): 539-546) [DOI: 10.7523/j.issn.2095-6134.2020.04.014]
- Chen H, Cheng L, Zhuang Q Z, Zhang K, Li N, Liu L and Duan Z X. 2022a. Structure-aware weakly supervised network for building extraction from remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 60: 5412712 [DOI: 10.1109/TGRS.2022.3217830]
- Chen H, Zhang L B, Ma J and Zhang J. 2019. Target heat-map network: an end-to-end deep network for target detection in remote sensing images. Neurocomputing, 331: 375-387 [DOI: 10.1016/j.neucom.2018.11.044]
- Chen J Y, Sun J G, Li Y Q and Hou C B. 2022b. Object detection in remote sensing images based on deep transfer learning. Multimedia Tools and Applications, 81(9): 12093-12109 [DOI: 10.1007/s11042-021-10833-z]
- Chen L C, Liu C S, Chang F L, Li S and Nie Z Y. 2021. Adaptive multi-level feature fusion and attention-based network for arbitrary-oriented object detection in remote sensing imagery. Neurocomputing, 451: 67-80 [DOI: 10.1016/j.neucom.2021.04.011]
- Cheng B, Li Z Z, Xu B T, Yao X, Ding Z Q and Qin T Q. 2021a. Structured object-level relational reasoning CNN-based target detection algorithm in a remote sensing image. Remote Sensing, 13(2): 281 [DOI: 10.3390/rs13020281]
- Cheng G and Han J W. 2016. A survey on object detection in optical remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 117: 11-28 [DOI: 10.1016/j.isprsjprs.2016.03.014]
- Cheng G, Han J W, Zhou P C and Xu D. 2019. Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection. IEEE Transactions on Image Processing, 28(1): 265-278 [DOI: 10.1109/TIP.2018.2867198]
- Cheng G, Si Y J, Hong H L, Yao X W and Guo L. 2021b. Cross-scale feature fusion for object detection in optical remote sensing images. IEEE Geoscience and Remote Sensing Letters, 18(3): 431-435 [DOI: 10.1109/lgrs.2020.2975541]
- Cheng G, Yan B W, Shi P Z, Li K, Yao X W, Guo L and Han J W. 2022. Prototype-CNN for few-Shot object detection in remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 60: 5604610 [DOI: 10.1109/TGRS.2021.3078507]
- Cheng X T and Zhang C Y. 2022. C2-YOLO: rotating object detection network for remote sensing images with complex backgrounds// 2022 International Joint Conference on Neural Networks (IJCNN). Padua: IEEE: 1-8 [DOI: 10.1109/IJCNN55064.2022.9891999]
- Cui Z Y, Leng J X, Liu Y, Zhang T L, Quan P and Zhao W. 2021. SKNet: detecting rotated ships as keypoints in optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 59(10): 8826-8840 [DOI: 10.1109/TGRS.2021.3053311]
- Ding J, Xue N, Long Y, Xia G S and Lu Q K. 2019. Learning RoI transformer for oriented object detection in aerial images//32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE: 2844-2853 [DOI: 10.1109/CVPR.2019.00296]
- Dong R C, Xu D Z, Zhao J, Jiao L C and An J G. 2019. Sig-NMS-based Faster R-CNN combining transfer learning for small target detection in VHR optical remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 57(11): 8534-8545 [DOI: 10.1109/TGRS.2019.2921396]
- Feng X X, Han J W, Yao X W and Cheng G. 2021. TCANet: triple context-aware network for weakly supervised object detection in re-

- mote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(8): 6946-6955 [DOI: 10.1109/TGRS.2020.3030990]
- Fu C Y, Liu W, Ranga A, Tyagi A and Berg A C. 2017. DSSD: deconvolutional single shot detector. arXiv Preprint arXiv: 1701.06659 [DOI: 10.48550/arXiv.1701.06659]
- Fu K, Chang Z H, Zhang Y, Xu G L, Zhang K S and Sun X. 2020. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161: 294-308 [DOI: 10.1016/j.isprsjprs.2020.01.025]
- Fu Y M, Wu F G and Zhao J S. 2018. Context-aware and depthwise-based detection on orbit for remote sensing image//24th International Conference on Pattern Recognition (ICPR). Beijing: IEEE: 1725-1730 [DOI: 10.1109/ICPR.2018.8545815]
- Gao L, Gao H, Wang Y H, Liu D and Momanyi B M. 2023. Centerness and repulsion: constraints to improve remote sensing object detection via RepPoints. *Remote Sensing*, 15(6): 1479 [DOI: 10.3390/rs15061479]
- Ge Z, Liu S T, Wang F, Li Z M and Sun J. 2021. YOLOX: exceeding YOLO series in 2021. arXiv Preprint arXiv: 2107.08430 [DOI: 10.48550/arXiv.2107.08430]
- Girshick R. 2015. Fast R-CNN//2015 IEEE International Conference on Computer Vision. Santiago: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]
- Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus: IEEE: 580-587 [DOI: 10.1109/CVPR.2014.81]
- Han W X, Kuerban A, Yang Y C, Huang Z T, Liu B H and Gao J. 2021. Multi-vision network for accurate and real-time small object detection in optical remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 19: 6001205 [DOI: 10.1109/LGRS.2020.304422]
- He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]
- He K M, Zhang X Y, Ren S Q and Sun J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9): 1904-1916 [DOI: 1109/TPAMI.2015.2389824]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- He Y Q, Sun X, Gao L R and Zhang B. 2018. Ship detection without sea-land segmentation for large-scale high-resolution optical satellite images//2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia: IEEE: 717-720 [DOI: 10.1109/IGARSS.2018.8519391]
- Hou J B, Zhu X B and Yin X C. 2021. Self-adaptive aspect ratio anchor for oriented object detection in remote sensing images. *Remote Sensing*, 13(7): 1318 [DOI: 10.3390/rs13071318]
- Hou Y J, Shi G, Zhao Y X, Wang F, Jiang X, Zhuang R J, Mei Y F and Ma X J. 2022. R-YOLO: a YOLO-based method for arbitrary-oriented target detection in high-resolution remote sensing images. *Sensors*, 22(15): 5716 [DOI: 10.3390/s22155716]
- Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE: 2261-2269 [DOI: 10.1109/CVPR.2017.243]
- Huang H, Huo C L, Wei F L and Pan C H. 2019. Rotation and scale-invariant object detector for high resolution optical remote sensing images//2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Yokohama: IEEE: 1386-1389 [DOI: 10.1109/IGARSS.2019.8898495]
- Huang Z C, Li W, Xia X G, Wang H, Jie F R and Tao R. 2022. LO-Det: lightweight oriented object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5603515 [DOI: 10.1109/TGRS.2021.3067470]
- Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90 [DOI: 10.1145/3065386]
- Lan X T, Guo Z H and Li C H. 2021. Attention and feature fusion for aircraft target detection in optical remote sensing images. *Chinese Journal of Liquid Crystals and Displays*, 36(11): 1506-1515 (兰旭婷, 郭中华, 李昌昊. 2021. 基于注意力与特征融合的光学遥感图像飞机目标检测. 液晶与显示, 36(11): 1506-1515) [DOI: 10.37188/cjlc.2021.0088]
- Lei J J, Luo X W, Fang L Y, Wang M Y and Gu Y F. 2020. Region-enhanced convolutional neural network for object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(8): 5693-5702 [DOI: 10.1109/tgrs.2020.2968802]
- Li C Y, Li L L, Jiang H L, Weng K H, Geng Y F, Li L, Ke Z D, Li Q Y, Cheng M, Nie W Q, Li Y D, Zhang B, Liang Y F, Zhou L Y, Xu X M, Chu X X, Wei X M and Wei X L. 2022a. YOLOv6: a single-stage object detection framework for industrial applications. arXiv Preprint arXiv: 2209.02976 [DOI: 10.48550/arXiv.2209.02976]
- Li C Z, Xu C Y, Cui Z, Wang D, Zhang T and Yang J. 2019a. Feature-attentioned object detection in remote sensing imagery//2019 IEEE International Conference on Image Processing (ICIP). Taipei, China: IEEE: 3886-3890 [DOI: 10.1109/ICIP.2019.8803521]
- Li G Q, Hao X Y, Zha L L and Chen A B. 2022b. An outstanding adaptive multi-feature fusion YOLOv3 algorithm for the small target detection in remote sensing images. *Pattern Analysis and Applications*, 25(4): 951-962. [DOI: 10.1007/s10044-022-01072-5]
- Li J F, Li Z F, Chen M X, Wang Y L and Luo Q H. 2022c. A new ship detection algorithm in optical remote sensing images based on improved R³Det. *Remote Sensing*, 14(19): 5048 [DOI: 10.3390/rs14195048]
- Li J X, Zhang Z L, Tian Y, Xu Y P, Wen Y H and Wang S C. 2022d. Target-guided feature super-resolution for vehicle detection in remote sensing images. *IEEE Geoscience and Remote Sensing Letters*,

- ters, 19: 8020805 [DOI: 10.1109/lgrs.2021.3112172]
- Li K, Cheng G, Bu S H and You X. 2018a. Rotation-insensitive and context-augmented object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 56(4): 2337-2348 [DOI: 10.1109/TGRS.2017.2778300]
- Li M J, Guo W W, Zhang Z H, Yu W X and Zhang T. 2018b. Rotated region based fully convolutional network for ship detection//2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Valencia: IEEE: 673-676 [DOI: 10.1109/IGARSS.2018.8519094]
- Li Q P, Mou L C, Liu Q J, Wang Y H and Zhu X X. 2018c. HSF-Net: multiscale deep feature embedding for ship detection in optical remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 56(12): 7147-7161 [DOI: 10.1109/TGRS.2018.2848901]
- Li Q Y, Chen Y S and Zeng Y. 2022e. Transformer with transfer CNN for remote-sensing-image object detection. *Remote Sensing*, 14(4): 984 [DOI: 10.3390/rs14040984]
- Li S, Xu Y L, Zhu M M, Ma S P and Tang H. 2019b. Remote sensing airport detection based on end-to-end deep transferable convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 16(10): 1640-1644 [DOI: 10.1109/LGRS.2019.2904076]
- Li X, Deng J Y and Fang Y. 2022f. Few-shot object detection on remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5601614 [DOI: 10.1109/TGRS.2021.3051383]
- Li X B, Jiang B T, Yang Y B, Fu Y Z and Yue W Z. 2019. A survey on object detection technology in optical remote sensing images. *Spacecraft Recovery and Remote Sensing*, 40(4): 95-104 (李晓斌, 江碧涛, 杨渊博, 傅雨泽, 岳文振. 2019. 光学遥感图像目标检测技术综述. 航天返回与遥感, 40(4): 95-104) [DOI: 10.3969/j.issn.1009-8518.2019.04.011]
- Li Y S, Zhang Y J, Huang X and Yuille A L. 2018d. Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146: 182-196 [DOI: 10.1016/j.isprsjprs.2018.09.014]
- Li Y Y, Huang Q, Pei X, Chen Y Q, Jiao L C and Shang R H. 2021. Cross-layer attention network for small object detection in remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 2148- 2161 [DOI: 10.1109/jstars.2020.3046482]
- Li Y Y, Huang Q, Pei X, Jiao L C and Shang R H. 2020. RADet: refine feature pyramid network and multi-layer attention network for arbitrary-oriented object detection of remote sensing images. *Remote Sensing*, 12(3): 389 [DOI: 10.3390/rs12030389]
- Li Z, Wang Y C, Zhang N, Zhang Y X, Zhao Z K, Xu D D, Ben G L and Gao Y X. 2022g. Deep learning-based object detection techniques for remote sensing images: a survey. *Remote Sensing*, 14(10): 2385 [DOI: 10.3390/rs14102385]
- Li Z X, Yang L and Zhou F Q. 2024. FSSD: feature fusion single shot multibox detector. *arXiv Preprint arXiv: 1712.00960* [DOI: 10.48550/arXiv.1712.00960]
- Liao Y R, Wang H N, Lin C B, Li Y, Fang Y Q and Ni S Y. 2022. Research progress of deep learning-based object detection of optical remote sensing image. *Journal on Communications*, 43(5): 190-203 (廖育荣, 王海宁, 林存宝, 李阳, 方宇强, 倪淑燕. 2022. 基于深度学习的光学遥感图像目标检测研究进展. 通信学报, 43(5): 190-203) [DOI: 10.1109/j.issn.1000-436x.2022071]
- Lin H J, Hao M, Luo W Q, Yu H Y and Zheng N S. 2023. BEARNet: a novel buildings edge-aware refined network for building extraction from high-resolution remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 20: 6005305 [DOI: 10.1109/LGRS.2023.3272353]
- Lin Y T, Feng P M, Guan J, Wang W W and Chambers J. 2021. IENet: interacting embankment one stage anchor free detector for orientation aerial object detection. *arXiv Preprint arXiv: 1912.00969* [DOI: 10.48550/arXiv.1912.00969]
- Liu S, Zhang L, Lu H C and He Y. 2022. Center-boundary dual attention for oriented object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5603914 [DOI: 10.1109/TGRS.2021.3069056]
- Liu T Y, Li W G and Guan J H. 2020. Deep learning based object detection in optical remote sensing images: a survey. *Radio Communications Technology*, 46(6): 624-634 (刘天颖, 李文根, 关佶红. 2020. 基于深度学习的光学遥感图像目标检测方法综述. 无线电通信技术, 46(6): 624-634) [DOI: 10.3969/j.issn.1003-3114.2020.06.002]
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot multibox detector. *arXiv Preprint arXiv: 1512.02325* [DOI: 10.48550/arXiv.1512.02325]
- Liu W C, Ma L and Chen H. 2018. Arbitrary-oriented ship detection framework in optical remote-sensing images. *IEEE Geoscience and Remote Sensing Letters*, 15(6): 937-941 [DOI: 10.1109/LGRS.2018.2813094]
- Liu Z K, Hu H G, Weng L B and Yang Y P. 2017. Rotated region based CNN for ship detection//2017 IEEE International Conference on Image Processing (ICIP). Beijing: IEEE: 900- 904 [DOI: 10.1109/ICIP.2017.8296411]
- Long H, Chung Y N, Liu Z B and Bu S H. 2019. Object detection in aerial images using feature fusion deep networks. *IEEE Access*, 7: 30980-30990 [DOI: 10.1109/access.2019.2903422]
- Long Y, Gong Y P, Xiao Z F and Liu Q. 2017. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5): 2486-2498 [DOI: 10.1109/tgrs.2016.2645610]
- Lu X C, Ji J, Xing Z Q and Miao Q G. 2021. Attention and feature fusion SSD for remote sensing object detection. *IEEE Transactions on Instrumentation and Measurement*, 70: 5501309 [DOI: 10.1109/TIM.2021.3052575]
- Ma T L, Mao M Y, Zheng H H, Gao P, Wang X D, Han S M, Ding E R, Zhang B C and Doermann D. 2021. Oriented object detection with transformer. *arXiv Preprint arXiv: 2106.03146* [DOI: 10.48550/arXiv.2106.03146]
- Nie G T and Huang H. 2021. A survey of object detection in optical re-

- mote sensing images. *Acta Automatica Sinica*, 47(8): 1749-1768
 (聂光涛, 黄华. 2021. 光学遥感图像目标检测算法综述. 自动化学报, 47(8): 1749-1768) [DOI: 10.16383/j.aas.c200596]
- Qing Y H, Liu W Y, Feng L Y and Gao W J. 2021. Improved YOLO network for free-angle remote sensing target detection. *Remote Sensing*, 13(11): 2171 [DOI: 10.3390/rs13112171]
- Qiu H Q, Li H L, Wu Q B, Meng F M, Ngan K N and Shi H C. 2019. A2RMNet: adaptively aspect ratio multi-scale network for object detection in remote sensing images. *Remote Sensing*, 11(13): 1594 [DOI: 10.3390/rs11131594]
- Qu J S, Su C, Zhang Z W and Razi A. 2020. Dilated convolution and feature fusion SSD network for small object detection in remote sensing images. *IEEE Access*, 8: 82832-82843 [DOI: 10.1109/ACCESS.2020.2991439]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]
- Redmon J and Farhadi A. 2017. YOLO9000: better, faster, stronger// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE: 6517-6525 [DOI: 10.1109/CVPR.2017.690]
- Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement. arXiv Preprint 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): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Ren Y, Zhu C R and Xiao S P. 2018. Deformable Faster R-CNN with aggregating multi-layer features for partially occluded object detection in optical remote sensing images. *Remote Sensing*, 10(9): 1470 [DOI: 10.3390/rs10091470]
- Schilling H, Bulatov D, Niessner R, Middelmann W and Soergel U. 2018. Detection of vehicles in multisensor data via multibranch convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(11): 4299-4316 [DOI: 10.1109/JSTARS.2018.2825099]
- Shi G, Zhang J S, Liu J M, Zhang C X, Zhou C S and Yang S Y. 2021a. Global context-augmented objection detection in VHR optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12): 10604-10617 [DOI: 10.1109/TGRS.2020.3043252]
- Shi L K, Tang Z J, Wang T T, Xu X, Liu J and Zhang J. 2021b. Aircraft detection in remote sensing images based on deconvolution and position attention. *International Journal of Remote Sensing*, 42(11): 4241-4260 [DOI: 10.1080/01431161.2021.1892858]
- Shi W X, Tan D L and Bao S L. 2020. Feature enhancement SSD algorithm and its application in remote sensing images target detection. *Acta Photonica Sinica*, 49(1): 0128002 (史文旭, 谭代伦, 鲍胜利. 2020. 特征增强 SSD 算法及其在遥感目标检测中的应用. 光子学报, 49(1): 0128002) [DOI: 10.3788/gzxb20204901.0128002]
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. arXiv Preprint arXiv: 1409.1556 [DOI: 10.48550/arXiv.1409.1556]
- Song Z N, Sui H G and Li Y C. 2021. A survey on ship detection technology in high-resolution optical remote sensing images. *Geomatics and Information Science of Wuhan University*, 46(11): 1703-1715 (宋志娜, 眭海刚, 李永成. 2021. 高分辨率可见光遥感图像舰船目标检测综述. 武汉大学学报(信息科学版), 46(11): 1703-1715) [DOI: 10.13203/j.whugis20200481]
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE: 1-9 [DOI: 10.1109/cvpr.2015.7298594]
- Tang P, Wang X G, Bai X and Liu W Y. 2017. Multiple instance detection network with online instance classifier refinement//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE: 3059-3067 [DOI: 10.1109/CVPR.2017.326]
- Tang R and Dong G G. 2022. Bridge-over-water detection via modulated deformable convolution and attention mechanisms. *EURASIP Journal on Advances in Signal Processing*, 2022(1): 56 [DOI: 10.1186/s13634-022-00891-2]
- Teng Z, Duan Y N, Liu Y, Zhang B P and Fan J P. 2022. Global to local: clip-LSTM-based object detection from remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5603113 [DOI: 10.1109/TGRS.2021.3064840]
- Tian Z Z, Zhan R H, Wang W, He Z Q, Zhang J and Zhuang Z W. 2020. Object detection in optical remote sensing images by integrating object-to-object relationships. *Remote Sensing Letters*, 11(5): 416-425 [DOI: 10.1080/2150704X.2020.1722330]
- Wan F, Liu C, Ke W, Ji X Y, Jiao J B and Ye Q X. 2019. C-MIL: continuation multiple instance learning for weakly supervised object detection//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE: 2194-2203 [DOI: 10.1109/CVPR.2019.00230]
- Wang B L, Zhao Y Q and Li X L. 2022a. Multiple instance graph learning for weakly supervised remote sensing object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5613112 [DOI: 10.1109/TGRS.2021.3123231]
- Wang C Y, Bochkovskiy A and Liao H Y M. 2022b. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv Preprint arXiv: 2207.02696 [DOI: 10.48550/arXiv.2207.02696]
- Wang G Q, Zhuang Y, Wang Z R, Chen H, Shi H and Chen L. 2019a. Spatial Enhanced-SSD for multiclass object detection in remote sensing images//2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama: IEEE: 318-321 [DOI: 10.1109/IGARSS.2019.8898526]
- Wang G Q, Zhuang Y, Chen H, Liu X, Zhang T, Li L L, Dong S and Sang Q B. 2022c. FSOD-Net: full-scale object detection from optical remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5602918 [DOI: 10.1109/TGRS.2021.3064599]

- Wang H, Li H, Qian W L, Diao W H, Zhao L J, Zhang J H and Zhang D B. 2021a. Dynamic pseudo-label generation for weakly supervised object detection in remote sensing images. *Remote Sensing*, 13(8): 1461 [DOI: 10.3390/rs13081461]
- Wang J K, He X H, Shao F M, Lu G L, Jiang Q Y and Hu R Z. 2022d. Multi-size object detection in large scene remote sensing images under dual attention mechanism. *IEEE Access*, 10: 8021-8035 [DOI: 10.1109/access.2022.3141059]
- Wang J W, Ding J, Guo H W, Cheng W S, Pan T and Yang W. 2019b. Mask OBB: a semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images. *Remote Sensing*, 11(24): 2930 [DOI: 10.3390/rs11242930]
- Wang P J, Sun X, Diao W H and Fu K. 2020. FMSSD: feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5): 3377-3390 [DOI: 10.1109/TGRS.2019.2954328]
- Wang X B, Jiang Y Y, Luo Z B, Liu C L, Choi H and Kim S. 2019c. Arbitrary shape scene text detection with adaptive text region representation//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 6442-6451 [DOI: 10.1109/CVPR.2019.00661]
- Wang X L, Ban Y, Guo H M and Hong L. 2019d. Deep learning model for target detection in remote sensing images fusing multilevel features//2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama: IEEE: 250-253 [DOI: 10.1109/IGARSS.2019.8898759]
- Wang X L, Liang M and Liu T. 2022. Feature enhanced single-stage remote sensing image object detection model. *Journal of Xidian University*, 49(3): 160-170 (汪西莉, 梁敏, 刘涛. 2022. 特征增强的单阶段遥感图像目标检测模型. 西安电子科技大学学报, 49(3): 160-170) [DOI: 10.19665/j.issn1001-2400.2022.03.018]
- Wang Y, Jia Y N and Gu L Z. 2021b. EFM-Net: feature extraction and filtration with mask improvement network for object detection in remote sensing images. *Remote Sensing*, 13(20): 4151 [DOI: 10.3390/rs13204151]
- Wang Y W, Guo Y and Shao X Y. 2022. Target detection in remote sensing images based on improved cascade algorithm. *Acta Optica Sinica*, 42(24): 2428004 (王友伟, 郭颖, 邵香迎. 2022. 基于进级联算法的遥感图像目标检测. 光学学报, 42(24): 2428004) [DOI: 10.3788/aos202242.2428004]
- Wang Z Q, Zhou Y, Wang F T, Wang S X and Xu Z Y. 2021c. SDGH-Net: ship detection in optical remote sensing images based on Gaussian heatmap regression. *Remote Sensing*, 13(3): 499 [DOI: 10.3390/rs13030499]
- Wei H R, Wan G and Ji S P. 2023. ParallelTracker: a transformer based object tracker for UAV videos. *Remote Sensing*, 15(10): 2544 [DOI: 10.3390/rs15102544]
- Wei H R, Zhang Y, Chang Z H, Li H, Wang H Q and Sun X. 2020. Oriented objects as pairs of middle lines. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169: 268-279 [DOI: 10.1016/j.isprsjprs.2020.09.022]
- Wu L S, Fang L Y, Yue J, Zhang B, Ghamisi P and He M. 2022. Deep bilateral filtering network for point-supervised semantic segmentation in remote sensing images. *IEEE Transactions on Image Processing*, 31: 7419-7434 [DOI: 10.1109/TIP.2022.3222904]
- Wu Y F, Zhao W, Zhang R F and Jiang F. 2021. AMR-Net: arbitrary-oriented ship detection using attention module, multi-scale feature fusion and rotation pseudo-label. *IEEE Access*, 9: 68208-68222 [DOI: 10.1109/ACCESS.2021.3075857]
- Wu Z Z, Weise T, Wang Y and Wang Y J. 2020. Convolutional neural network based weakly supervised learning for aircraft detection from remote sensing image. *IEEE Access*, 8: 158097-158106 [DOI: 10.1109/ACCESS.2020.3019956]
- Xu X K, Feng Z J, Cao C Q, Li M Y, Wu J, Wu Z Y, Shang Y J and Ye S B. 2021. An improved swin transformer-based model for remote sensing object detection and instance segmentation. *Remote Sensing*, 13(23): 4779 [DOI: 10.3390/rs13234779]
- Xu Z J and Bai X. 2022. Small ship target detection method for remote sensing images based on dual feature enhancement. *Acta Optica Sinica*, 42(18): 1828002 (徐志京, 柏雪. 2022. 基于双重特征增强的遥感舰船小目标检测. 光学学报, 42(18): 1828002) [DOI: 10.3788/AOS202242.1828002]
- Xu Z Z, Xu X, Wang L, Yang R and Pu F L. 2017. Deformable ConvNet with aspect ratio constrained NMS for object detection in remote sensing imagery. *Remote Sensing*, 9(12): 1312 [DOI: 10.3390/rs9121312]
- Yang J Z, Wu J L and Zhi J. 2021. Aircraft target detection in remote sensing image based on multi-scale circle frequency filter and convolutional neural network. *Journal of Electronics and Information Technology*, 43(5): 1397-1404 (杨钧智, 吴金亮, 智军. 2021. 基于多尺度圆周频率滤波与卷积神经网络的遥感图像飞机目标检测方法研究. 电子与信息学报, 43(5): 1397-1404) [DOI: 10.11999/jeit200144]
- Yang X, Hou L P, Zhou Y, Wang W T and Yan J C. 2021. Dense label encoding for boundary discontinuity free rotation detection//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE: 15814-15824 [DOI: 10.1109/CVPR46437.2021.01556]
- Yang X, Sun H, Fu K, Yang J R, Sun X, Yan M L and Guo Z. 2018a. Automatic ship detection in remote sensing images from Google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sensing*, 10(1): 132 [DOI: 10.3390/rs10010132]
- Yang X, Sun H, Sun X, Yan M L, Guo Z and Fu K. 2018b. Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network. *IEEE Access*, 6: 50839-50849 [DOI: 10.1109/ACCESS.2018.2869884]
- Yang X and Yan J C. 2022. On the Arbitrary-oriented object detection: classification based approaches revisited. *arXiv Preprint arXiv: 2003.05597* [DOI: 10.48550/arXiv.2003.05597]
- Yang X, Yang J R, Yan J C, Zhang Y, Zhang T F, Guo Z, Sun X and Fu K. 2019. SCRDet: towards more robust detection for small, cluttered and rotated objects//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE: 8231- 8240

- [DOI: 10.1109/ICCV.2019.00832]
- Yang Y, Pan Z X, Hu Y X and Ding C B. 2022. PistonNet: object separating from background by attention for weakly supervised ship detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 5190- 5202 [DOI: 10.1109/JSTARS.2022.3184637]
- Yao X W, Feng X X, Han J W, Cheng G and Guo L. 2021. Automatic weakly supervised object detection from high spatial resolution remote sensing images via dynamic curriculum learning. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1): 675-685 [DOI: 10.1109/TGRS.2020.2991407]
- Ye X H, Xiong F C, Lu J F, Zhou J and Qian Y T. 2020. \mathcal{F} -Net: feature fusion and filtration network for object detection in optical remote sensing images. *Remote Sensing*, 12(24): 4027 [DOI: 10.3390/rs12244027]
- Yin R H, Zhao W, Fan X D and Yin Y F. 2020. AF-SSD: an accurate and fast single shot detector for high spatial remote sensing imagery. *Sensors*, 20(22): 6530 [DOI: 10.3390/s20226530]
- Ying X, Wang Q, Li X W, Yu M, Jiang H, Gao J, Liu Z Q and Yu R G. 2019. Multi-attention object detection model in remote sensing images based on multi-scale. *IEEE Access*, 7: 94508-94519 [DOI: 10.1109/access.2019.2928522]
- Yuan Z, Fang W, Zhao Y M and Sheng V S. 2021. Research of insect recognition based on improved YOLOv5. *Journal on Artificial Intelligence*, 3(4): 145-152 [DOI: 10.32604/jai.2021.026902]
- Yue Y B, Chen F C and Chen G L. 2022. Statistical analysis of pre-seismic anomalies from FY-2G satellite infrared remote sensing images. *National Remote Sensing Bulletin*, 26(12): 2655-2664 (乐应波, 陈福春, 陈桂林. 2022. FY-2G 卫星红外遥感图像中的震前异常统计分析. 遥感学报, 26(12): 2655-2664) [DOI: 10.11834/jrs.20210251]
- Zhang C G, Xiong B L and Kuang G Y. 2020. A survey of ship detection in optical satellite remote sensing images. *Chinese Journal of Radio Science*, 35(5): 637-647 (张财广, 熊博莅, 匡纲要. 2020. 光学卫星遥感图像舰船目标检测综述. 电波科学学报, 35(5): 637-647) [DOI: 10.13444/jcjrs.2020040603]
- Zhang G J, Lu S J and Zhang W. 2019a. CAD-Net: a context-aware detection network for objects in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12): 10015-10024 [DOI: 10.1109/TGRS.2019.2930982]
- Zhang H, Liu F, Fan Y Y, Tan F X and Qian Y R. 2022a. FAFFENet: frequency attention and feature fusion enhancement network for multiscale remote sensing target detection. *Journal of Applied Remote Sensing*, 16(1): 014512 [DOI: 10.1117/1.JRS.16.014512]
- Zhang S X, Song F, Liu X Y, Hao X Y, Liu Y J, Lei T and Jiang P. 2023. Text semantic fusion relation graph reasoning for few-shot object detection on remote sensing images. *Remote Sensing*, 15(5): 1187 [DOI: 10.3390/rs15051187]
- Zhang T, Zhuang Y, Wang G Q, Dong S, Chen H and Li L L. 2022b. Multiscale semantic fusion-guided fractal convolutional object detection network for optical remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5608720 [DOI: 10.1109/tgrs.2021.3108476]
- Zhang W, Wang S H, Thachan S, Chen J Z and Qian Y T. 2018. Deconv R-CNN for small object detection on remote sensing images//2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia: IEEE: 2483-2486 [DOI: 10.1109/IGARSS.2018.8517436]
- Zhang Y L, Yuan Y, Feng Y C, Lu X Q. 2019b. Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8): 5535-5548 [DOI: 10.1109/TGRS.2019.2900302]
- Zheng Y B, Sun P, Zhou Z T, Xu W Y and Ren Q. 2021. ADT-Det: adaptive dynamic refined single-stage transformer detector for arbitrary-oriented object detection in satellite optical imagery. *Remote Sensing*, 13(13): 2623 [DOI: 10.3390/rs13132623]
- Zheng Z, Zhong Y F, Ma A L, Han X B, Zhao J, Liu Y F and Zhang L P. 2020. HyNet: hyper-scale object detection network framework for multiple spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166: 1-14 [DOI: 10.1016/j.isprsjprs.2020.04.019]
- Zhong Y F, Han X B and Zhang L P. 2018. Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 138: 281-294 [DOI: 10.1016/j.isprsjprs.2018.02.014]
- Zhong Y F, Zheng Z, Ma A L, Lu X Y and Zhang L P. 2020. COLOR: cycling, offline learning, and online representation framework for airport and airplane detection using GF-2 satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12): 8438-8449 [DOI: 10.1109/TGRS.2020.2987907]
- Zhou L L and El-Deiry W S. 2009. Multispectral fluorescence imaging. *Journal of Nuclear Medicine*, 50(10): 1563-1566 [DOI: 10.2967/jnmed.109.063925]
- Zhou L M, Zheng C, Yan H X, Zuo X Y, Liu Y, Qiao B J and Yang Y. 2022. RepDarkNet: a multi-branched detector for small-target detection in remote sensing images. *ISPRS International Journal of Geo-Information*, 11(3): 158 [DOI: 10.3390/ijgi11030158]
- Zhu F Z, Wang Y Y, Cui J Y, Liu G X and Li H L. 2023. Target detection for remote sensing based on the enhanced YOLOv4 with improved BiFPN. *The Egyptian Journal of Remote Sensing and Space Science*, 26(2): 351-360 [DOI: 10.1016/j.ejrs.2023.04.003]
- Zhu W T, Xie B R, Wang Y, Shen J and Zhu H W. 2020. Survey on aircraft detection in optical remote sensing images. *Computer Science*, 47(S2): 165-171, 182 (祝文韬, 谢宝蓉, 王琰, 沈霁, 朱浩文. 2020. 光学遥感图像中的飞机目标检测技术研究综述. 计算机科学, 47(S2): 165-171, 182) [DOI: 10.11896/jskx.190500176]
- Zhu X K, Lyu S C, Wang X and Zhao Q. 2021. TPH-YOLOV5: improved YOLOV5 based on transformer prediction head for object detection on drone-captured scenarios//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal: IEEE: 2778-2788 [DOI: 10.1109/ICCVW54120.2021.00312]
- Zhu Y X, Du J and Wu X Q. 2020. Adaptive period embedding for rep-

resenting oriented objects in aerial images. IEEE Transactions on
Geoscience and Remote Sensing, 58(10): 7247-7257 [DOI: 10.

1109/TGRS.2020.2981203]

Progress of research on deep learning algorithms for object detection in optical remote sensing images

XU Danqing, WU Yiquan

College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics,
Nanjing 211106, China

Abstract: Among all applications of optical remote sensing images, object detection has always been given more attention by researchers. Object detection has a wide application prospect in military and civilian fields. This study reviews the progress of research on object detection algorithms in optical remote sensing images on the basis of deep learning. The characteristics of remote sensing objects are different from those of conventional objects. First, remote sensing equipment has a long imaging distance, so it can cover a large range. The images may have objects with large scale and shape changes. Second, in remote sensing images, the background tends to occupy a large area. As a result, some objects are often submerged in the complex background, and detectors cannot distinguish these objects effectively. Last, in remote sensing images, the objects do not only have a small size and changeable direction. Sometimes, remote sensing objects are densely distributed, posing challenges to the detection of optical remote sensing objects.

This study introduces the development process of optical remote sensing object detection algorithms from template matching, prior knowledge, and machine learning to deep learning. Then, the process of optical remote sensing object detection based on deep learning, including data preprocessing, feature extraction, detection, and postprocessing, is introduced in detail. Classical deep learning-based object detection algorithms, including the one-stage algorithms represented by YOLO and SSD and the two-stage algorithm represented by Faster RCNN, are summarized. Afterward, in accordance with the characteristics of optical remote sensing image objects, various improved algorithms for addressing the optical remote sensing image object detection problems of scale diversity, direction diversity, shape diversity, small size, feature similarity, background complexity, distribution density, and weak features are systematically summarized. Non-strong supervised learning-based optical remote sensing image object detection methods and other advanced algorithms, such as Transformer-based algorithms, transfer learning-based algorithms, knowledge graph-based algorithms, and prior knowledge-based algorithms, are also summarized. In addition, open-source optical remote sensing image datasets and the performance of object detection evaluation indexes are introduced.

The Mean Average Precision (mAP) of advanced algorithms on the NWPU-VHR10 dataset can exceed 90%. On the DOTA dataset, the mAP of each advanced algorithm decreases considerably, and the advanced algorithms proposed in recent years continue to improve their performance on this dataset. Multiscale fusion with a feature pyramid network has become the mainstream method of advanced algorithms, which can detect multiscale objects effectively.

Many improved algorithms have been proposed to solve the abovementioned problems in optical remote sensing image object detection, and good detection results have been achieved. However, the research on object detection in large-scale remote sensing images and similar object detection between classes remains lacking. Furthermore, this study proposes future development directions, such as improved deep learning networks, lightweight networks, weakly supervised learning, small-object detection, and improved rotary detection mechanisms.

Key words: optical remote sensing image, object detection, deep learning, object characteristics, object detection process, object detection framework, incomplete supervised learning, dataset, evaluation index

Supported by National Natural Science Foundation of China (No. 61573183)