

# 光学遥感图像目标检测算法综述

聂光涛<sup>1</sup> 黄华<sup>2</sup>

**摘要** 目标检测技术是光学遥感图像理解的基础问题,具有重要的应用价值. 本文对遥感图像目标检测算法发展进行了梳理和分析. 首先阐述了遥感图像目标检测的特点和挑战; 之后系统总结了典型的检测方法, 包括早期的基于手工设计特征的算法和现阶段基于深度学习的方法, 对于深度学习方法首先介绍了典型的目标检测模型, 进而针对遥感图像本身的难点详细梳理了优化改进方案; 接着介绍了常用的检测数据集, 并对现有方法的性能进行比较; 最后对现阶段问题进行总结并对未来发展趋势进行展望.

**关键词** 光学遥感, 目标检测, 手工设计特征, 深度学习, 数据集

**引用格式** 聂光涛, 黄华. 光学遥感图像目标检测算法综述. 自动化学报, 2021, 47(8): 1749–1768

**DOI** 10.16383/j.aas.c200596

## A Survey of Object Detection in Optical Remote Sensing Images

NIE Guang-Tao<sup>1</sup> HUANG Hua<sup>2</sup>

**Abstract** Object detection in optical aerial images is a fundamental problem in the field of remote sensing and shows great importance in the application. The performance of the early hand-craft-feature algorithm is limited, while deep learning is the primary method for object detection at present. However, due to the characteristics of the remote sensing image itself, it is difficult for the existing detection algorithms to perform well on these images. In this paper, we first describe the characteristics and challenges of the object detection task in aerial images. We then summarize the typical detection methods, including early hand-craft feature extraction methods and current deep learning methods, especially for the deep learning algorithm enhancement upon the characteristics of the aerial images. Then the commonly used detection datasets are introduced, and the performances of the existing methods are compared. Finally, we summarize the current deficiencies and analyze the trends of future studies.

**Key words** Remote sensing, object detection, hand-craft features, deep learning, dataset

**Citation** Nie Guang-Tao, Huang Hua. A survey of object detection in optical remote sensing images. *Acta Automatica Sinica*, 2021, 47(8): 1749–1768

光学遥感, 是指传感器工作波段在可见光波段, 即 0.38 ~ 0.76 微米范围的遥感技术, 在此波段下采集的图像即为光学遥感图像. 光学遥感图像包含大量细节信息, 直观反映地物目标的形状、颜色和纹理等, 便于人眼直接观察. 光学遥感图像目标检测是指利用特定的算法从图像中搜索并标记出感兴趣的目标, 例如飞机、坦克、船只、车辆等. 光学遥感目标检测作为一种基础技术手段, 在遥感图像分析中具有重要的作用, 是城市规划、土地利用、交通疏导、军事监测等应用的重要基础. 随着对地观测技术的

发展, 获取到的高分辨率遥感图像数据规模越来越大. 高分辨率的遥感图像提供了更高的图像质量和更丰富的细节信息, 为目标检测在遥感领域的发展提供了大的机遇.

较早的光学遥感图像目标检测算法均是基于手工设计特征的方式, 首先提取候选区域, 之后针对待测目标人工设计特征, 最后结合分类器进行分类确定目标类别. 候选区域提取需要设置大量的滑动窗口, 产生大量的冗余计算, 时间复杂度高; 人工设计的特征主要根据目标视觉信息 (颜色、纹理、边缘等) 进行提取, 可解释性较强, 但是特征表达能力弱、鲁棒性差、适应范围小, 难以应用在复杂多变的环境中. 综上所述, 早期检测算法远不能满足实际应用需求.

随着深度学习的发展, 神经网络提取的深层特征有着更强的语义表征能力和判别性, 目标检测性能获得了进一步的提升. 然而由于遥感图像自身的特点, 现有检测手段尚不能在遥感图像上取得好的

收稿日期 2020-07-27 录用日期 2020-12-01  
Manuscript received July 27, 2020; accepted December 1, 2020  
国家自然科学基金 (61936011) 资助  
Supported by National Natural Science Foundation of China (61936011)

本文责任编辑 张军平

Recommended by Associate Editor ZHANG Jun-Ping  
1. 北京理工大学计算机科学与技术学院 北京 100081 2. 北京师范大学人工智能学院 北京 100875

1. School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081 2. School of Artificial Intelligence, Beijing Normal University, Beijing 100875

表现. 针对遥感图像的特点, 提出了大量改进方案, 有效提升了目标检测在遥感领域的性能.

目前针对光学遥感图像的目标检测已经有一定数量的综述文献发表. Wang 等<sup>[1]</sup>主要针对海上特定目标(舰船)检测进行总结梳理; Cheng 等<sup>[2]</sup>总结了现阶段遥感图像通用目标检测方法, 以基于手工设计特征的研究为主, 对当前主流的基于深度学习的方法介绍较少; 鉴于此, 刘小波等<sup>[3]</sup>和 Li 等<sup>[4]</sup>专门梳理了基于深度学习的光学遥感图像目标检测研究; 除了光学遥感图像, 张号逵等<sup>[5]</sup>还重点梳理了基于深度学习的高光谱遥感图像分类方法, 实现了在光谱图像上地物的检测和精准定位.

与现有综述不同, 本文对光学遥感图像中通用目标检测方法进行全面、细致的梳理, 既包括早期基于手工设计特征的方法, 也包括当前深度学习的方法, 同时重点对基于深度学习的方法及其针对遥感图像特点的改进进行更为精细的梳理, 此外还额外评估各个改进方法的性能定量改善, 并根据评估结果指出现存的问题和未来的发展方向.

## 1 遥感图像目标检测的特点与挑战

针对自然图像通用目标检测算法的探索已持续多年, 从传统的基于手工特征提取到现阶段的深度学习, 性能获得了大幅度的提升. 然而现阶段在常规图像中性能极佳的通用检测算法, 在遥感图像上表现并不理想. 主要原因是由于遥感图像自身的特点(如图 1 所示)为目标检测带来了较大的难度和挑战. 遥感图像的特点和面临的挑战主要总结如下:

1) 超大图像尺寸: 遥感图像具有超大的图像尺寸和覆盖面积. 典型检测算法针对的图像尺寸较小, 难以直接应用在遥感领域. 同时, 遥感图像中的背景占了较大比重, 目标区域范围较小, 典型检测算法对每一区域同等对待, 计算效率极低.

2) 方向变化大: 遥感图像从空中视角进行拍摄, 场景为俯视图, 目标以多种角度分布在场景中, 现有多数算法对角度的适应性不高, 在处理多方向问题时不够鲁棒. 此外经典的水平框定位方式在对多方向目标进行定位时, 包围框不够紧凑, 定位不够精细.

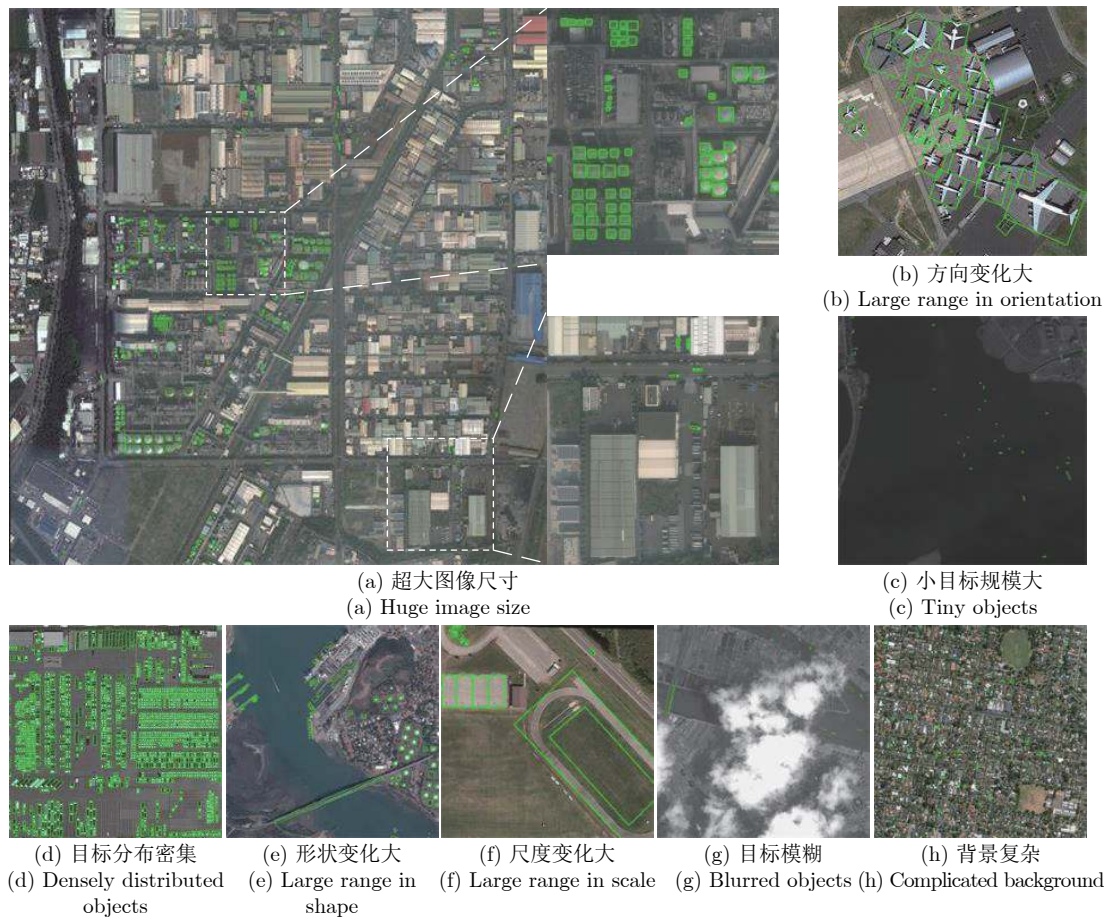


图 1 遥感图像目标检测的特点与挑战

Fig. 1 Characteristics and challenges of object detection in remote sensing images

3) 小目标规模大: 遥感图像小目标的数量占比较大. 小目标在现有检测算法中极容易因特征降采样而丢失, 从而造成漏检. 小目标检测的方法在常规自然图像检测中仍没有得到很好的解决, 遥感领域更是加大了检测的难度.

4) 目标分布密集: 遥感图像存在较大规模密集分布的目标, 目标之间容易产生相互干扰, 从而导致较大的定位误差, 同时漏检、误检的问题也极易产生.

5) 目标形状变化大: 遥感图像中目标涵盖类别广泛、形状差异大, 大量长宽比异常的狭长目标 (如桥梁、道路等) 脱离检测算法适应的形状检测范围, 导致定位精度较低.

6) 目标尺度变化大: 遥感图像分辨率范围广, 不同类别的目标或者同一类别目标在不同分辨率采集下的尺寸差异悬殊, 对检测算法的尺度适应能力要求极高, 同时兼顾微小目标和超大目标的检测算法, 具有更大的难度.

7) 目标模糊: 由于受传感器和环境因素 (风雨、云雾等) 的影响, 采集到的部分遥感图像目标具有外观模糊、边界不清晰的特点, 单从目标本身特征上难以进行有效判别, 从而加大了分类和定位的难度.

8) 背景复杂: 大量遥感图像目标存在于复杂的背景中, 一方面由于场景本身比较复杂, 另一方面容易受到噪声等其他因素的影响. 当目标本身特征不明显时, 背景信息甚至比目标本身还要突出, 容易造成大量的漏检和误检问题.

## 2 基于手工设计特征的目标检测

在遥感图像领域, 传统基于手工设计特征的目标检测算法与常规图像基本类似. 传统基于手工设计特征的目标检测流程包括候选区域提取、特征提取、分类器设计和后处理. 首先采用候选区域提取方式从输入图像中提取潜在的目标区域, 之后对每个区域进行特征提取, 进而对提取的特征进行分类, 最后对得到的所有候选框进行过滤、合并等后处理, 得到最终结果. 下文针对候选区域提取、特征提取、分类器设计三个重要流程进行相关研究的介绍.

### 2.1 候选区域提取

在图像中预先挑选出感兴趣目标可能出现的区域, 舍弃出现目标可能性较低的区域, 可以去除大量无关背景, 进而减少运算量, 因此通常作为目标检测前的预处理手段. 现存遥感图像候选区域提取方法大致分为五类: 基于区域先验的方法、基于模板匹配的方法、基于特征分类的方法、基于选择性

搜索的方法和基于视觉显著性的方法.

#### 2.1.1 基于区域先验的方法

基于区域先验的方法是借鉴人类对目标所在区域的先验知识, 提取目标的潜在区域. 典型应用包含舰船检测中的海陆分离、飞机检测中的机场检测等.

海陆分离已有的研究工作大都借助港口的地理坐标信息等先验信息定位到海陆边界, 进而完成海面区域提取<sup>[6]</sup>. Antelo 等<sup>[7]</sup>采用主动轮廓方法完成海面的分离, 通过构造能量函数并进行能量最小化迭代运算, 使得轮廓曲线逐渐逼近海陆分界线, 实现海陆边界精细定位, 进而完成海陆分离过程; Liu 等<sup>[8]</sup>在主动轮廓分割法的基础上, 同时引入水平集方法, 完成海面的分离. 此外还有大量基于阈值分割的方法提取海面区域, 此类方法主要是利用海面与陆地的灰度或纹理统计特征存在本质的不同, 来进行海陆背景的区分. 例如储昭亮等<sup>[9]</sup>利用最大似然准则自动计算分割阈值, 将像素分为前景像素和背景像素, 从而实现像素级别的海陆分割判定; You 等<sup>[10]</sup>在自动计算分割阈值的基础上, 引入大津 (OTSU) 算法<sup>[11]</sup>进行粗分割, 找出最大连通的海面区域, 建立海面分布模型, 进一步分割海面和陆地.

鉴于机场中跑道的直线性特点, 机场检测研究工作的主要核心思想是对机场跑道进行检测. Liu 等<sup>[12]</sup>利用 Canny 边缘检测算子提取图像边缘特征, 后续使用霍夫变换检测直线段, 从而判断跑道位置, 进而对机场位置进行定位. Tao 等<sup>[13]</sup>改进 SIFT 特征描述子, 结合先验知识确定机场的位置. Yao 等<sup>[14]</sup>分析了现有方法逐像素判断的缺陷, 利用霍夫变换判断是否存在潜在机场, 后续采用基于显著性区域提取的方式对候选区域提取 SIFT 特征, 通过分类判断目标区域位置.

基于区域分割的方法只能针对极特殊目标, 需要大量的人工先验信息, 例如水陆分界线清晰且颜色特征差异大、机场跑道呈直线分布等, 难以应用在通用目标检测上以及其他复杂场景中, 应用范围大大受限.

#### 2.1.2 基于模板匹配的方法

基于模板匹配方法的核心思想是对感兴趣目标建立模板库, 然后在待检图像中逐像素或者逐窗口, 根据相似度量与模板进行匹配.

Xu 等<sup>[15]</sup>基于霍夫变换过程生成舰船样本形状库, 在实际图像上采用滑动窗口方法计算各个窗口区域与形状库特征相似性, 从而判断是否包含目标图像. Xu 等<sup>[16]</sup>在形状库上进行改进, 在训练方法中引入迭代最大间隔, 使得形状模型更加鲁棒, 进一步提升检测性能. Harvey 等<sup>[17]</sup>对数据集中舰船样

本进行多角度旋转, 增加模板库的多样性, 从而提升检测的泛化能力, 同时在匹配算法中, 设计了一种灰度形态学命中算法来加速匹配过程. Liu<sup>[18]</sup> 等提出由粗到精 (Coarse to fine) 的飞机检测方法, 先利用模板匹配找出飞机可能存在的区域, 再对每个区域利用主成分分析和核密度函数进行识别, 完成整个检测过程.

基于模板匹配的图像预检方法可以在特定场景中找出感兴趣目标, 然而模板的设计一般存在鲁棒性差的缺陷, 难以应对复杂多变的遥感图像场景, 同时匹配的过程采用遍历的方式, 计算效率较低.

### 2.1.3 基于特征分类的方法

基于特征分类的方法属于有监督学习算法, 基本思想是对滑动窗口进行特征提取, 然后设计分类器, 对滑动图像块进行快速评价, 最后选取得分较高的图像区域作为候选区域.

Zhang 等<sup>[19]</sup> 利用滑动窗口法在图像中生成不同大小、不同长宽比的窗口, 并提取每个窗口的视觉特征, 利用级联的支持向量机 (SVM) 对滑动窗口进行打分, 完成候选区域的提取过程. BING<sup>[20]</sup> 在此基础上将窗口提取到的特征和 SVM 分类器的权重进行二值化, 从而极大地提升了算法处理速度. Liu 等<sup>[21]</sup> 在 BING 算法基础上, 针对舰船方向性问题构造旋转矩形框空间, 并设计级联线性分类器模型对空间内的候选框进行快速打分, 完成对舰船目标潜在区域的提取过程.

基于特征分类的方法采用滑动窗口暴力搜索的方式进行区域提取, 计算效率较低, 且定位精度较差; 同时算法大幅度依赖提取的特征表征能力和分类器性能, 难以取得较高的精度.

### 2.1.4 基于选择性搜索的方法

基于选择性搜索 (Selective search) 的方法旨在搜索到图像中所有潜在目标, 从而进行候选区域的提取. 其中, 分割是最核心的处理环节, 在目标区域提取过程中利用过分割结合相似性合并的策略, 最终得到不同尺寸的所有目标范围.

如图 2 所示, 选择性搜索方法<sup>[22]</sup> 首先采用基于图 (Graph) 的分割方法<sup>[23]</sup> 进行过分割, 生成大量初始小区域, 之后通过计算相邻区域间的相似度, 将最相似的两个区域进行合并, 重复上述过程直至满足提取的候选区域预定数量. Cheng 等<sup>[24-25]</sup> 基于选择性搜索方法提取舰船的候选区域, 有效滤除了大部分背景.

选择性搜索方法采用分割方式来解决目标不同尺度的问题, 减少了滑动窗口暴力搜索带来的时间消耗; 面对复杂类别同时使用颜色、纹理等多种特

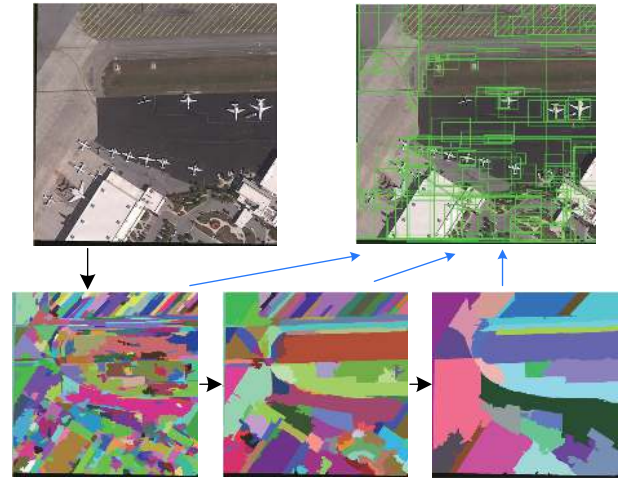


图 2 选择性搜索方法流程

Fig. 2 The process of selective search method

征进行相似度度量, 泛化能力较强. 但是算法复杂度依然较高, 难以满足实时性要求.

### 2.1.5 基于视觉显著性的方法

基于视觉显著性的方法参考人类视觉系统中的快速聚焦能力, 优先聚焦在视觉特征明显的显著性区域, 忽略无关背景, 从而缩小目标检测范围. 该类方法主要借鉴显著性检测算法来完成候选区域的提取.

相对于背景区域, 目标区域视觉特征有着明显的不同, 基于视觉特征进行显著性检测是较为经典的方法. Song 等<sup>[26]</sup> 有效利用了颜色特性, 融合方向和梯度信息来提取视觉特征, 进而进行视觉显著性的计算, 并依据显著性检测结果提取舰船候选区域.

此外, 基于频域特性进行显著性检测的方法也是较为常用的方法. 如利用傅里叶变换、余弦变换、小波变换将空域信息映射到频域, 进而在频域空间利用频域特性完成视觉显著性的检测. 例如 Zhang 等<sup>[27]</sup> 利用傅里叶变换提取飞机候选区域, Ding 等<sup>[28]</sup> 和 Xu 等<sup>[29]</sup> 利用频域特性提取舰船候选区域.

基于视觉显著性的方法在特定场景中有着不错的性能, 但是只能应用在简单背景环境下, 且目标本身的视觉特征相对明显, 例如平静水面上的舰船. 一旦目标特征不明显、背景复杂或者目标区域范围较大, 数量较多, 基于视觉显著性的方法便难以取得较高的性能.

在候选区域提取研究领域, 尽管已有的工作取得了一定的进展, 但是大部分工作采用遍历搜索的方式, 存在冗余计算, 处理速度较慢. 这些工作针对的场景有限, 只能处理比较简单的特定目标任务, 无法快速处理遥感领域复杂多变的数据, 同时定位精度较低.

## 2.2 特征提取

特征提取是遥感目标检测中最为关键的步骤, 直接影响着检测算法的性能和效率, 因此, 特征的设计一直是目标检测领域关注的重点和难点. 遥感图像目标检测常用的特征有颜色、纹理、边缘形状和上下文. 提取的用于目标检测的特征, 可以是单个视觉特征, 也可以融合多个视觉特征. 多特征融合的方式可以实现不同特征的优劣互补, 进行更全面的特征表达. 多特征融合的方式一般采用直接级联操作或者加权融合方式, 对于融合后的高维特征可选择性采取降维等处理方式提取主要信息, 减少运算量.

### 2.2.1 颜色特征

颜色特征主要是描述颜色在图像中的全局分布, 即不同颜色在整幅图像中所占的比例, 适用于对颜色敏感但对空间位置分布不敏感的目标. 相比其他特征, 颜色特征计算过程较为简单, 且不依赖目标的尺寸、方向或者图像视角的变化, 具有较强的鲁棒性. 常用的颜色特征有颜色矩、颜色相关图和颜色直方图等.

Cucchiara 等<sup>[30]</sup>在提取候选区域后, 转换到 HSV 颜色空间, 并对每个区域提取颜色特征, 来判断目标的位置信息和移动关系. Morillas 等<sup>[31]</sup>提取候选区域的颜色特征作为辅助信息, 与其他视觉特征进行结合, 来进行检测和识别. GIST 特征<sup>[32]</sup>是一种场景特征描述, 在空间包络基础上发展得到, 对颜色比较敏感, Li 等<sup>[33]</sup>利用 GIST 特征来对候选区域进行检测和识别.

颜色特征表征能力有限, 单纯凭借颜色特征很难对目标进行区分, 一般都会结合其他视觉特征来进一步提高特征表达能力.

### 2.2.2 纹理特征

纹理特征是描述图像中同质现象的视觉特征, 体现了物体表面的具有缓慢变化或者周期性变化的结构组织排列属性. 纹理特征描述了像素及其周围空间邻域像素的灰度分布, 同时也描述了特定图像区域包含的目标的表面特性. 高分辨率遥感图像目标包含了丰富的纹理结构, 不同种类的目标, 例如机场、港口、建筑物、绿地、海面等都具有不同的纹理结构, 合适的纹理特征能够有效提高目标检测性能.

Brekke 等<sup>[34]</sup>根据海面区域和海面浮油区域之间不同的纹理特性, 来做漏油检测. Selvi 等<sup>[35]</sup>利用纹理特征来进行舰船目标的检测识别.

纹理特征在一些特定场景下对特殊目标具有较

高的判别能力, 但是应用范围依然有限, 大部分检测识别问题不能通过纹理去解决.

### 2.2.3 边缘形状特征

边缘形状特征反映了目标的边缘和形状信息, 与颜色、纹理等全局特征相比, 更能反映感兴趣目标的本质特征. 为了便于目标检测, 通常要求边缘形状特征具有尺度、平移、旋转等不变性. 常用的边缘形状特征有矩特征、SIFT 特征和 HOG 特征等.

遥感目标的检测可以提取感兴趣目标的区域矩或者轮廓矩进行形状描述. 区域矩对整体区域的全局进行表述, 轮廓矩对候选区域内目标的边界进行表述. 区域矩识别区域范围大, 包含的信息量多, 有助于后续对目标进行分类, 应用广泛. Hsieh 等<sup>[36]</sup>提取目标的 Zernike 矩, 并结合小波系数、边缘和距离特征, 对各个特征进行加权融合, 提高了飞机的识别率.

SIFT 特征用于描述图像局部信息, 对尺度变化和角度变化保持不变, 同时对视角变化和噪声具有一定的适应性. SIFT 特征提取速度快, 广泛应用于目标检测. Sun 等<sup>[37]</sup>提取滑动窗口中的 SIFT 特征, 利用稀疏词袋模型 (Bag of words) 进行分类. Tong 等<sup>[38]</sup>利用 SIFT 作为舰船候选区域的特征, 进而根据提取的特征进行分类识别.

HOG 特征利用梯度或边缘的方向密度分布描述了区域内目标的轮廓和边缘信息, 弱化光照变化等环境上的干扰, 提取速度快, 特征泛化能力强, 广泛应用于遥感图像的目标检测识别, 如舰船检测识别<sup>[24, 29, 39-41]</sup>、飞机检测识别<sup>[42-44]</sup>等. Shi 等<sup>[39]</sup>提取舰船的候选区域后, 对每个区域提取 HOG 特征, 并采用 AdaBoost 分类器对候选区域进行筛选分类, 进一步提高算法的精确度. Gong 等<sup>[40]</sup>对候选区域提取多尺度 HOG 特征, 然后对多尺度特征进行融合, 进而进行分类识别. 鉴于 HOG 特征不具备旋转不变性的特点, Zhang 等<sup>[42]</sup>设计了对旋转不敏感的旋转不变性的特征, 对单个窗口进行网格划分, 提取网络内 HOG 特征, 然后基于部件模型的思想对窗口内网络特征进行融合, 生成旋转不变特征.

边缘形状特征提取速度快, 对目标的表达能力强, 应用范围广泛, 是人工特征提取的主要方式. 虽然取得了相对较好的性能, 但是距离实际应用仍差距较远.

### 2.2.4 上下文特征

上下文特征主要表征图像中不同实例之间的顺序、拓扑、相邻等空间位置关系. 上下文特征需要借助高层语义知识, 一般作为颜色、形状等其他视觉特征的辅助信息, 可进一步提高算法对目标的判别

能力。

Hinz 等<sup>[45]</sup>在道路检测中利用车辆、绿地和道路之间的上下文关系,对特征进行增强从而提升道路提取的精度。Gu 等<sup>[46]</sup>利用马尔科夫随机场来对空间目标物体之间的位置关系进行建模,从而对目标进行分类。在地物分类和检测中,大量研究<sup>[47-49]</sup>在像素基础上,扩充一个矩形窗口,以窗口的特征作为当前像素的特征表示,从而引入了上下文视觉特征。

上下文特征通常作为补充信息。在目标区域自身特征提取的基础上,结合周边的上下文信息来进一步增强区域的特征表达能力。该特征需要配合其他特征使用,无法单独用于检测。

### 2.3 分类器设计

现有主流的目标分类识别任务均采用有监督的机器学习方式,需要设计合适的分类器对目标进行分类,分类之前需要利用特征和标签对分类器进行训练,建立目标特征和目标类别之间的关系。分类器基本采用机器学习中较为流行的经典模型,如逻辑回归、支持向量机(SVM)和贝叶斯分类器等。

逻辑回归是最简单的分类器,将提取的特征映射到线性可分的空间,之后采用线性判别器完成分类过程。Corbane 等<sup>[50]</sup>提取候选区域内的 Radon 变换、小波变换等多个视觉特征,并将特征进行级联,利用逻辑回归完成目标分类和识别。

支持向量机的分类策略是将待分类样本通过核函数映射到线性可分的高维空间,并找到最优的分类超平面使得支持向量在特征空间上的间隔最大。Gong 等<sup>[40]</sup>设计了一个判别式的可见部件模型,利用隐支持向量机训练模型参数,缓解了遥感目标多尺度问题和旋转问题。Zhu 等<sup>[51]</sup>提取不变矩多维特征,对多个矩特征进行组合优化,利用 SVM 分类器进行分类识别,有效克服了单一特征鲁棒性不强的缺点。Wang 等<sup>[52]</sup>也利用 SVM 分类器对提取到的特征进行分类和识别。

贝叶斯分类器采用后验概率来进行判断,确定概率最高的类别为检测结果。Antelo 等<sup>[7]</sup>提取矩特征,使用线性判别分析对特征进行降维,最后用贝叶斯分类器完成候选区域分类。

分类器的设计相对成熟,特征提取的性能对分类结果具有重要影响,因此现有研究更倾向于增强特征表达能力,从根本上提高目标判别性。

虽然传统的基于手工设计特征的目标检测识别研究精心设计了每个环节,在特定的应用场景下取得了不错的效果,但仍存在较多的问题和局限。首先,传统基于候选区域提取的方式只能粗略提取水

平矩形区域,定位精度差;其次,人工设计的特征提取方式大多利用底层视觉信息,特征表述能力弱,无法准确表达图像中复杂的高层语义信息,类别间通用性差;最后,提取的候选区域数量大、冗余计算量大、时间复杂度高,难以满足实时系统的应用需求。综上,传统方法的性能还远远无法满足实际应用的需求。

## 3 基于深度学习的目标检测

相对于手工特征提取,深度学习直接将候选区域提取、特征学习和分类器整合在一起,实现了端到端的检测。深度学习在海量数据驱动下学习到的特征具有更强的语义表征能力,性能相对于手工提取方式得到大幅度的提升,同时在神经网络前向传播过程中避免了大量窗口的冗余计算,提升了检测速度。

下文首先简要概括典型的目标检测模型。后续针对遥感图像的特点,详细梳理了一系列改进方案,使其能够有效应用在遥感图像的检测上。

### 3.1 典型目标检测模型

基于深度学习的目标检测方法有两种分类标准,依据是否需要进行感兴趣区域(Region of interest, ROI)提取,划分为双阶段检测(Two-stage)和单阶段检测(One-stage);依据是否需要预设锚点框(Anchor),划分为基于锚点框的检测(Anchor based)和基于关键点的检测(Point based),基于关键点的检测也称为 Anchor free 检测。

双阶段的目标检测分两个阶段完成整个检测流程。首先要提取感兴趣的区域,之后针对每个区域进行进一步的检测和识别,典型方法代表为 Faster R-CNN<sup>[53]</sup>。双阶段的目标检测方法取得了较高的精度,然而由于需要首先提取感兴趣区域,并对每个区域分别进行分类和回归,增加了额外的计算量,速度上不够快,对于实时性要求较高的系统难以应用。单阶段的目标检测在一个阶段完成整个检测过程,典型方法以 YOLO<sup>[54]</sup>和 SSD<sup>[55]</sup>为代表。单阶段目标检测速度较快,基本满足实时系统的要求,但是检测精度略低于多阶段目标检测方式。

大部分检测方法需要提取锚点框,以锚点框为初始检测框,进行进一步的微调。通过回归真实框和锚点框的中心点位置偏移,以及对应宽、高的缩放比例,调整锚点框的位置、形状和尺寸,使其逐渐与真实框重合。基于锚点框检测的优势是网络输出值都是在锚点框基础上的相对值,值域范围小,训练比较容易,收敛速度快。然而由于检测过程需要

基于锚点框进行匹配, 锚点框的设计需要针对不同任务进行大量的人工干预, 针对特定任务需要具备一定的先验知识, 调参过程比较繁琐; 一些长宽比罕见的目标难以匹配上, 进而导致漏检; 同时大量的锚点框也会导致内存占用大、时间复杂度高等问题. 针对锚点框检测带来的问题, 近期基于关键点的检测方式<sup>[56-58]</sup>逐渐流行. 该类方法直接基于像素层面进行目标的分类和回归, 避免了锚点框的引入, 缓解了锚点框带来的一系列问题. 基于关键点的检测方法免除了人工繁琐的锚点设计过程, 避免了检测过程中重合度 (Intersection over union, IoU) 计算等匹配过程, 减少了运算量, 并取得了较高的精度, 目前已经成为了当前研究的热点.

现阶段几乎所有的基于深度学习的目标检测方法均是基于上述典型模型进行扩展的, 这些模型的发展变化也反映出当下目标检测整体的发展趋势. 尽管这些典型方法在自然图像上取得了不错的效果, 却无法直接应用在遥感领域. 遥感图像和常规自然图像相比, 有其独有的特点, 需要针对性进行算法的改进, 下文将具体阐述针对遥感领域的目标检测算法改进.

### 3.2 遥感目标检测模块改进

本节针对遥感图像存在的特殊问题分别进行分析, 对不同的改进算法进行归类 and 总结.

#### 3.2.1 针对超大图像尺寸的改进

一般情况下自然图像处理尺寸不会过大, 即便高清的图像, 宽高像素数很少会超过 5 000. 在尺寸较大的自然图像上进行目标检测时, 适当对原图进行缩放, 待测目标的特征也不会弱化太多, 绝大多数目标都可以正常进行检测. 然而在遥感图像上却有所不同, 遥感图像具有超大的图像尺寸, 宽高像素数可达数万级别, 直接进行目标检测需要过大的内存空间, 同时计算量过大, 现有硬件还不足以支持; 若将其缩放到合适的尺度, 则会因为缩放比例因子过小, 导致多数目标在缩放过程中“消失”, 从而造成较高的漏检率. 因此现有目标检测方法很难在保证检测质量的同时, 保持较快的速度.

为了处理图像尺度过大的问题, 最为常用的方法是采用分块切割的方式, 将大图切割成小图, 再分别对每张小图进行后续的检测识别, 最后把所有的检测结果进行拼接, 完成全图的最终检测<sup>[59-65]</sup>. 这种分块的方式避免了整张图像送入网络输入中带来的内存过大问题, 同时保持原有的目标尺度, 提高了检测效果, 然而该方法会将处于分块边缘的目标切割, 从而将完整目标一分为二, 影响定位精度. 为

了解决这个问题, Wang 等<sup>[66]</sup>以一定的重叠率对原图进行切割, 从而增加了分块小图边缘目标完整性的可能, 但是却增加了子图像的数量, 使得大图的处理时间变得冗余, 同时依然无法避免目标检测结果一分为二的问题. 为了减少大图检测的时间冗余, 基于图像中绝大多数区域为背景、待测目标只集中于较小区域的现状, R<sup>2</sup>-CNN<sup>[67]</sup>设计了轻量级主干 Tiny-Net 来进行特征提取, 并采用先判断、后定位的方式, 将不含目标的子图像块进行滤除, 从而减小后续检测识别过程的计算负担, 该方法保证了检测效果, 同时提高了检测效率, 但是依然会对重叠区域进行多次检测. 算法 YOLT<sup>[68]</sup>对切割后的子图, 采用多个轻量级模型进行检测, 并将检测结果进行融合, 保持检测精度的同时, 一定程度上提升了检测的速度. 以上方法均需要对原始遥感图像进行分块处理, 这种切割的方式始终不是一种最优的方式, 但目前来看很难通过非切割的办法进行完美的检测.

Zhang 等<sup>[69]</sup>在直接对原图进行处理的方向上迈进了一小步, 该方法将整张原图作为输入, 并基于像素级别进行分类, 对置信度较高的像素, 以该像素为中心提取候选区域, 并经过非极大值抑制过程进一步减少重合度高的区域, 最后基于每个单独的候选区域完成检测识别. 这种方式有效减少了切割分块过程带来的冗余计算量, 但是同时也存在较大的问题: 原始图像仍然不能过大, 在过大图像尺寸下这种方法同样无法应用; 候选框大小是基于待测目标的先验尺度信息和形状信息得到的, 面对尺度、形状多变的多类目标检测任务时, 则难以进行有效处理.

综上所述, 现阶段对超大尺寸的遥感图像检测, 图像分块结合快速过滤的方式是当前从速度和精度上最优的处理方式.

#### 3.2.2 针对目标方向多样化的改进

相对于自然图像, 遥感图像均是俯视视角拍摄得到的, 方向性问题更加突出. 同一类型的不同目标经常以多样化的方向在一张图像上分布, 而经典的卷积神经网络不具备旋转不变性, 提取的特征对旋转极其敏感, 同一个目标在不同方向下提取的特征完全不同, 给方向多样化的目标检测带来很大难度.

针对目标方向多样化问题, 有效的改进思路有两种: 数据扩充和增加旋转不变子模块.

数据扩充方式是对原有训练数据, 分别旋转多个不同的角度, 进而从角度上进行数据扩充, 以增强卷积神经网络对角度的适应能力. 现阶段大多数遥感图像目标检测方法<sup>[60-64]</sup>均默认对训练数据进行

角度扩充, 有的方式直接将扩充多个角度后的图像数据合成一个更大的数据集进行训练; 有的方式直接在训练过程中对单幅图像采用随机旋转处理. 数据增强方式在一定程度上提升了样本角度的多样性, 对处理多方向目标检测问题起到一定作用, 但是基于旋转扩充的数据增强方式起到的作用有限, 并不能从根本上解决卷积神经网络对旋转敏感的问题.

为了从根本上解决目标检测中角度多样性的问题, 现有研究通过在特征提取过程中增加对旋转不敏感的模块, 来对不同角度的目标进行角度归一化, 进而提高特征表示在不同角度下对各个类别的鲁棒性. RICNN<sup>[25]</sup> 模型优化一个新的目标函数, 通过引入正则化约束, 使样本的特征表示在旋转前后相互接近, 实现旋转不变性. Cheng 等<sup>[70]</sup> 通过优化一个新的目标函数显式地在 CNN 特征上添加了一个旋转不变正则化器, 强制旋转前后训练样本的特征表示紧密映射, 以实现旋转不变性. ORSI<sup>[71]</sup> 检测器采用了一种新颖的空频信道特征 (SFCF), 综合考虑了频域内构造的旋转不变信道特征和原始的空间信道特征来应对旋转问题. 以上方法通过引入旋转不变子模块, 减弱了卷积神经网络特征提取对角度的敏感性, 从根本上缓解了方向多样化问题对检测带来的影响.

### 3.2.3 针对目标尺度过小的改进

与自然图像相比, 遥感图像中小目标的数量规模更大. 小目标的总像素数只有几十到几百个像素, 随着卷积神经网络的下采样, 特征图尺度不断减小, 小目标会在多次下采样中, 从深层特征中消失, 进而造成很大程度的漏检. 小目标检测一直是现有深度学习算法中的难点, 目前仍有大量工作针对小目标检测做改进. 最直观的方法是提高特征图的尺寸, 尺寸的提高主要包括两种方式: 引入浅层特征和深层特征上采样.

针对浅层特征引入的改进, Yang 等<sup>[60]</sup> 通过增加浅层特征金字塔的数量和规模来提升小目标的检测精度, 并采用稠密连接的结构增强小目标的特征表达能力. YOLT<sup>[68]</sup> 增加上采样, 并通过恒等映射将中间浅层特征与输出特征进行拼接, 同时增加预测网格的个数, 从而提高小目标的适应性, 并融合多个尺度来寻找更小的目标. Chen 等<sup>[72]</sup> 在最浅层特征上结合语义信息进行增强, 并与深层特征进行融合, 提高小目标检测率. Wang 等<sup>[73]</sup> 在结合浅层信息后, 同时改进损失函数来增加小目标的训练权重. Li 等<sup>[63]</sup> 利用反卷积层进行浅层特征和深层特征的融合, 在已有检测精度下, 进一步增强小目标的

检测能力. 上述引入浅层特征的方式, 对于小目标会引入过多的背景噪声. 为了减少噪声的影响, Fu 等<sup>[62]</sup> 在进行浅层特征和深层特征融合时, 采用一个平衡因子来平衡特征融合的权重, 但是对于因子的取值只能凭借检测任务的先验, 面对不同任务鲁棒性较差.

深层特征上采样主要以反卷积的形式实现的. Zhang 等<sup>[74]</sup> 在两阶段 Faster RCNN 基础上, 对前一阶段提取的每个候选区域通过反卷积进行上采样, 放大特征图尺寸, 改善小目标检测效果. Schilling 等<sup>[64]</sup> 利用反卷积层扩大深层特征的尺度, 融合浅层特征共同完成检测过程. 然而基于反卷积的运算会引入更多的参数. 为了减少额外的参数量, Liu 等<sup>[75]</sup> 采用膨胀卷积操作来代替卷积, 在相同感受野的前提下, 减少了参数, 然而膨胀卷积会造成局部信息的丢失. 为了解决这个问题, Ying 等<sup>[76]</sup> 使用像素注意力机制来进行局部信息的融合, 弥补了膨胀卷积的不足, 改善了小目标检测问题. 然而, 只有在深层特征中依然存在小目标的前提下, 上采样操作才有意义. 如果本身小目标在深层特征中因为降采样而“丢失”了, 上采样也并不能挽回丢失的特征信息. 因此也有一部分工作<sup>[64]</sup> 同时结合引入浅层特征以及深层特征上采样过程, 实现优势互补, 但是随着特征图尺度的增加, 计算复杂度也有了一定程度的增加, 在增强小目标检测能力的同时, 增加了检测过程的时间消耗.

### 3.2.4 针对目标密集分布的改进

遥感图像在一些特定场合下目标分布过于密集, 例如停车场中的车辆、港口中的船只等. 在常规视角下的自然图像中, 也存在密集分布的问题, 例如拥挤街道中密集行人检测等. 相对于自然图像的密集检测, 遥感图像反而相对容易, 因为在俯视视角下, 较难出现同类目标相互遮挡的问题, 因此无需考虑遮挡因素. 遥感图像中密集检测的困难因素主要在于密集分布的目标尺度较小, 因此很大程度上, 可以借鉴上述小目标检测的改进方案. 除此之外, 对于密集分布的目标, 提取到的特征极容易出现边界模糊、高密度目标特征聚集的问题, 从而导致漏检、误检和定位错误. 现有研究大多通过特征增强方式来提升单个目标的区分度.

Long 等<sup>[77]</sup> 将多种传统方法和深度学习方法进行组合, 可以有效处理高密度, 小尺寸的物体在复杂背景下的检测, 但是流程极其复杂. DAPNet<sup>[78]</sup> 通过判断图像中目标的分布密度, 来进行自适应的区域生成策略. Long 等<sup>[79]</sup> 提出了一种基于无监督分数的边界盒回归 (USB-BBR) 算法, 并结合非最



大抑制算法对检测到的目标区域的边界盒进行优化, 提高密集分布目标的定位精度. SCRDet<sup>[80]</sup> 增加了像素注意力机制和通道注意力机制来同时实现小目标和密集目标的特征增强. R<sup>3</sup>Det<sup>[81]</sup> 通过增加一个特征图微调模块, 尽量保证目标的特征和目标的中心对齐, 提高密集目标的定位精度. SCRDet++<sup>[82]</sup> 借助语义分割网络的中间特征来指导检测识别的特征提取过程, 间接使用注意力机制来进行密集目标的特征增强, 从而提升密集目标的边界区分程度.

上述特征增强手段, 都引入了额外的运算模块, 提升了算法的复杂度, 在牺牲算法效率的基础上, 提高了对于密集目标检测和定位的精度. 现有算法大多需要根据实际需求对性能和效率进行权衡.

### 3.2.5 针对目标形状差异大的改进

遥感图像相对于常规图像, 不同类之间的目标形状差异过大. 例如车辆、储油罐、飞机等基本接近常规形状矩形, 但桥梁、道路、船只等目标过于狭长, 有着极其异常的长宽比. 面对形状的多样化, 容易因为锚点框预设不合适, 造成漏匹配问题, 从而导致漏检. 现有较多的方法是检测阶段提高锚点框的种类和数量, 以进行不同目标形状的适配. 同时还可以采用可变形卷积网络和基于关键点的检测模型来提高对形状的适应能力.

大量研究直接通过增加预设锚点框的种类, 例如增加不同的尺度、不同的长宽比, 甚至增加不同角度的锚点框来粗暴地提升算法对目标形状的泛化能力<sup>[77, 81, 83-86]</sup>. 尽管取得了明显的效果, 但是对于形状过于狭长的目标, 很小的角度偏差就会导致重合度的大幅度下降, 极容易导致漏匹配, 依然难以从根本上解决形状差异的问题; 同时大规模增加锚点框的数量, 也会带来额外的计算复杂度提升以及额外的内存占用. 现阶段依然是以牺牲效率为代价换取对不同目标形状的适应性.

此外, 少数研究<sup>[87-88]</sup> 在锚点框基础上引入可变形卷积模块 (Deformable convolution) 来自动适配不同的目标形状, 由于可变形卷积模块引入了大量的参数, 增加了训练难度, 因此目前并没有明显的效果提升; 还有一些方案<sup>[89-90]</sup> 引入形状注意力模块来尽可能保持特征对形状的不变性, 然而不能从根本上解决锚点框的漏匹配问题.

另一种有潜力的解决方式是最近兴起的基于关键点的检测方式, 直接省去了锚点框与目标匹配的过程, 而是基于像素直接对形状进行预测, 从而解决了形状差异大的问题. 目前已经有了有一定数量的基于关键点的检测算法<sup>[91-96]</sup> 对遥感领域目标检测进行了初步尝试, 从精度来看, 基本和基于锚点框检

测器的精度持平, 还有较大的潜在优化空间. Pan 等<sup>[97]</sup> 在基于关键点检测的基础上, 增加了特征选择模块和动态增强检测头, 使得检测能够进一步对目标形状进行自适应调整, 但是精度的提升依然很小, 不能从根本上解决形状差异带来的影响.

### 3.2.6 针对目标尺度变化大的改进

遥感图像中目标的尺度变化范围相对于自然图像更大, 包括不同种类目标在同一分辨率下的尺度差异, 例如车辆、船只、飞机等尺度较小, 操场、球场、机场等尺度过大; 同时还包括同类目标在不同采集分辨率下的尺度差异 (由于遥感图像本身采集的分辨率范围比较大, 不同分辨率下同一目标尺度相差也较大). 算法检测性能同时对大尺度和小尺度目标均保持较高的精度, 是一个较难的问题. 目前的研究主要通过引入多尺度信息来解决目标尺度变化大的问题.

Guo 等<sup>[98]</sup> 和 Zhang 等<sup>[99]</sup> 直接采用多尺度的候选区域网络和多尺度的检测器来检测不同尺度的目标. Zhang 等<sup>[69]</sup> 将第一阶段提取的候选区域调整为三个不同尺寸图, 并分别使用三个网络进行并行预测, 最终通过投票机制确定类别, 增强了不同尺度的适应能力. Zhang 等<sup>[100]</sup> 在已有 Faster R-CNN 基础上, 在不同的特征图上并行引入不同的候选区域提取网络, 来独立对不同尺度目标进行检测. Wang 等<sup>[66]</sup> 使用残差连接的自编码器多尺度结构来适应不同尺度的变化. Hou 等<sup>[101]</sup> 采用多个并行的 ROI Pooling 模块将特征映射到不同的尺度, 并采用多阶段级联的方式进行检测. 之前介绍的小目标改进模块中, 通过融合浅层特征和深层特征, 在提升小目标检测精度的同时, 尽量不影响常规尺寸目标的检测, 也间接提升了对尺度的泛化能力.

在多尺度信息特征融合中, 特征金字塔 (Feature pyramid network, FPN)<sup>[102]</sup> 通过融合底层视觉特征和高层语义信息, 并在不同尺度特征上分别进行预测, 在常规图像中获得了广泛的应用. Yang 等<sup>[60, 83]</sup> 和 Zou 等<sup>[103]</sup> 基于 FPN 采用稠密特征金字塔网络进一步加强不同尺度特征之间的联系. Fu 等<sup>[104]</sup> 在 FPN 自顶向下后额外又加入自底向上的连接, 将底层特征信息和高层视觉信息进行了二次融合, 从而进一步增强特征表达能力. 还有大量研究<sup>[84, 89, 105-109]</sup> 直接对特征融合模块的网络结构进行改进, 从而增强特征表示. Yan 等<sup>[110]</sup> 在多尺度检测基础上, 平衡不同尺度目标对于损失函数的训练权重, 加强算法对不同尺度的鲁棒性.

多尺度信息的引入有效解决了不同尺度目标的泛化能力, 但是对于极端尺度的目标, 例如数个像

素到数十像素的微小目标、接近全图尺寸的几十万像素大目标, 依然难以取得较好的效果. 为了缓解这个问题, 现有大多数算法在训练方式上, 采用图像金字塔进行多尺度训练和多尺度测试, 进一步降低漏检率, 提高检测精度, 但是增加了大量的计算开销.

### 3.2.7 针对目标外观模糊的改进

遥感图像在成像过程中, 受环境因素影响, 容易受到云雾遮挡、以及雨雪天气的干扰. 虽然现有针对遥感图像进行预处理的方式 (例如去雾等) 来滤除干扰信息, 但是难免会降低图像质量, 尤其对于本身较弱、较小的目标, 容易造成目标外观模糊的问题. 单纯针对外观模糊的目标本身, 提取的特征表征能力也较弱, 很难进行精确的目标检测, 因此可以结合区域上下文信息来进一步强化特征, 提高分类准确率.

现有很多基于区域上下文信息来提升遥感图像目标检测性能的方法. Li 等<sup>[111]</sup>提出了一种双通道特征融合网络, 可以沿着两个独立的路径学习局部属性和上下文属性, 并将特征进行融合增强判别力. Chen 等<sup>[112]</sup>以 FPN 为基础提取 ROI 特征图, 然后将整幅图的特征图和 ROI 的特征图融合, 得到上下文信息完成检测. Ma 等<sup>[113]</sup>设计了一种融合局部上下文特征和目标-目标关系上下文特征的上下文信息融合子网络, 来处理目标对象外观的复杂性. CAD-Net<sup>[108]</sup>通过学习目标的全局上下文特征和局部上下文特征来增强目标的特征表达, 同时引入注意力机制来集中检测区域目标特征.

基于区域上下文的方法通过结合目标之外的其他信息以及图像的全局信息进行辅助, 增强了特征表示, 在一定程度上缓解了目标外观模糊的问题, 是成像质量不高的一种有效的处理方式. 然而只能处理与周边环境有一定共生关系的目标, 相对独立的目标较难处理, 且提升的幅度有限. 针对目标外观模糊的问题, 最有效的办法依旧是提高成像质量和改进图像预处理手段, 从根源上避免模糊目标的产生.

### 3.2.8 针对复杂背景的改进

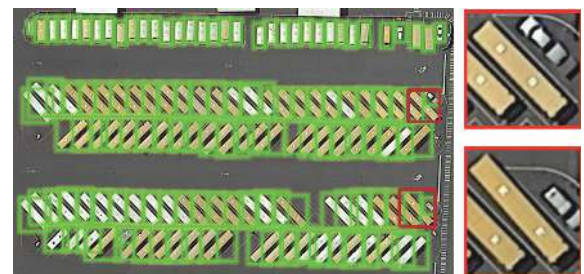
遥感图像覆盖范围广, 待测目标类别有限, 因而背景区域占据了绝大部分. 复杂背景中的目标, 难免容易与背景产生混淆, 导致目标的漏检以及背景区域的误判. 例如海上高速行驶的船只伴随的尾迹容易误判成船只, 大风大浪影响下的小型船只容易漏检, 坐落在沟渠中的桥梁难以从复杂背景中区分成漏检等. 针对复杂背景的改进方式, 主要考虑如何弱化背景、同时增强目标信息的特征.

注意力机制是对弱化背景、增强目标信息很有效的处理方式. 在船只检测中, 为了避免背景中船坞带来的影响, Zhang 等<sup>[114]</sup>采用语义分割的思想分别提取船只、海洋和陆地的特征掩膜, 之后采用像素注意力机制来对各个类别进行加权, 提升海陆分离的能力, 但是这种基于先验知识的方式不具备普适性. Li 等<sup>[115]</sup>提取多尺度特征后, 对每个特征图单独使用注意力机制进行特征增强, 从而消除背景噪声的影响. 此外还有大量研究<sup>[80, 82, 89, 105, 107, 116-117]</sup>也是对融合后的多尺度特征图采用注意力机制来弱化背景, 突出目标区域. Li 等<sup>[118]</sup>在特征图融合过程中设计了显著性金字塔融合策略来弱化背景带来的影响, 增强目标特征的表达能力, 并结合全局注意力机制增强语义信息, 完成最终的检测过程.

注意力机制较好地解决了复杂背景带来的问题, 但是引入了额外的掩膜计算, 造成了网络计算量的增大.

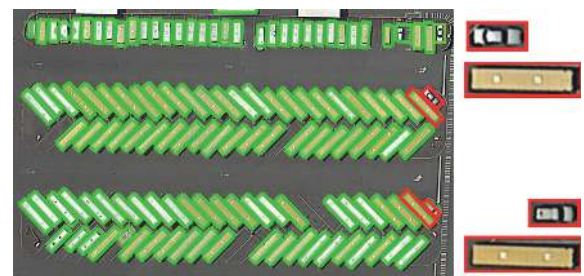
### 3.2.9 针对常规水平框定位粗糙的改进

相对于常规角度图像, 遥感图像目标具有方向性, 且方向具有随机性. 上述介绍的常规的水平框检测方式在处理方向性目标时, 难以进行紧凑、精准的定位, 尤其当数个狭长的目标 (货车、舰船) 以倾斜的角度密集排列时, 水平框检测方式更容易产生混淆 (如图 3 (a) 所示). 因此常规的水平框定位



(a) 水平框检测效果

(a) Results of horizontal bounding box



(b) 旋转框检测效果

(b) Results of rotated bounding box

图 3 水平框检测与旋转框检测对比

Fig. 3 Comparison of horizontal detection results and rotated detection results

方式对于遥感图像目标检测并不是最优的处理手段. 为了进一步提高目标检测的定位能力, 遥感领域方向性目标检测逐渐受到关注. 利用旋转框替代水平框的表示方法, 设计旋转框检测模型, 用更加紧凑的包围框来进行方向性目标的检测, 对定位能力的提升有着重要的价值 (如图 3 (b) 所示).

如图 4 所示, 常见的旋转框参数表示方案有两种, 分别是五参数法和八参数法. 由于五参数法参数量更少, 基于五参数法的研究更多. 五参数法主要包括目标中心点位置坐标, 以及目标宽度、高度和角度五个参数, 用来表征旋转矩形信息, 其中对于角度的周期一般选取  $180^\circ$  或者  $90^\circ$ ; 八参数法直接回归旋转框的四个顶点坐标, 因此可以进行任意四边形的回归, 回归更加灵活但是参数量更多.

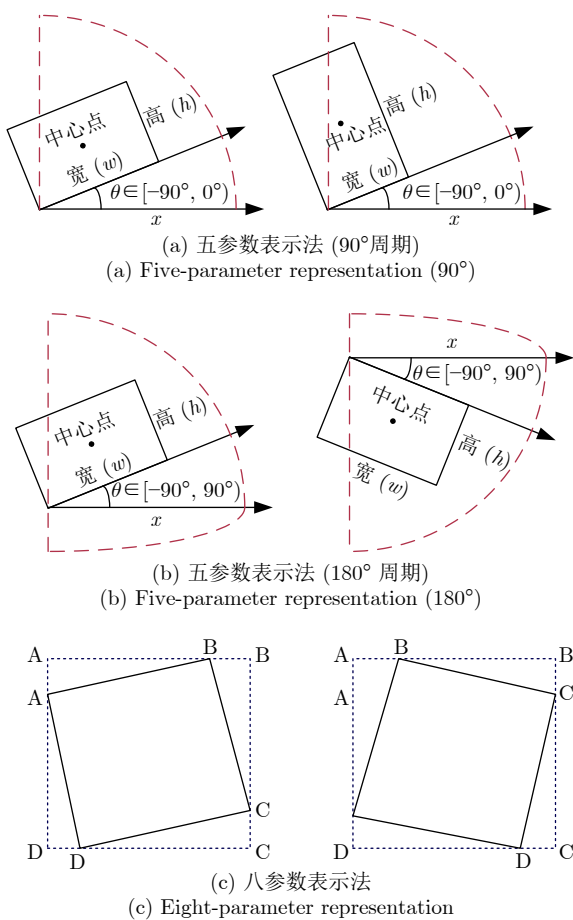


图 4 旋转框参数表示方案  
Fig. 4 Parameter representation of rotated boxes

现有旋转框目标检测架构大多基于上述参数表示方案, 多数方法基于已有的水平框检测架构做了针对性改进. 文献 [119] 直接在 Faster R-CNN 的基础上基于上述参数表示方案引入新的回归参数, 来进行旋转框的检测. RRPN<sup>[120]</sup> 人工定义了大量带

有角度的锚点框来进行旋转目标的回归, 同时在检测识别阶段提出了旋转特征池化操作 (RROI pooling) 来将特征进行归一化, 取得了方向性目标检测的初步效果. R<sup>2</sup>CNN<sup>[121]</sup> 依然选取水平锚点框来进行第一阶段感兴趣区域的提取, 在 (Region proposal network, RPN) 阶段生成旋转候选区域并采用多尺度池化操作来增强检测识别特征的泛化能力, 在第二阶段基于水平候选区域来进行旋转框的回归, 进一步提升了性能并减小了内存消耗. RoI transformer<sup>[86]</sup> 在 RPN 和 RCNN 之间插入了一个轻量级模块, 将 RPN 生成的水平区域转换成旋转区域, 尽可能减少旋转目标检测带来的计算复杂度. 为了进一步增强特征表达能力, 上述方法<sup>[86, 120-122]</sup> 同时进行水平框和旋转框的回归, 二者相互促进, 共同提升精度.

### 3.2.10 针对旋转框边界突变问题的改进

基于常见的参数表示进行旋转框坐标回归时候会伴随典型的边界突变问题. 边界突变问题是旋转框的独有特点, 由于同一旋转框在一种参数表示方案下会有多种表示值, 进而通过强加规则将具体表示值限制在特定范围, 使得同一旋转框在一种参数表示方案下只对应唯一的参数表示值, 导致了所在范围的边界上的表示值产生了突变, 影响边界值的检测精度.

不同参数表示方案的边界突变问题示意图如图 5 所示, 对于五参数表示法, 如果角度周期为  $180^\circ$ , 边界值  $0^\circ$  和  $180^\circ$  实际重合, 却在数值上有 180 的差距, 产生了突变, 影响角度的预测精度; 如果周期为  $90^\circ$ , 在边界上除了角度产生突变以外, 宽和高也会在边界发生交换, 产生突变, 因此周期为  $90^\circ$  的时候产生的突变更为复杂, 角度和形状的预测都会受到影响. 如果是八参数表示法, 根据四个顶点的起止顺序不同会有四种表示方式, 通过人为设置规则定好起始点可以确定唯一表示方式, 随之也产生边界问题, 在边界上起始点发生转移, 产生突变, 影响边界上顶点位置的预测. 总之, 边界突变问题对旋转框坐标回归精度带来了极大的负面影响.

现有研究人员针对旋转框回归的边界突变问题, 做了针对性改进, 主要改进方式包括两类: 采用新的参数表示方案和改进损失函数.

对于新的参数表示方案, APE<sup>[94]</sup> 采用一对具有周期性的三角函数来表征角度信息, 避免了边界突变. Gliding vertex<sup>[123]</sup> 首先提取外接水平框, 进而在水平框的边界上引入一个比例因子对旋转框顶点进行定位, 解决了八参数表示方案顶点顺序的问题, 但是在待测目标是水平框时候依然存在边界问题, 为了缓解边界问题引入了一个控制因子来决策是否

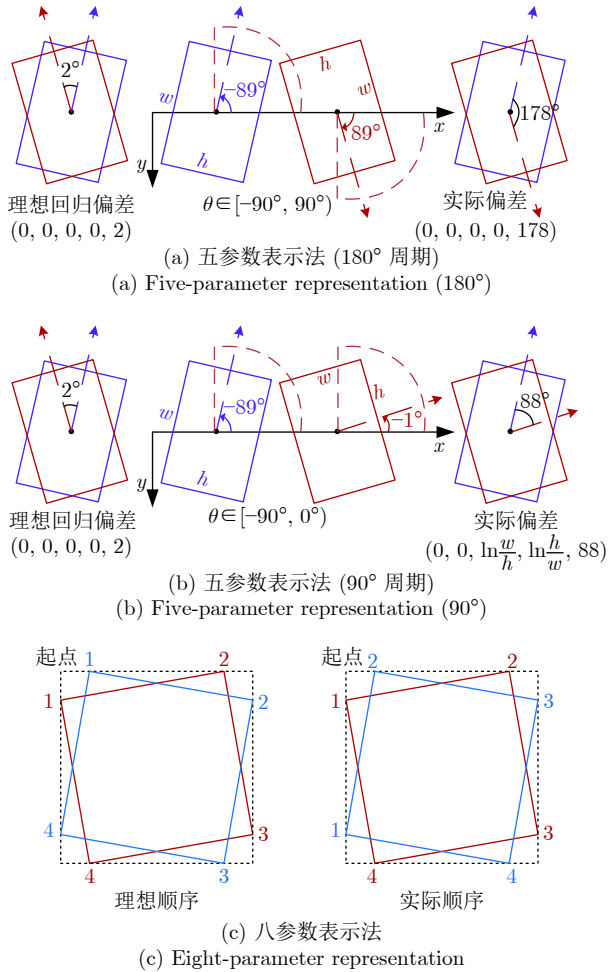


图 5 边界突变问题示意说明

Fig. 5 Illustration of boundary mutation

需要回归旋转框. SARD<sup>[124]</sup> 直接对网络输出结果进行了强制的标准化, 尽管发现了边界问题, 但是解决方式只限于强加规则, 并没有有效缓解该问题. Mask OBB<sup>[105]</sup> 和 RADet<sup>[107]</sup> 借鉴实例分割的方式, 基于分割产生的类别掩膜来生成最小外接矩形, 避免了边界问题, 但是引入了额外的语义分割计算, 计算复杂度较高. CSL<sup>[125]</sup> 采用角度分类的方式来替代回归, 并针对性设计了软标签来应对边界问题, 从根源上解决了边界问题, 取得了很好的效果, 但是采用角度分类的方式也导致了输出参数量过大.

对于损失函数的改进, SCRDet<sup>[80, 82]</sup> 设计了 IOU-Smooth 损失函数直接弱化训练时候的边界样本. Modulated loss<sup>[126]</sup> 在边界上采用所有突变值计算损失函数并选取损失最小值进行训练, 来缓解边界问题的影响. 基于改进损失函数的方法针对边界问题进行了很大程度的缓解, 但是没有从根源上完全解决边界问题.

综上, 针对遥感图像本身存在的特殊问题和挑

战, 对不同的改进算法进行了细致的分类和总结. 但有两点需要注意:

首先, 各个特殊问题并不是独立存在的, 多数问题之间具备很强的相关性, 甚至是共生的. 例如: 超大图像尺寸和覆盖面积的检测中, 目标相对尺寸明显减小, 此时和小目标检测便有了联系; 多数小目标在图像中是密集分布的, 同时容易受到成像因素影响导致外观模糊, 也更容易受到复杂背景的干扰.

其次, 针对各个问题的解决方式很大程度上是共通的. 例如: 注意力机制可以解决多个特殊问题; 小目标检测性能的提升也会间接提高目标检测算法对不同尺度的泛化能力; 旋转框检测方式的创新性引入, 直接将卷积神经网络在特征提取过程中对旋转敏感的劣势化为优势, 从而可以利用旋转敏感性进行更精确的旋转框回归; 旋转框紧凑包围目标的优势也在一定程度上缓解了密集分布目标检测困难的问题; 大多基于水平框的改进方式也可直接应用在旋转框检测的特征增强中.

因此, 基于不同问题的改进是可以相互借鉴的, 不能割裂开来, 而应统一成一个整体.

## 4 数据集和算法性能比较

手工特征提取算法性能的验证, 分类器的训练以及深度学习方法的训练测试, 均需要大量数据作为支撑. 本章首先按照时间顺序回顾了一下遥感图像目标检测常用数据集, 之后简单介绍了普遍使用的评测指标, 并基于两个常用数据集对现有遥感图像通用目标检测算法的性能进行比较.

### 4.1 遥感图像目标检测数据集

按照时间顺序, 现有遥感图像目标检测数据集包括: TAS 数据集<sup>[127]</sup>、SZTAKI-INRIA 数据集<sup>[128]</sup>、NWPU VHR-10 数据集<sup>[129]</sup>、VEDAI 数据集<sup>[130]</sup>、UCAS-AOD 数据集<sup>[131]</sup>、DLR 3K Vehicle 数据集<sup>[132]</sup>、HRSC2016 数据集<sup>[133]</sup>、RSOD 数据集<sup>[79]</sup>、DOTA 数据集<sup>[119]</sup> 和 DIOR 数据集<sup>[4]</sup> 等. 随着时间推移, 数据集规模越来越大, 检测类别数越来越多, 标注框涉及水平框和旋转框, 检测任务越来越复杂, 实例数越来越多, 逐渐接近实际复杂场景. 除了使用这些数据集之外, 还有一些研究人员自己从 GoogleEarth 收集数据或者购买 QuickBird 等卫星数据.

### 4.2 通用目标检测算法性能比较

现有的遥感图像中最常用的评测指标是精确度 (Precision) 和召回率 (Recall). 精确度反映所有被检测出来的目标中, 检测正确的概率; 召回率反映所有待检测目标中被成功检测到的概率, 二者是

此消彼长的关系. 以召回率为横坐标、精确度为纵坐标画出的曲线称为精度-召回率曲线 (Precision-recall curve), 简称 PR 曲线. 曲线下对应的面积称为平均精度 (Average precision), 简称为 AP, 用来表征单一类别的检测性能. 每个检测类别都对应一个 AP 值, 所有类别的平均 AP 值即 Mean average precision, 简称为 mAP, 用来评测算法在数据集整体上的性能.

为了比较算法在复杂场合下的表现和对多类目标的检测鲁棒性, 本章节主要比较遥感图像通用目标检测算法的性能. 鉴于传统的基于手工设计特征的方式处理的目标比较单一, 处理的任务比较简单、特殊, 算法适应性和鲁棒性很差, 因此不在本章考虑范围内; 对于检测单类目标的深度学习算法, 亦不进行考虑. 此外, 针对常规水平框定位粗糙、基于旋转框定位更加精细的现状, 本章同时对比了水平框检测性能和旋转框检测性能. 针对水平框检测, 选取 NWPU VHR-10 和 DOTA 两个数据集进行测试, NWPU VHR-10 类别数较少, 任务相对简单, 是早期深度学习算法广泛使用的数据集; DOTA 数据集更加复杂, 检测难度更大, 是现阶段的主要评测数据集, 检测算法性能对比结果如表 1 所示. 针对旋转框检测, 也选取 DOTA 数据集来测试, 性能对比结果如表 2 所示 (注: Res-ResNet, ResX-ResNeXt, HG-Hourglass).

通过对比表格中的数据进行分析后可以看出:

1) 面对较简单的检测任务, 例如 NWPU VHR-10 数据集, 当前算法已经达到极高的精度, mAP 在 90 以上, 性能基本饱和, 后续提升空间较小; 然而在处理更复杂和更有挑战性的问题时, 例如 DOTA 数据集, 性能相对理想情况还有较大的差距, 很多问题目前还无法得到有效解决, 现有研究成果还无法满足高标准应用场合需求.

2) 针对遥感图像目标自身特点进行的一系列改进是有效的, 相对于基本检测模型 Faster RCNN<sup>[53, 119]</sup>、R-P-Faster-RCNN<sup>[134]</sup> 和 SBL<sup>[135]</sup>, mAP 值均有 20 左右的提升, 效果显著.

3) 鉴于旋转框对目标的定位更加精细, 现阶段针对旋转框检测的研究数量逐年增长, 意味着旋转框检测在遥感领域应用潜力不断提高. 相似算法对比中, 旋转框检测的 mAP 值要略低于水平框, 说明旋转框在进行评测时候更为严格, 也表明了旋转框的检测难度更高.

4) 随着研究的深入, 选取的主干网络模型层数越来越深. 不同网络模型检测性能差别很大, 网络

层数越深, 意味着非线性拟合能力越强, 从而性能越好. 遥感图像目标检测性能提升很大程度依赖深度学习自身的发展.

5) DOTA 数据集中大多数图像有着超大的图像尺寸和覆盖面积, 因此所有在 DOTA 上进行评测的方法都涉及基于超大图像尺寸的改进, 图像分块的方式已经成为 DOTA 上公认的标准预处理方式.

6) 多尺度特征金字塔进行检测的方式, 可以有效处理目标尺度范围变化大的问题, 现阶段基本成为标准方法, 为每个模型所共用.

7) 水平框检测中不存在边界突变问题, 反而要针对目标方向变化的问题设计方向标准化模块; 旋转框检测时候恰好要利用旋转敏感特征来进行角度的回归, 因此定位分支不需要进行方向标准化, 反而需要考虑因旋转导致的边界突变问题.

8) 现有针对密集分布、外观模糊的目标检测研究成果并不多, 说明以现有方法手段仍难以找到该问题的有效处理方式.

## 5 现存问题和发展趋势

现阶段遥感图像目标检测仍然存在很多问题, 限制了其在实际场合下的进一步推广. 现存主要问题包括:

1) 针对超大图像尺寸的遥感图像目标检测, 现有方法尚不能直接对图像全局进行检测, 典型的分块处理方式效率较低, 引入过多冗余的计算, 同时容易割裂目标, 造成精度上的损失.

2) 针对遥感图像中目标密集分布和外观模糊等常见特点, 目前还没有较好的针对性处理方式. 具备该特点的目标大量存在, 难以进行精确的检测, 导致现有算法在实际应用中受限.

3) 现有目标检测方法大多是基于目标本身视觉特征, 缺少根据图像整体和上下文进行理解和推理的过程. 尽管个别研究存在引入上下文信息和全局信息的方式, 但依然还是针对视觉特征, 缺乏高层语义知识的引导, 可解释性差.

4) 现阶段有了较大规模的数据集, 解决了目标检测对大容量训练集的需求问题, 但是现有数据集的质量还有待提升, 大量实际存在的小目标、模糊目标没有标注出来, 限制了现有算法潜力.

5) 通用的水平框目标检测方式, 对遥感图像中方向性目标定位模糊, 大量研究采用旋转框检测方式, 使包围框更加紧凑. 但是旋转框检测的定位精度相对于水平框明显降低, 存在较多定位不准的目标.

6) 针对遥感图像数据规模较大的问题, 现有方法的处理速度较慢, 在海量数据下难以满足实时性

表 1 水平框检测算法性能对比  
Table 1 Performance comparison of horizontal box detection algorithms

| 算法                                      | 主干       | 改进模块 |      |      |      |      |      |      | mAP |       |
|---|----------|------|------|------|------|------|------|------|-----|-------|
|   |          | 超大覆盖 | 方向多样 | 尺度过小 | 密集分布 | 形状差异 | 尺度变化 | 外观模糊 |     | 复杂背景  |
| NWPU VHR-10数据集                          |          |      |      |      |      |      |      |      |     |       |
| RICNN <sup>[25]</sup>                   | AlexNet  |      | ✓    |      |      |      |      |      |     | 73.10 |
| R-P-Faster-RCNN <sup>[134]</sup>        | VGG-16   |      |      |      |      |      |      |      |     | 76.50 |
| Def.R-FCN <sup>[87]</sup>               | Res-101  |      |      |      |      | ✓    |      |      |     | 79.10 |
| Def.Faster-RCNN <sup>[88]</sup>         | Res-50   |      |      |      |      | ✓    | ✓    |      |     | 84.40 |
| RICADet <sup>[111]</sup>                | ZF       |      | ✓    |      |      |      |      | ✓    |     | 87.12 |
| RDAS512 <sup>[72]</sup>                 | VGG-16   |      |      | ✓    |      |      | ✓    |      |     | 89.50 |
| Multi-Scale CNN <sup>[98]</sup>         | VGG-16   |      |      |      |      |      | ✓    |      |     | 89.60 |
| CAD-Net <sup>[108]</sup>                | Res-101  |      |      |      |      |      | ✓    | ✓    |     | 91.50 |
| SCRDet <sup>[80]</sup>                  | Res-101  |      |      | ✓    | ✓    |      | ✓    |      | ✓   | 91.75 |
| DOTA数据集                                 |          |      |      |      |      |      |      |      |     |       |
| FR-H <sup>[53]</sup>                    | Res-101  | ✓    |      |      |      |      |      |      |     | 60.46 |
| SBL <sup>[135]</sup>                    | Res-50   | ✓    |      |      |      |      |      |      |     | 64.77 |
| FMSSD <sup>[73]</sup>                   | VGG-16   | ✓    |      | ✓    |      |      | ✓    |      |     | 72.43 |
| ICN <sup>[84]</sup>                     | Res-101  | ✓    |      |      |      |      | ✓    |      |     | 72.50 |
| IoU-Adaptive <sup>[110]</sup>           | Res-101  | ✓    |      |      |      |      | ✓    |      |     | 72.72 |
| EFR <sup>[106]</sup>                    | VGG-16   | ✓    |      |      |      |      | ✓    |      |     | 73.49 |
| SCRDet <sup>[80]</sup>                  | Res-101  | ✓    |      | ✓    | ✓    |      | ✓    |      | ✓   | 75.35 |
| FADet <sup>[80]</sup>                   | Res-101  | ✓    |      |      |      |      | ✓    | ✓    | ✓   | 75.38 |
| MFIAR-Net t <sup>[117]</sup>            | Res-152  | ✓    |      |      |      |      | ✓    |      | ✓   | 76.07 |
| Mask OBB <sup>[105]</sup>               | ResX-101 | ✓    |      |      |      |      | ✓    |      | ✓   | 76.98 |
| A <sup>2</sup> RMNet <sup>[90]</sup>    | Res-101  | ✓    |      |      |      |      | ✓    | ✓    |     | 78.45 |
| OWSR <sup>[109]</sup>                   | Res-101  | ✓    |      |      |      |      | ✓    |      | ✓   | 78.79 |
| Parallel Cascade R-CNN <sup>[101]</sup> | ResX-101 | ✓    |      | ✓    |      |      | ✓    |      |     | 78.96 |
| DM-FPN <sup>[99]</sup>                  | Res-101  | ✓    |      | ✓    |      |      | ✓    |      |     | 79.27 |
| SCRDet++ <sup>[82]</sup>                | Res-101  | ✓    |      | ✓    | ✓    |      | ✓    |      | ✓   | 79.35 |

要求.

针对以上问题, 结合现阶段自然图像中深度学习目标检测方法的发展趋势, 本文对后续该领域的发展趋势进行展望:

1) 对超大尺寸遥感图像进行目标检测, 采用基于图像整体进行感兴趣区域提取的方式来替代分块处理方式, 可以快速滤除大部分背景区域, 从而避免计算冗余, 提高算法效率.

2) 设计更好的特征增强方法, 来加强小目标、密集目标、模糊目标等较难检测目标的特征表示, 降低复杂背景的干扰, 从而进一步提高整体精度、增强鲁棒性.

3) 针对现有检测方法缺少知识推理的过程、以及可解释性不足的问题, 引入知识和推理模块 (例如知识图谱、图卷积等) 来辅助进行目标检测, 具有较大的研究潜力.

4) 针对数据标注存在的质量问题, 可采用弱监督学习、半监督学习、迁移学习等算法来解决, 以缓解算法对于数据集标注的高标准要求.

5) 遥感图像目标检测中旋转框具有更加精细的定位能力, 代替水平框进行检测是大势所趋, 但是旋转框检测方式会带来新的问题, 例如边界突变问题等, 导致旋转框定位的精度和水平框相比存在差距. 针对旋转框检测进一步提高定位能力, 具有较高的意义.

6) 轻量化网络结构的设计和算法推理过程的加速, 对于遥感图像目标检测在实际工程的应用具有重要的意义和价值. 在检测精度基本满足实际需求后, 算法轻量化以及加速将成为后续极具价值的方向, 例如可借鉴剪枝、量化等加速操作, 同时网络架构搜索 (NAS) 也是极具潜力的方向.

7) 持续跟进深度学习发展趋势, 快速融合最新

表 2 旋转框检测算法性能对比  
Table 2 Performance comparison of rotated box detection algorithms

| 算法                                   | 主干       | 改进模块 |      |      |      |      |      |      |      | mAP   |
|--------------------------------------|----------|------|------|------|------|------|------|------|------|-------|
|                                      |          | 超大覆盖 | 尺度过小 | 密集分布 | 形状差异 | 尺度变化 | 外观模糊 | 复杂背景 | 边界问题 |       |
| FR-O <sup>[119]</sup>                | Res-101  | ✓    |      |      |      |      |      |      |      | 52.93 |
| IENet <sup>[91]</sup>                | Res-101  | ✓    |      |      | ✓    | ✓    |      |      | ✓    | 57.14 |
| TOSO <sup>[96]</sup>                 | Res-101  | ✓    |      |      | ✓    |      |      |      | ✓    | 57.52 |
| R-DFPN <sup>[83]</sup>               | Res-101  | ✓    |      |      |      | ✓    |      |      |      | 57.94 |
| R <sup>2</sup> CNN <sup>[121]</sup>  | Res-101  | ✓    |      |      |      |      |      |      |      | 60.67 |
| RRPN <sup>[120]</sup>                | Res-101  | ✓    |      |      |      |      |      |      |      | 61.01 |
| Axis Learning <sup>[95]</sup>        | Res-101  | ✓    |      |      | ✓    | ✓    |      |      |      | 65.98 |
| ICN <sup>[84]</sup>                  | Res-101  | ✓    |      |      |      | ✓    |      |      |      | 68.20 |
| RADet <sup>[107]</sup>               | ResX-101 | ✓    | ✓    |      |      | ✓    |      | ✓    | ✓    | 69.09 |
| RoI-Transformer <sup>[86]</sup>      | Res-101  | ✓    |      |      |      | ✓    |      |      |      | 69.56 |
| P-RSDet <sup>[92]</sup>              | Res-101  | ✓    |      |      | ✓    |      |      |      |      | 69.82 |
| CAD-Net <sup>[108]</sup>             | Res-101  | ✓    |      |      |      | ✓    | ✓    |      |      | 69.90 |
| O <sup>2</sup> -DNet <sup>[93]</sup> | HG-104   | ✓    |      |      | ✓    |      |      |      |      | 71.04 |
| AOOD <sup>[103]</sup>                | Res-101  | ✓    | ✓    |      |      | ✓    |      |      |      | 71.18 |
| Cascade-FF <sup>[116]</sup>          | Res-152  | ✓    | ✓    |      |      | ✓    |      | ✓    |      | 71.80 |
| SCRDet <sup>[80]</sup>               | Res-101  | ✓    | ✓    | ✓    |      | ✓    |      | ✓    | ✓    | 72.61 |
| SARD <sup>[124]</sup>                | Res-101  | ✓    |      |      |      | ✓    |      |      | ✓    | 72.95 |
| GLS-Net <sup>[118]</sup>             | Res-101  | ✓    |      |      |      | ✓    |      | ✓    |      | 72.96 |
| DRN <sup>[97]</sup>                  | HG-104   | ✓    |      |      | ✓    |      |      |      |      | 73.23 |
| FADet <sup>[80]</sup>                | Res-101  | ✓    |      |      | ✓    | ✓    |      | ✓    |      | 73.28 |
| MFIAR-Net <sup>[117]</sup>           | Res-152  | ✓    |      |      |      | ✓    |      | ✓    |      | 73.49 |
| R <sup>3</sup> Det <sup>[81]</sup>   | Res-152  | ✓    |      | ✓    |      | ✓    |      |      |      | 73.74 |
| RSDet <sup>[126]</sup>               | Res-152  | ✓    |      |      |      | ✓    |      |      | ✓    | 74.10 |
| Gliding Vertex <sup>[123]</sup>      | Res-101  | ✓    |      |      |      | ✓    |      |      | ✓    | 75.02 |
| Mask OBB <sup>[105]</sup>            | ResX-101 | ✓    |      |      |      | ✓    |      | ✓    | ✓    | 75.33 |
| FFA <sup>[104]</sup>                 | Res-101  | ✓    |      |      |      | ✓    |      |      |      | 75.70 |
| APE <sup>[94]</sup>                  | ResX-101 | ✓    |      |      | ✓    |      |      |      | ✓    | 75.75 |
| CSL <sup>[125]</sup>                 | Res-152  | ✓    |      |      |      | ✓    |      |      | ✓    | 76.17 |
| OWSR <sup>[109]</sup>                | Res-101  | ✓    |      |      |      | ✓    |      | ✓    |      | 76.36 |
| SCRDet++ <sup>[82]</sup>             | Res-101  | ✓    | ✓    | ✓    |      | ✓    |      | ✓    | ✓    | 76.81 |

的先进成果, 例如更好的网络基础架构、最新的目标检测方式等, 进一步提高检测性能, 降低时间开销。

## 6 结论

光学遥感图像目标检测是遥感图像理解的基础研究内容, 具有很大的应用价值. 本文首先总结了遥感图像中目标检测的特点和挑战; 后续概括陈述了早期基于手工设计特征的算法, 重点对候选区域提取和特征提取进行了总结和分析; 之后对深度学习算法进行了系统、全面的总结, 重点梳理了针对遥感图像目标检测特点的改进方案, 详细分析了不同改进方案的优缺点, 在常用数据集下定量评估了

不同改进方案取得的性能提升; 最后对现阶段遥感图像目标检测依然存在的问题进行了梳理, 并对未来的发展趋势进行了展望。

## References

- 1 Wang Yan-Qing, Ma Lei, Tian Yuan. State-of-the-art of ship detection and recognition in optical remotely sensed imagery. *Acta Automatica Sinica*, 2011, **37**(9): 1029–1039  
(王彦清, 马雷, 田原. 光学遥感图像舰船目标检测与识别综述. 自动化学报, 2011, **37**(9): 1029–1039)
- 2 Cheng G, Han J W. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, **117**: 11–28
- 3 Liu Xiao-Bo, Liu Peng, Cai Zhi-Hua, Qiao Yu-Lin, Wang Ling, Wang Min. Research progress of optical remote sensing image

- object detection based on deep learning. *Acta Automatica Sinica*, DOI: 10.16383/j.aas.c190455  
(刘小波, 刘鹏, 蔡之华, 乔禹霖, 王凌, 汪敏. 基于深度学习的光学遥感图像目标检测研究进展. 自动化学报, DOI: 10.16383/j.aas.c190455)
- 4 Li K, Wan G, Cheng G, Meng L Q, Han J W. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, **159**: 296–307
  - 5 Zhang Hao-Kui, Li Ying, Jiang Ye-Nan. Deep learning for hyperspectral imagery classification: The state of the art and prospects. *Acta Automatica Sinica*, 2018, **44**(6): 961–977  
(张号逵, 李映, 姜晔楠. 深度学习在高光谱图像分类领域的研究现状与展望. 自动化学报, 2018, **44**(6): 961–977)
  - 6 Long Gang, Chen Xue-Quan. A method for automatic detection of ships in harbor area in high-resolution remote sensing image. *Computer Simulation*, 2007, **24**(5): 198–201  
(隆刚, 陈学全. 高分辨率遥感图像港内舰船的自动检测方法. 计算机仿真, 2007, **24**(5): 198–201)
  - 7 Antelo J, Ambrosio G, Gonzalez J, Galindo C. Ship detection and recognition in high-resolution satellite images. In: Proceedings of the 2009 IEEE International Geoscience and Remote Sensing Symposium. Cape Town, South Africa: IEEE, 2009. IV-514-IV-517
  - 8 Liu G, Zhang Y S, Zheng X W, Sun X, Fu K, Wang H Q. A new method on inshore ship detection in high-resolution satellite images using shape and context information. *IEEE Geoscience and Remote Sensing Letters*, 2014, **11**(3): 617–621
  - 9 Chu Zhao-Liang, Wang Qing-Hua, Chen Hai-Lin, Xu Shou-Shi. Ship auto detection method based on minimum error threshold segmentation. *Computer Engineering*, 2007, **33**(11): 239–241  
(储昭亮, 王庆华, 陈海林, 徐守时. 基于极小误差阈值分割的舰船自动检测方法. 计算机工程, 2007, **33**(11): 239–241)
  - 10 You X, Li W H. A sea-land segmentation scheme based on statistical model of sea. In: Proceedings of the 4th International Congress on Image and Signal Processing. Shanghai, China: IEEE, 2011. 1155–1159
  - 11 Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, **9**(1): 62–66
  - 12 Liu D H, He L H, Carin L. Airport detection in large aerial optical imagery. In: Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal, Canada: IEEE, 2004. Article No. V-761
  - 13 Tao C, Tan Y H, Cai H J, Tian J W. Airport detection from large IKONOS images using clustered SIFT keypoints and region information. *IEEE Geoscience and Remote Sensing Letters*, 2011, **8**(1): 128–132
  - 14 Yao X W, Han J W, Guo L, Bu S H, Liu Z B. A coarse-to-fine model for airport detection from remote sensing images using target-oriented visual saliency and CRF. *Neurocomputing*, 2015, **164**: 162–172
  - 15 Xu J, Fu K, Sun X. An invariant generalized Hough transform based method of inshore ships detection. In: Proceedings of the 2011 International Symposium on Image and Data Fusion. Tengchong, China: IEEE, 2011. 1–4
  - 16 Xu J, Sun X, Zhang D B, Fu K. Automatic detection of inshore ships in high-resolution remote sensing images using robust invariant generalized Hough transform. *IEEE Geoscience and Remote Sensing Letters*, 2014, **11**(12): 2070–2074
  - 17 Harvey N R, Porter R, Theiler J. Ship detection in satellite imagery using rank-order grayscale hit-or-miss transforms. In: Proceedings of SPIE 7701, Visual Information Processing XIX. Orlando, USA: SPIE, 2010. Article No. 770102
  - 18 Liu G, Sun X, Fu K, Wang H Q. Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior. *IEEE Geoscience and Remote Sensing Letters*, 2013, **10**(3): 573–577
  - 19 Zhang Z M, Warrell J, Torr P H S. Proposal generation for object detection using cascaded ranking SVMs. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR 2011). Colorado Springs, USA: IEEE, 2011. 1497–1504
  - 20 Cheng M M, Zhang Z M, Lin W Y, Torr P. BING: Binarized normed gradients for objectness estimation at 300fps. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014. 3286–3293
  - 21 Liu Z K, Wang H Z, Weng L B, Yang Y P. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. *IEEE Geoscience and Remote Sensing Letters*, 2016, **13**(8): 1074–1078
  - 22 Uijlings J R R, Van De Sande K E A, Gevers T, Smeulders A W M. Selective search for object recognition. *International Journal of Computer Vision*, 2013, **104**(2): 154–171
  - 23 Felzenszwalb P F, Huttenlocher D P. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 2004, **59**(2): 167–181
  - 24 Cheng G, Zhou P C, Yao X W, Yao C, Zhang Y B, Han J W. Object detection in VHR optical remote sensing images via learning rotation-invariant HOG feature. In: Proceedings of the 4th International Workshop on Earth Observation and Remote Sensing Applications. Guangzhou, China: IEEE, 2016. 433–436
  - 25 Cheng G, Zhou P C, Han J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, **54**(12): 7405–7415
  - 26 Song Z N, Sui H G, Wang Y J. Automatic ship detection for optical satellite images based on visual attention model and LBP. In: Proceedings of the 2014 IEEE Workshop on Electronics, Computer and Applications. Ottawa, Canada: IEEE, 2014.
  - 27 Zhang L B, Zhang Y Y. Airport detection and aircraft recognition based on two-layer saliency model in high spatial resolution remote-sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, **10**(4): 1511–1524
  - 28 Ding Z H, Yu Y, Wang B, Zhang L M. An approach for visual attention based on biquaternion and its application for ship detection in multispectral imagery. *Neurocomputing*, 2012, **76**(1): 9–17
  - 29 Xu F, Liu J H. Ship detection and extraction using visual saliency and histogram of oriented gradient. *Optoelectronics Letters*, 2016, **12**(6): 473–477
  - 30 Cucchiara R, Grana C, Piccardi M, Prati A, Sirotti S. Improving shadow suppression in moving object detection with HSV color information. In: Proceedings of the 2001 IEEE Intelligent Transportation Systems (ITSC 2001). Oakland, USA: IEEE, 2001. 334–339
  - 31 Morillas J R A, García I C, Zölzer U. Ship detection based on SVM using color and texture features. In: Proceedings of the 2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP). Cluj-Napoca, Romania: IEEE, 2015. 343–350
  - 32 Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal*



- of Computer Vision*, 2001, **42**(3): 145–175
- 33 Li Z C, Itti L. Saliency and gist features for target detection in satellite images. *IEEE Transactions on Image Processing*, 2011, **20**(7): 2017–2029
- 34 Brekke C, Solberg A H S. Oil spill detection by satellite remote sensing. *Remote Sensing of Environment*, 2005, **95**(1): 1–13
- 35 Selvi M U, Kumar S S. A novel approach for ship recognition using shape and texture. *International Journal of Advanced Information Technology*, 2011, **1**(5): 23–29
- 36 Hsieh J W, Chen J M, Chuang C H, Fan K C. Aircraft type recognition in satellite images. *IEE Proceedings-Vision, Image, and Signal Processing*, 2005, **152**(3): 307–315
- 37 Sun H, Sun X, Wang H Q, Li Y, Li X J. Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model. *IEEE Geoscience and Remote Sensing Letters*, 2012, **9**(1): 109–113
- 38 Tong S, Sun K, Shi B H, Chen J Y. A ship target automatic recognition method for sub-meter remote sensing images. In: Proceedings of the 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA). Guangzhou, China: IEEE, 2016.
- 39 Shi Z W, Yu X R, Jiang Z G, Li B. Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, **52**(8): 4511–4523
- 40 Gong C, Han J W, Guo L, Qian X L, Zhou P C, Yao X W, et al. Object detection in remote sensing imagery using a discriminatively trained mixture model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2013, **85**: 32–43
- 41 Yin Y Y, Liu N, Li C Y, Wan W B, Fang T. Coarse-to-fine ship detection using visual saliency fusion and feature encoding for optical satellite images. In: Proceedings of the 2016 International Conference on Audio, Language and Image Processing (ICALIP). Shanghai, China: IEEE, 2016.
- 42 Zhang W C, Sun X, Fu K, Wang C Y, Wang H Q. Object detection in high-resolution remote sensing images using rotation invariant parts based model. *IEEE Geoscience and Remote Sensing Letters*, 2014, **11**(1): 74–78
- 43 Li B Y, Cui X G, Bai J. A cascade structure of aircraft detection in high resolution remote sensing images. In: Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Beijing, China: IEEE, 2016.
- 44 Zheng J B, Xi Y, Feng M C, Li X X, Li N. Object detection based on BING in optical remote sensing images. In: Proceedings of the 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). Datong, China: IEEE, 2016. 504–509
- 45 Hinz S, Stilla U. Car detection in aerial thermal images by local and global evidence accumulation. *Pattern Recognition Letters*, 2006, **27**(4): 308–315
- 46 Gu W, Lv Z H, Hao M. Change detection method for remote sensing images based on an improved Markov random field. *Multimedia Tools and Applications*, 2017, **76**(17): 17719–17734
- 47 Hosseini M, Saradjian M, Javahery A, Nadi S. Noise removal from land cover maps by post processing of classification results. In: Proceedings of the 3rd International Conference on Recent Advances in Space Technologies. Istanbul, Turkey: IEEE, 2007.
- 48 Thapa R B, Murayama Y. Urban mapping, accuracy, & image classification: A comparison of multiple approaches in Tsukuba City, Japan. *Applied Geography*, 2009, **29**(1): 135–144
- 49 O'Donnell T, Goyne K W, Miles R J, Baffaut C, Anderson S H, Sudduth K A. Identification and quantification of soil redoximorphic features by digital image processing. *Geoderma*, 2010, **157**(3-4): 86–96
- 50 Corbane C, Najman L, Pecoul E, Demagistri L, Petit M. A complete processing chain for ship detection using optical satellite imagery. *International Journal of Remote Sensing*, 2010, **31**(22): 5837–5854
- 51 Zhu X F, Ma C W. The study of combined invariants optimization method on aircraft recognition. In: Proceedings of the 2011 Symposium on Photonics and Optoelectronics (SOPO). Wuhan, China: IEEE, 2011. 1–4
- 52 Wang D H, He X, Wei Z H, Yu H L. A method of aircraft image target recognition based on modified PCA features and SVM. In: Proceedings of the 9th International Conference on Electronic Measurement and Instruments. Beijing, China: IEEE, 2009. 4-177–4-181
- 53 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137–1149
- 54 Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 779–788
- 55 Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, et al. SSD: Single shot MultiBox detector. In: Proceeding of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 21–37
- 56 Law H, Deng J. CornerNet: Detecting objects as paired keypoints. In: Proceedings of the 15th European Conference on Computer Vision (ECCV 2018). Munich, Germany: Springer, 2018. 765–781
- 57 Duan K W, Bai S, Xie L X, Qi H G, Huang Q M, Tian Q. CenterNet: Keypoint triplets for object detection. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019. 6568–6577
- 58 Tian Z, Shen C H, Chen H, He T. FCOS: Fully convolutional one-stage object detection. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019. 9627–9636
- 59 Yan Z G, Song X, Zhong H Y, Zhu X Z. Object detection in optical remote sensing images based on transfer learning convolutional neural networks. In: Proceedings of the 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS). Nanjing, China: IEEE, 2018. 935–942
- 60 Yang X, Sun H, Sun X, Yan M L, Guo Z, Fu K. Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network. *IEEE Access*, 2018, **6**: 50839–50849
- 61 He Y Q, Sun X, Gao L R, Zhang B. Ship detection without sea-land segmentation for large-scale high-resolution optical satellite images. In: Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018). Valencia, Spain: IEEE, 2018. 717–720
- 62 Fu Y M, Wu F G, Zhao J S. Context-aware and depthwise-based detection on orbit for remote sensing image. In: Proceedings of the 24th International Conference on Pattern Recognition (ICPR). Beijing, China: IEEE, 2018. 1725–1730
- 63 Li M J, Guo W W, Zhang Z H, Yu W X, Zhang T. Rotated region based fully convolutional network for ship detection. In: Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018). Valencia, Spain: IEEE, 2018. 673–676

- 64 Schilling H, Bulatov D, Niessner R, Middelmann W, Soergel U. Detection of vehicles in multisensor data via multibranch convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, **11**(11): 4299–4316
- 65 Li X B, Wang S J. Object detection using convolutional neural networks in a coarse-to-fine manner. *IEEE Geoscience and Remote Sensing Letters*, 2017, **14**(11): 2037–2041
- 66 Wang C, Bai X, Wang S, Zhou J, Ren P. Multiscale visual attention networks for object detection in VHR remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 2019, **16**(2): 310–314
- 67 Pang J M, Li C, Shi J P, Xu Z H, Feng H J.  $R^2$ -CNN: Fast tiny object detection in large-scale remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, **57**(8): 5512–5524
- 68 Van Etten A. You only look twice: Rapid multi-scale object detection in satellite imagery. arXiv preprint arXiv: 1805.09512, 2018
- 69 Zhang F, Du B, Zhang L P, Xu M Z. Weakly supervised learning based on coupled convolutional neural networks for aircraft detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, **54**(9): 5553–5563
- 70 Cheng G, Han J W, Zhou P C, Xu D. Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection. *IEEE Transactions on Image Processing*, 2019, **28**(1): 265–278
- 71 Wu X, Hong D F, Tian J J, Chanussot J, Li W, Tao R. OR-SIm detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, **57**(7): 5146–5158
- 72 Chen S Q, Zhan R H, Zhang J. Geospatial object detection in remote sensing imagery based on multiscale single-shot detector with activated semantics. *Remote Sensing*, 2018, **10**(6): Article No. 820
- 73 Wang P J, Sun X, Diao W H, Fu K. FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, **58**(5): 3377–3390
- 74 Zhang W, Wang S H, Thachan S, Chen J Z, Qian Y T. Deconv R-CNN for small object detection on remote sensing images. In: Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018). Valencia, Spain: IEEE, 2018. 2483–2486
- 75 Liu W C, Ma L, Wang J, Chen H. Detection of multiclass objects in optical remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 2019, **16**(5): 791–795
- 76 Ying X, Wang Q, Li X W, Yu M, Jiang H, Gao J, et al. Multi-attention object detection model in remote sensing images based on multi-scale. *IEEE Access*, 2019, **7**: 94508–94519
- 77 Long H, Chung Y, Liu Z B, Bu S H. Object detection in aerial images using feature fusion deep networks. *IEEE Access*, 2019, **7**: 30980–30990
- 78 Cheng L, Liu X, Li L L, Jiao L C, Tang X. Deep adaptive proposal network for object detection in optical remote sensing images. arXiv preprint arXiv: 1807.07327, 2018
- 79 Long Y, Gong Y P, Xiao Z F, Liu Q. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, **55**(5): 2486–2498
- 80 Yang X, Yang J R, Yan J C, Zhang Y, Zhang T F, Guo Z, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019. 8232–8241
- 81 Yang X, Liu Q Q, Yan J C, Li A. R3Det: Refined single-stage detector with feature refinement for rotating object. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2021.
- 82 Yang X, Yan J C, Yang X K, Tang J, Liao W L, He T. SCRDet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing. arXiv preprint arXiv: 2004.13316, 2020
- 83 Yang X, Sun H, Fu K, Yang J R, Su X, Yan M L, et al. Automatic ship detection in remote sensing images from Google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sensing*, 2018, **10**(1): Article No. 132
- 84 Azimi S M, Vig E, Bahmanyar R, Körner M, Reinartz P. Towards multi-class object detection in unconstrained remote sensing imagery. In: Proceedings of the 14th Asian Conference on Computer Vision (ACCV 2018). Perth, Australia: Springer, 2018. 150–165
- 85 Wang X B, Jiang Y Y, Luo Z B, Liu C L, Choi H, Kim S. Arbitrary shape scene text detection with adaptive text region representation. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 6449–6458
- 86 Ding J, Xue N, Long Y, Xia G S, Lu Q K. Learning RoI transformer for oriented object detection in aerial images. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 2844–2853
- 87 Xu Z Z, Xu X, Wang L, Yang R, Pu F L. Deformable ConvNet with aspect ratio constrained NMS for object detection in remote sensing imagery. *Remote Sensing*, 2017, **9**(12): Article No. 1312
- 88 Ren Y, Zhu C R, Xiao S P. Deformable Faster R-CNN with aggregating multi-layer features for partially occluded object detection in optical remote sensing images. *Remote Sensing*, 2018, **10**(9): Article No. 1470
- 89 Li C Z, Xu C Y, Cui Z, Wang D, Zhang T, Yang J. Feature-attended object detection in remote sensing imagery. In: Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP). Taipei, China: IEEE, 2019. 3886–3890
- 90 Qiu H Q, Li H L, Wu Q B, Meng F M, Ngan K N, Shi H C. A<sup>2</sup>RMNet: Adaptively aspect ratio multi-scale network for object detection in remote sensing images. *Remote Sensing*, 2019, **11**(13): Article No. 1594
- 91 Lin Y T, Feng P M, Guan J, Wang W W, Chambers J. IENet: Interacting embranchment one stage anchor free detector for orientation aerial object detection. arXiv preprint arXiv: 1912.00969, 2019
- 92 Zhou L, Wei H R, Li H, Zhang Y, Sun X, Zhao W Z. Objects detection for remote sensing images based on polar coordinates. arXiv preprint arXiv: 2001.02988, 2020
- 93 Wei H R, Zhang Y, Chang Z H, Li H, Wang H Q, Sun X. Oriented objects as pairs of middle lines. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, **169**: 268–279
- 94 Zhu Y X, Du J, Wu X Q. Adaptive period embedding for representing oriented objects in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, **58**(10): 7247–7257
- 95 Xiao Z F, Qian L J, Shao W P, Tian X W, Wang K. Axis

- learning for orientated objects detection in aerial images. *Remote Sensing*, 2020, **12**(6): Article No. 908
- 96 Feng P M, Lin Y T, Guan J J, He G J, Shi H F, Chambers J. TOSO: Student's-t distribution aided one-stage orientation target detection in remote sensing images. In: Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020). Barcelona, Spain: IEEE, 2020. 4057–4061
- 97 Pan X J, Ren Y Q, Sheng K K, Dong W M, Yuan H L, Guo X W, et al. Dynamic refinement network for oriented and densely packed object detection. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 11207–11216
- 98 Guo W, Yang W, Zhang H J, Hua G. Geospatial object detection in high resolution satellite images based on multi-scale convolutional neural network. *Remote Sensing*, 2018, **10**(1): Article No. 131
- 99 Zhang X D, Zhu K, Chen G Z, Tan X L, Zhang L F, Dai F, et al. Geospatial object detection on high resolution remote sensing imagery based on double multi-scale feature pyramid network. *Remote Sensing*, 2019, **11**(7): Article No. 755
- 100 Zhang S, He G H, Chen H B, Jing N F, Wang Q. Scale adaptive proposal network for object detection in remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 2019, **16**(6): 864–868
- 101 Hou J Y, Ma H B, Wang S J. Parallel cascade R-CNN for object detection in remote sensing imagery. *Journal of Physics: Conference Series*, 2020, **1544**: Article No. 012124
- 102 Lin T Y, Dollár P, Girshick R, He K M, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 2117–2125
- 103 Zou F H, Xiao W, Ji W T, He K K, Yang Z X, Song J K, et al. Arbitrary-oriented object detection via dense feature fusion and attention model for remote sensing super-resolution image. *Neural Computing and Applications*, 2020, **32**(18): 14549–14562
- 104 Fu K, Chang Z H, Zhang Y, Xu G L, Zhang K S, Sun X. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, **161**: 294–308
- 105 Wang J W, Ding J, Guo H W, Cheng W S, Pan T, Yang W. Mask OBB: A semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images. *Remote Sensing*, 2019, **11**(24): Article No. 2930
- 106 Fu K, Chen Z, Zhang Y, Sun X. Enhanced feature representation in detection for optical remote sensing images. *Remote Sensing*, 2019, **11**(18): Article No. 2095
- 107 Li Y Y, Huang Q, Pei X, Jiao L C, Shang R H. RADet: Refine feature pyramid network and multi-layer attention network for arbitrary-oriented object detection of remote sensing images. *Remote Sensing*, 2020, **12**(3): Article No. 389
- 108 Zhang G J, Lu S J, Zhang W. CAD-Net: A context-aware detection network for objects in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, **57**(12): 10015–10024
- 109 Xu C Y, Li C Z, Cui Z, Zhang T, Yang J. Hierarchical semantic propagation for object detection in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, **58**(6): 4353–4364
- 110 Yan J Q, Wang H Q, Yan M L, Diao W H, Sun X, Li H. IoU-adaptive deformable R-CNN: Make full use of IoU for multi-class object detection in remote sensing imagery. *Remote Sensing*, 2019, **11**(3): Article No. 286
- 111 Li K, Cheng G, Bu S H, You X. Rotation-insensitive and context-augmented object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, **56**(4): 2337–2348
- 112 Chen C Y, Gong W G, Chen Y L, Li W H. Object detection in remote sensing images based on a scene-contextual feature pyramid network. *Remote Sensing*, 2019, **11**(3): Article No. 339
- 113 Ma W P, Guo Q Q, Wu Y, Zhao W, Zhang X R, Jiao L C. A novel multi-model decision fusion network for object detection in remote sensing images. *Remote Sensing*, 2019, **11**(7): Article No. 737
- 114 Zhang Y K, You Y N, Wang R, Liu F, Liu J. Nearshore vessel detection based on Scene-mask R-CNN in remote sensing image. In: Proceedings of the 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC). Guiyang, China: IEEE, 2018. 76–80
- 115 Li Q P, Mou L C, Jiang K Y, Liu Q J, Wang Y H, Zhu X X. Hierarchical region based convolution neural network for multiscale object detection in remote sensing images. In: Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018). Valencia, Spain: IEEE, 2018. 4355–4358
- 116 Hou L P, Lu K, Xue J, Hao L. Cascade detector with feature fusion for arbitrary-oriented objects in remote sensing images. In: Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME). London, UK: IEEE, 2020. 1–6
- 117 Yang F, Li W T, Hu H W, Li W Y, Wang P. Multi-scale feature integrated attention-based rotation network for object detection in VHR aerial images. *Sensors*, 2020, **20**(6): Article No. 1686
- 118 Li C Y, Luo B, Hong H L, Su X, Wang Y J, Liu J, et al. Object detection based on global-local saliency constraint in aerial images. *Remote Sensing*, 2020, **12**(9): Article No. 1435
- 119 Xia G S, Bai X, Ding J, Zhu Z, Belongie S, Luo J B, et al. DOTA: A large-scale dataset for object detection in aerial images. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 3974–3983
- 120 Ma J Q, Shao W Y, Ye H, Wang L, Wang H, Zheng Y B, et al. Arbitrary-oriented scene text detection via rotation proposals. *IEEE Transactions on Multimedia*, 2018, **20**(11): 3111–3122
- 121 Jiang Y Y, Zhu X Y, Wang X B, Yang S L, Li W, Wang H, et al.  $R^2$ CNN: Rotational region CNN for orientation robust scene text detection. arXiv preprint arXiv: 1706.09579, 2017
- 122 Cao L, Zhang X, Wang Z, et al. Multi angle rotation object detection for remote sensing image based on modified feature pyramid networks. *International Journal of Remote Sensing*, 2021: 5253–5276
- 123 Xu Y C, Fu M T, Wang Q M, Wang Y K, Chen K, Xia G S, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, **43**(4): 1452–1459
- 124 Wang Y S, Zhang Y, Zhang Y, Zhao L J, Sun X, Guo Z. SARD: Towards scale-aware rotated object detection in aerial imagery. *IEEE Access*, 2019, **7**: 173855–173865
- 125 Yang X, Yan J C. Arbitrary-oriented object detection with circular smooth label. In: Proceedings of the 16th European Conference on Computer Vision (ECCV 2020). Glasgow, UK: Springer, 2020. 677–694

- 126 Qian W, Yang X, Peng S L, Guo Y, Yan J C. Learning modulated loss for rotated object detection. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2021.
- 127 Heitz G, Koller D. Learning spatial Context: Using stuff to find things. In: Proceedings of the 10th European Conference on Computer Vision (ECCV 2008). Marseille, France: Springer, 2008. 30–43
- 128 Maggiori E, Tarabalka Y, Charpiat G, Alliez P. Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark. In: Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth, USA: IEEE, 2017. 3226–3229
- 129 Cheng G, Han J W, Zhou P C, Guo L. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, **98**: 119–132
- 130 Razakarivony S, Jurie F. Vehicle detection in aerial imagery: A small target detection benchmark. *Journal of Visual Communication and Image Representation*, 2016, **34**: 187–203
- 131 Zhu H G, Chen X G, Dai W Q, Fu K, Ye Q X, Jiao J B. Orientation robust object detection in aerial images using deep convolutional neural network. In: Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP). Quebec City, Canada: IEEE, 2015. 3735–3739
- 132 Liu K, Mattyus G. Fast multiclass vehicle detection on aerial images. *IEEE Geoscience and Remote Sensing Letters*, 2015, **12**(9): 1938–1942
- 133 Liu Z K, Yuan L, Weng L B, Yang Y P. A high resolution optical satellite image dataset for ship recognition and some new baselines. In: Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017). Porto, Portugal: SciTe Press, 2017. 324–331
- 134 Han X B, Zhong Y F, Zhang L P. An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery. *Remote Sensing*, 2017, **9**(7): Article No. 666
- 135 Sun P, Chen G, Shang Y. Adaptive saliency biased loss for object detection in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, **58**(10): 7154–7165



**聂光涛** 北京理工大学计算机科学与技术学院博士研究生。2015 年获得西安工业大学电子信息工程学院学士学位。2017 年获得北京理工大学自动化学院硕士学位。主要研究方向为计算机视觉, 目标检测和机器学习。

E-mail: nieguangtao@bit.edu.cn

**(NIE Guang-Tao** Ph. D. candidate at the School of Computer Science and Technology, Beijing Institute of Technology. He received the his bachelor degree from the School of Electronics and Information Engineering, Xi'an Technological University in 2015, and the master degree from the School of Automation, Beijing Institute of Technology in 2017. His research interest covers computer vision, object detection, and machine learning.)



**黄 华** 北京师范大学人工智能学院教授。分别于 1996 年和 2006 年获得西安交通大学学士学位和博士学位。主要研究方向为图像视频处理, 计算摄影学和计算机图形学。本文通信作者。E-mail: huahuang@bnu.edu.cn

**(HUANG Hua** Professor at the School of Artificial Intelligence, Beijing Normal University. He received his bachelor and Ph. D. degrees from Xi'an Jiaotong University, in 1996 and 2006, respectively. His research interest covers image and video processing, computational photography, and computer graphics. Corresponding author of this paper.)