



中国科学院大学  
University of Chinese Academy of Sciences

# 硕士学位论文

## 宽幅遥感图像舰船目标检测算法研究

作者姓名: 陈玺

指导教师: 何斌研究员

中国科学院长春光学精密机械与物理研究所

学位类别: 工学硕士

学科专业: 机械电子工程

培养单位: 中国科学院长春光学精密机械与物理研究所

2022年6月

**Research on ship target detection algorithm in**  
**wide-range remote sensing image**

A thesis submitted to  
University of Chinese Academy of Sciences  
in partial fulfillment of the requirement  
for the degree of  
Master of Engineering  
in Mechatronic Engineering  
By  
Chen Xi  
Supervisor: Professor He Bin

Changchun Institute of Optics, Fine Mechanics and Physics,  
Chinese Academy of Sciences

June 2022



## 摘要

近年来遥感对地观测技术疾速发展，各类高分辨率的光学成像卫星不断涌现，现今可获取的遥感图像与之前相比分辨率更好，可从图像中更容易直接捕捉到各种地物信息的样貌特征，在军事国防和民用经济建设领域都有重要应用。舰船作为水上运输的重要载体，主要在宽阔的河流海面上运行，由于其运行附近鲜有地面设备，周围地物信息较少，以水域为主，容易被空中的探测器捕捉到，所以成为对地观测技术需要探测的一个常用的目标对象。对遥感图像进行目标检测对海上救援、军事应用、船只管理等有重大意义。对遥感舰船目标的检测，已成为遥感对地观测技术的一个重要分支。

目前，本领域存在的问题主要是遥感图像的获取受到多方面的制约，如拍摄的时间，天气，光照，以及海上的种种影响因素都对目标的检测造成阻碍。同时，随着成像分辨率的提高，算法要处理的数据量也显著增加，对算法的速度也带来了更高的要求。如何能克服各种影响，最大程度从大量数据中快速，准确完成检测任务，成为本领域研究的主要方向。

针对现有的数据资源，本文针对大幅宽遥感图像进行舰船检测，在保证检测精度的同时，加快整体算法的检测速度。

本文主要内容主要可分为以下两部分：

### 1. 基于谱残差显著性模型的候选区域提取算法研究

为快速缩小检测视野，减少整体算法的计算量，本文设计了一种候选区域提取算法。以谱残差显著性算法基础计算显著图，结合算法特点和遥感图像特性在尺度选择和分割阈值的计算方法方面进行改进，有效克服了云雾和海浪等背景带来的干扰，同时可以最大限度减少目标遗漏。最后结合少量特征滤除存在的噪声等明显虚警。

### 2. 针对宽幅遥感图像中小目标及复杂背景下的轻量化舰船目标检测算法研究

本文针对大幅面遥感图像提出一种舰船目标检测方法，整体算法分为候选区域判断和目标精细检测。由于宽幅图像无法直接作为检测网络输入，需要将

图像裁剪为可输入的大小。在候选区域定位方面采取了基于第一部分研究的基于图像显著性的区域提取方法，判断裁剪后的子图像内部是否存在疑似目标，以保证进行后续检测。在检测网络方面，为保证检测的快速性，选择以 yolov4-tiny 轻量化目标检测模型为主干，同时为了改善数据中存在大量舰船小目标导致算法的检测精度下降的问题。在网络中加入注意力机制和数据增强模块，同时针对图像中大量的小目标和数据增强所带来的尺度和形状上的变化，采用了自适应选取先验框的方法，增强了模型对于此类样本的检测效果。与原始的 yolov4-tiny 相比，改进的算法在保证检测速度的条件下，增强了对目标的检测能力。在保证检测精确度的同时，大大提高了召回率，减少了漏检。同时，整体算法相较于纯基于深度学习的目标检测方法，在保证精度的同时，降低了计算量，提高了检测效率，可作为实际工程参考。

**关键词：** 大幅宽遥感图像，舰船检测，频域视觉显著性，yolov4-tiny

## Abstract

The rapid development in recent years, remote sensing for earth observation technology, all kinds of high resolution optical imaging satellites constantly emerging, now available on the remote sensing image resolution is better compared to before, can be more easily in the image directly capture the appearance characteristics of all kinds of feature information, in the field of military defense and civil economic construction has important application. As an important carrier of water transportation, ships mainly operate on the sea surface of wide rivers. Due to the lack of ground equipment and little information about other ground objects, ships mainly operate on water and are easy to be captured by aerial detectors. Therefore, they are a common target object to be detected by earth observation technology. Target detection in remote sensing images is of great significance to maritime rescue, military application and ship management. The detection of remote sensing ship target has become an important branch of remote sensing earth observation technology. At present, the main problem in this field is that the acquisition of remote sensing images is restricted by many aspects, such as the shooting time, weather, illumination, and various influencing factors on the sea, which hinder the detection of targets. At the same time, with the improvement of imaging resolution, the amount of data to be processed by the algorithm also increases significantly, which brings higher requirements on the speed of the algorithm. How to overcome all kinds of influences and maximize the rapid and accurate completion of detection tasks from a large amount of data has become the main research direction in this field.

Aiming at the existing data resources, this paper carries out ship detection for large and wide remote sensing images, and speeds up the detection speed of the whole algorithm while ensuring the detection accuracy.

The main content of this paper can be divided into the following two parts:

1. Research on candidate region extraction algorithm based on significance model of spectral residual

In order to reduce the detection field quickly and reduce the computational

burden of the whole algorithm, a candidate region extraction algorithm is designed in this paper. The saliency map is calculated on the basis of spectral residual saliency algorithm, and the calculation method of scale selection and segmentation threshold is improved in combination with the characteristics of algorithm and remote sensing image, which can effectively overcome the interference caused by cloud, waves and other background and minimize target omission. Finally, the noise and other obvious false alarms are filtered by a few features.

## 2. Research on detection algorithm of small and medium targets in wide remote sensing image and light ship target in complex background

This paper proposes a ship target detection method based on large area remote sensing images. The whole algorithm is divided into candidate region judgment and target fine detection. Because the wide image cannot be directly used as the detection network input, the image needs to be clipped to the size that can be input. In terms of candidate region location, a region extraction method based on image saliency based on the first part of the study is adopted to judge whether there is a suspected target in the clipped sub-image to ensure subsequent detection. In terms of detection network, in order to ensure the speed of detection, yOLOV4-TINY lightweight target detection model is selected as the main body, and the detection accuracy of the algorithm is reduced due to the large number of ships and small targets in the data. Attention mechanism and data enhancement module are added in the network, and adaptive prior box selection method is adopted to deal with the large number of small targets in the image and the changes in size and shape brought by data enhancement. The detection effect of the model for such samples is enhanced. Compared with the original yolov4-tiny network, the improved algorithm enhances the detection ability of the target under the condition of ensuring the detection speed. While ensuring the accuracy of detection, the recall rate is greatly improved and the missed detection is reduced. At the same time, compared with the pure target detection method based on deep learning, the overall algorithm not only ensures the accuracy, but also reduces the amount of calculation and improves the detection efficiency, which can be used as

a reference in practical engineering.

**Key words:** Large and wide remote sensing image, ship detection, frequency domain visual saliency, yolov4-tiny network

## 目 录

<b>第一章 绪论 .....</b>	<b>1</b>
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 基于图像处理的传统算法.....	3
1.2.2 基于深度学习的遥感目标检测算法.....	5
1.2.3 目前存在的问题和研究方向.....	6
1.3 本文内容安排.....	7
<b>第二章 遥感图像目标检测关键技术研究 .....</b>	<b>9</b>
2.1 引言 .....	9
2.2 遥感图像的候选区域提取.....	9
2.2.1 图像分析法.....	9
2.2.2 基于视觉显著性的方法.....	11
2.2.3 基于深度学习的方法.....	13
2.2.4 小结.....	15
2.3 基于深度学习的目标检测.....	16
2.3.1 深度学习概述 .....	16
2.3.2 卷积神经网络 .....	17
2.3.3 目前常用的目标检测网络.....	21
2.3.4 小结 .....	26
<b>第三章 基于谱残差模型的候选区域提取算法 .....</b>	<b>27</b>
3.1 引言 .....	27
3.2 谱残差显著性模型介绍.....	27
3.3 优化算法.....	29
3.3.1 全局显著图自适应尺度选择.....	29
3.3.2 基于直方图的自适应阈值分割.....	33
3.3.3 基于图像分析法的后处理.....	35
3.4 本章小结.....	37
<b>第四章 宽幅遥感图像舰船检测算法 .....</b>	<b>39</b>

4.1 引言 .....	39
4.2 算法整体介绍 .....	39
4.3 宽幅遥感图像的候选区域预筛选 .....	40
4.3.1 图像切分 .....	40
4.3.2 候选区域筛选 .....	41
4.4 目标精细检测 .....	42
4.4.1 主干网络介绍 .....	42
4.4.2 检测算法改进 .....	45
4.4.3 评价指标 .....	51
4.4.4 网络训练 .....	54
4.5 实验结果与分析 .....	57
4.5.1 候选区域预筛选阶段的结果与分析 .....	57
4.5.2 检测网络的结果与分析 .....	57
4.5.3 整体算法评估 .....	60
4.6 本章小结 .....	61
<b>第五章 总结与展望 .....</b>	<b>63</b>
5.1 全文总结 .....	63
5.2 后续工作展望 .....	64

## 表目录

表 3.1 选用特征及相关参数 .....	31
表 4.1 机器学习目标检测的分类标准 .....	51
表 4.2 实验平台参数.....	53
表 4.3 快速目标定位算法结果.....	56
表 4.4 优化方法对算法的提升.....	58
表 4.5 各网络性能对比.....	59
表 4.6 各算法计算时间对比.....	61

## 图目录

图 1.1 某港口遥感图像.....	1
图 1.2 传统检测方法处理流程.....	3
图 2.1 基于灰度的区域提取.....	9
图 2.2 基于面积和形状的区域提取.....	10
图 2.3 FT 算法处理流程 .....	11
图 2.4 SR 算法生成显著图.....	12
图 2.5 不同阈值对提取疑似区域的影响.....	13
图 2.6 RPN 网络流程图 .....	14
图 2.7 RPN 网络结构 .....	15
图 2.8 卷积神经网络结构.....	17
图 2.9 卷积计算过程.....	18
图 2.10 最大池化过程.....	19
图 2.11 全连接层分类实现过程.....	20
图 2.12 卷积层到全连接层的实现过程.....	20
图 2.13 Faster R-CNN 网络结构 .....	21
图 2.14 Faster R-CNN 网络处理流程 .....	22
图 2.15 YOLO-v3 网络结构 .....	24
图 2.16 Resnet 网络结构.....	25
图 2.17 FPN 特征金字塔和 Yolo head .....	26
图 3.1 候选区域提取算法流程.....	28
图 3.2 显著性算法流程.....	29
图 3.3 图 3.2 输入在不同尺度下显著图.....	30
图 3.4 输入图像及其梯度幅值图.....	31
图 3.5 在低尺度下的图像及梯度幅值图.....	32
图 3.6 显著图对比.....	32
图 3.7 显著图的灰度直方图.....	33
图 3.8 不同分割算法下的全局显著图.....	34
图 3.9 基于二值图像的区域提取.....	35
图 3.10 二值图像的后处理.....	36
图 4.1 整体算法流程.....	39

图 4.2 宽幅图像按固定大小分块.....	39
图 4.3 候选区域的提取和判定.....	40
图 4.4 yolov4 网络结构.....	42
图 4.5 yolov4-tiny 网络结构 .....	43
图 4.6 CSPnet 结构 .....	43
图 4.7 Mish 激活函数 .....	44
图 4.8 Mosaic 数据增强的输入图像 .....	45
图 4.9 Mosaic 数据增强效果图 .....	46
图 4.10 特征图的先验框分布.....	46
图 4.11 CBAM 注意力机制模块的处理流程 .....	48
图 4.12 CBAM 注意力机制模块的实现方式 .....	49
图 4.13 注意力机制模块在 yolo 网络的添加位置.....	49
图 4.14 IoU 的计算过程 .....	50
图 4.15 P-R 曲线 .....	52
图 4.16 采用的数据集.....	53
图 4.17 数据集制作.....	54
图 4.18 余弦退火学习率变化.....	55
图 4.19 网络训练.....	55
图 4.20 检测结果.....	57

# 第一章 绪论

## 1.1 研究背景及意义

舰船作为水上的主要运输载体，在交通运输，国土防御，资源勘探，事故搜救等领域都发挥着重要作用，各领域对舰船目标进行检测在军事和民用方面的需求也越来越迫切，检测的精度和速度对于实际应用方面更是极其关键的影响因素。当今，针对海上的舰船目标的检测技术突飞猛进，得益于空间遥感成像技术的快速发展。空间遥感成像技术是通过利用航天航空等空间平台，以电磁波为介质对采集的信号进行数字化成像的技术。近几十年来，空间遥感技术突飞猛进，成像分辨率在空间，时间，光谱上均有较大提高，可以从中区分出各种地物信息。图 1.1 展示了某港口的遥感图像，可以从中清楚分辨出靠岸的舰船目标。遥感图像的类型主要是由成像所采用的传感器来确定的，根据谱段的不同，可分为红外遥感图像，多光谱遥感图像，可见光遥感图像和合成孔径雷达（SAR）图像等。近年来随着空间成像技术的不断革新，遥感图像的成像分辨率不断提高，我们可以从高分辨率的遥感图像中直观分辨出各种地物的形态，在环境监测，资源勘探以及军事侦察，民用救援等领域上发挥越来越大的作用。



图 1.1 某港口遥感图像

Figure 1.1 Remote sensing image of a port

基于遥感图像的目标检测是空间遥感技术的重要研究方向。当前国际安全情况瞬息万变，同时，我国也正处于发展上升期，对海上安全的需求也愈发增加。在成像方面，我国遥感成像卫星技术虽起步较晚，但通过近几十年来的不懈努力，我国已经发射了“资源一号”，高分系列卫星等成像探测卫星，技术水平不断提高。影像空间分辨率从百米级提升到了米级，成像幅宽也越来越大，分辨率越来越高。无疑会对提高目标检测算法精度带来便利。但同时数据量也在急剧增长，为算法的检测速度和实时性也带来了巨大的挑战。

## 1.2 国内外研究现状

遥感技术的发展起源于 1957 年，当时苏联将世界上第一颗人造卫星送入了太空。1970 年我国发射了第一颗人造地球卫星东方红一号，自此也开始了对遥感技术的独立探索。利用获取的遥感图像进行信息的提取成为遥感技术的主要发展的一个领域。

SAR 图像和红外图像是最早采集到的图像数据类型，进而最早开始研究。目前在该领域提出的检测算法日趋成熟，在各发达国家均研制出高效的检测识别系统。在高光谱图像的检测领域，到目前为止，由于对高光谱成像技术的发展有限，造成相关数据难以获取、同时图像的空间分辨率不高，对高光谱图像的舰船目标检测研究难以成为热点研究方向。

基于全色可见光类型的图像进行舰船目标检测是目前一个热门研究领域。首先近年来随着半导体成像器件制造的技术提升和广泛应用，使得全色可见光图像的分辨率不断提高，同时数据获取更加容易。其地物信息明显同时符合人的视觉感官认知，舰船的特征也是具有良好的矩形度和对称性，易于设计算法对其进行特征描述和匹配。

国外的学者和机构提出了 eCognition<sup>[1]</sup>、SCORPIUS<sup>[2]</sup>等遥感图像信息探测系统。可以对传回的图像信息进行解译从而在各方面进行利用。我国的各大军事侦察机构和各大军事院校，也对遥感图像的目标检测进行了许多的研究，提出了大量相关的方法，主要基于数字图像处理技术对图像中感兴趣的部分进行提取和分析。近 10 年来，随着深度学习的迅猛发展，越来越多的学者将深度学

习的方法在本领域上进行了实践。为本领域的发展提供了新的途径和策略。所以，本领域的研究现状将从以上提到的两个方面来介绍。

### 1.2.1 基于图像处理的传统算法

传统方法主要是基于数字图像处理理论和经典机器学习知识，利用目标的特性，根据人工设计的特征或浅层的模型来训练特征最后通过机器学习分类器对目标进行检测。

处理流程为 two-step 的方式，首先是从图像中定位疑似目标区域（CTAL），其次通过分类器将第一步定位的区域做进一步分类（TCC）。如图 1.2 所示：

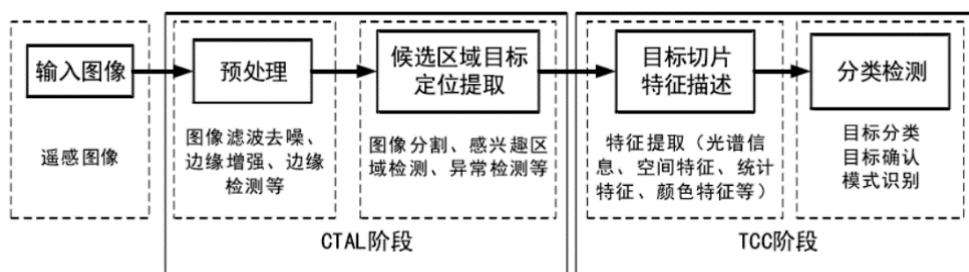


图 1.2 传统检测方法处理流程

**Figure 1.2 Processing flow of traditional detection methods**

在图像中定位疑似目标区域阶段，主要包括对输入图像的预处理以及对候选目标的定位提取。预处理包括对图像的滤波，增强等方法。预处理主要为了实现对图像进行降噪和增强，以提高输入图像的质量。图像滤波常用来处理成像传感器因各种原因引起的噪声干扰，同时，针对不同类型的图像，也会设计其他相应滤波器进行去噪处理。针对图像中的云雾等非噪声干扰，目前也存在一些算法获得很好的处理效果。何凯明等提出了基于暗原色的去雾算法<sup>[3,4]</sup>，能够对云雾进行很好的抑制，被许多研究人员作为遥感数据预处理的必要环节。

候选目标的定位提取可以理解为对输入图像提取感兴趣部分特征或者异常检测，最后通过阈值分割的方式来获得疑似区域的轮廓和在图像中的位置。主要作用是从图像中快速分离无关背景信息，并对可能性大的区域进行提取作为目标的候选区域，该过程对传统检测算法的检测效率至关重要。Proia and Page 以贝叶斯模型为基础设计针对小目标舰船的检测算法，有效提高检测精度<sup>[5]</sup>；

Xia 通过基于目标的几何结构，设计融合的轮廓特征，该特征对于海上的感兴趣目标能够做到较好的特征表述，从而高效地对海面的感兴趣区域进行提取<sup>[6]</sup>；

近些年来，有学者基于生物的注意力机制提出了一类视觉显著性算法。此类算法对生物对外界事物视觉感知的原理进行分析，通过设计相关算法模拟动物视觉感知的过程来对数字图像进行处理，以达到与生物感知相近的处理结果。目前提出了许多基于视觉显著性的模型如结合多个视觉感知特征的 Ft<sup>[7]</sup>模型，Itti<sup>[8]</sup>模型，基于视觉感知原理设计算法的 SR<sup>[9]</sup>模型，AC<sup>[10]</sup>模型，基于局部特征统计的算法 COV<sup>[11]</sup>，多方面结合的 GBVS<sup>[12]</sup>模型等。

分类检测的任务是对候选区域提取的结果进行精筛选，进一步剔除候选区域中存在的高疑似虚警，一般为与目标形似的云层，岛屿，海浪等背景。实现方式主要是对实际目标的正样本进行特征设计，之后对设计的特征训练分类器，最终通过分类器进行分类。常用的特征可分为人工设计的特征和基于区域统计的描述子。

常用人工设计的特征可以分为几类：以描述目标轮廓的几何特征，包括目标对象区域的长宽、周长、面积以及基于基本特征衍生出的长宽比，周径比等相关特征；反映图像与背景间灰度差异的灰度特征，包括对象的灰度值，对象区域的灰度均值方差以及对象区域与周围的灰度差异等；还有表述目标结构的纹理特征，包括对目标区域计算的灰度共生矩阵，能量、熵、对比度等结果作为特征进行表述<sup>[13]</sup>。

基于区域统计的特征描述子主要以 HOG 特征为代表<sup>[14]</sup>，HOG 又称方向梯度直方图（Histogram of Oriented Gradient）。主要通过计算图像各区域的梯度方向直方图来实现。研究表明，区域的梯度方向直方图可以很好地描述局部的特征，包括其纹理和形状。并且对同种类目标具有很强的匹配性，目前广泛应用于人脸识别等领域。HOG 特征的计算方式也为后续的基于深度学习的特征提取提供了思路，并沿用至今。

同时，还有多尺度分析，图像金字塔等<sup>[17]</sup>辅助优化方法，对目标检测效果的提升有很大助力。文献<sup>[15]</sup>基于传统的 Itti 模型，结合灰度共生矩阵和舰船的外形特征来对舰船目标进行检测，对单一目标检测效果较好，但无法解决紧靠

着的舰船的检测问题。Li<sup>[13]</sup>等通过采用自适应阈值的谱残差法结合融合特征的方式实现对复杂背景的舰船提取。Zhao<sup>[16]</sup>等首先获取根据统计学的基尼系数对不同尺度的显著图进行评估，获取多尺度下最优尺度的显著性图，之后引入径向梯度变换的方法提取舰船目标。Ju<sup>[17]</sup>等提出了一种在多尺度处理的船舶检测算法，首先在每个尺度上搜索潜在目标，设计相关特征判定目标是否为虚警，同时，加入了基于尺度的惩罚项，即同一目标两个以上的连续尺度下目标特征均保持不变，才会将其作为正例。若在连续的尺度下，判定出现变化，则将其置为虚警。在多尺度下检验目标的真实性，使得检测精度大大增加。

### 1.2.2 基于深度学习的遥感目标检测算法

传统算法以人工设计的能形象描述特定目标的特征为核心，在能够有效检测出目标特征的同时，也可以大大降低计算量。但也因为其轻量化的特点，人工选择的特征模型的容量往往较小，对于表征性差的目标，对其特征的描述就会比较模糊，比如在复杂场景干扰下的目标检测，采用传统的方法，检测精度就会快速下降。虽然也不断提出各种新的人工设计特征，但由于其本身的局限性也难以取得大的突破。同时，随着科技的发展，各种硬件的计算能力也取得了飞速提高，传统方法的快速性优势也渐渐减弱。所以，基于近年来兴起的深度学习算法来进行目标检测成为一种新的潮流。深度学习方法可以有效解决人工设计的不足，还可以实现自动提取特征。同时，由于其模型具有很大的容量，可以对目标进行详尽的特征描述，大大提升了检测的精确度。

现有基于深度学习网络框架的检测方法由于其处理步骤的不同可分为以下两类：

(1)双阶段的检测方法：此类方法按照传统方法的结构来进行。首先对图像整体进行遍历，生成多个候选框，再对候选框进行分类，判断区域内是否存在目标。获得数目更少，存在目标可能性更大的建议框，最后通过神经网络对目标进行精细定位和分类的操作。常用的检测模型有 R-CNN，Fast R-CNN，Mask R-CNN 等。

(2)单阶段的检测方法：以既定位置和大小在图像各区域生成先验框，在通

过判别网络，对先验框进行验证，找到存在物体的框并以预设的基准位置对位置参数进行细调节。常用的检测模型主要以 yolo 系列为代表。

近年来，深度学习的检测方法开始在遥感图像的目标检测领域进行实践应用，基于深度学习网络提出了许多改进策略，同时也制作了许多遥感图像数据集促进深度学习技术在本领域的发展。Li<sup>[18]</sup>针对深度学习网络对于小目标检测效果不佳的情况，采用先统一尺度后再检测的策略，预设一个检测器效果较好的尺度，在采用显著性相关算法对各种尺度的目标进行整合，统一到既定尺度后再进行检测。提高了检测精度，但也在一定程度下增加了网络的计算量。Yu 等<sup>[19]</sup>则通过与上述方法相反的方式在获取目标候选区域后，通过图像裁剪的方法获得一个目标在多个尺度大小的图像，通过增加多尺度的训练数据提高网络的检测能力。Wang 等<sup>[20]</sup>通过设计算法计算输入图像的最优尺度，将图像作最优尺度变换后再进行检测。文献<sup>[21]</sup>通过数据增强的方式获取不同尺度的数据以提高网络检测的适应性。Leclerc<sup>[22]</sup>等人通过迁移学习的方法基于 Inception 和 ResNet 架构的预训练权重进行网络训练，网络的精度较直接采用同一架构时显著提升，体现了迁移学习的优越性。文献<sup>[23]</sup>将民船样本作为数据集，主要为数据扩充提供了新的思路。文献<sup>[24]</sup>采用无监督学习方法对数据进行训练，一定程度上改善了正负样本分配不均的问题。Hu 等<sup>[25]</sup>为解决 Fast R-CNN 搜索过程较长的问题，采用传统视觉显著性的方法进行替换该步骤，能较好完成该部分网络的任务，降低了计算时间。文献<sup>[26]</sup>针对 Faster R-CNN 边界框采用全局遍历的方式导致生成数目过多存在计算浪费的问题，首先对控制边界框的参数进行整合再采用 K-Means 聚类算法对收集参数进行聚类，缩小边界框参数的取值范围使得产生的边界框数目更少，质量更高，提高了整体网络的检测效率。

### 1.2.3 目前存在的问题和研究方向

目前，对遥感舰船检测方面发展迅速，但仍存在一些问题，总结如下：

(1)对于传统的目标检测方法，其特点是计算量小，易于实现，但鲁棒性较差，对特征设计和重要参数的设置要求较高。对于不同种类的目标，也往往需要对特征进行大幅度修改。同时，人工特征的模型容量过小，也限制了其对于复杂场景下目标检测的性能提升。

(2)对于深度学习的方法，相较于传统方法的最大优势在于检测精度的提升。但其在应用性方面与传统方法相比仍有差距。首先，基于深度学习的方法目前多用来检测幅宽较小的自然图像，此类图像成像分辨率高，检测的目标往往存在于图中明显的区域且占比较大。对于卫星拍摄的宽幅遥感图像，有效目标的数据量较小。若直接采用此类方法会导致大量计算资源浪费。另外，算法本身对计算资源要求较高，对于卫星拍摄的宽幅影像，其由于分辨率越来越高，数据量也显著增加。对现有深度学习方法而言，若直接对全部图像进行处理，则很难保证算法的实时性，无法对算法进行迁移，必须要经过简化或修改。

基于目前存在的问题，如何更高效地使用遥感数据，提高算法的速度和准确率，最终实现星上的实时处理是本领域发展的必然趋势。鉴于深度学习的快速发展，采用深度学习为主题框架的目标检测方法成为已渐渐成为主流，不断改进以简化网络结构，提高检测效率同时降低计算量是算法提升的中心环节。鉴于此，本文提出了一种针对宽幅图像的小目标舰船检测算法，从输入数据量和检测网络复杂度两方面提高检测效率，首先利用传统方法的快速性将显著性算法作为数据预处理的一部分对输入数据进行预筛选以降低数据输入量，其次采用基于 Yolov4-tiny 的轻量化深度学习目标检测网络，从而高效快速检测目标。

### 1.3 本文内容安排

本论文主要对高分辨率宽幅遥感图像下的舰船目标的检测算法进行研究，在保证检测效率的同时，显著降低处理的数据量和计算量。具体工作是采用谱残差显著性算法对输入图像进行预筛选（首先将宽幅图像分成多个部分），剔除纯背景区域，显著降低输入数据量，其次采用基于 Yolov4-tiny 的轻量化深度学习目标检测网络，通过自适应先验框获取，Mosaic 数据增强，通道注意力机制等方式进行优化，提高对舰船对于复杂背景和小目标的检测能力。以获得高效快速的检测效果。

本论文共五章，除第一章绪论外，其余章节安排如下：

第二章主要对遥感图像目标检测主要技术进行介绍，包括候选目标区域提取和基于深度学习的目标检测技术两个大方面。分析各类方法的特点和适用场

景，为后续方法选用提供理论依据。

第三章提出基于谱残差显著性模型的候选区域提取方法。以谱残差显著性模型为基础，结合算法特点和遥感图像特性对尺度选择和分割阈值的计算方法进行改进，有效克服了云雾和海浪等背景带来的干扰，同时可以最大限度减少目标遗漏。

第四章提出一种针对宽幅遥感图像的舰船目标检测算法。将宽幅图像分割为多个子图，通过显著性算法对子图是否存在疑似目标进行判断，降低无效数据的输入，后续的目标检测网络主要基于 yolov4-tiny 架构对目标进行精细鉴别。针对数据集中存在小目标通过加入自适应先验框获取，Mosaic 数据增强，通道注意力机制等方式对网络进行优化。同时在训练中采用 label smooth 等训练策略使模型更好地收敛。进而实现对舰船目标较好的检测效果。计算时间较纯深度学习的方法明显降低。

第五章是对本文的总结与展望。首先对全文内容进行回顾和总结，并分析目前所做工作中存在的可继续改进的部分，最后分析目前本领域存在的问题和制约因素，展望了之后的研究方向。

## 第二章 遥感图像目标检测关键技术研究

### 2.1 引言

本章主要对当今遥感图像目标检测的关键技术分为两个部分进行介绍：疑似目标的区域提取方法和基于深度学习的目标检测技术，代表着当前目标检测技术的两个主要环节，本章将通过对遥感图像目标检测的关键技术的介绍，分析其特点和适用场景。并从中选择适宜宽幅遥感图像目标检测的技术方法。

### 2.2 遥感图像的候选区域提取

候选区域提取是目标检测中必不可少的环节，可以理解为一种粗略的区域检测。其主要作用是通过目标具有的几何，统计等能明显区分目标和背景的特征从而快速从输入图像数据中提取包含这些特征的部分作为疑似区域，从而降低后续处理的数据量和计算量。如何快速，准确地从大量数据中提取出含有目标的“有效”数据是对候选区域提取算法的一个重要要求。目前常用的方法大致分为三类：图像分析法，基于视觉显著性模型的方法，基于深度学习的方法。

#### 2.2.1 图像分析法

传统的候选区域提取方法往往采用图像分析的方式。通过分析图像的中背景和目标的某些特征存在的差异性来进行处理，设计合适的筛选方式，通过该特征将目标与背景分开。最经典的方法为基于灰度的图像分割。通过目标与背景的灰度范围存在差异从而实现对候选区域的提取。图 2.1 展示了阈值分割方法通过灰度分布的不同来对候选目标进行提取。通过阈值分割后的二值图像来区分候选区域和背景。



图 2.1 基于灰度的区域提取

Figure 2.1 Region extraction based on gray scale

同时，也可以采用较复杂的特征来表述候选目标，对于舰船目标，其采用的表述特征主要有面积，长宽比，紧致度，对称性等。下图显示了将区域的面积，形状等特性作为特征对目标区域的分析过程。可以看出，若合理设计目标区域特征进行候选区域提取，可以有效减少虚警，获得较好的检测效果<sup>[29]</sup>。

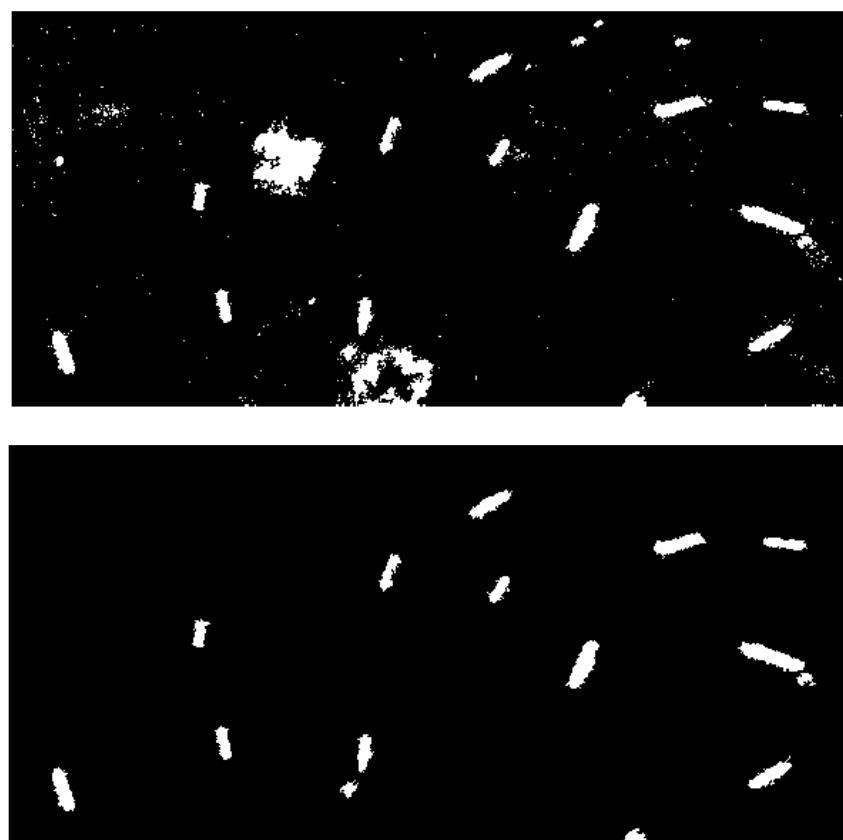


图 2.2 基于面积和形状的区域提取

**Figure 2.2 Area extraction based on area and shape**

图像分析法往往需要对于目标的特征有一定的先验知识，以此来确定采用合适的特征进行分析。此类方法可解释性强，计算复杂度较低，实现难度较小。已被广泛用于相关工程实践当中。吴飞等人对将舰船的头部作为特征提取的目标，设计特征并进入支持向量机进行训练，并通过图像分析法利用灰度与边缘特征提出明显虚警<sup>[27]</sup>。文献<sup>[28]</sup>使用了 SUSAN 算子，该算子可以描述目标的局部近似特征，结合局部邻域内的区域面积、占空比和灰度、方差均值等特征来提取舰船目标区域。但此类方法在背景较为简单的情况下效果较好，在复杂背

景的场景下，对采用特征的鲁棒性要求往往较高。目前提出的方法对于特定场景下有较好效果，难以获取各种场景中均有较好适应性的特征表述。

### 2.2.2 基于视觉显著性的方法

基于视觉显著性的方法是近年来提出的泛化能力较强的一类候选区域提取方法。显著性模型是一种提取场景中具有特征信息的算法模型，其理论来源于视觉注意机制。视觉注意机制本质上是一种生物机制，以人类举例，在日常生活中，我们往往要接触各式各样的事物，我们真正注意的，只有我们感兴趣的部位，即我们认为的显著性成分。而大量的无关信息被剔除。通过分析模拟生物注意到显著性目标的机理，设计出许多的显著性算法模型。

目前，现有视觉显著性模型有三类，第一类为基于视觉感知特征的算法模型，如 FT，Itti 模型等。

FT 算法由 Achanta 等人提出，主要利用了图像的色度特征。FT 算法从图像的颜色特征入手，保留了对视觉刺激性较高的颜色所在的区域，再通过低通滤波器，剔除了高频背景对图像的影响。算法流程如 2.3 图所示：

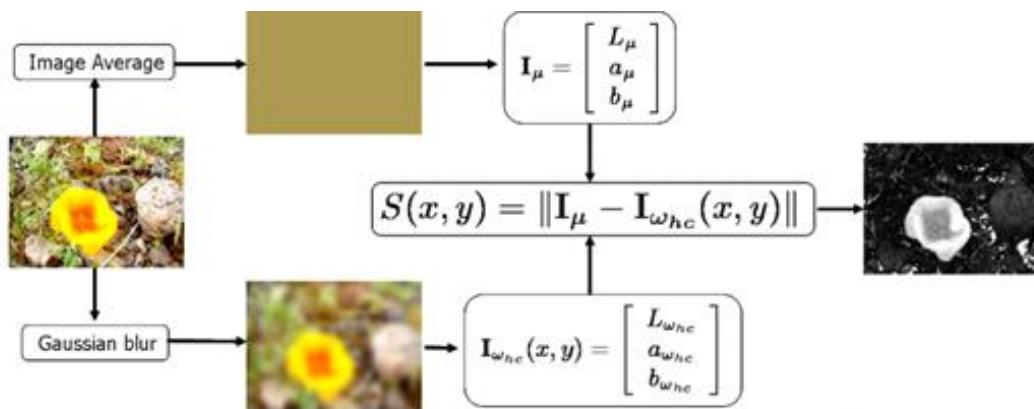


图 2.3 FT 算法处理流程

Figure 2.3 FT algorithm processing flow

Itti 视觉显著性模型是分析了动物对于不同亮度位置物体的关注机理设计的一种视觉注意力模型<sup>[1]</sup>。该模型从图像的颜色、亮度和方向的三个角度，利用高斯金字塔计算图像的亮度特征图、颜色特征图和方向特征图，结合多尺度的特征图来计算相应特征的显著图，最后合并得到最终结果。由于需要计算多个

特征类别以及多个尺度的特征图。 Itti 模型的计算量很大，计算速度较其他类型的显著性模型不具有优势。

第二类为纯计算方法，如谱残差法 SR 等。主要通过注意力机制，得知显著性高的成分类型，再通过设计算法保留图像中所需的成分。

在谱残差法的原理上来讲，作者从信息论的角度把图像分为有用信息和冗余信息。通过频域变换和滤波来提取显著性成分。图 2.4 展示了使用谱残差法对遥感图像计算显著图的效果。

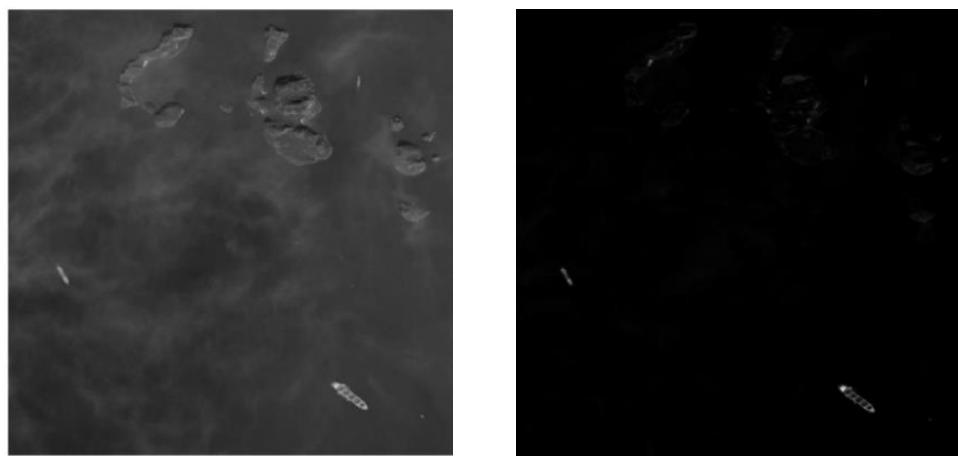


图 2.4 SR 算法生成显著图

**Figure 2.4 SR algorithm generates saliency graph**

最后一类为前两类的结合算法，典型算法为基于图论的方法 GBVS<sup>[12]</sup>。 Harel 等人在经典的 Itti 模型<sup>[6]</sup>上，提出了 GBVS 模型。GBVS 方法与 Itti 模型一样存在计算时间较长的问题。朱丹<sup>[34]</sup>等对 GBVS 进行了改进，大大降低了算法复杂度。

视觉显著性的方法较图像分析法来说，不需要对目标特征进行分析从而人工选择特定的特征，算法可以自动提取出输入图像内的显著性成分（往往是目标成分）。基于视觉显著性的方法相较于图像分析法计算复杂度略有提升，但泛化能力大大提高。显著性方法往往需要与阈值分割一同使用，由于注意力机制的存在，显著性算法提取的区域为图像中“显著”的区域，而并非只是我们需要的目标区域，这就导致了在复杂场景下，难以确定合适的阈值，这也导致了

采用显著性方法提取的疑似区域存在较多的虚警。这对之后的精细检测无疑增加了计算负荷。图 2.5 显示了不同阈值下对显著图的疑似目标区域提取效果。可以看出，阈值对于提取疑似区域的影响较大，左图阈值过小虚警较多，右图阈值偏大，目标的信息部分被掩盖。合适的阈值可以获得更好的提取效果及更少的虚警，对计算阈值算法的设计也至关重要。

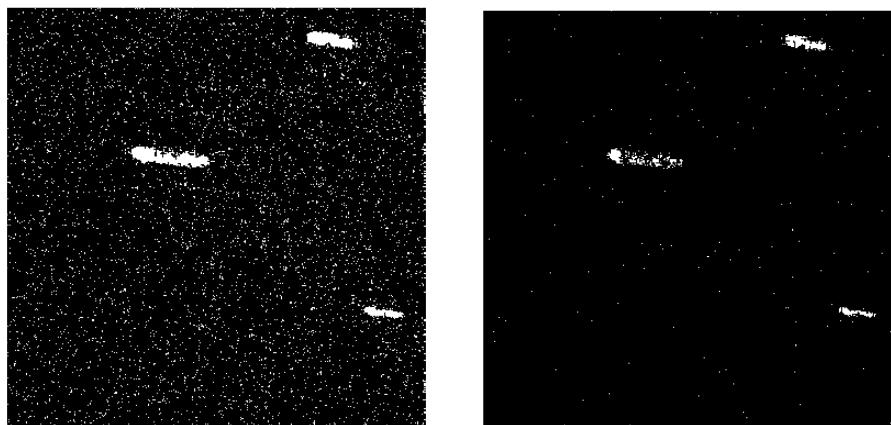


图 2.5 不同阈值对提取疑似区域的影响

**Figure 2.5 Influence of different thresholds on extraction of suspected regions**

同时，视觉显著性方法的提出往往是基于自然图像等基本不受噪声影响、图像颜色纹理变换较平滑的图像类型。处理此类图像采用显著性方法的效果较好。而宽幅遥感图像由于存在传感器噪声，抖动等各种对成像的影响，导致图像质量不佳，使得采用显著性算法处理的虚警数目进一步增加。同时对算法参数的设定也更加复杂。

### 2.2.3 基于深度学习的方法

在目标检测领域，出现了越来越多基于深度学习框架的方法，其处理流程大致与传统方法类似，候选区域提取作为目标检测的重要步骤在深度学习框架中也以模块化实现。基于深度学习方法的候选区域提取在实现方式和输出结果等与前两类方法都有很大差别。<sup>[30,31]</sup>

常用的基于深度学习网络目标检测方法大致可分为单阶段和双阶段两大类。而多出的一部分即为候选区域提取。深度学习网络的候选区域提取部分的处理输入为特征提取网络的输出，称为特征层。处理方案为根据预训练好的网络模

型，遍历特征层，区分目标和背景。

Faster RCNN 中的 RPN（Region Proposal Network）网络作为经典的候选区域提取网络之一。深度学习网络的输出框是以预设的锚框（anchor）为基础调节中心位置和长宽而得到的。由于需要遍历输入图片的各个位置，同时锚框的大小也需要存在与真实框接近的尺寸以使模型快速收敛。检测精细度较高的网络往往生成大量的预设框，从而增加计算量。RPN 网络的任务，就是对预设框是否存在疑似目标进行预筛选以减少后续进入判别网络中预设框的数目，减少整体的计算量。

RPN 的流程图如下：

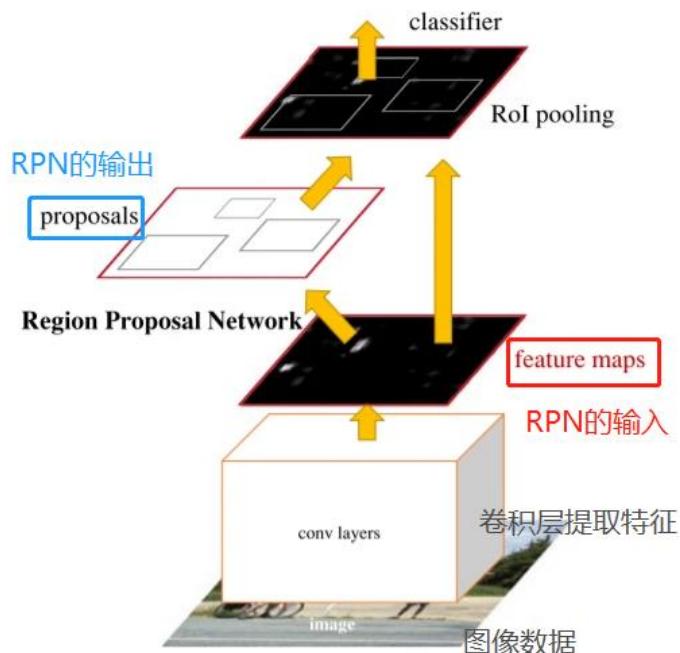


图 2.6 RPN 网络流程图

Figure 2.6 RPN network flow chart

具体步骤为：

- (1) 在特征图的每个特征点生成 9 个预设框。
- (2) 判断每个预设框是否包含目标，进行二分类。
- (3) 对预设框进行微调，使得调整后的预设框和真实框更加接近。
- (4) Proposal Layer 生成 proposals（可能包含目标的区域）。

RPN 具体处理网络结构如下图：

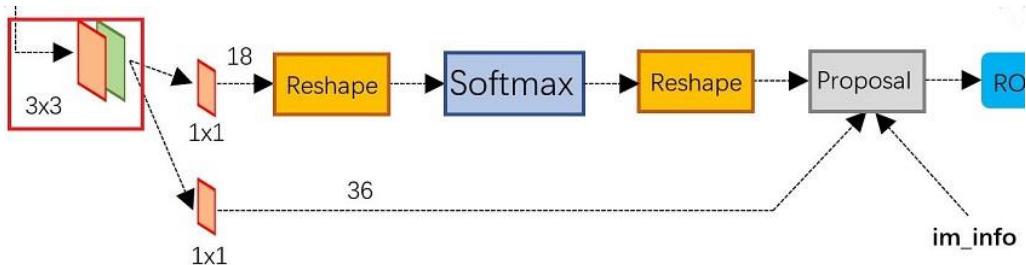


图 2.7 RPN 网络结构

Figure 2.7 RPN network structure

将之前特征提取网络输出的特征层（feature maps）作为网络输入。首先经过  $3 \times 3$  的卷积核进行卷积操作，得到的结果分别通过上下两个处理流程，获得对预设框是否存在物体的判断和位置上的调整。上方首先通过  $1 \times 1$  的 18 通道卷积层，其中 18 为  $9 \times 2$ ，9 代表特征图中每个位置的特征点中有 9 个不同大小的候选框，2 为置信度参数，判断该位置为目标或背景（两类），之后通过判断网络判定其类别。下方为  $1 \times 1$  的 36 通道卷积层，代表 9 个候选框的 4 维位置信息（候选框坐标位置和大小）。最后的输出（Propasal）称为建议框，是对目标进行的一个较为精细的预测结果。作为候选区域进行精细判断。

基于深度学习理论的候选区域网络可以明显提升网络的检测的精度。同时，此类方法往往需要结合前后的处理流程，不能独立作为模块使用，往往作为深度学习特征学习的一部分，其本身的参数量和计算量也远大于上述其他方法。这也是双阶段处理算法较单阶段算法运行时间大大增加的原因。

## 2.2.4 小结

本节对目标检测的候选区域提取部分进行了介绍，对于近年来常用的三类方法进行了简单分析，其中，图像分析法实现简单，解释性强。但在复杂场景下鲁棒性较差，并依赖人工特征选取，视觉显著性方法可以较为出色描述图像中的显著成分，有效分离目标和背景，但对于目标本身特征没有针对性表述，同时难以确定最佳阈值，导致提取虚警过多。增加后续计算量。基于深度学习的方法可以提高网络精度，但加入该模块较单阶段目标检测方法，计算量和检

测时间也会大大增加，在大幅宽场景下会造成大量计算浪费，同时，不易修改与迁移。

## 2.3 基于深度学习的目标检测

近年来，硬件研发水平不断提高，更精密的制造工艺也带来了巨大的性能提升，同时成本也在逐渐下降，这给更复杂模型的实现提供了更多可能。最具代表性的就是以卷积神经网络代表的深度学习的发展，在各个领域都取得了迅速的发展，在计算机视觉领域，基于深度学习算法不断推出新的模型框架，在各项指标上均领先于传统方法。本节将从深度学习的概念，卷积神经网络理论，常用的目标检测网络三个方面进行介绍，为舰船检测的方法奠定理论基础。

### 2.3.1 深度学习概述

深度学习由机器学习发展而来，目前已成为机器学习的一个重要的应用方向。受大数据和硬件开发的驱动，深度学习正迎来其快速发展的时期。深度学习的“深度”集中在模型的层数之“深”。深层次的模型理论上具有更高的泛化能力，拟合模型的鲁棒性更强，对测试的其他数据产生的预测结果可以保持在一个既定的区间，与简单的网络结构相比，深度网络的参数量显著增加，理论上在泛化能力方面就越好，能够处理更加复杂的任务。深度学习的出现，革新了各种情况下处理问题的方式，为许多领域的发展提供了新的思路。

深度学习的理论主要体现在卷积神经网络为代表的深度神经网络。在目标检测领域，2012年Krizhevsky等人提出的AlexNet<sup>[35]</sup>开创了深度神经网络架构的先河，其在各项指标上均领先于当时的其他方法，向世人提供了卷积神经网络在特征提取和分类任务的新思路，并促进了卷积神经网络的腾飞式发展。目前卷积神经网络在图像目标的分类和基于类别的图像分割、图像中的目标检测、目标匹配、对视频流的目标跟踪等领域获得广泛应用。各种专业的目标检测网络也在不断更新换代，基于深度学习的目标检测算法相较于传统的算法而言，主要体现在由于网络模型容量增加所带来的精度上的普遍提升，同时，同一网络模型使用不同的数据集进行训练即可在不同领域上进行应用，而不需要像传

统方法一样重新设计特征。体现了其良好的迁移性。基于深度学习的方法正逐步替代传统的目标检测方法，成为主流的目标检测方案。

### 2.3.2 卷积神经网络

卷积神经网络是深度学习网络目标检测模型的一种重要实现手段，深度网络大多是卷积层的堆叠加入少量池化层，激活函数层和全连接层共同实现。其结构如图所示：

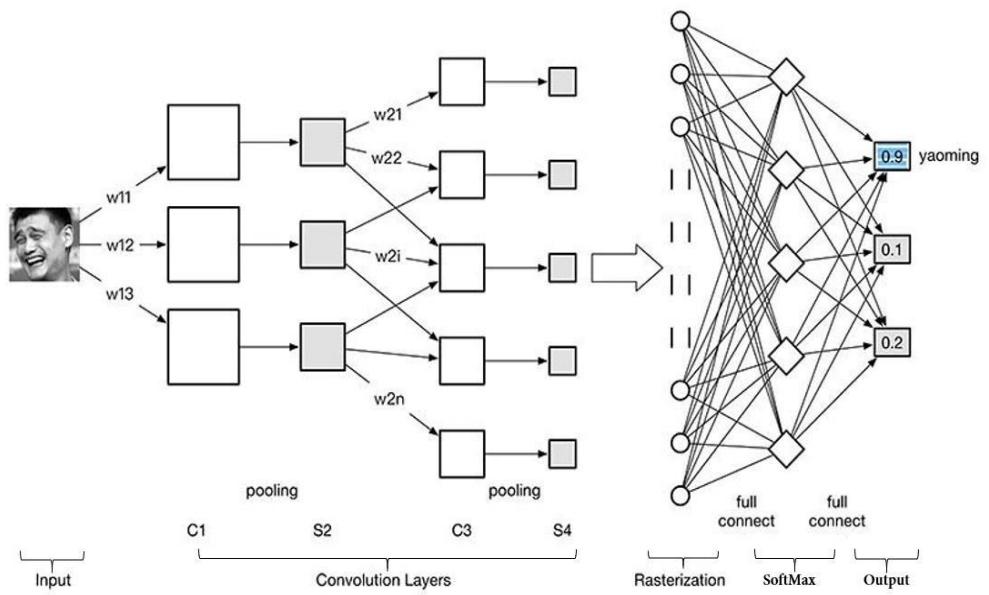


图 2.8 卷积神经网络结构

Figure 2.8 Convolutional neural network architecture

#### 2.3.2.1 卷积层

卷积层是卷积神经网络的核心部分，由一组可进行参数更新的卷积核（类似于数字图像处理中的各种算子）构成。通过对图像局部的卷积运算，可以提取该部分的特征。

深度学习网络的卷积部分基本运算多为对二维矩阵局部与卷积核对应位置相乘最后相加得到。如下图所示，输出的结果为输入与卷积核进行卷积得到。后续的结果与卷积的步长有关。步长代表卷积核每次移动的长度。下图表示步长为 1 的卷积。第二此处理的部分为第一次处理的区域输入向右移动一个单位得到。

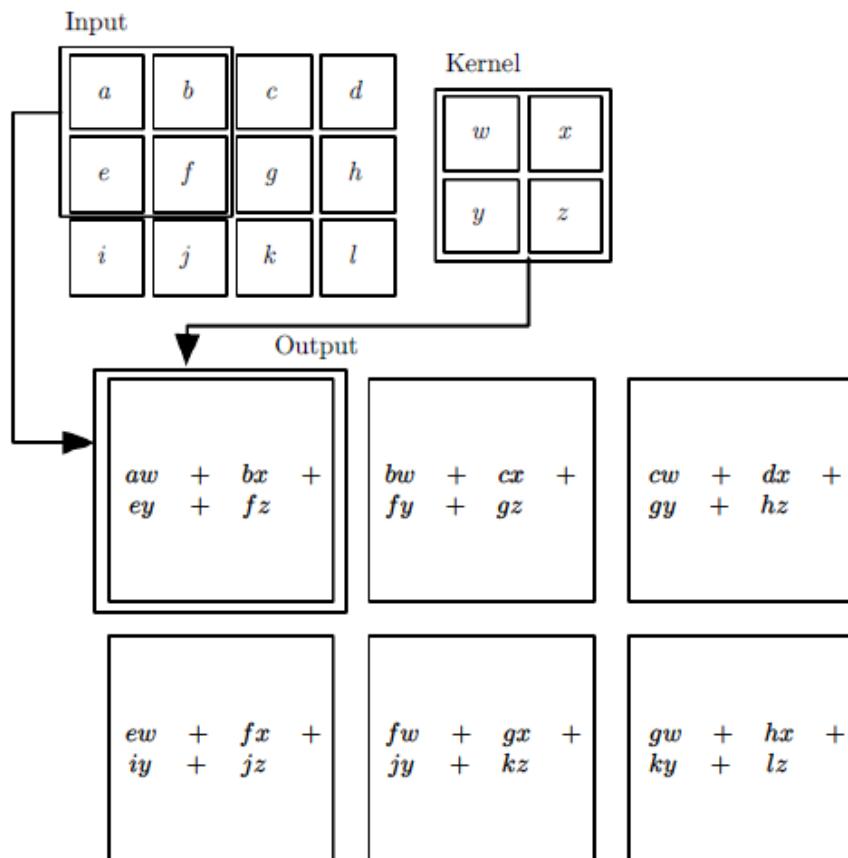


图 2.9 卷积计算过程

**Figure 2.9 Convolution procedure**

卷积层代表网络对输入局部特征的学习，而图像的特征也往往存在于局部的信息之中，采用卷积核的方式来提取特征，与全连接的方式相比较大降低了参数量，提高算法的效率，同时，在同一位置也可以存在多个通道的卷积核，以提取不同类型的特征。

### 2.3.2.2 池化层

池化层实际上可理解为一种非线性采样。主要作用为对局部特征区域进行浓缩，降低输出的大小，从而减少参数量，同时，对数据的“粗略”处理对防止模型的过拟合也有一定作用。另外，亦可为特征提供一定的平移不变性。实现方式主要有两方面：按对子区域的处理方式分为最大池化和平均池化。最大池化的计算步骤如下图所示。

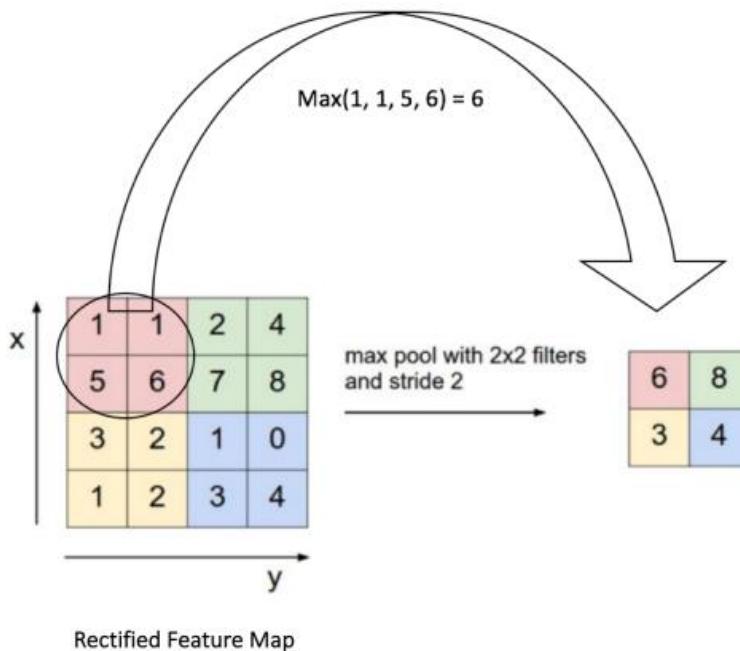


图 2.10 最大池化过程

Figure 2.10 Maximum pooling process

若大小为  $1000 \times 1000$  像素的图像经过  $10 \times 10$  的卷积核，步长为 1 的条件下进行卷积，输出特征图的大小为  $991 \times 991$  像素，然后使用  $2 \times 2$  大小的池化核进行下采样，即每 4 个点组成的小方块中，取最大的一个作为输出，正如上图所示情况。最终输出的图像大小降低为  $496 \times 496$  像素。在对特征图进行池化操作时经常会出现特征图的边长不能被池化核大小整除的情况，从而出现计算问题，需要对图像边缘进行增减以达到图像尺寸能够被池化核整除。处理方案如下：

- (1) 保留图像边缘，将缺少的部分以 0 填充；
- (2) 忽略边缘，即将多出的不足以满足池化核大小的边缘删除，或在设计算法时候改变输入图像的大小为特定值。

### 2.3.2.3 激活函数

激活函数是模型实现非线性拟合的重要手段。复杂的网络模型少不了对输入数据进行非线性映射，以达到非线性拟合的效果，将输入的特征空间转换为另一个空间，从而进行分类等操作，而激活函数就作为进行拟合的索引。主要的激活函数有：sigmoid, tanh, maxout, ReLU 等。

### 2.3.2.4 全连接层

全连接层（fully connected layers, FC）在整个卷积神经网络中起到分类器的作用，其思想在很早提出的多层感知机上已有应用，如图 2.11 所示，两层 4096 的特征层相互连接，最终输入一个  $n$  维的列向量，代表对目标所做出的类别概率预测。

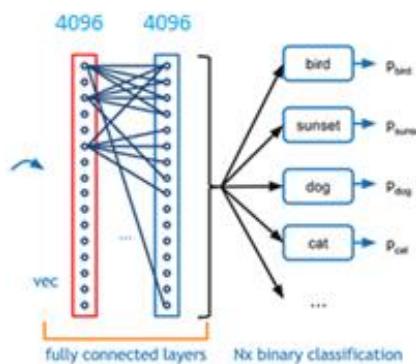


图 2.11 全连接层分类实现过程

**Figure 2.11 Full connection layer classification implementation process**

上图表示全连接层的实现过程。全连接网络的输入为一个  $3 \times 3 \times 5$  的矩阵。通过  $3 \times 3 \times 5 \times 4096$  的卷积层获得  $1 \times 4096$  的输出。具体实现过程如下：

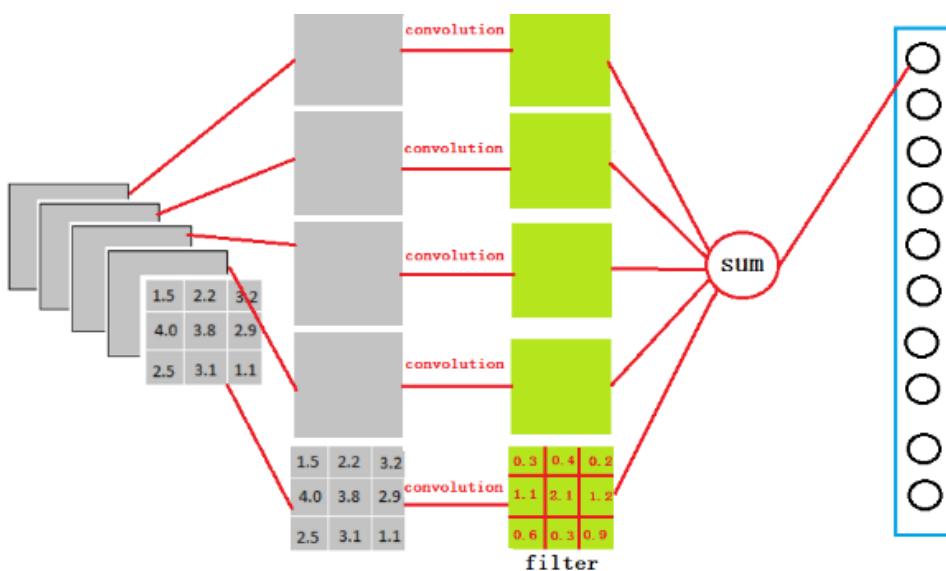


图 2.12 卷积层到全连接层的实现过程

**Figure 2.12 The realization process of convolution layer to full connection layer**

获得第一层  $1 \times 4096$  的输入后，就得到了输入层，之后通过隐藏层。从前一层获取输入，并整合每一个输入参数分配权重，将选定特征进行非线性组合。全连接层一般用于网络的末尾，用于输出分类结果。

### 2.3.3 目前常用的目标检测网络

基于卷积神经网络的基本内容，近年来也提出了许多类型的目标检测网络。按照处理流程可大致分为两类。分别是两阶段检测器和一阶段检测器。

两阶段检测器以 R-CNN 为代表，主要将目标检测分为两部分，首先生成建议区域，类似于传统算法的目标区域定位，之后对建议区域进行检测操作，对目标是否符合特征进行判定并生成最优候选框。一阶段检测器以 YOLO 和 SSD<sup>[51]</sup>为代表，其主要将二阶段的处理步骤集成成为一部分。下面以两种经典网络结构 Faster R-CNN 和 YOLO-v3 为例，详细阐述当前常用网络的实现过程。

#### 2.3.3.1 二阶段目标检测算法 Faster R-CNN

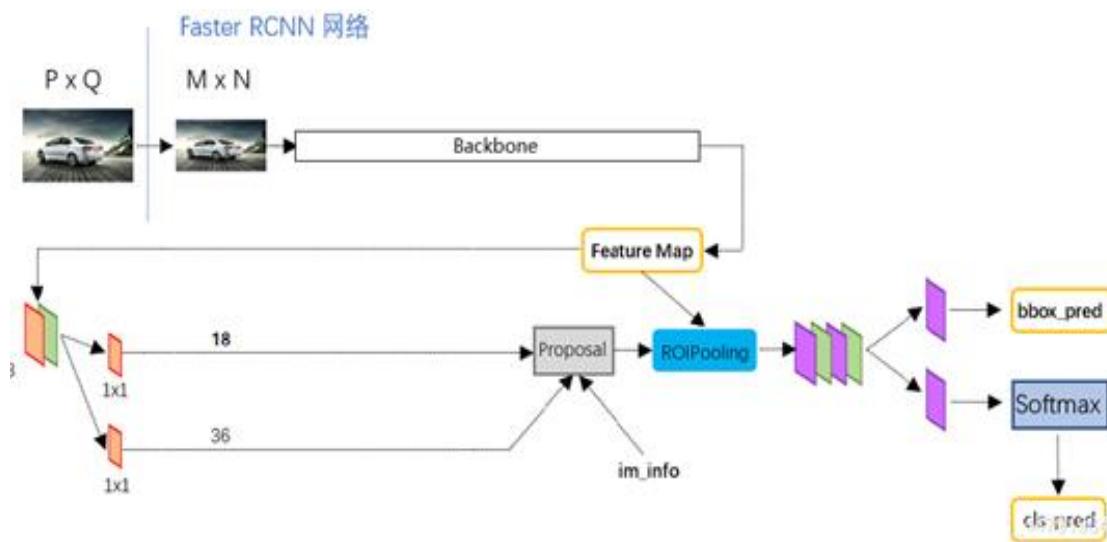


图 2.13 Faster R-CNN 网络结构

Figure 2.13 Faster R-CNN network structure

R-CNN 方法最早于 2014 年由 Girshick<sup>[41]</sup>提出，经过不断发展改良，目前已经提出基于 R-CNN 框架下的 Fast R-CNN，Faster R-CNN，Mask R-CNN 等改进模型，它们的实现方式不尽相同。本节以经典的 Faster R-CNN 算法为例，对二阶段检测算法进行介绍。

Faster-RCNN 作为一种两阶段的算法，与一阶段的算法相比，两阶段的算法更加复杂且速度较慢，但是检测精度会更高。双阶段相较于单阶段检测算法多了疑似区域提取的步骤，在 Faster-RCNN 网络结构中体现在 RPN 网络上，通过对大量预设框进行预筛选，获取存在疑似目标并且尺度合适的建议框。更利于检测网络的训练，使得网络更快收敛，提高了算法的检测精度，但也牺牲了一部分算法的运行效率。其处理流程如下：

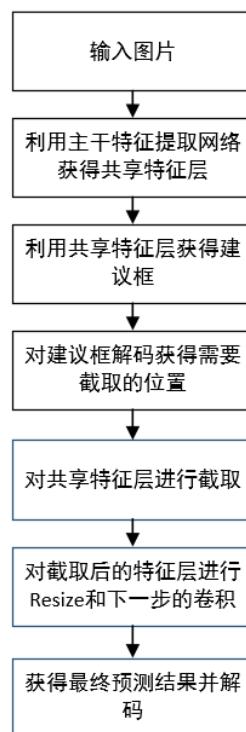


图 2.14 Faster R-CNN 网络处理流程

Figure 2.14 Faster R-CNN network processing process

首先，对输入图像进行标准大小的 resize，进入特征提取网络获取共享特征层。特征提取网络对应图 2.13 的 Backbone，Backbone 不是固定的，可以从主要应用的特征提取网络中选择。特征提取网络一般由多层的卷积层构成，对输入图像进行处理后生成可供后续判别网络处理的特征层，目前已提出如 VGG-16，Resnet，Xception<sup>[42]</sup>等特征提取网络可供选择。网络获得最后一层的输出为公用特征层。

获取的公用特征层有两个作用，如图 2.13 所示，将特征提取网络输出的特

征图首先是进行一次  $3 \times 3$  的卷积后，分为上下两个处理流程，上边部分进行一个 18 通道的  $1 \times 1$  卷积，下边部分进行 36 通道的  $1 \times 1$  卷积。获得 Proposal 建议框的相应参数。对应图 2.12 中的第 2 部分。两部分  $1 \times 1$  卷积的所代表的训练参数分别是：

36 通道的  $1 \times 1$  卷积用于公用特征层上每个网格点中每个先验框的位置参数调整。36 代表每个网格点中存在 9 个先验框的 4 维位置信息，包括框的中心坐标与长宽调整参数。

18 通道的  $1 \times 1$  的卷积主要对公用特征层上每个网格点上 9 个预测框内部是否存在包含物体特征的信息进而判断是否有疑似目标存在于该位置。

公用特征层的另一个作用是进行 ROI pooling。ROI 表示之前网络定位的兴趣区域。ROI pooling 层能实现对网络的显著加速，并提高检测精度。

ROI pooling 具体操作如下：

- (1) 根据输入图像，将感兴趣区域的坐标映射到 feature map 对应位置；
- (2) 将映射的区域按超参数设置的固定大小划分为多个子区域；
- (3) 对每个子区域进行最大池化（max pooling）操作；获得固定大小的输出作为新的局部特征图。极大降低了数据量，减少了后续网络的计算量。

最后会对 ROI pooling 层的输出（局部特征层）进行分类和回归操作，判断目标是否符合特征并确定其位置。（确定输出建议框的参数）进行完 ROI pooling 之后会通过分类预测网络，一般是以卷积神经网络和全连接层构成。输出目标的类别和目标框的坐标参数。完成目标检测的整体流程。

### 2.3.3.2 单阶段目标检测算法 YOLO-V3

YOLO 系列算法由 Redmon<sup>[43]</sup>等人提出，将目标检测任务简化为候选框定位和计算类别概率的回归问题。此类方法不需要像双阶段方法那样生成候选区域再检测，可以直接从输入中得到目标类别和具体位置。可以大大提高检测速度，为网络实现的实时性做出重大贡献。本节以 YOLO 系列的 YOLO-v3<sup>[44]</sup>为例，具体介绍此类方法的实现过程。

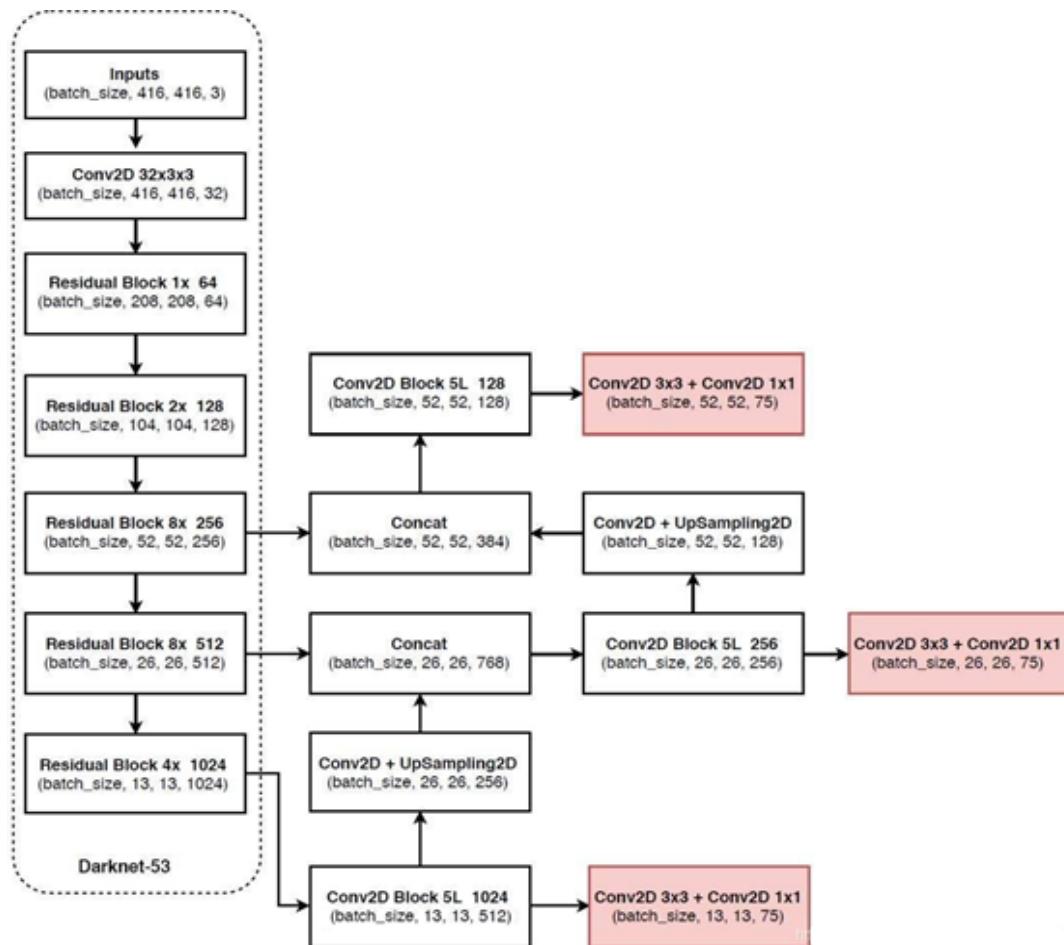


图 2.15 YOLO-v3 网络结构

Figure 2.15 YOLO-v3 network structure

YOLO-v3 网络的特征提取网络为 Darknet53，Darknet53 主要有两个特点：

Darknet53 第一个重要特点是使用了残差网络 Residual<sup>[36,37]</sup>，Darknet53 中的残差卷积对输入进行一次卷积核大小为  $3 \times 3$ 、步长为 2 的卷积操作，压缩输入进来的特征层的宽和高，卷积计算后可获得一个特征层。之后再对该特征层进行一次  $1 \times 1$  的卷积和一次  $3 \times 3$  的卷积，并把这个结果加上之前得到的特征层，通过多次类似卷积操作计算残差并进行叠加，使得梯度随着网络的加深不会出现保持极小的情况导致网络很难收敛，从而大幅度的加深了网络。图 2.16 展示了 Resnet 的结构。计算得到的残差部分与前面一层的输出进行叠加，保证通过残差网络后梯度不会急剧下降，缓解了因深度增加可能会导致的梯度消失。

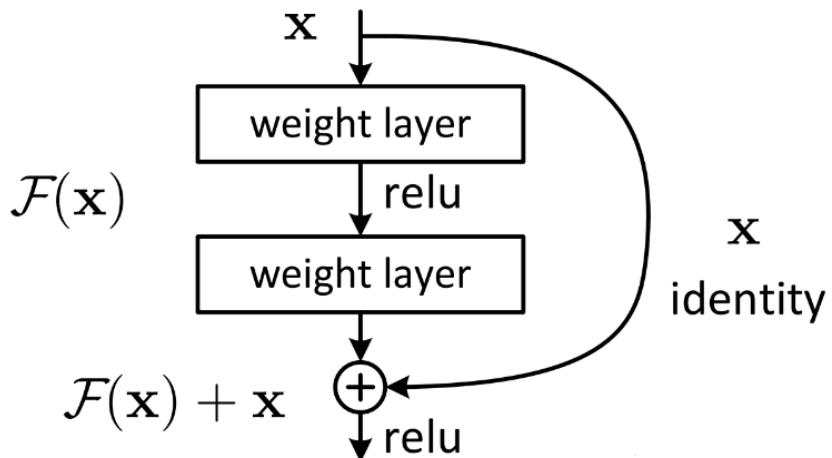


图 2.16 Resnet 网络结构

**Figure 2.16 Resnet network structure**

Darknet53 在卷积部分采用了新的 DarknetConv2D 结构，加入了 l2 正则化，和 BN 层进行权重标准化操作。在激活函数的使用上，最后采用 Leaky ReLU 激活函数代替常用的 ReLU 函数。Leaky ReLU 将输入为负数时以一定比例降低输出而不是像 ReLU 函数全设为零，使得仍有梯度返回。以实现公式可以表示为：

$$y_i = \begin{cases} x_i & \text{if } x_i \geq 0 \\ \frac{x_i}{a_i} & \text{if } x_i < 0 \end{cases} \quad \dots (2.1)$$

通过 Darknet53 获取特征层后，距离获取最终的检测结果还需要经过两个过程。首先通过构建 FPN 特征金字塔对目标特征进行在多尺度下进行增强。其次是预测，通过 Yolo head 模块获取预测结果。图 2.17 为特征 FPN 特征金字塔的结构，其中红色的部分为 Yolo head 的输出结果。

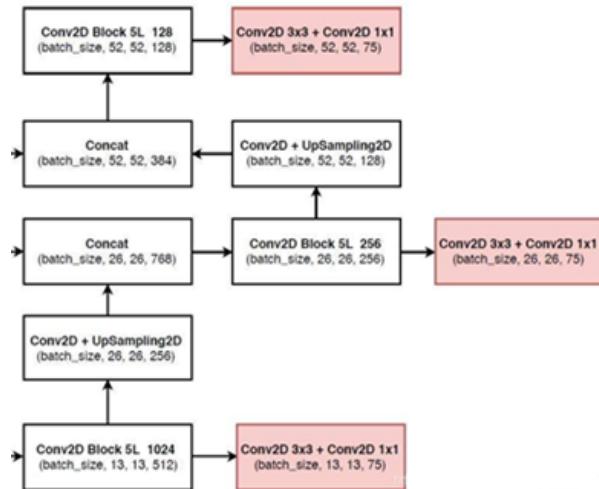


图 2.17 FPN 特征金字塔和 Yolo head

Figure 2.17 FPN features pyramid and Yolo head

特征金字塔的原理为通过提取不同尺度的特征来完善对特征的描述，不同尺度的叠加类似于金字塔结构以此得名。特征金字塔在 yolo 网络中可分为三层。从特征提取网络不同的位置的输出进行处理后构成。三个特征层的 shape 分别为(52,52,256)、(26,26,512)、(13,13,1024)。在获得三个有效特征层后，就可以利用其来完成 FPN 层的构建，构建方式为：从小 shape 的特征层开始先进行 5 次卷积操作，将计算后的结果之后，通过 Yolo Head 获得当前尺度的预测结果。

**Yolo Head** 图 2.17 中红色的部分。主要对输出的特征层进行一次  $3 \times 3$  卷积加上一次  $1 \times 1$  卷积，得到未解码的目标预测结果。其输出结果的含义如下：

假设输入特征层 shape 为  $a$ ，输出类别数为  $b$ ，则 Yolo Head 的输出为  $(a, a, (b+5) \times 3)$ 。其中 5 代表 5 维的参数，具体指先验框的位置（四维）和是否有目标，3 代表每个特征层的每个特征点先验框的个数。计算之后，会采用上采样的方式，与高 shape 的特征层进行融合，继续进行预测。

#### 2.3.4 小结

本节介绍了深度学习的相关知识，并由此引入目前常用的基于深度学习的目标检测网络。介绍了其特点和实现过程，为后续使用模型的架构提供相关的理论依据。

## 第三章 基于谱残差模型的候选区域提取算法

### 3.1 引言

通过第二章的介绍可知，候选目标区域提取是遥感图像目标检测的一个重要环节。算法能否快速准确提取出候选区域，对于整体的检测速度和精度都有重要的影响，遥感图像的幅宽越大，其影响也会愈发明显。目前对于遥感图像舰船目标候选目标区域提取算法所要解决的主要问题有两个，首先是算法的计算复杂度，关乎算法的运行效率，对算法的选择也要以检测目标的实际情况制定策略。文献<sup>[52-54]</sup>对飞行器目标进行了快速算法的研究，其在对图像中飞行器的定位是通过首先提取出图像中机场的位置来实现的，通过先找机场，从而找到临近的飞机位置，再对这些部分进行精细检测。由于机场面积大，形状规则，易于被定位算法提取。所以该策略广泛用于飞行器检测的算法中。对于舰船目标，其分布不止在港口附近，需要对全图进行搜索，所选择算法更需要以较快的速度从图像中找到疑似目标。其次是对图像中各种干扰的鲁棒性，遥感图像中的舰船检测中主要存在云雾，海杂波等地形干扰和传感器噪声和振动等成像过程中引入的干扰。针对以上问题，本章在谱残差显著性模型的基础上通过分析遥感图像的特点和算法的局限性从尺度和阈值选取上进行改进，再对显著性区域进行分析，并结合图像分析法去除明显虚警，最终实现对图像中疑似区域的提取。

### 3.2 谱残差显著性模型介绍

谱残差法（SR）是 Hou<sup>[13]</sup> 提出的一种基于频域分析的视觉显著性检测模型，经研究发现，图像的信息经傅里叶变换后都保存在图像的幅度谱信息中，各种图像的幅度谱均存在一定的线性规律，表现在平均对数幅度谱具有局部线性的特征。而图像平均对数幅度谱是在整体上存在波动的，用一幅图像的对数幅度谱减去其平均对数幅度谱，就得到了图像中存在的波动，即显著成分。这就是谱残差法的理论依据。谱残差法通过对图像频域变换后进行差分来可以有

效去除图像中的冗余信息，即海面，云层等背景信息，同时算法计算简单，实现速度也快于大多显著性模型。本文以谱残差模型为框架，结合视觉显著性的特点和同时分析遥感图像的存在的干扰。通过自适应尺度选择和阈值分割方法得到全局显著图，在通过图像分析法等对图像进行后处理过滤明显虚警，完成对图像的候选目标定位。

本章的算法流程图如下：

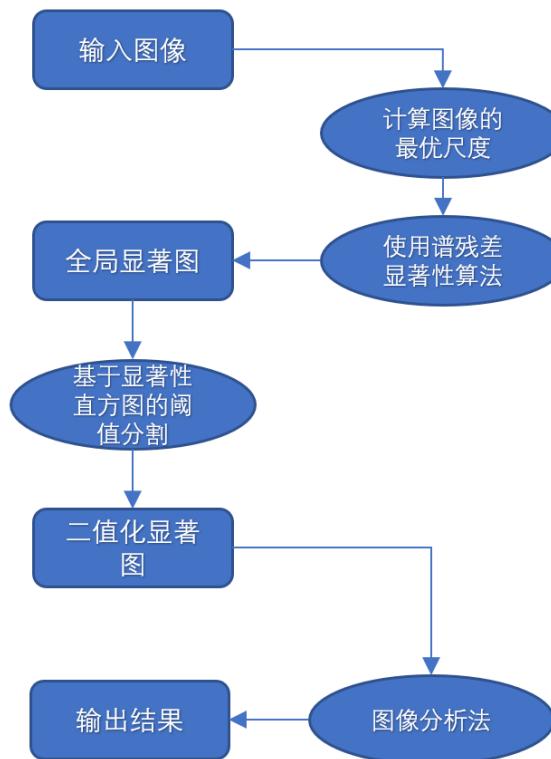


图 3.1 候选区域提取算法流程

**Figure 3.1 The flow of candidate region extraction algorithm**

算法显著图的计算需要通过基于谱残差模型的计算方法实现，过程如下：

- (1) 将图像通过傅里叶变换进行频域转换，获得图像的幅度谱  $A(f)$  及其对数  $L(f)$ ，相位谱  $P(f)$  等参数。

$$A(f) = |F[I(x)]|, \quad \dots (3.1)$$

$$P(f) = \varphi(F[I(x)]), \quad \dots (3.2)$$

$$L(f) = \log(A(f)), \quad \dots (3.3)$$

(2) 在频域下采用均值滤波器 $h_n(f)$ 进行平滑，提取图像中的冗余信息。

$$V(f) = L(f)h_n(f), \quad \dots (3.4)$$

(3) 使用幅度谱的对数减去平滑后的结果得到谱残差。用谱残差代替幅度谱进行反变换。

$$R(f) = L(f) - V(f), \quad \dots (3.5)$$

$$S(x) = |F^{-1}[\exp\{R(f) + iP(f)\}]|^2, \quad \dots (3.6)$$

(4) 最后对复原图像进行高斯滤波，得到显著图。生成的显著图一般需要进行二值分割得到良好的可视效果。其流程如图 3.2 所示：

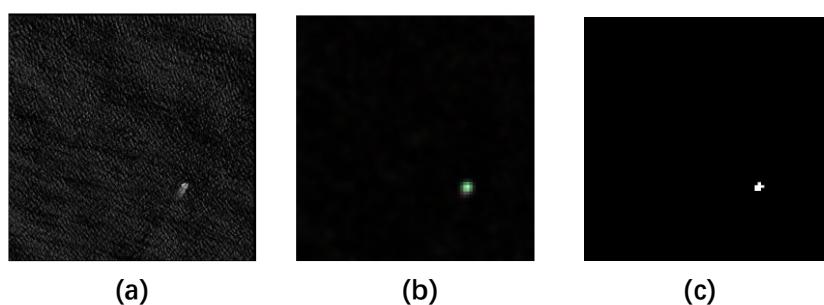


图 3.2 显著性算法流程

**Figure 3.2 Saliency algorithm flow**

其中，图 3.2 的 (a) 图代表输入图像，(b) 图代表通过显著性算法计算的显著图。(c) 图表示 (b) 图经阈值分割后的结果。

经典的谱残差模型主要应用于自然图像的显著性成分提取，遥感图像相较于自然图像来说，目标的特征没有自然图像丰富，同时，背景往往存在不同种类的噪声影响。直接采用会导致较多的虚警，使得算法效率下降。本文以谱残差模型为基础，从预处理，阈值选取等方面对算法进行优化，使得在复杂海况下获得较好的效果。

### 3.3 优化算法

#### 3.3.1 全局显著图自适应尺度选择

尺度代表一幅图像被放大或缩小的程度，显著性算法提取的目标成分是基于人的视觉注意力机制，所以在不同尺度下关注的显著成分会存在偏差，如同我们以不同的远近观察一幅景象，注意到的事物也会有所不同。在图像处理中

往往通过插值（默认为双线性插值）的方法改变图像的大小以达到改变尺度的效果。一般只进行下采样。

显著性算法的效果与图像的分辨率尺度息息相关，不同尺度下关注的显著性目标可能会存在差异<sup>[45]</sup>。特别是在复杂场景下的图像影响尤为明显，简单背景的图像尺度的敏感性不强。可以通过降低尺度以达到降低计算量的效果。同时高分辨率下存在的高频噪声，海杂波等背景也会导致虚警数目的增加，降低尺度也会对噪声进行抑制，同时实现速度也快于采用滤波方法。图 3.3 为图 3.2 输入图像在不同尺度下（分别为原图的大小的 0.8, 0.5, 0.1 倍）的显著图，由于图像中的海况复杂，在低尺度下目标较显著，同时噪声被抑制，效果较好。但尺度过小，也可能导致信息丢失，产生漏检等问题，需要根据实际情况确定最优尺度。为解决复杂背景图像自适应降低尺度的问题，本文提出一种基于图像梯度的自适应尺度选择算法。来获取最优尺度的显著图。

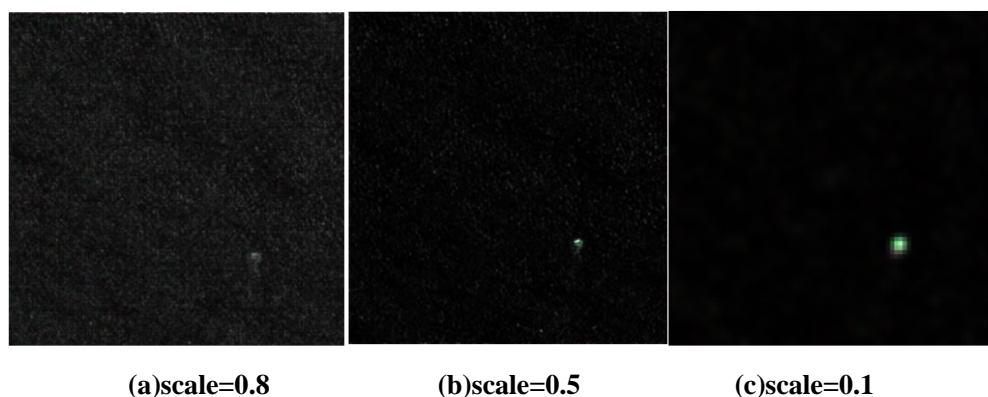


图 3.3 图 3.2 输入在不同尺度下显著图

**Figure 3.3 Significance diagram of input at different scales**

图像的梯度计算方法如下：

对于图像矩阵  $I$  中的任意一点  $I(x, y)$ ，其对应的梯度幅值  $A(x, y)$  可表示为：

$$A(x, y) = \sqrt{g_x(x, y)^2 + g_y(x, y)^2}, \quad \dots (3.7)$$

其中， $g_x(x, y)$ ， $g_y(x, y)$  分别为像素点  $I(x, y)$  分别在  $x$ ， $y$  轴方向的梯度值。其求取是通过图像矩阵分别与模板  $T$  及其转置矩阵卷积得到。

$$T = [-1, 0, 1], \quad \dots (3.8)$$

图像幅值矩阵 $A$ 的均值 $mean$ 代表图像的平均梯度幅值，其可以很好地反映图像纹理的复杂程度，进而判断图像是否存在复杂背景和纹理。如图 3.4 所示。

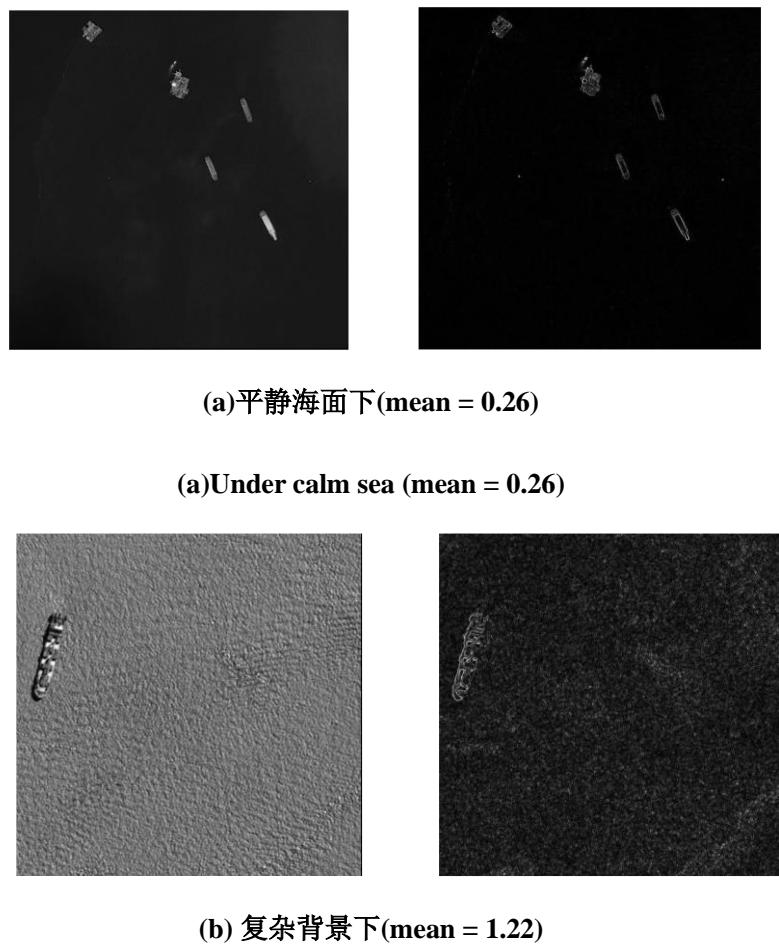


图 3.4 输入图像及其梯度幅值图

**Figure 3.4 image and its gradient amplitude map**

在背景信息复杂的图像中，其梯度矩阵的均值 $mean$ 远远大于平静海面的图像。而复杂图像中存在大量的高频成分会大大降低实际目标的显著性，需要对图像进行低尺度降维，使背景平滑，从而降低图像的平均梯度，减弱背景干扰。图 3.5 为图 3.4 (b) 图像的降维后的效果，此时尺度为 0.1，梯度均值为 0.52。与图 2 相比背景被明显压制，目标也更加突出。

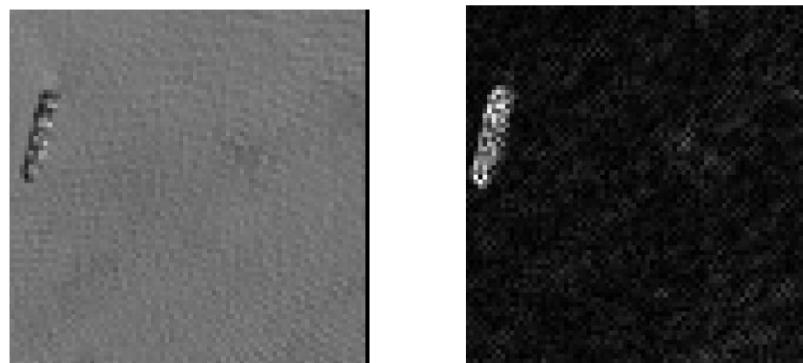


图 3.5 在低尺度下的图像及梯度幅值图

**Fig.3.5 image and its gradient amplitude map**

按照既定尺度范围的限制，以及对不同情况下样本的数据采集。测试得到图像的平均梯度在 0.2 附近时效果较好，定义其为标准平均梯度 $mean_{std}$ 。根据图像的平均梯度与标准值的接近程度来匹配最合适的尺度。设定尺度参数为  $S = [0.75, 0.5, 0.25, 0.1]$ ，选择的最优尺度索引 $c^*$ 为：

$$c^* = \arg(\min(|mean_{S(c)} - mean_{std}|)), \quad \dots (3.9)$$

式中， $mean_{S(c)}$ 是在 $c^*$ 索引代表的尺度下图像梯度的均值。最接近标准值 $mean_{std}$ 的尺度为最优尺度。图 3.6 分别表示图 3.4 (b) 中图像的显著图以及其在自适应最优尺度下的显著图。右图即为在最优尺度下的显著图，能够明显突出疑似目标。

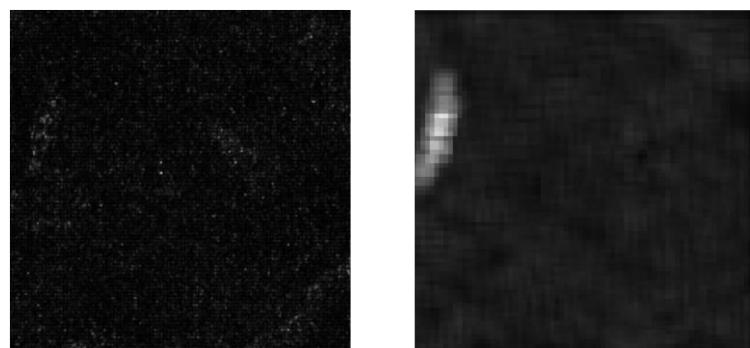


图 3.6 显著图对比

**Fig.3.6 Comparison of saliency maps**

### 3.3.2 基于直方图的自适应阈值分割

根据谱残差法的算法原理，取得的显著值集中分布在 0 和 1 附近。同时由于遥感图像低频背景部分占比较大，导致低显著值的部分显著值极小且占比很大。其显著性直方图如图 3.7 所示，图像的显著值集中在 0-10 的范围（归一化为显著值为 0.04）。使用常用的均值，方差等参数作为阈值不能很好地反映图像显著值分布的变化规律，而不同目标的显著值往往是多峰分布，常用的 OTSU<sup>[38]</sup> 阈值适用于二分类，往往导致所求阈值往往较大，容易遗漏目标。为了获得较好的分类效果同时尽可能避免遗漏。根据图像显著值的分布特征，本文采用一种基于直方图的阈值分割方法。鉴于图像中舰船目标所占体量较小。在总体显著图中，一般设定大小在前 10% 的像素点为基准作为阈值。其阈值对应显著图直方图大小在前 10% 对应的最小灰度级  $m$  所代表的灰度级数。

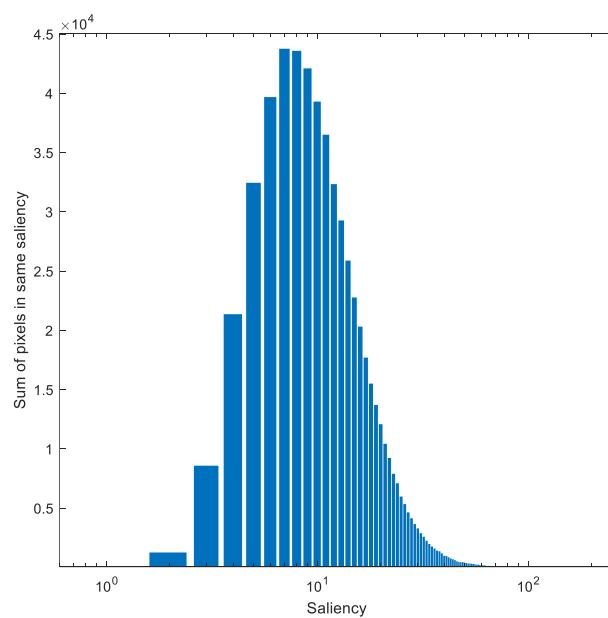
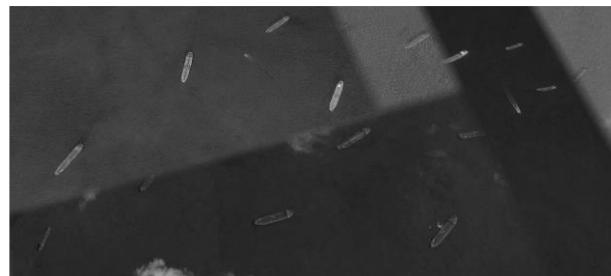


图 3.7 显著图的灰度直方图

**Fig.3.7 Grayscale histogram of salient map**

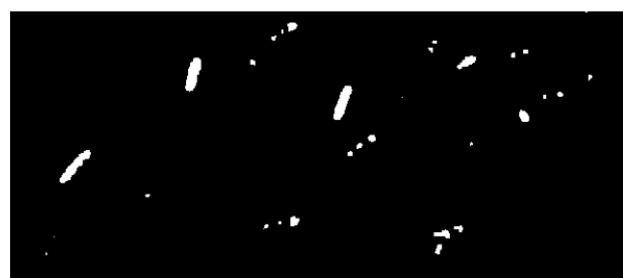
最后，将二值化的显著图还原为原来大小，即获得最优显著图。图 3.8 显示了某存在复杂背景图像的显著图通过各阈值分割方法的效果对比。可以看出，本文算法虽然存在一些冗余信息，但完整保留了疑似目标区域，没有出现目标遗漏，同时在控制虚警数目上也获得了较好效果。



(a) 原图



(b) 本文分割方法



(c) 文献[13]方法



(d) OTSU 方法

图 3.8 不同分割算法下的全局显著图

Fig.3.8 Global saliency graph under different segmentation algorithms

### 3.3.3 基于图像分析法的后处理

由于遥感图像中存在噪声，海杂波云层等影响因素，获得的二值化显著图存在大量虚警。需要进行后处理操作以去除明显虚警。本文采用一种基于区域特征的预筛选方法。本方法基于候选区域提取部分的图像分析法。主要针对舰船目标的特征判断图内是否存在疑似目标，通过选取目标和背景的特点选取差异较大且鲁棒性强的特征进行分离。最终选择采用面积、长宽比和紧密度对显著区域进行评估，若最终显著图内不存在符合要求的区域，则说明该部分为纯背景区域。

区域特征的提取主要基于对阈值分割后的显著图进行处理，分割后灰度值为 1 的部分为显著性区域。所有的特征均从显著图中存在的各个独立的不相连的区域计算而来。如图 3.9 所示。

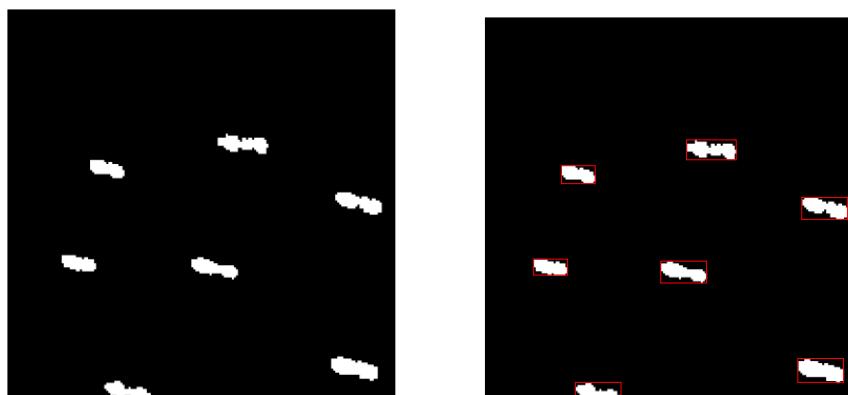


图 3.9 基于二值图像的区域提取

**Fig.3.9 Region extraction based on binary image**

选取的相应特征定义如下：

- (1) 长宽比  $l$ ，定义为疑似区域轮廓等效椭圆的长轴与短轴之比。
- (2) 面积  $a$ ，定义为疑似区域所包含的像素点个数。
- (3) 紧密度  $c$ ，定义如下：

$$c = \frac{Perimeter^2}{4\pi \times Area} , \quad \dots (3.10)$$

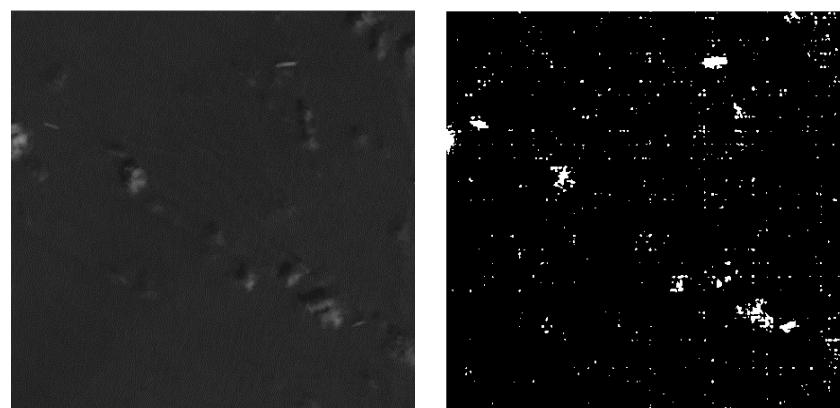
具体设置和区分标准如表 3.1 所示：

表 3.1 选用特征及相关参数

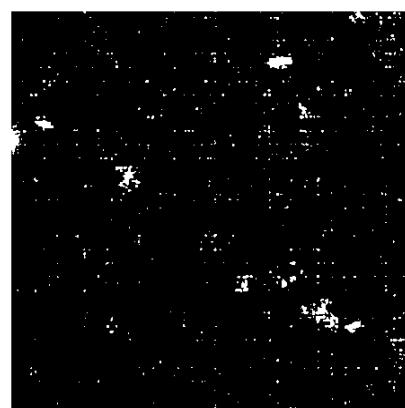
**Table3.1 Selection features and related parameters**

参数	疑似目标	虚警
长宽比 $l$	$1.5 < l < 10$	$l > 10$
紧密度 $c$	$1.3 < c < 5$	其它
面积 $a$	$60 < a < 2000$	其它

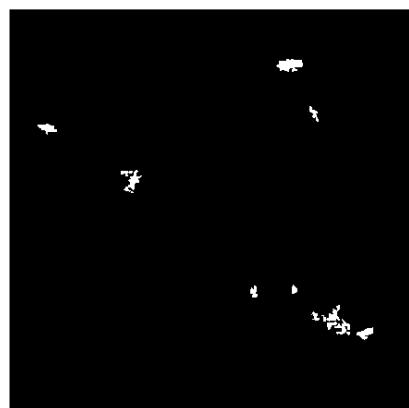
处理效果如图 3.10 所示：



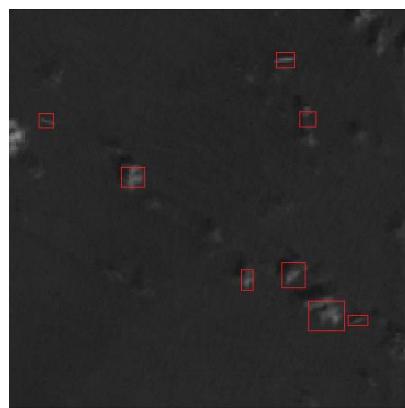
(a) 输入图像



(b) 二值化显著图



(c) 后处理后结果



(d) 区域提取结果

图 3.10 二值图像的后处理

**Fig.3.10 Post-processing of binary images**

从图中可以看出，本文算法的区域提取结果采取了较保守的策略，提取出

的区域中保留了一部分云层等背景区域，但在保证疑似目标没有遗漏的前提下最大限度减少了区域提取的数量。

### 3.4 本章小结

本章介绍了一种基于谱残差显著性模型并针对遥感图像进行改进优化的候选区域提取算法。首先通过计算图像的梯度判断图像的纹理复杂度，进而设置合适的尺度计算显著图，之后，根据谱残差计算显著值的值域分布，采用了根据显著值直方图提取前 10% 大小的显著性点作为阈值的选择方式，按比例保留显著性区域，最大限度避免了因阈值选取造成的目标遗漏。最后通过基于图像分析法的后处理剔除大量虚警。在存在海浪、云层、噪声等干扰下，能较好完成疑似区域的提取任务。



## 第四章 宽幅遥感图像舰船检测算法

### 4.1 引言

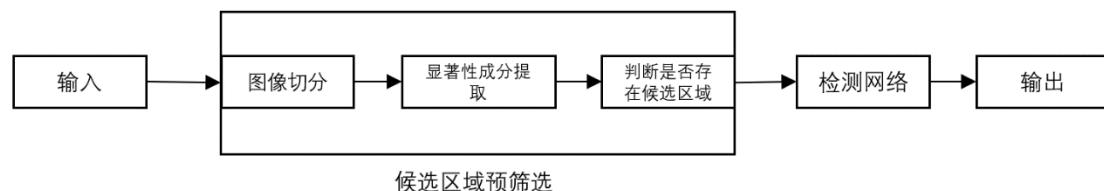
随着光学结构设计和制造工艺的不断进步，光学遥感器成像分辨率有了很大的提高。像素密度增加，单个像素能够存储更多的细节信息，使得目标的结构更加清晰和丰富，便于后续目标筛选与检测。同时，对遥感图像中目标的实时检测成为了本领域一个重要的发展方向。

目前对宽幅遥感图像进行目标检测主要存在的问题是解决算法和速度和精度的合理规划。宽幅遥感图像具有大幅宽，多背景的特点，同时，图像中的舰船目标比较分散，难以与其他图像中固定的地标物相关联，导致算法需要对全图进行遍历，导致算法的计算量剧增，为了提高速度，近年来提出的宽幅图像的舰船检测方法<sup>[45-50]</sup>仍以传统的方法为主，包括区域识别和精细检测两部分。但传统方法在检测器上依赖于特征的选择，鲁棒性较差，在精度上也低于深度学习的方法，基于深度学习的目标检测方法发展迅速，但由于网络结构较大，同时数据集较难获取。在宽幅图像的舰船检测提出的完整算法较少。近年来，随着各种轻量化网络的提出，在保证一定精度的同时，检测速度大大优于普通的检测网络，为深度学习方法在宽幅图像的检测提供了可能。

本文结合传统算法的区域提取和轻量化的目标检测网络，提出一种对宽幅图像中舰船目标检测的完整处理算法。

### 4.2 算法整体介绍

算法可分为两大部分，对宽幅图像的候选区域预筛选和目标精细检测。算法流程如图 4.1 所示：



**图 4.1 整体算法流程****Fig.4.1 Overall algorithm flow**

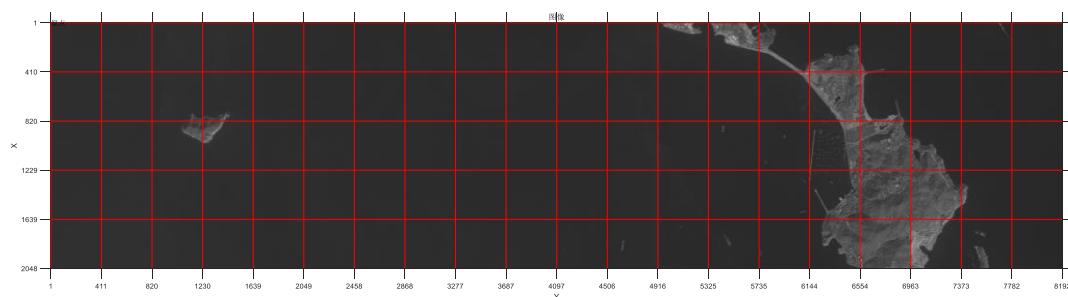
在候选目标预筛选步骤，采用了本文在第三章提出的算法作为候选区域提取方法，针对遥感图像的幅宽较大无法直接输入检测网络，加入了图像切分和候选区域判断等处理步骤。在目标精细检测部分，本文采用了 yolov4-tiny 轻量化目标检测网络为主干，针对宽幅图像中存在的大量舰船小目标进行优化，提高了网络对舰船的检测能力。

### 4.3 宽幅遥感图像的候选区域预筛选

宽幅图像由于存在大片纯海域背景区域，而这些成分对于目标检测是冗余成分，本步骤的主要作用即为从图像中快速分辨纯背景海域和可能存在目标的疑似区域，以减少对后续步骤的数据输入。

#### 4.3.1 图像切分

宽幅遥感图像较一般图像大小而言，由于检测网络的输入大小有限，不能直接对整个图像进行计算。需要对宽幅图像进行切分，将切分的子图分别计算。本文所使用图像大小为  $2048 \times 8192$  pixels，而检测网络输入的图片大小为  $416 \times 416$  pixels，再进行后续操作。图 4.2 为对宽幅图像的分割示意图。

**图 4.2 宽幅图像按固定大小分块****Fig.4.2 The division of wide image**

本文将单幅遥感图像按既定大小分割为 100 幅子图像，边角部分不足既定大小的，将图像尺寸改为既定大小作为输出。之后的候选区域筛选和目标检测网络的处理均基于子图进行。

#### 4.3.2 候选区域筛选

本文候选区域筛选模块的作用为提取每幅子图像的显著性区域并判断是否含有疑似目标，将可能含有疑似目标的区域的子图作为输出，同时作为后续检测网络的输入。本部分的任务较传统方法有所不同：

首先，对于宽幅图像切分后的输入图片，存在含有目标的正样本和只有背景的负样本，需要对正负样本进行判断。其次，传统方法的提取的目标区域，直接作为分类器的输入，而深度学习网络需要将整幅图像作为输入。

所以，本部分的任务可简化为对子图是否存在候选区域进行判断，基于第三章的候选区域提取方法来实现，本步骤的实现流程如下图：

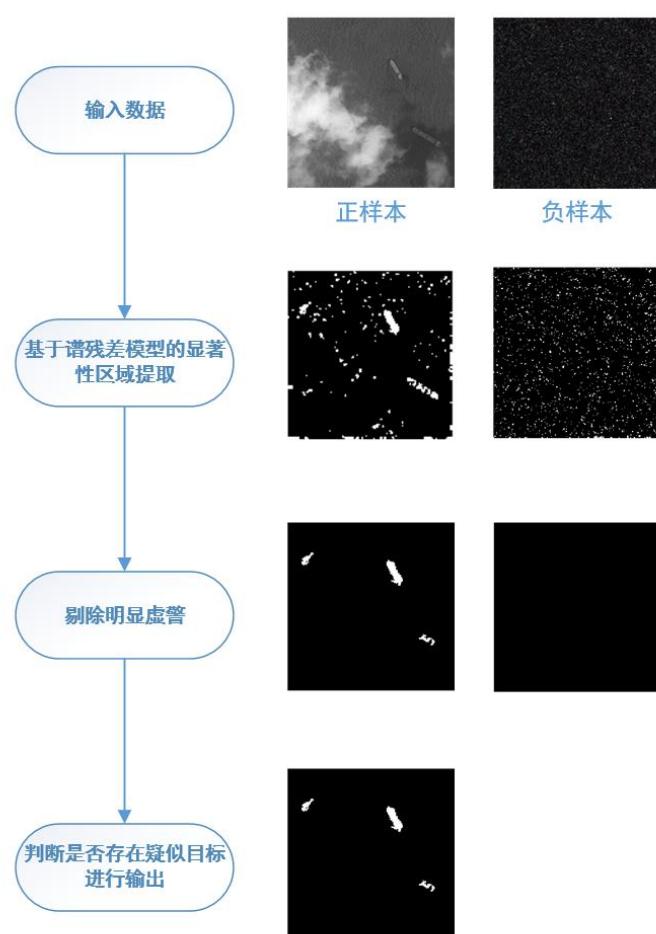


图 4.3 候选区域的提取和判定

Fig.4.3 Extraction and determination of candidate regions

对于遥感图像特点，本文在方法的使用上也做了一部分适应性改进，以对算法进行简化。

平静背景下，由于疑似目标较为突出，在降低尺度下其显著性不会发生太大改变。所以对所有输入图像均降低尺度，以提高计算速度。将尺度简化为 0.2 和 0.25。设定平均梯度为 0.5。大于阈值的判定为复杂背景，以 0.2 的尺度进行计算。其他图像以 0.25 的尺度进行计算，以降低计算量。同时，对显著图的分析以存在为判断标准，较之前的遍历进一步降低了计算量。

将裁切的子图经过显著性算法进行区域提取，再对输出的二值化显著图进行判断，标准为是否含有显著成分。将存在疑似区域的二值图对应的原图像作为检测网络的输入，若输出的二值化显著图中没有提取区域，则该部分判定为背景。不做后续处理。

## 4.4 目标精细检测

### 4.4.1 主干网络介绍

基于第二章的介绍，单阶段检测网络较二阶段网络处理速度更快，同时随着网络结构的不断优化，检测精度也不断提高，更适合对大数据量输入图像的检测。本文采用了 yolov4-tiny 轻量化网络模型作为主干，yolov4-tiny 为 yolov4 网络的精简版，在实现方法上保持一致。yolov4 目标检测网络采用了近年 CNN 领域中最优秀的优化策略，从数据处理、主干网络、网络训练、激活函数、损失函数等各个方面都有着不同程度的优化，较上一代的 yolov3 有巨大的性能提升。结构如下图所示：

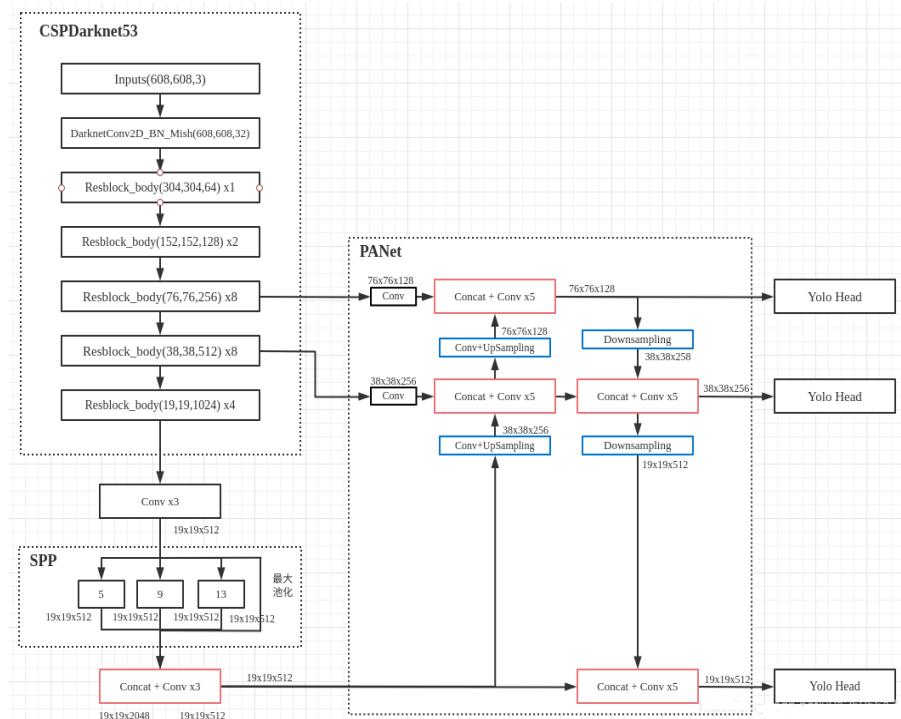


图 4.4 yolov4 网络结构

Fig.4.4 Yolov4 network structure

yolov4-tiny 作为 yolov4 的精简版，参数只有 600 万。是 yolov4 的十分之一，属于轻量化模型，这使得检测速度提升很大。整体网络结构只有 38 层，与 yolov4 相比，yolov4-tiny 在目标的分类与回归判别网络减少了最密集的一个特征层，是大大降低了网络的参数的主要原因之一。其结构图如图 4.5 所示。经实际测试，yolov4-tiny 模型在 COCO 数据集上获得了 40.2% 的 AP50、371FPS 的检测结果，相较于其他版本的轻量化模型性能优势显著。

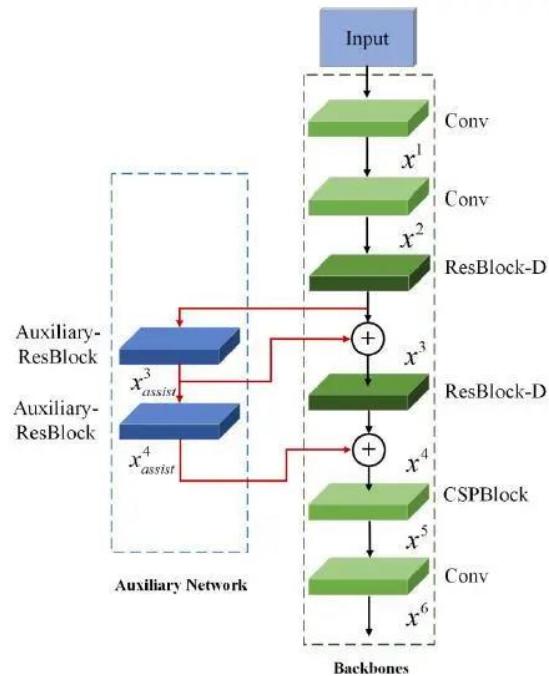


图 4.5 yolov4-tiny 网络结构

Fig.4.5 Yolov4-tiny network structure

yolov4 系列的处理结构与上一代的 v3 版本一致，分为特征提取网络，特征金字塔，回归预测网络 yolo head。但在许多部分都进行了改进和优化。

首先是对特征提取网络的优化，基于 yolov3 的特征提取网络 Darknet53 (backbone 为 Resnet)，在 yolov4 提出了 CSP Darknet53 结构。如图 4.6 所示：

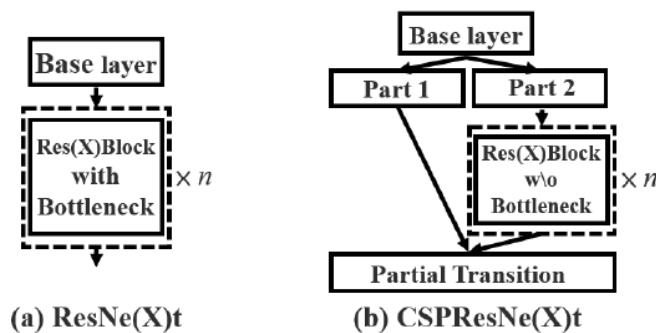


图 4.6 CSPnet 结构

Fig.4.6 CSPnetwork structure

CSP Darknet 结构较为简单，在 yolov3 的残差块进行拆分，分为左右两部分进行不同处理：主干部分继续进行残差块的堆叠操作；另一部分进行少量操

作后直接连接到最后。加入 CSP 结构较原来多了一个大的残差边。较原本的残差网络提高了对特征的提取能力。

同时网络的激活函数采用了 Mish 激活函数代替了原来的 Leaky ReLU。如图 4.7 所示：

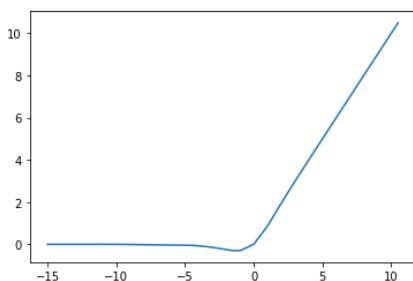


图 4.7 Mish 激活函数

Fig.4.7 Mish activation function

Mish 激活函数的定义如下：

$$\text{Mish} = x \times \tanh(\ln(1 + e^x)) \quad \dots (4.1)$$

Mish 的性能主要带来了以下优化：

- (1)无上界有下界：无上界主要为避免因训练急剧下降带来的梯度饱和。有下界属性有助于对模型的正则化，防止模型出现过拟合。
- (2)非单调函数：这种性质能够保持小的负值，使网络梯度流稳定性提高。防止因梯度为 0 造成拟合效果变差等各种问题。
- (3)连续性和光滑性：Mish 函数是连续且光滑的，保证了模型较强的泛化性能。

#### 4.4.2 检测算法改进

应用向遥感图像的舰船目标检测算法在检测性能上最重要的指标是算法的召回率，召回率代表检测算法对所有样本的检出率。漏检的数目越少，召回率越高。在实际应用中，精确率不足会造成对后续分析资源的浪费，但召回率不足则有可能遗漏重要目标，可能造成巨大的负面影响。所以，设计算法需要在保证精度的同时，尽量提高召回率。目前对宽幅遥感图像进行检测主要存在的问题有两点，一方面是存在海杂波和云雾等复杂背景信息干扰，另一方面是舰

船小目标居多，检测网络对小目标检测效果较差。所以本文主要从解决这两个问题出发，对检测网络进行优化，进而提高网络的召回率。提出以下三点对网络进行改进。

### 1) Mosaic 数据增强

Mosaic 数据增强方法的主要思想是将多张图片（一般为 4 张）进行随机裁剪，再拼接到一张图上作为训练数据。图 4.9 为图 4.8 输入图像采用 Mosaic 数据增强方法的输出结果。采用此方法，在一张图片下可以存在多个不同尺度的目标，同时其检测的背景也是千变万化的，可以提高网络对不同复杂背景和多尺度目标的学习能力。在训练时也会将这 5 幅图像都作为训练集，变相扩充了训练集。

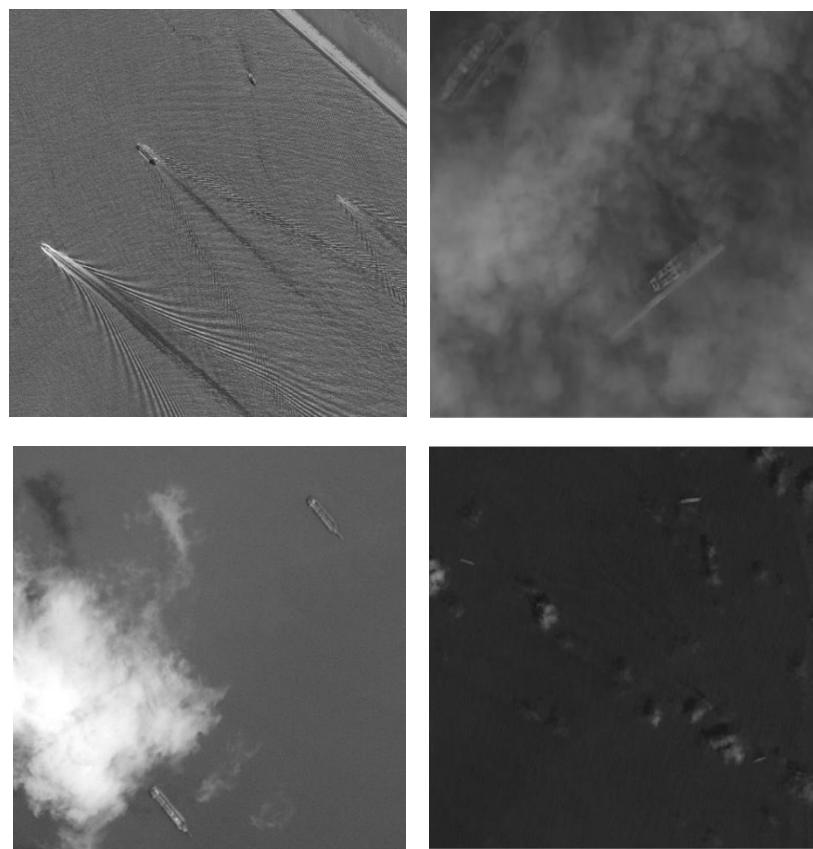


图 4.8 Mosaic 数据增强的输入图像

**Fig.4.8 Input images of mosaic data enhanced**

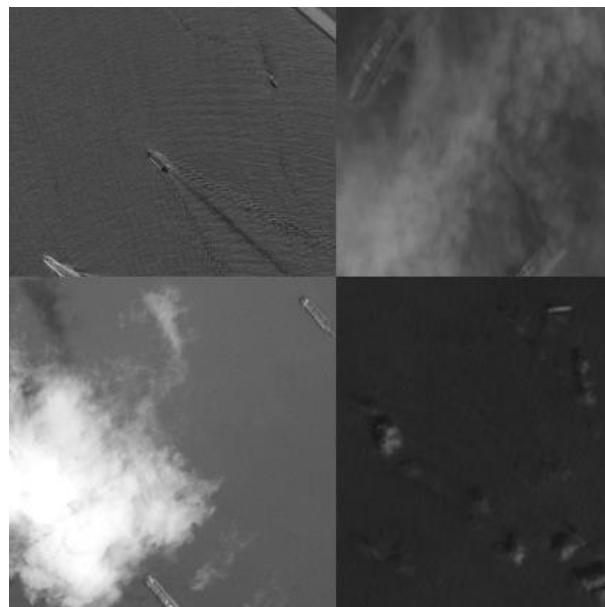


图 4.9 Mosaic 数据增强效果图

**Fig.4.9 Mosaic data enhanced rendering**

## 2) 自适应先验框获取

在 yolo 网络中，在检测阶段存在一些预设的锚框，称为先验框。如图 4.10 所示，图片被分成  $13 \times 13$  的网格，代表同等大小的特征图，即红色框代表的部分。每部分记录着输入图片在该位置的特征，其中蓝色部分代表着该位置特征图对应的先验框。黄色框代表真实框，检测网络的作用为调整先验框的参数，使得调整后的框与真实框的损失达到最小。所以，预设的先验框的尺寸和大小是至关重要的，相当于预测框的一个参考，基于这个参考，使得模型对预测框调整的幅度减小，使得模型更容易收敛。

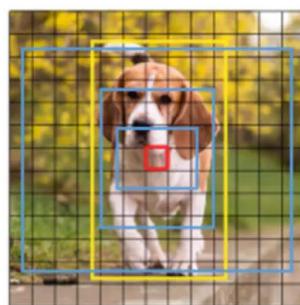


图 4.10 特征图的先验框分布

**Fig.4.10 The prior box distribution of feature graph**

本实验舰船检测的数据集，较 VOC 等数据集相比，小目标居多，属于小目标数据集，真实框的长宽主要分布在 10 像素到 50 像素之间，长宽比也大于其他数据集，同时在加入 Mosaic 数据增强的操作后，其中目标的大小会进一步减少。需要设定合适的先验框使网络更容易训练。

对于不同类型的数据集，预设框的大小也存在差异，需要重新确定预设框的位置。对于 yolov4-tiny 网络，共有  $13 \times 13$ ,  $26 \times 26$  两种尺度的特征图，每种尺度有三类预设框。预设框的大小由输入训练集中 ground truth (真实框) 聚类得到。同时为保证预设框符合网络的处理方式，对预设框的选择方法进行改进。

基本的预设框通过对正样本框的长宽进行按所需数目进行 K-means 聚类直接获取。这样虽然保证了生成的预设框为最接近所有训练集框样本的形状和大小，但没有考虑到长宽比的情况，而网络的损失函数为 CIOU，长宽比对损失有很大的影响。同时考虑到加入 mosaic 图像增强机制后的训练图像对目标的形状和大小进一步改变。所以对预设框的选择在尺度和形状上需要有更大的范围。所以，基于此采用了一种自适应先验框的获取方法。

步骤如下：

- 1) 采用 K-means 聚类方法获取需要预设框个数 2 倍类别（此处为 12）的聚类结果作为候选。
- 2) 在尺度层面上选取其中面积最大和最小的框作为预设框。代表两个尺度的边界。其余的 10 个候选框按面积大小归入两个尺度。
- 3) 在形状层面上，要使每个尺度都存在所有类型的长宽比要求 (1:1, 1:2, 2:1)。首先计算之前筛选出的预设框的长宽比。之后再从每个尺度中计算最接近其余两个长宽比的候选框作为预设框。

### (3) CBAM 注意力机制

提高特征提取网络对目标特征的学习程度也可以通过增加网络对特征的“注意程度”来实现，即加入注意力机制模块。注意力机制不仅应用于传统方法上，近年来在机器学习领域也成为深度学习常用的一个小技巧，它有多种多样的实现形式，尽管实现方式多样，但是每一种注意力机制的实现的核心都是类似的，就是注意力。注意力机制的核心重点就是让网络关注到它更需要关

注的地方。当我们使用卷积神经网络去处理图片的时候，我们会更希望卷积神经网络去注意应该注意的地方，而不是什么都关注，我们不可能手动去调节需要注意的地方，这个时候，如何让卷积神经网络去自适应的注意重要的物体变得极为重要，而网络认为“重要”的方式主要通过为主要目标分配高权重来达到。注意力机制就是实现网络自适应注意的一个方式。一般而言，注意力机制可以分为通道注意力机制，空间注意力机制，以及二者的结合。本算法在网络的检测部分加入了CBAM注意力模块，以提升整体的检测效果。

CBAM<sup>[39]</sup>将通道注意力机制和空间注意力机制进行一个结合，相比于SENet<sup>[40]</sup>等只关注通道的注意力机制可以取得更好的效果。其实现示意图如图4.11所示，CBAM会对输入进来的特征层，分别进行通道注意力机制的处理和空间注意力机制的处理。

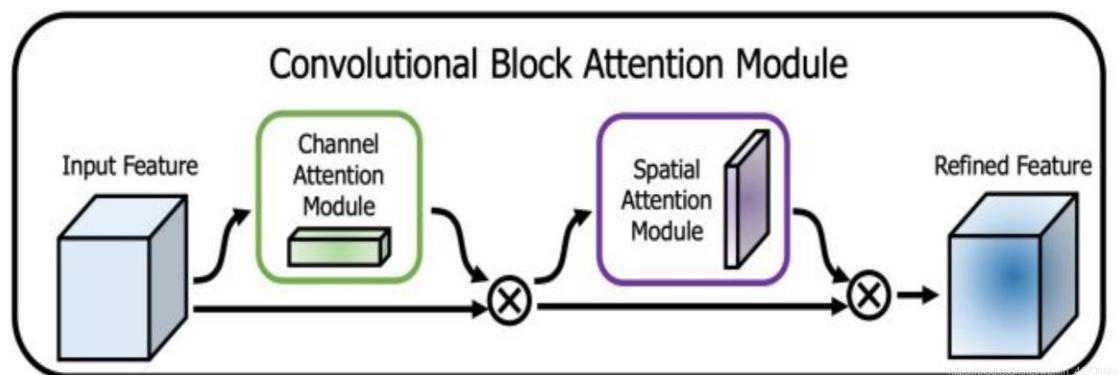


图 4.11 CBAM 注意力机制模块的处理流程

Fig.4.11 Processing flow of CBAM attention mechanism module

图4.12是通道注意力机制和空间注意力机制的具体实现方式：图像的上半部分为通道注意力机制，通道注意力机制的实现可以分为两个部分，我们会对输入进来的单个特征层，分别进行全局平均池化和全局最大池化。之后对平均池化和最大池化的结果，利用共享的全连接层进行处理，我们会对处理后的两个结果进行相加，然后通过sigmoid函数固定取值范围，此时我们获得了输入特征层每一个通道的权值（0-1之间）。在获得这个权值后，我们将这个权值乘上原输入特征层即可。

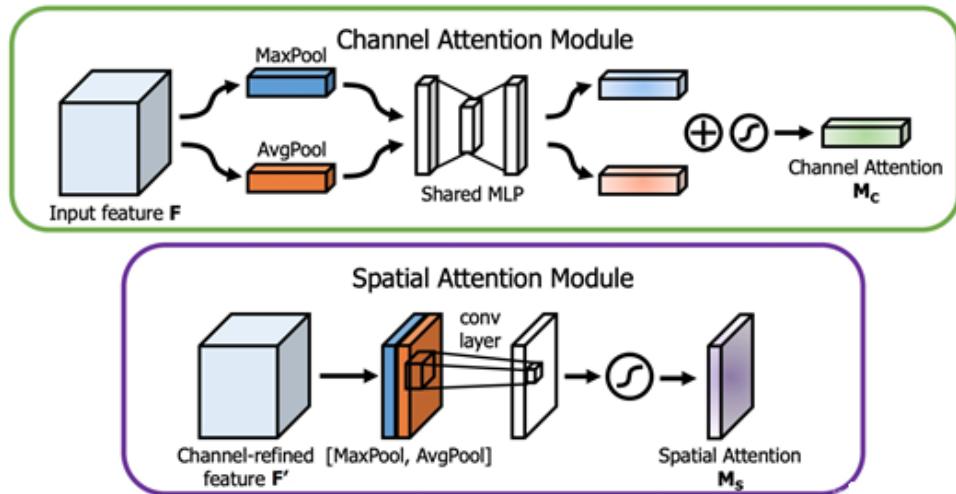


图 4.12 CBAM 注意力机制模块的实现方式

**Fig.4.12 CBAM attention mechanism module implementation**

图像的下半部分为空间注意力机制，对输入进来的特征层，在每一个特征点的通道上取最大值和平均值。之后将这两个结果进行一个堆叠，利用一次通道数为 1 的卷积调整通道数，然后通过 sigmoid 函数进行归一化，从而获得了输入特征层每一个特征点的权值（0-1 之间）。在获得这个权值后，我们将这个权值乘以原输入特征层，不同的权值即代表网络对不同特征层的“注意程度”。

本文在特征提取网络的输出与 FPN 特征金字塔之间加入了注意力机制模块，以对特征层进行增强。添加位置如图 4.13 所示：

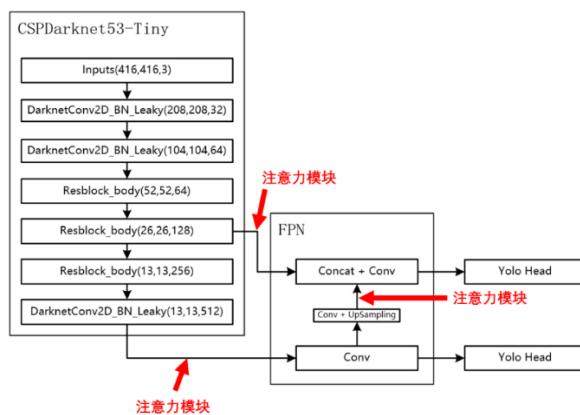


图 4.13 注意力机制模块在 yolo 网络的添加位置

**Fig.4.13 Added location of attention mechanism module in YOLO network**

#### 4.4.3 评价指标

##### (1) 每秒传输帧数 FPS

FPS 在目标检测领域指处理模块每秒处理图像的帧数，FPS 为衡量检测快慢的一个指标。反映相同条件下的各种方法计算量的大小，FPS 越高说明算法计算越快。

##### (2) 交并比 IoU

交并比是衡量深度学习目标检测网络精确度的一个重要指标，表示目标检测网络产生的候选框（candidate bound）与数据集标注框（ground truth bound）之间的相似度，以它们的交叠率作为指标。计算公式如下：

$$\text{IoU} = \frac{\text{area}(C) \cap \text{area}(G)}{\text{area}(C) \cup \text{area}(G)} \quad \dots (4.2)$$

图 4.14 代表 IoU 的计算过程，如图所示，IoU 体现了候选框与数据集标注框的重叠程度和偏移量，以此来衡量候选框对于标注的真实结果的差异性。若两类重合部分较多，未重合部分较少，则 IoU 较大，说明网络输出的候选框与真实框接近，检测精度较高。当它们完全重合，则 IoU 为 1。

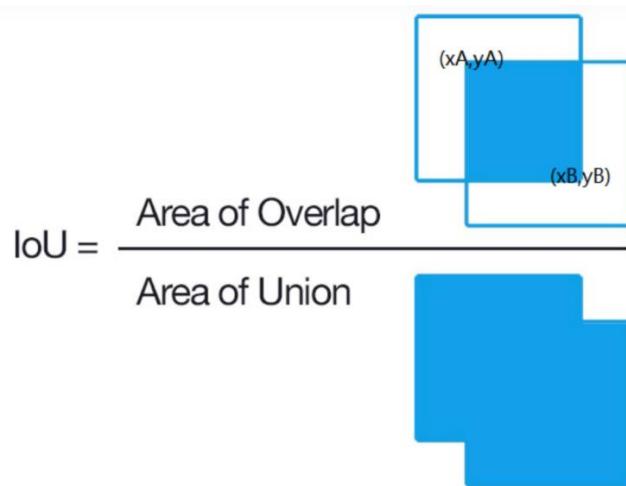


图 4.14 IoU 的计算过程

**Fig.4.14 Calculation process of IoU**

##### (3) 精确率和召回率 (precision & recall)

精确率和召回率是用来评估检测算法的两类重要指标。在目标检测领域中，

精确率是指算法输出的所有候选框中有多少是正确的，这反应出了算法对正类的检测能力；召回率是指在所有预标注框中有多少被检测出来，体现算法对正反两类样本的分辨能力。

在基于机器学习的分类领域存在以下标准：

**表 4.1 机器学习目标检测的分类标准**

**Table 4.1 Classification criteria of machine learning object detection**

类别	预测为正类	预测为负类
实际为正类	TP（真正例）	FN（假反例）
实际为反类	FP（假正例）	TN（真反例）

预测的标准在深度学习领域是根据其交并比大小来确定。常用的阈值T为0.5, 0.3等，检测框的IoU大于阈值则判定为正例。其中每种情况的判断标准如下：

**TP:** 检测正确的数量。

**FP:** 表示错误检测的数量。

**FN:** 没有检测到的标注位置候选框的数量，即漏检的数量。

**TN:** 该量无意义。

公式 4.3,4.4 说明了精度和召回率的计算方法：

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \dots (4.3)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \dots (4.4)$$

理想情况是检测精度和召回率都越高越好，但是两者一般不可兼得，如果希望提升精度，就会出现算法为了保证高精度的指标，减少返回更少的检测框，对正样本有更高的识别要求，从而减少误检保证精度，但这也会间接会提高漏检的数目，导致召回率的降低；同样，如果更偏好召回率，算法就会尽可能多增加返回检测框并降低目标检测的要求，以保证召回率，同样的，检测精度也

会有所降低，所以大多数情况下两者是不可兼得的，要根据实际任务的要求选择参数和指标。本文的任务是检测宽幅遥感图像中的舰船目标，在目标检测领域属于小目标检测的范畴，同时结合实际工程的要求，减少漏检提高召回率是本文对算法优化的主要方向。

#### (4) P-R 曲线与 AP, mAP

P-R 即 Precision-Recall 曲线，表示当阈值改变时，算法模型得到的 Precision 和 Recall 发生变化的情况。是一种常用的评估检测算法性能的方法，若曲线能反映出在 recall 增加的同时，precision 保持较高，表明这个算法的性能较好。相对的，随着 recall 值提升，但 precision 却损失严重，说明该算法的性能很差，通过精度来换取了召回率。P-R 曲线的示意图如下：

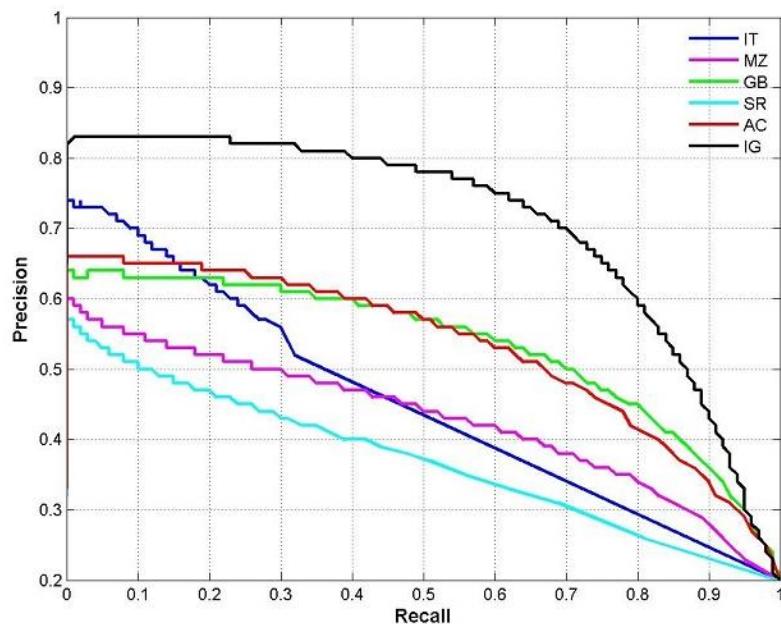


图 4.15 P-R 曲线

Fig.4.15 P - R curve

AP ( average precision) 是指 PR 曲线与坐标第一象限围成的面积，mAP 是指各类别 AP 的平均值。计算 mAP 首先需要分别计算绘制各类别的 PR 曲线，得到 AP，再求均值就得到了 mAP。本实验主要涉及到单一舰船检测，正样本只有一类，所以得到的 AP 与 mAP 相同。

#### 4.4.4 网络训练

本文的实验平台的软硬件参数如表 4.2 所示：

表 4.2 实验平台参数

Table 4.2 Experimental platform parameters

运行平台	Vscode
编程语言及版本	Python 3.8.8 64-bit
框架	TensorFlow2.2.0
GPU	NVIDIA Quadro RTX4000
CPU	Intel Core i7-10700H(2.9GHz)

本文实验的数据集来源于某型号空间相机获取的全色宽幅遥感图像。幅宽为  $2048 \times 8192$  pixels，按照检测网络的输入大小裁剪为  $416 \times 416$  pixels。并从中选取含有舰船目标的部分作为训练数据。由于用于训练的正样本有限，采用镜像、加入高斯噪声、灰度变化等方式进行数据扩充，最终得到 1861 张训练图像，如图 4.16 所示。剩余图像通过候选区域提取模块后，提取到 371 张含有目标的正样本。作为测试集。本文实验主要模拟实际场景下对卫星实拍获取的遥感图像进行舰船检测。其中图像中的舰船多为海上目标，并存在海杂波，云雾和成像噪声等相关干扰，能有效验证模型的鲁棒性。

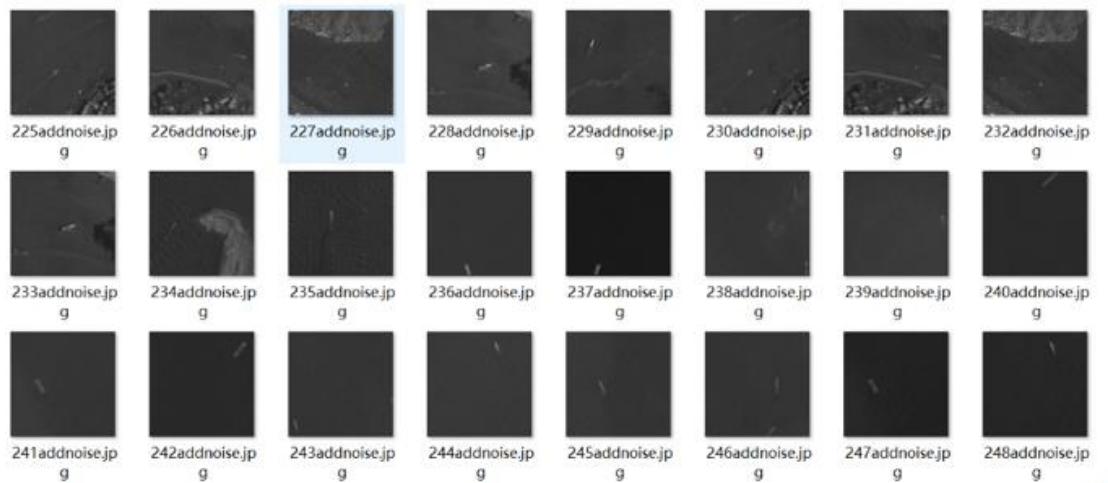


图 4.16 采用的数据集

Fig.4.16 Dataset

本实验采用 VOC 数据集的 XML 格式，所有数据均通过 LabelImg 软件进行标注及标签文件的获取。数据集文件与标签文件如图 4.17 所示：



```

<?xml version="1.0"?>
<annotation>
  <folder>JPEGImages</folder>
  <filename>1addnoise.jpg</filename>
  <path>D:\CX\舰船检测\JPEGImages\1addnoise.jpg</path>
  - <source>
    <database>Unknown</database>
  </source>
  - <size>
    <width>416</width>
    <height>416</height>
    <depth>1</depth>
  </size>
  <segmented>0</segmented>
  - <object>
    <name>ship</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    - <bndbox>
      <xmin>239</xmin>
      <ymin>317</ymin>
      <xmax>294</xmax>
      <ymax>343</ymax>
    </bndbox>
  </object>
  - <object>
    <name>ship</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
  </object>
</annotation>

```

图 4.17 数据集制作

**Fig.4.17 Dataset making**

本文实验将图 4.16 采用的正样本数据集以 8:2 的比例分为训练集和验证集。并从候选区域定位算法的输出中共选取到 372 张含有目标的正样本图像，将这些图像作为测试集，共存在 618 个舰船目标。为了提高模型的收敛速度，采用迁移学习的方法，引入基于相应模型在 ImageNet 的预训练权重，使网络很快收

敛。并在训练开始的数个 Epoch 采用冻结主干参数，采用 Adam 优化器对网络进行迭代更新，衰减权重为 0.94，检测器的置信度阈值为 0.5。拟定共训练 150 个 Epoch，同时，为加快训练速度，根据如果验证集的损失趋于稳定则停止提前停止网络训练。为更好地训练模型，还采用了 Cosine scheduler 余弦退火学习率和 Label smoothing 标签平滑的训练技巧。余弦退火衰减法使学习率采取图 4.18 的变化模式，可以使模型容易跳出局部最小值。Label Smoothing 平滑就是将标签进行平滑处理，将正负样本的置信度为 0, 1 修改为 0.1, 0.9。让模型对样本的分类存在一定的偏差，从而防止模型的过拟合。

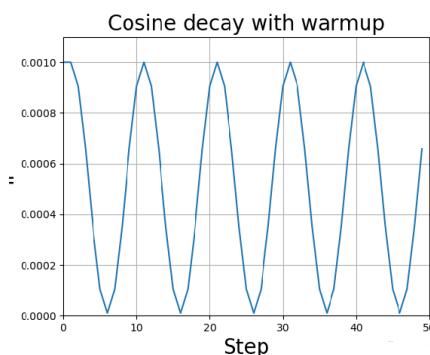


图 4.18 余弦退火学习率变化

Fig.4.18 Cosine annealing learning rate change

训练过程如图 4.19 所示：

```
Setting learning rate to 9.99999834657247e-06.
Epoch 106/200
297/297 [=====] - 26s 87ms/step - loss: 1.8118 - val_loss: 1.99
Setting learning rate to 9.9999971501935e-06.
Epoch 107/200
297/297 [=====] - 26s 87ms/step - loss: 1.7834 - val_loss: 1.96
Setting learning rate to 9.99999747378752e-06.
Epoch 108/200
297/297 [=====] - 26s 88ms/step - loss: 1.7229 - val_loss: 1.97
Setting learning rate to 9.9999971501935e-06.
Epoch 109/200
297/297 [=====] - 26s 87ms/step - loss: 1.8771 - val_loss: 1.95
Setting learning rate to 9.99999834657247e-06.
Epoch 110/200
297/297 [=====] - 26s 88ms/step - loss: 1.7846 - val_loss: 1.98
Setting learning rate to 9.99999912721506e-06.
Epoch 111/200
297/297 [=====] - 26s 87ms/step - loss: 1.7553 - val_loss: 1.94
Setting learning rate to 9.99999975876818e-06.
Epoch 112/200
297/297 [=====] - 26s 87ms/step - loss: 1.7933 - val_loss: 1.96
Setting learning rate to 1e-05.
Epoch 113/200
297/297 [=====] - 26s 87ms/step - loss: 1.7661 - val_loss: 1.96
Setting learning rate to 9.99999975876818e-06.
Epoch 114/200
```

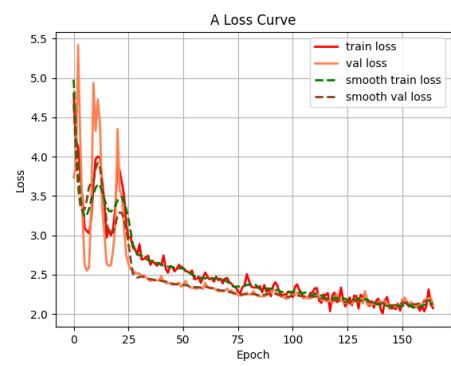


图 4.19 网络训练

Fig.4.19 network training

## 4.5 实验结果与分析

本文对 50 张幅宽为  $2048 \times 8192$  pixels 遥感图像作为输入，共截取  $416 \times 416$  pixels 大小的图像 5000 张，将 5000 张图像作为测试集，对整体算法进行测试。

### 4.5.1 候选区域预筛选阶段的结果与分析

本文的候选区域定位算法对 5000 张图片的处理结果如表 4.3 所示：

表 4.3 快速目标定位算法结果

Table 4.3 Fast target location algorithm results

方法	实际正样本数	算法输出的总数目	未检出实际的正样本数	平均运行时间/s	平均帧率 /fps
本文方法	372	624	12	1.05e-3	850

从实验结果可知，本文提出的基于谱残差模型的候选区域目标定位判别算法，作为一种降低冗余数据的输入方法，滤除 87.5% 的冗余数据。并保留了 97% 的有效数据，同时具有较高的计算速度，体现了传统方法的计算优势并保证较高的筛选精度，最重要的是极大降低了输入后续深度学习目标检测网络的数据量和整体算法的计算时间，同时变相减少了检测算法因引入大量负样本带来的潜在误检。

### 4.5.2 检测网络的结果与分析

实验的测试集来源于在 50 张宽幅全色图像中提取出的 372 张正样本，共 618 个舰船目标。图 4.9 展示了本文提出的目标检测模型在不同场景下的检测效果。分别代表海杂波干扰，云雾干扰和小目标检测。本文算法均可以从中正确检测到目标。

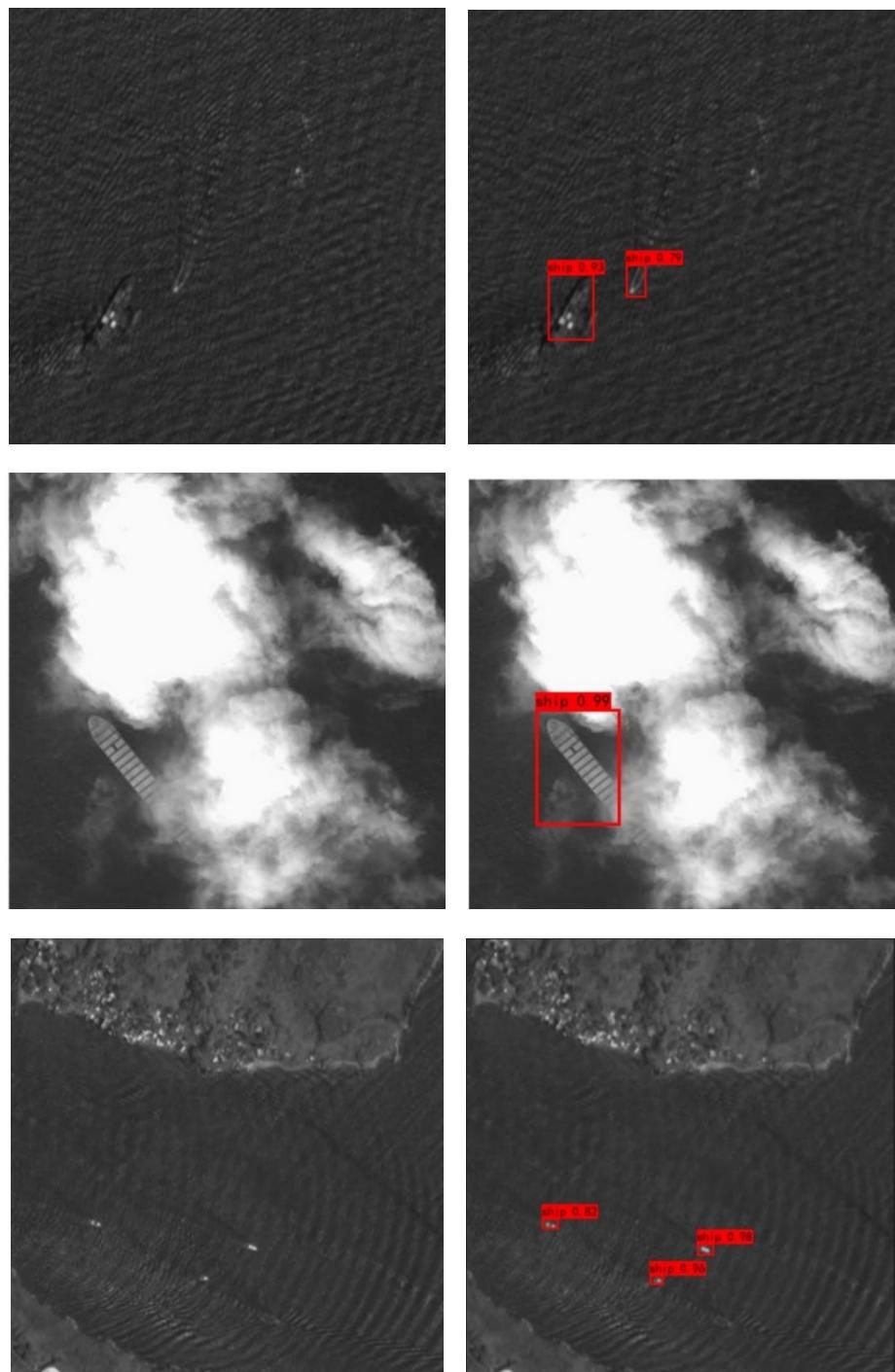


图 4.20 检测结果

Fig.4.20 detection result

在 iou 阈值为 0.5 的条件下，检测出 564 个正确目标，漏检 54 个目标，误检 66 个目标，精度和召回率的计算结果如下：

$$recall = \frac{TP}{TP + FP} = \frac{564}{564 + 54} = 91.26\% \quad \dots (4.5)$$

$$precision = \frac{TP}{TP + FN} = \frac{564}{564 + 66} = 89.52\% \quad \dots (4.6)$$

考虑到网络为轻量化检测网络在精度上略有不足，同时数据集正样本小目标较多，其计算 IoU 偏差更大，将 IoU 阈值放宽为 0.3，此时检测出 581 个舰船目标，漏检 37 个目标，误检 49 个目标。较阈值为 0.5 时漏检，误检明显减少，各指标显著提高。说明大量的正例被检测出来，进一步证明了算法的检测效果。

$$recall = \frac{TP}{TP + FP} = \frac{581}{581 + 37} = 94.01\% \quad \dots (4.7)$$

$$precision = \frac{TP}{TP + FN} = \frac{581}{581 + 49} = 92.22\% \quad \dots (4.8)$$

为进一步分析本文提出优化方法对算法的影响，结合优化方案进行了消融实验。如表 4.4 所示：

**表 4.4 优化方法对算法的提升**

**Table 4.4 Optimization method to improve the algorithm**

方法	CBAM 注意力 机制	Mosaic 图 像增强	自适应先验 框获取	精度	召回率	AP50
Yolov4-tiny				94.25%	84.95%	88.33%
改进 1	√			92.66%	85.76%	89.97%
改进 2	√	√		88.44%	87.86%	90.01%
本文方法	√	√	√	89.52%	91.26%	92.14%

为提升算法的检测速度，本文采取了轻量化的检测模型，较常规检测模型而言，其在精度上必然会存在损失，在实际应用中，对检测网络召回率的要求最高，因为召回率体现了模型对影像数据中舰船的检索能力，召回率越高，说明检测模型对目标的检索与识别能力越强。所以本文所采用的优化方法以提高召回率为首要目标，通过对表 4.4 数据的分析，本文所采用的 CBAM 注意力机制和 Mosaic 图像增强对于目标特征的学习都起到一定的帮助，而对先验框的自

适应获取方法有利于网络的拟合，促进模型的收敛。最终较改进前大大增加了召回率。获得了预期效果。

同时，本文在计算量和检测精度对提出的算法与其他常用检测模型进行对比。在实验数据集及相关训练参数相同的前提下分别对本文方法，yolov3，yolov4，yolov4-tiny 网络进行训练，表 4.5 表示了各模型在目标检测比较重要的参数召回率，来分析各模型的效果：

**表 4.5 各网络性能对比**

**Table 4.5 Comparison of network performance**

Model	recall	AP	FPS
Yolov4	94.34%	97.07%	20
Yolov4-tiny	84.95%	88.33%	125
Yolov3	92.78%	94.89%	28
Ours-net	91.26%	92.14%	105

本文算法为基于 yolov4-tiny 轻量化模型框架的改进算法，从表 4.4 可以看出，在 AP 和召回率等关于精度的指标上，本文所用优化方法在 AP 上，由于网络的深度有限，本文算法在精度上较 yolov3 和 yolov4 网络差距稍大，导致 AP 略有差距。但在召回率方面较原模型效果提升明显，召回率反映网络对目标的检测能力，重要性更高。同时，本文算法的召回率接近 yolov3 的网络水平，较 yolov4 网络差距不大。可以体现说明本文方法对舰船小目标有较好的检测效果。

#### 4.5.3 整体算法评估

本文为对宽幅遥感图像进行的舰船检测，将从算法的精度和计算时间对算法整体进行总结和评价。

在精度方面，检测网络对候选区域提取算法提取出的 624 个目标全部进行检测，代表算法的整体精度。在 IoU 阈值为 0.5 时，检测出 564 个正确目标，漏检 54 个目标，误检 103 个目标，同时，未检出 12 幅正样本图像，其中含有 20 个目标，全部当作漏检，整体算法的召回率的计算结果如下：

$$\text{recall} = \frac{TP}{TP + FP} = \frac{564}{564 + 54 + 20} = 88.40\% \quad \dots (4.9)$$

可以看出，整体算法的召回率有所降低，但仍保持较高水平，可满足实际需求。在时间方面，本文通过基于传统方法的候选区域提取降低了数据冗余，同时采用轻量化的检测网络，与直接采用深度学习网络相比，在计算时间上有所减少。表 4.6 显示了本文方法与 yolov3, yolov4, yolov4-tiny 网络处理整幅宽幅图像的平均计算时间对比，其中 FPS\_W 代表处理整幅宽幅图像的平均计算时间。

**表 4.6 各算法计算时间对比**

**Table 4.6 Comparison of calculation time of each algorithm**

Model	FPS	FPS_W
Yolov4	20	0.2
Yolov4-tiny	125	1.25
Yolov3	28	0.28
Ours-net	105	1.05
整体算法	—	3.71

可以看出，在加入了候选区域预筛选模块后，整体算法在保证精度的同时在检测速度上有了较大提升。

#### 4.6 本章小结

本章对宽幅遥感图像舰船检测方法进行总体论述，包括结合第三章提出的候选目标提取方法进行候选区域预筛选，以及基于 yolov4-tiny 模型的轻量化检测网络两部分。提出的检测网络模型 yolov4-tiny 轻量化模型为主干，以提高检测的召回率为目标，通过加入数据增强，注意力机制，自适应先验框获取等优化方法，构建了高效快速的轻量化检测网络。在以宽幅图像中提取的正样本构建数据集与其他类型网络相对比，在各种场景下均能高效检测目标，同时较原网络大大提高了召回率。同时，由于结合第三章研究的基于谱残差显著性模型的区域提取模块，整体算法在检测时间上获得较大缩减，为在图像中所占比例极小时目标检索困难、计算复杂的难题提供了一种较好的解决方案，实现了对大幅面的遥感数据中的舰船小目标的高效检测。



## 第五章 总结与展望

### 5.1 全文总结

随着空间遥感技术的不断发展和对地监测需求的不断提高，空间成像系统也在不断更新换代，成像分辨率大大提升。目标检测在对图像的解译方面起到至关重要的作用，得到的相关数据对国民经济和军事领域都有重要的应用价值。随着成像技术的提升，图像质量虽然不断提高但也伴随着数据量的增加，对检测算法的发展也带来了新的机遇和挑战。本文提出了一种在宽幅遥感图像中对舰船目标高效快速的目标检测方法，基于深度学习的目标检测模型目前已成为目标检测领域的主流方案，其在检测精度上大大领先传统方法，同时在复杂场景下也能保持较强的鲁棒性。但由于深度学习网络体量较大，且存在移植等问题，目前难以实现星载实时检测，本文的主要工作是基于深度神经网络框架，结合传统的显著性候选区域提取方法，提高整体算法对宽幅图像舰船目标检测的处理速度，为深度学习框架在星载实时检测提供解决方案。

本文主要做了以下两方面工作：

首先，提出一种基于谱残差显著性模型的候选区域提取方法。以谱残差显著性模型为基础，结合算法特点和遥感图像特性对尺度选择和分割阈值的计算方法进行改进，有效克服了云雾和海浪等背景带来的干扰，同时尽可能减少目标遗漏。

其次，提出一种针对宽幅遥感图像的舰船目标检测算法。将宽幅图像分割为多个子图，通过上述提出的显著性算法对子图是否存在疑似目标进行判断，降低无效数据的输入，后续的目标检测网络主要基于 yolov4-tiny 为主干架构对目标进行精细鉴别。并针对数据集中存在的小目标和复杂背景通过加入自适应先验框获取，Mosaic 数据增强，通道注意力机制等方式对网络进行优化。同时在训练中采用标签平滑(label smooth)等训练技巧使模型更好地收敛。

实验结果表明，本文的候选区域提取算法可以有效提出图像的疑似区域，目标检测网络较原网络提高了对舰船小目标的检测能力。整体算法在保证检测

精度的同时计算时间较纯深度学习的方法明显降低。可以对实际工程应用提供参考。

## 5.2 后续工作展望

未来对于遥感图像目标检测的必定是向着更快速更准确的方向发展，现有的研究成果也必须从实际的工程实践出发，向机载，星载的方向靠拢。

目前本领域存在的主要制约因素有以下几方面，首先是实际遥感数据集获取困难，不利于网络的训练，其次是星载，机载的硬件平台要求较高，现有的可进行深度学习算法移植的移动端硬件平台难以满足现有需求。

所以，本课题的后续工作展望也围绕以上所提情况继续研究。首先，继续对算法进行优化，以提高算法的速度和精度。其次，获取更大的数据集，对网络权重进行进一步优化。最后，对程序内部进行进一步优化，提高运行效率，并尝试在硬件端进行移植，测试实际效果。

## 参考文献

- [1] Flanders D . Pereverzoff H B & J . Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction[J]. Canadian Journal of Remote Sensing, 2003.
- [2] GUINDON B. Computer-based aerial image understanding: A review and assessment of its application to planimetric information extraction from very high resolution satellite images[J]. Canadian journal of remote sensing, 1997, 23(1) : 38-47.
- [3] He K, Jian S, Tang X. Single image haze removal using dark channel prior[C]. IEEE Conference on Computer Vision & Pattern Recognition, 2009
- [4] He K, Sun J, Tang X. Guided image filtering[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(6): 1397-1409
- [5] Proia N, Page V. Characterization of a Bayesian ship detection method in optical satellite images[J]. IEEE Geoscience & Remote Sensing Letters, 2010, 7(2): 226-230
- [6] Y. Xia S H W, And L. H. Yue. A novel algorithm for ship detection based on dynamic fusion model of multi-feature and support vector ma-chine[J]. Proc. 6th ICIG, 2011: 521–526
- [7] R Achanta, S Hemami, F Estrada, et al. Frequency-tuned salient region detection[J]. IEEE, 2009.
- [8] Itti L . A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Trans, 1998, 20.
- [9] HOU X , ZHANG L . Saliency Detection: A Spectral Residual Approach[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2007.
- [10] Achanta R , Estrada F , Wils P , et al. Salient Region Detection and Segmentation[C]// computer vision systems;ICVS 2008. 2008.
- [11] ERDEM E, ERDEM A. Visual saliency estimation by nonlinearly integrating features using region covariances[J]. Journal of Vision, 2013, 13(4):11.

- [12]B Schölkopf, Platt J , Hofmann T . Graph-Based Visual Saliency. MIT Press, 2007.
- [13] 李庆峰,何斌,王文胜,等.基于谱残差和梯度纹理融合特征的舰船检测[J].液晶与显示,2019,34(08):803-815.
- [14] Dalal N , Triggs B . Histograms of Oriented Gradients for Human Detection[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE, 2005.
- [15] Qiuguo YE , Zha X , Haibin L I . High Resolution Remote Sensing Image Ship Detection Based on Visual Significance[J]. Hydrographic Surveying and Charting, 2018.
- [16] 赵浩光, 王平, 董超,等. 结合多尺度视觉显著性的舰船目标检测[J]. 光学精密工程, 2020, 028(006):1395-1403.
- [17] Jubelin G , Khencha F A . Multiscale algorithm for ship detection in mid, high and very high resolution optical imagery[C]// Geoscience & Remote Sensing Symposium. IEEE, 2014.
- [18] Li Z , You Y , Liu F . Multi-Scale Ships Detection in High-Resolution Remote Sensing Image Via Saliency-Based Region Convolutional Neural Network[C]// IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019.
- [19] Yu Y, Ai H, He X J, et al. Ship Detection in Optical Satellite Images Using Haar-Like Features and Periphery-Cropped Neural Networks[J]. IEEE Access. 2018.6: 71 122-71 131
- [20] Wang Y L, DongZ P, Zhu Y. Multiscale Block Fusion Object Detection Method for Large - Scale High-Resolution Remote Sensing Imagery[J]. IEEE Access, 2019, 7: 99 530-99 539
- [21] He T, Zhang Z, Zhang H, et al. Bag of Tricks for Image Classification with Convolutional Neural Networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019
- [22] LECLERC M, THARMARASA R, FLORE A M C, et al. Ship Classification Using Deep Learning Techniques for Maritime Target Tracking[C] // 2018 21st International Conference on Information Fusion (FUSION). IEEE, 2018 : 737-744.
- [23] HUANG J, JIANG Z G, ZHANG H P, et al. Remote sensing image ship target detection based on convolutional neural network[J]. Journal of Beijing University of Aeronautics and

- Astronautics, 2017 ,43(9) : 1841-1848.
- [24] WANG J C,TAN X C,WANG Z H,et al. Research on remote sensing image target recognition method based on Faster R-CNN[J]. Geo-Information Science,2018,20(10) : 1500-1508.
- [25] HU G, YANG Z, HAN J,et al. Aircraft detection in remote sensing images based on saliency and convolution neuralnetwork [J]. EURASIP Journal on Wireless Communications and Networking,2018,2018(1) : 26.
- [26] ZHAO C H,ZHOU Y. Ship target detection and recognition based on the improved Faster R-CNN[J]. Journal of Shenyang University,2018 (5) : 366-371,380.
- [27]吴飞, 汪渤, 周志强, 等.基于船头特征提取和轮廓定位的港口舰船检测[J].北京理工大学学报,2018,38(04):387-392.
- [28]胡俊华, 徐守时, 陈海林, 等.基于局部自相似性的遥感图像港口舰船检测[J].中国图象图形学报, 2009, 14(4):591-597.
- [29]聂婷. 大幅宽光学遥感图像目标检测技术研究[D].中国科学院大学(中国科学院长春光学精密机械与物理研究所),2019.
- [30]徐芳,刘晶红,孙辉,王腾龙,王宣.光学遥感图像海面船舶目标检测技术进展[J].光学精密工程,2021,29(04):916-931.
- [31] 王彦情, 马雷, 田原. 光学遥感图像舰船目标检测与识别综述[J]. 自动化学报, 2011, 37(9).1029-1039.
- [32]冈萨雷斯主编.数字图像处理[M].第 2 版.阮秋琦等译.北京:电子工业出版社, 2003.3.176-220.
- [33]尼克松主编.特征提取与图像处理[M].第 2 版.李实英等译.北京:电子工业出版社, 2010. 92-95.
- [34]朱丹,王斌,张立明.基于直线邻近平行性和 GBVS 显著性的遥感图像机场目标检测[J].红外与毫米波学报,2015,34(03):375-384.
- [35] A. Krizhevsky, I. Sutskever, G. E. Hinton. Imagenet classification with deep convolutional neural networks[C]. Advances in neural information processing systems, Cambridge, MA, USA,2012, 1097-1105.

- [36] K. He, X. Zhang, S. Ren, et al. Identity mappings in deep residual networks[C]. European conference on computer vision, Amsterdam, The Netherlands, 2016, 630-645
- [37] K. He, X. Zhang, S. Ren, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 2016, 770-778
- [38] Otsu N . A Threshold Selection Method from Gray-Level Histograms[J]. IEEE Transactions on Systems Man & Cybernetics, 2007, 9(1):62-66.
- [39] Woo S , Park J , Lee J Y , et al. CBAM: Convolutional Block Attention Module[J]. Springer, Cham, 2018.
- [40] Jie H , Li S , Gang S , et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, PP(99).
- [41] Girshick R . Fast R-CNN[J]. Computer Science, 2015.
- [42] Chollet F . Xception: Deep Learning with Depthwise Separable Convolutions[J]. IEEE, 2017.
- [43] Redmon J , Farhadi A . YOLO9000: Better, Faster, Stronger[J]. IEEE, 2017:6517-6525.
- [44] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.
- [45] 刘娟妮, 彭进业, 李大湘, 王平. 基于谱残差和多分辨率分析的显著目标检测 [J]. 中国图象图形学报, 2011, 16 (02) :244-249.
- [46] 徐芳. 可见光遥感图像海面目标自动检测关键技术研究[D]. 中国科学院大学(中国科学院长春光学精密机械与物理研究所), 2018.
- [47] 徐芳, 刘晶红, 曾冬冬, 王宣. 基于视觉显著性的无监督海面舰船检测与识别[J]. 光学精密工程, 2017, 25(05):1300-1311.
- [48] 王慧利. 高分辨率光学遥感图像中典型人造目标检测技术研究[D]. 中国科学院长春光学精密机械与物理研究所, 2017.
- [49] 董超. 可见光遥感图像海面舰船目标检测技术研究[D]. 中国科学院大学(中国科学院长春光学精密机械与物理研究所), 2020. DOI:10.27522/d.cnki.gkcg.2020.000037.
- [50] 王文胜. 宽幅光学遥感图像舰船飞机目标检测识别技术研究[D]. 中国科学院大学(中国科学院长春光学精密机械与物理研究所), 2018.

- [51] 邱世赟. 基于深度学习的高分辨率光学遥感图像舰船检测[D]. 西安电子科技大学, 2018.
- [52] 仉长涛. 基于卷积神经网络的遥感图像飞机目标检测研究[D]. 河北工业大学, 2020. DOI: 10.27105/d.cnki.ghbgu.2020.000886.
- [53] 晁安娜, 刘坤. 基于卷积神经网络的遥感图像飞机目标识别[J]. 微型机与应用, 2017, 36(22): 66-69+73. DOI: 10.19358/j.issn.1674-7720.2017.22.018.
- [54] 张作省. 大面幅光学遥感影像飞行器目标快速检测[D]. 中国科学院大学(中国科学院长春光学精密机械与物理研究所), 2018.



## 致谢

时光荏苒，研究生阶段的三年即将结束。回想起来，从选题到论文完成，一直离不开引导我、激励我、帮助我的各位老师、朋友、亲人，在此请接收我最诚挚的谢意！

感谢我的导师何斌教授，您务实的学术精神、忘我的工作态度、严谨的治学作风深深地影响和教育了我。同时，也要感谢实验室中陈长征老师，张星祥老师和毕国玲老师对我的帮助和指导。在工作之余，老师们也教给我许多为人处世的道理，以开阔的视角解读我在工作上遇到的困难。

感谢宋向宇、龙勇机和张欣三位师兄和聂婷师姐在我硕士期间从选题到论文答辩期间对我的无私支持与帮助，以及 353 实验室的贺玉坤、沈凯、王上、王朋朋、单博闻、曹明辉、赵洪强、王宇哲、张成成、张墨、王子荣、曹业豪、戴得恩等师兄师弟们，感谢这几年来对我的鼓励与照顾。谢谢你们三年来的陪伴，留下了许多美好的回忆。需要感谢的人在此无法一一罗列，愿我们努力的人都有一个美好的明天！

最后，我要感谢我的父母，纵使三年来见面的机会越来越少，但我内心深深明白，你们才是我心中最温暖的港湾和最坚实的依靠。

三言两语，无法言明我内心的感激之情。我一直庆幸自己在人生的各个阶段的经历中，虽没有轰轰烈烈的壮举，但也一直遇到志趣相投的伙伴和正直务实的导师，才让我一直顺顺利利走完自己的求学之路。最后感谢一路上陪伴我的人。谢谢你们让我的生活充满快乐与幸福。我也即将踏出校门，进入社会开始全新的生活，也希望自己在以后的工作和生活中继续努力，砥砺前行！

陈玺

2022-3-20



