



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

博士学位论文

面向复杂背景下小目标检测的深度神经网络研究

作者姓名: 鞠默然

指导教师: 罗海波 研究员 中国科学院沈阳自动化研究所

学位类别: 工学博士

学科专业: 模式识别与智能系统

培养单位: 中国科学院沈阳自动化研究所

2021 年 6 月

Research on Deep Neural Network for Small Target Detection in
Complex Background

A dissertation submitted to
University of Chinese Academy of Sciences
in partial fulfillment of the requirement
for the degree of
Doctor of Philosophy
in Pattern Recognition and Intelligent System

By
Ju Moran
Supervisor : Professor Luo Haibo

Shenyang Institute of Automation Chinese Academy of Sciences

June 2021

中国科学院大学

研究生学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

鞠默然

日

期：

2021年5月25日

中国科学院大学

学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：

鞠默然

日

期：

2021年5月25日

导师签名：

罗海峰

日

期：

20210525

摘 要

作为计算机视觉领域中的一个研究热点, 目标检测已广泛应用于智能交通、智能监控、行为识别和医学图像辅助诊断等多个领域。其中, 对地小目标检测和对空红外弱小目标检测可分别在对地作战和对空作战中用于侦察目标, 有助于实现“先敌发现, 先敌锁定和先敌命中”, 对国防事业具有重要的应用价值。然而, 对地小目标和对空红外弱小目标在图像中占有的像素非常少, 特征不明显。因此, 对地小目标检测和红外弱小目标检测更具有挑战性。随着计算机技术的高速发展, 基于深度学习的目标检测算法成为目标检测领域中的主流方法。基于深度学习的目标检测算法可以将特征提取、特征融合、目标定位和分类等多个任务融合到一个网络中, 实现对目标端到端的检测。

综上所述, 基于深度学习的对地小目标检测和对空红外弱小目标检测是一个具有显著应用价值且具有挑战性的研究课题。因此, 本文分别对基于深度学习的对地小目标和对空红外弱小目标检测算法展开研究, 本文的主要贡献如下:

1. 分析并总结了影响小目标检测的关键因素。由于对地小目标和对空红外弱小目标包含像素少、特征不明显, 易受周围环境的影响。因此, 与常规目标相比, 对地小目标和对空红外弱小目标检测更具有挑战性。为了设计适合于对地小目标和对空红外弱小目标检测的算法, 本文首先对影响小目标检测的关键因素进行了分析和总结。并针对各个因素, 提出相应的解决方法。

2. 提出了面向小目标检测的图像预处理方法。在对地小目标检测任务中, 输入图像的大部分区域都是背景, 只有少部分区域是目标。根据此特点, 提出了一种基于图像裁剪的对地小目标预处理方法。利用该方法, 可以在不增加显存的条件下, 实现网络对目标多尺度的训练, 使网络变得更加鲁棒。另外, 针对红外弱小目标信号“弱”、“小”且常常伴随着噪声干扰的特点, 提出了基于局部对比度的红外弱小目标图像预处理方法。该方法可以利用简单的卷积操作和最大池化操作来实现, 可以快速、有效地提高目标与背景之间的对比度。

3. 提出了一种基于超分辨率重建的对地小目标检测网络 SuperDet。作为影

响小目标检测的一个关键因素,提高小目标的分辨率有助于提高检测网络对小目标的检测性能。**SuperDet** 通过结合超分辨率重建技术和目标检测技术来提高对对地小目标的检测性能。**SuperDet** 主要由两部分组成,分别是超分辨率模块和对地小目标检测模块。超分辨率模块实现了低分辨率图像到高分辨率图像的转换。将重建后的高分辨率图像作为输入送到对地小目标检测模块完成对对地小目标位置和类别的预测。为了选择合适的尺度来检测小目标,利用尺度匹配策略为对地小目标设计主干网络。随后提出了加权感受野融合模块,通过自适应融合具有不同感受野的特征来增加对地小目标周围的上下文信息。另外,针对对地小目标检测中正负样本数量不均衡的问题,提出了正样本选择策略。最后,分别采用 GIOUS loss、Focal loss 和 Binary Cross Entropy loss 来对小目标的位置、置信度和类别进行回归。为了实现对 **SuperDet** 端到端的训练,采用多任务损失函数对超分辨率重建任务的损失和目标检测任务的损失进行回归。实验结果表明, **SuperDet** 具有较为出色的对地小目标检测能力。

4. 提出了一种基于图像过滤的对空红外弱小目标检测网络 **FilterDet**。**FilterDet** 主要由两部分组成,分别是过滤模块和红外弱小目标检测模块。其中,过滤模块实现了对背景的抑制和对红外弱小目标的增强,可以过滤掉图像中的噪声和干扰。红外弱小目标检测模块将过滤后的图像作为输入,完成对红外弱小目标位置的预测。根据尺度匹配策略,采用沙漏结构来设计红外弱小目标检测模块的主干网络。另外,采用加权感受野融合模块来增加小目标周围的上下文信息。利用空间注意力机制来增强网络的表达能力,使红外弱小目标的特征响应得到加强。**FilterDet** 将图像过滤任务和红外弱小目标检测任务融合到了一个网络中,实现了对红外弱小目标端到端的检测。实验结果表明, **FilterDet** 具有较为出色的对空红外弱小目标检测性能。

关键词: 卷积神经网络, 对地小目标检测, 对空红外弱小目标检测, 超分辨率重建, 图像过滤

Abstract

As a research hotspot in the field of computer vision, target detection has been widely used in intelligent transportation, intelligent monitoring, action recognition and medical image aided diagnosis. Ground small target detection and aerial infrared small target detection can be used to perform target reconnaissance in to-air operations and to-ground operations respectively, which is helpful to realize “first enemy detection, first enemy lock and first enemy hit”, and has important application value to national defense. Compared with generic targets, ground small targets and aerial infrared dim and small targets cover only a small part of an image and the features of these targets are inconspicuous. Therefore, ground small target detection and aerial infrared dim and small target detection are more challenging. With the rapid development of computer technology, target detection algorithm based on deep learning has become the mainstream method in the field of target detection. That is because the target detection algorithms based on deep learning fuse multiple tasks such as feature extraction, feature fusion, target location and classification into a network, which can perform target detection in an end-to-end way.

To sum up, ground small target detection and aerial infrared dim and small target detection based on deep learning is a research subject with significant application value and challenges. Therefore, this dissertation researches on the ground small target detection algorithm based on deep learning and aerial infrared dim and small target detection algorithm based on deep learning respectively. The main contributions of this dissertation are as follows:

1. This dissertation analyzes and summarizes the factors affecting the detection performance of the small targets. Ground small targets and aerial infrared dim and small targets cover few pixels in an image and their features are not obvious. In addition, they are easily affected by the surrounding environment. Therefore, compared with the general target detection, ground small target detection and aerial

infrared dim and small target detection are more challenging. In order to design algorithms suitable for ground small target detection and aerial infrared dim and small target detection, this dissertation firstly analyzes and summarizes the key factors affecting small target detection and then puts forward solutions in view of each factor.

2. Image preprocessing methods for small target detection are proposed. In ground small target detection task, most regions of the input image are background and the targets only cover a few regions. According to this characteristic, an image cropping method for ground small target image preprocessing is introduced. By this method, the network can achieve multi-scale training on the same target without the increasement of memory, which is helpful to make the network more robust. In addition, aiming at the characteristics that the signal of infrared dim and small targets is dim and the targets are often accompanied by noise and interference, a local contrast method for infrared dim and small target image preprocessing is introduced. This image preprocessing method can be achieved by using simple convolution and max-pooling operation, which can enhance the contrast between the targets and background efficiently.

3. Aiming at ground small target detection, a ground small target detection network based on super-resolution reconstruction, called SuperDet, is proposed. As a key factor affecting the detection performance of small targets, the resolution of target plays an important role. Improving the resolution of small targets is conducive to improving the detection performance of the network. Therefore, a ground small target detection network SuperDet is proposed by combining the super-resolution reconstruction technology with the target detection algorithm. SuperDet consists of two parts, namely, super-resolution reconstruction module and target detection module. The super-resolution reconstruction module aims to recover a high resolution image from its low resolution counter part. The target detection module takes the reconstructed high resolution image as input and predict the location and the category of the ground small targets. In order to select appropriate scales for ground small

target detection, the scale matching strategy is proposed. According to the scale matching strategy, we design the backbone for the ground small targets. In addition, a weighted receptive field fusion module is introduced to increase the context information around the targets by merging the features with different receptive field. What is more, a positive sample selection strategy is proposed to alleviate the imbalance between the positive samples and negative ones in ground small target detection. Finally, GIOU Loss, Focal Loss and Binary Cross Entropy Loss are used to regress the position, confidence and category of the targets. To perform the end-to-end training of SuperDet, multi-task loss function is used to regress the loss of super-resolution reconstruction task and ground small target detection. The experimental results show that SuperDet has excellent detection ability for ground small targets.

4. Aiming at aerial infrared dim and small target detection, an infrared dim and small target detection network based on image filtering, called FilterDet, is proposed. FilterDet consists of two parts, namely image filtering module and infrared dim and small target detection module. The image filtering module can filter out the noise and enhance the targets. The infrared dim and small target detection module takes the filtered image as input and predicts the location of the infrared dim and small targets. According to the scale matching strategy, the hourglass structure is used to design the backbone of infrared dim and small target detection module. In addition, the weighted receptive field fusion module is used to increase the context information around the small targets. Spatial attention mechanism is applied to improve the expressive ability of the network, which enhances the response of infrared dim and small targets. FilterDet integrates the image filtering task and the infrared dim and small target detection task into one network, which realizes the end-to-end detection of the infrared dim and small targets. The experimental results show that FilterDet has excellent detection performance for aerial infrared dim and small targets.

Key Words: Convolutional Neural Network, Ground Small Target Detection, Aerial Infrared Dim and Small Target Detection, Super-Resolution Reconstruction, Image Filtering

目 录

第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	4
1.2.1 目标检测算法研究现状.....	4
1.2.2 小目标检测算法研究现状.....	10
1.2.3 对地小目标检测算法研究现状.....	12
1.2.4 红外弱小目标检测算法研究现状.....	12
1.2.5 基于深度学习的超分辨率重建算法研究现状.....	15
1.3 研究中的难点和挑战.....	19
1.4 论文研究内容与章节安排.....	20
第 2 章 影响小目标检测的因素	25
2.1 引言.....	25
2.2 影响小目标检测的关键因素.....	25
2.3 本章小结.....	31
第 3 章 面向小目标检测的图像预处理方法	33
3.1 引言.....	33
3.2 基于图像裁剪的对地小目标图像预处理方法.....	33
3.2.1 多尺度训练.....	33
3.2.2 图像裁剪的原理和流程.....	34
3.2.3 实验与分析.....	39
3.3 基于局部对比度的红外弱小目标图像预处理方法.....	42
3.3.1 局部对比度测量方法.....	43
3.3.2 局部对比度测量方法的改进与快速实现.....	44
3.3.3 实验与分析.....	48
3.4 本章小结.....	50
第 4 章 基于超分辨率重建的对地小目标检测算法	51

4.1 引言.....	51
4.2 超分辨率重建网络设计.....	52
4.2.1 一种 ESPCN 超分辨率重建网络的改进方法	52
4.2.2 实验与分析.....	54
4.3 对地小目标检测网络设计.....	56
4.3.1 尺度匹配策略.....	56
4.3.2 加权感受野融合模块.....	63
4.3.3 特征融合机制.....	66
4.3.4 对地小目标检测网络的结构.....	68
4.3.5 正样本选择策略.....	68
4.4 基于超分辨率重建的对地小目标检测网络结构.....	70
4.5 多任务损失函数.....	72
4.6 实验与分析.....	74
4.6.1 实验条件与评价方法.....	74
4.6.2 消融实验.....	76
4.6.3 DOTA 数据集上的验证.....	78
4.6.4 VEDAI 可见光数据集上的验证	79
4.6.5 VEDAI 红外数据集上的验证	80
4.7 本章小结.....	82
第 5 章 基于图像过滤的对空红外弱小目标检测算法	85
5.1 引言.....	85
5.2 图像过滤模块的设计.....	86
5.3 对空红外弱小目标检测网络设计.....	89
5.3.1 主干网络设计.....	89
5.3.2 特征融合机制.....	91
5.3.3 空间注意力机制.....	92
5.3.4 对空红外弱小目标检测网络的结构.....	94
5.4 基于图像过滤的对空红外弱小目标检测网络结构.....	95
5.5 多任务损失函数.....	97
5.6 实验与分析.....	97

5.6.1 实验条件与评价方法.....	97
5.6.2 消融实验.....	100
5.6.3 合成的单帧红外弱小目标数据集上的验证.....	102
5.6.4 真实红外弱小目标序列上的验证.....	104
5.7 本章小结.....	109
第 6 章 总结与展望	111
6.1 本文工作总结.....	111
6.2 未来展望.....	112
参考文献.....	115
致 谢.....	125
作者简历及攻读学位期间发表的学术论文与研究成果	129

图目录

图 1.1	目标检测的任务	1
图 1.2	对地目标检测和对空目标检测	2
图 1.3	对空红外弱小目标图像	3
图 1.4	传统目标检测算法的流程	5
图 1.5	RCNN 目标检测算法的流程	7
图 1.6	Faster RCNN 目标检测算法的工作流程	8
图 1.7	YOLO V3 目标检测算法	9
图 1.8	改进 RCNN 的小目标检测算法	10
图 1.9	小人脸检测算法	11
图 1.10	检测网络结构	15
图 1.11	SRCNN 结构图	16
图 1.12	SRCNN 与 FSRCNN 的对比	17
图 1.13	ESPCN 的结构	17
图 1.14	VDSR 的结构	18
图 1.15	遥感图像车辆目标与自然场景中的车辆对比	19
图 1.16	对地小目标和红外弱小目标对比	20
图 1.17	研究内容的框架图	21
图 2.1	相同场景下不同分辨率图像的对比	25
图 2.2	人脸分辨率对检测结果的影响	26
图 2.3	不同大小目标在相同特征图上对比	27
图 2.4	未增加上下文信息和增加上下文信息船图像的对比	27
图 2.5	增大感受野对小人脸检测的影响	28
图 2.6	目标在不同特征图上的位置回归	29
图 2.7	VGG-19 不同层特征图的可视化	30
图 2.8	红外弱小目标	30
图 2.9	全文设计的框架图	31
图 3.1	多尺度训练	33

图 3.2	图像裁剪过程中的 4 种情况	35
图 3.3	最小区域的延长方式	35
图 3.4	多尺度训练与基于图像裁剪的对地小目标图像预处理方法的对比	39
图 3.5	YOLO V3 和 YOLO V3*在 VEDAI 数据集上可视化结果	41
图 3.6	GSTDet 和 GSTDet*在 VEDAI 数据集可视化结果	42
图 3.7	对比度机制	43
图 3.8	局部对比度测量中的图像块	43
图 3.9	局部对比度测量方法与改进的局部对比度测量方法的结果对比	46
图 3.10	红外弱小目标周围背景的局部对比度的计算	46
图 3.11	改进的局部对比度测量方法的实现	47
图 3.12	PR 曲线对比	49
图 4.1	超分辨率重建与目标检测网络的结合	51
图 4.2	亚像素卷积层	52
图 4.3	原 ESPCN 与改进的 ESPCN 的对比	53
图 4.4	不同算法下图像重建效果比较	55
图 4.5	YOLO V3 结构图	57
图 4.6	等比例间隔原则(density=4)	58
图 4.7	每个尺度的锚框	59
图 4.8	Darknet53 结构单元	61
图 4.9	对地小目标检测网络 GSTDet 的主干	63
图 4.10	膨胀卷积。(a) 膨胀系数=1 (b) 膨胀系数=2 (c) 膨胀系数=4	63
图 4.11	尺度自适应三叉戟网络的结构	65
图 4.12	加权感受野融合模块的结构图	65
图 4.13	多尺度预测与 FPN 的对比	67
图 4.14	对地小目标检测网络 GSTDet 中的特征融合	67
图 4.15	对地小目标检测网络 GSTDet 的结构图	68
图 4.16	正样本选择策略	69
图 4.17	SuperDet 的结构	70

图 4.18	检测框的回归	72
图 4.19	SuperDet 在 DOTA 数据集上的可视化结果	79
图 4.20	SuperDet 在 VEDAI 数据集上的可视化结果	80
图 4.21	GSTDet 与 SuperDet 在红外图像上的可视化检测结果	82
图 5.1	红外弱小目标图像	85
图 5.2	对空红外弱小目标检测的框架	86
图 5.3	图像过滤模块输出的置信度图	89
图 5.4	对空红外弱小目标检测网络 ISTDet 的主干	91
图 5.5	对空红外弱小目标检测网络 ISTDet 中的特征融合	91
图 5.6	空间注意力机制	93
图 5.7	空间注意力机制的作用	94
图 5.8	对空红外弱小目标检测网络 ISTDet 的结构图	95
图 5.9	FilterDet 的结构	96
图 5.10	真实的红外弱小目标序列。(a) 序列 1 (b) 序列 2 (c) 序列 3	98
图 5.11	单帧红外弱小目标图像	100
图 5.12	单帧红外弱小目标数据集上 PR 曲线对比	103
图 5.13	FilterDet 在单帧红外图像上的检测结果	104
图 5.14	在 3 个红外弱小目标序列上的 PR 曲线对比	106
图 5.15	FilterDet 在序列 1 上的可视化检测结果	107
图 5.16	FilterDet 在序列 2 上的可视化检测结果	108
图 5.17	FilterDet 在序列 3 上的可视化检测结果	109

表目录

表 3.1	算法的伪代码	38
表 3.2	YOLO V3 在 VEDAI 数据集上的对比结果: AP (%)	40
表 3.3	GSTDet 在 VEDAI 数据集上的对比结果: AP (%)	40
表 3.4	改进的局部对比度测量方法的伪代码	48
表 3.5	不同输入的实验结果对比	49
表 4.1	不同算法 PSNR 值在 VEDAI 数据集上的对比结果	54
表 4.2	不同密度下的锚框尺寸范围	59
表 4.3	每一个尺度上锚框的尺寸	60
表 4.4	不同密度范围下的实验结果	62
表 4.5	普通卷积和膨胀卷积感受野的对比	64
表 4.6	VEDAI 数据集中各类目标的数据统计	76
表 4.7	在 VEDAI 数据集上的消融实验	77
表 4.8	SuperDet 在 VEDAI 数据集上的消融实验	78
表 4.9	不同检测算法在 DOTA 数据集上的检测结果	78
表 4.10	不同检测算法在 VEDAI 数据集上的检测结果	80
表 4.11	可见光图像与红外图像检测结果的对比	81
表 5.1	图像过滤模块的配置	87
表 5.2	置信度 ground truth 的生成伪代码	88
表 5.3	对空红外弱小目标检测网络 ISTDet 的主干配置	90
表 5.4	特征融合后 ISTDet 主干网络的配置	92
表 5.5	3 个红外弱小目标序列的详细信息	99
表 5.6	在 3 个序列上的消融实验	101
表 5.7	单帧红外弱小目标数据集上实验结果	102
表 5.8	序列 1 实验结果	105
表 5.9	序列 2 实验结果	105
表 5.10	序列 3 实验结果	105

第1章 绪论

1.1 研究背景及意义

作为人类获取外界信息的主要途径,图像可以给人类带来最直观的视觉感受。随着计算机技术的不断发展,出现了用计算机和相机来代替人眼对目标进行识别、检测和跟踪的计算机视觉系统。计算机视觉系统提高了工作效率,解决了人工成本高的问题,并保证了获取信息的准确性。计算机视觉的三大主流方向包括目标检测^[1]、图像分割^[2]和目标跟踪^[3-4]。其中,目标检测的任务是在图像中找到目标,对每一个目标给出一个精确的目标边框,并确定目标的类别,如图 1.1 所示。目前,目标检测已经广泛地应用到了电力设备自动监测^[5-6]、自动驾驶^[7-10]、智能交通^[11-15]、人脸识别^[16-17]、计算机辅助诊断^[18-19]等领域中。

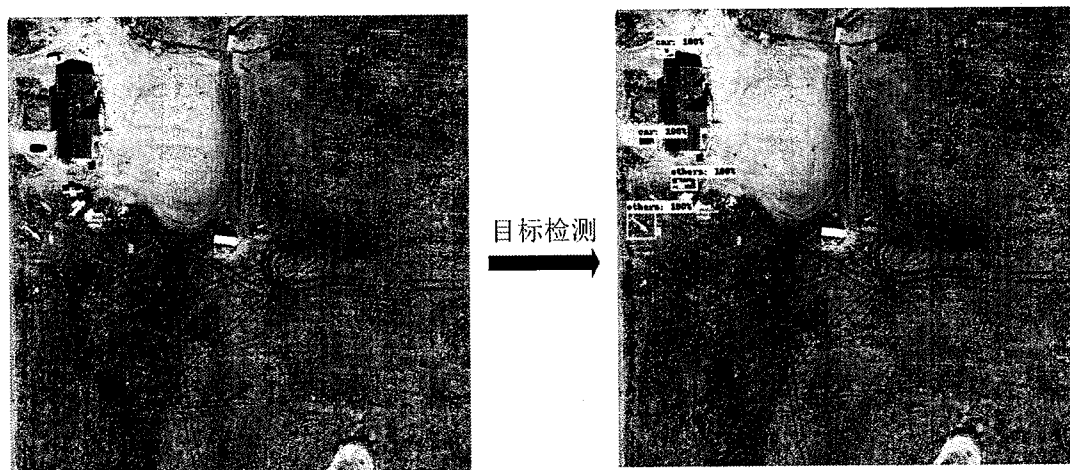


图 1.1 目标检测的任务

Figure 1.1 · The task of target detection

在现代战争中,军事系统的智能化体现了一个国家的军事实力,军事装备的智能化和信息化是各国军队发展的重要方向。其中,对空目标和对地目标自动检测技术是实现这一目标的重要环节。如图 1.2 所示,图 1.2(a)为对地目标检测,图 1.2(b)为对空目标检测。

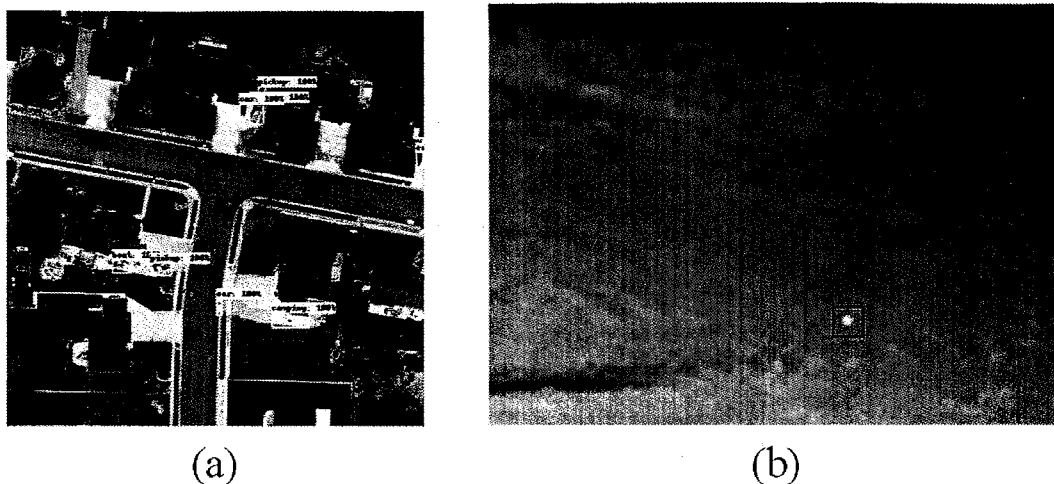


图 1.2 对地目标检测和对空目标检测

Figure 1.2 Ground target detection and aerial target detection

在对地作战中，快速的获取战场态势和敌方目标信息，是克敌制胜的关键。随着科学技术的高速发展，高分辨率卫星、预警机、航空侦察机和无人机在地面目标侦察中发挥了重要作用。为了能尽早的识别并锁定目标，帮助己方在战场上建立优势，需要在尽量远的距离探测、识别和跟踪地面目标，这导致目标在视场中尺寸较小，缺少边缘信息和纹理信息，给检测任务带来了挑战。因此，如何实现对地小目标快速、准确的检测成为了研究的热点^[20]。

在对空作战中，红外成像是空中目标探测的常用手段。作为一种精确制导方式，红外成像制导利用背景与目标之间的热辐射差异进行成像，并实施目标检测、跟踪与制导^[21]。在现代战争中，“先敌发现，先敌锁定和先敌命中”是取得战争胜利的重要保障。因此，实现对远距离空中目标的精准检测是克敌制胜的必要条件。远距离的空中目标在红外成像探测系统中的成像往往只有几个像素，而且强度弱、信噪比低，通常被定义为红外弱小目标。图 1.3 为几种典型场景下的红外弱小目标成像结果，目标与背景之间的对比度低，常伴有较多噪声，目标通常以“亮点”的形式呈现，不具备细致的纹理结构和形状特征。

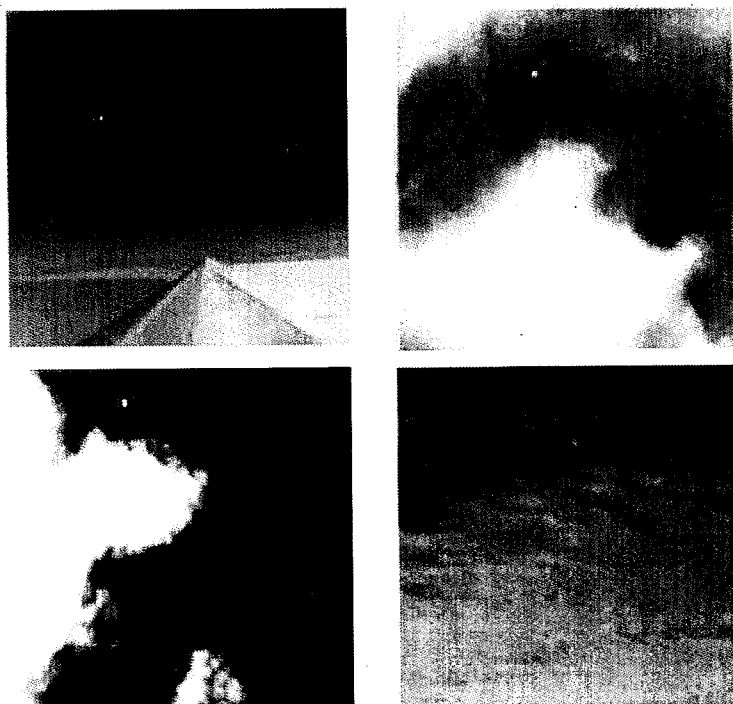


图 1.3 对空红外弱小目标图像

Figure 1.3 Image of aerial infrared dim and small target

综上所述,对地小目标检测和对空红外弱小目标检测可分别在对地作战和对空作战中用于侦察目标,对国防安全等领域具有重要的应用价值。但是,对地小目标和对空红外弱小目标在图像中占有像素少、特征不明显,且目标所处环境复杂多变,这都给检测任务带来了挑战。

传统的目标检测算法首先生成候选区域,然后对候选区域进行特征提取,最后使用分类器对目标进行分类。虽然,传统的目标检测算法取得了一定的成绩,但由于这些算法是基于手工设计的特征,不能表征图像的高维特征信息。因此,手工提取的特征直接限制了算法的性能。

2006 年, Hinton 等人^[22]利用无监督预训练方法来初始化网络参数,有效的解决了梯度消失的问题,为以后深度学习的发展奠定了基础。2012 年, Hinton 团队^[23]设计了卷积神经网络 AlexNet, 以巨大的优势战胜了基于传统手工特征的目标识别算法, 获得了 ImageNet 图像识别比赛的冠军, 使更多的学者加入到深度学习的研究队伍中。目前, 已经有许多基于深度学习的网络模型, 例如: AlexNet^[23]、VGG^[24]、GoogleNet^[25]、ResNet^[26]和 DenseNet^[27]等。随着计算机硬

件技术的不断发展,特别是图形处理器(Graphic Processing Unit, GPU)的进步,计算机的计算能力得到了迅速的提升,促使人工神经网络的深度变得越来越深,可以提取到更深层次的目标特征,为深度学习的进一步发展奠定了基础。由于深度学习具有强大的特征表征能力,可以提取目标中深层次的语义信息,使得基于深度学习的目标检测算法成为了计算机视觉领域中的一个研究热点^[28]。基于深度学习的目标检测算法可以将特征提取、特征融合、目标定位和目标分类等多个任务融合在一个网络中,实现对目标端到端的检测。目前,深度学习已在语音识别、计算机视觉和自然语言处理等多个领域得到广泛应用。基于深度学习的目标检测算法已经广泛应用到日常生活中的各个领域^[29],例如,行为检测^[30-31]、人脸识别^[32-33]、智能交通^[34-36]、智能监控^[37-39]和自动驾驶^[40-42]等。

本文分别针对基于深度学习的对地小目标检测算法和基于深度学习的对空红外弱小目标检测算法进行研究。对于对地小目标检测算法,本文研究的对地小目标的像素数量占整幅图像的 0.1%到 1.5%之间,以 512×512 的图像为例,目标大小介于 16×16 到 64×64 之间。红外弱小目标检测是利用红外成像系统对远距离的目标进行探测,由于远距离成像,红外弱小目标与背景之间的对比度低,信号强度弱。本文研究的对空红外弱小目标的像素数量不超过整幅图像的 0.1%^[43],即对于一张尺寸为 512×512 的图像,红外弱小目标的尺寸通常不超过 16×16 。由于目标在图像中占有的像素少,红外弱小目标除了亮度信息几乎不存在其他的图像特征。红外弱小目标检测的主要任务是区分背景区域与目标区域。

1.2 国内外研究现状

本节先对目标检测算法的研究现状进行总结,其中包括传统的目标检测算法的研究现状和基于深度学习的目标检测算法的研究现状。然后,分别介绍小目标检测算法和对地小目标检测算法的研究现状。最后,分别对传统的红外弱小目标检测算法和基于深度学习的红外弱小目标检测算法的研究现状进行总结。

1.2.1 目标检测算法研究现状

传统的目标检测算法主要是利用手工设计的特征对目标进行检测。传统目标检测算法的检测流程如图 1.4 所示,主要包括以下 4 步:获取目标候选区域、特

征的提取、分类器分类和目标的筛选。

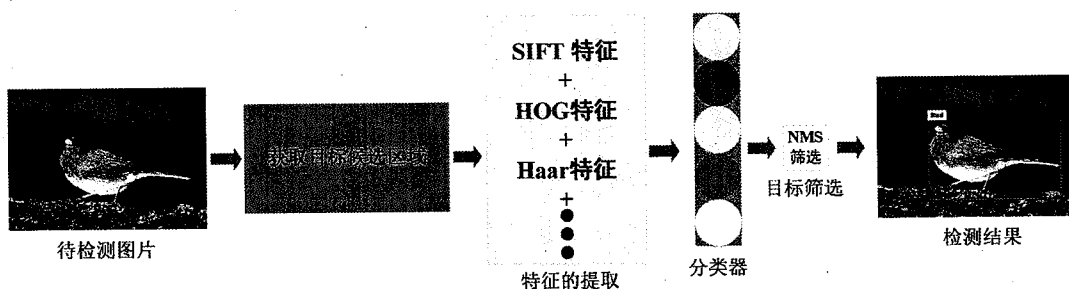


图 1.4 传统目标检测算法的流程

Figure 1.4 The workflow of traditional target detection method

第一步：获取目标候选区域。目标候选区域是对目标位置的初始预测，可能出现在图像中的任何一个位置，并且尺寸大小也不相同。滑动窗口遍历算法是一种比较经典的候选区域生成算法。该算法采用不同大小的窗口依次在图像上进行滑动，每滑动一步，就生成一个目标候选区域。基于滑动窗口的目标检测算法穷举了图像上所有目标可能出现的位置，大部分的目标候选区域都是负样本，因此算法冗余、效率低，检测结果的准确率不高。并且，目标候选区域数量特别大，对分类器的检测速度影响很大。后来，有学者提出了选择性搜索算法(Selective Search)^[44]，通过计算区域集中元素的相似度，不断地进行相似区域的合并，来生成候选目标区域。其中，元素的相似度采用颜色、尺寸和纹理等多个参数来进行计算，这使生成目标区域的复杂度增加。

第二步：候选区域的特征提取。作为目标检测中的关键步骤，选择表达能力强的特征有助于提升检测的准确率。Viola 等人^[45]提出使用图像的积分图作为人工分类的特征。Lowe 等人^[46]提出了尺度不变特征(Scale invariant feature transform, SIFT)，利用关键点方向和位置信息来建立关键点，因 SIFT 特征具有尺度不变性而得到广泛应用。Papageorgiou 等人^[47-48]提出了 Haar 特征，该特征反映了图像灰度的变化。Haar 特征分为：边缘特征、对角线特征、中心特征和线性特征，Haar 特征最早被应用于人脸检测中。Dalal 等人^[49]利用梯度方向直方图(Histogram of oriented gradient, HOG)对目标进行检测。这些人工设计的特征主要是依据梯度、纹理和形状等信息，往往不具有通用性。

第三步：分类器分类。常用的分类器包括 Softmax 分类器、支持向量机(SVM)分类器和 AdaBoost 分类器等。最具有代表性的目标分类器是由 Viola 等人^[45]提出的级联式 Adaboost 分类器，该分类器由多级 Adaboost 所构成，每一级又由多个弱分类器组成。前面的分类器用来判别目标是前景还是背景，后面的分类器用来判别目标的种类，应用这种级联方式极大地提高了目标检测的效率。

第四步：目标筛选。经过分类器后，图像上可能出现多个重叠度很高的目标框，需要进一步对这些目标框进行筛选，选择最优的检测结果。2006 年，Neubeck 等人^[50]提出了非极大值抑制(Non-Maximum Suppression, NMS)算法来过滤质量相对低的目标框。该方法根据置信度对每一个目标框进行排序，通过迭代对比最大置信度目标框与剩余目标框之间的交并比(Intersection over Union, IOU)来过滤掉质量低的目标框。然而，NMS 会将相邻的两个目标框混淆，从而错误地过滤掉正确的目标框。后来，Bodla 等人^[51]提出了 Soft-NMS，该方法不是直接过滤掉得分低的目标框，而是通过降低该目标框的得分来防止目标框的误删。

与传统的目标检测算法相比，基于深度学习的目标检测算法具有强大的特征表征能力。传统的目标检测算法采用手工设计的特征，而基于深度学习目标检测算法可以根据不同的任务自主地学习目标的特征。并且，网络提取的特征更加抽象，实现了对目标端到端、快捷、准确地检测。

基于深度学习的目标检测算法可以分为：双阶段(two-stage)目标检测算法和单阶段(one-stage)目标检测算法。其中，基于双阶段的目标检测算法将目标检测分为两个阶段，第一阶段是提取候选区域的特征信息，将图像分成前景区域和背景区域。然后再经过检测网络进一步完成对候选目标的位置和类别的预测和判断，是一种由粗到细的目标检测算法；基于单阶段的目标检测算法直接通过网络产生目标的位置和类别信息，是一种端到端的目标检测算法。因此，单阶段目标检测算法具有更快的检测速度，而双阶段目标检测算法具有更高的检测精度。

基于深度学习的双阶段目标检测算法：Girshick 等人^[52]提出了 RCNN 目标检测算法。该算法的检测流程如图 1.5 所示，首先利用颜色、纹理、相似度等特征生成候选区域，然后利用卷积神经网络对每个候选区域进行深度特征的提取。最后利用 SVM 分类器对这些深度特征进行分类并对 bounding box 进行回归。RCNN

与传统目标检测算法最大的不同是将手工设计的特征转换成了基于卷积神经网络(Convolutional neural network, CNN)的深度特征。然而, 由于需要提取大量的候选区域的特征, 并且候选区域之间存在大量的重叠, 导致检测速度特别慢。He 等人^[53]对 RCNN 网络进行了改进, 提出了 SPPnet。SPPnet 不需要将所有的候选区域都输入到卷积神经网络, 只需要将整张图片输入到网络中, 然后根据候选区域的位置, 在特征图的相应位置找到对应的窗口, 大幅度提高了检测速度。并且, SPPnet 引入了空间金字塔池化(Spatial Pyramid Pooling), 可以适应任何大小的图片作为输入。随后, Girshick 等人^[54]在 SPPnet 和 RCNN 的基础上提出了 Fast RCNN。Fast RCNN 在网络中增加了感兴趣区域池化层(Region of Interest Pooling, ROI Pooling), ROI Pooling 可以使具有不同分辨率的输入图片输出相同大小的特征图。并且, Fast RCNN 利用 Softmax 分类器对目标进行分类, 通过引入类内竞争来提升目标检测的精度。但是, Fast RCNN 仍然没有实现网络端到端的训练。

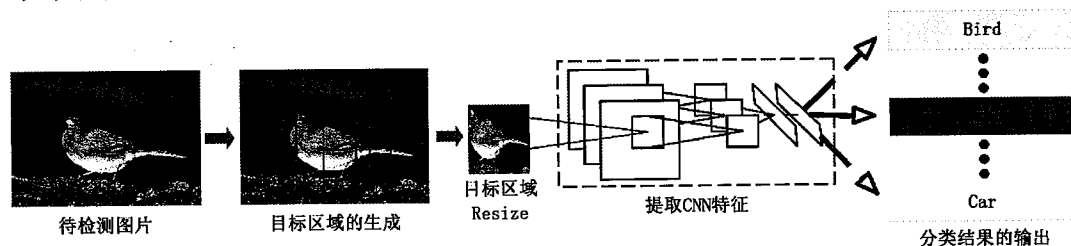


图 1.5 RCNN 目标检测算法的流程

Figure 1.5 The workflow of RCNN target detection method

Ren^[55]等人提出了 Faster RCNN, Faster RCNN 引入锚点框(anchor box)来初始化候选框, 然后利用区域候选网络(Region Proposal Network, RPN)来生成目标候选区域。Faster RCNN 的检测流程如图 1.6 所示, 通过使用 RPN 来代替 RCNN 中生成候选区域的方法, 使得目标的检测精度和速度得到进一步的提高, 同时实现了对网络端到端的训练。Faster RCNN 奠定了双阶段目标检测算法的基本框架。He 等人^[56]提出了 Mask-RCNN, Mask-RCNN 可以同时被应用到目标检测和分割任务中。另外, Mask-RCNN 采用 ROI Align 进一步提高网络的检测精度。2018 年, Cai 等人^[57]提出了级联 RCNN(Cascade RCNN)网络, 将双阶段目标检测算法扩展为多阶段目标检测算法, 通过对目标进行多阶段分类和回归, 使回归框定位

更精确，分类更准确。

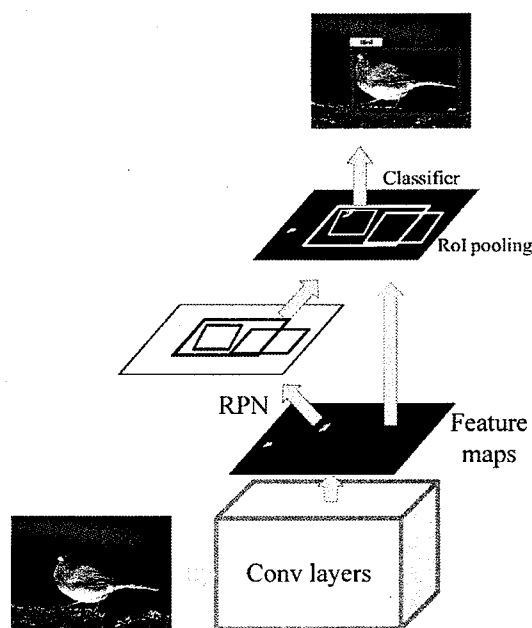


图 1.6 Faster RCNN 目标检测算法的工作流程

Figure 1.6 The workflow of Faster RCNN target detection method

基于深度学习的单阶段目标检测算法：基于深度学习的单阶段目标检测算法不用生成目标候选区域，可以直接对目标的类别和位置进行回归，极大提高了目标检测的速度。Redmon 等人^[58]提出了 YOLO (You Only Look Once) 目标检测算法。YOLO 直接将图片划分成小网格，并在每个小网格上对目标的位置信息和类别进行回归。由于不需要生成候选区域，所以 YOLO 具有极快的目标检测速度。但是对于小目标和邻近目标的检测率较低。随后 Redmon 等人^[59]又提出了 YOLO 9000，YOLO 9000 采用 Darknet19 网络架构，利用跳跃连接将目标的浅层特征与深层的语义特征相融合，从而提高了网络对目标的检测率。并且，YOLO 9000 借鉴了 Faster RCNN 中 anchor box 的概念，利用 Kmeans 聚类算法对数据集中的目标进行聚类分析，获得 anchor box 的长和宽，使选取的 anchor box 更加符合数据集中目标大小的分布。

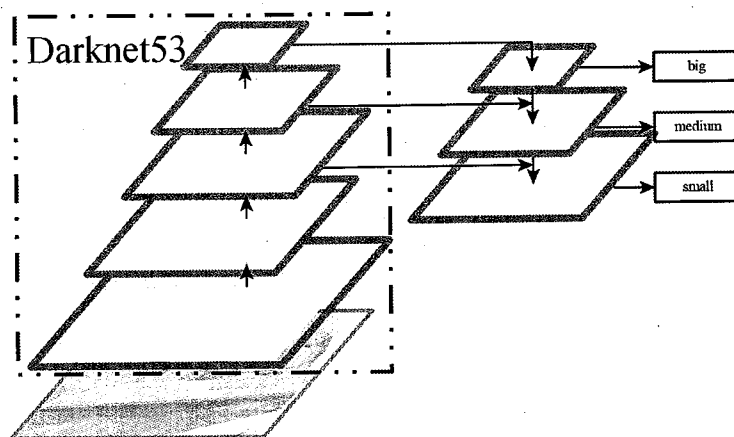


图 1.7 YOLO V3 目标检测算法

Figure 1.7 YOLO V3

2018 年, Redmon 等人^[60]又提出了 YOLO V3, YOLO V3 借鉴了残差网络 (Residual network)^[26]的思想, 通过建立 Darknet53 网络来提取目标特征, 并利用特征金字塔网络 (Feature Pyramid Network, FPN)^[61]完成对大、中、小 3 个尺度目标的检测, 使其对各尺度的目标都有较高的检测精度, YOLO V3 的结构如图 1.7 所示。由于 Darknet53 采用了 3×3 和 1×1 的小卷积核, 所以, YOLO V3 具有极快的检测速度, YOLO V3 可以实现对目标的高精度实时检测。2015 年, Liu 等人^[62]提出了 SSD (Single Shot Multi-box Detector) 的单阶段目标检测算法。SSD 采用了不同尺度和长宽比的先验框, 有利于目标的定位回归。并且, 对目标采用多尺度的检测, 利用大特征图检测小目标, 用小特征图来检测大目标, 使得 SSD 对不同尺度的目标都有较好的检测效果。Lin 等人^[63]提出了 RetinaNet, 他们分析了单阶段目标检测算法检测精度低于双阶段目标检测算法的原因, 认为是由于背景与目标分布不均衡导致的, 于是提出了 Focal loss 来替代原 loss 函数, Focal loss 通过降低简单样本的权重, 使得训练模型可以聚焦于困难样本的回归。该算法极大提高了对目标检测的准确率, 并且对检测速度没有影响。2018 年, Etten^[64]提出了 You Only Look Twice 算法。将单阶段目标检测算法应用到了高分遥感图像的目标检测中。Tian 等人^[65]提出了一种全卷积的目标检测算法 FCOS (Fully Convolutional One-Stage Object Detection), FCOS 不需要预设锚点框, 避免了与锚点框相关的复杂计算, 利用像素预测的方式解决目标检测问题。Law 等人^[66]

提出了 CornerNet, 利用单个卷积神经网络将框的左上角及右下角两个点组成一对关键点, 不需要设计锚点框。另外, CornerNet 引入了 corner pooling 来提升对角点的定位效果。Zhou 等人^[67]提出了 CenterNet, 利用目标中心点来呈现目标。CenterNet 在中心点位置回归出目标的一些属性, 包括: 尺寸、方向和姿态等。该算法将目标检测问题转换成一个标准的关键点估计问题, 只需将图像输入到全卷积神经网络, 就可以得到一个热力图。然后, 根据热力图上的峰值点来判断目标中心点的位置。最后, 通过每个特征图的峰值点位置来预测目标的宽高信息。FCOS、CornerNet 和 CenterNet 都是基于无锚框(anchor-free)的目标检测算法, 由于无锚框的目标检测算法不需要预设锚点框, 近年来受到的关注度逐渐提升, 其它的代表算法还包括 Densebox^[68]、Unitbox^[69]和 DeNet^[70]等。

1.2.2 小目标检测算法研究现状

目前, 大多数的小目标检测模型是由常规目标检测模型改进得到的。Lin 等人^[61]提出了 FPN(Feature Pyramid Network)网络架构。FPN 对小的特征图进行上采样, 然后与大的特征图进行融合。由于大特征图中具有更多的细粒度特征, 如位置信息, 小特征图中具有更多的粗粒度特征, 如语义信息。二者信息的融合既可以提高网络对小目标的定位能力又可以提高网络对小目标的分类能力。Chen 等人^[71]改进了 RCNN 目标检测算法, 通过融合一个上下文信息的模块, 来提升对小目标的检测能力。改进 RCNN 目标检测算法的结构如图 1.8 所示, 利用两个输入通道将小目标区域的特征和包含小目标上下文信息的特征进行融合。另外, 作者利用 COCO 和 SUN^[72]数据集制作了一个极具挑战性的小目标检测数据集。

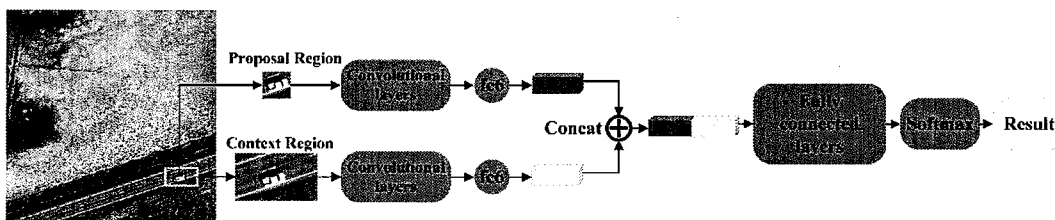


图 1.8 改进 RCNN 的小目标检测算法

Figure 1.8 Improved RCNN for small target detection

Eggert 等人^[73]利用改进的 Faster RCNN 对交通指示标志进行检测。同时, 提

出了一种新的 anchor box 的生成机制, 该机制可以为每个检测尺度选择合适的 anchor box。另外, 作者通过实验分析了不同分辨率的特征图与目标尺寸之间的关系, 并利用高分辨率特征图来对小目标进行检测。

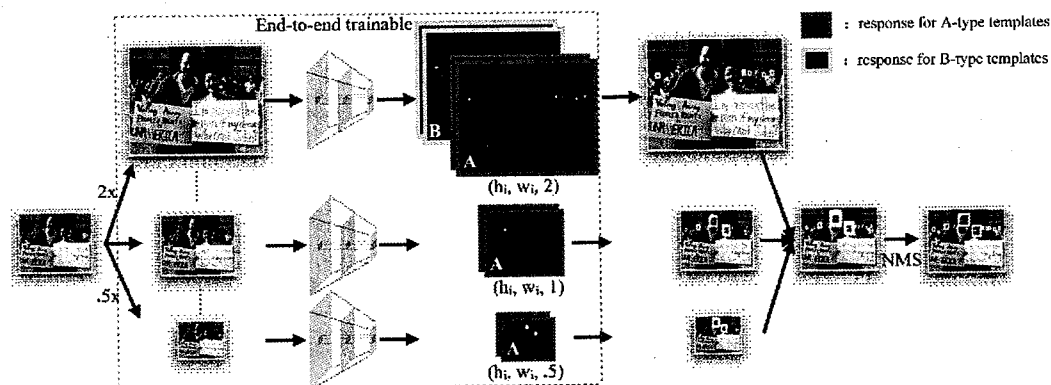


图 1.9 小人脸检测算法

Figure 1.9 The method for tiny face detection

Hu 等人^[74]对小人脸检测进行了研究, 详细分析了图像的分辨率、目标的尺度和空间上下文信息对小人脸检测的影响。如图 1.9 所示, 作者采用不同分辨率的输入图片来检测尺寸不同的人脸, 最后将检测的不同尺寸的人脸还原到原图像上, 并利用非极大值抑制来过滤质量较差的人脸。

以上基于深度学习的小目标检测算法都是基于双阶段的目标检测算法, 这些算法大多不适用于对目标的实时检测。

另外, Krishna 等人^[75]改进了 Faster RCNN, 并借鉴超分辨的思想对小目标进行检测。Fu 等人^[76]提出了 DSSD 网络来对小目标进行检测。DSSD 是对 SSD 目标检测算法的改进, 作者利用反卷积(Deconvolution)对低分辨率特征图进行上采样, 并与高分辨率特征图进行特征融合, 来提高网络对小目标的检测能力。Li 等人^[77]将生成对抗网络(Generative Adversarial Network, GAN)应用到了小目标检测中, 在网络内部实现了将小目标低分辨率特征转换成高分辨率特征的目的。该算法取得了较好的小目标检测结果。但是, 该算法是基于双阶段的目标检测算法, 并且增加了 GAN 网络, 因此, 对小目标的检测速度会受到限制。

1.2.3 对地小目标检测算法研究现状

目前,大多数对地小目标的数据集是由高分辨率卫星和无人机拍摄的,对地小目标检测算法多为遥感图像目标检测算法。

传统的遥感图像目标检测算法也是利用滑动窗口来获取目标候选区域,然后利用手工特征对目标进行检测^[78-83]。Zhong 等人^[84]提出了基于级联卷积神经网络的遥感图像车辆目标检测算法。该算法包括两个子网络,分别是车辆候选区域生成网络和遥感图像车辆目标检测网络。该算法对两个子网络采用单独训练的策略,避免两个网络的非共享层重复训练。Deng 等人^[85]借鉴 Faster RCNN 的思想,提出了一种快速精确的无人机地面车辆目标检测算法。该算法首先设计一个车辆候选区域网络(accurate vehicle proposal network, AVPN),通过融合深层和浅层的特征来实现对小车辆目标的精准检测。另外,该算法还设计了一个车辆属性自学习网络(vehicle attributes learning network, VALN)来学习遥感图像车辆目标的类别和方向信息。Tayara 等人^[86]设计了一个端到端的全卷积回归网络 FCRN 来检测遥感图像车辆目标,并对车辆目标数量进行统计。FCRN 采用先下采样后上采样的结构来提高对遥感图像小车辆目标位置的回归精度。Zhang 等人^[87]提出了一种实时的无人机地面车辆检测网络 DAGN。DAGN 利用注意力机制模块来融合不同尺度的特征,学习不同尺度特征之间的相关性。另外,Zhang 等人提出了高斯非极大值抑制(Gaussian non-maximum suppression)方法来提高对密集车辆的检测效果。Terrail 等人^[88]对 Faster RCNN 进行了改进,提出了 Faster RER-CNN。该算法不仅能检测遥感图像目标车辆的位置,还能检测目标车辆的方向信息。

1.2.4 红外弱小目标检测算法研究现状

随着红外成像技术的不断发展,红外弱小目标检测已经成为目标检测领域的研究热点,并广泛应用在预警、制导和空间监视等国防领域。

经典的红外弱小目标检测算法如形态学滤波算法^[89]只对目标亮度高,背景单一的场景下的目标监测有效。Max-mean 和 Max-median 算法^[90]利用红外图像来对图像中的背景进行预测,然后用原红外图像与预测的背景图像求差得到最后的目标图像,达到增强小目标响应的目的,该算法检测的目标图像中常常包含一些背景的边缘信息。Zeng 等人^[91]基于数学形态学提出了顶帽变换(Top-Hat)方法,

该方法首先设计合适的结构元素,然后通过形态学的开闭运算滤除图像中小于结构元素的奇异点,改善对红外弱小目标的检测效果,最后对原图像和预测的背景图像进行差分处理,检测图像中的弱小目标。当背景中包含较多杂波和噪声、目标信号强度较弱时,上述算法的检测性能和实时性都比较差。

受人类视觉系统(Human Visual System, HVS)的启发,许多基于HVS的红外弱小目标检测算法被提出。Chen等人^[92]提出了一种基于局部对比度测量(Local Contrast Measure, LCM)的红外弱小目标检测算法,主要是利用小目标的灰度值要比目标周围邻域的平均灰度值高的特点。该算法使用中心单元的最大灰度值和相邻单元的平均灰度值的比值作为局部对比度测量。利用 3×3 的小区域来遍历图像,并计算每个小区域的对比度,把具有最大对比度的区域作为目标。由于该算法要通过多次改变区域的大小来适应多尺度的目标,所以实时性比较差。Han等人^[93]对LCM算法进行了改进,提出了基于改进的局部对比度测量(Improved Local Contrast Measure, ILCM)的红外弱小目标检测算法。为了提高红外图像的信杂比,ILCM首先使用DoG滤波器对红外图像进行滤波,然后,利用中心单元灰度值的均值和最大值的乘积来替换LCM算法中的中心单元的最大灰度值,有效降低了虚警率。随后,Deng等人^[94]提出了一种附加权值的局部差异测量(Weighted local difference measure, WLDM)红外弱小目标检测算法。该算法将图像中的每一个像素作为中心,并将该像素邻域内的熵作为权重,与该像素在不同尺度下的局部差分最大值相乘,得到最后的对比度,有利于提高对红外弱小目标的检测精度。

Hu等人^[95]利用高斯模型来建立目标的字典,通过主成分分析并基于重构残余对红外图像进行分类。该算法过分依赖目标字典的模型。Gao等人^[96]提出了红外块图像(Infrared Patch Image, IPI)模型来进行红外弱小目标的检测,该模型基于红外图像的非局部自相关特性进行目标检测。IPI算法首先利用滑动窗口来获得图像块,然后通过矢量化来为每一个图像块进行建模。该算法的本质是通过采用鲁棒主成分分析技术(Robust principal component analysis, RPCA),分别用低秩矩阵和稀疏矩阵来描述背景图像和前景图像,达到分离目标和背景的目的。随后,一种附加权重的IPI模型(Weighted Infrared Patch image: WIPI)^[97]被提出,该算法

利用权重分配使预测的背景图像更加贴近于真实的图像背景。

虽然,基于深度学习的目标检测算法在计算机视觉领域已经取得了许多突破,但是大部分算法都是针对常规目标。目前针对基于深度学习的红外弱小目标检测算法的研究还比较少。

Yang 等人^[98]将红外弱小目标检测任务与语义分割联系起来,为解决红外弱小目标检测提供了新的思想。该算法利用递归操作和残差模块来构建全卷积递归网络,学习复杂背景下的红外弱小目标的特征,并采用红外弱小目标仿真实验数据集对递归网络进行训练,实现了对真实红外弱小目标的检测。

Chen^[99]等人将 YOLO V3 目标检测算法应用到了红外末制导中,提出了基于 YOLO V3 的红外末制导典型目标检测算法。该算法通过优化损失权重,提高了网络对红外目标的定位和分类能力。虽然该算法将基于 CNN 的目标检测算法应用到了红外目标检测,但是检测的目标相对较大,并不适用于红外弱小目标检测。

Liu^[100]等人提出了融合全卷积神经网络和视觉目标显著性的单阶段红外弱小目标检测算法。该算法首先设计了一个轻量化的卷积神经网络,通过对原始红外图像的分割实现对弱小目标的增强和图像背景的抑制。随后,利用红外弱小目标的显著性特征进一步抑制虚警,实现对红外弱小目标的检测。

Wu^[101]等人通过对红外弱小目标位置分布进行分类来完成对红外弱小目标的检测,提出了一种基于深度卷积神经网络的红外图像小目标检测算法。如图 1.10 所示,该检测网络主要由全卷积神经网络和分类网络构成。其中,全卷积神经网络是用来对红外图像中的小目标进行初始筛选,实现背景的抑制和目标的增强。分别将原红外图像和全卷积神经网络输出的图像输入到分类网络中,对红外弱小目标进行进一步的筛选。另外,网络通过引入注意力机制模块(Squeeze and Excitation Networks, SEnet)来学习不同通道之间的相关性,增强网络的表达能力。

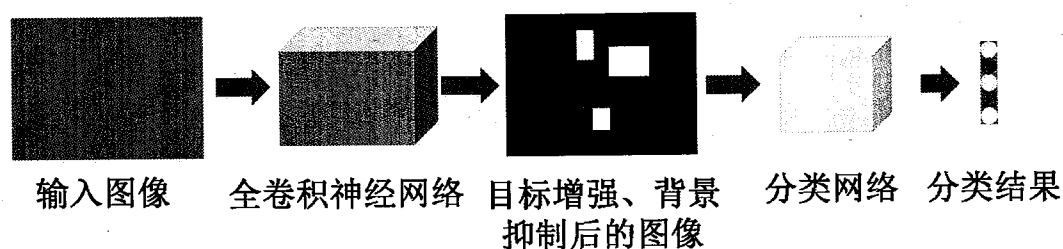


图 1.10 检测网络结构

Figure 1.10 Structure of the detection network

另外, Jiang 等人^[102]提出了基于卷积神经网络的红外舰船目标检测算法。Chen 等人^[103]提出了 Nv-Net 红外目标检测算法, 利用基于卷积神经网络的分割技术来检测红外目标。Shi 等人^[104]设计了一个自动编码和解码的网络来检测红外弱小目标。

1.2.5 基于深度学习的超分辨率重建算法研究现状

小目标在图像中占用的像素少, 携带的特征信息不明显。利用超分辨率重建技术可以将低分辨率图像重建为高分辨率图像, 例如, 将一张分辨率为 100×100 的图像提高到分辨率为 200×200 的图像, 并且要求图像的清晰度尽量不变, 即输出的高分辨率图像比输入的低分辨率图像包含更多的细节信息^[105]。因此, 将超分辨率重建技术与小目标检测技术相结合, 可以通过提高小目标的分辨率来提高对小目标的检测性能。

超分辨率重建技术已经被广泛应用于卫星遥感图像和医学图像处理中。本节将介绍基于深度学习的单张图像超分辨率(Single Image Super-Resolution, SISR)重建技术的研究现状。

SRCNN^[106]是第一个将深度学习应用于超分辨率重建技术上的卷积神经网络。该网络结构简单, 仅仅包含三个卷积层, 网络的结构如下图所示。

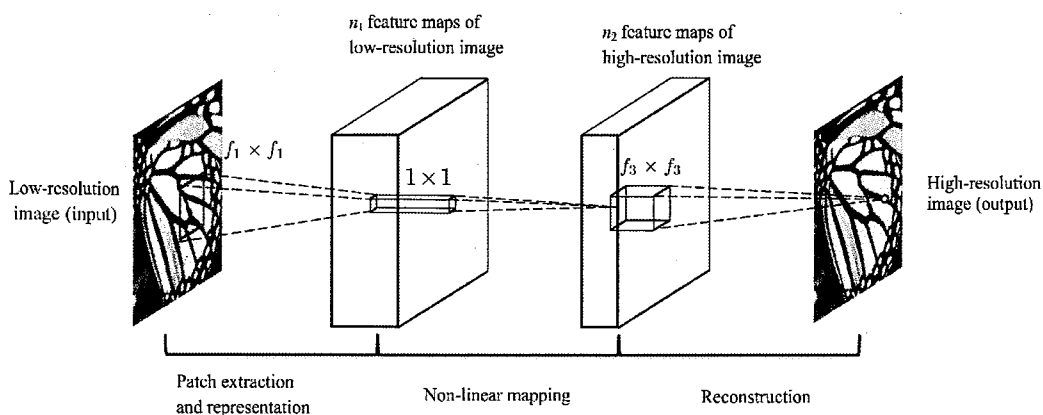


图 1.11 SRCNN 结构图

Figure 1.11 The structure of SRCNN

SRCNN 首先采用双三次插值将输入的低分辨率图像放大成高分辨率图像的尺寸，然后将插值后的图像送入到由 3 个卷积构成的网络中，这 3 个卷积的卷积核大小分别为 9×9 、 1×1 和 5×5 。利用这 3 个卷积完成图像块特征的提取与表示、非线性映射和高分辨率图像的重建。作者采用均方误差(Mean-Squared Error, MSE)作为损失函数。SRCNN 为基于深度学习的超分辨率重建方法提供了新的思路。由于该算法仅仅采用了 3 个卷积层，限制了网络对特征的表达能力和泛化能力，生成的高分辨率图像效果一般。

2016 年，Dong 等人^[107]对 SRCNN 进行了改进，提出了如图 1.12 所示的超分辨率重建网络FSRCNN。该网络主要在 SRCNN 网络基础上做了两方面的改进，一是利用反卷积层来代替插值操作，完成了对输入的低分辨率图像的放大。通过这种方式，可以使网络采用低分辨率的图像作为输入。二是改变了卷积层的数量，增加了输出特征的维数，并通过串联多个卷积核为 3×3 的卷积层来代替 SRCNN 中 9×9 和 5×5 的大卷积核。在增强网络表达能力的同时，又提高了重建速度。

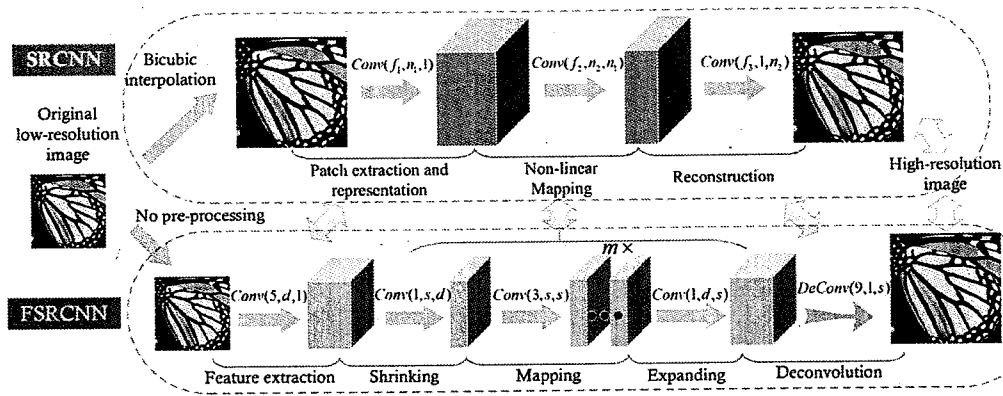


图 1.12 SRCNN 与 FSRCNN 的对比

Figure 1.12 The comparison of SRCNN and FSRCNN

2016 年, Shi^[108]等人提出了 ESPCN 网络, 该网络利用亚像素卷积层(Sub-pixel convolution layer)来实现高分辨率图像的重建。亚像素卷积层可以直接实现低分辨率特征图到高分辨率特征图的转换, 不需要 SRCNN 的预处理方式, 减少了算法的复杂度。

如图 1.13 所示, 亚像素卷积层是一种快速、高效、无参, 利用像素重排进行上采样的操作。亚像素卷积层实现了将维度为 $H \times W \times r^2 C$ 的特征图转换成维度为 $rH \times rW \times C$ 的特征图。即将低分辨率特征图中 r^2 个通道的像素重新排列成 $r \times r$ 的区域, 进而将维度为 $H \times W \times r^2 C$ 的低分辨率图排列成维度为 $rH \times rW \times C$ 高分辨率图。ESPCN 在网络的末段利用亚像素卷积层来实现上采样, 有利于在低分辨率空间中保留更多的纹理特征。另外, ESPCN 可以实时地完成对高分辨率图像的重建。

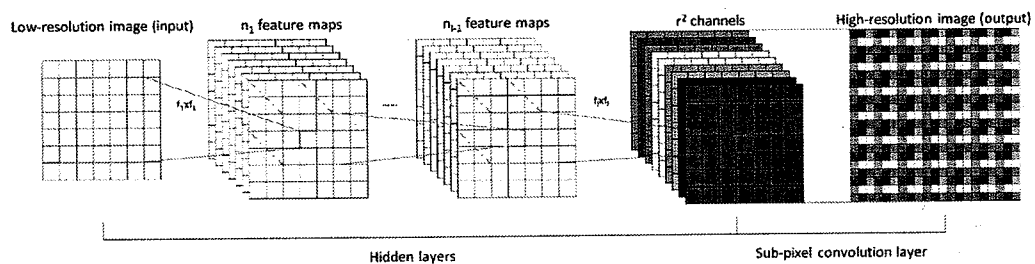


图 1.13 ESPCN 的结构

Figure 1.13 The structure of ESPCN

Kim^[109]等人提出 VSDR 超分辨率重建网络,该模型引入了残差网络(Residual network)的结构,解决了网络过深无法训练的问题。VSDR 的结构如图 1.14 所示,通过利用残差网络结构使 VSDR 网络的深度更深,一共包含 20 个卷积层。另外, VSDR 在每次卷积前都对图像进行补 0 操作,可以保证所有的特征图都具有相同的尺寸,解决了图像经过卷积后越变越小的问题。另外, VSDR 采用不同倍数的图像进行训练,使 VSDR 可以同时适用不同倍数的超分辨率重建任务。

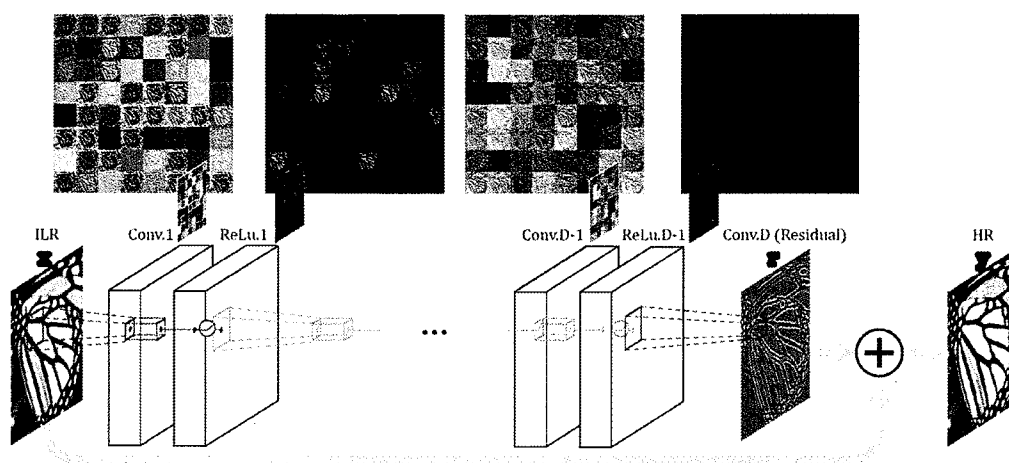


图 1.14 VSDR 的结构

Figure 1.14 The structure of VSDR

Mao 等人^[110]提出了 RED-Net(Very deep Residual Encoder-Decoder Networks)。该网络的结构是对称的,利用多个卷积层作为网络的编码器来提取特征,利用多个反卷积层作为解码器来重建高分辨率图像。另外,RED-Net 同样借鉴了 Residual network 的思想,每一个卷积层都与相应的反卷积层通过跳跃连接的方式来实现相同尺寸特征图的特征融合。通过这种方式使 RED-Net 的网络更深,达到了 30 层。

Tong 等人^[111]提出了 SRDensNet,该网络主要借鉴了 DenseNet 的思想,利用稠密连接将每一层的特征都串联起来,可以防止梯度消失并增强特征复用。SRDensNet 的框架可以简要概括为:利用一个卷积层来提取浅层的特征,然后利用 DenseNet 的结构实现深度特征的提取,然后利用反卷积层对高分辨率图像进行重建。

2017 年 Ledig 等人^[112]首次将生成对抗网络 (Generative Adversarial Network,

GAN)应用到了超分辨率重建中,提出了 SRGAN 网络。文中还分析了超分辨率重建网络采用均方差作为损失函数带来的不足,虽然均方差损失能够提高峰值信噪比,但是通常重建的图像中会缺失高频细节信息,影响人的视觉感受。

1.3 研究中的难点和挑战

本文主要是针对对地小目标检测算法和对空红外弱小目标检测算法进行研究。与自然场景中的目标相比,对地小目标检测存在以下难点^[113-114]:

(1) 目标特征少。图 1.15 对比了自然场景中的车辆和对地小车辆目标。对地小目标在图像中占有的像素非常少,分辨率低,通常为几十个或者十几个像素,有时甚至几个像素。所以对地小目标特征通常不明显,携带的信息特别少,缺少形状和纹理等特征信息。

(2) 背景复杂,目标数目多。对地小目标图像背景复杂,目标数量较多。当环境中掺杂噪声时,目标的特征将变得更加模糊。复杂环境将导致对地小目标的检测率降低,虚警率升高。

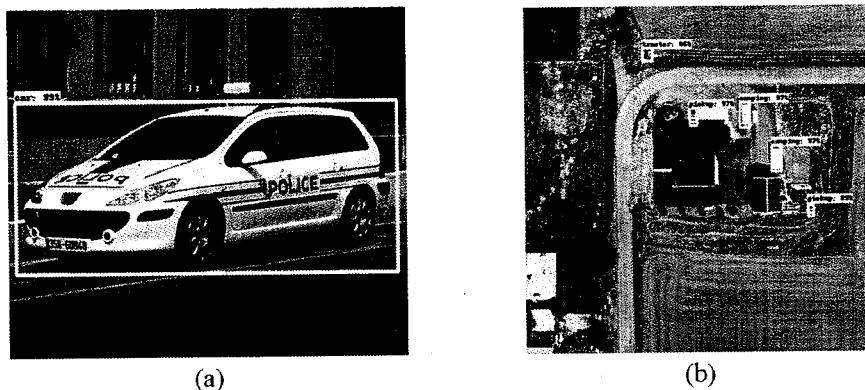


图 1.15 遥感图像车辆目标与自然场景中的车辆对比

Figure 1.15 Comparison between the vehicles in remote sensing image and the ones in natural scene

对地小目标和红外弱小目标的对比如图 1.16 所示,其中,图 1.16(a)为对地小目标,图 1.16(b)为红外弱小目标。与自然场景中的目标和对地小目标相比,红外弱小目标检测具有以下特点^[115]:

(1) 相比于自然场景中的目标和对地小目标,红外弱小目标不具备细致的纹

理结构和丰富的颜色信息，使得红外弱小目标检测变得更加困难。

(2) 如图 1.16(b)所示，红色框代表红外弱小目标的位置。红外弱小目标通常是利用红外成像系统远距离采集得到的，因此对比度和信噪比都比较低。

(3) 红外弱小目标在红外图像上通常表现为一个小斑点，往往只占几个像素，不具备细致的纹理结构和形状特征。由于成像距离远，使检测到的目标信号很弱，并且伴随着噪声的干扰。最终，使目标呈现出“弱小”的特点。相比于对地小目标，红外弱小目标的检测尺度范围更小。

(4) 数据集少。目前，针对红外弱小目标的公开数据集较少。

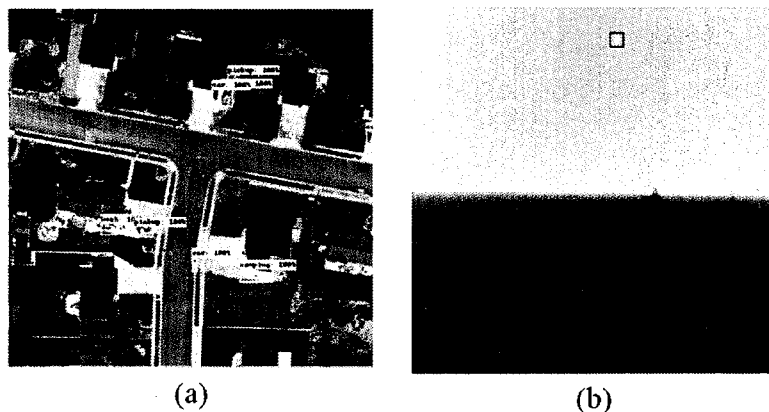


图 1.16 对地小目标和红外弱小目标对比

Figure 1.16 Comparison between infrared dim and small targets and ground small targets

综上所述，无论是对地小目标检测还是对空红外弱小目标检测都是目标检测领域中的挑战。对地小目标检测和红外弱小目标检测都可用于监视远距离目标，有助于实现“先敌发现，先敌锁定和先敌命中”，对国防事业具有重要的应用价值。因此，对地小目标检测技术和对空红外弱小目标检测技术是一项非常困难又极具实际意义的研究课题。

1.4 论文研究内容与章节安排

本文分别从对地小目标检测和对空红外弱小目标检测的应用背景和意义出发，分析了它们在国防领域中的需求。围绕基于深度学习的对地小目标检测算法和基于深度学习的红外弱小目标检测算法开展研究。通过对影响对地小目标检测

和对空红外弱小目标检测的关键因素进行分析,提出了相应的解决策略。本文研究内容的框架如图 1.17 所示。

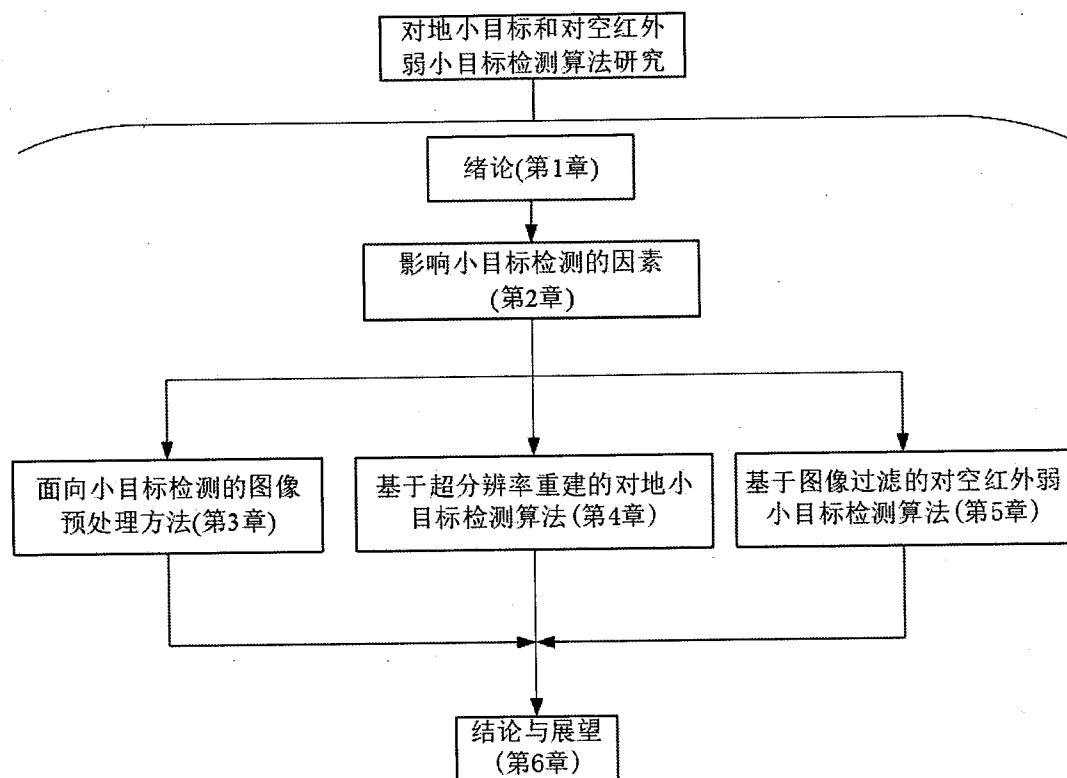


图 1.17 研究内容的框架图

Figure 1.17 The framework of research content

全文内容安排如下:

第1章: 绪论。本章首先介绍对地小目标检测算法、对空红外弱小目标检测算法和深度学习的研究背景及意义。阐述了当前国内外目标检测算法的研究现状, 包括: 目标检测算法、小目标检测算法, 对地小目标检测算法、红外弱小目标检测算法的研究现状。另外, 还介绍了基于深度学习的超分辨率重建算法的研究现状。然后, 分别分析了对地小目标检测和对空红外弱小目标检测面临的难点和挑战。最后, 列出本文的主要研究内容与章节安排。

第2章: 影响小目标检测的因素。对地小目标和对空红外弱小目标的特点就是在图像中占有像素少、目标尺寸小、特征不明显, 易受周围环境的影响。因此, 对地小目标和对空红外弱小目标检测更具有挑战性。为了设计适合于对地小目标

和对空红外弱小目标的算法,本章首先对影响小目标检测的关键因素进行了分析和总结。并针对各个因素,提出相应的解决方法,为后面的对地小目标检测算法和对空红外弱小目标检测算法的设计奠定基础。

第3章:面向小目标检测的图像预处理方法。本章首先提出了一种基于图像裁剪的对地小目标图像预处理方法。在对地小目标图像中,大部分区域都是背景,只有少部分是目标区域。根据此特点,裁掉图像中多余且不包含目标的区域,可以在不增加显存的条件下,实现网络对同一目标多尺度的训练,使网络变得更加鲁棒。另外,根据红外弱小目标信号“弱”、“小”且常常伴随有噪声干扰的特点,提出了基于局部对比度的红外弱小目标图像预处理方法。该方法对局部对比度测量算法进行了改进,提高了目标与背景之间的对比度。另外,该方法可以利用卷积和最大池化操作来实现,因此,它可以实现对图像快速、高效的处理。

第4章:基于超分辨率重建的对地小目标检测算法。本章将超分辨率重建技术与目标检测技术相融合,提出了基于超分辨率重建的对地小目标检测网络 SuperDet。SuperDet 主要由两部分构成,分别是超分辨率模块和对地小目标检测模块。超分辨率模块用来实现将低分辨率的图像转换成高分辨率的图像。对地小目标检测模块将重建的高分辨率图像作为输入,完成对目标类别和位置的预测。为了给对地小目标选择合适的网络尺度,提出了尺度匹配策略,为对地小目标检测模块设计了主干网络。为了增加目标周围的上下文信息,本章提出了加权感受野融合模块(weighted receptive field fusion module),可以自适应地融合具有不同感受野的特征。本文分别采用 GIOU loss、Focal loss 和 Binary Cross Entropy loss 来对对地小目标的位置、置信度和类别进行回归。另外,本章还提出了一种正样本选择策略,可以缓解正负样本数量不均衡的问题。为了实现对 SuperDet 端到端的训练,采用多任务损失函数对超分辨率重建任务的损失和目标检测任务的损失进行回归。在 VEDAI 数据集上进行消融实验,实验结果证明了 SuperDet 中每一个设计都有助于提高算法对对地小目标的检测能力。为了验证 SuperDet 对对地小目标的检测性能,分别在 VEDAI 可见光数据集、VEDAI 红外数据集和 DOTA 数据集上与目前流行的目标检测算法进行对比实验,实验结果表明, SuperDet 具有更好的对地小目标检测性能。

第5章：基于图像过滤的对空红外弱小目标检测方法。红外弱小目标不仅信号强度弱，并且常常受到环境中杂波和噪声的干扰。基于此，本章提出了基于图像过滤的红外弱小目标检测网络 FilterDet 来抑制图像中的噪声和干扰。FilterDet 主要由两部分组成，分别是过滤模块和红外弱小目标检测模块。其中，过滤模块实现了对噪声的过滤和对目标的增强。红外弱小目标检测模块将过滤后的图像作为输入完成对红外弱小目标的检测。根据尺度匹配策略，采用沙漏结构作为红外弱小目标检测模块的主干网络。利用加权感受野融合模块来增加小目标周围的上下文信息。另外，通过引入空间注意力机制来增强网络的表达能力，使红外弱小目标的特征响应得到加强。为了实现对 FilterDet 端到端的训练，采用多任务损失函数对图像过滤任务的损失和目标检测任务的损失进行回归。为了验证 FilterDet 对红外弱小目标的检测性能，分别在合成的单帧红外弱小目标数据集和 3 个真实的红外弱小目标序列上进行对比实验。实验结果表明，FilterDet 具有较为出色的红外弱小目标检测性能。

第6章：结论与展望。总结本文的主要工作和创新点，并对未来的工作进行展望。

第2章 影响小目标检测的因素

2.1 引言

由于成像距离远，对地小目标和对空红外弱小目标在图像中占有的像素非常少、特征不明显，且受环境因素影响大。因此，本章主要分析影响小目标检测的关键因素，根据这些因素来设计对地小目标检测算法和对空红外弱小目标检测算法，有助于提高算法对对地小目标和对空红外弱小目标的检测性能。

2.2 影响小目标检测的关键因素

本节根据对地小目标和红外弱小目标的特点，总结了影响小目标检测的因素包括：分辨率低、正负样本数量不均衡、小目标周围的上下文信息、小目标检测尺度的选择和特征融合。对于红外弱小目标，增强目标响应并抑制背景噪声是提高红外弱小目标的关键。

1. 分辨率低：小目标在图像中占有的像素非常少，分辨率低，缺少细节特征。图 2.1 对比了相同场景下分辨率分别为 512×512 和 1024×1024 的图像，很明显，分辨率为 1024×1024 的图像中的小汽车包含了更多的细节特征。因此，提高小目标的分辨率将有助于提高算法对小目标的检测率。

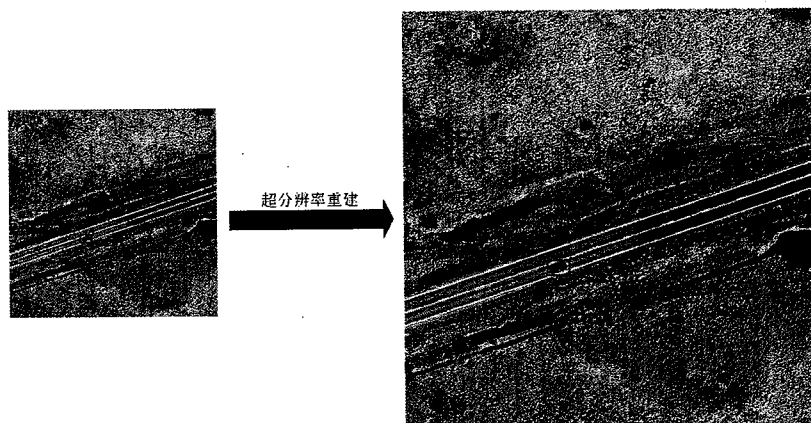


图 2.1 相同场景下不同分辨率图像的对比

Figure 2.1 Comparison of different resolution images in the same scene

Hu 等人^[74]通过对人脸数据集进行分析和实验,发现对小人脸进行 2 倍上采样(双线性插值)之后,有助于检测小人脸。实验结果如图 2.2 所示,当人脸的尺寸小于 43×34 时,提高小人脸的分辨率会有效提高检测算法对小人脸的检测率。

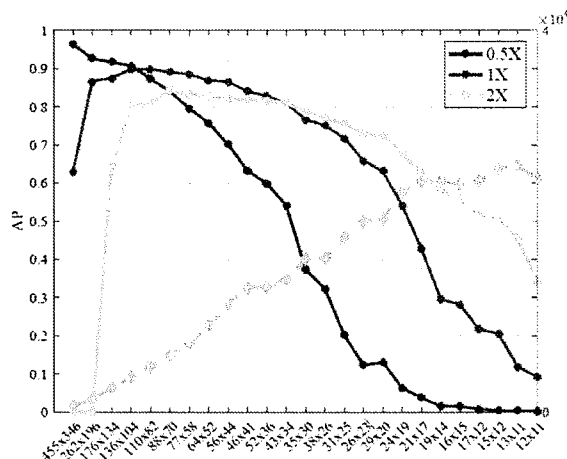


图 2.2 人脸分辨率对检测结果的影响

Figure 2.2 The influence of face resolution on detection results

另外,用具有不同分辨率的目标来训练网络,使网络能够学习同一个目标,不同尺度、不同分辨率的特征和信息,有助于提高检测网络的鲁棒性^[59]。然而,在图片进行多尺度变换时,会增大图像的分辨率,需要占用更多的计算机显存。因此,研究如何在不增加计算机显存的情况下,对小目标进行多尺度训练不仅可以提高网络对小目标的检测性能,还可以节省显存资源。用具有不同分辨率的目标来训练网络适用于对地小目标检测,并不适用于红外弱小目标检测。因为红外弱小目标通常非常小,尺度变化不大。

2. 正负样本的数量不均衡:在小目标检测任务中,由于小目标在图像中占有的像素非常少,导致图像的大部分区域都是背景。这使大部分的样本都成为负样本,只有少数的为正样本,造成了正负样本的极度不均衡,不利于卷积神经网络的训练。特别是红外弱小目标,通常只占有几个像素,且经常会受到复杂环境中噪声和杂波的影响。图 2.3 对比了不同大小的目标在相同特征图中正负样本分配的情况,图中的红框代表锚框(anchor box),特征图中的每一个小网格都分配一个锚框,通常将锚框与目标之间交并比(IOU)大于某个阈值的锚框作为正样本,其它的锚框作为负样本。很明显,小目标检测任务中负样本的数量更多,正样本

的数量更少，存在正负样本数量不均衡的问题。

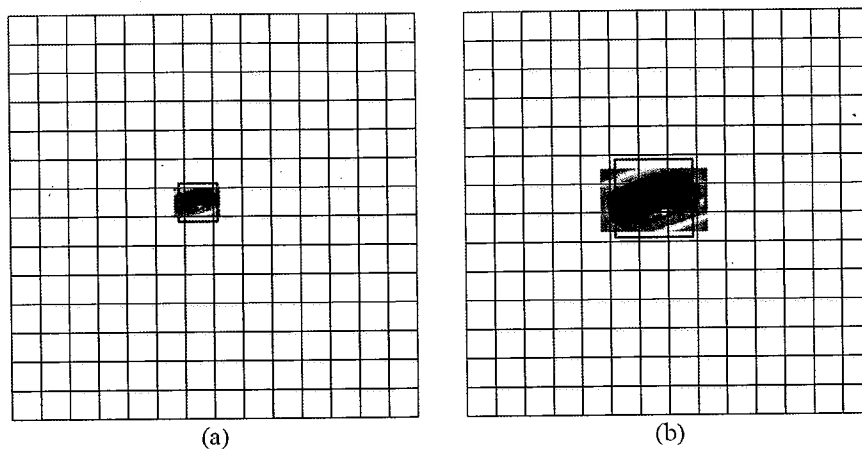


图 2.3 不同大小目标在相同特征图上对比

Figure 2.3 Comparison of targets with different sizes on the same feature map

以 VEDAI 数据集^[117]为例，平均每张图片有 5.5 个目标。并且，目标的大小主要分布在 10×10 像素到 60×60 像素之间，统计结果显示超过 99.5% 的初始锚框都属于负样本。因此，缓解正负样本数量的不平衡，将可有效提高卷积神经网络对小目标的检测性能。

3. 小目标周围的上下文信息：目标周围的上下文信息是影响目标检测的一个关键因素，对小目标检测尤其重要，因为小目标通常只占有图像中的小部分区域。由于小目标区域的特征信息不明显，可区分性不大，仅仅通过小目标区域很难完成小目标的识别任务。如图 2.4 所示，仅仅通过图(a)中的信息来识别船只有一定困难，但是通过增加船只周围的上下文信息，如天空和海面，将会很容易的识别图像中的船只。

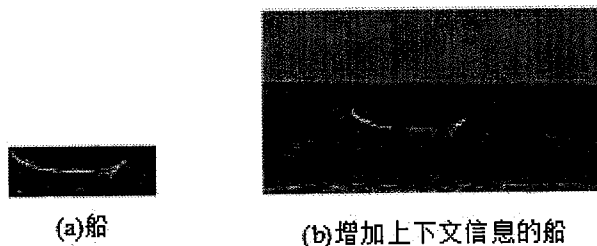


图 2.4 未增加上下文信息和增加上下文信息船图像的对比

Figure 2.4 The Comparison of ship images without and with context information

Hu 等人^[74]通过实验证明了目标周围的上下文信息对人脸检测十分重要，特别是对占有像素特别少，特征较为模糊的小人脸，有助于它们的定位和分类。文中指出扩大感受野是增加目标周围上下文信息的有效途径。如图 2.5 所示，通过融合具有更多感受野的特征，可以增加人脸周围的上下文信息，提高对人脸的检测性能，当感受野从 27×27 逐渐增加到 291×291 时，相应人脸检测的 mAP 值从 49.6% 增加到 68.5%。

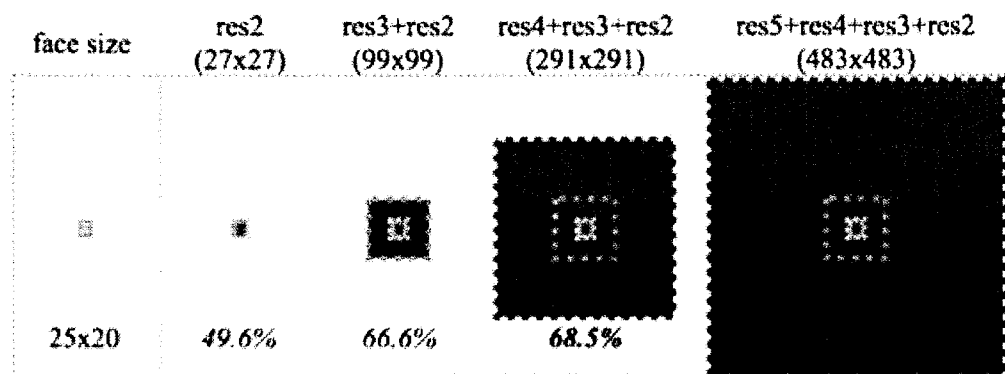


图 2.5 增大感受野对人脸检测的影响

Figure 2.5 The impact of expanding receptive field on tiny face detection

因此，通过增加感受野来增加小目标周围的上下文信息，可以有效的提高卷积神经网络对小目标的检测性能。

4. 尺度选择：对于卷积神经网络，大特征图(浅层特征图)适合小目标检测，小特征图(深层特征图)适合大目标检测。如图 2.6 所示，红框代表锚框，蓝框代表目标框，特征图中包含目标中心的网格负责该目标的回归。很明显，浅层特征图中的锚框中心更接近于目标中心，更适合于目标位置回归。然而，浅层的特征图感受野小，包含目标的语义信息少，不适合于目标的识别。另外，在相同的位置预测差距下，小目标的预测框与标签的交并比要小于大目标预测框与标签的交并比，小目标对位置的回归要求更高。因此，计算目标大小和网络尺度之间具体的数值关系，并为小目标检测选择合适的检测尺度至关重要。

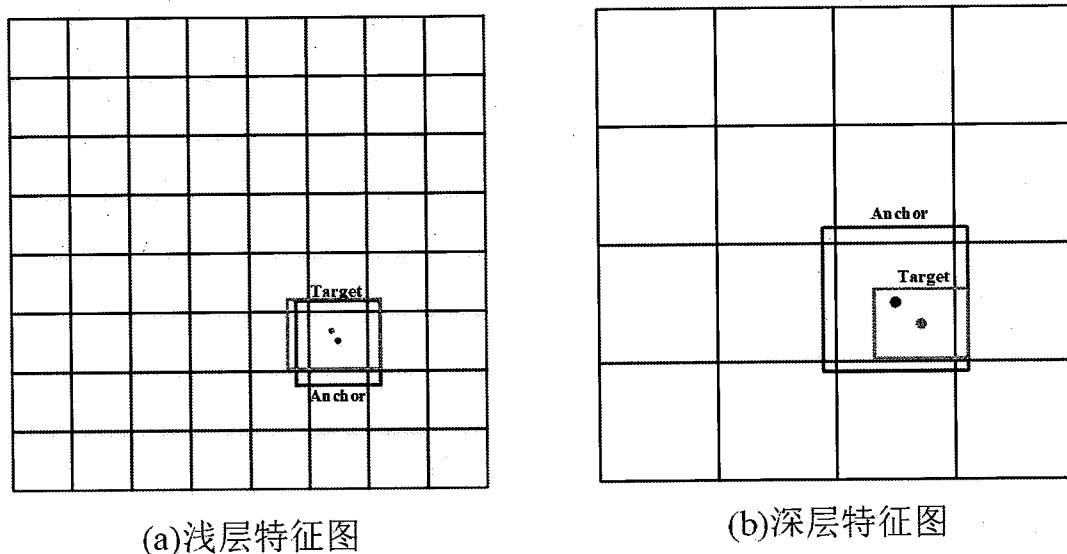


图 2.6 目标在不同特征图上的位置回归

Figure 2.6 Position regression of the target on different feature maps

5. 特征融合：在卷积神经网络中，浅层特征包含较多细粒度信息，有助于目标的定位。深层特征包含较多的语义信息，有助于目标的分类。将 VGG-19 不同层的特征图进行可视化，结果如图 2.7 所示，浅层特征图包含更多的目标纹理、轮廓等细节特征信息，这些信息有助于网络对小目标的准确定位。深层特征包含了更多的目标语义信息，更具有代表性，可用于去除环境和噪声等外界因素的干扰。因此，将浅层特征与深层特征进行特征融合，可以有效提高卷积神经网络对小目标的检测性能^[118]。

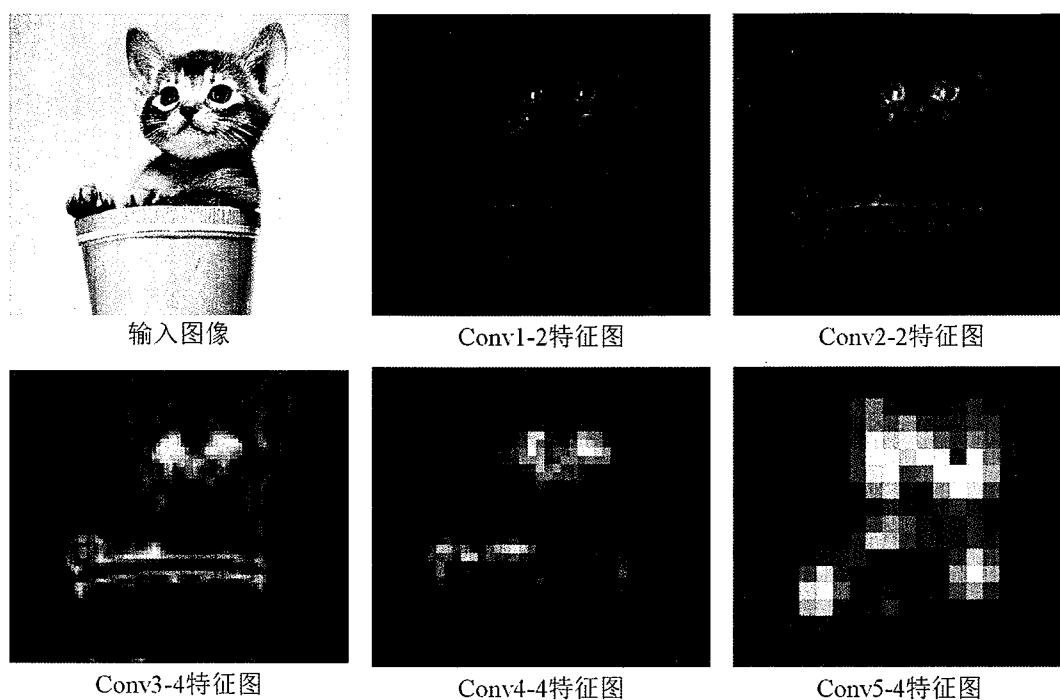


图 2.7 VGG-19 不同层特征图的可视化

Figure 2.7 Visualization of feature maps at different layers of VGG-19

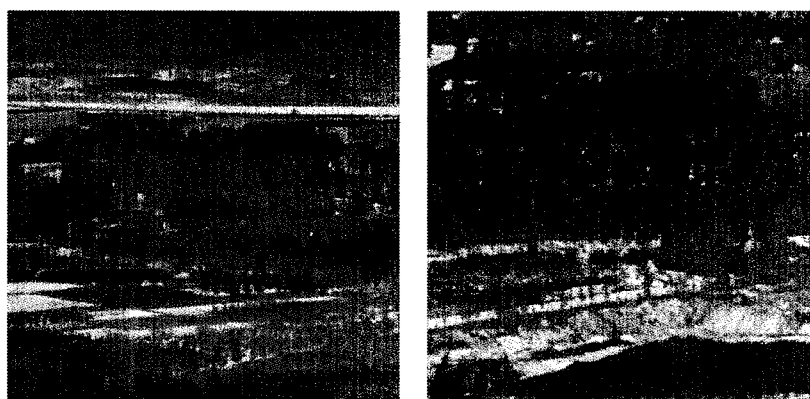


图 2.8 红外弱小目标

Figure 2.8 Infrared dim and small target

6. 目标信号弱，背景掺杂噪声：这个影响因素主要是针对红外弱小目标。如图 2.8 所示，相比于对地小目标，红外弱小目标几乎不含有亮度信息之外的其它特征信息，缺乏细致的纹理结构和丰富的颜色信息，通常以“亮点”的形式呈现。信号弱主要是因为红外弱小目标的成像距离较远，使红外探测器接收到的目标辐

射能量较低,与背景的对比度低。因此,抑制背景噪声,增强红外弱小目标的响应可有效提高卷积神经网络对红外弱小目标的检测性能。

上述6项是影响小目标检测算法性能的关键因素。本文根据这些因素提出相应的解决策略,并结合这些策略来设计对地小目标和对空红外弱小目标检测算法,整体的设计架构如图2.9所示。

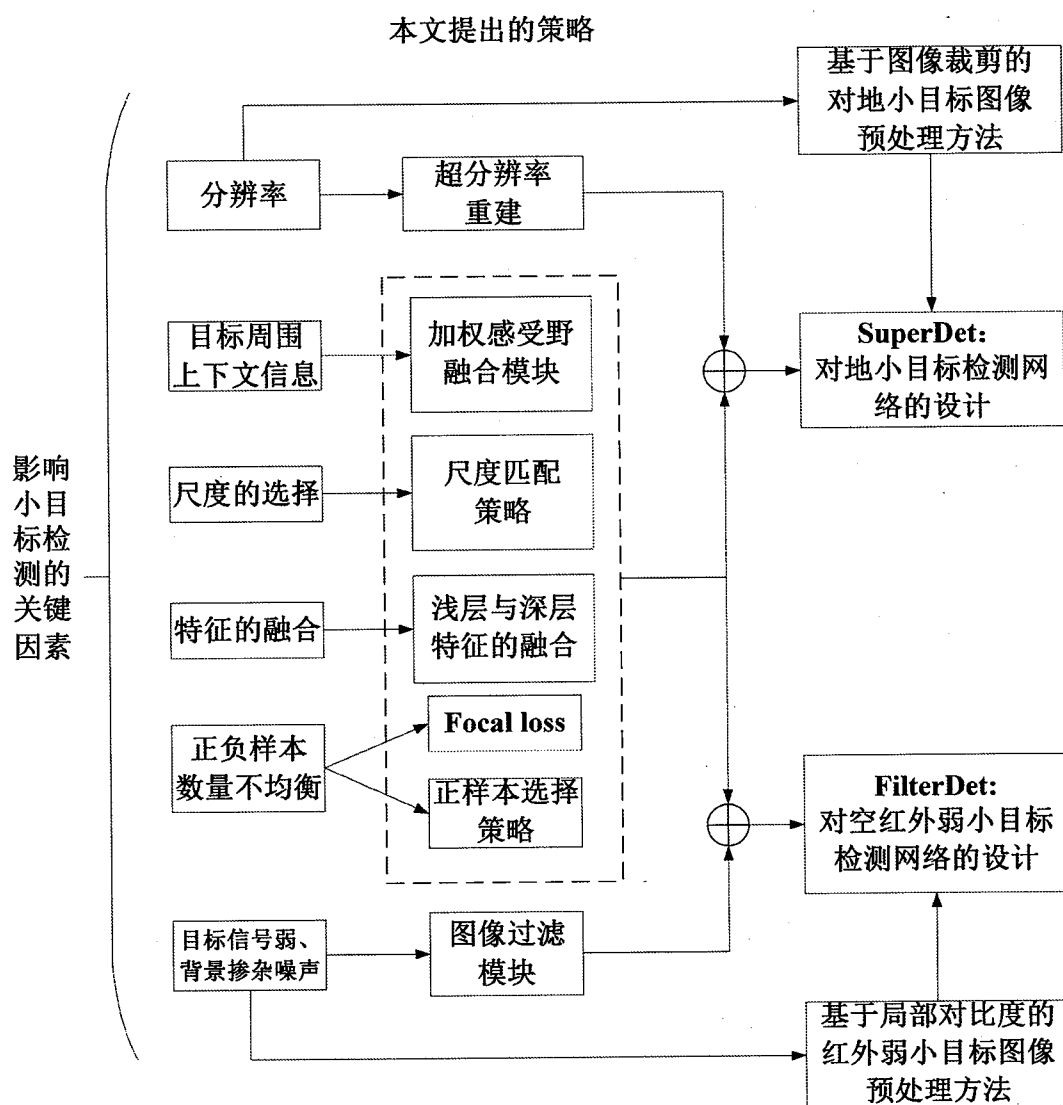


图 2.9 全文设计的框架图

Figure 2.9 The architecture design of this dissertation

2.3 本章小结

本章根据对地小目标和对空红外弱小目标在图像中占有像素少、特征不明显,

易受周围环境影响等特点，总结了影响小目标检测的关键因素，包括：分辨率、目标周围的上下文信息、尺度选择、特征融合、正负样本数量不均衡和目标信号弱、背景掺杂噪声。根据这 6 项影响小目标检测的关键因素，提出相应的解决方案，为对地小目标检测算法和对空红外弱小目标检测算法的设计奠定基础。最后，结合这些因素给出了本文的设计框架。

第3章 面向小目标检测的图像预处理方法

3.1 引言

正如第2章所分析的,利用不同分辨率的目标来训练网络,有助于提高检测网络的鲁棒性。然而,在对目标进行多尺度变换时,会增大图像的尺寸,需要占用更多的计算机显存。在对地小目标检测任务中,输入图像的大部分区域都是背景,只有少部分区域是目标。根据此特点,本章首先提出了一种基于图像裁剪的对地小目标图像预处理方法。利用该方法,可以在不增加显存的条件下,实现网络对同一目标多尺度的训练,使网络变得更加鲁棒。另外,针对红外弱小目标信号弱,背景掺杂噪声的特点,本章提出了一种基于局部对比度的红外弱小目标图像预处理方法。该方法可以有效的增强红外弱小目标的响应,提高目标与背景之间的对比度。

3.2 基于图像裁剪的对地小目标图像预处理方法

3.2.1 多尺度训练

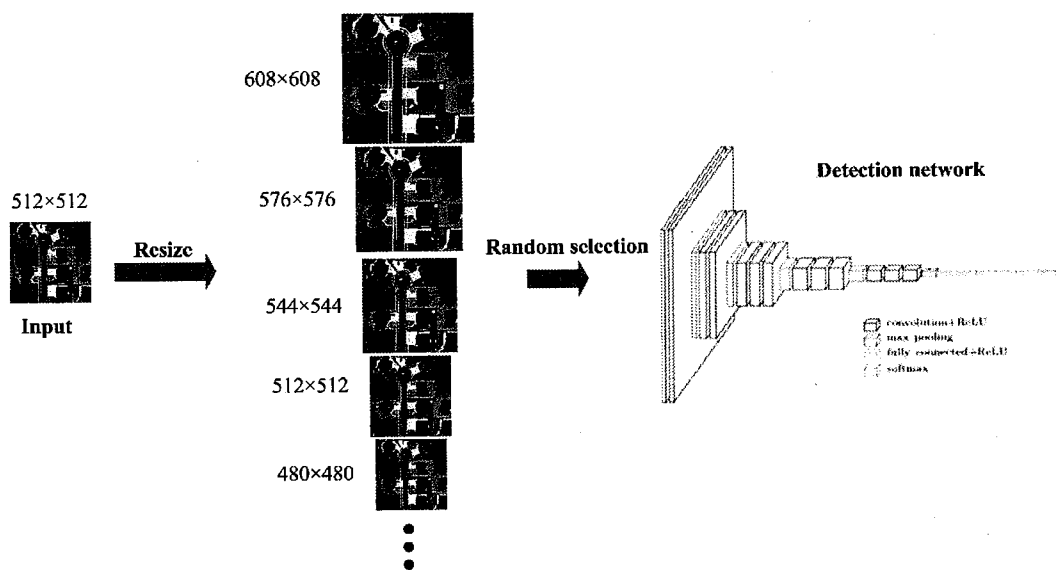


图 3.1 多尺度训练

Figure 3.1 Multi-scale training

Redmon 等人^[59]提出将图片进行多尺度变换，再输入到检测网络进行训练。这种多尺度训练(multi-scale training)策略使网络能够学习同一个目标，不同尺度的特征和信息，有助于提高检测网络的鲁棒性。

多尺度训练的具体过程如图 3.1 所示，以分辨率为 512×512 的输入图像为例，在训练过程中，输入图像的尺寸会被随机地调整，使网络可以接受同一目标多个尺度的训练。虽然这种策略有助于提高检测网络的鲁棒性，但是当输入图像的尺寸被调整到大于 512×512 时，不得不消耗更多的显存资源。

3.2.2 图像裁剪的原理和流程

在对地小目标检测任务中，图像中大部分的区域都是背景，只有少部分的区域属于小目标。如果想要将对地小目标放在一个更宽的尺度范围进行训练，不得不将输入图像的尺寸调整成更宽的范围，这将消耗更多的显存。针对这个问题，本节提出了一种基于图像裁剪的对地小目标图像预处理方法。该方法可以裁剪掉图像中多余的背景，使对地小目标可以在更宽的尺度范围进行训练，并且不需要消耗多余的显存资源。

假设输入图像的宽和高分别是 W 和 H 。首先，根据数据集中目标的边框标签(label)，找到图像中包含所有对地小目标的最小区域。假设最小区域的左上角坐标是 $(x_{\min}, y_{\min})$ ，右下角的坐标是 $(x_{\max}, y_{\max})$ 。则最小区域的长和宽可以分别表示为 $W_{\text{region}} = x_{\max} - x_{\min}$ 和 $H_{\text{region}} = y_{\max} - y_{\min}$ 。需要考虑一种特殊的情况，当图像中的目标彼此之间都离得很近，将会造成裁剪之后的区域太小，丢失过多的上下文信息，不利于对对地小目标的检测。为了防止出现该问题，规定裁剪之后的最小区域要大于 $W/2 \times H/2$ 。因此，将最小区域的宽 W_{region} 与 $W/2$ 进行对比，将最小区域的高 H_{region} 与 $H/2$ 进行对比，将会得到以下四种情况，如图 3.2 所示，其中，蓝色框表示最后的裁剪图像。

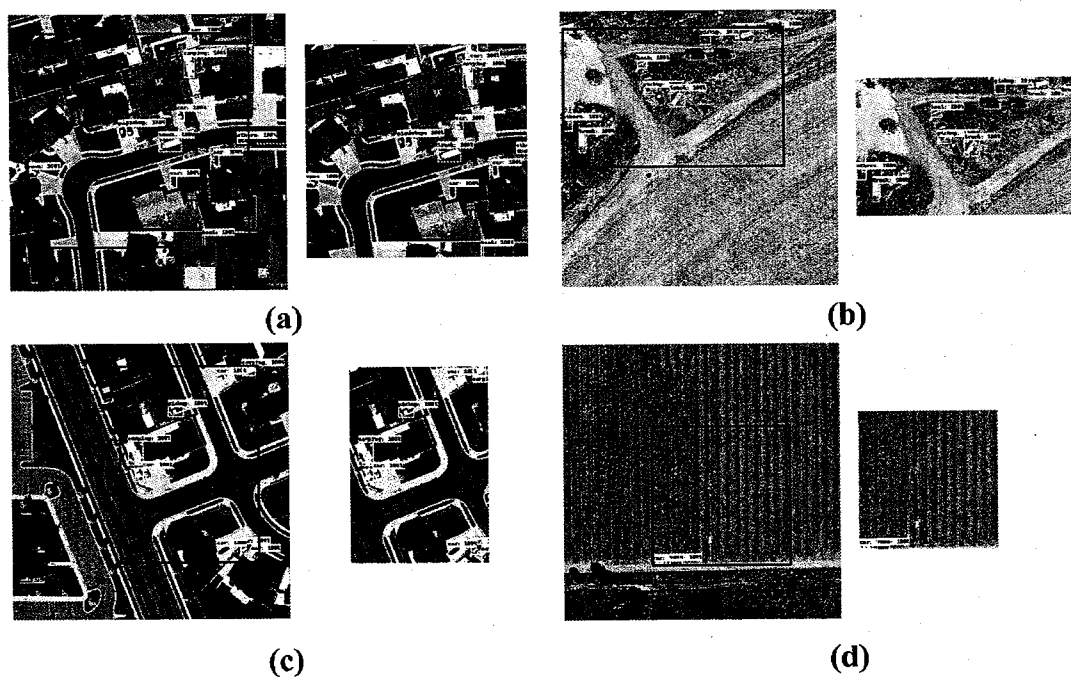


图 3.2 图像裁剪过程中的 4 种情况

Figure 3.2 The 4 cases for image cropping method

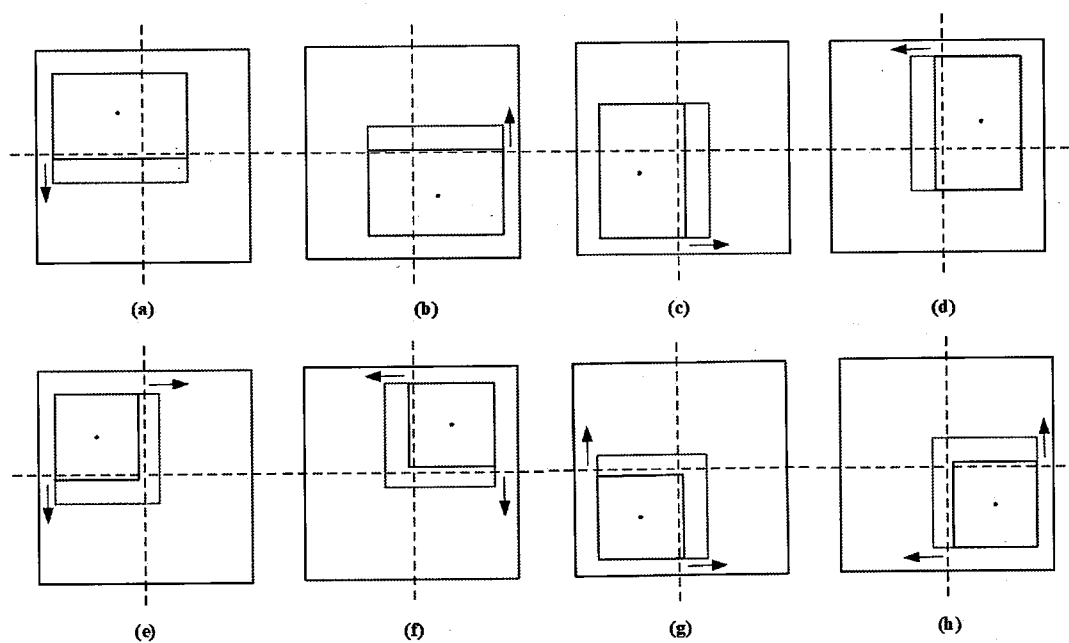


图 3.3 最小区域的延长方式

Figure 3.3 The extension of the smallest area

情况 1: $W_{region} \geq W/2, H_{region} \geq H/2$

如图 3.2(a)所示, 在这种情形下, 包含所有目标的最小区域的宽和高分别大于 $W/2$ 和 $H/2$ 。因此, 将区域 $W_{region} \times H_{region}$ 从原图像上裁剪下来, 作为最后的训练图像。

情况 2: $W_{region} \geq W/2, H_{region} < H/2$

如图 3.2(b)所示, 在这种情形下, 包含所有目标的最小区域的宽大于等于 $W/2$, 而高却小于 $H/2$ 。当 $\frac{y_{max} - y_{min}}{2} \leq H/2$ 时, 则以最小区域左上角点作为基准点, 将最小区域的高向下延长至 $H/2$, 如图 3.3(a)所示, 红色框为延长之后的最小区域, 红色框下的黑色框是原最小区域。当 $\frac{y_{max} - y_{min}}{2} > H/2$ 时, 则以最小区域右下角点作为基准点, 将最小区域的高向上延长至 $H/2$, 如图 3.3(b)所示。将区域 $W_{region} \times H/2$ 从原图像中裁剪下来, 作为最后的训练图像。

情况 3: $W_{region} < W/2, H_{region} \geq H/2$

如图 3.2(c)所示, 在这种情形下, 包含所有目标的最小区域的高大于等于 $H/2$, 而宽却小于 $W/2$ 。当 $\frac{x_{max} - x_{min}}{2} \leq W/2$ 时, 则以最小区域左上角点作为基准点, 将最小区域的宽向右延长至 $W/2$, 如图 3.3(c)所示。当 $\frac{x_{max} - x_{min}}{2} > W/2$ 时, 则以最小区域右下角点作为基准点, 将最小区域的宽向左延长至 $W/2$, 如图 3.3(d)所示。将区域 $W/2 \times H_{region}$ 从原图像中裁剪下来, 作为最后的训练图像。

情况 4: $W_{region} < W/2, H_{region} < H/2$

如图 3.2(d)所示, 在这种情形下, 包含所有目标的最小区域的高小于 $H/2$, 宽也小于 $W/2$ 。当 $\frac{x_{max} - x_{min}}{2} \leq W/2, \frac{y_{max} - y_{min}}{2} \leq H/2$ 时, 则以最小区域左上角点作为基准点, 将最小区域的宽向右延长至 $W/2$, 将最小区域的高向下延长

至 $H/2$ ，如图 3.3(e)所示。当 $\frac{x_{\max} - x_{\min}}{2} > W/2$, $\frac{y_{\max} - y_{\min}}{2} \leq H/2$ 时，则以最小区域右上角点作为基准点，将最小区域的宽向左延长至 $W/2$ ，将最小区域的高向下延长至 $H/2$ ，如图 3.3(f)所示。当 $\frac{x_{\max} - x_{\min}}{2} \leq W/2$, $\frac{y_{\max} - y_{\min}}{2} > H/2$ 时，则以最小区域左下角点作为基准点，将最小区域的宽向右延长至 $W/2$ ，将最小区域的高向上延长至 $H/2$ ，如图 3.3(g)所示。当 $\frac{x_{\max} - x_{\min}}{2} > W/2$, $\frac{y_{\max} - y_{\min}}{2} > H/2$ 时，则以最小区域右下角点作为基准点，将最小区域的宽向左延长至 $W/2$ ，将最小区域的高向上延长至 $H/2$ ，如图 3.3(h)所示。将区域 $W/2 \times H/2$ 从原图像中裁剪下来，作为最后的训练图像。

算法如表 3.1 所示。

表 3.1 算法的伪代码

Table 3.1 Pseudo code of the proposed method

Algorithm: Image cropping method in data preprocessing.

Input: T as bounding box ground truth, (x_{top}, y_{top}) as the left top coordinate of T , (x_{bottom}, y_{bottom}) as the right bottom coordinate of T . H and W as the height and the width of the input image.

Output: The smallest region containing all the targets.

1. $x_{min} = \min(x_{top})$, $x_{max} = \max(x_{bottom})$, $y_{min} = \min(y_{top})$, $y_{max} = \max(y_{bottom})$
 2. $W_{region} = x_{max} - x_{min}$, $H_{region} = y_{max} - y_{min}$
 3. if $W_{region} \geq W/2$, $H_{region} \geq H/2$ then
 4. Crop $W_{region} \times H_{region}$, as shown in Figure 3.2 (a).
 5. else if $W_{region} \geq W/2$, $H_{region} < H/2$ then
 6. Crop $W_{region} \times H/2$, as shown in Figure 3.2 (b).
 7. else if $W_{region} < W/2$, $H_{region} \geq H/2$ then
 8. Crop $W/2 \times H_{region}$, as shown in Figure 3.2 (c).
 9. else if $W_{region} < W/2$, $H_{region} < H/2$ then
 10. Crop $W/2 \times H/2$, as shown in Figure 3.2 (d).
 11. end
-

在裁剪完数据集中的图像后，大部分图像的背景被裁剪。因此，大多数的背景（负样本）没有参加训练，可能会使虚警的数量增多。为了缓解这个问题，将裁剪的图像和原图像混合在一起作为最终的训练集，这样也可以保证原尺度的目标包含在训练集中。在训练的过程中所有输入的图像都将被调整成相同的尺寸，与原数据集中的图像进行对比，裁剪后图像中的对地小目标可以在更宽的尺度范围内进行训练，并且不需要消耗多余的显存。

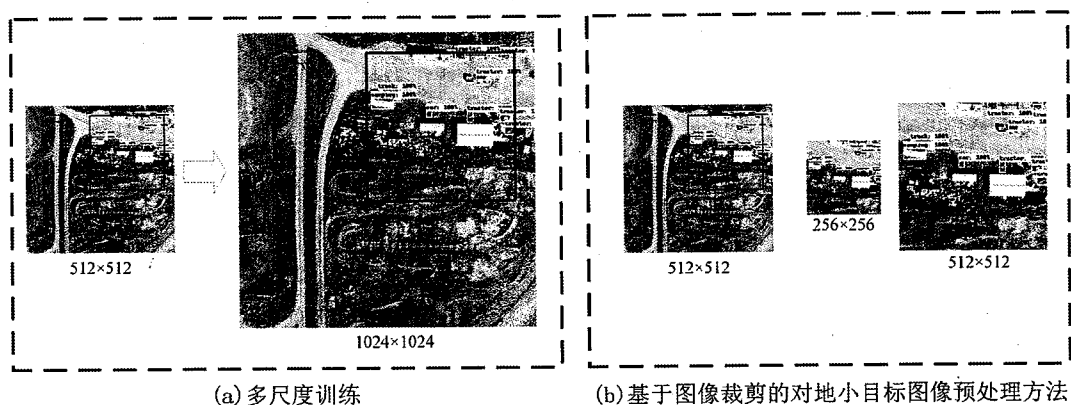


图 3.4 多尺度训练与基于图像裁剪的对地小目标图像预处理方法的对比

Figure 3.4 Comparison between multi-scale training and ground small target image cropping method-based data preprocessing

如图 3.4 所示,对多尺度训练和基于图像裁剪的对地小目标图像预处理方法进行对比。基于多尺度训练方法,如果要将目标的大小放大 1 倍,不得不将图像的尺寸扩大到 1024×1024 ,需要消耗更多的显存。而基于图像裁剪的对地小目标图像预处理方法,在训练的过程中,裁剪后的图像被调整为 512×512 ,目标的大小放大了 1 倍,但不需要消耗多余的显存。

基于图像裁剪的对地小目标图像预处理方法可以通过两种方式来实现。第一种方式是在训练之前就将数据集进行裁剪处理,然后将原始图像和裁剪图像进行混合。第二种方式是将此预处理方法在训练阶段的数据增广过程中进行实现。

3.2.3 实验与分析

为了证明基于图像裁剪的对地小目标图像预处理方法的效果,在 VEDAI 数据集^[117]上,分别利用 YOLO V3 和 GSTDet 进行对比实验,实验结果如下列各表。其中,表 3.2 和表 3.3 分别是 YOLO V3 和 GSTDet 在 VEDAI 数据集上的检测结果。表中*号表示结合了基于图像裁剪的对地小目标图像预处理方法的 YOLO V3 和 GSTDet。

表 3.2 YOLO V3 在 VEDAI 数据集上的对比结果: AP (%)

Table 3.2 Performance Comparisons for YOLO V3 on VEDAI dataset: AP (%)

method	Car	Pickup	Camping	Truck	Boat	Van	Other	Tractor	Plane	mAP
YOLO V3	75.4	74.2	66.2	47.3	58.3	44.7	23.6	66.7	100	61.8
YOLO V3*	79.6	70.9	69.7	46.7	52.3	56.2	44.1	65.4	100	65.0

表 3.3 GSTDet 在 VEDAI 数据集上的对比结果: AP (%)

Table 3.3 Performance Comparisons for GSTDet on VEDAI dataset: AP (%)

method	Car	Pickup	Camping	Truck	Boat	Van	Other	Tractor	Plane	mAP
GSTDet	86.9	80.2	82.9	57.8	75.7	71.4	39.2	73.3	100	74.1
GSTDet*	85.8	83.0	84.4	56.8	75.9	74.4	43.2	82.1	100	76.2

由表 3.2、表 3.3 可知, 当 YOLO V3 和 GSTDet 与基于图像裁剪的对地小目标图像预处理方法相结合时, 都有助于提高检测算法对对地小目标的检测性能。由于此预处理方法应用在训练阶段, 所以并不影响各检测算法对目标的检测速度。实验结果证明了基于图像裁剪的对地小目标图像预处理方法可以应用在不同的检测算法上, 实现了目标多尺度训练, 且不需要多余的显存资源, 可以有效提高检测算法对对地小目标的检测能力。

定性对比结果: 在 VEDAI 数据集上, 分别对 YOLO V3 和 YOLO V3*、GSTDet 和 GSTDet*进行了可视化的对比, 对比结果如图 3.5 和图 3.6。图 3.5 和图 3.6 分别为 YOLO V3 和 YOLO V3*、GSTDet 和 GSTDet*在 VEDAI 数据集上的对比结果。如图 3.5、图 3.6 所示, 相比于 GSTDet*和 YOLO V3*, GSTDet 和 YOLO V3 存在漏检和错检的情况。

通过 YOLO V3 和 GSTDet 在 VEDAI 数据集上的定性和定量对比实验, 证明了基于图像裁剪的对地小目标图像预处理方法适用于不同的目标检测算法。

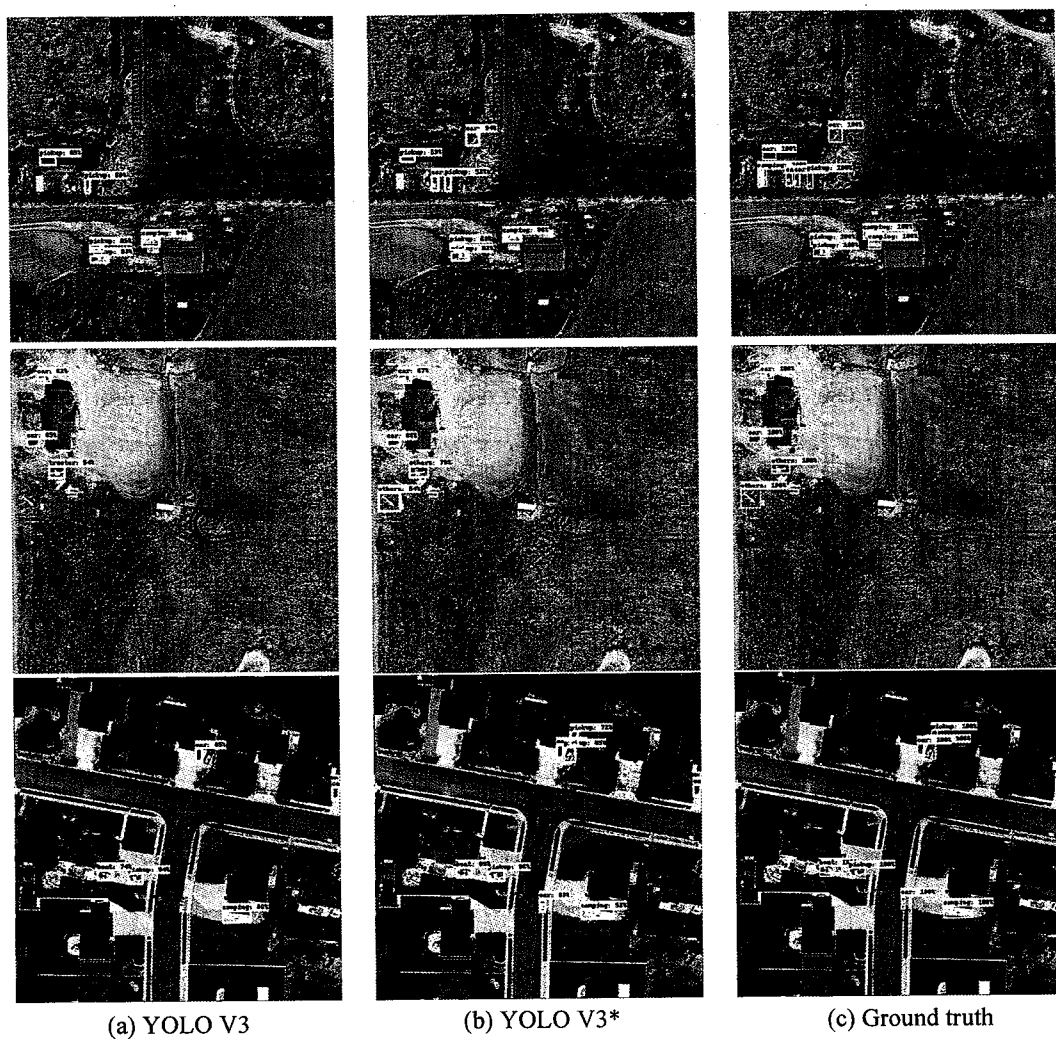


图 3.5 YOLO V3 和 YOLO V3*在 VEDAI 数据集上可视化结果

Figure 3.5 The visual results of YOLO V3 and YOLO V3* in VEDAI dataset



图 3.6 GSTDet 和 GSTDet*在 VEDAI 数据集可视化结果

Figure 3.6 The visual results of GSTDet and GSTDet* in VEDAI dataset

3.3 基于局部对比度的红外弱小目标图像预处理方法

本节主要是对基于局部对比度测量的红外弱小目标检测算法进行改进,并将改进的局部对比度测量方法作为图像预处理应用在对空红外弱小目标检测网络中。改进的局部对比度测量方法可以通过卷积和最大池化操作来实现,算法的处理效率得到了提高。

3.3.1 局部对比度测量方法

受人类视觉系统的启发, Chen 等人^[92]提出了基于局部对比度测量(Local contrast measure, LCM)的红外弱小目标检测算法。该算法主要是利用红外弱小目标与背景之间的对比度差异来实现对红外弱小目标的检测。如图 3.7 所示, I 代表整张图像, T 代表目标区域, 图像块 B 与目标 T 之间的绿色区域代表局部背景区域, 图像块 B 可以在图像 I 上滑动。

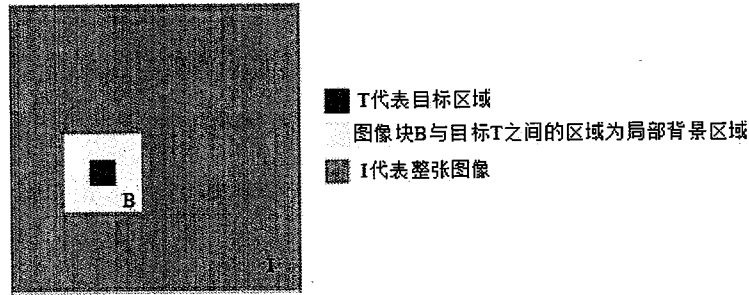


图 3.7 对比度机制

Figure 3.7 Local contrast mechanism

将图像块分成 9 个子块, 其中 1~8 代表局部背景区域, 0 代表目标区域, 每个子块包含 3×3 个像素, 如图 3.8 所示。

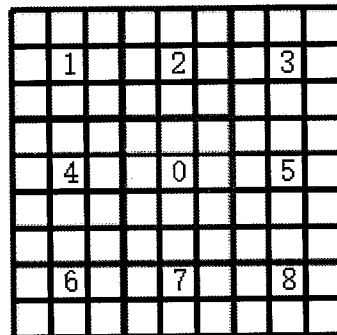


图 3.8 局部对比度测量中的图像块

Figure 3.8 Image patch for local contrast measure

局部对比度测量算法是通过计算当前目标位置与局部背景区域之间的对比度来获取对比度图。局部背景区域中第 i 个子块的平均灰度值 $m_i (i=1 \sim 8)$ 可以表

示为:

$$m_i = \frac{1}{9} \sum_{j=1}^9 I_j \quad (3.1)$$

其中, I_j 表示第 i 个子块中的第 j 个像素的灰度值。

则目标区域与第 i 个局部背景区域子块之间的对比度定义为:

$$c_i = \frac{L}{m_i} \quad (3.2)$$

其中, L 代表目标区域中最大的灰度值。

另外, 在红外图像中, 红外弱小目标的亮度一般要大于周围环境的亮度。然而, 有时这种差异并不是十分明显, 为了进一步增加目标的响应, 局部对比度测量值定义如下:

$$C = \min L \times c_i^n = \min \frac{L^2}{m_i} = \frac{L^2}{\max m_i} \quad (3.3)$$

这意味着局部对比度值 C 越大, 红外弱小目标越可能出现。

3.3.2 局部对比度测量方法的改进与快速实现

原局部对比度测量算法在计算局部对比度值时, 采用的是目标区域最大的灰度值来计算目标灰度, 当目标区域存在幅度较大的脉冲噪声时, 容易造成虚警和错检。另外, 原局部对比度测量算法采用局部背景子块中最大的平均灰度值来计算局部背景的灰度值, 没有充分利用局部背景子块的整体信息, 处理后的红外弱小目标会被深灰色的像素所包围。本节对局部对比度算法进行改进, 分别采用目标区域的平均灰度值和局部背景子块的平均灰度值代替目标区域最大的灰度值和局部背景子块中最大的平均灰度值来计算局部对比度测量值。

首先, 计算目标区域的平均灰度值, 如下式:

$$n = \frac{1}{9} \sum_{i=1}^9 I_i \quad (3.4)$$

其中, I_i 表示目标区域中第 i 个像素的灰度值。

然后, 计算局部背景区域所有子块的平均灰度值。

$$b = \frac{1}{72} \sum_{j=1}^{72} I_j \quad (3.5)$$

其中, I_j 表示局部背景区域中第 j 个像素的灰度值。

定义目标区域与局部背景区域子块之间的对比度 c 为

$$c = \frac{n}{b} \quad (3.6)$$

则改进的局部对比度测量值定义为

$$C = c \times L = \frac{n \times L}{b} \quad (3.7)$$

其中, L 代表目标区域中最大的灰度值。

通过采用目标区域的平均灰度值来代替最大灰度值可以防止由于目标区域存在幅度较大的脉冲噪声造成的错检和虚警, 采用局部背景子块的平均灰度值代替最大灰度值可以充分利用弱小目标周围的局部背景信息。原局部对比度测量方法和改进的局部对比度测量方法的对比结果如图 3.9 所示, 原局部对比度测量方法和改进的局部对比度测量方法都可以提高目标的强度, 但是原对比度测量方法处理后的红外弱小目标会被深灰色的像素所包围, 这是因为在原对比度测量方法中, 局部背景只是单纯的利用 8 个子块中的最大值来表示。如图 3.10 所示, 当计算目标附近背景的对比度值时, 背景会出现在中心 0 子块内, 即 L 值会很小, 目标会出现在局部背景 1~8 子块中, 即 $\max m_i$ 会很大, 因此求得的对比度 C 值会很小。在改进的局部对比度测量方法中, 用 8 个子块的平均值来表示局部背景, 可以有效的去除对比度图中红外弱小目标周围的深灰色的像素。

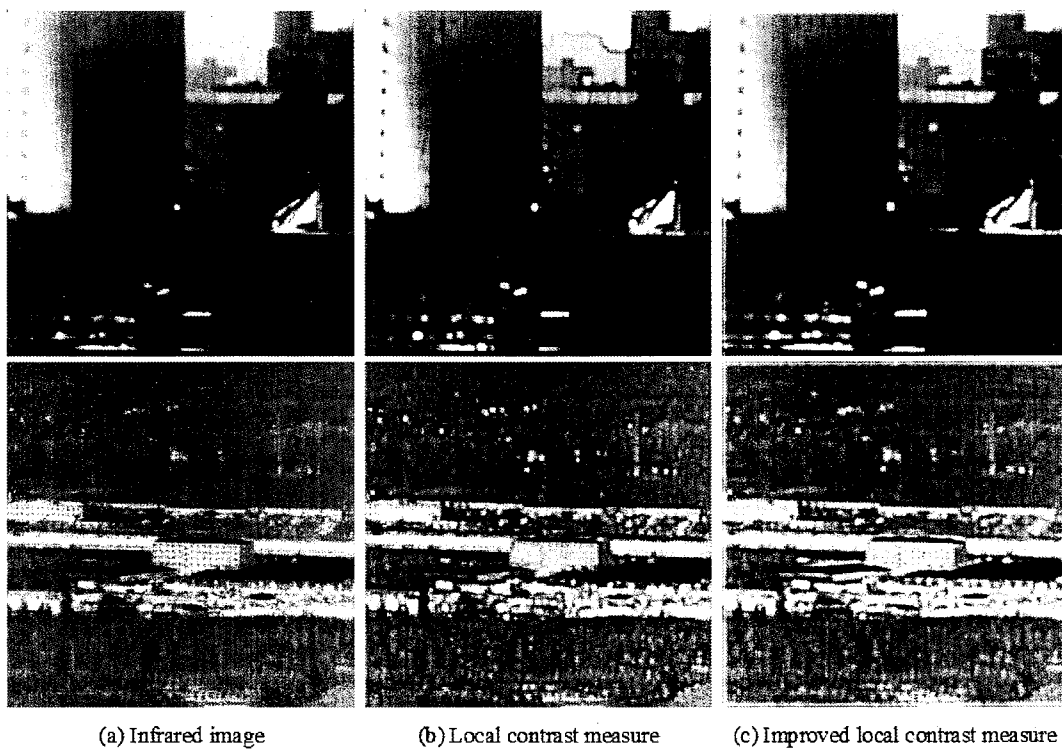


图 3.9 局部对比度测量方法与改进的局部对比度测量方法的结果对比

Figure 3.9 Comparison between local contrast measure and proposed local contrast measure

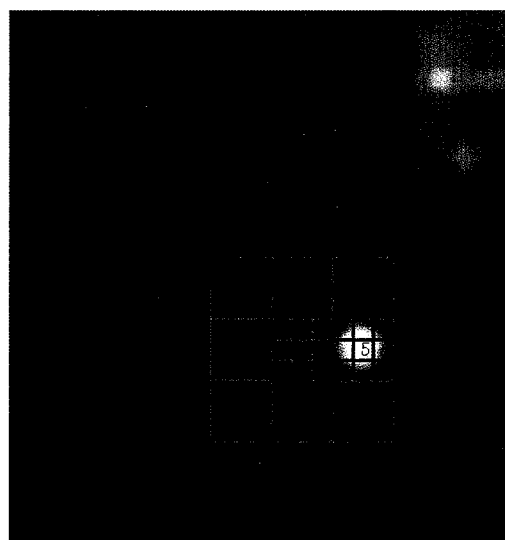


图 3.10 红外弱小目标周围背景的局部对比度的计算

Figure 3.10 Calculation of local contrast of background around infrared small and dim target

另外,与原局部对比度测量方法相比,改进的局部对比度测量方法可以由卷积和最大池化操作来实现,如图 3.11 所示。利用一个 3×3 的卷积来计算目标区域的平均灰度值,卷积的参数如下:

$$\begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix} \quad (3.8)$$

利用一个 9×9 的卷积来计算局部背景的平均灰度值,卷积的参数如下:

$$\begin{bmatrix} 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 \\ 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 \\ 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 \\ 1/72 & 1/72 & 1/72 & 0 & 0 & 0 & 1/72 & 1/72 & 1/72 \\ 1/72 & 1/72 & 1/72 & 0 & 0 & 0 & 1/72 & 1/72 & 1/72 \\ 1/72 & 1/72 & 1/72 & 0 & 0 & 0 & 1/72 & 1/72 & 1/72 \\ 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 \\ 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 \\ 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 & 1/72 \end{bmatrix} \quad (3.9)$$

最后,利用一个步长为 1 的 3×3 的最大池化来计算目标区域的最大灰度值。

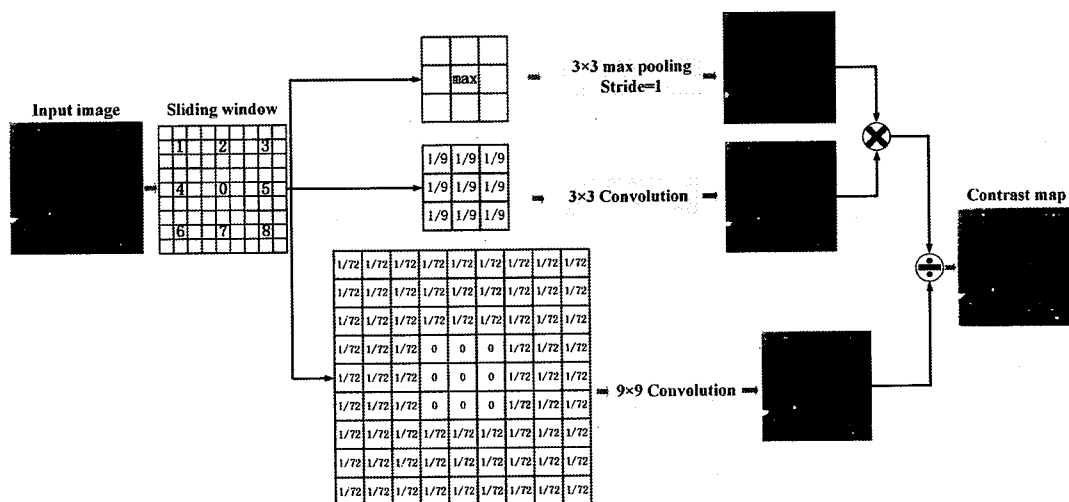


图 3.11 改进的局部对比度测量方法的实现

Figure 3.11 The implementation of improved local contrast measure

与原局部对比度测量方法相比,利用卷积和最大池化操作实现的改进局部对比度测量方法对红外图像处理速度更快,效率更高。改进的局部对比度测量方法

的流程如表 3.4 所示。

表 3.4 改进的局部对比度测量方法的伪代码

Table 3.4 Pseudo code of the improved local contrast measure

Algorithm: Proposed local contrast measure.

Input: The input infrared image.

Output: The contrast map and the saliency map of the input image.

1. Compute the gray mean n of the central red box in the sliding image patch by a 3×3 convolution.

$$n = \frac{1}{9} \sum_{i=1}^9 I_i$$

2. Compute the gray mean b of the local background in the sliding image patch by a 9×9 convolution.

$$b = \frac{1}{72} \sum_{j=1}^{72} I_j$$

3. Compute the maximum L of the gray value of the central red box in the sliding image patch by a 3×3 max-pooling with stride=1.

4. The contrast map can be generated by

$$C = \frac{n \times L}{b}$$

3.3.3 实验与分析

为了证明基于局部对比度的红外弱小目标图像预处理方法的有效性, 分别在 3 个红外弱小目标序列^[119]和合成的单帧红外弱小目标数据集上, 以原红外图像、经原局部对比度测量算法处理的红外图像和经改进的局部对比度测量算法处理的红外图像作为输入, 利用 FilterDet 进行对比实验, 实验结果如表 3.5 所示。相比于原图像, 基于局部对比度测量的红外弱小目标图像预处理方法可以通过增强红外弱小目标的强度来提高网络对红外弱小目标的检测性能。相比于原局部对比度测量方法, 改进的局部对比度测量方法采用目标区域的平均灰度值来代替最大灰度值, 可以缓解由于目标区域存在幅度较大的脉冲噪声造成的错检和虚警, 采用局部背景子块的平均灰度值代替最大灰度值可以充分利用弱小目标周围的局

部背景信息，使检测性能进一步提高。

表 3.5 不同输入的实验结果对比

Table 3.5 Comparison of experimental results with different input

输入	原红外图像	原局部对比度测量方法	改进的局部对比度测量方法
	AP(%)	AP(%)	法 AP(%)
Single frame image set	89.79	89.78	91.64
Sequence1	98.09	98.58	98.79
Sequence2	86.90	87.92	90.27
Sequence3	95.38	96.80	97.80

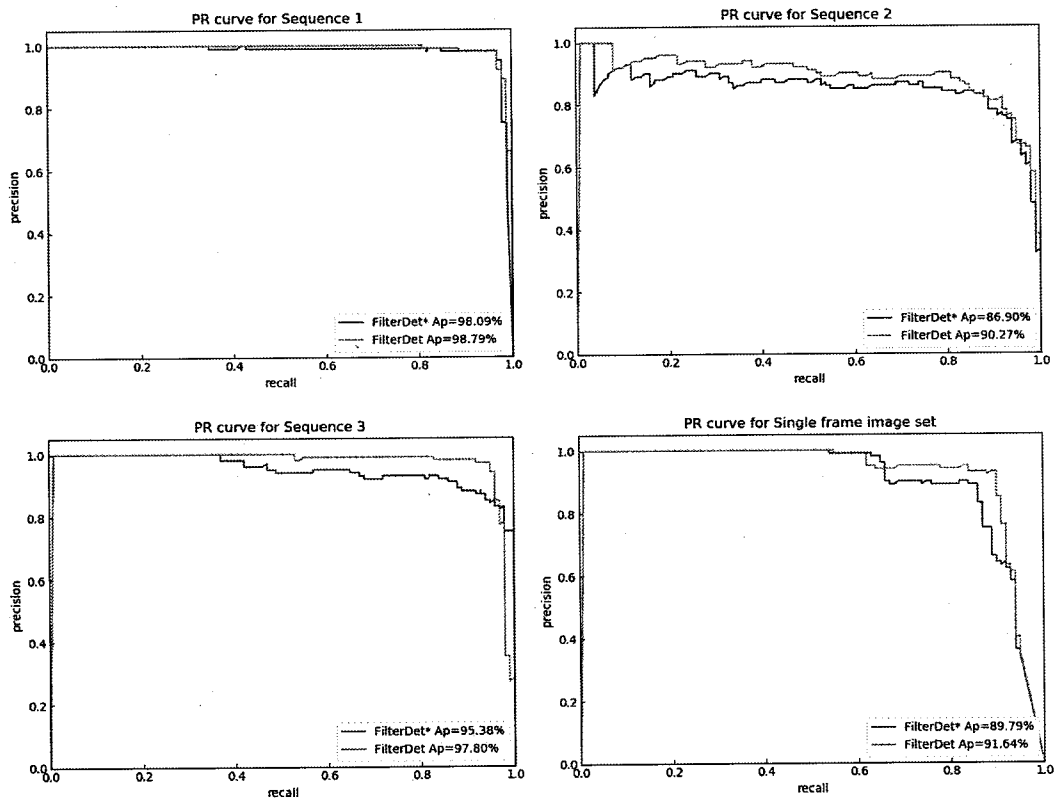


图 3.12 PR 曲线对比

Figure 3.12 The comparison of PR curve

另外，以原红外图像作为输入的 FilterDet 和以经过改进的局部对比度测量算

法处理的红外图像作为输入的 FilterDet 的 PR 曲线对比如图 3.13 所示。其中, FilterDet*代表网络的输入为未经基于局部对比度测量方法处理的红外图像, FilterDet 代表网络的输入为经过基于局部对比度测量方法处理的红外图像。结果表明,基于局部对比度的红外弱小目标图像预处理方法可以有效的提高网络对红外弱小目标检测能力。

3.4 本章小结

在对地小目标检测任务中,图像中大部分的区域都是背景,只有少部分的区域属于小目标。受多尺度训练的启发,本章提出了基于图像裁剪的对地小目标图像预处理方法,该方法可以裁剪掉图像中多余的背景,实现对地小目标多尺度训练,增强检测网络的鲁棒性,并且不需要消耗多余的显存。另外,针对红外弱小目标信号强度弱、常常受到环境中杂波和噪声干扰的特点,提出了基于局部对比度的红外弱小目标图像预处理方法,利用卷积和最大池化操作实现了改进的局部对比度测量算法,可以增强红外弱小目标信号,提高目标与背景之间的对比度。

第4章 基于超分辨率重建的对地小目标检测算法

4.1 引言

正如第2章所分析的,分辨率作为影响小目标检测的一个关键因素,对小目标检测起到至关重要的作用。因此,提高对地小目标的分辨率可以提高检测网络对对地小目标的检测能力。因为对地小目标在图像中占有像素少,携带的特征信息不明显。超分辨率重建技术可以将低分辨率图像重建成高分辨率图像,可以恢复目标更多的细节信息。因此,将超分辨率重建技术与对地小目标检测技术进行结合,将有助于检测对地小目标。

超分辨率重建技术中最常用的方法就是插值法,例如:双三次插值、双线性插值和最邻近插值。尽管这些方法实现起来很简单,但是它们在重建质量上仍然存在许多不足,一些细节特征常常被忽略,保证不了重建图像的质量。随着深度学习的发展,基于卷积神经网络的超分辨率重建技术取得了不错的成绩,重建后的高分辨率图像的质量也越来越高。因此,本章将采用基于深度学习的超分辨率重建网络来实现低分辨率对地小目标的重建。

本章的主要思想是将基于深度学习的超分辨率重建网络与对地小目标检测网络相结合来构造基于超分辨率重建的对地小目标检测网络 SuperDet,从而提高算法对对地小目标的检测率。

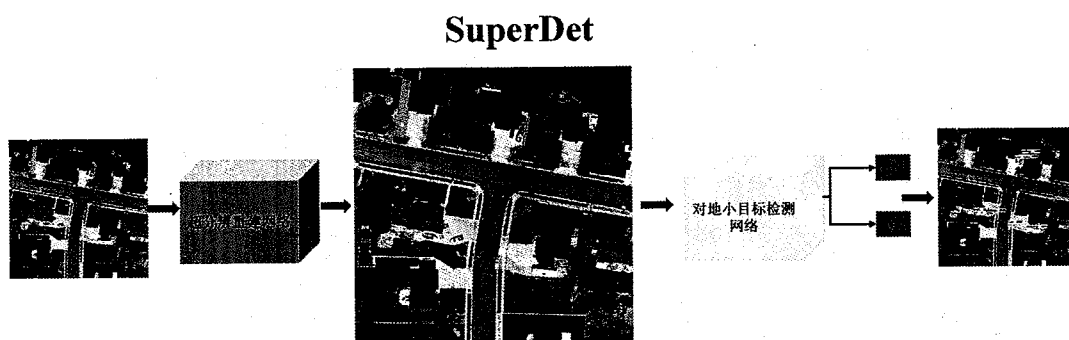


图 4.1 超分辨率重建与目标检测网络的结合

Figure 4.1 The combination of super-resolution reconstruction and the detection network

如图 4.1 所示, 将低分辨率的图像送入超分辨率重建网络, 实现高分辨率图像的重建。然后, 将高分辨率图像输入到对地小目标检测网络中实现对目标的定位和分类。为了实现 SuperDet 网络端到端的训练, 采用多任务损失来训练整个网络。

4.2 超分辨率重建网络设计

SuperDet 通过将超分辨率重建网络与对地小目标检测网络相结合来提高网络对小目标检测性能。本节的主要目的是为 SuperDet 设计合适的超分辨率重建网络。

4.2.1 一种 ESPCN 超分辨率重建网络的改进方法

为了保证结合超分辨率重建技术的对地小目标检测网络的时效性, 需要选择高效的超分辨率重建网络。虽然 VSDR^[109]、RED-Net^[110]、SRDensNet^[111]和 SRGAN^[112]都获得了高质量的高分辨率图像。但是, 这些网络层数较多、结构复杂, 不适合与对地小目标检测网络相结合。

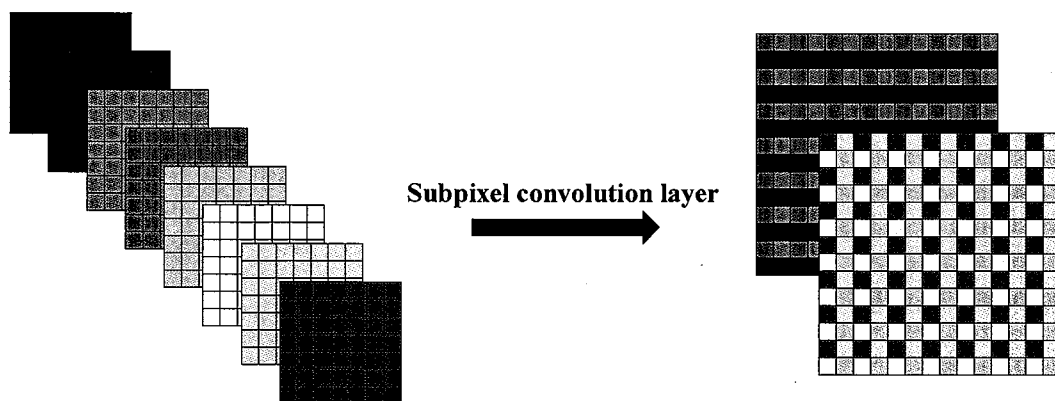


图 4.2 亚像素卷积层

Figure 4.2 Subpixel convolution layer

在本节中, 采用亚像素卷积层^[108]作为基础, 因为亚像素卷积层是一种快速、高效、无参数的上采样的操作。如图 4.2 所示, 利用亚像素卷积层可以将维度为 $H \times W \times r^2 C$ 的特征图转换成维度为 $rH \times rW \times C$ ($r=2$) 的特征图。借鉴 Residual network 的思想, 对原 ESPCN 超分辨率重建网络进行改进, 通过残差操作保留原图像中的细节特征。原 ESPCN 与改进后的 ESPCN 的结构对比如图 4.3 所示,

其中, $Conv(I, O, F)$ 代表卷积操作, I, O, F 分别代表输入通道数、输出通道数和卷积核的大小。在原 ESPCN 网络中, 首先是将输入的图像由 RGB 空间转换到 YCbCr 空间, 只对转换后的亮度通道(Y)进行重建。在输出端, 再将 YCbCr 空间转换成 RGB 空间。本章主要是将超分辨率重建技术应用在目标检测任务中, 通过提高对地小目标的分辨率来提高对对地小目标的检测性能。在目标检测任务中, 采用的图像是 RGB 空间。因此, 在改进的 ESPCN 中, 直接对图像的 RGB 空间进行重建, 可以避免由 RGB 空间与 YCbCr 空间相互转换带来的计算量。同时, 相比于单通道图像的重建, 三通道图像的重建可以恢复更多的图像细节信息。为了保证结合超分辨率重建技术的对地小目标检测网络的时效性, 需要设计高效的超分辨率重建网络, 因此, 将原 ESPCN 网络中卷积核为 5×5 的卷积的通道数由 64 降为 32。另外, 增加两个通道数为 16 的 3×3 卷积来增加网络的感受野, 提取更多语义信息。最后, 利用 Up-sample 层将输入的低分辨率图像变成与高分辨率图像相同的尺寸, 并用残差操作将 Up-sample 层输出的特征图与亚像素卷积层输出的特征图进行特征融合作为超分辨率重建网络的输出。

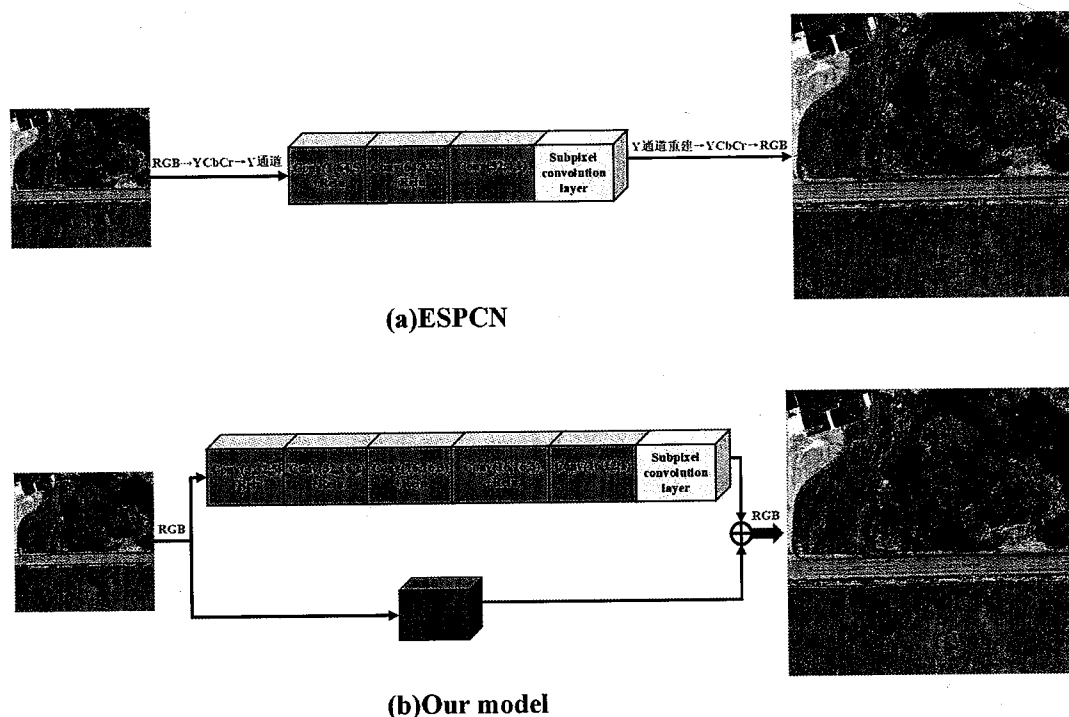


图 4.3 原 ESPCN 与改进的 ESPCN 的对比

Figure 4.3 The comparison between improved ESPCN and original ESPCN

4.2.2 实验与分析

为了验证改进的 ESPCN 超分辨率重建网络的性能, 在 VEDAI 数据集上进行对比实验。VEDAI 数据集中同时包含了 1024×1024 高分辨率图像和 512×512 低分辨率的图像。因此, 将 512×512 低分辨率的图像作为网络的输入, 将 1024×1024 高分辨率图像作为标准图(ground truth)。训练集一共有 1200 张图片, 测试集一共有 46 张图片。

实验环境: 在 Ubuntu16.0 操作系统下, 采用 pytorch 0.4.1 深度学习框架, 内存为 64G, NVIDIA GeForce GTX TITAN X 显卡。

评价指标: 采用峰值信噪比(Peak Signal to Noise Ratio, PSNR)来评价超分辨率重建模型的性能。PSNR 定义为信号最大功率和噪声功率的比值。PSNR 的计算公式如下:

$$PSNR = 10 \times \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (4.1)$$

其中, MAX_I 表示图像中像素的最大值, MSE 为预测的高分辨率图像与标准图的均方误差。

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I(i, j) - K(i, j))^2 \quad (4.2)$$

其中, 输出的高分辨率图像的大小为 $m \times n$, $I(i, j)$ 表示预测高分辨图像在点 (i, j) 处的值, $K(i, j)$ 表示标准图在点 (i, j) 处的值。

将改进的 ESPCN 分别与 Bicubic 和原 ESPCN 进行对比, 表 4.1 为 PSNR 值的对比结果。

表 4.1 不同算法 PSNR 值在 VEDAI 数据集上的对比结果

Table 4.1 PSNR comparison of different methods on VEDAI dataset

Method	Bicubic	ESPCN	Improved ESPCN
Input	512×512	512×512	512×512
Output	1024×1024	1024×1024	1024×1024
PSNR (dB)	31.8	32.9	33.6

如表 4.1 所示, Bicubic、ESPCN 和改进 ESPCN 的平均 PSNR 值分别为 31.8dB、32.9dB 和 33.6dB。改进 ESPCN 的 PSNR 值较 Bicubic 和 ESPCN 算法分别提高

了 1.8dB 和 0.7dB。

为了对比改进的 ESPCN 与其它算法在主观视觉上的差异, 随机选择了 4 张图像在 Bicubic、ESPCN 和改进 ESPCN 上进行了重建的可视化对比, 如图 4.4 所示。从图 4.4 中可以看出, 改进的 ESPCN 重建的高分辨率图像的细节更加清晰和真实, 观感更加舒适, 重建的图像更接近标准图。

无论是定性对比还是定量对比, 都证明了改进的 ESPCN 具有更好的高分辨率图像重建能力。

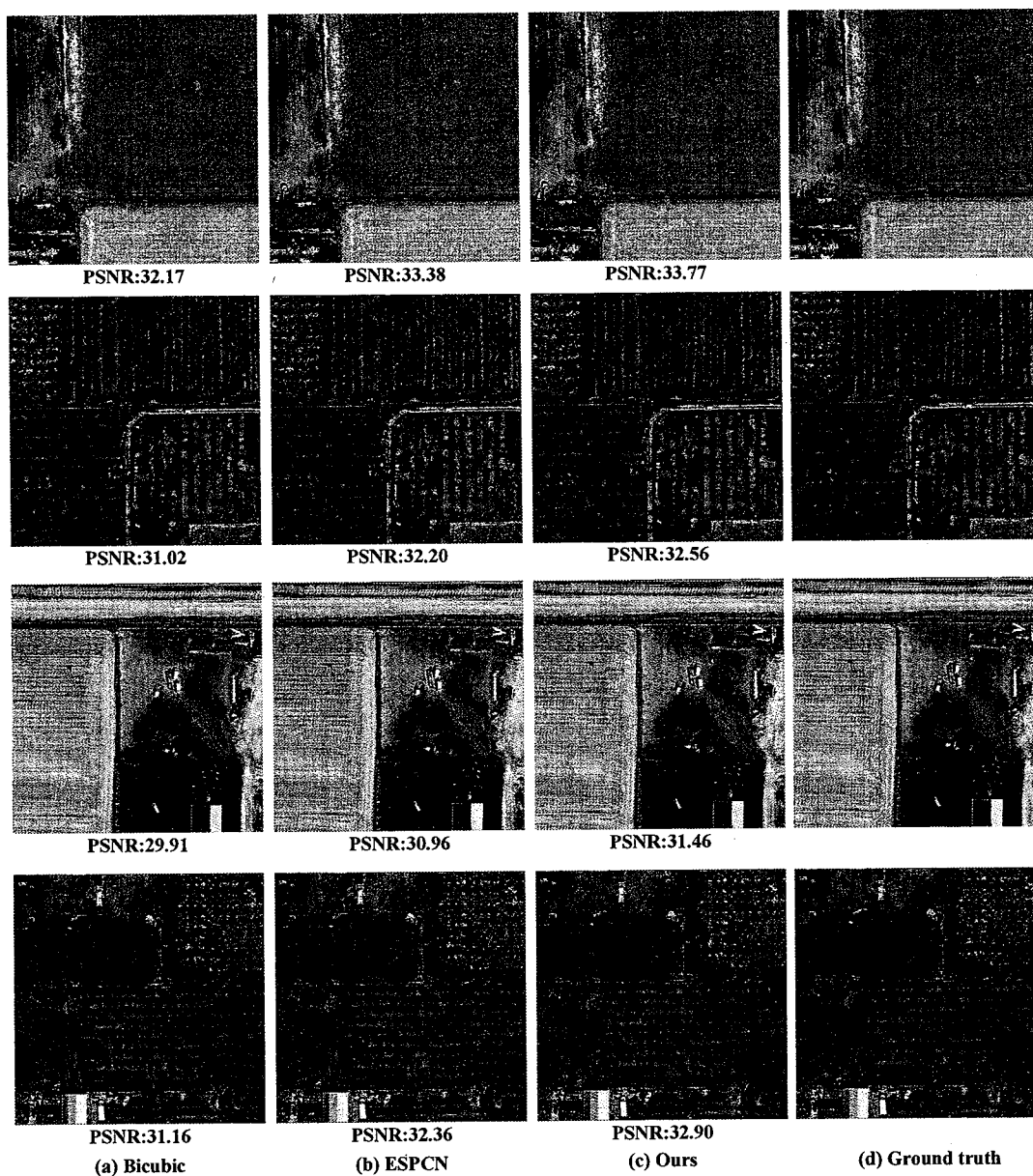


图 4.4 不同算法下图像重建效果比较

Figure 4.4 Comparison of image reconstruction effect with different methods

4.3 对地小目标检测网络设计

本节的主要目的是为 SuperDet 设计合适的对地小目标检测网络。结合第 2 章影响小目标的关键因素,设计了一个适用于对地小目标检测的网络,命名为 GSTDet(Ground small target detection network)。首先,为了给对地小目标选择合适的检测尺度,提出了基于等比例间隔原则的尺度匹配策略。根据尺度匹配策略,确定了用于对地小目标检测的 GSTDet 主干网络。随后,以膨胀卷积作为基本单元,设计了加权感受野融合模块,通过提高感受野来增加对地小目标周围的上下文信息。另外,采用深层特征和浅层特征相融合的方式来提高网络对对地小目标的定位和分类能力。最后,为了缓解正负样本数量的不均衡,采用 Focal loss 来设计置信度损失函数。同时,提出了一种正样本选择策略,通过增加网络训练过程中正样本的数量,来缓解正负样本数量的不均衡。

4.3.1 尺度匹配策略

目标检测算法,例如 SSD、DSSD、YOLO V3 和 RFBnet^[120]等,都是采用多个尺度来检测尺寸大小不同的目标。以 YOLO V3 为例,采用 3 个尺度来检测不同尺寸的目标,用 8 倍下采样特征图来检测小目标,16 倍下采样特征图来检测中目标,用 32 倍下采样特征图来检测大目标,如图 4.5 所示。目标大小的分布由 K-means 聚类算法对数据集进行聚类分析来确定。将聚类结果平均分成 3 份,分为小、中、大 3 类。然后,利用不同的尺度来检测不同大小的目标。虽然,YOLO V3 在小目标检测方面取得了不错的效果,但是 YOLO V3 并没有给出目标大小和网络尺度之间具体的数值关系,只是将聚类的结果平均分给了预先确定好的尺度。本章主要的研究对象是对地小目标检测,要设计一个适合于对地小目标检测的网络,首先要为对地小目标选择合适的尺度。因此,本章提出了基于等比例间隔原则的尺度匹配策略(Scale matching strategy),通过保证不同尺度的目标能够近似匹配相同数量的锚框来确定目标大小和网络尺度之间的关系。

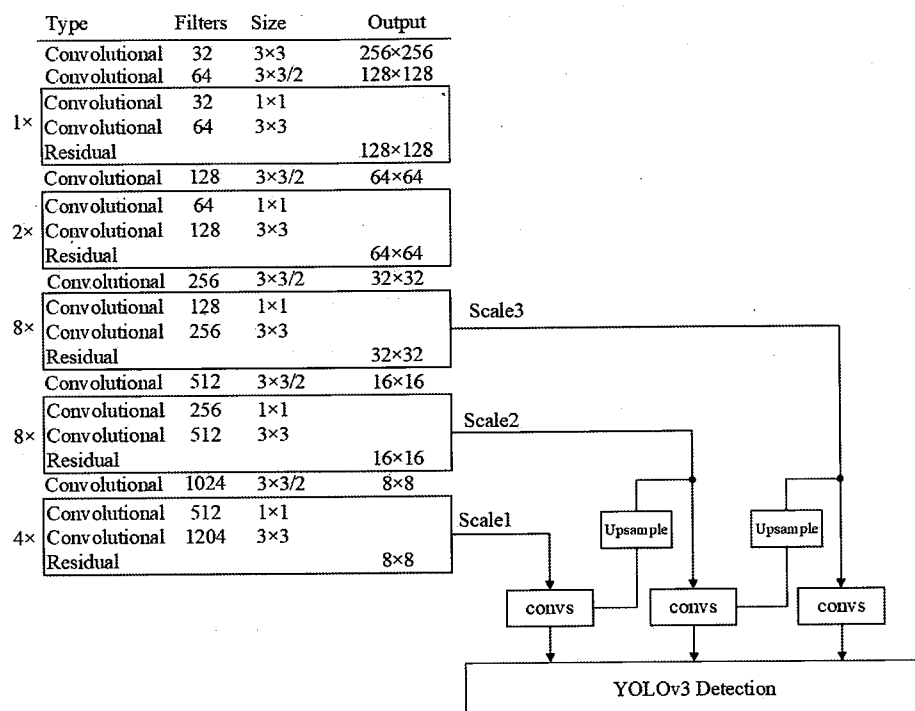


图 4.5 YOLO V3 结构图

Figure 4.5 The structure of YOLO V3

Zhang 等人^[121]提出了等比例间隔原则 (Equal-proportion interval principle), 目的是保证不同尺度的锚框在图像上都有相同的密度。如公式 4.3 所示, anchor size 表示锚框的大小, stride 表示特征图的步长, 特征图的步长(stride)与网络下采样的比例相同, density 表示锚框在图像上的密度。

$$density = \frac{anchor\ size}{stride} \quad (4.3)$$

如图 4.6 所示, 特征图的步长决定了锚框在图像上的间隔。例如, 特征图的步长为 4, 锚框的大小为 16×16, 这表示在图像上每隔 4 个像素就会有一个锚框。通过让不同尺度的锚框在图像上都有相同的密度, 可以使不同尺度的目标匹配相同数目的锚框。

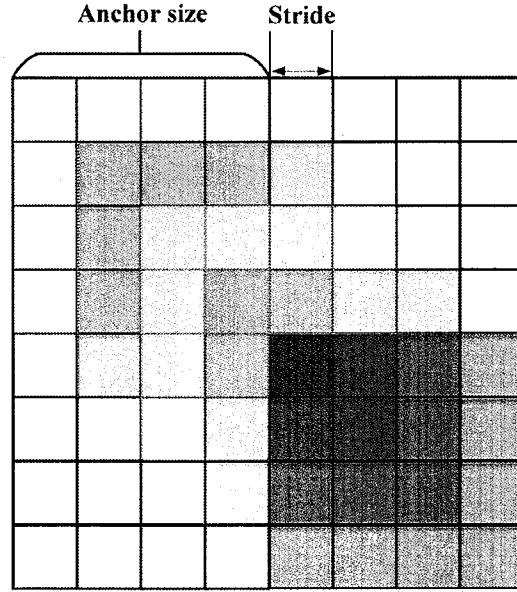


图 4.6 等比例间隔原则(density=4)

Figure 4.6 Equal-proportion interval principle (density=4)

虽然, Zhang 等人提出的等比例间隔原则保证了不同尺度的锚框在图像上都有相同的密度。但是, 不同尺度的锚框尺寸变化范围并不连续, 因为当密度固定时, 由公式 4.3 可知, 相应尺度的锚框尺寸也固定了。为了既能保证不同尺度的目标匹配相同数目的锚框, 又能保证不同尺度的锚框尺寸变化范围连续, 需要使不同尺度的锚框在图像上都有相同的密度变化范围。为了保证不同尺度的锚框在图像上都有相同的密度变化范围, 需要满足公式 4.4 和 4.5。

$$\frac{A_n}{2^n} = \frac{A_{n+1}}{2^{n+1}} \quad (4.4)$$

$$\frac{B_n}{2^n} = \frac{B_{n+1}}{2^{n+1}} \quad (4.5)$$

其中, B_n 和 B_{n+1} 分别代表锚框在尺度 n 和尺度 $n+1$ 上的最大值。 A_n 和 A_{n+1} 分别代表锚框在尺度 n 和尺度 $n+1$ 上的最小值。 2^n 代表尺度 n 上的下采样值。于是每一个尺度上的密度变化范围可以表示为 $[\frac{A_n}{2^n}, \frac{B_n}{2^n}]$ 。为了保证锚框的尺寸在不同尺度上可以保持连续, B_n 应该与 A_{n+1} 相等。于是可以得到公式 4.6。

$$\frac{A_n}{2^n} = \frac{B_n}{2^{n+1}} \quad (4.6)$$

因此, 每一个尺度上的密度变换范围可以表示为 $[\frac{A_n}{2^n}, \frac{2 \times A_n}{2^n}]$ 。根据不同密度的取值, 每一个尺度的锚框的尺寸范围如表 4.2 所示, 其中(2-4)代表密度的变化范围是 2 到 4。

表 4.2 不同密度下的锚框尺寸范围

Table 4.2 The range of anchor size of different density

n	Scale	The range of anchor size (2-4)	The range of anchor size (3-6)	The range of anchor size (4-8)	The range of anchor size (5-10)
1	2×	4~8	6~12	8~16	10~20
2	4×	8~16	12~24	16~32	20~40
3	8×	16~32	24~48	32~64	40~80
4	16×	32~64	48~96	64~128	80~160
5	32×	64~128	96~192	128~256	160~320

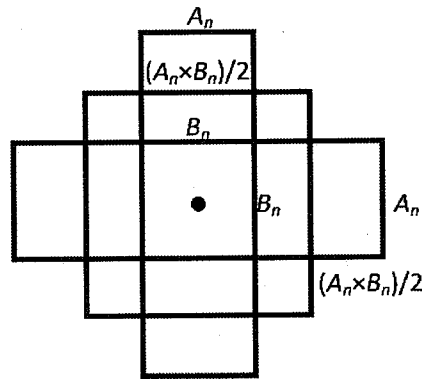


图 4.7 每个尺度的锚框

Figure 4.7 The anchors for each scale

如图 4.7 所示, 在每一个尺度上选择 3 个锚框, 锚框的宽和高分别为: $A_n \times B_n$ 、 $B_n \times A_n$ 和 $\frac{A_n + B_n}{2} \times \frac{A_n + B_n}{2}$ 。则在不同的密度变化范围内, 各尺度的锚框尺寸如表 4.3 所示。

表 4.3 每一个尺度上锚框的尺寸

Table 4.3 The aspect ratio of anchor boxes for each scale

Scale	Anchor box (2-4)	Anchor box (3-6)	Anchor box (4-8)	Anchor box (5-10)
2×	(4, 8), (6, 6), (8, 4)	(6, 12), (9, 9), (12, 6)	(8, 16), (12, 12), (16, 8)	(10, 20), (15, 15), (20, 10)
4×	(8, 16), (12, 12), (16, 8)	(12, 24), (18, 18), (24, 12)	(16, 32), (24, 24), (32, 16)	(20, 40), (30, 30), (40, 20)
8×	(16, 32), (24, 24), (32, 16)	(24, 48), (36, 36), (48, 24)	(32, 64), (48, 48), (64, 32)	(40, 80), (60, 60), (80, 40)
16×	(32, 64), (48, 48), (64, 32)	(48, 96), (72, 72), (96, 48)	(64, 128), (96, 96), (128, 64)	(80, 160), (120, 120), (160, 80)
32×	(64, 128), (96, 96), (128, 64)	(96, 192), (144, 144), (192, 96)	(128, 256), (192, 192), (256, 128)	(160, 320), (240, 240), (320, 160)

针对对地小目标检测，本文采用 Darknet53^[60]来设计 GSTDet 的主干部分。Darknet53 作为 YOLO V3 提取目标特征的主干网络(Backbone)，不仅结构简单，而且检测效果好。与 Resnet-101 相比，Darknet53 不仅具有更好的检测性能，而且检测速度是 Resnet-101 的 1.5 倍。Darknet53 的检测性能与 Resnet-152 相当，但是，它的检测速度是 Resnet-152 的 2 倍。

Darknet53 由 5 个残差块构成，借鉴了残差(Resnet)神经网络的思想^[26]。每个残差块由多个残差单元组成。如图 4.8(a)所示，残差单元是通过输入与两个卷积进行残差操作构建的。其中， I 和 O 分别代表残差单元输入和输出的通道数。 $Conv(A, B, C)$ 是由卷积、批归一化(batchnormalization)和 Leaky Relu 激活函数组成的，其中， A 为卷积核的大小， B 为输入通道数， C 为输出通道数，如图 4.8(b)所示。残差单元可以用公式 4.7 来表示。

$$x_{l+1} = \phi_{l+1}(x_l : w_{l+1}) + x_l \quad (4.7)$$

其中， x_l 和 x_{l+1} 分别代表第 l 层和 $l+1$ 层的输出， ϕ_{l+1} 代表第 $l+1$ 层中两个 $Conv$ 操作， w_{l+1} 代表两个 $Conv$ 操作中的权重。

通过引入残差单元使得网络深度可以更深，避免梯度消失。YOLO V3 对输入图片进行了 5 次降采样，并分别在最后 3 个降采样特征图上对不同尺寸的目标进行检测。

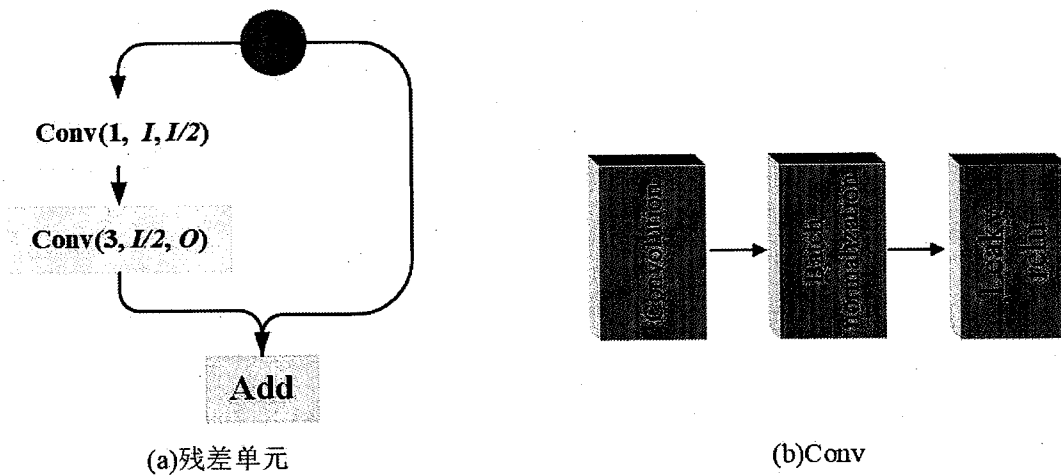


图 4.8 Darknet53 结构单元

Figure 4.8 Structural unit of Darknet53

在不同的密度范围下，同一目标可能被分配到不同的尺度上。因此，需要同时为对地小目标选择合适的尺度和锚框密度。本文研究的对地小目标的像素数量占整幅图像的 0.1% 到 1.5% 之间，以 512×512 的图像为例，目标大小介于 16×16 到 64×64 之间。

根据表 4.2 中每个尺度锚框尺寸的变化范围，选取 Darknet53 中相应的尺度作为 GSTDet 的主干网络。对于密度范围为(2-4)和(3-6)，4 倍下采样、8 倍下采样和 16 倍下采样的尺度适合对地小目标的检测。因此，选择 Darknet53 中 32 倍下采样之前的结构作为 GSTDet 的主干网络。对于密度范围分别为(4-8)、(5-10)和(6-12)，2 倍下采样、4 倍下采样和 8 倍下采样的尺度适合对地小目标检测。因此，选择 Darknet53 中 16 倍下采样之前的结构作为 GSTDet 的主干网络。如果选择的密度范围过低，一些小目标将被分配到大尺度上进行检测。如果选择的密度范围过高，一些相对大的目标将被分配到小尺度上检测，会影响网络对目标的检测性能。实验结果也验证了我们的想法，如表 4.4 所示。

表 4.4 不同密度范围下的实验结果

Table 4.4 The experimental results of different range of density

The range of density	(2-4)	(3-6)	(4-8)	(4-8)	(5-10)	(5-10)	(6-12)	(6-12)
Scale	4×, 8×, 16×	4×, 8×, 16×	2×, 4×, 8×	4×, 8×	2×, 4×, 8×	4×, 8×	2×, 4×, 8×	4×, 8×
mAP	71.7%	72.4%	67.8%	72.7%	67.3%	72.4%	66.4%	72.3%
FPS	28.1	28.3	29.6	50.3	29.4	49.5	28.6	49.5

如表 4.4 所示, 当密度为(4-8)、(5-10)和(6-12)时, 具有 4 倍下采样和 8 倍下采样尺度的主干网络要比具有 2 倍下采样、4 倍下采样和 8 倍下采样的尺度的主干网络具有更好的对地小目标检测能力。当包含 2 倍下采样检测尺度时, 目标的检测性能会明显下降, 这是因为 2 倍下采样检测尺度提取的特征缺少语义信息、感受野小, 不适合进行目标分类。密度范围为(4-8), 且具有 4 倍下采样和 8 倍下采样尺度的网络具有最好的小目标检测性能。原因如下: 浅层的特征中包含目标细粒度的特征, 有助于小目标的定位, 深层的特征包含更多语义信息, 有助于小目标的分类。如果密度范围过低, 一些小目标将被分配到大尺度进行检测, 例如密度范围为(2-4)和(3-6), 它们使用 16 倍下采样尺度对小目标进行检测, 不适合小目标的定位。当密度范围过高时, 一些相对大的目标将会被分配到小尺度上, 例如密度范围为(5-10)和(6-12)。因此, 选择 Darknet53 中 16 倍下采样之前的结构作为 GSTDet 的主干网络, 如图 4.9 所示, 采用密度范围为(4-8)的 4 倍下采样和 8 倍下采样尺度对对地小目标进行检测。

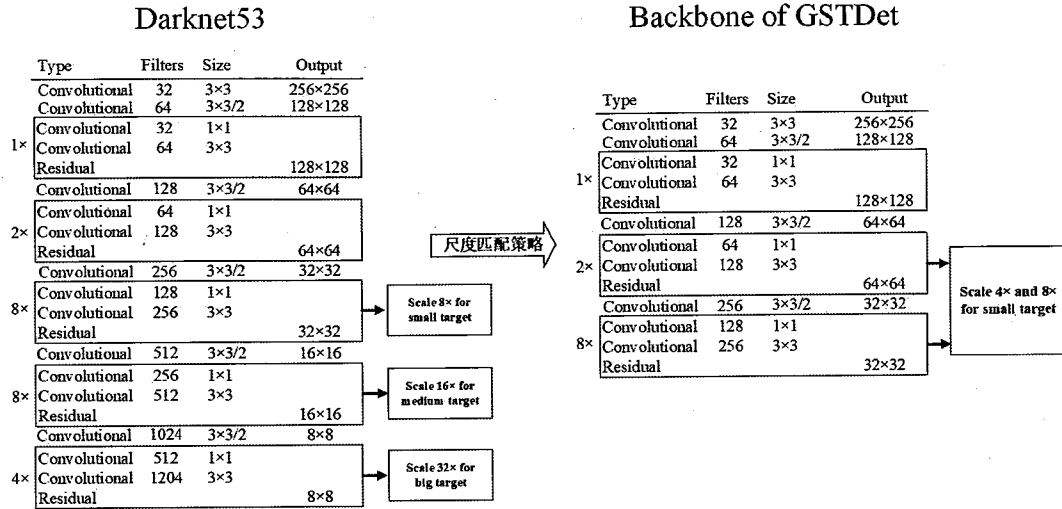


图 4.9 对地小目标检测网络 GSTDet 的主干

Figure 4.9 The backbone of GSTDet for ground small target detection

4.3.2 加权感受野融合模块

作为影响小目标检测的另一个关键因素，小目标周围的上下文信息对小目标检测至关重要。扩大感受野可以有效增加小目标周围的上下文信息。膨胀卷积^[122]可以在不改变特征图分辨率的前提下，通过改变卷积的膨胀系数来改变感受野。如图 4.10 所示，膨胀卷积利用不同的膨胀系数来扩张卷积核的大小，进而改变感受野。

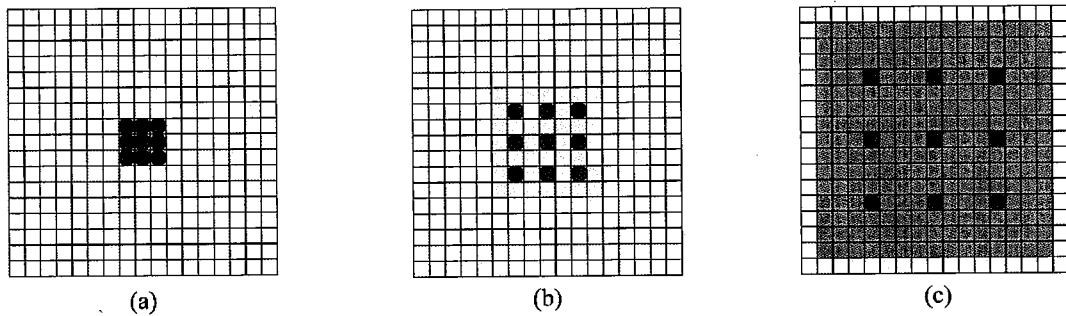


图 4.10 膨胀卷积。(a) 膨胀系数=1 (b) 膨胀系数=2 (c) 膨胀系数=4

Figure 4.10 Dilated convolution. (a) dilated rate=1 (b) dilated rate=2 (c) dilated rate=4

利用公式 4.8 来计算网络的感受野。

$$RF_k = RF_{k-1} + \left((f_k - 1) \times \prod_{i=1}^{k-1} s_i \right) \quad (4.8)$$

其中, RF_k 和 RF_{k-1} 分别代表第 k 层和第 $k-1$ 层的感受野, f_k 代表卷积核的大小, s_i 为第 i 层的步长。

表 4.5 对比了由 3 个 3×3 普通卷积构成网络的感受野和由 3 个膨胀系数分别为 1、2 和 4 的 3×3 膨胀卷积构成网络的感受野。对于 3 个 3×3 普通卷积构成的网络, 其感受野为 7×7 。对于由 3 个膨胀系数分别为 1、2 和 4 的 3×3 膨胀卷积构成的网络, 其感受野为 15×15 , 感受野扩大了一倍多。

表 4.5 普通卷积和膨胀卷积感受野的对比

Table 4.5 The receptive field comparison between normal convolution and dilated convolution

Layer	Network	Size/Stride	RF	Network	Size/Stride/Dilated rate	RF
0	Convolution	$3 \times 3/1$	3×3	Dilated Convolution	$3 \times 3/1/1$	3×3
1	Convolution	$3 \times 3/1$	5×5	Dilated Convolution	$3 \times 3/1/2$	7×7
2	Convolution	$3 \times 3/1$	7×7	Dilated Convolution	$3 \times 3/1/4$	15×15

Li 等人^[123]提出了尺度自适应三叉戟网络(Scale-Aware Trident Networks), 该网络利用具有不同膨胀系数的膨胀卷积来预测不同大小的目标, 然后将每个分支预测的结果利用非极大值抑制进行筛选, 结构如图 4.11 所示。

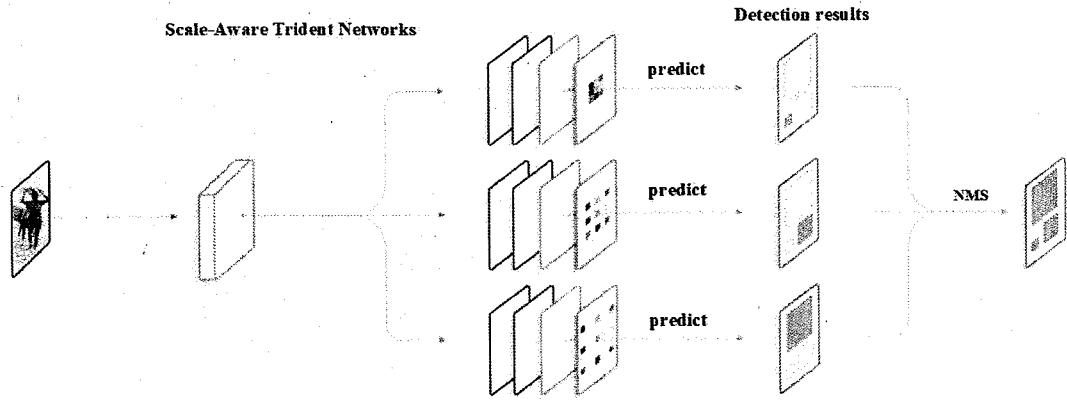


图 4.11 尺度自适应三叉戟网络的结构

Figure 4.11 The structure of Scale-Aware Trident Networks

受^[123]的启发，以膨胀卷积作为基础单元，设计了加权感受野融合模块 (Weighted receptive field fusion module)。尺度自适应三叉戟网络通过利用不同膨胀系数的膨胀卷积来预测不同大小的目标。不同于尺度自适应三叉戟网络，加权感受野融合模块利用不同膨胀系数的膨胀卷积来增大感受野，通过自适应融合具有不同感受野的特征来增强目标周围的上下文信息。

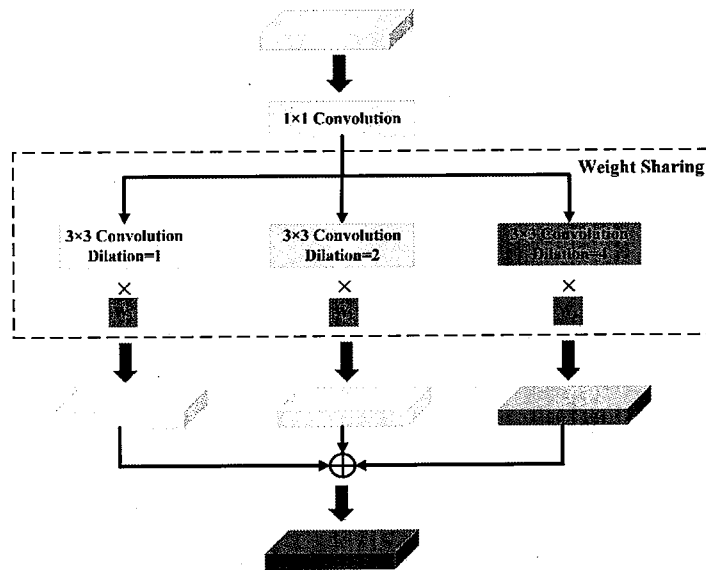


图 4.12 加权感受野融合模块的结构图

Figure 4.12 The structure of weighted receptive field fusion module

加权感受野融合模块的结构如图 4.12 所示，首先利用一个 1×1 卷积来对输

入的特征进行降维，然后用 3 个膨胀系数分别为 1、2 和 4 的 3×3 卷积对感受野进行扩张。为了进一步提高网络对地小目标的检测能力，将具有不同感受野的特征进行融合。然而，具有不同感受野的特征对最终融合后的特征的贡献是不相同的。因此，为每个感受野的特征都增加了一个可以自动学习的权重系数。用 W_i 来表示每个感受野的权重，用 F_i 来表示每个感受野的特征。于是加权感受野融合模块最后输出的融合特征可以用下式计算。

$$F_{fusion} = \sum_{i=0}^2 W_i \cdot F_i \quad (4.9)$$

为了防止训练过程中权重系数变化剧烈、不稳定，采用下式对权重系数进行归一化处理。

$$W_j^* = \frac{W_j}{\sum_{i=0}^2 W_i + \varepsilon} \quad (4.10)$$

其中， ε 值为 0.0001，防止分母为 0。

另外，在加权感受野融合模块中，对膨胀系数分别为 1、2 和 4 的 3×3 卷积之间采用权值共享的策略。权值共享有两方面的优势：第一个优势是可以减少冗余的参数，使网络变得更加高效。第二个优势是相同的参数经过具有不同感受野的特征训练之后，可以使网络变得更加鲁棒。

4.3.3 特征融合机制

卷积神经网络中浅层的特征包含目标细粒度的信息，对目标的定位十分重要。深层的特征包含更多的语义信息，对目标的分类非常重要。将深层特征与浅层特征进行融合有助于提高网络对目标的检测能力。

特征金字塔网络 FPN 是最先提出将深层特征和浅层特征相融合的网络。特征金字塔网络 FPN 采用一种自上而下(top-down)的结构，利用上采样操作将低分辨率特征图放大到与高分辨率特征图相同的尺寸，然后进行特征的融合。实验结果表明，融合了深层和浅层特征的网络具有更好的目标检测性能。多尺度预测与 FPN 结构图的对比如图 4.13 所示。

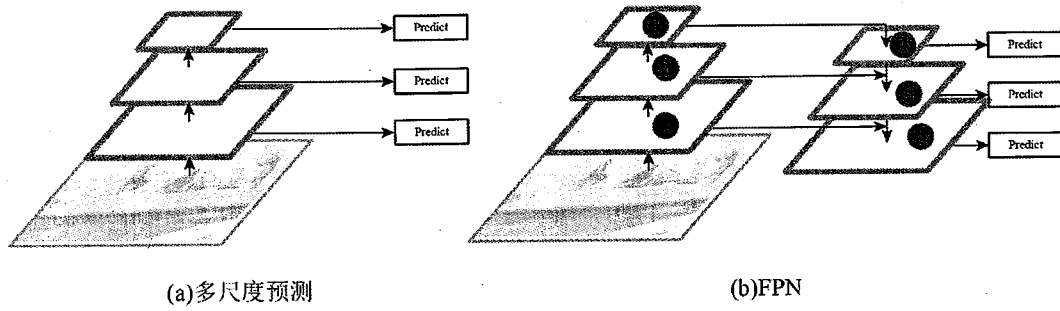


图 4.13 多尺度预测与 FPN 的对比

Figure 4.13 The comparison of multi-scale prediction and FPN

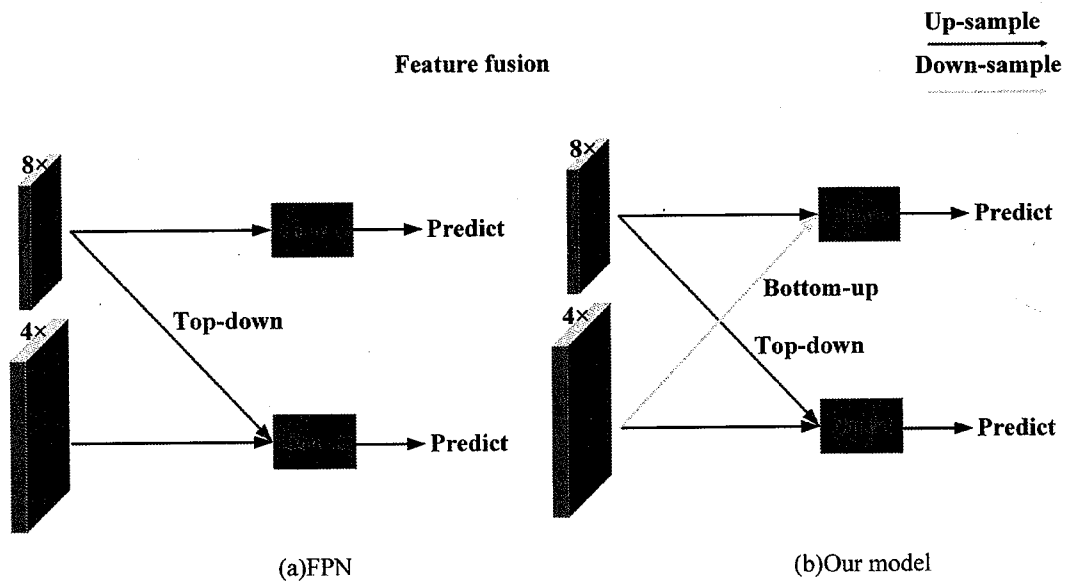


图 4.14 对地小目标检测网络 GSTDet 中的特征融合

Figure 4.14 Feature fusion of GSTDet for ground small target detection

特征金字塔网络采用自上而下的结构进行特征融合，提升了目标检测性能。但是，特征金字塔仅仅利用上采样操作，使浅层的高分辨率特征图与深层的低分辨率特征图进行融合，如图 4.13(b)所示，特征图 4 通过上采样操作与特征图 2 相融合得到特征图 5，特征图 5 并没有融合特征图 1 的浅层特征，特征图 4 也没有融合特征图 2 的浅层特征。对于对地小目标检测网络 GSTDet，根据尺度匹配策略，采用尺度为 4 倍下采样和 8 倍下采样的特征图来检测对地小目标。如图 4.14 所示，同样采用自上而下的 FPN 结构对 8 倍下采样的特征图进行上采样操

作，与 4 倍的下采样特征图进行融合。另外，采用自下而上(bottom-up)的结构对 4 倍下采样的特征图进行下采样操作，与 8 倍的下采样特征图进行融合，使浅层的特征与深层的特征得到充分融合。

4.3.4 对地小目标检测网络的结构

对地小目标检测网络 GSTDet 的结构如图 4.15 所示。根据尺度匹配策略，采用 Darknet53 中 4 倍下采样特征图和 8 倍下采样特征图作为检测对地小目标的两个尺度。然后，分别在两个输出尺度上连接加权感受野融合模块，来增加对地小目标周围的上下文信息。对 4 倍下采样特征图进行下采样，对 8 倍下采样特征图进行上采样，利用串联(Concat)操作来融合不同尺度的特征信息。最后，用 2 个 1×1 的卷积对检测结果进行输出。其中， $N_{output} = N_{boxes} \times (N_{classes} + 4 + 1)$ ， N_{output} 为输出检测结果的维度，4 代表预测的目标框的长、宽和中心坐标的偏移量，1 代表预测目标的置信度， N_{boxes} 表示特征图上每一个小网格上锚框的数量， $N_{classes}$ 代表目标种类的数量。

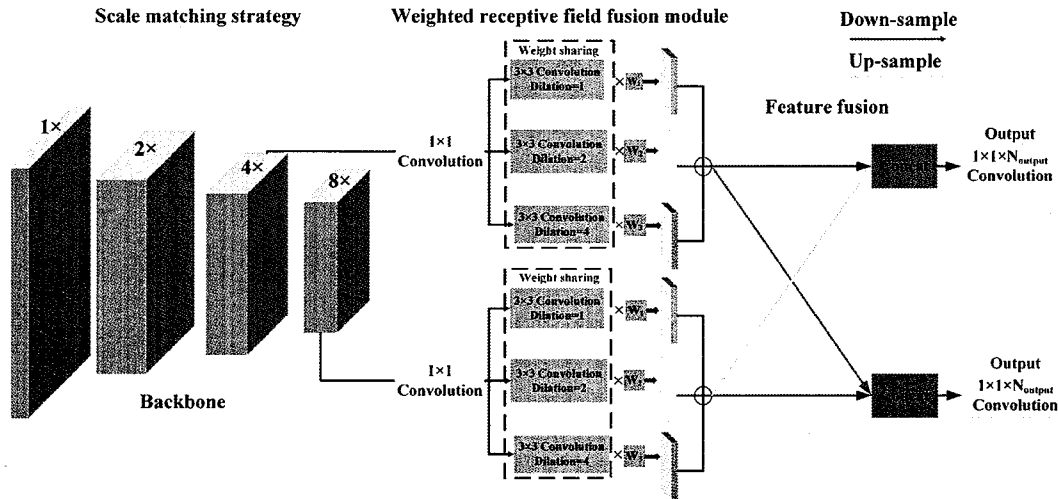


图 4.15 对地小目标检测网络 GSTDet 的结构图

Figure 4.15 The structure of GSTDet for ground small target detection

4.3.5 正样本选择策略

在对地小目标检测任务中，由于目标在图像中占有的像素非常少，使得负样本的数量较多，正样本的数量较少，存在正负样本数量的不均衡的问题。为了缓解对地小目标检测中正负样本数量不均衡的问题，本节提出了正样本选择策略

(Positive anchors selection strategy), 如图 4.16 所示, 其中, 蓝色框为目标框, 红色框为锚点框。YOLO V3 在选择正样本时, 将图像分成若干个小网格, 然后将由 K-means 聚类算法得到的锚点框分配到每一个网格中, 由目标中心所在的小网格来负责该目标。计算目标与该小网格上每个锚点框之间的交并比, 选择交并比最大的锚点框作为正样本, 负责对目标进行回归。在正样本选择策略中, 同样让包含目标中心的小网格来负责该目标。不同的是, 在所有的尺度上, 选择目标框与锚点框之间 $IOU \geq 0.3$ 的所有锚点框作为正样本。如果所有锚点框与目标框之间的交并比都小于 0.3, 则选择交并比最大的锚点框作为正样本。通过这种方式, 将会增加正样本的数量。正样本选择策略具有两个优势: 第一个优势是正样本的数量将增加, 可以缓解目标检测过程中正负样本数量不均衡的问题; 第二个优势是同一个目标可以同时由不同尺度下的锚点框负责, 有助于提高网络的鲁棒性。

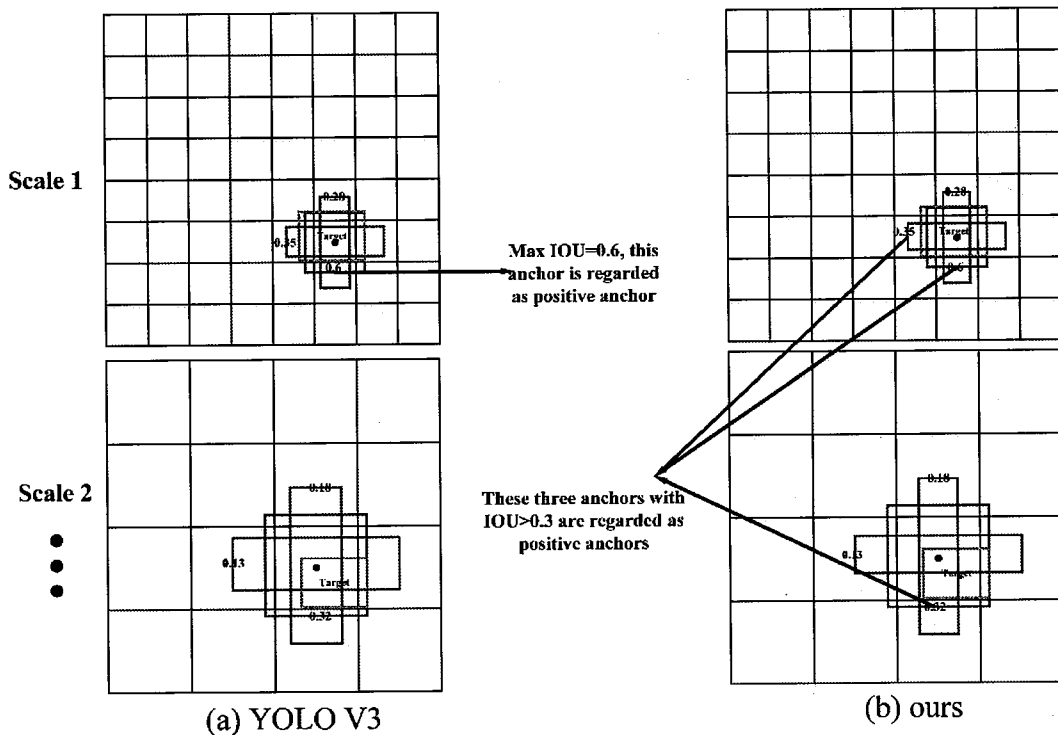


图 4.16 正样本选择策略

Figure 4.16 The positive samples selection strategy

网络对图像中的每一个目标都可能输出多个预测框。为了过滤掉冗余的预测框, 提高目标检测质量, 利用非极大值抑制对检测结果进行处理。非极大值抑制首先对输出的预测框按照置信度的大小进行排列, 然后选取置信度值最高的预测

框作为检测结果,同时删除与这个预测框交并比大于阈值的其它预测框,通过不断的迭代可以有效过滤掉冗余的预测框。

4.4 基于超分辨率重建的对地小目标检测网络结构

在本节中,主要是将超分辨率重建技术应用到对地小目标检测中。因此,将4.2节设计的改进 ESPCN 超分辨率重建网络与4.3节对地小目标检测网络 GSTDet 相结合,建立基于超分辨率重建的对地小目标检测网络,命名为 SuperDet。SuperDet 的结构如图 4.17 所示。

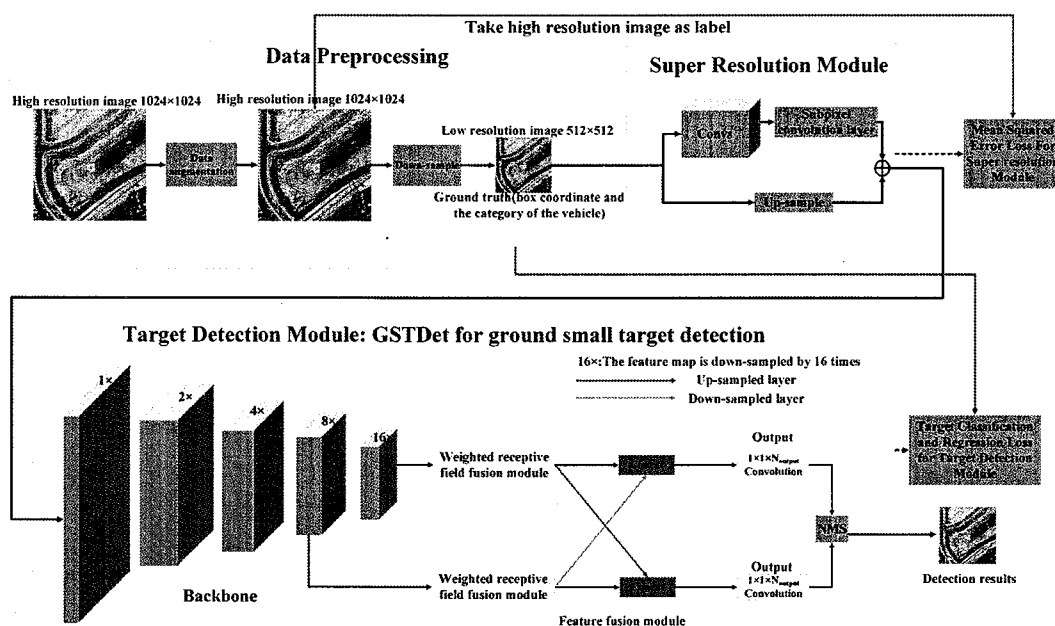


图 4.17 SuperDet 的结构

Figure 4.17 The structure of SuperDet

如图4.17所示,SuperDet主要由3部分构成,数据预处理(Data Preprocessing)、超分辨率模块(Super Resolution Module, SRM)和对地小目标检测模块(Target Detection Module, TDM)构成。超分辨率模块和对地小目标检测模块是 SuperDet 内部相连的两个模块。目标的检测流程如下:将预处理后的低分辨率图像作为输入,送到超分辨率模块中,使低分辨率图像可以重建成高分辨率图像,然后将重建后的高分辨率图像作为输入送到对地小目标检测模块,完成对地小目标位置和类别的预测。其中,超分辨率模块采用4.2节中改进 ESPCN 网络,而对地小目

标检测模块采用 4.3 节中的对地小目标检测网络 GSTDet。为了实现对 SuperDet 网络端到端的训练,采用多任务损失来设计损失函数。

数据预处理部分: SuperDet 的主要目的是利用超分辨率重建技术来提高检测网络对对地小目标的检测性能。为了能够实现从低分辨率图像到高分辨率图像的重建,需要将高分辨率图像作为训练超分辨率模块的标准图。如图 4.17 所示,在训练的过程中,首先对高分辨率图像进行处理。为了增加网络的泛化能力,对输入的高分辨率图像采用随机水平翻转、随机颜色变换、随机裁剪和随机映射等数据增广的方式进行处理。然后,将经过数据增广处理的高分辨率图像作为训练超分辨率模块的 ground truth。并对高分辨率图像进行下采样作为网络输入的低分辨率图像。SuperDet 经过训练之后,可以实现将低分辨率图像重建成为高分辨率图像。因此,在测试阶段,只需要将低分辨率的图像作为 SuperDet 的输入即可。

超分辨率模块: 此模块主要是实现高分辨率图像的重建。采用 4.2 节中改进的 ESPCN 网络作为超分辨率模块。因为亚像素卷积层是一种快速、高效、无参,利用像素重排进行上采样的操作。并且,改进 ESPCN 网络结构简单、实时性好,适合于与检测网络相结合。超分辨率重建网络,例如:SRCNN、FSRCNN 和 ESPCN 等,都是将输入的图像从 RGB 空间转换成 YCbCr 空间,只是对转换后的亮度通道(Y)进行预测。然后,将预测的 Y 通道与经过插值处理的 Cb 和 Cr 通道进行合并,输出最后的高分辨率图像。采用 Y 通道预测的原因是相比于色差信息,人类视觉对亮度的变化更加敏感^[106]。在本节中,将超分辨率重建网络与对地小目标检测网络相结合来构建 SuperDet,其中对地小目标检测模块的输入是 RGB 空间的图像。因此,对于 SuperDet 中的超分辨率模块,采用 RGB 空间的图片作为输入,这样就避免了 RGB 空间与 YCbCr 空间相互转换带来的计算量。同时,相比于单通道图像的重建,三通道图像的重建可以使预测的高分辨率图像更接近于 ground truth。

对地小目标检测模块: 采用 4.3 节对地小目标检测网络 GSTDet 作为对地小目标检测模块。由于对地小目标检测模块的输入是高分辨率的图像,目标的高和宽是原低分辨率图像的 2 倍,根据尺度匹配策略,将原对地小目标检测网络 GSTDet 的检测尺度由 4 倍下采样和 8 倍下采样变成 8 倍下采样和 16 倍下采样。

4.5 多任务损失函数

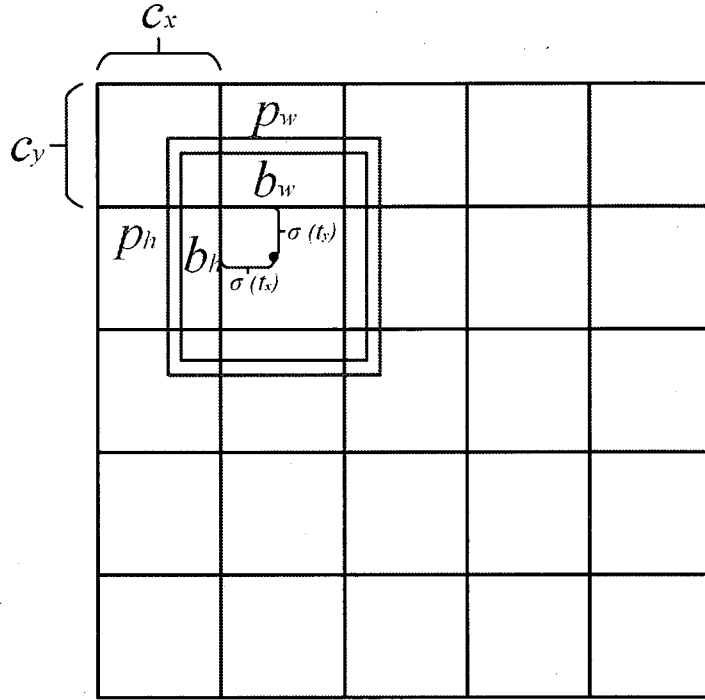


图 4.18 检测框的回归

Figure 4.18 The regression of the bounding box

如图 4.18 所示，对检测框坐标的回归采用预测相对位置^[59]的方法，其中，红色框为锚框，蓝色框为预测的目标框。预测相对位置可以使回归过程更加稳定。检测框的回归公式如 4.11-4.14 所示。

$$b_x = \sigma(t_x) + c_x \quad (4.11)$$

$$b_y = \sigma(t_y) + c_y \quad (4.12)$$

$$b_w = p_w e^{t_w} \quad (4.13)$$

$$b_h = p_h e^{t_h} \quad (4.14)$$

其中， c_x 和 c_y 为特征图中目标中心所在网格左上角的坐标， p_w 和 p_h 为预设锚框映射到特征图中的宽和高， t_x 、 t_y 、 t_w 和 t_h 为网络预测相对的位置， b_x 、 b_y 、 b_w 和 b_h 为检测框的输出结果， $\sigma(\cdot)$ 代表 sigmoid 函数。

SuperDet 主要由超分辨率模块和对地小目标检测模块构成，因此，SuperDet 的损失函数也由两部分组成，即超分辨率重建损失和对地小目标检测损失。

对于超分辨率重建损失，采用均方误差损失，如下式。

$$Loss_{SRM} = \frac{1}{4HWC} \sum_{c=1}^C \sum_{x=1}^{2H} \sum_{y=1}^{2W} (H_{x,y,c}^* - S_{x,y,c})^2 \quad (4.15)$$

其中， $H \times W \times C$ 代表低分辨率图像的维度，经过超分辨率模块后，低分辨率图像将转换成维度为 $2H \times 2W \times C$ 的高分辨率图像。 S 代表超分辨率模块输出的高分辨率图像， H^* 代表 ground truth。

经过超分辨率模块之后，输出的高分辨率图像被送到对地小目标检测模块中用来预测对地小目标的位置信息和类别信息。

为了提高模型的检测精度，利用 GIOU loss^[124]和 Focal loss^[63]来设计多尺度目标检测器的损失函数。与交并比(Intersection over Union, IOU)相比，GIOU 可以反映出预测框与目标框之间的远近关系。并且，GIOU loss 可以有效避免因为预测框与 ground truth 不相交而导致损失函数梯度为 0 的问题。因此，利用 GIOU loss 来对预测框的位置进行回归。GIOU 可以表示为

$$GIOU_{B_{GT}, B_P} = \frac{|B_{GT} \cap B_P|}{|B_{GT} \cup B_P|} - \frac{|B \setminus (B_{GT} \cup B_P)|}{|B|} \quad (4.16)$$

式中： B_{GT} 表示目标框的 ground truth， B_P 表示预测框， B 表示包围 B_{GT} 和 B_P 最小的矩形框。则 GIOU loss 可以表示为

$$Loss_{GIOU} = 1 - GIOU_{B_{GT}, B_P} \quad (4.17)$$

另外，在相同的预测偏离情况下，小目标预测框与 ground truth 的交并比要比大目标的交并比变化大，即小目标代价应该大于大目标。因此，将目标框位置损失乘以一个尺度变化系数 s ^[60]。

$$s = 2 - \frac{GT_w \times GT_h}{Width \times Height} \quad (4.18)$$

其中， GT_w 和 GT_h 分别代表目标框真实的宽和高， $Width$ 和 $Height$ 分别代表图像的宽和高。 s 值的变化范围在 1 到 2 之间，目标框尺寸越小，对应的 s 就越大。

于是，目标框位置损失可以表示为

$$Loss_{GIOU} = s \times (1 - GIOU_{B_{GT}, B_P}) \quad (4.19)$$

正负样本数量不均衡作为影响小目标检测的一个关键因素，影响网络对小目标的检测性能。并且，当训练样本中包含了大量容易分辨的负样本时，会导致这些简单负样本对 loss 函数起主要作用。为了增强网络对困难样本的预测，采用

Focal loss 来对目标的置信度进行回归，缓解正负样本数量不均衡的问题。

$$Loss_{conf} = \begin{cases} -\alpha(1-y_p)^\gamma \times \log y_p, & y_{GT}=1 \\ -(1-\alpha)y_p^\gamma \times \log(1-y_p), & y_{GT}=0 \end{cases} \quad (4.20)$$

式中： y_{GT} 表示目标置信度的 ground truth， y_p 表示预测目标的置信度， α 和 γ 分别设为 0.25 和 2。

对于分类损失，采用二进制交叉熵，如式 3.18 所示。

$$Loss_{cls} = C_{GT} \log C_p - (1 - C_{GT}) \log(1 - C_p) \quad (4.21)$$

式中： C_{GT} 表示目标类别的 ground truth， C_p 表示预测目标的类别。

用 $Loss_{TDM}$ 来表示对地小目标检测模块的损失，如公式 4.22。

$$Loss_{TDM} = \sum_{n=1}^2 Loss_{GLOU}^n + Loss_{conf}^n + Loss_{cls}^n \quad (4.22)$$

式中： n 代表网络包含的尺度。

则 SuperDet 的总 loss 可以表示为

$$Loss_{SuperDet} = Loss_{TDM} + Loss_{SRM} \quad (4.23)$$

4.6 实验与分析

为了证明 SuperDet 的检测性能，首先，对 SuperDet 进行消融实验，验证尺度匹配策略、加权感受野融合模块、特征融合机制和超分辨率模块的有效性。然后，分别在 DOTA 数据集^[125]、VEDAI 可见光数据集和 VEDAI 红外数据集上验证对地小目标检测网络 SuperDet 对对地小目标的检测性能。

4.6.1 实验条件与评价方法

实验环境和训练细节：随着深度学习的发展，许多科技公司 and 高校开发了深度学习开发框架，推动了深度学习在工程中的应用。其中比较著名的深度学习开发框架包括：Caffe^[126]、Pytorch^[127]、TensorFlow^[128]、Keras^[129]、Torch^[130]、MXNet^[131]、CNTK^[132]。本文选择的深度学习框架为 Pytorch 0.4.1，因为 Pytorch 开发框架使用灵活，具有简洁、高效和易用性强等优点。所有的模型都采用随机梯度下降法（stochastic gradient descent, SGD）在 1 块 NVIDIA GeForce GTX TITAN X 显卡上进行训练。利用随机水平翻转、颜色变换、随机剪切和随机仿射变换来进行数

据增广。采用余弦学习率策略(cosine learning rate)^[133]来训练各检测模型,权重衰减和动量分别设置为 0.0005 和 0.9。

在 DOTA 数据集和 VEDAI 数据集上对所有的模型均训练 80 个轮次(epoch)。为了使训练过程更加稳定,在训练最初的 4 个轮次采用“预热(warm-up)”策略,从第 4 个轮次到第 80 个轮次采用余弦学习率策略使学习率从 0.0001 下降到 0.000001。

评价指标:采用平均精准度(Average Precision, AP)来测量每一类目标的检测精度。平均精准度是从召回率和准确率两个角度来衡量检测算法的准确性,是评价检测模型准确性的直观评价标准,可以用来分析单个类别的检测效果^[134],计算公式如 4.24 所示。

$$AP = \int_0^1 p(r) dr \quad (4.24)$$

其中, r 代表召回率, $p(r)$ 代表在召回率为 r 时的检测精度。平均精度表示对于某一类目标在不同召回率下的检测精度的积分。

利用平均准确率均值(mean Average Precision, mAP)来测量多类目标的平均检测精度, mAP 值越高表示模型在全部类别中检测的综合性能越高,计算公式如 4.25 所示。

$$mAP = \frac{\sum_{i=1}^n AP(i)}{n} \quad (4.25)$$

其中, i 代表目标的类别, n 为类别数, $AP(i)$ 为目标类别为 i 时的平均精度。

利用每秒帧数(Frame per second, FPS)来衡量目标检测器的检测速度。

数据集介绍:为了验证本节提出的 SuperDet 对对地小目标的检测性能,分别在 VEDAI 数据集和 DOTA 数据集上进行实验。

VEDAI 数据集是由大小为 1024×1024 的卫星图像构成的,这些图像是由大视场卫星图像分割得到的,包含了各式各样的车辆目标,背景和混淆目标。数据集中包含可见光和红外两种图像。所有的图像都采用了与地面相同的距离,是典型应用于监视侦察的场景。此外,数据集中还包含了对 1024×1024 图像进行下采样得到的大小为 512×512 的图像,为评估提供更小、更具挑战性的目标图像。平均每张 VEDAI 数据集中的图像都包含约 5.5 个车辆目标,并且它们只占图像总

像素的 0.7%，是典型的小目标检测数据集。数据集中共有 9 类对象，包括：airplane, boat, car, truck, tractor, camping, pickup, vans, others。9 类对象中每种类别的数量如表 4.6 所示。

表 4.6 VEDAI 数据集中各类目标的数据统计

Table 4.6 Statistics of each target in VEDAI dataset

Class name	Boat	Camping	Car	Others	Pickup	Tractors	Truck	Vans	Airplane
Total numbers	170	390	1340	200	950	190	300	100	47

利用 VEDAI 数据集中大小为 1024×1024 的高分辨率图像作为超分辨率模块的标准图用于训练 SuperDet，用大小为 512×512 的低分辨率图像用于测试 SuperDet。

DOTA 数据集^[125]是一个应用于目标检测的遥感数据集，一共有 2806 张由不同场景拍摄的航拍图像，包含多个类别，其中篮球场、网球场和码头等类别都是大目标，因此，本文只选择 small-vehicle 的类别作为检测对象。由于 DOTA 数据集中图像的尺寸不统一，分布在 800×800 到 4000×4000 之间，因此，将包含 small-vehicle 的图像裁剪成大小为 512×512 的图像。裁剪后的图像中有 713 张图像用于训练，135 张图像用于测试。将经过数据增广，大小为 512×512 的图像作为超分辨率模块的 ground truth，并将此 512×512 的图像下采样至 256×256 ，作为 SuperDet 的输入。

4.6.2 消融实验

为了验证尺度匹配策略、加权感受野融合模块、特征融合机制和正样本选择策略对对地小目标检测网络 GSTDet 的检测结果的影响。在 VEDAI 数据集上对 GSTDet 进行消融实验来证明以上每个方法的有效性，实验结果如表 4.7 所示。在训练和检测阶段，所有的图片都以 512×512 的分辨率输入到网络。表中的 Baseline 是指由尺度匹配策略所确定的基于 Darknet 53 的主干网络，即选择 Darknet 53 中 4 倍下采样和 8 倍下采样作为小目标检测的尺度。Weighted receptive field fusion module*表示权值不进行共享的加权感受野融合模块。

表 4.7 在 VEDAI 数据集上的消融实验

Table 4.7 Ablation study on VEDAI dataset

Baseline	Feature fusion	Weighted receptive field fusion module*	Weighted receptive field fusion module	Positive anchors selection strategy	mAP	FPS
✓					69.8%	68.5
✓	✓				70.6%	62.5
✓	✓	✓			72.1%	50.1
✓	✓		✓		72.7%	50.3
✓	✓		✓	✓	74.1%	50.3

Baseline: 由尺度匹配策略确定的主干网络取得了 69.8% 的 mAP 值, 并且检测速度为 68.5FPS。

特征融合: 与 Baseline 相比, 通过对深层特征和浅层特征的融合, 使检测网络对地小目标的检测能力进一步提升, mAP 值由 69.8% 提升到了 70.6%。检测速度仍然保持 62.5FPS。

权值不共享的加权感受野融合模块: 通过采用加权感受野融合模块, 可以自适应地融合具有不同感受野的特征, 增加对地小目标周围的上下文信息。结合了加权感受野融合模块的网络 mAP 值提高了 1.5 个百分点, 并且仍然保持着 50.1FPS 的检测速度。

权值共享的加权感受野融合模块: 通过采用权值共享策略, mAP 值由 72.1% 提高到了 72.7%。实验结果进一步证明了相同的参数在不同的感受野下进行训练, 可以使检测网络变得更加鲁棒。

正样本选择策略: 采用了正样本选择策略, mAP 值提高了 1.4 个百分点, 检测速度几乎没有变化。该结果说明了正样本选择策略可以缓解对地小目标检测中正负样本数量不均衡的问题。

SuperDet 是由超分辨率模块与对地小目标检测模块构成的。为了进一步验证超分辨率模块对地小目标检测的影响, 在 VEDAI 数据集上对 SuperDet 进行了消融实验, 实验结果如表 4.8 所示。表中的 Baseline 是指由尺度匹配策略所确定的基于 Darknet 53 的主干网络, 即选择 Darknet 53 中 8 倍下采样和 16 倍下采样作为对地小目标检测的尺度。

表 4.8 SuperDet 在 VEDAI 数据集上的消融实验

Table 4.8 The ablation study of SuperDet on VEDAI dataset

Baseline	With weighted receptive field fusion module	With feature fusion module	With SRM	mAP	FPS
√				70.0%	53.9
√	√			72.1%	47.7
√	√	√		73.5%	41.3
√	√	√	√	77.6%	29.3

如表 4.8 所示, Baseline 在 VEDAI 数据集上取得了 70.0% 的 mAP 值, 并且检测速度为 53.9FPS。当结合了加权感受野融合模块之后, 对地小目标周围的上下文信息变得更加丰富, 网络的 mAP 值提高了 2.1 个百分点, 且检测速度仍然保持在 47.7FPS。通过将浅层的特征与深层的特征进行融合, 使 mAP 从 72.1% 提高到了 73.5%。特别的是, 当引入超分辨率模块之后, mAP 值上升了 4.1 个百分点, 较原检测网络有明显的提高, 检测速度为 29.3FPS。实验证明了引入超分辨率模块有助于提升网络对对地小目标的检测性能。

4.6.3 DOTA 数据集上的验证

为了证明 SuperDet 对对地小目标的检测性能, 首先在 DOTA 数据集上进行了对比实验, 对比结果如表 4.9 所示。

表 4.9 不同检测算法在 DOTA 数据集上的检测结果

Table 4.9 Results of different detectors on DOTA dataset

Detector	YOLO V2	Tiny-YOLO V3	YOLO V3	GSTDet	SuperDet
Input	512	512	512	256	256
FPS	58.3	76.4	14.7	130.0	95.3
mAP (%)	72.7	73.2	88.3	89.0	91.3

如表 4.9 所示, 以大小为 256×256 作为输入的 SuperDet 取得的 mAP 值比以大小为 256×256 作为输入的 GSTDet 取得的 mAP 值提高了 2.3 个百分点。这证明了低分辨率图像经过 SuperDet 超分辨率模块之后, 可以有效地对低分辨率图像进行重建, 恢复更多目标的细节, 从而提高网络对对地小目标的检测性能。另

外, SuperDet 仍然保持在 95.3FPS 的检测速度。

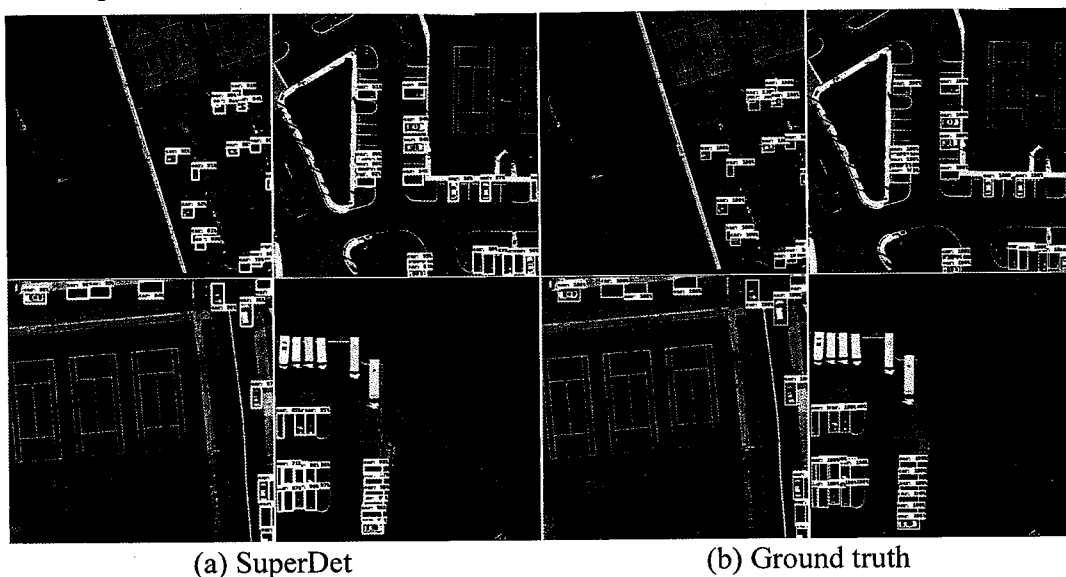


图 4.19 SuperDet 在 DOTA 数据集上的可视化结果

Figure 4.19 Visualization of the detection results of SuperDet on DOTA dataset

在 DOTA 数据集上对 SuperDet 也进行了定性评估,可视化的结果如图 4.19 所示。通过与 ground truth 进行对比可以看出,即使车辆十分密集,SuperDet 也可以精确地对车辆目标进行分类和定位。可视化结果再次证明 SuperDet 对对地小目标具有较好的检测效果。

4.6.4 VEDAI 可见光数据集上的验证

为了进一步证明 SuperDet 对对地小目标的检测性能,在 VEDAI 数据集上进行了对比实验。SuperDet 与其他算法的对比结果如表 4.10 所示。表中 Faster RCNN 和 Faster RER-CNN 是双阶段目标检测算法的代表。YOLO V2, YOLO V3, SSD 和 DAGN 是单阶段目标检测算法的代表。其中 Faster RER-CNN 和 DAGN 是专用于对地小车辆目标检测的算法。

表 4.10 不同检测算法在 VEDAI 数据集上的检测结果

Table 4.10 Results of different detectors on VEDAI dataset

Detector	Faster RCNN	SSD	YOLO V2	YOLO V3	Faster RER-CNN	DAGN	SuperDet
Input	512	512	512	512	512	512	512
FPS	6.3	22.0	55.2	31.5	2.7	25.1	29.3
mAP	67.0%	46.1%	50.3%	61.6%	70.8%	67.1%	77.6%

如表 4.10 所示, 在 VEDAI 数据集上, 与 Faster RER-CNN 和 DAGN 相比, mAP 值分别提高了 6.8 个百分点和 10.5 个百分点。并且, SuperDet 的检测速度是 Faster RER-CNN 的 11 倍, 是 DAGN 的 1.17 倍。在 VEDAI 数据集上的实验结果证明了 SuperDet 具有很好的小目标检测能力。

为了定性地对 SuperDet 进行评估, 将此网络的检测结果进行可视化, 如图 4.20 所示。通过与 ground truth 进行对比可以看出, 即使在复杂的背景下, SuperDet 也可以精确地对车辆目标进行分类和定位。

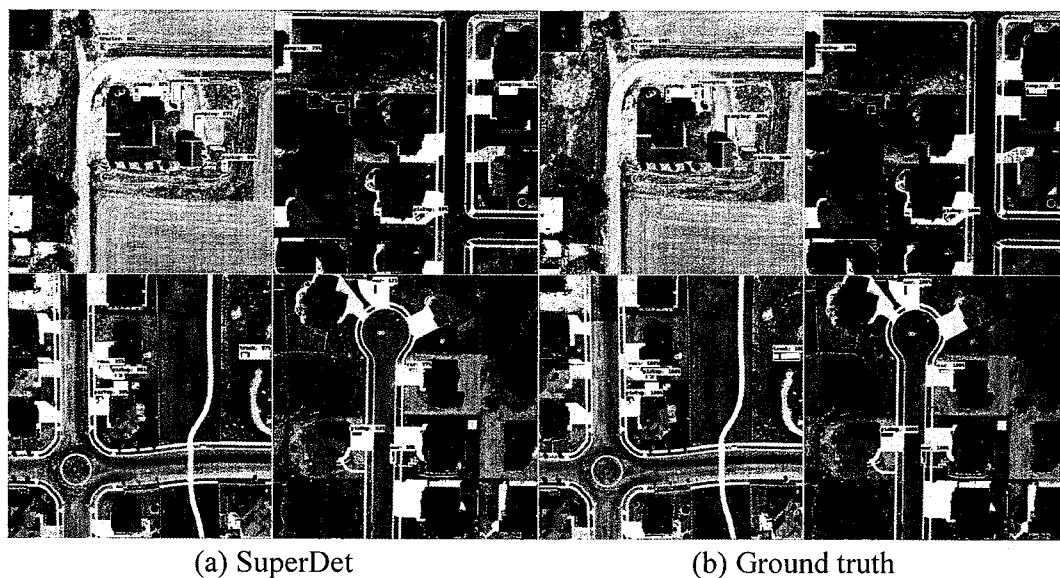


图 4.20 SuperDet 在 VEDAI 数据集上的可视化结果

Figure 4.20 Visualization of the detection results of SuperDet on VEDAI dataset

4.6.5 VEDAI 红外数据集上的验证

SuperDet 同样适合于红外对地小目标检测。在 VEDAI 红外数据集上对

SuperDet 进行训练和测试, 实验结果如表 4.11 所示, 表中 Pi、Ca、Tr、Bo、Va、Ot、Tra 和 Pl 分别代表 Pickup、Camping、Truck、Boat、Van、Other、Tractor 和 Plane。

表 4.11 可见光图像与红外图像检测结果的对比

Table 4.11 The comparison of the detection results between color images and infrared images (mAP: %)

method	Car	Pi	Ca	Tr	Bo	Va	Ot	Tra	Pl	mAP	FPS
GSTDet (infrared)	80.7	77.0	78.1	58.6	54.3	80.9	33.6	58.9	100	69.1	50.0
SuperDet (infrared)	86.6	84.3	76.7	69.8	60.8	82.7	28.6	65.3	100	72.8	30.1
SuperDet (color)	88.6	82.4	77.3	70.0	64.1	93.8	47.2	77.7	100	77.6	30.3

如表 4.11 所示, 在 VEDAI 数据集上, 几乎每一类红外对地小目标的 AP 值都要低于可见光对地小目标的 AP 值。相比于可见光对地小目标检测, mAP 值下降了 4.8 个百分点。相比于 GSTDet, SuperDet 结合了超分辨率重建技术, 使 mAP 值提高了 3.7 个百分点。这证明了利用超分辨率重建技术来恢复对地小目标的细节特征, 可以有效地提高网络对红外对地小目标的检测性能。

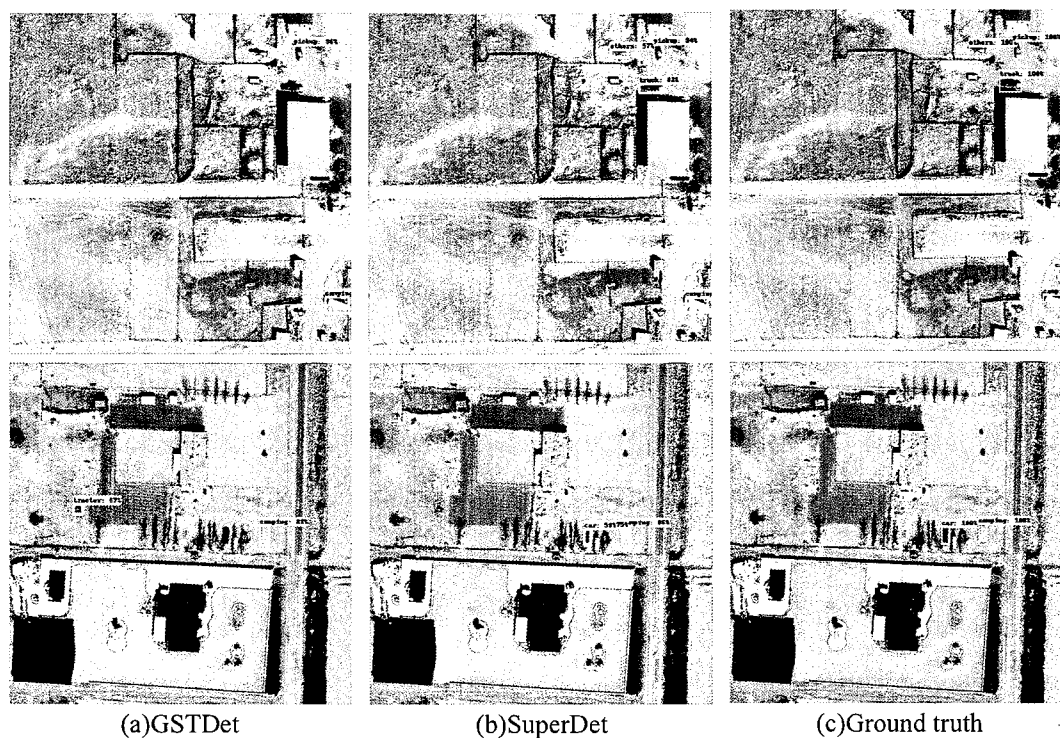


图 4.21 GSTDet 与 SuperDet 在红外图像上的可视化检测结果

Figure 4.21 Visual detection results of GSTDet and SuperDet on infrared images

为了定性对比 GSTDet 与 SuperDet 在红外图像上的检测效果,分别对二者的检测结果进行了可视化。如图 4.21 所示,相比于 GSTDet, SuperDet 可以有效地降低对地小目标的错检率和漏检率。

4.7 本章小结

作为影响小目标检测的一个关键因素,提高小目标的分辨率有助于提高网络对小目标的检测性能。本章将基于深度学习的超分辨率重建技术与基于深度学习的目标检测算法相融合,提出了基于超分辨率重建的对地小目标检测网络 SuperDet。该网络主要由两部分构成,分别是超分辨率模块和对地小目标检测模块。其中,超分辨率模块实现了低分辨率图像到高分辨率图像的转换,对地小目标检测模块则是由 GSTDet 构成的,将重建之后的图像送到对地小目标检测模块中,完成最后对地小目标的检测。为了实现对 SuperDet 端到端的训练,采用多任务损失函数对超分辨率重建任务的损失和目标检测任务的损失进行回归。为了

验证 SuperDet 对对地小目标的检测性能,分别在 DOTA 数据集和 VEDAI 数据集上进行了对比实验。实验结果表明,结合超分辨率重建技术的 SuperDet 具有更为出色的对地小目标检测能力。

第5章 基于图像过滤的对空红外弱小目标检测算法

5.1 引言

作为空中目标探测的常用手段,红外成像制导利用背景与目标之间的热辐射差异进行成像,可实现在有烟尘、浓雾和强光干扰的环境下对目标进行昼夜监视。红外弱小目标检测可以在对空作战中用于侦察目标,锁定目标。正如第2章所分析的,目标信号弱,背景掺杂噪声是影响红外弱小目标检测的一个关键因素。如图5.1所示,当小目标处于较远的位置时,在红外图像上通常表现为一个小亮点,往往只占几个像素,不具备纹理结构和形状特征。由于成像距离远,使检测到的目标信号很弱,并且伴随着噪声的干扰。最终,使目标呈现出“弱小”的特点。因此,红外弱小目标检测是目标检测领域中的一个挑战。本章将结合红外弱小目标信号弱,背景掺杂噪声的特点,设计增强红外弱小目标响应、抑制背景响应的方法,并结合红外弱小目标检测网络完成对红外弱小目标的检测。

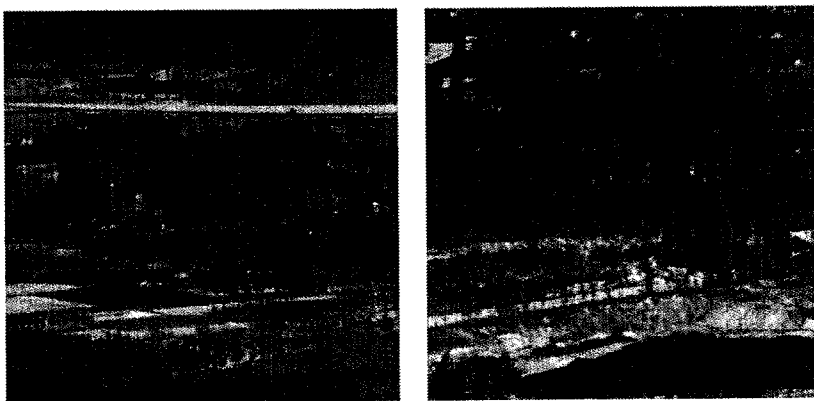


图 5.1 红外弱小目标图像

Figure 5.1 Image of infrared dim and small target

为了过滤背景噪声、降低虚警率,本章设计了图像过滤模块对红外弱小目标的图像进行过滤。本章的主要思想是将图像过滤模块与红外弱小目标检测模块相结合来构造基于图像过滤的对空红外弱小目标检测网络 FilterDet,从而提高算法对红外弱小目标的检测率。

本章的主要设计框架如图 5.2 所示, 首先, 将红外图像送入到过滤模块进行过滤。然后, 将过滤后的红外图像输入到红外弱小目标检测模块中, 完成对红外弱小目标的检测。

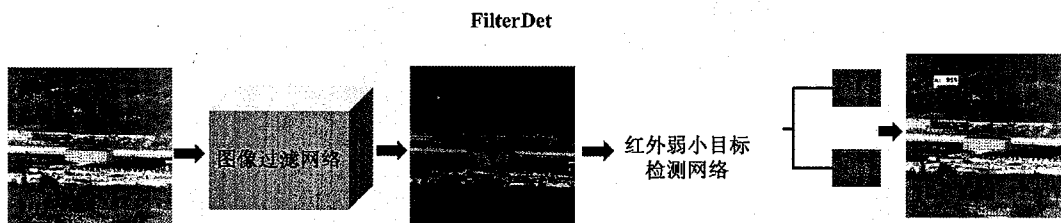


图 5.2 对空红外弱小目标检测的框架

Figure 5.2 The framework of aerial infrared dim and small target detection

5.2 图像过滤模块的设计

FilterDet 通过将图像过滤模块与对空红外弱小目标检测网络相结合来提高网络对红外弱小目标的检测性能。本节的主要目的是为 FilterDet 设计合适的图像过滤模块(Image filtering module, IFM)。图像过滤模块的任务是通过学习来分辨红外图像中的前景与背景, 对图像中的目标和背景进行二分类。图像过滤模块输出的是一个二分类的置信度图, 图中背景的置信度趋于 0, 目标的置信度趋于 1。将置信度图作为掩膜, 与原红外图像相乘, 可以有效地过滤掉背景噪声, 并保留红外目标信息。

图像过滤模块以残差单元作为基本单元, 采用沙漏结构^[135]来设计图像过滤模块。先对输入图像进行下采样操作来提取特征, 然后利用上采样操作将特征图恢复成原图像的尺寸, 图像过滤模块的详细配置如表 5.1 所示。由于浅层特征中包含了有利于定位的细粒度特征, 深层特征中包含了有利于分类的语义特征。因此, 对深层和浅层中具有相同分辨率的特征图进行融合来提高图像过滤模块的泛化能力, 如表中的第 8 层与第 4 层的特征融合、第 11 层与第 2 层的特征融合和第 14 层与第 0 层的特征融合。

表 5.1 图像过滤模块的配置

Table 5.1 Configuration of Image Filtering Module

Layer	Layer name	Size/Stride	Filters	Output
0	Convolution	3/1	32	256×256
1	Convolution	3/2	64	128×128
2	<i>Residual</i> (64,64)			128×128
3	Convolution	3/2	128	64×64
4	<i>Residual</i> (128,128)			64×64
5	Convolution	3/2	256	32×32
6	<i>Residual</i> (256,256)			32×32
7	Upsample			64×64
8	<i>Residual</i> (256,128)			64×64
9	Add	Layer8+Layer4		64×64
10	Upsample			128×128
11	<i>Residual</i> (128,64)			128×128
12	Add	Layer11+Layer2		128×128
13	Upsample			256×256
14	<i>Residual</i> (64,32)			256×256
15	Add	Layer14+Layer0		256×256
16	Convolution	1/1	1	256×256

图像过滤模块输出的是图像的置信度图，表示每个位置包含目标的可能性，值越大表示目标存在的可能性越大，输出的置信度图与输入的红外图像具有相同的分辨率。为了使图像过滤模块通过训练来学习分辨红外图像中的前景与背景，并获得置信度图，首先需要确定原图的置信度 ground truth。利用目标框 ground truth 来生成置信度 ground truth，生成的具体方式如下：如果一个像素 $GT_{h,w}$ 在目标框 ground truth 内，就将这个像素 $GT_{h,w}$ 分配为正样本 1，其它的为负样本 0，具体过程如表 5.2 所示。

表 5.2 置信度 ground truth 的生成伪代码

Table 5.2 Pseudo code of the generation of confidence ground truth

Algorithm: The generation of the ground-truth confidence.

Input: The label of the infrared small targets.

Output: The ground-truth confidence.

1. Suppose the width and the height of the input infrared image are H and W . $GT_{h,w}$ is the value of the ground truth confidence in the position (h, w) . T as bounding box ground truth, (x_{top}, y_{top}) as the left top coordinate of T , (x_{bottom}, y_{bottom}) as the right bottom coordinate of T .
 2. for $h=1:H$
 3. for $w=1:W$
 4. if $w \geq x_{top}$ and $h \geq y_{top}$ and $w \leq x_{bottom}$ and $h \leq y_{bottom}$
 5. $GT_{h,w} = 1$
 6. else
 7. $GT_{h,w} = 0$
 8. end
 9. end
 10. end
-

将置信度图进行了可视化, 如图 5.3 所示。图 5.3(a)代表输入的红外图像, 图 5.3(b)表示输出的置信度图。置信度图中每一个像素的取值范围都在 0 到 1 之间, 值越大表示该像素点是目标的可能性越大。因此, 利用置信度图来过滤输入图像, 将输入图像与置信度图进行点乘, 得到过滤之后的图像, 如图 5.3(c)所示。很明显经过置信度图处理后的红外图像, 目标响应得到了增强, 并且有效的抑制了背景响应。

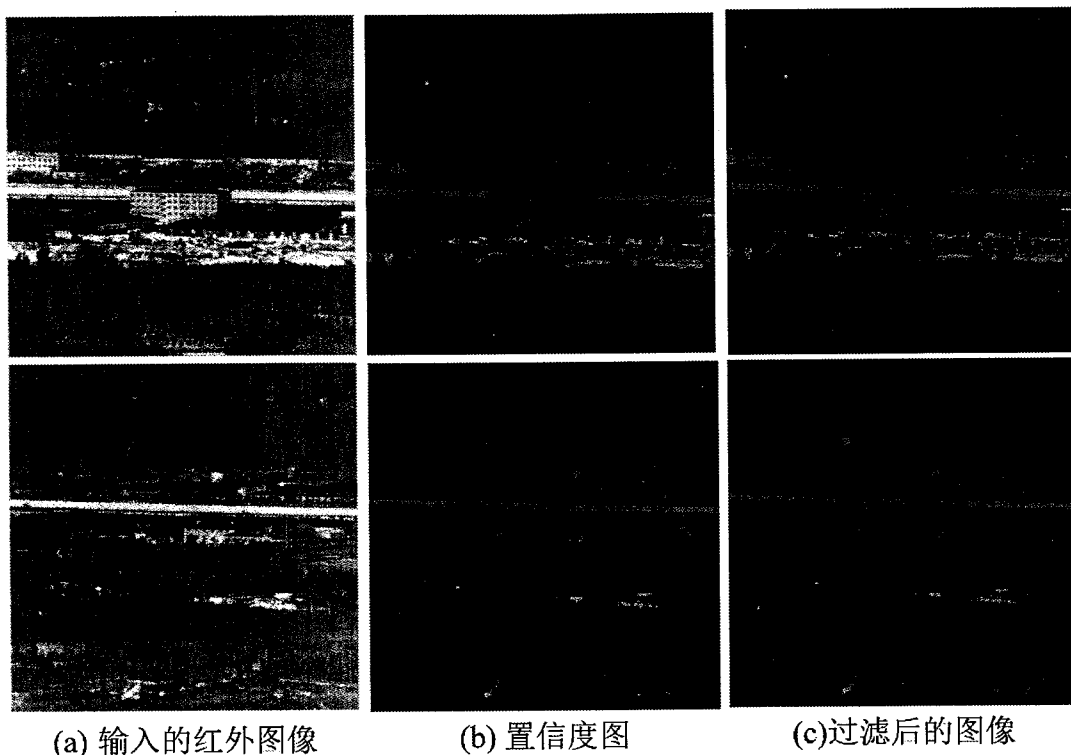


图 5.3 图像过滤模块输出的置信度图

Figure 5.3 Confidence map of image filtering module output

5.3 对空红外弱小目标检测网络设计

本节的主要目的是为 FilterDet 设计合适的对空红外弱小目标检测网络。结合第 2 章影响小目标的关键因素, 设计了一个适用于对空红外弱小目标检测的网络, 命名为 ISTDet(Infrared small target detection network)。首先, 利用尺度匹配策略确定了用于对空红外弱小目标检测的 ISTDet 主干网络。随后, 利用加权感受野融合模块来增加小目标周围的上下文信息。另外, 增加空间注意力机制来增强网络的表达能力, 使红外弱小目标的特征响应得到增强。最后, 采用深层特征和浅层特征相融合的方式来提高网络对红外弱小目标的定位和分类能力。

5.3.1 主干网络设计

对空红外弱小目标与对地小目标相比, 目标尺寸更小, 通常以一个“亮点”的形式呈现。对于一个尺寸为 256×256 的图像, 红外弱小目标的尺寸通常不超过 8×8 。根据尺度匹配策略, 应该选择 1 倍下采样的尺度来检测, 然而 1 倍下采样

尺度提取特征的能力有限, 缺少语义信息, 不适合目标检测。因此, 采用沙漏 (Hourglass) 结构来设计针对红外弱小目标检测的 ISTDet 的主干网络, 通过这种沙漏结构不仅可以使红外弱小目标在 1 倍下采样的尺度上检测, 还可以保证网络中包含充足的语义信息。网络的配置如表 5.3 所示, 其中, $Residual(I, O)$ 表示残差单元, I 和 O 分别代表残差单元输入和输出的通道数。ISTDet 的主干网络如图 5.4 所示。

表 5.3 对空红外弱小目标检测网络 ISTDet 的主干配置

Table 5.3 The configuration of ISTDet backbone for aerial infrared dim and small target detection

Layer	Layer name	Size/Stride	Filters	Output
0	Convolution	3/1	32	256×256
1	Convolution	3/2	64	128×128
2	$Residual(64, 64)$			128×128
3	Convolution	3/2	128	64×64
4	$Residual(64, 64)$			64×64
5	Convolution	3/2	256	32×32
6	$Residual(64, 64)$			32×32
7	Upsample			64×64
8	$Residual(64, 64)$			64×64
9	Upsample			128×128
10	$Residual(64, 64)$			128×128
11	Upsample			256×256
12	$Residual(64, 64)$			256×256

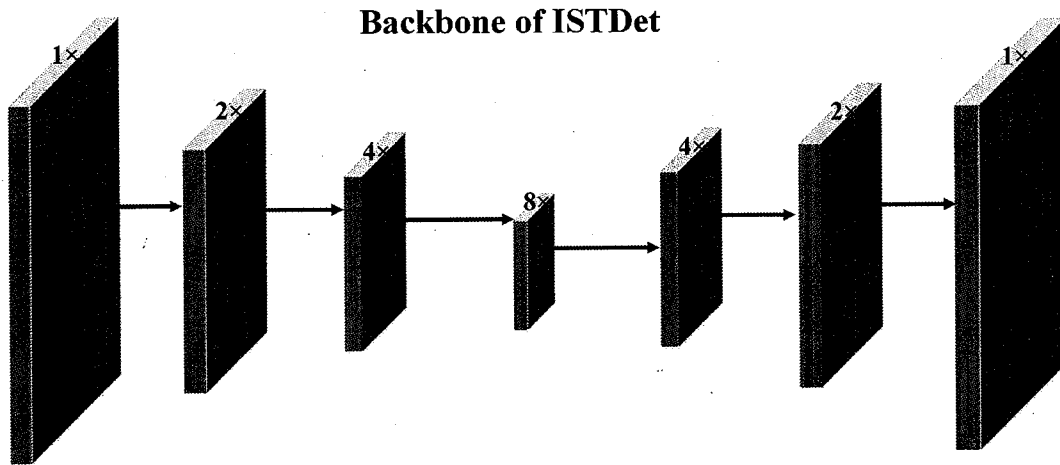


图 5.4 对空红外弱小目标检测网络 ISTDet 的主干

Figure 5.4 The ISTDet backbone for aerial infrared dim and small target detection

5.3.2 特征融合机制

浅层的特征包含目标细粒度的信息，对目标的定位十分重要。深层的特征包含更多的语义信息，对目标的分类非常重要。对于对空红外弱小目标检测网络 ISTDet，将主干网络中具有相同分辨率的深层特征图和浅层特征图进行融合，来提高网络对红外弱小目标的检测能力，特征融合后的主干网络如图 5.5 所示。

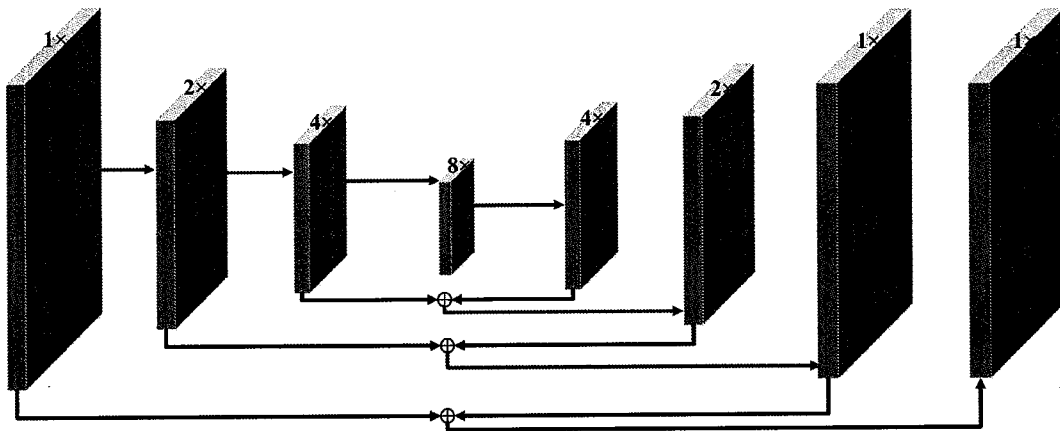


图 5.5 对空红外弱小目标检测网络 ISTDet 中的特征融合

Figure 5.5 Feature fusion of ISTDet for aerial infrared dim and small target detection

网络的详细配置如表 5.4 所示，其中， $Residual(I, O)$ 表示残差单元， I 和 O 分别代表残差单元输入和输出的通道数。

表 5.4 特征融合后 ISTDet 主干网络的配置

Table 5.4 Configuration of ISTDet backbone with feature fusion

Layer	Layer name	Size/Stride	Filters	Output
0	Convolution	3/1	32	256×256
1	Convolution	3/2	64	128×128
2	<i>Residual</i> (64,64)			128×128
3	Convolution	3/2	128	64×64
4	<i>Residual</i> (64,64)			64×64
5	Convolution	3/2	256	32×32
6	<i>Residual</i> (64,64)			32×32
7	Upsample			64×64
8	<i>Residual</i> (64,64)			64×64
9	Add	Layer8+Layer4		64×64
10	Upsample			128×128
11	<i>Residual</i> (64,64)			128×128
12	Add	Layer11+Layer2		128×128
13	Upsample			256×256
14	<i>Residual</i> (64,64)			256×256
15	Add	Layer14+Layer0		256×256

5.3.3 空间注意力机制

在红外弱小目标图像中，目标的信号往往很弱，图像的信噪比低，目标与背景之间的对比度低。本节采用空间注意力机制(Spatial attention mechanism)^[136]来对图像中不同区域之间相关性进行建模，将丰富的上下文信息解码为局部特征，增强网络的表达能力，使红外弱小目标的特征响应得到加强。因此，在对空红外弱小目标检测网络 ISTDet 中，增加了空间注意力机制，空间注意力机制的结构如图 5.6 所示。

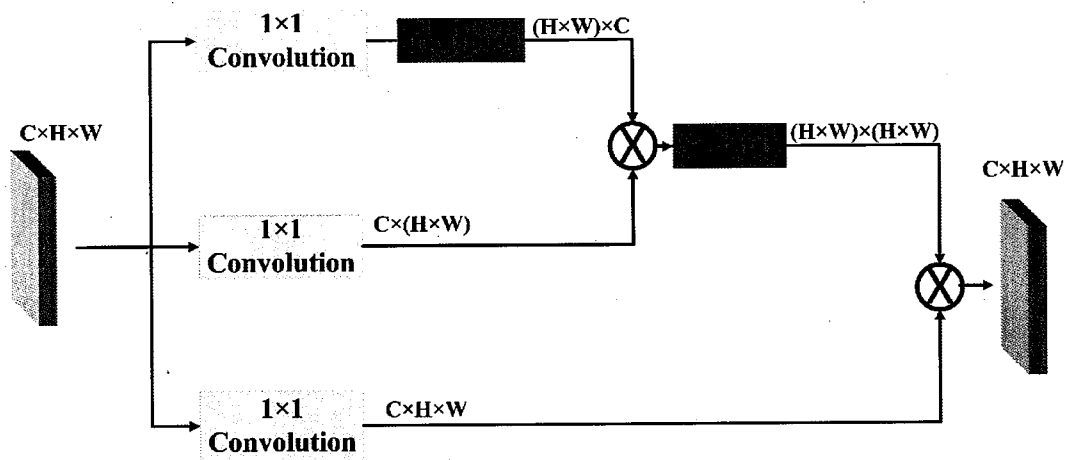


图 5.6 空间注意力机制

Figure 5.6 Spatial attention mechanism

首先,把自适应感受野融合模块输出的特征 $F \in \mathbb{R}^{C \times H \times W}$ 输入到 3 个 1×1 卷积,将第 1 个 1×1 卷积输出的特征经过维度变换和转置变换为 $F \in \mathbb{R}^{(H \times W) \times C}$, 将第 2 个 1×1 卷积输出的特征经过维度变换为 $G \in \mathbb{R}^{C \times (H \times W)}$ 。然后对 F 和 G 进行矩阵乘法,并用 softmax 函数来获得空间注意力机制矩阵 $H \in \mathbb{R}^{(H \times W) \times (H \times W)}$, H 反映了特征图上各个像素点之间的相关性。

$$h_{ji} = \frac{\exp(F_i \cdot G_j)}{\sum_{i=1}^{H \times W} \exp(F_i \cdot G_j)} \quad (5.1)$$

其中, h_{ji} 代表特征图上第 i 个位置与第 j 个位置的相关性,两个位置的特征表达越相似,二者的相关性就越强。

最后,将第 3 个 1×1 卷积输出的特征与 H 进行矩阵乘法和维度变换,得到基于空间注意力机制的输出特征 $F_{out} \in \mathbb{R}^{C \times H \times W}$ 。通过与空间注意力矩阵 H 相乘,使输出特征 F_{out} 不同区域之间的相关性得到增强。

为了说明空间注意力机制的作用,对增加空间注意力机制前后的特征图进行可视化对比,如图 5.7 所示。其中,图 5.7(a)为红外弱小目标图像,红框表示红外弱小目标的位置,图 5.7(b)为未增加空间注意力机制的可视化特征图,图 5.7(c)

为增加空间注意力机制的可视化特征图。由图 5.7 可以看出，增加空间注意力机制的红外弱小目标响应更强，这是由于引入空间注意力机制模块后，不同区域之间的相关性和紧凑性得到增强，有利于红外弱小目标的检测。

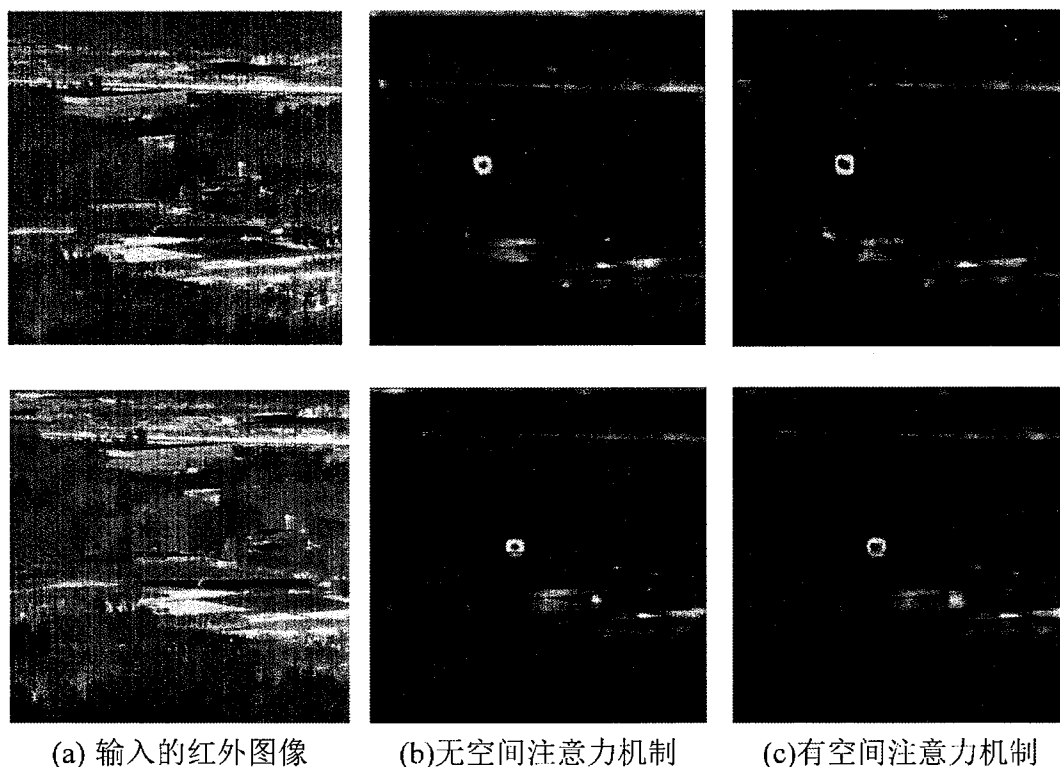


图 5.7 空间注意力机制的作用

Figure 5.7 The effect of spatial attention mechanism

5.3.4 对空红外弱小目标检测网络的结构

对空红外弱小目标检测网络 ISTDet 的结构如图 5.8 所示。根据尺度匹配策略，采用沙漏结构来设计针对红外弱小目标检测的 ISTDet 的主干网络，使红外弱小目标在 1 倍下采样的尺度上检测。将主干网络中具有相同分辨率的深层特征图和浅层特征图进行融合。然后，利用加权感受野融合模块，来增加小目标周围的上下文信息。采用空间注意力机制来增强网络的表达能力，使红外弱小目标的特征响应得到加强。最后，用 1 个 1×1 的卷积对检测结果进行输出。

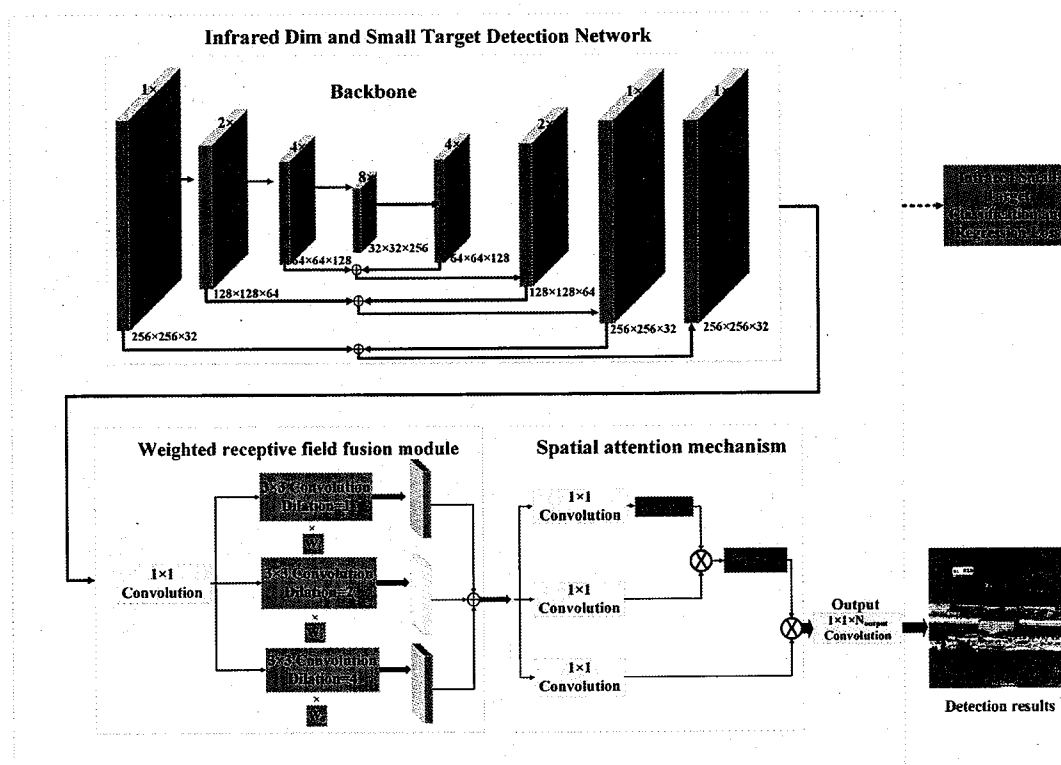


图 5.8 对空红外弱小目标检测网络 ISTDet 的结构图

Figure 5.8 The structure of ISTDet for aerial infrared dim and small target detection

5.4 基于图像过滤的对空红外弱小目标检测网络结构

在本节中，主要是将图像过滤模块应用到对空红外弱小目标检测任务中。因此，将 5.2 节的图像过滤模块与 5.3 节对空红外弱小目标检测网络 ISTDet 相结合，建立基于图像过滤的对空红外弱小目标检测网络，命名为 FilterDet。FilterDet 的结构如图 5.9 所示。

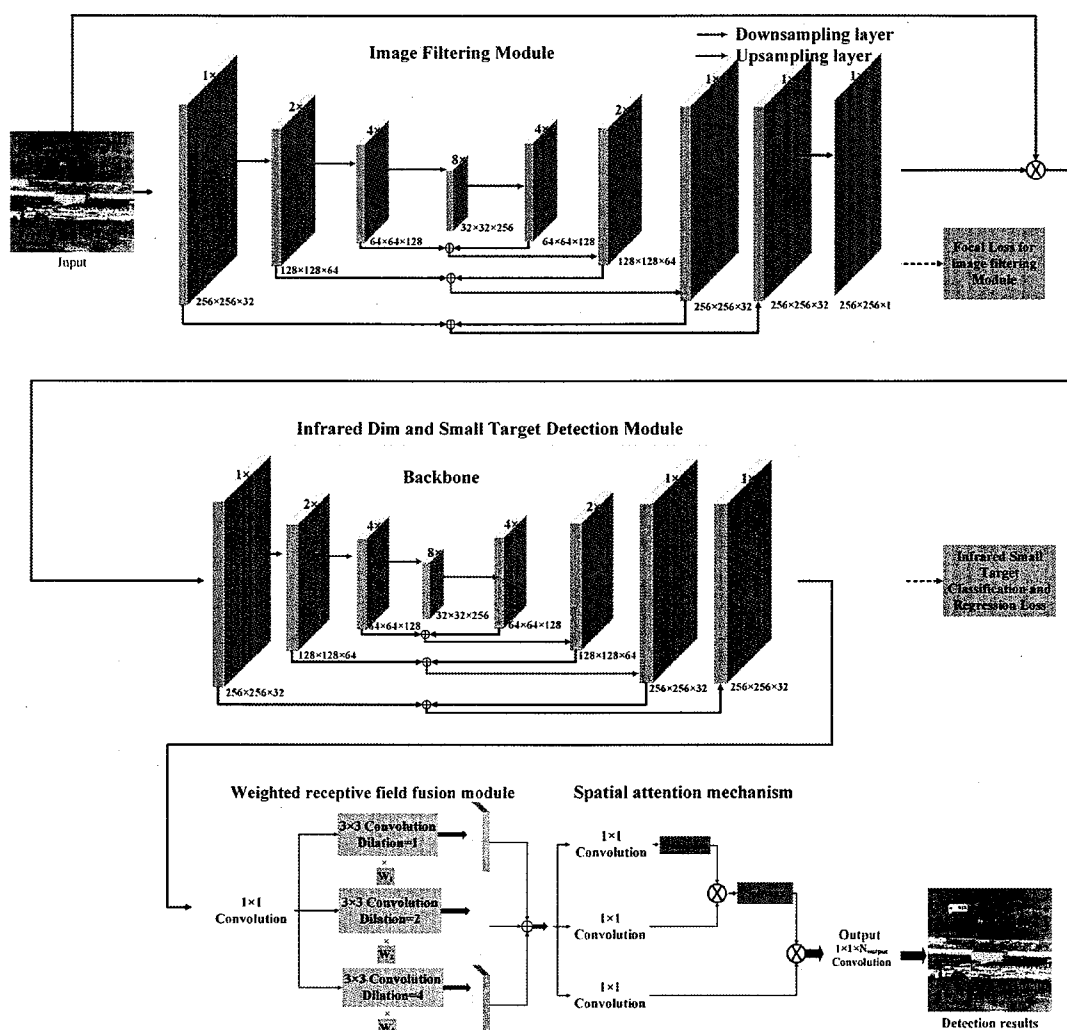


图 5.9 FilterDet 的结构

Figure 5.9 The structure of FilterDet

如图 5.9 所示, FilterDet 主要由 2 部分构成, 图像过滤模块(Image Filtering Module, IFM)和红外弱小目标检测模块(Infrared Dim and Small Target Detection Module, IDSTDM)。图像过滤模块和红外弱小目标检测模块是 FilterDet 内部相连的两个模块。其中, 红外弱小目标检测模块采用 5.3 节中的红外弱小目标检测网络 ISTDet。目标的检测流程如下: 将包含红外弱小目标的图像作为输入, 送到图像过滤模块中, 完成对红外图像的过滤。然后, 将经过过滤的图像作为输入送到红外弱小目标检测模块中, 完成目标位置和类别的预测。最后, 采用多任务损失函数来实现 FilterDet 网络端到端的训练。

5.5 多任务损失函数

FilterDet 主要由图像过滤模块和红外弱小目标检测模块构成。因此, FilterDet 的损失函数也由两部分组成, 即图像过滤损失和红外弱小目标检测损失。

对于图像过滤损失, 采用如式 5.2 所示的 Focal loss:

$$Loss_{IFM} = \begin{cases} -\alpha(1-y_p)^\gamma \times \log y_p, & y_{GT}=1 \\ -(1-\alpha)y_p^\gamma \times \log(1-y_p), & y_{GT}=0 \end{cases} \quad (5.2)$$

式中: y_{GT} 表示置信度图的真实值, y_p 表示预测的置信度图, α 和 γ 分别设为 0.25 和 2。

经过图像过滤模块之后, 将输入图像与置信度图进行点乘, 输出过滤图像。然后将过滤图像送到红外弱小目标检测模块中用来预测对空红外弱小目标的位置信息和类别信息。红外弱小目标检测模块的损失函数与 GSTDet 网络的相同, 即利用 GIOU loss 对目标的位置进行回归, 用 Focal loss 来对目标的置信度进行回归, 用二进制交叉熵来对目标的类别进行回归, 用 $Loss_{IDSTD}$ 来表示红外弱小目标检测模块的损失, 如公式 5.3 所示:

$$Loss_{IDSTD} = Loss_{GIOU} + Loss_{conf} + Loss_{cls} \quad (5.3)$$

则 FilterDet 的总 loss 可以表示为

$$Loss_{FilterDet} = Loss_{IDSTD} + Loss_{IFM} \quad (5.4)$$

5.6 实验与分析

为了评估 FilterDet 的检测性能, 首先, 对 FilterDet 进行消融实验, 验证空间注意力机制和图像过滤模块的有效性。然后, 分别在 3 个真实的红外弱小目标序列和单帧红外弱小目标数据集上验证红外弱小目标检测网络 FilterDet 对红外弱小目标的检测性能。

5.6.1 实验条件与评价方法

实验环境和训练细节: 所有的模型都采用随机梯度下降 (stochastic gradient descent, SGD) 法在 1 块 NVIDIA GeForce GTX TITAN X 显卡上进行训练, 深度学习框架采用 Pytorch 0.4.1。采用余弦学习率策略来训练各检测模型, 权重衰减

和动量分别设置为 0.0005 和 0.9。

在 3 个红外弱小目标序列和单帧红外弱小目标数据集上对所有的模型均训练 80 个轮次(epoch)。为了使训练过程更加稳定,在训练最初的 4 个轮次采用“warm-up”策略,从第 4 个轮次到第 100 个轮次采用余弦学习率策略使学习率从 0.0001 下降到 0.000001。

评价指标:采用平均精准度(Average Precision, AP)来测量每一类目标的检测精度。利用每秒帧数(Frame per second, FPS)来衡量目标检测器的检测速度。

数据集介绍:本文所使用的红外弱小目标数据集是来自 Hui 等人^[119]通过外场实地拍摄和数据准备加工提供的一套以一架和多架固定机翼无人机目标为探测对象的红外弱小目标数据集。本文选取 3 个具有代表性的红外弱小目标序列来测试所提出的红外弱小目标检测网络的性能。这 3 个红外弱小目标序列的样本如图 5.10 所示,红色框代表目标的位置,详细信息如表 5.5 所示。

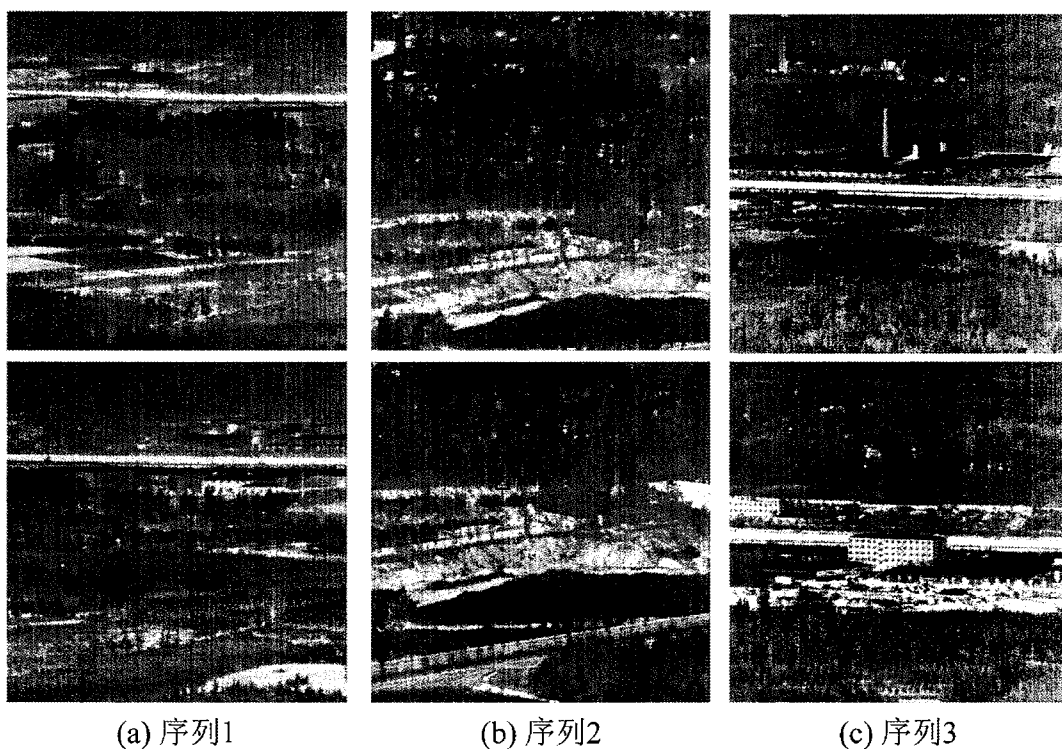


图 5.10 真实的红外弱小目标序列。(a) 序列 1 (b) 序列 2 (c) 序列 3

Figure 5.10 Real infrared dim and small target sequences. (a) Sequence 1. (b) Sequence 2.

(c) Sequence 3

表 5.5 3 个红外弱小目标序列的详细信息

Table 5.5 Detailed information of three infrared dim and small target sequences

	Image size	Frame	The details of the targets and background
Sequence1	256×256	3000	A long imaging distance, Ground background, Long time
Sequence2	256×256	763	Target from near to far, Dim and small target, Ground background
Sequence3	256×256	751	Maneuvering target, Ground background

为了进一步证明 FilterDet 对红外弱小目标的检测能力,本节也在单帧红外弱小目标数据集上进行了实验。单帧红外弱小目标数据集是在数百张不包含弱小目标的背景图像上,加入高斯信号构成的仿真数据集。首先在每张背景图像上生成坐标点,然后将二维高斯信号作为红外小目标样本添加到背景图像中^[137]。二维高斯信号用式 5.5 来描述:

$$s(i, j) = s_E \cdot \exp\left(-\frac{1}{2} \left[\frac{(i-x_0)^2}{\sigma_x^2} + \frac{(j-y_0)^2}{\sigma_y^2} \right]\right) \quad (5.5)$$

其中, (x_0, y_0) 代表红外弱小目标的中心在红外图像上的坐标, $s(i, j)$ 为红外弱小目标在位置 (i, j) 的像素值。 s_E 代表红外弱小目标的强度, 值为介于 0 到 1 之间的随机数, σ_x 和 σ_y 是高斯分布标准差。

合成的单帧红外弱小目标图像如图 5.11 所示, 红框代表红外弱小目标的位置。

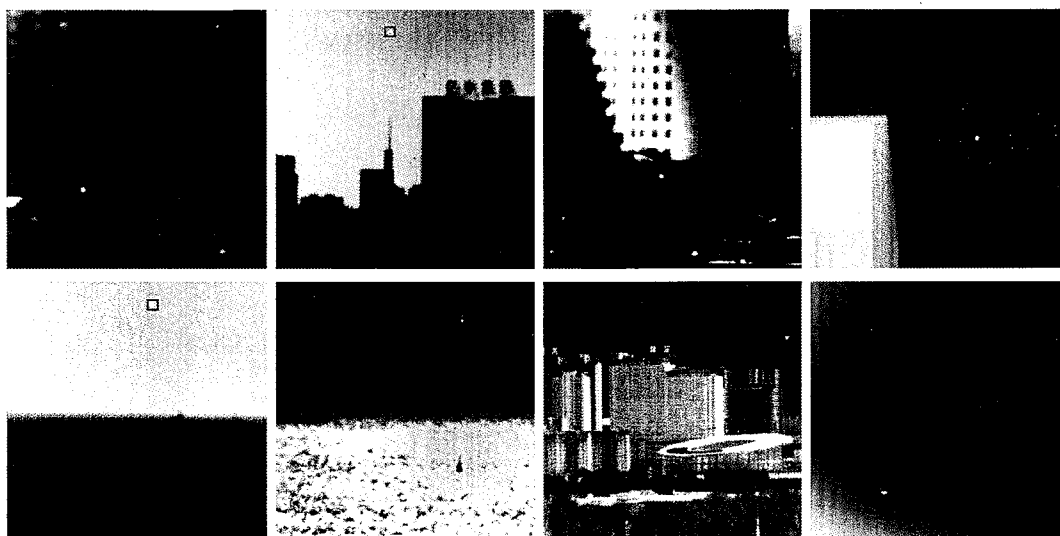


图 5.11 单帧红外弱小目标图像

Figure 5.11 Single frame infrared image set

5.6.2 消融实验

FilterDet 是由图像过滤模块与红外弱小目标检测模块构成的。为了验证尺度匹配策略、加权感受野融合模块、正样本选择策略、空间注意力机制和图像过滤模块对红外弱小目标检测的影响,在上述的 3 个真实红外弱小目标序列上进行消融实验。在训练和检测阶段,所有的图片都以 256×256 的分辨率输入到网络。实验结果如表 5.6 所示。表中的 Baseline 代表由沙漏结构设计的红外弱小目标检测网络 ISTDet 的主干网络, PASS 代表 Positive anchors selection strategy, WRFFM 代表 Weighted receptive field fusion module, SAM 代表 Spatial attention mechanism, IFM 代表 Image filtering module。

表 5.6 在 3 个序列上的消融实验

Table 5.6 Ablation study on 3 sequences

Baseline	PASS	WRFFM	SAM	IFM	Sequence 1		Sequence 2		Sequence 3	
					AP(%)	FPS	AP(%)	FPS	AP(%)	FPS
√					94.32	56.31	82.53	55.24	91.60	55.37
√	√				94.74	56.98	84.58	54.76	92.29	55.11
√	√	√			96.26	47.49	84.83	45.64	93.86	46.31
√	√	√	√		96.58	41.66	85.96	40.75	94.59	40.63
√	√	√	√	√	98.09	32.05	86.90	32.89	95.38	31.25

Baseline:如表 5.6 所示, Baseline 在序列 1 上获得了 94.32%的 AP 值, 检测速度为 56.31FPS, 在序列 2 上获得了 82.53%的 AP 值, 检测速度为 55.24FPS, 在序列 3 上获得了 91.60%的 AP 值, 检测速度为 55.37FPS。

正样本选择策略: 与 Baseline 相比, 在序列 1 上, AP 值由 94.32%提高到 94.74%。在序列 2 上, AP 值由 82.53%提高到 84.58%。在序列 3 上, AP 值由 91.60%提高到 92.29%。这是由于正样本选择策略可以通过增加正样本数量来缓解小目标检测中正负样本数量不均衡的问题。

加权感受野融合模块: 通过增加加权感受野融合模块, 在序列 1 上, AP 值由 94.74%提高到 96.26%。在序列 2 上, AP 值由 84.58%提高到 84.83%。在序列 3 上, AP 值由 92.29%提高到 93.86%。通过采用加权感受野融合模块, 可以自适应的融合具有不同感受野的特征, 增加小目标周围的上下文信息。

空间注意力机制: 通过引入空间注意力机制, 在序列 1 上, AP 值由 96.26%提高到 96.58%。在序列 2 上, AP 值由 84.83%提高到 85.96%。在序列 3 上, AP 值由 93.86%提高到 94.59%。正如 5.3.3 节讨论的, 空间注意力机制可以增强网络的表达能力, 使红外弱小目标的特征响应得到加强。

图像过滤模块: 很明显, 融合了图像过滤模块, 在序列 1、序列 2 和序列 3 上 AP 值分别提高了 1.51、0.94 和 0.79 个百分点。虽然检测速度下降了, 但仍然保持在 31FPS, 可以满足红外弱小目标的实时检测。图像过滤模块可以增强目标响应, 并抑制背景的干扰, 有助于提高网络对红外弱小目标的检测能力。

5.6.3 合成的单帧红外弱小目标数据集上的验证

表 5.7 为 ISTDet 在合成的单帧红外弱小目标数据集上的检测结果, 其中 YOLO V3、RFBnet 和 RefineDet^[138] 是基于 CNN 目标检测算法中的代表, 对不同尺度的目标都有不错的检测效果。并且, YOLO V3、RFBnet 和 RefineDet 都具有专门检测小目标的尺度。YOLO V3 借鉴了特征金字塔(Feature Pyramid Network, FPN)的结构。RFBnet 是一个快速且高效的目标检测器, 它模仿了人类视觉系统中感受野的结构。作为单阶段目标检测算法, RefineDet 同时融合了单阶段目标检测算法和双阶段目标检测算法的优势。表中 Input 为输入图像的分辨率, FilterDet* 代表网络的输入为未经基于局部对比度测量方法处理的红外图像, FilterDet 代表网络的输入为经过基于局部对比度测量方法处理的红外图像。

表 5.7 单帧红外弱小目标数据集上实验结果

Table 5.7 The experimental results for single frame infrared image set

Detector	YOLO V3	RFBnet	RefineDet	FilterDet*	FilterDet
Input	256×256	300×300	320×320	256×256	256×256
AP	81.61%	83.58%	84.91%	89.79%	91.64%
FPS	63.86	71.87	45.33	31.15	31.46

如表 5.7 所示, FilterDet 取得了 89.79% 的 AP 值, 检测速度保持在 31.15FPS。与 YOLO V3、RFBNet 和 RefineDet 相比, AP 值分别提高了 8.18、6.21 和 4.88 个百分点。经过基于局部对比度测量方法处理之后, 目标与背景之间的对比度得到增强, 使 AP 值进一步提高到 91.64%。检测速度保持在 31.46FPS, 可以实现对红外弱小目标的实时检测。

另外, FilterDet 与其它算法在单帧红外弱小目标数据集上的 PR 曲线对比如图 5.12 所示。结果表明, FilterDet 具有更好的红外弱小目标检测能力。

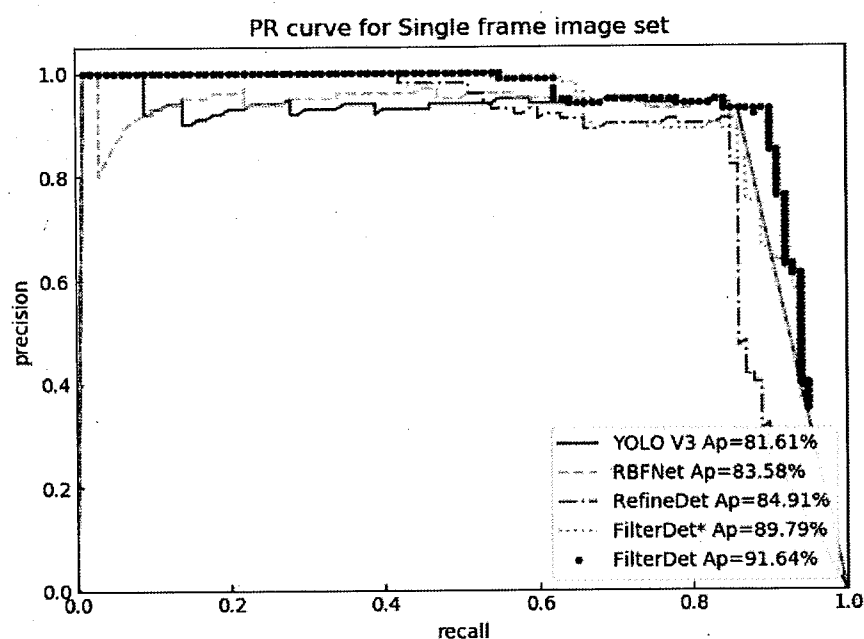


图 5.12 单帧红外弱小目标数据集上 PR 曲线对比

Figure 5.12 The comparison of PR curve on Single frame infrared image set

为了对 FilterDet 进行定性评估, 将 FilterDet 在单帧红外弱小目标数据集上的检测结果进行可视化, 如图 5.13 所示。可视化结果表明, FilterDet 可以有效的检测出不同背景下的红外弱小目标。

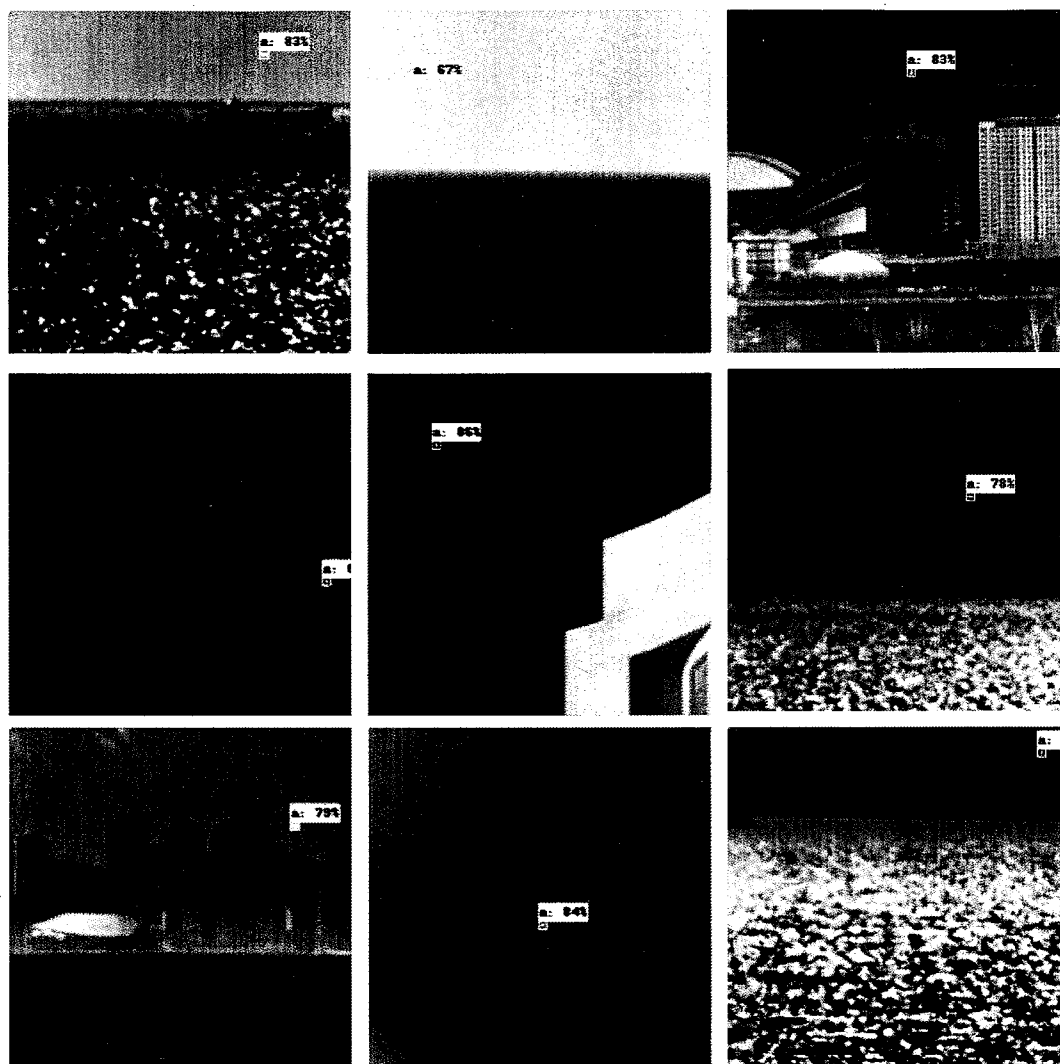


图 5.13 FilterDet 在单帧红外图像上的检测结果

Figure 5.13 The detection results of FilterDet in Single frame infrared image set

5.6.4 真实红外弱小目标序列上的验证

为了进一步证明 FilterDet 对红外弱小目标的检测能力,在 3 个真实的红外弱小目标序列上进行了对比实验。表 5.8、表 5.9 和表 5.10 分别为红外弱小目标检测网络在序列 1、序列 2 和序列 3 上的检测结果。

表 5.8 序列 1 实验结果

Table 5.8 The experimental results for sequence 1

Detector	YOLO V3	RFBnet	RefineDet	FilterDet*	FilterDet
Input	256	300	320	256	256
AP	72.51%	76.34%	88.06%	98.09%	98.79%
FPS	63.86	71.87	45.24	32.05	31.19

表 5.9 序列 2 实验结果

Table 5.9 The experimental results for sequence 2

Detector	YOLO V3	RFBnet	RefineDet	FilterDet*	FilterDet
Input	256	300	320	256	256
AP	64.55%	62.20%	74.24%	86.90%	90.27%
FPS	63.38	71.43	44.51	32.89	32.31

表 5.10 序列 3 实验结果

Table 5.10 The experimental results for sequence 3

Detector	YOLO V3	RFBnet	RefineDet	FilterDet*	FilterDet
Input	256	300	320	256	256
AP	76.66%	78.37%	82.03%	95.38%	97.80%
FPS	62.91	71.16	45.69	31.25	32.04

在序列 1 中, FilterDet 取得了 98.09% 的 AP 值, 检测速度仍然保持在 32.05FPS。与 YOLO V3、RFBNet 和 RefineDet 相比, AP 值分别提高了 25.58、21.75 和 10.03 个百分点。经过基于局部对比度测量方法处理之后, AP 值进一步提高到 98.79%。检测速度保持在 31.19FPS。

在序列 2 中, FilterDet 取得了 86.90% 的 AP 值, 检测速度仍然保持在 32.89FPS。与 YOLO V3、RFBNet 和 RefineDet 相比, AP 值分别提高了 22.35、24.70 和 12.66 个百分点。经过基于局部对比度测量方法处理之后, AP 值进一步提高到 90.27%。检测速度保持在 32.31FPS。

在序列 3 中, FilterDet 取得了 95.38% 的 AP 值, 检测速度仍然保持在 31.25FPS。与 YOLO V3、RFBNet 和 RefineDet 相比, AP 值分别提高了 18.72、17.01 和 13.35 个百分点。经过基于局部对比度测量方法处理之后, AP 值进一步提高到 97.80%。检测速度保持在 32.04FPS。

FilterDet 在序列 1 上取得了最高的 AP 值, 因为序列 1 是一个长序列, 共有 3000 张样本图像, 网络可以得到充分的训练。红外弱小目标检测网络在序列 2 上的 AP 值较序列 1 和 3 低, 这是由于序列 2 的目标小且信号特别微弱, 并且小目标呈现出由远到近的变化过程。实验结果表明通过引入空间注意力机制, 可以增强不同区域之间的相关性和紧凑性, 使网络对红外弱小目标的检测性能进一步提高。本章提出的红外弱小目标检测网络无论在精度和速度方面都要高于 YOLO V3、RFBnet 和 RefineDet。实验结果证明了 FilterDet 对红外弱小目标具有较好的检测性能, 并且适合应用于实时检测的场合。

另外, FilterDet 与其它算法在 3 个真实红外弱小目标序列上的 PR 曲线对比如图 5.14 所示。结果表明, FilterDet 具有更好的红外弱小目标检测能力。

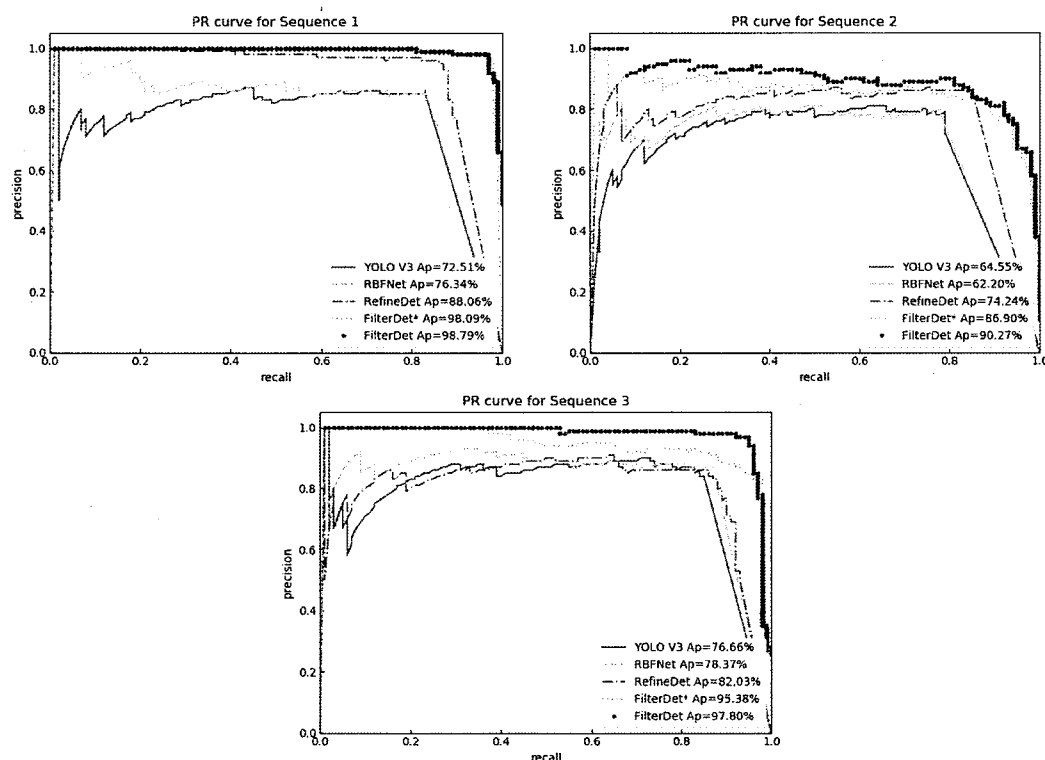


图 5.14 在 3 个红外弱小目标序列上的 PR 曲线对比

Figure 5.14 The comparison of PR curve on three infrared dim and small target sequences

为了对 FilterDet 进行定性评估, 将 FilterDet 在 3 个序列上的检测结果进行可视化, 并对红外弱小目标处的局部区域进行放大, 如图 5.15、5.16 和 5.17 所示。

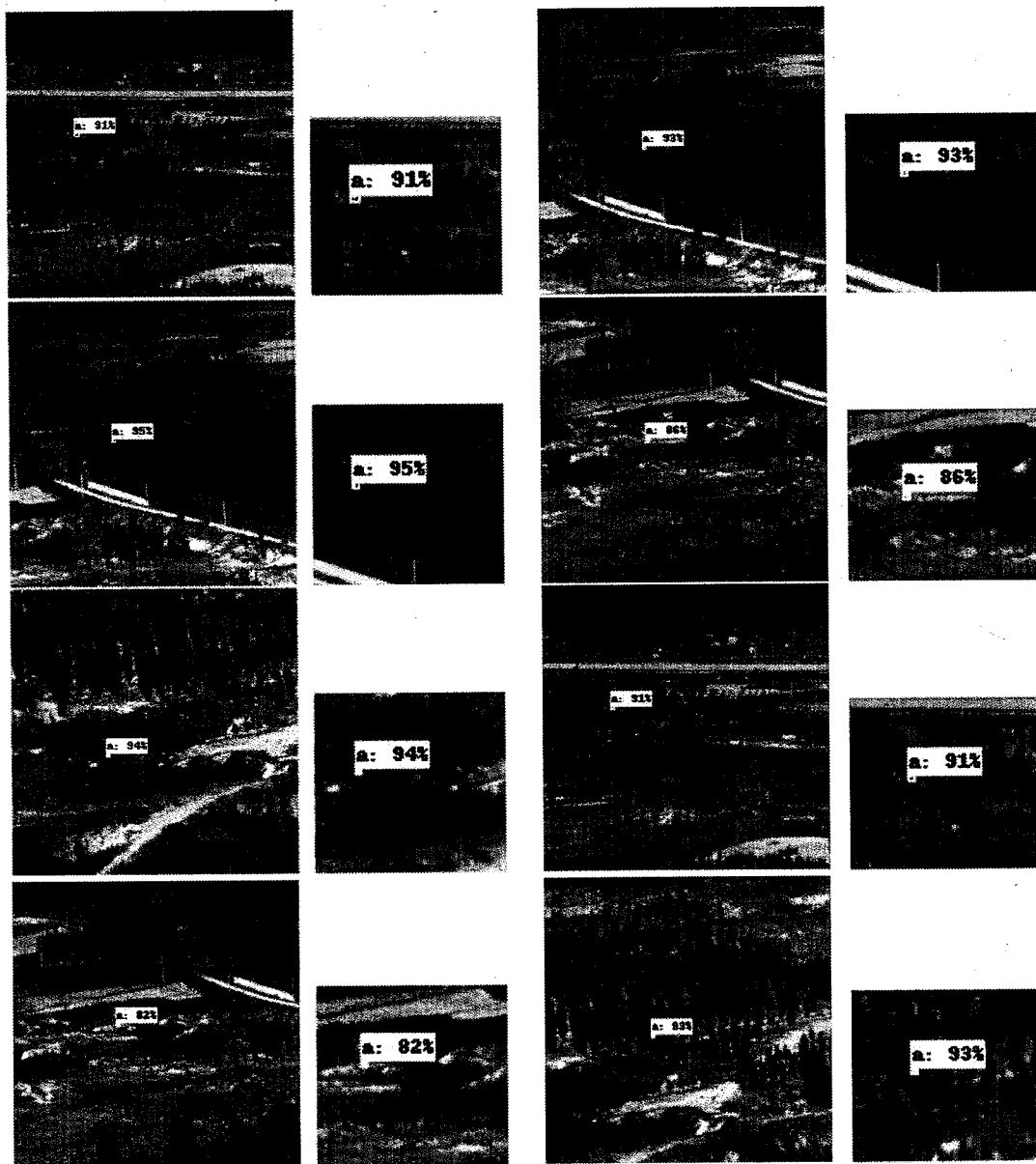


图 5.15 FilterDet 在序列 1 上的可视化检测结果

Figure 5.15 Visual detection results of FilterDet on Sequence 1

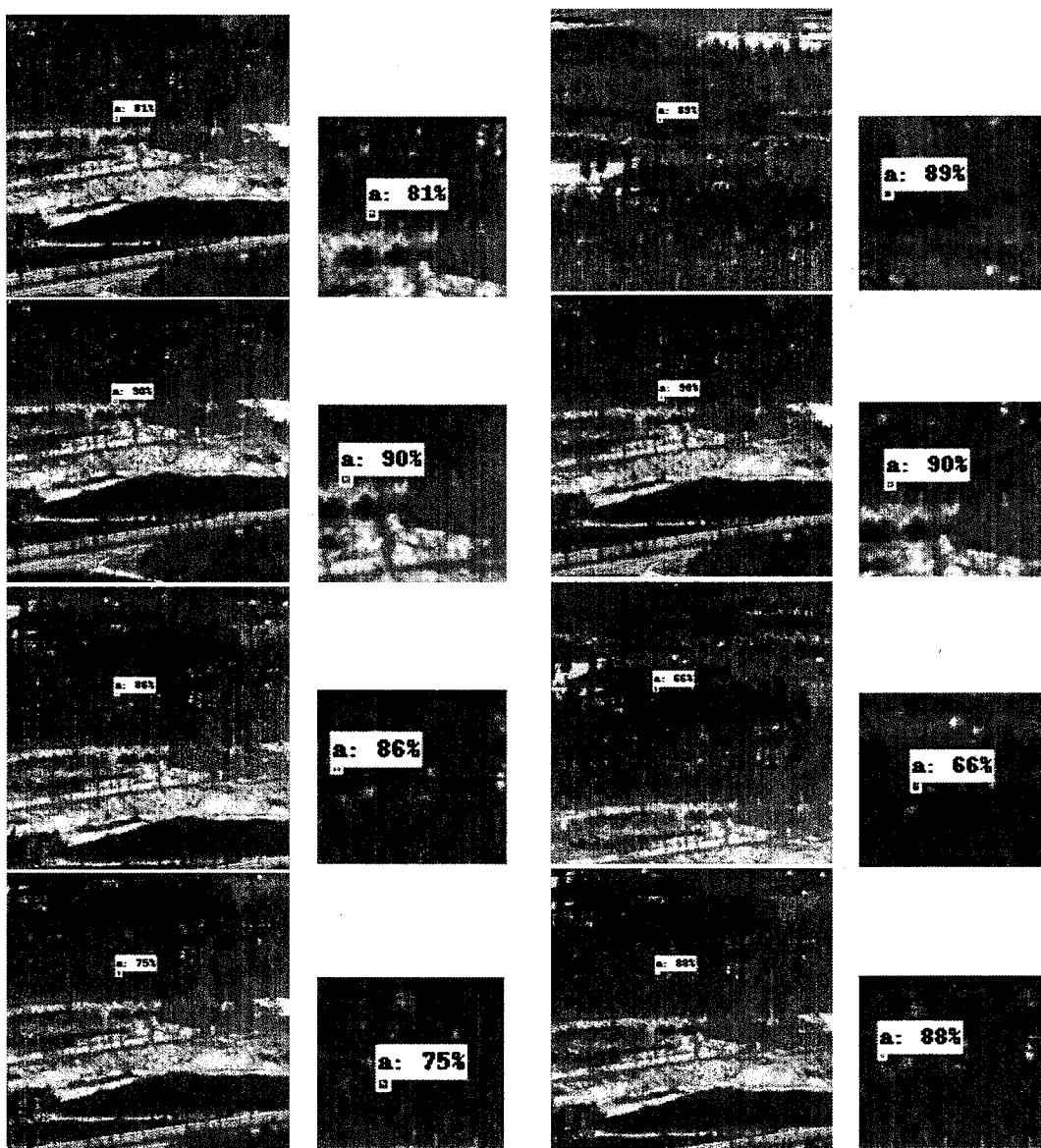


图 5.16 FilterDet 在序列 2 上的可视化检测结果

Figure 5.16 Visual detection results of FilterDet on Sequence 2

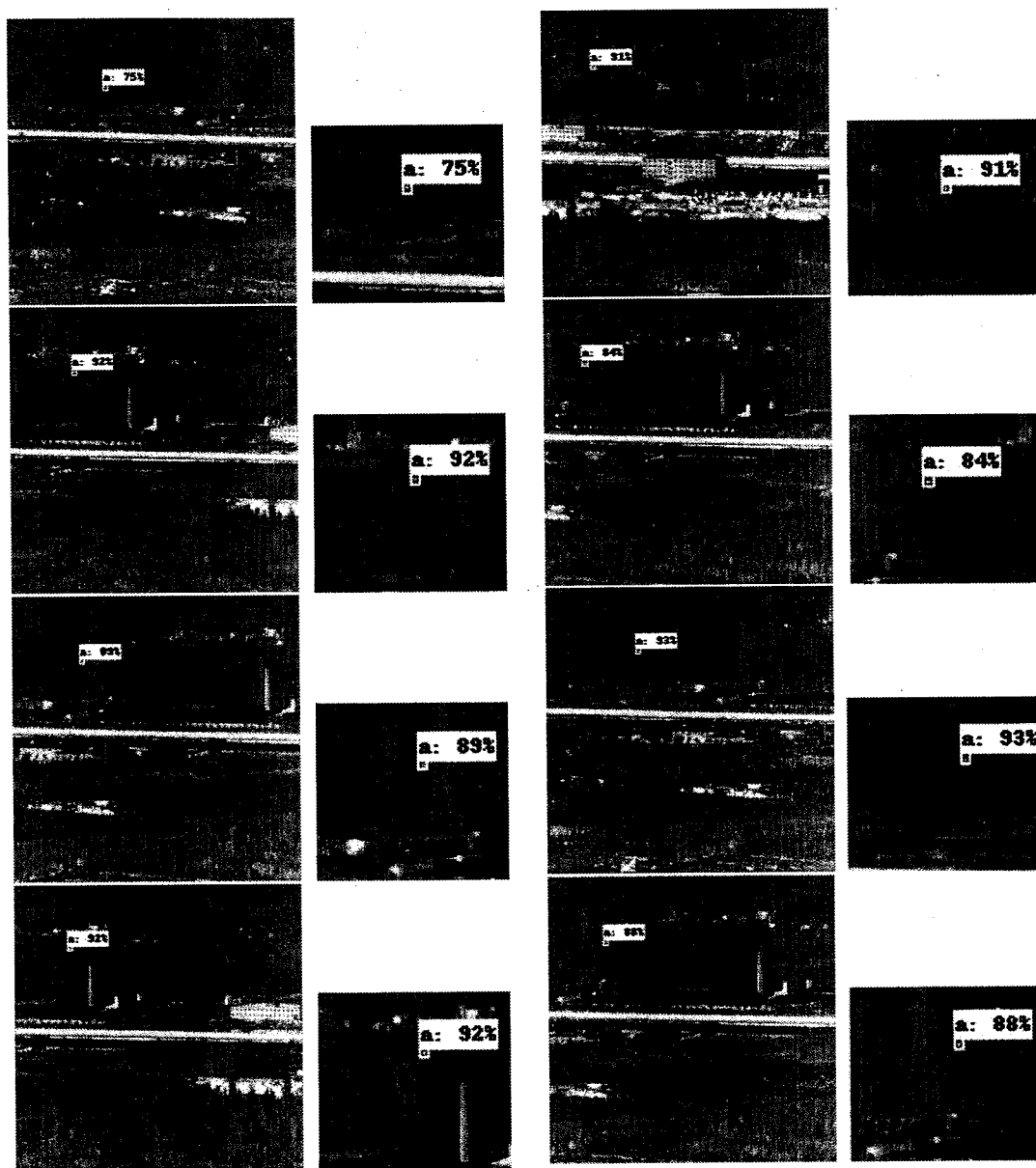


图 5.17 FilterDet 在序列 3 上的可视化检测结果

Figure 5.17 Visual detection results of FilterDet on Sequence 3

图 5.15、图 5.16 和图 5.17 分别显示了红外弱小目标检测网络在序列 1、序列 2 和序列 3 上的检测结果。实验结果表明，FilterDet 可以有效的检测出不同复杂背景下的红外弱小目标。

5.7 本章小结

红外弱小目标检测可以在对空作战中用于侦察目标，锁定目标。然而，红外

弱小目标不仅信号强度弱,并且常常受到环境中杂波和噪声的干扰,使得红外弱小目标检测成为了目标检测领域中的一个挑战。本章设计了一个用于对空红外弱小目标的检测的卷积神经网络 FilterDet,该网络是由图像过滤模块和红外弱小目标检测模块构成的。图像过滤模块用来增强目标响应,过滤背景噪声,减少虚警和错检。红外弱小目标检测模块则是由 ISTDet 构成的,将过滤后的图像输入到红外弱小目标检测模块中,完成最后红外弱小目标的检测。采用多任务损失函数来实现 ISTDet 端到端的训练。为了验证 FilterDet 对红外弱小目标的检测性能,分别在 3 个真实的红外弱小目标序列和单帧红外弱小目标数据集上进行了对比实验。实验结果表明,结合图像过滤模块的 FilterDet 具有较为出色的对空红外弱小目标检测能力。

第6章 总结与展望

6.1 本文工作总结

对地小目标检测和对空红外弱小目标检测都可用于监视远距离目标,有助于实现“先敌发现,先敌锁定和先敌命中”,对国防事业具有重要的应用价值。然而,对地小目标和对空红外弱小目标在图像中只占有很少的一部分,特征不明显。因此,对地小目标检和对空红外弱小目标检测是一个具有显著应用价值且具有挑战性的研究课题。随着计算机技术的高速发展,基于深度学习的目标检测算法成为目标检测领域中主流的方法。基于深度学习的目标检测算法可以将特征提取、特征融合、目标定位和分类等多个任务融合到一个网络中,实现对目标进行端到端的检测。本文分别对基于深度学习的对地小目标和红外弱小目标检测算法开展了研究,本文的主要贡献和创新点如下:

1. 分析并总结了影响小目标检测的关键因素,包括:分辨率、目标周围的上下文信息、尺度的选择、特征的融合、正负样本数量不均衡和目标信号弱、背景掺杂噪声等。根据这6项影响小目标检测的因素,提出了相应的解决策略,为对地小目标检测算法和对空红外弱小目标检测算法的设计奠定基础。

2. 在对地小目标图像预处理方面,提出了一种基于图像裁剪的对地小目标图像预处理方法。利用该方法,可以在不增加显存的条件下,实现网络对同一目标多尺度的训练,使网络变得更加鲁棒。另外,针对红外弱小目标信号“弱”、“小”且常常伴随着噪声干扰的特点,提出了基于局部对比度的红外弱小目标图像预处理方法。该方法可以利用简单的卷积操作和最大池化操作来实现,可以快速、有效地提高目标与背景之间的对比度。

3. 提出了一种基于超分辨率重建的对地小目标检测网络 SuperDet。SuperDet 通过结合超分辨率重建技术和目标检测技术来提高对地小目标检测性能。SuperDet 主要由两部分组成,分别是超分辨率模块和对地小目标检测模块。其中,超分辨率模块实现了低分辨率图像到高分辨率图像的转换。对地小目标检测模块完成对对地小目标位置和类别的预测。为了选择合适的尺度来检测对地小目标,

利用尺度匹配策略为对地小目标设计主干网络。随后提出了加权感受野融合模块,通过自适应的融合具有不同感受野的特征来增加对地小目标周围的上下文信息。另外,针对对地小目标检测中正负样本数量不均衡的问题,提出了正样本选择策略。最后,分别采用 GIOU loss、Focal loss 和 Binary Cross Entropy loss 来对对地小目标的位置、置信度和类别进行回归。SuperDet 将超分辨率重建任务和对地小目标检测任务融合到了一个网络中,实现了对对地小目标端到端的检测。

4. 提出了一种基于图像过滤的对空红外弱小目标检测网络 FilterDet。FilterDet 主要由两部分组成,分别是过滤模块和红外弱小目标检测模块。其中,过滤模块实现了对背景的抑制和对红外弱小目标的增强,过滤掉了图像中的噪声和干扰。红外弱小目标检测模块将过滤后的图像作为输入,完成对红外弱小目标位置的预测。根据尺度匹配策略,采用沙漏结构来设计红外弱小目标检测模块的主干网络。另外,采用加权感受野融合模块来增加小目标周围的上下文信息。利用空间注意力机制来增强网络的表达能力,使红外弱小目标的特征响应得到加强。FilterDet 将图像过滤任务和红外弱小目标检测任务融合到了一个网络中,实现了对红外弱小目标端到端的检测,具有较为出色的对空红外弱小目标检测性能。

6.2 未来展望

虽然本文对基于深度学习的对地小目标检测算法和基于深度学习的对空红外弱小目标检测算法开展了较为深入的研究,分别提出了用于对地小目标检测的网络 SuperDet 和用于对空红外弱小目标检测的网络 FilterDet,取得了较为理想的效果,但仍然存在着可改进的空间。因此,在未来的工作中,将针对以下问题展开更为深入的研究和探索:

1. 目前,基于深度学习的对地小目标检测算法和基于深度学习的对空红外弱小目标检测算法都依赖于数据集。设计基于弱监督或者无监督的对地小目标检测算法和对空红外弱小目标检测算法,可以利用少量数据来训练网络,实现对目标的检测。

2. 在对地小目标检测方面,本文提出将超分辨率重建技术与目标检测技术相结合。但是,要实现从低分辨率图像到高分辨率图像的重建需要依赖于包含高

分辨率图像的数据集。提出基于无监督的超分辨重建算法,可以减少对高分辨率图像的依赖。

3. 在对空红外弱小目标检测方面开展基于关键点的检测方法。红外弱小目标信号弱,环境中掺杂噪声,通常以“亮点”的方式呈现。因此,将红外弱小目标检测任务作为关键点检测任务,可实现基于无锚框(anchor free)、端到端的红外弱小目标检测。

4. 在基于局部对比度的红外弱小目标预处理方法中,只是对于中心大小为 3×3 的区域进行对比度运算。针对尺度变化的红外弱小目标,可以采用多尺度策略来实现红外弱小目标的检测。