

结合强化学习自适应候选框挑选的SAR目标检测方法

杜 兰^{*①} 王梓霖^① 郭昱辰^② 杜宇昂^① 严俊坤^①

^①(西安电子科技大学雷达信号处理国家重点实验室 西安 710071)

^②(西安电子科技大学前沿交叉研究院 西安 710071)

摘要: 大场景合成孔径雷达(SAR)图像相对于通用光学图像,复杂背景杂波对目标特征提取影响更大,由于传统基于候选框的深度目标检测算法会在整张特征图上产生大量冗余候选框,因而在SAR图像复杂背景杂波影响下会产生大量的虚警,降低目标检测精度。针对该问题,该文基于Faster R-CNN检测模型,提出结合强化学习自适应候选框挑选的SAR目标检测方法。该方法能够通过强化学习自适应搜索特征图中可能含有目标的区域,并挑选搜索区域内的候选框继续进行分类、回归。通过准确搜索到含有目标的区域,可以减少复杂背景杂波的影响并减少传统强化学习应用于检测问题的计算量。所提方法利用强化学习序列决策的特点,能够根据图像信息通过强化学习迭代搜索自适应确定图像中可能含有目标的搜索区域的位置。同时,该方法通过在强化学习中使用距离约束,可以根据之前的搜索结果自适应调整下一次搜索区域的尺寸。基于实测数据的实验结果表明,所提方法能够提升传统深度学习目标检测方法的检测性能。

关键词: 合成孔径雷达; 目标检测; 强化学习; Faster R-CNN算法

中图分类号: TN957.51

文献标识码: A

文章编号: 2095-283X(2022)05-0884-13

DOI: 10.12000/JR22121

引用格式: 杜兰, 王梓霖, 郭昱辰, 等. 结合强化学习自适应候选框挑选的SAR目标检测方法[J]. 雷达学报, 2022, 11(5): 884-896. doi: 10.12000/JR22121.

Reference format: DU Lan, WANG Zilin, GUO Yuchen, *et al.* Adaptive region proposal selection for SAR target detection using reinforcement learning[J]. *Journal of Radars*, 2022, 11(5): 884-896. doi: 10.12000/JR22121.

Adaptive Region Proposal Selection for SAR Target Detection Using Reinforcement Learning

DU Lan^{*①} WANG Zilin^① GUO Yuchen^② DU Yuang^① YAN Junkun^①

^①(National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China)

^②(Academy of Advanced Interdisciplinary Research, Xidian University, Xi'an 710071, China)

Abstract: Compared with optical images, the background clutter has a greater impact on feature extraction in Synthetic Aperture Radar (SAR) images. Due to the traditional redundant region proposals on the entire feature map, these algorithms generate large quantities of false alarms under the influence of clutter in SAR images, thereby lowering the target detection accuracy. To address this issue, this study proposes a Faster R-CNN model-based SAR target detection method, which uses reinforcement learning to realize adaptive region proposal selection. This method can adaptively locate areas that may contain targets on the feature map using the sequential decision-making characteristic of reinforcement learning and simultaneously adjust the scope of the next search area according to previous search results using distance constraints in reinforcement learning. Thus, this method can reduce the impact of complex background clutter and the computation of reinforcement learning. The experimental results based on the measured data indicate that the proposed method improves the detection performance.

Key words: Synthetic Aperture Radar (SAR); Target detection; Reinforcement learning; Faster R-CNN

收稿日期: 2022-06-22; 改回日期: 2022-08-24; 网络出版: 2022-09-02

*通信作者: 杜兰 dulan@mail.xidian.edu.cn *Corresponding Author: DU Lan, dulan@mail.xidian.edu.cn

基金项目: 国家自然科学基金(U21B2039)

Foundation Item: The National Natural Science Foundation of China (U21B2039)

责任编辑: 徐丰 Corresponding Editor: XU Feng

1 引言

合成孔径雷达(Synthetic Aperture Radar, SAR)可对地面区域进行大面积的主动微波成像,能够实现全天时、全天候对地观测,在军事领域和民用领域得到广泛应用。近年来,随着SAR系统和成像算法逐渐成熟,越来越多高质量的SAR图像出现,如何对SAR图像进行准确高效的目标检测是现在研究热点之一。

目前传统的SAR目标检测方法大都围绕经典的恒虚警率(Constant False Alarm Rate, CFAR)算法展开研究,这是一种基于灰度特征的传统SAR目标检测方法。其中双参数CFAR方法^[1,2]是一种经典的局部自适应目标检测方法。该方法通过预设滑动窗口遍历SAR图像实现目标检测,对窗口内像素灰度与自适应阈值进行对比以区分目标和杂波。自适应阈值由预先设置的恒定虚警率和窗口中的杂波分布确定。文献^[3]提出了一种基于双边微调统计的CFAR检测方法,该方法提出了一种基于双边阈值的策略,自动裁剪窗口内的样本来剔除异常值,提高了在海洋场景下的检测性能。此类方法需要SAR图像中目标与杂波具有较高对比度来拟合杂波的统计分布,因此只适用于简单场景,当场景较为复杂时,会造成检测性能的降低。

近年来,在光学领域,由于网络深度的增加以及计算能力的不断提高^[4],基于深度学习的目标检测方法^[5-9]开始占据主流,取得了不错的效果。鉴于深度学习在光学图像目标检测领域的成功,研究者也将深度学习运用在SAR目标检测当中^[10]。基于深度学习的SAR检测方法相较于浅层模型方法特征提取能力强、对于复杂SAR图像场景的检测效果好。文献^[11]将底层纹理、边缘特征与高层深度特征进行融合,提高了SAR舰船目标的检测性能。目前一些基于深度学习的SAR目标检测方法是由基于区域的卷积神经网络(Region-CNN, R-CNN)结合候选区域来设计的两阶段检测方法。文献^[12]使用多分辨率卷积特征,基于Faster R-CNN方法,对原有的卷积神经网络VGG-16的结构进行改进,提升了对小型目标的检测效果。文献^[13]在Faster R-CNN算法基础上,使用SAR分类数据集对网络进行预训练并扩充训练数据,解决了SAR目标检测训练样本不足的问题。此类方法在对区域候选网络(Region Proposal Network, RPN)产生的大量候选框进行挑选时,往往采用非极大值抑制(Non-Maximum Suppression, NMS)方法。在大场景SAR图像检测中,由于目标特征易受杂波影响,可鉴别性更差,并且单个目标占据图像比例更小,导致基于

Faster R-CNN的目标检测方法在整张特征图上产生的候选框会包含大量杂波,而NMS方法无法在筛选候选框时有效去除杂波,导致目标检测产生大量虚警。

近几年,在人工智能领域中,强化学习^[14,15]得到了广泛关注。强化学习根据当前自身状态(State)并结合策略(Policy)做出相应的动作(Action),通过与外界的交互获得不同的奖励(Reward)来不断更新自身的状态并调整策略,最终形成解决某一问题的最优策略。将具有解译能力的深度学习与具有决策能力的强化学习结合而形成的深度强化学习,可以更好地与其他基于深度学习算法相结合,应用在目标检测^[16-18]中以提高检测性能。在光学领域,文献^[16]根据当前收集到的图像信息,使用强化学习从5个预定义的固定形状和尺寸的候选区域中选择最有可能包含目标的区域,并通过迭代不断缩小候选区域,最终框定目标。但由于预设的候选区域并不能完全覆盖所有目标,此方法精度较低。

受强化学习启发,本文基于Faster R-CNN检测模型^[7],设计了新的强化学习模型实现自适应候选框挑选,有效减少冗余候选框数量。提出方法在RPN与感兴趣区域(Regions-of-Interest, RoI)池化层之间加入深度强化学习网络,对于RPN产生的大量初始候选框,深度强化学习网络综合相关信息进行迭代搜索,在特征图上不断找到可能含有目标的搜索区域,并挑选搜索区域内的候选框输入后续的检测器进行分类、回归,实现对搜索区域内的目标检测。由于深度强化学习网络是基于循环神经网络(Recurrent Neural Network, RNN)设计的,因此可以在迭代过程中捕捉到图像的上下文信息并确定可能含有目标的搜索区域的位置坐标。并且,本方法在强化学习中对产生的搜索区域尺寸添加距离约束,使得搜索区域尺寸可以根据之前迭代过程中的搜索区域以及目标检测结果进行自适应调整。针对大场景SAR图像中目标数量较多,分布情况较为复杂且易受杂波影响的特点,提出方法通过使用强化学习自适应确定搜索区域实现了对初始候选框的自适应挑选,提升了对背景杂波的鉴别能力并减少传统强化学习应用于检测问题的计算量。所提方法自适应确定搜索区域的能力包括两方面的自适应,一是通过综合利用图像的特征信息和上下文信息自适应确定搜索区域的位置坐标;二是通过搜索区域尺寸约束自适应调整下一次搜索区域的范围尺寸。基于实测数据的实验结果表明,所提方法能够有效减少SAR目标检测的虚警数量,提升传统深度学习目标检测方法的检测性能。

2 背景介绍

2.1 Faster R-CNN

Faster R-CNN^[7]是目前比较常用的目标检测框架,用RPN来获取区域候选。如图1所示,Faster R-CNN的结构可以分为4个主要部分:特征提取网络、RPN、RoI池化层以及检测器。其中,特征提取网络由卷积层、激活函数和池化层组成,用于提取输入图像的特征映射作为输出。后续的RPN和检测器将使用输出的特征映射完成候选框的生成以及分类、回归,实现卷积共享。RPN在特征图中每个点上设置 k 个锚框,对每个锚框进行二分类和初步位置修正并使用NMS算法进行候选框筛选作为初始的候选框;RoI池化层则负责收集原始的特征图和候选框,将其整合后提取出候选框对应位置的特征映射;最后输入检测器中进行目标分类和边界框的位置修订。

2.2 强化学习

强化学习的过程通常可以用图2中的马尔可夫决策过程^[14](Markov Decision Process, MDP)来描述:智能体(agent)在环境当中,拥有其对当前环境感知的状态量 S ;通过策略 $\pi: S \rightarrow A$ 从动作集 A 中选择动作 $a \in A$,根据动作的不同,环境出现基于一定概率的改变并更新状态量 S ;在状态发生变化的同时,环境会根据状态变化通过奖励函数给予智能体相应的奖励 r 。这样智能体在与环境交换信

息的过程中,依据得到的奖励来不断调整策略,最终得到最优策略。

智能体能够与环境进行交互,每个状态是智能体对当前环境的感知,并且智能体只能通过动作来影响环境,而策略能够指导智能体在当前状态下做出何种动作,可以将其表示为一个由状态到动作的映射。奖励是在当前环境状态下,对智能体所进行动作的即时评价,这个评价也是策略优化的主要依据,可以表示为奖励函数的形式。当智能体做出一个带来较低奖励的动作时,当下次遇到相同的环境状态时,调整过的策略就可能选择其他的动作来争取获得更高的奖励。强化学习的目标可以确定为:学习一个最优策略,来最大化期望累计奖励。因此,设置合适的状态、动作以及奖励函数,是强化学习获得最优策略的关键。

3 结合强化学习的SAR目标检测方法

3.1 总体框架

本方法基于传统Faster R-CNN方法,整体框架如图3所示。除深度强化学习网络之外,其余各部分与Faster R-CNN模型相似。特征提取网络采用VGG-16,尺寸为 $h_{ori} \times w_{ori} \times 3$ 的输入SAR图像在经过特征提取后生成 $h \times w \times 512$ 的特征图, h 和 w 与输入图像的尺寸有关。在获取特征图后,RPN会在特征图的每个像素点上生成 k (本文设置 $k=9$)个锚框,并通过softmax二分类器和回归器获得更精确的初始候选框,初始候选框将构成深度强化学习网络输入的一部分。对于RPN生成的大量初始候选框,我们使用深度强化学习网络进行区域搜索实现对初始候选框的挑选,将可能含有目标的候选框送入RoI池化层进行整合。RoI池化层会将挑选出来的候选框对应位置的特征映射划分为 7×7 的网格,并对每一个网格进行最大值池化处理,以输出固定尺寸的特征向量。这些特征向量将被送入检测器,通过全连接层和softmax计算候选框的具体类别,输出类别概率预测向量,同时利用边界框回归,获取更准确的检测目标框。其中深度强化学习网络采用神经网络实现,此网络可以在特征图上找到一个可能含有目标的区域,并将该区域内的候选框送入后续的检测器进行分类回归,然后根据检测结果再找到另一个可能含有目标的区域,继续将该区域内的候选框送入检测器,如此迭代,完成候选框的自适应挑选。

深度强化学习网络部分参考基于卷积门控循环单元^[19](Convolutional Gated Recurrent Unit, Conv-GRU)进行设计。Conv-GRU作为一种计算需求相对较低的循环神经网络,对网络输入输出具

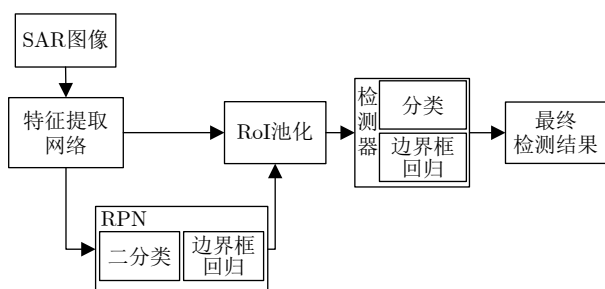


图1 Faster R-CNN结构

Fig. 1 Faster R-CNN structure

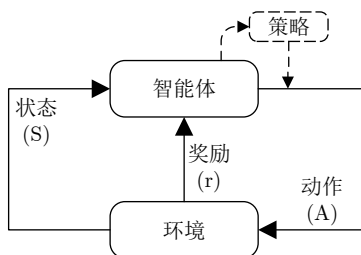


图2 强化学习原理

Fig. 2 Principles of reinforcement learning

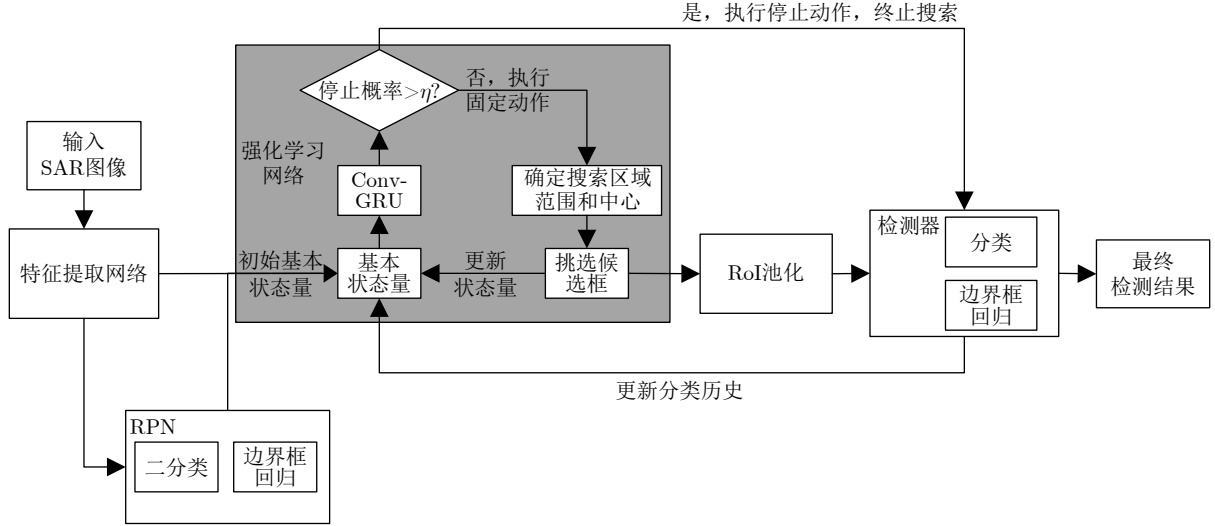


图3 结合强化学习的SAR目标检测方法整体框架

Fig. 3 Framework of SAR target detection method using reinforcement learning

有记忆能力，能够捕捉迭代过程中的依赖关系，是能够实现深度强化学习的网络之一。Conv-GRU智能体的方程式如下：

$$O_t = \sigma[W_{so} * S_t + W_{ho} * H_{t-1} + b_o] \quad (1)$$

$$\tilde{H}_t = W_{sh} * S_t + W_{hh} * (O_t \odot H_{t-1}) + b_h \quad (2)$$

$$Z_t = \sigma[W_{sz} * S_t + W_{hz} * H_{t-1} + b_z] \quad (3)$$

$$H_t = (1 - Z_t) \odot H_{t-1} + Z_t \odot \tanh[\tilde{H}_t] \quad (4)$$

$$\tilde{A}_t = \text{ReLU}[W_{ha} * H_t + b_a] \quad (5)$$

$$A_t = \tanh[W_{aa} * \tilde{A}_t + b_a] \quad (6)$$

其中，*表示卷积乘法， \odot 表示Hadamard乘积，权重和偏差分别表示为 W 和 b ，所有输入和输出的空间尺寸是 $h \times w$ 。 O_t 为重置门，决定如何将新的输入信息与之前的记忆信息相结合。 Z_t 为更新门，用于控制记忆信息的保留程度。 \tilde{H}_t 为候选隐藏状态，包含了当前时刻的输入信息和上一时刻保留的信息。 H_t 为最终的隐藏状态，通过更新门决定如何组合过去的隐藏状态和当前的候选隐藏状态。在第 t 次迭代时到达Conv-GRU的输入是强化学习的状态量 S_t 和上一次迭代输出的隐藏状态 H_{t-1} ，其中状态量由图像特征和RPN输出组成，Conv-GRU可根据图像信息来进行候选区域搜索。输出是对应两个动作量 A_t ，分别决定是否进行候选区域搜索以及搜索区域的位置和尺寸。深度强化学习可基于Conv-GRU调整策略，完成状态到动作的映射。

网络可训练的参数部分包括特征提取网络部分、RPN部分、检测器部分以及深度强化学习部分。

3.2 结合强化学习的候选框挑选方法

下面详细介绍如何利用深度强化学习网络进行

区域搜索来实现候选框的自适应挑选。在训练阶段，第 t 次迭代时，智能体根据当前状态和策略决定是否终止搜索。如果搜索没有停止，就执行固定动作 a_t^f 自动获得新的搜索区域的位置 z_t 和尺寸参数 p_t ；RoI观测量 R_t 在以 z_t 为中心的搜索区域中被更新，搜索区域内所有由RPN生成的初始候选框均被发送到ROI池化层，然后在检测器中进行分类和边界框回归预测。根据训练数据的标注与预测结果计算奖励 r 及检测结果更新强化学习基本状态量 S_t 。基于新状态，在第 $t+1$ 次迭代时采取新操作，并重复该过程直到发出停止搜索动作 a_t^d ，然后收集整个搜索轨迹中的预测结果并计算总的累积奖励 $\sum r$ 。训练的目标即为累积奖励的最大化，并据此不断优化策略，最终得到最优策略 $\pi_\theta(a_t|s_t)$ ，具体优化方法在3.2.3节中描述。而在测试过程中，搜索策略将被固定，在第 t 次迭代时，智能体根据当前状态 s_t 和已经训练好的策略 $\pi_\theta(a_t|s_t)$ 决定是否搜索以及搜索区域的位置及尺寸，然后选择候选框送入后续检测部分并更新相应的状态量。基于新状态，在第 $t+1$ 次迭代时采取新操作，并重复该过程直到发出停止搜索动作 a_t^d ，最后收集整个搜索轨迹中的预测结果。算法1展示了强化学习自适应候选框挑选测试过程的伪代码。下面分别介绍状态、动作以及奖励函数的具体设置。

3.2.1 状态

状态量 s_t 是一个数组，包含3部分： $s_t = (R_t, S_t, H_t)$ ，其中 $R_t \in \{0, 1\}^{h \times w \times k}$ 是RoI观测量， $S_t \in \mathbb{R}^{h \times w \times (d+2k+N+1)}$ 是基本状态量，而 $H_t \in \mathbb{R}^{h \times w \times 300}$ 是Conv-GRU的隐藏状态， d 为VGG-16的输出特征维度， N 是要检测对象类别的数量。

算法 1 自适应候选框挑选方法

Alg. 1 Adaptive region proposal selection

输入: 状态 $s_t = \{R_t, S_t, H_t\}$, 最大迭代次数为 $M = 10$
 输出: 搜索区域中心坐标和范围参数 $\{z_t, p_t\}$

```

1: 初始化状态量  $s_t = \{R_{t-1}, S_{t-1}, H_t\}$ 
2: for  $t = 1 \rightarrow M$  do //  $t$  表示迭代次数
3:   if  $(\pi(a_t^d = 1|s_t) > \eta)$  then //  $\eta$  为概率阈值, 取 0.5
4:     break // 结束迭代搜索
5:   else
6:      $z_t \leftarrow \max \pi(a_t^d = 0, a_t^f = z_t|s_t)$ 
7:      $p_t \leftarrow z_t, z_{t-1}$  // 计算搜索区域中心坐标和范围
8:   end if
9:   输出  $z_t, p_t$  // 挑选候选框并送入后续检测器部分进行检测
10:   $R_t \leftarrow R_{t-1}, S_t \leftarrow S_{t-1}$  // 更新状态  $R_t, S_t$ 
11: end for

```

RoI 观测量 R_t 是一个大小为 $h \times w \times k$ 的二元量, 其中当相应候选框在搜索区域内时, 对应的坐标 (i, j, l) 值为 1, 然后转入到网络的 RoI 池化和检测器部分进行分类。 R_t 初始为全零量。在固定动作 a_t^f 之后, 固定位置 z_t 相邻区域的一部分 R_t 将被更新, 模型将在此区域内使用 RPN 输出的全部初始候选框进行目标检测。这个相邻区域设置为一个以 z_t 为中心的矩形区域, 区域的边长可以进行自适应的调整。将此矩形区域内对应的所有 R_t 项设置为 1, 表示此区域内的候选框已经被探测过了。

基础状态量 S_t 包括 $V_t^1 \in \mathbb{R}^{h \times w \times d}$, $V_t^2 \in \mathbb{R}^{h \times w \times k}$, $V_t^3 \in \mathbb{R}^{h \times w \times k}$ 和 $V_t^4 \in \mathbb{R}^{h \times w \times (N+1)}$ 。将 V_0^1 设置为与输入 RPN 相同的基本特征映射, 将 V_0^2 设置为 RPN 的二分类量。RPN 的回归量被用于 V_t^3 , V_0^3 设置为 $[0, 1]$ 归一化偏移量 $[\Delta x_1, \Delta y_1, \Delta x_2, \Delta y_2]$ 。 V_t^2 和 V_t^3 分别对应着特征图每个像素位置上预设的 k 个 anchor 的二分类和回归结果。当某一位置在之前的迭代过程中被访问后, 使用 R_t 更新这些量, 将 V_t^1 , V_t^2 和 V_t^3 中的对应位置设置为 -1, 表示这些位置已被访问过, 这样有利于强化学习在下一迭代过程中对输入状态量的分析, 也有利于强化学习策略的训练, 防止在重复的位置上多次进行搜索。 V_t^4 表示候选框检测结果的历史记录, 将 V_0^4 设置为 0。强化学习网络执行固定动作获得搜索区域后, 将区域内的所有候选框送入检测器进行分类预测, 然后使用 NMS 对已分类的候选框进行筛选。对筛选后的候选框进行边界框回归预测, 并将输出结果即最终检测框的中心坐标和类别概率向量记录在 V_t^4 的相应空间位置, 作为下一次迭代过程的输入状态之一。这样做可以为强化学习提供之前迭代

过程中的检测结果历史信息, 有利于强化学习的决策。

3.2.2 动作

动作分为两种类型, 分别对应于 A_t 的两个通道: 固定动作 a_t^f 和停止动作 a_t^d 。停止动作是二值的, 其中 $a_t^d = 1$ 对应于终止搜索。如果 $a_t^d = 0$, 则执行固定动作 $a_t^f = (z_t, p_t)$, 其中 z_t 是下一个搜索区域的中心坐标, p_t 为下一个搜索区域的尺寸参数, 可结合特征图尺寸进一步确定搜索区域尺寸。下面将描述随机策略 π_θ 如何对动作进行选择。

随机策略 $\pi_\theta(a_t|s_t)$ 使用 A_t 进行动作的选择。令 A_t 中的双通道分别表示为代表停止动作的 A_t^d 和代表固定动作的 A_t^f 两个层。将停止层 A_t^d 调整为向量 $\mathbf{d}_t \in \mathbb{R}^{625}$ 。然后, 计算出终止于状态 s_t 概率为

$$\pi_\theta(a_t^d = 1|s_t) = \sigma[\mathbf{w}_d^T \mathbf{d}_t + t] \quad (7)$$

其中, \mathbf{w}_d^T 是可训练的权重向量的转置。

对于固定动作, 使用 softmax 函数, 将固定层 A_t^f 转换为概率映射 \hat{A}_t^f 。在智能体没有执行停止动作的情况下, 固定动作选择位置 $z(t) = (i, j)$ 的概率为 $\hat{A}_t^f[z_t]$ 。因此, 在状态 s_t 下搜索区域的中心坐标为 $z(t)$ 的概率可以表示为

$$\pi_\theta(a_t^d = 0, a_t^f = z_t|s_t) = (1 - \sigma[\mathbf{w}_d^T \mathbf{d}_t + t]) \hat{A}_t^f[z_t] \quad (8)$$

同时, 为了确定搜索区域尺寸, 我们设计了一种新的基于距离的约束。该约束能够通过迭代过程中搜索区域位置的变化, 自动调整搜索区域的尺寸, 在准确搜索到含有目标区域的同时减少传统强化学习的计算量。具体来说: 模型利用本次迭代中选择的中心坐标 $z(t)$ 与上一次迭代选择的中心坐标 $z(t-1)$ 对搜索区域尺寸进行调整, 形成参数 p_t 来确定搜索区域尺寸。参数 p_t 的取值区间为 $(0, 1)$, 计算方法如下:

$$p_t = \begin{cases} \frac{2\Delta - h_{t-1}}{h_0}, & \Delta < \frac{h_0 + h_{t-1}}{2} \\ 1, & \Delta \geq \frac{h_0 + h_{t-1}}{2} \end{cases} \quad (9)$$

其中, $\Delta = |z(t) - z(t-1)|$ 为两次搜索区域中心之间的距离, h_0 和 w_0 为初始搜索区域尺寸 (设 $h < w$), 与输入图像宽高比相同, h_{t-1} 和 w_{t-1} 为 $t-1$ 次迭代时的搜索区域尺寸。第 t 次迭代时的搜索区域尺寸计算方式如下:

$$h_t = h_0 \times p_t \quad (10)$$

$$w_t = w_0 \times p_t \quad (11)$$

此设置可实现在两次迭代过程中, 当搜索区域之间距离较远时, 搜索尺寸扩大; 当搜索区域之间距离较近时, 搜索尺寸缩小, 减少搜索区域的重叠, 提高搜索效率。

3.2.3 奖励

深度强化学习网络产生的搜索策略在搜索候选框时, 应在保证较高交并比(Intersection over Union, IoU)的同时, 尽量减少候选框的数量。这样可以在减少虚警(false positive)数量的同时, 节约处理时间。奖励函数以此为标准进行设定。

将奖励函数分为固定动作奖励和停止动作奖励部分。其中, 对于固定动作, 其奖励函数由两部分组成: 第1部分为每次执行固定动作都将获得的较小的负奖励 $-\beta$ (经过实验 β 设置为0.075); 第2部分为智能体执行固定动作时获得的正奖励, 这个正奖励与在当前图像的任何真实标注数据(ground truth) g_i 的IoU相关, 即如果固定动作在当前图像与

$$r_t^f = \begin{cases} -\beta + \sum_i I[g_i : \text{IoU}_t^i > \text{IoU}^i \geq \tau] \frac{\text{IoU}_t^i - \text{IoU}^i}{\text{IoU}_{\max}^i}, & p_t < 1 \\ -\beta + \max_{\text{IoU} \geq \tau} \frac{\text{IoU} - \text{IoU}^i}{\text{IoU}_{\max}^i}, & p_t = 1 \end{cases} \quad (12)$$

其中, 当 $p_t < 1$ 时, 表示强化学习网络认为在上次迭代中搜索区域的附近目标分布较密集, 因此使用对多目标奖励较高的函数形式; 而当 $p_t = 1$, 搜索区域距离较远时, 则使用对单目标奖励较高的函数形式。

对于停止动作, 在搜索终止后, 智能体会受到一个能够反映搜索轨迹质量的最终奖励:

$$r_t^d = \sum_i I[g_i : \text{IoU}_{\max}^i \geq \tau] \frac{\text{IoU}^i - \text{IoU}_{\max}^i}{\text{IoU}_{\max}^i} \quad (13)$$

其中, IoU^i 经过更新, 已经成为关于标注数据 g_i 在整个搜索迭代过程中所产生的最大IoU值。如果 g_i 没有被最大限度覆盖, 则会给予智能体一个随着 IoU^i 减小而不断增大的负的奖励值。并且, 如果 g_i 已经被最大限度的覆盖了, 即 $\text{IoU}^i = \text{IoU}_{\max}^i$ 的时候, 本次停止动作的奖励值将变为0。

在训练过程中通过累积奖励最大化来优化策略, 结合REINFORCE^[20]方法进行梯度更新, 使用50条搜索轨迹来逼近真实梯度, 并使用Adam^[21]优化器来更新策略参数。

3.3 训练

整个模型需要训练的参数可分为两部分: 一部分是原始Faster R-CNN部分的参数, 包括特征提取网络、RPN和检测器部分参数; 另一部分是强化学习搜索策略部分的参数。两部分参数采用交替训练的方式进行优化: 当强化学习部分参数进行更新后(此时原始Faster R-CNN部分参数保持不变), 模型将固定强化学习部分参数并使用该强化学习策

任何 g_i 的IoU增加了, 智能体都会获得正向奖励。对于每个标注数据 g_i , 设置 IoU^i 为在过去0, 1, \dots , $t-1$ 次迭代过程中产生的最大的IoU值, 并且在 $t=0$ 时, $\text{IoU}^i = 0$ 。当 $t \geq 1$ 时, 设置 IoU_t^i 为在第 t 次迭代过程中, 对于每个标注数据 g_i 所产生的最大IoU值。并检查是否满足 $\text{IoU}_t^i > \text{IoU}^i \geq \tau$ 。其中, 依据数据集PASCAL VOC的正阈值设置 $\tau = 0.5$ 。如果满足上述情况, 将对相应的标注数据 g_i 给出正向奖励 $(\text{IoU}_t^i - \text{IoU}^i) / \text{IoU}_{\max}^i$ 并在之后更新 $\text{IoU}^i = \text{IoU}_t^i$ 。此处, IoU_{\max}^i 指的是对于第 i 个标注数据 g_i , 在所有可能区域内获得的最大IoU值(即最终预测的关于标注数据 g_i 的真实IoU值)。综上所述, 在第 t 次迭代时给出的固定动作奖励为

略进行自适应候选框挑选, 挑选出的候选框将被送入后续检测器进行目标分类和回归, 并以此来更新Faster R-CNN部分的参数。模型通过两部分参数交替更新, 反复迭代至收敛。

4 实验与分析

首先对实验所用数据集及评价准则进行介绍, 然后与传统Faster R-CNN及其相关改进方法进行对比分析, 对提出的创新点进行实验验证, 证明所提方法的有效性。实验平台为Ubuntu系统, 代码基于Tensorflow编写。训练方面, 本方法采用经过预训练的VGG-16网络对Faster R-CNN的共享卷积层部分进行参数初始化; 其他网络层则使用零均值、标准差为0.01的高斯分布进行参数随机初始化。

4.1 实验设置

4.1.1 实验数据集介绍

本文使用MiniSAR数据集^[22]进行实验, 该数据集是美国桑迪亚实验室在2006年公开的SAR图像数据集, 包含复杂场景的SAR实测图像数据。在本文实验中, 共使用9幅SAR图像(7幅用于训练, 2幅用于测试), 设置车辆为感兴趣目标。在此数据集中, 由于图像的尺寸过大, 无法直接输入网络。因此, 首先将数据集中的原始图像裁剪成许多大小为300像素 \times 300像素的子图像, 并使用这些子图像进行网络训练。与训练类似, 在测试过程中通过滑动窗口, 将原始的测试图像也裁剪为大小为300像素 \times 300像素的子图像, 滑动窗口的步长设置为200像素。

对测试子图像进行检测后, 再将检测结果恢复到原始大图中。在恢复过程中, 我们对子图像中的检测结果进行NMS删除重复数据, 以获得最终结果。

4.1.2 评价准则

实验选择F1-score和接收机性能(Receiver Operating Characteristic, ROC)曲线作为检测性能的评价准则。F1-score的计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$F1\text{-score} = \frac{2PR}{P + R} \quad (16)$$

其中, TP (True Positives)是检测结果中正确的目标个数, FP (False Positives)为虚警, 是检测结果中错误的目标个数, FN (False Negatives)是漏警, 是未检测到的目标个数, P (Precision)是准确率, R (Recall)是召回率。

ROC曲线的绘制参考文献[23], 曲线描述了真阳率(True Positive Rate, TPR)和假阳率(False Positive Rate, FPR)之间的关系。TPR和FPR的计算公式如下:

$$TPR = \frac{TP}{TP + FN} \quad (17)$$

$$FPR = \frac{FP}{FP + TN} \quad (18)$$

ROC曲线和坐标轴下的面积(Area Under Curve, AUC)用于辅助ROC评估。通常, AUC越大, 性能越好。

4.2 检测结果分析

表1对比了不同方法的实验结果, 其中Gaussian-CFAR表示文献[13]中的方法; Faster R-CNN方法基于文献[7]; SSD方法使用文献[24]; Faster R-CNN+CBAM在Faster R-CNN中加入通道注意

表 1 不同方法实验结果

Tab. 1 Experimental results of different methods

方法	P	R	F1-score
Gaussian-CFAR	0.3789	0.7966	0.5135
Faster R-CNN	0.8156	0.9268	0.8677
SSD	0.8451	0.9106	0.8750
Faster R-CNN+CBAM	0.8409	0.9268	0.8818
本文方法-尺寸固定	0.8369	0.9350	0.8832
本文方法	0.8686	0.9350	0.9006

力和空间注意力模块[25]聚焦重要特征来辅助检测; 本文方法-尺寸固定表示只使用强化学习确定搜索区域位置, 而搜索区域的尺寸不能自适应变化, 生成固定尺寸($h_0 \times w_0$)的搜索区域来完成候选框挑选, 本文中取 $h_0 = h_{ori} \times 0.25$, $w_0 = w_{ori} \times 0.25$ 。

从表1可以看出, Gaussian-CFAR的F1-score非常低, 因为此方法仅使用SAR图像本身的对比度等信息进行检测, 需要拟合杂波分布, 只适用于简单场景, 当场景较为复杂时, 会产生大量虚警, 造成检测性能的降低。而传统Faster R-CNN方法虽然检测效果好于Gaussian-CFAR, 但在深度学习相关方法中F1-score最低, 主要原因为虚警很多, 很多背景杂波被误判为目标, 导致检测的准确率降低; SSD方法则通过使用多尺度特征进行预测, 相较于传统Faster R-CNN方法虚警明显减少, 获得了不错的性能; Faster R-CNN+CBAM方法在添加了注意力模块之后, 能够让网络更加关注目标区域, 相较于传统Faster R-CNN方法准确率有所提升, 虚警明显减少; 本文方法相较于其他方法, 在准确率和召回率上都有一定提升, 尤其在减少虚警方面, 相较于其他方法提升明显; 当搜索区域尺寸固定时, 本文方法则实现了与Faster R-CNN+CBAM相似的检测结果, 相较于传统Faster R-CNN方法的F1-score有一定提升, 但由于无法自适应确定搜索区域尺寸, 准确率相比本文原始方法有所下降。本文方法的F1-score相较于传统Faster R-CNN方法提升了0.0329, 准确率提升了0.0530。因此, 本文方法通过强化学习进行区域搜索来实现对候选框的自适应挑选, 可以有效减少虚警, 提升检测性能。

图4展示了各种方法的ROC曲线和相应的AUC

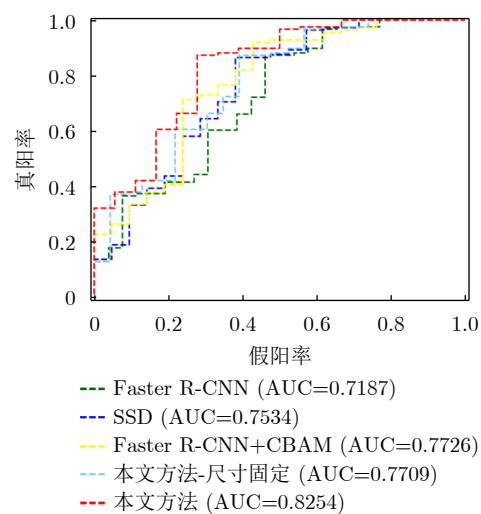


图 4 ROC曲线对比分析

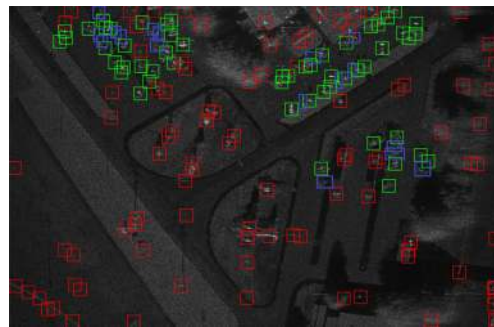
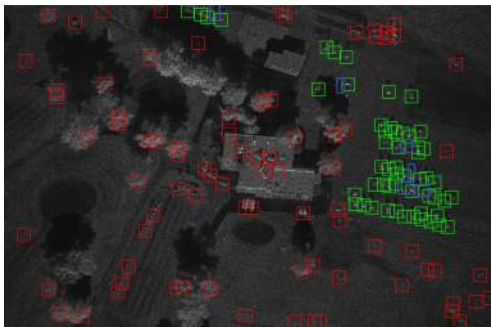
Fig. 4 ROC curves comparative analysis

值，由于CFAR检测效果相比于深度学习方法差距较大，研究价值不高，因此只对深度学习相关方法进行了ROC曲线刻画。我们以检测结果中的每一个边界框为对象，设置IoU阈值来判断边界框是否检测到正确目标，并根据其分类得分由高到低进行排序，通过依次将每个边界框划分为正例，来计算不同的TPR和FPR值，最终得到完整的ROC曲线。从中可以发现本文所提方法可以在保持较低FPR的同时获得较高的TPR，因此也获得了最高的AUC值，检测性能最好。

图5分别展示了上述4种方法在MiniSAR数据集上两张图像的测试结果：图中绿色框表示检测正确的目标，红色框表示检测错误的目标(虚警)，蓝色框表示未检测到的目标(漏警)。由图5可知，在复杂背景杂波的大场景SAR图像中，传统Faster R-CNN

方法检测结果虚警较多；本文方法的检测结果中虚警最少，检测效果最好，并且在目标数量较多，排布较密集且有一定规律时，能够实现目标的精确检测。

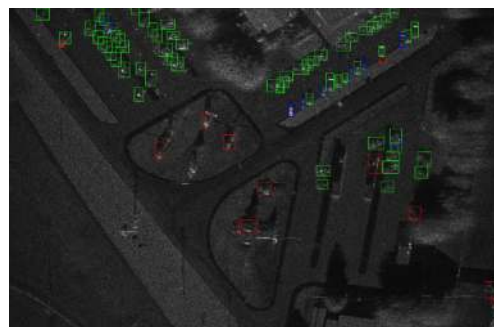
所提方法检测效果较好的原因主要有两方面：第一，采用深度强化学习网络确定搜索区域对候选框进行自适应挑选，能够对RPN生成的大量初始候选框进行鉴别，剔除其中的大量虚警框；第二，实验所使用数据集中的部分车辆目标分布和排列存在一定规律，本文方法能够在强化学习训练过程中学习这些规律，并在迭代搜索过程中保存这些上下文信息，获得更好的检测效果。我们计算了每张训练图像在训练过程中 $p_t = 1$ 所占比例来进行实验验证。在使用强化学习确定搜索区域时， $p_t = 1$ 表示本次迭代确定的搜索区域与上一次迭代确定的搜索



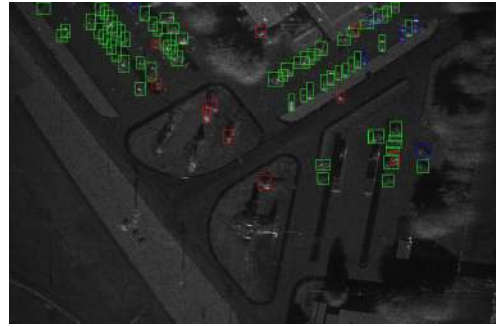
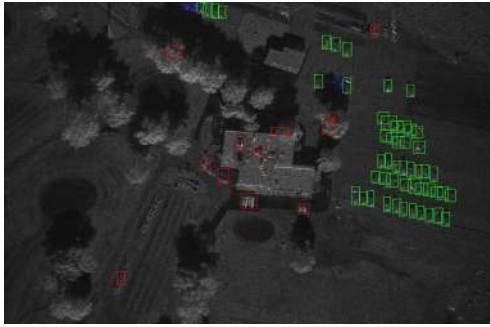
(a) Gaussian-CFAR的检测结果
(a) Detection results of Gaussian-CFAR



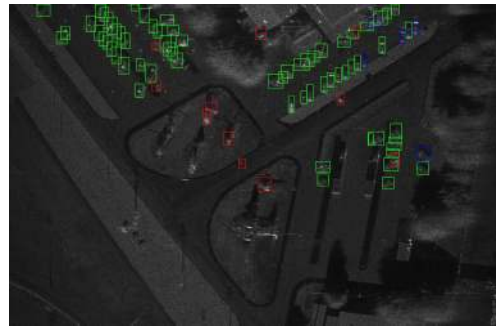
(b) 传统Faster R-CNN的检测结果
(b) Detection results of Faster R-CNN



(c) SSD的检测结果
(c) Detection results of SSD



(d) Faster R-CNN+CBAM的检测结果
(d) Detection results of Faster R-CNN+CBAM



(e) 本文方法-尺寸固定的检测结果
(e) Detection results of proposed method (fixed size)



(f) 本文方法的检测结果
(f) Detection results of proposed method
□ 检测正确的目标 □ 虚警 □ 漏警

图 5 SAR图像目标检测结果

Fig. 5 Detection results of SAR images

区域之间距离较远, $p_t < 1$ 则表示本次迭代确定的搜索区域在上一次迭代确定的搜索区域附近。因此, 如果 $p_t = 1$ 所占比例较小, 则说明搜索区域分布较为集中, 图像中的目标数量较多且分布密集; 如果 $p_t = 1$ 所占比例较大, 则说明搜索区域较为分散, 图像中的目标数量较少且分布稀疏。以图6中两张训练图像样本为例进行分析, 图中白色框为固定动作所产生的搜索区域, 左上角数字为区域的生成顺序。图6(a)中目标较少, 因此强化学习只执行了3次搜索动作, 搜索区域较为分散, p_t 值全部为1; 而图6(b)中由于目标数量较多、分布密集且排列整齐, 因此强化学习共执行了7次搜索动作, 且

在目标密集分布的区域多次生成搜索区域, $p_t = 1$ 所占比例为0.33, 相比于图6(a) $p_t = 1$ 所占比例明显减小。这说明, 强化学习能够学习到目标分布的规律性并应用于检测: 在检测目标较少、分布稀疏的图像时, 强化学习生成的搜索区域也较为分散, 可以在较少搜索次数内准确找到目标所在区域; 而在检测目标密集分布的图像时, 强化学习则会在目标密集分布的区域多次生成搜索区域来检测出该区域内的全部目标。

4.3 搜索区域对检测的影响

为了研究搜索区域对检测的影响, 我们对强化

学习的搜索过程进行了分析，并与不使用搜索区域的Faster R-CNN方法进行了对比。

图7展示了测试图像的强化学习可视化搜索过程。图7(a)为原始图像，图7(b)–图7(e)展示了本文方法的迭代搜索过程。其中，白色边框为固定动作所产生的搜索区域，左上角数字为区域的生成顺序，其中心位置坐标和尺寸确定方法与3.2节所述相同；绿色边框则表示已检测到目标的边界框。注意白色边框表示所挑选初始候选框中心点的区域集合，而绿色边框为经过回归的最终目标边界框，因此白色边框并不一定完全包裹目标和绿色边框。

图7(b)表示在强化学习第1次迭代过程中，策略执行固定动作确定的搜索区域，之后中心点在搜索区域内的候选框将被挑选并送入后续的检测器部分。图7(c)展示了在搜索区域内检测到的两个目标。图7(d)表示在强化学习第2次迭代过程中，

策略执行固定动作确定的搜索区域，由于距离约束的作用，该次搜索区域尺寸进行自适应调整，较上次迭代有所减小，减少了区域内初始候选框数量，在能够检测到目标的同时，降低了强化学习部分的计算量。图7(e)展示了在搜索区域内检测到的所有目标。两次迭代后，强化学习策略判断SAR图像内全部目标已经检测完成，停止迭代，完成检测流程。

由图7可视化搜索过程可知，强化学习通过训练得到的策略，能够让搜索区域更靠近待检测目标。在迭代搜索过程中，除了利用图像的特征信息进行决策外，深度强化学习网络还能够记录之前迭代步骤中产生的上下文信息，通过上下文信息进行不同搜索区域之间的信息交换，帮助决策，能够有效减少杂波对检测器的干扰，提高检测的准确性。

为了进一步分析搜索区域对检测的影响，我们

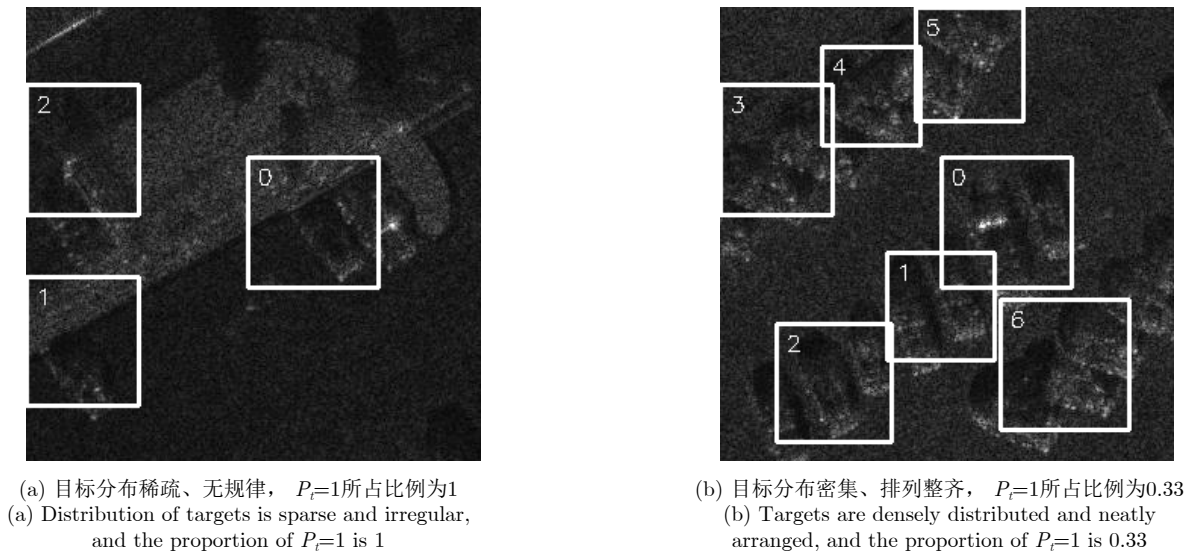


图 6 训练图像样本示例(白色框为固定动作所产生的搜索区域)

Fig. 6 Training image example (The white box indicates the search area generated by fixed action)

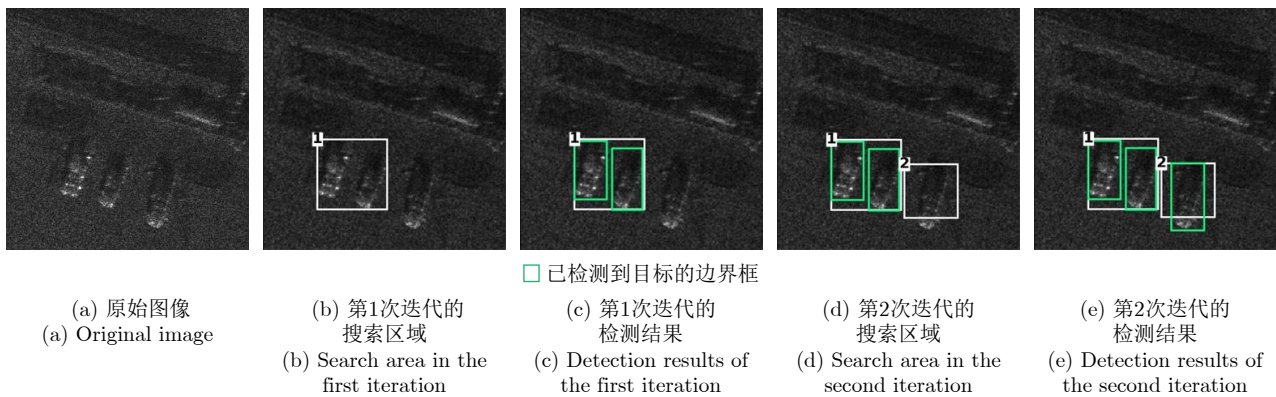


图 7 可视化搜索过程(白色框为固定动作所产生的搜索区域)

Fig. 7 Visualization of search (The white box indicates the search area generated by fixed action)

分别对不使用搜索区域的Faster R-CNN方法和可以自适应确定搜索区域的本文方法的RoI分布进行了对比。图8展示了两种方法产生的RoI分布对比。图8(a)为待检测的原始图像，图8(b)为传统Faster R-CNN产生的RoI分布，图中高亮区域为RoI的中心位置。图8(c)为传统Faster R-CNN的检测结果，绿色框表示检测正确的目标，红色框表示虚警。图8(d)

为本文方法产生的RoI分布，图中白框为强化学习产生的搜索区域，所生成RoI的中心位置均在白框之内。图8(e)为本文方法的检测结果。结合图8(b)和图8(d)可以发现，相较于传统的Faster R-CNN方法，经过本文方法的处理，RoI会更集中地产生在目标附近，因图片边缘和复杂背景杂波而产生的RoI大量减少。

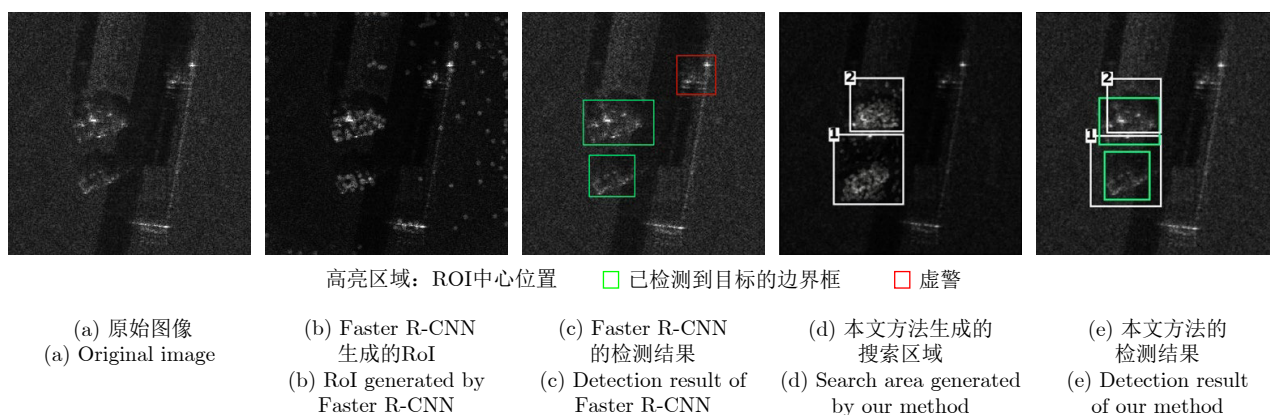


图 8 RoI分布对比(白色框为固定动作所产生的搜索区域)

Fig. 8 Visualization of RoI (The white box indicates the search area generated by fixed action)

结合表1的检测结果及RoI分布可以分析，Faster R-CNN方法由于采用NMS法对RPN在整张特征图上产生的大量初始候选框进行处理，会导致SAR图像的边缘和难鉴别的背景杂波位置生成大量RoI，从而造成检测结果中出现较多虚警。而所提方法则依靠强化学习策略及深度强化学习网络积累的上下文信息来确定含有目标的搜索区域，并且只在搜索区域内生成RoI，能够让RoI尽可能集中在目标周围，有效减少图片边缘及复杂背景杂波导致的虚警，获得较好的检测效果。

4.4 运行时间分析

运算速度也是衡量目标检测算法性能的重要指标之一，本节对Gaussian-CFAR, Faster R-CNN, Faster R-CNN+CBAM, SSD以及所提方法的运行时间进行比较分析，并分别对所提方法采用固定的搜索区域尺寸、自适应搜索区域尺寸进行实验，我们取所有测试图像的平均测试时间作为单张图片的测试时间，结果如图9所示。

由图9可知，Gaussian-CFAR无论在运算速度还是准确率方面与其他方法均具有较大差距。SSD和Faster R-CNN方法的检测速度较快，但F1-score与其他基于深度学习方法相比较低；添加CBAM注意力模块后，在检测速度略有降低的同时，提升了检测精度；对于本文方法，当采用人工设置的固定

的搜索区域尺寸($h_0 \times w_0$)时，本文方法的检测速度会降低，这是由于在两次强化学习迭代过程中，当搜索区域尺寸设置较大时，如果两次搜索区域较近，会造成搜索区域的重叠，增加计算量；而当搜索区域尺寸设置较小时，又会造成检测区域过小，需要强化学习迭代更多轮次来找到所有目标。并且，不合理的尺寸设定也会对检测精度造成影响。因此，本文方法在强化学习迭代过程中通过自适应调整搜索区域的尺寸，获得了最高的F1-score，能够在保证较高检测精度的同时，提高检测速度。

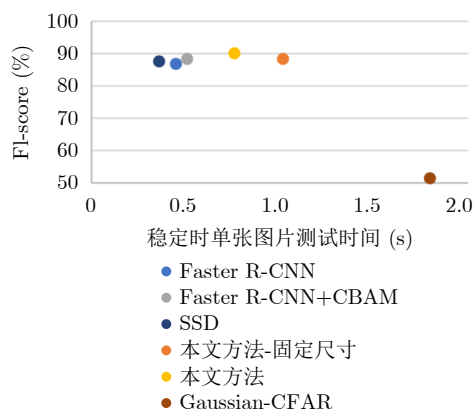


图 9 运行时间与F1-score关系

Fig. 9 Runtime versus F1-score

5 结语

本文针对SAR目标检测任务, 结合强化学习方法对Faster R-CNN模型进行改进, 提出新的候选框挑选方法来解决传统Faster R-CNN模型在检测过程中易受SAR图像复杂背景杂波影响而产生大量虚警的问题。本文所提方法利用强化学习序列决策的特点, 对大场景SAR图像中可能含有目标的区域进行迭代搜索, 通过强化学习策略确定搜索区域的位置坐标和尺寸, 最终实现对大量初始候选框的自适应挑选。实验结果表明, 所提方法能够找到含有待检测目标的区域, 提升对复杂背景杂波的鉴别能力, 有效减少虚警。除此之外, 通过在强化学习内部添加距离约束, 对搜索区域尺寸进行自适应调整, 能够在进一步提高检测性能的同时, 提高运算速度。与其他主流目标检测方法进行对比实验, 结果表明本文方法能够在增加较少运算量的同时, 大幅提升SAR目标检测精度。

参 考 文 献

- [1] BURL M C, OWIRKA G J, and NOVAK L M. Texture discrimination in synthetic aperture radar imagery[C]. Twenty-Third Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, 1989: 399–400. doi: [10.1109/ACSSC.1989.1200821](https://doi.org/10.1109/ACSSC.1989.1200821).
- [2] NOVAK L M, BURL M C, and IRVING W W. Optimal polarimetric processing for enhanced target detection[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1993, 29(1): 234–244. doi: [10.1109/7.249129](https://doi.org/10.1109/7.249129).
- [3] AI Jiaqiu, MAO Yuxiang, LUO Qiwu, *et al.* Robust CFAR ship detector based on bilateral-trimmed-statistics of complex ocean scenes in SAR imagery: A closed-form solution[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2021, 57(3): 1872–1890. doi: [10.1109/TAES.2021.3050654](https://doi.org/10.1109/TAES.2021.3050654).
- [4] LECUN Y, BOTTOU L, BENGIO Y, *et al.* Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324. doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [5] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 580–587. doi: [10.1109/cvpr.2014.81](https://doi.org/10.1109/cvpr.2014.81).
- [6] GIRSHICK R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1440–1448. doi: [10.1109/iccv.2015.169](https://doi.org/10.1109/iccv.2015.169).
- [7] REN Shaoqing, HE Kaiming, GIRSHICK R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks[C]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [8] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 779–788. doi: [10.1109/cvpr.2016.91](https://doi.org/10.1109/cvpr.2016.91).
- [9] LIU Wei, ANGUELOV D, ERHAN D, *et al.* SSD: Single shot MultiBox detector[C]. 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 21–37. doi: [10.1007/978-3-319-46448-0_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [10] 杜兰, 刘彬, 王燕, 等. 基于卷积神经网络的SAR图像目标检测算法[J]. *电子与信息学报*, 2016, 38(12): 3018–3025. doi: [10.11999/JEIT161032](https://doi.org/10.11999/JEIT161032).
DU Lan, LIU Bin, WANG Yan, *et al.* Target detection method based on convolutional neural network for SAR image[J]. *Journal of Electronics & Information Technology*, 2016, 38(12): 3018–3025. doi: [10.11999/JEIT161032](https://doi.org/10.11999/JEIT161032).
- [11] AI Jiaqiu, TIAN Ruitian, LUO Qiwu, *et al.* Multi-scale rotation-invariant Haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(12): 10070–10087. doi: [10.1109/TGRS.2019.2931308](https://doi.org/10.1109/TGRS.2019.2931308).
- [12] ZHANG Shaoming, WU Ruize, XU Kunyuan, *et al.* R-CNN-based ship detection from high resolution remote sensing imagery[J]. *Remote Sensing*, 2019, 11(6): 631. doi: [10.3390/rs11060631](https://doi.org/10.3390/rs11060631).
- [13] 杜兰, 魏迪, 李璐, 等. 基于半监督学习的SAR目标检测网络[J]. *电子与信息学报*, 2020, 42(1): 154–163. doi: [10.11999/JEIT190783](https://doi.org/10.11999/JEIT190783).
DU Lan, WEI Di, LI Lu, *et al.* SAR target detection network via semi-supervised learning[J]. *Journal of Electronics & Information Technology*, 2020, 42(1): 154–163. doi: [10.11999/JEIT190783](https://doi.org/10.11999/JEIT190783).
- [14] SUTTON R S and BARTO A G. Reinforcement Learning: An Introduction[M]. 2nd ed. Cambridge: MIT Press, 2018: 1–526.
- [15] CHUNG J. Playing Atari with deep reinforcement learning[J]. arXiv: 1312.5602. doi: [10.48550/arXiv.1312.5602](https://doi.org/10.48550/arXiv.1312.5602).
- [16] BUENO M B, GIRO-I-NIETO X, MARQUÉS F, *et al.* Hierarchical Object Detection with Deep Reinforcement Learning[M]. HEMANTH D J and ESTRELA V V. Deep Learning for Image Processing Applications. 2017: 164–1276. doi: [10.3233/978-1-61499-822-8-164](https://doi.org/10.3233/978-1-61499-822-8-164).
- [17] CAICEDO J C and LAZEBNIK S. Active object localization with deep reinforcement learning[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 2488–2496. doi: [10.1109/iccv.2015.286](https://doi.org/10.1109/iccv.2015.286).

- [18] PIRINEN A and SMINCHISESCU C. Deep reinforcement learning of region proposal networks for object detection[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 6945–6954. doi: [10.1109/cvpr.2018.00726](https://doi.org/10.1109/cvpr.2018.00726).
- [19] CHUNG J, GULCEHRE C, CHO K H, *et al.* Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv: 1412.3555, 2014. doi: [10.48550/arXiv.1412.3555](https://doi.org/10.48550/arXiv.1412.3555).
- [20] WILLIAMS R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning[J]. *Machine Learning*, 1992, 8(3/4): 229–256. doi: [10.1023/A:1022672621406](https://doi.org/10.1023/A:1022672621406).
- [21] KINGMA D P and BA J. Adam: A method for stochastic optimization[C]. 3rd International Conference on Learning Representations, San Diego, USA, 2015.
- [22] GUTIERREZ D. *MiniSAR*: A review of 4-inch and 1-foot resolution Ku-band imagery[EB/OL]. <https://www.sandia.gov/app/uploads/sites/124/2022/04/SAND2005-3706P-miniSAR-flight-SAR-images.pdf>, 2005.
- [23] AI Jiaqiu, MAO Yuxiang, LUO Qiwu, *et al.* SAR target classification using the multikernel-size feature fusion-based convolutional neural network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 5214313. doi: [10.1109/TGRS.2021.3106915](https://doi.org/10.1109/TGRS.2021.3106915).
- [24] 魏迪. 基于半监督卷积神经网络的SAR图像目标检测方法研究[D]. [硕士学位论文], 西安电子科技大学, 2020. doi: [10.27389/d.cnki.gxadu.2020.001557](https://doi.org/10.27389/d.cnki.gxadu.2020.001557).
WEI Di. The research on target detection of SAR image based on semi-supervised convolutional neural network[D]. [Master dissertation], Xidian University, 2020. doi: [10.27389/d.cnki.gxadu.2020.001557](https://doi.org/10.27389/d.cnki.gxadu.2020.001557).
- [25] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional block attention module[C]. 15th European Conference on Computer Vision, Munich, Germany, 2018: 3–19. doi: [10.1007/978-3-030-01234-2_1](https://doi.org/10.1007/978-3-030-01234-2_1).

作者简介

杜 兰, 博士, 教授, 主要研究方向为雷达目标识别、雷达信号处理、机器学习。

王梓霖, 硕士生, 主要研究方向为SAR图像目标检测、机器学习等。

郭昱辰, 博士, 讲师, 主要研究方向为智能雷达目标检测和识别。

杜宇昂, 博士生, 主要研究方向为SAR图像解译、机器学习与人工智能等。

严俊坤, 博士, 教授, 主要研究方向为单站雷达智能信号处理和网络化雷达协同探测。

(责任编辑: 高华)