



# 基于改进 RPN 的 Faster-RCNN 网络 SAR 图像车辆目标检测方法

曹磊 王强 史润佳 蒋忠进

(东南大学毫米波国家重点实验室, 南京 210096)

**摘要:** 针对传统 Faster-RCNN 方法中候选区域生成网络(RPN)模块在进行目标检测时对目标特征提取不够充分的问题,提出一种基于改进 RPN 的 Faster-RCNN 网络 SAR 图像车辆目标检测方法.首先基于 VGG-16 网络提取出图片的多层特征,然后利用卷积核对最深的 3 个特征层作进一步的特征提取和正则化处理,最后对处理后的 3 个特征层进行信息融合.利用 MSTAR 数据集中车辆目标 SAR 图像和自然背景 SAR 图像,通过图像分割和贴图的方式制作了 SAR 场景数据集,对所改进网络进行训练和测试.实验结果表明,在 SAR 图像车辆目标检测中,与传统 RPN 相比,改进 RPN 收敛速度更快,不仅将检测结果的查准率从 97.7% 提高到了 99.7%,虚警率明显降低,而且泛化性能更强.针对训练范围以外的目标,能将查准率由 98.0% 提高到 99.0%.

**关键词:** SAR 图像; 车辆目标检测; 卷积神经网络; Faster-RCNN; 候选区域生成网络

**中图分类号:** TN 957.51 **文献标志码:** A **文章编号:** 1001-0505(2021)01-0087-05

## Method for vehicle target detection on SAR image based on improved RPN in Faster-RCNN

Cao Lei Wang Qiang Shi Runjia Jiang Zhongjin

(State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China)

**Abstract:** Aiming at the problem that the region proposal network (RPN) module couldn't adequately extract target features when performing target detection in the traditional Faster-RCNN method, a method for vehicle target detection on the SAR image based on the improved RPN in Faster-RCNN was proposed. First, the multi-layer features of the image were extracted based on the VGG-16 network. Then the deepest three feature layers were further extracted and regularized using convolution kernels. Finally, the information fusion was performed on the three processed feature layers. Using the vehicle target SAR image and the natural background SAR image in the MSTAR data set, the SAR scene data set was created by image segmentation and texture, and the improved network was trained and tested. Experimental results show that in the detection of the vehicle targets on SAR images, compared with the traditional RPN, the improved RPN has a faster convergence speed, improving the accuracy of the test results from 97.7% to 99.7%, with a lower false alarm rate, and has stronger generalization performance. For targets outside the training range, the accuracy can be increased from 98.0% to 99.0%.

**Key words:** SAR images; vehicle target detection; convolutional neural network; Faster-RCNN; region proposal network (RPN)

近年来,深度学习迅猛发展,被应用到军事、地探、医疗等各个领域,并取得了很好的效果<sup>[1]</sup>.其中,卷积神经网络(CNN)<sup>[2]</sup>以其在图像处理方面优异的检测与识别能力受到了普遍的重视. CNN

不仅大量用于光学图像处理,也在 SAR 图像自动解译中表现出色<sup>[3-8]</sup>,能够高效准确地进行目标检测与识别<sup>[9]</sup>.

Girshick 等<sup>[10]</sup>提出了基于区域卷积神经网络

收稿日期: 2020-06-12. 作者简介: 曹磊(1994—)男,硕士生;蒋忠进(联系人)男,博士,副教授, zjjiang@seu.edu.cn.

引用本文: 曹磊,王强,史润佳,等.基于改进 RPN 的 Faster-RCNN 网络 SAR 图像车辆目标检测方法[J].东南大学学报(自然科学版), 2021, 51(1): 87-91. DOI: 10.3969/j.issn.1001-0505.2021.01.012.

(R-CNN) 的目标检测框架. 自此, 目标检测由基于视觉特征和统计特征为主的时期进入基于深度学习的时期, 并获得了快速的发展. 从 R-CNN 到 SPP-NET<sup>[11]</sup>、Fast-RCNN<sup>[12]</sup>, 再到 Faster-RCNN<sup>[13]</sup>, 网络的性能越来越好, 尤其 Faster-RCNN 是真正意义上的端到端深度学习目标检测算法. Faster-RCNN 最大的创新点在于设计了候选区域生成网络(RPN), 由 RPN 来提取特征和完成候选区域的筛选, 并对目标进行检测和定位, 然后将筛选出的候选框送至分类识别层进行目标分类操作, 明显提高了目标检测与识别的效率.

近年来, Faster-RCNN 逐渐成为一种重要的 SAR 图像目标检测和识别方法<sup>[14-16]</sup>. 在某些情况下, 无需知道车辆目标的具体类别, 只需完成目标检测即可. 本文便是采用 Faster-RCNN 进行 SAR 图像中的车辆目标检测. 在传统 Faster-RCNN 中, RPN 模块仅采用图像的顶层特征完成目标检测, 由于顶层特征比较抽象, 对 SAR 图像的细节信息丢失较为严重, 不利于目标与地物的区分. 因此, 本文以 VGG-16 网络作为基础网络提取多个特征层, 并对 RPN 进行改进, 将细节信息比较丰富的较低层特征与信息抽象的顶层特征进行处理和融合, 用于 SAR 图像中的车辆目标检测, 并取得了良好效果.

## 1 改进的 RPN 与目标检测

传统 RPN 的结构如图 1 所示. 首先采用 VGG-16 网络, 经过每层卷积核卷积操作之后会得到一系列特征图, 这里称之为特征层. 在本文中, 一共生成了 5 个特征层, 此处分别命名为  $L_1$ 、 $L_2$ 、 $L_3$ 、 $L_4$  和  $L_5$ , 每个特征层宽度、高度和厚度不同. 将最深特征层  $L_5$  进行  $3 \times 3$  卷积操作后得到特征层  $L_5^c$ , 作为最终的候选框提取层.

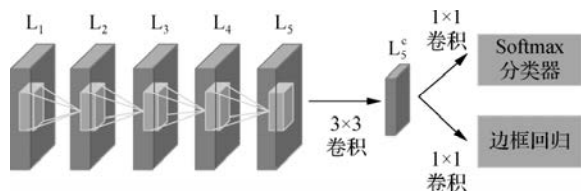


图 1 传统 RPN 结构图

在训练期间, 将候选框提取层  $L_5^c$  进行滑动处理, 提取候选框. 该过程中, 如文献[16]所述, 需要为每个像素点设置多个长宽比和大小均不同的锚点, 计算出各个锚点对应的原图区域与目标真实边框的交并比(IoU), 即两区域交集面积与并集面积的比值. 将 IoU 高于 0.7 的视为正样本, 低于 0.3 的视为负样本, 正负样本作为训练样本, 即候选区

域, 其他样本不参与训练. 然后将候选区域及其边框信息分别送进二分类网络和边框回归网络, 用于训练网络的目标辨识和定位功能.

RPN 网络主要学习 2 部分信息: ①类别信息, 根据候选区域特征信息计算候选区域类别信息, 判断其中是否包含目标, 若包含即为前景类, 否则为背景类; ②位置信息, 即通过候选区域边框回归出目标位置边框. 事实上, RPN 并未直接对位置信息本身进行训练学习, 而是学习候选区域边框相对于目标位置边框的偏移量.

传统 RPN 仅仅以最深特征层作为候选框提取层, 这会导致经过多次下采样后, 最深特征层里的图像细节信息损失较多, 目标特征信息损失较大, 带来一定的目标定位误差.

本文改进版 RPN 结构如图 2 所示. 首先, 将细节信息较丰富的  $L_3$ 、 $L_4$  特征层用  $3 \times 3$  卷积核作进一步特征提取及 L2 正则化处理, 得到 2 个新的特征层  $L_3^c$  和  $L_4^c$ . 然后, 特征层  $L_5$  经过  $3 \times 3$  卷积核卷积后得到特征层  $L_5^c$ , 将  $L_5^c$  与特征层  $L_3^c$  和  $L_4^c$  进行信息融合得到新的特征层  $L_6$ , 作为最后的候选框提取层.  $L_6$  的厚度与特征层  $L_5^c$  相同, 因此这种方式可以在不增加候选框提取层厚度的情况下, 利用较低特征层中的图像细节信息并结合较高特征层中的抽象特征信息, 提高目标辨识和定位的可靠性.

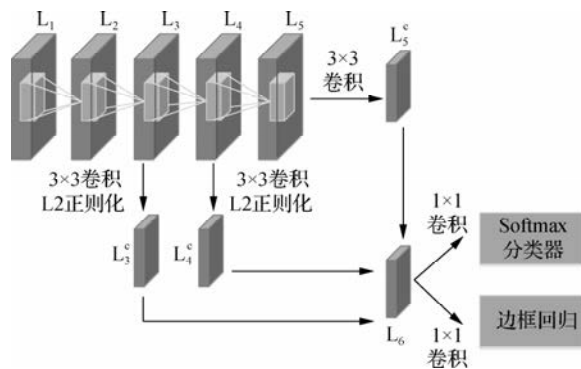


图 2 改进 RPN 结构图

本文将特征层  $L_5^c$  与  $L_3^c$  及  $L_4^c$  进行信息融合得到特征层  $L_6$ . 这里是指 3 个长宽相等的特征层对应位置的特征值相加, 得到拥有相同特征值个数新的特征层. 在候选框提取层确定以后, 将采用 RPN 网络提取候选框, 分别送入分类网络和边框回归网络进行目标辨识功能训练和边框回归功能训练, 其方法与传统的 RPN 网络训练相同.

在测试阶段, 每张测试集图片均会被检测出许多感兴趣区域(ROI), 而每一个目标也会被不止一个 ROI 包围. 本文采用非极大值抑制(NMS)方法

去除冗余的 ROI。首先将 ROI 按照目标辨识中所得到的分数进行降序排列,选中其中分数最高的 ROI,并遍历其他 ROI,若其他 ROI 与所选中 ROI 的 IoU 大于某阈值(本文实验中阈值设为 0.8),则认定此 ROI 与所选中 ROI 是针对同一目标,将此 ROI 删除;然后为剩下的其他 ROI 重复上述去冗余操作,确保每一个目标只被一个 ROI 包围。

## 2 实验数据集的制作

由于包含军事目标的大场景 SAR 图像较为稀缺,因此本文在 MSTAR 数据集基础上,制作带车辆目标的大场景数据集,用于对卷积神经网络的训练与测试。MSTAR 数据集是美国国防高级研究计划局(DARPA)提供的实测 SAR 图像数据集,其中包含了大量的车辆目标图像和少量的大场景图像。本文将这些大场景图像作为车辆目标检测的背景使用,由于大场景图像尺寸过大,为  $1500 \times 800$  像素,在制作训练数据时,将其分割成多个大小为  $300 \times 300$  像素的小场景图像,如图 3 所示。

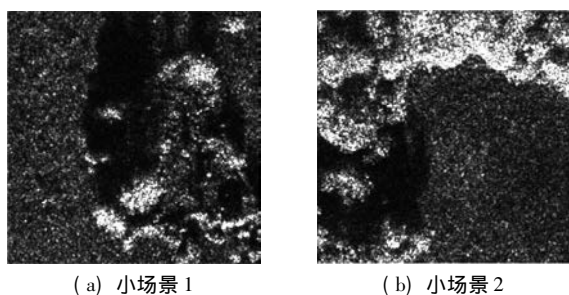


图3 没有目标的小场景 SAR 图像

MSTAR 数据集中共有 10 类车辆目标的 SAR 图像,图像大小为  $128 \times 128$  像素。这些目标多是不同型号坦克、装甲车在  $0 \sim 360^\circ$  方位角下的静止图像。本文挑选其中 8 类目标图像作预处理后,将其贴入小场景背景图像,生成训练数据集。图 4 展示了 8 类车辆目标 2S1、BMP2、BRDM\_2、BTR\_60、BTR\_70、T62、T72、ZSU\_23\_4 的 SAR 图像样本。

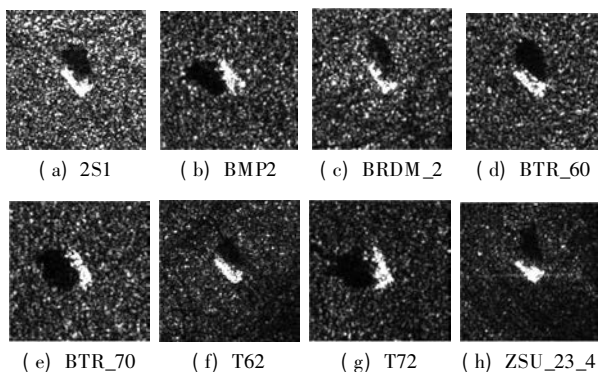


图4 8类车辆目标图像样本

本文的贴图操作是由编写的程序自动完成,不需要手工介入,因此可以快速大批量地生成实验数据集。程序会自动将目标贴入坐标等重要信息写入标签,因此本文并未采用 Pascal VOC 数据集中标签的格式,而是自拟了一种更简洁的标签格式,格式中包含目标的种类、位置和尺寸等信息。图 5 展示了贴上目标以后的小场景背景图像,可以看出,背景与目标的融合效果很好。

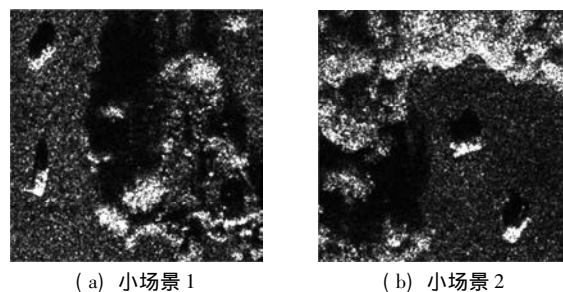


图5 贴上目标的小场景 SAR 图像

在制作数据集时,在每张小场景背景图中随机贴 2 类军事目标,如图 5(a) 和 (b) 所示,共生成 20 294 个训练样本和 894 个测试样本。为了验证改进 RPN 的目标检测性能,测试样本中不乏环境地物非常复杂的场景。

## 3 代码实现与实验分析

本文的目标检测代码基于 Python 语言和 TensorFlow 框架编写。为了缩短训练时间,用于提取图片特征的基础网络是已在 ImageNet 数据集上经过预训练的 VGG-16 网络。本文直接利用 VGG-16 网络的模型参数对所提出的改进 RPN 网络进行初始化,然后再利用 MSTAR 实验数据集进行训练。在目标检测代码中,本文为每个像素点设置了 9 个锚点,采用的长宽比包括  $1:1$ ,  $1:2$ ,  $2:1$ 。操作系统是 64 位 Windows 10,计算机处理器 Intel(R) Core(TM) i7-8700 CPU @ 3.20 GHz,内存容量 16 GB,显卡是 NVIDIA GeForce RTX2070 (8 GB)。

训练中 8 类目标场景图、共 20 294 个训练样本前 800 次迭代的损失函数值见图 6。由图可知,改进 RPN 的收敛速度比传统 RPN 更快。

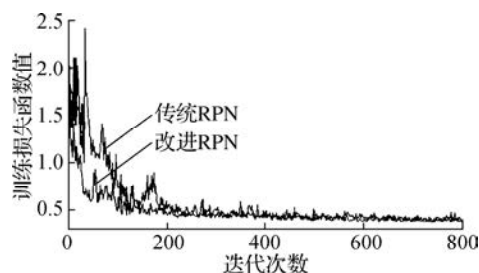


图6 训练中的前 800 次损失函数值

在评价指标方面,本文首先统计检出个数  $N_{TP}$ 、漏检个数  $N_{FN}$ 、虚警个数  $N_{FP}$  等直观结果,然后采用查全率  $r_R$ 、查准率  $r_P$  及查全率与查准率的调和平均数  $F_1$  三个指标衡量检测效果,即

$$r_R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (1)$$

$$r_P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (2)$$

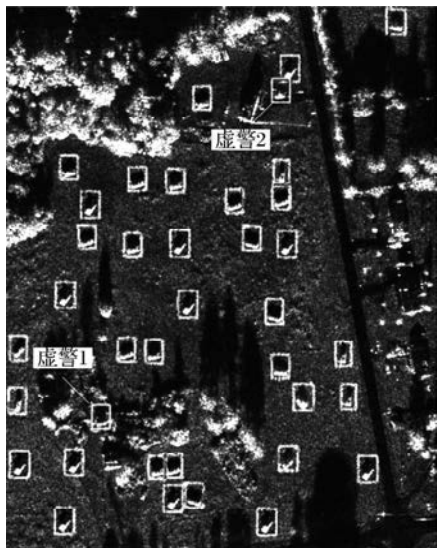
$$F_1 = \frac{2r_R r_P}{r_R + r_P} \quad (3)$$

如表 1 所示,测试集中的目标总数共 1 788 个,传统 RPN 与改进 RPN 基本均不会漏检,查全率  $r_R = 100\%$ 。但是改进 RPN 在虚警方面有所改善,传统 RPN 的查准率  $r_P = 97.7\%$ ,而改进 RPN 的  $r_P$  达 99.7%。传统 RPN 的调和平均数  $F_1 = 0.988$ ,而改进 RPN 的  $F_1 = 0.998$ 。

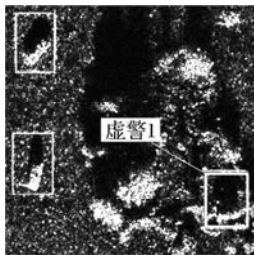
表 1 检测性能比较

方法	$N_{TP}$	$N_{FN}$	$N_{FP}$	$r_R / \%$	$r_P / \%$	$F_1$
传统 RPN	1 788	0	42	100	97.7	0.988
改进 RPN	1 788	0	4	100	99.7	0.998

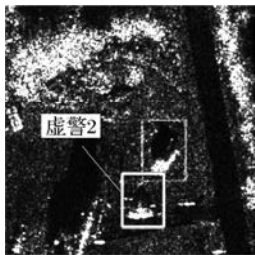
图 7 展示了传统 RPN 方法检测结果图,这里的大场景图是第 2 节中检测后的小场景按照分割



(a) 带有多个目标的复杂背景



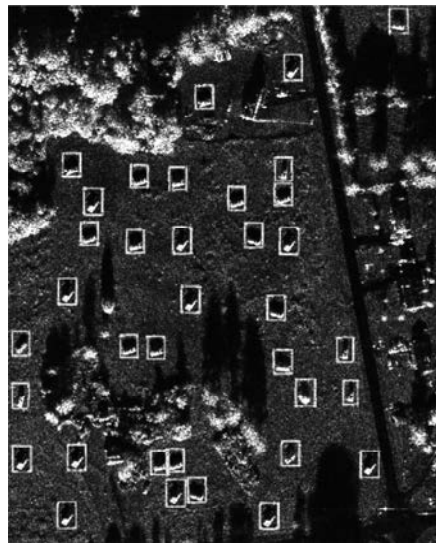
(b) 虚警 1 放大图



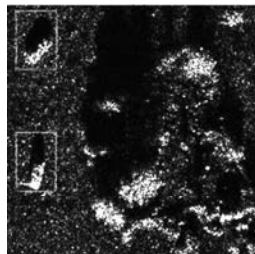
(c) 虚警 2 放大图

图 7 传统 RPN 检测结果图

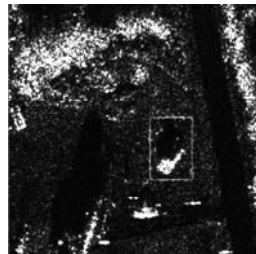
顺序和边界重新组合而成,其中 2 个虚警目标称为虚警 1 和虚警 2。改进 RPN 方法检测结果如图 8 所示,同样的区域则未出现这 2 个虚警。因此,传统 RPN 对于复杂地域(如树林、石块区等地物分布复杂区域)的检测虚警率更高。



(a) 带有多个目标的复杂背景



(b) 虚警 1 消失



(c) 虚警 2 消失

图 8 改进 RPN 检测结果图

为测试改进 RPN 的泛化能力,本文采用 8 类目标中的 5 类目标场景图作为训练集,共 13 453 个样本,剩余 3 类目标的场景图作为测试集。该训练集均匀地包含 2S1、BMP2、BRDM\_2、BTR\_60、BTR\_70 五类车辆目标,测试集则有 1 206 个,仅包含 T62、T72、ZSU\_23\_4 三类目标。传统 RPN 和改进 RPN 泛化能力测试的实验结果对比见表 2。

表 2 泛化性能比较

方法	$N_{TP}$	$N_{FN}$	$N_{FP}$	$r_R / \%$	$r_P / \%$	$F_1$
传统 RPN	2 410	2	49	99.9	98.0	0.989
改进 RPN	2 411	1	25	99.9	99.0	0.994

由表 2 分析可知 2 种方法对于未训练过的车辆目标的检出效果差异不大,传统 RPN 与改进 RPN 查全率  $r_R$  均超过 99.9%,仅有  $4 \times 10^{-6}$  的差值。但改进 RPN 的虚警个数仅约为传统 RPN 的 1/2,两者的查准率  $r_P$  分别为 99.0% 和 98.0%。综

上所述,改进RPN的泛化能力也相对更好。

## 4 结论

1) 与传统RPN相比,改进RPN在进行网络训练时具有更快的收敛速度,节省一定训练时间。

2) 2个版本的RPN在目标检测方面的性能相当,但改进RPN的虚警更少,将查准率由97.7%提高到99.7%。

3) 与传统RPN相比,改进RPN具有更好的泛化能力,针对训练范围以外的目标,能将查准率由98.0%提高到99.0%。

## 参考文献 (References)

- [1] 郑远攀,李广阳,李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用, 2019, 55(12): 20-36.  
Zheng Y P, Li G Y, Li Y. Survey of application of deep learning in image recognition[J]. *Computer Engineering and Applications*, 2019, 55(12): 20-36. (in Chinese)
- [2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90. DOI: 10.1145/3065386.
- [3] 徐丰,王海鹏,金亚秋. 深度学习在SAR目标识别与地物分类中的应用[J]. 雷达学报, 2017, 6(2): 136-148.  
Xu F, Wang H P, Jin Y Q. Deep learning as applied in SAR target recognition and terrain classification[J]. *Journal of Radars*, 2017, 6(2): 136-148. (in Chinese)
- [4] 田壮壮,占荣辉,胡杰民,等. 基于卷积神经网络的SAR图像目标识别研究[J]. 雷达学报, 2016, 5(3): 320-325.  
Tian Z Z, Zhan R H, Hu J M, et al. SAR ATR based on convolutional neural network[J]. *Journal of Radars*, 2016, 5(3): 320-325. (in Chinese)
- [5] 邹浩,林赞,洪文. 采用深度学习的多方位角SAR图像目标识别研究[J]. 信号处理, 2018, 34(5): 513-522.  
Zou H, Lin Y, Hong W. Research on multi-aspect SAR images target recognition using deep learning[J]. *Journal of Signal Processing*, 2018, 34(5): 513-522. (in Chinese)
- [6] 薛媛. 基于深度神经网络的SAR自动目标识别方法研究[D]. 成都: 电子科技大学, 2019.  
Xue Y. *Research on SAR automatic target recognition based on deep neural network*[D]. Chengdu: University of Electronic Science and Technology of China, 2019. (in Chinese)
- [7] 刘彬. 基于卷积神经网络的SAR图像目标检测及分类方法研究[D]. 西安: 西安电子科技大学, 2017.
- [8] 张笑. 基于深度学习的SAR图像目标识别算法研究[D]. 南京: 南京航空航天大学, 2018.  
Zhang X. *Research on SAR image target recognition based on deep learning*[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2018. (in Chinese)
- [9] Shang R H, Wang J M, Jiao L C, et al. SAR targets classification based on deep memory convolution neural networks and transfer parameters[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(8): 2834-2846. DOI: 10.1109/jstars.2018.2836909.
- [10] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// *27th IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA, 2014: 580-587. DOI: 10.1109/CVPR.2014.81.
- [11] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916. DOI: 10.1109/tpami.2015.2389824.
- [12] Girshick R. Fast R-CNN[C]// *Proceedings of the IEEE Conference on Computer Vision*. Santiago, Chile, 2015: 1440-1448.
- [13] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149. DOI: 10.1109/tpami.2016.2577031.
- [14] 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.  
Du L, Liu B, Wang Y, et al. Target detection method based on convolutional neural network for SAR image[J]. *Journal of Electronics & Information Technology*, 2016, 38(12): 3018-3025. (in Chinese)
- [15] 王思雨,高鑫,孙皓,等. 基于卷积神经网络的高分辨率SAR图像飞机目标检测方法[J]. 雷达学报, 2017, 6(2): 195-203.  
Wang S Y, Gao X, Sun H, et al. An aircraft detection method based on convolutional neural networks in high-resolution SAR images[J]. *Journal of Radars*, 2017, 6(2): 195-203. (in Chinese)
- [16] 常沛,夏勇,李玉景,等. 基于CNN的SAR车辆目标检测[J]. 雷达科学与技术, 2019, 17(2): 220-224, 231.  
Chang P, Xia Y, Li Y J, et al. SAR vehicle target detection based on CNN[J]. *Radar Science and Technology*, 2019, 17(2): 220-224, 231. (in Chinese)