



# 一种基于场景合成和锚点约束的 SAR 目标检测网络

## 摘要

随着深度学习方法在计算机视觉领域的崛起,如何将其应用于具有全天时、全天候等优点的 SAR 图像也成为一大研究重点.相较于传统图像,SAR 图像由于其难判读、应用人群较少等原因难以获得大量标注数据.本文提出一种基于场景合成和锚点约束的 SAR 目标检测方法.通过区域生长算法和阈值法对 SAR 车辆目标及其阴影进行分割,然后随机嵌入 SAR 复杂场景中的合理区域来合成目标检测数据集.针对 SAR 车辆目标的几何特性、图像分辨率参数,对 Faster-RCNN 中的锚点大小进行约束,减少不符合 SAR 车辆目标检测框尺寸的候选框,大量约简冗余计算,提升训练、测试效率及精度.

## 关键词

SAR 图像;目标检测;区域生长;数据生成;锚点约束

中图分类号 TP751

文献标志码 A

收稿日期 2019-10-21

资助项目 国家自然科学基金(61871413、61801015)

## 作者简介

金啸宇,男,硕士生,主要从事 SAR 目标检测方向研究.BUCT\_jin@163.com

尹婧(通信作者),女,博士,副教授,主要从事极化 SAR 图像处理方向研究.yinq@mail.buct.edu.cn

## 0 引言

SAR 图像由 SAR (Synthetic Aperture Radar,合成孔径雷达)系统产生,是一种主动的对地观测系统,可以安装于飞机、无人机、卫星等多种平台上,并且由于其成像所用波段相较于可见光而言受天气、昼夜等影响较少,因此可以实现全天时、全天候的观测,同时也具有一定的地表穿透能力,故而其广泛应用于海情监视、灾区勘探等领域.

在获取 SAR 图像后需要对其中的信息进行提取,而整个信息提取的过程对目标所在位置的自动检测是必要步骤.因为在获取目标位置后才便于进行后续的如切片、识别、场景理解等工作.传统的 SAR 图像检测算法中,基于统计模型的算法,以 CFAR 算法<sup>[1-2]</sup>(Constant False-Alarm Rate)的应用最为广泛.该算法主要根据雷达杂波背景的程度,建立目标场景的像素值分布模型,通过自适应地调整目标的检测阈值,从而判断目标像素的类别.此类方法也在不断优化和改进,如使用空间增强描述符和基于密度的空间聚类的舰船检测方法<sup>[3]</sup>.然而该方法要求场景中目标与背景具有较高的对比度,在场景较为简单的情况下,能够获得较好的检测结果,如海洋场景目标检测任务,但是在针对陆地场景进行检测时由于存在多种干扰目标(房屋、树木等),通常难以获得令人满意的检测结果.

卷积神经网络自提出以来,因其高效、高精度等优点被广泛应用于计算机视觉领域.而在目标检测方向上,同样有众多基于卷积神经网络的检测算法被提出,其中较为经典的是 Girshick 等提出的 R-CNN<sup>[4]</sup>网络以及众多基于 R-CNN 网络的改进型网络.在 R-CNN 网络中,首先使用选择搜索的方法提取备选区域,然后使用 CNN 网络进行特征提取并输入到 SVM 中进行分类,并同时备选区域做回归以获得更好的检测结果.为优化 R-CNN 网络中的大量冗余计算以及分步式训练过程,Girshick 又提出了 Fast R-CNN 网络<sup>[5]</sup>,该网络整合了整个流程并且共享了卷积操作,大大降低了计算量.在此基础上,Ren 等提出了 Faster R-CNN<sup>[6]</sup>,将备选区域(在该文章中称之为锚点,即 anchor)的提取过程也融合到网络模型中,使得模型可以进行一体化的训练,同时也提高了性能.在 SAR 图像领域,同样有研究人员致力于使用深度神经网络代替传统算法进行目标检测,如使用改进型卷积的舰船检测方法<sup>[7]</sup>、基于卷积神经网络的高分辨率图像检测<sup>[8]</sup>、使用级联网络的大场景检测方法<sup>[9]</sup>等.

1 北京化工大学 信息科学与技术学院,北京,100029

2 中国科学院空天信息创新研究院,北京,100094

然而由于 SAR 图像数据较少,并且解译难度大,导致其带标签数据较少,训练容易过拟合.针对该情况,本文使用区域生长算法<sup>[10]</sup>对切片数据(该切片包含目标、阴影以及一小块背景区域)进行图像分割,取出其主要目标以及阴影区域,接着按一定分布规律替换目标场景中的部分像素点,以此生成数据集以供模型进行训练.在此基础上,由于产生的数据集较为充分,并且其目标大小基本为固定值,可以对所选用的模型进行相适应的改进,即固定锚点的大小.基于实测数据集的实验结果表明,上述操作能有效降低网络的训练时间.

## 1 算法介绍

本节对算法做具体介绍,1.1 节介绍数据集生成方法,1.2 节介绍目标检测网络.

### 1.1 数据集生成方法

总体算法流程如图 1 所示.

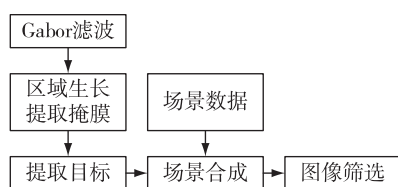


图 1 数据集生成流程

Fig. 1 Dataset generation process

依上述算法流程,首先需要对其对 MSTAR 坦克切片数据集的图片进行滤波,其原因在于目标图像要素基本可以分为阴影、坦克本体以及背景,且其像素灰度值区别较大,故可以使用 Gabor 算法依照多个角度进行滤波;然后使用区域生长算法提取掩膜.

区域生长算法的基本思想是将具有相似性质的像素合并为一个区域.首先需要对其对目标图像选取一个种子像素点作为生长的起点,然后根据像素点的相似性将种子像素点周围相似性高的像素点合并到分割区域中,然后将新并入的像素点选为新的种子像素点并重复上述过程,直到没有新像素点被并入分割区域,获得掩膜;然后将掩膜覆盖至场景中进行合成并对合成结果进行筛选.

#### 1.1.1 Gabor 滤波<sup>[11]</sup>

考虑到 Gabor 滤波器可以在频域不同尺度、不同方向上提取相关特征<sup>[12]</sup>,为了增强边界信息,弱化边界内部信息,首先使用 Gabor 滤波器对图像进行多角度的滤波增强,然后融合多角度的滤波结果.

Gabor 滤波器公式如下:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = e^{-\frac{x'^2 + y'^2/2}{2\sigma^2}} \cos\left(2\pi \frac{x'}{\lambda} + \psi\right),$$

其中参数  $\lambda, \theta, \psi, \sigma, \gamma$  分别代表正弦函数波长、Gabor 核函数的方向、相位偏移、高斯函数的标准差和滤波器长度.在本文的实验中,使用步长为  $\pi/3$  的角度偏移,融合了 6 步总计  $2 \times \pi$  的多角度图像,其结果如图 2 所示.

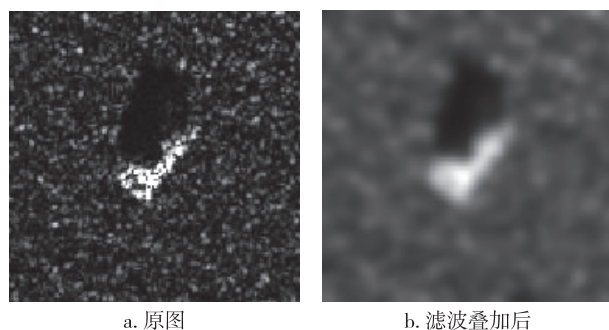


图 2 Gabor 滤波结果示意图

Fig. 2 Gabor filtering result

在图 2 中,左图是原始切片图像,右图是滤波叠加后的结果,可以看到在经过不同的相移滤波叠加后图像的原始三要素:阴影、坦克本体和背景已经有了更加明显的区分度,同时其对应区域的像素值分布也变得更为均匀,便于选择区域生长算法的种子像素点.

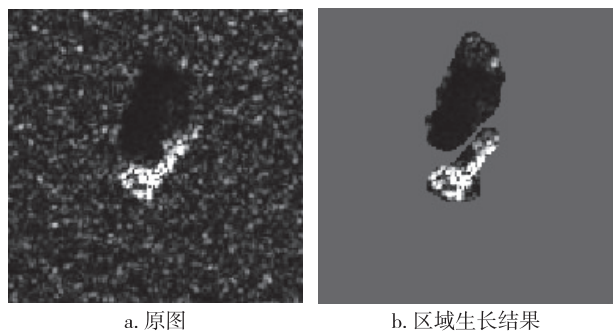
#### 1.1.2 区域生长算法

由于在上一步的滤波算法处理后,图像要素的分布已经具有较高的区分度,故可以直接按照其直方图分布选取对应波峰所对应的像素点的位置.考虑到在生成数据集时需要将阴影和坦克本体共同放置到场景中,需要选取两个种子像素点,分别对应于阴影和坦克本体所在区域.

同时,为了将坦克和阴影区域与背景区分开,必须设置一个生长终止条件.在观察图像直方图分布后,选取波谷中心点与波峰的平均距离(即灰度值差 20)作为生长终止条件.最终的分割结果如图 3 所示,灰色区域为背景,纯黑色部分为阴影区域,黑白相间区域是原图中坦克像素区域的复制,可以看到区域生长算法较好地地区分了原图中的三要素.

#### 1.1.3 战场环境模拟

在获得了切片数据的阴影-坦克掩膜后,随机替换战场环境中的部分像素点以生成数据集,同时对生成的数据集进行一定的筛选工作,剔除生成的坦克位于不合适的位置如森林、建筑物的图像.



a. 原图 b. 区域生长结果

图3 区域生长结果示意图

Fig. 3 Regional growing result

## 1.2 基于 Faster R-CNN 的检测网络

在现有的检测方法中,大致可以分为单步式和多步式的网络.单步式的网络如 SSD<sup>[13]</sup>、YOLO<sup>[14]</sup> 等的基本思路是将图像分块,然后进行边框回归以及分类;多步式的网络基本是 R-CNN 的改进型,其基本思路是首先生成锚点,然后对锚点进行边框回归以及分类.单步式网络的分类直接输出的是对应每个类别的分类概率,多步式网络的第一步输出是目标是否是背景的概率,即进行的是一个二分类任务,然后在获得了目标的大致位置后,将其输入到后续的网络中,再进行位置的精细调整以及对应多个类别的分类.图 4 为基于锚点约束的 Faster R-CNN 的示意图.考虑到本文中的改进点基本在 RPN 网络,故下文中重点关注 RPN 网络的部分.

### 1.2.1 RPN 网络基本思路

RPN 网络的基本思路为利用 CNN 网络提取的特征,对输入的锚点进行边框回归以及分类操作.

在具体的实现中,我们使用 resnet<sup>[15]</sup> 作为基础网络,该网络引入了跳跃连接,使得上一个残差块的

信息能直接流入下一个残差块,该操作提高了信息流通性,加快了网络的收敛速度,同时也避免了由于网络过深引起的梯度消失和退化问题,为之后的优化操作提供了基础.

在使用 resnet 提取特征后,将其输出的特征逐层连接至 RPN 网络.首先经过一个  $3 \times 3$  的卷积层,随后分两路通过  $1 \times 1$  的卷积层,一路输出为  $[h, w, anchors\_per\_location, 2]$ ,其中  $h, w$  分别为特征图的高和宽,这一路的输出对应的是 *anchor* 的分类结果,即区分该 *anchor* 对应的是前景还是背景,另外一路的输出为  $[h, w, anchors\_per\_location, 4]$ ,第 4 个维度的值分别为:  $dx, dy, dh, dw$ ,其中  $dx, dy$  代表中心点的位移,  $dh, dw$  代表高和宽的缩放量.

为使网络能同时对上述的输出进行反向传播, RPN 网络使用多任务损失函数,该函数包括分类损失函数以及边框回归损失函数,其定义式为

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, y_i) + \lambda \frac{1}{N_{reg}} \sum_i y_i L_{reg}(t_i, t_i^*),$$

其中  $\lambda$  表示平衡系数,用于平衡两个损失函数的比重,  $i$  代表数据集中某一图像的索引,  $L_{cls}$  代表分类损失函数,由于 RPN 网络的任务是进行二分类(前景和背景),故而使用对数似然函数,  $L_{reg}$  代表边框回归函数,该函数使用  $L2$  范数作为损失函数.

### 1.2.2 RPN 网络训练

为训练 RPN 网络,首先需要对数据集进行一定的处理再输入到网络中.具体的操作步骤是先获得真实的目标标签以及所在位置,然后依照不同尺度以及不同长宽比生成大量的锚点,将锚点与真实的目标框进行 IoU 计算,与真实目标框的 IoU 交叠大于 0.7 的锚点设置为前景, IoU 交叠小于 0.3 的锚点

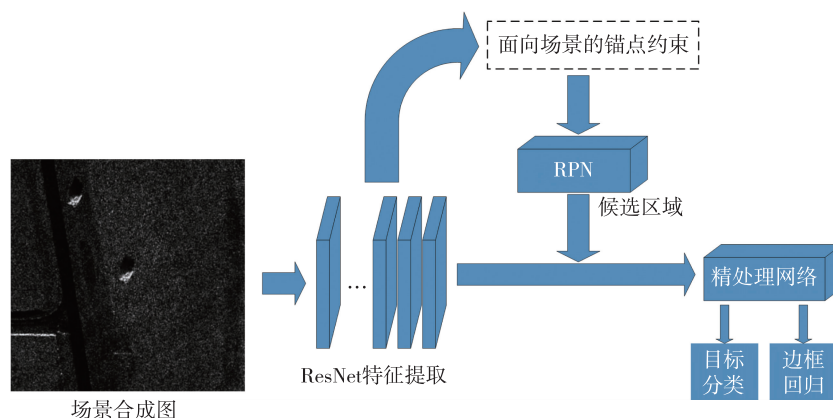


图4 基于锚点约束的 Faster R-CNN 网络

Fig. 4 Sketch of Faster R-CNN based on anchor constraint

设置为背景,参与 RPN 网络的训练.针对本实验的情况,考虑到数据集中的图像是由切片数据和场景图像合成,其目标的大小基本固定,故在生成锚点时使其大小固定为  $128 \times 128$  像素.选择该大小的原因在于使用 32 倍数大小的数据有利于 cuda 加速,并且与真实的切片数据大小相同(真实切片数据尺寸均为  $128 \times 128$  像素).

## 2 实验结果与分析

本节对实验中所用的数据集以及具体的实验设置进行了描述,所用系统是 Win10, CPU 为 Intel i5-8400 2.8 GHz, 内存大小为 8 GB, 所用 GPU 为 GTX 960, 显存大小为 6 GB, 使用的深度学习框架是 Tensorflow.

### 2.1 数据以及实验设置

本文的实验所使用的数据集为 MSTAR 数据集,该数据集大多是静止车辆的切片 SAR 图像数据,其中包含车辆目标在不同方位角下获取到的图像.该数据集同时包含推荐使用的训练集以及测试集,训练集为雷达工作俯仰角为  $17^\circ$  时获得的目标图像数据,共有 3 大类,测试集为俯仰角为  $15^\circ$  时获取的图像数据,同样包含 3 大类.除此之外,该数据集还包含一定的环境场景数据,其图像是雷达在条带模式下得到的大小不完全相同的背景图像,该数据集即是本文中合成场景时所使用的背景图像.数据集的总体分布为:训练集

700 张,测试集 200 张,验证集 100 张.图 5 是数据集展示,其大小为  $512 \times 512$  像素.

### 2.2 生成数据集对比试验

本节通过对比双阈值法和区域生长方法在进行区域分割任务时的表现,以显示本文采用方法的有效性.

在对比实验中,均需要首先对图像进行滤波处理,选择的滤波器为 Gabor 滤波器,其中带宽设置为 4,滤波器长度设为 0.5,波长设为 16,滤波器角度设为 0,角度分 6 步从 0 递增至  $2 \times \pi$ ,在获得 6 通道的数据后叠加并归一化至最大值为 255,以便于显示.

本文使用了双阈值分割方法<sup>[12]</sup>,在对 Gabor 滤波处理后的图像进行分析后发现,滤波增强后的图像的灰度直方图共分为 3 个部分,即低灰度、中灰度以及高灰度区域,分别对应的是原图中的阴影区域、背景以及目标所在区域,且大致呈现高斯分布特性,据此构造分割区域:

$$\begin{cases} T_s = \mu - k_1 \times \sigma, \\ T_t = \mu + k_2 \times \sigma, \end{cases}$$

其中  $T_s, T_t$  分别为阴影和目标区域分割的阈值, $\mu$  为滤波图像灰度均值, $\sigma$  为滤波图像方差, $k_1, k_2$  为调整因数.通过实验调整式中的因数,选择能使整个数据集的分割结果均较好的参数,其对比试验结果如图 6 所示.

从对比试验结果中可以看出,双阈值分割方法

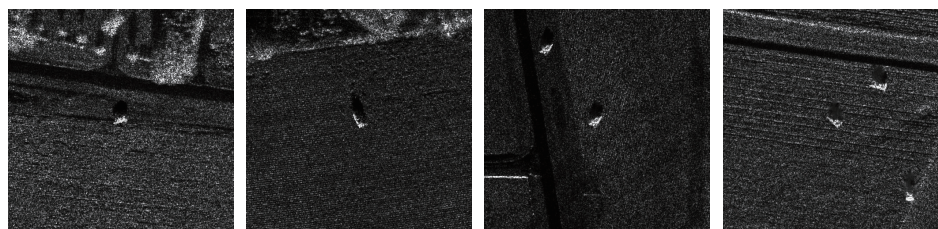
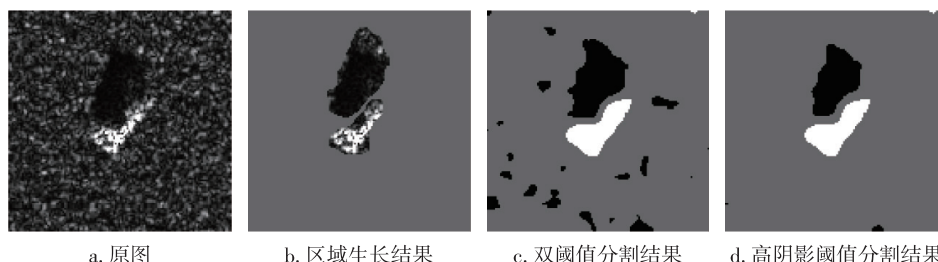


图 5 数据集部分样本示意图

Fig. 5 Example of MSTAR dataset



a. 原图

b. 区域生长结果

c. 双阈值分割结果

d. 高阴影阈值分割结果

图 6 区域生长法和阈值法的分割实验结果对比

Fig. 6 Segmentation result comparison between region growth and thresholding methods

对于图像灰度分布较为敏感,固定调整因子的情况下某些像素点容易超出阴影阈值而被划入到阴影区域中.为减少该现象的出现需要调高阴影分割阈值,但该操作会同时减少阴影区域.相对而言区域生长的结果更接近真实图像.

### 2.3 检测试验结果对比

在本实验中,主要对比的是约束锚点大小后检测速度的提升.原方法中由于需要多种不同尺度、不同形状的目标进行检测,故而在生成锚点时,依照作为基础特征提取网络的 resnet 的原图-特征图缩放比例,取其中第 2~6 个卷积池化模块的缩放比例,即 [4,8,16,32,64],对应的锚点大小 [32,64,128,256,512] 以及不同的宽高比例 [0.5,1,2],以图像大小为 [512,512] 为例,共形成 26 1888 个锚点.

考虑到生成数据集时使用的切片数据均为 128×128 像素大小,上述过程中产生的锚点事实上为网络添加了大量的冗余计算,因此本文约束锚点的宽度为 128 个像素点,同时约束其宽高比例为 1:1,如此仅产生 4 096 个锚点,大大加快了网络的收敛速度,对比结果展示于表 1.

表 1 Faster R-CNN、YOLO、锚点约束方法结果对比

Table 1 Result comparison between Faster R-CNN, YOLO, and anchor constraint method

	测试时间/s	召回率/%	准确率/%	检测数
Faster R-CNN	175	98.4	53.6	482
锚点约束	140	98.5	56.2	460
YOLOV3	15.7	86.4	91.1	243

在表 1 中,所有数据均是多次重新训练网络后得出的平均值,其中单个 epoch 的 step 为 50,训练时间为多个 epoch 平均后的结果,测试时间是使用训练好的权重对测试集所有图像进行一遍检测所需的平均时间.可以看到由于 Faster R-CNN 系列的方法具有富集式检测的特点,其召回率较高,但同时也容易出现误检导致其准确率较低;而对比实验中的 YOLOV3 网络的单步式检测以及较少的候选框数量使得其召回率稍低而准确率对比 Faster R-CNN 系列有所提升.对比改进前后方法的测试时间可以看到锚点固定的方法有效地降低了测试时间并提升检测性能.

### 3 结论

针对目标大小固定的检测问题,本文提出了一

种基于锚点约束的 Faster R-CNN 改进型网络.为验证本方法的有效性,首先使用 Gabor 滤波算法对 MSTAR 数据集中的切片目标图像进行滤波,接着使用区域生长算法对目标以及阴影区域分割,提取其掩膜,然后随机替换目标场景中的部分像素点以完成数据集制作,最后对比了原算法以及改进后的算法.其实验结果表明使用锚点约束的方法不仅能够有效降低训练测试时间,同时其检测性能也略有提高.

### 参考文献

#### References

- [1] 艾加秋,齐向阳,禹卫东.改进的 SAR 图像双参数 CFAR 舰船检测算法[J].电子与信息学报,2009,31(12):2881-2885  
AI Jiaqiu, QI Xiangyang, YU Weidong. Improved two parameter CFAR ship detection algorithm in SAR images [J]. Journal of Electronics & Information Technology, 2009, 31(12): 2881-2885
- [2] 王娜,时公涛,陆军,等.一种新的极化 SAR 图像目标 CFAR 检测方法[J].电子与信息学报,2011,33(2):395-400  
WANG Na, SHI Gongtao, LU Jun, et al. A new polarimetric SAR image CFAR target detecting method [J]. Journal of Electronics & Information Technology, 2011, 33(2): 395-400
- [3] Lang H T, Xi Y Y, Zhang X. Ship detection in high-resolution SAR images by clustering spatially enhanced pixel descriptor [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(8): 5407-5423
- [4] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587
- [5] Girshick R. Fast R-CNN [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448
- [6] 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
- [7] Li J W, Qu C W, Shao J Q. Ship detection in SAR images based on an improved faster R-CNN [C] // SAR in Big Data Era: Models, Methods and Applications (BIGSAR-DATA), 2017, DOI: 10.1109/BIGSARDATA.2017.8124934
- [8] 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 Siyu, GAO Xin, SUN Hao, et al. Aircraft target detection method for high resolution SAR image based on convolutional neural network [J]. Journal of Radars Sinica, 2017, 6(2): 195-203
- [9] 陈慧元,刘泽宇,郭炜炜,等.基于级联卷积神经网络

- 的大场景遥感图像舰船目标快速检测方法[J].雷达学报,2019,8(3):413-424  
CHEN Huiyuan, LIU Zeyu, GUO Weiwei, et al. Fast detection method of ship target based on cascaded convolutional neural network for large scene remote sensing image[J]. Journal of Radars Sinica, 2019, 8(3):413-424
- [10] 范伟.基于区域生长的彩色图像分割算法[J].计算机工程,2010,36(13):192-193  
FAN Wei. Color image segmentation algorithm based on region growing [J]. Computer Engineering, 2010, 36(13):192-193
- [11] 王璐,张帆,李伟,等.基于 Gabor 滤波器和局部纹理特征提取的 SAR 目标识别算法[J].雷达学报,2015,4(6):658-665  
WANG Lu, ZHANG Fan, LI Wei, et al. SAR target recognition algorithm based on Gabor filter and local texture feature extraction [J]. Journal of Radars Sinica, 2015, 4(6):658-665
- [12] 倪维平,严卫东,吴俊政,等.MSTAR 图像 2D Gabor 滤波增强与自适应阈值分割[J].光电工程,2013,40(3):87-93  
NI Weiping, YAN Weidong, WU Junzheng, et al. 2D Gabor filter enhancing and adaptive thresholding for MSTAR image[J]. Opto-Electronic Engineering, 2013, 40(3):87-93
- [13] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [C] // European Conference on Computer Vision, 2016:21-37
- [14] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:779-788
- [15] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:770-778
- [16] Sun N L, Xu S C, Cao M Y, et al. Segmenting and counting of wall-pasted cells based on Gabor filter [C] // IEEE Engineering in Medicine and Biology 27th Annual Conference, 2005:3324-3327

## SAR target detection network based on scenario synthesis and anchor constraint

JIN Xiaoyu<sup>1</sup> YIN Qiang<sup>1</sup> NI Jun<sup>1</sup> ZHOU Yongsheng<sup>1</sup> ZHANG Fan<sup>1</sup> HONG Wen<sup>2</sup>

<sup>1</sup> College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029

<sup>2</sup> Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094

**Abstract** Deep learning methods such as convolutional neural networks (CNN) have been widely used in fields of image processing and object recognition. However, the SAR images cannot yet be efficiently detected by CNN methods. Compared with traditional images, the SAR images have the advantage of all-day and all-weather acquisition, but they cannot obtain enough annotation due to the difficulty for interpretation and short of users. This paper proposes a SAR target detection method based on scene synthesis and anchor constraint. Firstly, the target and its shadow are segmented by region growing as well as threshold algorithm, and then the target detection data set is synthesized by randomly embedding the reasonable region into the SAR complex scene. Considering the SAR target's geometric characteristics and image resolution parameters, the anchor's size of Faster-RCNN is constrained to reduce the candidate frames that cannot meet the SAR target detection frame size, which massively reduce redundancy calculations so as to improve the efficiency and accuracy of training and testing process.

**Key words** SAR image; target detection; regional growth; dataset generation; anchor constraints