

# 基于级联 CNN 的 SAR 图像舰船目标检测算法

李健伟<sup>1†</sup>, 曲长文<sup>2</sup>, 彭书娟<sup>2</sup>

(1. 海军航空大学 研究生三队, 山东 烟台 264001; 2. 海军航空大学, 山东 烟台 264001)

**摘要:** 针对合成孔径雷达(SAR)图像中舰船目标稀疏的特点,提出一种基于级联卷积神经网络(CNN)的 SAR 图像舰船目标检测方法. 将候选区域提取方法 BING 与目标检测方法 Fast R-CNN 相结合,并采用级联 CNN 设计,可同时兼顾舰船检测的准确率和速度. 首先,针对 SAR 图像中相干斑噪声影响梯度检测的问题,在原有梯度算子的基础上增加平滑算子,并对图像尺寸个数和候选框个数进行适应性改进,使其提取到的候选窗口更快更准;然后,设计级联结构的 Fast R-CNN 检测框架,前端简单的 CNN 负责排除明显的非目标区域,后端复杂的 CNN 对高概率候选区域进行分类和位置回归,整个结构可以保证快速准确地对舰船这种稀疏目标进行检测;最后,设计一种联合优化方法对多任务的目标函数进行优化,使其更快更好地收敛. 在 SAR 图像舰船检测数据集 SSDD 上的实验结果显示,所提出的方法相比于原始 Fast R-CNN 和 Faster R-CNN 检测方法,检测精度从 65.2% 和 70.1% 提高到 73.5%,每张图像的处理时间从 2 235 ms 和 198 ms 下降到 113 ms.

**关键词:** 合成孔径雷达图像; 舰船; 检测; 级联; 卷积神经网络; Fast R-CNN

中图分类号: TN957.51

文献标志码: A

## A ship detection method based on cascade CNN in SAR images

LI Jian-wei<sup>1†</sup>, QU Chang-wen<sup>2</sup>, PENG Shu-juan<sup>1</sup>

(1. The 3rd Graduate Student Team, Naval Aeronautical University, Yantai 264001, China; 2. Naval Aeronautical University, Yantai 264001, China)

**Abstract:** According to the sparse characteristic of ships in synthetic aperture radar(SAR) images, a ship detection method based on the cascaded convolutional neural network(CNN) is proposed, which combines BING with Fast R-CNN in a cascaded way while taking the accuracy and speed into account. Firstly, a smoothing operator is added on the original gradient operator to tackle the speckle noise of SAR images. And the number of image size and candidate proposals are reduced according the distribution of ships in SAR images. After these improvements, the region proposal method gets more accurate without additional computations. Then, a cascade Fast R-CNN detection framework is designed to detect ships fastly and accurately. Its front simple CNN is responsible for rejecting the obvious background regions, and the back complex CNN is responsible for conducting classification and regression for the high probability candidate regions. The whole architecture makes the detection of the sparse ships in SAR images fastly and accurately. Finally, a joint optimization method is proposed to optimize the multi-objective function. The experiments on the dataset SSDD verify the superiority of the proposed method. The accuracy and speed boost from 65.2% / 70.1% and 2 235 ms / 198 ms of the Fast R-CNN and the Faster R-CNN to 73.5% and 113 ms, respectively.

**Keywords:** SAR images; ship; detection; cascade; CNN; Fast R-CNN

## 0 引 言

合成孔径雷达(Synthetic aperture radar, SAR)图像中的舰船目标检测分为大片海域和港口区域两种情况:大片海域的背景简单,较容易检测;港口区域的背景复杂,较难检测.目前,舰船目标检测算法都是专门针对某一种应用背景,无法同时兼顾两者,这是因为基于海杂波统计分布的建模方法丢失了船舶目标

的空间分布特征<sup>[1]</sup>.而基于机器学习的目标检测方法却不受这一限制,该方法通过设计有效的、能够区分舰船和非舰船的特征,可以很容易做到在 SAR 图像中判断舰船区域和非舰船区域(即海面或者岸边).目前,深度学习通过大量的数据自动学习到的特征是最有效的特征.

深度学习目标检测作为计算机视觉领域经典问

收稿日期: 2018-02-05; 修回日期: 2018-08-15.

基金项目: 国家自然科学基金项目(61571454).

责任编委: 林崇.

<sup>†</sup>通讯作者. E-mail: lgm\_jw@163.com.

题在近年来取得了非常大的进步,以 Faster R-CNN (Region based convolutional neural network)<sup>[2]</sup>、SSD (Single shot detector)<sup>[3]</sup>和 R-FCN (Region based fully convolutional neural network)<sup>[4]</sup>等为代表的深度学习目标检测算法,在通用物体检测(对多类目标检测)上取得了非常好的效果,受到了极大关注.为了提高整体检测的速度,上述方法基本都去掉了传统的候选区域提取模块,取而代之的是候选区域提取网络 (Region proposal network, RPN). 由于 RPN 可以端到端训练,也可与检测共享卷积网络,大大降低了候选区域提取的时间.

当考虑单一类别目标检测时(例如人脸和行人),传统的级联结构分类器是非常有效的方法,且在速度上具有先天的优势.级联结构分类器可以快速地去掉大量的背景区域,留下少量的潜在的目标区域进行精确分类,而级联结构本质上暗示着必须采用单独的候选区域提取模块,不能像上述方法那样利用 RPN 提取候选区域.

Viola & Jones<sup>[5]</sup>人脸检测器是级联检测的鼻祖,该方法仅利用简单的 Haar 特征,通过级联多个弱分类器提升分类效果,达到了实时检测人脸的效果,被应用于许多领域.后续的大量工作重点都集中于如何选取比较好的特征和如何设计比较好的分类器.直到以 CNN (Convolutional neural network)<sup>[6-7]</sup>为代表的深度学习技术横扫计算机视觉领域,人们自然地想到将级联与 CNN 相结合.例如, Li 等<sup>[8]</sup>设计的级联结构中有 6 个 CNN,其中 3 个 CNN 用于人脸/非人脸的二分类,另外 3 个 CNN 用于人脸区域的边框校正.但是该方法将级联与 CNN 结合得还不够彻底,虽然几个 CNN 模型进行了级联,但它们是分开优化的,没有像 Viola & Jones 那样进行联合优化. Li 等<sup>[9]</sup>利用逐层级联进行分割,也不能像 Viola & Jones 那样充分发挥级联效能.

候选区域提取是级联检测中的一个重要步骤,目前有 3 类方法:滑窗 (Sliding window)、分割 (Selective search, EdgeBoox 和 RIGOR 等)和似物性 (Objectness)<sup>[10]</sup>.传统的方法(如 Viola & Jones 等)就是利用基于滑窗的候选区域提取方法.滑窗策略是目标检测中的基本检测方式,需要遍历图像中的每个点以及以该点为起始点的不同大小的矩形窗口,然后依靠所采用的评分方式对检索窗口进行估分,从而判断当前检测位置是否为目标.显然,这样的遍历空间是巨大的,在实际应用中不易使用,因此需要采用简化策略.

Girshick<sup>[11]</sup>提出的 Fast R-CNN 所采用的候选区域提取方法是选择性搜索 (Selective search, SS)<sup>[12]</sup>,因为当时 BING (Binaried normed gradient)<sup>[13]</sup>还没被提出. Faster R-CNN 比 Fast R-CNN 更快的原因是用 RPN 代替了 SS,因为 SS 速度很慢,大概 2 s 一张图,这是整个检测方法的瓶颈.而 2014 年, Cheng 等<sup>[13]</sup>提出的 BING 相比于 SS,其速度和准确率都要好很多,但没有人将其与 Fast R-CNN 相结合,因为人们基本都采用 Fast R-CNN 的 RPN 网络而不采用其他候选区域提取方法.

尽管目前 Faster R-CNN 等方法的快速发展使级联得不到重视,但像 SAR 图像舰船目标这种比较稀疏的应用场景,级联结构会有更大的优势.本文将 BING 与 Fast R-CNN 相结合,针对 SAR 图像中舰船目标稀疏的特点,对 BING 进行针对性地改进,设计级联结构的 CNN,对原 Fast R-CNN 结构进行扩展并进行多任务联合优化,相比于 Faster R-CNN 能够大大提高检测速度和准确率.

## 1 级联模型设计与联合优化算法

### 1.1 数据集

基于深度学习的目标检测技术的关键是构建合适量的数据集及其标签.本文建立 SAR 图像数据集,称作 SSDD (SAR ship detection dataset) (下载链接: <https://pan.baidu.com/s/1dF6g3ZF>). 该数据集借鉴了 PASCAL VOC<sup>[14]</sup>数据集的构建流程,包括不同分辨率、极化、海况、大片海域和靠岸等条件的 SAR 图像. SSDD 一共有 1 160 个图像, 2 456 条船,平均每幅图像有 2.12 条船,图片长宽都在 500 像素左右.本文将这个数据集按照 7:2:1 的比例划分成训练集、测试集和验证集.图 1 给出了数据集中的一些示例.

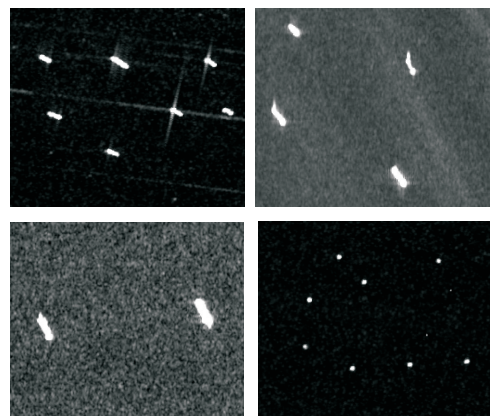


图 1 SSDD 中部分示例

SSDD 包含了不同条件下的 SAR 舰船目标,检测的目的是对每个 SAR 图像给出船的位置边框  $(x, y, w, h)$  以及置信度.其中:  $(x, y)$  代表左上角坐标,  $w$  代

表方框的宽度,  $h$  代表方框的高度. 标签制作过程是通过人工实现的, 而不是采用 AIS 信息.

## 1.2 基于BING的候选区域

似物性(Objectness)是一个反映图像窗口包含任一类别物体的衡量指标. 该方法的目的是产生一个小规模的候选窗口集, 也就是从图像中根据包含物体的概率大小抽样出一定数目的窗口集合. 相比于其他候选区域提取方法, 它通过降低搜索空间来提高计算效率, 在后端允许使用强分类器进一步提高检测的准确率.

Cheng等<sup>[13]</sup>提出的BING是一种简单而又高效的目标检测方法, 该方法基于这样一个假设: 物体具有严格完整的边界轮廓, 其梯度与背景有较大的差别. BING方法采用一种简单快速的64维梯度特征, 将其送入一个两阶段级联的SVM分类器进行学习, 利用学习到的参数有效地计算每个图像窗口的Objectness得分. BING方法能够进行快速检测的关键在于它引入了二值化近似计算, 这大大减少了特征提取和测试阶段的时间. 在PASCAL VOC数据集之上产生1000个候选框时准确率能够达到96.2%, 单台笔记本电脑处理每幅图像只需要0.03s, 训练时间也仅需20s, 而且具有很好的泛化能力, 也可适应SAR图像中的舰船目标.

在BING算法中, 为了找出图像中存在的物体, 先扫描各种量化尺度下的图像窗口(文献[13]中的窗口长宽可分别取10、20、40、80、160、320, 一共36种不同尺度), 每个窗口用一个线性模型  $w \in R^{64}$  进行打分, 即

$$s_l = \langle w, g_l \rangle, \quad (1)$$

$$l = (i, x, y). \quad (2)$$

其中:  $s_l$ 、 $w$ 、 $g_l$ 、 $l$ 、 $i$  和  $(x, y)$  分别表示图像窗口的滤波得分、模型的权重向量、图像窗口的梯度特征、位置、量化尺度和坐标. 然后, 在每个量化尺度上, 利用非极大值抑制原则从图像中筛选出一系列候选窗口. 最后, 定义图像窗口的Objectness得分  $o_l$  (即校正后的滤波器得分)为

$$o_l = v_i \cdot s_l + t_i, \quad (3)$$

其中  $v_i, t_i \in R$  表示在量化尺度  $i$  上支持向量机(SVM)分类器的学习系数和偏置项系数. 值得注意的是, 式(2)是用来矫正式(1)中的得分, 将最终的少数候选窗口进行重排. 为了加速图像的特征提取和测试过程, BING算法采取二值化近似方法, 充分利用计算机底层所有数据均由0和1表示这种机制, 通过

少数的操作快速实现大量复杂的卷积运算.

BING使用的是一维的模板  $[-1, 0, 1]$  来计算  $x$  和  $y$  方向的梯度. 由于SAR图像成像原理与普通光学成像原理不同, SAR图像中会有非常严重的相干斑噪声, 使图像的可解释性大大下降, 同时普通梯度算子也会检测到由相干斑噪声产生的无效的梯度, 对分类器造成严重的干扰. 针对这一问题, 本文在梯度模板  $[-1, 0, 1]$  后加上一个  $[1, 2, 1]$  的平滑算子, 对图像进行相干斑噪声梯度的抑制(等效于Sobel算子), 得到标准化梯度图(NG), 后续处理过程与文献[13]一致. 这样处理充分考虑了SAR图像中舰船的特点, 只是增加了一个梯度计算, 稍微增加了一些计算量.

本文将处理的图像的尺寸从36种改成16种(长宽可为40、80、160和320, 一共16种不同尺寸), 因为通过观察可以发现, 数据集SSDD中船的尺寸一般较小, 所以去掉了两个较小的图像尺寸.

由于不像PASCAL VOC那样有20类物体, SSDD仅有一类(舰船)目标, 平均每幅图像有2.12艘船, 没有必要产生1000个候选框, 实验结果发现, 500个候选窗口即可达到非常好的检测率. 因此, 这里每幅图仅产生500个候选窗口, 即选择前500的候选窗口并送入后续的分类网络.

## 1.3 级联结构设计

本文的级联结构设计是在Fast R-CNN基础上加以改进的. 首先, Fast R-CNN在整张图上进行卷积; 然后, 采用RoI(Region of interest)池化得到定长的特征向量, 例如不管窗口大小是多少, 都转换成  $7 \times 7$  的尺寸. Fast R-CNN还引入了一个重要的策略, 即在对窗口进行分类的同时, 还会对物体的边框进行回归, 使得检测框更加准确.

本文以VGG16<sup>[15]</sup>为例介绍级联结构的设计原理, VGG16的详细参数可参看其相关文献, 这里不再赘述. 图2和图3中阶段1~阶段3所用的CNN分别是VGG16的1~3、4~8和9~16层. VGG16是牛津大学计算机视觉组和Google DeepMind公司研发的深度卷积神经网络. VGG16探索了卷积神经网络的深度与其性能之间的关联, 通过反复堆叠  $3 \times 3$  的小型卷积核和  $2 \times 2$  的最大池化层, 成功地构建了16层的卷积神经网络. 相比之前的网络结构, 错误率大幅下降, 并取得了ILSVRC 2014比赛分类项目的第2名和定位项目的第1名. 同时VGG16拓展性很强, 泛化性能很好, 因此被用于很多地方.

图2为本文设计的级联CNN结构及其训练过程. 图2中卷积后面为后添加的模块, 包括在第1阶段添

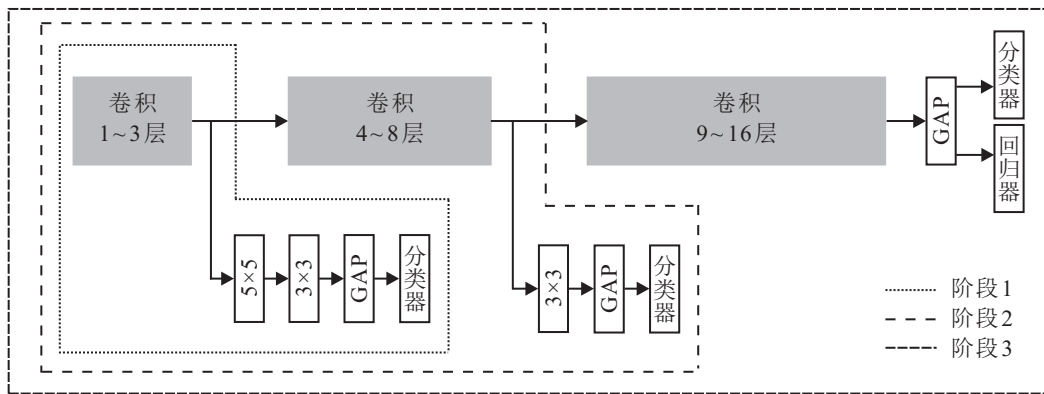


图2 级联CNN结构及其训练过程

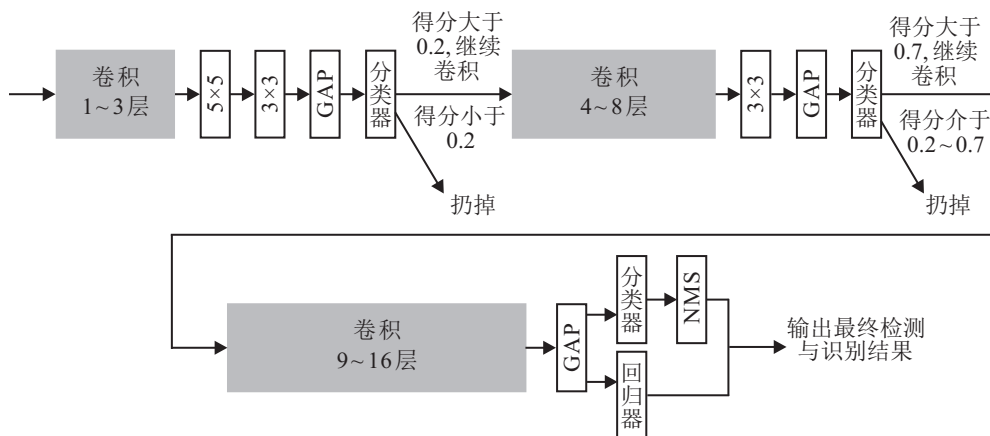


图3 级联CNN检测过程

加了一个 $5 \times 5$ 的卷积层、一个 $3 \times 3$ 的卷积层和一个全局平均池化层(Global average pooling, GAP)<sup>[16]</sup>。卷积层的目的是降低特征图的维度,避免较大特征图分类时产生较大的参数量和计算量。用GAP代替全连接层是因为全连接层几乎占据了VGG16中90%的参数量,去除全连接层后模型训练更快并且可以减轻过拟合。GAP对每一个特征图求出其平均数,然后将这些平均数组成一个特征向量,输入到Softmax层中进行分类。使用GAP能够强化特征图与类别的关系,且没有参数需要进行优化。

第2阶段类似于第1阶段,增加了一个 $3 \times 3$ 的卷积、一个GAP和一个分类器。

第3阶段增加了GAP、分类器和回归器,与Fast R-CNN的最后模块一致,同时进行分类与回归定位任务。

从图2中可以看到,训练过程分为3个阶段。

阶段1是一个3层的CNN,该阶段层数较少,只有3个隐藏层,所以其区分目标与背景能力较弱,但是其模型简单,计算速度快,可以用于快速排除SAR图像中是背景的滑动窗口区域,在该过程一般能排除90%以上的候选窗口。

阶段2是在阶段1的基础之上级联了一个5层的

CNN。该阶段通过增加的5层CNN提高了特征表达能力,所以能够更好地区分目标与背景,排除背景成分较高的候选窗口。

阶段3类似于阶段2,即又增加了8层的CNN,这样又进一步增加了网络的特征提取表达能力,从而能够更好地区分目标窗口和背景窗口。这一阶段留下来的窗口基本上都会含有大部分的目标成分和小部分的背景成分,只有通过更多层的网络才能更精确地进行鉴别。该阶段网络层数较深,计算量相对较大,但是因为前面两个阶段去掉了绝大多数的候选窗口,仅保留了极少数的窗口,所以速度也会很快。阶段3多了一个回归器,对海面目标的位置进行回归训练,用于预测目标的精确位置。

以上3个阶段利用同一个数据库(海面目标数据库)进行训练,迭代次数逐步增多,达到收敛,即可用于实际SAR图像中海面目标的检测和识别。

分成3个阶段进行优化是为了更好地理解训练过程,但是想得到更好的效果只能通过多阶段联合优化,这将在第1.4节进行探讨。

检测过程不像训练过程分为繁琐的3个阶段,这里只需要一个过程即可完成,如图3所示。图3为测试过程,这里力求实现高准确率和实时性。

首先将滑窗输入到训练好的 1~3 层进行分类, 由于该网络较浅, 对特征提取与表达的能力较差, 这里设定得分门限为 0.2 (门限是通过多次实验得到的最优的参数), 即分类器得分超过 0.2 的送到下一步骤继续卷积提取特征, 分类器得分小于 0.2 的直接丢弃. 送到下一步骤的候选窗口需要经过另外 5 层的 CNN 继续进行卷积, 提取特征, 进行分类. 该过程将分类器得分超过 0.7 的送到下一步, 小于 0.7 的直接扔掉. 最后, 对剩余的候选窗口继续进行 8 层的 CNN, 提取更好的特征, 进行更精确的分类, 并进行极大值抑制 (Non-maximum suppressing, NMS)<sup>[17-18]</sup>, 同时输出预测器得到的目标的边框坐标点.

### 1.4 联合优化算法

第 1.3 节中提到的级联 CNN 共有 4 个损失函数输出: 第 1 个代表第 1 阶段的分类损失函数; 第 2 个代表第 2 阶段的分类的损失函数; 后两个代表第 3 阶段的分类和回归的损失函数.

对于这 4 个损失函数, 有两种优化方法, 分别是分阶段优化和联合优化. 分阶段优化过程在第 1.3 节已进行论述, 但由于其计算繁琐, 且不能相互之间协同优化, 本文采用联合优化的方法. 联合优化 3 层 4 个输出, 每一个输出的损失函数可以作为其他损失函数的正则化项, 防止过拟合的同时提高准确率, 前端输出点可以提高额外的梯度信号, 防止梯度消失.

多任务损失函数如下所示:

$$L_{total} = \sum_{n=1}^N w_n L_n. \quad (4)$$

其中:  $N = 3$ ;  $L_1$  和  $L_2$  分别代表分类任务的损失函数, 二者的区别是使用的特征和判定门限不同;  $L_3$  代表分类和定位任务的损失函数. 即有

$$L_1 = -y_0 \log \hat{y}_0 - y_1 \log \hat{y}_1, \quad (5)$$

$$L_2 = -y_0 \log \hat{y}_0 - y_1 \log \hat{y}_1, \quad (6)$$

$y_0$  和  $y_1$  分别代表背景和船标签.

对于第 1 阶段, Softmax 交叉熵损失函数为

$$\hat{y}_1 = \text{softmax}(z) = \frac{\exp(z)}{\sum_{c \in \{0,1\}} \exp(z_c)}, \quad (7)$$

$z = f_1(x)$ ,  $f_1$  是第 1 阶段的输出. 第 2 阶段输出与第 1 阶段类似.

第 3 个输出用多任务损失函数联合优化分类和边框回归. 其中: 用  $u$  代表类别, 1 代表船, 0 代表背景,  $v$  是数据库里的真实标签边框. 令第 3 阶段损失函数为

$$L_3(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u = 1]L_{loc}(t^u, v). \quad (8)$$

其中:  $L_{cls}(p, u) = -\log p_u$  是分类损失;  $L_{loc}$  是边框回归损失;  $v = (v_x, v_y, v_w, v_h)$ ; 预测的边框为  $t^u = (t_x^u, t_y^u, t_w^u, t_h^u)$ ;  $[u = 1]$  代表  $u = 1$  时取 1, 其他情况取 0; 本文中令  $\lambda = 1$ . 定位误差采用平滑  $L_1$  损失函数

$$L_{loc}(t^u, v) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L_1}(t_i^u - v_i). \quad (9)$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1; \\ |x| - 0.5, & \text{otherwise.} \end{cases} \quad (10)$$

关于  $w_n$  参数的选取, 通过实验可以发现, 给前几个阶段更高的损失权重会提高整体的准确率, 这里令  $w_1 = 1.0, w_2 = 0.8, w_3 = 0.6$ .

### 1.5 应用细节

仿真环境: Ubuntu 14.04 操作系统 CPU Intel(R) Core(TM) i7-6770 K @ 4.00 GHz ×8 和 NVIDIA GTX1 080 GPU 8 G 显存 CUDA8.0 cuDNN5.0, Python 2.7, Pycharm 集成开发环境, 算法在 Fast R-CNN 的 caffe<sup>[19-23]</sup> 版本之上进行改进.

超参数: 学习策略为 step, gamma = 0.1, stepsize 为 30 000, Momentum 为 0.9, 训练 40 000 次迭代, 前 30 000 次学习率为 0.001, 后 10 000 次学习率为 0.000 1. 为了避免过拟合, 将 dropout 比率从 0.5 增加到 0.6, 训练时通过翻转来增加正样本的 RoI. 预训练与初始化: 将 VGG16 在 Imagenet<sup>[24]</sup> 的预训练之后的权重迁移过来, 新添加的层利用 Xavier 函数进行初始化.

## 2 实验结果与分析

### 2.1 BING 算法改进效果

用文献[13]的评价指标检测率 (Detection rate) 和候选窗口个数 (Number of windows) 曲线来对比不同算法的性能, 两种方法候选框个数与检测率的曲线如图 4 所示.

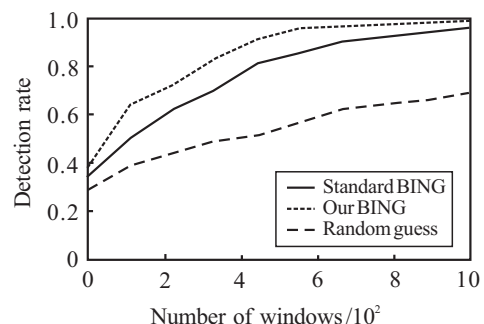


图 4 候选区域提取方法对比

本文所提出的方法在产生500个候选窗口时,检测概率能够达到96.5%;1000个候选窗口时,检测概率可以达到99.8%。相比于标准的BING算法,达到理想的检测概率(一般大于95%)所需要的候选窗口个数更少,在某一数量级候选窗口下得到的准确率更高。这是由于本文所用的SSDD数据集中目标种类少且背景较简单的原因,在实际应用时可将候选窗口选为500个。

算法耗时分析:标准的BING算法在PASCAL VOC数据集中平均用0.003 s处理完一幅图;改进的BING算法增加了一个平滑算子,但由于处理图像的尺度从36个降到16个,且SSDD数据集都是灰度图像,计算量有增有减,通过实验大概为0.0021 s一幅图像。相比于Fast R-CNN算法所用的选择性搜索(Selective search)方法,本文方法运行速度得到了极大的提升(文献[13]中,用的笔记本CPU是Intel i7-3940XM,主频3 GHz;本文则是Intel Core i7-6770K,频率为4 GHz)。

## 2.2 级联CNN处理效果

为了验证所提出的级联CNN算法的性能,这里将其与SS+Fast R-CNN、SS+级联CNN、BING+Fast

R-CNN、BING+级联CNN、Faster R-CNN进行对比(上述方法依次编号为方法1~方法6),主要是平均准确率和测试时每幅图像平均处理时间。

通过表1可以发现,选择性搜索(SS)耗时最严重,用BING替换之速度会得到极大地提升。级联CNN相比于原始Fast R-CNN其处理速度得到了提升,且其精度提升的会更大一些。将改进的BING与级联CNN(本文方法)相结合在准确率和速度上都得到了极大地提升,从而表明了本文方法的有效性。

表1 不同检测模型检测效果对比(基于VGG-16模型,利用SSDD数据集)

| 方法      | 1    | 2    | 3    | 4    | 5    | 6    |
|---------|------|------|------|------|------|------|
| 平均准确率/% | 65.2 | 66.1 | 65.4 | 68.3 | 70.1 | 73.5 |
| 平均时间/ms | 2235 | 2103 | 203  | 183  | 198  | 113  |

图5是SSDD数据集中几个典型图像处理结果。可以看出,本文方法不仅能够快速准确地检测到各种条件下的舰船目标,而且还具有很强的适应性,即对于那些靠近岛屿和近岸的目标也能很好地检测出来,不需要传统方法进行海陆分割和杂波建模等操作。

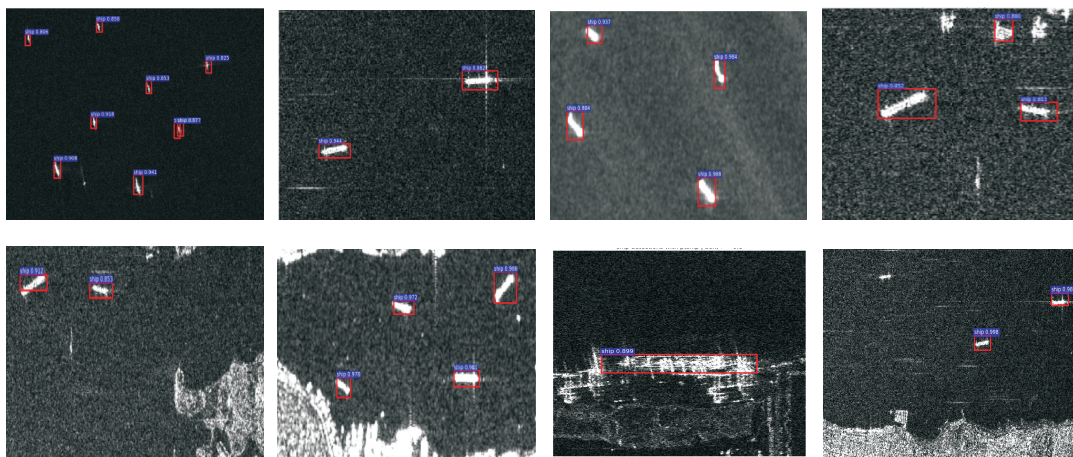


图5 部分样本处理结果

## 3 结论

本文提出了一种基于级联CNN的SAR图像舰船目标检测方法。针对SAR图像舰船目标的电磁散射特性,在原BING算法基础上增加了平滑算子,减弱了相干斑噪声对提取梯度的影响,并将多尺寸图像从36个降到16个,候选窗口从1000个降到500个,从而提高了提取候选窗口的速度和检测概率。针对SAR图像中舰船目标稀疏的特点,设计了级联CNN结构,前端简单的CNN负责快速排除明显非目标候选区域,后端复杂的CNN负责对存在目标的候选区域进

行精确识别和定位,并对网络结构的多个输出损失函数进行联合优化。通过在数据集SSDD上的实验表明,所提出的级联CNN舰船检测方法不仅具有很高的准确率,而且也有很快的处理速度。同时,此类方法不需要对大片海域和港口区域进行区分处理,具有很强的适应性。

## 参考文献(References)

- [1] Friedman W C C, Pichel K S, Clemente-Colon W G, et al. Automatic detection of ships in RADARSAT-1 SAR imagery[J]. Canadian J of Remote Sensing, 2001, 27(5):

- 568-577.
- [2] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Trans on Pattern Analysis & Machine Intelligence*, 2017, 39(6): 1137-1149.
- [3] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]. *Proc of the European Conf on Computer Vision*. Amsterdam, 2016: 21-37.
- [4] Dai J, Li Y, He K, et al. R-FCN: Object detection via region-based fully convolutional networks[C]. *Proc of the European Conf on Computer Vision*. Amsterdam, 2016: 1-12.
- [5] Viola P, Jones M. Robust real-time object detection[J]. *Int J of Computer Vision*, 2004, 57(2): 137-154.
- [6] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. *Int Conf on Neural Information Processing Systems*. Lake Tahoe, 2012: 580-587.
- [7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. *2014 IEEE Conf on Computer Vision and Pattern Recognition*. Columbus, 2014: 580-587.
- [8] Li H, Lin Z, Shen X, et al. A convolutional neural network cascade for face detection[C]. *2015 IEEE Conf on Computer Vision and Pattern Recognition*. Boston, 2015: 5325-5334.
- [9] Li X, Liu Z, Luo P, et al. Not all pixels are equal: Difficulty-aware semantic segmentation via deep layer cascade[C]. *2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 6459-6468.
- [10] Hosang J, Benenson R, Dollar P, et al. What makes for effective detection proposals[J]. *IEEE Trans on Pattern Analysis & Machine Intelligence*, 2016, 38(4): 814-821.
- [11] Girshick R. Fast R-CNN[C]. *2015 IEEE Conf on Computer Vision and Pattern Recognition*. Boston, 2015: 1440-1448.
- [12] Uijlings J R, Sande K E, Gevers T, et al. Selective search for object recognition[J]. *Int J of Computer Vision*, 2013, 104(2): 154-171.
- [13] Cheng M M, Zhang Z, Lin W Y, et al. BING: Binarized normed gradients for objectness estimation at 300 fps[C]. *2014 IEEE Conf on Computer Vision and Pattern Recognition*. Columbus, 2014: 3286-3293.
- [14] Everingham M, Eslami S M A, Van Gool L, et al. The pascal visual object classes challenge: A retrospective[J]. *Int J of Computer Vision*, 2015, 111(1): 98-136.
- [15] Zisserman A S K. Very deep convolutional networks for large-scale image recognition[C]. *IEEE Int Conf on Computer Vision*. Santiago, 2015: 311-317.
- [16] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. *2015 IEEE Conf on Computer Vision and Pattern Recognition*. Boston, 2015: 1-9.
- [17] Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining[C]. *Proc of the European Conf on Computer Vision*. Amsterdam, 2016: 761-769.
- [18] Wang X, Shrivastava A, Gupta A. A-Fast-RCNN: Hard positive generation via adversary for object detection[C]. *2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 3039-3048.
- [19] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[J]. *J of Machine Learning Research*, 2010, 3(5): 249-256.
- [20] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Trans on Pattern Analysis & Machine Intelligence*, 2014, 37(9): 1904-1916.
- [21] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. *2016 IEEE Conf on Computer Vision and Pattern Recognition*. Las Vegas, 2016: 770-778.
- [22] Huang G, Liu Z, Maaten L V, et al. Densely connected convolutional networks[C]. *2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 2261-2269.
- [23] Jia Yangqing, Shelhamer, Jeff Donahue, et al. Caffe: Convolutional architecture for fast feature embedding[C]. *Proc of the 2014 ACM Conf on Multimedia*. New York, 2014: 170-178.
- [24] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. *Int J of Computer Vision*, 2015, 115(3): 211-252.

### 作者简介

李健伟(1989—),男,博士生,从事SAR图像目标检测与识别、计算机视觉应用的研究, E-mail: ljm\_jw@163.com;  
曲长文(1963—),男,教授,博士生导师,从事电子对抗、SAR图像处理等研究, E-mail: qcw@sohu.com;  
彭书娟(1983—),女,工程师,从事电子对抗、SAR图像处理的研究, E-mail: peng\_shu\_juan@163.com.

(责任编辑: 李君玲)