

引用格式: QU Haicheng, SHEN Lei. SAR Image Denoising and Semantic Enhancement for Object Detection [J]. Acta Photonica Sinica, 2022, 51(4):0410003

曲海成, 申磊. 面向目标检测的 SAR 图像去噪和语义增强 [J]. 光子学报, 2022, 51(4):0410003

面向目标检测的 SAR 图像去噪和语义增强

曲海成, 申磊

(辽宁工程技术大学 软件学院, 辽宁 葫芦岛 125105)

摘要:为了解决合成孔径雷达图像中舰船目标容易在复杂背景下被淹没、相干斑噪声导致舰船边缘模糊和小尺度舰船目标在经过多次卷积后容易丢失的问题,提出像素级去噪和语义增强的检测模型。首先,利用预测掩码逐像素指导特征图,激活目标信息,抑制背景和相干斑噪声,检测相干斑噪声和相似舰船目标影响的图像。其次,利用语义增强模块增强特征图中语义信息,使得每层特征图都包含丰富的语义信息,进而判断小尺度舰船。最后,引入 Transformer Encoder 模块,提高舰船目标和特征图之间的上下文信息,增强舰船目标和图像之间的依赖关系。提出的模型能够有效减少漏检、误检情况,在公开数据集 SSDD 上进行测试,检测精度达到 96.73%,其中针对小尺度舰船检测精度达到 96.85%,大尺度舰船检测精度达到 96.41%,远海场景检测精度达到 98.53%,近海岸场景检测精度达到 90.00%,验证了该模型的有效性和泛化能力。

关键词:目标检测;图像处理;深度学习;SAR 图像;像素级去噪;语义增强;Transformer

中图分类号: TP751

文献标识码: A

doi: 10.3788/gzxb20225104.0410003

0 引言

合成孔径雷达(Synthetic Aperture Radar, SAR)是一种主动式微波传感器,可在能见度极差的环境下获取高分辨率的遥感图像,广泛应用在渔船监测和军事侦察中^[1]。合成孔径雷达的独特成像原理导致海洋杂波散射形成相干斑噪声^[2],使得舰船边缘在 SAR 图像中更加模糊。SAR 图像中的目标由大量亮点组成,岛礁、码头等形似舰船的物体使得 SAR 图像背景更加复杂,舰船目标容易淹没在复杂背景中。

传统 SAR 图像检测算法多为恒虚警率算法(Constant False-Alarm Rate, CFAR)及其衍生算法^[3]。为了解决泄露在局部窗口中的异质点影响杂波筛选的问题,AIJ 等提出自适应截断杂波统计的双参数恒虚警率算法(Truncated Statistics Log Normal Constant False-Alarm Rate, TS-LNCFAR)^[4],通过自适应阈值去除局部滑动窗口中的高强度异质点,保留更加真实的海洋杂波。为了解决多目标场景下算法性能较低的问题,LI T 等提出超像素级恒虚警检测算法(SuperPixel-Level Constant False-Alarm Rate, SuperPixel-Level CFAR)^[5]。以像素点为基本单元,对每一个超像素估计背景杂波分布的参数,并计算每一个像素点的检测统计量,大于阈值的定义为目标像素点,小于阈值的定义为海洋杂波像素点,从而更好地区分目标和海洋杂波,提高多目标场景检测效果。

深度学习在目标检测领域的应用推动了 SAR 图像目标检测的发展,LI J 等首次将 Faster R-CNN 应用于 SAR 图像^[6],利用特征融合和迁移学习^[7]对检测模型进行优化。T Y 等提出单阶段检测模型 Retinanet^[8],通过精简模型在损失一定精度的情况下提升了检测速度。近年来,CornerNet^[9]、CenterNet^[10]等基于无锚框检测模型减少了模型超参数,降低对硬件的要求。LANHAM M 等提出 Transformer^[11],在自然语言处理领域利用多头注意力机制达到了更好的效果。随后 Transformer 应用于目标检测,达到了比 Faster R-CNN 更

基金项目:国家自然科学基金(No.42071351),辽宁省教育厅基础科研项目(No.LJ2019JL010),辽宁工程技术大学学科创新团队资助项目(No.LNTU20TD-23)

第一作者:曲海成(1981—),男,副教授,博士,主要研究方向为遥感影像高性能计算、目标检测与识别。Email: quhaicheng@lntu.edu.cn

通讯作者:申磊(1995—),男,硕士研究生,主要研究方向为遥感图像目标检测、深度学习。Email: shenlei95821@163.com

收稿日期:2021-10-27; **录用日期:**2021-12-30

<http://www.photon.ac.cn>

好的检测效果。

传统 SAR 图像检测算法只能针对特定的数据,算法泛化能力较差。同时传统 SAR 图像需要人工提取特征,过程复杂且较难提取到充分的特征。基于深度学习的检测模型分为基于锚框的检测模型和无锚框检测模型。基于锚框的检测模型需要引入锚框数量、锚框比例等超参数,消耗大量的计算资源。在舰船尺度多样的 SAR 图像中,人工设计的锚框不能完美适合各种尺度的舰船目标,导致模型检测效果较差。无锚框的检测模型主要针对光学图像, SAR 图像中背景复杂且存在相干斑噪声,容易淹没舰船位置信息,降低模型检测精度。

为了解决上述问题,本文提出了像素级消噪和语义增强的检测模型。提出利用预测的真值掩码特征获得 $[0,1]$ 的注意力图,然后利用掩码注意力图逐像素指导特征图,抑制复杂背景中的物体和相干斑噪声,最后采用交叉熵损失优化预测的真值掩码特征。提出语义增强模块,利用非对称卷积层提取不同维度特征,同时将低层特征重新融合^[12]到下层网络,避免卷积操作导致信息丢失。语义增强模块能解决高 Intersection Over Union (IOU) 分数低分类置信度的候选框被抑制的问题。在语义增强模块里使用 Transformer Encoder 模块,用于提取特征图中的上下文信息^[13],保证特征图始终保持全局感受野。

1 像素级去噪和语义增强的检测模型

SAR 图像与光学图像不同, SAR 图像中的目标容易受复杂背景信息和噪声抑制,导致检测模型忽略目标信息,误将背景信息中的物体识别为舰船。为了消除背景信息和相干斑噪声的影响,提出像素级去噪和语义增强的检测模型。如图 1 所示,模型可以分为四部分:特征提取模块、特征金字塔模块、像素级去噪模块和检测模块。1) 特征提取模块:一般来说,较深的主干网络提取的特征图具有更丰富的语义信息,然而,简单的增加网络深度很容易导致小目标位置丢失。Resnet50 网络层数较浅,提取的特征图缺少语义信息,高质量的候选框因为较低的分类置信度被抑制,导致模型检测效果较差。因此选择 Resnet101 作为主干网络。2) 特征金字塔模块:首先将主干网络提取的特征图输入 Transformer Encoder 模块获得具有全局感受野和上下文信息的特征图。然后,在自上而下融合过程中,将对应的底层特征图重构到子网络,有助于解决梯度消失问题,最终获得 $\{P_3, P_4, P_5, P_6, P_7\}$ 层特征图。3) 像素级去噪模块:去噪模块生成与特征图对应的注意力图,然后将特征图与注意力图逐像素点相乘,增加目标区域像素的权重,降低非目标区域像素点的权重,使

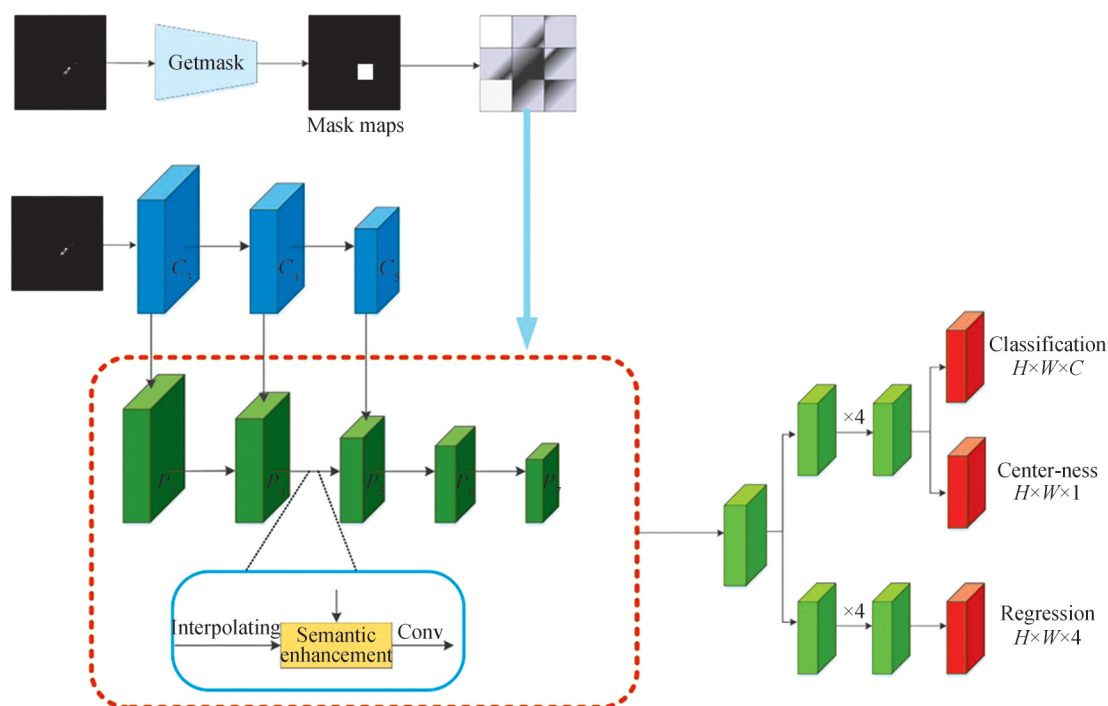


图1 网络结构

Fig.1 Network structure diagram

得目标区域和非目标区域具有更好的对比度。4)检测模块:检测模块用同一个检测头检测不同尺度的特征图。为了解决不同尺度特征图的回归框差异性,该模块引入可训练的尺度参数自动调整回归框。在训练过程中,检测模型总会出现距离中心点较远的低质量候选框。Center-ness表示该位置到中心点的距离,并采用交叉熵损失函数进行优化,获得高质量的候选框。

首先,利用Resnet101主干网络提取特征,Resnet101中conv2、conv3、conv4、conv5包含3、4、23、3个残差块,与conv1中的 7×7 卷积层和 3×3 池化层,共有110个卷积层。每一层的最后一个残差块将特征图缩减为原来的 $1/2$,通道数变为原来的2倍,最终获得 $\{C_2, C_3, C_4, C_5\}$ 尺度的特征图,为了保证模型的训练速度和显存的消耗, C_2 特征图没有输入检测头阶段。然后,在两层特征图之间采用对称卷积和非对称卷积提取不同维度特征,并对提取的特征做基于Transformer的编码^[14]处理,在增强语义信息的同时引入上下文信息。采用自上而下融合特征图,使得底层高分辨特征图包含丰富的语义信息。最后预测掩码特征图并采用Softmax函数构建注意力图,对每一层特征逐像素相乘,激活舰船目标信息,抑制背景信息。

算法总体流程为:

- 1)初始化主干网络的预训练参数,将SAR图像输入Resnet101提取特征,获得 $\{C_2, C_3, C_4, C_5\}$ 特征图。
- 2)采用金字塔思想,将主干网络提取的特征图输入语义增强模块,利用Transformer Encoder获得全局感受野和上下文信息,最终获得 $\{P_3, P_4, P_5, P_6, P_7\}$ 特征图。
- 3)去噪模块生成对应特征图的注意力图,逐像素更新特征图中像素点的权重,增强复杂背景下目标区的位置信息。同时,利用掩码损失函数优化注意力图。
- 4)检测头网络对输入的多尺度特征图进行目标分类和候选框回归。
- 5)利用中心度损失函数训练网络以获取更优的模型参数。中心度分数和分类置信度的乘积作为候选框分数,然后执行非极大值抑制算法筛选候选框,获得最终的检测结果。

1.1 无锚框

SAR图像中舰船目标稀疏,同时在不同分辨率下散射出的目标尺度不同,人工无法设计能适配所有尺度目标的锚框。为了提高对不同尺度目标的适配度设计大量锚框,需要消耗大量的计算资源。本文利用无锚框的检测模型,以中心点到每个边框的偏移量^[15]确定候选框的大小,去除了锚框数量、锚框比等冗余参数。中心点 (x, y) 到左边框、右边框、上边框、下边框的偏移量 l, r, t, b 表达式为

$$\begin{cases} l = x - x_{\min} \\ r = x_{\max} - x \\ t = y - y_{\min} \\ b = y_{\max} - y \end{cases} \quad (1)$$

式中, (x_{\min}, y_{\min}) 为真值框的左上角坐标, (x_{\max}, y_{\max}) 为真值框的右下角坐标。

特征图中 (x, y) 映射到原始图像中的 $\left[\left[\frac{s}{2}\right] + sx, \left[\frac{s}{2}\right] + sy\right]$ 感受区域。所有映射到真值框中的像素点均作为模型正样本,在训练过程中通过对偏移量回归实现对候选框的调整,感受区域映射如图2所示。

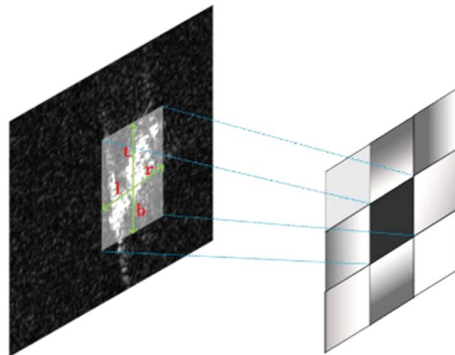


图2 感受区域映射

Fig.2 Receptive area mapping

1.2 像素级去噪

SAR图像中大部分背景都包含相干斑噪声,以注意力机制为主的去噪方法通过聚焦目标实现去噪。但是注意力机制在复杂背景中强化目标感知区域,并没有真正实现对噪声的抑制。因此提出像素级去噪模块,通过激活目标区域的位置信息活跃度,抑制非目标区域的噪声信息实现去噪。首先,特征图经过多个卷积层获得 $f_{\text{mask}} \in \mathbb{R}^{H \times W \times 2}$ 的掩码图;其次,经过Softmax函数将掩码图转化为 $[0, 1]$ 的注意力图;最后,对特征图的每一层做逐像素点乘,生成最终的特征图。针对预测掩码特征图,利用真值掩码做损失来优化注意力图,损失函数表达式为

$$L_{\text{mask}} = -\frac{1}{N_{\text{all}}} \sum_{i,j} f_{i,j}^* \log \left(\frac{\exp(f_{i,j})}{\sum_{i,j} \exp(f_{i,j})} \right) \quad (2)$$

式中, $f_{i,j}^*$ 为真值掩码, $f_{i,j}$ 为预测掩码, L_{mask} 为掩码损失。

由图3可见,注意力图中舰船目标区域权重较大,非目标区域权重较小,经过逐像素点相乘之后特征图中目标区域和非目标区域区分度更加明显。经过像素级消噪的特征图具有更广的目标感受区域,复杂背景中的舰船目标位置信息更加突出,形似舰船目标的物体也得到有效抑制。图4为经过Tensorboard的可视化特征图,图4(a)为未消噪的特征图,图4(b)为经过像素级消噪处理后的特征图。可以看出舰船目标在相干斑噪声和复杂背景下的特征图都达到了理想的去噪效果。

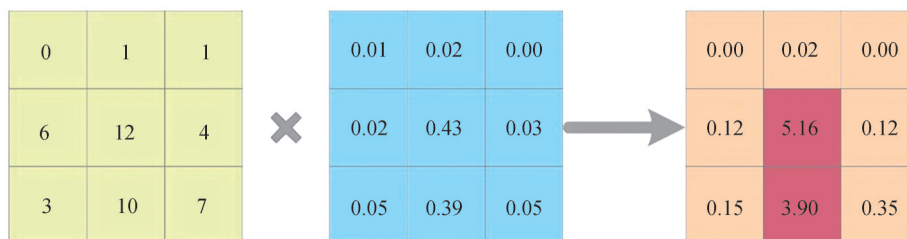


图3 像素级消噪示意图

Fig.3 Pixel-level denoising diagram

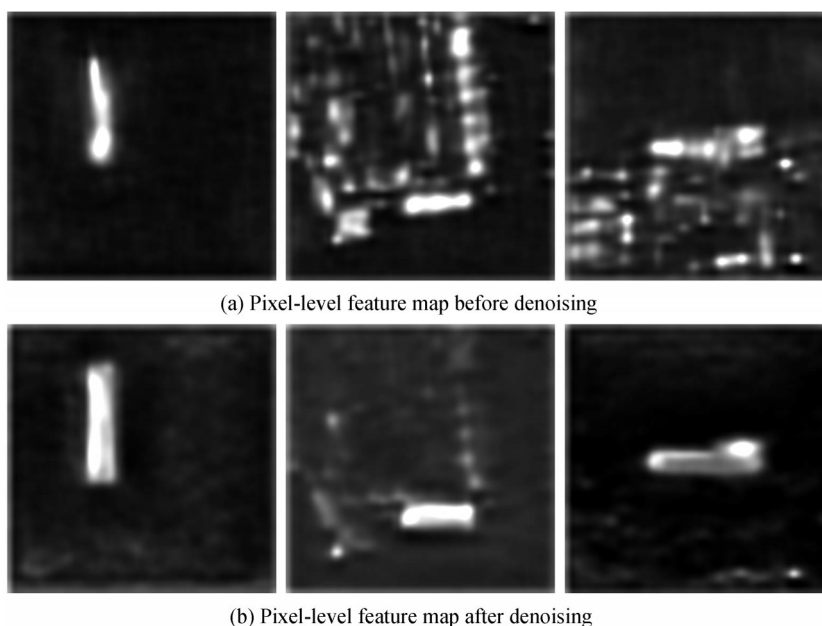


图4 像素级消噪效果

Fig.4 Pixel-level denoising effect

1.3 语义增强

语义信息对判别是否为舰船具有重要指导意义,丰富的语义信息能提高目标的分类置信度,防止高IOU分数低分类置信度的候选框被抑制。能够有效解决待检测目标漏检问题,同时防止形似舰船目标的物体被误检为目标。因此提出语义增强模块,首先,输入的特征图经过非对称卷积层,以不同维度提取多种特征;其次,利用Transformer Encoder增强舰船和特征图之间的上下文信息,提高模型对舰船目标的感知能力;最后,自上而下将语义信息融合到底层特征图,使得底层高分辨率特征图包含高层特征的语义信息。语义增强结构如图5所示。特征图经过多层卷积容易出现信息丢失的问题,在该模块中采用跳跃连接,将低层特征和主干网络提取的 $\{C_3, C_4, C_5\}$ 层特征图分别构造到对应子网络。

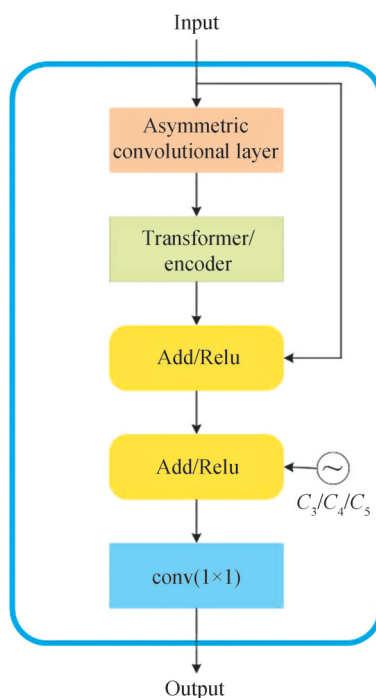


图5 语义增强结构

Fig.5 Semantic enhanced structure

1.3.1 非对称卷积层

非对称卷积层由 $\text{conv}(1 \times 3)$ 、 $\text{conv}(3 \times 1)$ 、 $\text{conv}(3 \times 3)$ 、 $\text{conv}(1 \times 1)$ 四个卷积组成,利用非对称卷积层增强网络的特征提取能力, 1×1 卷积实现在通道维度进行线性信息整合。同时,非对称卷积的引入降低了运算量,表现为将 $N \times N$ 的标准卷积核分为 $1 \times N$ 和 $N \times 1$ 的两层卷积,这样运算量由原来 N^2 减小到 $2N$,保证加深网络深度和模型检测精度。如图6所示,经过不同卷积核生成的特征图以通道维度连接,之后经过 1×1 卷积调整通道数,作为下一阶段的输入。

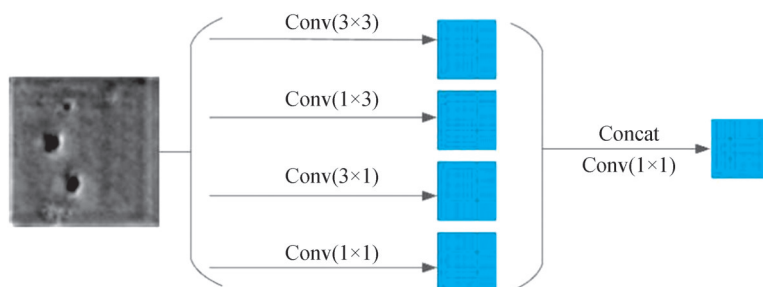


图6 非对称卷积层结构

Fig.6 Asymmetric convolutional layer structure

1.3.2 Transformer Encoder

上下文信息就像人类的感知,其作用是捕捉目标和特征图之间的相互信息,获得目标与邻域的差别,从而获得理想的检测效果。因此引入 Transformer Encoder 模块,将 $f_{\text{mask}} \in \mathbb{R}^{H \times W \times C}$ 的特征图拉伸为带有位置编码的 $H \times W \times C$ 大小的一维序列。以图片序列作为输入,经过多头注意力机制、正则化模块和前向传播模块,生成和舰船目标高度依赖的特征图。从图 7 可见,编码模块分为三步,首先,利用多头注意力机制捕获同一个图片序列中目标之间的语义特征,强化舰船目标的位置信息;其次,采用跳跃连接实现重用浅层特征来约束深层网络,利用正则化模块防止模型过拟合;最后,针对上层输入做前向传播,输出的特征图重新构造为原来尺度。

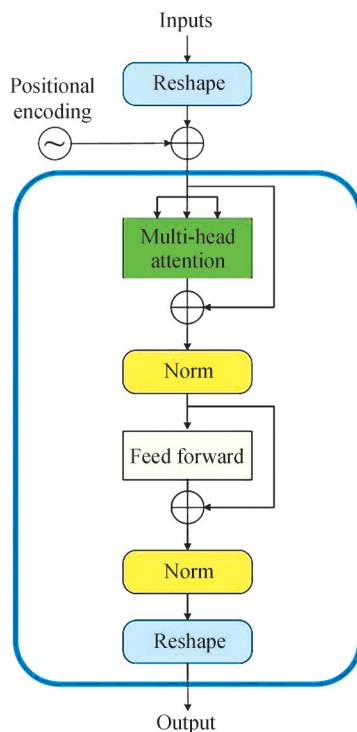


图 7 Transformer Encoder 结构

Fig.7 Structure of Transformer Encoder

该模块的位置编码通过三角线性变换获得,位置编码的表达式为

$$P_{\text{Positional Encoding}} = \begin{cases} \sin(p_{\text{sequence}}/10000^{2i/d_{\text{model}}}) \\ \cos(p_{\text{sequence}}/10000^{2i+1/d_{\text{model}}}) \end{cases} \quad (3)$$

式中, $P_{\text{Positional Encoding}}$ 为位置编码矩阵, i 为特征图序列位置, d_{model} 为特征序列长度, p_{sequence} 为特征序列位置。多头注意力在获得每个目标之间相关性的同时,也使每个目标包含其它目标的向量信息,在网络每一次迭代过程中不断学习舰船目标之间的位置信息,获得目标位置信息更加精准。多头注意力表达式为

$$M_{\text{attention}} = \text{softmax}\left(\frac{QK}{\sqrt{d_k}}\right)V \quad (4)$$

$$\begin{cases} Q = X \cdot W_Q \\ W = X \cdot W_W \\ V = X \cdot W_V \end{cases} \quad (5)$$

式中, $M_{\text{attention}}$ 为注意力图, d_k 为调节因子, W_Q 、 W_W 、 W_V 为权重参数,在模型迭代训练中不断学习。

1.4 损失函数

SAR 图像中舰船目标稀疏,在特征图中占比较小,目标区域和背景区域数量差别较大。一张特征图中

大部分候选区域为背景,容易出现正样本和负样本分配不均衡的问题。为了解决这个问题,采用 focal loss^[16] 优化分类损失,定义为

$$L_{\text{cls}} = \begin{cases} -\alpha(1-p_{i,j})^\gamma \log(p_{i,j}) & p_{i,j} = 1 \\ -(1-\alpha)p_{i,j}^\gamma \log(1-p_{i,j}) & p_{i,j} \neq 1 \end{cases} \quad (6)$$

式中, L_{cls} 为分类损失, $p_{(i,j)}$ 是在 i, j 处预测的类别, γ 为减少易分类样本的调节系数, α 为平衡正样本和负样本的平衡因子。根据 Faster R-CNN, 设置 $\gamma = 2, \alpha = 2.5$ 。预测生成的候选框与真值框的残差越大,模型检测效果越差。为了获得更加理想的候选框,选择 IOU_Loss 调整候选框,定义为

$$L_{\text{reg}} = -\log \left(\max \left(\frac{\text{Inter}(b_{i,j}, b_{i,j}^*)}{\text{Union}(b_{i,j}, b_{i,j}^*)} \right) \right) \quad (7)$$

式中, L_{reg} 为回归损失, $b_{i,j}$ 为预测候选框, $b_{i,j}^*$ 为真值框,Inter() 为交集运算,Union() 为并集运算。因此,总损失为

$$L_{\text{total}} = \frac{1}{N_{\text{all}}} \sum_{i,j} L_{\text{cls}}(p_{i,j}) + \frac{1}{N_{\text{pos}}} \sum_{i,j \in \text{pos}} L_{\text{reg}}(b_{i,j}, b_{i,j}^*) + \sum_{\text{level}} L_{\text{mask}}(m_{i,j}, m_{i,j}^*) \quad (8)$$

2 实验结果及分析

实验在 Ubuntu 16.04 操作系统下,基于深度学习框架 Tensorflow 进行,并通过 CUDA8.0 和 cuDNN5.0 加速模型训练速度。计算机的 CPU 为 Intel(R) Core(TM) i7-7700@3.6GHz×8, GPU 为 NVIDIA GTX1080Ti 显存为 11G。

2.1 数据集及预处理

将国内首个公开的数据集 SSDD 作为实验的训练集和验证集。SSDD 数据集的数据主要由 RadarSat-2、TerraSAR-X 和 Sentinel-1 传感器获取,共有 1 160 张图片,包含 2 456 艘舰船,平均每张图片 2.12 个舰船目标。传感器分辨率为 1~15 m,舰船尺度多样,包含近海岸、远海和相干斑噪声多个场景。为了验证模型的泛化能力,利用 SAR-Ship-Dataset 数据集进行对比。SAR-Ship-Dataset 数据集共有 43 819 张长和宽均为 256 像素的图片,包含 59 535 个舰船目标,具有更多的小尺度舰船。传感器分辨率分别为 3 m、5 m、8 m、10 m 和 25 m,成像模式包括 Strip-Map (UFS)、Fine Strip-Map 1 (FSI)、Full Polarization 1 (QPSI)、Full Polarization 2 (QPSII)、Fine Strip-Map 2 (FSII)、条带模式和宽幅模式。除此之外,HRSID 也被广泛应用于舰船检测领域,HRSID 共有 5 604 张图片,包含 16 951 个舰船目标,成像分辨率更高。为保证实验的公平性,实验数据均以 7:3 的比例划分为训练集和验证集,常用数据集介绍见表 1。

表 1 数据集参数
Table 1 Data set parameters

| Datasets | Sensors | Resolution/m | Polarization | Scenes |
|----------------------------------|-------------|-----------------|----------------|--|
| SSDD | RadarSat-2 | | | |
| | TerraSAR-X | 1~5 | HH, HV, VV, VH | |
| | Sentinel-1 | | | |
| SAR-Ship-Dataset ^[17] | Sentinel-1 | 3, 5, 8, 10, 25 | HH, HV, VV, VH | Offshore, Inshore, Coherent speckle noise |
| | GF-3 | | | |
| HRSID ^[18] | Sentinel-1B | | | |
| | TerraSAR-X | 0.5, 1, 3 | HH, HV, VV, VH | |
| | TanDEM | | | |

SSDD 数据集数据量较少,为了保证模型学习效果,提高模型泛化能力,对 SSDD 数据集进行数据增强操作。经过翻转、改变亮度、添加高斯噪声,最终训练集达到 3 248 幅图像。数据增强^[19] 为检测模型提供了充足的训练样本,提高了模型学习能力。SAR-Ship-Dataset 包含的图像数量远远大于 SSDD 数据集,可以在无需数据增强的条件下训练模型,因此实验中没有对 SAR-Ship-Dataset 数据集进行数据增强处理。

2.2 模型训练与评估

根据FCOS^[15]模型参数的设计,本文模型训练batch_size设置为1,采用MomentumOptimizer优化器,动量因子设置为0.9,学习率设为0.0005,每4万次迭代学习率衰减为原来的1/10,权重衰减系数为0.00004,在10万次迭代时模型达到完全收敛。

为了更加客观描述模型的性能,使用召回率(Recall)、精确率(Precision)和平均精度(Average Precision, AP)作为评价指标。针对不同方法对比,使用精确率-召回率(Precision-Recall, P-R)曲线直观表达模型效果,表达式为

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (9)$$

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (10)$$

$$P_A = \int_0^1 P(R) dR \quad (11)$$

式中, R 为召回率, P 为精确率, P_A 为平均精度, N_{TP} 为正确检测到舰船的数量, N_{FN} 为未正确检测到舰船的数量, N_{FP} 为误检为舰船目标的数量。

2.3 不同主干网络实验结果及分析

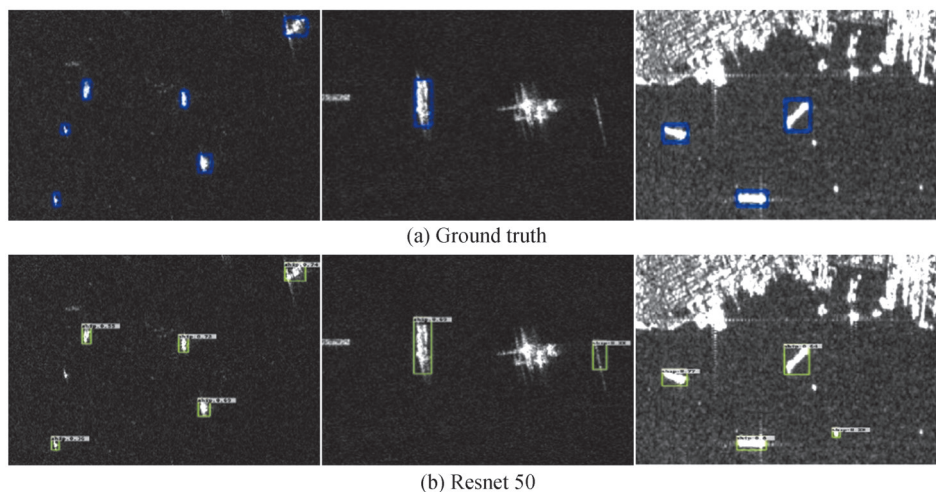
在特征提取阶段,不同主干网络提取特征的能力不同,Resnet50^[20]有4个残差块,conv2_x、conv3_x、conv4_x、conv5_x的卷积层数分别为3、4、6、3。Resnet101^[21]也包含四个残差块,每块的卷积层数分别为3、4、23、3。两种主干网络均在ImageNet数据集训练生成预训练模型参数,利用跳跃连接将原始特征图重新构造到新特征图中,保证网络在一定深度下不会出现过拟合现象。从表2可以看出,Resnet101网络比Resnet50网络召回率高0.51%,精确率高0.98%,平均精度高0.2%,Resnet101的检测效果优于Resnet50。

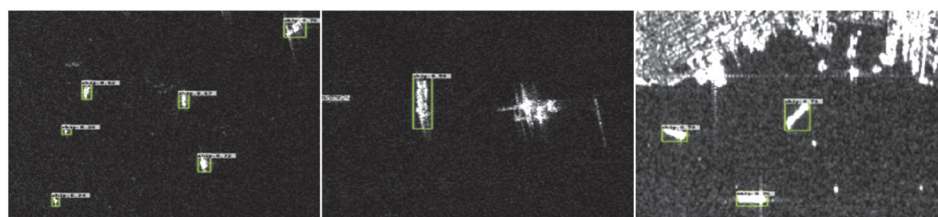
表2 不同骨干网络的检测结果

Table 2 Detection performance of different backbone networks

| Backbone | Recall/% | Precision/% | Average precision/% |
|-----------|----------|-------------|---------------------|
| Resnet50 | 97.07 | 95.13 | 96.53 |
| Resnet101 | 97.58 | 96.11 | 96.73 |

不同主干网络检测效果如图8所示,(a)为真值图,(b)为Resnet50主干网络检测效果,(c)为Resnet101主干网络检测效果。Resnet101具有更深的网络结构,能够提取丰富的抽象特征,具有足够的参数刻画模型在空间上的分布。所以经过每个残差块获得的特征图抽象特征越丰富,判别目标类别的准确率越高。使用Resnet101的主干网络,模型能够减小舰船目标漏检、误检情况,达到更高的检测精度。





(c) Resnet 101

图8 不同主干网络的检测效果

Fig.8 Different backbones network detection results

2.4 消融实验

SAR图像中小尺度舰船目标在经过多次卷积操作后容易丢失,同时受复杂背景信息影响,舰船目标位置信息模糊,因此增强位置信息和语义信息十分重要。从表3可见,特征图经过像素级消噪提高舰船目标区域权重,与原始模型相比平均精度提升了7.21%。语义增强模块提取更加丰富的语义信息,与原始模型相比平均精度提升了7.17%。在两个模块共同作用下,模型召回率达到97.58%,精确率达到96.11%,平均精度达到96.73%。

表3 不同模块的检测结果

Table 3 Detection performance of different models

| Pixel-level noise reduction | Semantic enhancement | Recall/% | Precision/% | Average precision/% |
|-----------------------------|----------------------|----------|-------------|---------------------|
| × | × | 89.94 | 94.26 | 89.23 |
| √ | × | 96.94 | 96.33 | 96.44 |
| × | √ | 96.93 | 95.84 | 96.40 |
| √ | √ | 97.58 | 96.11 | 96.73 |

提出了像素级消噪模块和语义增强模块,其中语义增强模块在增强特征图语义信息的同时也体现了金字塔网络的作用,在语义增强模块中将 3×3 的卷积拆分成 3×1 和 1×3 的卷积。卷积核为 3×3 的卷积包含 $3 \times 3 \times C_i \times C_o$ 的参数量,卷积核为 1×3 和 3×1 的卷积参数量之和为 $1 \times 3 \times C_i \times C_o + 3 \times 1 \times C_i \times C_o$,在该模块减少了33.3%的参数量。像素级消噪模块需要生成不同尺度的注意力图并在训练过程中利用掩码损失函数优化注意力图,因此该模块引入大量参数,本文方法在SSDD数据集的性能分析见表4。

表4 模型性能分析

Table 4 Model performance analysis

| Pixel-level noise reduction | Semantic enhancement | Params/M | Training time /FPS | Testing time /FPS |
|-----------------------------|----------------------|----------|--------------------|-------------------|
| × | × | 152.76 | 6.89 | 22.64 |
| √ | × | 160.84 | 6.49 | 20.50 |
| × | √ | 151.40 | 6.10 | 20.92 |
| √ | √ | 159.49 | 5.99 | 21.09 |

本文检测模型的平均精度指标具有一定的提升,原因有:1)像素级消噪模块增强舰船目标权重,抑制非舰船区域权重,使得舰船在复杂背景和相干斑噪声的场景中位置信息更加丰富,减少了舰船目标被漏检的数量;2)语义增强模块丰富了特征图中的语义信息,能够在分类分支被精准识别,避免非舰船物体被识别为舰船目标而导致模型虚警率增高;3)Transformer Ecoder模块能够保证特征图始终保持全局感受野,解决了因感受野不足导致大尺度舰船不能完全映射到原始图像的问题,使得候选框具有更好的回归效果。因此,本文方法能够有效解决模型中出现的漏检、误检问题。不同模块检测效果图如图9所示,其中(a)为真值图,(b)为原始模型检测效果,(c)为经过像素级消噪的效果,(d)为经过语义增强的效果,(e)为像素级消噪和语义增强的效果。

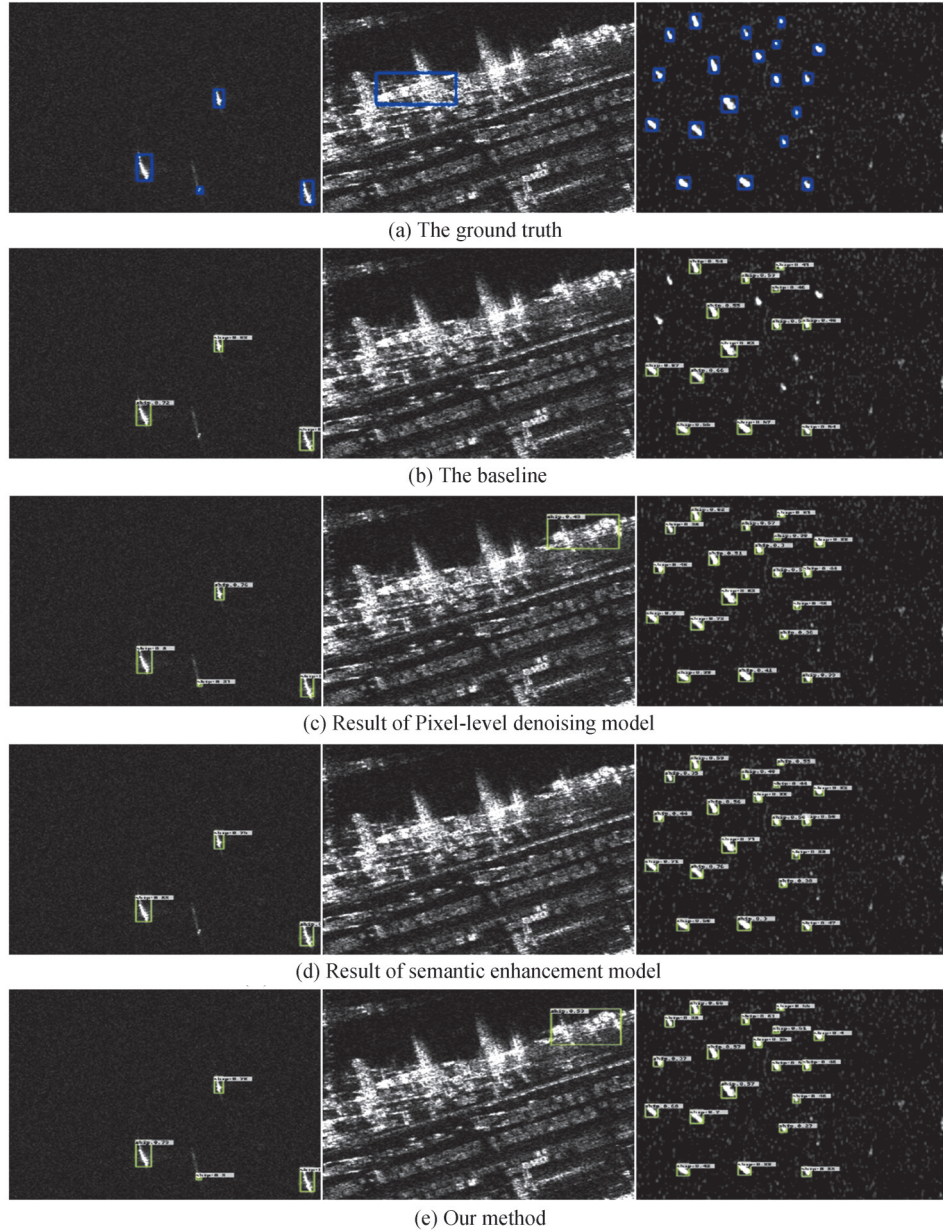


图9 不同模块的检测效果
Fig.9 Different modules detection effect

2.5 不同方法实验结果及分析

为了对比本文模型的检测效果,选择Faster R-CNN^[6]、Retinanet^[22]、FPN^[23]、WANG^[24]、R2FA-Det^[25]等主流的基于深度学习的检测模型。从表5中可以得出,在SSDD数据集下,本文模型检测精度比Faster R-CNN有9.47%的提升,比Retinanet有9.82%的提升,比FPN有6.04%的提升。与其它模型相比,本文模型的召回率、精确率和平均精度都达到了理想效果。

为了进一步验证本文方法的鲁棒性和泛化能力,将其与其它3种方法在SAR-Ship-Dataset数据集上的检测效果进行对比。从表6可以看出,本文方法在召回率和平均精度两个指标上有更好的提升。

Faster R-CNN、Retinanet为基于锚框的检测模型,受固定锚框参数影响不能完全适合多尺度舰船目标的SAR图像,降低了模型检测效果。FPN采用特征融合机制,提高了模型检测效果,但是受限于高层特征语义信息的不足,模型检测效果并不理想。提出的像素级去噪和语义增强的无锚框检测模型,利用中心点

表5 不同方法在SSDD数据集上的检测结果
Table 5 Detection performance of different methods on SSDD

| Methods | Backbone | Recall/% | Precision/% | Average precision/% |
|--------------------------|--------------|----------|-------------|---------------------|
| Faster R-CNN | Resnet101 | 88.54 | 88.09 | 87.26 |
| Retinanet | Resnet101 | 89.17 | 87.94 | 86.91 |
| FPN | Resnet101 | 92.22 | 86.41 | 90.69 |
| YOLOv4 | CSPDarknet53 | — | — | 95.60 |
| WANG ^[24] | ShuffleNetV2 | — | — | 94.70 |
| R2FA-Det ^[25] | Resnet50 | — | — | 94.72 |
| FBR-Net ^[26] | — | 92.79 | 94.01 | 94.10 |
| Ours | Resnet101 | 97.58 | 96.11 | 96.73 |

表6 不同方法在SAR-Ship-Dataset上的检测结果
Table 6 Detection performance of different methods on SAR-Ship-Dataset

| Methods | Backbone | Recall/% | Precision/% | Average precision /% |
|--------------|-----------|----------|-------------|----------------------|
| Faster R-CNN | Resnet101 | 80.24 | 81.39 | 76.35 |
| Retinanet | Resnet101 | 72.17 | 90.62 | 71.03 |
| FPN | Resnet101 | 83.40 | 90.80 | 81.38 |
| Ours | Resnet101 | 91.86 | 81.45 | 88.86 |

到边框的偏移量作为候选框参数,能自适应生成合适的候选框。同时像素级消噪能够抑制背景信息,突出舰船目标信息,使得特征图中舰船目标位置信息被增强。语义增强模块使得高层特征中语义信息更加丰富,经过特征融合后的特征图空间位置信息和语义信息更加平衡,在SAR图像舰船检测中具有更好的检测精度。为了更直观展现本文模型网络特征表达能力,采用不同模型的P-R对比,如图10所示。图11为不同模型检测效果,其中(a)为真值图,(b)为Faster R-CNN检测效果,(c)为Retinanet检测效果,(d)为FPN检测效果,(e)为本文模型检测效果。

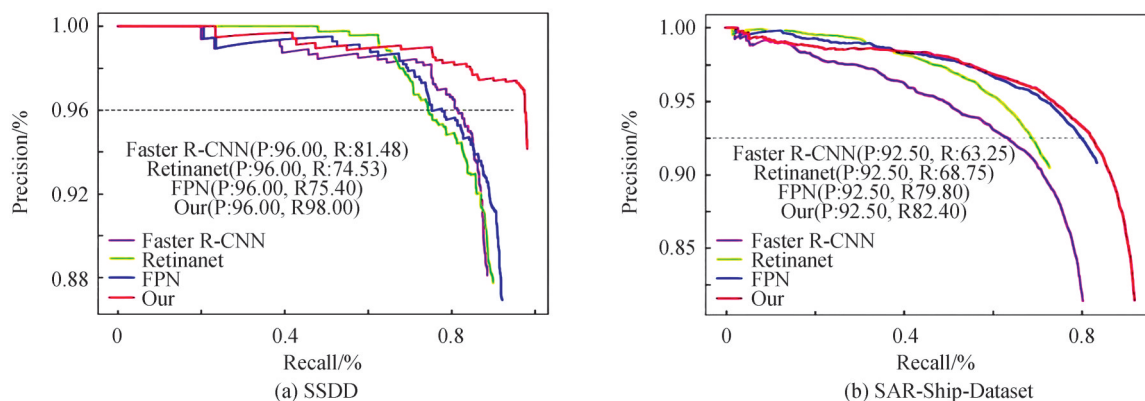
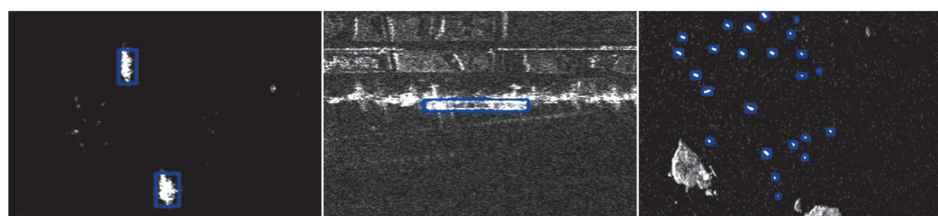


图10 不同对比方法的P-R图
Fig.10 P-R diagram of different comparison methods



(a) The ground truth

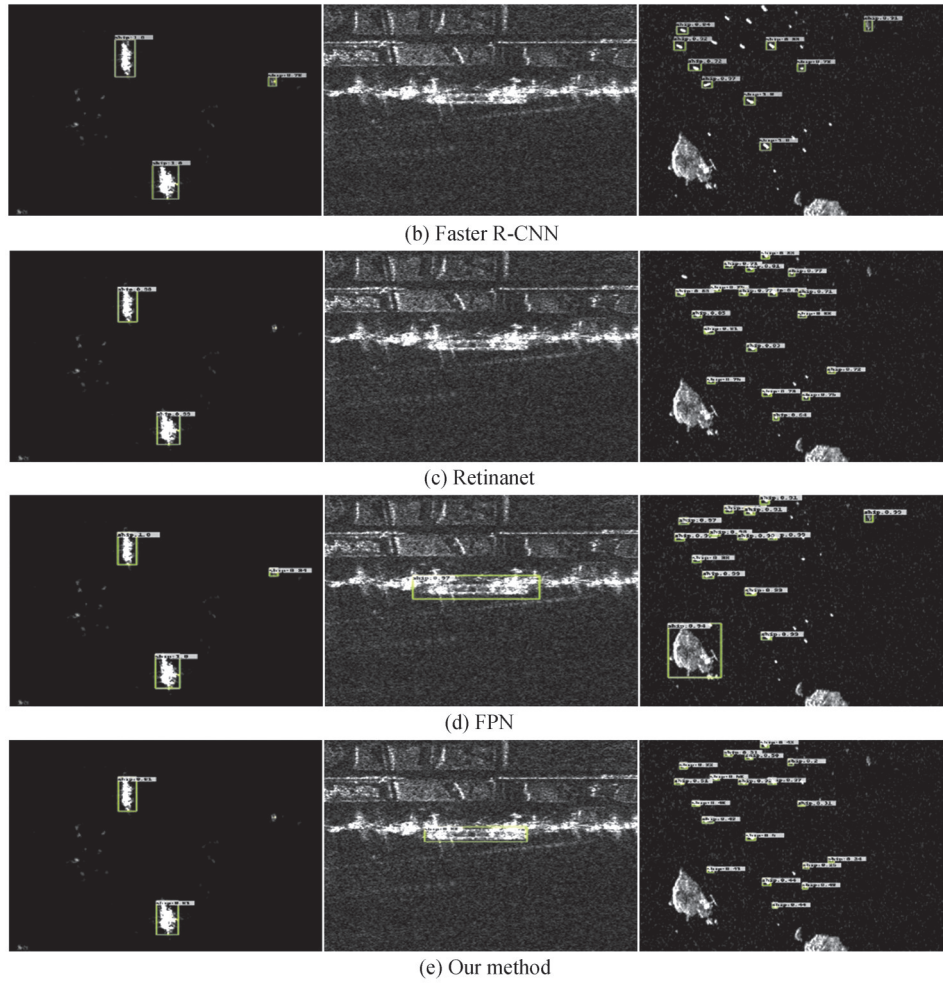


图11 不同方法的检测效果
Fig.11 Different methods detection effect

2.6 不同场景下实验结果及分析

合成孔径雷达成像原理独特,在不同场景下图像差别较大,因此为了验证在不同场景下模型检测效果,将验证集分为近海岸场景和远海场景,同时以 $60 \text{ px} \times 60 \text{ px}$ 将验证集划分为小尺度目标和大尺度目标。近海岸场景包含较多形似舰船目标的物体,舰船目标容易淹没在复杂背景中。远海场景受相干斑噪声影响,导致舰船目标边缘模糊。小尺度舰船在经过多次卷积后容易丢失,严重影响模型检测精度。从表7可见,本文模型针对多种不同场景均表现出较好的检测效果。

表7 不同场景的检测结果
Table 7 Detection performance of different scenes

| Scenes | Methods | Recall/% | Precision/% | Average precision/% |
|--------------------|----------|----------|-------------|---------------------|
| Inshore | Baseline | 79.01 | 87.07 | 77.79 |
| | Our | 91.98 | 89.76 | 90.00 |
| Offshore | Baseline | 92.78 | 96.01 | 92.24 |
| | Our | 99.04 | 97.78 | 98.53 |
| Small-scale target | Baseline | 90.79 | 90.78 | 90.66 |
| | Our | 97.60 | 96.78 | 96.85 |
| Large-scale target | Baseline | 89.84 | 94.65 | 89.13 |
| | Our | 97.37 | 92.50 | 94.64 |

本文模型能在多种场景下表现出较好的检测效果,是因为像素级消噪模块能抑制相干斑噪声和非目标信息,使得舰船目标在各种场景下都能突出位置信息。在远海场景中主要受相干斑噪声影响,经过像素级消噪模块处理,图像中更多的舰船目标被检测到,召回率提升明显。语义增强模块使得输入检测头的特征图语义信息更加丰富,同时 Transformer Encoder 模块增强目标和特征图之间的上下文信息,避免舰船目标被漏检。不同场景检测效果图如图 12 所示,其中(a)为近海岸场景检测效果,(b)为远海场景检测效果,(c)为大尺度舰船检测效果,(d)为小尺度舰标检测效果。

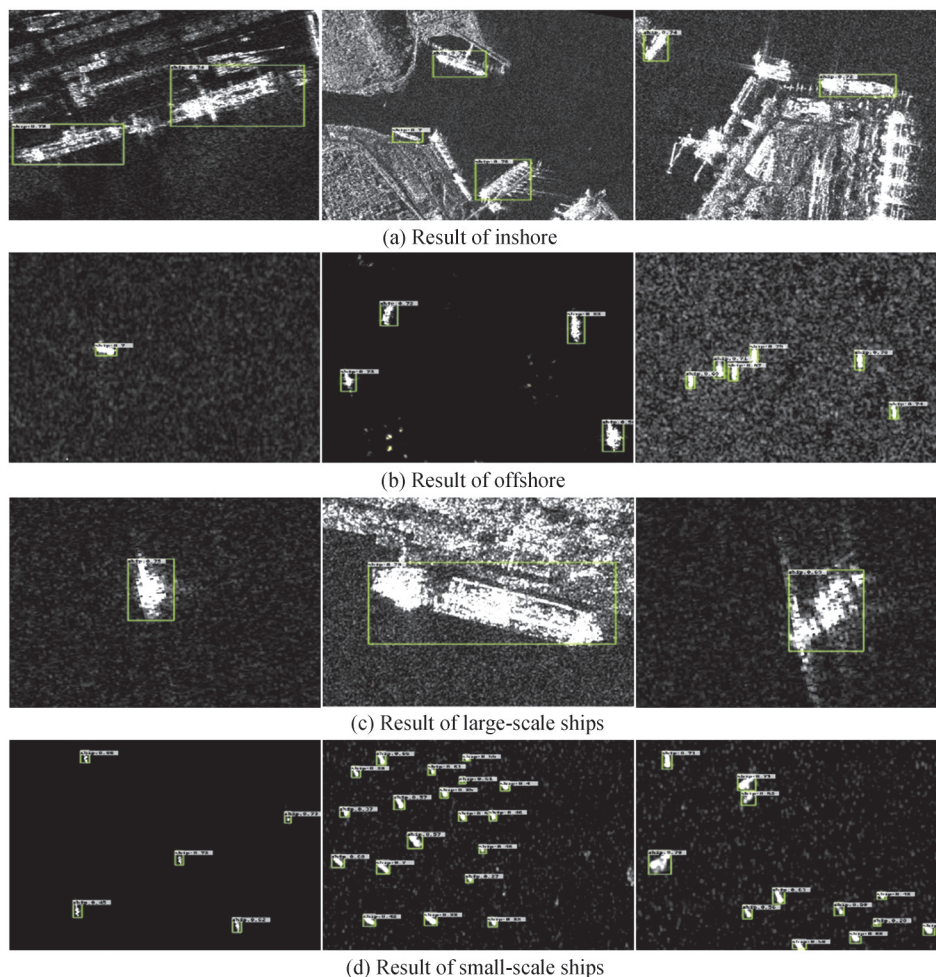


图 12 不同场景的检测效果
Fig.12 Different scenes detection effect diagram

3 结论

本文提出了像素级去噪和语义增强的检测模型,在公开遥感数据集 SSDD 上验证了该模型针对不同场景、不同尺度 SAR 图像舰船检测的有效性。无锚框的检测器节省更多超参数,避免了人工设计候选框不精确的情况,与基于锚框的检测模型相比检测效果更好。像素级消噪模块使相干斑噪声和近海岸物体信息影响的特征图中舰船目标位置突出,同时经过掩码损失函数优化掩码特征图,使特征图中目标位置更加精确。语义增强通过多种不同卷积核提取多种特征,并以通道维度进行特征融合,该模块生成的特征图语义信息更加丰富,最后采用自上而下将高层特征的语义信息融合到底层高分辨率特征图,包含丰富语义信息的特征图在分辨舰船目标时具有重要作用。编码模块通过模型不断迭代,学习舰船目标的位置信息,增强了舰船目标之间的依赖关系,在大尺度目标检测中表现优秀,经过位置信息增强,小尺度舰船检测也达到理想检测效果。本文提出的检测模型在近海岸、远海和带有海洋杂波的多种背景下获得较高的检测精度。但是由于候选框为水平框,候选框内不仅有舰船目标,同时还包含大量的背景信息,为了消除候选框中背景信息,

接下来的工作将着重研究基于旋转框的SAR图像舰船检测。

参考文献

- [1] ZHANG T, ZHANG X, SHI J, et al. Depthwise separable convolution neural network for high-speed SAR ship detection [J]. *Remote Sensing*, 2019, 11(21): 2483.
- [2] ZHU Lei, LI Jingman, PAN Yang, et al. SAR image despeckling algorithm using non-local means with adaptive filtering strength[J]. *Journal of Electronics & Information Technology*, 2021, 43(5): 1258-1266.
朱磊,李敬曼,潘杨,等.自适应调节滤波强度的SAR图像非局部平均抑斑算法[J].*电子与信息学报*, 2021, 43(5): 1258-1266.
- [3] HOU B, CHEN X, JIAO L. Multilayer CFAR detection of ship targets in very high resolution SAR images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2014, 12(4): 811-815.
- [4] AI J, YANG X, SONG J, et al. An adaptively truncated clutter-statistics-based two-parameter CFAR detector in SAR imagery[J]. *IEEE Journal of Oceanic Engineering*, 2017, 43(1): 267-279.
- [5] LI T, LIU Z, XIE R, et al. An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 11(1): 184-194.
- [6] LI J, QU C, SHAO J. Ship detection in SAR images based on an improved faster R-CNN[C]. 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA). IEEE, Beijing, China, 2017: 1-6.
- [7] HUANG Z, DUMITRU C O, PAN Z, et al. Classification of large-scale high-resolution sar images with deep transfer learning[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 18(1): 107-111.
- [8] WANG Y, WANG C, ZHANG H, et al. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery[J]. *Remote Sensing*, 2019, 11(5): 531.
- [9] LAW H, DENG J. Cornernet: detecting objects as paired keypoints [C]. *Proceedings of the European conference on computer vision (ECCV)*, Glasgow, UK 2018: 734-750.
- [10] DUAN K, BAI S, XIE L, et al. Centernet: keypoint triplets for object detection [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Korea, 2019: 6569-6578.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]. *Advances in Neural Information Processing Systems*, CA, USA, 2017: 5998-6008.
- [12] SHI Wenxu, JIANG Jinhong, BAO Shengli. Ship detection method in remote sensing image based on feature fusion [J]. *Acta Photonica Sinica*, 2020, 49(7): 0710004.
史文旭,江金洪,鲍胜利.基于特征融合的遥感图像舰船目标检测方法[J].*光子学报*, 2020, 49(7): 0710004.
- [13] QIAN Kun, ZHOU Huixin, QIN Hanlin, et al. Infrared dim-small target tracking based on guide filter and spatio-temporal context learning [J]. *Acta Photonica Sinica*, 2015, 44(9): 0910003.
钱琨,周慧鑫,秦翰林,等.基于引导滤波与时空上下文的红外弱小目标跟踪[J].*光子学报*, 2015, 44(9): 0910003.
- [14] WANG W, XIE E, LI X, et al. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions[J]. *arXiv preprint arXiv:2102.12122*, 2021.
- [15] TIAN Z, SHEN C, CHEN H, et al. Fcos: fully convolutional one-stage object detection [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Korea, 2019: 9627-9636.
- [16] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017: 2980-2988.
- [17] WANG Y, WANG C, ZHANG H, et al. A SAR dataset of ship detection for deep learning under complex backgrounds [J]. *Remote Sensing*, 2019, 11(7): 765.
- [18] WEI S, ZENG X, QU Q, et al. HRSID: a high-resolution SAR images dataset for ship detection and instance segmentation[J]. *IEEE Access*, 2020, 8: 120234-120254.
- [19] DING Bin, XIA Xue, LIANG Xuefeng, et al. Sea clutter data augmentation method based on deep generative adversarial network[J]. *Journal of Electronics & Information Technology*, 2021, 43(7): 1985-1991.
丁斌,夏雪,梁雪峰,等.基于深度生成对抗网络的海杂波数据增强方法[J].*电子与信息学报*, 2021, 43(7): 1985-1991.
- [20] THECKEDATH D, SEDAMKAR R R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks [J]. *SN Computer Science*, 2020, 1(2): 1-7.
- [21] XU Z, SUN K, MAO J. Research on ResNet101 network chemical reagent label image classification based on transfer learning [C]. 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), IEEE, Weihai, China, 2020: 354-358.
- [22] AHMAD M, ABDULLAH M, HAN D. Small object detection in aerial imagery using RetinaNet with anchor optimization [C]. 2020 International Conference on Electronics, Information, and Communication (ICEIC), IEEE, Barcelona, Spain, 2020: 1-3.
- [23] KIM S W, KOOK H K, SUN J Y, et al. Parallel feature pyramid network for object detection [C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018: 234-250.
- [24] WANG Yanni, SUN Xuesong, YU Lixian. Lightweight synthetic aperture radar ship detection algorithm with enhanced receptive field [J]. *Acta Photonica Sinica*, 2022, 51(2): 0210008.
王燕妮,孙雪松,余丽仙.增强感受野的轻量化合成孔径雷达船舶检测算法[J].*光子学报*, 2022, 51(2): 0210008.
- [25] CHEN S, ZHANG J, ZHAN R. R2FA-Det: delving into high-quality rotatable boxes for ship detection in SAR images

- [J]. Remote Sensing, 2020, 12(12): 2031.
- [26] FU J, SUN X, WANG Z, et al. An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(2): 1331-1344.

SAR Image Denoising and Semantic Enhancement for Object Detection

QU Haicheng, SHEN Lei

(College of Software, Liaoning Technical University, Huludao, Liaoning 125105, China)

Abstract: The imaging process of the synthetic aperture radar system is not affected by time and weather, and can achieve all-day and all-weather imaging of the target. It has a wide detection range and generates high-resolution images. Therefore, it is widely used in the military and civilian fields. In recent years, satellites of "GF-3" and satellites of "HJ-1" were successively launched to fill the gap in synthetic aperture radar technology in China. However, synthetic aperture radar images still have shortcomings in the target detection process. Ship targets in SAR images are sparse and of various scales. The anchor box-based detection model relies too much on manually designed candidate boxes, which cannot adapt to all ship targets, and the parameters of the candidate boxes consume a lot of computing resources. The background of the synthetic aperture radar image is complex, and the ship target can easily disappear in the complex background, which leads to the missed detection of the detection model. SAR images contain a large number of small-scale ship targets, which are easily lost after multiple convolutions. Coherent speckle noise present in SAR images causes blurring of ship edges. To address the impact of the above problems, this paper proposes a detection model for pixel-level denoising and semantic enhancement. First, the pixel-level denoising module uses the prediction mask to generate the attention map of $[0, 1]$, multiplies the feature map and the attention map pixel by pixel to achieve denoising, and optimizes the attention map using the cross-entropy loss. The denoising module can enhance the weight of the target area of the ship, suppress the weight of the non-target area, and enhance the difference between the ship target information and the background information in the feature map. Second, the semantic enhancement module enhances the semantic information contained in the feature map, and uses asymmetric convolutional layers to extract features of different dimensions, preventing candidate boxes with high IOU scores and low classification confidence from being suppressed. The transformer encoder is introduced in the semantic enhancement module to improve the context information between the ship target and the feature map, and enhance the dependency between the ship target and the image. Finally, the de-noised feature map with rich semantic information is fed into the detection head. In the public data set SSDD test, the model detection accuracy reaches 96.73%, the detection accuracy for small-scale ships reaches 96.85%, the detection accuracy for large-scale ships reaches 96.41%, the detection accuracy for distant sea scenes reaches 98.53%, and the detection accuracy for near-shore scenes reaches 90.00%. Compared with the detection effect of the current mainstream model in the SAR-Ship-Dataset dataset, the proposed model is verified to have a better detection effect. The experimental results show that the pixel-level denoising module uses the attention map to change the weight of the ship target area can better distinguish the target area from the non-target area. In the SAR image with complex background, the ship target position information is more obvious. The position information of small-scale ships is also enhanced, which solves the problem of small-scale ships loss after multi-layer convolution operations. The semantic enhancement module improves the model's ability to recognize ship targets and reduces the suppression of high-quality candidate boxes. Therefore, the model in this paper can effectively reduce the missed detection rate and false detection rate of the model.

Key words: Target detection; Image processing; Deep learning; SAR image; Pixel-level denoising; Semantic enhancement; Transformer

OCIS Codes: 100.3008; 110.2970; 280.6730