

基于 RetinaNet 的 SAR 图像舰船目标检测

刘洁瑜¹, 赵彤^{1†}, 刘敏²

(1. 火箭军工程大学 导弹工程学院, 陕西 西安 710025;

2. 湖南大学 电气与信息工程学院, 湖南 长沙 410082)

摘要:在合成孔径雷达图像舰船目标检测中,由于背景复杂多变,传统的基于人工特征的目标检测方法效果较差.基于深度学习中的单阶段目标检测算法 RetinaNet,结合合成孔径雷达图像本身特征信息较少的特点,采用了多特征层融合的思想,改进了网络特征提取能力,提出了相适应的损失函数的计算方法.采用 SAR 图像舰船目标检测数据集(SSDD)对网络进行训练,并通过样本增强和迁移学习的方法提升算法的鲁棒性和收敛速度.通过实验与其他基于深度学习的目标检测算法所得结果进行比较,结果表明本算法具有更高的检测精度.

关键词:合成孔径雷达(SAR)图像;舰船目标检测;深度学习;RetinaNet

中图分类号:TP751.1

文献标志码:A

Ship Target Detection in SAR Image Based on RetinaNet

LIU Jieyu¹, ZHAO Tong^{1†}, LIU Min²

(1. College of Missile Engineering, Rocket Force University of Engineering, Xi'an 710025, China;

2. College of Electrical and Information Engineering, Hunan University, Changsha 410082, China)

Abstract: In the detection of Synthetic Aperture Radar (SAR) image ship targets, the traditional artificial feature-based target detection method is less effective due to the complex background. Based on the single-stage target detection algorithm RetinaNet in deep learning, this paper combined the characteristics of SAR image with less feature information, adopted the idea of multi-feature layer fusion and proposed a more appropriate loss function calculation method. Then we used the SAR ship detection dataset(SSDD) data set to train the network, and improve the robustness and convergence speed of the algorithm through data augmentation and transfer learning. Finally, we compared the method with other target detection algorithms based on deep learning through experiments. The results show that our method has higher detection accuracy.

Key words: Synthetic Aperture Radar(SAR) image; ship target detection; deep learning; RetinaNet

SAR 图像是通过高频率的电磁波与合成孔径原理相结合对一片区域进行主动感知生成的高分辨率图像,其成像受外界环境影响较小,可以主动对目标

完成全天候和远距离的侦查^[1].因此,基于 SAR 图像的目标检测在敌舰侦察领域得到了广泛的应用.传统的 SAR 目标检测一般采用恒虚警率检测算法,其

* 收稿日期:2019-04-17

基金项目:国家自然科学基金青年基金资助项目(61503392), National Natural Science Foundation of China(61503392)

作者简介:刘洁瑜(1970—),女,陕西西安人,火箭军工程大学教授,博士

† 通讯联系人, E-mail:601080018@qq.com

采用检测器统计 SAR 图像中的每个像素点周围杂波的分布特性,自适应生成检测阈值,并将目标像素点从背景中提取出来,而后通过形态学处理和负样本剔除等操作得到目标中心,进而完成检测.此类方法需根据图像特征,人为地设定检测器和约束条件,因此鲁棒性和实用性较差^[2].尤其在舰船检测中,由于海洋潮汐等背景杂波的干扰,该类算法的检测精度会受到较大影响.

近年来,基于卷积神经网络(Convolutional Neural Network, CNN)的深度学习目标检测算法成为了研究热点. Girshick 等人^[3]提出了基于区域卷积神经网络的目标检测算法,使得基于深度学习的目标检测相对于传统方法在精度和速度上取得巨大突破,在 PASCAL VOC 2012 数据集上平均检测精度达到了 53.3%,开启了基于深度学习的目标检测的热潮.之后,陆续出现了 SPPNet^[4]、Faster R-CNN^[5]、Faster R-CNN^[6]、YOLO^[7-9]、SSD^[10]等算法.到目前为止,目标检测已经被构建为分类和回归的问题来进行解决——以 Faster R-CNN 为代表的两阶段方法和以 SSD 和 YOLO 为代表的单阶段方法,其中 Faster R-CNN 准确率更高,而 SSD 的速度占优.之后, Lin 等^[11]认为单阶段算法精度不及两阶段算法的本质原因是其正负样本的不均衡,并为此提出了 Focal loss 损失函数使得单阶段的检测网络在精度上有了很大的提升,其提出的 RetinaNet 也成为了当下最佳的检测网络方法之一.

得益于深度学习算法的发展, SAR 图像目标检测也迎来了很大变革.李健伟等^[12]构建了国内首个 SAR 图像舰船公开数据集 SSDD 并用其训练 Faster R-CNN 网络,通过对比传统方法,证明了深度学习算法的优越性. Shahzad 等人^[13]利用级联的全卷积神经网络实现了高分辨率 SAR 图像中对建筑物的有效检测.然而,基于深度学习的 SAR 图像检测中的各类算法在精度和速度上还有改进空间.

本文基于现阶段较为先进的 RetinaNet 检测算法进行改进,结合 SAR 图像本身信息相对 RGB 图像

较少的特点,在更底层的图像特征层上进行处理,融合了基础提取网络的多层特征信息;之后在多尺度的特征层对预选框进行回归.采用 SSDD 数据集对网络进行训练,并通过样本增强和迁移学习的方法提升算法的鲁棒性、收敛速度.最终通过实验与其他典型的深度学习算法进行对比,结果表明,本算法具有更好的效果.

1 算法结构

1.1 特征提取网络

本算法的特征提取部分与 RetinaNet 相同,采用了深度残差网(ResNet101)^[14].一般来说,神经网络的性能随着网络深度的增加而增加,然而如果简单地增加深度,会导致梯度弥散或梯度爆炸.针对该问题,解决的方法是采用 relu 激活函数和中间层的正则化(Batch Normalization)^[15],这样可以训练几十层的网络.但之后便会出现模型退化问题,即使深度继续增加,网络的准确率也不会上升甚至出现下降.而 ResNet 采用跨层连接的思想使用恒等映射直接将前一层输出传到最后,即使网络深度增加,也不会使网络的准确度下降.其网络结构如图 1(a)所示,其中每组括号代表一个结构块(block),其结构如图 1(b)所示(以 Conv3 为例).整个网络包含 101 个卷积层(这里省略了最后的全连接层).每个 block 中包含 3 个卷积层,其参数 $\{n \times n, m\}$ 中, n 代表卷积核的大小, m 代表卷积核的数量. Conv1 中的 stride 为卷积操作的步长,在每个 block 中第一个卷积层的步长为 2,其余步长均为 1,因此在每一个 block 结构中,特征图的尺寸会下降一半. Conv2 中的 maxpool 代表最大池化操作,其池化单元大小为 3×3 ,步长为 2.

1.2 特征融合结构

传统的 RetinaNet 采用了特征金字塔(Feature Pyramid Networks, FPN)^[16]的结构进行特征融合,其融合结构如图 2 所示. Conv5_9 层首先经过 1×1 的卷



图 1 特征提取网络参数
Fig.1 Feature extraction network parameter

积生成 Conv6 层,而后经过线性插值上采样得到和 Conv4_69 同样的尺度,之后再与 Conv4_69 经过 1×1 的卷积后进行融合生成 Conv7. 同样,Conv7 经过上采样生成和 Conv3_12 相同的尺度后与经过 1×1 的卷积的 Conv3_12 融合生成 Conv8. 在检测过程中,浅层的细节信息更有利于目标的定位,而高层的语义信息则更有益于目标的识别. FPN 巧妙地将 ResNet 浅层的细节信息和高层的语义信息结合起来,以更好地对局部信息进行约束. 然而在结合的过程中,FPN 仅仅将高层信息上采样后与浅层信息同纬度相加. 这样的结合方式相对来说较为粗糙,忽略了特征图经双线性插值引入的结构误差. 同时,由于 SAR 图像本身仅包含一个维度的有效信息,对误差的影响更加敏感,因此更有效地融合特征信息是提高检测效率的关键.

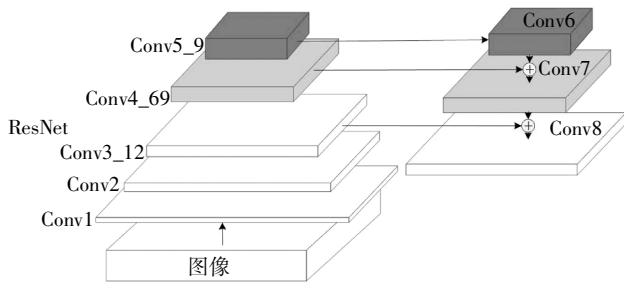


图 2 ResNet 的 FPN 结构

Fig.2 FPN structure in ResNet

本文提出了一种新的特征融合网络,其结构如图 3 所示. Fusion1 和 Fusion2 为特征融合单元, Fusion1 以适当的方式融合 3 层不同表征能力的特征图,提取了更多 Sar 图像的潜在特征,生成更准确的特征金字塔. Fusion2 将金字塔相邻的不同尺度的特征图进行融合,建立了深浅层之间的耦合关系. 其结构参数分别如表 1 (以 Conv6 和 Conv9 的生成为例)

和表 2 所示. Fusion1 对 ResNet 的 Conv2、Conv3、Conv4 和 Conv5 层后 3 个 block 分别进行 1×1 同纬度的卷积运算,而后并联在一起并再次通过 1×1 的卷积对有效信息进行融合,生成了 Conv6、Conv7、Conv8 和 Conv9. 这 4 个特征信息更丰富和准确的特征层,感受野逐渐变大,特征的细粒度也逐渐减小. 其中 Conv6 是从 Conv2 特征图中融合得到,包含更多的目标细节信息,因此其对小目标更敏感,对目标的定位更有效. Fusion2 结构在新的金字塔结构上建立了深层特征到浅层特征的链接. 对小尺寸的特征图首先进行双线性插值,而后通过 1×1 卷积跨通道地整合信息并降低特征层维度,通过 3×3 的卷积减少特征图的结构误差,最后与经过 1×1 卷积的低维特征图融合. 融合过程采用相加的方式, 3×3 的卷积是为消除混叠效应. 相比于原始网络的 FPN 结构,本网络的输出层提取了更浅层的特征信息,各特征图的尺度为原输出特征图的 4 倍.

表 1 Fusion1 的融合结构

Tab.1 Fusion structure of Fusion1

融合层	链接层	结构	融合
Conv6	Conv2_3	Conv $1 \times 1, 256$ BN	Cat Conv $1 \times 1, 256$ BN
	Conv2_6	Conv $1 \times 1, 256$ BN	
	Conv2_9	Conv $1 \times 1, 256$ BN	

Conv9	Conv5_3	Conv $1 \times 1, 2048$ BN	Cat Conv $1 \times 1, 2048$ BN
	Conv5_6	Conv $1 \times 1, 2048$ BN	
	Conv5_9	Conv $1 \times 1, 2048$ BN	

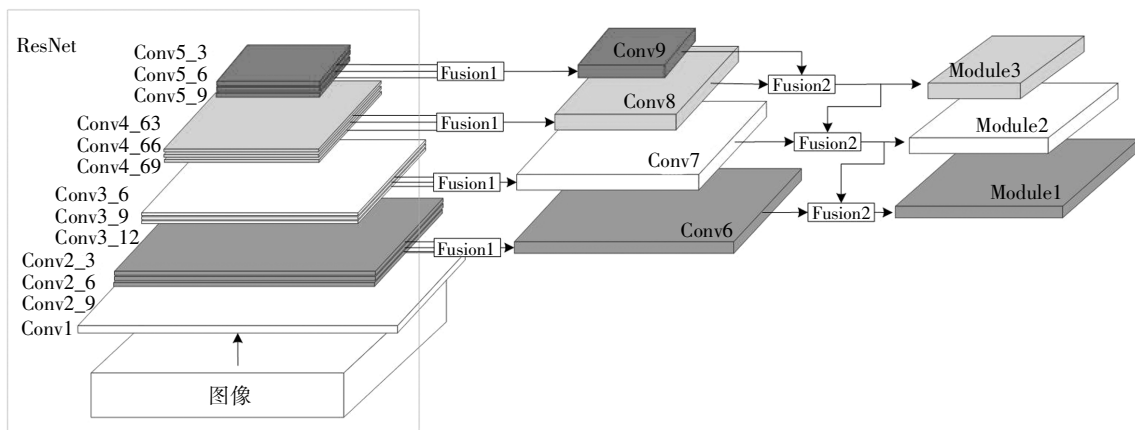


图 3 新的特征融合结构

Fig.3 New feature fusion structure

表2 Fusion2的融合结构
Tab.2 Fusion structure of Fusion2

融合层	链接层	结构	融合
Module3	Conv9	双线性插值	SUM Conv 3×3, 1 024
		Conv 1×1, 1 024	
	Conv 3×3, 1 024		
	Conv 1×1, 1 024		
Conv8	BN		
	双线性插值		
Module2	Module3	Conv 1×1, 512	SUM Conv 3×3, 512
		Conv 3×3, 512	
	Conv 1×1, 512		
	Conv7		
Conv6	BN		
	双线性插值		
Module1	Module2	Conv 1×1, 256	SUM Conv 3×3, 256
		Conv 3×3, 256	
	Conv 1×1, 256		
	Conv6		
Conv5	BN		
	双线性插值		

1.3 边框回归和分类

改进的融合结构在 Module1、Module2 和 Module3 三个特征层上分别通过 4 个 3×3 的卷积层后,进行位置回归和分类,生成固定数目的初始框。一般的回归部分,位置和种类的回归在同一个特征图上,而为了使两个不同的损失相互之间没有影响,增强特征的代表能力,RetinaNet 的位置回归和分类分开进行。以 Module2 为例,其回归结构如图 4 所示。在每个特征层的锚点上生成 9 个初始框,通过和真实框的匹配得到边界框。计算真实框和边界框之间的损失函数,并通过平均梯度下降法建立优化器,最终完成网络的训练。

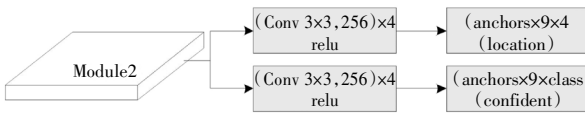


图4 Module2的回归结构
Fig.4 Regression structure of Module2

RetinaNet 的损失函数和其他检测算法相同,分为位置损失和分类损失。位置损失的计算如式(1)所示。

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smooth}_{L1}(l_i^m - \hat{g}_j^m) \quad (1)$$

式中: x_{ij}^k 表示第 k 类的第 i 个边界框和第 j 个真实框

之间的交并比(IOU); l_i^m 和 \hat{g}_j^m 分别代表边界框的 4 个位置参数和真实框的 4 个参数。其中 L1 loss 如式(2)所示。

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{其他} \end{cases} \quad (2)$$

考虑正负样本的不平衡问题,RetinaNet 采用了 Focal loss 作为分类损失。Focal loss 是在交叉熵损失函数(CE loss)中改进的,CE loss 如式(3)所示。

$$\text{CE}(y) = -\frac{1}{n} \sum_{i=1}^n [I\{y = y'_i\} \log(p_i)] \quad (3)$$

式中: n 代表边界框的总数, y'_i 为第 t 个边界框所对应的正确类别; p_i 为第 t 个边界框的预测类别; I 为符号函数; 大括号内为判断条件。Focal loss 在 CE loss 前加上了权重,如式 4 所示:

$$\text{FL}(y) = \alpha(1 - p_i)^\gamma \text{CE}(y) \quad (4)$$

式中: α 和 γ 均为权重因子,一般分别取 0.25 和 2。可以看出,当正负样本不均匀时,由负样本造成的损失会大幅度下降;而当样本均匀时,损失仅适当降低,从而减弱了样本不均造成的分类错误。

由于改进网络结构特征层的尺度是原图的 4 倍,因此初始框的数目为原图的 4 倍,属于背景的边界框与属于前景的边界框之间的比例会更大。这样,即使 Focal loss 可以适当平衡正负样本的不均衡,也会对损失函数的训练带来退化。同时,由于初始框的大量增加,网络计算量增加,训练效率也会受到较大影响。因此,在进行分类和回归前,采用了 OHEM (Online Hard Example Mining) 算法^[7]的思想,对正负样本进行筛选。OHEM 也是基于 CE 算法的改进,经证明^[10-11],其在平衡正负样本上具有较好的鲁棒性。

在本文算法中,首先计算每个边界框与真实框之间的 IOU;而后按照 IOU 从大到小对所有边界框进行排序,取其中前 2/3 数量的边界框作为训练样本;最后,调用 Focal loss 进行损失函数的计算和反向传递。

2 训练预处理

2.1 数据增广

在训练过程中,为了增加网络的鲁棒性,对输出的 SAR 图像进行增广。数据增广的方式有很多,比如水平翻转、裁剪、视角变换、jpeg 压缩、尺度变换、颜色变换和旋转等等。考虑到实际拍摄的角度不同

和 SAR 图像的单通道特性,这里我们选择采用水平翻转、尺度变化、旋转和亮度变化.在数据输入阶段,分别有 50% 的概率对图像进行随机类别的增广操作.

在进行水平翻转、尺度变化和旋转变换时,图像的标注框也要同时进行改变.在进行水平翻转和尺度变化时,标注框的类别信息不变,位置信息 (x, y, w, h) 进行对称变化和尺度变化.而在进行旋转变换时,标注框需要重新设定,如图 5 所示.首先通过旋转的角度,计算相应标注框的顶点旋转后的映射位置.而后,计算出每两个点之间的横向和纵向的最大距离,并以最大的横向距离和纵向距离作为新标注框的 (w, h) ,中心位置坐标不变.最后,选择适当的缩放尺度对边框的 (w, h) 进行修正.设 (w', h') 为新边框的大小, ε 为缩放因子, S 为原边框的面积,则

$$(w', h') = (w, h) \times \varepsilon \quad (5)$$

$$\varepsilon = \sqrt{2S/(w \times h)} \quad (6)$$

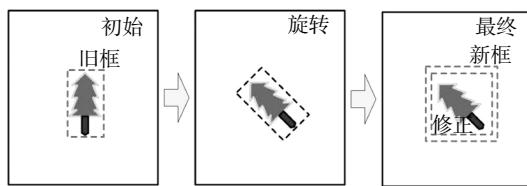


图 5 旋转变换时标注框的变化

Fig.5 Change in the label box

2.2 迁移学习

迁移学习是指针对新的神经网络重新使用预先训练的模型进行参数初始化.由于其能用较少的数据训练深度神经网络,能够提升网络的训练速度和精度,这使得它在深度学习领域非常流行.在本文中,首先使用 MSTAR 数据集在 ResNet101 上进行预训练,保存训练好的模型参数,而后将模型载入本算法的特征提取层,最后利用 SSDD 数据集对本网络的参数进行微调(Fine-tuning).

MSTAR 数据集是美国国防高等研究计划署(DARPA)支持的 MSTAR 计划所公布的实测 SAR 地面静止目标的公开数据集,包含多种车辆目标在各个方位角下获取到的目标图像.由于 SAR 图像和 RGB 图像之间存在较大差异,因此使用 MSTAR 数据集进行预训练可以避免出现负迁移的现象.将 MSTAR 数据集在 ResNet101 上进行训练,待网络完全收敛,去除 ResNet101 最后的全连接层,将前层的参数载入到本算法的特征提取层.在之后的 Fine-

tuning 过程中,不冻结特征提取层的参数,直接利用 SSDD 数据集对网络进行调整.

3 实验验证

实验在 Ubuntu16.04 系统的 Pytorch 框架下运行,并使用 CUDA8.0 和 cuDNN5.0 来加速训练.计算机搭载的 CPU 为 Corei7-8700k,显卡为 NVIDIA GTX1080Ti,内存为 32 G.数据集为 SSDD,共 1 160 张图像,2 456 艘舰船目标.本文将此数据集按照 7:3 的比例分为了训练集和测试集,在分类中,将图像按照复杂度(密集程度和尺度变化量)均匀地分给训练集和测试集以保证实验的有效性.非极大值抑制和预测概率的阈值均为 0.5.

将本算法与传统的 RetinaNet、SSD、FSSD、RFB 单阶段算法进行对比,采用平均精度和 FPS 作为性能指标,对比结果如表 3 所示(都经过预训练和数据增强).

表 3 算法性能对比表

Tab.3 Algorithm performance comparison table

方法	初始网络	平均精度/%	FPS
SSD300	Vgg16	84.7	83.30
	ResNet101	85.1	58.82
	DarkNet53	86.7	62.50
RFB300	Vgg16	85.0	58.82
	ResNet101	80.1	43.48
	DarkNet53	83.3	50.00
FSSD300	Vgg16	85.3	71.43
	ResNet101	85.6	55.56
	DarkNet53	85.8	62.50
RetinaNet321	ResNet101	88.5	58.82
本文算法	ResNet101	90.3	50.00

由表 3 可知,在精度上,经典的基于 Vgg16 的 SSD 网络精度相对较低,而其改进结构 RFB 和 FSSD 网络(基于 Vgg16)的检测精度虽然有所提高,但改进效果较小.当采用性能更好的特征提取层如 ResNet101 和 DarkNet53 时,SSD 网络的精度提升到了 86.7%,而 FSSD 网络提升较小,RFB 网络甚至出现精度下降.相对于 SSD 系列的算法,RetinaNet 在

精度上具有明显的优势. 而本文所提出的改进算法在 RetinaNet 的基础上仍有 1.8% 的提升, 取得了最高的检测精度. 在速度上, 可以看出特征提取层的影响较大, 其中基于 Vgg16 的网络速度较快, 基于 ResNet101 的速度较慢. 传统的基于 Vgg16 的 SSD 网络速度最快, 但是精度相对不足. 本文提出的算法在速度上达到 50 FPS, 相比原始 RetinaNet 仅仅下降 8.82 FPS (3 ms), 因此总体来看, 本文所提出的算法在此数据集中具有更好的效果.

为了证明改进算法的有效性, 分别将本文算法的改进部分(融合结构, 损失函数)与原始算法进行对比, 如表 4 所示.

表 4 改进部分的对比表

Tab.4 Improved part comparison

改进的融合结构	改进的损失函数	平均精度/%
		88.5
✓		90.0
	✓	88.1
✓	✓	90.3

由表 4 可以看出, 同时包含了这 2 种改进的算法精度达到最高; 只改进了损失函数算法的精度有所下降, 这是由于原始算法正负样本比例适中, 不会对 Focal loss 的计算带来退化所致. 而加入 OHEM 的思想, 网络强制性地去除了一些有利样本, 故精度有所下降. 当加入特征融合结合结构时, 可明显发现, 网络检测精度提高, 这是因为网络综合利用了底层的特征信息, 更高效地融合了深浅层的图像特征. 然而, 由于此时正负样本的不均衡程度增加, 损失函数存在部分退化, 此时, 加入 OHEM 的思想则对正负样本的均衡性进行调整, 再利用 Focal loss 进行损失计算和传递, 于是算法精度上升.

最后, 将改进的算法和原始的 RetinaNet 的检测结果可视化对比, 如图 6 所示. 通过对比 2 种算法和标注图, 可以看出, 在干扰因素较多、目标较小的时候, 传统的 RetinaNet 漏检、错检严重, 而本文算法都可以很好地识别出目标, 具有较好的鲁棒性.

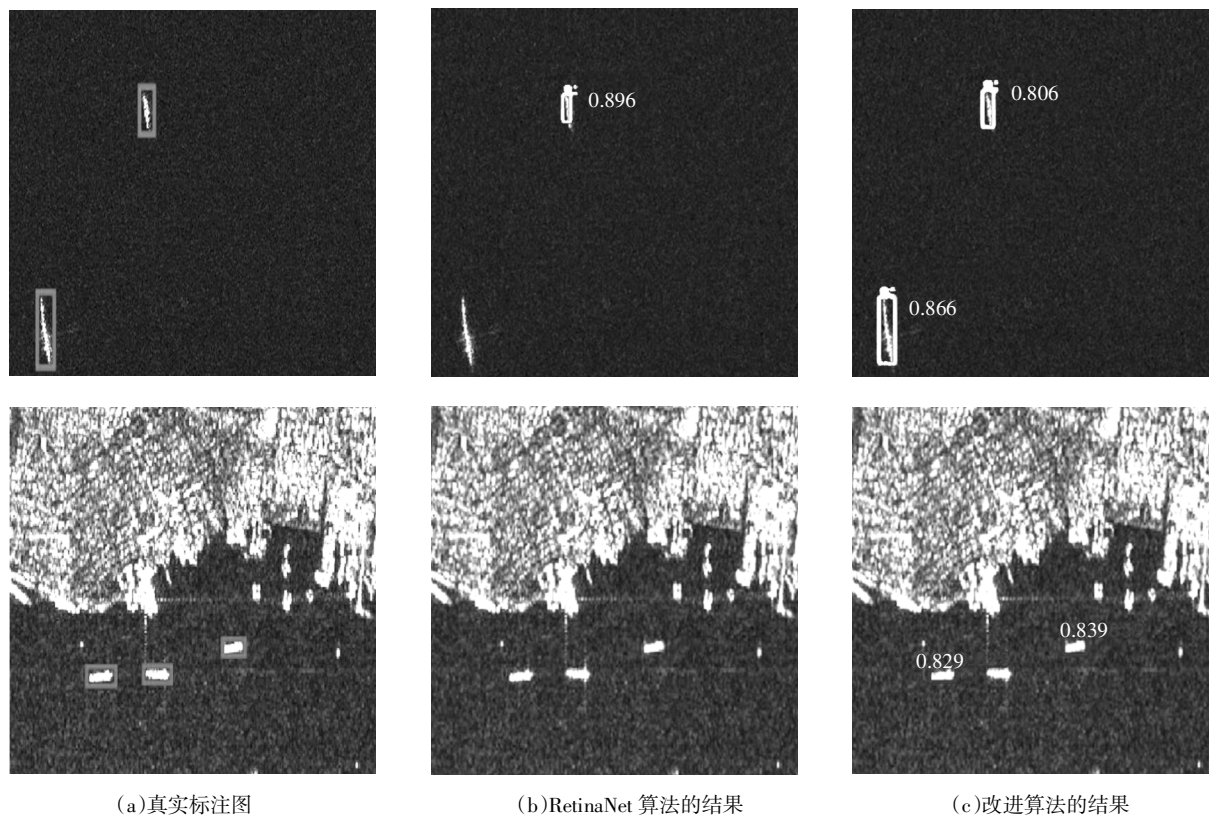


图 6 输出结果的可视化对比

Fig.6 Visual comparison of output

4 结论

针对传统的 SAR 图像舰船检测算法鲁棒性差、精度低、适应性不强等不足,本文提出了一种深度学习的 SAR 图像检测算法,该算法在 RetinaNet 的基础上针对 SAR 图像成像的特点对其 FPN 结构进行改进,增加了其特征数量,提升了融合效果。而后,针对数据量的增加出现的正负样本失衡问题,对损失函数的计算进行改进。最后,通过数据增强和迁移学习提升网络鲁棒性、收敛速度。实验中利用 SSDD 数据集对网络进行训练并对比了几个典型的单阶段检测算法,结果显示,本算法相比于传统的 RetinaNet 算法在 FPS 下降有限的情况下,准确率提升了 1.8%,具有更好的效果。

参考文献

- [1] HOLMBERG L, STEFÁNSSON V, GUNNERHED M. Internal conversion of the 208 keV E1 transition in 177 HF[J]. *Physica Scripta*, 1971, 4(1/2): 41—44.
- [2] 楚博策, 文义红, 陈金勇. 基于多特征融合的 SAR 图像舰船自主学习检测算法[J]. *无线电工程*, 2018, 48(2): 92—95.
CHU B C, WEN Y H, CHEN J Y. Self-learning detection algorithm for ship target in SAR images based on multi-feature fusion[J]. *Radio Engineering*, 2018, 48(2): 92—95. (In Chinese)
- [3] GIRSHICK R B, DONAHUE J, DARRELL T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, Los Alamitos: IEEE Computer Society Press, 2014: 580—587.
- [4] 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 & Machine Intelligence*, 2014, 37(9): 1904—1916.
- [5] GIRSHICK R B. Fast R-CNN [C]//*IEEE International Conference on Computer Vision*. Honolulu, Los Alamitos: IEEE Computer Society Press, 2015: 1440—1448.
- [6] REN S Q, HE K M, GIRSHICK R B, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(6): 1137—1149.
- [7] REDMON J, DIVVALA S, GIRSHICK R B, *et al.* You only look once: unified, real-time object detection [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2016: 1063—1069.
- [8] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2016: 6517—6525.
- [9] REDMON J, FARHADI A. YOLOv3: An incremental improvement [EB/OL]. [2018-04-08]. <https://arxiv.org/abs/1804.02767>.
- [10] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: Single shot multibox detector [C]//*Proceedings of European Conference on Computer Vision*. Heidelberg: Springer, 2016: 21—37.
- [11] LIN T Y, GOYAL P, GIRSHICK R B, *et al.* Focal loss for dense object detection [C]//*Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy, 2017: 2980—2988.
- [12] 李健伟, 曲长文, 彭书娟, 等. 基于卷积神经网络的 SAR 图像舰船目标检测 [J]. *系统工程与电子技术*, 2018, 40(9): 1953—1959.
LI J W, QU C W, PENG S J, *et al.* Ship detection in SAR images based on convolutional neural network [J]. *Systems Engineering and Electronics*, 2018, 40(9): 1953—1959. (In Chinese)
- [13] SHAHZAD M, MAURER M, FRAUNDORFER F, *et al.* Buildings detection in vhr sar images using fully convolution neural networks [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 57(2): 1—17.
- [14] 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*. Honolulu, Los Alamitos: IEEE Computer Society Press, 2016: 770—778.
- [15] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//*International Conference on International Conference on Machine Learning*. New York: ACM Press, 2015: 448—456.
- [16] LIN T Y, DOLLAR P, GIRSHICK R B, *et al.* Feature pyramid networks for object detection [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, Los Alamitos: IEEE Computer Society Press, 2017: 936—944.
- [17] SHRIVASTAVA A, GUPTA A, GIRSHICK R B. Training region-based object detectors with online hard example mining [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, Los Alamitos: IEEE Computer Society Press, 2016: 45—46.