基于YOLOv5的近岸SAR舰船目标检测方法

富　强[1],杨　威[1],陈　杰[1],郑世超[2],楚博策[1,3]

(1.北京航空航天大学 电子信息工程学院,北京 100191;2.上海卫星工程研究所,上海 200240;
3.中国电子科技集团公司 航天信息应用技术重点实验室,河北 石家庄 050081)

摘　要: 由于类似舰船的陆地目标的干扰和舰船的紧密排列,基于合成孔径雷达(SAR)图像的近岸舰船检测会出现较多漏检和误检,提出了一种基于YOLOv5网络的近岸舰船的检测方法。为提高近岸背景下的检测精度,使用了注意力机制模型和CSL的技术用于改进网络;分析了YOLOv5网络、注意力模型和CSL算法,基于YOLOv5进行了检测实验,引入注意力模型来改进网络;结合CSL算法,重构了YOLOv5旋转检测网络。通过调整训练参数和改进注意力,近岸目标检测网络的测试结果达到mAP 80%以上,证实了CSL+YOLOv5算法实现旋转检测的可行性。

关键词: 合成孔径雷达(SAR);舰船检测;卷积神经网络;YOLOv5网络;注意力机制模型;CSL算法

中图分类号: TN 183　　**文献标志码:** A　　**DOI:** 10.19328/j.cnki.2096-8655.2022.03.010

Detection Method for Nearshore SAR Ship Images Based on YOLOv5

FU Qiang[1], YANG Wei[1], CHEN Jie[1], ZHEN Shichao[2], CHU Boce[1,3]

(1.School of Electronic and Information Engineering, Beihang University, Beijing 100191, China; 2.Shanghai Institude of Satellite Engineering, Shanghai 200240, China; 3.Key Laboratory of Aerospace Information Applications, China Electronics Technology Group Corporation, Shijiazhuang 050081, Hebei, China)

Abstract: There are many missing and false alarms in the nearshore ship detection based on synthetic aperture radar (SAR) images, owing to the onshore ship-like objects and the close arrangement of ships. In view of this, in this paper, a detection method for nearshore ships is proposed based on YOLOv5. In order to improve the detection precision under nearshore backgrounds, the attention model and circle smooth label (CSL) algorithm are unified into the detection network. First, the YOLOv5 network, attention model, and CSL algorithm are analyzed. Second, the detection test is carried out based on YOLOv5, and the attention model is introduced to improve the network. Third, the YOLOv5 rotation detection network is reconstructed along with the CSL algorithm. Finally, tests are carried out to verify the reconstructed network. The results show that the detection network for nearshore targets could reach mAP above 80% by adjusting the training parameters and improving the attention, which confirms the feasibility of the CSL+YOLOv5 algorithm for rotation detection.

Key words: synthetic aperture radar(SAR); ship detection; convolutional neutral network; YOLOv5; attention model; circle smooth label algorithm

0　引言

合成孔径雷达(Synthetic Aperture Radar, SAR)是一种主动式微波成像传感器[1],它可以不受天气的影响,不分昼夜,同时可以获得更高质量、更高分辨率的大范围远距离图像。因此,利用SAR图像进行舰船检测已成为国内外研究的热点[1-2]。传统的基于SAR图像的舰船检测主要包括恒虚警率(Constant False Alarm Rate, CFAR)[3]算法和基于特征提取[4]的算法等。对于近岸舰船,要考虑舰船的密集停放,同时考虑陆地环境中大量的人工目

收稿日期: 2022-02-23; **修回日期:** 2022-03-22

作者简介: 富　强(1999—),男,本科,主要研究方向为遥感图像处理。

通信作者: 楚博策(1991—),男,硕士,工程师,主要研究方向为航天信息智能应用。

标,如建筑物、车辆等,其目标的特征与舰船相似,为准确的目标检测带来了很大的影响和挑战[5]。

近年来,随着深度学习的快速发展,这一技术得到了广泛的应用,并在图像识别领域取得了很大的进展,许多研究人员在舰船识别中也引入了深度学习。深度学习中的卷积神经网络(Convolutional Neutral Network,CNN)通过逐层学习和自主提取特征[6],可以充分利用图像中的信息,从而摆脱了传统方法的限制,在各种复杂多变的环境中具有良好的鲁棒性[4]。文献[7]设计了基于注意力机制特征重建网络的舰船目标检测算法,引入注意力旋转锚框以缓解密集排列目标引起的噪声问题,并利用特征重建模块来实现特征精炼。文献[8]通过网络轻量化和聚类算法改进 YOLOv3 网络,提升检测精度的同时加快了检测速度。

本文利用真实的开源 SAR 图像构建了一个近岸 SAR 舰船数据集,在此基础上进行了检测实验,精确评估了检测网络在近岸舰船上的性能。然后,引入注意力模块(Channel Attention Module, CAM)来提高网络的检测性能。在近岸测试集中,mAP 达到 80% 以上。此外,为了解决近岸地区舰船密集分布导致的检测性能下降的问题,引入了基于 YOLOv5 和 CSL 算法的旋转检测网络,并证明了其可行性。

1 YOLOv5 网络原理

YOLO[9]是一种基于深度学习神经网络的目标检测算法,属于一阶段检测网络,相比于以 SSD[10]为代表的双阶段检测网络,拥有运行速度很快的特点,适用于实时检测。YOLO 网络从 v1 开始不断改进更新[11-12],现在该系列最新的网络为 YOLOv5,在许多方面都有应用且达到了不错的效果。接下来是对 YOLOv5 的结构以及检测运用到的方法作介绍。YOLOv5 的整体网络结构如图 1 所示。

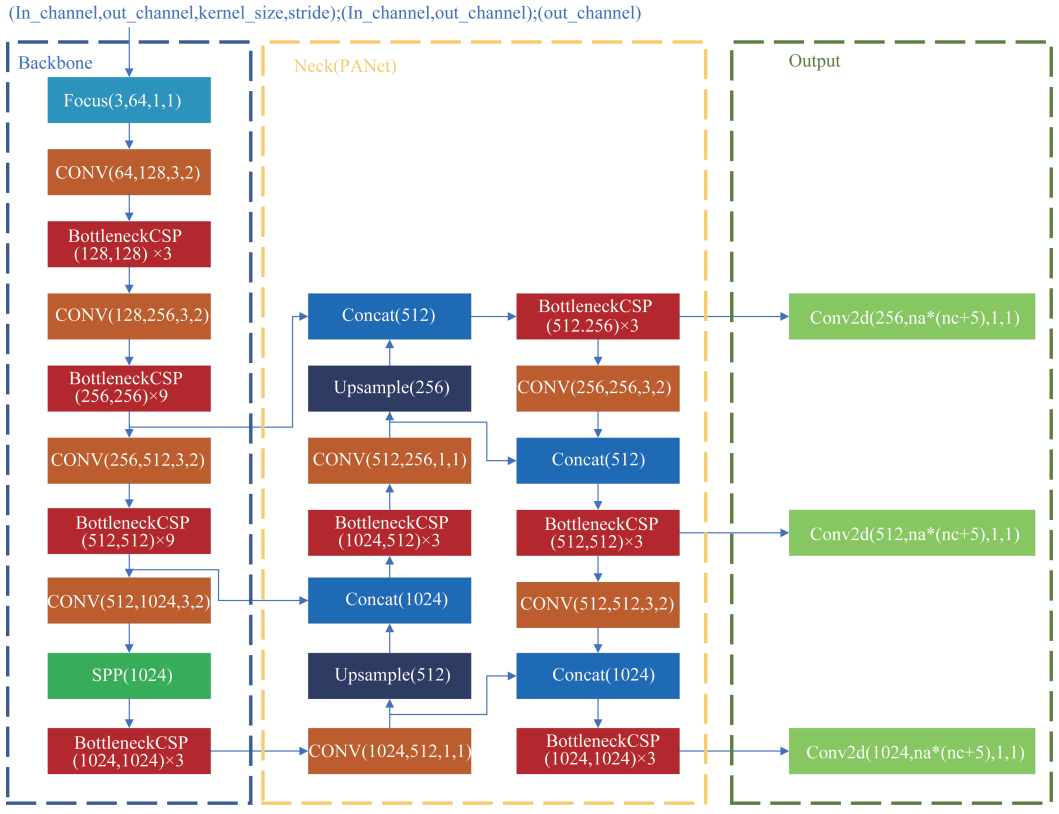


图 1 YOLOv5 网络结构[12]

Fig. 1 Network architecture of YOLOv5[12]

其中,骨干网络(Backbone)用以提取图像中的特征信息,颈部网络(Neck)用以整合不同尺度的特征信息并传输到输出层(Output),输出层用以生成预测结果。网络中的模块包含聚焦(Focus)模块、基础卷积(CONV)模块、跨阶段局部网络(CSP)模块、空间金字塔池化(SPP)模块、连接(Concat)模

块、上采样(Upsample)模块。相同的模块有不同的输入通道数(in_channel)、输出通道数(out_channel)、卷积核大小(kernel_size)、卷积步长(stride)。

1.1　CONV模块

CONV模块是YOLOv5中最基础的模块，由卷积模块、BN模块、Leaky ReLU激活函数组成。卷积模块是卷积神经网络的基础、Leaky ReLU函数，可以防止梯度饱和的问题，同时有效抑制梯度消失现象。

BN算法[13]引入了可学习参数γ、β，从而恢复出前一层网络所学习到的特征。BN的引入减少了初始参数的限制，可以选择较大的初始学习率，加快了训练速度；BN极大增强了网络泛化能力，减少过拟合，计算公式如下：

$$\begin{aligned} &\mathrm{BN}_{\gamma,\beta}(x_i)=\gamma_i\hat{x}_i+\beta_i \\ &\gamma_i=\sqrt{\mathrm{Var}(x_i)},\beta_i=E[x_i] \end{aligned} \tag{1}$$

式中：x_i为神经元输入；$E[x_i]$为各神经元输入值的平均值；$\mathrm{Var}(x_i)$为各神经元输入值的方差。

1.2　CSP结构

跨阶段局部网络[14](Cross Stage Partial Network，CSPNet)是减少目前需要的大量推理计算。以往在实际应用中能够获得较好结果的方法在很大程度上需要依赖庞大的计算量。CSPNet是一种网络设计思路，与残差结构[15]的思想相似，可以与已有的网络结合。实验中显示，随着CSPNet的引入，计算量有了大幅下降的同时，准确率没有下降，甚至有略微的提升。CSP模块结构如图2所示。

基础层的特征映射被CSP模块划分为2部分，即图2中左侧的单卷积部分和中间的深层网络部分，然后用跨阶段结构将合并这2部分。深层网络部分中的瓶颈模块(Bottleneck)为一个使用了残差结构的特征提取模块。

1.3　路径聚合网络结构

路径聚合网络(Path Aggregation Network，PANet，PANet)[17]提出的目的是在特征图金字塔网络[16](Feature Pyramid Network，FPN)的基础上进一步促进不同层之间特征信息的流动，其结构如图3所示。PANet在自顶往下的网络后，又增加了一个自底往上的金字塔结构。

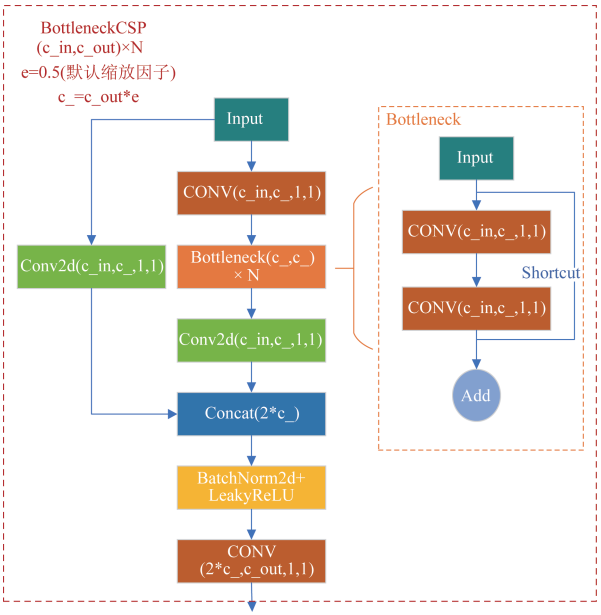


图2　CSP模块结构[12]

Fig. 2　Architecture of the CSP module[12]

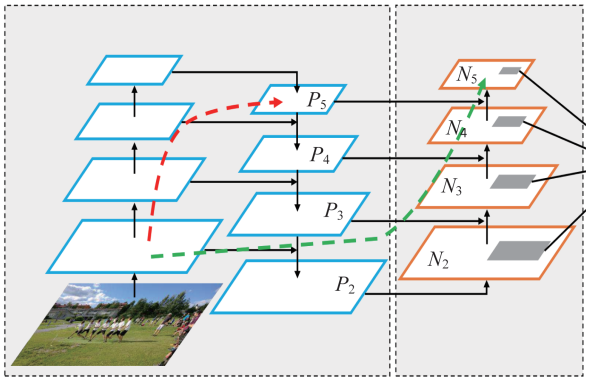


图3　PANet结构[17]

Fig. 3　Architecture of the PANet[17]

语义特征通过FPN层自顶往下传达，定位特征则通过PANet自底向上传达，不同的卷积层从不同的主干层进行融合，特征提取能力从而获得提高。

1.4　损失函数

目标检测最常用的损失函数是交并比(Intersection over Union，IoU)，计算公式如下：

$$I_{\mathrm{IoU}}=\frac{\mathrm{DR}\cap\mathrm{GT}}{\mathrm{DR}\cup\mathrm{GT}} \tag{2}$$

式中：DR(Detection Results)为检测结果；GT(Ground Truth)为实际结果。

但IoU无法反映两框距离的远近，因此提出了综合交并比(Generalized Intersection over Union，

GIoU)[18]。在GIoU基础上,距离交并比(Distance Intersection over Union, DIoU)[19]考虑了检测结果与真实值间的重叠率、距离以及尺度。更进一步地,完备交并比(Complete Intersection over Union, CIoU)考虑到了检测框的长宽比,在DIoU的公式中加入了一个影响因子,具体公式如下:

$$C_{\mathrm{CIoU}}=I_{\mathrm{IoU}}-\frac{\rho^2(b,b_{\mathrm{gt}})}{c^2}-\alpha\nu \tag{3}$$

式中:α为权重参数;ρ为2个点的欧氏距离;ν衡量检测值和真实值之间长宽比的相似性,定义如下:

$$\nu=\frac{4}{\pi^2}\left(\arctan\frac{w_{\mathrm{gt}}}{h_{\mathrm{gt}}}-\arctan\frac{w}{h}\right)^2 \tag{4}$$

式中:w为宽;h为高。

CIoU综合考虑了检测值和真实值之间的IoU、距离、长宽比等因素,包含的方面较为全面,使得检测值有较好的精度,因此YOLOv5中默认采用CIoU作为计算预测值损失的方法。

2 网络改进原理

2.1 注意力机制模型

计算机视觉的注意力机制的思想是通过寻找现有数据间的关联性突出某些重要特征,是一个在[0,1]区间连续分布的问题,属于软注意力。软注意力可以通过神经网络学习,经过前向传播和后向传播来得到注意力的权重。

压缩激发网络(Squeeze and Excitation Networks, SENet)[20]关注通道之间的关系,希望网络能够自发学习到特征图不同通道的重要程度,其结构如图4所示。

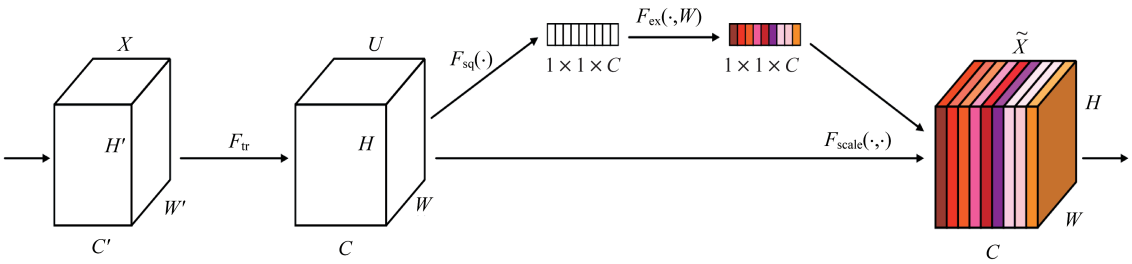
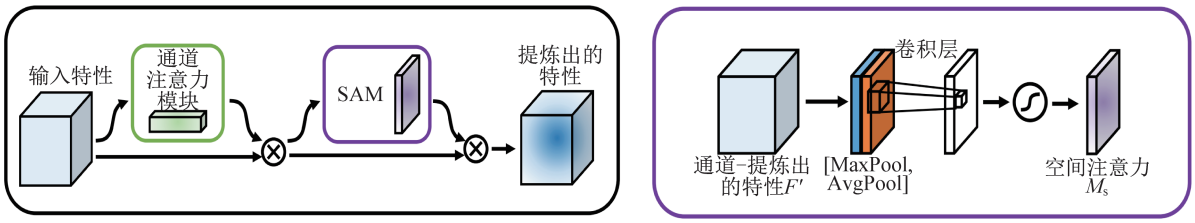


图4 SENet结构[20]

Fig. 4 Architecture of the SENet[20]

如图4所示,任意特征图X经过任意特征提取操作F_{tr}后获得特征图U,其中H、W、C分别为特征图的高度、宽度、通道数。SENet通过对特征图U进行压缩(Squeeze)和激发(Excitation)操作,即图中F_{sq}和F_{ex},得到1个一维向量,向量中的每个参数作为每个通道的评分,然后在通道上施加对应的权重,得到处理后的结果。其中压缩操作对每个通道做平均池化,得到各个通道的统计数值;激发操作学习通道的相关程度,并根据相关程度调整权重。

卷积块注意力模块(Convolutional Block Attention Module, CBAM)[21]结合了空间注意力和通道注意力,如图5(a)所示。输入特征图(Input Feature)经过通道CAM和空间注意力机制模块(Spatial Attention Module, SAM)获得提炼后特征图(Refined Feature)。通道CAM与SENet相似,不再赘述。图5(b)是SAM,和通道注意力类似,空间注意力可以学习特征图不同空间的重要程度。



(a) CBAM (b) SAM

图5 CBAM及SAM[21]

Fig. 5 Schematic diagram of the CBAM and SAM[21]

图 5(b)所示的 SAM 的具体操作如下：首先对于输入的特征图，即经过通道 CAM 处理的特征图(Channel-refined feature)，分别做基于通道的平均池化(AvgPool)、最大池化(MaxPool)，池化后的 2 个结果做基于通道的拼接操作；然后通过卷积(conv layer)，将通道数减少为 1；再通过激活函数，获得空间注意力特征图。将获得的空间注意力特征图施加在原来的特征图上得到最终的特征。与通道注意力类似，空间注意力可以学习特征图不同空间的重要程度。不同于只关注通道注意力机制的 SENet 模块，CBAM 模块往往可以取得更好的效果。

上述 2 种注意力机制模块都是轻量的通用模块，由于 2 种模块输入和输出的特征图大小是一致的，因此可以直接将模块添加用于 YOLOv5 网络的各层，将其融入到各种卷积神经网络中进行端到端的训练。本实验中，将 2 种模块分别添加与检测头前的位置。改进后的网络如图 6 所示。

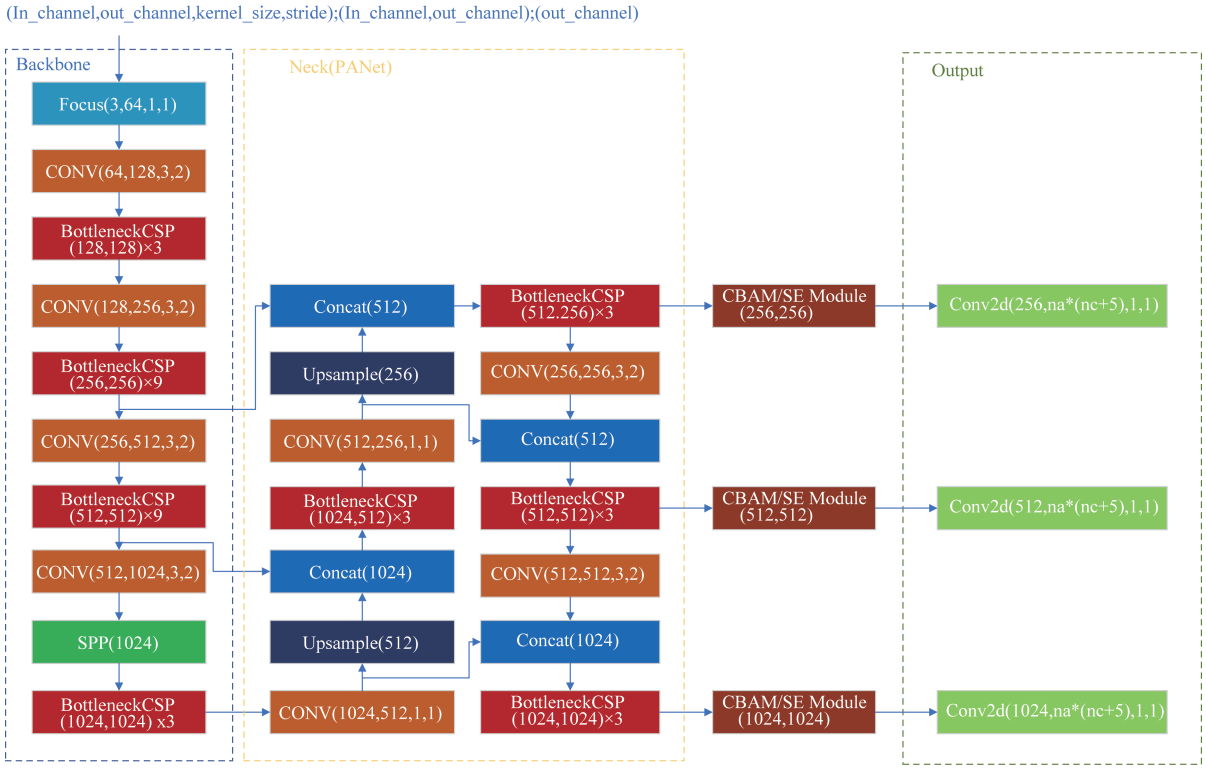


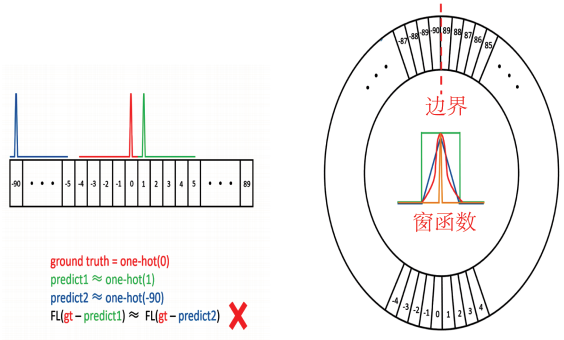
图 6　改进后 YOLOv5 结构

Fig. 6　Architecture of the improved YOLOv5

2.2　CSL 算法

为了处理边界问题，圆光滑标签(Circular Smooth Label，CSL)算法[22]将角度问题从原本的回归问题转化为分类问题，解决了角度的边界问题，可以很好地和长边定义法结合。CSL 将定义的角度进行划分，如图 7 所示。

仅仅把角度当作类别问题无法衡量预测角度和真实角度之间的距离，为了解决这个问题，CSL 设计了如下表达式：



(a) 独热标签　　(b) 圆光滑标签

图 7　角度的分类定义[22]

Fig. 7　Categorical definition of angles[22]

$$C_{\mathrm{CSL}}(x)=\begin{cases}g(x),\theta-r<x<\theta+r\\0,\quad \text{其他}\end{cases} \tag{5}$$

式中:$g(x)$为窗口函数;r为窗口半径。

满足如下几点:周期性,周期为$180/\omega$;对称性,关于θ对称;最大值为1;单调性,

$$0\leqslant g(\theta\pm\varepsilon)\leqslant g(\theta\pm\zeta)\leqslant 1,\ |\zeta|<|\varepsilon|<r \tag{6}$$

将角度问题转换为类别问题就是将一个连续问题转换为离散问题,这意味着转换过程中精度会有损失。CSL作者在文章中计算了损失以确定其对最终结果的影响,计算公式如下:

$$\mathrm{Max}(\mathrm{loss})=\omega/2 \tag{7}$$

$$E(\mathrm{loss})=\int_0^{\omega/2}x^{*}\frac{1}{\omega/2-0}\mathrm{d}x=\frac{\omega}{4} \tag{8}$$

式中:ω为相邻两类角度之间的差,比如将1°划分为一类时$\omega=1$,此时精度的最大损失、角度服从平均分布下的平均损失分别为0.5°和0.25°。这对于结果的影响很小,在实际应用中可以不予考虑。

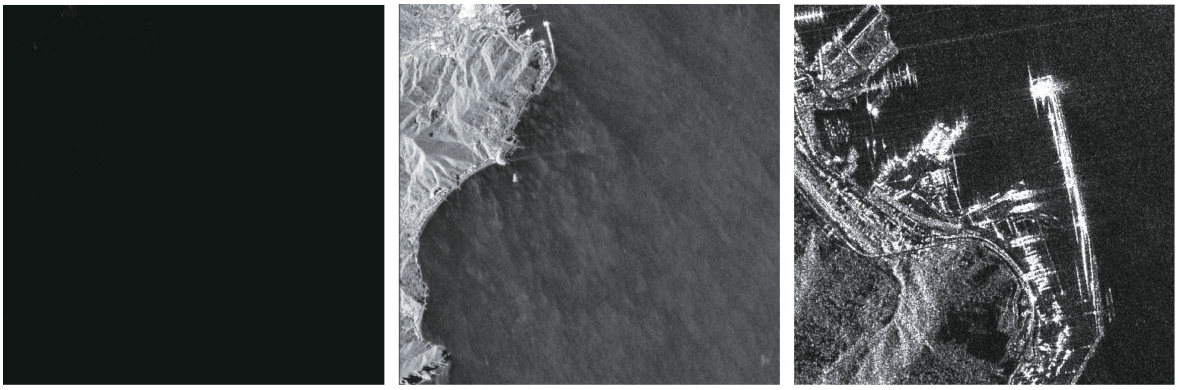
将该算法与YOLOv5网络相结合需对网络做如下改动:1) 数据输入部分,增加数据维度,在原有的检测框数据基础上增加角度数据信息;2) 网络检测部分,增加角度信息的预测;3) 损失函数部分,构建目标时增加角度信息列表,并在损失函数计算时增加角度信息损失的计算,计算函数与类别损失计算函数相同;4) 可视化部分,将绘制水平矩形框修改为绘制旋转矩形框。在YOLOv5网络的整体结构上并无改动。

3 检测实验与结果分析

3.1 数据集

本次实验计划使用SSDD[23]公开SAR舰船检测数据集。该数据集包含了多种不同的成像条件。图像主要来源于RadarSat-2、TerraSAR-X和Sentinel-1卫星传感器,一共4种极化模式,分辨率的范围为1~15 m。

通过位深度转换和对比度调整,将原数据调整为更适合用于网络训练和人工评估的图像。之后选取包含舰船目标的图像区域,按照所构建的神经网络要求,将图像进行切片为适合网络进行训练的图像尺寸,完成训练集的制作。原图和处理后图像如图8所示。



(a) 原图 (b) 对比度处理后 (c) 切片后

图8 处理前后数据对比

Fig. 8 Comparison of images before and after processing

数据集切片数量为1 386张。将1 386张图片分为两类,包含陆地的数据、不包含陆地的数据,其中包含陆地的切片数量为304张,不包含陆地的切片数量为1 082张。由于数据集数量较少,按照6:2:2的比例分别将上述两类数据分为训练集、验证集、测试集,并在测试集中去除不包含陆地的目标,着重分析网络对于近岸目标的检测性能。SSDD数据集分配见表1。

表1 SSDD数据集分配

Tab. 1 SSDD dataset distribution

数据集	样本数量/张		
	离岸	近岸	合计
训练集	649	182	831
验证集	216	61	277
测试集	0	61	61
合计	865	304	1 169

3.2　初步检测实验

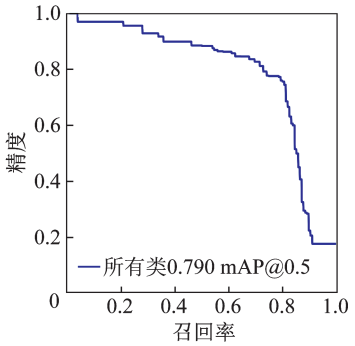
使用 YOLOv5 在 coco 数据集的预训练权重进行初步训练，处理批次大小设置为 8。处理批次过大会导致梯度下降速度缓慢，过小会导致损失函数难以收敛。在合理的范围内增加处理批次可以使显卡性能利用率提高，训练速度加快。迭代次数设置为 300，采用随机梯度下降作为梯度下降算法。初始学习率为 0.01，周期学习率为 0.2。YOLOv5 训练过程中，随着迭代次数增加，学习率减小，具体公式如下：

$$lr_n = lr_0\left(1-\frac{n}{m}(1-lrf)\right) \tag{9}$$

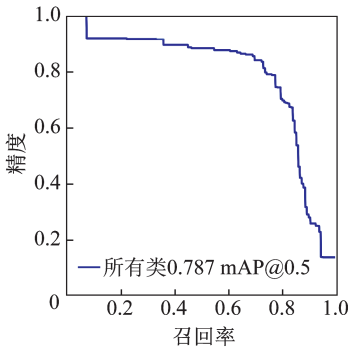
式中：m 为迭代总次数；n 为第 n 次迭代；lr_n 为第 n 次迭代的学习率；lr_0 为初始学习率；lrf 为周期学习率。最后一次迭代时，学习率为 $lr_0 \cdot lrf$。

均值平均精度（Mean Average Precision，mAP），用以衡量网络性能。计算时，置信度阈值设置为 0.001，NMS 置信度设置为 0.5；检测时，置信度阈值为 0.5，NMS 置信度为 0.5。测试结果如图 9 所示。

图 9 中，左侧为 200 次迭代训练后的检测结果，右图为 300 次迭代训练后的检测结果。然而两个结果之间差距不大，证明网络已经经过充分训练，虽然出现过拟合的现象，但下降幅度不大，过拟合不严重，证明了 YOLOv5 防止过拟合性能较为优秀。部分检测结果如图 10 所示。



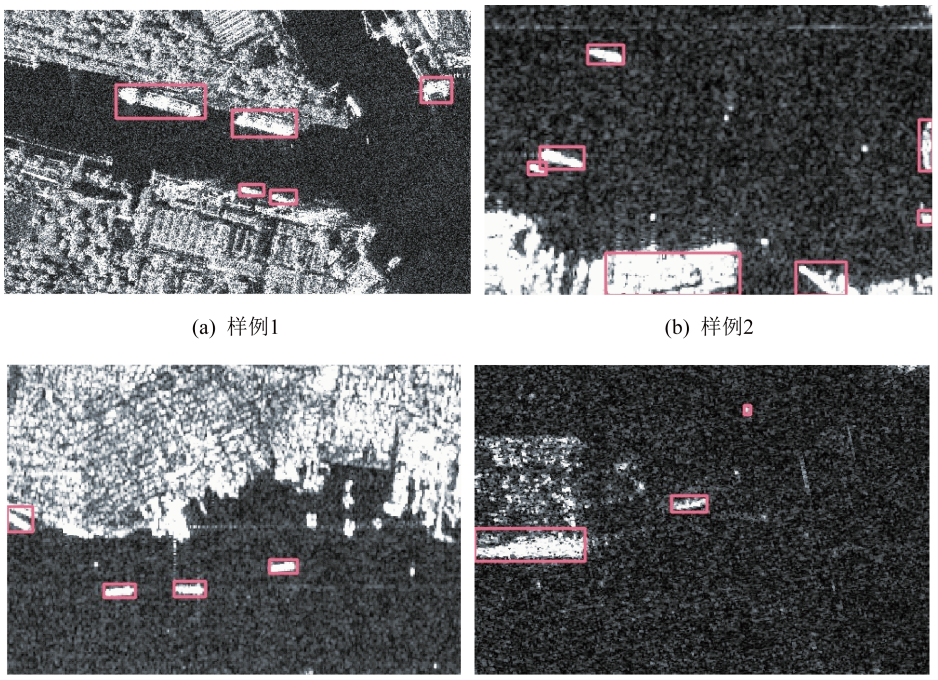
(a) 迭代200次



(b) 迭代300次

图 9　精度-召回率曲线

Fig. 9　Precision-recall curves



(a) 样例1　(b) 样例2

(c) 样例3　(d) 样例4

图 10　检测结果

Fig. 10　Detected results

3.3 注意力机制改进实验

为了进一步提高准确率,在YOLOv5网络结构中加入注意力机制模块,并重复实验。分别加入模块SELayer和CBAM,数量为1,位置为网络第8层,处理批次大小设置为8。采用随机梯度下降作为梯度下降算法,初始学习率为0.01,周期学习率为0.2,NMS阈值为0.5。由于增加了模块,模型参数增加,导致训练速度减慢,因此需要增加迭代次数,迭代次数设置为400。加入CAM后精度-召回率曲线如图11所示。

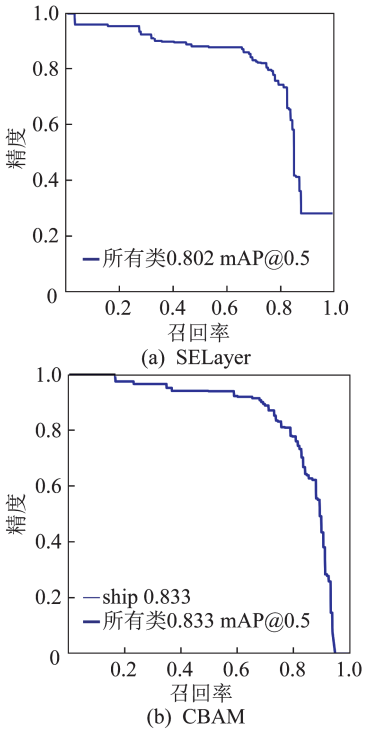


图11 加入CAM后精度-召回率曲线

Fig. 11 Precision-recall curves after adding CAM

图中可见:加入SELayer的网络,mAP值有小幅度提升,且准确率下降到一定程度后没有继续下降,证明SELayer模块在减少检测网络错检的效果上较为突出;加入CBAM模块的网络,mAP值有更明显的提升,证明了注意力机制模块在提升检测网络性能方面的作用,且CBAM较SELayer性能更好。

3.4 模型性能对比

YOLOv5、改进后YOLOv5与其他典型检测网络性能对比见表2,典型检测网络包括Faster R-CNN[24]和SSD[25]。

表2 检测结果对比

Tab. 2 Comparison of the detected results

网络	精度	召回率	mAP	检测用时/ms
Faster R-CNN	0.745	0.784	0.765	167.56
SSD	0.793	0.815	0.796	21.74
YOLOv5	0.778	0.801	0.790	11.62
YOLOv5_SE	0.798	0.826	0.802	14.51
YOLOv5_CBAM	0.821	0.853	0.833	15.61

由表2可知,原YOLOv5网络与典型检测相比,在准确率性能方面较为相近,但在检测速度方面有明显优势。增加CAM的改进后网络虽然增加了网络复杂度,因此增加了检测用时,但仍具有较高效率,同时在检测准确率方面也有提升。

3.5 结果分析与改进方案

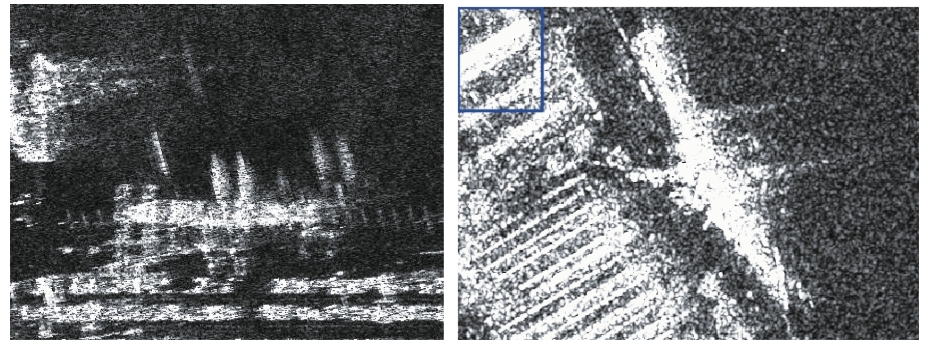
由以上收集的检测结果可以初步归纳检测错误原因,漏检的原因主要为检测目标附近的干扰、舰船的密集分布,其中干扰包含由于噪声导致的干扰和由于较为靠近陆地导致的陆地背景的干扰;错检的原因主要为将形状与舰船类似的目标错误判定为目标,将与舰船类似的目标命名为干扰目标,干扰目标可以分为离岸的干扰目标和陆地干扰目标。部分错检结果如图12所示。

针对以上问题提出如下解决方案:针对噪声导致的干扰,增加图像降噪的预处理环节;针对靠近陆地导致的陆地背景的干扰以及陆地的干扰目标,增加海陆分割的预处理环节;针对舰船密集分布的问题,加入旋转框检测网络。除了以上针对性措施之外,还可以通过增加训练集、提高网络的特征提取性能等方法,综合提升网络性能。

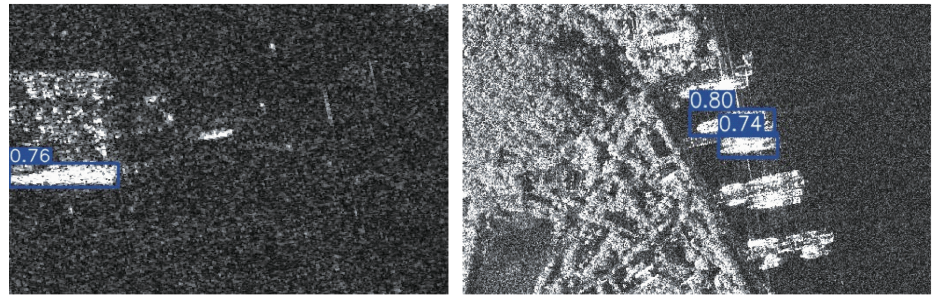
3.6 CSL旋转框检测实验

在以上实验中发现,错检漏检目标有部分原因来自于较为密集的目标,因此尝试实现旋转框的目标检测网络。计划使用CSL算法,实现YOLOv5旋转框检测网络改建。

训练时,处理批次大小设置为8,采用Adam算法作为梯度下降算法优化器。初始学习率为0.001,周期学习率为0.01,迭代次数为200次。最终loss的训练结果为box_loss=0.083 26,obj_loss=0.015 12,angle_loss=0.266,获得的部分结果如图13所示。



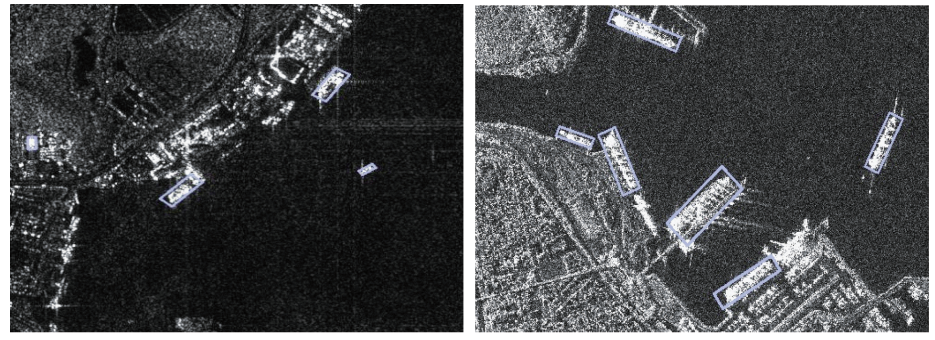
(a) 样例1　　(b) 样例2



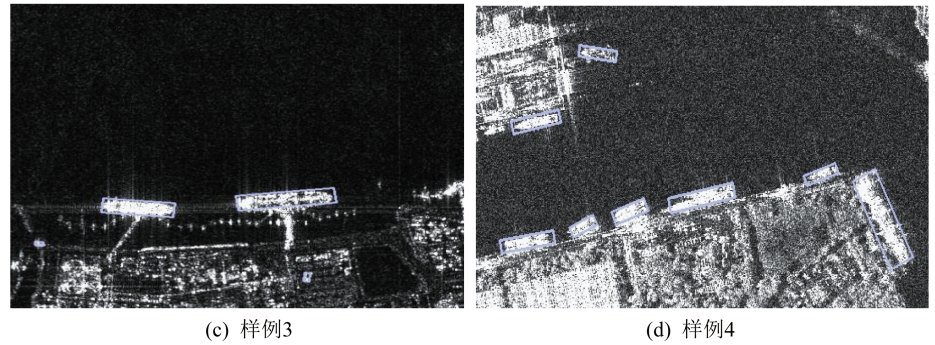
(c) 样例3　　(d) 样例4

图 12　部分错检结果

Fig. 12　Partial wrong detected results



(a) 样例1　　(b) 样例2



(c) 样例3　　(d) 样例4

图 13　部分旋转框检测结果

Fig. 13　Partial detected results of the rotating frame

由图 13 结果可知，在部分情况下旋转框可以较为精确地框定出目标范围，能够证实 CSL＋YOLOv5 算法进行旋转框检测的可行性。但是漏检、误检在大部分情况下较为严重，且目标的置信度普遍不高。判断原因主要在于网络未得到充分的训练，loss 在下降到一定程度后无法继续收敛，网络训

练陷入了局部最优。后续需要改变训练策略,进行进一步的训练,尝试将网络充分训练;或者对CSL算法进行改进,尝试寻找更优的旋转标签标注法和损失函数算法。

4 结束语

本文构建了着重于近岸目标的训练和测试集,开展以原YOLOv5网络为基础的近岸舰船目标检测网络实验和加入注意力机制模块的改进实验,并取得了较为理想的效果——在近岸测试集中,mAP达到了80%以上。证实了YOLOv5网络作为近岸舰船目标检测网络有着较为良好的性能,注意力机制模块在提升检测网路准确率方面也有着较好的效果。

针对近岸情况中舰船目标密集导致漏检的问题,开展旋转检测框实验,并且能够初步证明CSL+YOLOv5实现旋转框检测的可行性,但未能达到预期中提升检测性能的效果。分析原因后提出如下改进方案:尝试不同的学习率,梯度优化算法;修改原有的评估方法,得到适合旋转框目标的评估方法,指导训练网络的参数优化环节,从而实现网络的充分训练。或者对CSL算法进行改进,尝试寻找更优的旋转标签标注法和损失函数算法。

在总结错误原因后,除了旋转框检测之外,提出如下解决方案:针对噪声导致的干扰,增加图像降噪的预处理环节;针对靠近陆地导致的陆地背景的干扰以及陆地的干扰目标,增加海陆分割的预处理环节。除了以上针对性措施之外,还可以通过增加训练集、提高网络的特征提取性能等方法,综合提升网络性能。

参考文献

[1] 胡炎,单子力,高峰.基于Faster-RCNN和多分辨率SAR的海上舰船目标检测[J].无线电工程,2018(2):96-100.

[2] 蒋明哲.SAR图像舰船检测与分类方法研究[D].合肥:合肥工业大学,2016.

[3] 雷盼飞,苏清贺,杨桄.SAR图像舰船目标检测研究[J].影像技术,2011,23(4):40-45.

[4] 山鹏.SAR图像舰船目标检测及特征提取方法研究[D].哈尔滨:哈尔滨工程大学,2012.

[5] 王昌安.遥感影像中的近岸舰船目标检测和细粒度识别方法研究[D].武汉:华中科技大学,2019.

[6] 田东.基于卷积神经网络的深度学习算法研究与实现[D].上海:上海交通大学,2017.

[7] 牛戈,陈小前,季明江,等.基于注意力机制特征重建网络的舰船目标检测[J].上海航天,2021,38(4):128-136.

[8] 方青云,王兆魁.基于改进YOLOv3网络的遥感目标快速检测方法[J].上海航天,2019,36(5):21-27,34.

[9] 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. 2016: 779-788.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]// European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.

[11] REDMON J, FARHADI A. Yolo9000: better, faster, stronger[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2017: 7157-7163.

[12] REDMON J, FARHADI A. Yolov3: an incremental improvement [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7263-7271.

[13] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]// International Conference on Machine Learning. 2015: 448-456.

[14] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 390-391.

[15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[16] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2117-2125.

[17] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8759-8768.

[18] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 658-666.

(下转第90页)