基于改进YOLOv3的SAR舰船图像目标识别技术

姜浩风,张 顺,梅少辉
(西北工业大学 电子信息学院,陕西 西安 710072)

摘 要: 合成孔径雷达(SAR)图像自动目标识别(ATR)技术是人工图像解译的关键技术之一。针对传统的SAR舰船目标检测算法大多受限于场景且泛化能力较差的问题,设计了一种基于改进YOLOv3网络的检测模型。将YOLOv3与DenseNet网络融合,使用稠密网络模块代替用于提取中小尺度特征的残差网络模块,通过训练得到模型的最优权重,实现端到端的目标检测。使用综合交并比(GIoU)损失代替交并比(IoU)边界框回归损失,提供更加准确的边界框位置信息,提高检测精度,采用中国科学院空天信息研究院制作的SAR图像船舶检测数据集进行测试。测试结果表明:与原YOLOv3算法相比,改进后的YOLOv3检测准确率提高了1.4%。

关键词: 合成孔径雷达(SAR)舰船图像; 目标检测; YOLOv3; DenseNet; 多尺度先验框; 综合交并比(GIoU)
中图分类号: TP 751 **文献标志码:** A **DOI:** 10.19328/j.cnki.2096-8655.2022.03.009

Target Detection Technology for SAR Ship Images Based on Improved YOLOv3

JIANG Haofeng, ZHANG Shun, MEI Shaohui
(School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China)

Abstract: Automatic target recognition (ATR) of synthetic aperture radar (SAR) images is one of the key techniques for artificial image interpretation. However, traditional SAR ship target detection algorithms are mostly limited to the scene and poor generalization ability. In view of this, a detection model based on the improved YOLOv3 is designed. The YOLOv3 is fused with the DenseNet. The dense network block is used to replace the residual block so as to improve the feature extraction ability, and the optimal weight of the model is obtained through training so as to achieve end-to-end target detection. The GIoU loss is used to replace the IoU loss so as to provide more accurate boundary box location information and improve the detection accuracy. The detection dataset of SAR ship images produced by Aerospace Information Research Institute, Chinese Academy of Sciences are used for testing. The test results show that, compared with the original YOLOv3 algorithm, the detection accuracy of the improved YOLOv3 increases by 1.4%.

Key words: synthetic aperture radar (SAR) ship image; target detection; YOLOv3; DenseNet; multi-scale anchor box; GIoU

0 引言

舰船是水上交通运输的重要载具,对水面舰船的监测与管理是各国海事部门的一项重要工作。合成孔径雷达(Synthetic Aperture Radar, SAR)技术是一种主动式微波遥感探测技术,它利用脉冲压缩和合成孔径同时提高雷达距离向和方位向的分辨率,从而获得全天候、全天时、大面积、高分辨率SAR图像[1]。SAR图像自动目标识别(Automatic Target Recognition, ATR)技术旨在自动从SAR图像中判断出有无目标信息,提高SAR图像的解读效率与准确度。SAR系

收稿日期: 2022-03-04; **修回日期:** 2022-03-22
基金项目: 国家自然科学基金(62171381)
作者简介: 姜浩风(1998—),男,硕士研究生,主要研究方向为遥感图像处理和机器学习。
通信作者: 梅少辉(1984—),男,博士,副教授,主要研究方向为光电智能探测、遥感图像处理和模式识别与机器学习。

统可安装于飞机、卫星等平台，并且可全天候、全天时观察地面和海洋表面，因此SAR图像ATR技术对海洋监测领域的水面舰船检测与识别具有重要价值。

近年来，SAR图像ATR技术已经在全世界得到深入研究，并形成了固定的三级流程：检测、鉴别和分类[2]。检测模块主要基于检测算法获取包含SAR图像目标的切片，鉴别模块对于目标切片剔除虚警值，分类模块选取最佳决策机制判断类别。

传统的SAR图像舰船检测算法为基于统计建模的恒虚警率（Constant False Alarm Rate，CFAR）算法，包括双参数CFAR算法、基于K分布的CFAR算法[3]，以及最佳熵自动门限法、多极化方法等[4]。此类方法均需要对原始图像进行预处理，如杂波过滤、海陆分割，在检测过程中也需要根据先验知识对部分虚警目标进行过滤[5]。目前，随着人工智能领域的快速发展，卷积神经网络（Convolutional Neural Network，CNN）可以实现对图像高层特征的主动提取，避免了人工选取特征的复杂工作，具有良好的分类准确度和鲁棒性，为SAR图像目标检测提供了新的途径[6-7]。

在自然图像的分类识别任务上，研究人员提出了一些非常典型的深度学习网络模型，例如AlexNet[3]、VGG[4]、Googlenet[5]、ResNet[6]、DenseNet[7]、Inception[8]、MobileNet[9]、SqueezeNet[10]等。目前成熟的目标检测模型大多基于这一系列网络架构，例如two-stage的R-CNN[11]、Fast R-CNN[12]、Faster R-CNN[13]检测算法与one-stage的YOLO[14]系列检测算法。对于SAR图像的舰船检测，深度学习在检测速度与检测精度上都优于传统的检测方法，YOLOv3[15]相比其他目标检测模型，在保证准确率的同时进一步提高了检测效率。本文在YOLOv3网络的基础上，用稠密网络模块代替用于提取中小尺度特征的残差网络模块，改进网络对于较小尺度SAR舰船目标的检测性能，使用综合交并比（GIoU）度量损失代替交并比（IoU）边界框回归损失，提升边界框的检测精度。

1　网络模型及算法

1.1　YOLOv3网络

YOLOv3网络是一种单阶段目标检测网络，与两阶段目标检测网络相比，单阶段目标检测网络具有更快的检测速度。YOLOv3网络采用Darknet 53作为主干网络以提取输入图像的特征，同时结合了多种优秀的方法如多尺度检测、残差网络等，进一步提升了网络性能。YOLOv3首先将输入图像调整至416×416的大小，再将其分割成$S\times S$的网格，网格中每个单元格负责检测中心点落在该单元格中的目标。每个单元格会检测3个不同尺度的边界框并预测其置信度。预测量包括x、y、w、h与置信度，其中x、y为边界框中心坐标在X和Y方向的相对偏移量，w和h为边界框的宽和高，置信度通过阈值对预测结果进行取舍，表示边界框中预测目标类别的准确性。

1.2　DenseNet网络

密集连接网络（DenseNet）由HUANG等[16]于2016年提出，该网络并未参考ResNet的思想，通过残差级联实现增加网络层数，从而提升网络性能；也未参考Inception的思想，通过扩展网络宽度提升网络性能，而是着眼于特征方面，将特征在多个通道上以不同的方式进行连接实现特征重组，从而保证其特征信息的全面性，提高特征复用率。

通过对特征的重组利用不仅在一定程度上减轻了梯度弥散问题，而且在减少了参数数量的同时加强了特征的有效传递。对于一个m层的网络结构，DenseNet网络将0层至$(m-1)$层的输出进行非线性变换：

$$x_m = H_m([x_0, x_1, \cdots, x_{m-1}]) \tag{1}$$

式中：$H_m(\cdot)$为非线性变换函数；$[x_0, x_1, \cdots, x_{m-1}]$为将0层～$(m-1)$层输出的特征进行拼接；$x_m$为第$m$层的变换结果。

DenseNet以这种稠密连接的方式最大化地使用特征信息。因此本文采用DenseNet网络中的稠密模块来加强网络对于特征的提取。

1.3　YOLOv3与DenseNet网络的融合策略

ResNet网络与DenseNet网络都是应用于分类任务中，两者的框架结构极为相似，其网络都是由若干个单元模块堆叠而成，且单元模块都是由激活函数层、卷积层与批归一化层组成。因此，在原有结构的基础上，将残差模块（Residual Block，Res Block）替代为稠密模块（Dense Block），并改变其连接方式即可完成替代工作，如图1所示。其中Conv2D代表卷积模块，Contact代表全连接模块，Upsampling代表上采样模块。将YOLOv3网络中

对尺度 26×26 和 52×52 进行预测输入的两组残差模块替换为自定义的密集连接模块,构建出一个带有密集型连接模块的特征提取网络,使尺度 26×26 和 52×52 在进行预测之前能够接收密集连接块输出的多层卷积特征,增强特征的传递并促进特征复用和融合,进一步提升检测效果。

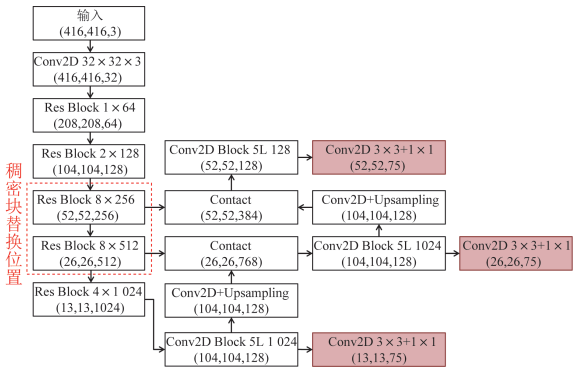
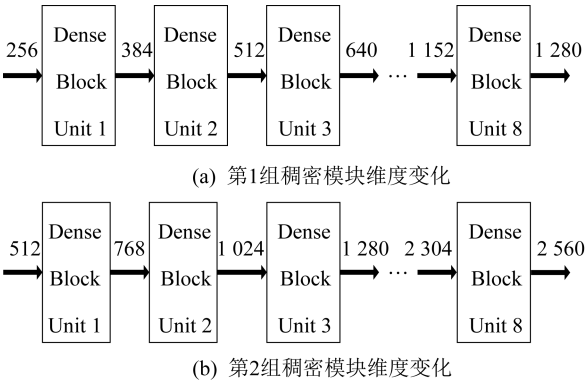


图 1　YOLOv3 结构图及稠密块位置

Fig. 1　Diagram of the YOLOv3 architecture and the dense block position

稠密模块中通道维数的变化过程如图 2 所示,在第 1 组 Dense Block 中共有 8 个 Dense Block Unit,分别代替对应位置的 Residual Block Unit。Dense Block 的输入为 52×52×256 的特征图,为了减少计算负担,在后 7 个 Dense Block Unit 中的输入前分别执行 1×1 的卷积操作实现降维,即每次增加的通道数 K 为 128:

$$X_L = H_L([X_0, X_1, \cdots, X_{L-1}])$$
$$L \in \{1, 2, \cdots, 8\} \tag{2}$$



(a) 第1组稠密模块维度变化

(b) 第2组稠密模块维度变化

图 2　稠密模块维度变化

Fig. 2　Dimension change of the dense block

式中:$H_L(\cdot)$ 为 BN+ReLU+Conv(1×1)+BN+ReLU+Conv(3×3)的组合函数表达,即附加 Bottleneck layer 的 Dense Block 的非线性组合函数;$X_0, X_1, \cdots, X_L$ 为每次线性组合操作后的结果。输入特征图经 8 组非线性函数组合获取到 52×52×1 280 特征图,随后,将其与特征交互部分中上采样获取的 52×52×128 特征图进行拼接操作,得到 52×52×1 408 的特征图,以此作为输出应用于检测小尺度的目标。

第 2 组 Dense Block 的输入通道数为 512,通道变化做出相同的降维预处理,即每次增加的通道数 K 为 256,经过 8 组 Dense Block Unit 后输出 26×26×2 560 的特征图,并将其与特征交互部分经上采样获取的 26×26×256 特图执行拼接操作,经通道合并后输出 26×26×2 816 特征图用于中型尺度目标检测。改进之后的模型 YOLOv3 网络结构如图 3 所示。

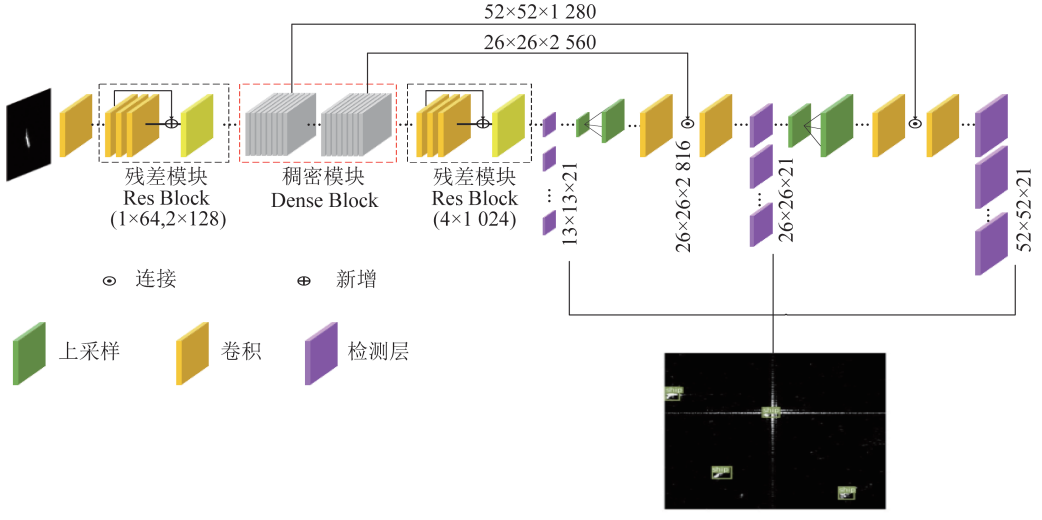


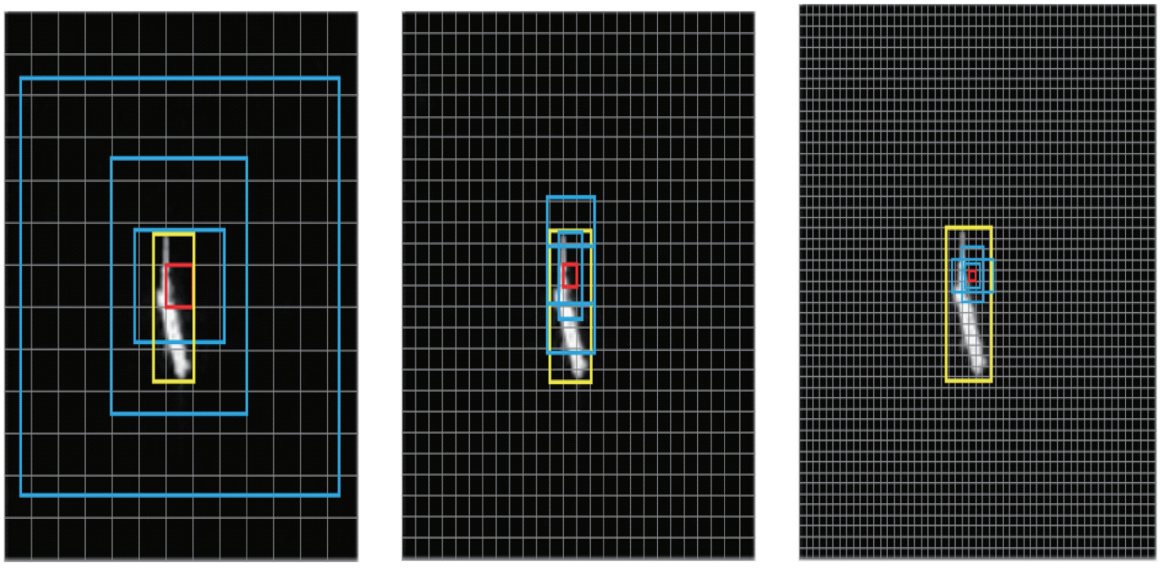
图 3　改进后的模型结构

Fig. 3　Structure of the improved model

1.4　多尺度先验框

随着输出特征图数量和尺度的变化，先验框的尺寸也需要进行相应调整。YOLOv3 延续了 YOLOv2[17] 的方法，采用 *K*-means[18] 聚类得到先验框的尺寸，为每种下采样尺度设定 3 种先验框，共聚类出 9 种尺寸的先验框：(10×13)、(16×30)、(33×23)、(30×61)、(62×45)、(59×119)、(116×90)、(156×198)、(373×326)。

在检测中，不同大小的目标分配不同尺度的先验框，最小的 13×13 特征图具有最大的感受野，应用较大的先验框(116×90)、(156×198)、(373×326)，适合检测较大的目标。中等的 26×26 特征图具有中等感受野，应用中等的先验框(30×61)、(62×45)、(59×119)，适合检测中等大小的目标。较大的 52×52 特征图具有较小的感受野，应用较小的先验框(10×13)、(16×30)、(33×23)，适合检测较小的目标。不同尺度下的特征图与先验框如图 4 所示。



(a) 13×13特征图　　(b) 26×26特征图　　(c) 52×52特征图

图 4　不同尺度下的特征图与先验框

Fig. 4　Characteristic diagrams and a priori frames at different scales

在目标检测的过程中，如果目标中心落在某网格中，即由该网格负责预测。YOLOv3 的输出结果通过预测不同网格的锚点所对应的偏移量完成检测目标预测框的回归，如图 5 所示。

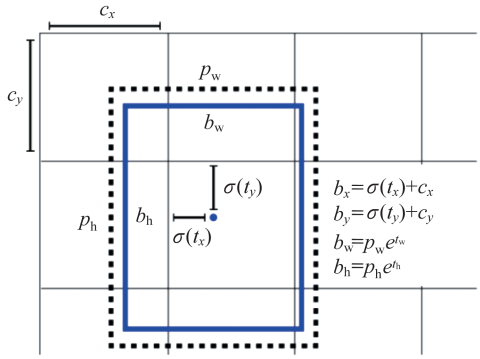


图 5　边框预测

Fig. 5　Frame prediction

图中：c_x、c_y 为负责预测网格的左上角坐标；t_x、t_y 为被测目标中心相对网格左上坐标的偏移量；p_w、p_h 为预设的先验框映射到特征图中的宽和高；b_x、b_y 为最终得到的边框中心坐标；b_w、b_h 为其相对特征图的宽与高。$\sigma(\cdot)$sigmoid 函数将，压缩到[0，1]区间内，确保目标中心处于执行预测的网格单元中，防止偏移过多；t_w、t_h 为尺度缩放的参数，经过指数运算还原后参与 b_w、b_h 的运算。

1.5　GIoU 改进

YOLOv3 模型使用 IoU 即交并比作为衡量检测定位性能的主要指标，IoU 计算公式如下：

$$I_{\mathrm{IoU}}=\frac{|A\cap B|}{|A\cup B|} \tag{3}$$

式中：A、B 分别为预测框与真实框的位置信息。

损失函数定义为

$$L_{\mathrm{IoU}}=1-I_{\mathrm{IoU}} \tag{4}$$

IoU作为指标存在2个问题:1) 若两框没有相交,即$I_{\mathrm{IoU}}=0$,此时IoU无法反映预测框与真实框之间的距离,同时损失函数无法进行梯度的反向传播,故无法通过梯度下降的方式进行训练。2) IoU仅能量化地表示两框之间的重合度大小,无法表示两框的空间位置关系。

针对以上2个问题,HAMID等[19]提出了使用GIoU作为新的指标,计算公式如下:

$$G_{\mathrm{GIoU}}=I_{\mathrm{IoU}}-\frac{\left|C-(A\cup B)\right|}{\left|C\right|} \tag{5}$$

式中:C为包含A与B的最小同类形状,其损失函数为

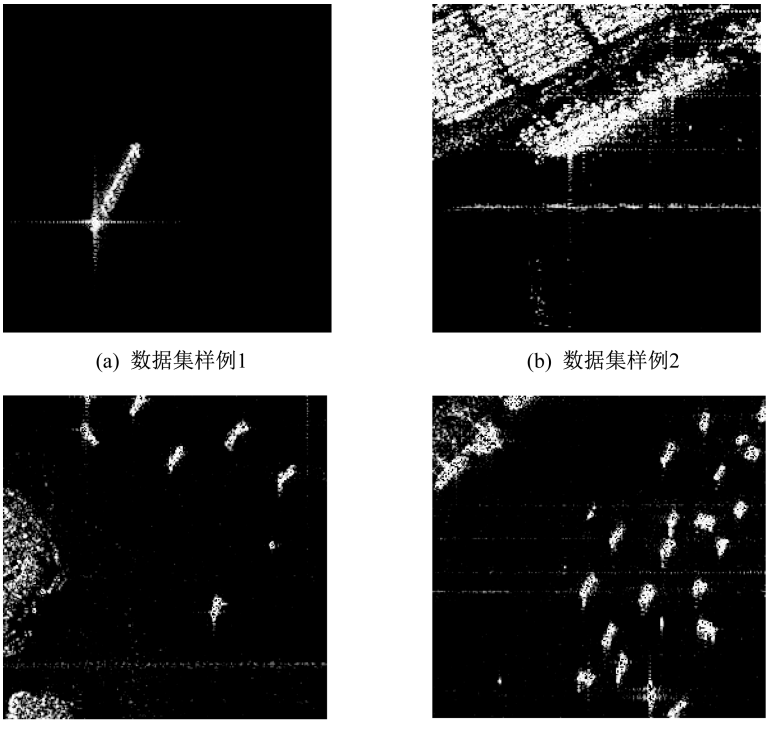
$$L_{\mathrm{GIoU}}=1-G_{\mathrm{GIoU}} \tag{6}$$

GIoU继承了IoU的尺度不变性,同时相较与IoU仅关注重合区域,GIoU还关注了非重合区域,可以更好地反映预测框与真实框的重合程度。

2 实验过程及结果

2.1 数据集介绍

本实验采用中国科学院空天信息研究院数字地球重点实验室研究员王超团队公开的SAR图像船舶检测数据集[20]。该数据集包括多源、多模式SAR图像,以我国国产高分三号SAR数据和Sentinel-1 SAR数据为主数据源,数据及其标注格式适用于目标检测任务。目前,该深度学习样本库包含43 819船舶切片及其标注信息,数据集样例如图6所示。



(a) 数据集样例1 (b) 数据集样例2

(c) 数据集样例3 (d) 数据集样例4

图6 SAR舰船数据集样例

Fig. 6 Examples of the SAR ship dataset

2.2 模型训练

本实验的训练环境为:Ubuntu16.04系统,python3.7,pytorch1.7.1框架,cuda10.1,cudnn8.0,GPU为GTX1080Ti。

本实验采用监督学习的方式,通过损失函数计算网络输出值与期望值之间的误差,通过误差反向传播调节网络内部的各项参数来减小误差,使网络逐渐收敛。在训练过程中选择SGD优化器,权重衰减(Weight Decay)设置为0.000 5,冲量(Momentum)设置为0.9,学习率设置为0.001,训练50 000批次后,损失函数趋近于0,表明模型收敛,对训练数据的拟合程度较好。损失函数曲线如图7所示。

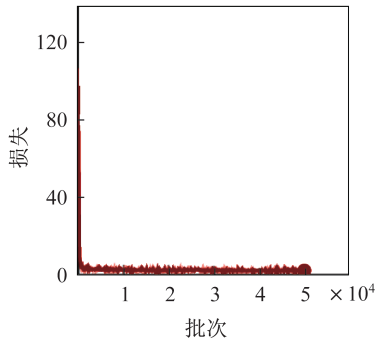


图 7　网络训练损失曲线

Fig. 7　Network training loss curve

2.3　实验结果

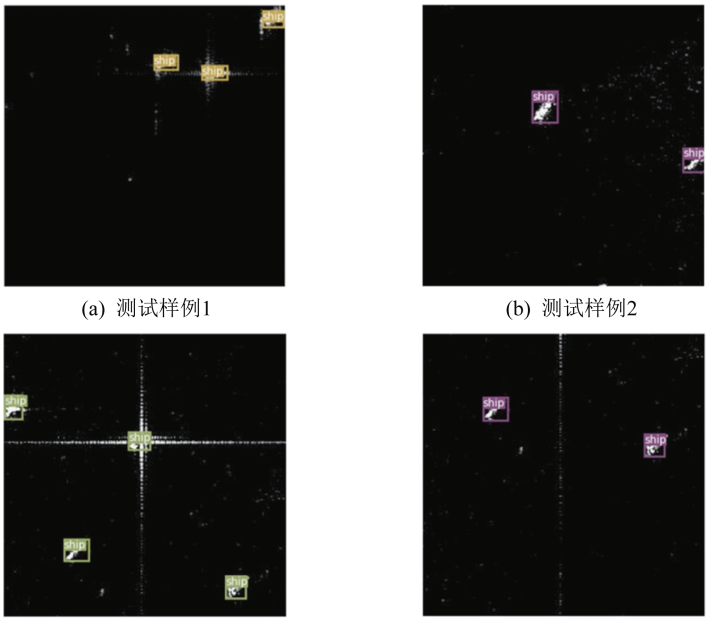
本实验采用的评价指标为主流的类平均准确率（Mean Average Precision，mAP）与每秒传输帧数（Frames Per Second，FPS）。mAP 是所有检测目标类别的平均值（Average Precision，AP），衡量模型在各类别上识别效果的平均水平。FPS 用于衡量模型的检测速率。本实验采用 Faster R-CNN 与原版 YOLOv3 作为对比算法，实验中训练样本与验证样本的比例为 9∶1，测试结果见表 1。

表 1　测试结果对比

Tab. 1　Comparison of the test results

模型	mAP/%	FPS
Faster R-CNN	84.63	11
YOLOv3	86.35	27
改进 YOLOv3	87.76	27

由表 1 可知，改进后的 YOLOv3 框架的检测性能在准确率上优于 Faster R-CNN 与原版 YOLOv3，相较原版 YOLOv3 模型，改进后模型的 mAP 提高了 1.4%，在检测速度上与原版 YOLOv3 一致，这是因为改进的方法并未对 YOLOv3 做模型结构与参数量上的简化。改进后的 YOLOv3 采用稠密模块替换残差模块来提高对小目标的检测精度。由于海面目标多为小型舰船，因此改进后的 YOLOv3 比原版具有更高的检测精度，检测结果图像如图 8 所示。



(a) 测试样例1　(b) 测试样例2

(c) 测试样例3　(d) 测试样例4

图 8　SAR 舰船检测结果

Fig. 8　Results of the SAR ship detection

3　结束语

本实验针对 SAR 图像自动目标识别问题展开了研究，设计了一种基于改进 YOLOv3 架构的 SAR 舰船检测模型。该模型通过将网络中部分残差模块替换为稠密连接模块，改进了网络对于小尺寸目标的检测性能，通过用 GIoU 取代 IoU 提高预测框的检索准确度。

实验结果表明，改进后的 YOLOv3 框架基于其对小目标检测的优秀性能，在 SAR 舰船检测的任务上取得了较好的结果。

参考文献

[1] 薛笑荣.SAR图像处理技术研究[M].北京:科学技术文献出版社,2017:13.

[2] 张强.SAR图像自动目标识别技术分析[J].信息技术,2018(12):157-160.

[3] 邢相薇.SAR图像舰船目标检测方法研究[D].长沙:国防科学技术大学,2009.

[4] 唐沐恩,林挺强,文贡坚.遥感图像中舰船检测方法综述[J].计算机应用研究,2011,28(1):29-36.

[5] 周舟,王海鹏,徐丰,等.基于通道剪枝的SAR图像舰船检测优化算法[J].上海航天,2020,37(4):48-54.

[6] CHEN S, WANG H, XU F, et al. Target classification using the deep convolutional networks for SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8): 1-12.

[7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]// Proceedings of the14th European Conference on Computer Vision (ECCV). 2016: 21-37.

[8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]// Advances in Neural Information Processing Systems. London: MIT Press, 2012: 1097-1105.

[9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [DB/OL]. arXiv: 1409.1556, 2014 [2022-01-15]. https://arxiv.org/pdf/1409.1556.pdf.

[10] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Press, 2016: 2818-2826.

[11] 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. Las Vegas: IEEE Press, 2016: 770-778.

[12] HUANG G, LIU Z, LAURENS V D N, et al. Densely connected convolutional networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE Press, 2017: 4700-4708.

[13] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// IEEE conference on computer vision and pattern recognition. Piscataway: IEEE Press, 2015: 1-9.

[14] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [DB/OL]. arXiv: 1704.04861, 2017 [2022-01-15]. https://arxiv. org/pdf/1704.04861.pdf.

[15] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size[DB/OL]. arXiv: 1602.07360, 2016 [2022-01-15]. https://arxiv. org/pdf/1602.07360.pdf.

[16] 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, Columbus: IEEE Press, 2017: 580-587.

[17] GIRSHICK R. Fast R-CNN [C]//IEEE International Conference on Computer Vision. Santiago, Chile: IEEE Press, 2015:1440-1448.

[18] REN S, HE K, GIRSHICK R, 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.

[19] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [DB/OL]. arXiv: 1506.02640, 2015 [2022-01-15]. https://arxiv.org/pdf/1506.02640.pdf.

[20] REDMON J, FARHADI A. YOLOv3: An incremental improvement [DB/OL]. arXiv: 1804.02767, 2018 [2022-01-15]. https://arxiv. org/pdf/1804.02767.pdf.

[21] HUANG G, LIU Z, LAURENS V, et al. Densely connected convolutional networks [C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 2261-2269.

[22] LLOYD S P. Least square quantization in PCM[J]. IEEE Transactions on Information Theory, 1982, 28(2):129-136.

[23] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]// IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE Press, 2017: 6517-6525.

[24] REZATOFIGHI H, TSOI N, GWAK JY, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]// IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA: IEEE Press, 2019.

[25] WANG Y, WANG C, ZHANG H. Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 SAR images[J]. Remote Sensing Letters, 2018, 9(8): 780-788.