基于深度学习的红外舰船目标识别

杨 涛，戴 军，吴钟建，金代中，周国家
（西南技术物理研究所，四川 成都 610041）

摘要：本文采用深度学习技术中的 YOLOv3（You Only Look Once Version 3）目标识别算法对红外成像仪从海面采集的红外图像中舰船进行识别。红外成像仪采集图像的频率高达 50 帧/s，为了能减少网络计算时间，本文借鉴 YOLOv3 的一些思想，采用全卷积结构和 LeakReLU 激活函数重新设计一个轻量化的基础网络，以此加快检测速度。输出层根据采集回来的红外图像的特点采用 Softmax 算法回归，在提高检测速度的同时，也兼顾了检测精度。

关键词：红外图像；目标识别；深度学习；YOLOv3
中图分类号：TN957.52，TP18 **文献标识码：**A **文章编号：**1001-8891(2020)05-0426-08

Target Recognition of Infrared Ship Based on Deep Learning

YANG Tao，DAI Jun，WU Zhongjian，JIN Daizhong，ZHOU Guojia
(*Institute of Southwest Technical Physics*, *Chengdu* 610041, *China*)

Abstract：In this study, the You Only Look Once Version 3 (YOLOv3) target recognition algorithm in deep learning technology is used to identify the ship in an infrared image collected using an infrared imager from the sea surface. The infrared imager captures images at a frequency of up to 50 frames per second. To reduce network computing time, a few ideas are generated based on YOLOv3; additionally, a full convolution structure and the LeakReLU activation function are used to redesign a lightweight basic network to accelerate detection. The output layer uses the softmax algorithm to regress according to the characteristics of the collected infrared images, which improves the detection speed and accounts for detection accuracy.

Key words：infrared image, target recognition, deep learning, YOLOv3

0 引言

近年来我国军事实力取得了巨大的进步，尤其是从辽宁舰成功服役后，我国的海军力量上了一个新的台阶，在现在的海上战争中，要有效地削弱敌方海军的战斗力，精确识别对方舰船并采用精确制导技术摧毁对方的舰船是现代战争常用的方式，但精确识别敌方舰船是一个难点，我们需要高效的目标识别算法才能准确地锁定对方的舰船。不仅如此，我国海岸线长达 18000 多公里，不仅是防止敌人入侵的重点地带，也是偷渡、走私、贩毒等违法犯罪的高发场所，要在如此之长的海岸线上重点监测海岸附近的海面船只也是一个难点，为了维护海洋资源的可持续发展，我国在 1995 年制定了“伏休制度”，在此期间，对海面船只的监控也是一件非常棘手的事。为了解决这些问题，我国在一些海岸线上安装了监控摄像头，但早期的摄像头都是可见光成像，只能在白天才能有效工作。随着红外成像技术的发展，红外成像由于其成像距离远、不受白天和夜间限制等优点得到了广泛的应用，但是红外成像只能得到灰度图像，再加上海天线、鱼鳞波、船只等目标航向姿势状态等干扰因素的影响，要在红外图像上高效地识别出舰船是行业一大挑战。

为了解决这个难题，学者们进行了大量的研究，许多优秀的目标识别算法被学者发明并得到了广泛的应用，在这些目标识别算法中，大致流程可分为图像预处理、图像特征提取、使用分类器对特征进行分类。传统的目标识别算法中，特征提取方法基本都是通过手工的方式设置参数提取特征，针对不同的图像检测任务，往往要设计许多不同的特征提取方法，如为了使检测系统适应图像尺度、旋转的变化，尺度不变特征变换（scale-invariant feature transform，SIFT）[1]，方

收稿日期：2019-06-18；**修订日期：**2019-07-22.
作者简介：杨涛（1992-），男，硕士研究生，主要从事目标检测、深度学习方面的研究。E-mail：304778654@qq.com。
通信作者：吴钟建（1967-），硕士，副研究员，硕士导师，主要从事目标跟踪、目标检测等方面的研究。E-mail：wjz209@126.com。

向梯度直方图变换（Histogram of oriented gradient, HOG）[2]被学者提出并得以广泛的应用。然而，这些传统的算法往往是在某些特定的场景下能获得很好的效果，但换一种场景可能就表现的很差。

随着近几年来神经网络的飞速发展，深度学习算法在计算机视觉领域得到广泛的应用，特别是目标识别中，大量的算法被提了出来，从 R-CNN[3]到 Faster-RCNN[4-7]等一系列基于提取候选区域（region of interests）的算法到 YOLO[8]（you only look once）、SSD[9-10]（single shot multibox detector）等端到端的检测算法大量涌现。其中 YOLOv3[11-12]以其速度快，准确率高得到了大部分人的青睐，在 VOC2012 数据集上，YOLOv3 在网络输入为 416×416，IOU 阈值设置为 0.5 时，检测平均精度值（mean average precision，mAP）达到了 57.9，超越了传统的目标识别算法。本次任务是要检测海面上过往的船只，目标成像达到 5×10 以上分辨率的目标。由于目标像素分辨率较小且经常存在一张图上有多个小目标的情形，YOLOv3 对小目标有良好的表现效果且计算开销相对较小。但是 YOLOv3 计算开销比较大，在嵌入式平台上实现实时检测成本太高，所以本次实验选择了 YOLOv3 作为基本算法进行改进。重新对 YOLOv3 提取特征的基础网络进行了设计，通过减少网络层数，减少卷积核的大小，借鉴 YOLOv3 的思想，采用全卷积和 LeakReLu 激活函数，增强网络拟合能力并减少计算开销，实现高效的舰船识别。

1 YOLOv3 检测的原理

1.1 YOLOv3（You Look Only Once Version3）

YOLOv3 把一张图经过一系列卷积或池化计算后，输出 3 个不同感受野 $S\times S\times[C+(B\times 5)]$的特征图。$S$ 为输出特征图的长和宽，C 为预测类别数量，表示网格内是属于哪一类物体的置信度，B 为每个网格最多预测目标的个数。5 表示网格预测每个目标的位置信息和位置的置信度的参数个数：每个目标的信息应该包含目标位置的中心坐标和目标尺寸信息，这里用(x,y)（w、h）来表示，这 4 个参数可以得到一个框，得到目标所在的区域。在 YOLOv3 中，目标中心位置(x,y)是相对于当前网格左上角顶点的偏移值，而（w、h）则是相对于整幅图像宽度和高度，都被归一化到(0,1)区间。得到 3 个特征图后，再把低分辨率的特征图采样到高分辨率的特征图上，这相当于把图片划分为了 $S\times S$ 个网格，如果目标物体的中心落入了某个网格，则这个网格就负责检测该物体。较高分辨率的特征图用来检测较小分辨率的目标，较小分辨率的特征图主要负责检测较大分辨率的目标，这大大提高了目标检测的准确率，每个网格还要预测目标位置信息的置信度（Confidence），即预测的框的准确率通过公式(1)计算得到，之后通过图像交并比（intersection-over-union, IOU）和非极大值抑制（non-maximum suppression, NMS）来排除重复的目标。IOU 为预测区域与物体真实区域（ground truth）的交集与两者并集之比。Confidence 和 IOU 的计算公式如下：

$$\text{Confidence}=P(\text{Object})\times\text{IOU} \tag{1}$$

$$\text{IOU}=\frac{\text{Area}_{\text{pred}}\cap\text{Area}_{\text{truth}}}{\text{Area}_{\text{pred}}\cup\text{Area}_{\text{truth}}} \tag{2}$$

式中：P(Object)为预测边界框包含目标物体的概率，如果目标中心落在边界框之内，则置信度为 1，相反，如果网格中不存在目标，则 P(Object)＝0；$\text{Area}_{\text{truth}}$为基于训练样本标注的目标真实区域的面积；$\text{Area}_{\text{pred}}$为预测的目标物体的边界框的面积。目标位置的置信度表示为预测目标与真实目标的 IOU（检测框和真实框的重叠程度）值乘以目标概率，最后再对输出的目标框使用非最大值抑制得到最后的结果，YOLOv3 实现了端到端的检测。YOLOv3 算法示意图如图 1，YOLOv3 的网络结构如图 2。

YOLOv3 的检测速度在 TitanX 上图像输入分辨率为 416×416 时检测帧率为 35 fps，在日常生活应用中，达到了实时性要求，但在普通计算芯片上，远远达不到这个速度。在检测精度方面，以 mAP 为评价参数，YOLOv3 达到了 55.3%，已经远远地高于传统的算法。

1.2 改进 YOLOv3 的红外舰船检测

YOLOv3 在目标检测中表现出了非常好的性能，但是，其网络结构比较复杂，卷积层数很深，这样可以使网络拟合能力大大增加，从而可以进行多种目标的检测，原作者是在 COCO 数据集上设计的基础网络，总共需要检测 91 种目标，计算开销特别大，特别是在普通 CPU 上无法实现实时性，但是在本次实验中，原则上只需要检测舰船这一种目标，但是为了减少海天线和鱼鳞波带来的干扰，我们实际检测了 3 种目标，即舰船、海天线、以及鱼鳞波。由于检测目标种类数变少了，所以网络的拟合能力可以不需要那么强，本文将对 YOLOv3 网络结构进行重新设计，使得在满足检测精度的同时，减少计算开销。在 YOLOv3 网络中，作者放弃了池化采样计算，采用全卷积计算，在对特征图下采样时采用了步长为 2 的卷积计算，变成了全卷积网络。本文拟借鉴原作者的思想，用全卷积网络作为特征提取层，为了减小计算开

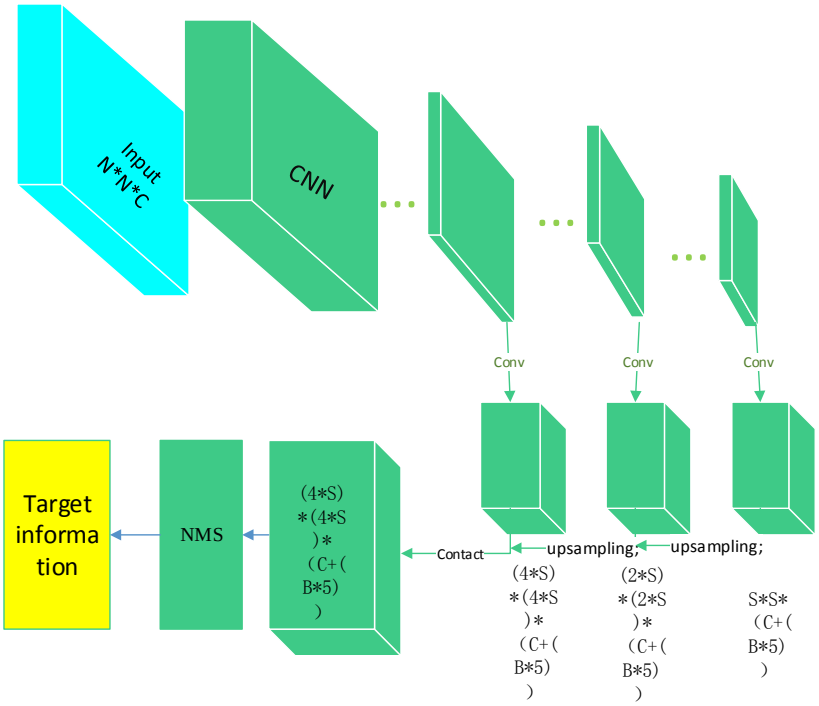

图 1 YOLOv3 算法示意图 Fig.1 YOLOv3 algorithm flow chart

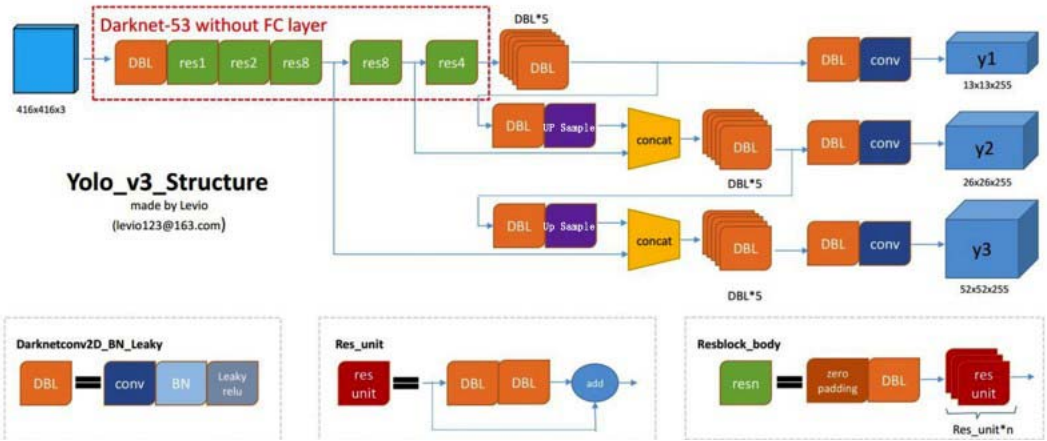

图 2 YOLOv3 网络结构图（特征提取网络为 Darknet-53） Fig.2 YOLOv3 network structure(Darknet-53)

销，采用小卷积核，在 13×13，26×26，52×52 三种尺寸的特征图上做预测。对采集回来的数据作分析，红外舰船图像背景相对比较简单，除了海面，天空外，基本只有海天线和鱼鳞波的干扰。在很多情形下，一张图片上存在多个舰船目标的情况，目标大小相差很大而且还存在舰船之间相互遮挡的问题。针对本次数据的特性，在数据集中特意增加了海天线和鱼鳞波类别，增强了该网络在鱼鳞波和海天线干扰下的鲁棒性，减少了虚警率。改进网络包括以下几点：

1）全网络激活函数采用 LeakReLU 激活函数，其数学计算公式如下：

$$y_i = \begin{cases} x_i & x_i \geq 0 \\ \dfrac{x_i}{a_i} & x_i < 0 \end{cases} \tag{3}$$

2）采用全卷积网络

本文借鉴了 YOLOv3 的原理，网络结构采用了全卷积和多尺度的思想，下采样采用控制卷积步长的方法来实现，增加了网络的拟合能力。

3）重新设计一个卷积网络，使其在保证精度的情况下减少计算开销，提高检测速度。整个网络的结构如图 3 所示：其中 Con2_BatNor_LR 包含 3 个连续的算法，卷积（Convolutional）、Batch Normalization，Leak ReLU。

①卷积计算

对图像进行卷积计算，提取图像特征，其中 fileter 为卷积核的个数，Size 为卷积核的尺寸，Stride 为卷积计算式的步长，如图 3 所示，本次采用 1×1，2×2，3×3 三种不同大小的卷积核，这 3 种卷积核尺寸都比

较小，相比与 5×5 或 7×7 的卷积核，可以大大减小计算量。

②Batch Normalization

在训练神经网络时，使用标准化输入可以提高训练的速度。方法是对训练数据集进行归一化的操作，把原来的数据减去其均值后，再除以其方差。这种标准化输入只是对输入进行了处理，中间的隐藏层没有什么变化。在神经网络中，第 l 层隐藏层的输入就是第 $l-1$ 层隐藏层的输出，如果对 l 层的输出进行标准化处理，从原理上来说可以提高 W_l 和 b_l 参数的训练速度和准确度。这种对各隐藏层的标准化处理就是 Batch Normalization。其计算公式如下：z_i 表示 l 层的输出的第 i 个参数，ε是为了防止分母为零，可设一个非常小的值，如 10^{-10}。Batch Normalization 计算由以下 3 步完成：

1）计算输出 l 层输出的均值μ，如式(4)：

$$\mu=\frac{1}{m}\sum_{i=1}^{m}z^i \tag{4}$$

2）计算 l 层输出的方差$\sigma^{2\prime}$，如式(5)：

$$\delta^2=\frac{1}{m}\sum_{i=1}^{m}(z^i-u)^2 \tag{5}$$

3）归一化输出，如式(6)：

$$z^i_{\text{norm}}=\frac{z^i-u}{\sqrt{\delta^2+\varepsilon}} \tag{6}$$

式中：m 表示 l 层输出参数的个数。

③Leaky Relu 函数

在早期的神经网络中，激活函数多用 ReLu 函数，ReLu 函数是将所有的负值都设为零，然而 Leaky ReLu 函数是给所有负值赋予一个非零斜率。LeakyReLU 函数在 ReLu 的基础上把小于零的数给一个很小的乘积因子，在近两年新提出的网络结构中，LeakyRelu 被广泛使用。其计算表达式如下：

$$y_i=\begin{cases}x_i & x_i\geq 0\\ \dfrac{x_i}{a_i} & x_i<0\end{cases} \tag{7}$$

式中：a_i 是属于(1，＋∞)上的常数。在输出的处理中如图 3 所示，依然拼接融合了 3 个不同尺度的特征层做最后的特征输出。

2　检测结果及性能分析

2.1　仿真平台介绍

此仿真平台操作系统为 64 位的 Windows 7，采用了 Tensorflow 深度学习框架搭建的改进 YOLOv3 网络，硬件方面 CPU 为 Intel Pentium G2030，GPU 采用了 GeForce GTX 1050。本次实验中红外舰船的图片数据共有 1330 张图，其中 1200 张为训练集，130 张为验证集。本次采用的损失函数为 YOLOv3 原有的损失函数，图 4 为训练过程中的损失下降过程，图 5 为验证集损失曲线图，可以看到数据迭代到 3000 次左右时模型趋于稳定，验证方式为每训练 10 次验证一次。

Type	Filter	Size/Stride	Output	Type	Filter	Size	Output
Con2_BatNor_LR	16	(3*3)	416*416				
Con2_BatNor_LR	16	(3*3)/(2*2)	208*208				
Con2_BatNor_LR	32	(2*2)	208*208				
Con2_BatNor_LR	32	(2*2)	208*208				
Con2_BatNor_LR	32	(2*2)/(2*2)	104*104				
Con2_BatNor_LR	64	(2*2)	104*104	Con2_BatNor_LR	256	(3*3)	52*52
Con2_BatNor_LR	64	(3*3)	104*104	Con2_BatNor_LR	256	(2*2)	52*52
Con2_BatNor_LR	128	(2*2)/(2*2)	52*52	Con2_BatNor_LR	75	(1*1)	52*52
Con2_BatNor_LR	256	(2*2)	52*52	YOLO			
Con2_BatNor_LR	256	(3*3)	52*52				
Con2_BatNor_LR	256	(2*2)/(2*2)	26*26	Con2_BatNor_LR	256	(2*2)	26*26
Con2_BatNor_LR	512	(3*3)	26*26	Con2_BatNor_LR	75	(1*1)	26*26
Con2_BatNor_LR	1024	(2*2)/(2*2)	13*13	YOLO			
Con2_BatNor_LR	256	(2*2)	13*13				
Con2_BatNor_LR	512	(2*2)	13*13				
Con2_BatNor_LR	75	(1*1)	13*13	YOLO			

图 3　改进 YOLOv3 的网络结构图　　Fig.3　Modified YOLOv3 network structure

注：Type 为计算流程，Con2_BatNor_LR 前面有介绍，Size/Stride 为卷积核的大小和卷积计算的步长，卷积步长为空是为默认值 1，Output 是卷积输出的 Feature Map 的尺寸大小。

Note: Type is the computational process. It has been introduced before. Size/Stride is the kernel size of the convolution and step. Default value is 1 if the convolution step is empty. Output is the Feature Map's size.

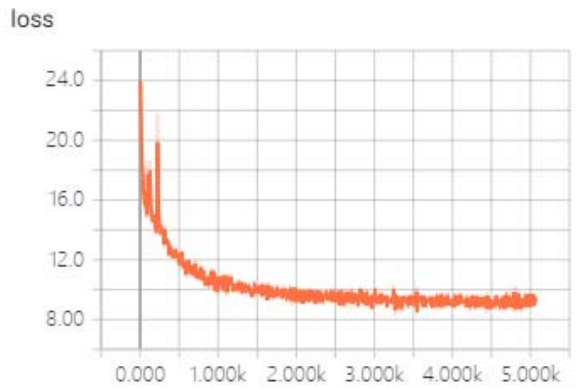

图 4 训练损失下降曲线图 Fig.4 Train loss descend

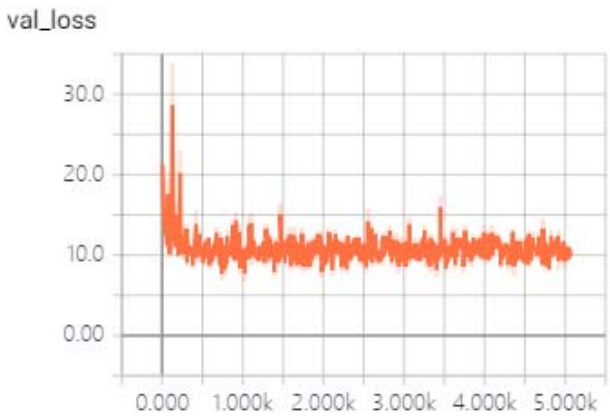

图 5 验证集损失下降曲线图 Fig.5 Validation set descend

2.2 测试结果

本次测试集总共有 130 张红外图片，包含 248 个舰船目标，还有一段视频，用于测试该网络的检测帧率。如图 6、图 7 所示，该网络在该平台上检测视频的速度高达 18 fps，而 YOLOv3 只有 6 fps，整整提高了 3 倍。该网络对于海杂波遮挡和海天线的干扰也有很强的抗干扰性，图 8、图 9、图 10。对测试图片检测后部分统计结果如表 1、表 2 所示，其中 confidence 是 YOLOv3 输出时使用非极大值抑制的阈值，Total nums 为模型检测目标的总数，Ground True 是测试数据上真实目标总数，True 是检测目标数和实际目标在能匹配上的数量，即可以认为检测出的目标是真实目标的数量。

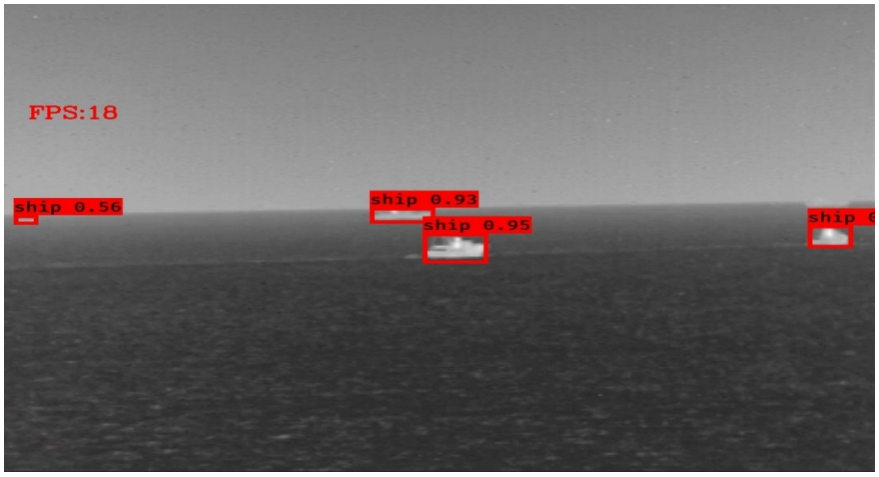

图 6 改进 YOLOv3 视频检测 Fig.6 Modified YOLOv3 video test

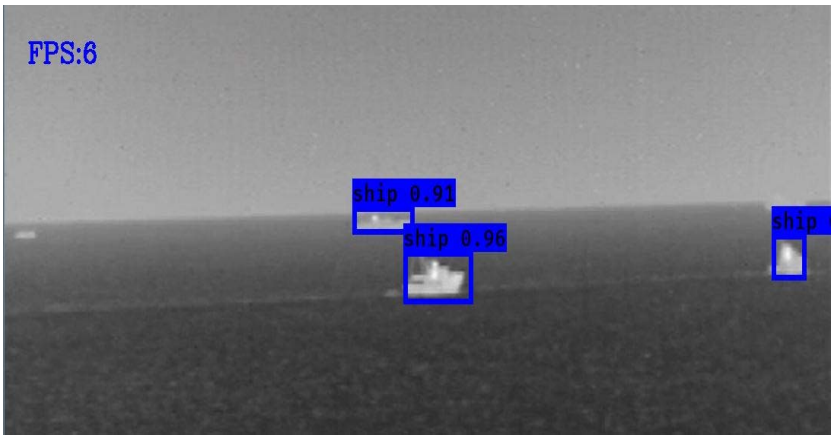

图 7 YOLOv3 视频检测 Fig.7 YOLOv3 video test

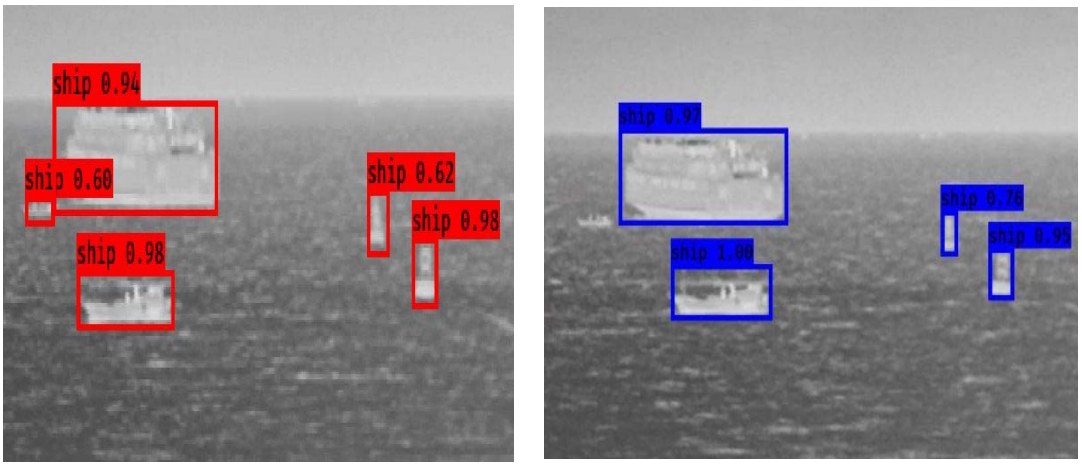

图 8　改进 YOLOv3（左）和 YOLOv3（右）对海杂波的抗干扰

Fig.8　Modified YOLOv3(left) and YOLOv3(right) anti-interference of the wave

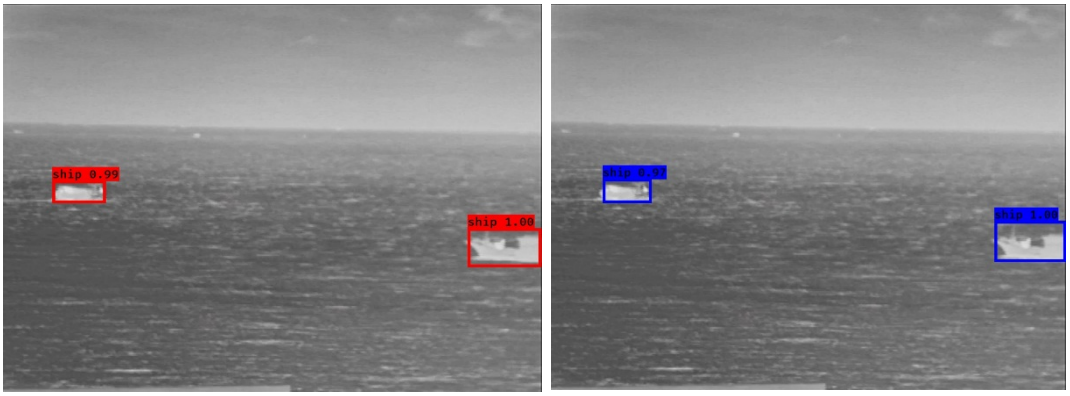

图 9　改进 YOLOv3（左）和 YOLOv3（右）对遮挡情况下的抗干扰

Fig.9　Modified YOLOv3(left) and YOLOv3(right) anti-interference of cover

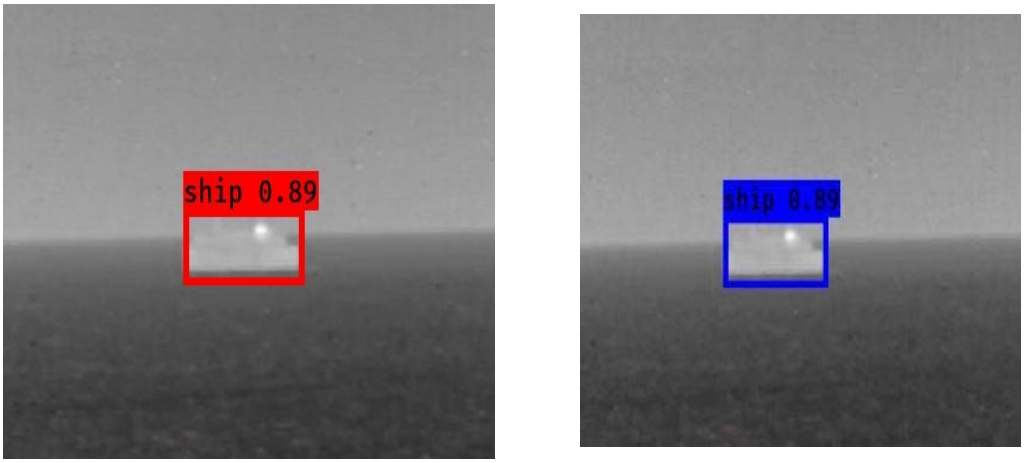

图 10　改进 YOLOv3（左）和 YOLOv3（右）对海天线的抗干扰

Fig.10　Modified YOLOv3(left) and YOLOv3(right) anti-interference of the sea-sky-line

表 1　改进 YOLOv3 在 IOU＝0.3 测试数据　　Table 1　Modified YOLOv3 test IOU=0.3

Confidence	Total nums	True	Ground True	Precision	Recall
0.1	217	147	148	0.99	0.68
0.15	196	143	148	0.97	0.73
0.2	174	132	148	0.89	0.76
0.25	156	127	148	0.86	0.81
0.3	142	121	148	0.82	0.85

YOLOv3 在测试集的统计结果如表 3、表 4 所示。通过表 1，表 3 对比，表 2 和表 4 的对比，我们可以看出，改进后的 YOLOv3 和 YOLOv3 检测精度的差距很小，为了更为直观地表达，我们使用表 1 和表 3，的数据绘制了准确率（Precision）和召回率（Recall）的 *P-R* 曲线（IOU＝0.3），如图 11 所示。从图中我们可以看出，改进 YOLOv3 和 YOLOv3 在检测精度上确实相差很小，大约只有 1%的检测精度损失。

表 2　改进 YOLOv3 在 IOU＝0.5 测试数据

Table 2　ModifiedYOLOv3 test IOU=0.5

Confidence	Total nums	True	Ground True	Precision	Recall
0.1	202	137	148	0.93	0.80
0.15	181	130	148	0.88	0.72
0.2	168	124	148	0.84	0.74
0.25	153	115	148	0.78	0.75
0.3	136	107	148	0.72	0.79

表 3　YOLOv3 在 IOU＝0.3 测试数据

Table 3　YOLOv3 test IOU=0.3

Confidence	Total nums	True	Ground True	Precision	Recall
0.1	165	119	148	0.80	0.72
0.2	135	109	148	0.74	0.81
0.3	126	104	148	0.72	0.82
0.4	120	100	148	0.68	0.83
0.5	114	97	148	0.66	0.85

表 4　YOLOv3 在 IOU＝0.5 测试数据

Table 4　YOLOv3 test IOU=0.5

Confidence	Total nums	True	Ground True	Precision	Recall
0.1	216	143	148	0.96	0.66
0.2	158	119	148	0.80	0.75
0.3	141	110	148	0.74	0.78
0.4	132	106	148	0.72	0.80
0.5	124	103	148	0.69	0.83

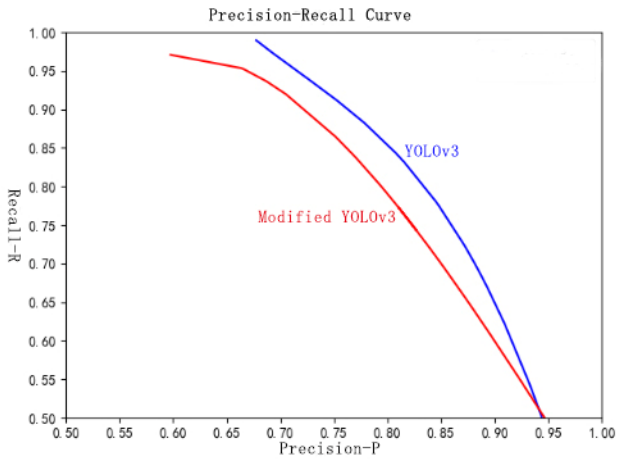

图 11　改进 YOLOv3 和 YOLOv3 的 *P-R* 曲线

Fig.11　Modified YOLOv3 and YOLOv3 *P-R* curves

3　结论

通过数据可以看出，改进 YOLOv3 的精度和原 YOLOv3 相差不大，改进 YOLOv3 比 YOLOv3 精度小于 1%，相差不大，但是改进 YOLOv3 在此平台上能实现 18 fps 的速度，而原 YOLOv3 只有 6 fps，速度提升了 3 倍，模型参数从原 YOLOv3 的 235 MB 减少到了 23.5 MB，这对于模型移植到嵌入式系统或者一个 GPU 实现多路视频检测非常有利。从结果来看，改进的 YOLOv3 在红外舰船目标检测应用中性能有了很大的改善，特别是速度和对硬件资源的要求方面有了很大的提升，对于工程化应用而言，在大大节约成本的同时，又可以提供可靠的实时性保障。此外，由于红外舰船图像数据稀缺，用于本文实验的图像数

据总共只有 1330 张，这对于深度学习而言，实在是太少了，相信如果有足够的数据来训练网络，实验结果还会有所提升。

参考文献:

[1] ZHAN J P, HUANG X J, SHEN Z X, et al. Target tracking method based on mean shift and Kalman filter[J]. *Journal of Chongqing University of Technology*: Science and Technology, 2010, **24**(3):76-80.

[2] 张世博, 李梦佳, 李乐, 等. 基于方向梯度直方图的行人检测与跟踪[J]. **北京石油化工学院学报**, 2013, **21**(4): 37-40.

ZHANG Shibo, LI Mengjia, LI Le, et al. Human detection and tracking based on HOG descriptor[J]. *Journal of Beijing Institute of Petro-chemical Technology*, 2013, **21**(4): 37-40.

[3] 张鹏. 基于卷积神经网络的光学遥感图像中机场目标识别研究[D]. 长沙: 国防科学技术大学, 2016.

ZHANG Peng. Airport Detection in Optical Remote Sensing Images with Convolution Neural Network[D]. Changsha: National University of Defense Technology, 2016.

[4] Girshick R. Fast R-CNN[C]//*IEEE International Conference on Computer Vision*, 2015: 1440-1448.

[5] REN S, HE K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//*International Conference on Neural Information Processing Systems*, 2015: 91-99.

[6] 戴陈卡, 李毅. 基于 Faster RCNN 以及多部件结合的机场场面静态飞机检测[J]. **计算机应用**, 2017, **37**(s2): 85-88.

DAI C K, LI Y. Aeroplane detection in static aerodrome based on faster RCNN and multi- part model[J]. *Journal of Computer Applications*, 2017, **37**(s2): 85-88.

[7] 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.

[8] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of CVPR*, 2015: 779-788.

[9] LIU W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//*European Conference on Computer Vision*, Springer International Publishing, 2016: 21-37.

[10] CAO Shuo, ZHAO Dean, LIU Xiaoyang, et al. Real-time robust detector for underwater live crabs based on deep learning[J]. *Computers and Electronics in Agriculture*, 2020: 172.

[11] Redmon J, Farhadi, A. YOLOv3: an incremental improvement[Z/OL][2018-04]. https://www.researchgate.net/publication/324387691_YOLOv3_ An_Incremental_Improvement.

[12] ZHANG Xiang, YANG Wei, TANG Xiaolin, et al. A fast learning method for accurate and robust lane detection using two-stage feature extraction with $YOLOv_3$[J]. *Sensors*, 2018, **18**(12): 4308.