

# 一种基于深层次多尺度特征融合 CNN 的 SAR 图像舰船目标检测算法

杨龙<sup>1</sup>, 苏娟<sup>1\*</sup>, 黄华<sup>2</sup>, 李响<sup>1</sup>

<sup>1</sup>火箭军工程大学核工程学院, 陕西 西安 710025;

<sup>2</sup>西北工业大学航海学院, 陕西 西安 710072

**摘要** 基于深度学习的目标检测技术在目标检测领域有强大的生命力,但是将其用于合成孔径雷达(SAR)图像舰船目标检测时并没有达到预期的效果。提出了一种基于卷积神经网络的 SAR 图像舰船目标检测算法用来检测多场景下的多尺度舰船目标,在单发多盒探测器检测框架的基础上,使用性能更好的 Darknet-53 作为特征提取网络,加入更深层次的特征融合网络,生成语义信息更加丰富的新的特征预测图。同时在训练策略上使用了一种新的二分类损失函数来解决训练过程中难易样本失衡的问题。在扩展的公开 SAR 图像舰船数据集上进行验证实验,实验结果表明,所提方法对复杂场景下不同尺寸的舰船目标的检测展现出了良好的适应性。

**关键词** 机器视觉; 合成孔径雷达; 神经网络; 舰船目标检测; 单发多盒探测器; 复杂场景

中图分类号 TP751.1

文献标志码 A

doi: 10.3788/AOS202040.0215002

## SAR Ship Detection Based on Convolutional Neural Network with Deep Multiscale Feature Fusion

Yang Long<sup>1</sup>, Su Juan<sup>1\*</sup>, Huang Hua<sup>2</sup>, Li Xiang<sup>1</sup>

<sup>1</sup> College of Nuclear Engineering, Rocket Force University of Engineering, Xi'an, Shaanxi 710025, China;

<sup>2</sup> School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

**Abstract** Object detection technology based on deep learning has shown excellent performance in the field of object detection; however, it has not yielded expected results when used for synthetic aperture radar (SAR) ship detection. Herein, an SAR ship detection method based on a convolutional neural network is proposed for multiscale ship detection in multiple scenarios. Based on the single shot multiBox detector, we use Darknet-53 as the feature extraction network. A deep feature fusion network is added to generate new feature prediction maps with rich semantic information. In addition, we use a new two-class loss function in the training strategy to deal with the imbalance in the difficult and easy samples in the training process. The verification experiments are performed on the expanded public SAR ship detection dataset. The experimental results indicate that our proposed method has a good adaptability to SAR ship detection at different sizes in complex scenes.

**Key words** machine vision; synthetic aperture radar; neural network; ship detection; single shot multiBox detector; complex background

**OCIS codes** 150.0155; 100.4996; 100.2000

## 1 引 言

与光学传感器不同,合成孔径雷达(SAR)是一种具有全天时、全天候数据获取能力的微波主动式传感器,在数据获取方面具有其独特的优势。近年来,SAR 图像在各个领域取得了广泛的应用<sup>[1]</sup>,

TerraSAR-X、Sentinel-1 和高分三号等卫星的发展更是为 SAR 图像的获取提供了广泛的数据来源。随着船舶海洋技术的发展,SAR 图像舰船目标的检测在军用和民用领域发挥着越来越关键的作用,如军用领域精确制导武器与民用领域海洋监测等。因此,研究 SAR 图像舰船目标的检测算法具有重要的

收稿日期: 2019-07-30; 修回日期: 2019-08-22; 录用日期: 2019-09-09

基金项目: 国家自然科学基金(41574008)

\* E-mail: yangl03@mail.nwpu.edu.cn

意义。

传统的 SAR 图像舰船检测算法大多针对特定场景,算法稳健性较差,对复杂港口下的舰船目标容易产生虚警和漏检,且处理步骤繁琐。针对信噪比低的场景,文献[2]提出了一种新的基于超像素级局部信息测量的极化 SAR 舰船自动检测方法,为了增强舰船目标与海杂波的对比如度,提出了三种超像素级的不同尺度,通过实验验证了该方法在信噪比低的场景下的有效性。针对海面密集型舰船检测,文献[3]提出了一种基于陷波滤波器的 SAR 舰船检测算法,为 SAR 图像海面舰船目标检测提供了一种新的思路。针对强杂波的复杂背景,文献[4]提出了一种基于局部对比度方差加权信息熵的海洋 SAR 图像舰船检测算法,该方法在强杂波和异质背景下都能获得稳定的检测性能。然而传统的 SAR 图像舰船检测算法大多需要对图片进行预处理,算法依然存在着泛化性差、效率低等缺点。

随着计算机硬件技术的发展,基于深度学习<sup>[5]</sup>的目标检测技术已经成为当前目标检测领域的主流方向,而卷积神经网络(CNN)<sup>[6]</sup>具有自带参数共享等天然优势,已经在目标检测领域取得了广泛的应用<sup>[7]</sup>。基于 CNN 的目标检测算法可以分为以 R-CNN<sup>[8]</sup>、Fast R-CNN<sup>[9]</sup>和 Faster R-CNN<sup>[10]</sup>为代表的双阶段目标检测算法,以及以 YOLO(You only look once)<sup>[11]</sup>和 SSD(Single Shot MultiBox Detector)<sup>[12]</sup>为代表的单阶段目标检测算法。双阶段检测算法流程包括候选区域提取与候选区域定位和分类两个阶段,检测精度高但是效率较低。单阶段检测算法直接在图像上划分区域进行分类与回归定位,兼顾了精度与效率。单阶段检测算法在 YOLO 和 SSD 的基础上又相继衍生出了 YOLOV2、YOLOV3、FSSD(Feature Fusion Single Shot Multibox Detector)、DSSD(Deconvolutional Single Shot Detector)、Retianet、RFBNet(Receptive Field Block Net for Accurate and Fast Object Detection)、RefineDet(Single-Shot Refinement Neural Network for Object Detection)等众多版本<sup>[13-19]</sup>,提升了检测精度和效率。将基于深度学习的方法应用到 SAR 图像舰船目标检测领域中,文献[20]提出了一种基于深度卷积神经网络的 SAR 图像舰船目标检测技术,在 Farster R-CNN 的基础上,提出了特征聚合、迁移学习等创新,检测的平均准确率较高;文献[21]针对 CNN 检测器不能有效利用图像空间信息等缺点,提出了一种新的用于

SAR 舰船目标检测的级联 CNN 制导视觉注意方法,该方法明显地减少了目标的漏检和错检;文献[22]提出了一种基于 Faster R-CNN 的密集连接多尺度神经网络,在多个公开的数据集上进行测试,取得了很好的检测效果。但是将深度学习的方法应用到 SAR 图像舰船目标检测领域时依然没有达到预期的效果,笔者认为主要有以下两个原因:一是目前用于 SAR 舰船检测的深度学习网络难以适应 SAR 舰船目标的多尺度性,尤其是对小目标的检测效果不佳,主要原因是从单一的特征层提取特征图来预测时忽略了上下文信息的关联性。CNN 底层的特征图分辨率较高,可以实现更精确的定位,高层特征图具有更深层次的语义信息,但是分辨率较低,不同卷积层的特征图目标分辨率对舰船检测有着不同的影响。二是使用 CNN 检测 SAR 图像舰船目标时,若目标为斑点噪声严重的海面 and 港口密集型舰船,则容易产生漏检和错检,在近海复杂背景下则容易对非舰船目标产生虚警。主要原因是在深度学习训练的过程中难易样本的不平衡。在双阶段检测算法中,解决这类问题的主要方法是利用在线困难样本挖掘(OHEM)<sup>[23]</sup>,在 OHEM 中,根据所考虑的每个样本的当前损失对训练样本进行采样,但是将其用于 SAR 图像舰船目标检测时仍然存在一定的缺陷,OHEM 并不能保证选择的负面样本包括与舰船相似的近海干扰目标以及港口密集型舰船目标。

针对 SAR 舰船目标的多尺度性给检测带来的困难,本文在 SSD 单阶段检测算法的基础上,提出了一种深层次多尺度特征融合的 CNN(DMFFCNN),通过多尺度特征融合的方式增强了网络对不同尺度舰船目标的适应能力。同时为了避免场景带来的干扰,本文使用了一种新的训练策略,使用了一种新的二分类损失函数,通过减少损失函数中简单样本的权重使网络在训练的过程中更加关注困难样本的挖掘,并且进行了多组对比分析实验进行实验验证。

## 2 算 法

### 2.1 网络结构

本文提出了一种基于深层次多尺度特征融合 CNN 的 SAR 图像舰船目标检测算法——DMFFCNN,网络结构如图 1 所示,其中 Conv 表示卷积操作,BN 表示批归一化处理,LeakyReLU 表示激活函数,concat 是一种特征融合的操作方式。其基本原理借鉴 SSD 目标检测算法,基于前馈卷积

网络,首先产生固定大小的默认框集合;然后基于所产生的默认框,对不同层次的特征图进行位置的回归和类别的预测;最后利用非极大值抑制算法对每个类别预测到的所有先验框集合进行筛选,去除多余的和概率较低的边框,生成最终的检测结果。CNN 不同层的特征图具有不同的空间分辨率和语义信息,其底层特征图特征丰富,空间分辨率高,适

合小目标的检测,深层特征图语义信息丰富,空间分辨率低,适合大目标检测。为了更好地解决海面 and 港口多尺度舰船目标的检测问题,本文首先使用性能更好的 Darknet-53 作为特征提取网络,输入图片尺寸为  $320 \times 320$ ,经过两次特征融合后生成新的特征预测图,同时使用了一种新的二分类损失函数改进训练策略,减小了复杂场景对舰船目标检测的影响。

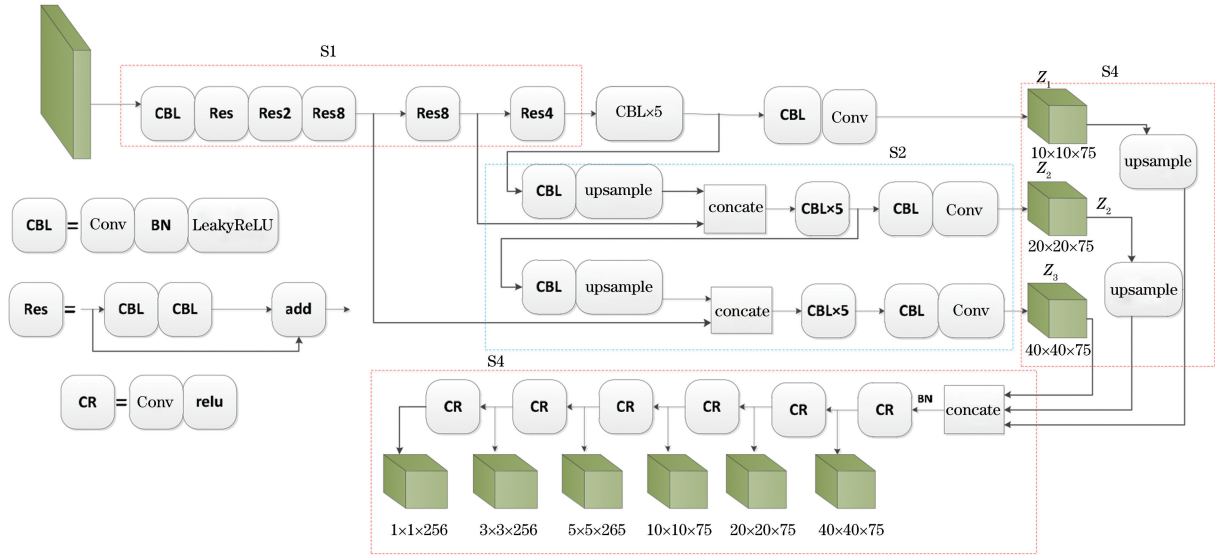


图 1 DMFFCNN 网络结构图

Fig. 1 Network structure of DMFFCNN

DMFFCNN 网络结构主要由图 1 所示 S1、S2、S3 和 S4 四部分组成,图中 CBL 模块表示 Conv + BN<sup>[24]</sup> + LeakyReLU, CBL × 5 表示 5 个 CBL 模块级联,具体为三个卷积核为  $1 \times 1$  的 CBL 模块中间加两个卷积核为  $3 \times 3$  的 CBL, Res 表示残差模块,图中 Res2、Res4、Res8 分别表示两个、四个、八个 Res 模块级联。其中 S1 为 Darknet-53,作为本文选取的基础网络架构,Darknet-53 是一个 53 层的卷积网络,由残差单元叠加而成,其浮点运算少,速度快,由文献[14]可知,Darknet-53 在兼顾分类精度与效率时,优于 ResNet-101、ResNet-152 以及 Darknet-19。S2 为第一次特征融合部分,结构如图 2(a)所示。由图 1 可知,有两个 Res8 模块和一个 Res4 模块,S2 模块从第一个 Res8、第二个 Res8 和之后的 Res4 层提取三个特征图,尺寸大小分别为  $40 \times 40$ ,  $20 \times 20$ ,  $10 \times 10$ ,对 Res4 层特征图进行上采样操作并与第二个 Res8 层特征图进行 concatenate 特征融合操作,对第二个 Res8 层特征图进行上采样操作并与第一个 Res8 层特征图进行 concatenate 特征融合操作,通过特征融合得到的三张含有丰富语义信息的特征图。在上采样之前加入 CBL × 5 既可以压缩特

征,加速收敛,又可以压缩维度,减少计算量。在 concatenate 特征融合操作之后加入  $3 \times 3$  的卷积减小上采样的混叠效应。具体计算过程为

$$Y_1 = W_1(X), \quad (1)$$

$$Y_f = W_i \{ T_c \{ \varphi_j [CBL_{1 \times 1} (Y_{f-1})] \} \}, \quad i = 2, 3, f = 2, 3, \quad (2)$$

$$Z_r = Conv_{1 \times 1} [CBL_{3 \times 3} (Y_f)], \quad r = 1, 2, 3, \quad (3)$$

式中:  $X$  表示 Res4 层特征图;  $W_i$  表示结构中的第  $i$  次 CBL<sub>5×5</sub> 操作;  $Y_f$  表示第  $f$  次 CBL<sub>5×5</sub> 操作后的特征图;  $\varphi_j$  表示 S2 模块中的第  $j$  次上采样操作,  $j$  取 1 和 2;  $T_c$  表示 S2 模块中的第  $c$  次 concatenate 特征融合操作,  $c$  取 1 和 2;  $Z_r$  表示 S3 模块中的第  $r$  个特征图; Conv<sub>1×1</sub> 表示卷积核为  $1 \times 1$  的卷积; CBL<sub>3×3</sub> 和 CBL<sub>1×1</sub> 表示卷积核分别为  $3 \times 3$  和  $1 \times 1$  的 CBL 模块计算。

S3 与 S4 为第二次特征融合部分,在 S2 部分得到三层特征图后,进行深层次的特征融合操作,结构如图 2(b)所示,对尺寸较小的两张特征图进行上采样操作后与尺寸大的特征图进行 concatenate 特征融合,再进行图 4 所示的卷积操作,提取新的 6 张特征图进行预测。计算过程为

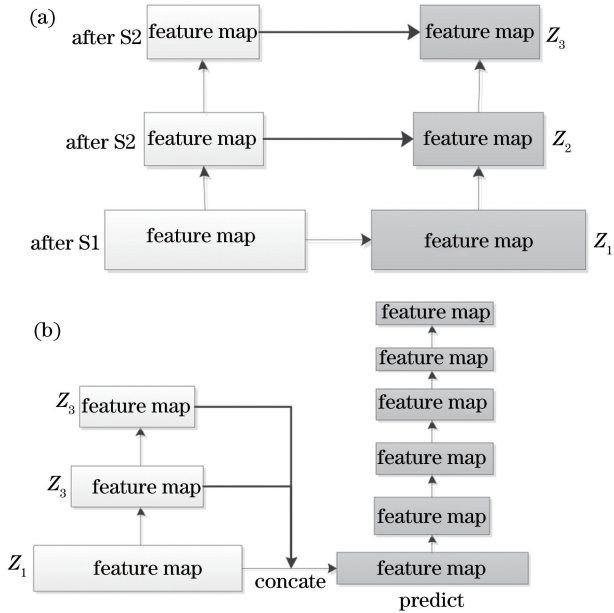


图 2 特征融合操作示意图。(a)第一次特征融合示意图；  
(b)第二次特征融合示意图

Fig. 2 Diagram of feature fusion. (a) The first feature fusion; (b) the second feature fusion

$$G = T[\varphi(Z_r)], \quad (4)$$

$$Z_h = Q(G), \quad (5)$$

$$F = \Phi_{\text{conf, loc}}(Z_h), \quad (6)$$

式中： $\varphi$  表示 S4 模块中上采样操作，采用双线性插值法； $T$  表示 S4 模块中 concat 特征融合操作； $G$  表示第二次特征融合后的首张特征图； $Q$  表示图 3 所示形成的特征图金字塔结构； $Z_h$  表示最终用来预测的  $h$  张特征预测图； $\Phi_{\text{conf, loc}}$  表示在形成的特征图上进行预测，conf 表示分类，loc 表示定位； $F$  表示预测结果。

用来预测的特征图大小分别为  $40 \times 40$ ,  $20 \times 20$ ,  $10 \times 10$ ,  $5 \times 5$ ,  $3 \times 3$ ,  $1 \times 1$ ，不同的特征图设置不同的先验框数目，先验框的尺寸随特征图大小的降低呈线性递增，即

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1}(k - 1), \quad k \in [1, m], \quad (7)$$

式中： $m$  为特征图的个数； $k$  表示  $m$  个特征图中的第  $k$  张特征图； $s_k$  表示先验框与图片的尺寸比例； $s_{\max}$  表示比例的最大值， $s_{\min}$  表示比例的最小值。文中  $s_{\max}$  取 0.9， $s_{\min}$  取 0.2，先验框数目分别为 4, 6, 6, 6, 4, 4，SSD 先验框尺度长宽比一般选取 1, 2, 3, 1/2, 1/3。

## 2.2 训 练

先验框匹配：训练阶段需要将样本的真实标签框与先验框进行匹配，遵循两个原则，第一个原则是

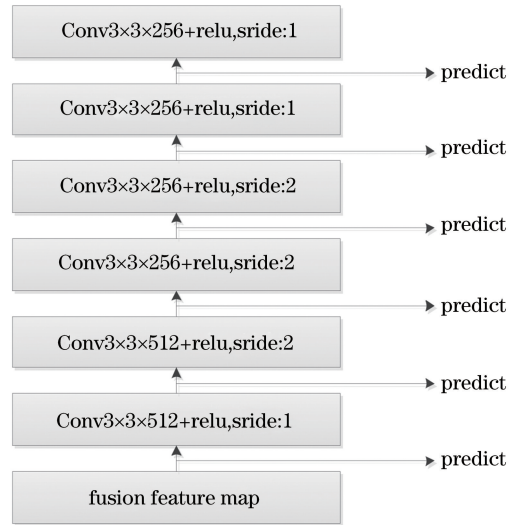


图 3 特征预测图生成器

Fig. 3 Feature prediction map generator

找出图片中的真实标签框与先验框重叠度 (IOU) 最大的先验框进行匹配，第二个原则是将图片中的真实标签框与 IOU 大于 0.5 的先验框进行匹配，当 IOU 大于 0.5 时为正样本，否则为负样本，但是负样本的数量仍然远远大于正样本数量。为了保证正负样本数量均衡，采用了困难样本挖掘技术 (hard negative mining)，对负样本进行抽样，抽样时按照置信度误差进行降序排列，选取误差较大的样本作为训练的负样本，保证正负样本数量比例接近 1:3。

损失函数：原始的 SSD 检测算法的损失函数定义为位置误差 ( $L_{\text{loc}}$ ) 与置信度误差 ( $L_{\text{conf}}$ ) 的加权和，详见文献[12]。

对于 SAR 图像舰船目标检测而言，置信度损失函数为 softmax 函数在两类目标情况下 (舰船与非舰船) 的损失函数，即标准的交叉熵损失函数

$$L_{\text{conf}} = -\text{lb}(p_t), \quad (8)$$

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & y = -1 \end{cases}, \quad (9)$$

式中： $p_t$  为舰船标签类的模型估计概率； $p$  为真值框的类别； $y$  表示真实标签框的类别，取值为 +1 或者 -1。但是该模型不会区分难易样本，对于 SAR 舰船样本，在复杂场景下，存在与舰船目标相似的干扰物。在斑点噪声严重的背景下，对于密集型舰船而言，很可能存在两艘舰船连在一起难以鉴别的情况。为了避免复杂场景对 SAR 舰船检测带来的干扰，本文使用了一种新的二分类损失函数：

$$L_{\text{conf}} = -(1 - p_t)^\lambda \text{lb}(p_t), \quad (10)$$

式中， $\lambda$  为权重因子。分类损失函数中通过引入参数  $\lambda$  可以调整简单样本的权重。当  $\lambda$  等于 0 时，损

失函数变为原始的标准交叉熵损失函数。当  $p_i$  接近 1 时,说明当前样本的分类损失很小,样本很容易实现分类正确,属于简单样本,会给当前样本损失值赋予较小的权重。当  $p_i$  接近 0 时,  $-\ln(p_i)$  趋近于无穷大,分类严重出错,属于困难样本,会给当前的分类损失值赋予较大的权重。图 4 所示为在不同参数条件下损失函数随  $p_i$  的变化曲线,可以看出通过引入权重因子  $\lambda$  减小简单样本在训练过程中所占的比例。

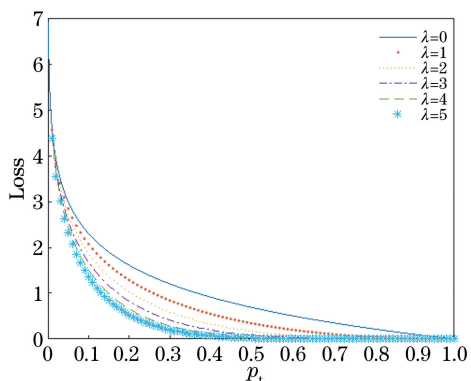


图 4 不同  $\lambda$  下分类损失函数曲线

Fig. 4 Classification loss function for different  $\lambda$

### 2.3 预 测

预测阶段,采用原始的 SSD 预测方法,对于每一个预测框,根据类别置信度确定目标类别与置信度值,过滤掉属于背景的预测框与置信度阈值低于 0.5 的预测框。对剩余的预测框进行解码降序排列,保留一定的预测框,最后使用非极大值抑制算法

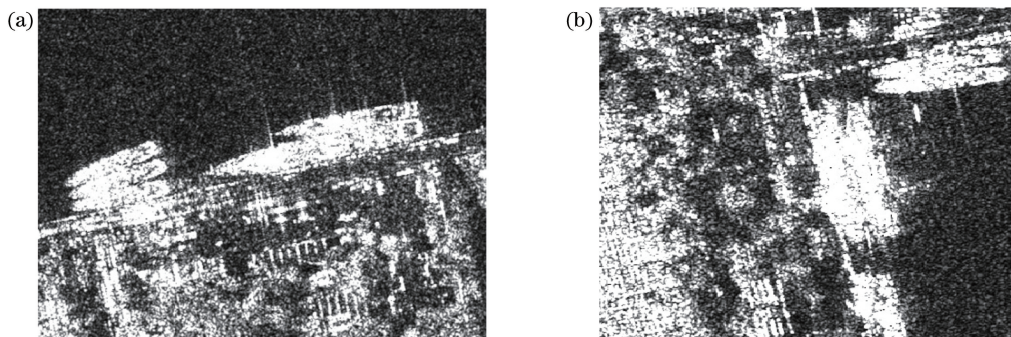


图 5 复杂背景密集型舰船切片

Fig. 5 Intensive ship slices in complex background

### 3.3 评价指标

本文使用的评价指标为平均准确率(AP)与检测速率(FPS)。平均准确率定义为

$$R_{AP} = \int_0^1 P(R) dR, \quad (11)$$

式中: $R$  表示召回率; $P$  表示准确率。

对其预测结果进行处理,消除没有目标的默认框和重复框,剩余的预测框就是检测结果。

## 3 实验结果与分析

### 3.1 实验平台与参数设置

本文所使用的计算机硬件平台是 CPU intel<sup>®</sup> Xeon (R) CPUE5-2603v4 @ 1.70 GHz  $\times$  6 和 NVIDIA GTX Titan GPU,在 Ubuntu16.04 系统上运行,采用的网络开发框架为 Pytorch。

使用在 ImageNet 数据集上预先训练好的 Darknet53 模型,训练的批尺寸(一次训练的样本数目)取 8,初始学习率为 0.001,学习率衰减权重为 0.0001,训练代数 epoch 取 300,参数更新方法为引入动量的梯度下降法,动量因子取 0.9。

### 3.2 数据集

公开的 SAR 舰船数据集 SSDD<sup>[20]</sup> 是模仿 PASCAL VOC 数据集建立的,共有 1160 幅图像,2456 条船,数据来源包括卫星 RadarSat-2, TerraSAR-X 和 Sentinel-1,分辨率包括 1, 3, 5, 7, 10 m,该数据集只标注了像素超过 3 pixel 的舰船。在此基础上,本文搜集了 800 张新的 SAR 舰船图像,裁成  $500 \times 500$  左右的尺寸,按照 PASCAL VOC 格式进行标注,主要包括近岸港口目标、密集型舰船目标(如图 5 所示)以及小目标舰船,数据来源包括高分三号卫星与 Sentinel 卫星,标注后加入到 SSDD 数据集当中,总共 1960 张图像,按照 7:3 的数量比例分为训练集和测试集。

检测速率(FPS)定义为每秒处理图像的帧数。

### 3.4 算法性能分析

为了更好地对所提算法进行评估,本文同时实现了 SSD、RFBNet、FSSD 算法,在本文数据集上进行训练与测试,作为本文的对比实验。图 6 所示为 4 种方法的精度随代数(epoch)的变化曲线。从图 6

中可以得出本文方法检测的平均准确率明显高于其他三种算法,随着训练的不断深入,模型能够稳定收敛。将四种方法(SSD、RFBNet、FSSD以及本文方法DMFFCNN)应用于SAR图像舰船目标检测时,本文方法取得了更好的效果。

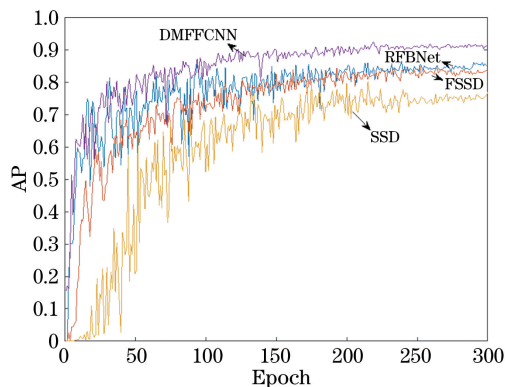


图 6 AP 变化曲线  
Fig. 6 Curves of AP

表 1 所示为 4 种方法的实验结果对比,从表 1 的数据中可以得出,将 SSD、FSSD 和 RFBNet 算法应用于 SAR 图像舰船目标检测时,对应的检测平均准确率分别为 76.1%、83.2% 与 85.2%。所提算法的平均准确率达到 91.2%,比原始的 SSD 检测算法高出 15.1 个百分点,比 RFBNet 检测算法高出 6 个百分点,比 FSSD 检测算法高出 8 个百分点,精度上取得较大的提升,但是不难发现,所提算法在网络结构上进行了更深层次的特征融合,导致检测时间增加,算法的检测效率有所下降。但是总的来说相比于双阶段目标检测算法,本文所提的单阶段目标检测算法在速度上仍然具有巨大的优势,能够满足实时检测的基本需求。

表 1 实验结果对比

Table 1 Comparison of experimental results

Method	Backbone	Input size	AP / %	Speed / (frame · s <sup>-1</sup> )
SSD <sup>[12]</sup>	VGG16	300	76.1	45
FSSD <sup>[15]</sup>	VGG16	300	83.2	36
RFBNet <sup>[18]</sup>	VGG16	300	85.2	39
DMFFCNN (λ = 0)	Darknet-53	320	91.2	29

为了进一步验证本文深层次特征融合的效果,在相同基础网络结构(backbone)的情况下,本文同时在网络结构中提取的 Z<sub>1</sub>、Z<sub>2</sub>、Z<sub>3</sub> 特征图上进行预测,计算结果如表 2 所示,检测的平均准确率达到 86.5%,在经过第二次特征融合后达到了 91.2%,两

次特征融合后检测精度明显提升。

表 2 实验结果对比

Table 2 Comparison of experimental results

Method	Backbone	Input size	AP / %	Speed / (frame · s <sup>-1</sup> )
Feature fusion-1	Darknet-53	320	86.5	35
Feature fusion-2	Darknet-53	320	91.2	29

在提出了深层次特征融合的网络结构后,为了进一步提升检测效果,提高本文算法对复杂场景适应能力,使用了新的分类损失函数,表 3 所示为本文方法使用二分类损失函数时在不同参数 λ 下的检测精度对比,可知在 λ = 3 时,本文算法的平均准确率最高,达到了 92.8%。因此,在训练的过程中,通过在损失函数中引入参数 λ 改变难易样本在训练中所占的权重有效地提高了算法的检测精度。

表 3 不同 λ 下的实验结果对比

Table 3 Comparison of experimental results for different λ

λ	AP / %
0	91.2
1	92.1
2	92.0
3	92.8
4	92.0
5	91.8

图 7 所示为 4 种典型的困难样本在本文四种算法下的可视化的检测效果对比。图 7(a)是包含靠岸舰船与海面舰船在内的复杂场景下的 SAR 舰船图片,可以看出,在原始的 SSD 算法下,靠岸舰船漏检率较高,RFBNet 与 FSSD 的效果有一定的提升,但是仍然存在漏检,本文算法对海面舰船和靠岸舰船的检测效果有较大的提升;图 7(b)是典型的海面密集型小尺寸舰船目标,可以发现,原始的 SSD 算法对小目标检测效果较差,漏检严重,RFBNet 与 FSSD 的检测效果稍好但是仍然存在漏检,本文算法对小目标检测的效果取得了提升;图 7(c)是典型的小尺寸港口舰船目标,受环境干扰严重,可以发现 SSD、RFBNet 与 FSSD 对其中的舰船均存在漏检,而本文算法能够精确检测,说明本文算法对复杂场景干扰下的 SAR 图像舰船目标检测适应力较强;图 7(d)是包含大尺寸与小尺寸舰船目标在内的 SAR 舰船图像,图像中有两艘舰船

紧密相邻,可以发现 SSD、RFBNet 与 FSSD 算法将左边两艘舰船都错检成了一艘,而本文算法能够较好地分辨出来。

图 8 所示为本文 DMFCNN 算法在  $\lambda=0$  与  $\lambda=3$  时的检测效果对比,图 8(a)~(d)均为既包含

舰船目标又包含类似舰船的场景干扰的 SAR 图像,对比图 8(a)和图 8(b),图 8(c)和图 8(d)可以发现,在  $\lambda=0$  时,图 8(a)、(c)均存在将干扰目标检测为舰船的虚警,当  $\lambda=3$  时,增大了困难样本在训练中的权重比例,使得误报明显减少。

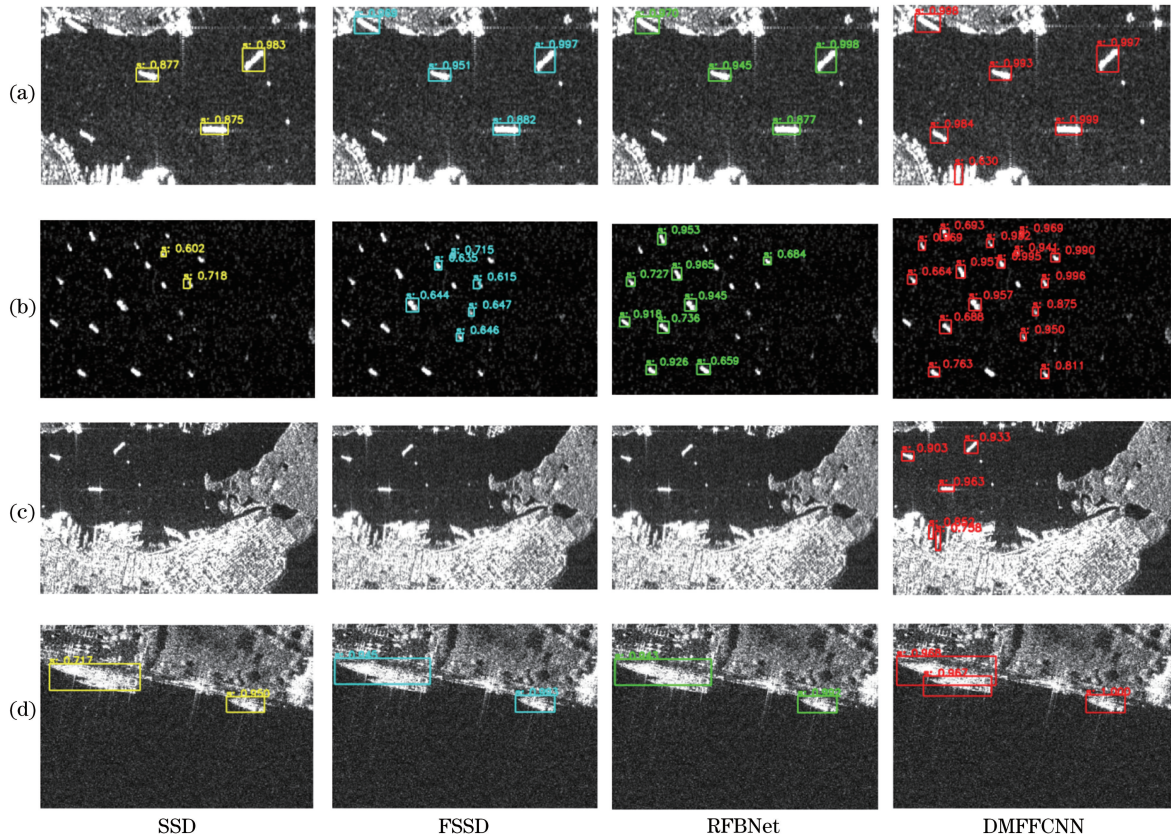


图 7 检测结果对比

Fig. 7 Comparison of detection results

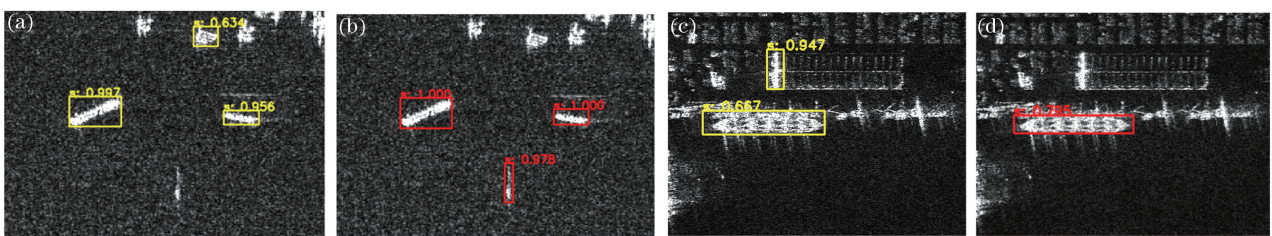


图 8  $\lambda=0$  与  $\lambda=3$  时检测效果对比。(a)  $\lambda=0$  时,场景干扰下的海面舰船的检测结果;(b)  $\lambda=3$  时,场景干扰下海面舰船的检测结果;(c)  $\lambda=0$  时,场景干扰下的港口舰船的检测结果;(d)  $\lambda=3$  时,场景干扰下的港口舰船的检测结果

Fig. 8 Comparison of detection results of DMFCNN when  $\lambda=0$  and  $\lambda=3$ . (a) Detection results of SAR surface ships in scene interference when  $\lambda=0$ ; (b) detection results of SAR surface ships in scene interference when  $\lambda=3$ ; (c) detection results of SAR inshore ships in scene interference when  $\lambda=0$ ; (d) detection results of SAR inshore ships in scene interference when  $\lambda=3$

## 4 结 论

针对当前将 CNN 应用于 SAR 图像舰船目标检测时存在的问题,提出了一种基于深层次多尺度

特征融合 CNN 的 SAR 图像舰船目标检测算法,并且使用了一种新的训练策略,在网络结构中通过对深层特征图进行上采样并与底层特征图相融合的方法,使得新的特征图能够同时兼顾底层定位信息与

深层语义信息。同时损失函数中使用了引入参数 $\lambda$ 的新的二分类损失函数,使得本文算法对复杂场景下的多尺度舰船目标的稳健性更好。实验结果表明,与当前多种深度学习的检测算法相比,本文算法的检测精度更高,同时检测效率也能够满足现代硬件条件下实时检测的需求。在未来的研究中,将进一步扩展 SAR 图像舰船目标的数据库,优化神经网络结构,提升检测精度与检测效率。

## 参 考 文 献

- [1] Yi W, Zeng Y, Yuan Z. Fusion of GF-3 SAR and optical images based on the nonsubsampling contourlet transform[J]. *Acta Optica Sinica*, 2018, 38(11): 1110002.  
易维, 曾湧, 原征. 基于 NSCT 变换的高分三号 SAR 与光学图像融合[J]. *光学学报*, 2018, 38(11): 1110002.
- [2] He J L, Wang Y H, Liu H W, et al. A novel automatic PolSAR ship detection method based on superpixel-level local information measurement [J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(3): 384-388.
- [3] Gao G, Gao S, He J, et al. Ship detection using compact polarimetric SAR based on the notch filter [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(9): 5380-5393.
- [4] Huo W B, Huang Y L, Pei J F, et al. Ship detection from ocean SAR image based on local contrast variance weighted information entropy[J]. *Sensors*, 2018, 18(4): 1196.
- [5] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1798-1828.
- [6] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]//*Advances in Neural Information Processing Systems 25 (NIPS 2012)*, December 3-6, 2012, Lake Tahoe, Nevada, United States. Canada: NIPS, 2012: 1097-1105.
- [7] Wang W X, Fu Y T, Dong F, et al. Infrared ship target detection method based on deep convolution neural network [J]. *Acta Optica Sinica*, 2018, 38(7): 0712006.  
王文秀, 傅雨田, 董峰, 等. 基于深度卷积神经网络的红外船只目标检测方法[J]. *光学学报*, 2018, 38(7): 0712006.
- [8] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 580-587.
- [9] Girshick R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1440-1448.
- [10] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [11] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.
- [12] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sze N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [13] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517-6525.
- [14] Redmon J, Farhadi A. YOLOv3: an incremental improvement [J/OL]. (2018-04-08) [2019-07-29]. <https://arxiv.org/abs/1804.02767>.
- [15] Li Z X, Zhou F Q. FSSD: feature fusion single shot Multibox detector [J/OL]. (2018-05-17) [2019-07-29]. <https://arxiv.org/abs/1712.00960>.
- [16] Fu C Y, Liu W, Ranga A, et al. DSSD: deconvolutional single shot detector [J/OL]. (2017-01-23) [2019-07-29]. <https://arxiv.org/abs/1701.06659>.
- [17] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2999-3007.
- [18] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection [C]//*Proceedings of the European Conference on Computer Vision (ECCV)*, September 8-14, 2018, Munich, Germa. New York: IEEE, 2018: 385-400.
- [19] Zhang S F, Wen L Y, Bian X, et al. Single-shot refinement neural network for object detection [C]//



- 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 4203-4212.
- [20] 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.  
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.
- [21] Zhao J P, Zhang Z H, Yu W X, et al. A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images [J]. *IEEE Access*, 2018, 6: 50693-50708.
- [22] Jiao J, Zhang Y, Sun H, et al. A densely connected end-to-end neural network for multiscale and multiscene SAR ship detection [J]. *IEEE Access*, 2018, 6: 20881-20892.
- [23] Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 761-769.
- [24] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning, July 6-11, 2015, Lille, France. USA: MIT Press, 2015, 37: 448-456.