北京航空航天大学学报
Journal of Beijing University of Aeronautics and Astronautics
ISSN 1001-5965,CN 11-2625/V

《北京航空航天大学学报》网络首发论文

题目：基于 EfficientDet 的无预训练 SAR 图像船舶检测器
作者：包壮壮，赵学军
DOI：10.13700/j.bh.1001-5965.2020.0255
收稿日期：2020-06-11
网络首发日期：2020-09-29
引用格式：包壮壮，赵学军．基于 EfficientDet 的无预训练 SAR 图像船舶检测器．北京航空航天大学学报. https://doi.org/10.13700/j.bh.1001-5965.2020.0255

基于 EfficientDet 的无预训练 SAR 图像船舶检测器

包壮壮[1]，赵学军[1]✉

（1. 空军工程大学 基础部，西安 710051）

✉通信作者 E-mail： 292457155@qq.com

摘要 现有的基于卷积神经网络的合成孔径雷达（synthetic aperture radar，SAR）图像船舶检测器并没有表现出其应有的出色性能。重要原因之一是依赖分类任务的预训练模型，没有有效的方法来解决 SAR 图像与自然场景图像之间存在的差异性。另一个重要原因是没有充分利用卷积网络各层的信息，特征融合能力不够强，难以处理包括海上和近海在内的多场景船舶检测，尤其是无法排除近海复杂背景的干扰。为此，提出了一种基于 EfficientDet 的无预训练目标检测器（Scratch-EfficientDet，SED），以解决多尺度、多场景的 SAR 船舶检测问题。在公开 SAR 船舶检测数据集上进行实验，SED 的检测精度指标 AP 达到 94.2%，与经典的深度学习检测器对比，超过最优的 RetineNet 模型 1.3 个百分点，在模型大小、算力消耗和检测速度之间达到平衡，验证了该模型在多场景条件下多尺度 SAR 图像船舶检测具有优异的性能。

关键词 船舶检测；合成孔径雷达；深度学习；卷积神经网络；目标检测

中图分类号 TN957.51；TP751 **文献标志码** A DOI: 10.13700/j.bh.1001-5965.2020.0255

Ship detector in SAR images based on EfficientDet without pre-training

BAO Zhuangzhuang[1], ZHAO Xuejun[1] ✉

（1. Department of Basic Science, Air Force Engineering University, Xi'an 710051, China）

✉E-mail： 292457155@qq.com

Abstract Existing synthetic aperture radar (SAR) image ship detectors based on convolutional neural networks does not show an excellent performance it should have. One of the important reasons is that they depend on the pre-training model of the classification tasks, and there is no effective method to solve the difference between the SAR image and the natural scene image. Another important reason is that the information of each layer of the convolutional network is not fully utilized, and the feature fusion ability is not strong enough to deal with the detection of ships in multiple-scene including sea and offshore, especially the interference of complex offshore background cannot be ruled out. Therefore, a without pre-training object detector (Scratch-EfficientDet, SED) based on EfficientDet is proposed to solve the problem of multi-scale and multi-scene SAR ship detection. Experiments are conducted on the public SAR ship detection data set, and the detection accuracy index AP of SED reaches 94.2%, which compared with the classic deep learning detector, has exceeded the best RetineNet model by 1.3 percentage points, achieved a balance between model size, computing power consumption and detection speed, verifying that the model can achieve an excellent performance on multi-scale SAR image ship detection in multi-scene.

Key words ship detection; synthetic aperture radar; deep learning; convolutional neural network; object detection

遥感图像船舶检测在民用和军用领域应用前景广泛，而合成孔径雷达（synthetic aperture radar，SAR）图像由于其全天候、全天时、可穿透的特点，是船舶检测的主要技术手段[1]。随着越来越多的

收稿日期：2020-06-11

作者简介：包壮壮 男，硕士研究生。主要研究方向：深度学习，目标检测。赵学军 男，博士，副教授，硕士生导师。主要研究方向：军用仿真的理论及技术，模式识别。

网络首发时间：2020-09-29 13:34:08 网络首发地址：https://kns.cnki.net/kcms/detail/11.2625.V.20200929.1010.001.html

SAR 卫星发射，原始的 SAR 图像数据日益增长，如何从中快速准确检测船舶成为亟需解决的问题[2,3]。

传统的 SAR 图像船舶检测通常分为四个阶段：地面遮掩、预处理、预筛选以及识别。地面遮掩将与海面成像结果悬殊的地面遮掩掉，从而减少探测面积[4]。预处理使用多种图像处理方法，如斑点噪声过滤等[5]，将原始的 SAR 图像转换成更易探测船舶的图像。预筛选阶段，将一些可能是船舶的像素提取为候选目标，在搜索这些候选像素的时候，主流方法是恒定虚警率法[6]和广义似然比测试法[7]。最后的识别阶段通过排除虚警区域同时接受包含真实目标区域来提取船只。这类方法利用 SAR 图像的统计信息，借助人类专家的先验知识，通过将图像转换到频域进一步提高了检测性能，但由于其核心思想是基于海杂波建模和参数估计进而找到代表船舶的像素点，同时由于人工标记特征的提取能力是有限的，在一些特殊场景，特别是在海岸线、港口等陆海交界的复杂环境下不能充分提取 SAR 图像船舶的特征，所以这些方法存在鲁棒性差，不能实现端到端操作，仍然需要人力干预的问题。

随着深度学习技术的发展，开始出现将卷积神经网络应用于 SAR 图像船舶检测的方法。由于目前主流的深度卷积神经网络模型是有监督的训练，大多包含数百万个参数，其需要大量人工标注的数据集进行训练。Li[8]和 Wang[9]分别基于 SAR 遥感卫星图像建立了 SAR 船舶检测数据集，同时运用主流目标检测器获得了基线（baseline）结果，Wang 运用 SSD512（Single Shot MultiBox Detector，单射多框检测器，512 指输入图像长宽为 512 像素）[10]、Faster RCNN（Faster Regions with CNN Features，更快的区域特征检测卷积神经网络）[11]、RetinaNet（视网膜网络）[12]分别获得了 AP（Average Precision，平均准确率）值为 89.43%、88.26%、91.36%的检测精度。从部分实验结果可以看到依然存在较多漏检、误检的情况，检测精度还有较大的提升空间。我们分析原因如下：（1）遥感图像和自然场景图像之间存在的跨领域不适用性，深度学习目标检测器大多基于自然场景的大型数据集上训练好的参数微调得来，这不能解决 SAR 成像机制导致的固有缺陷，如相干斑噪声，运动物体存在拖影、重叠、阴影等，即没有解决两者之间的差异性；（2）卷积神经网络由浅到深，每一层都提取成百上千的特征，浅层特征语义信息少，位置信息多，深层特征与之相反，如何充分利用、融合这些特征是解决假目标和背景干扰，实现多场景、多尺度检测的关键。

由于遥感图像与自然场景图像相比具有目标背景复杂、尺寸多变、有方向性等特点[13]，在普通场景下可行的迁移学习（利用在大型通用数据集上训练好的参数权重作为网络初始化，再用新的小样本训练）并未取得理想的成绩。随着脱离预训练的检测器性能达到甚至超过有预训练（pre-training）的网络[14,15]，我们尝试运用从头开始（scratch）训练的方式解决跨领域不适用的问题。SAR 图像的固有缺陷要求更高效的特征提取方法，为充分利用高低级特征图（feature map）的语义信息，在 FPN（Feature Pyramid Networks，特征金字塔网络）[16]结构的基础上进行改进，删减只有一个输入的节点，同时将尺度相同的输入和输出连接起来，形成类似 ResNet（Residual Network，残差网络）[17]的结构，再增加向下的路径，并将其整体看作一个层操作，多次连接从而加强特征融合。

为了实现这些目标，提出一种基于 EfficientDet（Scalable and Efficient Object Detection，可扩展且高效的目标检测器）[18]的脱离预训练参数神经网络检测器（Scratch-EfficientDet，SED），以实现多尺度和多场景的 SAR 船舶检测。该方法通过使用高效的 EfficientNet-D0 作为主干网络提取特征，在相较于先前的经典网络，在提升保证性能的同时大大降低了模型参数、算力消耗和训练时长，在测试集上进行测试的速度达到每秒 20 张图像，基本满足实时性的要求，同时对数据进行归一化操作从而增加训练过程中梯度传输的稳定性以实现脱离预训练模型参数的收敛，最终在 SAR 船舶检测数据集上取得了最优结果。

1 相关工作

单阶段检测器：现有的目标检测器根据它们是否使用感兴趣区域候选步骤（region-of-interest proposal step）分为两阶段法[11,19,20]（使用）和单阶段法[10,12,21]（不使用）。虽然普遍认为两阶段法具有更高的精度，但单阶段法通过使用预设的锚框（bounding box）可以更高效简单。本文遵循单阶段

检测器设计，方便针对任务调整结构，提升精度。

Scratch 训练：文献[14]使用深度连接网络的监督机制，首先实现了从头开始训练并取得了与有预训练网络接近的性能，但无法随意调整主干网络以克服跨领域不适用性。ScratchDet（Training ASingle-Shot Object Detector from Scratch，从头训练的检测器）[15]分析了下采样步长大小对小目标检测精度的影响，推迟下采样操作的同时使用批归一化（BatchNormalization，BN）来稳定训练过程中的梯度传播。但 BN 操作需要使用大的小批量尺寸（mini-batch size），对实验环境要求较高，在本文中，我们探讨更简单高效的梯度稳定方法来避免这一缺点同时优化从头开始训练过程。

多尺度特征融合：不论是自然场景，还是遥感 SAR 图像，目标检测的难点之一就是如何更有效地处理多尺度特征。较早的检测器只是简单的使用骨干网络提取的金字塔特征层[10]甚至只是最后一层[19]进行类别和位置预测。FPN[16]首次使用自上而下的途径组合多尺度特征，PANet（Path Aggregation Network，路径汇聚网络）[23]在其基础上，额外增加了一条自底向上的路径来进一步融合特征。随着自动机器学习（Auto Machine Learning，AutoML）的发展，NAS-FPN（Neural Architecture Search- Feature Pyramid Networks，神经网络架构搜索-特征金字塔网络）使用神经架构搜索自动设计了特征网络拓扑结构。虽然 NAS-FPN 具有优异的性能，但其消耗了大量算力，且生成的 FPN 网络不规律。EfficientDet 以更直观，可解释的方式优化了多尺度特征融合。

2　方法

2.1 组归一化

BN 在文献[25]中首次出现，目的是解决如下两个问题。一是在深度神经网络训练的过程中，每个批次具有不同的分布，增加了模型训练的难度。二是内部变量转换问题（Internal Covariate Shift，ICS)：在训练的过程中，激活函数会改变各层数据的分布，随着网络的加深，这种差异会越来越大，从而出现梯度弥散，使模型难以收敛。Zhu 等人[15]通过在 SSD 的主干网络和检测子网络分别添加 BN 进行从头开始训练的实验证实了 BN 可以在优化过程中大幅缓解梯度的波动幅度，从而保证了更大的学习率和更快的收敛。但 BN 的问题也是显而易见的，由于要在批次中获得更普适的均值和方差，它需要足够大的单卡批大小，例如 ScratchDet 使用了 128 的批大小。这需要较高的硬件条件才能实现，所以我们尝试使用组归一化（Group Normalization，GN）[26]。

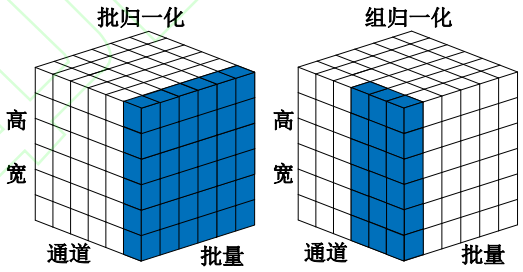


图 1. 数据归一化的方式。立方体表示特征图张量，蓝色方块表示用于归一化的量
Fig.1 Methods of data normalization. Cube represent feature map tensors, and blue squares represent tensors used for normalization

神经网络的输入数据通常有四个维度，B（batch）、C（channel）、H（height）、W（width）。训练过程中，显存能储存的数据量即批大小是有限的，在图像处理任务中可能是个位数。为了解决这一缺陷，GN 通过在通道维度进行数据的归一化，将通道分为几组后，在组内计算均值和方差以进行归一化，如图 1。

GN 在计算均值和标准差时，将每个特征图的通道维度分为 G 组，则每组中有 C/G 个通道，再对属于细分通道的像素求均值和标准差。每组通道独立地使用与其对应的参数进行归一化，所以 GN 的运算不受批大小的影响，并且精度比 BN 更加稳定。GN 的推导过程如下：

$$\hat{\boldsymbol{x}}_i = \frac{1}{\sigma_i}(\boldsymbol{x}_i - \mu_i) \tag{1}$$

式中，$\boldsymbol{x}$ 为特征图计算的张量，i 为索引号，即 $\boldsymbol{x}_i = (\boldsymbol{x}_{iB}, \boldsymbol{x}_{iC}, \boldsymbol{x}_{iH}, \boldsymbol{x}_{iW})$，$\hat{\boldsymbol{x}}_i$ 为经过归一化处理的张量。式（1）中的 μ 和 σ 分别是均值和标准差，计算方法如下：

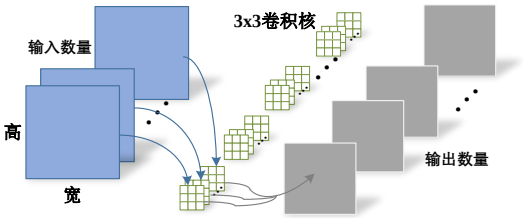
$$\mu_i(\boldsymbol{x}) = \frac{1}{(C/G)HW}\sum_{c=gC/G}^{(g+1)C/G}\sum_{h=1}^{H}\sum_{w=1}^{W}\boldsymbol{x}_{bchw} \tag{2}$$

$$\sigma_i(\boldsymbol{x}) = \sqrt{\frac{\sum_{c=gC/G}^{(g+1)C/G}\sum_{h=1}^{H}\sum_{w=1}^{W}\left(\boldsymbol{x}_{bchw} - \mu_i(\boldsymbol{x})\right)^2 + \varepsilon}{(C/G)HW}} \tag{3}$$

式中，ε 为一个极小的常数，保证 $\sigma_i \geq 0$，b,c,g,h,w 为索引号，B,C,G,H,W 为取值范围，G 是人为设定的分组数量，C/G 是每组的通道数。GN 由于并不在批次维度进行归一化，所以在小批次训练时的表现比 BN 更加优秀[26]。当想要使用小批次数据实现模型的从头开始训练时，GN 作为稳定优化过程中梯度变化，防止后向传播时发生梯度弥散的手段起着重要的作用。BN 与 GN 的应用对比将在本文 3.2 节以消融实验的形式呈现。

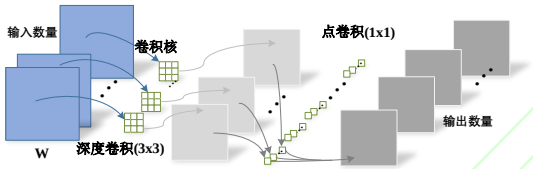
2.2 特征提取

遥感 SAR 图像中的船只绝大多数是相对尺寸（$\sqrt{w_{bbox}\times h_{bbox}}/\sqrt{w_{img}\times h_{img}}$，$w_{bbox}$、$h_{bbox}$ 表示框的宽和高，w_{img}、h_{img} 表示图像的宽和高）小于 0.2 的小目标对象[9]，这就要求作为特征提取部分的主干网络具有更强的提取能力，通常的做法是使用更复杂的模型[17]。而为了避免梯度弥散的发生，保证从头开始训练，同时减少下采样次数，尽可能增大深层特征图的感受野，提高小目标检测能力，又要求网络尽可能简化。这本身是相互矛盾的，但深度可分离卷积（depthwise separable convolution，DWConv）[27]和倒残差模块（Inverted Residuals，IRes）[28]的出现，一定程度上实现了两者兼具。DWConv 将普通卷积每个卷积核与每张特征图按位相乘再相加的步骤分离进行，先在通道维度进行按位相乘的卷积运算（depthwise conv，DWConv），此时 channel 不变，再用 1*1 卷积与第一步的结果进行卷积（pointwise conv，PW），见图 2。通过调整1×1 卷积个数改变通道，从而使得 DWConv 计算量约为普通卷积的 $1/(kernel\ size)^2$，损失精度仅为 1%。



(a)　普通卷积

(a)　Common convolution



(b)　深度可分离卷积

(b)　Depthwise separable convolution

图 2. 两种卷积对比

Fig.2 Comparison of two convolutions

残差模块[17]可以有效地复用之前的数据特征，如图 3，其输入量经过 1×1 卷积压缩，再使用 3×3 卷积提取特征，最后用 1×1 卷积将通道数增大，同时将输入与输出再次相加，形成如同沙漏的“压缩-卷积-扩张”的数据流图，从而使得卷积层集中精力学习输入、输出之间的残差。文献[28]直接将 DWConv 应用到残差块中并不能提升性能，原因是 DWConv 的特征提取能力受限于输入的通道数量，而倒残差模块的数据流图是“扩张-卷积-压缩”，类似纺锤的形状，在卷积操作前先进行扩张，保证了特征提取能力。

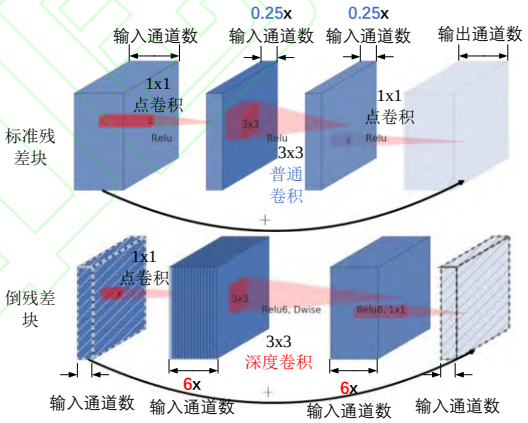


图 3. 残差模块和倒残差模块数据流图对比

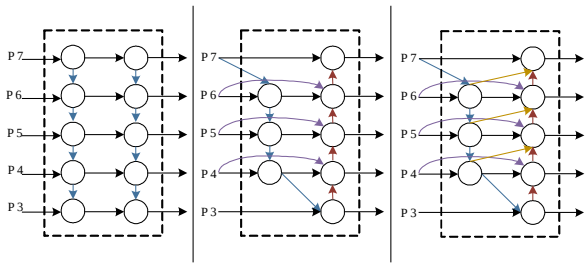
Fig.3 Comparison of data flow graph between residual blocks and inverted residuals blocks

2.3　特征融合

特征融合将不同分辨率的特征图信息充分利用，可以实现对多尺度目标较好的检测结果。网络自底向上的前向传播过程中，随着下采样次数的不断增加，语义信息越来越多，而位置信息逐渐减少。尽管更深的特征图拥有更多的语义信息，但是它们的分辨率较低，经过 5 次下采样后，原始图像中 32×32 像素的物体只有 1×1 大小，所以更深的特征图对小尺寸目标检测精度低。

EfficientDet 使用的 BiFPN（Efficient Bidirectional Cross-scale Connections and Weighted Feature Fusion，高效的双向跨尺度连接和加权特征融合），如图 4b，使用以下技巧来提高性能：（1）拥有

自顶向下和自底向上两条路径融合特征；（2）忽略只有一个输入的节点并加入跳跃连接以轻量化网络；（3）可学习权重自动加权融合过程的输入特征。我们为了更加充分地利用不同层级的语义和位置信息，在 BiFPN 的基础上改进（如图 4c），增加了跨级的数据流，实验证明这提升了网络性能。



(a)FPN (b)BiFPN (c)改进 BiFPN

图 4. 特征融合网络设计（ P_i 表示主干网络中分辨率为输入图像($1/2^i$)的特征图）

Fig.4 The design of feature fusion network(P_i represents a feature map with a resolution of the input image ($1/2^i$) in the backbone network)

本文选择倒残差模块作为主干网络基础结构，运用数据归一化手段对优化过程中梯度稳定性进行优化，降低主干网络下采样次数，使网络可以脱离预训练，最终实现端到端的目标检测，模型如图 5。

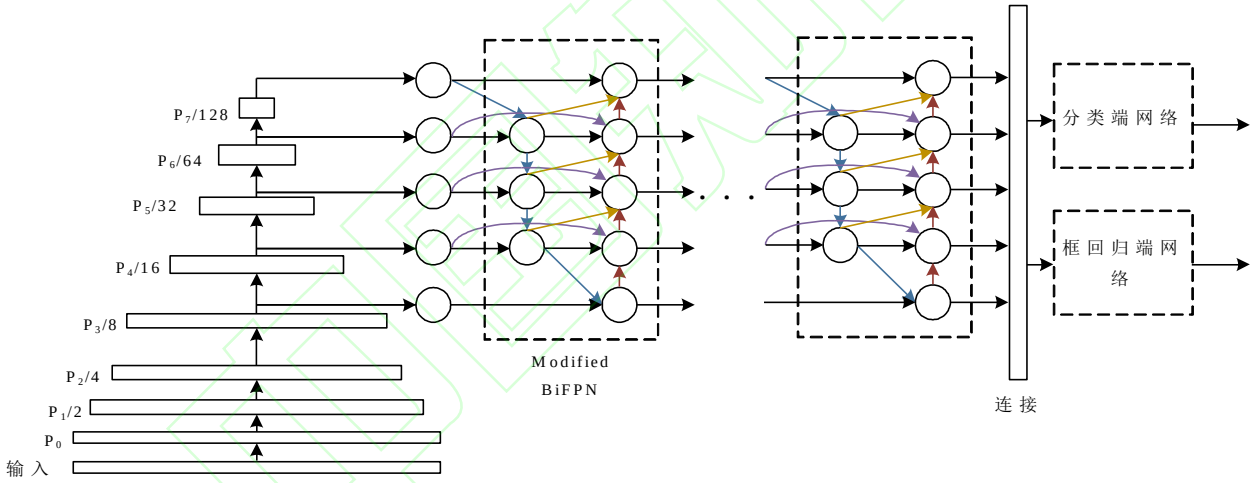


图 5. 网络结构示意图

Fig.5 Schematic diagram of the network structure

3 实验与分析

3.1 实验准备

3.1.1 数据集

本文中使用的数据主要来源于中国的高分 3 号 SAR 卫星和欧洲航天局的 Sentinel-1 SAR 卫星，总共使用了 102 张高分 3 号图像和 108 张 Sentinel-1 图像[9]。数据集包括 43,819 张切割后的船舶图像。高分 3 号的成像模式包括超细带状图（Ultrafine Strip-Map，UFS），精细带状图 1（Fine Strip-Map 1，FSI），全极化 1（Full Polarization 1，QPSI），精细带状图 2（Fine Strip-Map 2，FSII）和全极化 2（Full Polarization 2，QPSII），它们的分辨率分别为 3m，5m，8m，10m 和 25m。Sentinel-1 的成像模式是分辨率从 1.7×4.3 到 3.6×4.9m 的宽视场成像的 S3 Strip-Map（SM），S6 以及分辨率为 22m 的干涉测量宽幅模式（interference wide，IW）。船舶目标数据及其标注示例如图 6 所示，参照

MS COCO（Microsoft COCO: Common Objects in Context，微软语义场景通用目标）[29]格式制作数据集，并划分 70%为训练集，20%为验证集，10%为测试集。

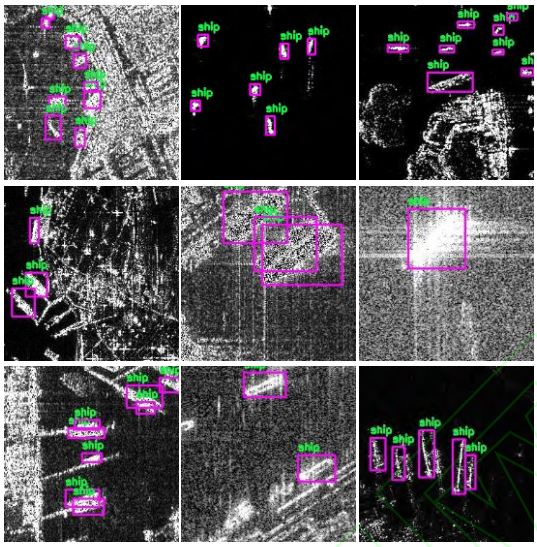


图 6. 复杂背景下的船舶数据集可视化
Fig.6 Visualization of a ship dataset in complex background

3.1.2 网络训练

实验平台系统是 Ubuntu 18.04，图形处理单元（GPU）是 NVIDIA RTX 2080Ti，深度学习框架是 pytorch。为了训练收敛，采用 Adam Optimizer（亚当优化器）梯度下降方法[30]，它使用了梯度的一阶矩估计和二阶矩估计来自适应地调整每个参数的学习步长。Adam 衰减系数分别为 0.9 和 0.999。若无特殊说明，图像输入大小均为 512x512，批大小均为 12。在主干网络和子网络同时设置 BN 或 GN，位置为 DWConv 操作后，激活函数前。损失函数使用聚焦损失函数（Focal Loss）[12]。

3.1.3 评价指标

船舶检测需要返回目标框位置信息和二分类置信度。评估指标主要有准确率 P （Precision）、召回率 R （Recall）和 AP 值，定义如下：

$$\mathrm{P} = \frac{TP}{TP + FP} \tag{4}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{5}$$

$$\mathrm{AP} = \int_0^1 \mathrm{P}(\mathrm{R})d\mathrm{R} \tag{6}$$

式中，TP 为指对象本来为正例，网络识别为正例，FP 为指对象本来为负例，网络识别为正例，FN 为对象本来为正例，网络识别为负例。因此 $TP + FP$ 是检测到的所有船舶数量，$TP + FN$ 表示实际船舶总数。由于实际中 P 关于 R 的函数是离散的，不同的预测框和标注框的交并比（Intersection over Union，IoU）下存在不同的 P 和 R ，所以 AP 在计算时由 IoU 划分包络曲线。$AP_{0.5}$ 表示当预测框和标注框的 IoU ≥ 0.5 时认为识别正确计算此时的 AP 值，$AP_{0.5:0.95}$ 表示预测框和标注框的 IoU 从 0.5 到 0.95，间隔 0.05 取值，在此 IoU 取值下，认为识别正确，计算 AP 并取平均值。$AP_{0.5:0.95}$ 比 $AP_{0.5}$ 更严格。

表1 消融实验
Table 1 Ablation experiment

组成	Lr=1e-3					
Pre-training	√					
BN	√	√	√			
GN				√	√	
BiFPN	√	√		√		√
Modified-BiFPN			√		√	
$AP_{0.5}$	92.3	93.4	93.5	93.7	94.2	Nan
$AP_{0.5:0.95}$	60.0	59.9	60.6	63.3	64.7	Nan

3.2 结果与分析

3.2.1 消融实验

如前所述，SED 具有 GN 和 Modified-BiFPN 两个核心构件，为了评估这两个组件对性能提升的作用，如表 1 开展消融实验，最左侧的 baseline 为 EfficientDet-D0 基于预训练参数迁移学习得到的结果。可以看到右侧没有 BN 或 GN 的实验条件下，模型从头开始训练无法收敛，而训练过程中有归一化优化数据操作，即存在 BN 或者 GN 可以使以 EfficientNet 作为 backbone 的模型实现脱离预训练的收敛。

由于实验硬件的限制，每个小批量尺寸最大只有 12，BN 的效果被削弱，而在 channel 维度进行归一化的 GN 操作则没有这些限制，并同样能够优化梯度传播。在表 1 中可以看到使用 GN 相比于 BN 在 $AP_{0.5}$ 指标下提升 0.3 和 0.7（2、4 列比较，3、5 列比较,），在 $AP_{0.5:0.95}$ 指标下提升 3.4 和 4.1（2、4 列比较，3、5 列比较）。在其他条件相同的情况下，本文提出 Modified-BiFPN 组件，通过增加 FPN 的数据路径，在 $AP_{0.5}$ 指标下提升幅度较小，为 0.1 和 0.5（2、3 列，4、5 列比较），在更严苛的 $AP_{0.5:0.95}$ 指标下提点 0.7 和 1.4（2、3 列，4、5 列比较），证明了更好的融合不同层级之间的数据对性能的提升。

3.2.2 对比实验

为了更好地评估 SED 的性能，使用了五个模型，即 SSD、Faster RCNN、RetinaNet、EfficientDet-D0 和 EfficienDet-D4 开展对比实验。为保证结果公平，均在本实验平台重新进行训练，除本文提出的 SED 模型外，其他均采用有预训练模型的迁移学习进行训练，受限于平台硬件性能，即使将训练时的批大小缩小至 1 依然无法训练 EfficientDet 模型系列中最优的 D7，所以采用能够训练的 D4 模型作为比较对象。表 2 所示为五种经典模型和本文提出的 SED 实验结果对比，从表 2 的数据中可以得出，SED 模型大小只有 15.4MB，是所有对比模型中最小的，虽然训练时长、测试时长、每秒处理图像数量（FPS）各个单项并不是最优，但综合各项指标，尤其是代表精度性能，最重要的 $AP_{0.5}$ 和 $AP_{0.5:0.95}$ 值方面取得了最佳成绩。图 7 中不同模型对相同图像的检测结果可视化对比也证实了本文所提的 SED 在多场景的情况下取得了更好的结果。

SSD512 和 SSD300 的不同是输入尺寸的不同，由表 2 可以看到越大的输入尺寸，结果越好，这是因为更多的小目标在大图像中经过多次下采样后在深层特征图中依然可以提供足够的位置信息。Faster RCNN 和 SSD 主要用于评估 FPN 的语义多尺度特征对性能的影响。基于 ResNet50 迁移学习的 Faster RCNN（双阶段法）和 RetinaNet（单阶段法）的结果对比用来验证单阶段法在 SAR 图像类中更具优势。RetinaNet 和 SSD 对比，评估 Focal Loss 损失函数对结果的影响。EfficientDet 作为最新最高效、精度最高的目标检测器，算法 D0 的 AP 值达到了 92.3%，而 D4 的 AP 值达到了 93.4%。但是经过对训练损失函数的分析，发现 D4 由于批大小只有 1，在训练过程中损失值下降极不平滑，波动性极大，在脱离预训练参数的条件下难以收敛，所以决定通过第 2 节所述的技巧在 D0 的基础上优化，检测精度取得了进一步的提升。这一方面原因是成功脱离预训练进行实验，对多场景检测结果的优化，因为没有采用自然场景下的初始化参数，可以更大程度地避免 SAR 图像成像机制所产生的固有缺陷对检测结果的影响，令模型更好地学习 SAR 图像特征，在复杂的多场景下，尤其是陆海交界处，获得更好的结果；另一方面是 Modified-BiFPN 对小目标检测做出的贡献，通过增加不同层级之

间的数据流通管道，将浅层的、精细的位置信息和深层的、粗糙的语义信息更好地融合，在小目标占比高的 SAR 图像船舶数据中，获得了更好的性能提升。

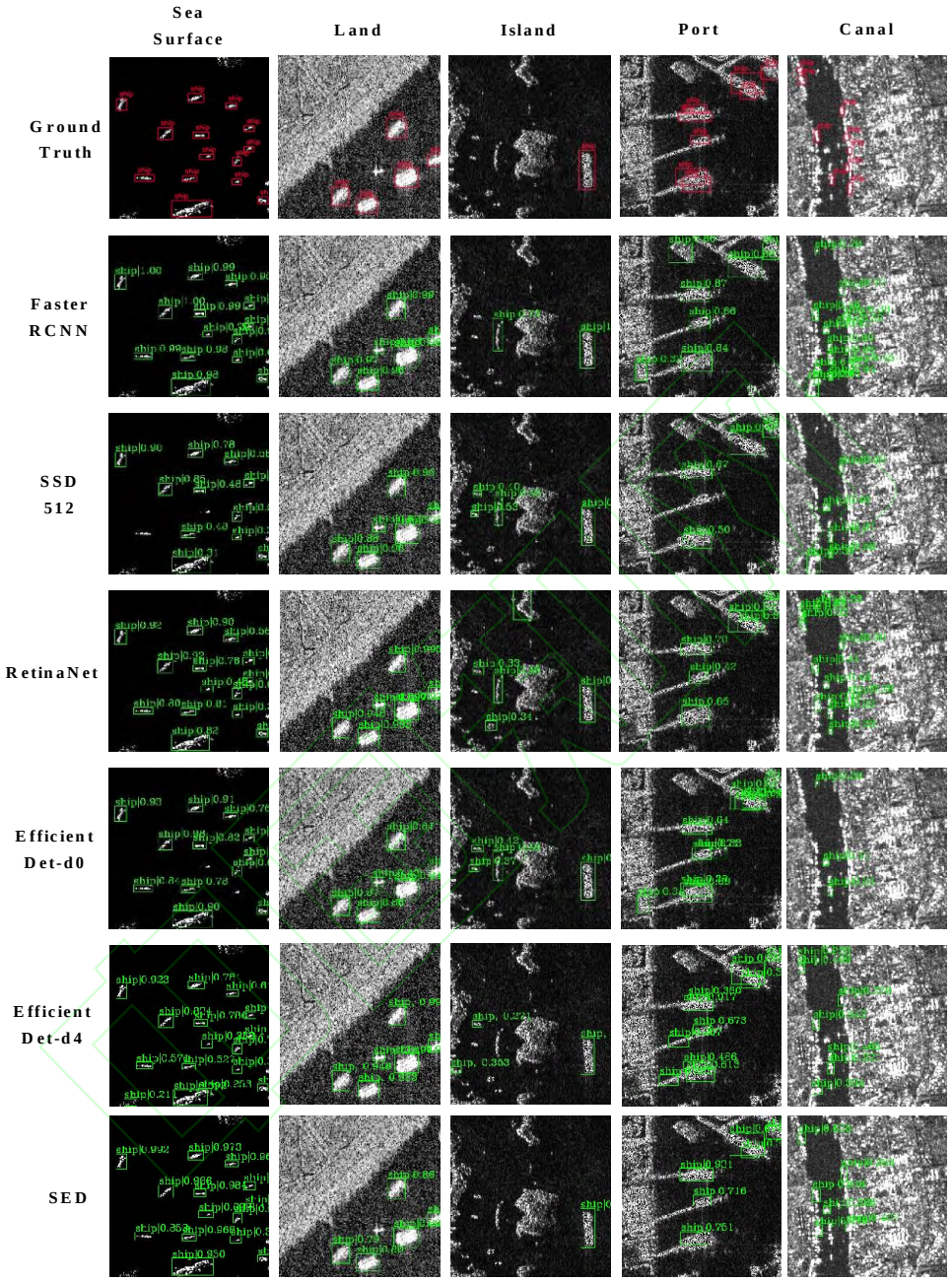


图 7. 不同模型预测结果可视化
Fig.7 Visualize prediction results of different models

由表 2 下面的 4 行数据可知，由于 SSD300 的输入图像大小为 300×300，在训练时间、测试时长和 FPS（每秒检测图像张数）方面取得了最优解，但在输入图像大小均为 512×512 的其他模型中，EfficientDet-D0 的训练时长最短，FPS 达到 30.5，SED 是在 D0 的基础上进行改进，由于添加了 GN 和从头开始训练的原因，训练时长稍有增加，FPS 也下降到 19.3，但基本满足实时性的要求，同时模型大小最小，便于在一些有算力限制的场景下部署。

表2 不同模型结果对比

Table 2 Compare results of different models

模型名称	SSD 300	SSD 512	Faster RCNN (R50)	RetinaNet (R50)	EfficientDet -D0(pre-training)	EfficientDet -D4(pre-training)	SED
$AP_{0.5}$(%)	88.5	89.6	91.8	92.9	92.3	93.4	94.2
$AP_{0.5:0.95}$(%)	49.1	51.4	54.9	57.1	60.0	62.7	64.7
训练时长(分钟)	10	43	23	15	14	195	19
测试时长(秒)	77	126	114	115	144	326	227
FPS(帧/秒)	56.6	36.3	38.6	38.0	30.5	13.5	19.3
大小(MB)	190.0	195.0	247.6	303.2	15.7	83.2	15.4

在图 7 中显示了在不同复杂场景中获得的检测结果。当背景是远离陆地的海面时，所有模型都表现出了有效的检测精度。当船舶接近陆地，岛屿和港口时，尽管模型仍然能够检测到船舶，但除了 SED 外，其他模型均存在不同程度的误报或漏报情况。漏报的发生主要是因为 SSD 虽然采用了多层卷积层提取的特征，但只是简单地叠加信息，而 Faster RCNN 和 RetinaNet 虽然采用了 FPN 结构融合特征，但信息复用不够充分，EfficientDet 的检测结果虽然已经很好，但由于依赖了分类模型的预训练模型参数，位置回归不够准确。本文提出的 SED 不管是在分类准确度还是预测框回归上均取得了最佳效果。

4 结论

在公开的 SAR 图像船舶检测数据集上的实验证明了我们提出的 SED 模型在多尺度和多场景 SAR 船舶检测中的有效性。使用 GN 作为梯度优化手段可以在较小的迷你批大小的条件下实现脱离预训练的模型收敛，一定程度上解决了跨领域不适用性的问题，从而实现多场景下的有效检测。Modified-BiFPN 用更典型简洁的结构将具有更多语义信息和更高分辨率的不同特征数据合并。低级的特征图适用于检测小型船舶，而高级的特征图适用于检测大型船舶，多次使用 Modified-BiFPN 使该模型更适合于多尺度 SAR 船舶检测。SED 在与其他模型的对比中，不仅检测精度达到了最优，模型大小也是最小的，训练难度低，虽然测试速度略有下降，但依然满足实时性的要求。在之后的研究中将进一步对网络进行优化，并考虑将生成式对抗网络加入检测器当中，增强网络的鲁棒性能，并且进一步提高检测精度。

参考文献（References）

[1] Kanjir U, Greidanus H, Environment K O J R S O. Vessel detection and classification from spaceborne optical images: A literature survey[J]. 2018, 207: 1-26.

[2] Wang Y, Wang C, Zhang H, et al. Automatic ship detection based on retinanet using multi-resolution gaofen-3 imagery[J]. 2019, 11(5): 531.

[3] El-Darymli K, Gill E W, Mcguire P, et al. Automatic target recognition in synthetic aperture tadar imagery: A State-of-the-Art Review[J]. IEEE access, 2016, 4: 6014-6058.

[4] Yang C-S, Park J-H, Harun-Al Rashid A J T J O N. An improved method of land masking for synthetic aperture radar-based ship detection[J]. 2018, 71(4): 788-804.

[5] Molina D E, Gleich D, Datcu M J I G, et al. Gibbs random field models for model-based despeckling of SAR images[J]. 2009, 7(1): 73-77.

[6] Xianxiang Qin, Shilin Zhou, Huanxin Zou, et al. A CFAR detection algorithm for generalized gamma distributed background in high-Resolution SAR images[J]. 2013, 10(4): 806-810.

[7] Zhao J, Zhang Z, Yu W, et al. A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images[J]. 2018, 6: 50693-50708.

[8] Li J, Qu C, Shao J. Ship detection in SAR images based on an improved faster R-CNN[C]. 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), 2017: 1-6.

[9] Wang Y, Wang C, Zhang H, et al. A SAR dataset of ship detection for deep learning under complex backgrounds[J]. 2019, 11(7): 765.

[10] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multiBox detector[C]. European Conference on Computer Vision, 2016, 21-37.

[11] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]. Advances in neural information processing systems, 2015: 91-99.

[12] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. 2017, PP(99): 2999-3007.

[13] Nogueira K, Penatti O A, Dos Santos J a J P R. Towards better exploiting convolutional neural networks for remote sensing scene classification[J]. 2017, 61: 539-556.

[14] Shen Z, Liu Z, Li J, et al. Dsod: Learning deeply supervised object detectors from scratch[C]. Proceedings of the IEEE international conference on computer vision, 2017: 1919-1927.

[15] Zhu R, Zhang S, Wang X, et al. ScratchDet: Training single-shot object detectors from scratch[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2019: 2268-2277.

[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] 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.
[18] Tan M, Pang R, Le Q V J a P A. Efficientdet: Scalable and efficient object detection[J]. arXiv preprint arXiv:1911.09070, 2019.
[19] Girshick R. Fast r-cnn[C]. Proceedings of the IEEE international conference on computer vision, 2015: 1440-1448.
[20] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]. Proceedings of the IEEE international conference on computer vision, 2017: 2961-2969.
[21] Redmon J, Farhadi A J a P A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
[22] Tian Z, Shen C, Chen H, et al. Fcos: Fully convolutional one-stage object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2019: 9627-9636.
[23] 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.
[24] Ghiasi G, Lin T-Y, Le Q V. Nas-fpn: Learning scalable feature pyramid architecture for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 7036-7045.
[25] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.
[26] Wu Y, He K. Group normalization[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
[27] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
[28] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 4510-4520.
[29] Chen X, Fang H, Lin T, et al. Microsoft COCO captions: data collection and evaluation Server[J]. arXiv preprint arXiv:1504.00325, 2015.
[30] Kingma D P, Ba J J a P A. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.