

基于改进SSD的合成孔径声纳图像感兴趣小目标检测方法

李宝奇^{1,2}, 黄海宁^{1,2,3*}, 刘纪元^{1,2}, 刘正君^{1,2}, 韦琳哲^{1,2,3}

(1. 中国科学院声学研究所, 北京 100190; 2. 中国科学院先进水下信息技术重点实验室, 北京 100190;
3. 中国科学院大学, 北京 100049)

摘要: 轻量化目标检测模型SSD-MV3(Single Shot Detection-MobileNet V3)因输入图像尺寸限制无法直接检测高分辨率大尺寸合成孔径声纳(Synthetic Aperture Sonar, SAS)图像中感兴趣小目标. 为此, 本文提出了一种新的目标检测方法HRSSD(High Resolution Single Shot Detection), 该方法通过冗余切割确保SSD-MV3输入图像尺寸的规范以及感兴趣小目标的完整, 并利用二次非极大值抑制保证检测结果的唯一. 此外, 提出了一种尺度、空间和通道注意力机制联合的特征提取模块, 并利用该模块重新设计了SSD-MV3的基础网络和附加特征提取网络, 记作SSD-MV3P(Single Shot Detection-MobileNet V3 Pro), 使得SSD-MV3P能更有效的感知感兴趣小目标特征信息. 实验结果表明, 在感兴趣小目标检测数据集SST(Sonar Small Targets)上, SSD-MV3P的平均检测精度(mean Average Precision, mAP)比SSD-MV3提升4.39%. HRSSD实现了高分辨率大尺寸SAS图像感兴趣小目标的检测, 并且保证了同一位置上检测结果的完整性和唯一性.

关键词: 合成孔径声纳; 感兴趣小目标检测; 轻量化目标检测模型; 注意力机制; 二次非极大值抑制

基金项目: 国家自然科学基金(No.11904386)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2024)03-0762-10

电子学报URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220925

Interested Small Target Detection Method Based on Improved SSD for Synthetic Aperture Sonar Image

LI Bao-qi^{1,2}, HUANG Hai-ning^{1,2,3*}, LIU Ji-yuan^{1,2}, LIU Zheng-jun^{1,2}, WEI Lin-zhe^{1,2,3}

(1. Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China;

2. Key Laboratory of Science and Technology on Advanced Underwater Acoustic Signal Processing,

Chinese Academy of Sciences, Beijing 100190, China; 3. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The efficient object detection model SSD-MV3 (Single Shot Detection-MobileNet V3) cannot directly detect the interested small targets in high-resolution SAS (Synthetic Aperture Sonar) images due to the input image size limit. To this end, this paper proposes a novel object detection method, HRSSD (High Resolution Single Shot Detection), which ensures the specification of SSD-MV3 input image size and the integrity of the interested small targets through redundant cutting algorithm, and guarantees the unique detection result by using secondary non-maximum suppression. Furthermore, an improved feature block with a combination of scale, space and channel attention mechanism is proposed, and the basic network and additional feature network of SSD-MV3 are redesigned as SSD-MV3P (Single Shot Detection-MobileNet V3 Pro). Thus, SSD-MV3P can more effectively perceive the feature information of interested small targets. The experimental results show that the mAP (mean Average Precision) of SSD-MV3P is 4.39% higher than that of SSD-MV3 on the interested small target detection dataset SST (Sonar Small Target). HRSSD realizes the detection of the interested small targets in high-resolution SAS images, and ensures the integrity and uniqueness of the detection result at the same location.

Key words: synthetic aperture sonar; interested small target detection; efficient object detection model; attention mechanism; secondary non maximum suppression

Foundation Item(s): National Natural Science Foundation of China (No.11904386)

1 引言

合成孔径声纳(Synthetic Aperture Sonar, SAS)通过对基阵多个回波信号进行相干处理,合成等效虚拟长孔径来获得海底声纳图像^[1].与侧扫声纳图像相比,SAS图像最为显著的优点是方位向分辨率较高,且理论分辨率与目标距离以及采用的声波频段无关^[2].不过,感兴趣小目标实际物理尺寸较小,导致其有效像素点少,在SAS图像中的相对像素比也较低,这给后续的目标检测和识别带来了挑战^[3,4].

传统的SAS图像感兴趣小目标检测^[5]先从SAS图像中提取感兴趣(Regions of Interest, ROI)目标“切片”^[6];接着,对切片内目标做分割^[7]处理,然后,进一步提取分割图像的特征;最后,利用分类器^[8]对提取特征进行分类识别.不过,切片选取算法鲁棒性差、目标分割和特征提取算法的设计十分困难且容易造成有用信息丢失、浅层分类器也容易达到性能瓶颈.以深度学习^[9,10]为核心的人工智能算法引起了全世界高度重视,它从根本上改变了人们处理复杂和不确定性问题的方法,与传统的机器学习和信号处理方法相比,深度学习模拟人类视觉神经系统的层次体系,含有更多的隐含单元层,通过对原始数据逐层的非线性变换,可以得到更高层次的、更加抽象的特征表达,高层次的表达能够强化输入数据的区分能力,同时削弱不相关因素的不利影响.因此,通过借鉴深度学习技术提高SAS图像感兴趣小目标的检测识别精度具有重要的理论和实践意义^[11-13].目前,基于卷积神经网络(Convolutional Neural Networks, CNN)的目标检测模型在一个框架内实现了目标检测和识别的统一.目标检测模型可分为基于候选区域和基于回归两类,其中基于候选区域的模型主要代表有R-CNN(Region-CNN)^[14]、Fast R-CNN(Fast Region-CNN)^[15]和Faster R-CNN(Faster Region-CNN)^[16]等,基于回归思想的模型主要代表有YOLO(You Only Look Once)^[17]和SSD(Single Shot Detection)^[18],SSD模型较好的兼顾了检测精度和检测速度.不过,两类目标检测模型输入图像的尺寸都相对较小,如基于候选区域的模型检测考虑Faster R-CNN原始图像输入大小是1 000×600像素,SSD原始图像输入大小是300×300像素或512×512像素;YOLO原始图片输入大小是416×416像素或544×544像素.因此,简单的下采样操作,无法保证上述目标检测模型对感兴趣小目标的检测效果.Adam^[19]提出了基于YOLO的高分辨率大尺寸卫星图像目标检测算法YOLT(You Only Look Twice),YOLT首先将高分辨率大尺寸卫星图像切割,然后用两个输入图像尺寸不同的YOLO模型分别检测图像中大尺寸目标和小尺寸目标,最后通过二次非极大值抑制算法Hard-NMS对检测结果做进一步筛选.

为了兼顾SSD目标的检测精度和速度,Howard等人提出了轻量化系列卷积神经网络MobileNet V1^[20]、MobileNet V2^[21,22]和MobileNet V3^[23],其中MobileNet V3的特征提取模块IRB⁺(Inverted Residual Block Plus)延续了MobileNet V1和MobileNet V2的轻量化设计,并在此基础上引入了注意力机制SE(Squeeze and Excitation)^[24]组件,SE组件通过对通道信息进行筛选来提高模块的特征提取能力.不过,IRB⁺捕获所有通道的依赖关系是低效并且是不必要的.Woo等^[25]提出了结合通道(Channel)和空间(Spatial)的注意力机制组件CBAM(Convolutional Block Attention Module),相比于SE-Net只关注通道注意力机制可以取得更好的结果.Wang等^[26]提出了跨通道交互的极轻量注意力组件ECA(Efficient Channel Attention).ECA去除了原来SE组件中的全连接层,直接在GAP(Global Average Pooling)之后的特征上通过一个可以权重共享的一维卷积进行学习.另外,Li等人^[27]提出了SKNet(Selective Kernel Net)分支注意力组件.SKNet针对多尺度输出设计了一个特征选择机制,即通过学习能保留最有价值的多尺度特征层,但计算复杂度较高.SAS图像像素高、感兴趣小目标像素点少.现有目标检测模型主要是为小尺寸图像设计,若将声纳图像缩放后直接送入目标检测模型,感兴趣小目标的有效像素点将降低甚至消失,无法实现感兴趣小目标的检测.另外,虽然轻量化目标检测模型较好的兼顾了计算成本与模型性能之间的平衡,但对于感兴趣小目标,其检测精度会存在一定程度的下降.

针对上述方法存在的问题,本文提出了一种高分辨大尺寸SAS图像感兴趣小目标检测的方法HRSSD(High Resolution SSD).HRSSD在研究感兴趣小目标成像特点和尺寸的基础上,采用冗余切割算法将原始SAS图像切割成适合目标检测模型输入的小尺寸图像,再利用轻量化目标检测模型对切割后的图像逐个检测,最后选用最大值抑制算法对检测结果进行确认,从而实现高分辨大尺寸SAS图像感兴趣小目标检测识别.在此基础上,结合分支注意力、空间注意力和通道注意力,提出了一种改进的注意力组件,嵌入到轻量化目标检测模型SSD-MV3的特征提取模块中,进一步提高了HRSSD对高分辨率大尺寸SAS图像感兴趣小目标的检测精度.

2 基于改进SSD的高分辨率大尺寸SAS图像感兴趣小目标检测方法

2.1 HRSSD算法

SAS图像分辨率高,感兴趣小目标相对像素比低,直接将原始SAS图像输入到SSD-MV3,无法保证感兴趣小目标的检测效果.HRSSD采用先切割、后检测、再

确认的策略对 SAS 图像感兴趣小目标进行检测识别。HRSSD 算法流程如图 1 所示,包括以下三步:首先对高分辨率大尺寸 SAS 图像进行逐行、逐列冗余切割;然后,利用改进的感兴趣小目标检测模型 SSD_MV3P 对图像切片分别进行检测识别;最后,选用 Soft-NMS^[28]对检测结果进行二次非极大值抑制,并将最终结果在高分辨率大尺寸 SAS 图像上显示。

假设高分辨率大尺寸 SAS 图像 I 的尺寸为 $N \times N$,经切割算法处理后的图像切片数学描述如下:

$$I \Rightarrow I_{r,c,h,s} \quad (1)$$

其中 $I_{r,c,h,s}$ 表示图像切片, r 表示 SAS 图像 I 中的列坐标, c 表示 SAS 图像 I 的中行坐标, h 表示图像切片的尺寸, s 表示图像切片的割切步长。为了保证感兴趣小目标在图像切片中的完整性,相邻图像切片在横向和纵向存在重叠,其中单边重叠比例为 20%(保证感兴趣小目标在图像切片中有完整的存在)。SAS 图像 I 四个顶点图像存在两边重叠,其余边沿图像是三边重叠,其余图像切片是四边重叠。

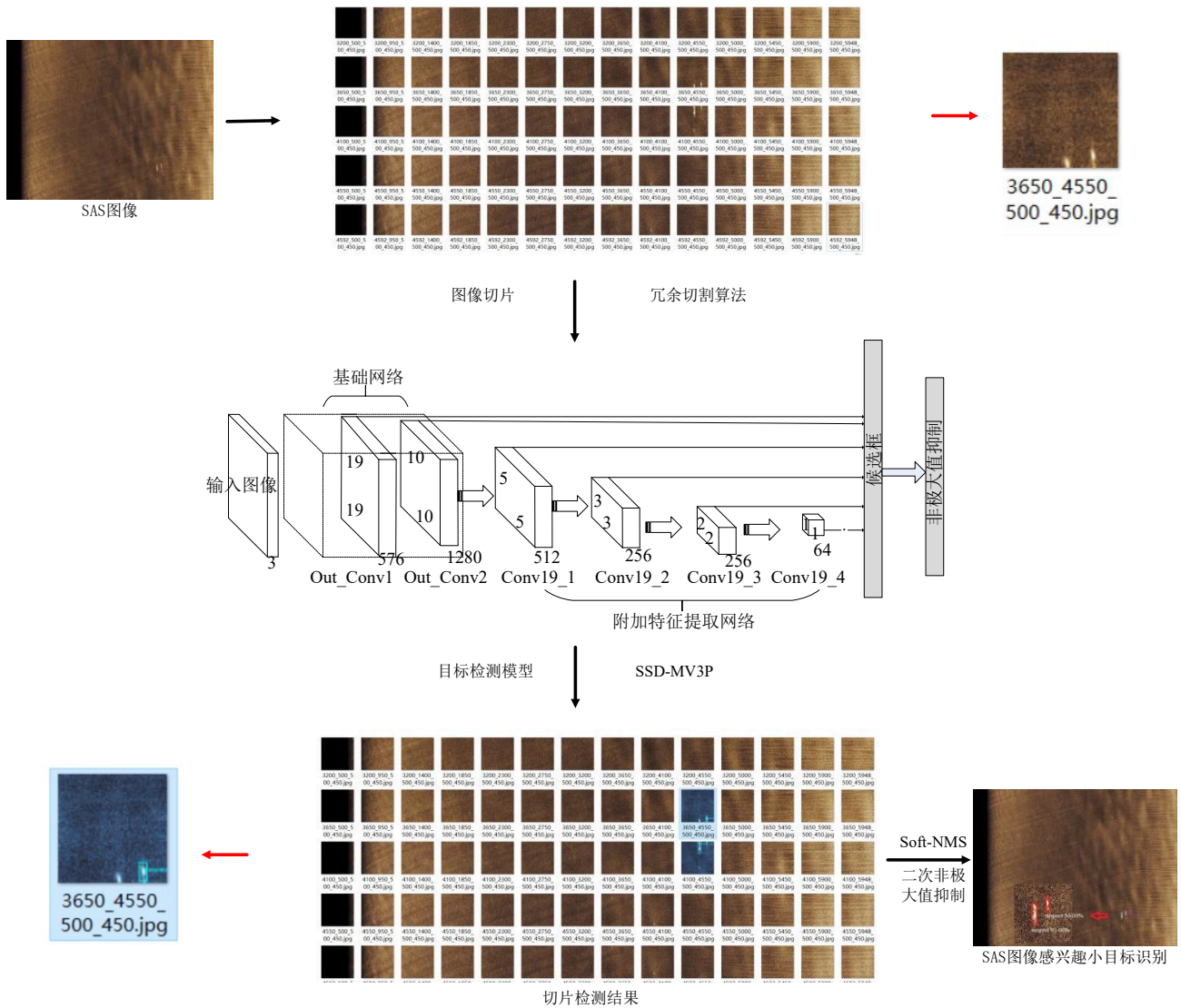


图 1 HRSSD 算法流程

图像切片 $I_{r,c,h,s}$ 经 SSD_MV3 生成感兴趣小目标检测结果,包括目标类型、概率和位置信息,记作:

$$I_{r,c,h,s} \Rightarrow B_{r,c,i} \quad (2)$$

其中, $B_{r,c,i}$ 表示图像切片的检测信息, r 表示 SAS 图像 I 中的列坐标, c 表示 SAS 图像 I 的中行坐标, i 表示目标的类型。

检测结果 $B_{r,c,i}$ 经非极大值抑制算法 Soft-NMS 剔除重叠建议框,最终得到每个类别中每个位置上回归修正后得分最高的包围框,记作:

$$\hat{B}_{r,c,i} = \begin{cases} B_{r,c,i}, & iou(S_i, b_{r,c,i}) < T \\ B_{r,c,i}(1 - iou(S_i, b_{r,c,i})), & iou(S_i, b_{r,c,i}) \geq T \end{cases} \quad (3)$$

其中, S 为当前循环中得分最高的候选框, $b_{r,c,i}$ 为当前位

置上的检测信息, T 为IoU阈值, $\widehat{B}_{r,c,i}$ 为二次非极大值抑制后的感兴趣小目标检测结果,并最终在SAS图像中显示.

2.2 IRB++模块

SSD-MV3的特征提取模块IRB⁺利用SE组件仅从通道维度对输入信息进行筛选,如图2(a)所示. IRB⁺⁺(Inverted Residual Block Plus and Plus)模块从多尺度、空间和通道三个维度对输入信息进行选择,如图2(b)所示,其中改进的注意力组件记作BSC(Branch Spatial

Channel Attention). IRB⁺⁺模块沿用IRB⁺模块先“扩张”后“压缩”的策略,并由扩张层、分割层、多尺度(Selective Dilated Kernel, SDK)组件、空间(Spatial Attention, SA)组件、拼接层、通道(Channel Attention, CA)组件和压缩层组成,其中扩张层负责输入特征通道扩张;分割层负责通道切割;多尺度组件负责输入信息的多尺度分解;空间组件负责选取多尺度层中有用的空间信息;拼接层负责通道整合;通道可选择组件通过学习权重选择包含重要信息的通道;压缩层负责将特征通道压缩成与输入特征一致的数量.

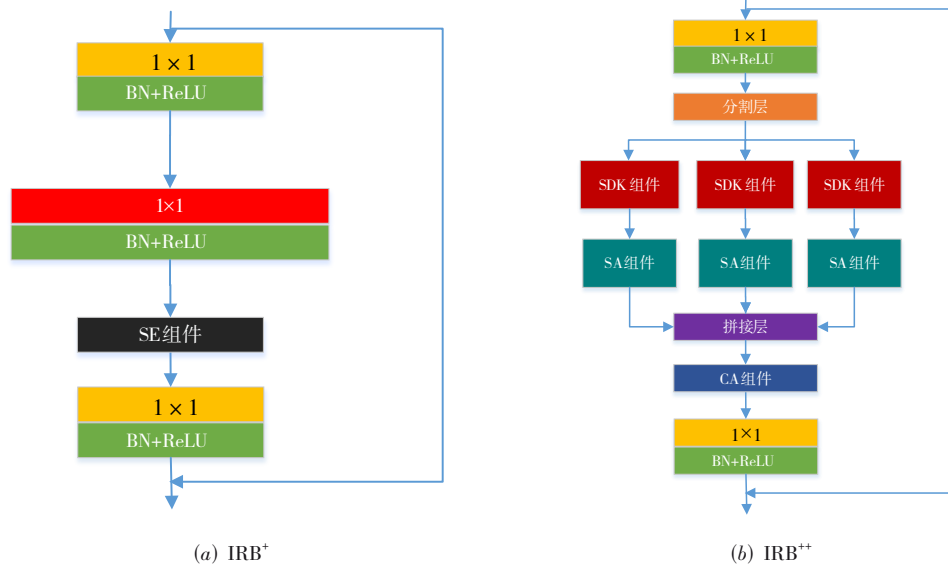


图2 基于改进注意力机制的特征提取模块

对于一个任意的输入特征 $\mathbf{D} \in \Phi^{H \times H \times M}$, 其中 $H \times H$ 为输入特征的尺寸, M 为输入特征的通道数. 输入特征 \mathbf{D} 进入IRB⁺⁺模块的两个支路网络: 左侧支路负责感兴趣小目标特征提取和选择; 右侧支路保持输入特征 \mathbf{D} 不变, 并最后与上侧支路网络的输出特征相加. 对于下侧支路网络, 输入特征 \mathbf{D} 首先经过扩张层, 其输出特征的数学表达式为:

$$\mathbf{D}_{\text{ex}} = F_{\text{ex}}(\mathbf{D}), \quad \mathbf{D} \in \Phi^{H \times H \times M} \quad (4)$$

其中, \mathbf{D} 为原始输入特征, \mathbf{D}_{ex} 为经过扩张层后的特征, F_{ex} 为通道扩张操作, 扩张层的卷积核尺寸为 1×1 , 输出通道数量为 M^+ .

接着, 输出特征 \mathbf{D}_{ex} 送入IRB⁺⁺分割层和多尺度组件, 其输出特征的数学表达式为:

$$\mathbf{D}_{\text{branches}}^k = F_{\text{branches}}(\mathbf{D}_{\text{ex}}^k), \quad k = 1, 2, 3, \dots, K \quad (5)$$

其中, \mathbf{D}_{ex}^k 为分割后的特征, k 表示第 k 个分支, 共 K 个分支, 具体是按通道均匀分割; $\mathbf{D}_{\text{branches}}^k$ 为多尺度组件输出的特征图, 特征图尺寸为 $H \times H$, 通道数为 M^+/K , F_{branches} 为多尺度深度可分离空洞卷积操作. 在多尺度组件中,

卷积核的实际覆盖范围与空洞率之间的关系如下:

$$R_{\text{rdf}} = R + (\kappa - 1)(R - 1) \quad (6)$$

其中, R_{rdf} 为该层卷积核实际覆盖范围, R 为该层卷积核尺寸, κ 为该层空洞率大小. 例如, 一个卷积核尺寸为 3×3 、空洞率为 2 的空洞卷积层^[29,30], 卷积核实际覆盖范围为 $R_{\text{rdf}} = 5$. 增大空洞率 κ 可以扩大卷积层的感受野, 即降低模型的计算成本和减少模型的参数, 本文 $\kappa = 1, 2$, 即 $R_{\text{rdf}} = 3, 5$.

然后, 输出特征 $\mathbf{D}_{\text{branches}}^k$ 送入空间选择组件, 其输出特征的数学表达式为:

$$\mathbf{s}_{\text{spatial}}^k = \sigma(f_{3 \times 3}(F_{\text{concat}}(\text{AvgPool}(\mathbf{D}_{\text{branches}}^k); \text{MaxPool}(\mathbf{D}_{\text{branches}}^k)))) \quad (7)$$

$$\mathbf{D}_{\text{spatial}}^k = \mathbf{s}_{\text{spatial}}^k \cdot \mathbf{D}_{\text{branches}}^k \quad (8)$$

其中, $\mathbf{s}_{\text{spatial}}^k$ 为第 k 个分支的空间权重, σ 为激活函数, $f_{3 \times 3}$ 为卷积核(尺寸为 3×3)的卷积操作, F_{concat} 为通道拼接操作, AvgPool 为平均池化操作, MaxPool 为最大值池化操作. $\mathbf{D}_{\text{spatial}}^k$ 为空间选择组件的输出, 输出特征图尺寸为 $H \times H$, 通道数为 M^+/K .

随后,利用拼接层对 K 个 $\mathbf{D}_{\text{spatial}}^k$ 在通道维度上拼接,其输出特征的数学表达式为:

$$\mathbf{D}_{\text{concat}} = F_{\text{concat}}(\mathbf{D}_{\text{spatial}}^k), k = 1, 2, 3, \dots, K \quad (9)$$

其中, $\mathbf{D}_{\text{concat}}$ 为拼接后的输出特征,输出特征图的尺寸为 $H \times H$,通道数为 M^+ ,与扩张后的通道数相等, F_{concat} 为通道拼接操作。

接着,将 $\mathbf{D}_{\text{concat}}$ 送入通道选择组件,其输出特征的数学表达式为:

$$\mathbf{s}_{\text{channel}} = \text{MLP}(\text{AvgPool}(\mathbf{D}_{\text{concat}})) + \text{MLP}(\text{MaxPool}(\mathbf{D}_{\text{concat}})) \quad (10)$$

$$\mathbf{D}_{\text{channel}} = \mathbf{s}_{\text{channel}} \cdot \mathbf{D}_{\text{concat}} \quad (11)$$

其中, $\mathbf{D}_{\text{channel}}$ 为通道选择组件的输出特征,输出特征图的尺寸为 $H \times H$,通道数为 M^+ , $\mathbf{s}_{\text{channel}}$ 为 $\mathbf{D}_{\text{concat}}$ 的通道权重系数,MLP 共享全连接网络,AvgPool 为平均池化操作,MaxPool 为最大值池化操作。

然后,对 $\mathbf{D}_{\text{channel}}$ 进行通道压缩,数学表达式为:

$$\mathbf{D}' = F_{\text{sq}}(\mathbf{D}_{\text{channel}}), \mathbf{D}' \in \Phi^{H \times H \times M} \quad (12)$$

其中, \mathbf{D}' 为通道压缩后的特征,特征图的尺寸为 $H \times H$,通道数为 M ,与输入特征 \mathbf{D} 的通道数相等; F_{sq} 为通道压缩操作,卷积核尺寸为 1×1 。

通过上面的计算,最后可以得到 IRB⁺⁺ 模块的输出特征数学表达式为:

$$\mathbf{D} = \mathbf{D} + \mathbf{D}' \quad (13)$$

其中, \mathbf{D} 为 IRB⁺⁺ 模块的输出特征, $\mathbf{D} \in \Phi^{H \times W \times M}$,特征图尺寸为 $H \times H$,通道数为 M 。

2.3 SSD-MV3P 模型

在 SSD-MV3 模型内,利用 IRB⁺⁺ 替换 IRB⁺ 模块,生成新的轻量化目标检测网络,记做 SSD-MV3P. SSD-MV3P 与 SSD-MV3 的网络结构保持一致,如图 3 所示,包括基础特征提取网络,附加特征提取网络,候选框生成和检测结果预测四个部分^[22],SSD-MV3P 主要网络结构如表 1 所示。

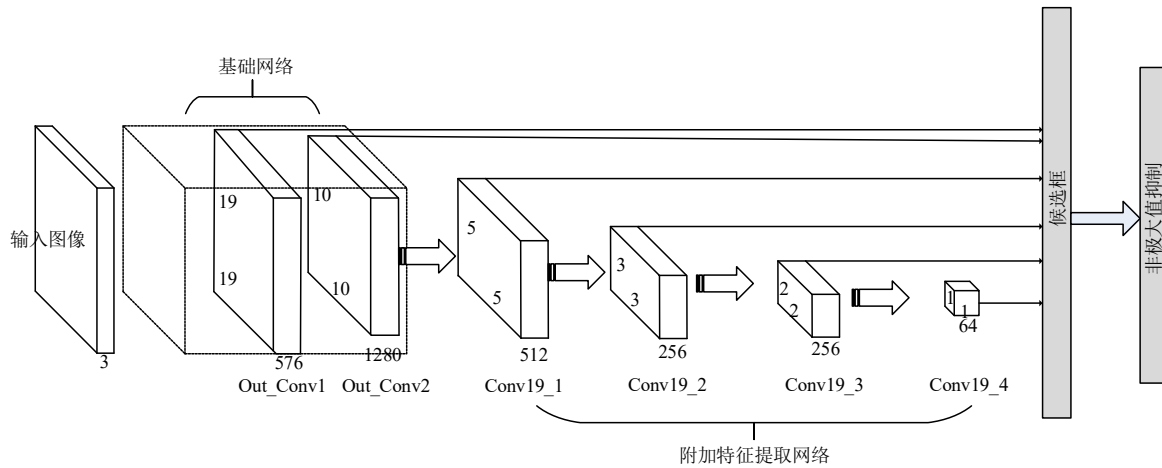


图3 SSD-MV3P 目标检测模型

表 1 中,SSD-MV3P 基础网络由前 14 个卷积层组成,附加特征提取网络由后 4 个卷积层组成,其中各卷积层输入通道、中间通道、输出通道、步长和激活函数(RE 表示 ReLU 激活函数,HS 表示 H-Swich^[24] 激活函数)与 SSD-MV3 保持一致;不同的是 SSD-MV3P 利用 IRB⁺⁺ 替换 SSD-MV3 中 IRB⁺ 模块,IRB⁺⁺ 模块对应的分支数量为 2、卷积核尺寸为 3×3 和 5×5 . 基础网络中 Out_Conv1 和 Out_Conv2,附加特征提取网络 Conv19_1、Conv19_2、Conv19_3 和 Conv19_4 用于候选框的生成,其中 Out_Conv1、Out_Conv2、Conv19_1、Conv19_2、Conv19_3 和 Conv19_4 的输入特征图尺寸为 19×19 、 10×10 、 5×5 、 3×3 、 2×2 和 1×1 . 候选框生成部分根据预先定义的尺度(scales)和纵横比(aspect ratios)从上述 6 个尺度的特征层中提取数量和大小不同的候选框;卷积预测部分则是对候选框内目标的类型和位置进行判

断,并利用非极大值抑制算法对候选框内目标进行优化。

3 仿真试验

为了验证 HRSSD 的有效性以及相关参数对其性能的影响,设计实验 1,以轻量化目标检测模型 SSD-MV3 为研究对象,比较分析不同注意力机制的性能差异,并验证 HRSSD 方法的有效性. 设计实验 2,以 IRB⁺⁺ 为研究对象,设计消融实验比较分析多尺度注意力、空间注意力和通道注意力对 BSC 组件的作用. 设计实验 3,以注意力机制 IRB⁺⁺ 为研究对象,比较分析多尺度组件不同的多分支数量对检测精度的影响. 设计实验 4,以 HRSSD 为研究对象,比较分析二次非极大值抑制算法对 HRSSD 性能的影响. 实验平台操作系统为 RedHat 7.5、环境管理软件为 Anaconda3、深度学习框架为

表 1 SSD-MV3P主要网络结构

网络	编号	操作	输入通道	中间通道	输出通道	分支数量(卷积核)	步长	激活函数
基础网络	Conv1	conv2d	3	—	16	1(3×3)	2	HS
	Conv2	IRB ⁺⁺	16	16	16	2(3×3、5×5)	2	RE
	Conv3	IRB ⁺⁺	16	72	24	2(3×3、5×5)	2	RE
	Conv4	IRB ⁺⁺	24	88	24	2(3×3、5×5)	1	RE
	Conv5	IRB ⁺⁺	24	96	40	2(3×3、5×5)	2	RE
	Conv6	IRB ⁺⁺	40	240	40	2(3×3、5×5)	1	RE
	Conv7	IRB ⁺⁺	40	240	40	2(3×3、5×5)	1	RE
	Conv8	IRB ⁺⁺	40	120	48	2(3×3、5×5)	1	HS
	Conv9	IRB ⁺⁺	48	144	48	2(3×3、5×5)	1	HS
	Conv10	IRB ⁺⁺	48	288	96	2(3×3、5×5)	2	HS
	Conv11	IRB ⁺⁺	96	576	96	2(3×3、5×5)	1	HS
	Conv12	IRB ⁺⁺	96	576	96	2(3×3、5×5)	1	HS
	Out_Conv1	conv2d	96	—	576	1(1×1)	2	HS
	Out_Conv2	conv2d	576	—	1280	1(1×1)	2	HS
附加网络	Conv19_1	IRB ⁺⁺	1 280	512	256	2(3×3、5×5)	2	HS
	Conv19_2	IRB ⁺⁺	512	128	256	2(3×3、5×5)	2	HS
	Conv19_3	IRB ⁺⁺	256	256	128	2(3×3、5×5)	2	HS
	Conv19_4	IRB ⁺⁺	256	64	64	2(3×3、5×5)	2	HS

Torch1.3.1 和 Torchvision 0.4.2 等; BatchSize 等于 64、学习率等于 0.001、所有模型均在 GPU 上进行训练和测试.

3.1 数据集

为了检验 HRSSD 对高分辨率大尺寸 SAS 图像感兴趣小目标的检测识别性能, 建立了一个 SAS 感兴趣小目标数据集: SST (Sonar Small Targets). SST 包含三种类型的目标: 非目标、疑似物和圆柱物, 共计 2 490 幅图像, 其中训练集包含 2 221 幅图像, 测试集包含 269 幅图像. SST 组成如表 2 所示.

表 2 SST 数据集

目标	数量/幅
非目标	279
疑似物	1 923
圆柱物	288
总计	2 490

3.2 实验 1: 不同注意力机制对 SSD-MV3P 性能的影响

在 HRSSD 框架内, 本实验比较分析 SE、ECA、CBMA 和本文 BSC 注意力组件在数据集 SST 上的性能差异. SSD-MV3 的特征提取模块为 IRB⁺, 采用 SE 注意力组件; SSD-MV3E 的特征提取模块为 IRB_E, 采用 ECA 注意力组件; SSD-MV3C 的特征提取模块为 IRB_C, 采用 CBAM 注意力组件; SSD-MV3P 的特征提取模块为 IRB⁺⁺, 采用 BSC 注意力组件. 分别记录检测模型在迭代 1 000 次时对感兴趣小目标数据集 SST 的 mAP

数值、参数大小和平均检测时间. 实验结果如表 3 所示.

表 3 不同注意力机制对 SSD-MV3P 性能的影响

模型	注意力组件	mAP/%	Parameters/MB	Time/ms
SSD-MV3	SE	83.27	10.9	24.02
SSD-MV3E	ECA	85.31	8.4	22.89
SSD-MV3C	CBAM	85.15	9.8	34.54
SSD-MV3P	BSC	87.66	15.5	47.83

从表 3 可以发现, SSD-MV3P 的检测精度比 SSD-MV3、SSD-MV3E 和 SSD-MV3C 分别高 4.39%、2.35% 和 2.51%, 其中 SSD-MV3E 和 SSD-MV3C 两个模型的检测精度接近; 模型参数比 SSD-MV3、SSD-MV3E 和 SSD-MV3C 分别增加 4.6 MB、7.1 MB 和 5.7 MB; 检测时间比 SSD-MV3、SSD-MV3E 和 SSD-MV3C 分别增加 23.81 ms、24.94 ms 和 13.29 ms. 鉴于目标模型和运算时间运行在相对较低水平, 可以满足 SAS 图像感兴趣小目标检测的实时性要求, SSD-MV3P 更适合 SAS 图像感兴趣小目标检测识别.

为了验证 HRSSD 的有效性, 利用 HRSSD 对三个场景下的高分辨率大尺寸 SAS 图像感兴趣小目标进行检测识别, 检测结果如图 4 所示.

从图 4 的检测结果可以看出, 基于 SSD-MV3P 的 HRSSD 算法对感兴趣小目标能实现准确检测识别, 并且检测结果完整、无重复.

3.3 实验 2: 消融实验

本实验比较 IRB⁺⁺ 模块中多尺度组件 (SDK)、空间

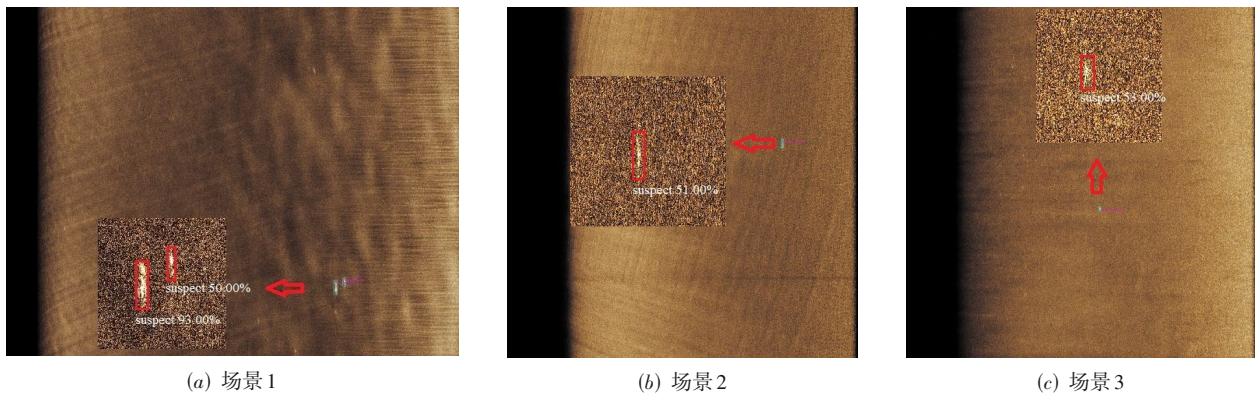


图4 高分辨率大尺寸SAS图像感兴趣小目标检测效果图

选择组件(SA)和通道选择组件(CA)对SSD-MV3P性能影响,以SSD-MV3为参考,将SDK、SA和CA三个注意力组件替换图2(a)中的SE组件形成三个新的模型,分别记录三个检测模型在迭代1000次时对感兴趣小目标数据集SST的mAP数值、参数大小和平均检测时间.实验结果如表4所示.

表4 MECA不同数量多尺度卷积核对SSD-MV3P性能的影响

注意力组件	mAP/%	Parameters/MB	Time/ms
SE	83.27	10.9	24.02
SDK	85.81	14.2	27.28
SA	84.06	8.4	24.18
CA	84.64	9.8	29.62

从表4可以发现,SDK组件、SA组件和CA组件的检测精度比SE组件分别提升2.54%、0.79%和1.37%;SDK组件比SE组件模型参数增加3.3MB,SA组件和CA组件模型参数比SE组件分别减少2.5MB和1.1MB;SDK、SA和CA三个注意力组件比SE组件的检测时间分别增加3.26ms、0.16ms和5.6ms.从实验结果可知,SDK、SA和CA比SE在检测精度上均有提升,尤其是SDK.主要原因是SDK组件是多分支和通道注意力的组合,SA和CA利用最大值池化和平均值池化两种信息获取空间和通道的权重,SE组件仅利用平均值池化信息,因此在信息输入上SDK、SA和CA比SE更丰富.考虑SDK、SA和CA中存在多分支操作,其运算时间比SE均略有增加,尽管SA和CA的模型参数比SE少.

3.4 实验3:IRB⁺⁺多尺度组件不同分支数量对SSD-MV3P性能的影响

本实验比较IRB⁺⁺多尺度组件不同分支数量对SSD-MV3P性能影响,多尺度组件中分支数量设置为1、2和4,分别记录检测模型在迭代1000次时对感兴趣小目标数据集SST的mAP数值、参数大小和平均检测时间.实验结果如表5所示.

从表5可以发现,随IRB⁺⁺多尺度组件分支数量的

表5 IRB⁺⁺多尺度组件不同分支数量对SSD-MV3P性能影响

分支数量	mAP/%	Parameters/MB	Time/ms
1	85.89	15.4	37.43
2	87.66	15.5	47.83
4	88.24	15.8	66.24

增加,SSD-MV3P检测精度逐渐增加,当多尺度卷积核数量等于4时,SSD-MV3P的检测精度为88.24%.SSD-MV3P的模型参数并没有随多尺度系数的增加有明显的变化,但检测时间随多尺度系数增加也不断增大,当分支数量等于4时,SSD-MV3P的检测时间已经达到66.24ms,比分支数量等于1时SSD-MV3P的检测时间几乎增加了一倍.显然,IRB⁺⁺多尺度组件分支数量的选取应兼顾检测精度和检测时间,本文IRB⁺⁺多尺度组件分支数量为2.

3.5 实验4:二次非极大值抑制算法对HRSSD性能的影响

本实验比较分析二次非极大值抑制算法在HRSSD中对SAS图像感兴趣小目标检测结果的影响,其中无二次非极大值抑制:直接将每个切片的检测结果不加处理的显示在原始SAS图像上;二次非极大值抑制:对切片的检测结果进行二次非极大值抑制处理,将抑制后的检测结果显示在原始SAS图像上.目标检测模型为实验1训练后的SSD-MV3P,输入图像为图4场景1,图像尺寸为4592×5948像素,切割图像尺寸为500×500像素,冗余比例为20%.实验结果如图5所示.

冗余切割的目的是保证图像切片能完整覆盖感兴趣小目标,但同时不可避免的使相邻图像切片之间存在重叠部分,进而导致同一目标被多次检测.二次非极大值抑制的目的就是去除当前位置检测精度较低的结果,保证检测结果的唯一性和完整性.无二次非极大值抑制的检测结果如图5(a)所示,左侧的小目标被检测了两次,其红色包围框覆盖完整且置信度高(93.0%),绿色包围框局部覆盖且置信度相对较低(66.0%).二次

非极大值抑制的检测结果如图5(b)所示,左侧的小目标仅剩一个置信度高(93.0%)且完整的红色包围框.由此可知,基于Soft-NMS的二次非极大值抑制算法对高分辨率大尺寸SAS图像感兴趣小目标重复的同一类型

检测结果具有良好的抑制作用.另外,在实际情况中,若出现因切片重叠导致同一位置出现检测结果不一致时,可利用检测结果的置信度对目标进行最终确认,即选取置信度高的检测结果.

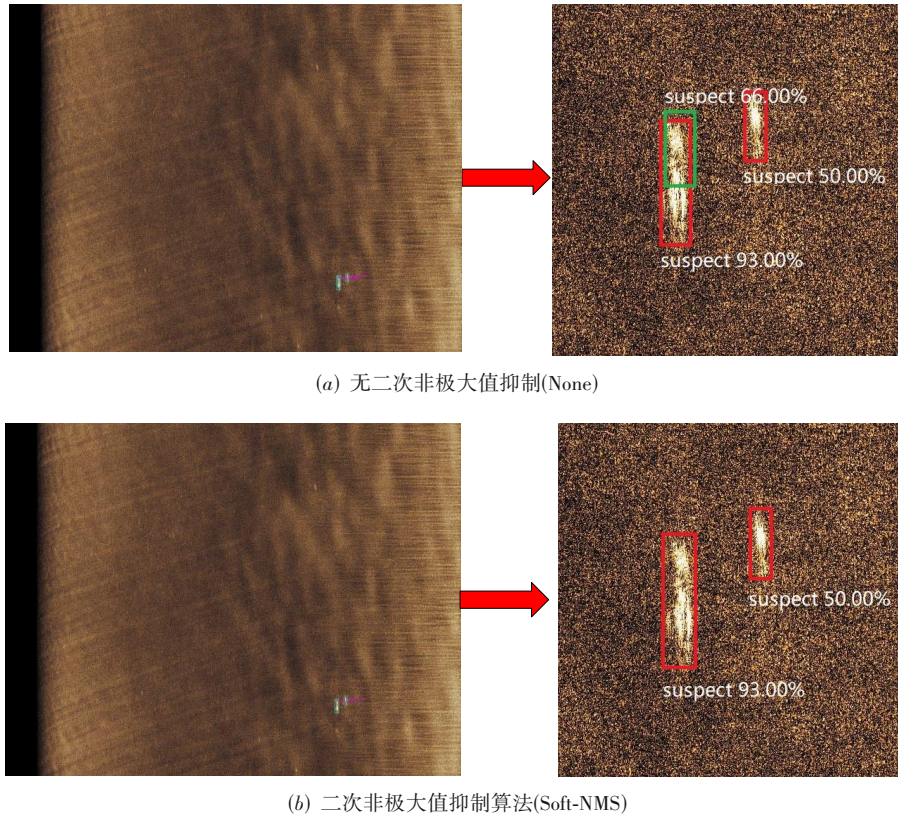


图5 高分辨率大尺寸SAS图像感兴趣小目标检测效果图

3.6 讨论

HRSSD首先对原始SAS图像冗余切割;接着,利用改进的SSD-MV3P对图像切片中感兴趣小目标的检测识别;最后,选用Soft-NMS确保同一位置上目标检测结果的完整性和唯一性.实验从mAP、参数大小和运算时间三个方面比较了不同注意力组件性能上的差异,分析了BSC组件中多尺度、空间和通道三个因素对提高感兴趣小目标检测精度的作用,也进一步研究了BSC组件中多尺度分支数量的选取对模型检测精度和检测时间的影响. IRB⁺中SE组件(图2(a))通过自动学习的方式,获取输入特征的每个通道的重要程度,然后用这个重要程度去给每个特征赋予一个权重值,进而强化对当前任务有用的特征通道,抑制对当前任务用处不大的特征通道. IRB⁺中BSC组件(图2(b))从多尺度、空间和通道三个维度捕获输入特征的有用信息,三个因素彼此独立,均能有效提高模型的检测精度,相互叠加后更是显著提高模型对感兴趣小目标的感知能力.虽然,本文利用空洞卷积减少了SSD-MV3P的模型计算成

本,但BSC中多尺度、空间和通道均采用分支结构,分支结构按顺序计算,因此检测时间有所增加. Soft-NMS是常用的目标检测后处理算法,它不是直接删除所有IoU大于阈值的框,而是降低其置信度,设定一个置信度阈值,得分大于该阈值的检测框得以保留.在实验4中,通过比较分析验证了Soft-NMS对高分辨率大尺寸感兴趣小目标检测结果能实现有效确认.

4 结论

高分辨率大尺寸SAS图像感兴趣小目标检测任务具有重要的理论研究和实际应用价值.本文提出了一种基于SSD的高分辨率大尺寸SAS图像感兴趣小目标检测方法HRSSD,HRSSD采用先切割、后检测、再确认的策略,保证了感兴趣小目标检测结果的完整性和唯一性.在此基础上,提出了一种改进的特征提取模块IRB⁺,IRB⁺模块融合了尺度、空间和通道三种注意力选择机制,可最大限度的提取感兴趣小目标有用特征.与现有方法相比,本文方法实现了高分辨率大尺寸SAS

图像感兴趣小目标的自主检测识别,并且检测精度有进一步的提升。

对于高分辨率大尺寸SAS图像感兴趣小目标检测识别任务,下一步的研究重点包括:(1)研究适合高分辨率大尺寸SAS图像感兴趣小目标的更高效注意力组件;(2)研究更复杂背景条件下的高分辨率大尺寸SAS图像感兴趣小目标自主识别方法。

参考文献

- [1] HAYES M P, GOUGH P T. Synthetic aperture sonar: A review of current status[J]. *IEEE Journal of Oceanic Engineering*, 2009, 34(3): 207-224.
- [2] SYNNE S A V, HANSEN R E, SAEBO T O. Spatial coherence of speckle for repeat-pass synthetic aperture sonar micronavigation[J]. *IEEE Journal of Oceanic Engineering*, 2021, 46(4): 1330-1345.
- [3] THOMAS B, HUNTER A, DUGELAY S. Phase wrap error correction by random sample consensus with application to synthetic aperture sonar micronavigation[J]. *IEEE Journal of Oceanic Engineering*, 2021, 46(1): 221-235.
- [4] BROWN D C, GERG I D, BLANFORD T E. Interpolation kernels for synthetic aperture sonar along-track motion estimation[J]. *IEEE Journal of Oceanic Engineering*, 2020, 45(4): 1497-1505.
- [5] FEI T, KRAUS D, ZOUBIR A M. Contributions to automatic target recognition systems for underwater mine classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(1): 505-518.
- [6] WILLIAMS D P. Fast target detection in synthetic aperture sonar imagery: A new algorithm and large-scale performance analysis[J]. *IEEE Journal of Oceanic Engineering*, 2015, 40(1): 71-92.
- [7] WANG L, LI M H, YE X F, et al. Specific target recognition and segmentation algorithm for real-time side scan sonar images[C]//2015 IEEE International Conference on Mechatronics and Automation (ICMA). Piscataway: IEEE, 2015: 2649-2653.
- [8] FANDOS R, ZOUBIR A M. Optimal feature set for automatic detection and classification of underwater objects in SAS images[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2011, 5(3): 454-468.
- [9] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [10] KWOK R. Deep learning Powers a motion-tracking revolution[J]. *Nature*, 2019, 574(7776): 137-138.
- [11] WILLIAMS D P. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks[C]//2016 23rd International Conference on Pattern Recognition (ICPR). Piscataway: IEEE, 2016: 2497-2502.
- [12] MCKAY J, GERG I, MONGA V, et al. What's mine is yours: Pretrained CNNs for limited training sonar ATR [EB/OL]. (2022-06-29) [2022-06-29]. <https://arxiv.org/abs/1706.09858.pdf>.
- [13] WILLIAMS D P. On the use of tiny convolutional neural networks for human-expert-level classification performance in sonar imagery[J]. *IEEE Journal of Oceanic Engineering*, 2021, 46(1): 236-260.
- [14] 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. Piscataway: IEEE, 2014: 580-587.
- [15] GIRSHICK R. Fast R-CNN[EB/OL]. (2015-04-30)[2022-05-27]. <http://arxiv.org/abs/1504.08083>.
- [16] 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.
- [17] JREDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[EB/OL]. (2015-06-08) [2022-05-09]. <https://arxiv.org/abs/1506.02640>.
- [18] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[EB/OL]. (2015-12-08) [2022-05-08]. <https://arxiv.org/abs/1512.02325>.
- [19] Adam V E. You only look twice: rapid multi-scale object detection in satellite imagery[EB/OL]. (2018-05-24) [2022-05-24]. <http://arxiv.org/abs/1805.09512>.
- [20] HOWARD A G, ZHU M L, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2022-04-17]. <https://arxiv.org/abs/1704.04861>.
- [21] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4510-4520.
- [22] 李宝奇, 黄海宁, 刘纪元, 等. 基于改进SSD的合成孔径声呐图像水下多尺度目标轻量化检测模型[J]. *电子与信息学报*, 2021, 43(10): 2854-2862
- LI B Q, HUANG H N, LIU J Y, et al. Synthetic aperture sonar underwater multi-scale target efficient detection

model based on improved single shot detector[J]. Journal of Electronics & Information Technology, 2021, 43(10): 2854-2862 (in Chinese)

- [23] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [24] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 1314-1324.
- [25] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[EB/OL]. (2018-07-11) [2022-05-18]. <https://arxiv.org/abs/1807.06521.pdf>.
- [26] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[OL]. (2019-10-08) [2022-04-07]. <http://arxiv.org/abs/1910.03151>.
- [27] LI X, WANG W H, HU X L, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 510-519.
- [28] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS—Improving object detection with one line of code[EB/OL]. (2017-04-14) [2022-05-08]. <http://arxiv.org/abs/1704.04503>.
- [29] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [30] QIN Y J, LUO H Y, ZHAO F, et al. NDGCN: Network in network, dilate convolution and graph convolutional networks based transportation mode recognition[J]. IEEE Transactions on Vehicular Technology, 2021, 70(3): 2138-2152.



黄海宁 男,1969年出生于河北,中国科学院声学研究所研究员,博士生导师,国务院政府津贴专家。目前担任中科院声学研究所首席科学家,中科院先进水下信息技术重点实验室主任,中国声学学会理事。主要从事水声信号与信息处理、水声通信与网络、极地声学等方面的研究。

E-mail: hhn@mail.ioa.ac.cn



刘纪元 男,1963年出生于辽宁,中国科学院声学研究所研究员,博士生导师。中国图形图像学会视觉与传感专业委员会会员,国家“863”计划“基于无人平台的合成孔径声纳系统研究”项目首席专家。主要研究领域包括水声信号处理、高分辨率水下成像技术等。

E-mail: Ljy@mail.ioa.ac.cn



刘正君 女,1982年出生于黑龙江,中国科学院声学研究所助理研究员。从事水声图像处理,水下目标检测、识别和跟踪等方面研究。

E-mail: liuzhengjun@mail.ioa.ac.cn



韦琳哲 男,1991年出生于河北廊坊,中国科学院声学研究所副研究员。主要从事合成孔径声纳成像、图像处理及目标识别研究工作。

E-mail: gman200619003@126.com

作者简介



李宝奇 男,1985年出生于天津,中国科学院声学研究所副研究员。主要从事水声信号处理、声纳图像目标识别、智能信息处理等方面的研究。

E-mail: libaoqi@mail.ioa.ac.cn