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

李宝奇 黄海宁* 刘纪元 刘正君 韦琳哲
(中国科学院声学研究所 北京 100190)
(中国科学院先进水下信息技术重点实验室 北京 100190)

摘 要：针对轻量化目标检测模型SSD-MV2对合成孔径声呐(SAS)图像水下多尺度目标检测精度低的问题，该文提出一种新的卷积核模块-可扩张可选择模块(ESK)，ESK具有通道可扩张、通道可选择和模型参数少的优点。与此同时，利用ESK模块重新设计了SSD的基础网络和附加特征提取网络，记作SSD-MV2ESK，并为其选择了合理的扩张系数和多尺度系数。在合成孔径声呐图像水下多尺度目标检测数据集SST-DET上，SSD-MV2ESK在模型参数基本相等的条件下，检测精度比SSD-MV2提升4.71%。实验结果表明，SSD-MV2ESK适用于合成孔径声呐图像水下多尺度目标检测任务。

关键词：合成孔径声呐；图像水下多尺度目标检测；SSD；MobileNet V2；多通道可选择；深度可分离空洞卷积
中图分类号：TN959.72; TP391 **文献标识码**：A **文章编号**：1009-5896(2021)10-2854-09
DOI: 10.11999/JEIT201042

Synthetic Aperture Sonar Underwater Multi-scale Target Efficient Detection Model Based on Improved Single Shot Detector

LI Baoqi HUANG Haining LIU Jiyuan LIU Zhengjun WEI Linzhe
(*Institute of Acoustics*, *Chinese Academy of Sciences*, *Beijing* 100190, *China*)
(*Key Laboratory of Science and Technology on Advanced Underwater Acoustic Signal Processing*, *Chinese Academy of Sciences*, *Beijing* 100190, *China*)

Abstract: In view of the problem that the efficient detection model SSD-MV2 (Single Shot Detector MobileNet V2) has low detection accuracy to underwater multi-scale targets in Synthetic Aperture Sonar (SAS) images, a novel feature extraction module Extended Selective Kernel (ESK) is proposed in this paper. ESK has the advantages of channel scalability, channel selection and few model parameters. At the same time, the basic network and additional feature extraction network of SSD are redesigned by using ESK module, which is named SSD-MV2ESK, and a set of reasonable expansion coefficient and multi-scale coefficient are selected for SSD-MV2ESK. On SST-DET, the mAP of SSD-MV2ESK is 4.71% higher than that of SSD-MV2 when the model parameters are basically the same. The experimental results show that SSD-MV2ESK is suitable for SAR underwater multi-scale target detection task in embedded platform.

Key words: Synthetic Aperture Sonar (SAS); Underwater multi-scale target detection; Single Shot Detector (SSD); MobileNet V2; Channel selectable; Depthwise separable dilated convolution

1 引言

合成孔径声呐 (Synthetic Aperture Sonar, SAS) 是一种高分辨率水下成像声呐，其基本原理是利用小孔径基阵的移动形成虚拟大孔径，从而获得方位向的高分辨率。与普通侧扫声呐相比，SAS 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[1,2]。合成孔径声呐图像目标检测任务在水下无人平台自主导航和搜索发挥着重要作用[3,4]。考虑水下目标尺寸的多样性，即合成孔径声呐图像中目标的尺度差别较大，这会进一步增加目标检测的难度。

通过将深度学习[5–7]模型卷积神经网络 (Convolutional Neural Networks, CNN) [8–10]嵌入到目标

收稿日期：2020-12-14；改回日期：2021-05-29；网络出版：2021-08-27
*通信作者： 黄海宁 hhn@mail.ioa.ac.cn
基金项目：国家自然科学基金(11904386)，国家基础科研计划重大项目(JCKY2016206A003)，中国科学院青年创新促进会(2019023)
Foundation Items: The National Natural Science Foundation of China(11904386), The State Administration of Science, Technology and Industry Program (JCKY2016206A003), The Youth Innovation Promotion Association of Chinese Academy of Sciences (2019023)

检测模型之中，目标检测精度在过去几年中不断提高，结合CNN的目标检测算法可分为基于候选区域和基于回归两类。基于候选区域的算法主要有R-CNN(Region-based Convolutional Neural Networks)[11], Fast R-CNN[12]和Faster R-CNN[13,14]等，此类算法检测速度有待提高。为了提高模型的检测速度，一些研究者开展了无区域建议的目标检测研究，主要采用回归的思想。Redmon等人[15]提出了一种无区域建议的目标检测模型YOLO (You Only Look Once)。YOLO 通过采用空间限制，大大提高了效率，能够达到实时的效果。但是YOLO的检测精度不如Faster R-CNN。针对YOLO存在的不足，Liu等人[16]提出SSD (Single Shot Detector)模型。SSD通过融合6个尺度的特征来提高目标检测的精度。虽然SSD单幅图像检测精度比YOLO有大幅的提高，不过检测速度依然较慢。为了缩短SSD的检测时间，Iandola等人[17]提出了基于FireModule的轻量化SqueezeNet网络。FireModule主要是利用1×1的卷积层对输入特征降维来降低模型的参数和计算量，同时也利用Inception[18]结构提高FireModule的特征提取能力。Howard等人[19]提出了轻量化的卷积神经网络MobileNet V1。MobileNet V1用深度可分离卷积 (Depthwise Separable Convolution, DSC)替换标准卷积来减少模型的参数和计算量，它在不影响目标检测精度的条件下能极大地提高SSD的检测速度。不过，DSC的输出很容易变为0，并且无法恢复。为此，Sandler等人[20]提出了MobileNet V1的改进版本MobileNet V2。MobileNet V2在深度可分离卷积的基础上引入了ResNet中的shortcut connection结构，并设计了新的特征提取模块IRB (Inverted Residual Block)。新模块将原来的先“压缩”后“扩张”调整为先“扩张”后“压缩”，同时为了降低激活函数在高维信息向低维信息转换时的丢失和破坏(DSC的输出很容易变为0)，将最后卷积层的激活层由非线性更改为线性。由于IRB卷积核尺寸单一，同时无法对特征进行有效区分，降低了模型对合成孔径声呐图像水下目标的适应能力。

在卷积神经网络卷积核选取和多尺度特征增强方面，Hu等人[21]提出了SE(Squeeze and Excitation)特征提取模块。SE模块首先对卷积得到的特征进行Squeeze操作，得到全局特征，然后对全局特征进行Excitation操作，得到不同特征的权重，最后乘以对应通道的特征得到最终特征。本质上，SE模块是在特征维度上做选择，这种注意力机制让模型可以更加关注信息量最大的特征，而抑制那些不重要的特征。在此基础上，Li等人[22]提出了SK (Selective Kernel)模块可以针对目标物体的大小选择不同的感受野。输入特征首先经过SK模块多尺度卷积层(使用分组卷积方式提升计算效率)，然后融合所有尺度的特征图，并计算不同尺度不同通道的权重，最后将多个尺度的特征融合成一个与输入特征通道数相等的输出特征，SK模块提高了网络对图像目标的特征提取能力和适应能力。虽然SK模块多尺度卷积层采用分组卷积降低了模型的参数和计算量，但参数依然较多、计算量依然较大。为了保证输出通道与输入通道一致，SK模块的多个尺度的特征相加融合成一个，这必然会造成多尺度特征无法准确区分，进而降低SK模块的特征提取能力。

受深度学习在计算机视觉领域取得突破进展的启发，近年来，国内外的研究学者利用深度学习技术提高SAS图像水下目标识别的准确率。Williams[23]利用深度卷积神经网络对SAS图像目标进行分类识别，提高了SAS图像目标的分类准确率。McKay等人[24]在深度卷积神经网络的基础上，通过迁移学习进一步提高了SAS图像水下目标的分类准确率。Williams[25]通过分析深度卷积神经网络的计算复杂度，选取参数更少的网络来对水下目标进行分类识别。上述3种SAS图像水下目标识别方法主要是利用CNN对SAS图像进行分类识别，因此无法获取图像内目标的位置信息。

针对上述轻量化目标检测方法及其改进方法对SAS图像水下多尺度目标检测精度低的问题，本文提出了一种可扩张、可选择卷积核模块 (Expand Selective Kernel, ESK)，ESK通过优化不同尺度特征层之间的融合方式来提高模块的特征提取能力和利用深度可分离空洞卷积降低模块的参数。接着，利用ESK模块重新设计了SSD的基础网络和附加特征提取网络，并为其选取了合理的参数。最后，在SSD框架内实现对合成孔径图像水下多尺度目标准确的检测。

2　基于改进SSD的合成孔径声呐水下目标检测模型

本节首先介绍新特征提取模块ESK，接着介绍改进SSD模型结构，最后对网络参数的选取进行了分析。

2.1　可扩张可选择卷积核模块

ESK模块借鉴IRB模块的“扩张压缩”残差结构和SK模块的动态选择机制：“扩张压缩”残差结构能有效增加深层网络的梯度传播，动态选择机制允许每个神经元根据输入信息的尺度自适应地调整

其感受野大小[22]，获取信息量最大的特征，增加对水下多尺度目标的适应性。此外，利用深度可分离空洞卷积(Depthwise Separable Dilated Convolution, DSDC) [26]替换分组卷积减少模型的计算成本，DSDC首先将标准卷积分解成DSC和点卷积，然后在DSC中引入一个称作空洞率[27,28]的新参数，并利用扩张率控制卷积核处理数据时各值的间距。同时，通过优化不同尺度卷积层的输入特征数量和融合方式来提高ESK模块的特征提取能力。IRB模块、SK模块、ISK模块和ESK模块的结构关系如图1所示。

图1(a)为IRB模块，模块采用了反残差网络结构，即先对通道采取先“扩张” 后“压缩”的策略，同时删除了最后一个卷积层的激活函数，保留特征的多样性。图1(b)为SK模块，SK模块包括分裂层、多尺度分组卷积层、融合层和选择层4个部分：分裂层是将输入特征分别送入多尺度卷积层；多尺度分组卷积层负责提取输入特征的不同尺度特征；融合层是将多尺度卷积层输出的结果进行叠加融合；选择层是计算多尺度多通道特征的权重系数，与多尺度特征相乘得到输出特征。图1(c)为SK模块的IRB结构，记作ISK。ISK模块由扩张层、分裂层、多尺度分组卷积层、融合层、选择层和压缩层组成。ISK是利用SK模块直接替换IRB模块中的深度可分离卷积。图1(d)为本文提出的ESK特征提取模块，ESK模块由扩张层、切割层、多尺度深度可分离空洞卷积层、拼接层、选择层和压缩层组成。与ISK的主要区别为切割层、多尺度深度可分离空洞卷积层和拼接层。切割层负责将通道放大后的输入特征按多尺度卷积核个数等分后分别送入不同尺度深度可分离空洞卷积层；多尺度深度可分离空洞卷积层负责提取输入信息不同尺度上的特征信息；拼接层负责将多尺度深度可分离卷积层的输出特征在通道上拼接合并。

对于一个任意的输入特征$\boldsymbol{F}\in\boldsymbol{\Phi}^{H\times W\times M}$，其中$H\times W$为输入特征的尺寸，$M$为输入特征的通道数。输入特征$\boldsymbol{F}$进入ESK模块的两个支路网络：左侧支路负责多尺度特征提取和选择；右侧支路保持输入特征$\boldsymbol{F}$不变，并最后与左侧支路网络的输出特征相加。对于左侧支路网络，输入特征$\boldsymbol{F}$首先经过扩张层，其输出特征的数学表达式为

$$\text{E}:\ \boldsymbol{F}\to\boldsymbol{U}\in\boldsymbol{\Phi}^{H\times W\times(k\times M)} \tag{1}$$

其中，$\boldsymbol{F}$为原始输入特征，$\boldsymbol{U}$为经过扩张层后的特征，扩张层的卷积核尺寸为1×1，卷积核的数量为输入特征通道的k倍，即$k\times M$。

随后，输出特征$\boldsymbol{U}$经切割层送入多尺度深度可分离空洞卷积层，其输出特征的数学表达式为

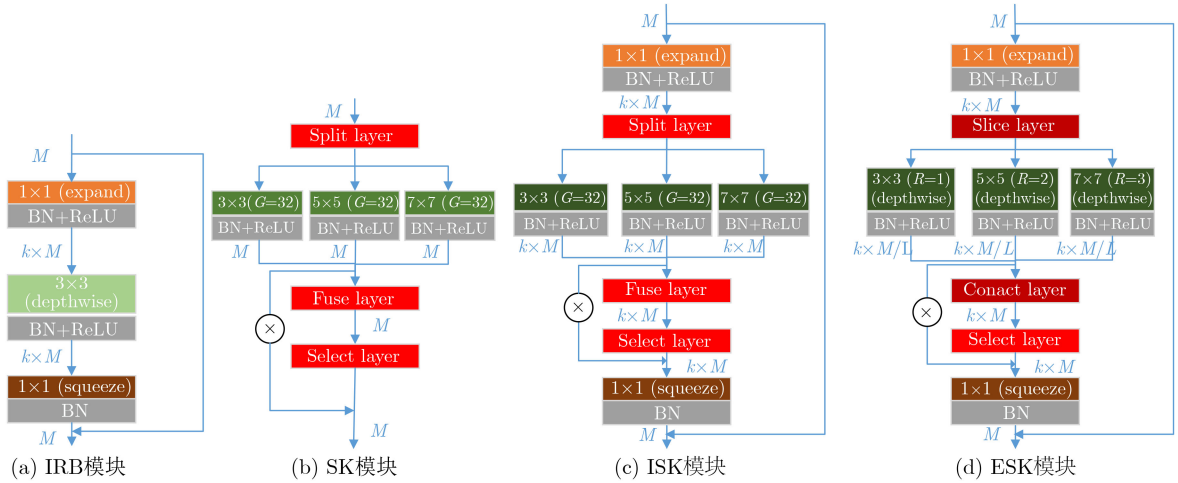
$$\text{D}:\ \boldsymbol{U}\to\boldsymbol{V}_l\in\boldsymbol{\Phi}^{H\times W\times(k\times M/L)},\ l=1,2,\cdots,L \tag{2}$$

其中，$\boldsymbol{V}_l$为深度可分离空洞卷积层输出的特征图，特征图尺寸为$H\times W$，通道数为$k\times M/L$，L为多尺度实际空洞滤波器(pRactical Dilated Filter, RDF)的类型数，例如RDF为3, 5和7，则$L=3$。RDF尺寸与空洞率之间的关系为

$$K_{\text{rdf}}=\kappa+(\kappa-1)(R-1) \tag{3}$$

其中，K_{rdf}为该层RDF尺寸，κ为该层卷积核尺寸，R为该层空洞率大小。例如，一个卷积核尺寸为3×3，空洞率$R=2$的空洞卷积层，RDF的实际覆盖范围为5×5，即$K_{\text{rdf}}=5$。进一步增大空洞率R来扩大卷积层的感受野。因此，ESK可以用更少的参数和计算量实现与ISK相当的特征提取能力。

接着，对L个多尺度深度可分离空洞卷积层的输出$\boldsymbol{V}_l$在通道项进行拼接融合，其输出特征的数学表达式为



(a) IRB模块 (b) SK模块 (c) ISK模块 (d) ESK模块

图 1 ESK特征提取模块

$$\mathrm{C}:\ \boldsymbol{V}_l \to \boldsymbol{V} \in \boldsymbol{\Phi}^{H\times W\times(k\times M)},\ l=1,2,\cdots,L \tag{4}$$

其中，$\boldsymbol{V}$为拼接融合后的输出特征，输出特征图的尺寸$H\times W$，通道数为$k\times M$与扩张后的通道数相等。

然后，对拼接融合后的输出特征$\boldsymbol{V}$的进行通道选择，多尺度通道选择系数的数学表达式为

$$\boldsymbol{s}=\mathrm{softmax}(f_{\mathrm{cs}}(f_{\mathrm{c}}(P_{\mathrm{g}}(\mathbf{V})))) \tag{5}$$

其中，$\boldsymbol{s}$为多尺度通道的选择系数，$\boldsymbol{s}\in\boldsymbol{\Phi}^{1\times(k\times M)}$；$P_{\mathrm{g}}()$为全局池化函数，输出特征维度为$\boldsymbol{\Phi}^{1\times(k\times M)}$；$f_{\mathrm{c}}$为第1全连接层，输出特征维度为$\boldsymbol{\Phi}^{1\times d}$，其中$d=32$；$f_{\mathrm{cs}}$为第2全连接层，输出特征维度为$\boldsymbol{\Phi}^{1\times(k\times M)}$；softmax()为归一化指数函数。多尺度通道选择系数(式(5))与输出特征相乘得到选择后的输出特征，其数学表达式为

$$\boldsymbol{V}'=\boldsymbol{s}\cdot\boldsymbol{V} \tag{6}$$

其中，$\boldsymbol{V}'$为通道选择后的多尺度通道特征。

接着，对$\boldsymbol{V}'$进行通道压缩，数学表达式为

$$\mathrm{S}:\ \boldsymbol{V}'\to\boldsymbol{F}'\in\boldsymbol{\Phi}^{H\times W\times M} \tag{7}$$

其中，$\boldsymbol{F}'$为通道压缩后的多尺度通道特征。

通过上面的计算，最后可以得到ESK模块的输出特征数学表达式为

$$\boldsymbol{G}=\boldsymbol{F}+\boldsymbol{F}' \tag{8}$$

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

2.2 基于可扩张可选择卷积核模块的SSD合成孔径声呐图像水下目标检测模型设计

基于ESK模块的SSD水下目标检测模型结构如图2所示，记作SSD-MV2ESK，包括基础网络(MobileNet V2ESK)、附加特征提取网络 (ESKAN)、Default boxes生成和卷积预测4个部分。

基础网络MobileNet V2ESK与MobileNet V2的网络结构保持一致，利用ESK模块替换IRB模块实现。SSD-MV2ESK附加特征提取网络一共提取6个尺度的特征，MobileNet V2ESK中的第14层Conv14和第19层Conv19的输出作为附加特征提取网络的第1特征层和第2特征层，特征图尺寸为19×19和10×10；Conv19_1，Conv19_2，Conv19_3和Conv19_4作为附加特征提取网络的第3—第6尺度特征层，4个特征层的输出特征图尺寸为5×5, 3×3, 2×2和1×1。Default Boxes生成部分根据预先定义的scales和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。SSD-MV2ESK与目标检测模型SSD的训练过程[16]一样。

2.3 MV2ESK和SSD-MV2ESK分析

对于M个尺寸为$D_H\times D_W$的输入特征图$\boldsymbol{F}$，经尺寸为$D_K\times D_K$的卷积核操作后，输出N个尺寸为$D_H\times D_W$的特征图$\boldsymbol{G}$，其中M是输入通道数，N是输出通道数，D_H和D_W是输入(出)的特征图的宽度和高度。为了便于计算和分析，限定输入通道数M等于输出通道数N，输入(出)特征图尺寸D_H等于D_W，图1(a)的$D_K\times D_K$等于3×3，图1(b)、图1(c)和图1(d)的多尺度通道数$L=3$。同时，省略参数或计算量较少的网络层，例如shortcut connection层、BN层、分裂层、分割层、融合层和拼接层。

IRB模块的生成特征图$\boldsymbol{G}$的计算成本为

$$M\times k\times M\times D_H\times D_H+3\times3\times k\times M\times D_H\times D_H + k\times M\times M\times D_H\times D_H \tag{9}$$

其中，第1项为扩张层的计算成本，第2项为深度可分离卷积层的计算成本，第3项为压缩层的计算成本。

SK 模块生成特征图$\boldsymbol{G}$的计算成本为

$$(3\times3+5\times5+7\times7)/32\times M\times M\times D_H\times D_H + k\times M\times32+32\times k\times M \tag{10}$$

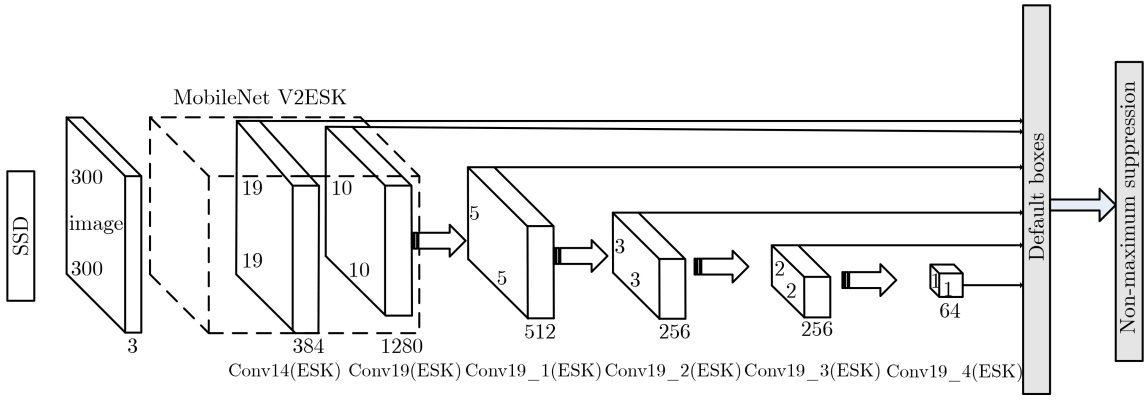


图 2 基于ESK模块的SSD目标检测模型

其中，第1项为多尺度卷积层的计算成本，第2和第3项为选择层的计算成本。

ISK模块生成特征图$\boldsymbol{G}$的计算成本为

$$\begin{aligned}&M\times k\times M\times D_H\times D_H+(3\times 3+5\times 5+7\times 7)\\&\times k\times M\times k\times M\times D_H\times D_H/32+k\times M\times 32\\&+32\times k\times M+k\times M\times M\times D_H\times D_H\end{aligned}\tag{11}$$

其中，第1项为扩张层的计算成本，第2项为多尺度分组卷积层的计算成本，第3和第4项为选择层的计算成本，第5项为压缩层的计算成本。

ESK 模块生成特征图$\boldsymbol{G}$的计算成本为

$$\begin{aligned}&M\times k\times M\times D_H\times D_H+(3\times 3+3\times 3+3\times 3)\\&\times k\times M/L\times D_H\times D_H+k\times M\times 32\\&+32\times k\times M+k\times M\times M\times D_H\times D_H\end{aligned}\tag{12}$$

其中，第1项为扩张层的计算成本，第2项为多尺度深度可分离卷积层的计算成本，第3和第4项为选择层的计算成本，第5项为压缩层的计算成本。

ISK模块与IRB模块的计算成本比值为

$$\begin{aligned}&\frac{M\times k\times M\times D_H\times D_H+(3\times 3+5\times 5+7\times 7)\times k\times M\times k\times M\times D_H\times D_H/32+k\times M\times 32+32\times k\times M+k\times M\times M\times D_H\times D_H}{M\times k\times M\times D_H\times D_H+3\times 3\times k\times M\times D_H\times D_H+k\times M\times M\times D_H\times D_H}\\&=\frac{83/32\times k\times M+2\times M}{3\times 3+2\times M}+\frac{64}{(2\times M+9)\times D_H\times D_H}\end{aligned}\tag{13}$$

ESK模块与IRB模块的计算成本比值为

$$\begin{aligned}&\frac{M\times k\times M\times D_H\times D_H+(3\times 3+3\times 3+3\times 3)\times k\times M/L\times D_H\times D_H+k\times M\times 32+32\times k\times M+k\times M\times M\times D_H\times D_H}{M\times k\times M\times D_H\times D_H+3\times 3\times k\times M\times D_H\times D_H+k\times M\times M\times D_H\times D_H}\\&=1+\frac{64}{(2\times M+9)\times D_H\times D_H}\end{aligned}\tag{14}$$

当多尺度系数$L=3$时，对于式(13)，当M取值较大时，ISK模块的计算成本与IRB的计算成本比值约等于$k+1$；对于式(14)，ESK模块的计算成本与IRB的计算成本比值基本相同。

对于由ESK模块组成的SSD-MV2ESK网络，由式(12)可以发现，扩张系数k对SSD-MV2ESK模块的计算成本影响较大，而多尺度系数L基本上对SSD-MV2ESK模块的计算成本没有影响。除上述因素外，SK, ISK和ESK在Pytorch深度学习框架内通过一个循环结构实现多尺度卷积组的设计，即每次仅进行一个尺度的卷积运算，因此，多尺度系数L会影响SSD-MV2ESK的计算时间。鉴于上述原因，扩张系数k主要影响 SSD-MV2ESK的模型参数，多尺度系数L主要影响SSD-MV2ESK的计算时间。为了平衡SSD-MV2ESK模型检测精度、参数大小和检测时间，基础网络中ESK模块的扩张系数$k=4$，多尺度系数L=4，附加特征提取网络中Conv19_1, Conv19_2, Conv19_3和Conv19_4的扩张系数以此为0.2, 0.25, 0.5和0.25，多尺度系数依次为4, 2, 2和1。

3 仿真实验

为了验证ESK的有效性以及扩张系数和多尺度系数对SSD-MV2ESK性能的影响，实验以mAP、参数大小和平均检测时间作为模型定量评价指标。设计实验1，以SSD-MV2(基础网络为MobileNet V2，特征提取网络为IRBAN)为参考，比较分析不同轻量化目标检测模型之间的性能差异。考虑ESK模块在SSD-MV2ESK基础网络和附加特征提取网络中占的比重，实验仅分析基础网络中扩张系数和多尺度系数对SSD-MV2ESK性能的影响。设计实验2，以基础网络MobileNet V2ESK中扩张系数为研究对象，比较分析不同扩张系数对SSD-MV2ESK性能的影响。设计实验3，以基础网络MobileNet V2ESK中多尺度系数为研究对象，比较分析不同多尺度系数对SSD-MV2ESK性能的影响。为了进一步验证ESK模块对水下多尺度目标的适用性，设计实验4，以单尺度的MobileNet V2ESK分类结果为参考，比较分析不同多尺度系数的MobileNet V2ESK对水下多尺度目标特征提取效果。实验平台基于Dell PowerEdge R730深度学习服务器，操作系统为RedHat Enterprise linux 7.5、环境管理软件为Anaconda3、深度学习框架为Torch 1.3.1和Torchvision 0.4.2等；CPU处理器为Intel E5-2603 V4、内存大小是32 GB、GPU计算单元为两个V100(16 GB)。输入图像的尺寸被剪切为300像素×300像素、BatchSize=32、学习率=0.001、所有模型均在V100(16 GB)上进行训练和测试。

3.1 实验数据集

为了更好地检验SSD-MV2ESK对合成孔径声呐图像水下多尺度目标的检测性能，本文建立了一个水下多尺度目标检测数据集：SST-DET。SST-DET数据集主要为高频合成孔径声呐图像，采集地点包括千岛湖、丹江口等地，包括3种水下目标：圆柱形目标、线缆和疑似物，共计704幅图像，其中633幅用于模型训练，71幅图像用于模型

测试，如表1所示。圆柱体和疑似物目标的像素比约为0.05，线缆目标在图像某一个方向上的像素比大于0.5。从这个角度认定圆柱体和疑似物为小尺寸目标，线缆属于大尺寸目标。

3.2 实验1：目标检测算法的性能比较

本实验比较分析SSD-SQ[21], SSD-MV1[19], SSD-MV2[20], SSD-MV2ISK[22]与本文目标检测方法SSD-MV2ESK在数据集SST-DET上的性能差异。SSD-SQ的基础网络为SqueezeNet，特征提取网络为OAN；SSD-MV1的基础网络为MobileNet V1，特征提取网络为OAN；SSD-MV2的基础网络为MobileNet V2，特征提取网络为IRB-AN；SSD-MV2ISK的基础网络为基于ISK模块的MobileNet V2ISK网络，附加特征提取网络为基于ISK模块的ISKAN网络；SSD-MV2ESK的基础网络为MobileNet-V2ESK，特征提取网络为ESKAN。分别记录检测模型在迭代1000次时对SST-DET测试数据集的mAP数值、参数大小和平均检测时间。

从表2可以发现，SSD-MV2ESK的检测精度比SSD-SQ, SSD-MV1, SSD-MV2和SSD-MV2ISK分别高16.18%, 7.62%, 4.71%和2.21%；模型参数比SSD-SQ和SSD-MV2分别高6 MB和0.1 MB，比SSD-MV1和SSD-MV2SK分别低15.1 MB和46.8 MB；检测时间比SSD-SQ, SSD-MV1和SSD-MV2分别高35.42 ms, 36.2 ms和28.77 ms，比SSD-MV2SK减少32.96 ms。SSD-MV2ESK检测精度最高为75.08%，SSD-SQ的检测精度最低为58.90%；SSD-SK的模型参数最大为59.4 MB，SSD-SQ的模型参数最小为7.51 MB；SSD-SK的检测时间最多为79.63 ms，SSD-MV1的检测时间最少为10.47 ms。虽然SSD-MV2ISK比SSD-MV2的检测精度提高2.5%，但模型的参数和检测时间均大幅提升。综合考虑检测精度(mAP)、参数大小和平均检测时间3个因素，SSD-MV2ESK优于其他检测模型，更适合基于合成孔径声呐图像水下多尺度目标检测任务。

为了更直观地说明SSD-MV2ESK对合成孔径图像水下多尺度目标的检测效果，利用训练20000次的SSD-MV2ESK模型分别对3种水下目标图像进行检测，检测结果如图3所示。从图3可以看出，SSD-MV2ESK模型对3种水下多尺度目标能实现准确检测。

表 1　合成孔径声呐水下多尺度目标检测数据集组成

目标	训练(幅)	测试(幅)
圆柱形目标	103	8
线缆	275	30
疑似物	255	33
总计	633	71

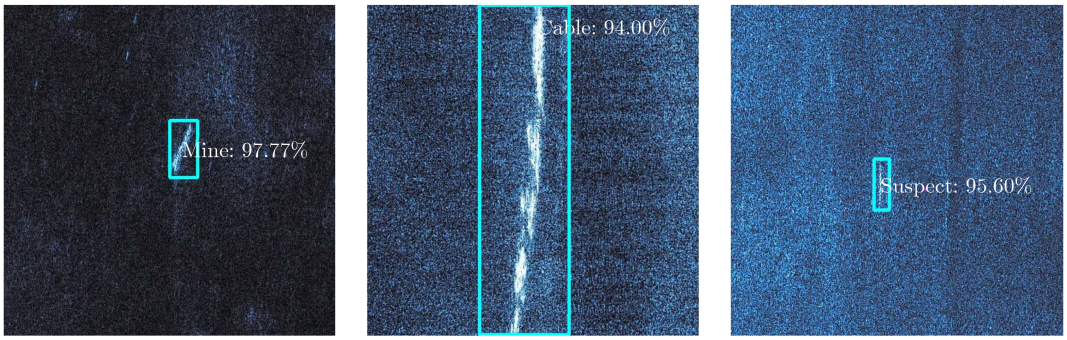
3.3 实验2：基础网络扩张系数对SSD-MV2ESK性能的影响

本实验比较基础网络不同扩张系数对SSD-MV2ESK性能的影响。基础网络SSD-MV2ESK的多尺度系数等于1，扩张系数分别为1, 5, 10, 15, 20和40。附加特征提取网络的扩张系数依次为0.2, 0.25, 0.5和0.25，多尺度系数依次为4, 2, 2和1。记录模型迭代1000次时模型对SST-DET测试数据集的mAP数值、平均检测时间和参数大小。

从表3可以看出，SSD-MV2ESK随基础网络扩张系数的增加检测精度逐渐增加，当扩张系数等于40时，SSD-MV2ESK的检测精度已经达到85.29%。另外，SSD-MV2ESK模型参数随基础网络扩张系数增加也不断增大，当扩张系数等于40时，SSD-MV2ESK的模型参数已经达到256 MB，不过SSD-MV2ESK的检测时间并没有随扩张系数的增加有明显的变化。扩张系数等于10、多尺度系数等于1时SSD-MV2ESK的检测精度与扩张系数等于4、多尺度系数等于4时SSD-MV2ESK 的检测精度基本相同(实验1)，不过模型参数已经达到30.5 MB，明显高于扩张系数等于4、多尺度系数等于4时SSD-MV2ESK模型参数12.6 MB。换句话说，虽然增大扩张系数能提高SSD-MV2ESK的检测精度，但模型参数的增加也比较明显，在相同的检测精度条件下，仅依靠增加扩张系数的SSD-MV2ESK比扩张系数和多尺度系数结合的SSD-MV2ESK要付出更多的存储空间。

表 2　目标检测模型性能比较

模型	基础网络	特征提取提取网络	mAP(%)	模型参数(MB)	检测时间(ms)
SSD-SQ[21]	SQNet	OAN	58.90	6.6	11.25
SSD-MV1[19]	MobileNet V1	OAN	67.46	27.7	10.47
SSD-MV2[20]	MobileNet V2	IRBAN	70.37	12.5	17.90
SSD-MV2ISK[22]	MobileNet V2ISK	ISKAN	72.87	59.4	79.63
SSD-MV2ESK	MobileNet V2ESK	ESKAN	75.08	12.6	46.67



(a) 圆柱形目标 (b) 线缆 (c) 疑似物

图 3 SSD-MV2ESK对合成孔径声呐水下多尺度目标的检测效果图

表 3 基础网络扩张系数对SSD-MV2ESK性能的影响

扩张系数	mAP(%)	模型参数(MB)	检测时间(ms)
1	66.27	6.4	28.20
5	74.39	14.9	31.34
10	74.72	30.5	31.36
15	78.32	52.4	31.77
20	81.67	80.6	32.28
40	85.29	256.0	32.92

3.4 实验3：基础网络多尺度系数对SSD-MV2ESK性能的影响

本实验比较基础网络不同多尺度系数对SSD-MV2ESK性能的影响。基础网络SSD-MV2ESK的扩张系数等于4，多尺度系数分别为1, 2和4(3, 5, 6没有数据，主要是因为它们无法保证每个多尺度输入通道数为整数)。记录模型迭代1000次时对SST-DET测试数据集的mAP数值、平均检测时间和参数大小。

从表4可以看出，SSD-MV2ESK随基础网络多尺度系数的增加检测精度增大，模型参数基本不变，检测时间存在明显增加。在深度学习Pytorch框架内，多尺度深度可分离空洞卷积层采用循序并行化结构设计，但在前向传播过程中每个多尺度卷积层是依次进行的，导致SSD-MV2ESK运算时间随多尺度系数增加。虽然增加多尺度系数能提高SSD-MV2ESK模型的检测精度，但模型的运算时间也存在明显的增加，因此，结合扩张系数和多尺度系数对SSD-MV2ESK是更好的选择。

表 4 基础网络多尺度系数对SSD-MV2ESK性能的影响

多尺度系数	mAP(%)	模型参数(MB)	检测时间(ms)
1	70.46	12.5	30.89
2	71.81	12.5	36.29
4	75.08	12.6	46.46

3.5 实验4：不同多尺度系数的ESK模块对水下多尺度目标适应性比较

为了进一步说明ESK模块对水下多尺度目标的适用性，在MobileNet V2的基础上，利用ESK模块设计3个轻量化目标分类网络MobileNet V2_4_1, MobileNet V2_4_2和MobileNet V2_4_4，3个模型分类网络的扩张系数等于4、多尺度系数分别为1, 2和4。用于分类测试实验的合成孔径图像水下多尺度目标分类数据集为SAS-DET中的703幅图像，其中训练样本集个数638幅图像，测试样本集数量65幅图像。模型训练Batchsize等于32、学习率等于0.01、迭代次数等于100。记录3个模型迭代100次过程中最高的分类准确率，实验结果如表5所示。

表 5 模型分类准确率(%)

分类网络	MobileNet V2_4_1	MobileNet V2_4_2	MobileNet V2_4_4
准确率	72.72	77.27	78.78

从表5可以发现，MobileNet V2_4_4的最高分类准确率比MobileNet V2_4_1和MobileNet V2_4_2分别高6.06%和1.51%。由于3个网络的扩张系数相同，即输入给多尺度深度可分离卷积组的特征数量是一样的，具有更多尺度的MobileNet V2_4_2和MobileNet V2_4_4比单尺度MobileNet V2_4_1的分类准确率高间接地明ESK模块对水下多尺度目标具有更好的适应性。

3.6 讨论

实验从mAP、平均检测时间和参数大小3个方面比较了本文合成孔径声呐图像水下多尺度目标检测方法SSD-MV2ESK与经典轻量化目标检测算法(SSD-SQ和SSD-MV1)和最新算法(SSD-MV2和SSD-MV2ISK)性能上的差异，也进一步分析了基

础网络的扩张系数和多尺度系数的选取如何影响SSD-MV2ESK的性能，同时也间接验证了ESK模块对水下多尺度目标的适用性。ESK模块结合IRB模块和SK模块的优点，并利用深度可分离空洞卷积替换分组卷积和优化输入特征分配和结合方式改善了模块的性能。通过原理和计算成本定量分析发现扩张系数对SSD-MV2ESK模型的计算成本影响较大，多尺度系数对模型计算成本影响较小。不过，由于在深度学习框架Pytorch内多尺度卷积组采用循序计算方式，将导致SSD-MV2ESK模块运算时间随多尺度系数增加而增加。虽然单独增加扩张系数和多尺度系数均能提高SSD-MV2ESK对水下多尺度目标的检测精度，但为了平衡检测精度、模型大小和计算时间，SSD-MV2ESK的扩张系数和多尺度系数均等于4。

SSD-SQ通过通道压缩来降低模型的计算量和参数，SSD-MV1通过标准卷积分解来降低模型的计算量和参数。通道压缩会引起激活函数在高维信息向低维信息转换时特征的丢失和破坏，SSD-SQ的较低的检测精度(58.90%)也证实了这一观点。SSD-MV2利用IRB特征提取模块重新设计了SSD-MV1的基础网络和附加特征提取网络，通过“扩张压缩”来提高深度可分离卷积层的特征提取性能，同时降低模型的参数。SSD-MV2比SSD-MV1的模型参数大幅降低，减少15.2 MB，同时检测精度也有提升，提升2.5%。SSD-MV2ISK则是利用ISK模块替换SSD-MV2中的IRB模块，虽然SSD-MV2ISK的检测精度比SSD-MV2有所提升，但代价是成倍的模型参数和计算时间。SSD-MV2ESK通过合理选择扩张系数和多尺度系数，可以较好地平衡检测精度、模型大小和计算时间。更重要的是，在相同的扩张系数和多尺度系数条件下，SSD-MV2ESK比SSD-MV2ISK检测精度更高、模型参数更少、检测时间更短。

对于合成孔径图像水下多尺度目标而言，在保持检测精度的同时需兼顾模型参数大小和检测时间，结合实验1、实验2和实验3的结果，显然结合扩张系数和多尺度系数的SSD-MV2ESK更适合合成孔径声呐多尺度目标的检测任务。实验4也进一步地证明了ESK模块对水下多尺度目标的适用性。

4　结论

合成孔径声呐图像水下多尺度目标检测任务具有重要的理论研究和实际应用价值。在SSD检测模型框架内，本文提出了一种多通道、通道可扩张且可选择的卷积模块ESK，并利用ESK重新设计了SSD的基础网络和附加特征提取网络。ESK有效提升SSD对合成孔径声呐图像水下目标的检测精度，并经理论分析和仿真实验证明了ESK特征提取模块对SAS图像水下多尺度目标的有效性。

对于基于SSD的合成孔径声呐图像水下多尺度目标检测任务，改进Default Boxes生成策略同样能提升SSD模型的性能。下一步的研究重点包括：(1) 研究适合捕获合成孔径声呐图像水下多尺度目标特征的Default Boxes生成策略；(2)研究更加轻量化的合成孔径声呐图像水下多尺度目标特征提取模块。

参 考 文 献

[1] HAYES M P and GOUGH P T. Synthetic aperture sonar: A review of current status[J]. *IEEE Journal of Oceanic Engineering*, 2009, 34(3): 207–224. doi: 10.1109/JOE.2009.2020853.

[2] 吴浩然, 张非也, 唐劲松, 等. 基于参考距离史的多子阵SAS成像算法[J]. 电子与信息学报, 2021, 43(3): 650–656. doi: 10.11999/JEIT200620.

WU Haoran, ZHANG Feiye, TANG Jinsong, *et al*. A imaging algorithm based on the reference range history for the multiple receivers synthetic aperture sonar[J]. *Journal of Electronics & Information Technology*, 2021, 43(3): 650–656. doi: 10.11999/JEIT200620.

[3] WANG Peng, CHI Cheng, ZHANG Yu, *et al*. Fast imaging algorithm for downward-looking 3D synthetic aperture sonars[J]. *IET Radar, Sonar & Navigation*, 2020, 14(3): 459–467.

[4] SUN Sibo, CHEN Yingchun, QIU Longhao, *et al*. Inverse synthetic aperture sonar imaging of underwater vehicles utilizing 3-D rotations[J]. *IEEE Journal of Oceanic Engineering*, 2020, 45(2): 563–576. doi: 10.1109/JOE.2019.2891281.

[5] HINTON G. Where do features come from?[J]. *Cognitive Science*, 2014, 38(6): 1078–1101. doi: 10.1111/cogs.12049.

[6] LECUN Y, BENGIO Y, and HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436–444. doi: 10.1038/nature14539.

[7] SCHMIDHUBER J. Deep learning in neural networks: An overview[J]. *Neural Networks*, 2015, 61: 85–117. doi: 10.1016/j.neunet.2014.09.003.

[8] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84–90. doi: 10.1145/3065386.

[9] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al*. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 770–778.

[10] XIE Saining, GIRSHICK R, DOLLÁR P, *et al*. Aggregated residual transformations for deep neural networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017: 5987–5995.

[11] 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, Washington, USA, 2014: 580–587.

[12] GIRSHICK R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, USA, 2015: 1440–1448.

[13] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904–1916. doi: 10.1109/TPAMI.2015.2389824.

[14] REN Shaoqing, HE Kaiming, 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. doi: 10.1109/TPAMI.2016.2577031.

[15] 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), Las Vegas, USA, 2016: 779–788.

[16] LIU Wei, ANGUELOV D, ERHAN D, *et al*. SSD: Single shot MultiBox detector[C]. The 14th European Conference, Amsterdam, The Kingdom of the Netherlands, 2016: 21–37.

[17] IANDOLA F N, HAN Song, MOSKEWICZ M W, *et al*. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size[C]. The 5th International Conference on Learning Representations, Toulon, France, 2017.

[18] SZEGEDY C, LIU Wei, JIA Yangqing, *et al*. Going deeper with convolutions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015: 1–9.

[19] HOWARD A G, ZHU Menglong, CHEN Bo, *et al*. Mobilenets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. https://arxiv.org/abs/1704.04861, 2017.

[20] SANDLER M, HOWARD A, ZHU Menglong, *et al*. MobileNetV2: inverted residuals and linear bottlenecks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 4510–4520.

[21] HU Jie, SHEN Li, ALBANIE S, *et al*. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011–2023. doi: 10.1109/TPAMI.2019.2913372.

[22] LI Xiang, WANG Xiang, HU Xiaolin, *et al*. Selective kernel networks[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019: 510–519.

[23] WILLIAMS D P. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks[C]. The 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexican, 2016: 2497–2502.

[24] MCKAY J, GERG I, MONGA V, *et al*. What's mine is yours: Pretrained CNNs for limited training sonar ATR[C]. OCEANS 2017 - Anchorage, Anchorage, USA, 2017: 1–7.

[25] 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. doi: 10.1109/JOE.2019.2963041.

[26] 李宝奇, 贺昱曜, 强伟, 等. 基于并行附加特征提取网络的SSD地面小目标检测模型[J]. 电子学报, 2020, 48(1): 84–91. doi: 10.3969/j.issn.0372-2112.2020.01.010.
LI Baoqi, HE Yuyao, QIANG Wei, *et al*. SSD with parallel additional feature extraction network for ground small target detection[J]. *Acta Electronica Sinica*, 2020, 48(1): 84–91. doi: 10.3969/j.issn.0372-2112.2020.01.010.

[27] 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. doi: 10.1109/TPAMI.2017.2699184.

[28] WANG Panqu, CHEN Pengfei, YUAN Ye, *et al*. Understanding convolution for semantic segmentation[C]. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, USA, 2018: 1451–1460.

李宝奇：男，1985年生，特别研究助理，研究方向为水声信号处理、目标检测、识别和跟踪、深度学习理论.

黄海宁：男，1969年生，研究员，研究方向为水声信号与信息处理、目标探测、水声通信与网络等.

刘纪元：男，1963年生，研究员，研究方向为水声信号处理、数字信号处理和水声成像与图像处理等.

刘正君：女，1982年生，助理研究员，研究方向为水声信号处理等.

韦琳哲：男，1991年生，助理研究员，研究方向为水声信号处理等.

责任编辑：马秀强