◇ 研究报告 ◇

# 基于改进MobilenetV2网络的声光图像融合水下 目标分类方法<sup>\*</sup>

巩文静<sup>1,2,3</sup> 田 杰<sup>1,3</sup> 李宝奇<sup>1,3</sup> 刘纪元<sup>1,3†</sup>

(1 中国科学院声学研究所 北京 100190)

(2 中国科学院大学 北京 100049)

(3 中国科学院先进水下信息技术重点实验室 北京 100190)

摘要:针对小样本条件下水下目标分类准确率低、计算资源量大的问题,提出一种声光图像融合目标分类 方法。首先,对MobilenetV2网络进行改进,去掉第9层之后的网络层,并将该层卷积通道数改为128,通过 Flatten层进行数据降维,增加一个全连接层得到分类结果;其次,设计一种融合网络结构,将声光图像成对输 入网络进行特征提取,在中间层利用通道拼接算法实现特征图融合,使用融合特征进行目标分类。在真实数据 集上对网络进行训练,结果表明,改进的MobilenetV2网络对水下目标的分类性能更好,融合网络的分类准确 率相比融合前有所提高,更加适用于水下目标分类任务。

关键词: 改进 MobilenetV2; 声学图像; 光学图像; 图像融合; 水下目标分类

中图法分类号: TB566 文献标识码: A 文章编号: 1000-310X(2022)03-0462-09 DOI: 10.11684/j.issn.1000-310X.2022.03.017

# Acoustic-optical image fusion underwater target classification method based on improved MobilenetV2

GONG Wenjing<sup>1,2,3</sup> TIAN Jie<sup>1,3</sup> LI Baoqi<sup>1,3</sup> LIU Jiyuan<sup>1,3</sup>

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

(2 University of Chinese Academy of Sciences, Beijing 100049, China)

(3 Key Laboratory of Science and Technology on Advanced Underwater Acoustic Signal Processing, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: To solve the problems of low accuracy and high consumption of underwater target classification under the condition of small samples, an acoustic-optical image fusion classification method is proposed. Firstly, the Mobilenetv2 network is improved by removing the network layer after layer 9 and changing the channels to 128. After the dimension reduction by flatten layer, a dense layer is added to get classification results. Secondly, a fusion network structure is designed, which inputs acoustic image and optical image in pairs for feature extraction. Then, the extracted feature maps are combined in the middle layer and used for target classification. Using the real data to train the network, the results show that the improved Mobilenetv2 network has better classification performance for underwater targets, and the accuracy of fusion network is improved compared with that before fusion, indicating the applicability in underwater target classification.

Keywords: Improved Mobilenetv2; Acoustic image; Optical image; Image fusion; Underwater target classification

作者简介: 巩文静(1996-), 女, 山东泰安人, 博士研究生, 研究方向: 信号与信息处理。

<sup>2021-04-28</sup> 收稿; 2021-07-12 定稿

<sup>\*</sup>中国科学院国防科技重点实验室基金项目(CXJJ-20S035)

<sup>†</sup>通信作者 E-mail: ljy@mail.ioa.ac.cn

# 0 引言

水下目标成像及分类可以通过不同的成像技 术实现,利用光学传感器获得的图像分辨率较高、 目标较为直观<sup>[1]</sup>,在目标分类研究中有重要应用。 Gleason等<sup>[2]</sup>使用监督分类的方法对水下光学图像 中的目标和海床进行分类,将海床的局部地形或高 度作为附加特征添加到分类器中,具有一定的有效 性;Pramunendar等<sup>[3]</sup>提出了一种应用于反向传播 神经网络的图像增强模型,通过选择合适的插值方 法和反向传播神经网络结构提高图像分辨率,取得 了较高的分类准确率;王士龙等<sup>[4]</sup>提取目标的边界 矩,利用改进的FCM聚类算法较好地实现了对水 下目标的分类识别。

虽然以上利用光图实现水下目标分类已经取 得了一定成果,但是受成像条件复杂性的影响,水下 光成像作用距离近,图像对比度较差。由于声呐的 探测距离较远,成像范围较大<sup>[5-6]</sup>,声成像技术的 不断发展使得利用声呐图像进行目标分析成为可 能<sup>[7-9]</sup>。Sinai等<sup>[10]</sup>利用C-V轮廓算法将声呐图像 分割为目标及阴影两部分,通过人工提取几何特征 来实现水下目标分类;Williams<sup>[11]</sup>利用卷积神经网 络将水声图像分为有目标和无目标两类,取得了满 意的效果;朱可卿等<sup>[12]</sup>使用高斯混合模型对声图 的阴影部分进行提取,设计融合分类器实现对水下 沉底小目标的分类,分类性能较好。

然而,受声呐自身技术参数的限制以及水下噪 声、混响的影响,声成像的清晰度较低,且声图获取 较为困难。现有水下目标分类方法的研究大多依靠 单一的光学或声学图像数据集,通过人工提取特征 或使用卷积神经网络完成对目标的分类。但是,单 一数据集对目标的描述具有一定限制,且卷积神经 网络模型较为复杂,网络的计算和分类速度较慢。 因此,如何降低模型复杂度,节约计算资源,获得更 好的分类效果,都是亟待解决的关键问题。

为了改善网络的分类性能,适应小样本背景下 的水下目标分类任务,主要从以下角度解决上述问 题。首先,选择轻量化的MobilenetV2网络并对其 结构进行改进,减小网络的参数量,进一步提高网络 运算效率。其次,在改进网络的基础上设计并行网 络结构,将采集的声、光学图像真实数据集同时输 入网络,采用中间层融合策略,利用融合特征得到最 终的分类结果。该方法规避了单一数据集对目标描 述的限制,充分利用声、光学两种图像各自的优势 以及MobilenetV2网络参数少、轻量化的特点,在节 约网络计算资源的同时,提高了算法的分类准确率。

# 1 改进MobilenetV2网络

#### 1.1 MobilenetV2网络模型

MobileNet是 Google于 2017年提出的新型轻量化网络<sup>[13]</sup>, MobileNetV2与其相比,具有较少的网络参数数量和更低的运算成本,相比普通的全卷积网络能够减少8~9倍的计算量,网络性得到了进一步改善,与VGG16等常用网络相比具有低消耗和实时性等优点,符合目标分类任务的要求<sup>[14]</sup>。MobileNetV2网络包括普通卷积(Conv)、反向残差结构的深度分离卷积(Bottleneck)和平均池化(Avgpool)几部分,网络结构如图1所示。





Bottleneck结构是 MobileNetV2 网络的核心部 分,每个 Bottleneck 由两个普通卷积和一个深度分 离卷积 (Dwise)组成<sup>[15]</sup>。该结构首先通过1×1的卷 积进行维度扩展,再用3×3的深度分离卷积提取特 征,最后使用1×1的卷积来压缩数据<sup>[16]</sup>,两个普通卷积分别使用ReLU6和Linear函数进行激活,深度分离卷积使用标准化BN层<sup>[17]</sup>和线性整流函数ReLU6<sup>[18]</sup>进行正则化和激活。图2(a)和图2(b)分

别表示步长为1和步长为2时的Bottleneck网络结构,当步长为1时,需要将该网络结构的输出与上一层的输出进行叠加,实现不同位置的信息整合。



图 2 Bottleneck 网络结构图

Fig. 2 Bottleneck network structure

# 1.2 改进的 MobileNetV2 网络

MobilenetV2网络使用ImageNet数据集进行 训练,数据集图片数量达到140万张,而水下目标 图像采集较为困难,获取的数据数量较少,直接使 用原网络进行训练无法得到较好的拟合效果,且 ImageNet数据集共包含图像1000个类别,目标种 类与水下目标差别悬殊,无法直接进行迁移学习。 MobilenetV2网络第9层的输出通道由320增加到 1280,通道数的增加会消耗更多的计算资源;网络使用平均池化 Avgpool 降采样来减少特征数量,更多地保留图像的背景信息,不完全适用于水下目标分类任务。

因此,为了增强网络在水下场景的适用性,充 分发挥深度分离卷积在特征提取中的优势,提高目 标分类的精确度,本文在MobileNetV2网络的基础 上进行了如下改进:(1)为了在训练网络时进一步 减少计算资源、节约内存空间,在保证精度的前提 下充分考虑参数量和运算成本,借鉴文献[14]的方 法,通过多次实验对比,本文去掉第9层之后的网络 层,并将该卷积层通道数由1280改为128。(2)为了 适应水下目标分类任务,在保留目标特征信息的同 时提升网络的收敛速率,本文使用Flatten层进行数 据降维,将三维的输出转化为一维后,添加Dropout 层改善网络拟合,丢弃率设为0.5,最后增加一个全 连接层,得到最终的分类结果。

图3为改进MobileNetV2网络的结构示意图, 网络包括特征提取和分类两个部分。特征提取网络 包括1个普通卷积、7个具有反向残差结构的深度 分离卷积和1个普通卷积,通过Flatten层将三维特 征图转换为一维后,使用Dropout层改善网络拟合。 分类网络使用全连接层结构,从而得到每一个目标 属于各个类别的概率。



## 图 3 改进 Mobilenetv2 网络结构图 Fig. 3 Network structure of improved Mobilenetv2

网络的参数量和计算量作为两个重要的指标, 通常用来对网络模型的复杂度进行评估,参数量对 应的是算法的空间复杂度,表示对计算机内存资源 的消耗;计算量对应时间复杂度,关系到网络的运算 时间。参数量和计算量主要来网络中的自卷积层和 全连接层,计算过程可以表示为

$$\begin{cases}
P_{\rm cnn} = \sum_{l=1}^{D_1} K_l^2 \cdot C_{l-1} \cdot C_l, \\
F_{\rm cnn} = \sum_{l=1}^{D_1} M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l,
\end{cases}$$
(1)

$$P_{\text{dense}} = F_{\text{dense}} = \sum_{l=1}^{D_2} C_{l-1} \cdot C_l, \qquad (2)$$

其中,P 和 F 分别代表模型的参数量和计算量,下标 $cnn和dense分别表示卷积层和全连接层,<math>M_l 和 K_l$ 分别表示输入图片的尺寸和网络使用的卷积核大 小, $C_{l-1} 和 C_l$ 为卷积运算中输入、输出特征图的通 道数, $D_1$ 、 $D_2$ 为网络中的卷积层与全连接层个数。 使用式(1)~(2)对原始及改进后的MobileNetV2 网络模型进行复杂度计算,计算结果如表1所示。

# 表1 原始及改进 MobileNetV2 网络复杂度 Table 1 Network complexity of original and improved Mobilenetv2

网络模型	参数量/M	计算量/M
MobileNetV2 网络	3.4	300
改进 MobileNetV2 网络	1.9	230

MobileNetV2网络模型需要训练的参数数量 约为3.4 M,改进后网络模型的参数数量约为1.9 M, 与原始网络相比,模型参数数量减少了近一倍。与 此同时,改进的MobileNetV2网络计算量为230 M, 相比原始网络的计算量也有一定数量的减少。由此 可见,改进后的网络复杂度有所减小,能够提高网络 运算效率,进一步节约计算资源。

# 2 声光图像融合分类网络

声学图像能够大范围获取,效率较高,光学图 像的高分辨率能够实现对目标细节的描述。为了实 现二者的优势互补,提出一种声光图像融合分类网 络模型。目前,对异源图像的联合处理网络主要有 输入前融合和输入后融合两种<sup>[19]</sup>,前者是将图像进 行融合处理后再输入特征提取网络,此种方式通常 需要改变第一层卷积的数量,使得训练结果变差;后 者是对图像进行特征提取之后,将特征提取网络的 中间层信息融合<sup>[20]</sup>,能够保证网络训练的准确性。

本文使用输入后融合的思想,将水下目标的声、 光两种图像并行输入网络进行特征提取,在某一层 将两个模块输出的特征图进行融合,实现两种图像的信息交流。特征提取使用的网络主干为1.2节中改进的MobileNetV2网络,在网络的特征提取过程中,图像的原始信息更多地体现在网络的浅层特征当中,网络的深层特征较为抽象,具有更多的分类信息。因此,根据网络的结构特点,本文选择在网络的深层位置进行特征融合,将网络最后一个卷积层的输出作为待融合特征,使用融合操作的结果实现水下目标分类,从而达到更高的分类准确率。

融合分类网络模型如图4所示,该网络由特征 提取、特征融合、融合特征提取、分类4个部分组成。 声学图像和光学图像分别送入改进的 MobileNetV2 网络,特征提取部分包括一个普通卷积、具有反向 残差结构的深度分离卷积及其之后的卷积层。在网 络的最后一个卷积层位置,将输出的特征图按通道 对应实现特征融合,这里应用的融合算法是通道拼 接 (concatenate),融合过程的数学表达式为

output =  

$$H_2(H_1^{\text{optical}}(X_{\text{optical}}), H_1^{\text{acoustic}}(X_{\text{acoustic}})),$$
 (3)

其中, X<sub>optical</sub> 和 X<sub>acoustic</sub> 表示输入的光学图像和声 学图像; H<sub>1</sub><sup>optical</sup> 和 H<sub>1</sub><sup>acoustic</sup> 表示光学图像和声学 图像从输入到最后一个卷积层之间的特征提取网 络; H<sub>2</sub> 代表融合操作的通道拼接算法; output 为融 合后输出的新特征, 用以实现目标分类。分类过程 可以表示为

$$K = S(\text{output}), \tag{4}$$

其中, K为分类结果, S代表 Softmax 分类函数, 将 网络最后一层的输出转化为输入图像属于各类别 的概率, 公式为

$$S_i = e^i \Big/ \sum_j e^j. \tag{5}$$





# 3 数据获取及预处理

水下目标声图由成像声呐获得,三维成像声呐 具有图像的深度信息,同时可以获得更清晰的目 标轮廓<sup>[21]</sup>,因此使用前视三维声呐获取声图。本 文所用的水下目标数据集由千岛湖实验获得,实验 装置布放如图5所示。数据采集装置由一个绿激光 水下摄像机和一个前视三维声呐组成,二者成对获 取水下目标图像。其中,绿激光水下摄像机型号为 WWA-6226,波长为532 nm,分辨率1920×1080,最 大可视范围设置为8 m;声呐设备为高频三维成像 声呐,工作频率为300 kHz,带宽30 kHz,波束开角 45°,波束数目为128×128,实验过程中最大工作深 度设置为30 m。

三维成像声呐得到的数据为三维图像,携带方 位、距离和散射强度等信息。数据获取过程中,受水 下复杂环境以及设备自身限制的影响,可能包含一 定的噪点,首先使用式(6)所示过程对归一化后的 原始数据进行滤波处理。

$$p_i = \begin{cases} p_i, & f(p_i) \ge k, \\ 0, & \notin \mathbb{H}, \end{cases}$$
(6)

其中,  $p_i = (x_i, y_i, z_i)$ 为三维图像中每个点的坐标,  $f(p_i)$ 为该点的散射强度, k为滤波阈值。将原始三 维图像及滤波后的图像以点云形式进行可视化, 效 果如图6所示。







图 6 水下三维图像可视化效果图 Fig. 6 Visualization of underwater 3D image



(b) 水下目标的声学图像图 7 部分实验数据集图像Fig. 7 Images of some experimental data sets

为了方便数据处理,将三维图像距离向上散射 强度最强的点投影到二维平面,用其深度值作为该 点的像素值,得到目标的深度图。将摄像机拍摄的 图像作为光学图像数据集,三维图像处理后得到的 深度图作为声学图像数据集。两个数据集中的图 像成对存在,数据格式为三通道RGB图像,共2196 张,包括铁框、蛙人、绳子、绳子拖曳的小球、桥墩5 类水下目标,部分数据集图像如图7所示。

# 4 实验结果与分析

为了验证所提网络在水下目标分类任务中的 可行性,设计了以下实验。首先采集水下目标图 像,制作实验数据集。(1)在光学图像数据集上进 行实验,对比VGG16、Resnet50、MobileNetV2以及 改进的MobileNetV2网络对5种水下目标的分类性 能,验证了改进MobileNetV2网络在水下目标分类 任务中的适用性; (2) 利用改进的 MobileNetV2 网 络以及提出的声光图像融合分类网络模型对水下 目标进行分类,验证融合网络结构对水下目标分类 准确率的提高;(3)使用不同融合算法在不同位置 进行融合,利用融合网络对目标图像进行分类,讨 论融合位置及融合算法对分类准确率的影响;(4)在 数据缺失的条件进行融合网络的分类实验,记录网 络的分类准确率,验证网络的鲁棒性。以下实验使 用的所有网络均基于 Keras 深度学习框架搭建,并 利用 CUDNN 进行加速处理。实验计算机 CPU 为6 核 i7-10750H、Win10 操作系统、GPU 为 RTX2070。

在采集的5类水下目标图像中,随机抽取20% 的目标数据作为测试集,余下的作为训练集。为了 确保目标分类的准确性,在抽取数据时需要将声学 图像和光学图像数据一一对应。训练集和测试集的 样本组成如表2所示。

#### 表2 训练集和测试集样本数量

Table 2Sample number of training setand test set

目标类别	训练集	测试集
铁框	211	53
蛙人	451	113
绳子	243	61
绳子+球	115	29
桥墩	736	184

# 4.1 不同分类网络在本文数据集上的分类性能

分别使用 VGG16、Resnet50、MobileNetV2 以 及改进的 MobileNetV2 网络对本文数据集 (以光学 图像为例)进行分类,验证几种分类网络对水下目 标的分类性能。将水下目标图像输入网络进行训 练和测试,在网络训练前,应用高斯分布 $G(\mu, \sigma^2)$ 对网络中的所有参数进行随机初始化,其中 $\mu = 0$ ,  $\sigma = 1$ 。采用 Optimizers 优化器对整个网络的参数 进行优化,学习率设为0.00001。在训练数据中每次 随机抽取 16 张图像训练网络,迭代次数为 100, 网络 采用代价函数选用分类交叉熵。几种模型的分类结 果可见表 3, 其中,分类时间是测试过程中对一张图 像得出分类结果所用的平均时间。

#### 表3 目标分类网络性能比较

Table 3Comparison of performance oftarget classification network

单通道网络	准确率/%	分类时间/ms	参数量/M
VGG16	93.7	12	65.08
Resnet50	80.3	7	45.75
Mobilenetv2	92.2	3	2.28
改进 Mobilenetv2	93.1	2	1.92

由表3可以看出,分类准确率最高的是VGG16 网络,可以达到93.7%,但是其分类时间最长,参 数量最大,计算代价较高。Resnet50网络的分类时 间和计算代价有所减少,但分类准确率较低,MobilenetV2网络与之相比准确率有所提高,在参数量 上也具有一定优势。综合考虑分类准确率、时间、参 数量几种指标,MobilNetV2网络要优于另外两种 网络模型。相比于原始MobilNetV2网络,改进后的 网络得到的分类准确率更高,且分类时间、参数量 均有减小,说明本文做出的改进对模型性能有一定 的提升,更加适用于水下目标分类任务。

#### 4.2 融合前后网络性能比较

分别使用改进的MobileNetV2网络对声学图 像和光学图像进行分类,之后应用本文提出的融合 分类网络将对应的声学图像和光学图像成对输入 网络进行训练和测试,对水下目标图像进行分类。 网络超参数的设置均与前述实验一致,训练过程曲 线如图8所示。



Fig. 8 Change curve of training process

由代价函数变化曲线可以看出,几种模型的函数值在整个训练过程中不断下降,最终都逐渐趋于 平稳,改进后的 MobileNetV2 网络在两个数据集上 具有较小的损失值,融合网络的损失值最小,相比 单通道的分类网络具有更好的性能。图8(b) 为训练 过程不同分类模型的分类准确率变化曲线,将测试 集输入训练好的网络进行分类,检验模型的分类性 能,改进前后与融合网络在本文数据集上的分类准 确率如表4所示。

由表4可以看出,改进MobilNetV2网络对两种图像的分类准确率分别为87.9%和93.1%,网络

#### 表4 不同网络的分类准确率

Table 4Classification accuracy of differ-<br/>ent networks

数据集	网络类型	准确率/%
声图	改进 MobilNetV2	87.9
光图	改进 MobilNetV2	93.1
声图 + 光图	本文声光融合网络	96.5

模型的拟合情况较好,能适应水下目标分类的小样本数据。本文提出的融合分类网络对水下目标图像的分类准确率达到96.5%,相比融合前的网络模型对声学图像和光学图像的分类准确率分别提高8.6%和3.4%,具有良好的分类性能。

# 4.3 不同融合方式下网络性能比较

本文在对声光图像进行融合时使用的算法是 通道拼接(concatenate),在网络的深层位置进行 特征融合。除此之外,融合算法还包括直接叠加 (add)、最大化运算(maximum)等,融合位置也可 以选在网络的浅层部分。分别使用add、maximum、 concatenate 三种融合算法,在改进MobileNetV2网 络的第一个卷积层后进行特征融合,将该融合网络 记为浅层融合网络,本文提出的融合网络记为深层 融合网络,利用声、光学图像分别在两种融合网络 上进行实验,训练过程代价函数及准确率的变化如 图9所示。

两种融合网络在不同的融合算法下代价函数 值均下降并收敛,融合算法的选择对网络性能的影 响不明显,深层融合网络的收敛速率更快。将测试 集数据输入训练好的融合分类网络,得到最终分类 结果,融合网络对水下目标的分类结果可见表5。

表5 不同融合网络对本文数据集的分类性能 Table 5 Classification performance of different fusion networks on the dataset in this paper

融合方式	准确率/%	分类时间/ms	参数量/M
浅层融合-add	95.4	3	1.91
浅层融合 -maximum	94.9	3	1.91
浅层融合-concatenate	95.7	3	1.92
深层融合-add	96.4	4	3.81
深层融合 -maximum	96.3	4	3.81
深层融合-concatenate	96.5	4	3.82

由表5可以看出,浅层融合和本文提出深层融合的平均准确率分别达到95.3%和96.4%,相比使用单通道网络对水下目标图像的分类准确率均有明显提高,深层位置的融合网络分类准确率更高。由于深层融合需要的参数更多,运算成本也略有增加。同一融合位置下不同融合算法的选择对分类结果的影响较小。



图 9 融合网络训练过程变化曲线 Fig. 9 Change curve of fusion network training process

## 4.4 融合网络鲁棒性实验

在实际水下应用中,由于水下摄像机和声呐的 作用距离不同,远距离条件下只有声呐能够获取到 有效数据,且水下环境的复杂性及实验过程的各种 不可控因素可能会导致其中一台设备无法正常工 作。考虑到最差的情况,假设其中一台设备先效,将 该设备获得的图像记为一个全0数组,另一设备获 得的为正常图像,将其输入本文的融合网络进行测 试,两种情况下不同数据集在网络中的分类结果如 表6所示。

#### 表6 不同数据集下的融合网络分类准确率

Table 6Classification accuracy of fusionnetworks under different data sets

数据集	准确率/%
仅声图 (光图缺失)	85.7
仅光图 (声图缺失)	90.3
声图 + 光图 (正常)	96.5

由表6可以看出,在光图缺失的情况下,融合网络的分类准确率为85.7%,声图缺失情况下的分类

准确率达到90.3%,相比正常数据下的分类准确率 有所降低,原因可能是融合后的特征图中包含无效 数据,对网络具有一定干扰作用。由此可见,当一台 设备失效时,融合网络的分类准确率会受到一定影 响,但是最低仍能达到85%以上,而普通的单路网 络在此种情况下会失去分类能力,说明本文提出的 融合网络具有一定的容错能力和鲁棒性。

# 5 结论

本文主要以下从两个角度改善网络的分类性 能,以适应小样本条件下的水下目标分类任务。首 先,将改进的MobilenetV2网络作为基础网络,以 减小网络训练过程的计算开销和内存占用,使用 实验采集的真实数据进行网络训练,改善网络拟合 效果。其次,将改进网络作为融合网络的分支,使 用中间层融合策略将水下目标的声、光学图像特 征图进行融合,实现各自的优势互补,进一步提高 分类准确率。实验结果表明,与其他常用分类模型 相比,MobilenetV2网络的分类准确率较提高,在 参数量上和分类时间上也具有一定优势。改进的 MobilenetV2网络与原网络相比,参数量及计算资 源消耗减少,分类准确率进一步提高,在水下目标分 类任务中具有更好的性能。相比融合前的网络,融 合网络模型的学习曲线收敛更快,且准确率更高,在 add、maximum、concatenate 三种融合算法下,融合 网络的分类准确率均有不同程度的提升。在单路数 据缺失的情况下,融合网络的分类准确率仍能达到 85%以上,具有一定的鲁棒性。

#### 参考文献

- 林森, 赵颖.水下光学图像中目标探测关键技术研究综述 [J]. 激光与光电子学进展, 2020, 57(6): 26-37.
   Lin Sen, Zhao Ying. Review on key technologies of target exploration in underwater optical images[J]. Laser & Optoelectronics Progress, 2020, 57(6): 26-37.
- [2] Gleason A C R, Shihavuddin A, Gracias N, et al. Improved supervised classification of underwater military munitions using height features derived from optical imagery[C]. OCEANS 2015-MTS/IEEE Washington, Washington, DC, USA, 2015: 1–5.
- [3] Pramunendar R A, Wibirama S, Santosa P I. Fish classification based on underwater image interpolation and back-propagation neural network[C]. 2019 5th International Conference on Science and Technology (ICST), Yogyakarta, Indonesia, 2019: 1–6.
- [4] 王士龙,徐玉如,万磊,等.基于边界矩和改进FCM聚类的水下目标识别[J].系统工程理论与实践,2012,32(12):2809-2815.

Wang Shilong, Xu Yuru, Wan Lei, et al. Underwater targets recognition based on contour moment and modified FCM algorithm[J]. System Engineering-Theory & Practice, 2012, 32(12): 2809–2815.

- [5] 孙宝申. 高分辨侧扫及合成孔径声纳图像与回波中类似水雷 目标计算机检测与分类 [J]. 应用声学, 2013, 32(4): 305–311. Sun Baoshen. Computer aided detection and classification of mine-like objects in high-resolution side scan and synthetic aperture sonar images and waveforms[J]. Journal of Applied Acoustics, 2013, 32(4): 305–311.
- [6] 刘纪元. 合成孔径声呐技术研究进展 [J]. 中国科学院院刊, 2019, 34(3): 283-288.
- [7] Raj M V, Murugan S S. Underwater image classification using machine learning technique[C]. 2019 International Symposium on Ocean Technology (SYMPOL), Ernakulam, India, 2019: 166–173.
- [8] Sperle M, Negri E, Ternes C. Automatic classification of sidescan sonar images for mapping marine mineral resources[C]. 2015 IEEE/OES Acoustics in Underwater Geosciences Symposium (RIO Acoustics), Rio de Janeiro, Brazil, 2015: 1–5.
- [9] Cho H, Gu J, Yu S. Robust sonar-based underwater object recognition against angle-of-view variation[J]. IEEE Sensors Journal, 2016, 16(4): 1013–1025.

- [10] Sinai A, Amar A, Gilboa G. Mine-like objects detection in side-scan sonar images using a shadows-highlights geometrical features space[C]. OCEANS 2016 MTS/IEEE Monterey, Monterey, CA, 2016: 1–6.
- [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), Cancun, 2016: 2497–2502.
- [12] 朱可卿,田杰,黄海宁.水下高分辨率声图中小目标的深度网络分类方法 [J]. 声学学报, 2019, 44(4): 595-603.
  Zhu Keqing, Tian Jie, Huang Haining. Underwater objects classification method in high-resolution sonar images using deep neural network[J]. Acta Acustica, 2019, 44(4): 595-603.
- [13] Howard A G, Zhu M, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. IEEE Access, 2017, 6: 1–14.
- [14] 孟琭, 徐磊, 郭嘉阳. 一种基于改进的 MobileNetV2 网络语义 分割算法 [J]. 电子学报, 2020, 48(9): 1769–1776.
  Meng Lu, Xu Lei, Guo Jiayang. Semantic segmentation algorithm based on improved MobileNetV2[J]. Acta Electronica Sinica, 2020, 48(9): 1769–1776.
- [15] Sandler M, Howard A, Zhu M, et al. MobileNetV2: inverted residuals and linear bottlenecks[C].
  2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 4510–4520.
- [16] Bousbai K, Merah M. A comparative study of hand gestures recognition based on MobileNetV2 and ConvNet models[C]. 2019 6th International Conference on Image and Signal Processing and their Applications (ISPA), Mostaganem, Algeria, 2019: 1–6
- [17] Thakkar V, Tewary S, Chakraborty C. Batch normalization in convolutional neural networks—A comparative study with CIFAR-10 data[C]. 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), Kolkata, India, 2018: 1–5.
- [18] Hoang V, Hoang V, Jo K. Realtime multi-person pose estimation with RCNN and depthwise separable convolution[C]. 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh City, Vietnam, 2020: 1–5.
- [19] 康硕, 柯臻铮, 王璇, 等. 基于红外和可见光图像融合的铺丝 缺陷检测方法 [J]. 航空学报, 2022, 43(3): 425187. Kang Shuo, Ke Zhenzheng, Wang Xuan, et al. Detection method of defects in automatic fiber placement based on infrared and visible image fusion[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(3): 425187.
- [20] Wolpert A, Teutsch M, Sarfraz M S, et al. Anchor-free small-scale multispectral pedestrian detection[J]. arXiv: 2008.08418.
- [21] 刘天宝, 吴晓潭, 黄勇, 等. 三维成像声呐图像重建研究 [J]. 声 学技术, 2015, 34(4): 358–361.
  Liu Tianbao, Wu Xiaotan, Huang Yong, et al. Image reconstruction for 3D acoustical imaging sonar[J]. Technical Acoustics, 2015, 34(4): 358–361.