

[引用格式] 杜雪, 廖泓舟, 张勋. 基于深度卷积特征的水下目标智能识别方法[J]. 水下无人系统学报, 2019, 27(3): 260-265.

# 基于深度卷积特征的水下目标智能识别方法

杜 雪, 廖泓舟, 张 勋

(哈尔滨工程大学 海洋装置与控制技术研究所, 黑龙江 哈尔滨, 150001)

**摘 要:** 无人水下航行器(UUV)由于本身的便利性和自主性在水下探测方面相比传统探测具有很大优势, 对 UUV 水下目标智能识别方法的研究具有重要意义。针对水下环境的高噪声、低对比度的特点, 文中首先使用中值滤波和局部增强处理进行水下图像预处理, 基于水下图像的小样本特点, 提出借鉴牛津大学视觉几何组网络(VGGNet)的逐层递增卷积层思想, 利用深度卷积神经网络(DCNN)设计水下智能识别框架并利用大数据集 Cifar-10 进行一次训练, 以学习图像通用特征; 同时使用迁移学习和数据增强技术进行二次学习, 以学习水下目标特有特征, 解决水下数据集不足的情况, 防止过拟合。通过仿真试验进行识别效果验证, 仿真结果表明, 在特定测试集下提出的水下目标智能识别方法在识别效果与自动化程度方面相比传统识别算法具有明显优势。

**关键词:** 无人水下航行器; 智能识别; 深度卷积神经网络; 迁移学习

**中图分类号:** TJ6; TP391.41; TP183

**文献标识码:** A

**文章编号:** 2096-3920(2019)03-0260-06

**DOI:** 10.11993/j.issn.2096-3920.2019.03.004

## Underwater Target Recognition Method Based on Deep Convolution Feature

DU Xue, LIAO Hong-zhou, ZHANG Xun

(Marine Equipment and Control Technology Research Institute, Harbin Engineering University, Harbin 150001, China)

**Abstract:** With the help of artificial intelligence technology, the intelligent recognition technology for unmanned undersea vehicle(UUV) to detect underwater target is presented. According to the characteristics of high noise and low contrast in underwater environment, median filtering and local enhancement are adopted to preprocess underwater images. Based on the characteristics of small sample of underwater images, the idea of incremental convolution layer by layer is proposed drawing from the visual geometry group net(VGGNet), and an underwater intelligent recognition framework is designed by using deep convolution neural network (DCNN) for training on the large data set Cifar-10 and learning general features of the image. In addition, migration learning and data enhancement technology are used for transferring knowledge to learn the unique features of underwater targets, so as to solve the problem of insufficient underwater data sets and to prevent over-fitting. Simulation results show that the proposed underwater target intelligent recognition technology has obvious advantages over the traditional algorithm in recognition effect and automation degree on specific test set.

**Keywords:** unmanned undersea vehicle(UUV); intelligent recognition; deep convolution neural network(DCNN); migration learning

收稿日期: 2018-09-30; 修回日期: 2018-12-25.

基金项目: 国家青年自然科学基金项目资助(51709062); 哈尔滨市科技创新人才研究专项基金资助(2017RAQXJ150); 中央高校基本科研专项基金(HEUCFM180402).

作者简介: 杜 雪(1987-), 女, 博士, 副教授, 主要研究方向为水下无人航行器协同导航与智能控制.

## 0 引言

水下无人航行器(unmanned undersea vehicle, UUV)作为各种水下任务的辅助手段,可避免作业人员在较危险的水下环境中工作,且能很好的适应水下环境和开展水下任务<sup>[1-2]</sup>。在民用领域,其可用于水下电缆铺设、水下管道质量检测、水下设备的维护与维修、渔场监测等,例如美国 Ocean Seaver 公司生产的 Iver2 UUV 已经被广泛用于环境监测和海洋勘探等领域<sup>[3]</sup>;在军用领域则可用于侦察、布雷、灭雷和援潜救生等,比如美国海军著名的先进无人搜索系统(advanced unmanned search system, AUSS)能潜入海底 6 100 m,自动搜索水下目标并通过水下摄像机进行拍照,实现对水下军事目标的搜索和跟踪任务<sup>[4]</sup>。这些都是依据 UUV 对水下目标的识别而完成的。

目前水下目标的识别主要可以通过声视觉和光视觉 2 种方式。其中,声视觉系统在远距离目标识别跟踪上占有绝对优势,但由于受到声波的盲区和多通道效应的影响,在近距离目标识别领域的发展受到限制;相反,UUV 可通过光视觉充分利用人工智能技术进行水下目标的探测和描述,逐渐成为近距离水下目标识别跟踪的主要研究方向<sup>[5]</sup>。传统的水下目标识别主要分为被识别信号获取、信号特征提取、目标识别 3 个过程<sup>[6]</sup>。其中的特征提取是一个既费时费力又需大量经验的繁琐处理过程,整个水下目标识别系统属于一个半自动化过程,并于近年向全自动识别方向发展<sup>[7]</sup>。基于此,对水下智能目标识别方法的研究具有重要意义。借助于人工智能技术的蓬勃发展,将基于深度学习的图像识别技术引入水下,构建水下端到端的通用图像智能识别技术,提高 UUV 在“看”方面的智能化水平,使 UUV 学会“看得更多,看得更准”,将对 UUV 的智能化发展有着重大意义,同时在军民融合的时代背景下,能够进一步促进 UUV 的多用途化。

近几年,国内外学者开始将深度学习技术引入水下进行目标智能识别。国外的公开研究主要是将深度学习技术用于水下声呐图像识别,如 Kim 等<sup>[8]</sup>提出了将深度卷积神经网络用于前视声呐图像,对声呐图像中的遥控水下航行器(rem-

otely operated vehicle, ROV)进行识别跟踪,显著提高了识别准确度;Matias<sup>[9]</sup>也将深度卷积神经网络用于前视声呐中多种目标物的检测和分类,从而达到水下多目标检测分类的最好水平。

国内的研究主要是海洋科学领域,比如渔场监控任务,通过视频采集技术借助深度学习理论对海洋生态系统进行无干扰监控和测量。Li 等<sup>[10-11]</sup>利用深度卷积神经网络对水下自由游动的鱼类进行了准确的检测与分类,实现了无干扰情况下对海洋生态的监测与调查研究,但其使用的网络结构并非真正意义上的深度网络,此外可用的数据量需足够大。同时,国内对声呐图像进行深度识别也在尝试中,Zhu 等<sup>[12]</sup>利用深度卷积神经网络对声呐图像进行特征提取,再通过支持向量机(support vector machine, SVM)进行分类任务,试验结果表明效果优于传统方向梯度直方图(histogram of oriented gradient, HOG)方法。王强等<sup>[13]</sup>将卷积神经网络用于水下 3 类实测目标噪声数据的分类识别,并与传统的 SVM 方法对比,可知 2 种方法所得结果比较接近。但以上研究存在以下问题:一是将深度学习用于水下目标识别方法尚不成熟,国内外的研究都是将其作为应用于水下的一种新尝试,尚无完整的框架或理论;二是国内外在声目标方面研究较多,但相关水下光视觉图像的深度学习智能识别算法研究较少。针对以上问题,文中从水下光视觉图像出发,提出了一套水下光视觉图像智能识别框架并进行了仿真验证,结果表明,在特定测试集下提出的水下目标智能识别方法在识别效果与自动化程度方面相比传统识别算法具有明显优势。

## 1 水下图像预处理

### 1.1 图像滤波去噪

水下光学图像相比水上光学图像具有更高的椒盐噪声,文中先使用基于空域的中值滤波(median filter)进行图像去噪处理,该方法对于非线性信号的处理有很好的效果,运算快,效率高,中值滤波去噪原理如下。

1) 中值输出:假定数组  $x_1, x_2, \dots, x_n$ , 并从大到小将这  $n$  个数值依次排列

$$x_1 \leq x_2 \leq \dots \leq x_n \quad (1)$$

数组  $x_1, x_2, \dots, x_n$  的中值为

$$y = \text{Med}\{x_1, x_2, \dots, x_n\} = \begin{cases} x_{[(n+1)/2]} & n \text{ 为奇数} \\ [x_{(n/2)} + x_{(n/2+1)}]/2 & n \text{ 为偶数} \end{cases} \quad (2)$$

2) 滤波窗口: 一维情况下可直接用  $y$  代替滤波窗口的中心值。但因图像为二维, 故需将窗口扩展到二维, 用二维数组来表示数字图像中各点的灰度值, 因此滤波窗口为二维的中值滤波定义为

$$y_{i,j} = \text{Med}_A\{x_{i,j}\} = \text{Med}\{x_{i+r,j+s}, (r,s) \in A(i,j) \in I^2\} \quad (3)$$

3) 窗口尺寸: 对于实际问题, 不同的图像常常需要使用不同形状和大小的窗口来处理。二维中值滤波的窗口形状有很多种, 根据图像的不同, 窗口形状也大相径庭。经过多次试验发现, 文中窗口尺寸选择为  $5 \times 5$  比较合适。

## 1.2 图像局部增强

水下拍摄时, 光的散射会对水下图像产生很大影响, 此外湍流、气泡和浮游生物等随机因素也会影响水下图像的质量。水下图像局部增强是指通过提高图像对比度、亮度及清晰度等, 增加图像视觉效果, 增强图像质量。对比度有限的自适应直方图均衡(contrast limited adaptive histogram equalization, CLAHE) 算法<sup>[14]</sup>是直方图均衡化(HE)算法和自适应直方图均衡(AHE)算法的集成与改进, 水下图像局部增强效果较好, 因此, 文中使用 CLAHE 算法对图像进行局部增强处理, 为后续的智能识别做铺垫。

## 2 水下目标智能识别方法

### 2.1 水下图像智能识别原理

水下图像智能识别的核心是深度学习算法, 文中使用的是深度卷积神经网络。深度学习已在陆上图像识别领域快速发展, 各种人工智能(artificial intelligence, AI)应用纷纷落地, 技术的快速发展既取决于深度学习理论不断发展, 也得益于各个领域数据的常年积累, 出现了各种各样的大数据集。目前大部分 AI 在某种层面上可以说是一种数据驱动型算法。而水下图像虽然种类多, 但单一图像数据量少, 普遍存在数据集不足的情况。为了解决这一问题, 提出使用迁移学习(transfer learning, TL)和数据增强技术防止产生少量数据集过拟合效应。其中, TL 的目标是将从一个

环境中学到的知识用来帮助新环境中的学习任务<sup>[15-16]</sup>, 其提供了一种解决数据不足问题的新方法, 允许迁移已有的知识来解决目标领域中仅有少量已标注样本数据的学习问题, 也就是运用已有的知识对不同但相关领域问题进行求解。迁移学习可以定义为: 给定一个源领域  $D_S$  和学习任务  $T_S$ , 一个目标领域  $D_T$  和学习任务  $T_T$ , 其中  $D_S \neq D_T$  和  $T_S \neq T_T$ , 迁移学习的目的是使用  $D_S$  和  $T_S$  中的已有知识, 帮助提高目标领域  $D_T$  中目标预测函数  $f_T(\cdot)$  的学习<sup>[17-18]</sup>。数据增强是对数据进行随机上下左右翻转、颜色、光照变换和多尺度变换, 可以有效增加水下数据集的数量。

### 2.2 水下图像智能识别框架

水下识别框架分为 2 个部分, 一是 Cifar-10 数据集训练设计好的深度卷积神经网络; 二是权值迁移, 增加一个全连接层并按分类结果改变输出神经元个数, 进行水下图像二次训练并识别。水下图像智能识别框架如图 1 所示。

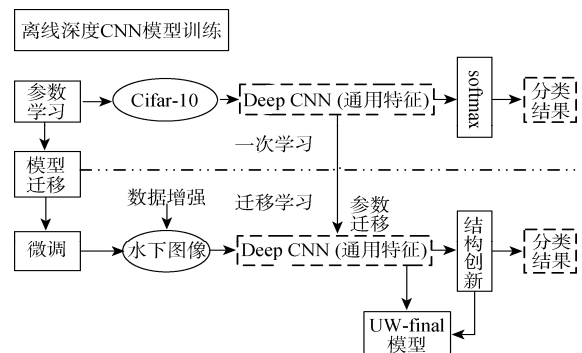


图 1 水下图像智能识别框架  
Fig. 1 Intelligent recognition framework of underwater images

### 2.3 深度卷积神经网络设计方法

水下深度神经网络框架使用 TensorFlow 进行设计, TensorFlow 是一种采用数据流图用于数值计算的开源库, 其计算通过“节点”和“线”来描述。

深度神经网络最显著的特点在于网络中可用于学习的参数量相比传统神经网络成倍剧增, 更多的参数意味着更强的学习能力, 也意味着更大的学习量。水下深度神经网络的设计思想借鉴牛津大学视觉几何组网络(visual geometry group net, VGGNet)。因此, 硬件计算能力和网络学习所需样本量也成倍剧增达到与之相匹配的程度, 在深度网络中样本不足会导致明显的过拟合效应。对

于特征提取能力强的卷积网络, 网络深度的增加体现在卷积核的骤增, 而实际操作中可以在某一层上增加卷积核个数即增加网络宽度, 也可以通过增加网络层数来增加卷积核个数, 即增加网络深度。由于卷积核按照局部感受野提取特征, 一个卷积核提取一种相应特征, 增加网络宽度能在一定程度上提高特征提取的能力, 但只在宽度上进行扩展, 特征量虽多但是比较具体, 如点、线和边等具体特征, 而卷积层之间的映射相当于特征之间的融合, 随着深度的扩展, 底层特征经过逐层映射, 特征越来越抽象, 更利于目标的识别。因此, 深度卷积神经网络(deep convolutional neural networks, DCNN)的设计原则应遵循深度扩展优于宽度扩展, 同时增加每层宽度的原则。

卷积核大小的增加只是深度卷积的一个方面, 卷积核的尺寸对网络的特征提取能力也有着重要的影响, 传统 DCNN 中常用的卷积尺寸有  $5 \times 5$ 、 $7 \times 7$  等, 不同尺寸的卷积核在不同的问题上有不一样的识别效果。随着网络深度和卷积核个数的增加, 小尺寸卷积核更适用于深度网络, 因为 1 个  $5 \times 5$  的感受野等于 2 个  $3 \times 3$  感受野, 如图 2 所示。因此, 1 个  $5 \times 5$  的卷积层可以被 2 个  $3 \times 3$  的卷积层替代, 网络信息传递一样的情况下网络的非线性得到增强, 同时降低了网络参数, 类似的  $7 \times 7$  的卷积层可以被 3 个  $3 \times 3$  的卷积层代替。

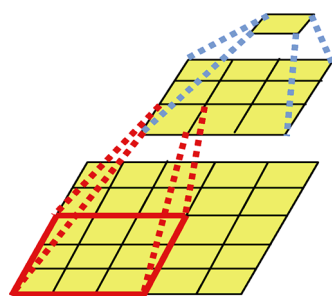


图 2 感受野等价图

Fig. 2 Equivalent graphs of receptive field

DCNN 的另一个特殊结构是池化层, 其作用主要是降低特征图大小来简化网络计算复杂度和进行特征压缩来提取主要特征。实际工程经验表明, 池化层紧跟卷积层效果比较好, 平均池化和最大值池化最常用, 二者无显著差别。

DCNN 在数据量不匹配时容易出现过拟合效应, 采用 Dropout 策略和  $L1$ 、 $L2$  正则化能在一定

程度上降低深度网络的过拟合效应。此外, DCNN 由于网络参数的骤增和样本分布的差异性易导致网络难以训练, 采用批归一化(batch normalization, BN)方法能加快网络的收敛性。

综上所述, 水下深度神经网络的设计过程遵循优先进行深度扩展, 即增加卷积核的层数, 其次进行宽度扩展, 即增加各层的卷积核个数。卷积核尺寸优先选择小尺寸, 如  $3 \times 3$ 。数据集足够的情况下可以考虑增加  $1 \times 1$  卷积核来增强网络的非线性, 以便获得更好的学习特征。深度网络设计的优劣以网络深度增加过程中一次训练识别率的高低为基准, 然后进行二次迁移学习, 以最终的目标识别率为评价标准。

## 2.4 试验验证

采用 Cifar-10 作为源领域数据集, 共 60 000 张  $32 \times 32$  彩色图片, 分为 10 个类别, 其中 50 000 张训练图片, 10 000 张测试图片, 如图 3 所示。目标领域数据集是 5 类不同水下鱼类资源实时视频, 分别是鼻鱼、豆娘鱼、蝴蝶鱼、刺尾鲷和立旗鲷, 资源来自于台湾鱼类资料库, 实际过程将视频按帧截取为静态图片, 为了保持样本平衡, 每一类截取 300 张训练图片, 100 张测试图片, 训练图片经过随机翻转、缩放及对比度增强等数据增强手段得到大量扩充。原始采集的视频图像如图 4 所示。

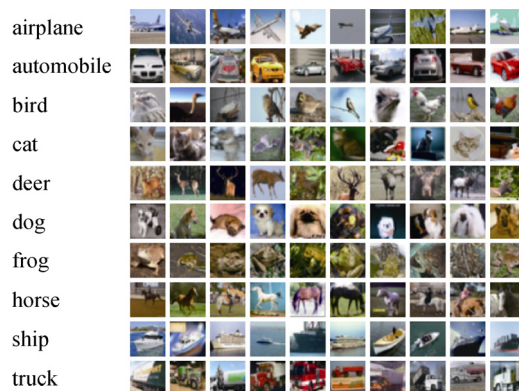


图 3 源领域数据集(Cifar-10)

Fig. 3 Dataset in source domain(Cifar-10)

首先对上述目标领域数据集进行图像滤波和局部增强预处理, 将其作为二次迁移训练的输入数据。按照 3.3 节的设计原则进行网络设计, 同时水下目标识别初始模型应该与 Cifar-10 数据集相匹配, 随着网络层数的递增, Cifar-10 训练集准确率逐渐增加, 当超过网络 10 层后训练准确率达到

100%，并不再变动，出现了过拟合现象。为了防止过拟合，水下目标识别初始模型层数设计为 10 层。一次训练过程中 DCNN 结构如表 1 所示。



图 4 目标领域数据集(5 类)  
Fig. 4 Dataset in target domain(five categories)

表 1 深度卷积神经网络参数列表		
Table 1 Parameter list of deep convolution neural network (DCNN)		
层名	参数	属性
层 1	Conv1+BN	1) Conv1: 3 × 3
	Relu1	2) Conv1 大小:64
	MaxPool1	3) pool1: 2 × 2, 2
层 2	Conv2+BN	1) Conv2: 3 × 3
	Relu2	2) Conv2 大小:128
	MaxPool2	3) pool2: 2 × 2, 2
层 3	Conv3+Conv4	1) Conv3/4: 3 × 3
	BN	2) Conv 大小: 256
	Relu4	3) pool4: 2 × 2, 2
层 4	Conv5+Conv6	1) Conv5/6: 3 × 3
	BN	2) Conv 大小: 512
	Relu6	3) pool6: 2 × 2, 2
层 5	Conv7+Conv8	1) Conv7/8: 3 × 3
	BN	2) Conv 大小: 512
	Relu8	3) pool8: 2 × 2, 2
层 6	FC9	—
	Dropout1	
层 7	Relu9	—
	FC10	
	Dropout2	
层 7	Relu10	—
	Relu10	

使用源领域数据集训练 10 层 DCNN，实际操作中迭代次数越长效果越好，使用 Relu 激活函数收敛性最快，但是效果不一定最好，相反传统的 Tanh 激活函数在更长的迭代次数下训练效果更好但不容易收敛，二者各有优势。试验中迭代次

数为 1 万次时损失值不再显著变化终止训练，预训练完成，将网络权重和模型进行迁移，增加一个全连接层用于目标识别，根据识别目标种类的多少将输出神经元由 10 个改为 5 个，目标领域训练集按批进行数据增强后输入 DCNN 进行训练。为了加快训练速度，训练过程中每迭代 100 次进行 1 次日志记录，训练过程中部分可视化结果如图 5~图 7 所示。图中，最大值和最小值分别指网络中间层的最大和最小输出值。

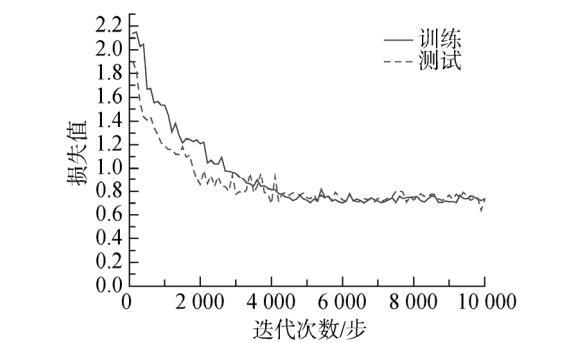


图 5 训练过程误差损失函数变化曲线  
Fig. 5 Curves of error loss function in training process

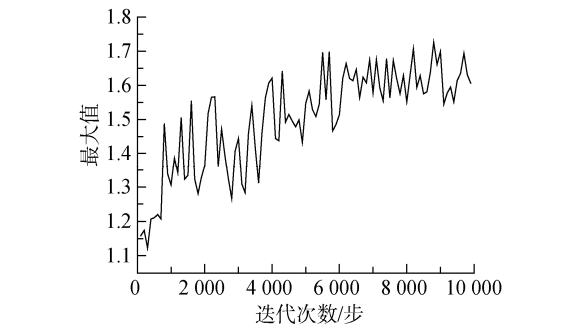


图 6 训练过程卷积层 3 最大输出值变化量  
Fig. 6 Maximum output value change of convolution layer 3 in training process

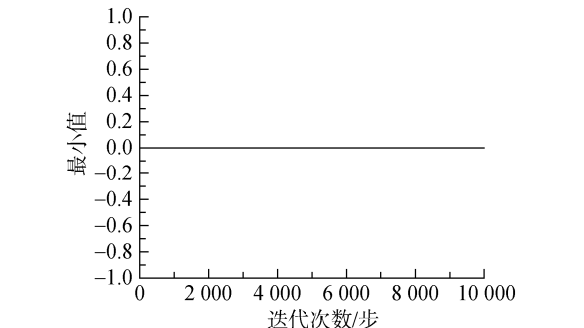


图 7 训练过程卷积层 3 最小输出值变化量  
Fig. 7 Minimum output value change of convolution layer 3 in training process



图5是训练过程中误差损失函数变化趋势,图6和图7是训练过程中任意选择的网络层(试验随机选择为第3层)的最大权重和最小权重的变化过程,由图5~图7可发现,随着迭代次数的增加,训练集和测试集的损失函数都呈现明显的下降趋势并且损失值在5000步左右开始稳定,证明网络训练效果好,同时随机抽取的第3层网络的最大输出值趋于波动上升中,证明权重得到有效更新,进一步佐证了网络训练的有效性,同时由于网络使用了relu激活函数,最小输出值一直维持为0。为验证识别效果,同时将提出的水下图像智能识别方法与传统HOG、尺度不变特征转换(scale-invariant feature transform, SIFT)算法进行粗略比较,HOG在60%左右,SIFT在62%左右,水下图像智能识别方法准确率在78%左右,准确率明显提高,可见基于深度卷积特征的智能识别方法在识别效果与自动化程度方面都具有明显的优势。

### 3 结束语

采用深度学习方法对水下图像进行智能识别,提高了UUV在“看”方面的智能化水平,对UUV的智能化发展有着重大的意义,同时在军民融合的时代背景下,促进了UUV的多用途化。

文中针对水下图像高噪声、低对比度的特点,先进行水下图像预处理,面对水下数据集不足的情况,提出了迁移学习与数据增强方法,试验验证了水下智能识别方法的有效性,在一定程度上优于传统的水下图像识别算法,但是提出的识别方法在准确率上还有很大上升空间,这是一次水下图像智能识别的崭新尝试,而寻求更长的训练时间,更多的水下数据,更加平衡的样本都将在工程意义上提高水下智能识别方法的准确率,这也是未来研究的主要目标。

#### 参考文献:

- [1] Zhang W K, Wang G X, Xu G H, et al. Development of Control System in Abdominal Operating ROV[J]. Chinese Journal of Ship Research, 2017, 12(2): 124-132.
- [2] Apgar J F, Edelson J S, Furtado J, et al. ORCA-IX: An Autonomous Underwater Vehicle[D]. Massachusetts Institute of Technology. Massachusetts, US: Massachusetts Institute of Technology, 2006.
- [3] Zhang L, Jiang D P, Zhao J X. The Basic Control System of an Ocean Exploration AUV[J]. Applied Mechanics and Materials, 2013(411/412/413/414): 1757-1761.
- [4] Ferri G, Munafo A, Lepage K D. An Autonomous Underwater Vehicle Data-Driven Control Strategy for Target Tracking[J]. IEEE Journal of Oceanic Engineering, 2018, 99: 1-21.
- [5] Jaffe M. Underwater Optical Imaging: Status and Prospects[J]. Oceanography, 2001, 14(3): 66-76.
- [6] Lin Y, Lv F, Zhu S, et al. Large-scale Image Classification: Fast Feature Extraction and SVM Training[C]/IEEE Conference on Computer Vision and Pattern Recognition, Providence. Colorado: IEEE, 2011: 1689-1696.
- [7] Wolek A, Dzikowicz B R, McMahon J, et al. At-Sea Evaluation of an Underwater Vehicle Behavior for Passive Target Tracking[J]. IEEE Journal of Oceanic Engineering, 2018, 99: 1-10.
- [8] Kim J, Cho H, Pyo J, et al. The Convolution Neural Network Based Agent Vehicle Detection Using Forward-looking Sonar Image[C]/2016' IEEE/OES Oceans. Shanghai: IEEE, 2016: 1-5.
- [9] Matias V T. End-to-end Object Detection and Recognition in Forward-looking Sonar images with Convolutional Neural Networks[C]/2016' IEEE/OES Autonomous Underwater Vehicles. Tokyo: IEEE, 2016: 144-150.
- [10] Qin H, Li X, Liang J, et al. Deep Fish: Accurate Underwater Live Fish Recognition with a Deep Architecture[J]. Neurocomputing, 2016, 187: 49-58.
- [11] Li X, Shang M, Hao J, et al. Accelerating Fish Detection and Recognition by Sharing CNNs with Objectness Learning [C]/2016' IEEE/OES Oceans. Shanghai: IEEE, 2016: 1-5.
- [12] Zhu P, Isaacs J, Fu B, et al. Deep Learning Feature Extraction for Target Recognition and Classification in Underwater Sonar Images[C]/2017' IEEE 56th Annual Conference on Decision and Control (CDC). Melbourne: IEEE, 2017: 2724-2731.
- [13] 王强, 曾向阳. 深度学习方法及其在水下目标识别中的应用[C]/中国声学学会水声学分会 2015 年学术会议论文集. 武汉: 中国声学学会, 2015.
- [14] Zuiderveld K. Contrast Limited Adaptive Histogram Equalization[M]. Graphic Gems IV. San Diego: Academic Press Professional, 1994: 474-485.
- [15] Pan S J, Yang Q. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge & Data Engineering, 2010, 22(10): 1345-1359.
- [16] Cheng P M, Malhi H S. Transfer Learning with Convolutional Neural Networks for Classification of Abdominal Ultrasound Images[J]. Journal of Digital Imaging, 2017, 30(2): 234.
- [17] 龙明盛. 迁移学习问题与方法研究[D]. 北京: 清华大学, 2014.
- [18] Li X X, Deng Z H, Choi K S. Generalized Hidden-Mapping Transductive Transfer Learning for Recognition of Epileptic Electroencephalogram Signals[J]. IEEE Transactions on Cybernetics, 2018, 99: 1-15.

(责任编辑: 杨力军)