

基于迁移学习的水声目标识别^①



邓 晋¹, 潘安迪¹, 肖 川¹, 刘珊珊²

¹(复旦大学 计算机科学技术学院, 上海 201203)

²(中国电子科技集团公司第二十三研究所, 上海 200437)

通讯作者: 邓 晋, E-mail: dengjin1995@gmail.com

摘 要: 海洋声场环境的时变空变特性、水声目标发声机理的多源性以及其他噪声源的干扰, 给水声目标的检测和识别带来很多困难. 常规的目标识别手段主要是基于音频时频域特征分析, 在复杂海洋环境下的难以获取有效的表征特征及鲁棒的识别效果. 为了解决这些问题, 本文提出了基于迁移学习的水声目标识别, 分别利用预训练网络 VGG 和 VGGish 提取深层声学特征及模型微调, 实现水声目标的分类识别. 实验表明, 本文提出的识别算法有效提升了识别准确率, 减少了训练时间, 基于微调的迁移学习算法在水声目标识别上平均准确率为 92.48%, 取得了当前最好的识别结果.

关键词: 迁移学习; 水声识别; 信号特征表征; 音频分类; 舰船噪声

引用格式: 邓晋, 潘安迪, 肖川, 刘珊珊. 基于迁移学习的水声目标识别. 计算机系统应用, 2020, 29(10): 255–261. <http://www.c-s-a.org.cn/1003-3254/-7538.html>

Transfer Learning for Acoustic Target Recognition

DENG Jin¹, PAN An-Di¹, XIAO Chuan¹, LIU Shan-Qi²

¹(School of Computer Science, Fudan University, Shanghai 201203, China)

²(The 23rd Research Institute of China Electronics Technology Group Corporation, Shanghai 200437, China)

Abstract: The time-varying and space-varying characteristics of the marine sound field environment, the multi-source nature of the sound mechanism of underwater acoustic targets, and interference from other noise sources have brought many difficulties to the detection and identification of acoustic targets. Conventional target recognition methods are mainly based on the audio time-frequency domain analysis, it is difficult to obtain effective features and robust recognition effects. In order to solve these problems, transfer learning based acoustic target recognition is proposed. The pre-trained networks VGG and VGGish are used to extract deep acoustic feature analysis and fine-tune respectively. Experiments show that the proposed algorithm effectively improves the recognition accuracy and reduces the training time. The fine-tuned transfer learning algorithm has an average accuracy rate of 92.48% in acoustic target recognition, which achieved the state-of-the-art recognition result.

Key words: transfer learning; acoustic recognition; signal feature characterization; audio classification; ship noise

水声目标识别技术是一种利用声呐接收的被动目标辐射噪声、主动目标回波以及其他传感器信息提取目标特征并判别目标类型或舰型的信息处理技术^[1]. 水声目标识别包括声呐技术、环境感知计算、监听技

术^[2]等. 在水声领域, 目标分类与识别是国际公认的难题, 目前尚无有效的方法实现水声目标的准确识别. 海水是声音在水下传播的介质, 但海水的分层特性、时变特性以及区域不均匀性, 使声音在船舶过程中受信

① 基金项目: 国家自然科学基金 (61671156)

Foundation item: National Natural Science Foundation of China (61671156)

收稿时间: 2019-12-21; 修改时间: 2020-01-19; 采用时间: 2020-02-25; csa 在线出版时间: 2020-09-30

道畸变、混响和杂波的影响,导致信号强度衰减,特征被掩盖,进而给水声目标特征识别和分类带来很大的挑战。远航、风、波、洋湍流提高了海洋环境中噪声水平,重要信号被遮蔽,使目标信号的探测和识别性能严重下降^[3]。因此,如何有效探测和估计长距离低信噪比下的目标、以及从船舶辐射噪声信号中提取特征^[4],是水下定位、识别和跟踪的迫切问题,实际上,水声目标识别一直是国际公认的难题。

本文的组织结构如下:第1节介绍基于传统声学特征水声目标识别概况;第2节介绍基于迁移学习的水声目标识别背景及方法;第3节实验结果及分析;第4节文章总结及未来工作展望。

1 基于传统声学特征的水声目标识别

水中目标辐射噪声组成复杂,主要声源包括:机械噪声、螺旋桨噪声和水动力噪声^[5]。一般的水声目标识别都是基于人工特征的,通常包括:时频特征^[2,6,7]、线谱^[4,8-11]、DEMON谱^[5]等等。在人工提取特征后使用统计学模型、模式分类或专家系统等方法进行分类识别。算法1为基于传统声学特征的水声目标识别算法。

算法1. 基于传统声学特征的水声目标识别算法

- 1) 预处理,将水听器阵列采集到的声呐信号通过信号增强,提高信噪比。
- 2) 特征提取,提取帧级别的时域或频域信息,诸如:线谱分布、连续谱能量谱分布、特定频段线谱数量、LOFAR谱和DEMON等。
- 3) 分类识别,基于水声目标特征的分类识别方法主要有统计分类、模型匹配、神经网络和专家系统等方法。

在过去的20年中,通过发展先进的水声信号处理技术,对舰船辐射噪声的特征提取方式进行了广泛多样的探索^[12]。这些方法里面比较经典的有,利用振荡性质, Duffing 振子^[10]和随机共振理论^[11]检测船舶辐射噪声的线谱。由于音频信号的非平稳特性,时频分析技术更适用于非平稳信号,这种方法的优点是能结合时域中的非平稳信息和频域信息,例如短时傅立叶变换(STFT)^[6,7],小波变换(WT)^[6,13,14]和 Hilbert-Huang 变换(HHT)^[15,16]。考虑到船舶辐射噪声的非线性特性,许多方法被用来提取非线性特征,包括相空间重构^[17,18],基于信号调制的 DEMON 谱分析^[5],基于序贯检测思想的 LOFAR 分析^[19],基于分形的方法^[20,21],和复杂性度量^[22]等。

2 基于迁移学习的水下声源识别

在机器学习领域,迁移学习^[20]被定义为针对一个有标记的源域 $D_s = \{x_i, y_i\}_{i=1}^n$ 和目标域 $D_t = \{x_j, y_j\}_{j=n+1}^{n+m}$,这两个领域的分布 $P(x_s)$ 和 $P(x_t)$ 不同,即 $P(x_s) \neq P(x_t)$,迁移学习的目的就是要借助 D_s 的知识,来学习目标域 D_t 的知识(标签)。通常有算法2所示的4种方法实现迁移学习。

算法2. 迁移学习的4种基本方法

- 1) 基于样本迁移,根据一定的权重生成规则,对数据样本进行重用。
- 2) 基于特征迁移,通过特征变幻的方法相互迁移,来减少源域和目标域之间的差距;或者将源域和目标域的数据特征变换到统一特征空间中,然后利用传统机器学习方法进行识别。
- 3) 基于关系迁移,关注源域和目标域的样本之间的关系,挖掘不同领域之间的相似性。
- 4) 基于模型迁移,从源域和目标域中找到他们之间共享的参数,以实现知识的迁移。

迁移学习的常见动机是目标任务中缺少足够的训练数据,当使用神经网络时,通过共享预训练网络的权重参数,可以显著减少目标任务模型中需要训练参数的数量,从而以较小的数据集实现有效的学习^[23]。一个流行的例子是在计算机视觉中迁移学习用于图像语义的分割,该网络可以利用在图像分类^[24]任务中学习到的诸如图像基本形状等丰富的信息并进一步用于图像分割任务。

针对图像任务, Visual Geometry Group (VGG)^[25]模型是由 Google 开发和维护的图像分类模型, VGG16 是最好的图像分类模型之一,它经过 1000 种物体的训练,因此可以识别 1000 种物体。

针对音频任务,数据集 AudioSet^[26]是基于人们上传到 YouTube 的大量数据及所提供的标签。其包含 567 个不同类型的音频数据,总时长约 567 小时。Google 已经发布了基于该数据集训练的模型 VGGish^[27],它在网络的较低层处理频谱图,并在较高层时使用多个卷积层和完全连接层获取深层表征用于后续分类。

深度学习技术可以从原始数据中学习特征,在提取信号样本的固有特征方面具有优势,特别是在复杂环境下,这种特征提取上面的优势相比传统特征提取方法更为突出。迁移学习可以将在学习一项任务时获得相应的知识用来解决相关的任务,通常可以将源任务的部分参数用于目标任务。受此启发,基于频谱与图片的二维相似性,我们提出了两类迁移学习方案,一是分别利用基于 ImageNet^[28]及 AudioSet 分别预训练的

模型作为特征提取器,将学习到深度特征进一步用于分类;二是对上述模型进行微调。

图1是基于迁移学习的水下声源识别算法的整体流程,主要分为3个部分:(1)信号预处理,对输入信号进行分段规整及傅里叶变换;(2)进行深度特征提取并选取相应分类器进行分类;(3)对相应预训练网络进行微调。

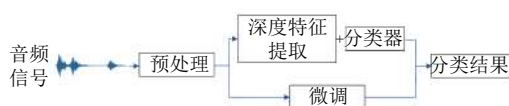


图1 算法整体流程

2.1 预处理

如图2(a)所示,舰船辐射噪声能量主要集中在8000 Hz以下的低频部分^[29],基于这种能量分布规律,为了更好地刻画低频成分的分布规律,我们借鉴了语音识别中的梅尔频谱。其中,目标音频信号采样率为52 734 Hz,使用具有50%重叠的90 ms汉明窗加窗函数,输出的梅尔频谱存储为3×224×224的图片格式,以便后续操作后续用于VGG16网络的输入。此外,VGGish预训练网络的输入采样频率为16 000 Hz,我们对实验音频数据进行了下采样。窗口长度(STFTlength)为25 ms,重叠长度(Hopsize)为10 ms,输出频谱为96×64,嵌入(Embedding)编码大小为128。如图2所示,图2(a)表示音频信号的功率谱,图2(b)表示音频信号的经傅里叶变换后的线性功率谱,图2(c)表示音频信号的梅尔频谱。

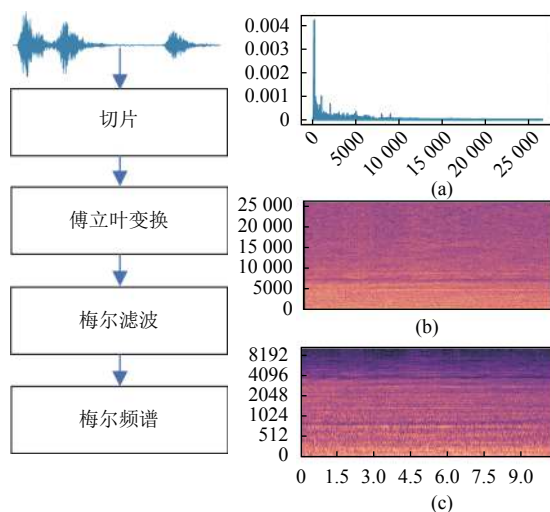


图2 VGG网络预处理及谱示意图

2.2 使用预训练网络提取深度特征用于分类

如图3为使用预训练网络特征提取器的迁移学习框架。本文分别使用基于音频预训练的VGGish,以及基于图像预训练的VGG16网络分别作为特征提取器,分别将网络输出层的输出作为特征向量,其特征向量长度分别为128和4096。

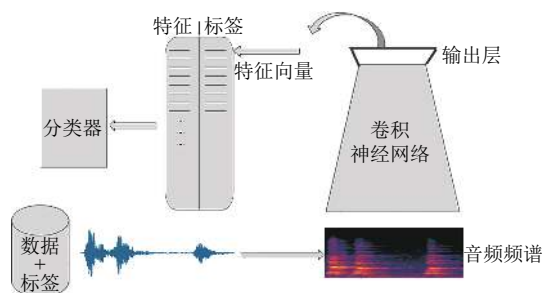


图3 使用预训练网络作为特征提取器

本文使用scikit-learn^[30]的SVM和kNN的算法实现,以及开源ELM的算法实现,分别进行实验并对比VGGish和VGG16抽取的深度特征向量在不同分类器(SVM,kNN和ELM)中的识别性能。

(1) SVM

支持向量机(Support Vector Machine, SVM)^[31]是在分类与回归分析中分析数据的监督式学习模型与相关的学习算法。原始数据是在有限维空间中分布的,但一般区分的集合在该空间中往往线性不可分。为此,SVM将原有限维空间映射到维数更高的空间中,并在高维或无限维空间中构造超平面或超平面集合用以划分不同的集合。为了保持计算负荷合理,人们选择适合该问题的核函数 $k(x,y)$ 来定义SVM方案使用的映射,以确保用原始空间中的变量可以很容易计算点积。在本文中,我们分别采用线性核函数(LinearSVM)和高斯核函数作为支持向量机的核函数,高斯核函数表示如式(1),其中 σ 是带宽。

$$K(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (1)$$

(2) kNN

k最近邻(k-Nearest Neighbor, kNN)^[32]分类算法是一种用于分类和回归的非参数统计方法。在这两种情况下,输入包含特征空间(Feature Space)中的 k 个最接近的训练样本。在kNN分类中,输出是一个分类族群。一个对象的分类由其邻居的“多数表决”确定, k 个

最近邻居 (k 为正整数, 通常较小) 中最常见的分类决定了赋予该对象的类别. 若 $k = 1$, 则该对象的类别直接由最近的一个节点赋予. k NN 是基于实例 (Instance-based) 的学习方法, k NN 算法没有训练的步骤, 因为数据集就是模型本身, 在有新样本输入时, 将测量样本与训练集中每个点的距离, 选择 k 个最接近的点作为 k 最近邻, 新样本的类别即为 k 个点中数量最多的类别, 可以描述为以下公式:

$$C_{\text{test}} = \max(C_1, C_2, C_3, \dots, C_n) \tag{2}$$

其中, C_i 为 K 近邻点中类别为 i 的点的个数, k 和点之间的距离度量方式是可选的参数, 经过实验, 本选取欧式距离作为度量, k 取值为 5.

(3) ELM

极限学习机 (Extreme Learning Machine, ELM)^[33] 实际上也属于前馈神经网络, 极限学习机的出现主要是为了解决神经网络训练速度慢的问题. 传统的前馈神经网络 (如 BP 神经网络) 需要人为设置大量的网络训练参数, 此算法却只需要设定网络的结构, 而不需设置其他参数, 因此具有简单易用的特点. 其输入层到隐藏层的权值是一次随机确定的, 算法执行过程中不需要再调整, 而隐藏层到输出层的权值只需解一个线性方程组来确定, 因此可以提升计算速度.

极限学习机原理如下: 其中 W_1 是输入向量到隐藏节点层的权重矩阵, σ 是激活函数, W_2 是隐藏节点层到输出向量的权重矩阵.

$$\hat{Y} = W_2 \sigma(W_1 x) \tag{3}$$

ELM 对超参数敏感, 我们对其进行了参数搜索, 并使用 ReLU 作为非线性激活函数. 其中, ELM 的主要超参数是隐藏的单元数: 100、250、500、1200、1800、2500. 本文使用开源 ELM 实现, 该代码可从 <https://github.com/zygmuntz/Python-ELM> 获取.

2.3 微调

如图 4 所示, 实验选择了基于 VGG 与训练的 VGG16 网络进行微调, 其中 VGG16 网络的输入为 [3,224,224], VGGISH 的输入为 [96,64]. 本文使用 Python 的 librosa 扩展包计算音频的梅尔频谱图作为网络的输入. 考虑到实验所用数据集音频时长, 我们无法微调整个网络, 因此在实验中只微调了嵌入层 (Embedding layers), 并使用一些正则化技术训练一个简单的多层感知机 (MLP, Multiayer Perceptron), 使用 ReLU 激活函数, Batch

Norm 正则化和 Dropout 的前连接隐藏层, 最终输出为 5 类.

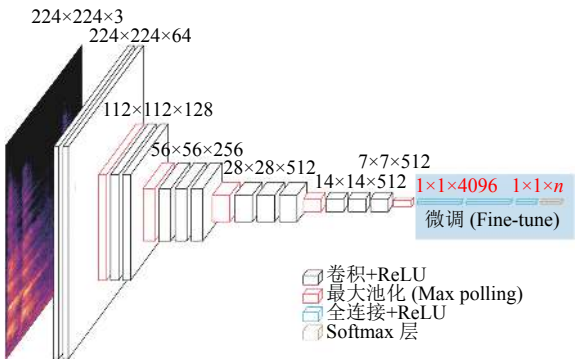


图 4 使用音频频谱对 VGG16 预训练网络进行微调

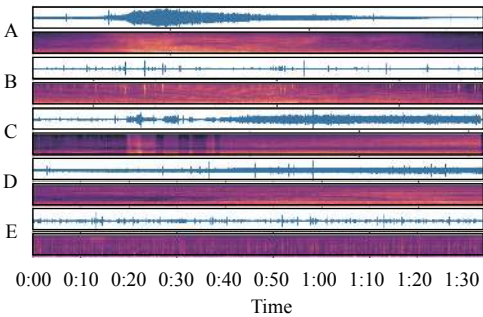


图 5 A~E 五类音频波形及梅尔频谱

3 实验及结果分析

3.1 数据集

本文使用的船只噪声数据来自 2012 年至 2013 年间在西班牙海岸不同地区录制的 ShipsEar^[29] 数据集. 该数据集共有在 15 s 到 10 分钟以内的 11 种舰船类型共计 90 条记录, 按照数据集原文的标注, 根据舰船种类的可以将它们合并为 A、B、C、D 各 4 个大类, 以及 E 类 1 个背景噪声, 如图 5 所示为选取 5 个类的音频波形及其梅尔频谱实例, 其中, 横轴为时间, 纵轴分别为振幅和梅尔频率. 详细划分如表 1 所示.

表 1 ShipsEar 数据集详情

类别	数据集
A	Fishingboats; Trawlers; Musselboats; Tugboats; Drafgers
B	Motorboats; Pilotboats; Sailboats
C	Passengerferries
D	Oceanliner; Ro-Rovessels
E	Backgroundnoiserecordings

3.2 评价指标

我们将模型预测值与真实标签进行比较, 获取测

试中所有样例的真正例 (TP), 假正例 (FP), 真负例 (TN) 和假负例 (FN) 个数. 并针对每一个实验结果分别计算准确率、召回率及 $F1$ 函数, 用以全面而准确地度量来实验结果. 其中, 准确率是所有正确被检索的结果 (TP) 占有所有实际被检索到的 ($TP+FP$) 的比例; 召回率是所有正确被检索的结果 (TP) 占有所有应该检索到的结果 ($TP+FN$) 的比例; $F1$ 是准确率和召回率的调和平均数. 这些指标可以用以下公式表示:

$$accuracy = \frac{TP}{TP+FP+TN+FN} \quad (4)$$

$$recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} \quad (6)$$

3.3 实验结果和分析

本文选取当前在 ShipEar 数据上使用神经网络取得的最好分类结果的论文: 基于深度玻尔兹曼机的水声目标识别^[3] 作为对照文章. 该方法是一种可通过输入数据集学习概率分布的随机生成神经网络.

(1) 基于预训练网络特征提取的识别效果

如表 2 所示, 将 VGGish 及 VGG16 分别直接作为特征提取器, 得到的特征向量分别作为不同分类器的输入. 在使用 VGG16 的特征向量与最近邻分类算法 kNN 组合时, 取得了最好的分类结果, 平均最优准确率为 90.26%, 与当前最好的结果 DBM^[3] 只差了不到 0.5 个百分点. 此外, VGGish 特征向量在与支持向量机 SVM 组合时也取得了准确率为 87.78% 比较理想的分类结果.

表 2 不同训练网络特征向量与不同分类器的性能指标					
预训练网络	性能指标	LinearSVM	SVM	kNN	ELM
VGGISH	Acc	0.7767	0.8778	0.8736	0.8610
	Rec	0.7527	0.8553	0.8791	0.8560
	F1	0.7629	0.8709	0.8719	0.8612
VGG16	Acc	0.8504	0.8672	0.9026	0.8238
	Rec	0.8492	0.8436	0.9044	0.8185
	F1	0.8489	0.8638	0.9046	0.8282

(2) 基于微调预训练网络的识别效果

本文将基于音频数据集预训练的 VGGish 及基于图片数据集预训练的 VGG16 分别去掉原始 Softmax 层之后, 加入不同的全连接层和新的 Softmax 输出层, 如表 3 所示, 在 VGGish 上取得的最优准确率为 89.22%, 在 VGG16 上最优准确率为 92.48%. 相对当前最优的

DBM 网络的 90.70%, 本文提出的算法有比较明显的性能提升.

表 3 微调 VGGISH、VGG16 及 DBM 网络性能指标		
网络	隐层网络结构	平均准确率
VGGISH	-64-5	0.8627
	-128-5	0.8922
	-128-32-5	0.8871
VGG16	-512-5	0.9177
	-1024-5	0.9248
	-1024-512-5	0.9106
DBM	-500-5	0.8860
	-500-500-5	0.9030
	-500-500-200-5	0.9070

如图 6 所示, 对于添加一个神经元个数为 1024 隐层的 VGG16 网络, 当使用对数损失函数时, 损失函数随训练次数的变化, 在大约 200 次迭代后即收敛, 相比 DBM^[3] 的收敛速度大约提升了 5 倍.

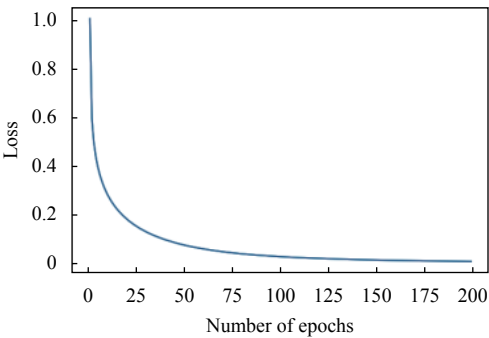


图 6 VGG16 在训练集上的错误率

如图 7 所示, 横轴为真实值, 纵轴为预测值, 可以看出, 针对类别 E, 即环境噪声的平均预测准确率为 98% 左右, 且不存在漏判, 即没有识别出将舰船噪声并当做环境音; 算法对于针对 A、B、C、D 四类舰船噪声的识别准确率分别为 90%、83%、92%、93%. 其中, 针对 B 类相对较低的准确率, 仅为 83%, 通过对 ShipsEar 数据集的分析, 我们认为是 B 类数据的量相对于 A、C、D 类原始数据的相对不均衡, 这个问题可以通过进一步扩充数据集得到避免. 总的来说, 我们提出的基于迁移学习的水声目标识别算法, 多次试验平均最优准确率为 92.48%, 取得了当前最好的分类效果, 在算法训练和数据处理上也有比较明显的优势.

4 结论与展望

在水声目标识别受限于海洋环境的复杂性, 而难

以提取有效的目标特征并进行分类的背景下,本文首次采用基于迁移学习的水声目标识别方法,分别采用预训练网络特征提取和微调两种迁移学习方法,并应用 ShipsEar 数据集,进行了算法验证.结果表明,与当前最好的 DBM 分类算法相比,我们取得了 92.48% 的最好平均准确率.总的来说,通过实验我们验证了迁移学习应用于水声目标分类中的可行性,并且迁移学习可以解决在音频领域的数据稀疏问题.此外,这种思路也许可以被用于很多不同的音频任务,具有广阔的应用前景.

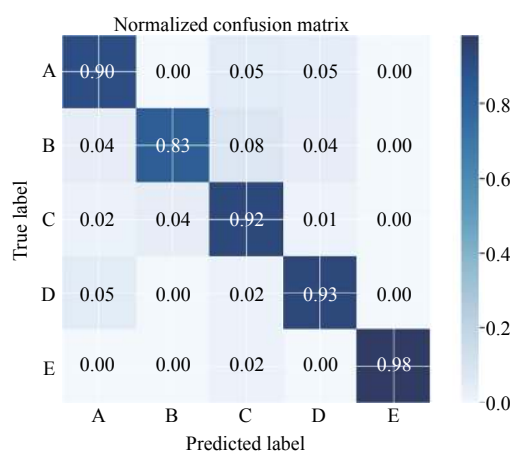


图7 5个类的分类混淆矩阵

参考文献

- 1 方世良, 杜栓平, 罗昕炜, 等. 水声目标特征分析与识别技术. 中国科学院院刊, 2019, 34(3): 297–305. [doi: 10.16418/j.issn.1000-3045.2019.03.007]
- 2 Chu S, Narayanan S, Kuo CCJ. Environmental sound recognition with time-frequency audio features. IEEE Transactions on Audio, Speech, and Language Processing, 2009, 17(6): 1142–1158. [doi: 10.1109/TASL.2009.2017438]
- 3 Xie JW, Chen J, Zhang J. DBM-based underwater acoustic source recognition. Proceedings of 2018 IEEE International Conference on Communication Systems (ICCS). Chengdu, China. 2018. 366–371.
- 4 Liu J, He Y, Liu Z, *et al.* Underwater target recognition based on line spectrum and support vector machine. Proceedings of 2014 International Conference on Mechatronics, Control and Electronic Engineering (MCE-14). Paris, France. 2014.
- 5 Chung KW, Sutin A, Sedunov A, *et al.* DEMON acoustic ship signature measurements in an urban harbor. Advances in Acoustics and Vibration, 2011, 2011: 952798. [doi: 10.1155/2011/952798]
- 6 Sejdić E, Djurović I, Jiang J. Time-frequency feature representation using energy concentration: An overview of recent advances. Digital Signal Processing, 2009, 19(1): 153–183. [doi: 10.1016/j.dsp.2007.12.004]
- 7 Wang Q. Underwater bottom still mine classification using robust time-frequency feature and relevance vector machine. International Journal of Computer Mathematics, 2009, 86(5): 794–806. [doi: 10.1080/00207160701704572]
- 8 Cai YB, Zhang MZ, Shi XZ, *et al.* The feature extraction and classification of ocean acoustic signals based on wave structure. Acta Electronica Sinica, 1999, 27(6): 129–130.
- 9 Li N, Li XK. A stochastic resonance of line-spectrum extraction method for underwater target recognition. Journal of Information & Computational Science, 2014, 11(18): 6437–6446.
- 10 Zheng SY, Guo HX, Li YA, *et al.* A new method for detecting line spectrum of ship-radiated noise using duffing oscillator. Chinese Science Bulletin, 2007, 52(14): 1906–1912. [doi: 10.1007/s11434-007-0271-4]
- 11 Fu ZK, Xing JC, Zhu RD, *et al.* A new method of detecting line-spectrum based on auto-correlation with stochastic resonance theory. Proceedings of the 2013 25th Chinese Control and Decision Conference (CCDC). Guiyang, China. 2013. 1104–1107.
- 12 Ma YC, Chin CS, Woo WL, *et al.* An acoustic annoyance study of hard disk drive for laptop. IEEE Transactions on Magnetics, 2016, 52(8): 3301609.
- 13 Azimi-Sadjadi MR, Yao D, Huang Q, *et al.* Underwater target classification using wavelet packets and neural networks. IEEE Transactions on Neural Networks, 2000, 11(3): 784–794. [doi: 10.1109/72.846748]
- 14 Averbuch A, Zheludev V, Neittaanmäki P, *et al.* Acoustic detection and classification of river boats. Applied Acoustics, 2011, 72(1): 22–34. [doi: 10.1016/j.apacoust.2010.09.006]
- 15 Huang NE, Shen SSP. Hilbert-Huang Transform and Its Applications: Volume 16. 2nd ed. London: World Scientific, 2014.
- 16 Lin L, Chu FL. Feature extraction of AE characteristics in offshore structure model using Hilbert-Huang transform. Measurement, 2011, 44(1): 46–54. [doi: 10.1016/j.measurement.2010.09.002]
- 17 Kennel MB, Brown R, Abarbanel HDI. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Physical Review A, 1992, 45(6): 3853–3864.

- 3403–3411. [doi: [10.1103/PhysRevA.45.3403](https://doi.org/10.1103/PhysRevA.45.3403)]
- 18 Meng QX, Yang S, Piao SC, *et al.* Research on chaotic character of ship-radiated noise based on phase space reconstruction. Proceedings of 2012 International Conference on Image Analysis and Signal Processing. Hangzhou, China. 2012. 1–5.
- 19 de Moura NN, Filho ES, de Seixas JM. Independent component analysis for passive sonar signal processing. In: Silva SR, eds. Advances in Sonar Technology. Austria: In-Teh, 2009.
- 20 Pan SJ, Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345–1359. [doi: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191)]
- 21 Yang S, Li ZS, Wang XL. Ship recognition via its radiated sound: The fractal based approaches. The Journal of the Acoustical Society of America, 2002, 112(1): 172–177. [doi: [10.1121/1.1487840](https://doi.org/10.1121/1.1487840)]
- 22 Wang Y, Liu QY. Multi-scale permutation entropy as a tool for complexity analysis of ship-radiated noise. Proceedings of 2016 IEEE/OES China Ocean Acoustics (COA). Harbin, China. 2016. 1–4.
- 23 Choi K, Fazekas G, Sandler M, *et al.* Transfer learning for music classification and regression tasks. Proceedings of 18th International Society for Music Information Retrieval Conference. 2017. 141–149.
- 24 Oquab M, Bottou L, Laptev I, *et al.* Learning and transferring mid-level image representations using convolutional neural networks. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 1717–1724.
- 25 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Bengio Y, LeCun Y, eds. 3rd International Conference on Learning Representations. San Diego, CA, USA. 2015.
- 26 Gemmeke JF, Ellis DPW, Freedman D, *et al.* Audio set: An ontology and human-labeled dataset for audio events. Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA. 2017. 776–780.
- 27 Hershey S, Chaudhuri S, Ellis DPW, *et al.* CNN architectures for large-scale audio classification. Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA. 2017. 131–135.
- 28 Deng J, Dong W, Socher R, *et al.* Imagenet: A large-scale hierarchical image database. Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA. 2009. 248–255.
- 29 Santos-Domínguez D, Torres-Guijarro S, Cardenal-López A, *et al.* Shippsear: An underwater vessel noise database. Applied Acoustics, 2016, 113: 64–69. [doi: [10.1016/j.apacoust.2016.06.008](https://doi.org/10.1016/j.apacoust.2016.06.008)]
- 30 Pedregosa F, Varoquaux G, Gramfort A, *et al.* Scikit-learn: Machine learning in python. The Journal of Machine Learning Research, 2011, 12: 2825–2830.
- 31 Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Processing Letters, 1999, 9(3): 293–300. [doi: [10.1023/A:1018628609742](https://doi.org/10.1023/A:1018628609742)]
- 32 Keller JM, Gray MR, Givens JA. A fuzzy K-nearest neighbor algorithm. IEEE Transactions on Systems, Man, and Cybernetics, 1985, SMC-15(4): 580–585.
- 33 Huang GB, Zhu QY, Siew CK. Extreme learning machine: Theory and applications. Neurocomputing, 2006, 70(1–3): 489–501.