

# 基于改进视觉词袋模型的水声目标识别<sup>①</sup>



潘安迪, 肖川, 陈曦

(复旦大学 计算机科学技术学院, 上海 201203)

通讯作者: 潘安迪, E-mail: [panandi@qq.com](mailto:panandi@qq.com)

**摘要:** 水声目标识别的任务是通过采集到水声目标的信号来对目标进行分类, 在海洋勘探, 监听技术等领域有着非常重要和广泛的应用. 由于海洋环境的复杂性, 以及船只目标发动机的多样性以及噪声的存在, 水声目标识别是一个困难的任务. 传统的特征提取方法无法提取到足够有效的特征表示, 充分地表示目标. 为了解决这个问题, 本文提出了一种基于改进的视觉化词袋模型的水声识别算法, 通过使用视觉化词袋模型对频谱图进行高维的特征提取, 然后使用了自然语言处理领域中常见的词频-逆文件频率 (TF-IDF) 算法来对得到的特征向量进行权重调整, 然后输入到多层感知机中, 对水声目标进行分类识别. 实验结果表明, 本文提出的识别算法取得了 92.53% 的正确率, 相比于当前效果最好的深度玻尔兹曼机 (DBM) 算法有了明显的提升.

**关键词:** 支持向量机; 水声识别; 视觉词袋; TF-IDF; 音频分类

引用格式: 潘安迪, 肖川, 陈曦. 基于改进视觉词袋模型的水声目标识别. 计算机系统应用, 2021, 30(5): 170-175. <http://www.c-s-a.org.cn/1003-3254/7901.html>

## Underwater Acoustic Target Recognition Based on Improved Bag of Visual Words

PAN An-Di, XIAO Chuan, CHEN Xi

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

**Abstract:** Underwater acoustic target recognition is to classify the targets through collecting the signals of the underwater acoustic targets and has very important and extensive applications in the fields of ocean exploration and monitoring technology. Due to the complexity of the marine environment, the diversity of target ship engines, and the background noise, underwater acoustic target recognition is difficult. Traditional feature extraction methods cannot extract effective feature representations to fully represent the targets. In order to solve this problem, we propose an underwater acoustic target recognition algorithm based on the improved bag of visual words. Specifically, this algorithm adopts the bag of visual words to extract high-dimensional features in a spectrogram and then adjusts the weights of the obtained feature vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm commonly used in the field of natural language processing. Furthermore, the vectors are input to a Multi Layer Perceptron (MLP) to classify and recognize the underwater acoustic targets. The experimental results show that the recognition algorithm proposed in this study achieves an accuracy of 92.53%, which is a significant improvement in comparison with the best Deep Boltzmann Machine (DBM) algorithm at present.

**Key words:** Support Vector Machine (SVM); underwater acoustic target recognition; bag of visual words; Term Frequency-Inverse Document Frequency (TF-IDF); audio classification

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: 2020-09-03; 修改时间: 2020-09-25; 采用时间: 2020-10-13; csa 在线出版时间: 2021-04-28

## 1 介绍

水声目标识别的任务是通过采集到水声目标的信号来对目标进行识别和检测<sup>[1]</sup>。海水是一种复杂的介质,具有时变特性,充斥复杂噪声的特性<sup>[2-4]</sup>。因此捕捉到的声学信号在时间和相位上拥有时变的特点<sup>[2]</sup>。这也使得水声目标识别是一项具有挑战性的工作。

在水声目标识别领域早期,大部分分类任务都是基于特征工程与分类器构造的<sup>[5]</sup>。常被水声目标识别使用的特征有梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient, MFCC),时频谱特征等等。从特征工程的角度来讲,有人设计手工特征,如Li等<sup>[2]</sup>通过比较小波包分解后不同频带的分形维数(Fractal Dimension, FD)值来设计船声信号特征;有人使用特征组合和特征选择方法来组合、抽取以构成更加鲁棒的特征,如Jiang等<sup>[6]</sup>将功率谱(power spectra)和方谱特征(square spectrum feature)结合起来,并且使用主成分分析法降维,获得更加有判别力的特征。分类器构造方面,有人引入集成算法,如Sun等<sup>[7]</sup>使用自适应提升算法(AdaBoost)构造并集成多个有偏好的分类器,以提高最终分类准确率;有人致力于核函数的构造,如Sherin等<sup>[8]</sup>使用BAT算法进行核函数的构造与选择。整体而言,在水声目标识别领域中,抽取的特征能否有效地表示样本决定了模型的性能。

近年来,由于深度学习可以从数据中自动学习特征,并且学得特征比传统特征更能表示样本,成为了近年来的研究热点,如Xie等<sup>[9]</sup>基于时频谱,使用了深度玻尔兹曼机。然而,深度学习方法在拥有大量数据时才更好地发挥优越性,显然,常见的水声目标识别数据集无法满足这一要求。因此,半监督学习方法被引入。Yang等<sup>[10]</sup>使用无标签数据预训练了深度信念网络,利用竞争性学习方法以及标签数据对神经元进行聚类,最后进行微调。邓晋等<sup>[11]</sup>使用了两类迁移学习方法,基于ImageNet<sup>[12]</sup>和AudioSet训练分类模型作为特征提取器,对于抽取的深度特征进行进一步分类;另一种方法是,将上述模型作为预训练模型,使用水声目标识别数据集进行微调得到最终的分类模型。

在现实的海洋环境声学场景中,往往船只的声音信号伴随着很大的背景噪声<sup>[13,14]</sup>,传统的低层次特征<sup>[15-18]</sup>对噪声会比较敏感,因为低层次特征容易直接被声音信号中的物理性质影响。在图像分类任务<sup>[19-21]</sup>中,视觉词袋模型是一个有效的模型,常常被用于计算机

视觉中的目标识别任务。与另外一种高层次特征SIFT<sup>[22]</sup>相比,视觉化词袋模型更适合水声目标识别,因为水声音频信号频谱沿着时间轴的变化是显著的,但是在尺度、位置以及形态上面没有显著的特征。因此,本文提出了一种基于改进视觉化词袋模型的方法来进行水声目标识别,对高层次特征进行提取,通过建立视觉词本,统计一个音频频谱图中含有的每个视觉化词条的数量,将这个向量转化为TF-IDF向量,得到最终的特征,输入到多层感知机中,进行分类。

本文后续的组织结构如下:第2节介绍本文提出的算法流程以及算法的原理;第3节介绍本文进行的实验,包括实验配置以及实验结果和结果分析;第4节对文章进行总结。

## 2 算法设计

本文提出的算法的整体架构如图1所示。首先,对音频信号进行预处理,将音频信号转换为梅尔频谱,然后对梅尔频谱图做归一化操作,将梅尔频谱图转换为灰度图谱。然后需要建立视觉词汇库,在每个类别中随机抽取10个样本,对样本进行上述的预处理操作,转换为灰度图,然后将灰度图中的一列作为一个视觉词汇,将所有视觉词汇打乱,使用K-means聚类<sup>[23]</sup>将所有视觉词汇聚类为600个簇心。簇心的数目即视觉词汇库的规模。最后,训练集中的音频数据都需要进行预处理,然后将灰度图中的每一列归类到视觉词汇库中的某个词汇,从而得到一个600维的向量,每个维度表示的是该样本对应的灰度图中含有某个词汇的数量。

### 2.1 预处理

由于船只噪声信号能量主要分布在8000 Hz以下的低频带,为了更好地表现能量沿着频率轴的变化,我们选择梅尔频谱来表现信号的时频特性。目标音频信号采用的采样率为52 734 Hz,使用50%重叠的90 ms汉明窗加窗函数,然后通过归一化操作将梅尔频谱图转换为灰度图,公式描述如下,得到224×224的图片格式。

$$GI(f,t) = \frac{S(f,t) - \min(S)}{\max(S) - \min(S)} \quad (1)$$

### 2.2 建立视觉词汇库

词袋模型(Bag of Words, BOW)<sup>[24]</sup>是自然语言处理(natural language processing)领域中常用的算法,用来表示文本信息。词袋模型假设一个文档,忽视文档的单

词顺序与语法, 仅将文本看作词语的集合, 用词语的统计信息来表示一个文档. 在本论文中我们将灰度图中的一列作为一个词汇, 从而得到视觉词袋模型 (Bags of Visual Words, BOVW), 可以将频谱图表示为一个统计向量<sup>[25]</sup>. 首先, 我们需要建立视觉词汇库. 从每个类别

中随机选择 10 个样本, 转化为灰度频谱图, 然后把灰度频谱图的每一列作为一个词汇, 使用 K-means 聚类算法, 使用余弦距离作为度量标准, 将 50 个样本中所有的词汇进行聚类, 聚类为  $k$  个类别,  $k$  为视觉词汇库的规模, 即为其中含有的词汇数量.

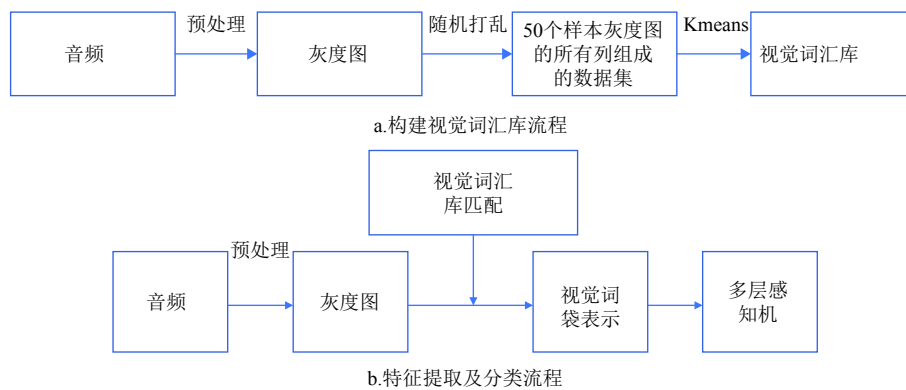


图 1 算法整体流程图

算法 1. 建立视觉词汇库

输入: 训练集音频  
输出:  $K$  个标准视觉词汇

```

 $(S_1, S_2, \dots, S_{50}) \leftarrow (A, B, C, D, E)$ 
#在训练集 A,B,C,D,E 五个类别分别随机抽取 10 个样本
 $((V_{1,1}, V_{1,2}, \dots, V_{1,224}), \dots, (V_{50,1}, V_{50,2}, \dots, V_{50,224})) \leftarrow \text{Preprocess}(S_1, S_2, \dots, S_{50})$ 
#对样本进行上述的预处理步骤, 然后将灰度图中的每一列作为一个视觉词汇
 $\text{Data}(V) \leftarrow \text{Shuffle}(((V_{1,1}, V_{1,2}, \dots, V_{1,224}), \dots, (V_{50,1}, V_{50,2}, \dots, V_{50,224})))$ 
#将 50 个样本中的所有视觉词汇随机打乱, 作为聚类的数据集
 $W_1, W_2, \dots, W_K \leftarrow \text{Kmeans}(\text{Data}(V))$ 
#使用 K-means 聚类算法将数据集聚类为  $K$  个簇,  $K$  个簇心即为  $K$  个视觉词汇,  $K$  为视觉词汇库的规模
    
```

2.3 视觉词袋表示

建立视觉词汇库之后, 我们将提取音频的视觉词袋表示特征. 首先我们将音频转化为灰度图, 然后将灰度图中的每一个列作为视觉词汇, 将视觉词汇与视觉词汇库中的  $K$  个标准视觉词汇进行对比, 找到欧式距离最近的标准视觉词汇  $W_i$ , 这一列视觉词汇即属于视觉词汇  $W_i$ . 在得到了灰度图中的每一列所属的视觉词汇之后, 通过统计每一个视觉词汇出现的次数, 得到一个  $K$  维向量, 这个向量即为一个音频的视觉词袋表示. 在视觉词袋模型上, 本文还使用 TF-IDF 方法做出了额外的改进, 用来凸显重要性高的视觉词汇, 对重要性较

低的视觉词汇则使用较低的权重. 视觉词袋表示的特征提取流程如算法 2.

算法 2. 视觉词袋特征提取

输入: 样本音频  
输出: 视觉词袋特征向量

```

 $(V_1, V_2, \dots, V_{224}) \leftarrow G \leftarrow \text{Preprocess}(S)$ 
#对音频进行预处理, 得到灰度图, 将灰度图中的每一列当成一个视觉词汇
 $(N_1, N_2, \dots, N_K) \leftarrow (W_i \dots) \leftarrow \text{match}(((V_1, V_2, \dots, V_{224})))$ 
#在视觉词汇库中找到欧式距离最近的标准词汇  $W_i$ , 将这一列看作词汇  $W_i$ , 统计灰度图中包含的各个视觉词汇的数量, 得到一个  $K$  维向量, 向量的第  $i$  维是视觉词汇  $W_i$  的数量
 $(F_1, F_2, \dots, F_K) \leftarrow \text{Mutiply}((N_1, N_2, \dots, N_K), (TDF, TDF, \dots, TDF))$ 
#计算逆文件频率 (IDF) 向量, 将两个向量对应位置的值相乘, 得到  $K$  维的 BOVW-TF-IDF 向量
    
```

TF-IDF (Term Frequency-Inverse Document Frequency, 词频-逆向文件频率)<sup>[26]</sup> 是一种用于信息检索与文本挖掘的常用加权技术. 其实 TF 是词频 (Term Frequency), 表示了词语在文本中出现的频率, IDF (Inverse Document Frequency) 是逆文档频率, 表示了词语在语料库中出现的频率. TF-IDF 是一种统计方法, 用以评估一个词汇对于一个语料库中的其中一份文档的重要程度. 词汇的重要性随着它文档中出现的次数成正比增加, 但同时会随着它在语料库中出现的频率成反比下降, 例如

常用的“的”字. 在视觉词汇中, 词汇同样有重要程度的区别, 比较重要的视觉词汇更能显著地描述样本, 重要程度低的视觉词汇可能与音频信号中共有的场景以及噪声有关. 所以本论文中将视觉词袋模型方法提取的向量转化为 TF-IDF 向量, 表示为如公式:

$$IDF_i = \log\left(\frac{N_{\text{all}}}{N_i + 1}\right) \quad (2)$$

$$TF_{ij} = \frac{n_{ij}}{l_i} \quad (3)$$

$$TF-IDF = TF \times IDF \quad (4)$$

$IDF_i$  可以衡量视觉词汇  $W_i$  在视觉词汇库中的重要程度, 在音频文件转换得到的灰度图中出现的频率越高, 重要程度越低, 反之, 则重要程度越高.

## 2.4 分类器

感知机 (perceptron)<sup>[27]</sup> 是最原始的前馈神经网络. 它通过权重  $w$  和偏置项  $b$ , 可以将一个多维输入  $X$  映射为一个二元的值, 是一种简单的二元线性分类器. 多层感知机 (multilayer perceptron) 是感知机的推广. 感知机由输入层部分、隐藏层部分和输出层部分组成, 隐藏层部分可以是一层或者多层的. 每一层由多个感知机单元组成, 输入为上一层的所有输出. 同时, 若多层感知器的每一个单元都是线性的, 那么任意多层感知机都可被等价为一个单层感知机. 所以, 多层感知机中, 每一个单元都选择了非线性的激活函数. 最终, 使用衡量预测值与实际值之间误差的损失函数, 通过反向传播算法, 一个多层感知机可以在训练集上面拟合.

在本文中, 我们将设置感知机的参数为: 输入层的神经元数目为 600, 即为输入特征的维数, 也是聚类算法的簇心数目, 视觉词汇库的规模; 中间隐藏层的神经元数目为 300; 输出层神经元数目与类别数目一致, 为 5 个神经元. 在输入层和隐藏层之后都使用批归一化 (batch normalization) 来调整数据的分布. 本文使用的感知机原理图简化如图 2.

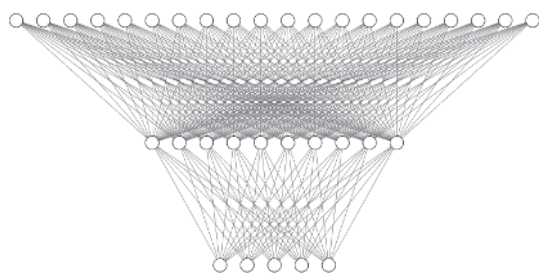


图2 多层感知机原理图

本文还测试了将支持向量机<sup>[28]</sup> 和随机森林<sup>[29]</sup> 作为后端分类器的性能, 结果显示多层感知机要优于其他两个分类器, 实验结果在第 3.4 节中给出.

## 3 实验及结果分析

### 3.1 实验平台

本文实验在 Ubuntu20.04 系统下, 利用 Python 语言, Scikit-Learn<sup>[30]</sup> 机器学习框架完成. 硬件环境为 CPU Intel I7, 内存 16 GB, 显卡 Nvidia GeForce RTX1080Ti.

### 3.2 数据集

本文中使用的船舶噪声数据集来自于挪威科技大学在 2012-2013 年录制的 ShipsEar 数据集. ShipsEar 数据集包含不同种类的船只录音以及海洋环境自然背景噪声, 旨在用于训练船只识别检测系统. 该数据集总共包含 90 条记录, 每条记录的长度从 15 秒到 10 分钟不等, 船只总共有 11 类, 根据原论文的分类, 将 11 种船只噪声和海洋环境自然背景噪声合并为 A, B, C, D, E 五类, 其中 E 类为自然背景噪声, 详细的分类如表 1 所示.

表1 ShipsEar 数据集详情

类别	内容
A	Fishing boats; Trawlers; Mussel boats; Tugboats; Drafgers
B	Motorboats; Pilot boats; Sailboats
C	Passenger ferries
D	Ocean liner; Ro-Ro vessels
E	Background noise recordings

### 3.3 评价指标

在评估模型性能过程中, 经常需要使用多种不同指标进行评估. 大部分的评价指标只能片面的反映模型某一部分的性能. 常用的分类任务评价指标有精确率 (Precision), 召回率 (Recall), 准确率 (Accuracy), F1. 本文的数据类别比较均衡, 用 Accuracy 即可较好地评估模型的好坏. 将模型的预测结果和真实的标签进行对比, 真正例 (True Positive, TP) 为被模型预测为正的正样本, 假正例 (False Positive, FP) 为被模型预测为正的负样本, 假负例 (False Negative, FN) 为被模型预测为负的正样本, 真负例 (True Negative, TN) 为被模型预测为负的负样本.

$$Accuracy = \frac{TP}{TP + FP + TN + FN} \quad (5)$$

### 3.4 结果分析

本论文选择现在 ShipEar 船舶噪声数据集上使用

神经网络达到最好性能的论文: 基于深度玻尔兹曼机的水声目标识别算法 (dbm-based underwater acoustic source recognition) 作为对比模型. 该论文提出了一种通过数据集用深度学习学习特征的方法进行水声目标识别.

为了体现本文提出的使用词频-逆文件频率改进的视觉词袋特征 (BOVW-TF-IDF) 的作用, 本实验选择了对比使用词频-逆文件频率改进的视觉词袋特征 (BOVW-TF-IDF) 与视觉词袋特征 (BOVW) 的效果, 分类器都是多层感知机. 如表 2, 可以看出, BOVW-TF-IDF 特征相比传统 BOVW 特征能更好地表示水声目标识别信号, 可能是因为 TF-IDF 算法可以将每个类别中都会出现的噪声以及海洋背景声的权重降低了, 使得能表示船只特性的词汇更为突出.

表 2 与 DBM 算法性能对比

算法	Accuracy
DBM	0.9070
BOVW	0.9205
<b>BOVW-TF-IDF</b>	<b>0.9253</b>

在第一个实验中, 我们对现在基于深度玻尔兹曼机的水声识别算法性能和本文提出的算法的正确率进行对比. 如表 2 所示, 本文提出的算法相对基于深度玻尔兹曼机的水声识别算法有比较明显的提升, 证明了本文提出的特征表示方法是很有价值的, 可以达到深度学习模型提取特征的效果.

本文选择了 3 种分类器进行对比, 结果见表 3, 输入的都是 TF-IDF 特征, 根据 3 种分类器的正确率, 最终选择了多层感知机作为末端分类器. 可以看出多层感知机的分类准确率要高于其他两个分类器, 所以本文选择多层感知机作为末端分类器.

表 3 不同分类器的分类正确率对比

末端分类器	Accuracy
<b>MLP</b>	<b>0.9253</b>
SVM	0.9164
Random Forest	0.9198

分类器的正确率也会受词汇库的规模影响, 如图 3 所示, 在词汇库规模超过 600 之后, 识别的正确率有降低的趋势. 可能是因为声音的频谱图比现实中的更抽象, 600 的词汇库规模已经可以有效地表示水声目标, 达到比较高的正确率

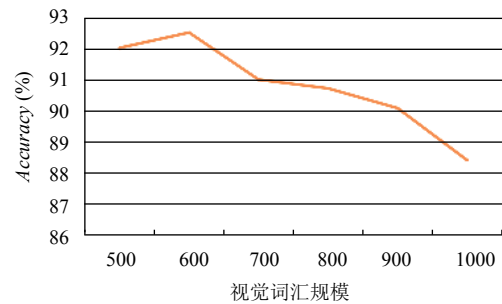


图 3 不同词汇规模对模型性能的影响对比

#### 4 结论与展望

本文针对水声目标识别问题, 提出了一种基于改进视觉化词袋模型的算法, 该算法通过视觉化词袋方法提取出了频谱图中的统计信息作为高阶表示, 并且使用了自然语言处理中常用的词频-逆文件频率方法来调整特征的权重, 能够更有效地利用重要的视觉特征, 降低环境声以及噪声的影响, 并在 ShipEar 数据集上取得了 92.53% 的平均准确率, 相较于现在效果最好的 DBM 算法有了显著的提升, 满足了现实应用场景的准确率需求. 本文提出的算法的思路可以广泛运用于音频分类的领域.

#### 参考文献

- 杨宏晖, 戴健, 孙进才, 等. 用于水声目标识别的自适应免疫特征选择算法. 西安交通大学学报, 2011, 45(12): 28-32, 110.
- Li HT, Cheng YS, Dai WG, *et al.* A method based on wavelet packets-fractal and SVM for underwater acoustic signals recognition. Proceedings of the 2014 12th International Conference on Signal Processing (ICSP). Hangzhou, China. 2014. 2169-2173.
- 戴健, 杨宏晖, 王芸, 等. 用于水声目标识别的自适应遗传样本选择算法. 声学技术, 2013, 32(4): 332-335.
- Jiang JJ, Shi T, Huang M, *et al.* Multi-scale spectral feature extraction for underwater acoustic target recognition. Measurement, 2020, 166: 108227. [doi: 10.1016/j.measurement.2020.108227]
- Wei ZX, Ju Y, Song M. A method of underwater acoustic signal classification based on deep neural network. Proceedings of the 2018 5th International Conference on Information Science and Control Engineering (ICISCE). Zhengzhou, China. 2018. 46-50.
- Jiang WH, Tong F, Dong YZ, *et al.* Modulation recognition of non-cooperation underwater acoustic communication signals using principal component analysis. Applied

- Acoustics, 2018, 138: 209–215. [doi: [10.1016/j.apacoust.2018.03.033](https://doi.org/10.1016/j.apacoust.2018.03.033)]
- 7 Sun SL. Local within-class accuracies for weighting individual outputs in multiple classifier systems. *Pattern Recognition Letters*, 2010, 31(2): 119–124. [doi: [10.1016/j.patrec.2009.09.017](https://doi.org/10.1016/j.patrec.2009.09.017)]
- 8 Sherin BM, Supriya MH. Selection and parameter optimization of SVM kernel function for underwater target classification. *Proceedings of 2015 IEEE Underwater Technology (UT)*. Chennai, India. 2015. 1–5.
- 9 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.
- 10 Yang HH, Shen S, Yao X H, *et al.* Competitive deep-belief networks for underwater acoustic target recognition. *Sensors*, 2018, 18(4): 952. [doi: [10.3390/s18040952](https://doi.org/10.3390/s18040952)]
- 11 邓晋, 潘安迪, 肖川, 等. 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. [doi: [10.15888/j.cnki.csa.007538](https://doi.org/10.15888/j.cnki.csa.007538)]
- 12 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.
- 13 Robinson SP, Lepper PA, Ablitt J. The measurement of the underwater radiated noise from marine piling including characterisation of a “soft start” period. *Proceedings of OCEANS 2007 - Europe*. Aberdeen, UK. 2007. 1–6.
- 14 Merchant ND, Pirota E, Barton TR, *et al.* Monitoring ship noise to assess the impact of coastal developments on marine mammals. *Marine Pollution Bulletin*, 2014, 78(1–2): 85–95.
- 15 Wang XY, Han TX, Yan SC. An HOG-LBP human detector with partial occlusion handling. *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision*. Kyoto, Japan. 2009. 32–39.
- 16 Tan XY, Triggs B. Fusing Gabor and LBP feature sets for kernel-based face recognition. *Proceedings of the 3rd International Conference on Analysis and Modeling of Faces and Gestures International Conference on Analysis & Modeling of Faces & Gestures*. Rio de Janeiro, Brazil. Springer-Verlag, 2007. 235–249.
- 17 Yang ZQ, Li QH, Li Y, *et al.* A LBP texture analysis based liveness detection for face authentication. *Proceedings of the 2nd International Conference on Machine Learning for Cyber Security*. Xi'an, China. 2019. 17–28.
- 18 Gao X, Qu CM. Pedestrian detection algorithm based on HOG and residual network. *Journal of Hleilongjiang University of Technology (Comprehensive Edition)*, 2019, 19(4): 72–77.
- 19 Gu XW, Angelov PP. Semi-supervised deep rule-based approach for image classification. *Applied Soft Computing*, 2018, 68: 53–68.
- 20 Wan WT, Zhong YY, Li TP, *et al.* Rethinking feature distribution for loss functions in image classification. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. 2018. 9117–9126.
- 21 Zhang C, Pan X, Li HP, *et al.* A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and & Remote Sensing*, 2018, 140: 133–144.
- 22 Nam SH, Kim WH, Mun SM, *et al.* A SIFT features based blind watermarking for DIBR 3D images. *Multimedia Tools and & Applications*, 2018, 77(7): 7811–7850.
- 23 Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 2010, 31(8): 651–666. [doi: [10.1016/j.patrec.2009.09.011](https://doi.org/10.1016/j.patrec.2009.09.011)]
- 24 Zhang Y, Jin R, Zhou ZH. Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and & Cybernetics*, 2010, 1(1–4): 43–52.
- 25 Yang J, Jiang YC, Hauptmann AG, *et al.* Evaluating Evaluating bag-of-visual -words representations in scene classification. *Proceedings of International Workshop on Multimedia Information Retrieval*. Augsburg, Germany. 2007. 197–206.
- 26 Dessì D, Helaoui R, Kumar V, *et al.* TF-IDF vs word embeddings for morbidity identification in clinical notes: An initial study. *Proceedings of the 1st Workshop on Smart Personal Health Interfaces co-located with 25th International Conference on Intelligent User Interfaces*. Cagliari, Italy. 2020. 1–12.
- 27 Gardner MW, Dorling SR. Artificial neural networks (the multilayer perceptron) —A review of applications in the atmospheric sciences. *Atmospheric Environment*, 1998, 32(14–15): 2627–2636.
- 28 Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. *Neural Processing Letters*, 1999, 9(3): 293–300.
- 29 Liaw, A, Watthew M. Classification and regression by random forest. *R News*, 2002, 3: 18–22.
- 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.