

# 基于多层优选卷积的水声信号样本自动标注方法

王红滨<sup>1</sup>, 张帅<sup>1</sup>, 何鸣<sup>1</sup>, 陈夏可<sup>2</sup>

(1. 哈尔滨工程大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001; 2. 黑龙江科技大学 计算机科学与信息工程学院, 黑龙江 哈尔滨 150022)

**摘要:**针对深度学习在水声研究领域的应用中面临大数据量要求和现有样本量限制的问题,本文提出了一种多层优选卷积网络模型。通过基于相似度的优选方法选出最佳卷积核,以提取更具代表性的特征。利用探索层特征融合策略,叠加多层卷积输出,获取更全面的特征信息。采用约减策略优化模型,有效缩短运算时间。通过优选、特征融合和注意力机制,有效解决此类问题。实验结果表明,该模型在数据集上取得的最好的标注准确率为高基线模型 1.12%;同时运行时间减少了 43.5%。因此,该模型的使用提高了水声信号标注的准确率和效率。

**关键词:**水声信号; 自动标注; 声纹识别; 多层优选卷积模型; 时间优化; 注意力机制; 特征融合

DOI: 10.11990/jheu.202206048

网络出版地址: <https://link.cnki.net/kcms/detail/23.1390.U.20240305.1031.003>

中图分类号: U661.31 文献标志码: A 文章编号: 1006-7043(2024)04-0758-06

## Automatic labeling method for underwater acoustic signal samples based on multilayer optimal convolution

WANG Hongbin<sup>1</sup>, ZHANG Shuai<sup>1</sup>, HE Ming<sup>1</sup>, CHEN Xiake<sup>2</sup>

(1. College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China; 2. College of Computer Science and Information Engineering, Heilongjiang University of Science and Technology, Harbin 150022, China)

**Abstract:** The application of deep learning in underwater acoustic research often faces problems such as large data volume requirements and current sample size limitations. Herein, the best convolution kernel is selected using the similarity-based optimization method to extract representative features. Then, by exploring the layer feature fusion strategy, the multilayer convolution output is superimposed to obtain comprehensive feature information. This study proposes a multilayer optimized convolutional network model that can effectively solve such problems through optimization, feature fusion, and attention mechanisms. Finally, a reduction strategy is used to optimize the model, which effectively shortens the operation time. The experimental results reveal that the best annotation accuracy of the model on the data set is 1.12% of the high baseline model, and the running time is reduced by 43.5%. Therefore, this model improves the accuracy and efficiency of underwater acoustic signal labeling.

**Keywords:** underwater acoustic signal; automatic annotation; voiceprint recognition; multilayer optimal convolution model; time optimization; attention mechanism; character merger

水下声信号标注的发展缓慢,导致水声信号的模式识别任务面临水下声信号数据集少,数据集的数据量不够大以及数据集不够准确的现状。并且由这些数据集训练出的模型泛化能力较低,对细微的噪声和干扰敏感,不能正确进行模式识别。传统的标注方法通过人工对声音信号进行标注,不仅费时

费力,经济效益也不高,同时受标注人员专业性的限制,其标注的准确性往往也不能达到要求。

声纹识别技术<sup>[1-3]</sup>通过对一种或多种声音信号的特征进行分析,可辨别声音信号,判断某一个声音信号是否由某物或者某人发出。在声纹识别领域,一般通过声学模型提取出声纹的声学特征,再根据得到的声学特征采用识别算法,对声音进行识别分类、标注。一般常用的声纹特征为线性预测倒谱系数(linear prediction cepstral coefficients, LPCC)<sup>[4-5]</sup>、梅尔频率倒谱系数(Mel Frequency cepstrum coeffi-

收稿日期: 2022-06-13. 网络出版日期: 2024-03-05.  
基金项目: 基础科研项目(JCKY2019604C004).  
作者简介: 王红滨,男,副教授;  
何鸣,男,助理研究员,博士.  
通信作者: 何鸣, E-mail: heming@hrbeu.edu.cn.

cient, MFCC)<sup>[6-7]</sup>等。

传统的识别技术一般基于统计模型,例如隐马尔可夫模型(Hidden Markov model, HMM)<sup>[8]</sup>,高斯混合模型(Gaussian mixture model, GMM)<sup>[8]</sup>,高斯混合模型-通用背景模型(Gaussian mixture model-universal background model, GMM-UBM)<sup>[9]</sup>等,属于早期常用的声纹识别方法。这些方法模型在现在的部分场景中虽然还可以继续使用,但是它们使用场景过于单一、使用范围小,不能使用在一些场景复杂的应用中,而现在的声纹识别多要求应用于复杂的场景中<sup>[10]</sup>。因此,众多的深度学习方法被应用到声纹的模式识别任务中。基于卷积神经网络(convolutional neural networks, CNN)<sup>[11]</sup>、循环神经网络(recurrent neural network, RNN)<sup>[12]</sup>等多种神经网络的各种声纹识别系统层出不穷。其中有百度提出的一种端到端的 Deep Speaker<sup>[13-15]</sup>,谷歌推出的 Google Cloud 声纹识别平台。同时,在国际声纹识别比赛中,各种基于深度学习的声纹识别模型屡次刷新最好成绩,目前单网络模型中识别率最好的是 ECAPA-TDNN<sup>[16]</sup>模型,该模型是一种基于 CNN 进行改进的模型。同时,还有将不同的网络模型进行融合,例如将 CNN 和 RNN 相结合的 RBFNN-CNN<sup>[17]</sup>。

本文使用声纹识别技术<sup>[1]</sup>进行水下声信号数据自动标注,利用传统的声学模型计算出水下声信号的声学特征。针对水声信号自动标注任务,基于深度学习的方法,提出了多层优选卷积(multilayer optimal convolution, MOC),并进行水声信号特征嵌入的提取,完成分类及标注。

## 1 基于优选卷积的多层优选卷积模型

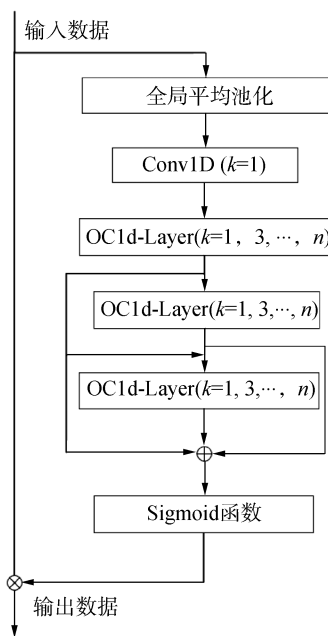
### 1.1 优选卷积的 OCA-Block

SE-Block 通过使用挤压<sup>[18]</sup>与激励操作<sup>[19]</sup>得到不同通道的权重,模型可自主判定哪些特征嵌入的重要性。但却有可能导致部分重要的特征嵌入丢失,从而降低了模型的性能<sup>[20]</sup>。为了减少该操作带来的特征嵌入的损失,本文使用 1 个一维卷积层取代 SE-Block 中的第 1 个全连接层,减少了使用全连接层带来巨大的参数量。本文通过使用卷积层,可有效地减少参数计算量,提升了模型的计算速度,减少运算时间。相较于全连接层,使用一维卷积层能够捕捉到输入更多的低级特征,更深层的卷积层可以基于此进一步提取更高级语义特征,因而能够大大减少特征嵌入损失。

本文在一维卷积层后引入优选卷积层(optimal convolution layer, OC-Layer), OC-Layer 利用优选算法找出每层中最合适的卷积核,通过优选卷积操作能够在声学信号特征中进行小范围内跨信道交互,

以便获取更优、更具有代表性的特征嵌入。考虑到不同的 OC-Layer 操作之后得到的特征信息不同,因此采用多层特征融合思想,将多层 OC-Layer 的输出进行聚合。

基于以上想法,本文提出了一种优选卷积层注意力机制(optimal convolution attention mechanism, OCA),本文称为 OCA-Block,本文采用一维优选卷积层即为 OC1d-Layer。模块结构如图 1 所示, $k$  为卷积核大小, $n$  为奇数,表示在优选层中有  $k = (1, 3, \dots, n)$  的卷积核。



注: Conv1D 为一维卷积

图 1 OCA-Block 结构图

Fig. 1 OCA-block structure diagram

### 1.2 使用优选算法的 OC1d-Layer

本文在每个 OC1d-Layer 中设置了多个大小不同的卷积核。卷积核尺寸与处理过的声学模型有关,利用其信号的声学特征维度设置卷积层的最大核尺寸。本文数据集中的声学特征为梅尔频谱,频域维度尺寸为 80 维, $k=7$  时最大的卷积核尺寸为:

$$k = \text{lbc} \quad (1)$$

式中: $c$  为声学特征的频域维度的大小; $k$  为卷积核尺寸,向上取整,若  $k$  为偶数,则  $k+1$ 。

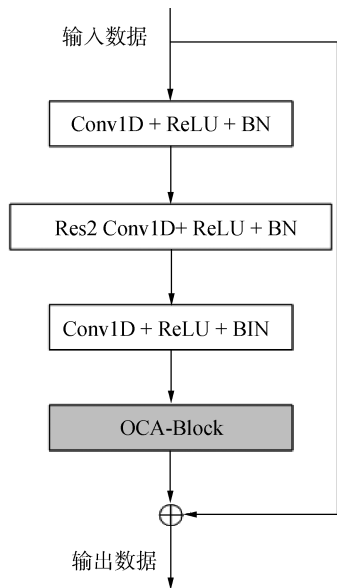
本文对更大尺寸的卷积核也进行了研究。在每个 OC1d-Layer 中都有  $(n+1)/2$  个卷积核即  $(1, 3, \dots, n)$ 。针对这些不同尺寸的卷积核,本文使用相似度方法计算每 2 个不同卷积核运算出来的输出,找出相似度最优的 2 个卷积核,取卷积核尺寸大的作为该层的卷积核。因为 2 个不同尺寸的卷积核,大的卷积核卷积后产生的数据参数较少。后面的 OC1d-Layer 以此类推。本文使用了余弦相似度、汉明距离、斯皮尔曼相关系数<sup>[21-23]</sup>这 3 个相似度算法进行

实验研究。

针对 OCA-Block, 考虑到 OC1d-Layer 作为整个模型中的子模块, 层数不宜过深, 经过实验验证, 3 层是比较合理的设置。

### 1.3 优化运算时间的 OCA-Res2Block

本文通过将注意力机制以及残差思想<sup>[24-26]</sup>进行结合, 提出了如图 2 所示的优选卷积残差模块, 称为 OCA-Res2Block。在该模块中首先使用了一个 Res2Block 模块, 然后在该模块后面接入一个基于注意力的 OCA-Block。简单地增加模块数量并不一定能够对模型的性能产生积极影响, 反而可能会降低模型的运行效率。本文进一步研究了 OCA-Res2Block 模块的个数, 并展开了一系列探索性实验。



注: 1. Conv1D 为一维卷积, 2. BN 为批处理归一化, 3. ReLU 为激活函数。

图 2 OCA-Res2Block 结构

Fig. 2 OCA-Res2Block structure diagram

### 1.4 基于优选卷积的 MOC

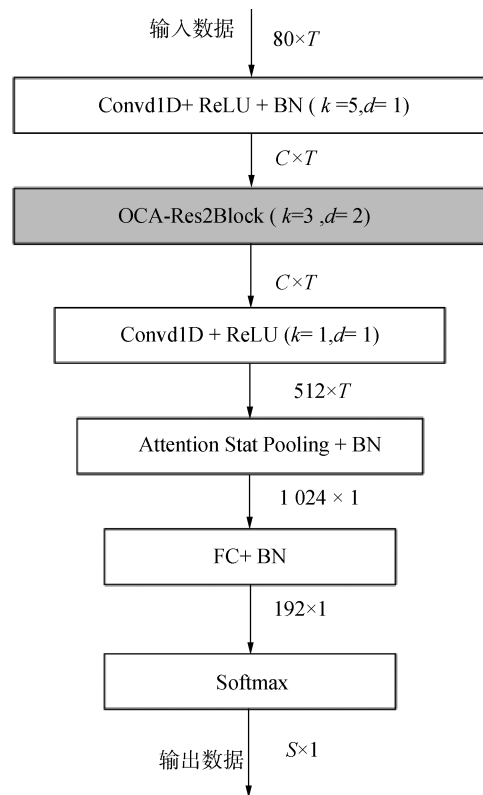
本文中多种声纹识别系统<sup>[27-28]</sup>进行了研究, 并提出了图 3 所示的 MOC, 将多层特征融合和注意力机制结合。

## 2 实验结果与分析

### 2.1 数据集与实验设备

本文中采取浙江省德清县对河口水库音频作为的数据集。为了保证数据集具有普遍性, 该数据集考虑到了时间、温度、天气、距离、深度、频率等多种环境因素。整个数据集包含 3 个时间段 (上午、中午、下午); 5 种深度 (1、3、5、7、9 m); 2 种频率 (单频 CW、调频 LFM); 2 种距离 (75、150 m) 即声源到声音接收器的距离; 32 个通道; 每种音频数据录制 90 s。本文从整个数据集中抽取 2 组数据, 同时为

了便于实验将 90 s 的完整数据切分为 900 份 0.1 s 的片段, 即 2 组数据包含 75 MB、150 MB 的 CW、LFM, 同时每组数据有 32 个通道。由于使用设备的原因在进行梅尔特征提取时, 采样率设置为 1 000 000。在实验数据集中, 本文将数据集设置为训练集与测试集 2 种, 所占比例为 80%:20%。



注: 1.  $C, T$  分别为输入的大小, 2.  $k$  为卷积核大小, 3.  $d$  为空洞卷积的膨胀率, 4.  $S$  为类别数, 5. FC 为全连接层

图 3 MOC 结构

Fig. 3 MOC structure diagram

### 2.2 实验评估指标

在实验中, 本文将标注准确率作为评价指标。标注准确率为在标注任务过程之中, 标注正确的样本占总测试样本的比值:

$$A_{cc} = \sum S_t / L(S) \quad (2)$$

式中:  $S_t$  为在数据集  $S$  中标注为  $t$ , 即标注正确的数据个数;  $L(S)$  为数据集  $S$  的总样本数。

### 2.3 结果及分析

本文中正常水声信号和其中的噪音进行分类标注。为从不同角度探索模型优化的可能性, 进行了 4 种类型的实验。分别关注于简化 OCA-Res2Block 模块数量以实现时间优化、OC1d-Layer 的层数、OC1d-Layer 中卷积核的选择, 以及不同卷积核相似度计算方法。每种类型的实验都包括了多个实验组。作为参照, 本文选用了当前单网络模型中识别率最高的 ECAPA-TDNN 作为基线模型。在

MOC 的默认设置中,配置了一个 OCA-Block, OC1d-Layer 的层数为 3, OC1d-Layer 中的卷积核尺寸设置为  $K=7$ , 即每个 OC1d-Layer 包含 1、3、5、7 共 4 种尺寸的卷积核, 相似度计算采用斯皮尔曼相关系数。

2.3.1 OCA-Res2Block 的时间优化实验

本文以运行时间及数据标注的准确率作为评判

指标, 采用 MOC, 分别设置一个 OCA-Res2Block、2 个 OCA-Res2Block, 标记为  $F_1$ 、 $F_2$ , 除了 OCA-Res2Block 个数, 其他参数为默认设置参数, 包括 OC1d-Layer 的层数设置为 3, OC1d-Layer 中的卷积核尺寸设置为  $K=7$ 。实验时间优化实验结果如表 1 所示。

表 1 OCA-Res2Block 时间优化实验结果表

Table 1 OCA-Res2Block time optimization experiment results

模型	基线		$F_1$		$F_2$	
	t/s	准确率/%	t/s	准确率/%	t/s	准确率/%
75M-CW	6 488.3	95.28	<b>3 665.8</b>	<b>96.35</b>	8 221.6	95.77
75M-LFM	6 663.7	97.92	<b>3 762.5</b>	<b>99.04</b>	9 026.5	99.04
150M-CW	5 524.4	92.45	<b>3 994.6</b>	<b>93.23</b>	9 058.1	92.69
150M-LFM	5 810.8	97.53	<b>3 544.2</b>	<b>98.70</b>	8 255.8	98.44

由表 1 可得,  $F_1$  模型的运算时间远低于  $F_2$  模型和基线模型, 运算时间分别优化了 55.4%、43.5%, 但是  $F_1$  在准确率方面与  $F_2$  模型最高的准确率持平, 比基线模型最多高出 1.17%。同时可以发现  $F_2$  模型的运算时间比基线模型高出 21.1%, 但是其准确率要高于基线模型。由此说明使用  $F_1$  模型即本文提出的 OCA-Res2Block 模块的使用注意力机制和优选算法是有效的, 以及适当的减少模型模块的构成能够减轻模型的计算压力和运行时间。同时可以看到在这 2 组数据中提出的 MOC 模型比基线模型从运行时间和准确率 2 个方面都有较大提升, 即 MOC 全面优于基线模型。

2.3.2 OCA-Block 中 OC1d-Layer 层数设置实验

在该类实验中, 本文用数据标注的准确率作为评判指标, 在实验过程中, 本文将 OC1d-Layer 层数分别设置为 3、5、7 层标记为  $O_3$ 、 $O_5$ 、 $O_7$ , 除了 OC1d-Layer 层数, 其他参数为默认设置参数, 包括 OC1d-Layer 的层数设置为 3, OC1d-Layer 中的卷积核尺寸设置为  $K=7$ 。实验结果如表 2 所示。

表 2 OC1d-Layer 层数实验结果表

Table 2 OC1d-Layer layer experiment results

模型\数据	75M-LFM	150M-CW	150M-LFM	75M-CW
基线	95.28	97.92	92.45	97.53
$O_3$	<b>96.35</b>	<b>99.04</b>	<b>93.23</b>	<b>98.70</b>
$O_5$	95.83	98.83	<b>93.23</b>	98.44
$O_7$	95.83	98.56	<b>93.23</b>	98.44

由以上结果可分析出: 在 MOC 中将优选卷积层的层数设置为 3 层时, 在距离为 75 M 时 CW 与 LFM 中最大的准确率分别为 96.35% 以及 99.04%; 在距离为 150 M 时 CW 与 LFM 中最大的准确率分别为

93.23% 及 98.70%。说明本文中提出 3 层 OC1d-Layer 有效, 随意加深层数对模型并无益处。因为并行的逻辑层只增加收敛速度, 而本模型并没有随便增加层数。并且过多的层数可能会导致模型本身更加复杂, 如果训练数据本就不多, 很容易造成过拟合的问题。通过实验结果可以发现无论几层 OC1d-Layer 的 MOC 的准确率都比基线模型的准确率更高, 即说明 MOC 优于基线模型的。

该类实验说明本文中提出的 OC1d-Layer 有效, 并且从实验的结果可以推测出层数的阈值为 3。

2.3.3 OC1d-Layer 中优选算法的对比实验

在该类实验中, 本文用数据标注的准确率作为评判指标, 并验证优选算法的有效性以及最佳的卷积核尺寸。在实验过程中, 除了卷积核尺寸和个数, 其他参数为默认设置参数, 包括 OC1d-Layer 的层数设置为 3, OC1d-Layer 中的卷积核尺寸设置为  $K=7$ 。在实验中, 本文设置 OC1d-Layer 中的卷积核尺寸以及卷积核数如表 3 所示。

表 3 卷积核尺寸以及卷积核数表

Table 3 Convolution kernel size and convolution kernel number table

标记	每层卷积核数	每层卷积核尺寸
$K_5$	3	1、3、5
$K_7$	4	1、3、5、7
$K_9$	5	1、3、5、7、9
$K_{11}$	6	1、3、5、7、9、11

优选算法实验结果如表 4 所示。

由以上结果可得, 将 OC1d-Layer 中的最大卷积核尺寸设置为 7, 在这 2 组数据中标注准确率均达到最优,  $K_7$  模型优于基线模型, 本文提出的优选算法有效。同时, 也可以发现在 75 M 的 CW 数据中,

$K_5$ 、 $K_{11}$  的标注准确率低于基线模型的标注准确率, 说明了本文中设置的  $K_7$  是当前最优的, 也证明了在 OC1d-Layer 中设置卷积核尺寸以及个数应该按照声学模型产生的声信号特征进行设置。

表 4 优选算法实验结果表

模型	75M-LFM	150M-CW	150M-LFM	75M-CW
基线	95.28	97.92	92.45	97.53
$K_5$	95.25	98.56	92.71	97.92
$K_7$	<b>96.35</b>	<b>99.04</b>	<b>93.23</b>	<b>98.70</b>
$K_9$	95.28	98.56	92.84	98.44
$K_{11}$	94.24	98.77	92.84	98.44

### 2.3.4 OC1d-Layer 中相似度算法的对比实验

该类实验主要将 3 种相似度算法进行研究。本文用数据标注的准确率作为评判指标, 在实验过程中, 采用的相似度函数分别为余弦相似度、汉明距离、斯皮尔曼相关系数标记为  $M_c$ 、 $M_h$ 、 $M_s$ , 除了相似度算法, 其他参数为默认设置参数, 包括 OC1d-Layer 的层数设置为 3, OC1d-Layer 中的卷积核尺寸设置为  $K=7$ 。相似度算法实验结果如表 5 所示。

表 5 相似度算法实验结果表

模型\数据	75M-LFM	150M-CW	150M-LFM	75M-CW
基线	95.28	97.92	92.45	97.53
$M_c$	95.86	<b>99.04</b>	92.71	<b>98.70</b>
$M_h$	96.03	98.42	92.84	97.92
$M_s$	<b>96.35</b>	<b>99.04</b>	<b>93.23</b>	98.44

由以上结果可分析出, 在 75 M 的 LFM 数据中, 余弦相似度与斯皮尔曼相关系数算法准确率为 99.04%; 75 M 的 CW 数据中, 斯皮尔曼相关系数算法准确率为 96.35%。而在 150 M 的 CW 数据中, 斯皮尔曼相关系数算法准确率为 93.23%; 在 150 M 的 LFM 数据中, 余弦相似度算法准确率为 98.70%。在 OC1d-Layer 中使用不同的相似度算法得到的准确率不同, 但在相同的数据集中, 准确率的差别不到 1%。从整体来说, 在这 3 种算法中, 斯皮尔曼相关系数算法较为合适。本文使用斯皮尔曼相关系数算法作为相似度算法合理。在模型中无论使用哪种算法, 其标注准确率都高于基线模型。

## 3 结论

1) 本文使用优选、注意力机制、多层特征融合等策略, 实现了 MOC 在运算时间和标注准确率方面

的预期目标。

2) 本文 MOC 在河口水库采集的音频数据集上取得了标注准确率为 99.04%, 高于基线模型的 97.92%; 同时运行时间较少了 43.5%。本文提出的方法使用时间更短, 并且对水声数据的标注准确率更高。

3) 相同环境下, LFM 波形数据的标注准确率高于 CW 波形数据。该模型的使用提高了水声信号标注的准确率和效率。

在以后研究中, 将进一步研究 MOC 的改进, 同时考虑将 2 个网络模型并行使用的研究, 进一步提高水下声信号自动标注的标注性能。

## 参考文献:

- [1] YANG H, BYUN S H, LEE K, et al. Underwater acoustic research trends with machine learning: active SONAR applications[J]. Journal of ocean engineering and technology, 2020, 34(4): 277-284.
- [2] SHAN Shuaijie, LIU Jianbao, DUN Yaowu. Prospect of voiceprint recognition based on deep learning[J]. Journal of physics: conference series, 2021, 1848(1): 012046.
- [3] KHDIER H Y, JASIM W M, ALIESAWI S A. Deep learning algorithms based voiceprint recognition system in noisy environment[J]. Journal of physics: conference series, 2021, 1804(1): 012042.
- [4] MINI P P, THOMAS T, GOPIKAKUMARI R. EEG based direct speech BCI system using a fusion of SMRT and MF-CC/LPCC features with ANN classifier[J]. Biomedical signal processing and control, 2021, 68: 102625.
- [5] 付进, 许婉琰, 王燕等. 复倒谱域水声信道多途抑制技术[J]. 哈尔滨工程大学学报, 2015, 36(9): 1188-1193.  
FU Jin, XU Wanyan, WANG Yan, LIANG Guolong. Anti-multipath technique of underwater acoustic channel in complex cepstrum domain[J]. Journal of Harbin Engineering University, 2015(9): 1188-1193.
- [6] ZHU Qiang, WANG Zhong, DOU Yunfeng, et al. Whispered speech conversion based on the inversion of mel frequency cepstral coefficient features[J]. Algorithms, 2022, 15(2): 68.
- [7] 王大宇, 王志欣, 张光普. 基于改进谱减算法的水声通信信号检测方法[J]. 应用科技, 2020, 47(3): 69-73.  
WANG Dayu, WANG Zhixin, ZHANG Guangpu. Detection method of underwater acoustic communications signal based on improved spectral subtraction algorithm[J]. Applied science and technology, 2020, 47(3): 69-73.
- [8] SHIVAPRASAD S, SADANANDAM M. Optimized features extraction from spectral and temporal features for identifying the Telugu dialects by using GMM and HMM[J]. Ingénierie des systèmes d'information, 2021, 26(3): 275-283.
- [9] ZHANG Jing. Research on cantonese phonetic feature extraction algorithm based on GMM-UBM[C]//2019 Chinese Control and Decision Conference (CCDC). Piscataway, NJ: IEEE, 2019: 4034-4038.

- [10] 杨爽,曾向阳. 基于多尺度稀疏简单循环单元模型的水声目标识别方法[J]. 哈尔滨工程大学学报,2022,43(7):958-964.  
YANG Shuang, ZENG Xiangyang. Underwater acoustic target recognition method based on the multi-scale sparse simple recurrent unit model[J]. Journal of Harbin Engineering University,2022,43(7):958-964.
- [11] GU Jiuxiang, WANG Zhenhua, KUEN J, et al. Recent advances in convolutional neural networks[J]. Pattern recognition, 2018, 77(C): 354-377.
- [12] REJAIBI E, KOMATY A, MERIAUDEAU F, et al. MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech[J]. Biomedical signal processing and control, 2022, 71: 103107.
- [13] LI C, MA X, JIANG B, et al. Deep speaker: an end-to-end neural speaker embedding system[J]. Arxiv, 2017. DOI:10.48550/arXiv.1705.02304.
- [14] 汤礼颖,贺利乐,何林等. 一种卷积神经网络集成的多样性度量方法[J]. 智能系统学报,2021,16(6):1030-1038.  
TANG Liying, HE Lile, HE Lin, et al. Diversity measuring method of a convolutional neural network ensemble[J]. CAAI Transactions on Intelligent Systems, 2021, 16(6): 1030-1038.
- [15] 王同,苏林,任群言等. 循环神经网络在浅海声速-声源联合反演中的应用[J]. 哈尔滨工程大学学报,2021,42(8):1133-1139.  
WANG Tong, SU Lin, REN Qunyan, WANG Wenbo, MA Li. Application of the sound speed profile and sound source location in shallow waters[J]. Journal of Harbin Engineering University, 2021, 42(8): 1133-1139.
- [16] DESPLANQUES B, THIENPOND J, DEMUYNCK K. ECAPA-TDNN: emphasized channel attention, propagation and aggregation in TDNN based speaker verification[J]. Arxiv:2005.07143.
- [17] BOU NASSIF A, ALNAZZAWI N, SHAHIN I, et al. A novel RBFNN-CNN model for speaker identification in stressful talking environments[J]. Applied sciences, 2022, 12(10): 4841.
- [18] JIN Xin, XIE Yanping, WEI Xiushen, et al. Delving deep into spatial pooling for squeeze-and-excitation networks[J]. Pattern recognition, 2022, 121: 108159.
- [19] 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.
- [20] 赵文清,杨盼盼. 双向特征融合与注意力机制结合的目标检测[J]. 智能系统学报,2021,16(6):1098-1105.  
ZHAO Wenqing, YANG Panpan. Target detection based on bidirectional feature fusion and an attention mechanism[J]. CAAI Transactions on Intelligent Systems, 2021, 16(6): 1098-1105.
- [21] MAY J O, LOONEY S W. On sample size determination when comparing two independent spearman or Kendall coefficients[J]. Open journal of statistics, 2022, 12(2): 291-302.
- [22] GRABOWSKI S, KOWALSKI T M. Algorithms for all-pairs Hamming distance based similarity[J]. Software: practice and experience, 2021, 51(7): 1580-1590.
- [23] LI Jing, LIN Song, YU Kai, et al. Quantum K-nearest neighbor classification algorithm based on Hamming distance[J]. Quantum information processing, 2021, 21(1): 18.
- [24] INDIA M, SAFARI P, HERNANDO J. Self multi-head attention for speaker recognition[C]//Interspeech 2019. ISCA; ISCA, 2019: 1906-2015.
- [25] LEI Zeyu, WANG Yan, LI Zijian, et al. Attention based multilayer feature fusion convolutional neural network for unsupervised monocular depth estimation[J]. Neurocomputing, 2021, 423: 343-352.
- [26] 李晓峰,邢金明. 融合时空多特征表示的运动人体目标跟踪算法[J]. 应用科技,2020,47(4):26-31.  
LI Xiaofeng, XING Jinming. Tracking algorithm of a moving human body target using multi-feature representation of fused time and space[J]. Applied science and technology, 2020, 47(4): 26-31.
- [27] HU Zhangfang, SI Xingtong, LUO Yuan, et al. Speaker recognition based on 3DCNN-LSTM[J]. Speaker recognition based on 3DCNN-LSTM, 2021, 29. 0(2. 0).
- [28] LI Lantian, LIU Ruiqi, KANG Jiawen, et al. CN-Celeb: multi-genre speaker recognition[J]. Speech communication, 2022, 137: 77-91.

### 本文引用格式:

王红滨,张帅,何鸣,等. 基于多层优选卷积的水声信号样本自动标注方法[J]. 哈尔滨工程大学学报,2024,45(4):758-763.  
WANG Hongbin, ZHANG Shuai, HE Ming, et al. Automatic labeling method for underwater acoustic signal samples based on multilayer optimal convolution[J]. Journal of Harbin Engineering University, 2024, 45(4): 758-763.