

DOI: 10.3969/j.issn.1671-0673. 2026. 01. 003

融合多模态对比学习的仿嘀嗒声水声 仿生通信信号识别方法

杨硕硕, 王 彬, 郑 娄, 孟钰婷

(信息工程大学, 河南 郑州 450001)

摘要: 为提升复杂水声环境中短样本仿嘀嗒声仿生通信信号识别性能, 提出一种融合多模态对比学习(MCL)的识别方法。首先, 通过引入时频与融合特征模态, 直接从信号细微结构中提取判别性特征, 降低了对长序列统计稳定性的依赖。其次, 利用对比学习机制拉近同类样本、推远异类样本, 迫使模型聚焦于对多径干扰不敏感的鲁棒性特征。最后, 文本模态的引入为特征学习提供了语义约束, 引导模型关注类别本质属性。仿真与湖试实验结果表明, 该方法在短样本及水声多径信道条件下的识别性能显著优于现有基于时延差统计与统计—卷积神经网络算法。其中, 湖试实验在 0.7 s 短样本条件下的识别准确率达到 88.70%。

关键词: 水声仿生通信; 信号识别; 特征融合; 多模态学习; 对比学习

中图分类号: TN911.7

文献标识码: A

文章编号: 1671-0673(2026)01-0019-08

Recognition of Hydroacoustic Biomimetic Camouflage Click Communication Train Using Fused Multimodal Contrastive Learning

YANG Shuoshuo, WANG Bin, ZHENG Lou, MENG Yuting

(Information Engineering University, Zhengzhou 450001, China)

Abstract: To enhance the recognition performance of short-sample click-like bionic hydroacoustic communication signals in complex underwater environments, a recognition method based on fused multimodal contrastive learning is proposed. Firstly, by incorporating time-frequency and fused feature modalities, discriminative features are directly extracted from fine-grained signal structures, reducing reliance on long-sequence statistical stability. Secondly, the samples of the same class are pulled together and those of different classes are pushed apart to the model to focus on robust features that are insensitive to multipath interference using the contrastive learning mechanism. Finally, the introduction of a text modality provides semantic constraints for feature learning, guiding the model to capture category-related essential attributes. Simulation and lake trial results demonstrate that the proposed method significantly outperforms existing algorithms based on time-delay statistics and statistical-convolutional neural networks under short-sample and multipath hydroacoustic channel conditions. Specifically, the lake trial achieves a recognition accuracy of 88.70% with a short sample length of 0.7 s.

Key words: hydroacoustic biomimetic communication; signal recognition; feature fusion; multimodal learning; contrastive learning

水声仿生通信是一种新兴的水下隐蔽通信方式, 其中仿嘀嗒声仿生通信信号(Biomimetic Camouflage Click Communication Train, BCCCT)主要模

拟鲸目动物典型声学特征, 并直接使用原始或者模拟的鲸目动物真实嘀嗒声信号(Real Click Train, RCT)作为通信载体, 将信息调制在相邻嘀嗒

收稿日期: 2025-10-16; 修回日期: 2025-11-15

作者简介: 杨硕硕(2001—), 男, 硕士生, 主要研究方向为水声信号分析处理。E-mail: 2110456710@qq.com

声信号的时延差间隔中,在隐蔽通信中具有重要应用价值^[1-3]。水声仿生通信信号识别问题是指在复杂的海洋环境中有效区分仿生通信信号与自然生物声学信号问题,在软件无线电水声通信调制解调器设计和海洋哺乳生物目标跟踪等领域具有重要意义。本文主要研究在短样本和复杂水声环境下如何区分仿嘀嗒声通信信号与鲸类嘀嗒声的问题。本文所述复杂水声环境,主要是指对仿嘀嗒声通信信号识别构成严峻挑战的复杂水声多径信道,其引起的码间干扰会严重破坏信号的统计特征,且环境噪声会掩盖信号的细微特征。

目前,仿嘀嗒声水声仿生通信信号的识别研究尚处于起步阶段,识别方法主要可分为两类。第1类是时延差统计方法(Time Delay Statistics, TDS),首先对单个嘀嗒声进行定位并计算相邻嘀嗒声的时延差间隔,进而根据时延差间隔是否出现等间隔阶梯分布的统计特征,判断信号是否为BCCCT^[4-5]。第2类为统计-卷积神经网络方法(Statistics and Convolutional Neural Network, S-CNN),在统计相邻嘀嗒声时延差间隔之后,将其转化为时延差概率密度图并输入卷积神经网络(Convolutional Neural Network, CNN)进行分类^[6]。现有的两类方法都是以时延差统计特征为基础,将等间隔阶梯分布或者等间隔多峰分布特征作为BCCCT的判别依据。然而,实际海洋环境中RCT属于短突发信号,持续时间通常短于1 s,为维持高隐蔽性,BCCCT的长度也相应较短,导致难以获取稳定的统计特征。此外,水声多径信道也会破坏上述统计特征。尽管文献[6]提出基于自相关的去多径方法,但其效果依赖阈值选取,难以适应复杂多变的水声信道。因此,在复杂水声多径信道环境下对短突发的BCCCT的准确识别成为亟待解决的实际问题。

针对上述问题,本文借鉴对比学习思想^[7],提出了一种融合多模态对比学习(Multimodal Contrastive Learning, MCL)的仿嘀嗒声通信信号识别方法。本文建立了融合时频模态、融合特征模态、文本模态的多模态对比学习框架。在训练阶段,对目标样本分别从上述3个模态提取嵌入向量,并通过对比学习方法优化模型参数;在识别阶段,使用描述类别的自然语言替代传统标签,将所有描述性文本经字节对编码后输入文本模态编码器,获得文本模态的嵌入向量,待识别样本的时频图输入时频编码器后获得时频模态的嵌入向量;最后通过计算时频模态嵌入向量与所有候选文本模态嵌入向量的余弦

相似度,将相似度最高的类别作为识别结果。

该方法的设计旨在直接应对短样本与复杂多径信道带来的挑战。首先,通过引入时频模态和融合特征模态,模型能够直接学习信号在时频域上的细微结构差异与多种声学特征的联合分布,而非依赖在短样本中难以稳定建立、且在多径信道下极易被破坏的宏观时延差统计特征,从而从根本上提升了特征的鲁棒性。其次,对比学习机制通过拉近同类样本、推远异类样本的方式,在嵌入空间中强制模型学习到了对多径干扰不敏感的、更具判别性的本质特征表示。最后,文本模态的引入为特征学习提供了语义监督,可向模型注入先验知识,进一步引导特征学习过程,稳定识别性能。仿真和湖试实验验证了所提方法在短样本和复杂水声多径信道条件下,相比基线方法具有更高的识别准确率和鲁棒性。

1 问题模型

1.1 信号特点

1.1.1 RCT特点

RCT是鲸目动物导航及目标定位时发出的声音,具有脉冲窄、带宽大、指向性强等特点,并且嘀嗒声信号常出现在哨声信号间隙,或与哨声同时出现,不掺杂有哨声的纯净的嘀嗒声持续时间较短,最长为几秒钟左右,大多在1 s以内,属于短突发信号^[8]。图1是来自沃特金斯海洋哺乳动物声音数据库^[9-10]的虎鲸叫声信号的RCT时频图。其中:红框内的为较纯净的嘀嗒声信号;红框两侧为哨声信号。

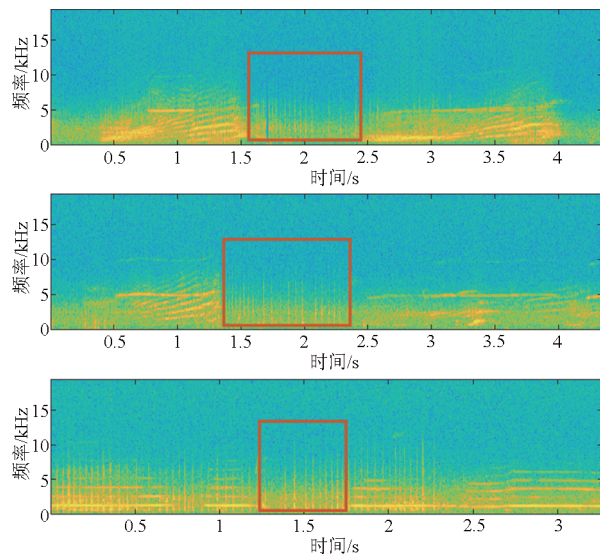


图1 RCT时频图

1.1.2 BCCCT 特点

典型的BCCCT采用时延差编码,如图2所示,这种信号使用从真实鲸目动物叫声中截取的单个嘀嗒声信号Click_{*i*}作为信号码元,将信息调制到相邻嘀嗒信号的时间间隔 d_i 中^[11-12]。

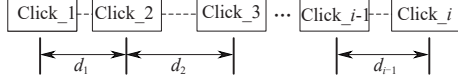


图2 时延差编码原理

设 d_{base} 表示基础编码时间, Δd 表示单位编码时间,每个时延差包含 n 比特信息,则第 i 个和第 $(i+1)$ 个嘀嗒声之间的时延差 d_i 可表示为

$$d_i = d_{\text{base}} + k_i \Delta d, k_i = 0, 1, \dots, 2^n - 1 \quad (1)$$

当采用特定方法定位单个嘀嗒声脉冲信号,并计算相邻嘀嗒声信号的时延差间隔概率密度分布图时,BCCCT会出现等间隔多峰统计特征,而RCT则呈现出随机均匀分布,不具备等间隔多峰特性,该差异是现有识别方法的主要依据。

1.2 问题模型

假设待分类的接收信号为

$$s_r(k) = s_i(k) \otimes h(k) + n(k) \quad (2)$$

式中: $s_i(k)$ 为RCT或BCCCT; $h(k)$ 表示水声信道冲激响应; $n(k)$ 表示加性高斯白噪声。

当前主流的仿嘀嗒声通信信号识别方法都是以相邻嘀嗒声时延差间隔统计特征作为基础。以文献[6]方法为例,其采用基于双滑动窗口的嘀嗒声信号提取方法定位嘀嗒声信号位置,之后统计所有相邻嘀嗒声时延差间隔 $\{T_i\}_{i=1}^{N_T}$,并采用高斯核概率密度函数表征 T_i 的分布规律,计算公式为

$$P(t) = \frac{1}{\sqrt{2\pi} N_T h} \sum_{i=1}^{N_T} \exp \left[-\frac{1}{2} \left(\frac{t - T_i}{h} \right)^2 \right] \quad (3)$$

式中: N_T 为 T_i 的总个数; h 为带宽。 h 计算公式为

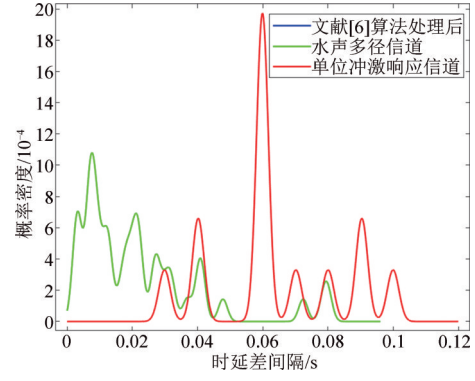
$$h = 0.15\sigma / \sqrt{N_T} \quad (4)$$

式中, σ 表示 T_i 的标准差。

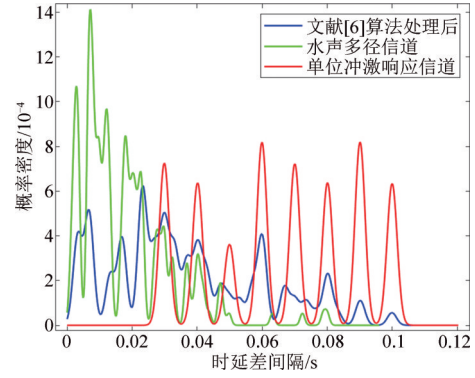
为分析现有方法局限性,设置BCCCT参数 $d_{\text{base}} = 30 \text{ ms}$ 、 $\Delta d = 10 \text{ ms}$ 、 $n = 3 \text{ bit}$,由于时延差间隔统计特征是现有TDS与S-CNN识别方法的核心依据,使用文献[6]中的方法分别对样本长度为1、4、7、10 s的BCCCT提取时延差间隔统计特征。仿真信道 $h(k)$ 分别设为单位冲激响应信道和水声多径信道,水声多径信道的信道系统函数如式(5)所示:

$$H(z) = 1 + 0.502 3z^{-9.547} + 0.440 3z^{-14.719} + 0.351 4z^{-28.647} + 0.109z^{-58.302} \quad (5)$$

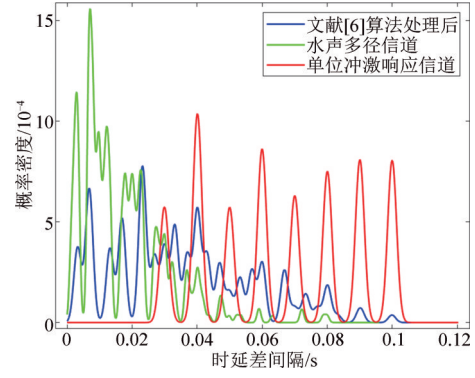
实验结果如图3所示。



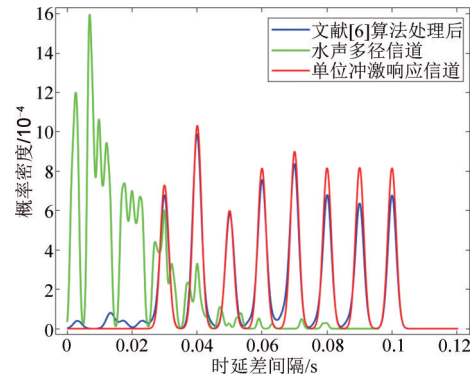
(a) 样本长度为1 s (蓝线与绿线重合)



(b) 样本长度为4 s



(c) 样本长度为7 s



(d) 样本长度为10 s

图3 BCCCT时延差概率密度图

图3中红色线代表的单位冲激响应信道下的时延差间隔统计特征,当样本长度为1 s时,未呈现等间隔多峰特征;当样本长度达到4 s及以上时,则

呈现明显的等间隔多峰统计特征。绿色线为水声多径信道下的时延差间隔统计特征,不同样本长度下等间隔多峰特征均被多径信道破坏。蓝色线为采用文献[6]中的去多径方法处理后的多径信号的时延差间隔统计特征,在样本长度达到 10 s 后才恢复等间隔多峰特征。从上述实验结果可以看出,当样本长度较短,如小于 1 s 时,时延差间隔统计特征难以有效提取,并且水声多径信道也会破坏时延差分布的等间隔多峰统计特征。

综上所述,基于时延差间隔统计特征的识别方法在短样本和水声多径信道条件下难以取得稳定的识别效果。虽然 RCT 及 BCCCT 在宏观上均表现为能量集中的宽带脉冲,在时域波形和粗粒度频谱上具有很高的相似性,但其细微的时频结构差异蕴含了关键的区分信息。这些细微差异对水声多径信道的敏感度远低于时延差统计特征,且对样本长度没有很高的要求。因此,本文选取时频模态作为目标样本的主要表示模态,以直接捕捉上述细微差异;并添加由 7 类对样本长度依赖较低的时域、频域及倒谱域特征组成的融合特征模态作为补充,从更多维度捕捉两类信号之间的细微差异。同时,采用文本模态取代传统的类别标签,以提升识别方法的灵活性和可扩展性。为综合运用 3 个模态信息,采用对比学习方法两两计算交叉熵损失以优化模型参数。基于上述思路,本文提出融合多模态对比学习的仿嘀嗒声通信信号识别方法。

2 方法基本原理

本文采用融合多模态对比学习框架,旨在从根本上解决基于时延差统计特征的方法在短样本与复杂水声多径信道下面临的瓶颈。其核心原理在于该方法摒弃了对相邻嘀嗒声之间时序结构的依赖,转而挖掘并强化信号固有的、对信道扰动不敏感的本质特征。具体而言,首先,通过引入时频模态和高维融合特征模态,模型能够直接从少量信号的细微结构中学习判别性信息,这些特征不依赖于长信号的统计稳定性,因而天然适用于短样本条件;其次,对比学习机制通过拉近同类样本、推远异类样本在嵌入空间中的距离,迫使编码器忽略由多径效应引起的信号畸变,并聚焦于由信号本身决定的鲁棒性特征;最后,文本模态的引入为特征学习提供了强语义约束,进一步引导模型关注与类别相关的本质属性。

融合多模态对比学习的仿嘀嗒声通信信号识

别方法训练和识别框架如图 4 和图 5 所示。训练阶段联合使用时频、融合特征、文本等 3 个模态,通过对比学习方法优化编码器参数;识别阶段采用时频和文本两个模态,计算待识别信号时频嵌入向量与候选文本嵌入向量的余弦相似度,并将最大值对应的类别作为识别结果。

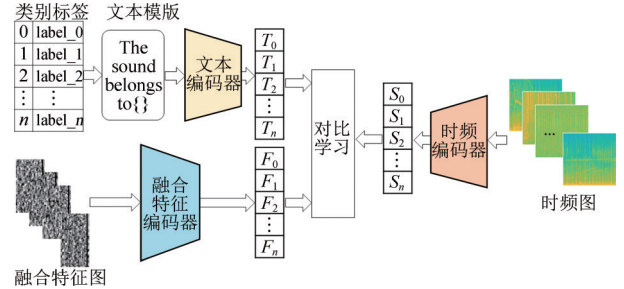


图4 所提方法训练框架

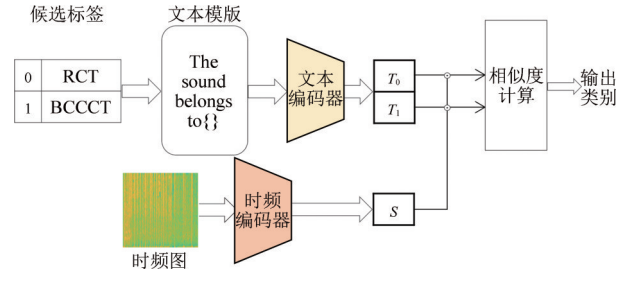


图5 所提方法识别框架

2.1 时频模态

时频模态首先计算音频信号短时傅里叶变换 (Short-Time Fourier Transform, STFT), 即:

$$\text{STFT}(\tau, \omega) = \sum_{k=-\infty}^{\infty} s_r(k) w(k - \tau) e^{-j\omega k} \quad (6)$$

式中, $w(\cdot)$ 表示窗函数。随后根据 STFT 计算对数时频谱图:

$$N_{\text{spec}} = 10 \lg |\text{STFT}(\tau, \omega)|^2 \quad (7)$$

将时频图输入时频编码器, 获得时频嵌入向量。时频编码器以 ResNet-50 为基础, 并将其全局平均池化层替换为注意力池化层^[13]。

2.2 融合特征模态

融合特征模态首先对信号进行融合特征提取预处理。对 $s_r(k)$ 进行分帧处理:

$$s_r^i(k) = s_r(k + K_{\text{shift}} \times (i - 1)) \quad (8)$$

式中: $s_r^i(k)$ 为第 i 帧子信号; K_{shift} 为帧移步长; $i = 1, \dots, N_f$; N_f 为总分帧数; $k = 0, \dots, K_{\text{win}} - 1$; K_{win} 为帧长。

对每帧子信号分别提取以下 7 类特征。

1) 过零率 (Zero Crossing Rate, ZCR) 表征频率信息, 反映信号波形单位时间过零次数^[14]:

$$R_i^{\text{ZC}} = 0.5 \left\{ \left| \text{sgn}[s_r^i(k)] - \text{sgn}[s_r^i(k - 1)] \right| \right\} \quad (9)$$

式中, $\text{sgn}[\cdot]$ 表示取符号运算。

2) 频谱质心(Spectral Centroid, SC)表示频率对信号频谱的加权平均值,反映了信号在频域上能量分布的中心位置^[15]:

$$C_i^S = \frac{\sum_{k=0}^{K/2-1} k \cdot |X_i(k)|}{\sum_{k=0}^{K/2-1} |X_i(k)|} \quad (10)$$

式中: $X_i(k)$ 为第*i*帧子信号的频谱; K 为FFT点数。

3) 频谱延展度(Spectral Spread, SS)衡量频谱能量围绕质心的分散程度^[16]:

$$S_i^S = \sqrt{\frac{\sum_{k=0}^{K/2-1} (k - C_i^S)^2 |X_i(k)|}{\sum_{k=0}^{K/2-1} |X_i(k)|}} \quad (11)$$

4) 谐波比(Harmonic Ratio, HR)和基频(Fundamental Frequency, FF)反映信号谐波特性, RCT实际鲸目动物嘀嗒声信号往往掺杂有一定的哨声信号,相比BCCCT更容易提取到谐波比和基频特征,计算公式分别为^[16-17]:

$$R_i^H = \max_{K_{\min} \leq k \leq K_{\max}} \{F_i(k)\} \quad (12)$$

$$F_i^F = \frac{\int_{\text{sampling}}}{\arg\max_{K_{\min} \leq k \leq K_{\max}} \{F_i(k)\}} \quad (13)$$

5) 谱熵(Spectral Entropy, SE)衡量信号在频域内的能量分布的不确定性或复杂度^[17]:

$$\begin{cases} E_i^S = - \sum_{b=0}^{H-1} p_b \log_2 p_b; \\ p_b = E_b / \sum_{b=0}^{K/2-1} E_b. \end{cases} \quad (14)$$

式中, E_b 表示第($b+1$)个子带能量。

6) 频谱滚降点(Spectral Rolloff, SR)标识了信号中主要能量分布的频率范围与能量开始快速衰减的频率点之间的界限,能够反映信号的能量分布特性^[18]:

$$\arg\min_{r \in \{0.1, \dots, K/2-1\}} \sum_{k=0}^r |X_i(k)| \geq C \cdot \sum_{k=0}^{K/2-1} |X_i(k)| \quad (15)$$

式中, C 为指定比例,一般取值范围为0.85~0.9。满足式(15)的*r*值所对应的频率值即为频谱滚降点。

7) 梅尔频率倒谱系数(Mel Frequency Cepstral Coefficients, MFCC)特征计算公式为^[16]

$$\text{MFCC}_i(r) = \sum_{j=1}^M \left\{ \cos \left(\frac{\pi(j-5)r}{M} \right) \times \ln \left[\sum_{k=0}^{K/2-1} |X_i(k) H_j(k)| \right] \right\} \quad (16)$$

式中: $r = 1, 2, \dots, R$ (R 为阶数); $H_j(k)$ 为第*j*个Mel三角滤波器的频率响应; M 为Mel三角滤波器的总数。Mel三角滤波器由一组三角滤波器叠加而成,模拟了人耳对高频声音不敏感的非线性特征,结构如图6所示^[19]。

得到上述7类特征后,为了使网络在训练过程中能同时学习不同维度的特征信息,从而更好地理解 and 识别特征图像中的复杂模式,本文采用早融合的方式构建融合特征图像^[20-21]。将从所有子信号中提取的7类特征按照时间顺序拼接为如图7所示的融合特征图像,括号中的数字表示特征维度。

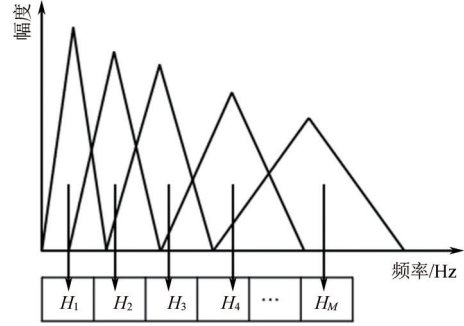


图6 Mel滤波器

1:	ZCR(1)	SC(1)	SS(1)	MFCC(36)	HR(1)	FF(1)	SE(1)	SR(1)
2:	ZCR(1)	SC(1)	SS(1)	MFCC(36)	HR(1)	FF(1)	SE(1)	SR(1)
...								
80:	ZCR(1)	SC(1)	SS(1)	MFCC(36)	HR(1)	FF(1)	SE(1)	SR(1)
81:	ZCR(1)	SC(1)	SS(1)	MFCC(36)	HR(1)	FF(1)	SE(1)	SR(1)

图7 融合特征图像结构图

将融合特征图像输入融合特征编码器,获得融合特征嵌入向量。本文采用添加了并行卷积注意力模块的ResNet网络^[22]作为融合特征编码器。

2.3 文本模态

本文采用含有类别信息的描述性文本作为文本模态的输入,基本格式如下:“The sound belongs to {BCCCT/RCT}”。对描述性文本进行字节对编码后,输入文本编码器获得文本模态嵌入向量。文本编码器采用基于Transformer架构的模型^[13],其核心包含12层编码器层、8个注意力头以及512维的隐藏层。

2.4 对比学习

获得3个编码模态的嵌入向量后,采用对比学习方法,利用3个模态的信息共同优化编码器。优化目标是增大同一样本不同模态嵌入向量间的相似性,减小不同样本嵌入向量间的相似性,具体实现如算法1所示。

算法1 MCL算法

输入: 一个训练批次中的融合特征模态、时频模态、文本模态嵌入向量,分别表示为 $\mathbf{F}(S_b \times 1024)$ 、 $\mathbf{S}(S_b \times 1024)$ 、 $\mathbf{T}(S_b \times 1024)$,其中括号内为维度, S_b 表示批大小。

1. 对所有嵌入向量进行归一化,如 $\hat{\mathbf{F}}(i,:) = \mathbf{F}(i,:)/\|\mathbf{F}(i,:)\|_2$,获得 $\hat{\mathbf{F}}(S_b \times 1024)$ 、 $\hat{\mathbf{S}}(S_b \times 1024)$ 、 $\hat{\mathbf{T}}(S_b \times 1024)$;

2. 在3个模态之间两两计算相似度矩阵,如 $I_{FS}^{Sim} = \hat{F} \times \hat{S}^T \times FS$,其中FS(1)为可学习的放缩系数,获得 $I_{FS}^{Sim}(S_b \times S_b)$ 、 $I_{FT}^{Sim}(S_b \times S_b)$ 、 $I_{ST}^{Sim}(S_b \times S_b)$ 、 $I_{FS}^{Sim}(S_b \times S_b)$ 、 $I_{TF}^{Sim}(S_b \times S_b)$ 、 $I_{TS}^{Sim}(S_b \times S_b)$ 共6个相似度矩阵;

3. 根据相似度矩阵计算多模态交叉熵损失 $L = (L_{FS} + L_{FT} + L_{ST} + L_{SF} + L_{TF} + L_{TS})/6$,其中

$$L_{FS} = \left(-I_{FS}^{Sim}(i,i) + \ln \sum_{j=1}^{bs} \exp(I_{FS}^{Sim}(i,j)) \right);$$

4. 根据交叉熵损失优化模型参数。

3 仿真与湖试实验

为全面评估所提方法在短样本和复杂水声多径信道条件下的性能,本文以TDS和S-CNN方法为基线,设计并开展仿真与湖试实验,综合论证所提方法在短样本和水声多径信道条件下识别性能的有效性和稳健性。MCL算法基于PyTorch1.13框架与Python3.8环境实现。模型训练在一张NVIDIA Quadro RTX P5000显卡上进行,批大小设置为16,训练共计50轮次。优化器选用Adam,其初始学习率为 1×10^{-5} ,一阶和二阶矩估计指数衰减率分别设置为0.9和0.999,权重衰减系数设置为 1×10^{-5} 。

3.1 数据集构建

本文采用的RCT源自沃特金斯海洋哺乳动物声音数据库^[9]。为模拟实际海洋信道环境,基于Bellhop声场传播模型构建包含浅海与深海两类典型场景的10种水声信道,仿真信道参数见表1。

表1 仿真信道参数

单位:km

信道	水深	声源深度	接收机深度	水平距离
h_1	2	0.20	0.50	10
h_2	2	0.50	0.50	8
h_3	2	0.50	0.70	9
h_4	0.2	0.07	0.15	2
h_5	0.2	0.07	0.07	2
h_6	0.2	0.07	0.07	4
h_7	2	1.20	1.20	10
h_8	0.2	0.10	0.10	2
h_9	0.2	0.10	0.10	10
h_{10}	2	1.20	0.50	2

其声速剖面分布如图8所示。各信道的冲激响应均为具有不同幅度与时延的多径结构,具体响应可通过Bellhop工具及表1所列参数计算得到。其中: $h_1 \sim h_6$ 用于构造增强训练集; $h_7 \sim h_{10}$ 用于测试。

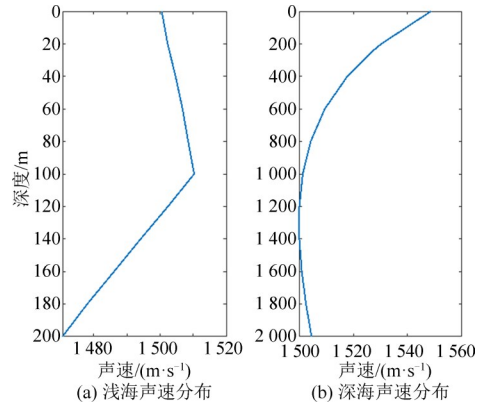


图8 声速剖面图

在实验中所用训练集组成如表2所示,其中信噪比表示信号与添加的高斯白噪声的功率之比,当信噪比为“ ∞ ”时表示不添加噪声,增加信道 h_0 ,其信道系统函数为单位冲激函数,表示无多径的信道。通过对原始嘀嗒声数据集进行信道增强^[23],生成包含约10 h音频的训练数据集,较原始数据量提升56倍。信号采样率统一设为48 kHz,样本长度设为0.7 s。

表2 训练集组成

信号	编码位数/bit	信噪比/dB	信道	总时长/s
BCCCT	1,2,3,4,5,6	-10, -5, 0, 5, 10, 15, 20, 25	$h_0, h_1, h_2, h_3, h_4, h_5, h_6$	19 152
RCT	1,2,3,4,5,6	-10, -5, 0, 5, 10, 15, 20, 25, ∞	$h_0, h_1, h_2, h_3, h_4, h_5, h_6$	17 512

3.2 仿真实验

为验证所提方法在短样本和复杂水声多径信道条件下的有效性和稳健性,本文分别在 h_7 、 h_8 、 h_9 、 h_{10} 信道条件下构造测试数据集(h_7 和 h_{10} 为深海信道, h_8 和 h_9 为浅海信道),信号样本长度固定为0.7 s,信噪比范围为-15~30 dB。不同信道条件下的识别准确率如图9~图12所示。相较对比方法,基于MCL的本文方法在浅海和深海信道条件下识别准确率明显提升,说明该方法在短样本长度和复杂水声多径信道条件下具有更强的有效性和稳健性。

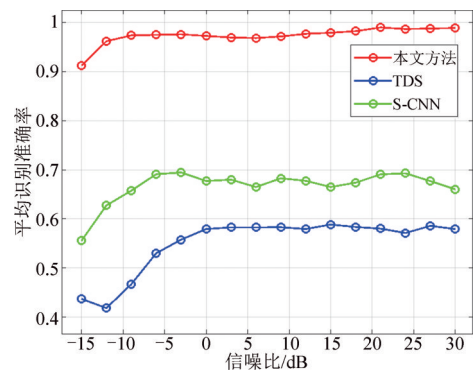
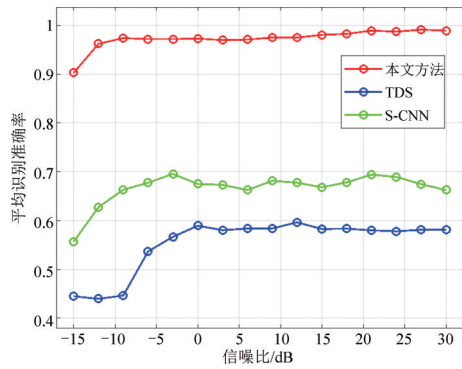
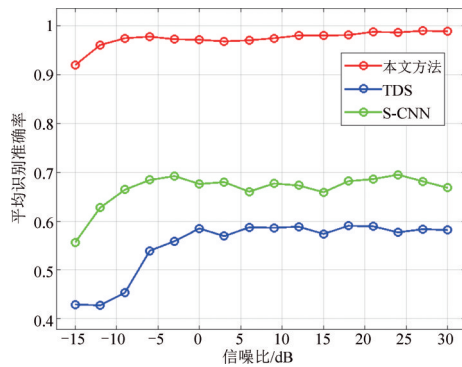
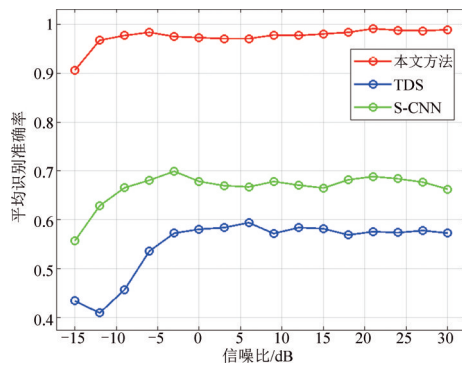


图9 信道 h_7 条件下的识别准确率

图10 信道 h_8 条件下的识别准确率图11 信道 h_9 条件下的识别准确率图12 信道 h_{10} 条件下的识别准确率

为验证MCL框架中各组件的有效性,设计消融实验。以完整的MCL模型(记为 M_0)作为基准,通过系统性地移除特定模态,构建了两个对比模型。 M_1 模型移除融合特征模态,仅使用时频与文本模态; M_2 模型移除时频模态,仅使用融合特征与文本模态。两对比模型均在与 M_0 相同的训练集上,采用完全一致的超参数进行训练,以确保性能差异仅源于模型结构的变化。测试在仿真信道 $h_7 \sim h_{10}$ 、信噪比 $-15 \sim 30$ dB下进行,识别准确率取平均值,结果如表3。由表3可知,完整模型 M_0 取得最优性能,表明3模态协同工作的有效性。模型 M_1 的识别准确率较 M_0 下降2.42个百分点,表明融合特征模态提供了有价值的补充信息,能够进一步提升模型性能。模型 M_2 的性能下降明显,准确率

较 M_0 降低40.60个百分点,充分证明时频模态是模型中最关键的特征表示来源,其提供的时频结构信息对区分BCCCT和RCT具有决定性作用。综上,消融实验结果表明,时频模态在识别任务中贡献最大,是模型性能的主要支撑;融合特征模态作为有效的补充,能进一步提升识别精度;而多模态特征的融合策略对于实现高性能识别至关重要。

表3 消融实验结果

模型名称	模型配置	平均准确率/%
M_0	时频+融合特征+文本(完整模型)	97.40
M_1	时频+文本(移除融合特征模态)	94.98
M_2	融合特征+文本(移除时频模态)	56.80

3.3 湖试实验

3.3.1 实验条件

为了验证本文所提方法在实际环境中的有效性,在丹江口水库开展湖试实验,如图13所示。实验区域声速剖面图如图14所示。

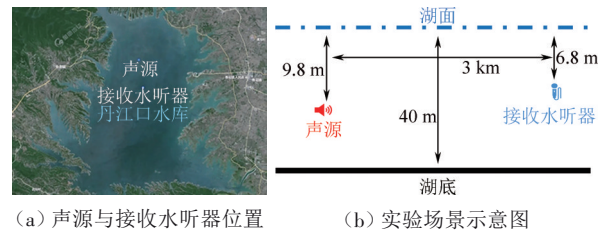


图13 丹江口水库实验场景

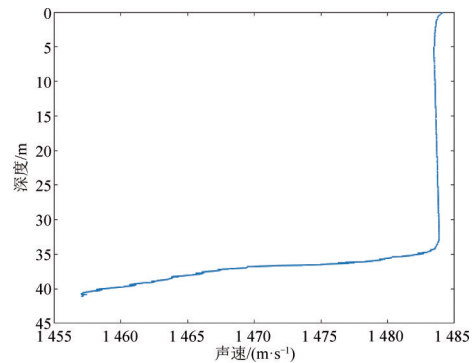


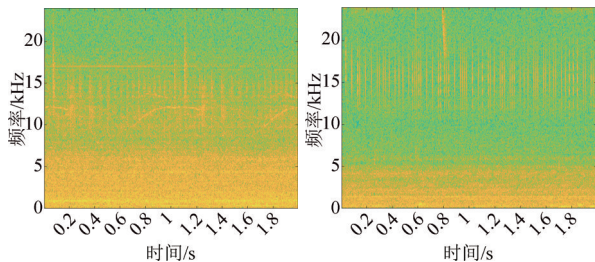
图14 实验水域声速剖面图

实验发送信号包含BCCCT和RCT。BCCCT的码元为从沃特金斯海洋哺乳动物声音数据库^[9]中截取并处理后的真实海豚嘀嗒声脉冲。信息调制方式采用时延差编码,基础编码时间设置为1 ms,编码位数1~5 bit,单位编码时间为15、55、95 ms。实验共发送了总时长为435 s的BCCCT信号,包含了不同编码参数的信号序列,以充分测试识别性能。RCT同样源自沃特金斯数据库,选取了多种鲸目动物在自然状态下发出的较纯净的嘀嗒声,总

时长约为 140 s。

3.3.2 实验结果分析

接收信号时频图示例如图 15 所示。接收信号多径效应不明显,在 5 kHz 以下存在低频干扰噪声。



(a) 接收 RCT 信号时频图 (b) 接收 BCCCT 信号时频图

图 15 接收信号时频图

该湖试场景主要用于验证方法在短样本和实际背景噪声下的性能,而方法的多径鲁棒性已在前文的仿真实验中得以充分验证。为消除 BCCCT 与 RCT 样本数量不均衡对评估效果的影响,本文分别计算 BCCCT 与 RCT 的识别准确率后,取平均值作为最终准确率。最终,所提方法在湖试中取得 88.70% 的识别准确率,显著优于 TDS(80.19%)和 S-CNN(50.23%)方法,性能提升主要源于 3 方面改进。首先,本文方法通过时频模态和融合特征模态,直接从信号细微结构中提取本质特征,避免了传统方法对时延差统计特征的依赖,从而有效克服了短样本条件下统计特征不稳定的问题。其次,对比学习机制通过拉近同类样本、推远异类样本,使模型能够聚焦于信号的本质特性,有效抑制了多径效应和噪声带来的干扰。最后,文本模态提供的语义监督引导模型关注与类别相关的判别性特征,进一步提升了特征学习的有效性。

综上,所提方法通过多模态特征提取、对比学习和语义监督的协同作用,在短样本和实际信道条件下实现了显著的性能提升。

4 结束语

针对短样本与复杂水声多径信道环境下仿啁啾声通信信号识别难题,提出一种融合多模态对比学习的识别方法。通过构建时频、融合特征和文本 3 模态表示体系,结合对比学习机制,有效提升信号识别的准确率与鲁棒性。仿真与湖试实验验证该方法在复杂水声多径信道和短样本条件下的优越性能,为水下隐蔽通信信号识别提供新的技术途径。未来将进一步探索多模态融合与自监督学习在水声信号处理中的深度融合、应用扩展。

参考文献:

- [1] 王梦佳. 海豚哨声分类识别与通信技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2019: 1-87.
- [2] 贾於清. 仿鲸目哨声隐蔽水声通信方法研究[D]. 杭州: 浙江大学, 2021: 1-96.
- [3] 申家琪. 仿鲸目动物啁啾声隐蔽水声通信技术研究[D]. 海口: 海南大学, 2022: 1-69.
- [4] JIANG J J, QIAO F, LI Y, et al. Recognition method for the bionic camouflage click communication trains modulated by time delay difference [J]. The Journal of the Acoustical Society of America, 2022, 152(1): No.491.
- [5] 桂阳. 非合作仿生水声通信信号检测及识别方法研究[D]. 南京: 东南大学, 2023: 1-89.
- [6] YAO Q W, JIANG J J, YU X L, et al. Recognition method for biomimetic camouflage communication signal imitating cetacean click in underwater multipath channels [J]. Applied Acoustics, 2024, 222: No.110039.
- [7] GUZHOV A, RAUE F, HEES J, et al. Audioclip: extending clip to image, text and audio [C]. //Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, USA: IEEE, 2022: 976-980.
- [8] JIANG J J, LI S Y, WANG X Q, et al. Study of the relationship between sound signals and behaviors of a sperm whale during the hunting process [J]. Applied Acoustics, 2021, 174: No.107745.
- [9] WATKINS W A. Watkins marine mammal sound database [EB/OL]. (2020-03-09) [2025-10-16]. <https://cis.whoi.edu/science/B/whalesounds/index.cfm>.
- [10] WATKINS W A, FRISTRUP K, DAHER M A, et al. Sound database of marine animal vocalizations structure and operations [J]. Woods Hole Oceanographic Institution, 1992. DOI:10.1575/1912/854.
- [11] 刘淞佐, 刘冰洁, 尹艳玲, 等. M 元仿海豚叫声隐蔽水声通信 [J]. 哈尔滨工程大学学报, 2014, 35(1): 119-125.
- [12] LIU S, QIAO G, ISMAIL A. Covert underwater acoustic communication using dolphin sounds [J]. The Journal of the Acoustical Society of America, 2013, 133(4): 300-306.
- [13] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision [DB/OL]. (2021-02-26) [2025-10-16]. <https://arxiv.org/abs/2103.00020>.
- [14] KEDEM B. Spectral analysis and discrimination by zero-crossings [J]. Proceedings of the IEEE, 1986, 74(11): 1477-1493.
- [15] FLORES-FUENTES W, RIVAS-LOPEZ M, SERGIYENKO O, et al. Combined application of power spectrum centroid and support vector machines for measurement improvement in optical scanning systems [J]. Signal Processing, 2014, 98: 37-51.

(下转第 34 页)