

基于多分主干外部注意力网络的水声信号识别^①



王 越¹, 李佐勇², 颜佳泉², 胡 蓉¹

¹(福建理工大学 计算机科学与数学学院 福建省大数据挖掘与应用技术重点实验室, 福州 350118)

²(闽江学院 计算机与大数据学院 福建省信息处理与智能控制重点实验室, 福州 350121)

通信作者: 颜佳泉, E-mail: yanjiaquan@mju.edu.cn; 胡 蓉, E-mail: 896410521@qq.com

摘 要: 水声信号识别近年来备受关注, 由于海洋信道具有时变空变性、信号传播的衰落特性和水下目标声源具有复杂多变性, 水声信号识别任务面临巨大挑战. 传统的水声信号识别方法难以充分获取目标的表征信息且不具备良好的抗噪声能力, 识别效果有待提升. 针对上述问题, 本文提出一种基于多分支外部注意力网络 (multi-branch external attention network, MEANet) 的水声信号识别方法, 可以在复杂海洋环境下充分获取水声信号的特征并进行识别. MEANet 由多分支主干网络、通道、空间注意力模块和外部注意力模块组成. 首先, 输入数据通过多个并行的主干网络分支, 提取水声信号不同层级的特征信息; 其次, 辅以通道、空间注意力模块对水声信号的通道和空间维度分别进行加权, 调节不同通道和空间位置对特征表示的重要性; 最后, 整合外部注意力模块, 以外部记忆单元和附加计算来引导网络的特征提取和预测, 从而显著提高模型的识别率和鲁棒性. 实验结果表明, 本文提出的 MEANet 在 ShipsEar 数据集上的水声信号识别率达到 98.84%, 显著优于其他对比算法, 证实了其有效性.

关键词: 水声信号识别; 多分支主干; 注意力机制; 音频分类; 时频分析

引用格式: 王越, 李佐勇, 颜佳泉, 胡蓉. 基于多分主干外部注意力网络的水声信号识别. 计算机系统应用, 2024, 33(4): 263–270. <http://www.c-s-a.org.cn/1003-3254/9477.html>

Underwater Acoustic Signal Recognition Based on Multi-backbone External Attention Network

WANG Yue¹, LI Zuo-Yong², YAN Jia-Quan², HU Rong¹

¹(Fujian Provincial Key Laboratory of Big Data Mining and Applications, School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou 350118, China)

²(Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, School of Computer and Data Science, Minjiang University, Fuzhou 350121, China)

Abstract: In recent years, underwater acoustic target recognition has received considerable attention. However, due to the time-varying and space-varying nature of the underwater acoustic channel, as well as the complex and variable characteristics of the underwater target sound sources, water sound signal recognition tasks face significant challenges. Traditional methods for water sound signal recognition struggle to capture sufficient representation information of the targets and lack robustness against noise, resulting in suboptimal recognition performance. To address these issues, this study proposes a water sound signal recognition method based on the multi-branch external attention network (MEANet), which can effectively extract features and perform recognition in complex marine environments. MEANet consists of multiple branches for the backbone network, channel and spatial attention modules, and external attention modules. Firstly, the study feeds the input data through multiple parallel branches of the backbone network to extract features at different levels from the water sound signals. Secondly, it employs the channel and spatial attention modules to weight the

① 基金项目: 福建省自然科学基金青年创新项目 (2021J05202); 福建省发树慈善基金会资助研究专项 (MFK23006); 国家自然科学基金 (61972187); 福建省卫生健康重大科研专项 (2021ZD01004)

收稿时间: 2023-09-14; 修改时间: 2023-10-25, 2023-11-09; 采用时间: 2023-12-20; csa 在线出版时间: 2024-03-01

CNKI 网络首发时间: 2024-03-07

channels and spatial dimensions of the water sound signals. Finally, the external attention module integrates external memory units and additional computations to guide feature extraction and prediction, significantly improving the recognition rate and robustness of the model. Experimental results demonstrate that the proposed MEANet achieves a recognition rate of 98.84% on the ShipsEar dataset, outperforming other comparative algorithms.

Key words: underwater acoustic target recognition; multi-branch backbone; attention mechanism; audio classification; time-frequency analysis

1 引言

水声信号识别是指对水下传播的声音信号进行分析、处理和识别的过程。水声信号识别对于海洋开发、国防安全有着重大意义,是水声信号处理领域的重要研究方向之一。然而,由于海洋环境的复杂性、水下目标的多样性和目标数据的小样本特性,以及异类特征相似性和同类特征相异性,造成了水声信号识别系统的迁移能力和泛化能力较差。随着海洋的战略地位日益突显,各国都在积极开发利用海洋资源和空间,为了应对这些挑战,对水声信号识别研究提出了新要求。

在水声信号识别领域早期,传统水声信号识别基本是采用基于浅层学习的方法^[1],包括预处理、特征提取、特征选择和分类器设计4个环节。特征提取方法是水声信号识别任务的重中之重,其有效性直接影响水声信号的识别性能,因此,浅层学习无法充分获取目标的表征信息且不具备良好的抗噪能力。深度学习的出现给水声信号识别领域提供了一个新视角。邓晋等人^[2]基于迁移学习,利用预训练网络VGG和VGGish提取深层声学特征并进行模型微调,取得92.48%的识别率。Ke等人^[3]使用RSSD对船舶辐射噪声进行分类,从音频数据中提取高共振分量。然而,它需要大量的数据进行预训练,这在小数据集上表现不佳。Han等人^[4]利用联合一维卷积神经网络与长短期记忆网络两种网络串行的方式,首次将联合网络运用到水声信号识别中。随着深度学习的发展,信号识别的网络构架逐渐成熟,特征融合及注意力机制也被研究人员引入其中。Hong等人^[5]提出一种使用融合特征和18层残差网络(ResNet18)的新型分类方法。Yang等人^[6]提出了一种新的协同深度学习方法,这种新方法将深度长短期记忆网络(DLSTM)和深度自动编码器神经网络(DAE)结合在一起,并通过无监督学习以获取更好的结果。徐承等人^[7]基于Transformer的网络结构的ViT模块进

一步在时间维度上对不同帧特征进行相关性计算和深度特征融合,最终构建水声特征空间到类别空间的映射关系,完成目标识别任务。Xie等人^[8]提出了一种自适应广义识别系统——AGNet,采用具有并行卷积注意模块的卷积神经网络作为分类器,展示了在各种水下环境中的鲁棒性。Ma等人^[9]针对不平衡水声数据集,提出了一种基于三角函数的加权交叉熵损失函数,并将所提出的损失函数应用于嵌入了注意力机制的多尺度残差卷积神经网络。Feng等人^[10]提出了一种基于无卷积架构的UATR-Transformer,可以从声谱图中感知全局和局部信息,从而提高准确性。Liu等人^[11]提出了一种基于双注意力网络(DAN)和多分辨率卷积神经网络(DAN-MCNN)的新方法,并基于多类特征子集设计了三维聚合特征,提出的DAN模块由位置注意模块和通道注意模块组成,捕捉样本的全局依赖性和局部特征。

本文提出一种基于多分支外部注意力网络(multi-branch external attention network, MEANet)的水声信号识别方法。该网络采用多分支主干网络结构以更好地捕获输入数据的局部细节和全局特征,同时加速模型的收敛速度。辅以通道注意和空间注意力模块来降低海洋背景噪声和多目标噪声的特征权重,更好地捕捉输入数据中的空间结构和上下文信息。最后,该方法还在全连接层前添加外部注意力模块。该模块在复杂度低于自注意力的前提下,可以挖掘整个水声数据集的潜在关系,增强模型的感知能力和泛化能力,从而提高模型的识别效果。与此同时,将不同的数据增强方法分别应用到一维的时域、频域数据和二维的时频图上,实验结果表明,本文提出的MEANet能够有效地提高水声信号识别的准确率。

2 本文方法

如图1所示,本文方法主要包含数据预处理、数据增强、MEANet网络训练以及最后分类结果4个步

骤. 首先进行数据的预处理操作, 去除空白段并切片为时域信号、频域信号和时频域信号, 对它们进行相应

的数据增强后, 输入到 MEANet 网络中进行训练, 最终输出分类结果.

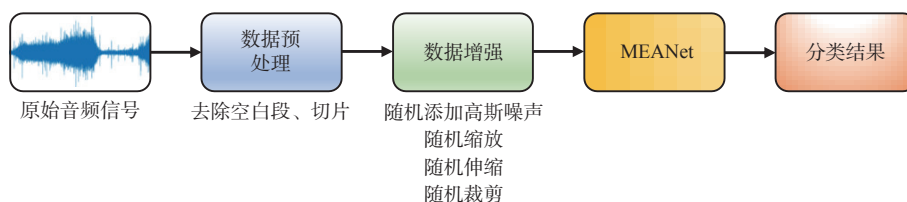


图1 本文方法概述

2.1 数据预处理

在数据预处理中, 音频文件的切分是通过 Matlab 软件将 ShipsEar 数据集原始的声信号 (.wav 格式), 裁剪掉其中的空白段后, 将其保存为时间序列的 .mat 文件. 在实验中, 每个样本长度为 2048, 样本总数可以从式 (1) 中获得. 生成样本后, 我们将总样本的 80% 作为训练集, 将总样本中的 20% 作为测试集.

$$N = \text{floor}\left(\frac{L}{2048}\right) \quad (1)$$

其中, L 是每个信号的长度, N 是总样本数, floor 是向负无穷方向舍入的意思. 图 2 为水声信号时域图.

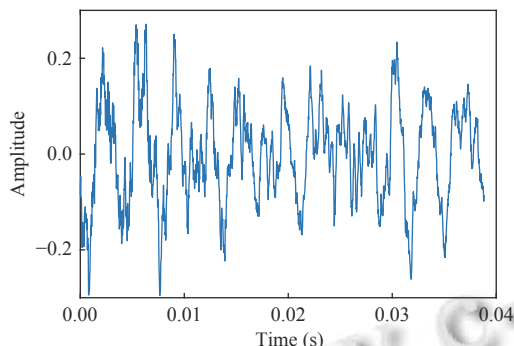


图2 水声信号时域图

对于频域输入, 使用 FFT 函数将每个样本 x_i 从时域转换到式 (2) 所示的频域. 在此操作后, 数据长度将减半, 新样本可以表示为:

$$x_i^{FFT} = FFT(x_i) \quad (2)$$

其中, 算子 $FFT(\cdot)$ 表示将 x_i 变换到频域, 并取结果的前半部分.

在频域部分, 另外讨论了小波变换的处理方法, 小波变换通常产生两种类型的系数: 近似系数和细节系数. 近似系数表示信号中的低频成分或整体特征. 它们对信号进行粗略的平滑和降采样, 以保留信号的主要

信息. 细节系数表示信号中的高频成分或细节信息. 在本文中我们选取信号的近似系数来进行实验, 选定 Hear 小波函数, 变换层数设置为 1.

$$x_i^{WT} = WT(x_i) \quad (3)$$

其中, 算子 $WT(\cdot)$ 表示将初始信号 x_i 通过 Hear 小波函数变换为近似系数. 图 3 为水声信号频域图.

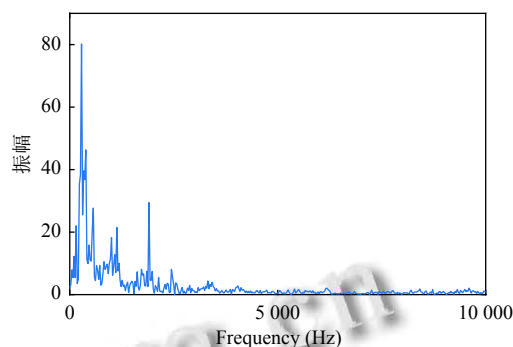


图3 水声信号频域图 (FFT)

对于时频域输入, 对每个样本 x_i 应用短时傅立叶变换 ($STFT$), 以获得式 (3) 所示的时频表示. 为了保证相邻文件的相关性, 我们采用汉宁窗 (Hanning window) 函数减少频谱泄露效应, 根据选取的水声信号的特点, 窗口长度设置为 64, 能够兼顾时间分辨率和频率分辨率, 此操作后, 将生成时频表示 (一个 33×33 图像), 如下所示:

$$x_i^{STFT} = STFT(x_i), i = 1, 2, \dots, N \quad (4)$$

其中, 算子 $STFT(\cdot)$ 表示将 x_i 变换到时频域. 图 4 为水声信号时频图.

如果二维卷积神经网络输入一个很小的二维信号, 会导致模型, 无法提取适当的特征, 因此, 我们将 $STFT$ 的信号大小调整为 330×330 . 时域和频域输入保持不变.

2.2 数据增强

数据增强对于使训练数据集更加多样化,缓解小样本问题带来的学习困难非常重要.在本文中,我们分别对时域、频域和时频域的数据做了实验,为了针对不同维度的输入数据进行数据增强,所以本文分别使用了一维和二维的数据增强方法来进行不同的实验.

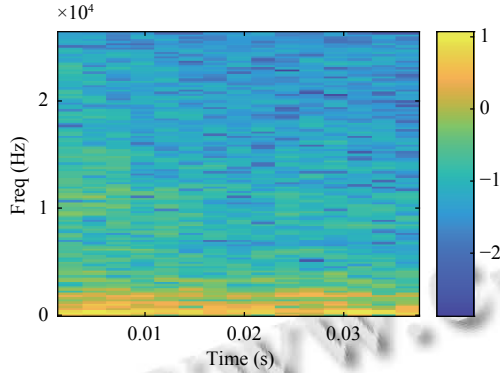


图4 水声信号时频图

2.2.1 一维输入数据增强

RandomAddGaussian: 该策略将高斯噪声随机添加到输入信号中,公式如下:

$$x := x + n \quad (5)$$

其中, x 是一维输入信号, n 由高斯分布 $N(0, 0.01)$ 生成.

RandomScale: 该策略将输入信号与随机因子随机相乘,其公式如下:

$$x := \beta \times x \quad (6)$$

其中, x 是一维输入信号, β 是高斯分布 $N(0, 0.01)$ 之后

的定标器.

RandomStretch: 该策略将信号重采样成随机比例,并通过置零和截断来确保长度相等.

RandomCrop: 该策略随机地覆盖部分信号,其公式如下:

$$x := \text{mask} \times x \quad (7)$$

其中, x 是一维输入信号, mask 是随机位置的子序列为零的二进制序列.在本文中,子序列的长度为 10.

2.2.2 二维输入数据增强

RandomScale: 该策略将输入信号与随机因子随机相乘,其公式如下:

$$x := \beta \times x \quad (8)$$

其中, x 是二维输入信号, β 是高斯分布 $N(0, 0.01)$ 之后的定标器.

RandomCrop: 该策略随机地覆盖部分信号,其公式如下:

$$x := \text{mask} \times x \quad (9)$$

其中, x 是二维输入信号, mask 是随机位置的子序列为零的二进制序列.在本文中,子序列的长度为 20.

2.3 模型架构

为了在复杂海洋环境下充分获取水声信号的特征并识别,本文设计了一个多分支网络来提取水声信号信息,它主要由基于多分支结构的主干网络和注意力模块组成.

2.3.1 多分支主干网络

MEANet 模型结构图如图 5 所示.

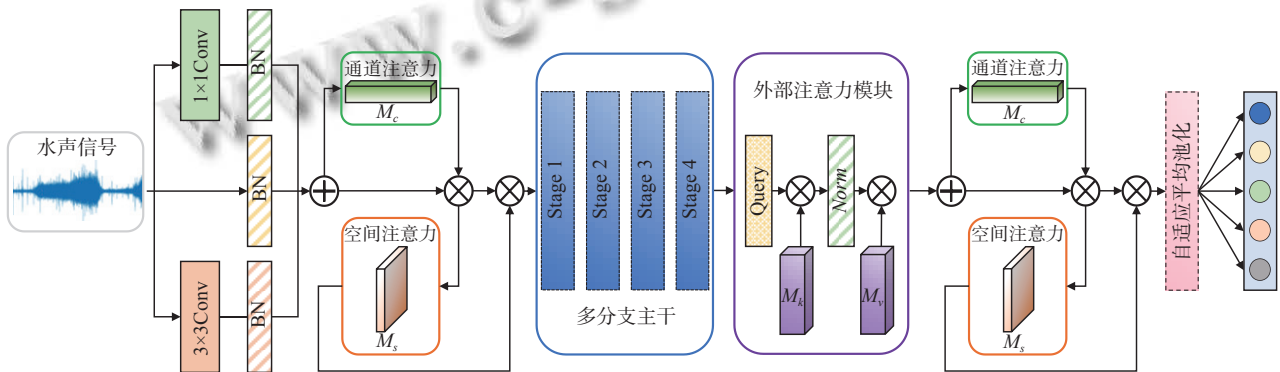


图5 MEANet 模型结构图

多分支主干由卷积分支、恒等连接和批量归一化层组成.卷积分支的卷积核大小分别为 1×1 和 3×3 ,在

每个分支中,卷积层之后是批量归一化层.其中 3×3 卷积层可以用于提取更细节的特征, 1×1 卷积层可以用于

降低特征的维度并提高计算效率. 这样一来, 每个分支就可以提取一些独特的特征, 然后将这些特征进行融合和整合, 得到更加丰富和完整的特征表示, 从而提高网络的性能并提高模块化和可扩展性. 与此同时, 主干网络汇总的结果可以流入注意力机制网络, 另一方面, 它跳过注意力模块进入下一步操作, 多分支主干的结构可以堆叠多次, 如图 5 中的 Stage 1–Stage 4 分别叠加了 4、6、16、1 层.

2.3.2 注意力机制模块

注意力机制是指模拟人类的视觉系统, 选择性地聚焦于一部分关键信息, 忽略其他冗余信息, 已逐渐成为改进深度神经网络的一个发展方向. 迄今为止, 注意力模块已被证明有助于提高多种任务准确性^[12–15]. 本文将注意力模块耦合进主干网络中. 它包括通道注意力模块 (CAM)、空间注意力模块 (SAM) 和外部注意力模块 (EAM)^[16].

本文中, 通道注意力和空间注意力模块作串行使用, 将上一层输出的特征作为通道注意力模块的输入, 经过通道注意力模块后得到的权值, 作为空间注意力模块的输入, 得到新的特征值继续参与模型训练.

- 通道注意力模块: 在卷积神经网络中, 通道数是一个很关键的因素, 模型通常会依赖大量的通道特征进行学习. 然而, 在实际应用中, 通道中的一部分特征可能是冗余的噪音或者其他无用信息. 通道注意力模块通过自适应地学习每个通道的权重, 对输入特征的通道进行加权, 使得网络更加关注那些重要的特征通道, 从而提高了网络的表现. 具体来说, 将上一层网络的输出作为通道注意力模块的输入 $H \times W \times C$, 经过平均池化层 (AvgPool) 和最大池化层 (MaxPool) 得到新的特征, 如式 (9) 所示, 输出原始船舶信号中有意义的特征:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (10)$$

其中, $M_c(F)$ 的大小是 F 的通道数, 即 C ; F 特征的大小为 $C \times H \times W$; $\sigma(\cdot)$ 表示 Sigmoid 激活函数; W_0 为 $\frac{C}{r} \times C$ 的矩阵; W_1 为 $C \times \frac{C}{r}$ 的矩阵.

- 空间注意力模块: 该模块关注原始特征图上不同空间位置的重要性. 它基于注意力机制, 自适应地学习每个空间位置的权重, 以增强那些重要的位置的特征信息. 该模块主要是获取通道注意力模块特征作为

输入, 而后在平均池化层 (AvgPool) 和最大池化层 (MaxPool) 中选择合适的池化层进行池化后再进行卷积, 得到的值能说明在原始船舶信号中哪些区域特征是值得关注的, 如式 (10) 所示:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (11)$$

其中, $\sigma(\cdot)$ 表示 Sigmoid 激活函数, $f^{7 \times 7}(\cdot)$ 表示卷积操作且 7×7 是卷积核大小.

- 外部注意力模块: 与通道注意力和空间注意力不同, 外部注意力不是在特征图内部生成, 而是依赖于外部信息和附加计算来引导网络的特征提取和预测. 外部注意力模块可以自适应学习权重, 用于对输入水声信号特征进行加权, 帮助模型确定哪些特征对分类任务最有用, 从而增强了模型对关键特征的关注. 对于不同频率或时域特性, 模块可以根据其重要性自动分配不同的权重. 此外, 外部注意力模块可以帮助模型抑制噪声或将关注点放在与噪声无关的信号部分上, 协助模型整合包括低频和高频成分的多尺度信息. 具体来说, 外部注意力由两个部分组成: 一个或多个可学习的查询向量和与输入数据进行比对的外部记忆单元 $M \in \mathbb{R}^{S \times d}$. 首先, 通过计算查询向量和外部记忆单元中每个记忆向量间的关系, 生成注意力权重, 接着, 将注意力权重利用加权平均或加权池化的方式应用于输入数据的不同维度, 以生成新的调节后的特征表征, 计算公式如下:

$$A = (\alpha)_{i,j} = Norm(FM^T) \quad (12)$$

$$F_{out} = AM, M \in \mathbb{R}^{S \times d} \quad (13)$$

其中, $a_{i,j}$ 是 M 的第 i 个像素和第 j 行之间的相似性, M 是独立于输入的可学习参数, 其充当整个训练数据集的存储器. A 是从中学习到的数据集级的先验知识推出的注意力图. 最后, 通过 A 中的相似性来更新 M 的输入特征.

在实践中, 使用两个不同的存储单元 M_k 和 M_v 作为 key 和 value, 以提高网络的性能. 此时, 外部注意力的计算公式为:

$$A = Norm(FM_k^T) \quad (14)$$

$$F_{out} = AM_v \quad (15)$$

与余弦相似度不同, 由于注意力图是通过矩阵乘

法计算的,注意力图对输入特征的尺度比较敏感.而在自注意力^[17]中采用 Softmax 对注意力图进行双重归一化,使得 $\sum_j \alpha_{i,j} = 1$. 为了避免这个问题,我们选择文献[18]中提出的双归一化,它对行和列进行归一化.这种双重归一化表示为:

$$(\tilde{\alpha})_{i,j} = FM_k^T \quad (16)$$

$$\hat{\alpha}_{i,j} = \frac{\exp(\tilde{\alpha}_{i,j})}{\sum_k \exp(\tilde{\alpha}_{k,j})} \quad (17)$$

$$\alpha_{i,j} = \frac{\hat{\alpha}_{i,j}}{\sum_k (\hat{\alpha}_{i,k})} \quad (18)$$

3 实验及结论分析

3.1 数据集设置

水声信号的所有数据均来自 ShipsEar 数据集^[19]. 该数据集给出了 2012 年和 2013 年期间,研究人员在西班牙大西洋沿岸录制的许多不同类别船只的声音. 总共包含 11 种船舶类型和 1 个背景噪声共 91 条数据. 每个水声信号数据的长度在 15 s–10 min 之间,采样频率为 52 734 Hz. 去除掉空白段和干扰过大部分,使用 Matlab 将音频信号转换为时域信号后,按照固定的 2 048 采样点对所有的声音片段进行分割,共得到 93 409 个标记的声音样本. 不同类型船舶辐射噪声的详细划分如表 1 所示.

表 1 数据集分类

Class	Type
ClassA	Background noise
ClassB	Fishboats, trawlers, mussel boats, tugboats, dredger
ClassC	Motorboats, pilot boats, sailboats
ClassD	Passengers
ClassE	Ocean liners, RO-RO

许多研究人员使用信号处理方法将时间序列映射到不同的域以提高性能. 然而,哪种输入类型更适合水声目标识别仍是一个悬而未决的问题. 本文讨论了不同输入类型对模型性能的影响. 当输入为时域数据和频域数据时,本文将模型改写,把二维卷积层、池化层和归一化层等改为一维版本,以适应输入的变化,实验结果表明,本文方法在输入为一维数据时仍然有效.

3.2 输入格式不同的比较

我们首先比较分析了时域、频域和时频域作为模

型输入时的不同性能,其中,频域输入时采用快速傅里叶变换和小波变换进行实验. 将 5 大类共 93 409 条数据的 80% 划分为训练集,20% 为测试集,训练设置如下: 一维信号 batch_size 大小为 64,二维信号 batch_size 大小为 32,优化器选择 adamW,初始学习率为 0.000 1,迭代次数为 100. 选取 CNN、BiLSTM 和 ResNet 进行对比,结果如表 2 所示.

表 2 不同输入格式的识别结果

方法	输入类型	准确率 (%)
CNN	时域	92.28
	频域 (FFT/WT)	86.45/83.34
	时频域	91.54
BiLSTM	时域	95.27
	频域 (FFT/WT)	91.68/84.19
	时频域	93.26
ResNet18	时域	96.50
	频域 (FFT/WT)	92.14/90.95
	时频域	97.76
MEANet	时域	97.70
	频域 (FFT/WT)	92.85/91.03
	时频域	98.84

从表 2 可以看出,当分类方法为 CNN 和 BiLSTM 时,时域信号作为输入可以取得最好的分类效果,分别为 92.28% 和 95.27%. 同时,采用 ResNet 和 MEANet 作为分类方法时,时频域输入 (STFT) 可以取得最好的结果,分别为 97.76% 和 98.84%. 这意味着对于 CNN 和 BiLSTM 这种较浅的模型来说,更擅长处理一维时域信号,对于更复杂的图像它们的效果不佳. 而更深的 ResNet 模型和 MEANet 能够从包含信息更多的二维图像中提取到更有用的信息,并取得更高的识别率.

3.3 有数据增强与无数据增强的比较

从表 3–表 6 可以看出,采用数据增强后,在时域、频域 (WT) 和时频域输入上均优于未采用数据增强的效果,而在频域上的结果有所偏差. 相较于快速傅里叶变换 (FFT),采用小波变换 (WT) 方法的识别性能稍逊,在无数据增强方法的情况下性能下降得较为明显,采用数据增强方法后,识别性能与 (FFT) 相当. 以频域 (FFT) 作为输入格式时,采用 CNN 模型在有数据增强时效果更优,而在 BiLSTM、ResNet 和 MEANet 上则不如未采用数据增强的结果,其原因可能是频域信号在快速傅里叶变换后数据段太短,只保留一半长度,所具有的特征不够显著,在加噪、乘随机因子和覆盖变换后反而会对模型产生误导,造成识别率下降.

表3 时域信号数据增强对比 (%)

方法	无数据增强准确率	有数据增强准确率
CNN_1D	87.07	92.28
BiLSTM	94.13	95.27
ResNet18	93.03	96.50
MEANet	95.84	97.70

表4 频域信号数据增强对比 (FFT)

方法	无数据增强准确率	有数据增强准确率
CNN_1D	84.01	86.45
BiLSTM	93.30	91.68
ResNet18	91.87	92.14
MEANet	93.76	92.85

表5 频域信号数据增强对比 (WT)(%)

方法	无数据增强准确率	有数据增强准确率
CNN_1D	76.19	83.34
BiLSTM	82.19	84.19
ResNet18	74.38	90.95
MEANet	82.48	91.03

表6 时频域信号数据增强对比 (%)

方法	无数据增强准确率	有数据增强准确率
CNN_2D	91.08	91.54
BiLSTM	92.88	93.26
ResNet50	97.41	97.76
MEANet	98.52	98.84

3.4 模型之间的比较

对比模型间的性能时,在输入类型为时频图时,我们额外引入 ResNet50 进行对比,从表7、表8和图6中可以看出,本文提出的模型在3种输入类型以及是否做数据增强上均取得最好结果,在输入类型为频域时,短时傅里叶变换方法优于小波变换方法.在以 $STFT$ 作为输入,并采用数据增强时达到了 98.84% 的识别率,充分说明了本文提出的 MEANet 性能更优,且数据增强在大部分情况下的有效性.

表7 无数据增强的识别结果 (%)

输入类型	模型	准确率
时域	CNN_1D	87.07
	BiLSTM_1D	94.13
	ResNet18_1D	93.03
	MEANet	95.84
频域 (FFT/WT)	CNN_1D	84.01/76.19
	BiLSTM_1D	93.30/82.19
	ResNet18_1D	91.87/74.38
	MEANet	93.76/82.48
时频域	CNN_2D	91.08
	BiLSTM_2D	92.88
	ResNet18_2D	97.81
	ResNet50_2D	97.41
	MEANet	98.52

表8 有数据增强的识别结果 (%)

输入类型	模型	准确率
时域	CNN_1D	92.28
	BiLSTM_1D	95.27
	ResNet18_1D	96.50
	MEANet	97.70
频域 (FFT/WT)	CNN_1D	86.45/83.34
	BiLSTM_1D	91.68/84.19
	ResNet18_1D	92.14/90.95
	MEANet	92.85/91.03
时频域	CNN_2D	91.54
	BiLSTM_2D	93.26
	ResNet18_2D	97.78
	ResNet50_2D	97.76
	MEANet	98.84

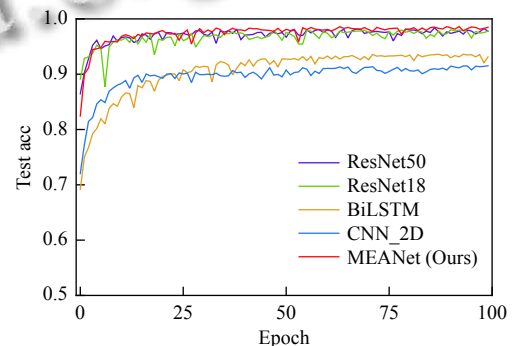


图6 不同模型间识别率对比

4 结论与展望

本文提出一种基于多分支外部注意力网络 (MEANet) 的水声信号识别算法,该方法采用多分支最优架构设计,充分提取不同样本特征,引入残差结构以充分利用时频特征,避免梯度消失或梯度爆炸.本文还引入通道、空间注意力模块,增强不同通道特征差异性,从而抑制海洋背景噪声的干扰,提高模型的适用性.为了提高分类的识别率,本文另外引入一种新颖的、轻量级但有效的外部注意力模块,采用两个外部记忆力单元,能够从输入学习更多的代表性特征,增强模型的感知能力和泛化能力,相比于自注意力模块,还可以降低计算成本.本文同时探究了输入格式以及数据增强对于水声信号识别的影响,对比了具有代表性的时域、频域和时频域这3种输入格式的识别结果,以及一些有效的数据增强方法.实验结果表明,在嘈杂水声背景环境下,MEANet 可以有效地检测并提高水声信号的识别率,相比于时域和频域,时频域信号作为输入的效果最好,数据增强方法在时域和时频域数据上可以起到

提高识别率的作用。

尽管本文提出的方法可以有效提升实验效果,但仍有以下问题有待进一步深入研究: 1) 本文仅使用一种数据集,在其他数据集上的性能有待验证; 2) 时频分析方法还有很多种,如 MFCC、Mel 谱图和 CQT 等,接下来的工作可以对比研究不同时频分析方法和特征融合对识别率的影响。

参考文献

- 1 Berger CR, Zhou SL, Preisig JC, *et al.* Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing. *IEEE Transactions on Signal Processing*, 2010, 58(3): 1708–1721. [doi: [10.1109/TSP.2009.2038424](https://doi.org/10.1109/TSP.2009.2038424)]
- 2 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)]
- 3 Ke XQ, Yuan F, Cheng E. Underwater acoustic target recognition based on supervised feature-separation algorithm. *Sensors*, 2018, 18(12): 4318. [doi: [10.3390/s18124318](https://doi.org/10.3390/s18124318)]
- 4 Han XC, Ren CX, Wang LM, *et al.* Underwater acoustic target recognition method based on a joint neural network. *PLoS One*, 2022, 17(4): e0266425. [doi: [10.1371/journal.pone.0266425](https://doi.org/10.1371/journal.pone.0266425)]
- 5 Hong F, Liu CW, Guo LJ, *et al.* Underwater acoustic target recognition with ResNet18 on ShipsEar dataset. *Proceedings of the 4th IEEE International Conference on Electronics Technology*. Chengdu: IEEE, 2021. 1240–1244.
- 6 Yang HH, Xu GH, Yi SZ, *et al.* A new cooperative deep learning method for underwater acoustic target recognition. *Proceedings of the 2019 OCEANS*. Marseille: IEEE, 2019. 1–4.
- 7 徐承, 李勇, 张梦, 等. 基于特征融合和自注意力机制的水下目标识别. *移动通信*, 2022, 46(6): 91–98. [doi: [10.3969/j.issn.1006-1010.2022.06.015](https://doi.org/10.3969/j.issn.1006-1010.2022.06.015)]
- 8 Xie Y, Ren JW, Xu J. Adaptive ship-radiated noise recognition with learnable fine-grained wavelet transform. *Ocean Engineering*, 2022, 265: 112626. [doi: [10.1016/j.oceaneng.2022.112626](https://doi.org/10.1016/j.oceaneng.2022.112626)]
- 9 Ma YX, Liu MQ, Zhang Y, *et al.* Imbalanced underwater acoustic target recognition with trigonometric loss and attention mechanism convolutional network. *Remote Sensing*, 2022, 14(16): 4103. [doi: [10.3390/rs14164103](https://doi.org/10.3390/rs14164103)]
- 10 Feng S, Zhu XQ. A Transformer-based deep learning network for underwater acoustic target recognition. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 1505805.
- 11 Liu CW, Hong F, Feng HH, *et al.* Underwater acoustic target recognition based on dual attention networks and multiresolution convolutional neural networks. *Proceedings of the 2021 OCEANS*. San Diego: IEEE, 2021. 1–5.
- 12 Xie SN, Liu SN, Chen ZY, *et al.* Attentional ShapeContextNet for point cloud recognition. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 4606–4615.
- 13 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 14 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 3–19.
- 15 Chen L, Zhang HW, Xiao J, *et al.* SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 5659–5667.
- 16 Guo MH, Liu ZN, Mu TJ, *et al.* Beyond self-attention: External attention using two linear layers for visual tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(5): 5436–5447.
- 17 Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. New Orleans: ACL, 2018. 464–468.
- 18 Guo MH, Cai JX, Liu ZN, *et al.* PCT: Point cloud transformer. *Computational Visual Media*, 2021, 7(2): 187–199. [doi: [10.1007/s41095-021-0229-5](https://doi.org/10.1007/s41095-021-0229-5)]
- 19 Santos-Domínguez D, Torres-Guijarro S, Cardenal-López A, *et al.* ShipsEar: 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)]

(校对责编: 牛欣悦)