◇ 研究报告 ◇

## 基于深度神经网络的水声信号恢复方法研究\*

## 王全东<sup>1,2</sup> 郭良浩<sup>1†</sup> 闫 超<sup>1</sup>

(1 中国科学院声学研究所 声场声信息国家重点实验室 北京 100190)

(2 中国科学院大学 北京 100049)

**摘要** 针对干扰或噪声环境下水声目标信号难以获取的问题,该文提出研究基于深度神经网络的自适应水声 被动信号波形恢复方法。在单阵元情况下,该方法提取对数功率谱特征作为输入,采用深度神经网络回归模型 自适应学习目标信号的自身特征,输出降噪后的对数功率谱特征并还原时域波形。在多阵元情况下,提出阵列 深度神经网络降噪方法,将部分或全部阵元特征拼接为长向量作为输入,从而利用空域信息。为全面利用阵列 丰富的时频域信息,该文提出一种两阶段特征融合深度神经网络,在第一阶段将阵列分为若干个子阵,将每个 子阵分别用阵列深度神经网络进行处理,在第二阶段将第一阶段的各子阵处理结果与阵列接收信号同时输入 一个深度神经网络进行融合学习。实验表明,所提出的单阵元和两阶段融合深度神经网络取得了显著优于常 规波束形成的恢复结果,能够准确估计目标信号波形和功率并显著提高输出信噪比。 关键词 水声被动降噪,深度神经网络,阵列处理,特征映射

中图法分类号: O427.9 文献标识码: A 文章编号: 1000-310X(2019)06-1004-11 DOI: 10.11684/j.issn.1000-310X.2019.06.014

#### Underwater acoustic target waveform recovery based on deep neural networks

WANG Quandong<sup>1,2</sup> GUO Lianghao<sup>1</sup> YAN Chao<sup>1</sup>

(1 State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China)

(2 University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract A deep neural network (DNN) based method is proposed to recover underwater acoustic target signal waveform under noise. In single sensor condition, the log-power spectral (LPS) feature is extracted as input, and a DNN regression model is employed to adaptively learn the inherent pattern of target signal and output the enhanced LPS to recover the waveform. In multi-sensor condition, an array-based DNN which uses the concatenated feature from partial or all sensors as input is proposed to exploit spatial information. To fully use the rich temporal and spatial information from the array, we propose a two-stage DNN. In the first stage, the array is split into sub-arrays and each sub-array is processed by an array-based DNN, while in the second stage, the sub-array enhanced features and noisy array features are input to a DNN for integration. Experiments show that our single-sensor and two-stage DNN achieved far better recovery results than conventional beamforming, can accurately recover the target waveform and power and significantly improve the output signal-to-noise ratio.

Key words Underwater acoustic signal recovery, Deep neural network, Array processing, Feature mapping

<sup>2019-01-28</sup> 收稿; 2019-03-25 定稿

<sup>\*</sup>国家自然科学基金项目(61571436)

作者简介: 王全东 (1991-), 男, 黑龙江哈尔滨人, 博士研究生, 研究方向: 信号与信息处理。

<sup>†</sup>通讯作者 E-mail: glh2002@mail.ioa.ac.cn

### 0 引言

海洋环境下,水声目标信号常常被强干扰或背 景噪声所掩盖,造成水声目标识别十分困难。因此, 为提高识别的有效性和可靠性,抑制干扰因素,还原 目标信号波形显得尤为重要。水声被动信号恢复技 术旨在研究处理被动接收信号,适应海洋环境的复 杂性,抑制干扰和背景噪声的同时保证目标信号不 失真,从而恢复目标信号波形,对国家安全、海洋监 测及开发等领域有重要意义。

在单阵元观测情况下,目前学者已提出了很多 算法来解决信号降噪问题。谱减法最早被Boll<sup>[1]</sup>提 出,是将噪声谱从接收信号谱中减去,但会引入不 自然的人工噪声;基于最小均方误差的降噪算法包 括Ephraim等<sup>[2]</sup>提出的最小均方误差估计器,以及 Cohen等<sup>[3]</sup>提出的最优对数幅度谱估计法,这些方 法能降低人工噪声,但是需要对噪声谱进行估计,在 信噪比较低时难以取得满意效果;其他经典算法包 括经验模态分解<sup>[4]</sup>、小波变换<sup>[5-6]</sup>、奇异谱分解<sup>[7]</sup> 和单阵元盲源分离<sup>[8-9]</sup>,这些算法通常需要对噪声 的能量大小、概率密度分布和统计特性做出较多 的先验假设,然而在实际海洋条件下这些假设很难 满足。

在多阵元观测情况下,算法可以同时利用空域 和频域信息,因而可以取得较单阵元更显著的降噪 效果<sup>[10]</sup>。过去几十年来,水声阵列信号降噪的研究 工作主要集中在波束形成理论上。波束形成技术通 过噪声干扰抵消达到阵列降噪的目的,是改善信噪 比的有效手段。一般地,波束形成类方法分常规波 束形成 (Conventional beamforming, CBF) 和自适 应波束形成。常规波束形成可在保持信号幅度不变 的同时,衰减干扰噪声信号。该类方法鲁棒性强,实 现简单,但在干扰方位与目标方位临近的情况下难 以抑制强干扰。自适应波束形成包括最小方差无失 真响应波束形成(Minimum variance distortionless response, MVDR)、广义旁瓣消除器(Generalized sidelobe canceller, GSC)、最大信噪比波束形成器 等。MVDR由Capon<sup>[11]</sup>最先提出,对期望信号进行 无失真约束,同时使得总的输出功率最小。该类方 法具有较好的方位分辨率和较强的干扰抑制能力, 但其需要已知精确的期望信号方位,对参数失配非 常敏感。Cox等<sup>[12]</sup>提出对白噪声增益进行约束,从 而导出了对角加载 MVDR,提高了对基阵误差的稳 健性。对角加载 MVDR 因操作简单而广泛应用, Du 等<sup>[13]</sup>针对加载量不易确定的问题提出了全自动对 角加载算法,进一步提高了鲁棒性。Griffths等<sup>[14]</sup> 引入了GSC作为MVDR的一种等价形式,GSC将 固定波束形成器与自适应滤波器结合,显著降低了 MVDR计算量。针对GSC的方位敏感问题,研究 人员提出利用特征分析技术提取干扰噪声子空间, 降低了阻塞矩阵的信号泄漏<sup>[15]</sup>。总的来说,自适 应波束形成取得了显著进展,但其依然对目标信号 方位和阵元位置有着较高的要求。在浅海垂直阵 条件下,在多途作用下阵元之间的相关性较低,尽 管距离较近时可采用聚焦波束形成来进行降噪和 定位[16-17],但远距离情况下依然较难准确获得目 标信号在各阵元的到达时间差(Time difference of arrival, TDOA),因而很难通过波束形成达到降噪 的目的。

最近几年,深度学习理论在图像处理、语音识别、人工智能等重要领域取得了成功。深度神经网络 (Deep neural network, DNN) 作为一种卓越的信号处理模型,具有强大的建模能力,是由数据驱动的,因而能将已有的数据应用起来完成分类、回归等任务<sup>[18]</sup>。目前已有研究人员将浅层神经网络和深层神经网络用于水声定位研究中<sup>[19-21]</sup>,取得了显著进展。这种监督学习的方法也可以用于估计带噪信号和目标信号之间的复杂的非线性映射,具备更强的能力去从噪声中分辨目标信号,对比传统无监督方法已显示出了卓越的性能提升<sup>[22]</sup>。目前已有采用 DNN 进行主动声呐信号降噪的方法<sup>[23]</sup>,但是这种方法只适用于单阵元接收信号,也仅在仿真中对加性白噪声进行了抑制,对于其他真实情况下干扰或噪声的抑制还没有进行研究。

本文提出一种适用于水声被动目标信号波形恢复的方法,此方法能够自适应地学习目标信号自身的时频特征,有效去除不符合目标信号特征分布的噪声,在保真目标信号的同时,提高输出信噪比(Signal to noise ratio, SNR)。本方法不限定噪声的统计分布(各向同性噪声,临近方向干扰),也对阵列排布没有限制,对训练数据集中不存在的信号具有一定的泛化能力,因而可以应用在各种阵列条件下。首先,在单阵元被动接收情况下,本方法采用DNN作为回归模型,估计带嗓信号到纯净信号之间的映射函数,将时间上扩帧的带嗓对数功率谱特

征 (Log-power spectra, LPS) 映射为纯净的 LPS,再 结合带噪相位利用重叠相加法<sup>[24]</sup>将信号变换回时 域。本文进一步采用 DNN 同时估计纯净 LPS 和理 想比率掩蔽(Ideal ratio mask, IRM)以提高DNN 输出的稳健性。第二,提出阵列DNN方法,将多个 阵元的 LPS 特征拼接为一个长向量作为 DNN 的输 入,从而利用阵列空间信息来恢复目标信号。同时 研究了不同阵元规模情况下如何有效利用阵列信 号中的时间信息和空域信息。第三,在阵列DNN的 基础上提出了两阶段融合 DNN 模型, 第一阶段将 阵列分为若干个子阵,分别利用每个子阵训练阵列 DNN,第二阶段将第一阶段降噪后的特征以及所有 阵元的带噪特征一起输入一个 DNN 进行特征融合 训练,这样的两阶段 DNN 处理可看作一个组合模 型,第一阶段主要利用时间信息,第二阶段主要强调 空间信息,因而能全面利用各阵元包含的丰富时频 域信息。实验证明单阵元DNN能明显提高输出信 号质量,抑制噪声,最后的两阶段DNN 模型可以取 得上述模型中最佳的处理效果。

本文内容安排如下:第一部分介绍神经网络回 归理论和单阵元 DNN 模型。第二部分阐述提出的 阵列 DNN 模型和两阶段融合模型。实验结果和分 析为第三部分。最后一个部分为结论。

### 1 单阵元DNN 模型

#### 1.1 信号模型

考虑一个任意的M阵元的接收阵。阵列接收信 号经过短时傅里叶变换(Short time Fourier transform, STFT)后在频域表示为M维向量 $Z(\omega_l, k)$ ,

$$\mathbf{Z}(\omega_l, k) = \mathbf{S}(\omega_l, k) + \mathbf{N}(\omega_l, k),$$
$$l = 1, 2, \cdots, L, \tag{1}$$

其中, $Z(\omega_l, k) = [z_1(\omega_l, k) z_i(\omega_l, k) \cdots z_M(\omega_l, k)]^T$ 代表接收向量, $S(\omega_l, k) = [s_1(\omega_l, k) s_i(\omega_l, k) \cdots s_M(\omega_l, k)]^T$  代表M维信号向量, $N(\omega_l, k) = [n_1(\omega_l, k) n_i(\omega_l, k) \cdots n_M(\omega_l, k)]^T$ 为M维噪 声或干扰向量, $(\cdot)^T$ 代表转置。 $z_i(\omega_l, k), s_i(\omega_l, k)$ 和 $n_i(\omega_l, k)$ 分别代表第*i*阵元的接收信号、目标信号和 噪声, $i = 1, \cdots, M$ 。 $\omega_l$ 为离散频率,l为频率索引, L为处理的频率点数,k为时间帧索引。为简洁描述,文章余下部分省略 $\omega_l$ 。当M = 1时,接收信号简 化为单阵元信号。

#### 1.2 神经网络回归模型

在人工神经网络理论中,无限逼近定理<sup>[25-26]</sup> 为DNN强大的映射学习能力提供了理论基础。无 限逼近定理表明,一个仅有一个隐藏层的前馈多层 感知机含有一个任意的连续S型函数和有限神经元 数量,可以在ℝ<sup>n</sup>的子集上以任意精度逼近任意连 续的函数。

令 $\sigma$ 为任意连续S型函数,如 sigmoid 函数。 $I_n$ 代表n维单位超立方体 $[0,1]^n$ 。 $I_n$ 上的实值连续函 数空间表示为 $C(I_n)$ 。无限逼近定理表述如下<sup>[26]</sup>: 函数形式

$$G(\boldsymbol{x}) = \sum_{j=1}^{N} \beta_j \sigma(\boldsymbol{y}_j^{\mathrm{T}} \boldsymbol{x} + \theta_j)$$
(2)

在  $C(I_n)$  中 是 稠 密 (dense) 的, 其 中 N 为 整 数,  $\beta_j, \theta_j \in \mathbb{R}$ 为实参数,  $y_j$  是实向量。换言之,给定 任意  $f \in C(I_n)$  和  $\varepsilon > 0$ ,总有一个函数 G(x),对于 所有  $x \in I_n$  满足  $|G(x) - f(x)| < \varepsilon$ 。

人工神经网络的函数逼近误差是有界的<sup>[25]</sup>。 对于一个单隐藏层含有N个S型神经元节点的前 馈网络,网络逼近函数的积分平方误差在cf/N内, cf 取决于被逼近函数的傅里叶变换的模。分析公 式(2)可知,最简单的回归网络结构含有一个输入 层, 对应于形式  $y_i^{\mathrm{T}} x + \theta_i$ , 一个隐藏的S型函数层, 对应于 $\sigma(\boldsymbol{y}_{i}^{\mathrm{T}}\boldsymbol{x} + \theta_{i})$ 和一个线性输出层来对所有神 经元的值进行加权求和作为最后的标量输出 $G(\mathbf{x})$ 。 最近,上述向量对标量的多层感知机映射已被拓展 到高维的向量对向量的深度回归模型<sup>[22]</sup>,也即输出 不再是一个标量而是一个多维向量,可以利用其映 射能力来研究水声被动信号恢复问题。在本文中所 用深度神经网络的结构如下:输入向量 x 由接收信 号的L维LPS特征构成,LPS特征即为 $\log(P_Z)$ ,其 中PZ为功率谱,输出也为一个向量,其中包含恢复 出的目标信号的LPS特征,中间不再只有一个隐藏 层,隐藏层节点数为N。

#### 1.3 单阵元 DNN

在单阵元条件下,本文提出用DNN回归模型 来进行水声目标信号恢复。如图1所示,其处理过程 包括两个阶段:训练阶段和测试阶段。对于训练阶 段,首先构造一个训练数据集,然后提取训练所需的 特征,设计好DNN结构并进行训练。DNN的输入为 接收信号的LPS进行扩帧后的特征,输出为纯净的 目标信号的当前时间帧的LPS特征。DNN的目标 函数 E 为输出的 LPS 特征  $\hat{S}_k$  与纯净目标信号 LPS 特征  $\bar{S}_k$  的均方误差 (Mean square error, MSE),表示如下:

$$\boldsymbol{E} = \frac{1}{NN} \sum_{k=1}^{NN} \left\| \hat{\boldsymbol{S}}_k(\bar{\boldsymbol{Z}}_{k\pm\tau}, \boldsymbol{W}, \boldsymbol{b}) - \bar{\boldsymbol{S}}_k \right\|_2^2, \quad (3)$$

其中, NN为批处理个数,  $\hat{S}_k$ 和 $\bar{S}_k$ 的维数均为 L,分别为输出的LPS和纯净的LPS,  $\bar{Z}_{k\pm\tau} =$  $[\bar{Z}_{k-\tau}, \bar{Z}_{k-\tau+1}, \cdots, \bar{Z}_k, \bar{Z}_{k+1}, \cdots, \bar{Z}_{k+\tau}]$ 为扩帧拼 接的输入带噪LPS特征,扩帧窗长 $R = 2\tau + 1$ , 而(W, b)为DNN的权值和偏置的集合。扩帧能使 DNN同时学习时间和频域尺度上的信息,有助于 DNN分辨目标信号和噪声干扰。DNN可以用随机 梯度下降法来最小化目标函数直至收敛完成训练。 在此基础上本文采用多目标(multi-target)训练,将 IRM特征作为正则项与纯净LPS一同作为输出以 期进一步提高DNN的泛化能力<sup>[27]</sup>。L维IRM特征 定义如下:

$$\mathbf{IRM}_{k} = \sqrt{\frac{\mathrm{e}^{\bar{\boldsymbol{S}}_{k}}}{\mathrm{e}^{\bar{\boldsymbol{S}}_{k}} + \mathrm{e}^{\bar{\boldsymbol{N}}_{k}}}},\tag{4}$$

其中, $\bar{N}_k$ 代表噪声的LPS特征。IRM的取值范围 是[0,1],其物理意义为归一化的输入SNR。



图1 单阵元 DNN 处理框图

Fig. 1 The block diagram of the single-sensor DNN method

加入 IRM 后的目标函数变为  

$$E_{\text{IRM}} = \frac{1}{NN} \sum_{k=1}^{NN} \left\| \hat{S}_k(\bar{Z}_{k\pm\tau}, \boldsymbol{W}, \boldsymbol{b}) - \bar{S}_k \right\|_2^2$$

$$+ \alpha \frac{1}{NN} \sum_{k=1}^{NN} \left\| \widehat{\mathbf{IRM}}_k(\bar{Z}_{k\pm\tau}, \boldsymbol{W}, \boldsymbol{b}) - \mathbf{IRM}_k \right\|_2^2, \quad (5)$$

其中, $\widehat{IRM}_k$ 为估计的 IRM 特征,  $\alpha$ 为 IRM 目标权 值。采用均值方差归一化变换<sup>[28]</sup> 用全局的均值和 方差将输入向量和输出向量逐维变换为零均值、单 位方差的向量再传输给 DNN 进行训练。 在测试阶段,分别提取需降噪的信号的LPS和相位,将带噪LPS输入DNN处理,输出恢复的LPS特征。有研究表明,带噪相位本身就是对纯净相位的一种最小均方误差估计<sup>[2]</sup>,因而可以将恢复的特征与带噪相位结合进行反短时傅里叶变换(Inverse short-time Fourier transform, ISTFT),并用重叠相加法<sup>[24]</sup>恢复为时域波形。

#### 2 基于阵列的 DNN 模型

#### 2.1 阵列DNN

针对阵列接收信号降噪问题,本文提出阵列 DNN算法,将多个阵元的带噪特征拼接在一起形 成一个长向量作为DNN输入。在受限的DNN宽度 (隐层单元数)下,若将K(K ≤ M)个阵元信号分 别扩帧后再拼接,输入维数将达到L×R×K,可能 会造成输入维数过高,难以完成稳定的训练或造成 过拟合。因此,本文控制了时间窗长的大小使得输 入维数在不同阵元数条件下可比,并研究了不同阵 元数的降噪性能。可将任一阵元设置为参考阵元, 从阵列带噪信号中估计参考阵元中的目标信号波 形。利用多目标训练,阵列DNN的目标函数表示为

$$\boldsymbol{E}_{\mathrm{Arr}} = \frac{1}{NN} \sum_{k=1}^{NN} \left\| \hat{\boldsymbol{S}}_{k}(\bar{\boldsymbol{Y}}_{k\pm\tau}, \boldsymbol{W}, \boldsymbol{b}) - \bar{\boldsymbol{S}}_{k} \right\|_{2}^{2} + \alpha \frac{1}{NN} \sum_{k=1}^{NN} \left\| \widehat{\mathbf{IRM}}_{k}(\bar{\boldsymbol{Y}}_{k\pm\tau}, \boldsymbol{W}, \boldsymbol{b}) - \mathbf{IRM}_{k} \right\|_{2}^{2}, \quad (6)$$

其中,当K < M时, $\bar{Y}_{k\pm\tau} = [\bar{Z}_{1,k\pm\tau}, \cdots, \bar{Z}_{K,k\pm\tau}]$ 为K个阵元各自扩帧后的拼接输入向量, $\bar{Z}_{i,k\pm\tau}$ 为 扩帧的第i个阵元的带噪LPS,当K = M时, $\tau = 0$ 各阵元不进行扩帧操作。这里的IRM目标由参考 阵元的信号和噪声计算而来。由于不同阵元接收的 是来自不同水声传播路径混合的信号,所以它们具 有相关且互补的特征,因此可以让DNN学习到阵 列中的空域信息。这种将多个阵元信号同时送入 DNN并以最小均方误差作为损失函数的处理方式 与波束形成类似,但不同点在于阵列DNN能够自 适应学习非线性的映射函数。

#### 2.2 两阶段融合 DNN

如上文所述,阵元数较多情况下,阵列DNN仅 将一部分阵元信号进行拼接作为输入时可以进行 扩帧利用时间信息,但不能利用所有阵元的空间信 息,而当阵列DNN拼接所有阵元的信号作为输入 时,为控制输入特征总维数,将不会采取扩帧的操作 而直接将所有阵元的当前时间帧拼接在一起,但这 样不能充分利用时间信息,所以总体上由于输入维 数限制,阵列DNN不能同时充分利用时间和空间 上的信息。为解决这一问题,本文提出一种两阶段 阵列 DNN 融合的处理方法。这是一个组合模型,其 结构如图2所示。第一阶段是增强阶段,将M个阵 元分为Q组,每一组K个阵元组成一个子阵,扩帧 后采用阵列 DNN 进行训练, 记为增强 DNN。这里 DNN 不进行 IRM 的训练,以便输出给下一个阶段。 这一阶段的每个DNN 强调利用时间信息,但没有 利用所有的阵元信息。第二阶段为融合阶段,在第 一阶段的Q个DNN之上进一步挖掘阵列的空间信 息,记为融合 DNN。将增强阶段的 Q 个增强 DNN 输出和所有阵元带噪LPS特征拼接在一起作为融 合 DNN 的输入去映射目标信号的 LPS 和 IRM。融 合DNN的目标函数定义为

$$\boldsymbol{E}_{\text{INT}} = \frac{1}{NN} \sum_{k=1}^{NN} \left\| \hat{\boldsymbol{S}}_{k}(\bar{\boldsymbol{X}}_{k}, \boldsymbol{W}, \boldsymbol{b}) - \bar{\boldsymbol{S}}_{k} \right\|_{2}^{2} + \alpha \frac{1}{NN} \sum_{k=1}^{NN} \left\| \widehat{\text{IRM}}_{k}(\bar{\boldsymbol{X}}_{k}, \boldsymbol{W}, \boldsymbol{b}) - \text{IRM}_{k} \right\|_{2}^{2}, \quad (7)$$

其中,  $\bar{X}_k = [\tilde{S}_{1,k}, \cdots, \tilde{S}_{Q,k}, \bar{Z}_{1,k}, \cdots, \bar{Z}_{M,k}]$ 为融 合DNN的输入向量, 包含Q个增强DNN的当前帧 输出和所有阵元的带噪当前帧LPS,  $\tilde{S}_{i,k}$ 为上一阶 段的第i个DNN的输出。通过两阶段融合DNN模 型, 可以充分利用DNN的映射能力并挖掘各阵元 的时间信息和空间信息来进行目标信号恢复。



图2 两阶段融合DNN结构图



#### 3 实验结果和分析

为验证本文所提方法的降噪能力,本文进行了 七组对比实验。本实验采用一次海试实验中的船 辐射噪声作为目标信号,其实验条件如下:声源深 度为5 m,水深152 m,八阵元接收垂直阵深度为 128~135 m, 阵元间隔1m; 沉积层厚24m, 声 速1572~1593 m/s,密度1.76 g/cm<sup>3</sup>;水底为流体 半空间, 声速 5200 m/s, 密度 1.8 g/cm<sup>3</sup>; 深度剖面 为典型负梯度。本文选用时间段为2009年1月31 日01:43-02:05,期间目标匀速逐渐远离接收阵。本 文将公开数据集SWellEX-96<sup>[29]</sup> S5实验中的信号 作为要去除的噪声,选取垂直阵前八个阵元的信 号,所用时间段为1996年5月10日23:15-23:58,其 有两个噪声源,较深的源发射信号为65个单频信 号和FM chirp信号,较浅的源发射另外9个单频 信号,所有单频信号在49~400 Hz之间。依据公 式(1)合成阵列接收信号,单阵元接收SNR分别为 -10 dB、-5 dB、0 dB、5 dB、10 dB。训练信号是 将船辐射噪声数据的前17 min和SWellEX-96数据 的前38 min拆分成多个时间片段,进行随机组合按 信噪比混合在一起,得到一个约90h记录时间的训 练集。测试集信号为船辐射噪声最后6 min 数据和 SWellEX-96 最后6 min 数据混合而成,因而是训练 数据集中不存在的,其信噪比设置与训练集相同。 训练数据中的目标信号和噪声的时频图分别为图3 和图4,可以看出目标信号与噪声的频域分布不同, 目标信号和噪声都既包含连续谱也包含强线谱。



图3 前17 min 目标信号时频谱



本实验的分析频带为0~500 Hz,频域分辨率为1.95 Hz,LPS特征维数为257(一帧),IRM 维数也为257 维。降噪所用的DNN结构为输入层为线性层,中间为两层隐藏层,每层2048节点数,S型函数为sigmoid,输出层也为线性层。DNN初始学习率为0.00002,采用随机梯度下降法训练,批处理数量

为32。训练IRM时, $\alpha$ 为0.05。为检验算法降噪和 保真的效果,采用两种客观描述参数。





第一个参数为输出 SNR, 定义为  
SNR = 10 lg 
$$\left\{ \frac{1}{T} \sum_{k} \frac{\sum_{l} P_{tar}(\omega_{l}, k)}{\sum_{l} P_{noi}(\omega_{l}, k)} \right\},$$
 (8)

其中,T为输出时间帧总数,P<sub>tar</sub>(ω<sub>l</sub>,k)和P<sub>noi</sub>(ω<sub>l</sub>,k) 分别代表输出信号中目标信号的功率谱和输出信 号中噪声的功率谱。需要指出的是,由于DNN为非 线性处理,不能通过分别输入纯目标信号特征和纯 噪声特征分别得到输出的目标信号特征和残余噪 声特征来直接计算输出SNR,所以这里目标信号和 噪声功率的估计方法如下:首先根据IRM确定带噪 信号时频图上噪声占主导的时频区域,IRM值小于 0.2处视为噪声区域,计算输出信号中这一区域的 功率作为噪声功率,其他区域的功率作为目标信号 功率。

第二个参数为输出功率的均方误差 MSE, 物理 意义为输出信号的功率估计误差, 定义为

 $MSE = \frac{1}{T} \sum_{k} \sum_{l} [P_{test}(\omega_l, k) - P_{clean}(\omega_l, k)]^2, (9)$ 其中,  $P_{test}(\omega_l, k)$  和  $P_{clean}(\omega_l, k)$  分别代表待测试的 输出信号的功率谱和纯净目标信号的功率谱。输 出 SNR 和 MSE 是一组互补的参数,比如输出 SNR 增高的时候, MSE 不一定减小,即输出能量未必跟 原来相同。因此不同算法对比时,以下三种情况都 可以认为达到了更好的恢复效果: SNR 增大,同时 MSE 减小; MSE 相当的情况下, SNR 增大; 以及 SNR 相当的情况下, MSE 显著减小。

#### 3.1 常规波束形成

本实验将第四个阵元(ch4)作为参考阵元,并 在所有对比算法中估计第四个阵元中的目标信号。 作为对比,采用水声环境下常用的CBF对阵列数 据进行处理。常规波束形成算法流程简述如下:首 先根据阵列位置和俯仰角扫描范围(0°~180°)确 定阵列流形向量,然后利用较高频段的接收数据 (300~500 Hz)来进行波束扫描,从而获得每一帧信 号的扫描方位谱(接收协方差矩阵利用21个快拍来 进行估计),根据方位谱峰值所在的位置确定每一帧 接收信号的俯仰角,并据此进行常规波束形成获得 输出波形和功率谱。本文的俯仰角的零度定义为垂 直向上方向。

作为测试示例,0dB带噪测试信号的对数功 率谱如图5所示,其中的噪声成分的对数功率谱为 图6,真实目标信号的对数功率谱为图7。观察到 噪声信号几乎将目标信号掩盖。下面给出CBF处 理结果。图8展示了0dB条件下CBF波束扫描后 的入射俯仰角估计结果,图9为10dB条件下的扫 描结果。对比这两图可知,在较低信噪比和浅水多 途条件下,较难获得准确的俯仰角估计结果。图10 为0 dB下根据图8的俯仰角估计结果,利用CBF 获得的输出信号的对数功率谱,可以看到噪声的强 线谱依然清晰可见,没有得到很有效的去除,且由 于入射俯仰角估计错误一些时间段的谱出现了异 常。表1为各算法输出结果的两项客观描述参数在 -10 dB ~ 10 dB 的平均结果。对比表1的第二行 的第四号阵元的接收信号(记为ch4 mix)和第三行 CBF输出结果,CBF算法能够一定程度上提高输出 SNR,但是在多途作用下各阵元接收到的目标信号 是不同的,因而CBF仅是将目标信号成分在不同阵 元间进行了平均,难以取得理想平面波假设下的阵 增益,导致CBF结果与参考阵元接收的目标信号相 关性较差,所以导致了MSE参数反而上升了。



图 5 0 dB 测试带噪信号对数功率谱 (ch4) Fig. 5 The LPS of the testing signal at 0 dB (ch4)



图 6 0 dB 测试信号中的噪声的对数功率谱 (ch4) Fig. 6 The LPS of the noise in the testing signal at 0 dB (ch4)



图 7 0 dB 测试信号中的真实目标信号的对数功率 谱 (ch4)

Fig. 7 The LPS of the target signal in the testing signal at 0 dB (ch4)





Fig. 8 The pitch angle scanning (PAS) result of the target signal at 0 dB

#### 3.2 单阵元 DNN

下面利用 DNN 来进行目标信号恢复。单阵元 DNN 利用了第四阵元信号的时间信息和频域信息, 其输入为扩帧 LPS, 维数为 257 × 11 × 1, 输出为当



图 9 10 dB下目标信号入射俯仰角扫描结果 Fig. 9 The PAS result of the target signal at 10 dB



图 10 0 dB 下的 CBF 输出的对数功率谱

Fig. 10 The LPS output by the CBF at 0 dB  $\,$ 

# 表1 各个对比算法的输出参数在 –10 dB ~ 10 dB 的平均结果

Table 1 The average objective measurements for all algorithms at  $-10 \text{ dB} \sim 10 \text{ dB}$ 

对比算法		输出 SNR/dB	MSE
ch4	mix	0.97	214.99
CBF		2.29	313.13
ch4	noIRM	7.68	16.74
ch4	IRM	7.44	7.89
ch45	noIRM	7.90	14.96
ch3456	noIRM	7.66	6.16
8ch	noIRM	7.91	14.55
8ch	IRM	7.69	5.88
2stage	IRM	9.15	6.00

前帧257维LPS,记为'ch4 noIRM'。如图11所示, 将带噪信号输入单阵元DNN后输出的对数功率谱 很接近图7,这显示出DNN具有强大的建模能力, 显著抑制了噪声成分,目标信号恢复效果明显。对 比表1的第三行和第四行,单阵元DNN输出SNR 比CBF高5dB以上的同时,显著降低了功率估计 的MSE,表明本处理方法对测试信号有较强的泛化 能力。





#### 3.3 阵列DNN

根据第2.1的分析,由于DNN输入维数的限制, 在利用较多阵元数时,需要降低扩帧的窗长来保 证 DNN 训练的稳定。因此本实验研究了利用两个 阵元(阵元4和阵元5,记为'ch45')、四个阵元(阵元 3~ 阵元6,记为'ch3456')和八个阵元(阵元1~ 阵 元8,记为'8ch')的情况,分别采用窗长R等于5、3、 1,均没有训练IRM特征,标记为'noIRM'。三者的 输出结果列于表1的第六、第七、第八行,可以看出 三个阵列 DNN 的恢复效果均优于单阵元 DNN, 表 明使用阵元DNN利用了阵列的空间信息,提高了 DNN估计的准确度。四阵元系统'ch3456'的输出 SNR与两阵元系统和八阵元系统相当,但是取得了 三者之中最小的 MSE, 更好地估计了目标波形的功 率大小。这说明在阵列 DNN 的框架下, 阵元数越多 不一定越好,综合利用空间信息和时间信息才能取 得更佳的恢复效果。

#### 3.4 单目标训练对比多目标训练

在DNN输出层不仅可以输出目标信号LPS, 也可以估计IRM,即输出总维数为514。表1的第五 行和第九行为单阵元DNN和八阵元DNN同时训 练LPS和IRM的结果,标记为'IRM'。在同等情况 下,标记'IRM'的系统在保持输出SNR相当时输出 MSE的要明显优于标记'noIRM'的DNN系统,说 明多目标映射时,IRM作为目标函数中的正则项, 与LPS一起训练,能提升DNN估计LPS的鲁棒性, 更好地恢复目标信号功率。下面简述一下IRM正则 项提高性能的原因。类似于文献[12]中在目标函数 中加入正则项来提高MVDR的鲁棒性,本文式(5)、 式(6)、式(7)加入右边第二项的正则化项后使网络 进行多目标或多任务学习,由于IRM特征的物理意 义为归一化的输入信噪比,所以相当于在估计LPS 特征这个主要任务上多加了一个估计输入信噪比 的附加任务。在机器学习理论中,通常来讲,用共享 的神经网络来同时学习一个或多个额外的适当任 务可以同时提高全部任务的学习能力,而在实际使 用网络时可以将用于学习额外任务的参数去除。总 的来说, IRM 特征估计作为正则项可以限制深度神 经网络参数,引入损失函数的归纳偏差,迫使学习 算法在两个任务的交叉表征区域找到最终解,避免 单任务下在一个更大表征区域求得局部最优解,因 而可以降低网络过拟合风险使得网络更好、更快地 收敛。

#### 3.5 两阶段融合 DNN

根据第3.3节的分析,为了全面利用八个阵 元信号的时间和空间信息,本文训练了两阶段 融合系统。第一阶段训练了两个增强DNN,分 别为'ch3456'和'ch1234'(由阵元1~ 阵元4训练)。 'ch1234'与'ch3456'性能接近,因此未列入表1。第 二阶段中,将第一阶段的两个DNN的LPS输出和 所有八个阵元的带噪信号拼接在一起,形成一个维 数为257×10的向量作为输入,输出为LPS和IRM 特征,共514维。图12为两阶段DNN处理0dB测 试信号后的对数功率谱,可以看到图7中目标信 号在50~100 Hz之间、360 Hz和420 Hz处的线 谱在图12中得到了很好的突出和保留,目标信号 在180~240 Hz之间的连续谱也得到了有效恢复, 说明本方法对目标信号和噪声的频域特性没有限 制,因而适用性较好。两阶段DNN的平均输出参 数结果如表1的最后一行所示,记为'2stage'。特征 融合后,其MSE 略好于'ch3456',但输出SNR优于 'ch3456'1 dB 以上。这说明,本方法能在两个四 阵元阵列 DNN 基础上, 进一步综合利用八个阵元 的时间空间信息,取得了所有单阵元和阵列对比算 法中最佳的信号恢复效果。其原因为,在融合阶段 DNN 的映射能力被再次利用,由于每个阵元的带噪 信号中含有一部分较高信噪比的时频点,将其与降 噪后的特征一同送入DNN本质上提高了训练的丰 富度,因此融合DNN能从中学习并将第一阶段没 有处理好的时频点进一步恢复。



图 12 0 dB下两阶段融合 DNN 输出的对数功率谱 Fig. 12 The LPS output by the two-stage DNN at 0 dB

#### 3.6 DNN 的恢复性能随输入信噪比变化情况

为着重讨论DNN目标信号恢复算法在不同输入SNR条件下的性能变化,在图13和图14中对比了CBF、单阵元DNN和两阶段DNN的两种描述参数。由于CBF的MSE远高于本文的DNN算法,故均方误差由对数MSE,即10log(MSE)来衡量。 三种算法的输出SNR随输入SNR增大而增大,对数MSE随输入SNR增大而减小。然而CBF的对数 MSE 随输入SNR增大而减小。然而CBF的对数 MSE 始终高出DNN输出结果30以上,说明CBF 的功率估计远不及DNN准确。随着输入SNR增 大,两阶段DNN的输出SNR与CBF的输出SNR 的差距从10 dB减小到了2 dB,DNN方法比CBF 在低SNR下有着更加突出的优势。此外,两阶段



图 13 不同输入 SNR 下三种算法的输出 SNR 对比 Fig. 13 Output SNR comparisons between three methods under different SNRs



图 14 不同输入 SNR 下三种算法的对数 MSE 对比 Fig. 14 Log-MSE comparisons between three methods under different SNRs

DNN在低SNR条件下比高SNR条件下展示出相 对单阵元DNN更加明显的优势,这表明两阶段 DNN算法融合阵列的丰富时域空域信息对于低 SNR接收信号有着更重要的意义。

#### 3.7 DNN的恢复性能随频率变化情况

为体现本文方法能够自适应学习目标信号的 时频特征,计算了以频率为变量的输出SNR,其定 义如下:

$$\operatorname{SNR}(\omega_l) = 10 \log \left\{ \frac{\sum_k P_{\operatorname{tar}}(\omega_l, k)}{\sum_k P_{\operatorname{noi}}(\omega_l, k)} \right\}, \quad (10)$$

因为DNN只输出单路LPS, 所以 $P_{tar}(\omega_l, k)$ 和 $P_{noi}(\omega_l, k)$ 两者难以分开, 故采用真实目标信号的功率谱 $P_{clean}(\omega_l, k)$ 来近似 $P_{tar}(\omega_l, k)$ , 用 $P_{test}(\omega_l, k) - P_{clean}(\omega_l, k)$ 来近似 $P_{noi}(\omega_l, k)$ ,从而近似估计单频SNR。

图15对比了在0dB测试条件下的CBF、单阵 元DNN和两阶段DNN在全部频点上输出的SNR。 从图中可以看到,带噪信号的SNR在噪声频点处 (如200 Hz左右、350 Hz左右和400 Hz左右等)有 非常明显的谷点,也即这些频率的信号质量非常差。 经过CBF,这些频点的SNR有所增加,然而还是有 明显的谷点。经过本文DNN处理后,这些噪声谷 点被基本消除,其中两阶段DNN在各个频点的输 出SNR最高。此外,可以看到DNN输出的SNR在 原目标信号的功率较强的频率点上出现了峰值(如 100 Hz左右和360 Hz左右),较好地恢复了目标信 号原有的频率分布。



图 15 0 dB条件下全部频点处 CBF、单阵元 DNN 和两阶段 DNN 的输出 SNR 对比

Fig. 15 The output SNR comparisons between CBF, single-sensor DNN and two-stage DNN at different frequencies at 0 dB

为体现 DNN 在不同频率下的恢复性能差异, 在图 15基础上计算了全部频率下三种算法的 SNR 增益,如图 16 所示。可以看到,三种算法在噪声强线 谱频点上均有较大的 SNR 增益的峰值,其中两阶段 DNN 的 SNR 增益最高。结合图 15 可以得到以下结 论, DNN 恢复算法在噪声较强的频点上具有更高的 SNR 增益,所以能抑制噪声并准确恢复目标信号的 频率分布。对比图 7、图 10 和图 12 中真实目标信号 的对数功率谱、CBF 和两阶段 DNN 的处理结果,可 以看出两阶段 DNN 输出信号比 CBF 输出信号更接 近真实目标信号。由此说明本文 DNN 处理方法能 够自适应地学习目标信号自身的时频特征,有效去 除不符合目标信号特征分布的噪声。



图 16 0 dB条件下全部频点处 CBF、单阵元 DNN 和两阶段 DNN 的输出 SNR 增益对比

Fig. 16 The output SNR gain comparisons between CBF, single-sensor DNN and two-stage DNN at different frequencies at 0 dB

#### 4 结论

本文提出了一种基于 DNN 的水声目标信号被 动恢复方法。DNN作为回归模型,具有强大的建模 能力,建立了带噪对数功率谱特征到纯净对数功率 谱特征的多维映射函数。本方法在单阵元下利用时 间和频谱信息恢复了目标信号,并利用多目标训练 提高了鲁棒性;在接收阵情况下采用阵元DNN同 时将阵列信号映射到参考阵元,提高了空域信息利 用能力,并探讨了时间窗和阵元数的限制关系;在此 基础上提出两阶段处理算法来融合降噪后的特征 和所有空域带噪特征,进一步提高了目标信号恢复 能力。实验数据处理结果表明,本方法能显著去除 噪声,输出比常规波束形成更高的SNR并准确估计 目标信号能量。本方法优势在于能够从大量训练数 据中学习目标信号的先验信息,将其有效应用于测 试信号,对于训练集中不存在的信号有一定的泛化 能力,且本方法对噪声分布和阵列排布没有要求,能 够有效恢复线谱和连续谱信号。本文实验的训练数 据和测试数据有一定相关性,但在实际应用中,目标 信号可能会叠加不同种类的噪声或干扰,因此如何 提高在其他恶劣条件下的目标波形恢复的稳健性 需要进一步研究。此外,由于本文的实验为垂直阵 数据,阵元之间的相关性较差,因此DNN在利用更 多阵元数时,没有体现出类似于波束形成在平面波 条件下的增益变化,下一步将研究提高阵元相关性 来进一步提高阵列 DNN 处理的增益。

参考文献

- Boll S. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1979, 27(2): 113–120.
- [2] Ephraim Y, Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984, 32(6): 1109–1121.
- [3] Cohen I, Berdugo B. Speech enhancement for nonstationary noise environments[J]. Signal Processing, 2001, 81(11): 2403-2418.
- [4] 杨宏. 经验模态分解及其在水声信号处理中的应用 [D]. 西安: 西北工业大学, 2015.
- [5] 韩建辉,杨日杰,王伟.基于小波变换的 Power-Law 水声瞬态 信号检测研究 [J].系统仿真学报,2008,20(13):3514–3516, 3520.

- [6] 汤春瑞, 刘丹丹, 曹家年. 基于表面波变换的水声瞬态信号降 嗓方法 [J]. 大连海事大学学报, 2008, 34(3): 51–54. Tang Chunrui, Liu Dandan, Cao Jianian. Denoising method for underwater acoustic transient signals based on surfacelet transform[J]. Journal of Dalian Maritime University, 2008, 34(3): 51–54.
- [7] 李亚安,王洪超,陈静,基于奇异谱分解的水声信号降噪方法研究 [J]. 系统工程与电子技术, 2007, 29(4): 524–527.
  Li Ya'an, Wang Hongchao, Chen Jing. Research of noise reduction of underwater acoustic signals based on singular spectrum analysis[J]. Systems Engineering and Electronics, 2007, 29(4): 524–527.
- [8] 刘佳,杨士莪,朴胜春. 基于 EEMD 的地声信号单通道盲源 分离算法 [J]. 哈尔滨工程大学学报, 2011, 32(2): 194–199. Liu Jia, Yang Shi'e, Piao Shengchun. The single channel seismic-acoustic signal blind-source separation method based on EEMD[J]. Journal of Harbin Engineering University, 2011, 32(2): 194–199.
- [9] 董姝敏.水声信号处理的盲信号分离方法研究[D].哈尔滨: 哈尔滨工程大学, 2012.
- [10] 鄢社锋,马远良. 传感器阵列波束优化设计及应用 [M]. 北京: 科学出版社, 2009.
- [11] Capon J. High-resolution frequency-wavenumber spectrum analysis[J]. Proceedings of the IEEE, 1969, 57(8): 1408–1418.
- [12] Cox H, Zeskind R, Owen M. Robust adaptive beamforming[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1987, 35(10): 1365–1376.
- [13] Du L, Li J, Stoica P. Fully automatic computation of diagonal loading levels for robust adaptive beamforming[J].
   IEEE Transactions on Aerospace & Electronic Systems, 2010, 46(1): 449–458.
- [14] Griffiths L, Jim C. An alternative approach to linearly constrained adaptive beamforming[J]. IEEE Transactions on Antennas and Propagation, 1982, 30(1): 27–34.
- [15] Wang Q, Ren S. Robust generalized sidelobe canceller with an eigenanalysis-based blocking matrix[C]. Int. Conf. Electr. Eng. Autom. Control (ICEEAC), Nanjing, 2017: 410–415.
- [16] 梅继丹,惠俊英,惠娟.聚焦波束形成声图近场被动定位技术 仿真研究 [J].系统仿真学报,2008,20(5):1328-1333.
  Mei Jidan, Hui Junying, Hui Juan. Research on simulation of near field passive ranging with underwater acoustic image by focused beam-forming[J]. Journal of System Simulation, 2008, 20(5): 1328-1333.

[17] 梅继丹,王珺琳,惠俊英. 垂直矢量阵声图被动定位技术研究[J]. 兵工学报, 2010, 31(3): 369–374.
Mei Jidan, Wang Junlin, Hui Junying. Research on the underwater acoustic image passive locating using vertical

linear array[J]. Acta Armamentarii, 2010, 31(3): 369-374.

- [18] Deng L, Hinton G, Kingsbury B. New types of deep neural network learning for speech recognition and related applications: an overview[C]. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), Vancouver, 2013: 8599–8603.
- [19] Niu H, Ozanich E, Gerstoft P. Ship localization in Santa Barbara Channel using machine learning classifiers[J]. Journal of the Acoustical Society of America, 2017, 142(5): EL455–EL460.
- [20] Niu H, Reeves E, Gerstoft P. Source localization in an ocean waveguide using supervised machine learning[J]. Journal of the Acoustical Society of America, 2017, 142(3): 1176–1188.
- [21] Huang Z, Xu J, Gong Z, et al. Source localization using deep neural networks in a shallow water environment[J]. Journal of the Acoustical Society of America, 2018, 143(5): 2922–2932.
- [22] Xu Y, Du J, Dai L, et al. A regression approach to speech enhancement based on deep neural networks[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2015, 23(1): 7–19.
- [23] 任宇飞,李宇,黄海宁.深度学习的主动声纳信号增强方法研究 [J]. 网络新媒体技术, 2017, 6(3): 14–19.
  Ren Yufei, Li Yu, Huang Haining. Research on active sonar signal enhancement based on deep learning[J]. Journal of Network New Media, 2017, 6(3): 14–19.
- [24] Crochiere R. A weighted overlap-add method of shorttime Fourier analysis/synthesis[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1980, 28(1): 99–102.
- [25] Barron A R. Universal approximation bounds for superpositions of a sigmoidal function[J]. IEEE Transactions on Information Theory, 1993, 39(3): 930–945.
- [26] Cybenko G. Approximation by superpositions of a sigmoidal function[J]. Mathematics of Control, Signals and Systems, 1989, 2(4): 303–314.
- [27] Sun L, Du J, Dai L, et al. Multiple-target deep learning for LSTM-RNN based speech enhancement[C]. Handsfree Speech Commun. Microphone Arrays (HSCMA), San Francisco, 2017: 136–140.
- [28] Povey D, Ghoshal A, Boulianne G, et al. The Kaldi speech recognition toolkit[C]. IEEE Workshop Autom. Speech Recognit. Understanding, Hawaii, 2011.
- [29] Murray J, Ensberg D. The swellex-96 experiment[EB/ OL]. [2019-01-01]. http://swellex96.ucsd.edu/index.htm.