

声源定位问题中卷积神经网络结构研究

韦永刚, 肖瑶

北京信息科技大学信息与通信工程学院, 北京

收稿日期: 2024年10月21日; 录用日期: 2024年11月20日; 发布日期: 2024年11月29日

摘要

声源定位是声源信号处理中非常重要的研究目标。传统方法容易受到噪声和混响的干扰。随着深度学习算法在诸多领域的成功应用, 本文探究了使用深度学习算法解决声源定位问题。本文对使用卷积神经网络结构实现基于麦克风信号的声源定位性能分析, 并基于仿真实验探究相同房间条件和声源条件下, 不同卷积层和卷积核数量对于声源定位性能的影响。实验表明, 声音信号的基于相位变换加权的广义互相关特征作为卷积神经网络输入信号, 在声音信噪比10 dB~40 dB, 混响在200~600 ms的常规房间条件设定下, 相比于其他方法其声源定位准确率最高, 且卷积网络中包含6个卷积层, 首层卷积层卷积核为4时其网络定位精度和计算效率之间取得了较好的平衡。

关键词

深度学习, 卷积神经网络, 声源定位, 广义互相关特征

Research on Convolutional Neural Network Structures in Sound Source Localization

Yonggang Wei, Yao Xiao

School of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing

Received: Oct. 21st, 2024; accepted: Nov. 20th, 2024; published: Nov. 29th, 2024

Abstract

Sound source localization is a crucial research objective in sound source signal processing. Traditional methods are prone to interference from noise and reverberation. With the successful application of deep learning algorithms in many fields, this paper explores the use of deep learning algorithms to solve the problem of sound source localization. This study analyzes the performance of sound source localization based on microphone signals using a convolutional neural network (CNN)

structure. Through simulation experiments, we investigate the impact of different numbers of convolutional layers and convolutional kernels on sound source localization performance under the same room and sound source conditions. The experiments show that the sound signal after the generalized cross-correlation phase transform operation is used as the input signal of the convolutional neural network, under typical room conditions with a signal-to-noise ratio of 10 dB~40 dB and reverberation times of 200~600 ms, this method achieves the highest localization accuracy compared to other methods. Furthermore, when the network contains 6 convolutional layers and the first layer has 4 convolutional kernels, a good balance between localization accuracy and computational efficiency is achieved.

Keywords

Deep Learning, Convolutional Neural Network, Sound Source Localization, Generalized Cross Correlation Function

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着人工智能领域的快速发展,声源定位技术成为了基于语音的人机交互研究中的关键技术。原理上,麦克风阵列通过分析多路信号的时空差异实现声源定位,常用的传统方法包括:基于高分辨空间谱的估计算法[1],其具有较好的声源定位准确率,然而实现比较复杂,对于声音环境要求较高;基于波束形成的方法[2],需要声源信号信息作为先验参数,对应用场合要求较高;基于时延的声源定位方法[3]以及基于时延估计的GCC (Generalized Cross-Correlation, GCC, 广义互相关函数) [4]算法,算法使用了加权函数,提高了算法的性能同时增加了算法的复杂度。机器学习的发展也给声源定位提供了新的思路,相较于传统方法,机器学习算法比较简单,具有较高的定位性能,且算法运行时间在秒量级,针对不同的应用场景条件,只需要调整些许参数,重新训练网络,便可获得较高的定位性能。

进一步,深度学习的迅速发展和在诸多领域的成功应用为声源定位提供了灵感[5][6]。文献[7]中采用LS-SVM算法实现声源定位,取得了较好的性能。文献[8]提出了使用BP(后向传播)网络实现声源定位,其能够在高混响、低信噪比环境中仍然能够实现较高的声源定位准确率和较低的均方根差(RMSE);文献[9][10][11]提出了基于卷积神经网络的声源定位方法,以基于信号间相位差信息(IPD),GCC,或者相位变换广义互相关函数(GCC-PHA)提取的声音信号特征作为神经网络输入信息,声源位置坐标作为真值训练网络,其卷积神经网络共包含若干卷积层,且实现了较高的声源定位性能;然而,不同应用场景,不同声音条件下,对能够以较低训练成本实现最优声源定位性能的卷积神经网络的层数和卷积核尺寸尚未讨论。

本文提出一种使用GCC-PHAT [12]提取声源特征的基于卷积神经网络(CNN)的声源定位方法,主要探究在不同声音条件下的最优卷积神经网络结构。本文使用TIMIT数据集作为声源信号,使用RoomSim [13]对房间内声音信号的传播函数进行仿真,若干麦克风接收经过房间传播的声音信号;对麦克风的接收信号进行基于GCC-PHAT的特征提取,提取后的特征作为卷积神经网络的输入,声源信号的位置作为网络真值。实验表明,卷积神经网络的精度和鲁棒性优于传统的TDOA,SRP-PHAT等算法,并通过仿真实验分析了卷积层数、隐藏层数等参数对模型性能的影响规律,发现5个卷积层,首层卷积层卷积核为4时,定位精度和计算效率之间取得了较好的平衡,更适合实际应用场景。

2. 算法基础理论

本文提出的声源定位算法由数据预处理, 声源特征提取和卷积神经网络训练。

2.1. 卷积神经网络

现阶段, 卷积神经网络(Convolutional Neural Network, CNN)是一种典型的深度学习网络, 其最先应用于计算机视觉领域进行表象和视频处理[14] [15]。与传统的神经网络相比, CNN 通过引入卷积操作和池化操作, 能够有效地减少了参数数量, 提高训练效率, 同时保留了空间结构信息。

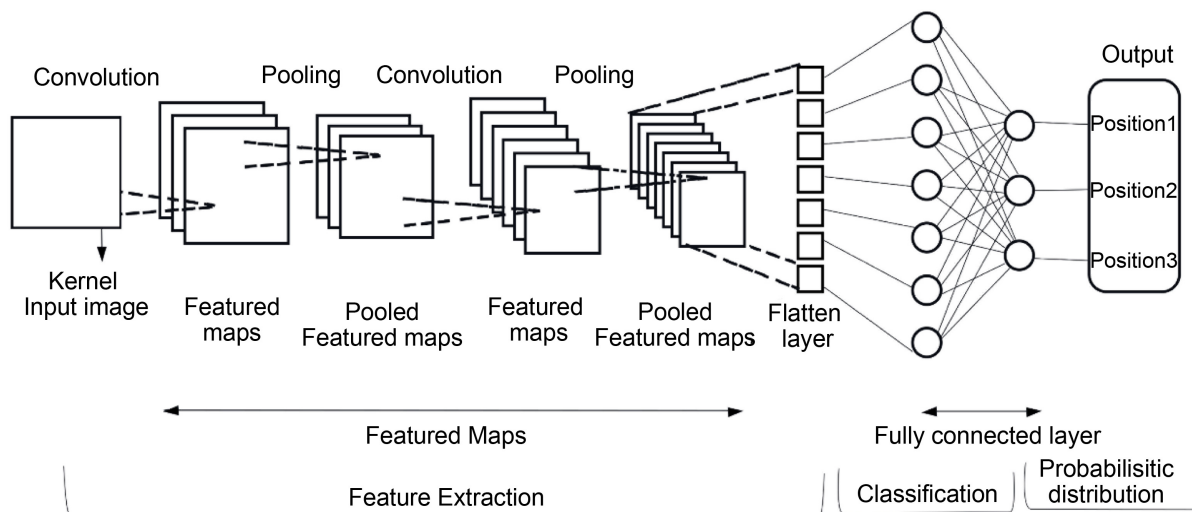


Figure 1. The architecture of the convolutional neural network

图 1. 卷积神经网络结构

卷积神经网络的主要由卷积层, 激活函数, 池化层和全连接组成(如图 1 所示), 其分别的功能为:

1) 卷积层是 CNN 的核心组成部件, 其主要功能是提取输入数据的局部特征。卷积操作通过卷积核(也称滤波器或特征检测)在输入数据上滑动, 逐点计算内积, 从而生成特征图。卷积核是一个小的矩阵, 通过与输入数据的各部分(感受野)进行矩阵乘法, 然后累加得到一个单一的输出值。

卷积层具有两个显著特点: 一是权重共享, 即同一个卷积核在整个输入图像上进行卷积操作, 大大减少了需要训练的参数数量; 二是局部连接, 即卷积核只关注局部区域, 使得模型可以捕捉到图像的局部特征, 如边缘、角点等。在每个卷积层之后, 通常会应用一个非线性激活函数, 使模型能够处理非线性数据。ReLU(Rectified Linear Uni)是最常用的激活函数, 它将所有负值设为零, 正值保持不变, 通过引入非线性激活函数, CNN 能够学习和表达更加复杂的特征和模式。

2) 池化层用于减少特征图的尺寸, 从而降低计算量和防止过拟合。常见的池化方法有最大池化(Max Pooling)和平均池化(Average Pooling)。最大池化通过从特征图的每个子区域中选取最大值来保留主要特征, 例如, 2×2 的最大池化会将 2×2 的区域缩小为 1×1 区域, 取该区域内的最大值。在经过多层卷积和池化层后, 最终会将特征图展(Flatten), 并连接到一个或多个全连接层。全连接层类似于传统的神经网络, 节点与节点之间完全连接, 用于结合提取的特征并完成最终的分类或回归任务。

3) 在神经网络学习的过程中, 输入特征在前向传播过程中, 依次通过卷积层、激活函数、池化层和全连接层, 最后输出预测结果, 其中每层的输出作为下一层的输入, 不断传递和处理数据, 逐步提取出高级特征, 基于神经网络的最终输出与真值的差距计算损失函数, 通过计算损失函数的梯度, 反向传播

训练 CNN 模型, 更新每层的权重和偏置参数, 重复此过程直至网络的输出值的准确率达到需求[16]。

在实际使用中, 对于如何选择卷积层的层数, 卷积核的尺寸以及激活函数的选择需要根据目标任务灵活选择[17]。

2.2. 相位变换的广义互相关时延估计

目前使用深度学习方法解决问题, 对原始信号进行预处理提取重要特征依旧是非常重要的。在传统方法中, 基于到达时间差(TDOA)的声源定位方法依赖于声音信号到达不同麦克风的时间差来获取声源位置, 但这种方法易受到噪声房间混响的影响; 当对声音信号进行加权处理时, 进而突出信号峰值, 减弱混响和噪声对信号峰值的影响; 其中 GCC-PHAT 是一种常用的时延估计算法, 通过计算麦克风对之间接收信号的广义互相关函数来估计声源到达延迟, 在互相关的基础上, GCC-PHAT 引入了相位变换加权函数, 对信号的互功率谱进行归一化, 增强了高频成分, 提高了对噪声和混响的鲁棒性, 其数学原理为:

$$R_{ij}^{\text{GCC-PHAT}}(\tau) = \int_{-\infty}^{\infty} \frac{G_{ij}(f)}{|G_{ij}(f)|} e^{j2\pi f\tau} df \quad (1)$$

其中 $G_{ij}(f)$ 为麦克风 i 和 j 接收信号的互功率谱, 时延估计为在 $R_{ij}^{\text{GCC-PHAT}}(\tau)$ 取得最大值时对应的 τ [15], GCC-PHAT 具有较好抗混响能力, 所以在本文中 choice GCC-PHAT 作为预处理函数获取卷积神经网络的输入信号。

2.3. 网络模型训练

卷积神经网络结构的选择和参数的选择需要根据任务需求而确定, 因此, 为确定最佳的卷积层数和隐藏层数, 本文用控制变量实验方法, 从 2 个卷积核、2 个隐藏层的卷积神经网络结构开始, 逐步增加卷积核数和隐藏层数, 使用同等声源条件下的数据集进行训练和测试, 多次实验反复调整后确定最优性能的卷积网络结构, 包括隐藏层层数和每层的卷积核数; 在保证较高定位精度的同时, 又不会引入过多计算开销和过拟合风险。

3. 实验仿真与结果分析

3.1. 房间定位模型

本文考虑一个大小为 $6\text{ m} \times 6\text{ m} \times 6\text{ m}$ 的长方体封闭房间, 用以模拟实际的室内声源传播和定位场景。使用 RoomSim 仿真工具对室内混响环境进行仿真, 实验设置混响时间为 200 ms, 400 ms, 600 ms, 对应的墙壁吸收系数根据 Sabine 公式的逆过程计算得到。房间中心位置(3 m, 3 m, 3 m)处放置一个由 $M = 6$ 个麦克风组成的圆形阵列, 阵列半径为 0.5 m, 所有麦克风高度均为 3 m, 接收房间内的声音信号, 如图 1 所示。房间内存在一个随机位置的单点声源 s , 其位置用极坐标 (r_s, θ_s, ϕ_s) 表示, 其中 r_s 为声源到麦克风阵列中心的距离, θ_s 和 ϕ_s 分别表示俯仰角和方位角。声源位置在房间内随机生成, 但与麦克风阵列的距离被约束在 0.5 m 至 3 m 之间, 即麦克风与墙壁之间的位置。

图 2 为麦克风阵列在 x - y 平面上针对 500 Hz、1000 Hz 和 2000 Hz 三个不同频率的波束形成图, 可以看出随着频率的增加, 阵列变得更加具有方向性, 波束图中表现出更窄且更多的波瓣。在 500 Hz 时, 阵列显示出相对宽广且均匀的灵敏度, 能够从较大范围的方向捕捉声音。然而, 随着频率增加到 1000 Hz 和 2000 Hz, 波束形成器产生了更尖锐且更明显的波瓣, 表明其方向性增强, 更能够隔离特定方向的声音, 同时衰减其他方向的声音。麦克风阵列对于频率的依赖性行为表明了阵列在低频时提供广泛覆盖, 而在高频时提供聚焦灵敏度的能力, 适用于需要方向性声音捕捉的应用场景。

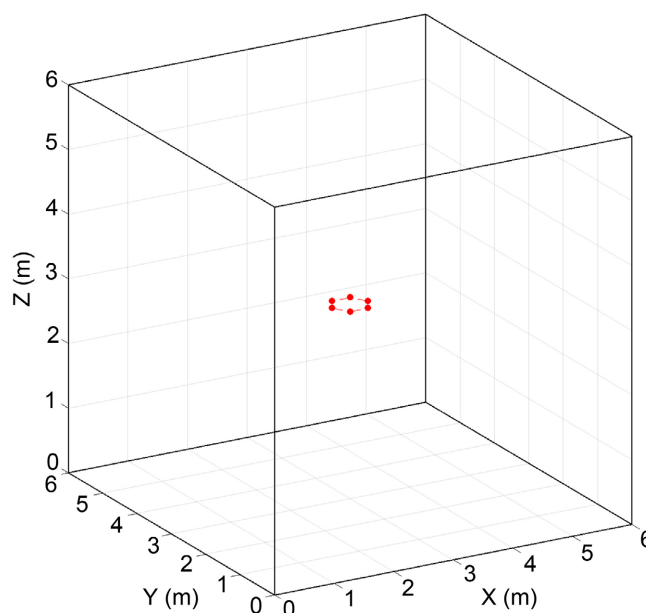


Figure 2. Room model and microphone array distribution
图 2. 房间模型和麦克风阵列分布

将总方位角度 360 度的房间以 20 度作为一个步进划分为 $K = 18$ 个方位区间, 以一个固定的 A 点作为参考点, 将每个方位区域进行唯一的类别编码为 $k = 1, 2, 3, \dots, 18$, 即卷积神经网络的输出为类别, 因此将声源定位问题转化为分类问题。令麦克风接收信号为 $x(t)$, 对于第 k 个子区域, 提取一组与声源位置相关的特征 f_k 。给定特征 f_k , 声源位于第 k 个子区域的后验概率为 $p(C = k | f_k)$, 其中 C 为子区域索引变量。声源定位问题可以表示为最大后验概率估计:

$$\hat{k} = \arg \max p(C = k | f_k), k = 1, 2, 3, \dots, K \quad (2)$$

通过卷积神经网络来估计后验概率, 然后选择后验概率最大的子区域作为声源位置的估计。在实际应用中, 子区域的大小和数量 K 可以根据需要的定位精度和计算复杂度进行调整。

3.2. 模型训练

基于卷积神经网络的声源定位方法分为训练阶段和测试阶段。声源信号采用的是 TIMIT 数据集, 随机选择其中的 6000 条语音数据, 作为房间仿真实验中的声源信号, 为了模拟实际的房间条件和声音传播, 在训练数据的预处理阶段, 我们引入了 3 种信噪比(SNR)水平的高斯白噪声: 30 dB, 20 dB 和 10 dB, 在每种 SNR 条件下, 分别在房间响应信号中加入 200, 400, 600 ms 的混响, 即选择随机的位置处的语音信号与对应位置的房间脉冲响应进行卷积, 并叠加噪声信号; 然后对麦克风接收信号进行 GCC-PHAT 计算, 实验使用 256 点 STFT, 汉明窗, 然后对相关特征信号进行分帧处理, 每两个子帧之间重叠 50%, 随即将各帧 GCC-PHAT 特征组合成一个二维特征矩阵 F , F 即为卷积神经网络的输入数据, 网络的输出即为每个仿真数据的声源位置所在区间的类别标签。在实验中, 数据集按照 8:1:1 的比例随机划分为训练集、验证集和测试集。在测试阶段, 将测试数据的特征矩阵输入到卷积神经网络中, 网络的输出为声源在各个区间内的后验概率, 选择后验概率最大的区间作为声源位置的估计结果。

3.3. 定位结果分析

我们考察了在 9 种不同的混响时间(RT60)和信噪比(SNR)组合下, 4 种算法的定位准确率, 如下表 1

所示。

Table 1. Localization accuracy of different sound source localization methods under different simulation settings
表 1. 不同声源定位方法在不同仿真设定下的定位准确率

RT60 (ms)	SNR (dB)	TDOA (%)	SRP-PHAT (%)	FCN (%)	CNN (%)
200	40	59.2	91.2	90.1	92.5
400	40	56.3	90.5	88.2	91.7
600	40	55.0	89.4	85.7	91.2
200	20	46.8	87.2	83.3	90.3
400	20	44.7	88.9	82.1	89.8
600	20	41.9	85.9	79.8	87.5
200	10	35.6	83.2	78.5	86.5
400	10	30.5	77.8	76.9	85.4
600	10	29.9	74.2	74.8	84.7

其中, 定位准确率定义为预测正确的声源区间数占总测试区间数的百分比。从实验结果可以看出, 所提出的 CNN 方法在各种环境条件下均取得了最高的定位准确率, 且优势在高混响、低 SNR 情况下更为明显。以 RT60 = 600 ms, SNR = 10 dB 为例, CNN 的准确率达到 84.7%。同时在相对理想的低混响、高 SNR 环境(RT60 = 200 s, SNR = 40 dB)下, CNN 的性能仍然保持领先。这表明, CNN 方法更准确地提取声源特征以及拟合声源特征与声源位置的关系, 更不容易被房间混响和噪声的影响。

相比之下, 传统的 TDOA 算法对噪声和混响最为敏感, 准确率随着环境恶化而急剧下降, 在强混响低信噪比条件下甚至降至 29.9%。这主要是因为该方法过于依赖理想的自由场传播假设, 难以应对实际中的多径效应和背景干扰。SRP-PHAT 虽然对混响的适应性有所改善, 但在高噪声下的表现仍不够理想。

FCN 作为一种全连接神经网络, 其输入信号为一维向量, 因此声源特征提取能力和非线性拟合能力相对有限, 导致在复杂声学环境下的泛化性能比 CNN 差。本文方法在实现更优定位性能的同时, 也不可避免地引入了一定的计算开销, 主要体现在离线训练阶段对大量数据和算力的需求。考虑到声源定位任务对精度的苛刻要求和实际声学环境的多变性, 这种性能与效率的权衡是值得的。

3.4. 卷积神经网络结构对算法准确度的影响分析

由表 1 可以看出在同一个卷积层网络结构下, 不同声音条件下的声源定位准确率不同, 因此为了适应更多的房间和声音条件, 确保尽可能保持更高的预测准确率, 本文对如何选择卷积网络结构进行研究, 即选择卷积层层数和卷积核的个数, l 层卷积层的输出特征图可表示为:

$$x(l) = f(l)(w(l) * x(l-1) + b(l)) \quad (3)$$

其中 $x(l-1)$ 为上一层输出, $w(l)$ 和 $b(l)$ 分别为本层卷积核和偏置, $*$ 表示卷积操作, $f(l)$ 为激活函数, 可见卷积层数影响特征提取的深度和抽象级别。本实验从卷积层数为 2 开始测试, 首先固定卷积核个数为 2, 通过调整最大迭代次数和目标损失, 得到不同卷积层数下预测准确值的变化情况。

如图 3 所示, 随卷积层数增加, 预测准确率呈先升后平缓趋势。进一步分析发现, 卷积层数为 6 时的 CNN 网络性能最优, 预测准确率最高, 收敛速度最快, 泛化能力最强; 当卷积层层数到达 10, 12 时, 网络会很快到达过拟合状态, 同时验证集的验证损失远远大于训练损失, 表明卷积层层数过大时无法提

高预测性能, 且是训练网络的时间消耗过大。因此, 固定卷积层层数为 6, 改变首层隐藏层的卷积核的个数进行实验, 如图 4 所示, 随着卷积核数量的增加, 预测准确率先升后降, 比如在卷积核数量为 6 时, 其性能反而小于卷积核为 4 时, 表明过多的卷积核并不会提高预测性能。进一步分析发现, 卷积核数量

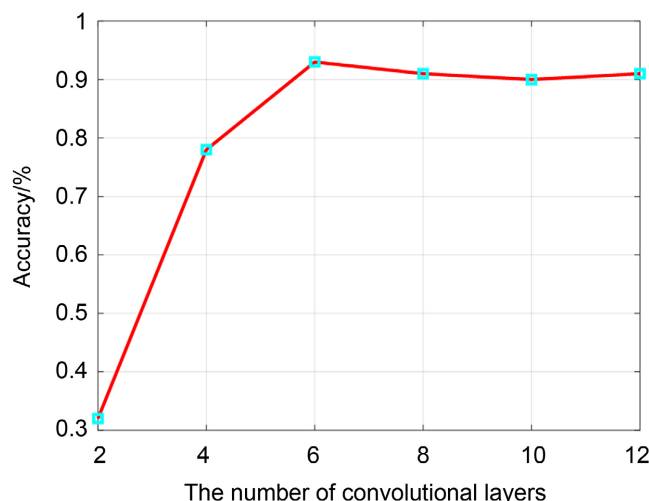


Figure 3. The relationship between the number of hidden layers and prediction accuracy

图 3. 不同隐藏层层数与预测准确率的关系

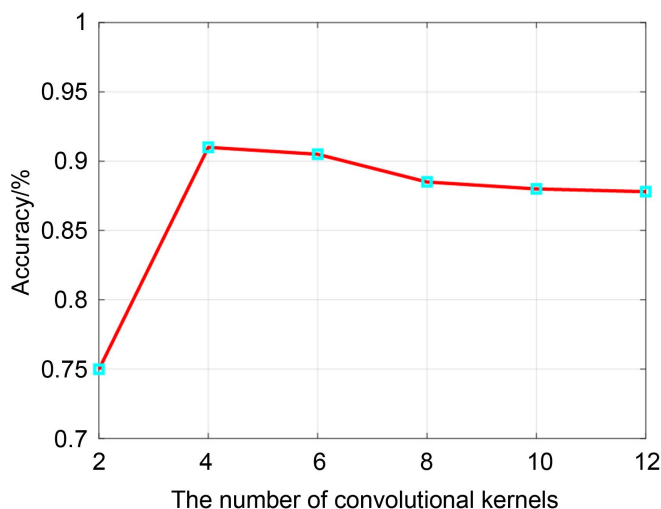


Figure 4. The relationship between the number of convolutional kernels in the first layer and prediction accuracy

图 4. 首层卷积核个数与预测准确率的关系

为 4 时, CNN 网络在定位精度和实时性之间达到了最佳平衡, 发挥卷积层的特征提取能力和隐藏层的非线性映射能力, 同时兼顾训练难度和计算开销, 为实际应用提供了有益参考, 因此上述实验中 CNN 的结构为包含 6 层卷积层, 首层卷积层的卷积核为 4, 所有卷积层的卷积核尺寸均为 3, 所有池化层均为均值池化。

4. 结论

本文针对传统麦克风阵列声源定位算法在复杂声学环境下性能显著下降的问题, 提出了一种基于卷

积神经网络(CNN)的声源定位新方法。该方法利用大量通过房间声学模拟生成的不同混响和噪声条件下的仿真数据, 提取麦克风对之间的广义互相关相位变换(GCC-PHAT)函数作为 CNN 的输入特征, 以期从信号中学习并与声源空间位置相关的判别信息。将定位问题转化为声源所在角度区间的分类任务, 充分发挥 CNN 在模式识别方面的优势。

本文还通过仿真实验系统地研究了卷积层层数, 卷积核的尺寸等参数对定位性能的影响。实验发现, 随着卷积层数的增加, 声源定位准确率呈现先升高后降低的趋势, 6 层卷积时性能最优。卷积核尺寸的增加也呈现类似规律, 首层卷积层卷积核数量为 4 时可以在精度和效率间达到最佳平衡。综合分析, 卷积层数为 6、首层卷积层卷积核数量为 4 时, CNN 网络在声源定位任务上取得了最佳表现。该结构在保证较高定位精度的同时, 训练速度快、参数量适中, 更有利于实际应用。

在 8 种不同混响时间和信噪比组合的测试条件下, 所提出的 CNN 定位方法均取得了优于传统的时间延迟估计(TDOA), SRP-PHAT 以及全连接网络(FCN)的性能, 尤其在高混响低信噪比情况下优势更加明显, 自适应地提取声学信号中与声源方位相关的判别式特征, 并对混响和噪声引入的干扰更加鲁棒。

基金项目

中国科学院应用海洋学重点实验室 2021 年度开放课题——南海中尺度涡旋及其声传播特性研究(OOST2021-04)。

参考文献

- [1] Alameda-Pineda, X. and Horaud, R. (2014) A Geometric Approach to Sound Source Localization from Time-Delay Estimates. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **22**, 1082-1095. <https://doi.org/10.1109/taslp.2014.2317989>
- [2] Yook, D., Lee, T. and Cho, Y. (2016) Fast Sound Source Localization Using Two-Level Search Space Clustering. *IEEE Transactions on Cybernetics*, **46**, 20-26. <https://doi.org/10.1109/tcyb.2015.2391252>
- [3] Li, X., Deng, Z.D., Rauchenstein, L.T. and Carlson, T.J. (2016) Contributed Review: Source-Localization Algorithms and Applications Using Time of Arrival and Time Difference of Arrival Measurements. *Review of Scientific Instruments*, **87**, Article 041502. <https://doi.org/10.1063/1.4947001>
- [4] Yang, X.Y., Liu, Y., Lian, Y.C., et al. (2020) GCC-Phat Based Acoustic Source Localization Using Incremental Learning with Adaptive Threshold. *IEEE Access*, **8**, 71702-71716.
- [5] Jekaterýnczuk, G. and Piotrowski, Z. (2023) A Survey of Sound Source Localization and Detection Methods and Their Applications. *Sensors*, **24**, Article 68. <https://doi.org/10.3390/s24010068>
- [6] Soumitro, C. and Emanuel, A.P. (2019) Multi-Speaker DOA Estimation Using Deep Convolutional Networks Trained with Noise Signals. 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, 20-23 October 2019, 64-68.
- [7] Ju, C.Y., Wu, C.L., Chen, W.J., et al. (2022) Neural Networks Based Subspace Learning Method for Direction-of-Arrival Estimation. 2022 *IEEE International Conference on Consumer Electronics-Taiwan*, Taipei, 6-8 July 2022, 1-2.
- [8] 高春艳, 赖光金, 吕晓玲, 等. 基于卷积神经网络的移动机器人声源定位方法综述[J]. 科学技术与工程, 2024, 24(7): 2617-2624.
- [9] Gao, C.Y., Lai, G.J., Lü, X.L., et al. (2024) Review of CNN-Based Methods for Mobile Robot Sound Source Localization. *Science Technology and Engineering*, **24**, 2617-2624.
- [10] 焦琛, 张涛, 孙建红. 基于卷积神经网络的室内麦克风阵列声源定位算法[J]. 激光与光电子学进展, 2020, 57(8): 187-192.
- [11] Jiao, C., Zhang, T. and Sun, J.H. (2020) Convolutional Neural Network Based Indoor Microphone Array Sound Source Localization. *Advances in Laser and Optoelectronics*, **57**, 195-200.
- [12] Kwon, B., Park, Y. and Park, Y. (2010) Analysis of the GCC-PHAT Technique for Multiple Sources. *ICCAS 2010*, Gyeonggi-do, 27-30 October 2010, 2070-2073. <https://doi.org/10.1109/iccas.2010.5670137>
- [13] Alien, J.B. and Berkley, D.A. (1976) Image Method for Efficiently Simulating Small-Room Acoustics. *The Journal of the Acoustical Society of America*, **60**, S9-S9. <https://doi.org/10.1121/1.2003643>

-
- [14] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) Imagenet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. <https://doi.org/10.1145/3065386>
 - [15] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015 *International Conference on Learning Representations*, San Diego, 7-9 May 2015, 1-4.
 - [16] Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>
 - [17] Zhang, Y.M., Guo, Q.L., Lu, Y., *et al.* (2023) Nonlinear Weighted GCC-Phat for Acoustic Source Localization. *IEEE Signal Processing Letters*, **30**, 1-5.