引用格式：宋昊, 刘雪洁, 俞胜锋, 等. 基于深度学习的双耳声源定位算法研究[J]. 声学技术, 2022, **41**(4): 602-607. [SONG Hao, LIU Xuejie, YU Shengfeng, et al. Binaural localization algorithm based on deep learning[J]. Technical Acoustics, 2022, **41**(4): 602-607.] **DOI**: 10.16300/j.cnki.1000-3630.2022.04.018

基于深度学习的双耳声源定位算法研究

宋 昊[1]，刘雪洁[2]，俞胜锋[3]，钟小丽[3]

(1. 广东工业大学管理学院，广东广州 510000；2. 华南师范大学物理与电信工程学院，广东广州 510006；
3. 华南理工大学物理与光电学院，广东广州 510640)

摘要：针对多种定位因素存在复杂关联且不易准确提取的问题，提出了以完整双耳声信号作为输入的、基于深度学习的双耳声源定位算法。首先，分别采用深层全连接后向传播神经网络(Deep Back Propagation Neural Network，D-BPNN)和卷积神经网络(Convolutional Neural Network， CNN)实现深度学习框架；然后，分别以水平面15°、30°和45°空间角度间隔的双耳声信号进行模型训练；最后，采用前后混乱率、定位准确率与训练时长等指标进行算法有效性分析。模型预测结果表明，CNN模型的前后混乱率远低于D-BPNN；D-BPNN模型的定位准确率能够达到87%以上，而CNN模型的定位准确率能够达到98%左右；在相同实验条件下，CNN模型的训练时长大于D-BPNN，且随着水平面角度间隔的减小，两者训练时长之间的差异愈发显著。

关键词：双耳声源定位；深度学习；卷积神经网络

中图分类号：O429 **文献标志码：**A **文章编号**：1000-3630(2022)-04-0602-06

Binaural localization algorithm based on deep learning

SONG Hao[1], LIU Xuejie[2], YU Shengfeng[3], ZHONG Xiaoli[3]

(1. *School of Management, Guangdong University of Technology, Guangzhou* 510000*, Guangdong, China*;
2. *School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou* 510006*, Guangdong, China*;
3. *School of Physics and Optoelectronics, South China University of Technology, Guangzhou* 510640*, Guangdong, China*)

Abstract: Due to existence of complicated relationships between multiple localization cues, which causes them hard to be extracted accurately, a deep learning-based binaural sound source localization algorithm with complete binaural sound signals as input is proposed. Firstly, the deep fully connected back propagation neural network (D-BPNN) and the convolutional neural network (CNN) are used to implement the deep learning framework respectively. And then, binaural sound source signals with uniform azimuthal spacing of 15°, 30° and 45° in horizontal plane are applied to model training respectively. Finally, indicators such as front-back confusion rate, localization accuracy and training duration are used to investigate effectiveness of the models. The model prediction results show that the front-back confusion rate of the CNN model is much lower than that of D-BPNN model. The localization accuracy of the D-BPNN model can reach more than 87%, while the localization accuracy of the CNN model is about 98%. Under the same experimental conditions, the training time of CNN model is longer than that of D-BPNN model; Moreover, this difference in training time becomes more and more obviously as the azimuthal spacing in the horizontal plane decreases.

Key words: binaural localization algorithm; deep learning; convolutional neural network (CNN)

0 引 言

在双耳听觉中，人类能够通过接收到的双耳声信号反推出声源的空间方位，即实现双耳声源定位。研究表明，双耳声源定位的主要因素包括耳间差异：包括双耳时间差(Interaural Time Difference, ITD)、双耳声级差(Interaural Level Difference, ILD)和单耳谱特征[1-2]。通常，ITD是低频声源的主要定位因素，ILD是中、高频声源的主要定位因素，而单耳谱特征对于中垂面以及混乱锥定位至关重要。由于现实声场景的复杂性，准确的声源定位往往是多种定位因素综合作用的结果[3]。

已有的双耳声源定位模型分为两大类：基于听觉系统的模型和基于机器学习的模型[4]。前者通过较为详尽地再现声信号传输和分析生理和心理过

收稿日期： 2021-03-01； **修回日期：** 2021-05-04
基金项目： 广东省自然科学基金项目(2021A1515011871，2021A1515012630)
作者简介： 宋昊(2000—)，男，广东广州人，研究方向为信息技术和声信号处理。
通信作者： 钟小丽, E-mail: xlzhong@scut.edu.cn

程，从而达到模拟人类声源定位功能的目的。然而，受限于声源定位的生理和心理过程的研究进展，目前基于听觉系统的双耳声源定位模型只能表征较为简单的声场景，例如基于谱因素的中垂面定位[4-5]。本质上，双耳声源定位是一种基本的大脑机能，因此基于机器学习(即采用计算机模拟人脑行为)的双耳声源定位模型得到了重视[6]。Gill 等[7]将单耳谱特征输入单隐层(含 9 个神经元)前馈型神经网络，以预测声源的仰角方位。Chung 等[8]以 ITD 和单耳谱特征作为输入，采用浅层全连接后向传播神经网络(Back Propagation Neural Network, BPNN)预测声源的空间方位。Jin 等[9]先采用耳蜗模型提取双耳定位因素，再采用浅层时延神经网络预测声源的空间方位。可见，已有基于神经网络的定位模型通常以定位因素作为模型输入进行训练和预测。由于声源定位是多种定位因素综合作用的结果，且不同声场景下这种综合作用可能不同，目前对其尚无定论。因此，以定位因素作为模型输入的定位模型需要较完备的先验知识，且模型适用性取决于定位因素的选取。此外，已有基于神经网络的定位模型属于全连接的浅度学习(即只包含一个或者两个隐层)，这制约着预测效果的提升。随着计算机计算能力以及数据量的提升，同时受益于神经生物学家对动物大脑解剖研究的成果，2006 年 Hinton 等[10]提出了深度学习的概念。深度学习源于机器学习的范畴。相比于机器学习中各种浅层的学习模型，比如支持向量机、最大熵方法等，深度学习神经网络能表达现实中各种复杂数据的内部结构，已成为一种广泛使用的工程方法[11-12]。2019 年丁建策等[13]提出了利用深度神经网络(Deep Neural Network, DNN)预测声源方位角的算法。该方法采用子带双耳特征和双耳信号互相关特征共计 57 维特征作为输入，取得了较好的预测效果，为后续声源距离的预测提供了可靠的信息。Ding 等[14]进一步提出了多目标 DNN 算法。该方法提取了双耳信号中的子带特征和统计特性共计 393 维特征作为输入，可同时预测声源的距离和方位角；由于新算法降低了方位角变化对距离估计的影响，其距离预测的准确性高于现有的同类算法。

本文提出了基于深度学习的双耳声源定位算法，并采用完整的双耳声信号作为输入，避免了人为提取特征的繁琐过程。首先，实现了基于卷积神经网络(Convolutional Neural Network, CNN)和基于深层后向传播神经网络(Deep Back Propagation Neural Network, D-BPNN)的深度学习框架，并采用不同空间声源间隔的双耳声信号作为输入进行训练与预测，最后采用前后混乱率、定位准确率等指标比较了两种深度学习模型的有效性。

1　算法介绍

本文的声信号处理流程如图 1 所示。首先，将单通道声信号 $E_0(t)$ 分别与左右耳脉冲响应 H_L 和 H_R 进行卷积，合成双耳声信号 E_L 和 E_R；然后，将预处理后的 E_L 和 E_R 输入深度神经网络进行训练；最后，采用训练好的算法模型进行预测，得到声源空间方位的分类输出(即方位预测)。

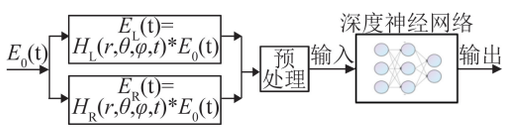


图 1　声信号处理流程图

Fig.1　Acoustic signal processing flow diagram

卷积神经网络 CNN 是一种典型的深度学习框架。目前，CNN 已被广泛应用于声信号处理，例如遇险信号识别[15]、混响时间估计[16]、水下声源距离预测[17]和海床类型识别[18]等。图 1 中的深度神经网络主要采用 CNN 实现。在深度学习领域，卷积神经网络 CNN 和循环神经网络(Recurrent Neural Network, RNN)都是常用的处理音频信号的算法。两者的区别在于：(1) RNN 具有对时间进行扩展以及多个时间输出计算的能力，而 CNN 能够在空间上拓展并对特征进行卷积；(2) RNN 可以用于描述时间上连续状态的输出(具备记忆功能)，而 CNN 用于静态输出；(3) RNN 的层次结构深度有限，而 CNN 的层数能够达到 100 层以上。本文的研究目标是实现对处于不同空间方位的声源的准确定位(分类)，而不考虑声信号在时间上的连续特征。因此，本文选用 CNN 作为实现深度学习的框架。此外，目前采用全连接后向传播神经网络 BPNN 的定位模型多为浅层网络。为了和 CNN 模型进行对比，本文通过增加隐层的层数，将浅层 BPNN 改进为 D-BPNN。因此，图 1 中的深度神经网络分别采用 CNN 和 D-BPNN 实现。

1.1　数据准备

以人头中心为坐标原点，建立顺时针球坐标系。定义正前方的水平方位角为 0°，正右方的水平方位角为 90°。

采用虚拟声技术合成双耳声信号数据库，其中双耳脉冲响应来自 MIT HRTF 数据库[19]。该数据库含有 KEMAR 人工头远场(距离声源 1.4 m)710 个声源空间方位的双耳脉冲响应，数据长度为 512 点

(44.1 kHz采样, 16 bit量化)。单通道声信号采用中英文混合的单声道语音信号(频段范围为150 Hz~4 000 Hz，采样频率为44.1 kHz)。首先，采用短时过零率语音端点检测算法对初始语音信号进行检测，依据检测结果将其分别截断成8 300段(100 ms每段)短时语言片段。然后，依据约10∶1的比例随机将各方位的8 300段短时语言片段划分为训练信号与测试信号。最后，根据图1中的虚拟声合成方式得到不同声源方位的双耳时域声信号。为了加快训练的收敛速度，对所有合成的双通路信号进行归一化运算，将其幅值限制在[−1，1]范围内。

为了探讨不同声源空间间隔的影响，图1中的网络输入分别采用水平面15°、30°和45°空间角度间隔的双耳声信号。表1是不同空间角度间隔时，深度神经网络所采用的数据情况。

表1 深度神经网络采用的数据
Table1 Data used for DNN

空间角度间隔/(°)	所需分类的空间方位数目	训练集数据量	测试集数据量
15	24	180 169	19 031
30	12	90 467	9 133
45	8	60 226	6 174

1.2 基于D-BPNN模型的双耳声源定位算法

D-BPNN模型主要由两个全连接层、一个随机失活层和一个扁平层组成，网络结构如图2所示。

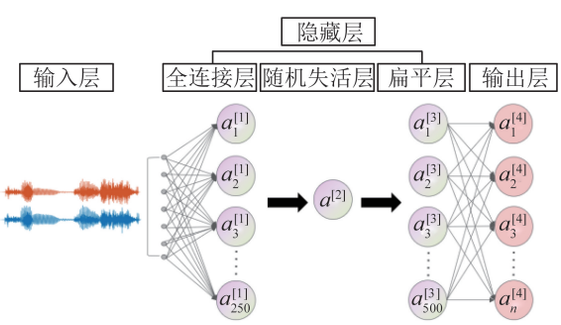


图2 深层全连接后向传播神经网络结构图
Fig.2 The structure of D-BPNN

输入信号首先进入一个含250个神经元的全连接层$a^{[1]}$，随后经过丢弃概率为0.3的随机失活层$a^{[2]}$，最后经过扁平层$a^{[3]}$后进入输出层$a^{[4]}$输出，图2中输出层神经元个数n取决于所需分类的空间方位数目(见表1)。其中，第一个全连接层的激活函数为线性整流函数(Rectified Linear Units, ReLU)。与其他激活函数相比，ReLU激活函数具有提升网络训练速度、防止梯度消失及增加网络非线性能力的优点。最后一个全连接层采用Softmax函数输出，将多个输出映射到[0，1]区间内，从而实现空间方位的多分类任务。

采用Adam优化算法并使用交叉熵损失函数进行网络训练。交叉熵损失函数定义为

$$F=-\sum_{k=1}^{m}[y_{k1}\lg(\hat{y}_{k1})+\ldots+y_{kn}\lg(\hat{y}_{kn})] \tag{1}$$

其中：m代表样本数，即训练集数据的总数目；n代表输出分类数，即所需分类的空间方位数目；$\hat{y}_{kn}$代表第k个样本预测为第n个分类的概率。

1.3 基于CNN模型的双耳声源定位算法

CNN模型主要由三个卷积层和四个全连接层组成，网络结构如图3所示。需要说明的是，卷积层中的滤波器皆以一维卷积的形式在两个输入通道上分别做卷积，同时所有激活函数均为ReLU函数。此外，在激活函数层、卷积层、全连接层之间都加入归一化层，实现了在神经网络层的中间进行预处理的操作。

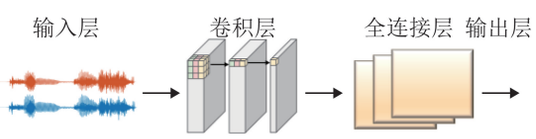


图3 卷积神经网络结构图
Fig.3 The structure of Convolutional Neural Network

输入信号首先经过连续三个卷积层进行特征提取。其中第一个卷积层共有128个滤波器(卷积核)，维度为1×300，步长为30；第二个卷积层共有128个滤波器，卷积核的大小为1×40，步长为2；第三个卷积层共有64个滤波器，卷积核的大小为1×20，步长为2。输入信号经过三个卷积层后，将特征向量再输入四个全连接层。其中，前三个全连接层的神经元个数分别为2 048、1 024、128，最后一个全连接层为输出层，其神经元个数取决于所需分类的空间方位数目(见表1)。最后采用Softmax函数计算出输入数据属于每个类别的概率值，并选取概率值最大的类别作为预测方位。

此外，为了在训练时抑制过拟合，提高网络的泛化能力，全连接层中均使用Dropout方法；同时，网络在训练时采用Adam优化算法与交叉熵损失函数。

CNN模型仿真算法采用PyTorch框架实现。PyTorch框架是目前最为流行的深度学习框架之一，基于Torch框架，广泛用于自然语言处理等领域，拥有极强的易用性与灵活性。D-BPNN模型仿真算法采用Keras框架实现。Keras框架是一种高层神经网络应用程序编程接口，使用TensorFlow、Theano及CNTK作为后端，具有拓展性强的优点。上述两种模型的仿真实验均采用相同的数据集以及硬件仿

真环境。硬件环境包括：Intel(R) Core(TM) i7-10750H CPU@ 2.60GHz 2.59GHz 处理器以及 NVIDIA Quadro T2000 显卡。

一共进行了 6 组仿真实验，即 2 种算法模型(D-BPNN 模型和 CNN 模型)，每种执行 3 种空间角度间隔(15°、30°和 45°)。对 D-BPNN 模型和 CNN 模型分别进行了 50 次和 20 次迭代训练后，观察到模型训练准确率达到稳定或围绕某一中值上下轻微波动，此时判定网络训练成功。图 4 是 D-BPNN 模型训练准确率随迭代次数的变化。由图 4 中可知，对于任何一种空间角度间隔的输入，经过 50 次迭代训练后训练准确率都趋于平稳；此时，3 种空间角度间隔(15°、30°和 45°)的训练准确率分别达到 73.59%、77.18%、81.76%。

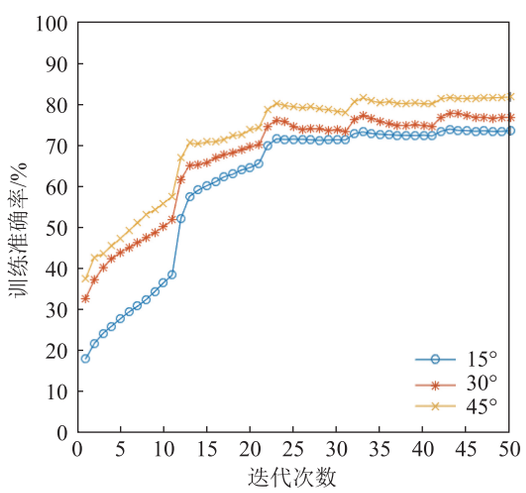


图 4　深层全连接后向传播神经网络的训练准确率
Fig.4　Training accuracy of D-BPNN

图 5 是 CNN 模型训练准确率随迭代次数的变化。图 5 中可见，对于任何一种空间角度间隔的输入，经过 20 次迭代训练后训练准确率都趋于平稳；此时，3 种空间角度间隔(15°、30°和 45°)的训练准确率分别达到 96.07%、98.22%、98.93%。

2　实验结果和讨论

如表 1 所示，当网络输入的空间角度间隔为 15°、30°和 45°时，测试信号分别为 19 031、9 133 和 6 174 个。

2.1　模型定位效果

在人类听觉中，由于前后镜像方位(例如方位角 θ=30°和 θ'=150°)具有相似的 ITD 和 ILD，因此容易出现前后混淆现象，即处于 θ=30°的声源被感知处于 θ'=150°，反之亦然。在听觉定位主观实验中，如果被试出现了前后混淆现象，通常是先校正混

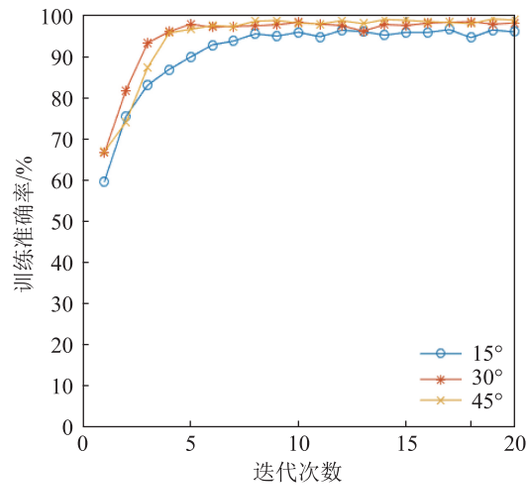


图 5　卷积神经网络的训练准确率
Fig.5　Training accuracy of CNN

淆，即将发生混淆的方位进行空间的镜像反演，然后再进行定位准确率计算[20]。假设某个空间方位角 θ 共有 N 个测试信号，经模型预测后，有 X_1 个测试信号的空间方位角预测为 θ，即方位角预测正确；有 X_2 个测试样本的空间方位角预测为 θ'(θ 与 θ' 互为镜像方位)，即出现前后混乱。那么，方向 θ 的前后混乱率 R 和定位准确率 P 分别为

$$R(\theta)=\frac{X_2}{N} \tag{2}$$

$$P(\theta)=\frac{X_1+X_2}{N} \tag{3}$$

在每一组实验条件下，对所有测试方位的前后混乱率和定位准确率取平均，得到该实验条件下的平均前后混乱率和平均定位准确率，如表 2 所示。可以看出 CNN 模型的平均前后混乱率远低于 D-BPNN 模型，其中前者的前后混乱率均低于 2.24%。这表明，对于前后镜像方位，当耳间差异(ITD 和 ILD)无法为定位提供有效信息时，CNN 模型比 D-BPNN 模型能更好地通过自学习提取单耳谱特征，从而可以更好地区分前后镜像方位。

为测试不同信噪比情况下的定位效果，在 1.1 节中生成的双耳信号中加入不同信噪比的高斯白噪声。合成的双耳带噪信号的信噪比分别为 0 dB、10 dB、20 dB，分别采用 CNN 模型和 D-BPNN 模型进行方位角预测。结果表明，CNN 模型的定位准确率分别为 34.68%、76.92%、97.36%；而 D-BPNN 模型的定位准确率分别为 32.85%、66.29%、85.71%。可见，在训练与测试环境不匹配的情况下，两种模型的预测准确率均有下降，但是 CNN 模型的鲁棒性优于 D-BPNN 模型。

2.2　模型训练用时

为了进一步进行模型比对，在相同实验环境下

对模型的训练时长进行了测算，结果如表2所示。由表2中可见，无论是D-BPNN模型还是CNN模型，随着输入空间角度间隔的减小，训练时长都呈现上升趋势；对于相同的空间角度间隔，CNN模型的训练时长高于D-BPNN模型。进一步的计算结果表明，两者训练时长的差异随着空间角度间隔的减小而增大，例如随着空间角度间隔从45°变为15°，CNN模型训练时长高于D-BPNN模型时长的比例从7.34%增加到33.36%。

表2 基于D-BPNN和CNN的双耳声源定位算法的结果
Table 2 Results of D-BPNN and CNN based binaural localization algorithms

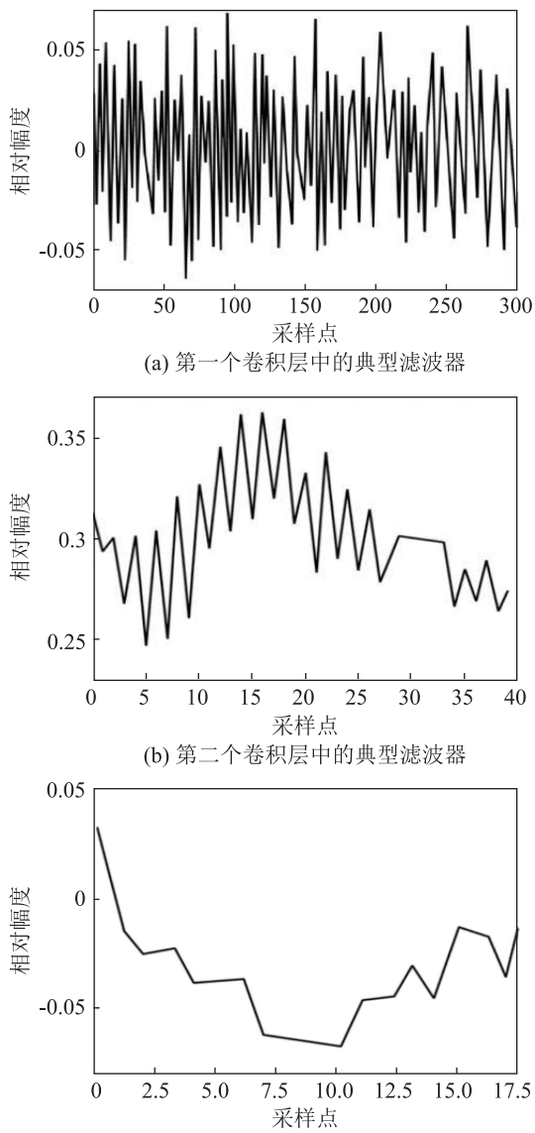
模型名称	空间角度间隔/(°)	平均前后混乱率/%	平均定位准确率/%	训练时长/s
D-BPNN	15	21.35	93.29	2 262.93
	30	28.00	89.27	1 114.51
	45	27.53	86.88	740.63
CNN	15	2.24	97.75	3 395.60
	30	0.94	98.89	1 210.26
	45	1.04	99.51	799.33

2.3 CNN模型滤波器的分析

通过前后混乱率和定位准确率指标的对比研究发现，在同等实验条件下CNN模型的预测效果优于D-BPNN模型。为了进一步探究这一现象的内部机制，对空间角度间隔为15°的输入信号CNN模型中的三个卷积层的典型滤波器进行了展示，如图6所示。限于篇幅，各层中仅选取一个典型滤波器进行分析。

滤波器代表CNN模型中对应卷积核的参数权重，用于提取相应的深度神经网络内部特征。在训练效果良好的模型中，滤波器图形往往展现出平滑的滤波特性；卷积核的参数权重在第一个卷积层中体现出可解释性，但是这种可解释性随着层次的加深而逐渐消失。

图6(a)中显示出类似于语音信号的波形，这是CNN模型对输入信号进行特征提取的过程，展现了第一个卷积层对其卷积核参数权重的可解释性。这类似于双耳听觉定位时，人类的神经系统自动根据双耳接收信号对各类参数差异进行判断。图6(b)、6(c)分别为第二层、第三层的典型滤波器的结果。与图6(a)相比，图6(b)的特征趋于抽象，主要表现为密集的波动。图6(c)的特征则是在图6(b)特征上的进一步提取和抽象，主要表现为平缓的波动。可见，随着网络卷积层次的加深，更加抽象、基本的类别信息将被抽取；同时，对卷积核参数权重的可解释性也逐渐降低。



(a) 第一个卷积层中的典型滤波器
(b) 第二个卷积层中的典型滤波器
(c) 第三个卷积层中的典型滤波器

图6 卷积神经网络的典型滤波器
Fig.6 Typical filters in CNN

3 结 论

本文针对多种双耳定位因素存在复杂关联的问题，提出了两种基于深度学习的算法模型：深层全连接后向传播神经网络D-BPNN模型和卷积神经网络CNN模型；并采用前后混乱率、定位准确率、训练时长等指标，比较了两种深度学习模型在不同空间角度间隔情况下的仿真效果。

实验结果表明：随着输入信号的空间角度间隔的减小，D-BPNN模型的定位准确率逐渐递增；而CNN模型的定位准确率趋于稳定，达到98%左右。当D-BPNN或CNN训练信号空间角度间隔减小时，训练时长呈现上升趋势；同一空间角度间隔时，

CNN 模型的训练时长均高于 D-BPNN 模型；随着水平面空间角度间隔从 45°变为 15°，CNN 模型时长高于 D-BPNN 模型时长的比例从 7.34% 增加到 33.36%。

综上所述，在相同的实验条件下，虽然卷积神经网络在对双耳听觉声源定位算法中比 BP 神经网络需要耗费相对更多的训练时长，但其拥有更高的定位准确率、更强的泛化能力与更低的前后混乱率。实际应用时，可根据用时和精度的具体需求进行算法选择。

参 考 文 献

[1] BLAUERT J. Spatial hearing: the psychophysics of human sound localization[M]. Cambridge: MIT Press, 1997.

[2] MIDDLEBROOKS J C. Sound localization[J]. Handbook of Clinical Neurology, 2015, **129**: 99-116.

[3] ALTMANN C F, UEDA R, BUCHER B, et al. Trading of dynamic interaural time and level difference cues and its effect on the auditory motion-onset response measured with electroencephalography[J]. NeuroImage, 2017, **159**: 185-194.

[4] BAUMGARTNER R, MAJDAK P, LABACK B. Modeling sound-source localization in sagittal planes for human listeners [J]. The Journal of the Acoustical Society of America, 2014, **136**(2): 791-802.

[5] LANGENDIJK E H A, BRONKHORST A W. Contribution of spectral cues to human sound localization[J]. The Journal of the Acoustical Society of America, 2002, **112**(4): 1583-1596.

[6] BIANCO M J, GERSTOFT P, TRAER J, et al. Machine learning in acoustics: theory and applications[J]. The Journal of the Acoustical Society of America, 2019, **146**(5): 3590.

[7] GILL D, TROYANSKY L, NELKEN I. Auditory localization using direction-dependent spectral information[J]. Neurocomputing, 2000, **32-33**: 767-773.

[8] CHUNG W, CARLILE S, LEONG P. A performance adequate computational model for auditory localization[J]. The Journal of the Acoustical Society of America, 2000, **107**(1): 432-445.

[9] JIN C, SCHENKEL M, CARLILE S. Neural system identification model of human sound localization[J]. The Journal of the Acoustical Society of America, 2000, **108**(3 Pt 1): 1215-1235.

[10] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, **18**(7): 1527-1554.

[11] 倪俊帅, 赵梅, 胡长青 . 基于深度学习的舰船辐射噪声多特征融合分类[J]. 声学技术, 2020, **39**(3): 366-371.
NI Junshuai, ZHAO Mei, HU Changqing. Multi-feature fusion classification of ship radiated noise based on deep learning[J]. Technical Acoustics, 2020, **39**(3): 366-371.

[12] 罗笑雪,柯雨璇,郑成诗,等 . 联合谱和空间特征的深度学习语音增强研究[J]. 声学技术, 2019, **38**(5): 467-468.
LUO Xiaoxue, KE Yuxua, ZHENG Chengshi, et al. Deep learning-based speech enhancement using both spectral and spatial features[J]. Technical Acoustics, 2019, **38**(5): 467-468.

[13] 丁建策, 厉剑, 彭任华, 等 . 室内两步法监督式学习双耳声源距离估计[J]. 声学学报, 2019, **44**(4): 405-416.
DING Jiance, LI Jian, PENG Renhua, et al. Two-stage supervised binaural distance estimation in room environments[J]. Acta Acustica, 2019, **44**(4): 405-416.

[14] DING J C, KE Y X, CHENG L J, et al. Joint estimation of binaural distance and azimuth by exploiting deep neural networks[J]. The Journal of the Acoustical Society of America, 2020, **147**(4): 2625.

[15] GAVIRIA J F, ESCALANTE-PEREZ A, CASTIBLANCO J C, et al. Deep learning-based portable device for audio distress signal recognition in urban areas[J]. Applied Sciences, 2020, **10**(21): 7448.

[16] GAMPER H, TASHEV I J. Blind reverberation time estimation using a convolutional neural network[C]//2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC). Tokyo, Japan. IEEE, 2018: 136-140.

[17] CHEN R, SCHMIDT H. Model-based convolutional neural network approach to underwater source-range estimation[J]. The Journal of the Acoustical Society of America, 2021, **149**(1): 405-420.

[18] NEILSEN T B, ESCOBAR-AMADO C D, ACREE M C, et al. Learning location and seabed type from a moving mid-frequency source[J]. The Journal of the Acoustical Society of America, 2021, **149**(1): 692.

[19] GARDNER W G, MARTIN K D. HRTF measurements of a KEMAR[J]. The Journal of the Acoustical Society of America, 1995, **97**(6): 3907-3908.

[20] 谢菠荪 . 头相关传输函数与虚拟听觉[M]. 北京: 国防工业出版社, 2008.