doi:10.6041/j.issn.1000-1298.2022.05.026

基于深度神经网络的猪咳嗽声识别方法

沈明霞[1,2]　王梦雨[1,2]　刘龙申[1,2]　陈　佳[1,2]　太　猛[1,2]　张　伟[1,2]
(1.南京农业大学工学院,南京 210031;2.江苏智慧牧业装备科技创新中心,南京 210031)

摘要:猪只呼吸道疾病易传染,影响猪的养殖生产效率,咳嗽是呼吸道疾病的显著症状之一,为识别猪只咳嗽声,提出了一种基于深度神经网络的识别方法。对声音信号进行谱减法去噪和双门限端点检测后分别提取梅山猪咳嗽及喷嚏、鸣叫、呼噜声的滤波器组(Log_filter bank, logFBank)和梅尔频率倒谱系数(Mel frequency cepstral coefficents, MFCC)特征,每种特征与其一阶及二阶差分组合作为卷积神经网络(Convolutional neural networks, CNNs)和深层前馈序列记忆神经网络(Deep feed forward sequential memory networks, DFSMN)咳嗽声识别模型的输入,进行多分类训练。对比不同特征提取方法及不同迭代次数对模型效果的影响,实验结果表明,以 MFCC 作为特征输入的 CNNs 模型效果较优,测试集上咳嗽声识别精确率为 97%,召回率为 96%,F1 值为 98%,总体识别准确率为 96.71%。表明该模型有效可行,可为生猪福利养殖中猪咳嗽声识别提供技术支持。

关键词:梅山猪;咳嗽声识别;滤波器组;梅尔频率倒谱系数;深度神经网络
中图分类号:TP391.4;S828　　**文献标识码**:A　　**文章编号**:1000-1298(2022)05-0257-10　　**OSID**:

Recognition Method of Pig Cough Based on Deep Neural Network

SHEN Mingxia[1,2]　WANG Mengyu[1,2]　LIU Longshen[1,2]　CHEN Jia[1,2]　TAI Meng[1,2]　ZHANG Wei[1,2]
(1. *College of Engineering*, *Nanjing Agricultural University*, *Nanjing* 210031, *China*
2. *Jiangsu Smart Animal Husbandry Equipment Technology Innovation Center*, *Nanjing* 210031, *China*)

Abstract: Respiratory diseases of pigs are easily contagious, which affects pig breeding efficiency. Cough is one of the obvious symptoms of respiratory diseases. An algorithm based on deep neural network was proposed to accurately identify pig coughs. Log_filter bank (logFBank) and Mel frequency cepstral coefficents (MFCC) were extracted respectively after spectral subtraction denoising and double threshold endpoint detection of the sound signal. Then the two kinds of extracted features and their first and second order differences were used as inputs to the convolutional neural networks (CNNs) and the deep feed forward sequence memory neural networks (DFSMN) for multi-classification training. The effects of the different features and different iteration times on the effectiveness of the model were compared. Except the accuracy of cough recognition, the recognition effects of other pig sounds, such as sneezing, which was easily confused with cough were also analyzed. The experimental resulst showed that when the number of training rounds reached 200, the CNNs model with MFCC as feature had a good effect. The recognition precision of cough on test set was 97%, the cough recognition recall rate was 96%, the F1-score was 98%, and accuracy reached 96.71%. It was showed that the model was effective and feasible, and can provide technical support for pig cough recognition in pig welfare breeding.

Key words: Meishan pigs; cough recognition; Log_filter bank; Mel frequency cepstral coefficents; deep neural network

0　引言

养猪业是畜牧业的重要组成部分[1],猪养殖生产中,对猪呼吸道疾病诊疗是主要工作基础[2-3]。咳嗽是猪呼吸道疾病的主要早期症状[4],因此通过识别咳嗽声进行猪早期呼吸道疾病预警成为当前关

收稿日期:2021-05-14　修回日期:2021-06-22
基金项目:江苏省重点研发计划(现代农业)重点项目(BE2019382)和政府间国际科技创新合作重点专项(2017YFE0114400)
作者简介:沈明霞(1964—),女,教授,博士生导师,主要从事机器视觉和信息农业研究,E-mail:mingxia@njau.edu.cn

注点[5]。

国内外科研人员在识别猪只咳嗽声方面的研究主要分为 3 类:一是基于一般声学分析算法的咳嗽声识别,GUARINO 等[6]使用滤波器组结合振幅解调进行特征提取,通过动态时间规整算法(DTW)将这些特征向量与参考集进行比较,咳嗽声识别正确率为 85.5%。EXADAKTYLOS 等[7]利用功率谱密度(PSD)研究病猪咳嗽声的频率特征,利用模糊 C 均值聚类算法对咳嗽和其它声音分类,咳嗽声识别正确率为 82%。徐亚妮等[8]提取梅山母猪声音信号的功率谱密度作为特征,提出了模糊 C 均值聚类算法的改进方法,对梅山猪的咳嗽声与尖叫声进行聚类分析,取到了较好的分类效果。二是基于传统机器学习的咳嗽声识别算法研究,原始的信号输入往往只经过一层处理,模型结构简单、易于学习,但对于较为复杂的信号处理能力具有局限性[9-12]。文献[13-15]分别构建基于不同特征提取(LPCC、MFCC)的隐马尔可夫(HMM)模型和矢量量化(VQ)模型对猪咳嗽声进行识别,都取得了一定成果。三是基于深度学习技术的猪咳嗽声识别,深层神经网络被广泛应用于医学影像分析、农作物识别、人脸识别等各个领域[16-19]。YIN 等[20]提出了一种基于 AlexNet 模型和谱图特征的分类算法,总体识别准确率达到 96.8%,F1 值达到 96.2%。SHEN 等[2]通过将 MFCC 与多层 CNN 融合得到 MFCC-CNN 特征,总体识别准确率为 96.6%,使用 softmax 和线性支持向量机(SVM)分类器进行分类后,咳嗽识别准确率分别提高了 7.21% 和 3.86%,F1 值分别提高了 10.37% 和 5.21%。黎煊等[1]提出了基于双向长短时记忆网络-连接时序分类的声学模型(BLSVM-CTC),利用五折交叉验证试验,猪咳嗽声识别率为 92.40%,误识别率为 3.55%,总识别率为 93.77%。

现有研究成果表明将声音信号转换成声谱图,利用神经网络进行识别,可有效提高识别精度[21-23]。本文通过谱减法对截取的猪只有效声音段进行去噪,从而消除环境中的干扰噪声,利用将声音信号的短时能量和短时过零率作为特征参量的双门限端点检测法进行猪只有效声音信号获取,分别提取猪只咳嗽、鸣叫、喷嚏以及呼噜声的 logFBank 以及 MFCC 两种声音特征后,输入 CNNs 和 DFSMN 两类神经网络模型进行猪咳嗽声识别训练,比较两种特征提取方法对模型效果的影响,以及不同迭代次数的影响,最后进行 4 种声音信号的四分类模型训练。

1 数据采集与预处理

1.1 实验对象

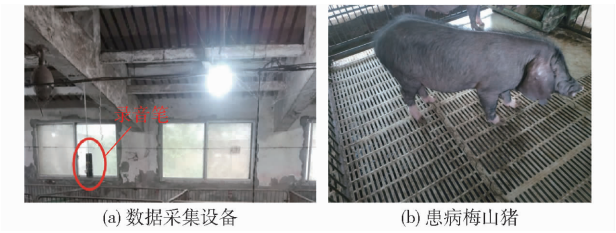
实验数据采集自江苏省镇江市句容江苏农博园梅山猪育种中心,采集时间为 2018 年 6 月,采集对象为 10 只 2—3 月龄、体质量(50±5)kg 的梅山猪,其中 5 只体温升高至 40.5℃,表现为连续性咳嗽,呼吸增快,经专业饲养人员诊断为患病。另外 5 只为健康猪只,无咳嗽症状。

1.2 实验设备

录音设备为韩国现代数字 E66 型智能录音笔,最大内存 16 GB,音频储存格式为 wav,采样精度为 16 位,采样频率为 48 kHz,每 2 h 保存为一个录音文件,以年-月-日-时-分-秒格式命名。

1.3 数据采集

分别将患病猪只及健康猪只放置在 2 个相邻猪舍内,将 2 支录音笔分别悬挂于猪舍内距离地面高度约 1.6 m 处,共采集到有效录音数据 57 h。图 1 为实验数据采集现场图。



(a) 数据采集设备 (b) 患病梅山猪

图 1 实验数据采集现场图

Fig. 1 Pig house of experimental data collection

1.4 数据预处理与数据集划分

本次实验中猪舍采集到的音频包含猪只的咳嗽、鸣叫、喷嚏、呼噜声等可用数据以及人声、金属器皿碰撞等其它环境噪声,利用音频处理软件 Adobe Audition(AU)进行有声段截取,共截取有效声音(咳嗽声、呼噜声、鸣叫声及喷嚏声)536 段。

谱减法是声音信号处理过程中常用的去噪方法之一,将声音信号类比为目标信号与噪声信号的叠加,获取声音信号的前几帧或静默段作为噪声信号,获取平均噪声能量,实现声音信号的去噪处理。图 2 为谱减法处理前后时长为 10 s 的连续咳嗽声降噪效果波形对比图,由图 2b 可知,猪连续声音信号噪声得到明显削减,并通过人耳试听进一步验证,猪声音样本没有失真。

利用将声音信号的短时能量和短时过零率作为特征参量的双门限端点检测法对去噪后的 536 段录音信号进行端点检测,共得到咳嗽声 1 058 个,其它声音样本 420 个,其中喷嚏声 184 个,鸣叫声 121 个,呼噜声 115 个。验证集和测试集的数目分别设

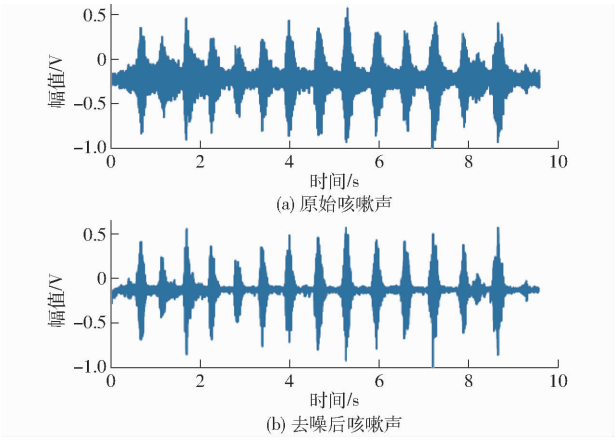


图 2　谱减法去噪前后猪咳嗽声波形图

Fig. 2　Sound waveform of pig cough before and after spectral subtraction denoising

置为本类声音样本个数的 10%，每次实验进行随机选取，剩余样本设置为训练集。表 1 为数据集划分结果。

表 1　数据集划分结果

Tab. 1　Data set partitioning results　　个

声音类别	声音特征	训练集样本数	验证集样本数	测试集样本数	总数
咳嗽	声音短且连续，强劲有力，喉音明显	848	105	105	1 058
喷嚏	声音短且快速，鼻音明显	148	18	18	184
鸣叫	声音尖锐刺耳	97	12	12	121
呼噜	声音低沉，类似哼叫	93	11	11	115

2　猪咳嗽声识别

2.1　猪声音特征参数提取

滤波器组(FBank)特征由于相邻滤波器组有重叠，相邻的特征高度相关，MFCC 特征是在 logFBank(FBank 取对数)特征的基础上再进行离散余弦变换(DCT)得到，具有更好的判别度。分别提取上述得到的声音信号的 logFBank 和 MFCC 两种音频特征。通过实验分析比较这两种声学特征对咳嗽声识别模型效果的影响，提取的特征存储为 npy 文件。图 3 为算法整体流程图，检测样本格式等信息正确后逐个提取特征并保存，否则跳过此样本。

为了补偿声音信号中被压抑的高频部分，突出高频的共振峰，需要在频域上面乘以一个系数，预加重系数设置为 0.97。预加重对噪声无影响，提高了输出信噪比，公式为

$$s_n' = s_n - ks_{n-1} \tag{1}$$

式中　s_n'——预加重后的声音信号

s_n——n 时刻声音采样值

s_{n-1}——$n-1$ 时刻声音采样值

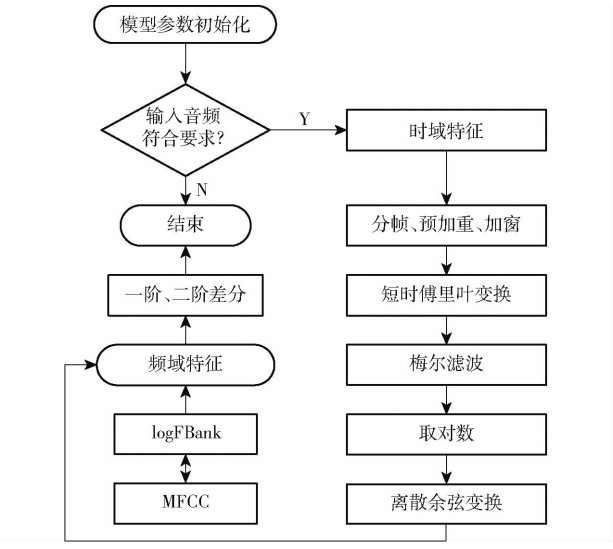


图 3　猪只声音特征提取算法流程图

Fig. 3　Flow chart of pig sound feature extraction algorithm

k——预加重系数

由于原始音频样本时间长度不固定，但具有微观上的短时平稳性，为了便于傅里叶变换，加窗切分为固定长度的小片段。为了保证信号不失真，根据奈奎斯特采样定律，采样频率设置为 48 kHz，帧长设置为 25 ms。为了避免窗边界对信号的遗漏，帧与帧之间设置一部分重叠区域，即帧移设置为 10 ms。采用汉明窗消除每一帧信号在其两端出现的信号不连续问题，从而使傅里叶变换之后取得更高质量的频谱。汉明窗计算公式为

$$W(n) = 0.538\,36 - 0.461\,64\cos\frac{2\pi n}{N-1}$$
$$(n = 0, 1, \cdots, N) \tag{2}$$

式中　$W(n)$——汉明窗函数

N——信号总的采样点个数

n——采样点序号

加窗过程为汉明窗函数与原信号作乘积变换，公式为

$$S'(n) = W(n)S(n) \tag{3}$$

式中　$S(n)$——原始帧信号

$S'(n)$——加窗后信号

为了方便深度神经网络学习，需要将时域信号转换到频域，本实验在每一帧进行 2 048 点短时傅里叶变换(Short time Fourier transform, STFT)，计算公式为

$$X_n(W) = \sum_{n=-\infty}^{\infty} x(n)W(n - mR)\mathrm{e}^{-\mathrm{j}\omega n} \tag{4}$$

式中　$X_n(W)$——短时傅里叶变换频域值

$x(n)$——第 n 个采样点的采样值

m——汉明窗长度　　ω——角频率

R——窗口随时间滑动的距离

W——汉明窗函数

短时傅里叶变换后得到的幅度谱与每一个梅尔刻度滤波器进行频率相乘累加，滤波器组个数设置为 64。为了更接近人耳听声原理，滤波后取对数，最终得到 logFBank 特征。MFCC 特征梅尔刻度滤波器组个数设置为 128，离散余弦变换（DCT）后，返回倒谱数 64 个。最终将时域特征转换到频域，得到 4 种声音信号的各两种声音特征，分别再对其求一阶差分和二阶差分，得到 3 个特征进行零-均值标准化（z-score）处理，转换成三维数组作为咳嗽声识别模型的输入。

图 4 为时频域转换结果，从左到右依次为呼噜声、喷嚏声、鸣叫声、咳嗽声，图中 Δ 为一阶差分。通过对比分析 4 种声音各自的时域和频域特征图可发现，不同声音信号的语谱图存在明显差异，4 种声音信号的 logFBank 及 MFCC 特征也存在差异，这是本实验进行分类识别的基础。

2.2 猪咳嗽声识别模型

采用提取的 logFBank 和 MFCC 及其一阶、二阶差分组合作为特征参数，分别经过 CNNs 和 DFSMN

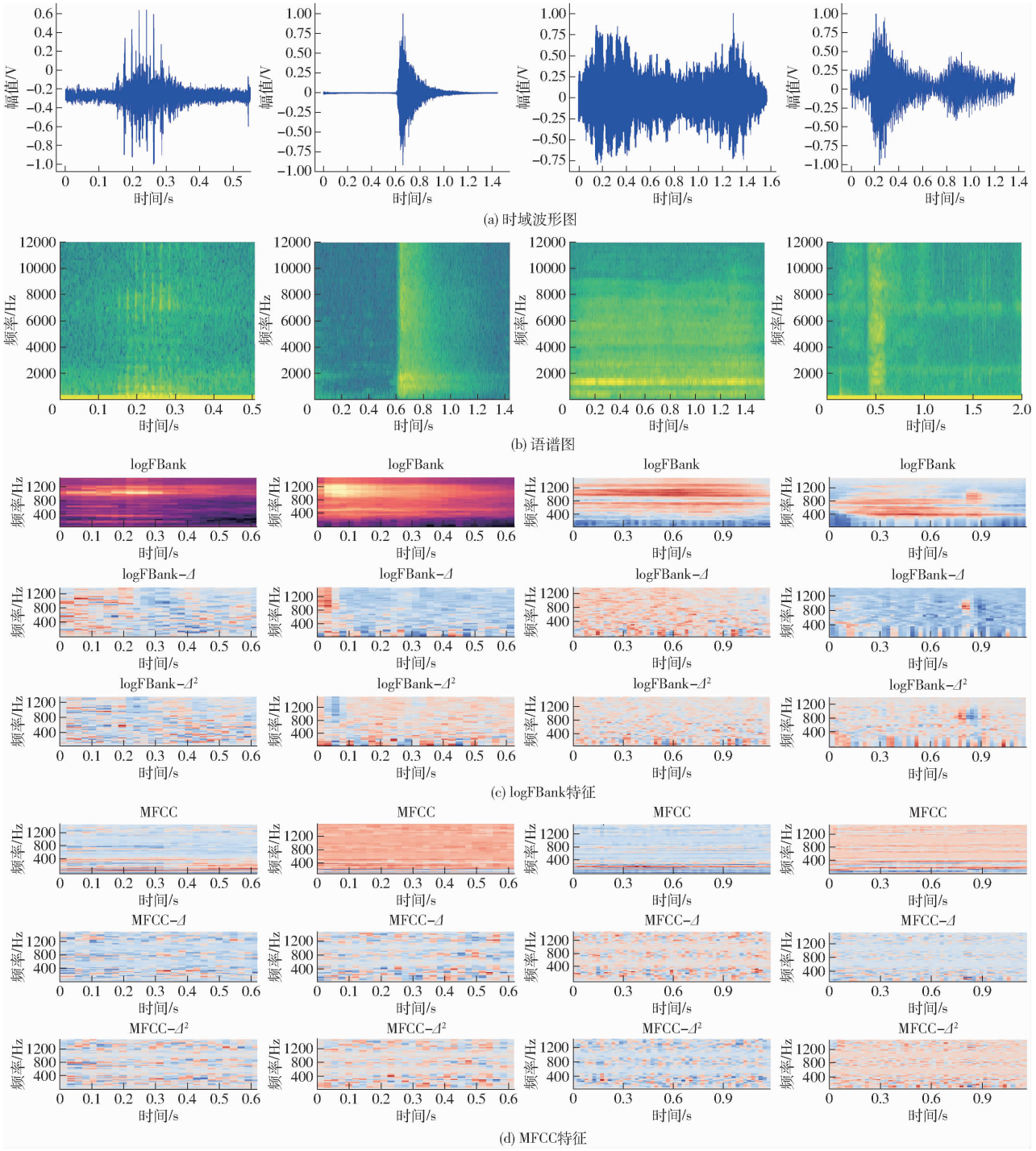


图 4 猪声音时频域特征转换

Fig. 4 Feature conversion of pig sound in time-frequency domain

两种深度学习模型训练后比较总体精确度、咳嗽声和非咳嗽声识别精确度、召回率及 F1 值等指标，综合判定模型性能。图 5 为深度神经网络模型整体训练流程图。

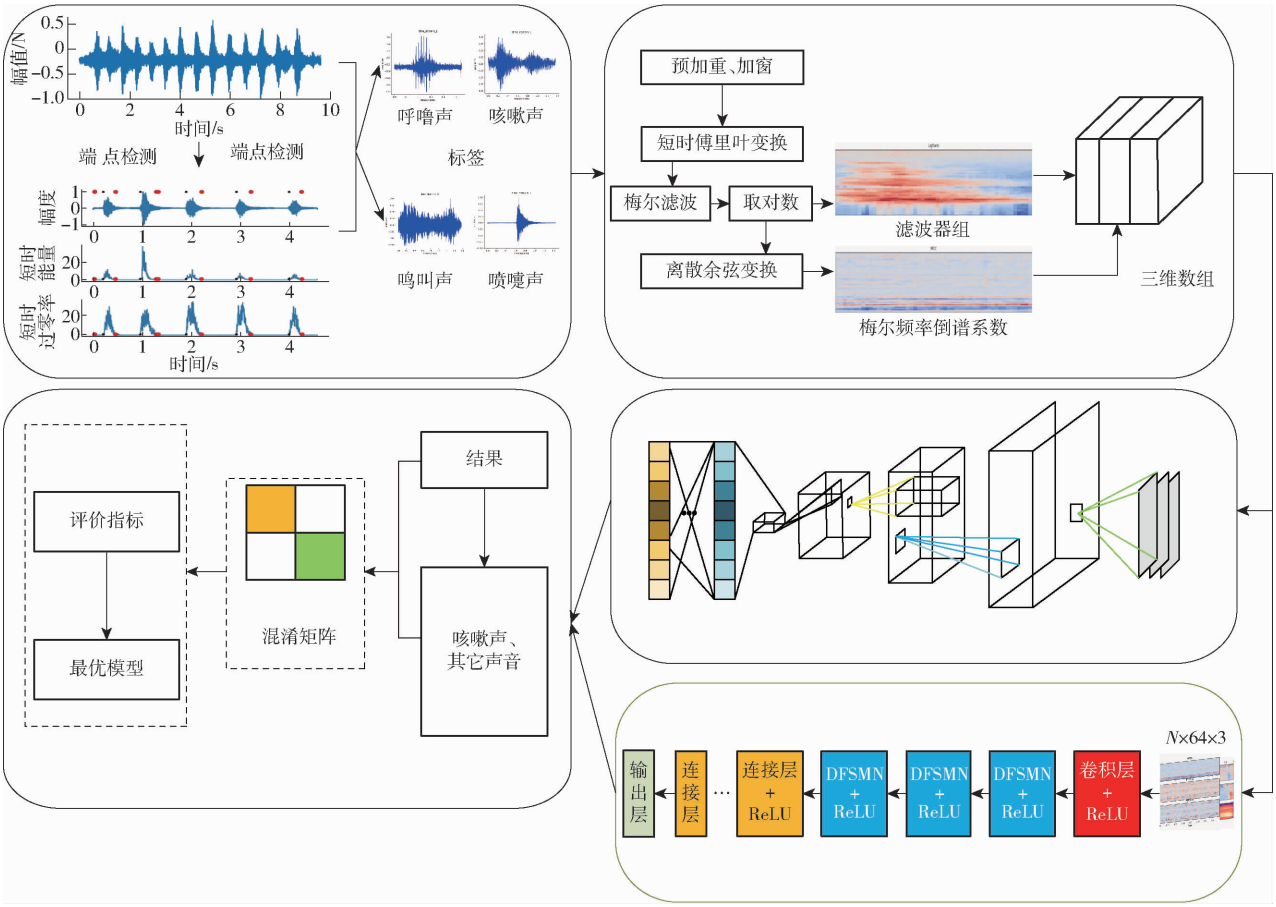


图 5　模型训练整体流程图

Fig. 5　Overall process of model training

2.2.1　基于 CNNs 的猪咳嗽声识别模型

卷积神经网络是目前比较流行的一种深度学习算法，在图像特征提取、语音处理等领域效果优异。一个深层 CNNs 模型通常包含卷积层、激活层、池化层和全连接层。卷积层是用若干个卷积核进行卷积运算，即卷积核在二维特征数据上按特定步长滑动计算权重矩阵和扫描所得的数据矩阵的乘积，得到一个输出，一层所有卷积核的个数即为本层卷积输出通道数。激活层通常使用 ReLU 函数，用于非线性运算，使得神经网络能更好地解决更加复杂的问题。

池化层通常在各个卷积层之间，降低各个特征的维度，减少过拟合，通常有最大池化（Max pooling）和平均池化（Average pooling）两种形式，最大池化即把上层结果特征元素的最大值作为输出，平均池化即将特征平均值作为下层输入，本文选择最大池化层。全连接层通常在模型尾部，所有神经元都有权重连接，用于特征分类。图 6 为基于 CNNs 的猪只咳嗽声识别模型的整体结构，包括 3 层卷积加最大池化层以及 3 层全连接层。

本文使用的 CNNs 模型包括 3 层卷积层，输入通道数分别为 3、32、32，输出通道数分别为 32、32、32，卷积核（kernel-size）均为 3，步长（stride）默认为 1，无填充（padding-mode）。激活层均采用 ReLU 函数，利用最大池化层（MaxPool2d）提取重要信息，减少计算量，kernel-size 和 stride 均为 2。全连接层输入输出通道数分别为 576 × 256、256 × 256、256 × 2。模型具体参数结构如表 2 所示。

2.2.2　基于 DFSMN 的猪咳嗽声识别模型

前馈序列记忆神经网络（DFSMN）通过进一步在 cFSMN 的记忆模块之间添加跳转连接（Skip connection），从而可以将低层记忆模块的输出直接累加到高层记忆模块里，有利于解决网络深度造成的梯度消失问题。DFSMN 记忆单元公式为

$$\widetilde{P}_t^l = H(\widetilde{P}_t^{l-1}) + P_t^l + \sum_{i=0}^{N_1^l} a_i^l \odot P_{t-s_1 i}^l + \sum_{j=1}^{N_2^l} c_i^l \odot P_{t+s_2 j}^l \tag{5}$$

其中　　$P_t^l = V^l h_t^l + b^l$

式中　P_t^l——第 l 个线性投影层的线性输出

V^l——第 l 个隐含层和第 l 个线性投影层的权值

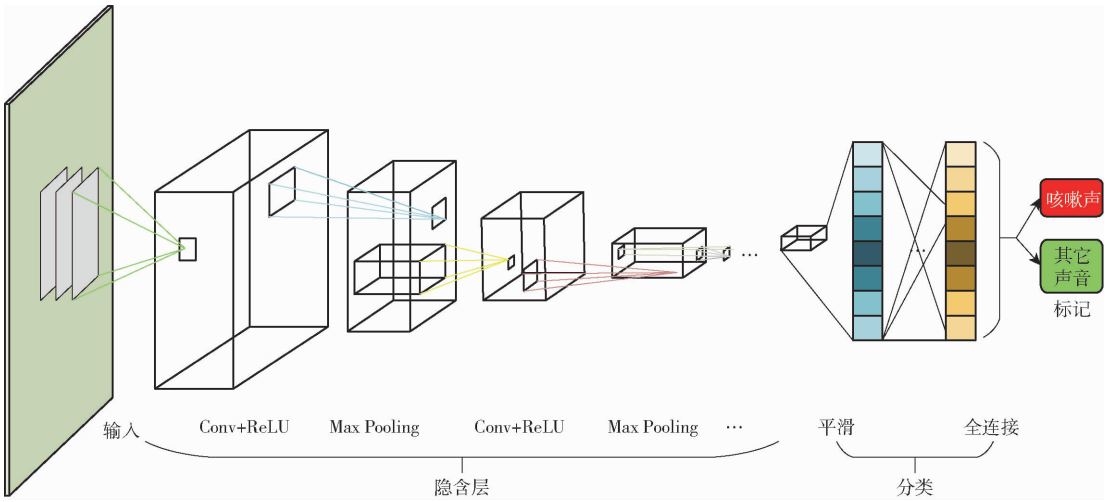


图 6 基于 CNNs 的咳嗽声识别模型结构

Fig. 6 Structure of cough recognition model based on CNNs

表 2 CNNs 模型结构参数

Tab. 2 Model structure parameters of CNNs

类别	层序号	输出通道数	步长	窗口尺寸
卷积层	1	32	1	3×3
最大池化层	3	32	2	2×2
卷积层	4	32	1	3×3
最大池化层	6	32	2	2×2
卷积层	7	32	1	3×3
最大池化层	9	32	2	2×2
全连接层	10	256		
全连接层	12	256		
全连接层	14	4		

h_t^l——第 l 个隐含层的输出

b^l——第 l 个隐含层和第 l 个线性投影层的偏置

$\widetilde{P}_t^l$——第 l 层记忆块输出

$H(\cdot)$——记忆模块之间的跳转连接函数

$\widetilde{P}_t^{l-1}$——第 $l-1$ 层记忆模块 t 时刻的输出

N_1^l——第 l 个记忆块的历史信息的阶数

N_2^l——第 l 个记忆块的未来信息的阶数

$P_{t-s_1 i}^l$——第 l 个记忆块的后向 i 个步幅因子数的线性投影层输出

$P_{t+s_2 j}^l$——第 l 个记忆块的前向 j 个步幅因子数的线性投影层输出

a_i^l——第 l 个记忆块的第 i 个后向可学习的系数因子

c_i^l——第 l 个记忆块的第 i 个前向可学习的系数因子

s_1——反向滤波器的编码步幅因子

s_2——前向滤波器的编码步幅因子

在声音信号预处理中，相邻帧信息由于重叠而具有很强的冗余性，DFSMN 将步长因子添加到记忆块中，以消除这种冗余。总延迟与每个记忆块中的步长和前向过滤器顺序有关，计算公式为

$$\tau = \sum_{l=1}^{L} N_2^l s_2 \tag{6}$$

式中 τ——记忆模块的延迟

L——记忆模块总数

训练的 DFSMN 含 1 个线性投影层、2 个隐含层，第 1 个隐含层记忆块使用 20 阶的 FIR 滤波器，隐含层单元采用 ReLU 函数。图 7 为 DFSMN 的内部结构。图中 $H = \{h_t^1, h_t^2, \cdots, h_t^5\}$ 表示相关隐含层的输出，$p = \{p_t^1, p_t^2, p_t^3, p_t^4\}$ 表示线性投影层输出，$\tilde{p} = \{\tilde{p}_t^1, \tilde{p}_t^2, \tilde{p}_t^3\}$ 表示记忆模块，$V = \{V^1, V^2, V^3, V^4\}$ 表示非线性隐含层与线性隐含层连接系数因子，$W = \{W^1, W^2, W^3, W^4\}$ 表示两层非线性隐含层之间的连接系数因子。

咳嗽声识别模型输入层接着一层卷积网络，DFSMN 部分由 3 层 DFSMN 组成，紧接着 3 个 ReLU 激活函数层和 3 层全连接层，最后是输出部分。卷

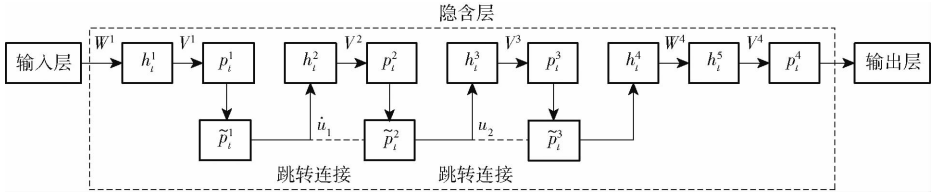


图 7 DFSMN 结构

Fig. 7 Structure of DFSMN

积层输入通道数为每种特征及其一阶和二阶差分组合，设置为 3，输出通道数依据经验值即卷积核个数，设置为 2 的倍数 32，卷积核通道数依据输入通道数设置为 3；DFSMN 层模型输入特征维度由上层卷积运算后进行数组重新组合得到，具体输入分别为 1 984、256、256 个单元，输出特征维度与下一层矩阵维度匹配，可进行矩阵运算，输出均为 256 个单元；全连接层同理，输入分别为 9 728、256、256 个单元，输出为 256、256、4 个单元。图 8 为基于 DFSMN 的咳嗽声识别模型框架，图中 InF、OutF 分别表示输入、输出特征，In-Conv、L-Conv、R-Conv 分别表示输入、左半部分、右半部分网络的一维卷积层。

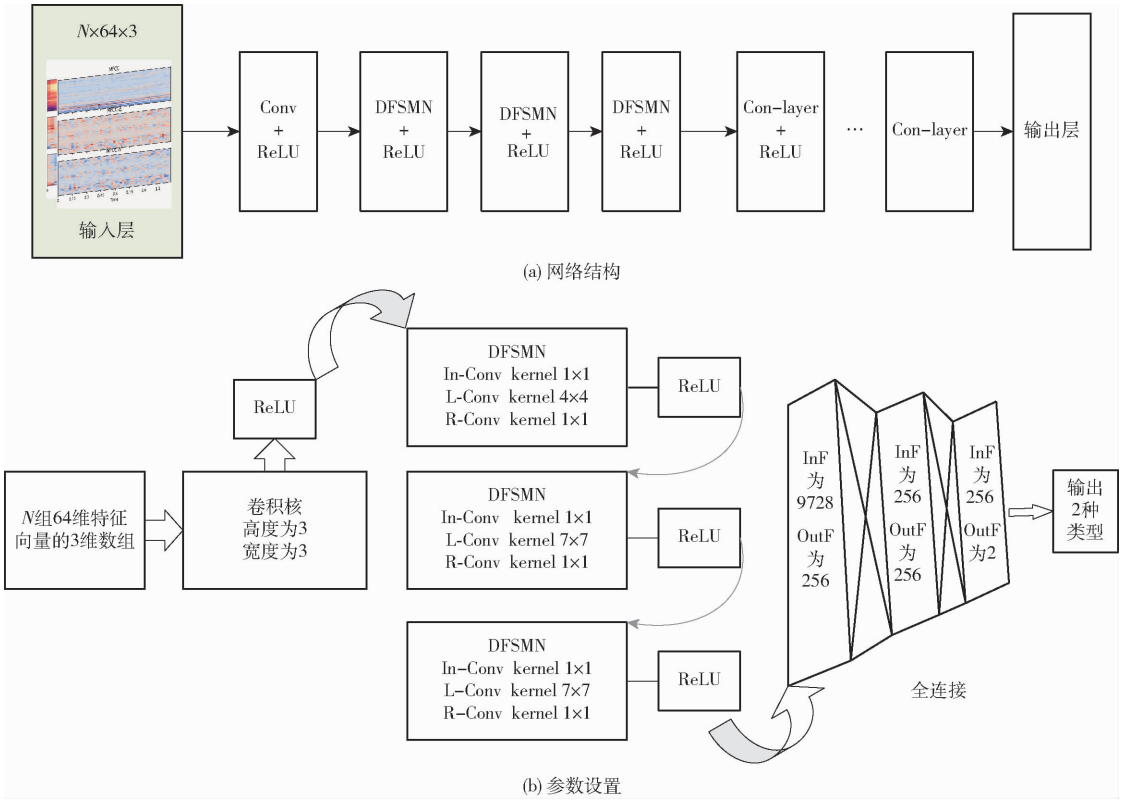


图 8　基于 DFSMN 的咳嗽声识别模型

Fig. 8　Cough recognition model based on DFSMN

2.2.3　模型训练步骤与参数设置

深度神经网络模型训练步骤通常随机生成初始模型参数，算法代入设定好的学习率和迭代次数 N；由损失函数获得损失值(loss)；计算导数后由优化函数(optimizer)得到下一次迭代的优化参数；不断迭代直至迭代 N 次。

本文的神经网络具体操作步骤如下：①整体网络框架构建，包括训练集、验证集、测试集样本数据加载。由于原始音频时长不一，特征提取后帧数不同，在输入模型之前需要对数据进行长度统一，本文取最大长度 40 帧为阈值，不足 40 部分由零填充，超过 40 强制截取最大长度。创建以时间为名的文件夹，用来存储模型结果。②超参数初始化，每轮模型训练中，利用 shuffle 函数打乱样本顺序。训练一次取出样本数据 64 个，全部数据参与训练完成一次迭代周期，初始轮数设置为 100 轮。设定随机梯度下降(Stochastie gradient descent, SGD)为优化器，学习率初始化为 0.001，动量设置为 0.9，学习衰减率为 0.98。优化器公式为

$$g_t = \nabla f(\theta_{t-1}) \tag{7}$$

$$\Delta\theta_t = -\eta g_t \tag{8}$$

式中　θ_{t-1}——模型参数

η——学习率

g_t——损失值关于参数的梯度

f——代价函数　　∇——梯度算子

损失函数为交叉熵损失(CrossEntropyLoss)，损失函数公式为

$$H(p,q) = -\sum_x p(x)\lg q(x) \tag{9}$$

式中　H——损失函数值

p——样本真实分布

q——模型预测的样本分布

$p(x)$——期望输出

$q(x)$——实际输出

③开始训练后每次迭代完成计算损失并更新参数，每 50 个迭代周期降低一次学习率。④训练结束对比验证集和测试集的训练效果，手动调整相对应的超参数，当训练集和验证集的损失值、精确率浮动误

差不超过 2% 时,停止训练。

3 实验结果与分析

本实验均在一台配置为 1 块 GTX1080Ti 显卡、有效内存 31.1 GB、2 块 Xeon Gold 5118CPU 的工作站进行训练,所涉及算法均基于 pytorch 深度学习开发库。

3.1 模型评价指标

通过混淆矩阵形式清楚表示分类识别结果的咳嗽声正确识别个数、咳嗽声误识别个数、其它声音正确识别个数与误识别个数,并通过准确率(Accuarcy)、精确率(Precision)、召回率(Recall)和 F1 值来衡量猪咳嗽声识别模型的性能。

3.2 不同迭代次数和特征提取方法对模型效果的影响

3.2.1 不同迭代次数

影响深度学习模型性能的因素通常包括学习率、迭代次数、初始化参数等,数据集数量、迭代次数、模型深度对模型效果的影响尚未发现普遍规律。

实验设置从 100 轮开始每次增加 50 轮直至 250 轮,MFCC - CNNs 模型测试集准确率在 100 ~ 200 轮之间不断上升,在训练轮数 200 时咳嗽声精确率达到 97%,其它声音精确率为 93%,召回率咳嗽声为 96%,其它声音为 93%,F1 值咳嗽声为 98%,其它声音为 93%,总体识别准确率为 96.71%。随着训练轮数继续增加,模型准确率并不会继续提高,识别准确率稳定在 91% 以上。logFBank - CNNs 模型性能较稳定,测试集的准确率基本稳定在 94% 左右。MFCC - DFSMN 模型最高准确率在 200 轮时达到 92.46%,总体识别率较低,模型总体性能较差。logFBank - DFSMN 模型咳嗽声识别准确率稳定在 90% 以上,最高达到 95.89%。

表 3 为不同迭代次数对不同模型在验证集和测试集上咳嗽声和其它声音的识别效果对比,表中 P_1、P_2、R_1、R_2、F_1、F_2、A 分别表示咳嗽声和其它声音的精确率、召回率、F1 值以及总体识别准确率。综上分析在训练轮数为 200 时,MFCC - CNNs 模型对咳嗽声的识别效果最好。不同特征与神经网络的组合随训练次数不同,识别准确率存在上下波动情况,波动幅度不超过 4%,总体稳定性较好。

3.2.2 不同特征提取方法

表 4、5 为两种神经网络模型训练次数为 200 时不同特征输入下咳嗽声和其它声音的精确率、召回率、F1 值和总体识别准确率的比较。模型在两种特征输入下,测试集和验证集的结果基本相同。

表 3 不同迭代次数识别效果

Tab.3 Experiment results of different iteration times

%

模型	训练轮数	数据集	P_1/P_2	R_1/R_2	F_1/F_2	A
MFCC - CNNs	100	验证集	91/86	95/76	93/81	89.72
		测试集	91/87	96/88	93/85	91.78
	150	验证集	94/78	90/85	92/81	89.04
		测试集	94/78	90/87	92/85	95.89
	200	验证集	97/93	97/92	97/93	96.49
		测试集	97/93	96/93	98/93	96.71
	250	验证集	93/89	96/80	94/85	91.78
		测试集	93/89	96/82	94/86	90.41
logFBank - CNNs	100	验证集	94/92	97/83	95/87	93.15
		测试集	94/92	97/83	95/87	94.52
	150	验证集	93/87	95/80	94/84	91.09
		测试集	93/87	95/80	94/84	94.52
	200	验证集	92/86	95/78	93/82	90.41
		测试集	92/86	95/78	93/82	91.78
	250	验证集	92/86	95/78	93/82	95.89
		测试集	92/86	95/78	93/82	95.89
MFCC - DFSMN	100	验证集	88/87	96/66	92/75	87.67
		测试集	88/87	96/66	92/75	90.41
	150	验证集	93/100	99/80	96/89	94.52
		测试集	93/100	99/80	96/89	86.98
	200	验证集	91/82	95/76	92/78	88.35
		测试集	91/82	95/76	92/78	92.46
	250	验证集	88/90	97/66	92/76	88.35
		测试集	88/90	97/66	92/76	83.56
logFBank - DFSMN	100	验证集	94/92	97/85	96/89	93.83
		测试集	94/92	97/85	96/89	91.09
	150	验证集	92/82	93/78	92/80	89.04
		测试集	92/82	93/78	92/80	95.89
	200	验证集	88/80	93/68	91/74	86.30
		测试集	88/80	93/68	91/74	91.78
	250	验证集	95/90	96/88	96/89	93.83
		测试集	95/90	96/88	96/89	90.41

表 4 MFCC 特征识别效果

Tab.4 Recognition effect of MFCC %

模型	数据集	P_1/P_2	R_1/R_2	F_1/F_2	A
CNNs	验证集	97/93	97/92	97/93	96.49
	测试集	97/93	96/93	98/93	96.71
DFSMN	验证集	95/100	99/85	97/92	95.89
	测试集	95/99	99/86	97/93	88.35

表 5 logFBank 特征识别效果

Tab.5 Recognition effect of logFBank %

模型	数据集	P_1/P_2	R_1/R_2	F_1/F_2	A
CNNs	验证集	95/88	95/88	95/88	93.15
	测试集	95/89	95/90	95/89	94.52
DFSMN	验证集	93/92	97/80	95/86	92.47
	测试集	93/92	97/81	95/87	95.89

由图 9a 可以明显发现，以 MFCC 为特征的两种模型准确率均高于以 logFBank 为特征的模型，且 MFCC－CNNs 的准确率最高。由图 9 可以看出，CNNs 模型在迭代周期数为 25 时，准确率还处于上升阶段，此时损失值曲线处于不断下降状态，波动较小，而后逐渐平稳，模型收敛。DFSMN 模型准确率与损失值趋于平稳状态要早于 CNNs 模型，最终 logFBank－CNNs、logFBank－DFSMN、MFCC－CNNs、MFCC－DFSMN 4 种模型验证集训练损失值分别为 0.032、0.044、0.025、0.028。

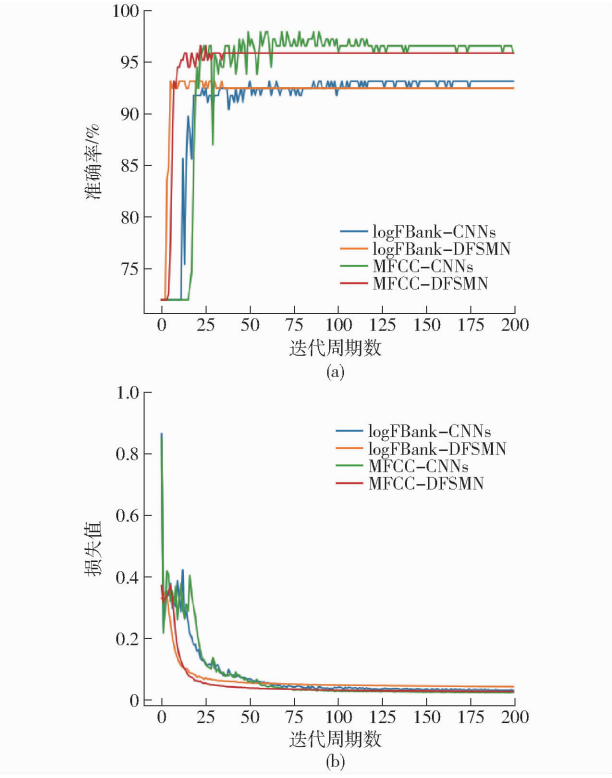


图 9　4 种模型验证集准确率、损失值变化曲线

Fig. 9　Accuracy and loss curves of four models

3.3　MFCC－CNNs 对 4 种声音的分类效果

猪只的咳嗽声和喷嚏声由人耳试听存在一定相似度，为了更加准确地对猪只的不同声音进行区分，在二分类的基础上将标签数增加至 4 个，检验 MFCC－CNNs 模型对 4 种声音分类识别的效果。表 6 为训练 100 轮后测试集上的模型评价指标结果。由测试集结果可知，总体识别准确率为 92.47%。相较咳嗽声，其它 3 种声音召回率和 F1 值都比较低，考虑原因是样本数量不均衡或者样本数量较少导致。图 10 为 MFCC－CNNs 模型训练过程，图 10a 表明验证集准确率尚未达到训练集水平，图 10b 表明损失值均随着网络迭代周期数逐步下降，最终趋于稳定，损失值分别为 0.009、0.052。

表 6　MFCC－CNNs 四分类效果对比

Tab. 6　Comparison of four classification effects of MFCC－CNNs　%

数据集	精确率	召回率	F1 值	总体准确率
咳嗽声	92	95	93	92.47
喷嚏声	93	72	81	
鸣叫声	100	75	86	
呼噜声	64	82	72	

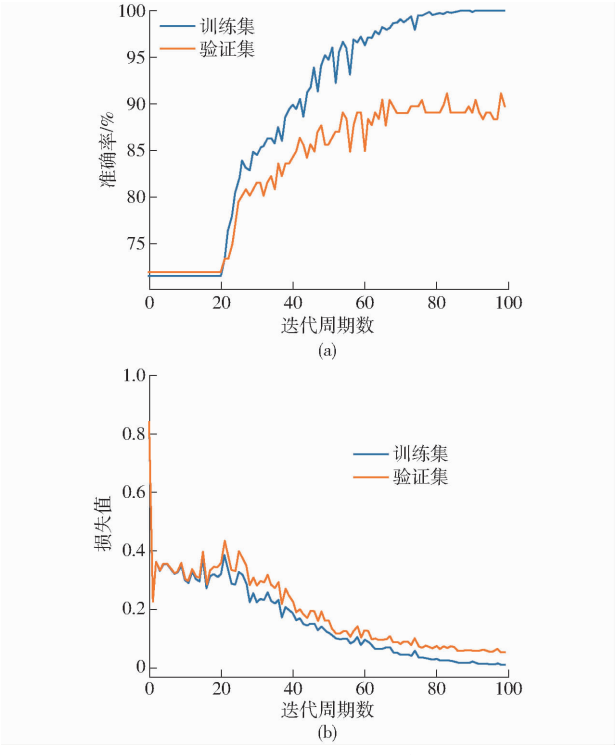


图 10　MFCC－CNNs 训练过程

Fig. 10　MFCC－CNNs training process

4　结论

(1)提出了一种基于 MFCC－CNNs 的深度神经网络猪咳嗽声识别方法。通过传统谱减法处理后原始声音噪声明显减小，利用双门限端点检测法得到有效声音信号后分别提取 logFBank 和 MFCC 特征，结果显示以 MFCC 为特征的 CNNs 咳嗽声识别模型准确率最高，咳嗽声识别精确率为 97%，召回率为 96%，F1 值为 98%，总体识别准确率为 96.71%。

(2)将深层卷积神经网络与前馈序列记忆神经网络引入猪咳嗽声识别领域，通过设置不同迭代次数的对比实验可知，不同迭代次数对二分类模型结果的影响最大误差不超过 4%，且以 MFCC 为特征输入的模型准确率普遍高于以 logFBank 为特征输入的模型。

参 考 文 献

[1] 黎煊,赵建,高云,等.基于连续语音识别技术的猪连续咳嗽声识别[J].农业工程学报,2019,35(6):174 - 180.
LI Xuan, ZHAO Jian, GAO Yun, et al. Recognition of continuous pig cough based on continuous speech recognition technology [J]. Transactions of the CSAE, 2019, 35(6):174 - 180. (in Chinese)

[2] SHEN Weizheng, TU Ding, YIN Yanling, et al. A new fusion feature based on convolutional neural network for pig cough recognition in field situations[J]. Information Processing in Agriculture, 2020, 8(4):573 - 580.

[3] 黎煊,赵建,高云,等.基于深度信念网络的猪咳嗽声识别[J].农业机械学报,2018,49(3):179 - 186.
LI Xuan, ZHAO Jian, GAO Yun, et al. Recognition of pig cough voice based on deep belief network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(3):179 - 186. (in Chinese)

[4] BAUMANN B, BILKEI G. Emergency-culling and mortality in growing/fattening pigs in a large Hungarian "farrow-to-finish" production unit[J]. Deutsche Tierärztliche Wochenschrift, 2002, 109(1):26 - 33.

[5] DAVIS J A. Pathological cry, stridor, and cough in infants[J]. Archives of Disease in Childhood, 1983, 58(4):319 - 320.

[6] GUARINO M, JANS P, COSTA A, et al. Field test of algorithm for automatic cough detection in pig houses[J]. Computers and Electronics in Agriculture, 2008, 62(1):22 - 28.

[7] EXADAKTYLOS V, SILVA M, AERTS J M, et al. Real-time recognition of sick pig cough sounds[J]. Computers and Electronics in Agriculture, 2008, 63(2):207 - 214.

[8] 徐亚妮,沈明霞,闫丽,等.待产梅山母猪咳嗽声识别算法的研究[J].南京农业大学学报,2016,39(4):681 - 687.
XU Yani, SHEN Mingxia, YAN Li, et al. Study on the sound recognition algorithm of cough of Meishan sows[J]. Journal of Nanjing Agricultural University, 2016, 39(4):681 - 687. (in Chinese)

[9] BAI Zhongxin, ZHANG Xiaolei. Speaker recognition based on deep learning: an overview[J]. Neural Networks, 2021, 140: 65 - 99.

[10] 王海坤,潘嘉,刘聪.语音识别技术的研究进展与展望[J].电信科学,2018,34(2):1 - 11.
WANG Haikun, PAN Jia, LIU Cong. Progress and prospect of speech recognition technology[J]. Telecommunication Science, 2018, 34(2):1 - 11. (in Chinese)

[11] CHEDAD A, MOSHOU D, AERTS J M, et al. Recognition system for pig cough based on probabilistic neural networks[J]. Journal of Agricultural Engineering Research, 2001, 79(4):449 - 457.

[12] RAMAIAH N P, IJJINA E P, MOHAN C K. Illumination invariant face recognition using convolutional neural networks[C]// IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES). IEEE, 2015.

[13] 赵建,黎煊,刘望宏,等.基于 DNN - HMM 声学模型的连续猪咳嗽声识别[J].农业工程技术,2020,13(3):186 - 193.
ZHAO Jian, LI Xuan, LIU Wanghong, et al. Recognition of continuous pig cough sound based on DNN - HMM acoustic model [J]. Agricultural Engineering Technology, 2020, 13(3):186 - 193. (in Chinese)

[14] 龚永杰,黎煊,高云,等.基于矢量量化的猪咳嗽声识别[J].华中农业大学学报,2017,36(3):119 - 124.
GONG Yongjie, LI Xuan, GAO Yun, et al. Recognition of pig cough sound based on vector quantization[J]. Journal of Huazhong Agricultural University, 2017, 36(3):119 - 124. (in Chinese)

[15] 张振华,田建艳,王芳,等.猪咳嗽声特征参数提取与识别的研究[J].黑龙江畜牧兽医,2017(23):18 - 22.
ZHANG Zhenhua, TIAN Jianyan, WANG Fang, et al. Study on extraction and recognition of pig cough sound characteristic parameters[J]. Heilongjiang Animal Husbandry and Veterinary Medicine, 2017(23):18 - 22. (in Chinese)

[16] 沈明霞,太猛,CEDRIC Okinda,等.基于深层卷积神经网络的初生仔猪目标实时检测方法[J].农业机械学报,2019,50(8):270 - 279.
SHEN Mingxia, TAI Meng, CEDRIC Okinda, et al. Real-time detection method of newborn piglets based on deep convolution neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(8): 270 - 279. (in Chinese)

[17] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems, 2012.

[18] LAKSHMI N A, JEEVA J B. A computer aided diagnosis for detection and classification of lung nodules[C]//2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO). IEEE, 2015.

[19] YIN Y, WU J M, ZHENG H X. Ncfm: accurate handwritten digits recognition using convolutional neural networks[C]//2016 International Joint Conference on Neural Networks (IJCNN). IEEE, 2016.

[20] YIN Yanling, TU Ding, SHEN Weizhing, et al. Recognition of sick pig cough sounds based on convolutional neural network in field situations[J]. Information Processing in Agriculture, 2020,8(3):369 - 379.

[21] 苍岩,罗顺元,乔玉龙.基于深层神经网络的猪声音分类[J].农业工程学报,2020,36(9):195 - 204.
CANG Yan, LUO Shunyuan, QIAO Yulong. Swine voice classification based on deep neural network[J]. Transactions of the CSAE, 2020, 36(9):195 - 204. (in Chinese)

[22] XIE Jie, ZHU Mingying. Handcrafted features and late fusion with deep learning for bird sound classification[J]. Ecological Informatics, 2019, 52:74 - 81.

[23] SEUNG J L, SEONG J J, JEE Y L, et al. Classification of snoring sound based on a recurrent neural network[J]. Expert Systems with Applications, 2019, 123:237 - 245.