潘伟豪, 盛卉子, 王春宇, 等. 基于欠定盲源分离和深度学习的生猪状态音频识别 [J]. 华南农业大学学报, 2024, 45(5): 730-742.
PAN Weihao, SHENG Huizi, WANG Chunyu, et al. Pig state audio recognition based on underdetermined blind source separation and deep learning[J]. Journal of South China Agricultural University, 2024, 45(5): 730-742.

基于欠定盲源分离和深度学习的生猪状态音频识别

潘伟豪[1]，盛卉子[1]，王春宇[1]，闫顺丕[2]，周小波[1]，辜丽川[1]，焦　俊[1]
(1 安徽农业大学 信息与人工智能学院, 安徽 合肥 230036； 2 安徽喜乐佳生物科技有限公司, 安徽 亳州 233500)

摘要:【目的】为解决群养环境下生猪音频难以分离与识别的问题，提出基于欠定盲源分离与 ECA-EfficientNetV2 的生猪状态音频识别方法。【方法】以仿真群养环境下 4 类生猪音频信号作为观测信号，将信号稀疏表示后，通过层次聚类估计出信号混合矩阵，并利用 l_p 范数重构算法求解 l_p 范数最小值以完成生猪音频信号重构。将重构信号转化为声谱图，分为进食声、咆哮声、哼叫声和发情声 4 类，利用 ECA-EfficientNetV2 网络模型识别音频，获取生猪状态。【结果】混合矩阵估计的归一化均方误差最低为 3.266×10^{-4}，分离重构的音频信噪比在 3.254~4.267 dB 之间。声谱图经 ECA-EfficientNetV2 识别检测，准确率高达 98.35%；与经典卷积神经网络 ResNet50 和 VGG16 对比，准确率分别提升 2.88 和 1.81 个百分点；与原 EfficientNetV2 相比，准确率降低 0.52 个百分点，但模型参数量减少 33.56%，浮点运算量 (FLOPs) 降低 1.86 G，推理时间减少 9.40 ms。【结论】基于盲源分离及改进 EfficientNetV2 的方法，轻量且高效地实现了分离与识别群养生猪音频信号。

关键词: 猪；盲源分离；声谱图；音频识别；稀疏重构；卷积神经网络
中图分类号: TN912.34；TP183；S828　　**文献标志码:** A　　**文章编号:** 1001-411X(2024)05-0730-13

Pig state audio recognition based on underdetermined blind source separation and deep learning

PAN Weihao[1], SHENG Huizi[1], WANG Chunyu[1], YAN Shunpi[2], ZHOU Xiaobo[1], GU Lichuan[1], JIAO Jun[1]
(1 School of Information and Artificial Intelligence, Anhui Agricultural University, Hefei 230036, China;
2 Anhui Xilejia Biotechnology Co., Ltd., Bozhou 233500, China)

Abstract:【Objective】In order to solve the problem of difficult separation and recognition of pig audio under group rearing environment, we propose a method of pig state audio recognition based on underdetermined blind source separation and ECA-EfficientNetV2. 【Method】Four types of pig audio signals were simulated as observation signals in group rearing environment. After the signals were sparsely represented, the signal mixing matrix was estimated by hierarchical clustering, and the l_p-paradigm reconstruction algorithm was used to solve for the minimum of l_p-paradigm to complete the reconstruction of pig audio signals. The reconstructed signals were transformed into acoustic spectrograms, which were divided into four categories, namely, eating sound, roar sound, hum sound and estrous sound. The audio was recognized using the ECA-EfficientNetV2 network

收稿日期:2023–12–07　　**网络首发时间:**2024–07–15 10:24:53
首发网址:https://link.cnki.net/urlid/44.1110.s.20240711.1442.004
作者简介:潘伟豪，硕士研究生，主要从事人工智能、信号处理等研究，E-mail: panweihao17114212@stu.ahau.edu.cn；通信作者：焦　俊，教授，博士，主要从事人工智能、智能控制和物联网等研究，E-mail: jiaojun2000@ahau.edu.cn
基金项目:安徽省重点研究与开发计划 (2023n06020051，202103B06020013)；安徽省研究生质量工程项目 (2022lhpysfjd023，2022cxcyjs010)

model to obtain the state of the pigs. 【Result】The normalized mean square error of the hybrid matrix estimation was as low as 3.266×10^{-4}, and the signal-to-noise ratios of the separated reconstructed audio ranged from 3.254 to 4.267 dB. The acoustic spectrogram was recognized and detected by ECA-EfficientNetV2 with an accuracy of up to 98.35%, and the accuracy improved by 2.88 and 1.81 percentage points compared with the classical convolutional neural networks ResNet50 and VGG16, respectively. Compared with the original EfficientNetV2, the accuracy decreased by 0.52 percentage points, but the amount of the model parameters reduced by 33.56%, the floating-point operations (FLOPs) reduced by 1.86 G, and inference time reduced by 9.40 ms. 【Conclusion】The method based on blind source separation and improvement of EfficientNetV2 lightly and efficiently realizes separating and recognizing audio signals of group-raised pigs.

Key words: Pig; Blind source separation; Spectrogram; Audio recognition; Sparse reconstruction; Convolutional neural network

生猪音频包含丰富的可利用信息[1]。然而，如何在嘈杂的群养环境中分离出各类生猪音频信号并有效识别是智慧养殖的难点问题，解决该问题也是智慧饲养的趋势。

国内外的盲源分离算法研究主要集中在军事通信、语音信号处理、生物医学信号处理等领域。Ghani 等[2] 利用一种基于投射追踪的盲源分离技术，成功将无人机声音与其他环境声音区分开来。He 等[3] 提出一种针对非平稳信号的时变卷积盲源分离算法，其采用变分贝叶斯推理方法和高斯过程，将非平稳源逐帧从时变卷积信号中分离出来，最终可有效分离时变混合语音信号。Adam 等[4] 利用基于快速独立分量分析的盲源分离技术，去除噪声对脑电图、肌电图等生物信号的影响，获得了较好的试验结果。

国内外学者在生猪音频识别方面已有相应进展，张振华[5] 利用隐马尔科夫模型对生猪打斗声、咳嗽声、饿叫声和抽搐声进行识别，总体识别率为 89.25%。沈明霞等[6] 提出一种基于深度神经网络的识别方法，提取梅山猪咳嗽及喷嚏、鸣叫、呼噜声的滤波器组与梅尔频率倒谱系数特征，识别生猪的咳嗽声，识别准确率高达 97%。Ji 等[7] 将声学和视觉特征融合，提取均方根能量、梅尔频率倒谱系数等特征，精准检测猪咳嗽声，准确率达 96.45%。在生猪音频盲源分离方面研究较少，彭硕等[8] 利用基于稀疏分量分析的欠定盲源分离方法，成功分离了 3 类混合猪声信号，但其研究尚未拓展到处理更多类别的生猪音频分离问题，也未考虑到后续重构音频识别问题。

本研究提出一种基于欠定盲源分离及改进 ECA-EfficientNetV2 的生猪状态音频识别方法。利用欠定盲源分离技术，从混杂音频中分离出哼叫声、进食声、咆哮声、发情声 4 类生猪状态音频信号，再采用 ECA-EfficientNetV2 模型识别音频，旨在实现对猪只生活健康状态的监测和识别。

1 材料与方法

1.1 试验整体设计流程

基于欠定盲源分离及 ECA-EfficientNetV2 的生猪状态音频识别方法总体流程如图 1 所示。

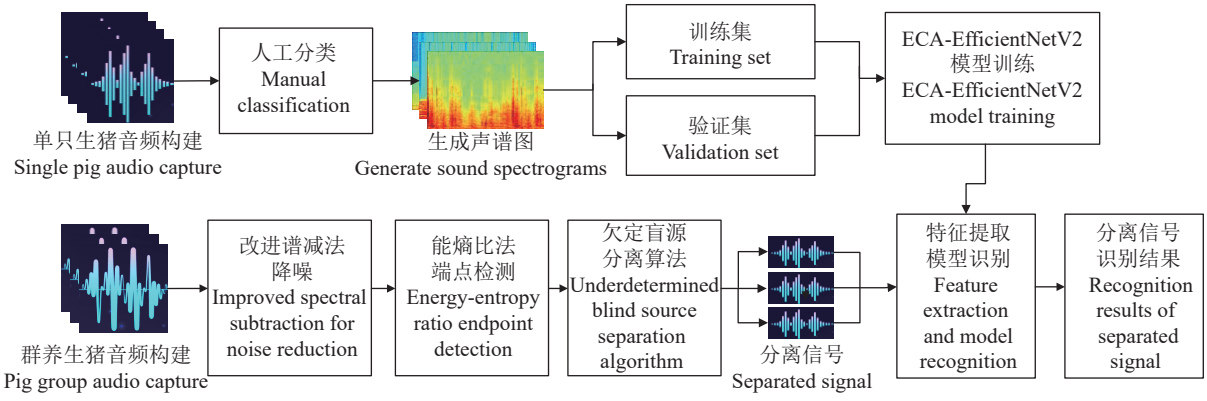


图 1 生猪音频识别试验总体流程图

Fig. 1 Overall flow chart of the pig audio recognition test

首先采集生猪不同状态的音频信号，将音频转化为具有时频特征的声谱图，构建声谱图数据集，再训练 ECA-EfficientNetV2，实现不同生猪状态音频识别。仿真群养环境下生猪音频后，采用改进谱减法降噪及能熵比法端点检测对音频预处理。得到降噪的生猪混合音频，利用欠定盲源分离算法对混合音频分离重构。最终将重构音频转化为声谱图后利用 ECA-EfficientNetV2 模型进行识别，从音频中获取生猪当前状态信息。

算法运行硬件环境为 Windows11 系统，内存为 32 GB，CPU 为 AMD Ryzen7 5800H with Radeon Graphics 3.2 GHz，GPU 为 NVIDIA Geforce RTX 3070。软件环境为开发平台 Pycharm 社区版 2022.3，使用语言为 Python3.8.16，神经网络模型在 Pytorch2.0.0 框架下运行，显卡为 Cuda 11.8 版。

1.2　生猪音频信号获取及预处理

1.2.1　*音频数据获取*　研究所用生猪音频采集于安徽某生猪养殖基地，采集装置为飞利浦 VTR5110 录音笔，单通道录制，采样点数设置为 16 bit，采样率为 44 100 Hz，音频保存格式为 WAV 格式。为获取较纯净的生猪音频，将长白猪依次单独圈养在封闭的 4 m×4 m 的房间内，房内无人无猪时声音分贝低于 12 dB。经人工分类，将采集的音频信号分为进食声、哼叫声、咆哮声、发情声、噪声和无声段。

1.2.2　*基于多窗谱估计的谱减法降噪*　多窗谱估计谱减法[9] 在传统谱减法的基础上，采用多个不同长度的分析窗口。每个窗口捕捉语音和噪声成分的不同时频特性，可更准确地估计噪声谱[10]，其具体算法流程图如图 2 所示。

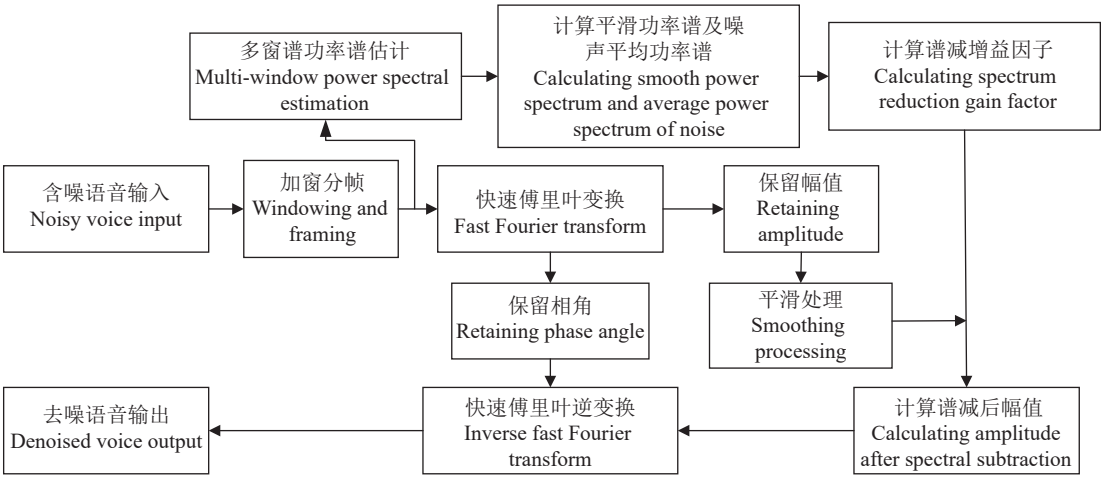


图 2　基于多窗谱估计的改进谱减法算法的流程图

Fig. 2　Flow chart of improved spectral subtraction algorithm based on multi-window spectral estimation

设含噪语音加窗分帧后为$x_i(z)$，将$x_i(z)$进行快速傅里叶变换 (Fast Fourier transform, FFT)，得出其幅度谱和相位谱，并且由此可计算出平均幅度谱：

$$\left|\overline{X_i}(k)\right| = \frac{1}{2Z+1}\sum_{j=-Z}^{Z}\left|X_{i+j}(k)\right|, \tag{1}$$

式中，$\left|\overline{X_i}(k)\right|$为平均幅度谱，$Z$ 为帧数，$\sum_{j=-Z}^{Z}\left|X_{i+j}(k)\right|$表示以第 i 帧为中心，取前后 Z 帧的幅度谱值相叠加，共有 2Z+1 帧进行平均，计算平均幅度谱。音频初始的数帧中，并无可用的生猪音频段，大多为仅含噪声的音频段。设此时仅含噪声的音频片段共有 NIS 帧，则可算出噪声的平均功率谱密度，如下式：

$$P_n(k) = \frac{1}{\mathrm{NIS}}\sum_{i=1}^{\mathrm{NIS}}P_y(k,i), \tag{2}$$

式中，$P_n(k)$为平均功率谱密度，$P_y(k,i)$为平滑功率谱密度，其由$x_i(z)$进行多窗谱估计后得出。

求出功率谱密度后，利用谱减关系计算增益因子，如下式：

$$g(k,i) = \begin{cases} [P_y(k,i)-\alpha P_n(k)]/P_y(k,i), & P_y(k,i)-\alpha P_n(k) \geqslant 0, \\ \beta P_n(k)/P_y(k,i), & P_y(k,i)-\alpha P_n(k) < 0, \end{cases} \tag{3}$$

式中，α为过减因子，β为增益补偿因子。

得出增益因子$g(k,i)$后，则可计算出谱减后的幅度谱，再通过离散傅里叶逆变换 (Inverse discrete Fourier transform, IDFT) 得到降噪音频，如下式：

$$\hat{x}_i(z) = \mathrm{IDFT}\left\{\left|\hat{X}_i(k)\right|\exp[j\theta_i(k)]\right\}, \tag{4}$$

式中，$\theta_i(k)$为相位谱，$\left|\hat{X}_i(k)\right|$为谱减后的幅度谱，

$\hat{x}_i(z)$为降噪音频，j 为傅里叶变换中的虚数单位。

1.2.3 基于能熵比的端点检测 基于能熵比的端点检测[11]利用信号的改进能量和谱熵 2 种参数的比值，确定有声段的起始位置和终止位置。有声段内，信号的熵相对较高；非有声段内，信号的熵相对较低[12]。

设加窗分帧后第 i 帧音频信号为$\boldsymbol{x}_i(m)$，FL 为音频信号的帧长度，则音频的每帧能量如下式：

$$\mathrm{AMP}_i = \sum_{m=1}^{\mathrm{FL}} x^2{}_i(m)。 \tag{5}$$

为缓和AMP_i的剧烈变化，引入常量 a，并将短时能量对数化计算得改进能量，如下式：

$$\mathrm{LE}_i = \lg(1 + \mathrm{AMP}_i/\mathrm{a})。 \tag{6}$$

将改进能量LE_i和谱熵H_i构成能熵比，如下式：

$$\mathrm{EEF}_i = \sqrt{1 + |\mathrm{LE}_i/H_i|}。 \tag{7}$$

1.2.4 数据集制作 声谱图[13]的作用是将时域的音频信号转为频域表示，更直观地观察和分析音频信号的频谱结构。制作声谱图流程如图 3 所示。流程中的伪彩色映射指将音频信号在时频域上的频率、能量、时间等特征信息以彩色图的形式展现出来。

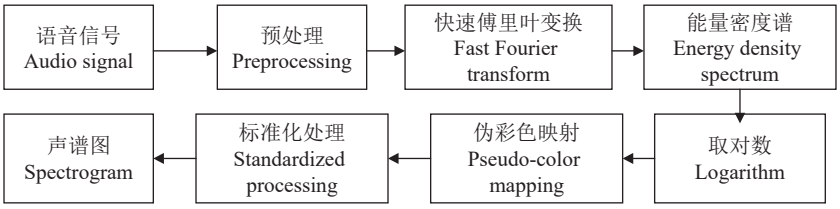
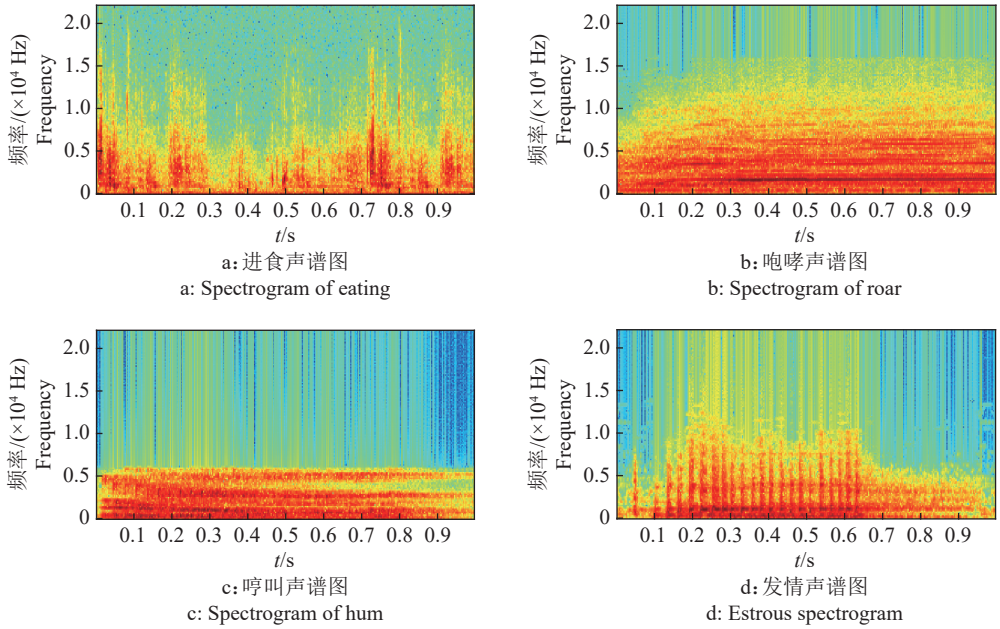


图 3 声谱图生成流程图

Fig. 3 Flow chart of spectrogram generation

ECA-EfficientNetV2 模型所需 4 种类型生猪声谱图样例如图 4 所示。利用不同窗函数、不同窗长生成具有不同时频特征信息的声谱图，其中宽带声谱图的时间分辨率较高，窄带声谱图的频率分辨率较高[14]，2 种声谱图如图 5 所示。

通常宽带声谱图以 3 ms 左右为 1 帧，窄带声谱图以 20 ms 左右为 1 帧进行分帧加窗处理[15]。根据 FFT 公式，当窗函数长度 T 为 3 ms 时，对应带宽约为 293 Hz，T 为 20 ms 时，对应带宽约为 44 Hz。图 5 显示的是同一音频下采用 44 Hz 的带宽和以 300 Hz 的带宽分帧制成的窄带声谱图和宽带声谱图。不同的窗函数，可提取不同的时频特征[16]。本

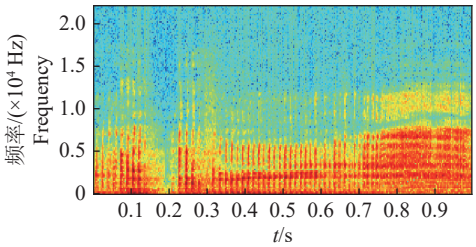


图中每个坐标点的颜色代表此时间点此频率上的能量大小，颜色越深，能量越高

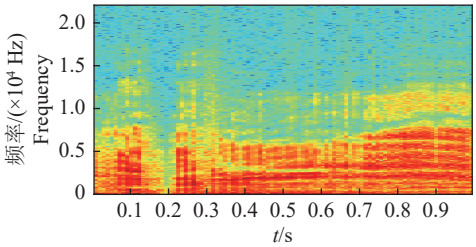
The color of each coordinate point in the graph represents the energy magnitude at this frequency at this time point, and the darker the color, the higher the energy

图 4 不同状态的生猪音频声谱图示例

Fig. 4 Examples of audio spectrograms of pigs in different states



a：宽带声谱图
a: Wide spectrogram



b：窄带声谱图
b: Narrowband spectrogram

图中每个坐标点的颜色代表此时间点此频率上的能量大小，颜色越深，能量越高

The color of each coordinate point in the graph represents the energy magnitude at this frequency at this time point, and the darker the color, the higher the energy

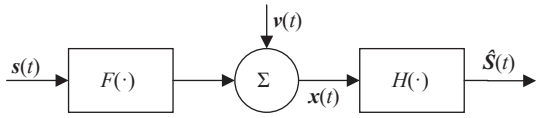
图 5　同音频下宽带声谱图与窄带声谱图

Fig. 5　Wideband and narrowband spectrograms under the same audio

研究使用汉明窗、汉宁窗和布莱克曼窗分别处理生成声谱图，将含不同特征的声谱图数据集利用 ECA-EfficientNetV2 模型训练，提高模型鲁棒性。

1.3　基于稀疏重构的欠定盲源分离

1.3.1　欠定盲源分离整体流程　典型盲源分离处理模型[17] 如图 6 所示。



$\boldsymbol{s}(t)$ 为源信号矢量，$F(\cdot)$ 为混合系统，$\boldsymbol{v}(t)$ 为噪声矢量，$\boldsymbol{x}(t)$ 为观测信号，$H(\cdot)$ 为解混系统，$\hat{\boldsymbol{S}}(t)$ 为解混信号，即重构后的信号

$\boldsymbol{s}(t)$ is the source signal vector, $F(\cdot)$ is the hybrid system, $\boldsymbol{v}(t)$ is the noise vector, $\boldsymbol{x}(t)$ is the observed signal, $H(\cdot)$ is the demixed system, $\hat{\boldsymbol{S}}(t)$ is the demixed signal, that is, the reconstructed signal

图 6　盲源分离处理模型

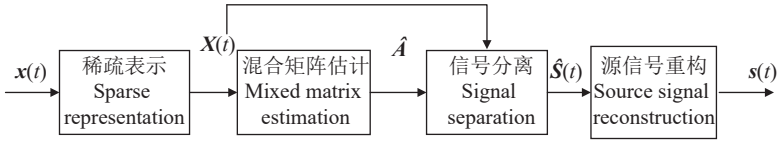
Fig. 6　Blind source separation processing model

研究仅考虑信号幅度衰减，不考虑时延性和传播路径问题，此时欠定盲源分离模型[18] 如下式：

$$\boldsymbol{x}_i(t)=\sum_{k=1}^{N}\boldsymbol{a}_{ik}\boldsymbol{s}_k(t)+\boldsymbol{v}_i(t),\quad 1\leqslant i\leqslant M, \tag{8}$$

式中，t 为某一时刻，N 为源信号数量，M 为观测信号数量，$\boldsymbol{a}_{ik}$ 为第 i 个时频点上第 k 个源信号对应的混合矩阵，表示信号的衰减幅度，$\boldsymbol{x}_i(t)$ 为第 i 个传感器获取到的观测信号，$\boldsymbol{s}_k(t)$ 为第 k 个源信号，$\boldsymbol{v}_i(t)$ 为第 i 个噪声信号噪声 $\boldsymbol{v}(t)$ 利用“1.2”节中改进谱减法进行降噪处理后得到的信号。

盲源分离中的欠定盲源分离处理如图 7 所示。



$\boldsymbol{x}(t)$ 为观测信号，$\boldsymbol{X}(t)$ 为稀疏表示后的稀疏向量，$\hat{\boldsymbol{A}}$ 为混合矩阵估计值，$\hat{\boldsymbol{S}}(t)$ 为分离出的信号，$\boldsymbol{s}(t)$ 为最终重构出的信号

$\boldsymbol{x}(t)$ is the observed signal, $\boldsymbol{X}(t)$ is the sparse vector after sparse representation, $\hat{\boldsymbol{A}}$ is the mixing matrix estimate, $\hat{\boldsymbol{S}}(t)$ is the separated signal, and $\boldsymbol{s}(t)$ is the final reconstructed signal

图 7　欠定盲源分离流程图

Fig. 7　Flow chart of underdetermined blind source separation

稀疏表示时，音频信号经过短时傅里叶变换从时域转为时频域，提升信号稀疏性[19]。稀疏表示后，聚类各特征点估计出混合矩阵，类心矢量方向即对应相应源信号的混合矢量方向[20]。利用信号的稀疏性，使用稀疏重构法[21] 对信号进行重构，分离各状态的生猪音频信号。

1.3.2　单源点检测　比较稀疏表示后混合信号实部与虚部，对信号进行单源点检测[22]，剔除低能点，使信号更具稀疏性。式 (8) 中，使用谱减法进行降噪操作后，任一时频点 (t,f) 上，可改写为式 (9)：

$$\boldsymbol{x}_i(t,f)=\sum_{k=1}^{N}\boldsymbol{a}_{ik}\boldsymbol{s}_k(t,f)=\boldsymbol{A}\boldsymbol{s}(t,f), \tag{9}$$

式中，$\boldsymbol{s}_k(t,f)$ 为某时频点上的第 k 个源信号，$\boldsymbol{A}$ 为混合矩阵向量，$\boldsymbol{s}(t,f)$ 为源信号向量。此时观测信号 $\boldsymbol{x}(t,f)$ 的实部为：

$$\mathrm{Re}\left[\boldsymbol{x}(t,f)\right]=\sum_{k=1}^{N}\boldsymbol{a}_{ik}\mathrm{Re}\left[\boldsymbol{s}_k(t,f)\right], \tag{10}$$

虚部为：

$$\mathrm{Im}\left[\boldsymbol{x}(t,f)\right]=\sum_{k=1}^{N}\boldsymbol{a}_{ik}\mathrm{Im}\left[\boldsymbol{s}_k(t,f)\right]。 \tag{11}$$

当时频点为单源点时，实部与虚部关系如式 (12)，$\boldsymbol{x}_1(t,f)$ 为观测信号 1 的向量，$\boldsymbol{x}_2(t,f)$ 为观测信号 2 的向量：

$$\frac{\mathrm{Re}\left[\boldsymbol{x}_2(t,f)\right]}{\mathrm{Re}\left[\boldsymbol{x}_1(t,f)\right]}-\frac{\mathrm{Im}\left[\boldsymbol{x}_2(t,f)\right]}{\mathrm{Im}\left[\boldsymbol{x}_1(t,f)\right]}=0, \tag{12}$$

由于噪声和计算误差的影响，很少有时频点满足式 (12) 的条件，因此降低检测条件，使用式 (13) 初步检测单源点:

$$\left|\frac{\mathrm{Re}\left[\boldsymbol{x}_2(t,f)\right]}{\mathrm{Re}\left[\boldsymbol{x}_1(t,f)\right]}-\frac{\mathrm{Im}\left[\boldsymbol{x}_2(t,f)\right]}{\mathrm{Im}\left[\boldsymbol{x}_1(t,f)\right]}\right|<\varepsilon_1, \tag{13}$$

式中，ε_1 为阈值条件，值接近 0。当满足式 (13) 时，特征点可视为单源点。

1.3.3 混合矩阵估计 将剔除大部分低能点的特征单源点使用聚类算法进行聚类，得到混合矩阵的估计值[23]。研究使用层次聚类算法[24] 对特征点进行聚类。聚类算法步骤如下：

1) 将每个特征点视为 1 个簇，并计算每个特征点之间的距离。

2) 合并 2 个簇之间距离最小的 2 个簇，形成 1 个新的簇。

3) 计算新的簇和当前其他簇之间的距离。

4) 反复重复步骤 2) 和步骤 3)，直到所有特征点合并完成。

使用平均距离计算簇之间的距离，表示 2 个簇中任意的 2 个点距离相加取和，设平均值 d_{avg} 为 2 个簇之间的距离，如式 (15)：

$$d_{\mathrm{avg}}(C_i,C_j)=\frac{1}{|C_i|\,|C_j|}\sum_{p\in C_i}\sum_{q\in C_j}|p-q|, \tag{14}$$

式中，C_i 和 C_j 为任意 2 个簇，p 和 q 为 2 个簇中任意 2 个点。

1.3.4 生猪音频信号重构 由于求解 l_0 范数最小化问题是 NP 难问题，因此使用 l_p 范数类算法中的求解非凸函数最小化算法 ($0<p<1$) 求解 l_p 范数最小化[25]，如式 (15)：

$$\min\|\boldsymbol{s}\|_p \text{ s.t. } \boldsymbol{x}=\boldsymbol{A}\boldsymbol{s}, \tag{15}$$

式中，$\boldsymbol{x}$ 和 $\boldsymbol{s}$ 分别表示观测信号向量和最终重构的音频向量。t 时刻，l_p 范数最小化的可能解为$\hat{\boldsymbol{S}}^{(K)}(t)$，式 (16) 中，$K$ 为取得局部最小值的次数，最多有 C_N^M 个可能解，$\hat{\boldsymbol{S}}_{K_M}^{(K)}(t)$为 t 时刻第 K 次取得局部最小值时第 M 个观测信号分解的源信号估计向量，将其余解$\hat{\boldsymbol{S}}_j^{(K)}(t)$均设为 0：

$$\hat{\boldsymbol{S}}^{(K)}(t)=\begin{cases}\left[\hat{\boldsymbol{S}}_{K_1}^{(K)}(t),\hat{\boldsymbol{S}}_{K_2}^{(K)}(t),\cdots,\hat{\boldsymbol{S}}_{K_M}^{(K)}(t)\right]^{\mathrm{T}}=\hat{\boldsymbol{A}}_K^{-1}\boldsymbol{x}(t),\\ \quad(K_1,K_2,\cdots K_M\in\{1,2,\cdots,N\}),\\ \hat{\boldsymbol{S}}_j^{(K)}(t)=0,\ (j\neq K_1,K_2,\cdots K_M),\end{cases} \tag{16}$$

式中，$\hat{\boldsymbol{A}}_K^{-1}$ 为C_N^M 个 $M\times M$ 维子矩阵的逆矩阵 (K=1, 2, ⋯, C_N^M)。

式 (16) 的$\hat{\boldsymbol{S}}^{(K)}(t)$对应 l_p 范数 J_K 为式 (17) 所示：

$$J_K=\sum_{i=1}^{M}\left|\hat{\boldsymbol{S}}_i^{(K)}(t)\right|^p,(K=1,2,\cdots,C_N^M), \tag{17}$$

最终确定 l_p 范数最小解$\hat{\boldsymbol{S}}_{\min}=\arg\min J_K$，此解即为源信号 $\boldsymbol{s}(t)$ 的估计$\hat{\boldsymbol{S}}(t)$。

1.3.5 盲源分离评价指标 采用信噪比[26] 和归一化均方误差[27] 2 种指标评价欠定盲源分离算法的分离质量。归一化均方误差用于评价混合矩阵估计的准确度，表达式如下：

$$\mathrm{NMSE}=10\times\lg\left(\frac{\sum_{i=1}^{L}(\hat{\boldsymbol{a}}_i-\boldsymbol{a}_i)^2}{\sum_{i=1}^{L}(\boldsymbol{a}_i)^2}\right), \tag{18}$$

式中，$\hat{\boldsymbol{a}}_i$为混合矩阵估计值，$\boldsymbol{a}_i$为原观测信号矩阵的值，L 为混合矩阵数量。归一化均方误差的值越小，代表估计的精确率越高。

盲源分离中信噪比用于判断原信号和重构信号的相似度，重构信号与源信号的差值类比为信噪比中的噪声，如下式：

$$\mathrm{SNR}_i=10\times\lg\left(\frac{\sum_{t=1}^{T}\boldsymbol{s}_i^2(t)}{\sum_{t=1}^{T}\left[\boldsymbol{s}_i(t)-\boldsymbol{y}_i(t)\right]^2}\right), \tag{19}$$

式中，$\boldsymbol{s}_i(t)$为第 i 个源信号，$\boldsymbol{y}_i(t)$为对应的重构信号。信噪比的值越大，源信号和重构信号之间的差异性越小，代表信号的重构效果越好。

1.4 基于 ECA-EfficientNetV2 的生猪声谱图识别

1.4.1 ECA-EfficientNetV2 网络模型 EfficientNetV2[28] 本质是基于卷积神经网络的一类模型。其为 EfficientNet 的改进模型，降低了原始模型的参数量，并引入了渐进学习方法，动态调节训练图像的尺寸。

由于 SE 模块跨通道交互过程中可导致模型降维，影响模型预测能力。针对生猪声谱图的特点，将 EfficientNetV2-S 架构进一步简化，并将 SE 注意力机制替换为更轻量且能有效避免模型降维影响的 ECA(Efficient channel attention) 注意力机制[29] 模块。架构如表 1 所示，将 Stage 2 和 Stage 3 中的 Fused-MBConv 内 Expansion ratio 调整为 2 倍，并将 Stage 1 中的 Fused-MBConv 模块和 Stage 4、Stage 5、Stage 6 中的 MBConv 模块的重复次数 Layers 进行缩减，降低模型参数量，改进后的网络模型结构如图 8 所示。

表 1　ECA-EfficientNetV2 架构表

Table 1　ECA-EfficientNetV2 architecture table

阶段 Stage	操作 Operator	层数 Layers	步长 Stride
0	Conv3×3	1	2
1	Fused-MBConv1, k3×3	1	1
2	Fused-MBConv2, k3×3	4	2
3	Fused-MBConv2, k3×3	4	2
4	MBConv4, k3×3, ECA	5	2
5	MBConv6, k3×3, ECA	7	1
6	MBConv6, k3×3, ECA	12	2
7	Conv1×1, Pooling, FC	1	

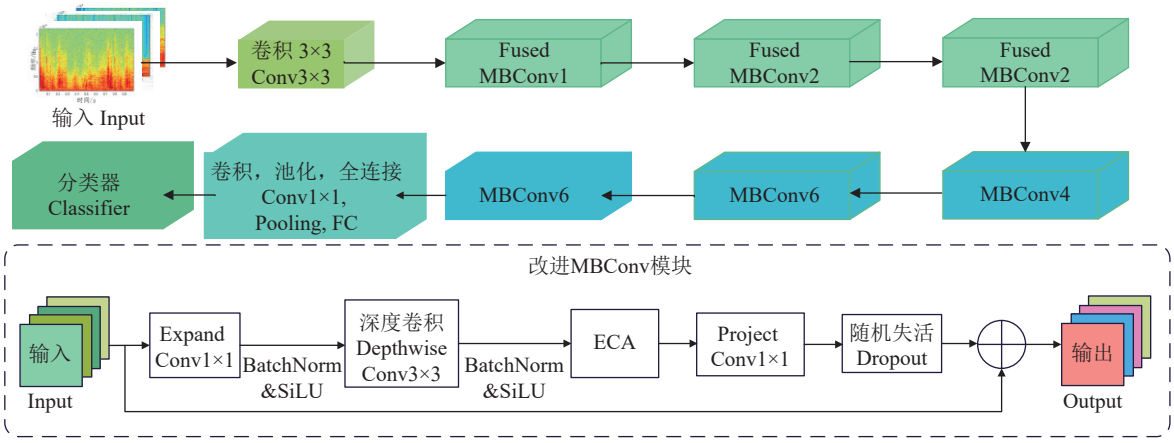


图 8　ECA-EffficientNetV2 网络模型结构图

Fig. 8　Structure of the ECA-EffficientNetV2 network model

根据数据集特征及硬件条件调整网络训练参数，其中 batch-size 调整为 16，学习率选用 0.01，momentum 为 0.9，表 1 中最终阶段随机失活率 dropout-rate 为 0.2，MBConv 卷积结构 Dropout 层随机丢弃率 drop-connect-rate 为 0.2。

1.4.2　网络性能评价指标　采用精确率 (Precision)、准确率 (Accuracy)、召回率 (Recall)、F1 分数 (F1-score)、浮点运算量 (Floating-point operations, FLOPs)、参数量及推理时间评价模型能力，计算公式如下：

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}, \tag{20}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}}, \tag{21}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}}, \tag{22}$$

$$\text{F1-score}=\frac{2\text{TP}}{2\text{TP}+\text{FP}+\text{FN}}, \tag{23}$$

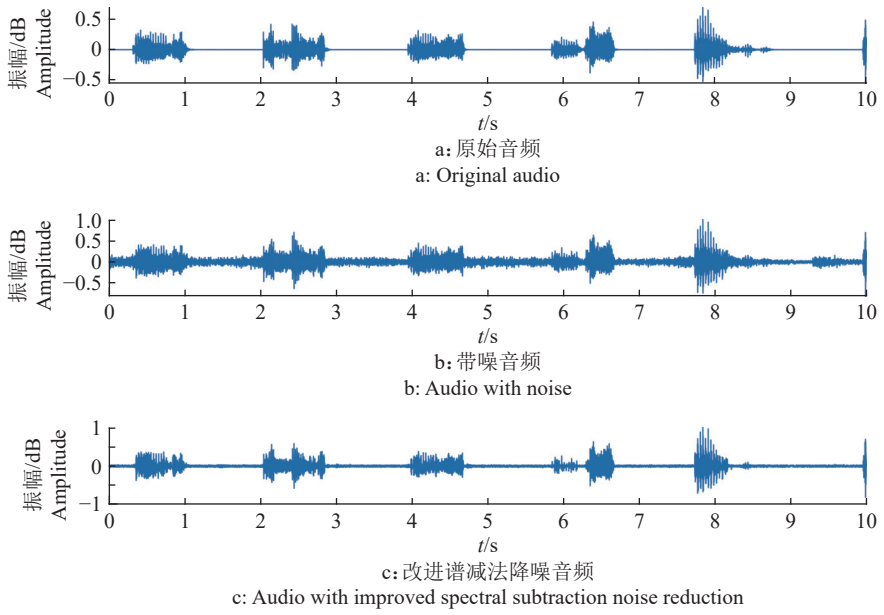
式中，TP 为模型正确预测正例样本数，FP 为模型错误预测正例样本数，FN 为模型错误预测负例样本数，TN 为模型正确预测负例样本数。

2　结果与分析

2.1　改进谱减法降噪及端点检测结果分析

直接采集猪棚内音频，虽可获得真实的环境噪声，但无法有效判断降噪算法对含噪音频的处理效果。为了定量分析降噪算法的降噪效果，在纯净生猪音频信号中添加 IKS 风噪数据集中排风扇的排风声和英国荷兰 TNO 感知研究所语音研究单位发布的金属门、金属围栏等金属碰撞产生的噪声信号，再对含噪音频降噪处理，对比降噪前后信噪比，判断降噪算法效果。信噪比为 15 dB 时，降噪前后如图 9 所示。

将多窗谱估计谱减法和经典谱减法分别在信噪比为 0、5、15 dB 下进行对比试验，前后信噪比如表 2 所示，当降噪前信噪比设为 0 dB 时，改进谱减法较传统谱减法信噪比提升 2.36 dB，较降噪前提升 4.67 dB。当降噪前信噪比设为 5 dB 时，改进谱减法较传统谱减法信噪比提升 1.45 dB，较降噪前提升 2.42 dB。当降噪前信噪比设为 15 dB 时，改进谱减法较传统谱减法信噪比提升 0.37 dB，较降噪前提升 0.58 dB。表 2 数据可看出多窗谱估计的改进



a：原始音频
a: Original audio

b：带噪音频
b: Audio with noise

c：改进谱减法降噪音频
c: Audio with improved spectral subtraction noise reduction

图 9　降噪前后音频波形图对比

Fig. 9　Audio waveform comparison before and after noise reduction

表 2　谱减法与改进谱减法降噪效果对比

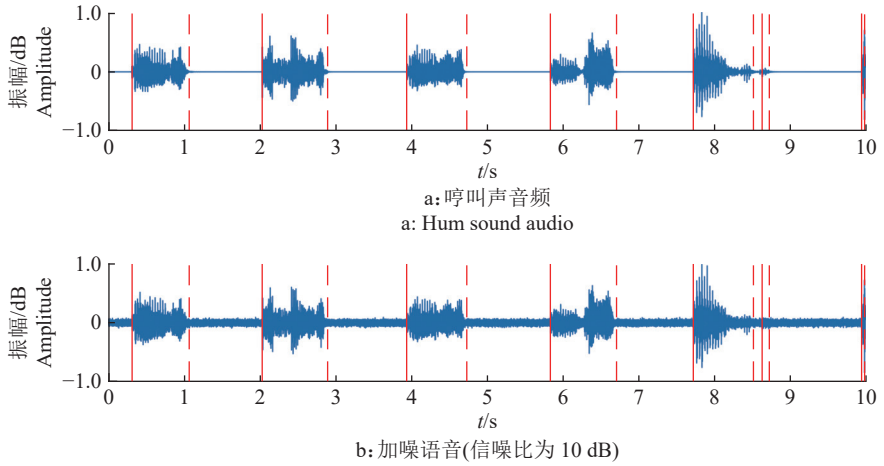
Table 2　Comparison of noise reduction effects of spectral subtraction and improved spectral subtraction

降噪前信噪比/dB SNR before noise reduction	谱减法降噪后信噪比/dB SNR after spectral subtraction	改进谱减法降噪后信噪比/dB SNR after improved spectral subtraction for noise reduction
0	2.31	4.67
5	5.97	7.42
15	15.21	15.58

谱减法较经典谱减法对于带噪生猪音频降噪效果更好，信噪比提升更大。原信噪比越低时，降噪后信噪比提升越明显。原信噪比越高，噪声的干扰越小，降噪后信噪比并无明显提升。

利用能熵比法端点检测含噪生猪音频，结果如图 10 所示。即使在含噪情况下，能熵比法的端点检测仍可检测音频有声段并将其起始点和终止点准确地标出。

若信噪比太低，噪声干扰严重，能熵比法端点检测则无法再准确判断起始和终止点，如图 11 所



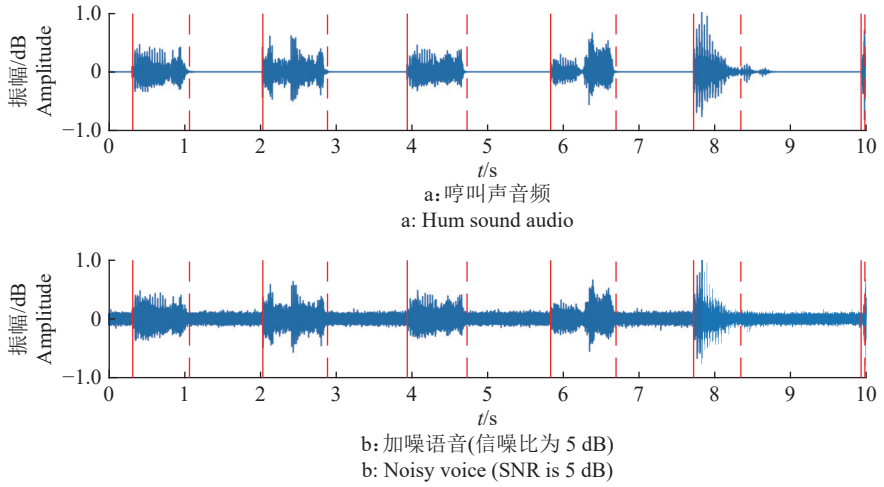
a：哼叫声音频
a: Hum sound audio

b：加噪语音(信噪比为 10 dB)
b: Noisy voice (SNR is 10 dB)

红色实线和虚线分别为当前有声段起始点和终止点
The solid and dashed red lines are the start point and end point of the current audible segment respectively

图 10　能熵比法端点检测波形图

Fig. 10　Waveforms of energy entropy ratio endpoint detection



红色实线和虚线分别为当前有声段起始点和终止点

The solid and dashed red lines are the start point and end point of the current audible segment respectively

图 11　低信噪比下的能熵比端点检测波形图

Fig. 11　Waveforms of energy entropy ratio endpoint detection at low signal-to-noise ratio

示。信噪比为 5 dB 时，第 5 小段音频无法准确判断有声段的起始点和终止点，因此本研究所用端点检测方法适用于高信噪比音频，仅具有一定抗噪性。

2.2　欠定盲源分离结果分析

利用已知的混合矩阵对 4 种生猪音频进行混合，再通过欠定盲源分离算法重构 4 类单只生猪音频，通过对比重构音频和原始音频，对分离算法的性能进行定量评价。本研究以哼叫声、进食声、咆哮声、发情声各 10 s 的单声道音频为试验对象，用已知的混合矩阵混合成 2 个观测信号。4 种状态的生猪音频波形图如图 12 所示。

设定原混合矩阵如下：

$$\boldsymbol{A}=\begin{bmatrix}0.258\,8 & 0.743\,1 & 0.987\,7 & 0.913\,5\\ -0.965\,9 & -0.669\,1 & -0.156\,4 & 0.406\,7\end{bmatrix}。\tag{24}$$

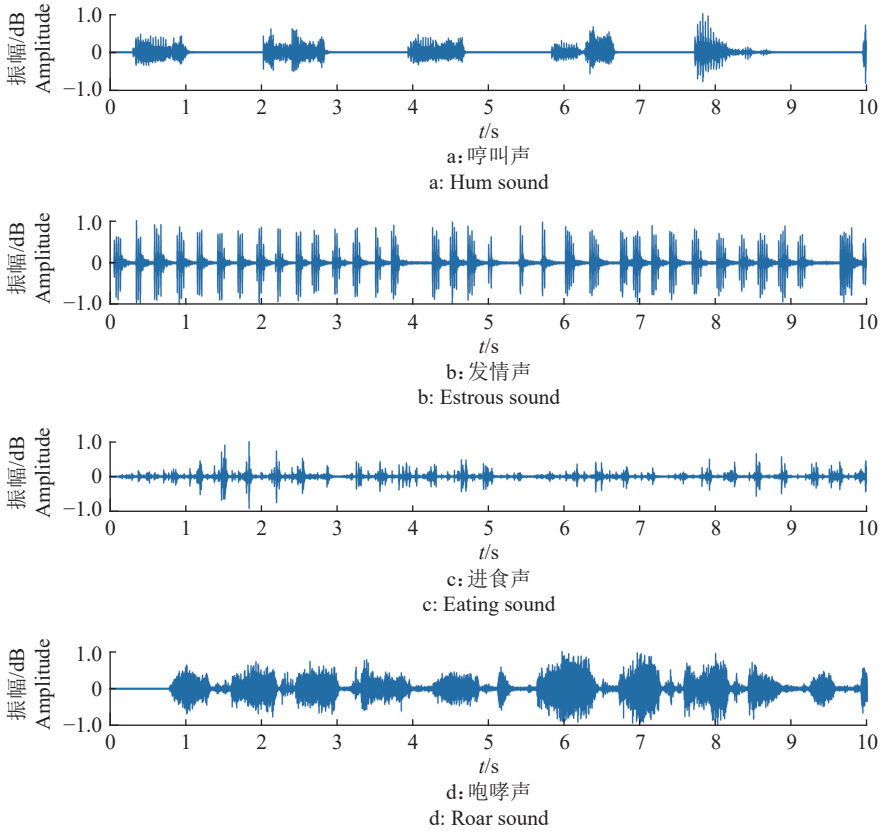


图 12　不同状态的原始生猪音频波形图

Fig. 12　Original audio waveforms of pigs in different states

生成的 2 个观测信号波形图如图 13 所示。

根据“1.3.2”中单源点检测算法，设定式 (13) 中 ε_1=0.01，剔除低能点前后对比图如图 14 所示。检测后的图 14b 可清晰观测到 4 条直线分布特征，且较图 14a 单源点检测前更具稀疏特性。

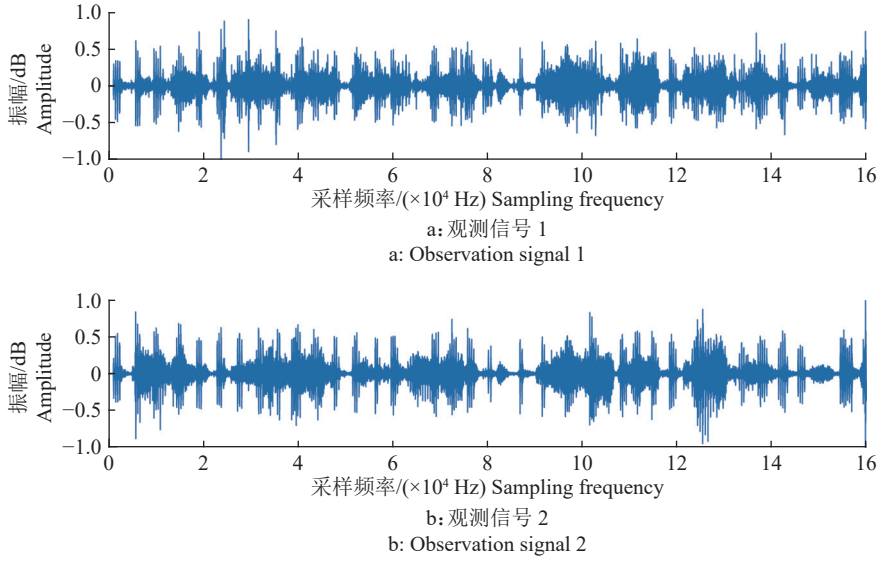


图 13　生猪观测信号波形图

Fig. 13　Waveforms of pig observation signals

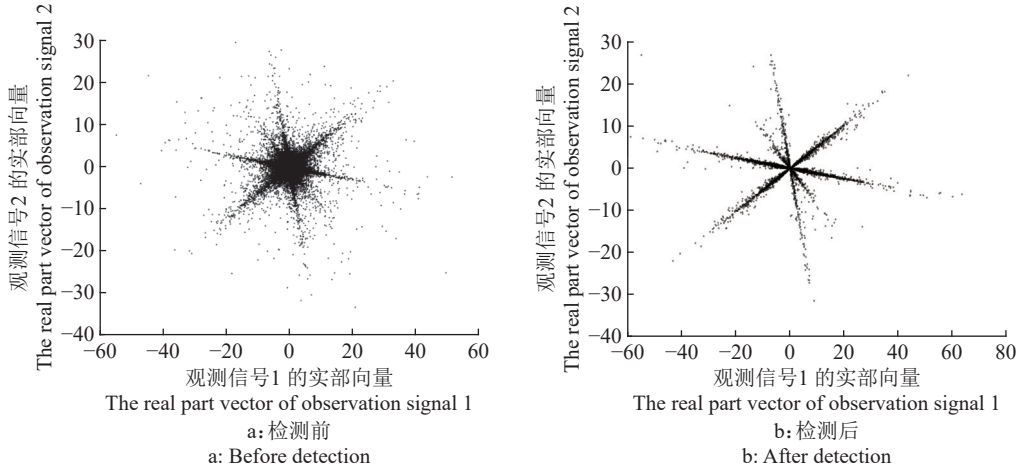


图 14　单源点检测前后的观测信号散点图

Fig. 14　Scatter plots of observation signals before and after single-source point detection

利用层次聚类算法聚类特征点，聚类迭代次数设为 80 轮，用式 (18) 评价聚类后估计的混合矩阵与原矩阵的相似度，值越小则估计精确率越高，如图 15 所示。图 15 表明，归一化均方误差 NMSE 随着迭代次数的增加逐渐降低，在第 44 次后 NMSE 无限接近于 0，取第 63 次时 NMSE 最低值 3.266×10^{-4}，此时估计精确率最高。

图 15 中，可观察到曲线在第 0 到 43 次迭代时产生了振荡，结合图 14b 可看出，有 2 条直线十分靠近，且经过单源点检测后的散点图十分稀疏，随着迭代次数增加层次聚类的距离算法易将 2 条相近线上的特征点聚类成 1 条，从而产生误差，导致曲线产生振荡。

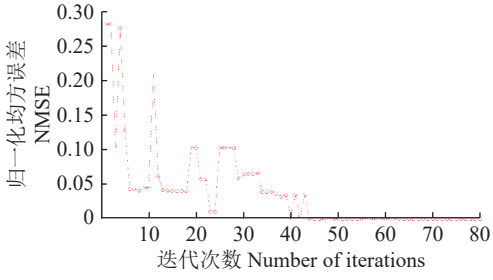


图 15　NMSE 迭代曲线图

Fig. 15　NMSE iteration graph

本研究选用 NMSE 最低值时的混合矩阵进行后续试验，此时混合矩阵如下：

$$\hat{\boldsymbol{A}}=\begin{bmatrix}0.256\,6 & 0.753\,9 & 0.992\,3 & 0.891\,3\\ -0.966\,5 & -0.657\,0 & -0.124\,1 & 0.453\,5\end{bmatrix}。\quad(25)$$

获取混合矩阵后，使用 l_p 范数类算法对信号重构，p 值取 0.5 时重构信号波形图如图 16 所示。

分别计算 4 种波形重构前后信噪比，求其平均值。p 不同取值时，4 种源信号和重构信号平均信噪比在 3.254~4.267 dB 之间变化，p 取 0.3 时平均信噪比最小，即 3.254 dB。p 取 0.8 时平均信噪比最大，即 4.267 dB，此时 l_p 范数分离算法重构的波形最接近源信号波形，分离效果最佳。

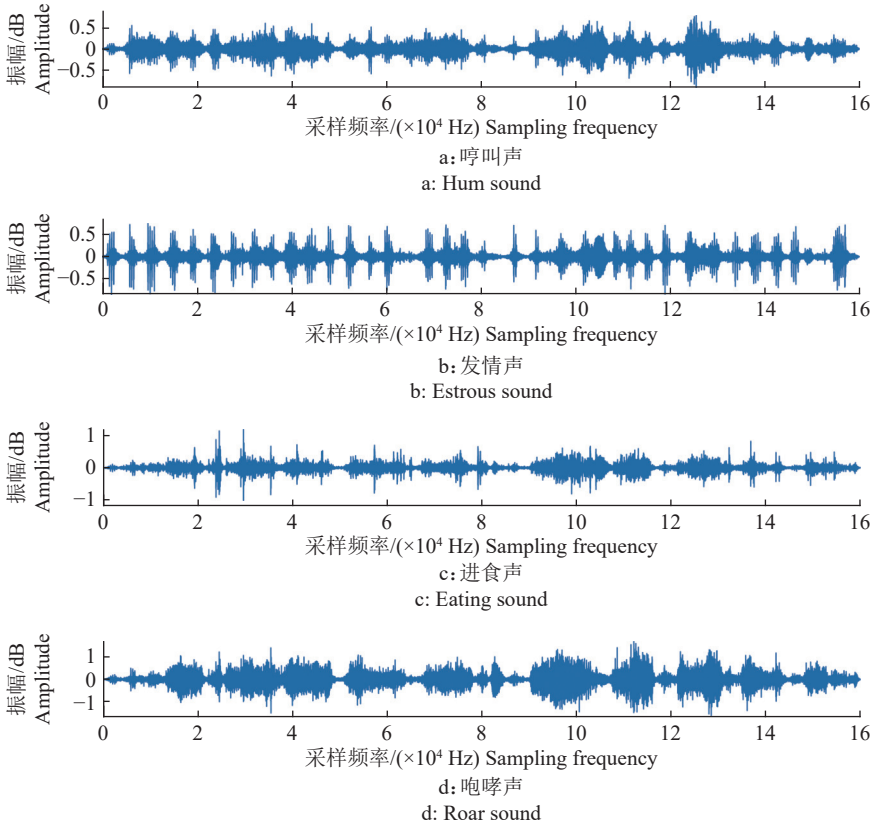


图 16　重构信号波形图

Fig. 16　Waveforms of reconstructed signals

为了对盲源分离效果进行定量分析，本研究并未考虑生猪音频信号到达不同传感器的时延性和衰减性都不同的问题。且由于本研究使用的欠定盲源分离算法依赖信号的稀疏性，当信号的稀疏性较弱时，无法从观测信号获得高质量的重构信号，后续研究中会进一步研究解决上述问题。

2.3　ECA-EfficientNetV2 模型分类结果分析

利用 ECA-EfficientNetV2 网络模型训练声谱图数据集。数据集图片共 2 700 张，进食、哼叫、咆哮 3 类每个类别 720 张，发情类 540 张，数据集以 8∶2 分为训练集和验证集。每轮迭代中训练集与验证集的准确率和损失值结果如图 17 所示。在迭代

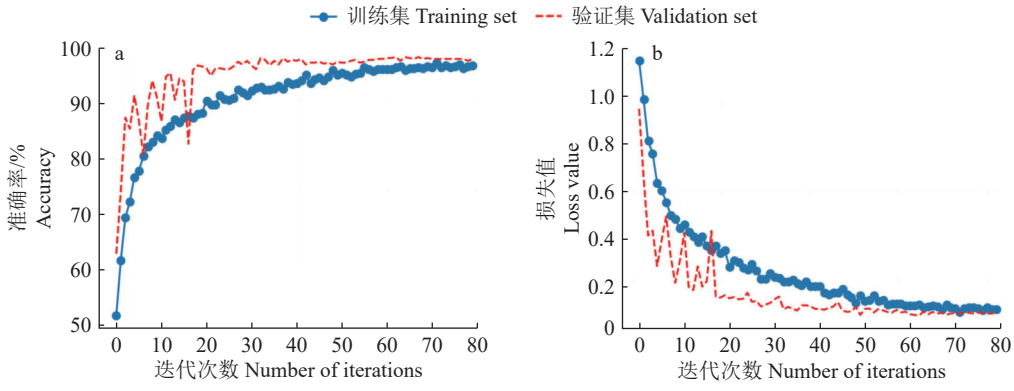


图 17　训练集与验证集的准确率和损失值对比曲线

Fig. 17　Comparison curves of accuracy and loss value between training set and verification set

80 个 epoch 时曲线已经接近平稳, 并无太大的上下波动, 最终准确率在 98% 左右波动, 损失值在 0.08 左右波动。

利用“1.4.2”的评价指标进行定量分析, 使用相同数据集, 设置图像输入大小为 224×224, 迭代次数为 80, 学习率都为 0.01, 与同为卷积神经网络的经典网络模型 ResNet50 和 VGG16 以及原 EfficientNetV2-S 进行对比, 具体结果如表 3 所示。

表 3 数据显示, ECA-EfficientNetV2 相较经典模型 ResNet50、VGG16, 其准确率、精确率等指标都优于经典模型, 与原 EfficientNetV2 模型相比虽准确率下降了 0.52 个百分点, 但模型参数量降低了 33.56%, 浮点运算量 FLOPs 降低了 1.86 G, 平均推理时间减少了 9.40 ms, 使模型更轻量化且推理速度更快, 证明了本研究的生猪声谱图识别方法的有效性, 并为后续应用于边缘节点计算打下基础。但在推理时间上与经典模型 ResNet50、VGG16 相比, 所需时间更长, 表明 EfficientNetV2 网络模型仍需进一步优化。在后续研究中, 将采集更多患病生猪样本进行试验, 并进一步优化网络模型, 为生猪健康智能养殖提供更有意义的技术支持。

表 3 不同模型在生猪声谱图上的预测性能比较

Table 3 Comparison of prediction performance of different models on pig spectrograms

模型 Model	准确率/% Accuracy	精确率/% Precision	召回率/% Recall	F1分数/% F1-score	参数量/M Parameter quantity	FLOPs/G	推理时间/ms Inference time
ResNet50	95.47	95.35	95.56	95.45	25.56	8.21	11.16
VGG16	96.54	96.73	96.58	96.54	138.30	30.97	10.11
EfficientNetV2-S	98.87	98.69	98.86	98.85	21.46	5.73	34.34
ECA-EfficientNetV2	98.35	98.33	98.37	98.35	14.26	3.87	24.94

3 结论

本研究提出一种基于盲源分离算法与 ECA-EfficientNetV2 网络模型相结合的生猪音频状态分类方法, 所得主要结论如下:

1) 欠定盲源分离方面, 研究使用的改进谱减法降噪算法比原谱减法降噪算法的降噪效果更优。聚类得到的混合矩阵估计与原混合矩阵的 NMSE 最低可达 3.266×10^{-4}。l_p 范数重构时, 在 p 取不同值时, 重构出的信号与源信号有不同的差异, 当 p 取值为 0.8 时, 此时重构信号与源信号的差异性最小, 利用信噪比作为评价指标, p 为 0.8 时信噪比为 4.267 dB, 重构信号质量最佳。

2) 生猪音频识别方面, ECA-EfficientNetV2 相较于经典网络模型 ResNet50、VGG16 和原 EfficientNetV2 模型, 具有更轻量化的模型参数, 且准确率也相较 ResNet50 和 VGG16 提高了 2.88 和 1.81 个百分点, 与原 EfficientNetV2 相比准确率降低 0.52 个百分点, 但模型参数量和浮点运算量 FLOPs 与其余模型相比均为最低, 且推理时间较原模型减少 9.40 ms, 为后续应用于边缘节点计算打下基础。

参考文献:

[1] 杨亮, 王辉, 陈睿鹏, 等. 智能养猪工厂的研究进展与展望[J]. 华南农业大学学报, 2023, 44(1): 13-23.

[2] GHANI S H, KHAN W. Extraction of UAV sound from a mixture of different sounds[J]. Acoustics Australia, 2020, 48(3): 363-373.

[3] HE P, QI M, LI W, et al. A general nonstationary and time-varying mixed signal blind source separation method based on online Gaussian process[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2020, 34(11): 2058015.

[4] ADAM A M, FAROUK R M, EL-DESOUKY B S. Generalized gamma distribution for biomedical signals denoising[J]. Signal, Image and Video Processing, 2023, 17(3): 695-704.

[5] 张振华. 基于音频分析技术的猪异常检测[D]. 太原: 太原理工大学, 2017.

[6] 沈明霞, 王梦雨, 刘龙申, 等. 基于深度神经网络的猪咳嗽声识别方法[J]. 农业机械学报, 2022, 53(5): 257-266.

[7] JI N, SHEN W Z, YIN Y L, et al. Investigation of acoustic and visual features for pig cough classification[J]. Biosystems Engineering, 2022, 219: 281-293.

[8] 彭硕, 陶亮, 查文文, 等. 基于稀疏分量分析的生猪音频欠定盲源分离研究[J]. 畜牧兽医学报, 2023, 54(7): 2794-2809.

[9] HU Y, LOIZOU P C. Speech enhancement based on wavelet thresholding the multitaper spectrum[J]. IEEE Transactions on Speech and Audio Processing, 2004, 12(1): 59-67.

[10] 杨稷, 沈明霞, 刘龙申, 等. 基于音频技术的肉鸡采食量检测方法研究[J]. 华南农业大学学报, 2018, 39(5): 118-124.

[11] 赵欢, 王纲金, 赵丽霞. 一种新的对数能量谱熵语音端

点检测方法[J]. 湖南大学学报 (自然科学版), 2010, 37(7): 72-77.
[12] 朱国俊, 唐振博, 冯建军, 等. 启动方式对混流泵噪声特性的影响[J]. 农业工程学报, 2023, 39(13): 34-42.
[13] 杜晓冬, 滕光辉, 刘慕霖, 等. 基于轻量级卷积神经网络的种鸡发声识别方法[J]. 农业机械学报, 2022, 53(10): 271-276.
[14] CAI W, LI M. A unified deep speaker embedding framework for mixed-bandwidth speech data [EB/OL]. arXiv: 2012.00486[2023-12-01]. https://ui.adsabs.harvard.edu/abs/2020arXiv201200486C.
[15] 丁楠. 基于特征学习的语音情感识别研究[D]. 南京: 南京邮电大学, 2023.
[16] CANDAN C. An automated window selection procedure For DFT based detection schemes to reduce windowing SNR loss[EB/OL]. arXiv: 1710.10200[2023-12-01]. https://ui.adsabs.harvard.edu/abs/2017arXiv171010200C.
[17] 张逸, 陈书畅, 刘必杰, 等. 基于盲源分离的工业谐波源负荷分类识别方法[J]. 中国电机工程学报, 2024, 44(10): 3850-3862.
[18] 邹亮, 张鹏, 陈勋. 基于三阶统计量的欠定盲源分离方法[J]. 电子与信息学报, 2022, 44(11): 3960-3966.
[19] 凌康杰, 岳学军, 刘永鑫, 等. 基于移动互联的农产品二维码溯源系统设计[J]. 华南农业大学学报, 2017, 38(3): 118-124.
[20] 王晶, 李炜, 洪心睿, 等. 基于改进密度聚类算法的语音信号欠定盲分离[J]. 信息与控制, 2023, 52(6): 784-796.
[21] KEMIHA M, KACHA A. Single-channel blind source separation using adaptive mode separation-based wavelet transform and density-based clustering with sparse reconstruction[J]. Circuits Systems and Signal Processing, 2023, 42(9): 5338-5357.
[22] WANG M, CAI X X, ZHU K F. Underdetermined mixed matrix estimation of single source point detection based on noise threshold eigenvalue decomposition[C]//Proceedings of the 8th International Conference on Communications, Signal Processing, and Systems. Singapore: Springer, 2020: 704-711.
[23] LI Y B, NIE W, YE F, et al. A complex mixing matrix estimation algorithm in under-determined blind source separation problems[J]. Signal, Image and Video Processing, 2017, 11(2): 301-308.
[24] 冯建英, 石岩, 王博, 等. 基于聚类分析的数据挖掘技术及其农业应用研究进展[J]. 农业机械学报, 2022, 53(S1): 201-212.
[25] XIE Y, XIE K, XIE S L. Underdetermined blind separation of source using l_p-norm diversity measures[J]. Neurocomputing, 2020, 411: 259-267.
[26] XU P F, JIA Y J, JIANG M X. Blind audio source separation based on a new system model and the Savitzky-Golay filter[J]. Journal of Electrical Engineering, 2021, 72(3): 208-212.
[27] FISLI S, DJENDI M. Hybrid PSO-NLMS (HPSO-NLMS) algorithm for blind speech quality enhancement in time domain[J]. Applied Acoustics, 2021, 177: 107936.
[28] TAN M, LE Q V. EfficientNetV2: Smaller models and faster training[EB/OL]. arXiv: 2104.00298[2023-12-01]. https://ui.adsabs.harvard.edu/abs/2021arXiv210400298T.
[29] WANG X K, JIA X, ZHANG M Y, et al. Object detection in 3D point cloud based on ECA mechanism[J]. Journal of Circuits, Systems and Computers, 2023, 32(5): 2350080.

【责任编辑 庄 延】