doi:10.6041/j.issn.1000-1298.2025.07.056

基于音频的肉鸽雌雄智能鉴别模型研究

黄伟锋[1]　陈品岚[2]　卫洁茹[2]　张世昂[3]　付　晶[4]　梁雅妍[4]　朱立学[2]

(1. 仲恺农业工程学院自动化学院, 广州 510225; 2. 仲恺农业工程学院机电工程学院, 广州 510225;
3. 仲恺农业工程学院创新创业学院, 广州 510225; 4. 仲恺农业工程动物科技学院, 广州 510225)

摘要: 肉鸽雌雄鉴别是种鸽养殖、配对过程的关键工作。为实现肉鸽雌雄智能化检测,本研究提出了一种基于肉鸽鸣叫声的雌雄鉴别方法,并设计了基于 EGLKA 模块的肉鸽鸣叫声端点检测模型 EGVAN_VAD 和雌雄鉴别模型 EGVAN。该方法将肉鸽一维音频信号转换为二维的梅尔语谱图作为 EGVAN_VAD 模型的输入数据,检测肉鸽鸣叫声的音频区域,进而使用噪声门限和巴特沃斯低通滤波器过滤环境中的稳态和非稳态噪声,再将降噪后音频的梅尔语谱图作为 EGVAN 模型的输入数据,采用 EGVAN、LSTM、GRU、TDNN、VAN 模型对不同鸽龄(包括 1~3 月、3~6 月和大于 6 月)的肉鸽鸣叫声进行识别。结果显示,EGVAN_VAD 模型的准确率、召回率分别为 93.4%、94.3%,每检测 4 s 时长的音频耗时 10 ms,较其他端点检测模型综合性能最优。EGVAN 模型对肉鸽雌性和雄性鸣叫声的综合识别准确率最高,分别为 90.7%、90.3%,每检测 3 s 时长的音频耗时 14.1 ms。经系统测试,对 3~6 月鸽龄的雌雄鉴别成功率可达 99.5%,对各个鸽龄段平均鉴别成功率为 89.0%。研究表明,EGVAN 对雌雄肉鸽鸣叫声具有优异的识别性能。本研究为利用音频技术实现其他单态鸟类性别的智能化精准识别提供了技术参考。

关键词: 肉鸽; 音频技术; 深度学习; 雌雄鉴别

中图分类号: S836.4; TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)07-0596-12　　**OSID**:

Intelligent Identification Model for Male-female Pigeons Based on Audio

HUANG Weifeng[1]　CHEN Pinlan[2]　WEI Jieru[2]　ZHANG Shiang[3]　FU Jing[4]　LIANG Yayan[4]　ZHU Lixue[2]

(1. *School of Automation, Zhongkai University of Agriculture and Engineering, Guangzhou* 510225, *China*
2. *School of Mechanical and Electrical Engineering, Zhongkai University of Agriculture and Engineering, Guangzhou* 510225, *China*
3. *School of Innovation and Entrepreneurship, Zhongkai University of Agriculture and Engineering, Guangzhou* 510225, *China*
4. *School of Animal Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou* 510225, *China*)

Abstract: Sex identification of pigeons is a crucial task in the breeding and pairing process. To enable intelligent sex identification of pigeons, a method based on pigeon vocalizations was proposed and the endpoint detection model EGVAN_VAD and the sex classification model EGVAN, were developed both based on the EGLKA module. The method transformed one-dimensional audio signals into two-dimensional Mel spectrograms as input to the EGVAN_VAD model for detecting pigeon vocal segments, followed by noise reduction by using a thresholding method and a Butterworth low-pass filter to eliminate both steady-state and transient environmental noise. The denoised Mel spectrogram was then used as input for the EGVAN model, which was compared with LSTM, GRU, TDNN, and VAN models for recognizing pigeon vocalizations across different age groups (1~3 months, 3~6 months, and more than six months). Experimental results indicated that the EGVAN_VAD model achieved an accuracy of 93.4% and the recall of 94.3%, with processing time of 10 ms per 4-second audio segment, outperforming other endpoint detection models in overall performance. The EGVAN model achieved the highest recognition precision for female and male calls, reaching 90.7% and 90.3%, respectively, with processing time of 14.1 ms per 3-second audio segment. For pigeons aged 3~6 months, the identification success rate reached 99.5%, while the average success rate across all age groups was 89.0% following systematic testing. These findings demonstrated that the EGVAN model exhibited excellent performance in

收稿日期: 2024-04-17　修回日期: 2024-06-22

基金项目: 广州市科技计划项目(2023B03J0862)

作者简介: 黄伟锋(1979—),男,副教授,主要从事视觉检测技术研究,E-mail: wfhuang@zhku.edu.cn

通信作者: 朱立学(1970—),男,教授,主要从事智能农业装备研究,E-mail: zhulixue@zhku.edu.cn

recognizing both male and female pigeon calls. The research result can provide a technical reference for applying audio-based methods to intelligent and accurate sex identification in other monomorphic bird species.

Key words: pigeons; audio technology; deep learning; male-female identification

0　引言

肉鸽属于单态鸟，白羽肉鸽的体型和羽毛颜色没有明显的性别差异。肉鸽属于一夫一妻制的鸟类，鸽群中若雌雄比例失衡，会对产蛋率和受精蛋质量产生负面影响。因此，准确鉴别肉鸽雌雄对成功实施种鸽配对、肉鸽繁殖、饲养和销售有至关重要的意义。除了通过人工饲养经验进行粗略的雌雄鉴别，肉鸽产业目前可用的雌雄检测技术有泄殖腔检测、行为特征检测和 PCR 基因检测等[1-4]，但上述技术存在一定的缺点，泄殖腔检测是一种入侵式性别检测方法，会导致肉鸽出现强烈的应激反应，操作技术要求高且损害动物福利；行为特征检测不仅高度依赖检测人员的技术水平，还费时费力；PCR 基因检测技术成本高，抽血过程会对肉鸽造成伤害，且须配备专门设备进行性别分析，不便在肉鸽产业大规模应用[5-6]。

鸣叫声作为动物之间信息传递方式，能够反映动物生存状态的多种信息，进而可分析动物的生长和健康状况[7]。基于机器学习算法的动物声音信息识别技术主要利用深度学习和信号处理等方法，对动物声音进行分析，来实现自动分类识别[8-13]。基于声音的非侵入性雌雄鉴别方法已被证明对许多单态鸟类有效[1,14-16]。SADEGHI 等[17]采用高斯径向基函数的支持向量机（SVM）对鸡鸣信号进行分类，通过对时间、频率和时频域等多个特征值进行公鸡和母鸡的识别分类，研究发现，使用时频域特征的分类准确率最高达到 90.74%。LI 等[18]采用基于 ResNet－50 神经网络的方法对黄雏鸡的叫声进行性别检测，并比较了语谱图、倒谱图和 MFCC + Logfbank 3 种音频特征。结果显示，将 MFCC + Logfbank 音频特征与 ResNet－50 网络相结合的方法测试准确率为 83%。YIN 等[19]从幼鸭发声中提取梅尔语谱图特征作为 Res2Net 模型输入，并在模型中引入 SE 注意力机制和 Ghost 模块，通过改进后的模型进行性别分类，最终模型的准确率达到 96.67%。

目前，以音频数据作为输入数据进行肉鸽雌雄鉴别的研究极少。针对雌雄肉鸽体型、外观雌雄特征不明显、生产中不易识别的问题，本研究提出一种基于音频特征和深度学习技术的肉鸽雌雄非侵入式鉴别方法。通过采集肉鸽日常鸣叫的音频数据，构建一个包含雄性和雌性个体的音频数据集，并提取其梅尔语谱特征图作为改进的视觉注意力网络（Visual attention network，VAN）[20]模型的输入。通过进一步比较不同音频特征的神经网络模型性能，分析不同鸽龄肉鸽鸣叫声的特征，选择最优模型组合，并采用效果最佳的检测方法开展肉鸽雌雄鉴别试验。

1　材料与方法

1.1　试验设置

试验数据采集于仲恺农业工程学院海珠校区实验室，采集目标为鸽龄大于 1 个月的肉鸽，品种为白羽王鸽，共采集了 52 只肉鸽鸣叫声音数据，其中雌性 28 只，雄性 24 只。肉鸽被放入隔音箱中的鸽笼采集鸣叫声，其真实性别均已采用聚合酶链式反应（Polymerase chain reaction，PCR）技术完成了测定。数据采集时间为 2023 年 10—11 月，每日采集时间为 09:00—17:00。

隔音箱长×宽×高为 2.5 m×1.5 m×2.2 m，壁厚 10 cm，被分隔为大小相同的两个空间。箱壁由隔音板、阻尼隔音毡、吸音棉、聚酯纤维隔音毡及吸音布等材料制成，可有效隔离外部噪声。隔音箱安装有静音风机和暖白色 LED 阵列光源，以保持空气流通以及和外界相似的养殖环境。使用联想 B601 型数码录音笔和 RODE VideoMicro Ⅱ型枪式麦克风作为录音设备，采样率为 48 000 Hz，位深度为 16 位。根据肉鸽生产阶段的划分，建议分为 1～6 月鸽龄的育成期（其中 5～9 周鸽龄的童鸽期，10～25 周鸽龄的青年鸽期）和大于 6 月鸽龄的产鸽期（成年鸽期）[21]。在声音采集过程中发现，1～3 月鸽龄肉鸽处于从“吱”变为“鸣”和“咕”的变声期。因此试验中，52 只肉鸽根据不同的鸽龄和性别分为 6 组，分别为变声期（1～3 月）、青年鸽期（3～6 月）与成年鸽期（大于 6 个月）。每组内再按照 1、2、4 只肉鸽进行组合，采集不同数量饲养情况下肉鸽的鸣叫声。音频采集系统如图 1 所示。

1.2　肉鸽声音数据集构建

1.2.1　肉鸽鸣叫声端点检测数据集构建

肉鸽音频数据通常包含各种活动语音和非活动语音，当标注肉鸽鸣叫声端点检测数据集时，仅需要

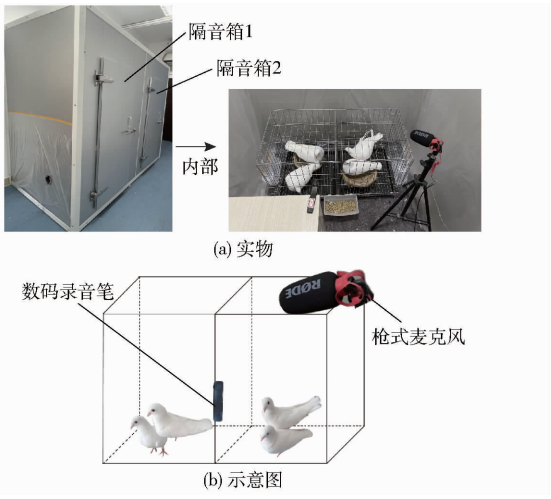


图 1 音频采集系统

Fig. 1 Audio acquisition system

对音频信号中的肉鸽活动语音进行标注,即标注具有肉鸽鸣叫声的区域部分。在标注过程中,使用 WavePad 音频编辑软件查看鸣叫声波形,然后标注每一段肉鸽鸣叫声(需要标注鸣叫声活动区域的起始位置和结束位置)。对于非活动语音的区域片段,如静默、噪声等,则不需要进行标注。在数据选取方面,肉鸽鸣叫声重叠与不重叠情况均作为数据集样本。当肉鸽叫声重叠时,端点检测数据集标注重叠鸣叫声的并集区域,即从鸣叫声开始到最后的肉鸽鸣叫声结束。最终标注了肉鸽鸣叫声数据集音频 4 137 段,其中训练集 3 525 段,测试集 612 段,标注数据保存为 json 文件格式。

1.2.2 肉鸽鸣叫声雌雄分类数据集构建

试验所用 52 只肉鸽中,鸽龄 1 ~ 3 月、3 ~ 6 月以及大于 6 月的 24 只肉鸽鸣叫声被用作训练集和验证集。其余 28 只肉鸽鸣叫声被用作测试集,其中 1 ~ 3 月和 3 ~ 6 月鸽龄的雌性肉鸽数量各为 5 只,大于 6 月鸽龄的雌性肉鸽数量为 6 只,不同鸽龄雄性肉鸽数量均为 4 只。不同鸽龄的雌性和雄性肉鸽鸣叫声数据集见表 1。在训练集和验证集中,1 ~ 3 月、3 ~ 6 月以及大于 6 月鸽龄的雌性和雄性肉鸽鸣叫声总样本数量分别为 1 270、1 650,将鸣叫声样本以比例 4∶1划分为训练集和验证集。另外,1 ~ 3 月、3 ~ 6 月以及大于 6 月鸽龄的雌性肉鸽叫声测试集样本总数量为 1 264,雄性为 1 473。

在音频采集过程中,本研究关注雌性和雄性肉鸽的鸣叫声以及肉鸽日常的扇翅声和采食声,根据声音类型分为目标声音和非目标声音,其中目标声音为雌性、雄性肉鸽鸣叫声两个类别(“雌性”和“雄性”),非目标声音(扇翅声和采食声)定义为一个类别(“其他”)。加入“其他”类别是为了降低非目标声音对目标声音识别的影响。因此,数据集添加了 1 400 个属于“其他”类别的声音样本作为训练集和验证集,再单独选取 1 000 个声音样本作为测试集。在正常笼养肉鸽鸣叫声的数据采集中发现,肉鸽初始鸣叫声的振幅较小,随后逐渐增大,一般在 2 s 后叫声振幅达到平稳,因此每个鸣叫声样本时长截取为 3 s。图 2 为肉鸽声音数据集音频归一化后的波形图。如图 2a、2d 和图 2b、2e 所示,雌性和雄性肉鸽均会发出“鸣”和“咕”两种不同但有规律的鸣叫声,无法直接通过人的听觉或观察声波图判断肉鸽的雌雄。因此,本研究采用了音频特征提取和深度学习的方法对肉鸽鸣叫声进行分类。

表 1 肉鸽鸣叫声数据集

Tab. 1 Data set of pigeon chirping

鸽龄/月	性别	数据	肉鸽数量/只	鸣叫声数量
1 ~ 3	雌性	训练集 + 验证集	4	463
		测试集	5	434
	雄性	训练集 + 验证集	4	492
		测试集	4	343
3 ~ 6	雌性	训练集 + 验证集	4	366
		测试集	5	368
	雄性	训练集 + 验证集	4	501
		测试集	4	377
>6	雌性	训练集 + 验证集	4	441
		测试集	6	460
	雄性	训练集 + 验证集	4	657
		测试集	4	751

1.3 音频特征雌雄鉴别系统流程

本研究基于改进的视觉注意力网络识别肉鸽鸣叫声进行雌雄分类,系统通过 2 个进程并行运行识别,主要流程如图 3 所示。系统进程 1 对输入的音频进行梅尔特征提取,然后通过鸣叫声活动区域端点检测模型获取肉鸽鸣叫声区域。进程 1 通过队列的方式将端点检测结果和音频传输给进程 2,当进程 2 根据进程 1 的传输结果检测到肉鸽鸣叫声时,截取包含鸣叫声区域的 3 s 音频进行降噪和滤波处理,得到较为纯净的肉鸽鸣叫声音频段。进一步提取滤波后音频的梅尔特征作为肉鸽声音特征雌雄鉴别模型中,最后输出肉鸽性别。

1.4 预处理

1.4.1 降噪

在音频采集过程中,风机和肉鸽日常活动会产生一定的噪声。风机噪声振幅通常小于 100,并且分布在各个频段,而肉鸽鸣叫声的频率较低,低于 1 000 Hz,如图 4 所示。

为了降低噪声对模型性能的影响,采用噪声门限和低通滤波器对原始音频进行降噪处理。设置噪声门限是为了过滤环境中的稳态噪声,其过

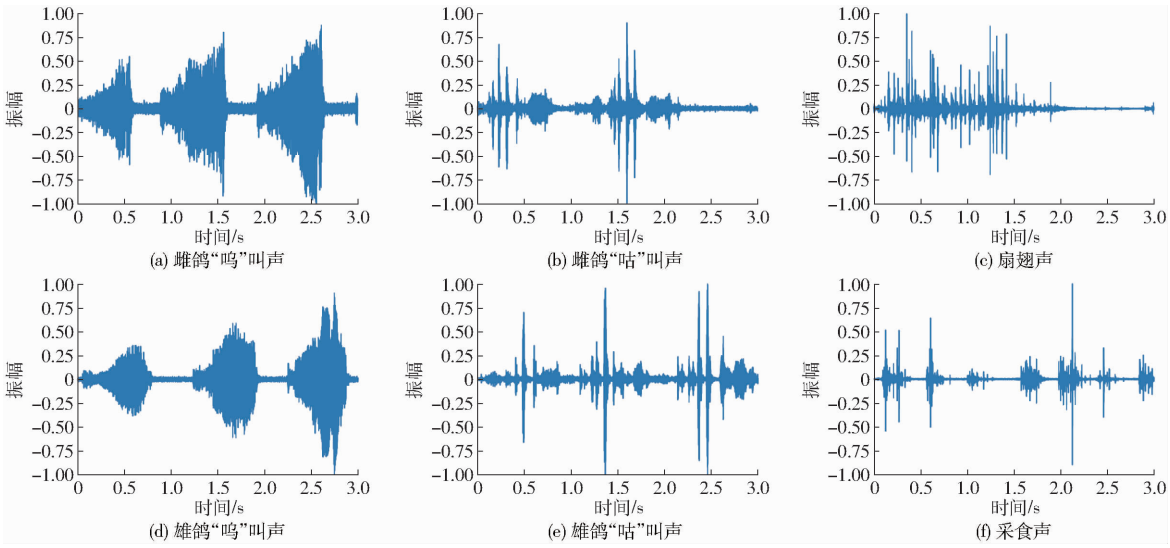


图 2　数据集音频波形图

Fig. 2　Audio waveform of data set

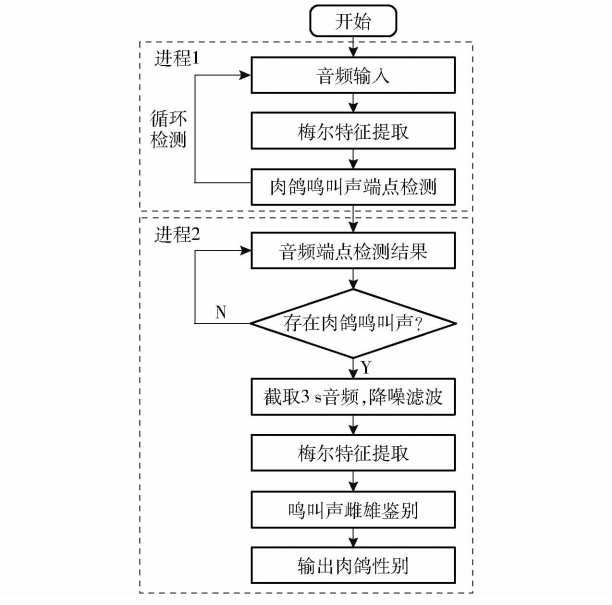


图 3　肉鸽声音特征雌雄鉴别系统流程图

Fig. 3　Flowchart of male-female identification of sound characteristics for pigeons

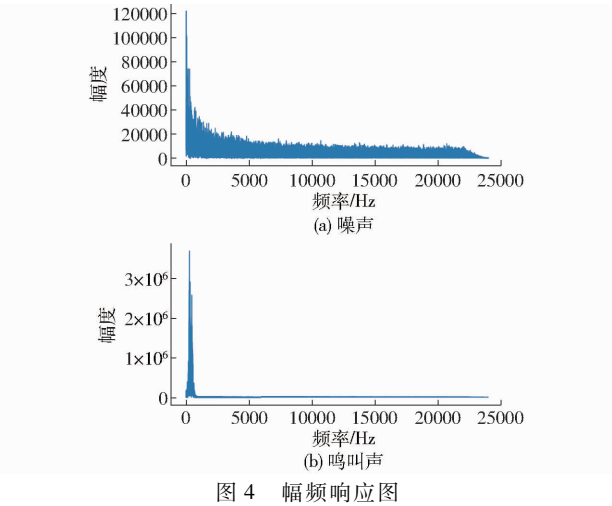


图 4　幅频响应图

Fig. 4　Amplitude-frequency response graphs

滤方程为

$$y(n)=\begin{cases} y(n) & (y(n)\geqslant Y) \\ \left[\dfrac{y(n)}{10}\right] & (y(n)<Y) \end{cases} \tag{1}$$

其中　$Y=$

$$\begin{cases} \max(y_m) & (\max(y_m)<100) \\ 100 & (\max(y_m)\geqslant 100) \end{cases} \quad (m=\mathrm{index}(\min(E_i))) \tag{2}$$

$$E_i=\frac{\sum_{k=1}^{N} y_i(k)^2}{N} \quad (i\in(1,2,\cdots,S)) \tag{3}$$

式中　$y(n)$——原始音频振幅

n——采样点

Y——噪声门限阈值

E_i——非语音音频段的能量均值

y_i——非语音音频段 1/4 至 3/4 音频区域

N——y_i的采样点数

S——非语音音频段数量

y_m——最小能量均值所对应的音频段

无限脉冲响应(Infinite impulse response,IIR)滤波器更加适合在计算机中进行设计和实现[18]。本研究通过构建截止频率为 1 000 Hz 的巴特沃斯低通滤波器,以巴特沃斯低通滤波器的输出作为 IIR 滤波器多项式系数向量,并将噪声门限过滤后的音频输入到 IIR 滤波器进一步降噪[22]。

1.4.2　分帧与窗口

肉鸽鸣叫声作为一种非平稳信号,从整段波形中直接分析其音频特征比较困难,因此需要结合频域信息进行分析。由于短时间(10 ~ 30 ms)内声音信号具有短时平稳性[23-24],为了在频域上更好地表

示和分析肉鸽鸣叫声信号,采用分帧、加窗及短时傅里叶变换(Short time Fourier transform,STFT)的方法提取语谱特征。本研究使用的分帧长度为 21.3 ms,即 1 024 个采样点,帧移为 320 个采样点,对 3 s 的音频进行分帧共获得 451 帧。分帧后,由于其 STFT 结果不是理想的离散频谱,会存在波形间断和频谱泄漏。因此,需要添加窗函数在帧的开始和结束处提供平滑过渡,减少由于截断导致的频谱泄漏。本研究使用窗口长度与分帧长度一致的汉明窗和汉宁窗函数,并比较无窗口、汉明窗和汉宁窗的肉鸽鸣叫声语谱图,如图 5 所示。由图 5 可以看出,图 5a 中没有使用窗口函数的语谱图,其帧与帧之间的语谱特征平滑性差,图 5b 中使用汉明窗函数的语谱图有明显改善,图 5c 中使用汉宁窗函数的语谱图平滑性最优。因此,本研究使用汉宁窗作为肉鸽叫声语谱特征的窗口函数。

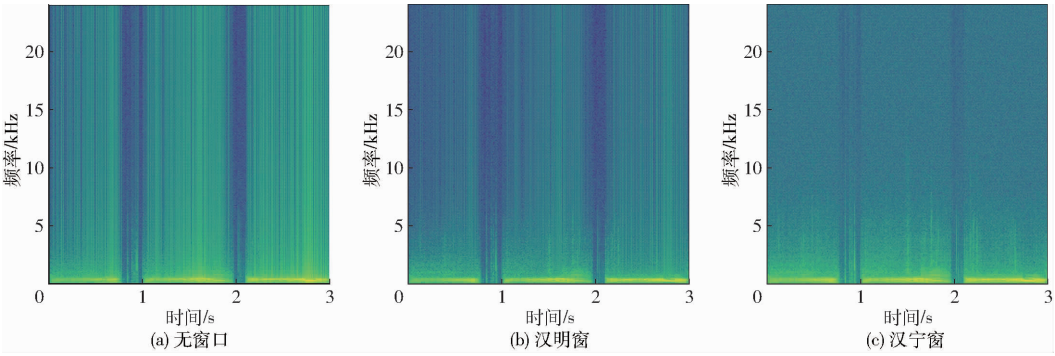


图 5 语谱图对比

Fig. 5 Comparison of spectrograms

1.4.3 梅尔语谱图特征

肉鸽音频经过 STFT 特征提取后会生成高维度的语谱图。为了降低肉鸽鸣叫声识别模型的计算复杂度、减少语谱图的冗余信息,本研究使用 64 维的梅尔滤波器组对语谱图进行维度降低和特征选择,将信号中的能量集中在关键的频率区域,同时降低其他频率的能量。根据梅尔计算,利用语谱图和梅尔滤波器组获取梅尔语谱特征,即

$$S_{i,m}=\sum_{k}^{F}(X_i(k))^2H_m(k) \tag{4}$$

式中 $S_{i,m}$——分帧后第 i 帧频谱与第 m 个梅尔滤波器的梅尔语谱特征输出

$X(k)$——频谱幅度谱

$H_m(k)$——第 m 个梅尔滤波器的频谱响应

F——频谱数量

最后可获得 451 × 64 梅尔语谱特征的二维矩阵,梅尔语谱特征经过对数变换所获得的梅尔语谱图如图 6 所示。

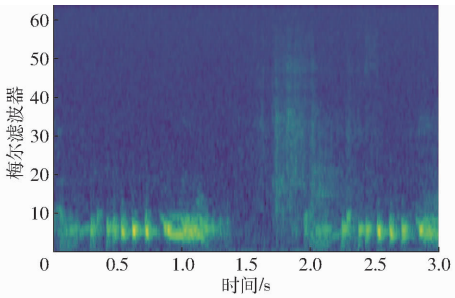


图 6 肉鸽叫声梅尔语谱图

Fig. 6 Meir phonogram of pigeon calls

1.5 肉鸽声音识别模型

经过上述处理和特征提取,可将肉鸽鸣叫音频转换为二维梅尔语谱图特征。即对肉鸽音频进行分类任务转换为对音频梅尔语谱图进行分类的任务。注意力机制可以提高处理图像信息的有效性[25],能够有选择性地聚焦音频的梅尔语谱特征信息,优化分类任务需求。本研究以 VAN 为基础,构建用于肉鸽鸣叫音频分类的模型。VAN 为图像处理任务构建大内核注意力(Large kernel attention,LKA),LKA 融合了卷积和自注意力的优势,充分考虑了局部结构信息、远程特征依赖性以及适应性[20]。VAN 中最主要两个部分为 LKA 和多层感知器(Multilayer perceptron,MLP),而 MLP 容易导致在训练集上结果良好,但泛化能力较差出现过拟合情况,并且对图像中的局部特征和空间关系处理能力较差。故本研究剪裁了 MLP 模块,主要采用 VAN 中的 LKA 模块。以肉鸽鸣叫声的整个梅尔语谱图作为类别特征,而模型在提取特征的过程中需要进一步关注图像全局特征信息。为此,本研究在 LKA 后面引入有效全局注意力机制(Effective global attention mechanism,EGAM)模块,并基于 LKA 与 EGAM 构建了 EGLKA 模块,进一步构建一种简单有效全局视觉注意力网络(EGVAN)。EGVAN 具有简单的层次结构,有 4 个阶段并且每个阶段使用一个 EGLKA,输出空间分辨率不断降低,即$\frac{W}{2}\times\frac{H}{2}$、$\frac{W}{4}\times\frac{H}{4}$、$\frac{W}{8}\times\frac{H}{8}$和$\frac{W}{16}\times\frac{H}{16}$,其中 W 和 H 表示梅尔语谱图像的高度和宽度。随

着分辨率降低，输出通道数也在增加，4 个阶段输出通道数分别为 64、128、256、512。对每个阶段的输入进行下采样，并使用步长来控制下采样率。在下采样之后，每一个阶段中其他所有模块输出保持相同的空间分辨率和通道数。最后，网络对第 4 阶段的输出结果先使用平均池化以调整特征维度，再使用全连接层输出最终识别结果（图 7）。网络使用 Adam 优化器进行训练。

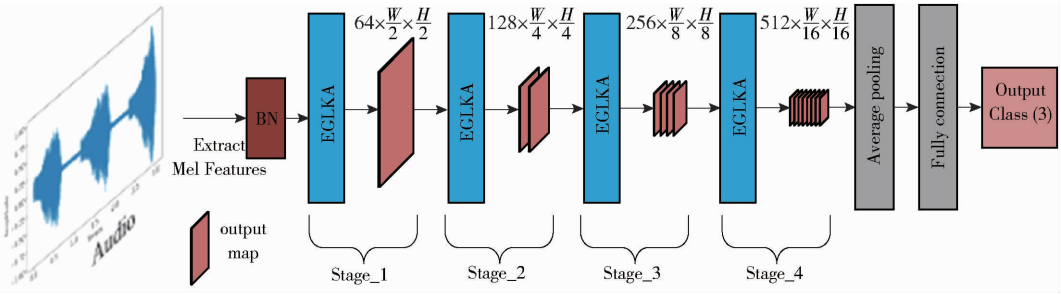


图 7　EGVAN 网络结构图

Fig. 7　EGVAN voice recognition network structure

1.5.1　EGAM 模块

EGAM 模块由 ESE[26] 与 GAM[27] 注意力机制构成，如图 8 所示。ESE 采用单个卷积层实现通道特征的提取和加权，减少了计算复杂度。与 SE 相比，其有效避免了由于维度变换导致的信息损失。GAM 通过结合通道和空间注意力机制，增强对全局特征的提取，更准确地理解长距离特征的空间依赖性。其中，通道注意力采用双层 MLP 和缩减率为 r 的编码器-解码器结构来增强跨维度通道的空间依赖性，但特征维度的缩减可能会导致通道特征丢失。因此，本研究使用 ESE 作为 GAM 的通道注意力部分，相较于 MLP 需要通过较为复杂的结构来学习通道间的特征依赖关系，ESE 直接在特征图的通道维度上进行操作，有效避免了在维度变换中可能发生的特征损失。在空间注意力方面，通过使用两个卷积层来融合空间特征信息，为了避免参数显著增加，将 7×7 卷积层替换为尺寸更小的 3×3 卷积层，从而提高网络实时性。

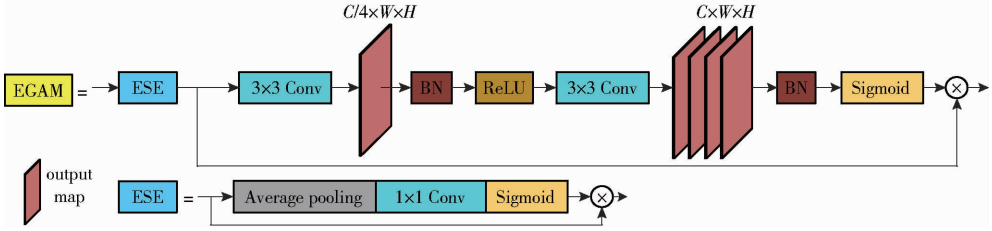


图 8　EGAM 模块基本组成部分

Fig. 8　Basic components of EGAM module

1.5.2　EGLKA 模块

大的核卷积注意力具有更强的特征提取能力，但会带来大量计算量和参数。为了克服此缺点，可通过分解单个通道特征图的大核卷积为小核卷积和膨胀卷积。如图 9 所示（图中橙色表示卷积核大小，黄色表示卷积核中心，灰色表示使用 0 填充，粉红色表示卷积后的特征，蓝色表示特征），对于单个通道特征图 5×5 标准卷积被分解为 3×3 卷积、3×3 膨胀率为 2 的膨胀卷积。一个 $K\times K$ 的大核卷积，可以分解为 $(2d-1)\times(2d-1)$ 的小核卷积、$\left[\frac{K}{d}\right]\times\left[\frac{K}{d}\right]$ 膨胀率为 d 的膨胀卷积，其中[ ]为向上取整符号。通过上述分解能在计算成本和参数较低的情况下获取远程特征之间的联系，并对点的重要性进行估计。

根据深度可分离卷积，一个具有多通道的特征图标准卷积可分解为一个深度卷积和一个点卷积。LKA 结合了单通道大的核卷积分解与深度可分离卷积的方法，将多通道特征图大核标准卷积分解为空间局部卷积（深度卷积，DW－Conv）、空间远程卷积（深度膨胀卷积，DW－D－Conv）和通道卷积（点卷积，1×1 Conv）3 个部分。因此，LKA 模块可以写为

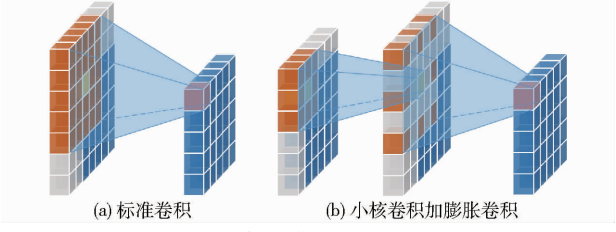


图 9　单通道大核卷积分解

Fig. 9　Single-channel large kernel convolution decomposition

$$\text{Output}=\text{Conv}_{1\times1}(\text{DW}-\text{D}-\text{Conv}(\text{DW}-\text{Conv}(F)))\otimes F \tag{5}$$

式中 F——输入特征图

当卷积核为 K、膨胀率为 d、通道数为 C 时，模块参数量 P 为

$$P(K,d)=(2d-1)^2C+\left[\frac{K}{d}\right]^2C^2+C^2 \quad (6)$$

本研究使用 $K=21$、$d=3$ 的 LKA 模块，对应分解为 5×5 深度卷积和 7×7 膨胀率为 3 深度膨胀卷积，因此参数量为 $25C+50C^2$，远小于 $K=21$ 标准卷积的参数量（$441C^2$）。EGLKA 结合 EGAM 和 LKA 模块（图 10），依次通过下采样、批量归一化、1×1 Conv、GeLU 激活函数、LKA 和 EGAM 来提取特征。

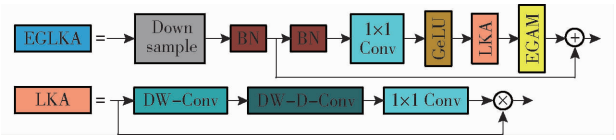


图 10 EGLKA 模块基本组成部分

Fig. 10 Basic components of EGLKA module

1.6 肉鸽鸣叫声端点检测模型

语音端点检测（Voice activity detection，VAD）是否准确，将很大程度上直接影响系统的识别准确率[28]，故肉鸽鸣叫声端点检测准确性将直接影响声音识别模型对雌雄肉鸽鸣叫声的识别结果。而音频采集设备在收集肉鸽鸣叫声数据集时，会同时获取其他声音和大量无声片段。为减小数据量、提高信噪比、增强系统的自动化程度，需要使用 VAD 从音频中准确提取肉鸽鸣叫声音区间。常用的 VAD 有基于短时能量和过零率的双门限检测[29]、RNN[30]和 DNN[31]等。基于短时能量和过零率的双门限检测对于非稳态噪声信号环境 VAD 的结果较差。由于肉鸽日常活动引起的非稳态噪声信号也具有较高短时能量，因此，使用对低信噪比环境适应性更强的深度学习 VAD 更具鲁棒性，在稳态和非稳态的噪声环境中均有较好的检测效果。

为准确检测肉鸽鸣叫声端点，本研究采用了 EGLKA 构建肉鸽鸣叫声的 VAD 网络（EGVAN_VAD）。与 EGVAN 相比，EGVAN_VAD 具有更加简单的层次结构，其减少了 Stage_3 和 Stage_4 部分的网络结构，在 Stage_2 输出结果后，使用与 EGVAN 中 Stage_4 输出连接一样的平均池化、全连接层。但由于 EGVAN_VAD 是用于检测肉鸽鸣叫声，因此模型输入和输出的特征维度结构与 EGVAN 不一致。自注意力（Self-attentive）VAD 网络[32]通过以梅尔特征图的一列作为特征中心，进一步在特征中心的左右两边各取 3 列对称特征构建 7×梅尔滤波器组维度的网络特征输入矩阵，不同的列均以此方式构建为输入特征。本研究 EGVAN_VAD 采用与 Self-attentive VAD 相同处理方式的特征输入，其中使用了 64 维的梅尔滤波器组，因此肉鸽音频端点检测梅尔特征图中的一列输入特征为 7×64 的二维矩阵。从输入特征维度可以得出，当特征经过 Stage_2 输出特征为 2×16，此特征 W、H 均已小于 LKA 模块中的核卷积大小。若在模型后再添加 Stage_3 和 Stage_4 会限制网络对输入特征的提取能力，并大幅增加模型参数量，因此网络减少了 Stage_3 和 Stage_4 部分。而 EGVAN_VAD 在全连接层后使用 LogSoftmax 调整数据，获得最终输出的端点检测结果，如图 11 所示。

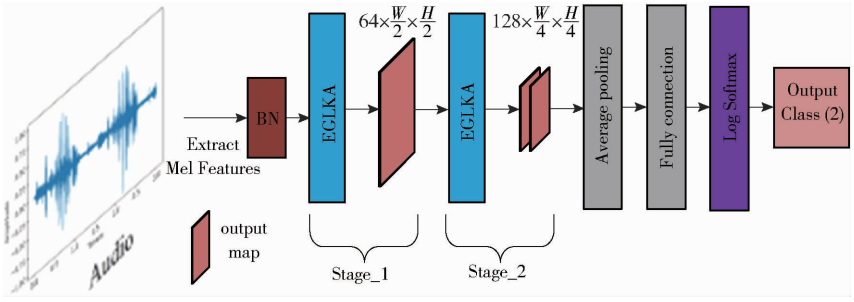


图 11 EGVAN_VAD 鸣叫声端点检测网络结构图

Fig. 11 EGVAN_VAD call endpoint detection network structure

2 结果与分析

2.1 试验环境

模型训练所用硬件配置为 Intel（R）Xeon（R）Gold 5218 CPU @ 2.30 GHz 处理器、Quadro RTX 5000 显卡 16 GB 显存。操作系统为 Ubuntu18.04，配置 CUDA 10.2 和 CUDNN 8.6.0，在系统安装的 Anaconda 虚拟环境内进行模型训练，环境内安装有 PyTorch 1.8.1、Scipy 1.7.3 等依赖库，编程语言使用 Python 3.7。

2.2 评价指标

在肉鸽鸣叫声端点检测测试中，为评价 EGVAN_VAD 模型性能及有效性，使用准确率（A）、精确率（P）、召回率（R）作为性能评估指标。

在肉鸽声音识别测试中，使用雌雄鸣叫声识别准确率（P_v）作为性能评估指标。最终根据 EGVAN_

VAD 检测为肉鸽鸣叫声且正确的音频段数与 EGVAN 成功鉴别肉鸽雌雄的音频段数计算雌雄鉴别成功率(MF_{dsr})。

2.3　肉鸽鸣叫声端点检测结果

EGVAN_VAD 鸣叫声端点检测模型训练参数：学习率设置为 0.001，批量大小设置为 256，训练轮次为 100，选择准确率最高的权重文件作为模型权重。

为评价所构建的 EGVAN_VAD 模型对肉鸽鸣叫声端点检测性能，建立了不同肉鸽鸣叫声端点检测模型的对比试验，比较了 EGVAN_VAD、增强型深度神经网络(BDNN)、单声道音频分离的注意力感知卷积架构(ACAM)、Self-attentive VAD 和 VAN 模型对肉鸽鸣叫声端点检测性能。5 种不同的端点检测模型性能结果见表 2，其中检测时间为模型检测 4 s 音频特征所需的时间。

表 2　不同端点检测模型性能结果对比

Tab.2　Comparison of performance results for different endpoint detection models

模型	A/%	P/%	R/%	参数量	检测时间/ms
BDNN	88.0	94.0	79.5	5.000×10^5	8.5
ACAM	89.9	88.4	90.6	2.300×10^5	14.1
Self-attentive VAD	93.5	92.9	93.5	6.000×10^5	10.2
VAN	93.7	93.9	92.6	1.077×10^7	21.0
EGVAN_VAD	94.1	93.4	94.3	3.400×10^5	10.0

由表 2 可知，EGVAN_VAD 对肉鸽鸣叫声端点检测的 A、R 最高，分别为 94.1%、94.3%，而 P 分别比 BDNN、VAN 低 0.6、0.5 个百分点，但 A 和 R 分别比 BDNN、VAN 高 6.1、0.4 个百分点和 14.8、1.7 个百分点。综合 A、P、R 可知，EGVAN_VAD 检测效果最好，其次是 VAN，BDNN 最差，而 Self-attentive VAD 略低于 EGVAN_VAD。本文模型参数量为 3.400×10^5，仅为 VAN 参数量的 3.2%，且小于 BDNN、Self-attentive VAD，ACAM 在 5 种模型中参数量最小，为 EGVAN_VAD 参数量的 67.7%，但 EGVAN_VAD 的 A、P、R 较 ACAM 分别提高 4.2、5.0、3.7 个百分点。本文模型的检测速度为 10 ms，与 Self-attentive VAD 检测速度差距较小，较 ACAM、VAN 模型检测时间降低 4.1、11 ms，BDNN 检测速度在 5 种模型中最优。

综合上述分析，本文模型对肉鸽鸣叫声检测性能优于其他模型，具有较好的检测精度和速度，且参数量较低。

为进一步说明本文模型(EGVAN_VAD)对肉鸽鸣叫声端点的检测效果，对比其他 4 种模型对肉鸽“鸣”和“咕”鸣叫声端点的检测结果，如图 12 所示，图中绿色垂直虚线是叫声端点检测起点，红色垂直实线是端点检测终点，每一组绿色和红色垂线区域为一段音频端点检测区域，且前后对应。结果显示，5 种不同的端点检测模型对肉鸽“鸣”鸣叫声区域具有较好检测结果(图 12a)。从图 12b 来看，BDNN 检测结果存在 1 次鸣叫声检测为 2 次，ACAM、VAN 也出现同样的检测结果，并且 BDNN 存在部分鸣叫声区域漏检，而 Self-attentive VAD 将 1 次“咕”鸣叫声检测为多次，出现此情况主要由于“咕”鸣叫声呈现出多个尖峰波形，导致模型将尖峰波形之间误识别为非目标波段。由于肉鸽鸣叫声持续时长均大于 300 ms，而非目标声音误检测为鸣叫声的情况音频时长一般较短，为过滤误检测音频段，设置端点检测时长阈值为 300 ms。对于 BDNN、Self-attentive VAD “咕”鸣叫声检测结果，经过过滤后会导致肉鸽初始鸣叫声端点检测区域不完整，会对鸣叫声雌雄鉴别结果产生影响。图 12a 中 5 s 后和图 12b 中 6 s 后为肉鸽活动而引起的噪声音频，ACAM、Self-attentive VAD 出现将噪声音频误检测为肉鸽鸣叫声。从 EGVAN_VAD 检测结果可见，模型“鸣”和“咕”鸣叫声均有较好的端点检测结果，并且能有效过滤非鸣叫声音频段。结果表明，基于 EGLKA 模块所设计的端点检测模型比 VAN 模型具有更强的特征提取能力，并且得益于 Stage_3 和 Stage_4 部分裁剪，避免过度卷积和维度下降而降低模型特征提取能力。

因此，使用 EGVAN_VAD 作为肉鸽鸣叫声端点检测模型可以更好地为肉鸽鸣叫声雌雄鉴别模型提供准确的肉鸽鸣叫声音频。

图 12 音频经过噪声门限和巴特沃斯低通滤波器降噪结果如图 13 所示。由图 13 可知，滤波后，肉鸽鸣叫声影响不大，而一些振幅较低的环境噪声在滤波过程中被滤除。对于一些由肉鸽日常活动而出现的高振幅高能量噪声音频，滤波后也减少了噪声音频的杂波。从而提高肉鸽鸣叫声质量和清晰度，降低噪声对肉鸽鸣叫声雌雄鉴别模型的影响。

2.4　肉鸽声音识别模型训练结果

模型训练参数学习率设置为 0.001，批次大小设置为 32，训练类别数量为 3。训练过程采用早停策略(Early stopping)。该策略通过监测验证集准确率，当连续 100 个训练轮次未出现提升，即验证集准确率未增加时终止训练，并在这 100 个轮次中选择具有最高验证集准确率的轮次作为训练的最终结果。

为评价所构建的 EGVAN 模型以及 RNN 与 DNN 方法模型对肉鸽鸣叫声的识别性能，进行了不同肉鸽鸣叫声识别模型对比试验，比较 EGVAN、

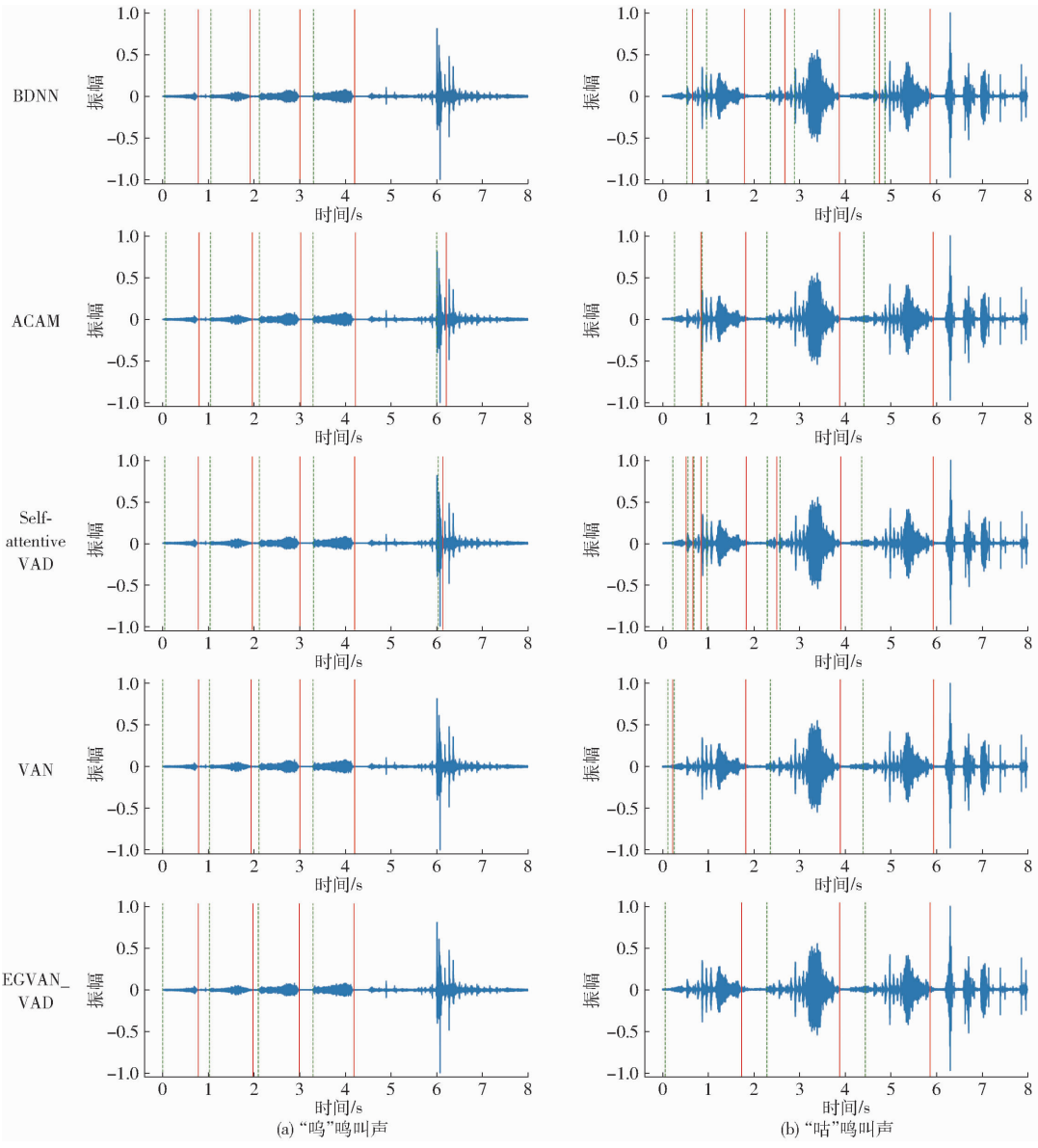


图 12 不同端点检测模型检测结果

Fig. 12 Detection results of different endpoint detection models

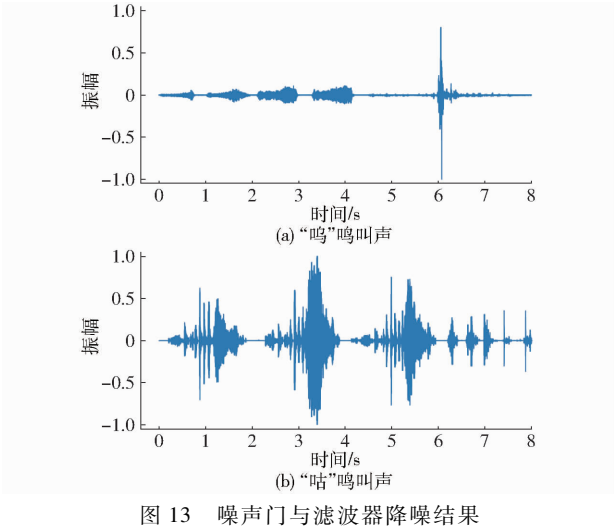


图 13 噪声门与滤波器降噪结果

Fig. 13 Noise gate and filter noise reduction results

LSTM、门控循环单元(GRU)、TDNN 和 VAN 网络模型对肉鸽鸣叫声识别性能。其中 LSTM、GRU 属于 RNN 类型,TDNN 属于 DNN 类型,VAN 和 EGVAN 属于 CNN 类型。5 种不同模型在验证集上的识别结果见表 3。由表 3 可知,EGVAN 模型识别效果最好,平均识别准确率为 99.0%,其中雄性和“其他”类别识别准确率为 5 个模型中最高,分别为 98.6%、100%,雌性类别识别准确率略低于 TDNN 与 VAN。与 VAN 相比,EGVAN 参数量降低 44.3%,为 6.00×10^6,检测时间降低 31.6%,为 14.1 ms,具有较好的检测速度,仅次于 TDNN。而 LSTM 识别准确率和检测时间最差,其次为 GRU,但 GRU 参数量最小。

表 4 为不同模型在测试集上的识别结果。针对测试集中所有声音数据,EGVAN 模型表现最佳,平均识别准确率达到91.3%。在雌性和雄性肉鸽的

表 3　不同模型在验证集上的识别结果

Tab.3　Identification results of different models on verification sets

模型	P_v/%			参数量	检测时间/ms
	雌性	雄性	其他		
LSTM	94.7	95.1	92.8	1.38×10^6	25.5
GRU	98.4	96.8	97.5	1.09×10^6	21.3
TDNN	99.1	96.2	98.2	2.60×10^6	6.2
VAN	98.9	98.6	98.8	1.08×10^7	20.6
EGVAN	98.4	98.6	100	6.00×10^6	14.1

表 4　不同模型在测试集上的识别准确率结果对比

Tab.4　Comparison of recognition results of different models on test set　%

模型	1～3 月		3～6 月		大于 6 月		其他	平均值
	雌性	雄性	雌性	雄性	雌性	雄性		
LSTM	86.4	78.4	93.5	95.8	91.6	74.4	90.2	87.2
GRU	93.3	78.7	97.8	97.1	93.9	72.2	95.5	89.8
TDNN	87.1	83.4	98.4	96.6	93.3	74.3	95.3	89.9
VAN	88.5	88.6	95.1	98.7	74.3	78.8	92.9	88.1
EGVAN	88.7	86.3	97.8	98.7	85.5	85.8	96.6	91.3

鸣叫声识别方面，VAN 模型平均识别准确率分别为 86.0%、88.7%，EGVAN 模型平均识别准确率分别为 90.7%、90.3%，较 VAN 模型分别提升 4.7、1.6 个百分点。在不同鸽龄的肉鸽鸣叫声中，各模型对于 3～6 月鸽龄的音频数据识别效果最佳。其中，EGVAN 模型在这个鸽龄段的雌性和雄性音频的识别准确率分别为 97.8%、98.7%，平均识别准确率为 5 种模型中最高，达到 98.3%。对大于 6 月鸽龄的雌性类别音频数据，LSTM、GRU 和 TDNN 模型表现较好，但对雄性类别的识别效果较差，两个类别的准确率相差较大，表明模型的识别结果不稳定。VAN 与 EGVAN 模型在 1～3 月和大于 6 月鸽龄的音频数据识别准确率较为稳定，但 VAN 模型对大于 6 月鸽龄的肉鸽音频数据识别准确率较低。EGVAN 模型对“其他”类别识别准确率最高，达到 96.6%。

在 5 个识别模型中，EGVAN 模型的整体性能最好，其次为 TDNN 模型。EGVAN 模型在处理梅尔语谱图图像分类任务中充分发挥了其对特征局部结构和长距离依赖性提取的优越性能。LSTM 和 GRU 模型参数量较优，但二者准确率和检测速度不如 EGVAN，这说明 EGVAN 模型在实现准确率和模型轻量化方面取得了良好平衡。综合而言，EGVAN 模型表现出较好的性能，有效减少参数量，取得比 VAN 更高的准确率，同时显著提高了检测速度。

2.5　基于音频的肉鸽雌雄鉴别结果

为验证系统模型的可行性，使用 EGVAN_VAD 鸣叫声端点检测模型采集测试集中 1～3 月、3～6 月和大于 6 月鸽龄肉鸽 100 段鸣叫声，并将检测结果传输到 EGVAN 中截取 3 s 的音频段进行雌雄鉴别。根据 EGVAN 的输入要求，EGVAN_VAD 的输入音频时长需大于 3 s，并且为了在不明显增加系统端点检测推理时间的情况下，可以一定程度地调节截取 3 s 音频的区域，因此本研究将 EGVAN_VAD 输入音频时长设定为 4 s，当检测到肉鸽鸣叫声时，音频截取以端点检测结果为起始点，若起始时间大于 1 s，则截取 1～4 s 的音频。此外，端点检测到的肉鸽鸣叫声持续时长需大于 300 ms，否则不作为端点检测结果。将肉鸽鸣叫声雌雄鉴别系统搭载在 CPU 和 GPU 分别为 AMD Ryzen 7 3700X 8－Core Processor 3.6 GHz 和 NVIDIA GeForce GTX 1650 的 Windows 系统下运行，最终系统基于声音特征的肉鸽雌雄鉴别结果如表 5 所示。

表 5　基于声音特征的肉鸽雌雄鉴别结果

Tab.5　Male-female identification results of pigeons based on sound characteristics

鸽龄/月	性别	端点检测（音频段）		声音识别（音频段）				雌雄鉴别成功率/%
				成功鉴别		错误鉴别		
		目标	非目标	目标	非目标	目标	非目标	
1～3	雌	98	2	88	2	10	0	89.8
	雄	98	2	80	2	17	1	81.6
3～6	雌	94	6	93	6	1	0	98.9
	雄	99	1	99	1	0	0	100
>6	雌	96	4	79	4	16	1	82.3
	雄	97	3	79	3	17	1	81.4
总计		582	18	518	18	61	3	89.0

由表 5 可知，肉鸽鸣叫声端点检测准确率较好，对于各个鸽龄段检测到的 100 段鸣叫声，总共将 18 段非目标音频检测为鸣叫声，正确检测鸣叫声 582 段，准确率为 97.0%。而对于端点检测错误的非目标音频，EGVAN 均正确识别为非目标声音。对于端点检测正确的目标音频，EGVAN 对 3～6 月鸽龄段的肉鸽雌雄鉴别结果最优，成功率可达 99.5%，对大于 6 月鸽龄段的雌雄鉴别结果最差，成功率仅为 81.9%，存在较多的错误雌雄鉴别结果，而 1～3 月鸽龄段雌雄鉴别结果成功率为 85.7%。此外，存在将肉鸽鸣叫声识别为非目标音频。最终系统各个鸽龄段肉鸽正确鉴别鸣叫声雌雄音频段为 518，错误鉴别音频段为 61，雌雄鉴别平均成功率为 89.0%。在鉴别速度方面，系统检测 4 s 一段的音频总耗时 0.223 s，远小于 4 s，可实现实时监测。

上述结果表明，不同鸽龄的肉鸽鸣叫声呈现出

不同的雌雄区分度。其中 3 ~ 6 月鸽龄的肉鸽鸣叫声识别准确率最高,表明在这一鸽龄段,雌性和雄性肉鸽鸣叫声之间的差异性最为显著,其区分度较高。其次是 1 ~ 3 月鸽龄的鸣叫声,大于 6 月鸽龄的鸣叫声识别效果最差,表明在这两个鸽龄段,雌性和雄性肉鸽鸣叫声间差异性不明显,甚至可能存在交叉,导致识别结果较差。种鸽需要保证体重达标才能发挥良好繁殖性能,25 周鸽龄基本已达到体成熟和性成熟,处于上笼配对期[21]。为了防止种鸽早配和做好配对准备,在种鸽发情前应进行雌雄鉴别[33]。因此,肉鸽雌雄鉴别最佳鸽龄为性成熟前,而不同企业种鸽配对多集中在 4 ~ 7 月鸽龄。而系统对 3 ~ 6 月鸽龄肉鸽雌雄鉴别成功率最高,可解决实际生产中青年鸽雌雄难以准确识别的问题,有助于降低肉鸽饲养过程中对雌雄鉴别分类的劳动强度,有望减少青年鸽雌雄鉴别配对关键作业对饲养员工作经验的依赖,有利于肉鸽产业的长远健康发展。此外,本研究也为其他单态鸟类的性别智能精准检测识别提供重要参考。

3 结论

(1)基于 EGLKA 模型构建了肉鸽鸣叫声端点检测模型 EGVAN_VAD,试验使用 A、P、R、参数量和每段音频检测时间作为评价指标,评价模型对鸣叫声的检测效果。结果表明,EGVAN_VAD 模型评价指标分别为 94.1%、93.4%、94.3%、3.400×10^5 和 10 ms,该模型的检测效果综合性能最好,能够满足在音频信号中对肉鸽鸣叫声的检测识别。

(2)对 1 ~ 3 月、3 ~ 6 月和大于 6 月鸽龄雌性和雄性的肉鸽鸣叫音频数据使用 5 种模型进行了对比试验。结果表明,使用梅尔语谱特征训练的 EGVAN 肉鸽鸣叫声雌雄鉴别模型平均识别准确率最高,达到 91.3%。在雌、雄鸣叫声方面,平均识别准确率分别为 90.7%、90.3%。而且在 3 个不同鸽龄段的测试集中,3 ~ 6 月鸽龄雌性和雄性的音频数据平均识别准确率最高(98.3%),表明此鸽龄段雌雄肉鸽的鸣叫声差异明显。

(3)经系统整体测试,对 3 ~ 6 月鸽龄肉鸽雌雄鉴别成功率为 99.5%,对 3 个鸽龄段平均鉴别成功率为 89.0%,处理 4 s 音频耗时 0.223 s。在 3 个不同鸽龄段中,3 ~ 6 月鸽龄段正是肉鸽饲养过程中需要进行雌雄识别配对的关键时期。说明使用深度学习算法的肉鸽鸣叫音频识别方法可解决实际生产中肉鸽雌雄难以准确鉴别的问题。

参 考 文 献

[1] 徐善金,虞德兵,何宗亮,等. 鸽的雌雄鉴别技术[J]. 江西农业学报,2009,21(2):98 - 100.
XU Shanjin, YU Debing, HE Zongliang, et al. Techniques of sexual discrimination in pigeons[J]. Acta Agriculturae Jiangxi, 2009,21(2):98 - 100. (in Chinese)

[2] VOLODIN I A, VOLODINA E V, KLENOVA A V, et al. Gender identification using acoustic analysis in birds without external sexual dimorphism[J]. Avian Research, 2015, 6: 1 - 17.

[3] 刘国乾, 龙航宇, 朱健润, 等. 鸽性别早期经济快速鉴定方法的研究[J]. 中国畜牧杂志, 2021,57(8):114 - 118.
LIU Guoqian, LONG Hangyu, ZHU Jianrun, et al. Study on early economic rapid identification method of pigeon sex[J]. Chinese Journal of Animal Science, 2021,57(8):114 - 118. (in Chinese)

[4] 刘成康,刘勇,王巧华,等. 种蛋受精与性别信息无损检测技术与装备研究进展[J]. 农业机械学报,2024,55(5):1 - 20.
LIU Chengkang, LIU Yong, WANG Qiaohua, et al. Progress and prospect in non-destructive assessment technologies and equipment of fertilization and gender determination in hatching eggs[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(5):1 - 20. (in Chinese)

[5] CUAN K, LI Z, ZHANG T, et al. Gender determination of domestic chicks based on vocalization signals[J]. Computers and Electronics in Agriculture, 2022, 199: 107172.

[6] WANG Y, WANG S, GAO Y, et al. An approach to rapidly identify the gender of the pigeon by using cross-priming amplification with immune-chromatographic strip[J]. Talanta, 2024, 269: 125452.

[7] LI N, REN Z, LI D, et al. Automated techniques for monitoring the behaviour and welfare of broilers and laying hens: towards the goal of precision livestock farming[J]. Animal, 2020, 14(3): 617 - 625.

[8] CUAN K, ZHANG T, HUANG J, et al. Detection of avian influenza-infected chickens based on a chicken sound convolutional neural network[J]. Computers and Electronics in Agriculture, 2020, 178: 105688.

[9] 宗超,曹晏飞,曹孟冰,等. 笼养和栖架养殖模式下蛋鸡的发声特征[J]. 农业工程学报,2021,37(6):135 - 141.
ZONG Chao, CAO Yanfei, CAO Mengbing, et al. Vocalization characteristics of laying hens under cage and perch systems[J]. Transactions of the CSAE,2021,37(6):135 - 141. (in Chinese)

[10] 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.
DU Xiaodong, TENG Guanghui, LIU Mulin, et al. Recognition method of breeding birds' vocalization based on lightweight convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(10):271 - 276. (in Chinese)

[11] XIE J, ZHANG L, ZHANG J, et al. Cross-corpus open set bird species recognition by vocalization[J]. Ecological Indicators, 2023, 154: 110826.
[12] MORGAN M M, BRAASCH J. Open set classification strategies for long-term environmental field recordings for bird species recognition[J]. The Journal of the Acoustical Society of America, 2022, 151(6): 4028 – 4038.
[13] 祝鹏飞,滕光辉,刘健,等.基于深度学习模型 PANNS – CNN 的肉种鸡惊吓声识别研究[J].中国家禽,2023,45(11):105 – 111.
ZHU Pengfei, TENG Guanghui, LIU Jian, et al. Recognition of screams of broiler breeders based on deep leaing model PANNS – CNN[J]. China Poultry,2023,45(11):105 – 111. (in Chinese)
[14] GONCHAROVA M V, KLENOVA A V, BRAGINA E V. Development of cues to individuality and sex in calls of three crane species: when is it good to be recognizable? [J]. Journal of Ethology, 2015, 33: 165 – 175.
[15] BAPTISTA L F, GAUNT S L L. Bioacoustics as a tool in conservation studies[J]. Behavioral Approaches to Conservation in the Wild, 1997: 212 – 242.
[16] PEREIRA E M, NAEAES I D A, GARCIA R G. Vocalization of broilers can be used to identify their sex and genetic strain [J]. Engenharia Agrícola, 2015, 35: 192 – 196.
[17] SADEGHI M, BANAKAR A. Gender determination of fowls by using bioacoustical data mining methods and support vector machine[J]. J. Agr. Sci. Tech, 2017, 19: 1041 – 1055.
[18] LI Z, ZHANG T, CUAN K, et al. Sex detection of chicks based on audio technology and deep learning methods[J]. Animals, 2022, 12(22): 3106.
[19] YIN J J, LI W G, LIU Y F, et al. Sex identification of ducklings based on acoustic signals[J]. Poultry Science, 2024, 103(6):103711.
[20] GUO M H, LU C Z, LIU Z N, et al. Visual attention network[J]. Computational Visual Media, 2023, 9(4): 733 – 752.
[21] 汤青萍,卜柱,穆春宇,等.肉鸽实用选育技术[J].中国家禽,2018,40(4):69 – 72.
[22] 耿磊,吴寒冰,张芳,等.基于 TCN 和 Transformer 的鸡胚心跳混淆信号分类方法[J].农业机械学报,2023,54(8):296 – 308.
GENG Lei, WU Hanbing, ZHANG Fang, et al. Classification method of heartbeat confusion signals of hatching eggs based on TCN and Transformer[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(8): 296 – 308. (in Chinese)
[23] 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.
SHEN Mingxia, WANG Mengyu, LIU Longshen, et al. Recognition method of pig cough based on deep neural network[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(5):257 – 266. (in Chinese)
[24] 谢卓钒,李鼎昭,孙海信,等.面向鸟鸣声识别任务的深度学习技术[J].生物多样性,2023,31(1):137 – 144.
XIE Zhuofan, LI Dingzhao, SUN Haixin, et al. Deep learning techniques for bird chirp recognition task[J]. Biodiversity Science,2023,31(1):137 – 144. (in Chinese)
[25] LU S, LIU M, YIN L, et al. The multi-modal fusion in visual question answering: a review of attention mechanisms[J]. PeerJ Computer Science, 2023, 9: e1400.
[26] LEE Y, PARK J. Centermask: real-time anchor-free instance segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 13906 – 13915.
[27] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions[J]. arXiv Preprint, arXiv:2112.05561, 2021.
[28] 单蒙,米吉提·阿不里米提,艾斯卡尔·艾木都拉.基于小样本学习的语音端点检测[J].现代电子技术,2022,45(24):145 – 150.
SHAN Meng, MIJIT Ablimit, ASKAR Hamdulla. Voice activity detection based on few-shot learning[J]. Modern Electronics Technique,2021,45(24):145 – 150. (in Chinese)
[29] 顾小平,吴浩,陈佳豪,等.基于 DE – VRF 的猪声音分类识别[J].计算机应用与软件,2023,40(3):241 – 247.
GU Xiaoping, WU Hao, CHEN Jiahao, et al. Pig sound classification and recognition based on DE – VRF[J]. Computer Applications and Software,2023,40(3):241 – 247. (in Chinese)
[30] LAVECHIN M, GILL M P, BOUSBIB R, et al. End-to-end domain-adversarial voice activity detection[J]. arXiv Preprint, arXiv:1910.10655, 2019.
[31] MIHALACHE S, IVANOV I A, BURILEANU D. Deep neural networks for voice activity detection[C]// 2021 44th International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2021: 191 – 194.
[32] JO Y R, MOON Y K, CHO W I, et al. Self-attentive VAD: context-aware detection of voice from noise[C]// ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021: 6808 – 6812.
[33] 陈益填.种鸽的配对方法和配对前后的管理技术[J].中国家禽,2005,27(17):28 – 30.