



祝万军,王蕾,李鹏,袁超,陈金鑫,唐瑜嵘,沈明霞. 基于改进 TransformerCNN 的轻量化笼养鸡发声识别模型[J]. 中国农业大学学报,2025,30(08):132-140.
ZHU Wanjun, WANG Lei, LI Peng, YUAN Chao, CHEN Jinxin, TANG Yurong, SHEN Mingxia. A lightweight cage chicken vocalization recognition model based on improved TransformerCNN[J]. *Journal of China Agricultural University*, 2025, 30(08): 132-140.
DOI: 10.11841/j.issn.1007-4333.2025.08.12

基于改进 TransformerCNN 的轻量化笼养鸡发声识别模型

祝万军¹ 王蕾² 李鹏¹ 袁超¹ 陈金鑫¹ 唐瑜嵘¹ 沈明霞^{1*}

(1. 南京农业大学 人工智能学院/农业农村部养殖装备重点实验室, 南京 210031;

2. 南通市海门区农业技术推广中心, 江苏 南通 226199)

摘要 为解决传统笼养鸡发声检测难度高、实时性差等问题,本研究提出一种基于轻量化 TransformerCNN 的鸡只发声检测模型。首先,通过引入 EfficientViT 轻量网络,替换传统骨干网络,减少参数数量和计算复杂度,从而提升检测模型的实时性;其次,将 ReLU 替换为 Hardswish 激活函数,在不增加额外的权重和偏置参数的前提下,提升模型的识别准确率;最后,采用线性自注意力机制替代传统的 Softmax 操作,进一步增强了模型对鸡只发声特征的捕捉能力,提高检测的鲁棒性。结果表明:1)相较于基础的 TransformerCNN 模型,引入 EfficientViT、线性自注意力和 Hardswish 激活函数后,模型精确度略有下降,但在推理速度、计算复杂度和实时性方面表现更佳。2)与其他轻量化骨干网络相比,EfficientViT 在推理速度、计算量和模型存储需求上具有明显优势,适用于实时分类任务。3)与主流分类模型对比,TC+EfficientViT 在检测精度、实时性和模型体积方面均有优势,满足高效、实时的部署需求。综上,在智能福利化养殖环境中,通过无接触式检测鸡只发声,能够建立笼养鸡群体发声数据库,为进一步分析鸡只健康状况提供关键数据。

关键词 笼养鸡;发声识别;轻量化模型;实时检测;EfficientViT

中图分类号 S381.4;TP391.4

文章编号 1007-4333(2025)08-0132-09

文献标志码 A

A lightweight cage chicken vocalization recognition model based on improved TransformerCNN

ZHU Wanjun¹, WANG Lei², LI Peng¹, YUAN Chao¹, CHEN Jinxin¹,
TANG Yurong¹, SHEN Mingxia^{1*}

(1. College of Artificial Intelligence/Key Laboratory of Breeding Equipment of Ministry of Agriculture and Rural Affairs, Nanjing Agricultural University, Nanjing 210031, China;

2. Agricultural Technology Promotion Center in Haimen District, Nantong 226199, China)

Abstract To solve the problems of high difficulty and poor real-time detection of traditional cage chicken vocalization, this paper proposes a chicken vocalization detection algorithm based on lightweight TransformerCNN. Firstly, by introducing the EfficientViT lightweight network to replace the original backbone network, the number of parameters and computational complexity are reduced, and the real-time performance of network algorithm detection is improved; Secondly, replacing ReLU with Hardswish activation function can improve the recognition accuracy of the model without adding additional weights and bias parameters. Finally, by combining linear self-attention mechanism instead of traditional Softmax operation, the model's ability to capture chicken vocalization features is further enhanced, improving the

收稿日期: 2025-02-20

基金项目: 江苏省科技项目(BE2022379)

第一作者: 祝万军(ORCID: 0009-0000-3942-8184), 硕士研究生, E-mail: zhuwanjun@stu.njau.edu.cn

通讯作者: 沈明霞(ORCID: 0000-0001-8649-6083), 教授, 主要从事机器视觉和信息农业研究, E-mail: mingxia@njau.edu.cn

robustness of detection. The results show that: 1) Compared with the basic TransformerCNN model, the introduction of EfficientViT, linear self-attention, and Hardswish activation function significantly improves the inference speed and computational complexity of the model. Although the average accuracy slightly decreases, it performs better in real-time performance and lightweight nature. 2) Compared with other lightweight backbone networks, EfficientViT has significant advantages in inference speed, computational complexity, and model storage requirements, making it suitable for real-time audio classification tasks. 3) Compared with mainstream audio classification models, TC+EfficientViT has improved detection accuracy, real-time performance, and model size, which can better meet the requirements of efficient and real-time deployment. In conclusion, in the intelligent welfare breeding environment, through contactless detection of chicken vocalization, a database of vocalizations in cage chicken populations can be established, providing key data for further analysis of chicken health status.

Keywords cage chicken farming; voice recognition; lightweight model; real-time detection; efficientViT

笼养鸡的发声信息可以评估其健康状况和生产性能^[1]。不同健康状态下的鸡只声音具有不同的特征变化,例如在呼吸道疾病、热应激等状态下,鸡只发声的频率、时长、响度等特征随之改变。目前,在规模化养殖场中,检测鸡只呼吸道疾病、热应激等仍然需要人工观测。随着智能福利化养殖的兴起^[2],无接触式、智能检测鸡只发声变得至关重要,这有助于及时发现鸡只是否出现不良反应,进而预防鸡只疾病的发生^[3]。

国内外专家学者已相继展开关于鸡只发声检测方面的研究。目前在识别鸡只发声方面的研究主要分两类:一是基于传统机器学习的鸡只异常声音识别算法研究。Carroll等^[4]对梅尔频率倒谱系数(Mel-frequency cepstral coefficients, MFCC)向量进行聚类并检查聚类索引在时间窗口内的分布来从音频中导出特征,使用C4.5决策树对特征进行分类,啰音的识别准确率为73.4%,召回率为51.4%。Rizwan等^[5]利用13个频带的梅尔尺度滤波器组进行特征提取,利用改进的极限学习机(ELM)和支持向量机(SVM)对这些特征进行分类,啰音识别准确率分别为97.1%和97.6%,召回率分别为81.5%和86.6%。两者的召回率差别巨大的重要原因是选择的训练文件是相对响亮的啰音,而真实场景下有许多较安静的啰音混杂在其中未被检测到,这导致了召回率较低^[6]。二是基于深度学习技术的鸡只异常声音识别。袁超等^[7]提出利用提取滤波器组(Filter bank, FBank)和梅尔频率倒谱系数(Mel-frequency cepstral coefficients, MFCC)特征组合,利用VGG16和ResNet18神经网络对10~19、20~29、30日龄的3种日龄的特征进行分类,咳嗽声识别的准确率分别为94.29%、

97.65%和98.66%。Cuan等^[8]提出将鸡叫声的Logfbank、梅尔频率倒谱系数(MFCC)、MFCC Delta和MFCC Delta-Delta 4个特征组合成总特征图,利用CSCNN卷积神经网络识别健康鸡和病鸡的声音,识别准确率在染病第4天为93.56%。Carion等^[9]首次将Transformer机制(一种在自然语言处理领域广泛应用的架构)引入检测任务中,相较于基于卷积神经网络的方法,其网络结构更为简洁。

上述研究大多针对平养鸡,但在笼养环境下,由于鸡舍内空间有限,由于鸡舍空间有限,传统模型部署通常依赖主机等体积较大的设备,当模型参数量较大时,不仅部署困难,还会导致检测速度较慢,难以实现实时检测。为解决以上问题,本研究在TransformerCNN双通道模型的基础上,结合EfficientViT骨干网络、线性自注意力机制和Hardswish激活函数,提出了一种轻量化的发声检测模型。为笼养鸡发声的实时识别提供有效的算法支持。

1 材料与方法

1.1 试验对象和音频数据采集

数据于2023年9月8日—10月8日在河北省石家庄市峪口鸡场完成采集。试验对象为笼养京粉6号蛋鸡,鸡笼尺寸800 cm×420 cm×2 100 cm,养殖密度为15~20只/笼,一个鸡舍约2万只鸡。

音频采集设备为海康威视DS-2FP1021型网络拾音器(如图1所示),拾音范围约5 m,音频存储格式为wav,采集精度16位,采样频率48 kHz,每1 h保存为一个录音文件,命名格式为“年-月-日-时-分-秒”,一个拾音器能采集的范围约100只鸡。

试验对象为养殖舍中5周龄的京粉6号蛋鸡。为降低噪声和公鸡嘈杂的打鸣声影响,设置采集时间为每日04:00—05:00,该时段舍内熄灯,鸡只杂乱发声大幅降低。记录多日音频数据,共采集录音数据30 h。

1.2 数据预处理与数据划分

本试验中采集到的音频包含鸡只的咳嗽声、鸣叫声等可用数据以及风机、投饲机等机器运作时的环境噪声,各声音类别特征如表1所示。为了提高识别模型的性能,需对原始音频数据进行预处理以



图1 海康威视网络拾音器示意图
Fig. 1 Schematic diagram of Hikvision network microphone

提高数据可用性^[10-12]。

表1 各声音类别特征分析
Table1 The characteristics of different sound categories

声音类型 Sound type	特征描述 Feature description	时域特征 Time domain characteristics	能量分布 Energy distribution
咳嗽声 Coughing	短促急促,30~500 ms	波峰明显,起伏急促	瞬时能量高,集中
鸣叫声 Chirping	节奏平稳,约2 s	波形平滑,起伏有节奏	能量均匀
其他声音 Other voices	持续缓慢,随机性强	波形平缓,缺乏峰值	能量分散

1.2.1 声音降噪

频谱门控法是音频信号处理过程中常用的降噪方法之一,用于去除或抑制音频信号中的持续性背景噪声^[13]。在工厂环境下录制的声音包含鸡只的叫声和环境噪声,鸡只声音相对环境噪声的频率高,所以在降噪时针对频率范围不同可以有效降低环境噪声,并且由于养鸡场的现场音频噪音大且复

杂,设置固定的阈值进行去噪易去除目标声音^[14-16]。频谱门控是一种针对声音频率进行降噪的方法,并且在设置噪声阈值时,是随着时间的推移,不断更新估计噪声阈值,对于鸡场环境下长时间录制音频的情况,有很好的效果。如图2所示,选择15 min音频做对比试验,由原始音频和降噪后音频的声谱图对比可知,声音片段的噪点明显减少。

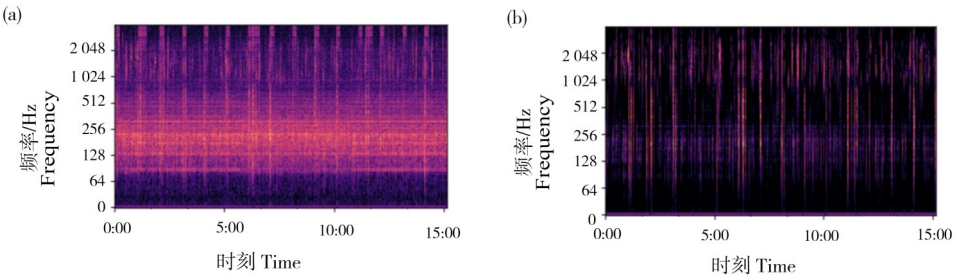


图2 原始音频(a)与降噪后音频(b)的对比图
Fig. 2 Comparison between the original audio (a) and the denoised audio (b)

1.2.2 鸡只发声数据集划分

对录制的声音进行人工分类,共得到咳嗽声1 121个,鸣叫声1 205个和其他声音351个。其中,

其他声音包括风机、投饲机等环境制造的声音,3种声音不存在混叠。数据集以7:2:1样本占比设置训练集、验证集和测试集(表2)。

表 2 数据集样本数划分结果

Table 2 Data set partitioning results

声音类别 Sound class	训练集 Training set	验证集 Verification set	测试集 Test set	总数 Total
咳嗽声 Coughing	1 050	300	150	1 500
鸣叫声 Chirping	2 100	600	300	3 000
其他声音 Other voices	704	200	102	1 006

1.3 声音特征参数提取

在声音降噪后,需要做连续的加窗、分帧处理,再提取每一帧的声音特征参数作为模型输入。在特征参数的选取上,由于呼吸道疾病、热应激等异常情况会对家禽声道造成破坏性影响,从而影响它们的呼吸行为和发声特征,并且会增加某些频率(谐波)的强度,抑制了其他频率。基于此,选择梅尔子带的能量分布(Energy of mel sub bands, MFENs)、梅尔频率倒谱系数(Mel-frequency cepstral coefficients, MFCCs)等可以反映鸡只声音谐波的特征。其中,梅尔子带能量分布与 MFCCs 的区别如图 3 所示,在 MFCCs 的计算中,取对数的目的是考虑非线性尺度上的能量分布,离散傅里叶变换的目的是对能量带之间进行去相关操作,但是在笼养鸡场环境下,即便经过降噪处理,音频中仍然含有部分噪声,因此,上述两步计算会降低系统性能。综上,梅尔子带能量分布从理论上比 MFCCs 更能表现鸡只不同发声^[17-18]。

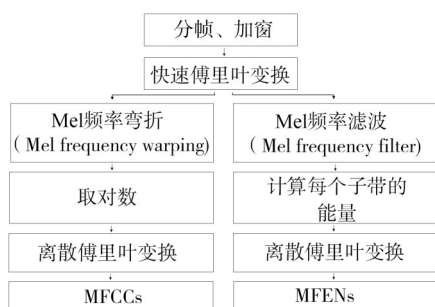


图 3 MFCC 和 Mel 子带的能量分布的计算步骤

Fig. 3 Calculation steps for energy distribution of MFCC and Mel sub bands

1.4 轻量化模型处理

使用 1.4 中提取的特征参数作为输入,对模型进行轻量化处理,通过比较模型训练后的准确度、

召回率等指标,综合判断模型性能。

1.4.1 EfficientViT 主干网络

针对鸡只声音识别在实际场景中的应用需求,在现有研究基础上,提出了一种轻量化的实时声音识别模型^[19]。基于现有 Transformer 框架,通过将其主干网络更换为 EfficientViT 网络,优化了计算效率。EfficientViT 采用了一系列优化策略,如深度可分离卷积、逆残差结构等,减少了网络的参数量和计算复杂度。这一替换不仅降低了模型的复杂度,且提高了模型检测的实时性,使其更适用于实时检测场景。

EfficientViT^[20]结构由 1 个初始模块(Stem)和 4 个阶段(Stages)组成,每个阶段在逐步降低特征图尺寸的同时增加了通道数,如图 4(a)所示。与传统卷积神经网络不同,EfficientViT 在其第 3 和第 4 阶段引入了 Efficient block 模块,如图 4(b)所示。Efficient block 模块的特色在于其包含 1 个轻量级的多尺度线性注意模块(Multi-scale linear attention, MSA),以及 1 个具有深度卷积(Depth wise convolution, DWConv)的前馈神经网络(Feedforward neural network, FNN)^[21]。

1.4.2 Hardswish 激活函数

在原 EfficientViT 模型中,使用的是 ReLU 激活函数。ReLU 的函数表达式如公式(1)所示,通过对输入值进行判断,若 >0 则输出该值,否则输出 0。相较于其他激活函数,ReLU 在计算上具有更高的效率,并能保证网络模型的稀疏性,避免过拟合。然而,ReLU 存在“神经元死亡”的问题,即当输入值始终 <0 时,神经元无法被激活,无法参与训练过程,进而影响模型的学习能力^[22]。

ReLU 的函数表达式:

$$R_e(x) = \max(0, x) = \begin{cases} 0, & x \leq 0 \\ x, & x \geq 0 \end{cases} \quad (1)$$

式中: x 为输入值,指神经网络中某一层的输入信

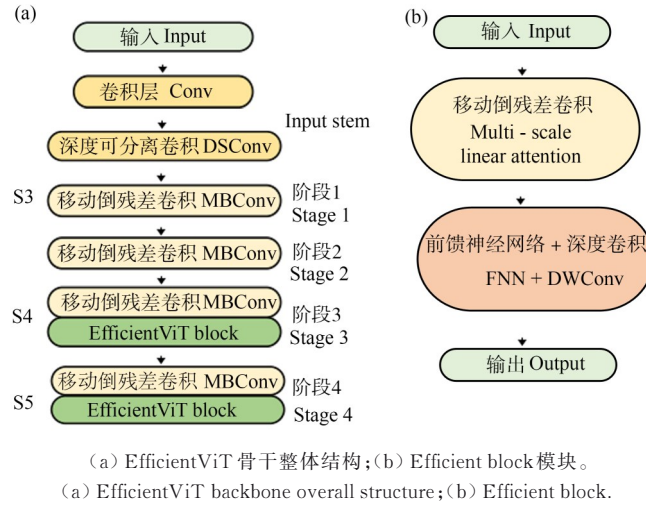


图4 EfficientViT 骨干结构图

Fig. 4 EfficientViT backbone structure diagram

号; R_e 为 ReLU 激活函数的输出值, 指经过 ReLU 函数处理后的结果。

为了解决这一问题, 本研究引入了 Hardswish 激活函数。Hardswish 作为 Swish 激活函数的变体, 通过分段线性函数替代了 Sigmoid 函数, 降低了模型的计算成本, 优化了性能和计算效率, 同时缓解了梯度消失和非单调性问题。其表达式如式(2)所示。通过将 Hardswish 激活函数应用于 EfficientViT 主干网络, 能够进一步增强模型的轻量化特性, 提高计算效率, 同时保留 Swish 激活函数的优势, 优化模型的特征提取能力。

Hardswish 激活函数的表达式:

$$H(x) = x \times \frac{R_e(x+3)}{6} = \begin{cases} x, & x \geq 3 \\ \frac{x(x+3)}{6}, & -3 < x < 3 \\ 0, & x \leq -3 \end{cases} \quad (2)$$

式中: $H(x)$ 为 Hardswish 激活函数的输出值; R_e 为 ReLU 激活函数的输出值; x 为输入值, 指神经网络中某一层的输入信号。

1.4.3 线性自注意力 (Linear attention mechanism)

在鸡只发声检测任务中, 使用线性自注意力机制的主要原因在于其能够减少计算复杂度, 并提高实时推理能力。传统的自注意力机制由于计算全局关系, 其复杂度随着输入序列长度的增加而快速增长, 这在实时应用中会带来较大的计算负担。而

线性自注意力机制通过低秩近似方法降低了计算复杂度, 使得模型在处理长时间音频序列时更加高效, 尤其适用于鸡只发声这种时长较短且多变的音频信号。此外, 线性自注意力机制能够专注于局部特征的提取, 从而在识别短暂且局部的叫声特征时表现出更好的效率与精度。这使得它在要求实时性和低计算资源消耗的鸡只发声检测任务中具有优势^[23-24]。

对于特征表示 $\mathcal{O}(x)$, Transformer 自注意力如公式(3)所示:

$$Q_i = \sum_{j=1}^N \frac{\mathcal{O}(Q_i) \mathcal{O}(K_j)^T}{\sum_{j=1}^N \mathcal{O}(Q_i) \mathcal{O}(K_j)^T} V_j \quad (3)$$

式中: Q, K, V 分别为输入序列的 Query、Key 和 Value 矩阵; i 为第 i 个 Q 或 K 或 V 的输出位置; j 为从序列的第 j 个 Key 或 Value 的输入; N 为输入序列时间长度; T 为对矩阵的转置运算; $\mathcal{O}(\cdot)$ 为特征映射函数。

线性自注意力如下:

$$Q_i = \frac{\mathcal{O}(Q_i) \left(\sum_{j=1}^N \mathcal{O}(K_j)^T V_j \right)}{\mathcal{O}(Q_i) \left(\sum_{j=1}^N \mathcal{O}(K_j)^T \right)} \quad (4)$$

式中各物理量含义同式(3)。

线性自注意力通过2个独立的核函数代替 Softmax 操作, 近似解决了 Transformer 高计算复杂度问题。将自注意力计算看作核特征映射的线性点积, 利用矩阵乘积的结合性, 使计算顺序由 $(QK^T)V$ 变换为 $Q(K^TV)$,

其计算复杂度将由 $O(N^2d)$ 转变为 $O(Nd^2)$ (d 是输入数据的特征向量维度)。

1.4.4 基于轻量化双通道的鸡只咳嗽声识别模型

不同的识别模型,处理的过程和优势是不同的。对于鸡只发声检测任务,CNN模型的作用和优势在于提取从输入的特征参数里的局部特征,而Transformer模块的优势在于能够捕捉长距离时间依赖关系。鸡只发声具有时间尺度,每一小部分的语音片段都有相应的特征参数,应用CNN模块将特征参数提取局部信息,经过Transformer模块处理,可以获得音频在时间尺度上的全局特征,弥补了CNN的不足。具体步骤如下:

在图5的模型架构中,利用CNN对输入的发声

信号进行局部特征抽取,并通过自适应最大池化层将提取出的多维特征映射到固定维度的二维向量。Transformer编码器处理发声序列,其输出经过平均池化以获得反映全局上下文信息的特征向量。将CNN模块产生的线性特征向量与Transformer的平均池化特征向量进行拼接,形成一个融合局部和全局信息的联合特征向量。这个综合特征向量通过全连接层进一步转换和压缩,最后经过线性自注意力(Linear self-attention)层得到各分类的概率分布,从而实现了对鸡只发声的分类识别。双通道识别模型结合了CNN在局部特征提取方面和Transformer在建模时间依赖性方面优势。

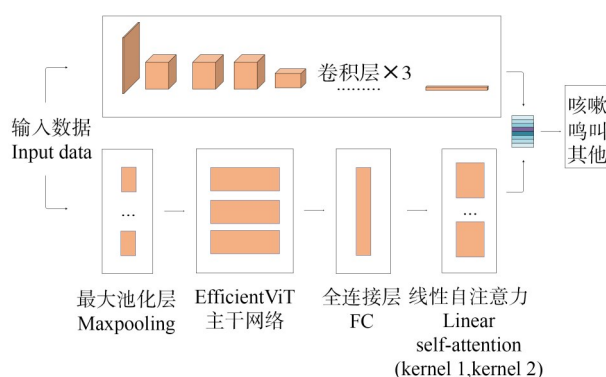


图5 轻量化TransformerCNN模型结构图

Fig. 5 Lightweight 'TransformerCNN' model structure diagram

2 结果与分析

2.1 模型评价指标

所有模型训练均在相同的试验环境下进行。训练平台配置为Intel i9-10900K处理器,搭载NVIDIA GTX 3090显卡(显存24 Gb),操作系统为Ubuntu 18.04,且所有训练均基于PyTorch深度学习框架。在模型训练前,进行参数配置,采用随机梯度下降(SGD)优化器计算模型训练过程中的损失函数梯度,并动态更新网络模型权重。损失函数选择交叉熵(Cross entropy),初始学习率设定为0.01,训练迭代次数为40次,批次大小为32,学习率优化策略为每5次迭代后将学习率衰减为原来的一半。为评估鸡只发声识别模型的性能,本研究使用准确率(Accuracy)、召回率(Recall)和F1值等指标,通过消融试验、对比不同轻量化主干网络试验和对比不同模型试验来说明、比较模型的性能。

2.2 消融试验分析

为验证各改进模块的有效性,本研究基于TransformerCNN基础模型进行了消融试验,如表3所示。结果表明,当引入EfficientViT骨干网络替代原Transformer结构时,模型浮点计算量降低57.1%,推理速度提升至15 AIT/s,但平均准确率下降1.8%,表明轻量化设计需权衡特征表达能力。进一步引入线性自注意力机制后,模型召回率提升1.1%,推理速度提高至30 AIT/s,说明该机制能够有效捕捉时序依赖特征。最终结合Hardswish激活函数,模型综合性能最优,准确率达97.0%,召回率为91.8%,浮点计算量压缩至3.2 G,推理速度提升至44 AIT/s。消融试验证明,模块间的协同优化可在轻量化前提下维持较高的检测精度。

2.3 不同轻量化主干网络对模型性能的影响分析

为筛选最优骨干网络,本研究对比了不同轻量化架构的性能,如表4所示。TC+EfficientViT模型在准确率、推理速度及模型体积上均表现最

表3 应用不同改进模块消融试验分析
Table 3 Analysis of ablation using different improved modules

主干网络 EfficientViT	线性自注意力 Linear attention mechanism	激活函数 Hardswish	准确率/% Accuracy	召回率/% Recall	F1值/% F1	推理效率 AIT/s	浮点运算/G FLOPs
×	×	×	97.6	92.4	95.0	12	4.5
✓	×	×	95.8	88.3	91.5	15	4.0
✓	✓	×	96.1	89.4	92.0	30	3.6
✓	✓	✓	97.0	91.8	94.4	44	3.2

注: ×表示去除对应模块; ✓表示含有对应模块。
Note: × indicates that the corresponding module is removed, ✓ indicates that the corresponding module is contained.

佳,其准确率为97.0%,推理速度达44 AIT/s,模型体积为34 Mb,浮点计算量仅为原TransformerCNN的30.5%。尽管准确率较基础模型下降0.6个百分点,但其推理速度提升267%,模型参数量减少32%,优于准确率为92.3%、推理速度20 AIT/s的MobileNetV3,以及准确率96.4%、推理速度33 AIT/s的ShuffleNetV2。结果表明,EfficientViT通过多尺度线性注意力与深度卷积的协同设计,在复杂声学特征提取与计算效率方面表现良好。

表4 应用不同轻量级骨干网络对比分析
Table 4 Comparison analysis results of lightweight backbone networks

模型名称 Model name	准确率/% Accuracy	召回率/% Recall	F1值/% F1	推理速度/ (AIT/s)	浮点计算/G FLOPs	模型大小/Mb Model size
TransformerCNN	97.6	92.4	95.0	12	10.5	50
TC+ MobileNetV3	92.3	93.2	92.7	20	4.0	40
TC+ ShuffleNetV2	96.4	94.1	95.2	33	3.5	42
TC+ FasterNet	94.8	92.4	93.6	30	5.2	35
TC+ EfficientVit	97.0	91.8	94.4	44	3.2	34

2.4 不同模型性能对比分析

由表5可知,经相关主流音频分类模型对比分析,TC+EfficientViT准确率为97.0%,超过Faster R-CNN的91.2%和LSTM的90.7%,接近VGG16的96.5%。其推理速度达44 AIT/s,为VGG16的3.7倍,模型体积为34 Mb,仅为VGG16

的28.3%。相较于准确率94.8%、推理速度15 AIT/s的VGGish,本模型在保持相近精度的同时,计算效率提升193%。结果显示,基于轻量化双通道架构的设计策略,能够在降低计算量的同时,满足规模化养殖场景对实时检测与边缘部署的需求。

表5 相关主流分类模型对比分析
Table 5 Comparison analysis of mainstream models

模型名称 Model name	准确率/% Accuracy	召回率/% Recall	F1值/% F1	推理速度/ (AIT/s)	浮点计算/G FLOPs	模型大小/Mb Model size
FasterR-CNN	91.2	90.5	90.8	28	15.2	100
VGG16	96.5	92.9	94.6	12	13.4	120
VGGish	94.8	93.5	94.1	15	7.5	80
LSTM	90.7	90.4	90.6	16	5.5	60
TC+ EfficientVit	97.0	91.8	94.4	44	3.2	34

3 讨论

本研究提出的轻量化 TransformerCNN 模型在笼养鸡发声检测任务中展现出一定优势。现从模型性能、计算复杂度以及实际应用与局限性三方面进行分析。

1) 模型性能分析。通过消融试验分析了模型中不同模块的作用。引入线性自注意力机制和 Hardswish 激活函数后,模型的召回率和精度得到了提升,推理速度也得到了优化,浮点计算量减少,表明这些改进有效提升了模型在复杂特征提取上的能力,同时减少了计算资源消耗。

相比袁超等^[7]基于 VGG16 的鸡咳嗽声检测模型,本研究在分类精度仅下降 0.65% 的情况下,将模型体积从 120 Mb 缩小至 34 Mb,推理速度提升 2 倍以上,降低了边缘设备的部署门槛。并且,本研究模型在噪声环境下的鲁棒性表现良好,进一步提升了实际应用的可靠性。与 Cuan 等^[8]提出的 CSCNN 相比,本研究在发声识别任务上的准确率提升近 4%,表明线性自注意力机制能够有效抑制噪声,并增强对局部关键特征的聚焦能力,从而提升模型对异常声音的检测能力。

2) 计算复杂度分析。在计算复杂度方面,Liao 等^[25]的 TransformerCNN 在家猪声音分类任务中取得了 96.05% 的准确率,并在 AUC 和召回率方面表现出较高的识别能力。但是该模型计算复杂度较高,未针对边缘设备的资源限制进行优化,难以满足实时检测需求。为解决 TransformerCNN 计算复杂度较高的问题,本研究引入 EfficientViT 轻量化骨干网络,将计算量从 10.5 G FLOPs 降至 3.2 G,使推理速度从 12 AIT/s 提升至 44 AIT/s,提高了实时性。同时,本研究进一步优化了模型参数量,使模型在笼养鸡舍环境下表现出稳定的检测效果。基于上述优化,后续模型可实现鸡只发声模式的实时监测,从而动态评估鸡只的热应激状态。

3) 实际应用与局限性分析。在模型对各类型声音的检测性能上,鸡只咳嗽声和鸣叫声的识别率均表现良好。其中,鸣叫声的识别准确率较高,能够很好地满足鸣叫声次数统计的需求,为后续鸡只热应激分析提供了可靠的数据支持。咳嗽声的识别准确率略低于鸣叫声,可能是由于咳嗽声的特征参数与环境噪声存在一定相似性,从而导致部分识别错误。由于咳嗽声通常用于判断鸡只是否存在呼吸道疾病,且主要以次数统计为依据,因此本模型能较好地满足检测需

求。此外,本研究模型能够在噪声较大的环境中维持较高的识别精度,进一步证明了其在实际养殖场景中的应用潜力。综上,本研究提出的轻量化模型在鸡只声音检测中表现良好,能够满足实际应用需求。

本研究的创新性体现在 2 个方面。第一,将轻量化 Transformer 架构应用于鸡只声学检测任务,并结合多尺度线性注意力与深度卷积,使计算效率与特征表达能力达到平衡。第二,采用 Hardswish 激活函数优化特征提取能力,避免了 ReLU 对音频负值特征的损失,从而提高了模型的分类精度。综上所述,本研究提出的 TransformerCNN 轻量化模型在笼养鸡发声检测中展现出良好的性能,具有较强的鲁棒性与部署灵活性,为鸡群健康监测与智能化管理提供了有力支持。

4 结论

本研究提出了一种基于轻量化 TransformerCNN 模型的笼养鸡发声检测方法,并验证了不同轻量化骨干网络对模型性能的影响,在笼养鸡发声识别任务中实现了高精度、低计算复杂度以及实时检测需求,为鸡只发声的监控和管理提供一个高效且适用的解决方案。

参考文献 References

- [1] Manteuffel G, Puppe B, Schön P C. Vocalization of farm animals as a measure of welfare[J]. *Applied Animal Behaviour Science*, 2004, 88(1/2): 163-182
- [2] Ganapathy K. Diagnosis of infectious bronchitis in chickens [J]. *In Practice*, 2009, 31(9): 424-431
- [3] 辛翔飞, 郑麦青, 文杰, 王济民. 2023 年我国肉鸡产业形势分析、未来展望与对策建议[J]. *中国畜牧杂志*, 2024, 60(3): 312-317
- Xin X F, Zheng M Q, Wen J, Wang J M. Situation analysis, future prospect, and countermeasures of China's broiler industry in China in 2023[J]. *Chinese Journal of Animal Husbandry*, 2024, 60(3): 312-317 (in Chinese)
- [4] Carroll B T, Anderson D V, Daley W, Harbert S, Britton D F, Jackwood M W. Detecting symptoms of diseases in poultry through audio signal processing [C]. In: *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Atlanta: IEEE, 2014: 1132-1135
- [5] Rizwan M, Carroll B T, Anderson D V, Daley W, Harbert S, Britton D F, Jackwood M W. Identifying rale sounds in chickens using audio signals for early disease detection in poultry [C]. In: *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Washington: IEEE, 2016: 55-59
- [6] Mahdavian A, Minaei S, Marchetto P M, Almasganj F, Rahimi S, Yang C. Acoustic features of vocalization signal in poultry health monitoring [J]. *Applied Acoustics*, 2021, 175: 107756
- [7] 袁超, 沈明霞, 姚文, 刘龙申, 陈佳. 基于发声特征和深度学习的白羽肉鸡全生命周期咳嗽检测方法[J]. *南京农业大学学报*, 2023, 46(5): 975-985
- Yuan C, Shen M X, Yao W, Liu L S, Chen J. Detection method for whole-life cycle cough of white-feathered broiler based on vocalization

- characteristics and deep learning [J]. *Journal of Nanjing Agricultural University*, 2023, 46(5): 975-985 (in Chinese)
- [8] Cuan K X, Zhang T M, Huang J D, Fang C, Guan Y. Detection of avian influenza-infected chickens based on a chicken sound convolutional neural network[J]. *Computers and Electronics in Agriculture*, 2020, 178: 105688
- [9] Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers[C]. In: *Computer Vision - ECCV 2020*. Cham: Springer International Publishing, 2020: 213-229
- [10] Sun Z G, Zhang M, Liu J G, Wang J Q, Wu Q Y, Wang G T. Research on white feather broiler health monitoring method based on sound detection and transfer learning [J]. *Computers and Electronics in Agriculture*, 2023, 214: 108319
- [11] Carpentier L, Berckmans D, Youssef A, Berckmans D, van Waterschoot T, Johnston D, Ferguson N, Earley B, Fontana I, Tullo E, Guarino M, Vranken E, Norton T. Automatic cough detection for bovine respiratory disease in a calf house[J]. *Biosystems Engineering*, 2018, 173: 45-56
- [12] 汪开英, 赵晓洋, 何勇. 畜禽行为及生理信息的无损监测技术研究进展 [J]. *农业工程学报*, 2017, 33(20): 197-209
- Wang K Y, Zhao X Y, He Y. Review on noninvasive monitoring technology of poultry behavior and physiological information [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2017, 33(20): 197-209 (in Chinese)
- [13] Inouye J M, Blemker S S, Inouye D I. Towards undistorted and noise-free speech in an MRI scanner: Correlation subtraction followed by spectral noise gating [J]. *The Journal of the Acoustical Society of America*, 2014, 135(3): 1019-1022
- [14] van Maercke D, Salomons E. The Harmonoise sound propagation model: further developments and comparison with other models[J]. *Proceedings NAG-DAGA conference, Rotterdam, The Netherlands*. 2009
- [15] Hill E M, Koay G, Heffner R S, Heffner H E. Audiogram of the chicken (*gallus gallus domesticus*) from 2 Hz to 9 kHz [J]. *Journal of Comparative Physiology A*, 2014, 200(10): 863-870
- [16] Manning L, Chadd S A, Baines R N. Key health and welfare indicators for broiler production [J]. *World's Poultry Science Journal*. 2007, 63 (1): 46-62
- [17] 杜晓冬, 滕光辉, Norton T, 王朝元, 刘慕霖. 基于声谱图纹理特征的蛋鸡发声分类识别[J]. *农业机械学报*, 2019, 50(9): 215-220
- Du X D, Teng G H, Norton T, Wang C Y, Liu M L. Classification and recognition of laying hens' vocalization based on texture features of spectrogram [J]. *Transactions of the Chinese Society of Agricultural Machinery*, 2019, 50(9): 215-220 (in Chinese)
- [18] Li L, Doroslovački M, Loew M H. Approximating the gradient of cross-entropy loss function[J]. *IEEE Access*, 2020, 8: 111626-111635
- [19] Aydin A, Bahr C, Viazzi S, Exadaktylos V, Buyse J, Berckmans D. A novel method to automatically measure the feed intake of broiler chickens by sound technology [J]. *Computers and Electronics in Agriculture*, 2014, 101: 17-23
- [20] Liu X Y, Peng H W, Zheng N X, Yang Y Q, Hu H, Yuan Y X. EfficientViT: Memory efficient vision transformer with cascaded group attention [C]. In: *CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver: IEEE, 2023: 14420-14430
- [21] Chollet F. Xception: Deep learning with depthwise separable convolutions [C]. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017: 1251-1258
- [22] Sun L J, Wu C Y, Ching W K, Akutsu T. On the compressive power of autoencoders with linear and ReLU activation functions [J]. *Neural Computation*, 2025, 37(2): 235-259
- [23] Katharopoulos A, Vyas A, Pappas N, Fleuret F. Transformers are RNNs: Fast autoregressive transformers with linear attention [C]. In: *Proceedings of the 37th International Conference on Machine Learning*. Vienna: Proceedings of Machine Learning Research, 2020, 199: 5156-5165
- [24] Xiong Y, Zeng Z, Chakraborty R, Tan M, Fung G, Li Y, Singh V. Nyströmformer: A nyström-based algorithm for approximating self-attention [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(16): 14138-14148
- [25] Liao J, Li H X, Feng A, Wu X, Luo Y J, Duan X L, Ni M, Li J. Domestic pig sound classification based on TransformerCNN[J]. *Applied Intelligence*, 2023, 53(5): 4907-4923

责任编辑: 秦梅



第一作者简介: 祝万军, 南京农业大学人工智能学院硕士研究生。主要从事智慧畜禽养殖技术的研究, 农业农村部养殖装备重点实验室成员。



通讯作者简介: 沈明霞, 南京农业大学人工智能学院教授, 博士生导师, 兼任江苏省自动化学会副主任委员, 中国畜牧医学会信息技术分会常务理事, 农业农村部养殖装备重点实验室主任。主要从事智慧畜禽养殖, 长期致力于自动化技术、微电子技术及通信技术等在现代农业领域的应用研究。主持了 863、科技部 农业科技成果转化资金项目、135 国家重点研发专项等多项国家省部级项目, 带领一支多学科交叉的精细养殖技术与装备研发与推广团队, 2014 年被授予江苏省科教系统“巾帼标兵岗”。