

基于多层注意力机制的蜜蜂音频识别方法

廖越颖 黄闽英

西南民族大学 计算机科学与人工智能学院 四川 成都 610065

摘要: 传统蜜蜂活动监测依赖人工检查, 存在工作量大、易干扰蜜蜂等问题。计算机视觉和无线物联网技术的发展推动了自动化监测, 但图像识别受光照和环境影响, 难以监测隐蔽天敌。相比之下, 声音识别不受光线限制, 能捕捉生物声音, 监测范围更广, 精度更高。本研究提出两种方案提高天敌监测精度: 一是采用声音识别技术, 弥补图像监测不足; 二是引入多特征融合, 结合SSD特征提取、散射图生成、FCN分析和HGCN注意力机制优化模型。实验表明, 该方法较传统方法效益提升5.8%至44%。

关键词: 蜜蜂天敌监测; 声音识别; 多特征融合

本文针对传统蜜蜂监测效率低、易干扰蜂群的问题, 提出一种融合听觉、视觉和人工特征的蜜蜂声音识别方法。利用SVM与FCN提取多模态特征, 采用HGCN进行特征融合, 并构建多层次注意力机制聚焦关键特征, 在噪声环境下有效提升监测准确度。实验结果显示该方法显著优于传统技术。

1 相关工作

1.1 蜜蜂监测研究

Soares^[1] (2022) 结合 MFCC、时间与频率特征, 使用 SVM 和随机森林分类蜂王存在, 准确率达 0.99, 优于 CNN, 但受限于数据集不足。Phan^[2] 同年通过超参数调优结合 MFCC 与多种算法, 在 BUZZ1 和 BUZZ2 数据集上显著提高准确率, ET 方法在 BUZZ2 上达 96.9%。Borgianni^[3] (2023) 在 Beehive 和 NU-Hive 数据集上测试 DenseNet121、ResNet50、InceptionV3, InceptionV3 准确率最高 (72%), 建议结合集成技术提升性能。Libal^[4] (2024) 利用 MFCC 和 LASSO 进行特征选择, 并采用自动编码器神经网络, 实现 90% 分类准确率, 为蜂群情绪检测提供新思路。研究表明, 优化特征提取、降维及集成方法可有效提升深度学习在蜜蜂声音识别中的准确性与鲁棒性。

1.2 声音识别研究

声音识别涵盖语音、音乐、情感、环境及生物声音, 广泛用于数据训练、环境分类、数据增强和模式识别。Kong 等^[5] (2020) 提出 Wavegram-CNN 与

Wavegram-Logmel-CNN, 结合 PANN 平衡性能与计算复杂度, 但跨任务适应性待验证。Toffa^[6] (2021) 融合 LBP 与音频特征, 适用于数据稀缺环境。Nanni^[7] (2021) 结合频谱图、信号增强与预训练 CNN 提升分类准确性。Verbitskiy^[8] (2022) 通过加宽因子的 CNN 和数据增强优化评估集性能, 但独立测试仍需验证。整体来看, 研究者们提升了音频模式识别性能, 但仍需优化数据依赖性、泛化能力和计算资源, 以扩展应用场景。

2 基于多层注意力机制的蜜蜂声音识别模型

本文提出一种基于多层注意力机制的蜜蜂声音识别模型, 包括数据预处理、多特征提取和特征融合三个阶段。首先对音频信号进行预加重、分帧加窗、端点检测和增强处理; 然后提取听觉、视觉和人工特征, 分别通过SVM与FCN网络进行深度特征提取; 最后采用融合类型注意力和节点注意力的HGCN模型动态分配权重, 实现特征融合, 提升模型的鲁棒性与识别精度。

2.1 数据预处理

该阶段进行信号处理和声音增强。因为一般的声音信号都属于非平稳的信号, 需要对其进行预处理。将音频通过预加重、分帧加窗、端点检测变成可用信号后, 为了减少噪声的干扰, 采取声音增强方法, 再将可用的信号做增强处理, 本次实验使用了两种声音增强方法: 时移增强和音高变换增强。

2.2 基于多特征的提取方法

本方法结合了音频的听觉特征、视觉特征还有人工特征, 再分别使用SVM和FCN模型来提取特征。然后通过HGCN层融合, 通过计算不同特征之间的关联性动态地分配权重, 使模型聚焦于更有价值的特征。

2.2.1 听觉特征

音频文件首先分别被转化成SSD (统计频谱描述

通讯作者: 黄闽英 (1975-), 女, 副教授, 研究方向: 信息管理与人力资源大数据分析。E-mail: hmy101@swun.edu.cn。

作者简介: 廖越颖 (1999-), 女, 硕士研究生, 研究方向为: 农业信息化。E-mail: lyy1209366209@163.com。

符), 分为两步, 首先是时域信号转换为频域, 然后计算计算频谱特征。

通过短时傅里叶变换(STFT)将时域信号 $x(n)$ 转换为频域信号 $X(n)$ 。

$$FT\{x(t)\}(m, w) = X(m, w) = \sum_{n=-\infty}^{\infty} x(n) \cdot w(n-m) \cdot e^{-j\omega n} \quad \#(1)$$

分别计算频谱质心、频谱带宽、频谱对比度、频谱平坦度等特征。

频谱质心 $C(t)$:

$$C(t) = \frac{\sum_{k=0}^{N-1} f(k) \cdot |S(t, k)|}{\sum_{k=0}^{N-1} |S(t, k)|} \quad \#(2)$$

频谱带宽 $B(t)$:

$$B(t) = \sqrt{\frac{\sum_{k=0}^{N-1} (f(k) - C(t))^2 \cdot |S(t, k)|}{\sum_{k=0}^{N-1} |S(t, k)|}} \quad \#(3)$$

频谱对比度 $SC(b)$:

$$SC(b) = 10 \log_{10} \left(\frac{\text{峰值能量}}{\text{谷值能量}} \right) \quad \#(4)$$

频谱平坦度 $SF(t)$:

$$SF(t) = \frac{\left(\prod_{k=0}^{N-1} |S(t, k)| \right)^{\frac{1}{N}}}{\frac{1}{N} \sum_{k=0}^{N-1} |S(t, k)|} \quad \#(5)$$

频谱滚降点 $R(t)$:

$$R(t) = f(k) \text{ such that } \sum_{j=0}^k |S(t, j)| = 0.85 \sum_{j=0}^{N-1} |S(t, j)| \quad \#(6)$$

频谱熵 $H(t)$:

$$H(t) = - \sum_{k=0}^{N-1} p(k) \log p(k) \quad \#(7)$$

其中, $f(k)$ 是频率索引 k 对应的频率值、 $|S(t, k)|$ 是频谱强度、 N 是频谱中的频率分量数、 $p(k) = \frac{|S(t, k)|}{\sum_{j=0}^{N-1} |S(t, j)|}$ 是归一化后的频谱强度。

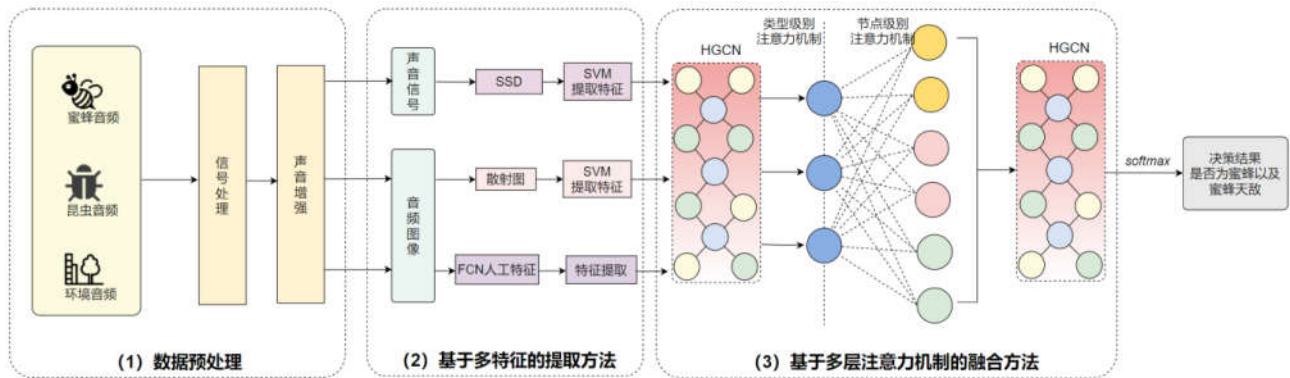


图1 SSF-TN总体框架

2.2.2 视觉特征

音频文件通过散射图转换为视觉特征, 并计算纹理描述符进行视觉特征提取, 最后采用SVM模型实现特征提取, 有效实现音频分类与相似度评估。

1) 音频信号转化为散射图

散射图通过分层小波变换与非线性变换相结合来生成表示。假设音频信号 $x(t)$ 被输入到散射网络中:

第一层小波变换: 首先, 使用尺度参数为 j 的小波函数(其中 θ 表示方向)对音频信号 $x(t)$ 进行卷积:

$$U_1(t, j, \theta) = |x * \psi_{j, \theta}(t)| \quad \#(8)$$

其中 $*$ 表示卷积运算, U_1 表示第一层散射系数。该步骤提取了输入信号的高频特征。

第二层小波变换: 对第一层变换后的结果进行二次小波变换, 用于捕捉不同尺度间的依赖关系:

$$U_2(t, j_1, \theta_1, j_2, \theta_2) = |U_1 * \psi_{j_2, \theta_2}(t)| \quad \#(9)$$

通过多级小波变换, 可以捕获到更深层的高频信息, 从而构建出散射图。

低通滤波: 对每一层输出应用低通滤波器 ϕ_J (尺度为 J)以获得平滑的散射系数, 并形成最终的散射图表示:

$$S(x) = U * \phi_J(t) \quad \#(10)$$

通过多层变换, 散射图表示能够保留信号的稳定特征, 适用于进一步的特征提取。

2) 纹理描述符的计算

为了从散射图中提取视觉特征, 可以计算一组纹理描述符来描述散射图中的纹理信息, 本论文使用的是多重局部相位量化(MLPQ)。

首先, 对于输入散射图或图像的每个像素点, 首先将其局部邻域定义为一个小的窗口(3×3), 在此窗口内, 通过傅里叶变换获取每个像素的局部相位信息。设局部邻域窗口内的像素值为 $f(x, y)$, 对其应用二维离散傅

里叶变换 (DFT) :

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi \left(\frac{ux}{M} + \frac{vy}{N} \right)} \quad \#(11)$$

其中 M 和 N 是窗口的行和列数, u, v 是频域坐标, $F(u,v)$ 的相位角度可以作为量化基础。

然后计算每个频域分量的相位角度:

$$\theta(u,v) = \arg(F(u,v)) \quad \#(12)$$

$\arg(\cdot)$ 表示复数的相位角函数。

再将相位角度进行量化以减少噪声的影响, 并增强稳定的纹理信息。设定相位量化的级数(4级), 将相位角度 $\theta(u,v)$ 映射到离散化的量化级别:

$$Q(\theta) = \frac{\theta(u,v)}{\frac{2\pi}{K}} \quad \#(13)$$

其中 K 表示量化级数, $Q(\theta)$ 为量化后的相位索引。

将量化后的相位索引转化为二进制编码, 用于构成最终的特征描述符。

而对于每个局部邻域的二进制编码, 使用直方图累积相邻窗口的二进制模式, 构建 MLPQ 描述符。计算散射图或图像中每个窗口内的量化相位编码直方图:

$$H(k) = \sum_{i=1}^M \sum_{j=1}^N \delta(Q(\theta_{ij}) - k) \quad \#(14)$$

其中 k 为量化级别, δ 为指示函数, 当 $Q(\theta_{ij}) = k$ 时, $\delta = 1$, 否则为0。

最后, 将所有局部窗口的直方图合并, 生成完整的 MLPQ 纹理描述符向量:

$$MLPQ = [H(0), H(1), \dots, H(K-1)] \quad \#(15)$$

3) 使用 SVM 提取特征

从图像中提取人工提取的纹理描述符特征集, 和 SSD 特征集, 都分别用单独的 SVM 进行分类, 结果用识别准确率来评估。论文使用的基本分类器是一个带有径向基函数 RBF 核的单对多 SVM, 并使用网格搜索法可得最优参数为惩罚因子 $C = 1000$, 和核函数参数 $\gamma = 0.1$ 。在分类步骤之前, 特征被线性归一化为 $[0, 1]$ 。一个集合体中任意一组 SVM 的输出结果通过总和规则 (sum rule) 进行组合:

$$Sum Rule(v) = \arg \max_{k=1, \dots, c} \sum_{i=1}^n P(\omega_x | l_i(v)) \quad \#(16)$$

v 要分类的模式, n 表示分类器的数量, l_i 表示任务中第 i 个分类器的输出标签, 其中可能的类别标签为 $\Omega = \omega_1, \omega_2, \dots, \omega_c$, c 是类的数量, $P(\omega_x | l_i(v))$ 是根据第 i 个分类器对属于类 ω_x 的模式 v 的概率的估计。

最后的集合体决策是获得最大支持的类别:

$$\sum_{cl=1}^N score(cl, j, x) \quad \#(17)$$

N 是集合体中分类器的数量, $score(cl, j, x)$ 是分类器 cl 对给定类 j 的输出 x 。

2.2.3 人工特征

本方法采用迁移学习, 将预训练的 FCN 特征提取层微调用于蜜蜂音频数据分类。通过网格搜索确定最优架构, 共4个卷积层, 核大小依次为 3×3 (最后一层为 1×1), 使用自适应激活函数和 same padding, 卷积层后添加 maxpooling 和全局平均池化, dropout 设为 0.4 防止过拟合, 实现对不同输入大小的有效分类。

设输入音频数据的特征图维度为 $X \in R^{T \times F}$, 对于卷积运算, 可以通过以下公式表示:

$$Y^{(l)} = f(W^{(l)} * X^{(l-1)} + b^{(l)}) \quad \#(18)$$

$W^{(l)}$ 表示第 l 层的卷积核, 大小为 3×3 , $X^{(l-1)}$ 表示第 $l-1$ 层的输入特征图, $b^{(l)}$ 是偏置项, 经过三层卷积输出的特征大小为 $T \times F \times 3$ 。

在每个卷积层之后应用池化操作, 除了最后一层:

$$Y^{(l)} = \max(X_{i,j,k}^{(l-1)}) \quad \#(19)$$

在卷积层4之后, 使用全局平均池化来减少特征图的大小:

$$Y = \frac{1}{n} \sum_{i=1}^n X_i \quad \#(20)$$

输出大小: $1 \times 1 \times 3$ 。

为了避免过拟合, 在全局平均池化层之后添加 dropout 层, 比例为 0.4:

p 为 0.4, 表示有 40% 的神经元会被随机丢弃。

最后, 输出层使用 softmax 激活函数进行分类, 输出类别概率:

$$P(c|x) = \frac{\exp(z_c)}{\sum_{k=1}^K \exp(z_k)} \quad \#(21)$$

其中 z_c 是类别 c 的得分, K 是类别总数。

2.3 基于多层注意力机制的特征融合

1) 第一层 HGCN

设音频文件经过前两步特征提取后得到的特征向量分别为 X_{bee} 、 $X_{cricket}$ 和 $X_{environment}$, 这些特征向量首先会通过一个图卷积网络 (HGCN) 生成嵌入表示, 每一层图卷积的特征嵌入计算公式为:

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad \#(22)$$

其中, $H^{(l)}$ 是第 l 层的特征表示, 输入层有 $H^{(0)} = X$, $W^{(l)}$ 是第 l 层的可学习权重矩阵。

经过多层图卷积后, 得到最终的特征嵌入。假设图卷积网络共包含 L 层, 则最终的特征嵌入表示为:

$$H_{embedding} = H^{(L)} \# (23)$$

对于每种特征，可以应用上述图卷积操作，得到其对应的嵌入表示：

$$H_{bee} = HGCN(X_{bee}) \# (24)$$

$$H_{cricket} = HGCN(X_{cricket}) \# (25)$$

$$H_{environment} = HGCN(X_{environment}) \# (26)$$

在特征嵌入之后，接下来是类型级别注意力和节点级别注意力的计算。

2) 类型级别注意力

类型级别注意力机制的目的是在不同的特征类型之间分配权重，以确定不同特征类型在最终融合中的重要性。假设特征嵌入后的不同类型特征表示为 H_{bee} 、 $H_{cricket}$ 和 $H_{environment}$ 。

对每种类型的特征进行线性变换，以映射到相同的特征空间，便于后续计算注意力，对于每种类型，计算其注意力系数 β_{bee} 、 $\beta_{cricket}$ 和 $\beta_{environment}$ ，以反映不同特征类型的重要性，公式如下：

$$\beta_{bee} = \frac{\exp(w^T \tanh(\tilde{H}_{bee}))}{\sum_{type \in \{bee, cricket, environment\}} \exp(w^T \tanh(\tilde{H}_{type}))} \# (27)$$

$$\beta_{cricket} = \frac{\exp(w^T \tanh(\tilde{H}_{cricket}))}{\sum_{type \in \{bee, cricket, environment\}} \exp(w^T \tanh(\tilde{H}_{type}))} \# (28)$$

$$\beta_{environment} = \frac{\exp(w^T \tanh(\tilde{H}_{environment}))}{\sum_{type \in \{bee, cricket, environment\}} \exp(w^T \tanh(\tilde{H}_{type}))} \# (29)$$

其中， w 是类型级注意力的可学习权重向量，用于映射特征的非线性关系。

然后使用注意力系数对不同类型的特征进行加权融合，得到最终的类型级融合特征 H_{fusion} ：

$$H_{fusion} = \beta_{bee} \cdot H_{bee} + \beta_{cricket} \cdot H_{cricket} + \beta_{environment} \cdot H_{environment} \# (30)$$

3) 节点级别注意力

接着是节点级别注意力机制，用于在图结构中为每个节点分配权重，以便更有效地聚合邻居节点的信息。若 H_{fusion} 是类型融合后的特征，以下是节点级注意力的计算步骤。

对每个节点特征向量 $h_i \in H_{fusion}$ 进行线性变换，得到新的特征表示。然后对于每个节点 i 和其邻居节点 j ，计算它们之间的注意力系数 e_{ij} ：

$$e_{ij} = \text{LeakyReLU}(a^T [h_i \| h'_j]) \# (31)$$

其中 a 是节点级注意力的可学习权重向量， $\|$ 表示向量连接操作， LeakyReLU 是用于非线性激活的函数。

然后通过 softmax 函数对邻居节点的注意力系数进行归一化，得到注意力权重 α_{ij} ：

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})} \# (32)$$

其中 $N(i)$ 表示节点 i 的邻域节点集合。

最后，根据注意力权重 α_{ij} 聚合邻居节点的特征，得到节点 i 的最终输出特征 h_i^{out} ：

$$h_i^{out} = \sigma \left(\sum_{j \in N(i)} \alpha_{ij} h'_j \right) \# (33)$$

4) 第二层HGCN

输出的特征 h_i^{out} 再经过一层HGCN网络后，得到 $H^{(l+1)}$ 将包含经过异质图卷积后的更新特征。

为了将节点特征转化为图级特征，使用全局平均池化，则对于特征矩阵 $H^{(l+1)}$ ，可以得到图级特征表示 H_{global} ：

$$H_{global} = \frac{1}{N} \sum_{i=1}^N h_i^{(l+1)} \# (34)$$

其中 $h_i^{(l+1)}$ 表示经过HGCN更新后的每个节点特征， N 为节点总数。

再把上一步得到的将全局池化得到的特征 H_{global} 传入一个全连接层，并通过 softmax 激活函数得到最终的分类结果。类别数量为 C ，全连接层的权重矩阵为 W_{fc} ，偏置为 b_{fc} ，则决策结果可以表示为：

$$y = \text{softmax}(W_{fc} H_{global} + b_{fc}) \# (35)$$

最终，类别 c 的决策通过选择 y 中概率最大的元素来确定：

$$\tilde{c} = \arg \max_c y_c \# (36)$$

通过以上步骤，在完成类型级别注意力和节点级别注意力之后，特征经过了两次加权融合，使得最终的节点表示既包含了不同特征类型的重要性信息，也包含了邻居节点的局部信息。这种多级注意力机制能够有效地提升模型的表示能力。再经过一层 HGCN，得到更新后的节点特征后，通过全局池化层生成图级特征，最后使用全连接层和 softmax 激活函数生成类别概率分布，进而确定最终的分类决策。

3 实验设置与结果分析

3.1 数据集准备和处理

本研究选用的三个类别的数据集分别是论文里的公开数据，该数据集在物种识别预测研究中被广泛使用。本研究选用的类别名与数量信息具体如表3-1所示。

表1 数据集统计

类别	数量
Bee ^[9]	7000
InsectSet66 ^[10]	5250
UrbanSound8K ^[11]	8732

3.2 评价指标

本文参考文献选取了常用的评价指标：准确率（正确预测比例）、精确率（预测为正的准确程度）、召回率（正样本识别能力）和F1值（精确率与召回率的调和平均）。

3.3 基线方法

本文选择mel_FCN^[12]、mlpq_scatter_SVM^[13]、SSD_SVM^[14]和mel_Inceptionv3^[15]四种代表性基线模型进行对比验证。其中mel_FCN利用梅尔频谱图和FCN模型；mlpq_scatter_SVM采用散射图特征与SVM分类；SSD_SVM基于SSD声觉特征与SVM分类；mel_Inceptionv3则结合梅尔频谱图与Inceptionv3模型进行分类。

3.4 实验结果及分析

本章节首先将本文提出的方法从准确率、精确度、召回率和F1值4个方面，与基线方法在数据集上进行对比试验，验证SSF-TN的性能；然后是消融实验，验证重要模块对方法性能的贡献。

3.4.1 对比实验

表2显示，SSF-TN在Accuracy、Precision、Recall和F1值上均显著优于基线方法（较mel_Inceptionv3提升5.8%-5.9%）。SSF-TN融合声觉、视觉与人工特征，通过HGCN的类型级和节点级注意力机制挖掘特征关联，显著提升蜜蜂音频识别准确度和可靠性。

表2 蜜蜂声音识别方法性能对比

模型	准确率	精确率	召回率	F1-Score
mel_FCN	0.55	0.54	0.56	0.55
mlpq_scatter_SVM	0.89	0.91	0.89	0.87
SSD_SVM	0.9019	0.9007	0.9019	0.9010
mel_Inceptionv3	0.9249	0.9266	0.9249	0.9250
SSF_TN	0.9822	0.9846	0.9822	0.9822

3.4.2 消融实验

为验证SSF-TN有效性，本研究设计了两类消融实验：一是移除节点级或类型级注意力机制，结果显示SSF_TN性能显著提升，证实注意力机制能突出关键信息；二是移除FCN特征，结果表明SSF_TN性能提升明显，说明FCN单独无法充分捕捉多模态特征的关联性。综上，多特征与多层次注意力融合显著增强了模型性能。

表3 消融实验结果

方法	准确率	精确率	召回率	F1-Score
SSF_TN	0.9822	0.9846	0.9822	0.9822
SSF_TN -Node	0.9654	0.9528	0.9528	0.9523
SSF_TN -Type	0.94	0.90	0.92	0.94
SSF_TN -HGCN	0.9024	0.9013	0.9024	0.9016

续表：

方法	准确率	精确率	召回率	F1-Score
SSF_TN-SSD	0.9014	0.8926	0.9028	0.8926
SSF_TN-散射图	0.8872	0.8874	0.8872	0.8872
SSF_TN-FCN	0.86	0.89	0.86	0.84

4 总结

本文探讨蜜蜂声音识别方法，提出基于多特征融合的方案，以解决现有研究局限。针对图像监测受环境影响的问题，采用声音识别技术，提升监测范围和实时性。为克服单一特征提取的不足，结合 SSD 特征提取、散射图转换和 FCN 模型分析音频图像，并通过迁移学习优化视觉特征利用率。HGCN 模型融合不同特征，结合多层次注意力机制增强特征权重分配，提高蜜蜂天敌声音识别精度，尤其在噪声环境下显著提升监测效果。

实验表明，SSF-TN 在大规模数据集上的性能优于基线方法，验证了其有效性。未来将扩展至实际蜂箱环境，解决监测挑战，并探索更多音频特征与先进技术，优化系统性能，在更广泛数据集上进行对比研究。

参考文献

[1]Soares B S, Luz J S, de Macêdo V F, et al. MFCC-based descriptor for bee queen presence detection[J]. Expert Systems with Applications, 2022, 201: 117104.

[2]Phan T T H, Nguyen-Doan D, Nguyen-Huu D, et al. Investigation on new Mel frequency cepstral coefficients features and hyper-parameters tuning technique for bee sound recognition[J]. Soft Computing, 2023, 27(9): 5873-5892.

[3]Borgianni L, Ahmed M S, Adami D, et al. Spectrogram Based Bee Sound Analysis with DNNs: a step toward Federated Learning approach[C]//2023 4th International Symposium on the Internet of Sounds. IEEE, 2023: 1-8.

[4]Libal U, Biernacki P. MFCC selection by LASSO for honey bee classification[J]. Applied Sciences, 2024, 14(2): 913.

[5]Kong Q, Cao Y, Iqbal T, et al. Panns: Large-scale pretrained audio neural networks for audio pattern recognition[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 2880-2894.

[6]Toffa O K, Mignotte M. Environmental sound classification using local binary pattern and audio features collaboration[J]. IEEE Transactions on Multimedia, 2020, 23: 3978-3985.

[7]Nanni L, Maguolo G, Brahnam S, et al. An ensemble

of convolutional neural networks for audio classification[J]. Applied Sciences, 2021, 11(13): 5796.

[8]Verbitskiy S, Berikov V, Vyshegorodtsev V. Eranns: Efficient residual audio neural networks for audio pattern recognition[J]. Pattern Recognition Letters, 2022, 161: 38-44.

[9]Zgank A. Reduced MFCC Feature Extraction Dimension for Acoustic Classification of Bee Swarm Activity[C]//2022 ELEKTRO (ELEKTRO). IEEE, 2022: 1-4.

[10]Faiß M. InsectSet47 & InsectSet66: Expanded datasets for automatic acoustic identification of insects (Orthoptera and Cicadidae)[J]. Zenodo, 2023.

[11]Malaviya P, Kumar Y, Modi N. Advancements in Environmental Sound Classification: Evaluating Machine Learning and Deep Learning Approaches on the UrbanSound8k[C]//2023 Seventh International Conference on

Image Information Processing (ICIIP). IEEE, 2023: 900-905.

[12]García-Ordás M T, Rubio-Martín S, Benítez-Andrades J A, et al. Multispecies bird sound recognition using a fully convolutional neural network[J]. Applied Intelligence, 2023, 53(20): 23287-23300.

[13]Gupta U, Gupta D. Least squares structural twin bounded support vector machine on class scatter[J]. Applied Intelligence, 2023, 53(12): 15321-15351.

[14]Yan Y. Using the Improved SSD Algorithm to Motion Target Detection and Tracking[J]. Computational Intelligence and Neuroscience, 2022, 2022(1): 1886964.

[15]Truong T H, Du Nguyen H, Mai T Q A, et al. A deep learning-based approach for bee sound identification[J]. Ecological Informatics, 2023, 78: 102274.