第 39 卷 第 3 期
2020 年 5 月

应用声学
Journal of Applied Acoustics

Vol. 39, No. 3
May, 2020

◇ 研究报告 ◇

贝叶斯优化卷积神经网络公共场所异常声识别*

曾 宇† 户文成

(北京市劳动保护科学研究所 北京 100054)

摘要: 针对公共场所异常声的感知和识别问题,提出一种基于贝叶斯优化卷积神经网络的识别方法。提取声信号的Gammatone倒谱系数、倍频程功率谱、短时能量和谱质心,组合成声信号的特征图。构建卷积神经网络作为分类器,利用递增的卷积核设置和池化操作处理不同尺度的特征。基于贝叶斯优化算法优化卷积神经网络的模型参数,对包括火苗噼啪声、婴儿啼哭声、烟花燃放声、玻璃破碎声和警报声的5种公共场所异常声进行识别。该方法的识别结果与基于不同的特征提取和分类器方案得到的识别结果进行比较,结果表明该方法的识别效果优于其他特征提取和分类器方案的识别效果。最后分析了该方法在不同信噪比噪声干扰下的识别结果,验证了该方法的有效性。

关键词: 公共场所;异常声识别;Gammatone倒谱系数;贝叶斯优化;卷积神经网络

中图法分类号: TP391.4; TP183 **文献标识码:** A **文章编号:** 1000-310X(2020)03-0409-08

DOI: 10.11684/j.issn.1000-310X.2020.03.013

Recognition of abnormal sound in public places based on Bayesian optimal convolutional neural network

ZENG Yu HU Wencheng

(*Beijing Municipal Institute of Labour Protection, Beijing* 100054*, China*)

Abstract: Aiming at the problem of abnormal sound perception and recognition in public places, a recognition method based on Bayesian optimal convolution neural network is proposed. The Gammatone cepstrum coefficients, octave power spectrum, short-term energy and spectral centroid of sound signal are extracted and combined to form the characteristic map of sound signal. Using convolution neural network as classifier, different convolution kernel settings and pooling operations are adopted to deal with different scales of features. Based on Bayesian optimization algorithm, the model parameters of convolution neural network are optimized. Five kinds of abnormal sounds in public places, including crackling of fire, crying of infants, fireworks, broken glass and alarms, are identified. Finally, the recognition results of different feature extraction and classifier schemes are compared, and the advantages of this method are illustrated. The recognition results of this method under noise jamming are analyzed, and the validity of this method is verified.

Keywords: Public place; Abnormal sound recognition; Gammatone cepstrum coefficients; Bayesian optimization; Convolutional neural network

2019-07-11 收稿; 2019-11-28 定稿

*北京市财政项目 (PXM2019_178304_000003), 北京市劳动保护科学研究所自立课题 (H194)

作者简介: 曾宇 (1979–), 男, 山西太原人, 助理研究员, 研究方向: 噪声与振动控制。

†通信作者 E-mail: zengyu@bmilp.com

0 引言

近年来随着公共场所安全问题复杂性的提高，公共场所的异常监控和危险预警得到了越来越多的关注[1]。公共场所环境中的声信号包含了大量的安全信息，异常事件的发生常会伴随特定的异常声。声频监控系统用于实时公共安全监控，所需的数据存储和传输条件都低于视频监控系统，同时也能更好地保护隐私。公共场所异常声的识别作为公共场所声频监控的关键技术之一，具有重要的研究意义和实用价值。

对于异常事件的声音识别，学者们进行了一系列的研究。韦娟等[2]对公共场所异常声进行总体平均经验模态分解并提取各层信号的Mel倒谱系数(Mel-frequency cepstrum coefficient, MFCC)、短时能量和能量比，采用改进的决策导向无环图支持向量机(Support vector machine, SVM)对枪声、爆炸声、玻璃破碎声、说话声和脚步声进行识别。胡涛等[3]将公共场所异常声分帧后各帧的Mel倒谱系数及其一阶、二阶差分按照时间先后顺序沿着不同方向排列分别形成二维和一维特征图，采用卷积神经网络对爆炸声、玻璃破碎声、枪声、警报声、开关门声和哭声进行识别。李伟红等[4]提出改进的极点对称模态分解特征提取方法，采用支持向量机对爆炸声、尖叫声、枪声与玻璃破碎声进行识别。罗森林等[5]以Mel倒谱系数为特征，将分别使用高斯混合模型(Gaussian mixed model, GMM)和支持向量机获得的识别结果进行融合，对两类枪声进行识别。刘鑫锦等[6]提取岩石脆性破坏时声信号的Mel倒谱系数、谱质心和过零率作为特征，采用高斯混合模型对颗粒弹射和岩板劈裂情况进行检测。苏国韶等[7]提取岩爆过程声信号的波形持续时间、主频及短时能量作为特征，基于随机森林对颗粒弹射、岩板劈裂和块片弹射情况进行识别。张铁民等[8]提取鸡叫声的短时过零率和短时能量，采用模糊神经网络对禽流感病鸡进行识别。韩磊磊等[9]提取生猪异常声的Mel 倒谱系数及其一阶、二阶差分，采用支持向量机对生猪打斗声、咳嗽声、喷嚏声、饥饿声和呛水声进行识别。杨元威等[10]基于KS检验和ReliefF算法对高压断路器故障时的声信号进行特征提取和选择，采用支持向量机对线圈电源低压、电磁铁卡阻、合闸弹簧疲劳、脱扣延迟和传动阻尼增大情况进行检测。王丰华等[11]对变压器噪声信号的Mel 倒谱系数进行主成分分析，基于降维后的特征采用矢量量化算法对变压器铁芯未压紧故障进行检测。

公共场所中，火灾可能导致严重的财产损失和人员伤亡，烟花爆竹燃放也在国内数百个城市中被禁止，上述研究中没有对此两类安全事件中的异常声进行分析和识别。此外不同分类器的模型参数设置对识别结果有影响，上述研究中鲜有对异常声分类器模型的参数优化。本文针对公共场所异常声的感知和识别问题，提出一种基于贝叶斯优化卷积神经网络的识别方法。提取异常声信号的Gammatone倒谱系数(Gammatone cepstrum coefficient, GTCC)、短时能量、倍频程功率谱和谱质心，经过信息融合形成特征图，整合公共场所异常声的时域、频域和倒谱特性。以卷积神经网络为分类器，设计递增的卷积核尺度和池化操作以处理不同尺度的特征，构建批量归一化层和丢弃层以提高网络模型的泛化能力。提取该卷积神经网络的网络结构参数和网络训练参数，基于贝叶斯优化算法对卷积神经网络的模型参数进行优化，对包括火苗噼啪声、婴儿啼哭声、烟花燃放声、玻璃破碎声和警报声的5种公共场所异常声进行识别。最后分析比较了基于不同的特征提取和分类器方案得到的识别结果，并对本文方法在不同信噪比噪声干扰下的识别效果进行验证。

1 公共场所异常声的特征提取

1.1 公共场所异常声的特征表示

公共场所异常声属于环境声，由于环境声与语声的相似性，语声识别中的典型特征参数也常用于环境声识别中。倍频程功率谱分析是最常用的声信号处理方法之一，倍频程功率谱谱线少频带宽，符合人耳知觉频带低频部分较窄、高频部分较宽的特点，表征了环境声信号的声学特性。

Mel倒谱系数是语声识别和说话人识别的有效特征之一，但其在低信噪比环境下识别效果较差。Gammatone滤波器可以模拟人耳基底膜的时频分析功能，在噪声条件下具有较强的抗干扰性，滤波效果更好，且经过Gammatone滤波后的信号能够更好增强目标识别系统的鲁棒性。将Mel倒谱系数计算中的滤波器替换为Gammatone滤波器得到的Gammatone倒谱系数已应用于声音识别中，在不同背景噪声环境下取得比Mel倒谱系数更好的识别效

果[12−14]。Gammatone 滤波器的时域表达式如下：

$$g_i(t) = At^{a-1}\,\mathrm{e}^{-2\pi B_w(f_i)t}\cos(2\pi f_i t + \varphi_i)U(t),\quad t \geqslant 0,\ \ 1 \leqslant i \leqslant a, \tag{1}$$

式 (1) 中，A 和 a 分别为滤波器增益和阶数，$U(t)$ 为阶跃函数，f_i 和 φ_i 分别为中心频率和相位，$B_w(f_i)$ 为等效矩阵带宽函数，其表达式为 $B_w(f_i) = 24.7 + 0.108f_i$。提取 Gammatone 倒谱系数时，首先对声信号进行加窗分帧，对每帧信号进行快速傅里叶变换，然后通过 Gammatone 滤波器组进行滤波，最后进行离散余弦变换。

异常声往往是突发的，瞬间爆发力较强，能量随时间变化比较明显。声信号的短时能量是信号在一帧时间内的能量值，在一定程度上能反映出信号在时域的幅度变化情况。对于声信号 $x_i(n)$，其短时能量为

$$E_i = \sum_{n=1}^{N} |x_i(n)|^2, \tag{2}$$

式 (2) 中，i 为帧号，N 为帧长。

谱质心是描述音色属性的重要信号特征之一。它是一定频率范围内通过能量加权平均的频率，关联着信号的基频特性；同时它也体现了声音的明亮度，声音明亮度随谱质心增加而增高。对于声信号 $x_i(n)$，其谱质心为

$$\mathrm{SC}_i = \frac{\sum_{k=1}^{K} S_{ik} f_k}{\sum_{k=1}^{K} f_k}, \tag{3}$$

式 (3) 中，i 为帧号，K 为离散傅里叶变换的长度，S_{ik} 为频率 f_k 处的功率谱值。

本文提取公共场所异常声的 Gammatone 倒谱系数、倍频程功率谱、短时能量和谱质心，将这些特征组合成特征向量，即：

$$\begin{aligned}
&\boldsymbol{V} = [\mathbf{GTCC}, \boldsymbol{E}, \mathbf{Oct}, \mathbf{SC}],\\
&\mathbf{GTCC} = [\mathrm{GTCC}_1, \mathrm{GTCC}_2, \cdots, \mathrm{GTCC}_{N_m}],\\
&\boldsymbol{E} = [E_1, E_2, \cdots, E_{N_e}],\\
&\mathbf{Oct} = [\mathrm{Oct}_1, \mathrm{Oct}_2, \cdots, \mathrm{Oct}_{N_o}],\\
&\mathbf{SC} = [\mathrm{SC}_1, \mathrm{SC}_2, \cdots, \mathrm{SC}_{N_c}],
\end{aligned} \tag{4}$$

式 (4) 中，$\boldsymbol{V}$ 为公共场所异常声特征向量，$\mathbf{GTCC}$ 为 Gammatone 倒谱系数，$\boldsymbol{E}$ 为短时能量，$\mathbf{Oct}$ 为倍频程功率谱，$\mathbf{SC}$ 为谱质心，N_m、N_e、N_o 和 N_c 分别为 Mel 倒谱系数、短时能量、倍频程功率谱和谱质心的个数。

1.2　公共场所异常声的特征图提取

二维的特征图所包含的信息量大于一维的特征向量，以特征图作为分类器的输入会增加单次分类器训练所包含的信息量，提高训练效率。常用的特征图生成方法包括短时傅里叶变换、Mel 谱图等[3,15−16]，短时傅里叶变换和 Mel 谱图分别表征了信号的频域特性和倒谱特性，但没有体现频域特性和 Mel 倒谱特性的交叉效应。本文整合公共场所异常声的时域特性、频域特性和倒谱特性，构建公共场所异常声的特征图表示，步骤如下：

(1) 对声频文件进行处理，如果声频文件采样率不同或长度不同则进行重采样并裁剪到同样长度，之后得到各声频文件的时域信号 $S_i(n)$，式中 i 为声频文件标识号；

(2) 对时域信号 $S_i(n)$ 进行分帧加窗，帧信号时长为 10∼30 s，得到分帧时域信号 $s_{ij}(m)$，式中 j 为帧标识号；

(3) 计算时域信号 $S_i(n)$ 的第 j 个分帧信号的特征向量 $\boldsymbol{V}_{ij}$，该特征向量长度为 28，包含 13 个 Gammatone 倒谱系数、13 个倍频程功率谱、1 个短时能量和 1 个谱质心；

(4) 生成时域信号 $S_i(n)$ 的特征矩阵：

$$\boldsymbol{P}_i(k,j) = \boldsymbol{V}_{ij}(k),\ \ 1 \leqslant k \leqslant 28,\ 1 \leqslant j \leqslant N_f, \tag{5}$$

式 (5) 中，$\boldsymbol{P}_i$ 为特征矩阵，k 和 j 分别为特征矩阵 $\boldsymbol{P}_i$ 的行标识和列标识，N_f 为帧数；

(5) 对时域信号 $S_i(n)$ 的特征矩阵 $\boldsymbol{P}_i$ 进行归一化，得到归一化的特征矩阵 $\bar{\boldsymbol{P}}_i$，即：

$$\begin{aligned}
&\bar{\boldsymbol{P}}_i(k,j) =\\
&\frac{2\boldsymbol{P}_i(k,j) - (\max_m \boldsymbol{P}_i(k,j) + \min_m \boldsymbol{P}_i(k,j))}{\max_m \boldsymbol{P}_i(k,j) - \min_m \boldsymbol{P}_i(k,j)},\\
&\qquad 1 \leqslant k \leqslant 28,\ 1 \leqslant j \leqslant N_f;
\end{aligned} \tag{6}$$

(6) 保存时域信号 $S_i(n)$ 的归一化特征矩阵 $\bar{\boldsymbol{P}}_i$ 作为该信号对应的公共场所异常声声频文件的特征图。

2　公共场所异常声的感知识别

2.1　卷积神经网络

卷积神经网络是一类包含卷积计算且具有深度结构的多层神经网络，是研究和应用最广泛的深度神经网络。卷积神经网络的局部连接、权值共享等特性使之可以有效地降低网络的复杂度，减少训

练参数的数目，使模型对平移、扭曲、缩放具有一定程度的不变性，并具有强鲁棒性和容错能力，且也易于训练和优化[17−18]。

卷积神经网络的基本结构由输入层、卷积层、激活层、池化层、全连接层及输出层构成，此外还可以增加批量归一化层和丢弃层以进一步强化模型的泛化能力。卷积层包含多个卷积核，神经元通过卷积核与上一层的局部区域相关联，每个卷积核以局部权值矩阵的形式遍历作为输入的特征图，进行内积运算，同一卷积核实现权值共享，加上偏置完成特征映射，即：

$$x_i^l = \sum_{j\in \boldsymbol{M}_k} \boldsymbol{X}_j^{l-1}\boldsymbol{W}_{ji}^l + \boldsymbol{b}_i^l,\quad 1\leqslant i\leqslant N_c^l, \tag{7}$$

式(7)中，x_i^l为第l层的第i个卷积核的特征映射，N_c^l为第l层的卷积核数量，$\boldsymbol{X}_j^{l-1}$是第$l-1$层的第j个特征映射图，$\boldsymbol{M}_k$为第$l-1$层中与第l层第i个卷积核进行卷积运算的所有特征映射，$\boldsymbol{W}_{ji}^l$为权值矩阵，$\boldsymbol{b}_i^l$为偏置。$\boldsymbol{W}_{ji}^l$和$\boldsymbol{b}_i^l$会随着网络训练而自动变化，$\boldsymbol{b}_i^l$的各分量的初值均为0，$\boldsymbol{W}_{ji}^l$采用Glorot初始化，即$\boldsymbol{W}_{ji}^l$的各分量独立采样自均匀分布，该均匀分布均值为0，方差如下：

$$\text{var} = \frac{2}{\dim W_1 * \dim W_2 * N_{ch} + \dim W_1 * \dim W_2 * N_c^l}, \tag{8}$$

式(8)中，var为方差，$\dim W_1$和$\dim W_2$为卷积核的两个维度，N_{ch}为输入通道数。激活层实现卷积神经网络对非线性特征的检测，典型的激活函数有sigmoid、tanh和ReLU等，其中ReLU激活函数是最为常用的，其形式为

$$\text{ReLU}(x) = \max_x(0, x). \tag{9}$$

池化层旨在通过降低特征映射的分辨率来实现移位不变性，典型的池化类型有最大池化、均值池化等，其中最大池化函数为

$$\text{Pool}_{\max}(\boldsymbol{X}) = \max(\boldsymbol{X}), \tag{10}$$

式(10)中，$\boldsymbol{X}$是维度为$\dim P_1 * \dim P_2$的矩阵，$\dim P_1$和$\dim P_2$为池化窗口的尺度。全连接层与卷积层类似，差别在于该层的所有神经元均与前一层的所有神经元相连，即：

$$x_i^l = \boldsymbol{X}^{l-1}\boldsymbol{W}_i^l + \boldsymbol{b}^l,\quad 1\leqslant i\leqslant N_f^l, \tag{11}$$

式(11)中，x_i^l为第l层的第i个特征映射，N_f^l为第l层的特征映射数，$\boldsymbol{X}^{l-1}$是第$l-1$层的特征映射图，$\boldsymbol{W}_i^l$和$\boldsymbol{b}^l$分别为权值矩阵和偏置。批量归一化层对输出数据进行规范化处理，即：

$$y_i = \gamma\frac{x_i - \mu_b}{\sqrt{\sigma_b^2 + \varepsilon}} + \beta, \tag{12}$$

式(12)中，x_i为输入的第i个特征映射图，y_i为x_i规范化处理的结果，γ为缩放系数，β为偏移，γ和β会随着网络训练而自动变化，γ和β的初值分别为1和0，μ_b和σ_b为数据的均值和方差，ε为一个非常小的数，旨在提高方差σ_b极小时归一化处理的稳定性。丢弃层会按照一定比例将部分输入数据赋值为0，从而避免了所有的神经元同步优化其权重以致收敛到同样的结果，防止过拟合的发生。

本文设计了包含多个卷积层和池化层的卷积神经网络，采用递增的卷积核设置和池化操作处理不同尺度的特征，并增加批量归一化层和丢弃层以避免过拟合，该网络的结构如下：

(1) 输入层；

(2) 卷积层，卷积核数量为NC，卷积核的两个维度相等，均为$\dim W$；

(3) 批量归一化层；

(4) 激活层，激活函数为ReLU函数；

(5) 池化层，池化类型为最大池化，池化窗口的两个维度分别为1和$\dim P$；

(6) 卷积层，卷积核数量为NC，卷积核的两个维度相等，都为$2 * \dim W$；

(7) 批量归一化层；

(8) 激活层，激活函数为ReLU函数；

(9) 池化层，池化类型为最大池化，池化窗口的两个维度分别为1和$\dim P$；

(10) 卷积层，卷积核数量为NC，卷积核的两个维度相等，都为$4 * \dim W$；

(11) 批量归一化层；

(12) 激活层，激活函数为ReLU函数；

(13) 丢弃层，丢弃率为rDrop；

(14) 全连接层，神经元数为分类数；

(15) 输出层。

卷积神经网络训练时，采用随机梯度下降法对卷积层和全连接层的权值和偏差、批量归一化层的缩放系数和偏移等参数进行调整，随机梯度下降法可表示为

$$\boldsymbol{\theta}_{l+1}=\boldsymbol{\theta}_l-r\mathrm{Learn}*\nabla E_R(\boldsymbol{\theta}_l)+\mathrm{mom}D*(\boldsymbol{\theta}_l-\boldsymbol{\theta}_{l-1}),$$
$$E_R(\boldsymbol{\theta}_l)=E(\boldsymbol{\theta}_l)+\frac{1}{2}*\mathrm{L2Reg}*\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W},\quad(13)$$

式(13)中，l为迭代数，$\boldsymbol{\theta}$为调整的向量，$\boldsymbol{W}$为权值向量，rLearn为学习率，mom D为随机梯度下降动量，L2Reg为L2正则化强度，$E(\boldsymbol{\theta})$为损失函数。

本文选取4个网络结构参数NC、$\dim W$、$\dim P$、rDrop和3个网络训练参数rLearn、mom D、L2Reg作为卷积神经网络的设计变量。

2.2　贝叶斯优化卷积神经网络

贝叶斯优化是一种全局优化算法，通过设计恰当的概率代理模型和采集函数，贝叶斯优化框架只需经过少数次目标函数评估即可获得理想解，非常适用于求解目标函数表达式未知、非凸、多峰和评估代价高昂的复杂优化问题[19−20]。

贝叶斯优化算法以贝叶斯定理为理论基础，该定理表示为

$$p(f|D_{1:t})=\frac{p(D_{1:t}|f)p(f)}{p(D_{1:t})},\quad(14)$$

式(14)中，f为未知的目标函数或参数模型中的参数，$D_{1:t}=\{(x_1,y_1),(x_2,y_2),\cdots,(x_t,y_t)\}$为已评估点集合，$x_t$为决策向量，$y_t=f(x_t)+\varepsilon$为观测值，$\varepsilon$为观测误差，$p(D_{1:t}|f)$为$y$的似然分布，$p(D_{1:t})$为边际化$f$的边际似然分布，$p(f)$为$f$的先验概率，$p(f|D_{1:t})$为$f$的后验概率，后验概率分布是通过已评估点集合对先验进行修正后未知目标函数或参数模型中的参数的置信度。贝叶斯优化算法使用概率代理模型拟合真实的目标函数，根据采集函数选择下一个评估点。常用的概率代理模型包括贝塔−伯努利模型、线性模型、高斯过程、随机森林等，其中高斯过程具有高度的灵活性、可扩展性和可分析性，是贝叶斯优化中应用最广泛的概率代理模型。高斯过程是多元高斯概率分布的范化，由均值函数和半正定的协方差函数构成，即：

$$y=gp(m(x),k(x,x')),\quad(15)$$

式(15)中，$m(x)$为均值函数，$k(x,x')$为协方差函数。采用高斯过程对一系列离散数据对(x_i,y_i)进行函数拟合时，$m(x)$通常设置为0，$k(x,x')$通常采用Matern协方差函数，即：

$$k(x,x')=\sigma_f^2\left[1+\sqrt{5}\frac{r}{\sigma_l}+\frac{5}{3}\left(\frac{r}{\sigma_l}\right)^2\right]\mathrm{e}^{-\sqrt{5}\frac{r}{\sigma_l}},\quad(16)$$

式(16)中，r为x和x'的欧拉距离，σ_f为特征偏差，σ_l为特征长度，σ_f和σ_l会随着高斯过程拟合而自动变化，σ_l的初值为x_i的标准差，σ_f的初值为y_i的标准差除以$\sqrt{2}$。常用的采集函数生成策略包括基于提升概率的策略、基于提升概率和提升量的策略、置信边界策略、基于信息的策略等，基于提升概率和提升量的策略构造的采集函数如下：

$$\alpha_t(x;D_{1:t})=\begin{cases}(v^*-\mu_t(x))\phi\left(\dfrac{v^*-\mu_t(x)}{\sigma_t(x)}\right)\\ \quad+\sigma_t(x)\phi\left(\dfrac{v^*-\mu_t(x)}{\sigma_t(x)}\right), & \sigma_t(x)>0,\\ 0, & \sigma_t(x)=0,\end{cases}\quad(17)$$

式(17)中，$\alpha_t(x;D_{1:t})$为采集函数，v^*为当前最优函数值，$\phi(x)$为标准正态分布累积密度函数，$\mu_t(x)$和$\sigma_t(x)$分别为均值和标准差。

本文基于贝叶斯优化算法对卷积神经网络模型的7个设计变量进行优化，概率代理模型选用高斯过程模型，采集函数通过基于提升概率和提升量的策略构造，优化变量的约束条件如表1所示。

表1　优化变量的约束条件

Table 1　Constraints on optimizing variables

优化变量	变量标识	变量数据类型	取值范围
卷积核数量	NC	整数	4,8,12,16
卷积核维度	$\dim W$	整数	2,3,4
池化窗口维度	$\dim P$	整数	2,3,4
丢弃层丢弃率	rDrop	浮点数	[0.1,0.3]
学习率	rLearn	浮点数	[0.001,0.1]
随机梯度下降动量	mom D	浮点数	[0.8,0.95]
L2正则化强度	L2Reg	浮点数	$[10^{-10},10^{-2}]$

3　公共场所异常声识别结果与比较

本文所用的异常声来源于Freesound网站，包括火苗噼啪声、玻璃破碎声、婴儿啼哭声、烟花燃放声和警报声，声音文件的样本数为1000个，其中每类声音文件的样本数均为200，声音文件长度均为5 s，采样频率均为44.1 kHz。为了比较不同特征提取和分类器方案的识别效果，选取Mel倒谱系数(MFCC)、Mel倒谱系数+Gammatone倒谱系数(MFCC+GTCC)作为特征提取的比较对象，选取

高斯混合模型(GMM)、支持向量机(SVM)作为分类器的比较对象，高斯混合模型和支持向量机也都采用贝叶斯优化进行模型参数优化，高斯混合模型的优化参数为阶数，支持向量机的优化参数为惩罚系数和径向基核尺度，优化过程最长时间均为20 h。训练集、验证集和预测集的分割比例为6 ∶ 2 ∶ 2，即训练集、验证集和预测集中的样本数分别为600、200和200。每类异常声的200个样本中，120个样本用于训练分类器，40个样本用于贝叶斯优化分类器参数，40个样本用于预测。

3.1 异常声识别结果评价指标

每类异常声识别结果的评价指标包括准确率、召回率、F值，其计算表达式为

$$P_i=\frac{\mathrm{TP}_i}{\mathrm{TPFP}_i},\ R_i=\frac{\mathrm{TP}_i}{\mathrm{TPFN}_i},\ F_i=2\frac{P_iR_i}{P_{i+}R_i},\quad(18)$$

式(18)中，P_i、R_i和F_i分别为第i种异常声识别的准确率、召回率和F值，TP_i为预测集中第i种异常声被正确识别出的数量，TPFP_i为预测集中被预测为第i种异常声的数量，TPFN_i为预测集中第i种异常声的数量。

本文的异常声识别问题为多分类问题，以异常声识别的准确率、召回率和F值分别求均值所得到的平均准确率、平均召回率和平均F值作为异常声识别结果的综合评价指标。

3.2 异常声识别结果

不同特征提取和分类器方案的异常声识别结果如图1~3所示，本文方法对不同类别异常声的识别结果如图4所示。本文方法对5种异常声识别的平均准确率、平均召回率和平均F值均为最高，分别为91.3%、91.5%和91.0%，其识别效果优于其他特征提取和分类器方案。主要原因在于本文方法整合了声信号的时域、频域和倒谱域特征，与单独使用Mel 倒谱特征或整合使用Mel和Gammatone两种倒谱特征相比，可以更全面地表征公共场所异常声的特性。此外本文方法的分类器通过递增的卷积核尺度和池化操作设计可以处理公共场所异常声不同尺度的特征，而文中的高斯混合模型和支持向量机分类器在多尺度特征分析方面有所不足。

为了考察本文方法在噪声干扰下的识别效果，在声音文件中加入信噪比分别为−10 dB、−6 dB、0 dB、10 dB的高斯白噪声，本文方法在不同信噪比噪声干扰下的识别结果如图5所示。识别结果随着

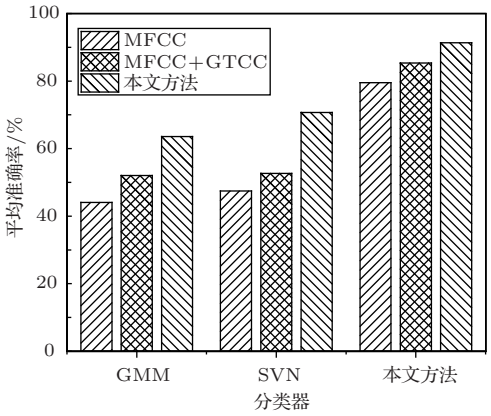

图1 异常声识别平均准确率

Fig. 1 Average precise ratio of abnormal sound recognition

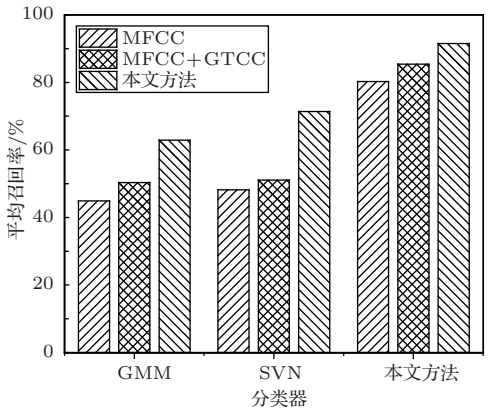

图2 异常声识别平均召回率

Fig. 2 Average recall ratio of abnormal sound recognition

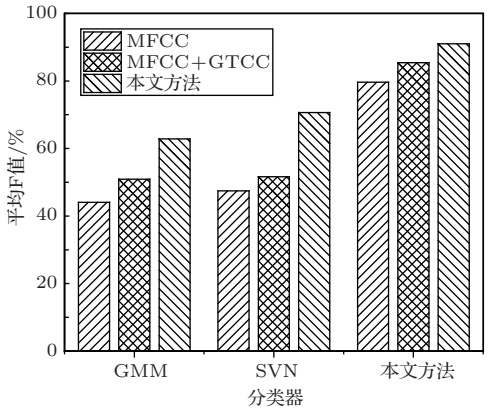

图3 异常声识别平均F值

Fig. 3 Average F score of abnormal sound recognition

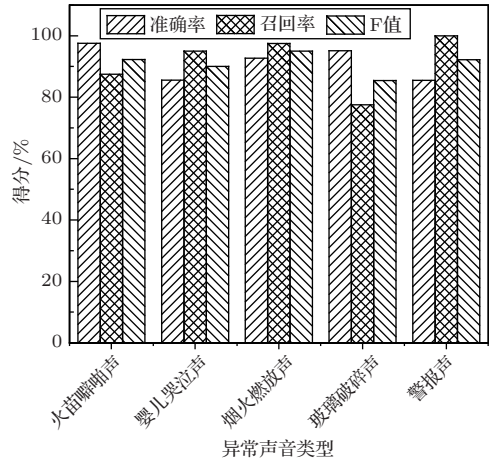

图 4　本文方法的异常声识别结果

Fig. 4　Results of abnormal voice recognition based on my method

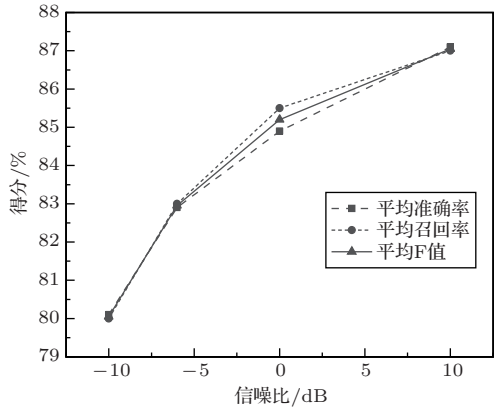

图 5　本文方法在噪声干扰下的异常声识别结果

Fig. 5　Results of abnormal voice recognition under different SNR

信噪比的增大而提高，信噪比为 −10 dB 时平均准确率、平均召回率和平均 F 值分别为 80.1%、80.0% 和 80.0%，本文方法在噪声干扰下识别效果较好。主要原因在于本文方法的特征提取部分用抗干扰性更好、鲁棒性更强的 Gammatone 倒谱代替 Mel 倒谱，而批量归一化层和丢弃层的构建也增强了分类器的泛化能力。因此本文方法可以有效地对火苗噼啪声、玻璃破碎声、婴儿啼哭声、烟花燃放声和警报声 5 种异常声进行识别。

4　结论

本文针对公共场所异常声的感知和识别问题，提出一种基于贝叶斯优化卷积神经网络的识别方法。提取声信号的 Gammatone 倒谱系数、倍频程功率谱、短时能量和谱质心，组合成声信号的特征图。设计了包含多个卷积层和池化层的卷积神经网络作为分类器，采用递增的卷积核设置和池化操作处理不同尺度的特征，并增加批量归一化层和丢弃层以避免过拟合。采用高斯过程模型和基于提升概率和提升量的策略构建概率代理模型和采集函数，基于贝叶斯优化算法对卷积神经网络模型的设计变量进行优化，对包括火苗噼啪声、婴儿啼哭声、烟花燃放声、玻璃破碎声和警报声的 5 种公共场所异常声进行识别。该方法的识别结果与基于 MFCC 或 MFCC+GTCC 的特征提取、基于 GMM 或 SVM 的分类器得到的识别效果进行比较，结果表明该方法的识别效果优于其他特征提取和分类器方案的识别效果。最后分析了该方法在不同信噪比噪声干扰下的识别结果，验证了该方法的有效性。

参　考　文　献

[1] 袁宏永, 苏国锋, 付明. 城市安全空间构建理论与技术研究 [J]. 中国安全科学学报, 2018, 28(1): 185–190.
Yuan Hongyong, Su Guofeng, Fu Ming. Study and application of architecture method and technology of urban safety space[J]. China Safety Science Journal, 2018, 28(1): 185–190.

[2] 韦娟, 岳凤丽, 仇鹏, 等. 基于 EEMD 的异常声音多类识别算法 [J]. 华中科技大学学报 (自然科学版), 2018, 46(7): 117–121.
Wei Juan, Yue Fengli, Qiu Peng, et al. Abnormal sound multiclass recognition algorithm based on EEMD[J]. Journal of Huazhong University of Science and Technology(Natural Science Edition), 2018, 46(7): 117–121.

[3] 胡涛, 张超, 程炳, 等. 卷积神经网络在异常声音识别中的研究 [J]. 信号处理, 2018, 34(3): 357–367.
Hu Tao, Zhang Chao, Cheng Bing, et al. Research on abnormal audio event detection based on convolutional neural networks[J]. Journal of Signal Processing, 2018, 34(3): 357–367.

[4] 李伟红, 田真真, 龚卫国, 等. 改进的 ESMD 用于公共场所异常声音特征提取 [J]. 仪器仪表学报, 2016, 37(11): 2429–2437.
Li Weihong, Tian Zhenzhen, Gong Weiguo, et al. Developed ESMD for the feature extraction of abnormal sound in public places[J]. Chinese Journal of Scientific Instrument, 2016, 37(11): 2429–2437.

[5] 罗森林, 王坤, 谢尔曼, 等. 融合 GMM 及 SVM 的特定音频事件高精度识别方法 [J]. 北京理工大学学报, 2014, 34(7): 716–722.
Luo Senlin, Wang Kun, Xie Erman, et al. High-precision specific audio event recognition method combining SVM

and GMM[J]. Transactions of Beijing Institute of Technology, 2014, 34(7): 716–722.

[6] 刘鑫锦, 苏国韶, 冯夏庭, 等. 基于声音信号的室内岩爆动态预测方法 [J]. 岩土力学, 2018, 39(10): 3573–3580.
Liu Xinjin, Su Guoshao, Feng Xiating, et al. Dynamic prediction method of laboratory rockburst using sound signals[J]. Rock and Soil Mechanics, 2018, 39(10): 3573–3580.

[7] 苏国韶, 刘鑫锦, 闫召富, 等. 岩爆预警与烈度评价的声音信号分析 [J]. 爆炸与冲击, 2018, 38(4): 716–724.
Su Guoshao, Liu Xinjin, Yan Zhaofu, et al. Sound signal analysis for warning and intensity evaluation of rockburst[J]. Explosion and Shock Waves, 2018, 38(4): 716–724.

[8] 张铁民, 黄俊端. 基于音频特征和模糊神经网络的禽流感病鸡检测 [J]. 农业工程学报, 2019, 35(2): 168–174.
Zhang Tiemin, Huang Junduan. Detection of chicken infected with avian influenza based on audio features and fuzzy neural network[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(2): 168–174.

[9] 韩磊磊, 田建艳, 张苏楠, 等. 基于决策树支持向量机和模糊推理的生猪异常声音识别 [J]. 畜牧与兽医, 2019, 51(3): 38–44.
Han Leilei, Tian Jianyan, Zhang Sunan, et al. Porcine abnormal sounds recognition using decision-tree-based support vector machine and fuzzy inference[J]. Animal Husbandry & Veterinary Medicine, 2019, 51(3): 38–44.

[10] 杨元威, 关永刚, 陈士刚, 等. 基于声音信号的高压断路器机械故障诊断方法 [J]. 中国电机工程学报, 2018, 38(22): 6730–6737.
Yang Yuanwei, Guan Yonggang, Chen Shigang, et al. Mechanical fault diagnosis method of high voltage circuit breaker based on sound signal[J]. Proceedings of the CSEE, 2018, 38(22): 6730–6737.

[11] 王丰华, 王邵菁, 陈颂, 等. 基于改进 MFCC 和 VQ 的变压器声纹识别模型 [J]. 中国电机工程学报, 2017, 37(5): 1535–1543.
Wang Fenghua, Wang Shaojing, Chen Song, et al. Voiceprint recognition model of power transformers based on improved MFCC and VQ[J]. Proceedings of the CSEE, 2017, 37(5): 1535–1543.

[12] 周萍, 沈昊, 郑凯鹏. 基于 MFCC 与 GFCC 混合特征参数的说话人识别 [J]. 应用科学学报, 2019, 37(1): 24–32.
Zhou Ping, Shen Hao, Zheng Kaipeng. Speaker recognition based on combination of MFCC and GFCC feature parameters[J]. Journal of Applied Sciences, 2019, 37(1): 24–32.

[13] 程小伟, 王健, 曾庆宁, 等. 噪声环境下稳健的说话人识别特征研究 [J]. 声学技术, 2017, 36(5): 479–483.
Cheng Xiaowei, Wang Jian, Zeng Qingning, et al. A study of robust speaker recognition feature under noisy environment[J]. Technical Acoustics, 2017, 36(5): 479–483.

[14] 林正青, 邱梦然. 水中目标窄带噪声识别的听觉外周模型 [J]. 声学学报, 2016, 41(6): 881–890.
Lin Zhengqing, Qiu Mengran. An auditory periphery model for underwater target narrow-band noise recognition[J]. Acta Acustica, 2016, 41(6): 881–890.

[15] 李靓, 孙存威, 谢凯, 等. 基于深度学习的小样本声纹识别方法 [J]. 计算机工程, 2019, 45(3): 262–267, 272.
Li Jing, Sun Cunwei, Xie Kai, et al. Small sample voiceprint recognition method based on deep learning[J]. Computer Engineering, 2019, 45(3): 262–267, 272.

[16] 张少康, 田德艳. 水下声目标的梅尔倒谱系数智能分类方法 [J]. 应用声学, 2019, 38(2): 267–272.
Zhang Shaokang, Tian Deyan. Intelligent classification method of Mel frequency cepstrum coefficient for underwater acoustic targets[J]. Journal of Applied Acoustics, 2019, 38(2): 267–272.

[17] 陈超, 齐峰. 卷积神经网络的发展及其在计算机视觉领域中的应用综述 [J]. 计算机科学, 2019, 46(3): 63–73.
Chen Chao, Qi Feng. Review on development of convolutional neural network and its application in computer vision[J]. Computer Science, 2019, 46(3): 63–73.

[18] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述 [J]. 计算机学报, 2017, 40(6): 1229–1251.
Zhou Feiyan, Jin Linpeng, Dong Jun. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017, 40(6): 1229–1251.

[19] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述 [J]. 软件学报, 2018, 29(10): 3068–3090.
Cui Jiaxu, Yang Bo. Survey on Bayesian optimization methodology and applications[J]. Journal of Software, 2018, 29(10): 3068–3090.

[20] 任婷玉, 梁中耀, 刘永, 等. 基于贝叶斯优化的三维水动力-水质模型参数估值方法 [J]. 环境科学学报, 2019, 39(6): 2024–2032.
Ren Tingyu, Liang Zhongyao, Liu Yong, et al. The parameters estimation method based on Bayesian optimization for complex water quality models[J]. Acta Scientiae Circumstantiae, 2019, 39(6): 2024–2032.