DOI:10.19651/j.cnki.emt.2212417

基于CNN-RNN集成的隧道事故异常声音识别

郎巨林 郑 晟
(太原理工大学数字化融合监控实验室 太原 034000)

摘 要: 为提高公路隧道事故异常声音识别的准确率,并针对卷积神经网络只关注局部信息问题,提出了一种基于CNN-RNN集成的声音识别模型。该模型采用Stacking集成策略将CNN的强特征表达能力和RNN的强记忆能力相结合,并使用门控循环单元减少循环神经网络的计算复杂度,将SIREN正弦周期函数作为RNN的隐式激活函数,增强模型对声音数据的拟合能力,设计多通道卷积细化特征提取的精度,实现全局化特征提取。在异常声音数据集上评估了所提声音识别模型的识别性能,实验结果表明:提出的声音模型的识别性能高于其他模型,且更加稳健,可有效识别公路隧道事故的异常声音。

关键词: 集成学习;Stacking;CNN;RNN;声音识别
中图分类号: TP391.9 **文献标识码:** D **国家标准学科分类代码:** 510.99

Tunnel accident abnormal sound recognition based on CNN-RNN integration

Lang Julin Zheng Sheng
(Digital Fusion Monitoring Laboratory, Taiyuan University of Technology, Taiyuan 034000, China)

Abstract: In order to improve the accuracy of abnormal sound recognition of highway tunnel accident and to solve the problem that convolutional neural networks only pay attention to local information. An integrated voice recognition model based on CNN-RNN is proposed. The model used the Stacking integration strategy to combine the strong feature expression ability of CNN and the strong memory ability of RNN. The gated cyclic memory unit was used to reduce the computational complexity of RNN. SIREN sinusoidal periodic function was used as the implicit activation function of RNN to enhance the fitting ability of the model to sound data. The precision of multi-channel convolution refinement feature extraction was designed to achieve global feature extraction. The performance of the proposed sound recognition model was evaluated on the abnormal sound data set. Experimental results show that the proposed sound model has higher recognition performance than other models and is more robust, which can effectively identify the abnormal sound of highway tunnel accidents.

Keywords: integrated learning;Stacking;CNN;RNN;sound recognition

0 引 言

随着社会经济和生产的发展,截至2020年底,我国公路隧道规模和数量居于世界之首[1],我国公路隧道内的事故发生率高于非隧道路段[2]。事故发生后,若不能及时处理还会造成更为严重的二次伤害和损失,如后续车辆的追尾,严重时甚至危及生命等。故公路隧道交通事故的感知,报警与定位是减少人员伤亡和财产损失的重要举措。目前我国对于公路隧道的监控手段主要为视频监控,但视频监控存在"死角"问题,且易受光线影响。当事故发生时,视频监控可能不会及时检测到异常事件。为更好地实现对公路隧道的监控,引入了异常声音识别技术。通过识别异常声音的种类来判断隧道事故类型和事故发生的严重程度,从而开展针对性救援工作,既减小了损失也减少了不必要的资源浪费。

近年来,随着城市智慧化建设的需要和工厂自动化的升级需求,声音检测的应用越来越广泛且声音检测技术也逐渐完善,同时也涌现出许多重要的研究成果。如Zulfiqar等[3]通过附加人工噪声结合不同深度卷积神经网络(convolutional neural networks, CNN)进行频谱分析,实现7种异常呼吸音的分类,其模型的识别结果优于传统方法。薛珊等[4]提出了一种基于卷积神经网络的声音识别模型用

收稿日期:2022-12-17

于识别无人机,并通过与支持向量机的对比实验证明了其模型识别无人机的可行性且识别性能优于支持向量机。李传坤等[5]采用密集卷积神经网络架构通过频谱位移改善频谱特征图的质量,且模型在公开数据集上表现良好。曾金芳等[6]提出了多级残差网络模型用于环境声音分类。高晓利等[7]将卷积神经网络与深度循环神经网络(recurrent neural network, RNN)结合用于汽车发动机声纹个体识别,实验表明模型识别效果良好。以上研究都是利用卷积神经网络在图像识别的优势,通过识别声音的频谱特征实现对声音信号的分类,从而将声音识别问题转换为图像识别问题。但声学数据属于时序序列,而卷积神经网络主要关注特征图的局部信息,忽略了样本数据内部帧与帧之间的潜在特征联系。当噪声干扰较大时,会出现局部特征误差现象,极易影响模型的识别性能。RNN 模型对具有时序特征的数据有较强的"记忆能力",被广泛用于序列数据的特征识别,因此为了更好地挖掘声音数据中的时序特征,本文引入了 RNN 模型并与 CNN 模型进行集成。

目前 Stacking 集成学习算法因其可以将多个分类模型提取的特征进行集成,提高模型的总体识别性能而被广泛应用于分类识别领域。Waqas 等[8]使用 Stacking 集成方法对风电机组进行故障检测和分类,实验表明该集成方法可有效识别风力发电机的故障。Cheng 等[9]将多尺度卷积神经网络和隐马尔科夫模型作为基学习器,通过集成学习实现对遥感场景图像的分类,结果表明基于 Stacking 集成学习模型分类性能比其他先进方法表现更加优异。以上研究证明了 Stacking 集成学习在分类识别问题中比单一模型的表现更佳。

由于公路隧道事故声音在时频域上具有独特的特征,因此事故声音识别实质上是对事故声音时频域特征的学习,因此针对强噪声干扰下产生的局部特征误差问题,本文提出了一种基于 RNN-CNN 集成学习的声音识别模型。该模型采取 Stacking 集成策略将局部特征提取能力较强的 CNN 和记忆能力较强的 RNN 并行连接,实现了声音样本的多尺度时空特征提取,并将声学数据的上下文特征信息联系起来,从而避免因局部特征误差降低模型识别的准确率。公路隧道发生的事故有汽车追尾,撞壁,紧急刹车,严重时会发生爆炸,同时还有人的呼救声和因通风机故障造成隧道内氧气稀薄等。根据以上事故,将车辆碰撞声,紧急刹车声,人的呼救声,异常鸣笛声,通风机故障声和爆炸声 6 类异常声音作为识别目标。

1　改进的 Stacking 集成模型

1.1　Stacking 集成学习策略的基本原理

集成学习的基本思想是"取长补短",将不同分类模型的优势相结合,得到比单一模型更好的分类性能。Stacking 是集成学习的集成策略之一,模型结构主要分为两层。图 1 为模型的结构示意图。第一层通过不同的基学习器提取样本数据的不同特征,得到不同的预测结果并进行拼接,然后将拼接后的预测结果作为输入训练第二层元学习器模型。

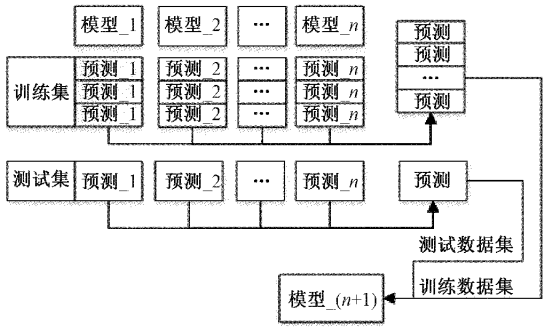


图 1　Stacking 模型结构示意图

1.2　基于 Stacking 集成的声音识别模型

为发挥 CNN 模型在图像识别方面的优势和 RNN 模型对序列数据记忆的优势,将 CNN 和 RNN 作为基学习器提取声学数据的局部特征和上下文信息,将全连接神经网络作为元学习器学习基学习器的训练结果。本文模型的流程如图 2 所示。

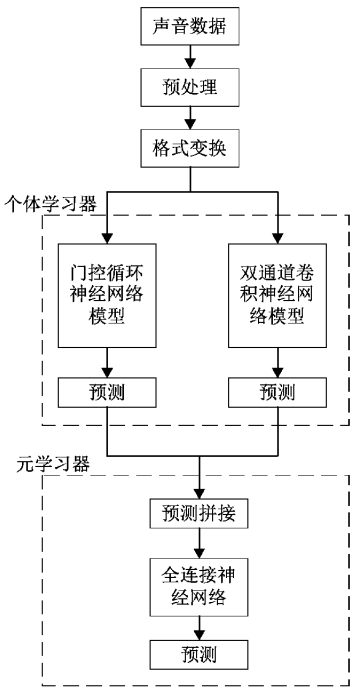


图 2　Stacking 模型流程图

1.3　RNN 模型设计

门控循环单元(gate recurrent unit,GRU)是一种改进的循环神经网络,不仅解决了 RNN 在长期记忆中易出现的梯度问题,而且计算量小,因此被广泛应用于声音识别领域[10-11]。GRU 由复位门和更新门组构成。复位门用于确定如何将新的输入信息与内存中的信息相结合;更新门用于保存到当前时间步的有用信息。Shewalkar 等[12]评估了 RNN、

LSTM 和 GRU 在缩减的 TED-LIUM 数据集上的性能，结果表明 LSTM 表现性能最好，GRU 优化更快且性能接近 LSTM。因此为减少本文模型的计算量，本文的循环网络采用 GRU 作为隐藏层单元。考虑到本文模型的计算代价，RNN 模型采用 GRU 作为隐藏层的基本单元。

将声音数据的序列信息以向量形式作为门控循环神经网络模型的输入并提取声音样本的时序特征。图 3 为门控循环神经网络模型的流程图，模型采用两层门控循环单元获取声学数据的时序特征，然后通过两层全连接层，将提取的时序特征展开，最后使用 Softmax 分类器得到声音的识别结果。

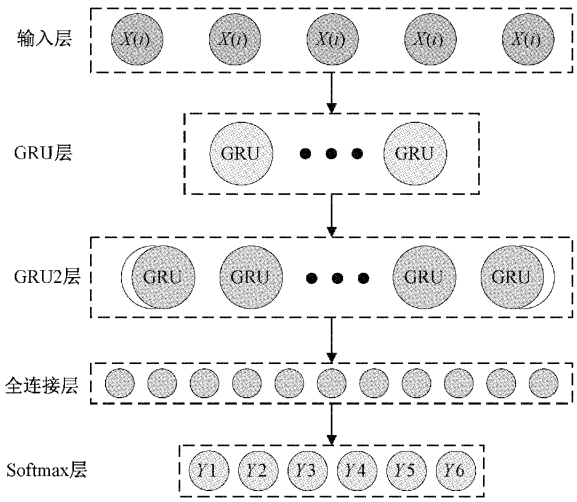


图 3　门控循环神经网络结构

正弦周期激活函数是由 Sitzmann[13] 等提出的。他们通过实验证明了正弦表示网络非常适合表示复杂的自然信号及其衍生物，不仅可以对信号进行细节建模，还能表示信号的时空导数。其表达式如下：

$$y = \sin(\boldsymbol{W}_i^{\mathrm{T}} f(x_i) + \boldsymbol{b}_i) \tag{1}$$

式中：y，$f(x_i)$，$\boldsymbol{W}_i^{\mathrm{T}}$ 和 $\boldsymbol{b}_i$ 分别为模型第 i 层的输出，输入，权重参数矩阵和偏置矩阵。GRU 采用正弦周期函数作为隐式神经网络的激活函数，增强循环神经网络对时间细节的建模能力。本文为了获取声音信号的细节特征，在 GRU 中引入 SIREN 架构，将正弦周期激活函数作为 GRU 单元的隐藏激活函数，改进的 GRU-SIREN 结构如图 4 所示。本文 RNN 模型中采用两层 GRU-SIREN 层，单元个数分别为 64 和 32。

1.4　CNN 模型设计

卷积神经网络由输入层，卷积层，池化层，全连接层和输出层构成。为获取声音样本的多尺度特征，本文采用具有多通道输入的卷积神经网络模型。提取原始声音数据的多尺度特征后，拼接成新的特征向量，将音频数据的梅尔频谱图作为模型的一个特征输入；将短时能量，谱质心和过零率等低层特征拼接为新的特征向量作为模型的另一个特征输入。通过设计不同内核大小的卷积层和池化层提取样本

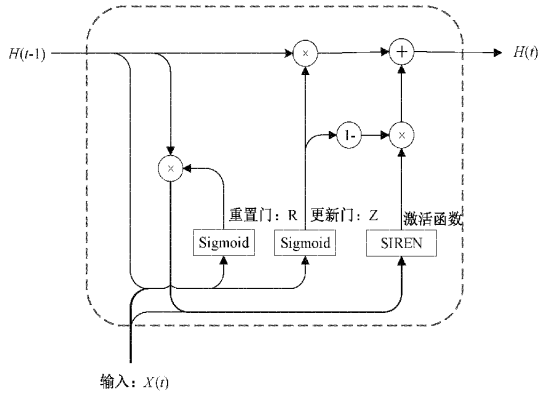


图 4　门控循环记忆单元结构

的深层特征，最后使用 softmax 分类器得到模型的分类结果。其结构示意图如图 5 所示。CNN 模型的频谱图输入通道使用大小为{3×3}的卷积核，由于拼接而成的多维特征之间的联系较小，因此多维特征输入通道使用大小为{3×1}的卷积核，使模型可准确提取音频的相关特征。

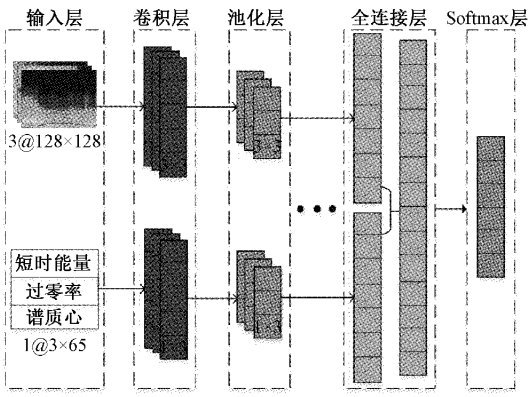


图 5　卷积神经网络结构

1.5　声音特征提取

1)梅尔频谱图

梅尔频谱图模拟人耳的听觉感知将人类不敏感的高频信号转换为对数频率，常作为特征输入用于声音识别模型的训练[14-16]，将声音分类问题直接转换成图像分类问题。其频率转换如式(1)所示。

$$m = 2\,595\lg\left(1+\frac{f}{700}\right) \tag{2}$$

其中，f 为信号的原始频率，m 为转换后的梅尔对数频率。

2)短时能量

短时能量实质上就是声音信号一帧内的幅值的平方，其体现了信号不同时刻的强弱程度。第 n 帧信号 $x_n(m)$ 的短时能量 E_n 为：

$$E_n = \sum_{m=0}^{N-1} x_n^{\,2}(m) \tag{3}$$

3)过零率

过零率为声音信号在每一帧中通过零点的次数，在一

定程度上可以反映出信号的频率信息,是较为基础的时域特征。其计算公式是:

$$Z_n = \frac{1}{2}\sum_{m=0}^{N-1} |\,\mathrm{sgn}[x_n(m)] - \mathrm{sgn}[x_n(m-1)]\,| \quad (4)$$

式(3)中的 sgn[]是符号函数,即:

$$\mathrm{sgn}[x] = \begin{cases} 1, & (x \geqslant 0) \\ -1, & (x \geqslant 0) \end{cases} \quad (5)$$

4)谱质心

谱质心是音频信号的频率分布和能量分布的重要信息,反映了声音信号中主谐波的基频值的特性。谱质心的计算如式(6)所示。

$$SC = \sum_{n=1}^{N} f(n) \cdot P(E(n)) \quad (6)$$

式中:$f(n)$ 为声音信号的频率,$E(n)$ 为信号在傅里叶变换后对应频率的谱能量,$P(n)$ 的计算公式如式(7)所示。

$$P(E(n)) = \frac{\sum_{n=1}^{N} E(n)}{\sum_{1}^{N} E(n)} \quad (7)$$

1.6 评价指标

本文研究的隧道事故异常声音识别属于分类问题,为客观评价模型分类的性能指标,本文选取准确率,F1 值和混淆矩阵作为评价指标。准确率较直观的反映了模型分类能力的强弱。F1 值是在精确率和召回率的基础上而提出的,可较全面的评价不同模型的优劣,F1 值越大,则模型的性能效果越好。混淆矩阵描述了模型分类的详细信息,可以更好地分析模型的分类性能。准确率和 F1 值的计算公式分别为:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$F1 = \frac{P \times R \times 2}{P + R} \quad (9)$$

式中:TP 为真阳性事件,TN 为真阴性事件,FP 为假阳性事件,FN 为假阴性事件。式(9)中 P 和 R 分别表示模型的精确率和召回率。准确率,$F1$ 值和混淆矩阵常作为分类模型的评价指标[17-19],用于衡量模型的分类性能。

2 实验与结果分析

2.1 数据搜集与处理

本文模型识别的声音信号有车辆碰撞声,紧急刹车声,异常鸣笛声,爆炸声,人为喊叫声和通风机故障声。由于隧道事故异常声音信号的数据较少,故实验从 findsounds.com,aigei.com,ear0.com 等网站上爬取异常声音信号,并通过添加合适的噪声得到模拟的隧道异常声音信号。各类声音信号的数量详情如图 6 所示。图中紧急刹车声的数据相对较少,为防止由于数据量不平衡导致模型过拟合现象的出现,对刹车声信号做数据增强处理。在公路隧道中,刹车声往往会伴随有车辆碰撞声,故将刹车声 x_1 与裁剪后的部分碰撞声 x_2 相叠加,得到新的刹车声音信号 X,使刹车声的数据量与其他声信号数量达到均衡。叠加的公式如下:

$$X = \alpha x_1 + (1-\alpha)x_2 \quad (10)$$

式中:α 为叠加因子,取值范围为(0,1)。

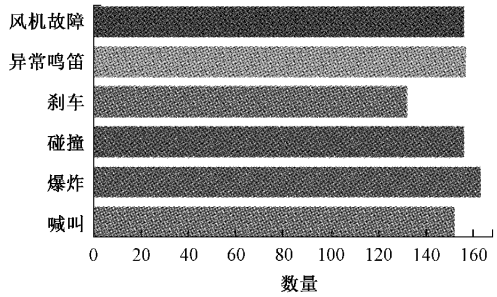


图 6 各类声音信号数量

经过对刹车声信号的处理后,异常鸣笛声 158 条,爆炸声 164 条,刹车声 156 条,车辆碰撞声 156 条,喊叫声 153 条,通风机故障声 156 条,各类声音信号的数量都处于(150,165)区间内。所有声音数据都为 wav 格式,并通过填零扩充方式使声音信号都统一为最长的音频长度。

2.2 实验设计与结果

1) 数据预处理

本文将增强后的数据集进行二次预处理。首先将数据集按照 9∶1 的比例划分为训练集和测试集。然后在 Python 环境中使用 libroas 库提取声音数据的时序信息用于循环神经网络的输入,提取梅尔频谱图,短时能量,过零率和谱质心,尺寸分别为{128×128},{1×65},{1×65},{1×65}。将短时能量,过零率和谱质心拼接成尺寸为{3×65}的新特征向量图。则卷积神经网络模型的双通道输入分别为梅尔频谱图和新的特征向量图。为防止模型过拟合现象的发生,在训练个体学习器时,将训练集做 3 折交叉验证。

2) 实验设计、结果及分析

实验在 Pycharm 环境中使用以 Tensorflow 为后端的深度学习库 Keras 搭建并训练模型。

为验证本文模型对隧道事故异常声音的识别性能,实验将本文模型分别与 CNN,RNN 和文献[7]中的方法在本文制作的声音数据集上进行识别性能对比,并使用 F1 值和模型准确率作为评价准则。其中 CNN 和 RNN 模型分别为改进的基学习模型。由于公路隧道噪声多且复杂,故实验对声音数据集添加不同程度的低信噪比噪声用于模拟隧道环境,如表 1 所示,不同模型在不同低信噪比环境下的公路隧道异常声音识别的 $F1$ 值和准确率。图 7 为本文模型在不同信噪比下的混淆矩阵。

为验证卷积神经网络双通道输入的有效性和循环神经网络隐式正弦激活函数在时域信号应用中的实用性,实验分别采用常见的 SVM,KNN,RF 分类模型和本文改进的

表 1　不同模型的识别性能

$F1$ 值\ACC	CNN	RNN	文献[7]	本文模型
−10 dB	0.67\0.732	0.62\0.644	0.84\0.827	0.89\0.896
−5 dB	0.74\0.814	0.65\0.692	0.87\0.901	0.91\0.913
0 dB	0.78\0.879	0.71\0.793	0.90\0.910	0.91\0.922
5 dB	0.85\0.926	0.80\0.887	0.92\0.933	0.93\0.952
10 dB	0.90\0.937	0.89\0.917	0.93\0.942	0.95\0.965

RNN 和 CNN 作为基学习器与本文模型进行分析比较。　表 2 为不同集成学习模型的识别准确率和 $F1$ 值。

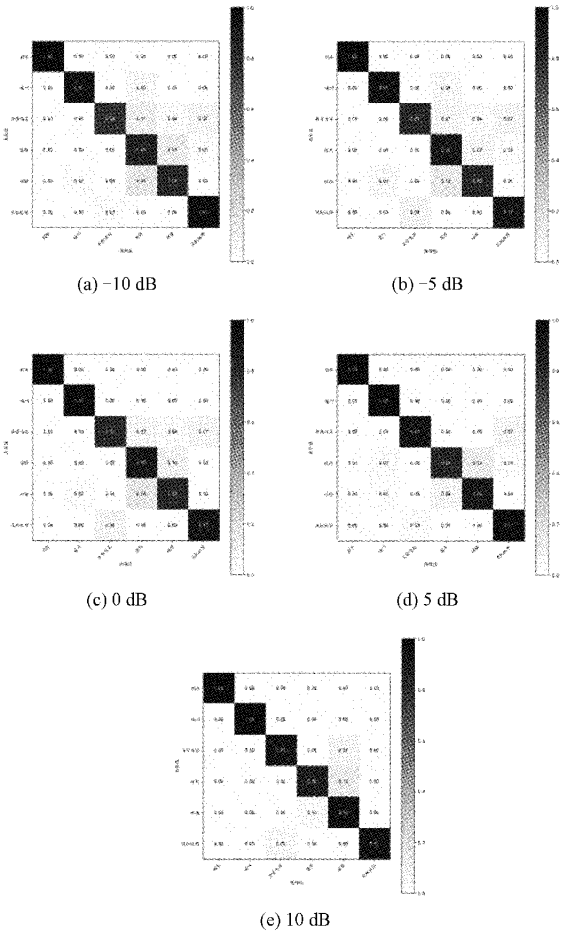
表 2　不同集成模型的识别性能

$F1$ 值\ACC	SVM+RF	KNN+SVM	KNN+RF	本文模型
−10 dB	0.85\0.854	0.81\0.815	0.88\0.837	0.89\0.896
−5 dB	0.86\0.863	0.82\0.824	0.85\0.852	0.91\0.913
0 dB	0.88\0.885	0.82\0.836	0.87\0.886	0.91\0.922
5 dB	0.90\0.936	0.89\0.904	0.90\0.923	0.93\0.952
10 dB	0.93\0.941	0.90\0.912	0.91\0.935	0.95\0.965

由表 1 中数据可得，在不同低信噪比环境下，根据各个模型 $F1$ 值和准确率的大小得出：本文模型对隧道异常声音的识别性能均高于 CNN，RNN 和文献[7]中的方法。在极低信噪比环境下，本文模型仍能保持接近 90% 的识别准确率，且 $F1$ 值达到 0.89，表明模型的精确率和召回率都相对较高，模型的异常声音识别性能良好。在噪声环境相对较好的情况下，模型的识别准确率达到了 96% 以上，$F1$ 值为 0.95，这表明噪声的干扰可直接影响模型的识别性能。而 RNN，CNN，文献[7]和本文模型在 −10 dB 和 10 dB 环境下的识别准确率的差值分别为 0.273，0.205，0.115，0.069，本文模型的抗噪声干扰能力高于其他模型。

通过比较表 2 中不同集成模型的识别准确率和 $F1$ 值，表明本文模型的基学习器比其他分类模型的识别性能更好。且通过计算 −10 dB 和 10 dB 环境下集成模型的识别准确率差值，并与表 3 的数据相比较，也从侧面验证了集成模型对噪声变化具有一定鲁棒性。

由图 7(a)可得，在 −10 dB 下，对紧急刹车声，通风机故障声和人为呼救声的识别准确率都达到了 95% 以上，表明本以上 3 种声音的可识别性较高；对于爆炸声和车辆碰撞声的识别准确率分别为 85% 和 82%，且 12% 的爆炸声被预测为车辆碰撞声，16% 的车辆碰撞声被预测为爆炸声，表明这两种声音的相似度较高。异常鸣笛声的识别准确率最低为 78%，可能是由于异常鸣笛声易受噪声干扰。通过综合比较不同信噪比下模型对不同声音的识别准确率，在不同信噪比下，模型对紧急刹车声，人为喊叫声和通风机故障声的识别较准确，但对于异常鸣笛声，爆炸声和碰撞声的识别准确率有所不同，故这三种声音受噪声影响较大。



(a) −10 dB　(b) −5 dB　(c) 0 dB　(d) 5 dB　(e) 10 dB

图 7　不同信噪比下模型的混淆矩阵

3 结 论

本文采取 Stacking 集成策略将 CNN 模型和 RNN 模型作为基学习器对异常声音数据集进行训练并合并预测结果,然后使用全连接神经网络作为元学习器对基学习器的预测结果进行二次预测。通过对比实验得出:在低信噪比环境下,本文模型对隧道异常声音的识别准确率较高,且模型的识别性能强于现有的声音识别方法,可有效地识别隧道事故异常声音事件。

虽然本文模型简化了部分算法,但还存在模型训练参数多,训练时间长的问题,因此下一步研究会侧重减少模型的训练参数和训练时间。

参考文献

[1] 交通运输部. 2020 年交通运输行业发展统计公报[R/OL]. [2021-05-19]. https://xxgk.mot.gov.cn/2020/jigou/zhghs/202105/t20210517_3593412.html. Ministry of Transport, PRC. Statistics report of transportation in China 2020.

[2] 苏保锋,胡江碧. 高速公路隧道事故风险规避新技术应用分析和探讨[J]. 隧道建设(中英文),2022,42(3):363-371.

[3] ZULFIQAR R, MAJEED F, IRFAN R, et al. Abnormal respiratory sounds classification using deep CNN through artificial noise addition[J]. Frontiers in Medicine, 2021, 8: 714811.

[4] 薛珊,李广青,吕琼莹,等. 基于卷积神经网络的反无人机系统声音识别方法[J]. 工程科学学报,2020,42(11):1516-1524.

[5] 李传坤,郭锦铭,李剑,等. 基于频谱位移模块的环境声音识别方法[J]. 电子测量技术,2022,45(5):62-67.

[6] 曾金芳,李友明,杨恢先,等. 基于多级残差网络的环境声音分类方法[J]. 数据采集与处理,2021,36(5):960-968.

[7] 高晓利,李捷,王维,等. 基于 CRNN 的汽车发动机声纹个体识别方法[J]. 火力与指挥控制,2021,46(3):150-153, 159.

[8] WAQAS K P, BYUN Y C. Multi-fault detection and classification of wind turbines using stacking classifier[J]. Sensors, 2022, 22(18):6955.

[9] CHENG X, LEI H. Remote sensing scene image classification based on mmsCNN-HMM with stacking ensemble model[J]. Remote Sensing, 2022, 14(17):4423.

[10] 朱凌建,陈剑虹,王裕鑫,等. 基于 GRU 神经网络的脉搏波波形预测方法研究[J]. 电子测量与仪器学报,2022,36(5):242-248.

[11] 阳雨妍,宋爱国,沈书馨,等. 基于 CNN-GRU 的遥操作机器人操作者识别与自适应速度控制方法[J]. 仪器仪表学报,2021,42(3):123-131.

[12] SHEWALKAR A, NYAVANANDI D, LUDWIG S A. Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU[J]. Journal of Artificial Intelligence and Soft Computing Research, 2019, 9(4): 235-245.

[13] SITZMANN V, MARTEL J, BERGMAN A W, et al. 2020. Implicit neural representations with periodic activation functions[J]. In Proceedings of the 34th International Conference on Neural Information Processing Systems(NIPS'20),2020, 626: 462-7473.

[14] 刘杰,朱正伟. 基于稀疏轻量卷积神经网络的管道泄漏检测[J]. 电子测量技术,2022,45(19):131-135.

[15] 安鑫,代子彪,李阳,等. 基于 BERT 的端到端语音合成方法[J]. 计算机科学,2022,49(4):221-226.

[16] 杨智伦,朱铮涛,陈树雄,等. 改进 CNN 的供水管道泄漏声音识别[J]. 国外电子测量技术,2023,42(1):153-158.

[17] SRIVASTAVA R, KUMAR P. GSO-CNN-based model for the identification and classification of thyroid nodule in medical USG images[J]. Network Modeling Analysis in Health Informatics and Bioinformatics, 2022, 11(1):1-14.

[18] AHMED H M, ALQAHTANI H, ELKAMCHOUCHI D H, et al. Hyperparameter tuned deep autoencoder model for road classification model in intelligent transportation systems[J]. Applied Sciences, 2022, 12(20): 10605.

[19] 许毓婷,孙浩然,高勋,等. 基于 LIBS 技术结合 PCA-SVM 机器学习对猪肉部位的识别研究[J]. 光谱学与光谱分析,2021,41(11):3572-3576.

作者简介

郎巨林,硕士研究生,主要研究方向为智能系统监控。

E-mail:con_langjl@163.com

郑晟(通信作者),硕士,副教授,主要研究方向为智能控制技术和装置。

E-mail:3145212204@qq.com