DOI：10.13718/j.cnki.xdzk.2023.12.015
徐科，姚凌云，姚静怡，等. 基于改进 VGG-16 网络的交通声音事件分类方法研究 [J]. 西南大学学报(自然科学版)，2023，45(12)：145-156.

基于改进 VGG-16 网络的交通声音事件分类方法研究

徐科，　姚凌云，　姚静怡，　姚敦辉

西南大学 工程技术学院/丘陵山区农业装备重庆市重点实验室，重庆 400715

摘要：交通声音事件分类是提升城市智慧交通系统环境感知能力的关键技术之一. 针对传统交通系统的环境声音感知能力弱、效率低、鲁棒性低、可分类数量少等问题，研究了一种基于 VGG 卷积神经网络的交通声音事件分类方法，该方法使用语谱图(spectrogram image features，SIF)作为交通声学特征，建立并优化了卷积神经网络(convolutional neural networks，CNN)，从而实现交通声音的智能分类. 首先，使用实验室采集的 10 种交通声音，构建了交通声音数据集. 其次，利用语谱图方法对交通声音进行声学特征提取，搭建 VGG-16 分类算法主模型，通过双卷积层融合算法和块间直连通道对网络进行改进，得到了 VGG-TSEC 网络. 该优化网络的交通声音事件分类准确率可达 97.18%，与优化前相比准确率提升 4.68%，其权重参数降低 72.76%，占用空间降低 384MB. 同时，将该优化模型与 K 邻近(KNN)、支持向量机(SVM)等机器学习方法进行对比，其准确率分别提高了 19.68%和 4.41%. 结果表明，VGG-TSEC 交通声音分类方法可以实现警笛音、事故碰撞、行人尖叫、卡车等交通声音的高效分类，为交通声音事件分类提供参考.

关　键　词：交通声音事件分类；卷积神经网络；交通声音；语谱图特征；深度学习

中图分类号：U495　　**文献标志码**：A

文 章 编 号：1673-9868(2023)12-0145-12

开放科学(资源服务)标识码(OSID)：

Research on Traffic Sound Event Classification Method Based on Improved VGG-16 Network

XU Ke，　YAO Lingyun，　YAO Jingyi，　YAO Dunhui

College of Engineering and Technology，Southwest University/Chongqing Key Laboratory of Agriculture Equipment in Hilly Areas，Chongqing 400715，China

Abstract：Traffic sound event classification is the most important step to improve the environmental perception ability of transportation system. Aiming at the problems of traditional traffic system，such as weak sound perception，inefficiency，low robustness and few detectable types，a traffic sound event classification method based on VGG was studied. This method used Spectrogram image features (SIF) as traffic

收稿日期：2022-09-06
基金项目：国家自然科学基金项目(52175121).
作者简介：徐科，硕士研究生，主要从事声音事件检测、深度学习、汽车听觉研究.
通信作者：姚凌云，博士，教授，博士研究生导师.

sound features established the Convolutional neural networks (CNN) to complete intelligent classification of traffic sounds. Firstly, a traffic sound dataset was constructed using 10 sounds collected in the laboratory. Then, the SIF method was used to extract the acoustic features of traffic sounds, and the main model of VGG-16 classification algorithm was built. Finally, the VGG-TSEC network is improved by fusion algorithm with two convolution layer and inter-block channel algorithm. The final experiment shows that traffic sound event classification accuracy of the optimized network can reach 97.18%, which is 4.68% higher than that of before optimization. The weight parameter is reduced by 72.76% and the resource consumption is reduced by 384MB. At the same time, the optimization model is compared with machine learning such as K-nearest neighbor (KNN) and support vector machine (SVM), and the final accuracy was improved by 19.68% and 4.41%, respectively. The results show that the VGG-TSEC traffic sound classification method can achieve efficient classification of traffic sounds such as siren sounds, accident collisions, pedestrian screams, and trucks sounds, etc., which can provide a reference for the traffic sound event classification.

Key words: traffic sound event classification; convolutional neural network; traffic sound; spectrogram image feature; deep learning

近年来，随着机器学习理论的发展，大量人工智能(artificial intelligence, AI)项目应运而生，促使交通环境感知系统朝着多传感器融合的智能化方向发展[1-4]. 目前，交通环境信息的感知主要依靠激光雷达、毫米波雷达和视觉传感器等机器视觉技术，几乎没有听觉技术的应用. 然而，听觉能力对城市智慧交通系统十分关键，交通环境中的声音事件(如喇叭声、警笛声、车辆碰撞声、轮胎制动声等)携带着大量声音信息. 研究交通声音事件分类方法，对于完善道路安全和不同背景下的声音检测方法有重要的实际意义[5]和应用价值.

交通环境中的声音事件(sound event)是指一段独立完整且能引起人们感知注意的短时连续声音信号[6-7]. 声音事件检测(sound event detection, SED) 是交通环境感知的核心技术之一，主要包括声音事件分类(sound event classification, SEC)和声音事件定位(sound event location, SEL). 传统的声音事件分类主要借鉴语音识别和模式匹配，将语音识别技术迁移应用到声音事件分类领域. 例如使用基于矢量量化的识别技术、动态时间规整(dynamic time warping, DTW)技术、隐马尔可夫模型(hidden Markov models, HMM)、高斯混合模型(gaussian mixture model, GMM)、支持向量机(support vector machine, SVM)等技术.

目前，交通声音事件分类相关研究以模式识别理论方法为主，即特征提取，模式匹配. Karpis[8]研究了基于声学信号检测特种车辆(例如警车消防车)的方法，实现了警车、消防车的初步检测. Choi等[9]针对音频监控问题，采用GMM分类器在不同背景噪声环境下对9种异常声音(尖叫声、汽笛声、撞击声等)进行识别，并自动更新模型参数达到对环境的自适应，识别效率有所提高. Li等[10]以HMM识别模型为基础，采用环境中的大量声学事件训练HMM模型，并通过将未知声学事件的MFCC特征与背景池对比，提取目标声学事件的声音，该算法在不牺牲识别性能的情况下简化了模型的复杂度. Lefebvre等[11]在2017年使用声学信号并采用支持向量回归方法实现了交通流量测量. 朱强华等[12]以MFCC特征和SVM作为声音特征和分类器对交通声音分类(警车、消防车、救护车、汽笛声等)进行了研究，通过优化MFCC和SVM算法，完成无人车交通声音分类任务，但其所建模型在信噪比减小的情况下，分类准确率大大降低. 2020年，Zhang等[13]提出了一种基于稀疏自动编码器的车辆声音事件分类方法分析交通状况，其检测准确率达到94.9%.

上述研究主要采用梅尔频率倒谱系数(mel-frequency cepstrum coefficient, MFCC)等声学特征和传统机器学习方法(machine learning, ML)作为声音事件的模型分类器. 然而传统机器学习仅适用于小样本，在处理大样本、高维度的数据时准确率会大幅降低[14]. 此外，实际交通环境噪音较大，MFCC声音特征提取对噪声十分敏感，较大程度上影响了机器学习的性能[15]. 而近几年由于人工智能快速发展，基于深度学习的算法在声音识别方面表现出巨大优势，具有学习能力强、覆盖范围广等优点，通过神经网络对声音事件进行特征提取和学习，可以获得更好的分类效果[16].

鉴于此，本文以 SIF 特征提取法作为交通声音的声学特征，将卷积神经网络引入交通声音事件分类研究，在 VGG 卷积神经网络中搭建了双卷积层融合算法以及块间直连通道，提出一种基于改进 VGG-16 卷积神经网络的交通声音事件分类算法. 该算法对麦克风系统采集到的交通声音进行预处理，将快速傅里叶变换得到的时频域谱图作为声音的特征，神经网络则负责交通声音的深层特征进行学习，完成交通环境的声音事件分类任务. 实验结果表明，本文提出的 VGG-TSEC 块间直连算法在交通声音测试集上的准确率为 97.18%，分类性能优于随机森林、K 邻近(KNN)和支持向量机(SVM)等传统机器学习算法.

1　声音特征提取算法

交通声音是一维时域信号，直接输入神经网络会导致信号帧的丢失，进而影响模型精度. 研究表明，语谱图在声音特征标记领域具有较高的噪声鲁棒性优势，然而语谱图特征在兼顾时域和频域信息的同时，容易造成特征泄漏. 因此本文充分考虑语谱图的优势，使用二维语谱图提取交通声音声学特征，增强了特征的噪声鲁棒性. 语谱图特征提取示意图见图 1，该特征是交通声音信号的时频域谱图. 声学特征的提取步骤包括预加重、分帧、加窗、短时傅里叶变换等[17]. 首先使用预加重减小信号在传递过程中高频部分的损失. 对预加重后的目标声音进行分帧，分帧后的片段表现出短时连续性. 汉宁窗是一种窗函数，其定义见式(1)，该窗函数用于分帧后的信号处理，可消除各帧两端信号的不连续性，其中 $M-1$ 是窗函数的周期. 对加窗后的信号进行快速傅里叶变换(fast fourier transform，FFT)，计算方法见式(2)，将傅里叶变换后得到的时频域谱图碎片按像素帧顺序排列，短时碎片连接即可得到长时稳定的二维特征谱图[18]，部分交通声音事件的时域信号和时频谱的特征见图 2.

$$w_{(n)}=\begin{cases}0.5\left[1-\cos\left(\dfrac{2\pi n}{M-1}\right)\right] & 0\leqslant n\leqslant M-1\\ 0 & \text{其他}\end{cases} \tag{1}$$

$$X_{\hat{n}}(e^{j\hat{w}})=\sum_{m=-\infty}^{+\infty} w\left[\hat{n}-m\right]x[m]e^{-j\hat{w}m} \tag{2}$$

式中，$X_{\hat{n}}(e^{j\hat{w}})$表示时间 n 和频率 w 的二维函数；$x[m]$为输入的交通声音信号；$w\left[\hat{n}-m\right]$表示真实的序列信息.

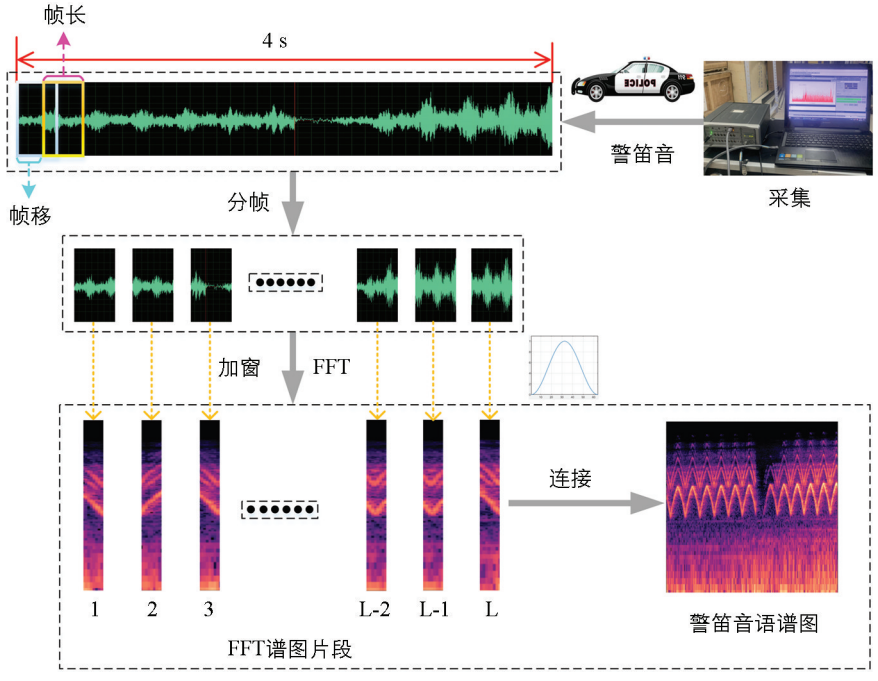


图 1　语谱图特征提取示意图

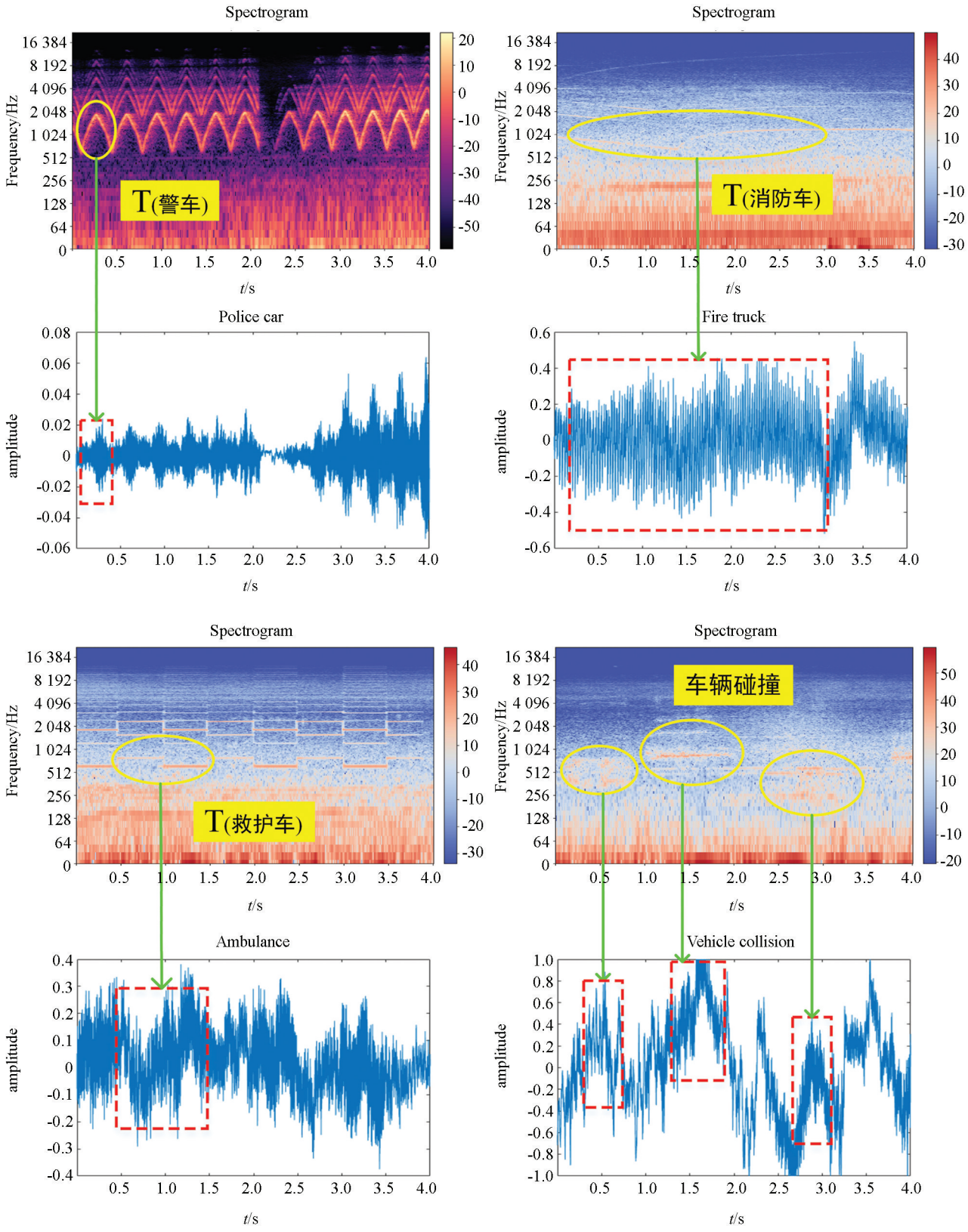


图2 4类交通声音语谱图(警车、消防车、救护车、车辆事故碰撞声)

2 基于改进VGG网络的交通声音分类方法

2.1 VGG网络结构

语谱图特征提取是将交通声音时域信号从一维映射到二维的图像分类问题. 相比于传统的图像分类方

法，深度学习方法可以通过自主学习来提取深层语义特征，加强特征与分类器之间的联系. 目前应用比较成功的卷积神经网络有 AlexNet，GoogleNet，VGG，Inception 和 ResNet 系列，此外还包括其他新兴的轻量级网络，如胶囊网络和 MobileNet 等. 研究表明，CNN 结构太深易引起模型过拟合，发生训练退化；结构太浅则容易导致特征提取不充分，无法表达图像的深层次信息. 试验对比以上经典结构模型，选择具有 16 个权重层的 VGG-16 网络作为本研究的基线结构，VGG-16 网络结构见图 3，包括 13 个卷积层、3 个全连接层、5 个最大池化层，以及 Softmax 输出层.

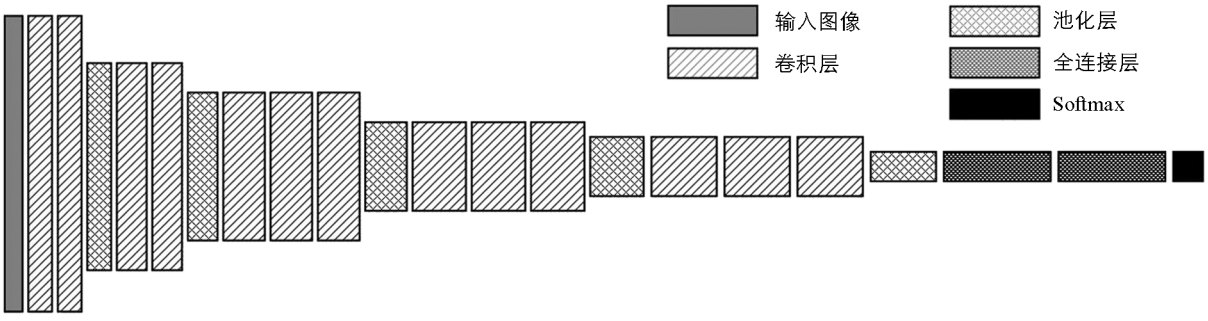


图 3　VGG-16 网络结构

VGG 网络最主要的特点是卷积层采用多个 3×3 卷积核堆叠构成，池化层则采用 2×2 的小卷积核. 在感受野相同的情况下，通过较小参数代价的小卷积核能获得更优的非线性结果. 全连接层主要负责卷积特征的融合，起到分类器的作用. 随着网络层数的加深，需要求解的参数数目也随之增加，其中大部分参数来自全连接层. 全连接层的 $C\times1$ 维向量输入 Softmax 层，该层输出的数值表示该样本所属类别的概率，数值越大，则可信度越高. Softmax 函数见式(3).

$$f(Z_i)=\frac{e^{Z_i}}{\sum_{i=1}^{C}e^{Z_i}} \tag{3}$$

式中，Z_i 为第 i 个节点的输出值；C 为分类数目；$f(Z_i)$为分类类别为 i 的概率；在模型测试中，Softmax 层会选择 $f(Z_1)$，$f(Z_2)$，$f(Z_3)$，…，$f(Z_c)$中概率最大的类别作为样本的预测标签.

2.2　VGG 卷积神经网络参数优化

传统的 VGG-16 网络中 Softmax 层分类的数目达到 1 000 个，3 层全连接层结构为(4 096，4 096，1 000). 第一层全连接层输入参数来自池化层，共 25 088 个神经元，所需参数为 102 764 544 个. VGG 中 3 层全连接神经网络共计 1.236 亿个参数，占总参数量的 89.33%. 由此可知，全连接层冗余的参数浪费了系统资源，且容易发生过拟合现象. 本文采用两层卷积核为 1×1 的卷积层与全局平均池化层融合代替全连接层，降低网络模型权重参数的同时使模型趋于轻量化. 改进后的 VGG-TSEC 参数数目为 0.377 亿个，降低 72.76%. 全连接层改进前后结构对比见表 1.

表 1　VGG-16 全连接层改进前后结构表

指标	改进前 VGG-16				优化后 VGG-16			
网络层	FC1-4096	FC2-4096	FC3-1000	Softmax-1000	C14-64	C15-32	GAP	Softmax-10
参数量	102 764 544	16 781 312	4 097 000	0	16 448	2 080	0	330

卷积层是模型中最为核心的一层，由若干个神经元组成. 假设卷积层第 l 层直接相连的输入张量为 $x^l\in R^{n\times p\times q}$，其中 p，q 分别为矩阵高度和宽度. 第 l 层激活函数输出 a^l 的计算见式(4).

$$a_l=f(W_lX_l+b_l) \tag{4}$$

式中，f 为激活函数；$W_l\in R^{m\times h\times h}$ 为卷积核的滤波器，m 为滤波器的个数，h 为滤波器的尺寸；b_l 为卷积层的偏置. 常见的激活函数有 Tanh，ReLU，Leaky ReLU，Sigmod 等. 其中 ReLU 激活函数收敛较快，然而 ReLU 存在神经元“死亡”问题，权重迭代见式(5)，在使用较大学习率时，w_{ij} 会取到负值，使激活函数输出为 0，因此梯度下降对这些神经元无效. 与之相反，Leaky ReLU 在 $x\leqslant 0$ 时斜率不为 0. 经实验测试 Leaky ReLU 的交通声音分类性能优于 ReLU[19]，因此本研究选取 Leaky ReLU 作为卷积层激活函数.

$$w_{ij}=w_{ij}-\eta\frac{\partial E}{\partial w_{ij}} \tag{5}$$

式中，w_{ij} 为浅层第 i 个神经元与深层第 j 个神经元之间的连接权重；E 为网络激活函数.

此外，神经网络在训练过程中极易出现“训练集优，测试集差”的过拟合情况，致使模型的泛化能力差. 因此引入批量归一化层(batch normalization，BN)，加快模型训练的收敛速度，增强模型泛化能力.

首先在训练阶段，BN 层可以对输入网络的时频谱进行预处理，批处理(mini-batch)输入 x：$B=\{x_1, x_2, \cdots, x_m\}$，计算批处理均值 μ_B，见式(6).

$$\mu_B=\frac{1}{m}\sum_{train=1}^{m}x_{train} \tag{6}$$

批处理数据方差 σ_B^2：

$$\sigma_B^2=\frac{1}{m}\sum_{train=1}^{m}(x_{train}-\mu_B)^2 \tag{7}$$

规范化处理：

$$\hat{x}_{train}=\frac{x_{train}-\mu_B}{\sqrt{\sigma_B^2+\varepsilon}} \tag{8}$$

式中，ε 是一个很小的数，防止分母为零.

尺度变换和偏移：

$$\hat{y}_{train}=\gamma\hat{x}_{train}+\beta=BN_{\gamma,\beta(x_{train})} \tag{9}$$

式中，γ 为尺度因子，β 为平移因子.

测试阶段，BN 层计算输出 $\hat{x}_{test}$，见式(10).

$$\hat{y}_{test}=\frac{\hat{x}_{test}-\mu_r}{\sqrt{\sigma_r^2+\varepsilon}}*\gamma+\beta \tag{10}$$

式中，μ_r，σ_r^2，γ，β 均来自训练阶段统计或优化的结果，测试阶段直接使用，不会进行更新.

2.3 块间直连通道

特征信息在网络层传输过程中，各网络块内信息经多层卷积神经网络的特征提取，易产生特征堆叠，从而使特征模糊化，因此本文分别在 BlockA1 与 BlockA2，BlockB1 与 BlockB2 块间引入 2 个直连通道，A1 处部分原始输入的信息传入 B1 处，B1 处的卷积特征传输至 B3 处，使网络中的 SIF 特征提取层次不同，这样可以在抑制梯度消失的同时加快网络收敛速度. 改进后的网络结构见图 4，其中 2 个直连通道分别跨连接核心卷积块. 模型左侧为声音特征输入端，右侧则为 Softmax 分类的概率. 网络中 A1，B1，C 3 个核心卷积块的结构和参数见图 5. A1，A2 结构类似，包含 2 个卷积层，1 个批量归一化层和 1 个最大池化层；B1，B2，B3 包含 3 个卷积层，1 个最大池化层；C 由 2 个单核卷积层融合全局平均池化层构成.

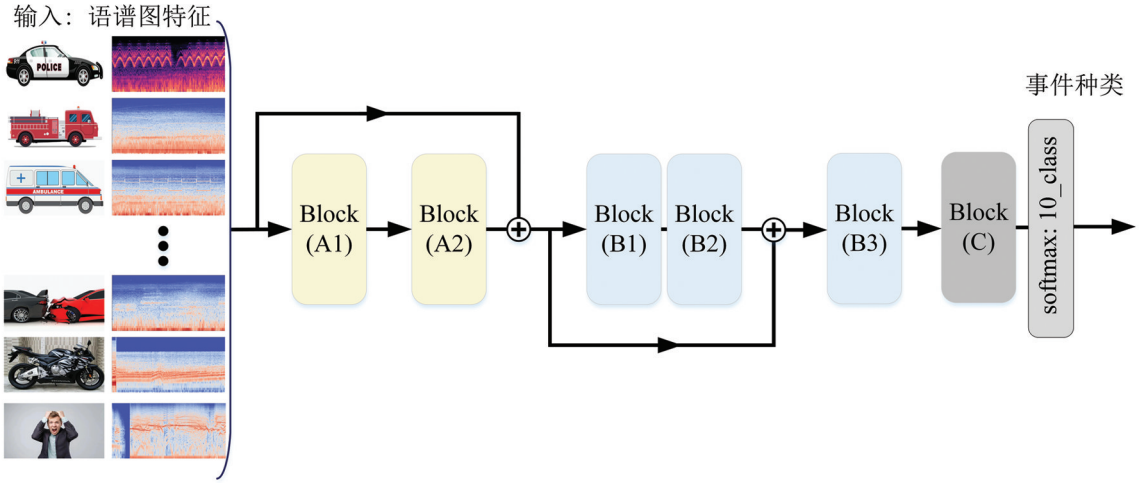


图 4 VGG-TSEC 网络结构

3　实验结果与分析

3.1　建立实验数据集

3.1.1　采集交通声音

交通环境中的声音事件类别复杂，如警笛声、汽车喇叭声、车辆刹车制动声、发动机加速声、事故碰撞声、行人尖叫声等. 鉴于典型声音事件对交通系统影响最大，因此研究选择警车、消防车、救护车警笛声、车辆事故碰撞声、公共汽车、城市警报声、摩托车、倒车提示音、行人尖叫声、卡车共 10 种交通声音作为主要研究对象，场景为交通声学场景. 根据《车用电子警报器》(GB 8108-1999)，设置车用电子警报器的音响频率和重复变调周期，见表 2.

表 2　警笛音频率和周期

音调名称	音响频率/Hz	周期/s	车型
紧急调频调	600^{0}_{-50}～$1\ 500^{+50}_{0}$	0.333～0.385	警车
连续调频调	600^{0}_{-50}～$1\ 500^{+50}_{0}$	3.00～5.00	消防车
双音转换调	f_1：800±50，f_2：1 000±50	1.67～2.50	救护车

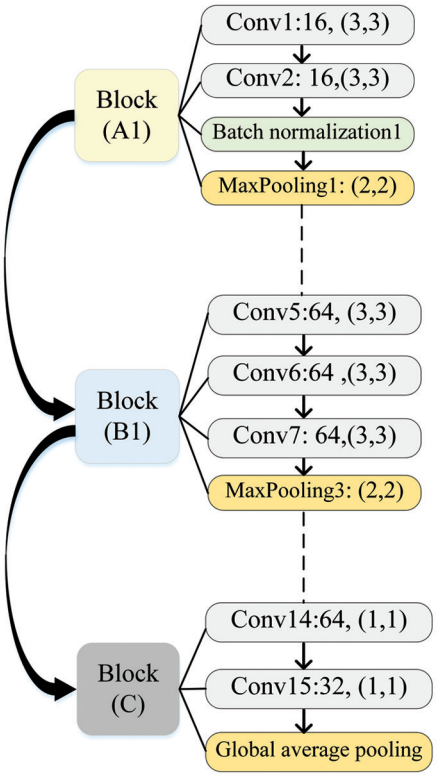
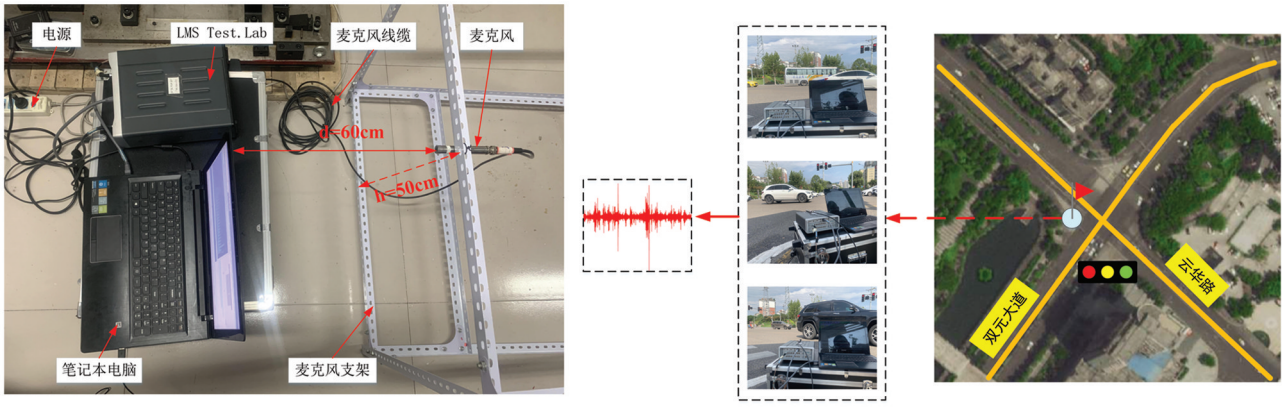


图 5　卷积块结构示意

交通声音采集过程存在不安全、困难度高等问题，因此在保证实验室声音采集有效的情况下，叠加实际路口交通环境背景音，以模拟真实路况下的交通声音. 本文声音事件的采集工作分 2 个阶段进行：第一阶段在实验室内采集交通声音；第二阶段采集交通背景噪音. 首先在实验室使用采集设备采集声音. 交通声音采集示意图见图 6，采集设备为 LMS Test. Lab，Grass 专业级声学麦克风和笔记本电脑等. 实验室采集获得交通声音约 10 h. 第二步采集交通背景噪音，采集地点为重庆市北碚区云华路与双元大道某岔路口（路口常见卡车、轿车、摩托车、行人等）. 采集现场见图 6，获得交通背景噪音约 10 h.



a. LMS Test. Lab采集设备　　b. 交通背景噪音采集

图 6　交通声音采集示意图

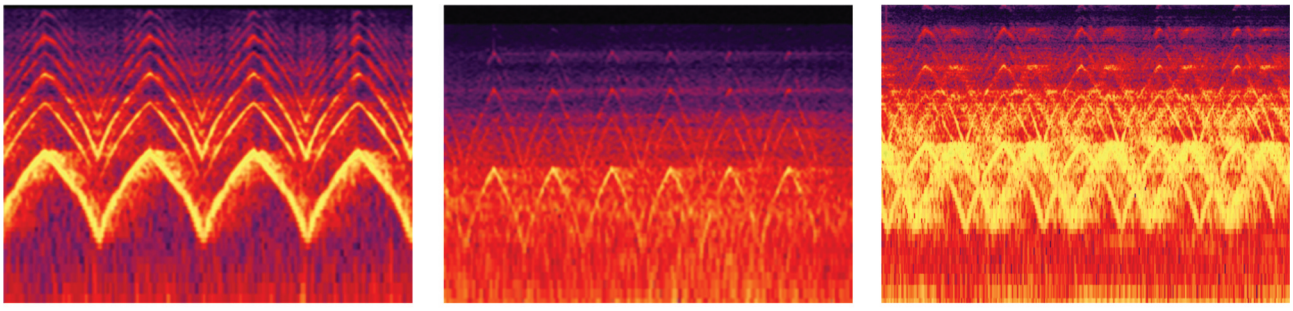
3.1.2　数据集扩增

训练深度学习模型需要大量样本，低样本复杂度极易导致模型过拟合. 选择多普勒频移、声音扭曲、卷积混响、相位改变、延迟等扩增方式，对采集到的交通声音进行数据集扩增. 两通道的声音增益和衰减设置见表 3，警笛音原始声音频谱图见图 7(a)，卷积混响扩增后的频谱图见图 7(b)和图 7(c). 扩增后得到交通声音数据约 45 h. 参考警笛音的周期性，对扩增后的音频声音信号进行切片处理，得到训

练集32 740个，验证集和测试集4 092个，各类交通声音的具体编号、数量等信息见表4和图8.

表3 声音通道增益和衰减

类型	Ch1(声音事件)	Ch2(交通背景音)	类型	Ch1(声音事件)	Ch2(交通背景音)
a	+0dB	+0dB	c	−20dB	+10dB
b	−15dB	+10dB	d	−25dB	+10dB



a b c

图7 原始声音(a),Bitter hall way(b), Corner verbation(c)

表4 各类交通声音的标签信息

标签	声音类型	数量	标签	声音类型	数量
0	救护车	2 980	5	警车	3 050
1	公共汽车	4 680	6	倒车提示音	4 120
2	城市警报声	5 020	7	尖叫声	4 670
3	消防车	2 920	8	卡车	5 810
4	摩托车	4 560	9	车辆碰撞声	3 110

0. 救护车；1. 公共汽车；2. 城市警报声；3. 消防车；4. 摩托车；5. 警车；6. 倒车提示音；7. 尖叫声；8. 卡车；9. 车辆碰撞声.

图8 实验中的交通声音

3.2 实验环境配置和模型评估指标

实验训练与测试计算机的物理环境配置：CPU为Intel(R) Xeon(R) Platinum 8259L，GPU为NVIDIA Quadro GP100，显存16GB，主存192GB；软件环境：Ubuntu操作系统，Tensorflow-gpu 2.5.0深度学习框架，CUDAtoolkit 11.2.0，CUDNN 8.1.0.77，keras 2.5.0，基于python 3.8.12的Pycharm开发环境. 学习率为0.001，最大迭代次数300，选用交叉熵损失函数和Adam模型优化器.

此外，引入准确率(A)、精确率(P)、召回率(R)、F_1分数等指标对试验结果进行评价，具体计算方法见式(11)−(14)：

$$A=\frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

$$P(PPV)=\frac{TP}{TP+FP} \tag{12}$$

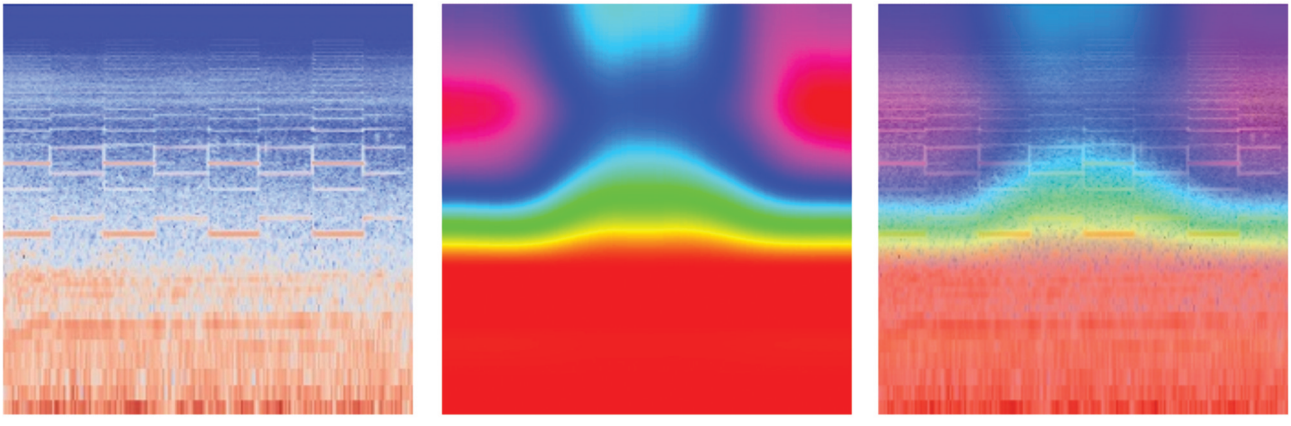
$$R(TPR)=\frac{TP}{TP+FN} \tag{13}$$

$$F_1=\frac{P*R}{P+R} \tag{14}$$

式中，TP 为真阳性；FP 为假阳性；TN 为真阴性；FN 为假阴性.

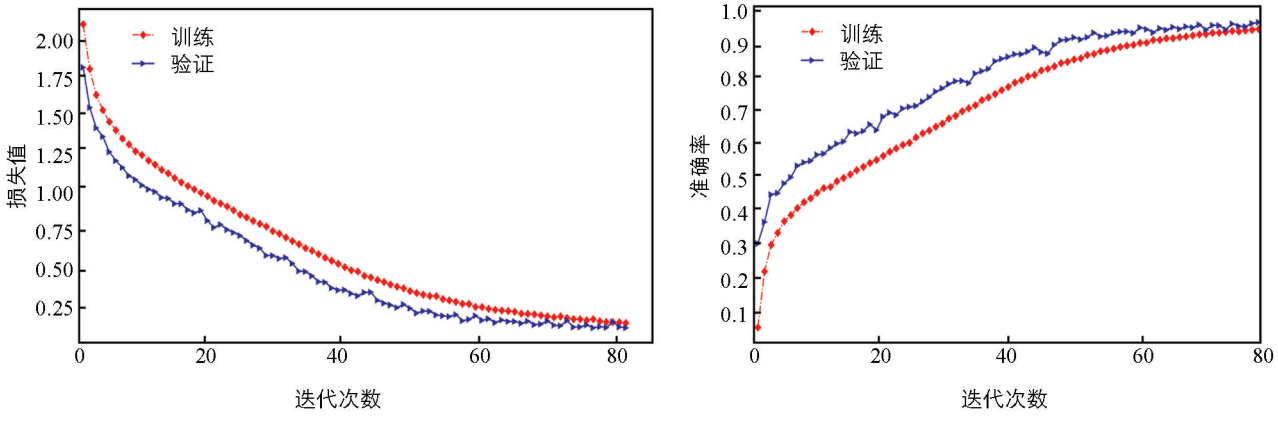
3.3　分类结果及分析

使用交通数据集验证改进后的模型. 通过 CAM 方法获取声音 SIF 热力图，见图 9，从热力图高亮部分可以看出该区域包含模型的重要特征. VGG-TSEC 模型训练的准确率和损失值曲线见图 10，在迭代了 80 次后，模型趋于收敛、重合，此时训练集、验证集的准确率和损失值基本保持不变. VGG-TESC 混淆矩阵见图 11，混淆矩阵高亮对角线代表模型预测正确数，通过该图可以直观地分析模型分类能力，如公共汽车和卡车的分类性能较差，警笛音的分类结果较好. 交通声音验证集结果见表 5，VGG-TSEC 在测试集上的准确率为 97.18%，声音事件分类精确率、召回率、F_1 分数均处于较高水平.



a　　b　　c

图 9　救护车警笛音语谱图(a)，卷积层后的热力图(b),(c)



a. 模型训练集和验证损失值　　b. 模型训练集和验证准确率

图 10　VGG-TESC 训练集和验证集性能表征

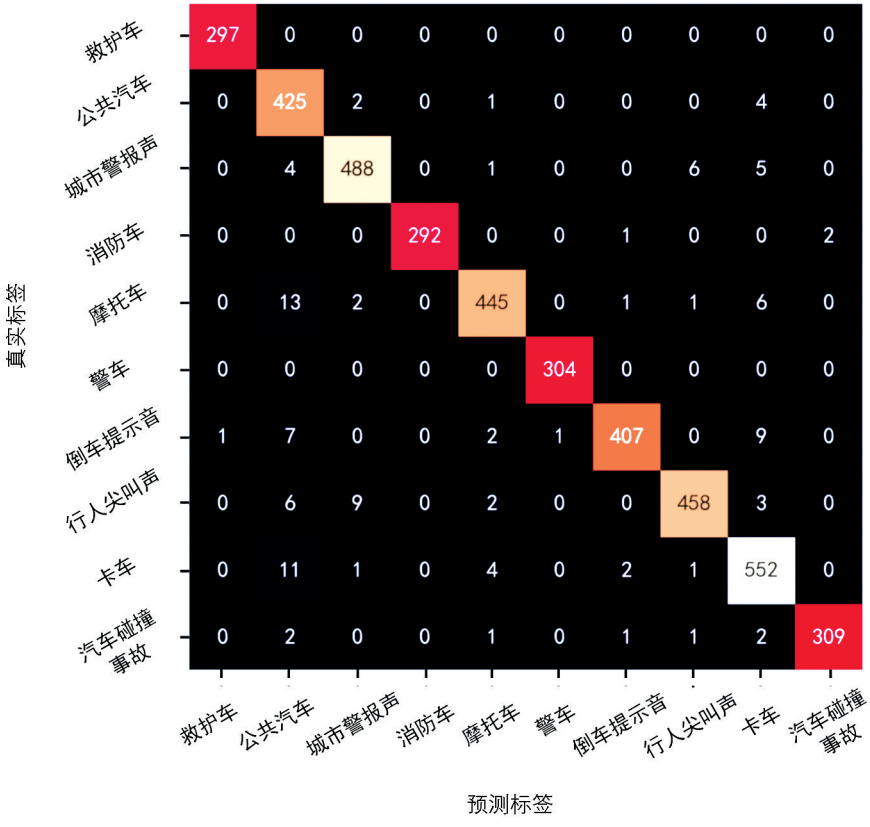


图11 VGG-TESC混淆矩阵

表5 交通声音验证集结果

声音事件	准确率/%	召回率/%	F_1分数/%	数量/次
救护车	99.66	1.00	99.83	297
公共汽车	90.81	98.37	94.44	432
城市警报声	97.21	96.82	97.01	504
消防车	1.00	98.98	99.48	295
摩托车	97.58	95.08	96.32	468
警车	99.67	1.00	99.83	304
倒车提示音	98.78	95.31	97.02	427
行人尖叫声	98.07	95.81	96.93	478
卡车	95.00	96.67	95.83	571
车辆碰撞	99.35	97.78	98.56	316

3.4 与其他模型性能的对比

将LeNet-5，AlexNet，VGG等卷积神经网络用于验证交通声音数据集，结果见表6. 本文提出的VGG-TSEC模型的精确率、召回率和F_1分数相比其他网络有了显著提高. 实验表明，本文提出的VGG-TSEC比原VGG网络分类准确率高4.68%. 各机器学习方法的交通事件分类性能见表7，对比VGG-TESC模型与其他研究结果，传统机器学习方法在交通声音事件分类领域的识别准确率较低，而本文所优化的VGG-TSEC分类准确率比随机森林、K邻近(KNN)和支持向量机(SVM)分别提高了17.05%，22.25%，11.59%.

表 6　各网络声音分类的综合性能

网络	训练速度 ms/step	训练用时/min	占用空间/MB	准确率/%
LeNet-5	7	16.72	0.46	87.90
AlexNet	158	47.27	155.94	77.56
VGG-16	24	92.82	527.79	92.54
ResNet-34	213	327.38	57.21	93.05
VGG-TSEC	22	64.85	7.52	97.18

表 7　各机器学习方法的交通声音事件分类性能

声音特征和分类器	准确率/%
径向基神经网络＋A-计权＋MFCC＋SVM[12]	89.33
PCA＋MFCC＋SVM[20]	70.82
LPC＋MFCC＋SVM[21]	86.25
A-Weighting-Mel filters＋MFCC＋SVM[21]	88.25
多尺度 RBF＋MFCC＋SVM[21]	92.77
HMFCC＋SVM[22]	72.00
径向基神经网络＋HMFCC＋SVM[22]	90.80
KNN[17]	77.50
随机森林[23]	82.70
SVM＋MFCC[24]	66.67
改进小波包变换＋EMD-MFCC＋SVM[24]	87.08
VGG-TSEC＋STFT-Spectrogram	97.18

4　结论

交通声音事件分类旨在识别环境中的声音事件类别，为交通系统提供更多的声音信息. 结论如下：本文针对现有交通系统环境声音感知能力不足、效率低、鲁棒性低、可分类数量少等问题，基于 VGG-16 改进并提出了 VGG-TSEC 交通声音事件分类方法，提高了复杂交通环境下的声音事件分类的准确率，丰富了不同环境背景下的声音事件分类方法.

1) 本文所提出 VGG-TSEC 交通声音事件分类方法的平均准确率达到 97.18%，与 AlexNet，VGG-16，ResNet34 等网络相比，模型性能显著提高.

2) 实验表明，双卷积层融合算法优化后的模型参数量降低了 72.75%，使得网络时空效率均得到了明显提升，为后续移动端的部署奠定基础.

3) 创新性地引入块间直连通道算法，避免深层网络中图形特征堆叠，抑制梯度消失，加快网络收敛速度.

参考文献：

[1] 姚洁，邱劲. 基于 SSA-BP 算法的道路交通流量预测研究 [J]. 西南大学学报(自然科学版)，2022，44(10)：193-201.

[2] LI L Z，OTA K，DONG M X. Humanlike Driving：Empirical Decision-Making System for Autonomous Vehicles [J]. IEEE Transactions on Vehicular Technology，2018，67(8)：6814-6823.

[3] MISHRA S K，DAS S. A Review on Vision Based Control of Autonomous Vehicles Using Artificial Intelligence Techniques [C] //2019 International Conference on Information Technology (ICIT). December 19-21，2019，Bhubaneswar，India. IEEE，2020：500-504.

[4] KUUTTI S, BOWDEN R, JIN Y C, et al. A Survey of Deep Learning Applications to Autonomous Vehicle Control [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(2): 712-733.

[5] GLOAGUEN J R, CAN A, LAGRANGE M, et al. Road Traffic Sound Level Estimation from Realistic Urban Sound Mixtures by Non-Negative Matrix Factorization [J]. Applied Acoustics, 2019, 143: 229-238.

[6] XIA X J, TOGNERI R, SOHEL F, et al. Auxiliary Classifier Generative Adversarial Network with Soft Labels in Imbalanced Acoustic Event Detection [J]. IEEE Transactions on Multimedia, 2019, 21(6): 1359-1371.

[7] VESPERINI F, GABRIELLI L, PRINCIPI E, et al. Polyphonic Sound Event Detection by Using Capsule Neural Networks [J]. IEEE Journal of Selected Topics in Signal Processing, 2019, 13(2): 310-322.

[8] KARPIS O. System for Vehicles Classification and Emergency Vehicles Detection [J]. IFAC Proceedings Volumes, 2012, 45(7): 186-190.

[9] CHOI W, RHO J, HAN D K, et al. Selective Background Adaptation Based Abnormal Acoustic Event Recognition for Audio Surveillance [C] //2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance. September 18-21, 2012, Beijing, China. IEEE, 2012: 118-123.

[10] LI Q, LIU X M, YANG X Y, et al. Abnormal Event Detection Method in Multimedia Sensor Networks [J]. International Journal of Distributed Sensor Networks, 2015, 11(11): 154658.

[11] LEFEBVRE N, CHEN X D, BEAUSEROY P, et al. Traffic Flow Estimation Using Acoustic Signal [J]. Engineering Applications of Artificial Intelligence, 2017, 64: 164-171.

[12] 朱强华, 郑铁然, 韩纪庆. 行车环境下基于二值语谱图的声学事件检测 [C] //第十二届全国人机语音通讯学术会议(NCMMSC2013)论文集. 贵阳, 2013: 377-381.

[13] ZHANG X D, CHEN Y S, TANG G C. Research on Traffic Acoustic Event Detection Algorithm Based on Sparse Autoencoder [J]. MATEC Web of Conferences, 2020, 308: 05002.

[14] YAN C X, LUO M N, LIU H, et al. Top-k Multi-Class SVM Using Multiple Features [J]. Information Sciences, 2018, 432: 479-494.

[15] JIN S P, WANG X F, DU L L, et al. Evaluation and Modeling of Automotive Transmission Whine Noise Quality Based on MFCC and CNN [J]. Applied Acoustics, 2021, 172: 107562.

[16] 黎煊, 赵建, 高云, 等. 基于连续语音识别技术的猪连续咳嗽声识别 [J]. 农业工程学报, 2019, 35(6): 174-180.

[17] WANG X B, YING T, TIAN W. Spectrum Representation Based on STFT [C] //2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). October 17-19, 2020, Chengdu, China. IEEE, 2020: 435-438.

[18] COX R W, TONG R Q. Two- and Three-Dimensional Image Rotation Using the FFT [J]. IEEE Transactions on Image Processing, 1999, 8(9): 1297-1299.

[19] 刘坤华, 钟佩思, 徐东方, 等. 基于双曲正切函数的修正线性单元 [J]. 计算机集成制造系统, 2020, 26(1): 145-151.

[20] 孔鸿运. 行车环境下鲁棒的声学事件检测方法 [D]. 哈尔滨: 哈尔滨工业大学, 2013.

[21] 裴孝中, 郑铁然, 韩纪庆. 行车噪声环境下基于人耳频率选择特性的声学特征提取方法 [J]. 智能计算机与应用, 2015, 5(3): 16-18.

[22] 毛锦, 李林聪, 刘凯, 等. 无人驾驶汽车行车环境下鲁棒性声学特征提取算法 [J]. 中国公路学报, 2019, 32(6): 169-175.

[23] PAL M. Random Forest Classifier for Remote Sensing Classification [J]. International Journal of Remote Sensing, 2005, 26(1): 217-222.

[24] ZHANG M L, ZHOU Z H. ML-KNN: A Lazy Learning Approach to Multi-Label Learning [J]. Pattern Recognition, 2007, 40(7): 2038-2048.

责任编辑 孙文静