DOI:10.16652/j.issn.1004-373x.2018.14.018

基于卷积神经网络的交通声音事件识别方法

张文涛[1,2]，韩莹莹[1,2,3]，黎　恒[3]
（1.桂林电子科技大学 电子工程与自动化学院，广西 桂林　541004；2.光电信息处理重点实验室，广西 桂林　541004；
3.广西交通科学研究院，广西 南宁　530000）

摘　要：针对公路交通声音事件识别中传统语音算法识别效率低、鲁棒性差的问题，提出一种基于卷积神经网络的交通声音事件识别方法。首先通过Gammatone滤波器对声音数字信号进行子带滤波，得到音频信号耳蜗谱图，然后将其代入卷积神经网络模型对声音事件类型进行识别。利用上述方法对公路交通环境下的四种音频事件做了检测处理，并与经典的隐马尔科夫模型和目前广泛使用的深层神经网络进行比较。实验结果表明，使用卷积神经网络模型能够更加准确地对交通声音事件进行识别，且在噪声环境下具有更好的鲁棒性。

关键词：Gammatone滤波器；卷积神经网络；音频事件识别；公路交通环境；声音数字信号；子带滤波
中图分类号：TN912.3-34　　**文献标识码**：A　　**文章编号**：1004-373X(2018)14-0070-04

Traffic sound event recognition method based on convolutional neural network

ZHANG Wentao[1,2], HAN Yingying[1,2,3], LI Heng[3]
(1. School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China;
2. Key Laboratory of Optoelectronic Information Processing, Guilin 541004, China; 3. Guangxi Transportation Research Institute, Nanning 530000, China)

Abstract: In allusion to the problems of low recognition rate and poor robustness of the traditional acoustic algorithm in road traffic sound event recognition, a traffic sound event recognition method based on convolutional neural network is proposed. The sub-band filtering of sound digital signals is performed by using the Gammatone filter, so as to obtain the cochleogram of audio signals, which is then substituted into the convolutional neural network model for recognition of the sound event type. Four audio events in road traffic environment are detected by using the above method, and the results are compared with those of the classic hidden Markov model and deep neural network widely used at present. The experimental results show that the convolutional neural network model can recognize the traffic sound event more accurately, and has better robustness in noisy environment.

Keywords: Gammatone filter; convolutional neural network; audio event recognition; road traffic environment; sound digital signal; sub-band filtering

0　引　言

随着我国交通事业高速发展，对道路监控和信息处理分析提出了更高的要求。目前，国内外道路交通事件检测技术主要以视频为手段，依赖于视频采集的质量，技术难度大，生产成本高且受环境因素影响大。声音是人类信息的重要来源，而且其方便采集，检测范围广。声音事件识别已应用于环境安全监控、场景分析、证据提取、声源定位和突发事件检测等领域，且发挥着重要作用。

目前，对于声音事件识别一些学者已经做出一些研究[1-3]。McLoughlin等提出基于声谱图的前端特征并结合支持向量机(Support Vector Machine，SVM)[1]和深度神经网络(Deep Neural Network，DNN)对声音事件进行分类。Kucukbay等提出使用梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficients，MFCC)[2]和SVM分类器结合5-折交叉验证方法识别办公环境中的16种声音。Diment等提出基于隐马尔科夫模型(Hidden Markov

收稿日期：2017-09-01　　**修回日期**：2017-10-27
基金项目：国家自然科学基金(61565004)；广西自然科学基金(2014GXNSFGA118003)；桂林市科学研究与技术开发课题(20140127-1；20150133-3)
Project Supported by National Natural Science Foundation of China (61565004), Guangxi Natural Science Foundation (2014GXNSFGA118003), Scientific Research and Technology Development Project of Guilin (20140127-1, 20150133-3)

Model，HMM）的声音事件检测系统[3]，对办公环境下的声学场景和事件进行分类检测。

以上研究在声音事件识别中都取得了一定成效。但传统的SVM算法在大规模训练样本和多分类问题上难以实现。DNN模型又存在参数数量膨胀、学习时间过长等问题。且在真实的公路交通环境中存在复杂多变的噪声，公路隧道中甚至出现声音反射和回响，对声音事件识别产生较大干扰。因此针对公路交通环境需要找出一种新的识别方法。Ossama等人首次将卷积神经网络（Convolutional Neural Network，CNN）应用于语音识别中[4]，与DNN模型相比识别率取得明显改善。本文将卷积神经网络应用于公路交通声音事件识别。针对拥堵、事故等公路事件，利用卷积神经网络对车辆碰撞声、汽车鸣笛、人员呼救和车门关闭四种声音进行分类，从而判断发生的对应事件。

1　基于Gammatone滤波器的耳蜗谱图提取

人耳可以准确地对公路交通环境声音事件进行判断。在人类听觉系统中，声音信号经过耳蜗基底膜的频带分解作用后，沿听觉通路进入大脑听觉中枢神经系统。耳蜗谱图特征仿照人耳感知声音的过程，是常用的时频二维图像特征表示方法。本文使用Gammatone滤波器组来模拟耳蜗模型，实现子带滤波，最终得到耳蜗谱图。Gammatone滤波器是一个标准的耳蜗听觉滤波器，其时域脉冲响应为[5]：

$$g(f,t)=At^{n-1}\mathrm{e}^{-2\pi bt}\cos(2\pi ft+\varphi)U(t),t\geqslant 0 \quad (1)$$

式中：A为滤波器增益；i表示第i个滤波器；f为中心频率；φ为相位，本文取$\varphi_i=0$；n为滤波器阶数；b是衰减因子，该因子决定相应的滤波器的带宽$b=1.019\mathrm{ERB}(f)$，$\mathrm{ERB}(f)$是等效矩形带宽，它与中心频率f的关系为：

$$\mathrm{ERB}(f)=24.7(4.37f/100+1) \quad (2)$$

本文采用一组64个4阶Gammatone滤波器，其中心频率在350~4 000 Hz之间。Gammatone滤波器会保留原有的采样率，因此在时间维度上设置响应频率为100 Hz，将产生10 ms的帧移，可用于短时声音特征提取。当声音信号通过Gammatone滤波器时，输出信号的响应$G_m(i)$的表达式如下：

$$G_m(i)=\left[|g|(i,m)]^{1/2},i=0,1,2,\cdots,N;m=0,1,2,\cdots,M-1\right] \quad (3)$$

式中：N为通道数；M为采样后的帧数。

$G_m(i)$构成代表输入声音信号频域上分布变化的矩阵，本文采用耳蜗谱图来描述信号频域分布变化。与语谱图相比，耳蜗谱图的物理意义更明确，具有更高的低频分辨率，因此在声音识别领域更具应用价值[6]。综上所述，本文采用耳蜗谱作为样本进行神经网络模型的训练与测试。

2　卷积神经网络

卷积神经网络[7]最初由Yann LeCun等人提出，应用于简单的手写字符识别，逐渐扩展到人脸检测[8]、动作识别[9]和标志识别[10]等领域。近几年，卷积神经网络作为具有优秀深度学习能力的深层网络结构，被应用于声音识别领域。

卷积神经网络是一种多层神经网络，数据以特征图的形式输入网络，然后依次进行卷积与池化处理，具体过程在相应的卷积层与池化层完成，层与层之间采用局部连接和权值共享的方式。

在卷积层中，输入的特征图被一个可学习的卷积核进行卷积。卷积操作公式如下：

$$x_k^l=f\left(\sum_{i\in W_k}x_i^{l-1}*H_{i,k}^l+b_k^l\right) \quad (4)$$

式中：x_k^l代表l层的第k个特征图；W_k代表$l-1$层的第k个特征图；$H_{i,k}^l$表示第l层第i个特征图的第k个卷积核；b_k^l为偏置项；f是激活函数；“*”代表卷积符号。激活函数一般采用 relu，tanh等饱和非线性函数。所有的输入特征图经卷积操作后输出一定数量的新特征图。新特征图的数量由卷积层中卷积滤波器数目决定。

经卷积层后得到的新特征图进入池化层进行池化操作。一方面使特征图变小，简化网络计算复杂度；另一方面进行特征压缩，提取主要特征。池化层的一般形式如下：

$$x_k^l=f(\beta_k^l\mathrm{down}(x_k^{l-1})+b_k^l) \quad (5)$$

式中：down（·）代表池化层；x_k^l代表l层的第k个特征图；β_k^l与b_k^l为偏置项。如果分割成若干个$a\times a$的区域，那输出图片的尺寸在不同维度上都是输入图片的$1/n$。

3　实验与分析

3.1　实验性能评价指标

根据CLEAR 2007测评[11]，本文使用三个指标评估所提出的方法：准确率（Precision Rate，P）、召回率（Recall Rate，R）和F-值（F-Measure，F）。准确率表明方法的查准率，召回率表明方法的查全率，F-值为准确率和召回率的调和平均值，计算公式分别为：

$$P=\frac{t}{e},\ R=\frac{t}{g},\ F=\frac{2PR}{P+R} \quad (6)$$

式中：t表示正确检测的声音事件数；e表示输出的声音事件总数；g表示标注的声音事件总数。

3.2 实验数据采集

本实验在真实公路交通环境下，使用模拟声级计配合麦克风阵列，分别在20 dB，10 dB，0 dB三种信噪比下对音频数据进行采集，采样频率为8 kHz。表1中总结了每种事件类别的统计数据，共有86 400段，每种声音片段长度为1~3 s。

表1 各类音频事件数量

Tab. 1 Numbers of various audio events

样本	20 dB		10 dB		0 dB	
	训练	测试	训练	测试	训练	测试
撞车声	3 600	1 200	3 600	1 200	3 600	1 200
汽车鸣笛声	3 600	1 200	3 600	1 200	3 600	1 200
人员呼救声	3 600	1 200	3 600	1 200	3 600	1 200
车门关闭声	3 600	1 200	3 600	1 200	3 600	1 200
打雷声	3 600	1 200	3 600	1 200	3 600	1 200
汽车发动机	3 600	1 200	3 600	1 200	3 600	1 200

按照第1节中提到的方法提取耳蜗谱。抽取每种声音耳蜗谱中的3/4作为训练集，剩下的1/4为测试集。并对每种声音的种类进行标注。本文将撞车声、汽车鸣笛、人员呼救和车门关闭四种音频信号作为目标声音事件，因为这些声音事件的出现一般意味着发生交通事故或拥堵。其余两种声音事件作为干扰声。

3.3 卷积神经网络模型建立

为了研究基于卷积神经的交通声音事件识别模型性能，首先需要建立卷积神经网络。卷积神经网络结构确定的过程包括建立模型、训练模型和测试模型三部分。实验使用Matlab的Parallel Computing Toolbox工具箱和 Neural Network Toolbox工具箱创建和训练卷积神经网络。基于Pascal GP104核心的NVIDIA GTX1080搭建训练平台，使用GPU阵列进行计算。

图1所示为基于卷积神经网络的声音事件识别流程图，包括训练过程与测试过程。训练过程中，利用随机分布函数对卷积核和权重进行随机初始化，而对偏置进行全0初始化。为了加快训练过程则使用标准的梯度下降算法调整权值与阈值。

通过网络前向传播和反向传播反复交叉处理的方式来训练卷积神经网络，直到代价函数小于0.01为止。

3.4 基于CNN的交通声音事件识别方法

本文针对公路交通环境下声音信号的特殊性，选取网络结构如图2所示，包含2个卷积层、2个池化层、2个归一化层和3个全连接层。

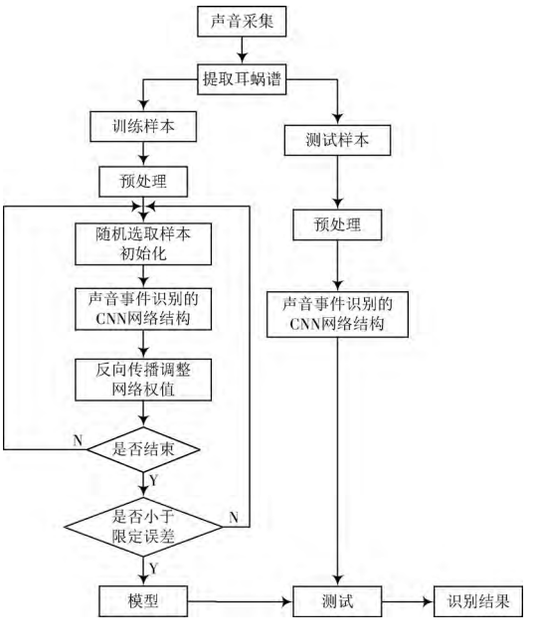

图1 声音事件识别流程图

Fig. 1 Flow chart of sound event recognition

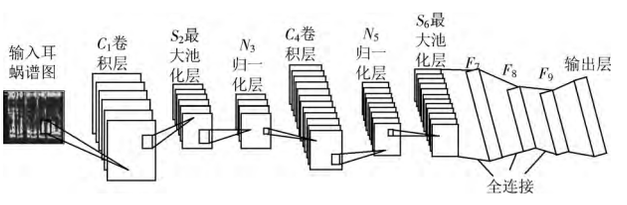

图2 CNN网络结构

Fig. 2 Network structure of CNN

1）输入层。本文将耳蜗谱图作为特征图输入。输入之前先对耳蜗谱图进行预处理，预处理过程包括平滑和裁剪，把耳蜗谱图处理成32×32的图像。

2）卷积层。卷积层为特征提取层。均使用3×3的卷积核对输入图像进行卷积，C_1卷积核个数为10，卷积移动步长为1，为保证充分提取耳蜗谱图的特征，激活函数使用tanh函数。

3）池化层。卷积层后加入池化层可降低体征维数，避免过拟合。本文采用最大池化方式，池化域大小均为2×2。

4）归一化层。在下一个卷积层前加入归一化层，对局部做减和除归一化，迫使相邻特征进行局部竞争。保证性能的稳定性，并提取耳蜗谱的低阶和高阶统计特征。

5）输出层。通过Softmax回归算法将特征映射到目标的四个分类。Softmax是Logistic回归在多分类问题上的推广。在Softmax回归函数中$y=j$的概率为：

$$p(y=j|x;\theta)=\frac{e^{\theta_j^T x}}{\sum_{l=1}^{k} e^{\theta_j^T x}} \tag{7}$$

式中：θ表示全部的模型参数；x为输入；y为输出；j代表类别。

3.5　实验结果对比与分析

将20 dB，10 dB，0 dB三种不同信噪比下的实验样本分别代入第3.4节确定CNN模型进行训练与测试实验，并与经典HMM模型[12]和目前得到广泛应用的DNN模型进行对比。实验结果如表2~表4所示。

表2　20 dB下对比实验结果

Tab. 2　Results of comparison experiment in 20 dB　%

样本	CNN			DNN			HMM		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
车辆碰撞	99.3	100.0	99.6	95.2	98.3	96.7	74.5	81.9	78.0
汽车鸣笛	98.6	99.8	99.2	93.7	97.8	95.7	72.9	80.8	76.6
人员呼救	98.0	99.4	98.7	94.4	98.2	96.2	73.6	81.5	77.3
车门关闭	96.4	99.2	97.8	94.0	97.1	95.5	71.8	80.2	75.7
平均	98.1	99.6	98.8	94.3	97.9	96.1	73.2	81.1	76.9

表3　10 dB下对比实验结果

Tab. 3　Results of comparison experiment in 10 dB　%

样本	CNN			DNN			HMM		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
车辆碰撞	96.7	99.2	97.9	92.9	97.0	94.9	43.6	43.7	43.6
汽车鸣笛	95.2	98.4	96.8	91.8	95.6	93.7	41.6	41.4	41.5
人员呼救	95.7	99.3	97.4	91.3	95.8	93.5	43.3	43.3	43.3
车门关闭	95.3	98.0	96.6	91.0	94.9	92.9	41.2	40.8	41.0
平均	95.7	98.7	97.2	91.8	95.8	93.8	42.4	42.3	42.4

表4　0 dB下对比实验结果

Tab. 4　Results of comparison experiment in 0 dB　%

样本	CNN			DNN			HMM		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
车辆碰撞	90.1	96.7	93.3	84.9	91.3	88.0	17.6	10.8	13.4
汽车鸣笛	86.9	93.3	90.0	83.3	90.0	86.5	16.1	9.8	12.2
人员呼救	86.4	92.9	89.5	84.2	90.2	87.1	16.4	10.0	12.4
车门关闭	84.1	90.4	87.1	83.6	89.1	86.2	16.5	9.9	12.4
平均	86.0	92.4	89.1	84.0	90.1	86.9	16.7	10.1	12.6

从3个表中可以看出，在3种不同信噪比情况下，对于车辆碰撞声、汽车鸣笛、人员呼救和车门关闭四种声音识别，CNN模型与DNN模型的指标均明显高于HMM模型，且CNN模型的识别率可达到99.3%，召回率可达100%。信噪比发生变化时，CNN模型的平均F-值相比其他两种模型所受影响最小。在0 dB的情况下，识别率突破90%。由此可以得出，相比于其他两种模型，卷积神经网络模型可以更加准确地对公路交通环境下的声音事件进行识别且鲁棒性更好。

4　结　论

本文将卷积神经网络应用到公路交通环境声音识别中。先将声音信号经Gammatone滤波器转化为耳蜗谱图，后把耳蜗谱图输入卷积神经网络进行分类识别。并与经典隐马尔科夫模型和广泛使用的深层神经网络进行了对比，基于卷积神经网络的方法在识别性与鲁棒性上有明显提高。在后续研究中，将继续优化卷积神经网络结构，进一步对混合声音事件进行识别。

参　考　文　献

[1] MCLOUGHLIN I, ZHANG H, XIE Z, et al. Robust sound event classification using deep neural networks [J]. IEEE/ACM transactions on audio, speech, and language processing, 2015, 23(3): 540-552.

[2] KUCUKBAY S E, SERT M. Audio-based event detection in office live environments using optimized MFCC-SVM approach [C]// Proceedings of IEEE International Conference on Semantic Computing. Anaheim: IEEE, 2015: 475-480.

[3] DIMENT A, HEITTOLA T, VIRTANEN T. Sound event detection for office live and office synthetic AASP challenge [J/OL]. [2013-12-01]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.706.807&rep=rep1&type=pdf.

[4] ABDEL-HAMID O, MOHAMED A, JIANG H, et al. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition [C]// Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Kyoto: IEEE, 2012: 4277-4280.

[5] LU B, DIBAZAR A, BERGER T W. Noise-robust acoustic signature recognition using nonlinear Hebbian learning [J]. Neural networks, 2010, 23(10): 1252-1263.

[6] TJANDRA A, SAKTI S, NEUBIG G, et al. Combination of two-dimensional cochleogram and spectrogram features for deep learning-based ASR [C]// Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Brisbane: IEEE, 2015: 4525-4529.

[7] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural computation, 1989, 1(4): 541-551.

[8] 汪济民，陆建峰.基于卷积神经网络的人脸性别识别[J].现代电子技术，2015，38(7)：81-84.
WANG Jimin, LU Jianfeng. Face gender recognition based on convolutional neural network [J]. Modern electronics technique, 2015, 38(7): 81-84.

（下转第78页）

3 结 论

在宽范围高精度测温系统中，针对K型热电偶输出的热电势E与温度t之间的非线性关系，建立PID神经网络对其进行非线性校正，采用粒子群算法进行寻优取值。选取E-t曲线中特殊拐点的标准值作为训练样本，对寻优过程中的惯性权值改进，增强其全局寻优能力。仿真结果表明：控制系统稳定性较高，精确性提高；训练所得温度值与标准值之间误差均在1%以内，且二者拟合效果好，收敛速度快。精确温度信号获取为火箭发射塔架的健康评估提供可靠依据。

参 考 文 献

[1] 段艳明.基于PSO算法和BP神经网络的PID控制研究[J].计算机技术与发展，2014，24(8)：238-241.
DUAN Yanming. Reserch of PID control based on BP neural network and PSO algorithm [J]. Computer technology and development, 2014, 24(8): 238-241.

[2] 朴海国，王志新.基于CPSO的PID神经网络及偏航电机控制策略[J].电机与控制学报，2010，14(9)：55-62.
PIAO Haiguo, WANG Zhixin. Control strategy of CPSO-based PID neural network and a yaw motor [J]. Electric machines and control, 2010, 14(9): 55-62.

[3] GAO M Y, CHEN S X, CHENG L L, et al. Online measurement of battery internal resistance based on AC impedance method [J]. Advanced materials research, 2013, 718-720: 773-778.

[4] 屈毅，宁铎，赖展翅，等.温室温度控制系统的神经网络PID控制[J].农业工程学报，2011，27(2)：307-311.
QU Yi, NING Duo, LAI Zhanchi, et al. Neural networks based on PID control for greenhouse temperature [J]. Transactions of the Chinese Society of Agricultural Engineering, 2011, 27(2): 307-311.

[5] 于立君，陈佳，刘繁明，等.改进粒子群算法的PID神经网络解耦控制[J].智能系统学报，2015，10(5)：699-704.
YU Lijun, CHEN Jia, LIU Fanming, et al. An improved particle swarm optimization for PID neural network decoupling control [J]. CAAI transactions on intelligent systems, 2015, 10(5): 699-704.

[6] MOHANDES M A. Modeling global solar radiation using particle swarm optimization [J]. Solar energy, 2012, 86(11): 3137-3145.

[7] 沈锡.基于粒子群优化算法的船舶航向PID控制[D].大连：大连海事大学，2011.
SHEN Xi. Ship course PID control based on particle swarm optimization [D]. Dalian: Dalian Maritime University, 2011.

[8] 周西峰，林莹莹，郭前岗.基于粒子群算法的PID神经网络解耦控制[J].计算机技术与发展，2013，23(9)：158-161.
ZHOU Xifeng, LIN Yingying, GUO Qiangang. PID neural network decoupling control based on particle swarm optimization [J]. Computer technology and development, 2013, 23(9): 158-161.

[9] 应进.基于粒子群算法的航空发动机多变量控制研究[D].南昌：南昌航空大学，2011.
YING Jin. Research on multi-variable control of aero engine based on particle swarm optimization [D]. Nanchang: Nanchang Hangkong University, 2011.

[10] 俞凯耀，席东民.人工鱼群算法优化的PID神经网络解耦控制[J].计算机仿真，2014，31(10)：350-353.
YU Kaiyao, Xi Dongmin. Optimized PID neural network decoupling control based on artificial fish optimization [J]. Computer simulation, 2014, 31(10): 350-353.

作者简介：苏淑靖(1971—)，女，副教授，硕士生导师。主要研究方向为感知与探测、信号处理。
吕楠楠(1992—)，女，硕士研究生。主要研究方向为数据处理、电路与系统。
翟成瑞(1964—)，男，教授。主要从事测试计量技术研究。

(上接第73页)

[9] JI S, XU W, YANG M, et al. 3D convolutional neural networks for human action recognition [J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(1): 221-231.

[10] 黄琳，张尤赛.应用深层卷积神经网络的交通标志识别[J].现代电子技术，2015，38(13)：101-106.
HUANG Lin, ZHANG Yousai. Traffic signs recognition applying with deep-layer convolution neural network [J]. Modern electronics technique, 2015, 38(13): 101-106.

[11] TEMKO A, NADEU C, MACHO D, et al. Acoustic event detection and classification [M]// WAIBEL A, STIEFELHAGEN R. Computers in the human interaction loop. Berlin: Springer, 2009: 61-73.

[12] TEMKO A, MALKIN R, ZIEGER C, et al. CLEAR evaluation of acoustic event detection and classification systems [C]// Proceedings of the 1st international evaluation conference on classification of events, activities and relationships. Berlin: Springer, 2006: 311-322.

作者简介：张文涛(1976—)，男，山东济南人，博士，博士生导师，教授。研究方向为纳米计量及激光技术。
韩棠棠(1992—)，女，河北保定人，硕士研究生。研究方向为声音识别、机器学习。
黎 恒(1982—)，男，广西南宁人，博士。研究方向为视频信号处理、机器学习。