

---

# 声学所建立基于小波谱图和深度卷积网络的音频场景识别新框架

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/3207.html>

*本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！*

声学所建立基于小波谱图和深度卷积网络的音频场景识别新框架。近年来，网络上音频的种类和数量呈现爆发式增长。相比语音和音乐，音频中的环境信息更加多样，受到了国内外研究者的广泛关注。

音频场景分类的目的是识别出录制音频的特定场景，从而使穿戴式设备、智能机器人感知周围的环境信息并做出相应反射。传统的音频分类任务大都使用短时特征，如梅尔频率倒谱系数等。然而，环境信息一般隐藏在不同时长的背景声中，因此提取多分辨率多尺度的特征是音频场景分类的关键。

近期，中国科学院声学研究所语言声学与内容理解重点实验室博士生陈航艇及其导师张鹏远等人提出了一种基于小波变换和深度卷积神经网络的场景识别新方法。相关研究成果已被Proceedings of the Annual Conference of the International Speech Communication Association(2018)收录。

研究人员提出的音频场景分类系统包括前端和后端两个模块。前端采用不同时长和频率分布的小波滤波器来提取小波谱图，此滤波器组可以同时捕捉到转瞬即逝的音频事件和长时的韵律节奏。后端的深度卷积神经网络结合小卷积核和池化操作，将“高维语义”从原始特征中抽象出来。

在相关场景数据集上的实验结果表明，使用小波谱图和深度卷积神经网络的音频场景识别表现明显优于传统的短时特征。相比短时特征，长时的小波谱图的时间分辨率较低，一定程度上降低了存储和计算的要求。

该研究为音频场景的分析提供了新的前后端架构，为环境声音的研究提供了新思路。

参考文献:CHEN Hangting, ZHANG Pengyuan, BAI Haichuan, YUAN Qingsheng, BAO Xiuguo, YAN Yonghong. Deep Conventional Neural Network with Scalogram for Audio Scene Modeling. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, v 2018-September, p 3304-3308. DOI: 10.21437/Interspeech.2018-1524.

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发